Efficient atom selection strategy for iterative sparse approximations

Clément Dorffer, Angélique Drémeau, Cédric Herzet

Efficient atom selection strategy for iterative sparse approximations

Clément Dorffer[1][*], Angélique Drémeau[1] and Cédric Herzet[2]

[1]Lab-STICC UMR 6285, CNRS, ENSTA Bretagne, Brest, F-29200, France.

[2]INRIA Centre Rennes-Bretagne Atlantique and Lab-STICC 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 $\mathcal{A}$ consists of finding $k \ll m$ coefficients x_i and atoms $a_i \in \mathcal{A}$ such that $y \approx \sum_{i=1}^{k} a_i x_i$. The dictionary $\mathcal{A}$ can be either discrete, *i.e.,* 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 [16–19], 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_{select}, verifies

$$a_{\text{select}} \in \arg\max_{a \in \mathcal{A}} \langle r, a \rangle , \qquad (1)$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product[1] 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[2] of (1) may become resource consuming when the number of atoms in $\mathcal{A}$ 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 $\bar{\mathcal{A}} \subseteq \mathcal{A}$,

$$\max_{a \in \bar{\mathcal{A}}} \langle r, a \rangle \leq \max_{a \in \mathcal{A}} \langle r, a \rangle .$$

Hence, letting $\tau \triangleq \max_{a \in \bar{\mathcal{A}}} \langle r, a \rangle$, we have $\forall a \in \mathcal{A}$:

$$\langle r, a \rangle < \tau \Rightarrow a \notin \arg\max_{\tilde{a} \in \mathcal{A}} \langle r, \tilde{a} \rangle . \qquad (2)$$

In other words, if $a \in \mathcal{A}$ 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:

$$\max_{a \in \mathcal{R}} \langle r, a \rangle < \tau \Rightarrow \forall a \in \mathcal{A} \cap \mathcal{R} : a \notin \arg\max_{\tilde{a} \in \mathcal{A}} \langle r, \tilde{a} \rangle , \quad (3)$$

where $\mathcal{R}$ is some arbitrary subset of $\mathbb{R}^m$. In the sequel we will refer to $\mathcal{R}$ 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 $\mathcal{A} \cap \mathcal{R}$ will attain the maximum of $\langle r, a \rangle$. The entire set $\mathcal{A} \cap \mathcal{R}$ 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 $\mathcal{R}$. These choices are motivated by the following requirements: *i)* $\mathcal{R}$ should lead to any easy evaluation of $\max_{a \in \mathcal{R}} \langle r, a \rangle$; *ii)* $\mathcal{R}$ should approximate as tightly as possible some part of $\mathcal{A}$ 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 $\mathcal{R}$. In particular, we consider "sphere", "dome" and "slice" geometries. Sphere and dome regions can be formally expressed as

$$\begin{aligned} \mathcal{B}_{t,\epsilon} &\triangleq \{ a \ : \ \|a - t\|_2 \leq \epsilon \} & \text{(sphere)} \\ \mathcal{D}_{t,\epsilon} &\triangleq \{ a \ : \ \langle t, a \rangle \geq \epsilon, \|a\|_2 = 1 \} & \text{(dome)}, \end{aligned}$$

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, $\max_{a \in \mathcal{R}} \langle r, a \rangle$ admits a simple

---

[1]In case of possibly negative coefficients in x, it is common to use the absolute value of the inner product. However, one can also deal with negative coefficients using eq. (1) with a doubled size dictionary containing atoms a_i and their negatives $-a_i$.

[2]In the discrete setting, the standard approach consists of evaluating $\langle r, a \rangle$ for all $a \in \mathcal{A}$, leading to a complexity scaling as $\mathcal{O}(\text{card}(\mathcal{A})m)$. In the continuous setting, a classic approach consists of running gradient ascent algorithms initialized on a fine discretization of the dictionary.

