HAL
archives-ouvertes.fr

Highly sparse representations from dictionaries are unique and independent of the sparseness measure

Rémi Gribonval, Morten Nielsen

▶ To cite this version:

Rémi Gribonval, Morten Nielsen. Highly sparse representations from dictionaries are unique and independent of the sparseness measure. Applied and Computational Harmonic Analysis, Elsevier, 2007, 22 (3), pp.335–355. <10.1016/j.acha.2006.09.003>. <inria-00544779>

HAL Id: inria-00544779
https://hal.inria.fr/inria-00544779

Submitted on 7 Feb 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

© 2007 Elsevier. This is the author version of an article published in an Elsevier journal. The original publication is available at www.sciencedirect.com with DOI: 10.1016/j.acha.2006.09.003

HIGHLY SPARSE REPRESENTATIONS FROM DICTIONARIES ARE UNIQUE AND INDEPENDENT OF THE SPARSENESS MEASURE

R. GRIBONVAL AND M. NIELSEN

Abstract. The purpose of this paper is to study sparse representations of signals from a general dictionary in a Banach space. For so-called localized frames in Hilbert spaces, the canonical frame coefficients are shown to provide a near sparsest expansion for several sparseness measures. However, for frames which are not localized, this no longer holds true and sparse representations may depend strongly on the choice of the sparseness measure. A large class of admissible sparseness measures is introduced, and we give sufficient conditions for having a unique sparse representation of a signal from the dictionary w.r.t. such a sparseness measure. Moreover, we give sufficient conditions on a signal such that the simple solution of a linear programming problem simultaneously solves all the non-convex (and generally hard combinatorial) problems of sparsest representation of the signal w.r.t. arbitrary admissible sparseness measures.

Sparse representation, redundant dictionary, sparseness measure, localized frame, incoherent dictionary, linear programming, non-convex optimization.

1. Introduction

Given a redundant signal (or image) dictionary, every signal or image y has infinitely many possible representations, and it is common to choose one according to some *sparseness measure*. When the dictionary is indeed a (Schauder) basis of the underlying Banach space, each signal has a unique representation and it does not matter which sparseness measure is used. However, in the redundant case, it is not clear when the sparsest representation is unique and how it is influenced by the choice of the sparseness measure.

A **dictionary** for a separable Banach space X is a family of unit vectors $\{g_k\}_{k\in K}$ with dense span in X (K is a finite or countable index set). In the finite dimensional case, when $X = \mathbb{R}^N$ or $X = \mathbb{C}^N$, we can think of g_k as the k-th column of the $N \times K$ matrix $\mathbf{D}$, and any vector $y \in X$ has at least one representation $y = \mathbf{D}x$ with coefficient vector $x \in \mathbb{R}^K$ (resp. $x \in \mathbb{C}^K$), and whenever $\mathbf{D}$ is **redundant** ($K > N$), y has infinitely many such representations. Among this infinite number of possible representations, it is often desirable to choose one with the *sparseness* property, however there are several notions of sparseness such as the ℓ^0 and ℓ^1 sparseness measures $\|x\|_0 := \sharp\{k \ : \ x_k \neq 0\}$ and $\|x\|_1 := \sum_k |x_k|$. If $\|x\|_f$ denotes some other "sparseness measure" (we will define and study several sparseness measures in this paper), one can consider the optimization problem for the "**f-sparsest representation**"

$$\text{Minimize } \|x\|_f \text{ subject to } y = \sum_k x_k g_k. \tag{1}$$

For this class of optimization problems, two rather natural questions arise

1/ when is the f-sparsest representation of y unique?

This work is supported in part by the European Union's Human Potential Programme, under contract HPRN-CT-2002-00285 (HASSIP) and in part by the Danish Technical Science Foundation, Grant no. 9701481.

© 2007 Elsevier. This is the author version of an article published in an Elsevier journal. The original publication is available at www.sciencedirect.com with DOI: 10.1016/j.acha.2006.09.003

2/ if it is unique, how much does it depend on the choice of the sparseness measure f?

The main purpose of this paper is to show that when a signal y has a very sparse representation (in the sense of the ℓ^0 "norm" which measures the total number of nonzero coefficients) then this representation is simultaneously the unique sparsest representation for any sparseness measure in a fairly large class $\mathcal{M}$ which we define in Section 2. In particular, the standard ℓ^1 sparseness measure belongs to this class and we show that, when y has a very sparse representation, solving the ℓ^1 minimization problem indeed simultaneously solves all the nonlinear/combinatorial optimization problems corresponding to sparseness measures in this class.

One of the most common criteria to select a representation is the energy *i.e.*, the representation which is often chosen is the one with the smallest ℓ^2 norm. When the dictionary is a frame in a Hilbert space, the minimum energy representation is nicely expressed in terms of analysis coefficients, *i.e.*, inner products of the signal with the canonical dual frame. For *localized* frames (such as wavelet frames or Gabor frames), the canonical frame coefficients provide a representation which, in addition to being generally easy to compute, is *nearly the sparsest one* for many sparseness measures besides the ℓ^2 norm. However, for *incoherent* frames (such as the union of a wavelet basis and a Wilson basis, or perhaps more simply, the union of the Dirac and the Fourier systems), the representation of minimum energy can be far from optimal, and one has to consider alternate strategies to select a "good" representation.

In the early 1990's, the Matching Pursuit and Basis Pursuit strategies were introduced with the purpose of getting good representations of signals with dictionaries where the frame representation was not satisfying. Soon, it was experimentally noticed that, when y has a sufficiently sparse expansions (in the sense of the ℓ^0 "norm") in the Dirac/Fourier dictionary, Basis Pursuit can exactly recover it. In a series of recent results [19, 20, 21, 14, 15, 17, 27, 13, 26], the experimental observation was turned into a theorem and extended to unions of "incoherent" bases as well as to more general "incoherent" dictionaries. Theorems in the same spirit were also recently proved, under slightly stronger assumptions, for exact recovery with Matching Pursuits [22, 23, 45, 29]. The Basis Pursuit results have essentially the following flavor : if y has a sufficiently sparse expansion x (in the sense of the ℓ^0 "norm"), then x is *simultaneously* the *unique* ℓ^0-sparsest and ℓ^1-sparsest representation of y, thus it can be recovered through linear programming [3, 41], which solves the ℓ^1-minimization problem.

In between the ℓ^0 and the ℓ^1 sparseness measures lie the ℓ^τ "norms" and it seemed only natural that by some sort of "interpolation", the Basis Pursuit results should extend to *simultaneous uniqueness* of the ℓ^τ-sparsest representations for y with a sufficiently sparse representation. It turns out that the interpolation can be done and extends to a much larger setting. Letting Σ_m be the set of all m-term expansions from the dictionary $\mathbf{D}$, that is to say

$$\Sigma_m := \{y = \mathbf{D}x, \|x\|_0 \leq m\} \tag{2}$$

one can define a characteristic number of the dictionary $m_1(\mathbf{D})$ as the supremum of all integers m with the following property :

for every element $y \in \Sigma_m$, any representation $y = \mathbf{D}x$ with $\|x\|_0 \leq m$ is the (unique) sparsest ℓ^1-representation of y; that is, there is no other representation x' of y with smaller (or equal) ℓ^1-norm of the coefficient sequence $\|x'\|_1 \leq \|x\|_1$.

© 2007 Elsevier. This is the author version of an article published in an Elsevier journal. The original publication is available at www.sciencedirect.com with DOI: 10.1016/j.acha.2006.09.003

The main result of our paper is the following theorem, which generalizes naturally the recent series of Basis Pursuit results [19, 20, 21, 14, 15, 17, 27, 13, 26].

Theorem 1. *Let* $\mathbf{D}$ *be an arbitrary dictionary in a separable finite or infinite dimensional Banach space and* $m \leq m_1(\mathbf{D})$ *an integer. Then, for any* $y \in \Sigma_m$ *and any representation* $y = \mathbf{D}x$ *with* $\|x\|_0 \leq m$, x *is simultaneously the (unique)* f*-sparsest representation of* y *for all sparseness measures* $f \in \mathcal{M}$. *In particular it is the* ℓ^τ*-sparsest representation for all* $0 \leq \tau \leq 1$.

Thus, if y has a *highly sparse representation* x (with at most $m_1(\mathbf{D})$ elements from the dictionary), this representation must indeed be the f-sparsest representation for *all* sparseness measures. The interesting consequence is that the combinatorial/highly non-linear search for the highly sparse representation of such vectors y can be replaced with a polynomial time computation based on linear programming [3, 41], which solves the ℓ^1-optimization problem.

The problem of finding a sparse signal representation from a redundant dictionary has been shown to be important in many application fields, from signal coding to blind source separation or signal denoising. In practice, the assumption that the analyzed signal y has an *exact* highly sparse representation is too strong. It is more reasonable to only assume that y can be well *approximated* by such a highly sparse expansion. From this point of view, the results in this paper on the uniqueness properties of exact (highly) sparse representations and the way they depend on the sparseness measure are not direct applicable. However, the analysis of the ideal case which is carried out here is a necessary first step, and we address the more realistic approximate case in a follow-up paper [25].

The structure of the paper is as follows. In Section 2, we define and study some properties of a fairly large class $\mathcal{M}$ of admissible sparseness measures which include the ℓ^τ "norms" ($0 \leq \tau \leq 1$) as a special case. In Section 3 we consider frames in infinite dimensional Hilbert space. We compare the ℓ^τ sparseness of the *sparsest synthesis coefficients* with that of the frame representation obtained through the *canonical analysis coefficients*. For *localized* frames, the two are shown to be equivalent up to constants. But for *incoherent* frames we illustrate the fact that the canonical analysis coefficients do not provide sparse representations. At the end of the section, we briefly discuss the proper definition of sparse representations in arbitrary dictionaries which may not be frames. In Section 4 we give some general conditions under which any expansion $y = \mathbf{D}_I x$ from a given sub-dictionary $\mathbf{D}_I := \{g_k, k \in I\}$ is bound to be the unique f-sparsest representation of y in the whole dictionary. The general conditions depend on the sparseness measure $f \in \mathcal{M}$ as well as the index set $I \subset K$ which corresponds to the sub-dictionary. We illustrate with an example the fact that for a given I, the conditions may be satisfied for some sparseness measure $f \in \mathcal{M}$ and violated for some other one $g \in \mathcal{M}$. In Section 5 we prove our main theorems and obtain necessary and sufficient conditions $\mathrm{card}(I) \leq m_f(\mathbf{D})$ on the *cardinality* of the sub-dictionary $\mathbf{D}_I$ which ensure that for all sparseness measures $g \in \mathcal{M}$ "between" a given $f \in \mathcal{M}$ and the ℓ^0 sparseness measure, the representation is unique and independent of g. We conclude this paper in Section 6 with a focus on dictionaries in Hilbert spaces, for which we discuss concrete estimates of the numbers $m_f(\mathbf{D})$ which appear in the "highly sparse" conditions. The estimates depend on the structure of the dictionary and are essentially based on properties of its Gram matrix.

© 2007 Elsevier. This is the author version of an article published in an Elsevier journal. The original publication is available at www.sciencedirect.com with DOI: 10.1016/j.acha.2006.09.003

2. Sparseness measures and sparse representations

In this section, we introduce the class $\mathcal{M}$ of admissible sparseness measures which will be used throughout this paper and study some of its important properties which we will need later on. We will discuss a bit later the motivation for its definition. At the end of the section, we will show that the notion of a f-sparse representation $y = \mathbf{D}x$ is well defined even when X is a separable infinite dimensional Banach space, provided that $f \in \mathcal{M}$.

2.1. **A class of sparseness measures.**

Definition 1. *We let $\mathcal{M}$ the set of all non-decreasing functions $f : [0, \infty) \to [0, \infty)$, not identically zero, with $f(0) = 0$ and such that $t \mapsto f(t)/t$ is non-increasing on $(0, \infty)$.*

Examples of f in this class include the power functions $f_\tau(t) = t^\tau$, $0 \leq \tau \leq 1$ (we use the convention that $t^0 = 0$ if $t = 0$ and $t^0 = 1$, $t > 0$), and it is not difficult to check that every other non-identically zero, non-decreasing, concave function f with $f(0) = 0$ is in $\mathcal{M}$ too. It is straightforward to check that $\mathcal{M}$ is

- convex;
- stable by composition (if $f, g \in \mathcal{M}$ then $f \circ g \in \mathcal{M}$);
- stable when taking the minimum (if $f, g \in \mathcal{M}$ then $\min(f, g) \in \mathcal{M}$);
- stable when taking the maximum (if $f, g \in \mathcal{M}$ then $\max(f, g) \in \mathcal{M}$).

