REVISTA DE LA
UNIÓN MATEMÁTICA ARGENTINA
Volumen 50, Número 2, 2009, Páginas 217–225

OPTIMAL NON-LINEAR MODELS

AKRAM ALDROUBI, CARLOS CABRELLI, AND URSULA MOLTER

Abstract. This paper is a survey about recent results on sparse representations and optimal models in different settings. Given a set of functions, we show that there exists an optimal collection of subspaces minimizing the sum of the square of the distances between each function and its closest subspace in the collection. Further, this collection of subspaces gives the best sparse representation for the given data, in a sense defined later, and provides an optimal model for sampling in a union of subspaces.

1. Introduction

A new paradigm for signal sampling and reconstruction recently developed by Lu and Do [19] (see also [9]) starts from the point of view that signals live in some union of subspaces $\mathcal{M} = \cup_{i\in I} V_i$, instead of a single vector space $\mathcal{M} = V$ such as the space of band-limited functions also known as the Paley-Wiener space. This new paradigm is general and includes (when $\mathcal{M} = V$) the classical Shannon sampling theory and its extensions [6], as well as sampling of signals with finite rate of innovation (see e.g., [20, 16]). In the new framework, when we have more than one subspace, the signal space model $\mathcal{M} = \cup_{i\in I} V_i$ is non-linear and the techniques for reconstructing a signal $f \in \cup_{i\in I} V_i$ from its samples $\{f(x_j)\}_j$ are involved and the reconstruction operators are non-linear.

Since for each class of signals the starting point of this new theory is the knowledge of the signal space $\mathcal{M} = \cup_{i\in I} V_i$, the first step for implementing the theory is to find an appropriate signal model $\mathcal{M} = \cup_{i\in I} V_i$ from a set of observed data $\mathcal{F} = \{f_1, \dots, f_m\}$. For the classical sampling theory, the problem of finding the shift-invariant space model $\mathcal{M} = V$ from a set of observed data has been studied and solved in [4],[3]. For the new sampling paradigm, the problem consists in proving the existence and finding subspaces $V_1, \cdots, V_l$, that minimize the expression

$$e(\mathcal{F}, \{V_1, \dots, V_l\}) = \sum_{i=1}^{m} \min_{1\le j\le l} d^2(f_i, V_j), \tag{1}$$

2000 *Mathematics Subject Classification.* Primary 41A65, 42C15 ; Secondary 68P30, 94A20.
Key words and phrases. Sampling, Sparsity, Compressed Sensing, Frames.
The research of Akram Aldroubi is supported in part by NSF Grant DMS-0504788. The research of Carlos Cabrelli and Ursula Molter is partially supported by Grants: PICT2006-00177, PIP112-200801-00398, UBACyT X058 and X108.

over all choices of l subspaces, from an appropriate class $\mathcal{C}$ in some Hilbert space $\mathcal{H}$. Here $\mathcal{F} = \{f_1, \dots, f_m\} \subset \mathcal{H}$ is a set of observed data and d is the distance function in $\mathcal{H}$.

It is well known that the problem of sampling and reconstruction of signals with finite rate of renovation is closely related to the developing theory of compressed sensing (see e.g., [12, 11, 13, 14, 15, 21] and the references therein). Compressed sensing proposes to find a vector $x \in \mathbb{R}^N$ from the knowledge of the values, when applied to x, of a relatively small set of functionals $\{\psi_k : k = 1, \dots, p\}$ (where $p << N$). Obviously, the problem of finding x from the set $\{y_k = \langle x, \psi_k \rangle : k = 1, \dots, p\}$ is ill-posed. However, it becomes meaningful if x is assumed to be sufficiently sparse.

A typical assumption of sparsity is that x has at most n non-zero components ($\|x\|_0 \leq n$), where $n \leq 2p << N$. As a consequence of this assumption of sparsity, the vector x belongs to some union of subspaces, each of which is generated by exactly n vectors from the canonical basis of $\mathbb{R}^N$. In matrix formulation this problem can be stated as follows: find $x \in \mathbb{R}^N$ with $\|x\|_0 \leq n$ from the matrix equation $y = Ax$ where A is a $p \times N$ matrix and y is a given vector in $\mathbb{R}^p$.

