

# OMP AND CONTINUOUS DICTIONARIES: IS $k$ -STEP RECOVERY POSSIBLE?

Clément Elvira<sup>†</sup>, Rémi Gribonval<sup>†</sup>, Charles Soussen<sup>‡</sup> and Cédric Herzet<sup>†</sup>

<sup>†</sup> Inria, Université de Rennes, CNRS, IRISA, F-35000 Rennes, France

<sup>‡</sup> L2S, CentraleSupélec-CNRS-Université Paris-Saclay, 91192 Gif-sur-Yvette, France

## ABSTRACT

In this work, we present new theoretical results on sparse recovery guarantees for a greedy algorithm, orthogonal matching pursuit (OMP), in the context of continuous parametric dictionaries, *i.e.*, made up of an infinite uncountable number of atoms. We build up a family of dictionaries for which  $k$ -step recovery is possible with OMP for 1-dimensional parameters. In higher dimension, algebraic conditions become necessary and will lead us to revisit some well-known  $k$ -step discrete analyses. Finally, a toy-example illustrates the level of tightness of our sufficient conditions.

**Index Terms**— Sparse representation, continuous dictionaries, orthogonal matching pursuit, exact recovery.

## 1. INTRODUCTION

Over the last decade, sparse representations have sparked a surge of interest in both signal processing and statistics communities. It consists in finding a “simple” representation of a signal  $\mathbf{y}$  living in a Hilbert space  $\mathcal{H}$  as a linear combination of a few unit-normed elements from a dictionary  $\mathcal{A} \subset \mathcal{H}$ , that is

$$\mathbf{y} = \sum_{\ell=1}^k c_{\ell} \mathbf{a}_{\ell} \quad \text{where } c_{\ell} \in \mathbb{R}, \mathbf{a}_{\ell} \in \mathcal{A}. \quad (1)$$

Here, the notion of sparsity refers to the fact that few elements are used in the decomposition of  $\mathbf{y}$  compared to the ambient dimension. In the sequel, we restrict our attention to *parametric dictionaries* defined as

$$\mathcal{A} = \{\mathbf{a}(\theta) \mid \theta \in \Theta\} \quad (2)$$

where  $\Theta$  is the parameter set and  $\mathbf{a}$  a continuous mapping.

Sparse representations have proven to be of great interest in many application fields, leading to, *e.g.*, the famous compression standards JPEG or MP3 [1]. Although this problem has been shown to be NP-hard [2], numerous practical procedures, along with their theoretical analyses, have been proposed in the literature. However, most contributions address the sparse-representation problem in the “discrete setting”, *i.e.*, where the dictionary contains a finite number of elements, see [3]. Among the most popular, one may mention methods

based on convex relaxation, *e.g.*, *Basis Pursuit* (BP) [4] or the *Lasso* [5], nonconvex minimization [6, 7] and greedy algorithms. The latter are procedures that iteratively increment the support of the estimated signal starting from the empty support, by selecting one new atom per iteration. In this paper, we focus on *Orthogonal Matching Pursuit* (OMP) [8] described in Alg. 1.

A question of broad interest is the  $k$ -step recovery of sparse supports. Given a finite subset  $\mathcal{S}^* \triangleq \{\theta_{\ell}^*\}_{\ell=1}^k \subset \Theta$ ,<sup>1</sup> is it possible to recover  $\mathcal{S}^*$  in  $k$ -steps from any observation  $\mathbf{y}$  that is a linear combination of atoms  $\mathbf{a}(\theta_{\ell}^*)$  for  $\theta^* \in \mathcal{S}^*$ ? In the discrete setting, the first thorough analysis of the so-called *uniform recovery* (irrespective of the weights  $(c_{\ell})_{\ell=1}^k$  of the atoms) of a support  $\mathcal{S}^*$  by OMP is due to Tropp [9]. Using our formalism, he showed that OMP recovers  $\mathcal{S}^*$  from any combination of linearly independent atoms  $\{\mathbf{a}(\theta_{\ell}^*)\}_{\ell=1}^k$  if and only if

