REGION-BASED RELAXATIONS TO ACCELERATE GREEDY APPROACHES

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

[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 the standard "Gaussian deconvolution" problem show the computational gain induced by the proposed approach.

Index Terms— Sparse approximation, atom selection, low-complexity methods.

1. PROBLEM STATEMENT

We consider the problem of sparsely approximating a signal vector $y \in \mathcal{H}$ in some dictionary $\mathcal{A}$, where $\mathcal{H}$ is a Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\|\cdot\|$. In the last decades, many works have addressed this problem in a "*discrete*" setting, where the dictionary $\mathcal{A}$ contains a *finite* number of elements. In the recent years, several contributions have focused on the more involved problem of sparse representations in a "*continuous*" dictionary [1–9]. In this setup, the dictionary $\mathcal{A}$ is typically defined as

$$\mathcal{A} = \{ \boldsymbol{a}(\theta) : \theta \in \Theta \} \tag{1}$$

where $\boldsymbol{a} : \Theta \to \mathcal{H}$ is some continuous function and Θ is some compact subset of $\mathbb{R}^d$.

Sparse representations in continuous dictionaries may seem intractable from both a theoretical and practical point of view since $\mathcal{A}$ contains an *infinite* uncountable number of elements. Nevertheless, recent theoretical works showed that, under some mild conditions, sparse linear combinations of elements of $\mathcal{A}$ may be recovered as the solution of convex optimization problems over the space of Radon measures [1–4]. From a practical point of view, several contributions have also proposed methodologies to access to the solution of these convex optimization problems, see [5–9].

In this paper, we assume $\mathcal{H}$ is defined over the real field $\mathbb{R}$ and the elements of the sparse representation to be positive.[1]

---

[1]This can be done without loss of generality as the standard problem can be exposed as a particular case of the positive one.

We focus on the efficient implementation of a particular step of the procedures (see *e.g.*, [6–8]) based on the "Frank-Wolfe" (FW) algorithm [10], namely the so-called "atom selection" step:

$$\text{Find} \ \ \boldsymbol{a}_{\max} \in \arg\max_{\boldsymbol{a} \in \mathcal{A}} \langle \boldsymbol{a}, \boldsymbol{r} \rangle, \tag{2}$$

where $\boldsymbol{r} \in \mathcal{H}$ is some "iteration-dependent" element.[2] We note that (2) is also central to the implementation of many standard greedy procedures, *e.g.*, MP [12] or OMP [13].

Although greedy and FW-based procedures are commonly assumed to be low-complexity methodologies in the sparse-representation literature, we see from (2) that identifying $\boldsymbol{a}_{\max}$ nevertheless requires to sweep over all the elements of $\mathcal{A}$. If the dictionary contains a huge number of elements (and in particular if $\mathcal{A}$ is continuous), finding the true maximizer can thus lead to an overwhelming complexity. Quite surprisingly, little attention has been brought to the practical implementation of (2) in the current literature. Most of the contributions simply suppose that "an efficient scheme to evaluate (2) exists". A noticeable exception is [14], where the authors propose to approximate the solution of (2) as the output of a low-dimensional interpolation scheme.

In this paper, we propose an efficient methodology to perform (2) with a prescribed accuracy. Our procedure is inspired from our work [15]: it consists in performing simple tests allowing to identify groups of atoms *not* attaining the maximum value of (2). Although motivated in the context of continuous dictionaries, our methodology applies to both the continuous and discrete settings. It may provide important computational gain as soon as the number of atoms in $\mathcal{A}$ is large. Moreover, in the context of continuous dictionaries, and considering "inexact" versions of the FW algorithm (see [16]), the proposed framework can also be exploited to derive upper bounds on the maximum value of (2) that can, in turn, be used to check the "δ-accuracy" of the selected atoms.

2. THE PROPOSED STRATEGY

Hereafter, we first expose our atom rejection strategy in Section 2.1. Then, in Section 2.2, we propose a strategy to ensure

---

[2]We note that if $\mathcal{A}$ is a "continuous" dictionary, the continuity of $\boldsymbol{a} : \Theta \to \mathcal{H}$ and the compactness of Θ ensure the existence of a maximizer by virtue of the Weierstrass theorem, see *e.g.*, [11, Proposition A.8].

that the selected atom achieves the maximum of (2) with a prescribed accuracy.

