Sparse Representations in Unions of Bases

Rémi Gribonval, Morten Nielsen

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Sparse Representations in Unions of Bases

Rémi Gribonval and Morten Nielsen

N° 4642

Novembre 2002

THÈME 3

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--4642--FR+ENG

Sparse Representations in Unions of Bases

Rémi Gribonval* and Morten Nielsen†

Thème 3 — Interaction homme-machine,
images, données, connaissances
Projet METISS

Rapport de recherche n° 4642 — Novembre 2002 — 15 pages

Abstract: The purpose of this paper is to generalize a result by Donoho, Huo, Elad and Bruckstein on sparse representations of signals in a union of two orthonormal bases. We consider general (redundant) dictionaries in finite dimension, and derive sufficient conditions on a signal for having unique sparse representations in such dictionaries. The special case where the dictionary is given by a union of several orthonormal bases is studied in more detail. In particular, it is proved that the result of Donoho and Huo, concerning the replacement of a combinatorial optimization problem with a linear programming problem when searching for sparse representations, has an analog for dictionaries that may be highly redundant.

Key-words: dictionary, mutually incoherent bases, Grassmannian frame, linear programming, sparse representation, nonlinear approximation

* Email : remi.gribonval@inria.fr
† Email: mnielsen@math.auc.dk,
Department of Mathematical Sciences,
Aalborg University,
Fredrik Bajers Vej 7G,
DK-9220 Aalborg East,
Denmark.

Unité de recherche INRIA Rennes
IRISA, Campus universitaire de Beaulieu, 35042 Rennes Cedex (France)
Téléphone : +33 2 99 84 71 00 — Télécopie : +33 2 99 84 71 71

Représentations parcimonieuses dans des dictionnaires constitués de l'union de bases orthonormales

Résumé : On généralise les résultats de Donoho, Huo, Elad et Bruckstein sur les décompositions parcimonieuses de signaux dans l'union de deux bases orthonormales en dimension finie. On fournit tout d'abord –pour des dictionnaires redondants arbitraires– une condition suffisante sur le signal pour que ses représentations les plus parcimonieuses soient uniques. On s'intéresse ensuite au cas de dictionnaires formés de l'union de deux (ou plus) bases orthogonales. On prouve que pour certains signaux ayant une représentation suffisamment parcimonieuse, il est possible de calculer celle-ci par programmation linéaire au lieu d'effectuer une recherche combinatoire exhaustive. Des exemples de dictionnaires très redondants où nos résultats s'appliquent sont donnés.

Mots-clés : dictionnaire, bases mutuellement incohérentes, frame Grassmannien, programmation linéaire, décomposition parcimonieuse, approximation non-linéaire

1 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 $\mathbf{D} = [g_1, \ldots, g_K]$ for a dictionary.*

By a representation of s in $\mathbf{D}$ we mean a (column) vector $\alpha = (\alpha_k) \in \mathbb{R}^K$ (resp. in $\mathbb{C}^K$) such that $s = \mathbf{D}\alpha$. We notice that when $K > N$, the vectors of $\mathbf{D}$ 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 non-zero 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 [4, 3] we will measure the sparsity of a representation $s = \mathbf{D}\alpha$ by two quantities: the ℓ^0 and the ℓ^1 norm of α, resp. (the ℓ^0-norm simply counts the number of non-zero entries of a vector). This leads to the following two minimization problems to determine the sparsest representation of s:

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

and

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

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 details in [3] and later been refined in [4] in the special case where the dictionary $\mathbf{D}$ is the union of *two* orthonormal bases. Below we generalize the results of [4, 3] to arbitrary dictionaries $\mathbf{D}$. The case where $\mathbf{D}$ 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 [4] valid for $L = 2$.