$$\forall \theta \in \Theta \setminus \mathcal{S}^*, \quad \|\mathbf{G}^{-1} \mathbf{g}_{\theta}\|_1 < 1, \quad (3 - \text{ERC})$$

where  $\mathbf{G} \in \mathbb{R}^{k \times k}$  and  $\mathbf{g}_{\theta} \in \mathbb{R}^k$  are defined by  $\mathbf{G}[i, j] = \langle \mathbf{a}(\theta_i^*), \mathbf{a}(\theta_j^*) \rangle$  and  $\mathbf{g}_{\theta}[i] = \langle \mathbf{a}(\theta), \mathbf{a}(\theta_i^*) \rangle_{\mathcal{H}}$  for any  $\theta$ . Condition (3 – ERC) is called *exact recovery condition* (ERC) and can be computationally tested when  $\Theta$  is a finite set. Interestingly, (3 – ERC) is also connected to the correct identification of  $\mathcal{S}^*$  by other greedy procedures [10, 11] and by convex relaxation [12, 13, 14]. However, condition (3 – ERC) is of limited practical interest since it requires to know the support beforehand. To circumvent this issue, weaker but easier-to-evaluate sufficient conditions of success have been developed. One of the most popular is based on the *mutual coherence*  $\mu$  of the dictionary and imposes

$$k < \frac{1}{2}(1 + \mu^{-1}) \quad \text{where } \mu = \sup_{\theta \neq \theta' \in \Theta} |\langle \mathbf{a}(\theta), \mathbf{a}(\theta') \rangle_{\mathcal{H}}|. \quad (4)$$

In the last few years, several works have tackled the problem of sparse representation in *continuous* dictionaries, *i.e.*, where  $\mathcal{A}$  is made up of an *uncountable* number of atoms [15, 16, 17]. It was shown that a continuous version of the Lasso can be expressed as a convex optimization problem over the space of Radon measures, called the Beurling Lasso [18, 19]

<sup>1</sup>Note that  $\mathcal{S}^*$  usually refers to a set of indices in the discrete setting.

---

**Algorithm 1:** Orthogonal Matching Pursuit (OMP)

---

**Input:** Observation  $\mathbf{y}$ , atoms  $\mathbf{a}(\cdot)$  and sparsity  $k$ .

```
1 for  $t \leftarrow 1$  to  $k$  do
2    $\mathbf{r} = \mathbf{y} - \sum_{\ell=1}^{t-1} \hat{c}_\ell \mathbf{a}(\hat{\theta}_\ell)$ ;
3    $\hat{\theta}_t \in \arg \max_{\theta \in \Theta} |\langle \mathbf{a}(\theta), \mathbf{r} \rangle_{\mathcal{H}}|$ ;
4    $\hat{c}_1, \dots, \hat{c}_t = \arg \min_{(c_1, \dots, c_t) \in \mathbb{R}^t} \left\| \mathbf{y} - \sum_{\ell=1}^t c_\ell \mathbf{a}(\hat{\theta}_\ell) \right\|_{\mathcal{H}}$ ;
5 end
Output:  $\hat{\theta}_1, \dots, \hat{\theta}_k$  and  $\hat{c}_1, \dots, \hat{c}_k$ 
```

---

while a continuous version of BP was proposed in [16]. However, most conditions resulting from classical recovery analyses are non-realistic for thin grids on  $\Theta$  since: i) condition (3 – ERC) becomes difficult to assess; ii) the coherence  $\mu$  tends to 1 so (4) becomes  $k = 1$ . For now, most contributions related to support recovery were done on the side of convex-based approaches. More particularly, for some families of continuous dictionaries, support recovery holds for BP and Beurling Lasso whenever  $\mathbf{y}$  is a positive linear combination of atoms or the parameters  $\theta_\ell^*$  are sufficiently separated [16, 17, 19].

In this work, we study the recovery properties of OMP in the context of continuous dictionaries. OMP is seen here as a generic algorithm that holds both in the discrete and continuous settings. In particular, we leverage the expression of (3 – ERC) to rephrase  $k$ -step recovery as a property of the kernel induced by the inner product between atoms. We then build up a class of dictionaries for which  $k$ -step support recovery is possible under some assumptions.

