A Generalized Uncertainty Principle and Sparse Representation in Pairs of $\Re^N$ Bases

Michael Elad[*]   and   Alfred M. Bruckstein[†]

June 14th, 2001

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 $\underline{\phi}_1, \underline{\phi}_2, \ldots \underline{\phi}_N$ are N orthogonal vectors of unit length , i.e $\langle \underline{\phi}_i, \underline{\phi}_j \rangle = \delta_{ij}$, we have that

$$\underline{S} = \begin{bmatrix} \underline{\phi}_1 \ \underline{\phi}_2 \ \cdots \ \underline{\phi}_N \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix} = \sum_{i=1}^{N} \alpha_i \underline{\phi}_i$$

with coefficients α_i given by $\langle \underline{\phi}_i, \underline{S} \rangle$, since

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix} = \begin{bmatrix} \underline{\phi}_1^T \\ \underline{\phi}_2^T \\ \vdots \\ \underline{\phi}_N^T \end{bmatrix} \underline{S}.$$

[*]Jigami Research Division, Jigami Corporation, Haifa 32000, Israel

[†]Department of Computer Science, Technion–Israel Institute of Technology Technion City, Haifa 32000, Israel

Suppose now that we have two different bases for $\Re^N$: $\boldsymbol{\Phi} = \{\underline{\phi}_1 \ \underline{\phi}_2 \ \cdots \ \underline{\phi}_N\}$ and $\boldsymbol{\Psi} = \{\underline{\psi}_1 \ \underline{\psi}_2 \ \cdots \ \underline{\psi}_N\}$. 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 S in a joint, overcomplete set of vectors, say

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

(that can be called a "dictionary" concatenating the $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ bases)?

Indeed, we can consider a signal

$$\underline{S}^* = \gamma_k \underline{\phi}_k + \gamma_l \underline{\psi}_l$$

that has a very sparse representation in terms of the joint set of vectors forming the $\boldsymbol{\Phi}$ and the $\boldsymbol{\Psi}$ bases, but, in general, $\underline{S}^*$ will have highly non-sparse representations in either of this bases alone.

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 $\boldsymbol{\Phi}$ and the $\boldsymbol{\Psi}$) every signal has multiple representations. Of those multiple representations choosing one based on sparsity is a difficult optimization problem. Indeed suppose we have

$$\underline{S} = \{\underline{\phi}_1 \ \underline{\phi}_2 \ \cdots \ \underline{\phi}_N \ \underline{\psi}_1 \ \underline{\psi}_2 \ \cdots \ \underline{\psi}_N\} \begin{bmatrix} \gamma_1^\phi \\ \gamma_2^\phi \\ \vdots \\ \gamma_N^\phi \\ \cdots \\ \gamma_1^\psi \\ \gamma_2^\psi \\ \vdots \\ \gamma_N^\psi \end{bmatrix} = [\boldsymbol{\Phi} \ \boldsymbol{\Psi}]\underline{\gamma} = \sum_{i=1}^N \gamma_i^\phi \underline{\phi}_i + \sum_{i=1}^N \gamma_i^\psi \underline{\psi}_i$$

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

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

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 $\underline{S} = [\boldsymbol{\Phi} \ \boldsymbol{\Psi}]\underline{\gamma}$ and $\|\underline{\gamma}\|_0 < Function\{[\boldsymbol{\Phi} \ \boldsymbol{\Psi}]\}$, then this sparse representation is the unique solution of not only (P_0) as defined above, but also of

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

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

$$Function\{[\boldsymbol{\Phi} \ \boldsymbol{\Psi}]\} = \frac{1}{2} \left(1 + M^{-1}\right)$$

where

$$M = \text{Sup}\{|\langle \underline{\phi}_i, \underline{\psi}_j \rangle| \ , \forall \ \ i, j \ \}$$

The results of Donoho and Huo are based on exploiting an uncertainty principle showing that a signal cannot have representations that are simultaneously sparse in two orthonormal bases. This then leads to a result showing that sufficiently sparse representations in "dictionaries" concatenating the two bases must be unique, showing the uniqueness of the solution of problem (P_0) as defined above. Subsequently for "very" sparse representations (considerably sparser than the ones ensuring uniqueness !) it is shown that solving (P_1) leads also to the unique solution of (P_0).

In the general case discussed here Donoho and Huo showed in [2] that uniqueness of the solution of of the (P_0)-problem is ensured for $\|\underline{\gamma}\|_0 < \frac{1}{2}\left(1 + \frac{1}{M}\right)$, and that in this case (P_1) provides this unique solution as well.

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' $((P_0)$ and $(P_1))$ solutions coincide for $\|\underline{\gamma}\|_0 < \frac{(\sqrt{2}-0.5)}{M}$. This way we enlarge the class of signals for which we can apply a simple, linear-programming based search for the optimal sparse representation.

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 $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$. Suppose a signal $\underline{S}$ has the

representations:

$$\underline{S} = \begin{bmatrix} \underline{\phi}_1 \ \underline{\phi}_2 \ \cdots \ \underline{\phi}_N \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix} = \mathbf{\Phi}\underline{\alpha} = \begin{bmatrix} \underline{\psi}_1 \ \underline{\psi}_2 \ \cdots \ \underline{\psi}_N \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{bmatrix} = \mathbf{\Psi}\underline{\beta}$$

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

$$\alpha_i = \langle \underline{S}, \underline{\phi}_i \rangle \quad , \quad \beta_j = \langle \underline{S}, \underline{\psi}_j \rangle$$

