SPARSE DECOMPOSITIONS IN “INCOHERENT” DICTIONARIES

R. Gribonval

IRISA (INRIA & CNRS UMR 6074),
Campus de Beaulieu, F-35760 Rennes Cedex, FRANCE
E-mail : remi.gribonval@inria.fr

M. Nielsen

Dept. of Mathematical Science, Aalborg University,
Fredrik Bjers Vej 7G, DK-9220 Aalbord East, DENMARK
E-mail : mnielsen@math.auc.dk

ABSTRACT

The purpose of this paper is to generalize a result by Donoho, Huo, Elad and Bruckstein on sparse representations of signals/images in a union of two orthonormal bases. We consider general (redundant) dictionaries in finite dimension, and derive sufficient conditions on a signal/image for having a unique sparse representation in such a dictionary. 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. The special case where the dictionary is given by a union of several orthonormal bases is studied in more detail and some examples are given.

1. INTRODUCTION

Images and signals, when considered as vectors in a high-dimensional Hilbert space $\mathcal{H} = \mathbb{R}^N$ (resp. $\mathcal{H} = \mathbb{C}^N$), can be uniquely represented by their set of coefficients in any given basis. For many applications –such as denoising [1], compression [2] or source separation [3] –, the goodness of such a representation is related to its *sparsity* : a good representation will be such that many coefficients are almost zero while only a few ones are non negligible.

It has been observed that in many cases, natural images or sounds contain superimposed structures of very different nature (edges and textures, transients and stationary parts) that do not necessary have sparse representations in a single basis. The dictionary approach [4] to signal and image representation considers more general types of expansions, where the basis is replaced by a so-called dictionary. 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 $\mathbf{D} = [g_1, \dots, g_K]$ for a dictionary.

A representation of an image $y \in \mathcal{H}$ in a dictionary $\mathbf{D}$ is *any* (column) vector $\alpha = (\alpha_k)_{k=1}^K$ such that $y = \mathbf{D}\alpha$. When $K > N$, the dictionary forms a family which is linearly dependent, so there exists infinitely many representations for a single image. This allows for the choice of “the best” representation with a criterion that may depend on the application. In this paper we will consider two criteria related to the notion of sparsity. Following [5, 6] we will measure the sparsity of a representation $y = \mathbf{D}\alpha$ 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 y:

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

and

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

The optimization problem (1) is combinatorial and generally computationally intractable [7]. On the other hand, (2) is much easier to handle through the use of linear programming (LP), which is the ground of the Basis Pursuit (BP) approach [8] to dictionary signal and image processing. None of these problems is strictly convex, so it is important to know when they have a unique solution. Moreover, it has been observed in practice [8] that solving (2) often gives the solution to (1), hence it is a question whether this can be grounded mathematically. In the special case where $\mathbf{D}$ is the union of *two* orthonormal bases, these issues have been studied in details in [5] and later been refined in [6].

In this paper[1], we generalize the results of [5, 6] [2] to arbitrary dictionaries $\mathbf{D}$. The flavour of the main results is the following: we provide sufficient conditions that can be checked on a computed solution α of the ℓ^1 problem (2) to guarantee that : (i) it is indeed *the unique* solution of (2); (ii) it is *also* the unique solution of (1).

In Section 2 we introduce the notations and in Section 3-4 we give the results for arbitrary dictionaries. The case where $\mathbf{D}$ is the union of $L \geq 2$ orthonormal bases for $\mathcal{H}$ is studied in detail in section 5. This leads to a natural generalization of the recent results from [6] valid for $L = 2$.

[1] The detailed proofs are available in our preprint [9].

[2] A parallel work done independently by Donoho and Elad [10] also addresses the question of generalizing previous Basis Pursuit analysis results general dictionaries. Though there are some similarities between this work to the work in [10], somewhat different perspective on the problem is adopted.

2. NOTATIONS

We will make an extensive use of the following definitions. The **support** of a coefficient vector $\alpha = (\alpha_k) \in \mathbb{R}^K$ (resp. $\mathbb{C}^K$) is $S(\alpha) := \{k, \alpha_k \neq 0\}$. The **kernel** Ker($\mathbf{D}$) $:= \{x, \mathbf{D}x = 0\}$ of the dictionary will play a special role as well as the integer quantity (called **spark** of $\mathbf{D}$ in [10])

$$Z(\mathbf{D}) := \min_{x \in \mathrm{Ker}(\mathbf{D}), x \neq 0} \|x\|_0. \tag{3}$$

The **coherence** of a dictionary is

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

and it will serve to estimate $Z(\mathbf{D})$ in some of our results. It is not difficult to check that, when $\mathbf{D}$ contains at least an orthonormal basis and some additional unit vector, the value of M satisfies $1/\sqrt{N} \leq M(\mathbf{D}) \leq 1$. When the lower bound $M = 1/\sqrt{N}$ is achieved we say that $\mathbf{D}$ is perfectly incoherent.

