Efficient atom selection strategy for iterative sparse approximations

Clément Dorffer1,‡, Angélique Drémeau1 and Cédric Herzet2

1 Lab-STIC UMR 6285, CNRS, ENSTA Bretagne, Brest, F-29200, France.
2 INRIA Centre Rennes-Bretagne Atlantique and Lab-STIC UMR 6285, CNRS, IMT-Atlantique, Rennes, F-35000, France.

Abstract—We propose a low-computational strategy for the efficient implementation of the “atom selection step” in sparse representation algorithms. The proposed procedure is based on simple tests enabling to identify subsets of atoms which cannot be selected. Our procedure applies on both discrete or continuous dictionaries. Experiments performed on DOA and Gaussian deconvolution problems show the computational gain induced by the proposed approach.

1 Problem statement

Sparsely approximating a signal vector \( y \in \mathbb{R}^m \) in a dictionary \( A \) consists of finding \( k \ll m \) coefficients \( x_i \) and atoms \( a_i \in A \) such that \( y \approx \sum_{i=1}^{k} a_i x_i \). The dictionary \( A \) can be either discrete, composed of a finite number of elements, or “continuous”, i.e., having an infinite uncountable number of atoms.

Sparse approximations have proven to be relevant in many application domains and a great number of procedures to find “good” sparse approximations have been proposed in the literature: convex relaxation [1][10], greedy algorithms [11][15], Bayesian approaches [16][12], etc. Many popular instances of these procedures rely on the same “atom selection” step, i.e., the new atom added to the support at each iteration, says \( a_{\text{select}} \), verifies

\[
\alpha_{\text{select}} \in \arg \max_{a \in A} \langle r, a \rangle,
\]

where \( \langle \cdot, \cdot \rangle \) denotes the inner product and \( r \) is the current “residual”, i.e., the original signal from which the contributions from the previously selected atoms have been removed. The Frank-Wolfe algorithm [1], the matching pursuit [11] or the orthogonal matching pursuit [12] procedures are popular instances of algorithms using (1). It is worth noting that (1) constitutes the core of most sparse-approximation algorithms in the context of continuous dictionaries [6][15] where some standard “matrix-vector” operations available in the discrete setting are no longer possible.

A brute-force evaluation of (1) may become resource consuming when the number of atoms in \( A \) is large since it requires the exploration of the whole dictionary to find the atom the most correlated with \( r \). In this work, we propose a strategy to alleviate the complexity of (1). Our procedure is inspired from work [20]: it consists of performing simple tests allowing to identify group of atoms not attaining the maximum value of (1). Interestingly, the proposed approach provides a rigorous framework to recently-proposed procedures based on some approximations of continuous dictionaries [6][15]. In the rest of this abstract, we do not elaborate on the connections with [6][15] (details will be provided during the conference) but rather focus on the description of the proposed methodology.

2 Proposed strategy

Our proposed selection strategy is based on the following observations. If \( \tilde{A} \subseteq A \),

\[
\max_{a \in \tilde{A}} \langle r, a \rangle \leq \max_{a \in A} \langle r, a \rangle.
\]

Hence, letting \( \tau \triangleq \max_{a \in \tilde{A}} \langle r, a \rangle \), we have \( \forall a \in A \):

\[
\langle r, a \rangle < \tau \Rightarrow a \notin \arg \max_{\tilde{A}} \langle r, \tilde{a} \rangle.
\]

In other words, if \( a \in \tilde{A} \) is an atom which satisfies the inequality in the left-hand side of (2), then this atom is surely not the one to be selected by (1). Elaborating on this observation, we further have:

\[
\max_{a \in \tilde{R}} \langle r, a \rangle < \tau \Rightarrow \forall a \in \tilde{A} \cap \tilde{R} : a \notin \arg \max_{\tilde{A}} \langle r, \tilde{a} \rangle,
\]

where \( \tilde{R} \) is some arbitrary subset of \( \mathbb{R}^m \). In the sequel we will refer to \( \tilde{R} \) as “region”. The operational meaning of (3) is as follows: if the inequality in the left-hand side is satisfied, one is ensured that no atom in \( \tilde{A} \cap \tilde{R} \) will attain the maximum of \( \langle r, a \rangle \). The entire set \( \tilde{A} \cap \tilde{R} \) can thus be ignored, enabling us to reduce the number of candidate atoms to be tested in the selection (1).

Implication (3) constitutes the basis of our complexity reduction method. More specifically, we consider (3) with some particular choices of region \( \tilde{R} \). These choices are motivated by the following requirements: i) \( \tilde{R} \) should lead to any easy evaluation of \( \max_{a \in \tilde{R}} \langle r, a \rangle \); ii) \( \tilde{R} \) should approximate as tightly as possible some part of \( \tilde{A} \) since larger regions typically lead to inequalities more difficult to satisfy.

In our contribution, we show that the first requirement is satisfied for some particular geometries of region \( \tilde{R} \). In particular, we consider “sphere”, “dome” and “slice” geometries. Sphere and dome regions can be formally expressed as

\[
B_{\epsilon, l} \triangleq \{a : \|a - t\|_2 \leq \epsilon \} \quad \text{(sphere)}
\]
\[
D_{\epsilon, l} \triangleq \{a : \langle t, a \rangle \geq \epsilon, \|a\|_2 = 1 \} \quad \text{(dome)}
\]

whereas the mathematical characterization of the slice regions is more involved and not detailed in this abstract. We show that for these choices of regions, \( \max_{a \in \tilde{R}} \langle r, a \rangle \) admits a simple
analytical expression. In particular, evaluating $\langle r, a \rangle$ for sphere and dome regions basically requires the computation of one single inner product $(t, r)$.