Therefore, it must also contains non-concave functions such as

$$f(t) := \max(t/2, \min(t, t^0)) = \begin{cases} t, & 0 \leq t \leq 1 \\ 1, & 1 \leq t \leq 2 \\ t/2, & 2 \leq t < \infty \end{cases} . \tag{3}$$

By analogy with the ℓ^τ 'norms' we define for $f \in \mathcal{M}$ and any sequence $x = (x_k)_{k \in K}$ (where K is a finite or countable index set) the "f-norm"

$$\|x\|_f := \sum_k f(|x_k|). \tag{4}$$

By abuse of notation we will denote $\|x\|_\tau = \|x\|_{f_\tau} = \sum_k |x_k|^\tau$, $0 \leq \tau \leq 1$, and similar abuses of notation will be made throughout this paper with other quantities which depend on f.

In the definition of the class $\mathcal{M}$ of admissible sparseness measures, we impose several properties on f. Most of them are rather natural, because we want the f-sparsest representation of any signal y, in the sense of Eq. (1), to have few large components and most components concentrated around zero. In order to favor small components rather than large ones, it is only natural to impose that f be non-decreasing, and we need $f(0) = 0$ to ensure that the series $\sum_k f(|x_k|)$ is summable for some sequences (x_k) with an infinite number of entries. The condition $t \mapsto f(t)/t$ non-increasing is perhaps less intuitive. Besides being technically necessary to get the most important results of this paper (see Section 5), it also implies that $d(x, y) := \|x - y\|_f$ defines a metric on the underlying vector space[1]. To see that, we merely have to check the triangle inequality (we let the reader check that the other axioms of metrics are trivially satisfied), which is given by the following proposition.

[1]Note that $\|x\|_f$ can be a norm only if we have $f(\lambda x) = \lambda f(x)$, which implies $f(t) \propto t$, in which case we get a multiple of the ℓ^1 norm.

© 2007 Elsevier. This is the author version of an article published in an Elsevier journal. The original publication is available at www.sciencedirect.com with DOI: 10.1016/j.acha.2006.09.003

Proposition 1. *$\mathcal{M}$ is* strictly *contained in the set $\mathcal{S}$ of non-decreasing functions $f : [0,\infty) \to [0,\infty)$, not identically zero, with $f(0) = 0$ which are* sub-additive.

Proof. Let us first show that for any $f \in \mathcal{M}$ we have a triangle inequality

$$f(|u+v|) \leq f(|u|+|v|) \leq f(|u|) + f(|v|). \tag{5}$$

It will follow that we have the claimed inclusion, and we will show later that it is strict. The leftmost inequality comes from the fact that $|u+v| \leq |u|+|v|$ and that f is non-decreasing. Because $f(t)/t$ is non-increasing, we easily derive $f(|u|) \geq f(|u|+|v|) \cdot |u|/(|u|+|v|)$ and $f(|v|) \geq f(|u|+|v|) \cdot |v|/(|u|+|v|)$, and we obtain the rightmost inequality by summation. To see that the inclusion is strict, we will consider a simple example, which was kindly pointed out to us by G. Gribonval. Denoting $\lfloor t \rfloor$ the largest integer such that $\lfloor t \rfloor < t \leq \lfloor t \rfloor + 1$, we consider

$$f(t) := \begin{cases} 0, & t = 0 \\ 1 + \lfloor t \rfloor, & t > 0 \end{cases}. \tag{6}$$

For $u, v > 0$ we have $\lfloor u+v \rfloor \leq \lfloor u \rfloor + \lfloor v \rfloor + 1$ and it follows that $f(u+v) \leq f(u) + f(v)$. The same inequality is trivial if u and/or v is zero, hence $f \in \mathcal{S}$. However, because f has strictly positive jumps at the positive integers, so does $f(t)/t$, hence $f \notin \mathcal{M}$. □

Remark 1. *Though it will not be used in this paper, it is interesting to notice that any sparseness measure $f \in \mathcal{M}$ is continuous on $(0,\infty)$. To see that, simply notice that since $f(t)$ is non-decreasing, $f(t^-)$ and $f(t^+)$ are well defined and satisfy $f(t^-) \leq f(t^+)$ for every $t > 0$. But since $f(t)/t$ is non-increasing, we also have $f(t^-) \geq f(t^+)$. As a result, any sparseness measure f which is continuous at zero is indeed a modulus of continuity as defined in* [11, Chapter 2.6, p.41].

Remark 2 (Statistical interpretation)**.** *In finite dimension, the f-sparsest representation problem* (1) *is related to the statistical problem of Bayesian estimation of unknown parameters (x_k) given the noiseless observation $y = \mathbf{D}x$ and the prior probability density function $P_h(x) = \frac{1}{Z_h} \exp\left(-h(\|x\|_f)\right)$, where Z_h is a normalizing constant and $h : [0,\infty) \to [0,\infty)$ is an increasing function such that $P_h(x)$ is a well-defined probability density on $\mathbb{R}^K$ (resp. $\mathbb{C}^K$). In the Bayesian interpretation, the fact that $f(t)/t$ is non-increasing is related to the marginals densities $P(x_k)$ being sharply "peaked" at zero: this is satisfied,* e.g.*, for the generalized Gaussians $P(x_k) \propto \exp(-|x_k|^\tau)$, $0 < \tau \leq 1$, which include the Laplacian ($\tau = 1$). However, smoothed versions of the Laplacian such as $f(t) = t - \log(1+t)$ –which are sometimes used to make numerical optimization algorithms more robust, see* e.g. [46] *– do not generally satisfy this property around zero. Interestingly, as soon as the function $h(u)/u$ is not constant,e the probability density function P_h is non separable and corresponds to non independent random variables (x_k).*

2.2. **Sparse representations in infinite dimension.** When X is a separable infinite dimensional Banach space, it is not always clear when the notion of *representation* $y = \mathbf{D}x$ makes sense, because the convergence of the series $\sum_k x_k g_k$ might not be unconditional. When X is a Hilbert space and $\mathbf{D}$ is actually a *frame* for X, we know the series is unconditionally convergent as soon as $x \in \ell^2$, but for general dictionaries and Banach spaces it is not necessarily the case. Nevertheless, thanks to properties of the sparseness class defined above, *sparse* representations (with $\|x\|_f < \infty$) are well defined even when the dictionary has no special structure. We have the following result for elements $y \in X$ which have an *f-sparse representation* from $\mathbf{D}$ for some sparseness measure $f \in \mathcal{M}$.

© 2007 Elsevier. This is the author version of an article published in an Elsevier journal. The original publication is available at www.sciencedirect.com with DOI: 10.1016/j.acha.2006.09.003

Lemma 1. *Let $y \in X$ and assume there exists $x \in \ell^\infty$ with $\|x\|_f < \infty$, for some $f \in \mathcal{M}$ such that, for some enumeration $\phi(k)$ of the infinite index set K,*

$$\|y - \sum_{k=1}^{n} x_{\phi(k)} g_{\phi(k)}\| \to 0.$$

Then $\|x\|_1 < \infty$ and $y = \mathbf{D}x = \sum_k x_k g_k$ where the series is unconditionally convergent.

Proof. Since $f \in \mathcal{M}$, $f(t)/t$ is non-increasing and we have, for all k,

$$f(|x_k|)/|x_k| \geq f(\|x\|_\infty)/\|x\|_\infty = c(x) > 0.$$

Hence

$$\sum_k |x_k| \leq \frac{1}{c(x)} \sum_k f(|x_k|) < \infty$$

and, because $\|g_k\| = 1$, we can conclude that the series $\sum_k x_k g_k$ is absolutely convergent. □

It follows that if y admits *at least* one representation x with $\|x\|_f < \infty$, we can consider the f-sparsest representation problem (1) just as in the finite dimensional case, and the same natural questions arise. Is the representation unique? Does it depend on the choice of the sparseness measure?

3. Sparseness of frame representations

Frames are perhaps the most widely studied family of signal or image dictionaries. A dictionary $\mathbf{D} = \{g_k\}_k$ in a Hilbert space $\mathcal{H}$ is called a **frame** if there exist two constants $0 < A, B < \infty$ such that, for any $y \in \mathcal{H}$, $A\|y\|^2 \leq \sum_k |\langle y, g_k\rangle|^2 \leq B\|y\|^2$. If $A = B$ then $\mathbf{D}$ is called a tight frame. Equivalently, $\mathbf{D}$ is a frame if the **synthesis operator**

$$\begin{array}{rccl} \mathbf{D}: & \ell^2 & \to & \mathcal{H} \\ & x & \mapsto & \mathbf{D}x := \sum_k x_k g_k \end{array} \tag{7}$$

is bounded and onto. To every frame corresponds a canonical dual frame $\widetilde{\mathbf{D}} = \{\widetilde{g_k}\}_k$ which is also a frame such that for every $y \in \mathcal{H}$:

$$y = \sum_k \langle y, g_k\rangle \, \widetilde{g_k} = \sum_k \langle y, \widetilde{g_k}\rangle \, g_k = \mathbf{D}\widetilde{\mathbf{D}}^\star y. \tag{8}$$

The sequence $\{\langle y, \widetilde{g_k}\rangle\}_k = \widetilde{\mathbf{D}}^\star y$ (where $(\cdot)^\star$ denotes the adjoint of an operator) is called the (canonical) frame representation of y and is obtained through the **(canonical) analysis operator** $\widetilde{\mathbf{D}}^\star$. For tight frames we have $\widetilde{\mathbf{D}} = \frac{1}{A}\mathbf{D}$. Among all possible representations $y = \mathbf{D}x$, the canonical frame representation is the one with minimum energy, *i.e.*,

$$\widetilde{\mathbf{D}}^\star y = \arg \min_{x|\mathbf{D}x=y} \|x\|_2. \tag{9}$$

However, in many signal and image processing applications, it is known that the energy $\|x\|_2$ of a representation might not be the most appropriate criterion to select a "good" representation. It is now well understood that "sparse" representations [38] might do a much better job for applications as diverse as compression [10] , feature extraction [18, 34] or blind source separation [46, 24]. Whether the frame representation provides a "sparse enough" representation or not, where the "sparseness" is measured, *e.g.*, with some ℓ^τ norm for $\tau < 2$, is thus a theoretical problem which has a practical impact. In this section we try to address this problem. With this aim, we start by recalling/defining two families of **sparseness classes** defined respectively in terms of sparseness of the *optimal*

© 2007 Elsevier. This is the author version of an article published in an Elsevier journal. The original publication is available at www.sciencedirect.com with DOI: 10.1016/j.acha.2006.09.003

synthesis coefficients and of the canonical frame representation. We study the conditions which ensure equality of these two families and discuss some simple frames where the conditions are not satisfied.

3.1. **Sparseness classes.**

Definition 2. *Let* $\mathbf{D}$ *be a frame for* $\mathcal{H}$ *and* $0 \leq \tau \leq 2$. *We define the* **synthesis sparseness class**

$$\mathcal{K}^\tau(\mathbf{D}) := \left\{ y \in \mathcal{H} : \exists x \in \ell^2,\ \|x\|_\tau < \infty, y = \mathbf{D}x \right\} \tag{10}$$

and the **analysis sparseness class**

$$\mathcal{H}^\tau(\mathbf{D}) := \left\{ y \in \mathcal{H} : \|\widetilde{\mathbf{D}}^\star y\|_\tau < \infty \right\}. \tag{11}$$

We define a (quasi)norm on $\mathcal{K}^\tau(\mathbf{D})$ *as follows*

$$|y|_{\mathcal{K}^\tau(\mathbf{D})} := \inf_{x|y=\mathbf{D}x} \|x\|_\tau. \tag{12}$$

For $\mathcal{H}^\tau(\mathbf{D})$ *we define*

$$|y|_{\mathcal{H}^\tau(\mathbf{D})} := \|\widetilde{\mathbf{D}}^\star y\|_\tau. \tag{13}$$

Clearly, we have the inclusion $\mathcal{H}^\tau(\mathbf{D}) \subset \mathcal{K}^\tau(\mathbf{D})$ together with the (quasi)norm inequality $|y|_{\mathcal{K}^\tau(\mathbf{D})} \leq |y|_{\mathcal{H}^\tau(\mathbf{D})}$. However, it is not generally clear if the reversed inclusion and the corresponding reversed inequality holds: the freedom on the synthesis coefficients –which comes from the redundancy of the frame– might make it possible for some y to get a much sparser representation than the canonical frame expansion.

