rates taken so high that further increasing them produced no visible changes in the figure. As can be seen, the $\gamma_{a,b}$ obtained in that way turns from a well-behaved function for the values $b = 1.5, 2.5$ into a quite irregularly behaved one when b approaches 2 or 3.

REFERENCES

[1] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.

[2] K. Gröchenig, *Foundations of Time-Frequency Analysis*. Boston, MA: Birkhäuser, 2001.

[3] A. Ron and Z. Shen, "Weyl-Heisenberg frames and Riesz bases in $L_2(\mathbf{R}^d)$," *Duke Math. J.*, vol. 89, no. 2, pp. 237–282, 1997.

[4] A. J. E. M. Janssen, "Zak transforms with few zeros and the tie," in *Advances in Gabor Analysis*, H. G. Feichtinger and T. Strohmer, Eds. Boston, MA: Birkhäuser, 2003, pp. 31–70.

[5] Y. I. Lyubarskiĭ, "Frames in the Bargmann space of entire functions," in *Entire and Subharmonic Functions*. Providence, RI: Amer. Math. Soc., 1992, pp. 167–180.

[6] K. Seip, K. Seip, and R. Wallstén, "Density theorems for sampling and interpolation in the Bargmann-Fock space I; II," *J. Reine Angew. Math.*, vol. 429, pp. 91–106, 1992.

[7] A. J. E. M. Janssen and T. Strohmer, "Hyperbolic secants yield Gabor frames," *Appl. Comput. Harmon. Anal.*, vol. 12, pp. 259–267, 2002.

[8] A. J. E. M. Janssen, "On generating tight Gabor frames at critical density," *J. Fourier Anal. Appl.*, vol. 9, no. 2, pp. 175–214, 2003.

[9] H. G. Feichtinger and N. Kaiblinger, "Varying the time-frequency lattice of Gabor frames," *Trans. Amer. Math. Soc.*, to be published.

[10] A. Cavaretta, W. Dahmen, and C. A. Micchelli, "Stationary Subdivision," *Mem. Amer. Math. Soc.*, vol. 53, no. 453, pp. 1–186, 1991.

[11] A. J. E. M. Janssen, "The duality condition for Weyl-Heisenberg frames," in *Gabor Analysis and Algorithms*, H. G. Feichtinger and T. Strohmer, Eds. Boston, MA: Birkhäuser, 1998, pp. 33–84.

[12] ——, "From continuous to discrete Weyl-Heisenberg frames through sampling," *J. Fourier Anal. Appl.*, vol. 3, pp. 583–597, 1997.

[13] N. Kaiblinger, "Approximation of the Fourier transform and the Gabor dual function from samples," preprint, 2003.

Sparse Representations in Unions of Bases

Rémi Gribonval, *Member, IEEE*, and Morten Nielsen

***Abstract*—The purpose of this correspondence is to generalize a result by Donoho and Huo and Elad and Bruckstein on sparse representations of signals in a union of two orthonormal bases for $\mathbb{R}^N$. We consider general (redundant) dictionaries for $\mathbb{R}^N$, and derive sufficient conditions for having unique sparse representations of signals in such dictionaries. The special case where the dictionary is given by the union of $L \geq 2$ orthonormal bases for $\mathbb{R}^N$ is studied in more detail. In particular, it is proved that the result of Donoho and Huo, concerning the replacement of the ℓ^0 optimization problem with a linear programming problem when searching for sparse representations, has an analog for dictionaries that may be highly redundant.**

***Index Terms*—Dictionaries, Grassmannian frames, linear programming, mutually incoherent bases, nonlinear approximation, sparse representations.**

I. INTRODUCTION

We consider vectors (also referred to as signals) in $\mathcal{H} = \mathbb{R}^N$ (resp., $\mathcal{H} = \mathbb{C}^N$). The goal is to find an efficient representation of a signal $s \in \mathcal{H}$. One well-known way to do this is to take an orthonormal basis $\Phi = \{\phi_1, \ldots, \phi_N\}$ for $\mathcal{H}$ and use the Fourier coefficients $\{\langle s, \phi_k \rangle\}_{k=1}^N$ to represent s. This approach is simple and works reasonably well in many cases. However, one can also consider a more general type of expansion where the orthonormal basis is replaced by a so-called dictionary for $\mathcal{H}$.

Definition 1: A **dictionary** in $\mathcal{H} = \mathbb{R}^N$ (resp., $\mathcal{H} = \mathbb{C}^N$) is a family of $K \geq N$ unit (column) vectors $\{g_k\}$ that spans $\mathcal{H}$. We will use the matrix notation $\boldsymbol{D} = [g_1, \ldots, g_K]$ for a dictionary.

By a representation of s in $\boldsymbol{D}$ we mean a (column) vector $\alpha = (\alpha_k) \in \mathbb{R}^K$ (resp., in $\mathbb{C}^K$) such that $s = \boldsymbol{D}\alpha$. We notice that when $K > N$, the vectors of $\boldsymbol{D}$ are no longer linearly independent and the representation of s is not unique. The hope is that among all possible representations of s there is a *very sparse* representation, i.e., a representation with few nonzero coefficients. The tradeoff is that we have to *search* all possible representations of s to find the sparse representations, and then determine whether there is a unique sparsest representation. Following [1] and [2], we will measure the sparsity of a representation $s = \boldsymbol{D}\alpha$ by two quantities: the ℓ^0 and the ℓ^1 norm of α, respectively (the ℓ^0-norm simply counts the number of nonzero entries of a vector). This leads to the following two minimization problems to determine the sparsest representation of s:

$$\text{minimize} \quad \|\alpha\|_0 \quad \text{subject to } s = \boldsymbol{D}\alpha \tag{1}$$

and

$$\text{minimize} \quad \|\alpha\|_1 \quad \text{subject to } s = \boldsymbol{D}\alpha. \tag{2}$$

It turns out that the optimization problem (2) is much easier to handle than (1) through the use of linear programming (LP), so it is important to know the relationship between the solution(s) of (1) and (2), and to determine sufficient conditions for the two problems to have the same unique solution. This problem has been studied in detail in [1] and later has been refined in [2] in the special case where the dictionary $\boldsymbol{D}$ is the union of *two* orthonormal bases. In what follows, we generalize the results of [1] and [2] to arbitrary dictionaries.[1] The case where $\boldsymbol{D}$ is the union of $L \geq 2$ orthonormal bases for $\mathcal{H}$ is studied in detail. This leads to a natural generalization of the recent results from [2] valid for $L = 2$.

