On the exponential convergence of matching pursuits in quasi-incoherent dictionaries — Source link ↗

Rémi Gribonval, Pierre Vandergheynst

Institutions: French Institute for Research in Computer Science and Automation, École Polytechnique Fédérale de Lausanne

Related papers:

- Greed is good: algorithmic results for sparse approximation

- Matching pursuits with time-frequency dictionaries

- Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition

- Atomic Decomposition by Basis Pursuit

- Compressed sensing

On the exponential convergence of Matching Pursuits in quasi-incoherent dictionaries

Rémi Gribonval, Pierre Vandergheynst

▶ To cite this version:

[10] S. Natarajan, "Large deviations, hypothesis testing, and source coding for finite Markov chains," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 3, pp. 360–365, May 1985.

[11] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of minimum phase systems," *IEEE Trans. Inf. Theory*, vol. 36, no. 2, pp. 312–321, Mar. 1990.

[12] S. M. Sowelam and A. H. Tewfik, "Waveform selection in radar target classification," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 1014–1029, May 2000.

[13] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: A time domain approach," *IEEE Trans. Inf. Theory*, vol. 40, no. 2, pp. 340–349, Mar. 1994.

[14] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*. New York: Wiley, 1968, pt. 1.

On the Exponential Convergence of Matching Pursuits in Quasi-Incoherent Dictionaries

Rémi Gribonval, *Member, IEEE,* and
Pierre Vandergheynst, *Member, IEEE*

Abstract—The purpose of this correspondence is to extend results by Villemoes and Temlyakov about exponential convergence of Matching Pursuit (MP) with some structured dictionaries for "simple" functions in finite or infinite dimension. The results are based on an extension of Tropp's results about Orthogonal Matching Pursuit (OMP) in finite dimension, with the observation that it does not only work for OMP but also for MP. The main contribution is a detailed analysis of the approximation and stability properties of MP with quasi-incoherent dictionaries, and a bound on the number of steps sufficient to reach an error no larger than a penalization factor times the best m-term approximation error.

Index Terms—Dictionary, greedy algorithm, matching pursuit (MP), nonlinear approximation, sparse representation.

I. INTRODUCTION

In a Hilbert space $\mathcal{H}$ of finite or infinite dimension, we consider the problem of getting m-term approximants of a function f from a possibly redundant *dictionary* $\mathcal{D} = \{g_k, k \in \mathbb{Z}\}$ of unit norm basis functions also called *atoms*. It will often be convenient to see a dictionary as a synthesis operator (or, in finite dimension, as a matrix)

$$\mathbf{D} : \boldsymbol{c} = (c_k) \mapsto \mathbf{D}\boldsymbol{c} = \sum_k c_k g_k$$

that maps sequences to vectors in $\mathcal{H}$. A special class of dictionaries that is widely used in signal and image processing is the family of frames: a dictionary $\mathcal{D}$ is a frame for $\mathcal{H}$ if, and only if, $\mathbf{D}$ is a bounded operator from ℓ^2 onto $\mathcal{H}$ [2]. However, in this correspondence, we consider dictionaries that may not be frames, hence $\mathbf{D}$ shall be defined essentially on sequences $\boldsymbol{c}$ with a finite number of nonzero entries. For any

index set I (not necessarily finite) we will also consider the restricted synthesis operator

$$\mathbf{D}_I : \boldsymbol{c} \mapsto \mathbf{D}_I \boldsymbol{c} = \sum_{k \in I} c_k g_k$$

that corresponds to the subset $\mathcal{D}_I = \{g_k, k \in I\}$ of the full dictionary.

When $\mathcal{D}$ is an orthonormal basis for $\mathcal{H}$, it is well known how to get the best m-term approximant to any f: the solution is to keep the m atoms of the basis which have the largest inner products $|\langle f, g_k \rangle|$ with f. However, for arbitrary redundant dictionaries, the problem becomes NP-hard [3]. In the recent years, much effort has been made to understand what structure should be imposed on f (for a given dictionary) or on the dictionary itself so that good approximants can be obtained with computationally feasible algorithms.

One of the first algorithms that appeared in the signal processing community for approximating signals from a redundant dictionary was the Matching Pursuit (MP) algorithm of Mallat and Zhang [25], which iteratively decomposes the analyzed function f into an m-term approximant $f_m = \sum_{n=1}^m \alpha_n g_{k_n}$ and a residual $r_m = f - f_m$. MP is also known as Projection Pursuit in the statistics community [10], [22] and as a Pure Greedy Algorithm [27] in the approximation community. In finite dimension, MP is known to converge exponentially, i.e., for some $0 < \beta < 1$

$$\|r_m\|^2 = \|f_m - f\|^2 \leq \beta^m \cdot \|f\|^2, \qquad m \geq 1.$$

In infinite-dimensional Hilbert spaces, Jones [24] proved that MP is still convergent, i.e., $\|f_m - f\| \to 0$, but gave no estimate of the speed of convergence. DeVore and Temlyakov [4] exhibited a "bad" dictionary $\mathcal{D}$ where there exists a "simple" function (sum of two dictionary elements) for which MP gives "bad" approximations (i.e., with a slow convergence $\|f_m - f\| \geq C m^{-1/2}$). On the positive side, Villemoes [30] showed that for Walsh wavelet packets, MP on "simple" functions ($f = c_i g_i + c_j g_j$ any sum of any two wavelet packets) was exponentially convergent (just as MP in finite dimension) with $\|f_m - f\|^2 \leq (3/4)^m \|f\|^2$. Temlyakov obtained similar results [26]: in particular, for f a function on the erval $[0, 1)$ taking constant values on a partition of $[0, 1)$ into n disjoint intervals, and $\mathcal{D}$ a highly redundant dictionary containing all (normalized) characteristic functions of intervals $I \subset [0, 1)$

$$\|f_m - f\|_2 \leq (1 - 1/n)^{m/2} \|f\|_2.$$

In this correspondence, we extend Villemoes and Temlyakov results about MP to more general dictionaries and "simple functions," as stated in the following featured theorem.

Featured Theorem 1: Let $\mathcal{D}$ be a dictionary in a finite- or infinite-dimensional Hilbert space and I an index set such that the stability condition (SC)

$$\eta(I) := \sup_{k \notin I} \|(\mathbf{D}_I)^\dagger g_k\|_1 < 1 \tag{1}$$

is met, where $(\cdot)^\dagger$ denotes pseudoinversion.[1] Then, for any $f = \sum_{k \in I} c_k g_k \in \mathrm{span}(g_k, k \in I)$, MP

1) picks up only correct atoms at each step: $(\forall n, k_n \in I)$;
2) if I is a finite set, then the residual r_m converges exponentially to zero.

