

REGULARIZED WASSERSTEIN BARYCENTER

Elsa Cazelles¹ & Jérémie Bigot² & Nicolas Papadakis³

¹ *Institut de Mathématiques de Bordeaux. elsa.cazelles@u-bordeaux.fr*

² *Institut de Mathématiques de Bordeaux. jeremie.bigot@math.u-bordeaux1.fr*

³ *CNRS. nicolas.papadakis@math.u-bordeaux1.fr*

Résumé.

Le barycentre dans l'espace de Wasserstein permet de définir une notion de moyenne de Fréchet sur un ensemble de probabilités. Cependant, selon les données considérées, ce barycentre peut être irrégulier. Dans ce papier, nous décidons d'introduire une régularisation convexe sur les barycentres de Wasserstein pour des mesures aléatoires définies sur \mathbb{R}^d . Nous prouvons l'existence et l'unicité de tels barycentres pour une large classe de fonctions régularisantes. Nous sommes également en mesure d'obtenir un résultat de stabilité du barycentre régularisé en terme de divergence de Bregman associé à la régularisation convexe, notamment lorsque l'on dispose d'observations aléatoires. En particulier, nous démontrons la convergence en divergence de Bregman du barycentre empirique régularisé de n mesures de probabilité aléatoires de loi \mathbb{P} vers le vrai barycentre généralisé de \mathbb{P} , et nous discutons de sa vitesse de convergence. Cette approche est adaptée pour l'étude de probabilités aléatoires discrètes ou bien absolument continues par rapport à la mesure de Lebesgue. À titre d'exemple, nous nous concentrons sur l'analyse de mesures de probabilité réelles. Ainsi nous présentons un algorithme de minimisation efficace basé sur une descente de gradient accélérée pour calculer le barycentre de Wasserstein régularisé. Finalement, nous appliquons cette analyse statistique à des ensemble de données simulées et réelles.

Mots-clés. Espace de Wasserstein, Barycentre de mesures de probabilité, Moyenne de Fréchet, Divergence de Bregman, Analyse convexe, Convergence de barycentre empirique, Régularisation

Abstract.

The concept of barycenter in the Wasserstein space allows the definition of a notion of Fréchet mean of a set of probability measures. However, depending on the considered data, such barycenters may be irregular. In this paper, we thus introduce a convex regularization of Wasserstein barycenters for random measures supported on \mathbb{R}^d . We prove the existence and uniqueness of such barycenters for a large class of regularizing functions. A stability result of regularized barycenters in terms of Bregman divergence associated to the convex regularization term is also given. Hence, we study the case of data made of i.i.d. random probability measures. In particular, we prove the convergence in Bregman divergence of the regularized empirical barycenter of a set of n random probability measures towards its

population counterpart, and we discuss its rate of convergence. This approach is shown to be appropriate for the study of discrete or absolutely continuous random measures. We then focus on the analysis of probability measures supported on the real line. In this setting, we propose an efficient minimization algorithm based on accelerated gradient descent for the computation of regularized Wasserstein barycenters. This approach is finally applied to the statistical analysis of simulated and real data sets.

Keywords. Wasserstein space, Barycenter of probability measures, Fréchet mean, Bregman divergence, Convex analysis, Convergence of empirical barycenter, Regularization

Présentation

Le barycentre dans l'espace de Wasserstein - ou moyenne de Fréchet- introduit par Agueh et Carlier [1] est un outil efficace pour synthétiser les informations issues d'un ensemble de mesures de probabilité. Il correspond à la notion de barycentre euclidien appliquée à un espace non-linéaire muni d'une métrique non-euclidienne. Soient Ω un ensemble de \mathbb{R}^d et $\mathcal{P}_2(\Omega)$ l'ensemble des mesures de probabilité définies sur Ω de moment d'ordre deux fini. On note W_2 la distance de Wasserstein sur $\mathcal{P}_2(\Omega)$ associée au coût quadratique sur Ω . Le barycentre empirique de Wasserstein d'un ensemble de n mesures de probabilité ν_1, \dots, ν_n est donc défini comme un minimiseur de

$$\mu \mapsto \frac{1}{n} \sum_{i=1}^n W_2^2(\mu, \nu_i), \text{ pour } \mu \in \mathcal{P}_2(\Omega). \quad (1)$$

Lorsqu'on souhaite manipuler des données réelles, le transport optimal entre les différentes distributions est souvent irrégulier. Pour illustrer ce propos, nous considérons un ensemble de données provenant de l'étude de potentiels d'actions neuronales qui est disponible sur le site MBI : <http://mbi.osu.edu/2012/stwdescription.html>. Au cours d'une expérience, il est demandé à un singe de déplacer sa main d'un point S à un point N. Les pics de tension du neurone affecté au mouvement de la main du singe sont enregistrés pendant 5 secondes et l'expérience est répétée $n = 60$ fois. Chaque suite de potentiels est alors lissée en utilisant un noyau Gaussien. Pour chaque essai $1 \leq i \leq n$, ν_i désigne la mesure de densité de probabilité égale à la somme des différentes gaussiennes renormalisées. Les résultats sont présentés dans la Figure 1(a). La densité de probabilité du barycentre de Wasserstein est tracée dans la Figure 1(b). Cette approche amène clairement à une estimation très irrégulière du modèle de densité de potentiel électrique.

