AN ℓ_0 SOLUTION TO SPARSE APPROXIMATION PROBLEMS WITH CONTINUOUS DICTIONARIES

M. Boudineau, H. Carfantan

IRAP, Université de Toulouse/CNRS/CNES
14 Avenue É. Belin, 31 400 Toulouse, France

S. Bourguignon*

LS2N, École Centrale de Nantes/CNRS
1 rue de la Noë, 44321, Nantes Cedex 3, France

ABSTRACT

We address sparse approximation in the particular case where the dictionary is built upon the discretization of a continuous parameter. The resulting dictionary being highly correlated, equivalence between ℓ_0 and suboptimal solutions (*e.g.* greedy algorithms and convex relaxation) is not guaranteed. To tackle this issue, continuous parameter estimation has been proposed using a dictionary based on polar interpolation [1, 2]. Alternately, the exact ℓ_0-norm optimization problem can be addressed on moderate size problems through *Mixed Integer Programming* (MIP) [3]. We propose to merge these two approaches in a new MIP formulation adapted to polar interpolation. Improvements on polar interpolation and refinements on its use in the ℓ_1-norm framework are also proposed. Methods are evaluated on simulated spike train deconvolution problems, where the proposed ℓ_0-norm approach with continuous dictionary achieves the best results, although with higher computing time.

Index Terms— Sparse approximation, continuous dictionary, ℓ_0 norm, polar interpolation, spike train deconvolution.

1. INTRODUCTION

Sparse approximation (SA) of a signal $\boldsymbol{y} \in \mathbb{R}^N$ consists in solving the problem:

$$\mathcal{P}^{\mathcal{D}} \ : \ \text{estimate sparse } \boldsymbol{x} \text{ s.t. } \boldsymbol{y} \approx \mathbf{H}\boldsymbol{x} \tag{1}$$

where $\mathbf{H} \in \mathbb{R}^N \times \mathbb{R}^J$ is a *dictionary* of *atoms* $\boldsymbol{h}_j$ (column vectors), and sparse $\boldsymbol{x}$ means that few x_j are non-zero. SA has received much attention in the past decades, and can be formulated as a bi-objective optimization problem, where the reconstruction error $\|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|^2$ and the sparsity level (or ℓ_0 "norm" $\|\boldsymbol{x}\|_0 = \text{Card}\{j \,|\, x_j \neq 0\}$) are simultaneously minimized. As the ℓ_0 norm makes this problem combinatorial and NP-hard [4], many suboptimal approaches have been proposed, *e.g.,* greedy algorithms, which iteratively include new atoms in the initially-empty solution [5, 6], or the well-known convex relaxation [7, 8] with the ℓ_1 norm $\|\boldsymbol{x}\|_1 = \sum_j |x_j|$. The latter leads for example to the penalized problem:

$$\mathcal{P}^{\mathcal{D}}_{2+1}(\lambda) \ : \ \min_{\boldsymbol{x}} \|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|^2 + \lambda \|\boldsymbol{x}\|_1 \tag{2}$$

where the regularization parameter λ controls the trade-off between sparsity and reconstruction error. Solutions of such suboptimal approaches are guaranteed to be equivalent to the ℓ_0-norm one under certain conditions (*e.g.* [9]), barely summarized with a low sparsity level and low correlation between atoms. However, these properties are generally not satisfied for many inverse problems, where the dictionary $\mathbf{H}$, with atoms $\boldsymbol{h}_j = \boldsymbol{h}(\tau_j)$, results from the discretization of a continuous parameter $\tau \in \mathcal{G} = \{\tau_1, \ldots, \tau_J\}$. Examples include frequencies for spectral analysis [2, 10] or spike locations for spike train deconvolution [11, 12]. Indeed, to reduce model errors caused by discretization of the continuous parameter, the discretization step must be small, leading to a highly correlated dictionary and possibly to bad performance of suboptimal approaches [10, 13].

Two main directions have been proposed to tackle this issue. The first one considers a continuous dictionary $\{\boldsymbol{h}(\tau)\}_\tau$, so that τ can be estimated continuously. Then, the *atomic* norm, the continuous analog of the ℓ_1 norm, can be used to reformulate the problem as a semi-definite program [14, 15]. Alternately, linear approximations of $\boldsymbol{h}_j(\delta_j) = \boldsymbol{h}(\tau_j + \delta_j)$ are proposed in [1, 2] to estimate continuous shift parameters δ_j, with adaptation of classical ℓ_1-formulations and greedy algorithms. The second direction considers the exact resolution of $\mathcal{P}^{\mathcal{D}}$ in an ℓ_0 framework thanks to *Mixed Integer Programs* (MIPs), as recently proposed in [3]. Yet the solutions of such problem necessarily suffer from discretization error.

