On Sparse Representation in Pairs of Bases

Arie Feuer, *Senior Member, IEEE,* and Arkadi Nemirovski

***Abstract*—In previous work, Elad and Bruckstein (EB) have provided a sufficient condition for replacing an l_0 optimization by linear programming minimization when searching for the unique sparse representation. We establish here that the EB condition is both sufficient and necessary.**

***Index Terms*—Dictionary, sparse representation, tight frame.**

I. Introduction

In their recent publication, [1], Elad and Bruckstein (EB) address the following problem:

Given two orthogonal matrices $\Phi, \Psi \in \mathbb{R}^{N\times N}$ and a vector $\underline{S} \in \mathbb{R}^N$ consider the following two optimization problems:

(P_0) Minimize

$$\|\underline{\gamma}\|_0 = \sum_{k=1}^{2N} \gamma_k^0$$

subject to

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

and

(P_1) Minimize

$$\|\underline{\gamma}\|_1 = \sum_{k=1}^{2N} |\gamma_k|$$

subject to

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

where by $\|\underline{\gamma}\|_0$ we refer to the number of nonzero entries of vector $\underline{\gamma}$.

The columns of each matrix Φ, Ψ constitute an orthonormal basis in $\mathbb{R}^N$. When viewed together as a set of $2N$ vectors they are referred to in [1] as a dictionary (we also note that this set of vectors is in fact a tight frame with frame bound 2). The vector $\underline{\gamma} \in \mathbb{R}^{2N}$ is a representation of vector (signal) $\underline{S}$ in this dictionary. A question of obvious interest is what can be gained by representing $\underline{S}$ in the dictionary rather than in either basis. The answer lies in the possibility of getting sparser representations where the measure of sparseness is the norm $\|\cdot\|_0$ (which just counts the nonzero entries of the vector). Clearly, one can always find $\underline{\gamma}$ so that $\|\underline{\gamma}\|_0 \leq \min\{\|\underline{\alpha}\|_0, \|\underline{\beta}\|_0\}$ where $\underline{S} = \Phi\underline{\alpha} = \Psi\underline{\beta}$ (note that $\left[\begin{smallmatrix}\underline{\alpha}\\0\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}0\\\underline{\beta}\end{smallmatrix}\right]$ are also representations of the signal in the dictionary). Hence, problem (P_0) is in fact an attempt to find the sparsest representation of a given vector (signal) in a particular dictionary—this is of much interest in, e.g., signal compression problems.

In [1], the following result on the uniqueness of this representation is given.

Theorem 1 [1, Theorem 2]: Let

$$M = \max_{1\leq i, j\leq N} \left(\left|\langle \underline{\phi}_i, \underline{\psi}_j\rangle\right|\right) \tag{1}$$

where $\{\underline{\phi}_i\}, \{\underline{\psi}_j\}$ are the columns of Φ and Ψ, respectively. If the signal $\underline{S}$ is to be represented using the concatenated dictionary $[\Phi, \Psi]$ ($2N$ vectors), for any two feasible distinct representations denoted by $\underline{\gamma}_1, \underline{\gamma}_2 \in \mathbb{R}^{2N}$, we have that

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

This means that the uniqueness of a sparse representation is ensured by

$$\|\underline{\gamma}\|_0 < \frac{1}{M}. \tag{3}$$

Clearly, condition (3) guarantees the uniqueness of the solution of problem (P_0). However, finding this representation, namely, solving (P_0) is a problem we do not know how to solve efficiently. As pointed out in [2], under certain conditions, solving (P_0) can be traded with solving (P_1) a much easier problem. The sufficient condition found by Donoho and Huo is that if

$$\|\underline{\gamma}\|_0 < \frac{1}{2}\left(1 + \frac{1}{M}\right) \tag{4}$$

then (P_0) and (P_1) are equivalent problems in the sense that the (unique) solution of (P_1) is also the (unique) solution of (P_0).

