When does OMP achieve support recovery with continuous dictionaries?

Clément Elvira, Rémi Gribonval, Charles Soussen, Cedric Herzet

When does OMP achieve support recovery with continuous dictionaries?

Clément Elvira[1], Rémi Gribonval[1], Charles Soussen[2] and Cédric Herzet[1]

[1] *Univ Rennes, Inria, CNRS, IRISA*
F-35000 Rennes, France
[2] *L2S, CentraleSupélec-CNRS-Université Paris-Saclay*
91192 Gif-sur-Yvette, France

Abstract

This paper presents new theoretical results on sparse recovery guarantees for a greedy algorithm, Orthogonal Matching Pursuit (OMP), in the context of continuous parametric dictionaries. Here, the continuous setting means that the dictionary is made up of an infinite (potentially uncountable) number of atoms. In this work, we rely on the Hilbert structure of the observation space to express our recovery results as a property of the kernel defined by the inner product between two atoms. Using a continuous extension of Tropp's Exact Recovery Condition, we identify two key notions of *admissible kernel* and *admissible support* that are sufficient to ensure exact recovery with OMP. We exhibit a family of admissible kernels relying on *completely monotone functions* for which admissibility holds for any support in the one-dimensional setting. For higher dimensional parameter spaces, an additional notion of *axis admissibility* is shown to be sufficient to ensure a form of "delayed" recovery. An additional algebraic condition involving a finite subset of (known) atoms further yields exact recovery guarantees. Finally, a coherence-based viewpoint on these results provides recovery guarantees in terms of a *minimum separation* assumption.

Keywords: sparse representation, continuous dictionaries, Orthogonal Matching Pursuit, exact recovery

1. Introduction

Sparse representation is a fundamental problem in signal processing. It consists in decomposing a signal $\boldsymbol{y}$ belonging to some vector space $\mathcal{H}$ as the linear combination of a few elements of some set $\mathcal{A} \subseteq \mathcal{H}$, that is

$$\boldsymbol{y} = \sum_{\ell=1}^{k} c_\ell \, \boldsymbol{a}_\ell \quad \text{where } c_\ell \in \mathbb{R}^*, \, \boldsymbol{a}_\ell \in \mathcal{A}. \tag{1.1}$$

Sparsity refers to the fact that the number of elements involved in the decomposition (1.1) should be much smaller than the ambient dimension of $\boldsymbol{y}$. The set $\mathcal{A}$ is commonly referred to as *dictionary* and its elements as *atoms*. In the sequel, we will assume that $\mathcal{A}$ is defined as:

$$\mathcal{A} = \{\boldsymbol{a}(\theta) \mid \theta \in \Theta\} \tag{1.2}$$

where $\Theta \subseteq \mathbb{R}^D$ and $\boldsymbol{a} : \Theta \to \mathcal{H}$ is some continuous function.

Over the past decade, sparse representations have proven to be of great interest in many applicative fields. As a consequence, numerous practical procedures, along with their theoretical analyses, have been proposed in the literature. Most contributions addressed the sparse-representation problem in the "*discrete*" setting, where the dictionary contains a *finite* number of elements, see [1]. Recently, several works tackled the problem of sparse representations in "*continuous*" dictionaries, where $\mathcal{A}$ is made up of an infinite *uncountable* number of atoms but $\boldsymbol{a} : \Theta \to \mathcal{H}$ enjoys some continuity property, see *e.g.*, [2–4]. We review the contributions most related to the present work in Section 2.

Before dwelling over the state of the art, we briefly describe the scope of our paper. In this work, we focus on the continuous setting and assume that $\mathcal{H}$ is a Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\|\cdot\|$. We derive "exact recovery" conditions for "Orthogonal Matching Pursuit" (OMP) [5], a natural adaptation to the continous setting of a popular greedy procedure of the literature (see Algorithm 1). The main question addressed in this paper is as follows: let $\{\theta^\star_\ell\}_{\ell=1}^k$ be k elements of Θ and assume $\boldsymbol{y}$ obeys (1.1) with $\boldsymbol{a}_\ell = \boldsymbol{a}(\theta^\star_\ell)$. Under which conditions can OMP with $\boldsymbol{y}$ as input correctly identify the set $\{\theta^\star_\ell\}_{\ell=1}^k$?

We note that, in the context of continuous dictionaries, the fact that OMP could correctly identify a set of k atoms may seem surprising in itself. Indeed, inspecting Algorithm 1, we see that this entails that OMP must identify one correct atom at *each* iteration of the algorithm, that is $\widehat{\theta}_t \in \{\theta^\star_\ell\}_{\ell=1}^k \; \forall t \in [\![1, k]\!]$. However, the following simple example suggests that such a requirement may never be met for continuous dictionaries:

Example 1 (The Gaussian deconvolution problem)**.** *Let* $\Theta = \mathbb{R}$ *and* $\mathcal{H} = L_2(\mathbb{R})$ *be the space of square integrable functions on* $\mathbb{R}$*. Assume* $\boldsymbol{a}(\cdot)$ *is defined as*

$$\begin{aligned} \boldsymbol{a} : \mathbb{R} &\longrightarrow L_2(\mathbb{R}) \\ \theta &\longmapsto \tfrac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}(\cdot - \theta)^2}. \end{aligned} \tag{1.3}$$

Algorithm 1: Orthogonal Matching Pursuit (OMP)

Input: Observation $\boldsymbol{y}$, family of atoms $\boldsymbol{a}(\cdot)$ and maximum number of iteration q.

1 $\boldsymbol{r} = \boldsymbol{y}$ `// residual vector`

2 **for** $t = 1$ **to** q **do**

3     $\widehat{\theta}_t \in \arg\max_{\theta\in\Theta} |\langle \boldsymbol{a}(\theta), \boldsymbol{r}\rangle|$ `// atom selection`

4     $(\widehat{c}_1, \ldots, \widehat{c}_t) = \underset{(c_1,\ldots,c_t)\in\mathbb{R}^t}{\arg\min} \left\| \boldsymbol{y} - \sum_{\ell=1}^{t} c_\ell\, \boldsymbol{a}\left(\widehat{\theta}_\ell\right) \right\|$ `// least-squares update`

5     $\boldsymbol{r} = \boldsymbol{y} - \sum_{\ell=1}^{t} \widehat{c}_\ell\, \boldsymbol{a}\left(\widehat{\theta}_\ell\right)$ `// residual vector`

6     $\widehat{k} = t$;

7   **if** $\boldsymbol{r} = \boldsymbol{0}_{\mathcal{H}}$ **then**

8   quit the loop ;

9   **end**

10 **end**

Output: parameters $\widehat{\mathcal{S}} \triangleq \{\widehat{\theta}_1, \ldots, \widehat{\theta}_{\widehat{k}}\}$ and coefficients $\widehat{c}_1, \ldots, \widehat{c}_{\widehat{k}}$

Suppose $\boldsymbol{y}$ results from the positive linear combination of $k = 2$ distinct atoms, that is $\boldsymbol{y} = c_1\boldsymbol{a}(\theta_1^\star) + c_2\boldsymbol{a}(\theta_2^\star)$, $\theta_1^\star \neq \theta_2^\star$, $c_1 > 0$, $c_2 > 0$. Then, even in this very simple case, OMP never selects an atom in $\{\theta_1^\star, \theta_2^\star\}$ at the first iteration. Indeed, particularizing step 3 of Algorithm 1 to the present setup, we have that, at the first iteration, OMP will select the parameter θ maximizing

$$|\langle \boldsymbol{a}(\theta), \boldsymbol{y}\rangle| = \frac{c_1}{2\sqrt{\pi}} \mathrm{e}^{-\frac{1}{4}(\theta-\theta_1^\star)^2} + \frac{c_2}{2\sqrt{\pi}} \mathrm{e}^{-\frac{1}{4}(\theta-\theta_2^\star)^2}. \tag{1.4}$$

Now, since the right-hand side of (1.4) is twice continuously differentiable, first-order optimality conditions tell us that any maximizer of $|\langle \boldsymbol{a}(\theta), \boldsymbol{y}\rangle|$ must satisfy

$$(\theta - \theta_1^\star)c_1\mathrm{e}^{-\frac{1}{4}(\theta-\theta_1^\star)^2} + (\theta - \theta_2^\star)c_2\mathrm{e}^{-\frac{1}{4}(\theta-\theta_2^\star)^2} = 0. \tag{1.5}$$

Since $\theta_1^\star \neq \theta_2^\star$, $c_1 \neq 0$, $c_2 \neq 0$, this equality cannot be verified by either $\theta_1^\star$ or $\theta_2^\star$. As a consequence, OMP necessarily selects some $\theta \notin \{\theta_1^\star, \theta_2^\star\}$.

□

Nevertheless, we show in this paper that exact recovery is possible with OMP for some particular dictionaries $\mathcal{A}$. Our recovery conditions are expressed in terms of the *kernel function* $\kappa(\theta, \theta')$ associated to the inner product between two atoms, *i.e.*,

$$\kappa(\theta, \theta') \triangleq \langle \boldsymbol{a}(\theta), \boldsymbol{a}(\theta')\rangle. \tag{1.6}$$

In particular, we show that if $\kappa\big(\theta, \theta'\big)$ and a subset $\mathcal{S}^\star = \{\theta_\ell^\star\}_{\ell=1}^k$ are *admissible* (see Section 3.2), then some form of exact recovery is possible with OMP. We emphasize moreover that these admissibility properties are satisfied for a large family of kernels of the form:

$$\kappa(\theta, \theta') = \varphi\Big(\|\theta - \theta'\|_p^p\Big) \qquad 0 < p \leq 1, \tag{1.7}$$

where $\|\cdot\|_p$ is the ℓ_p (quasi-)norm and φ is a *completely monotone* function (see Definition 3). For example, this family encompasses the well-known Laplace kernel. Hereafter, we will refer to kernels taking the form (1.7) as "CMF kernels".

The main focus of the paper is on exact recovery where OMP is required to (unambiguously) identify a subset $\mathcal{S}^\star \triangleq \{\theta_\ell^\star\}_{\ell=1}^k$ in at most k iterations. Our conditions of success guarantee that exact recovery must occur for *any* nonzero values of the weighting coefficients $\{c_\ell\}_{\ell=1}^k$.

A first (perhaps surprising) outcome of our analysis is as follows. If $\Theta \subseteq \mathbb{R}$ and the dictionary is defined by a CMF kernel (1.7), OMP correctly identifies *any* subset of size k in exactly k iterations for *any* $k \in \mathbb{N}$ (see Theorem 2). We emphasize that no separation is needed. To our knowledge, this is the first recovery result of combinations of atoms when no sign constraint is imposed on the coefficients.

When $\Theta \subseteq \mathbb{R}^D$ with $D > 1$ and the dictionary is defined by a CMF kernel (1.7), we show that such an exact recovery result no longer holds (see Example 4). For dictionaries based on CMF kernels, under an additional hypothesis of *axis admissibility*, we demonstrate that a form of *delayed* recovery (that is, in more than k iterations) holds. The number of iterations sufficient to identify the support is upper-bounded by k^D (see result *ii)* of Theorem 4).

Moreover, under axis admissibility, sufficient and necessary conditions for exact recovery of a given subset $\mathcal{S}^\star$ can be written in terms of a *finite* number of atoms of the dictionary (see result *iii)* of Theorem 4). We leverage this result to prove that exact recovery is possible a soon as the elements of the subset $\{\theta_\ell^\star\}_{\ell=1}^k$ obey some "minimum separation" condition (see Theorem 6).

The rest of this paper is organized as follows: Section 2 draws connections with the sparse recovery literature. Sections 3.1 and 3.2 elaborate on the "continuous" dictionary setup and the notion of exact/delayed recovery. Our results are presented in Sections 3.3 and 3.4. Appendix A presents all technical details and concluding remarks are given in Section 4.

Notations

The following notations will be used in this paper. The symbols $\mathbb{R}, \mathbb{R}^*, \mathbb{R}_+, \mathbb{R}_+^*$ refer to the set of real, non null, non-negative and positive numbers, respectively. Boldface lower and upper cases, *e.g.*, $\mathbf{g}$, $\mathbf{G}$ are used to denote vectors and matrices, respectively. We use the notation $[i]$ to refer to the ith element of a vector, and $[i, j]$ for the (i, j)-element of a matrix. We use italic boldface letters, *e.g.*, $\boldsymbol{y}$ or $\boldsymbol{a}$ to denote elements of a Hilbert space $\mathcal{H}$. All-one and all-zero column vectors in $\mathbb{R}^k$ are denoted $\mathbf{1}_k$ and $\mathbf{0}_k$, respectively. When there is no ambiguity, the notations $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$ refer to the inner product and its induced norm on $\mathcal{H}$, while $\|\cdot\|_p$ with $p > 0$ refers to the classical ℓ_p (pseudo-) norm on $\mathbb{R}^D$ for some $D > 0$. Finally, the calligraphic letters $\mathcal{S}, \mathcal{G}$ are used to describe finite subsets of the parameter space Θ, while $[\![m, n]\!]$ denotes the set of integers i such that $m \leq i \leq n$ and $\mathrm{card}(\cdot)$ denotes the cardinality. The main notations used in this paper are summarized in Table E.1 (see Appendix E).

2. Related works and state of the art

Over the last decade, sparse representations have sparked a surge of interest in the signal processing and statistics / machine learning communities. A question of broad interest which has been addressed by many scientists is the identification of the "sparsest" representation of an input signal $\boldsymbol{y}$ (that is, the representation involving the smallest number of elements of $\mathcal{A}$). Since this problem has been shown to be NP-hard [6], many sub-optimal procedures have been proposed to approximate its solution.[1] Among the most popular, one can mention methodologies based on a *convex relaxation* and *greedy algorithms*.

Greedy procedures have a long history in the signal processing and statistical literature, which can be traced back to (at least) the 60's [7]. In the signal processing community, the most popular instances of greedy algorithms are known under the names of *Matching Pursuit* (MP) [8], *Orthogonal Matching Pursuit* (OMP) [5] (also known as *Orthogonal Greedy Algorithm (OGA)* [9, 10]) and *Orthogonal Least Squares* (OLS) [11]. Although these algorithms were already known under different names in other communities [12], they have been "rediscovered" many times, see *e.g.*, [13–15]. Extensions to more general cost functions and kernel dictionaries are discussed in [16].

Sparse representations based on the resolution of convex optimization problems were initially proposed in geophysics [17] for seismic exploration. These methods have been popularized in the signal processing community by the seminal work by Chen *et al.* [18] and by Tibshirani in Statistics [19]. Well-known instances of convex-relaxation approaches for sparse representations are *Basis Pursuit* (BP) [18] and *Lasso* [19], also known as *Basis Pursuit Denoising*, which correspond to different convex optimization problems. Many algorithmic solutions to solve efficiently these problems have been proposed, see *e.g.*, [20–22].

All the "early" contributions mentioned above have been made in the *discrete* setting, where the dictionary contains a *finite* number of atoms. Although [8] already defines MP for continuous dictionaries, greedy sparse approximation in the context of dictionaries made of an infinite (possible uncountable) number of atoms has only been studied more recently [15, 23, 24]. Practical procedures to implement OMP for continuous dictionaries can be found in [25, 26].

On the side of convex relaxation approaches, it was shown in [27] that a *continuous* version of Lasso can be expressed as a convex optimization problem

[1] The term "sub-optimal" has to be understood in the following sense: these procedures only find the sparsest solution of the input vector $\boldsymbol{y}$ under some restricted conditions.

over the space of Radon measures, see [28]. This problem was later referred to as the *Beurling Lasso* (BLasso) [27]. A continuous version of BP was also proposed in [3] for specific continuous dictionaries by exploiting similar ingredients. Different strategies to find the solution of this problem (to some accuracy) were proposed in a series of papers, see [2, 3, 29, 30].

Because the approaches mentioned above (both in the discrete and continuous settings) are *sub-optimal* procedures looking for the sparsest representation of some $\boldsymbol{y}$, many theoretical works have been carried out to analyze their performance. In the rest of this section, we review the contributions of the literature most related to the present work. We organize our presentation into two parts, dealing respectively with the discrete and the continuous cases. In the discrete setting, we restrict our attention to contributions addressing the performance of MP, OMP, *i.e.*, the procedures the most connected to the framework of this paper. In the continuous setting, recovery analysis, including stability and robustness to noise, has only been addressed for convex-relaxation approaches. We review these conditions below and draw some similarities and differences with the guarantees derived for OMP in this paper in Section 3.

2.1. Discrete setting

Exact Recovery Condition. In the discrete setting, the cardinal of the parameter space Θ is finite, written $|\Theta| < +\infty$. The first thorough analysis of the recovery of some subset $\mathcal{S}^\star \triangleq \{\theta_\ell^\star\}_{\ell=1}^k$ by OMP is due to Tropp in [31]. Recall that since Θ is finite, recovering $\mathcal{S}^\star$ is equivalent to identifying the *support* of a sparse vector, that is to say the set of indices of its nonzero entries. Rephrasing Tropp's results with the formalism of this paper, if the atoms $\{\boldsymbol{a}(\theta_\ell^\star)\}_{\ell=1}^k$ are linearly independent, a necessary and sufficient condition for the exact recovery[2] of $\mathcal{S}^\star$ by OMP is

$$\forall \theta \in \Theta \setminus \mathcal{S}^\star : \left\| \mathbf{G}^{-1} \mathbf{g}_\theta \right\|_1 < 1, \tag{2.1 – ERC}$$

where $\mathbf{G} \in \mathbb{R}^{k\times k}$ and $\mathbf{g}_\theta \in \mathbb{R}^k$ are defined as follows[3]

$$\begin{aligned} \mathbf{G}[i,j] &\triangleq \kappa\big(\theta_i^\star, \theta_j^\star\big), \\ \mathbf{g}_\theta[i] &\triangleq \kappa(\theta, \theta_i^\star). \end{aligned} \tag{2.2}$$

In particular, if (2.1 – ERC) is not satisfied, there exists some linear combination $\boldsymbol{y}$ of the atoms $\{\boldsymbol{a}(\theta_\ell^\star)\}_{\ell=1}^k$, such that OMP with $\boldsymbol{y}$ as input fails at the first iteration. Condition (2.1 – ERC) is also related to the success of MP and OLS. In [32], the authors showed that (2.1 – ERC) is necessary and sufficient for the recovery of $\mathcal{S}^\star$ by OLS. As for MP, (2.1 – ERC) ensures that the procedure only selects atoms in $\mathcal{S}^\star$ but does not imply exact recovery in k steps since the same atom can be selected many times (no orthogonal projection is performed

[2] In the sense of [31]. We define rigorously the notions of exact/delayed recovery for continuous dictionaries in Section 3.1.

[3] Recall that κ is a kernel function acting as an inner product between atoms, see (1.6).

when updating the coefficients), see *e.g.*, [23]. Interestingly, it also ensures correct identification of any support of size k by convex relaxation procedures, see [33–35].

Coherence. Tropp's condition is of limited practical interest to characterize the recovery of all supports of size k since it requires to verify that (2.1 – ERC) holds for any $\mathcal{S}^\star$ with $\mathrm{card}(\mathcal{S}^\star) = k$. In order to circumvent this issue, other sufficient conditions of success, weaker but easier to evaluate in practice, have been proposed in the literature. One of the most popular conditions is based on the *coherence* μ of the normalized dictionary:

$$k < \frac{1}{2}\left(1 + \frac{1}{\mu}\right) \tag{2.3}$$

where

$$\mu \triangleq \sup_{\substack{\theta,\theta' \in \Theta \\ \text{s.t. } \theta \neq \theta'}} |\kappa(\theta, \theta')| \tag{2.4}$$

Condition (2.3) guarantees that any support of size k can be recovered in k steps by OMP/OLS.

The coherence of the dictionary can be seen as a particular measure of "proximity"[4] between the atoms $\{\boldsymbol{a}(\theta)\}_{\theta\in\Theta}$. Other exact recovery conditions, based on different proximity measures, have been proposed in the literature. In [31], the author derived recovery conditions based on "cumulative coherence", whereas in [10, 36–42], guarantees based on "restricted isometry constants" were proposed.

2.2. Continuous setting

General setup. Sparse representations in continuous dictionaries are basically characterized by two main ingredients: *i)* a parameter set Θ, assumed to be uncountable and metrizable, *e.g.*, $\Theta = \mathbb{R}$; *ii)* an "atom" function $\boldsymbol{a} : \Theta \to \mathcal{H}$, assumed to enjoy some continuity properties, *i.e.*, if θ and θ' are "close" to each other then so are $\boldsymbol{a}(\theta)$ and $\boldsymbol{a}(\theta')$. This type of dictionary appears in numerous signal processing tasks, *e.g.*, super-resolution where one aims at recovering fine-scale details from an under-resolved input signal [3] or sparse spike deconvolution in seismic inversion [17].

