

SUPPRESSION DE LIGNE DE BASE ET DÉBRUITAGE DE CHROMATOGRAMMES PAR PÉNALISATION ASYMÉTRIQUE DE POSITIVITÉ ET DÉRIVÉES PARCIMONIEUSES (BEADS)

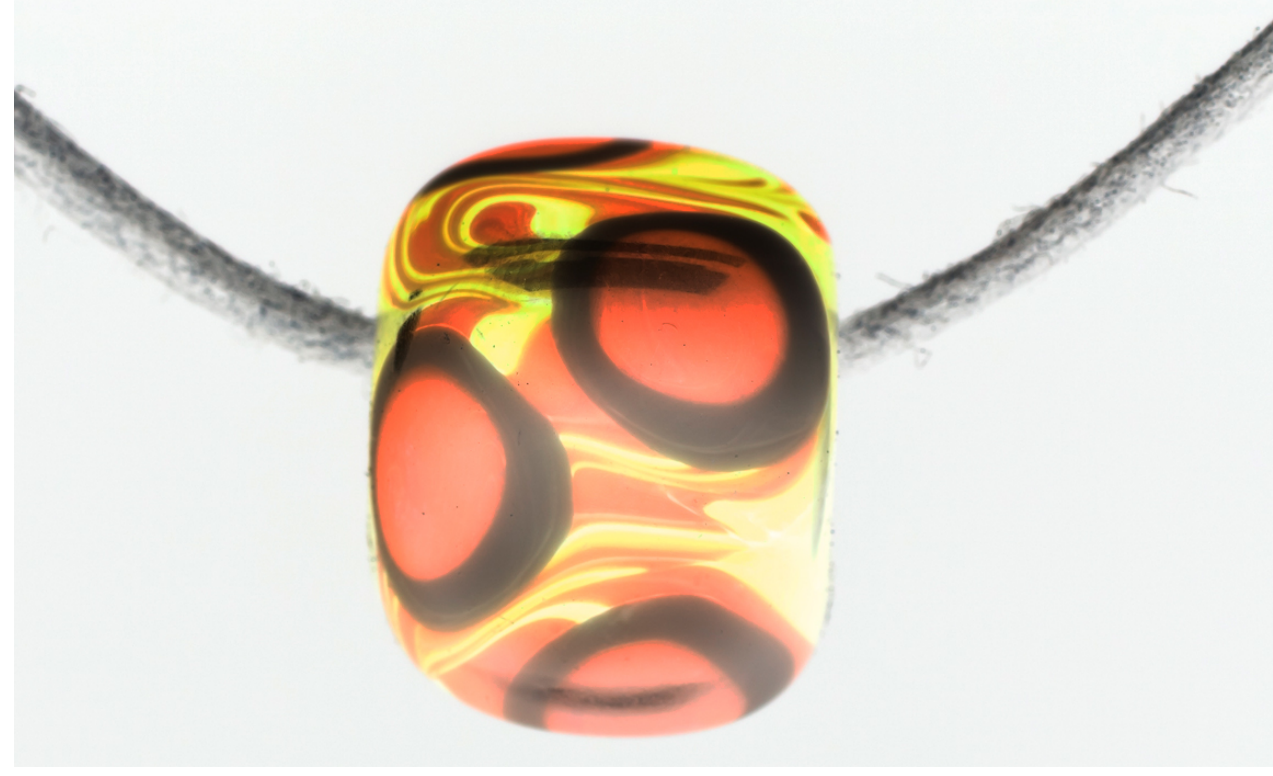
Laurent Duval¹, Aurélie Pirayre¹, Xiaoran Ning² et Ivan W. Selesnick²

¹ IFP Energies nouvelles, 92852 Rueil-Malmaison Cedex, France

² Polytechnic School of Engineering, New York University, Brooklyn, NY 11201, USA

BEADS, en très rapide

- Question : où passe le collier derrière la perle ?
- Réponse : linéarité, parcimonie, positivité, asymétrie



