On Theorem 10 in "On Polar Polytopes and the Recovery of Sparse Representations"

Bob L. Sturm, *Member, IEEE,*, Boris Mailhé, *Member, IEEE* and Mark D. Plumbley, *Member, IEEE*

Abstract—It is shown that Theorem 10 (Non-Nestedness of ERC) in [Plumbley, IEEE Trans. Info. Theory, vol. 53, pp. 3188, Sep. 2007] neglects the derivations of the exact recovery conditions (ERCs) of constrained ℓ_1-minimization (BP) and orthogonal matching pursuit (OMP). This means that it does not reflect the recovery properties of these algorithms. Furthermore, an ERC of BP more general than that in [Tropp, IEEE Trans. Info. Theory, vol. 50, pp. 2231, Oct. 2004] is shown.

Index Terms—Sparse representation, basis pursuit, orthogonal matching pursuit, compressed sensing

I. INTRODUCTION

Theorem 10 (Non-Nestedness of ERC) of Plumbley [1] claims that if the exact recovery condition (ERC) of Tropp [2] — derived for both ℓ_1-minimization (BP) [3] and orthogonal matching pursuit (OMP) [4] — is satisfied for all representations having s non-zeros in a dictionary, it is not necessarily satisfied for representations having $k < s$ non-zeros in the same dictionary. In this correspondence, we show that this claim does not reflect the recovery properties of BP or OMP. We first provide background and notation. Then we briefly review BP and OMP, and the ERCs derived by Tropp [2]. We extend these conditions to the case of non-unit-norm dictionaries, and find an ERC of BP more general than that of Tropp [2]. We finally review and discuss Theorem 10 of Plumbley [1].

A. Background

Consider a dictionary of N not necessarily unit-norm atoms $\mathcal{D} := \{\boldsymbol{\varphi}_i \in \mathbb{C}^m\}_{i \in \Omega}$, $\Omega := \{1, 2, \ldots, N\}$, which we express as the matrix $\boldsymbol{\Phi} = [\boldsymbol{\varphi}_1 | \boldsymbol{\varphi}_2 | \ldots | \boldsymbol{\varphi}_N]$. We say $\mathcal{D}$ is *overcomplete* when the cardinality of the index set $|\Omega| = N > m$, and $\operatorname{span}(\mathcal{D}) = \mathbb{C}^m$. Any *measurement* $\mathbf{u} = \boldsymbol{\Phi}\mathbf{x}$ is thus expressable exactly as a linear combination of at most m linearly independent atoms from $\mathcal{D}$. Given $\mathbf{u} = \boldsymbol{\Phi}_{I^*}\mathbf{y}^*$, where the columns of $\boldsymbol{\Phi}_{I^*}$ are the atoms indexed by $I^* \subset \Omega$, the

exact-sparse problem [2] entails recovering I^* from $\mathbf{u}$ by

$$\hat{I} := \arg \min_{I \subset \Omega, \mathbf{y} \in \mathbb{C}^{|I|}} |I| \text{ subject to } \mathbf{u} = \boldsymbol{\Phi}_I \mathbf{y}. \qquad \text{(P0)}$$

It is of interest to know when an algorithm paired with a dictionary guarantees a solution to (P0) for any $\mathbf{u}$. A condition that guarantees such behavior is called a *stability condition* when $\hat{I} \subseteq I^*$, or ERC when $\hat{I} = I^*$. In [2], Tropp derives ERCs of two algorithms solving (P0): BP [3] and OMP [4]. We see the same ERC applies to both algorithms. In later work, Grinbonval and Vandergheynst [5] show that the same ERC is a stability guarantee for the class of general matching pursuit algorithms, i.e., any algorithm with alternating steps of (weak) greedy atom selection, and an update of the solution in the span of the selected atoms.

