Interpolating and translation-invariant approximations of parametric dictionaries

Frédéric Champagnat
ONERA, The French Aerospace Lab
Palaiseau, France
frederic.champagnat@onera.fr

Cédric Herzet
INRIA Rennes-Bretagne Atlantique
Rennes, France
cedric.herzet@inria.fr

***Abstract*—In this paper, we address the problem of approximating the atoms of a parametric dictionary $\mathcal{A} = \{\boldsymbol{a}(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$, commonly encountered in the context of sparse representations in "continuous" dictionaries. We focus on the case of translation-invariant dictionaries, where the inner product between $\boldsymbol{a}(\boldsymbol{\theta})$ and $\boldsymbol{a}(\boldsymbol{\theta}')$ only depends on the difference $\boldsymbol{\theta} - \boldsymbol{\theta}'$. We investigate the following general question: is there some low-rank approximation of $\mathcal{A}$ which interpolates a subset of atoms $\{\boldsymbol{a}(\boldsymbol{\theta}_j)\}_{j=1}^J$ while preserving the translation-invariant nature of the original dictionary? In this paper, we derive necessary and sufficient conditions characterizing the existence of such an "interpolating" and "translation-invariant" low-rank approximation. Moreover, we provide closed-form expressions of such a dictionary when it exists. We illustrate the applicability of our results in the case of a two-dimensional isotropic Gaussian dictionary. We show that, in this particular setup, the proposed approximation framework outperforms standard Taylor approximation.**

***Index Terms*—Sparse representations, continuous dictionaries, translation invariance, interpolating approximations.**

I. Introduction

Sparse representations aim at representing a signal of interest as the linear combination of a few elements of a dictionary $\mathcal{A}$. Recently, this problem has been reformulated in a "continuous" setting, where the elements of $\mathcal{A}$ are continuously indexed by some parameter $\boldsymbol{\theta}$:

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

where Θ is a square interval of $\mathbb{R}^d$, $\mathcal{H}$ is a Hilbert space over the real field $\mathbb{R}$ with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\|\cdot\|$, and $\boldsymbol{a}(\cdot)$ some continuous function from Θ to $\mathcal{H}$, see *e.g.*, [1]–[4]. In this paper, we adopt the common hypothesis that $\|\boldsymbol{a}(\boldsymbol{\theta})\| = 1 \ \forall \boldsymbol{\theta} \in \Theta$.

Continuous dictionaries contain an infinite uncountable number of elements and induce therefore new difficulties. As a matter of fact, trivial operations in the discrete setting may become challenging in the continuous framework. As a simple example, one can mention the well-known "atom selection" problem [5], [6]:

$$\text{Find } \boldsymbol{\theta}^\star = \arg\max_{\boldsymbol{\theta} \in \Theta} \langle \boldsymbol{a}(\boldsymbol{\theta}), \boldsymbol{r} \rangle \quad \text{for some } \boldsymbol{r} \in \mathcal{H}, \tag{2}$$

The authors thank the project "ANR BECOSE" (ANR-15-CE23-0021) for its support.

which only involves the evaluation of a finite number of inner products in the discrete setting but can turn out to be a difficult optimization task in the continuous framework.

In order to circumvent this problem, several contributions of the literature have proposed to tackle the infinite-dimensional nature of continuous dictionaries by resorting to "low-rank" approximations of $\boldsymbol{a}(\boldsymbol{\theta})$. More formally, the idea consists in approximating the elements of $\mathcal{A}$ as linear combinations of a few vectors $\{\boldsymbol{v}_k\}_{k=1}^K$:

$$\hat{\boldsymbol{a}}(\boldsymbol{\theta}) \triangleq \sum_{k=1}^K \boldsymbol{v}_k c_k(\boldsymbol{\theta}) \tag{3}$$

for some functions $\{c_k : \Theta \to \mathbb{R}\}_{k=1}^K$. This approach was for example used in [7]–[9] to transform "continuous" sparse-representation problems into approximate (but tractable) finite-dimensional ones. The quality of the approximation obtained with this strategy obviously depends on the choice of the vectors $\{\boldsymbol{v}_k\}_{k=1}^K$ and functions $\{c_k\}_{k=1}^K$. Several options were considered in [7], [9], [10]. In [7], the authors introduced the "Taylor" and "polar" approximations: the former is based on a Taylor decomposition of $\boldsymbol{a}(\boldsymbol{\theta})$; the latter is constructed so that $\hat{\boldsymbol{a}}(\boldsymbol{\theta})$ has a unit norm $\forall \boldsymbol{\theta} \in \Theta$ and interpolates $\boldsymbol{a}(\boldsymbol{\theta})$ for some $\{\boldsymbol{\theta}_j\}_{j=1}^3$. In [9], the authors suggested to use a singular-value decomposition of $\boldsymbol{a}(\boldsymbol{\theta})$ to identify the approximation subspace minimizing the projection error in a $\|\cdot\|$-sense. In [10], the authors pointed out some connection between the approximations considered in [7] and [9].