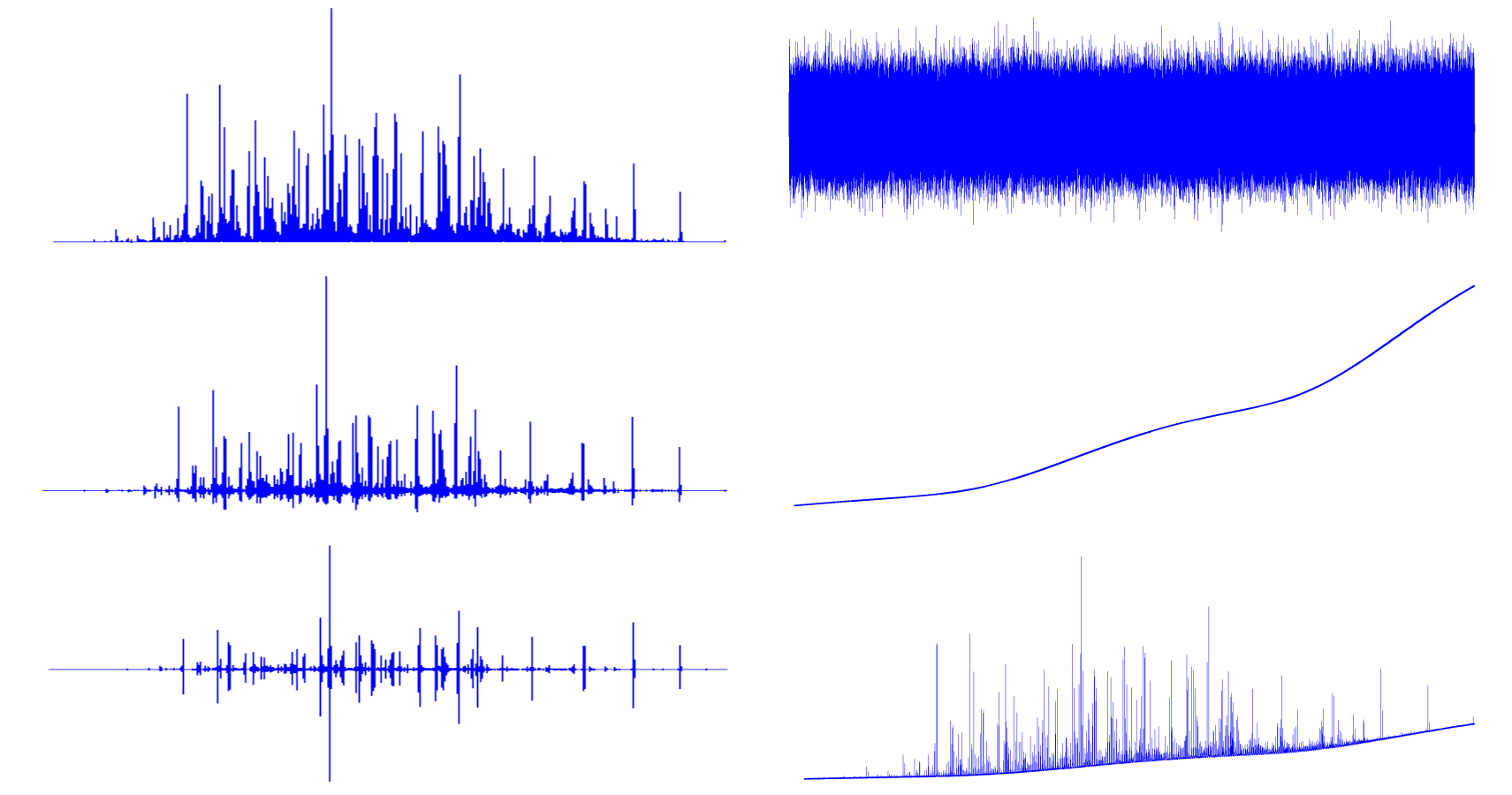
Boîte à outils Matlab™ : <http://lc.cx/beads>

Le fond du fond



- Sujet : arrière-plan et évaluation/comparaison quantitatives
- Biais (instrumental), tendance (saisonnière), fond
- Physico-chimie : bruit de fond, dérive, ligne de base, continu
- Modélisation paramétrique peu fréquente ou fiable

Signaux élémentaires



Éléments-clefs

- Séparation conjointe de trois composantes morphologiques
- Parcimonie des dérivées successives, positivité relative
- Résolution par minimisation-majoration
- Matrices de filtrage en bande : simplicité algorithmique et signaux longs

Hypothèses

- Régularité : sans éléments saillants \mathbf{p} (estimés), le fond peut être estimé $\hat{\mathbf{l}}$ par filtrage passe-bas
- $\hat{\mathbf{l}} = \mathbf{L}(\mathbf{c} - \hat{\mathbf{p}})$ (\mathbf{L} : filtre passe-bas),
- formulé par le résidu $\|\mathbf{c} - \hat{\mathbf{s}}\|_2^2 = \|\mathbf{H}(\mathbf{c} - \hat{\mathbf{p}})\|_2^2$,
- $\mathbf{H} = \mathbf{I} - \mathbf{L}$: filtre passe-haut.

Notations des signaux

Décomposition : $\mathbf{c} = \mathbf{p} + \mathbf{l} + \mathbf{b}$, $(\mathbf{c}, \mathbf{p}, \mathbf{l}, \mathbf{b}) \in (\mathbb{R}^N)^4$.

- \mathbf{p} : pics sans bruit, dérivées première et seconde (gauche),
- \mathbf{b} : bruit (haut, droite),
- \mathbf{l} : ligne de base (centre, droite),
- \mathbf{c} : observation (bas, droite).

