



HAL
open science

On the strong uniqueness of highly sparse expansions from redundant dictionaries

Rémi Gribonval, Morten Nielsen

► **To cite this version:**

Rémi Gribonval, Morten Nielsen. On the strong uniqueness of highly sparse expansions from redundant dictionaries. Independent Component Analysis and Blind Signal Separation, Fifth International Conference, ICA 2004, Sep 2004, Granada, Spain. pp.201-208, 10.1007/978-3-540-30110-3_26 . inria-00567341

HAL Id: inria-00567341

<https://inria.hal.science/inria-00567341v1>

Submitted on 20 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.

On the strong uniqueness of highly sparse representations from redundant dictionaries

R. Gribonval¹ and M. Nielsen² *

¹ IRISA-INRIA, Campus de Beaulieu, F-35042 Rennes CEDEX, France,
remi.gribonval@inria.fr,

² Department of Mathematical Sciences, Aalborg University, Fredrik Bajers Vej 7G,
DK-9220 Aalborg East, Denmark,
mnielsen@math.aau.dk

Abstract. A series of recent results shows that if a signal admits a sufficiently sparse representation (in terms of the number of nonzero coefficients) in an “incoherent” dictionary, this solution is unique and can be recovered as the unique solution of a linear programming problem. We generalize these results to a large class of sparsity measures which includes the ℓ^p -sparsity measures for $0 \leq p \leq 1$. 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 w.r.t. arbitrary admissible sparsity measures. Our results should have a practical impact on source separation methods based on sparse decompositions, since they indicate that a large class of sparse priors can be efficiently replaced with a Laplacian prior without changing the resulting solution.

1 Introduction

Sparse decompositions of signals in redundant dictionaries provide quite a successful practical tool for blind source separation (BSS), including the degenerate case where there are more sources than sensors [18]. In this paper, we prove that estimators based on sparse decompositions are relatively robust to the choice of the sparse prior within a fairly large class. Our results directly apply to some noise-free single sensor BSS problems [1], but further work is needed to extend them to the case of multiple sensors and noisy measurements.

Given a redundant signal (or image) dictionary, every signal y has infinitely many possible representations, and it is common to choose one according to some *sparsity measure*. When the dictionary is indeed a basis, each signal has a unique representation and it does not matter which sparsity 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 sparsity measure.

* This work was 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.

A **dictionary** in a real or complex Hilbert space \mathbb{R}^N or \mathbb{C}^N is a family of $K \geq N$ unit vectors $\{g_k\}_{k \in K}$ which spans the entire space. One can think of g_k as the k -th column of a $N \times K$ matrix $\mathbf{D} = [g_1, \dots, g_K]$, and any vector $y \in \mathbb{R}^N$ (resp. \mathbb{C}^N) has at least one representation $y = \sum_k x_k g_k = \mathbf{D}x$ with coefficient vector $x \in \mathbb{R}^K$ (resp. $x \in \mathbb{C}^K$). When \mathbf{D} is **redundant** ($K > N$), among the infinite number of representations of a vector y , it is often desirable to choose a *sparse* one. However, sparsity can be measured with diverse quantities such as the ℓ^p measures $\|x\|_p := \sum_k |x_k|^p$ for $0 \leq p \leq 1$ (with the convention $t^0 = 1$ if $t > 0$ and $0^0 = 0$). In this paper, we consider a large class \mathcal{M} of **admissible sparsity measures** $\|x\|_f := \sum_k f(|x_k|)$ where $f : [0, \infty) \rightarrow [0, \infty)$ is non-decreasing, not identically zero, $f(0) = 0$ and $t \mapsto f(t)/t$ is non-increasing on $(0, \infty)$. To each sparsity measure corresponds the “*f*-sparsest representation” optimization problem

$$\text{minimize } \|x\|_f \text{ subject to } y = \sum_k x_k g_k. \quad (1)$$

We address two rather natural questions related to this class of problems:

- 1/ when is the *f*-sparsest representation of y unique?
- 2/ if it is unique, does it depend on the choice of the sparsity measure f ?

Our main result³ is that when a signal y has a very sparse representation (in terms of the total number $\|x\|_0$ of nonzero coefficients), this representation is the *simultaneous and unique* sparsest representation with any admissible sparsity measure. More precisely we have the following theorem.