The present paper aims at merging these two directions so that the estimation of continuous nonlinear parameters τ can be performed in the ℓ_0 framework. More precisely, we consider linearization of the continuous dictionary thanks to the polar interpolation proposed in [1] and we propose a new constrained MIP formulation to estimate both amplitude and shift parameters with ℓ_0-norm-based sparsity. Additionally, we bring critical corrections to the polar interpolation in [2] in the case of real-valued x_j, and improvements of the ℓ_1-norm-based estimation method in [2] are brought.

*This work was partially supported by French National Research Agency through JCJC program (ANR project MIMOSA ANR-16-CE33-0005).

The paper is organized as follows. Section 2 presents the polar interpolation for continuous estimation in a general setting, with our corrections in the real-valued case. Then, Section 3 describes the adaptation of sparse approximation methods to continuous dictionaries, both in the ℓ_1 and in the proposed ℓ_0 framework. Section 4 compares the efficiency of both approaches on simulated sparse deconvolution problems and the conclusion in Section 5 closes the paper.

2. POLAR APPROXIMATION FOR CONTINUOUS DICTIONARIES

2.1. Dictionary with polar interpolation

To avoid the loss of precision caused by the discretization $\tau_j \in \mathcal{G}$, one may consider the following continuous sparse estimation problem, with $\boldsymbol{h}_j(\delta_j) = \boldsymbol{h}(\tau_j + \delta_j)$, $\tau_j \in \mathcal{G}$:

$$\mathcal{P}^{\mathcal{C}}: \begin{array}{c} \text{estimate } (\boldsymbol{x}, \boldsymbol{\delta}) \text{ s.t. } \boldsymbol{y} \approx \sum_j x_j \boldsymbol{h}_j(\delta_j) \\ \text{with sparse } \boldsymbol{x} \text{ and } |\delta_j| \leq \frac{\Delta}{2} \end{array}, \tag{3}$$

where Δ is the sampling step of the grid $\mathcal{G}$. This problem is more difficult due to the non-linearity of $\boldsymbol{h}_j$ in δ_j. However, if the discretization grid is fine enough, a linearization procedure of $\boldsymbol{h}_j(\delta_j)$ can be considered and classical SA methods can be adapted. In this perspective, two linearization schemes were proposed in [1]. While the most intuitive one is a Taylor expansion ($\boldsymbol{h}_j(\delta_j) \approx \boldsymbol{h}_j(0) + \delta_j \boldsymbol{h}_j'(0)$ for example), it was shown in [1] that a *polar interpolation* gives better results for translation-invariant signals. The nonlinear function $\boldsymbol{h}_j(\cdot)$ can then be approximated by:

$$\boldsymbol{h}_j(\delta_j) \approx \boldsymbol{c}_j + r\cos(\varphi_j)\boldsymbol{u}_j + r\sin(\varphi_j)\boldsymbol{v}_j,\ \varphi_j = \frac{2\theta}{\Delta}\delta_j \tag{4}$$

where constants (r, θ) and $[\boldsymbol{c}_j, \boldsymbol{u}_j, \boldsymbol{v}_j]$ are calculated from the basis elements $[\boldsymbol{h}_j(-\frac{\Delta}{2}), \boldsymbol{h}_j(0), \boldsymbol{h}_j(\frac{\Delta}{2})]$ (see [16] for analytic expressions $|\varphi_j| \leq \theta$. Then, the nonlinear model $x_j \boldsymbol{h}_j(\delta_j)$ in problem (3) can be approximated with the linear one $\alpha_j \boldsymbol{c}_j + \beta_j \boldsymbol{u}_j + \gamma_j \boldsymbol{v}_j$ with a change of variables from (x_j, δ_j) to $(\alpha_j, \beta_j, \gamma_j)$. Denoting by $\mathbf{C}$ (resp. $\mathbf{U}$, $\mathbf{V}$) the matrix with column vectors $\boldsymbol{c}_j$ (resp. $\boldsymbol{u}_j$, $\boldsymbol{v}_j$), problem (3) can be rewritten as:

$$\begin{array}{l} \text{Estimate } (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) \text{ s.t. } \boldsymbol{y} \approx \mathbf{C}\boldsymbol{\alpha} + \mathbf{U}\boldsymbol{\beta} + \mathbf{V}\boldsymbol{\gamma} \\ \text{s.t.} \left\{ \begin{array}{llll} \boldsymbol{\alpha} \text{ is sparse} & & & \\ r\cos\theta & \leq & \beta_j/\alpha_j & \forall j \in [\![1; J]\!] \\ \beta_j^2 + \gamma_j^2 & = & r^2\alpha_j^2 & \forall j \in [\![1; J]\!] \end{array} \right. \end{array}. \tag{5}$$

