Sparse Modeling and Compressive Sensing

Ivan Ralašić
University of Zagreb, Faculty Of Electrical Engineering and Computing
ivan.ralasic@fer.hr

***Abstract*—Sparse modeling and compressive sensing are novel methods for signal representation and acquisition. Sparse signal representations are manifestation of the parsimony principle also known as the Occam's razor which states that the simplest and most concise explanation of a natural phenomenon is in most cases the best one possible. Sparse structure appears to be an inherent property of many natural signals when observed in an appropriate basis. Compressive sensing represents a signal acquisition technique that exploits underlying sparse signal structure and enables accurate signal recovery from an incomplete set of measurements. In this overview, basis representation and sparse decomposition fundamentals and sparse recovery problem formulations are covered and bridged with different practical applications of compressive sensing framework. The most recent advances in the compressive sensing theory and the state-of-the-art applications are presented.**

***Index Terms*—sparse modeling, compressive sensing, basis representation, signal processing, signal acquisition, optimization.**

I. Introduction

FUNDAMENTAL problem in sparse modeling and compressive sensing is to obtain accurate recovery of an unobserved high-dimensional signal from a reduced number of measurements. Traditional approach to signal acquisition is based on the classical Shannon-Nyquist theorem [1], [2] stating that in order to preserve information about a signal, one must sample the signal at a rate which is at least twice the signal's bandwidth, defined as the highest frequency in the signal's spectrum. Compressive sensing defines conditions for successful signal recovery based on informational content in contrast to the traditional signal bandwidth viewpoint. Traditional sample-and-compress framework is efficient and used in many real-world applications, however, the fact that we are able to compress the acquired data suggests that Shannon-Nyquist theorem is too pessimistic and it does not take advantage of any specific underlying structure that the signal may possess. In practice, the traditional sampling scheme produces tremendous number of samples and it must be followed by a compression step in order to successfully store, process and transmit obtained information. The compression step uses different basis representations to obtain concise signal representation and it essentially keeps only the significant basis coefficients while disregarding the others. The described procedure is known as transform coding and it takes advantage of underlying sparse signal structure. A natural question arises, can we combine the acquisition and compression into one-step process that will enable us to make signal acquisition more efficient.

Candés, Romberg and Tao considered signal reconstruction from an incomplete frequency samples in [3] and presented a new kind of nonlinear sampling theorem. They state that exact signal recovery from an incomplete set of measurements is possible by solving a convex optimization problem under certain constraints. Donoho [4] defined compressive sensing on an example of arbitrary unknown vector $\boldsymbol{x} \in \mathbf{R}^N$ (digital image or signal) where $\boldsymbol{x}$ is known to be compressible by transform coding in a certain basis. He stated that accurate signal reconstruction can be obtained using a nonlinear reconstruction procedure on a reduced number of measurements m which can be dramatically smaller than the signal dimensionality n. This led to considerable research interest in the signal processing community and numerous papers followed [5]–[8].

Compressive sensing gained more popularity with the emergence of first practical applications. Compressed sensing MRI study by Lustig et al. [9] reviewed the requirements for successful compressive sensing reconstruction and described their natural fit to MRI. CS-MRI offered significant scan time reductions with benefits for patients and health care economics. In [10], authors presented a new approach for building simpler, smaller, and cheaper digital cameras that can operate efficiently across a broader spectral range than conventional cameras and established a natural branch of CS named compressive imaging (CI). Their approach fuses a new camera architecture based on a digital micro-mirror device (DMD) with the new compressive sensing mathematical framework and instead of measuring pixel samples of the scene, they measure inner products between the scene and a set of random test functions. This leads to sub-Nyquist image acquisition that enables one to stably reconstruct an image from fewer measurements than the number of reconstructed pixels. Tropp et al. [11] proposed a new type of sampling system called a random demodulator that can be used to acquire sparse, band-limited signals. The major advantage of the random demodulator is that it bypasses the need for a high-rate ADC since demodulation is much easier to implement than high-rate sampling. Above mentioned are just some of the numerous applications of compressive sensing and we will give an extended overview later in the text.

In this paper, we will provide an overview of basis representation fundamentals, sparse signal decomposition, sparse recovery problem formulations and different optimization procedures which lead to a definition of complete compressive sensing framework.

Paper is organized as follows: section II represents an introduction to sparse signal representations where basis representation and sparse recovery fundamentals are covered, section III is an introduction to compressive sensing theory, section IV introduces overcomplete dictionaries and dictionary learning methods, section IV gives an overview of sparse recovery algorithms used in compressive sensing, and finally, section V bridges the theory presented in previous sections with practical applications of compressive sensing.

II. Sparse Signal Modeling

A. Basis Representation Fundamentals

To introduce the notion of sparsity, we rely on a basis decomposition that results in a low-dimensional representation of the observed signal. Every signal $\boldsymbol{x} \in \mathbf{R}^N$ is representable in terms of N coefficients $\{s_i\}_{i=1}^N$ in a given basis $\{\psi_i\}_{i=1}^N$ for $\mathbf{R}^N$ as:

$$\boldsymbol{x} = \sum_{i=1}^{N} \psi_i s_i \tag{1}$$

Arranging the ψ_i as columns into the $N \times N$ matrix $\boldsymbol{\Psi}$ and the coefficients s_i into the $N \times 1$ coefficient vector $\boldsymbol{s}$, we can write that $\boldsymbol{x} = \boldsymbol{\Psi s}$, with $\boldsymbol{s} \in \mathbf{R}^N$. We say that signal $\boldsymbol{x}$ is K-sparse in the basis $\boldsymbol{\Psi}$ if there exists a vector $\boldsymbol{s} \in \mathbf{R}^N$ with only $K \ll N$ nonzero entries such that $\boldsymbol{x} = \boldsymbol{\Psi s}$. Additionally, by a compressible representation, we mean that the coefficient's magnitudes, when sorted, have a fast power-law decay. Many natural signals are sparse or compressible when observed in an appropriate basis. If we use a frame $\boldsymbol{\Psi}$ containing N unit-norm column vectors of length L with $L < N$ (i.e., $\boldsymbol{\Psi} \in \mathbf{R}^{L \times N}$), then for any vector $\boldsymbol{x} \in \mathbf{R}^L$ there exist infinitely many decompositions $\boldsymbol{s} \in \mathbf{R}^N$ such that $\boldsymbol{x} = \boldsymbol{\Psi s}$. In a general setting, $\boldsymbol{\Psi}$ is called an overcomplete sparsifying dictionary [12].

B. Motivating Example

As an illustrative example, we can consider the case where our overcomplete dictionary is union of two particular orthobases: the identity (spike) basis and the Fourier (sine) basis $\boldsymbol{\Psi} = [\boldsymbol{I} \quad \boldsymbol{F}]$ (see Fig. 1) where $\boldsymbol{I}$ is $N \times N$ identity matrix and $\boldsymbol{F}$ is $N \times N$ normalized discrete Fourier matrix with entries defined with:

$$\boldsymbol{F}(m, l) = \frac{1}{\sqrt{N}} e^{j2\pi(m-1)(l-1)/N} \tag{2}$$

Identity and Fourier basis are mutually fully incoherent in the sense that it takes N spikes to build up a single sinusoid and also it takes N sinusoids to build up a single spike. Now we can create a signal which is a mixture of spikes and

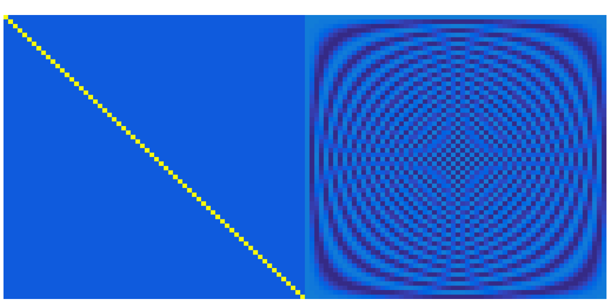

Fig. 1. Overcomplete dictionary created by concatenation of identity (*left*) and Fourier (*right*) basis.

sinusoids. As we know that the first half of our matrix $\boldsymbol{\Psi}$ contains spike functions and the second half corresponds to sine functions, we can construct random sparsity pattern with sparsity K for the vector $\boldsymbol{s}$ (of size $2N$) such that some of the non-zero entries fall into the first half and some in the second half of the vector, and then compute $\boldsymbol{x} = \boldsymbol{\Psi s}$ to obtain a signal which is a mixture of impulses and sinusoids (see Fig. 2a).