B. Notation

The ith element of a vector is $[\mathbf{x}]_i$. The support of a vector $\operatorname{supp}(\mathbf{x})$, the set of locations at which $\mathbf{x}$ is non-zero. The ℓ_0-pseudonorm of $\mathbf{x}$ is $\|\mathbf{x}\|_0 := |\operatorname{supp}(\mathbf{x})|$. We call $\mathbf{x}$ s-sparse if $\|\mathbf{x}\|_0 = s$; and we say $\mathbf{u}$ has an s-sparse representation in $\mathcal{D}$ if $\mathbf{u} = \boldsymbol{\Phi}\mathbf{x}$ where $\mathbf{x}$ is s-sparse. For a subset $A \subset \Omega$, $\operatorname{rank}(A) := \operatorname{rank}(\boldsymbol{\Phi}_A)$, and $\operatorname{span}(A) := \operatorname{span}\{\boldsymbol{\varphi}_i : i \in A\}$. We consider only dictionaries $\mathcal{D}$ containing unique atoms, i.e., $\operatorname{span}(\{i\}) \cap \operatorname{span}(\{j\}) = \{\mathbf{0}\}$ for all $i, j \in \Omega, i \neq j$. We define the diagonal matrix $\mathbf{N}_A$ such that $[\mathbf{N}_A]_{ii} := \|\boldsymbol{\Phi}_A \mathbf{e}_i\|_2$, $i \in \{1, \ldots, |A|\}$, where $\mathbf{e}_i$ is the ith standard basis vector. When $A = \Omega$, we write simply $\mathbf{N} := \mathbf{N}_\Omega$. Define the subspace $\mathcal{V}^* := \operatorname{span}(I^*)$. We define $\Omega_F(\mathcal{V}^*) \subseteq \Omega$ as the indices of atoms in $\mathcal{D}$ that participate non-trivially in the feasible set of solutions $\mathcal{X}(\mathcal{V}^*, \mathcal{D}) := \{\mathbf{x} \in \mathbb{C}^N : \boldsymbol{\Phi}\mathbf{x} \in \mathcal{V}^*\}$. Formally, this set is defined $\Omega_F(\mathcal{V}^*) := \{i \in \Omega : \exists\mathbf{x}(\mathbf{x} \in \mathcal{X}(\mathcal{V}^*, \mathcal{D}) \wedge i \in \operatorname{supp}(\mathbf{x}))\}$. Every atom indexed by $\Omega_F(\mathcal{V}^*)$ thus has a non-zero weight in at least one element of $\mathcal{X}(\mathcal{V}^*, \mathcal{D})$.

II. BP, OMP AND ERCs

A. Unit-norm Dictionaries

Assume the columns of $\boldsymbol{\Phi}$, or the atoms in $\mathcal{D}$, have the same norm. Without loss of generality, we assume they all have unit-norm. BP [3] is the principle of replacing the non-convex cost in (P0) with the "nearest" convex cost — the ℓ_1-norm of the solution — i.e., posing (P0) as

$$\min_{\mathbf{x} \in \mathbb{C}^N} \|\mathbf{x}\|_1 \text{ subject to } \mathbf{u} = \boldsymbol{\Phi}\mathbf{x}. \qquad (\ell_1)$$

One ERC [2] states a sufficient condition for (ℓ_1) to solve (P0) for any $\mathbf{u} \in \mathcal{V}^*$, i.e., $\operatorname{supp}(\mathbf{x}) = I^*$, is

$$\max_{i \in \Omega \setminus I^*} \|\boldsymbol{\Phi}_{I^*}^\dagger \boldsymbol{\varphi}_i\|_1 < 1. \qquad \text{(ERC-T)}$$

OMP takes an iterative approach to solve (P0). Given $I_k \subset \Omega$, OMP augments this set by

$$I_{k+1} \leftarrow I_k \cup \arg\max_{n \in \Omega} |\langle \mathbf{u}^{\perp I_k}, \boldsymbol{\varphi}_n \rangle| \tag{1}$$

where $\mathbf{u}^{\perp I_k} := (\mathbf{I} - \boldsymbol{\Phi}_{I_k} \boldsymbol{\Phi}_{I_k}^\dagger)\mathbf{u}$. This selection criterion comes from the desire to minimize in a greedy way the residual error (before orthogonal projection), i.e.,

$$\arg\min_{n \in \Omega} \left[\min_{\alpha \in \mathbb{C}} \|\mathbf{u}^{\perp I_k} - \alpha \boldsymbol{\varphi}_n\|_2^2 \right] = \arg\max_{n \in \Omega} |\langle \mathbf{u}^{\perp I_k}, \boldsymbol{\varphi}_n \rangle|^2. \tag{2}$$

