BEHAVIOR OF GREEDY SPARSE REPRESENTATION ALGORITHMS ON NESTED SUPPORTS

Boris Mailhé[†] *Bob Sturm*[⋆] *Mark D. Plumbley*[†]

[†]Centre for Digital Music, Queen Mary University of London
[⋆] Audio Analysis Lab, Aalborg University Copenhagen, Denmark

ABSTRACT

In this work, we study the links between the recovery properties of sparse signals for Orthogonal Matching Pursuit (OMP) and the whole General MP class over nested supports. We show that the optimality of those algorithms is not locally nested: there is a dictionary and supports I and J with J included in I such that OMP will recover all signals of support I, but not all signals of support J. We also show that the optimality of OMP is globally nested: if OMP can recover all s-sparse signals, then it can recover all s'-sparse signals with s' smaller than s. We also provide a tighter version of Donoho and Elad's spark theorem, which allows us to complete Tropp's proof that sparse representation algorithms can only be optimal for all s-sparse signals if s is strictly lower than half the spark of the dictionary.

***Index Terms*—** Sparsity, Compressed sensing, Basis Pursuit, Orthogonal Matching Pursuit, Performance analysis and bounds

1. INTRODUCTION

In the method of sparse representations, one tries to represent a signal as a linear combination of only a few vectors called *atoms* selected from a set called a *dictionary* [1]. That requires the dictionary to be well adapted to the signal to represent. When the signal is too complex to be represented on any orthonormal basis, one has to use an *overcomplete* dictionary that contains more atoms than the dimension of the signal. Representations are then not unique anymore, and finding the sparsest one becomes an NP-hard problem [2]. One way to solve the problem efficiently is to use greedy algorithms such as Orthogonal Matching Pursuit (OMP) [3]. Those algorithms are generally suboptimal, however a property called the Exact Recovery Condition (ERC) characterizes the sparse supports that are guaranteed to be successfully identified by OMP for a given dictionary [4].

This work was supported by the EPSRC Project EP/G007144/1 Machine Listening using Sparse Representations and by the EU FET-Open project FP7-ICT-225913-SMALL.

BLS is supported in part by Independent Postdoc Grant 11-105218 from Det Frie Forskningsråd; and EPSRC Platform Grant EP/E045235/1 at the Centre for Digital Music of Queen Mary University of London.

This paper addresses the link between the recovery properties of nested supports: if one support satisfies the ERC, do its sub-supports also satisfy it? If all supports of a given size satisfy the ERC, do all smaller supports also satisfy it? Our main result is that if OMP can recover all s_1-sparse signals, then it can recover all s_2-sparse signals for any $s_2 < s_1$, despite earlier hints of the opposite (Theorem 10 in [5]).

We first review the state of the art, then prove ancillary results on the link between the size of the support and the spark of the dictionary, and finally prove the main results.

2. STATE OF THE ART

Consider an overcomplete dictionary $\mathbf{\Phi} = (\boldsymbol{\varphi}_n)_{1\leq n\leq N}$ of N atoms $\boldsymbol{\varphi}_n \in \mathbb{C}^M$ with $N > M$. We will only work with normalized dictionaries: $\forall n \in [1, N],\ \|\boldsymbol{\varphi}_n\|_2 = 1$. We will also assume that the dictionary has rank M. For a given signal $\mathbf{y} \in \mathbb{C}^M$, the exact-sparse problem [4] is the one of finding the coefficients $\hat{\mathbf{x}}$ defined as

$$\hat{\mathbf{x}} \triangleq \underset{\mathbf{x}\in\mathbb{C}^N | \mathbf{y}=\mathbf{\Phi x}}{\operatorname{argmin}} \|\mathbf{x}\|_0 \tag{P0}$$

where $\|\mathbf{x}\|_0$ is the number of nonzero coefficients in $\mathbf{x}$: if $\Gamma \subseteq [1, N]$ is the support of $\mathbf{x}$, then $\|\mathbf{x}\|_0 \triangleq \operatorname{card}(\Gamma)$. The problem (P0) was proven to be NP-hard [2] but several practical algorithms have been proposed [6]. In this work we will focus on the General MP class as it is described by Gribonval and Vandergheynst [7].