In Section II, we provide conditions for a solution α of the problem

$$\text{minimize} \quad \|\alpha\|_\tau \quad \text{subject to } s = \boldsymbol{D}\alpha \tag{3}$$

to be indeed the unique solution, with $0 \leq \tau \leq 1$ and an arbitrary dictionary $\boldsymbol{D}$. We put a special emphasis on sufficient conditions of the type $\|\alpha\|_0 < f(\boldsymbol{D})$ and prove a sufficient condition for $\tau \in \{0, 1\}$ with $f(\boldsymbol{D}) = (1 + 1/M(\boldsymbol{D}))/2$ where

$$M(\boldsymbol{D}) := \max_{k \neq k'} |\langle g_k, g_{k'} \rangle| \tag{4}$$

is the **coherence** of the dictionary. The special case where $\boldsymbol{D}$ is the union of $L \geq 2$ bases is studied in Section III, leading to explicit sufficient conditions for $\tau = 0$ with

$$f(\boldsymbol{D}) = \left(1/2 + \frac{1}{2(L-1)}\right) / M(\boldsymbol{D})$$

and for $\tau \in \{0, 1\}$ with

$$f(\boldsymbol{D}) = \left(\sqrt{2} - 1 + \frac{1}{2(L-1)}\right) / M(\boldsymbol{D}).$$

[1] Parallel work done independently by Donoho and Elad [3] also addresses the question of generalizing previous results to general dictionaries. Though there are some similarities between this work to the work in [3], a somewhat different perspective on the problem is adopted and the proofs use different techniques.

Manuscript received November 13, 2002; revised August 4, 2003.

R. Gribonval is with IRISA-INRIA, Campus de Beaulieu, F-35042 Rennes Cedex, France (remi.gribonval@inria.fr).

M. Nielsen is with the Department of Mathematical Sciences, Aalborg University, DK-9220 Aalborg East, Denmark (mnielsen@math.auc.dk).

Communicated by G. Battail, Associate Editor At Large.

Digital Object Identifier 10.1109/TIT.2003.820031

In the case $L = 2$ we simply recover the main result from [2], and for $L \leq 6$ we obtain a condition that is *less restrictive* than the condition for arbitrary dictionaries. In Section IV, we construct highly redundant dictionaries where the results of the present correspondence give fairly relaxed conditions for (1) and (2) to have a unique solution.

II. Sparse ℓ^τ Representations, $0 \leq \tau \leq 1$

Any vector $s \in \mathcal{H}$ has (possibly many) representations $s = \boldsymbol{D}\alpha$ with coefficient vector $\alpha \in \mathbb{R}^K$ (resp., $\alpha \in \mathbb{C}^K$).

Definition 2: The **support** of a coefficient vector $\alpha = (\alpha_k) \in \mathbb{R}^K$ (resp., $\mathbb{C}^K$) is

$$S(\alpha) := \{k, \alpha_k \neq 0\}. \tag{5}$$

The **kernel** of the dictionary will play a special role

$$\mathrm{Ker}\,(\boldsymbol{D}) := \{x, \boldsymbol{D}x = 0\} \tag{6}$$

as well as the integer quantity (called *spark* of the dictionary in [3])

$$Z(\boldsymbol{D}) := \min_{x \in \mathrm{Ker}\,(\boldsymbol{D}), x \neq 0} \|x\|_0. \tag{7}$$

By refining ideas from [2] we have the following lemma.

Lemma 1: Let $\boldsymbol{D}$ be a (possibly redundant) dictionary and $S \subset \{1, \ldots, K\}$ a set of indexes. For $0 \leq \tau \leq 1$ define

$$P_\tau(S, \boldsymbol{D}) := \max_{x \in \mathrm{Ker}\,(\boldsymbol{D}), x \neq 0} \frac{\sum\limits_{k \in S} |x_k|^\tau}{\sum\limits_{k} |x_k|^\tau} \tag{8}$$

where we use the convention $0^0 = 0$ and $x^0 = 1, x \neq 0$.

1) If $P_\tau(S, \boldsymbol{D}) < 1/2$ then, for all α such that $S(\alpha) \subset S$, α is *the unique* solution to the problem (3) with $s := \boldsymbol{D}\alpha$.
2) If $P_\tau(S, \boldsymbol{D}) = 1/2$ then, for all α such that $S(\alpha) \subset S$, α is *a* solution to the problem (3) with $s := \boldsymbol{D}\alpha$.
3) If $P_\tau(S, \boldsymbol{D}) > 1/2$ there exists α such that $S(\alpha) \subset S$ and β such that $\|\beta\|_\tau < \|\alpha\|_\tau$ and $\boldsymbol{D}\alpha = \boldsymbol{D}\beta$.

Proof: The lemma was used without being stated explicitly in [1] and [2], in the special case $\tau = 1$ and with $\boldsymbol{D}$ a union of two orthonormal bases. The proof follows the same steps as in [1] and [2]. Under the assumption $P_\tau(S, \boldsymbol{D}) < 1/2$ and $S(\alpha) \subset S$, what we need to prove is that for all $x \in \mathrm{Ker}(\boldsymbol{D})$,

$$\sum_k |\alpha_k + x_k|^\tau > \sum_k |\alpha_k|^\tau .$$

This is equivalent to showing

$$\sum_{k \notin S} |x_k|^\tau + \sum_{k \in S} \left(|\alpha_k + x_k|^\tau - |\alpha_k|^\tau \right) > 0.$$