3.2. **Conditions for equality between sparseness classes; localized frames.** As $\mathcal{K}^\tau(\mathbf{D}) = \mathbf{D}\ell^\tau$ (see [28]), it is not difficult to see that a necessary and sufficient condition to get $\mathcal{H}^\tau(\mathbf{D}) = \mathcal{K}^\tau(\mathbf{D})$ is that the operator

$$\begin{array}{rccc} \widetilde{\mathbf{D}}^\star\mathbf{D}: & \ell^2 & \to & \ell^2 \\ & x & \mapsto & \widetilde{\mathbf{D}}^\star\mathbf{D}x \end{array} \tag{14}$$

map continuously ℓ^τ into ℓ^τ. We know that, since $\mathbf{D}$ is a frame, $\widetilde{\mathbf{D}}^\star\mathbf{D}$ does map ℓ^2 boundedly into ℓ^2. If the same holds true when $\tau = 2$ is replaced with $\tau = 1$, then we get the same result for $1 \leq \tau \leq 2$ by the real or complex method of interpolation [2, 1], which simply corresponds to applying Schur's lemma. Fortunately, for $0 \leq \tau \leq 1$, there is an easy characterization of the operator norm from ℓ^τ to ℓ^τ.

Lemma 2. *Let* $\mathbf{T}$ *be a doubly infinite matrix which we may write columnwise* $\mathbf{T} = [\mathbf{T}_k]$. *For* $0 \leq \tau \leq 1$, *we have*

$$|||\mathbf{T}|||_\tau := \sup_{x \neq 0} \frac{\|\mathbf{T}x\|_\tau}{\|x\|_\tau} = \sup_k \|\mathbf{T}_k\|_\tau. \tag{15}$$

Proof. First, remember that in this paper we denote $\|x\|_\tau := \sum_k |x_k|^\tau$ for $0 \leq \tau \leq 1$. Using the (quasi)triangle inequality in ℓ^τ, $0 \leq \tau \leq 1$, we obtain the result using the following inequalities

$$\begin{aligned} \|\mathbf{T}x\|_\tau &= \|\sum_k x_k \mathbf{T}_k\|_\tau \leq \sum_k \|x_k \mathbf{T}_k\|_\tau \\ &\leq \sum_k |x_k|^\tau \|\mathbf{T}_k\|_\tau \leq \|x\|_\tau \cdot \sup_k \|\mathbf{T}_k\|_\tau. \end{aligned}$$

© 2007 Elsevier. This is the author version of an article published in an Elsevier journal. The original publication is available at www.sciencedirect.com with DOI: 10.1016/j.acha.2006.09.003

□

As a consequence, we obtain the following characterization.

Lemma 3. *Let* $\mathbf{D}$ *be a frame,* $\widetilde{\mathbf{D}}$ *be its canonical dual frame, and* $0 \leq \tau \leq 1$. *The synthesis and analysis sparseness classes* $\mathcal{K}^\eta(\mathbf{D})$ *and* $\mathcal{H}^\eta(\mathbf{D})$ *coincide for all* $\tau \leq \eta \leq 2$ *if, and only if*

$$\sup_k \sum_l |\langle g_k, \widetilde{g}_l \rangle|^\tau < \infty. \tag{16}$$

Closely connected to the above condition are two properties of frames which were recently defined and studied by K. Gröchenig [32, 31] and seem shared by a number of classical frames.

Definition 3 (Gröchenig [32])**.** *Let* $\mathbf{D} = \{g_k\}_{k\in K}$ *be a frame for* $\mathcal{H}$*, and let* $\mathbf{B} = \{e_n\}_{n\in\mathcal{N}}$ *be a Riesz basis for* $\mathcal{H}$ *with dual system* $\widetilde{\mathbf{B}} = \{\widetilde{e}_n\}_{n\in\mathcal{N}}$. *Assume that both* K *and* $\mathcal{N}$ *are* separated *index sets in* $\mathbb{R}^d$, i.e., $\inf_{k,l\in K:k\neq l} |k-l| \geq \delta > 0$, *and likewise for* $\mathcal{N}$. *For* $s > 0$, *we say that* $\mathbf{D}$ *is* s*-localized with respect to* $\mathbf{B}$ *if*

$$\max\left(|\langle g_k, e_n\rangle|, |\langle g_k, \widetilde{e_n}\rangle|\right) \leq C(1+|k-n|)^{-s} \tag{17}$$

for all $k \in K$ *and* $n \in \mathcal{N}$, *where* $|\cdot|$ *denotes any of the equivalent norms on* $\mathbb{R}^d$. *We say that* $\mathbf{D}$ *is* intrinsically localized *with decay* $s > 0$ *if*

$$|\langle g_k, g_l\rangle| \leq C(1+|k-l|)^{-s} \tag{18}$$

for all $k, l \in K$.

Gröchenig proved that any localized frame is intrinsically localized, but it is not known so far whether the reciprocal is true or not. Many classical frames which come up in signal and image processing, such as Gabor frames [30] and wavelet frames [6, 35, 40, 39, 7, 5] are localized in the above sense, with $\mathbf{B}$ a Wilson or a wavelet basis. The authors suspect that curvelet frames [12] also have the localization property. For localized frames, K. Gröchenig kindly pointed out to us the following property, which somehow nicely extends Lemma 3 (note that Proposition 2 below can be extended to weighted sparseness classes using Gröchenig's theory).

Proposition 2 (Gröchenig)**.** *Assume that* $\mathbf{D}$ *is an* $(s+d+\epsilon)$*-localized frame w.r.t.* $\mathbf{B}$ *for some* $\epsilon > 0$. *Let* $\widetilde{\mathbf{D}} = \{\tilde{g}_k\}_{k\in K}$ *be the dual frame to* $\mathbf{D}$. *We have, for any* $d/(s+d+\epsilon) < \tau < 2$,

$$\mathcal{K}^\tau(\mathbf{B}) = \mathcal{K}^\tau(\widetilde{\mathbf{B}}) = \mathcal{H}^\tau(\mathbf{D}) = \mathcal{K}^\tau(\widetilde{\mathbf{D}}) = \mathcal{H}^\tau(\widetilde{\mathbf{D}}) = \mathcal{K}^\tau(\mathbf{D}).$$

Proof. The result is a direct corollary of Gröchenig's main theorem on localized frames (see [32, Theorem 3.5]), which in particular implies that the dual frame is also $(s+d+\epsilon)$-localized. We will prove the result directly for $d/(s+d+\epsilon) < \tau \leq 1$ and let the reader check that interpolation (Schur's test) can be used to extend it to $1 < \tau < 2$. First we show that $\mathcal{K}^\tau(\mathbf{B}) \subset \mathcal{H}^\tau(\mathbf{D})$. Let $f \in \mathcal{K}^\tau(\mathbf{B})$. Then $f = \sum c_n e_n$ with $\|\{c_n\}\|_{\ell^\tau} < \infty$, and we easily get $\langle f, \tilde{g}_k\rangle = \sum_{n\in\mathcal{N}} c_n \langle e_n, \tilde{g}_k\rangle$. Since we assume $(s+d+\epsilon)\tau > d$, we have by [32, Lemma 2.1]

$$\sup_n \sum_{k\in K} (1+|k-n|)^{-(s+d+\epsilon)\tau} \leq C_{(s+d+\epsilon)\tau} < \infty$$

and the bi-infinite matrix $(\langle e_n, \tilde{g}_k\rangle)_{n\in\mathcal{N},k\in K}$ is thus bounded from $\ell^\tau(\mathcal{N})$ to $\ell^\tau(K)$. Hence we have $\{\langle f, \tilde{g}_k\rangle\}_{k\in K} \in \ell^\tau(K)$ and the claim follows. We have already noticed the trivial inclusion $\mathcal{H}^\tau(\mathbf{D}) \subset \mathcal{K}^\tau(\mathbf{D})$, and using the same argument as above we see that $\mathcal{K}^\tau(\mathbf{D}) \subset$

© 2007 Elsevier. This is the author version of an article published in an Elsevier journal. The original publication is available at www.sciencedirect.com with DOI: 10.1016/j.acha.2006.09.003

$\mathcal{K}^\tau(\mathbf{B})$. Hence we have the chain $\mathcal{K}^\tau(\mathbf{B}) \subset \mathcal{H}^\tau(\mathbf{D}) \subset \mathcal{K}^\tau(\mathbf{D}) \subset \mathcal{K}^\tau(\mathbf{B})$. We can repeat the above chain of arguments with $\widetilde{\mathbf{D}}$ in place of $\mathbf{D}$, since by Gröchenig's results it is also $s+d+\epsilon$)-localized. We obtain the inclusions $\mathcal{K}^\tau(\mathbf{B}) \subset \mathcal{H}^\tau(\widetilde{\mathbf{D}}) \subset \mathcal{K}^\tau(\widetilde{\mathbf{D}}) \subset \mathcal{K}^\tau(\mathbf{B})$, and finally we get $\mathcal{K}^\tau(\mathbf{B}) = \mathcal{H}^\tau(\mathbf{D}) = \mathcal{K}^\tau(\widetilde{\mathbf{D}}) = \mathcal{H}^\tau(\widetilde{\mathbf{D}}) = \mathcal{K}^\tau(\mathbf{D})$. To complete the proof, we just notice that $\mathbf{D}$ is $(s+d+\epsilon)$-localized w.r.t. $\widetilde{\mathbf{B}}$ and we can repeat the arguments above for the dual system $\widetilde{\mathbf{B}}$. □

It is not difficult to modify the above arguments to obtain a similar result for tight intrisically localized frames.

Proposition 3. *Assume that* $\mathbf{D}$ *is a* tight, intrinsically localized frame *with decay* $(s+d+\epsilon)$ *for some* $\epsilon > 0$. *Then, for any* $d/(s+d+\epsilon) < \tau < 2$, *we have* $\mathcal{H}^\tau(\mathbf{D}) = \mathcal{K}^\tau(\widetilde{\mathbf{D}})$.

3.3. **"Truly" redundant frames and incoherent dictionaries.** The above analysis shows that for classical frames –such as Gabor and wavelet frames– the canonical frame representation, which has minimum energy among all possible representations, also provides a "near ℓ^τ-sparsest" representation for several values of $\tau \neq 2$. For the above mentioned examples of localized frames, the reason for this good behavior of the frame representation is simply that by their very design, the frames are "close" to some orthonormal basis, so in a sense they are not so redundant.

Recently, there has been an increasing interest in signal and image representations based on "truly" redundant systems which are not so close to orthonormal bases. Typically, in order to get sparse representations of signals or images which display features of very different nature, it seems natural to combine several sufficiently different orthonormal bases into a redundant dictionary and to look for the sparsest representation. The approach –or variants thereof– has been used successfully with the Gabor multiscale dictionary [36], the union of a wavelet basis and a local Fourier basis for audio coding [8, 9], as well as with tight curvelet frames and local Fourier bases for image segmentation and source separation [42].

It turns out that with such "truly" redundant dictionaries, the frame representation no longer provides a near ℓ^τ-sparsest representation for any $\tau \neq 2$. Let us illustrate this fact with the example of $\mathbf{D} = [\mathbf{\Psi}, \mathbf{G}]$ the union of a "nice" wavelet frame $\mathbf{\Psi} = \{\psi^l_{j,k}\}$ and a "nice" Gabor frame $\mathbf{G} = \{g_{n,m}\}$ (the proof readily extends to any other pair of systems which give rise to two different sparseness classes). By nice we mean that they should be sufficiently localized w.r.t. a smooth wavelet basis with vanishing moments [10] (resp. w.r.t. a regular Wilson basis [33]) so that we can identify, with equivalent norms

$$\begin{aligned} \mathcal{K}^\tau(\mathbf{\Psi}) &= \mathcal{H}^\tau(\mathbf{\Psi}) = B^\alpha_{\tau,\tau}, \qquad \alpha = \frac{1}{\tau} - \frac{1}{2} \\ \mathcal{K}^\tau(\mathbf{G}) &= \mathcal{H}^\tau(\mathbf{G}) = M^{\tau,\tau} \end{aligned}$$

where $B^\alpha_{\tau,\tau}$ is a Besov space [44] and $M^{\tau,\tau}$ a modulation space [30]. It is quite obvious that when we consider the union of the two systems, we have

$$\begin{aligned} \mathcal{K}^\tau(\mathbf{D}) &= \mathcal{K}^\tau(\mathbf{\Psi}) + \mathcal{K}^\tau(\mathbf{G}) = B^\alpha_{\tau,\tau} + M^{\tau,\tau}, \\ \mathcal{H}^\tau(\mathbf{D}) &= \mathcal{H}^\tau(\mathbf{\Psi}) \cap \mathcal{H}^\tau(\mathbf{G}) = B^\alpha_{\tau,\tau} \cap M^{\tau,\tau}, \end{aligned}$$

and we conclude using the fact that $B^\alpha_{\tau,\tau} \neq M^{\tau,\tau}$ except for $\tau = 2$ (see, *e.g.*, discussions on the different behavior of Besov and modulation spaces w.r.t. dilation and modulation in [16], as well as embeddings of Besov spaces into modulation spaces in [37]).

