OMP with Highly Coherent Dictionaries

Raja Giryes and Michael Elad
Computer Science Department
Technion - IIT 32000, Haifa, ISRAEL
Email: [raja,elad]@cs.technion.ac.il

***Abstract*—Recovering signals that has a sparse representation from a given set of linear measurements has been a major topic of research in recent years. Most of the work dealing with this subject focus on the reconstruction of the signal's representation as the means to recover the signal itself. This approach forces the dictionary to be of low-coherence and with no linear dependencies between its columns. Recently, a series of contributions show that such dependencies can be allowed by aiming at recovering the signal itself. However, most of these recent works consider the analysis framework, and only few discuss the synthesis model. This paper studies the synthesis and introduces a new mutual coherence definition for signal recovery, showing that a modified version of OMP can recover sparsely represented signals of a dictionary with very high correlations between pairs of columns. We show how the derived results apply to the plain OMP.**

I. Introduction

Much attention has been given to the problem of recovering a sparse signal from a given set of linear measurements in the recent decade. In the basic setup, an unknown signal $\mathbf{x} \in \mathbb{R}^d$ passes through a given linear transformation $\mathbf{M} \in \mathbb{R}^{m \times d}$ (including $m < d$) with an additive noise $\mathbf{e} \in \mathbb{R}^m$, providing $\mathbf{y} = \mathbf{M}\mathbf{x} + \mathbf{e}$. The signal $\mathbf{x}$ is assumed to have a k-sparse representation $\boldsymbol{\alpha} \in \mathbb{R}^n$ under a given dictionary $\mathbf{D} \in \mathbb{R}^{d \times n}$, i.e. $\mathbf{x} = \mathbf{D}\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}$ has at most k non-zero entries. Most existing work dealing with the problem of estimating $\mathbf{x}$ from $\mathbf{y}$ focuses on the recovery of the signal's representation, assuming that this would lead to the desired signal recovery. This approach forces $\mathbf{D}$ to be incoherent, and in particular, with no linear dependencies between small groups of its atoms.

Recently, a series of papers have shown that such dependencies can be permitted by aiming at estimating the signal itself [1], [2], [3], [4], [5]. Indeed, in [3], [4], [5] it is even suggested that such linear dependencies should be encouraged. However, these contributions consider the "analysis" framework. A first clue that this is not unique to the analysis model but rather applicable also to the "synthesis" appears in [1]. Though its results are for signals from the analysis model, the recovery conditions rely on the $\mathbf{D}$-RIP, a synthesis model property.

The work reported in [6] is different and daring, as it addresses the synthesis model, presenting a variation of CoSaMP that targets the recovery of the signal directly. In their theoretical study, they use the $\mathbf{D}$-RIP to analyze the algorithm's performance assuming the existence of an efficient near-optimal projection scheme, like in [4]. However, the availability of such a projection is questionable in the general case. Another recent work that exploits the $\mathbf{D}$-RIP in the context of the synthesis is [7], proposing stable signal recovery conditions for the basic synthesis ℓ_0-minimization problem.

It is interesting to note that in [6] it is observed that orthogonal matching pursuit (OMP) [8], though not backed up theoretically, achieves some success in recovering signals in the presence of high coherence in the dictionary. In this work we make the first steps to explain this behaviour. We propose a slightly modified version of OMP, OMP$_{\epsilon,2}$, and analyze its performance in the noiseless case ($\mathbf{e} = 0$). Instead of using the $\mathbf{D}$-RIP, we rely on a new property of $\mathbf{M}$ and $\mathbf{D}$: The ϵ-coherence μ_ϵ, which generalizes the definition of the regular coherence μ. Using this definition we show that if $k \leq \frac{1}{2}(1+\frac{1}{\mu_\epsilon}) - O(\epsilon)$ then the OMP$_{\epsilon,2}$ signal recovery error is $O(\epsilon)$. This result implies that OMP$_{\epsilon,2}$ achieves an almost exact reconstruction in the case of very high correlations within pairs of dictionary columns. We draw also the connection between OMP and OMP$_{\epsilon,2}$. Note that our conditions do not include the need for an efficient projection, as needed in [6] .

The organization of this paper is as follows. Section II introduces the ϵ-coherence along with other new definitions. In Section III a modified version of OMP is introduced to support high correlation between pairs of columns. In Section IV the algorithm is analyzed using the ϵ-coherence providing some performance guarantees for the noiseless case. In Section V the derived results are demonstrated empirically.

II. New Coherence Definition

We start with some notation. The largest singular value of a matrix $\mathbf{M}$ is denoted by $\sigma_{\mathbf{M}}$. The i-th column/element of a matrix/vector $\mathbf{D}$/$\mathbf{x}$ is denoted by $\mathbf{d}_i$/$\mathbf{x}_i$, and the sub-matrix/vector with the entries of the support set T by $\mathbf{D}_T$/$\boldsymbol{\alpha}_T$. By abuse of notation, $\boldsymbol{\alpha}_T$ corresponds both to the sub-vector with these entries alone and to the zero padded one. We denote by $\mathbf{W}_\mathbf{D}$ a diagonal matrix that contains the norms of the columns of $\mathbf{D}$ on its diagonal, i.e, $\mathbf{W}_{i,i} = \|\mathbf{d}_i\|_2$.

We turn to introduce some definitions which serve as building blocks in our proposed algorithm and theoretical study. As in [9] the columns of $\mathbf{MD}$ are assumed to be normalized, since if this is not the case a simple scaling can be applied.