For $0 \leq \tau \leq 1$, we have the quasi-triangle inequality $|a+b|^\tau \leq |a|^\tau + |b|^\tau$, from which we can derive the inequality $|\beta + y|^\tau - |\beta|^\tau \geq -|y|^\tau$. It is thus sufficient to prove that for all $x \in \mathrm{Ker}\,(\boldsymbol{D})$

$$\sum_{k \notin S} |x_k|^\tau - \sum_{k \in S} |x_k|^\tau > 0$$

or equivalently

$$\sum_{k \in S} |x_k|^\tau < \frac{1}{2} \sum_k |x_k|^\tau .$$

But this is exactly the assumption $P_\tau(S, \boldsymbol{D}) < 1/2$. To prove the result for $P_\tau(S, \boldsymbol{D}) = 1/2$ we copy the above line of arguments and simply replace strict inequalities with large ones. To prove the result for $P_\tau(S, \boldsymbol{D}) > 1/2$, it is sufficient to take some $x \in \mathrm{Ker}\,(\boldsymbol{D})$ so that

$$\sum_{k \in S} |x_k|^\tau > \left(\sum_k |x_k|^\tau \right) /2$$

and to consider for $k \in S$, $\alpha_k := -x_k$, $\beta_k := 0$ and for $k \notin S$, $\alpha_k := 0$, $\beta_k := x_k$. Because $x = \beta - \alpha \in \mathrm{Ker}\,(\boldsymbol{D})$ one easily checks that $\boldsymbol{D}\beta = \boldsymbol{D}\alpha$. Obviously

$$\sum_k |\beta_k|^\tau = \sum_{k \notin S} |x_k|^\tau < \sum_{k \in S} |x_k|^\tau = \sum_k |\alpha_k|^\tau$$

and $S(\alpha) \subset S$. □

Lemma 1 will be most useful to look for sufficient conditions on S that ensure uniqueness of the sparsest ℓ^τ expansion, i.e., conditions such that $P_\tau(S, \boldsymbol{D}) < 1/2$. Of particular interest are sufficient conditions that take the form

$$\text{if } \mathrm{card}\,(S) < f(\boldsymbol{D}), \text{ then } P_\tau(S, \boldsymbol{D}) < 1/2 \tag{9}$$

which correspond to uniqueness results of the form

$$\text{if } \|\alpha\|_0 < f(\boldsymbol{D}), \text{ then } \alpha \text{ is the unique solution to } (3).$$

The following lemma shows that such conditions are intimately related to the spark $Z(\boldsymbol{D})$. We denote by $\lceil x \rceil$ the smallest integer not smaller than x, i.e., $\lceil x \rceil - 1 < x \leq \lceil x \rceil$.

Lemma 2:

1) If $\|\alpha\|_0 < Z(\boldsymbol{D})/2$, then α is the unique solution to (1). In other words, (9) holds for $\tau = 0$ with $f(\boldsymbol{D}) = Z(\boldsymbol{D})/2$.
2) If (9) holds true for some $0 \leq \tau \leq 1$ with some $f(\boldsymbol{D})$, then it holds true with the same constant $f(\boldsymbol{D})$ for $\tau = 0$.
3) If (9) holds true for $\tau = 0$ with some $f(\boldsymbol{D})$, then $f(\boldsymbol{D}) \leq \lceil Z(\boldsymbol{D})/2 \rceil$.

Proof: For any S we observe that

$$P_0(S, \boldsymbol{D}) \leq \max_{x \in \mathrm{Ker}(\boldsymbol{D}), x \neq 0} \frac{\mathrm{card}(S)}{\|x\|_0} \leq \frac{\mathrm{card}(S)}{Z}.$$

The first statement immediately follows.

The second statement is almost trivial. Assuming that condition (9) holds true with τ and f, we know that when $\|\alpha\|_0 < f$, for all $\beta \neq \alpha$ such that $\boldsymbol{D}\beta = \boldsymbol{D}\alpha$ we have $\|\beta\|_\tau > \|\alpha\|_\tau$. Assume β satisfies $\boldsymbol{D}\beta = \boldsymbol{D}\alpha$ and $\|\beta\|_0 \leq \|\alpha\|_0$: then in particular $\|\beta\|_0 < f$ so β is *also* the unique minimizer of the same ℓ^τ problem, hence, $\beta = \alpha$. It follows that α is indeed the unique minimizer of the ℓ^0 problem.

To conclude, let us prove the third statement. By definition, there exists $x \in \mathrm{Ker}\,(\boldsymbol{D})$ such that $\|x\|_0 = Z$. We can split its support $S(x)$ into two disjoint sets S_1 and S_2 of same cardinality $Z/2 = \lceil Z/2 \rceil$ (if Z is even) or with

$$\mathrm{card}\,(S_1) = (Z-1)/2 \quad \text{and} \quad \mathrm{card}\,(S_2) = (Z+1)/2 = \lceil Z/2 \rceil$$

(if Z is odd). Obviously, $P_0(S_2, \boldsymbol{D}) \geq 1/2$, hence (9) cannot hold true for $\tau = 0$ with $f > \mathrm{card}\,(S_2) = \lceil Z/2 \rceil$. □

There are two consequences of this lemma. The first one is that we need to estimate $Z(\boldsymbol{D})$. The second is that if we are able to prove that (9) holds for some $0 \leq \tau \leq 1$ with some constant $f(\boldsymbol{D})$, then it will also hold for $\tau = 0$ with the same constant. This fact will be extensively used to find sufficient conditions so that a solution to the ℓ^1 problem also solves uniquely the ℓ^0 problem.