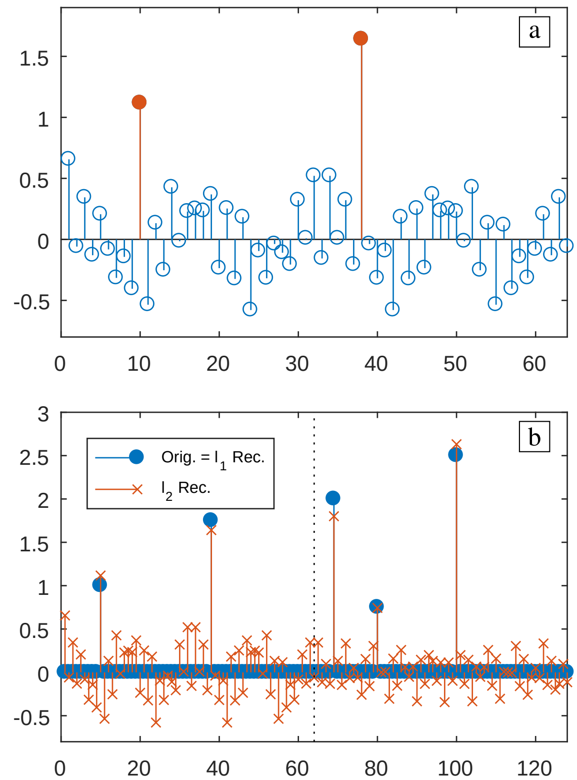

Fig. 2. Example: a) arbitrary signal which consists of a mixture of two spikes (*red*) and three sines of different frequency (*blue*), b) spectrum of the original sparse signal (*blue*) and minimum energy decomposition using l_2-norm. Notice that the spectrum obtained using l_2-norm is not sparse. The vertical dashed line separates spike (*left*) and sine (*right*) part of the spectrum.

We can apply $\boldsymbol{s} = \frac{1}{2}\boldsymbol{\Psi}^*\boldsymbol{x}$ and get a basis representation that corresponds to the minimum energy decomposition of our signal into a coefficient vector $\boldsymbol{s}$ that represents signal $\boldsymbol{x}$. Minimum energy decomposition corresponds to l_2-norm minimization. Unfortunately, minimum energy decomposition almost never yields the sparsest possible solution. The reason for this is that a vector has minimum energy when its total energy is distributed over all the coefficients of the vector. l_2-norm minimization gives us a solution that is dense, and has small values diffused over all coefficients (see Fig. 2b).

In our example, we synthetically produced a signal with only K coefficient representation in overcomplete basis $\boldsymbol{\Psi}$ and we want to find a decomposition that yields a signal decomposition that is K sparse. Since the goal of finding the sparsest possible representation of signal $\boldsymbol{x}$ over some basis $\boldsymbol{\Psi}$ is equivalent to finding the solution with the smallest number of nonzero elements in the basis coefficient vector $\boldsymbol{s}$, we will use l_0-pseudo-norm to find the solution. Sparse signal recovery can be formulated as finding minimum-cardinality solution to a constrained optimization problem.

C. Sparse Signal Recovery

We will use the following notation in this section to formalize the sparse signal recovery: $\boldsymbol{x} = (x_1, ..., x_N) \in \mathbf{R}^N$ is an unobserved sparse signal, $\boldsymbol{y} = (y_1, ..., y_M) \in \mathbf{R}^M$ is a vector of measurements (observations), and $\boldsymbol{A} = \{a_{i,j}\} \in \mathbf{R}^{M \times N}$ is a design matrix.

The simplest problem we are going to start with is the noiseless signal recovery from a set of linear measurements, i.e., solving for $\boldsymbol{x}$ the system of linear equations:

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \tag{3}$$

It is usually assumed that $\boldsymbol{A}$ is a full-rank matrix, and thus for any $\boldsymbol{y} \in \mathbf{R}^M$, the above system of linear equations has a unique solution. Note that, when the number of unknown variables, i.e., dimensionality of the signal, exceeds the number of observations ($M < N$), the above system is underdetermined, and has infinitely many solutions. In order to recover the signal $\boldsymbol{x}$, it is necessary to further constrain, or regularize the problem. This is usually done by introducing an objective function, or regularizer $R(\boldsymbol{x})$ to an existing loss function. Regularizer encodes additional properties of the signal, with lower values corresponding to a more desirable solution. Signal recovery is then formulated as a constrained optimization problem:

$$\min_{\boldsymbol{x} \in \mathbf{R}^N} R(\boldsymbol{x}) \quad s.t. \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \tag{4}$$

Since we want to exploit underlying sparse structure of the observed signal, $R(\boldsymbol{x})$ can be defined as the number of nonzero elements, or the cardinality of $\boldsymbol{x}$, also called the l_0-norm. In general, l_q-norms for particular values of q, denoted $\|\boldsymbol{x}\|_q$, or more precisely their q-th power $\|\boldsymbol{x}\|_q^q$, are frequently used as regularizers $R(\boldsymbol{x})$ in constrained optimization problems.

Using the cardinality function (l_0-norm), we can now rewrite the problem of sparse signal recovery from noiseless linear measurements as:

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_0 \quad s.t. \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \tag{5}$$

The above optimization problem is NP-hard and no known algorithm can solve it efficiently in polynomial time. Therefore, approximations have to be introduced. Under appropriate conditions the optimal solution can be recovered efficiently by certain approximate techniques.

First approach to approximation is a heuristic-based search such as greedy search. In the greedy search method, one can start with a zero vector and keep adding nonzero coefficients one by one, selecting at each step the coefficient that leads to the best improvement in the objective function. In general, such heuristic search methods are not guaranteed to find the global optimum. However, in practice, they are simple to implement, computationally efficient and under certain conditions they are even guaranteed to recover the optimal solution.

An alternative approximation technique is the relaxation approach based on replacing an intractable objective function or constraint by a tractable one. In other words, one can either solve the exact problem approximately using greedy methods, or solve an approximate problem exactly using relaxations of the original problem. In the following section, we will discuss l_q-norm based relaxations, and show that the l_1-norm occupies a unique position among them, combining convexity with sparsity. Convex problems are easier to solve than general optimization problems because of the important property that any local minima of a convex function is also a global one [13].

D. Convex Relaxations of Sparse Recovery Problem

We will now focus on different l_q-norms as possible relaxations of the l_0-norm. These functions are convex for $q \geq 1$ and non-convex for $q < 1$. For example, l_2-norm (Euclidean norm) would be a natural first choice as a relaxation of l_0-norm because of its convexity and the existence of closed form solution. However, we already showed that solution obtained using l_2-norm reconstruction in not sparse.

From a geometric point of view, solving the sparse recovery problem is equivalent to "blowing up" l_q-balls with the center at the origin, i.e., increasing their radius, starting from 0, until they touch the hyperplane $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$. The resulting point is the minimum l_q-norm vector that is also a feasible point, i.e. it is the optimal solution of the sparse recovery problem (see Fig. 3b).

Note that when $q \leq 1$, l_q-balls have sharp "corners" on the coordinate axis, corresponding to sparse vectors, since some of their coordinates are zero, but l_q-balls for $q > 1$ do not have this property. Thus, for $q \leq 1$, l_q-balls are likely to meet the hyperplane $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ at the corners, thus producing sparse solutions, while for $q > 1$ the intersection practically never occurs at the axes, and thus solutions are not sparse (see Fig. 3a).

Within the family of l_q-norms, only those with $q \geq 1$ are convex, but only those with $0 < q \leq 1$ are sparsity-enforcing. The only function within that family that has both useful properties is therefore $\|\boldsymbol{x}\|_1$, i.e. the l_1-norm [13]. This unique combination of sparsity and convexity is the reason for the widespread use of l_1-norms in the modern sparse signal recovery field. Optimization problem using l_1-norm now becomes:

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \quad s.t. \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \tag{6}$$

III. Compressive Sensing Theoretical Framework

Compressive sensing (CS) offers a theoretical framework for simultaneous sensing and compression of finite-dimensional vectors, that relies on dimensionality reduction. In compressive sensing, the signal is not measured via standard point samples but rather through the projection by a measurement matrix $\boldsymbol{\Phi}$:

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} = \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{s} = \boldsymbol{A}\boldsymbol{s} \tag{7}$$

where $\boldsymbol{\Phi}$ is an $M \times N$ measurement matrix and $\boldsymbol{y} \in \mathbf{R}^M$ is a set of M measurements or samples where M can be much smaller than the original dimensionality of the signal, hence the name compressive sensing. We introduce $\boldsymbol{A} = \boldsymbol{\Phi}\boldsymbol{\Psi}$ and refer to it as the design matrix for compressive sensing (4). The central problem of compressive sensing is reconstruction