A related problem consists in finding an approximation to the vector y using a sparse vector x. Formally, this problem can be stated as follows: find $\min_x \|x\|_0$ subject to the constraint $\|Ax - y\|_2 \leq \varepsilon$ for some given ε. The above two problems, their analysis, extensions, and efficient algorithms for finding their solutions can be found in [1, 2, 10, 12, 11, 13, 14, 15, 18, 22] and the references therein.

If in the above problems the matrix A is also an unknown to be found together with the set of unknown vectors $\{x_i : i = 1, \dots, m\} \subset \mathbb{R}^N$, then these problems become the problems of finding a *dictionary* A from the data $\{y_i : i = 1, \dots, m\} \subset \mathbb{R}^p$ obtained by sampling the sparse vectors $\{x_i : i = 1, \dots, m\} \subset \mathbb{R}^N$ see e.g., [2, 1, 18]. In this context, the columns of A are called *atoms* of A. Under appropriate assumptions on the data and dictionary, the problem has a unique solution up to a permutation of the columns of A [2, 1]. Finding the solution to this problem by exhaustive methods is computationally intractable, but the K-SVD algorithm described in [1] provides a computationally effective search algorithm.

The problem of finding the signal model for signals with finite rate of innovation consists of finding a set $\mathcal{M} = \cup_{i \in I} V_i$, formed by subspaces V_i that are infinite dimensional, in general, but usually structured, e.g., each V_i is a shift-invariant space. However, the signal modeling problem as described by (1) is closely related to the dictionary problem for sparse data, described in the previous paragraph.

To see this relation, let us formulate the dictionary problem as follows: given a class of signals, determine if there exists a dictionary of small size, such that each of the signals can be represented with minimal sparsity.

More precisely, assume that we have a class of m signals, where m is a very large number. We want to know whether there exists a dictionary, such that every signal in the class is a linear combination of at most n atoms in the dictionary. Clearly, to make the problem meaningful and realistic the length of the dictionary should be small compared with m.

It follows, that if for a given set of data such a dictionary exists, then the data can be partitioned into subsets each of which belongs to a subspace of dimension at most n (i.e. to the subspace generated by the atoms that the signal uses in its representation). That is, each subset of the partition can be associated to a low dimensional subspace.

Conversely, if our class of signals can be partitioned into l subsets, such that the signals in each subset belong to a subspace of dimension no bigger than n, then by choosing a set of generators from each of the subspaces, we can construct a dictionary of length at most ln with the property that each of the signals can be represented using at most n atoms in the dictionary.

This suggests that the problem of finding a dictionary where the signals have sparse representation can be solved by finding a small collection of low dimensional subspaces containing our signals, and viceversa.

So, we will say that the class of signals is (l, n)-*sparse* if there exist l subspaces of dimension at most n, such that the signals in our class belong to the union of these l subspaces. From the above discussion, it is clear that if our data is (l, n)-sparse then there exists a dictionary of length at most ln.

Now assume that for a given l and n our data is not (l, n)-sparse. Then we can still try to determine if there exists a collection of optimal subspaces providing the needed sparsity. More precisely, if $\varepsilon > 0$ is given, we want to determine if there exists a collection of l subspaces of dimension at most n, such that the *total error*, that is the sum of the squares of the distance of each signal to the union of the subspaces, is not larger than ε, (see formula (1)). In that case we will say that our data is (l, n, ε)-*sparse*.

As before it is clear that if our data is (l, n, ε)-sparse, then a dictionary of length at most ln exists such that every signal in our class can be approximated using a linear combination of at most n atoms from the dictionary, with total error not larger than ε.

Note that this definition of sparsity is an intrinsic property of the data and the space where they belong to, and does not depend on any fixed dictionary.

A relevant and important question is then, given a class of signals and a small number n, which is the minimun possible ε such the data is (l, n, ε)-sparse?

In this paper we present a survey about a general scheme, developed in detail in [5], that allows to solve the problem described in (1), thereby finding the signal model for the new signal sampling paradigm described in [19], finding a new method for solving the segmentation subspace problem that is optimal in the presence of noise [23], and solving the (l, n, ε)-sparsity problem (in the sense defined above) for a given set of data, in different contexts. Specifically, given a set $\mathcal{F}$ of m functions and numbers l, n such that $n, l < m$, it can beshown that there exist no more than l subspaces of dimension no bigger than n that provide the minimum ε such that the functions in $\mathcal{F}$ are (l, n, ε)-sparse. When the minimum ε is zero, the data is (l, n)-sparse. We also present an algorithm to find the solution subspaces, which has been successfully applied in [8]. Note that there may be several choices of

subspaces that produce the same error. We can identify all those solutions which are called *optimal solutions*.

