An ℓ_0 solution to sparse approximation problems with continuous dictionaries

Mégane Boudineau*, Hervé Carfantan*, Sébastien Bourguignon†
*IRAP, Université de Toulouse, CNRS, UPS, CNES, Toulouse, France
†LS2N, École Centrale de Nantes/CNRS, Nantes, France

Paper's objective

Sparsity with a continuous dictionary

$\mathcal{P}^{\mathcal{C}}$: Estimate $(\boldsymbol{x}, \boldsymbol{\delta})$ s.t. $\boldsymbol{y} \approx \sum_j x_j \boldsymbol{h}_j(\delta_j)$ with sparse $\boldsymbol{x}$ and $|\delta_j| \leq \frac{\Delta}{2}$

- *Sparse amplitudes*: $\boldsymbol{x} \in \mathbb{R}^J$ s.t. few $x_j \neq 0$
- *Continuous atoms*: $\boldsymbol{h}_j(\delta_j) = \boldsymbol{h}(\tau_j + \delta_j)$
- *Discretization grid*: $\mathcal{G}$ s.t. $\tau_j \in \mathcal{G}$
- *Discretization step*: $\Delta = \tau_{j+1} - \tau_j$

Main difficulties

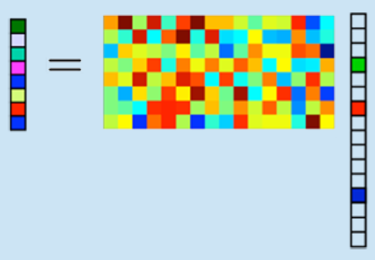
- Nonlinearity of $\boldsymbol{h}_j(\cdot)$ *w.r.t.* δ_j
- Sparse $\boldsymbol{x}$

Classical approach: Linear sparse approximation

Discrete model: impose $\delta_j = 0$

$\mathcal{P}^{\mathcal{D}}$: Estimate sparse $\boldsymbol{x}$ s.t.

$$\boldsymbol{y} \approx \mathbf{H}\,\boldsymbol{x} = \sum_{j=1}^{J} \boldsymbol{h}_j x_j \quad (\boldsymbol{h}_j = \boldsymbol{h}_j(0) = \boldsymbol{h}(\tau_j))$$



- Linear combinaison of *atoms* $\boldsymbol{h}_j$ among a *dictionary* $\mathbf{H}$
- Sparse: few $x_j \neq 0$

$$\mathcal{P}^{\mathcal{D}}_{2/0}(K_0): \min_{\boldsymbol{x}} \|\boldsymbol{y} - \mathbf{H}\,\boldsymbol{x}\|^2 \text{ s.t. } \|\boldsymbol{x}\|_0 \leq K_0$$

$\|\boldsymbol{x}\|_0 = \text{Card}(j \mid x_j \neq 0) \to \ell_0$ pseudo-norm

$\Longrightarrow$ NP-hard problem

Suboptimal approaches

- Greedy algorithms: iterative choice of $\boldsymbol{h}_j$
- Convex relaxation: with $\|\boldsymbol{x}\|_1 = \sum_j |x_j|$

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

☺ (Greedy, ℓ_1) $\Leftrightarrow$ ℓ_0 for low sparsity level or a low-correlated *dictionary* $\mathbf{H}$

☹ No equivalence for correlated *dictionary* $\mathbf{H}$

Exact ℓ_0 resolution [1]

- Introduction of binary variables b_j s.t.

$$x_j = 0 \Leftrightarrow b_j = 0$$

- *Big-M* hypothesis: $|x_j| \leq M b_j$

$\Rightarrow$ Reformulation of $\mathcal{P}^{\mathcal{D}}_{2/0}(K_0)$:

Mixed Integer Program (MIP)

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

References

[1] 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.

[2] 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.

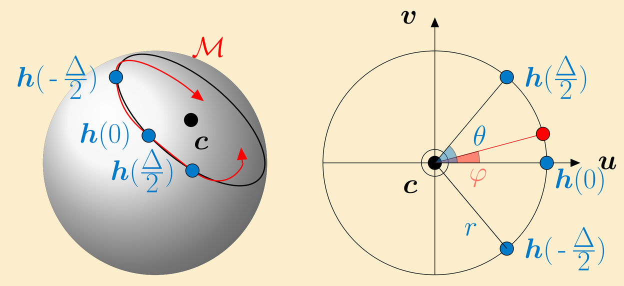
[3] K. Fyhn, M. F. Duarte, and S. H. Jensen, "Compressive parameter estimation for sparse translation-invariant signals using polar interpolation," *IEEE Trans. Signal Process.*, vol. 63, no. 4, 2015.