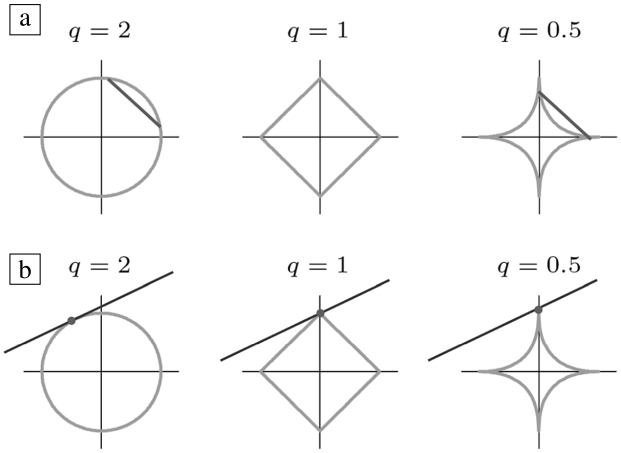

Fig. 3. a) Among the family of l_q-norms, only l_1-norm has both necessary properties for successful recovery of sparse signals, i.e. it induces sparsity and is convex. b) Optimization of sparse recovery problem as inflation of the origin-centered l_q-balls until they meet the set of feasible points $\boldsymbol{Ax} = \boldsymbol{y}$. [13]

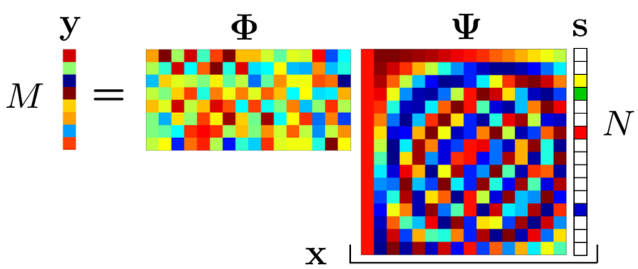

Fig. 4. Matrix form representation of compressive sensing problem. Measurement vector is obtained through projection of an unknown signal $\boldsymbol{s}$ that is sparse in some basis $\boldsymbol{\Psi}$ by a measurement matrix $\boldsymbol{\Phi}$. [14]

of high-dimensional sparse signal representation $\boldsymbol{x}$ from a low-dimensional linear observation $\boldsymbol{y}$, also called the measurement vector.

Ideally, the measurement matrix $\boldsymbol{\Phi}$ is designed to reduce the number of measurements M as much as possible while allowing for recovery of a wide class of signals and providing non-adaptive sampling scheme. However, the fact that $M < N$ renders the matrix rank-deficient, meaning that it has a nonempty null-space, which in turn implies that for any particular signal $\boldsymbol{x}_0 \in \mathbf{R}$, an infinite number of signals yields the same measurements $\boldsymbol{y}_0 = \boldsymbol{\Phi x}_0 = \boldsymbol{\Phi x}$ for chosen measurement matrix $\boldsymbol{\Phi}$. The motivation behind the design of the matrix $\boldsymbol{\Phi}$ is to allow for distinct signals ($\boldsymbol{x}$, $\boldsymbol{x'}$) within a class of signals of interest to be uniquely identifiable using sparse optimization techniques from their measurements ($\boldsymbol{y}$, $\boldsymbol{y'}$), even though $M \ll N$ [12].

A. Uniqueness of Compressive Sensing Recovery Problem

In this section we will discuss the solution uniqueness for the l_0- and l_1- norm minimization problems. The main design criterion for matrix $\boldsymbol{A}$ is to enable the unique identification of a signal of interest $\boldsymbol{x}$ from its measurements $\boldsymbol{y} = \boldsymbol{Ax}$. Clearly, when we consider the class of K-sparse signals Σ_K, the number of measurements has to be $M > K$ for any matrix design, since the identification problem has K unknowns.

We will now determine properties of $\boldsymbol{A}$ that guarantee that distinct signals $\boldsymbol{x}, \boldsymbol{x'} \in \Sigma_K$, $\boldsymbol{x} \neq \boldsymbol{x'}$, lead to different measurement vectors $\boldsymbol{Ax} \neq \boldsymbol{Ax'}$. In other words, we want each vector $\boldsymbol{y} \in \mathbf{R}^M$ to be matched to at most one vector $\boldsymbol{x} \in \Sigma_K$ such that $\boldsymbol{y} = \boldsymbol{Ax}$.

B. Sensing Matrix Properties

1) Spark: A key relevant property of the matrix in this context is its spark [15]. Given an $M \times N$ matrix $\boldsymbol{A}$, its *spark*, is defined as the minimal number of linearly dependent columns. Spark is closely related to the Kruskal's rank (*krank*) defined as the maximal number k such that every subset of k columns of the matrix $\boldsymbol{A}$ is linearly independent [16]. We can now write the relation between *spark* and *krank* as:

$$spark(\boldsymbol{A}) = krank(\boldsymbol{A}) + 1 \quad \text{and} \quad rank(\boldsymbol{A}) \geq krank(\boldsymbol{A}) \tag{8}$$

By definition, the vectors in the null-space of the matrix $\boldsymbol{Ax} = \boldsymbol{0}$ must satisfy $\|\boldsymbol{x}\|_0 \geq spark(\boldsymbol{A})$, since these vectors linearly combine columns from $\boldsymbol{A}$ to give the zero vector, and at least *spark* such columns are necessary by definition. Sparse recovery solution uniqueness via spark can be stated as: if $spark(\boldsymbol{A}) > 2K$, then for each measurement vector $\boldsymbol{y} \in \mathbf{R}^M$ there exists at most one signal $\boldsymbol{x} \in \Sigma_K$ such that $\boldsymbol{y} = \boldsymbol{Ax}$ [13]. The singleton bound yields that the highest spark of an matrix $\boldsymbol{A} \in \mathbf{R}^{M \times N}$ with $M < N$ is less than or equal to $M + 1$ and using the before stated theorems we get the requirement $M \geq 2K$.

While *spark* is useful notion for proving the exact recovery of a sparse optimization problem, it is NP-hard to compute since one must verify that all sets of columns of a certain size are linearly independent. Thus, it is preferable to use properties of $\boldsymbol{A}$ which are easily computable to provide recovery guarantees.

2) Coherence: The coherence $\mu(\boldsymbol{A})$ of a matrix is the largest absolute inner product between any two columns of matrix $\boldsymbol{A}$:

$$\mu(\boldsymbol{A}) = \max_{1 \leq i \neq j \leq N} \frac{\langle a_i, a_j \rangle}{\|a_i\|_2 \|a_j\|_2} \tag{9}$$

For any matrix $\boldsymbol{A}$,

$$spark(\boldsymbol{A}) \geq 1 + \frac{1}{\mu(\boldsymbol{A})} \tag{10}$$

It can be shown that $\mu(\boldsymbol{A}) \in \left[\sqrt{\frac{N-M}{M(N-1)}}, 1\right]$. The lower bound is known as the Welch bound. Note that when $N >> M$, the lower bound is approximately $\mu(\boldsymbol{A}) \geq \frac{1}{\sqrt{M}}$ [15]. In our example with overcomplete dictionary with spikes and sines basis, the coherence exactly corresponds to the Welch bound. That confirms our statement that spikes and sines are mutually completely incoherent. In [17], a method to design sensing matrices with minimum coherence to a given sparsifying orthogonal basis was proposed. They provided a mathematical proof of the optimality in terms of coherence minimization for the proposed sensing matrices.

3) Restricted Isometry Property: The prior properties of the CS design matrix provide guarantees of uniqueness when the measurement vector $\boldsymbol{y}$ is obtained without error. There can be two sources of error in the measurements: inaccuracies due to noise at sensing stage (in the form of additive noise $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + noise$) and inaccuracies due to mismatches between the design matrix used during recovery and that implemented during acquisition (in the form of multiplicative noise $\boldsymbol{A'} = \boldsymbol{A} + \boldsymbol{A}_{noise}$). Under these sources of error, it is no longer possible to guarantee uniqueness, but it is desirable for the measurement process to be tolerant to both types of error. To be more formal, we would like the distance between the measurement vectors for two sparse signals $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ and $\boldsymbol{y'} = \boldsymbol{A}\boldsymbol{x'}$ to be proportional to the distance between the original signal vectors $\boldsymbol{x}$ and $\boldsymbol{x'}$. Such a property allows us to guarantee that for small enough noise, two sparse vectors that are far apart from each other cannot lead to the same noisy measurement vector. This behavior has been formalized into the restricted isometry property (*RIP*) [18]–[21].