© 2007 Elsevier. This is the author version of an article published in an Elsevier journal. The original publication is available at www.sciencedirect.com with DOI: 10.1016/j.acha.2006.09.003

As we have just seen, in "truly" redundant frames, one cannot hope to simply use the canonical frame representation in order to get a sparse representation. In the many applications where it is desirable to get a sparse representation of the data, it is thus necessary to explicitly look for a sparse representation, which corresponds to numerically solving an optimization problem such as (1). We have already mentioned that it is not clear then whether or not the problem has a unique solution and if the solution depends on the choice of the sparseness measure. In the next sections we try to answer these questions.

4. General Uniqueness Conditions

In this section we are going to provide some general sufficient conditions on a representation $y = \mathbf{D}x$ which ensure that x is the unique f-sparsest representation (resp. a f-sparsest representation) of y, where $f \in \mathcal{M}$ is an arbitrary admissible sparseness measure as defined in Section 3.1. The conditions depend only on the set of elements of the dictionary which are used in the representation, *i.e.*, they depend on properties of the **support** $I(x)$ of the coefficient vector $x = (x_k) \in \mathbb{R}^K$ (resp. $\mathbb{C}^K$) :

$$I(x) := \{k, x_k \neq 0\}. \tag{19}$$

The **kernel** of the dictionary will play a special role. In finite dimension, it is simply defined as

$$\mathrm{Ker}(\mathbf{D}) := \{z, \mathbf{D}z = 0\}. \tag{20}$$

In the infinite dimensional case, so as to avoid problems with convergence of infinite series, the useful definition will be

$$\mathrm{Ker}_f(\mathbf{D}) := \{z, \|z\|_f < \infty, \mathbf{D}z = 0\} \tag{21}$$

which relies on Lemma 1 to ensure unconditional convergence of $\mathbf{D}z = \sum_k z_k g_k$. In finite dimension, we have $\mathrm{Ker}_f(\mathbf{D}) = \mathrm{Ker}(\mathbf{D})$ for all sparseness measures, but in infinite dimension $\mathrm{Ker}_f(\mathbf{D}) \subset \mathrm{Ker}_1(\mathbf{D})$ may actually depend on f. When $\mathrm{Ker}_f(\mathbf{D}) = \{0\}$, it is straightforward to check that $y = \mathbf{D}x_1 = \mathbf{D}x_2$ with $\|x_i\|_f < \infty$ implies $x_1 = x_2$, therefore any representation with finite f-norm is the unique f-sparsest one.

For $f \in \mathcal{M}$, $I \subset K$ a set of indices and z with $0 < \|z\|_f < \infty$ we can now define

$$\theta_f(I, z) := \frac{\sum_{k \in I} f(|z_k|)}{\|z\|_f}. \tag{22}$$

Note that for the specific case of the ℓ^τ norms, i.e. for $f_\tau(t) = |t|^\tau$, we have $\theta_{f_\tau}(I, z) = \theta_{f_\tau}(I, \lambda z)$ for any scalar λ. However, for other functions f this is generally *not* the case.

By refining ideas from [14, 27] we have the following two lemmas.

Theorem 2. *Let* $\mathbf{D}$ *be an arbitrary dictionary in a separable Banach space* X *of finite or infinite dimension,* $f \in \mathcal{M}$ *be a sparseness measure and* $I \subset K$ *be an index set.*

(1) *If for all* $z \in Ker_f(\mathbf{D}) \backslash \{0\}$

$$\theta_f(I, z) < 1/2, \tag{23}$$

then, for all x, y *such that* $\|x\|_f < \infty$, $y = Dx$ *and* $I(x) \subset I$, x *is* the unique f*-sparsest representation of* y.

© 2007 Elsevier. This is the author version of an article published in an Elsevier journal. The original publication is available at www.sciencedirect.com with DOI: 10.1016/j.acha.2006.09.003

(2) *If for all* $z \in Ker_f(\mathbf{D})\backslash\{0\}$

$$\theta_f(I,z) \le 1/2, \tag{24}$$

then, for all x, y *such that* $\|x\|_f < \infty$, $y = Dx$ *and* $I(x) \subset I$, x *is* an f*-sparsest representation of* y.

The conditions given by Theorem 2 are sharp, which is shown by the next result.

Theorem 3. *Let* $\mathbf{D}$ *be an arbitrary dictionary in a separable Banach space* X *of finite or infinite dimension,* $f \in \mathcal{M}$ *a sparseness measure and* $I \subset K$ *an index set.*

(1) *If* $\theta_f(I,z) > 1/2$ *for some* $z \in Ker_f(\mathbf{D})\backslash\{0\}$*, then there exist* x *and* x' *such that* $\mathbf{D}x = \mathbf{D}x'$, $I(x) \subset I$ *and* $\|x'\|_f < \|x\|_f$.
(2) *If* $\theta_f(I,z) = 1/2$ *for some* $z \in Ker_f(\mathbf{D})\backslash\{0\}$*, then there exist* $x \neq x'$ *such that* $\mathbf{D}x = \mathbf{D}x'$, $I(x) \subset I$ *and* $\|x'\|_f = \|x\|_f$.

Proof of Theorem 2. (1) By assumption, x is a representation of y with $\|x\|_f < \infty$. If x' is another representation of y with $\|x'\|_f < \infty$, then by Lemma 1 $\|x\|_1 < \infty$, $\|x'\|_1 < \infty$. As $\mathbf{D}x = \mathbf{D}x'$ we actually have $z := x' - x \in \mathrm{Ker}_f(\mathbf{D})$. Thus, under the assumption $I(x) \subset I$, what we need to prove is that for all $z \in \mathrm{Ker}_f(\mathbf{D})$ with $z \neq 0$, $\sum_k f(|x_k + z_k|) > \sum_k f(|x_k|)$. This is equivalent to showing

$$\sum_{k\notin I} f(|z_k|) + \sum_{k\in I} \big(f(|x_k + z_k|) - f(|x_k|)\big) > 0. \tag{25}$$

From the triangle inequality (5), we derive the inequality $f(|x_k + z_k|) - f(|x_k|) \ge -f(|z_k|)$. Thus, we will get (25) if we can prove that for all $z \in \mathrm{Ker}_f(\mathbf{D})\backslash\{0\}$,

$$\sum_{k\notin I} f(|z_k|) - \sum_{k\in I} f(|z_k|) > 0,$$

or equivalently

$$\sum_{k\in I} f(|z_k|) < \frac{1}{2}\|z\|_f$$

which is exactly the assumption $\theta_f(I,z) < 1/2$.

(2) Copy the above line of arguments replacing strict inequalities with weak ones.

□

Remark 3. *In finite dimension, one can easily check that the proof only requires that* f *be in the class* $\mathcal{S}$ *of non-decreasing sub-additive functions as defined in Proposition 1. In the infinite dimensional case, however, we need to be sure that the series* $y = \sum_k x_k g_k$ *converges unconditionally, and this is perhaps the main reason for restricting to the class* $\mathcal{M}$ *of admissible sparseness measures.*

Proof of Theorem 3. (1) Let $z \in \mathrm{Ker}_f(\mathbf{D})$ satisfy $\sum_{k\in I} f(|z_k|) > \frac{1}{2}\|z\|_f$ and consider for $k \in I$, $x_k := -z_k$, $x'_k := 0$ and for $k \notin I$, $x_k := 0$, $x'_k := z_k$. Since $x' - x = z \in \mathrm{Ker}_f(\mathbf{D})$ we have $\mathbf{D}x' = \mathbf{D}x$, and obviously $I(x) \subset I$ and $\|x'\|_f = \sum_{k\notin I} f(|z_k|) < \sum_{k\in I} f(|z_k|) = \|x\|_f$.

(2) Copy the above line of arguments with $z \in \mathrm{Ker}_f(\mathbf{D})$ such that $\sum_{k\in I} f(|z_k|) = \frac{1}{2}\|z\|_f$.

□

© 2007 Elsevier. This is the author version of an article published in an Elsevier journal. The original publication is available at www.sciencedirect.com with DOI: 10.1016/j.acha.2006.09.003

We can define a function of the index set I which (almost) completely characterizes the uniqueness of f-sparsest expansions. For $f \in \mathcal{M}$, $\mathbf{D}$ a dictionary in X a Banach space and $I \subset K$ a set of indices, we define

$$\Theta_f(I, \mathbf{D}) := \sup_{z \in \mathrm{Ker}_z(\mathbf{D}) \setminus \{0\}} \theta_f(I, z) \tag{26}$$

and by convention we set $\Theta_f(I, \mathbf{D}) = 0$ for all I if $\mathrm{Ker}_z(\mathbf{D}) = \{0\}$. Our previous lemmas have an immediate corollary.

Corollary 1. *Let $\mathbf{D}$ be a dictionary in a separable Banach space X, $f \in \mathcal{M}$ be a sparseness measure and $I \subset K$ be an index set.*

1. *If $\Theta_f(I, \mathbf{D}) < 1/2$ then, for all x, y such that $y = Dx$ and $I(x) \subset I$, x is* the unique *f-sparsest representation of y.*
2. *If $\Theta_f(I, \mathbf{D}) > 1/2$, then there exists x and x' such that $\mathbf{D}x = \mathbf{D}x'$, $I(x) \subset I$ and $\|x'\|_f < \|x\|_f$.*

For each sparseness measure $f \in \mathcal{M}$, the functional $\Theta_f(\cdot, \mathbf{D})$ gives a complete characterization of the uniqueness of the f-sparsest representation of expansions from the sub-dictionary $\mathbf{D}_I = [g_k]_{k \in I}$. However, the evaluation of $\Theta_f(I, \mathbf{D})$ for a given index set I is not trivial in general, and it is not clear when the condition $\Theta_f(I, \mathbf{D}) < 1/2$ is simultaneously satisfied for all $f \in \mathcal{M}$, *i.e.*, when the unique f-sparsest representation is the same for all sparseness measures f. The following example shows that f-sparsest representations do not necessarily coincide and that estimating $\Theta_f(I, \mathbf{D})$ for some sparseness measure $f \in \mathcal{M}$ does not tell much about $\Theta_g(I, \mathbf{D})$ for other measures $g \in \mathcal{M}$.

Example 1. *Let $\mathbf{B} = [g_1, \ldots, g_N]$ be an orthonormal basis in dimension N, $g_{N+1} := \sum_{k=1}^{N} \frac{1}{\sqrt{N}} g_k$ and $\mathbf{D} = [\mathbf{B}, g_{N+1}]$. Clearly, the kernel of $\mathbf{D}$ is the line generated by the vector $z = (1, \ldots, 1, -\sqrt{N})$. Let us consider $I = \{1 \leq k \leq L\}$ an index set where $L \leq N$. We have*

$$\Theta_1(I, \mathbf{D}) = \frac{L}{N + \sqrt{N}} < \frac{L}{N+1} = \Theta_0(I, \mathbf{D})$$

As a result, we have $\Theta_1(I, \mathbf{D}) < 1/2 < \Theta_0(I, \mathbf{D})$ when

$$(N+1)/2 < L < (N + \sqrt{N})/2.$$

On the other hand, let us now consider $J = \{1 \leq k \leq L\} \cup \{N+1\}$. As

$$\begin{aligned} \Theta_1(J, \mathbf{D}) &= \frac{L + \sqrt{N}}{N + \sqrt{N}} \\ \Theta_0(I, \mathbf{D}) &= \frac{L+1}{N+1} \end{aligned}$$

we obtain $\Theta_0(J, \mathbf{D}) < 1/2 < \Theta_1(J, \mathbf{D})$ whenever

$$(N - \sqrt{N})/2 < L < (N-1)/2.$$

Remark 4 (Exact recovery with Basis Pursuit *vs* exact recovery with Matching Pursuit)**.** *When $\mathbf{D}$ is a dictionary in a Hilbert space, Tropp* [45] *proved that if the so-called* Exact Recovery Condition

$$\max_{l \notin I} \|D_I^{\dagger} g_l\|_1 < 1 \tag{27}$$

is satisfied ($\mathbf{D}_I^{\dagger}$ denotes the pseudo-inverse of $\mathbf{D}_I = [g_k]_{k \in I}$), then for any x and y such that $y = \mathbf{D}x$ and $I(x) = I$, both Orthogonal Matching Pursuit and Basis Pursuit "exactly

© 2007 Elsevier. This is the author version of an article published in an Elsevier journal. The original publication is available at www.sciencedirect.com with DOI: 10.1016/j.acha.2006.09.003

recover" the expansion x. For Basis Pursuit, "exact recovery" means that x must be the unique ℓ^1-sparsest representation of y, hence we know by Corollary 1 that, in finite dimension, the Exact Recovery Condition (27) *implies $\Theta_1(I,\mathbf{D}) < 1/2$. However, Example 1 shows that the converse is not true. Indeed, for I defined as above and $l = N+1 \notin I$ we have $\|D_I^+ g_l\|_1 = \sum_{k=1}^{L} \frac{1}{\sqrt{N}} = \frac{L}{\sqrt{N}}$. Hence for $\sqrt{N} \leq L < (N+\sqrt{N})/2$ we have $\Theta_1(I,\mathbf{D}) < 1/2$ and $\max_{l\notin I} \|D_I^+ g_l\|_1 \geq 1$.*

5. Uniqueness of highly sparse expansions

Example 1 shows that, for arbitrary index sets I, not much can be said about the simultaneity of the f-sparsest representation for different sparseness measures. In this section, we will show that the picture completely changes when we look for conditions on the *cardinality* of I so that $\theta_f(I,z) < 1/2$ for any $z \in \mathrm{Ker}_f(\mathbf{D})\backslash\{0\}$. Let us immediately state the main results of this section. The first result gives the theorem announced in the introduction, which is the natural generalization to a series of recent results [19, 20, 21, 14, 15, 17, 27, 13, 26]. To state it we define $m_1(\mathbf{D})$ (respectively $\overline{m_1}(\mathbf{D})$) as the supremum of all integers m such that for every $y \in \Sigma_m$, any representation $y = \mathbf{D}x$ with $\|x\|_0 \leq m$ is *the unique* (respectively *an*) ℓ^1-sparsest representation of y.