In [1] and [2], the case of $\boldsymbol{D} = [\boldsymbol{B}_1, \boldsymbol{B}_2]$ was considered where $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ are two orthonormal matrices corresponding to orthonormal bases. Donoho and Huo proved an *uncertainty principle*

$Z([\boldsymbol{B}_1, \boldsymbol{B}_2]) \geq 1 + 1/M$ [1, Theorem VII.3] and obtained the sufficient condition (9) for $\tau \in \{0, 1\}$ with

$$f([\boldsymbol{B}_1, \boldsymbol{B}_2]) = \frac{1}{2}(1 + 1/M(\boldsymbol{D})). \tag{10}$$

Elad and Bruckstein improved the uncertainty principle by getting [2, Theorem 1] $Z([\boldsymbol{B}_1, \boldsymbol{B}_2]) \geq 2/M$. Thus, they obtained the sufficient condition (9) for $\tau = 0$ with the less restrictive constant

$$f([\boldsymbol{B}_1, \boldsymbol{B}_2]) = 1/M(\boldsymbol{D}). \tag{11}$$

Eventually, Elad and Bruckstein used another technique to obtain condition (9) for $\tau = 1$ (and, thus, for $\tau = 0$) with

$$f([\boldsymbol{B}_1, \boldsymbol{B}_2]) = \Big(\sqrt{2} - 1/2\Big)/M(\boldsymbol{D}) \approx 0.914/M(\boldsymbol{D}). \tag{12}$$

Feuer and Nemirovsky [4] recently proved that the above constant is essentially the best one to get condition (9) for $\tau = 1$, in particular it cannot be replaced with the less restrictive constant (11).

Next we show that the result with the most restrictive of the constants, that is, (10), extends to the case of arbitrary dictionaries. In the next section, we will consider results for dictionaries built by taking the union of $L \geq 2$ orthonormal bases.

Theorem 1: For any dictionary, if

$$\|\alpha\|_0 < \frac{1}{2}(1 + 1/M(\boldsymbol{D})) \tag{13}$$

then α is the (unique) solution to both the ℓ^0 and the ℓ^1 minimization problems.

Proof: As already noticed, we will just need to show that (9) holds for $\tau = 1$ with $f := (1 + 1/M)/2$.

Consider $x \in \mathrm{Ker}\,(\boldsymbol{D})$. For every k we have

$$x_k g_k = -\sum_{k' \neq k} x_{k'} g_{k'}$$

hence, taking the inner product of both hand sides with g_k, $|x_k| \leq M(\boldsymbol{D}) \cdot \sum_{k' \neq k} |x_{k'}|$. It follows that

$$(1 + M) \cdot |x_k| \leq M \cdot \|x\|_1. \tag{14}$$

Summing over $k \in S$ we get $P_1(S, \boldsymbol{D}) \cdot \|x\|_1 \leq \frac{\mathrm{card}(S) M}{1+M} \cdot \|x\|_1$, so

$$P_1(S, \boldsymbol{D}) \leq \frac{\mathrm{card}\,(S)}{1 + 1/M} < 1/2$$

as soon as $\mathrm{card}\,(S) < (1 + 1/M)/2$. □

Note that the above line of arguments can be modified slightly to prove that for arbitrary dictionaries we have the generalized uncertainty principle

$$Z(\boldsymbol{D}) \geq 1 + 1/M(\boldsymbol{D}). \tag{15}$$

Notice that, as soon as the dictionary contains an orthonormal basis and an additional unit vector, the value of M is at least $1/\sqrt{N}$. To see that, let us assume, without loss of generality, that the orthonormal basis corresponds to the first N vectors of $\boldsymbol{D}$. By

$$\sum_{k=1}^{N} |\langle g_{N+1}, g_k \rangle|^2 = \|g_{N+1}\|^2 = 1$$

we see that $\max_{k=1}^{N} |\langle g_{N+1}, g_k \rangle|^2 \geq 1/N$, hence the inequality

$$M(\boldsymbol{D}) \geq 1/\sqrt{N}. \tag{16}$$

III. Sparse Representations in Unions of Bases

We now switch to the special case of $\boldsymbol{D}$ a union of L orthonormal bases, i.e., $\boldsymbol{D} = [\boldsymbol{B}_1, \ldots, \boldsymbol{B}_L]$ where $\boldsymbol{B}_l$ is an orthonormal matrix, $1 \leq l \leq L$. First, we concentrate on getting a result of the type (9) for $\tau = 0$. This will correspond to getting a sharper generalized uncertainty principle by getting a lower bound on $Z(\boldsymbol{D})$.

Lemma 3: Let $\boldsymbol{D}$ be a union of L orthonormal bases. Let

$$x = \begin{bmatrix} x^1 \\ \ldots \\ x^L \end{bmatrix} \in \mathrm{Ker}\,(\boldsymbol{D})$$

with $x^l \in \mathbb{R}^N$ (resp., $\mathbb{C}^N$) and assume $x \neq 0$. Then

$$\sum_{l=1}^{L} \frac{1}{1 + M(\boldsymbol{D})\|x^l\|_0} \leq L - 1. \tag{17}$$

Consequently

$$Z(\boldsymbol{D}) \geq \Big(1 + \frac{1}{L-1}\Big) \frac{1}{M(\boldsymbol{D})}. \tag{18}$$

Proof: Because $x \in \mathrm{Ker}\,(\boldsymbol{D})$, for every l we have $B_l x^l = -\sum_{l' \neq l} B_{l'} x^{l'}$, hence, $x^l = -\sum_{l' \neq l} B_l^T B_{l'} x^{l'}$. Denoting $X^l \in \mathbb{R}^N$ the vector having as entries the absolute values of those of the original vector x^l, we have for all $x \in \mathrm{Ker}\,(\boldsymbol{D})$ and $1 \leq l \leq L$

$$X^l \leq M(\boldsymbol{D}) \cdot \sum_{l' \neq l} \|X^{l'}\|_1 \cdot 1_N \tag{19}$$

where $1_N \in \mathbb{R}^N$ is a column vector with all entries equal to one. For each l, summing over the nonzero coordinates of X^l we obtain $\|X^l\|_1 \leq M\|X^l\|_0 \cdot \sum_{l' \neq l} \|X^{l'}\|_1$. It follows that

$$(1 + M\|X^l\|_0) \cdot \|X^l\|_1 \leq M\|X^l\|_0 \cdot \sum_{l'} \|X^{l'}\|_1$$

hence

$$\|x^l\|_1 \leq \frac{M\|x^l\|_0}{1 + M\|x^l\|_0} \cdot \|x\|_1.$$

Summing over l we obtain $\|x\|_1 \leq \Big(\sum_l \frac{M\|x^l\|_0}{1+M\|x^l\|_0}\Big) \cdot \|x\|_1$ from which we get

$$\sum_{l=1}^{L} \frac{M\|x^l\|_0}{1 + M\|x^l\|_0} \geq 1.$$

This is easily rewritten $\sum_{l=1}^{L} g(M\|x^l\|_0) \leq L - 1$ with $g(y) := 1/(1 + y)$ and gives (17). By the convexity of g and the fact that $M\|x\|_0 = \sum_{l=1}^{L} M\|x^l\|_0$, we have $g(M\|x\|_0/L) \leq (L - 1)/L$ hence $M\|x\|_0 \geq \frac{L}{L-1}$ and (18) follows. □