Inapplicability of existing analyses. The continuity of $\boldsymbol{a}(\cdot)$ invalidates most previous analyses in the context of discrete dictionaries. In particular, all exact recovery conditions based on coherence or restricted isometry constants turn out to be violated whenever dealing with continuous dictionaries. As for the coherence condition (2.3), it is easy to see that the continuity of $\boldsymbol{a}(\cdot)$ implies the continuity of $\kappa(\cdot,\cdot)$ with respect to both its arguments. This, in turn, implies that $\mu = 1$ and the coherence-based condition (2.3) is never met, even for $k = 1$!

[4] $\mu = 0$ if all the atoms are orthogonal and $\mu \simeq 1$ if some atoms are very correlated.

Minimum separation assumption. In order to circumvent this issue, some specific exact recovery conditions for continuous dictionaries have been proposed in the literature, see *e.g.*, [3, 4]. In the context of convex-relaxation approaches, these conditions originate from the analysis of the associated optimality conditions. In [4], the authors derived dual certificates generalizing the work done by Fuchs for the Lasso [33] in an infinite-dimensional setup. The comparison with the discrete case goes even deeper: it is shown that, as in the discrete case, the solution of BP for continuous dictionaries is, in some sense, the limit of the solution of BLasso. Two families of results rise up from these studies. Under certain conditions, BP recovers any positive combination of atoms. However, when the observation is made up of a signed combinations of atoms, a *minimum separation* condition becomes necessary to ensure recovery. More particularly, a signal supported in $\mathcal{S}^\star$ is the unique solution of BP for continuous dictionaries provided that

$$\min_{\substack{\ell,\ell'\in[\![1,k]\!]\\ \text{s.t.}\quad \ell\neq\ell'}} |\theta^\star_{\ell'} - \theta^\star_\ell| > C, \tag{2.5}$$

where C is a constant depending on the problem, see for instance [3, 4, 43]. Surprisingly, for $D = 1$, this separation condition is no longer needed when dealing with positive linear combinations of atoms [27, 44, 45]. In particular, under certain conditions of linear independence on the atoms $\{\boldsymbol{a}(\theta^\star_\ell)\}_{\ell=1}^k$, the signal supported in $\mathcal{S}^\star$ is the unique solution of BP. When $\Theta \subset \mathbb{R}^D$ with $D > 1$, a separation condition still exists for BP [3], but also depends on $\mathcal{S}^\star$, see [46].

3. Main results

In this section, we present the main results of the paper. All the proofs and technical details are postponed to Section A.

3.1. Definitions of support recovery

Dictionary. We first elaborate on the ingredients characterizing the "continuous" setup considered in this paper. First, the metric space Θ is usually assumed to contain an infinite uncountable number of elements, *e.g.*, Θ can be a hyper-rectangle of $\mathbb{R}^D$, a torus, etc. Hereafter, for the sake of conciseness, we will restrict our attention to the case where $\Theta = \mathbb{R}^D$. Nevertheless, our results can be straightforwardly adapted to other definitions of Θ. A second ingredient is the continuity of the function $\boldsymbol{a} : \Theta \to \mathcal{H}$, that is

$$\lim_{\theta'\to\theta} \|\boldsymbol{a}(\theta') - \boldsymbol{a}(\theta)\| = 0 \quad \forall\theta\in\Theta. \tag{3.1}$$

In this paper, we will moreover consider normalized atoms:

$$\|\boldsymbol{a}(\theta)\| = 1 \quad \forall\theta\in\Theta. \tag{3.2}$$

In the sequel, recovery conditions will be expressed as a function of the symmetric kernel $\kappa\big(\theta,\theta'\big)$:

$$\kappa(\theta,\theta') \triangleq \langle \boldsymbol{a}(\theta), \boldsymbol{a}(\theta')\rangle \qquad \forall\theta,\theta'\in\Theta. \tag{3.3}$$

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

$$\text{“\textit{unit norm}”}: \qquad \kappa\big(\theta,\theta\big) = 1 \quad \forall\theta\in\Theta, \tag{3.4a}$$

$$\text{“\textit{continuity}”}: \lim_{\theta'\to\theta} \kappa\big(\theta,\theta'\big) = 1 \quad \forall\theta\in\Theta. \tag{3.4b}$$

Moreover, we also have from the Cauchy-Schwarz inequality that

$$\big|\kappa\big(\theta,\theta'\big)\big| \leq 1, \qquad \forall\theta,\theta'\in\Theta. \tag{3.5}$$

Lastly, in this work, we restrict our attention to either compact domains Θ or kernels that vanish at infinity, *i.e.*,

$$\forall\,\varepsilon>0, \forall\theta\in\Theta\,,\ \exists K\subset\Theta \text{ compact:} \quad \sup_{\theta'\in K^c} \kappa(\theta',\theta) < \varepsilon, \tag{3.6}$$

where K^c is the complement of K. When the observation $\boldsymbol{y}$ is made up of a linear combination of atoms, Condition (3.6) along with the continuity of $\theta'\mapsto\kappa(\theta,\theta')$ is sufficient to guarantee that the maximizer of the inner product exists at each iteration of OMP, see Line 3 in Algorithm 1.

Notion of support. First, let $\mathcal{S}^\star \triangleq \{\theta^\star_\ell\}_{\ell=1}^k$ be a subset of Θ made of pairwise distinct parameters. We say that $\mathcal{S}^\star$ is *a* support of $\boldsymbol{y}$ if

$$\boldsymbol{y} = \sum_{\ell=1}^{k} c_\ell\, \boldsymbol{a}(\theta^\star_\ell) \tag{3.7}$$

with $c_1,\cdots,c_k\neq 0$.
In the sequel, we provide sufficient conditions on the kernel κ to ensure that *any finite collection of pairwise distinct atoms is linearly independent.* In such a case, the notion of support is uniquely defined: if $\mathcal{S}^\star$ is *a* support of $\boldsymbol{y}$, then without ambiguity $\mathcal{S}$ can be called *the* support of $\boldsymbol{y}$, written $\mathrm{supp}(\boldsymbol{y})$. When $\mathrm{supp}(\boldsymbol{y}) = \mathcal{S}^\star$ and $\mathrm{card}(\mathcal{S}^\star)=k$, we also say that $\boldsymbol{y}$ is "k-sparse", and that it is "supported in" $\mathcal{S}^\star$.

Notions of recovery. Given a dictionary made of linearly independent atoms and a k-sparse observation $\boldsymbol{y}\neq\mathbf{0}_{\mathcal{H}}$ with support $\mathcal{S}^\star\triangleq\mathrm{supp}(y)$, we say that OMP achieves *exact recovery* if and only if (when run with a maximum number of iteration $q\geq\mathrm{card}(\mathcal{S}^\star)$) we have

$$\emptyset \neq \arg\max_{\theta\in\Theta} |\langle \boldsymbol{a}(\theta), \boldsymbol{r}\rangle| \subseteq \mathcal{S}^\star, \tag{3.8}$$

during the first $\mathrm{card}(\mathcal{S}^\star)$ iterations[5] and the residual (see Line 5 in Algorithm 1) becomes $\boldsymbol{r}=\mathbf{0}_{\mathcal{H}}$ in the next iterations. The left-hand side of (3.8) ensures that a

[5]See Line 6 in Algorithm 1.

maximizer of $|\langle \boldsymbol{a}(\theta), \boldsymbol{r}\rangle|$ *does* exist at each iteration. The right-hand side entails that any maximizer belongs to the support $\mathcal{S}^\star$, that is

$$\max_{\theta\in\mathcal{S}^\star} |\langle \boldsymbol{a}(\theta), \boldsymbol{r}\rangle| > |\langle \boldsymbol{a}(\theta'), \boldsymbol{r}\rangle| \qquad \forall \theta' \in \Theta \setminus \mathcal{S}^\star. \tag{3.9}$$

The set of maximizers appearing in (3.8) might not reduce to a singleton; if it is a subset of $\mathcal{S}^\star$, the definition means that all choices for $\widehat{\theta}$ lead to exact recovery. However, we consider that the situation where

$$\max_{\theta\in\mathcal{S}^\star} |\langle \boldsymbol{a}(\theta), \boldsymbol{r}\rangle| = |\langle \boldsymbol{a}(\theta'), \boldsymbol{r}\rangle| \quad \text{for some } \theta' \in \Theta \setminus \mathcal{S}^\star \tag{3.10}$$

leads to a failure.

The success of OMP in *more* than k steps, referred to as *delayed recovery*, has also been considered [47, 48]. In this context, we say that OMP achieves *q-delayed recovery* with $q \geq \operatorname{card}(\mathcal{S}^\star)$ (necessarily) if, when OMP is run with input $\boldsymbol{y}$ and a maximum number of iterations q, the inferred support contains $\operatorname{supp}(\boldsymbol{y})$. In this case, we have by (3.7) and linear independence between atoms:

$$\widehat{\theta}_\ell \notin \mathcal{S}^\star \Longrightarrow \widehat{c}_\ell = 0. \tag{3.11}$$

We now say that OMP achieves exact recovery of $\mathcal{S}^\star$ if for all observations $\boldsymbol{y}$ supported in $\mathcal{S}^\star$, OMP with input $\boldsymbol{y}$ achieves exact recovery. Similarly, we say that OMP achieves q-delayed recovery of $\mathcal{S}^\star$ with $q \geq \operatorname{card}(\mathcal{S}^\star)$ (necessarily) if for all observations $\boldsymbol{y}$ supported in $\mathcal{S}^\star$, OMP with input $\boldsymbol{y}$ achieves q-delayed recovery.

3.2. Exact recovery of a given support: sufficient conditions

In this section, we introduce a set of sufficient conditions ensuring exact recovery of some support $\mathcal{S}^\star$ by OMP. These conditions are the basis of our results on "CMF dictionaries" stated in the next two sections.

We first notice that, in the context of continuous dictionaries, condition (2.1 – ERC) is still necessary and sufficient for exact recovery of a support $\mathcal{S}^\star$. However, this condition may become intractable in practice since Θ now contains an infinite uncountable number of elements. In particular, the standard formulation

$$\max_{\theta\in\Theta\setminus\mathcal{S}^\star} \left\| \mathbf{G}^{-1}\mathbf{g}_\theta \right\|_1 < 1, \tag{3.12}$$

equivalent to (2.1 – ERC) in the discrete setting, does no longer hold in the case of continuous dictionaries as the supremum

$$\sup_{\theta\in\Theta\setminus\mathcal{S}^\star} \left\| \mathbf{G}^{-1}\mathbf{g}_\theta \right\|_1 \tag{3.13}$$

is always at least 1. In order to circumvent this problem, we define below the notions of admissibility for the dictionary $\mathcal{A}$ and the support $\mathcal{S}^\star$, easier to evaluate than (2.1 – ERC), but ensuring exact recovery (see Theorem 1). In the next section, we will show that some families of dictionaries (whose kernel is defined by a "*completely monotone function*") verify these abstract conditions.

Consider a kernel κ and a support $\mathcal{S}^\star \triangleq \{\theta_\ell^\star\}_{\ell=1}^k$. Our definitions of admissibility for kernel κ and support $\mathcal{S}^\star$ read:

Definition 1 (Admissible kernel). *A kernel κ is said to be* admissible *if: a) it is continuous; b) it satisfies (3.4a); c) the parameter set Θ is compact or (3.6) holds; and d) $0 \leq \kappa\big(\theta, \theta'\big) < 1$ for any $\theta \neq \theta'$.*
By extension, a dictionary $\mathcal{A} = \{\boldsymbol{a}(\theta) \mid \theta \in \Theta\}$ is said to be admissible *if the kernel defined by $\kappa(\theta, \theta') = \langle \boldsymbol{a}(\theta), \boldsymbol{a}(\theta')\rangle$ for all $\theta, \theta' \in \Theta$ is admissible.*

Definition 2 (Admissible support given a kernel). *A support $\mathcal{S}^\star$ of size k is* admissible *with respect to the kernel κ if the following holds for any non-empty subset $T \subseteq [\![1, k]\!]$ and any positive coefficients $c_\ell > 0, \ell \in T$ such that $\sum_{\ell \in T} c_\ell < 1$:*

i) The set of global maximizers of

$$\begin{aligned} \psi \colon \Theta &\longrightarrow \mathbb{R}_+ \\ \theta &\longmapsto \sum_{\ell \in T} c_\ell\, \kappa(\theta, \theta_\ell^\star)\,, \end{aligned} \tag{3.14}$$

is a subset of $\mathcal{S} = \{\theta_\ell^\star\}_{\ell \in T}$.

ii) If $t' \in [\![1, k]\!] \setminus T$ satisfies $\psi(\theta_\ell^\star) - \kappa(\theta_\ell^\star, \theta_{t'}^\star) \leq 0$ for all $\ell \in T$, then

$$\forall\, \theta \in \Theta\,, \quad \psi(\theta) - \kappa(\theta, \theta_{t'}^\star) \leq 0. \tag{3.15}$$

By extension, the support $\mathcal{S}^\star$ is said to be admissible *with respect to the dictionary $\mathcal{A} = \{\boldsymbol{a}(\theta) \mid \theta \in \Theta\}$ if $\mathcal{S}^\star$ is admissible with respect to the kernel defined by $\kappa(\theta, \theta') = \langle \boldsymbol{a}(\theta), \boldsymbol{a}(\theta')\rangle$ for all $\theta, \theta' \in \Theta$.*

We can make the following general comments on these definitions. First, $0 \leq \kappa(\theta, \theta')$ implies that the inner product between two atoms of $\mathcal{A}$ is always nonnegative, whereas $\kappa(\theta, \theta') < 1$ simply guarantees that any pair of distinct atoms of the dictionary is linearly independent[6] (remember we assume $\kappa(\theta, \theta) = 1$ for all $\theta \in \Theta$). Property *i)* in Definition 2 ensures that a correct atom selection always occurs when the residual $\boldsymbol{r}$ is a positive combination of the atoms of the support. Indeed, if $\boldsymbol{r} = \sum_{\ell=1}^k c_\ell\, \boldsymbol{a}(\theta_\ell^\star)$ with $c_1 \ldots c_k > 0$ and the kernel is admissible then from Definition 1

$$|\langle \boldsymbol{a}(\theta), \boldsymbol{r}\rangle| = \sum_{\ell=1}^k c_\ell\, \kappa(\theta, \theta_\ell^\star). \tag{3.16}$$

[6] As a consequence of the Cauchy-Schwarz inequality.

In such a case, Property *i)* of Definition 2 then implies

$$\arg\max_{\theta\in\Theta} |\langle \boldsymbol{a}(\theta), \boldsymbol{r}\rangle| \subseteq \mathcal{S}^\star. \tag{3.17}$$

Property *ii)* of Definition 2 does not have such a simple interpretation but we will see that it allows to significantly shorten the proof of recovery results. Moreover, we let the reader check that the following result is a direct consequence of the definition of support admissibility:

Fact 1. *If $\mathcal{S}^\star$ is admissible with respect to κ then each $\mathcal{S} \subset \mathcal{S}^\star$ is admissible.*

We now state our first result:

Theorem 1. *Assume kernel κ is admissible (see Definition 1) and support $\mathcal{S}^\star$ is admissible with respect to kernel κ (see Definition 2). Then:*

i) OMP achieves exact recovery of $\mathcal{S}^\star$.

ii) For any k-sparse observation $\boldsymbol{y}$ whose support is included in $\mathcal{S}^\star$, OMP with input $\boldsymbol{y}$ only selects parameters in $\mathrm{supp}(\boldsymbol{y})$.

Proof. See Appendix A.1. □

In the next section, we provide some families of *admissible* dictionaries such that *all* supports are admissible. This yields exact recovery of *any* support.

3.3. Recovery of k-sparse signals with CMF dictionaries

In this section, we show that exact recovery is possible for dictionaries whose associated kernel takes the form

$$\kappa(\theta, \theta') = \varphi\Big(\|\theta - \theta'\|_p^p\Big), \qquad \forall \theta, \theta' \in \Theta \subset \mathbb{R}^D \tag{3.18}$$

where $\|\cdot\|_p$, $0 \le p < 1$ is the ℓ_p (quasi-)norm and φ is some function. More particularly, we will show in Theorem 2 that for dictionaries verifying (3.18) and $\Theta = \mathbb{R}$, OMP achieves exact recovery of any support.
We will see that a way to build admissible kernels κ and supports (see Definitions 1 and 2) is to ensure that φ decreases sufficiently fast when two parameters are distant from each other. One appropriate framework to describe such kernels is the set of *completely monotone function (CMF)* defined below:

Definition 3 ([49], Def. 2.1)**.** *A function $\varphi : \mathbb{R}_+ \longmapsto \mathbb{R}$ is completely monotone on $[0, +\infty[$ if and only if: it is infinitely differentiable on $]0, +\infty[$, its derivatives obey*

$$(-1)^n \varphi^{(n)}(x) > 0 \quad \forall x, n \in \mathbb{R}_+^* \times \mathbb{N}, \tag{3.19}$$

where $\varphi^{(n)}$ is the n-th derivative of φ, and it is right continuous at 0.

Example 2. *The following functions are completely monotone [50]:*

- *the function $x \mapsto \mathrm{e}^{-\lambda x}$ for $\lambda > 0$ which gives birth to the Laplace kernel,*

- *the function* $x \mapsto \frac{1}{1+\lambda x}$ *for* $\lambda > 0$,
- *ratios of modified Bessel functions of the first kind,*
- *a subset of the confluent hypergeometric functions (Kummer's function),*
- *a subset of the Gauss hypergeometric functions.*

Our results exploit the following characterization of CMF in terms of Laplace transforms of measures, referred to as the Bernstein-Widder theorem [51, 52]:

Lemma 1 ([53], Theorem 7.11)**.** *A function* φ *is completely monotone on* $[0, +\infty[$ *if and only if there exists a nonzero (unsigned) finite measure* ν *on Borel sets of* $[0, +\infty[$ *that is not proportional to the Dirac measure in* 0*, written* δ_0*, such that*

$$\varphi(x) = \int_{[0,+\infty[} \mathrm{e}^{-ux}\,\mathrm{d}\nu(u) \tag{3.20}$$

where the integral converges for all $x \geq 0$.

One immediately sees from the integral representation (3.20) that the Laplace kernel is a CMF whose corresponding measure is a Dirac measure $\nu = \delta_\lambda$ with $\lambda > 0$. However, the case ν proportional to δ_0 is prohibited since it leads to $\varphi'(x) = 0$ for all $x \in \mathbb{R}_+$ which is in contradiction with Definition 3 for a CMF. Since $\varphi(0) = 1$, ν can be any probability measure such that $\nu \neq \delta_0$.

By definition, any CMF is (strictly) positive, decreasing and convex. In the sequel, we state recovery results for dictionaries whose kernels satisfy (3.18) with φ a CMF such that $\varphi(0) = 1$, so as to ensure the "unit-norm" hypothesis (3.4a).

Definition 4 (CMF kernel)**.** *The class of* CMF kernels *in dimension* $D \geq 1$, *denoted* $\mathcal{K}_{\mathtt{CMF}}(D)$, *consists of all kernels* $\kappa : \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}_+^*$ *such that*

$$\kappa(\theta, \theta') = \varphi\Big(\|\theta - \theta'\|_p^p\Big) \quad \forall \theta, \theta' \in \mathbb{R}^D \tag{3.21}$$

where φ *is a CMF verifying* $\varphi(0) = 1$*, and* $0 < p \leq 1$.

Definition 5 (CMF dictionary)**.** *A CMF dictionary in dimension* $D \geq 1$ *is a collection of atoms* $\mathcal{A} = \{\boldsymbol{a}(\theta) \mid \theta \in \Theta\}$*, where* $\Theta \subset \mathbb{R}^D$*, such that there exists* $\kappa \in \mathcal{K}_{\mathtt{CMF}}(D)$ *satisfying* $\langle \boldsymbol{a}(\theta), \boldsymbol{a}(\theta') \rangle = \kappa(\theta, \theta')$ *for all* $\theta, \theta' \in \Theta$.