The equality constraint in (5) expresses that (β_j, γ_j) must be the cosine and sine of angle φ_j, multiplied by radius $r\,|\alpha_j|$. It guarantees a one-to-one mapping between (x_j, δ_j) and $(\alpha_j, \beta_j, \gamma_j)$. The inequality constraint in (5) is then equivalent to the former one $|\varphi_j| \leq \theta$. Finally, apart from approximation errors due to polar interpolation in model (4), the solution of problem (3) can be obtained by solving (5) with:

$$x_j = \alpha_j \text{ and } \delta_j = \frac{\Delta}{2\theta} \text{ atan2}\left(\frac{\gamma_j}{r\alpha_j}, \frac{\beta_j}{r\alpha_j}\right), \tag{6}$$

where $\text{atan2}(y, x) = \varphi \in]-\pi; \pi[$ with $x = R\cos(\varphi)$, $y = R\sin(\varphi)$ and $R = \sqrt{x^2 + y^2}$.

2.2. Convexification of the feasible set

The resolution of problem (5) in an optimization framework is difficult, in particular because the feasible set is not convex. This is due to (i) the quadratic equality constraint, and (ii) the non-linear constraint on β_j/α_j. For (i), [1] proposed to replace it with its convex relaxation (*i.e.* an inequality constraint). For (ii), the constraint is linear only if α_j is assumed of known sign as in [1]. Actually, for each j, the variable space, say Ω_j, for $(\alpha_j, \beta_j, \gamma_j)$ described by constraints in (5) is the union of two distinct cone surfaces (or cone sections with the convex relaxation of (i)), one for $\alpha_j \geq 0$ and the other for $\alpha_j \leq 0$. To get a convex feasible set, it is proposed in [2][1], to replace the search of one signal $\alpha_j \boldsymbol{c}_j + \beta_j \boldsymbol{u}_j + \boldsymbol{\gamma}_j \boldsymbol{v}_j$ in the non-convex space Ω_j with the search of two signals: one described by $(\alpha_j^+, \beta_j^+, \gamma_j^+)$ in the positive cone portion Ω_j^+, the other by $(-\alpha_j^-, -\beta_j^-, -\gamma_j^-)$ in Ω_j^-, with positive values for α_j^+ and α_j^-. Our formulation is quite similar to that of [2], but we correct two critical issues:

- First, we impose only one significant signal in Ω_j: at least α_j^+ or α_j^- is zero, to be coherent with problem $\mathcal{P}^{\mathcal{C}}$.
- Second, we do *not* impose any sign for $(\beta_j^+, \beta_j^-, \gamma_j^+, \gamma_j^-)$ whereas both are assumed to be positive in [2]. Indeed, the latter assumption forces (β_j, γ_j) to have the same sign than α_j, which restricts the model to shifts $\delta_j \geq 0$ (see Eq. (6)).

With the following variable substitutions: $\widetilde{\boldsymbol{\zeta}} = [\boldsymbol{\zeta}^{+T}, \boldsymbol{\zeta}^{-T}]^T$ for $\boldsymbol{\zeta} = \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$ and $\widetilde{\mathbf{Z}} = [\mathbf{Z}, -\mathbf{Z}]$ for $\mathbf{Z} = \mathbf{C}, \mathbf{U}, \mathbf{V}$, solving (5) is then equivalent to :

