

Décomposition parcimonieuse multicanal basée sur le transport optimal

Jean MALLÉJAC¹ Jérôme IDIER² Charles SOUSSEN¹

¹Université Paris-Saclay, CNRS, CentraleSupélec, L2S, 91192 Gif-sur-Yvette

²CNRS, École Centrale Nantes, Nantes Université, LS2N, UMR 6004, 44000 Nantes

Résumé – La décomposition parcimonieuse multicanal consiste à représenter une séquence de signaux dans un dictionnaire rassemblant des formes élémentaires. Afin de favoriser des décompositions proches entre canaux voisins, les approches traditionnelles utilisent les normes mixtes. Cependant, elles tendent à imposer un support commun aux représentations parcimonieuses. Nous proposons une approche plus flexible basée sur le transport optimal. La régularisation inter-canal proposée prend en compte la continuité de la « trajectoire » décrite par les atomes du dictionnaire sélectionnés à travers les canaux.

Abstract – Multichannel sparse decomposition consists in representing a sequence of signals in a dictionary gathering elementary forms. In order to promote close decompositions between neighboring channels, traditional approaches use mixed norms. However, the latter tend to impose a common support on all sparse representations. We propose a more flexible approach based on optimal transport. The proposed inter-channel regularization takes into account the continuity of the “trajectory” described by selected dictionary atoms across channels.

1 Introduction

On s’intéresse à l’analyse d’une séquence spatiale ou temporelle de signaux. Ces données multicanaux (où un canal est associé à une position spatiale ou temporelle donnée) apparaissent dans de nombreux contextes applicatifs allant de la spectroscopie optique et l’imagerie hyperspectrale [14, 3] à la sismologie [15]. La décomposition parcimonieuse multicanal vise à décomposer chaque signal en un mélange linéaire de composantes élémentaires, et à suivre ces composantes à travers les canaux.

La décomposition parcimonieuse est typiquement formalisée par le biais d’un dictionnaire. Ce dernier rassemble un ensemble de signaux élémentaires (appelés atomes), possible-ment présents dans les signaux observés. Ainsi, chaque signal est représenté comme une combinaison linéaire d’un faible nombre d’atomes. Ce problème a été largement abordé dans la littérature. Les méthodes les plus classiques reposent sur les techniques d’optimisation ℓ_0 et ℓ_1 [16].

Dans le cas de la décomposition parcimonieuse multicanal, il est naturel d’utiliser une régularisation couplant les décompositions entre canaux voisins. Pour cela, les régularisations basées sur les normes mixtes ont été abondamment exploitées [2]. La norme $\ell_{2,1}$ permet de modéliser la parcimonie simultanée, en imposant de sélectionner un même atome sur tous les canaux. Des approches généralisées s’appuient sur l’activation d’un groupe d’atomes entre canaux consécutifs, *via* la définition de voisinages locaux. Ce modèle s’avère plus flexible, un même atome n’étant plus activé pour la totalité des canaux [12]. Au-delà du choix délicat de la structure de voisinage, la parcimonie par groupe modélise l’activation d’un groupe d’atomes, par opposition à l’approche de type transport optimal (TO) proposée, qui modélise le déplacement des atomes sélectionnés à travers les canaux.

L’objectif de l’article est de proposer une régularisation du problème de décomposition multicanal avec des pénalités

basées sur le TO, permettant de modéliser le déplacement continu d’atomes entre canaux voisins. La théorie du transport optimal permet de définir des distances entre distributions de probabilité. En particulier, la distance de Wasserstein permet de quantifier le transport « horizontal » correspondant à une translation du support d’une distribution. L’idée sous-jacente à notre approche est d’appliquer la distance de Wasserstein aux représentations parcimonieuses pour deux canaux consécutifs, afin de modéliser le déplacement des atomes. On peut noter que des idées proches ont été introduites pour la localisation de sources acoustiques [10] et la régression multi-tâches en imagerie cérébrale [11].

Dans cet article, on se restreint à des signaux à valeurs positives, qui peuvent être traités comme des distributions de probabilités (modulo une normalisation). Ce n’est pas le cas de signaux dont le signe est arbitraire, cependant, une représentation dans le domaine fréquentiel rend possible l’utilisation des distances de type TO, appliquée aux densités spectrales de puissance de deux signaux [5].