Ainsi, le barycentre associé aux données n'est visuellement pas représentatif de leur moyenne. Pour pallier à ces défauts et en s'inspirant de travaux de problèmes inverses [2], nous introduisons une régularisation convexe du barycentre. Dans le but d'analyser un ensemble de données, nous considérons qu'elles peuvent être décrites comme un ensemble de probabilités aléatoires et alors correspondre à l'étude d'une suite de probabilités discrètes

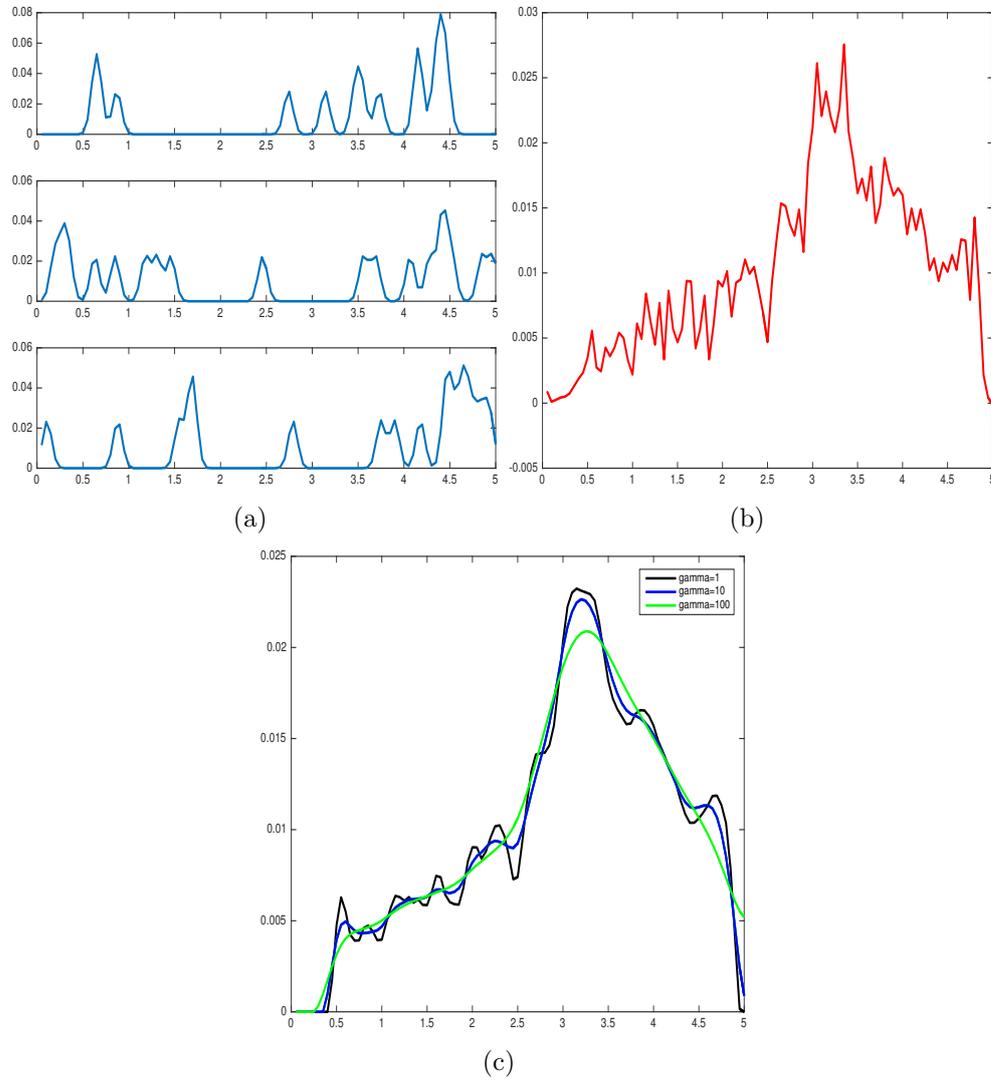


Figure 1: (a) Un exemple de 3 potentiels électriques lissés parmi $n = 60$. Chaque ligne représente une expérience et la densité obtenue en lissant chaque pic avec un noyau Gaussien de largeur à mi-hauteur 50 millisecondes. (b) Densité de probabilité du barycentre empirique de Wasserstein pour cet ensemble de données. (c) Densité de probabilité du barycentre régularisé empirique pour différentes valeurs de γ .

ou bien à densité (ie. histogrammes). Ce type de données est largement utilisé dans des domaines tels que les neurosciences, l'étude de la biodémographie et la génomique, l'économie ou encore l'imagerie médicale. Nous nous concentrons sur des méthodes de statistique du premier ordre dans le but d'estimer, à partir de telles données, la densité de probabilité de la moyenne d'une telle population, c'est-à-dire la fonction minimisant la fonctionnelle

$$f \mapsto \frac{1}{n} \sum_{i=1}^n W_2^2(\mu_f, \nu_i) + \gamma E(f) \text{ avec } f \in \mathcal{D}_+(\Omega), \quad (2)$$

où $\mathcal{D}_+(\Omega)$ est l'espace des densités de probabilité sur Ω , μ_f désigne la mesure de densité f , $E : \mathcal{D}_+(\Omega) \rightarrow \mathbb{R}_+$ est une fonction de pénalité lisse strictement convexe et $\gamma \geq 0$ est un paramètre de régularisation. La Figure 1(c) représente la densité de probabilité du barycentre régularisé des potentiels d'action pour les valeurs $\gamma = 1, 10, 100$.