A matrix $M \times N$ $\boldsymbol{A}$ is said to satisfy the (K, δ)-restricted isometry property ((K, δ)-RIP) if, for all $\boldsymbol{x} \in \Sigma_K$:

$$(1-\delta)\|\boldsymbol{x}\|_2^2 \leq \|\boldsymbol{\Phi}\boldsymbol{x}\|_2^2 \leq (1+\delta)\|\boldsymbol{x}\|_2^2 \tag{11}$$

where Σ_K represents the set of all length-N vectors with K non-zero coefficients and $\delta \in (0, 1)$. In words, the (K, δ)-RIP ensures that all sub-matrices of $\boldsymbol{A}$ of size $M \times K$ are close to an isometry, and therefore are distance-preserving. This property suffices to prove that the recovery is stable to presence of additive noise and the RIP also leads to stability with respect to the multiplicative noise.

The definition of the RIP is closely related to the frame theory and Riesz bases that were developed in [22] to provide theoretical framework for band-limited signal reconstruction from irregularly spaced samples. RIP is consistent with the thought of spreading energy behind random sensing matrices so measurement $\boldsymbol{y}$ does not shrink or expand too much comparing to the original signal $\boldsymbol{x}$. RIP can be viewed as a modified frame theory:

$$A\|\boldsymbol{x}\|^2 \leq \|\boldsymbol{\Phi}\boldsymbol{x}\|^2 \leq B\|\boldsymbol{x}\|^2 \tag{12}$$

where the frame bounds are $0 < A \leq B \leq 2$.

In [23], authors provided a simple compressive sensing framework which applies all the standard compressive sensing models with addition of some new ones. They established a probabilistic and RIP-less CS theory that states that the RIP property is not necessarily needed to accurately recover nearly sparse vectors from noisy compressive measurements. This represents a significant advance with application to many real world situations where RIP may be hard to check or even might not hold.

C. Asymptotic Structure in Compressive Sensing

CS theory in the previous sections presents a traditional view on the compressive sensing. In recent years, novel compressive sensing theory was introduced by Roman, Adcock and Hansen in [24]–[27]. They introduce new pillars of compressive sensing, namely *asymptotic incoherence*, *asymptotic sparsity* and *multilevel sampling* which replace the traditional notion of sparsity (there are K important coefficients in the vector to be recovered and their location is arbitrary), incoherence (values in the measurement matrix should be dense and uniformly spread out) and uniform random subsampling. Traditional compressive sensing states that the signal sampling strategy is completely independent of the location of the nonzero coefficients. In their flip test, Adcock et al. show that sparsity alone does not dictate the compressive sensing reconstruction quality and that the optimal sampling strategy must indeed depend on the signal structure.

Fig. 5. Flip test performed on *Barb* image. Left image shows recovery from direct wavelet coefficients while right image shows reconstruction using flipped coefficients. Measurement matrix used in the flip test is Walsh-Hadamard matrix, and the transformation basis is Haar's wavelet. Flipped coefficients recovery is substantially worse than the direct coefficient recovery.

Assume that $\boldsymbol{x}$ is an image vector, $\boldsymbol{\Phi}$ is some measurement matrix and $\boldsymbol{\Psi}$ is a transformation basis matrix. We can define a new matrix $\boldsymbol{A} = \boldsymbol{R}_\Omega\boldsymbol{\Phi}\boldsymbol{\Psi}$, where $\boldsymbol{R_\Omega}$ is reduction matrix corresponding to some subsampling pattern $\Omega \subseteq \{1, ..., N\}$ with $|\Omega| = M$, and let our measurement vector be $\boldsymbol{y} = \boldsymbol{R}_\Omega\boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{s} = \boldsymbol{A}\boldsymbol{s}$. We can now solve standard sparse optimization problem in the form of Eq. 6 and obtain sparse representation $\boldsymbol{s}$ of the image. Assuming that sparsity is the correct model as traditional CS theory states, we should be able to flip the reconstructed coefficients to obtain $\boldsymbol{s}_f$ since $\|\boldsymbol{s}\|_0 = \|\boldsymbol{s}_f\|_0$. Next, we can define new measurement vector using the flipped coefficients $\boldsymbol{y'} = \boldsymbol{A}\boldsymbol{s}_f$ and again solve the sparse optimization problem using the new measurement vector $\boldsymbol{y'}$ and get the estimation of the sparse vector $\boldsymbol{s}'_f$ which we flip again to get $\boldsymbol{s'}$. The image estimation $\boldsymbol{x'}$ is then calculated as $\boldsymbol{x'} = \boldsymbol{\Psi}\boldsymbol{s'}$. If sparsity is the precise model, estimation using the direct coefficient recovery should be equal to flipped coefficients recovery $\boldsymbol{x} = \boldsymbol{x'}$ which is not the case, as shown in Fig. 7. Notice that this implies that the standard notion of sparsity defined by traditional RIP property does not hold.

D. Structure in Measurement Matrix

In the previous sections, we assumed use of randomized compressive sensing matrices whose entries are obtained independently from a standard probability distribution. While the proposed matrix choice satisfies the conditions imposed on the CS measurement matrix, it has numerous disadvantages when it comes to real-world applications. One disadvantage is the cost of multiplying arbitrary matrices with high-dimensional signals. Furthermore, often the physics of the sensing modality and the capabilities of the measurement device limit the freedom in choosing measurement matrices.

To overcome these problems, structured measurement matrices were proposed and [28] provides an extensive overview. There are numerous ways of introducing structure into the measurement matrix design. One idea is to simply select a basis that is incoherent to the sparsity basis $\boldsymbol{\Psi}$, and obtain CS measurements by selecting a subset of the coefficients of the signal in the chosen basis [6]. Use of subsampled circulant matrices as CS measurement matrices was inspired by applications in channel estimation and multiuser detection [29]–[31]. Furthermore, separable matrices provide computationally efficient alternatives to measure high-dimensional data. These matrices use Kronecker product [32] to create sparsity basis which simultaneously exploits the sparsity properties of a multidimensional signal along each of its dimensions [33], [34].

IV. Overcomplete Dictionaries and Dictionary Learning

Sparsifying basis that leads to sparse representations of signals can either be chosen as a pre-specified set of basis functions, or designed by adapting the basis functions to fit a given set of signal examples. We refer to the first approach as analytic approach in which a mathematical model is formulated to efficiently represent the data. This procedure leads to implicit dictionaries which have efficient implementations, but lack adaptability. Some examples of implicit dictionaries are: Fourier basis, wavelets [35], curvelets [36], contourlets [37], shearlets [38] and bandelets [39].

The second approach suggests using machine learning techniques to infer the dictionary from a set of examples [40]–[43]. In this case, the dictionary is typically represented as an explicit matrix, and a training algorithm is employed to adapt the matrix coefficients to the examples and the final goal is to find the dictionary $\boldsymbol{D}$ that yields the sparsest representation of the training signals.

The ultimate sparse-coding objective is to find both $\boldsymbol{D}$ and $\boldsymbol{X}$ that yield the sparsest representation of the data $\boldsymbol{Y}$, subject to some acceptable approximation error ϵ:

$$\min_{\boldsymbol{D},\boldsymbol{X}} \sum_{i=1}^{N} \|\boldsymbol{x}_i\|_1 \quad s.t. \quad \|\boldsymbol{Y} - \boldsymbol{D}\boldsymbol{X}\|_2 \leq \epsilon \tag{13}$$

Note that this problem formulation looks very similar to the classical sparse signal recovery problem, only with two modifications: dictionary $\boldsymbol{D}$ is now included as an unknown variable that we must optimize over, and there are multiple observed samples and the corresponding sparse signals rather than just one. Several different dictionary learning algorithms have been introduced over the past years: method of optimal directions (MOD) [44], K-SVD [45] and online-dictionary learning [46] are the most popular ones. First two methods are batch methods which attempt to learn a dictionary directly from the full data set which makes them computationally inefficient. Third method belongs to a family of online dictionary learning methods where learning is done incrementally, one training sample (or a batch of samples) at a time, which makes them scale better in applications where a very large number of training examples exists.

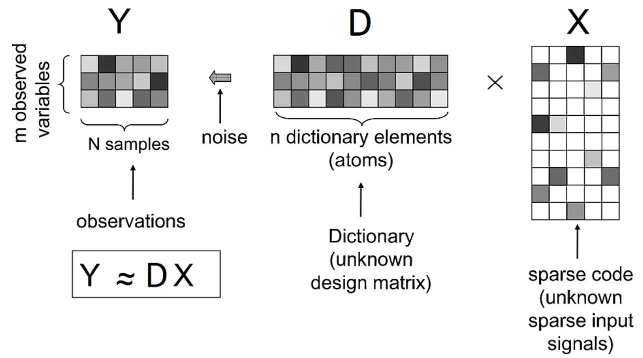

Fig. 6. Dictionary learning matrix form representation [13].

