This is a pre-print of an article published in Circuits Systems and Signal Processing.
The final authenticated version is available online at: https://doi.org/10.1007/s00034-018-0909-2

A Tutorial on Sparse Signal Reconstruction and its Applications in Signal Processing

Ljubiša Stanković, Ervin Sejdić†, Srdjan Stanković*, Miloš Daković*, and Irena Orović**

Abstract

Sparse signals are characterized by a few nonzero coefficients in one of their transformation domains. This was the main premise in designing signal compression algorithms. Compressive sensing as a new approach employs the sparsity property as a precondition for signal recovery. Sparse signals can be fully reconstructed from a reduced set of available measurements. The description and basic definitions of sparse signals, along with the conditions for their reconstruction, are discussed in the first part of this paper. The numerous algorithms developed for the sparse signals reconstruction are divided into three classes. The first one is based on the principle of matching components. Analysis of noise and nonsparsity influence on reconstruction performance is provided. The second class of reconstruction algorithms is based on the constrained convex form of problem formulation where linear programming and regression methods can be used to find a solution. The third class of recovery algorithms is based on the Bayesian approach. Applications of the considered approaches are demonstrated through various illustrative and signal processing examples, using common transformation and observation matrices. With pseudo-codes of the presented algorithms and compressive sensing principles illustrated on simple signal processing examples, this tutorial provides an inductive way through this complex field to researchers and practitioners starting from the basics of sparse signal processing up to the most recent and up-to-date methods and signal processing applications.

Index terms— Sparse signals, Compressive sensing, Signal sampling, Signal representation, Signal reconstruction, Discrete Fourier transform

1 Introduction

A discrete signal can be considered as sparse in a certain representation domain if the number of nonzero values in that domain is much smaller than the total signal length. Common situation in the case of real-world signals is that the number of significant coefficients is small as compared to the number of other components. These coefficients could be neglected or set to zero. The sparse signals are considered as compressible signals.

When the sparse signal representation is obtained as the result of a linear signal transformation, then the signal samples may be considered as observations (or measurements) of the sparse representation domain. In recent years, it has been shown that sparse signals can be reconstructed from a reduced set of measurements, much smaller than the conventional number of samples required by the signal bandwidth [3,4,8,13,17,18,26,27,32,34,36,37,41,42,44,48,70,72,74,76,81,90,112,113]. A reduced set of measurements can result from a rational and unwasteful sensing strategy aiming to acquire the lowest possible number of measurements (compressive sensing), but it can also result from a physical unavailability to take a complete set of measurements/samples. Moreover, in various applications it happens that arbitrarily samples are heavily corrupted by disturbances, and these are in turn omitted and considered as unavailable. Regardless of the reasons for dealing with a reduced set of measurements, all of the mentioned circumstances can be considered within the unified compressive sensing theoretical framework. It allows us to accurately recover sparse signals using only available samples, provided that the measurement process is incoherent, i.e., the relative contribution from nonzero transform domain

*Ljubiša Stanković (ljubisa@ac.me), Srdjan Stanković (srdjan@ac.me), Miloš Daković (milos@ac.me), and Irena Orović (ienao@ac.me) are with Faculty of Electrical Engineering, University of Motenegro, Podgorica, Motenegro.

†Ervin Sejdić (esejdic@pitt.edu) is with Electrical and Computer Engineering, University of Pittsburgh, Pittsburgh, PA, USA

This is a pre-print of an article published in Circuits Systems and Signal Processing.
The final authenticated version is available online at: https://doi.org/10.1007/s00034-018-0909-2

coefficients is different for each measurement. By exploiting the sparsity and coherence, the problem of recovering the entire signal is observed as a problem of finding the sparsest representation in certain transform bases that corresponds to the available set of measurements.

Many techniques have been adopted to produce the sparsest approximation from redundant dictionaries. The simplest yet efficient class of algorithms belongs to the so-called greedy iterative methods [33], wherein the algorithm selects the transform basis function that participates the most in total signal energy in each iteration. The most common among these algorithms are Matching Pursuit (MP) [58], Orthogonal Matching Pursuit (OMP) [110,111], and Compressive Sampling Matching Pursuit (CoSaMP) [63,64]. The approaches in the second class are based on the convex optimization algorithms. The basic formulations of problems in this class are formulated using the Basis Pursuits (BP) [29], which relaxes the sparsity condition by the ℓ_1-norm and solves the problem through linear programming, using for example, the Basis Pursuit Denoising (BPDN) [55] or the Least Absolute Shrinkage and Selection Operator (LASSO) [23,57,107]. The third class covers Bayesian methods that observe the unknowns as stochastic variables with assigned probability distributions [6,10,45,109,114].

The reconstruction of sparse signals from a reduced set of measurements is the topic of this tutorial. We will start by reviewing the basic concepts and constraints of compressive sensing theory needed to understand signal reconstruction algorithms. Further, several reconstruction algorithms belonging to the three mentioned classes are explained and summarized. The benefits and efficiency of using these algorithms are demonstrated through the variety of signal processing applications presented in the last part of this tutorial.

2 Reconstruction of Sparse Signals

Basic definitions needed to understand various reconstruction approaches will be introduced first. Conditions that have to be met in order to reconstruct the full signal from a reduced set of measurements will be presented in this section as well.

2.1 Measurements

Consider a set of N coefficients $X(k)$, for $k = 0,1,...,N-1$. Denote the vector with elements $X(k)$ as $\mathbf{X}$. This vector is considered as sparse if the number of its nonzero coefficients, denoted by K, is much smaller than the total number of samples N,

$$X(k) = 0 \text{ for } k \notin \mathbb{K} = \{k_1, k_2, ..., k_K\} \text{ and } K \ll N. \tag{1}$$

The number of nonzero coefficients is commonly denoted by $\|\mathbf{X}\|_0 = K$. Function $\|\mathbf{X}\|_0$ is typically referred to as the ℓ_0-norm (norm-zero) or the ℓ_0-pseudo-norm of the vector $\mathbf{X}$ although it does not satisfy the norm properties.

A measurement of vector $\mathbf{X}$ is defined as a linear combination of its elements $X(k)$. The mth measurement will be denoted by $y(m)$. If M measurements $y(0), y(1), \ldots, y(M-1)$ are available they can be written in the form of a system of M linear equations

$$y(m) = \sum_{k=0}^{N-1} \alpha_k(m) X(k), \quad m = 0,1,\ldots,M-1, \quad M < N, \tag{2}$$

where $\alpha_k(m)$ are the weighting (measurement) coefficients. In a matrix form we can write

$$\begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(M-1) \end{bmatrix} = \begin{bmatrix} \alpha_0(0) & \alpha_1(0) & \cdots & \alpha_{N-1}(0) \\ \alpha_0(1) & \alpha_1(1) & \cdots & \alpha_{N-1}(1) \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_0(M-1) & \alpha_1(M-1) & \cdots & \alpha_{N-1}(M-1) \end{bmatrix} \begin{bmatrix} X(0) \\ X(1) \\ \vdots \\ X(N-1) \end{bmatrix}$$

$$\mathbf{y} = \mathbf{A}\mathbf{X}.$$

The matrix $\mathbf{A}$, whose elements are $\alpha_k(m)$, is called the measurement matrix.

This is a pre-print of an article published in Circuits Systems and Signal Processing.
The final authenticated version is available online at: https://doi.org/10.1007/s00034-018-0909-2

2.2 Common Measurement Matrices

Some common measurement matrices used in practical applications and theoretical considerations will be presented here.

Randomness of measurement matrices is a favorable property in compressive sensing. Matrices with random elements are often used. The most common measurement matrix with random elements is a matrix with Gaussian-distributed numbers as its elements

$$\alpha_k(m) \sim \frac{1}{\sqrt{M}}\mathcal{N}(0,1).$$

The mean value of these numbers is zero and the variance is one. They are normalized with $1/\sqrt{M}$ so that the energy of each column of the measurement matrix is one. Bernoulli random matrices, whose elements take the value $1/\sqrt{M}$ or $-1/\sqrt{M}$, are also used in compressive sensing.

In signal processing, the measurement matrices are usually formed as partial matrices of the common (inverse) linear signal transforms. Consider a signal $\mathbf{x}$ and its linear transform $\mathbf{X}$. An analysis model (direct transformation) is

$$\mathbf{X} = \mathbf{\Phi x}, \tag{3}$$

where $\mathbf{\Phi}$ is a full size $(N \times N)$ transformation matrix. The synthesis model is of the form

$$\mathbf{x} = \mathbf{\Psi X}, \tag{4}$$

where $\mathbf{\Psi}$ is a full size $(N \times N)$ inverse transformation matrix. A signal sample $x(n)$ can also be written as

$$x(n) = \sum_{k=0}^{N-1} \psi_k(n) X(k). \tag{5}$$

where $\psi_k(n)$ are the elements of the inverse transformation matrix. Assuming a linear transform, the signal samples $x(n)$ can be represented as linear combinations of the transform coefficients $X(k)$. Therefore, the signal samples can be considered as the measurements of $\mathbf{X}$.

An arbitrary set of signal samples with elements $x(n_{m+1}) = y(m)$, where $m = 0,1,\ldots,M-1$ is a set of measurements where the measurement matrix elements $\alpha_k(m)$ are equal to $\psi_k(n_{m+1})$. The measurement matrix $\mathbf{A}$ can be obtained from the inverse transformation matrix $\mathbf{\Psi}$ by keeping only the rows that correspond to the positions of the measurements $y(m)$ within the signal vector $\mathbf{x}$.

In signal processing, the most common transform is the discrete Fourier transform DFT.[1] The coefficients of its direct transform matrix $\mathbf{\Phi}$ are defined as $\phi_k(n) = \exp(-j2\pi nk/N)$. The inverse DFT matrix coefficients are

$$\psi_k(n) = \frac{1}{N} e^{j2\pi nk/N}.$$

The signal samples $y(m) = x(n_{m+1})$, for $m = 0,1,\ldots,M-1$, where

$$n_m \in \mathbb{M} = \{n_1, n_2, \ldots, n_M\} \subset \{0,1,\ldots,N-1\},$$

and

$$y(m) = x(n_{m+1}) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{j2\pi n_m k/N}$$

[1] Consider a signal $x(t)$ of a duration T and its samples $x(n\Delta t)$ satisfying the sampling theorem. The periodic extension of this signal can be written in a Fourier series (FS) form

$$x(t) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{j2\pi k \frac{t}{T}},$$

where the FS coefficients $X(k)$ are equal to the DFT coefficients if we use the notation $x(n)$ for $x(n\Delta t)$, and $\Delta t = T/N$ as the sampling interval. When the sampling theorem is satisfied then

$$X(k) = \frac{N}{T} \int_0^T x(t) e^{-j2\pi kt/T} dt = \sum_{n=0}^{N-1} x(n) e^{-j2\pi kn/N}, \quad k = 0,1,2,\ldots,N-1.$$

This is the DFT of a signal $x(n)$.

This is a pre-print of an article published in Circuits Systems and Signal Processing.
The final authenticated version is available online at: https://doi.org/10.1007/s00034-018-0909-2

can be considered as the measurements of coefficients $X(k)$. The measurement matrix is defined in this case by

$$\mathbf{A}=\frac{1}{N}\begin{bmatrix}1 & e^{j2\pi n_1/N} & \dots & e^{j2\pi n_1(N-1)/N}\\ 1 & e^{j2\pi n_2/N} & \dots & e^{j2\pi n_2(N-1)/N}\\ \vdots & \vdots & \ddots & \vdots\\ 1 & e^{j2\pi n_M/N} & \dots & e^{j2\pi n_M(N-1)/N}\end{bmatrix}. \tag{6}$$

This is a partial inverse DFT matrix. In compressive sensing theory, it is common to normalize the measurement matrix so that the energy of its columns (diagonal elements of $\mathbf{A}^H\mathbf{A}$ matrix) is equal to one. Then the factor $1/N$ in $\mathbf{A}$ should be replaced by $1/\sqrt{M}$.

In order to increase randomness in the Fourier transform matrix, the measurements may be taken at any random instant. Then the measurement vector elements are $y(m-1)=x(t_m)$, where t_m, $m=1,2,\dots,M$, are random instants within the considered time interval T. The measurement matrix follows then from the Fourier series definition

$$x(t)=\sum_{k=0}^{N-1}X(k)e^{j2\pi kt/T}.$$

It has been assumed that the Fourier series coefficients are within $0\leq k\leq N-1$. The measurement matrix is

$$\mathbf{A}=\begin{bmatrix}1 & e^{j2\pi t_1/T} & \dots & e^{j2\pi t_1(N-1)/T}\\ 1 & e^{j2\pi t_2/T} & \dots & e^{j2\pi t_2(N-1)/T}\\ \vdots & \vdots & \ddots & \vdots\\ 1 & e^{j2\pi t_M/T} & \dots & e^{j2\pi t_M(N-1)/T}\end{bmatrix} \tag{7}$$

with a possible normalization factor $1/\sqrt{M}$. This measurement matrix is a partial random inverse Fourier transform matrix.

Other signal transforms and other random distributions can be used to form an appropriate measurement matrix. For example, if a signal is sparse in the discrete cosine transform (DCT) domain [2] (or in the discrete sine transform (DST) domain [60]), then this transform will be used to perform measurements and form the corresponding measurement matrix.

The signal of interest $x(n)$ can also be measured in an indirect way. Consider, for example, a signal which is sparse in the domain of the wavelet transform or the short-time Fourier transform. Sparsity domain coefficients $\mathbf{X}$ are then related to the signal samples within the window used for their calculation. If we consider M signal samples from one window only, they do not contain any information about the transform coefficients calculated using signal samples outside this window. The reconstruction of all transform coefficients $\mathbf{X}$ is not possible with such a set of samples. However, if we form a linear combination of all signal samples $x(n)$ with random nonzero coefficients $b_k(m)$ then a measurement $y(m)$ defined as

$$y(m)=b_0(m)x(0)+b_1(m)x(1)+...+b_{N-1}(m)x(N-1), \tag{8}$$

will contain information about all signal samples and all sparsity domain coefficients $\mathbf{X}$. In a matrix form the relation between the signal $\mathbf{x}$ and the indirect measurement vector $\mathbf{y}$ is given by

$$\mathbf{y}=\mathbf{B}\mathbf{x},$$

where $b_k(n)$ are elements of the matrix $\mathbf{B}$. Since the signal is related to its sparsity domain by $\mathbf{x}=\mathbf{\Psi X}$, the measurements are related to the sparsity domain transform of the signal as

$$\mathbf{y}=\mathbf{B\Psi X}=\mathbf{AX},$$

where

$$\mathbf{A}=\mathbf{B\Psi}.$$

These cases can be analyzed in a similar way as in the case of direct measurements.

Very often in practical applications the signal sparsity cannot be revealed in a single orthonormal transform basis. Instead, we need to use larger and diversified dictionaries of atoms. In that case,

the proper choice of the dictionary that sparsifies the signal is essential in order to model the signal as a linear combination of a few dictionary atoms [1, 66, 77, 120]. One common solution is to use the overcomplete dictionaries made from several prespecified sets of linear transforms as it is the case with the oversampled DFT, Gabor frames, wavelet and curvelet frames, as well as the concept of concatenated transform matrices. The second approach that recently evolved, assumes the design of the dictionary through the process of adaptive atoms learning based on the set of training signals to which it is supposed to serve [1,77]. Generally there are many techniques that can be used for adaptive dictionary learning mainly based on the machine learning algorithms, starting from the maximum likelihood analysis, optimal directions, principal component analysis (PCA), parametric training methods, and particularly the K-SVD algorithm [1] are examples of flexible iterative algorithm that can work with any pursuit method.

2.3 Direct Search Solution

We will start the reconstruction analysis with the simplest case of a sparse signal representation $\mathbf{X}$ with only one nonzero coefficient. This signal has sparsity $K = 1$. Hence, we are looking for a single nonzero coefficient $X(i)$ with unknown value and position within $\mathbf{X}$. A direct sensing approach would imply sensing/measuring the set of N independent values of $X(k)$, for $k = 0,1,\dots,N-1$. However, if N is large and there is only $K = 1$ nonzero coefficient in $\mathbf{X}$ we can definitely solve the problem with a significantly reduced number of measurements.

1) Consider one available measurement of $\mathbf{X}$ denoted by $y(0)$:

$$y(0) = \sum_{k=0}^{N-1} X(k)\alpha_k(0), \tag{9}$$

where $\alpha_k(0), k = 0,1,\dots,N-1$ are the weighting coefficients. This equation represents an N-dimensional hyperplane in the space of transform domain variables $X(0)$, $X(1)$, ... , $X(N-1)$, where only one of the variables $X(i)$ is nonzero. Any cross-section of this hyperplane with any of the coordinate axes could be a solution to the reconstruction problem for $K = 1$. Assuming that all weighting coefficients are nonzero $\alpha_k(0) \neq 0$, a solution will exist for any k. Thus, one measurement can produce N possible solutions with sparsity $K = 1$:

$$X(k) = y(0)/\alpha_k(0), \qquad k = 0,1,\dots,N-1.$$

It means that with one measurement we cannot solve the problem. We need one more measurement, at least.

2) Assume that two measurements $y(0)$ and $y(1)$ are available. Now we have two hyperplanes and two sets of possible solutions:

$$X(k) = y(0)/\alpha_k(0), \qquad k = 0,1,\dots,N-1,$$
$$X(k) = y(1)/\alpha_k(1), \qquad k = 0,1,\dots,N-1.$$

The unique solution is achieved if these two sets of solutions have only one common value at $k = i$:

$$X(i) = y(0)/\alpha_i(0) = y(1)/\alpha_i(1).$$

Statement: A common value for two measurements $X(i) = y(0)/\alpha_i(0)$ and $X(i) = y(1)/\alpha_i(1)$ is unique if

$$\alpha_i(0)\alpha_k(1) - \alpha_i(1)\alpha_k(0) = \det\begin{bmatrix}\alpha_i(0) & \alpha_k(0)\\ \alpha_i(1) & \alpha_k(1)\end{bmatrix} \neq 0 \tag{10}$$

for any $i \neq k$.

This statement will be proven by contradiction. Assume that two solutions with sparsity $K = 1$ can be obtained. The nonzero coefficient of the first solution is $X(i)$ and the nonzero coefficient of the second solution is $X(k)$ at $k \neq i$. Both of them are obtained based on two measurements $y(0)$ and $y(1)$. Then these two solutions satisfy the measurement equation with only one nonzero coefficient

$$\begin{bmatrix}y(0)\\ y(1)\end{bmatrix} = \begin{bmatrix}\alpha_i(0) & \alpha_k(0)\\ \alpha_i(1) & \alpha_k(1)\end{bmatrix}\begin{bmatrix}X(i)\\ 0\end{bmatrix}$$

and

$$\begin{bmatrix} y(0) \\ y(1) \end{bmatrix} = \begin{bmatrix} \alpha_i(0) & \alpha_k(0) \\ \alpha_i(1) & \alpha_k(1) \end{bmatrix} \begin{bmatrix} 0 \\ X(k) \end{bmatrix}. \tag{11}$$

By subtracting these two equations we obtain:

$$\begin{bmatrix} \alpha_i(0) & \alpha_k(0) \\ \alpha_i(1) & \alpha_k(1) \end{bmatrix} \begin{bmatrix} X(i) \\ -X(k) \end{bmatrix} = 0. \tag{12}$$

For $\alpha_i(0)\alpha_k(1) - \alpha_i(1)\alpha_k(0) \neq 0$ follows $X(i) = X(k) = 0$. Therefore, two different nonzero solutions $X(i)$ and $X(k)$ in this case cannot exist. This concludes the proof.

It means that $\text{rank}(\mathbf{A}_2) = 2$ for any $\mathbf{A}_2$ being a 2×2 submatrix of the matrix of coefficients (measurement matrix) $\mathbf{A}$. Here we will introduce the spark of a matrix $\mathbf{A}$ in notation $\text{spark}\{\mathbf{A}\}$. While the rank of a matrix is equal to the largest number of independent columns/rows the spark of a matrix is the smallest number of dependent columns/rows. If any column of matrix $\mathbf{A}$ is with all zero values then $\text{spark}\{\mathbf{A}\} = 1$. If there is no an all-zero column and there are two linearly dependent columns (in this case a singular 2×2 submatrix $\mathbf{A}_2$) then $\text{spark}\{\mathbf{A}\} = 2$. In our example, the reconstruction is unique if $\text{spark}\{\mathbf{A}\} > 2$, meaning that there are no two dependent columns in the measurement matrix $\mathbf{A}$.

Generalization: Consider now a sparse vector $X(k)$ with sparsity K. Using a subset of K measurements, denoted as $\mathbf{y}_K^{(1)}$, we can get a set of solutions for possible nonzero coefficients. For each assumed positions of K nonzero components in $X(k)$ we form the vector $\mathbf{X}_K$ with K nonzero elements $k \in \mathbb{K} = \{k_1, k_2, ..., k_K\}$. The system $\mathbf{y}_K^{(1)} = \mathbf{A}_K \mathbf{X}_K$ is solved for each possible set of positions $\{k_1, k_2, ..., k_K\}$. Note that there are $\binom{N}{K}$ combinations for different positions. We get then $\binom{N}{K}$ possible solutions $\mathbf{X}_K$. The same process should be done with another subset of K measurements $\mathbf{y}_K^{(2)}$. Again, we solve the system $\mathbf{y}_K^{(2)} = \mathbf{A}_K \mathbf{X}_K$ with $\binom{N}{K}$ combinations of different nonzero positions $\{k_1, k_2, ..., k_K\}$. If these two sets of $\binom{N}{K}$ solutions have only one common member $\mathbf{X}_K$, then this is the solution of our problem. The uniqueness of the solution will be discussed next.

Let us assume that two different solutions for $\mathbf{X}$ of sparsity K exist and that $M = 2K$ measurements are available within $\mathbf{y}$. Then, the solution is unique if the determinants of all $\mathbf{A}_{2K}$ submatrices of $\mathbf{A}$ are different from zero (all possible $\mathbf{A}_{2K}$ submatrices are nonsingular). If we denote these two nonzero parts of the solutions by $\mathbf{X}_K^{(1)}$ and $\mathbf{X}_K^{(2)}$, then both of them satisfy the measurement equation

$$\begin{bmatrix} \mathbf{A}_K^{(1)} & \mathbf{A}_K^{(2)} \end{bmatrix} \begin{bmatrix} \mathbf{X}_K^{(1)} \\ \mathbf{0}_K \end{bmatrix} = \mathbf{y} \quad \text{and} \quad \begin{bmatrix} \mathbf{A}_K^{(1)} & \mathbf{A}_K^{(2)} \end{bmatrix} \begin{bmatrix} \mathbf{0}_K \\ \mathbf{X}_K^{(2)} \end{bmatrix} = \mathbf{y}. \tag{13}$$

where $\mathbf{A}_K^{(1)}$ and $\mathbf{A}_K^{(2)}$ are $M \times K$ submatrices of $\mathbf{A}$ corresponding to the elements in $\mathbf{X}_K^{(1)}$ and $\mathbf{X}_K^{(2)}$.