Après quelques rappels sur la théorie du transport optimal (Section 2), la méthode proposée est présentée en Section 3, où deux formes alternatives de régularisation basées TO sont développées. Ces régularisations rendent possible la reconstruction de nombreuses trajectoires continues d’atomes entre les canaux. Pour limiter leur nombre, nous proposons d’ajouter une contrainte de parcimonie. La Section 5 compare l’approche proposée avec plusieurs méthodes concurrentes.

2 Pré-requis sur le transport optimal

Notations : On note $\mathcal{M}_+^1(\mathbb{R})$ l’ensemble des distributions de probabilité univariées, notées ν . Pour ν_1 et $\nu_2 \in \mathcal{M}_+^1(\mathbb{R})$, on note $\Pi(\nu_1, \nu_2)$ l’ensemble des distributions bivariées γ dont les distributions marginales coïncident avec ν_1 et ν_2 . Lorsque ν_1 et ν_2 sont des distributions discrètes (c’est-à-dire une somme

de K_1 et K_2 distributions de Dirac), γ est représentée par une matrice de couplage $P \in \mathbb{R}_+^{K_1 \times K_2}$.

Par extension, on définit $\Pi(\nu_1, \dots, \nu_d)$ l'ensemble des distributions de probabilités multivariées dans $\mathbb{R}^d$, dont les marginales coïncident avec $\nu_1, \dots, \nu_d$. Pour des distributions discrètes, on note $\mathbb{P} \in \mathbb{R}_+^{K_1 \times \dots \times K_d}$ le tenseur de couplage.

GMM_d désigne l'ensemble des mélanges d'un nombre fini (quelconque) de gaussiennes dans $\mathbb{R}^d$, avec des pondérations dont la somme vaut 1. Enfin, on note $\langle \mathbb{A}, \mathbb{B} \rangle_F$ le produit de Frobenius entre deux matrices ou tenseurs de même dimension.

2.1 Transport entre deux distributions

Soit ν_1 et ν_2 deux distributions univariées. La distance de Wasserstein $W_2(\nu_1, \nu_2)$ est une mesure classique de transport. Elle est définie par :

$$W_2^2(\nu_1, \nu_2) := \inf_{\gamma \in \Pi(\nu_1, \nu_2)} \int_{\mathbb{R}^2} c(y_1, y_2) d\gamma(y_1, y_2),$$

avec la fonction de coût $c(y_1, y_2) = (y_1 - y_2)^2$. Dans le cas de distributions gaussiennes $\nu_1 \in \mathcal{N}(\mu_1, \sigma_1^2)$, $\nu_2 \in \mathcal{N}(\mu_2, \sigma_2^2)$, $W_2(\nu_1, \nu_2)$ prend une forme particulièrement simple :

$$W_2^2(\nu_1, \nu_2) = (\mu_1 - \mu_2)^2 + (\sigma_1 - \sigma_2)^2. \quad (1)$$

Considérons maintenant deux mélanges de gaussiennes (GMMs, pour *Gaussian mixture models*) $\nu_1 = \sum_{i=1}^{K_1} x_1^i \nu_1^i$ et $\nu_2 = \sum_{i=1}^{K_2} x_2^i \nu_2^i$ avec x_1, x_2 des vecteurs de poids normalisés. Malheureusement, il n'existe pas de méthode efficace pour calculer $W_2(\nu_1, \nu_2)$ de façon exacte. Par ailleurs, le plan de transport optimal γ n'est pas nécessairement un mélange de gaussiennes bivariées. Une distance entre GMMs alternative à W_2 a été proposée dans [8, 9] en ajoutant la contrainte que le plan de transport γ est un GMM :

$$MMW_2^2(\nu_1, \nu_2) := \inf_{\gamma \in \Pi(\nu_1, \nu_2) \cap \text{GMM}_2} \int_{\mathbb{R}^2} c(y_1, y_2) d\gamma(y_1, y_2).$$

De façon remarquable, $MMW_2(\nu_1, \nu_2)$ peut être calculé très efficacement lorsque le nombre de gaussiennes est limité [9]. Le calcul se ramène à la résolution d'un problème de transport optimal dans le cas discret. Il s'agit d'un problème de programmation linéaire en dimension $K_1 K_2$:

$$MMW_2^2(\nu_1, \nu_2) = \min_{P \in \Pi(x_1, x_2)} \langle P, C \rangle_F \quad (2)$$

où les coûts de transport $C_{i,j} = W_2^2(\nu_1^i, \nu_2^j)$ se calculent en appliquant l'expression analytique (1).

