THE CONTINUOUS JOINT SPARSITY PRIOR FOR SPARSE REPRESENTATIONS: THEORY AND APPLICATIONS

Moshe Mishali and Yonina C. Eldar

Department of Electrical Engineering
Technion - Israel Institute of Technology, Haifa, Israel
moshiko@tx.technion.ac.il , yonina@ee.technion.ac.il

ABSTRACT

The classical problem discussed in the literature of compressed sensing is recovering a sparse vector from a relatively small number of linear non-adaptive projections. In this paper, we study the recovery of a continuous set of sparse vectors sharing a common set of locations of their non-zero entries. This model includes the classical sparse representation problem, and also its known extensions. We develop a method for joint recovery of the entire set of sparse vectors by the solution of just one finite dimensional problem. The proposed strategy is exact and does not use heuristics or discretization methods. We then apply our method to two applications: The first is spectrum-blind reconstruction of multi-band analog signals from point-wise samples at a sub-Nyquist rate. The second application is to the well studied multiple-measurement-vectors problem which addresses the recovery of a finite set of sparse vectors.

Index Terms— Joint sparsity prior, multiband sampling, multiple-measurement vector (MMV), nonuniform periodic sampling, sparse representation.

1. INTRODUCTION

Digital applications have developed rapidly over the last few decades. This on-going development lead to increased data rates entering the signal processing units resulting in the need for constant speedup in the operating rates of all digital hardware components. Indeed, the core frequency of DSP processors as well as other peripheral devices has been greatly accelerated. However, the essential stage of the analog-to-digital (A2D) conversion, which is part of the front-end of every typical digital application, remains a major bottleneck. This is partially due to technical limitations, but mostly because of a traditional assumption that the sampling stage must acquire the data at the Nyquist rate. When prior knowledge on the signal structure exists, this assumption is often quite restrictive.

The literature of compressed sensing (CS) addresses this problem from a discrete point of view [1],[2]. This theory suggests that a sparse vector $\mathbf{x} \in \mathbb{R}^n$, with no more than $K << n$ non-zero entries, can be recovered from a relatively small number of linear non-adaptive measurements $\mathbf{y} \in \mathbb{R}^m$. The vector $\mathbf{y}$ is produced by a linear transformation $\mathbf{y} = \mathbf{A}\mathbf{x}$. The matrix $\mathbf{A}$ is usually thought of as the sampling operator, although both $\mathbf{x}$ and $\mathbf{y}$ are discrete. Clearly, m must be at least twice the number of non-zero values in $\mathbf{x}$, namely $m \geq 2K$, since the number of unknowns is $2K$: the non-zero values and their locations. Recovery of a sparse solution $\mathbf{x}$ from measurements $\mathbf{y}$ is known to be NP-hard. Instead, CS ideas suggest tractable techniques [3],[4] for the recovery of $\mathbf{x}$ from $\mathbf{y}$, where $m = cK$ measurements are required with a typical value of c that is greater than 2. Nevertheless, the number of measurements satisfies $m << n$ which still represents a substantial saving.

An extension of the basic CS model is when a sparse matrix $\mathbf{X}$, with no more than K non-identical zero rows, is sampled with the linear system $\mathbf{Y} = \mathbf{A}\mathbf{X}$. This model is usually referred to as Multiple-Measurement-Vectors (MMV), while the basic model is named Single-Measurement-Vector (SMV). Clearly, each of the column vectors of $\mathbf{X}$ is sparse. However, in an MMV problem we have an additional assumption that the non-zero values share common locations in all the columns of $\mathbf{X}$. Theoretical results for MMV show that in some cases $\mathbf{X}$ can be determined from $\mathbf{Y}$ using less than $m = 2K$ measurements per column [5]. In addition, practical techniques for MMV tend to recover $\mathbf{X}$ from $m = cK$ measurements, where the value of c is lower than in the SMV case [5],[6].