In this paper we focus on families of parametric dictionaries exhibiting some form of translation invariance. More specifically, we consider the case where

$$\big\langle \boldsymbol{a}(\boldsymbol{\theta}), \boldsymbol{a}(\boldsymbol{\theta}') \big\rangle = \big\langle \boldsymbol{a}(\boldsymbol{\theta} - \boldsymbol{\tau}), \boldsymbol{a}(\boldsymbol{\theta}' - \boldsymbol{\tau}) \big\rangle, \tag{4}$$

$\forall \boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta$ and $\boldsymbol{\tau} \in \mathbb{R}^d$ such that $\boldsymbol{\theta} - \boldsymbol{\tau}, \boldsymbol{\theta}' - \boldsymbol{\tau} \in \Theta$, that is the inner product between two atoms of the dictionary only depends on $\boldsymbol{\theta} - \boldsymbol{\theta}'$. This setup is ubiquitous in many physical, chemical or biological problems where the observed signal is the linear combination of shifted copies of the system's impulse response, see *e.g.*, [3], [11].

The present paper is mainly of theoretical nature. We address the following general question: is there some low-rank approximation of the form (3) which: *(i)* interpolates the original dictionary over some subset of parameters $\{\boldsymbol{\theta}_j\}_{j=1}^J$;

(ii) preserves the translation invariance of the original dictionary? A formal statement of the question addressed in this paper is given in Section II. In Section III, we introduce necessary and sufficient conditions for this question to have a positive answer. Moreover, we provide closed-form expressions to build such a dictionary if it exists. In Section IV, we show how the proposed conditions particularize to the case of parametric dictionaries with separable kernels. Finally, in Section V, we illustrate the approximation performance of the proposed approach when the target dictionary is made up of Gaussian isotropic atoms. We show that, as far as the considered setup is concerned, the proposed methodology outperforms standard Taylor approximation by orders of magnitude.

II. Problem statement

We focus on low-rank approximations of the form (3) obeying the two following properties:

- *Interpolation*:

$$\hat{\boldsymbol{a}}(\boldsymbol{\theta}) = \boldsymbol{a}(\boldsymbol{\theta}) \quad \text{for } \boldsymbol{\theta} \in \{\boldsymbol{\theta}_j\}_{j=1}^{J}. \tag{5}$$

- *Translation invariance*:

$$\left\langle \hat{\boldsymbol{a}}(\boldsymbol{\theta}), \hat{\boldsymbol{a}}(\boldsymbol{\theta}') \right\rangle = \lambda_0 + \sum_{\ell=1}^{L} \lambda_\ell \cos(\boldsymbol{\omega}_\ell^{\mathrm{T}}(\boldsymbol{\theta} - \boldsymbol{\theta}')) \tag{6}$$

  for some $\{\lambda_\ell\}_{\ell=0}^{L}$ and $\{\boldsymbol{\omega}_\ell\}_{\ell=1}^{L}$.

Property (5) imposes that the approximation $\hat{\boldsymbol{a}}(\boldsymbol{\theta})$ perfectly interpolates the original atom $\boldsymbol{a}(\boldsymbol{\theta})$ for some values of the parameters $\{\boldsymbol{\theta}_j\}_{j=1}^{J}$. Property (6) enforces the inner product of the approximated atoms to obey a "raised-cosine" kernel. In particular, we note that satisfying (6) ensures the translation-invariance of $\hat{\boldsymbol{a}}(\boldsymbol{\theta})$.

In the rest of this section, we answer the two following questions: 1) is there some low-rank approximation $\hat{\mathcal{A}} = \{\hat{\boldsymbol{a}}(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$ verifying properties (5) and (6)? 2) if such an approximation exists, how to build it? Before presenting our main result, we make two important remarks, which will allow us to give a more formal statement of the two above questions. First, in order to simplify our exposition, we assume that $\{\lambda_\ell\}_{\ell=0}^{L}$ and $\{\boldsymbol{\omega}_\ell\}_{\ell=1}^{L}$ are such that the right-hand side of (6) defines a semidefinite kernel. This assumption makes sense since the kernel induced by any family of parametric atoms must necessarily be positive semidefinite.