Formulation

L'estimation de $\hat{\mathbf{p}}$ et $\hat{\mathbf{l}}$ peut s'obtenir (\mathbf{D}_i : opérateurs de différences finies) via :

$$\hat{\mathbf{p}} = \arg \min_{\mathbf{p}} \left\{ F(\mathbf{p}) = \frac{1}{2} \|\mathbf{H}(\mathbf{c} - \mathbf{p})\|_2^2 + \lambda_0 \sum_{n=0}^{N-1} \theta_\epsilon(x_n; r) + \sum_{i=1}^M \lambda_i \sum_{n=0}^{N_i-1} \phi([\mathbf{D}_i \mathbf{p}]_n) \right\},$$

$$\hat{\mathbf{l}} = (\mathbf{D}_0 - \mathbf{H})(\mathbf{c} - \hat{\mathbf{p}}).$$

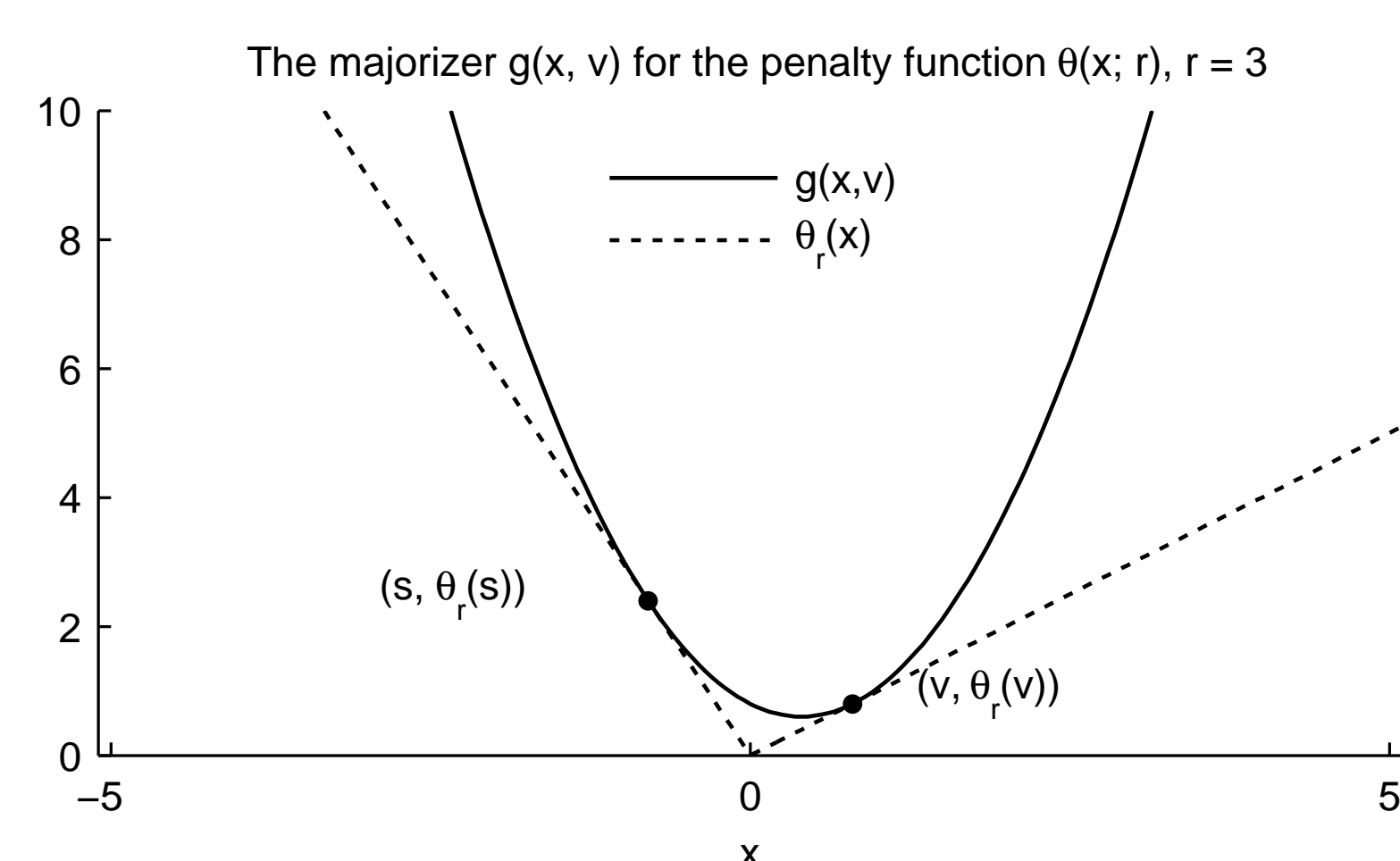
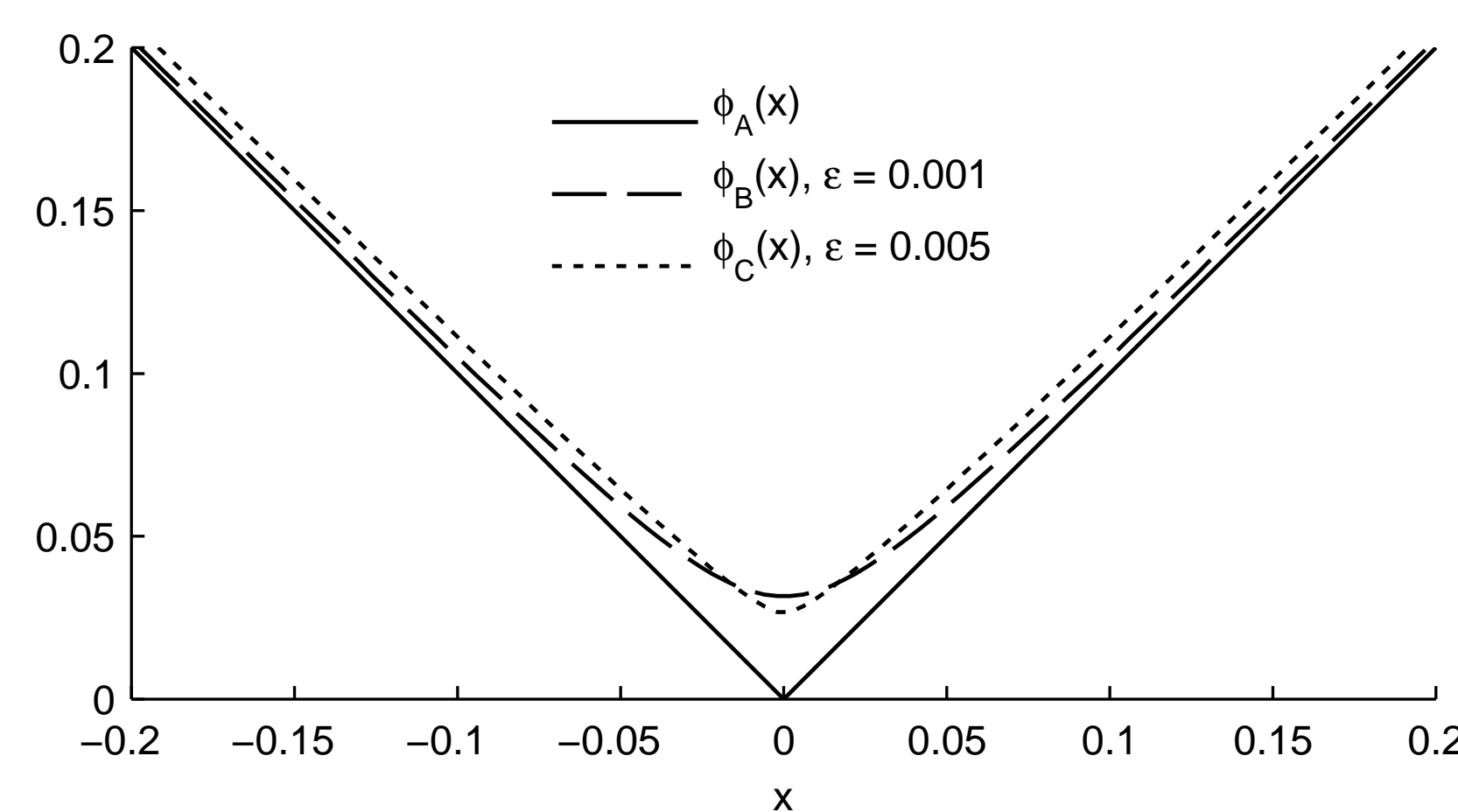
Une majorante de $F(\mathbf{p})$ [1] est $G(\mathbf{p}, \mathbf{v})$:

$$\frac{1}{2} \|\mathbf{H}(\mathbf{c} - \mathbf{p})\|_2^2 + \lambda_0 \mathbf{p}^T [\mathbf{T}(\mathbf{v})] \mathbf{p} + \lambda_0 \mathbf{b}^T \mathbf{p} + \sum_{i=1}^M \left[\frac{\lambda_i}{2} (\mathbf{D}_i \mathbf{p})^T [\mathbf{A}(\mathbf{D}_i \mathbf{v})] (\mathbf{D}_i \mathbf{p}) \right] + c(\mathbf{v}),$$

minimisée explicitement par :

$$\mathbf{p} = \left[\mathbf{H}^T \mathbf{H} + 2\lambda_0 \mathbf{T}(\mathbf{v}) + \sum_{i=1}^M \lambda_i \mathbf{D}_i^T [\mathbf{A}(\mathbf{D}_i \mathbf{v})] \mathbf{D}_i \right]^{-1} (\mathbf{H}^T \mathbf{H} \mathbf{c} - \lambda_0 \mathbf{b}).$$