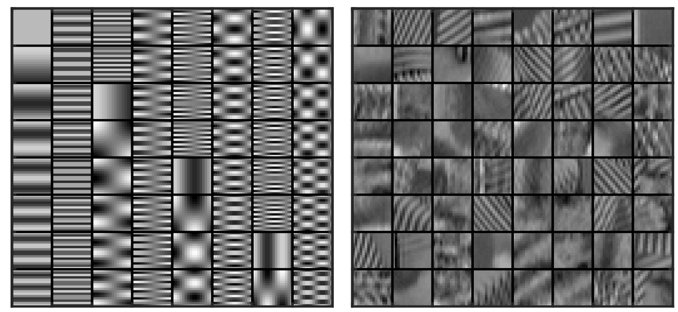

Fig. 7. Dictionary learning performed on *Barb* image. Left image shows part of initial overcomplete DCT dictionary and right image shows part of the learned dictionary using online dictionary learning algorithm. Notice that the high-frequency content that is image specific is captured in the learned dictionary.

In [47], an efficient and flexible dictionary structure is proposed which bridges the gap between implicit dictionaries, which have efficient implementations yet lack adaptability, and explicit dictionaries, which are fully adaptable but costly in terms of computational resources. The proposed sparse dictionary is based on a sparsity model of the dictionary atoms over a base dictionary, and takes the form $\boldsymbol{D} = \boldsymbol{\Psi}\boldsymbol{A}$ where $\boldsymbol{\Psi}$ is a fixed base dictionary and $\boldsymbol{A}$ is sparse. The sparse dictionary model suggests that each atom of the dictionary has itself a sparse representation over some prespecified base dictionary $\boldsymbol{\Psi}$ (see Fig. 8).

V. Overview of Algorithms for Sparse Recovery

In this section we will provide a brief overview of algorithms for sparse recovery. We can write the sparse recovery problem in the case of noise as a convex l_1-norm relaxation of the standard l_0-norm minimization:

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \quad s.t. \quad \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2 \leq \epsilon \tag{14}$$

Furthermore, l_1-norm minimization can be written in alternative unconstrained form as:

$$\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_1 \tag{15}$$

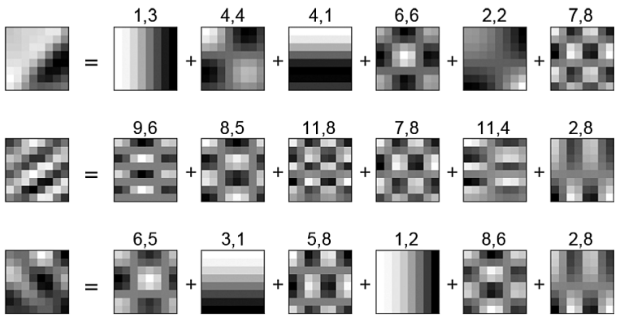

Fig. 8. Some atoms of sparse dictionary trained over an overcomplete DCT dictionary using Sparse K-SVD algorithm. Dictionary atoms are represented using 6 coefficients each [47].

or, for an appropriate parameter $t(\epsilon)$, denoted as t, the same problem can be rewritten as:

$$\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 \quad s.t. \quad \|x\|_1 \leq t \tag{16}$$

The above l_1 regularized problem in the latter two forms is known as the *LASSO* (Least Absolute Shrinkage and Selection Operator) [48].

A comprehensive overview of sparse recovery algorithms is given in [49]. They categorize sparse optimization algorithms into four groups: greedy strategy approximation, constrained optimization, proximity algorithm-based optimization and homotopy algorithm-based sparse optimization.

In the greedy strategy approximation for solving sparse representation problem, the target task is mainly to solve the sparse representation method with l_0-norm minimization. Because of the fact that this problem is an NP-hard problem , the greedy strategy provides an approximate solution to alleviate this difficulty. The greedy strategy searches for the best local optimal solution in each iteration with the goal of achieving the optimal holistic solution. Some of the most commonly-used greedy methods are: matching pursuit [50], orthogonal matching pursuit (OMP) [51], stage-wise OMP (StOMP) [52], compressive sampling matching pursuit (CoSaMP) [53] and several others.

As mentioned before, in the constrained optimization strategy, the core idea is to explore a suitable way to transform a non-differentiable optimization problem into a differentiable optimization problem by replacing the l_1-norm minimization term, which is convex but non-smooth, with a differentiable optimization term, which is convex and smooth. Some of constrained optimization strategy based algorithms are: gradient projection sparse reconstruction (GPSR) [54], least angle regression for *LASSO* (LARS) [55], truncated Newton based interior-point method (TNIPM) [56], alternating direction method (ADM) [57] and others.

In the proximity algorithm-based optimization strategy for sparse representation, the main task is to reformulate the original problem into the specific model of the corresponding proximal operator such as the soft thresholding operator, hard thresholding operator, and resolvent operator, and then exploits the proximity algorithms to address the original sparse optimization problem. Iterative shrinkage thresholding algorithm (ISTA) [58], fast iterative shrinkage thresholding algorithm (FISTA) [59], sparse reconstruction by separable approximation (SpaRSA) [60] and generalized Nesterov's algorithm (NESTA) [61] are examples of proximity algorithm-based optimization methods.

The main idea of homotopy algorithm-based sparse representation is to solve the original optimization problems by tracing a continuous optimization problems by tracing a continuous parameterized path of solutions along with varying parameters. In contrast to LARS and OMP, the homotopy method is more favorable for sequentially updating the sparse solution by adding or removing elements from the active set. Some of the most representative homotopy based reconstruction methods are *LASSO* homotopy and basis pursuit denoising (BPDN) [49].

In [62], authors adapt the kernel trick usually used in machine learning applications to easily adapt linear algorithms to nonlinear situations. They show how the kernel trick can be used to adapt the traditional paradigm of reconstructing a linearly sparse signal from a linear set of measurements to the case of reconstructing a nonlinearly sparse signal from either nonlinear or linear measurements. The key idea is that a signal that is nonlinearly sparse can, with a proper choice of kernel, become linearly sparse in feature space. Then, reconstruction can be performed on random measurements in feature space, which can be obtained from the usual random measurements for some kernels.

Recently, machine learning methods for learning data representations started to gain more interest in the area of compressive sensing. A non-iterative and extremely fast algorithms based on deep neural networks have recently been proposed in [63]–[65]. These papers adapt different designs of deep neural network for application in compressive sensing reconstruction of images. In [63], authors propose a novel convolutional neural network (CNN) architecture inspired by the work of Dong et al. [66] which takes CS measurements of an image as input and outputs an intermediate image reconstruction. They showed that the images reconstructed using convolutional network outperform state-of-the-art iterative CS reconstruction algorithms under certain conditions. Furthermore, to demonstrate the performance of their *ReconNet* non-iterative reconstruction algorithm they presented a proof of concept real-time visual tracking system that can recover semantically informative images even at very low measurement rates. The convolutional neural network (*ReconNet*) design is shown in Fig. 9. Signal recovery framework proposed in [64] learns the inverse transformation from the measurement vectors $\boldsymbol{y}$ to signals $\boldsymbol{x}$ using a special deep network architecture. When trained on a set of representative images, the network learns both a representation for the signals $\boldsymbol{x}$ and an inverse map approximating a greedy or convex recovery algorithm. Finally, Yao et al. propose a novel deep learning sparse recovery algorithm based on state-of-the-art deep residual networks in [67].

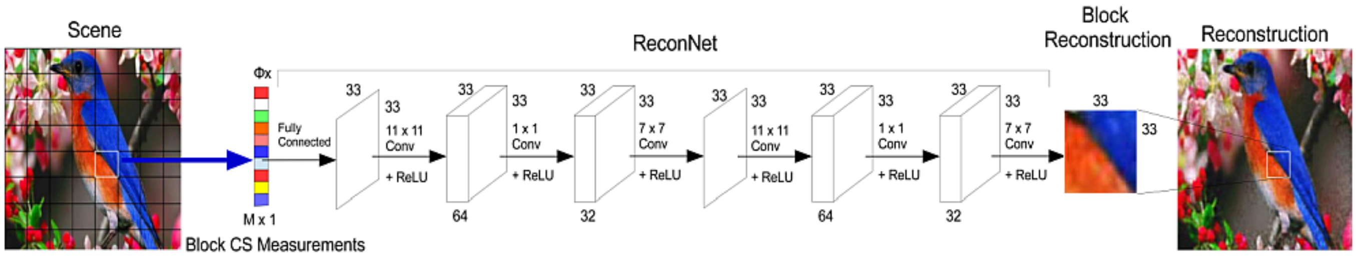

Fig. 9. Overview of deep neural network design for non-iterative block CS image recovery algorithm. The scene is divided into non-overlapping blocks. Each block is reconstructed by feeding in the corresponding CS measurements to the deep network. The reconstructed blocks are then arranged appropriately to form the image reconstruction. The input to the network is an M-dimensional vector of compressive measurements, denoted by $\boldsymbol{\Phi x}$, where $\boldsymbol{\Phi}$ is the measurement operator of size $M \times M$ where M is the number of measurements and $\boldsymbol{x}$ is the vectorized image block [63].

