Invided Talk: International Conference on Image Processing (ICIP) 2001, Tsaloniky, Greece, November 2001.

ON SPARSE SIGNAL REPRESENTATIONS

Michael Elad

Jigami Research Division
Jigami Corporation
Technion City, Haifa 32000, Israel

Alfred M. Bruckstein

Department of Computer Science
The Technion – Israel Institute of Technology
Haifa 32000, Israel

ABSTRACT

An elementary proof of a basic uncertainty principle concerning pairs of representations of $\Re^N$ vectors in different orthonormal bases is provided. The result, slightly stronger than stated before, has a direct impact on the uniqueness property of the sparse representation of such vectors using pairs of orthonormal bases as overcomplete dictionaries. The main contribution in this paper is the improvement of an important result due to Donoho and Huo concerning the replacement of the l_0 optimization problem by a linear programming minimization when searching for the unique sparse representation.

1. INTRODUCTION

Given a real vector $\underline{S}$ in $\Re^N$ it has a unique representation in every basis of this space. Indeed, if $\underline{\phi}_1, \underline{\phi}_2, \dots \underline{\phi}_N$ are N orthogonal vectors of unit length , i.e $\langle \underline{\phi}_i, \underline{\phi}_j \rangle = \delta_{ij}$, we have that

$$\underline{S} = \sum_{i=1}^{N} \alpha_i \underline{\phi}_i$$

with coefficients α_i given by $\langle \underline{\phi}_i, \underline{S} \rangle$. Suppose now that we have two different bases for $\Re^N$: $\mathbf{\Phi} = \{\underline{\phi}_1\ \underline{\phi}_2\ \dots\ \underline{\phi}_N\}$ and $\mathbf{\Psi} = \{\underline{\psi}_1\ \underline{\psi}_2\ \dots\ \underline{\psi}_N\}$. Then every vector $\underline{S}$ has representations both in terms of $\{\underline{\phi}_i\}$ and in terms of $\{\underline{\psi}_i\}$. Let us write then

$$\underline{S} = \left[\underline{\phi}_1\ \underline{\phi}_2\ \cdots\ \underline{\phi}_N \right] \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix} = \left[\underline{\psi}_1\ \underline{\psi}_2\ \cdots\ \underline{\psi}_N \right] \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{bmatrix}$$

A question posed by Donoho and his co-workers is the following: is there some benefit in representing $\underline{S}$ in a joint, overcomplete set of vectors, say

$$\left\{ \underline{\phi}_1\ \underline{\phi}_2\ \cdots\ \underline{\phi}_N\ \underline{\psi}_1\ \underline{\psi}_2\ \cdots\ \underline{\psi}_N \right\}$$

that can be called a "dictionary" concatenating the $\mathbf{\Phi}$ and $\mathbf{\Psi}$ bases? Sparse representations can have advantages in terms of compression of signals and/or in terms of understanding the underlying processes that generated them. The problem that arises, however, is that in terms of "dictionaries" of overcomplete set of vectors (as obtained by concatenating the basis vectors of $\mathbf{\Phi}$ and the $\mathbf{\Psi}$) every signal has multiple representations. Of those multiple representations choosing one based on sparsity is a difficult optimization problem. Indeed suppose we have

$$\underline{S} = [\mathbf{\Phi}\ \mathbf{\Psi}]\,\underline{\gamma} = \sum_{i=1}^{N} \gamma_i^{\phi} \underline{\phi}_i + \sum_{i=1}^{N} \gamma_i^{\psi} \underline{\psi}_i$$