2.1. Greedy algorithms: the General MP class

A greedy algorithm solves the problem (P0) by growing the support Γ_i of the solution at each iteration i, and removing the selected atoms from the signal to form a temporary *residual* that the algorithm decomposes in the next iterations. Initially, the support Γ_0 is empty and the residual $\mathbf{r}_0$ is the signal $\mathbf{y}$ itself. General MP is a class of greedy algorithms that share two characteristics. First, for an algorithm from the General MP class, only one atom of index n_i is added at iteration i and it is the atom with the highest correlation with the residual:

$$\Gamma_i \triangleq \Gamma_{i-1} \cup \underset{n\in[1,N]}{\operatorname{argmax}} \left|\langle \mathbf{r}_{i-1}, \boldsymbol{\varphi}_n\rangle\right| . \tag{1}$$

The second requirement for a General MP algorithm is that the i^{th} approximation of the signal is in the span of the subdictionary $\mathbf{\Phi}_{\Gamma_i}$ containing atoms indexed by Γ_i only:

$$\mathbf{y} - \mathbf{r}_i \in \operatorname{span}(\mathbf{\Phi}_{\Gamma_i}) \ . \tag{2}$$

Several stopping criteria can be used, but in all cases the algorithm must stop if the residual becomes 0. The different General MP algorithms such as Matching Pursuit (MP) [8], Gradient Pursuit (GP) [9], OMP [3], their local variants LocGP and LocOMP [10], only differ in how they compute the new residual. For OMP, the update of the residual is performed by projecting the signal orthogonally to all previously selected atoms:

$$\mathbf{r}_i \triangleq \mathbf{y} - \mathbf{\Phi}_{\Gamma_i}^{\dagger}\mathbf{y} \tag{3}$$

where $\mathbf{\Phi}_{\Gamma_i}^{\dagger}$ is the pseudoinverse of the subdictionary $\mathbf{\Phi}_{\Gamma_i}$. Within the General MP class, OMP has the unique property that the residual is always orthogonal to all previously selected atoms. As a consequence, OMP can never select the same atom twice.

2.2. Exact recovery conditions

The ERC was first derived by Tropp for OMP [4]. In that case it guarantees that OMP will find the best support and how many iterations it will take.

Theorem 1 (Exact Recovery Condition for OMP [4])**.** *For any normalized dictionary* $\mathbf{\Phi}$ *and support* Γ*, if*

$$\max_{n \notin \Gamma} \left\| \mathbf{\Phi}_{\Gamma}^{\dagger} \boldsymbol{\varphi}_n \right\|_1 < 1 \tag{ERC}$$

then for any sparse signal of sparsest support Γ*, OMP will recover* Γ *in* $\operatorname{card}(\Gamma)$ *iterations.*

See [11] for the case where the dictionary is not normalized. Gribonval and Vandergheynst later pointed that the ERC also entails a weaker stability result for the whole General MP class.

Theorem 2 (Exact Recovery Condition for General MP [7])**.** *For any normalized dictionary* $\mathbf{\Phi}$ *and support* Γ*, if the ERC holds, then for any sparse signal with a support included in* Γ*, the support recovered by any General MP algorithm is also included in* Γ*.*

Theorem 1 can be derived from Theorem 2 and the OMP property to never select the same atom twice (it is done in Tropp's proof of Theorem 1 [4]).