Theorem 4. *Let $\mathbf{D}$ be an arbitrary dictionary in a separable Banach space X of finite or infinite dimension, and $f \in \mathcal{M}$ be a sparseness measure.*

1. *Assume $y = \mathbf{D}x$ with $\|x\|_0 \leq m_1(\mathbf{D})$ and $\|x\|_0 < \infty$ (the latter assumption is needed to deal with the case $m_1(\mathbf{D}) = \infty$). Then x is simultaneously* the unique *f-sparsest representation of y for any $f \in \mathcal{M}$.*
2. *Assume $y = \mathbf{D}x$ with $\|x\|_0 \leq \overline{m_1}(\mathbf{D})$ and $\|x\|_0 < \infty$. Then x is simultaneously* an *f-sparsest representation of y for any $f \in \mathcal{M}$.*

Thus, if y has a *highly sparse representation* x –with at most $m_1(\mathbf{D})$ (respectively $\overline{m_1}(\mathbf{D})$) elements from the dictionary– this representation must indeed be the (resp. an) f-sparsest representation for *all* admissible sparseness measures. The interesting consequence is that the highly sparse representation of such vectors y can simply be computed using linear programming, which solves the ℓ^1-optimization problem. Theorem 4 is indeed only a special (but striking) case of a more general result. First we must define $m_f(\mathbf{D})$ (respectively $\overline{m_f}(\mathbf{D})$) as the supremum of all integers m such that for every $y \in \Sigma_m$, any representation $y = \mathbf{D}x$ with $\|x\|_0 \leq m$ is *the unique* (respectively *an*) f-sparsest representation of y.

Proposition 4. *Let $\mathbf{D}$ be an arbitrary dictionary in a separable Banach space X of finite or infinite dimension, and $f \in \mathcal{M}$ be a sparseness measure.*

1. *Assume $y = \mathbf{D}x$ with $\|x\|_0 \leq m_f(\mathbf{D})$ and $\|x\|_0 < \infty$. Then x is simultaneously* the unique *$g \circ f$-sparsest representation of y for any $g \in \mathcal{M}$.*
2. *Assume $y = \mathbf{D}x$ with $\|x\|_0 \leq \overline{m_f}(\mathbf{D})$ and $\|x\|_0 < \infty$. Then x is simultaneous* a *$g \circ f$-sparsest representation of y for any $g \in \mathcal{M}$.*

Proposition 4 is the core of this paper and the rest of this section is devoted to its proof. We will study in more details in the next section which integer(s) m satisfy the assumptions of Proposition 4.

5.1. Some notations.

For any sequence $z = \{z_k\}_{k\in K}$ we let $|z|^\star$ be the decreasing rearrangement of $|z|$, *i.e.*, $|z|_k^\star = |z_{\phi(k)}|$ where ϕ is one to one and $|z|_k^\star \geq |z|_{k+1}^\star$. With a

© 2007 Elsevier. This is the author version of an article published in an Elsevier journal. The original publication is available at www.sciencedirect.com with DOI: 10.1016/j.acha.2006.09.003

slight abuse of notation, we define the "growth function"

$$\theta_f(m,z) := \sup_{\mathrm{card}(I)\leq m} \theta_f(I,z) \tag{28}$$

for any $f \in \mathcal{M}$, $m \geq 1$ and z with $0 < \|z\|_f < \infty$. One can easily check that indeed

$$\theta_f(m,z) = \max_{\mathrm{card}(I)\leq m} \theta_f(I,z) = \frac{\sum_{k=1}^m f(|z|_k^\star)}{\|z\|_f}. \tag{29}$$

5.2. **Proofs.** From the results of the previous section, we have the lemma.

Lemma 4. (1) *An integer m satisfies $m \leq m_f(\mathbf{D})$ if, and only if, for all $z \in Ker_f(\mathbf{D})\backslash\{0\}$*

$$\theta_f(m,z) < 1/2.$$

(2) *An integer m satisfies $m \leq \overline{m_f}(\mathbf{D})$ if, and only if, for all $z \in Ker_f(\mathbf{D})\backslash\{0\}$*

$$\theta_f(m,z) \leq 1/2.$$

We leave the easy proof to the reader. Given this lemma we see that all we need in order to prove Proposition 4 is to prove that if $\theta_f(m,z) < 1/2$ (resp. $\theta_f(m,z) \leq 1/2$) for all $z \in \mathrm{Ker}_f(\mathbf{D})\backslash\{0\}$, then $\theta_{g\circ f}(m,z) < 1/2$ (resp. $\theta_{g\circ f}(m,z) \leq 1/2$) for all $z \in \mathrm{Ker}_{g\circ f}(\mathbf{D})\backslash\{0\}$, and all $g \in \mathcal{M}$. What we will do is prove the following property of growth functions, which is even stronger than needed.

Theorem 5. *Let $x \neq 0$ be any sequence, and $f, g \in \mathcal{M}$. For all integers m we have*

$$\theta_0(m,x) \leq \theta_{g\circ f}(m,x) \leq \theta_f(m,x) \leq \theta_1(m,x). \tag{30}$$

If, for some m, $0 < \theta_{g\circ f}(m,x) = \theta_f(m,x) < 1$ then, for all m, $\theta_{g\circ f}(m,x) = \theta_f(m,x)$.

Proof. First, we notice from the property (29) of growth functions, it is sufficient to prove the desired inequalities for non-increasing sequences. Let us show that it is also sufficient to prove that for any $h \in \mathcal{M}$, $x \neq 0$ non-increasing and $m \geq 0$

$$\theta_h(m,x) \leq \theta_1(m,x). \tag{31}$$

Assuming (31) is true for all h, we can write

$$\theta_{g\circ f}(m,x) = \theta_g(m, f(x)) \leq \theta_1(m, f(x)) = \theta_f(m,x) \tag{32}$$

where we used the fact that since x is a non-increasing sequence and f is a non-decreasing function, $f(x)$ is also a non-increasing sequence. Now from (32) we get

$$\theta_0(m,x) = \theta_{f_0\circ(g\circ f)}(m,x) \leq \theta_{g\circ f}(m,x)$$

and the conclusion is reached.

So let us now prove (31). It will follow from the fact that $(\sum_{k=1}^m h(x_k))/(\sum_{k=1}^m x_k)$ is a non-decreasing sequence. For a given m and all $1 \leq k \leq m$, $x_k \geq x_{m+1}$ implies

© 2007 Elsevier. This is the author version of an article published in an Elsevier journal. The original publication is available at www.sciencedirect.com with DOI: 10.1016/j.acha.2006.09.003

$h(x_k)x_{m+1} \le x_k h(x_{m+1})$, hence we have

$$\begin{aligned}
\sum_{k=1}^{m} h(x_k)\ x_{m+1} &\le \sum_{k=1}^{m} x_k\ h(x_{m+1}) \\
\frac{h(x_{m+1})}{\sum_{k=1}^{m} h(x_k)} &\ge \frac{x_{m+1}}{\sum_{k=1}^{m} x_k} \\
\frac{\sum_{k=1}^{m+1} h(x_k)}{\sum_{k=1}^{m} h(x_k)} &\ge \frac{\sum_{k=1}^{m+1} x_k}{\sum_{k=1}^{m} x_k} \\
\frac{\sum_{k=1}^{m+1} h(x_k)}{\sum_{k=1}^{m+1} x_k} &\ge \frac{\sum_{k=1}^{m} h(x_k)}{\sum_{k=1}^{m} x_k}
\end{aligned}$$

Taking the limit as $m \to \infty$ (in the case where x is finitely supported it is sufficient to take $m+1 = \|x\|_0$) we get that for all m

$$\frac{\|x\|_h}{\|x\|_1} \ge \frac{\sum_{k=1}^{m+1} h(x_k)}{\sum_{k=1}^{m+1} x_k} \ge \frac{\sum_{k=1}^{m} h(x_k)}{\sum_{k=1}^{m} x_k} \tag{33}$$

which obviously implies $\theta_h(m,x) \le \theta_1(m,x)$.

Now, assume $0 < \theta_h(m,x) = \theta_1(m,x) < 1$ for some m. Then the inequalities in Eq. (33) are indeed equalities, so for all $p \ge 1$,

$$\frac{\|x\|_h}{\|x\|_1} = \frac{\sum_{k=1}^{m+p} h(x_k)}{\sum_{k=1}^{m+p} x_k} = \frac{\sum_{k=1}^{m} h(x_k)}{\sum_{k=1}^{m} x_k} \tag{34}$$

and it follows that $\theta_h(m+p,x) = \theta_1(m+p,x)$, $p \ge 1$. Moreover, because the inequalities in (33) are indeed equalities, the equality can be carried over to all inequalities in the proof of (33) to get

$$h(x_k)x_{m+1} = x_k h(x_{m+1}), 1 \le k \le m$$

Because $\theta_1(m,x) < 1$ we must have $x_{m+1} \ne 0$, thus for $l \le m$ we obtain

$$\frac{\sum_{k=1}^{l} h(x_k)}{\sum_{k=1}^{l} x_k} = \frac{h(x_{m+1})}{x_{m+1}}.$$

which, combined with (34) shows that indeed

$$\frac{\sum_{k=1}^{l} h(x_k)}{\sum_{k=1}^{l} x_k} = \frac{\|x\|_h}{\|x\|_1}, \forall l.$$

We have proved that if $0 < \theta_h(m,x) = \theta_1(m,x) < 1$, then $\theta_h(m,x) = \theta_1(m,x)$ for all m. To conclude, we notice that $\theta_{g\circ f}(m,x) = \theta_f(m,x)$ is equivalent to $\theta_g(m,f(x)) = \theta_1(m,f(x))$. □

Remark 5. *In Theorem 5 the assumption that $f \in \mathcal{M}$, and not merely $f \in \mathcal{S}$ (see Proposition 1), is essential. Indeed, the simple fact that we want to get $\theta_f(m,x) \le \theta_1(m,x)$ for all m and x implies that, for $a < b$, we must have $f(b)/(f(a)+f(b)) \le b/(a+b)$,* i.e., *$1 + f(a)/f(b) \ge 1 + a/b$, and this means exactly that $f(t)/t$ is non-increasing.*

© 2007 Elsevier. This is the author version of an article published in an Elsevier journal. The original publication is available at www.sciencedirect.com with DOI: 10.1016/j.acha.2006.09.003

5.3. **Summary.** We have now proved Proposition 4 which can be restated as follows:

$$\begin{aligned} m_f(\mathbf{D}) &\leq m_{g\circ f}(\mathbf{D}),\ \forall g \in \mathcal{M} \\ \overline{m_f}(\mathbf{D}) &\leq \overline{m_{g\circ f}}(\mathbf{D}),\ \forall g \in \mathcal{M}. \end{aligned}$$

From the results in this section, we get even more relations.