Choosing $\underline{\gamma}$ involves solving an underdetermined set of equations with N equations and 2N unknowns, and hence must be done subject to additional requirements on the solution. The additional requirement for sparsity would be to minimize the support of $\underline{\gamma}$, i.e., minimize the number of places where $\underline{\gamma}$ is nonzero. Hence we need to solve the problem

$$(P_0) \qquad \text{Minimize } \|\underline{\gamma}\|_0 \qquad \text{subject to} \qquad \underline{S} = [\mathbf{\Phi}\ \mathbf{\Psi}]\underline{\gamma}$$

where $\|\underline{\gamma}\|_0$ is the size of the support of $\underline{\gamma}$. The main result of Donoho and Huo is that in case $\underline{S}$ has a "very" sparse representation i.e. when $\exists \underline{\gamma}$ so that $\underline{S} = [\mathbf{\Phi}\ \mathbf{\Psi}]\underline{\gamma}$ and $\|\underline{\gamma}\|_0 < Func.\{[\mathbf{\Phi}\ \mathbf{\Psi}]\}$, then this sparse representation is the unique solution of not only (P_0) as defined above, but also of

$$(P_1) \qquad \text{Minimize } \|\gamma\|_1 \qquad \text{subject to} \qquad \underline{S} = [\mathbf{\Phi}\ \mathbf{\Psi}]\gamma$$

where $\|\underline{\gamma}\|_1 = \sum_j |\gamma_j|$, the l_1 -norm of $\underline{\gamma}$. This is an important result stating that "very sparse" representations can be found by solving not a combinatorial search problem as implied by (P_0) but by solving a much simpler linear programming problem as implied by minimizing to l_1-norm of $\underline{\gamma}$. The bound defining sparsity as provided by Donoho and Huo is $Func.\{[\mathbf{\Phi}\ \mathbf{\Psi}]\} = \frac{1}{2}\left(1 + M^{-1}\right)$, where $M = \text{Sup}\{|\langle \underline{\phi}_i, \underline{\psi}_j \rangle|\ , \forall\ i, j\ \}$.

Here we follow the path of work of Donoho and Huo, and improve their bounds. First, we prove an "improved" uncertainty principle leading to better bounds yielding uniqueness of the (P_0) solution. The result is that uniqueness of the (P_0)-solution is achieved for $\|\underline{\gamma}\|_0 < \frac{1}{M}$. Our main contribution in this paper is an improvement of the result concerning the replacement of the (P_0) minimization problem with the (P_1) convex minimization, while achieving the same solution. We show that the two problems' solutions coincide for $\|\underline{\gamma}\|_0 < \frac{(\sqrt{2}-0.5)}{M}$.

2. THE BASIC UNCERTAINTY PRINCIPLE

We shall first prove a basic "uncertainty principle" concerning pairs of representations of a given vector $\underline{S}$ (or signal) in two given orthonormal bases $\mathbf{\Phi}$ and $\mathbf{\Psi}$. Suppose a signal $\underline{S}$ has the representations:

$$\underline{S} = \mathbf{\Phi}\underline{\alpha} = \mathbf{\Psi}\underline{\beta}$$

and, w.l.o.g. we assume that $\underline{S}^T\underline{S} = 1$, i.e. we have normalized the l_2 energy of the signal to 1. We have $\alpha_i = \langle \underline{S}, \underline{\phi}_i \rangle$ and $\beta_j = \langle \underline{S}, \underline{\psi}_j \rangle$. Now let us write

$$1 = \underline{S}^T\underline{S} = \underline{\alpha}^T \mathbf{\Phi}^T \mathbf{\Psi} \underline{\beta}. \tag{1}$$

Writing $M = Sup\left\{|\underline{\phi}_i^T \underline{\psi}_j|, \forall(i,j)\right\}$, all the entries in the matrix $\mathbf{\Phi}^T\mathbf{\Psi}$ are smaller than M. We further have the (Parseval) energy preservation property $1 = \underline{S}^T\underline{S} = \sum \alpha_i^2 = \sum \beta_i^2$. Assuming that $\underline{\alpha}\|_0 = A$ and $\|\underline{\beta}\|_0 = B$, we get that