Notice that for $L = 2$ the condition (17) can be rewritten $\sqrt{\|x^1\|_0 \|x^2\|_0} \geq 1/M$ as in [2, Theorem 1]. There are examples of pairs of bases with $Z = 2/M$, so the generalized uncertainty principle (18) is sharp for $L = 2$. For $L \geq 3$ it is an open problem whether there exist examples of L orthonormal bases for which MZ is arbitrarily close to $1 + 1/(L - 1)$. Using (18) together with Lemma 2 we have the following corollary.

Corollary 1: Let $\boldsymbol{D}$ be a union of L orthonormal bases. If

$$\|\alpha\|_0 < \left(\frac{1}{2} + \frac{1}{2(L-1)}\right) \frac{1}{M(\boldsymbol{D})} \tag{20}$$

then the unique solution to the ℓ^0 problem is α.

For $L = 2$, we find again the least restrictive condition (11) of Elad and Bruckstein. As we increase the number L of bases while keeping M constant (we will see in Section IV that it is indeed possible to have up to $L = N + 1$ orthonormal bases with perfect separation $M = 1/\sqrt{N}$, for N a power of two), Condition (20) gets more and more restrictive. It is only natural that we have to pay a price for increasing the redundancy of the dictionary. For small enough values of L, (20) is less restrictive than (13). For $L \geq 1 + 1/M$, however, the bound in Corollary 1 becomes more restrictive than the general result from Theorem 1, so the latter should be used in this case.

Let us now consider the ℓ^1 minimization problem with unions of orthonormal bases. For pairs of bases, the general result of Theorem 1 was improved in [2] to get the less restrictive sufficient condition $\|\alpha\|_0 < (\sqrt{2} - 0.5)/M(\boldsymbol{D})$. The authors in [2] indeed proved a stronger result which can be stated as follows: if we denote

$$\alpha = \begin{bmatrix} \alpha^1 \\ \alpha^2 \end{bmatrix}$$

with $\alpha^l \in \mathbb{R}^N$ (resp., $\mathbb{C}^N$) and $K_l := \|\alpha^l\|_0, l = 1, 2$, then a sufficient condition to ensure $P_1(S(\alpha), \boldsymbol{D}) < 1/2$ is that

$$2M^2(\boldsymbol{D})K_1K_2 + M(\boldsymbol{D})\max(K_1, K_2) - 1 < 0. \tag{21}$$

Next we generalize this result to a union of L bases.

Theorem 2: Let $\boldsymbol{D}$ be a union of L orthonormal bases. Denote

$$\alpha = \begin{bmatrix} \alpha^1 \\ \dots \\ \alpha^L \end{bmatrix}$$

with $\alpha^l \in \mathbb{R}^N$ (resp., $\mathbb{C}^N$). Without loss of generality, we can assume that the bases $\boldsymbol{B}_l$ have been numbered so that $\|\alpha^1\|_0 \leq \cdots \leq \|\alpha^L\|_0$. If

$$\sum_{l \geq 2} \frac{M\|\alpha^l\|_0}{1 + M\|\alpha^l\|_0} < \frac{1}{2(1 + M\|\alpha^1\|_0)} \tag{22}$$

then α is the (unique) solution to the ℓ^1 minimization problem.

Proof: We follow [2, proof of Theorem 3] and start similarly to the proof of Lemma 3. Consider

$$x = \begin{bmatrix} x^1 \\ \dots \\ x^L \end{bmatrix} \in \mathrm{Ker}\,(\boldsymbol{D})$$

with $x^l \in \mathbb{R}^N$ (resp., $\mathbb{C}^N$). For every l we have

$$B_l x^l = -\sum_{l' \neq l} B_{l'} x^{l'}$$

hence,

$$x^l = -\sum_{l' \neq l} B_l^T B_{l'} x^{l'}.$$

Denoting $X^l \in \mathbb{R}^N$ the vector having as entries the absolute values of those of the original vector x^l, we have for all $x \in \mathrm{Ker}\,(\boldsymbol{D})$ and $1 \leq l \leq L$

$$X^l \leq M(\boldsymbol{D}) \sum_{l' \neq l} \mathbf{1}_{N \times N} X^{l'} \tag{23}$$

where $\mathbf{1}_{N \times N}$ is an N-by-N matrix with all entries equal to one. By definition, X^l also satisfies $X^l \geq 0$. In addition, for all $x \in \mathrm{Ker}\,(\boldsymbol{D})$ with $\|x\|_1 = \sum_{l=1}^L \|x^l\|_1 = 1$, we have

$$\sum_{l=1}^L 1_N^T X^l = 1 \tag{24}$$

and

$$\sum_{l=1}^L 1_{S(\alpha^l)}^T X^l = \sum_{k \in S(\alpha)} |x_k|, \tag{25}$$

where $1_N \in \mathbb{R}^N$ is a vector with all entries equal to one and $1_S \in \mathbb{R}^N$ is a vector with ones on the index set S and zeroes elsewhere. Thus, it is sufficient to show that under the condition (22) and the constraints (23)–(24) and $X^l \geq 0$ we have

$$\max_{X^1, \dots, X^L} \sum_{l=1}^L 1_{S(\alpha^l)}^T X^l < \frac{1}{2}.$$

Let us proceed as in [2]: by replacing the equality constraints (24) with two inequalities, we now have a classical linear programming problem, which can be put into canonical form

$$\min_{\text{primal}} := \min C^T Z \qquad \text{subject to } \boldsymbol{A}Z \geq B, Z \geq 0$$