Lissage et majoration



$$g_0(x, v) = \begin{cases} \frac{1+r}{4|v|} x^2 + \frac{1-r}{2} x + |v| \frac{1+r}{4}, & |v| > \epsilon, \\ \frac{1+r}{4\epsilon} x^2 + \frac{1-r}{2} x + \epsilon \frac{1+r}{4}, & |v| \leq \epsilon, \end{cases}$$

$$\sum_{n=0}^{N-1} g_0(x_n, v_n) = \mathbf{p}^T [\mathbf{T}(\mathbf{v})] \mathbf{p} + \mathbf{b}^T \mathbf{p} + c(\mathbf{v}) \geq \sum_{n=0}^{N-1} \theta_\epsilon(x_n; r).$$

Filtres

- Filtres à phase nulle, non-causaux, récurrents
 - Ordre $2d$, fréquence de coupure f_c
 - Produit de matrices en bande (solveurs rapides)
- $$\mathbf{H} = \mathbf{A}^{-1} \mathbf{B} \approx \mathbf{B} \mathbf{A}^{-1} \quad (\text{matrices finies})$$

Algorithme

Entrées : \mathbf{c} , $r \geq 1$, \mathbf{A} , \mathbf{B} , λ_i , $i = 0, \dots, M$

1. $[\mathbf{b}]_n = \frac{1-r}{2}$
 2. $\mathbf{d} = \mathbf{B}^T \mathbf{B} \mathbf{A}^{-1} \mathbf{c} - \lambda_0 \mathbf{A}^T \mathbf{b}$
 3. $\mathbf{p} = \mathbf{c}$ (Initialisation)
- Itérer
4. $[\mathbf{T}]_{n,n} = \begin{cases} \frac{1+r}{4|x_n|}, & |x_n| \geq \epsilon \\ \frac{1+r}{4\epsilon}, & |x_n| < \epsilon \end{cases}$
 5. $[\mathbf{A}]_{i,n,n} = \frac{\phi'([\mathbf{D}_i \mathbf{p}]_n)}{[\mathbf{D}_i \mathbf{p}]_n}$, $i = 0, \dots, M$,
 6. $\mathbf{M} = 2\lambda_0 \mathbf{T} + \sum_{i=1}^M \lambda_i \mathbf{D}_i^T \mathbf{A}_i \mathbf{D}_i$
 7. $\mathbf{Q} = \mathbf{B}^T \mathbf{B} + \mathbf{A}^T \mathbf{M} \mathbf{A}$
 8. $\mathbf{p} = \mathbf{A} \mathbf{Q}^{-1} \mathbf{d}$
- Jusqu'à convergence
9. $\mathbf{l} = \mathbf{c} - \mathbf{p} - \mathbf{B} \mathbf{A}^{-1} (\mathbf{c} - \mathbf{p})$
- Sorties : \mathbf{p} , \mathbf{l}

Ingrédients

$$[\mathbf{T}(\mathbf{v})]_{n,n} = \begin{cases} \frac{1+r}{4|v_n|}, & |v_n| \geq \epsilon, \\ \frac{1+r}{4\epsilon}, & |v_n| < \epsilon, \end{cases}$$