$$1 = \sum_{i'=1}^{A}\sum_{j'=1}^{B} \alpha_{i'} \langle \underline{\phi}_{i'}, \underline{\psi}_{j'} \rangle \beta_{j'} \leq M \cdot \sum_{i'=1}^{A} |\alpha_{i'}| \sum_{j'=1}^{B} |\beta_{j'}| \tag{2}$$

where i' runs over the nonzero support of $\underline{\alpha}$ and j' runs over the nonzero support of $\underline{\beta}$. Next, in order to bound the above expression from above we can maximize it over all α_i and β_j, bearing in mind that these unknowns must be positive and with l_2 unit norm. This problem can be maximized using simple Lagrange analysis. The optimal result is $\alpha_i = \frac{1}{\sqrt{A}}$, $\beta_i = \frac{1}{\sqrt{B}}$. Returning to our derivation of the uncertainty relations in Equation 2 we now have:

$$1 \leq M \cdot \sum_{i'=1}^{A}\sum_{j'=1}^{B} |\alpha_{i'}||\beta_{j'}| \leq M\sqrt{AB}$$

We have obtained the following result: If we have two representations of a signal $\underline{S}$ in the bases $\mathbf{\Phi} = [\underline{\phi}_1\ \underline{\phi}_2\ \cdots\ \underline{\phi}_N]$ and $\mathbf{\Psi} = [\underline{\psi}_1\ \underline{\psi}_2\ \cdots\ \underline{\psi}_N]$ and the coefficient vectors $\underline{\alpha}$ and $\underline{\beta}$ have supports of sizes $\|\underline{\alpha}\|_0 = A$ and $\|\underline{\beta}\|_0 = B$ then $\sqrt{AB} \geq \frac{1}{M}$.

Using the well–known inequality between the geometric and arithmetic means, $\frac{A+B}{2} \geq \sqrt{AB}$, we also have that

$$\frac{A+B}{2} \geq \left(\sqrt{AB}\right) \geq \frac{1}{M} \quad \Rightarrow \quad A + B \geq \frac{2}{M}.$$

Donoho and Huo obtained, by emulating arguments for the sinusoidal and spike bases a weaker result stating that

$$A + B \geq \left(1 + \frac{1}{M}\right)$$

(see [2]), and clearly, the new bound is higher since $M \leq 1$. Note that the value of M is crucial in the above arguments. For any pair of orthonormal bases $\mathbf{\Phi}$ and $\mathbf{\Psi}$ of $\Re^N$ we have that $M \geq \frac{1}{\sqrt{N}}$.

3. UNIQUENESS OF SPARSE REPRESENTATIONS

A direct consequence of the uncertainty relation derived above is the following fact: if we have a "sparse" representation in terms of the $[\mathbf{\Phi}\ \mathbf{\Psi}]$ dictionary, it is unique. How sparse the representation should be to achieve such uniqueness depends crucially on the bound provided by the uncertainty principle. The connection follows easily from the following line of argumentation (taken from [2]).

Suppose $\underline{\gamma}_1$ and $\underline{\gamma}_2$ are the coefficient vectors of two different representations of the same signal $\underline{S}$, i.e. $\underline{S} = [\mathbf{\Phi}\ \mathbf{\Psi}]\underline{\gamma}_1 = [\mathbf{\Phi}\ \mathbf{\Psi}]\underline{\gamma}_2$. Then clearly

$$[\mathbf{\Phi}\ \mathbf{\Psi}][\underline{\gamma}_1 - \underline{\gamma}_2] = \underline{0} \Rightarrow \mathbf{\Phi}\underline{\gamma}^{\phi}_{\Delta} = -\mathbf{\Psi}\underline{\gamma}^{\psi}_{\Delta} = \underline{\Lambda}$$