$$\begin{array}{l} \text{Estimate } (\widetilde{\boldsymbol{\alpha}}, \widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\gamma}}) \text{ s.t. } \boldsymbol{y} \approx \widetilde{\mathbf{C}}\widetilde{\boldsymbol{\alpha}} + \widetilde{\mathbf{U}}\widetilde{\boldsymbol{\beta}} + \widetilde{\mathbf{V}}\widetilde{\boldsymbol{\gamma}} \\ \text{s.t.} \left\{ \begin{array}{ll} \widetilde{\boldsymbol{\alpha}} \text{ is sparse} & \\ \widetilde{\alpha}_j \geq 0,\ \widetilde{\beta}_j \geq \widetilde{\alpha}_j r\cos\theta & \forall j \in [\![1; 2J]\!] \\ \widetilde{\beta}_j^2 + \widetilde{\gamma}_j^2 \leq r^2\widetilde{\alpha}_j^2 & \forall j \in [\![1; 2J]\!] \\ \alpha_j^+ \cdot \alpha_j^- = 0 & \forall j \in [\![1; J]\!] \end{array} \right. \end{array}. \tag{7}$$

3. SPARSE SOLUTIONS FOR APPROXIMATE DICTIONARIES

3.1. Solution for ℓ_1 relaxation

The use of the ℓ_1 norm was initially proposed by [1, 2] to address sparse approximation with dictionaries obtained after polar interpolation. Its adaptation to solve (7) reads:

$$\begin{array}{c} \mathcal{P}_{2+1}^{\mathcal{C}}(\lambda): \min_{\widetilde{\boldsymbol{\alpha}}, \widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\gamma}}} \left\| \boldsymbol{y} - \widetilde{\mathbf{C}}\widetilde{\boldsymbol{\alpha}} - \widetilde{\mathbf{U}}\widetilde{\boldsymbol{\beta}} - \widetilde{\mathbf{V}}\widetilde{\boldsymbol{\gamma}} \right\|^2 + \lambda \left\| \widetilde{\boldsymbol{\alpha}} \right\|_1 \\ \text{s.t. } \forall j \in [\![1; 2J]\!] \left\{ \begin{array}{c} \widetilde{\alpha}_j \geq 0,\ \widetilde{\beta}_j \geq \widetilde{\alpha}_j r\cos\theta \\ \widetilde{\beta}_j^2 + \widetilde{\gamma}_j^2 \leq r^2\widetilde{\alpha}_j^2 \end{array} \right. \end{array}. \tag{8}$$

[1]Actually, such a substitution is proposed in the case of complex-valued variables which can be easily adapted to the real-valued case.

The constraints $\alpha_j^+\alpha_j^- = 0$ in (8) are not accounted for in this problem, because they make optimization much more difficult. On the contrary, the ℓ_0-norm formulation based on MIP proposed in § 3.2 offers a natural framework for such constraints. Let us remark, however, that the obtained solution with ℓ_1 penalization always satisfied $\alpha_j^+\alpha_j^- = 0$ in our tests.

3.2. Exact ℓ_0 solution with MIP

Recently, the reformulation of the ℓ_0-norm problem as a *Mixed Integer Program* (MIP) has been proposed [3] (although earlier works can be found, *e.g.,* in [17]). It relies on introducing binary variables b_j, such that $b_j = 0 \Leftrightarrow x_j = 0$. Thus, $\|\boldsymbol{x}\|_0 = \sum_j b_j$. If the amplitudes x_j are assumed to be bounded by a given value M, called *big-M*, the former equivalence can be written as a set of linear inequality constraints: $-Mb_j \leq x_j \leq Mb_j$. Therefore, the minimization of the approximation error $\|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|^2$ for K_0-sparse solutions can be written as:

$$\mathcal{P}_{2/0}^{\mathcal{D}}(K_0)\ : \quad \min_{\boldsymbol{x},\boldsymbol{b}} \|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|^2 \text{ s.t.} \begin{cases} \boldsymbol{b} \in \{0,1\}^J \\ -M\boldsymbol{b} \leq \boldsymbol{x} \leq M\boldsymbol{b} \\ \sum_{j=1}^J b_j \leq K_0 \end{cases}. \tag{9}$$

Optimizing both continuous ($\boldsymbol{x}$) and integer ($\boldsymbol{b}$) variables, with a quadratic objective and linear constraints, is a *Mixed Integer Quadratic Program*. Even if $\mathcal{P}_{2/0}^{\mathcal{D}}(K_0)$ is still NP-hard, the MIP resolution benefits from recent progress in linear and discrete programming, and can be solved exactly in a reasonable time for small-to-medium sized problem [3].