For initialization, $I_0 := \emptyset$ and $\mathbf{u}^{\perp I_0} = \mathbf{u}$; and usually, OMP is made to stop once it has found a certain number of atoms, or $\|\mathbf{u}^{\perp I_k}\|_2 / \|\mathbf{u}\|_2 \le \delta$. An ERC of OMP [2] states that (ERC-T) is a sufficient condition for OMP with $\mathcal{D}$ to solve (P0) for any $\mathbf{u} \in \mathcal{V}^*$, i.e., $I_k \subseteq I^*$.

B. Dictionaries having atoms of different lengths

The formulation of (ℓ_1) implicitly assumes all atoms in $\mathcal{D}$ have the same norm. When they do not, then (ℓ_1) does not mimic (P0). Consider a dictionary of three non-identical atoms in $\mathbb{C}^2$, and let $\mathbf{x}$ lie in the span of $\boldsymbol{\varphi}_3$. We can shrink the length of $\boldsymbol{\varphi}_3$ so that (ℓ_1) produces a solution using the other two atoms simply because the weight assigned to $\boldsymbol{\varphi}_3$ in the constraints becomes larger than the sum of the other two. Thus, to apply the principle of BP to (P0) and take into account that the dictionary has atoms of any length, we must pose the problem instead [6]

$$\min_{\mathbf{x} \in \mathbb{C}^N} \|\mathbf{N}\mathbf{x}\|_1 \text{ subject to } \mathbf{u} = \boldsymbol{\Phi}\mathbf{x}. \tag{$W\ell_1$}$$

Now, no matter how much we shrink an atom in relation to the others, $\mathbf{N}$ balances all atom contributions in the constraints. Figure 1 illustrates how the solutions to (ℓ_1) and $(W\ell_1)$ can differ for a dictionary having atoms of different lengths. We generate a dictionary by sampling $N = 128$ atoms from the uniform spherical ensemble in $m = 48$ dimensions. We then scale the norm of each atom by sampling from a uniform distribution in $[1, 10]$. We select at random $s = 15$ atoms from the dictionary, and form $\mathbf{u}$ by linearly combining them with weights $\mathbf{x}$ sampled from a normal distribution. Finally, using CVX [7], we solve for $\mathbf{u} = \mathbf{D}\mathbf{x}$ the convex optimization problems posed by (ℓ_1) or $(W\ell_1)$. We see that the solution to (ℓ_1) does not match $\mathbf{x}$, while that of $(W\ell_1)$ does.
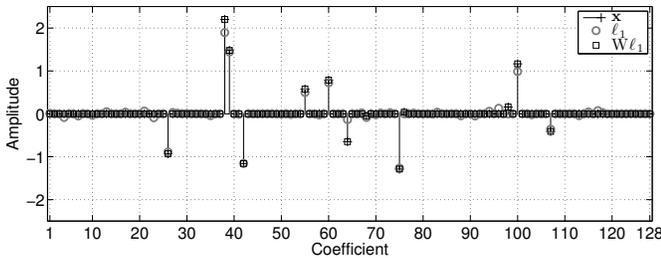

Fig. 1. For a dictionary having atoms of different lengths, we see the true solution $\mathbf{x}$ is recovered by (ℓ_1) but not $(W\ell_1)$.

We now derive an ERC for BP for the explicit cost in $(W\ell_1)$, and prove it is more general than the ERC in [2].