There are no nonzero solutions for $\mathbf{X}_K^{(1)}$ and $\mathbf{X}_K^{(2)}$ if the determinant of $2K \times 2K$ matrix $\mathbf{A}_{2K} = \begin{bmatrix} \mathbf{A}_K^{(1)} & \mathbf{A}_K^{(2)} \end{bmatrix}$ is nonzero. If all possible submatrices $\mathbf{A}_{2K}$ (including all lower order submatrices) of measurement matrix $\mathbf{A}$ are nonsingular, then two solutions of sparsity K cannot exist, and the solution is unique. The proof is straightforward, following the case for $K = 1$.

Based on the previous analysis, the solution for a K sparse problem is unique if

$$\text{spark}\{\mathbf{A}\} > 2K.$$

In general the number of measurements is $M \geq 2K$. Again we should consider each of possible vectors $\mathbf{X}_K$ with K nonzero elements $k \in \mathbb{K} = \{k_1, k_2, ..., k_K\}$. In this case we have M equations with K unknowns. Since there are more equations than the unknowns the system is solved in the mean square sense as

$$\mathbf{X}_K = \left(\mathbf{A}_K^H \mathbf{A}_K\right)^{-1} \mathbf{A}_K^H \mathbf{y}. \tag{14}$$

Details and derivation of this formula are given in Section 3.1. This system has to be solved for $\binom{N}{K}$ combinations of possible nonzero positions $\{k_1, k_2, ..., k_K\}$. Among all solutions, the desired one is the solution producing a minimal squared error calculated as $e^2 = \|\mathbf{y} - \mathbf{A}_K \mathbf{X}_K\|_2^2$. In theory it should be zero for the true solution of the problem. However, there are $\binom{N}{K}$ possible combinations.

For any reasonable N and K the number of combinations $\binom{N}{K}$ is extremely large and computationally not feasible since this combinatorial problem is an NP hard problem (with NonPolynomial number of combinations).

As direct search is not a viable option, several alternative reconstruction algorithms have been proposed. We will start with algorithms based on the direct problem solution with a minimal number of nonzero coefficients that avoid a direct search combinatorial approach. In the second class of algorithms the convex relaxation of the ℓ_0-norm to the ℓ_1-norm is used with appropriate iterative solvers. Finally, the Bayesian approach to the sparse signals reconstruction is presented. Before we present the commonly used reconstruction algorithms, we will look at some important conditions imposed on the signal reconstruction.

2.4 Conditions for Reconstruction

For an arbitrary K sparse $\mathbf{X}$ and $M \geq 2K$ its measurements the reconstruction solution is unique if all measurement submatrices $\mathbf{A}_{2K}$ are nonsingular for all possible positions $\{k_1, k_2, ..., k_{2K}\}$ and for any combination of $2K$ indices from all possible measurement indices $m \in \{0, 1, 2, \ldots, M-1\}$. It means that $\text{rank}(\mathbf{A}_{2K}) = 2K$ for all $\mathbf{A}_{2K}$. Based on the previous analysis, we can again state that the solution for a K sparse problem is unique if

$$\text{spark}\{\mathbf{A}\} > 2K.$$

Note that for any square matrix its determinant is equal to the product of its eigenvalues $\det\{\mathbf{A}_{2K}\} = d_1 d_2 \cdot \ldots \cdot d_{2K}$. The uniqueness condition $\text{rank}(\mathbf{A}_{2K}) = 2K$ can be rewritten as

$$\min_i |d_i| > 0$$

for all submatrices $\mathbf{A}_{2K}$ of the measurement matrix $\mathbf{A}$.

If a matrix $\mathbf{A}_{2K}$ is of an order of $M \times 2K$ then the rank of this matrix can be checked by considering $\mathbf{A}_{2K}^T \mathbf{A}_{2K}$. It means that there is no need for combinations over measurements (to form a quadratic matrix $2K \times 2K$ from $M \times 2K$ matrix) if $M > 2K$ since the rank of $M \times 2K$ matrix $\mathbf{A}_{2K}$ can be checked by calculating the rank of a $2K \times 2K$ matrix $\mathbf{A}_{2K}^T \mathbf{A}_{2K}$ using

$$\text{rank}(\mathbf{A}_{2K}) = \text{rank}(\mathbf{A}_{2K}^T \mathbf{A}_{2K}).$$

In numerical and practical applications we would not be satisfied if any of the determinant $\det\{\mathbf{A}_{2K}\}$ or $\det\{\mathbf{A}_{2K}^T \mathbf{A}_{2K}\}$ is close to zero. In this case, the theoretical condition for a unique solution would be satisfied, however the analysis and possible inversion would be highly sensitive to noise in the measurements. Thus, a practical requirement is that the determinant is not just different from zero, but that it sufficiently differs from zero so that an inversion stability and noise robustness is achieved.

Restricted Isometry Property: From the matrix theory it is known that the norm of a matrix $\mathbf{A}_{2K}$ satisfies

$$\lambda_{\min} \leq \frac{\|\mathbf{A}_{2K}\mathbf{X}_{2K}\|_2^2}{\|\mathbf{X}_{2K}\|_2^2} = \frac{\mathbf{X}_{2K}^T \mathbf{A}_{2K}^T \mathbf{A}_{2K} \mathbf{X}_{2K}}{\mathbf{X}_{2K}^T \mathbf{X}_{2K}} \leq \lambda_{\max} \tag{15}$$

where $\lambda_{\min}$ and $\lambda_{\max}$ are the minimal and the maximal eigenvalue of the (Gram) matrix $\mathbf{A}_{2K}^T \mathbf{A}_{2K}$ and $\|\mathbf{X}\|_2^2 = |X(0)|^2 + \ldots |X(N-1)|^2$ is the squared ℓ_2-norm (norm-two) of $\mathbf{X}$.

The isometry property for a linear transformation matrix $\mathbf{A}$ holds if

$$\|\mathbf{A}\mathbf{X}\|_2^2 = \|\mathbf{X}\|_2^2 \quad \text{or} \quad \frac{\|\mathbf{A}\mathbf{X}\|_2^2}{\|\mathbf{X}\|_2^2} = 1$$

The restricted isometry property (RIP) for a matrix $\mathbf{A}_{2K}$ holds if

$$1 - \delta_{2K} \leq \frac{\|\mathbf{A}_{2K}\mathbf{X}_{2K}\|_2^2}{\|\mathbf{X}_{2K}\|_2^2} \leq 1 + \delta_{2K} \tag{16}$$

for any $2K$-sparse vector $\mathbf{X}_{2K}$, where $0 \leq \delta_{2K} < 1$ is the isometric constant [7, 15, 24, 28, 35, 116]. From Eqs.(15) and (16) we can write

$$\delta_{2K} = \max\{1 - \lambda_{\min},\ \lambda_{\max} - 1\}.$$

Commonly, the isometric constant is defined by $\lambda_{\max} - 1$ and it is calculated as the maximal eigenvalue of matrix $\mathbf{A}_{2K}^T\mathbf{A}_{2K} - \mathbf{I}$. Normalized energies of the columns of matrix $\mathbf{A}$ (that are equal to the diagonal elements of $\mathbf{A}_{2K}^T\mathbf{A}_{2K}$) are assumed. Otherwise, the normalization factors should be added. For complex-valued matrices the Hermitian transpose should be used in $\mathbf{A}_{2K}^H\mathbf{A}_{2K}$.

For a K-sparse vector $\mathbf{X}$ and a measurement matrix $\mathbf{A}$ the RIP is satisfied if relation (16) holds for all submatrices $\mathbf{A}_K$ with $0 \leq \delta_K < 1$. The solution for K-sparse vector is unique if the measurement matrix satisfies the RIP for $2K$-sparse vector $\mathbf{X}$ with $0 \leq \delta_{2K} < 1$. Note that if the RIP is satisfied then $\lambda_{\min} > 0$. It means that all $\mathbf{A}_{2K}^T\mathbf{A}_{2K}$ submatrices are nonsingular.

The restricted isometry property for small δ_{2K} is closer to the isometry property and improves the solution stability. It can be related to the matrix conditional number. The conditional number of a matrix $\mathbf{A}_{2K}^T\mathbf{A}_{2K}$ is defined as the ratio of its maximal and minimal eigenvalues

$$\operatorname{cond}\left\{\mathbf{A}_{2K}^T\mathbf{A}_{2K}\right\} = \frac{\lambda_{\max}}{\lambda_{\min}}.$$

If a matrix $\mathbf{A}_{2K}$ satisfies the restricted isometry property with δ_{2K} then

$$\operatorname{cond}\left\{\mathbf{A}_{2K}^T\mathbf{A}_{2K}\right\} \leq \frac{1+\delta_{2K}}{1-\delta_{2K}}.$$

With small values of δ_{2K} the conditional number is close to one, meaning stable invertibility and low sensitivity to the input noise (small variations of the input signal (measurements) do not cause large variations of the result).

Incoherence condition: The mutual coherence (coherence index) of a matrix $\mathbf{A}$ is defined as the maximal absolute value of the normalized scalar product of its two columns

$$\mu = \max|\mu(m,k)|, \text{ for } m \neq k$$

where

$$\mu(m,k) = \frac{\sum_{i=0}^{M-1}\alpha_m(i)\alpha_k^*(i)}{\sum_{i=0}^{M-1}|\alpha_k(i)|^2} \tag{17}$$

and $\alpha_k(i)$ are the elements of the kth column of matrix $\mathbf{A}$. Note that $\mu(m,k)$ are the off-diagonal elements of matrix $\mathbf{A}^H\mathbf{A}$ normalized with the corresponding diagonal elements. This index plays an important role in the analysis of measurement matrices. The coherence index should be as small as possible (meaning high incoherence). With smaller values of coherence index the matrix $\mathbf{A}^H\mathbf{A}$ is closer to the identity matrix.

The reconstruction of a K-sparse vector $\mathbf{X}$ from M measurements is unique if

$$K < \frac{1}{2}\left(1 + \frac{1}{\mu}\right). \tag{18}$$

The presented reconstruction condition based on the coherence index will be derived in Section 3.3.

The coherence index μ cannot be arbitrarily small for an $M \times N$ matrix $\mathbf{A}$ ($M < N$). The Welch upper bound relation holds

$$\mu \geq \sqrt{\frac{N-M}{M(N-1)}}. \tag{19}$$

We can use the mutual coherence to determine the spark lower bound

$$\operatorname{spark}(\mathbf{A}) \geq \left(1 + \frac{1}{\mu}\right). \tag{20}$$

If $\mathbf{X}$ is a solution of a system of equations $\mathbf{y} = \mathbf{A}\mathbf{X}$ such that

$$\|\mathbf{X}\|_0 = K < \frac{1}{2}\left(1 + \frac{1}{\mu}\right) \leq \frac{1}{2}\operatorname{spark}(\mathbf{A}), \tag{21}$$

then $\mathbf{X}$ is necessarily the sparsest possible solution.

The RIP constant can be related to the coherence index as

$$\delta_K \leq (1-K)\mu.$$

This is a pre-print of an article published in Circuits Systems and Signal Processing.
The final authenticated version is available online at: https://doi.org/10.1007/s00034-018-0909-2

3 Norm-Zero Based Reconstruction

The formal signal reconstruction approach based on the signal sparsity states that the signal can be reconstructed from its measurements defined by vector $\mathbf{y}$, by finding the sparsest vector $\mathbf{X}$ that corresponds to the measurements $\mathbf{y}$. Hence, by introducing the notation for the number of components based on the ℓ_0-norm $K = \|\mathbf{X}\|_0$, the fundamental minimization problem can be formulated as:

$$\min \|\mathbf{X}\|_0 \text{ subject to } \mathbf{y} = \mathbf{A}\mathbf{X}. \tag{22}$$

Generally, the ℓ_0-norm is not suitable for minimization methods. However, a class of algorithms is based on the minimization of the number of coefficients $K = \|\mathbf{X}\|_0$ in an implicit way. For instance, in certain applications we can predict the number of components or we are able to estimate the position of nonzero coefficients. In this way the computational complexity will be significantly reduced as compared to the direct search method [58, 111].

3.1 Reconstruction with Known or Estimated Positions

Let us initially assume that the positions of nonzero coefficients are already estimated and known. By denoting the set of these positions by $\mathbb{K} = \{k_1, k_2, \ldots, k_K\}$ and the vector of unknown coefficients by $\mathbf{X}_K = [X(k_1), X(k_2), \ldots, X(k_K)]^T$, we reduce the undetermined system $\mathbf{y} = \mathbf{A}\mathbf{X}$ to the overdetermined system

$$\mathbf{y} = \mathbf{A}_K \mathbf{X}_K, \tag{23}$$

where $\mathbf{A}_K$ is an $M \times K$ matrix ($K < M$) obtained from the measurements matrix $\mathbf{A}$ with columns corresponding to the zero-valued transform coefficients being omitted. Its solution, in the least mean-squared sense, follows from the minimization of difference of the available measurements and their values produced by the reconstructed coefficients $\mathbf{X}_K$, defined by

$$\begin{aligned} e^2 &= \|\mathbf{y} - \mathbf{A}_K \mathbf{X}_K\|_2^2 = (\mathbf{y} - \mathbf{A}_K \mathbf{X}_K)^H (\mathbf{y} - \mathbf{A}_K \mathbf{X}_K) \\ &= \|\mathbf{y}\|_2^2 - 2\mathbf{X}_K^H \mathbf{A}_K^H \mathbf{y} + \mathbf{X}_K^H \mathbf{A}_K^H \mathbf{A}_K \mathbf{X}_K, \end{aligned} \tag{24}$$

where H denotes the Hermitian transpose conjugate. Using symbolic derivation over the vector of unknowns, we obtain the minimum of error e^2 from

$$\frac{\partial e^2}{\partial \mathbf{X}_K^H} = -2\mathbf{A}_K^H \mathbf{y} + 2\mathbf{A}_K^H \mathbf{A}_K \mathbf{X}_K = 0$$

with its solution being equal to

$$\mathbf{X}_K = \left(\mathbf{A}_K^H \mathbf{A}_K\right)^{-1} \mathbf{A}_K^H \mathbf{y} = \text{pinv}(\mathbf{A}_K)\mathbf{y}. \tag{25}$$

The pseudoinverse of matrix $\mathbf{A}_K$ is defined by $\text{pinv}(\mathbf{A}_K) = \left(\mathbf{A}_K^H \mathbf{A}_K\right)^{-1} \mathbf{A}_K^H$.

3.2 Estimation of Nonzero Coefficient Positions

The positions of nonzero coefficients in $\mathbf{X}$ are not known in general. These positions can be estimated using the *matching pursuit* (MP) approach to the sparse signal reconstruction. The *orthogonal matching pursuit* (OMP) is an improved greedy reconstruction algorithm compared to the MP. The reconstruction of nonzero coefficients $\mathbf{X}$ is iteratively updated by projecting the vector of measurements $\mathbf{y}$ onto the columns of matrix $\mathbf{A}$ that corresponds to the current support set (see Algorithm 1). The idea behind this approach is described next.

We begin with an initial estimate given by:

$$\mathbf{X}_0 = \mathbf{A}^H \mathbf{y} = \mathbf{A}^H \mathbf{A}\mathbf{X}. \tag{26}$$

An intuitive idea for the initial estimate comes from the fact that the measurements are obtained as linear combinations of the sparsity domain coefficients with the rows of the measurement matrix $\mathbf{A}$ as

This is a pre-print of an article published in Circuits Systems and Signal Processing.
The final authenticated version is available online at: https://doi.org/10.1007/s00034-018-0909-2

Algorithm 1 Norm-zero based OMP reconstruction

Input:

- Measurement vector $\mathbf{y}$
- Measurement matrix $\mathbf{A}$
- Number of selected coefficients in each iteration r, by default $r = 1$
- Required precision ε

$\mathbb{K} \leftarrow \emptyset$
$\mathbf{e} \leftarrow \mathbf{y}$
while $\|\mathbf{e}\|_2 > \varepsilon$ **do**
  $(k_1, k_2, \ldots, k_r) \leftarrow$ positions of r highest values in $\mathbf{A}^H\mathbf{e}$
  $\mathbb{K} \leftarrow \mathbb{K} \cup \{k_1, k_2, \ldots, k_r\}$
  $\mathbf{A}_K \leftarrow$ columns of matrix $\mathbf{A}$ selected by set $\mathbb{K}$
  $\mathbf{X}_K \leftarrow \text{pinv}(\mathbf{A}_K)\mathbf{y}$
  $\mathbf{y}_K \leftarrow \mathbf{A}_K\mathbf{X}_K$
  $\mathbf{e} \leftarrow \mathbf{y} - \mathbf{y}_K$
end while
$\mathbf{X} \leftarrow \begin{cases} \mathbf{0} & \text{for positions not in } \mathbb{K} \\ \mathbf{X}_K & \text{for positions in } \mathbb{K} \end{cases}$

Output:

- Reconstructed signal coefficients $\mathbf{X}$

weights. It means that the projection of the measurements $\mathbf{y}$ to the measurement matrix $\mathbf{A}$ can be used to estimate the positions of nonzero coefficients. In an ideal case matrix $\mathbf{A}^H\mathbf{A}$ should be an identity matrix and the initial estimate $\mathbf{X}_0$ would fully correspond to the exact solution $\mathbf{X}$. However with a reduced number of measurements this cannot be achieved. Still the requirement is that the diagonal elements of $\mathbf{A}^H\mathbf{A}$ are as large as possible with respect to the other nondiagonal elements.

The first element of the set $\mathbb{K}$ is estimated as the position of $\mathbf{X}_0$ maximum

$$k_1 = \arg\max |\mathbf{A}^H\mathbf{y}|.$$

Then, the system (23) is solved to find minimum of $\|\mathbf{y} - \mathbf{A}_K\mathbf{X}_K\|_2^2$ with respect to $\mathbf{X}_K$ by using Eq.(25) with $\mathbb{K} = \{k_1\}$. The reconstructed vector $\mathbf{y}_1 = \mathbf{A}_1\mathbf{X}_1$ is calculated. If $\mathbf{y}_1 = \mathbf{y}$ the signal is of sparsity $K = 1$ and $\mathbf{X}_1$ is the problem's solution. If this is not the case the estimated component is removed from $\mathbf{y}$ and the signal $\mathbf{e}_1 = \mathbf{y} - \mathbf{y}_1$ is formed.

Then the second nonzero position is estimated as

$$k_2 = \arg\max |\mathbf{A}^H\mathbf{e}_1|$$

and the set $\mathbb{K} = \{k_1, k_2\}$ is formed. It is very important to use both coefficients $X(k_1)$ and $X(k_2)$ in this calculation and to reestimate the first coefficient $X(k_1)$ as well. We repeat the process with $K = 2$ as outlined above to obtain the reconstruction $\mathbf{X}_2$, measurements $\mathbf{y}_2$, and the error vector $\mathbf{e}_2 = \mathbf{y} - \mathbf{y}_2$. If $\mathbf{e}_2$ is a zero vector, the solution is found as $\mathbf{y} = \mathbf{y}_2$. If this is not the case, we will continue to estimate k_3 using $\mathbf{e}_2$ and find $\mathbb{K} = \{k_1, k_2, k_3\}$, followed by $\mathbf{X}_3$ and $\mathbf{y}_3$ calculations.

The procedure is continued until zero (or acceptable) error is achieved.

Example 1. Signal

$$x(t) = 0.3 + 0.4\exp(-j12\pi t) + 0.5\exp(j40\pi t)$$

is considered within $0 \leq t < 1$. This signal is uniformly sampled at $t_n = n\Delta t = n/N$, with $N = 64$. A random subset $\mathbf{y}$ of $M = 16$ samples $y(i) = x(n_i)$ at n_i, $i = 1, 2, \ldots, M$, is available, Fig.1 (first row, left). The measurement matrix $\mathbf{A}$ is a partial inverse DFT matrix corresponding to the samples defined by time indices n_i, $i = 1, 2, \ldots, M$. Since the inverse DFT matrix is scaled with $1/N$ the initial estimate in the DFT matrix case is calculated as $\mathbf{X}_0 = N\mathbf{A}^H\mathbf{y}$.

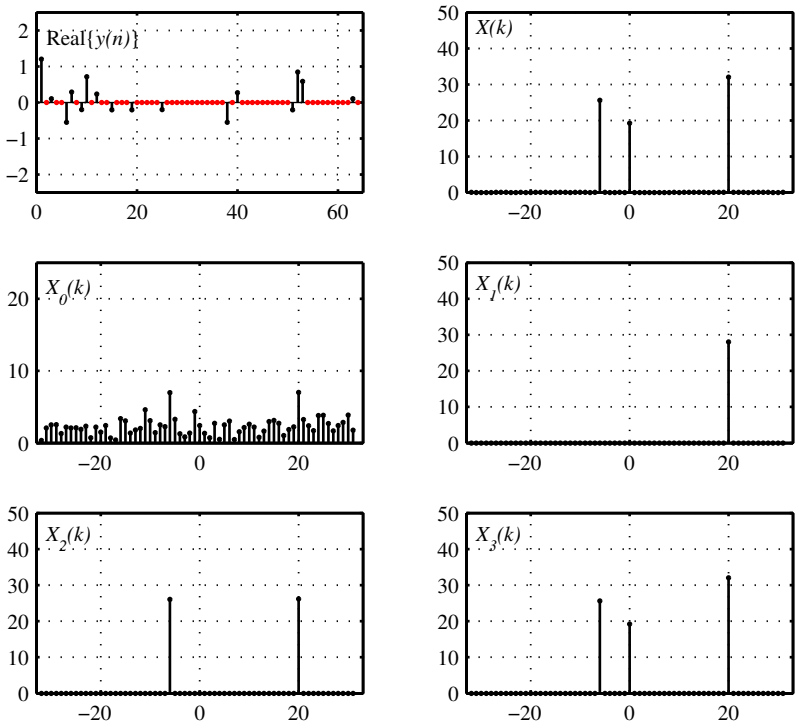

Figure 1: Iterative signal reconstruction: Available samples $y(n)$; Exact DFT coefficients $X(k)$; Initial estimation of the DFT coefficients $X_0(k)$; The estimated DFT coefficient in the first iteration $X_1(k)$; The estimated DFT coefficients in the second, and the third iteration, denoted by $X_2(k)$, and $X_3(k)$, respectively.