Now let us write

$$1 = \underline{S}^T\underline{S} = [\alpha_1 \ \alpha_2 \ldots \alpha_N] \begin{bmatrix} \underline{\phi}_1^T \\ \underline{\phi}_2^T \\ \vdots \\ \underline{\phi}_N^T \end{bmatrix} [\underline{\psi}_1 \ \underline{\psi}_2 \ldots \underline{\psi}_N] \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{bmatrix} = \underline{\alpha}^T \mathbf{\Phi}^T \mathbf{\Psi} \underline{\beta}$$

hence

$$1 = \begin{matrix} [\alpha_1 \ \alpha_2 \ldots \alpha_N] \end{matrix} \begin{bmatrix} \underline{\phi}_1^T\underline{\psi}_1 & \underline{\phi}_1^T\underline{\psi}_2 & \cdots & \underline{\phi}_1^T\underline{\psi}_N \\ \underline{\phi}_2^T\underline{\psi}_1 & \underline{\phi}_2^T\underline{\psi}_2 & \cdots & \underline{\phi}_2^T\underline{\psi}_N \\ \cdot & \cdot & & \cdot \\ \underline{\phi}_N^T\underline{\psi}_1 & \underline{\phi}_N^T\underline{\psi}_2 & \cdots & \underline{\phi}_N^T\underline{\psi}_N \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{bmatrix} = \sum_{i=1}^N \sum_{j=1}^N \alpha_i < \underline{\phi}_i, \underline{\psi}_j > \beta_j \qquad (1)$$

Writing

$$M = Sup \ \{ \ |\langle \underline{\phi}_i, \underline{\psi}_j \rangle|, \forall(i,j) \ \} = Sup \ \{ \ |\underline{\phi}_i^T\underline{\psi}_j|, \forall(i,j) \ \},$$

all the entries in the matrix $\mathbf{\Phi}^T\mathbf{\Psi}$ are smaller than M. We further have the (Parseval) energy preservation property

$$1 = \underline{S}^T\underline{S} = \sum \alpha_i^2 = \sum \beta_i^2$$

Now assume that

$$\|\underline{\alpha}\|_0 = A \quad \text{and} \quad \|\underline{\beta}\|_0 = B$$

Then Equation (1) becomes

$$1 = \sum_{i'=1}^A \sum_{j'=1}^B \alpha_{i'} \langle \underline{\phi}_{i'}, \underline{\psi}_{j'} \rangle \beta_{j'},$$

where i' runs over the nonzero support of $\underline{\alpha}$ and j' runs over the nonzero support of $\underline{\beta}$. Now we can write that

$$\begin{aligned} 1 \ &= \ \sum_{i'=1}^A \sum_{j'=1}^B \alpha_{i'} \langle \underline{\phi}_{i'}, \underline{\psi}_{j'} \rangle \beta_{j'} \\ &\leq \ \sum_{i'=1}^A \sum_{j'=1}^B |\alpha_{i'}| \ |\langle \underline{\phi}_{i'}, \underline{\psi}_{j'} \rangle| \ |\beta_{j'}| \end{aligned}$$

$$\leq \sum_{i'=1}^{A} \sum_{j'=1}^{B} |\alpha_{i'}| \cdot M \cdot |\beta_{j'}| \qquad (2)$$

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

$$= M \left(\sum_{i'=1}^{A} |\alpha_{i'}| \right) \left(\sum_{j'=1}^{B} |\beta_{j'}| \right)$$

Next, in order to bound the above expression from above let us find

$$\text{Maximize} \quad \left(\sum_{i=1}^{A} \alpha_i \right) \left(\sum_{j=1}^{B} \beta_j \right)$$

$$\text{subject to} \quad \alpha_i, \beta_j > 0 \ , \ \sum_{i=1}^{A} \alpha_i^2 = 1 \ , \ \sum_{j=1}^{B} \beta_j^2 = 1$$

Since this problem is separable, we maximize the following

$$\text{Maximize} \quad \left(\sum_{i=1}^{A} \alpha_i \right) \quad \text{subject to} \quad \alpha_i > 0 \quad \sum_{i=1}^{A} \alpha_i^2 = 1$$

$$\text{Maximize} \quad \left(\sum_{j=1}^{B} \beta_j \right) \quad \text{subject to} \quad \beta_j > 0 \quad \sum_{j=1}^{B} \beta_j^2 = 1$$

To solve these problems let us consider the following Lagrangian (note that we do not explicitly enforce the positivity constraint, but verify its correctness at the end of the process)

$$L(\underline{\alpha}) = \sum_{i=1}^{A} \alpha_i + \lambda \left[1 - \sum_{i=1}^{A} \alpha_i^2 \right]$$

$$\frac{\partial L}{\partial \alpha_i} = 1 - 2 \cdot \lambda \cdot \alpha_i = 0 \quad \rightarrow \quad \alpha_i = \frac{1}{2\lambda}$$

Next, using the unit-norm constraint we get

$$\sum_{i=1}^{A} \alpha_i^2 = \sum_{i=1}^{A} \left(\frac{1}{2\lambda} \right)^2 = \frac{A}{4\lambda^2} = 1 \quad \rightarrow \quad \lambda = \frac{\sqrt{A}}{2} \quad \rightarrow \quad \alpha_i = \frac{1}{\sqrt{A}}$$

Therefore the maximal value of $\sum_{i'=1}^{A} \alpha_{i'}$ is

$$\sum_{i=1}^{A} \alpha_i = \frac{A}{\sqrt{A}} = \sqrt{A}$$

and the maximal value of $\sum_{i'=1}^{A} \alpha_{i'} \sum_{j'=1}^{B} \beta_{j'}$ is

$$\frac{1}{\sqrt{A}} \cdot \frac{1}{\sqrt{B}} \cdot (A \cdot B) = \sqrt{AB}.$$

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

$$\sqrt{AB} \geq \frac{1}{M}$$

where $M = \text{Sup}\{|\langle \phi_i, \psi_j \rangle|, \quad \forall (i,j)\}$. 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 eq. (7.1) of [2]). Clearly $\frac{2}{M} = \frac{1}{M} + \frac{1}{M} \geq 1 + \frac{1}{M}$, since $M \leq 1$. In [2] it is said that their "... *general bound (i.e. (7.1)) can be a factor of two away from sharpness w.r.t. those (earlier particular, i.e. sinusoidal and spike bases) cases. Its generality can be an advantage in some cases*".