It is important to remark here that an optimal solution can have less than l subspaces, and the dimensions of the subspaces can be less than n. Since the minimization we consider is over unions of no more than l subspaces, where the dimension of the subspaces is no bigger than n, some of the optimal solutions for a given (l, n) (that is, some of the solutions that give the smallest ε) will yield the minimum $l_0 \leq l$ such that the data is (l_0, n, ε)-sparse, that is l is set to be just an upper bound for the number of allowable subspaces. Furthermore, the number n constraining the dimension of the subspaces is also only an upper bound, that is, an optimal solution can have subspaces of dimension strictly less than n.

Since it is possible to have several optimal solutions (all giving exactly the same ε), some of these solutions may have more subspaces of smaller dimensions and others may have less subspaces of bigger dimensions (always smaller or equal than n). A good variable to look at might be the sum of the dimensions of the subspaces involved in the unions. Since with our methods we are in fact able to find all the optimal solutions, we can choose the most convenient one for each particular application we have in mind.

2. Setting of the Problem

Let $\mathcal{H}$ be a Hilbert space. For $x, y \in \mathcal{H}$ let us denote by $d(x, y) = \|x - y\|_{\mathcal{H}}$. Given a finite subset $\mathcal{F} \subset \mathcal{H}$ and a closed subspace C of $\mathcal{H}$, we denote by $E(\mathcal{F}, C)$ the total distance of the data set $\mathcal{F}$ to the subspace C, i.e.

$$E(\mathcal{F}, C) = \sum_{f \in \mathcal{F}} d^2(f, C). \tag{2}$$

We set $E(\mathcal{F}, C) = 0$ for $\mathcal{F} = \emptyset$ and any subspace C of $\mathcal{H}$.

Let $\mathcal{C}$ be a family of closed subspaces of $\mathcal{H}$ containing the zero subspace. We will say that $\mathcal{C}$ has the Minimal Approximation Property (MAP) if for any finite set $\mathcal{F}$ of vectors in $\mathcal{H}$ there exist a subspace $C_0 \in \mathcal{C}$ that minimizes $E(\mathcal{F}, C)$ over all the subspaces $C \in \mathcal{C}$. That is,

$$E(\mathcal{F}, C_0) = \min_{C \in \mathcal{C}} E(\mathcal{F}, C) \leq E(\mathcal{F}, C), \qquad \forall\, C \in \mathcal{C}. \tag{3}$$

Any subspace $C_0 \in \mathcal{C}$ satisfying (3) will be called an *optimal subspace for* $\mathcal{F}$. Note that if $\mathcal{F} = \emptyset$ then every subspace in $\mathcal{C}$ is optimal. We will choose the zero subspace in that case. For the rest of this section we will assume that the class $\mathcal{C}$ has the Minimal Approximation Property.

Next, since we are interested in models that are union of subspaces, we will arrange the subspaces in finite bundles that will be our main objects, and define the distance (error) between a bundle and a set of vectors.

To do this, let us fix $m, l \in \mathbb{N}$ with $1 \leq l \leq m$ and let $\mathcal{F} = \{f_1, ..., f_m\}$ be a finite set of vectors in $\mathcal{H}$.

Define $\mathcal{V}$ to be the set of sequences of elements in $\mathcal{C}$ of length l, i.e.

$$\mathcal{V} = \mathcal{V}(l) = \big\{\{V_1, \ldots, V_l\} : V_i \in \mathcal{C}, 1 \leq i \leq l\big\}.$$

We will call these finite sequences *bundles*. For $\mathbf{V} \in \mathcal{V}$ with $\mathbf{V} = \{V_1, ..., V_l\}$, we define,

$$e(\mathcal{F}, \mathbf{V}) = \sum_{f \in \mathcal{F}} \min_{1 \le j \le l} d^2(f, V_j). \tag{4}$$

Note that $e(\mathcal{F}, \mathbf{V})$ is a non-linear function of $\mathcal{F}$. Hence, for the problems described in the introduction, what we want is to minimize e over all possible bundles of subspaces. To show that all these problems have indeed a (constructive) solution, we need some definitions (for details we refer the reader to [5]).