The iterative procedure for nonzero coefficient indices determination is used. The recovered signal is calculated for detected DFT positions during the iterations according to the presented algorithm. The recovered DFT values in the rth iteration are denoted by $X_r(k)$ and presented in Fig.1. After first iteration, the strongest component in the initial transform is detected and its amplitude is estimated. Its value differs from the true one, since the other four undetected components degrade the estimation accuracy. After three steps, all components are detected, respectively. In this final iteration matching between the reconstructed signal and the available signal samples is achieved and the algorithm is stopped. The DFT of the recovered signal is presented as $X_3(k)$ in the last subplot of Fig.1. Its values are the same as in the DFT of the original signal, Fig.1 (first row, right). A simple and efficient uniqueness check in this case is proposed [88]. □

3.3 Coherence Index Interpretation

One way to estimate the positions of nonzero coefficients (and to solve the signal reconstruction problem) is based on using the initial estimate $\mathbf{X}_0 = \mathbf{A}^H\mathbf{y} = \mathbf{A}^H\mathbf{A}\mathbf{X}$. Since the ideal (identity matrix) form of $\mathbf{A}^H\mathbf{A}$ can not be achieved, a measure of closeness to the identity matrix can be the ratio of the maximal value of nondiagonal elements and the diagonal elements of this matrix. Now we will analyze the elements of $\mathbf{A}^H\mathbf{A}$. The diagonal elements are

$$a_{kk} = \sum_{i=0}^{M-1} \alpha_k(i)\alpha_k^*(i) = \langle \alpha_k, \alpha_k^* \rangle ,$$

while the elements outside the diagonal are

$$a_{mk} = \sum_{i=0}^{M-1} \alpha_m(i)\alpha_k^*(i) = \langle \alpha_m, \alpha_k^* \rangle , \qquad m \neq k,$$

This is a pre-print of an article published in Circuits Systems and Signal Processing.
The final authenticated version is available online at: https://doi.org/10.1007/s00034-018-0909-2

where $\langle \alpha_m, \alpha_k^* \rangle$ denotes the scalar product of α_m and α_k^*.

Consider a K-sparse signal coefficients $\mathbf{X}$. Their measurements are $y(m) = \sum_{k \in \mathbb{K}} X(k)\alpha_k(m)$, with $\mathbb{K} = \{k_1, k_2, \ldots, k_K\}$. The worst case of influence of other components to the strongest component in the initial estimate occurs when the remaining $K-1$ components are equally strong (with equal amplitudes $X(k) = 1$, $k \in \{k_1, k_2, \ldots, k_K\}$). Then the amplitude of the strongest component is 1 (assuming normalized measurement matrix with $a_{kk} = 1$ and normalized signal component amplitude). The worst case for the detection of this component is when all other components maximally reduce the value of considered component in the initial estimate $X_0(i) = \sum_{k \in \mathbb{K}} \sum_{m=0}^{M-1} \alpha_k(m)\alpha_i^*(m)$. The influence of the kth component on the mth position is $\sum_{m=0}^{M-1} \alpha_k(m)\alpha_i^*(m)$. Its maximal possible value is μ. In the worst case the amplitude of the considered strongest component is then $1-(K-1)\mu$. At the position where there is no component in $X(k)$ the maximal possible contributions of all K components (being equal to μ) sum up in phase to produce the maximal possible disturbance $K\mu$. Detection of the strongest component is always successful if the worst possible amplitude, at the considered component position, is greater than the largest possible disturbance, [84],

$$1-(K-1)\mu > K\mu$$

producing the condition given in Eq.(21):

$$K < \frac{1}{2}\left(1 + \frac{1}{\mu}\right). \tag{27}$$

After the first component is successfully detected then the previous relation will certainly hold for $K-1$ sparsity, guaranteeing the exact (unique) solution.

3.4 CoSaMP Reconstruction Algorithm

A slightly different reconstruction approach is based on the compressive sampling matching pursuit (CoSaMP). CoSaMP is an algorithm where a signal with a desired sparsity K is obtained in an iterative way [65]. The measurement vector $\mathbf{y}$ is projected onto the columns of the measurement matrix $\mathbf{A}$ and $2K$ positions with the highest projection magnitudes are selected. The set of the selected positions is expanded with nonzero positions in the current estimate of the sparse signal $\mathbf{X}$. A solution in the least squared sense is found and K coefficients with the highest magnitudes are selected as the reconstructed signal $\mathbf{X}$. The measurement vector is adjusted by subtracting the current solution and the iterative procedure is repeated. The CoSaMP reconstruction procedure is presented in Algorithm 2. The procedure can be performed with a predefined number of iterations or by obtaining a desired norm value of the residual measurement vector $\mathbf{e}$.

3.5 Noise in the Initial Estimate

Consider a K-sparse vector $\mathbf{X}$ with elements $X(k)$, $k = 0,1,2,\ldots,N-1$, whose nonzero values are at $\mathbb{K} = \{k_1, k_2, \ldots, k_K\}$. The measurements of this sparse vector are

$$y(m) = \sum_{k \in \mathbb{K}} C_k \alpha_k(m), \quad m = 0,1,\ldots,M-1, \tag{28}$$

where $X(k) = C_k$ are unknown nonzero elements. The initial estimate $\mathbf{X}_0 = \mathbf{A}^H\mathbf{y}$ at a position $i \notin \mathbb{K}$ (non-component positions) is

$$X_0(i) = \sum_{k \in \mathbb{K}} C_k \sum_{m=0}^{M-1} \alpha_k(m)\alpha_i^*(m).$$

For $k \neq i$ and common measurement/basis functions, the expected value of the scalar product is zero,

$$\mathrm{E}\Big\{ \sum_{m=0}^{M-1} \alpha_k(m)\alpha_i^*(m) \Big\} = 0.$$

For a random set of $1 \ll M \ll N$ available measurements, according to the central limit theorem we can also consider that the variable $X_0(i)$ is Gaussian. As a consequence of having a reduced number M

Algorithm 2 CoSaMP reconstruction algorithm

Input:

- Measurement vector $\mathbf{y}$
- Measurement matrix $\mathbf{A}$
- Desired sparsity K

1: $\mathbf{X} \leftarrow \mathbf{0}_{N\times 1}$
2: $\mathbf{e} \leftarrow \mathbf{y}$
3: **repeat**
4: $\quad \mathbb{T}_1 \leftarrow$ positions of $2K$ highest values in $\mathbf{A}^H\mathbf{e}$
5: $\quad \mathbb{T}_2 \leftarrow$ positions of nonzero coefficients in $\mathbf{X}$
6: $\quad \mathbb{T} \leftarrow \mathbb{T}_1 \cup \mathbb{T}_2$
7: $\quad \mathbf{A}_T \leftarrow$ columns from matrix $\mathbf{A}$ selected by set $\mathbb{T}$
8: $\quad \mathbf{b} \leftarrow \text{pinv}(\mathbf{A}_T)\mathbf{y}$
9: $\quad$ Put K coefficients with highest magnitude from $\mathbf{b}$ to the corresponding positions in $\mathbf{X}$ and set remaining coefficients to zero.
10: $\quad \mathbf{e} \leftarrow \mathbf{y} - \mathbf{A}\mathbf{X}$
11: **until** stopping criterion is satisfied

Output:

- Reconstructed K-sparse signal vector $\mathbf{X}$

of available values in $\mathbf{y}$ the initial estimate can be considered as a noisy version of the true coefficients $X(k)$. For $i \in \mathbb{K}$ their mean-value is

$$E\{X_0(i)\} = E\left\{\sum_{k\in\mathbb{K}} C_k \sum_{m=0}^{M-1} \alpha_k(m)\alpha_i^*(m)\right\} = C_i \sum_{m=0}^{M-1} \alpha_i(m)\alpha_i^*(m).$$

The variance of the random variable $X_0(i)$ can be calculated for various basis/measurement functions [103]. It corresponds to the noise of variance σ^2 in the transform domain representation $\mathbf{X}$. This noise depends on the number of missing samples.

For example, for the partial DFT matrix, with the DFT coefficients C_k, and the initial estimate in the DFT matrix case calculated as $\mathbf{X}_0 = N\mathbf{A}^H\mathbf{y}$, the mean value is obtained as:

$$\text{E}\{X_0(i)\} = MC_i. \tag{29}$$

The coefficients C_i corresponds to the signal amplitude in the time domain. The variance at the non-component positions $i \notin \mathbb{K}$ is given by [84, 96]

$$\sigma^2 = \text{var}\{X_0(i)\} = \sum_{k\in\mathbb{K}} |C_k|^2 M\frac{N-M}{N-1}. \tag{30}$$

The value $\sum_{k\in\mathbb{K}} |C_k|^2$ can be estimated as the mean energy of the measurements/available samples using $\frac{1}{M}\sum_{m=0}^{M-1} |y(m)|^2$.

When all signal samples are available, $M = N$ in the case of the DFT as sparsity domain, the variance of this noise will be zero.

3.6 Single Iteration Reconstruction

The estimated variance can be further used to estimate the positions of the K signal components. Namely, we can define the probability that all $(N-K)$ nonsignal components are below a certain threshold τ as:

$$P(\tau) = \left(1 - \exp(-\frac{\tau^2}{\sigma^2})\right)^{N-K}. \tag{31}$$

Algorithm 3 SIRA - Single iteration reconstruction algorithm

Input:

- Measurement vector $\mathbf{y}$
- Measurement matrix $\mathbf{A}$
- Probability P for components detection
- Variance σ^2

1: $\tau \leftarrow \sqrt{-\sigma^2 \log(1-P^{\frac{1}{N}})}$
2: $\mathbf{X}_0 = \mathbf{A}^H \mathbf{y}$
3: $\mathbb{K} \leftarrow \{k : |X_0(k)| > \tau\}$
4: $\mathbf{A}_K \leftarrow$ columns of matrix $\mathbf{A}$ selected by set $\mathbb{K}$
5: $\mathbf{X}_K \leftarrow \text{pinv}(\mathbf{A}_K)\mathbf{y}$
6: $\mathbf{X} \leftarrow \begin{cases} \mathbf{0} & \text{for positions not in } \mathbb{K} \\ \mathbf{X}_K & \text{for positions in } \mathbb{K} \end{cases}$

Output:

- Reconstructed signal coefficients $\mathbf{X}$

In order to estimate the positions of signal components, we may fix the value for $P(\tau) = P$ and calculate the threshold according to:

$$\tau = \sqrt{-\sigma^2 \log(1-P^{\frac{1}{N-K}})} \approx \sigma \sqrt{-\log(1-P^{\frac{1}{N}})}. \tag{32}$$

Based on the threshold derived from the variance of the random variable $X_0(i)$, we can define a simple single-iteration reconstruction algorithm (SIRA), [67, 68, 99, 103, 106] (Algorithm 3).

Example 2. Signal

$$x(t) = \sin\left(12\pi t + \frac{\pi}{3}\right) + 0.6\cos\left(32\pi t + \frac{\pi}{4}\right)$$

is considered within $0 \leq t < 1$. This signal is uniformly sampled at $t_n = n\Delta t = n/N$, with $N = 64$. A random subset $\mathbf{y}$ of $M = 32$ samples $y(i) = x(n_i)$ at n_i, $i = 1,2,\ldots,M$, at n_i, $i = 1,2,\ldots,M$ is available, Fig.2 (first row, left). The measurement matrix $\mathbf{A}$ is a partial DFT matrix corresponding to the samples defined by time indices n_i, $i = 1,2,\ldots,M$.

The one-step procedure for nonzero coefficient indices determination is used. The threshold $\tau = 9.5$ is calculated with $P = 0.99$ using (32), while σ^2 is estimated based on the mean energy of the available samples and (30). The recovered signal is calculated for detected DFT positions. The recovered DFT values are denoted as $X_r(k)$ and presented in Fig.2 (second row, right). Its values are the same as in the DFT of the original signal, Fig.2 (first row, right). A simple and efficient uniqueness check in this case is proposed [88]. □

3.7 Additive Noise

Assume now that an input additive noise $\varepsilon(m)$ exists in the measurements

$$\mathbf{y} + \boldsymbol{\varepsilon} = \mathbf{A}\mathbf{X}.$$

The equations for noisy measurements and detected positions of sparsity coefficients $k \in \mathbb{K} = \{k_1, k_2, ..., k_K\}$ are

$$\mathbf{y} + \boldsymbol{\varepsilon} = \mathbf{A}_K \mathbf{X}_R, \tag{33}$$

where $\mathbf{X}_R$ is the resulting noisy K sparse vector. The solution is obtained as

$$\mathbf{X}_R = \left(\mathbf{A}_K^H \mathbf{A}_K\right)^{-1} \mathbf{A}_K^H (\mathbf{y} + \boldsymbol{\varepsilon}) = \mathbf{X}_K + \mathbf{X}_N \tag{34}$$

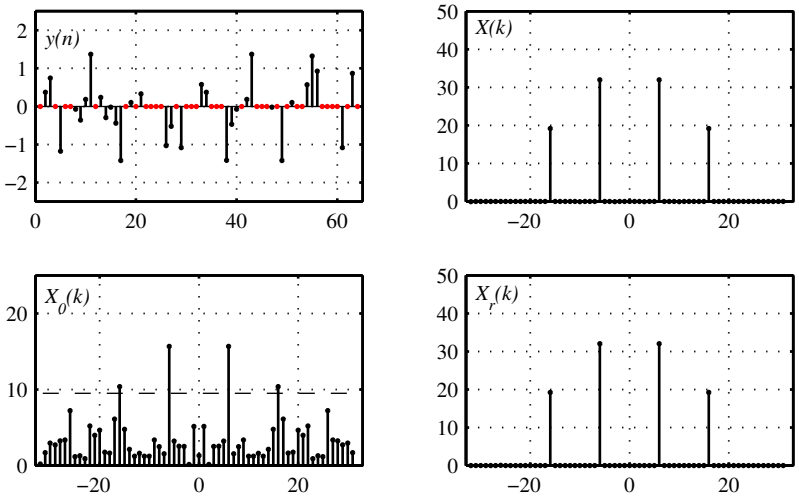

Figure 2: Single iteration reconstruction: Available samples $y(n)$; Exact DFT coefficients $X(k)$; Initial estimation of the DFT coefficients $X_0(k)$; The estimated DFT coefficients $X_r(k)$.

where $\mathbf{X}_R = \mathbf{X}_K + \mathbf{X}_N$. The true coefficients are $\mathbf{X}_K = \left(\mathbf{A}_K^H \mathbf{A}_K\right)^{-1} \mathbf{A}_K^H \mathbf{y}$ and $\mathbf{X}_N = \left(\mathbf{A}_K^H \mathbf{A}_K\right)^{-1} \mathbf{A}_K^H \boldsymbol{\varepsilon}$ is the noise influence on the reconstructed coefficients.

From (33) we can write $\|\mathbf{y}\|_2 = \|\mathbf{A}_K \mathbf{X}_K\|_2$ and $\|\boldsymbol{\varepsilon}\|_2 = \|\mathbf{A}_K \mathbf{X}_N\|_2$. If $\|\mathbf{A}_K \mathbf{X}_K\|_2$ is normalized with $\|\mathbf{X}_K\|_2$ and $\|\mathbf{A}_K \mathbf{X}_N\|_2$ with $\|\mathbf{X}_N\|_2$ the matrix $\mathbf{A}_K$ norms are obtained. The ratio of these norms is

$$\frac{\frac{\|\mathbf{y}\|_2}{\|\mathbf{X}_K\|_2}}{\frac{\|\boldsymbol{\varepsilon}\|_2}{\|\mathbf{X}_N\|_2}} = \frac{\frac{\|\mathbf{A}_K \mathbf{X}_K\|_2}{\|\mathbf{X}_K\|_2}}{\frac{\|\mathbf{A}_K \mathbf{X}_N\|_2}{\|\mathbf{X}_N\|_2}} \leq \text{cond}(\mathbf{A}_K) = \sqrt{\frac{d_{\max}}{d_{\min}}}$$

This ratio is used for the condition number definition. Its maximal value is $\sqrt{d_{\max}/d_{\min}}$ and the minimal value of this ratio is $\sqrt{d_{\min}/d_{\max}}$, where $d_{\max}$, and $d_{\min}$ are the maximal and the minimal eigenvalue of $\mathbf{A}_K^H \mathbf{A}_K$, respectively. Therefore, the solution of noisy system (33) satisfies

$$\frac{d_{\min}}{d_{\max}} \frac{\|\mathbf{y}\|_2^2}{\|\boldsymbol{\varepsilon}\|_2^2} \leq \frac{\|\mathbf{X}_K\|_2^2}{\|\mathbf{X}_N\|_2^2} \leq \frac{d_{\max}}{d_{\min}} \frac{\|\mathbf{y}\|_2^2}{\|\boldsymbol{\varepsilon}\|_2^2}$$

or in [dB]

$$SNR_i - 10\log(\frac{d_{\max}}{d_{\min}}) \leq SNR \leq SNR_i + 10\log(\frac{d_{\max}}{d_{\min}}),$$

where the input and output signal-to-noise ratios are defined as $SNR_i = 10\log(\|\mathbf{y}\|_2^2 / \|\boldsymbol{\varepsilon}\|_2^2)$ and $SNR = 10\log(\|\mathbf{X}_K\|_2^2 / \|\mathbf{X}_N\|_2^2)$. We assumed normalized columns energy of $\mathbf{A}$ for notation simplicity.

For small additive noise, a simplified analysis can be performed. It will produce the exact signal-to-noise value. The signal-to-noise (SNR) ratio, for a full set of signal samples $x(n)$ is defined as:

$$SNR_i = 10\log \frac{\sum_{n=0}^{N-1} |x(n)|^2}{\sum_{n=0}^{N-1} |\varepsilon(n)|^2}.$$

In the case of a reduced set of measurements $y(m)$, $m \in \mathbb{M}$, the noise energy is reduced to:

$$E_{\varepsilon A} = \sum_{n \in \mathbb{M}} |\varepsilon(n)|^2.$$

The correct amplitude in the signal transform at an index i, in the case as if all signal samples were available, would be NC_i in (29), where C_i is the amplitude of the of i-th signal component. Hence, in order to compensate the resulting transform for the known bias in amplitude when only M samples are

available, given by Eq.(29), the coefficients should be multiplied by N/M. The noise is multiplied by the same factor, while the energy of noise in the reconstructed signal is increased to $E_{\varepsilon A}N^2/M^2$. The SNR in the recovered signal is

$$SNR = 10\log\frac{\sum_{n=0}^{N-1}|x(n)|^2}{\frac{N^2}{M^2}\sum_{n\in\mathbb{M}}|\varepsilon(n)|^2}. \tag{35}$$

Since only K out of N coefficients are used in the reconstruction, the energy of the reconstruction error is reduced by the factor of K/N as well, [103,104]. The energy of noise in the recovered signal is

$$E_{\varepsilon R} = \frac{K}{N}\frac{N^2}{M^2}\sum_{n\in\mathbb{M}}|\varepsilon(n)|^2 = \frac{K}{M}\sum_{n=0}^{N-1}|\varepsilon(n)|^2.$$

Using the fact that the variances in measurements are the same $\frac{1}{M}\sum_{n\in\mathbb{M}}|\varepsilon(n)|^2 = \frac{1}{N}\sum_{n=0}^{N-1}|\varepsilon(n)|^2$, the SNR in the recovered signal is

$$SNR = SNR_i - 10\log\left(\frac{K}{M}\right) \tag{36}$$

It means that assuming smaller K in the reconstruction (ideally equal to the signal sparsity), may improve the results in the case of additive input noise.

3.8 Nonsparse Signal Reconstruction

According to the results in Section 3.5, the missing samples can be represented as noise in the initial estimate. Assume that we use a reconstruction algorithm for a signal of sparsity K on a signal whose DFT coefficients $\mathbf{X}$ are not sparse (or not sufficiently sparse). The sparse signal with K nonzero coefficients equal to the largest K coefficients of $\mathbf{X}$ is denoted by $\mathbf{X}_K$. Suppose that the number of components K and the measurements matrix satisfy the reconstruction conditions so that a reconstruction algorithm can detect (one by one or all at once) the largest K components whose values are (C_1, C_2,...,C_K) and perform signal reconstruction to get $\mathbf{X}_R$. The remaining $N-K$ components (C_{K+1},C_{K+2},...,C_N) will be treated as a noise in these K largest components. Variance from a signal component, according to (30), is $|C_i|^2 M(N-M)/(N-1)$. After the reconstruction this variance is multiplied by $(N/M)^2$, according to the analysis in previous subsection, producing

$$|C_i|^2\frac{N^2}{M^2}\frac{M(N-M)}{N-1} \cong |C_i|^2 N\frac{N-M}{M}.$$

The total energy of noise in the K largest reconstructed components $\mathbf{X}_R$ will be

$$\|\mathbf{X}_R - \mathbf{X}_K\|_2^2 = KN\frac{N-M}{M}\sum_{i=K+1}^{N}|C_i|^2.$$

Denoting the energy of remaining signal, when the K largest components are removed from the original signal, by

$$\|\mathbf{X}-\mathbf{X}_K\|_2^2 = \sum_{i=K+1}^{N}|NC_i|^2$$

we get

$$\|\mathbf{X}_R - \mathbf{X}_K\|_2^2 = K\frac{N-M}{MN}\|\mathbf{X}-\mathbf{X}_K\|_2^2. \tag{37}$$

If the signal is sparse, i.e., $\mathbf{X}=\mathbf{X}_K$, then $\|\mathbf{X}_R-\mathbf{X}_K\|_2^2 = 0$. The same result follows if $N=M$.

In the case of additive input noise in measurements, with a variance σ_ε^2, a general expression is obtained in the form [84,89]

$$\|\mathbf{X}_R - \mathbf{X}_K\|_2^2 = K\frac{N-M}{MN}\|\mathbf{X}-\mathbf{X}_K\|_2^2 + \frac{K}{M}N^2\sigma_\varepsilon^2.$$

Similar analysis can be done for other than the partial DFT measurement matrices.

