An $\ell_0$ solution to sparse approximation problems with continuous dictionaries

Mégane Boudineau*, Hervé Carfantan* and Sébastien Bourguignon†

*IRAP, Université de Toulouse/CNRS/CNES, 14 avenue Edouard Belin, 31400 Toulouse, France.
†LS2N, Ecole Centrale de Nantes/CNRS 1 rue de la Noé, 44321, Nantes Cedex 3, France.

Email: Megane.Boudineau@irap.omp.eu, Herve.Carfantan@irap.omp.eu, Sebastien.Bourguignon@ec-nantes.fr

Abstract—We propose an $\ell_0$-norm solution to sparse approximation problems with a continuous estimation of localization parameters relying on a Mixed Integer Programming formulation with dictionary interpolation in the case of real valued amplitudes. We also present a linear relaxation of constraints for polar interpolation. Results are illustrated on a classical spike train deconvolution problem.

I. INTRODUCTION

Linear sparse approximation problems generally consist in approximating a signal $y \in \mathbb{R}^N$ as the best sparse linear combination of atoms $h_j$ chosen among a dictionary $H = \{ h_j \}_{j=1}^J$. A particular case is when the dictionary is built upon the fine discretization of a continuous parameter, i.e. $h_j = h(\nu_j)$ where $\nu_j$ belong to a regular grid. The resulting dictionary being highly correlated, equivalence between $\ell_0$ and suboptimal solutions (e.g. greedy algorithms and convex relaxation) is not guaranteed. Here, we tackle this issue combining two approaches: (i) the use of continuous dictionaries linearized using interpolation as introduced in [1], and (ii) an exact $\ell_0$-norm solution thanks to a Mixed Integer Programming (MIP) formulation as proposed in [2] in the linear case.

A continuous dictionary is built upon $J$ continuous atoms $h(\nu_j) = h_j(\delta_j)$, where continuous parameters $\nu_j$ can be written $\nu_j + \delta_j$ with $|\delta_j| \leq \Delta/2$. Then, sparse approximation amounts to estimate $\{ \delta_j \}_j$ and sparse $x = \{ x_j \}_j$ such that $y \approx \sum_j h_j(\delta_j)x_j$. To face the non-linearity introduced by the shift parameter $\delta_j$, [1] proposed to build a linear approximation of the continuous dictionary. It consists in linearizing the signal $x H_j(\delta_j)$ as the combination of $K$ columns such as $H x_j$. Constraints on the resulting amplitudes $\tilde{x}_j \in \mathcal{H}_j$ have to be taken into account to translate the variable substitutions from $(x_j, \delta_j)$ to $\tilde{x}_j$ and the constraints $|\delta_j| \leq \Delta/2$. Among linear approximations proposed up to now, such as Taylor expansion [1], K-SVD [3] or Minimax [4] approximation, the polar interpolation [1] provides a weak approximation error. Thanks to these approximations, the problem can now be addressed as a linear sparse approximation problem, under the constraints $\tilde{x}_j \in \mathcal{H}_j$.

II. MIP FORMULATION

In the non negative case where $x_j \geq 0$ with feasible set $\mathcal{H}_j^+$, this problem can be addressed with the sparsity-constrained formulation: $\min_{x_j} \| y - \sum_j h_j(\delta_j)x_j \|_2^2$ s.t. $\tilde{x}_j \in \mathcal{H}_j^+$ and $\| \tilde{x}_j \|_0 \leq L_0$ for a given maximal model order $L_0$. Then, following [2], it is straightforward to formulate this problem as a MIP, by introducing binary variables $\{ b_j \}_j$ ruling the nullity of $\tilde{x}_j$. The only difference with [2] is the presence of constraints $\tilde{x}_j \in \mathcal{H}_j^+$, which can easily be taken into account by MIP solvers when convex.

In the general case $x_j \in \mathbb{R}$, the feasible set $\mathcal{H}_j$ cannot be convex. In particular, in agreement with [5], we showed in [6] for polar interpolation, and in [7] for other interpolations, that instead of $\tilde{x}_j$, it is necessary to consider $(\tilde{x}_j^+, \tilde{x}_j^-) \in \mathcal{H}_j^{+2}$ such that $\tilde{x}_j = \tilde{x}_j^+ - \tilde{x}_j^-$. Moreover, we demonstrated the key role of an additional constraint, not taken into account in [5], that guarantees that $\tilde{x}_j^+$ and $\tilde{x}_j^-$ can not simultaneously be non-zero.

Then, we proposed the following MIP formulation:

$$\min \| y - \tilde{H}(\tilde{x}^+ - \tilde{x}^-) \|_2^2 \text{s.t.}$$

$$\begin{align}
& -Mb_j^+ \leq \tilde{x}_j^+ \leq Mb_j^+ \\
& b_j^+ + b_j^- \leq 1 \\
& (\tilde{x}_j^+, \tilde{x}_j^-) \in \mathcal{H}_j^{+2} \\
& \sum b_j^+ + b_j^- \leq L_0 \\
& \end{align}$$

with $\tilde{H} = \{ \tilde{h}_j \}_j$ and where $b_j^+ + b_j^- \leq 1$, associated with other constraints, ensures that $\tilde{x}_j^+$ and $\tilde{x}_j^-$ can not be non-zero at the same time. This formulation can easily be taken into account by MIP solvers for convex spaces $\mathcal{H}_j^+$, and be efficiently solved for small to medium-sized problems.