Theorem 3 (ERC converse [4])**.** *For any dictionary* $\mathbf{\Phi}$ *and support* Γ *such that* $\mathbf{\Phi}_{\Gamma}$ *has full rank, if the ERC does not hold, then there exists a signal* $\mathbf{y} \in \operatorname{span}(\mathbf{\Phi}_{\Gamma})$ *such that the first atom selected by any General MP algorithm does not belong to* Γ*.*

The full rank hypothesis is implicit in Tropp's work because he only considers supports that are the sparsest ones for some signal, and those supports must have full rank. Interestingly, that hypothesis is not needed for the proof of Theorem 2, but to our knowledge, the proof of Theorem 3 requires the factorization of the pseudoinverse $\mathbf{\Phi}_{\Gamma}^{\dagger} = \left(\mathbf{\Phi}_{\Gamma}^T\mathbf{\Phi}_{\Gamma}\right)^{-1}\mathbf{\Phi}_{\Gamma}^T$ to hold.

3. UNIVERSAL RECOVERY AND SPARK

We now consider conditions for all sparse signals of a given support size to be recovered by OMP. If a support satisfies the ERC, then all sparse signals with that support are recovered by OMP. If all supports of size s satisfy the ERC, then all s-sparse signals are recovered by OMP, which implies that a signal can only have at most one s-sparse representation [4].

Besides Donoho and Elad introduced the notion of *spark*. $\operatorname{spark}(\mathbf{\Phi})$ is the support size of the sparsest vector in the null space of $\mathbf{\Phi}$. Donoho and Elad showed that for s such that $2s \geq \operatorname{spark}(\mathbf{\Phi})$, there is a signal that has two s_1 and s_2-sparse representations with $s_1 \leq s$ and $s_2 \leq s$ [12].

Tropp suggested that those two uniqueness results could be combined to conclude on how large s can be to allow universal recovery by OMP and proposed a proof sketch, but did not formalize it [4]. That might be because those two results do not work well together as they are presented: assuming that all supports of size s satisfy the ERC does not imply that the supports of size s_1 and s_2 also do. [1] We now formulate a stronger version of the spark theorem that will let us conclude the proof of the upper bound.

Theorem 4 (Spark and uniqueness of sparse representations)**.** *For any* s *such that*

$$\frac{\operatorname{spark}(\mathbf{\Phi})}{2} \leq s \leq M \tag{4}$$

there is a signal with two s*-sparse representations on two different full-rank supports.*

Proof. We first prove the case where $s < \operatorname{spark}(\mathbf{\Phi})$. By definition of the spark, there is a coefficient vector $\mathbf{x}$ such that $\|\mathbf{x}\|_0 = \operatorname{spark}(\mathbf{\Phi})$ and $\mathbf{\Phi}\mathbf{x} = \mathbf{0}$. Let $\Gamma \triangleq \operatorname{supp}(\mathbf{x})$. Let $\Gamma_1, \Gamma_2 \subset \Gamma$ be any exact covering of Γ (i.e. $\Gamma_1 \cup \Gamma_2 = \Gamma$) with $\operatorname{card}(\Gamma_1) = \operatorname{card}(\Gamma_2) = s$. Since $s < \operatorname{card}(\Gamma)$, Γ_1 and Γ_2 must be different to cover Γ. We can now define the two vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ by their individual coefficients $[\mathbf{x}_1]_m$ and

[1] We will see later in Theorem 8 that this implication actually holds, but the proof of it will require the upper bound on s that we are currently trying to prove, therefore it cannot be used yet.

$[\mathbf{x}_1]_m$:

$$[\mathbf{x}_1]_m \triangleq \begin{cases} [\mathbf{x}]_m, & m \in \Gamma_1 \backslash \Gamma_2 \\ [\mathbf{x}]_m/2, & m \in \Gamma_1 \cap \Gamma_2 \\ 0, & \text{else} \end{cases} \tag{5}$$

$$[\mathbf{x}_2]_m \triangleq \begin{cases} -[\mathbf{x}]_m, & m \in \Gamma_2 \backslash \Gamma_1 \\ -[\mathbf{x}]_m/2, & m \in \Gamma_1 \cap \Gamma_2 \\ 0, & \text{else.} \end{cases} \tag{6}$$