Hence in this case we have two vectors $\underline{\gamma}^{\phi}_{\Delta}$ and $\underline{\gamma}^{\psi}_{\Delta}$ (defined as the upper N values in $\underline{\gamma}_1 - \underline{\gamma}_2$ and the lower N values in $\underline{\gamma}_2 - \underline{\gamma}_1$) that are different from $\underline{0}$ and represent the same vector $\underline{\Lambda}$ in two orthogonal bases. Now the basic uncertainty principle states that if

$$\|\underline{\gamma}^{\phi}_{\Delta}\|_0 = A \ \text{ and } \ \|\underline{\gamma}^{\psi}_{\Delta}\|_0 = B$$

then we must have $A + B \geq \left(2\sqrt{AB} \geq\right) \frac{2}{M}$. Suppose that the original representations were both sparse, i.e. $\|\underline{\gamma}_1\|_0 < F$ and $\|\underline{\gamma}_2\|_0 < F$. Then we must necessarily have

$$\|\underline{\gamma}_1 - \underline{\gamma}_2\|_0 \ < \ \|\underline{\gamma}_1\|_0 + \|\underline{\gamma}_2\|_0 < \ 2F$$

On the other hand we have $\|\underline{\gamma}_1 - \underline{\gamma}_2\|_0 = \|\underline{\gamma}^{\phi}_{\Delta}\|_0 + \|\underline{\gamma}^{\psi}_{\Delta}\|_0 = A + B$. Hence sparsity of both $\underline{\gamma}_1$ and $\underline{\gamma}_2$ with bound F implies that $A + B < 2F$. But by the uncertainty principle we have

$$A \ + \ B \ \geq \ \frac{2}{M}$$

In conclusion, if F would be M^{-1} or smaller, we would contradict the uncertainty principle if we would assume two different sparse representations. Hence we have the following uniqueness theorem: If a signal $\underline{S}$ has a sparse representation in the dictionary $[\mathbf{\Phi}\ \mathbf{\Psi}]$ so that

$$\underline{S} = [\mathbf{\Phi}\ \mathbf{\Psi}]\underline{\gamma} \ \text{ and } \|\underline{\gamma}\|_0 < \tfrac{1}{M}$$

then this sparse representation is necessarily unique. This bound is a better bound than the one implied by the uncertainty principle stated in [2], which is $\frac{1}{2}\left(1 + M^{-1}\right)$. This means that the uniqueness result will be true for much less sparse representations than those required by the bound provided in Ref. [2].

4. SPARSE REPRESENTATIONS VIA L_1 OPTIMIZATION

The next question that naturally arises is: if a signal $\underline{S}$ has a sparse representation in a dictionary $[\mathbf{\Phi}\ \mathbf{\Psi}]$, how should we find it? Solving the l_0 optimization problem defined as

$$(P_0) \quad \text{Minimize} \quad \|\underline{\gamma}\|_0 = \sum_{k=0}^{2N} \gamma_k^0 \quad \text{subject to} \quad \underline{S} = [\mathbf{\Phi}\ \mathbf{\Psi}]\underline{\gamma}$$

involves an unfeasible search problem. However, it was discovered experimentally that solving the l_1 optimization problem

$$(P_1) \quad \text{Minimize} \quad \|\underline{\gamma}\|_1 = \sum_{k=0}^{2N} |\gamma_k| \quad \text{subject to} \quad \underline{S} = [\mathbf{\Phi}\ \mathbf{\Psi}]\underline{\gamma}$$

often provides the sparse representation. Donoho and Huo proved in [2] that the strong sparsity condition $\|\underline{\gamma}\|_0 < \frac{1}{2}\left(1 + M^{-1}\right)$ is a condition ensuring that the solution of the problem (P_1) yields the sparse solution of (P_0) too. This is a wonderful result, since (P_1) is essentially a linear programming problem!