Proposition 5. *For any $f, g \in \mathcal{M}$ and any dictionary $\mathbf{D}$*

$$m_0(\mathbf{D}) \geq m_{g\circ f}(\mathbf{D}) \geq m_f(\mathbf{D}) \geq m_1(\mathbf{D}), \tag{35}$$

$$\overline{m_0}(\mathbf{D}) \geq \overline{m_{g\circ f}}(\mathbf{D}) \geq \overline{m_f}(\mathbf{D}) \geq \overline{m_1}(\mathbf{D}), \tag{36}$$

$$m_f(\mathbf{D}) \leq \overline{m_f}(\mathbf{D}) \tag{37}$$

$$\overline{m_f}(\mathbf{D}) \leq m_f(\mathbf{D}) + 1. \tag{38}$$

Proof. The inequalities (35)-(36)-(37) are immediately obtained from Lemma 4 and Theorem 5, for finite or infinite values of the numbers $m_h(\mathbf{D})$ which are involved. The last inequality (38) is trivial if the right hand side is infinite (with the notational convention $\infty \leq \infty + 1$!), let us prove it when $m_f(\mathbf{D})$ is finite. From Lemma 4 we know that for some $z \in \mathrm{Ker}_f(\mathbf{D})\backslash\{0\}$,

$$\theta_f(m_f(\mathbf{D}) + 1, z) \geq 1/2.$$

If $\theta_f(m_f(\mathbf{D}) + 1, z) > 1/2$ we have $\overline{m_f}(\mathbf{D}) = m_f(\mathbf{D})$. To conclude, let us treat the case when $\theta_f(m_f(\mathbf{D}) + 1, z) = 1/2$. Given the expression of $\theta_f(m, z)$ (see Eq. (29)), we must have $|z|^\star_{m_f(\mathbf{D})+2} > 0$ (else $\theta_f(m_f(\mathbf{D}) + 1, z) = 1$), hence

$$\theta_f(m_f(\mathbf{D}) + 2, z) = \theta_f(m_f(\mathbf{D}) + 1, z) + \frac{f(|z|^\star_{m_f(\mathbf{D})+2})}{\|z\|_f} > 1/2,$$

and it follows that $\overline{m_f}(\mathbf{D}) = m_f(\mathbf{D}) + 1$. □

We see that the smallest and the largest of these numbers are respectively $m_1(\mathbf{D})$ and $m_0(\mathbf{D})$. They are of particular interest to characterize uniqueness of sparse expansions, so it seems appropriate to name them : we will call $m_1(\mathbf{D})$ the **strong sparseness number** while $m_0(\mathbf{D})$ is named the **weak sparseness number**. We devote the next section to estimating these two sparseness numbers.

6. Explicit sparseness conditions

In this section we want to estimate the numbers $m_f(\mathbf{D})$ and $\overline{m_f}(\mathbf{D})$ which provide optimal ℓ^0-sparseness conditions to ensure simultaneity of f-sparsest representations over a range of different sparseness measures. In particular, we are interested in estimating the strong and weak sparseness numbers $m_1(\mathbf{D})$ and $m_0(\mathbf{D})$. The goal is to estimate them based on computable characteristics of the dictionary.

For some dictionaries, it may happen that Proposition 4 (resp. Theorem 4) is almost trivial because $m_f(\mathbf{D}) = \overline{m_f}(\mathbf{D}) = 1$ (resp. $m_1(\mathbf{D}) = \overline{m_1}(\mathbf{D}) = 1$). However, we will see that there are many useful dictionaries for which Proposition 4 is not trivial, in the sense that $m_1(\mathbf{D}) \gg 2$. Thus, non-trivial highly sparse expansions can be recovered using Basis Pursuit.

In Section 6.1 we will focus on estimates of $m_f(\mathbf{D})$ in terms of the so-called *spark* and *spread* of the dictionary. The spread of a given dictionary turns out to give a lower bound for $m_f(\mathbf{D})$, but it is not always easy to compute, so we give several computable estimates of the spread in Section 6.2 for dictionaries in a Hilbert space. We will see that the structure of the dictionary determines how good the estimates of Section 6.2 are. In

© 2007 Elsevier. This is the author version of an article published in an Elsevier journal. The original publication is available at www.sciencedirect.com with DOI: 10.1016/j.acha.2006.09.003

Section 6.3 we consider estimates of the spark and spread for special dictionaries built by taking the union of several *mutually incoherent bases*, very much in the spirit of the example of the union of a Gabor and a wavelet frame discussed previously in Section 3.3. Eventually, we discuss in Section 6.4 a few alternative techniques which can be used to estimate the sparseness numbers.

6.1. **Explicit bounds for $m_f(\mathbf{D})$ in terms of the spark and spread of $\mathbf{D}$.** We begin by introducing some notation. We will need the following function:

$$\Theta_f(m,\mathbf{D}) := \sup_{I,\mathrm{card}(I)\leq m} \Theta_f(I,\mathbf{D}) = \sup_{z\in \mathrm{Ker}_f(\mathbf{D})\backslash\{0\}} \theta_f(m,z) \tag{39}$$

with the convention that $\Theta_f(m,\mathbf{D}) = 0$ for all m if $\mathrm{Ker}_f(\mathbf{D})\backslash\{0\} = \emptyset$.

Lemma 5. *For any dictionary $\mathbf{D}$ and sparseness measure $f \in \mathcal{M}$ we have*

$$\max\{m, \Theta_f(m,\mathbf{D}) < 1/2\} \leq m_f(\mathbf{D}) \tag{40}$$

Proof. First, assume that $\Theta_f(m,\mathbf{D}) < 1/2$. By definition, for any $z \in \mathrm{Ker}_f(\mathbf{D})\backslash\{0\}$, we have $\theta_f(m,z) \leq \Theta_f(m,\mathbf{D}) < 1/2$. Applying Lemma 4, it follows that $m \leq m_f(\mathbf{D})$, and we get the left hand side inequality. □

We now introduce the spark and the spread of a dictionary. These two quantities will play an important role in estimating $m_f(\mathbf{D})$. The **spark** $Z_0(\mathbf{D})$ of the dictionary $\mathbf{D}$ is defined by $Z_0(\mathbf{D}) := 1/\Theta_0(1,\mathbf{D})$ and the **spread** $Z_1(\mathbf{D})$ is defined by $Z_1(\mathbf{D}) := 1/\Theta_1(1,\mathbf{D})$. The spark was introduced by Donoho and Elad in [13] and the spread was introduced by the authors in [26]. When these quantities are finite one can easily verify that

$$Z_0(\mathbf{D}) = \inf_{x\in \mathrm{Ker}_0(\mathbf{D})\backslash\{0\}} \|x\|_0 \tag{41}$$

and

$$Z_1(\mathbf{D}) = \inf_{x\in \mathrm{Ker}_1(\mathbf{D})\backslash\{0\}, \|x\|_\infty=1} \|x\|_1. \tag{42}$$

Denoting $\lfloor t \rfloor$ the largest integer such that $\lfloor t \rfloor < t \leq \lfloor t \rfloor + 1$, we have the following lower estimate for $m_f(\mathbf{D})$.

Lemma 6. *For any sparsity measure $f \in \mathcal{M}$ and any dictionary $\mathbf{D}$, if $Z_1(\mathbf{D}) < \infty$ then*

$$m_f(\mathbf{D}) \geq \left\lfloor \frac{1}{2\cdot\Theta_f(1,\mathbf{D})} \right\rfloor \geq \lfloor Z_1(\mathbf{D})/2 \rfloor. \tag{43}$$

If $Z_0(\mathbf{D}) < \infty$ then

$$m_0(\mathbf{D}) = \big\lfloor Z_0(\mathbf{D})/2 \big\rfloor. \tag{44}$$

In infinite dimension, (43) *still holds true when $Z_1(\mathbf{D}) = \infty$, and* (44) *is valid when $Z_0(\mathbf{D}) = \infty$. In finite dimension, $\max(Z_1(\mathbf{D}), Z_0(\mathbf{D})) < \infty$ except if $\mathbf{D}$ is a basis of linearly independent vectors, in which case all $m_f(\mathbf{D})$ equal the dimension.*

Proof. First, let us deal with the case where $Z_1(\mathbf{D}) < \infty$ and $Z_0(\mathbf{D}) < \infty$. It is not difficult to check that Θ_f is sub-additive, *i.e.*, $\Theta_f(k+l,\mathbf{D}) \leq \Theta_f(k,\mathbf{D}) + \Theta_f(l,\mathbf{D})$, $k,l \geq 1$, so in particular

$$\Theta_f(m,\mathbf{D}) \leq m\cdot\Theta_f(1,\mathbf{D}), \qquad m \geq 1 \tag{45}$$

© 2007 Elsevier. This is the author version of an article published in an Elsevier journal. The original publication is available at www.sciencedirect.com with DOI: 10.1016/j.acha.2006.09.003

The left hand side inequality in (43) is obvious from (40) and (45), and the right hand side inequality follows immediately from Theorem 5. The equality (44) was stated in [27, 13], but let us give the proof for the sake of completeness. Remember that the spark is $Z_0(\mathbf{D}) = \inf\left\{\|x\|_0,\ x \in \mathrm{Ker}(\mathbf{D}), x \neq 0\right\}$. As the infimum of a set of integer numbers, Z_0 is itself an integer and is indeed a minimum, *i.e.*, there exists $x \in \mathrm{Ker}(\mathbf{D})$ such that $\|x\|_0 = Z_0$. Letting $I = I(x)$ the support of this sequence x we can split I into two disjoint sets I_1 and I_2 of same cardinality $Z_0/2 = \lfloor Z_0/2 \rfloor + 1$ (if Z_0 is even) or with $\mathrm{card}(I_1) = (Z_0 - 1)/2$ and $\mathrm{card}(I_2) = (Z_0 + 1)/2 = \lfloor Z_0/2 \rfloor + 1$ (if Z_0 is odd). Obviously $\theta_0(I_2, \mathbf{D}) = \frac{\mathrm{card}(I_2)}{\|x\|_0} \geq 1/2$ (see Eq. (22)), hence by the very definition of $m_0(\mathbf{D})$, we have $m_0(\mathbf{D}) < \mathrm{card}(I_2) = \lfloor Z_0/2 \rfloor + 1$.

The spark is infinite if, and only if, $\mathrm{Ker}_0(\mathbf{D}) = \{0\}$, which, in infinite dimension, implies that $m_0(\mathbf{D})$ is also infinite. The spread is infinite if, and only if, $\mathrm{Ker}_1(\mathbf{D}) = \{0\}$, in which case $\mathrm{Ker}_f(\mathbf{D}) \subset \mathrm{Ker}_1(\mathbf{D}) = \{0\}$ for all sparseness measures f (by Lemma 1) which, in infinite dimension, implies that $m_f(\mathbf{D}) = \infty$ for all f. □

Lemma 6 gives an exact estimate of $m_0(\mathbf{D})$ in terms of the spark, but the problem with the spark is that its numerical computation is generally combinatorial [13]. For some special dictionaries however, we will see in Section 6.3 that the spark can be estimated analytically. At the other end of the scale of numbers $\{m_f(\mathbf{D}),\ f \in \mathcal{M}\}$ is $m_1(\mathbf{D})$. Lemma 6 does not give an exact estimate of $m_1(\mathbf{D})$ in terms of $Z_1(\mathbf{D})$. There is a good reason for this as the following example illustrates.

Example 2. *With the dictionary considered in Example 1, it is easy to check that*

$$\begin{aligned} Z_1(\mathbf{D}) &= \sqrt{N} + 1 \\ Z_0(\mathbf{D}) &= N + 1 \\ \Theta_1(m, \mathbf{D}) &= \frac{m - 1 + \sqrt{N}}{N + \sqrt{N}} \\ m_1(\mathbf{D}) &= 1 + \left\lfloor \frac{N - \sqrt{N}}{2} \right\rfloor. \end{aligned}$$

hence for large N, $m_1(\mathbf{D}) \approx N/2 \approx \sqrt{N}\lfloor Z_1(\mathbf{D})/2 \rfloor$ is much larger than its lower estimate (43).

6.2. **Lower estimates for the spread $Z_1(\mathbf{D})$.** From Lemma 6 and the inequality (35) we know that the spread $Z_1(\mathbf{D})$ can be used to get an "easy" –though sometimes too pessimistic, see Example 2– lower estimate of $m_f(\mathbf{D})$ for any sparseness measure $f \in \mathcal{M}$. The following lemma gives a general estimate for the spread when we have a dictionary in a *Hilbert space*.