The stability condition (1) may look fairly abstract, but for so-called *quasi-incoherent* dictionaries, one can obtain more explicit sufficient conditions [28]. For such dictionaries, we derive estimates of the rate of

---

[1]Basic reminders on pseudoinversion are given in Section II-E

exponential convergence of MP, and we obtain the following featured theorem.

Featured Theorem 2: Let $\mathcal{D}$ be a dictionary in a finite- or infinite-dimensional Hilbert space and let $\mu : \max_{k \neq l} |\langle g_k, g_l \rangle|$ be its *coherence*. For any finite index set I of size card $(I) = m < (1 + 1/\mu)/2$ and any $f = \sum_{k \in I} c_k g_k \in \mathrm{span}\,(g_k, \ k \in I)$, MP

1) picks up only correct atoms at each step: $(\forall n, \ k_n \in I)$;
2) converges exponentially

$$\|f_n - f\|^2 \leq ((1 - 1/m)(1 + \mu))^n \|f\|^2.$$

The proof of this theorem is based on an argument given by Tropp [28] where the condition (1) is called "Exact Recovery Condition" (ERC) because it ensures that Orthonormal Matching Pursuit (OMP) and Basis Pursuit (BP) exactly recover any

$$f = \sum_{k \in I} c_k g_k \in \mathrm{span}\,(g_k, \ k \in I).$$

We have chosen to rename the ERC a "stability condition." Indeed, for MP one cannot strictly speak about recovery, however, the theorem is definitely a stability result since all residuals remain in the subspace span $(g_k, \ k \in I) \subset \mathcal{H}$. Tropp's result was the last of a series of "recovery" results: first with the BP "algorithm"—which was introduced [1] as an alternative to MP since the latter cannot resolve close atoms—under some assumptions on both the analyzed function and the dictionary [6]–[9], [20], [19]; then with variants of the MP [12], [13]. After the first draft of this manuscript was submitted for publication, it came to our attention that Donoho, Elad, and Temlyakov also consider stability and recovery properties of MP in incoherent dictionaries [5]. We discuss in more details in Section V how our results are connected to other approaches.

The previous theorems only explain the behavior of MP on *exact* expansions, i.e., they require that the approximated function f be exactly expressed as an expansion from a "good" set of atoms. However, real signals or images almost never have such a simple expansion in practical dictionaries. Fortunately, just as for OMP [28], the analysis of MP as an approximation algorithm can be carried out by taking into account how well a function is *approximated* by an expansion from a good set of atoms. In particular, our results lead to the following theorem (with the notations of Featured Theorem 2)

Featured Theorem 3: Let $\{f_n\}$ be a sequence of approximants to $f \in \mathcal{H}$ produced with MP with g_{k_n} the corresponding atoms. Let $m < (1 + 1/\mu)/4$ and let $f_m^\star = \sum_{k \in I_m^\star} c_k g_k$ be a best m-term approximant to f from $\mathcal{D}$, i.e.,

$$\|f_m^\star - f\| = \sigma_m(f) := \inf\{\|f - D_I c\|, \ \mathrm{card}(I) \leq m, c \in \mathbb{C}^I\}.$$

Then, there is a number N_m such that

1) the error after N_m steps of MP satisfies

$$\|f_{N_m} - f\| \leq \sqrt{1 + 4m}\ \sigma_m;$$

2) during the first N_m steps, MP picks up atoms from the best m-term approximant: $k_n \in I_m^\star$;
3) if $\sigma_m^2 < 3\sigma_1^2/m$ then N_m is no larger than

$$N_m \leq 2 + m \cdot \frac{4}{3} \cdot \ln \frac{3\sigma_1^2}{m\sigma_m^2}.$$

In the course of this correspondence, we actually prove slightly more general results (Theorems 1–4) and particularize them to get our featured results (Featured Theorems 1 and 3). The structure of this correspondence is as follows. In Section II, we recall the definition of MP and several variants thereof, and prove the stability result (Featured Theorem 1). In Section III, we particularize this result to a special class

of dictionaries, *quasi-incoherent dictionaries*. This allows us to obtain constraints on the dictionary so that the stability condition is met and we also give estimates on the rate of convergence of MP in these cases (Featured Theorem 2). Finally, in Section IV, we explore the approximation properties of various flavors of MP. In particular, we show that greedy algorithms may robustly select atoms participating in a near best m-term approximation and give the resulting approximation bounds (Featured Theorem 3).

The proof of Featured Theorem 1 is merely a rewriting of Tropp's proof with the observation that it does not only work for OMP but also for MP. Thus, the main contribution of this correspondence is in the study of the approximation and stability properties of greedy algorithms with quasi-incoherent dictionaries.

II. MATCHING PURSUIT(S) ON "SIMPLE" EXPANSIONS

In this section, we first recall the definition of MP and several variants thereof, then we prove the stability of all these variants, in the sense of Featured Theorem 1.

A. Matching Pursuit (MP)

MP is an iterative algorithm that builds n-term approximants f_m and residuals $r_n = f - f_n$ by adding one term at a time in the approximant. It works as follows. At the beginning, we set $f_0 = 0$ and $r_0 = f$; assuming f_n and r_n are defined, we set

$$|\langle r_n, g_{k_{n+1}} \rangle| = \sup_k |\langle r_n, g_k \rangle| \tag{2}$$

$$f_{n+1} = f_n + \langle r_n, g_{k_{n+1}} \rangle g_{k_{n+1}} \tag{3}$$

and compute a new residual as $r_{n+1} = f - f_{n+1}$.

B. Weak MPs

When the dictionary is infinite, the supremum in (2) may not be attained, so one may have to consider the so-called *weak* selection rule

$$|\langle r_n, g_{k_{n+1}} \rangle| \geq \alpha \sup_k |\langle r_n g_k \rangle| \tag{4}$$

with some fixed $0 < \alpha \leq 1$ independent of n. Corresponding variants of MP will be called Weak MP with weakness parameter α, or in short Weak (α) MP or even Weak MP when the value of α does not need to be specified.