To show that (P_1) provides also the (P_0) solution one has to prove that if $\|\underline{\gamma}\|_0 < F$ and $[\mathbf{\Phi}\ \mathbf{\Psi}]\underline{\gamma} = \underline{S}$, then, if there exists some other representation $[\mathbf{\Phi}\ \mathbf{\Psi}]\underline{\tilde{\gamma}} = \underline{S}$, we must have $\|\underline{\tilde{\gamma}}\|_1 \geq \|\underline{\gamma}\|_1$. Here we show how Donoho and Huo's bound can be improved. Following [2] we have that if:

$$[\mathbf{\Phi}\ \mathbf{\Psi}]\underline{\gamma} = \underline{S} = [\mathbf{\Phi}\ \mathbf{\Psi}]\underline{\tilde{\gamma}} \ \Rightarrow \ [\mathbf{\Phi}\ \mathbf{\Psi}][\underline{\tilde{\gamma}} - \underline{\gamma}] = 0.$$

Therefore the difference vector $\underline{x} = \tilde{\underline{\gamma}} - \underline{\gamma}$ satisfies

$$\mathbf{\Phi}\underline{x}^{\phi} = \mathbf{\Psi}(-\underline{x}^{\psi}) \tag{3}$$

where we define $\underline{x}^{\phi}$ as the first N entries in $\underline{x}$, and $\underline{x}^{\psi}$ are the last N entries in $\underline{x}$. We need to show that for every nonzero vector $\underline{x}$ that obeys Equation (3) we shall have that

$$\sum_{k=1}^{2N} |\gamma_k + x_k| - \sum_{k=1}^{2N} |\gamma_k| \geq 0.$$

Hence we need to show that

$$\sum_{\text{off support of}\underline{\gamma}} |x_k| \quad + \sum_{\text{on support of}\underline{\gamma}} (|\gamma_k + x_k| - |\gamma_k|) \geq 0$$

Due to $|v+m| \geq |v| - |m|$ we have $|v+m| - |v| \geq |v| - |m| - |v| = -|m|$. Using this inequality we can write that

$$\frac{1}{2}\sum_{\text{all sup.}} |x_k| - \sum_{\text{on sup.}} |x_k| \geq 0 \;\Rightarrow\; \frac{\sum_{\text{on sup.}} |x_k|}{\sum_{\text{all sup.}} |x_k|} \leq \frac{1}{2}. \tag{4}$$

Equations (3) and (4) can re-interpreted as an optimization problem of the following form: Minimize $\frac{1}{2}\sum_{\text{all}} |x_k| - \sum_{\text{on}} |x_k|$ subject to $\mathbf{\Phi}\underline{x}^{\phi} = \mathbf{\Psi}(-\underline{x}^{\psi})$. This problem should be solved for various values of F ($F = \|\underline{\gamma}\|_0$) and all profiles of non-zero entries in $\underline{\gamma}$. The maximal F that yields a minimum that is still above zero will be the bound on the sparsity of $\underline{\gamma}$, ensuring equivalence between (P_0) and (P_1).

Working with the above minimization problem is complicated because od several reasons (i) we need an explicit condition to avoid the trivial solution $\underline{x} = \underline{0}$; (ii) the problem should involve only the absolute values of the entries in the vector $\underline{x}$; (iii) the orthonormal matrices $\mathbf{\Psi}$ and $\mathbf{\Phi}$ should appear implicitly through the parameter M they define; and (iv) the sensitivity to the location of non-zero elements in the support of $\underline{\gamma}$ should be removed.

The first problem is solved by introducing an additional constraint of the form $\sum_{\text{all}} |x_k| = 1$. It is easily verified that this constraint does not change the sign of the result. As to the other problems, they are solved via the definition of an alternative minimization problem. In this new problem we minimize the same function, but pose a weaker set of constraints, as follows

$$\text{Minimize} \quad \left[\frac{1}{2}\sum_{\text{all}} |x_k| - \sum_{\text{on}} |x_k|\right] \quad \text{subject to:}$$
$$|x^{\phi}| \leq M \cdot \mathbf{1}_{N\times N}|x^{\psi}|, \;\; |x^{\psi}| \leq M \cdot \mathbf{1}_{N\times N}|x^{\phi}|, \;\; \sum_{\text{all}} |x_i| = 1$$

where $\mathbf{1}_{N\times N}$ is an N by N matrix containing ones in all its entries.