In this paper, we study a general model that includes both the SMV and MMV formulations. In our model, a continuous set of sparse vectors sharing a common non-zero locations set is sampled by a matrix $\mathbf{A}$, producing a corresponding continuous set of measurement vectors. We show that theoretical results of MMV are easily generalized to this scenario. However, treating the continuous case is inherently more complicated as the dimension of the problem is infinite.

Our main contribution is the development of an exact method to transform the continuous problem into a finite dimensional one using the prior knowledge of joint sparsity. Our strategy does not involve heuristics or discretization methods. Theoretically, this approach allows for the exact recovery of the continuous set of vectors by solving a finite dimensional problem. Moreover, we show that the casted problem is always an MMV, and thus known practical CS techniques can be employed to solve it.

To demonstrate our approach, we present an application to spectrum-blind reconstruction of multi-band analog signals from sub-Nyquist samples. Specifically, we consider the problem of sampling and reconstruction a multi-band signal when the band locations are unknown. To-date, reconstruction methods for this class of signals are non-blind as the spectral structure of the signal is required in the design of the reconstruction stage. We show that the spectrum-blind reconstruction problem can be reformulated and solved using the continuous framework developed in this paper. The sampling rate required to allow perfect reconstruction is shown to be dramatically less than the Nyquist rate. Evidently, this application introduces a solution to the A2D conversion bottleneck in the form of a truly analog compressed sensing technique.

Finally, we discuss the computational advantages of our method for the solution of a redundant MMV, namely a solution matrix $\mathbf{X}$ with non-full column rank. As we show, our approach leads to a new reconstruction strategy for this problem that reduces the dimension

978-1-4244-1714-8/07/$25.00 ©2007 IEEE

of the original MMV and the computational efforts in finding $\mathbf{X}$.

This paper is organized as follows. In Section 2 the problem is formulated and we discuss several relevant priors. Transformation of the continuous problem into a finite dimensional MMV is described in Section 3. Applications to multi-band signals and MMV are discussed in Sections 4 and 5, respectively.

2. CONTINUOUS JOINT SPARSITY MODEL

2.1. Problem formulation

Consider the following parametric linear system:

$$\mathbf{y}(\lambda) = \mathbf{A}\mathbf{x}(\lambda), \quad \lambda \in \Lambda, \tag{1}$$

in which $\mathbf{A}$ is a known $m \times n$ matrix with $m < n$. For every $\lambda \in \Lambda$, the vector $\mathbf{y}(\lambda)$ is given while $\mathbf{x}(\lambda)$ is unknown. The parameter λ belongs to some known set Λ. We denote $\mathbf{x}(\Lambda)$ as the set of vectors $\{\mathbf{x}(\lambda)\}_{\lambda\in\Lambda}$ and assume that there exists a solution set $\bar{\mathbf{x}}(\Lambda)$ for (1). Our goal is to determine $\bar{\mathbf{x}}(\Lambda)$ from the data $\mathbf{y}(\Lambda)$ and the knowledge of $\mathbf{A}$.

For every λ, the linear system of (1) has fewer equations than unknowns. Thus, the solution set $\bar{\mathbf{x}}(\Lambda)$ is not uniquely determined unless some prior is used. In what follows, we introduce different priors that can be utilized in order to determine $\bar{\mathbf{x}}(\Lambda)$ from (1).

2.2. Uniqueness priors

2.2.1. *Null orthogonality prior*

Suppose $\bar{\mathbf{x}}(\lambda)$ belongs to the orthogonal of the null space of $\mathbf{A}$. In this case, multiplying (1) from both sides by the pseudo-inverse $\mathbf{A}^\dagger$ of $\mathbf{A}$ recovers $\bar{\mathbf{x}}(\lambda)$ exactly. If this prior holds for every $\lambda \in \Lambda$ then (1) determines $\bar{\mathbf{x}}(\Lambda)$.

