

# Group Lasso Wasserstein sans grille

Paul CATALA<sup>1</sup>, Vincent DUVAL<sup>2,3</sup>, Gabriel PEYRÉ<sup>1</sup>

<sup>1</sup>DMA, ENS, PSL, UMR 8553, CNRS,  
45 rue d'Ulm, 75005 Paris, France

<sup>2</sup>Mokaplan, INRIA Paris  
2 rue Simone Iff, 75012 Paris, France

<sup>3</sup>CEREMADE, Université Paris-Dauphine, PSL, UMR 7534, CNRS  
Place du Maréchal de Lattre de Tassigny, 75016 Paris, France

pcatala@dma.ens.fr, vincent.duval@inria.fr, gabriel.peyre@ens.fr

**Résumé** – On s'intéresse à la reconstruction simultanée de sources ponctuelles (positions et amplitudes) dans plusieurs canaux similaires, étant données des observations basse fréquences et bruitées (sur chaque canal). Le group Lasso régularise ce problème en imposant le même support parcimonieux aux solutions à travers les différents canaux, ce qui est souvent trop restrictif. Une relaxation de cette contrainte a récemment été proposée par [6], contraignant les supports à être proches au sens d'une distance de transport optimal. Nous étendons cette approche en proposant une méthode sans discrétisation, qui permet de retrouver plus précisément les positions sans utiliser de grille. Cette technique utilise la hiérarchie de Lasserre afin de remplacer la contrainte de positivité du problème dual par une optimisation convexe matricielle sous contrainte semi-définie.

**Abstract** – We consider in this paper the problem of simultaneously recovering pointwise sources across several similar tasks, given some low-pass measurements. The group Lasso regularize this problem by enforcing a common sparse support to the solutions in each task, which is often too restrictive in real applications. A relaxation of this constraint has recently been proposed in [6], which constrain the supports to be close with respect to an optimal transport distance. In this paper, we extend this approach by introducing an off-the-grid method, which allows to recover accurately the positions without resorting to some discretizing grid. This technique uses the Lasserre hierarchy in order to replace the positivity constraint of the dual problem by a convex semidefinite program.

## 1 Introduction

Les problèmes de super-résolution peuvent se poser ainsi: étant donné des observations basse fréquence d'un signal (obtenues typiquement par convolution avec un filtre passe-bas), comment reconstruire exactement le signal original, *i.e.* avec une résolution meilleure que celle de l'instrument de mesure, limitée par des phénomènes comme la diffraction ou le bruit photonique.

Bien que ce problème soit mal posé dans le cas général, il devient néanmoins soluble lorsque le signal original est parcimonieux, c'est-à-dire composé d'un petit nombre de sources ponctuelles. Les problèmes de super-résolution parcimonieuse ont de nombreuses applications dans plusieurs domaines de l'imagerie, par exemple en microscopie par fluorescence [5].

**Super-résolution sur les mesures de Radon.** Formellement, on modélise les sources ponctuelles par des Diracs  $a\delta_x$ , de position  $x \in \mathbb{T}^d$  et d'amplitude  $a \in \mathbb{R}$ . On note  $\mathcal{M}(\mathbb{T}^d)$  l'ensemble des mesures de Radon sur le tore. On considère alors le problème inverse suivant: reconstruire la mesure discrète  $\mu_0 = \sum_{i=1}^r a_i \delta_{x_i} \in \mathcal{M}(\mathbb{T}^d)$  à partir des observations bruitées  $u = \mathcal{F}\mu + w \in \mathbb{C}^m$  où  $\mathcal{F}$  est un opérateur linéaire et  $w$  modélise le bruit d'acquisition. Si  $\mathcal{F}$  est invariant par transla-

tion (*i.e.* est une convolution), quitte à tronquer les trop hautes fréquences, sans perte de généralité, on peut supposer que  $\mathcal{F} : \mathcal{M}(\mathbb{T}^d) \rightarrow \mathbb{C}^m$  est la transformée de Fourier

$$(\mathcal{F}\mu)_k \stackrel{\text{def.}}{=} \int_{\mathbb{T}^d} e^{-2i\pi\langle k, x \rangle} d\mu(x), \quad \forall k \in \Omega_c \stackrel{\text{def.}}{=} \llbracket -f_c, f_c \rrbracket^d. \quad (1)$$

Cela correspond à un filtre passe-bas idéal, de fréquence de coupure  $f_c \in \mathbb{N}^*$ . Comme détaillé dans [2], le type de méthodes considérées se généralise par approximation spectrale à une large classe d'opérateurs.