Definition 2.1 (ϵ-coherence): Let $0 \leq \epsilon < 1$, $\mathbf{M}$ be a fixed measurement matrix and $\mathbf{D}$ be a fixed dictionary. The ϵ-coherence $\mu_\epsilon(\mathbf{M}, \mathbf{D})$ is defined as

$$\mu_\epsilon(\mathbf{M}, \mathbf{D}) = \max_{1 \leq i < j \leq n} |\langle \mathbf{M}\mathbf{d}_i, \mathbf{M}\mathbf{d}_j \rangle| \quad s.t. \quad \frac{|\langle \mathbf{d}_i, \mathbf{d}_j \rangle|^2}{\|\mathbf{d}_i\|_2^2 \|\mathbf{d}_j\|_2^2} < 1 - \epsilon^2. \tag{1}$$

For calculating $\mu_\epsilon(\mathbf{M},\mathbf{D})$, one may compute the Gram matrices $\mathbf{G_{MD}} = \mathbf{D}^*\mathbf{M}^*\mathbf{MD}$ and $\mathbf{G_D} = \mathbf{W_D^{-1}}\mathbf{D}^*\mathbf{D}\mathbf{W_D^{-1}}$. The ϵ-coherence is simply the value of the largest off-diagonal element in absolute value in $\mathbf{G_{MD}}$, corresponding to an entry in $\mathbf{G_D}$ that is smaller in its absolute value than $\sqrt{1-\epsilon^2}$. Note that for $\mathbf{D}=\mathbf{I}$, the ϵ-coherence coincides with the regular coherence $\mu(\mathbf{M})$ and we have $\mu_\epsilon(\mathbf{M},\mathbf{I}) = \mu(\mathbf{M})$. When it is clear to which $\mathbf{M}$ and $\mathbf{D}$ we refer, we use simply μ_ϵ.

Definition 2.2 (ϵ-independent support set): Let $0 \le \epsilon < 1$, $\mathbf{D}$ be a fixed dictionary. A support set T is ϵ-independent with respect to a dictionary $\mathbf{D}$ if $\forall i \neq j \in T$, $\frac{|\langle \mathbf{d}_i, \mathbf{d}_j\rangle|^2}{\|\mathbf{d}_i\|_2^2\|\mathbf{d}_j\|_2^2} < 1-\epsilon^2$.

Definition 2.3 (ϵ-closure): Let $0 \le \epsilon < 1$ and $\mathbf{D}$ be a fixed dictionary. The ϵ-closure of a given support T is defined as $clos_{\epsilon,2}(T) = \{i | \exists j \in T,\ \frac{|\langle \mathbf{d}_i, \mathbf{d}_j\rangle|^2}{\|\mathbf{d}_i\|_2^2\|\mathbf{d}_j\|_2^2} \ge 1-\epsilon^2\}$.

The ϵ-closure of a support T extends it to include each column in $\mathbf{D}$ which is "ϵ-correlated" with elements included in T. Obviously, $T \subseteq clos_{\epsilon,2}(T)$. Note that the last two definitions are related to a given dictionary $\mathbf{D}$. If $\mathbf{D}$ is clear from the context, it is omitted.

III. ϵ-Orthogonal Matching Pursuit

In order to treat the ϵ dependencies in a dictionary we propose the ϵ-orthogonal matching pursuit (OMP$_{\epsilon,2}$) presented in Algorithm 1, which is a modification of OMP [8]. OMP$_{\epsilon,2}$ is the same as the regular OMP but with the addition of the ϵ-closure step. The methods coincide if $\epsilon = 0$ as OMP's orthogonality property guarantees not selecting the same vector twice and thus the ϵ-closure step in OMP$_{\epsilon,2}$ has no effect.

IV. Algorithm Recovery Guarantees

We start with the following Lemma.

Lemma 4.1: Let $\mathbf{x} = \mathbf{D}\boldsymbol{\alpha}$, T be the support of $\boldsymbol{\alpha}$, $\tilde{T}$ be a support set such that $T \subseteq clos_{\epsilon,2}(\tilde{T})$, $\beta_i = \frac{\langle \mathbf{d}_i, \mathbf{d}_{F(i,\mathbf{D}_T)}\rangle}{\left\|\mathbf{d}_{F(i,\mathbf{D}_T)}\right\|_2^2}$ and $\tilde{i} = F(i,\mathbf{D}_T)$ is a function of i such that $\frac{\left|\langle \mathbf{d}_i, \mathbf{d}_{\tilde{i}}\rangle\right|^2}{\|\mathbf{d}_i\|_2^2\left\|\mathbf{d}_{\tilde{i}}\right\|_2^2} \ge 1-\epsilon^2$. If there are several possible $\tilde{i}$ for a given i, choose any one of those and proceed. For the construction

$$\tilde{\mathbf{x}} = \sum_{i\in T\cap\tilde{T}} \mathbf{d}_i\boldsymbol{\alpha}_i + \sum_{i\in T\setminus\tilde{T}} \beta_i \mathbf{d}_{F(i,\mathbf{D}_T)}\boldsymbol{\alpha}_i, \tag{2}$$

we have

$$\|\mathbf{x}-\tilde{\mathbf{x}}\|_2^2 \le \left\|\mathbf{W}_{\mathbf{D}_T}\boldsymbol{\alpha}_{T\setminus\tilde{T}}\right\|_1^2 \epsilon^2. \tag{3}$$