Let us denote by $\Pi = \Pi_l$ the set of all l-sequences $P = \{\mathcal{F}_1, \cdots, \mathcal{F}_l\}$ of subsets of $\mathcal{F}$ satisfying the property that for all $1 \le i, j \le l$,

$$\mathcal{F}_i \subset \mathcal{F}, \quad \mathcal{F} = \cup_{s=1}^{l} \mathcal{F}_s, \quad \text{and} \quad \mathcal{F}_i \cap \mathcal{F}_j = \emptyset \text{ for } i \neq j.$$

Note that we allow some of the elements of $P \in \Pi$ to be the empty set. By abuse of language we will still call the elements of Π_l *partitions* (of $\mathcal{F}$).

For $P \in \Pi_l$, $P = \{\mathcal{F}_1, ..., \mathcal{F}_l\}$ and $\mathbf{V} \in \mathcal{V}, \mathbf{V} = \{V_1, ..., V_l\}$ we define,

$$\Gamma(P, \mathbf{V}) = \sum_{i=1}^{l} E(\mathcal{F}_i, V_i). \tag{5}$$

So Γ measures the error between a fixed partition P and a fixed bundle $\mathbf{V}$.

Note that when trying to compute $e(\mathcal{F}, \mathbf{V})$, for each $f \in \mathcal{F}$ we first have to find the subspace $V_{j(f)}$ in $\mathbf{V}$ (see remark after the definition of $e(\mathcal{F}, \mathbf{V})$) that is closest to f and then compute $d^2(f, V_{j(f)})$. While for Γ, since a partition is given and we just compute the distance of each function to its corresponding space (not the closest one necessarily). The surprising fact is that e and Γ can indeed be compared. (see Lemma 1, [5]).

This comparison actually allows us to prove our main theorem.

Theorem 1. *Let $\mathcal{H}$ be a Hilbert space, m, l positive integers with $l \le m$ and $\mathcal{F} = \{f_1, ..., f_m\}$ a set of vectors in $\mathcal{H}$. Then there exists a bundle $\mathbf{V}_0 \in \mathcal{V}$ such that*

$$e(\mathcal{F}, \mathbf{V}_0) = \inf\{e(\mathcal{F}, \mathbf{V}) : \mathbf{V} \in \mathcal{V}\}.$$

Remark. If $0 < l_1 < l_2$, then for any $\mathbf{V} \in \mathcal{V}(l_1)$, $\mathbf{V} = \{V_1, \ldots, V_{l_1}\}$ the bundle $\mathbf{V}' = \{V_1, \ldots, V_{l_1}, \{0\}, \ldots, \{0\}\}$ belongs to $\mathcal{V}(l_2)$ and therefore the error decreases (or no increases) when l (the number of subspaces) increases. Note that in case that the number of subspaces equals the number of data, the error is zero, since we can pick for each data signal the subspace spanned by itself.

It is important to remark here that optimal bundles can have the zero subspace as some of its components. So, if l_0 is the number of subspaces that have dimension greater than zero, in some optimal bundle $\mathbf{V}_0$, then the bundle with l_0 components obtained after the $l - l_0$ zero components are removed from $\mathbf{V}_0$, is also an optimal bundle for the Problem when $\mathcal{V}(l)$ is replaced by $\mathcal{V}(l_0)$. Thus as mentioned in the introduction, the number l is simply a set to be an a priori upper bound on the number of subspaces, and the optimal solution(s) can have any number of subspaces $l_0 \le l$.

As a consequence, we show the following important particular cases:

2.1. **Best Approximation by Bundles of SIS (Shift Invariant Spaces).** For the purpose of this paper, a shift-invariant space will be a subspace of $L^2(\mathbb{R}^d)$ of the form:

$$S(\Phi) := \text{closure}_{L_2} \text{ span}\{\varphi_i(x-k) : \ i = 1, \dots, n, \ k \in \mathbb{Z}^d\} \tag{6}$$

where $\Phi = \{\varphi_1, ..., \varphi_n\}$ is a set of functions in $L^2(\mathbb{R}^d)$. These spaces are often used as standard signal and image models. For example, if $n = 1$, $d = 1$ and $\phi(x) = \text{sinc}(x)$, then the underlying space is the space of band-limited functions (often used in communications). We have the following theorem.