**BLASSO** Une approche possible pour résoudre ce problème est la régularisation en variation totale, qui généralise en dimension infinie la régularisation  $\ell^1$ : on recherche  $\mu_0$  comme solution du problème suivant, appelé BLASSO [1],

$$\min_{\mu \in \mathcal{M}(\mathbb{T}^d)} \frac{1}{2\lambda} \|\mathcal{F}\mu - u\|^2 + |\mu|(\mathbb{T}^d) \quad (2)$$

où  $\lambda \in \mathbb{R}_+$  dépend du niveau de bruit et la variation totale d'une mesure est définie comme

$$|\mu|(\mathbb{T}^d) \stackrel{\text{def.}}{=} \sup \left\{ \int_{\mathbb{T}^d} \eta d\mu ; \eta \in \mathcal{C}(\mathbb{T}^d) \text{ et } \|\eta\|_\infty \leq 1 \right\}, \quad (3)$$

où  $\mathcal{C}(\mathbb{T}^d)$  est l'ensemble des fonctions continues sur  $\mathbb{T}^d$  à valeurs réelles.

Dans le cas multi-canal, on dispose de plusieurs observations, et l'on recherche plusieurs mesures (une par canal). Pour simplifier l'exposition, on considère deux observations ( $u = \Phi\mu_0 + w, v = \Phi\nu_0 + \varepsilon$ ) et on recherche deux mesures ( $\mu, \nu$ ). En pratique, les sources à reconstruire sont liées à des phénomènes physiques ou biologiques, et l'on souhaite contraindre leurs positions à travers les canaux. En imagerie cérébrale par exemple, ces sources peuvent être liées à une position dans le cerveau [11], ou bien en génomique, une position dans le génome [10]. Une approche populaire, le group-Lasso, impose que les supports soient les mêmes à l'aide d'une norme  $\ell^1 - \ell^2$ . La généralisation sans grille (sur des mesures) de cette idée correspond à utiliser dans (2), à la place de  $|\mu|(\mathbb{T}^d)$ , une variation totale vectorielle

$$\|(\mu, \nu)\|(\mathbb{T}^d) \stackrel{\text{def}}{=} \sup \left\{ \int \eta d\mu + \int \xi d\nu ; \forall x, \left\| \begin{pmatrix} \eta(x) \\ \xi(x) \end{pmatrix} \right\| \leq 1 \right\} \quad (4)$$

où  $\eta, \xi \in \mathcal{C}(\mathbb{T}^d)$ . Dans certains cas, cela conduit à de meilleurs résultats que plusieurs estimations séparées [4], voir Figure 1.

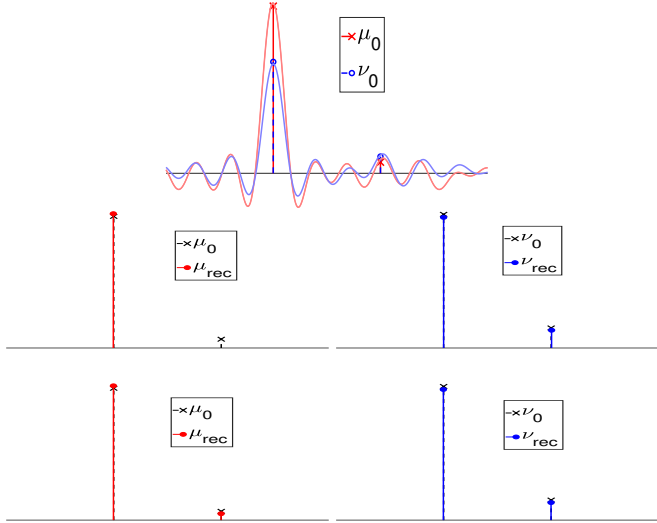


Figure 1 – Reconstruction jointe avec un group-Blasso (*bas*), et avec deux Blasso indépendants (*haut*), dans le cas de deux mesures ayant le même support. On voit que la méthode jointe parvient à reconstruire un spike de faible amplitude dissimulé dans le bruit, là où le Blasso seul échoue.  $\lambda = 1.5$ ,  $\|w\| = 0.15\|u_0\|$ , où  $u_0 = \mathcal{F}\mu_0$ .