2.1. Atom rejection procedure

Our proposed selection strategy is based on the following observations. If $\mathcal{S}$ is a (closed) subset of $\mathcal{A}$, then

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

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

$$\langle \boldsymbol{a}, \boldsymbol{r} \rangle < \tau \Rightarrow \boldsymbol{a} \notin \arg\max_{\tilde{\boldsymbol{a}}\in\mathcal{A}} \langle \tilde{\boldsymbol{a}}, \boldsymbol{r} \rangle . \quad (3)$$

In other words, if $\boldsymbol{a} \in \mathcal{A}$ is an atom which satisfies the inequality in the left-hand side of (3), then this atom is surely not the one to be selected by (2). Elaborating on this observation, we further have:

$$\max_{\boldsymbol{a}\in\mathcal{R}} \langle \boldsymbol{a}, \boldsymbol{r} \rangle < \tau \Rightarrow \forall \boldsymbol{a} \in \mathcal{A} \cap \mathcal{R} : \boldsymbol{a} \notin \arg\max_{\tilde{\boldsymbol{a}}\in\mathcal{A}} \langle \tilde{\boldsymbol{a}}, \boldsymbol{r} \rangle , \quad (4)$$

where $\mathcal{R}$ is some arbitrary subset of $\mathcal{H}$. In the sequel we will refer to $\mathcal{R}$ as "region". The operational meaning of (4) 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 $\langle \boldsymbol{a}, \boldsymbol{r} \rangle$. 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 (2).

Test (4) constitutes the basis of our complexity reduction method, described in Algorithm 1. Our goal is to identify subsets of the dictionary, say $\mathcal{A}_{\mathrm{not}}$, that do *not* contain $\boldsymbol{a}_{\max}$. This goal is achieved by applying the discriminative test (4) on a set of regions $\{\mathcal{R}_l\}_{l=1}^{L}$. Once a set $\mathcal{A}_{\mathrm{not}}$ has been identified, we are left with the simpler problem

$$\text{Find } \boldsymbol{a}_{\max} \in \arg\max_{\boldsymbol{a}\in\mathcal{A}\backslash\mathcal{A}_{\mathrm{not}}} \langle \boldsymbol{a}, \boldsymbol{r} \rangle . \quad (5)$$

The computation savings induced by (5) stands in the reduction of the number of atoms to be tested to find $\boldsymbol{a}_{\max}$. In the discrete setting, the computational reduction is a direct function of the cardinality of $\mathcal{A}$ and $\mathcal{A}\backslash\mathcal{A}_{\mathrm{not}}$. In the continuous setting, both $\mathcal{A}$ and $\mathcal{A}\backslash\mathcal{A}_{\mathrm{not}}$ usually contain an infinite number of atoms. Nevertheless the search of $\boldsymbol{a}_{\max}$ over the reduced domain $\mathcal{A}\backslash\mathcal{A}_{\mathrm{not}}$ may be much faster than over the entire space $\mathcal{A}$.

The choice of the regions $\{\mathcal{R}_l\}_{l=1}^{L}$ is a degree of freedom of the procedure, different choices leading to different performance. The complexity associated to the evaluation of the left-hand side of (4) is a direct function of the geometry of $\mathcal{R}$. The following result, proved in our companion paper [15, Theorem 1], shows that for "dome" and "sphere" regions, test (4) can be carried out at the cost of *one* inner product in $\mathcal{H}$:

Algorithm 1 Efficient atom selection strategy

Inputs: residual $\boldsymbol{r}$, subset of atoms $\mathcal{S}$, set of regions $\{\mathcal{R}_l\}_{l=1}^{L}$
Init: : $\mathcal{A}_{\mathrm{not}} = \emptyset$
Evaluate $\tau = \max_{\boldsymbol{a}\in\mathcal{S}} \langle \boldsymbol{a}, \boldsymbol{r} \rangle$
for all $1 \leq l \leq L$ **do**
    if $\max_{\boldsymbol{a}\in\mathcal{R}_l} \langle \boldsymbol{a}, \boldsymbol{r} \rangle < \tau$ **then**
        $\mathcal{A}_{\mathrm{not}} = \mathcal{A}_{\mathrm{not}} \cup (\mathcal{A} \cap \mathcal{R}_l)$
    end if
end for
Find $\boldsymbol{a}_{\max} \in \arg\max_{\boldsymbol{a}\in\mathcal{A}\backslash\mathcal{A}_{\mathrm{not}}} \langle \boldsymbol{a}, \boldsymbol{r} \rangle$

