Sparseness Via ℓ_2 –Norm Minimization?

F. Marti-Lopez.

NeuroPhysics Department, Cuban Neuroscience Center; #15202 of 25th Ave., Playa, Havana City, PO Box 6412, Cuba
(e-mail: felipe@cneuro.edu.cu).

***Abstract-* in this paper, the common general opinion that ℓ_2 −norm can not lead to sparse adaptive representations of discretized signals is refuted by formulating, a certain heuristic ℓ_2 −norm signal decomposition procedure that empirically provides them as contrexamples, whichever the signal and the dictionary are.**

***Index Terms-*. sparseness, adaptive signal representation, Method of Frames, Basis Pursuit, bilevel linear least squares problems in finite dimensions.**

I. INTRODUCTION

Finding the sparsest adaptive representation of arbitrary data over some given dictionary of atoms has been desired since long in diverse applications; for instance, in statistical, image and signal processing applications. The emphasis here has been always on reaching such an adaptive representation, where most atoms are expected to be disregarded. Unfortunately, that process literally requires comparison type search through all the combinations of atoms, what is often impracticable, because process complexity exponentially grows with the atomic length [2, 3, 5].

For overcoming such a serious difficulty, several non-combinatorial heuristic decomposition methods have arisen – for instance, Matching Pursuit (MP), Basis Pursuit (BP) and all their further variants- that empirically can provide sparse enough approximate adaptive representations over certain classes of dictionaries [2]. Specially since BP first formulation in 1995, its continuous improvement in providing ℓ_1 −norm representations has made sparseness an indispensable issue in the adaptive representation heuristics [2, 3, 5, 6]. In this sense, BP has became a sort of emblematic decomposition method for the sparse adaptive representation of discretized signals [5, 6].

Mathematically, the aim of BP is to solve the well-studied convex programming problem with linear equality constraints [2, 3]

$$\min_{x\in\{x\in\mathbf{R}^n \mid Ax=b\}} \|x\|_1 ,$$

where $\|\cdots\|_1$ denotes the ℓ_1 −norm and the real matrix A, $A\in\mathbf{R}^{m\times n}$, is a given overcomplete dictionary from n atoms with length m each, $m<n$; the real vector b, $b\in\mathbf{R}^m$, is any given discretized signal from an arbitrary, but determined family of atoms and the real vector x, $x\in\mathbf{R}^n$, is the array of unknown adaptive linear representation coefficients to be found.

Unfortunately, the uniqueness of BP signal representations is not guaranteed for arbitrary dictionaries or signals [2, 3, 5, 6].

On the other hand, in spite of the apparently inherent lack of sparseness of the adaptive signal representations the Method of Frames (MOF) provides [4], it is still one of the most used signal decomposition methods, maybe because of its theoretical, algebraic simplicity and the special clarity of its underlying geometric idea [2, 4, 10, 11].

In mathematical terms, the goal of MOF is to solve the also well-studied positive definite quadratic minimization problem with linear equality constraints [2, 4]

$$\min_{x\in\{x\in\mathbf{R}^n \mid Ax=b\}} \|x\|_2^2 , \qquad (1)$$

where $\|\cdots\|_2$ denotes the ℓ_2 −norm.

Concerning (1), one additionally knows, from the mathematical programming theory [7], that any such problem has always a unique solution x^{MOF},

$$x^{MOF} = \arg\min_{x\in\{x\in\mathbf{R}^n \mid Ax=b\}} \|x\|_2^2 = A^+ b , \qquad (2)$$

where A^+ denotes the Moore-Penrose pseudoinverse matrix [8] of A.

Whereas the emblematic BP, unlike MOF, is a sparseness-revealing method [2] and the latter formally differs from the former just on the type of norm each of them uses, it is certainly hard not to impute the abovementioned lack of sparseness of MOF solutions to the use of ℓ_2 −norm. As a matter of fact, although no proof has been given never before, the common general opinion is that ℓ_2 −norm can not lead to sparse adaptive representations of discretized signals [2, 3].

So, can ℓ_2 −norm lead to sparseness or not?

The purpose of this work is just to answer definitively that question by showing a certain non-combinatorial heuristic procedure that empirically provides sparse enough ℓ_2-norm representations, whichever the discretized signal and the dictionary are, as contrexamples to the common general opinion.

II. SPARSE ℓ_2-NORM SIGNAL REPRESENTATIONS

Let $\widetilde{X}$ be the set

$$\widetilde{X} = \{x \in \mathbf{R}^n \mid sign(x_j^{MOF})x_j \geq 0, \\ j = 1,2,\cdots,n\},$$

where x^{MOF} denotes the corresponding unique signal representation MOF provides.