A legitimate question is as follows: for any kernel $\kappa \in \mathcal{K}_{\mathtt{CMF}}(D)$, is there a Hilbert space $\mathcal{H}$ and a dictionary $\mathcal{A}$ such that

$$\langle \boldsymbol{a}(\theta), \boldsymbol{a}(\theta') \rangle = \kappa(\theta, \theta') \tag{3.22}$$

holds for any $\theta, \theta' \in \mathbb{R}^D$? The answer is positive (see Appendix B.2) since the elements of $\mathcal{K}_{\mathtt{CMF}}(D)$ are positive definite functions and Mercer's Theorem [54] ensures the existence of a Hilbert space $\mathcal{H}$ and a mapping $\boldsymbol{a} : \Theta \longrightarrow \mathcal{H}$ such that κ acts as an inner product in $\mathcal{H}$. The next example exhibits a family of atoms in $\mathcal{H} = L_2(\mathbb{R})$ which is a CMF dictionary in $\mathbb{R}$.

Example 3. *Let $\Theta = \mathbb{R}$ and consider the dictionary $\mathcal{A}$ defined by*

$$\begin{aligned} \boldsymbol{a} : \mathbb{R} &\longrightarrow L_2(\mathbb{R}) \\ \theta &\longmapsto t \mapsto \sqrt{2\lambda}\, \mathrm{e}^{-\lambda(t-\theta)} \mathbb{I}\{t \geq \theta\} \end{aligned} , \tag{3.23}$$

for some $\lambda > 0$, where $\mathbb{I}\{t \geq \theta\}$ is the "indicator" function which is equal to 1 if $t \geq \theta$ and 0 otherwise. Straightforward calculations both show that $\|\boldsymbol{a}(\theta)\| = 1$ for any θ and the inner product in $\mathcal{H} = L_2(\mathbb{R})$ between two atoms writes $\langle \boldsymbol{a}(\theta), \boldsymbol{a}(\theta') \rangle = \mathrm{e}^{-\lambda|\theta-\theta'|}$. The latter function corresponds to the so-called "Laplace kernel". One immediately sees that such a kernel is an element of $\mathcal{K}_{\mathtt{CMF}}(1)$ by taking $\varphi : x \mapsto \mathrm{e}^{-\lambda x}$ and $p = 1$.

,

A first surprising result holds in the context of CMF dictionaries with $\Theta = \mathbb{R}$:

Theorem 2. *Consider $\mathcal{A}$ an* arbitrary *CMF dictionary in $\mathbb{R}$. Then for* any *support $\mathcal{S}^\star \subset \Theta \subset \mathbb{R}$, the two following properties hold:*

i) OMP achieves exact recovery of $\mathcal{S}^\star$.

ii) For any k-sparse observation $\boldsymbol{y}$ whose support is included in $\mathcal{S}^\star$, OMP with input $\boldsymbol{y}$ only selects parameters in $\mathrm{supp}(\boldsymbol{y})$.

Proof. See Appendix A.2.1. □

In essence, Theorem 2 identifies a class of dictionaries for which exact recovery holds *for any support of any finite size k*. To our knowledge, this is the first occurrence of such a "universal" recovery result of signed combinations of atoms. We have already seen that BP is able to recover any positive linear combination of atoms [27] for some families of dictionary. However, a separation condition becomes necessary if one allows negative coefficients (see Section 2). The novelty of Theorem 2 is a *separation-free* recovery result for any signed finite linear combination of atoms. In light of the existing links between Tropp's ERC and recovery guarantees for ℓ_1 minimization [34], it would be interesting to understand whether the guarantees expressed in Theorem 2 extend to sparse spike recovery with total variation norm minimization [4, 27].

3.4. Support recovery with "CMF" kernels

The universal exact recovery result stated in Theorem 2 no longer holds in dimension $D > 1$. In the next example, we show that if $D > 1$, there always exist configurations of parameters $\{\theta_\ell^\star\}_{\ell=1}^k$ such that OMP fails at the first iteration:

Example 4. *In dimension $D \geq 3$, consider $\mathcal{S}^\star \triangleq \{\theta_\ell^\star\}_{\ell=1}^k \subsetneq \Theta = \mathbb{R}^D$ where $3 \leq k \leq D$ and $\theta_\ell = \Delta \mathbf{e}_\ell$ with $\mathbf{e}_\ell$ the ℓ-th canonical basis vector and $\Delta > 0$. This configuration defines a support $\mathcal{S}^\star$ of size k with $\left\|\theta_\ell^\star - \theta_j^\star\right\|_p^p = 2\Delta^p$ for all $j \neq \ell$, and $\|\theta_\ell^\star - \mathbf{0}_D\|_p^p = \Delta^p$ for all ℓ.*

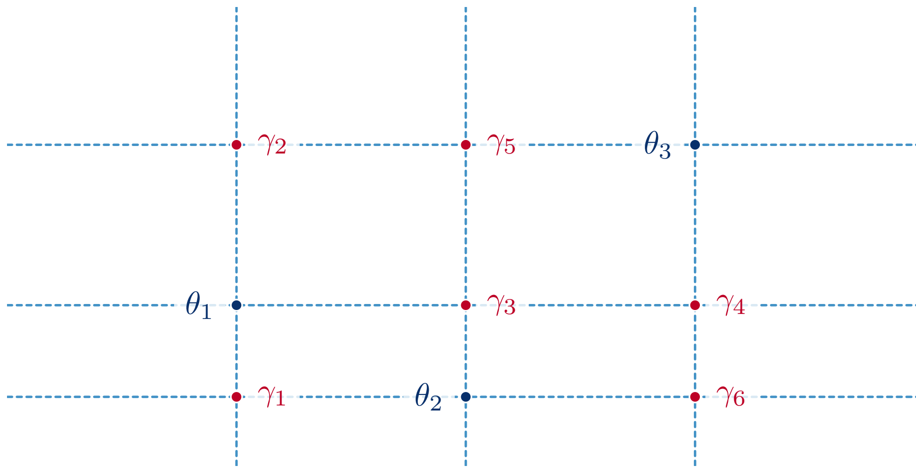

Figure 1: Illustration in dimension $D = 2$ with $k = 3$ of the definition of the set augmenter `Cart` defined in (3.28). The blue points, denoted θ_ℓ for $\ell = 1, 2, 3$, form the support $\mathcal{S}^\star$. The red points, denoted γ_ℓ, $\ell = 1 \dots 6$ represent the elements of $\texttt{Cart}(\mathcal{S}^\star) \setminus \mathcal{S}^\star$.

Let $\boldsymbol{a} : \mathbb{R}^D \mapsto \mathcal{H}$ define a CMF dictionary in $\mathbb{R}^D$ with kernel $\kappa = \varphi\left(\|\cdot - \cdot\|_p^p\right)$. Let us show that, if Δ is sufficiently small, there always exists a linear combination of the atoms $\{\boldsymbol{a}(\theta)\}_{\theta \in \mathcal{S}^\star}$ such that OMP selects a parameter not *in $\mathcal{S}^\star$ at the first iteration. In particular, let us consider the case where all coefficients c_ℓ are equal and $\boldsymbol{y} = \sum_{\ell=1}^k c_\ell \boldsymbol{a}(\theta_\ell^\star)$. We then have*

$$\frac{\langle \boldsymbol{a}(\mathbf{0}_D), \boldsymbol{y}\rangle}{\langle \boldsymbol{a}(\theta_\ell^\star), \boldsymbol{y}\rangle} = \frac{k\varphi(\Delta^p)}{1 + (k-1)\varphi(2\Delta^p)}. \tag{3.24}$$

Then, $\theta = \mathbf{0}_D$ will be preferred to all "ground-truth" parameters $\theta_\ell^\star$ at the first iteration of OMP as soon as (3.24) is larger than 1, or, equivalently,

$$(k-1)\varphi(2\Delta^p) - k\varphi(\Delta^p) + 1 < 0. \tag{3.25}$$

Let us show that (3.25) holds whenever Δ^p is "sufficiently small". We first consider the case where $\varphi(t) = \mathrm{e}^{-\lambda t}$ with $\lambda > 0$. Condition (3.25) then writes

$$(k-1)x^2 - kx + 1 < 0 \tag{3.26}$$

with $x = \varphi(\Delta^p)$. As $k \geq 3$, the left-hand side of (3.26) is a second order polynomial with two distinct roots, namely $(k-1)^{-1}$ and 1. Therefore, OMP prefers $\mathbf{0}_D$ as soon as $(k-1)^{-1} < x < 1$ or, equivalently, when $\Delta^p < \lambda^{-1}\log(k-1)$. The latter condition acts as a necessary separation *condition such that OMP does not fail at the first iteration. It is possible to draw similar conclusions whenever φ is a CMF function differentiable at zero and $\varphi(0) = 1$. The proof requires extra work detailed in Appendix C.*

Before stating our next results, we introduce some additional notations.

Definition 6 (Cartesian grid)**.** *A finite set* $\mathcal{S} \subsetneq \Theta$ *is a* Cartesian grid *in dimension* $D \geq 1$ *if there exists* D *one-dimensional finite sets* $\{\mathcal{S}_d\}_{d=1}^{D}$ *such that*

$$\mathcal{S} = \prod_{d=1}^{D} \mathcal{S}_d \tag{3.27}$$

where $\prod$ *denotes the Cartesian product.*

Moreover, we define the following "set augmenter" operator that, given a set $\mathcal{S}$, returns the smallest Cartesian grid containing $\mathcal{S}$

$$\texttt{Cart}(\mathcal{S}) \triangleq \prod_{d=1}^{D} \Big\{ \theta[d] \mid \theta \in \mathcal{S} \Big\}. \tag{3.28}$$

The reader can check that $\mathcal{S} \subseteq \texttt{Cart}(\mathcal{S})$ for any finite set $\mathcal{S} \subseteq \Theta$ and that $\texttt{Cart}$ is idempotent. We illustrate the definitions of $\texttt{Cart}(\mathcal{S})$ in Fig. 1 for $\mathcal{S} = \{\theta_1, \theta_2, \theta_3\}$ in dimension $D = 2$.

Definition 7 (Axis admissible grid given a kernel)**.** *Consider* κ *a CMF kernel in the sense of Definition 4.*
A Cartesian grid $\mathcal{G}$ *is axis admissible with respect to* κ *if and only if: for each Cartesian grid* $\mathcal{G}' = \prod_{d'=1}^{D} \mathcal{S}_{d'} \subset \mathcal{G}$*, for coefficient vector* $\mathbf{c} = (c_1 \ldots c_q)^{\mathrm{T}} \in \mathbb{R}^q \setminus \{\mathbf{0}_q\}$ *where* $q = \mathrm{card}(\mathcal{G}')$*, for each* $d \in [\![1, D]\!]$ *and each* $\theta_0 \in \Theta$ *such that* $\theta_0 \perp \mathbf{e}_d$ *where* $\mathbf{e}_d$ *is the* d*-th vector of the canonical basis, all maximizers of*

$$\begin{aligned} f_d \colon \mathbb{R} &\longrightarrow \mathbb{R}_+ \\ t &\longmapsto \left| \sum_{\ell=1}^{\mathrm{card}(\mathcal{G}')} c_\ell \kappa(\theta_0 + t\mathbf{e}_d, \theta_\ell^\star) \right|, \end{aligned} \tag{3.29}$$

belong to $\mathcal{S}_d$*.*
By extension, a Cartesian grid $\mathcal{G}$ *is said to be axis admissible with respect to a CMF dictionary* $\mathcal{A}$ *if it is axis admissible with respect to the induced kernel.*

Note that we have restricted the notion of axis admissibility to CMF kernels. Although the notion could be defined for any kernel, we will see below (see the proof of Corollary 1 in Appendix A.2.2) that axis admissibility is mainly useful for CMF kernels. Moreover, we let the reader check that the following result is a direct consequence of the definition of axis admissibility.

Fact 2. *If a cartesian grid* $\mathcal{G}$ *is axis admissible with respect to* κ*, the each cartesian grid* $\mathcal{G}'$ *such that* $\mathcal{G}' \subset \mathcal{G}$ *is also axis admissible.*

In higher dimension, the notion of exact recovery is first proved for axis admissible grids:

Theorem 3. *Let* $\mathcal{A}$ *be an* arbitrary *CMF dictionary and* $\mathcal{G}$ *be a Cartesian grid. If* $\mathcal{G}$ *is axis admissible with respect to* $\mathcal{A}$ *then* *i) OMP achieves exact recovery of $\mathcal{G}$.*

ii) For any k-sparse observation $\boldsymbol{y}$ with $\operatorname{supp}(\boldsymbol{y}) \subset \mathcal{G}$, OMP with input $\boldsymbol{y}$ selects parameters in $\texttt{Cart}(\operatorname{supp}(\boldsymbol{y}))$ until the residual vanishes.

Proof. See Appendix A.2.2. □

In other words, Theorem 3 ensures that for any admissible grid $\mathcal{G}$ and observation $\boldsymbol{y}$ supported in $\mathcal{S}^\star = \mathcal{G}$, an iteration of OMP with input $\boldsymbol{y}$ will identify an element of $\mathcal{G}$ irrespectively of the (nonzero) coefficients in the representation of $\boldsymbol{y}$ (see (3.7)).

Axis admissible Cartesian grids are the cornerstone of Theorem 4 below, our main recovery result in higher dimensions. Indeed, for a given support $\mathcal{S}^\star$ of size k, a form of delayed recovery becomes possible provided that the extended support $\texttt{Cart}(\mathcal{S}^\star)$ is axis admissible. Moreover, exact recovery will be achieved by preventing from selecting parameters in $\texttt{Cart}(\mathcal{S}^\star)\setminus\mathcal{S}^\star$ by means of a classical ERC condition for finite dictionaries (see (2.1 – ERC)). This rationale is formalized in the following result:

Theorem 4. *Let $\mathcal{A}$ be an* arbitrary *CMF dictionary in $\mathbb{R}^D$, $\mathcal{G}$ be an axis admissible grid with respect to $\mathcal{A}$, and $\mathcal{S}^\star \subseteq \mathcal{G}$ be a support of size k. We have:*

i) OMP achieves exact recovery of $\mathcal{G}$.

ii) OMP achieves q-delayed recovery of $\mathcal{S}^\star$ with $q = k^D$.

iii) Assume that

$$\max_{\theta\,\in\,\texttt{Cart}(\mathcal{S}^\star)\setminus\mathcal{S}^\star} \left\|\mathbf{G}^{-1}\mathbf{g}_\theta\right\|_1 \;<\; 1 \qquad \text{(3.30-CMF-ERC)}$$

where $\theta^\star_\ell, 1 \le \ell \le k$ denote the distinct elements of $\mathcal{S}^\star$, $\mathbf{G} \in \mathbb{R}^{k\times k}$ is the matrix with entries $\mathbf{G}[i,j] = \left\langle \boldsymbol{a}(\theta^\star_i), \boldsymbol{a}\left(\theta^\star_j\right)\right\rangle$ for $i,j \in [\![1,k]\!]$, and $\mathbf{g}_\theta \in \mathbb{R}^k$ is the vector with entries $\mathbf{g}_\theta[i] = \langle \boldsymbol{a}(\theta), \boldsymbol{a}(\theta^\star_i)\rangle$ for all $i \in [\![1,k]\!]$ and all $\theta \in \Theta$. Then OMP achieves exact recovery of $\mathcal{S}^\star$.
Vice-versa, if (3.30-CMF-ERC) does not hold, there exists (at least) one observation $\boldsymbol{y}$ with $\operatorname{supp}(\boldsymbol{y}) \subseteq \mathcal{S}^\star$ such that OMP with $\boldsymbol{y}$ as input selects some $\theta \in \texttt{Cart}(\mathcal{S}^\star) \setminus \mathcal{S}^\star$ at the first iteration.

Proof. Theorem 4 is a consequence of Theorem 3. We first state a result concerning the linear independence of a collection of atoms from a CMF dictionary:

Lemma 2. *Consider $\mathcal{A} = \{\boldsymbol{a}(\theta) \mid \theta \in \Theta\}$ an arbitrary CMF dictionary in dimension $D \ge 1$. Then for any support $\mathcal{S} = \{\theta_\ell\}_{\ell=1}^k \subset \Theta$ of k distinct parameters, the symmetric matrix $\mathbf{G}_\mathcal{S} \in \mathbb{R}^{k\times k}$ whose entries are*

$$\forall\, i,j \in [\![1,k]\!]\,, \qquad \mathbf{G}_\mathcal{S}[i,j] = \langle \boldsymbol{a}(\theta_i), \boldsymbol{a}(\theta_j)\rangle \tag{3.31}$$

is invertible, or, equivalently, the atoms $\{\boldsymbol{a}(\theta)\}_{\theta\in\mathcal{S}}$ are linearly independent.

Proof. See Appendix B.2. □

Property i). This property is a simple rephrasing of the statement of Theorem 3.

Property ii). Property *ii)* is also a rephrasing of Property *ii)* of Theorem 3: first, if $\boldsymbol{y}$ is supported in $\mathcal{S}^\star$ with $\mathcal{S}^\star \subseteq \mathcal{G}$ then OMP with input $\boldsymbol{y}$ only selects parameters in $\texttt{Cart}(\text{supp}(\boldsymbol{y}))$ at each iteration.
Note now that the atoms $\{\boldsymbol{a}(\theta) \mid \theta \in \texttt{Cart}(\mathcal{S}^\star)\}$ are linearly independent by Lemma 2. Therefore, the decomposition of $\boldsymbol{y}$ over $\texttt{Cart}(\mathcal{S}^\star)$ is unique and $\mathcal{S}^\star$ may be identified from the non-zero coefficients of $\widehat{\mathbf{c}} \in \mathbb{R}^{\text{card}(\texttt{Cart}(\mathcal{S}^\star))}$ (see (3.11)). Hence OMP identifies $\mathcal{S}^\star$ in at most $\text{card}(\texttt{Cart}(\mathcal{S}^\star)) \leq k^D$ steps.

Property iii). Let $\mathcal{S}^\star \triangleq \{\theta_\ell^\star\}_{\ell=1}^k \subset \mathcal{G}$ be a support of size k and $\boldsymbol{y}$ be supported in $\mathcal{S}^\star$. Let $\mathbf{G} \in \mathbb{R}^{k\times k}$ be the matrix whose i,j entry is $\mathbf{G}[i,j] = \left\langle \boldsymbol{a}(\theta_i^\star), \boldsymbol{a}\big(\theta_j^\star\big)\right\rangle$ for all $i,j \in [\![1,k]\!]$. Since the parameters in $\mathcal{S}^\star$ are pairwise distinct, the matrix $\mathbf{G}$ is invertible by Lemma 2.
By Property *ii)* of Theorem 3, OMP always selects parameters in $\texttt{Cart}(\mathcal{S}^\star)$ at each iteration. The problem can then be reduced to studying classical (discrete) sparse recovery with the following finite dictionary $\widetilde{\mathcal{A}} \triangleq \{\boldsymbol{a}(\theta) \mid \theta \in \texttt{Cart}(\mathcal{S}^\star)\}$ which can be decomposed as the disjoint union $\widetilde{\mathcal{A}} = \mathcal{A}_{\texttt{good}} \cup \mathcal{A}_{\texttt{bad}}$ where $\mathcal{A}_{\texttt{good}} \triangleq \{\boldsymbol{a}(\theta_\ell^\star)\}_{\ell=1}^k$ and $\mathcal{A}_{\texttt{bad}} \triangleq \{\boldsymbol{a}(\theta) \mid \theta \in \texttt{Cart}(\mathcal{S}^\star) \setminus \mathcal{S}^\star\}$ are the set of "true" and spurious atoms, respectively. Recall now that by Tropp's original results in the discrete setting [31, Th. 3.1], condition (3.30-CMF-ERC) precisely prevents from selecting parameters in $\mathcal{A}_{\texttt{bad}}$, *i.e.*, wrong parameters. Moreover, the least-square update of the coefficients at each iteration (see Line 4 in Algorithm 1) prevents from selecting twice the same parameters. Hence OMP recovers $\mathcal{S}^\star$ in k iterations.