Theorem 1. *Let*

$$\mathcal{B}_{\boldsymbol{t},\epsilon} \triangleq \{\boldsymbol{a} \: : \: \|\boldsymbol{a} - \boldsymbol{t}\| \leq \epsilon\} \qquad \textit{(sphere)},$$
$$\mathcal{D}_{\boldsymbol{t},\epsilon} \triangleq \{\boldsymbol{a} \: : \: \langle \boldsymbol{a}, \boldsymbol{t} \rangle \geq \epsilon, \|\boldsymbol{a}\| = 1\} \quad \textit{(dome)}.$$

If $\mathcal{R} = \mathcal{B}_{\boldsymbol{t},\epsilon}$, then $\max_{\boldsymbol{a}\in\mathcal{R}} \langle \boldsymbol{a}, \boldsymbol{r} \rangle < \tau$ if and only if

$$\langle \boldsymbol{t}, \boldsymbol{r} \rangle < \tau - \epsilon \|\boldsymbol{r}\|.$$

Moreover, if $\mathcal{R} = \mathcal{D}_{\boldsymbol{t},\epsilon}$, then $\max_{\boldsymbol{a}\in\mathcal{R}} \langle \boldsymbol{a}, \boldsymbol{r} \rangle < \tau$ if and only if

$$\langle \boldsymbol{t}, \boldsymbol{r} \rangle < \tau,$$
$$\epsilon > \frac{\langle \boldsymbol{t}, \boldsymbol{r} \rangle \tau + \sqrt{\|\boldsymbol{t}\|^2 - \langle \boldsymbol{t}, \boldsymbol{r} \rangle^2}\sqrt{\|\boldsymbol{t}\|^2 - \tau^2}}{\|\boldsymbol{t}\|^2}.$$

$\mathcal{B}_{\boldsymbol{t},\epsilon}$ and $\mathcal{D}_{\boldsymbol{t},\epsilon}$ define a sphere and a dome in $\mathcal{H}$. Other geometries are possible, leading to different results. The analysis of different geometries of regions will be investigated in our future work.

2.2. Guarantees of δ-optimality

Although problem (5) may be structurally easier than its standard counterpart (2), finding an exact maximizer of $\langle \boldsymbol{a}, \boldsymbol{r} \rangle$ may still be complicated (in particular in the continuous setting). Hopefully, "inexact" versions of the FW algorithm exist [16]: in these variants, the atom selected at each iteration, say $\boldsymbol{a}_{\max}^{\delta}$, must verify

$$\delta \geq \max_{\boldsymbol{a}\in\mathcal{A}} \langle \boldsymbol{a}, \boldsymbol{r} \rangle - \langle \boldsymbol{a}_{\max}^{\delta}, \boldsymbol{r} \rangle \quad (6)$$

where δ is some prescribed accuracy level.

In practice, (6) may seem difficult to check since evaluating the right-hand side of this expression entails knowing $\max_{\boldsymbol{a}\in\mathcal{A}} \langle \boldsymbol{a}, \boldsymbol{r} \rangle$. Nevertheless, if one can identify some upper bound on $\max_{\boldsymbol{a}\in\mathcal{A}} \langle \boldsymbol{a}, \boldsymbol{r} \rangle$, say β, a sufficient condition for (6) to hold simply writes

$$\delta \geq \beta - \langle \boldsymbol{a}_{\max}^{\delta}, \boldsymbol{r} \rangle . \quad (7)$$

Algorithm 2 Region-based upper bound estimation

Inputs: residual r, atom subset $\mathcal{A}\backslash\mathcal{A}_{\text{not}}$

Build: a set of (dome or sphere) regions $\{\mathcal{R}_l\}_{l=1}^L$, such that

$$\mathcal{A}\backslash\mathcal{A}_{\text{not}} \subseteq \cup_{l=1}^L \mathcal{R}_l \qquad (12)$$

for all $1 \leq l \leq L$ **do**
    Compute $\beta_l = \max_{a \in \mathcal{R}_l} \langle a, r \rangle$ using (9) or (10).
end for
Set the upper bound: $\beta = \max_{1 \leq l \leq L} \beta_l$.
Check the relaxed condition (7).