with

$$Z = \begin{bmatrix} X^1 \\ \dots \\ X^L \end{bmatrix}$$

$$C^T = [-1_{S(\alpha^1)}, \dots, -1_{S(\alpha^L)}]$$

$$\boldsymbol{A} = \begin{bmatrix} -\boldsymbol{I}_N & M\mathbf{1}_{N\times N} & \dots & \dots & M\mathbf{1}_{N\times N} \\ M\mathbf{1}_{N\times N} & -\boldsymbol{I}_N & \ddots & & M\mathbf{1}_{N\times N} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & -\boldsymbol{I}_N & M\mathbf{1}_{N\times N} \\ M\mathbf{1}_{N\times N} & \dots & \dots & M\mathbf{1}_{N\times N} & -\boldsymbol{I}_N \\ 1_N^T & \dots & \dots & \dots & 1_N^T \\ -1_N^T & \dots & \dots & \dots & -1_N^T \end{bmatrix}$$

and

$$B^T = [0.1_N^T, \dots, 0.1_N^T, 1, -1].$$

What we need to prove is $\min_{\text{primal}} > -\frac{1}{2}$. The dual linear programming problem is

$$\max_{\text{dual}} := \max B^T U \qquad \text{subject to } \boldsymbol{A}^T U \leq C, U \geq 0$$

and we know [5] that $\max_{\text{dual}} = \min_{\text{primal}}$, so the desired result will be obtained if we can prove that there exists some $U \geq 0$ that satisfies $\boldsymbol{A}^T U \leq C$ and $B^T U > -1/2$.

We will look for such a U in a parametric form

$$U^T := [a_1 1_{S(\alpha^1)}^T, \dots, a_L 1_{S(\alpha^L)}^T, b, c]$$

with $a_l, b, c \geq 0$. Noticing that $B^T U = b - c$, the goal will be to choose a_l, b, c so that $b - c > -1/2$ and $\boldsymbol{A}^T U \leq C$.

Straightforward computations show that the condition $\boldsymbol{A}^T U \leq C$ is equivalent to the inequalities for $1 \leq l \leq L$

$$(b - c)1_N + M\mathbf{1}_{N\times N}\left(\sum_{l' \neq l} a_{l'} 1_{S(\alpha^{l'})}^T\right) + (1 - a_l)1_{S(\alpha^l)}^T \leq 0.$$

By the equality $\mathbf{1}_{N\times N} 1_S^T = \mathrm{card}\,(S) 1_N^T$ this becomes

$$\left(b - c + M\sum_{l'} a_{l'} \|\alpha^{l'}\|_0\right) 1_N^T - M a_l \|\alpha^l\|_0 1_N^T + (1 - a_l) 1_{S(\alpha^l)}^T \leq 0$$

where we used the fact that $\|\alpha^l\|_0 := \mathrm{card}\,(S(\alpha^l))$. Denoting $y^+ = \max(y, 0)$ and $y^- = \min(y, 0)$, the positive and negative parts of

$y \in \mathbb{R}$, the constraint is eventually expressed, for $1 \le l \le L$ such that $\|\alpha^l\|_0 \neq 0$, as

$$b - c + M \sum_{l'} a_{l'} \|\alpha^{l'}\|_0 \le M a_l \|\alpha^l\|_0 + (a_l - 1)^-. \tag{26}$$

The constraint for all l such that $\|\alpha^l\|_0 = 0$ (if there is any) is

$$b - c + M \sum_{l'} a_{l'} \|\alpha^{l'}\|_0 \le 0.$$

Now that the constraints have been established, let us build a_l, b, and c. We define $a_l = 1$ when $\|\alpha^l\|_0 = 0$ and use a "threshold" parameter θ to define $a_l(\theta) := \theta / \|\alpha^l\|_0$ when $1 \le \|\alpha^l\|_0 < \theta$ and $a_l(\theta) := (1 + M\theta)/(1 + M\|\alpha^l\|_0)$ when $\|\alpha^l\|_0 \ge \theta$. Let us also define $\Sigma(\theta) := \sum_l a_l(\theta)\|\alpha^l\|_0$. One can check case-by-case that for all θ, when $\|\alpha^l\|_0 \neq 0$, the constraint (26) becomes

$$b - c + M\Sigma(\theta) \le M\theta. \tag{27}$$

If there is some value of l for which $\|\alpha^l\|_0 = 0$, the associated constraint is stronger and becomes

$$b - c + M\Sigma(\theta) \le 0. \tag{28}$$

Obviously, (27) (resp., (28)) can always be satisfied by taking $b = g(\theta)^+$ and $c = -g(\theta)^-$ with $g(\theta) := M(\theta - \Sigma(\theta))$ (resp., $g(\theta) := -M\Sigma(\theta)$).

For $U = U(\theta)$ built in this parametric form, $U(\theta)$ satisfies the constraints $U \ge 0$ and $\boldsymbol{A}U \le C$, and we have $B^T U(\theta) = g(\theta)$. Thus, the problem is now whether

$$\max_\theta g(\theta) > -1/2. \tag{29}$$

Let us deal first with the case where $\|\alpha^l\|_0 \neq 0$ for all l. It is easy to check that

$$g(\|\alpha^1\|_0) = -\sum_{l \ge 2} \frac{M\|\alpha^l\|_0}{1 + M\|\alpha^l\|_0}(1 + M\|\alpha^1\|_0)$$

so the theorem is proved. Simple (but tedious) computations would show that indeed $\max_\theta g(\theta) = \max(g(0), g(\|\alpha^1\|_0))$, and that when the maximum is $g(0)$ it does not satisfy the constraint (29). So, in this case, the sufficient condition $g(\|\alpha^1\|_0) > -1/2$ is, in a sense, optimal for the type of argument we have presented.