Algorithm 1 Efficient selection strategy

Input: residual r, subset of atoms $\bar{\mathcal{A}}$, set of regions $\{\mathcal{R}_l\}_{l=1}^L$
Init: : $\mathcal{A}_{\text{removed}} = \emptyset$
Evaluate $\tau = \max\limits_{a \in \bar{\mathcal{A}}} \langle r, a \rangle$
for all $1 \leq l \leq L$ **do**
    if $\max\limits_{a \in \mathcal{R}_l} \langle r, a \rangle < \tau$ **then**
        $\mathcal{A}_{\text{removed}} = \mathcal{A}_{\text{removed}} \cup (\mathcal{A} \cap \mathcal{R}_l)$
    end if
end for
Find $a_{\text{select}} \in \arg \max\limits_{a \in \mathcal{A} \setminus \mathcal{A}_{\text{removed}}} \langle r, a \rangle$

---

analytical expression. In particular, evaluating $\max\limits_{a \in \mathcal{R}} \langle r, a \rangle$ for sphere and dome regions basically requires the computation of *one single* inner product $\langle t, r \rangle$.

We address the second requirement by proposing a methodology to automatically adapt the size of $\mathcal{R}$ (via a tuning of ϵ) 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 ϵ 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 $\bar{\mathcal{A}} \subset \mathcal{A}$ and L regions $\{\mathcal{R}_l\}_{l=1}^L$, and apply test (3) for each region. Each test allows to identify a set[3] of atoms which do not attain the maximum of $\langle r, a \rangle$. We evaluate (1) by working on a reduced dictionary:

$$a_{\text{select}} \in \arg \max_{a \in \mathcal{A} \setminus \mathcal{A}_{\text{removed}}} \langle r, a \rangle, \qquad (4)$$

where $\mathcal{A}_{\text{removed}}$ 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

$$\mathcal{O}(\underbrace{\text{card}(\bar{\mathcal{A}})m}_{\text{evaluation of } \tau} + \underbrace{Lm}_{\text{evaluation of (3)}} + \underbrace{\text{card}(\mathcal{A} \setminus \mathcal{A}_{\text{removed}})m}_{\text{evaluation of (4)}}).$$

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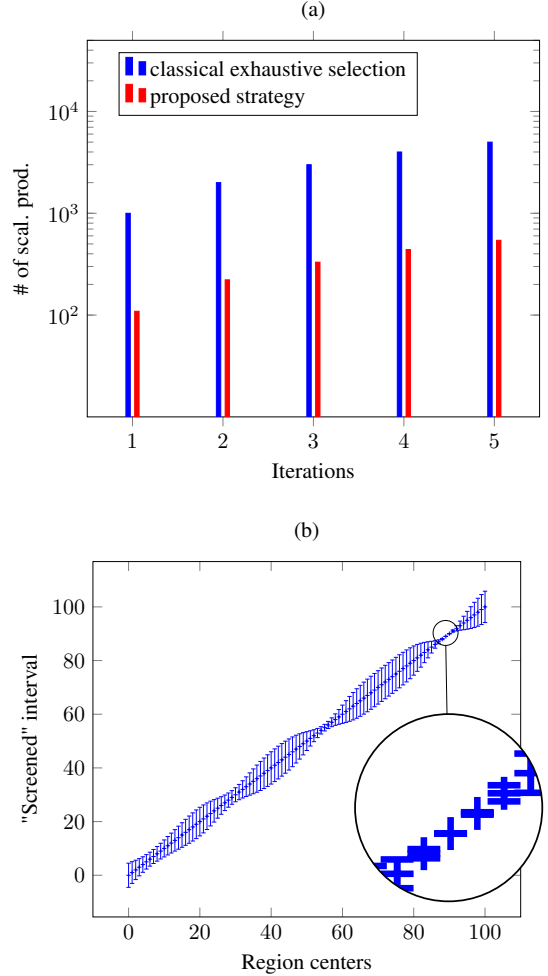
---

[3]Which may be empty in some cases.

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).