$$[\mathbf{b}]_n = \frac{1-r}{2},$$

$$[\mathbf{A}(\mathbf{v})]_{n,n} = \frac{\phi'(v_n)}{v_n},$$

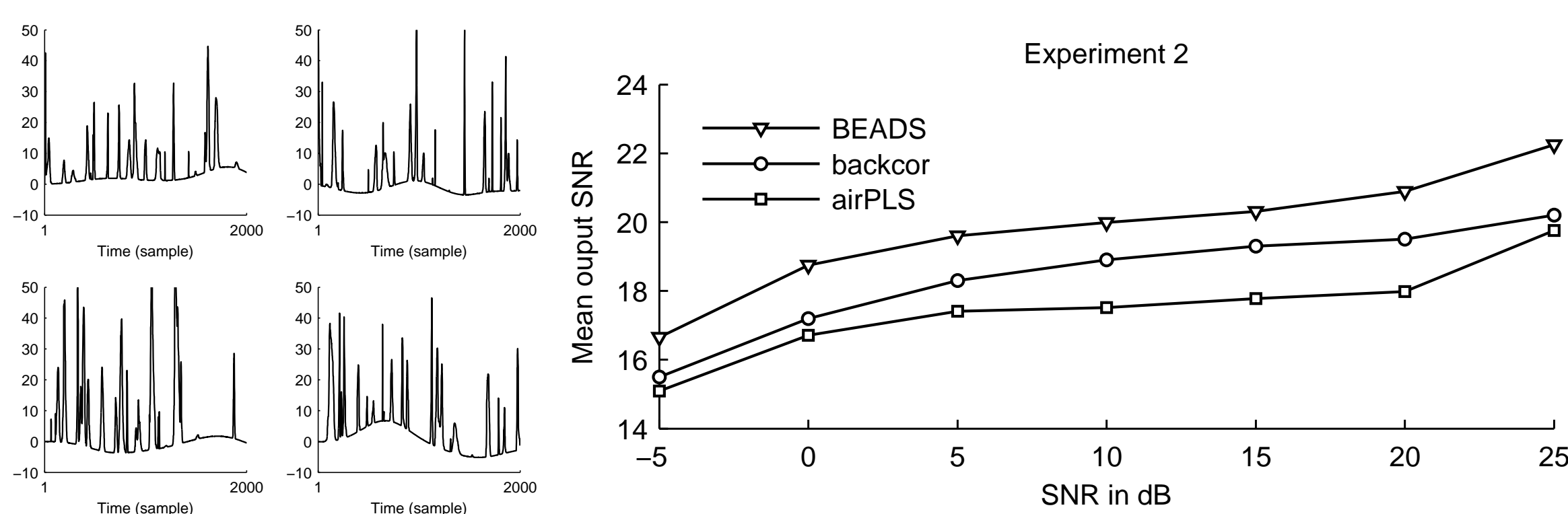
$$c(\mathbf{v}) = \sum_n \left[\phi(v_n) - \frac{v_n}{2} \phi'(v_n) \right].$$

Pénalités

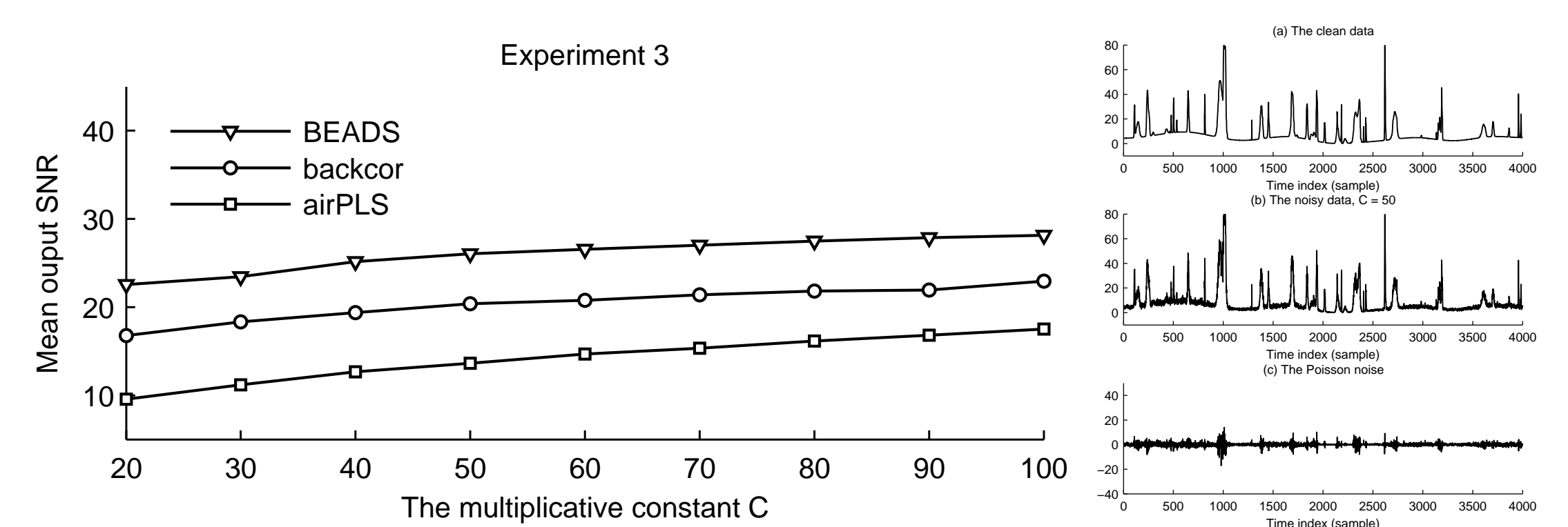
Pénalités promotrices de parcimonie :

- symétriques :
 - $\phi_A : |x|$ (non-dérivable),
 - $\phi_B : \sqrt{|x|^2 + \epsilon}$,
 - $\phi_C : |x| - \epsilon \log(|x| + \epsilon)$,
- asymétriques :
 - $\theta_\epsilon : \begin{cases} x, & x > \epsilon \\ \frac{1+r}{4\epsilon} x^2 + \frac{1-r}{2} x + \epsilon \frac{1+r}{4}, & |x| \leq \epsilon \\ -rx, & x < -\epsilon \end{cases}$