3.9 Noise Folding

The additive noise is assumed in the measurements $\mathbf{y}$ only, in Section 3.7, as

$$\mathbf{y} + \boldsymbol{\varepsilon} = \mathbf{AX}. \tag{38}$$

The reconstruction performance of the reconstruction algorithms in the case when the pre-measurement additive noise $\mathbf{z}$ exists in the sparse signal coefficients $\mathbf{X}$ will be considered next [5]. The measurements model is of the form

$$\mathbf{y}+\boldsymbol{\varepsilon} = \mathbf{A}(\mathbf{X} + \mathbf{z}). \tag{39}$$

The noise $\mathbf{z}$ exists in the signal coefficients $\mathbf{X}$ prior to taking the measurements. Here, the additive noise affecting the signal measurements is denoted by $\boldsymbol{\varepsilon}$ with covariance $\sigma_\epsilon^2\mathbf{I}$. The noise vector $\mathbf{z}$ is random with covariance $\sigma_z^2\mathbf{I}$. It is independent of $\boldsymbol{\varepsilon}$. Relation (39) can be written as

$$\mathbf{y} + \mathbf{v} = \mathbf{AX} \tag{40}$$

where

$$\mathbf{v} = \boldsymbol{\varepsilon} - \mathbf{Az}. \tag{41}$$

Therefore, the resulting noise $\mathbf{v}$ covariance matrix is

$$\mathbf{C} = \sigma_\epsilon^2\mathbf{I} + \sigma_z^2\mathbf{AA}^H. \tag{42}$$

When $\mathbf{A}$ is formed as a partial matrix from an orthonormal inverse transformation matrix, with normalized energy of columns the relation $\mathbf{AA}^T = \frac{N}{M}\mathbf{I}$ holds. The variance of $\mathbf{v}$ is then

$$\sigma_v^2 = \sigma_\epsilon^2 + \frac{N}{M}\sigma_z^2,$$

with the covariance matrix $\mathbf{C} = \sigma_v^2\mathbf{I}$. In this special case, models (38), (39) and (40) are equivalent with the difference that the noise variance of $\mathbf{v}$ is increased for $\frac{N}{M}\sigma_z^2$.

In a more general case, with $\mathbf{A}$ being an arbitrary random matrix with low coherence or low RIP, the noise $\mathbf{v}$ in (40) can be whitened multiplying the system with $\mathbf{C}^{-1/2}/\sigma_v$. In that case, the equivalent system $\tilde{\mathbf{y}} + \mathbf{u} = \mathbf{BX}$ is obtained, where $\tilde{\mathbf{y}} = \mathbf{C}^{-1/2}\mathbf{y}/\sigma_v$, $\mathbf{B} = \mathbf{C}^{-1/2}\mathbf{A}/\sigma_v$, and $\mathbf{u} = \mathbf{C}^{-1/2}\mathbf{v}/\sigma_v$. Under the reasonable assumption that $\|\mathbf{I} - \mathbf{AA}^T\|_2$ is small, the coherence and the RIP constants of $\mathbf{B}$ and $\mathbf{A}$ are similar [5].

We may conclude that the pre-measurement noise variance is increased by the factor of $\frac{N}{M}$. In the case when $M \ll N$ this leads to a large noise increase, or noise folding. For most common measuring schemes used in CS, the model with the pre-measurement noise $\tilde{\mathbf{y}} + \mathbf{u} = \mathbf{BX}$ is equivalent to the standard model (38) assuming only measurements noise. The measurement matrix is changed and the noise variance is increased by the factor $\frac{N}{M}$. It means that all previous relations and conclusions can be applied to this case as well.

4 Norm-One Based Reconstruction Algorithms

As mentioned above, minimizing the number of nonzero coefficients using the ℓ_0-norm is a nonconvex optmization problem that cannot be solved using well developed iterative algorithms and linear programming methods. Therefore, to avoid dealing with NP-hard problems, significant efforts have been undertaken to replace the nonconvex and discontinuous ℓ_0-norm with a convex and continuous norm that would be more appropriate for optimization. As a result, the ℓ_1-norm (norm-one) has been commonly employed in many signal reconstruction approaches. It has been shown that, under certain conditions, minimization of the ℓ_1-norm produces the same solution as the minimization of the ℓ_0-norm [26,27,34,40].

In the ℓ_1-norm based reconstructions the problem is formulated as

$$\min \|\mathbf{X}\|_1 \quad \text{subject to} \quad \mathbf{y} = \mathbf{AX}$$

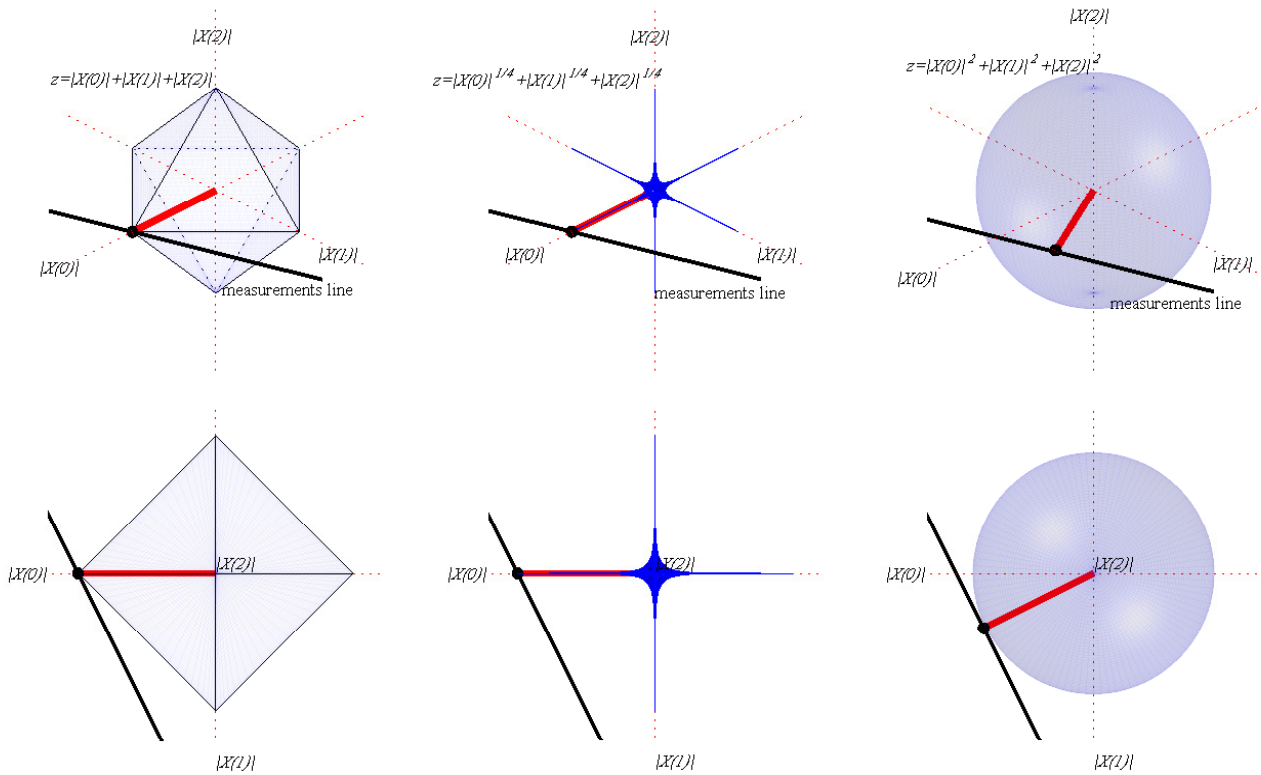

Figure 3: Illustration of the solution with the ℓ_1-norm, the $\ell_{1/4}$-norm (close to the ℓ_0-norm), the ℓ_2-norm for a three-dimensional case. In lower graphics, a view from the direction where the observations/measurement lines and the considered norm-balls touches, is presented.

where

$$\|\mathbf{X}\|_1 = \sum_{k=0}^{N-1} |X(k)|\,.$$

This is the so-called *basis pursuit* (BP) approach to sparse signal reconstruction.

As an illustration, consider the problem with sparsity $K = 1$, dimension $N = 3$ and two measurements $M = 2$ described by

$$\begin{aligned} y(0) &= X(0)\alpha_0(0) + X(1)\alpha_1(0) + X(2)\alpha_2(0) \\ y(1) &= X(0)\alpha_0(1) + X(1)\alpha_1(1) + X(2)\alpha_2(1). \end{aligned} \tag{43}$$

Each of the equations is a plane in the space of variables $X(0)$, $X(1)$ and $X(2)$, while both equations together represent a line in this space. This is illustrated in Fig. 3 where a "ball" with the constant cost function $z = |X(0)|^p + |X(1)|^p + |X(2)|^p$ is presented for the ℓ_1-norm and the $\ell_{1/4}$-norm that is close to the ℓ_0-norm ($p = 1$ and $p = 1/4$). The solution of the minimization problem is the intersection of the measurement line with the minimal possible "ball". It is obvious that in both cases $X(0)$ is indicated as a nonzero coefficient. Note also that the ℓ_2-norm ($p = 2$) will not produce a sparse result.

The solution is the same as long as the measurement line does not penetrate through the minimal "ball". For the ℓ_0-norm case the measurement line should not have zeros in the direction vector. For this illustrative example this condition is equivalent to the condition that all submatrices of the measurement matrix of order 2×2 have nonzero determinant (since these determinants are equal to the direction coefficients of the "measurement line"). For the ℓ_1-norm case the condition for measurement line direction is obviously more restrictive since it must not penetrate through the ℓ_1-"ball".

In general, the equivalence of the ℓ_0-norm and the ℓ_1-norm solutions is defined by the restricted isometry property.

For a K-sparse vector and a measurement matrix $\mathbf{A}$ the solution of the ℓ_0-norm minimization problem is the same as the solution of the corresponding ℓ_1-norm based minimization process if the measurement matrix satisfies

$$1-\delta_{2K} \leq \frac{\|\mathbf{A}_{2K}\mathbf{X}_{2K}\|_2^2}{\|\mathbf{X}_{2K}\|_2^2} \leq 1+\delta_{2K} \tag{44}$$

where $0 \leq \delta_{2K} < \sqrt{2}-1$ for all submatrices of order $2K$ of the measurement matrix $\mathbf{A}$ [24]. Note that for the ℓ_0-norm the restricted isometry constant range was $0 \leq \delta_{2K} < 1$.

The problem of the ℓ_1-norm based minimization can be reformulated in various ways. For example, its Lagrangian formulation is

$$F(\mathbf{X}) = \|\mathbf{y}-\mathbf{A}\mathbf{X}\|_2^2 + \lambda \|\mathbf{X}\|_1$$

where $F(\mathbf{X})$ is the function to be minimized.

Reformulation of the problem in a constrained form reads

$$\min \|\mathbf{X}\|_1 \quad \text{subject to} \quad \|\mathbf{y}-\mathbf{A}\mathbf{X}\|_2^2 < \varepsilon,$$

where ε is sufficiently small parameter. This is the ℓ_1-norm based minimization with quadratic constraint.

There are many ways to solve the stated problem, based on the constrained or Lagrangian form [40]. Many of them are developed within the regression theory. Here we will present in detail just one of them, based on the least absolute selection and shrinkage operator (LASSO) formulation and Lagrangian minimization form [108].

4.1 LASSO Minimization

The ℓ_1-norm based minimization can be formulated as the minimization of the error $\mathbf{y}-\mathbf{A}\mathbf{X}$ with a condition imposed on $\mathbf{X}$. The cost function

$$\begin{aligned} F(\mathbf{X}) &= \|\mathbf{y}-\mathbf{A}\mathbf{X}\|_2^2 + \lambda \|\mathbf{X}\|_1 \\ &= \|\mathbf{y}\|_2^2 - \mathbf{X}^T\mathbf{A}^T\mathbf{y} - \mathbf{y}^T\mathbf{A}\mathbf{X} + \mathbf{X}^T\mathbf{A}^T\mathbf{A}\mathbf{X} + \lambda\mathbf{X}^T\text{sign}\{\mathbf{X}\} \end{aligned}$$

is used here. The LASSO minimization problem formulation is

$$\mathbf{X} = \arg\min_{\mathbf{X}} \left\{ \|\mathbf{y}-\mathbf{A}\mathbf{X}\|_2^2 + \lambda \|\mathbf{X}\|_1 \right\}.$$

The function $\|\mathbf{X}\|_1$ promotes sparsity. It produces the same results (under the described conditions) as when $\|\mathbf{X}\|_p$, with p close to 0, is used, Fig. 3.

The minimization problem with the ℓ_1-norm constraint does not have a closed form solution. It is solved in iterative ways. In order to define an iterative procedure a nonnegative term

$$G(\mathbf{X}) = (\mathbf{X}-\mathbf{X}_s)^T(\alpha\mathbf{I}-\mathbf{A}^T\mathbf{A})(\mathbf{X}-\mathbf{X}_s)$$

having zero value at the solution $\mathbf{X}_s$ of the problem is added to the function $F(\mathbf{X})$. This term will not change the minimization solution. New cost function is

$$H(\mathbf{X}) = F(\mathbf{X}) + (\mathbf{X}-\mathbf{X}_s)^T(\alpha\mathbf{I}-\mathbf{A}^T\mathbf{A})(\mathbf{X}-\mathbf{X}_s),$$

where α is such that the added term is always nonnegative. It means $\alpha > \lambda_{\max}$, where $\lambda_{\max}$ is the largest eigenvalue of $\mathbf{A}^T\mathbf{A}$. The gradient of $H(\mathbf{X})$ is

$$\nabla H(\mathbf{X}) = \frac{\partial H(\mathbf{X})}{\partial \mathbf{X}^T} = -2\mathbf{A}^T\mathbf{y} + 2\mathbf{A}^T\mathbf{A}\mathbf{X} + \lambda\text{sign}\{\mathbf{X}\} + 2(\alpha\mathbf{I}-\mathbf{A}^T\mathbf{A})(\mathbf{X}-\mathbf{X}_s).$$

the solution of $\nabla H(\mathbf{X}) = \mathbf{0}$ is

$$\begin{aligned} -\mathbf{A}^T\mathbf{y} + \frac{\lambda}{2}\text{sign}\{\mathbf{X}\} - (\alpha\mathbf{I}-\mathbf{A}^T\mathbf{A})\mathbf{X}_s + \alpha\mathbf{X} &= \mathbf{0} \\ \mathbf{X} + \frac{\lambda}{2\alpha}\text{sign}\{\mathbf{X}\} &= \frac{1}{\alpha}\mathbf{A}^T(\mathbf{y}-\mathbf{A}\mathbf{X}_s) + \mathbf{X}_s. \end{aligned}$$

The corresponding iterative relation is of the form

$$\mathbf{X}_{s+1}+\frac{\lambda}{2\alpha}\text{sign}\{\mathbf{X}_{s+1}\}=\frac{1}{\alpha}\mathbf{A}^T(\mathbf{y}-\mathbf{A}\mathbf{X}_s)+\mathbf{X}_s.$$

The soft-thresholding rule is used as a solution for the scalar equation

$$x+\lambda\text{sign}(x)=y.$$

It is defined by the function $\text{soft}(y,\lambda)$ as

$$x=\text{soft}(y,\lambda)=\begin{cases} y+\lambda & \text{for} \quad y<-\lambda \\ 0 & \text{for} \quad |y|\leq\lambda \\ y-\lambda & \text{for} \quad y>\lambda \end{cases},$$

or

$$\text{soft}(y,\lambda)=\text{sign}(y)\max\{0,|y|-\lambda\}.$$

The same rule can be applied to each coordinate of vector $\mathbf{X}_{s+1}$,

$$\mathbf{X}_{s+1}=\text{soft}(\frac{1}{\alpha}\mathbf{A}^T(\mathbf{y}-\mathbf{A}\mathbf{X}_s)+\mathbf{X}_s,\frac{\lambda}{2\alpha}) \tag{45}$$

or

$$X_{s+1}(k)=\text{soft}(\frac{1}{\alpha}(a(k)-b(k))+X_s(k),\frac{\lambda}{2\alpha})$$

where $a(k)$ and $b(k)$ are coordinates of vectors $\mathbf{a}$ and $\mathbf{b}$ defined by $\mathbf{a}=\mathbf{A}^T\mathbf{y}$ and $\mathbf{b}=\mathbf{A}^T\mathbf{A}\mathbf{X}_s$. The Lagrangian constant λ is a balance between the error and the ℓ_1-norm value, while $\alpha=2\max\left\{\text{eig}\{\mathbf{A}^T\mathbf{A}\}\right\}$ is commonly used.

This is the Iterative Soft-Thresholding Algorithm (ISTA) for LASSO minimization. It can be easily modified to improve convergence in the Fast ISTA (FISTA), [12,32]. Note that this is just one of possible solutions of the minimization problem with the ℓ_1-norm.

One of the most popular software tools for the ℓ_1-norm based signal reconstruction is the ℓ_1-magic [25]. It is based on the primal-dual algorithm for linear programming [19].

Algorithm 4 LASSO – ISTA reconstruction

Input:

- Measurement vector $\mathbf{y}$
- Measurement matrix $\mathbf{A}$
- Regularization parameter α
- Sparsity promotion parameter λ

1: $\mathbf{X}\leftarrow\mathbf{0}_{N\times 1}$
2: **repeat**
3: $\quad\mathbf{s}\leftarrow\frac{1}{\alpha}\mathbf{A}^T(\mathbf{y}-\mathbf{A}\mathbf{X})+\mathbf{X}$
4: $\quad$**for** $k\leftarrow 1$ to N **do**
5: $\quad\quad X(k)\leftarrow\begin{cases} s(k)+\lambda & \text{for } s(k)<-\lambda \\ 0 & \text{for } |s(k)|\leq\lambda \\ s(k)-\lambda & \text{for } s(k)>\lambda \end{cases}$
6: $\quad$**end for**
7: **until** stopping criterion is satisfied

Output:

- Reconstructed signal coefficients $\mathbf{X}$

This is a pre-print of an article published in Circuits Systems and Signal Processing.
The final authenticated version is available online at: https://doi.org/10.1007/s00034-018-0909-2

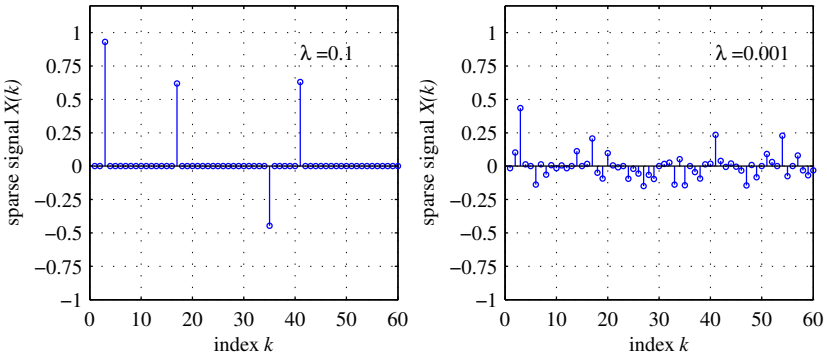

Figure 4: A sparse signal with $N = 100$ and $K = 4$ reconstructed using a reduced set of $M = 40$ measurements and LASSO iterative algorithm. The results for $\lambda = 0.1$ and $\lambda = 0.001$ are presented.

Example 3. Measurement matrix $\mathbf{A}$ is formed as a Gaussian random matrix of the size 40×100. Since there are 40 measurements the random variable $\mathcal{N}(0, \sigma^2)$ with $\sigma^2 = 1/40$ is used. The original sparse signal of total length $N = 100$ is $X(k) = \delta(k-3) + 0.7\delta(k-17) - 0.5\delta(k-35) + 0.7\delta(k-41)$ in the transformation domain. It is measured with a matrix $\mathbf{A}$ with 40 measurements stored in vector $\mathbf{y}$. All 60 signal values are reconstructed using these 40 measurements $\mathbf{y}$ and the matrix $\mathbf{A}$, in 1000 iterations. In the initial iteration $\mathbf{X}_0 = \mathbf{0}$ is used. Then, for each subsequent s the new values of $\mathbf{X}$ are calculated using (45), given data $\mathbf{y}$ and matrix $\mathbf{A}$. The value of $\alpha = 2\max\{\text{eig}\{\mathbf{A}^T\mathbf{A}\}\}$ is used. The results for $\lambda = 0.1$ and $\lambda = 0.001$ are presented in Fig. 4. For a very small $\lambda = 0.001$ the result is not sparse, since the constraint is too weak.

□

4.1.1 Norm-Two Based Minimization Solution

Here we will also explain why commonly used the ℓ_2-norm based minimization is not suitable for this application. If the cost function with the ℓ_2-norm is used

$$F(\mathbf{X}) = \|\mathbf{y} - \mathbf{A}\mathbf{X}\|_2^2 + \lambda \|\mathbf{X}\|_2^2$$

then the solution can be obtained in an analytic way. Minimization of the ℓ_2-norm constrained problem can be reformulated in Lagrangian form using a parameter λ as

$$\mathbf{X} = \arg\min_{\mathbf{X}} \left\{ \|\mathbf{y} - \mathbf{A}\mathbf{X}\|_2^2 + \lambda \|\mathbf{X}\|_2^2 \right\}.$$

The solution of

$$\begin{aligned} F(\mathbf{X}) &= \|\mathbf{y} - \mathbf{A}\mathbf{X}\|_2^2 + \lambda \|\mathbf{X}\|_2^2 \\ &= \|\mathbf{y}\|_2^2 - \mathbf{X}^H\mathbf{A}^H\mathbf{y} - \mathbf{y}^H\mathbf{A}\mathbf{X} + \mathbf{X}^H\mathbf{A}^H\mathbf{A}\mathbf{X} + \lambda\mathbf{X}^H\mathbf{X} \end{aligned}$$

minimization can be obtained in a closed form using the symbolic derivative operator $\partial F(\mathbf{X})/\partial \mathbf{X}^H = 0$. It is

$$\mathbf{X}_{ridge} = \left(\mathbf{A}^H\mathbf{A} + \mathbf{I}\lambda\right)^{-1}\mathbf{A}^H\mathbf{y} = \left(\mathbf{A}^H\mathbf{A} + \mathbf{I}\lambda\right)^{-1}\mathbf{X}_0, \tag{46}$$

where $\mathbf{X}_0 = \mathbf{A}^H\mathbf{y}$ is the initial transform defined by (26), for normalized measurement matrices. Parameter λ represent a balance between the error and the constraint.

The ℓ_2-norm minimization problem can also be formulated with the Lagrange multipliers as,

$$F(\mathbf{X}) = \|\mathbf{X}\|_2^2 + \mathbf{\Lambda}^H(\mathbf{y} - \mathbf{A}\mathbf{X}),$$

where $\mathbf{\Lambda} = [\lambda_1, \lambda_2, \ldots, \lambda_M]^T$ are the Lagrange multipliers. Its minimization, using the symbolic derivative operator, $\partial F(\mathbf{X})/\partial \mathbf{X}^H = 0$ produces

$$2\mathbf{X} - \mathbf{\Lambda}\mathbf{A}^H = 0 \text{ or } \mathbf{X} = \mathbf{A}^H\mathbf{\Lambda}/2.$$

If this value of $\mathbf{X}$ is replaced in the constraint relation $\mathbf{y} = \mathbf{A}\mathbf{X}$ we get $\mathbf{y} = \mathbf{A}\mathbf{A}^H\mathbf{\Lambda}/2$ or $\mathbf{\Lambda} = 2(\mathbf{A}\mathbf{A}^H)^{-1}\mathbf{y}$. Finally

$$\mathbf{X} = \mathbf{A}^H\mathbf{\Lambda}/2 = \mathbf{A}^H(\mathbf{A}\mathbf{A}^H)^{-1}\mathbf{y} = \mathbf{X}_0.$$

In most of the measurement matrices, like the partial DFT and the partial DCT matrix, $\mathbf{A}\mathbf{A}^H = \mathbf{I}$ holds, up to a constant equal to the energy of the measurement matrix rows. In the Gaussian measurement matrix this relation holds approximately for large N.