The first and second constraints simply use the fact that any given entry in one of the vectors ($\underline{x}^{\psi}$ or $\underline{x}^{\phi}$) cannot be grater than M multiplying the sum of the absolute entries in the other vector. Clearly, every feasible solution of the original constraint set is also a feasible solution of the new constraint set, but not vice-versa. Thus, if the minimum of the function is still positive using the new constraint set, it implies that it is surely positive using the original constraint set.

Looking closely at the the newly defined optimization problem, we can rewrite it as

$$\text{Minimize} \;\; \left[\frac{1}{2} - \mathbf{1}_{\gamma 1}^T \underline{X}_1 - \mathbf{1}_{\gamma 2}^T \underline{X}_2\right] \;\; \text{subject to:} \tag{5}$$
$$\underline{X}_1 \leq M \cdot \mathbf{1}_{N\times N} \;\; \underline{X}_2, \quad \underline{X}_2 \leq M \cdot \mathbf{1}_{N\times N}\underline{X}_1$$
$$\mathbf{1}_N^T(\underline{X}_1 + \underline{X}_2) = 1, \quad \underline{X}_1 \geq 0, \;\; \underline{X}_2 \geq 0$$

where we define $\underline{X}_1$ and $\underline{X}_2$ as the absolute values of the original vectors $\underline{x}^{\psi}$ and $\underline{x}^{\phi}$. This is the reason we added the fourth constraint regarding positivity of the unknowns. The notations $\mathbf{1}_{\gamma 1}$ and $\mathbf{1}_{\gamma 2}$ stand for vectors of length N, $[\mathbf{1}_{\gamma 1} \;\; \mathbf{1}_{\gamma 2}]$ being the $2N$ vector with ones where $\underline{\gamma}$ is non-zero. $\mathbf{1}_N$ is simply an N vector containing all ones. If we assume that there are K_1 non-zeros in $\mathbf{1}_{\gamma 1}$ and K_2 non-zeros in $\mathbf{1}_{\gamma 2}$, then $K_1 + K_2 = \|\gamma\|_0$. In the new formulation, we can assume w.l.o.g. that the K_1 and K_2 non-zeros are located at the beginning of the two vectors $\underline{X}_1$ and $\underline{X}_2$, due to the symmetrical form of the constraints.

The problem we have obtained is a classical Linear Programming (LP) problem, and as such, has a unique local minimum point which is also the unique global minimum point. Let us bring it to its canonical form

$$(P) \quad \text{Minimize} \;\; \underline{C}^T\underline{Z} \;\; \text{subject to} \;\; \mathbf{A}\underline{Z} \geq \underline{B} \;\; , \;\; \underline{Z} \geq 0.$$

The matrices $(A, \underline{Z}, \underline{B}, \underline{C})$ involved are defined as follows

$$\underline{Z}^T = \left[\underline{X}_1^T \;\; \underline{X}_1^T\right], \;\; \underline{C}^T = \left[-\mathbf{1}_{\gamma 1}^T \;\; -\mathbf{1}_{\gamma 2}^T\right]$$
$$\underline{B}^T = \left[0 \cdot \mathbf{1}_N^T \;\; 0 \cdot \mathbf{1}_N^T \;\; 1 \;\; -1\right]$$
$$\mathbf{A} = \begin{bmatrix} -\mathbf{I}_N & M \cdot \mathbf{1}_{N\times N} \\ M \cdot \mathbf{1}_{N\times N} & -\mathbf{I}_N \\ \mathbf{1}_N^T & \mathbf{1}_N^T \\ -\mathbf{1}_N^T & -\mathbf{1}_N^T \end{bmatrix}$$