C. Orthonormal MP

Moreover, once m atoms have been selected, the approximant

$$f_m = \sum_{n=0}^{m-1} \langle r_n, g_{k_{n+1}} \rangle g_{k_{n+1}}$$

is generally not the best approximant to f from the finite-dimensional subspace $\mathcal{V}_m := \mathrm{span}(g_{k_1}, \ldots, g_{k_m})$. OMP, respectively, Weak (α) OMP, replaces the update rule (3) with

$$f_{n+1} = P_{\mathcal{V}_{n+1}} f \tag{5}$$

where $P_{\mathcal{V}}$ is the orthonormal projector onto the finite dimensional subspace $\mathcal{V}$.

D. General MPs

More generally, one can consider the family of approximation algorithms based on the repeated application of two steps:

1) a (weak) selection step according to (4);
2) an update step where a new approximant $f_{n+1} \in \mathcal{V}_{n+1}$ is chosen.

Algorithms from this larger family will be called General MP, Weak (α) General MP, or Weak General MP. Examples of Weak (α) General

MP algorithms include the High Resolution Pursuits [23], [16], which were introduced to attenuate the lack of resolution of plain MP with time–frequency dictionaries in the time domain.

E. Stability of Weak (α) General MP

The major result of Tropp [28] is what he calls the "Exact Recovery Condition"

$$\eta(I) := \sup_{k \notin I} \|(\mathbf{D}_I)^\dagger g_k\|_1 < \alpha \qquad (6)$$

(where $(\cdot)^\dagger$ denotes pseudoinversion, see below): when the Exact Recovery Condition is met, Weak (α) OMP "exactly recovers" any linear combinations of atoms from the subdictionary $\mathcal{D}_I$, which means that Weak (α) OMP can only pick up correct atoms at each step. Tropp's proof indeed works for Weak (α) General MP, with the only difference that we no longer get exact recovery but only stability of the Pursuit, as stated in the following theorem.

Theorem 1: Let I be an index set (finite or infinite) with $\eta(I) < 1$. For any $f = \sum_{k \in I} c_k g_k$ and $\alpha > \eta(I)$, Weak (α) General MP picks up a correct atom at each step, i.e., for all $n \geq 1$, $k_n \in I$.

Note that we do not assume that the elements $\{g_k\}_{k \in I}$ are linearly independent for the result to hold true. Before giving the proof of the theorem, let us give a quick reminder on the notion of pseudoinverse. Most of this material can be found in the usual suspects [14], [21].

Let $\mathbf{A}$ be a linear operator and let Range $\mathbf{A}$ be its range. The pseudoinverse $\mathbf{A}^\dagger$ is the left inverse that is zero on $\{\mathrm{Range}\,\mathbf{A}\}^\perp$. It is also the left inverse of minimal sup norm. In the case of general p by q matrices, we will make use of the Moore–Penrose pseudoinverse. It is the unique q by p matrix that satisfies the following properties:

$$\mathbf{A}\mathbf{A}^\dagger\mathbf{A} = \mathbf{A}$$
$$\mathbf{A}^\dagger\mathbf{A}\mathbf{A}^\dagger = \mathbf{A}^\dagger$$
$$(\mathbf{A}\mathbf{A}^\dagger)^* = \mathbf{A}\mathbf{A}^\dagger \quad \text{and}$$
$$(\mathbf{A}^\dagger\mathbf{A})^* = \mathbf{A}^\dagger\mathbf{A}$$

where $(\cdot)^*$ denotes the adjoint. In particular, $\mathbf{A}\mathbf{A}^\dagger$ is an orthonormal projection onto Range $\mathbf{A}$. If the inverse of $\mathbf{A}^*\mathbf{A}$ exists, the Moore–Penrose pseudoinverse can simply be written

$$\mathbf{A}^\dagger = (\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*.$$

Proof of Theorem 1: Just as the proof of exactness of OMP by Tropp (which is a special case), we can show by induction that at each step MP picks up an atom $k_n \in I$, so the residual r_n remains in the finite-dimensional space $\mathcal{V}_I = \mathrm{span}(g_k, k \in I)$. Initially, we have by assumption $r_0 = f \in \mathcal{V}_I$. Assuming that $r_n \in \mathcal{V}_I$, we notice that the inner products $\{\langle r_n, g_k \rangle_{k \in I}\}$ between r_n and $\{g_k, \ k \in I\}$ are listed in the vector $\mathbf{D}_I^* r_n$ while those with $\{g_k, \ k \notin I\}$ are listed in $\mathbf{D}_{\bar{I}}^* r_n$ where $\bar{I} = \{k, k \notin I\}$. Thus, the atom $g_{k_{n+1}}$ is a correct one (i.e., $k_{n+1} \in I$) if and only if

$$\eta(I, r_n) := \frac{\|\mathbf{D}_{\bar{I}}^* r_n\|_\infty}{\|\mathbf{D}_I^* r_n\|_\infty} < \alpha.$$

The core of of the proof of [28, Theorem 3.1] yields $\eta(I, r_n) \leq \eta(I)$. From the assumption $\eta(I) < \alpha$, we can infer that $k_{n+1} \in I$ and $r_{n+1} \in \mathcal{V}_I$, and we get the theorem. $\square$

F. Recovery and Convergence

Suppose that the analyzed function f belongs to $\mathrm{span}(g_k, \ k \in I)$ where I satisfies $\eta(I) < 1$, and that we perform some Weak (α) General MP with $\alpha > \eta(I)$: Theorem 1 states that the Pursuit will only pick up correct atoms.

In the particular case of an *Orthogonal* Pursuit, since each residual r_n is orthogonal to previously selected atoms $g_{k_1}, \ldots g_{k_n}$, any atom can only be picked up once by the Pursuit. As a result, if in addition I is a finite set of cardinality m, the Orthogonal Pursuit exactly recovers f in m iterations: this is the main result formalized by Tropp and already present—though not with such a clear statement—in the results of Gilbert *et al.* [12], [13].

If the Pursuit we are performing on f is not orthogonal, it is known that convergence does *not* generally occur in a finite number of steps. However, if I is a finite set, the stability condition implies that the Pursuit is actually performed in the finite-dimensional space $\mathcal{V}_I$. In the case of Weak MP, it follows [25] that we have exponential convergence, just as stated in Featured Theorem 1. In the next section, we provide some tools to estimate the rate of this convergence, and it will turn out that they also make it possible to estimate the speed of convergence of (Weak) OMP.

III. MP IN QUASI-INCOHERENT DICTIONARIES

In the previous section, we have given fairly abstract conditions to ensure stability of Weak General MP, exact recovery with Weak OMP, and exponential convergence of Weak MP toward the approximated function. However, the quantity $\eta(I)$ that appears in the stability condition (6) is not very explicit, and we did not yet provide estimates for the rate of exponential convergence.