2.2 Transport multi-marginales

On considère maintenant d mesures de probabilités univariées $\nu = (\nu_1, \dots, \nu_d)$. Le transport multi-marginales consiste à trouver une loi jointe multivariée minimisant le critère :

$$W_2^2(\nu) := \inf_{\gamma \in \Pi(\nu)} \int_{\mathbb{R}^d} c(y_1, \dots, y_d) d\gamma(y_1, \dots, y_d), \quad (3)$$

où c est une fonction de coût arbitraire. Pour $c(y_1, \dots, y_d) = \frac{1}{2} \sum_{i,j=1}^d \lambda_i \lambda_j (y_i - y_j)^2$ avec $\lambda_1, \dots, \lambda_d$ des poids positifs et de somme 1, (3) est une reformulation du problème de barycentre de Wasserstein entre $\nu_1, \dots, \nu_d$, défini par

$\min_{\nu} \sum_{i=1}^d \lambda_i W_2^2(\nu_i, \nu)$. Le barycentre entre des GMMs n'étant pas nécessairement un GMM, une nouvelle formulation a été introduite dans [9] par analogie avec MMW_2 :

$$MMW_2^2(\nu) := \inf_{\gamma \in \Pi(\nu) \cap \text{GMM}_d} \int_{\mathbb{R}^d} c(y_1, \dots, y_d) d\gamma(y_1, \dots, y_d).$$

Le calcul de MMW_2^2 se réduit à résoudre le problème de transport multi-marginales discret :

$$MMW_2^2(\nu) = \min_{\mathbb{P} \in \Pi(x_1, \dots, x_d)} \langle \mathbb{P}, \mathbb{C} \rangle_F, \quad (4)$$

avec $\mathbb{C}$ le tenseur de coût d'ordre d défini par :

$$\forall i_1, \dots, i_d, \quad \mathbb{C}_{i_1, \dots, i_d} = W^2(\nu_1^{i_1}, \dots, \nu_d^{i_d}).$$

3 Décomposition multicanal

3.1 Problème direct

Dans la suite, on note $(z_n)_{n \in [1, N]}$ les données multicanal. Le signal z_n associé au canal n , est modélisé comme un mélange bruité de composantes élémentaires (voir Figure 1) :

$$z_n = Ax_n + b_n, \quad n \in \{1, \dots, N\},$$

où le dictionnaire $A = [a_1, \dots, a_M] \in \mathbb{R}^{T \times M}$ contient M atomes et b_n est un bruit blanc gaussien. La décomposition multicanal consiste à reconstruire $x_n \in \mathbb{R}^M$ à partir de z_n et A . Pour ce faire, on considère trois hypothèses de travail.

1) Le dictionnaire est gaussien *i.e.*, chaque atome a_m est une gaussienne univariée (pas nécessairement de variance constante) évaluée sur une grille régulière de T points, notée $\mathcal{t}$.

2) Les décompositions $x = (x_n)_{n \in [1, N]}$ sont positives et se somment à la même « masse » *i.e.*, appartiennent au simplexe :

$$\mathcal{X} = \{x \in \mathbb{R}_+^{MN} / \forall n \in [1, N], \|x_n\|_1 = \|x_1\|_1\}.$$

Cette hypothèse certes restrictive, permet de se placer dans le cadre du TO classique dit « équilibré ». Elle apparaît en démixage hyperspectral où l'on retrouve la contrainte de somme-à-un [1]. Le TO dit « non-équilibré » est une extension du TO classique qui permettrait de s'affranchir de cette contrainte [7].

3) Enfin, les signaux x_n sont supposés évoluer « lentement » en fonction du canal n .