As shown in [12, Lemma 2], this assumption also implies (under mild conditions) that any family of atoms $\hat{\mathcal{A}} = \{\hat{\boldsymbol{a}}(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$ verifying (6) spans a vector space of dimension $1_{\{\lambda_0>0\}} + 2L$, where $1_{\{.\}}$ is the "indicator" function whose value is equal to one if the statement between brackets is true and zero otherwise. As a consequence, simple dimensionality considerations impose that any approximation of the form (3) verifying (6) must necessarily satisfy $K \geq 1_{\{\lambda_0>0\}} + 2L$. Hereafter, we consider the most natural case where $K = 1_{\{\lambda_0>0\}} + 2L$. In particular, this choice entails that $\lambda_0 = 0$ if K is even and $\lambda_0 > 0$ if K is odd.

As a second remark, let us mention that if the family of interpolated points $\{\boldsymbol{a}(\boldsymbol{\theta}_j)\}_{j=1}^{J}$ is free, it can only be contained in a vector subspace of dimension greater than or equal to J. Hence, any approximation of the form (3) verifying (5) must necessarily satisfy $J \leq K$. We consider hereafter the most natural case where $J = K$.

In summary, in the following we suppose that

$$J = K$$
$$L = \lfloor K/2 \rfloor$$

where $\lfloor \cdot \rfloor$ returns the greatest integer less than or equal to its input argument. We also assume that the interpolated vectors $\{\boldsymbol{a}(\boldsymbol{\theta}_j)\}_{j=1}^{J}$ are linearly independent. In the sequel, in order to alleviate the presentation our results, we will always assume that the above hypotheses are verified without explicitly repeating them.

The main question addressed herefafter thus takes the following form:

Main Question. *Is there a rank-K approximation* (3) *that interpolates a family of K linearly independent atoms $\{\boldsymbol{a}(\boldsymbol{\theta}_j)\}_{j=1}^{K}$ and verifies* (6) *for some $\{\lambda_\ell\}_{\ell=0}^{\lfloor K/2 \rfloor}$ and $\{\boldsymbol{\omega}_\ell\}_{\ell=1}^{\lfloor K/2 \rfloor}$? If so, how to build such an approximation?*

The answer to this question is provided in Theorem 1 below. We present necessary and sufficient conditions that ensure the existence of the desired approximation. Moreover, when it exists, we provide the closed-form expressions of $\{\boldsymbol{v}_k\}_{k=1}^{K}$ and $\{c_k(\boldsymbol{\theta})\}_{k=1}^{K}$ defining the approximation.

III. Existence and construction

Our main result relates the existence of a low-rank approximation verifying (5)-(6) to some constraints on the Gram matrix of the family of interpolated atoms $\{\boldsymbol{a}(\boldsymbol{\theta}_j)\}_{j=1}^{K}$. More specifically, letting

$$\mathbf{G} \triangleq [\langle \boldsymbol{a}(\boldsymbol{\theta}_i), \boldsymbol{a}(\boldsymbol{\theta}_j) \rangle]_{i,j} \in \mathbb{R}^{K \times K}, \tag{7}$$

the following result holds:

Theorem 1. *Let $K \in \mathbb{N}$. If there exists a family of atoms* (3) *satisfying* (5)-(6)*, then*

$$\mathbf{G}(i,j) = \lambda_0 + \sum_{\ell=1}^{L} \lambda_\ell \cos(\boldsymbol{\omega}_\ell^{\mathrm{T}}(\boldsymbol{\theta}_i - \boldsymbol{\theta}_j)) \quad \forall i,j \tag{8}$$

for some $\{\lambda_\ell\}_{\ell=0}^{L}$, $\{\boldsymbol{\omega}_\ell\}_{\ell=1}^{L}$.

Conversely, assume there exist $\{\lambda_\ell\}_{\ell=0}^{L}$, $\{\boldsymbol{\omega}_\ell\}_{\ell=1}^{L}$ and $\{\boldsymbol{\theta}_j\}_{j=1}^{K}$ such that (8) *holds. Then, low-rank approximation* (3) *with*

$$c_k(\boldsymbol{\theta}) = \lambda_0 + \sum_{\ell=1}^{L} \lambda_\ell \cos(\boldsymbol{\omega}_\ell^{\mathrm{T}}(\boldsymbol{\theta} - \boldsymbol{\theta}_k)) \tag{9}$$

$$\boldsymbol{v}_k = \sum_{j=1}^{K} \boldsymbol{a}(\boldsymbol{\theta}_j)\, \mathbf{G}^{-1}(k,j), \tag{10}$$

verifies (6) *and interpolates $\{\boldsymbol{a}(\boldsymbol{\theta}_j)\}_{j=1}^{K}$.*

