

# ATOM SELECTION IN CONTINUOUS DICTIONARIES: RECONCILING POLAR AND SVD APPROXIMATIONS

Frédéric Champagnat<sup>\*</sup> and Cédric Herzet<sup>†</sup>

<sup>\*</sup> ONERA, The French Aerospace Lab, BP 80100 91123, Palaiseau Cedex, France

<sup>†</sup> INRIA Rennes-Bretagne Atlantique, Campus de Beaulieu, 35000 Rennes, France

## ABSTRACT

This paper deals with efficient atom selection procedure in a continuous dictionary, as required for instance in a Frank-Wolfe approach within a BLASSO problem for the one-dimensional deconvolution problem. We show that efficient maximization of a correlation between any given vector and an atom sweeping a continuous dictionary can be performed through a particular piece-wise linear approximation of dictionaries: the polar approximation. We finally identify the polar approximation as being optimal in a mean square error sense for dictionaries with raised-cosine Toeplitz kernels.

**Index Terms**— Sparse representation, Continuous dictionary, BLASSO, SVD

## 1. PROBLEM STATEMENT

Sparse representations in redundant dictionaries is a problem of great interest which has attracted the attention of many researchers during the last decades. Recently, this paradigm has been reformulated in a “continuous” setup, where the representation dictionary contains an *infinite* uncountable number of elements. In this context, the collection of atoms forming the dictionary  $\mathcal{A}$  typically takes the form:

$$\mathcal{A} = \{\mathbf{a}(\theta) : \theta \in \Theta\} \quad (1)$$

where  $\mathbf{a} : \Theta \rightarrow \mathcal{H}$  is some continuous bounded function,  $\Theta$  an interval of  $\mathbb{R}^d$  and  $\mathcal{H}$  some Hilbert space (with inner product  $\langle \cdot, \cdot \rangle$  and induced norm  $\|\cdot\|$ ).

In the last years, several theoretical works showed that sparse linear combinations of continuous atoms (say  $\mathbf{y} = \sum_{i=1}^K \alpha_i \mathbf{a}(\theta_i)$  with  $\alpha_i \in \mathbb{R}$  and  $\theta_i \in \Theta$ ) can be recovered under mild conditions as the solution a convex optimization problem over the space of Radon measure:<sup>1</sup>

$$\min_{\mu \in \mathcal{M}(\Theta)} \frac{1}{2} \|\mathbf{y} - A\mu\|^2 + \lambda \|\mu\|_{TV} \quad (2)$$

where  $\mathcal{M}(\Theta)$  is the space of Radon measures over  $\Theta$ ,  $A : \mathcal{M}(\Theta) \rightarrow \mathcal{H}$  is a “dictionary-dependent” bounded

linear operator and  $\|\mu\|_{TV}$  is the so-called “Total Variation” norm, see *e.g.*, [1–4]. Problem (2) is commonly referred to as “BLASSO” and can be seen as the continuous counterpart of the well-known LASSO problem [5].

Despite these nice theoretical results, finding the solution of BLASSO turns out to be a challenging issue since it entails the resolution of an optimization problem over the infinite-dimensional set of Radon measures  $\mathcal{M}(\Theta)$ . Hence, several works recently addressed the problem of designing tractable algorithms solving (2), see [2, 6–11]. This article is part of this trend. We propose an efficient procedure to solve BLASSO for “one-dimensional (1D) deconvolution” problems ( $d = 1$ ). Our procedure is based on a fast (approximated) implementation of the well-known Frank-Wolfe (FW) algorithm [12].

## 2. PRACTICAL IMPLEMENTATIONS OF BLASSO

We first review some algorithmic solutions proposed in the literature to address the BLASSO problem (2). We leverage these considerations to build a new fast solving methodology for BLASSO in the next section.