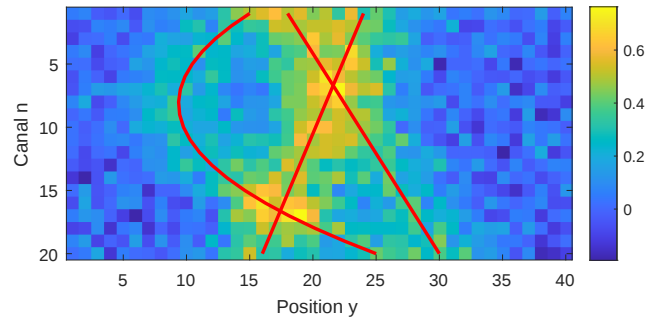


Figure 1 : Exemple d'une séquence de 20 mélanges gaussiens bruités à 10 dB. Les lignes correspondent aux signaux z_n . Les trajectoires des atomes sont affichées en rouge.

3.2 Régularisation par transport optimal

3.2.1 Approche variationnelle

Pour résoudre le problème de décomposition multicanal, nous proposons une approche variationnelle conjointe avec un terme de moindres carrés :

$$J(\mathbf{x}) = \sum_{n=1}^N \|z_n - A\mathbf{x}_n\|_2^2 \quad (5)$$

et une régularisation par transport optimal notée $R(\mathbf{x})$. Le problème de minimisation correspondant s'écrit :

$$\min_{\mathbf{x} \in \mathcal{X}} J(\mathbf{x}) + \lambda R(\mathbf{x}) \quad (6)$$

avec λ un hyperparamètre. Le signal reconstruit $A\mathbf{x}_n$ est un mélange d'atomes gaussiens identifiable à un GMM, modulo la discrétisation sur la grille $\underline{t}$. Notons $\mathcal{A}\mathbf{x}_n$ la combinaison linéaire des densités gaussiennes (non discrétisées).

3.2.2 TO entre paires de canaux consécutifs

Les deux mélanges successifs $\mathcal{A}\mathbf{x}_n$ et $\mathcal{A}\mathbf{x}_{n+1}$ étant supposés proches, on régularise de proche en proche *via* la distance de Wasserstein modifiée :

$$R_1(\mathbf{x}) = \sum_{n=1}^{N-1} MW_2^2(\mathcal{A}\mathbf{x}_n, \mathcal{A}\mathbf{x}_{n+1}).$$

Les poids $\mathbf{x}_n$ ne se somment pas à 1, mais l'hypothèse de conservation de la masse $\mathbf{x} \in \mathcal{X}$ permet de s'y ramener à une normalisation près indépendante de n , ce qui permet d'utiliser (2). Cette régularisation entre canaux consécutifs favorise les décompositions $\mathbf{x}_n$ et $\mathbf{x}_{n+1}$ ayant des supports proches sans pour autant forcer un support identique comme le ferait la parcimonie simultanée. Cependant, cette pénalisation « d'ordre 1 » tend à sélectionner des atomes identiques entre canaux consécutifs, puisque

$$R_1(\mathbf{x}) = 0 \Leftrightarrow \forall n, \mathcal{A}\mathbf{x}_n = \mathcal{A}\mathbf{x}_{n+1} \Leftrightarrow \forall n, \mathbf{x}_n = \mathbf{x}_{n+1}$$

(par identifiabilité des GMMs à une permutation près). Pour cette raison, nous construisons dans la suite une variante « d'ordre 2 ».

3.2.3 TO entre triplets de canaux consécutifs

L'objectif est ici de concevoir une pénalisation qui soit minimale pour les trajectoires rectilignes, obliques ou non. Cette nouvelle construction est basée sur un problème de transport optimal multi-marginales entre $d = 3$ mélanges consécutifs :

$$R_2(\mathbf{x}) = \sum_{n=1}^{N-2} MMW_2^2(\mathcal{A}\mathbf{x}_n, \mathcal{A}\mathbf{x}_{n+1}, \mathcal{A}\mathbf{x}_{n+2}).$$

Le coût de transport utilisé $c(y_1, y_2, y_3) = (y_1 - 2y_2 + y_3)^2$ est nul pour toute trajectoire localement rectiligne des atomes. Cependant, en pratique, l'optimisation du critère ainsi pénalisé conduit à des décompositions peu parcimonieuses, de nombreuses trajectoires étant approximativement équivalentes au sens de la pénalisation R_2 .