We propose to solve the continuous sparse estimation problem (7) in this framework. First, binary variables b_j^+ and b_j^- (equivalently $\widetilde{b}_j$) are introduced to control the sparsity of α_j^+ and α_j^- respectively (equivalently $\widetilde{b}_j = 0 \Leftrightarrow \widetilde{\alpha}_j = 0$). Second, the constraint that at least one of α_j^+ or α_j^- is zero can be simply written as the linear constraint $b_j^+ + b_j^- \leq 1$. Then, minimizing the approximation error for K_0-sparse solutions subject to the polar constraints in (7) can be written as the quadratically-constrained MIP:

$$\mathcal{P}_{2/0}^{\mathcal{C}}(K_0)\ :\ \min_{\widetilde{\boldsymbol{\alpha}},\widetilde{\boldsymbol{\beta}},\widetilde{\boldsymbol{\gamma}},\widetilde{\boldsymbol{b}}} \left\|\boldsymbol{y} - \widetilde{\mathbf{C}}\widetilde{\boldsymbol{\alpha}} - \widetilde{\mathbf{U}}\widetilde{\boldsymbol{\beta}} - \widetilde{\mathbf{V}}\widetilde{\boldsymbol{\gamma}}\right\|^2$$
$$\text{s.t.} \begin{cases} \widetilde{\boldsymbol{b}} \in \{0,1\}^{2J} & \mathbf{0} \leq \widetilde{\boldsymbol{\alpha}} \\ \widetilde{\boldsymbol{\alpha}} \leq M\widetilde{\boldsymbol{b}} & \widetilde{\boldsymbol{\alpha}} r\cos\theta \leq \widetilde{\boldsymbol{\beta}} \\ \sum_{j=1}^{2J} \widetilde{b}_j \leq K_0 & \widetilde{\boldsymbol{\beta}}^2 + \widetilde{\boldsymbol{\gamma}}^2 \leq r^2\widetilde{\boldsymbol{\alpha}}^2 \end{cases}.$$
$$\text{and} \quad \forall j \in [\![1;J]\!], \quad b_j^+ + b_j^- \leq 1 \tag{10}$$

3.3. Posterior improvements of polar approximation

It is well known that the ℓ_1 norm leads to underestimated amplitudes $\widehat{\boldsymbol{x}}$ in problem $\mathcal{P}_{2+1}^{\mathcal{D}}(\lambda)$ of eq. (2). This bias can be posteriorly corrected by least-squares estimation of the solution on the support $\mathcal{S} = \{j \text{ s.t. } \widehat{x}_j \neq 0\}$:

$$\widehat{\boldsymbol{x}}_{\mathcal{S}} = \arg\min_{\boldsymbol{x}_{\mathcal{S}}} \|\boldsymbol{y} - \mathbf{H}_{\mathcal{S}}\boldsymbol{x}_{\mathcal{S}}\|^2 \tag{11}$$

where subscript $\mathcal{S}$ indexes components in $\mathcal{S}$. The problem is worse in the case of ℓ_1 penalization for polar approximation $\mathcal{P}_{2+1}^{\mathcal{C}}(\lambda)$, as the bias in $\widetilde{\boldsymbol{\alpha}}$ will produce errors on the estimation of $\widetilde{\boldsymbol{\alpha}}$, $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\gamma}}$, and then on estimated amplitudes $\boldsymbol{x}$ and shifts $\boldsymbol{\delta}$ in eq. (6). Therefore, we propose to perform such a re-estimation step for the continuous dictionary case, on the solution of $\mathcal{P}_{2+1}^{\mathcal{C}}(\lambda)$. Amplitudes $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$ are re-estimated on the support $\mathcal{S} = \{j \text{ s.t. } \widehat{\alpha}_j \neq 0\}$ of the solution of (8):

$$\begin{array}{c}\min_{\boldsymbol{\alpha}_{\mathcal{S}},\boldsymbol{\beta}_{\mathcal{S}},\boldsymbol{\gamma}_{\mathcal{S}}} \|\boldsymbol{y} - \mathbf{C}_{\mathcal{S}}\boldsymbol{\alpha}_{\mathcal{S}} - \mathbf{U}_{\mathcal{S}}\boldsymbol{\beta}_{\mathcal{S}} - \mathbf{V}_{\mathcal{S}}\boldsymbol{\gamma}_{\mathcal{S}}\|^2 \\ \text{subject to constraints in (8)}\end{array}, \tag{12}$$