In Section 2 we provide conditions for a solution α of the problem

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

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

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

is the **coherence** of the dictionary. The special case where $\mathbf{D}$ is the union of $L \geq 2$ bases is studied in Section 3 leading to explicit sufficient conditions for $\tau = 0$ with $f(\mathbf{D}) = \left(1/2 + \frac{1}{2(L-1)}\right)/M(\mathbf{D}))$ and for $\tau \in \{0,1\}$ with $f(\mathbf{D}) = \left(\sqrt{2} - 1 + \frac{1}{2(L-1)}\right)/M(\mathbf{D}))$. In the case $L = 2$ we simply recover the main result from [4], and for $L \leq 6$ we obtain a condition that is *less restrictive* than the condition for arbitrary dictionaries. In Section 4 we construct highly redundant dictionaries where the results of the present paper give fairly relaxed conditions for (2) to have a unique solution.

2 Sparse ℓ^τ representations, $0 \leq \tau \leq 1$

Any vector $s \in \mathcal{H}$ has (possibly many) representations $s = \mathbf{D}\alpha$ 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

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

as well as the integer quantity

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

By refining ideas from [4] we have the following Lemma.

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

$$P_\tau(S, \mathbf{D}) := \max_{x \in Ker(\mathbf{D}), x \neq 0} \frac{\sum_{k \in S} |x_k|^\tau}{\sum_k |x_k|^\tau} \tag{8}$$

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

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

Proof. The lemma was used without being stated explicitely in [3, 4], in the special case $\tau = 1$ and with $\mathbf{D}$ a union of two orthonormal bases. The proof follows the same steps as in [3, 4]. Under the assumption $P_\tau(S, \mathbf{D}) < 1/2$ and $S(\alpha) \subset S$, what we need to prove is that for all $x \in \mathrm{Ker}(\mathbf{D})$, $\sum_k |\alpha_k + x_k|^\tau > \sum_k |\alpha_k|^\tau$. 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 $x \in \mathrm{Ker}(\mathbf{D})$

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

or equivalently

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

But this is exactly the assumption $P_\tau(S, \mathbf{D}) < 1/2$. To prove the result for $P_\tau(S, \mathbf{D}) = 1/2$ we copy the above line of arguments and simply replace strict inequalities with large ones. To prove the result for $P_\tau(S, \mathbf{D}) > 1/2$, it is sufficient to take some $x \in Ker(\mathbf{D})$ 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 $x = \beta - \alpha \in Ker(\mathbf{D})$ one easily checks that $\mathbf{D}\beta = \mathbf{D}\alpha$. 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 $P_\tau(S, \mathbf{D}) < 1/2$. Of particular interest are sufficient conditions that take the form

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

which correspond to uniqueness result of the form

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

Lemma 2 *Denote*

$$f_0(\mathbf{D}) := \begin{cases} Z(\mathbf{D})/2 & Z \text{ even} \\ Z(\mathbf{D}) + 1)/2 & Z \text{ odd.} \end{cases} \tag{10}$$

For $\tau = 0$, (9) holds true if, and only if, $f(\mathbf{D}) \leq f_0(\mathbf{D})$. If (9) holds true for some $0 \leq \tau \leq 1$ with some $f(\mathbf{D})$, then it holds true with $f(\mathbf{D})$ for $\tau = 0$, hence $f(\mathbf{D}) \leq f_0(\mathbf{D})/2$.

Proof. For any S we observe that

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

so we get that (9) holds true with any $f \leq Z/2$. If Z is odd and $f \leq (Z+1)/2$, for any S, $\mathrm{card}(S) < f$ implies $\mathrm{card}(S) \leq (Z-1)/2 < Z/2$ so (9) holds true. For the converse result, we consider the case Z even (we leave the case Z odd to the reader). We can take $x \in Ker(\mathbf{D})$ such that $\|x\|_0 = Z$. Assuming $f > Z/2$, there is $S \subset S(x)$ with $Z/2 \leq \mathrm{card}(S) < f$, and we get $P_0(S, \mathbf{D}) \geq 1/2$.