We see here that an elementary derivation provides tight bounds that are not by a factor of two away from sharpness. Indeed, as $M \to \frac{1}{\sqrt{N}}$ (as it happens for spikes and complex sinusoids) $\left(1 + \frac{1}{M}\right)$ goes to $1 + \sqrt{N}$ while $\frac{2}{M}$ goes to $2\sqrt{N}$, which is the bound claimed for this particular case in [2] (see e.g. (2.1).

In fact, the uncertainty result that we have obtained is even stronger since we got $\sqrt{AB} \geq \frac{1}{M}$. The use of the arithmetic mean instead of the geometric one looses tightness and is correct only for $A = B$. Also, this uncertainty result is of the form of the classical multiplicative uncertainty results (see e.g. [3]) $\sigma_t^2 \cdot \sigma_\omega^2 \geq \frac{1}{4}$.

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 $M \geq \frac{1}{\sqrt{N}}$. To see this, one simply notices that $\mathbf{\Phi}^T \mathbf{\Psi}$ is an orthonormal matrix, having the sum of squares of its entries equal to N. All entries cannot therefore be less than $\frac{1}{\sqrt{N}}$ since then we would have that (the sum of all squared entries) $\leq N^2$.

6

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 a dictionary that is the concatenation of two orthonormal bases $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$, 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 $\underline{S}$ in the dictionary $[\boldsymbol{\Phi}\ \boldsymbol{\Psi}]$, i.e.

$$\underline{S} = [\boldsymbol{\Phi}\ \boldsymbol{\Psi}]\underline{\gamma}_1 = [\boldsymbol{\Phi}\ \boldsymbol{\Psi}]\underline{\gamma}_2$$

Then clearly

$$[\boldsymbol{\Phi}\ \boldsymbol{\Psi}][\underline{\gamma}_1 - \underline{\gamma}_2] = \underline{0}$$

or

$$\boldsymbol{\Phi} \cdot [\underline{\gamma}_1 - \underline{\gamma}_2]^\phi + \boldsymbol{\Psi} \cdot [\underline{\gamma}_1 - \underline{\gamma}_2]^\psi = \underline{0} \quad \text{or}$$
$$\boldsymbol{\Phi}\underline{\gamma}^\phi_\Delta = -\boldsymbol{\Psi}\underline{\gamma}^\psi_\Delta = \underline{\Lambda}$$

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

$$\|\underline{\gamma}^\phi_\Delta\|_0 = A \quad \text{and} \quad \|\underline{\gamma}^\psi_\Delta\|_0 = B$$

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. that

$$\|\underline{\gamma}_1\|_0 < F \quad \text{and} \quad \|\underline{\gamma}_2\|_0 < F.$$

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 $[\boldsymbol{\Phi}\ \boldsymbol{\Psi}]$ so that

7

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

then this sparse representation is necessarily unique, (i.e., there cannot be two different $\underline{\gamma}$'s obeying $\|\underline{\gamma_0}\| < M^{-1}$ that represent the same signal).

The bound $1/M$ is a better bound than the one implied by the uncertainty principle stated in [2], which would yield $\frac{1}{2}(1 + M^{-1})$. 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   Finding 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 $[\boldsymbol{\Phi} \ \boldsymbol{\Psi}]$, how should we find it? Solving the l_0 optimization problem defined as

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

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

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

often provides the sparse representation. Donoho and Huo proved in [2] that the strong sparsity condition $\|\underline{\gamma}\|_0 < \frac{1}{2}(1 + M^{-1})$ 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

$$\|\underline{\gamma}\|_0 < F \quad \text{and} \quad [\boldsymbol{\Phi} \ \boldsymbol{\Psi}]\underline{\gamma} = \underline{S}$$

then, if there exists some other representation $[\boldsymbol{\Phi} \ \boldsymbol{\Psi}]\tilde{\underline{\gamma}} = \underline{S}$, we must have

$$\|\tilde{\underline{\gamma}}\|_1 \geq \|\underline{\gamma}\|_1.$$

In words, we need to show that the (unique) sparse $\underline{\gamma}$ is also "shortest" in the l_1 - metric. As was said above, Donoho and Huo have shown that for $F = \frac{1}{2}(1 + M^{-1})$ this requirement is met. We first describe below the derivation of this result, then, in the next section we show how this bound can be improved.

Following [2] we have that if:

$$[\boldsymbol{\Phi} \ \boldsymbol{\Psi}]\underline{\gamma} = \underline{S} = [\boldsymbol{\Phi} \ \boldsymbol{\Psi}]\tilde{\underline{\gamma}}$$

then

$$[\boldsymbol{\Phi} \ \boldsymbol{\Psi}][\tilde{\underline{\gamma}} - \underline{\gamma}] = 0.$$

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

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

8

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

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

In order to simplify notations in following analysis, we define

$$\sum_{all} |x_k| = \sum_{k=1}^{2N} |x_k|$$

$$\sum_{off} |x_k| = \sum_{\text{off support of}\gamma} |x_k|$$

$$\sum_{on} |x_k| = \sum_{\text{on support of}\gamma} |x_k|.$$

Since (due to $|v + m| \geq |v| - |m|$) we have

$$|v + m| - |v| \geq |v| - |m| - |v| = -|m|$$

we can write that

$$\sum_{\text{off}} |x_k| + \sum_{\text{on}} (|\gamma_k + x_k| - |\gamma_k|) \geq \sum_{\text{off}} |x_k| - \sum_{\text{on}} |x_k|$$

Plugging this into the above inequality we get

$$\sum_{\text{off}} |x_k| - \sum_{\text{on}} |x_k| \geq 0$$

We can simplify this inequality by the following step

$$\sum_{\text{off}} |x_k| + \sum_{\text{on}} |x_k| \geq 2 \cdot \sum_{\text{on}} |x_k|$$

which finally gives

$$\frac{1}{2} \sum_{\text{all}} |x_k| - \sum_{\text{on}} |x_k| \geq 0 \quad \text{or equivalently} \quad \frac{\sum \text{on } |x_k|}{\sum \text{all}_{|x_k|}} \leq \frac{1}{2}. \tag{4}$$

Therefore, if we shall show that for some value of the support size of $\underline{\gamma}$ (F), all $\underline{x}$ obeying (3) also satisfy the inequalities in (4), then the assertion $\|\tilde{\underline{\gamma}}\|_1 > \|\underline{\gamma}\|_1$ follows, and (P_1)-problem can replace the (P_0)-problem while searching for the sparse $\underline{\gamma}$.