2.2.2. *Vector sparsity prior*

An alternative prior is that $\bar{\mathbf{x}}(\lambda)$ is K-sparse, which means that it contains no more than K non-zero values. This prior is widely assumed in the CS literature. The following definition [7] and theorem provide a sufficient uniqueness condition for this prior:

Definition 1 *The Kruskal-rank of $\mathbf{A}$, denoted as $\sigma(\mathbf{A})$, is the maximal number q such that every set of q columns of $\mathbf{A}$ is linearly independent.*

Theorem 1 *If the vector $\bar{\mathbf{x}}$ is a K-sparse solution of $\mathbf{y} = \mathbf{A}\mathbf{x}$, with $\sigma(\mathbf{A}) \geq 2K$, then it is the unique sparsest solution of the system.*

Theorem 1 and its proof are given in [5],[8] with a slightly different notation of $\mathrm{Spark}(A)$ instead of the Kruskal-rank of $\mathbf{A}$. A direct corollary of Theorem 1 is that if $\bar{\mathbf{x}}(\lambda)$ is K-sparse for every $\lambda \in \Lambda$ and $\sigma(\mathbf{A}) \geq 2K$ then $\bar{\mathbf{x}}(\Lambda)$ is the unique solution set of (1). Note that the null orthogonality prior implies that $\bar{\mathbf{x}}(\lambda)$ belongs to some subspace of $\mathbb{R}^n$. Thus, it implies a strict linear relationship between all the entries of $\bar{\mathbf{x}}(\lambda)$, which typically results in a non-sparse vector. In contrast, the vector sparsity prior imposes a weaker constraint, since only the number of non-zero values in $\bar{\mathbf{x}}(\lambda)$ is restricted. In particular, this prior does not impose any limitation on the exact locations of these non-zero entries nor on their specific values.

2.2.3. *Joint sparsity prior*

Both previous priors are local since there is no relation between the constraints imposed on $\bar{\mathbf{x}}(\lambda)$ for different values of λ. In contrast, the joint sparsity prior is global: a common property is associated with *all* the vectors of the solution set $\bar{\mathbf{x}}(\Lambda)$. Specifically, it is assumed that there exists some set S of size no more than K, such that each $\bar{\mathbf{x}}(\lambda)$ is sparse and that the set of locations corresponding to non-zero values is a subset of S. Note that as in the vector sparsity prior, the exact knowledge of S is not assumed, only its size.

Formally, define the index set

$$I(\bar{\mathbf{x}}(\lambda)) = \{k \,|\, \bar{\mathbf{x}}_k(\lambda) \neq 0\}, \tag{2}$$

that contains the locations of the non-zero values of $\bar{\mathbf{x}}(\lambda)$. The set $I(\bar{\mathbf{x}}(\lambda))$ is finite and depends on λ, since the locations of the non-zero values in $\bar{\mathbf{x}}(\lambda)$ varies with λ. Next, define the union

$$S = I(\bar{\mathbf{x}}(\Lambda)) = \bigcup_{\lambda\in\Lambda} I(\bar{\mathbf{x}}(\lambda)). \tag{3}$$

The joint sparsity prior assumes that $|S| \leq K$, where $|S|$ denotes the size of S. In this case, $\bar{\mathbf{x}}(\Lambda)$ is called a K-sparse solution set. We emphasize that the joint sparsity prior implies vector sparsity but the converse is not true.

Observe that if Λ is a single element set then the joint sparsity prior coincides with the vector sparsity prior. The system (1) then represents the classical CS problem of recovering a single sparse vector from an undetermined system (SMV). If $\Lambda = \{\lambda_i\}$ is a finite set then (1) can be written as

$$\mathbf{Y} = \mathbf{A}\mathbf{X}, \tag{4}$$

where the ith column of $\mathbf{Y}$ is $\mathbf{y}(\lambda_i)$. Similarly, the columns of $\mathbf{X}$ form a solution set $\mathbf{x}(\Lambda)$. In this case, the joint sparsity prior on $\bar{\mathbf{X}}$ implies that the values in $\bar{\mathbf{X}}$ are restricted to some unknown set S of K rows. This is the standard MMV problem.