Now, the second part of the Lemma 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 $\mathbf{D}\beta = \mathbf{D}\alpha$ we have $\|\beta\|_\tau > \|\alpha\|_\tau$. Assume β satisfies $\mathbf{D}\beta = \mathbf{D}\alpha$ 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 and we can use the first part of the Lemma to conclude that $f \leq Z/2$ (resp. $(Z+1)/2$). □

There are two consequences of this Lemma. The first one is that we need to estimate $Z(\mathbf{D})$. The second is that if we are able to prove that (9) holds for some $0 \leq \tau \leq 1$ with $f(\mathbf{D})$, 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 [3, 4], the case of $\mathbf{D} = [\mathbf{B}_1, \mathbf{B}_2]$ was considered where $\mathbf{B}_1$ and $\mathbf{B}_2$ are two orthonormal matrices corresponding to orthonormal bases. Donoho and Huo proved an *uncertainty principle* $Z([\mathbf{B}_1, \mathbf{B}_2]) \geq 1 + 1/M$ [3, Th. VII.3] and obtained the sufficient condition (9) for $\tau = 1$ with

$$f([\mathbf{B}_1, \mathbf{B}_2]) = \frac{1}{2}(1 + 1/M(\mathbf{D})). \tag{11}$$

Elad and Bruckstein improved the uncertainty principle by getting [4, Th. 1] $Z([\mathbf{B}_1, \mathbf{B}_2]) \geq 2/M$. Thus, they obtained the sufficient condition (9) for $\tau = 0$ with the less restrictive constant

$$f([\mathbf{B}_1, \mathbf{B}_2]) = 1/M(\mathbf{D}). \tag{12}$$

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

$$f_1([\mathbf{B}_1, \mathbf{B}_2]) = \left(\sqrt{2} - 1/2\right)/M(\mathbf{D}) \approx 0.914/M(\mathbf{D}) \tag{13}$$

and raised the question was whether (9) indeed holds true for $\tau = 1$ with the even less restrictive constant (12).

Next we show that result with the most restrictive of the constants, that is to say (11), 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*

$$\|\alpha\|_0 < \frac{1}{2}(1 + 1/M(\mathbf{D})) \tag{14}$$

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 $x \in \mathrm{Ker}(\mathbf{D})$. 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, $|x_k| \leq M(\mathbf{D}) \sum_{k' \neq k} |x_{k'}|$. It follows that

$$(1 + M)\,|x_k| \leq M\|x\|_1. \tag{15}$$

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

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

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

Note that a direct consequence of Lemma 2 is that for arbitrary dictionaries we have the generalized uncertainty principle

$$Z(\mathbf{D}) \geq 1 + 1/M(\mathbf{D}). \tag{16}$$

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 $\mathbf{D}$. 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

$$M(\mathbf{D}) \geq 1/\sqrt{N}. \tag{17}$$

3 Sparse representations in unions of bases

We now switch to the special case of $\mathbf{D}$ a union of L orthonormal bases, *i.e.* $\mathbf{D} = [\mathbf{B}_1, \ldots, \mathbf{B}_L]$ where $\mathbf{B}_l$ 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 $Z(\mathbf{D})$.

Lemma 3 *Let* $\mathbf{D}$ *a union of* L *orthonormal bases. Let* $x = \begin{bmatrix} x^1 \\ \ldots \\ x^L \end{bmatrix} \in Ker(\mathbf{D})$ *with* $x^l \in \mathbb{R}^N$ *(resp.* $\mathbb{C}^N$*) and assume* $x \neq 0$*. Then*

$$\sum_{l=1}^{L} \frac{1}{1 + M(\mathbf{D})\|x^l\|_0} \leq L - 1. \tag{18}$$

Consequently

$$Z(\mathbf{D}) \geq \big(1 + \frac{1}{L-1}\big)\frac{1}{M(\mathbf{D})}. \tag{19}$$