In this section, we will show that we can use the so-called *cumulative coherence function*[2] of the dictionary to estimate $\eta(I)$—and check the Stability Condition—as well as the rate of exponential convergence of Plain MP.

A. Cumulative Coherence Function and Coherence

Definition 1: Let $\mathcal{D}$ be a dictionary. Its cumulative coherence function is defined for each integer $m \geq 1$ as

$$\mu_1(m) := \max_{I \mid \mathrm{card}(I) \leq m} \max_{k \notin I} \sum_{l \in I} |\langle g_l, g_k \rangle|. \qquad (7)$$

As a special case, for $m = 1$, the value of the cumulative coherence function is the so-called coherence of the dictionary

$$\mu = \mu_1(1) = \max_{k \neq l} |\langle g_l, g_k \rangle|. \qquad (8)$$

One easily observes that the cumulative coherence function is subadditive

$$\mu_1(k + l) \leq \mu_1(k) + \mu_1(l), \qquad \forall k, l$$

hence, we have $\mu_1(m) \leq \mu \cdot m, m \geq 1$. A dictionary is called incoherent if μ is small: typically, in finite dimension N, any dictionary that strictly contains an orthonormal basis has coherence $\mu \geq 1/\sqrt{N}$. The union of the Dirac and the Fourier bases is an incoherent dictionary where indeed $\mu = 1/\sqrt{N}$ and $\mu_1(m) = \mu \cdot m$. When the cumulative coherence function grows no faster than $\mu \cdot m$, we say that the dictionary is quasi-incoherent.

B. Explicit Stability Condition and Rate of Convergence

Using Neumann series, Tropp proved that whenever I is of size m such that $\mu_1(m - 1) < 1$, we have the upper bound

$$\eta(I) \leq \frac{\mu_1(m)}{1 - \mu_1(m - 1)}. \qquad (9)$$

From this estimate, we can derive the following theorem which shows that the cumulative coherence function μ_1 can provide both a practical Stability Condition for Weak General MP and an estimate of the rate of exponential convergence for Weak MP.

[2]Formerly known as the Babel function.

Theorem 2: Let m be an integer such that

$$\mu_1(m) + \mu_1(m-1) < 1. \tag{10}$$

Then for any index set I of size at most m, any $f \in \operatorname{span}(g_k, k \in I)$, and $\alpha > \mu_1(m)/(1 - \mu_1(m-1))$.

1) Weak (α) General MP picks up a correct atom at each step, i.e., for all $n \geq 1, k_n \in I$.
2) Weak (α) MP/OMP converge exponentially to f: more precisely, we have $\|f - f_n\|^2 \leq (\beta_m(\alpha))^n \cdot \|f\|^2$ with

$$\beta_m(\alpha) := 1 - \alpha^2(1 - \mu_1(m-1))/m. \tag{11}$$

Before we prove the theorem, we need a few lemmas.

Lemma 1: For any index set I with card $(I) = m$, the squared singular values of $\mathbf{D}_I$ exceed $1 - \mu_1(m-1)$.

The proof relies on Gershgorin Disc Theorem and can be found in [12], [6], [15], [28], see for example [28, Lemma 2.3]. The second important lemma is due to DeVore and Temlyakov [4]; it gives a lower estimate on the amount of energy of a signal that can be removed in one step of MP.

Lemma 2 (DeVore,Temlyakov): For any I and c

$$\sup_{k \in I} |\langle \mathbf{D}_I c, g_k \rangle| \geq \frac{\|\mathbf{D}_I c\|^2}{\|c\|_1}.$$

We can now prove Theorem 2.

Proof of Theorem 2: The stability result is trivial using the estimate (9) together with Theorem 1. Let us proceed with the exponential convergence of Weak (α) MP/OMP. From the stability part we know that at each step the residual $r_n = f - f_n$ of Weak (α) MP/OMP is in $\mathcal{V}_I$. Thus, $r_n = \mathbf{D}_I c_n$ for some sequence c_n with at most m nonzero elements. Denoting λ, the smallest nonzero singular value of $\mathbf{D}_I$, it follows using Lemma 1 that

$$\|c_n\|_1^2 \leq m \|c_n\|_2^2 \leq \frac{m}{\lambda^2} \|\mathbf{D}_I c_n\|_2^2$$
$$\leq \frac{m}{1 - \mu_1(m-1)} \|r_n\|^2.$$

Then, by Lemma 2, we obtain

$$\sup_{k \in I} |\langle r_n, g_k \rangle| \geq \frac{\|r_n\|^2}{\|c_n\|_1} \geq \|r_n\| \sqrt{\frac{1 - \mu_1(m-1)}{m}}.$$

We conclude by noticing that

$$\|r_{n+1}\|^2 \overset{(a)}{\leq} \|r_n\|^2 - |\langle r_n, g_{k_{n+1}} \rangle|^2$$
$$\leq \|r_n\|^2 - \alpha^2 \sup_k |\langle r_n, g_k \rangle|^2$$
$$\leq \left(1 - \alpha^2(1 - \mu_1(m-1))/m\right) \cdot \|r_n\|^2$$
$$\leq \beta_m(\alpha) \cdot \|r_n\|^2 \leq \cdots \leq (\beta_m(\alpha))^{n+1} \|r_0\|^2$$
$$= (\beta_m(\alpha))^{n+1} \|f\|^2.$$

Notice that (a) is an equality for MP and an inequality for OMP. □

The above estimate is valid for the whole range of admissible weakness parameter α: $\alpha = 1$ corresponds to the standard "full search" Pursuit while $\alpha = \mu_1(m)/(1 - \mu_1(m-1))$ gives the worst case estimate corresponding to the limiting case of the weakest allowable Pursuit. To avoid carrying unnecessary heavy notations throughout the rest of the

correspondence, from now on we will only consider the case of a full search Pursuit.

C. Estimates Based on the Coherence

For any dictionary, we have seen that the cumulative coherence function can be bounded using the coherence as $\mu_1(m) \leq \mu \cdot m$, $m \geq 1$. Thus, a sufficient condition to get the stability condition (10) with the cumulative coherence function becomes a condition based on the coherence:

$$m < \frac{1}{2}\left(1 + \frac{1}{\mu}\right). \tag{12}$$

If the dictionary is a union of incoherent orthonormal bases in finite dimension N [20], then indeed $\mu_1(m) = \mu \cdot m$ for $1 \leq m \leq N$ and (12) is equivalent to (10). In any case, the rate $\beta_m = \beta_m(1)$ of exponential convergence of a (full search) MP is estimated from above by

$$\beta_m := \beta_m(1) = 1 - (1 - \mu_1(m-1))/m \leq (1 - 1/m)(1 + \mu). \tag{13}$$

The combination of (3) with Theorem 2 yields our Featured Theorem 2.

