

Transport optimal de mesures en domaine fréquentiel

Laurent Condat

Univ. Grenoble Alpes, GIPSA-Lab, CNRS, 38000 Grenoble, France

laurent.condat@gipsa-lab.grenoble-inp.fr

Résumé – On définit de nouvelles distances convexes correspondant au transport optimal de mesures signées sur le cercle, exprimées en fonction d'un nombre fini de leurs coefficients de Fourier. Cela permet de manipuler des mesures numériquement, sans avoir à les discrétiser sur une grille de positions prédéfinies.

Abstract – We define new convex distances corresponding to optimal transport of signed measures on the circle, expressed in terms of a finite number of their Fourier coefficients. This makes it possible to deal with measures computationally, without the need to discretize them on a grid of predefined locations.

1 Introduction

Le transport optimal de mesures [1, 2] a connu ces dernières années un essor important, motivé par un large spectre d'applications, du traitement d'image [3] au traitement du langage [4].

Nous nous penchons ici sur le transport optimal de mesures 1-D signées définies sur le cercle $\mathbb{T} = \mathbb{R} \setminus \mathbb{Z}$ (identifié à l'intervalle $[0, 1[$), dont on ne connaît et manipule qu'un nombre fini $2M + 1$ de coefficients de Fourier, où $M \geq 1$ est un entier fixé. Définissons ainsi l'ensemble \mathcal{M} des mesures de Radon signées sur \mathbb{T} , ainsi que l'espace de Hilbert réel, de dimension $2M + 1$,

$$\mathbb{V} = \{v = (v_m)_{m=-M}^M \in \mathbb{C}^{2M+1} : (\forall m) v_{-m} = v_m^*\}, \quad (1)$$

où $*$ indique la conjugaison complexe, équipé du produit scalaire usuel $\langle v, v' \rangle = \sum_{m=-M}^M v_m v_m'^* \in \mathbb{R}$. Étant donnée $\mu \in \mathcal{M}$, ses premiers coefficients de Fourier, ou moments trigonométriques, forment la séquence $\mathcal{F}\mu = (\hat{\mu}_m)_{m=-M}^M \in \mathbb{V}$, où

$$\hat{\mu}_m = \int_0^1 e^{-j2\pi f m} d\mu(f). \quad (2)$$

Ainsi, \mathbb{V} est l'image de \mathcal{M} par l'opérateur \mathcal{F} . Naturellement, deux mesures différentes μ et μ' peuvent satisfaire $\mathcal{F}\mu = \mathcal{F}\mu'$, car $\mathcal{F}\mu$ n'encode qu'une information partielle sur μ . Retrouver μ , supposée par exemple discrète, à partir de $\mathcal{F}\mu$ est le problème classique des moments tronqués [5]. On peut aussi parler de *super-resolution* [6, 7] : on retrouve l'information hautes fréquences de μ , n'en connaissant que l'information basses fréquences. Les problèmes de minimisation de la variation totale de mesures sont parfois appelés *Beurling lasso* [8]. Sur les questions d'existence, unicité, et cardinalité de mesures discrètes, ne connaissant que leurs premiers coefficients de Fourier, voir [9].

On indique par $\|\mu\|_{\text{TV}}$ la norme de variation totale de $\mu \in \mathcal{M}$. Notons que l'on peut trouver une mesure μ de variation totale minimale, et qui est une mesure discrète concentrée en au plus $2M$ points de \mathbb{T} , à partir de $\mathcal{F}\mu$, en résolvant un problème d'optimisation convexe de dimension finie (voir eqs. (4),(5)),