Theorem 1. *Let \mathbf{D} be a dictionary. Assume m is an integer such that for any x and y with $y = \mathbf{D}x$ and $\|x\|_0 \leq m$, x is the unique ℓ^1 -sparsest representation of y . Then, for any x and y such that $y = \mathbf{D}x$ and $\|x\|_0 \leq m$, x is indeed the unique *f*-sparsest representation of y for any admissible sparsity measure. In particular it is the ℓ^p -sparsest representation for $0 \leq p \leq 1$.*

The interesting consequence is that if y has a *highly sparse representation* x (with at most m elements) from the dictionary, then the combinatorial/highly nonlinear search for the *f*-sparsest representation of y can be replaced with a polynomial time computation based on linear programming [2, 15], which solves the ℓ^1 -optimization problem.

This extends a series of recent results about recovery of sparse expansions from dictionaries by Pursuit algorithms. In the early 1990’s, the Matching Pursuit and Basis Pursuit strategies were introduced with the purpose of getting good representations of signals with redundant dictionaries. Soon, it was experimentally noticed that, when y has a sufficiently sparse expansions (in the sense of $\|x\|_0$) in the Dirac/Fourier dictionary, Basis Pursuit can exactly recover it. The experimental observation was turned into a theorem and extended to unions of “incoherent” bases as well as to more general “incoherent” dictionaries [8, 5–7, 13, 3, 11]. Theorems in the same spirit were also recently proved, under slightly stronger assumptions, for exact recovery with Matching Pursuits [9, 10, 16, 14].

³ We refer the reader to our technical report [12] for the proofs.

Previous Basis Pursuit results stated that if y has an expansion x with $\|x\|_0$ sufficiently small, then x is *simultaneously* the *unique* ℓ^0 -sparsest and ℓ^1 -sparsest representation of y [2, 15]. In between the ℓ^0 and the ℓ^1 sparsity measures lie the ℓ^p ones, and it seemed only natural that by some sort of “interpolation”, the Basis Pursuit results should extend to *simultaneous uniqueness* of the ℓ^p -sparsest representations. It turns out that the interpolation can be done and our result show that it extends to the much larger class of admissible sparsity measures.

The structure of the paper is as follows. In Section 2 we give general conditions on an index set $I \subset K$ such that any expansion $y = \mathbf{D}_I x$ from the sub-dictionary $\mathbf{D}_I := \{g_k, k \in I\}$ is the unique f -sparsest representation of y in the whole dictionary. An example shows that the admissible sparsity measures f for which the conditions are satisfied may depend on the considered index set I . In Section 3 we give our main theorems and obtain necessary and sufficient conditions $\text{card}(I) \leq m_f(\mathbf{D})$ which ensure that for all admissible sparsity measures $g \in \mathcal{M}$ “between” a given $f \in \mathcal{M}$ and the ℓ^0 sparsity measure, the highly sparse representation is unique and independent of g . To conclude this paper, we briefly discuss how the numbers $m_f(\mathbf{D})$ which appear in the “highly sparse” conditions can be estimated from the *coherence* of the dictionary.

2 Sufficient Uniqueness Conditions

In this section we provide sufficient conditions on a representation $y = \mathbf{D}x$ which ensure that x is the unique f -sparsest representation of y , where f is an arbitrary admissible sparsity measure. A crucial property of $f \in \mathcal{M}$ is the quasi-triangle inequality

$$f(|u + v|) \leq f(|u| + |v|) \leq f(|u|) + f(|v|) \quad (2)$$

which is an easy consequence of the fact that $t \mapsto f(t)/t$ is non-increasing, see [12]. The sufficient uniqueness conditions are expressed in terms of the **support** $I(x) := \{k, x_k \neq 0\}$ of the coefficient vector $x = (x_k) \in \mathbb{R}^K$ (resp. \mathbb{C}^K), *i.e.* they depend on the set of elements of the dictionary which are used in the representation. The **kernel** $\text{Ker}(\mathbf{D}) := \{z, \mathbf{D}z = 0\}$ of the dictionary will play a special role. For $f \in \mathcal{M}$, \mathbf{D} a dictionary and $I \subset K$ a set of indices, we define

$$\theta_f(I, z) := \frac{\sum_{k \in I} f(|z_k|)}{\|z\|_f} \quad \text{and} \quad \Theta_f(I, \mathbf{D}) := \sup_{z \in \text{Ker}(\mathbf{D}), z \neq 0} \theta_f(I, z) \quad (3)$$

The value of $\Theta_f(I, \mathbf{D})$ (almost) completely characterizes the uniqueness of f -sparsest expansions from \mathbf{D}_I , as expressed in the following lemma.