IV. MP AS AN APPROXIMATION ALGORITHM

So far we have considered the behavior of (Weak) MP on *exact* sparse expansions in the dictionary. However, the set of functions with an exact sparse expansion $f \in \operatorname{Range} \mathbf{D}_I$, card $(I) < \dim \mathcal{H}$ is negligible, hence, it is more interesting to know what is the behavior of Pursuits on more general vectors, typically on f "close enough" to some $f^\star$ with an exact sparse expansion.

A. Best M-Term Approximation

For any $f \in \mathcal{H}$ and m, the error of best m-term approximation to f from the dictionary is

$$\sigma_m(f) := \inf\{\|f - D_I c\|, \operatorname{card}(I) \leq m, c_k \in \mathbb{C}\}. \tag{14}$$

When there is no ambiguity about which f is considered, we will simply write σ_m. For $f \in \mathcal{H}$, let $f_m^\star = \sum_{k \in I_m} c_k g_k$ be a best m-term approximation to f, i.e., with card $(I_m) \leq m$ and $\|f - f_m^\star\| = \sigma_m$. If a best m-term approximant does not exist (because the infimum in the definition of σ_m is not reached), one can consider a *near best* m-term approximant by letting $\epsilon > 0$ and only requiring $\|f - f_m^\star\| = (1 + \epsilon)\sigma_m$. In any case, without loss of generality, we can assume that

1) the atoms $\{g_k, k \in I_m\}$ are linearly independent;
2) $f_m^\star$ is the orthogonal projection of f onto span $(g_k, k \in I_m)$;

else, we could easily replace $f_m^\star$ with a better m-term approximant to f by either changing the coefficients c_k or selecting a subset $I \subset I_m$ corresponding to linearly independent atoms with

$$\operatorname{span}(g_k, k \in I) = \operatorname{span}(g_k, k \in I_m).$$

B. Robustness Theorem

From Theorem 1, we know that if I_m satisfies the stability condition, then General MP performed on $f_m^\star$ is stable. The following theorem is a *robustness result* which shows that if f is "close enough" to $f_m^\star$, the atoms selected during "the first iterations" of a Pursuit will coincide with those which would be selected by a Pursuit on $f_m^\star$, which can be considered as the correct ones.

Theorem 3: Let $\{r_n\}_{n \geq 0}$ be a sequence of residuals computed by General MP to approximate some $f \in \mathcal{H}$. For any integer m such that $\mu_1(m-1) + \mu_1(m) < 1$, let $f_m^\star = \sum_{k \in I_m} c_k g_k$ be a best m-term approximation to f, and let $N_m = N_m(f)$ be the smallest integer such that

$$\|r_{N_m}\|^2 \leq \sigma_m^2 \cdot \left(1 + \frac{m \cdot (1 - \mu_1(m-1))}{[1 - \mu_1(m-1) - \mu_1(m)]^2} \right). \qquad (15)$$

Then, for $1 \leq n \leq N_m$, General MP picks up a correct atom, i.e., $k_n \in I_m$. If no best m-term approximant exists, the same results are valid provided that σ_m are replaced with $\|f - f_m^\star\| = (1 + \epsilon)\sigma_m$ in (15).

An analogue to this theorem was originally proved by Tropp [28] in the case of OMP, and the reader can easily prove it by following step by step Tropp's proof.

C. Comments on the Robustness Theorem

As already said, Theorem 3 was proved for OMP by Tropp [28], and we merely had to notice that it also works for MP. Our main contribution comes next with the detailed analysis of the approximation properties of MP through the consequences of the theorem. The statement of the robustness theorem relates several sequences of approximation errors. For a given f, both the sequence $\{\|r_n\|^2\}_{n \geq 1}$ of approximation errors with MP and that of best m-term approximation errors $\{\sigma_m^2\}_{m \geq 1}$ are decreasing, and the statement of the theorem shows that we should also consider a "penalized" version $\rho_m^2 := \sigma_m^2 \cdot (1 + \lambda_m)$ of the best m-term approximation error, using the penalty factor

$$\lambda_m := \frac{m \cdot (1 - \mu_1(m-1))}{[1 - \mu_1(m-1) - \mu_1(m)]^2}. \qquad (16)$$

The penalized error sequence is defined for any $m \geq 1$ such that $\mu_1(m-1) + \mu_1(m) < 1$, but it is no longer decreasing, since it blows up when $\mu_1(m-1) + \mu_1(m)$ approaches the value 1.

The theorem tells us that correct atoms (i.e., atoms that belong to the best m-term approximant) are picked up by MP until a good enough approximation error $\|r_n\|^2$ is achieved (compared to the penalized error ρ_m^2). The number of provably correct steps is at least

$$N_m = \min \left\{ n \mid \|r_n\|^2 \leq \rho_m^2 \right\}. \qquad (17)$$

For OMP, since each atom can be picked up at most once, we must have $N_m \leq m$ and the theorem thus guarantees that $\|r_m\|^2 \leq \rho_m^2$, i.e., in (at most) m steps an error no worse than ρ_m^2 is reached (this is the result of Tropp). Next we want to extend this result to MP by estimating how many steps of MP are sufficient to reach an error no worse than ρ_m^2: our goal is thus to obtain upper bounds on N_m.

D. Rate of Convergence

Let us start with some obvious remarks. For $m = 1$, it is not difficult to check that $\lambda_1 = 1/(1-\mu)^2 > 1$ and $\|r_1\|^2 = \sigma_1^2 \leq \rho_1^2$ hence, $N_1 \leq 1$. For $m \geq 2$ such that $\mu_1(m-1) + \mu_1(m) < 1$, the expresssion of an upper bound on N_m must depend on the value of σ_m. Indeed, when $\sigma_m = 0$, $f = f_m^\star$ has an exact m-term expansion and the analysis of the previous section shows that MP can loop forever within the set of correct atoms: the error $\|r_n\|$ decreases exponentially but never reaches zero, hence, $N_m = \infty$. On the contrary, as soon as σ_m is nonzero, the decrease to zero of the residual guarantees that $N_m < \infty$. Given these observations, it seems only natural that (for a given m) the smaller σ_m, the larger the bound on N_m. The bound expressed in the following theorem displays this behavior.