We have

$$\mathbf{x}_1 - \mathbf{x}_2 = \mathbf{x} \tag{7}$$

$$\mathbf{\Phi}(\mathbf{x}_1 - \mathbf{x}_2) = 0 \tag{8}$$

$$\mathbf{\Phi}\mathbf{x}_1 = \mathbf{\Phi}\mathbf{x}_2. \tag{9}$$

Then $\mathbf{y} \triangleq \mathbf{\Phi}\mathbf{x}_1$ has two sparse representations $\mathbf{x}_1$ and $\mathbf{x}_2$ on two different supports Γ_1 and Γ_2. Moreover the supports Γ_1 and Γ_2 are smaller than $\mathrm{spark}(\mathbf{\Phi})$, so both $\mathbf{\Phi}_{\Gamma_1}$ and $\mathbf{\Phi}_{\Gamma_2}$ have full rank.

If $\mathrm{spark}(\mathbf{\Phi}) \leq s \leq M$, then one can start with the same proof but the obtained supports are too small and need to be completed to reach the size s. Let Γ_1 and Γ_2 be two sets of size $\mathrm{spark}(\mathbf{\Phi})$-1 that exactly cover Γ. Let m_1 and m_2 be the only indices respectively in $\Gamma_2 \backslash \Gamma_1$ and $\Gamma_1 \backslash \Gamma_2$. Then $\boldsymbol{\varphi}_{m_1}$ is in the span of $\mathbf{\Phi}_{\Gamma_1}$ and $\boldsymbol{\varphi}_{m_2}$ is in the span of $\mathbf{\Phi}_{\Gamma_2}$. So $\mathbf{\Phi}_{\Gamma_1}$, $\mathbf{\Phi}_{\Gamma_2}$ and $\mathbf{\Phi}_{\Gamma}$ all span the same subspace of dimension $\mathrm{spark}(\mathbf{\Phi})$-1 and $\mathbf{\Phi} \backslash \{\boldsymbol{\varphi}_{m_1}\}$ has the same rank M as $\mathbf{\Phi}$. So for any $s \leq \mathrm{rank}(\mathbf{\Phi})$, which we assumed to be M, one can find a completion support $J \subset [1, N] \backslash \Gamma$ of size $s - \mathrm{spark}(\mathbf{\Phi}) + 1$ such that $\Gamma_1 \cup J$ has full rank s. Then $\Gamma_2 \cup J$ also has full rank s. Let $\mathbf{x}_1$ and $\mathbf{x}_2$ be defined as in Equations (5) and (6) and let $\mathbf{y} \triangleq \mathbf{\Phi}\mathbf{x}_1 = \mathbf{\Phi}\mathbf{x}_2$. Let $\mathbf{x}'$ be a vector of support J and let $\mathbf{y}' = \mathbf{y} + \mathbf{\Phi}\mathbf{x}' = \mathbf{\Phi}(\mathbf{x}_1 + \mathbf{x}') = \mathbf{\Phi}(\mathbf{x}_2 + \mathbf{x}')$. $\mathbf{y}'$ has two different s-sparse decompositions with different full-rank supports $\Gamma_1 \cup J$ and $\Gamma_2 \cup J$. □

We can now complete Tropp's proof of how large s can be until signals that cannot be recovered appear.