III. CONSTRAINTS RELAXATION FOR POLAR INTERPOLATION

Polar interpolation, introduced by [1], writes as $x, h_j(\delta_j) \approx x_0 c_j + x_1 r_j \cos(2\theta_j \delta_j / \Delta) u_j + x_2 r_j \sin(2\theta_j \delta_j / \Delta) w_j$ where $(r_j, \theta_j)$ and $h_j = \{ c_j, u_j, v_j \}$ are computed from $[h_j(-\Delta/2), h_j(0), h_j(\Delta/2)]$. The convex feasible set $\mathcal{H}_j^+$ is defined from the image of the variable change from $(x_1, \delta_j)$ to $\tilde{x}_j$ which is illustrated in Fig. 1(a) and can be written as:

$$\tilde{x}_j \in \mathbb{R}^3 \text{s.t.} \begin{align}
& \tilde{x}_{j,1} \geq 0, \tilde{x}_{j,2} \geq \tilde{x}_{j,1} r_j \cos(\theta_j) \\
& \tilde{x}_{j,3} \geq \tilde{x}_{j,2} + \tilde{x}_{j,3} r_j^2 \tilde{x}_{j,1} \\
& \end{align}$$

As this space is not convex, [1] proposed a convex relaxation, represented in Fig. 1(b), substituting the equality constraint with an inequality one, which actually leads to the convex hull of the image. However, even if this quadratic constraint can be formulated as a SOCP in convex programming and can be taken into account in MIP formulation, it strongly increases the computation time.

Consequently, we proposed in [7] a linear relaxation of $\mathcal{H}_j^+$ illustrated in Fig. 1(c). The proposed relaxation is very similar to the one of [1]: it allows to benefit from weak approximation error of polar interpolation while avoiding the use of quadratic constraints and then enables a faster resolution of MIP.

IV. EXPERIMENTS AND CONCLUSION

Some results of the proposed method with polar interpolation, mainly extracted from [6], [7], are shown in Fig. 2, Tab. I and Tab II on a classical example of spike train deconvolution. The results of the proposed method $P_{\ell_0}^2$ (continuous dictionary, $\ell_0$ constraint) is compared to the one of the discrete dictionary methods $P_{D,\ell_1}^2$ ($\ell_1$ penalization) and $P_{D,\ell_0}^2$ ($\ell_0$ constraint) and the $P_{\ell_0}^2$ with a continuous dictionary. Tab II illustrates saving of computing time using proposed linear constraints instead of quadratic ones; while both kinds of constraints give same detection results as in Tab I.
Fig. 1. Representation of different feasible set \( H^+ \) for polar interpolation with fixed \( \bar{x}_1 \geq 0 \), at an index \( j \) not specified: (a) image space of variable change (red), (b) \( H^+ \) proposed in [1] with a quadratic constraint based on a circle arc (yellow) and (c) our proposition (see [7]) with linear constraints based on tangents to the circle arc (green).

Fig. 2. Waveform (top-left), data with \( \| \epsilon \|_2^2 = 3.56 \) (top-right, noise \( \epsilon \) in gray and signal in black) and estimation results for various sparse approximation 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 \( \zeta^2 \) is given. The bottom panel shows all estimation results zoomed around the fourth spike.

### Table I

Exact recovery location (ELR - equal to one only if the four spikes are detected with location error not exceeding \( \Delta/2 \)) rate averaged over 200 tests for \( P^{2+1}_C \) (in percent). \( P^{2+1}_C \) is for \( P^{2+1}_C \) method without improvements we proposed in [6].

<table>
<thead>
<tr>
<th>SNR (dB)</th>
<th>( P^{2+1}_C )</th>
<th>( P^{2+1}_C )</th>
<th>( P^{2+1}_C/0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>24.5</td>
<td>31</td>
<td>67.5</td>
</tr>
<tr>
<td>20</td>
<td>31</td>
<td>39.5</td>
<td>90.5</td>
</tr>
<tr>
<td>30</td>
<td>30.5</td>
<td>34.5</td>
<td>88</td>
</tr>
</tbody>
</table>

### Table II

Mean resolution time averaged over 200 tests for \( P^{2+1}_C/0 \) with \( L_0 = 4 \) and polar interpolation. Comparison between the use of a quadratic constraint as in [1] and linear constraints as we proposed in [7].

<table>
<thead>
<tr>
<th>SNR (dB)</th>
<th>Mean resolution time for ( P^{2+1}_C/0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>796 s</td>
</tr>
<tr>
<td>20</td>
<td>302 s</td>
</tr>
<tr>
<td>30</td>
<td>218 s</td>
</tr>
</tbody>
</table>

### References