Proof. Because $x \in \mathrm{Ker}(\mathbf{D})$, 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 with the absolute values of the original vector x^l, we have for all $x \in \mathrm{Ker}(\mathbf{D})$ and $1 \leq l \leq L$

$$X^l \leq M(\mathbf{D}) \sum_{l' \neq l} \|X^{l'}\|_1 1_N \tag{20}$$

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 $\|X^l\|_1 \leq M\|X^l\|_0 \sum_{l' \neq l} \|X^{l'}\|_1$. It follows that $(1 + M\|X^l\|_0)\|X^l\|_1 \leq M\|X^l\|_0 \sum_{l'} \|X^{l'}\|_1$, hence

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

Summing over l we obtain $\|x\|_1 \leq \sum_l \frac{M\|x^l\|_0}{1+M\|x^l\|_0} \|x\|_1$ 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 (18). By the concavity 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 (19) follows. □

Notice that for $L = 2$ the condition (18) can be rewritten $\sqrt{\|x^1\|_0 \|x^2\|_0} \geq 1/\sqrt{M}$ as in [4, Th. 1]. There are example of pairs of bases with $Z = 2/M$, so the generalized uncertainty principle (19) is sharp for $L = 2$. For $L \geq 3$ it seems to be an open problem whether there exists examples of L orthonormal bases for which MZ is arbitrarily close to $1 + 1/(L-1)$. Using (19) together with Lemma 2 we have the Corollary.

Corollary 1 *Let* $\mathbf{D}$ *a union of* L *orthonormal bases. If*

$$\|\alpha\|_0 < \left(\frac{1}{2} + \frac{1}{2(L-1)}\right) \frac{1}{M(\mathbf{D})} \tag{21}$$

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

For $L = 2$, we find again the least restrictive condition (12) of Elad and Bruckstein. As we increase the number L of bases while keeping M constant (we will see in Section 4 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 (21) 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, (21) is less restrictive than (14). For $L \geq 1 + 1/M$ however, the bound in Corollary 2 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 [4] to get the less restrictive sufficient condition $\|\alpha\|_0 < (\sqrt{2} - 0.5)/M(\mathbf{D})$. The authors in [4] 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 $P_1(S(\alpha)) < 1/2$ is that

$$2M^2(\mathbf{D})K_1K_2 + M(\mathbf{D})\max(K_1, K_2) - 1 < 0. \tag{22}$$

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

Theorem 2 *Let* $\mathbf{D}$ *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* $\mathbf{B}_l$ *have been numbered so that* $\|\alpha^1\|_0 \leq \dots \leq \|\alpha^L\|_0$*. If*

$$\sum_{l \geq 2} \frac{M\|\alpha^l\|_0}{1 + M\|\alpha^l\|_0} < \frac{1}{2(1 + M\|\alpha^1\|_0)}. \tag{23}$$

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

Proof We follow the proof of [4, Th. 3] and start similarly to the proof of Lemma 3. Consider $x = \begin{bmatrix} x^1 \\ \dots \\ x^L \end{bmatrix} \in \mathrm{Ker}(\mathbf{D})$ 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 with the absolute values of the original vector x^l, we have for all $x \in \mathrm{Ker}(\mathbf{D})$ and $1 \leq l \leq L$

$$X^l \leq M(\mathbf{D}) \sum_{l' \neq l} \mathbf{1}_{N \times N} X^{l'}, \tag{24}$$

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 $x \in \mathrm{Ker}(\mathbf{D})$ with $\|x\|_1 = \sum_{l=1}^L \|x^l\|_1 = 1$, we have

$$\sum_{l=1}^{L} 1_N^T X^l = 1 \tag{25}$$

$$\sum_{l=1}^{L} 1_{S(\alpha^l)}^T X^l = \sum_{k \in S(\alpha)} |x_k| \tag{26}$$

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

(23) and the constraints (24)-(25) and $X^l \geq 0$ we have

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

Let us proceed as in [4]: by replacing the equality constraints (25) with two inequalities, we now have a classical linear programming problem, which can be put into canonical form