In [1], this condition is relaxed and a higher bound is given, again, as a sufficient condition

$$\|\underline{\gamma}\|_0 < \frac{\sqrt{2} - 0.5}{M}. \tag{5}$$

We are going to show that the condition in (5) is in fact both necessary and sufficient. So, this bound for the equivalency of problems (P_0) and (P_1) is tight.

II. Main Result

In this section, we are going to show that the bound presented in [1] is tight. To do that we first prove the following proposition.

Proposition 2: Let $L_0 < \frac{1}{M}$ be a given integer and

$$\Theta_{L_0}(\underline{x}) = \frac{S_{L_0}(\underline{x})}{\|\underline{x}\|_1} \tag{6}$$

where $S_{L_0}(\underline{x})$ is the sum of the L_0 largest absolute values of the entries of $\underline{x} \in \mathbb{R}^{2N}$. Then, (P_0) and (P_1) are equivalent if and only if

$$\sup_{[\Phi, \Psi]\underline{x}=0} \Theta_{L_0}(\underline{x}) < \frac{1}{2}. \tag{7}$$

Proof: Note that in the proof here we use Theorem 1 implicitly, claiming that since $L_0 < \frac{1}{M}$, $\underline{\gamma}$ is indeed the unique solution of (P_0) for $\underline{S} = [\Phi, \Psi]\,\underline{\gamma}$.

Suppose (7) holds. Then for any $\underline{\gamma}, \tilde{\underline{\gamma}} \in \mathbb{R}^{2N}$ such that $\|\underline{\gamma}\|_0 = L_0$ and $[\Phi, \Psi]\,\underline{\gamma} = [\Phi, \Psi]\,\tilde{\underline{\gamma}}$ we have

$$2S_{L_0}(\underline{\gamma} - \tilde{\underline{\gamma}}) < \|\underline{\gamma} - \tilde{\underline{\gamma}}\|_1. \tag{8}$$

Manuscript received June 23, 2002; revised January 21, 2003.

A. Feuer is with the Control and Robotics Laboratory, Department of Electrical Engineering, Technion–Israel Institute of Technology, 32000 Technion City, Haifa, Israel (e-mail: feuer@ee.technion.ac.il).

A. Nemirovsky is with the Department of Industrial Engineering and Management, Technion–Israel Institute of Technology, 32000 Technion City, Haifa, Israel (e-mail: nemitovs@ie.technion.ac.il).

Communicated by G. Battail, Associate Editor at Large.

Digital Object Identifier 10.1109/TIT.2003.811926

Let j_i, $i = 1, \ldots, 2N$, be a permutation of the indexes $1, \ldots, 2N$ such that $j_1, \ldots, j_{L_0}$ are the indexes of the L_0 entries of $\underline{\gamma}$ with the largest absolute values. Then, from (8) we have

$$2\sum_{i=1}^{L_0} |\gamma_{j_i} - \tilde{\gamma}_{j_i}| < \sum_{i=1}^{L_0} |\gamma_{j_i} - \tilde{\gamma}_{j_i}| + \sum_{i=L_0+1}^{2N} |\tilde{\gamma}_{j_i}|$$

$$\sum_{i=1}^{L_0} |\gamma_{j_i} - \tilde{\gamma}_{j_i}| < \sum_{i=L_0+1}^{2N} |\tilde{\gamma}_{j_i}|$$

$$\sum_{i=1}^{L_0} (|\gamma_{j_i} - \tilde{\gamma}_{j_i}| + |\tilde{\gamma}_{j_i}|) < \sum_{i=1}^{L_0} |\tilde{\gamma}_{j_i}| + \sum_{i=L_0+1}^{2N} |\tilde{\gamma}_{j_i}|$$

and since

$$|\gamma_{j_i}| < |\gamma_{j_i} - \tilde{\gamma}_{j_i}| + |\tilde{\gamma}_{j_i}|$$

we get

$$\sum_{i=1}^{L_0} |\gamma_{j_i}| < \sum_{i=1}^{2N} |\tilde{\gamma}_{j_i}|$$

namely,

$$\left\|\underline{\gamma}\right\|_1 < \left\|\underline{\tilde{\gamma}}\right\|_1$$

so that (P_0) and (P_1) are equivalent.