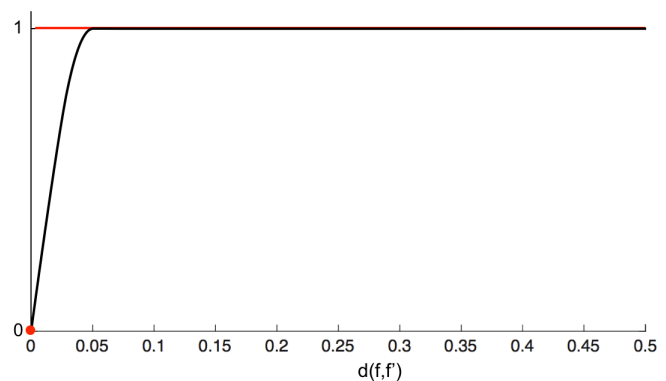


FIGURE 1 – Valeur de la distance atomique de Radon $d_{a0}(v, v')$, avec pour tout m , $v_m = e^{-j2\pi f m}$ et $v_m' = e^{-j2\pi f' m}$, en fonction de $d(f, f')$, pour $M = 10$, en noir. On observe que si $d(f, f') < 1/(2M)$, $d_{a0}(v, v') = \sin(\pi M d(f, f'))$. En revanche, si $d(f, f') \geq 1/(2M)$, $d_{a0}(v, v') = 1$, ce qui est la valeur attendue, soit la distance de Radon entre un Dirac en f et un Dirac en f' , en rouge.

puis en appliquant la méthode de Prony [9]. Enfin, on munit \mathbb{T} de la distance $d : (f, f') \in [0, 1]^2 \mapsto \min(|f - f'|, 1 - |f - f'|)$.

2 De nouvelles distances sur \mathbb{V}

2.1 Distance atomique de Radon

La distance de Radon (rééchelonnée) entre deux mesures μ et μ' de même masse peut être définie comme $\frac{1}{2}\|\mu - \mu'\|_{\text{TV}}$. Par ailleurs, on peut définir la *norme atomique* de $v \in \mathbb{V}$ par [9, 10] :

$$\|v\|_a := \inf \{\|\mu\|_{\text{TV}} : \mu \in \mathcal{M}, v = \mathcal{F}\mu\}. \quad (3)$$

Le mot *atomique* fait référence à la notion d'atomes, ou composants élémentaires, dont les éléments de \mathbb{V} sont constitués. En l'occurrence, il s'agit d'exponentielles complexes échantillonnées, correspondant aux coefficients de Fourier d'une mesure

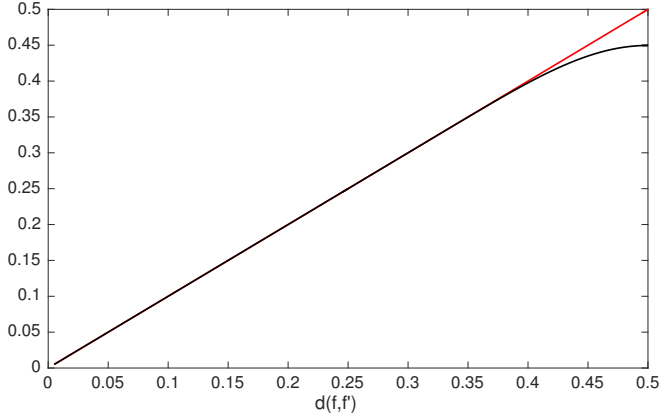


FIGURE 2 – Valeur de la distance atomique Wasserstein-1 $d_{a1}(v, v')$, avec pour tout m , $v_m = e^{-j2\pi f m}$ et $v'_m = e^{-j2\pi f' m}$, en fonction de $d(f, f')$, pour $M = 10$, en noir. On observe que cette fonction est égale à $d(f, f')$, en rouge, sauf pour les plus grandes valeurs, où cette distance est sous-estimée.

de Dirac. On peut exprimer la norme atomique sous la forme d'un problème d'optimisation semi-définie positive [9] :

$$\|v\|_a = \frac{2}{M+1} \text{tr}(V^+) - v_0 \quad \text{où} \quad (4)$$

$$V^+ := \arg \min_X \text{tr}(X) \quad \text{t.q.} \quad X \text{ est Toeplitz} \\ \text{et } X \succeq 0 \quad \text{et } X - V \succeq 0, \quad (5)$$

où tr indique la trace et $\succeq 0$ la positivité.