Theorem 4: Let $\{r_n\}_{n \geq 0}$ be a sequence of residuals computed by MP to approximate some $f \in \mathcal{H}$. For any integer m such that $\mu_1(m-1) + \mu_1(m) < 1,$, let $f_m^\star$ and $N_m = N_m(f)$ be defined as in Theorem 3. We have $N_1 \leq 1$, and for $m \geq 2$:

- if $\sigma_m^2 \leq 3\sigma_1^2/m$ then

$$2 \leq N_m \leq 2 + \frac{m}{1 - \mu_1(m-1)} \cdot \ln \frac{3\sigma_1^2}{m\sigma_m^2}; \qquad (18)$$

- else, $N_m \leq 1$.

We need some technical results before we proceed to the proof of the theorem. The first results concern the rate of decrease of the error $\{\|r_n\|\}$ for $1 \leq n \leq N_m$: the faster the decrease, the smaller the number of steps needed to reach the condition $\|r_n\| \leq \rho_m$.

Lemma 3: With the notations of Theorem 3, if $\{r_n\}$ is a sequence of residuals produced by MP, we have for $1 \leq n \leq N_m$

$$\|r_n\|^2 - \sigma_m^2 \leq \min_{0 \leq l \leq n} (\beta_m)^{n-l} \left(\|r_l\|^2 - \sigma_m^2 \right) \qquad (19)$$

with β_m as defined in (13).

Proof: We know from Theorem 3 that $k_{n+1} \in I_m$ for $0 \leq n < N_m$, hence we have

$$\|r_n\|^2 - \|r_{n+1}\|^2 = |\langle r_n, g_{k_{n+1}} \rangle|^2 = \sup_{k \in I_m} |\langle r_n, g_k \rangle|^2$$
$$= \sup_{k \in I_m} |\langle f_m^\star - f_n, g_k \rangle|^2$$
$$\geq (1 - \beta_m) \cdot \|f_m^\star - f_n\|^2$$

where the second line follows from the fact that $f - f_m^\star$ is orthogonal to $\mathcal{V}_{I_m}$ and the last inequality is, again, a consequence of Lemmas 1 and 2. Observing that $\|r_n\|^2 = \sigma_m^2 + \|f_m^\star - f_n\|^2$, we have

$$\|r_n\|^2 - \|r_{n+1}\|^2 = \|f_m^\star - f_n\|^2 - \|f_m^\star - f_{n+1}\|^2$$

and we obtain

$$\|f_m^\star - f_{n+1}\|^2 \leq \beta_m \cdot \|f_m^\star - f_n\|^2.$$

It follows that for $0 \leq l \leq n + 1$

$$\|f_m^\star - f_{n+1}\|^2 \leq (\beta_m)^{n+1-l} \cdot \|f_m^\star - f_l\|^2.$$

We can now conclude that

$$\|r_{n+1}\|^2 \leq \sigma_m^2 + (\beta_m)^{n+1-l} \cdot \left(\|r_l\|^2 - \sigma_m^2 \right)$$

which gives for $1 \leq n + 1 \leq N_m$ and $0 \leq l \leq n + 1$

$$\|r_{n+1}\|^2 - \sigma_m^2 \leq (\beta_m)^{n+1-l} \cdot \left(\|r_l\|^2 - \sigma_m^2 \right). \qquad \square$$

As a consequence of Lemma 3, we have the following relation between the numbers N_l.

Lemma 4: With the assumptions and notations of Theorem 3, for any $1 \leq k < m$ such that $N_k < N_m$ we have

$$N_m - N_k \leq 1 + \frac{m}{1 - \mu_1(m-1)} \cdot \left(\ln \frac{\sigma_k^2}{\sigma_m^2} + \ln \frac{1 + \lambda_k}{\lambda_m} \right). \qquad (20)$$

Proof: We let $l = N_k$ and $n = N_m - 1$ in (19) and use the very definition of N_m and N_k (cf. (17)) to obtain

$$\lambda_m \cdot \sigma_m^2 < \|r_{N_m-1}\|^2 - \sigma_m^2$$
$$\leq (\beta_m)^{N_m-1-N_k} \cdot \left(\|r_{N_k}\|^2 - \sigma_m^2 \right)$$
$$\leq (\beta_m)^{N_m-1-N_k} \cdot (1 + \lambda_k) \cdot \sigma_k^2.$$

It follows that

$$(1/\beta_m)^{N_m - 1 - N_k} \leq (\sigma_k^2/\sigma_m^2) \cdot (1 + \lambda_k)/\lambda_m$$

hence, we have $N_m - N_k \leq 1 + \Delta$ with

$$\Delta := \frac{1}{\ln \frac{1}{\beta_m}} \left(\ln \frac{\sigma_k^2}{\sigma_m^2} + \ln \frac{1 + \lambda_k}{\lambda_m} \right).$$

For $t \geq 0$, we have $\ln(1 - t) \leq -t$, hence, $1/\ln(1/(1 - t)) \leq 1/t$. Since $\beta_m = 1 - (1 - \mu_1(m - 1))/m$, it follows that

$$\frac{1}{\ln \frac{1}{\beta_m}} \leq \frac{m}{1 - \mu_1(m - 1)}$$

and we obtain (20) by combining the previous estimates. □

Theorem 4 will follow from Lemma 4 using the estimate of $(1 + \lambda_k)/\lambda_m$ provided by the following lemma.

Lemma 5: For all m such that $\mu_1(m - 1) + \mu_1(m) < 1$ and $1 \leq k < m$, we have

$$\lambda_m \geq m \tag{21}$$
$$\frac{\lambda_k}{\lambda_m} \leq \frac{k}{m} \cdot \frac{1 - \mu_1(k - 1)}{1 - \mu_1(m - 1)}. \tag{22}$$

Proof: For the first inequality, we write

$$\lambda_m = m \cdot \frac{1}{1 - \mu_1(m - 1) - \mu_1(m)} \cdot \frac{1 - \mu_1(m - 1)}{1 - \mu_1(m - 1) - \mu_1(m)}$$

and observe that the two rightmost factors are no less than 1. For the second inequality, consider $2 \leq l \leq m$: since $\mu_1(l - 2) + \mu_1(l - 1) \leq \mu_1(l - 1) + \mu_1(l)$, it is not difficult to check that

$$\frac{\lambda_{l-1}}{\lambda_l} \leq \frac{l - 1}{l} \cdot \frac{1 - \mu_1(l - 2)}{1 - \mu_1(l - 1)}.$$