### 2.1. FW-based Implementations

Among the different works addressing the resolution of (2), quite a fair number of them were based on variations of the well-known FW iterative algorithm, see [7, 9, 10]. A common feature of all these procedures stands in the following “atom selection” step:<sup>2</sup>

$$\mathbf{a}_{\max} \in \arg \max_{\mathbf{a} \in \mathcal{A}} \langle \mathbf{a}, \mathbf{r} \rangle, \quad (3)$$

where  $\mathbf{r} \in \mathcal{H}$  is some iteration-dependent vector. From a practical point of view, FW-based procedures thus trade the infinite-dimensional search over the space of Radon measure against the maximization of some linear function over the dictionary  $\mathcal{A}$ .

Although most contributions of the literature assume that “some efficient scheme to find  $\mathbf{a}_{\max}$  exists”, the search over

<sup>1</sup>The authors thank the ANR (ANR-15-CE23-0021) for its support.

<sup>2</sup>More specifically, the solution of (2) writes as  $\mu = \sum_{i=1}^K \beta_i \delta_{\theta_i}$  for some  $\beta_i \in \mathbb{R}$ , where  $\delta_{\theta_i}$  is the Dirac function with mass located in  $\theta_i$ .

<sup>3</sup>With a slight abuse of notation, we use the notation “ $\mathbf{a}$ ” to denote both an element of  $\mathcal{A}$  and the continuous bounded function  $\mathbf{a} : \Theta \mapsto \mathcal{H}$ .

the entire dictionary (3) may reveal to be a tedious operation since  $\mathcal{A}$  is commonly a highly non-convex set. In this paper, we propose a solution to this issue in the case of 1D-deconvolution problems. Our approach is based on some finite-dimensional approximations of the dictionary.

## 2.2. Dictionary approximations

A practical way of avoiding the infinite-dimensional nature of problem (2) consists in approaching  $\mathcal{A}$  by a set of finite-dimensional approximations.

This strategy can be formalized as follows. Let  $\Theta = \cup_{\ell=1}^L \Theta_\ell$  be a partition of the parameter space. Using definition (1),  $\mathcal{A}$  can then be expressed as  $\mathcal{A} = \cup_{\ell=1}^L \mathcal{A}_\ell$  where  $\mathcal{A}_\ell = \{\mathbf{a}(\theta) : \theta \in \Theta_\ell\}$ . For some proper choice of  $\Theta_\ell$ , all the atoms in  $\mathcal{A}_\ell$  may be “close” to some low-dimensional subspace of  $\mathcal{H}$ . In such a case,  $\mathcal{A}_\ell$  can be accurately approximated as<sup>3</sup>

$$\hat{\mathcal{A}}_\ell = \left\{ \sum_{k=1}^K c_k \mathbf{v}_k : \mathbf{c} \in \mathcal{C} \right\} \quad (4)$$

for some  $\mathbf{v}_k \in \mathcal{H}$  and  $\mathcal{C} \subseteq \mathbb{R}^K$ . An approximation of the overall dictionary  $\mathcal{A}$  can then be simply written as

$$\hat{\mathcal{A}} = \cup_{\ell=1}^L \hat{\mathcal{A}}_\ell. \quad (5)$$

We note that (4) is tantamount to approximating the atoms of  $\mathcal{A}_\ell$  in some linear  $K$ -dimensional subspace of  $\mathcal{H}$ . In the sequel, we will thus refer to (4)-(5) as “piece-wise linear approximation of  $\mathcal{A}$ ”.

From a computational point of view, working with piece-wise linear approximations may be far more efficient than exploiting the true dictionary. This idea has thus been used in several contributions, where different choices for  $\{\mathbf{v}\}_{\ell=1}^K$  and  $\mathcal{C}$  were suggested. We describe hereafter the “polar” and “SVD” approximations proposed in [6, 8].