To prove Equation (4) above consider the problem:

$$\text{Minimize} \quad \sum_{i=1}^{N} |x_i^{\phi}| + \sum_{i=1}^{N} |x_i^{\psi}| \quad \text{subject to} \quad \{ \; x_k^{\phi/\psi} = V \quad \text{and} \quad \mathbf{\Phi}[\underline{x}^{\phi}] = \mathbf{\Psi}[-\underline{x}^{\psi}] \; \}$$

If the minimum attained for all conditions $x_k^{\phi/\psi} = V$ is given by some $f(V) \geq |V|\widetilde{F}(M)$ then we will be able to say that

$$\frac{\sum\limits_{\text{on}} |x_i|}{\sum\limits_{\text{all}} |x_j|} = \sum_{\text{on}} \frac{|x_i|}{\sum\limits_{\text{all}} |x_j|} \leq \sum_{\text{on}} \frac{|V|}{|V| \cdot \widetilde{F}(M)} = \frac{1}{\widetilde{F}(M)} \cdot \|\underline{\gamma}\|_0$$

But we have assumed that $\|\underline{\gamma}\|_0 < F$ hence we shall have

$$\frac{\sum\limits_{\text{on}} |x_i|}{\sum\limits_{\text{all i}} |x_i|} \leq \frac{F}{\widetilde{F}(M)}.$$

If we shall have $F \leq \frac{1}{2}\widetilde{F}(M)$ then the required condition for the optimality of $\underline{\gamma}$ for the l_1 problem will be met. We shall next prove that $\widetilde{F}(M) \geq 1 + \frac{1}{M}$ and it will lead to the result that if $F \leq \frac{1}{2}(1 + \frac{1}{M})$ the ratio

$$\frac{\sum\limits_{\text{on}} |x_i|}{\sum\limits_{\text{all}} |x_i|} \leq \frac{1}{2}$$

as needed.

Since we have $\mathbf{\Phi}\underline{x}^{\phi} = -\mathbf{\Psi}\underline{x}^{\psi}$ we can write that $\underline{x}^{\phi} = -\mathbf{\Phi}^T\mathbf{\Psi}\underline{x}^{\psi}$. Suppose w.l.o.g. that the condition we have is that $x_k^{\phi} = V$. Then we have

$$\begin{aligned}
\|\underline{x}^{\phi}\|_{\infty} &= \max_{j} \|x_j^{\phi}\| = |x_{(v)}^{\phi}|_{\text{for some index v}} \\
&= \left| -[\mathbf{\Phi}^T\mathbf{\Psi}]_{\text{row v}} \cdot \underline{x}^{\psi} \right| \leq M \cdot \|\underline{x}^{\psi}\|_1
\end{aligned}$$

Hence

$$\|\underline{x}^{\psi}\|_1 \geq \frac{\|\underline{x}^{\phi}\|_{\infty}}{M} \geq \frac{|V|}{M}.$$

We also obviously have

$$\|\underline{x}^{\phi}\|_1 \geq |V|,$$

hence

$$\|\underline{x}^{\psi}\|_1 + \|\underline{x}^{\phi}\|_1 \geq |V|\left(1 + \frac{1}{M}\right)$$

Hence we may take

$$\widetilde{F}(M) = \left(1 + \frac{1}{M}\right)$$

This proves that if

$$\|\underline{\gamma}\|_0 < \frac{1}{2}\left(1 + \frac{1}{M}\right)$$

the very sparse representation will be provided by the solution of (P_1) too.

So far we have seen that if

$$\|\underline{\gamma}\|_0 < \frac{1}{M}$$

the sparse representation is unique and if

$$\|\underline{\gamma}\|_0 < \frac{1}{2}\left(1 + \frac{1}{M}\right)$$

the unique sparse representation is provided by solving a linear programming problem (P_1). Hence if

$$\frac{1}{2}\left(1 + \frac{1}{M}\right) \le \|\underline{\gamma}\|_0 < \frac{1}{M}$$

we have uniqueness of the (P_0) solution, but it is not sure that (P_1) will provide this solution. So, obviously there remains a gap, and the natural question is what happens for vectors of sparsity in this gap. A wealth of simulations we have done shown that for signals with (unique) sparse representation in this gap the equivalence between the (P_0) and the (P_1) solutions remains true. Motivated by these empirical results, we succeeded to prove that the gap is indeed much narrower.

5   Improving the Bound for l_1 Optimization

In the previous section we closely followed Donoho and Huo's work and reproduced their results. Next we take a different route in order to obtain an improved bound necessary for equivalence between (P_0) and (P_1). Equations (3) and (4) can re-interpreted as an optimization problem of the following form

$$\text{Minimize} \quad \left[\frac{1}{2}\sum_{\text{all}} |x_k| - \sum_{\text{on}} |x_k|\right] \quad \text{subject to} \quad \boldsymbol{\Phi}\underline{x}^\phi = \boldsymbol{\Psi}(-\underline{x}^\psi).$$

This problem should be solved for various values of F ($F = \|\underline{\gamma}\|_0$) and all profiles of non-zero entries in $\underline{\gamma}$. As F increases the minimum of the above expression decreases and tends to a negative value. 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). However, working with the above minimization problem is complicated because of several reasons:

1. We have not explicitly set conditions on $\underline{x}$ to avoid the trivial solution $\underline{x} = \underline{0}$,

2. The problem involves both the entries of the vector $\underline{x}$ and their absolute values,

3. The orthonormal matrices $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$ appear explicitly, and we would like them to influence the result only through the parameter M they define,

4. The problem is clearly sensitive not only to the number of non-zero elements in the support of $\underline{\gamma}$, but also to their position.