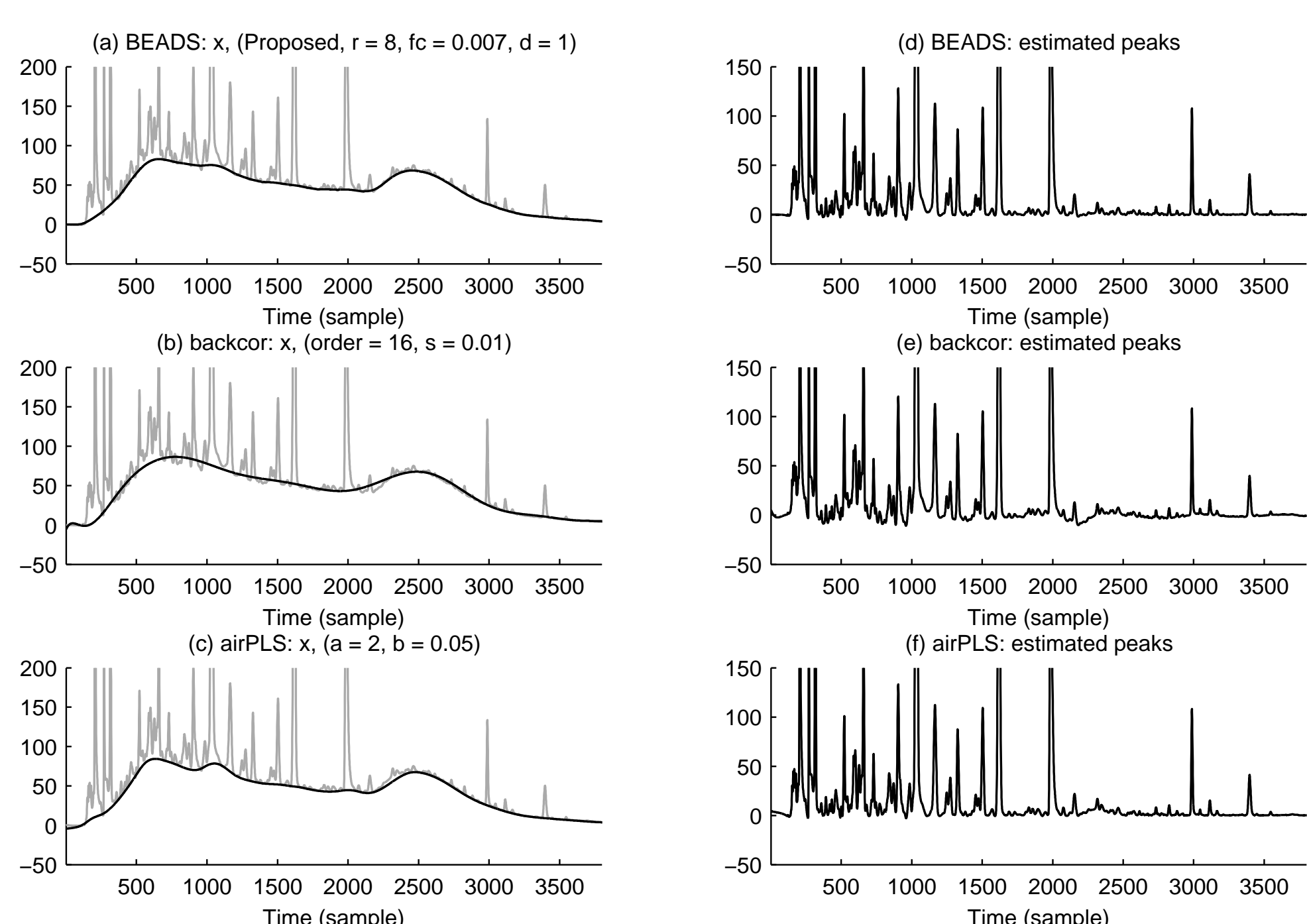
Évaluation comparative [1, 2, 3], cas gaussien