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

with $Z = \begin{bmatrix} X^1 \\ \cdots \\ X^L \end{bmatrix}$, $C^T = [-1_{S(\alpha^1)}, \ldots, -1_{S(\alpha^L)}]$,

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

and

$$B^T = [0.1^T_N, \ldots, 0.1^T_N, 1, -1].$$

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

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

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

We will look for such a U in a parametric form

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

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 $\mathbf{A}^T U \leq C$.

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

$$(b-c)1_N + M\mathbf{1}_{N\times N}\left(\sum_{l'\neq l} a_{l'} 1^T_{S(\alpha^{l'})}\right) + (1-a_l)1^T_{S(\alpha^l)} \leq 0.$$

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

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

where we used the notation $\|\alpha^l\|_0 := \mathrm{card}(S(\alpha^l)) = \|\alpha^l\|_0$. 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 $1 \leq l \leq L$ such that $\|\alpha^l\|_0 \neq 0$, as

$$b - c + M\sum_{l'} a_{l'}\|\alpha^{l'}\|_0 \leq Ma_l\|\alpha^l\|_0 + (a_l - 1)^-, \tag{27}$$

The constraint for all l such that $\|\alpha^l\|_0 = 0$ (if there is any) is $b - c + M\sum_{l'} a_{l'}\|\alpha^{l'}\|_0 \leq 0$. 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 $1 \leq \|\alpha^l\|_0 < \theta$ and $a_l(\theta) := (1+M\theta)/(1+M\|\alpha^l\|_0)$ when $\|\alpha^l\|_0 \geq \theta$. 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 (27) becomes

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

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

$$b - c + M\Sigma(\theta) \leq 0. \tag{29}$$

Obviously, (28) (resp. (29)) can always be satisfied by taking $b = g(\theta)^+$ and $c = -g(\theta)^-$ with $g(\theta) := M\big(\theta - \Sigma(\theta)\big)$ (resp. $g(\theta) := -M\Sigma(\theta)$).

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

$$\max_\theta g(\theta) > -1/2. \tag{30}$$

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

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

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 (30). 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

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

□

In the case of $L = 2$ bases, Eq. (23) is exactly the condition (22) proved in [4] where it is proved that a simpler sufficient condition is $\|\alpha\|_0 \leq (\sqrt{2} - 1/2)/M$. The general condition (23) 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 $\|\alpha\|_0 < f(\mathbf{D})$.

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

$$\|\alpha\|_0 < \left(\sqrt{2} - 1 + \frac{1}{2(L-1)}\right)\frac{1}{M(\mathbf{D})} \tag{31}$$

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 $l \geq 2$ 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 (23) can be rewritten

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

For any $c > 0$ consider the set $H_c := \{y, \sum_{l=1}^L y_l = c, \forall l\ y_l \geq 0\}$ 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 $\frac{\partial g}{\partial y_l^\star} = \lambda$, $1 \leq l \leq L$. For $l = 1$ this becomes $(1 + y_1^\star)^{-2}/2 = \lambda$, while for $2 \leq l \leq L$ 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 $y_l^\star \geq 0$ is given by

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

and we can check that $y_1^\star \leq y_2^\star = ... = y_L^\star$. 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

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

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

Table 1: Numerical values of the constant $\sqrt{2}-1+\frac{1}{2(L-1)}$ in Corollary 2 for small values of L.