where $\mathbf{C}_{\mathcal{S}} = [\text{sign}(\widehat{\alpha}_j)\boldsymbol{c}_j]_{j\in\mathcal{S}}$, $\mathbf{U}_{\mathcal{S}} = [\text{sign}(\widehat{\alpha}_j)\boldsymbol{u}_j]_{j\in\mathcal{S}}$, and $\mathbf{V}_{\mathcal{S}} = [\text{sign}(\widehat{\alpha}_j)\boldsymbol{v}_j]_{j\in\mathcal{S}}$, so that we impose the sign of the reevaluated $\boldsymbol{\alpha}_{\mathcal{S}}$ to be the same as the initial solution. Then, we compute the associated solution $(\boldsymbol{x}_{\mathcal{S}}, \boldsymbol{\delta}_{\mathcal{S}})$ thanks to identification equations (6). Finally, for both $\mathcal{P}_{2/0}^{\mathcal{C}}$ and $\mathcal{P}_{2+1}^{\mathcal{C}}$, a new dictionary $\mathbf{H}_{\boldsymbol{\delta}_{\mathcal{S}}}$ is computed with columns $\boldsymbol{h}_j(\delta_j)$, $j \in \mathcal{S}$, and the amplitudes $\boldsymbol{x}_{\mathcal{S}}$ are re-estimated in the least squares sense in a similar way than with eq. (11). It helps to correct the small amplitude errors due to the polar approximation (4).

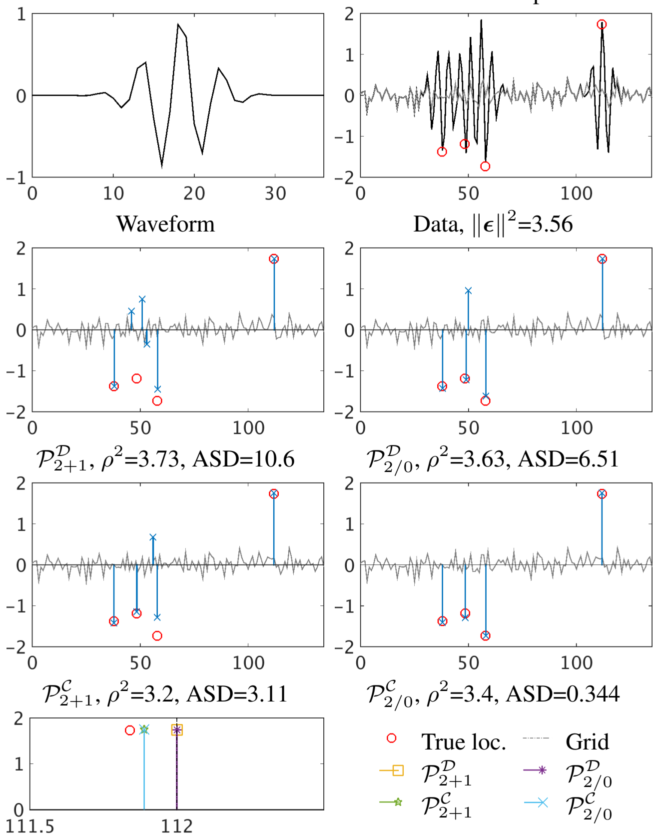
4. SIMULATION RESULTS

4.1. Description of signals and statistical tests

We illustrate the previous methods efficiency on simulated data corresponding to spike train deconvolution problems arising *e.g.* in seismic inversion or ultrasonic non-destructive testing [12, 11, 18]. It can be seen as a classical sparse approximation problem, where data $\boldsymbol{y}$ are modeled as the combination of K waveforms $\sum_{k=1}^K x_k\,\boldsymbol{h}(t-\tau_k)$, sampled at times $t_n = nT_s$, with $T_s = 1$, with additional white Gaussian noise $\boldsymbol{\epsilon}$. The used waveform is similar to that in [18] and its duration is $37T_s$ (see Figure 1). Data are simulated in $\mathbb{R}^{N=137}$ with $K = 4$ spikes, their locations τ_k are drawn uniformly in $[0; (J-1)]$ with $J = 100$ and their amplitudes x_k are drawn with random sign and uniform absolute value in $[0.5; M]$, with $M = 2$ (such M will be used in the MIP formulation (10)). 200 data sets are simulated for three signal-to-noise ratios (SNR), 10, 20 and 30 dB. The grid is naturally $\mathcal{G} = \{0, \ldots, J-1\}$, such that the interval length $\Delta = T_s = 1$. Therefore, discrete methods $\mathcal{P}^{\mathcal{D}}$ consider $\tau_k \in \mathcal{G}$ while continuous ones $\mathcal{P}^{\mathcal{C}}$ enable shifts δ_k with $|\delta_k| \leq \frac{1}{2}$. For fair comparison, we propose to tune the various sparsity-controlling parameters (λ for ℓ_1-norm problems, K_0 for ℓ_0-norm problems) on the same basis: they are tuned to find the sparsest solution such that the squared residual norm ρ^2 is at the noise level: $\rho^2 \sim \chi_0^2\sigma^2$, with σ^2 the noise variance and χ_0^2 defined such that the probability that $\rho^2/\sigma^2 < \chi_0^2$ is 95%.