Taking the product for $k + 1 \leq l \leq m$ we obtain the result. □

Proof of Theorem 4: From Lemma 5, we have $(1 + \lambda_k)/\lambda_m = 1/\lambda_m + \lambda_k/\lambda_m \leq (1 + k/(1 - \mu_1(m - 1)))/m$. Moreover, since $2\mu_1(m - 1) \leq \mu_1(m - 1) + \mu_1(m) < 1$, we have $1 - \mu_1(m - 1) > 1/2$, hence,

$$\ln \frac{1 + \lambda_k}{\lambda_m} \leq \ln \left(1 + \frac{k}{1 - \mu_1(m - 1)} \right) - \ln m$$
$$\leq \ln(2k + 1) - \ln m.$$

For $1 \leq k < m$, either $N_m \leq N_k$ or we can apply Lemma 4 and obtain

$$N_m \leq N_k + 1 + \frac{m}{1 - \mu_1(m - 1)} \left(\ln \frac{\sigma_k^2}{\sigma_m^2} + \ln \frac{2k + 1}{m} \right).$$

Taking $k = 1$ yields either $N_m \leq N_1 = 1$ (which is the second case of the theorem) or $N_m \geq N_1 + 1 = 2$ and

$$N_m \leq 2 + \frac{m}{1 - \mu_1(m - 1)} \cdot \ln \frac{3 \sigma_1^2}{m \sigma_m^2}$$

which is only possible if $3\sigma_1^2 > m\sigma_m^2$. □

Let us conclude by showing how Theorem 4 can be used to obtain Featured Theorem 3. If $m < (1 + 1/\mu)/4$, it is easy to check

that the condition $\mu_1(m - 1) + \mu_1(m) < 1$ is satisfied and that $m \leq \lambda_m \leq 4m$. As a result, $\rho_m^2 \leq (1 + 4m)\sigma_m^2$ and N_m defined in Theorem 3 satisfies the first and second statements of Featured Theorem 3. The third statement is checked using Theorem 4 and the easy fact that $1/(1 - \mu_1(m - 1)) < 4/3$.

V. DISCUSSION AND CONNECTIONS WITH OTHER RESULTS

Before concluding this correspondence, we would like to make it explicit how the work presented here extends the results of Villemoes and Temlyakov about MP in special dictionaries, as well as some other recent parallel work in incoherent dictionaries.

A. Connection With Villemoes' Result With Walsh Wavelet Packets

Without going into too much details, let us recall the definition of the Walsh wavelet packet dictionary (the material below is essentially taken from [30]). The dictionary, which is the collection of atoms

$$g_p(x) = g_{j,k,n}(x) := 2^{j/2} W_n(2^j x - k)$$

in $\mathcal{H} = L^2(\mathbb{R})$ obtained by dilations and translations of the Walsh system $\{W_n\}_{n \geq 0}$ on $L^2[0, 1]$, has coherence $\mu = \mu_1(1) = 1/\sqrt{2}$. In this dictionary, one can find four atoms $g_{p_d}, g_{p_u}, g_{p_l}, g_{p_r}$ with g_{p_l} orthogonal to g_{p_r} which satisfy

$$\begin{pmatrix} g_{p_d} \\ g_{p_u} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} g_{p_l} \\ g_{p_r} \end{pmatrix}. \tag{23}$$

As a result, we have

$$\mu_1(2) \geq |\langle g_{p_d}, g_{p_l} \rangle| + \langle |g_{p_d}, g_{p_r} \rangle| = \sqrt{2} > 1.$$

Hence, the hypothesis ((10)) of Theorem 2 is only valid for $m = 1$, i.e., only 1-term expansions from the Walsh wavelet packet dictionary can be stably recovered through MP! One should not be mislead by the meaning of such a "poor" result: it essentially means that one should be very careful about the relevance of the notion of a correct atom when such a notion is ambiguously defined. Let us consider a simple example. Assume we want to recover expansions from $\mathrm{span}(g_k, k \in I)$ with $I = \{p_l, p_r\}$. Since $g_{p_d} \in \mathrm{span}(g_{p_l}, g_{p_r})$, if MP is performed on $f := g_{p_d}$ it will pick up g_{p_d} as its first atom, and this is a "wrong" choice since only a choice $g_k, k \in I$ is considered a "good" one according to the terminology used in this correspondence. The fact that MP on a 2-term expansion can pick up a "wrong" atom is thus correctly predicted by the analysis, but the question is rather the relevance of the notion of "correct" versus "wrong" atom which is intrinsically ambiguous in this example.

In Villemoes' result, there is no statement about recovery of "good" versus "wrong" atoms, instead, the main point is the exponential convergence which comes from the stability of the pursuit in some subspace of dimension at most four: for every I of size two, there exists a reasonably small set $J \supset I$ which satisfies the stability condition $\eta(J) < 1$, hence, all the residuals remain in the finite-dimensional subspace $\mathrm{span}(g_p, p \in J)$. The latter is of finite dimension at most four hence the convergence is exponential. The rate of convergence is computed on a case by case basis.

B. Connection With Temlyakov's Result for Characteristic Functions

Temlyakov has long been studying the behavior of MP (which he calls Pure Greedy Algorithm) and its variants in several settings. In

particular, in the paper [26]—which was kindly pointed out to us by an anonymous referee—Temlyakov introduces several classes of "structured" dictionaries (λ-quasi-orthogonal, (μ, γ)-semistable) for which he is able to compare the rate of convergence of MP with that of the ideal but NP-hard best m-term approximation. The nature of the main results in [26] are far beyond the more restricted scope of the present work, but Temlyakov also considers some illustrative examples that have closer connections to the present analysis.

As already mentioned in the Introduction, if $\mathcal{D}$ is the collection of all (normalized) characteristic functions of intervals $I \subset [0, 1)$, and f is a function on the interval $[0, 1)$ taking constant values on a partition of $[0, 1)$ into n disjoint intervals, Temlyakov proved that

$$\|f_m - f\|_2 \leq (1 - 1/n)^{m/2} \|f\|_2$$

without any claim about MP recovering "good" pieces (the intervals on which f is constant), just about a rate of convergence. Temlyakov's example is just another example of the possible limitations of an analysis purely based on the (cumulative) coherence of the dictionary: in this case, it is not difficult to check that $\mu = \mu_1(1) = 1$ by considering two intervals arbitrarily close one to another, and $\mu_1(m) \geq m^{1/2}$ for $m \geq 2$ by letting $g_l = m^{1/2} \cdot \chi_{[l/m, (l+1)/m)}, 0 \leq l < m$ and $g_k = \chi_{[0,1)}$ in (7).