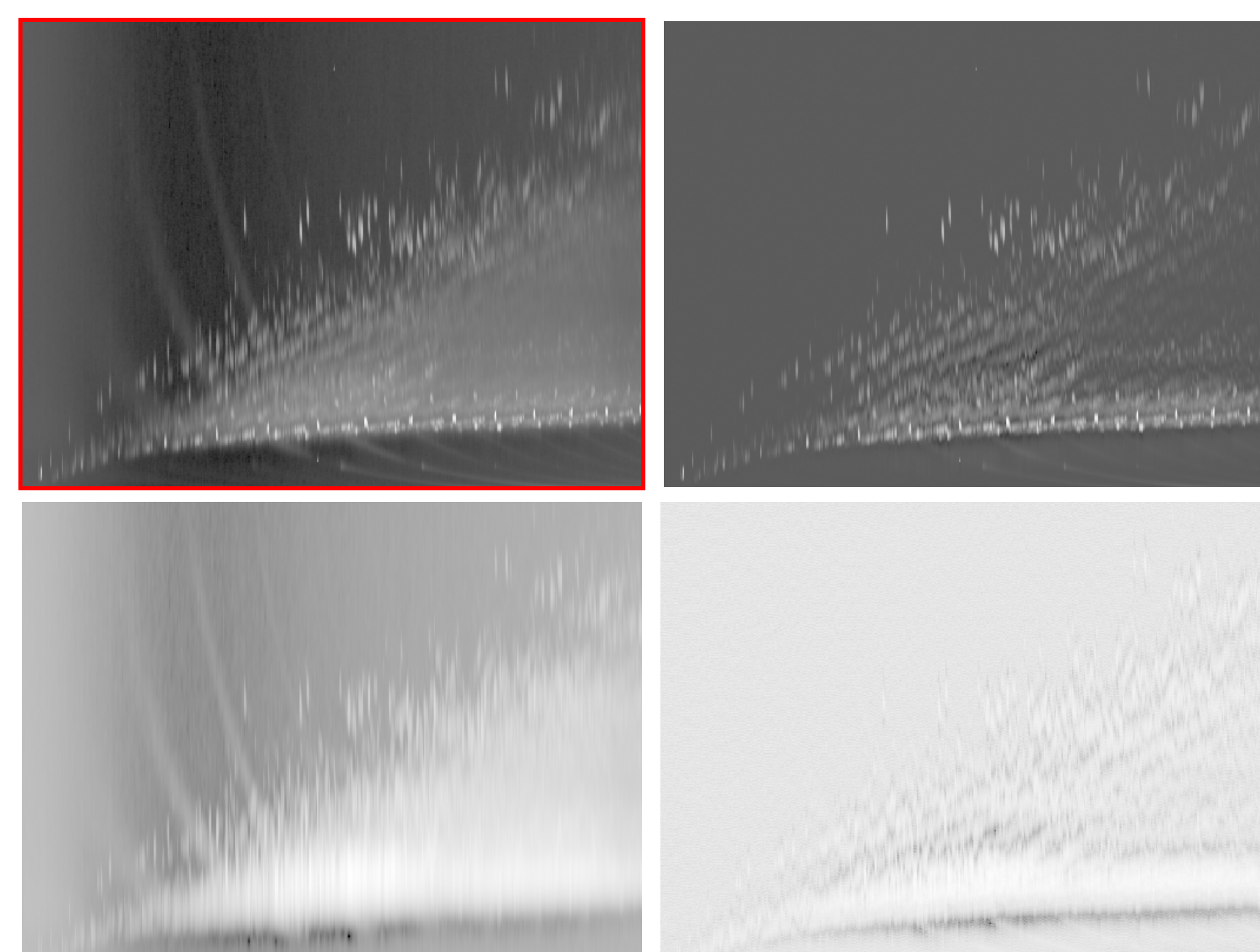
Évaluation comparative [1, 2, 3], cas Poisson



Chromatogrammes réels 1D



Chromatogrammes réels 2D [4]



Conclusions

- Bons résultats d'estimation conjointe, faible variance
- Faibles besoins en mémoire
- Couplage en déconvolution parcimonieuse en rapport ℓ_1/ℓ_2 [5]
- Extension en deux dimensions pour l'image
- Applications : chimie, biomédical (ECG, EEG)

Références

- [1] X. Ning, I. W. Selesnick, and L. Duval. Chromatogram baseline estimation and denoising using sparsity (BEADS). *Chemometr. Intell. Lab. Syst.*, 139:156–167, Dec. 2014.
- [2] V. Mazet, C. Carteret, D. Brie, J. Idier, and B. Humbert. Background removal from spectra by designing and minimising a non-quadratic cost function. *Chemometr. Intell. Lab. Syst.*, 76(2) :121–133, 2005.
- [3] Z.-M. Zhang, S. Chen, and Y.-Z. Liang. Baseline correction using adaptive iteratively reweighted penalized least squares. *Analyst*, 135(5) :1138–1146, 2010.
- [4] C. Vendevure, R. Ruiz-Guerrero, F. Bertoncini, L. Duval, and D. Thiébaud. Comprehensive two-dimensional gas chromatography for detailed characterisation of petroleum products. *Oil Gas Sci. Tech.*, 62(1) :43–55, 2007.
- [5] A. Repetti, M. Q. Pham, L. Duval, E. Chouzenoux, and J.-C. Pesquet. Euclid in a taxicab : Sparse blind deconvolution with smoothed ℓ_1/ℓ_2 regularization. *IEEE Signal Process. Lett.*, 22(5) :539–543, May 2015.