Theorem 5 (Spark and the ERC)**.** *If the ERC holds for all supports of a size $s \leq M$, then $s < \mathrm{spark}(\mathbf{\Phi})/2$, and hence, every s-sparse representation is unique and the sparsest possible.*

Proof. We prove this by contradiction. If $s \geq \mathrm{spark}(\mathbf{\Phi})/2$, then Theorem 4 applies and there is a signal $\mathbf{y}$ with two different s-sparse decompositions with different full-rank supports I and J. Since I has size s it satisfies the ERC so OMP run on $\mathbf{y}$ will only ever select atoms that belong to I. Moreover, since I has full rank, the s-sparse decomposition of $\mathbf{y}$ on $\mathbf{\Phi}_I$ is the only one with support included in I so OMP cannot stop before running all s iterations. So all the atoms of I are selected by OMP, including those that do not belong to J. But J has size s so it also satisfies the ERC hence we have a contradiction. □

4. NESTEDNESS OF RECOVERY ALGORITHMS

Theorem 10 of [5] states such that the ERC holds for all supports of some size s but fails for some support of size $s' < s$. That would indicate that there are dictionaries such that OMP can recover all s-sparse signals and not sparser ones. However, the proof uses a dictionary with non-normalized atoms. In that case, the ERC as defined in Equation (ERC) is not linked to the recovery properties of OMP [11]. So although mathematically correct, that theorem does not help understanding the behavior of algorithms. That is why we want to investigate the question.

There are three properties we want to consider: the ERC, the optimality of OMP and the stability of General MP. OMP is optimal for a support Γ and a dictionary $\mathbf{\Phi}$ if it can recover any sparse signal on $\mathbf{\Phi}$ with the support Γ. A General MP algorithm is stable for a support Γ if for any signal with the support Γ, the algorithm only select atoms that belong to $\mathbf{\Phi}_\Gamma$. We cannot talk about optimality for General MP because that class contains some notoriously bad residual updates such as doing nothing or adding the last atom instead of subtracting it [10], so one cannot guarantee that a General MP algorithm will recover the whole support.

4.1. Notions of nestedness

We first need to define two different notions of nestedness.

Definition 4.1 (Local Nestedness)**.** *For any $K \leq N$, let $\mathbb{E}_K$ be the set of all subsets of $[1, N]$ of cardinality K or less. A property $P(\Gamma)$ that applies to a support Γ is* locally nested *up to size K if*

$$\forall \Gamma \in \mathbb{E}_K, \ [P(\Gamma) \Rightarrow (\forall \Gamma' \subset \Gamma, \ P(\Gamma'))] \tag{10}$$

Definition 4.2 (Global Nestedness)**.** *For any $K \leq N$, let $\mathbb{F}_K$ be the set of all subsets of $[1, N]$ of cardinality K only. A property $P(\Gamma)$ that applies to a support Γ is* globally nested *up to size K if*

$$\forall K' \leq K, [(\forall \Gamma \in \mathbb{F}_{K'}, \ P(\Gamma)) \Rightarrow (\forall \Gamma' \in \mathbb{E}_{K'}, \ P(\Gamma'))] \tag{11}$$

A property is locally nested if when it is true for a set, it is true for all subsets. A property is globally nested if when it is true for all sets of a certain size, it is true for all smaller sets. Global nestedness is a weaker meta-property than local nestedness because its hypotheses are more restrictive. While local nestedness implies global nestedness, the opposite is false in general. The restriction to size K is there because the sparsest support is always smaller than the rank of the dictionary, so only those sets should be considered. In particular, for any dictionary, the full support $[1, N]$ is trivially stable for General MP, but that does not say anything about its subsets.

4.2. Results for OMP, General MP and the ERC

We now characterize the nestedness of the different considered properties.

Theorem 6 (Non local nestedness)**.** *There are dictionaries such that the optimality of OMP is not locally nested at size 3. Equivalently, there are dictionaries such that the ERC or the stability of General MP are not locally nested.*