Theorem 1 (ERC of BP ($W\ell_1$) with $\mathcal{D}$): A sufficient condition for $(W\ell_1)$ to solve (P0) for any $\mathbf{u} \in \mathcal{V}^*$ is

$$\sigma_{\mathrm{BP}}(\Omega, I^*) := \max_{i \in \Omega_F(\mathcal{V}^*) \setminus I^*} \frac{\|\mathbf{N}_{I^*} \boldsymbol{\Phi}_{I^*}^\dagger \boldsymbol{\varphi}_i\|_1}{\|\boldsymbol{\varphi}_i\|_2} < 1. \tag{ERC-$W\ell_1$}$$

Proof: First, consider $\mathbf{y}^*$ to be the non-zero elements of the true solution $\mathbf{x}^*$ to $(W\ell_1)$ such that $\mathbf{u} = \boldsymbol{\Phi}\mathbf{x}^* = \boldsymbol{\Phi}_{I^*}\mathbf{y}^*$. By the constraints of (ℓ_1), we know $\mathbf{y}^* = \boldsymbol{\Phi}_{I^*}^\dagger \mathbf{u}$. Since $\mathcal{D}$ is overcomplete, there is another $\mathbf{x}'$ and $\mathbf{y}'$ such that $\mathbf{u} = \boldsymbol{\Phi}\mathbf{x}' = \boldsymbol{\Phi}_{I'}\mathbf{y}'$ where $I' = \mathrm{supp}(\mathbf{x}')$. Now, we know

$$\begin{aligned}
\|\mathbf{N}_{I^*}\mathbf{y}^*\|_1 &= \|\mathbf{N}_{I^*} \boldsymbol{\Phi}_{I^*}^\dagger \boldsymbol{\Phi}_{I'}\mathbf{y}'\|_1 \\
&= \|\mathbf{N}_{I^*} \boldsymbol{\Phi}_{I^*}^\dagger \boldsymbol{\Phi}_{I'} \mathbf{N}_{I'}^{-1} \mathbf{N}_{I'}\mathbf{y}'\|_1 \\
&\le \|\mathbf{N}_{I^*} \boldsymbol{\Phi}_{I^*}^\dagger \boldsymbol{\Phi}_{I'} \mathbf{N}_{I'}^{-1}\|_{1,1} \|\mathbf{N}_{I'}\mathbf{y}'\|_1 \quad (3)
\end{aligned}$$

by using the definition of the induced matrix norm, and what it implies, i.e.,

$$\|\mathbf{A}\|_{p,q} := \max_{\mathbf{w}} \frac{\|\mathbf{A}\mathbf{w}\|_p}{\|\mathbf{w}\|_q} \implies \|\mathbf{A}\mathbf{w}\|_p \le \|\mathbf{A}\|_{p,q}\|\mathbf{w}\|_q. \tag{4}$$

Since $\|\mathbf{A}\|_{1,1}$ is the maximum sum of the magnitude elements of its columns, we know

$$\|\mathbf{N}_{I^*} \boldsymbol{\Phi}_{I^*}^\dagger \boldsymbol{\Phi}_{I'} \mathbf{N}_{I'}^{-1}\|_{1,1} = \max_{i \in I'} \frac{\|\mathbf{N}_{I^*} \boldsymbol{\Phi}_{I^*}^\dagger \boldsymbol{\varphi}_i\|_1}{\|\boldsymbol{\varphi}_i\|_2}. \tag{5}$$

Recall $\mathbf{N}$ is diagonal and positive, all elements of $\mathbf{y}'$ are non-zero, and so $\|\mathbf{N}_{I'}\mathbf{y}'\|_1 > 0$. Hence, dividing both sides of (3) by $\|\mathbf{N}_{I'}\mathbf{y}'\|_1$ produces

$$\frac{\|\mathbf{N}_{I^*}\mathbf{y}^*\|_1}{\|\mathbf{N}_{I'}\mathbf{y}'\|_1} \le \max_{i \in I'} \frac{\|\mathbf{N}_{I^*} \boldsymbol{\Phi}_{I^*}^\dagger \boldsymbol{\varphi}_i\|_1}{\|\boldsymbol{\varphi}_i\|_2}. \tag{6}$$