Optimization is run with IBM ILOG CPLEX V12.6.0 (a free unlimited version is available to students and academics) from a Matlab interface on a computer with Intel Xeon *E5-2680* processors (40 threads) with CPUs clocked at 2.8 GHz. We compare the results with two quality indices:

Fig. 1. Waveform (top-left), data with SNR=10dB (top-right, noise ϵ in gray and signal in black) and estimation results for various SA methods. Red circles show the true spike locations, blue crosses their estimated locations. For each method, the residual is plotted in gray line, and its norm ρ^2 and ASD (see text) are given. The bottom panel shows all estimation results zoomed around the fourth spike.



- ELR: the *Exact Location Recovery* is a binary index equal to one only if exactly $K = 4$ spikes are detected and if all location errors between the true spikes and the estimated ones do not exceed $\Delta/2$.
- ASD: the *Average Spike Distance* (inspired by the one used in neuroscience [19]) compares the estimated and true spike trains by computing the quadratic error between their convolutions with a Laplacian kernel (with standard deviation $\Delta/2$), therefore accounting for both amplitude and location estimation. It is less strict than ELR as the impact of small-valued false detections is reduced.

4.2. Results and analysis

First, a result example for $\mathcal{P}^{\mathcal{D}}_{2+1}$, $\mathcal{P}^{\mathcal{D}}_{2/0}$, $\mathcal{P}^{\mathcal{C}}_{2+1}$ and $\mathcal{P}^{\mathcal{C}}_{2/0}$ with SNR = 10dB is given in Figure 1. Note that three echoes of the waveform overlap, which makes the estimation problem difficult. Solutions to $\mathcal{P}^{\mathcal{D}}$ problems (discrete grid) present false detections, which compensate for discretization errors in order to achieve a low residual, and the ℓ_1-norm solution is worse than the ℓ_0-norm one. On all data sets, we observed a high number of false detections in almost all $\mathcal{P}^{\mathcal{D}}$ results, especially when SNR decreases. One false detection is also present for the continuous ℓ_1-norm solution $\mathcal{P}^{\mathcal{C}}_{2+1}$, but this solution achieves lower ASD than discrete methods. Only the solution of $\mathcal{P}^{\mathcal{C}}_{2/0}$ shows an Exact Location Recovery. On this example, ℓ_0-norm solutions give better results and lower ASD than with ℓ_1 relaxation, for both discrete and continuous problems. Similarly, the polar approximation gives better results than the use of the discrete dictionary. The bottom panel in Figure 1 corresponds to a zoom around the fourth spike location of the results of each method: discrete estimates with $\mathcal{P}^{\mathcal{D}}_{2+1}$ and $\mathcal{P}^{\mathcal{D}}_{2/0}$ give the same grid position and continuous estimates $\mathcal{P}^{\mathcal{C}}_{2+1}$ and $\mathcal{P}^{\mathcal{C}}_{2/0}$ give similar results, closer to the true location, highlighting the benefits of polar interpolation. Note that in this example the formulation of [2] would fail to find a correct location as it would necessary find a positive shift δ_j (see § 2.2), while the true one is negative.

Table 1. ELR rate (in percent) and ASD (mean and standard deviation in brackets) averaged over 200 tests for $\mathcal{P}^{\mathcal{C}}$.