Lemma 7. *For a general dictionary $\mathbf{D} = \{g_k\}$ in a Hilbert space, the* **coherence** *is defined* [14] *as*

$$M(\mathbf{D}) := \sup_{k \neq k'} |\langle g_k, g_{k'} \rangle|. \tag{46}$$

We have the lower estimate

$$Z_1(\mathbf{D}) \geq 1 + \frac{1}{M(\mathbf{D})}. \tag{47}$$

Proof. Consider $x \in \mathrm{Ker}_1(\mathbf{D})$. 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, $|x_k| \leq M(\mathbf{D}) \cdot \sum_{k' \neq k} |x_{k'}|$. It follows that

© 2007 Elsevier. This is the author version of an article published in an Elsevier journal. The original publication is available at www.sciencedirect.com with DOI: 10.1016/j.acha.2006.09.003

$(1+M)\cdot|x_k| \leq M\cdot\|x\|_1$. Taking the supremum over k we get $(1+M)\|x\|_\infty \leq M\cdot\|x\|_1$ and the result follows. □

Corollary 2. *Assume* $y=\sum_k c_k g_k$ *where*

$$\|c\|_0 \leq \lfloor (1+1/M(\mathbf{D}))/2 \rfloor.$$

Then c *is the unique and simultaneous* f*-sparsest representation of* y *for any* $f\in\mathcal{M}$*. In particular, it can be computed by linear programming, which solves the* ℓ^1*-problem.*

Corollary 2 was in germ in Donoho and Huo's early paper [14] on exact recovery of sparse expansion through Basis Pursuit, where it was only used for $\mathbf{D}$ a union of two orthonormal bases and $f(t)=t^\tau$, $\tau\in\{0,1\}$. In [27] and[13] it was extended to arbitrary dictionaries, and in [26] to $f(t)=t^\tau$, $\tau\in[0,1]$.

In practice, if one builds a dictionary $\mathbf{D}$ with the aim of using Basis Pursuit to recover highly sparse expansions, it is desirable to guarantee that $m_1(\mathbf{D})$ has a large value. If no other tool is available to estimate $m_1(\mathbf{D})$, the above Lemma shows that the dictionary should be designed so as to have as small a coherence/as large a spread as possible. For redundant dictionaries which contain an orthonormal basis in finite dimension, Lemma 8 below shows that the coherence cannot be arbitrarily small/the spread cannot be arbitrarily large.

Lemma 8. *In a finite dimensional Hilbert space of dimension* N*, assume* $\mathbf{D}$ *contains an orthonormal basis and at least one additional vector. Then*

$$1+\frac{1}{M(\mathbf{D})} \leq Z_1(\mathbf{D}) \leq 1+\sqrt{N}. \tag{48}$$

It follows that

1. $M(\mathbf{D})\geq 1/\sqrt{N}$*;*
2. *if* $M(\mathbf{D})=1/\sqrt{N}$*, then the spread is exactly* $Z_1(\mathbf{D})=1+\sqrt{N}$*.*

Proof. The lower estimate in (48) simply comes from the general one (47). To get the upper estimate, without loss of generality, we can assume that the orthonormal basis corresponds to the first N vectors of $\mathbf{D}$. Take

$$x_k := \begin{cases} \langle g_{N+1}, g_k\rangle, & 1\leq k\leq N \\ -1, & k=N+1 \\ 0, & k\notin\{1..N\}\cup\{N+1\} \end{cases}.$$

Obviously $x\in\mathrm{Ker}(\mathbf{D})$, $\|x\|_\infty=1$ and

$$\begin{aligned} \|x\|_1 &= 1+\sum_{k=1}^N |\langle g_{N+1}, g_k\rangle| \\ &\leq 1+\sqrt{N}\left(\sum_{k=1}^N |\langle g_{N+1}, g_k\rangle|^2\right)^{1/2} = 1+\sqrt{N}, \end{aligned}$$

hence, using the characterization (42) of the spread, we obtain the result. □

6.3. **Estimates for the spark $Z_0(\mathbf{D})$ and for $m_1(\mathbf{D})$ in unions of orthonormal bases.** We have already seen with Example 2 that the spread $Z_1(\mathbf{D})$ can be a pessimistic lower estimate for $m_1(\mathbf{D})$. To the contrary, we know from Lemma 6 that the spark $Z_0(\mathbf{D})$ provides an exact estimate of $m_0(\mathbf{D})$, thus an upper estimate for all numbers $m_f(\mathbf{D})$ including $m_1(\mathbf{D})$. For arbitrary general dictionaries, the computation of the spark is

© 2007 Elsevier. This is the author version of an article published in an Elsevier journal. The original publication is available at www.sciencedirect.com with DOI: 10.1016/j.acha.2006.09.003

combinatorial [13]. However, when $\mathbf{D} = [\mathbf{B}_1, \ldots, \mathbf{B}_L]$ is a finite union of orthonormal bases, it is possible to estimate the spark analytically. As a result we get a uniform estimate of the order of magnitude of $m_f(\mathbf{D})$ for all $f \in \mathcal{M}$.

It is perhaps not obvious that one can have a large number of orthonormal bases with small coherence $M(\mathbf{D})$, but this is possible (for certain values of the dimension N), and we will use the following theorem to build examples of such dictionaries. The dictionaries from Theorem 6 are called **Grassmannian dictionaries** due to the fact that their construction is closely related to the Grassmannian packing problem. See [4, 43] for details and for a proof of Theorem 6.

Theorem 6. *Consider $\mathcal{H} = \mathbb{R}^N$ with $N = 2^{j+1}$ and $j \geq 0$. 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 of distinct vectors belonging to $\mathbf{D}$: $|\langle u, v\rangle| \in \{0, N^{-1/2}\}$.*

For $N = 2^j$, $j \geq 0$ and $\mathcal{H} = \mathbb{C}^N$, one can find a dictionary $\mathbf{D}$ 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 $\mathbf{D}$: $|\langle u, v\rangle| \in \{0, N^{-1/2}\}$.

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 $\mathbb{C}^N$ –hence $\mathbf{D}$ contains the large number $N(N + 1)$ 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.

In [27] the authors showed that for $\mathbf{D}$ a union of L orthonormal bases with coherence $M(\mathbf{D})$, the spark satisfies

$$Z_0(\mathbf{D}) \geq \left(1 + \frac{1}{L-1}\right) \frac{1}{M(\mathbf{D})}. \tag{49}$$

In the case of $L = 2$ bases, the estimate is sharp in the sense that there are examples [14, 15] of pairs of bases with $Z_0 = 2/M = 2\sqrt{N}$. The typical examples are the Dirac/Fourier pair and the Haar/Walsh pair in dimension $N = 2^{2j}$. For unions of three or more bases, it is not known in general when the estimate is sharp. However, if $L > 1 + 1/M$, it is certainly not sharp since it is weaker than the general estimate $Z_0 \geq Z_1 \geq 1 + 1/M$.

Given an arbitrary orthonormal basis $\mathbf{B}_1$ in $\mathcal{H} = \mathbb{C}^{2^j}$, it is not difficult to check from Theorem 6 that it is possible to complete $\mathbf{B}_1$ with 2^j other bases $\mathbf{B}_l$ so that the resulting dictionary has minimum coherence. However, it does not seem clear whether or not such a completion is still possible when the first two (or, more generally, the first L) mutually incoherent bases are fixed. In the case of the Dirac and Fourier bases, the Chirp basis can be added [26] to get three incoherent bases, but it is not known whether the construction can go further. If we let $3 \leq L \leq 1 + \sqrt{N}$ be an integer for which the answer is yes, then the corresponding union of orthonormal bases $\mathbf{B}_l$ –which extends the Dirac/Fourier pair $[\mathbf{B}_1, \mathbf{B}_2]$– satisfies

$$\begin{aligned} \left(1 + \frac{1}{L-1}\right)\sqrt{N} &\leq Z_0([\mathbf{B}_1, \ldots, \mathbf{B}_L]) \\ &\leq Z_0([\mathbf{B}_1, \mathbf{B}_2]) = 2\sqrt{N}. \end{aligned}$$

Thus, for any L for which the construction is possible, the lower estimate (49) on the spark is sharp in the sense of order of magnitude.

When $\mathbf{D}$ is a union of L orthonormal bases, we can also get improved lower estimates for $m_1(\mathbf{D})$ which will show that the spark, despite not being sharp, also gives the right

© 2007 Elsevier. This is the author version of an article published in an Elsevier journal. The original publication is available at www.sciencedirect.com with DOI: 10.1016/j.acha.2006.09.003

order of magnitude. It was proved in [27] that for any expansion $y = \mathbf{D}x$ with

$$\text{card}(I(x)) < \left(\sqrt{2} - 1 + \frac{1}{2(L-1)}\right) \frac{1}{M(\mathbf{D})}, \tag{50}$$

x was necessarily the unique ℓ^1-sparsest representation of y. In light of the general theory developed in the previous sections (in particular Theorem 4) we see that this can be restated as

$$m_1(\mathbf{D}) \geq \left\lfloor \left(\sqrt{2} - 1 + \frac{1}{2(L-1)}\right) \frac{1}{M(\mathbf{D})} \right\rfloor \tag{51}$$

and that indeed the sparseness condition (50) is sufficient to ensure that the representation x is f-sparsest for any sparseness measure $f \in \mathcal{M}$. The lower estimate (51) can improve the general one based on the spread and the coherence (see Eqs. (43) and (47)) only if $2 \leq L \leq 6$ and M is small enough (see [27] for more explanations on the upper limit $L \leq 6$).

In [17] it has been shown that for $L = 2$ and some special pairs of incoherent bases (with $M([\mathbf{B}_1, \mathbf{B}_2)] = 1/\sqrt{N}$) in dimension $N = 2^{2j+1}$, the sufficient condition (50) cannot be improved, the lower estimate (51) is an equality and indeed $m_1 = \left\lfloor (\sqrt{2} - 1/2)\frac{1}{M} \right\rfloor = \left\lfloor (\sqrt{2} - 1/2)\sqrt{N} \right\rfloor$. The construction of [17] can easily be extended to the Dirac/Fourier pair and the Haar/Walsh one. Just as for the spread estimate above, it is not known when/if the lower estimate of $m_1(\mathbf{D})$ is sharp, but if the pair $[\mathbf{B}_1, \mathbf{B}_2]$ constructed in [17] (or the Dirac-Fourier pair or the Haar/Walsh one, ...) can be extended to a union of $3 \leq L \leq 6$ incoherent bases, then the corresponding union of orthonormal bases $\mathbf{B}_l$ satisfies the upper estimate

$$m_1([\mathbf{B}_1, \ldots, \mathbf{B}_L]) \leq m_1([\mathbf{B}_1, \mathbf{B}_2]) = \lfloor (\sqrt{2} - 1/2)\sqrt{N} \rfloor.$$

Thus, for any L for which the construction is possible, the specific lower estimate (51) as well as the general one $m_1(\mathbf{D}) \geq \lfloor (1 + \sqrt{N})/2 \rfloor$ are sharp in the sense of order of magnitude.

6.4. **Alternative estimates using the Babel function.** We conclude this paper by a brief discussion of some alternate estimates of the strong and weak sparseness numbers using the so-called Babel function. In some cases, the Babel function indeed gives stronger estimates of $m_1(\mathbf{D})$ than the ones considered so far. The **Babel function** of a dictionary $\mathbf{D}$ in a Hilbert space –which provides a natural generalization to the notion of coherence– was formally introduced by Tropp [45]

$$\mu(m, \mathbf{D}) := \max_k \max_{I, \text{card}(I)=m, k \notin I} \sum_{l \in I} |\langle g_l, g_k \rangle| . \tag{52}$$

The Babel function is related to the growth function $\Theta_1(m, \mathbf{D})$ that we have defined above (see Eq. (39)). Indeed, though they did not explicitly develop either the notion of Babel function or that of a growth function, Donoho and Elad implicitly used the two notions and proved the following inequality (see [13, Theorem 8]).

Lemma 9 (Donoho,Elad)**.** *For any dictionary* $\mathbf{D}$ *in a Hilbert space and any* $m \geq 1$*,*

$$\Theta_1(m, \mathbf{D}) \leq \mu(m, \mathbf{D}). \tag{53}$$

It follows that

$$m_1(\mathbf{D}) \geq \max\{m, \mu(m, \mathbf{D}) < 1/2\}. \tag{54}$$

© 2007 Elsevier. This is the author version of an article published in an Elsevier journal. The original publication is available at www.sciencedirect.com with DOI: 10.1016/j.acha.2006.09.003

Let us give Donoho and Elad's proof for the sake of completeness.