Proof: Note that $\mathbf{x}-\tilde{\mathbf{x}} = \sum_{i\in T\setminus\tilde{T}} \left(\mathbf{d}_i - \beta_i\mathbf{d}_{F(i,\mathbf{D}_T)}\right)\boldsymbol{\alpha}_i$ and $\left\|\mathbf{d}_i - \beta_i\mathbf{d}_{F(i,\mathbf{D}_T)}\right\|_2^2 = \|\mathbf{d}_i\|_2^2\left(1 - \frac{\left|\langle \mathbf{d}_i, \mathbf{d}_{\tilde{i}}\rangle\right|^2}{\|\mathbf{d}_i\|_2^2\left\|\mathbf{d}_{\tilde{i}}\right\|_2^2}\right) \le \|\mathbf{d}_i\|_2^2\,\epsilon^2$. The Cauchy-Schwartz inequality with some arithmetics gives

$$\begin{aligned}\|\mathbf{x}-\tilde{\mathbf{x}}\|_2^2 &= \left\|\sum_{i\in T\setminus\tilde{T}} \left(\mathbf{d}_i - \beta_i\mathbf{d}_{F(i,\mathbf{D}_T)}\right)\boldsymbol{\alpha}_i\right\|_2^2 \qquad (4)\\ &= \sum_{i,j\in T\setminus\tilde{T}} \left(\mathbf{d}_i - \beta_i\mathbf{d}_{F(i,\mathbf{D}_T)}\right)^* \left(\mathbf{d}_j - \beta_j\mathbf{d}_{F(j,\mathbf{D}_T)}\right)\boldsymbol{\alpha}_i\boldsymbol{\alpha}_j\\ &\le \sum_{i\in T\setminus\tilde{T}} \epsilon^2\|\mathbf{d}_i\|_2^2\,\boldsymbol{\alpha}_i^2 + \sum_{i\neq j\in T\setminus\tilde{T}} \epsilon^2\|\mathbf{d}_i\|_2\|\mathbf{d}_j\|_2\,\boldsymbol{\alpha}_i\boldsymbol{\alpha}_j.\end{aligned}$$

Algorithm 1 ϵ-Orthogonal Matching Pursuit

Require: $k, \mathbf{M}, \mathbf{D}, \mathbf{y}$ where $\mathbf{y} = \mathbf{Mx}+\mathbf{e}$, $\mathbf{x} = \mathbf{D}\boldsymbol{\alpha}$, $\|\boldsymbol{\alpha}\|_0 \le k$ and $\mathbf{e}$ is an additive noise.

Ensure: $\hat{\mathbf{x}}_{\text{OMP}_{\epsilon,2}}$: k-sparse approximation of $\mathbf{x}$.

Initialize estimate $\hat{\mathbf{x}}^0 = \mathbf{0}$, residual $\mathbf{r}^0 = \mathbf{y}$, support $\hat{T}^0 = \check{T}^0 = \emptyset$ and set $t = 0$.

while $t \le k$ **do**

$t = t+1$.

New support element: $i^t = \text{argmax}_{i\notin\check{T}^{t-1}} |\mathbf{d}_i^*\mathbf{M}^*(\mathbf{r}^{t-1})|$.

Extend support: $\hat{T}^t = \hat{T}^{t-1}\cup\{i^t\}$.

Calculate a new estimate: $\hat{\mathbf{x}}^t_{\text{OMP}_{\epsilon,2}} = \mathbf{D}_{\hat{T}^t}(\mathbf{MD}_{\hat{T}^t})^\dagger\mathbf{y}$.

Calculate a new residual: $\mathbf{r}^t = \mathbf{y} - \mathbf{M}\hat{\mathbf{x}}^t_{\text{OMP}_{\epsilon,2}}$.

Support ϵ-closure: $\check{T}^t = clos_{\epsilon,2}(\hat{T}^t)$.

end while

Form the final solution $\hat{\mathbf{x}}_{\text{OMP}_{\epsilon,2}} = \hat{\mathbf{x}}^k_{\text{OMP}_{\epsilon,2}}$.

By the definitions of the ℓ_1-norm and $\mathbf{W}_{\mathbf{D}_T}$ we have that the rhs (right-hand-side) of (4) is equal to the rhs of (3). □

Theorem 4.2: Let $0 \le \epsilon < 1$, $\mathbf{M}$ be a fixed measurement matrix, $\mathbf{D}$ be a fixed dictionary with ϵ-coherence $\mu_\epsilon = \mu_\epsilon(\mathbf{M},\mathbf{D})$ and $\mathbf{y} = \mathbf{Mx}$ be a set of measurements of $\mathbf{x} = \mathbf{D}\boldsymbol{\alpha}$ where $\boldsymbol{\alpha}$ is supported on T and $|T| = k$. Let $\tilde{T} \subseteq T$ be an ϵ-independent set such that $T \subseteq clos_{\epsilon,2}(\tilde{T})$ and $\tilde{\mathbf{x}} = \mathbf{D}\tilde{\boldsymbol{\alpha}}$ is constructed according to (2) such that $\tilde{\boldsymbol{\alpha}}$ is supported on $\tilde{T}$. If

$$k < \frac{1}{2}(1+\frac{1}{\mu_\epsilon}) - \frac{2\|\mathbf{W}_\mathbf{D}\tilde{\boldsymbol{\alpha}}\|_1 + \left\|\mathbf{W}_\mathbf{D}\boldsymbol{\alpha}_{T\setminus\tilde{T}}\right\|_1}{|\tilde{\boldsymbol{\alpha}}_{\min}|\,\mu_\epsilon}\sigma_\mathbf{M}\epsilon, \tag{5}$$

where $\tilde{\boldsymbol{\alpha}}_{\min}$ is the minimal non-zero entry in absolute value of $\tilde{\boldsymbol{\alpha}}$, then after k iterations at most, $\hat{\mathbf{x}}_{\text{OMP}_{\epsilon,2}}$ satisfies

$$\left\|\hat{\mathbf{x}}_{\text{OMP}_{\epsilon,2}} - \mathbf{x}\right\|_2^2 \le \left\|\mathbf{W}_{\mathbf{D}_{T\setminus\tilde{T}}}\boldsymbol{\alpha}_{T\setminus\tilde{T}}\right\|_1^2\epsilon^2 + \|\mathbf{W}_\mathbf{D}\tilde{\boldsymbol{\alpha}}\|_1^2\,\epsilon^2. \tag{6}$$