Linearization with polar interpolation [2]

Linearization of $\boldsymbol{h}(\cdot)$

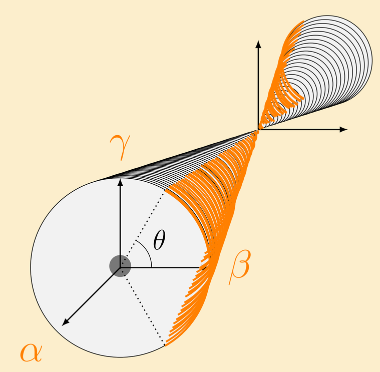
1. Taylor: $\boldsymbol{h}(\delta) \approx \boldsymbol{h}(0) + \delta \boldsymbol{h}'(0)$
2. Polar:

$\varphi = \frac{2\theta}{\Delta}\delta$
$\Downarrow$
$|\delta| \leq \frac{\Delta}{2} \Leftrightarrow |\varphi| \leq \theta$



$$x\,\boldsymbol{h}(\delta) \approx x\,\boldsymbol{c} + x\,r\cos\varphi\,\boldsymbol{u} + x\,r\sin\varphi\,\boldsymbol{v}$$
$$\approx \alpha\,\boldsymbol{c} + \beta\,\boldsymbol{u} + \gamma\,\boldsymbol{v}$$

Bijective change of variables

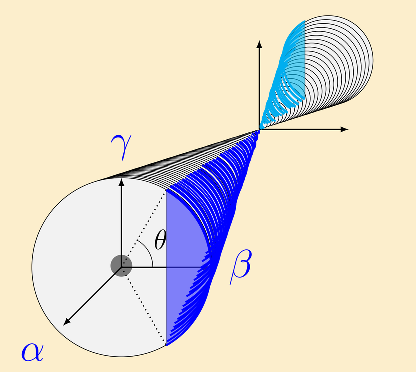


$$(\alpha, \beta, \gamma) = f(x, \delta)$$
$$(x, \delta) \in (\mathbb{R} \times]-\tfrac{\Delta}{2}; \tfrac{\Delta}{2}[)$$
$$\Leftrightarrow \Omega: \begin{cases} \alpha \in \mathbb{R} \\ \beta^2 + \gamma^2 = r^2\alpha^2 \\ \beta/\alpha \geq r\cos\theta \end{cases}$$

$$f^{-1}: x = \alpha,\ \delta = \frac{\Delta}{2\theta}\,\text{atan2}\left(\frac{\gamma}{r\alpha}, \frac{\beta}{r\alpha}\right)$$

⚠ Feasible set Ω not convex!

Feasible set's convex-relaxation



1. Relaxation: from cone's surface to cone's volume
2. Replace (α, β, γ) in Ω with
$\begin{cases} (\alpha^+, \beta^+, \gamma^+) \text{ in } \Omega^+ \\ (-\alpha^-, -\beta^-, -\gamma^-) \text{ in } \Omega^- \end{cases}$

Reformulation of $\mathcal{P}^{\mathcal{C}}$

Notations
- $\mathbf{C}, \mathbf{U}, \mathbf{V}$: matrix with column vectors $\boldsymbol{c}_j, \boldsymbol{u}_j, \boldsymbol{v}_j$
- $\widetilde{\mathbf{C}} = [\mathbf{C}, -\mathbf{C}]$ (same for $\mathbf{U}, \mathbf{V}$),
- $\widetilde{\alpha} = [\boldsymbol{\alpha}^{+T}, \boldsymbol{\alpha}^{-T}]^T$ (same for $\boldsymbol{\beta}, \boldsymbol{\gamma}$)