Sharpness of the result. When (3.30-CMF-ERC) is not met, one can rely on Tropp's original result in the discrete setting [31, Th. 3.10] to obtain that: if (3.30-CMF-ERC) does not hold, there exists $\boldsymbol{y}_{\texttt{bad}} \in \text{span}(\{\boldsymbol{a}(\theta_\ell^\star)\}_{\ell=1}^k)$ such that

$$\max_{\theta \in \texttt{Cart}(\mathcal{S}^\star)\setminus\mathcal{S}^\star} |\langle \boldsymbol{a}(\theta), \boldsymbol{y}_{\texttt{bad}}\rangle| > \max_{\theta^\star \in \mathcal{S}^\star} |\langle \boldsymbol{a}(\theta^\star), \boldsymbol{y}_{\texttt{bad}}\rangle|. \tag{3.32}$$

In other words, OMP with input $\boldsymbol{y}_{\texttt{bad}}$ selects an element of $\texttt{Cart}(\mathcal{S}^\star) \setminus \mathcal{S}^\star$ □

We emphasize that result *ii)* of Theorem 4 can be seen as a worst-case bound on the number of iterations needed to achieve delayed recovery. The surprisingly new result of *iii)* can be interpreted as follows: whereas the parameter space is a continuum, exact recovery is possible provided that 1) the support is a subset of some axis admissible grid and 2) a condition on a *finite* number of atoms is fulfilled. Indeed, (3.30-CMF-ERC) only depends on a finite subset (namely $\texttt{Cart}(\mathcal{S}^\star)$) of the elements of Θ. Since the cardinality of $\texttt{Cart}(\mathcal{S}^\star)$ is finite, the numerical evaluation of (3.30-CMF-ERC) is possible. We emphasize that even if (3.30-CMF-ERC) holds for a set $\mathcal{S}^\star$, OMP may not achieve exact recovery of supports $\mathcal{S}$ which are *strictly included in* $\mathcal{S}^\star$. Indeed, it has been shown in [55, Th. 6] that, in general, the ERC condition is not locally nested. In other words, if the ERC holds for a set $\mathcal{S}^\star$, there may exist a support $\mathcal{S} \subsetneq \mathcal{S}^\star$ such that the ERC does not hold for $\mathcal{S}$. In practice this means that OMP will achieve k-delayed recovery of $\mathcal{S}$, but not exact recovery (*i.e.*, in $\operatorname{card}(\mathcal{S}) < k$ iterations). Finally, Theorem 4 is also consistent with Theorem 2 since condition (3.30-CMF-ERC) is always satisfied in dimension 1 with the convention that $\max_{\theta\in\emptyset}(\cdot) = 0$. Indeed, one can obviously check that $\mathtt{Cart}(\mathcal{S}^\star) = \mathcal{S}^\star$ when $D = 1$.

Theorem 4 can be exploited for Generalized Laplace dictionaries.

Definition 8 (Generalized Laplace kernel and Laplace dictionary)**.** *The class of* Generalized Laplace kernels *in dimension $D \geq 1$, denoted $\mathcal{K}_{\mathtt{Lap}}(D)$ consists of all kernels $\kappa : \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}_+^*$ such that*

$$\kappa(\theta, \theta') = \mathrm{e}^{-\lambda\left\|\theta-\theta'\right\|_p^p} \quad \forall \theta, \theta' \in \mathbb{R}^D \tag{3.33}$$

where $\lambda > 0$ and $0 < p \leq 1$.
By extension, a Generalized Laplace dictionary *in dimension $D \geq 1$ is a collection of atoms $\mathcal{A} = \{\boldsymbol{a}(\theta) \mid \theta \in \Theta\}$ where $\Theta \subset \mathbb{R}^D$, such that there exists $\kappa \in \mathcal{K}_{\mathtt{Lap}}(D)$ satisfying $\langle \boldsymbol{a}(\theta), \boldsymbol{a}(\theta')\rangle = \kappa(\theta, \theta')$ for all $\theta, \theta' \in \Theta$.*

One immediately sees that $\mathcal{K}_{\mathtt{Lap}}(D) \subset \mathcal{K}_{\mathtt{CMF}}(D)$ (see Definition 4) since the function $t \mapsto \mathrm{e}^{-\lambda t}$ defined on $\mathbb{R}_+$ is a CMF. The next result shows that for Generalized Laplace of dictionary, any Cartesian grid is axis admissible:

Theorem 5. *Consider $\mathcal{A}$ an* arbitrary *Generalized Laplace dictionary in $\mathbb{R}^D$. Then* all *Cartesian grids $\mathcal{G}$ are admissible with respect to $\mathcal{A}$.*

Proof. See Appendix A.2.2. □

Corollary 1. *Consider $\mathcal{A}$ an* arbitrary *Generalized Laplace dictionary in $\mathbb{R}^D$. For any support $\mathcal{S}^\star$ of any size $k \geq 1$, the two following properties hold:*

i) OMP achieves q-delayed recovery of $\mathcal{S}^\star$ with $q = k^D$.

ii) OMP achieves exact recovery of $\mathcal{S}^\star$ as soon as

$$\max_{\theta\,\in\,\mathtt{Cart}(\mathcal{S}^\star)\setminus\mathcal{S}^\star} \left\|\mathbf{G}^{-1}\mathbf{g}_\theta\right\|_1 \;<\; 1. \tag{3.34}$$

Proof. By Theorem 5, $\mathcal{G} = \mathtt{Cart}(\mathcal{S}^\star)$ is axis admissible. The result is then a direct consequence of Theorem 4. □

In other words, for Generalized Laplace dictionaries, *delayed* recovery holds *without any separation assumption* on the parameters. We can draw connections with the results from [56] where it is shown that one can identify any mixture of arbitrarily close Gaussians with computational complexity polynomial in the dimension. Moreover, k^D is probably a pessimistic upper bound on the number of potential global maximizers. However, further investigations are required to see if one can reach a bound linear in k on the number of iterations, as in [57] for finite dictionaries.

Axis admissible grids and CMF dictionaries. Theorem 5 shows that any Cartesian grid is axis admissible with respect to a Generalized Laplace dictionary. Such a result does not hold for general CMF dictionaries without extra assumptions on the grid. Nevertheless, empirical evidence suggests that the admissible grid assumption is only a proof artifact. We conjecture that Theorem 4 remains valid even when the Cartesian grid $\mathcal{G}$ is not axis admissible. To support our conjecture, we show in Appendix D that Result *iii)* of Theorem 4 still holds for *any* CMF dictionary when $\mathcal{G} = \mathtt{Cart}(\{\theta_1^\star, \theta_2^\star\})$ for arbitrary $\theta_1^\star, \theta_2^\star \in \mathbb{R}^D$, even though such $\mathcal{G}$ is generally not axis-admissible. The proof in the general case is still under investigation.

Minimum separation condition – coherence-based analysis. Example 4 showed that for CMF dictionaries, a too small separation between the elements of $\mathcal{S}^\star$ can prevent OMP from achieving exact recovery (note that, as just seen, *delayed* recovery may still be guaranteed for Generalized Laplace dictionaries without any minimum separation). Corollary 1 gives a necessary and sufficient algebraic condition, namely (3.34), such that OMP achieves exact recovery for Generalized Laplace dictionaries. Such a condition can be seen as an ERC (see (2.1 – ERC) from Section 2.1) associated to the finite dictionary $\{\boldsymbol{a}(\theta) \mid \theta \in \mathtt{Cart}(\mathcal{S}^\star)\}$. A coarser sufficient condition is expressed in terms of μ, the coherence of the latter dictionary, defined as

$$\mu \triangleq \max_{\theta_1, \theta_2 \in \mathtt{Cart}(\mathcal{S}^\star)} |\langle \boldsymbol{a}(\theta_1), \boldsymbol{a}(\theta_2) \rangle| \quad \text{s.t.} \quad \theta_1 \neq \theta_2. \tag{3.35}$$

For finite dictionaries, exact recovery is met whenever $\mu < (2k-1)^{-1}$, see (2.3) in Section 2.1. In the setting of Generalized Laplace dictionaries, the latter result can be expressed in terms of a *minimum separation* condition.

Theorem 6. *Consider $\mathcal{A}$ a Generalized Laplace dictionary in $\mathbb{R}^D$ and $\lambda > 0$, $0 < p \leq 1$ the parameters of the associated kernel (cf Equation (3.33)). Let $\mathcal{S}^\star$ be support of size k and denote*

$$\Delta_0 \triangleq \min_{d \in [\![1, D]\!]} \left|\theta_j^\star[d] - \theta_i^\star[d]\right| \qquad s.t. \qquad \begin{cases} i, j \in [\![1, k]\!] \\ \theta_i^\star[d] \neq \theta_j^\star[d] \end{cases}, \tag{3.36}$$

the smallest distance between non equal coordinates. If

$$\Delta_0^p \geq \frac{\log(2k-1)}{\lambda} \tag{3.37}$$

then for any observation $\boldsymbol{y}$ supported in $\mathcal{S}^\star$, OMP with input $\boldsymbol{y}$ achieves exact recovery.

Proof. By definition of Δ_0 and of $\mathtt{Cart}(\mathcal{S}^\star)$, we have $\|\theta_1 - \theta_2\|_p^p \geq \Delta_0^p$ for all $\theta_1, \theta_2 \in \mathtt{Cart}(\mathcal{S}^\star)$ hence by definition of the coherence we have $\mu \leq \exp(-\lambda \Delta_0^p)$ and (3.37) implies that the sufficient recovery condition $\mu < (2k-1)^{-1}$ holds. $\square$

Theorem 6 states that OMP recovers any combination of sufficiently separated atoms. Although condition (3.37) is expressed in terms of minimal distance between parameters, it can be seen as a condition on the coherence between atoms. In contrast to the discrete case, this coherence guarantee is only related to a particular finite subset of the (continuous) Generalized Laplace dictionary: the finite dictionary composed of atoms located at the Cartesian grid associated to the support.

Furthermore, Condition (3.37) can be compared to the separation condition for off-the-grid super-resolution (see (2.5)). More particularly, for the Laplace kernel (see Definition 8 with $p = 1$) and by assimilating the parameter λ to the cut-off frequency λ_c of some low-pass filter, condition (3.37) is similar to the *separation condition* [3] $\Delta_T \geq \frac{C_{st}}{\lambda_c}$ (Δ_T is a distance on the D-dimensional torus). At first glance, one may find condition (3.37) more demanding since it depends on the number of atoms k. However, the so-called *separation condition* is expressed on a D-dimensional torus preventing also high values of k. For instance, on the 1-dimensional torus, the minimum separation condition for BP requires $k \leq \frac{\lambda_c}{C_{st}} - 1$.

4. Conclusion - discussion

In this work, we have shown that the study of the recovery properties of greedy procedures such *Orthogonal Matching Pursuit* (OMP) can be extended to the setting of continuous dictionaries where atoms smoothly depend on some parameters. Capitalizing on the formulation of OMP in terms of inner products between atoms, our results rely on the properties of the kernel implicitly defined by the inner product between atoms. More particularly, we have identified two key notions of *admissible kernel* and *admissible support* that are sufficient to ensure exact recovery irrespectively of the value of the coefficients involved in the representation. For the class of CMF dictionaries, we have shown that when the dimension of the parameter space is 1, all implicitly defined kernels are admissible as well as all supports. The novelty of this result is that no separation is needed even for signed combination of atoms.

Although exact recovery can also be ensured for CMF dictionaries with a parameter space of higher dimension, more conditions have to be imposed on the support to be recovered, as some supports may not be admissible anymore. The cornerstone of our analysis in dimension higher than one is the notion of *axis admissible* Cartesian Grid. Indeed, axis admissibility is sufficient to allow OMP to identify supports, leading to a form of "delayed recovery" for all support of size k included in some admissible Cartesian grid. For such supports, exact recovery can also be achieved whenever a condition on a finite number of (known) atoms is fulfilled. In the special case of Generalized Laplace dictonaries, any Cartesian grid is in fact axis admissible, and a simplified *coherence-based analysis* can be revisited, leading to exact recovery under a *minimal separation condition.*

We now review some prospects of this work.

Beyond axis-admissible grids for CMF kernels. Our analysis for multi-dimensional parameter sets relies on the notion of axis-admissible grids. While axis-admissibility holds for any grid with respect to Generalized Laplace dictionaries, this is apparently no longer the case with respect to more general CMF dictionaries. Even for grids which seem to violate the axis-admissibility condition with respect to a CMF dictionary, empirical evidence suggest that Theorem 4 remains valid. As a first step towards a better understanding of this phenomenon, we showed in Appendix D that, for supports of size 2, axis-admissibility is not necessary for the conclusion *iii)* of Theorem 4 to hold.

Connection with TV-minimization. In light of the existing links between Tropp's ERC [31] and recovery guarantees for ℓ_1 minimization [34], an interesting question is whether the guarantees developed in this paper can be extended to sparse spike recovery with total variation norm minimization (see Section 2). More particularly, one could benefit from the *null space properties for measures* [27] that is sufficient to characterize the solution of the continuous version of Basis Pursuit. Such a connection would allow us to propose support recovery results for signed combination of atoms with TV-norm minimization without separation conditions.

Robustness to estimation error. One advantage of greedy procedures over convex relaxations is that the associated recovery guarantees involve solutions provided by actual *algorithms* rather than merely expressed as the minimizer of some *optimization problem.* In the continuous setting, this has to be tampered with the fact that actually implementing OMP requires a maximization procedure (*e.g.* by resorting to a gradient ascent procedure), at each iteration. Our current analysis does not take into account the resulting numerical estimation error. One could envision overcoming this limitation by analyzing the behavior of OMP when a small error is systematically done when maximizing the inner product (see Line 3 of Algorithm 1). Note that such an approximation error may also be useful to account for discretized implementations of the latter step of OMP using a fine grid over the parameter set Θ.

A. Technical details

This section contains the core of the proofs related to results given in Section 3. Appendix A.1 first details the proof of Theorem 1 for *admissible* kernels and support, see Definitions 1 and 2 from Section 3.2. Appendix A.2 contains the proofs related to CMF dictionaries.

A.1. Proof of Theorem 1

The proof of Theorem 1 relies on the following proposition.

Proposition 1. *Assume kernel κ is* admissible *and $\mathcal{S}^\star = \{\theta^\star_\ell\}_{\ell=1}^k \subset \Theta$ is admissible with respect to κ (see Definitions 1 and 2). Consider $\mathbf{G} \in \mathbb{R}^{k\times k}$ and*

$\mathbf{g}_\theta \in \mathbb{R}^k$ *for* $\theta \in \Theta$ *defined as follows*

$$\mathbf{G}[i,j] \triangleq \kappa\big(\theta_i^\star, \theta_j^\star\big) \qquad \forall i,j \in [\![1,k]\!] \tag{A.1}$$

$$\mathbf{g}_\theta[i] \triangleq \kappa(\theta, \theta_i^\star) \qquad \forall i \in [\![1,k]\!]. \tag{A.2}$$

Then $\mathbf{G}$ *is invertible and*

$$\forall\, \theta \notin \mathcal{S}^\star, \quad \left\|\mathbf{G}^{-1}\mathbf{g}_\theta\right\|_1 < 1. \tag{A.3}$$

The proof of this Proposition is postponed to the end of the section. Theorem 1 is now a corollary of Proposition 1.

Proof of Theorem 1. Let $\mathcal{S}^\star$ be an admissible support made of k pairwise distinct parameters. The hypotheses of Theorem 1 are the same as in Proposition 1. Therefore, Eq. (A.3) ensures that (2.1 – ERC) given in Section 2 is fulfilled for all $\theta \notin \mathcal{S}^\star$. Tropp's original proof that OMP only selects parameters in $\mathcal{S}^\star$ can thus be repeated to show by induction that at each step: a) the residual observation $\boldsymbol{r}$ is a linear combination of atoms supported in $\mathcal{S}^\star$; b) (2.1 – ERC) implies that OMP can only select a parameter in $\mathcal{S}^\star$; c) the residual error is different from $\mathbf{0}_\mathcal{H}$ during the k first iterations.
In addition, the least squares update of the coefficients ensures that $\langle \boldsymbol{a}(\widehat{\theta}), \boldsymbol{r}\rangle = 0$ for any $\widehat{\theta} \in \widehat{\mathcal{S}}$. Consequently, OMP never selects twice the same parameter and Property i) is proved.

Consider now an observation $\boldsymbol{y}$ whose support is included in $\mathcal{S}^\star$. Notice that, since $\texttt{Cart}(\mathrm{supp}(\boldsymbol{y})) \subset \mathcal{S}^\star$, $\mathrm{supp}(\boldsymbol{y})$ is also admissible by Fact 1. Property ii) then simply follows by applying Property i) to $\mathrm{supp}(\boldsymbol{y})$. □

The proof of Proposition 1 relies on two technical lemmas. First, we show that Property *i)* of Definition 2 implies linear independence between atoms.

Lemma 3. *Consider an integer* $k \geq 1$ *and* $\mathcal{S}^\star = \{\theta_\ell^\star\}_{\ell=1}^k \subsetneq \Theta$ *a set of* k *distinct parameters. If* κ *is admissible (see Definition 1) and* $\mathcal{S}^\star$ *satisfies Property i) of support admissibility (see Definition 2) then the atoms* $\{\boldsymbol{a}(\theta_\ell^\star)\}_{\ell=1}^k$ *are linearly independent.*

Proof. Let $\mathbf{c} = (c_1, \ldots, c_k)^\mathrm{T} \in \mathbb{R}^k$ be such that $\boldsymbol{y} \triangleq \sum_{\ell=1}^k c_\ell \boldsymbol{a}(\theta_\ell^\star) = \mathbf{0}_\mathcal{H}$, and let T be the set of indices such that $c_t \neq 0$. We will prove by contradiction that T is empty.
Assuming that T is not empty we first prove by contradiction that the sign of the coefficients $c_t, t \in T$ is not constant. Otherwise we would have (without loss of generality) $c_t > 0$ for all $t \in T$. Since $\boldsymbol{y} = \mathbf{0}_\mathcal{H}$, the function $f : \theta \mapsto \langle \boldsymbol{a}(\theta), \boldsymbol{y}\rangle$ is identically zero, hence any point of Θ is a maximizer. Since all entries of $\mathbf{c}$ are assumed to be positive, Property *i)* of Definition 2 applies and the maximizers of f belong to $\mathcal{S}^\star$, hence $\Theta \subset \mathcal{S}^\star$ which contradicts the definition of $\mathcal{S}^\star$ and proves that (if T is not empty) the sign of $c_t, t \in T$ is not constant.
We can thus partition T into two non-empty disjoint subsets $T_+ = \{t \in T, c_t > 0\}$, $T_- = \{t \in T, c_t < 0\}$. Defining $\boldsymbol{y}_\pm = \sum_{t \in T_\pm} c_t \boldsymbol{a}(\theta_t^\star)$ we note that $\boldsymbol{y} =$

$\boldsymbol{y}_+ + \boldsymbol{y}_-$ where $\boldsymbol{y}_\pm$ is supported by $\mathcal{S}^\star_\pm = \{\theta^\star_t\}_{t\in T_\pm}$. One deduces from $\boldsymbol{y} = \mathbf{0}_\mathcal{H}$ that $\boldsymbol{y}_+ = -\boldsymbol{y}_-$. Since $\mathcal{S}^\star_+, \mathcal{S}^\star_- \subset \mathcal{S}^\star$, Property *i)* of Definition 2 applies to both $\boldsymbol{y}_+$ and $\boldsymbol{y}_-$ hence any maximizer of $f_+ : \theta \mapsto \langle \boldsymbol{a}(\theta), \boldsymbol{y}_+\rangle = \langle \boldsymbol{a}(\theta), -\boldsymbol{y}_-\rangle$ belongs to $\mathcal{S}^\star_+ \cap \mathcal{S}^\star_-$. To conclude, we recall that the set of global maximizers of f_+ is not empty by Definition 1 (see discussion around (3.6)), however we have $\mathcal{S}^\star_+ \cap \mathcal{S}^\star_- = \emptyset$ since $T_+ \cap T_- = \emptyset$ and $\mathcal{S}^\star$ is made of distinct parameters. This yields the desired contradiction.
We have thus proved by contradiction that $T = \emptyset$, i.e. $\boldsymbol{y} = \mathbf{0}_\mathcal{H}$ implies $\mathbf{c} = \mathbf{0}_k$ so the atoms are linearly independent. □

The next lemma is a collection of results related to the invertibility of Gramian matrices associated to linearly independent atoms.

Lemma 4. *Let $\mathcal{S} = \{\theta_\ell\}_{\ell=1}^k \subset \Theta$ be such that the atoms $\{\boldsymbol{a}(\theta_\ell)\}_{\ell=1}^k$ are linearly independent. Consider $\mathbf{G} \in \mathbb{R}^{k\times k}$ and $\mathbf{g}_\theta \in \mathbb{R}^k$ for $\theta \in \Theta$ defined as follows*

$$\mathbf{G}[i,j] \triangleq \kappa(\theta_i, \theta_j) \qquad \forall i,j \in [\![1,k]\!] \tag{A.4}$$
$$\mathbf{g}_\theta[i] \triangleq \kappa(\theta, \theta_i) \qquad \forall i \in [\![1,k]\!]. \tag{A.5}$$

Then the following properties hold.