In the sequel, we leverage (7) to derive accuracy guarantees on our atom selection method. More specifically, we emphasize that the "region-based" framework considered in this paper is prone to deriving simple upper bounds on $\max_{a \in \mathcal{A}} \langle a, r \rangle$.

We first note that $\max_{a \in \mathcal{A}} \langle a, r \rangle = \max_{a \in \mathcal{A}\backslash\mathcal{A}_{\text{not}}} \langle a, r \rangle$, so we can concentrate on the evaluation of an upper bound over $\mathcal{A}\backslash\mathcal{A}_{\text{not}}$. Next, if one can identify a set of (closed and bounded) regions $\{\mathcal{R}_l\}_{l=1}^L$ (possibly different from those considered in Section 2.1) such that $\mathcal{A}\backslash\mathcal{A}_{\text{not}} \subseteq \cup_{l=1}^L \mathcal{R}_l$, we have that

$$\max_{a \in \mathcal{A}\backslash\mathcal{A}_{\text{not}}} \langle a, r \rangle \leq \max_{1 \leq l \leq L} \max_{a \in \mathcal{R}_l} \langle a, r \rangle. \qquad (8)$$

Finally, $\max_{a \in \mathcal{R}_l} \langle a, r \rangle$ has a simple analytical expression for some particular geometries of $\mathcal{R}_l$:

Theorem 2. $\forall r \in \mathcal{H}$ we have

$$\max_{a \in \mathcal{B}_{t,\epsilon}} \langle a, r \rangle = \langle t, r \rangle + \epsilon \|r\|, \qquad (9)$$

$$\max_{a \in \mathcal{D}_{t,\epsilon}} \langle a, r \rangle = \begin{cases} \|r\| & \text{if } \epsilon \leq \frac{\langle t, r \rangle}{\|r\|} \\ f(\epsilon) & \text{otherwise} \end{cases}, \qquad (10)$$

where

$$f(\epsilon) \triangleq \epsilon \langle t, r \rangle + \sqrt{1-\epsilon^2}\sqrt{\|r\|^2 - \langle t, r \rangle^2}. \qquad (11)$$

A proof of this result can be found in our companion paper [15, Theorem 2]. We see that by combining Theorem 2 with (8), we obtain an easy computable upper bound on $\max_{a \in \mathcal{A}\backslash\mathcal{A}_{\text{not}}} \langle a, r \rangle$. The δ-accuracy of some candidate atom $a_{\max}^{\delta}$ can then be verified via (7). The proposed strategy is summarized in Algorithm 2.

3. EXPERIMENTS

In this section, we illustrate the performance achieved by the proposed methodologies in the case of a Gaussian deconvolution problem. In this setup, the "atom" function defining the dictionary in (1) writes (up to some normalization factor) as

$$a : \Theta \to \mathbb{R}^m \qquad (13)$$
$$\theta \mapsto \left[\exp\left(-\frac{(i\Delta - \theta)^2}{2\sigma^2} \right) \right]_{1 \leq i \leq m}$$
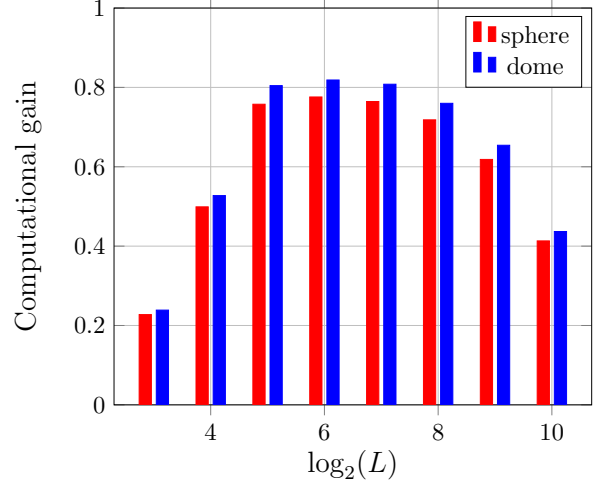

Fig. 1. Computational gain achieved by the atom rejection method described in Algorithm 1 as a function of the number L of regions for random realizations of r.

where $\sigma^2 = 0.01$ and $\Delta = 0.02$ are some parameters.