A proof of this result is available in the appendix of the paper. Theorem 1 provides necessary and sufficient conditions for our "Main Question" to have a positive answer. Interestingly, we see that the existence of a low-rank decomposition verifying (5)-(6) is exclusively conditioned on the existence of some particular factorization of the Gram matrix $\mathbf{G}$ (see condition (8)). Hence, Theorem 1 transforms the question of the existence of an interpolating and translation-invariant low-rank approximation into an algebraic problem where one must find a set of parameters $\{\lambda_\ell\}_{\ell=0}^L$, $\{\boldsymbol{\omega}_\ell\}_{\ell=1}^L$ verifying equation (8).

As a general remark, we mention that the existence of decomposition (8) will depend on the nature of the original dictionary $\mathcal{A}$ and the choice of the interpolation parameters $\{\boldsymbol{\theta}_j\}_{j=1}^K$. Providing a general answer to this algebraic problem is therefore a broad question which is out of the scope of this paper. In what follows, we provide nevertheless two setups of practical interest where the conditions of our theorem can be verified (and a low-rank interpolating and translation-invariant approximation thus exists).

The first setup corresponds to the case where $\Theta \subseteq \mathbb{R}$ and the parameters $\{\boldsymbol{\theta}_j\}_{j=1}^K$ are equally spaced. In this case, the Gram matrix in (7) has a Toeplitz structure and a factorization of the form (8) always exists [13]. Moreover, the parameters $\{\lambda_\ell\}_{\ell=0}^L$, $\{\boldsymbol{\omega}_\ell\}_{\ell=1}^L$ of this factorization can be evaluated via a generalized eigenvalue decomposition [13].

The generalization of previous setup to the multi-dimentional case (that is $\Theta \subseteq \mathbb{R}^d$ with $d > 1$ and the parameters $\{\boldsymbol{\theta}_j\}_{j=1}^K$ lie on a regular cartesian grid) is more tedious to handle. In particular, to the best of our knowledge, there is no general answer in the literature to the question of the existence of a factorization (8) of the block-Toeplitz matrices $\mathbf{G}$ appearing in this case, see [14]. However, as a first result in this direction, we show in the next section that such a factorization exists for families of atoms that have a translation-invariant and separable kernel.

To conclude this section, let us mention that the expressions of the vectors $\{\boldsymbol{v}_k\}_{k=1}^K$ and functions $\{c_k\}_{k=1}^K$ provided in Theorem 1 only involve simple operations and are thus easy to evaluate numerically. Hence, the construction a low-rank dictionary verifying (5)-(6) is straightforward once the set of paremeters $\{\lambda_\ell\}_{\ell=0}^L$ and $\{\boldsymbol{\omega}_\ell\}_{\ell=1}^L$ satisfying conditions (8) has been identified.

IV. Parametric dictionary with separable kernel

In this section, we particularize the general results stated in the previous section to a d-dimensional "separable" case. More particularly, we consider the setup where $\Theta = [-\frac{\Delta}{2}, \frac{\Delta}{2}]^d$ for some $\Delta > 0$. Moreover, we assume that the family of atoms to be approximated $\mathcal{A} = \{\boldsymbol{a}(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$ has a separable translation-invariant kernel, that is, without loss of generality,

$$\big\langle \boldsymbol{a}(\boldsymbol{\theta}), \boldsymbol{a}(\boldsymbol{\theta}')\big\rangle = \prod_{n=1}^{d} \rho(\boldsymbol{\theta}(n) - \boldsymbol{\theta}'(n)), \tag{11}$$

for some positive semidefinite function $\rho : [-\Delta, \Delta] \to \mathbb{R}$, where $\boldsymbol{\theta}(n)$ denotes the nth component of $\boldsymbol{\theta}$. As mentioned at the very beginning of the paper, we also assume that the elements of $\mathcal{A}$ have unit norm, so that $\rho(0) = 1$.