Assume  $\Theta$  is an interval of  $\mathbb{R}$  and  $\{\Theta_\ell\}_{\ell=1}^L$  some subintervals of  $\Theta$ , say  $\Theta_\ell = [\bar{\theta}_\ell - \frac{\Delta}{2}, \bar{\theta}_\ell + \frac{\Delta}{2}]$  where  $\{\bar{\theta}_\ell\}_{\ell=1}^L$  are points on a regular grid and  $\Delta$  is the space between two grid points. The “polar” decomposition presented in [6] considers the case where  $K = 3$  and define  $\mathcal{C}$  as

$$\mathcal{C} = \left\{ (1, \cos(\omega\tau), \sin(\omega\tau)) : \tau \in \left[-\frac{\Delta}{2}, \frac{\Delta}{2}\right] \right\} \quad (6)$$

for some  $\omega \in \mathbb{R}$ . The vectors  $\{\mathbf{v}_k\}_{k=1}^3$  are chosen so that

$$\mathbf{a}(\bar{\theta}_\ell) \in \hat{\mathcal{A}} \quad \text{and} \quad \mathbf{a}(\bar{\theta}_\ell \pm \frac{\Delta}{2}) \in \hat{\mathcal{A}}. \quad (7)$$

This choice entails in particular that perfect atom interpolation is achieved on the grid  $\{\bar{\theta}_\ell\}_{\ell=1}^L \cup \{\bar{\theta}_\ell \pm \Delta/2\}_{\ell=1}^L$ .

<sup>3</sup>The vectors  $\mathbf{v}_k$  typically differ for each subdictionary  $\hat{\mathcal{A}}_\ell$ . Nevertheless, we drop this dependence in our notations to keep our exposition as simple as possible.

The “SVD” approximation presented in [8] computes  $\{\mathbf{v}_k\}_{k=1}^K$  according to a different criterion. More specifically, the authors suggest to select  $\{\mathbf{v}_k\}_{k=1}^K$  to minimize the “mean square approximation error (MSAE)” between the atoms in  $\mathcal{A}_\ell$  and their projections onto  $V = \text{span}(\{\mathbf{v}\}_{\ell=1}^K)$  (see Section 3 for more details).

We note that the authors deviate nevertheless from the goal of achieving the minimum MSAE by constraining the coefficients  $c_k$  to belong to some pre-specified intervals of values. We elaborate on this topic in the next section. In particular, we emphasize that quasi-optimal MSAE can be achieved in some situations, by imposing polar constraints (6) on the coefficients  $c_k$ ’s.

## 3. EFFICIENT FW-IMPLEMENTATION WITH QUASI $\|\cdot\|$ -OPTIMAL APPROXIMATION

Dictionary approximations have been used to devise different “low-complexity” continuous sparse representation procedures in the literature. In [6], the authors used approximate dictionaries to build a discrete surrogate of the BLASSO problem. In [8], a modified version of orthogonal matching pursuit [13] was proposed.

In this paper, we show that fast (approximate) FW-based implementation of BLASSO can be devised by using approximated dictionaries. Our contribution is twofold. We first show that any piece-wise linear approximation enforcing polar constraints (6) allows for an efficient implementation of the atom selection step (3). Second, we emphasize that, in some particular cases, the  $\|\cdot\|$ -optimal approximation of  $\hat{\mathcal{A}}_k$  in the “SVD” subspace proposed in [8] admits a polar decomposition. Overall, combining these two ingredients, we propose a fast approximate implementation of FW for BLASSO.

### 3.1. Atom selection with polar decomposition

We focus on the atom selection step (3) when  $\mathcal{A}$  is approached by some piece-wise linear approximation (4)-(5). We consider the case where  $K = 3$  and assume that polar constraints (6) are imposed on the coefficients  $c_k$ . We make *no* assumptions on the choice of the vectors  $\{\mathbf{v}_k\}_{k=1}^3$ .

Using the partition of the approximate dictionary (5), we first have that

$$\max_{\mathbf{a} \in \hat{\mathcal{A}}} \langle \mathbf{a}, \mathbf{r} \rangle = \max_{\ell} \max_{\mathbf{a} \in \hat{\mathcal{A}}_\ell} \langle \mathbf{a}, \mathbf{r} \rangle. \quad (8)$$

Moreover, if the subdictionary  $\hat{\mathcal{A}}_\ell$  takes the form (4) and  $\mathcal{C}$  is defined as in (6), the inner maximization in the right-hand side of (8) can equivalently be rewritten as

$$\max_{\tau \in [-\frac{\Delta}{2}, \frac{\Delta}{2}]} \langle \mathbf{v}_1, \mathbf{r} \rangle + \cos(\omega\tau) \langle \mathbf{v}_2, \mathbf{r} \rangle + \sin(\omega\tau) \langle \mathbf{v}_3, \mathbf{r} \rangle. \quad (9)$$