Lemma 1. *Let \mathbf{D} be a dictionary, f an admissible sparsity measure, and $I \subset K$ an index set.*

1. *Assume that for all $z \in \text{Ker}(\mathbf{D})$ ($z \neq 0$), $\theta_f(I, z) < 1/2$, and let x, y such that $y = \mathbf{D}x$. If $I(x) \subset I$, x is the unique f -sparsest representation of y .*
2. *Assume that for some $z \in \text{Ker}(\mathbf{D})$, $\theta_f(I, z) \geq 1/2$. Then, there exists $x \neq x'$ such that $\mathbf{D}x = \mathbf{D}x'$, $I(x) \subset I$ and $\|x'\|_f \leq \|x\|_f$.*

The proof is a slight refinement of ideas from [5, 13], see [12].

Even though the value of $\Theta_f(I, \mathbf{D})$ essentially characterizes the uniqueness of the f -sparsest representation of expansions from the sub-dictionary $\mathbf{D}_I = \{g_k\}_{k \in I}$, its evaluation for a given index set I is not trivial in general. In particular, 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 sparsity measures f . The following example shows that f -sparsest representations do not necessarily coincide for different f , and that estimating $\Theta_f(I, \mathbf{D})$ for some admissible sparsity measure $f \in \mathcal{M}$ does not tell much about $\Theta_g(I, \mathbf{D})$ for other ones $g \in \mathcal{M}$.

Example 1. Let $\mathbf{B} = [g_1, \dots, g_N]$ 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 $z = (1, \dots, 1, \sqrt{N})$. Let us consider $I = \{1 \leq k \leq L\}$ an index set where $L \leq N$ and denote Θ_p for Θ_{f_p} where $f_p(t) = t^p$, $0 \leq p \leq 1$. Since

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

we have $\Theta_1(I, \mathbf{D}) < 1/2 < \Theta_0(I, \mathbf{D})$ whenever $(N + 1)/2 < L < (N + \sqrt{N})/2$. On the other hand, let us now consider $J = \{1 \leq k \leq L\} \cup \{N + 1\}$. As

$$\Theta_1(J, \mathbf{D}) = \frac{L + \sqrt{N}}{N + \sqrt{N}} \quad \text{and} \quad \Theta_0(J, \mathbf{D}) = \frac{L + 1}{N + 1}$$

we obtain $\Theta_0(J, \mathbf{D}) < 1/2 < \Theta_1(J, \mathbf{D})$ whenever $(N - \sqrt{N})/2 < L < (N - 1)/2$.

3 Uniqueness of highly sparse expansions

In the previous section, Example 1 illustrated the fact that, for arbitrary index sets I , not much can be said about the simultaneity of the f -sparsest representation for different admissible sparsity measures. In this section, we will show that the picture completely changes when we look for conditions on the *cardinal* of I so that $\Theta_f(I, \mathbf{D}) < 1/2$. Let us immediately state the main results of this section. The first result gives the theorem advertised in the introduction, which is the natural generalization to a series of recent results [8, 5–7, 13, 3, 11].

Theorem 2. *Let \mathbf{D} be a dictionary, and f an admissible sparsity measure. Let m be an integer and assume that whenever $y = \mathbf{D}x$ with $\|x\|_0 \leq m$, x is the ℓ^1 -sparsest representation of y . Then, whenever $y = \mathbf{D}x$ with $\|x\|_0 \leq m$, x is the simultaneous unique f -sparsest representation of y for any $f \in \mathcal{M}$.*

Theorem 2 is indeed only a special case of the following more general result.

Theorem 3. *Let \mathbf{D} be a dictionary, and f an admissible sparsity measure. Let m be an integer and assume that whenever $y = \mathbf{D}x$ with $\|x\|_0 \leq m$, x is the f -sparsest representation of y . Then, whenever $y = \mathbf{D}x$ with $\|x\|_0 \leq m$, x is the simultaneous unique $(g \circ f)$ -sparsest representation of y for any $g \in \mathcal{M}$.*

Note that one can easily check that if $f, g \in \mathcal{M}$ then $g \circ f \in \mathcal{M}$.