C. Some Other Related Results

After this work was submitted for publication, it came to our attention that Donoho, Elad, and Temlyakov had also been pursuing a parallel track studying the recovery of approximate sparse expansions with MP [5] and Basis Pursuit (other tracks being parallely followed by Tropp [29], Fuchs [11], and the authors of the present contribution [17], [18]). Altough of a broader nature than our results by the variety of algorithms considered, the contribution of [5] is performed in quite a more restricted setting since the structure of the dictionary is merely summarized in its coherence μ. By comparison, we have expressed all our results in terms of cumulative coherence, which can make a substantial difference [28].

VI. CONCLUSION

Nonlinear sparse approximations in redundant dictionaries opened brand new perspectives in data processing, mostly thanks to the freedom in designing atoms that match particular structures. Nevertheless, until recently, these methods suffered from a lack of constructive results regarding the approximation properties and stability of the associated decomposition algorithms. This correspondence provides insights that one of the most widely used heuristics, the Matching Pursuit algorithm, is stable, and offers good approximation properties when the dictionary is sufficiently incoherent. Extending these results to wider (and more useful) classes of dictionaries is a fundamental problem that we hope to address in forthcoming papers.

REFERENCES

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

[2] O. Christensen, *An Introduction to Frames and Riesz Bases.* Boston, MA: Birkhauser, 2003.

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

[4] R. A. DeVore and V. N. Temlyakov, "Some remarks on greedy algorithms," *Adv. Comput. Math.*, vol. 5, no. 2–3, pp. 173–187, 1996.

[5] D. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," Univ. South Carolina, Industrial Mathematics Institute, Columbia, Tech. Rep. 04:06, 2004.

[6] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization," in *Proc. Nat. Acad. Sci.*, vol. 100, Mar. 2003, pp. 2197–2202.

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

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

[9] A. Feuer and A. Nemirovsky, "On sparse representations in pairs of bases," *IEEE Trans. Inf. Theory*, vol. 49, no. 6, pp. 1579–1581, Jun. 2003.

[10] J. H. Friedman and W. Stuetzle, "Projection pursuit regression," *J. Amer. Statist. Assoc.*, vol. 76, pp. 817–823, 1981.

[11] J.-J. Fuchs, "Recovery of exact sparse representations in the presence of noise," in *Proc. 2004 IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP'04)*, vol. 2, Montreal, QC, Canada, May 2005, pp. 533–536.

[12] A. C. Gilbert, S. Muthukrishnan, and M. J. Strauss, "Approximation of functions over redundant dictionaries using coherence," in *Proc. 14th ACM-SIAM Symp. Discrete Algorithms (SODA'03)*, Baltimore, MD, Jan. 2003, pp. 243–252.

[13] A. C. Gilbert, S. Muthukrishnan, M. J. Strauss, and J. Tropp, "Improved sparse approximation over quasi-incoherent dictionaries," in *Proc. Int. Conf. Image Processing (ICIP'03)*, Barcelona, Catalonia, Spain, Sep. 2003, pp. 37–40.

[14] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd ed. Baltimore, MD: John Hopkins Univ. Press, 1989.

[15] R. Gribonval and E. Bacry, "Harmonic decomposition of audio signals with matching pursuit," *IEEE Trans. Signal Processing*, vol. 51, no. 1, pp. 101–111, Jan. 2003.

[16] R. Gribonval, E. Bacry, S. Mallat, P. Depalle, and X. Rodet, "Analysis of sound signals with high resolution matching pursuit," in *Proc. IEEE Conf. Time-Frequency and Time-Scale Analises (TFTS'96)*, Paris, France, Jun. 1996, pp. 125–128.

[17] R. Gribonval, R. M. Figueras i Ventura, and P. Vandergheynst, "A Simple Test to Check the Optimality of Sparse Signal Approximations," IRISA, Rennes, France, Tech. Rep., Nov. 2004. To appear in *EURASIP Signal Processing J.*

[18] R. Gribonval and M. Nielsen, "Beyond sparsity: Recovering structured representations by l1-minimization and greedy algorithms.-application to the analysis of sparse underdetermined ICA," *Adv. Comput. Math.*, to be published.

[19] ——, "Highly Sparse Representations from Dictionaries Are Unique and Independent of the Sparseness Measure," Dept. Math. Sci. , Aalborg Univ., Aalborg, Denmark, Tech. Rep. R-2003-16, Oct. 2003.

[20] ——, "Sparse decompositions in unions of bases," *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.

[21] R. A. Horn and C. R. Johnson, *Matrix Analysis.* Cambridge, U.K.: Cambridge Univ. Press, 1985.

[22] P. J. Huber, "Projection pursuit," *Ann. Statist.*, vol. 13, no. 2, pp. 435–475, 1985.

[23] S. Jaggi, W. C. Carl, S. Mallat, and A. S. Willsky, "High resolution pursuit for feature extraction," *Appl. Comput. Harmon. Anal.*, vol. 5, no. 4, pp. 428–449, Oct. 1998.

[24] L. K. Jones, "On a conjecture of Huber concerning the convergence of PP-regression," *Ann. Statist.*, vol. 15, pp. 880–882, 1987.

[25] S. Mallat and Z. Zhang, "Matching pursuit with time-frequency dictionaries," in *IEEE Trans. Signal Processing*, vol. 41, Dec. 1993, pp. 3397–3415.

[26] V. N. Temlyakov, "Greedy algorithms and m-term approximation with regard to redundant dictionaries," *J. Approx. Theory*, vol. 98, pp. 117–145, 1999.

[27] ——, "Weak greedy algorithms," *Adv. Comput. Math.*, vol. 12, no. 2, 3, pp. 213–227, 2000.

[28] J. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[29] ——, "Just Relax: Convex Programming Methods for Subset Selection and Sparse Approximation," Univ. Texas at Austin, Austin, TX, Tech. Rep. ICES Rep. 04-04, Feb. 2004.

[30] L. Villemoes, "Nonlinear approximation with walsh atoms," in *Proc. "Surface Fitting and Multiresolution Methods" Chamonix 1996*, A. Le Méhauté, C. Rabut, and L. L. Schumaker, Eds. Nashville, TN: Vanderbilt Univ. Press, 1997, pp. 329–336.