**Contributions** Un inconvénient majeur de cette méthode est qu'elle impose un recoupement parfait entre les supports des différentes sources (voir Figure 2). Cette hypothèse est en pratique irréaliste: les enregistrements biologiques par exemple varient d'un sujet à l'autre. Dans [6], les auteurs relâchent cette contrainte en utilisant un Lasso (régularisation  $\ell^1$ ) pénalisé avec une distance de transport optimal entre les canaux. Nous proposons de généraliser cette idée au Blasso, c'est-à-dire sans recourir à une discrétisation, ce qui permet d'obtenir des résultats plus précis, sans artefact de discrétisation, et avec une com-

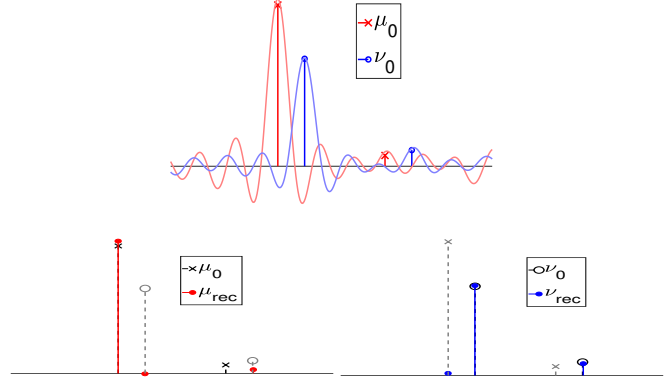


Figure 2 – Même contexte que précédemment, cette fois avec des mesures de supports disjoints. Les figures du bas montrent le résultat de la méthode jointe, moins performante ici que pour des mesures de même support.

plexité numérique indépendante de la taille de la grille.

## 2 Régularisation TV-Wasserstein

Dans toute la suite, on considère deux mesures discrètes *positives*  $\mu_0$  et  $\nu_0$  à super-résoudre étant donné les observations  $u = \mathcal{F}\mu_0 + w$  et  $v = \mathcal{F}\nu_0 + \varepsilon$ .

**Transport optimal** Le transport optimal permet de définir des similarités géométriques entre des mesures, qui tient compte des distances entre leur supports. Etant donné un coût  $c(x, y)$  sur  $\mathbb{T}^d \times \mathbb{T}^d$ , on définit la “métrique”  $W_c$  par

$$W_c(\mu, \nu) \stackrel{\text{def}}{=} \inf_{\gamma \in \Pi(\mu, \nu)} \langle c, \gamma \rangle \stackrel{\text{def}}{=} \int_{\mathbb{T}^d \times \mathbb{T}^d} c(x, y) d\gamma(x, y) \quad (5)$$

où  $\Pi(\mu, \nu)$  est l'ensemble des plans de transport entre  $\mu$  et  $\nu$ :

$$\Pi(\mu, \nu) \stackrel{\text{def}}{=} \left\{ \gamma \in \mathcal{M}_+(\mathbb{T}^d \times \mathbb{T}^d) ; \gamma_1 = \mu \text{ et } \gamma_2 = \nu \right\}, \quad (6)$$

avec  $\gamma_1$  et  $\gamma_2$  les deux marginales de  $\gamma$ , *i.e.*  $\gamma_1 = \gamma(\cdot, \mathbb{T}^d)$  et  $\gamma_2 = \gamma(\mathbb{T}^d, \cdot)$ . Si  $c$  est le carré d'une distance, alors  $W_c^{1/2}$  est la distance de Wasserstein-2. Pour les mesures produit, on notera  $\mathcal{F}_2$  l'opérateur de Fourier, c'est-à-dire, pour tout  $k_1, k_2 \in \Omega_c$ ,

$$(\mathcal{F}_2\gamma)_{k_1, k_2} \stackrel{\text{def}}{=} \int_{\mathbb{T}^d} e^{-2i\pi\langle k_1, x \rangle} e^{-2i\pi\langle k_2, y \rangle} d\gamma(x, y). \quad (7)$$

*Remark 1.* Pour généraliser à un nombre  $k > 2$  de sources  $\mu_1, \dots, \mu_k$ , on peut considérer comme dans [6] une pénalité de la forme  $\min_{\bar{\mu}} \sum_k W_c(\mu_k, \bar{\mu})$ .