Il est donc naturel de définir sur \mathbb{V} la *distance atomique de Radon*, notée d_{a0} , telle que pour tous v et v' de \mathbb{V} avec $v_0 = v'_0$, on ait

$$d_{a0}(v, v') := \min \left\{ \frac{1}{2} \|\mu - \mu'\|_{\text{TV}} : (\mu, \mu') \in \mathcal{M}^2, \right. \\ \left. \mathcal{F}\mu = v, \mathcal{F}\mu' = v' \right\} \quad (6)$$

$$= \min \left\{ \frac{1}{2} \|\eta\|_{\text{TV}} : \eta \in \mathcal{M}, \mathcal{F}\eta = v - v' \right\} \quad (7)$$

$$= \frac{1}{2} \|v - v'\|_a. \quad (8)$$

La fonction d_{a0} est convexe par rapport à la paire (v, v') . Elle permet d'évaluer, de manière convexe, si v et v' sont différents ou non. Sa capacité à discriminer v et v' provenant de mesures proches, au sens de la distance de Radon, est naturellement d'autant meilleure que M est grand. On illustre en Fig. 1 la valeur de d_{a0} prise en une paire d'exponentielles complexes échantillonnées, en fonction de la séparation de leurs fréquences.

2.2 Distance atomique de Wasserstein-1

On définit ensuite la *distance atomique de Wasserstein-1*, notée d_{a1} , entre deux éléments v et v' de \mathbb{V} tels que $v_0 = v'_0$:

$$d_{a1}(v, v') := \min \left\{ d_{\mathcal{W}1}(\mu, \mu') : (\mu, \mu') \in \mathcal{M}^2, \right. \\ \left. \mathcal{F}\mu = v, \mathcal{F}\mu' = v' \right\}, \quad (9)$$

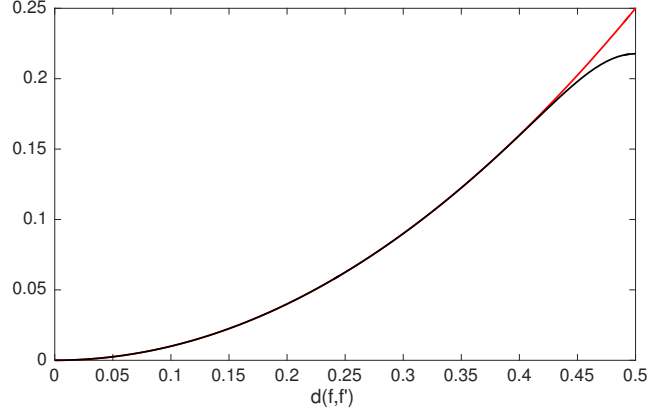


FIGURE 3 – Valeur de la fonction $\tilde{d}_{a2}(v, v')^2$, avec pour tout m , $v_m = e^{-j2\pi f m}$ et $v'_m = e^{-j2\pi f' m}$, en fonction de $d(f, f')$, pour $M = 10$, en noir. On observe que cette fonction est égale à $d(f, f')^2$, en rouge, sauf pour les plus grandes valeurs, où cette distance au carré est sous-estimée.

où l'on rappelle que la distance de Wasserstein-1 $d_{\mathcal{W}1}$ entre deux mesures positives μ et μ' sur \mathbb{T} est le minimum, parmi toutes les mesures positives ν sur \mathbb{T}^2 de marginales μ et μ' , du coût de transport $\int_{\mathbb{T}^2} d(f, f') d\nu(f, f')$ [1]. Il est connu que cette valeur ne dépend que de $\mu - \mu'$ et est égale au minimum sur le réel α de $1/(2\pi) \int_{\mathbb{T}} |F(z) - F'(z) - \alpha| dz$, où F et F' sont les fonctions de répartition de μ et μ' , respectivement [11]. Par conséquent, $d_{\mathcal{W}1}$ peut être étendue à une paire de mesures signées $(\mu, \mu') \in \mathcal{M}^2$, avec $\mu(\mathbb{T}) = \mu'(\mathbb{T})$, par $d_{\mathcal{W}1}(\mu, \mu') = d_{\mathcal{W}1}(\mu^+ + \mu'^-, \mu'^+ + \mu^-)$, où (η^+, η^-) est la décomposition de Jordan d'une mesure $\eta \in \mathcal{M}$ [12].