3. MAIN RESULTS

In this section we consider $0 \leq \tau \leq 1$ and an arbitrary dictionary $\mathbf{D}$. We provide conditions for a solution α of the problem

$$\text{minimize} \quad \|\alpha\|_\tau \quad \text{subject to } y = \mathbf{D}\alpha \tag{5}$$

to be indeed unique. Our first main result is the following Lemma which we proved in [9] by refining ideas from [6].

Lemma 1 ([9]) *Let* $\mathbf{D}$ *a (possibly redundant) dictionary and* $S \subset \{1, \ldots, K\}$ *a set of indices. For* $0 \leq \tau \leq 1$ *define*

$$P_\tau(S, \mathbf{D}) := \max_{x \in Ker(\mathbf{D}), x \neq 0} \frac{\sum_{k \in S} |x_k|^\tau}{\sum_k |x_k|^\tau} \tag{6}$$

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

1. *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* $\mathbf{D}\alpha = \mathbf{D}\beta$.
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* (5) *with* $y := \mathbf{D}\alpha$.
3. *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* (5) *with* $y := \mathbf{D}\alpha$.

The quantities $P_\tau(S, \mathbf{D})$ are not always completely straightforward to evaluate. Next we concentrate on the case $\tau \in \{0, 1\}$. In the special case $\tau = 0$ we have, for any S,

$$P_0(S, \mathbf{D}) \leq \max_{x \in \mathrm{Ker}(\mathbf{D}), x \neq 0} \frac{\mathrm{card}(S)}{\|x\|_0} = \frac{\mathrm{card}(S)}{Z(\mathbf{D})}. \tag{7}$$

It follows (we leave the proof to the reader) that we have :

Corollary 1 *For all* α *such that* $card(S(\alpha)) < Z(\mathbf{D})/2$, α *is* the unique *solution to the* ℓ^0 *problem (1) with* $y := \mathbf{D}\alpha$.

In the next sections we will prove lower bounds on $Z(\mathbf{D})$ in order to have explicit sufficient conditions for uniqueness of an ℓ^0 solution.

When we are given an image y and compute (with LP) a solution α to the ℓ^1 problem (2), we can use Corollary 1 to check whether it is also *a* (and indeed *the unique*) solution to the ℓ^0 problem (5). Another question is its uniqueness with respect to the ℓ^1 problem. While we can in theory check if $P_1(S(\alpha), \mathbf{D}) < 1/2$, this is not a very explicit test. In the next sections we will provide more explicit sufficient conditions to check uniqueness of an ℓ^1 solution.

4. ARBITRARY DICTIONARIES

In this section, we provide a general lower bound on $Z(\mathbf{D})$ based on the coherence $M(\mathbf{D})$ of an arbitrary dictionary.

In [5, 6], 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* [5, Th. VII.3]

$$Z([\mathbf{B}_1, \mathbf{B}_2]) \geq 1 + \frac{1}{M(\mathbf{D})} \tag{8}$$

which corresponds to the sufficient condition for uniqueness of the ℓ^0 solution

$$card(S(\alpha)) < \frac{1}{2}\left(1 + \frac{1}{M(\mathbf{D})}\right). \tag{9}$$

In addition, they proved that (11) is also a sufficient condition for uniqueness of the ℓ^1 solution. Elad and Bruckstein improved the uncertainty principle by getting [6, Th. 1]

$$Z([\mathbf{B}_1, \mathbf{B}_2]) \geq \frac{2}{M(\mathbf{D})} \tag{10}$$

resulting in a less restrictive sufficient condition for ℓ^0 uniqueness

$$card(S(\alpha)) < \frac{1}{M(\mathbf{D})}. \tag{11}$$

With a different technique, Elad and Bruckstein also obtained a sufficient condition

$$card(S) < \frac{\sqrt{2} - 1/2}{M(\mathbf{D})} \approx \frac{0.914}{M(\mathbf{D})} \tag{12}$$

for uniqueness of the ℓ^1 solution. In [11], Feuer and Nemirovsky proved that the latter condition cannot be relaxed.