Now, (9) has a closed form expression, as shown by the following lemma:

**Lemma 1.** Let  $f(\tau) = a \cos(\omega\tau) + b \sin(\omega\tau)$  with  $0 < \omega \leq 2\pi$ . Then,

$$\max_{\tau \in [-\frac{1}{2}, \frac{1}{2}]} f(\tau) = \begin{cases} \sqrt{a^2 + b^2} & \text{if } \rho > \bar{\eta} \\ a\bar{\eta} + |b|\sqrt{1 - \bar{\eta}^2} & \text{otherwise} \end{cases}$$

where  $\rho = \frac{a}{\sqrt{a^2 + b^2}}$  and  $\bar{\eta} = \cos(\omega/2)$ . Moreover, the unique minimizer of the problem writes

$$\arg \max_{\tau \in [-\frac{1}{2}, \frac{1}{2}]} f(\tau) = \begin{cases} \omega^{-1} \arccos(\rho) \operatorname{sign}(b) & \text{if } \rho > \bar{\eta} \\ \frac{1}{2} \operatorname{sign}(b) & \text{otherwise} \end{cases}$$

*Proof:* Using symmetry of the optimization domain and  $\sin(\omega\tau) = -\sin(-\omega\tau)$ , the problem can be rewritten as

$$\max_{\bar{\eta} \leq \eta \leq 1} a\eta + |b|\sqrt{1 - \eta^2} \quad (10)$$

by setting  $\eta = \cos(\omega\tau)$ . The final results are direct consequences of [14, Lemma 2].  $\square$

Lemma 1 thus provides an efficient procedure to solve (8) when  $\hat{\mathcal{A}}_\ell$  takes the form (4) and  $\mathcal{C}$  is defined as in (6). We show hereafter that such approximation spaces appear as quasi-optimal in a MSAE sense for translation invariant dictionaries.

### 3.2. $\|\cdot\|$ -optimal approximation

Since (4) enforces  $\hat{\mathcal{A}}_\ell$  to belong to some  $K$ -dimensional subspace of  $\mathcal{H}$  (namely  $V = \operatorname{span}(\{\mathbf{v}_k\}_{k=1}^K)$ ), we can raise the following questions:

- i) Given a  $K$ -dimensional subspace  $V$  and some atom  $\mathbf{a} \in \mathcal{A}_\ell$ , what is the “best” approximation of  $\mathbf{a}$  in  $V$ ?
- ii) How to choose some “good”  $K$ -dimensional approximation subspace for *all* the atoms in  $\mathcal{A}_\ell$ ?

The answers to these questions have been known in the literature for decades. The best approximation of  $\mathbf{a}$  in  $V$  with respect to the distance induced by  $\|\cdot\|$  is given by the so-called “orthogonal projector”:

$$P_V(\mathbf{a}) \triangleq \arg \min_{\tilde{\mathbf{a}} \in V} \|\mathbf{a} - \tilde{\mathbf{a}}\|. \quad (11)$$

$P_V(\mathbf{a})$  always exists and is unique since  $V$  is  $K$ -dimensional.

As for the second question, a common criterion to select a “good” approximation subspace is the minimization of the “average projection error”, that is

$$V_{\text{opt}}^K = \arg \min_{V: \dim(V)=K} \int_{\Theta_\ell} \|\mathbf{a}(\theta) - P_V(\mathbf{a}(\theta))\|^2 d\theta. \quad (12)$$

It can be seen [15], that the solution of this problem only depends on the kernel function specifying the inner product between two atoms of  $\mathcal{A}_\ell$ :

$$\kappa(\theta, \theta') \triangleq \langle \mathbf{a}(\theta), \mathbf{a}(\theta') \rangle. \quad (13)$$

More specifically, the following result shows that an orthogonal basis of  $V_{\text{opt}}^K$  can be found as the solution of an eigenvalue problem:

**Theorem 1** (Spectral decomposition [15]). *Let*

$$R: L_2(\Theta_\ell) \rightarrow L_2(\Theta_\ell) \\ u(\theta) \mapsto \int_{\Theta_\ell} \kappa(\theta, \theta') u(\theta') d\theta', \quad (14)$$

where  $L_2(\Theta_\ell)$  is the set of square-integrable functions on  $\Theta_\ell$ . Let  $\{\sigma_k^2\}_{k=1}^K$  be the  $K$  largest eigenvalues of  $R$  and  $\{u_k\}_{k=1}^K$  the associated orthonormal eigenfunctions. Then,

$$\mathbf{v}_k \triangleq \int_{\Theta_\ell} \mathbf{a}(\theta) u_k(\theta) d\theta, \quad k = 1 \dots K \quad (15)$$

is an orthogonal basis of  $V_{\text{opt}}^K$ . Moreover, we have  $\forall \theta \in \Theta_\ell$ :

$$P_{V_{\text{opt}}^K}(\mathbf{a}(\theta)) = \sum_{k=1}^K \sigma_k u_k(\theta) \mathbf{v}_k. \quad (16)$$

### 3.3. Raised-cosine kernels

In [8], the authors suggested to choose the “SVD” approximation vectors  $\{\mathbf{v}_k\}_{k=1}^3$  as in (15). Their dictionary approximation does however not attain the  $\|\cdot\|$ -optimal error because the decomposition coefficients of each atom  $\mathbf{a}(\theta)$  (in the definition of  $\mathcal{C}$ ) is not constrained to be equal to  $c_k = \sigma_k^2 u_k(\theta)$ , as in (16). This choice is probably due to the fact that  $u_k(\theta)$  is typically a complicated function of  $\theta$ .

In this section we emphasize that the decomposition coefficients of  $\mathbf{a}(\theta)$  (in some particular basis of  $V_{\text{opt}}^3$ ) may nevertheless have a desirable form in some cases. Our starting point is the following observation: for many standard 1D-deconvolution problems, kernel  $\kappa(\theta, \theta')$  can be accurately approximated as

$$\kappa(\theta, \theta') = a + b \cos(\omega(\theta - \theta')), \quad (17)$$

for some  $a \in \mathbb{R}$ ,  $b \in \mathbb{R}$  and  $\omega \in \mathbb{R}^+$ .

Indeed, for 1D-deconvolution problems the dictionary is translation invariant, so  $\kappa$  is a Toeplitz kernel: we have  $\kappa(\theta, \theta') = \tilde{\kappa}(\theta - \theta')$  where  $\tilde{\kappa}$  is the autocorrelation of any atom. For standard atoms,  $\tilde{\kappa}$  has a smooth symmetric bell shape, and can be accurately approximated by a raised-cosine, see Table 1.

The next theorem shows that any dictionary  $\mathcal{A}_\ell$  whose kernel satisfies (17), admits a “polar” decomposition for the projection operator onto  $V_{\text{opt}}^3$ :

**Lemma 2.** Assume the kernel function  $\kappa(\theta, \theta')$  associated to  $\mathcal{A}_\ell$  verifies (17). Then, there exists  $\{\mathbf{v}_i^* \in \mathcal{H}\}_{i=1}^3$  such that

$$P_{V_{\text{opt}}^3}(\mathbf{a}(\theta)) = \mathbf{v}_1^* + \cos(\omega\theta) \mathbf{v}_2^* + \sin(\omega\theta) \mathbf{v}_3^*. \quad (18)$$

*Proof:* Using (16), it is sufficient to show that the first three eigenfunctions of  $R$  (defined in (14)) can be written as linear combinations of  $\cos(\omega\theta)$ ,  $\sin(\omega\theta)$  and the constant function over  $\Theta_\ell$ , say  $1_{\Theta_\ell}(\theta)$ . Now, using a simple trigonometric identity, we obtain

$$\kappa(\theta, \theta') = a + b \cos(\omega\theta) \cos(\omega\theta') + b \sin(\omega\theta) \sin(\omega\theta'),$$

and therefore  $\text{span}(R) = \text{span}(1_{\Theta_\ell}(\theta), \cos(\omega\theta), \sin(\omega\theta))$ . Finally, since the first three eigenfunctions of  $R$  must belong to  $\text{span}(R)$ , we obtain the result.  $\square$