Theorem 2. *Let $\mathcal{F} = \{f_1, ..., f_m\}$ vectors in $L^2(\mathbb{R}^d)$, then there exist a bundle $\mathbf{S}_0 = \{S_1^0, ..., S_l^0\} \in \mathcal{S}$ such that*

$$e(\mathcal{F}, \mathbf{S}_0) = \sum_{i=1}^{m} \min_{1 \le j \le l} d_2^2(f_i, S_j^0) \le \sum_{i=1}^{m} \min_{1 \le j \le l} d_2^2(f_i, S_j) \tag{7}$$

over all bundles $\mathbf{S} = \{S_1, \dots, S_l\} \in \mathcal{S}$.

2.2. **Best Non-Linear Approximation by Bundles of Subspaces in $\mathbb{R}^N$.** Let $\mathcal{F} = \{f_1, \dots, f_m\}$ be a set of vectors in $\mathbb{R}^N$ and $n \le m$, and let us denote by $\mathfrak{L}_n$ the set of all subspaces of dimension smaller (or equal) than n.

Define $\mathfrak{B} = \mathfrak{B}(l)$ to be the set of non-empty bundles of length l in $\mathfrak{L}_n$. We then have:

Theorem 3. *Let $\mathcal{F} = \{f_1, \dots, f_m\}$ be vectors in $\mathbb{R}^N$, and let l and n be given ($l < m$, $n < N$), then there exist a bundle $\mathbf{V}_0 = \{V_1^0, \dots, V_l^0\} \in \mathfrak{B}$, such that*

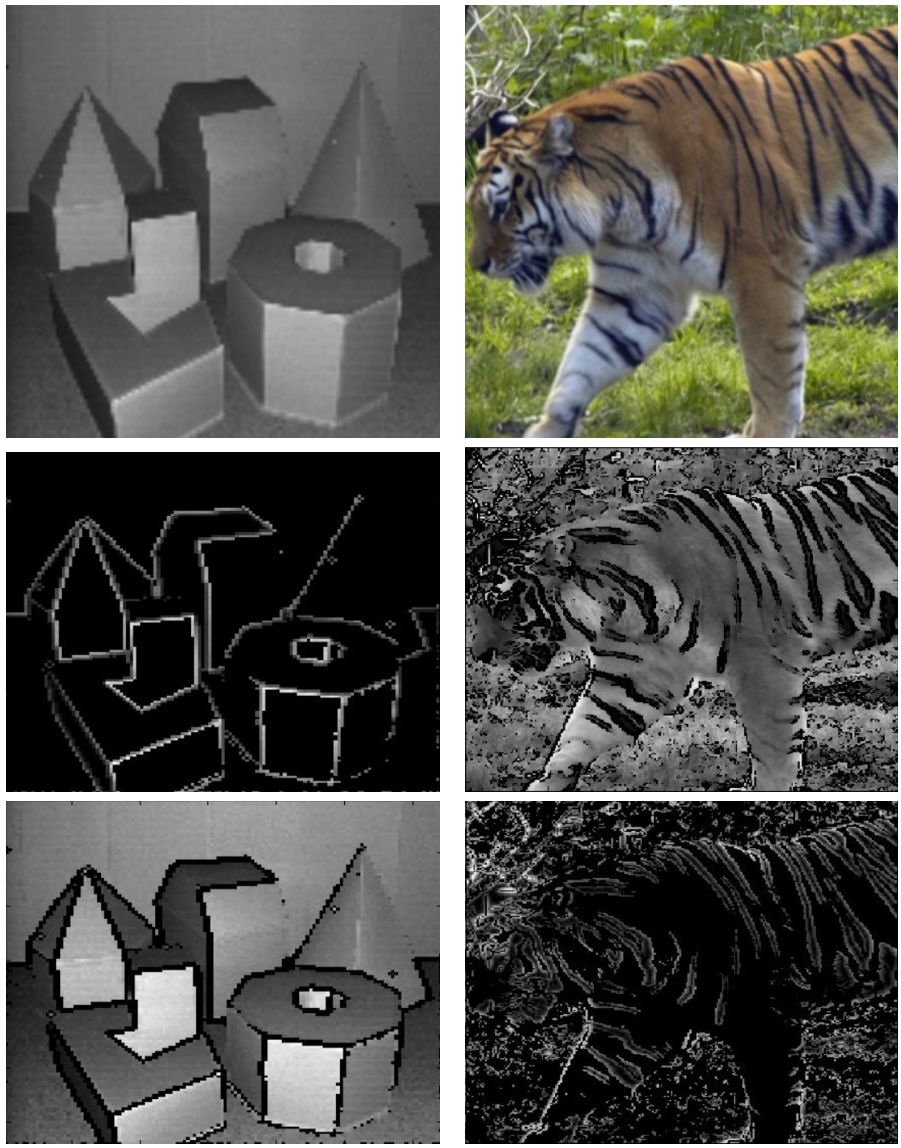
$$e(\mathcal{F}, \mathbf{V}_0) = \sum_{i=1}^{n} \min_{1 \le j \le l} d_2^2(f_i, V_j^0) = \inf\{e(\mathcal{F}, \mathbf{V}) : \mathbf{V} \in \mathfrak{B}\}.$$