Ainsi, on peut réécrire la distance atomique de Wasserstein-1 comme suit :

$$d_{a1}(v, v') = \min \left\{ \|\eta\|_{\text{TV}} : \eta \in \mathcal{M}, \mathcal{F}\eta = w \text{ avec} \right. \\ \left. j2\pi m w_m = v_m - v'_m, m = -M, \dots, M \right\}. \quad (10)$$

d_{a1} est convexe par rapport à la paire (v, v') . On peut l'exprimer sous la forme d'un problème d'optimisation semi-définie positive : soit $w = ((v_m - v'_m)/(j2\pi m))_{m=-M}^M$, avec $w_0 = 0$, et $W = \text{T}(w)$. Alors

$$d_{a1}(v, v') = \min_{X, \alpha} \left(\frac{2}{M+1} \text{tr}(X) + \alpha \right) \quad \text{t.q.} \quad X \text{ est Toeplitz} \\ \text{et } X \succeq 0 \quad \text{et } X - W + \alpha \text{Id} \succeq 0, \quad (11)$$

$$= \min_X \left(\frac{2}{M+1} \text{tr}(X) + i^+(W - X) \right) \quad \text{t.q.} \\ X \text{ est Toeplitz} \quad \text{et } X \succeq 0, \quad (12)$$

où i^+ indique la plus grande valeur propre et Id est l'identité.

La Fig. 2 montre que quand v et v' sont deux exponentielles complexes échantillonnées, $d_{a1}(v, v')$ approche très bien la distance entre leurs fréquences.

2.3 Distance atomique de Wasserstein-2

La distance de Wasserstein-2 $d_{\mathcal{W}2}$ entre deux mesures positives μ et μ' sur \mathbb{T} est la racine carrée du minimum, parmi toutes