1. *The square matrix obtained by selecting the rows and columns of $\mathbf{G}$ indexed by any index set $\mathcal{I} \subset [\![1,k]\!]$ is invertible.*
2. *If $\mathbf{G}^{-1}\mathbf{g}_\theta$ has nonnegative entries, then*
$$\left\|\mathbf{G}^{-1}\mathbf{g}_\theta\right\|_1 = \mathbf{1}_k^\mathrm{T}\mathbf{G}^{-1}\mathbf{g}_\theta. \tag{A.6}$$

Proof. *1)* Each considered submatrix is the Gramian matrix of a linearly independent family of atoms, hence it is positive definite and therefore invertible. *2)* By Item 1, $\mathbf{G}$ is invertible. As $\mathbf{G}^{-1}\mathbf{g}_\theta$ has nonnegative entries we immediately deduce $\left\|\mathbf{G}^{-1}\mathbf{g}_\theta\right\|_1 = \mathbf{1}_k^\mathrm{T}\mathbf{G}^{-1}\mathbf{g}_\theta$. □

Proof of Proposition 1. In short we show that if kernel κ is and support $\mathcal{S}^\star$ are admissible then $\{\boldsymbol{a}(\theta^\star_\ell)\}_{\ell=1}^k$ are linearly independent, $\mathbf{G}^{-1}\mathbf{g}_\theta$ has nonnegative entries and $\mathbf{1}_k^\mathrm{T}\mathbf{G}^{-1}\mathbf{g}_\theta < 1$. The result then follows from Lemma 4-(A.6).

First, Definition *i)* ensures that the atoms $\{\boldsymbol{a}(\theta^\star_\ell)\}_{\ell=1}^k$ are indeed linearly independent by resorting to Lemma 3. We then prove by induction on k that: for all supports of size k that are admissible with respect to κ, we have

a) $\mathbf{G}^{-1}\mathbf{1}_k$ has nonnegative entries,

b) $\forall\, \theta \in \Theta\,, \quad \mathbf{G}^{-1}\mathbf{g}_\theta$ has nonnegative entries,

c) $\forall \theta \in \Theta \setminus \mathcal{S}^\star, \quad \mathbf{1}_k^\mathrm{T}\mathbf{G}^{-1}\mathbf{g}_\theta < 1$.

Before starting the proof, recall first that by definition of support admissibility, if $\mathcal{S}^\star$ is admissible with respect to κ, then any support $\mathcal{S} \subset \mathcal{S}^\star$ is also admissible.

Initialization: $k = 1$. Both $\mathbf{G}$ and $\mathbf{g}_\theta$ are scalars. Definition 1-*b) and* 1-*d)* gives $\mathbf{G} = 1$ (cf (3.4a)) and $\mathbf{g}_\theta \geq 0$, respectively, so that items *a)* and *b)* are both fulfilled. Definition 1-*d)* also gives $\mathbf{1}_k^{\mathrm{T}}\mathbf{G}^{-1}\mathbf{g}_\theta = \mathbf{g}_\theta = \kappa(\theta, \theta_1^\star) < 1$ for $\theta \neq \theta_1^\star$, so item *c)* is true.

Induction. We assume items *a)*-*b)*-*c)* hold for all admissible supports $\mathcal{S}$ of cardinality $k-1 \geq 1$. Considering $\mathcal{S}^\star$ an arbitrary admissible support of size $k \geq 2$, we show that items *a)*-*b)*-*c)* also hold for $\mathcal{S}^\star$.
Recall that all subsets of $\mathcal{S}^\star$ are also admissible. We consider $\overline{\mathcal{S}} = \{\theta_\ell^\star\}_{\ell=1}^{k-1} \subset \mathcal{S}^\star$ and use over-lined notations for quantities related to $\overline{\mathcal{S}}$: denote $\overline{\mathbf{G}} \in \mathbb{R}^{(k-1)\times(k-1)}$, $\overline{\mathbf{g}}_\theta \in \mathbb{R}^{k-1}$ the quantities given by (A.1), (A.2) for $\overline{\mathcal{S}}$ and $\mathbf{G} \in \mathbb{R}^{k\times k}, \mathbf{g}_\theta \in \mathbb{R}^k$ the same quantities for $\mathcal{S}^\star$. Likewise, the notations $\overline{\mathbf{g}}_\ell \in \mathbb{R}^{k-1}, \mathbf{g}_{\ell'} \in \mathbb{R}^k$ for $\ell = 1\ldots k-1$, $\ell' = 1\ldots k$ will refer to the columns of $\overline{\mathbf{G}}, \mathbf{G}$, respectively. With these notations we have:

$$\mathbf{g}_\theta = \begin{pmatrix} \overline{\mathbf{g}}_\theta \\ \kappa(\theta, \theta_k^\star) \end{pmatrix} \in \mathbb{R}^k \quad \forall \theta \in \Theta \tag{A.7}$$

$$\mathbf{G} = \begin{pmatrix} \overline{\mathbf{G}} & \overline{\mathbf{g}}_k \\ \overline{\mathbf{g}}_k^{\mathrm{T}} & 1 \end{pmatrix} \in \mathbb{R}^{k\times k}. \tag{A.8}$$

By linear independence between atoms, both $\mathbf{G}$ and $\overline{\mathbf{G}}$ are invertible according to Lemma 4.

Item a). We show that the last entry of $\mathbf{u} \triangleq \mathbf{G}^{-1}\mathbf{1}_k$ is positive. Since the reasoning holds for any ordering of the $\theta_\ell^\star$'s, we then deduce that all the entries of $\mathbf{u}$ are positive. Block inversion results [58, Corr 2.8.9] gives

$$\mathbf{G}^{-1} = \begin{pmatrix} \overline{\mathbf{G}}^{-1} + s\overline{\mathbf{G}}^{-1}\overline{\mathbf{g}}_k\overline{\mathbf{g}}_k^{\mathrm{T}}\overline{\mathbf{G}}^{-1} & -s\overline{\mathbf{G}}^{-1}\overline{\mathbf{g}}_k \\ -s\overline{\mathbf{g}}_k^{\mathrm{T}}\overline{\mathbf{G}}^{-1} & s \end{pmatrix}, \tag{A.9}$$

where $s \triangleq \left(1 - \overline{\mathbf{g}}_k^{\mathrm{T}}\overline{\mathbf{G}}^{-1}\overline{\mathbf{g}}_k\right)^{-1}$. Notice that

$$\begin{aligned} \overline{\mathbf{g}}_k^{\mathrm{T}}\overline{\mathbf{G}}^{-1}\overline{\mathbf{g}}_k &\leq \|\overline{\mathbf{g}}_k\|_\infty \left\|\overline{\mathbf{G}}^{-1}\overline{\mathbf{g}}_k\right\|_1 \\ &\overset{\text{Definition 1-}d)}{\leq} \left\|\overline{\mathbf{G}}^{-1}\overline{\mathbf{g}}_k\right\|_1 \\ &\underset{\textit{Lemma 4-2)}}{\overset{\text{Induction } b) \text{ and}}{=}} \mathbf{1}_{k-1}^{\mathrm{T}}\overline{\mathbf{G}}^{-1}\overline{\mathbf{g}}_k \overset{\text{Induction } c)}{<} 1. \end{aligned} \tag{A.10}$$

Hence $s > 0$.
The last entry of $\mathbf{u} = \mathbf{G}^{-1}\mathbf{1}_k$ now writes $\mathbf{u}[k] = s(1 - \overline{\mathbf{g}}_k^{\mathrm{T}}\overline{\mathbf{G}}^{-1}\mathbf{1}_{k-1})$. By (A.10), we know that $\mathbf{1}_{k-1}^{\mathrm{T}}\overline{\mathbf{G}}^{-1}\overline{\mathbf{g}}_k < 1$ and $s > 0$ hence $\mathbf{u}[k] > 0$. Since this holds for any ordering of the $\theta_\ell^\star$'s, we then deduce that all the entries of $\mathbf{u}$ are positive. Hence $\mathbf{G}^{-1}\mathbf{1}_k$ has nonnegative entries.

Item b). Again, we first show that the last entry of $\mathbf{v} \triangleq \mathbf{G}^{-1}\mathbf{g}_\theta$ is positive. With the decomposition of $\mathbf{G}$ given in (A.9) in mind, the last entry of $\mathbf{G}^{-1}\mathbf{g}_\theta$ writes

$$\mathbf{v}[k] = s\Big(\mathbf{g}_\theta[k] - \overline{\mathbf{g}}_k^{\mathrm{T}}\overline{\mathbf{G}}^{-1}\overline{\mathbf{g}}_\theta\Big) = s\Big(\kappa(\theta, \theta_k^\star) - \overline{\mathbf{g}}_k^{\mathrm{T}}\overline{\mathbf{G}}^{-1}\overline{\mathbf{g}}_\theta\Big), \tag{A.11}$$

with $s > 0$ (see (A.10)).

Then, it is sufficient to show that $\kappa(\theta, \theta_k^\star) - \overline{\mathbf{v}}^{\mathrm{T}}\overline{\mathbf{g}}_\theta \geq 0$ where $\overline{\mathbf{v}} \triangleq \overline{\mathbf{G}}^{-1}\overline{\mathbf{g}}_k$ to show that $\mathbf{v}[k] \geq 0$. This will be achieved by studying this quantity seen as a function of θ. Consider $T \subset [\![1, k-1]\!]$ the (possibly empty) set defined by $T \triangleq \{i \mid \overline{\mathbf{v}}[i] \neq 0\}$ and define

$$\begin{aligned} \psi_1 \colon \mathbb{R} &\longrightarrow \mathbb{R}_+ \\ \theta &\longmapsto \overline{\mathbf{v}}^{\mathrm{T}}\overline{\mathbf{g}}_\theta = \textstyle\sum_{\ell=1}^{k-1} \overline{\mathbf{v}}[\ell]\kappa(\theta, \theta_\ell^\star) = \sum_{\ell \in T} \overline{\mathbf{v}}[\ell]\kappa(\theta, \theta_\ell^\star). \end{aligned} \tag{A.12}$$

Notice that

- the entries of $\overline{\mathbf{v}}$ are nonnegative by the induction hypothesis *b)*. Moreover, by the induction hypothesis *c)*, we have $\sum_{\ell=1}^{k-1} \overline{\mathbf{v}}[\ell] = \mathbf{1}_{k-1}^{\mathrm{T}}\overline{\mathbf{v}} < 1$.
- For $j \in [\![1, k-1]\!]$ and $\theta = \theta_j^\star$ we have $\overline{\mathbf{g}}_\theta = \overline{\mathbf{g}}_j = \overline{\mathbf{G}}\overline{\mathbf{e}}_j$, the j-th column of $\overline{\mathbf{G}}$ where $\overline{\mathbf{e}}_j$ is the j-th canonical vector of $\mathbb{R}^{k-1}$. Hence we have $\psi_1(\theta_j^\star) = \overline{\mathbf{g}}_k^{\mathrm{T}}\overline{\mathbf{G}}^{-1}\overline{\mathbf{g}}_j = \overline{\mathbf{g}}_k^{\mathrm{T}}\overline{\mathbf{e}}_j = \kappa\big(\theta_j^\star, \theta_k^\star\big)$.

If T is not empty, since $T \subset [\![1, k-1]\!] \subsetneq [\![1, k]\!]$, using that $\mathcal{S}^\star$ is admissible with respect to κ, we obtain from Property *ii)* of Definition 2 with $t' = k$

$$\kappa(\theta, \theta_k^\star) - \overline{\mathbf{v}}^{\mathrm{T}}\overline{\mathbf{g}}_\theta = \kappa(\theta, \theta_k^\star) - \psi_1(\theta) \geq 0$$

for all $\theta \in \Theta$. The same obviously holds if T is empty as $\psi_1(\theta)$ is identically zero and κ is non-negative by Definition 1-*d)*. Since this result does not depend on the ordering of the $\theta_\ell^\star$'s, the same reasoning applies for all $\ell \in [\![1, k]\!]$ so $\mathbf{G}^{-1}\mathbf{g}_\theta \in \mathbb{R}_+^k$, hence $\mathbf{G}^{-1}\mathbf{g}_\theta$ has nonnegative entries.

Item c). Since $\mathbf{G}$ is invertible by Lemma 4-1) we can define

$$\begin{aligned} \psi_2 \colon \Theta &\longrightarrow \mathbb{R} \\ \theta &\longmapsto \mathbf{1}_k^{\mathrm{T}}\mathbf{G}^{-1}\mathbf{g}_\theta \end{aligned} \tag{A.13}$$

and we just need to prove that $\psi_2(\theta) < 1$ for all $\theta \notin \mathcal{S}^\star$.
We have already shown in point *a)* that the vector $\mathbf{u} \triangleq \mathbf{G}^{-1}\mathbf{1}_k \in \mathbb{R}^k$ has nonnegative entries, and $\psi_2(\theta)$ writes

$$\psi_2(\theta) \triangleq \sum_{\ell=1}^{k} \mathbf{u}[\ell]\kappa(\theta, \theta_\ell^\star). \tag{A.14}$$

Consider $\varepsilon > 0$, and define $\widetilde{\mathbf{u}} \triangleq (\|\mathbf{u}\|_1 + \varepsilon)^{-1}\mathbf{u} \in \mathbb{R}^k$ and note that $\widetilde{\mathbf{u}}$ has also nonnegative entries. Let θ_0 be a maximizer of ψ_2. Then θ_0 is also a maximizer of

$$\widetilde{\psi}_2(\theta) \triangleq \sum_{\ell=1}^{k} \widetilde{\mathbf{u}}[\ell]\kappa(\theta, \theta_\ell^\star). \tag{A.15}$$

Denote $T \triangleq \{i \mid \widetilde{\mathbf{u}}[i] \neq 0\} \subset [\![1, k]\!]$ the set of nonzero entries of $\widetilde{\mathbf{u}}$. Since $\mathbf{G}\mathbf{u} = \mathbf{1}_k$ is nonzero and $\mathbf{G}$ is injective, T is non-empty and $\sum_{\ell \in T} \widetilde{\mathbf{u}}[\ell] < 1$ by construction, Property *i)* of support admissibility (see Definition 2) ensures that θ_0 belongs to $\mathcal{S}^\star$. In addition,

$$\forall\, \ell = 1 \ldots k\,, \quad \mathbf{1}_k^{\mathrm{T}}\mathbf{G}^{-1}\mathbf{g}_{\theta_\ell^\star} = \mathbf{1}_k\mathbf{e}_\ell = 1, \tag{A.16}$$

where $\{\mathbf{e}_\ell\}_{\ell=1}^{k}$ is the canonical basis of $\mathbb{R}^k$. Therefore, $\psi_2(\theta) < 1$ for all $\theta \notin \mathcal{S}^\star$. This ends the proof. $\square$

A.2. Proofs related to CMF dictionaries

All proofs in this section relies on the Bernstein-Widder theorem, see Lemma 1 (in Section 3.3).

A.2.1. Proof of Theorem 2 - Recovery in dimension 1

This section is dedicated to the proof of Theorem 2 and exploits the results of Theorem 1, *i.e.*, we show that when $\mathcal{A}$ is a CMF dictionary associated to kernel $\kappa(\theta, \theta')$, $\theta, \theta' \in \mathbb{R}$, then a) kernel κ is admissible in the sense of Definition 1 and b) all supports $\mathcal{S}^\star$ of size k are admissible with respect to kernel κ in the sense of Definition 2. To that aim, let $\mathcal{S}^\star = \{\theta_\ell^\star\}_{\ell=1}^{k}$ be k distinct parameters of Θ and κ be an element of $\mathcal{K}_{\texttt{CMF}}(1)$ (cf. Definition 4).

Kernel κ is admissible (cf. Definition 1). Recall that the kernel κ originates from a CMF that equals to 1 at 0, is continuous and strictly decreasing. Hence Definition 1 is verified by construction of $\mathcal{K}_{\texttt{CMF}}(1)$.

To show that $\mathcal{S}^\star$ is admissible with respect to κ, we consider a non-empty subset of indices $T \subseteq [\![1, k]\!]$ with $t = \mathrm{card}(T)$. Without loss of generality (up to a global renumbering), assume that $T = [\![1, t]\!]$. Let $c_1 \ldots c_t > 0$ be t positive coefficients satisfying $\sum_{\ell=1}^{t} c_\ell < 1$ and consider the function

$$\begin{aligned} \psi \colon \mathbb{R} &\longrightarrow \mathbb{R}_+ \\ \theta &\longmapsto \sum_{\ell=1}^{t} c_\ell \kappa(\theta, \theta_\ell^\star). \end{aligned} \tag{A.17}$$

Using the integral formulation of CMF (see Lemma 1), we have that ψ is twice differentiable [52, proof of Theorem 12a] at any $\theta \in \Theta \setminus \{\theta^\star_\ell\}_{\ell\in T}$ and

$$\psi''(\theta) = \sum_{\ell=1}^{t} p(1-p)c_\ell \int_{[0,+\infty[} u|\theta^\star_\ell - \theta|^{p-2} \mathrm{e}^{-u|\theta^\star_\ell-\theta|^p}\,\mathrm{d}\nu(u) \\ +\ p^2 c_\ell \int_{[0,+\infty[} u^2|\theta^\star_\ell - \theta|^{2(p-1)} \mathrm{e}^{-u|\theta^\star_\ell-\theta|^p}\,\mathrm{d}\nu(u) \quad \text{(A.18)}$$

for some probability measure $\nu \neq \delta_0$ (determined by the CMF associated to κ). We now successively verify Property *i)* and *ii)* of Definition 2.

Property i) of Definition 2. Recall first that the vanishing property of admissible kernels (see Definition 1) ensures that ψ admits at least one maximizer. We show now that if $\theta \in \Theta$ is a maximizer of ψ, then necessarily $\theta \in \mathcal{S}^\star$.
Consider $\theta \in \Theta \setminus \{\theta^\star_\ell\}_{\ell\in T}$ a critical point of ψ. Since $\theta \notin \{\theta^\star_\ell\}_{\ell\in T}$, ψ is twice differentiable with $\psi''(\theta)$ given in (A.18). Each integral term appearing in (A.18) is positive since ν is not null, satisfies $\nu \neq \delta_0$, and $\theta \notin \{\theta^\star_\ell\}_{\ell\in T}$. Since $p \in]0,1]$ and $c_\ell > 0$ for each ℓ, it follows that $\psi''(\theta) > 0$. Therefore, θ is a local *minimizer* of ψ. Hence any maximizer of ψ belongs to $\{\theta^\star_\ell\}_{\ell\in T}$.

The verification of Property *ii)* of Definition 2 relies on a technical lemma that exploits the notion of "sign changes of a finite sequence". The latter notion is defined as the number of times two consecutive elements of the finite sequence have opposite signs; For instance, the sequence $(1,1,-1,1)$ has two sign changes (third and fourth positions).

Lemma 5. *Let $P(u) \triangleq \sum_{\ell=1}^{k} c_\ell \mathrm{e}^{-\lambda_\ell u}$ be an exponential polynomial on $\mathbb{R}_+$ with $0 < \lambda_1 < \ldots < \lambda_k$ and nonzero $c_1,\ldots,c_k$. Assume that:*

- *the sequence $c_1 \ldots c_k$ has at most two sign changes;*
- *$P(0) < 0$ and $\lim_{u\longrightarrow+\infty} P(u) = 0_+$.*

Then there exists $u_0 > 0$ for which the following inequality holds

$$\int_{[0,+\infty[} f(u)P(u)\,\mathrm{d}\nu(u)\ \geq f(u_0)\int_{[0,+\infty[} P(u)\,\mathrm{d}\nu(u). \quad \text{(A.19)}$$

for any non-decreasing function f on $\mathbb{R}_+$ and any (unsigned) finite Borel measure ν on $\mathbb{R}_+$ such that the integrals converge.

The proof this lemma is postponed to Appendix B.1.