3.1 Sketch of the proof of Theorem 3

We will study in more details in the next section which integers m satisfy the assumptions of Theorem 3, but let us first sketch the proof. For any sequence $z = \{z_k\}_{k \in K}$, denote $|z|^\star$ a 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 slight abuse of notation, consider the “growth function”

$$\theta_f(m, z) := \max_{\text{card}(I) \leq m} \theta_f(I, z) = \frac{\sum_{k=1}^m f(|z|_k^\star)}{\|z\|_f} = \theta_f(m, |z|^\star) \quad (4)$$

defined for any $f \in \mathcal{M}$, $m \geq 0$ and $z \neq 0$. We have the following lemma [12].

Lemma 2. *For any $f, g \in \mathcal{M}$, $m \geq 0$ and $z \neq 0$ we have*

$$\theta_0(m, z) \leq \theta_{g \circ f}(m, z) \leq \theta_f(m, z) \leq \theta_1(m, z). \quad (5)$$

Let us just mention that the result relies crucially on the property that $t \mapsto f(t)/t$ is non-increasing, since the fact that $\theta_f(m, z) \leq \theta_1(m, z)$ for all m and z implies in particular that for any $a < b$ we must have $f(b)/(f(a) + f(b)) \leq b/(a + b)$, i.e., $1 + f(a)/f(b) \geq 1 + a/b$.

Theorem 3 is proved as follows: from Lemma 1, the assumption on m implies that, for all I with $\text{card}(I) \leq m$ and $z \in \text{Ker}(\mathbf{D})$ ($z \neq 0$), $\theta_f(m, z) < 1/2$. It follows from Lemma 2 that for all such I and z , and any $g \in \mathcal{M}$, $\theta_{g \circ f}(m, z) < 1/2$, which gives the desired result using again Lemma 1.

3.2 Explicit sparsity conditions

For any dictionary \mathbf{D} and sparsity measure f , one can consider the largest integer $m_f(\mathbf{D})$ that satisfies the assumption of Theorem 3, i.e., such that for any x and y such that $y = \mathbf{D}x$ and $\|x\|_0 \leq m$, x is indeed the unique f -sparsest representation of y . Another formulation of Theorem 3 is simply that for any $f, g \in \mathcal{M}$, $m_{g \circ f}(\mathbf{D}) \geq m_f(\mathbf{D})$. Indeed, it follows from Lemma 2 that

$$m_0(\mathbf{D}) \geq m_{g \circ f}(\mathbf{D}) \geq m_f(\mathbf{D}) \geq m_1(\mathbf{D}) \quad (6)$$

where m_p , $0 \leq p \leq 1$ is a shorthand for m_{f_p} with $f_p(t) := t^p$.

It is a challenge to compute the numbers $m_f(\mathbf{D})$ for an arbitrary dictionary (the computation of $m_0(\mathbf{D})$ is generally NP-hard). Let us however give a few examples of dictionaries where it is possible to get some non trivial bounds on the **strong sparsity number** $m_1(\mathbf{D})$ and the **weak sparsity number** $m_0(\mathbf{D})$ based on easily computable characteristics of the dictionary. Denoting $\lfloor t \rfloor$ the largest integer such that $\lfloor t \rfloor < t \leq \lfloor t \rfloor + 1$, we have the following lemma [12].

Lemma 3. *For any admissible sparsity measure $f \in \mathcal{M}$ and any dictionary \mathbf{D} ,*

$$m_f(\mathbf{D}) \geq m_1(\mathbf{D}) \geq \lfloor Z_1(\mathbf{D})/2 \rfloor \quad \text{and} \quad m_0(\mathbf{D}) = \lfloor Z_0(\mathbf{D})/2 \rfloor \quad (7)$$

where

$$Z_0(\mathbf{D}) := \inf_{z \in \text{Ker}(\mathbf{D}), z \neq 0} \|z\|_0 \quad \text{and} \quad Z_1(\mathbf{D}) := \inf_{z \in \text{Ker}(\mathbf{D}), \|z\|_\infty = 1} \|z\|_1. \quad (8)$$

are respectively called the **spark** and the **spread** of the dictionary.

The spark was introduced in [3] and its numerical computation is generally combinatorial. The spread was introduced by the authors in [11]. The above estimates are not quite explicit, but the next one is easily computable.