In particular, if T is an ϵ-independent set then $\boldsymbol{\alpha} = \tilde{\boldsymbol{\alpha}}$ and

$$\left\|\hat{\mathbf{x}}_{\text{OMP}_{\epsilon,2}} - \mathbf{x}\right\|_2^2 \le \|\mathbf{W}_\mathbf{D}\boldsymbol{\alpha}\|_1^2\,\epsilon^2. \tag{7}$$

Before proceeding we comment on the role of ϵ and $\tilde{T}$ in the theorem. If two columns are ϵ-correlated and we use the regular coherence μ, the condition in (5) cannot be met. The use of ϵ-coherence allows us to ignore these correlations and have a reduced coherence value. Thus, the value of ϵ determines the level of correlations the algorithm can handle. Condition (5) bounds this level by $\frac{\frac{1}{2}(\mu_\epsilon+1)-k\mu_\epsilon}{2\|\mathbf{W}_\mathbf{D}\tilde{\boldsymbol{\alpha}}\|_1+\left\|\mathbf{W}_\mathbf{D}\boldsymbol{\alpha}_{T\setminus\tilde{T}}\right\|_1}\frac{|\tilde{\boldsymbol{\alpha}}_{\min}|}{\sigma_\mathbf{M}}$. Remark that as ϵ approaches zero the value of μ_ϵ approaches μ_0, a mutual coherence of $\mathbf{D}$ that ignores the dependent columns.

The set $\tilde{T}$ is needed in the theorem because the columns of $\mathbf{D}_T$, which span $\mathbf{x}$, might be ϵ-correlated or even dependent. To avoid that, we select the maximal subset of T which is ϵ-independent and still includes T in its ϵ-closure. The construction of such a maximal subset is easy. We start by initializing $\tilde{T} = T$, and then sequentially for each index $i \in \tilde{T}$ update $\tilde{T} = \tilde{T}\setminus clos_{\epsilon,2}(\{i\})$. The resulting subset $\tilde{T}$ is guaranteed to be ϵ-independent and have $T \subseteq clos_{\epsilon,2}(\tilde{T})$.

The following key Lemma is used in the Theorem's proof.

Lemma 4.3: Under the same setup of Theorem 4.2, we have

$$\tilde{T} \subseteq \check{T}^k = clos_{\epsilon,2}(\hat{T}^k). \tag{8}$$

Proof: We prove by induction on the iteration $t \leq |\tilde{T}| = \tilde{k}$ that either $\tilde{T} \subseteq \check{T}^t$ or $\exists i \in \tilde{T}$ such that $i \in \check{T}^t$ and $i \notin \check{T}^{t-1}$. Since the induction guarantees that in each iteration a new element from $\tilde{T}$ is included in $\check{T}^t$, after $k \geq \tilde{k}$ iterations (8) holds.

The basis of the induction is $t = 1$. Define $\bar{T} = clos_{\epsilon,2}(\tilde{T})$. The basis holds if in the first iteration we select an element from $\bar{T}$. This is true due to the fact that $\forall i, j \in T$, $i \in clos_{\epsilon,2}(\{j\})$ iff $j \in clos_{\epsilon,2}(\{i\})$. Thus, we need to require

$$\max_{i\in\bar{T}} |\mathbf{d}_i^*\mathbf{M}^*\mathbf{y}| > \max_{i\in\bar{T}^C} |\mathbf{d}_i^*\mathbf{M}^*\mathbf{y}|. \tag{9}$$

First note that $\mathbf{y} = \mathbf{M}\tilde{\mathbf{x}} + \mathbf{M}(\mathbf{x} - \tilde{\mathbf{x}})$. Thus, using the triangle inequality, the Cauchy-Shwartz inequality and the facts that the ℓ_2-norm is multiplicative and $\|\mathbf{M}\mathbf{d}_i\|_2 = 1$, (9) holds if

$$\max_{i\in\bar{T}} |\mathbf{d}_i^*\mathbf{M}^*\mathbf{M}\tilde{\mathbf{x}}| > \max_{i\in\bar{T}^C} |\mathbf{d}_i^*\mathbf{M}^*\mathbf{M}\tilde{\mathbf{x}}| + 2\|\mathbf{M}(\mathbf{x} - \tilde{\mathbf{x}})\|_2. \tag{10}$$

In order to check when the last happens we shall bound its lhs (left-hand-side) from below and its rhs from above.