In order to solve the first problem, we introduce an additional constraint that prevents the trivial solution. Such a constraint could be posed on the l_2 or the l_1 norm of the unknown vector $\underline{x}$, i.e., $\underline{x}^T\underline{x} = 1$ or $\sum_{\text{all}} |x_k| = 1$. The newly added constraint will clearly not change the sign of the result, i.e., if for some F the minimization result becomes negative, then it would have been so without the constraint as well. Thus, the new constraint does not interfere with our goals in this optimization problem.

As to the other problems, we solve them 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

$$\text{Minimize} \quad \left[\frac{1}{2}\sum_{\text{all}} |x_k| - \sum_{\text{on}} |x_k|\right] \quad \text{subject to:} \quad \begin{array}{l} 1. \ |x^\phi| \leq M \cdot \mathbf{1}_{N\times N}|x^\psi| \\[2mm] 2. \ |x^\psi| \leq M \cdot \mathbf{1}_{N\times N}|x^\phi| \\[2mm] 3. \ \sum_{\text{all}} |x_i| = 1 \end{array}$$

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 further simplify it by noticing first that the sum over all the support that appears in the function is known to be 1 due to the third constraint. Also, only the absolute values of the unknown vector $\underline{x}$ play a role here. Exploiting these two properties we can rewrite the the new problem and get

$$\text{Minimize} \quad \left[\frac{1}{2} - \mathbf{1}_{\gamma_1}^T\underline{X}_1 - \mathbf{1}_{\gamma_2}^T\underline{X}_2\right] \quad \text{subject to:} \quad \begin{array}{l} 1. \ \underline{X}_1 \leq M \cdot \mathbf{1}_{N\times N}\underline{X}_2 \\[2mm] 2. \ \underline{X}_2 \leq M \cdot \mathbf{1}_{N\times N}\underline{X}_1 \\[2mm] 3. \ \mathbf{1}_N^T(\underline{X}_1 + \underline{X}_2) = 1 \\[2mm] 4. \ \underline{X}_1 \geq 0, \quad \underline{X}_2 \geq 0 \end{array} \quad (5)$$

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

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

(P denotes the fact that this is a primal LP form). The matrices $(A, \underline{Z}, \underline{B}, \underline{C})$ involved are defined as follows

$$\begin{aligned}
\underline{Z}^T &= \begin{bmatrix} \underline{X}_1^T & \underline{X}_1^T \end{bmatrix} \\
\underline{C}^T &= \begin{bmatrix} -\mathbf{1}_{\gamma_1}^T & -\mathbf{1}_{\gamma_2}^T \end{bmatrix} \\
\underline{B}^T &= \begin{bmatrix} 0 \cdot \mathbf{1}_N^T & 0 \cdot \mathbf{1}_N^T & 1 & -1 \end{bmatrix} \\
\mathbf{A} &= \begin{bmatrix} -\mathbf{I}_N & M \cdot \mathbf{1}_{N \times N} \\ M \cdot \mathbf{1}_{N \times N} & -\mathbf{I}_N \\ \mathbf{1}_N^T & \mathbf{1}_N^T \\ -\mathbf{1}_N^T & -\mathbf{1}_N^T \end{bmatrix}
\end{aligned}$$

Notice that the equality constraint in the original problem is replaced by two inequalities, in order to fit the required canonic form. 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 $\underline{C}^T \underline{Z} \geq -0.5$.

It is still difficult to give an analytic form to the solution of the LP problem we obtained. Instead, we shall exploit the dual LP problem of the form

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

(D here denotes that this is a dual LP form), 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.,

$$\text{Minimize} \ \{\underline{C}^T \underline{Z}\} = \text{Maximize} \ \{\underline{B}^T \underline{U}\}.$$

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 $\underline{U}$

$$\underline{U}^T = \begin{bmatrix} \mathbf{1}_{\gamma_1}^T & \alpha \mathbf{1}_{\gamma_2}^T & \beta & \gamma \end{bmatrix}.$$

13

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 we get

$$-\mathbf{1}_{\gamma_1} + \alpha M \cdot \mathbf{1}_{N \times N} \mathbf{1}_{\gamma_2} + (\beta - \gamma) \mathbf{1}_N \leq -\mathbf{1}_{\gamma_1}$$
$$-\alpha \mathbf{1}_{\gamma_2} + M \cdot \mathbf{1}_{N \times N} \mathbf{1}_{\gamma_1} + (\beta - \gamma) \mathbf{1}_N \leq -\mathbf{1}_{\gamma_2}$$

and rearranging these inequalities we get

$$\alpha M \cdot \mathbf{1}_{N \times N} \mathbf{1}_{\gamma_2} + (\beta - \gamma) \mathbf{1}_N \leq 0$$
$$(1 - \alpha) \mathbf{1}_{\gamma_2} + M \cdot \mathbf{1}_{N \times N} \mathbf{1}_{\gamma_1} + (\beta - \gamma) \mathbf{1}_N \leq 0.$$

The multiplication $\mathbf{1}_{N \times N} \mathbf{1}_{\gamma_2}$ is simply $K_2 \mathbf{1}_N$, and similarly, $\mathbf{1}_{N \times N} \mathbf{1}_{\gamma_1} = K_1 \mathbf{1}_N$. Thus, the first set of N inequalities is actually a simple scalar requirement

$$\alpha M K_2 + (\beta - \gamma) \leq 0$$

If $0 \leq \alpha \leq 1$ as assumed before, the second set of inequalities can be replaced by the following, scalar requirement

$$(1 - \alpha) + M K_1 + (\beta - \gamma) \leq 0$$

(For the $N - K_2$ places where $\mathbf{1}_{\gamma_2}$ has zeros, we get the inequality $M K_1 + (\beta - \gamma) \leq 0$, which is weaker than the above one).