Besides, let $\{j_1, j_2, \cdots, j_n\} \in \mathbf{N}$ be the sub-index array of the x^{MOF} components, ordered in such a way that $\left|x_{j_k}^{MOF}\right| \geq \left|x_{j_{k+1}}^{MOF}\right|$, $k = 1,2,\cdots,n-1$.

If one defines the matrix $T \in \mathbf{R}^{n\times n}$,

$$T = (t_{ij})_{n\times n} \mid t_{ij_k} = \begin{cases} sign(x_{j_k}^{MOF}), & i \leq k \\ 0, & i > k \end{cases},$$

then [8]

$$rank(T) = n,$$

whichever A and b are.

Now, denote by $x_{j:n}$ the vectors

$$x_{j:n} = (0 \quad 0 \quad \cdots \quad 0 \quad x_j \quad x_{j+1} \quad \cdots \\ \cdots \quad x_{n-1} \quad x_n)', \; j = 1,2,\cdots,n;$$

where $x \in \mathbf{R}^n$; and for simplicity, without loss of generality, assume x^{MOF} satisfies

$$\left|x_j^{MOF}\right| \geq \left|x_{j+1}^{MOF}\right|, \; j = 1,2,\cdots,n-1.$$

Thus, for any x, belonging to $\widetilde{X}$,

$$\left(\|x\|_1 \quad \|x_{2:n}\|_1 \quad \cdots \quad |x_n|\right)' = Tx.$$

Hence, the following sparseness-revealing link between ℓ_1- and ℓ_2-norms can be stated:

$$\|Tx\|_2^2 = \|x\|_1^2 + \sum_{k=2}^{n-1}\|x_{k:n}\|_1^2 + |x_n|^2.$$

Therefore, the positive definite bilevel linear least squares problem with bound constraints in finite dimensions

$$\min_{x \in Arg \min_{x\in\widetilde{X}}\|Ax-b\|_2^2} \|Tx\|_2^2, \qquad (3)$$

as a constrained bilevel convex programming problem with elliptic ℓ_2-norm object function, has a sparse unique solution, whichever A and b are [1, 7, 11, 12] (see figures 1-3).

So, at least every effective method for solving the problem (3) can be defined as a sparseness-revealing ℓ_2-norm signal decomposition method; that is to say, as a procedure for the sparse adaptive representation of discretized signals on ℓ_2-norm.

III. EMPIRICAL VALIDATION

For testing purposes, diverse arbitrary dictionaries were created as random matrices, whose elements were drawn from a $(-1,1)$ uniform distribution each, after considering three necessary study cases [5, 6]:

1. Dictionary with various (distinct) orthonormal bases.
2. Dictionary with multiple (repeated) orthonormal bases.
3. Dictionary with non-orthonormal basis.

The signals to be represented were constructed as linear combinations of the dictionary atoms with a few amount of random coefficients, also drawn from a $(-1,1)$ uniform distribution each (figure 1). Optionally, uniformly distributed noise was added to generated signal.

Whereas, by construction,

$$Arg \min_{x\in\widetilde{X}}\|Ax-b\|_2^2 = \\ = \{x \in \widetilde{X} \mid A'Ax = A'b\} \neq \phi, \qquad (4)$$

whichever A and b are, for a more conventional implementation of the above sparseness-revealing procedure, every bilevel problem (3) was restated and solved as its circumstantially equivalent monolevel *Quadratic Programming* problem [7],

$$\min_{x \in \{x \in \tilde{X} | A'Ax = A'b\}} \|Tx\|_2^2 . \qquad (5)$$

Of course, in order to avoid numerical difficulties with the possible high *sensitiveness* of x^{MOF} to data noise when dictionaries are *rank-deficient* or *ill-conditioned* matrices, a *robust*[1] approximating variant of MOF, named *Approximating Method of Frames* (AMOF) [10, 11], was used instead for approximating (2).

The performed simulation experiment confirmed every solution of (3) as a sparse enough representation of the respective signal on the corresponding dictionary, whichever those signal and dictionary are.

So, the required contrexamples have been completed and, therefore, the common general opinion is not true.

In conclusion:

1. Each elliptic $\ell_2 -$norm adaptive signal representation, discussed here, is simultaneously unique and sparse, whichever the signal and the dictionary are.
2. Unlike further existing signal decomposition methods, the discussed elliptic $\ell_2 -$norm procedure can provide robust adaptive sparse representations; that is to say, the dictionaries may not be overcomplete, orthonormally based or standardized and also may be even ill-conditioned or rank-deficient matrices.
3. There is no reason nor evidence to affirm that the discussed elliptic $\ell_2 -$norm decomposition method is the only one to provide such sparse signal representations.