Lemma 4. *The coherence of a dictionary $\mathbf{D} = \{g_k\}$ is defined [5] as*

$$M(\mathbf{D}) := \sup_{k \neq k'} |\langle g_k, g_{k'} \rangle|. \quad (9)$$

For any admissible sparseness measure $f \in \mathcal{M}$ we have the lower estimate

$$m_f(\mathbf{D}) \geq m_1(\mathbf{D}) \geq \lfloor (1 + 1/M(\mathbf{D})) / 2 \rfloor. \quad (10)$$

Proof. Consider $x \in \text{Ker}(\mathbf{D})$. For every k we have $x_k g_k = -\sum_{k' \neq k} x_{k'} g_{k'}$ hence, taking the inner product of both hand sides with g_k , $|x_k| \leq M(\mathbf{D}) \cdot \sum_{k' \neq k} |x_{k'}|$. It follows that $(1 + M) \cdot |x_k| \leq M \cdot \|x\|_1$. Taking the supremum over k we get $(1 + M)\|x\|_\infty \leq M \cdot \|x\|_1$ or equivalently $Z_1(\mathbf{D}) \geq 1 + 1/M$, and the result follows using Lemma 3.

When \mathbf{D} contains an orthonormal basis \mathbf{B} in dimension N , the coherence satisfies $M(\mathbf{D}) \geq 1/\sqrt{N}$, and it is possible to find up to $N+1$ orthonormal bases $\{\mathbf{B}_j\}_{j=1}^N$ such that their union $\mathbf{D} := [\mathbf{B}_1 \dots \mathbf{B}_{N+1}]$ is a dictionary of coherence $m(\mathbf{D}) = 1/\sqrt{N}$. For such highly redundant dictionaries, the lemma shows that $m_1(\mathbf{D}) \geq \lfloor (1 + \sqrt{N})/2 \rfloor$. Lemma 4 was in germ in Donoho and Huo's early paper [5] on exact recovery of sparse expansion through Basis Pursuit, where it was only used for \mathbf{D} a union of two orthonormal bases and $f(t) = t^p$, $p \in \{0, 1\}$. In [13] and [3] it was extended to arbitrary dictionaries, and in [11] to $f(t) = t^p$, $p \in [0, 1]$. Finer estimates of $m_1(\mathbf{D})$ can be obtained from the properties of the Gram matrix of \mathbf{D} , see [12].

4 Conclusion and statistical perspectives

We have studied sparse representation of signals using an arbitrary dictionary and a very general admissible sparsity measure $\|\cdot\|_f$. Given a dictionary and a signal y , we provided sufficient conditions for the minimization problem

$$\text{minimize } \|x\|_f \text{ subject to } y = \sum_k x_k g_k, \quad (11)$$

to have the same unique solution as the problem

$$\text{minimize } \|x\|_1 \text{ subject to } y = \sum_k x_k g_k, \quad (12)$$

and the conditions are *independent* of the particular admissible sparsity measure f . The latter minimization problem (12) can be solved using a linear programming technique, *i.e.*, by a polynomial time algorithm. For a dictionary in a Hilbert space we proved that the condition $\|x\|_0 \leq \lfloor 1/2(1 + 1/M) \rfloor$, where M is the coherence of the dictionary, is sufficient for (11) to have the same solution as (12) for any sparsity measure f . The results generalize previous results by Donoho and Elad [3] and by the authors [13], where only two types of sparsity measures were considered: the ℓ^0 -norm and the ℓ^1 -norm.

The f -sparsest representation problems (11) that we have considered in this paper are related to the statistical problem of Bayesian estimation of unknown parameters (x_k) given the noise-free observation $y = \mathbf{D}x$ and the prior probability density function $P_{f,h}(x) = \frac{1}{Z_{f,h}} \exp(-h(\|x\|_f))$, where $h : [0, \infty) \rightarrow [0, \infty)$ is an increasing function and $Z_{f,h}$ a normalizing constant such that $P_{f,h}(x)$ is a probability density on \mathbb{R}^K (resp. \mathbb{C}^K).

In this Bayesian estimation setting, our results have an interpretation in terms of the *robust estimation* with respect to modeling error. Assume that the prior on x has the above structure, where $\|\cdot\|_f$ is an admissible sparsity measure. Then, for any noise-free observation y that admits a sufficiently sparse representation (with $\|x\|_0 \leq m_1(\mathbf{D})$), it does not matter which admissible sparse prior we use to model the data and search for the sparsest representation: each admissible sparse model yields the same estimate, which is indeed the MAP estimate under the true prior. In particular, we can as well model the parameters with a Laplacian prior $P_1(x) \propto \exp(-\|x\|_1)$, and this relaxed model will recover the “good” parameters (x_k) .