3.3 Régularisation parcimonieuse

Afin d'obtenir des décompositions plus parcimonieuses, une pénalisation ℓ_p est ajoutée au critère :

$$\min_{\mathbf{x} \in \mathcal{X}} J(\mathbf{x}) + \lambda R(\mathbf{x}) + \mu \|\mathbf{x}\|_p^p, \quad (7)$$

avec $\|\mathbf{x}\|_p^p = \sum_n \|\mathbf{x}_n\|_p^p$ et $0 < p \leq 1$. Choisir $p = 1$ a pour effet de contraindre la masse des $\mathbf{x}_n$ à une valeur faible. Cela ne crée pas l'effet de parcimonie attendu, du fait de la contrainte $\mathbf{x} \in \mathcal{X}$ déjà présente. Il est donc nécessaire de choisir $p < 1$ en pratique, au prix d'une perte de garantie de convexité.

4 Algorithme et implémentation

Le calcul des pénalisations R_1 et R_2 implique la résolution de $N - 1$ et $N - 2$ problèmes de programmation linéaire, cf. (2) et (4). Déplions ces fonctions coût. En notant $\mathbf{p}$ le vecteur qui rassemble les éléments des matrices P_n intervenant dans le calcul de $MW_2^2(\mathcal{A}\mathbf{x}_n, \mathcal{A}\mathbf{x}_{n+1})$ (respectivement, des tenseurs $\mathbb{P}_n$ liés au calcul de $MW_2^2(\mathcal{A}\mathbf{x}_n, \mathcal{A}\mathbf{x}_{n+1}, \mathcal{A}\mathbf{x}_{n+2})$), le critère (7) devient :

$$\min_{(\mathbf{x}, \mathbf{p}) \in \mathcal{C}} J(\mathbf{x}) + \lambda L(\mathbf{p}) + \mu \|\mathbf{x}\|_p^p, \quad (8)$$

avec $L(\mathbf{p}) = \sum_{n=1}^{N-1} \langle P_n, C \rangle_F$ ou $\sum_{n=1}^{N-2} \langle \mathbb{P}_n, \mathbb{C} \rangle_F$. Les contraintes linéaires $\mathcal{C}$ apparaissant dans (8) sont (i) $\mathbf{x} \in \mathcal{X}$, et (ii) $\forall n, P_n \in \Pi(\mathbf{x}_n, \mathbf{x}_{n+1})$ ou $\mathbb{P}_n \in \Pi(\mathbf{x}_n, \mathbf{x}_{n+1}, \mathbf{x}_{n+2})$.

Le critère (8) inclut deux termes convexes et un terme non-convexe. Pour l'optimiser, nous utilisons un schéma ℓ_1 repondéré (IR ℓ_1) [4]. Il s'agit d'un schéma de Majoration-Minimisation dans lequel le terme $\|\mathbf{x}\|_p^p$, concave sur $\mathbb{R}_+^{MN}$, est majoré par une fonction linéaire en $\mathbf{x}$ au point courant $(\mathbf{x}^{(k)}, \mathbf{p}^{(k)})$ à l'itération k . En notant $\phi(t) = t^p$, l'étape de majoration s'écrit pour tout $\mathbf{x} \geq \mathbf{0}$:

$$\|\mathbf{x}\|_p^p = \sum_{\ell=1}^{MN} \phi(x_\ell) \leq \sum_{\ell=1}^{MN} \left\{ \phi(x_\ell^{(k)}) + \phi'(x_\ell^{(k)}) (x_\ell - x_\ell^{(k)}) \right\}.$$

En incorporant cette majoration dans le critère (8), on obtient le problème de programmation quadratique :

$$\mathcal{Q}(\mathbf{x}, \mathbf{p}; \mathbf{w}^{(k)}) : \min_{(\mathbf{x}, \mathbf{p}) \in \mathcal{C}} J(\mathbf{x}) + \lambda L(\mathbf{p}) + \mu \mathbf{x}^T \mathbf{w}^{(k)}$$

avec les poids $w_\ell^{(k)} := \phi'(x_\ell^{(k)})$. Il en résulte un problème de programmation quadratique (convexe). Les solutions algorithmiques proposées dans [13] pour $\mu = 0$ s'adaptent aisément. L'ensemble du schéma itératif est résumé par l'algorithme 1, l'essentiel du coût de calcul étant lié aux étapes de minimisation de $\mathcal{Q}$. Ces dernières sont résolues par un algorithme standard de point intérieur, *via* le solver CPLEX.