Proof. We just need an example of a dictionary and two supports $\Gamma' \subset \Gamma$ such that Γ satisfies the ERC and Γ' does not. OMP is always optimal for 1-sparse signals so Γ' must have size 2 at least and Γ size 3. For some $\theta \in [-\pi/2, \pi/2]$, let

$$\mathbf{\Phi} = \begin{pmatrix} 1 & 0 & \frac{\cos\theta}{\sqrt{2}} & 0 \\ 0 & 1 & \frac{\cos\theta}{\sqrt{2}} & 0 \\ 0 & 0 & \sin\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad \mathbf{y} = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \tag{12}$$

and $\Gamma = \{1, 2, 3\}$. $\boldsymbol{\varphi}_4 \perp \operatorname{span}(\mathbf{\Phi}_\Gamma)$ so $\left\| \mathbf{\Phi}_\Gamma^\dagger \boldsymbol{\varphi}_4 \right\|_1 = 0$ and Γ satisfies the ERC. Besides we have $\langle \mathbf{y}, \boldsymbol{\varphi}_1 \rangle = \langle \mathbf{y}, \boldsymbol{\varphi}_2 \rangle = 1$, $\langle \mathbf{y}, \boldsymbol{\varphi}_4 \rangle = 0$ and $\langle \mathbf{y}, \boldsymbol{\varphi}_3 \rangle = \cos\theta\sqrt{2}$. So for any θ such that $0 < |\theta| < \frac{\pi}{4}$, OMP will select $\boldsymbol{\varphi}_3$ first, although $\boldsymbol{\varphi}_3$ does not belong to the sparsest support $\Gamma' = \{1, 2\} \subset \Gamma$ of $\mathbf{y}$, so Γ' does not satisfy the ERC. □

Theorem 7 (Global nestedness for OMP)**.** *For any dictionary $\mathbf{\Phi}$, the optimality of OMP is globally nested up to the size $s = \operatorname{rank}(\mathbf{\Phi})$.*

Proof. Consider a dictionary $\mathbf{\Phi}$. Assume that there is a cardinality s such that OMP will recover all s-sparse signals. Theorem 5 applies so $s < \operatorname{spark}(\mathbf{\Phi})/2$. So all s-size supports have full rank, Theorem 3 applies and they all satisfy the ERC.

Suppose that there is a support Γ' of size $s' < s$ for which OMP is not optimal. Then Γ' does not satisfy the ERC. Γ' has full rank so Theorem 3 applies: there is a signal $\mathbf{y} \in \operatorname{span}(\mathbf{\Phi}_{\Gamma'})$ and an atom $\boldsymbol{\varphi} \notin \mathbf{\Phi}_{\Gamma'}$ such that OMP will select $\boldsymbol{\varphi}$ at the first iteration.

Now let Γ be a superset of Γ' of size s that does not contain $\boldsymbol{\varphi}$. Such a set exists because $\mathbf{\Phi}$ contains at least as many atoms as its spark so

$$s < \operatorname{spark}(\mathbf{\Phi})/2 \leq \operatorname{spark}(\mathbf{\Phi}) \leq \operatorname{card}(\mathbf{\Phi})$$

and $s < \operatorname{card}(\mathbf{\Phi})$. Γ has size s so it satisfies the ERC. Therefore Theorem 2 applies: since $\mathbf{y} \in \operatorname{span}(\mathbf{\Phi}_\Gamma)$, OMP will only select atoms that belong to $\mathbf{\Phi}_\Gamma$. However at the first iteration, OMP selects $\boldsymbol{\varphi} \notin \mathbf{\Phi}_\Gamma$, thus resulting in a contradiction. □

One should note that although the proof is valid for all cardinalities up to $\operatorname{rank}(\mathbf{\Phi})$, Theorem 7 is only useful up to $\operatorname{spark}(\mathbf{\Phi})/2 - 1$. For s between $\operatorname{spark}(\mathbf{\Phi})/2$ and $\operatorname{rank}(\mathbf{\Phi})$, global nestedness is trivially satisified because OMP cannot recover all s-sparse signals (because of Theorem 5) so the hypothesis in Formula (11) is always false, therefore the implication is true. For $s > \operatorname{rank}(\mathbf{\Phi})$, global nestedness breaks: OMP is trivially optimal at those sparsity levels since there are no such s-sparse signals, but Theorem 5 provides a whole range of smaller $s' < s$ for which OMP is not optimal.