Finally, it should be noted that, although approximate, each elliptic $\ell_2 -$norm representation allowed a good enough reconstruction of the simulated signal (figures 2 and 3). Besides, most of them were graphically sparser not merely than their homologous MOF representations, but, sometimes, also even than the original ones.

REFERENCES

[1] G. Anandalingam and T. Friesz, "Hierarchical optimization: an introduction", *Annals of Operations Research*, num. 34, pp. 1-11, 1992.

[2] S. .S. Chen., D. L. Donoho and M. A. Saunders, "Atomic decomposition by basis pursuit", *SIAM Review*, vol. 43, num. 1, pp. 129-159, 2001.

[3] S. S. Chen, *Basis Pursuit*. PhD thesis, California, Stanford University, 1995.

[4]I Daubechies, "Time-frequency localization operators: a geometric phase space approach", *IEEE Transactions on Information Theory*, vol. 34, pp. 605–612, 1988.

[5] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via L1", Proceedings of the National Academy of Science (PNAS), vol. 100, num. 5, pp. 2197-2202, 2003.

[6] M. Elad and A. Bruckstein, "A Generalized Uncertainty Principle and Sparse Representation in Pairs of Bases", *IEEE Transactions on Information Theory*, vol. 48, num. 9, pp. 605–612, 2002.

[7] R. Fletcher, *Practical Methods of Optimization*, second edition. New York. John Wiley and Sons, 1987.

[8] G. H. Golub and C. F. Van Loan, *Matrix Computations,* second edition. Baltimore, John Hopkins Pub. Co., 1989.

[9] T. Koenig, F. Marti-Lopez and P. Valdes-Sosa, "Topographic time-frequency decomposition of spontaneous EEG". *NeuroImage*, vol. 14, num. 1; pp. 383-390, 2001.

[10] F. Marti-Lopez and T. Koenig, "Approximating method of frames". *Digital Signal Processing*, num. 13, pp. 519-529, 2003.

[11] F. Marti-Lopez, *Splitting Iterations for the Robust Approximation of Linear Least Squares Solutions in Finite Dimensions.* (in Spanish). PhD thesis, Havana City, Cuban Neuroscience Center, 2003 (http://es.geocities.com/fmartilopez).

[12] L. N. Vicente and P. H. Calamai, "Bilevel and multilevel programmming: a bibliography review". *Journal of Global Optimization*, num. 5, pp. 291-306, 1994.

[1] "Robust methods of estimation are methods that work well not only under ideal conditions, but also under conditions representing a departure from an assumed distribution or model" [Launer R.L., Wilkinson G.H. : *Robustness in Statistics*, Academic Press, New York, 1979].

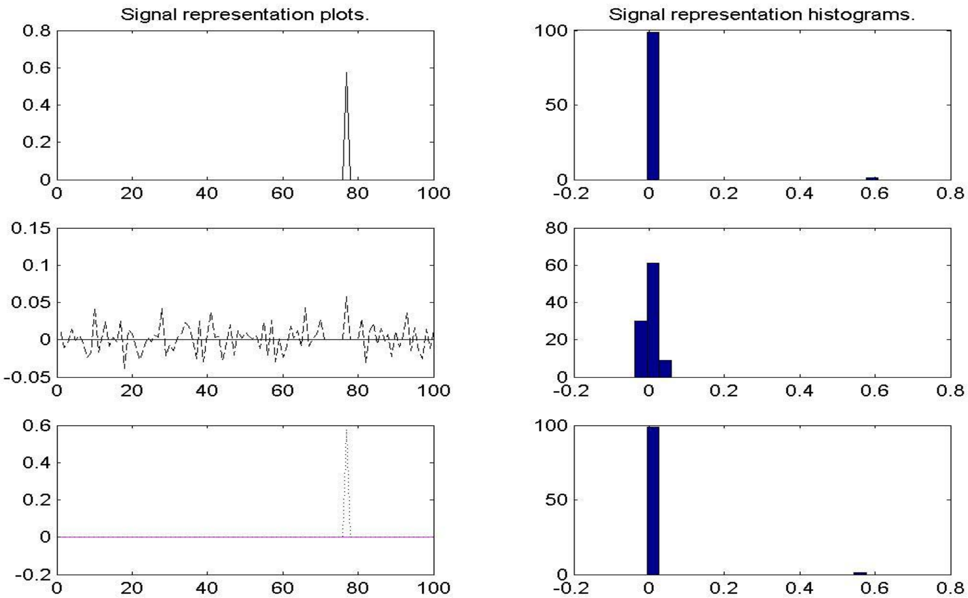

Fig. 1. This figure shows the plots of the original (top-left), the MOF (center-left) and the elliptic ℓ_2-norm (bottom-left) signal representations over the 10×100 dictionary from a first study-case example, without noise, together with the respective histograms (top-, center- and bottom-right).