For $(W\ell_1)$ to find $\mathbf{x}^*$ instead of $\mathbf{x}'$, we require $\|\mathbf{N}\mathbf{x}^*\|_1 < \|\mathbf{N}\mathbf{x}'\|_1$, or equivalently $\|\mathbf{N}_{I^*}\mathbf{y}^*\|_1 < \|\mathbf{N}_{I'}\mathbf{y}'\|_1$. Thus, a sufficient condition for the solution of $(W\ell_1)$ to involve only atoms indexed by I^*, and none by $I' \setminus I^* \subset \Omega_F(\mathcal{V}^*) \setminus I^*$, given any $\mathbf{u} \in \mathcal{V}^*$ is thus (ERC-$W\ell_1$). ∎

Corollary 1 (ERC of BP ($W\ell_1$) for unit-norm $\mathcal{D}$): If all atoms in $\mathcal{D}$ are unit-norm, then (ERC-$W\ell_1$) becomes

$$\max_{i \in \Omega_F(\mathcal{V}^*) \setminus I^*} \|\boldsymbol{\Phi}_{I^*}^\dagger \boldsymbol{\varphi}_i\|_1 < 1. \tag{ERC-ℓ_1}$$

Proof: This comes immediately from considering (ERC-$W\ell_1$) with atoms having unit-norm, i.e., $\mathbf{N} = \mathbf{I}$. ∎

Note that (ERC-ℓ_1) is different from (ERC-T) because, by the constraints of (ℓ_1), some atoms may not be in $\Omega_F(\mathcal{V}^*)$ — regardless of the atoms in the dictionary having unit norm. Overlooking the feasible set produces a less general and more strict ERC of BP in [2].

With its greedy iterative approach, OMP does not consider the feasible set. For a non-unit-norm dictionary, the atom selection criterion of OMP becomes

$$I_{k+1} \leftarrow I_k \cup \arg\max_{n \in \Omega} \frac{|\langle \mathbf{u}^{\perp I_k}, \boldsymbol{\varphi}_n \rangle|}{\|\boldsymbol{\varphi}_n\|_2}. \tag{7}$$

As before, this comes from the desire to minimize in a greedy way the residual error (before orthogonal projection), i.e.,

$$\arg\min_{n \in \Omega} \left[\min_{\alpha \in \mathbb{C}} \|\mathbf{u}^{\perp I_k} - \alpha \boldsymbol{\varphi}_n\|_2^2 \right] = \arg\max_{n \in \Omega} \frac{|\langle \mathbf{u}^{\perp I_k}, \boldsymbol{\varphi}_n \rangle|^2}{\|\boldsymbol{\varphi}_n\|_2^2} \tag{8}$$

For this case, we now adapt Tropp's proof of the ERC for OMP with unit-norm dictionaries [2].

Theorem 2 (ERC of OMP (7) with $\mathcal{D}$): A sufficient condition for OMP with $\mathcal{D}$ to solve (P0) for any $\mathbf{u} \in \mathcal{V}^*$ is

$$\sigma_{\text{OMP}}(\Omega, I^*) := \max_{i \in \Omega \setminus I^*} \frac{\|\mathbf{N}_{I^*} \mathbf{\Phi}_{I^*}^\dagger \boldsymbol{\varphi}_i\|_1}{\|\boldsymbol{\varphi}_i\|_2} < 1. \quad \text{(ERCOMP)}$$

Proof: Assume that $\|\mathbf{u}^{\perp I_k}\| > 0$ (otherwise OMP stops). In iteration $k+1$, OMP selects by (7) an element from I^* if

$$\max \left\{ \frac{|\langle \mathbf{u}^{\perp I_k}, \boldsymbol{\varphi}_n \rangle|}{\|\boldsymbol{\varphi}_n\|_2} \right\}_{n \in I^*} > \max \left\{ \frac{|\langle \mathbf{u}^{\perp I_k}, \boldsymbol{\varphi}_l \rangle|}{\|\boldsymbol{\varphi}_l\|_2} \right\}_{l \in \Omega \setminus I^*} \tag{9}$$