$\mathcal{A} = \{a(\mu) : \mu \in [0, 100]\}$ where $a(\mu)$ is a Gaussian function with mean μ 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_{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 $\{t_l\}_{l=1}^L$ by a regular subsampling of the dictionary. We let $\bar{\mathcal{A}} = \{t_l\}_{l=1}^L$. We see that the proposed method allows for a gain of 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, $\mathcal{A} \cap \mathcal{R}_l$, for each (sphere) region $\{\mathcal{R}_l\}_{l=1}^{L=100}$. We choose the centers of the regions $\{t_l\}_{l=1}^L$ by a regular subsampling of $\mu \in [0, 100]$. The size of each region (value of ϵ) is automatically tuned to verify (if possible) the inequality in the left-hand side (3) as discussed in Section 2. We set $\bar{\mathcal{A}} = \{t_l\}_{l=1}^{L=100}$. The figure shows the interval of value of μ 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 $\mu \approx 90$.

References

[1] Marguerite Frank and Philip Wolfe. An algorithm for quadratic programming. *Naval Research Logistics (NRL)*, 3(1-2):95–110, 1956.

[2] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[3] Ingrid Daubechies, Michel Defrise, and Christine De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on pure and applied mathematics*, 57(11):1413–1457, 2004.

[4] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.

[5] Nikhil Rao, Parikshit Shah, and Stephen Wright. Forward–backward greedy algorithms for atomic norm regularization. *IEEE Transactions on Signal Processing*, 63(21):5798–5811, 2015.

[6] Chaitanya Ekanadham, Daniel Tranchina, and Eero P Simoncelli. Recovery of sparse translation-invariant signals with continuous basis pursuit. *IEEE transactions on signal processing*, 59(10):4735–4744, 2011.

[7] Gongguo Tang, Badri Narayan Bhaskar, Parikshit Shah, and Benjamin Recht. Compressed sensing off the grid. *IEEE transactions on information theory*, 59(11):7465–7490, 2013.

[8] Angeliki Xenaki and Peter Gerstoft. Grid-free compressive beamforming. *The Journal of the Acoustical Society of America*, 137(4):1923–1935, 2015.

[9] Paul Catala, Vincent Duval, and Gabriel Peyré. A low-rank approach to off-the-grid sparse deconvolution. In *Journal of Physics: Conference Series*, volume 904, page 012015. IOP Publishing, 2017.

[10] Nicholas Boyd, Geoffrey Schiebinger, and Benjamin Recht. The alternating descent conditional gradient method for sparse inverse problems. *SIAM Journal on Optimization*, 27(2):616–639, 2017.

[11] Stéphane G Mallat and Zhifeng Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on signal processing*, 41(12):3397–3415, 1993.

[12] Yagyensh Chandra Pati, Ramin Rezaiifar, and Perinkulam Sambamurthy Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*, pages 40–44. IEEE, 1993.

[13] Joel A Tropp, Anna C Gilbert, and Martin J Strauss. Algorithms for simultaneous sparse approximation. part i: Greedy pursuit. *Signal Processing*, 86(3):572–588, 2006.

[14] Deanna Needell and Joel A Tropp. Cosamp: Iterative signal recovery from incomplete and inaccurate samples. *Applied and computational harmonic analysis*, 26(3):301–321, 2009.

[15] Karin C Knudson, Jacob Yates, Alexander Huk, and Jonathan W Pillow. Inferring sparse representations of continuous signals with continuous orthogonal matching pursuit. In *Advances in neural information processing systems*, pages 1215–1223, 2014.

[16] Cédric Févotte and Simon J Godsill. Blind separation of sparse sources using jeffrey's inverse prior and the em algorithm. In *International Conference on Independent Component Analysis and Signal Separation*, pages 593–600. Springer, 2006.

[17] Philip Schniter, Lee C Potter, and Justin Ziniel. Fast bayesian matching pursuit. In *Information Theory and Applications Workshop, 2008*, pages 326–333. IEEE, 2008.

[18] Charles Soussen, Jérôme Idier, David Brie, and Junbo Duan. From bernoulli–gaussian deconvolution to sparse signal restoration. *IEEE Transactions on Signal Processing*, 59(10):4572–4584, 2011.

[19] Angélique Drémeau, Cedric Herzet, and Laurent Daudet. Soft bayesian pursuit algorithm for sparse representations. In *Statistical Signal Processing Workshop (SSP), 2011 IEEE*, pages 341–344. IEEE, 2011.

[20] Cédric Herzet and Angélique Drémeau. Joint screening tests for lasso. In *ICASSP 2018-IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1–5, 2018.