1. Estimate $(\widetilde{\alpha}, \widetilde{\beta}, \widetilde{\gamma})$ s.t. $\boldsymbol{y} \approx \widetilde{\mathbf{C}}\widetilde{\alpha} + \widetilde{\mathbf{U}}\widetilde{\beta} + \widetilde{\mathbf{V}}\widetilde{\gamma}$

$$\text{s.t.} \begin{cases} \widetilde{\alpha} \text{ is sparse} \\ \widetilde{\alpha} \geq 0,\ \widetilde{\beta} \geq \widetilde{\alpha} r\cos\theta,\ \widetilde{\beta}^2 + \widetilde{\gamma}^2 \leq r^2\widetilde{\alpha}^2 \\ \boldsymbol{\alpha}^+ \cdot \boldsymbol{\alpha}^- = 0 \end{cases}$$

2. Reconstruct $\boldsymbol{\alpha} = \boldsymbol{\alpha}^+ - \boldsymbol{\alpha}^-$ (and so for $\boldsymbol{\beta}, \boldsymbol{\gamma}$)
3. Recover $(\boldsymbol{x}, \boldsymbol{\delta}) = f^{-1}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$

Differences with [3]:

- At least α_j^+ or α_j^- is zero.
- No imposed sign for $(\beta_j^+, \beta_j^-, \gamma_j^+, \gamma_j^-)$.

Solution of ℓ_1 relaxation [2, 3]

$$\mathcal{P}^{\mathcal{C}}_{2+1}(\lambda): \min_{\widetilde{\boldsymbol{\alpha}}, \widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\gamma}}} \left\|\boldsymbol{y} - \widetilde{\mathbf{C}}\widetilde{\boldsymbol{\alpha}} - \widetilde{\mathbf{U}}\widetilde{\boldsymbol{\beta}} - \widetilde{\mathbf{V}}\widetilde{\boldsymbol{\gamma}}\right\|^2 + \lambda \|\widetilde{\boldsymbol{\alpha}}\|_1$$
$$\text{s.t. } \widetilde{\boldsymbol{\alpha}} \geq 0,\ \widetilde{\boldsymbol{\beta}} \geq \widetilde{\boldsymbol{\alpha}} r\cos\theta,\ \widetilde{\boldsymbol{\beta}}^2 + \widetilde{\boldsymbol{\gamma}}^2 \leq r^2\widetilde{\boldsymbol{\alpha}}^2$$

$\Rightarrow$ Quadratically constrained quadratic program (QCP).

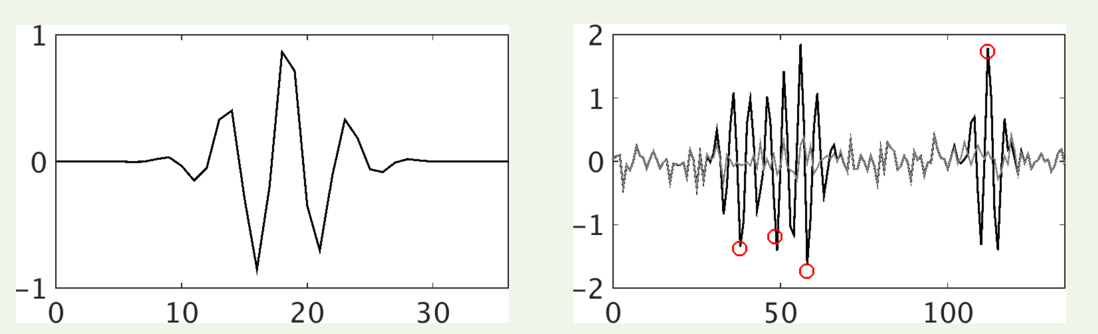
$+$ Posterior reevaluation of polar variables on estimated support ($\mathcal{P}^{\mathcal{C}*}_{2+1} \to \mathcal{P}^{\mathcal{C}}_{2+1}$).

Exact ℓ_0 resolution

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

$\Rightarrow$ Mixed integer quadratically constrained program (MIQCP).