Solving the two inequalities as equalities, we get

$$\alpha M K_2 = (1 - \alpha) + M K_1 \quad \Rightarrow \alpha = \frac{1 + M K_1}{1 + M K_2}$$

We see that, indeed, our assumption $0 \leq \alpha \leq 1$ is correct since we assumed $K_1 \leq K_2$, (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

$$(\beta - \gamma) = -M K_2 \cdot \frac{1 + M K_1}{1 + M K_2} < 0.$$

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

$$\underline{B}^T \underline{U} = (\beta - \gamma) = -M K_2 \cdot \frac{1 + M K_1}{1 + M K_2} \geq -\frac{1}{2}.$$

14

Thus

$$MK_2 \cdot \frac{1 + MK_1}{1 + MK_2} \leq \frac{1}{2} \quad \Rightarrow \quad 2M^2 K_1 K_2 + MK_2 - 1 \leq 0 \tag{6}$$

We have therefore obtained a requirement on K_1 and K_2 which is posed in terms of the parameter M. Thus, given a vector $\underline{S} \in \Re^N$, the two orthonormal bases $\mathbf{\Phi}$ and $\mathbf{\Psi}$, and their induced cross-product factor M, we solve the (P_1)-problem and obtain a candidate representation $\underline{\gamma}$. From this representation we can determine K_1 and K_2, the number of non-zeros of γ's two N-sections. Plugging them into the inequality (6) we immediately know whether this is also the (P_0)-solution as well, and hence also the sparsest representation.

We now proceed and produce a simpler sparsity condition on the representation $\underline{\gamma}$ ensuring that (P_1) produces the unique and sparse (P_0) solution. Assuming that we know K_2, the requirement on K_1 is

$$2M^2 K_1 K_2 + MK_2 - 1 \leq 0 \quad \Rightarrow \quad K_1 \leq \frac{1}{M} \cdot \frac{1 - MK_2}{2MK_2}.$$

Adding K_2 to both sides of the above inequality we get

$$K_1 + K_2 \leq \frac{1}{M} \cdot \frac{1 - MK_2}{2MK_2} + K_2 = \frac{1}{M} \cdot \frac{1 - MK_2 + 2\left(MK_2\right)^2}{2MK_2}.$$

Let us look at the term which multiplies the $1/M$ in the above bound. This term is a function of MK_2 which is known to be non-negative. The minimum of this term is given by

$$\begin{aligned}
f(x) &= \frac{1 - x + 2x^2}{2x} \\
\Rightarrow \quad \frac{df(x)}{dx} &= \frac{2x \cdot (4x - 1) - 2\left(1 - x + 2x^2\right)}{4x^2} = \frac{4x^2 - 2}{4x^2} = 0 \\
x_{1,2} &= \pm\sqrt{0.5}.
\end{aligned}$$

The negative solution is irrelevant, and the positive one is indeed the minimum (as easily verified with the second derivative), where we get that $f(\sqrt{0.5}) = \sqrt{2} - 0.5 = 0.9142$. Thus, our final result is the requirement

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

To summarize, we got that if $\|\underline{\gamma}\|_0 \leq (\sqrt{2} - 0.5)/M$ then the dual LP necessarily gives a value above -0.5, which in turn assures that the minimum of the primal problem is above -0.5 as well. Thus, it is guaranteed that the l_1 norm of the solution with this number of non-zeros is the smallest possible, so we have proved that solving the P_1 yields the solution of P_0 too.

The result obtained is better than the $0.5\left(1 + 1/M\right)$-bound asserted by Donoho and Huo. As an example, for $M = 1/\sqrt{N}$, we get that for $N = 16$ the old (DH) requirement is to have less than 2.5 non-zeros, while we (EB) require 3.65 and below. For $N = 64$ the old requirement (DH)

15

is 4.5 and below, while the new bound requires 7.3 or less non-zeros. As N goes to infinity the ratio between the two results becomes

$$\frac{\frac{(\sqrt{2}-0.5)}{M}}{0.5\left(1+\frac{1}{M}\right)} = \frac{(\sqrt{2}-0.5)\sqrt{N}}{0.5\left(1+\sqrt{N}\right)} \rightarrow 2\sqrt{2}-1 = 1.8284.$$

One question that remains at the moment unanswered is whether the $(\sqrt{2}-0.5)$ coefficient we got reflects true behavior of the P_0 problem versus the P_1 one, or is it emerging due to the still loose bounding approximation in our proof. For example, if we take the trivial (spike) and Hadamard bases for $N = 128$, we get that $1/M = 1/\sqrt{128} = 1/11.31$. Hence the uniqueness bound requires $\|\underline{\gamma}\|_0 \leq 1/M = 11.31$. On the other hand, the new bound found here, $(\sqrt{2}-0.5)/M$, requires less than 10.34 non-zero entries, which in practice means 10 and below in order to ensure that the P_1- and the P_0-problems coincide. An experiment that would test this case with 11 non-zero entries and produce a counter-example were P_0 and P_1 lead to different solutions could be beneficial. The search for such an example currently under way but so far we could not find any case like this.

It is interesting to note that if we assume that $K_1 = 0$, the requirement made in Equation (6) leads to the requirement $K_2 = \|\underline{\gamma}\|_0 \leq \frac{1}{M}$. Similarly, if $K_2 = K_1$ then Equation (6) leads to the requirement $2K_2 = \|\underline{\gamma}\|_0 \leq \frac{1}{M}$. Note that these two cases are the extremes in the range of K_1 ($0 \leq K_1 \leq K_2$). Thus, for these two cases the $1/M$ bound is valid and there is no gap with respect to the uniqueness bound.