The standard ridge regression, based on the ℓ_2-norm, minimizes the energy of solution $X(k)$ and not its sparsity.

4.2 Signal Reconstruction with a Gradient Algorithm

Let us consider a sparse vector $\mathbf{X}$ of the length N. Assume that $\mathbf{X}$ is obtained as a linear and invertible transformation of a signal $\mathbf{x}$. In this case the vector $\mathbf{X}$ and corresponding signal $\mathbf{x}$ are related via direct and inverse transformation matrices

$$\mathbf{x} = \mathbf{\Psi}\mathbf{X}, \tag{47}$$
$$\mathbf{X} = \mathbf{\Phi}\mathbf{x}. \tag{48}$$

The transformation matrices $\mathbf{\Psi}$ and $\mathbf{\Phi}$ are of size $(N \times N)$. Vectors $\mathbf{X}$ and $\mathbf{x}$ are of the same length (N coefficients/samples). Signal samples $x(n)$, $n = 0,1,\ldots,N-1$ can be considered as a full (complete) set of measurements of $\mathbf{X}$. The arbitrary signal $\mathbf{X}$ can be uniquely reconstructed from the full set of measurements $\mathbf{x}$ using Eq. (48).

Sparse $\mathbf{X}$ can be reconstructed from a reduced set of $M < N$ measurements. The measurements vector, denoted by $\mathbf{y}$, of length M is obtained by selecting M measurements from the full set of signal $\mathbf{x}$ samples (considered and denoted as measurements)

$$\mathbf{y} = [x(n_1), x(n_2), \ldots, x(n_M)]^T.$$

The set of selected indices $\{n_1, n_2, \ldots, n_M\}$ will be denoted as $\mathbb{M}$. The measurements with indices $n \notin \mathbb{M}$ can be considered as unavailable/missing measurements. The set of the indices of remaining measurements will be denoted by $\mathbb{Q}$

$$\mathbb{M} = \{n_1, n_2, \ldots, n_M\} \subset \{0,1,\ldots,N-1\}$$
$$\mathbb{Q} = \{n_{M+1}, n_{M+2}, \ldots, n_N\} = \{0,1,\ldots,N-1\} \setminus \mathbb{M}.$$

Note that $\mathbb{M} \cap \mathbb{Q} = \emptyset$ and $\mathbb{M} \cup \mathbb{Q} = \{0,1,\ldots,N-1\}$. The vector of missing measurements will be denoted by $\mathbf{y}_c$, where

$$\mathbf{y}_c = [y_c(0), y_c(1), \ldots, y_c(N-M-1)]^T = [x(n_{M+1}), x(n_{M+2}), \ldots, x(n_{N-1})]^T.$$

The measurement matrix $\mathbf{A}$ is obtained by selecting M rows from the inverse transformation matrix $\mathbf{\Psi}$ with indices that belongs to the set $\mathbb{M}$.

The ℓ_1-norm based reconstruction of the sparse signal

$$\min \|\mathbf{X}\|_1 \quad \text{subject to } \mathbf{y} = \mathbf{A}\mathbf{X}$$

can be reformulated in the following way. If we can reconstruct the missing measurements $\mathbf{y}_c$, then the full set of measurements is obtained. The direct transformation matrix $\mathbf{\Phi}$ is used to calculate reconstructed signal in its sparsity domain Eq. (47). The missing measurements $\mathbf{y}_c$ are obtained by minimizing $\|\mathbf{X}\|_1 = \|\mathbf{\Phi}\mathbf{x}\|_1$, where $\mathbf{x}$ is a complete vector of the measurements with elements $x(n_i) = y(i)$ for $n_i \in \mathbb{M}$ and $x(n_i) = y_c(i)$ for $n_i \in \mathbb{Q}$.

This minimization problem can be solved by varying all unavailable/missing samples $x(n_i)$, $n_i \in \mathbb{M}$ within the range of their possible values and selecting values which produced the minimal sparsity measure. The available samples are kept unchanged. This is a kind of direct search over the missing measurement/sample values. However, when the number of missing samples is large, then this direct search cannot be used due to its high calculation complexity.

This is a pre-print of an article published in Circuits Systems and Signal Processing.
The final authenticated version is available online at: https://doi.org/10.1007/s00034-018-0909-2

Another approach to minimization of the sparsity measure $\|\mathbf{X}\|_1 = \|\mathbf{\Phi x}\|_1$ is to use the gradient descent (or steepest descent) algorithm. The minimum of sparsity measure, with respect to the missing samples as minimization variables, is determined through an iterative procedure

$$\mathbf{y}_c^{(m+1)} = \mathbf{y}_c^{(m)} - \alpha \left. \frac{\partial \|\mathbf{\Phi x}\|_1}{\partial \mathbf{y}_c} \right|_{\mathbf{y}_c = \mathbf{y}_c^{(m)}},$$

where $\mathbf{y}_c^{(m)}$ is the vector of missing samples in the mth iteration and α is the iteration step. The ℓ_1-norm is used as the sparsity measure [85]. The gradient of sparsity measure, calculated at $\mathbf{y}_c = \mathbf{y}_c^{(m)}$, is denoted by

$$\mathbf{g}^{(m)} = \partial \|\mathbf{\Phi x}\|_1 / \partial \mathbf{y}_c|_{\mathbf{y}_c = \mathbf{y}_c^{(m)}}.$$

The iterative procedure convergence depends on the iteration step α.

The gradient of the sparsity measure in mth iteration can be estimated using finite difference method. For each missing sample n_i, $i = 1,2,\ldots,M$ we can form two full sets of measurements $z_1(n)$ and $z_2(n)$, where the value of the ith missing sample (located at position n_i in $z_1(n)$ and $z_2(n)$) is increased and decreased by some Δ

$$z_1(n) = \begin{cases} x^{(m)}(n) & \text{for } n \neq n_i \\ x^{(m)}(n) + \Delta & \text{for } n = n_i \in \mathbb{Q} \end{cases}$$

$$z_2(n) = \begin{cases} x^{(m)}(n) & \text{for } n \neq n_i \\ x^{(m)}(n) - \Delta & \text{for } n = n_i \in \mathbb{Q} \end{cases}$$

The n_ith coefficient of gradient vector $\mathbf{g}^{(m)}$ is then estimated as

$$g^{(m)}(n_i) = \frac{\|\mathbf{\Phi z}_1\|_1 - \|\mathbf{\Phi z}_2\|_1}{2\Delta}.$$

The iteration step α is commonly used in the form $\alpha = 2\Delta c$, where c is a constant equal to the energy of the inverse transform basis functions. Then, in the algorithm we can use $g^{(m)}(n_i) = ||\mathbf{\Phi z}_1||_1 - ||\mathbf{\Phi z}_2||_1$ and $\alpha = c$, since the variable iteration step is now included into the form of $g^{(m)}(n_i)$.

Coefficients of the gradient vector for available samples are zero valued, $g^{(m)}(n_i) = 0, n_i \in \mathbb{M}$.

In the initial iteration for missing measurements we can use zero values $\mathbf{y}_c^{(0)} = \mathbf{0}$, or the mean value of available samples/measurements. The reconstruction procedure is presented in Algorithm 5.

Due to specific form of the ℓ_1-norm in the very vicinity of its minimum (the solution) the algorithm will approach to the solution with a bias whose upper limit is proportional to the Δ. Its value can be estimated as $\Delta K/N$. This bias can be reduced to an acceptable precision by using very small Δ. However, calculation with a small Δ would be time consuming (requiring many iterations). Efficient implementation ([16,61,62,83,92]) can be achieved by using variable Δ. In the initial iterations Δ should be of the same order as the amplitude of the available samples $\Delta = \max\{|\mathbf{y}|\}$. When the algorithm reaches a stationary point, with a given Δ, the value of the mean squared error will assume an almost constant value, determined with the bias, and the values of missing samples will oscillate [92].

Example 4. Consider a signal $x(n) = 3\cos(2\pi\frac{n}{N} - \pi/5)$ with $N = 8$. The signal is sparse in the DFT domain. Assume that the missing measurements/samples are at positions $n \in \mathbb{Q} = \{1,6\}$. The signal is reconstructed using the gradient based algorithm in 60 iterations. The initial algorithm parameter $\Delta = 1$, step $\alpha = 1/N$ and the initial value of missing samples $x(1) = 0$ and $x(6) = 0$ are used. The values of missing samples in the first 20 iterations are represented by dots (connected by a line) in Fig. 6. After about 6 iterations the algorithm with $\Delta = 1$ does not significantly change the missing sample values (zoomed changes are shown in lower subplot within the figure). Close to the stationary point obtained for $\Delta = 1$ the gradient coordinates are almost zero-valued (with direction changes for almost π). After the step is reduced to $\Delta = 0.1$ in the 20th iteration, the algorithm resumes its fast approach toward the exact value, until reaching a new stationary state. With a new change of Δ to $\Delta = 0.01$ the approach is again continued. The stationary state bias for $\Delta = 1$ is lower than $\frac{K}{N}\Delta = 1/4$. For a signal of amplitude 3 it corresponds to the bias caused by an MSE lower than $-20\log(3/(1/4)) = -21.6$ [dB], Fig. 5. By each reduction of Δ to $\Delta/10$ the bias caused MSE will be lower for 20 [dB]. The reconstruction result and the MSE for the estimated missing values $x(1)$ and $x(6)$ are presented in Fig. 5.

This is a pre-print of an article published in Circuits Systems and Signal Processing.
The final authenticated version is available online at: https://doi.org/10.1007/s00034-018-0909-2

Algorithm 5 Gradient based reconstruction procedure

Input:

- Set of missing/omitted sample positions $\mathbb{Q}$
- Set of available sample positions $\mathbb{M}$
- Available samples (measurements) $\mathbf{y}$
- Transformation matrix $\mathbf{\Phi}$
- Step α

$m \leftarrow 0$
Set initial estimate signal vector $\mathbf{x}^{(0)}$ as $x^{(0)}(n_i) = y(i)$ for $n_i \in \mathbb{M}$ and $x^{(0)}(n_i) = 0$ for $n_i \in \mathbb{Q}$
$\Delta \leftarrow \max_n |x^{(0)}(n)|$
repeat
  repeat
    $\mathbf{x}^{(m+1)} \leftarrow \mathbf{x}^{(m)}$
    for $n_i \in \mathbb{Q}$ **do**
      $\mathbf{z}_1 \leftarrow \mathbf{x}^{(m)}$
      $z_1(n_i) \leftarrow z_1(n_i) + \Delta$
      $\mathbf{z}_2 \leftarrow \mathbf{x}^{(m)}$
      $z_2(n_i) \leftarrow z_2(n_i) - \Delta$
      $g(n_i) \leftarrow \|\mathbf{\Phi z_1}\|_1 - \|\mathbf{\Phi z_2}\|_1$
      $x^{(m+1)}(n_i) \leftarrow x^{(m)}(n_i) - \alpha\, g(n_i)$
    end for
    $m \leftarrow m + 1$
  until stopping criterion is satisfied
  $\Delta \leftarrow \Delta/3$
until required precision is achieved
$\mathbf{x} \leftarrow \mathbf{x}^{(m)}$
$\mathbf{X} \leftarrow \mathbf{\Phi x}$

Output:

- Reconstructed signal vector $\mathbf{X}$
- Full set of measurements $\mathbf{x}$

In order to approach the true signal values with a given precision Δ should be reduced. The fact that the reconstructed values oscillate when the stationary point is reached may be used as an indicator to reduce the step Δ. The oscillations of the solution are detected by measuring the angle between two successive gradient vectors

$$\beta_m = \arccos \frac{\sum_{i=0}^{N-1} g^{(m-1)}(i) g^{(m)}(i)}{\sqrt{\sum_{i=0}^{N-1} (g^{(m-1)}(i))^2} \sqrt{\sum_{i=0}^{N-1} (g^{(m)}(i))^2}}.$$

If the angle β_m is lower than, for example 170°, the calculation is continued with the same Δ. Otherwise the Δ is reduced. This is a possible stopping criterion in Algorithm 5.

A possible divergence in a gradient-based algorithm is related to the algorithm behavior for large steps Δ. Small step Δ significantly increase the number of iterations in the algorithm. It can be shown that the value of the estimated gradient $\mathbf{g}^{(m)}$ that is used to correct the missing signal samples does not depend on the value of step Δ if Δ is large. The missing signal values will be adapted for a value independent of Δ in that case. The missing samples will oscillate within the range of the available signal samples, until Δ is reduced. Then the missing samples will start approaching to the minimum of the sparsity measure.

The iterative algorithm should be stopped when Δ is small enough, or when the change of reconstructed missing samples between two consecutive Δ reductions is bellow the desired precision.

This is a pre-print of an article published in Circuits Systems and Signal Processing.
The final authenticated version is available online at: https://doi.org/10.1007/s00034-018-0909-2

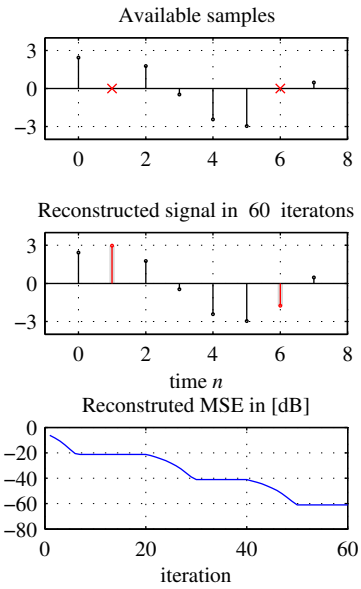

Figure 5: Gradient-based reconstruction of a sparse signal.

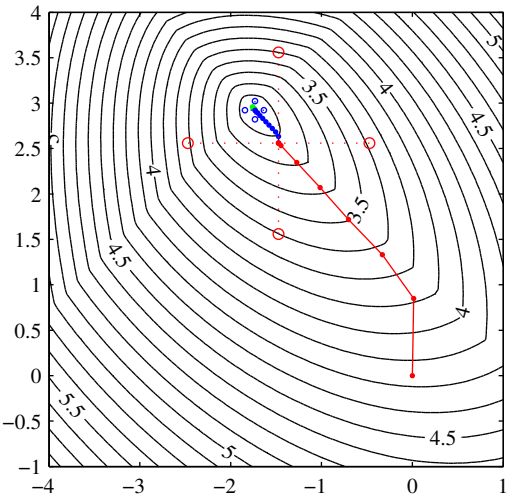

Figure 6: Illustration of a signal reconstruction using the adaptive gradient algorithm.

4.3 Total variations

The presented ℓ_1-norm based algorithms may be used to solve other convex problems. Here we will also mention the possibility that the coefficients in vector $\mathbf{X}$ are not sparse, but instead, for example, their first-order finite difference

$$Z(k) = X(k) - X(k-1)$$
$$\mathbf{Z} = \text{diff}\{\mathbf{X}\}$$

is sparse. A simple example is $X(k) = 1$ for $0 \leq k < N/2$ and $X(k) = -1$ for $N/2 \leq k < N$. In this case the minimization problem is formulated as

$$\min \|\mathbf{Z}\|_1 \qquad \text{subject to} \qquad \mathbf{y} = \mathbf{A}\mathbf{X},$$

since the first difference vector $\mathbf{Z}$ is sparse. This is the minimization of total variations in the space of $\mathbf{X}$. The result is the value of $\mathbf{Z}$ with the smallest possible number of nonzero elements. The value of $\mathbf{X}$ is then with a minimal variations (maximally flat). Its values follow from $X(0) = Z(0)$, $X(1) = Z(1) + X(0), \ldots$, $X(N-1) = Z(N-1) + X(N-2)$.

Total variations (TV) can be extended to two-dimensional signals in order to get maximally flat images. The TV of a 2D signal $X(k_1,k_2)$ can be defined as the sum of the magnitudes of the discrete gradient at each point:

$$\text{TV}(\mathbf{X}) = \sum_{k_1,k_2} \left\| D_{k_1,k_2} \right\|_2,$$

where the gradient approximation for the position (k_1,k_2) is denoted as D_{k_1,k_2}. It is defined as:

$$D_{k_1,k_2} = \begin{bmatrix} X(k_1+1,k_2) - X(k_1,k_2) \\ X(k_1,k_2+1) - X(k_1,k_2) \end{bmatrix}.$$

The discrete form of the TV is

$$\text{TV}(\mathbf{X}) = \sum_{k_1,k_2} \sqrt{(X(k_1+1,k_2) - X(k_1,k_2))^2 + (X(k_1,k_2+1) - X(k_1+1,k_2))^2}.$$

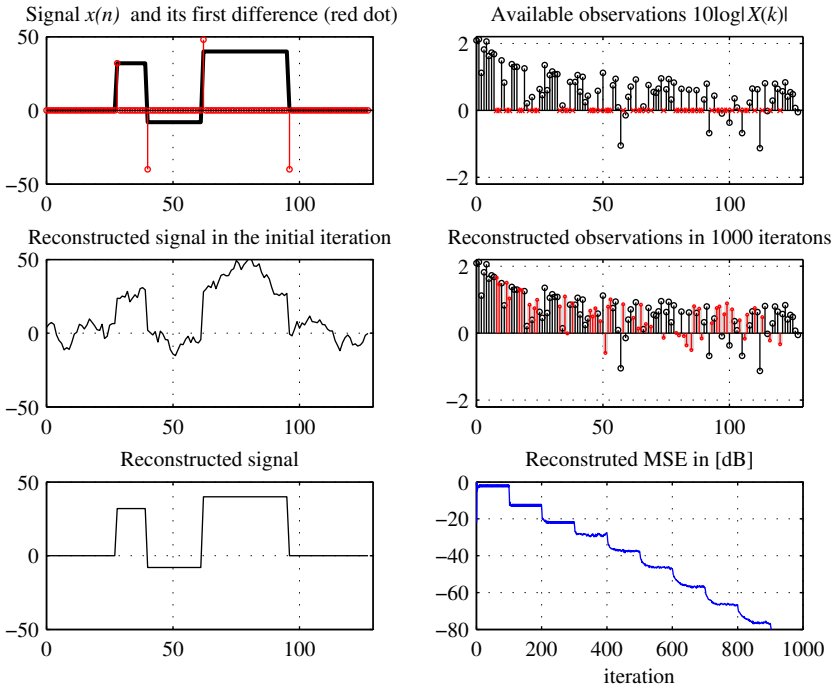

Figure 7: Illustration of a signal reconstruction whose first difference is sparse (with $K = 4$) using the adaptive gradient algorithm and a reduced set of $M = 80$ out of $N = 128$ measurements in the DCT domain, $X(k) = \text{DCT}\{x(n)\}$.

The TV minimization problem formulation reads

$$\min \text{TV}(\mathbf{X}) \text{ subject to } \mathbf{y} = \mathbf{AX}.$$

The signal and its transform can also switch their common roles. If a signal $\mathbf{x}$ is sparse then the transformation domain coefficients $\mathbf{X}$, where the signal is dense, can be considered as its measurement space. The minimization problem is then $\min \|\text{diff}\{\mathbf{x}\}\|_1$ subject to $\mathbf{Y} = \mathbf{Ax}$, with an appropriate measurement matrix $\mathbf{A}$ corresponding now to a direct signal transform.

Example 5. Consider a signal $x(n) = 32u(n-28) - 40u(n-40) + 48u(n-62) - 40u(n-96)$ within $0 \leq n \leq N-1$ and $N = 128$. Unit step function is denoted by $u(n)$. This signal is not sparse, but its first difference is sparse with sparsity $K = 4$, Fig.7 (top-left). The measurements are the DCT domain coefficients $X(k) = \text{DCT}\{x(n)\}$. Note that here the signal (its first difference) is sparse in the time domain where the measurement domain is the transform domain. A reduced set of randomly positioned $M = 80$ measurements $\mathbf{Y}$ is available, Fig.7 (top-right). The red cross indicates missing measurements. The signal reconstructed using available measurements and setting the missing ones to zero values is presented in Fig.7 (middle-left). The minimization problem $\min \|\text{diff}\{\mathbf{x}\}\|_1$ subject to $\mathbf{Y} = \mathbf{Ax}$ is solved using the gradient algorithm in 1000 iterations, with the algorithm step Δ being reduced to $\Delta/\sqrt{10}$ after each 100 iterations, respectively. The missing measurements are reconstructed and presented in Fig.7 (middle-right). The reconstructed signal is shown in Fig.7 (bottom-left), along with the reconstruction error through iterations Fig.7 (bottom-right). This kind of reconstruction promotes signal flatness (the first difference sparsity).

5 Bayesian-Based Reconstruction

A set of the output values and a set of the input values are given in the supervised learning. The main goal is to learn the model that describes the relation between the input and the output data. This model is then used for the next input values prediction. The measurement relation $\mathbf{y} = \mathbf{AX}$ can be understood, in this sense, as a linear combination of functions in $\mathbf{A}$ with parameters $\mathbf{X}$ that can adjust as inputs to

this model. Elements of the output vector $\mathbf{y}$ are

$$y(m) = \sum_{k=0}^{N-1} X(k)\alpha_k(m). \tag{49}$$

The problem can now be defined as a determination of Bayesian model for the estimation of parameters $\mathbf{X}$. The a priori knowledge about these parameters is that they are sparse [6, 10, 45, 109, 114].

Consider a noisy measurement

$$\mathbf{y} = \mathbf{A}\mathbf{X} + \varepsilon.$$

A Gaussian noise of zero-mean and variance σ^2 is denoted by ε. The probability density function of the measurement error $\varepsilon = \mathbf{y} - \mathbf{A}\mathbf{X}$ is Gaussian. Then the Gaussian likelihood model is defined as

$$p(\mathbf{y}|\mathbf{X},\sigma) = \frac{1}{\left(\sigma\sqrt{2\pi}\right)^M} e^{\frac{1}{2\sigma^2}\|\mathbf{y}-\mathbf{A}\mathbf{X}\|_2^2}. \tag{50}$$

Next we have introduce the sparsity promotion prior condition on $\mathbf{X}$. The Laplacian density function is a common sparsity promotion on $\mathbf{X}$ in Bayesian formulation. It corresponds to the ℓ_1-norm,

$$p(\mathbf{X}|\lambda) = (\frac{\lambda}{2})^N e^{-\lambda\|\mathbf{X}\|_1}. \tag{51}$$

However, instead of imposing a Laplacian prior on $\mathbf{X}$, a hierarchical prior has been used in the relevance vector machine approach. Thus prior exhibits similar properties. The zero-mean Gaussian prior distribution for $\mathbf{X}$ is assumed with hyperparameters d_i

$$p(\mathbf{X}|\mathbf{D}) = \prod_{i=0}^{N-1} \frac{1}{\sqrt{2\pi d_i^{-1}}} e^{-X(i)d_i X^*(i)} = e^{-\mathbf{X}^T\mathbf{D}\mathbf{X}} \prod_{i=0}^{N-1} \frac{1}{\sqrt{2\pi d_i^{-1}}}. \tag{52}$$

A diagonal matrix of hyperparameters is defined as $\mathbf{D} = diag(d_0, d_1, \ldots, d_{N-1})$. The role of hyperparameters is to promote sparsity.