The remaining of the paper is organized as follows. Section 2 elaborates on the “continuous” dictionary setup and the notion of support recovery. Section 3 presents our results. Finally, a numerical evaluation of the method is presented in Section 4 and concluding remarks are given.

## 2. CONTINUOUS DICTIONARIES AND SUPPORT RECOVERY

### 2.1. Continuous dictionaries

The parameter set  $\Theta$  is assumed to contain an infinite uncountable number of elements (e.g., a hyper-rectangle of  $\mathbb{R}^D$ ). However, we restrict our attention to the case where  $\Theta = \mathbb{R}^D$ . A second ingredient grounding the paradigm of continuous dictionaries is the continuity of the function  $\mathbf{a} : \Theta \rightarrow \mathcal{H}$ , that is:  $\forall \theta \in \Theta, \lim_{\theta' \rightarrow \theta} \|\mathbf{a}(\theta') - \mathbf{a}(\theta)\|_{\mathcal{H}} = 0$ . In this paper, we moreover assume that the atoms of the dictionary have a unit norm, i.e.,  $\|\mathbf{a}(\theta)\|_{\mathcal{H}} = 1$ .

In the sequel, our recovery conditions will be expressed as a function of the symmetric kernel  $\kappa(\theta, \theta')$  induced by the inner product between atoms, that is:

$$\kappa(\theta, \theta') \triangleq \langle \mathbf{a}(\theta), \mathbf{a}(\theta') \rangle_{\mathcal{H}} \quad \forall \theta, \theta' \in \Theta. \quad (5)$$

The *continuity* and *unit-norm* properties imply

$$\begin{aligned} \text{“unit norm”} : \quad & \kappa(\theta, \theta) = 1 \quad \forall \theta \in \Theta, \\ \text{“continuity”} : \quad & \lim_{\theta' \rightarrow \theta} \kappa(\theta, \theta') = 1 \quad \forall \theta \in \Theta. \end{aligned} \quad (6)$$

Note that combined with the Cauchy-Schwartz inequality, the unit norm assumption (6) gives  $|\kappa(\theta, \theta')| \leq 1$ . Lastly, we restrict our attention to evanescent kernels, i.e., such that for all  $\varepsilon > 0$  and  $\theta \in \Theta$ ,

$$\exists K_\varepsilon \subset \Theta \text{ compact} : \sup_{\theta' \in \Theta \setminus K_\varepsilon} |\kappa(\theta, \theta')| < \varepsilon. \quad (7)$$

Condition (7) guarantees that the maximizer in line 3 of OMP exists at each iteration, i.e., the supremum is attained.

### 2.2. Support recovery

We now define the notion of *uniform  $k$ -step* recovery. Let  $\mathcal{S}^* \triangleq \{\theta_\ell^*\}_{\ell=1}^k$  be a subset of  $\Theta$ . We say that  $\mathbf{y}$  is supported in  $\mathcal{S}^*$  if  $\mathbf{y}$  obeys (1) with  $\forall \ell \in \{1, \dots, k\}$ ,  $\mathbf{a}_\ell = \mathbf{a}(\theta_\ell^*)$  and  $c_\ell \in \mathbb{R}^*$ . When  $\mathbf{y}$  is supported in  $\mathcal{S}^*$  and  $\text{card}(\mathcal{S}^*) = k$ , we also say that  $\mathbf{y}$  is “ $k$ -sparse”. Given  $\mathbf{y}$ , we say that OMP achieves “ $k$ -step” recovery if and only if

$$\arg \max_{\theta \in \Theta} |\langle \mathbf{a}(\theta), \mathbf{r} \rangle_{\mathcal{H}}| \subseteq \mathcal{S}^*, \quad (8)$$

during the first  $k$  iterations, where  $\mathbf{r} \in \mathcal{H}$  refers to the residual computed by OMP (see Alg. 1, line 5). The right-hand side entails that any maximizer belongs to the support  $\mathcal{S}^*$ , i.e.,

$$\max_{\theta \in \mathcal{S}^*} |\langle \mathbf{a}(\theta), \mathbf{r} \rangle_{\mathcal{H}}| > |\langle \mathbf{a}(\theta'), \mathbf{r} \rangle_{\mathcal{H}}| \quad \forall \theta' \in \Theta \setminus \mathcal{S}^*. \quad (9)$$

In particular, we consider that the situation where equality occurs in (9) for some  $\theta' \notin \mathcal{S}^*$  leads to a failure.