In the case where $\|\alpha^l\|_0 = 0$ for some l (i.e., $\|\alpha^1\|_0 = 0$), we notice that $\Sigma(\theta)$ is a piecewise-linear increasing function, so $\max_\theta g(\theta) = g(0)$. Because $\|\alpha^1\|_0 = 0$ we conclude by estimating $g(0)$ as

$$-\sum_{l=1}^{L} \frac{M\|\alpha^l\|_0}{1 + M\|\alpha^l\|_0} = -\sum_{l \ge 2} \frac{M\|\alpha^l\|_0}{1 + M\|\alpha^l\|_0}(1 + M\|\alpha^1\|_0). \quad \square$$

In the case of $L = 2$ bases, (22) is exactly the condition (21) proved in [2] where it is proved that a simpler sufficient condition is $\|\alpha\|_0 \le (\sqrt{2} - 1/2)/M$. The general condition (22) is simple to check for any given α. However, in order to benefit from Lemma 2 and get a sufficient condition for α to simultaneously minimize the ℓ^0 and the ℓ^1 problems, let us look for a sufficient condition $\|\alpha\|_0 < f(\boldsymbol{D})$.

Corollary 2: For a dictionary that is the union of L orthonormal bases, if

$$\|\alpha\|_0 < \left(\sqrt{2} - 1 + \frac{1}{2(L-1)}\right)\frac{1}{M(\boldsymbol{D})} \tag{30}$$

then α is the (unique) solution to both the ℓ^0 and the ℓ^1 minimization problems. With the notations of Theorem 2, the same conclusion is reached if the above inequality is large but there exists an index $l \ge 2$ such that $(1 + M\|\alpha^l\|_0)/(1 + M\|\alpha^1\|_0) \neq \sqrt{2}$.

Proof: Denoting $y_l := M\|\alpha^l\|_0$ and $y = (y_l)_{l=1}^L$, condition (22) can be rewritten

$$g(y) := \sum_{l=2}^{L} \frac{y_l}{1 + y_l} - \frac{1}{2}\frac{1}{1 + y_1} \le 0.$$

For any $c > 0$ consider the set $H_c := \{y \mid y_l \ge 0, \sum_{l=1}^L y_l = c\}$ and let us compute

$$G(c) := \sup_{y \in H_c} g(y).$$

Using Lagrange multipliers, we know that any $y^\star$ that corresponds to an extremum of g under the constraint $\sum_l y_l = c$ will satisfy the equalities $\frac{\partial g}{\partial y_l^\star} = \lambda$, $1 \le l \le L$. For $l = 1$, this becomes $(1 + y_1^\star)^{-2}/2 = \lambda$, while for $2 \le l \le L$, this corresponds to $(1 + y_l^\star)^{-2} = \lambda$. Looking at the second partial derivatives of g we easily check that all extrema are indeed maxima. The only maximum that satisfies the additional constraint $y_l^\star \ge 0$ is given by

$$\begin{aligned} y_1^\star &= \lambda^{-1/2}/\sqrt{2} - 1 \\ y_l^\star &= \lambda^{-1/2} - 1, \qquad l \ge 2, \end{aligned}$$

and we can check that $y_1^\star \le y_2^\star = \cdots = y_L^\star$. Let us express λ as a function of c. By using the constraint we get

$$c = \sum_l y_l^\star = \lambda^{-1/2}\Big(1/\sqrt{2} + (L-1)\Big) - L$$

and it follows that

$$\sqrt{2}(1 + y_1^\star) = (1 + y_l^\star) = \lambda^{-1/2} = (L + c)(1/\sqrt{2} + (L-1)).$$

Then we get by direct computations

$$g(y^\star) = \frac{1}{\sqrt{2}(L+c)}\Big[(L-1)(\sqrt{2}c + \sqrt{2} - 2) - 1/\sqrt{2}\Big]$$

so the condition $G(c) = g(y^\star) \le 0$ is equivalent to $(L-1)(\sqrt{2}c + \sqrt{2} - 2) \le 1/\sqrt{2}$, that is,

$$c \le \sqrt{2} - 1 + \frac{1}{2(L-1)}.$$

To conclude, let us consider $y := (M\|\alpha^l\|_0)_{l=1}^L$ and assume the strict inequality (30) is satisfied. Then by the above computations $g(y) \le G(M\|\alpha\|_0) < 0$ hence the strict inequality (22) is satisfied. If (30) is satisfied as a large inequality and there exists some index $l \ge 2$ such that $(1 + M\|\alpha^l\|_0)/(1 + M\|\alpha^1\|_0) \neq \sqrt{2}$ (this is generally the case!), then $y \neq y^\star$ so we have $g(y) < G(M\|\alpha\|_0) \le 0$ and we get the same result. In both cases, we reach the conclusion using Theorem 2. $\square$

The sufficient conditions in Corollary 1 and 2 are very similar, but the latter is a bit more restrictive, with a gap $1/2 - (\sqrt{2} - 1) \approx 0.086$ in the constant in front of $1/M$. Table I lists the values of the constant in front of $1/M(\boldsymbol{D})$ in Corollary 2. For $L = 2$, we recover the constant $\sqrt{2} - 1/2$ from [2, Theorem 3]. For larger values of L, we get more restrictive constraints, i.e., with smaller constants. Indeed, for $L \ge 7$, one can check that for any value of M,

$$\left(\sqrt{2} - 1 + \frac{1}{2(L-1)}\right)/M < (1 + 1/M)/2$$

so the general sufficient condition in Theorem 1 is less restrictive than the specialized one in Corollary 2. For $L \le 6$ and small values of M (i.e., because of (16), in large dimension $N \gg 1$), the condition in Corollary 2 is less restrictive than that of Theorem 1, and we get an improved result. For large values of M and $L \le 6$, one has to check on a case-by-case basis which result is stronger.

TABLE I
NUMERICAL VALUES OF THE CONSTANT $\sqrt{2} - 1 + \frac{1}{2(L-1)}$ IN COROLLARY 2 FOR SMALL VALUES OF L

L	2	3	4	5	6	7
c	0.914	0.664	0.580	0.539	0.514	0.497

IV. HIGHLY REDUNDANT DICTIONARIES

Let us show how to apply the extended result (Theorem 2) to highly redundant dictionaries. It is perhaps not obvious that one can have a large number of orthonormal bases in $\mathbb{R}^N$ with a small coherence factor $M(\boldsymbol{D})$, but this is possible (for certain values of N), and we will use the following theorem to build examples of such dictionaries. We refer to [6] and [7] for a proof of Theorem 3.