Suppose (7) does not hold for some $\underline{x} \in \mathbb{R}^{2N}$ such that $[\Phi, \Psi]\,\underline{x} = 0$, namely, $2S_{L_0}(\underline{x}) \geq \|\underline{x}\|_1$. Let us use, again, the permuted indexes $\{j_i\}_{i=1}^{2N}$ so that this time $j_1, \ldots, j_{L_0}$ are the indexes of the L_0 entries of $\underline{x}$ with the largest absolute values. Then choose $\underline{\gamma}$ so that

$$\gamma_{j_i} = \begin{cases} -2x_{j_i}, & \text{for } 1 \leq i \leq L_0 \\ 0, & \text{otherwise} \end{cases}$$

and $\underline{\tilde{\gamma}} = \underline{\gamma} + \underline{x}$. Then, clearly, $\left\|\underline{\gamma}\right\|_1 = 2S_{L_0}(\underline{x})$ and $\left\|\underline{\tilde{\gamma}}\right\|_1 = \|\underline{x}\|_1$, so that

$$\left\|\underline{\gamma}\right\|_1 \geq \left\|\underline{\tilde{\gamma}}\right\|_1$$

and since $\underline{\gamma} \neq \underline{\tilde{\gamma}}$, clearly, (P_0) and (P_1) are not equivalent, which completes the proof. □

The next proposition is basically a reworded version of a similar result proven in [1]. However, we believe the proof here is somewhat more straightforward and provides the motivation for the main contribution of this note—a family of counterexamples which show that the bound in [1] is indeed tight.

Proposition 3: Let $L_0 < \frac{1}{M}$ be a given integer. Then, for all $\underline{x} \in \mathbb{R}^{2N}$ such that $[\Phi, \Psi]\,\underline{x} = 0$, we have

$$\Theta_{L_0}(\underline{x}) \leq \left(\sqrt{2 + ML_0} - 1\right)^2 \tag{9}$$

where $\Theta_{L_0}(\underline{x})$ is as in (6).

Proof: Since $[\Phi, \Psi]\,\underline{x} = 0$, $\underline{x}$ must be of the form

$$\underline{x} = \begin{bmatrix} I \\ A \end{bmatrix} \underline{v}$$

for some $\underline{v} \in \mathbb{R}^N$, where $A = -\Psi^T\Phi$ is also an orthogonal matrix. Let us now fix $\underline{v}$ and let us denote $\underline{w} = A\underline{v}$. Swapping, if necessary, $\underline{v}$ and $\underline{w}$, we may assume without loss of generality that among the L_0 largest (in absolute value) entries of $\underline{x}$, $m \geq \frac{L_0}{2}$ are the first m entries of $\underline{v}$, and the remaining $L_0 - m$ are entries of $\underline{w}$. Since, by its definition, $\Theta_{L_0}(\alpha\underline{x}) = \Theta_{L_0}(\underline{x})$ for all $\alpha > 0$, we may assume that

$$\sum_{i=1}^{m} |v_i| = 1. \tag{10}$$

Let us denote

$$P = \sum_{i=m+1}^{N} |v_i|.$$

Observe that

$$\begin{aligned} \|\underline{w}\|_\infty &= \|A\underline{v}\|_\infty \\ &\leq \max_{1\leq i,\, j\leq N} |A_{i,j}|\, \|\underline{v}\|_1 = M\, \|\underline{v}\|_1 \\ &\leq M(1+P). \end{aligned} \tag{11}$$