VI. Applications of Sparse Representations and Compressive Sensing

The impact that compressive sensing and sparse signal representations had on the signal processing community is immense. There are numerous applications trying to apply the theoretical CS framework to real world problems. Still, there is a gap between the theoretical knowledge and practical implementations. In this section, we will provide an overview of existing practical applications of sparse representations and compressive sensing.

Signal compression has been a topic of interest for many years thanks to the fact that most of the natural signals are compressible in some transformation basis, thus, sparse modeling seems to be a very natural approach to signal compression. Conventional transform coding methods use predefined basis dictionaries to obtain concise representations. While predefined dictionaries are computationally efficient, they are typically limited in the ability to sparsify the observed signals. On the other hand, the proposed dictionary learning methods enable discovery of underlying structure in signals of interest and tend to provide us with sparser representations than conventional methods. In [68]–[70] different overcomplete dictionaries are learned and applied to image compression problems and outperform conventional methods. Other applications of dictionary learning include image denoising [71], image inpainting [72], [73] and image super-resolution [74]–[76].

The previously mentioned applications use dictionary learning for purely reconstructive tasks and we call such learning methods unsupervised learning. The final goal of reconstructive dictionary learning is to minimize the residual of reconstructing the original signals. Sparse signal representations have another natural property of being discriminative in such a way that the obtained signal representation is sparsest possible. This lead to development of different classification methods based on dictionary learning where classification is based on the corresponding reconstruction error [77], [78] with applications on face recognition [79] (see Fig. 10) and other classification problems.

As already mentioned in the introduction, compressive sensing found its first application in MRI medical imaging because of the existing natural connection. MRI obeys two key requirements for successful application of CS: medical

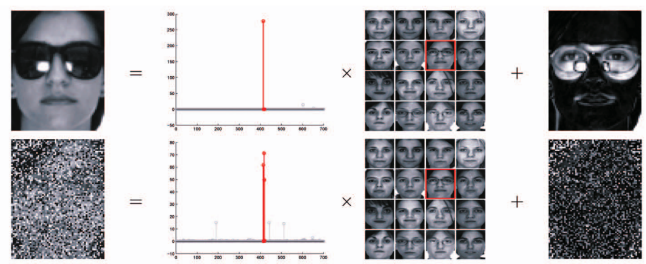

Fig. 10. Robust face recognition using sparse coding reconstruction error. Proposed method represents a test image as a sparse linear combination of all the training images and classifies test image based on the obtained reconstruction error. [79]

images are naturally compressible by sparse coding in appropriate transform domain and MRI scanners acquire encoded samples, rather than direct pixel samples. Compressive sensing improved acquisition speed and spatial resolution of MRI scans and novel applications like rapid 3D angiography, whole-heart coronary imaging and brain imaging emerged. In [80], Geethanath et al. give a detailed overview of applications of CS to MRI.

In compressive imaging the central idea is to design encoded, subsampled linear measurements which allow for accurate optimization-based decoding which assumes sparsity in some transform domain. Applications of compressive imaging include fluorescence microscopy, infrared microscopy, single-pixel imaging, light-field reconstructions and several others. Different designs for imaging devices based on compressive sensing have been proposed and each design tackles certain limitation of traditional imaging systems. For example, in [81], authors propose P2C2 (Programmable Pixel Compressive Camera) camera which consists of a normal 25fps, low resolution video camera in combination with a high resolution, high frame-rate modulating device (DMD or LCoS). This design offers inexpensive and efficient acquisition and recovery of high-resolution, high-speed frames from the captured low-res frames by exploiting temporal redundancy in the form of brightness constancy and spatial redundancy through transform domain sparsity in a convex optimization framework. Similar design was proposed in [82] by Gupta et al., where their goal was to build video cameras whose spatio-

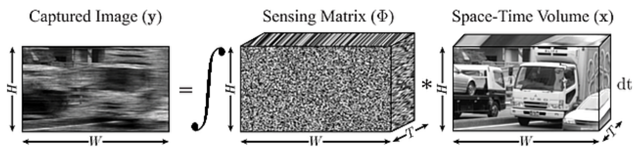

Fig. 11. Temporal compressive sensing model. A space-time volume $\boldsymbol{x}$ consisting of a set of T frames, with $H \times W$ pixels each is multiplied by a set of mask patterns embedded within the sensing matrix $\boldsymbol{\Phi}$. The sensor integrates over time, producing a single coded captured image $\boldsymbol{y}$ consisting of $H \times W$ pixels. [83]

temporal resolution can be changed post-capture depending on the scene (independently at each image location) eliminating the drawback of traditional cameras which offer a fixed spatio-temporal resolution. Proposed sampling scheme requires a fast per-pixel shutter on the sensor-array, which was implemented using a co-located camera-projector system.

In [83], a prototype compressive video camera that encodes scene movement using a translated binary photo-mask in the optical path was presented. The encoded recording can be used to reconstruct multiple output frames from each captured image, effectively synthesizing high-speed video (see Fig. 11). Similarly, coded aperture compressive temporal imaging (CACTI) device proposed in [84], uses translating coded aperture during exposure to modulate each temporal plane in the video stream by a shifted version of the code, attaining per-pixel modulation using no additional sensor bandwidth. This process can be viewed as a form of code division multiple access (CDMA). The temporal channels are isolated from the compressed data by inverting a highly underdetermined system of equations and using an iterative reconstruction algorithm, several high-speed video frames are obtained from a single coded measurement. CACTI device was used in [85] to estimate a sequence of video frames and to reconstruct an extended depth of field image from a single two-dimensional coded measurement.

Light field imagers such as the plenoptic and integral imagers measure projections of 4D light field scalar function onto a two dimensional sensor and therefore suffer from a spatial vs. angular resolution trade-off. Several architectures for compressive light-field imaging that require relatively few photon-efficient measurements to obtain a high-resolution estimate of the light field have been proposed in [86]–[88]. These architectures offer a significant improvement over traditional light-field imagers exploiting spatio-angular correlations inherent in the light fields of natural scenes. The basic principle of utilizing a coded aperture to obtain light field images is depicted in Fig. 12.

Inspired by different compressive imaging architectures, we have designed a simple setup for compressive imaging consisting of off-the-shelf components. A simple and portable plug-and-play compressive imaging architecture consists of digital camera and projector. Our compressive imaging method allows the use of low-resolution camera even with strong lens distortion for measurement acquisition. Results of our compressive imaging method are non-distorted, high-resolution images which are obtained using compressive sensing reconstruction process on acquired measurement data (see Fig. 13).

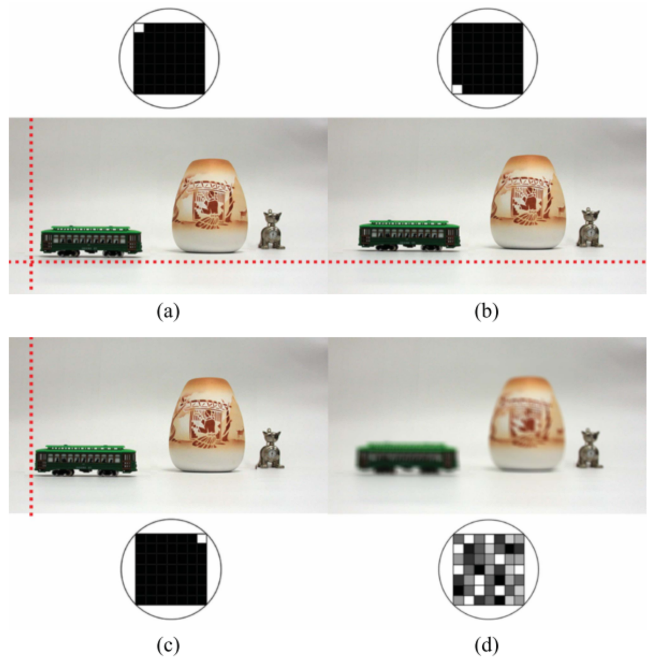

Fig. 12. Basic principle of utilizing a coded aperture to obtain light field images. (a)-(c) The angular images when only corner blocks of the aperture are left open. (d) Captured image with the randomly coded aperture used in the proposed compressive sensing light field camera. Angular images (a)-(c) can be de-multiplexed from (d) to obtain compressive sensing reconstruction of the captured light field [87]

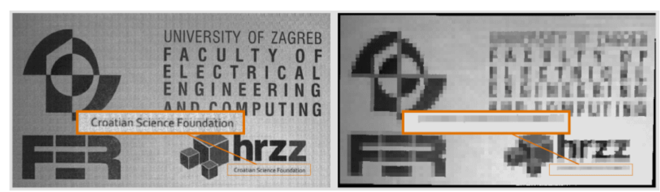

Fig. 13. Reconstruction of a two-dimensional scene using compressive imaging system composed of a projector and a camera. Left image shows the scene reconstruction, while right image shows single measurement from the camera sensor.