**Blasso-Wasserstein** On considère le problème

$$\min_{\mu, \nu} \frac{\|\mathcal{F}\mu - u\|^2}{4\lambda} + \frac{\|\mathcal{F}\nu - v\|^2}{4\lambda} + \frac{\mu(\mathbb{T}^d)}{2} + \frac{\nu(\mathbb{T}^d)}{2} + \tau W_c(\mu, \nu), \quad (8)$$

sous la contrainte  $\mu, \nu \in \mathcal{M}_+(\mathbb{T}^d)$ . Les mesures solutions de ce problème ont nécessairement la même masse, autrement on

aurait  $W_c(\mu, \nu) = \infty$ . Il est possible de relâcher cette contrainte en considérant le modèle plus général du transport non équilibré [3], qui ne sera pas discuté ici. On se ramène ainsi, en utilisant (5), à un problème sur les mesures de l'espace produit  $\mathbb{T}^d \times \mathbb{T}^d$ :

$$\min_{\gamma} \frac{\|\mathcal{F}\gamma_1 - u\|^2}{4\lambda} + \frac{\|\mathcal{F}\gamma_2 - v\|^2}{4\lambda} + \gamma_0 + \tau \langle c, \gamma \rangle \quad (\mathcal{P}_\lambda)$$

où  $\gamma_0 = \gamma(\mathbb{T}^d \times \mathbb{T}^d)$ ,  $\gamma_1 = \gamma(\cdot, \mathbb{T}^d)$  et  $\gamma_2 = \gamma(\mathbb{T}^d, \cdot)$ .

**Problème dual et optimalité** Il est utile de considérer le problème dual de  $(\mathcal{P}_\lambda)$ , qui fait intervenir des éléments de  $\text{Im } \mathcal{F}^*$ , c'est-à-dire des polynômes trigonométriques. Etant donné deux polynômes trigonométriques  $d$ -variés  $P$  et  $Q$ , on note  $(P \oplus Q)$  le polynôme  $(2d)$ -varié défini par  $(P \oplus Q)(x, y) = P(x) + Q(y)$  pour tout  $(x, y) \in \mathbb{T}^d \times \mathbb{T}^d$ . On a alors le résultat suivant:

**Proposition 1.** *Le dual de  $(\mathcal{P}_\lambda)$  est donné par*

$$\begin{aligned} \max_{p, q \in \mathbb{C}^m} \langle u - \lambda e_1, p \rangle + \langle v - \lambda e_1, q \rangle - \frac{\lambda}{2} \|p\|^2 - \frac{\lambda}{2} \|q\|^2 \quad (\mathcal{D}_\lambda) \\ \text{s.t. } \tau(\mathcal{F}^*p \oplus \mathcal{F}^*q) \leq c \end{aligned}$$

où  $e_1 \stackrel{\text{def}}{=} (1, 0, \dots, 0)$ . De plus, les relations d'optimalité montrent que, pour toute paire de solutions  $(\gamma_\lambda, (p_\lambda, q_\lambda))$ ,

$$\text{Supp } \gamma_\lambda \subset \{(x, y) ; (c - \tau \mathcal{F}_2^*(p \oplus q))(x, y) = 0\},$$

où l'on a noté  $\mathcal{F}^*p \oplus \mathcal{F}^*q = \mathcal{F}_2^*(p_\lambda \oplus q_\lambda)$ .

### 3 Implémentation

Dans cette section, on suppose que le coût  $c$  dans la similarité  $W_c$  est un polynôme trigonométrique, i.e.  $c \in \text{Im } \mathcal{F}_2^*$ . De cette manière, bien que  $(\mathcal{P}_\lambda)$  soit de dimension infinie, il ne dépend en fait que d'un nombre fini de moments trigonométriques de ses variables. Il peut donc être reformulé en fonction d'un vecteur de dimension finie  $z$ , sous la contrainte que ce vecteur soit représenté par une mesure, c'est-à-dire coïncide avec les moments d'une mesure. En augmentant la dimension de ce vecteur, on obtient une hiérarchie de problèmes de dimensions finies, donnant une approximation arbitrairement précise de la solution exacte. Cette méthode est connue sous le nom de hiérarchies de Lasserre [8].