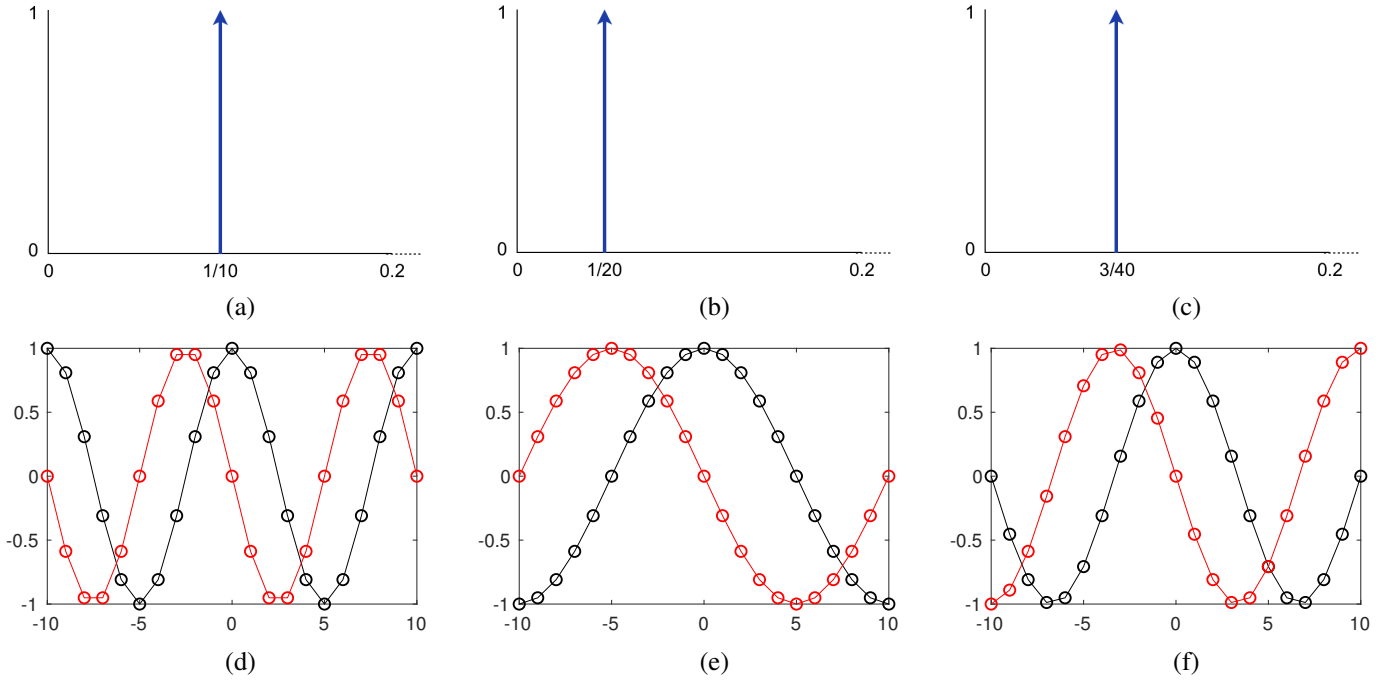


FIGURE 4 – On considère les deux mesures μ et μ' sur \mathbb{T} , qui sont des Diracs d'amplitude 1 en $1/10$ (a) et $1/20$ (b), respectivement. On dispose des séquences $v = \mathcal{F}\mu$ (d) et $v' = \mathcal{F}\mu'$ (e), avec $M = 10$, respectivement (l'index m est en abscisse, les parties réelle et imaginaire des valeurs sont en ordonnée, en noir et en rouge, respectivement). On calcule le barycentre atomique de Wasserstein $v_b \in \mathbb{V}$ qui minimise $\tilde{d}_{a2}(v, v_b)^2 + \tilde{d}_{a2}(v', v_b)^2$ (f). La mesure $\mu_b \in \mathcal{M}$ de variation totale minimale telle que $\mathcal{F}\mu_b = v_b$ est bien un Dirac d'amplitude 1 et de position $3/40$ (c), comme on s'y attend.

les mesures positives ν sur \mathbb{T}^2 de marginales μ et μ' , du coût de transport $\int_{\mathbb{T}^2} d(f, f')^2 d\nu(f, f')$ [1]. Par contre, contrairement à d_{W1} , il n'est pas sûr que l'extension à des mesures signées ait un sens. On ne définit donc pas la distance atomique de Wasserstein-2 entre deux éléments v et v' de \mathbb{V} .

En revanche, restreignons-nous au cas positif, et considérons en outre que l'un des deux vecteurs, disons v' , est une exponentielle complexe échantillonnée avec, pour tout $m = -M, \dots, M$, $v'_m = c.e^{-j2\pi f' m}$, pour certaines valeurs $f' \in \mathbb{T}$ et $c > 0$. Définissons l'opérateur \mathbb{T} , qui à $v \in \mathbb{V}$ associe la matrice Toeplitz hermitienne

$$\mathbb{T}(v) := \begin{bmatrix} v_0 & v_1 & \cdots & v_M \\ v_{-1} & v_0 & \cdots & v_{M-1} \\ \vdots & \vdots & \ddots & \vdots \\ v_{-M} & v_{-M+1} & \cdots & v_0 \end{bmatrix}. \quad (13)$$

On peut définir la fonction convexe, qui est le carré d'une distance, et qui à tout $v \in \mathbb{V}$, avec $v_0 = v'_0$ et $\mathbb{T}(v) \succcurlyeq 0$, associe le minimum sur $\mu \in \mathcal{M}$ positive, telle que $\mathcal{F}\mu = v$ du coût de transport $\int_{\mathbb{T}} d(f, f')^2 d\mu(f)$. Notez que lorsque $v_m = c.e^{-j2\pi f m}$, pour un certain $f \in \mathbb{T}$, l'unique mesure μ positive, telle que $\mathcal{F}\mu = v$, est un Dirac en f et la fonction vaut donc exactement $d(f, f')^2$. Cependant, on ne sait pas exprimer cette fonction comme un problème d'optimisation semi-définie positive.