Eigenvalues and eigenfunctions of  $\kappa(\theta, \theta')$  can be derived in closed form, but, for the sake of conciseness, we restrict ourselves to the result stated in Lemma 2. We note however that the vectors  $\{\mathbf{v}_k^*\}_{k=1}^3$  can be readily evaluated from the knowledge of  $a$ ,  $b$  and  $\omega$ .

We conclude this section by identifying a basis  $\{\mathbf{v}_k\}_{k=1}^3$  such that the piece-wise linear approximation (4) obeys

$$\hat{\mathcal{A}}_\ell = \left\{ P_{V_{\text{opt}}^3}(\mathbf{a}(\theta)) : \theta \in \Theta_\ell \right\} \quad (19)$$

when  $\mathcal{C}$  is defined as in (6). This result is a direct consequence of Lemma 2, it is stated in the following Theorem 2.

**Theorem 2.** Let  $\Theta_\ell = [\bar{\theta}_\ell - \frac{\Delta}{2}, \bar{\theta}_\ell + \frac{\Delta}{2}]$ . Assume the kernel function  $\kappa(\theta, \theta')$  associated to  $\mathcal{A}_\ell$  verifies (17) and let  $\{\mathbf{v}_i^* \in \mathcal{H}\}_{i=1}^3$  be the vectors defined in Lemma 2. If

$$\begin{aligned} \mathbf{v}_1 &= \mathbf{v}_1^* \\ \mathbf{v}_2 &= \cos(\omega\bar{\theta}_\ell)\mathbf{v}_2^* + \sin(\omega\bar{\theta}_\ell)\mathbf{v}_3^* \\ \mathbf{v}_3 &= \cos(\omega\bar{\theta}_\ell)\mathbf{v}_3^* - \sin(\omega\bar{\theta}_\ell)\mathbf{v}_2^* \end{aligned} \quad (20)$$

then approximation (4) with  $\mathcal{C}$  defined as in (6) obeys (19).

The combination of Lemma 1 + Theorem 2 leads to a fast implementation of atom selection step (3) with (quasi)  $\|\cdot\|$ -optimal piece-wise linear dictionary approximation.

## 4. NUMERICAL EXPERIMENTS

We explore hereafter the generality and accuracy of the (quasi)  $\|\cdot\|$ -optimal approximation in the case of translation invariant dictionaries. We have already noticed that in such case  $\kappa$  is a Toeplitz kernel defined by autocorrelation  $\tilde{\kappa}$  of any atom. For sufficiently smooth atoms (see below the discussion on the Laplacian atom),  $\tilde{\kappa}$  has a smooth symmetric bell shape near zero. A raised-cosine approximation of such a bell shaped function is built by matching  $\tilde{\kappa}$  at points  $\{0, \Delta/2, \Delta\}$ , and closed form expressions can be derived for  $a$ ,  $b$  and  $\omega$  in terms of  $\tilde{\kappa}(0)$ ,  $\tilde{\kappa}(\Delta/2)$  and  $\tilde{\kappa}(\Delta)$ .

The accuracy of such an approximation is measured in Table 1 using a  $PSNR$  metric:

$$PSNR = 10 \log_{10} \frac{(\tilde{\kappa}(0) - \tilde{\kappa}(\Delta))^2}{MSE},$$

$\sigma$	1	2	3	4	5
$PSNR_{gauss}$	30	71	96	113	127
$PSNR_{laplace}$	0.5	23	37	47	54

**Table 1.**  $PSNR$  of the approximation of  $\tilde{\kappa}$  on  $[-1, 1]$  by a raised-cosine, for Gaussian and Laplacian atoms.  $PSNR$  is computed w.r.t  $\sigma$ , the standard deviation of the atom.