**Hiérarchie SDP** Soit  $\ell \geq f_c$ , et  $\Omega_\ell = \llbracket -\ell, \ell \rrbracket^d$ . Posons, pour  $(s, t) \in \Omega_\ell^2$ ,

$$z_{(s,t)} = \int_{\mathbb{T}^d \times \mathbb{T}^d} e^{-2i\pi \langle s, x \rangle} e^{-2i\pi \langle t, y \rangle} d\gamma(x, y), \quad (9)$$

le moment trigonométrique d'ordre  $(s, t)$  de  $\gamma$ . On note aussi  $z_1 \stackrel{\text{def}}{=} z_{(\cdot, 0)}$  et  $z_2 \stackrel{\text{def}}{=} z_{(0, \cdot)}$  les moments de  $\gamma_1$  et  $\gamma_2$  respectivement, et  $z_0 \stackrel{\text{def}}{=} z_{(0, 0)}$ . Soit  $R_\ell(z) = (z_{(s-s', t-t')})_{s, s', t, t' \in \llbracket 0, \ell \rrbracket^d}$  la matrice de moment d'ordre  $\ell$  de  $\gamma$ . Ces matrices vérifient nécessairement  $R_\ell(z) \succeq 0$ . On considère le problème suivant:

$$\begin{aligned} \min_{z \in \mathbb{C}^{|\Omega_\ell|}} \frac{1}{4\lambda} \|z_1 - u\|^2 + \frac{1}{4\lambda} \|z_2 - v\|^2 + z_0 + \tau \langle \hat{c}, z \rangle \quad (\mathcal{P}_\lambda^{(\ell)}) \\ \text{s.t. } R_\ell(z) \succeq 0 \end{aligned}$$

où  $\hat{c}$  est le vecteur des coefficients du polynôme trigonométrique  $c$ , i.e.  $c = \mathcal{F}_2^* \hat{c}$ . On note  $z_\lambda$  la solution de ce problème. Les propositions suivantes sont des résultats classiques de la hiérarchie de Lasserre [9]. La première justifie que  $(\mathcal{P}_\lambda^{(\ell)})$  est bien une approximation de  $(\mathcal{P}_\lambda)$ .

**Proposition 2.**  $\min(\mathcal{P}_\lambda^{(\ell)}) \leq \min(\mathcal{P}_\lambda^{(\ell+1)}) \leq \min(\mathcal{P}_\lambda)$  et de plus  $\lim_{\ell \rightarrow +\infty} \min(\mathcal{P}_\lambda^{(\ell)}) = \min(\mathcal{P}_\lambda)$ .

Le second résultat permet de lier les minimiseurs des deux problèmes, dans le cas où  $\min(\mathcal{P}_\lambda^{(\ell)}) = \min(\mathcal{P}_\lambda)$  (on parle d'*effondrement* de la hiérarchie).

**Proposition 3.** *Si  $\min(\mathcal{P}_\lambda^{(\ell)}) = \min(\mathcal{P}_\lambda)$  pour un certain  $\ell$ , alors  $z_\lambda$  est un vecteur de moments de  $\gamma_\lambda$ , i.e.  $z_\lambda = \mathcal{F}_2 \gamma_\lambda$ .*

En particulier, si  $\gamma_\lambda$  est une mesure discrète,  $R_\ell(z_\lambda)$  étant une matrice de moment de  $\gamma_\lambda$ , elle vérifie alors  $\text{rang } R_\ell(z_\lambda) \leq \text{Card}(\text{Supp } \gamma_\lambda)$ . Un algorithme exploitant cette propriété de faible rang a été proposé dans [2], que nous souhaiterions appliquer à ce problème dans de futurs travaux.

**Extraction du support** Etant donnée la matrice de moments  $R_\ell(z)$ , il est possible d'en déduire la mesure sous-jacente par une méthode algébrique. Dans nos tests numériques, nous utilisons l'implémentation proposée dans [7].