Returning to the suggested search for a counter-example for the 11 non-zeros described above, it is clear now that choosing the non-zeros in one of the above two extremes will not lead to a successful hunt. Instead, one has to choose $K_2 \approx \frac{\sqrt{0.5}}{M} = 8$ and $K_1 = 11 - 8 = 3$ in order to maximize the chances for a success. Again, simulations done so far did not find such a special case.

Another important question is how much of the entire range of signals in $\Re^N$ are we covering by the sparsity condition we have formed? What is the "measure" of signals that require more than $(\sqrt{2}-0.5)/M$ none-zeros in their sparse representation? These and more questions will be the subject of future investigations.

6 Conclusions

Given a vector $\underline{S} \in \Re^N$, and given an orthonormal N by N matrix $\boldsymbol{\Phi}$ representing a basis, we can uniquely represent $\underline{S}$ using this basis and get $\underline{\alpha} = \boldsymbol{\Phi}^T \underline{S}$. Assuming that we have a second orthonormal basis $\boldsymbol{\Psi}$ and a second representation $\underline{\beta} = \boldsymbol{\Psi}^T \underline{S}$, the first question we have addressed in this paper is whether there is a lower bound on the number of non-zeros in these two representations. It was found that

$$\frac{\|\alpha\|_0 + \|\beta\|_0}{2} \geq \sqrt{\|\alpha\|_0 \cdot \|\beta\|_0} \geq \frac{1}{M},$$

where M is a scalar value denoting the maximal absolute value of cross-inner-products between vectors in $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$.

The second question answered in this paper refers to the uniqueness of sparse representation with overcomplete dictionaries. Assuming that the signal $\underline{S}$ is to be represented using a dictionary that contains $\mathbf{\Phi}$ and $\mathbf{\Psi}$ ($2N$ vectors), it is clear that there are numerous possible representations. It was established that for any two feasible representations denoted by $\underline{\gamma}_1$, $\underline{\gamma}_2 \in \Re^{2N}$, we have that

$$\|\underline{\gamma}_1\|_0 + \|\underline{\gamma}_2\|_0 \geq \frac{2}{M}.$$

Thus, there cannot be two different representations for the same vector having each lower than $1/M$ non-zero entries, i.e., for any given representation $\underline{\gamma}$, we have that uniqueness is ensured by

$$\|\underline{\gamma}\|_0 \geq \frac{1}{M}.$$

The main contribution of the paper concentrated to the way to find the sparse representation over an overcomplete dictionary as described above. Finding the optimal representation by minimizing the l_0-norm is a highly complicated non-convex combinatorial optimization search. An alternative approach based on l_1-norm minimization was proposed by Donoho and Huo and proved to lead to the same result for sufficiently sparse signals. Their result was that if the signal has a sparse representation with no more than $0.5\,(1 + 1/M)$ non-zero entries, minimization of the l_1-norm can replace the original l_0-norm minimization while ensuring the same result. Obviously, such a result is very valuable since l_1-norm minimization leads to a linear programming problem, which is a convex problem, while the l_0-norm minimization is a hard non-convex problem. In this paper we have improved this sparsity bound and found that if $\underline{S}$ has a sparse representation $\underline{\gamma}$ such that

$$\|\underline{\gamma}\|_0 \geq \frac{(\sqrt{2} - 0.5)}{M} = \frac{0.9142}{M},$$

then the l_1-norm minimization solution coincides with the minimization of the l_0-norm.

In fact, the bound obtained in this paper is better than $\frac{(\sqrt{2}-0.5)}{M}$, since we obtained the following requirement in terms of K_1 and K_2 (the number of non-zeros in the two N-sections of the representations)

$$2M^2 \cdot K_1 K_2 + M K_2 - 1 \leq 0.$$

Figure 1 shows graphically how this expression compares to the alternatives $K_1 + K_2 \leq \frac{(\sqrt{2}-0.5)}{M}$ and $K_1 + K_2 \leq \frac{1}{M}$. In this graph we have assumed $M = 1/\sqrt{128}$. Note that since $0 \leq K_1 \leq K_2$, 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.