Assuming w.l.o.g that the index of the largest entry in $\tilde{\boldsymbol{\alpha}}$ is 1, we have for the lhs of (10)

$$\begin{aligned}\max_{i\in\bar{T}} |\mathbf{d}_i^*\mathbf{M}^*\mathbf{M}\tilde{\mathbf{x}}| &\geq |\mathbf{d}_1^*\mathbf{M}^*\mathbf{M}\tilde{\mathbf{x}}| = |\sum_{l\in\tilde{T}} \mathbf{d}_1^*\mathbf{M}^*\mathbf{M}\mathbf{d}_l\tilde{\boldsymbol{\alpha}}_l| \quad (11)\\ &\geq |\mathbf{d}_1^*\mathbf{M}^*\mathbf{M}\mathbf{d}_1\tilde{\boldsymbol{\alpha}}_1| - \sum_{l\in\tilde{T}, l\neq 1} |\mathbf{d}_1^*\mathbf{M}^*\mathbf{M}\mathbf{d}_l\tilde{\boldsymbol{\alpha}}_l| \\ &\geq \tilde{\boldsymbol{\alpha}}_1 - \mu_\epsilon \sum_{l\in\tilde{T}, l\neq 1} |\tilde{\boldsymbol{\alpha}}_l| = (1 - (\tilde{k} - 1)\mu)\,|\tilde{\boldsymbol{\alpha}}_1|,\end{aligned}$$

where the first inequality is due to the triangle inequality; the second is due to the fact that $\|\mathbf{M}\mathbf{d}_i\|_2 = 1$, the definition of μ_ϵ and the Cauchy-Schwartz inequality; and the last is because $\tilde{\boldsymbol{\alpha}}_1$ is the largest element in $\tilde{\boldsymbol{\alpha}}$ and $|\tilde{T}| = \tilde{k}$.

We turn now to bound the rhs of (10) from above. Using the same considerations, we have

$$\begin{aligned}\max_{i\in\bar{T}^C} |\mathbf{d}_i^*\mathbf{M}^*\mathbf{M}\mathbf{x}| &= \max_{i\in\bar{T}^C} |\sum_{l\in\tilde{T}} \mathbf{d}_i^*\mathbf{M}^*\mathbf{M}\mathbf{d}_l\tilde{\boldsymbol{\alpha}}_l| \quad (12)\\ \leq \max_{i\in\bar{T}^C} \sum_{l\in\tilde{T}} |\mathbf{d}_i^*\mathbf{M}^*\mathbf{M}\mathbf{d}_l\tilde{\boldsymbol{\alpha}}_l| &\leq \sum_{l\in\tilde{T}} \mu_\epsilon\,|\tilde{\boldsymbol{\alpha}}_l| \leq |\tilde{\boldsymbol{\alpha}}_1|\,\tilde{k}\mu_\epsilon.\end{aligned}$$

Plugging (11) and (12) into (10) and then using Lemma 4.1 with the fact that $\|\mathbf{M}\|_2 = \sigma_{\mathbf{M}}$ gives us the condition

$$\tilde{k} < \frac{1}{2}(1 + \frac{1}{\mu_\epsilon}) - \frac{\sigma_{\mathbf{M}}}{\mu_\epsilon\tilde{\boldsymbol{\alpha}}_1} \left\|\mathbf{W}_{\mathbf{D}_T}\boldsymbol{\alpha}_{T\setminus\tilde{T}}\right\|_1 \epsilon, \tag{13}$$

for selecting an element from $\bar{T}$ in the first iteration.

Having the induction basis proven, we turn to the induction step. Assume that the induction assumption holds till iteration $t-1$. We need to prove that it holds also in the t-th iteration. Let $\bar{T}^t = clos_{\epsilon,2}(\tilde{T} \setminus \check{T}^{t-1}))$. This set includes the ϵ-closure of elements in $\tilde{T}$ for which an element was not selected in the previous iterations. For proving the induction step it is enough to show that in the t-th iteration we select an index from $\bar{T}^t$:

$$\max_{i\in\bar{T}^t} \left|\mathbf{d}_i^*\mathbf{M}^*\mathbf{r}^{t-1}\right| > \max_{i\in(\bar{T}^t)^C\setminus\check{T}^{t-1}} \left|\mathbf{d}_i^*\mathbf{M}^*\mathbf{r}^{t-1}\right|. \tag{14}$$

On the rhs we do not check the maximum over elements in $\check{T}^{t-1}$ because $\text{OMP}_{\epsilon,2}$ excludes these indices in the step of selecting a new element. As in the basis of the induction, in order to check when (14) holds we shall bound its lhs from below and its rhs from above. Let $\tilde{\mathbf{x}}^{t-1} = \sum_{i\in\tilde{T}\setminus\hat{T}^{t-1}} \mathbf{d}_i\tilde{\boldsymbol{\alpha}}_i + \sum_{i\in\hat{T}^{t-1}} \beta_i\mathbf{d}_{F(i,D_{\tilde{T}})}\tilde{\boldsymbol{\alpha}}_i$ be constructed as in (2) where we use the fact that $\tilde{\boldsymbol{\alpha}}$ is supported on $\tilde{T}$. Denoting $\tilde{\mathbf{r}}^{t-1} = (\mathbf{I} - \mathbf{M}\mathbf{D}_{\hat{T}^{t-1}}(\mathbf{M}\mathbf{D}_{\hat{T}^{t-1}})^\dagger)\mathbf{M}\tilde{\mathbf{x}}^{t-1}$ and using a similar argument like in (10) we have that (14) holds if

$$\begin{aligned}\max_{i\in\bar{T}^t} |\mathbf{d}_i^*\mathbf{M}^*\tilde{\mathbf{r}}^{t-1}| \;&> \max_{i\in(\bar{T}^t)^C\setminus\check{T}^{t-1}} |\mathbf{d}_i^*\mathbf{M}^*\tilde{\mathbf{r}}^{t-1}| \quad (15)\\ &\quad + 2\left\|\tilde{\mathbf{r}}^{t-1} - \mathbf{r}^{t-1}\right\|_2.\end{aligned}$$