Posterior distribution of $\mathbf{X}$ will be denoted by $p(\mathbf{X},\mathbf{D},\sigma^2|\mathbf{y})$. It could be found from

$$p(\mathbf{X},\mathbf{D},\sigma^2|\mathbf{y})p(\mathbf{y}) = p(\mathbf{y}|\mathbf{X},\mathbf{D},\sigma^2)p(\mathbf{X},\mathbf{D},\sigma^2). \tag{53}$$

The calculation of corresponding probabilities in this equation is not possible in an analytic way. Therefore, approximations should be carried out. To this aim the posterior for $\mathbf{X}$ is decomposed

$$p(\mathbf{X},\mathbf{D},\sigma^2|\mathbf{y}) = p(\mathbf{X}|\mathbf{y},\mathbf{D},\sigma^2)p(\mathbf{D},\sigma^2|\mathbf{y}) \tag{54}$$

and the probabilities in this relation are calculated. The first term

$$p(\mathbf{X}|\mathbf{y},\mathbf{D},\sigma^2) = \frac{p(\mathbf{y}|\mathbf{X},\sigma^2)p(\mathbf{X}|\mathbf{D})}{p(\mathbf{y}|\mathbf{D},\sigma^2)} \tag{55}$$

is obtained from the probabilities given by Eqs.(50) and (52). Using

$$p(\mathbf{X}|\mathbf{y},\mathbf{D},\sigma^2)p(\mathbf{y}|\mathbf{D},\sigma^2) = p(\mathbf{y}|\mathbf{X},\sigma^2)p(\mathbf{X}|\mathbf{D}) \tag{56}$$

and grouping the terms in exponent on the right side of Eq.(56) with Eq.(50) and Eq.(52). After appropriate matrix transformations we will get that the probability $p(\mathbf{X}|\mathbf{y},\mathbf{D},\sigma^2)$ is Gaussian. Its covariance and mean-value matrices are defined by [45, 109]

$$\Sigma = (\mathbf{A}^T\mathbf{A}/\sigma^2 + \mathbf{D})^{-1}, \tag{57}$$

$$\mathbf{V} = \Sigma\mathbf{A}^T\mathbf{y}/\sigma^2. \tag{58}$$

Note that the solution for the mean-value, for small noise variance σ^2, has the form of Eq.(25) since

$$\mathbf{V} = \Sigma\mathbf{A}^T\mathbf{y}/\sigma^2 = (\mathbf{A}^T\mathbf{A} + \sigma^2\mathbf{D})^{-1}\mathbf{A}^T\mathbf{y} \approx (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}. \tag{59}$$

This form is similar to Eq.(46), as well. Different weights for each $X(k)$, defined by matrix $\mathbf{D}$, are used here. The next step is in promoting sparsity. It is done by detecting candidates for zero-valued coefficients. These coefficients are then omitted from $\mathbf{X}$ and $\mathbf{A}$, until the desired K-sparse vector $\mathbf{X}_K$ is achieved. The corresponding measurement matrix $\mathbf{A}_K$ is used, resulting in $(\mathbf{A}_K^T\mathbf{A}_K)^{-1}\mathbf{A}_K^T\mathbf{y}$ as in Eq.(25).

We can see that in MP algorithms sparsity is promoted starting from its lowest order, by taking the strongest component. The sparsity is then increased until the true sparsity is achieved. Here in the Bayesian approach the process runs in opposite direction. The sparsity is reduced from its maximal value N to the true value K by using hyperparameters.

As we stated, in order to calculate the function $p(\mathbf{D},\sigma^2|\mathbf{y})$ and promote sparsity, some approximations are needed. One possible approximation is in replacing this function with its most-probable value. With this approximation the search for the hyperparameters $\mathbf{D}$ reduces to the maximization of $p(\mathbf{D},\sigma^2|\mathbf{y}) = p(\mathbf{y}|\mathbf{D},\sigma^2)p(\mathbf{D})p(\sigma^2)$. If the hyperparameters are unform only $p(\mathbf{y}|\mathbf{D},\sigma^2)$ should be minimized. Its value can be obtained in analytic form, as the marginal value

$$p(\mathbf{y}|\mathbf{D},\sigma^2) = \int p(\mathbf{y}|\mathbf{X},\sigma^2)p(\mathbf{X}|\mathbf{D})d\mathbf{X} \tag{60}$$

$$= \frac{1}{(2\pi)^{M/2}} \frac{1}{\sqrt{|\sigma^2\mathbf{I} + \mathbf{A}\mathbf{D}^{-1}\mathbf{A}^T|}} \exp(-\frac{1}{2}\mathbf{y}^H(\sigma^2\mathbf{I} + \mathbf{A}\mathbf{D}^{-1}\mathbf{A}^T)^{-1}\mathbf{y}). \tag{61}$$

This is the marginal likelihood. Its maximization is of type-II maximum likelihood method. Note that $\mathbf{D}$ and σ^2 that maximize this probability cannot be obtained in a closed form. A possible iterative procedure for their calculation is defined by [45,109]

$$d_i^{new} = \frac{\gamma_i}{V_i^2}$$

where V_i are the elements of the posterior mean vector $\mathbf{V}$ defined by Eq.(58) and coefficients γ_i are related to the diagonal elements Σ_{ii} of the covariance matrix $\mathbf{\Sigma}$, defined by Eq.(57), and the sparsity hyperparameters d_i as

$$\gamma_i = 1 - d_i\Sigma_{ii}.$$

After the coefficient γ_i is calculated the noise variance is estimated as

$$(\sigma^2)^{new} = \frac{\|\mathbf{y} - \mathbf{A}\mathbf{V}\|^2}{M - \sum_i \gamma_i}.$$

The final result are the coefficients, representing the resulting mean values of the coefficients. In addition, we get the variances of the estimated values. They follow from the covariance matrix. The iterative procedure is repeated until a required precision is obtained.

Note that in iterations, some of parameters d_i will tend to infinity. This indicates that the corresponding coefficients $X(i)$ tend to zero. These coefficients d_i and corresponding $X(i)$ are omitted from their matrices in next iterations. A small set of finite values d_i will remain in the final iteration. They indicate the positions of nonzero values in $X(i)$. Vector $\mathbf{X}_K$ is calculated with corresponding $\mathbf{A}_K$. The result of the iterative procedure are mean values of nonzero coefficients in vector $\mathbf{V}$,

$$\mathbf{X}_K = \mathbf{V} = \Sigma_K\mathbf{A}_K^T\mathbf{y}/\sigma^2$$

given by Eq.(58) and (59). The variances of coefficients $\mathbf{X}_K$ will be obtained as well. If a component position is wrongly detected and omitted it will influence the result accuracy as described in Section 3.8. This result for the variance in the reconstructed signal obtained here is the same as the one obtained by a quite simple derivation in Section 3.7. Note that after the positions of nonzero coefficients are found, using an iterative procedure in the Bayesian approach, coefficients with large hyperparameters are excluded along with the corresponding elements of matrix $\mathbf{D}$ and columns of $\mathbf{A}$. The hyperparameters for the nonzeroo coefficients are zero.

Example 6. A signal sparse in the discrete sine transform (DST) domain

$$x(n) = \sqrt{\frac{2}{N}} \sum_{i=1}^{K} X(k_i) \sin\Big(\frac{2\pi(2n+1)k_i}{4N}\Big) + \varepsilon(n), \quad k_i \neq 0$$

Algorithm 6 Bayesian based reconstruction

Input:

- Measurement vector $\mathbf{y}_{M\times 1}$
- Measurement matrix $\mathbf{A}_{M\times N}$

$d_i \leftarrow 1$ ▷ For $i = 1,2,\ldots,N$
$\sigma^2 \leftarrow 1$ ▷ Initial estimate
$T_h = 10^2$ ▷ Threshold
$\mathbf{p} = [1,2,\ldots,N]^T$
repeat
  $\mathbf{D} \leftarrow$ diagonal matrix with d_i values
  $\mathbf{\Sigma} \leftarrow (\mathbf{A}^T\mathbf{A}/\sigma^2 + \mathbf{D})^{-1}$
  $\mathbf{V} \leftarrow \mathbf{\Sigma}\mathbf{A}^T\mathbf{y}/\sigma^2$
  $\gamma_i \leftarrow 1 - d_i\Sigma_{ii}$ ▷ For each i
  $d_i \leftarrow \gamma_i / V_i$ ▷ For each i
  $\sigma^2 \leftarrow \dfrac{\|\mathbf{y} - \mathbf{AV}\|^2}{M - \sum_i \gamma_i}$
  $\mathbb{R} \leftarrow \{i : |d_i| > T_h\}$
  Remove columns from matrix $\mathbf{A}$ selected by $\mathbb{R}$
  Remove elements from array d_i selected by $\mathbb{R}$
  Remove elements from vector $\mathbf{p}$ selected by $\mathbb{R}$
until stopping criterion is satisfied
Reconstructed vector $\mathbf{X}$ nonzero coefficients are in vector $\mathbf{V}$ with corresponding positions in vector $\mathbf{p}$, $X_{p_i} = V_i$

Output:

- Reconstructed signal vector $\mathbf{X}$

with $N = 128$, $K = 10$, and noise variance $\sigma_\varepsilon^2 = 0.01$ is considered. The coefficients $X(k_i)$ values and positions are presented in Fig.8(a). They are considered as unknown and they should be found from the available measurements. Assume that only $M = N/2 = 64$ randomly positioned measurements $\mathbf{y} = [x(n_1), x(n_1), \ldots, x(n_M)]^T$ are available, Fig.8(b). The observation matrix $\mathbf{A}$ is obtained from the full DST matrix by keeping the rows corresponding to the measurement instants $n_1, n_2, \ldots, n_M$ only. In order to start the iterative algorithm the initial values $\mathbf{D} = \mathbf{I}$ and $\sigma = 0.1$ are assumed. The assumed threshold for considering hyperparameters extremely large is $T_h = 100$. Hyperparameters above this threshold are omitted from calculation (along with the corresponding values in $\mathbf{X}$, $\mathbf{A}$, $\mathbf{D}$ and $\mathbf{V}$). The results for the estimated mean value $\mathbf{V}$ in the first iteration are shown in Fig.8(c), along with the values of hyperparameters $\mathbf{V}$ in Fig.8(d). The hyperparameters whose value is above T_h are omitted (pruned) along with the corresponding values at the same positions in all other matrices. In the second iteration the values of remaining hyperparameters $\mathbf{V}$ are shown in Fig.8(e). After the elimination of hyperparameters above the threshold, the third iteration is calculated with the remaining positions of the hyperparameters. In this iteration all hyperparameters, except those whose values are close to one, are eliminated Fig.8(f). The remaining positions, after this iteration, correspond to the nonzero coefficients $X(k_i), i = 1,2,\ldots,K$ positions, with the corresponding pruned matrices $\Sigma_K, \mathbf{A}_K, \mathbf{D}_K$. The values of $X(k_i)$ are estimated using V_i given by

$$\mathbf{V}_K = \Sigma_K \mathbf{A}_K^T \mathbf{y}/\sigma^2 = (\mathbf{A_K}^T \mathbf{A_K} + \sigma^2 \mathbf{D_K})^{-1} \mathbf{A}_K^T \mathbf{y}$$

in the final iteration. If the measurements were noise-free this would be the exact recovery. The values of estimated $X(k_i), i = 1,2,\ldots,K$ are shown in Fig.8(g). The diagonal values of Σ_K are the variances of the estimated $X(k_i)$. The parameters d_i for the estimated nonzero coefficients tend to zero.

5.1 Bayesian-Based Reconstruction with Contiguous Structures/Patterns

Contiguous structure/pattern can be readily exploited in Bayesian-based reconstruction by imposing proper priors [53, 118, 119], which provides additional robustness in the signal reconstruction and parameter estimation. For example, in practice, the measured data may experience missing samples [96],

This is a pre-print of an article published in Circuits Systems and Signal Processing.
The final authenticated version is available online at: https://doi.org/10.1007/s00034-018-0909-2

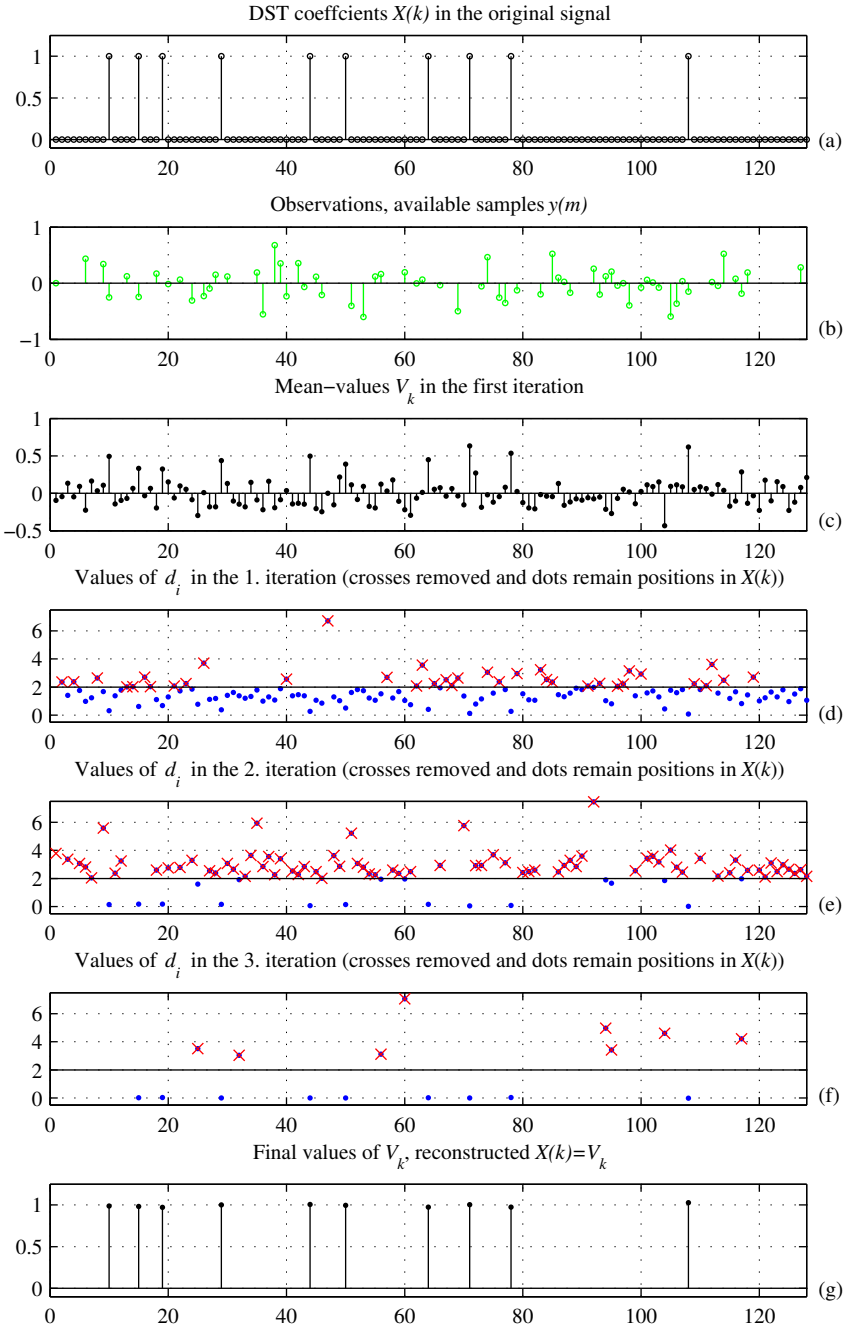

Figure 8: Bayesian reconstruction of a signal sparse in the DST domain. (a) Signal coefficients in the DST domain. (b) Available samples/measurements. (c) Distribution mean-values (estimation of coefficients) in the initial iteration. (d)-(f) Hyperparameters in the first, second, and third iteration with a threshold. (g) The final mean-value (the final estimated DST coefficients) at the positions of remaining hyperparameters from (f).

which induce noise-like artifacts in the time-frequency distributions [51]. As desirable signal components generally exhibit sparsity in a certain domain, structure-aware Bayesian method, together with the aid of time-frequency techniques, can effectively mitigate the missing-sample artifacts and robustly recover the useful signal [51, 52]. This approach has been successfully applied to the co-existence of radio telescope arrays with frequency-hoping wireless communication systems [50].

6 Applications in Signal Processing

6.1 Audio Signals

The presented reconstruction will be illustrated on audio signals [82, 87]. Audio signals are nonstationary with changing spectral content in time. In general, they are not sparse. The sparsity can be improved considering the localized segments of audio signals. These kind of signals can then be considered as approximately sparse. In order to improve the sparsity of audio signals, a windowed form of the DCT is used, as in the case of MDCT, widely employed in compression procedures involved in modern audio formats. Long duration audio signals $x(n)$ are analyzed with the DCT applied on consecutive blocks of windowed signals

$$x_m(n) = w(n)x(n + mN/2),$$

where $w(n)$ is a window within $0 \leq n \leq N-1$. The subsequent blocks are overlapped such that the second half of one block coincides with the first half of the subsequent block. It is important to note that such a block-based approach in the analysis and processing of audio signals is adequate for the presented sparse signal reconstruction algorithms involving pseudo-inversion of partial DCT matrices (since it reduces the dimensionality of the problem). Note that the block approach is also used in the DCT based image analysis as well. If the window form satisfies the condition $w(n) + w(n + N/2) = 1$ within the overlapping interval, $N/2 \leq n \leq N-1$, then the reconstruction is quite simple from the reconstructed windowed segments $x_{mR}(n)$ as

$$x_R(n) = \sum_m x_{mR}(n - mN/2).$$

Many window forms satisfy this condition [91]. The most commonly used window among them is the Hann window $w(n) = 0.5(1 + \cos(\frac{2\pi}{N}(n + \frac{N}{2}))) = \sin^2(\frac{\pi}{N}n)$. The condition that $w(n) + w(n + N/2) = 1$ within the overlapping interval, corresponds to well known Princen-Bradley condition $w^2(n) + w^2(n + N/2) = 1$ within the overlapping interval. If the same window is used on both, the analysis and reconstruction side. Then the window $w(n) = \sin(\frac{\pi}{N}n)$ would be used. In our example the window is used on the analysis side only.

Next we will assume that a reduced set of signal samples is available (sensed) or reliable. Various circumstances may cause the unavailability of audio signal samples. One illustrative example includes clicks and pops present in old recordings that significantly corrupt a certain percent of samples. After removal of these impulsive disturbances, these randomly positioned samples can be considered as unavailable, and reconstructed using the presented CS-based approach [83, 84, 100, 102].

The audio signal "mtlb.wav" from MATLAB is considered. Its form is shown in Fig.9. The signal is corrupted by impulsive noise in 15% of randomly positioned samples. Positions of the impulsive noise can easily be detected using a limiter (a method for more complex cases detection of impulsive noise with amplitudes within the signal values can be found in [93]). The signal samples at the positions of strong noise are considered as unavailable and the reconstruction of the signal is performed using the rest of samples on blocks with a Hann window of length $N = 500$ with half of window length overlapping. Reconstruction is performed using the presented OMP algorithm with various assumed sparsities K. The estimated error in the signal is calculated along with the one presented by (37). The estimated error is presented by "*" and the one expected in theory by a line. The agreement is high. The reconstructed signal segments are added up and the final reconstructed signal is presented in Fig.9 for the case of assumed sparsity $K = 150$.

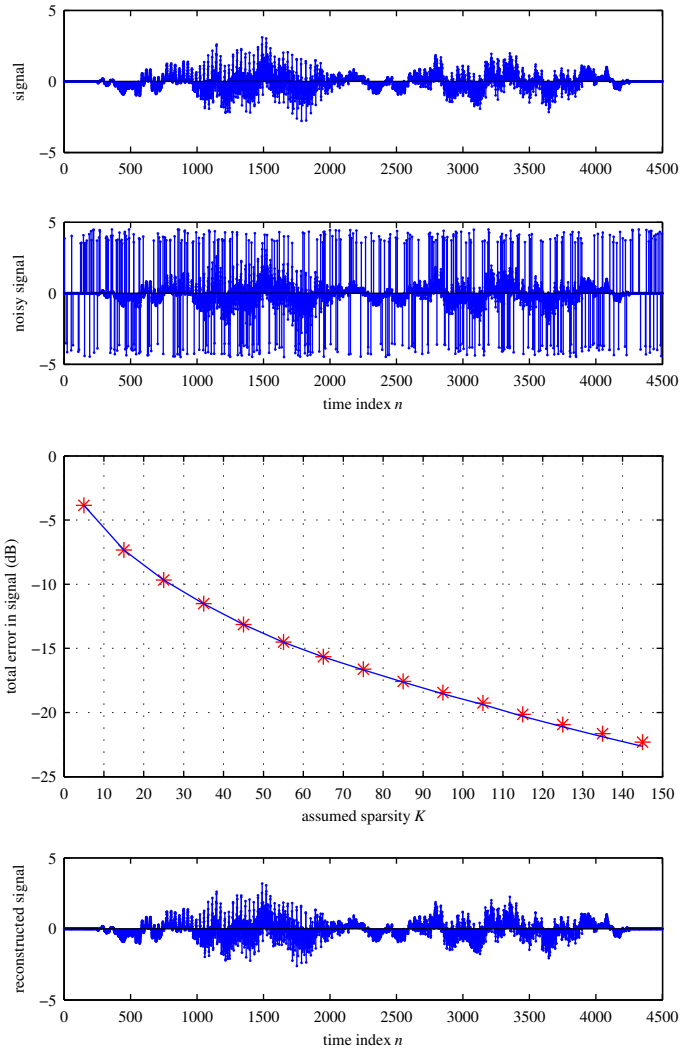

Figure 9: Error energy in the reconstruction (Section 3.2) of noisy non-sparse signal - calculated numerically and according to the theory (Section 3.8). Error is shown for various assumed sparsities.