We show hereafter that a low-rank approximation obeying properties (5)-(6) can be constructed in that particular setup. We consider the case where $K = 2L = 2^d$ and define the interpolating points as follows:

$$\{\boldsymbol{\theta}_j\}_{j=1}^K \triangleq \big\{\boldsymbol{\theta} = \tfrac{\Delta}{2}\mathbf{s} \colon \mathbf{s} \in \{-1, 1\}^d\big\}. \tag{12}$$

Without loss of generality, we assume that the elements of $\{\boldsymbol{\theta}_j\}_{j=1}^K$ are ordered such that $\boldsymbol{\theta}_j = -\boldsymbol{\theta}_{j+L}$, $j = 1, \ldots, L$. From Theorem 1, a low-rank decomposition of the form (3) verifying (5)-(6) exists if one can find $\{\lambda_\ell\}_{\ell=0}^L$, $\{\boldsymbol{\omega}_\ell\}_{\ell=1}^L$ such that condition (8) holds. We identify below such $\{\lambda_\ell\}_{\ell=0}^L$ and $\{\boldsymbol{\omega}_\ell\}_{\ell=1}^L$. The construction of the desired interpolating and translation-invariant low-rank approximation then directly follows from (9)-(10) in Theorem 1. We first set

$$\lambda_0 = 0,\ \lambda_\ell = \frac{1}{L} \quad \ell = 1, \ldots, L. \tag{13}$$

We consider moreover the following definition for $\{\boldsymbol{\omega}_\ell\}_{\ell=1}^L$:

$$\boldsymbol{\omega}_\ell = \omega \,\mathrm{sign}\,(\boldsymbol{\theta}_\ell) \quad \ell = 1, \ldots, L, \tag{14}$$

for some $\omega \in \mathbb{R}$. Simple trigonometric operations show that this choice leads to the following "symmetric separable" low-rank kernel:

$$\big\langle \hat{\boldsymbol{a}}(\boldsymbol{\theta}), \hat{\boldsymbol{a}}(\boldsymbol{\theta}')\big\rangle = \prod_{n=1}^{d} \cos\big(\omega(\boldsymbol{\theta}(n) - \boldsymbol{\theta}'(n))\big).$$

With this definition, we finally have that $\forall i, j$:

$$\langle \hat{\boldsymbol{a}}(\boldsymbol{\theta}_i), \hat{\boldsymbol{a}}(\boldsymbol{\theta}_j)\rangle = \prod_{n=1}^{d} \rho(\boldsymbol{\theta}_i(n) - \boldsymbol{\theta}_j(n))$$

if and only if

$$\omega = \frac{1}{\Delta}\arccos(\rho(\Delta)). \tag{15}$$

V. Numerical performance

We give below a numerical illustration of the performance achievable by the proposed approximation framework. The goal of this section is not to provide an exhaustive comparison between the proposed method and other procedures of the litterature, but rather to illustrate the potential benefits of taking the translation invariance of the dictionary into account.

We consider a target dictionary $\mathcal{A}$ of the form (11) with

$$\rho(\tau) = \exp\big(-\tau^2/2\big).$$

This corresponds to the case where $\boldsymbol{a}(\boldsymbol{\theta})$ is an isotropic Gaussian function with unit variance and mean equal to $\boldsymbol{\theta}$; $\mathcal{H}$ is then the set of square-integrable functions on $\Theta = [-\frac{\Delta}{2}, \frac{\Delta}{2}]^d$. We consider the case $d = 2$ hereafter.

We construct a low-dimensional interpolating and translation-invariant approximation of this dictionary with

$K = 2L = 4$ as described in Section IV. We refer to this approximation as "ITI" (interpolating translation-invariant) in what follows. In the rest of this section, we compare the accuracy of ITI to the two-dimensional version of the Taylor approximation considered in [7]. We focus on the Taylor approximation since the polar methodology introduced in [7] only tackles one-dimensional setups, and the SVD decomposition of [9] is not straightforwardly applicable to the "atom selection" problem considered below.

In Fig. 1, we evaluate the average square approximation error achieved by ITI and Taylor approximations (that is $\int_\Theta \|\boldsymbol{a}(\boldsymbol{\theta}) - \hat{\boldsymbol{a}}(\boldsymbol{\theta})\|^2 d\boldsymbol{\theta}$) as a function of Δ. We see that both approximation methods improve their performance when Δ decreases since the range of functions to be approximated becomes smaller. Nevertheless, it can be noticed that the rate of decrease of the proposed ITI approximation ($\sim \mathcal{O}(\Delta^4)$) is clearly superior to that of Taylor approximation ($\sim \mathcal{O}(\Delta^2)$). This shows that accounting for the translation-invariance of the original dictionary into the low-rank approximation may be beneficial to improve the method accuracy.