Property ii) of Definition 2. Here $T \neq [\![1,k]\!]$ hence $t = \mathrm{card}(T) \leq k-1$, and we consider $t' \in [\![1,k]\!] \setminus T$. Without loss of generality (up to a renumbering in the complement of T) we can assume that $T = [\![1,t]\!]$ and $t' = t+1$. In addition, the coefficients are such that

$$s \triangleq \sum_{\ell\in T} c_\ell < 1 \quad \text{(A.20)}$$

by hypothesis. Consider

$$\begin{aligned} \phi\colon \mathbb{R} &\longrightarrow \mathbb{R} \\ \theta &\longmapsto \psi(\theta) - \kappa\big(\theta, \theta^\star_{t+1}\big). \end{aligned} \tag{A.21}$$

From the definition of ψ in (A.17), ϕ writes

$$\phi(\theta) = \sum_{\ell=1}^{t} c_\ell \varphi\Big(|\theta - \theta^\star_\ell|_p^p\Big) - \varphi\Big(\big|\theta - \theta^\star_{t+1}\big|_p^p\Big) \tag{A.22}$$

for all $\theta \in \Theta$. Assume that $\phi(\theta^\star_\ell) \leq 0$ for every $\ell \in T$. See first that $\phi(\theta) = 0$ for all $\theta \in \{\theta^\star_\ell\}_{\ell=1}^t$ by choice of $c_1 \ldots c_t$. We will show below that $\phi(\theta) \leq 0$ for any $\theta \in \Theta \setminus \{\theta^\star_\ell\}_{\ell=1}^t$ that is a local maximizer of ϕ. As a consequence, $\phi(\theta) \leq 0$ for any $\theta \in \Theta$, therefore establishing that Property *ii)* of Definition 2 is satisfied.

Consider $\theta \in \Theta \setminus \{\theta^\star_\ell\}_{\ell=1}^t$ a local maximizer of ϕ. Assume first that (up to a renumbering within the set T) $|\theta^\star_1 - \theta| < |\theta^\star_2 - \theta| < \ldots < |\theta^\star_t - \theta|$. The equality cases will be addressed later. For the sake of readability, denote $\lambda_\ell \triangleq |\theta^\star_\ell - \theta|$ for all $\ell \in [\![1, t+1]\!]$ and recall that we have $\lambda_1 < \cdots < \lambda_t$. We distinguish three cases.

Case 1: $\lambda_{t+1} \leq \lambda_1 = \min_{1\leq i \leq t} \lambda_i$. Therefore we have $\max_{1\leq i\leq t} \mathrm{e}^{-u\lambda_i^p} \leq \mathrm{e}^{-u\lambda_{t+1}^p}$ for any $u \geq 0$. Hence (even without using the fact that θ is a local maximizer)

$$\phi(\theta) \leq \underbrace{\left(\sum_{i=1}^t c_i - 1\right)}_{<0 \text{ by hyp.}} \underbrace{\int_{[0,+\infty[} \mathrm{e}^{-u\lambda_{t+1}^p}\,\mathrm{d}\nu(u)}_{>0} < 0.$$

Note that Case 1 covers the situation where $\theta = \theta^\star_{t+1}$.

Case 2: $\lambda_{t+1} > \lambda_t = \max_{1\leq i\leq t} \lambda_i$. Given that $\theta \in \Theta \setminus \{\theta^\star_\ell\}_{\ell=1}^{t+1}$, we show below that $\phi(\theta) < C\phi''(\theta)$ for some positive constant $C > 0$. Since θ is a local maximizer of ϕ, we have $\phi''(\theta) \leq 0$ and therefore $\phi(\theta) < 0$.
Since $\theta \in \Theta \setminus \{\theta^\star_\ell\}_{\ell=1}^{t+1}$, ϕ is twice differentiable at θ. Using (A.18) we get $\phi'' = \phi''_1 + \phi''_2$ with

$$\begin{aligned} \phi''_1(\theta) &= p(1-p)\int_{[0,+\infty[} u\left(\sum_{i=1}^t c_i\lambda_i^{p-2}\mathrm{e}^{-u\lambda_i^p} - \lambda_{t+1}^{p-2}\mathrm{e}^{-u\lambda_{t+1}^p}\right)\mathrm{d}\nu(u) \\ \phi''_2(\theta) &= p^2\int_{[0,+\infty[} u^2\left(\sum_{i=1}^t c_i\lambda_i^{2(p-1)}\mathrm{e}^{-u\lambda_i^p} - \lambda_{t+1}^{2(p-1)}\mathrm{e}^{-u\lambda_{t+1}^p}\right)\mathrm{d}\nu(u). \end{aligned}$$

Using that $\frac{\lambda_{t+1}}{\lambda_i} > 1$ for each $1 \leq i \leq t$, we have

$$
\begin{aligned}
\phi_2''(\theta) &= \frac{p^2}{\lambda_{t+1}^{2(1-p)}} \int_{[0,+\infty[} u^2 \left(\sum_{i=1}^{t} c_i \left(\frac{\lambda_{t+1}}{\lambda_i} \right)^{2(1-p)} \mathrm{e}^{-u\lambda_i^p} - \mathrm{e}^{-u\lambda_{t+1}^p} \right) \mathrm{d}\nu(u) \\
&> \frac{p^2}{\lambda_{t+1}^{2(1-p)}} \int_{[0,+\infty[} u^2 \left(\sum_{i=1}^{t} c_i \mathrm{e}^{-u\lambda_i^p} - \mathrm{e}^{-u\lambda_{t+1}^p} \right) \mathrm{d}\nu(u). \qquad \text{(A.23)}
\end{aligned}
$$

Note now that

- the function $u \longmapsto u^2$ is increasing,
- $P(u) \triangleq \sum_{i=1}^{t} c_i \mathrm{e}^{-u\lambda_i^p} - \mathrm{e}^{-u\lambda_{t+1}^p}$ is an exponential polynomial whose sequence of coefficients is $(c_1, \ldots, c_t, -1)$ has exactly one sign change.
- As $\max_{1 \leq i \leq t} \lambda_i < \lambda_{t+1}$ by hypothesis, we have $P(u) > 0$ for sufficiently large u so $\lim_{u \to +\infty} P(u) = 0^+$.
- As $\sum_{i=1}^{t} c_i < 1$ we have $P(0) < 0$.

Then, by Lemma 5, there exists $u_0 > 0$ such that

$$
\phi_2''(\theta) > \frac{p^2}{\lambda_{t+1}^{2(1-p)}} u_0^2 \int_{[0,+\infty[} P(u)\, \mathrm{d}\nu(u) = \frac{p^2}{\lambda_{t+1}^{2(1-p)}} u_0^2 \phi(\theta). \qquad \text{(A.24)}
$$

This establishes that $\phi_2''(\theta) > C_2 \phi(\theta)$ where $C_2 > 0$ is a positive constant. The same rationale leads to $\phi_1''(\theta) > C_1 \phi(\theta)$ with $C_1 \geq 0$ ($C_1 = 0$ for $p = 1$ since ϕ_1'' is identically zero). Since $\phi'' = \phi_1'' + \phi_2''$, one gets $\phi''(\theta) > (C_1 + C_2)\phi(\theta)$, which concludes the case since $C = C_1 + C_2 > 0$.

Case 3: *other situations.* Here, $\lambda_1 < \lambda_{t+1} \leq \lambda_t$. There exists $i \in [\![1, t-1]\!]$ such that $\lambda_\ell < \lambda_{t+1}$ for $1 \leq \ell \leq i$ and $\lambda_\ell \geq \lambda_{t+1}$ otherwise. Denote $\varepsilon \triangleq 1 - s > 0$ where s has been defined in (A.20) and let $s_1 \triangleq \sum_{\ell=1}^{i} c_\ell + \frac{\varepsilon}{2}$ and $s_2 \triangleq \sum_{\ell=i+1}^{t} c_\ell + \frac{\varepsilon}{2}$ such that $s_1 + s_2 = 1$. One can write

$$
\begin{aligned}
\phi(\theta) &= s_1 \underbrace{\int_{[0,+\infty[} \left(\sum_{\ell=1}^{i} \frac{c_\ell}{s_1} \mathrm{e}^{-u\lambda_\ell^p} - \mathrm{e}^{-u\lambda_{t+1}^p} \right) \mathrm{d}\nu(u)}_{\triangleq\, \psi_1(\theta)} \\
&\qquad + s_2 \underbrace{\int_{[0,+\infty[} \left(\sum_{\ell=i+1}^{t} \frac{c_\ell}{s_2} \mathrm{e}^{-u\lambda_\ell^p} - \mathrm{e}^{-u\lambda_{t+1}^p} \right) \mathrm{d}\nu(u)}_{\triangleq\, \psi_2(\theta)}. \qquad \text{(A.25)}
\end{aligned}
$$

Notice now that

- $\lambda_{t+1} > \lambda_\ell$ for all $\ell \in [\![1, i]\!]$ and $\sum_{\ell=1}^{i} \frac{c_\ell}{s_1} < 1$. By resorting to case 2, we have $\psi_1(\theta) < 0$.

- $\lambda_{t+1} \leq \lambda_\ell$ for all $\ell \in [\![i+1, t]\!]$ and $\sum_{\ell=i+1}^{t} \frac{c_\ell}{s_2} < 1$. By resorting to case 1, we have $\psi_2(\theta) < 0$.

Hence we have $\phi(\theta) \leq s_1\psi_1(\theta) + s_2\psi_2(\theta) < 0$, which concludes the case.

Equality case. To deal with equality cases we denote $0 < \lambda_1 < \ldots < \lambda_{k'}$, with $k' \leq t$, the ordered distinct values in $\{|\theta_i^\star - \theta|\}_{i=1}^t$, and define d_i the sum of coefficients c_ℓ over all indices $1 \leq \ell \leq t$ such that $|\theta_\ell^\star - \theta| = \lambda_i$. We have $d_i > 0$ and $s = \sum_{i=1}^{k'} d_i = \sum_{\ell=1}^{k} c_\ell < 1$. Define $d_{k'+1} = -1$ and $\lambda_{k'+1} = \left|\theta_{t+1}^\star - \theta\right|$. We can then show that Property *ii)* of Definition 2 holds by applying the previous reasoning to $\lambda_1 \ldots \lambda_{k'+1}$ and $d_1 \ldots d_{k'+1}$.

Conclusion. We have shown that κ is admissible and $\mathcal{S}^\star$ is admissible with respect to κ. Theorem 1 applies and OMP recovers $\mathcal{S}^\star$ in at most k steps. $\square$

A.2.2. Proof of Theorems 3 and 5 - Recovery in Dimension D

Proof of Theorem 3. Let $\mathcal{G}$ be an axis admissible Cartesian grid. By definition (see Definition 6), there exists a family of finite one-dimensional sets $\{\mathcal{S}_d\}_{d=1}^D$ such that $\mathcal{G} = \prod_{d=1}^D \mathcal{S}_d$. Moreover, since the observation $\boldsymbol{y}$ is supported in $\mathcal{G} \triangleq \{\theta_\ell^\star\}_{\ell=1}^{\mathrm{card}(\mathcal{G})}$, there exists $c_1 \ldots c_{\mathrm{card}(\mathcal{G})} \neq 0$ such that $\boldsymbol{y} = \sum_{\ell=1}^{\mathrm{card}(\mathcal{G})} c_\ell \boldsymbol{a}(\theta_\ell^\star)$.

Correct selection at the first iteration. Consider now θ_m a global maximizer of the function $\theta \mapsto |\langle \boldsymbol{a}(\theta), \boldsymbol{y}\rangle|$. Then, $t = \theta_m[d]$ is also a maximizer of

$$\begin{aligned} f_d \colon \mathbb{R} &\longrightarrow \mathbb{R}_+ \\ t &\longmapsto |\langle \boldsymbol{a}(\theta_m + (t - \theta_m[d])\mathbf{e}_d), \boldsymbol{y}\rangle|, \end{aligned} \tag{A.26}$$

where $\mathbf{e}_d$ is the d-th element of the canonical basis of $\mathbb{R}^D$. Denoting $\theta_0 \triangleq \theta_m - \theta_m[d]\mathbf{e}_d$, we have $\theta_0 \perp \mathbf{e}_d$ by construction and f_d writes

$$f_d(t) = \left|\sum_{\ell=1}^{\mathrm{card}(\mathcal{G})} c_\ell \kappa(\theta_0 + t\mathbf{e}_d, \theta_\ell^\star)\right|. \tag{A.27}$$

Since $\mathcal{G}$ is axis admissible with respect to the kernel κ (see Definition 7), the maximizers of f_d belong to $\mathcal{S}_d$. Hence $\theta_m[d] \in \mathcal{S}_d$.

Finally, the latter result holds for all values of $d \in [\![1, D]\!]$; we deduce that θ_m belongs to $\prod_{d=1}^D \mathcal{S}_d = \mathcal{G}$. Hence OMP selects an element in $\mathcal{G}$ at the first iteration.

Correct selection at all iterations. Since the residual error after the first iteration is still supported in $\mathcal{G}$, the same rationale can be repeated at the second iteration. Therefore, a recursive application of the previous result ensures that OMP with input $\boldsymbol{y}$ only selects parameters in $\mathcal{G}$ at each iteration. Moreover, the Least Squares update of the coefficients (see Line 4 of Algorithm 1) prevents from selecting twice the same parameters. Hence OMP recovers $\mathcal{G}$ in $\mathrm{card}(\mathcal{G})$ steps.

Proof of Property ii). Let $\boldsymbol{y}$ be an observation whose support is included in $\mathcal{G}$. Since $\texttt{Cart}(\text{supp}(\boldsymbol{y})) \subset \mathcal{G}$ it follows from Fact 2 that $\texttt{Cart}(\text{supp}(\boldsymbol{y}))$ is also axis admissible. Reasoning as in the proof of Property *i)* yields that OMP with input $\boldsymbol{y}$ selects parameters in $\texttt{Cart}(\text{supp}(\boldsymbol{y}))$ at each iteration, until the residual vanishes. □

Proof of Theorem 5. Let $\mathcal{G} = \{\theta_\ell\}_{\ell=1}^q$ where $q = \text{card}(\mathcal{G})$ be an arbitrary Cartesian grid. Consider real coefficients $c_1, \ldots, c_q$ not all equal to 0. Let finally $d \in [\![1, D]\!]$ and $\theta_0 \in \mathbb{R}^D$ such that $\theta_0[d] = 0$ and define function f_d

$$\begin{aligned} f_d \colon \mathbb{R} &\longrightarrow \mathbb{R}_+ \\ t &\longmapsto \left| \sum_{\ell=1}^{q} c_\ell \kappa(\theta_0 + t\mathbf{e}_d, \theta_\ell) \right| = \left| \sum_{\ell=1}^{q} c_\ell \mathrm{e}^{-\lambda \|\theta_0 + t\mathbf{e}_d - \theta_\ell\|_p^p} \right|, \end{aligned} \tag{A.28}$$

as in the statement of Definition 7. One sees that f_d rewrites

$$\begin{aligned} f_d(t) &= \left| \sum_{\ell=1}^{q} c_\ell \mathrm{e}^{-\lambda |t - \theta_\ell[d]|^p - \lambda \sum_{j=1, j\neq d}^{D} |\theta_0[j] - \theta_\ell[j]|^p} \right| \\ &= \left| \sum_{\ell=1}^{q} \widetilde{c}_\ell \mathrm{e}^{-\lambda |t - \theta_\ell[d]|^p} \right| \end{aligned} \tag{A.29}$$

where $\widetilde{c}_\ell \triangleq c_\ell \mathrm{e}^{-\lambda \sum_{j=1, j\neq d}^{D} |\theta_0[j] - \theta_\ell[j]|^p}$ for all $\ell \in [\![1, k]\!]$.

Let $\mathcal{A}_0 = \{\boldsymbol{a}_0(u),\ u \in \mathbb{R}\}$ be a Generalized Laplace dictionary in dimension 1 in the sense of Definition 8. Then f_d may also be interpreted as the inner product between atom $\boldsymbol{a}_0(t) \in \mathcal{A}_0$ and input signal $\boldsymbol{y}_0 = \sum_{\ell=1}^{k} \widetilde{c}_\ell \boldsymbol{a}_0(\theta_\ell[d])$. By Theorem 2 the global maximizer of f_d belongs to $\{\theta_\ell[d]\}_{\ell=1}^k$. As this holds for any d and θ_0 such that $\theta_0[d] = 0$, this establishes that $\mathcal{G}$ is axis-admissible. □

B. Miscellaneous

B.1. Proof of Lemma 5

The key ingredient of Lemma 5 is the following lemma:

Lemma 6 (Laguerre's generalization of Descartes's rule of signs [59], p. 319)**.** *Let $a_1, \ldots, a_k$ be nonzero real coefficients and $0 < x_1 < \cdots < x_k$ be real numbers. Let z be the number of real roots of the function $P(t) = \sum_{\ell=1}^{k} a_\ell x_\ell^t$, and n_c be the number of changes in sign in the sequence of numbers $a_1, \ldots, a_k$. Then $z \leq n_c$.*

The sequence of coefficients $a_\ell = c_{k+1-\ell}$ with $\ell \in [\![1, k]\!]$ has only two sign changes by hypothesis. By applying Lemma 6 with $x_\ell = \mathrm{e}^{-\lambda_{k+1-\ell}}$, one sees that P has at most two real roots, so at most two sign changes on $\mathbb{R}_+$. However, P must satisfy the following constraints

i) P is continuous on $[0, +\infty[$,

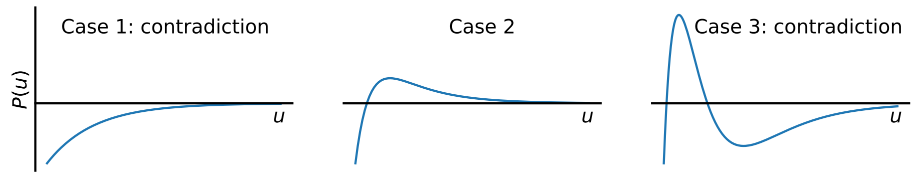

Figure B.2: Shape of P (see proof of Lemma 5) with constraints i) P is continuous, ii) $P(u) < 0$ and iii) $\exists u_0 > 0$ such that $P(u) > 0$ for all $u > u_0$. One see that the constraints cannot be satisfied in case 1 and 3.

ii) $P(0) < 0$,

iii) there exists $u_0 > 0$ such that for all $P(u) > 0$ for all $u > u_0$.

As illustrated in Figure B.2, these three constraints cannot be verified simultaneously if P has exactly 0 or 2 roots.

Thus P has exactly one sign change on $\mathbb{R}_+$ and there exists $u_0 > 0$ such that $u < u_0 \implies P(u) < 0$ and $u > u_0 \implies P(u) > 0$. One then has, for any non-decreasing function f and any (non-negative) measure ν on $\mathbb{R}_+$

$$\begin{aligned}\int f(u)P(u)\,\mathrm{d}\nu(u) &= \int_{[0,u_0]} \underbrace{f(u)}_{\text{non-decreasing}} \underbrace{P(u)}_{\leq 0} \mathrm{d}\nu(u) + \int_{]u_0,+\infty[} \underbrace{f(u)}_{\text{non-decreasing}} \underbrace{P(u)}_{\geq 0} \mathrm{d}\nu(u) \\ &\geq \int_{[0,u_0]} f(u_0)P(u)\,\mathrm{d}\nu(u) + \int_{]u_0,+\infty[} f(u_0)P(u)\,\mathrm{d}\nu(u) \\ &= f(u_0)\int_{[0,+\infty[} P(u)\,\mathrm{d}\nu(u).\end{aligned}$$

□

B.2. Proof of Lemma 2 - Invertibility of Gram matrices

In probability theory, the characteristic function of a given vector-valued random variable $\mathbf{Z} \in \mathbb{R}^D$ with probability distribution P is the function $\mathbf{x} \in \mathbb{R}^D \mapsto \mathbb{E}_{\mathbf{Z}\sim P}[\mathrm{e}^{\mathrm{i}\mathbf{x}^T\mathbf{Z}}]$ where $\mathbb{E}$ denotes the expected value and i the imaginary number. The main ingredient of the proof of Lemma 2 is Pólya's theorem on characteristic functions:

Theorem 7 (Pólya, 1949, see [60], Th. 1)**.** *Let Φ be a real-valued function defined on $\mathbb{R}$ such that*

- *Φ is continuous and even,*
- *Φ is convex on $\mathbb{R}_+$,*
- *$\Phi(0) = 1$,*
- *$\lim_{x\to+\infty} \Phi(x) = 0$.*

Then Φ is the characteristic function corresponding to a continuous cumulative distribution function F whose derivative F', the probability density function, exists, is an even function, and is continuous everywhere except possibly at the point $x = 0$.

Proof of Lemma 2. Let $0 < p \leq 1$. The outline of the proof is as follows. First, we show that for any $0 < p \leq 1$, $\theta, \theta' \in \Theta$, the quantity $\mathrm{e}^{-u\left\|\theta'-\theta\right\|_p^p}$ is related to the characteristic function of some D-dimensional random variable $\mathbf{Z}_u$. Then we deduce that the kernel is positive definite.