We address the second requirement by proposing a methodology to automatically adapt the size of $\mathcal{R}$ (via a tuning of $\epsilon$) in order to satisfy the inequality in the left-hand side of (3). We do not detail this procedure here but mention that the evaluation of the “optimal” value of $\epsilon$ has a negligible complexity.

We thus propose the following strategy (summarized in Algorithm 1) to speed up the computation of (1). We select a set of atoms which do not attain the maximum of $\langle r, a \rangle$. We evaluate (1) by working on a reduced dictionary:

$$a_{\text{select}} \in \arg \max_{a \in \mathcal{A} \setminus \overline{\mathcal{A}}} \langle r, a \rangle,$$

where $\overline{\mathcal{A}}$ denotes the set of atoms which have been removed by the tests (3).

For discrete dictionaries, the computational complexity of the proposed method (for sphere and dome regions) evolves as

$$O(\text{card}(\mathcal{A})) m + \frac{L m}{\text{evaluation of } \tau} + \text{card}(\mathcal{A} \setminus \overline{\mathcal{A}}) m).$$

This has to be compared to the complexity required by a brute evaluation of (1), i.e., $O(\text{card}(\mathcal{A})) m$.

In the next section, we propose to compare the efficiency and the computational gain allowed by the proposed methodology.

### 3 Experiments

We propose to challenge the proposed selection procedure and the classical exhaustive search on two different problems: a direction-of-arrival (DOA) estimation problem using a discrete dictionary and a Gaussian deconvolution problem with a continuous dictionary. Within the DOA estimation framework, we examine the number of scalar products required to achieve the selection step (1), with and without the proposed method. This provides a quantitative assessment of the computational gain induced by our method. The Gaussian deconvolution problem acts as a proof of concept, illustrating the interest of our method in continuous dictionaries.

In the DOA estimation problem, we consider a dictionary composed of $n = 1000$ normalized steering vectors of size $m = 100$, each corresponding to a different angle of arrival in $[-\pi/2, \pi/2]$. In the Gaussian deconvolution problem, $A = \{a(\mu): \mu \in [0, 100]\}$ where $a(\mu)$ is a Gaussian function with mean $\mu$ and variance $\sigma^2 = 10$. In both simulations, the observed signal $y$ is constructed as a random linear combination of 5 atoms of $\mathcal{A}$. Coefficients associated to the atoms are realizations of a uniform distribution on $[-1, 1]$.

In Fig. 1(a), we considered the DOA estimation problem. We applied $5$ iterations of orthogonal matching pursuit on $y$ (which requires to solve (1) at each iteration). Each column of the figure represents the (cumulated) number of inner products which have been computed with an exhaustive search of $a_{\text{select}}$ in the whole dictionary (blue), and with the proposed method described in Algorithm 1 (red). We set $L = 100$ and define the centers of the (dome) regions $\{r_j\}_{j=1}^{L}$ by a regular subsampling of the dictionary. We let $\mathcal{A} = \{r_j\}_{j=1}^{L}$. We see that the proposed method allows for a gain in complexity of one order of magnitude with respect to a brute-force approach.

In Fig. 1(b) we consider the Gaussian deconvolution problem and show the set of removed atoms, $\mathcal{A} \cap \mathcal{R}_i$, for each sphere region $\{\mathcal{R}_i\}_{i=1}^{L=100}$. We choose the centers of the regions $\{t_l\}_{l=1}^{L=100}$ by a regular subsampling of $\mu \in [0, 100]$. The size of each region (value of $\epsilon$) is automatically tuned to verify (if possible) the inequality in the left-hand side (3) as discussed in Section 2. We set $A = \{t_l\}_{l=1}^{L=100}$. The figure shows the interval of value of $\mu$ which can be removed from the dictionary for each test region. We see in Fig. 1(b) that by the end of the proposed elimination procedure, the search space is reduced to a small interval, around $\mu \approx 90$. 

---

**Algorithm 1 Efficient selection strategy**

**Input:** residual $r$, subset of atoms $\mathcal{A}$, set of regions $\{\mathcal{R}_i\}_{i=1}^{L}$

1. **Init:** $\mathcal{A}_{\text{removed}} = \emptyset$
2. Evaluate $\tau = \max_{a \in \mathcal{A}} \langle r, a \rangle$
3. **for all** $1 \leq l \leq L$ **do**
   - **if** $\max_{a \in \mathcal{R}_l} \langle r, a \rangle < \tau$ **then**
     - $\mathcal{A}_{\text{removed}} = \mathcal{A}_{\text{removed}} \cup (\mathcal{A} \cap \mathcal{R}_l)$
   - **end if**
4. **end for**
5. Find $a_{\text{select}} \in \arg \max_{a \in \mathcal{A} \setminus \mathcal{A}_{\text{removed}}} \langle r, a \rangle$

---

### Figure 1

(a) "discrete" DOA estimation problem - number of scalar products evaluated for the atom selection. (b) "continuous" Gaussian deconvolution problem - interval of atoms which can be ignored in problem $1$. 

---

3 Which may be empty in some cases.
References