Let $P_0 = \{\mathcal{F}_1^0, ..., \mathcal{F}_l^0\}$ be the best partition of $\mathcal{F}$ associated to the optimal bundle $\mathbf{V}_0 = \{V_1^0, ..., V_l^0\}$, using the theorem of Eckardt Young [17] for each $h = 1, ..., l$ and such that $\mathcal{F}_h^0 \neq \emptyset$, a set of generators forming an orthonormal base can be obtained for the optimal space V_h^0 in terms of the singular values and singular vectors of the matrix A_h associated to the subset $\mathcal{F}_h^0$. Further, $e(\mathcal{F}, \mathbf{V}_0)$, is given in terms of the singular values of A_h.

2.3. **The MAP property.** The main ingredient in both of the important special cases presented above, was that the space had the MAP property. Recall that a family $\mathcal{C}$ of closed subspaces of $\mathcal{H}$ containing the zero subspace has the MAP if for any finite set $\mathcal{F}$ of vectors in $\mathcal{H}$ there exist a subspace $C_0 \in \mathcal{C}$ that minimizes $E(\mathcal{F}, C)$ over all the subspaces $C \in \mathcal{C}$. That is,

$$E(\mathcal{F}, C_0) = \min_{C \in \mathcal{C}} E(\mathcal{F}, C) \le E(\mathcal{F}, C), \qquad \forall \ C \in \mathcal{C}. \tag{8}$$

It is therefore relevant to characterize when such a family has this property. In [7] the authors were able to find sufficient conditions on the subspaces in order to have the MAP property.



Top: Original Image. Middle ($\mathcal{F}_L$): Edges. Bottom ($\mathcal{F}_R$): Remaining points.

Precisely, let us parametrize $\mathcal{C}$ by the set of orthogonal projections $\Pi(\mathcal{H}) \subset \mathcal{B}(\mathcal{H})$ and let $\mathcal{Q} = \{I - P_C : C \in \mathcal{C}\} \subset \Pi(\mathcal{H})$. We form new classes $\mathcal{C}^+$ and $\mathcal{Q}^+$ by adding all positive operators to $\mathcal{C}$, $\mathcal{Q}$ respectively. Then, it is not difficult to show that ([7])

Theorem 4. *Assume that $\mathcal{C}^+$ is closed in the weak star operator topology. Then $\mathcal{C}$ satisfies the MAP.*

In fact, we have necessary and sufficient for the MAP in both the finite and infinite dimensional cases. For brevity, we give the theorem for the finite dimensional case. The infinite dimensional case is more involved and use geometric concepts such as as the tangent hull and the convex hull of $\mathcal{Q}^+$. For the finite dimensional case we have ([7]):

Theorem 5. *Suppose $\mathcal{H}$ has finite dimension. Then $\mathcal{Q}$ satisfies MAP if and only if $\overline{\mathcal{Q}} \subset \mathcal{Q}^+$.*

2.4. **Application to Signal Segmentation.** These results can be used to successfully segment images, as the following two examples show: ([8]).

Example. Segmentation of objects using $\mathcal{F} = \{(\nabla I_{ij}, I_{ij})\}$ where I_{ij} is the image intensity at position ij and ∇ is the discrete gradient at that position. The algorithm partitions the set $\mathcal{F}$ into two sets, $\mathcal{F}_L$ and $\mathcal{F}_R$. The data points for which the gradient dominates, are in $\mathcal{F}_L$, the others, in $\mathcal{F}_R$ obtaining essentially a partition into edges and "the rest".