SNR	%ELR			ASD		
(dB)	$\mathcal{P}^{\mathcal{C}*}_{2+1}$	$\mathcal{P}^{\mathcal{C}}_{2+1}$	$\mathcal{P}^{\mathcal{C}}_{2/0}$	$\mathcal{P}^{\mathcal{C}*}_{2+1}$	$\mathcal{P}^{\mathcal{C}}_{2+1}$	$\mathcal{P}^{\mathcal{C}}_{2/0}$
10	24.5	31	67.5	8.1 (16.6)	6.2 (9.2)	3.6 (7.9)
20	31	39.5	90.5	2.6 (5.2)	1.6 (4.1)	0.8 (4.3)
30	30.5	34.5	88	1.3 (5.1)	0.7 (2.3)	0.05(0.3)

The ELR rate and averaged ASD over the 200 data sets are given in Table 1, for each noise level. We also show the results given by $\mathcal{P}^{\mathcal{C}*}_{2+1}$, which corresponds to $\mathcal{P}^{\mathcal{C}}_{2+1}$ without the re-estimation step proposed in §3.3. The improvement due to this step is obvious, both in terms of ELR and ASD. All methods meet difficulties at low SNR. Best results are obtained for SNR=20 dB, and get slightly worse in terms of ELR for SNR=30 dB. This unexpected result may be explained by approximation errors due to polar interpolation, which cannot be neglected compared to noise anymore. In any case, the ℓ_0 solution shows the best ELR rate (up to 90.5% for SNR=20dB) and ASD, while the ℓ_1 one suffers from a high false detection rate or erroneous shifts estimation. However, we note that solving $\mathcal{P}^{\mathcal{C}}_{2/0}$ is much more time consuming (average computation time of 400s on 3.5 threads) compared to $\mathcal{P}^{\mathcal{C}}_{2+1}$ (2 seconds / 1.8 threads).

5. CONCLUSION

We proposed a new formulation of sparse approximation with continuous dictionaries based on polar interpolation, allowing one to estimate the continuous shift parameter in the framework of ℓ_0-norm-based sparsity. In our sparse deconvolution examples, it achieved the best solutions both in terms of location recovery and spike reconstruction, compared to classical discrete ℓ_1- and ℓ_0-norm or continuous ℓ_1-norm approaches.

6. REFERENCES

[1] C. Ekanadham, D. Tranchina, and E. P. Simoncelli, "Recovery of sparse translation-invariant signals with continuous basis pursuit," *IEEE Trans. Signal Process.*, vol. 59, no. 10, 2011.

[2] K. Fyhn, M. F. Duarte, and S. H. Jensen, "Compressive parameter estimation for sparse translation-invariant signals using polar interpolation," *IEEE Trans. Signal Process.*, vol. 63, no. 4, 2015.

[3] S. Bourguignon, J. Ninin, H. Carfantan, and M. Mongeau, "Exact sparse approximation problems via mixed-integer programming: Formulations and computational performance," *IEEE Trans. Signal Process.*, vol. 64, no. 6, 2016.

[4] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comp.*, vol. 24, no. 2, pp. 227–234, 1995.

[5] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, 1993.

[6] J.A. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, 2004.

[7] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer, 2010.

[8] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. IEEE*, vol. 98, no. 6, pp. 948–958, 2010.

[9] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[10] S. Bourguignon, H. Carfantan, and J. Idier, "A sparsity-based method for the estimation of spectral lines from irregularly sampled data," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, 2007.

[11] J. Mendel, "Some modeling problems in reflection seismology," *IEEE ASSP Mag.*, vol. 3, no. 2, pp. 4–17, 1986.

[12] M. S. O'Brien, A. N. Sinclair, and S. M. Kramer, "Recovery of a sparse spike time series by L1 norm deconvolution," *IEEE Trans. Signal Process.*, vol. 42, no. 12, pp. 3353–3365, 1994.

[13] S. Bourguignon, C. Soussen, H. Carfantan, and J. Idier, "Sparse deconvolution: Comparison of statistical and deterministic approaches," in *IEEE Workshop Stat. Sig. Proc.*, 2011.

[14] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," *Foundations of Computational mathematics*, vol. 12, no. 6, pp. 805–849, 2012.

[15] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7465–7490, 2013.

[16] C. Ekanadham, *Continuous basis pursuit and its applications*, Ph.D. thesis, New York University, 2012.

[17] D. Bienstock, "Computational study of a family of mixed-integer quadratic programming problems," *Mathematical Programming*, vol. 74, no. 2, pp. 121–140, Aug 1996.

[18] E. Carcreff, S. Bourguignon, J. Idier, and L. Simon, "Resolution enhancement of ultrasonic signals by up-sampled sparse deconvolution," in *Proc. IEEE ICASSP*, 2013.

[19] M. C. W. Van Rossum, "A novel spike distance," *Neural Comp.*, vol. 13, no. 4, pp. 751–763, 2001.