In Fig. 1, we first assess the performance of the atom-rejection procedure introduced in Section 2.1 (see Algorithm 1) on random realizations of r. In order to evaluate the computational gain induced by the proposed methodology, we consider a discrete setting where $\mathcal{A}$ contains a finite number of atoms. In particular, we set

$$\Theta = \left\{ \theta_{\min} + \frac{j}{2048}\theta_{\text{range}} : j = 1 \ldots 2048 \right\}$$

where $\theta_{\min} = 0$ and $\theta_{\text{range}} = 2$. The figure of merit represented in Fig. 1 then writes

$$\text{computational gain} = 1 - \frac{L + \text{card}\{\mathcal{A}\backslash\mathcal{A}_{\text{not}}\}}{\text{card}\{\mathcal{A}\}}. \qquad (14)$$

The numerator (resp. denominator) of the ratio in (14) represents the total number of inner products needed to identify $a_{\max}$ with Algorithm 1 (resp. a brute-force approach). The "computational gain" thus represents the proportion of inner products which can be saved by using the proposed methodology (as compared to a brute-force approach).

We consider both the cases where the regions $\{\mathcal{R}_l\}_{l=1}^L$ are defined as either spheres $\mathcal{B}_{t,\epsilon}$ or domes $\mathcal{D}_{t,\epsilon}$ (see Theorem 1). The centers of the regions correspond to a regular subsampling of the dictionary $\mathcal{A}$. The radius of the spheres are set to $\epsilon = 0.3$ and the radius of the domes to $\epsilon = 0.9550$. These radii have been chosen so that sphere and dome regions cover exactly the same subset of atoms of the dictionary. The vector r is built as a Gaussian random linear combination of 5 atoms of $\mathcal{A}$. The results are averaged over 50 realizations of r.

We see in Fig. 1 that the computational gain induced by our method raises up to 0.8 for both sphere and dome regions

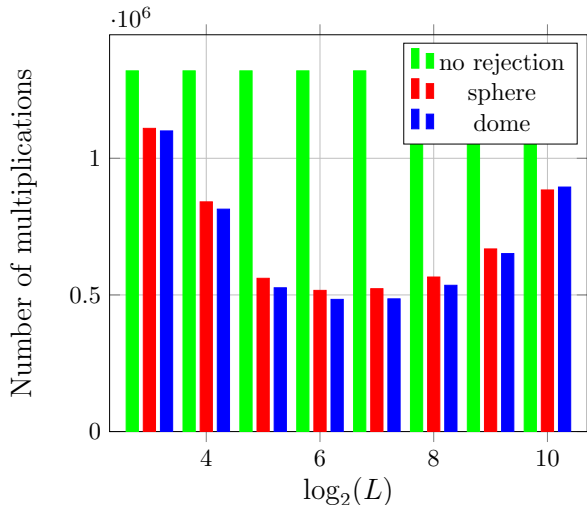**Fig. 2**. Total number of multiplications carried out during the FW algorithm as a function of the number L of regions.

Fig. 3. Percentage of saved multiplications to complete the optimization process by using the proposed atom-rejection procedure.

when $L = 128$. In other words, the proposed methodology allows for the savings of 80% of the inner products needed by a brute-force evaluation of (2). The gain is usually slightly larger for dome regions (blue) than for sphere regions (red). This is due to the fact that dome regions preserve the unit norm of the atoms and thus allow for a finer "covering" of the dictionary.