References

[1] M. Aharon, M. Elad, and A.M. Bruckstein. The k-svd: An algorithm for designing of overcomplete dictionaries for sparse representation. *IEEE Trans. On Signal Processing*, 54(11):4311–4322, November 2006.

[2] Michal Aharon, Michael Elad, and Alfred M. Bruckstein. On the uniqueness of overcomplete dictionaries, and a practical way to retrieve them. *Linear Algebra Appl.*, 416(1):48–67, 2006.

[3] A. Aldroubi, C. A. Cabrelli, D. Hardin, and U. M. Molter. Optimal shift invariant spaces and their parseval frame generators. *Applied and Computational Harmonic Analysis*, 23(2):273–283, 2007.

[4] A. Aldroubi, C. A. Cabrelli, D. Hardin, U. M. Molter, and A. Rodado. Determining sets of shift invariant spaces. In *Wavelets and Their Applications*, Chennai, India, 2004. Allied Publishers.

[5] A. Aldroubi, C. A. Cabrelli, and U. Molter. Optimal non-linear models for sparsity and sampling. *Journal of Fourier Analysis and Applications*, 14(5):793–812, December 2008.

[6] A. Aldroubi and K-H. Gröchenig. Non-uniform sampling in shift-invariant space. *Siam Review*, 43(4):585–620, 2001.

[7] A. Aldroubi and R. Tessera, Sparsity and the MAP, preprint.

[8] A. Aldroubi and K. Zaringhalam, Nonlinear least squares in R^n, *Acta Appl. Math.*, 107 (2009), no. 1-3, 325–337.

[9] Magalí Anastasio and Carlos Cabrelli. Sampling from a union of subspaces with frame generators. *Sampling Theory in Signal and Image Processing*, to appear 2009.

[10] R. Baraniuk, M. Davenport, Ronald A. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. Preprint, 2007.

[11] E. Candès and J. Romberg. Quantitative robust uncertainty principles and optimally sparse decompositions. *Foundations of Comput. Math.*, 6:227–254, 2006.

[12] E. Candès, J. Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. on Information Theory*, 52:489–509, 2006.

[13] E. Candès and Terence Tao. Near optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. on Information Theory*, 52:5406–5425, 2006.

[14] Ronald A. DeVore. Deterministic constructions of compressed sensing matrices. Preprint, 2007.

[15] David L. Donoho. Compressed sensing. *IEEE Trans. on Information Theory*, 52:1289–1306, 2006.

[16] Pier Luigi Dragotti, M. Vetterli, and T. Blu. Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets strang-fix? *IEEE Transactions on Signal Processing*, 55:1741–1757, 2007.

[17] C. Eckart and G. Young. The approximation of one matrix by another of lower rank. *Psychometrica*, 1:211–218, 1936.

[18] R. Gribonval and M. Nielsen. Sparse decompositions in unions of bases. *IEEE Trans. Inf. Theory*, 49:3320–3325, 2003.

[19] Y. Lu and M. N. Do. A theory for sampling signals from a union of subspaces. *IEEE Transactions on Signal Processing,* 2007.

[20] I. Maravic and M. Vetterli. Sampling and reconstruction of signals with finite rate of innovation in the presence of noise. *IEEE Transactions on Signal Processing*, 53:2788–2805, 2005.

[21] Holger Rauhut, K. Schass, and P. Vandergheynst. Compressed sensing and redundant dictionaries. Preprint, 2006.

[22] J. A. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Trans. Inf. Theory*, 50:2231–2242, 2004.

[23] R. Vidal, Y. Ma, and S. Sastry. Generalized principal component analysis (gpca). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:1–15, 2005.

Akram Aldroubi
Department of Mathematics,
Vanderbilt University,
Nashville, TN 37240, USA
akram.aldroubi@vanderbilt.edu

Carlos Cabrelli and Ursula Molter
Departamento de Matemática
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires
Ciudad Universitaria, Pabellón I
1428 Capital Federal, Argentina
and CONICET, Argentina
cabrelli@dm.uba.ar and umolter@dm.uba.ar

Recibido: 27 de marzo de 2009
Aceptado: 29 de octubre de 2009