Similarly, we can conclude that

$$\|\underline{v}\|_\infty \leq M\, \|\underline{w}\|_1. \tag{12}$$

Since we assumed that the first m entries in $\underline{v}$ are the largest, (10) implies that $\|\underline{v}\|_\infty \geq \frac{1}{m}$ and it follows from (12) that

$$\|\underline{w}\|_1 \geq \frac{1}{mM}. \tag{13}$$

Denoting $\theta = mM$, we have from (11) and (13)

$$\begin{aligned} \Theta_{L_0}(\underline{x}) &= \frac{\sum_{i=1}^{m} |v_i| + S_{L_0-m}(\underline{w})}{\|\underline{v}\|_1 + \|\underline{w}\|_1} \\ &\leq \frac{1 + (L_0 - m)\, \|\underline{w}\|_\infty}{\|\underline{v}\|_1 + \|\underline{w}\|_1} \\ &\leq \frac{1 + M(L_0 - m)(1+P)}{1 + P + \frac{1}{mM}} \\ &\leq \frac{1 + (ML_0 - \theta)(1+P)}{1 + P + \theta^{-1}}. \end{aligned} \tag{14}$$

Denoting the right-hand side of (14) by $f(\theta, P)$, we note that for $\frac{ML_0}{2} \leq \theta \leq ML_0$

$$\frac{\partial f(\theta, P)}{\partial P} = \frac{\theta^{-1}ML_0 - 2}{(1 + P + \theta^{-1})^2} \leq 0$$

so

$$\begin{aligned} f(\theta, P) &\leq \max_{\frac{ML_0}{2}\leq\theta\leq ML_0} \{f(\theta, 0)\} \\ &\leq \max_{\frac{ML_0}{2}\leq\theta\leq ML_0} \left\{\frac{1 + ML_0 - \theta}{1 + \theta^{-1}}\right\} \\ &= \left(\sqrt{2 + ML_0} - 1\right)^2. \end{aligned}$$

Hence,

$$\Theta_{L_0}(\underline{x}) \leq \left(\sqrt{2 + ML_0} - 1\right)^2.$$

□

Clearly, if $ML_0 < \sqrt{2} - 0.5$ then by this proposition

$$\Theta_{L_0}(\underline{x}) < \left(\sqrt{2 + \sqrt{2} - 0.5} - 1\right)^2 = \frac{1}{2}.$$

Hence, by Proposition 2, (P_0) and (P_1) are equivalent. This is a restatement of a result in [1].

Next we prove the following proposition.

Proposition 4: For every positive integer r, there exist orthogonal matrices $\Phi = [\phi_1 \cdots \phi_{2^{2r-1}}]$ and $\Psi = [\psi_1 \cdots \psi_{2^{2r-1}}]$ such that $\left|\phi_i^T \psi_j\right| \leq M \triangleq 2^{-r+0.5}$ and a vector $\underline{x} \neq 0$ satisfying $[\Phi, \Psi]\,\underline{x} = 0$ so that $\Theta_{L_0}(\underline{x}) \geq \frac{1}{2}$ for all integers L_0 for which $\sqrt{2} - 0.5 \leq ML_0 \leq 1$.

Proof: We will prove the proposition by construction.

Let $r > 0$ be any integer and we denote

$$\begin{aligned} m &= 2^r \\ k &= 2^{2r-1} \\ M &= \frac{1}{\sqrt{k}} \\ L_{\max} &= \frac{1}{M} \\ L_{\min} &= \frac{\sqrt{2} - 0.5}{M}. \end{aligned} \tag{15}$$

Furthermore, let

$$F^{(1)} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{16}$$

and

$$\begin{aligned} F^{(s)} &= \underbrace{F^{(1)} \otimes \cdots \otimes F^{(1)}}_{s\,\text{times}} \\ &= F^{(1)} \otimes F^{(s-1)} \end{aligned} \tag{17}$$

where $\otimes$ denotes the Kronecker product (these matrices are known as the Hadamard matrices).