We say that OMP achieves “uniform  $k$ -step recovery” if condition (8) is fulfilled for any observation  $\mathbf{y}$  supported in  $\mathcal{S}^*$ . “Uniform” refers to the fact that exact recovery is achieved for any choice of the nonzero coefficients  $c_1 \dots c_k$ .

## 3. UNIFORM RECOVERY FOR CONTINUOUS DICTIONARIES

In this section, we present a class of dictionaries for which uniform recovery is achieved, see Secs. 3.1 and 3.2. Sketches of proof are available in App. A.

Let  $\mathcal{H} = L_2(\mathbb{R}^D)$ ,  $\lambda > 0$  and consider the following dictionary  $\mathcal{A}$  made of functional atoms  $\mathbf{a}(\theta)$  given by

$$\mathbf{a}(\theta) : t \in \mathbb{R}^D \mapsto (2\lambda)^{\frac{D}{2}} e^{-\lambda \mathbf{1}_D^\top (t - \theta)} \mathbb{I}_{\{t \geq \theta\}}, \quad (10)$$

$\forall \theta \in \mathbb{R}^D$  where  $\mathbf{1}_D = (1 \dots 1)^\top \in \mathbb{R}^D$  and  $\mathbb{I}$  equals 1 if all component-wise inequalities hold and 0 otherwise. Straight-forward calculations show that the inner product in  $L_2(\mathbb{R}^D)$  between two atoms writes  $\langle \mathbf{a}(\theta), \mathbf{a}(\theta') \rangle_{\mathcal{H}} = e^{-\lambda \|\theta - \theta'\|_1}$  and corresponds to the so-called “Laplace kernel”.

### 3.1. Uniform recovery of $k$ -sparse supports

The next result shows that in the 1-dimensional case,  $k$ -step recovery is possible for all  $k$  with the continuous dictionary defined in (10):

**Theorem 1.** Let  $\Theta = \mathbb{R}$  and define  $\mathcal{A}$  as in (10). Then OMP uniformly recovers any support  $\mathcal{S}^* = \{\theta_\ell^*\}_{\ell=1}^k$  made of  $k$  distinct parameters in  $k$  iterations.

The novelty of Th. 1 is the following: although the parameter space is a continuum,  $k$ -step recovery is always achieved whatever the support. Surprisingly, such a dictionary does not require separation between parameters, as it is usually the case for signed combination of atoms [16, 17].

When  $\Theta \subset \mathbb{R}^D$  with  $D \geq 2$ , such a result does not hold, see for instance Sec. 4. However, inspired by the results proved in [16, 17], one ensures  $k$ -step recovery by imposing a separation condition between parameters. Define

$$\Delta \triangleq \min_{\substack{d=1 \dots D \\ i \neq j}} |\theta_j^*[d] - \theta_i^*[d]| \quad \text{s.t.} \quad \theta_i^*[d] \neq \theta_j^*[d], \quad (11)$$

as the smallest distance between non equal coordinates. When  $\Delta$  is too small, one may find a configuration of parameters such that OMP fails at the first iteration, see Sec. 4 for an illustration. Our next result circumvents this issue by imposing a minimum separation conditions that ensures the recovery of any  $k$ -sparse support:

**Theorem 2.** Let  $\Theta = \mathbb{R}^D$  with  $D \geq 2$  and define  $\mathcal{A}$  as in (10). OMP uniformly recovers any  $k$ -sparse support in  $k$  steps if

$$\Delta > \lambda^{-1} \log(2k - 1), \quad (12)$$

or, equivalently,

$$k < \frac{1}{2}(1 + e^{\lambda\Delta}). \quad (13)$$

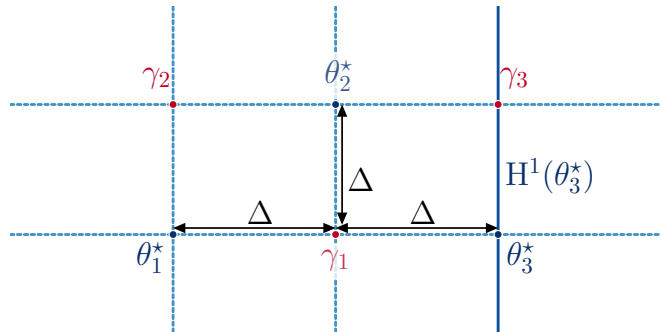
Th. 2 states that OMP uniformly recovers any combination of sufficiently separated atoms. One recognizes in (13) the flavor of mutual coherence conditions (e.g., (4)). The main difference with the discrete setting is that the coherence condition now concerns only a small part of the dictionary *via*  $\Delta$ , namely the atoms related to the support of the signal.

### 3.2. Uniform recovery of a given support

Although almost tight for some configurations of parameters, the separation condition given in Th. 2 is only sufficient. We now show that, given a  $k$ -sparse support, uniform recovery can be characterized under some “ERC-like” finite-dimensional condition (see Th. 3 below). We introduce first some additional notations. For  $d = 1 \dots D$ , let

$$\mathbf{H}^d(\theta) \triangleq \{\theta' \in \mathbb{R}^D \mid \theta'[d] = \theta[d]\}, \quad (14)$$

be the  $(D-1)$ -dimensional parametric affine hyperplane passing through the  $d$ th component of  $\theta \in \Theta$ . If  $\mathbf{y}$  is supported



**Fig. 1.** Illustration of the definitions of  $H^d(\theta_\ell^*)$  and the superset  $\Gamma$  see (15)) with  $D = 2$  and  $k = 3$ .  $\Gamma((\theta_\ell^*)_{\ell=1}^3)$  is composed of the  $\theta_\ell^*$ 's and 3 additional points denoted  $\gamma_\ell$ .

by  $\mathcal{S}^*$ , the hyperplanes  $H^d(\theta^*)$  for  $\theta^* \in \mathcal{S}^*$  are related to the non-differentiability of the function  $\theta \mapsto \langle a(\theta), r \rangle_{\mathcal{H}}$  (see line 3 in Alg. 1). We also define the set operator

$$\Gamma: \mathfrak{P}(\Theta) \longrightarrow \mathfrak{P}(\Theta)$$

$$\mathcal{S} \longmapsto \bigcap_{d=1}^D \left( \bigcup_{\theta \in \mathcal{S}} \mathbf{H}^d(\theta) \right) \quad (15)$$

where  $\mathfrak{P}(\Theta)$  is the set of all subsets of  $\Theta$ . One immediately sees that  $\mathcal{S} \subseteq \Gamma(\mathcal{S})$  for any  $\mathcal{S} \in \mathfrak{P}(\Theta)$ . Fig. 1 illustrates these definitions for  $\mathcal{S}^* = \{\theta_1^*, \theta_2^*, \theta_3^*\}$ . The set  $\Gamma(\mathcal{S}^*)$  will be used to describe the set of potential maximizers in line 3 of Alg. 1. We provide the following lemma (see [20] for a proof) and then state our main result.

**Lemma 1.** *If  $\text{card}(\mathcal{S}^*) < +\infty$  then  $\text{card}(\Gamma(\mathcal{S}^*)) < +\infty$ .*

**Theorem 3.** Let  $\Theta = \mathbb{R}^D$ ,  $\mathcal{A}$  be defined as in (10) and let  $\mathcal{S}^* = \{\theta_\ell^*\}_{\ell=1}^k$  be a set of  $k$  distinct parameters. Then OMP uniformly recovers  $\mathcal{S}^*$  in  $k$  steps if and only if