### 4 Résultats numériques

Nous présentons dans cette section quelques résultats numériques préliminaires. Dans nos simulations, nous utilisons  $c(x, y) = \sin^2(x - y)$  comme coût dans  $W_c$ . Ce choix correspond à une approximation simple du coût quadratique standard  $\|x - y\|^2$  par un polynôme trigonométrique de faible degré. Le problème  $(\mathcal{P}_\lambda^{(\ell)})$  est résolu pour  $\ell = f_c$ , et à l'aide du solveur SDPT3 [12]. Le support du couplage  $\gamma_\lambda$  est ensuite récupéré en utilisant la méthode algébrique mentionnée dans la section précédente, et les amplitudes correspondantes s'obtiennent en résolvant le système linéaire  $z_\lambda = \mathcal{F}_{2, x_\lambda} a$ , où

$$\mathcal{F}_{2, x} = \left( e^{-2i\pi \langle k, x_i \rangle} \right)_{k \in \Omega_c, i=1, \dots, r}, \quad x = (x_1, \dots, x_r),$$

et  $x_\lambda$  est le support reconstruit. Enfin, les deux mesures  $\mu_\lambda$  et  $\nu_\lambda$  s'obtiennent immédiatement par projection.

Les Figures 3 et 4 montrent l'influence du paramètre  $\tau$  sur la reconstruction, dans un régime sans bruit (Figure 3) et de faible bruit (Figure 4). Pour de faibles valeurs de  $\tau$ , le Blasso-Wasserstein donne des résultats similaires à deux Blasso indépendants. Pour de larges valeurs de  $\tau$ , les deux mesures solutions se confondent en une mesure relativement proche des deux mesures initiales. Le régime intermédiaire voit apparaître plusieurs Diracs parasites de faible amplitude.

La Figure 5 montre le résultat du Blasso-Wasserstein sur l'exemple présenté en Section 1: on voit notamment que la méthode parvient à reconstruire plus fidèlement le support de  $\mu_0$  là où le group Lasso sans grille échouait. En revanche, des erreurs sont commises sur le support de  $\nu_0$ .

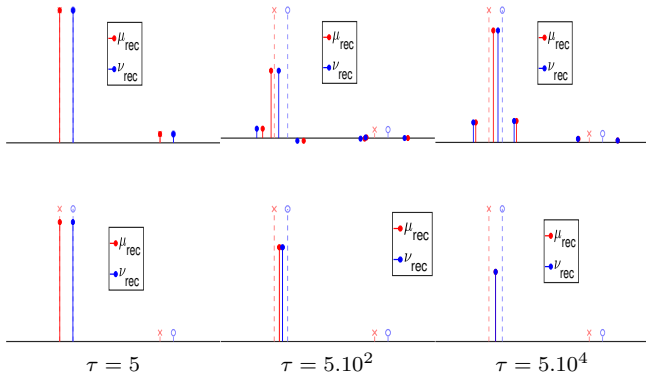


Figure 3 – Exemple de reconstruction sans bruit d’une mesure  $\mu_0$  et de sa translatée  $\nu_0$  (lignes pointillées), pour  $\lambda = 10^{-2}$  (haut) et  $\lambda = 1$  (bas), et pour des valeurs croissantes de  $\tau$ . Lorsque  $\tau$  est faible, la méthode se comporte comme deux Blasso indépendants.

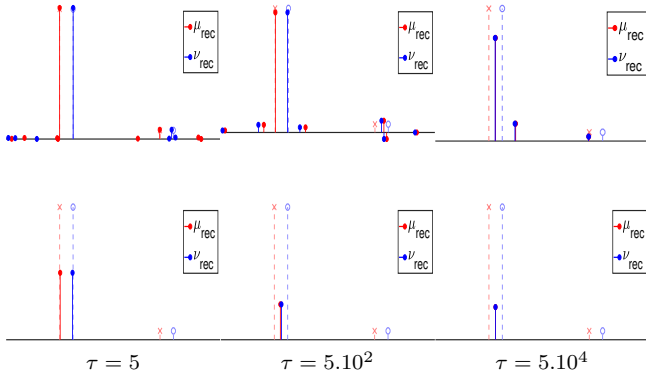


Figure 4 – Même exemple que ci-dessus, cette fois avec bruit, pour  $\lambda = 5.10^{-2}$  (haut) et  $\lambda = 5$  (bas). Le niveau de bruit est pris comme  $\|w\| = 0.05\|u_0\|$ , où  $u_0$  est l’observation non bruitée  $u_0 = \mathcal{F}\mu_0$ .