Characteristic function. We rely on Theorem 7 to show that the scalar-valued function

$$\begin{aligned} \Phi_u : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto \mathrm{e}^{-u|x|^p} \end{aligned}$$

is a characteristic function. Indeed Φ_u is even, continuous, $\Phi_u(0) = 1$ and $\lim_{x\to+\infty} \Phi_u(x) = 0$. Moreover, for $p \in]0, 1]$ and $x > 0$, we have

$$\begin{aligned} \Phi_u'(x) &= -up\, x^{p-1}\mathrm{e}^{-ux^p} \\ \Phi_u''(x) &= up\, x^{p-2}((1-p) + upx^p)\mathrm{e}^{-ux^p}. \end{aligned}$$

As $0 < p \leq 1$, one sees that $\Phi_u''(x) > 0$ for all $x > 0$ hence Φ_u is convex on $\mathbb{R}_+$. By Pólya's theorem, Φ_u is the characteristic function of some scalar random variable Z_u, i.e., $\mathbb{E}_{Z_u}\left[\mathrm{e}^{-\mathrm{i}xZ_u}\right] = \Phi_u(x) = \mathrm{e}^{-u|x|^p}$ for all $x \in \mathbb{R}$. By noticing that for all $u > 0$ and $x \geq 0$ we have $\Phi_u(x) = \Phi_1(u^{1/p}x)$, we can rewrite

$$\Phi_u(x) = \mathbb{E}_{Z_1}\left[\mathrm{e}^{-\mathrm{i}u^{1/p}xZ_1}\right] \quad \forall u > 0, x \geq 0. \tag{B.1}$$

Denote $\mathbf{Z}_1 = \left(Z_1^1, \ldots, Z_1^D\right)^{\mathrm{T}}$ a multivariate random variable made of D independent copies of Z_1. By independence of the Z_1^d, we have for all $\theta, \theta' \in \mathbb{R}^D$:

$$\begin{aligned} \mathbb{E}_{\mathbf{Z}_1}\left[\mathrm{e}^{-\mathrm{i}u^{1/p}\left(\theta-\theta'\right)^{\mathrm{T}}\mathbf{Z}_1}\right] &= \prod_{d=1}^{D} \mathbb{E}_{Z_1^d}\left[\mathrm{e}^{-\mathrm{i}u^{1/p}\left(\theta[d]-\theta'[d]\right)Z_1^d}\right] \\ &= \prod_{d=1}^{D} \mathrm{e}^{-u\left|\theta[d]-\theta'[d]\right|^p} = \mathrm{e}^{-u\left\|\theta-\theta'\right\|_p^p}. \end{aligned} \tag{B.2}$$

CMF kernels are positive definite. Since the kernel κ is a CMF kernel (see Definition 4), Lemma 1 ensures the existence of a positive measure ν such that for all $\theta, \theta' \in \mathbb{R}^D$:

$$\kappa(\theta, \theta') = \int_0^{+\infty} \mathrm{e}^{-u\left|\theta-\theta'\right|^p} \,\mathrm{d}\nu(u). \tag{B.3}$$

Consider now an arbitrary support $\mathcal{S} = \{\theta_\ell\}_{\ell=1}^k \subset \Theta$ of k distinct parameters and an arbitrary nonzero vector of coefficients $\mathbf{c} = (c_1 \ldots c_k)^{\mathrm{T}} \in \mathbb{R}^k \setminus \{\mathbf{0}_k\}$. Let

$\mathbf{G} \in \mathbb{R}^{k\times k}$ be the matrix with entries $\mathbf{G}[i,j] = \kappa(\theta_i, \theta_j)$ for all $i, j \in [\![1, k]\!]$. We show that $\mathbf{G}$ is positive definite. We have

$$\begin{aligned}
\mathbf{c}^{\mathrm{T}}\mathbf{G}\mathbf{c} &= \sum_{i=1}^{k}\sum_{j=1}^{k} c_i c_j \kappa(\theta_i, \theta_j) \\
&= \sum_{i=1}^{k}\sum_{j=1}^{k} c_i c_j \int_0^{+\infty} \mathrm{e}^{-u\|\theta_i-\theta_j\|_p^p}\,\mathrm{d}\nu(u) \\
&\overset{Eq.\ (\mathrm{B.2})}{=} \sum_{i=1}^{k}\sum_{j=1}^{k} c_i c_j \int_0^{+\infty} \mathbb{E}_{\mathbf{Z}_1}\left[\mathrm{e}^{-\mathrm{i}u^{1/p}(\theta_i-\theta_j)^{\mathrm{T}}\mathbf{Z}_1}\right]\mathrm{d}\nu(u).
\end{aligned}$$

One obtains by linearity of the expectation

$$\begin{aligned}
\mathbf{c}^{\mathrm{T}}\mathbf{G}\mathbf{c} &= \int_0^{\infty} \mathbb{E}_{\mathbf{Z}_1}\left[\sum_{i=1}^{k}\sum_{j=1}^{k} c_i c_j \mathrm{e}^{-\mathrm{i}u^{1/p}(\theta_i-\theta_j)^{\mathrm{T}}\mathbf{Z}_1}\right]\mathrm{d}\nu(u) \\
&= \int_0^{\infty} \mathbb{E}_{\mathbf{Z}_1}\left[\left(\sum_{i=1}^{k} c_i \mathrm{e}^{-\mathrm{i}u^{1/p}\theta_i{}^{\mathrm{T}}\mathbf{Z}_1}\right)\left(\sum_{j=1}^{k} c_j \mathrm{e}^{+\mathrm{i}u^{1/p}\theta_j{}^{\mathrm{T}}\mathbf{Z}_1}\right)\right]\mathrm{d}\nu(u) \\
&= \int_0^{\infty} \mathbb{E}_{\mathbf{Z}_1}\left[\left|\sum_{i=1}^{k} c_i \mathrm{e}^{-\mathrm{i}u^{1/p}\theta_i{}^{\mathrm{T}}\mathbf{Z}_1}\right|^2\right]\mathrm{d}\nu(u) \qquad \geq 0. \qquad (\mathrm{B.4})
\end{aligned}$$

Since this holds for any $\mathbf{c}$ this shows that $\mathbf{G}$ is positive semi-definite. To conclude we show that it is positive definite. Define the function ψ on $\mathbb{R}^D$ by $\psi : \mathbf{z} \mapsto \left|\sum_{i=1}^{k} c_i \mathrm{e}^{-\mathrm{i}\theta_i{}^{\mathrm{T}}\mathbf{z}}\right|^2$ and let also Ψ be defined in $\mathbb{R}_+$ by $\Psi(u) = \mathbb{E}_{\mathbf{Z}_1}\left[\psi(u^{1/p}\mathbf{Z}_1)\right]$.

Assume now that $\mathbf{c}^{\mathrm{T}}\mathbf{G}\mathbf{c} = \int_{[0,+\infty[} \Psi(u)\,\mathrm{d}\nu(u) = 0$. Since Ψ is continuous, ν is non-null and different from a Dirac in 0, there exists at least one $u_0 > 0$ such that $\Psi(u_0) = 0$. Note now that

- $\Psi(u_0) = \mathbb{E}_{\mathbf{Z}_1}\left[\psi(u_0^{1/p}\mathbf{Z}_1)\right]$ where the density of $\mathbf{Z}_1$ is absolutely continuous (except possibly in 0) and does not reduce to a Dirac in 0 (*cf.* Theorem 7).
- ψ is nonnegative and continuous (by construction as a linear combination of exponentials).

We deduce that there exists a vector $z_0 \in \mathbb{R}^D$, a radius $r > 0$ such that $\psi(\mathbf{z}) = 0$ for all $\mathbf{z} \in \mathcal{B}(\mathbf{z}_0, r) \subset \mathbb{R}^D$, the open ball of $\mathbb{R}^D$ centered at $\mathbf{z}_0$ with radius r. The proof of the following lemma is postponed to the end of the section. We now require the following lemma to conclude the proof

Lemma 7. *Let $\{\mathbf{x}_\ell\}_{\ell=1}^{k}$ be a set of $k \geq 1$ pairwise distinct vectors of $\mathbb{R}^D$ with $D \geq 1$. Then there exists a vector $\mathbf{y} \in \mathbb{R}^D$ such that $\mathbf{x}_i{}^{\mathrm{T}}\mathbf{y} \neq \mathbf{x}_j{}^{\mathrm{T}}\mathbf{y}$ for all $i \neq j$.*

Since $\theta_1^\star \dots \theta_k^\star$ are pairwise distinct, by Lemma 7, there exists $\mathbf{y} \in \mathbb{R}^D$ such that ${\theta_i^\star}^{\mathrm{T}}\mathbf{y} \neq {\theta_j^\star}^{\mathrm{T}}\mathbf{y}$ for all $i \neq j$. For small enough $t \in \mathbb{R}$ we have $\mathbf{z}_0 + t\mathbf{y} \in \mathcal{B}(\mathbf{z}_0, r)$, hence $\sum_{i=1}^k c_i \mathrm{e}^{-\mathrm{i}{\theta_i^\star}^{\mathrm{T}}(\mathbf{z}_0+t\mathbf{y})} = 0$ for t in a neighborhood of zero. As this is an analytic function, it is indeed zero for all $t \in \mathbb{R}$. Since the scalars ${\theta_i^\star}^{\mathrm{T}}\mathbf{y}$, $i \in [\![1,k]\!]$, are pairwise distinct the corresponding functions $t \mapsto \mathrm{e}^{-\mathrm{i}t{\theta_i^\star}^{\mathrm{T}}\mathbf{y}}$ are linearly independent, hence $\mathbf{c} = \mathbf{0}_k$ and therefore $\mathbf{G}$ is positive definite as claimed. □

Proof of Lemma 7. Consider the following finite set of vectors

$$\mathcal{N} \triangleq \{\mathbf{x}_i - \mathbf{x}_j \mid i,j \in [\![1,k]\!] \text{ s.t. } i \neq j\} \tag{B.5}$$

As the vectors $\mathbf{x}$ are pairwise distinct, each $\mathbf{n} \in \mathcal{N}$ is nonzero. Denote $H_{\mathbf{n}}$ the linear hyperplane whose normal vector is $\mathbf{n}$, and consider $H \triangleq \cup_{\mathbf{n}\in\mathcal{N}} H_{\mathbf{n}}$. Since H is the union of a finite number of hyperplanes, $\mathbb{R}^D \setminus H$ is not empty.
Consider $\mathbf{y} \in \mathbb{R}^D \setminus H$. Then, by construction of $\mathbf{y}$, $\mathbf{n}^{\mathrm{T}}\mathbf{y} \neq 0$ for all $\mathbf{n} \in \mathcal{N}$ and therefore ${\mathbf{x}_i}^{\mathrm{T}}\mathbf{y} \neq {\mathbf{x}_j}^{\mathrm{T}}\mathbf{y}$ for all $i \neq j$, which concludes the proof. □

C. Details related to Example 4

For the configuration in Example 4, OMP prefers $\mathbf{0}_D$ to the ground-truth parameters at the first iteration if

$$\frac{k\varphi(\Delta^p)}{1 + (k-1)\varphi(2\Delta^p)} > 1, \tag{C.1}$$

or, equivalently, if

$$(k-1)\varphi(2\Delta^p) - k\varphi(\Delta^p) + 1 < 0. \tag{C.2}$$

Consider the function $f : x \mapsto (k-1)\varphi(2x) - k\varphi(x) + 1$ defined for all $x \geq 0$. Since $\varphi(0) = 1$, we have $f(0) = 0$. As φ is positive and decreasing, it admits a limit $\lim_{x\to\infty} \varphi(x) < 1$, hence $\lim_{x\to+\infty} f(x) > 0$. Moreover for each $x > 0$

$$\begin{aligned} f'(x) &= 2(k-1)\varphi'(2x) - k\varphi'(x) \\ &= k\varphi'(x)\Big[2\big(1 - \tfrac{1}{k}\big)\tfrac{\varphi'(2x)}{\varphi'(x)} - 1\Big]. \end{aligned} \tag{C.3}$$

When the ratio $\frac{\varphi'(2x)}{\varphi'(x)}$ tends to 1 as t tends to 0, which is the case, *e.g.*, when φ is differentiable at 0 (the derivative being then necessarily nonzero as φ is a CMF), there exists $x_0 > 0$ such that $x < x_0$ implies $2\big(1 - \frac{1}{k}\big)\frac{\varphi'(2x)}{\varphi'(x)} - 1 > 0$ provided that $k \geq 3$ (which is the case in Example 4 since $k = 2k' \geq 4$). Moreover, $\varphi'(x) < 0$ for all $x > 0$ so $f'(x) < 0$ for $x < x_0$, *i.e.*, f is decreasing on $[0, x_0]$. Combining this result with $f(0) = 0$, we deduce that (C.1) holds whenever $\Delta^p < x_0$, *i.e.*, the wrong parameter $\mathbf{0}_D$ will be preferred to any of the $\{\theta_\ell^\star\}_{\ell=1}^k$. Therefore, the quantity $x_0^{1/p}$ can be seen as a necessary separation condition.

D. Exact recovery in higher dimensions - CMF kernel and $k = 2$

We first give the following lemma about CMF kernels.

Lemma 8. *Let κ be a CMF kernel in dimension $D \geq 1$, in the sense of Definition 4. For any $u, v, w \in \Theta$, the following result holds:*

$$\kappa(u,v)\,\kappa(v,w) \leq \kappa(u,w) \tag{D.1}$$

Proof. By definition of CMF kernels, there exists a CMF φ such that $\kappa(\cdot,\cdot) = \varphi\Big(\|\cdot - \cdot\|_p^p\Big)$ and $\varphi(0) = 1$. By [49, Lemma 4.3], for all $x, y > 0$, we have

$$\varphi(x)\varphi(y) \leq \varphi(0)\varphi(x+y). \tag{D.2}$$

Using this result with $x = \|u - v\|_p^p$ and $y = \|v - w\|_p^p$, we have

$$\kappa(u,v)\kappa(v,w) \ \leq \varphi\Big(\|u - v\|_p^p + \|v - w\|_p^p\Big). \tag{D.3}$$

Since the quasi-norm $\|\cdot\|_p^p$ satisfies a triangular inequality, we have $\|u - w\|_p^p \leq \|u - v\|_p^p + \|v - w\|_p^p$. As any CMF is decreasing, (D.1) follows. ☐

Lemma 9 (Exact recovery for CMF dictionaries when $k = 2$)**.** *Let $\mathcal{A}$ be a CMF dictionary in dimension $D \geq 1$ and κ the associated CMF kernel. Consider a support $\mathcal{S}^\star = \{\theta_1^\star, \theta_2^\star\}$ where $\theta_1^\star \neq \theta_2^\star$, and $\mathbf{G} \in \mathbb{R}^{2\times 2}$ the matrix defined by $\mathbf{G}[i,j] = \kappa\big(\theta_i^\star, \theta_j^\star\big)$. Assume that*

$$\forall\, \theta \in \mathtt{Cart}(\mathcal{S}^\star) \setminus \mathcal{S}^\star\,, \quad \left\|\mathbf{G}^{-1}\mathbf{g}_\theta\right\|_1 < 1 \tag{D.4}$$

where $\mathbf{g}_\theta \in \mathbb{R}^2$ is defined by $\mathbf{g}_\theta[i] = \kappa(\theta, \theta_i^\star)$ for $i = 1, 2$. Then for any $\boldsymbol{y} \in \mathrm{span}(\boldsymbol{a}(\theta_1^\star), \boldsymbol{a}(\theta_2^\star))$, OMP with input $\boldsymbol{y}$ recovers $\mathcal{S}^\star$ in at most 2 iterations.

Proof. By construction, the kernel κ is admissible in the sense of Definition 1. We show below that since (D.4) holds, the support $\mathcal{S}^\star$ is admissible with respect to κ in the sense of Definition 2. Therefore, by Theorem 1, for any observation $\boldsymbol{y} \in \mathrm{span}(\boldsymbol{a}(\theta_1^\star), \boldsymbol{a}(\theta_2^\star))$, OMP with input $\boldsymbol{y}$ recovers $\mathcal{S}^\star$ in at most 2 iterations.

Consider a non-empty subset of indices $T \subset \{1, 2\}$ and $t := \mathrm{card}(T)$. Let also $\{c_\ell\}_{\ell\in T}$ be such that $c_\ell > 0$ and $\sum_{\ell\in T} c_\ell < 1$. Define

$$\begin{aligned} \psi\colon \mathbb{R} &\longrightarrow \mathbb{R}_+ \\ \theta &\longmapsto \textstyle\sum_{\ell\in T} c_\ell \kappa(\theta, \theta_\ell^\star) \end{aligned} \tag{D.5}$$

We now check Property *i)* and Property *ii)* of Definition 2.

Property i). We distinguish two cases:

- If $t = 1$, we can assume without loss of generality that $T = \{1\}$. Since $\kappa(\theta, \theta_1^\star) < 1$ for all $\theta \neq \theta_1^\star$, one immediately sees that $\psi(\theta) = c_1\kappa(\theta, \theta_1^\star) < c_1 = \psi(\theta_1^\star)$ for all $\theta \neq \theta_1^\star$, so $\theta_1^\star$ is the unique global maximizer of ψ_1
- If $t = 2$, let θ_0 be is a maximizer of ψ. Then $\theta_0[d]$ is a maximizer of

$$\begin{aligned} \psi_d \colon \mathbb{R} &\longrightarrow \mathbb{R}_+ \\ x &\longmapsto \textstyle\sum_{\ell \in T} c_\ell \varphi\Big(|x - \theta_\ell^\star[d]|^p + \sum_{j \neq d} |\theta_0[j] - \theta_\ell^\star[d]|^p\Big), \end{aligned} \tag{D.6}$$

  where $d \in [\![1, D]\!]$. Reasoning as in the proof of Theorem 2-*i)* (see Appendix A.2.1) yields that for all $x \notin \{\theta_\ell^\star[d]\}_{\ell=1}^k$, ψ_d is twice differentiable in x and $\psi_d''(x) > 0$ so x can't be a maximizer. Hence $\theta_0[d] \in \{\theta_\ell^\star[d]\}_{\ell=1}^k$ for all $d \in [\![1, k]\!]$ and therefore $\theta_0 \in \mathtt{Cart}(\mathcal{S}^\star)$.

  Moreover, since (D.4) holds, we have

$$\max_{\theta \in \mathtt{Cart}(\mathcal{S}^\star) \setminus \mathcal{S}^\star} \psi(\theta) = \max_{\theta \in \mathtt{Cart}(\mathcal{S}^\star) \setminus \mathcal{S}^\star} |\langle \boldsymbol{a}(\theta), \boldsymbol{y} \rangle| < \max_{\theta^\star \in \mathcal{S}^\star} |\langle \boldsymbol{a}(\theta_\ell^\star), \boldsymbol{y} \rangle| = \max_{\theta^\star \in \mathcal{S}^\star} \psi(\theta_\ell^\star) \tag{D.7}$$

  Hence all maximizers of ψ belong to $\mathcal{S}^\star$.

Property ii). As $T \neq \emptyset$ and $T \neq \{1, 2\}$, T is a singleton. We assume without loss of generality that $T = \{1\}$, hence $\psi(\theta) = c_1\kappa(\theta, \theta_1^\star)$ for each θ, where $0 < c_1 < 1$.

If $\psi_1(\theta_1^\star) - \kappa(\theta_1^\star, \theta_2^\star) \leq 0$, then $c_1 - \kappa(\theta_1^\star, \theta_2^\star) = \psi_1(\theta_1^\star) - \kappa(\theta_1^\star, \theta_2^\star) \leq 0$ hence $c_1 \leq \kappa(\theta_1^\star, \theta_2^\star)$ and $\psi(\theta) \leq \kappa(\theta_1^\star, \theta_2^\star)\kappa(\theta, \theta_1^\star)$ for each θ. By Lemma 8 with $u = \theta$, $v = \theta_1^\star$, $w = \theta_2^\star$, we obtain for each $\theta \in \Theta$

$$\psi(\theta) - \kappa(\theta, \theta_2^\star) \leq \kappa(\theta, \theta_1^\star)\kappa(\theta_1^\star, \theta_2^\star) - \kappa(\theta, \theta_2^\star) \leq 0. \qquad \square$$

As a last result, we show that in dimension $D > 1$, there exists Cartesian grids given by $\mathcal{G} = \mathtt{Cart}(\mathcal{S}^\star)$ with $\mathrm{card}(\mathcal{S}^\star) = 2$ that are not axis admissible.