6.2 Image Reconstruction

The gradient-based reconstruction algorithm is applied to image reconstruction. In general, it can be assumed that common images are sparse in the two-dimensional discrete cosine transform (2D-DCT) domain [9, 14, 22, 39, 49, 73, 83, 121]. That is true for smaller image blocks used in the 2D-DCT image transform as well. Sparsity assumptions imply that the image can be reconstructed from a reduced set of measurements (pixels). Assume that an image is degraded by salt and pepper noise Fig.10 (middle). The disturbed pixels can be easily detected, omitted and marked as unavailable. The undisturbed pixels are considered as the available measurements. They are rearranged into a column vector form and denoted by $\mathbf{y}$. The sparsity domain is the 2D-DCT. The 2D-DCT coefficients are also rearranged into a column vector form and denoted by $\mathbf{X}$. The measurement matrix $\mathbf{A}$ follows from the 2D-DCT coefficients (rearranged into matrix form),

$$y(i) = x(m_i, n_i) = \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} v_k v_l \cos\left(\frac{2\pi(2m_i+1)k}{4N}\right) \cos\left(\frac{2\pi(2n_i+1)l}{4N}\right) X(k,l),$$

where $v_0 = \sqrt{1/N}$ and $v_k = \sqrt{2/N}$ for $k \neq 0$ and N is the size of the blocks used in the 2D-DCT analysis. Rows corresponding to the available pixels are retained from the full 2D-DCT matrix.

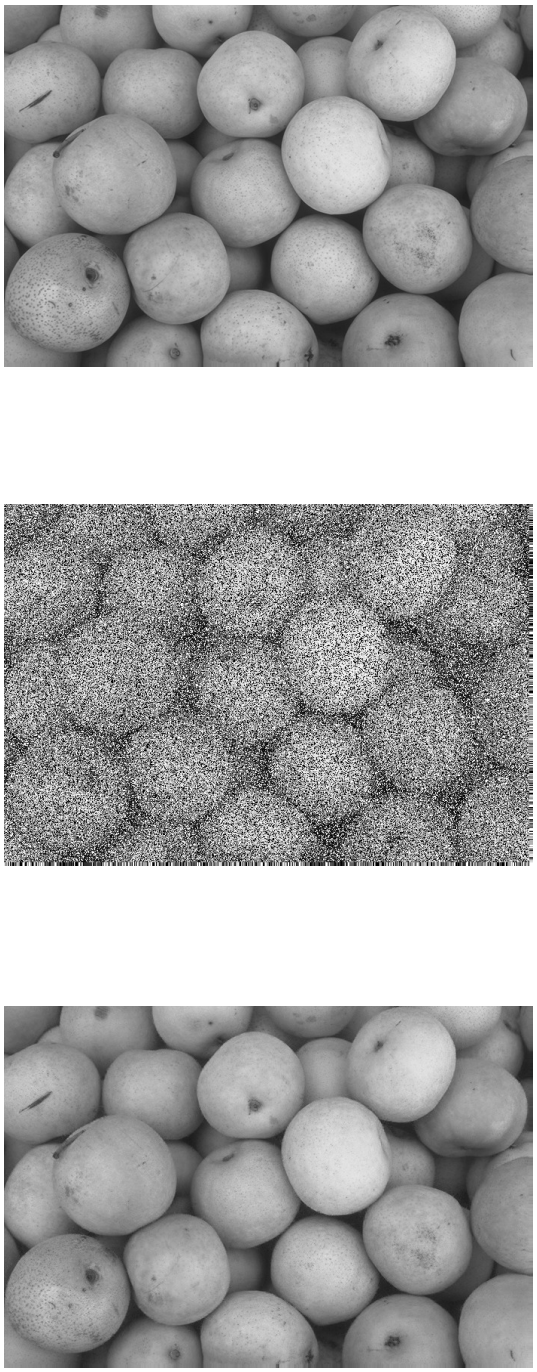

Figure 10: Image reconstruction from a reduced set of pixels: Original image (top). Image with 50% of salt and pepper noisy pixels considered as unavailable (middle). Reconstructed image (bottom).

The task then reduces itself to the described minimization problem. Find the signal, or in this case unavailable pixels, by minimizing the ℓ_1-norm of the transform, subject to the available samples, i.e., $\min \|\mathbf{X}\|_1$ subject to $\mathbf{y} = \mathbf{AX}$.

The unavailable pixels are iterated using the presented gradient algorithm, until the desired accuracy is achieved. The resulting reconstructed image is shown in Fig.10. The mean absolute error in the noisy and reconstructed image is $MAE = 40.77$, and $MAE = 0.30$, respectively. The peak signal-to-noise ratio is $PSNR = 27.85$[dB], and $PSNR = 44.78$[dB], respectively.

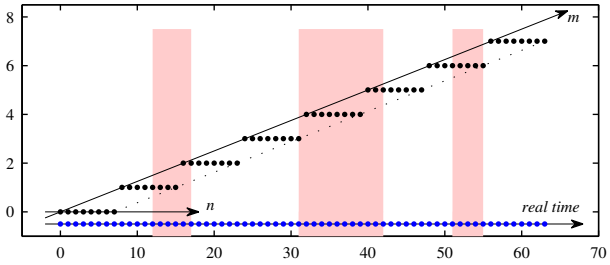

Figure 11: Illustration of one revisit (chirp series) discretization in coordinates m (chirp index, slow time) and n (time within one chirp, fast time), along with a real time. The case of $N_c = 8$ chirps in one revisit and $N = 8$ samples within a chirp is presented. The CIT is 64 samples. Unavailable or heavily corrupted data are marked by red.

6.3 Application to Radar (ISAR) Signal Processing

In inverse synthetic aperture radar (ISAR) a high resolution image of a target is obtained by using the two-dimensional Fourier transform (FT) of a received (and processed) signal. The ISAR image of a point target is a highly concentrated two-dimensional function whose peak position corresponds to the target's range and cross-range. For a number of scattering points, the radar image consists of several peaks at the range and cross-range positions. Usually the number (area) of nonzero values in the ISAR image is small as compared to the total number of signal samples. Thus, we may say that a common signal in ISAR is sparse in the two-dimensional Fourier domain. As such it can be reconstructed from much fewer samples than the sampling theorem requires. Unavailable, randomly positioned samples could also result from heavily corrupted parts of the signal, that are omitted and declared as unavailable, before the ISAR image recovery and calculation is done. Measurements and physical constraints of the target and radar interferences may also cause only some of the randomly positioned data to be received and measured. The fact that the two-dimensional FT domain is the domain of the ISAR signal sparsity is used in signal recovery [31,38,43,47,71,75,94,95,97,98,115].

The received signal, from the ith scattering point, after the distance compensation, is

$$q_i(m,n) = \sigma_i e^{j2\pi\beta_i m/N_c} e^{j2\pi\gamma_i n/N},$$

where β_i and γ_i are the parameters proportional to the cross-range and range. The number of chirps is denoted by N_c, while the number of samples within a chirp is N. The total signal for K scattering points is

$$q(m,n) = \sum_{i=1}^{K} q_i(m,n).$$

Assume that some samples or blocks of samples of the received radar signal are either unavailable or heavily corrupted so that they are omitted from the analysis [86,95]. Assume that the blocks of omitted signal samples are randomly positioned. The two-dimensional FT of this signal is then

$$\hat{Q}(k,l) = \sum_{m=0}^{M-1} \sum_{n\in\mathbb{M}_m} q(m,n) e^{-j(\frac{2\pi mk}{N_c} + \frac{2\pi nl}{N})}. \tag{62}$$

It can happen that the unavailable/corrupted data are: all within one chirp or spread over two or more chirps, including the possibility that a few chirps in a row are affected in this way, Fig. 11. These cases are included by using the notation $n \in \mathbb{M}_m$ where $\mathbb{M}_m$ is the set of available samples within the mth chirp. For some m it could also happen that $\mathbb{M}_m = \varnothing$, i.e., that there are no available samples within that chirp. The total number of available samples is $1 \ll M \leq N_c N$.

The full data set based image is shown in Fig.12(a). The delta-wing data set was collected using $T_r = 1/2000$ s. Each range profile is generated in 0.5 ms and each profile had 41 bins. The total data set contains 60000 samples. The delta-wing was at a range of 2 km and was rotating at 2 ^{o}/s. Only the data corresponding to the interval of 61 range bins (where the target was located) are used for calculation and presentation. Data blocks of a random duration are considered as unavailable/corrupted, like in Fig.12(b). The reconstruction results in this case are presented in Fig.12(d).

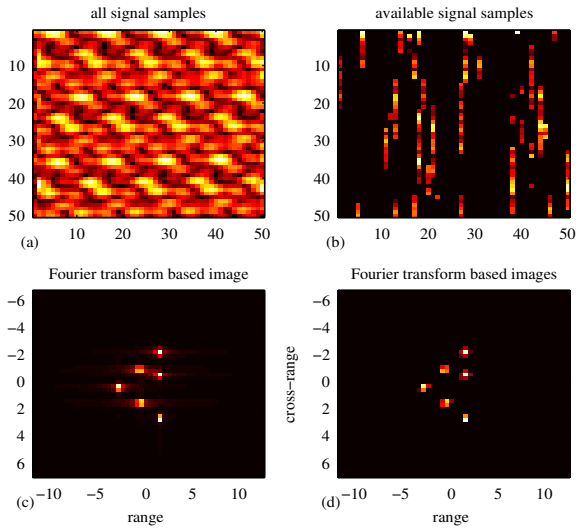

Figure 12: Analysis of radar data of 6-point scatterer based on 12.5% of the signal samples missing in blocks. In the reconstruction $K = 30$ is used.

Another illustrative example assumes the simulated 2D ISAR signal of an aircraft, exhibiting sparsity in the 2D DFT domain. The detailed description of the signal model is given in [71]. In Fig.13(a) we present the 2D DFT of the full data set. Further, it is assumed that only 25% of the data are available, where the available samples are randomly selected and represented as a measurement vector $\mathbf{y}$ having M samples. The initial 2D DFT is calculated using the values of available measurements. It is shown in Fig.13(b). Next, the reconstruction algorithm is applied to recover the missing data. The 2D version of the threshold based reconstruction algorithm (Algorithm 3) is applied, where the variance is calculated as:

$$\sigma^2 = EM\frac{(N_cN - M)}{(N_cN - 1)}, \tag{63}$$

where E represents the energy estimated as:

$$E = \sum_{m=0}^{M-1} \sum_{n\in\mathbb{M}(m)} |q(m,n)|/M \tag{64}$$

The 2D threshold is then calculated using (32) with $N = N_cN$, with P being set to the value 0.99. In order to obtain efficient results, the iterative version of the algorithm is employed. Namely after applying a threshold τ and detecting the support of signal components being above τ, the least-squared minimization is performed. Then, the influence of detected components is removed from the measurements. After updating the measurements, we need to recalculate the amplitude, variance, and threshold. Then **Algorithm 3** can be repeated again. The reconstruction results are shown in Fig.13(c).

6.4 ECG signal reconstruction

ECG signals belongs to the group of biomedical signals used for analysis and detection of different heart diseases [46,56,78]. Particularly, the specific waves of ECG signals called QRS complexes play an important role in medical diagnoses and treatments. It has been shown that the compact representation of ECG signals, especially the QRS complexes can be achieved using the Hermite transform allowing fast processing and compression of ECG signals [11,20,21,30,54,59,69,79,80,101,105,117]. Due to their specific shape, the Hermite basis functions are found to be suitable for the representation of QRS complexes. Therefore, these signals can be represented using just a few Hermite coefficients. Consequently, the Hermite transform representation allows us to exhibit the sparsity of QRS complexes, bringing

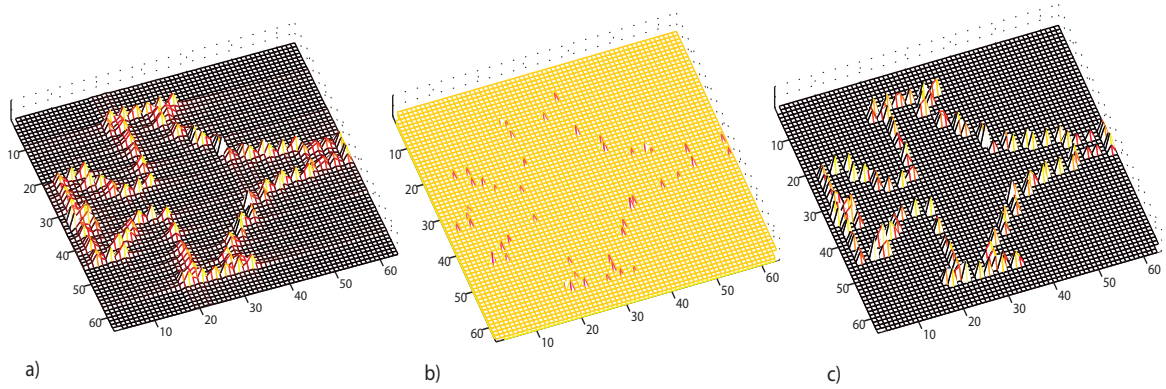

Figure 13: A single iteration based algorithm applied to the simulated radar signal: a) 2D DFT of the full data (signal) set; b) Initial 2D DFT of the signal with 25% of available samples, c) 2D DFT of the reconstructed signal

advantages for their application to compressive sensing, but also for the automatic recognition and classification of QRS complexes.

The Hermite basis functions are easily calculated using the following recursive formula [59,69,101]:

$$\psi_0(t_n) = \frac{1}{\sqrt[4]{\pi}} e^{-\frac{t_n^2}{2\sigma^2}}, \ \psi_1(t_n) = \frac{\sqrt{2}t_n}{\sqrt[4]{\pi}} e^{-\frac{t_n^2}{2\sigma^2}},$$
$$\psi_i(t_n) = t_n \sqrt{\frac{2}{i}} \psi_{i-1}(t_n) - \sqrt{\frac{i-1}{i}} \psi_{i-2}(t_n), \ i \geq 2,$$

where it is assumed that the Hermite basis functions are sampled at N points t_n being the roots of the N-th order Hermite polynomial. The Hermite transform coefficients are usually calculated using the Gauss-Hermite quadrature:

$$C_i = \frac{1}{N} \sum_{n=1}^{N} \frac{\psi_i(t_n)}{[\psi_{N-1}(t_n)]^2} x(t_n), i = 0,1,...,N-1,$$

where $x(t_n)$ is a signal sampled at the zero of the N-th order Hermite polynomial, while N is the length of the signal.

The Hermite transform $N \times N$ matrix $\mathbf{\Phi}$ has the elements $\phi(i,n) = \psi_i(t_n)/[(\psi_{N-1}(t_n))^2 N]$, for $i = 0,1,\ldots,N-1$, and $n = 1,2,\ldots,N$. Vector of the transform domain coefficients is given by

$$\mathbf{C} = \mathbf{\Phi x}.$$

In the case we are dealing with a smaller number of measurements compared to the signal length N, the measurement vector y can be written as follows:

$$\mathbf{y} = \mathbf{Bx} = \mathbf{B\Psi C},$$

where $\mathbf{\Psi}$ is the inverse Hermite transform matrix with elements $\psi_i(t_n)$.

The compressive sensing matrix is $\mathbf{A} = \mathbf{B\Psi}$. Usually, the CS matrix $\mathbf{A}$ can be efficiently constructed as a partial random inverse Hermite transform.

As already mentioned, in order to provide the most compact or the sparsest representation in the Hermite transform domain, the discrete signals of length N should be sampled at non-uniform points being proportional to the roots of the N-th order Hermite polynomial. Since the signals are usually sampled uniformly, we can apply the sinc interpolation formula to obtain the values at the desired

non-uniform points:

$$x(\lambda t_n) \approx \sum_{m=-K}^{K} x(m\Delta t) \frac{\sin\left(\pi(\lambda t_n - m\Delta t)/\Delta t\right)}{\pi(\lambda t_n - m\Delta t)/\Delta t},$$

where $n = 1,...,N$, $m = -K,...,K$ and Δt is the sampling period. Now the scaling factor σ (that allows stretching and compressing of the Hermite functions) should be adapted to the analyzed signal, in this case the ECG signal or its QRS complex. Instead of doing so, we can use a fixed value $\sigma = 1$ and introduce the signal time-axis scaling factor λ. The optimal value of λ should produce the sparsest representation. It is obtained as a solution of the following optimization problem:

$$\lambda_{opt} = \min_{\lambda} \|\mathbf{C}_\lambda\|_1 = \min_{\lambda} \|\mathbf{\Phi}\mathbf{x}_\lambda\|_1 = \min_{\lambda} \|\mathbf{\Phi}\mathbf{A}_\lambda\mathbf{x}\|_1,$$

where $\mathbf{\Phi}$ is the Hermite transform matrix, $\mathbf{x}_\lambda$ is the vector with elements $x(\lambda t_n)$, and $\mathbf{A}_\lambda$ is the matrix with elements

$$a_{nm} = \frac{\sin\left[\pi\left(\lambda t_n - (m-K-1)\Delta t\right)/\Delta t\right]}{\pi\left(\lambda t_n - (m-K-1)\Delta t\right)/\Delta t},$$

with $n, m = 1,2,...N$.

The optimization can be done using the presented steepest descent approach with iterative parameter λ variations.

A real ECG signal is considered in this example. The QRS complexes are extracted from the observed signal and the time domain representation of one of its QRS complexes is shown in Fig.14(a). Note that this is the original full signal that will be used as a reference in order to test the efficiency of the CS reconstruction. In order to achieve sparsity in the Hermite transform domain, the QRS complex should be sampled at the points proportional to the roots of the Hermite polynomial. The corresponding sparse Hermite domain representation is shown in Fig.14(b), proving that the QRS signals can be characterized by only a few low-order Hermite coefficients. The available measurements are depicted in Fig.14(c).

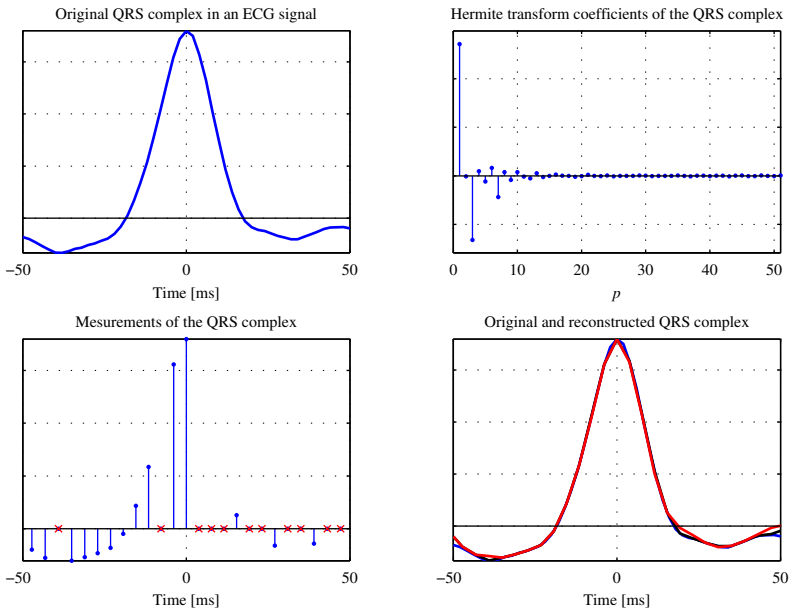

Figure 14: Compressed sensing applied to ECG signals: a) Original QRS complex extracted from an ECG signal, b) Hermite transform representation of the QRS complex, c) available measurements, d) reconstructed QRS complex.

The reconstruction of QRS signals is performed using the gradient algorithm (Algorithm 5), [21]. The amount of missing samples that are reconstructed in each QRS signal is approximately 45%. The

reconstructed QRS complex is shown in Fig.14(d), where we may observe good match with the original QRS signals. The value of parameter λ is $\lambda = 6.2/500 = 0.0124$ sec. Generally, when taking into account all QRS complexes from the same ECG signal, the optimal values of parameter λ_{opt} calculated using the presented optimization algorithm range between $6.1/500$ and $6.4/500$. Since these values are very close to each other, we use one of the values from this range for the representation and reconstruction of all QRS signals ($\lambda = 6.2/500$).

7 Conclusions

Signals that are sparse in one of their representation domains can be reconstructed from a reduced set of measurements. Conditions to reconstruct these signals are reviewed and discussed. Several methods for signal reconstruction are presented, along with the pseudocode algorithms and examples.

The presentation starts with the direct search method. It is simple and well defined, however computationally not feasible for practical problem dimensions. The second group of reconstruction methods is based on the minimization of the number of nonzero coefficients in the signal sparsity domain. These methods belong to the group of matching pursuit approaches. The next group of methods is based on the reformulation of the presented problem into a constrained convex problem. The problem is then solved by using well established linear programming tools. These methods belong to the group of basis pursuit approaches. Finally the Bayesian formulation of the reconstruction problem is presented, along with a specific algorithm to reconstruct a sparse signal from a reduced set of measurements.

A reduced set of measurements is interesting in practice from several points of view. It can result from the desire to sense a sparse signal with the lowest possible number of measurements (compressive sensing). It can also result from the physical unavailability to take a complete set of measurements/samples. Then the unavailable samples are reconstructed to the full signal, using the sparsity property. It also can happen that some arbitrarily positioned samples of a signal are heavily corrupted by disturbances making it better to omit them and consider them as unavailable in the analysis and to try to reconstruct the signal with a reduced set of uncorrupted samples. In this way efficient filtering and disturbance removal can be achieved. Practical areas of application are numerous and can be found in many recent papers.

Although compressive sensing and its applications have been intensively studied in the last decade, there are still open questions in this field, like for example, computationally efficient uniqueness tests and optimal sampling strategies. Adaptive dictionary learning in compressive sensing, based on the machine learning algorithms, is another challenging research direction. Structure-aware Bayesian approach, together with the time-frequency representations, can provide a good basis for improvement of the compressive sensing based reconstruction robustness. Although many applications of compressive sensing in signal processing are already well described and implemented, this field remains an open area for researches and engineers, with many improvements and new applications still to be discovered.

References

[1] 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, 2006, pp. 4311-4322.

[2] N. Ahmed, T. Natarajan and K. R. Rao, "Discrete Cosine Transform," in *IEEE Transactions on Computers*, vol. C-23, no. 1, 1974, pp. 90-93.

[3] M. G. Amin, *Compressive Sensing for Urban Radar*, CRC Press, 2014.

[4] D. Angelosante, G. B. Giannakis, and E. Grossi, "Compressed sensing of time-varying signals," *Proc. of the 16th International Conference on Digital Signal Processing (DSP '09)*, pp. 1-8, Santorini-Hellas, Greece, July 2009.

[5] E. Arias-Castro and Y. Eldar, "Noise folding in compressed sensing," *IEEE Signal Processing Letters*, vol. 18, no. 8, 2011, pp. 478-481.