VII. Conclusion

In this paper, we gave an introduction to the theory of sparse signal modeling and compressive sensing. We covered basis representation fundamentals using an illustrative example, sparse signal recovery basics, desired properties of compressive sensing matrices and uniqueness of CS recovery problem. Furthermore, we provided overview of algorithms for sparse recovery and overcomplete dictionary learning. Finally, we showed numerous practical applications of compressive sensing framework on real-word problems. Compressive sensing offers novel signal acquisition and reconstruction framework. The traditional Shannon-Nyquist sampling theory in which successful recovery of a signal only depends on the bandwidth of the signal is replaced by novel compressive sensing framework where informational content of the signal dictates recovery success.

Compressive sensing has become an alternative for the well established Shannon-Nyquist sampling theory. However, there are several aspects of CS where further improvements are needed including sparse recovery algorithms, theoretical formulations and guarantees and finally, fundamental limits which will lead to novel and revolutionary practical applications solving real-world problems.

REFERENCES

[1] C. E. Shannon, “Communication in the presence of noise,” *Proc. Institute of Radio Engineers*, vol. 37, no. 1, pp. 10–21, 1949.

[2] H. Nyquist, “Certain topics in telegraph transmission theory,” *Transactions of the American Institute of Electrical Engineers*, vol. 47, no. 2, pp. 617–644, 1928.

[3] E. J. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

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

[5] E. J. Candès, “Compressive sampling,” *Proceedings of the International Congress of Mathematicians*, pp. 1433–1452, 2006.

[6] E. Candès and J. Romberg, “Sparsity and incoherence in compressive sampling,” *Inverse Problems*, vol. 23, no. 3, pp. 969–985, 2007.

[7] R. Baraniuk, “Compressive Sensing [Lecture Notes],” *IEEE Signal Processing Magazine*, vol. 24, no. July, pp. 118–121, 2007.

[8] E. Candes and M. Wakin, “An Introduction To Compressive Sampling,” *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[9] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, “Compressed sensing MRI: A look at how CS can improve on current imaging techniques,” *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 72–82, 2008.

[10] M. F. Duarte, M. A. Davenport, D. Takbar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, “Single-pixel imaging via compressive sampling,” *IEEE signal processing magazine*, vol. 25, no. 2, pp. 83–91, 2008.

[11] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk, “Beyond Nyquist: Efficient sampling of sparse bandlimited signals,” *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 520–544, 2010.

[12] M. F. Duarte and Y. C. Eldar, “Structured Compressed Sensing: From Theory to Applications,” vol. 59, no. 170, pp. 4053–4085, jun 2011.

[13] I. Rish and G. Grabarnik, *Sparse Modeling: Theory, Algorithms, and Applications*, 2015.

[14] Visited 17.10.2017. [Online]. Available: http://baspfrontiers.epfl.ch/context.php

[15] D. L. Donoho and M. Elad, “Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization,” *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.

[16] J. B. Kruskal, “Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics,” *Linear algebra and its applications*, vol. 18, no. 2, pp. 95–138, 1977.

[17] M. P. Pereira, L. Lovisolo, E. A. B. Da Silva, and M. L. R. De Campos, “On the design of maximally incoherent sensing matrices for compressed sensing using orthogonal bases and its extension for biorthogonal bases case,” *Digital Signal Processing: A Review Journal*, vol. 27, no. 1, pp. 12–22, 2014.

[18] J. Moreira, “What is... a rip matrix?” pp. 1–4, 2014.

[19] J. Romberg, “Mathematics of Compressive Sensing : Random matrices are restricted isometries,” 2013.

[20] E. J. Candès, “The restricted isometry property and its implications for compressed sensing,” *Comptes Rendus Mathematique*, vol. 346, no. 9, pp. 589 – 592, 2008.

[21] J. D. Blanchard, C. Cartis, and J. Tanner, “Compressed sensing: How sharp is the restricted isometry property?” *SIAM Review*, vol. 53, no. 1, pp. 105–125, 2011.

[22] R. J. Duffin and A. C. Schaeffer, “A class of nonharmonic fourier series,” *Transactions of the American Mathematical Society*, vol. 72, no. 2, pp. 341–366, 1952.

[23] E. J. Candès and Y. Plan, “A probabilistic and RIPless theory of compressed sensing,” *IEEE Transactions on Information Theory*, vol. 57, no. 11, pp. 7235–7254, 2011.

[24] B. Adcock, a. C. Hansen, C. Poon, and B. Roman, “Breaking the coherence barrier: asymptotic incoherence and asymptotic sparsity in compressed sensing,” *arXiv preprint arXiv:1302.0561*, pp. 1–44, 2013.

[25] B. Adcock, A. C. Hansen, and B. Roman, “The Quest for Optimal Sampling: Computationally Efficient, Structure-Exploiting Measurements for Compressed Sensing,” *Journal of Physics A: Mathematical and Theoretical*, vol. 44, no. 8, pp. 143–167, 2015.

[26] B. Roman, A. Hansen, and B. Adcock, “On asymptotic structure in compressed sensing,” no. c, pp. 1–10, 2014.

[27] B. Adcock, B. C. Canada, A. C. Hansen, and B. Roman, “Compressed sensing with local structure : theory , applications and benefits.”

[28] K. Li and S. Cong, “State of the art and prospects of structured sensing matrices in compressed sensing,” *Frontiers of Computer Science*, vol. 9, no. 5, pp. 665–677, 2015.

[29] W. U. Bajwa, J. D. Haupt, G. M. Raz, S. J. Wright, and R. D. Nowak, “Toeplitz-structured compressed sensing matrices,” in *Statistical Signal Processing, 2007. SSP’07. IEEE/SP 14th Workshop on*. IEEE, 2007, pp. 294–298.

[30] J. Haupt, W. U. Bajwa, G. Raz, and R. Nowak, “Toeplitz compressed sensing matrices with applications to sparse channel estimation,” *IEEE transactions on information theory*, vol. 56, no. 11, pp. 5862–5875, 2010.

[31] W. Yin, B. Simon Morgan, J. Yang *et al.*, “Practical compressive sensing with toeplitz and circulant matrices.”

[32] K. Schäcke, “On the kronecker product,” 2013.

[33] Y. Rivenson and A. Stern, “Compressed imaging with a separable sensing operator,” *IEEE Signal Processing Letters*, vol. 16, no. 6, pp. 449–452, 2009.

[34] M. F. Duarte and R. G. Baraniuk, “Kronecker compressive sensing,” *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 494–504, 2012.

[35] I. Daubechies, *Ten lectures on wavelets*. SIAM, 1992.

[36] J.-L. Starck, E. J. Candès, and D. L. Donoho, “The curvelet transform for image denoising,” *IEEE Transactions on image processing*, vol. 11, no. 6, pp. 670–684, 2002.

[37] M. N. Do and M. Vetterli, “Contourlets: a directional multiresolution image representation,” in *Image Processing. 2002. Proceedings. 2002 International Conference on*, vol. 1. IEEE, 2002, pp. I–I.

[38] K. Guo and D. Labate, “Optimally sparse multidimensional representation using shearlets,” *SIAM journal on mathematical analysis*, vol. 39, no. 1, pp. 298–318, 2007.

[39] E. Le Pennec and S. Mallat, “Sparse geometric image representations with bandelets,” *IEEE transactions on image processing*, vol. 14, no. 4, pp. 423–438, 2005.