Now we choose

$$U = \frac{1}{\sqrt{k}} F^{(2r-1)} \tag{18}$$

and

$$\begin{aligned} \underline{v} &= \begin{bmatrix} \underline{1}_m \\ \underline{0}_{k-m} \end{bmatrix} \\ &= \underline{e}_1^{(\frac{k}{m})} \otimes \underline{1}_m \end{aligned} \tag{19}$$

where $\underline{1}_m$ and $\underline{0}_{k-m}$ are m- and $k-m$-dimensional vectors of ones and zeros, respectively, $\underline{e}_1^{(\frac{k}{m})}$ is the first column of the $\frac{k}{m} = 2^{r-1}$-dimensional identity vector. Then

$$\begin{aligned} U\underline{v} &= \frac{1}{\sqrt{k}} \left(F^{(\frac{k}{m})} \otimes F^{(m)}\right)\left(\underline{e}_1^{(\frac{k}{m})} \otimes \underline{1}_m\right) \\ &= \frac{1}{\sqrt{k}} \left(F^{(\frac{k}{m})} \underline{e}_1^{(\frac{k}{m})}\right) \otimes \left(F^{(m)} \underline{1}_m\right) \\ &= \frac{m}{\sqrt{k}} \underline{1}_{(\frac{k}{m})} \otimes \underline{e}_1^{(m)}. \end{aligned} \tag{20}$$

Define

$$\underline{x} = \begin{bmatrix} \underline{v} \\ U\underline{v} \end{bmatrix} \tag{21}$$

then we observe from (19)–(21) that $\underline{x}$ has a total of $m + \frac{k}{m}$ nonzero entries, m of which are equal to 1 and the remaining $\frac{k}{m} = 2^{r-1}$ are equal to $\frac{m}{\sqrt{k}} = \sqrt{2}$.

Hence, for any $L_{\min} \leq L_0 \leq L_{\max}$ (noting that $L_{\min} > 2^{r-1}$ and $L_{\max} < 2^r + 2^{r-1}$) we get

$$\begin{aligned} \Theta_{L_0}(\underline{x}) &= \frac{S_{L_0}(\underline{x})}{\|\underline{x}\|_1} \\ &= \frac{\frac{k}{m}\frac{m}{\sqrt{k}} + \left(L_0 - \frac{k}{m}\right)}{\frac{k}{m}\frac{m}{\sqrt{k}} + m} \\ &\geq \frac{2^{r-0.5} - 2^{r-1} + L_{\min}}{2^{r-0.5} + 2^r} \\ &= \frac{1}{2}. \end{aligned}$$

Namely, $\Phi = I$, $\Psi = -U$, and $\underline{x}$ as defined in (21) constitute the required example which completes the proof of the proposition. □

We can combine now the consequences of Propositions 3 and 4 and state our main result.

Theorem 5: (P_0) and (P_1) are guaranteed to be equivalent for any given two orthogonal matrices Φ, Ψ, and signal $\underline{S} = [\Phi, \Psi]\,\underline{\gamma}$ if and only if $M\left\|\underline{\gamma}\right\|_0 \leq \sqrt{2} - 0.5$.

III. Conclusion

Given a signal $\underline{S} \in \mathbb{R}^N$ one is interested in the sparsest representation of this signal in a given dictionary which consists of Φ and Ψ, two orthonormal bases in $\mathbb{R}^N$. This is an l_0-optimization problem which is very difficult to solve. An alternative approach was proposed in [2], where it was shown that if the signal representation has no more than $0.5(1 + 1/M)$ nonzero entries it can be found using an l_1-optimization approach which leads to linear programming methods and is much simpler to solve. This bound was improved in [1] to become $\left(\sqrt{2} - 0.5\right)/M > 0.5(1 + 1/M)$.

However, the question whether the bound is tight was left unanswered. We have established here that the Elad and Bruckstein bound is indeed tight.

References

[1] M. Elad and A. M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Trans. Inform. Theory*, to be published.

[2] D. L. Donoho and X. Huo, "Uncertainty principle and ideal atomic decomposition," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2845–2862, July 2001.