Algorithme 1 : Schéma itératif IR ℓ_1

Entrées : $\mathbf{z}, A, \mathbf{x}^{(0)}$

Sorties : $\mathbf{x}^{(k)}, \mathbf{p}^{(k)}$

1 **Pour** ($k = 0$; $k = k + 1$; $k < K$)

2 Mise à jour des poids $\mathbf{w}^{(k)} = \phi'(\mathbf{x}^{(k)})$;

3 $(\mathbf{x}^{(k+1)}, \mathbf{p}^{(k+1)}) \leftarrow \arg \min_{\mathbf{x}, \mathbf{p}} \mathcal{Q}(\mathbf{x}, \mathbf{p}; \mathbf{w}^{(k)})$;

5 Résultats numériques

La méthode présentée ci-dessus est testée sur un problème de déconvolution gaussien. Les signaux z_n sont des convolutions bruitées $z_n = g * x_n + b_n$ où g est la réponse impulsionnelle gaussienne. Ainsi, les atomes du dictionnaire A correspondent à des versions translatées de g , donc de différentes moyennes μ_m mais de même écart-type ($\sigma_m = \sigma$). Les données sont présentées en Figure 1 pour un niveau de bruit élevé (10 dB).

La régularisation proposée est comparée avec une approche favorisant la parcimonie de groupe [17]. Dans cette méthode, la parcimonie des décompositions x est encodée par les groupes $\{(m, n), (m, n + 1)\}$ pour tout m, n . Cela revient à favoriser de proche en proche des décompositions ayant un même support. Les résultats obtenus sont également comparés à ceux d’une autre approche basée TO [10] dans laquelle l’évolution dynamique des décompositions est modélisée par une distribution de vitesses. Cette régularisation prend en compte la position et la vitesse des atomes en définissant le coût de transport *via* un problème de contrôle optimal.

Les décompositions obtenues sont présentées en Figure 2. La parcimonie de groupe (a) ne permet pas de déplacement latéral des atomes et apparaît comme une régularisation trop rigide. Le résultat (b) basé TO dynamique [10] permet de suivre les atomes à travers les canaux. En revanche, les décompositions obtenues sont peu parcimonieuses. La régularisation R_1 permet d’obtenir des décompositions parcimonieuses (c) qui attestent d’un suivi précis des trajectoires des atomes, malgré une difficulté lors d’un croisement entre deux trajectoires. En effet, ce croisement ne correspond pas à un déplacement de masse optimal au sens du TO. Le résultat (d) montre que la régularisation R_2 couplée avec la pénalisation ℓ_p (avec $p = 0.9$) remédie à ce problème et favorise des trajectoires rectilignes par morceaux. Cependant, des effets de grilles apparaissent sur certaines trajectoires (trop lentes, proches de la verticale) reconstruites avec cette seconde pénalisation. En considérant une grille plus fine, super-résolue, ils tendraient à disparaître, au prix d’un coût calculatoire plus élevé.

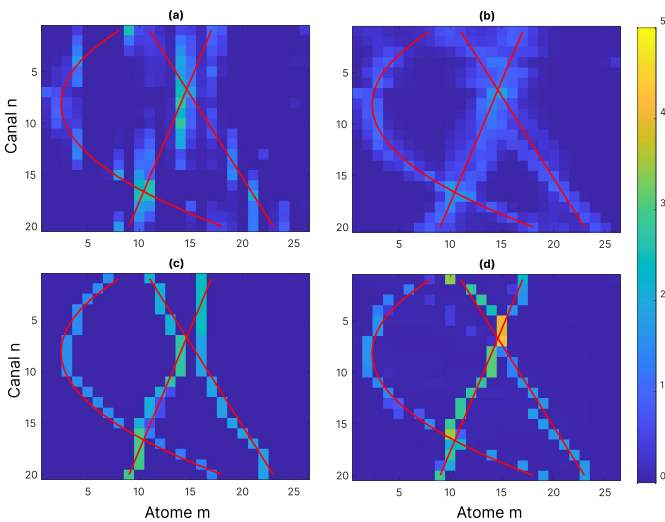


Figure 2 : Décomposition $(x_n)_{n \in [1, N]}$ du signal multicanal $(z_n)_{n \in [1, N]}$ dans un dictionnaire convolutif gaussien : (a) parcimonie de groupe [17], (b) TO dynamique [10], (c) et (d) méthode proposée, d’ordre 1 et 2.