Proof. We take $z \in \mathrm{Ker}_1(\mathbf{D})\backslash\{0\}$ and $I \subset K$ an index set of cardinality at most m. From $\mathbf{D}z = 0$ we derive $-z = (\mathbf{\Gamma} - \mathbf{Id})z$ where $\mathbf{\Gamma} = (\langle g_l, g_k\rangle) = \mathbf{D}^\star\mathbf{D}$ is the Gram matrix of $\mathbf{D}$. Denoting $\mathbf{H} = (H_{l,k})$ the $\mathrm{card}(I) \times \mathrm{card}(K)$ matrix formed from the rows of $\mathbf{\Gamma} - \mathbf{Id}$ corresponding to indices in the index set I, we have $\sum_{l\in I} |z_l| = \|\mathbf{H}z\|_1$. Viewing $\mathbf{H}$ as a matrix mapping from $\ell^1(K)$ into $\ell^1(I)$, we know that its operator norm is

$$\begin{aligned} |||\mathbf{H}|||_{1,1} &= \sup_u \frac{\|\mathbf{H}u\|_1}{\|u\|_1} = \max_k \sum_{l\in I} |H_{l,k}| \\ &= \max_k \sum_{l\in I} |\langle g_l, g_k\rangle - \delta_{l,k}| \\ &= \max_k \sum_{l\in I, l\neq k} |\langle g_l, g_k\rangle| \leq \mu(m, \mathbf{D}) \end{aligned}$$

where $\delta_{l,k}$ is the Kronecker symbol. The result follows readily. □

Using von Neumann series, Tropp [45] obtained a slightly stronger result.

Lemma 10 (Tropp)**.** *For any dictionary* $\mathbf{D}$ *in a Hilbert space and any* $m \geq 1$*,*

$$m_1(\mathbf{D}) \geq \max\{m, \mu(m-1, \mathbf{D}) + \mu(m, \mathbf{D}) < 1\}. \tag{55}$$

We refer the reader to Tropp's enjoyable paper for the proof.

It is easy to verify that for any dictionary, $\mu(m, \mathbf{D}) \leq m \cdot \mu(1, \mathbf{D})$, where $\mu(1, \mathbf{D})$ is nothing but the coherence $M(\mathbf{D})$ of the dictionary. Using this (sometimes crude) estimate in (55), we can recover the estimate obtained previously through the use of the spread (namely, by combining Eqs. (43) and (47)).

Corollary 3. *We have,*

$$m_1(\mathbf{D}) \geq \left\lfloor \frac{1}{2}\left(1 + \frac{1}{M(\mathbf{D})}\right)\right\rfloor. \tag{56}$$

for an arbitrary dictionary $\mathbf{D}$ *in a Hilbert space.*

We notice that Tropp's Babel function estimate (55) may provide a stronger estimate of $m_1(\mathbf{D})$ in the cases where the estimate $\mu(m, \mathbf{D}) \leq m \cdot M(\mathbf{D})$ is not tight. Again, we refer to the paper of Tropp for examples.

7. Conclusion

We have studied sparse representation of signals using an arbitrary dictionary in a Banach space. When the dictionary is a *localized frame* in a *Hilbert space*, we showed that the canonical frame representation provides a near sparsest representation for many ℓ^τ sparseness measures. For more general dictionaries in Banach spaces, we considered sparseness as measured by a very general sparseness measure $\|\cdot\|_f$. Given a dictionary and a signal y, we provided sufficient conditions for the minimization problem

$$\text{minimize } \|x\|_f \text{ subject to } s = \sum_k x_k g_k, \tag{57}$$

to have the same unique solution as the problem

$$\text{minimize } \|x\|_1 \text{ subject to } s = \sum_k x_k g_k, \tag{58}$$

© 2007 Elsevier. This is the author version of an article published in an Elsevier journal. The original publication is available at www.sciencedirect.com with DOI: 10.1016/j.acha.2006.09.003

and the conditions are *independent* of the particular sparseness measure f.

The latter minimization problem (58) can be solved using a linear programming technique, *i.e.*, by a polynomial time algorithm. For a dictionary in a Hilbert space we prove that the condition $\|x\|_0 \leq 1/2(1 + 1/M)$, where M is the coherence of the dictionary, is sufficient for (57) to have the same solution as (58) for any sparseness measure f.

The results generalize previous results by Donoho and Elad [13] and by the authors [27], where only two types of sparseness measures were considered: the ℓ^0-norm and the ℓ^1-norm.

Acknowledgments

This work was done while the first author was a guest of the Department of Mathematical Sciences at Aalborg University. The first author would like to thank the department and the WAVES project from the Danish Technical Research Council for making his visit possible.

Dedication

The first author dedicates this paper to his daughter, Ariane.

References

[1] C. Bennett and R. Sharpley. *Interpolation of operators*. Academic Press Inc., Boston, MA, 1988.
[2] J. Bergh and J. Löfström. *Interpolation Spaces, an introduction*. Number 223 in Comprehensive Studies in Mathematics. Springer-Verlag, 1976.
[3] D. Bertsekas. *Non-Linear Programming*. Athena Scientific, Belmont, MA, 2nd edition, 1995.
[4] 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.
[5] C. K. Chui, W. He, and J. Stöckler. Compactly supported tight and sibling frames with maximum vanishing moments. *Appl. Comput. Harmon. Anal.*, 13(3):224–262, 2002.
[6] I. Daubechies. *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.
[7] I. Daubechies, B. Han, A. Ron, and Z. Shen. Framelets: MRA-based constructions of wavelet frames. *Appl. Comput. Harmon. Anal.*, 14(1):1–46, 2003.
[8] L. Daudet. *Représentations structurelles de signaux audiophoniques : méthodes hybrides pour des applications à la compression*. PhD thesis, Université de Provence (Aix-Marseille I), 2000.
[9] L. Daudet and B. Torrésani. Hybrid representations for audiophonic signal encoding. *Signal Processing, special issue on Image and Video Coding Beyond Standards*, 82(11):1595–1617, 2002.
[10] R. A. DeVore, B. Jawerth, and V. Popov. Compression of wavelet decompositions. *Am. J. Math.*, 114(4):737–785, 1992.
[11] R. A. DeVore and G. G. Lorentz. *Constructive approximation*. Springer-Verlag, Berlin, 1993.
[12] D. Donoho and E. J. Candès. Curvelets: A surprisingly effective nonadaptive representation of objects with edges. Technical report, Stanford University, 1999.
[13] D. Donoho and M. Elad. Maximal sparsity representation via ℓ^1 minimization. *Proc. Nat. Aca. Sci.*, 100(5):2197–2202, Mar. 2003.
[14] D. Donoho and X. Huo. Uncertainty principles and ideal atomic decompositions. *IEEE Trans. Inform. Theory*, 47(7):2845–2862, Nov. 2001.
[15] M. Elad and A. Bruckstein. A generalized uncertainty principle and sparse representations in pairs of bases. *IEEE Trans. Inform. Theory*, 48(9):2558–2567, Sept. 2002.
[16] H. Feichtinger and G. Zimmermann. An exotic minimal banach space of functions. *Math. Nachr.*, 239-240:42–61, 2002.
[17] A. Feuer and A. Nemirovsky. On sparse representations in pairs of bases. *IEEE Trans. Inform. Theory*, 49(6):1579–1581, June 2003.
[18] D. Field and B. Olshausen. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609, 1996.

© 2007 Elsevier. This is the author version of an article published in an Elsevier journal. The original publication is available at www.sciencedirect.com with DOI: 10.1016/j.acha.2006.09.003

[19] J.-J. Fuchs. Une approche à l'estimation et l'identification simultanées. In *Actes du seizième colloque GRETSI*, volume 2, pages 1273–1276, Grenoble, Sept. 1997.

[20] J.-J. Fuchs. Detection and estimation of superimposed signals. In *Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP'98)*, volume III, pages 1649–1652, Seattle, 1998.

[21] J.-J. Fuchs. On sparse representations in arbitrary redundant bases. *IEEE Trans. Inform. Theory*, 50(6):1341–1344, June 2004.

[22] A. Gilbert, S. Muthukrishnan, and M. Strauss. Approximation of functions over redundant dictionaries using coherence. In *The 14th ACM-SIAM Symposium on Discrete Algorithms (SODA'03)*, pages 243–252, Baltimore, Maryland, USA, Jan. 2003.

[23] A. Gilbert, S. Muthukrishnan, M. Strauss, and J. Tropp. Improved sparse approximation over quasi-incoherent dictionaries. In *Int. Conf. on Image Proc. (ICIP'03)*, pages 37–40, Barcelona, Catalonia, Spain, Sept. 2003.

[24] R. Gribonval. Sparse decomposition of stereo signals with matching pursuit and application to blind separation of more than two sources from a stereo mixture. In *Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP'02)*, Orlando, Florida, May 2002.

[25] R. Gribonval, R. M. Figueras i Ventura, and P. Vandergheynst. A simple test to check the optimality of sparse signal approximations. *EURASIP Signal Processing, special issue on Sparse Approximations in Signal and Image Processing*, 86(3):496–510, Mar. 2006.

[26] R. Gribonval and M. Nielsen. Approximation with highly redundant dictionaries. In M. Unser, A. Aldroubi, and A. F. Laine, editors, *Proc. SPIE '03*, volume 5207 Wavelets: Applications in Signal and Image Processing X, pages pp. 216–227, San Diego, CA, Aug. 2003.

[27] R. Gribonval and M. Nielsen. Sparse decompositions in unions of bases. *IEEE Trans. Inform. Theory*, 49(12):3320–3325, Dec. 2003.

[28] R. Gribonval and M. Nielsen. Nonlinear approximation with dictionaries. I. Direct estimates. *J. Fourier Anal. and Appl.*, 10(1):51–71, 2004.

[29] R. Gribonval and P. Vandergheynst. On the exponential convergence of Matching Pursuits in quasi-incoherent dictionaries. *IEEE Trans. Information Theory*, 52(1):255–261, Jan. 2006.

[30] K. Gröchenig. *Foundations of Time-Frequency Analysis*. Applied and Numerical Harmonic Analysis. Birkhauser, Boston, MA, Dec. 2001.

[31] K. Gröchenig. Localized frames are finite unions of riesz sequences. *Adv. Comp. Math.*, 18(2–4):149–157, 2003.

[32] K. Gröchenig. Localization of frames, Banach frames, and the invertibility of the frame operator. *J. Fourier Anal. Appl.*, 10(2), 2004. to appear.

[33] K. Gröchenig and S. Samarah. Nonlinear approximation with local Fourier bases. *Constr. Approx.*, 16(3):317–332, 2000.

[34] K. Kreutz-Delgado, B. Rao, K. Engan, T.-W. Lee, and T. Sejnowski. Convex/schur-convex (csc) log-priors and sparse coding. In *6th Joint Symposium on Neural Computation*, pages 65–71, Institute for Neural Computation, May 1999.

[35] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, San Diego, CA, 1998.

[36] S. Mallat and Z. Zhang. Matching pursuit with time-frequency dictionaries. *IEEE Trans. Signal Process.*, 41(12):3397–3415, Dec. 1993.

[37] K. A. Okoudjou. Embeddings of some classical Banach spaces into modulation spaces. *Proc. Amer. Math. Soc.*, 132(6):1639–1647, 2004.

[38] B. D. Rao. Signal processing with the sparseness constraint. In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP'98)*, pages 1861–1864, Seattle, may 1998.

[39] A. Ron and Z. Shen. Affine systems in $L_2(\mathbb{R}^d)$. II. Dual systems. *J. Fourier Anal. Appl.*, 3(5):617–637, 1997. Dedicated to the memory of Richard J. Duffin.

[40] A. Ron and Z. Shen. Affine systems in $L_2(\mathbb{R}^d)$: the analysis of the analysis operator. *J. Funct. Anal.*, 148(2):408–447, 1997.

[41] A. Shrijver. *Theory of Linear and Integer Programming*. John Wiley, 1998.

[42] J.-L. Starck, M. Elad, and D. Donoho. Image decomposition : Separation of textures from piecewise smooth content. In M. Unser, A. Aldroubi, and A. Laine, editors, *Wavelet: Applications in Signal and Image Processing X, Proc, SPIE '03*, volume 5207, San Diego, CA, aug 2003.

[43] T. Strohmer and R. Heath. Grassmannian frames with applications to coding and communications. *Appl. Comp. Harm. Anal.*, 14(3):257–275, 2003.

© 2007 Elsevier. This is the author version of an article published in an Elsevier journal. The original publication is available at www.sciencedirect.com with DOI: 10.1016/j.acha.2006.09.003

[44] H. Triebel. *Theory of function spaces*, volume 78 of *Monographs in Mathematics*. Birkhäuser Verlag, Basel, 1983.
[45] J. Tropp. Greed is good : Algorithmic results for sparse approximation. *IEEE Trans. Inform. Theory*, 50(10):2231–2242, Oct. 2004.
[46] M. Zibulevsky and B. Pearlmutter. Blind source separation by sparse decomposition in a signal dictionary. *Neural Computation*, 13(4):863–882, 2001.

IRISA-INRIA, Campus de Beaulieu, F-35042 Rennes CEDEX, France
E-mail address: Remi.Gribonval@inria.fr

Department of Mathematical Sciences, Aalborg University, Fredrik Bajers Vej 7G, DK-9220 Aalborg East, Denmark
E-mail address: mnielsen@math.auc.dk