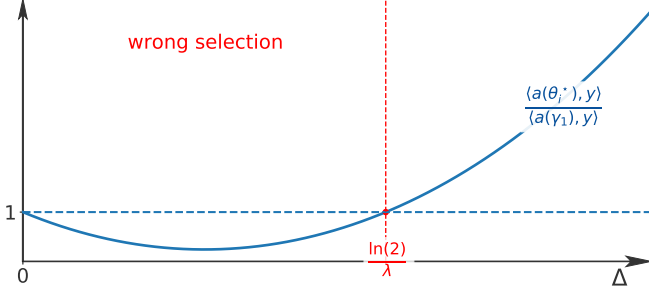
$$\max_{\theta \in \Gamma(\mathcal{S}^*) \setminus \mathcal{S}^*} \|\mathbf{G}^{-1} \mathbf{g}_\theta\|_1 < 1. \quad (16 - \text{ERC})$$

*In particular, if (16 – ERC) does not hold, there exists (at least) one linear combination  $\mathbf{y} = \sum_{\ell=1}^k c_{\ell} \mathbf{a}(\theta_{\ell}^*)$  such that OMP with  $\mathbf{y}$  as input, fails at the first iteration.*

Combined with Lem. 1, Th. 3 states that  $k$ -step recovery is achieved provided that a condition on a finite number of atoms is fulfilled. Indeed, Eq. (16 – ERC) only depends on  $\Gamma(\mathcal{S}^*)$  which is a *finite* subset of  $\Theta$ . Consequently, similarly to the discrete case, its numerical evaluation is possible.

## 4. NUMERICAL ILLUSTRATION

**Tightness of (13).** Although not stated in Th. 3, the rationale behind our result is that the local maximizers of the inner product at each iteration (see line 3 in Alg. 1) are contained in  $\Gamma(S^*)$ . We now leverage this remark to deduce a necessary



**Fig. 2.** Ratio of inner products (see (17)) as a function of  $\Delta$ , the distance between parameters.

condition for  $k$ -step recovery by evaluating inner products in well chosen points. Let  $\Delta > 0$ . We consider the following configuration of points in  $\Theta = \mathbb{R}^2$ :  $\theta_1^* = (-\Delta, 0)^T$ ,  $\theta_2^* = (0, \Delta)^T$  and  $\theta_3^* = (\Delta, 0)^T$  and let  $\gamma_1 = (0, 0)^T$ . By construction,  $k = 3$ , one has  $\|\theta_i^* - \theta_j^*\|_1 = 2\Delta$  and  $\|\theta_i^* - \gamma_1\|_1 = \Delta$  for all  $i \neq j$ . This configuration of points corresponds to Fig. 1 and  $\Delta$  has been defined in (11). Let  $\mathbf{r} = \mathbf{a}(\theta_1^*) + \mathbf{a}(\theta_2^*) + \mathbf{a}(\theta_3^*)$ , be composed of three equally weighted atoms. We now study the ratio

$$\frac{\langle \mathbf{a}(\theta_i^*), \mathbf{r} \rangle_{\mathcal{H}}}{\langle \mathbf{a}(\gamma_1), \mathbf{r} \rangle_{\mathcal{H}}} \quad i = 1, 2, 3. \quad (17)$$

According to (9), OMP fails at the first iteration when (17) is lower than 1 and uniform recovery will not be achieved.

Fig. 2 plots the ratio (17) seen as a function of  $\Delta$ , the distance between parameters. When  $\Delta \leq \lambda^{-1} \log(2)$ , one sees that OMP fails at the first iteration so  $\Delta > \lambda^{-1} \log(2)$  is a *necessary condition* here. Consequently, the sufficient condition  $\Delta > \lambda^{-1} \log(5)$  given in Th. 2 is not far from being tight. This observation is also confirmed by the numerical verification of (16 – ERC) which shows here that the latter condition is fulfilled as soon as  $\Delta > \lambda^{-1} \log(2)$ .

**Discussion.** An enhanced implementation of a continuous version of OMP has already been carried out in the context of sketching [21]. The main difference with the discrete setting is that the selection of  $\hat{\theta}_i$  at each iteration is now done up to the precision of the numerical method used to maximize the inner product (see line 3 in Alg. 1). Therefore,  $k$ -step recovery is achieved up to this precision. Although this small error had no impact during our numerical investigations, future works include a theoretical study of the behavior of OMP in this scenario.