Theorem 3: Let $N = 2^{j+1}$, $j \geq 0$, and consider $\mathcal{H} = \mathbb{R}^N$. There exists a dictionary $\boldsymbol{D}$ in $\mathcal{H}$ consisting of the union of $L = 2^j = N/2$ orthonormal bases for $\mathcal{H}$, such that for any pair u, v of distinct vectors belonging to $\boldsymbol{D}$: $|\langle u, v\rangle| \in \{0, N^{-1/2}\}$.

For $N = 2^j$, $j \geq 0$, and $\mathcal{H} = \mathbb{C}^N$, one can find a dictionary $\boldsymbol{D}$ in $\mathcal{H}$ consisting of the union of $L = N + 1$ orthonormal bases for $\mathcal{H}$, again with the perfect separation property that for any pair u, v of distinct vectors belonging to $\boldsymbol{D}$: $|\langle u, v\rangle| \in \{0, N^{-1/2}\}$.

The dictionaries from Theorem 3 are called **Grassmannian dictionaries** due to the fact that their construction is closely related to the Grassmannian packing problem, see [6] and [7] for details.

For $N = 2^{j+1}$, Theorem 3 tells us that we can take a dictionary $\boldsymbol{D}$ consisting of the union of $N + 1$ orthonormal bases in $\mathbb{C}^N$, that is, $\boldsymbol{D}$ contains the large number $N(N+1)/2$ of elements, but we still have coherence $M(\boldsymbol{D}) = N^{-1/2}$. We can extract from such a dictionary many examples of unions $\boldsymbol{D}_L$ of L bases $(2 \leq L \leq N+1)$ with the same coherence. For each example, we can apply Theorem 1 or Corollary 2 to conclude that α is the unique sparsest ℓ^0 and ℓ^1 representation of $s := \boldsymbol{D}_L \alpha$ as soon as

$$\|\alpha\|_0 < \max\left(\sqrt{N}/2 + 1/2, \sqrt{N}\left(\sqrt{2} - 1 + \frac{1}{2(L-1)}\right)\right\}.$$

V. CONCLUSION

We have studied sparse representations of signals using an arbitrary dictionary $\boldsymbol{D}$ in $\mathcal{H} = \mathbb{R}^N$ (resp., $\mathcal{H} = \mathbb{C}^N$). For any dictionary $\boldsymbol{D}$, $\tau \in \{0, 1\}$, and a given signal s we prove that α, with $s := \boldsymbol{D}\alpha$, is the unique solution to the optimization problem

$$\text{minimize} \quad \|\beta\|_\tau \quad \text{subject to } \boldsymbol{D}\beta = s \tag{31}$$

provided that $\|\alpha\|_0 < \frac{1}{2}(1 + 1/M(\boldsymbol{D}))$. So this condition on $\|\alpha\|_0$ ensures that the more difficult ℓ^0 minimization problem has exactly the same unique solution α as the ℓ^1 problem. This is of practical importance since (31) can be restated and solved as a linear programming minimization problem, thus giving us a feasible way to actually compute the minimizer.

When $\boldsymbol{D}$ is a union of $L \geq 2$ orthonormal bases for $\mathcal{H}$, we have derived the sufficient condition

$$\|\alpha\|_0 < \left(\sqrt{2} - 1 + \frac{1}{2(L-1)}\right)\frac{1}{M(\boldsymbol{D})}$$

for α with $s := \boldsymbol{D}\alpha$, to be the simultaneous unique minimizer in (31) for $\tau \in \{0, 1\}$. When $2 \leq L \leq 6$, this condition is generally less restrictive (and the result thus covers more cases) than the estimate for arbitrary dictionaries. For $L = 2$, we simply recover the main result from [2].

We also proved an uncertainty principle for unions of L orthonormal bases for $\mathcal{H}$ and derived a slightly less restrictive sufficient condition

$$\|\alpha\|_0 < \left(\frac{1}{2} + \frac{1}{2(L-1)}\right)\frac{1}{M(\boldsymbol{D})}$$

to ensure that the (most difficult) ℓ^0 minimization problem admits α as a unique solution.

The proofs of the above results are based on the techniques introduced in [1] and [2] so the main contribution of the present correspondence is to point out that we are not restricted to dictionaries that are the union of two orthonormal bases. We can consider more general dictionaries and still enjoy all the practical benefits from restating the problem as a linear programming minimization problem and get the ℓ^0 minimizer for "free" in cases where the output from the LP algorithm has few nonzero entries.

Finally, we should note that many *natural* and useful redundant dictionaries such as the discrete Gabor dictionary, unions of bi-orthogonal discrete wavelet dictionaries, etc., *cannot* be written as a union of two orthonormal bases and thus were not covered by the results in [1] and [2].

REFERENCES

[1] D. Donoho and X. Huo, "Uncertainty principles and ideal atomic decompositions," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2845–2862, Nov. 2001.

[2] M. Elad and A. Bruckstein, "A generalized uncertainty principle and sparse representations in pairs of bases," *IEEE Trans. Inform. Theory*, vol. 48, pp. 2558–2567, Sept. 2002.

[3] D. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization," *Proc. Nat. Acad. Sci.*, vol. 100, no. 5, pp. 2197–2202, Mar. 2003.

[4] A. Feuer and A. Nemirovsky, "On sparse representations in pairs of bases," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1579–1581, June 2003.

[5] D. Bertsekas, *Non-Linear Programming*, 2nd ed. Belmont, MA: Athena Scientific, 1995.

[6] A. R. Calderbank, P. J. Cameron, W. M. Kantor, and J. J. Seidel, "$\boldsymbol{Z}_4$-Kerdock codes, orthogonal spreads, and extremal euclidean line-sets," *Proc. London Math. Soc. (3)*, vol. 75, no. 2, pp. 436–480, 1997.

[7] T. Strohmer and R. Heath, "Grassmannian frames with applications to coding and communications," *Appl. Comp. Harm. Anal.*, vol. 14, no. 3, pp. 257–275, 2003.