Next we show that the result of Donoho and Huo indeed extends from the case of the union of two bases to the case of arbitrary dictionaries.

Theorem 1 ([9, 10]) *For any dictionary, we have the* generalized uncertainty principle

$$Z(\mathbf{D}) \geq 1 + 1/M(\mathbf{D}). \tag{13}$$

Moreover, if

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

then α is the unique *solution to the ℓ^0 and the ℓ^1 problems.*

5. UNIONS OF BASES

In this section we consider the special case where $\mathbf{D}$ is the 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, let us see that it is actually possible to have several orthonormal basis with small coherence factor $M(\mathbf{D})$. The proof of Theorem 2 can be found in [12, 13].

Theorem 2 *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 incoherence property: $u, v \in \mathcal{D}, u \neq v \Rightarrow |\langle u, v\rangle| \in \{0, N^{-1/2}\}$.

Let us now state the main results for dictionaries that are the union of orthonormal bases. First, we get another lower bound on $Z(\mathbf{D})$, another *generalized uncertainty principle*.

Lemma 2 ([9]) *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 $x \neq 0$. Then*

$$\sum_{l=1}^{L} \frac{1}{1 + M(\mathbf{D})\|x^l\|_0} \leq L - 1. \tag{15}$$

Consequently

$$Z(\mathbf{D}) \geq \big(1 + \frac{1}{L-1}\big)\frac{1}{M(\mathbf{D})}. \tag{16}$$

For $L = 2$ the condition (15) can be rewritten $\sqrt{\|x^1\|_0\|x^2\|_0} \geq 1/\sqrt{M}$ as in [6, Th. 1], and our generalized uncertainty principle (16) corresponds to that of Elad and Bruckstein (11). So, the above result recovers the result of Elad and Bruckstein for $L = 2$ and extends them to $L \geq 3$.

In a sense the uncertainty principle (16) is sharp for $L = 2$, because there are example of pairs of bases with $Z = 2/M$ [5, 6]. For L strictly larger than $1 + 1/M$, (16) is certainly not sharp since it is strictly weaker than the general lower estimate (13). For $3 \leq L \leq 1 + 1/M$, it is not known whether (16) is sharp.

Combining Lemma 2 and Corollary 1, we can check uniqueness of an ℓ^0 solution using the sufficient condition

$$card(S(\alpha)) < \Big(\frac{1}{2} + \frac{1}{2(L-1)}\Big)\frac{1}{M(\mathbf{D})}. \tag{17}$$

When $L \leq 1 + 1/M$, this sufficient condition is less restrictive than the one for general dictionaries (14). However, this less restrictive condition does not seem to be sufficient to ensure uniqueness of as an ℓ^1 solution. Instead we have the following result.

Theorem 3 ([9]) *Let $\mathbf{D}$ a union of L orthonormal bases. Denote $\alpha = \begin{bmatrix} \alpha^1 \\ \ldots \\ \alpha^L \end{bmatrix}$ 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 $\|\alpha^1\|_0 \leq \ldots \leq \|\alpha^L\|_0$. If*

$$\sum_{l \geq 2} \frac{M\|\alpha^l\|_0}{1 + M\|\alpha^l\|_0} < \frac{1}{2(1 + M\|\alpha^1\|_0)}. \tag{18}$$

then α is the unique *solution to the ℓ^1 problem. In the more restricted case where*

$$card(S(\alpha)) < \Big(\sqrt{2} - 1 + \frac{1}{2(L-1)}\Big)\frac{1}{M(\mathbf{D})} \tag{19}$$

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

The sufficient conditions (19) and (17) are very similar but differ by a gap $1/2 - (\sqrt{2} - 1) \approx 0.086$ in the constant in front of $1/M$. This is just the same as the difference between the conditions (12) and (11) proved by Elad and Bruckstein in the case of two bases. For $L = 2$, Feuer and Nemirovsky [11] proved that this gap cannot be bridged. If the same holds true for all $L \geq 3$, then the only values of L for which the new sufficient condition (19) is less restrictive than the general one of Theorem 1 are $2 \leq L \leq 6$, with some possible restrictions depending on the value of M.

6. CONCLUSION

In this paper we have considered two theoretical problems of sparse decomposition of signals or images in a redundant dictionary. We have shown that when the solution of the (combinatorial) ℓ^0 problem is very sparse, it is unique and coincides with that of the ℓ^1 problem, which can be found in polynomial time by linear programming.