[6] S. D. Babacan, R. Molina, A, K. Katsaggelos, "Bayesian Compressive Sensing Using Laplace Priors," *IEEE Transactions on Image Processing*, vol. 19, no. 1, 2010, pp. 53-63.

[7] A. S. Bandeira, E. Dobriban, D. G. Mixon and W. F. Sawin, "Certifying the Restricted Isometry Property is Hard," *IEEE Transactions on Information Theory*, vol. 59, no. 6, 2013, pp. 3448-3450.

[8] R. Baraniuk, "Compressive sensing," *IEEE Signal Processing Magazine*, vol. 24, no. 4, 2007, pp. 118–121.

[9] R. G. Baraniuk, T. Goldstein, A. C. Sankaranarayanan, C. Studer, A. Veeraraghavan, M. B. Wakin, "Compressive Video Sensing: Algorithms, architectures, and applications," *IEEE Signal Processing Magazine*, vol. 34, no. 1, 2017, pp. 52-66.

[10] D. Baron, S. Sarvotham, R. G. Baraniuk, "Bayesian Compressive Sensing via Belief Propagation," *IEEE Transactions on Signal Processing*, vol. 58, no. 1, 2010, pp. 269-280.

[11] J. Bazerque, G. Giannakis, "Distributed spectrum sensing for cognitive radio networks by exploiting sparsity," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, 2010, pp. 1847-1862.

[12] A. Beck, M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal Imaging Sci.*, vol. 2, no. 1, 2009, pp. 183-197.

[13] C. R. Berger, Z. Wang, J. Huang and S. Zhou, "Application of compressive sensing to sparse channel estimation," *IEEE Communications Magazine*, vol. 48, no. 11, 2010, pp. 164-174.

[14] J. M. Bioucas-Dias, M. A. T. Figueiredo, "A New TwIST: Two-Step Iterative Shrinkage/Thresholding Algorithms for Image Restoration," *IEEE Transactions on Image Processing*, vol. 16, no. 12, 2007, pp. 2992-3004.

[15] J. D. Blanchard, . Cartis, and J. Tanner, "Compressed Sensing: How Sharp Is the Restricted Isometry Property?," *SIAM Rev.*, vol. 53, no. 1, 2011, pp.105-125.

[16] T. Blumensath, M. E. Davies, "Gradient Pursuits," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, 2008, pp. 2370-2382.

[17] T. Blumensath, M. E. Davies, "Iterative Thresholding for Sparse Approximations," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, 2008, pp 629-654.

[18] J. Bobin, J. L. Starck, R. Ottensamer, "Compressed Sensing in Astronomy," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 5, 2008, pp. 718-726.

[19] S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[20] M. Brajovic, I. Orović, M. Daković, and S. Stanković, "On the Parameterization of Hermite Transform with Application to the Compression of QRS Complexes," *Signal Processing*, vol. 131, 2017, pp. 113-119.

[21] M. Brajovic, I. Orović, M. Daković, and S. Stanković, "Gradient-based signal reconstruction algorithm in the Hermite transform domain," *Electronics Letters*, vol. 52, no. 1, 2016, pp.41-43.

[22] M. Brajovic, I. Stanković, M. Daković, C. Ioana, and L. Stanković, "Error in the Reconstruction of Nonsparse Images," *Mathematical Problems in Engineering*, Volume 2018 (2018), Article ID 4314527, 10 pages, https://doi.org/10.1155/2018/4314527

[23] L. Breiman, "Better Subset Regression Using the Nonnegative Garrote," *Technometrics*, Taylor and Francis, Ltd., vol. 37, no. 4, 1995, pp. 373-384.

[24] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9-10, 2008, pp. 589-592.

[25] E. J. Candès, J. Romberg, "ℓ_1-magic : Recovery of Sparse Signals via Convex Programming, " Caltech, http://users.ece.gatech.edu/justin/ l1magic/downloads/l1magic.pdf, Oct.2005.

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

[27] E. J. Candès, M. Wakin, "An Introduction to Compressive Sampling," *IEEE Signal Processing Magazine,* vol. 25, no. 2, 2008, pp. 21-30.

[28] R. Chartrand and V. Staneva, "Restricted isometry properties and nonconvex compressive sensing," *Inverse Problems*, vol. 24, no. 3, 2008, pp. 035020-1-14.

[29] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic Decomposition by Basis Pursuit," *SIAM Review*, vol. 43, no. 1, 2001, pp. 129-159.

[30] D. Craven, B. McGinley, L. Kilmartin, M. Glavin, E. Jones, "Compressed Sensing for Bioelectric Signals: A Review," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 2, 2015, pp. 529-540.

[31] S. Costanzo, A. Rocha, M. D. Migliore, "Compressed Sensing: Applications in Radar and Communications," *The Scientific World Journal*, Volume 2016, Article ID 5407415, 2 pages, Editorial, 2016.

[32] I. Daubechies, M. Defrise, C. De Mol "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, 2004, pp. 1413-1457.

[33] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *Constructive approximation*, vol. 13, no. 1, 1997, pp. 57-98.

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

[35] D. L. Donoho, M. Elad, V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Transactions on Information Theory*, vol. 52, no. 1, 2006, pp. 6-18.

[36] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing,* Springer, 2010.

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

[38] J. Ender, "On compressive sensing applied to radar," *Signal Processing*, vol. 90, no. 5, 2010, pp. 1402-1414.

[39] N. Eslahi, A. Aghagolzadeh, "Compressive Sensing Image Restoration Using Adaptive Curvelet Thresholding and Nonlocal Sparse Regularization," *IEEE Transactions on Image Processing*, vol. 25, no. 7, 2016, pp. 3126-3140.

[40] M. A. Figueiredo, R. D. Nowak, 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, 2007, pp. 586-597.

[41] P. Flandrin, P. Borgnat, "Time-Frequency Energy Distributions Meet Compressed Sensing," *IEEE Transactions on Signal Processing*, vol.58, no.6, 2010, pp.2974-2982.

[42] M. Fornsaier, H. Rauhut, "Iterative thresholding algorithms," *Applied and Computational Harmonic Analysis*, vol. 25, no. 2, 2008, pp. 187- 208.

[43] M. A. Hadi, S. Alshebeili, K. Jamil, F. E. Abd El-Samie, "Compressive sensing applied to radar systems: an overview," *Signal, Image and Video Processing*, vol. 9, 2015, pp. 25-39.

[44] G. Hua, Y. Hiang, G. Bi, "When Compressive Sensing meets Data Hiding," *IEEE Signal Processing Letters*, vol. 23, no. 4, 2016, pp. 473-477.

[45] S. Ji, Y. Xue, L. Carin, "Bayesian Compressive Sensing," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, 2008, pp. 2346-2356.

[46] P. Lander and E. J. Berbari, "Principles and signal processing techniques of the high-resolution electrocardiogram," *Progress in Cardiovascular Diseases* vol. 35, no. 3, 1992, pp. 169-188.

[47] C. Li, G. Zhao, W. Zhang, Q. Qiu, H. Sun, "ISAR Imaging by Two-Dimensional Convex Optimization-Based Compressive Sensing," *in IEEE Sensors Journal*, vol. 16, no. 19, 2016, pp. 7088-7093.

[48] X. Li, G. Bi, "Time-frequency representation reconstruction based on the compressive sensing," *9th IEEE Conference on Industrial Electronics and Applications*, Hangzhou, 2014, pp. 1158-1162.

[49] X. Liao, K. Li, J. Yin, "Separable data hiding in encrypted image based on compressive sensing and discrete Fourier transform," *Multimedia Tools and Applications*, pp. 1-15, 2016.

[50] S. Liu, Y. D. Zhang, and T. Shan, "Detection of weak astronomical signals with frequency-hopping interference suppression," *Digital Signal Processing,* vol. 72, 2018, pp. 1-8.

[51] S. Liu, Y. D. Zhang, T. Shan, S. Qin, and M.G. Amin, "Structure-aware Bayesian compressive sensing for frequency-hopping spectrum estimation," *in Proc. SPIE 9857, Compressive Sensing V: From Diverse Modalities to Big Data Analytics*, pp. 98570N, May 2016.

[52] S. Liu, Y. D. Zhang, T. Shan, and R. Tao, "Structure-aware Bayesian compressive sensing for frequency-hopping spectrum estimation with missing observations," *IEEE Transactions on Signal Processing,* vol. 66, no. 8, 2018, pp. 2153-2166.

[53] S. Liu, J. B. Jia, and Y. J. Yang, "Image reconstruction algorithm for electrical impedance tomography based on block sparse Bayesian learning," *in Proc. IEEE International Conference on Imaging Systems and Techniques (IST),* Beijing, China, Oct. 18-20, 2017.

[54] Y. Liu, M. De Vos, S. Van Huffel, "Compressed Sensing of Multichannel EEG Signals: The Simultaneous Cosparsity and Low-Rank Optimization," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 8, 2015, pp. 2055-2061.

[55] W. Lu and N. Vaswani, "Regularized Modified BPDN for Noisy Sparse Reconstruction With Partial Erroneous Support and Signal Value Knowledge," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, 2012, pp. 182-196.

[56] S. Luo and P. Johnston "A review of electrocardiogram filtering," *Journal of Electrocardiology*, vol. 43, no. 6, 2010, pp. 486-496.

[57] X. Lv, G. Bi and C. Wan, "The Group Lasso for Stable Recovery of Block-Sparse Signal Representations," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, 2011, pp. 1371-1382.

[58] S. Mallat, Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, 1993, pp. 3397-3415.

[59] J. B. Martens, "The Hermite transform - Theory," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, no. 9, 1990, pp. 1595-1606.

[60] S. A. Martucci, "Symmetric convolution and the discrete sine and cosine transforms," *IEEE Transactions on Signal Processing,* vol. 42, no. 5, 1994, pp. 1038-1051.

[61] J. Music, T. Marasovic, V. Papic, I. Orović, and S. Stanković, "Performance of compressive sensing image reconstruction for search and rescue," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 11, 2016, pp. 1739-1743.

[62] J. Music, I. Orović, T. Marasovic, V. Papic, and S. Stanković, "Gradient Compressive Sensing for Image Data Reduction in UAV based Search and Rescue in the Wild," *Mathematical Problems in Engineering*, 2016.

[63] D. Needell, J. A. Tropp, "CoSaMP: Iterative signal recovery from noisy samples," *Applied and Computational Harmonic Analysis*, 2008. DOI: 10.1016/j.acha.2008.07.002.

[64] D. Needell, J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *ACM Technical Report*, 2008-01, California Institute of Technology, Pasadena, July 2008.

[65] D. Needell, J. A. Tropp, "CoSaMP: Iterative Signal Recovery from Incomplete and Inaccurate Samples," *Communicationt of the ACM,* vol. 53, no.12, 2010, pp.93-100.

[66] B. Ophir, M. Lustig, and M. Elad, "Multi-scale dictionary learningusing wavelets," *IEEE Journal Selected Topics Signal Processing,* vol. 5, no. 5, 2011, pp. 1014-1024.

[67] I. Orović, V. Papic, C. Ioana, X. Li, and S. Stanković, "Compressive Sensing in Signal Processing: Algorithms and Transform Domain Formulations," *Mathematical Problems in Engineering*, 2016.

[68] I. Orović, and S. Stanković, "Improved Higher Order Robust Distributions based on Compressive Sensing Reconstruction," *IET Signal Processing*, vol. 8, no. 7, 2014, pp. 738-748.

[69] I. Orović, S. Stanković, T. Chau, C. M. Steele, and E. Sejdic, "Time-frequency analysis and Hermite projection method applied to swallowing accelerometry signals," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, Article ID 323125, 7 pages.

[70] I. Orović, S. Stanković, and T. Thayaparan, "Time-Frequency Based Instantaneous Frequency Estimation of Sparse Signals from an Incomplete Set of Samples," *IET Signal Processing, Special issue on Compressive Sensing and Robust Transforms*, vol. 8, no. 3, 2014, pp. 239-245.

[71] C. Ozdemir, *Inverse Synthetic Aperture Radar Imaging With MATLAB Algorithms*, John Wiley & Sons, 2012.

[72] M. Panic, J. Aelterman, V. S. Crnojevic, A. Pizurica, "Compressed sensing in MRI with a Markov random field prior for spatial clustering of subband coefficients," *Proceedings of the EUSIPCO*, 2016, pp.562-566.

[73] V. M. Patel and R. Chellappa, "Sparse Representations and Compressive Sensing for Imaging and Vision," *Springer Briefs in Electrical and Computer Engineering*, 2013.

[74] G. Pope, "Compressive Sensing: a Summary of Reconstruction Algorithms," Eidgenossische Technische Hochschule, Zurich, Switzerland, 2008; http://e-collection.library.ethz.ch/eserv/eth:41464/eth-41464-01.pdf

[75] L. C. Potter, E. Ertin, J. T. Parker, M. Cetin, "Sparsity and Compressed Sensing in Radar Imaging," *Proceedings of the IEEE*, vol. 98, no. 6, 2010, pp. 1006-1020.

[76] S. Qaisar, R. M. Bilal, W. Iqbal, M. Naureen and S. Lee, "Compressive sensing: From theory to applications, a survey," *Journal of Communications and Networks*, vol. 15, no. 5, 2013 pp. 443-456.

[77] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proceeding of the IEEE,* vol. 98, no. 6, 2010, pp. 1045-1057.

[78] R. Sameni , G. D. Clifford "A Review of Fetal ECG Signal Processing Issues and Promising Directions," *The Open Pacing, Electrophysiology and Therapy Journal*, vol. 3, 2010, pp. 4-20.

[79] A. Sandryhaila, S. Saba, M. Puschel, J. Kovacevic, "Efficient compression of QRS complexes using Hermite expansion," *IEEE Transactions on Signal Processing*, vol. 60, no. 2, 2012, pp. 947-955.

[80] A. Sandryhaila, J. Kovacevic, M. Puschel, "Compression of QRS complexes using Hermite expansion," *IEEE int. Conference on Acoust., Speech and Signal Process., ICASSP*, Prague, 2011, pp. 581-584.

[81] E. Sejdic, "Time-Frequency Compressive Sensing", in *Time-Frequency Signal Analysis and Processing*, ed. B. Boashash, Academic Press, Nov. 2015. pp.424-429.

[82] I. Stanković, C. Ioana, and M. Daković, "On the reconstruction of nonsparse time-frequency signals with sparsity constraint from a reduced set of samples," *Signal Processing*, vol. 142, 2018, pp. 480-484.

[83] I. Stanković, I. Orović, M. Daković, and S. Stanković, "Denoising of Sparse Images in Impulsive Disturbance Environment," *Multimedia Tools and Applications*, 2017, doi:10.1007/s11042-017-4502-7.

[84] L. Stanković, *Digital Signal Processing with Applications: Adaptive Systems, Time-Frequency Analaysis, Sparse Signal Processing*, CreateSpace Independent Publishing Platform, 2015.

[85] L. Stanković, "A measure of some time–frequency distributions concentration," *Signal Processing*, vol. 81, 2001, pp. 621-631.

[86] L. Stanković, "On the ISAR Image Analysis and Recovery with Unavailable or Heavily Corrupted Data," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 3, 2015, pp.2093-2106.

[87] L. Stanković, and M. Brajovic, "Analysis of the Reconstruction of Sparse Signals in the DCT Domain Applied to Audio Signals," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 7, 2018, pp.1216-1231.

[88] L. Stanković, M. Daković, "On the Uniqueness of the Sparse Signals Reconstruction Based on the Missing Samples Variation Analysis," *Mathematical Problems in Engineering*, vol. 2015, Article ID 629759, 14 pages, doi:10.1155/2015/629759.

[89] L. Stanković, M. Daković, I. Stanković, and S. Vujovic, "On the Errors in Randomly Sampled Nonsparse Signals Reconstructed with a Sparsity Assumption," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 12, 2017, pp. 2453-2456.

[90] L. Stanković, M. Daković, S. Stanković, and I. Orović, "Sparse Signal Processing - Introduction," *Wiley Encyclopedia of Electrical and Electronics Engineering*, John Wiley & Sons, 2017.

[91] L. Stanković, M. Daković, T. Thayaparan, *Time-Frequency Signal Analysis with Applications*, Artech House, Boston, March 2013.

[92] L. Stanković, M. Daković, S. Vujovic, "Adaptive Variable Step Algorithm for Missing Samples Recovery in Sparse Signals,"*IET Signal Processing*, vol. 8, no. 3, 2014. pp. 246-256.

[93] L. Stanković, M. Daković, S. Vujovic, "Reconstruction of Sparse Signals in Impulsive Disturbance Environments," *Circuits, Systems and Signal Processing*, vol. 2016. pp. 1-28.

[94] L. Stanković, I. Orović, S. Stanković, and M. Amin, "Compressive Sensing Based Separation of Non-Stationary and Stationary Signals Overlapping in Time-Frequency," *IEEE Transactions on Signal Processing*, vol. 61, no. 18, 2013, pp. 4562-4572.

[95] L. Stanković, I. Stanković, and M. Daković, "Nonsparsity Influence on the ISAR Recovery from Reduced Data," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 6, 2016, pp. 3065-3070.

[96] L. Stanković, S. Stanković, M. G. Amin, "Missing Samples Analysis in Signals for Applications to L-Estimation and Compressive Sensing", *Signal Processing*, Elsevier, vol. 94, 2014, pp. 401-408.

[97] L. Stanković, S. Stanković, T. Thayaparan, M. Daković, and I. Orović, "Separation and Reconstruction of the Rigid Body and Micro-Doppler Signal in ISAR Part II - Statistical Analaysis," *IET Radar, Sonar and Navigation*, vol.9, no.9, 2015, pp.1155-1161.

[98] L. Stanković, S. Stanković, T. Thayaparan, M. Daković, and I. Orović, "Separation and Reconstruction of the Rigid Body and Micro-Doppler Signal in ISAR Part I- Theory," *IET Radar, Sonar and Navigation*, vol. 9, no. 9, 2015, pp. 1147-1154.

[99] S. Stanković, and I. Orović, "An Approach to 2D Signals Recovering in Compressive Sensing Context," *Circuits, Systems and Signal Processing*, vol. 36. no. 4, 2017, pp. 1700-1713.

[100] S. Stanković, I. Orović, and M. Amin, "L-statistics based Modification of Reconstruction Algorithms for Compressive Sensing in the Presence of Impulse Noise," *Signal Processing*, vol.93, no.11, 2013, pp. 2927-2931.

[101] S. Stanković, I. Orović, and A. Krylov, "Video Frames Reconstruction based on Time-Frequency Analysis and Hermite projection method," *EURASIP Journal on Advances in Signal Processing, Special Issue on Time-Frequency Analysis and its Application to Multimedia signals*, vol. 2010, Article ID 970105, 11 pages.

[102] S. Stanković, I. Orović, and E. Sejdic, *Multimedia Signals and Systems: Basic and Advanced Algorithms for Signal Processing,* Springer International Publishing, second edition, 2015.

[103] S. Stanković, I. Orović, and L. Stanković, "An Automated Signal Reconstruction Method based on Analysis of Compressive Sensed Signals in Noisy Environment," *Signal Processing*, vol. 104, 2014, pp. 43-50.

[104] S. Stanković, I. Orović, and L. Stanković, "Polynomial Fourier Domain as a Domain of Signal Sparsity," *Signal Processing*, vol. 130, 2017, pp. 243-253.

[105] S. Stanković, L. Stanković, and I. Orović, "Compressive sensing approach in the Hermite transform domain," *Mathematical Problems in Engineering*, vol. 2015 , Article ID 286590, 9 pages.

[106] S. Stanković, L. Stanković, and I. Orović, "A Relationship between the Robust Statistics Theory and Sparse Compressive Sensed Signals Reconstruction," *IET Signal Processing*, vol. 8, no. 3, 2014, pp. 223-229.

[107] R. Tibshirani, "Regression Shrinkage and Selection via the lasso," *Journal of the Royal Statistical Society* Series B (methodological) vol. 58, no. 1, 1996, pp. 267-88.

[108] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, K. Knight, "Sparsity and Smoothness via the Fused lasso", *Journal of the Royal Statistical Society,* Series B (statistical Methodology), vol. 67, no. 1, 2005, pp. 91-108.

[109] M. Tipping, "Sparse Bayesian Learning and the Relevance Vector Machine," *Journal of Machine Learning Research*, JMLR.org, 1, 2001, pp. 211-244.

[110] J. A. Tropp. "Greed is good: Algorithmic results for sparse approximation," *IEEE Transactions on Information Theory*, vol. 50, no. 10, 2004, pp.2231âĂŞ2242.

[111] J. A. Tropp, A.C. Gilbert, "Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, 2007, pp. 4655-4666.

[112] D. Vukobratovic and A. Pizurica, "Compressed sensing using sparse adaptive measurements," *Proc. of the Symposium on Information Theory in the Benelux (SITB '14)*, Eindhoven, The Netherlands, May 2014.

[113] Y. Wang, J. Xiang, Q. Mo, S. He, "Compressed sparse time-frequency feature representation via compressive sensing and its applications in fault diagnosis," *Measurement*, vol 68, 2015, pp. 70-81.

[114] L. Wang, L. Zhao, G. Bi and C. Wan, "Hierarchical Sparse Signal Recovery by Variational Bayesian Inference," *IEEE Signal Processing Letters*, vol. 21,no. 1, 2014, pp. 110-113.

[115] L. Zhang, M. Xing, C. W. Qiu J. Li, and Z. Bao, "Achieving higher resolution ISAR imaging with limited pulses via compressed sampling," *IEEE Geoscience and Remote Sensing Letters*, vol. 6, no. 3, 2009, pp. 567-571.

[116] T. Zhang, "Sparse Recovery with Orthogonal Matching Pursuit Under RIP," *IEEE Transactions on Information Theory,* vol. 57, no. 9, 2011, pp. 6215-6221.

[117] Z. Zhang, T. P. Jung, S. Makeig, and B. D. Rao, "Compressed Sensing of EEG for Wireless Telemonitoring With Low Energy Consumption and Inexpensive Hardware," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 1, 2013, pp. 221-224.

[118] Z. Zhang, and B.D. Rao, "Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 5, 2011, pp. 912-926.

[119] Z. Zhang, and B.D. Rao, "Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation," *IEEE Transactions on Signal Processing*, vol. 61, no. 8, 2013, pp. 2009-2015.

[120] L. Zhu, E. Liu, and J. H. McClellan. "Sparse-promoting full-waveform inversion based on online orthonormal dictionary learning," *Geophysiscs*, vol. 82, no. 2, 2017, pp. 87-107.

[121] Z. Zhu, K. Wahid, P. Babyn, D. Cooper, I. Pratt, and Y. Carter, "Improved Compressed Sensing-Based Algorithm for Sparse-View CT Image Reconstruction," *Computational and Mathematical Methods in Medicine*, vol. 2013.