Description of statistical tests



Waveform $\boldsymbol{h}(\cdot)$ — Example $\boldsymbol{y}$, SNR=10dB

Signals simulation description

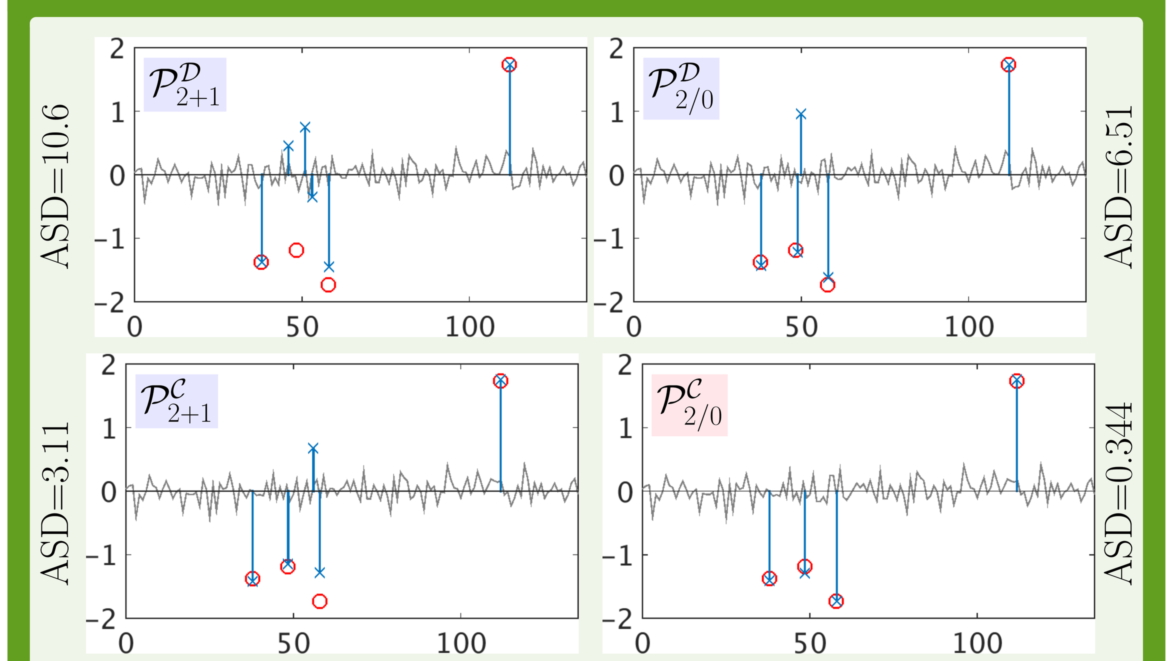
- Signal model: $y_n = \sum_{k=1}^{4} x_k\, h(n - \tau_k) + \epsilon_n$, with $\epsilon_n \sim \mathcal{N}(0, \sigma^2)$, $\tau_k \sim \mathcal{U}([0; 99[)$ and $x_k \sim \pm\mathcal{U}([0.5; 2[)$.
- 200 data sets simulated for SNR: 10,20,30 dB.
- Stopping criteria: sparsest solution s.t. the residual ρ^2 is at the noise level.

Two quality indices:

- ELR: the *Exact Location Recovery*.
ELR $= 1 \Leftrightarrow \left(\widehat{K} = 4 \text{ and } \forall k,\ |\widehat{\tau}_k - \tau_k| \leq \frac{\Delta}{2}\right)$
- ASD: the *Average Spike Distance*.
compares the estimated and true spike trains, accountes for both $\widehat{\tau}_k$ and $\widehat{x}_k$ estimation.

Optimization set-up: Optimization run with IBM ILOG CPLEX V12.6.0 from a Matlab interface.

A result example for SNR=10dB



Estimation results for various SA methods.
Red circle : true locations.
Blue crosses: estimated locations.

Statistics results

ELR rate and ASD averaged over 200 tests for $\mathcal{P}^{\mathcal{C}}$

SNR (dB)	% ELR $\mathcal{P}^{\mathcal{C}*}_{2+1}$	% ELR $\mathcal{P}^{\mathcal{C}}_{2+1}$	% ELR $\mathcal{P}^{\mathcal{C}}_{2/0}$	ASD(std) $\mathcal{P}^{\mathcal{C}*}_{2+1}$	ASD(std) $\mathcal{P}^{\mathcal{C}}_{2+1}$	ASD(std) $\mathcal{P}^{\mathcal{C}}_{2/0}$
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)

Computation time
- $\mathcal{P}^{\mathcal{C}}_{2+1}$: 2s on 1.8 threads.
- $\mathcal{P}^{\mathcal{C}}_{2/0}$: 400s on 3.5 threads.