In the problem defined in Equation (5) we wanted conditions so that the minimum will not be negative. After removing the $1/2$ from the function, the new requirement becomes $\underline{C}^T\underline{Z} \geq -0.5$. Since it is still difficult to give an analytic form to the solution of the LP problem we obtained, we shall exploit the dual LP problem of the form

$$(D) \quad \text{Maximize} \;\; \underline{B}^T\underline{U} \;\; \text{subject to} \;\; \mathbf{A}^T\underline{U} \leq \underline{C} \;\; , \;\; \underline{U} \geq 0$$

with the same matrices as in the primal problem (see e.g. [4]). The approach we are going to take is as follows: We know that the primal and the dual problems obtain the same optimal value ([4]), i.e., Minimize $\{\underline{C}^T\underline{Z}\}$ = Maximize $\{\underline{B}^T\underline{U}\}$. We require this optimal value to be higher than or equal to -0.5. In the dual problem, if we find a feasible solution $\underline{U}$ such that $\underline{B}^T\underline{U} \geq -0.5$, we guarantee that the maximal value is also above -0.5, and thus we fulfill the original requirement on the primal problem.

Let us consider the following parameterized form for a feasible solution for $\underline{U}$

$$\underline{U}^T = \left[\mathbf{1}_{\gamma 1}^T \;\; \alpha\mathbf{1}_{\gamma 2}^T \;\; \beta \;\; \gamma\right].$$

Using previous notations, there are K_1 non-zeros in $\mathbf{1}_{\gamma 1}$ and K_2 non-zeros in $\mathbf{1}_{\gamma 2}$. We assume w.l.o.g. that $K_1 \leq K_2$ (the problem is perfectly symmetric with respect to these two sections). We also assume that $0 \leq \alpha \leq 1$. This way, three parameters (α, β, γ) govern the entire solution $\underline{U}$, and we need to find requirements on them in order to guarantee that the proposed solution is indeed feasible. Substituting the proposed solution form into the constraint inequalities of the dual problem and using the fact that $\mathbf{1}_{N\times N}\mathbf{1}_{\gamma_2} = K_2\mathbf{1}_N$ we get

$$\alpha M K_2 + (\beta - \gamma) \le 0, \quad (1-\alpha) + MK_1 + (\beta - \gamma) \le 0$$

Solving the two inequalities as equalities, we get

$$\alpha M K_2 = (1-\alpha) + MK_1 \;\; \Rightarrow \alpha = \frac{1+MK_1}{1+MK_2}$$

We see that, indeed, our assumption $0 \le \alpha \le 1$ is correct since we assumed $K_1 \le K_2$, (note that $M > 0$). So far we have found an expression for the first parameter α. As to the other two, substituting α in the above equations we get

$$(\beta - \gamma) = -MK_2 \cdot \frac{1+MK_1}{1+MK_2} < 0.$$

Thus, we can choose $\beta = 0$ and γ will be the above expression multiplied by -1. This way we have satisfied all the inequality constraints, and obtained a solution $\underline{U}$ which is also non-negative in all its entries.

Now that we have established that the proposed solution is feasible, let us look at the value of the function. The function is simply $\underline{B}^T\underline{U} = (\beta - \gamma)$. So we should require

$$\underline{B}^T\underline{U} = (\beta - \gamma) = -MK_2 \cdot \frac{1+MK_1}{1+MK_2} \ge -\frac{1}{2}.$$

Thus

$$MK_2 \cdot \frac{1+MK_1}{1+MK_2} \le \frac{1}{2} \;\; \Rightarrow \;\; 2M^2K_1K_2 + MK_2 - 1 \le 0 \quad (6)$$

We have therefore obtained a requirement on K_1 and K_2 which is posed in terms of the parameter M.