In a given incoherent dictionary $\mathbf{D}$, most real signals or images y do not have an exact representation that is sparse enough to satisfy the sparsity constraints of our theorems.

However, they can be *approximated* by a linear combination y_m of m vectors from $\mathbf{D}$ [14]. The more redundant the dictionary, the more likely it is that any y will have good sparse approximations. However, the relation between sparsity and goodness of nonlinear approximations in redundant dictionaries is only partially understood yet [15].

For each class of signals/images, one may need to "learn" different dictionaries that yield sparse representations [16, 17] and it is not clear whether these dictionaries are close to be perfectly incoherent.

It is also an important practical issue to build efficient approximation algorithms in such redundant dictionaries. For incoherent dictionaries and $m \ll M(\mathbf{D})$, recent results [18] show that it is possible to compute in polynomial time a near-best m-term approximation $y_m^\star$ to any y. Another approach is to replace the linear programming problem (2) by a quadratic program (QP) for which Fuchs [19] obtained optimality results similar to those presented in this paper. When $\mathbf{D}$ is a finite union of bases, QP can be solved by a relaxation method [20], but the rate of convergence seems unknown. It would only seem natural that the rate of convergence depends on the coherence $M(\mathbf{D})$ of the dictionary.

7. REFERENCES

[1] D.L. Donoho, "Denoising by soft-thresholding," *IEEE Trans. Inform. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.

[2] Ronald A. DeVore, Björn Jawerth, and Vasil Popov, "Compression of wavelet decompositions," *Amer. J. Math.*, vol. 114, no. 4, pp. 737–785, 1992.

[3] M. Zibulevsky and B.A. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary," *Neural Computations*, vol. 13, no. 4, pp. 863–882, 2001.

[4] S. Mallat and Z. Zhang, "Matching pursuit with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[5] D.L. Donoho and Xiaoming Huo, "Uncertainty principles and ideal atomic decompositions," *IEEE Trans. Inform. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.

[6] M. Elad and A.M. Bruckstein, "A generalized uncertainty principle and sparse representations in pairs of bases," *IEEE Trans. Inform. Theory*, vol. 48, no. 9, pp. 2558–2567, Sept. 2002.

[7] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *Constr. Approx.*, vol. 13, no. 1, pp. 57–98, 1997.

[8] S. Chen, D.L. Donoho, and M.A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, Jan. 1999.

[9] R. Gribonval and M. Nielsen, "Sparse decompositions in unions of bases," Tech. Rep. 1499, IRISA, Nov. 2002, submitted to IEEE Trans. Inf. Th.

[10] D.L. Donoho and M. Elad, "Optimally sparse representation in general (non-orthogonal) dictionaries via ℓ^1 minimization," *Proc. Nat. Aca. Sci.*, Dec. 2002, submitted.

[11] A. Feuer and A. Nemirovsky, "On sparse representations in pairs of bases," *IEEE Trans. Inform. Theory*, Nov. 2002, accepted.

[12] 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)*, vol. 75, no. 2, pp. 436–480, 1997.

[13] T. Strohmer and R. Heath, "Grassmannian frames with applications to coding and communications," Tech. Rep., 2002, Preprint, submitted to Appl.Comp.Harm.Anal.

[14] Ronald A. DeVore, "Nonlinear approximation," in *Acta numerica, 1998*, pp. 51–150. Cambridge Univ. Press, Cambridge, 1998.

[15] R. Gribonval and M. Nielsen, "Nonlinear approximation with dictionaries. II. Inverse estimates," Tech. Rep., IMI, Dept of Mathematics, University of South Carolina, Columbia, SC 29208, 2002, in preparation.

[16] D.J. Field and B.A. Olshausen, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, pp. 607–609, 1996.

[17] M.S. Lewicki, "Efficient coding of natural sounds," *Nature Neuosci.*, vol. 5, no. 4, pp. 356–363, 2002.

[18] A.C. Gilbert, S. Muthukrishnan, and M.J. Strauss, "Approximation of functions over redundant dictionaries using coherence," in *SIAM Symposium on Discrete Algorithms (SODA'03)*, 2003.

[19] J.-J. Fuchs, "On sparse representations in arbitrary redundant bases," Tech. Rep., IRISA, Dec. 2002, submitted to IEEE Trans. Inf. Th.

[20] S. Sardy, A. Bruce, and P. Tseng, "Block coordinate relaxation methods for nonparametric wavelet denoising," *Journal of Computational and Graphical Statistics*, vol. 9, no. 2, 2000.