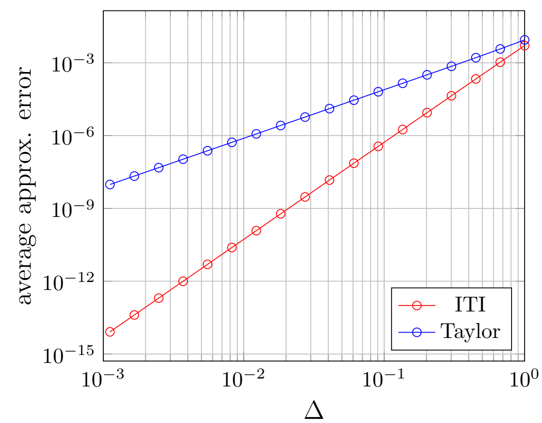

Fig. 1. Averaged approximation error versus approximation scale Δ.

In a second experiment, we assess the ability of the two approximations to solve accurately the "atom selection" problem (2). We consider the particular setup where $\boldsymbol{r} = \boldsymbol{a}(\boldsymbol{\theta})$ so that the solution of the problem simply reads $\boldsymbol{\theta}^\star = \boldsymbol{\theta}$. We then evaluate $\hat{\boldsymbol{\theta}}^\star$ by solving the following approximated "atom selection" problem as suggested in [10]

$$\hat{\boldsymbol{\theta}}^\star = \arg\max_{\boldsymbol{\theta}' \in \Theta} \langle \hat{\boldsymbol{a}}(\boldsymbol{\theta}'), \boldsymbol{r} \rangle = \arg\max_{\boldsymbol{\theta}' \in \Theta} \sum_{k=1}^{K} \langle \boldsymbol{v}_k, \boldsymbol{r} \rangle c_k(\boldsymbol{\theta}'),$$

and compute the bias $\mathbf{b}_{\boldsymbol{\theta}} \triangleq \hat{\boldsymbol{\theta}}^\star - \boldsymbol{\theta}$. Due to isotropy, the two components of the bias are identical and, due to separability, component $\mathbf{b}_{\boldsymbol{\theta}}(1)$ (resp. $\mathbf{b}_{\boldsymbol{\theta}}(2)$) is invariant with respect to $\boldsymbol{\theta}(2)$ (resp. $\boldsymbol{\theta}(1)$). Hence, we only represent the first component of the bias $\mathbf{b}_{\boldsymbol{\theta}}(1)$ as a function of $\boldsymbol{\theta}(1)$ in Fig. 2. We consider $\Delta = 1$ and $\boldsymbol{\theta}(2) = 0$ for this simulation.

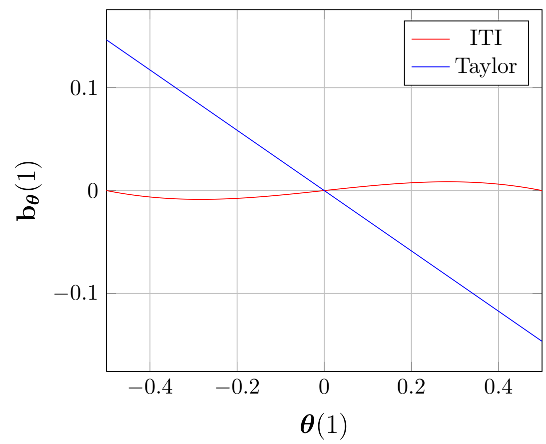

Fig. 2. Location estimation bias with respect to location.

We see that ITI and Taylor approximations have quite different behaviors: while the Taylor approximation is perfect at $\boldsymbol{\theta} = \mathbf{0}$, it linearly drifts away from the exact location of $\boldsymbol{\theta}$ when the latter moves to the domain boundary. Conversely, the ITI approximation is tight (by construction) at $\boldsymbol{\theta} = (\pm\Delta/2, \pm\Delta/2)$, so that the bias vanishes at the boundaries. Due to central symmetry in the dictionary kernel it also vanishes at $\boldsymbol{\theta} = \mathbf{0}$ and reaches a maximal bias at $\boldsymbol{\theta}(1) \approx 0.3\Delta$. The latter maximal bias is approximately seventeen times smaller than the maximal bias of the Taylor approximation. This shows that the good approximation performance emphasized in Fig. 1 has also a beneficial impact on the (approximated) resolution of the "atom selection" problem (2).