or equivalently $\|\mathbf{N}_{I^*}^{-1} \mathbf{\Phi}_{I^*}^H \mathbf{u}^{\perp I_k}\|_\infty > \|\mathbf{N}_{\Omega \setminus I^*}^{-1} \mathbf{\Phi}_{\Omega \setminus I^*}^H \mathbf{u}^{\perp I_k}\|_\infty$. Dividing the right side by the other, we see the condition is

$$\frac{\|\mathbf{N}_{\Omega \setminus I^*}^{-1} \mathbf{\Phi}_{\Omega \setminus I^*}^H \mathbf{u}^{\perp I_k}\|_\infty}{\|\mathbf{N}_{I^*}^{-1} \mathbf{\Phi}_{I^*}^H \mathbf{u}^{\perp I_k}\|_\infty} < 1. \tag{10}$$

Assuming OMP has selected k elements of I^*, $\mathbf{u}^{\perp I_k}$ must lie in $\mathcal{V}^*$, and so $\mathbf{u}^{\perp I_k} = (\mathbf{\Phi}_{I^*}^\dagger)^H \mathbf{\Phi}_{I^*}^H \mathbf{u}^{\perp I_k}$. Substituting this above

$$\frac{\|\mathbf{N}_{\Omega \setminus I^*}^{-1} \mathbf{\Phi}_{\Omega \setminus I^*}^H (\mathbf{\Phi}_{I^*}^\dagger)^H \mathbf{N}_{I^*} \mathbf{N}_{I^*}^{-1} \mathbf{\Phi}_{I^*}^H \mathbf{u}^{\perp I_k}\|_\infty}{\|\mathbf{N}_{I^*}^{-1} \mathbf{\Phi}_{I^*}^H \mathbf{u}^{\perp I_k}\|_\infty} < 1 \tag{11}$$

where we have inserted $\mathbf{N}_{I^*} \mathbf{N}_{I^*}^{-1}$. By (4), we see

$$\frac{\|\mathbf{N}_{\Omega \setminus I^*}^{-1} \mathbf{\Phi}_{\Omega \setminus I^*}^H (\mathbf{\Phi}_{I^*}^\dagger)^H \mathbf{N}_{I^*} \mathbf{N}_{I^*}^{-1} \mathbf{\Phi}_{I^*}^H \mathbf{r}_k\|_\infty}{\|\mathbf{N}_{I^*}^{-1} \mathbf{\Phi}_{I^*}^H \mathbf{r}_k\|_\infty}$$
$$\leq \|\mathbf{N}_{\Omega \setminus I^*}^{-1} \mathbf{\Phi}_{\Omega \setminus I^*}^H (\mathbf{\Phi}_{I^*}^\dagger)^H \mathbf{N}_{I^*}\|_{\infty, \infty}. \tag{12}$$

Since the right-hand side is the maximum sum along the magnitude rows, and is equivalent to the maximum sum along the magnitude columns of the transposed argument, we see

$$\|\mathbf{N}_{I^*} \mathbf{\Phi}_{I^*}^\dagger \mathbf{\Phi}_{\Omega \setminus I^*} \mathbf{N}_{\Omega \setminus I^*}^{-1}\|_{1,1} = \max_{i \in \Omega \setminus I^*} \frac{\|\mathbf{N}_{I^*} \mathbf{\Phi}_{I^*}^\dagger \boldsymbol{\varphi}_i\|_1}{\|\boldsymbol{\varphi}_i\|_2}. \tag{13}$$

Bounding this to be strictly less than 1 gives (ERCOMP). ∎

Corollary 2 (ERC of OMP (7) for unit-norm $\mathcal{D}$): If all atoms in $\mathcal{D}$ are unit-norm, then (ERCOMP) becomes (ERC-T).

Proof: This comes immediately from considering (ERCOMP) with atoms having unit-norm, i.e., $\mathbf{N} = \mathbf{I}$. ∎

We can appreciate the difference between (ERC-$W\ell_1$) and (ERCOMP) with the next theorem.

Theorem 3: If (ERCOMP) holds, then so does (ERC-$W\ell_1$); but not necessarily the converse. Similarly, if (ERC-T) holds, then so does (ERC-ℓ_1); but not necessarily the converse.