Clearly, a global prior results in a dependency between $\bar{\mathbf{x}}(\lambda)$ for different values of λ. In particular, the joint sparsity prior is more restrictive than vector sparsity. Thus, it is expected that uniqueness of $\bar{\mathbf{x}}(\Lambda)$ can be proved for S whose size is larger than the threshold of $\sigma(\mathbf{A})/2$. Indeed, an improved uniqueness threshold for MMV is given in [5]. Their arguments can be directly adopted to the case of a continuous set Λ, resulting in the following proposition.

Proposition 1 *Let $\bar{\mathbf{x}}(\Lambda)$ be a K-sparse solution set of (1). If*

$$K \leq \frac{\sigma(\mathbf{A}) + \mathrm{rank}(\mathbf{y}(\Lambda)) - 1}{2} \tag{5}$$

then $\bar{\mathbf{x}}(\Lambda)$ is the unique sparsest solution set of (1).

Here $\mathrm{rank}(\mathbf{y}(\Lambda))$ denotes the maximal dimension of any subspace spanned by a finite collection of vectors from $\mathbf{y}(\Lambda)$.

Proposition 1 has a notable drawback: the prior knowledge of K is not sufficient to set the number of measurements m, since (5) relates the design of $\mathbf{A}$ to the specific data set $\mathbf{y}(\Lambda)$. Therefore, if uniqueness must be assured for every K-sparse solution set, including ones with[1] $\mathrm{rank}(\mathbf{y}(\Lambda)) = 1$, then (5) reduces to the condition $\sigma(\mathbf{A}) \geq 2K$ of Theorem 1.

As pointed out, recovery of the sparsest solution set is known to be NP-hard. Several efficient sub-optimal techniques were proposed

[1]The degenerate case $\mathrm{rank}(\mathbf{y}(\Lambda)) = 0$ implies the sparsest solution set $\bar{\mathbf{x}}(\Lambda) = 0$.

in the CS literature for both SMV and MMV. These approaches involve matrix operations (e.g. inverse, projection, etc) that cannot be directly extended to the continuous case. A straightforward strategy would be to solve (1) on a finite dense grid $\{\lambda_i \in \Lambda\}$, and then construct the solution set $\bar{\mathbf{x}}(\Lambda)$ by some interpolation technique. However, this method cannot guarantee an exact recovery of $\bar{\mathbf{x}}(\Lambda)$. Moreover, the accuracy of such a solution depends on the density of the grid, which directly impacts the computational complexity.

In the next section, we present an exact method for the continuous case, in which the continuous joint sparsity prior is exploited in order to formulate a finite dimensional problem. This enables exact recovery of $\bar{\mathbf{x}}(\Lambda)$ since no heuristics or discretization techniques are used.

3. CONTINUOUS TO FINITE TRANSFORMATION

We begin by defining the $m \times m$ matrix

$$\mathbf{Q} = \int_{\lambda \in \Lambda} \mathbf{y}(\lambda)\mathbf{y}^H(\lambda) d\lambda, \tag{6}$$

where $\mathbf{y}^H(\lambda)$ denotes the conjugate transpose of $\mathbf{y}(\lambda)$. We assume that the integrand in (6) is integrable over the continuous set Λ. Since $\mathbf{Q}$ is positive semi-definite, it can be decomposed as

$$\mathbf{Q} = \mathbf{V}\mathbf{V}^H \tag{7}$$

with $\mathbf{V}$ having r orthogonal columns, where $r = \mathrm{rank}(\mathbf{Q})$. Next, define the finite dimensional linear system

$$\mathbf{V} = \mathbf{A}\mathbf{U}. \tag{8}$$

The following theorem relates the sparsest solution of (8) with $\bar{\mathbf{x}}(\Lambda)$.