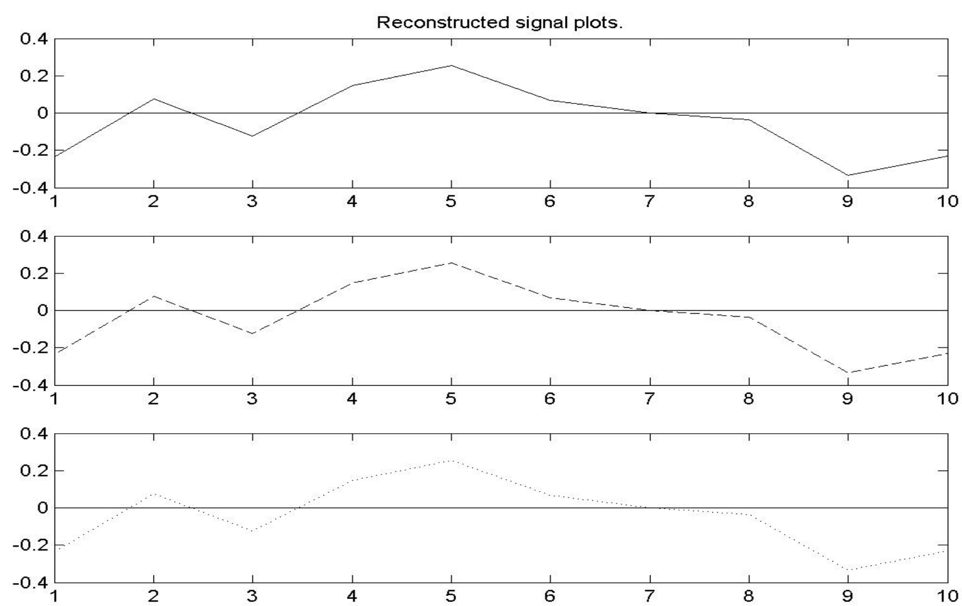

Fig. 2. This figure shows the plots corresponding to the original signal (top) and to the reconstructed ones, due to the MOF (center) and the elliptic ℓ_2-norm (bottom) representations over the dictionary from the above-mentioned first study-case example.

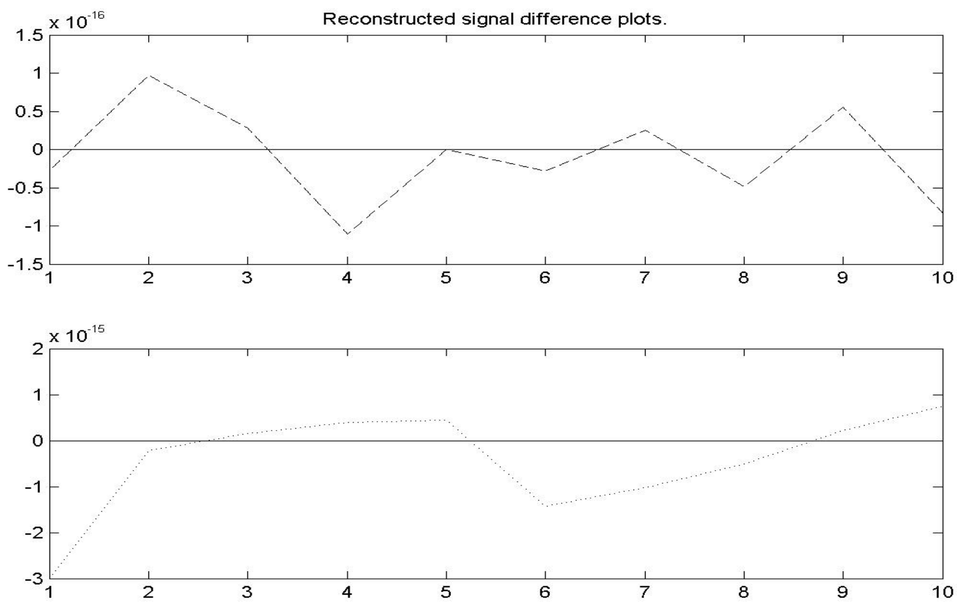

Fig. 3. This figure shows the plots of the difference between the original signal and the reconstructed ones, due to the MOF (top) and the elliptic L2-norm (bottom) representations over the dictionary from the above-mentioned first study-case example.