## 5. CONCLUSION

In this paper, we have studied the recovery properties of OMP in the context of parametric continuous dictionaries. We have built up a family of dictionaries, for which OMP achieves  $k$ -step recovery for a well designed continuous dictionary.

Our recovery results vary with the dimension of the parameter set. In particular, OMP recovers any  $k$ -sparse support in  $k$ -step in dimension 1 while an algebraic condition is necessary in higher dimensions. We leveraged this result to build a sufficient condition based on minimum separation between parameters. Numerical experiments on toy examples have shown that this condition is almost tight. Fortunately, all the results and the analyses presented in this paper are not limited to dictionaries given by (10). Although not detailed here, these results apply for a broader class of dictionaries that relies on completely monotone functions. In [20], we propose a broader class of kernels for which  $k$ -step recovery is achieved.

## 6. ACKNOWLEDGMENTS

Part of this work has been funded thanks to the Becose ANR project no. ANR-15-CE23-0021.

## A. PROOF SKETCHES

*Proof of Th. 3.* Let  $\mathcal{S}^*$  be  $k$ -sparse and  $\mathbf{y} = \sum_{\ell=1}^k c_{\ell} \mathbf{a}(\theta_{\ell}^*)$  with for all  $\ell$ ,  $c_{\ell} \in \mathbb{R}^*$ . The proof relies on the study of

$$f: \mathbb{R}^D \rightarrow \mathbb{R} \\ \theta \mapsto \sum_{\ell=1}^k c_{\ell} e^{-\lambda \|\theta_{\ell}^* - \theta\|_1}, \quad (18)$$

such that  $f(\theta) = \langle \mathbf{a}(\theta), \mathbf{y} \rangle_{\mathcal{H}}$  for all  $\theta \in \mathbb{R}^D$ . The proof relies on the following observation: for all  $\theta \in \mathbb{R}^D \setminus \bigcup_{\ell,d} \mathcal{H}^d(\theta_{\ell}^*)$ ,  $f$  is twice differentiable and its Hessian matrix  $\mathbf{H}_f$  verifies

$$\text{trace}(\mathbf{H}_f(\theta)) = D\lambda^2 f(\theta). \quad (19)$$

The complete proof is done by induction on  $D$  where one shows that whatever  $c_1 \dots c_k$ , the maximizers of  $f$  belong to  $\Gamma(\mathcal{S}^*)$ . We describe below the rationale of the induction.

a) Recall first that the trace of a symmetric matrix is the sum of its eigenvalues. Then, if  $f$  is twice differentiable in  $\theta_0$  and  $f(\theta_0) > 0$  then  $\theta_0$  cannot be a maximizer of  $f$  since  $\text{trace}(\mathbf{H}_f(\theta_0)) > 0$ . In a symmetric fashion,  $f(\theta_0)$  cannot be a minimizer of  $f$  if  $f(\theta_0) < 0$ . Therefore, if  $\theta_0$  is a maximizer of  $|f|$  then  $f$  is not differentiable in  $\theta_0$ , i.e.,  $\theta_0 \in \bigcup_{\ell,d} \mathcal{H}^d(\theta_{\ell}^*)$ . b) Given  $d \leq D$  and  $\ell \leq k$ , we now study the maximizers of  $f$  restricted to the  $(D-1)$ -dimensional hyperplane  $\mathcal{H}^d(\theta_{\ell}^*)$ . The latter function can be interpreted as a function of the form (18) defined on  $\mathbb{R}^{D-1}$  which is differentiable on  $\mathcal{H}^d(\theta_{\ell}^*) \setminus \left( \bigcup_{d' \neq d} \bigcup_{\ell'=1}^k \mathcal{H}^{d'}(\theta_{\ell'}^*) \right)$ . The recursion hypothesis applies so its maximizers belong to  $\bigcap_{d' \neq d} \left( \bigcup_{\ell'=1}^k \mathcal{H}^{d'}(\theta_{\ell'}^*) \right)$ . By repeating this reasoning for all  $\ell, d$ , we conclude that the maximizers of  $f$  belong to  $\bigcup_{d=1}^D \bigcap_{d' \neq d} \left( \bigcup_{\ell'=1}^k \mathcal{H}^{d'}(\theta_{\ell'}^*) \right) = \Gamma(\mathcal{S}^*)$  by set operations.  $\square$