6 Conclusion

Les régularisations de type TO proposées permettent un suivi efficace des trajectoires des atomes sélectionnés en minimisant une fonction de coût convexe. Cependant, elles ne favorisent intrinsèquement pas les décompositions parcimonieuses. En ajoutant une pénalisation ℓ_p légèrement non-convexe ($p = 0.9$), nous parvenons à dépasser cette difficulté au détriment d’une perte de garantie théorique sur l’absence de minima locaux. Par ailleurs, pour remédier à un coût calculatoire élevé, des astuces algorithmiques ont été proposées dans [13]. L’extension de la méthode à d’autres dictionnaires paramétriques (atomes non gaussiens) semble être envisageable [6]. L’utilisation de transport optimal non-équilibré [7] pourra être envisagée par la suite pour supprimer la contrainte de conservation de la masse et ainsi traiter des signaux plus variés.

Références

- [1] Y. ALTMANN, N. DOBIGEON, S. MCLAUGHLIN et J.-Y. TOURNERET : Démélange non linéaire d’images hyperspectrales à l’aide de fonctions radiales de base et de moindres carrés orthogonaux. *In Proc. GretsI*, 2011.
- [2] F. BACH, R. JENATTON, J. MAIRAL et G. OBOZINSKI : Optimization with sparsity-inducing penalties. *Found. Trends Mach. Learn.*, 4(1), 2012.
- [3] J. M. BIOUCAS-DIAS, A. J. PLAZA, N. DOBIGEON, M. PARENTE, Q. DU, P. D. GADER et J. CHANUSSOT : Hyperspectral unmixing overview : Geometrical, statistical, and sparse regression-based approaches. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, 5(2), 2012.
- [4] E. CANDÈS, M. WAKIN et S. BOYD : Enhancing sparsity by reweighted ℓ_1 minimization. *Journal of Fourier analysis and applications*, 14, 2008.
- [5] E. CAZELLES, A. ROBERT et F. TOBAR : The Wasserstein-Fourier distance for stationary time series. *IEEE Trans. Signal Process.*, 69, 2021.
- [6] A. CHAKRABORTY, A. BHATTACHARYA et D. PATI : Robust probabilistic inference via a constrained transport metric. Research report, 2023.
- [7] L. CHAPEL, R. FLAMARY, H. WU, C. FÉVOTTE et G. GASSO : Unbalanced optimal transport through non-negative penalized linear regression. *NeurIPS*, 34, 2021.
- [8] Y. CHEN, T. T. GEORGIU et A. TANNENBAUM : Optimal transport for Gaussian mixture models. *IEEE Access*, 7, 2019.
- [9] J. DELON et A. DESOLNEUX : A Wasserstein-type distance in the space of Gaussian mixture models. *SIAM J. Imaging Sciences*, 13(2), 2020.
- [10] F. ELVANDER, I. HAASLER, A. JAKOBSSON et J. KARLSSON : Multi-marginal optimal transport using partial information with applications in robust localization and sensor fusion. *Signal Process.*, 171, 2020.
- [11] H. JANATI, T. BAZEILLE, B. THIRION, M. CUTURI et A. GRAMFORT : Multi-subject MEG/EEG source imaging with sparse multi-task regression. *NeuroImage*, 220, 2020.
- [12] M. KOWALSKI, K. SIEDENBURG et M. DÖRFLER : Social sparsity ! Neighborhood systems enrich structured shrinkage operators. *IEEE Trans. Signal Process.*, 61(10), 2013.
- [13] J. MALLÉJAC, J. IDIER, C. SOUSSEN et X. LI : Joint decomposition of multichannel signals with multimarginal optimal transport. *In Proc. Eur. Sig. Proc. Conf.*, 2024.
- [14] H. MORTADA, V. MAZET, C. SOUSSEN, C. COLLET et L. POISSON : Parameterized source separation for delayed spectroscopic signals. *Signal Process.*, 158, 2019.
- [15] D. PEREG, I. COHEN et A. VASSILIOU : Multichannel sparse spike inversion. *J. Geophys. Eng.*, 14(5), 2017.
- [16] J. A. TROPP et S. J. WRIGHT : Computational methods for sparse solution of linear inverse problems. *Proc. IEEE*, 98(5), 2010.
- [17] Lei YUAN, Jun LIU et Jieping YE : Efficient methods for overlapping group lasso. *NeurIPS*, 24, 2011.