To see how strong is the robustness to modeling error, let us simply give an example. First, notice that the Laplacian prior is a model where we assume the independence of the x_k , since $P_1(x) = \prod_k P_1(x_k)$. However, as shown in [12], since the class \mathcal{M} of admissible sparsity measures is stable by $\min(\cdot)$ and $\max(\cdot)$, it contains some nontrivial measures such as $\|\cdot\|_f$ with

$$f(t) := \max(t/2, \min(t^{1/2}, t^0)) = \begin{cases} \sqrt{t}, & 0 \leq t \leq 1 \\ 1, & 1 \leq t \leq 2 \\ t/2, & 2 \leq t < \infty \end{cases} . \quad (13)$$

Moreover, the use of a “sufficiently increasing” function h to define $P_{f,h}$ can introduce a dependence between the coefficients x_k , since $P_{f,h}$ will no longer be the product of its marginals. Yet, if the solution to the true Bayesian estimation problem is sparse enough, it will be recovered with the Laplacian model, where the parameters are assumed independent!

The main limitation to the theory developed in this paper is certainly that the sparsity condition $\|x\|_0 \leq m_1(\mathbf{D})$ is quite restrictive, since the set of observations y that admits such a sparse representation is of Lebesgue measure zero in \mathbb{R}^K and probability zero under the sparse prior. A second limitation comes from the fact that the results do not apply to noisy data $y = \mathbf{D}x + n$. Recent results [17, 4] indicate that similar robustness properties can be proved even with approximate and noisy sparse representations, and the authors are also

investigating the problem of simultaneous sparse representation/approximation of several observations in a single dictionary, which is a widely-spread tool to perform blind source separation [18].

References

1. L. Benaroya, R. Gribonval, and F. Bimbot. Représentations parcimonieuses pour la séparation de sources avec un seul capteur. In *GRETSI 2001*, Toulouse, France, 2001. Article # 434.
2. D. Bertsekas. *Non-Linear Programming*. Athena Scientific, Belmont, MA, 2nd edition, 1995.
3. D. Donoho and M. Elad. Optimally sparse representation in general (non-orthogonal) dictionaries via ℓ^1 minimization. *Proc. Nat. Aca. Sci.*, 100(5):2197–2202, Mar. 2003.
4. D. Donoho, M. Elad, and V. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. Working draft, Feb. 2004.
5. D. Donoho and X. Huo. Uncertainty principles and ideal atomic decompositions. *IEEE Trans. Inform. Theory*, 47(7):2845–2862, Nov. 2001.
6. 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.
7. A. Feuer and A. Nemirovsky. On sparse representations in pairs of bases. *IEEE Trans. Inform. Theory*, 49(6):1579–1581, June 2003.
8. J.-J. Fuchs. On sparse representations in arbitrary redundant bases. Technical report, IRISA, Dec. 2003. to appear in *IEEE Trans. Inform. Theory*.
9. 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)*, Jan. 2003.
10. A. Gilbert, S. Muthukrishnan, M. Strauss, and J. Tropp. Improved sparse approximation over quasi-incoherent dictionaries. In *Int. Conf. on Image Proc. (ICIP'03)*, Barcelona, Spain, Sept. 2003.
11. 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.
12. R. Gribonval and M. Nielsen. Highly sparse representations from dictionaries are unique and independent of the sparseness measure. Technical Report R-2003-16, Dept of Math. Sciences, Aalborg University, Oct. 2003.
13. R. Gribonval and M. Nielsen. Sparse decompositions in unions of bases. *IEEE Trans. Inform. Theory*, 49(12):3320–3325, Dec. 2003.
14. R. Gribonval and P. Vandergheynst. Exponential convergence of Matching Pursuit in quasi-incoherent dictionaries. Technical report, IRISA, 2004. in preparation.
15. A. Shrijver. *Theory of Linear and Integer Programming*. John Wiley, 1998.
16. J. Tropp. Greed is good : Algorithmic results for sparse approximation. Technical report, Texas Institute for Computational Engineering and Sciences, 2003.
17. J. Tropp. Just relax: Convex programming methods for subset selection and sparse approximation. Technical Report ICES Report 04-04, UT-Austin, Feb. 2004.
18. M. Zibulevsky and B. Pearlmutter. Blind source separation by sparse decomposition in a signal dictionary. *Neural Computations*, 13(4):863–882, 2001.