*Other proofs.* Th. 1 is a corollary of Th. 3 by seeing that  $\Gamma(\mathcal{S}^*) = \mathcal{S}^*$  when  $\Theta = \mathbb{R}$  since all hyperplanes reduce to singletons. Similarly, Th. 2 is obtained using classical mutual coherence analysis applied to the finite dictionary  $\{\mathbf{a}(\theta) \mid \theta \in \Gamma(\mathcal{S}^*)\}$  that appears in Th. 3, see e.g., [9].  $\square$

## REFERENCES

- [1] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok, "Introduction to compressed sensing," in *Compressed Sensing: Theory and Applications*, chapter 1, pp. 1–64. Cambridge University Press, 2012.
- [2] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM Journal on Computing*, vol. 24, no. 2, pp. 227–234, 1995.
- [3] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Springer New York, 2013.
- [4] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.
- [5] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.
- [6] G. Gasso, A. Rakotomamonjy, and S. Canu, "Recovering sparse signals with a certain family of nonconvex penalties and DC programming," *IEEE Transactions on Signal Processing*, vol. 57, no. 12, pp. 4686–4698, Dec. 2009.
- [7] L. Chen and Y. Gu, "The convergence guarantees of a non-convex approach for sparse recovery," *IEEE Transactions on Signal Processing*, vol. 62, no. 15, pp. 3754–3767, Aug. 2014.
- [8] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, Nov. 1993, pp. 40–44 vol.1.
- [9] J. A. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
- [10] C. Soussen, R. Gribonval, J. Idier, and C. Herzet, "Joint  $k$ -Step Analysis of Orthogonal Matching Pursuit and Orthogonal Least Squares," *IEEE Transactions on Information Theory*, vol. 59, no. 5, pp. 3158–3174, May 2013.
- [11] R. Gribonval and P. Vandergheynst, "On the exponential convergence of matching pursuits in quasi-incoherent dictionaries," *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 255–261, Jan. 2006.
- [12] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Transactions on Information Theory*, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.
- [13] J.-J. Fuchs, "On sparse representations in arbitrary redundant bases," *IEEE Transactions on Information Theory*, vol. 50, no. 6, pp. 1341–1344, June 2004.
- [14] J. A. Tropp, "Just relax: convex programming methods for identifying sparse signals in noise," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 1030–1051, Mar. 2006.
- [15] C. Ekanadham, D. Tranchina, and E. P. Simoncelli, "Recovery of sparse translation-invariant signals with continuous basis pursuit," *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 4735–4744, Oct. 2011.
- [16] E. J. Candès and C. Fernandez-Granda, "Towards a Mathematical Theory of Super-resolution," *Communications on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, 2014.
- [17] V. Duval and G. Peyré, "Exact support recovery for sparse spikes deconvolution," *Found. Comput. Math.*, vol. 15, no. 5, pp. 1315–1355, Oct. 2015.
- [18] K. Bredies and H. K. Pikkariainen, "Inverse problems in spaces of measures," *ESAIM: Control, Optimisation and Calculus of Variations*, vol. 19, no. 1, pp. 190–218, 2012.
- [19] Y. de Castro and F. Gamboa, "Exact reconstruction using Beurling minimal extrapolation," *Journal of Mathematical Analysis and Applications*, vol. 395, no. 1, pp. 336 – 354, 2012.
- [20] C. Elvira, R. Gribonval, C. Herzet, and C. Soussen, "When does OMP achieves  $k$ -step recovery with continuous dictionaries?," *preprint*, 2018.
- [21] R. Gribonval, G. Blanchard, N. Keriven, and Y. Traonmilin, "Compressive Statistical Learning with Random Feature Moments," *preprint*, Dec. 2017.