Theorem 2 *Suppose $\bar{\mathbf{x}}(\Lambda)$ is a K-sparse solution set of (1). If $\sigma(\mathbf{A}) \geq 2K$ then the finite linear system of (8) represents an MMV with a unique sparsest solution matrix $\bar{\mathbf{U}}$ having $I(\bar{\mathbf{U}}) = I(\bar{\mathbf{x}}(\Lambda))$.*

The advantage of Theorem 2 is the ability to recover a continuous solution set $\bar{\mathbf{x}}(\Lambda)$ by solving only one finite dimensional problem. Essentially, the set $S = I(\bar{\mathbf{U}})$ is the key information needed for exact recovery of $\bar{\mathbf{x}}(\Lambda)$. To see this, observe that (1) can be written as

$$\mathbf{y}(\lambda) = \mathbf{A}_S\, \mathbf{x}_S(\lambda), \quad \lambda \in \Lambda, \tag{9}$$

where $\mathbf{A}_S$ is the sub-matrix of $\mathbf{A}$ containing only the columns whose indices belong to S. Similarly, $\mathbf{x}_S(\lambda)$ is a vector of length $|S|$ that consists of the entries of $\mathbf{x}(\lambda)$ in the locations described by S. It can be proved that under the conditions of Theorem 2, the pseudo-inverse of $\mathbf{A}_S$ satisfies $(\mathbf{A}_S)^\dagger \mathbf{A}_S = I$, which results in

$$\mathbf{x}_S(\lambda) = (\mathbf{A}_S)^\dagger \mathbf{y}(\lambda), \quad \forall \lambda \in \Lambda. \tag{10}$$

In addition, from (3)

$$\mathbf{x}_i(\lambda) = 0, \quad \forall \lambda \in \Lambda,\ i \notin S. \tag{11}$$

Thus, once the sparsest solution matrix $\bar{\mathbf{U}}$ is found, the solution set $\bar{\mathbf{x}}(\Lambda)$ is obtained by (10)-(11). Note that Theorem 2 is valid for any selection of $\mathbf{V}$ in (7) as the decomposition is not unique.

The result of Theorem 2 is achieved due to the global nature of the joint sparsity prior. In the next sections, we demonstrate the usage of Theorem 2 in two applications. One corresponds to a continuous set Λ, while the other has finite Λ. Obviously, the flow of (6)-(8) is also valid for the latter application, in which (1) is already an MMV and there are known techniques to solve it. Nevertheless, we show that our method has a potential for computational savings.

4. APPLICATION: MULTI-BAND SIGNALS

We first study an application of Theorem 2 to spectrum-blind reconstruction of analog multi-band signals. A detailed account of these results can be found in [12].

4.1. Formulation

Let $x(t)$ be an analog bandlimited signal, so that its Fourier transform $X(f)$ is supported on $\mathcal{F} = [0, 1/T]$. Evidently, $1/T$ is the Nyquist rate for $x(t)$. A multi-band signal has an additional property that $X(f)$ is also restricted to several disjoint intervals (called bands) in $\mathcal{F}$, which means $X(f) = 0$ outside the bands.

Consider the set $\mathcal{M}$ of multi-band signals consisting of no more than N bands, where each of the band widths is not greater than B. We wish to construct a system that enables perfect reconstruction of $x(t) \in \mathcal{M}$ from its samples at a low sampling rate. In addition, both sampling and reconstruction should be blind, so that neither can use information about the band locations.

In [9],[10] the authors describe a half-blind system: A blind multi-coset strategy is used for sampling, while a non-blind scheme is proposed for the reconstruction. Here we use the blind sampling strategy of [9] and develop a spectrum-blind reconstruction scheme using the joint sparsity prior discussed earlier. We also derive an expression for the sampling rate that is sufficient for perfect reconstruction. This sampling rate is typically far less than the Nyquist rate.