Afin d'approcher la fonction ci-dessus, on définit la fonction

suivante : pour tout $v \in \mathbb{V}$ avec $v_0 = v'_0$ et $\mathbb{T}(v) \succcurlyeq 0$,

$$\tilde{d}_{a2}(v, v')^2 = \min \{ \eta(\mathbb{T}) : \eta \in \mathcal{M} \text{ est positive, } \mathcal{F}\eta = w, \text{ avec } -4\pi^2 m^2 w_m = v_m - 2v'_m + v_m'^2 v_m^*, m = -M, \dots, M \}. \quad (14)$$

Cette fonction, qui est convexe par rapport à v , a une forme explicite : soit $w = ((v_m - 2v'_m + v_m'^2 v_m^*) / (-4\pi^2 m^2))_{m=-M}^M$, avec $w_0 = 0$, et $W = \mathbb{T}(w)$; alors

$$\tilde{d}_{a2}(a, v)^2 = i^+(-W). \quad (15)$$

La Fig. 3 montre que quand v et v' sont deux exponentielles complexes échantillonnées, $\tilde{d}_{a2}(v, v')^2$ approche très bien la distance au carré entre leurs fréquences.

3 Exemple d'application

La Fig. 4 illustre une application simple mais importante : on calcule ce qu'on appellera le barycentre atomique de Wasserstein $v_b \in \mathbb{V}$ de $v = \mathcal{F}\mu$ et $v' = \mathcal{F}\mu'$, pour $M = 10$, où μ et μ' sont des mesures de Dirac d'amplitude 1 de positions $f = 1/10$ et $f' = 1/20$, respectivement. v_b minimise $\tilde{d}_{a2}(v, v_b)^2 + \tilde{d}_{a2}(v', v_b)^2$. Ce problème d'optimisation convexe est résolu numériquement avec la version sur-relaxée [13] de l'algorithme de Chambolle-Pock [14]. On observe empiriquement que v_b , tout comme v et v' , est une exponentielle complexe échantillonnée de fréquence $3/40$, soit la moyenne (de Fréchet sur

\mathbb{T}) des fréquences f et f' de v et v' . Dit autrement, $v_b = \mathcal{F}\mu_b$, où μ_b est un Dirac d'amplitude 1 et de position $3/40$, qui est bien le barycentre de Wasserstein [15] de μ et μ' . Le même résultat a été obtenu avec d'autres valeurs de f, f', M .

Afin de mettre en évidence la prise en compte imparfaite de la distance d par \tilde{d}_{a2} , déjà illustrée en Fig. 3, on refait la même expérience avec $f = 0$ et $f' = 0.4$, mais on minimise cette fois $(9/10)\tilde{d}_{a2}(v, v_b)^2 + (1/10)\tilde{d}_{a2}(v', v_b)^2$, avec $M = 2$. Le résultat v_b est bien une exponentielle complexe échantillonnée, mais de fréquence 0.022 au lieu de 0.04. Le fait qu'on obtienne bien une exponentielle pure montre que la minimisation de \tilde{d}_{a2} au carré, qui porte sur des vecteurs de taille $2M + 1$, favorise bien des solutions parcimonieuses, tout comme la distance de Wasserstein-2 au carré entre mesures, qui vit elle dans l'espace \mathcal{M} de dimension infinie.