VI. Conclusions

In this paper, we address the problem of finding a "good" low-rank approximation of a parametric dictionary $\mathcal{A} = \{\boldsymbol{a}(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$. We focus on dictionaries satisfying a translation-invariance property and search for low-rank approximations preserving this feature. More specifically, we consider a family of translation-invariant approximations interpolating a subset of atoms $\{\boldsymbol{a}(\boldsymbol{\theta}_j)\}_{j=1}^K$ and investigate the two following questions: 1) is there some low-rank approximation $\hat{\mathcal{A}} = \{\hat{\boldsymbol{a}}(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$ verifying the desired interpolation and translation-invariance properties? 2) If such a dictionary exists, how to build it? Our main result is stated in Theorem 1. It provides a precise answer to these two questions. We show that the existence of $\hat{\mathcal{A}}$ is equivalent to the existence of some particular decomposition of the Gram matrix of the interpolated atoms. Upon the identification of such a decomposition, we provide closed-form expressions of the low-rank dictionary satisfying the desired interpolation and translation-invariance properties. We note that in the case of interpolated atoms on a regular grid, the Gram matrix problem decomposition is linked to the well-documented Vandermonde decomposition problem [14]. The Vandermonde decomposition is not fully resolved in the multi-dimensional case but we provide a solution to this problem in the particular case of a separable isotropic dictionary $\mathcal{A}$. Finally, we provide some numerical results showing that the proposed approximation exhibits better performance than its Taylor counterpart.

References

[1] 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.

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

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

[4] G. Tang, B. N. Bhaskar, and B. Recht, "Sparse recovery over continuous dictionaries-just discretize," in *2013 Asilomar Conference on Signals, Systems and Computers*, 2013, pp. 1043–1047.

[5] 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.

[6] 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.

[7] C. Ekanadham, D. Tranchina, and E. P. Simoncelli, "Recovery of sparse translation-invariant signals with continuous basis pursuit," *IEEE Trans. Signal Processing*, vol. 59, no. 10, pp. 4735–4744, Oct. 2011.

[8] K. Fyhn, M. F. Duarte, and S. H. Jensen, "Compressive parameter estimation for sparse translation-invariant signals using polar interpolation," *IEEE Trans. Signal Processing*, vol. 63, no. 4, pp. 870–881, Feb 2015.

[9] 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.

[10] F. Champagnat and C. Herzet, "Atom selection in continuous dictionaries: Reconciling polar and SVD approximations," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 5516–5520.

[11] Zai Yang, Jian Li, Petre Stoica, Lihua Xie, Rama Chellappa, and Sergios Theodoridis, *Chapter 11 - Sparse methods for direction-of-arrival estimation*, pp. 509–581, Academic Press, 2018.

[12] F. Champagnat and C. Herzet, "Interpolating and translation-invariant approximations of parametric dictionaries," Tech. Rep., INRIA-Rennes Bretagne Atlantique, http://people.rennes.inria.fr/Cedric.Herzet, 2020.

[13] Zai Yang and Lihua Xie, "Frequency-selective Vandermonde decomposition of Toeplitz matrices with applications," *Signal Processing*, vol. 142, no. C, pp. 157—-167, Jan. 2018.

[14] Zai Yang, Lihua Xie, and Petre Stoica, "Vandermonde decomposition of multilevel Toeplitz matrices with application to multidimensional super-resolution," *IEEE Trans. Information Theory*, vol. 62, pp. 3685–3701, 2016.

Appendix

In this appendix, we provide the technical details of the proof of Theorem 1. We denote

$$\hat{\rho}(\boldsymbol{\theta}-\boldsymbol{\theta}') \triangleq \lambda_0 + \sum_{\ell=1}^{L} \lambda_\ell \cos(\boldsymbol{\omega}_\ell^{\mathrm{T}}(\boldsymbol{\theta}-\boldsymbol{\theta}')) \tag{16}$$

and assume

$$\lambda_\ell > 0, \quad \boldsymbol{\omega}_\ell \neq 0, \quad \boldsymbol{\omega}_\ell \neq \boldsymbol{\omega}_{\ell'} \quad \forall \ell \neq \ell' \quad \forall \ell, \ell' \geq 1. \tag{17}$$

We first introduce a remarkable identity which must be verified by kernels of the form (16).

Lemma 1. *Let* $\hat{\rho}(\boldsymbol{\theta}-\boldsymbol{\theta}')$ *fulfil* (16) *and* (17). *Let* $R \triangleq 1_{\{\lambda_0>0\}} + 2L$. *Then, for any* $\{\boldsymbol{\theta}_j\}_{j=1}^R$ *such that*

$$\mathbf{G} = [\hat{\rho}(\boldsymbol{\theta}_i - \boldsymbol{\theta}_j)]_{i,j} \in \mathbb{R}^{R\times R} \tag{18}$$

is invertible, the following relation holds:

$$\hat{\rho}(\boldsymbol{\theta}-\boldsymbol{\theta}') = \sum_{i=1}^{R}\sum_{j=1}^{R} \mathbf{H}(i,j) c_i(\boldsymbol{\theta}) c_j(\boldsymbol{\theta}') \tag{19}$$

where $c_j(\boldsymbol{\theta}) \triangleq \hat{\rho}(\boldsymbol{\theta}-\boldsymbol{\theta}_j)$ *and* $\mathbf{H} = \mathbf{G}^{-1}$.