Proof: Since the set over which we evaluate (ERC-$W\ell_1$) or (ERC-ℓ_1) is contained in the set over which we evaluate (ERCOMP) and (ERC-T), i.e., $\Omega_F(\mathcal{V}^*) \setminus I^* \subseteq \Omega \setminus I^*$, then if (ERCOMP) holds, so must (ERC-$W\ell_1$); and if (ERC-T) holds, so must (ERC-ℓ_1). The converse, however, is not true. Consider the following dictionary [1]:

$$\mathbf{\Phi} := [\boldsymbol{\varphi}_1 | \boldsymbol{\varphi}_2 | \boldsymbol{\varphi}_3] := \begin{bmatrix} 1 & 0 & 1/\sqrt{3} \\ 0 & 1 & 1/\sqrt{3} \\ 0 & 0 & 1/\sqrt{3} \end{bmatrix}. \tag{14}$$

While (ERCOMP) does not hold for all atom pairs, i.e., $\sigma_{\text{OMP}}(\Omega, \{1,3\}) = \sigma_{\text{OMP}}(\Omega, \{2,3\}) = (\sqrt{3}+1)/2$, and $\sigma_{\text{OMP}}(\Omega, \{1,2\}) = 2/\sqrt{3}$, (ERC-$W\ell_1$) holds for all pairs, i.e., $\Omega_F(\text{span}\{1,2\}) \setminus \{1,2\} = \Omega_F(\text{span}\{1,3\}) \setminus \{1,3\} = \Omega_F(\text{span}\{2,3\}) \setminus \{2,3\} = \emptyset$ (with the logical extension of (ERC-$W\ell_1$) to define "success" when the set over which it is evaluated is empty). Finally, we can always expand $\Omega_F(\mathcal{V}^*) \setminus I^*$ with an atom not in $\text{span}(\Omega_F(\mathcal{V}^*))$ such that (ERC-ℓ_1) still holds, but (ERC-T) does not. ∎

III. THEOREM 10 OF PLUMBLEY [1]

Theorem 10 of Plumbley [1] says that if $\mathcal{D}$ satisfies (ERC-T) for all $\{I \subseteq \Omega : |I| = s\}$, $\mathcal{D}$ is not guaranteed to also satisfy (ERC-T) for all $\{I \subset \Omega : |I| = k < s\}$. In other words, that (ERC-T) is satisfied for all signals of one sparsity, it is not necessarily satisfied for all signals with a smaller sparsity. The proof is by counterexample. We first define the dictionary matrix

$$\mathbf{\Phi} := [\boldsymbol{\varphi}_1 | \boldsymbol{\varphi}_2] := \begin{bmatrix} 1 & \sqrt{2} \\ 0 & \sqrt{2} \end{bmatrix}. \tag{15}$$

When $s = 2$, (ERC-T) is trivially satisfied, since $\Omega \setminus I^* = \emptyset$, and thus $\max_{i \in \Omega \setminus I^*} \|\mathbf{\Phi}_{I^*}^\dagger \boldsymbol{\varphi}_i\|_1 = 0$. Define $\mathbf{x}^* := [\beta, 0]^T$. Thus, (ERC-T) fails since $|\boldsymbol{\varphi}_1^T \boldsymbol{\varphi}_2| = \sqrt{2} > 1$. Therefore, though $\mathcal{D}$ satisfies (ERC-T) for all 2-sparse representations, it does not for all 1-sparse representations. For the dictionary in (14), we see it trivially satisfies (ERC-T) for all 3-sparse representations; but when $\mathbf{x} = [1,1,0]^T$, (ERC-T) fails.