4 Conclusion

Ce premier travail ouvre la voie à la manipulation de mesures, dont on veut contrôler le coût de transport, par un nombre fini de coefficients de Fourier. L'intérêt réside dans l'absence de discrétisation des mesures elles-mêmes. De nombreuses questions théoriques se posent, concernant notamment les valeurs prises par les distances atomiques proposées. L'auteur prévoit d'utiliser ces distances pour formuler de nouvelles relaxations convexes de problèmes en traitement du signal et des images [16, 17]. En outre, il faudrait voir le lien entre la construction proposée et les approches existantes de traitement de signaux sonores par transport en domaine de Fourier [18, 19].

Références

- [1] C. Villani, *Topics in Optimal Transportation*, ser. Graduate studies in mathematics. American Mathematical Society, 2003.
- [2] F. Santambrogio, *Optimal Transport for Applied Mathematicians*. Birkhauser, 2015.
- [3] J. Rabin, G. Peyré, J. Delon, and M. Bernot, "Wasserstein barycenter and its application to texture mixing," in *Proc. of Scale Space and Variational Methods in Computer Vision (SSVM) 2011*, vol. Lecture Notes in Computer Science, vol 6667. Springer, Berlin, Heidelberg, 2012.
- [4] J. Alaux, E. Grave, M. Cuturi, and A. Joulin, "Unsupervised hyper-alignment for multilingual word embeddings," in *Proc. of ICLR*, May 2019.
- [5] R. E. Curto and L. A. Fialkow, "Recursiveness, positivity, and truncated moment problems," *Houston J. Math*, vol. 17, no. 4, pp. 603–635, 1991.
- [6] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Communications on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, Jun. 2014.
- [7] L. Condat and A. Hirabayashi, "Cadzow denoising upgraded : A new projection method for the recovery of Dirac pulses from noisy linear measurements," *Sampling Theory in Signal and Image Processing*, vol. 14, no. 1, pp. 17–47, 2015.
- [8] J.-M. Azaïs, Y. De Castro, and F. Gamboa, "Spike detection from inaccurate samplings," *Applied and Computational Harmonic Analysis*, vol. 38, no. 2, pp. 177–195, Mar. 2015.
- [9] L. Condat, "Atomic norm minimization for decomposition into complex exponentials," 2018, research report, GIPSA-Lab, Grenoble, France.
- [10] B. N. Bhaskar, G. Tang, and B. Recht, "Atomic norm denoising with applications to line spectral estimation," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5987–5999, Dec. 2013.
- [11] C. A. Cabrelli and U. M. Molter, "The Kantorovich metric for probability measures on the circle," *Journal of Computational and Applied Mathematics*, vol. 57, no. 3, pp. 345–361, Feb. 1995.
- [12] T. Fisher, "Existence, uniqueness, and minimality of the Jordan measure decomposition," 2012, report arXiv :1206.5449v2.
- [13] L. Condat, "A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms," *J. Optim. Theory Appl.*, vol. 158, no. 2, pp. 460–479, 2013.
- [14] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2011.
- [15] M. Agueh and G. Carlier, "Barycenters in the Wasserstein space," *SIAM J. Math. Anal.*, vol. 43, no. 2, pp. 904–924, 2011.
- [16] L. Condat, "A convex approach to K-means clustering and image segmentation," in *Proc. of EMMCVPR. In : M. Pelillo and E. Hancock eds., Lecture Notes in Computer Science vol. 10746, Springer, pp. 220–234, 2018, Venice, Italy, Oct. 2017.*
- [17] L. Condat, D. Kitahara, and A. Hirabayashi, "A convex lifting approach to image phase unwrapping," in *Proc. of IEEE ICASSP*, Brighton, UK, May 2019.
- [18] R. Flamary, C. Févotte, N. Courty, and V. Emiya, "Optimal spectral transportation with application to music transcription," in *Advances in Neural Information Processing Systems (Proc. of NIPS)*, 2016, pp. 703–711.
- [19] A. Rolet, V. Seguy, M. Blondel, and H. Sawada, "Blind source separation with optimal transport non-negative matrix factorization," *EURASIP Journal on Advances in Signal Processing*, vol. 53, 2018.