A simpler sparsity condition on the representation $\underline{\gamma}$ can be derived by bounding $K_1 + K_2$. It turns out that (see Ref. [5])

$$\|\underline{\gamma}\|_0 = K_1 + K_2 \le (\sqrt{2} - 0.5) \cdot \frac{1}{M}$$

Figure 1 shows graphically how the various discussed bounds compare. In this graph we have assumed $M = 1/\sqrt{128}$. Note that since $0 \le K_1 \le K_2$, only the upper-left part of the graphs is relevant, and thus we have masked the non-relevant zone. We can see that the $1/M$ bound (which is also the uniqueness bound) is valid at the extremes, whereas the $(\sqrt{2} - 0.5)/M$ is relevant in the middle of the K_1 zone.

The result obtained is better than the $0.5\,(1 + 1/M)$-bound asserted by Donoho and Huo. As an example, for $M = 1/\sqrt{N}$, we get that for $N = 64$ the old requirement (DH) is to have less than 4.5 non-zeros, while we (EB) require 7.3 and below. As N goes to infinity the ratio between the two bounds becomes 1.8284.

5. CONCLUSIONS

In this paper we presented a new uncertainty theorem on the minimal joint-sparseness of a vector represented by two orthonormal bases. Based on this theorem, we proved that, using a dictionary of the two orthonormal bases, a representation with less than $1/M$ non-zeros is guaranteed to be the unique sparse representation. The main contribution of this paper concentrated to the way

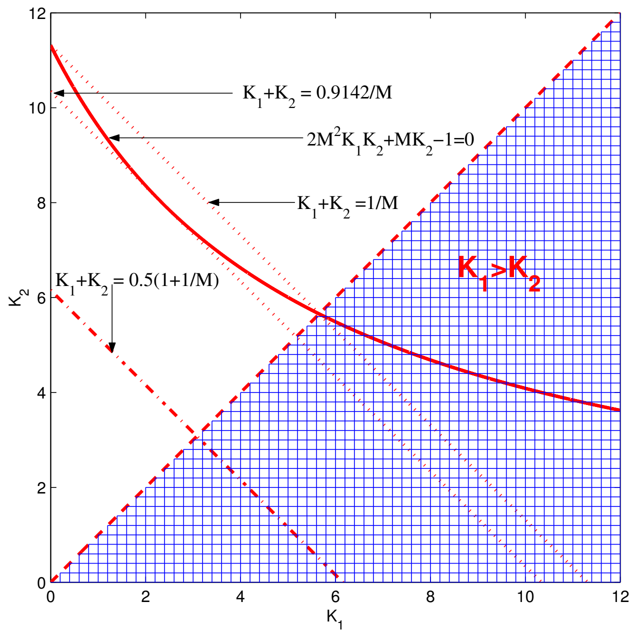

Fig. 1. Figure 1 - A graphic description of the bounds obtained.

to find the sparse representation over an overcomplete dictionary as described above. We have found that if there exists a sparse representation with less than $0.9142/M$ non-zeros, than this representation can be found using the minimization of the l_1-norm, which leads to solving a linear programming problem.

6. ACKNOWLEDGEMENTS

This work emerged as a follow-up on David Donoho's extremely interesting talk, based on [2], at the Technion (Israel Institute of Technology) in December 2000.

7. REFERENCES

[1] D.L.Donoho & P.B Stark (June 1989) *Uncertainty Principles and Signal Recovery*, SIAM Journal on Applied Mathematics, Vol 49/3,pp 906-931.

[2] D.L.Donoho & X.Huo (June 1999) *Uncertainty Principles and Ideal Atomic Decomposition*, www manuscript.

[3] S. Mallat(1998) **A Wavelet Tour of Signal Processing**, Academic Press, Second Edition.

[4] D. Bertsekas (1995) **Non-Linear Programming**, Athena Publishing

[5] M. Elad & A.M. Bruckstein (June 2001) *A Generalized Uncertainty Principle and Sparse Representation in Pairs of* $\Re^N$ *Bases*, CIS Report 2001-06, Technion, Israel