The first problem with Theorem 10 of Plumbley [1] is that (ERC-T) is undefined for a $|\Omega|$-sparse solution, i.e., $\max_{i \in \Omega \setminus \Omega} \|\mathbf{\Phi}_{I^*}^\dagger \boldsymbol{\varphi}_i\|_1$ is undefined since $\Omega \setminus \Omega = \emptyset$. We can sensibly extend the conditions in (ERC-T) to include the case where searching in an empty set automatically implies exact recovery. In the case of a $|\Omega|$-sparse solution then, we are guaranteed any sparse representation algorithm will always return atoms from the dictionary — which is not really a useful result. Furthermore, a solution using all the atoms of an overcomplete dictionary is *a priori* not sparse, and thus it is meaningless to consider the exact recovery condition for such solutions. The second problem is that Theorem 10 of Plumbley [1], focusing on the form of (ERC-T), neglects a key assumption made at the beginning of its derivation: that the dictionary has atoms with unit norms. If we first make the columns of (15) have unit-norm, then we see (ERC-T) is then satisfied for all 1-sparse representations.

Now, while Theorem 10 of Plumbley [1] is in essence true for (ERC-T) (with the extension to an empty set above), it does not translate into a result for recovery for either BP or OMP. It neglects the dependence of the guarantee of exact recovery on the recovery algorithm — a dependence that is subtle in the case of ($W\ell_1$) — and is not useful for non-normalized dictionaries — for which one must use (ERC-$W\ell_1$) or (ERCOMP) instead of (ERC-T). Furthermore, Theorem 10 of Plumbley [1] does not reflect the "nestedness" of recovery conditions for either OMP or BP. In fact, we show in [8] that for any dictionary, if OMP can recover all s-sparse signals, then it can recover all signals with sparsity $k < s$.

IV. CONCLUSION

We have shown Theorem 10 of Plumbley [1], while true for (ERC-T), does not reflect useful recovery conditions for either OMP or BP. The dictionary is not the only arbiter in the recovery of sparse representations; one must consider the dictionary *and* the algorithm. This has led us to develop ERCs for both OMP and BP that take into consideration that a dictionary may not have atoms of the same norm. Finally, we find an ERC more general for BP than that of Tropp [2].

REFERENCES

[1] 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.

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

[3] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, Aug. 1998.

[4] 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, pp. 40–44.

[5] 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.

[6] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. New York, NY, USA: Springer, 2010.

[7] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 1.21," http://cvxr.com/cvx, Apr. 2011.

[8] B. Mailhé, B. L. Sturm, and M. D. Plumbley, "Recovery of nested supports by greedy sparse approximation algorithms," in *Proc. Int. Conf. Acoustics, Speech, Signal Process.*, 2013 (accepted).

Mark Plumbley (S'88, M'90, SM'12) received the B.A. (Hons.) degree in electrical sciences in 1984 from the University of Cambridge, Cambridge, U.K., and the Ph.D. degree in neural networks in 1991, also from the University of Cambridge. From 1991 to 2001 he was a Lecturer at King's College London. He moved to Queen Mary University of London in 2002, and where he is now an EPSRC Leadership Fellow and Director of the Centre for Digital Music. His research focuses on the automatic analysis of music and other audio sounds, including automatic music transcription, beat tracking, and audio source separation, and with interest in the use of techniques such as independent component analysis (ICA) and sparse representations. Prof. Plumbley is a member of the IEEE SPS TC on Audio and Acoustic Signal Processing.

Bob L. Sturm (S'04, M'09) received the B.A. degree in physics from the University of Colorado, Boulder, in 1998, the M.A. degree in Music, Science, and Technology, from Stanford University, Stanford, CA, in 1999, a M.S. degree in multimedia engineering in the Media Arts and Technology program, University of California, Santa Barbara (UCSB), in 2004, and the M.S. and Ph.D. degrees in electrical and computer engineering from UCSB, in 2007 and 2009, respectively. He is currently an associate professor in the Audio Analysis Lab, AD:MT, Aalborg University Copenhagen, Denmark. He research interests include signal processing, sparse approximation, machine learning, and their applications to audio and music signals.

Boris Mailhé (S'09, M'12) was born in Lyon. He received his B.Sc. and M.Sc. in computer science from Rennes 1 University and Ecole Normale Suprieure Cachan in 2004 and 2005 and his Ph.D. in signal processing from Rennes 1 University in 2010. Since 2011, he is a Postdoctoral Research Assistant at Queen Mary, University of London. His research interests include large-scale sparse signal processing and non-convex optimization methods.