Notice that $\tilde{\mathbf{r}}^{t-1}$ is supported on $\hat{T}^{t-1}\cup(\tilde{T}\setminus\check{T}^{t-1})$, i.e., $\tilde{\mathbf{r}}^{t-1} = \mathbf{M}\mathbf{D}_{\hat{T}^{t-1}\cup(\tilde{T}\setminus\check{T}^{t-1})}\tilde{\boldsymbol{\alpha}}^{\mathbf{r}^{t-1}}$, and $\tilde{\boldsymbol{\alpha}}^{\mathbf{r}^{t-1}}_{\tilde{T}\setminus\check{T}^{t-1}} = \tilde{\boldsymbol{\alpha}}_{\tilde{T}\setminus\check{T}^{t-1}}$.

We want to show that the index of the maximal coefficient (in absolute value) of $\tilde{\mathbf{r}}^{t-1}$ belongs to $\tilde{T} \setminus \check{T}^{t-1}$ and hence we will be able to use almost the same derivation of the basis of the induction. We prove it by contradiction. Assume that the maximum is achieved for $i \in \hat{T}^{t-1}$. By the orthogonality property of the residual it is easy to see that $\mathbf{d}_i^*\mathbf{M}^*\tilde{\mathbf{r}}^{t-1} = 0$. Using similar considerations as in (11) we have $0 = \left|\mathbf{d}_i^*\mathbf{M}^*\tilde{\mathbf{r}}^{t-1}\right| \geq (1-(\tilde{k}-1)\mu_\epsilon)\left|\tilde{\boldsymbol{\alpha}}_i^{\mathbf{r}^{t-1}}\right|$ which implies $\tilde{k} \geq 1 + \frac{1}{\mu_\epsilon}$ and we get a contradiction to (5).

Let w.l.o.g. t be the maximal coefficient in $\tilde{\boldsymbol{\alpha}}_{i_t}^{\mathbf{r}^{t-1}}$. By the above observations $t \in \tilde{T} \setminus \check{T}^{t-1}$ and $\tilde{\boldsymbol{\alpha}}_t^{\mathbf{r}^{t-1}} = \tilde{\boldsymbol{\alpha}}_t$. Applying the same steps as in (11) and (12), we have

$$\begin{aligned}\max_{i\in\bar{T}^t} \left|\mathbf{d}_i^*\mathbf{M}^*\tilde{\mathbf{r}}^{t-1}\right| &\geq (1 - \mu_\epsilon(\tilde{k} - 1))\,|\tilde{\boldsymbol{\alpha}}_t|, \quad (16)\\ \max_{i\in(\bar{T}^t)^C\setminus\check{T}^{t-1}} \left|\mathbf{d}_i^*\mathbf{M}^*\tilde{\mathbf{r}}^{t-1}\right| &\leq \mu_\epsilon\tilde{k}\,|\tilde{\boldsymbol{\alpha}}_t|.\end{aligned}$$

Using norm inequalities and the projection property that implies $\left\|\mathbf{I} - \mathbf{M}\mathbf{D}_{\hat{T}^{t-1}}(\mathbf{M}\mathbf{D}_{\hat{T}^{t-1}})^\dagger\right\|_2 \leq 1$, we have

$$\begin{aligned}\left\|\tilde{\mathbf{r}}^{t-1} - \mathbf{r}^{t-1}\right\|_2 &\leq \left\|\mathbf{M}(\tilde{\mathbf{x}}^{t-1} - \mathbf{x})\right\|_2 \leq \sigma_{\mathbf{M}}\left\|\tilde{\mathbf{x}}^{t-1} - \mathbf{x}\right\|_2 \quad (17)\\ &\leq \sigma_{\mathbf{M}}\left\|\tilde{\mathbf{x}}^{t-1} - \tilde{\mathbf{x}}\right\|_2 + \sigma_{\mathbf{M}}\left\|\tilde{\mathbf{x}} - \mathbf{x}\right\|_2\end{aligned}$$

Using Lemma 4.1 with (17) and then combining it with (15) and (16) results with the condition

$$\tilde{k} < \frac{1}{2} + \frac{1}{2\mu_\epsilon} - \frac{\sigma_{\mathbf{M}}\epsilon}{|\tilde{\boldsymbol{\alpha}}_t|\,\mu_\epsilon}(\|\mathbf{W}_{\mathbf{D}}\tilde{\boldsymbol{\alpha}}_{\tilde{T}^{t-1}}\|_1 + \left\|\mathbf{W}_{\mathbf{D}}\boldsymbol{\alpha}_{T\setminus\tilde{T}}\right\|_1). \tag{18}$$