## 5 Conclusion et perspectives

Nous avons présenté dans cet article une extension aux méthodes sans grille de la régularisation Wasserstein proposée dans [6], qui permet de d’exploiter de manière flexible les similarités entre signaux pour des problèmes multi-canaux. Il serait intéressant pour de futurs travaux d’étudier l’impact du coût choisi dans  $W_c$  sur les performances de reconstruction. D’autre part, la méthode présentée pourrait s’étendre à une régularisation en transport optimal non équilibré (ce qui est l’idée initiale introduite dans [6]), ainsi qu’à du transport multi-marges, afin de gérer plus que deux canaux à la fois. Enfin, nous souhaiterions appliquer l’algorithme rapide introduit dans [2] pour résoudre  $(\mathcal{P}_\lambda^{(\ell)})$  efficacement, les solveurs SDP usuels étant peu efficaces voire inutilisables pour des matrices de grande taille.

**Remerciements** Ce projet est soutenu par une allocation de recherche Région Ile-de-France.

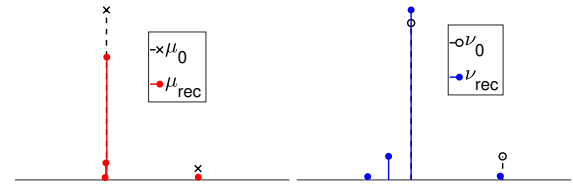


Figure 5 – Reconstruction de  $\mu_0$  (gauche) et  $\nu_0$  (droite) sur l’exemple utilisé en Fig. 2.  $\lambda = 1.3$ ,  $\tau = 10$ ,  $\|w\| = 0.15\|u_0\|$ .

## References

- [1] J. M. AZAÏS, Y. de CASTRO et F. GAMBOA : Spike detection from inaccurate sampling. *Applied and Computational Harmonic Analysis*, 38(2):177–195, 2015.
- [2] P. CATALA, V. DUVAL et G. PEYRÉ : A low-rank approach to sparse spikes deconvolution. *arXiv:1712.08800*, 2017.
- [3] L. CHIZAT, G. PEYRÉ, B. SCHMITZER et F.-X. VIALARD : Unbalanced optimal transport: Dynamic and kantorovich formulations. *Journal of Functional Analysis*, 274(11):3090 – 3123, 2018.
- [4] Carlos FERNANDEZ-GRANDA : Super-resolution of point sources via convex programming. *2015 IEEE 6th International Workshop CAMSAP*, pages 41–44, 2015.
- [5] S.T. HESS, T.P.K GIRIRAJAN et M.D. MASON : Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophysical Journal*, 91(11): 4258–4272, 2006.
- [6] H. JANATI, M. CUTURI et A. GRAMFORT : Wasserstein regularization for sparse multi-task regression. *ArXiv e-prints*, 2018.
- [7] C. JOSZ, J.B. LASSERRE et B. MOURRAIN : Sparse polynomial interpolation: Compressed sensing, super resolution, or prony? *arXiv:1708.06187*, 2017.
- [8] J. B. LASSERRE : Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.
- [9] J. B. LASSERRE : *Moments, positive polynomials and their applications*. Imperial College Press Optimization Series. Imperial College Press, London, 2010.
- [10] J. LAURENT, G. OBOZINSKI et J.-P. VERT : Group-lasso with overlap and graph lasso. *ICML*, pages 43–440, 2009.
- [11] J. OWEN, H.T. ATTIAS, K. SEKIHARA, S.S. NAGARAJAN et D.P. WIPF : Estimating the location and orientation of complex, correlated neural activity using MEG. *NIPS*, pages 1777–1784, 2009.
- [12] K.C. TOH, M.J. TODD et R.H. TUTUNCU : SDPT3 - a Matlab software package for semidefinite programming. *Optimization Methods and Software*, 11:545 – 581, 1999.