Example 5. *In dimension $D = 2$, consider a Cartesian grid $\mathcal{G} \triangleq \{\theta_\ell^\star\}_{\ell=1}^k \subset \Theta = \mathbb{R}^2$ where $k = 4$ and $\theta_1^\star = \mathbf{0}_2$, $\theta_2^\star = \Delta\mathbf{e}_1$, $\theta_3^\star = \Delta\mathbf{e}_2$, $\theta_4^\star = \boldsymbol{\Delta}\mathbf{1}_2$ with $\mathbf{e}_\ell$ the ℓ-th canonical basis vector of $\mathbb{R}^2$ and $\Delta > 0$.*
Let $\boldsymbol{a} : \mathbb{R}^D \mapsto \mathcal{H}$ define a CMF dictionary in $\mathbb{R}^D$ with kernel $\kappa = \varphi\Big(\|\cdot - \cdot\|_p^p\Big)$. Let us show that, whenever $\Delta > 0$ and for some families of kernels, there always exists a linear combination of the atoms such that the maximizers of the function f_1 defined in (3.29) with $\theta_0 = \mathbf{0}_2$ does not belong to $\mathcal{S}_1 = \{0, \Delta\}$.
To that aim, take $c_1 = c_2 = 1$ and $c_3 = c_4 = -\frac{1+\varphi(\Delta^p)}{\varphi(\Delta^p)+\varphi(2\Delta^p)}$. We detail below that $f_1(0) = f_1(\Delta) = 0$ while

$$f_1\big(\tfrac{\Delta}{2}\big) = 2\varphi\big(\tfrac{1}{2^p}\Delta^p\big)\left|1 - \frac{1+\varphi(\Delta^p)}{\varphi(\Delta^p)+\varphi(2\Delta^p)} \frac{\varphi\big(\Delta^p + \frac{1}{2^p}\Delta^p\big)}{\varphi\big(\frac{1}{2^p}\Delta^p\big)}\right|. \tag{D.8}$$

If φ and Δ are such that $f_1\big(\frac{\Delta}{2}\big) \neq 0$, one can conclude that there exists at least one point $t_0 \in \mathbb{R}$ such that $f_1(t_0) > f_1(0) = f_1(\Delta) = 0$ so the maximizers of

f_1 do not belong to $\{0, \Delta\}$ and therefore $\mathcal{G}$ is not axis admissible. For instance, this is the case when φ is the CMF defined by $\varphi : x \mapsto \frac{1}{1+x}$ (cf Example 2). The construction easily extends to $D \geq 2$ by zero-padding of the $\theta^\star_\ell$.

Details. We have

$$\begin{aligned} f_1(0) &= \left|\kappa(\mathbf{0}_2, \mathbf{0}_2) + \kappa(\mathbf{0}_2, \Delta\mathbf{e}_1) - \frac{1+\varphi(\Delta^p)}{\varphi(\Delta^p)+\varphi(2\Delta^p)}[\kappa(\mathbf{0}_2, \Delta\mathbf{e}_2) + \kappa(\mathbf{0}_2, \Delta\mathbf{1}_2)]\right| \\ &= \left|1 + \varphi(\Delta^p) - \frac{1+\varphi(\Delta^p)}{\varphi(\Delta^p)+\varphi(2\Delta^p)}[\varphi(\Delta^p) + \varphi(2\Delta^p)]\right| \\ &= 0. \end{aligned} \tag{D.9}$$

Similarly

$$\begin{aligned} f_1(\Delta) &= \left|\kappa(\Delta\mathbf{e}_1, \mathbf{0}_2) + \kappa(\Delta\mathbf{e}_1, \Delta\mathbf{e}_1) - \frac{1+\varphi(\Delta^p)}{\varphi(\Delta^p)+\varphi(2\Delta^p)}[\kappa(\Delta\mathbf{e}_1, \Delta\mathbf{e}_2) + \kappa(\Delta\mathbf{e}_1, \Delta\mathbf{1}_2)]\right| \\ &= \left|\varphi(\Delta^p) + 1 - \frac{1+\varphi(\Delta^p)}{\varphi(\Delta^p)+\varphi(2\Delta^p)}[\varphi(2\Delta^p) + \varphi(\Delta^p)]\right| \\ &= 0. \end{aligned} \tag{D.10}$$

Finally

$$\begin{aligned} f_1(\tfrac{\Delta}{2}) &= \left|\kappa\big(\tfrac{\Delta}{2}\mathbf{e}_1, \mathbf{0}_2\big) + \kappa\big(\tfrac{\Delta}{2}\mathbf{e}_1, \Delta\mathbf{e}_1\big) - \frac{1+\varphi(\Delta^p)}{\varphi(\Delta^p)+\varphi(2\Delta^p)}\big[\kappa\big(\tfrac{\Delta}{2}\mathbf{e}_1, \Delta\mathbf{e}_2\big) + \kappa\big(\tfrac{\Delta}{2}\mathbf{e}_1, \Delta\mathbf{1}_2\big)\big]\right| \\ &= \left|2\varphi\big(\tfrac{\Delta^p}{2^p}\big) - \frac{1+\varphi(\Delta^p)}{\varphi(\Delta^p)+\varphi(2\Delta^p)}\big[2\varphi\big(\Delta^p + \tfrac{\Delta^p}{2^p}\big)\big]\right| \\ &= 2\varphi\big(\tfrac{1}{2^p}\Delta^p\big)\left|1 - \frac{1+\varphi(\Delta^p)}{\varphi(\Delta^p)+\varphi(2\Delta^p)}\frac{\varphi\big(\Delta^p + \frac{1}{2^p}\Delta^p\big)}{\varphi\big(\frac{1}{2^p}\Delta^p\big)}\right|. \end{aligned} \tag{D.11}$$

Now, when φ is the CMF defined by $\varphi : x \mapsto \frac{1}{1+x}$, f_1 becomes

$$f_1(\tfrac{\Delta}{2}) = \frac{2}{1+\frac{\Delta^p}{2^p}}\left|1 - \frac{2+\Delta^p}{1+\frac{1+\Delta^p}{1+2\Delta^p}}\frac{1+\frac{\Delta^p}{2^p}}{1+\Delta^p+\frac{\Delta^p}{2^p}}\right|. \tag{D.12}$$

As the factor inside the absolute value in the right hand side is a nonzero rational function of $x = \Delta^p$, we have $f_1(\Delta/2) \neq 0$ except on a set of Lebesgue measure zero. Hence there exists $\Delta > 0$ such that $f_1(\Delta/2) > 0$.

E. Table of notations

NOTATION	COMMENT
	General notations
$\mathcal{H}$, $\boldsymbol{y}$	(Hilbert) observation space and observation
$\mathcal{A}, \boldsymbol{a}(\cdot)$	Dictionary $\mathcal{A}$ made of parametric atoms $\boldsymbol{a}$
μ	Coherence between atoms of a support
$\mathbf{c} \in \mathbb{R}^k$	Weighting coefficients
Θ, θ	Parameter set and element
Δ_0	Minimum separation between elements of a support, see (3.36)
$\mathcal{S}$, $\mathcal{S}^\star$	Set of parameters
$\mathrm{supp}(\boldsymbol{y})$	Support of observation $\boldsymbol{y}$ (see (3.7))
$\mathcal{G}$	Cartesian grid
k, ℓ	Number of atoms, most frequent index
$\mathtt{Cart}$	Set augmenter, see (3.28)
φ	CMF (see Definition 3)
κ	Kernel function $\Theta \times \Theta \to \mathbb{R}_+$
$\mathcal{K}_{\mathtt{CMF}}(D)$	Set of CMF kernels in dimension D
$\mathcal{K}_{\mathtt{Lap}}(D)$	Set of Laplace kernels in dimension D
	Technical notations
ψ	Nonnegative linear combination of kernel (see Item *i)*)
ϕ	Function related to Item *ii)*
T	Subset of $[\![1, k]\!]$ in Item *i)* and Item *ii)*
g	Inner product between some observation $\boldsymbol{y}$ and atom $\boldsymbol{a}(\theta)$ seen as a function of θ
f	Absolute value of g
$\mathbf{G}$, $\mathbf{g}_\ell$	Gram matrix related to a support $\mathcal{S}$, columns of $\mathbf{G}$
$\mathbf{g}_\theta$	parametric vector related to a support $\mathcal{S}$
$\mathbf{u}, \mathbf{v}$	Vector of $\mathbb{R}^k$ for some k often defined as $\mathbf{u}, \mathbf{v} = \mathbf{G}^{-1}\mathbf{g}_\theta$ for some $\theta \in \Theta$
$\mathbb{E}$	Expected value
i	Imaginary number
Φ	Usual notation for characteristic function

Table E.1: Table of notations (and commands).

Acknowledgments

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

References

[1] S. Foucart, H. Rauhut, A Mathematical Introduction to Compressive Sensing, Birkhäuser Basel, 2013. doi:10.1007/978-0-8176-4948-7.

[2] C. Ekanadham, D. Tranchina, E. P. Simoncelli, Recovery of sparse translation-invariant signals with continuous basis pursuit, IEEE Transactions on Signal Processing 59 (10) (2011) 4735–4744. doi:10.1109/TSP.2011.2160058.

[3] E. J. Candès, C. Fernandez-Granda, Towards a Mathematical Theory of Super-resolution, Communications on Pure and Applied Mathematics 67 (6) (2014) 906–956. doi:10.1002/cpa.21455.

[4] V. Duval, G. Peyré, Exact Support Recovery for Sparse Spikes Deconvolution, Vol. 15, Springer US, 2015. doi:10.1007/s10208-014-9228-6.

[5] Y. C. Pati, R. Rezaiifar, 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, 1993, pp. 40–44 vol.1. doi:10.1109/ACSSC.1993.342465.

[6] B. Natarajan, Sparse approximate solutions to linear systems, SIAM Journal on Computing 24 (2) (1995) 227–234. doi:10.1137/S0097539792240406.

[7] A. J. Miller, Subset selection in regression, Chapman and Hall, London, 2002.

[8] S. Mallat, Z. Zhang, Matching pursuits with time-frequency dictionaries, IEEE Transactions on Signal Processing 41 (12) (1993) 3397–3415. doi:10.1109/78.258082.

[9] R. A. DeVore, V. N. Temlyakov, Some remarks on greedy algorithms, Advances in Computational Mathematics 5 (1) (1996) 173–187. doi:10.1007/bf02124742.

[10] E. Liu, V. N. Temlyakov, The Orthogonal Super Greedy Algorithm and Applications in Compressed Sensing, IEEE Transactions on Information Theory 58 (4) (2012) 2040–2047. doi:10.1109/TIT.2011.2177632.

[11] S. Chen, S. A. Billings, W. Luo, Orthogonal least squares methods and their application to non-linear system identification, International Journal of Control 50 (5) (1989) 1873–1896. doi:10.1080/00207178908953472.

[12] J. H. Friedman, W. Stuetzle, Projection pursuit regression, Journal of the American Statistical Association 76 (376) (1981) 817–823. doi:10.1080/01621459.1981.10477729.

[13] P. J. Huber, Projection pursuit, The Annals of Statistics 13 (2) (1985) 435–475. doi:10.1214/aos/1176349519.

[14] L. Rebollo-Neira, D. Lowe, Optimized orthogonal matching pursuit approach, IEEE Signal Processing Letters 9 (4) (2002) 137–140. doi:10.1109/LSP.2002.1001652.

[15] V. N. Temlyakov, Greedy approximation, Acta Numerica 17 (2008) 235–409. doi:10.1017/s0962492906380014.

[16] P. Vincent, Y. Bengio, Kernel Matching Pursuit, Machine Learning 48 (1/3) (2002) 165–187. doi:10.1023/a:1013955821559.

[17] J. F. Claerbout, F. Muir, Robust modeling with erratic data, Geophysics 38 (5) (1973) 826–844. doi:10.1190/1.1440378.

[18] S. Chen, D. Donoho, M. Saunders, Atomic decomposition by basis pursuit, SIAM Journal on Scientific Computing 20 (1) (1998) 33–61. doi:10.1137/S1064827596304010.

[19] R. Tibshirani, Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society. Series B (Methodological) 58 (1) (1996) 267–288.
URL http://www.jstor.org/stable/2346178

[20] R. Tibshirani, I. Johnstone, T. Hastie, B. Efron, Least angle regression, The Annals of Statistics 32 (2) (2004) 407–499. doi:10.1214/009053604000000067.

[21] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM Journal on Imaging Sciences 2 (1) (2009) 183–202. doi:10.1137/080716542.

[22] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, Foundations and Trends® in Machine Learning 3 (1) (2011) 1–122. doi:10.1561/2200000016.

[23] R. Gribonval, P. Vandergheynst, On the exponential convergence of matching pursuits in quasi-incoherent dictionaries, IEEE Transactions on Information Theory 52 (1) (2006) 255–261. doi:10.1109/tit.2005.860474.

[24] L. Borup, R. Gribonval, M. Nielsen, Beyond coherence: Recovering structured time–frequency representations, Applied and Computational Harmonic Analysis 24 (1) (2008) 120–128. doi:10.1016/j.acha.2007.09.002.

[25] K. C. Knudson, J. Yates, A. Huk, J. W. Pillow, Inferring sparse representations of continuous signals with continuous orthogonal matching pursuit, Advances in Neural Information Processing Systems 27 (2014) 1215–1223.

[26] C. Dorffer, A. Drémeau, C. Herzet, Efficient atom selection strategy for iterative sparse approximations, in: iTWIST 2018 - International Traveling Workshop on Interactions between low-complexity data models and Sensing Techniques, Marseille, France, 2018, pp. 1–3.
URL https://hal.inria.fr/hal-01937501

[27] Y. de Castro, F. Gamboa, Exact reconstruction using Beurling minimal extrapolation, Journal of Mathematical Analysis and Applications 395 (1) (2012) 336 – 354. doi:10.1016/j.jmaa.2012.05.011.

[28] K. Bredies, H. K. Pikkarainen, Inverse problems in spaces of measures, ESAIM: Control, Optimisation and Calculus of Variations 19 (1) (2012) 190–218. doi:10.1051/cocv/2011205.

[29] N. Boyd, G. Schiebinger, B. Recht, The alternating descent conditional gradient method for sparse inverse problems, in: 2015 IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2015, pp. 57–60. doi:10.1109/CAMSAP.2015.7383735.

[30] Q. Denoyelle, V. Duval, G. Peyré, Support Recovery for Sparse Super-Resolution of Positive Measures, Journal of Fourier Analysis and Applications 23 (5) (2017) 1153–1194.

[31] J. A. Tropp, Greed is good: Algorithmic results for sparse approximation, IEEE Transactions on Information Theory 50 (10) (2004) 2231–2242. doi:10.1109/TIT.2004.834793.

[32] C. Soussen, R. Gribonval, J. Idier, C. Herzet, Joint k-Step Analysis of Orthogonal Matching Pursuit and Orthogonal Least Squares, IEEE Transactions on Information Theory 59 (5) (2013) 3158–3174. doi:10.1109/tit.2013.2238606.

[33] J. . Fuchs, On sparse representations in arbitrary redundant bases, IEEE Transactions on Information Theory 50 (6) (2004) 1341–1344. doi:10.1109/TIT.2004.828141.

[34] J. A. Tropp, Just relax: convex programming methods for identifying sparse signals in noise, IEEE Transactions on Information Theory 52 (3) (2006) 1030–1051. doi:10.1109/TIT.2005.864420.

[35] R. Gribonval, M. Nielsen, Sparse representations in unions of bases, IEEE Transactions on Information Theory 49 (12) (2003) 3320–3325. doi:10.1109/tit.2003.820031.

[36] S. Huang, J. Zhu, Recovery of sparse signals using OMP and its variants: convergence analysis based on RIP, Inverse Problems 27 (3) (2011) 035003. doi:10.1088/0266-5611/27/3/035003.

[37] R. Maleh, Improved RIP Analysis of Orthogonal Matching Pursuit, Tech. rep. (2011). arXiv:arXiv:1102.4311.

[38] Q. Mo, Y. Shen, A Remark on the Restricted Isometry Property in Orthogonal Matching Pursuit, IEEE Transactions on Information Theory 58 (6) (2012) 3654–3656. doi:10.1109/TIT.2012.2185923.

[39] J. Wang, B. Shim, On the Recovery Limit of Sparse Signals Using Orthogonal Matching Pursuit, IEEE Transactions on Signal Processing 60 (9) (2012) 4973–4976. doi:10.1109/TSP.2012.2203124.

[40] L. Chang, J. Wu, An Improved RIP-Based Performance Guarantee for Sparse Signal Recovery via Orthogonal Matching Pursuit, IEEE Transactions on Information Theory 60 (9) (2014) 5702–5715. doi:10.1109/TIT.2014.2338314.

[41] J. Wen, X. Zhu, D. Li, Improved bounds on restricted isometry constant for orthogonal matching pursuit, Electronics Letters 49 (23) (2013) 1487–1489. doi:10.1049/el.2013.2222.

[42] Q. Mo, A Sharp Restricted Isometry Constant Bound of Orthogonal Matching Pursuit, Tech. rep. (2015). arXiv:arXiv:1501.01708.

[43] C. Fernandez-Granda, Super-resolution of point sources via convex programming, Information and Inference: A Journal of the IMA 5 (3) (2016) 251–303. doi:10.1093/imaiai/iaw005.

[44] V. I. Morgenshtern, E. J. Candès, Super-resolution of positive sources: The discrete setup, SIAM Journal on Imaging Sciences 9 (1) (2016) 412–444. doi:10.1137/15m1016552.
URL https://doi.org/10.1137/15m1016552

[45] T. Bendory, Robust Recovery of Positive Stream of Pulses, IEEE Transactions on Signal Processing 65 (8) (2017) 2114–2122. doi:10.1109/TSP.2017.2656841.

[46] C. Poon, G. Peyré, Multi-dimensional sparse super-resolution (2017). arXiv:arXiv:1709.03157.

[47] T. Zhang, Sparse Recovery With Orthogonal Matching Pursuit Under RIP, IEEE Transactions on Information Theory 57 (9) (2011) 6215–6221. doi:10.1109/tit.2011.2162263.

[48] S. Foucart, Stability and Robustness of Weak Orthogonal Matching Pursuits, in: Springer Proceedings in Mathematics & Statistics, Springer New York, 2012, pp. 395–405. doi:10.1007/978-1-4614-4565-4_30.

[49] M. Merkle, Completely Monotone Functions: A Digest (2014) 347–364doi:10.1007/978-1-4939-0258-3_12.

[50] K. S. Miller, S. G. Samko, Completely monotonic functions, Integral Transforms and Special Functions 12 (4) (2001) 389–402.

[51] S. Bernstein, Sur les fonctions absolument monotones, Acta Mathematica 52 (1929) 1–66. doi:10.1007/BF02592679.

[52] D. V. Widder, Laplace Transform, hardcover Edition, Princeton Mathematical Press, 1941.

[53] H. Wendland, Scattered data approximation, 2005. doi:10.2277/0521843359.

[54] N. Aronszajn, Theory of reproducing kernels, Transactions of the American Mathematical Society 68 (3) (1950) 337–337. doi:10.1090/s0002-9947-1950-0051437-7.

[55] B. Mailhé, B. Sturm, M. D. Plumbley, Behavior of greedy sparse representation algorithms on nested supports, in: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 2013, pp. 5710–5714. doi:10.1109/ICASSP.2013.6638758.

[56] M. Belkin, K. Sinha, Learning Gaussian Mixtures with Arbitrary Separation, arXiv preprintarXiv:0907.1054.

[57] J. Wang, B. Shim, Exact Recovery of Sparse Signals Using Orthogonal Matching Pursuit: How Many Iterations Do We Need?, IEEE Transactions on Signal Processing 64 (16) (2016) 4194–4202. doi:10.1109/tsp.2016.2568162.

[58] D. Bernstein, Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory, Princeton University Press, 2005.

[59] H. Fejzić, C. Freiling, D. Rinne, Descartes' rule of signs, alternations of data sets, and balanced differences, The American Mathematical Monthly 116 (4) (2009) 316–327.
URL http://www.jstor.org/stable/40391091

[60] G. Pólya, Remarks on Characteristic Functions, in: Proceedings of the [First] Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, Calif., 1949, pp. 115–123.
URL https://projecteuclid.org/euclid.bsmsp/1166219202