Proof: From [12, Lemma 1], there exists a family of atoms $\hat{\mathcal{A}} = \{\hat{\boldsymbol{a}}(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$ verifying $\langle \hat{\boldsymbol{a}}(\boldsymbol{\theta}), \hat{\boldsymbol{a}}(\boldsymbol{\theta}') \rangle = \hat{\rho}(\boldsymbol{\theta}-\boldsymbol{\theta}')$ and, from [12, Lemma 2], $\dim(\mathrm{span}(\hat{\mathcal{A}})) = R$. Since $\{\boldsymbol{\theta}_j\}_{j=1}^R$ is such that $\mathbf{G}$ is invertible, the family $\{\hat{\boldsymbol{a}}(\boldsymbol{\theta}_j)\}_{j=1}^R$ is free. Therefore, any $\hat{\boldsymbol{a}}(\boldsymbol{\theta}) \in \hat{\mathcal{A}}$ can be expressed as

$$\hat{\boldsymbol{a}}(\boldsymbol{\theta}) = \sum_{j=1}^{R} \hat{\boldsymbol{a}}(\boldsymbol{\theta}_j)\alpha_j \tag{20}$$

where

$$\alpha_j = \sum_{i=1}^{R} \mathbf{H}(j,i)\langle \hat{\boldsymbol{a}}(\boldsymbol{\theta}_i), \hat{\boldsymbol{a}}(\boldsymbol{\theta})\rangle = \sum_{i=1}^{R} \mathbf{H}(j,i) c_i(\boldsymbol{\theta}). \tag{21}$$

Using (20)-(21), we have

$$\langle \hat{\boldsymbol{a}}(\boldsymbol{\theta}), \hat{\boldsymbol{a}}(\boldsymbol{\theta}') \rangle = \sum_{i=1}^{R}\sum_{j=1}^{R} \mathbf{H}(i,j) c_i(\boldsymbol{\theta}) c_j(\boldsymbol{\theta}').$$

We finally obtain (19) by noticing that $\langle \hat{\boldsymbol{a}}(\boldsymbol{\theta}), \hat{\boldsymbol{a}}(\boldsymbol{\theta}') \rangle = \hat{\rho}(\boldsymbol{\theta}-\boldsymbol{\theta}')$ by definition. □

We are now ready to prove our main result:

Proof of Theorem 1: The direct part of the theorem directly follows from (5)-(6) and the definition of matrix (8).

The converse part can be shown as follows. Assume that there exist $\{\lambda_\ell\}_{\ell=0}^L$ and $\{\boldsymbol{\omega}_\ell\}_{\ell=1}^L$ such that (8) holds. We then obtain from (8) and (9):

$$\mathbf{G}(i,j) = c_j(\boldsymbol{\theta}_i). \tag{22}$$

We note that since $\{\boldsymbol{a}(\boldsymbol{\theta}_j)\}_{j=1}^K$ is assumed to be free, Gram matrix $\mathbf{G}$ is invertible. Using (22), we thus have

$$\sum_{k=1}^{K} \mathbf{G}^{-1}(k,j) c_k(\boldsymbol{\theta}_i) = \delta_{i-j}. \tag{23}$$

Combining (3) and (10) with this identity leads to

$$\begin{aligned} \hat{\boldsymbol{a}}(\boldsymbol{\theta}_i) &= \sum_{k=1}^{K} \boldsymbol{v}_k c_k(\boldsymbol{\theta}_i) \\ &= \sum_{j=1}^{K} \boldsymbol{a}(\boldsymbol{\theta}_j) \sum_{k=1}^{K} \mathbf{G}^{-1}(k,j) c_k(\boldsymbol{\theta}_i) \\ &= \sum_{j=1}^{K} \boldsymbol{a}(\boldsymbol{\theta}_j)\delta_{i-j} = \boldsymbol{a}(\boldsymbol{\theta}_i). \end{aligned} \tag{24}$$

This shows that (5) is verified.

Moreover, using (3) and (9)-(10) we have that

$$\langle \hat{\boldsymbol{a}}(\boldsymbol{\theta}), \hat{\boldsymbol{a}}(\boldsymbol{\theta}') \rangle = \sum_{i=1}^{K}\sum_{j=1}^{K} \mathbf{G}^{-1}(i,j) c_i(\boldsymbol{\theta}) c_j(\boldsymbol{\theta}').$$

(6) then directly follows from Lemma 1. □