The proof ends by noticing that (18) is implied by (5). □

Proof of Theorem 4.2: Note that $\hat{\mathbf{x}}_{\text{OMP}_{\epsilon,2}} = \mathbf{D}_{\hat{T}^k}(\mathbf{M}\mathbf{D}_{\hat{T}^k})^\dagger\mathbf{y}$ and $\mathbf{y} = \mathbf{M}\mathbf{x}$. Using some basic algebraic steps we have

$$\begin{aligned}\left\|\hat{\mathbf{x}}_{\text{OMP}_{\epsilon,2}} - \mathbf{x}\right\|_2 &= \left\|\mathbf{D}_{\hat{T}^k}(\mathbf{M}\mathbf{D}_{\hat{T}^k})^\dagger\mathbf{M}\mathbf{x} - \mathbf{x}\right\|_2 \quad (19)\\ &= \left\|(\mathbf{D}_{\hat{T}^k}(\mathbf{M}\mathbf{D}_{\hat{T}^k})^\dagger\mathbf{M} - \mathbf{I})(\mathbf{I} - \mathbf{D}_{\hat{T}^k}\mathbf{D}_{\hat{T}^k}^\dagger)\mathbf{x}\right\|_2 \\ &\leq \left\|(\mathbf{I} - \mathbf{D}_{\hat{T}^k}\mathbf{D}_{\hat{T}^k}^\dagger)\mathbf{x}\right\|_2,\end{aligned}$$

where the last inequality is due to the fact that $\mathbf{D}_{\hat{T}^k}(\mathbf{M}\mathbf{D}_{\hat{T}^k})^\dagger\mathbf{M} - \mathbf{I}$ is a projection operator and thus its operator norm is smaller or equal to 1. Splitting $\mathbf{x}$ into $\tilde{\mathbf{x}}$

and $\mathbf{x}-\tilde{\mathbf{x}}$, and then using the triangle inequality and the fact that $\mathbf{I}-\mathbf{D}_{\hat{T}^k}\mathbf{D}^{\dagger}_{\hat{T}^k}$ is a projection with (19) give

$$\|\hat{\mathbf{x}}_{\text{OMP}}-\mathbf{x}\|_2 \leq \left\|(\mathbf{I}-\mathbf{D}_{\hat{T}^k}\mathbf{D}^{\dagger}_{\hat{T}^k})\tilde{\mathbf{x}}\right\|_2 + \|\mathbf{x}-\tilde{\mathbf{x}}\|_2 . \quad (20)$$

By Lemma 4.3, after k iterations (8) holds. Thus, Lemma 4.1 implies the existence of a vector $\hat{\mathbf{z}}^k$, with a representation supported on $\hat{T}^k$, satisfying $\left\|\tilde{\mathbf{x}}-\hat{\mathbf{z}}^k\right\|_2 \leq \left\|\mathbf{W}_{\mathbf{D}_{\tilde{T}}}\tilde{\boldsymbol{\alpha}}\right\|_1 \epsilon$. This and projection properties yield for the first element in the rhs

$$\left\|(\mathbf{I}-\mathbf{D}_{\hat{T}^k}\mathbf{D}^{\dagger}_{\hat{T}^k})\tilde{\mathbf{x}}\right\|_2 \leq \left\|\tilde{\mathbf{x}}-\hat{\mathbf{z}}^k\right\|_2 \leq \left\|\mathbf{W}_{\mathbf{D}_{\tilde{T}}}\tilde{\boldsymbol{\alpha}}\right\|_1 \epsilon. \quad (21)$$

For the second element we have using Lemma 4.1

$$\|\mathbf{x}-\tilde{\mathbf{x}}\|_2 \leq \left\|\mathbf{W}_{\mathbf{D}_{T\setminus\tilde{T}}}\boldsymbol{\alpha}_{T\setminus\tilde{T}}\right\|_1 \epsilon. \quad (22)$$

Plugging (22) and (21) in (20) results with (6). Notice that if T is an ϵ-independent set then $T=\tilde{T}$ and (7) follows immediately from (6) because the first term in its rhs vanishes and in the second one $\mathbf{W}_{\mathbf{D}_T}\boldsymbol{\alpha}_T = \mathbf{W}_{\mathbf{D}}\boldsymbol{\alpha}$ since $\boldsymbol{\alpha}_{T^C}=0$. □

Remark 4.4: Theorem 4.2 can be easily extended to the noisy case using the proof technique in [9].

Remark 4.5: If for a certain vector $\mathbf{x}$ supported on T, we get $\left|\check{T}^k\right| \leq d$ then the condition in (5) in Theorem 4.2 implies a perfect recovery by using a simple twist in $\text{OMP}_{\epsilon,2}$, setting $\check{\mathbf{x}}_{\text{OMP}_{\epsilon,2}} = \mathbf{D}_{\check{T}^k}(\mathbf{M}\mathbf{D}_{\check{T}^k})^{\dagger}\mathbf{y}$. Due to uniqueness conditions, in this case $\check{\mathbf{x}}_{\text{OMP}_{\epsilon,2}} = \mathbf{x}$. It can be easily shown that $|clos_{2\epsilon,2}(T)| \leq d$ is a sufficient condition for this to happen.

Remark 4.6: From the previous remark we conclude that if for any T such that $|T| \leq k$ we have $|clos_{2\epsilon,2}(T)| \leq d$ then the algorithm provides us always with a perfect recovery.

Remark 4.7: Theorem 4.2 applies also to the regular OMP if $\frac{|\langle \mathbf{d}_i,\mathbf{d}_j\rangle|^2}{\|\mathbf{d}_i\|_2^2\|\mathbf{d}_j\|_2^2} < 1-\epsilon^2$ implies $|\langle \mathbf{M}\mathbf{d}_i,\mathbf{M}\mathbf{d}_j\rangle|^2 < 1-\epsilon^2$. The latter property guarantees that in the step of selecting a new element, OMP does not choose an index from $\check{T}^t$. For a formal proof, the induction step in Lemma 4.3 needs to be modified showing that an element from $\check{T}^t$ is not chosen.