In our second experiment, we assess the overall computational savings allowed by our atom-rejection procedure when used in the selection step (2) of a FW algorithm [10]. More specifically, we address the resolution of a LASSO problem with the Gaussian dictionary defined in (13). The observation vector used as input of the LASSO problem, say y, is built as the Gaussian random linear combination of 5 atoms of the dictionary. The epigraph formulation of the LASSO problem, as suggested in [17, Chapter 4], is considered and standard FW algorithm is applied.[3] The optimization procedure is stopped when a primal-dual gap of 10^{-2} is attained. The complexity of the FW algorithm (with and without the proposed atom-rejection procedure) is measured by the number of multiplications[4] needed to complete the optimization process. The obtained results are illustrated in Fig. 2 as a function of the number of regions L used in the atom-rejection procedure. We see that for L varying between 2^5 and 2^8, the proposed methodology allows for significant complexity savings. Fig. 3 offers a different perspective on this result: it represents the percentage of multiplications saved by using the proposed methodology. We see that, for values of L ranging between 2^5 and 2^8, savings of roughly 60% can be achieved.

In our last experiment, we assess the quality of the δ-

optimality relaxed condition proposed in Section 2.2. More specifically, in Fig. 4 we illustrate the tightness of the upper bounds proposed in Theorem 2. We consider a continuous setup where $\Theta = \{\theta \in [0, 2]\}$. The curves represent the following figure of merit as a function of the number L of regions:

$$\text{bound tightness} = \frac{\beta}{\max_{a \in \mathcal{A}} \langle a, r \rangle}, \qquad (15)$$

where β is the bound defined in Algorithm 2.

We consider the following simulation setup. The vector r is built as the Gaussian random linear combination of 5 atoms of $\mathcal{A}$. The results are averaged over 50 realizations of r. We suppose that the atoms in $\mathcal{A}\backslash\mathcal{A}_{\text{not}}$ correspond to parameters in the interval $I_{\theta_{\max}} = [\theta_{\max} - \rho, \theta_{\max} + \rho]$, where $\theta_{\max}$ is the parameter associated to $a_{\max}$ (that is $a_{\max} = a(\theta_{\max})$). Two different values of ρ are considered in Fig. 4. The centers of the regions used in Algorithm 2 correspond to a regular subsampling of $I_{\theta_{\max}}$. The radius of all the regions is the same and is chosen $\forall L$ so that $\mathcal{A}\backslash\mathcal{A}_{\text{not}} \subseteq \cup_{l=1}^{L} \mathcal{R}_l$, that is the set of regions $\{\mathcal{R}_l\}_{l=1}^{L}$ covers the residual dictionary.

We see in Fig. 4 that, quite logically, the tightness of the proposed bound improves with the number of regions and the decrease of ρ. We also note that the dome regions lead again to superior performance than sphere regions.

4. CONCLUSIONS

In this paper, we propose an atom-rejection procedure that allows to reduce the computation complexity of the atom-selection step inherent to greedy algorithms. Our approach relies on the partition of the atoms into regions having suitable geometries. We show that the maximization of linear

---

[3]The penalization parameter "λ" used in the LASSO problem is set to $\lambda = 0.5 \max_{\theta \in \Theta} |\langle a(\theta), y \rangle|$.

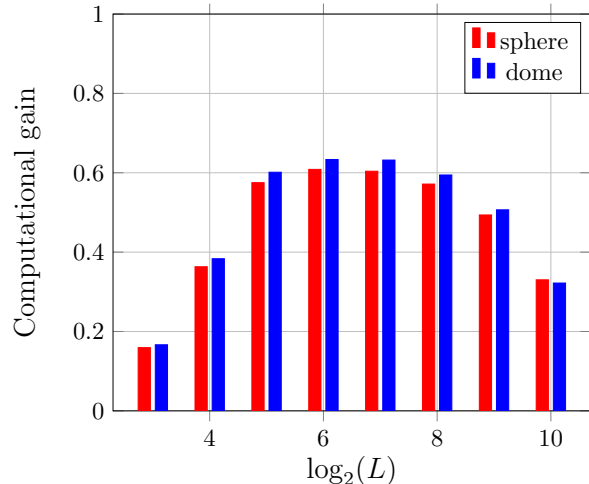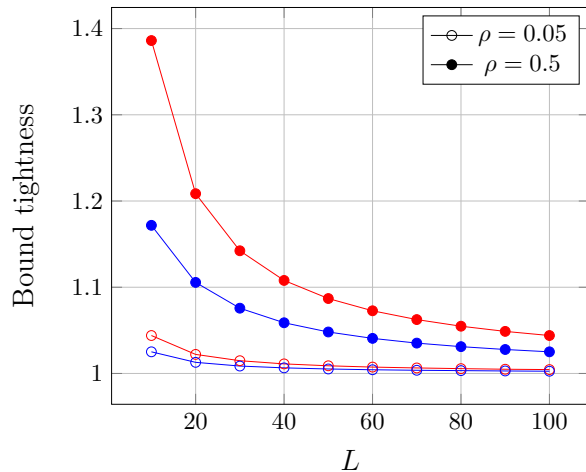[4]Multiplications dominate the complexity of floating-point operations.