4.2. Spectrum-blind sampling

Multi-coset sampling is a periodic non-uniform sampling on the Nyquist grid $x(t = nT)$. The grid is divided into blocks of L samples, in which p are kept and the rest are ignored. A sampling pattern $C = \{c_i\}_{i=1}^p$ describes the locations of the p samples that are kept in each block. This sampling strategy results in p uniform sequences

$$x_{c_i}[n] = \begin{cases} x(t = nT) & n = mL + c_i, \text{ for some } m \in \mathbb{Z} \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

Direct calculations show that [9]

$$X_{c_i}(e^{j2\pi fT}) = \frac{1}{LT} \sum_{r=0}^{L-1} \exp\left(j\frac{2\pi}{L} c_i r\right) X\left(f + \frac{r}{LT}\right), \quad \forall\, f \in \mathcal{F}_0 = \left[0, \frac{1}{LT}\right),\ 1 \leq i \leq p, \tag{13}$$

where $X_{c_i}(e^{j2\pi fT})$ is the discrete-time Fourier transform of $x_{c_i}[n]$. To conform with the formulation of (1) we express (13) in matrix form

$$\mathbf{y}(f) = \mathbf{A}\mathbf{x}(f), \quad \forall f \in \mathcal{F}_0, \tag{14}$$

where $X_{c_i}(e^{j2\pi fT})$ is the ith entry of the vector $\mathbf{y}(f)$, and $\mathbf{x}(f)$ contains L unknowns for each f

$$\mathbf{x}_i(f) = X\left(f + \frac{i}{LT}\right), \quad 0 \leq i \leq L-1, \quad f \in \mathcal{F}_0. \tag{15}$$

The matrix $\mathbf{A}$ depends on the parameters L, p and the set C but not on $x(t)$ and is defined by

$$\mathbf{A}_{ik} = \frac{1}{LT} \exp\left(j\frac{2\pi}{L} c_i k\right). \tag{16}$$

In the next section, we present a spectrum-blind reconstruction scheme for $\mathbf{x}(\mathcal{F}_0)$ from the data $\mathbf{y}(f)$, which is equivalent to recovery of $x(t)$ from the sampling sequences $x_{c_i}[n]$.

4.3. Spectrum-Blind Reconstruction

The following theorem shows that a specific parameter selection of L, p, C implies the continuous joint sparsity prior. In turn, this allows for the use of Theorem 2, which enables perfect reconstruction of the signal using only one finite dimensional MMV. This selection of L, p, C is also blind as it does not require knowledge of the band locations [12].

Theorem 3 *Let $x(t) \in \mathcal{M}$ be a multi-band signal. If:*

1. $L \leq 1/(BT)$,
2. $p \geq 4N$,
3. C *is a universal pattern, which means* $\mathbf{A}$ *is of full Kruskal-rank,*

then, the solution set $\mathbf{x}(\mathcal{F}_0)$ of (14) is $2N$-sparse and can be obtained by finding the sparsest solution matrix $\bar{\mathbf{U}}$ of (8).

Theorem 3 relies on the use of a universal pattern as defined in [9]. In [11] we show that if L is prime then any pattern is universal. Using Theorem 3 we can recover $\mathbf{x}(\mathcal{F}_0)$ by first creating the matrix $\mathbf{Q} = \int_{\mathcal{F}_0} \mathbf{y}(f)\mathbf{y}^H(f)df$, and then solving the MMV (8) with $\mathbf{A}$ given by (16). This approach enjoys all the properties of Theorem 2. In particular, an exact recovery of the signal is obtained from (10)-(11) without heuristics or discretization techniques. However, for practical implementation it is suggested to carry out the last step of finding the sparsest solution $\bar{\mathbf{U}}$ by any of the efficient techniques described in the CS literature. Empirically, these methods show a high correct recovery rate of $\bar{\mathbf{U}}$ when tested on random data.

A Multi-coset stage has an average sampling rate of $p/(LT)$ as can be seen from (12). Using the limitations that Theorem 3 imposes on the values of L, p, we get that an average sampling rate of at least $4NB$ is required for perfect reconstruction. As NB is the portion of the bandwidth occupied by the multi-band signal, the sampling rate dictated by Theorem 3 can be far less than the Nyquist rate.