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

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

To conclude, let us consider $y := (M\|\alpha^l\|_0)_{l=1}^L$ and assume the strict inequality (31) is satisfied. Then by the above computations $g(y) \leq G(M\|\alpha\|_0) < 0$ hence the strict inequality (23) is satisfied. If (31) is satisfied as a large inequality and there exists some index $l \geq 2$ 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 $g(y) < G(M\|\alpha\|_0) \leq 0$ and we get the same result. In both cases, we reach the conclusion using Theorem 2. □

The sufficient conditions in Corollary 1 and 2 are very similat, 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 1 list the values of the constant in front of $1/M(\mathbf{D})$ in Corollary 2. For $L = 2$ we recover the constant $\sqrt{2}-1/2$ from [4, Th. 3]. For larger values of L, we get more restrictive constraints, *i.e.* with smaller constants. Indeed, for $L \geq 7$, one can check that for any value of M, $(\sqrt{2}-1+\frac{1}{2(L-1)})/M < (1+1/M)/2$, so the general sufficient condition in Theorem 1 is less restrictive than the specialized one in Corollary 2. For $L \leq 6$ and small values of M (*i.e.*, because of (17), 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 $L \leq 6$, one has to check on a case by case basis which result is stronger.

4 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 $M(\mathbf{D})$, 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 [2, 5] 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 $\mathbf{D}$ in $\mathcal{H}$ consisting of the union of $L = 2^j = N/2$ orthonormal bases for $\mathcal{H}$, such that for any pair $u, v \in \mathcal{D}, u \neq v$: $|\langle u, v\rangle| \in \{0, N^{-1/2}\}$.*

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

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 [2, 5] for details.

For $N = 2^{j+1}$ the Theorem tells us that we can take a dictionary $\mathbf{D}$ consisting of the union of $N+1$ orthonormal bases in $\mathbb{C}^N$, that is $\mathbf{D}$ contains the large number $N(N+1)/2$ of elements, but we still have coherence $M(\mathbf{D}) = N^{-1/2}$. We can extract from such a dictionary many examples of unions $\mathbf{D}_L$ 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 $s := \mathbf{D}_L\alpha$ as soon as $\|\alpha\|_0 < \max\left(\sqrt{N}/2 + 1/2, \sqrt{N}(\sqrt{2} - 1 + \frac{1}{2(L-1)})\right)\}$.

5 Conclusion

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

$$\text{minimize} \quad \|\beta\|_\tau \quad \text{subject to } \mathbf{D}\beta = s, \tag{32}$$

provided that $\|\alpha\|_0 < \frac{1}{2}(1 + 1/M(\mathbf{D}))$. So this condition on $\|\alpha\|_0$ ensure 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 (32) can be restated and solved as a linear programming minimization problem, thus giving us a feasible way to actually compute the minimizer $s = \mathbf{D}\alpha$.

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

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

for α with $s := \mathbf{D}\alpha$, to be the simlutaneous unique minimizer in (32) 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 extimate for arbitrary dictionaries. For $L = 2$ we simply recover the main result from [4].

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

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

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 [4, 3] so the main contribution of the present paper 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 non-zero 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 [4, 3].

References

[1] D. Bertsekas. *Non-Linear Programming.* Athena Scientific, Belmont, MA, 2nd edition, 1995.

[2] A. R. Calderbank, P. J. Cameron, W. M. Kantor, and J. J. Seidel. Z_4-Kerdock codes, orthogonal spreads, and extremal Euclidean line-sets. *Proc. London Math. Soc. (3)*, 75(2):436–480, 1997.

[3] D.L. Donoho and Xiaoming Huo. Uncertainty principles and ideal atomic decompositions. *IEEE Trans. Inform. Theory*, 47(7):2845–2862, November 2001.

[4] M. Elad and A.M. Bruckstein. A generalized uncertainty principle and sparse representations in pairs of bases. *IEEE Trans. Inform. Theory*, 48(9):2558–2567, September 2002.

[5] T. Strohmer and R. Heath. Grassmannian frames with applications to coding and communications. Technical report, 2002. Preprint, submitted to Appl.Comp.Harm.Anal.

Unité de recherche INRIA Rennes
IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)

Unité de recherche INRIA Futurs : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38330 Montbonnot-St-Martin (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399