[40] B. A. Olshausen and D. J. Field, “Emergence of simple-cell receptive field properties by learning a sparse code for natural images,” *Nature*, vol. 381, no. 6583, p. 607, 1996.

[41] Y. Petrov, “Overcomplete dictionaries for sparse representation of signals,” *PloS one*, vol. 7, no. 10, p. e44439, 2012.

[42] R. Rubinstein and M. Zibulevsky, “Learning Sparse Dictionaries for Sparse Signal Representation,” vol. X, no. X, pp. 1–11.

[43] S. Christen, C. Studer, and G. Pope, “Dictionary Learning for Super-Resolution,” 2010.

[44] K. Engan, S. O. Aase, and J. H. Husoy, “Method of optimal directions for frame design,” in *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*, vol. 5. IEEE, 1999, pp. 2443–2446.

[45] M. Aharon, M. Elad, and A. Bruckstein, “K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation,” *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[46] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, “Online learning for matrix factorization and sparse coding,” *Journal of Machine Learning Research*, vol. 11, no. Jan, pp. 19–60, 2010.

[47] R. Rubinstein, M. Zibulevsky, and M. Elad, “Double sparsity: Learning sparse dictionaries for sparse signal approximation,” *IEEE Transactions on Signal Processing*, vol. 58, no. 3 PART 2, pp. 1553–1564, 2010.

[48] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[49] Z. Zhang, Y. Xu, J. Yang, X. Li, and D. Zhang, “A Survey of Sparse Representation: Algorithms and Applications,” *IEEE Access*, vol. 3, pp. 490–530, 2015.

[50] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM review*, vol. 43, no. 1, pp. 129–159, 2001.

[51] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, “Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition,” in *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*. IEEE, 1993, pp. 40–44.

[52] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, “Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit,” *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 1094–1121, 2012.

[53] 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.

[54] M. A. Figueiredo, R. D. Nowak, and S. J. Wright, “Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems,” *IEEE Journal of selected topics in signal processing*, vol. 1, no. 4, pp. 586–597, 2007.

[55] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani *et al.*, “Least angle regression,” *The Annals of statistics*, vol. 32, no. 2, pp. 407–499, 2004.

[56] L. F. Portugal, M. G. Resende, G. Veiga, and J. Júdice, “A truncated primal-infeasible dual-feasible network interior point method,” *Networks*, vol. 35, no. 2, pp. 91–108, 2000.

[57] J. Yang and Y. Zhang, “Alternating direction algorithms for \ell_1-problems in compressive sensing,” *SIAM journal on scientific computing*, vol. 33, no. 1, pp. 250–278, 2011.

[58] M. A. Figueiredo and R. D. Nowak, “A bound optimization approach to wavelet-based image deconvolution,” in *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, vol. 2. IEEE, 2005, pp. II–782.

[59] A. Beck and M. Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[60] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo, “Sparse reconstruction by separable approximation,” *IEEE Transactions on Signal Processing*, vol. 57, no. 7, pp. 2479–2493, 2009.

[61] S. Becker, J. Bobin, and E. J. Candès, “Nesta: A fast and accurate first-order method for sparse recovery,” *SIAM Journal on Imaging Sciences*, vol. 4, no. 1, pp. 1–39, 2011.

[62] H. Qi and S. Hughes, “Using the kernel trick in compressive sensing: Accurate signal recovery from fewer measurements,” in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. IEEE, 2011, pp. 3940–3943.

[63] K. Kulkarni, S. Lohit, P. Turaga, R. Kerviche, and A. Ashok, “ReconNet: Non-Iterative Reconstruction of Images from Compressively Sensed Random Measurements,” 2016.

[64] A. Mousavi and R. G. Baraniuk, “Learning to invert: Signal recovery via Deep Convolutional Networks,” in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, mar 2017, pp. 2272–2276.

[65] A. Mousavi, A. B. Patel, and R. G. Baraniuk, “A deep learning approach to structured signal recovery,” in *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, sep 2015, pp. 1336–1343.

[66] C. Dong, C. C. Loy, K. He, and X. Tang, “Image Super-Resolution Using Deep Convolutional Networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, 2016.

[67] H. Yao, F. Dai, D. Zhang, Y. Ma, S. Zhang, Y. Zhang, and Q. Tian, “DRˆ2-Net: Deep Residual Reconstruction Network for Image Compressive Sensing,” pp. 1–10, 2017. [Online]. Available: http://arxiv.org/abs/1702.05743

[68] M. Nejati, S. Samavi, N. Karimi, S. M. R. Soroushmehr, and K. Najarian, “Boosted dictionary learning for image compression,” *IEEE Transactions on Image Processing*, vol. 25, no. 10, pp. 4900–4915, 2016.

[69] I. Horev, O. Bryt, and R. Rubinstein, “Adaptive image compression using sparse dictionaries,” in *Systems, Signals and Image Processing (IWSSIP), 2012 19th International Conference on*. IEEE, 2012, pp. 592–595.

[70] K. Skretting and K. Engan, “Image compression using learned dictionaries by rls-dla and compared with k-svd,” in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1517–1520.

[71] M. Elad and M. Aharon, “Image Denoising Via Sparse and Redundant Representations Over Learned Dictionaries,” *IEEE Transactions on Image Processing*, vol. 15, no. 12, p. 2008, 2008.

[72] B. Shen, W. Hu, Y. Zhang, and Y.-j. Zhang, “Image inpainting via sparse representation,” *Icassp*, pp. 159–162, 2009.

[73] J. Mairal, G. Sapiro, and M. Elad, “Learning Multiscale Sparse Representations for Image and Video Restoration,” *Multiscale Modeling & Simulation*, vol. 7, no. 1, pp. 214–241, 2008.

[74] J. Yang, J. Wright, T. S. Huang, and Y. Ma, “Image super-resolution via sparse representation,” *IEEE transactions on image processing*, vol. 19, no. 11, pp. 2861–2873, 2010.

[75] K. I. Kim and Y. Kwon, “Single-image super-resolution using sparse regression and natural image prior,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 6, pp. 1127–1133, 2010.

[76] J. Yang, J. Wright, T. Huang, and Y. Ma, “Image super-resolution as sparse representation of raw image patches,” in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

[77] Q. Zhang and B. Li, “Discriminative K-SVD for dictionary learning in face recognition,” in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, jun 2010, pp. 2691–2698.

[78] Zhuolin Jiang, Zhe Lin, and L. S. Davis, “Label Consistent K-SVD: Learning a Discriminative Dictionary for Recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2651–2664, nov 2013.

[79] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Yi Ma, “Robust Face Recognition via Sparse Representation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, feb 2009.

[80] S. Geethanath, R. Reddy, A. S. Konar, S. Imam, R. Sundaresan, R. Babu D. R., and R. Venkatesan, “Compressed Sensing MRI: A Review.” *Critical reviews in biomedical engineering*, vol. 41, no. 3, pp. 183–204, 2013.

[81] D. Reddy, A. Veeraraghavan, and R. Chellappa, “P2C2: Programmable pixel compressive camera for high speed imaging,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 329–336, 2011.

[82] M. Gupta, A. Agrawal, and A. Veeraraghavan, “Flexible Voxels for Motion-Aware Videography,” *European Conference on Computer Vision (ECCV)*, pp. 100–114, 2010.

[83] R. Koller, L. Schmid, N. Matsuda, T. Niederberger, L. Spinoulas, O. Cossairt, G. Schuster, and A. K. Katsaggelos, “High spatio-temporal resolution video with compressed sensing,” *Optics Express*, vol. 23, no. 12, p. 15992, 2015.

[84] P. Llull, X. Liao, X. Yuan, J. Yang, D. Kittle, L. Carin, G. Sapiro, and D. J. Brady, “Coded aperture compressive temporal imaging,” *Optics Express*, vol. 21, no. 9, p. 10526, 2013.

[85] P. Llull, X. Yuan, X. Liao, J. Yang, L. Carin, G. Sapiro, and D. J. Brady, “Compressive extended depth of field using image space coding,” *Class. Opt. 2014*, p. CM2D.3, 2014.

[86] A. Ashok and M. A. Neifeld, “Compressive light field imaging,” *Proc. SPIE*, vol. 7690, p. 76900Q, 2010.

[87] S. D. Babacan, R. Ansorge, M. Luessi, P. R. Mataran, R. Molina, and A. K. Katsaggelos, “Compressive light field sensing,” *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4746–4757, 2012.

[88] K. Marwah, G. Wetzstein, Y. Bando, and R. Raskar, “Compressive light field photography using overcomplete dictionaries and optimized projections,” *ACM Transactions on Graphics*, vol. 32, no. 4, p. 1, 2013.