5. APPLICATION: MMV

Consider the MMV of (4) with the sparsest solution $\bar{\mathbf{X}}$. Applying (6)-(8) on (4) results in another MMV system, whose sparsest solution is $\bar{\mathbf{U}}$. The following proposition shows that this strategy can be advantageous when $\bar{\mathbf{X}}$ has linearly dependent columns.

Proposition 2 *If $\sigma(\mathbf{A}) \geq 2K$ and $\bar{\mathbf{X}}$ is the unique K-sparse solution matrix of (4), then:*

1. $r = \mathrm{rank}(\mathbf{Y}) = \mathrm{rank}(\bar{\mathbf{X}})$.
2. *Each of the matrices $\mathbf{V}, \bar{\mathbf{U}}$ has exactly r columns.*

If $\mathbf{Y}$ contains more than r columns, then Proposition 2 enables to reduce the number of unknowns by moving to another MMV problem with less columns. The advantage of the joint sparsity prior is used here to reduce the computational complexity of solving a given redundant MMV system. It can be easily seen that the dimensions of the MMV of (8) are the minimal possible. We point out that an alternative approach is to identify a subset of r linearly independent columns of $\mathbf{Y}$ and solve the MMV for this subset only. Then, we would have to deduce the set S from the sparsest solution of this subset, and construct $\bar{\mathbf{X}}$ using (10)-(11). However, this approach works only for the MMV scenario, since finding a subset of r independent vectors of $\mathbf{y}(\Lambda)$ is impractical for a continuous set Λ. Thus, the approach described in Section 3 is more general.

6. CONCLUSIONS

In this paper, we presented the joint sparsity prior for parametric undetermined linear systems. It was shown that this formulation covers known problems in the CS literature. However, a special treatment is needed when the linear system depends on a continuous parameter, since the dimension of the solution is infinite and applying discretization cannot allow for an exact recovery. We introduced a method to transform the continuous problem into a finite dimensional one, and demonstrated the use of this approach in two applications. Evidently, the essential contribution of the proposed method is for the continuous scenario, while for the finite case it was shown that some computational efforts can be saved.

7. REFERENCES

[1] E. J. Candès, J. Romberg and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory,* vol. 52, pp. 489-509.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory,* vol. 52, Iss.4, pp. 1289–1306, April 2006.

[3] S. S. Chen, D. L. Donoho and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Scientific Computing,* vol. 20, no. 1, pp. 33-61, 1999.

[4] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.,* vol. 41, no. 12, pp. 3397-3415, Dec. 1993.

[5] J. Chen, and X. Huo, "Theoretical results on sparse representations of multiple measurement vectors," *IEEE Trans. Signal Process,* vol. 54, pp. 4634-4643, Dec. 2006.

[6] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process,* vol. 53, pp. 2477-2488, Jul. 2005.

[7] J. B. Kruskal, "Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear Alg. Its Applic.,* vol. 18, no.2, pp. 95–138, 1977.

[8] D. L. Donoho and M. Elad, "Maximal sparsity representation via ℓ_1 minimization," *Proc. Natl. Acad. Sci.,* vol. 100, pp. 2197-2202, Mar. 2003.

[9] R. Venkataramani and Y. Bresler, "Perfect reconstruction formulas and bounds on aliasing error in sub-Nyquist nonuniform sampling of multiband signals," *IEEE Trans. Inform. Theory,* vol. 46, pp. 2173-2183, Sept. 2000.

[10] R. Venkataramani and Y. Bresler, "Optimal sub-nyquist nonuniform sampling and reconstruction of multiband signals," *IEEE Trans. Signal Process.,* vol. 49, no. 10, pp. 2301-2313, Oct. 2001.

[11] M. Mishali, Y. C. Eldar and M. Horowitz, "Properties of universal sampling patterns for multi-band signals," *IEEE Letters. on Signal Process., submitted for publication,* Mar. 2007.

[12] M. Mishali and Y. C. Eldar, "Blind multi-band signal reconstruction: Compressed sensing for analog signals," *IEEE Trans. on Signal Process.,* submitted for publication, Sep. 2007.