Since the ERC and the optimality of OMP are equivalent on full-rank supports, we also have the following immediate corollary.

Theorem 8 (Global nestedness for the ERC)**.** *For any dictionary $\mathbf{\Phi}$, the ERC is globally nested up to $s = \operatorname{spark}(\mathbf{\Phi}) - 1$.*

Proof. For any $s < \operatorname{spark}(\mathbf{\Phi})$, all supports of size s have full rank. If all supports of size s satisfy the ERC, then OMP is optimal for all of them. So Theorem 7 applies and OMP is also optimal for all smaller supports. Since those supports have full rank, Theorem 3 applies and they all satisy the ERC. □

If one had a proof of Theorem 3 that does not require a full rank support, then one could prove Theorem 8 up to $\operatorname{rank}(\mathbf{\Phi})$, but then again the really useful part is only up to $\operatorname{spark}(\mathbf{\Phi})/2 - 1$ and we already have it.

For the General MP class, we can only consider stability instead of optimality:

Theorem 9 (Global nestedness for General MP)**.** *For any dictionary and any General MP algorithm, the stability of the algorithm is globally nested up to $s = \operatorname{spark}(\mathbf{\Phi}) - 1$.*

Proof. With that constraint on s, all considered supports have full rank. In that case, the stability of General MP is equivalent to the ERC and a proof with the same structure as for Theorem 8 can be used. □

5. CONCLUSION

We have shown that the optimality of OMP is not locally nested but is globally nested. We have also extended those results to the whole General MP class. Those results shed a new light on pursuit algorithms and counter the intuition provided by earlier works. To complete our proofs, we also filled some gaps in the underlying theory and provided a formal proof that OMP can only be optimal for signals with strictly less than $\operatorname{spark}(\mathbf{\Phi})/2$ nonzero coefficients.

With our results, one can consider the ERC has a bound on the admissible error: if a superset of the support of a sparse signal satisfies the ERC, then it does not guarantee that OMP will recover the signal, but it still guarantees that it will not select too many wrong atoms. This work should also allow us to further investigate possible links between greedy algorithms and ℓ_1 minimization.

6. REFERENCES

[1] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, Academic Press, Elsevier, Amsterdam, 3rd edition, 2009.

[2] BK Natarajan, “Sparse approximate solutions to linear systems,” *SIAM Journal on Computing*, vol. 24, pp. 227–234, 1995.

[3] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, “Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition,” in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 1993, vol. 1, pp. 40–44.

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

[5] M. D. Plumbley, “On polar polytopes and the recovery of sparse representations,” *IEEE Trans. Info. Theory*, vol. 53, no. 9, pp. 3188–3195, Sep. 2007.

[6] J. A. Tropp and S. J. Wright, “Computational methods for sparse solution of linear inverse problems,” *Proc. IEEE*, vol. 98, no. 6, pp. 948–958, June 2010.

[7] R. Gribonval and P. Vandergheynst, “On the exponential convergence of matching pursuits in quasi-incoherent dictionaries,” *IEEE Trans. Information Theory*, vol. 52, no. 1, pp. 255–261, 2006.

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

[9] T. Blumensath and M. E. Davies, “Gradient pursuits,” *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2370–2382, June 2008.

[10] B. Mailhé, R. Gribonval, P. Vandergheynst, and F. Bimbot, “Fast orthogonal sparse approximation algorithms over local dictionaries,” *Signal Process. (accepted)*, 2011.

[11] B. L. Sturm, B. Mailhé, and M. D. Plumbley, “On ’theorem 10’ of ’on polar polytopes and the recovery of sparse representations’,” *IEEE Trans. Info. Theory (in review)*, 2012.

[12] D. L. Donoho and M. Elad, “Maximal sparsity representation via ℓ_1 minimization,” in *Proc. Natl. Acad. Sci.*, Mar. 2003, vol. 100, pp. 2197–2202.