Figure 2 graphically describes how the various bounds developed by Donoho and Huo (and thus denoted by DH) are compared to the new bounds denoted here (denoted EB).
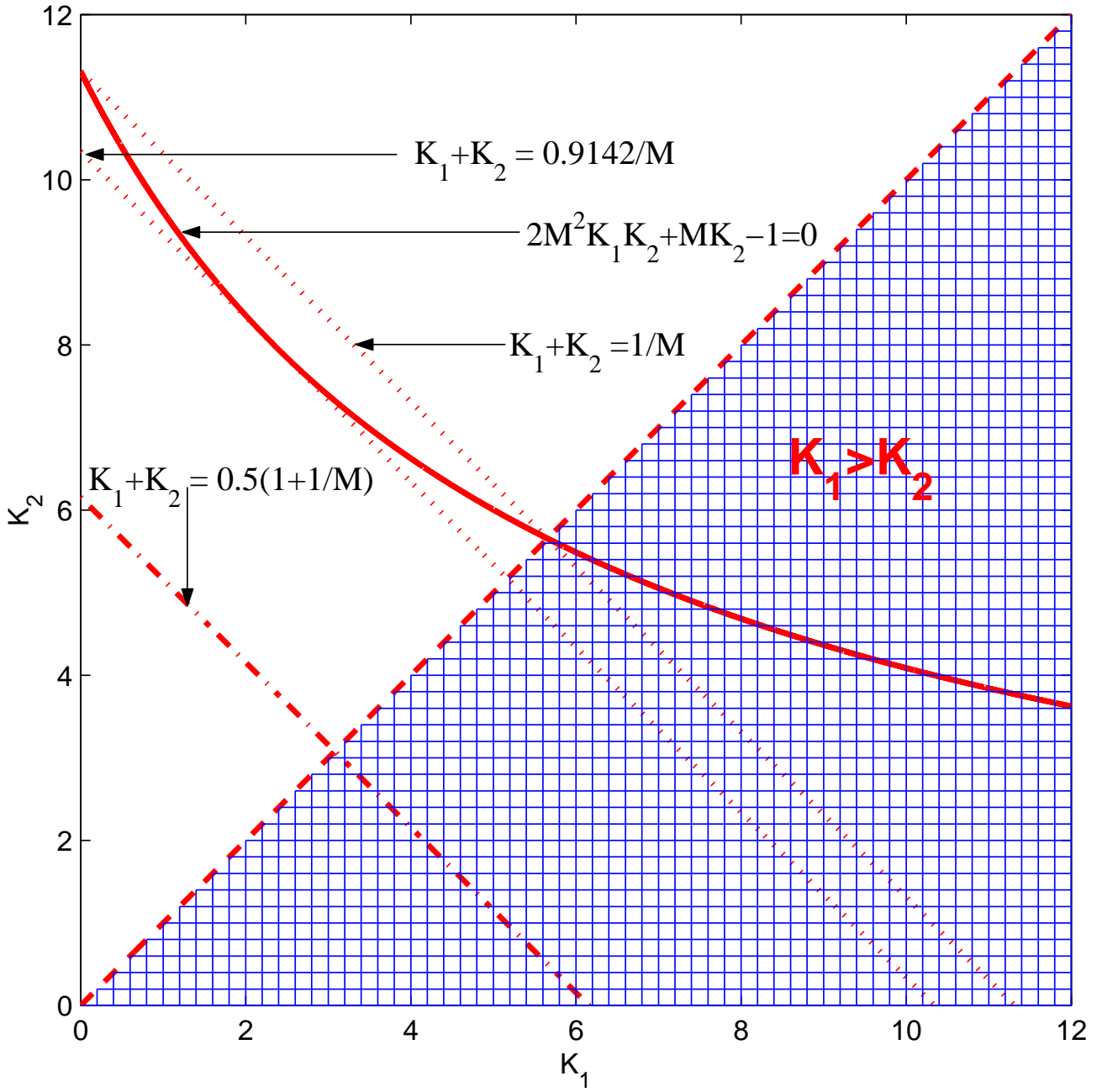
7 Acknowledgements

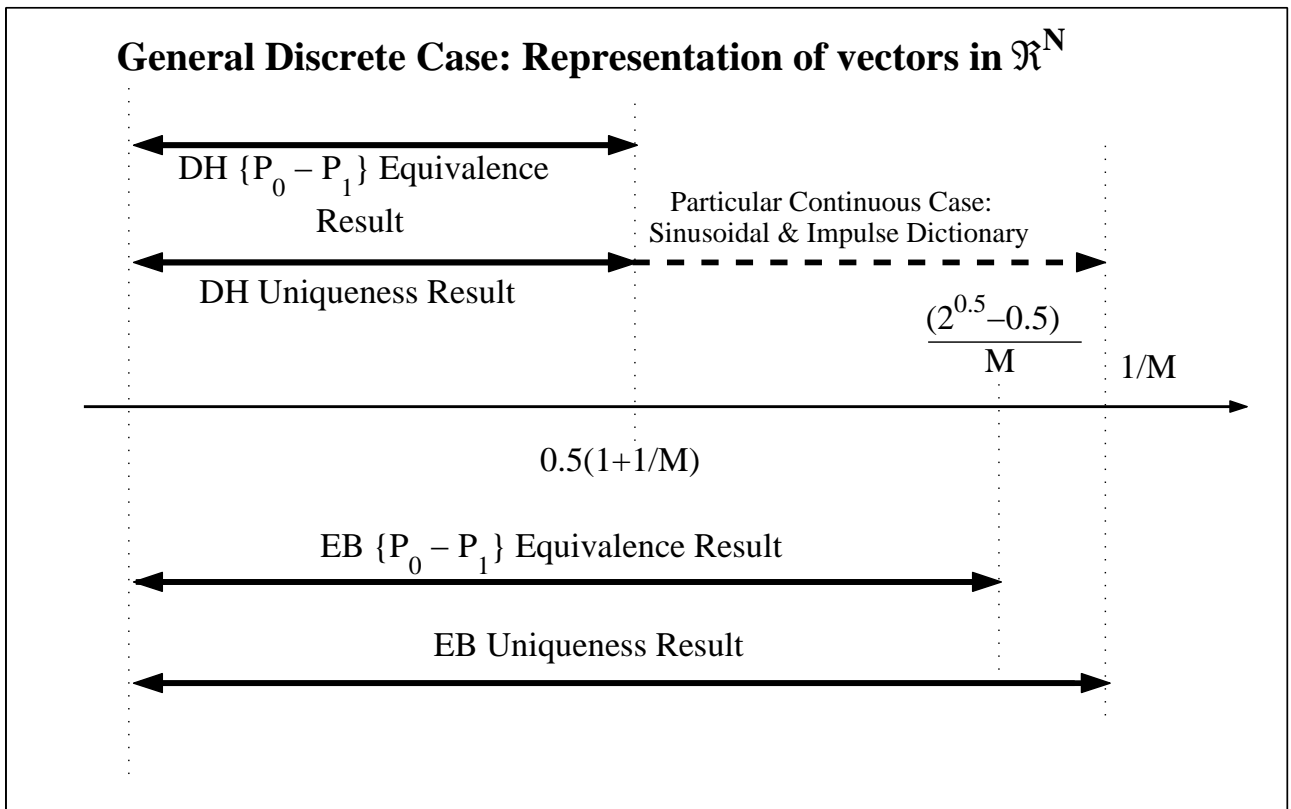Figure 1: Figure 1 - A graphic description of the bounds obtained.

Figure 2: Figure 2 - A graphic illustration of the bounds due to Donoho and Huo (DH) versus the ones developed in this paper (EB).

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