$\sigma$	0.5	1	1.5	2	2.5
bias	$3.3e-3$	$2.6e-4$	$6e-5$	$2.3e-5$	$1.1e-5$

**Table 2.** Maximal bias for estimation of  $\theta^*$  on  $[-1/2, 1/2]$  ( $\Delta = 1$ ) using (21), for Gaussian atoms. Bias is computed w.r.t  $\sigma$ , the standard deviation of the atom.

where  $MSE$  is the mean squared error between  $\tilde{\kappa}$  and its raised-cosine approximation. For the Gaussian atom, the approximation is already excellent at  $\sigma = 1$ . For the Laplacian atom, a raised-cosine approximation for  $\tilde{\kappa}$  is also available, although it has a derivative singularity at zero. The approximation of the Laplacian is less accurate than that of the Gaussian, but is already very good for  $\sigma = 2$ .

Therefore, for large enough  $\sigma$ , or, equivalently, small enough  $\Delta$ , (17) holds for Gaussian and Laplacian atoms. In such context, one may consider dictionary approximations (4) with  $\mathbf{v}_k$  defined by (20) and  $\mathcal{C}$  defined as in (6).

We next assess the ability of such approximations to achieve accurate atom selection. We study the basic problem of estimating the true location  $\theta^*$  of the atom as follows

$$\hat{\theta} = \arg \max_{\theta \in \Theta_\ell} \langle \mathbf{v}_1 + \cos(\omega\theta)\mathbf{v}_2 + \sin(\omega\theta)\mathbf{v}_3, \mathbf{a}_{\theta^*} \rangle, \quad (21)$$

where  $\{\mathbf{v}_i \in \mathcal{H}\}_{i=1}^3$  fulfills the assumptions of Theorem 2. We use Lemma 1 to compute  $\hat{\theta}$ , then define

$$\overline{bias} = \max_{\theta^* \in \Theta_\ell} |\hat{\theta}(\theta^*) - \theta^*|. \quad (22)$$

Table 2 shows that the approximation recovers the true location of the atom with very good accuracy (less than 0.004 the width of the approximation cell), improving with increasing smoothness of the atom.

## 5. CONCLUSION

This paper deals with efficient atom selection procedure in a translation invariant continuous dictionary, as required in a Frank-Wolfe approach within a BLASSO problem for 1D-deconvolution. The main results are Lemma 1 and Theorem 2 whose combination leads to a fast implementation of atom selection step (3) with (quasi)  $\|\cdot\|$ -optimal piece-wise linear dictionary approximation. Such approximations are shown to be accurate for typical Gaussian and Laplacian atoms, even if the latter has a singularity at zero on its derivative.

## 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] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing.*, Applied and Numerical Harmonic Analysis. Birkhäuser, 2013.
- [6] Chaitanya Ekanadham, Daniel Tranchina, and Eero P. Simoncelli, “Recovery of sparse translation-invariant signals with continuous basis pursuit,” *Signal Processing, IEEE Transactions on*, vol. 59, no. 10, pp. 4735–4744, Oct. 2011.
- [7] 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.
- [8] Karin C. Knudson, Jacob L. Yates, Alexander C. Huk, and Jonathan W. Pillow, “Inferring sparse representations of continuous signals with continuous orthogonal matching pursuit,” in *Advances in Neural Information Processing*, 2014.
- [9] 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.
- [10] 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.
- [11] Paul Catala, Vincent Duval, and Gabriel Peyré, “A low-rank approach to off-the-grid sparse deconvolution,” in *Journal of Physics: Conference Series*. IOP Publishing, 2017, vol. 904, p. 012015.
- [12] Marguerite Frank and Philip Wolfe, “An algorithm for quadratic programming,” *Naval Research Logistics (NRL)*, vol. 3, no. 1-2, pp. 95–110, 1956.
- [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] Cédric Herzet and Angélique Drémeau, “Joint screening tests for Lasso,” *arXiv:1710.09809*, 2017.
- [15] Erhard Schmidt, “Zur theorie der linearen und nicht-linearen integralgleichungen. iii. teil,” *Mathematische Annalen*, vol. 65, no. 3, pp. 370–399, Sept. 1908.