Fig. 4. Tightness of the upper bounds defined in Theorem 2.

functionals over these regions has simple analytical solutions which can, in turn, be exploited to reduce the cost of greedy procedures. Our experiments, carried out in the context of the standard Gaussian deconvolution problem, show that large computational gains can be achieved when dealing with large dictionaries.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] Yohann de Castro and Fabrice Gamboa, "Exact reconstruction using Beurling minimal extrapolation," *Journal of Mathematical Analysis and Applications*, vol. 395, no. 1, pp. 336 – 354, 2012.

[2] Emmanuel J. Candès and Carlos Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Comm. Pure Appl. Math*, vol. 67, no. 6, pp. 906–956, June 2014.

[3] Vincent Duval and Gabriel Peyré, "Exact Support Recovery for Sparse Spikes Deconvolution," *Foundations of Computational Mathematics*, vol. 15, no. 5, pp. 1315–1355, 2015.

[4] Quentin Denoyelle, Vincent Duval, and Gabriel Peyré, "Support Recovery for Sparse Super-Resolution of Positive Measures," *Journal of Fourier Analysis and Applications*, vol. 23, no. 5, pp. 1153–1194, 2017.

[5] Gongguo Tang, Badri Narayan Bhaskar, Parikshit Shah, and Benjamin Recht, "Compressed sensing off the grid," *IEEE transactions on information theory*, vol. 59, no. 11, pp. 7465–7490, 2013.

[6] Kristian Bredies and Hanna Katriina Pikkarainen, "Inverse problems in spaces of measures," *ESAIM: Control, Optimisation and Calculus of Variations*, vol. 19, no. 1, pp. 190–218, 2013.

[7] Nikhil Rao, Parikshit Shah, and Stephen Wright, "Forward–backward greedy algorithms for atomic norm regularization," *IEEE Transactions on Signal Processing*, vol. 63, no. 21, pp. 5798–5811, 2015.

[8] Nicholas Boyd, Geoffrey Schiebinger, and Benjamin Recht, "The alternating descent conditional gradient method for sparse inverse problems," *SIAM Journal on Optimization*, vol. 27, no. 2, pp. 616–639, 2017.

[9] Sebastian Semper, Florian Roemer, Thomas Hotz, and Giovanni Del Galdo, "Grid-free direction-of-arrival estimation with compressed sensing and arbitrary antenna arrays," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 3251–3255.

[10] Marguerite Frank and Philip Wolfe, "An algorithm for quadratic programming," *Naval Research Logistics (NRL)*, vol. 3, no. 1-2, pp. 95–110, 1956.

[11] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, USA, 2003.

[12] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[13] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Ann. Asilomar Conf. Signals, Systems, and Computers*, 1993.

[14] 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*, 2014, pp. 1215–1223.

[15] Cédric Herzet, Clément Dorffer, and Angélique Drémeau, "Gather and Conquer: Region-based Strategies to Accelerate Safe Screening Tests," working paper or preprint, <hal-01913331>, 2018.

[16] Martin Jaggi, "Revisiting Frank-Wolfe: Projection-free sparse convex optimization," in *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*. 2013, ICML'13, pp. I–427–I–435, JMLR.org.

[17] Quentin Denoyelle, *Theoretical and Numerical Analysis of Super-Resolution Without Grid*, Ph.D. thesis, Université Paris-Dauphine, 2018.