V. Numerical Simulation

We turn to check numerically the recovery performance of OMP and $\text{OMP}_{\epsilon,2}$ for sparse signals under a dictionary that contains pairs of correlated columns. We generate a dictionary $\mathbf{D}=[\mathbf{D}^1,\mathbf{D}^2]$ where $\mathbf{D}^1,\mathbf{D}^2 \in \mathbb{R}^{d\times d}$, $d=1000$, $\mathbf{D}^1$ contains sparse columns with 2 non-zero entries which are ± 1 with probability 0.5 like in [7] and $\mathbf{D}^2$ is constructed such that each of its columns $\mathbf{d}_i^2$ is ϵ-correlated to the corresponding column $\mathbf{d}_i^1$. Each entry of the measurement matrix $\mathbf{M} \in \mathbb{R}^{m\times d}$ is distributed according to a normal Gaussian distribution, where $m{=}\lfloor \gamma d \rfloor$ and γ is the sampling rate – a value in the range $(0,1]$. We set k to be $\lfloor \rho m \rfloor$ ($\rho \ll 1$) and measure the recovery rate of the representation $\boldsymbol{\alpha}$ and the signal $\mathbf{x}$ for various values of $\gamma \in \{0.1, 0.2, \ldots, 0.9\}$ and $\rho \in \{0.02, 0.04, \ldots, 0.2\}$.

Figure 1 presents the recovery performance over 100 realizations per each parameter setting. We use the observation in Remark 4.5 and present the recovery rate of $\text{OMP}_{\epsilon,2}$ for both $\hat{\mathbf{x}}_{\text{OMP}_{\epsilon,2}}$ and $\check{\mathbf{x}}_{\text{OMP}_{\epsilon,2}}$. As expected from Theorem 4.2, for the first we do not get a perfect recovery but only an error

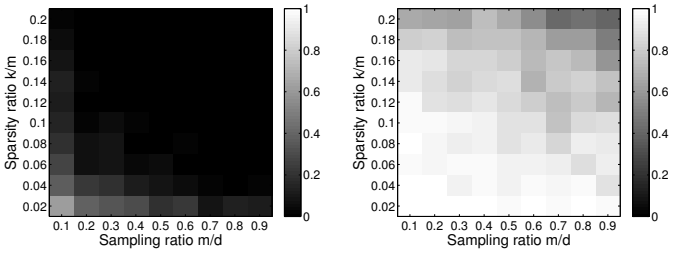

Fig. 1. $\text{OMP}_{\epsilon,2}$ recovery rate for $\hat{\mathbf{x}}_{\text{OMP}_{\epsilon,2}}$ (left) and $\check{\mathbf{x}}_{\text{OMP}_{\epsilon,2}}$ (right) for the synthetic experiment described in Section V. Color attribute: fraction of realizations in which a perfect recovery is achieved.

of an order of ϵ (due to lack of space we do not present the recovery error). However, as observed in Remark 4.5 when we take an ϵ-closure on the achieved support we get an almost perfect recovery. As high correlations between columns in $\mathbf{D}$, indeed imply high correlations between columns in $\mathbf{MD}$ in the common case, the recovery performance we present for $\text{OMP}_{\epsilon,2}$ are the same as for OMP as predicted in Remark 4.7. This provides a partial explanation for the reason that OMP achieves recovery in the experiments in [6].

VI. Conclusion

In this paper we have proposed a variant of the OMP algorithm – the ϵ-OMP ($\text{OMP}_{\epsilon,2}$) – for recovering signals with sparse representations under dictionaries with pairs of highly correlated columns. We have shown, both theoretically and empirically, that $\text{OMP}_{\epsilon,2}$ succeeds in recovering such signals and that the same holds for OMP. These results are a first step for explaining its success for coherent dictionaries.

Acknowledgment

R. Giryes thanks the Azrieli Foundation for the Azrieli Fellowship. This research was supported by European Community's FP7- ERC program, grant agreement no. 320649.

References

[1] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Appl. Comput. Harmon. Anal*, vol. 31, no. 1, pp. 59 – 73, 2011.

[2] S. Nam, M. Davies, M. Elad, and R. Gribonval, "The cosparse analysis model and algorithms," *Appl. Comput. Harmon. Anal.*, vol. 34, no. 1, pp. 30 – 56, 2013.

[3] T. Peleg and M. Elad, "Performance guarantees of the thresholding algorithm for the cosparse analysis model," *IEEE Trans. on Information Theory*, vol. 59, no. 3, pp. 1832–1845, Mar. 2013.

[4] R. Giryes, S. Nam, M. Elad, R. Gribonval, and M. E. Davies, "Greedy-like algorithms for the cosparse analysis model," *to appear in the Special Issue in Linear Algebra and its Applications on Sparse Approximate Solution of Linear Systems*, 2013.

[5] R. Rubinstein, T. Peleg, and M. Elad, "Analysis K-SVD: A dictionary-learning algorithm for the analysis sparse model," *IEEE Trans. on Signal Processing*, vol. 61, no. 3, pp. 661–677, Feb. 2013.

[6] M. A. Davenport, D. Needell, and M. B. Wakin, "Signal space CoSaMP for sparse recovery with redundant dictionaries," *CoRR*, vol. abs/1208.0353, 2012.

[7] R. Giryes and M. Elad, "Can we allow linear dependencies in the dictionary in the synthesis framework?" in *ICASSP 2013*.

[8] S. Chen, S. A. Billings, and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification," *International Journal of Control*, vol. 50, no. 5, pp. 1873–1896, 1989.

[9] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.