Dans ce papier, nous nous intéressons en premier lieu au barycentre régularisé généralisé, c'est-à-dire la moyenne de Fréchet régularisée d'une mesure de probabilité \mathbb{P} sur l'espace des mesures de probabilités. Nous prouvons l'existence et l'unicité du barycentre régularisé. Le théorème de Prokhorov permet de prouver l'existence de ce barycentre et l'unicité provient de la convexité de la fonction de régularisation. Dans un second temps, nous considérons la moyenne de Fréchet régularisée associée à $\mathbb{P}_n = \frac{1}{n} \sum \delta_{\nu_i}$, où les ν_1, \dots, ν_n sont des probabilités aléatoires issues d'une même distribution \mathbb{P} . En considérant un barycentre qui serait également absolument continu par rapport à la mesure de Lebesgue, nous obtenons un résultat de stabilité du barycentre régularisé en terme de divergence de Bregman symétrisée associée à la régularisation. Comme E est différentiable, nous rappelons que la divergence de Bregman est définie pour $h_1, h_2 \in \mathcal{P}_{2,ac}(\Omega)$ l'ensemble des densités de probabilité sur Ω admettant un moment d'ordre deux par

$$D_E(h_1, h_2) = E(h_1) - E(h_2) - \langle \nabla E(h_2), h_1 - h_2 \rangle.$$

En particulier, nous avons accès à la stabilité lorsque les mesures aléatoires proviennent d'observations indépendantes identiquement distribuées (Figure 2). La preuve repose sur une inégalité impliquant les sous-différentielles de la distance de Wasserstein et de la fonction de régularisation, obtenue par le biais de l'analyse convexe.

Le principal intérêt de cette communication réside dans les propriétés statistiques du barycentre régularisé. Dans la suite, nous nous concentrons sur des barycentres absolument continus par rapport à la mesure de Lebesgue. En effet, nous montrons que le barycentre régularisé $f_{\mathbb{P}_n}^\gamma$ converge en divergence de Bregman associé à la régularisation E vers le vrai barycentre généralisé $f_{\mathbb{P}}^\gamma$ où

$$f_{\mathbb{P}}^\gamma \in \operatorname{argmin}_{f \in \mathcal{P}_{2,ac}(\Omega)} J_{\mathbb{P}}^\gamma(f) = \int W_2^2(\mu_f, \nu) d\mathbb{P}(\nu) + \gamma E(f).$$

Autrement dit, nous prouvons que le barycentre régularisé de n observations i.i.d. de distribution \mathbb{P} converge vers le vrai barycentre de \mathbb{P} . Pour cela, nous introduisons le

barycentre généralisé régularisé. Nous disposons alors d'un taux de convergence entre les barycentres régularisés de \mathbb{P}_n et de \mathbb{P} de l'ordre de $1/\gamma^2 n$, où γ correspond à la constante de régularisation:

Théorème 0.1. *Si Ω est un compact de \mathbb{R}^d , on a*

$$E_{\mathbb{P}}(D_E^2(f_{\mathbb{P}_n}^\gamma, f_{\mathbb{P}}^\gamma)) \leq \frac{CJ(1, \mathcal{H})\|H\|_{\mathbb{L}_2(\mathbb{P})}}{\gamma^2 n} \quad (3)$$

avec C une constante positive, $\mathcal{H} = \{h_\mu : \nu \in \mathcal{P}_2(\Omega) \mapsto W_2^2(\mu, \nu) \in \mathbb{R}; \mu \in \mathcal{P}_2(\Omega)\}$ la classe de fonctions d'enveloppe H et

$$J(\delta, \mathcal{H}) = \sup_Q \int_0^\delta \sqrt{1 + \log N(\epsilon \|H\|_{\mathbb{L}_2(Q)}, \mathcal{H}, \mathbb{L}_2(Q))} d\epsilon$$

où le supremum est pris sur l'ensemble des mesures de probabilité discrètes Q telle que $\|H\|_{\mathbb{L}_2(Q)} > 0$. N désigne l'entropie métrique.

La théorie des processus empiriques permet de majorer la divergence de Bregman. Quant à la convergence entre le vrai barycentre de \mathbb{P} et son barycentre régularisé, elle est obtenue par Γ -convergence.

Afin de se convaincre de la nécessité d'une régularisation sur le barycentre, nous présentons cette approche en l'appliquant à un jeu de données sous forme d'histogrammes. Par descente de gradient projetée, ainsi qu'une accélération de type FISTA [4], nous obtenons numériquement le barycentre régularisé de nos données. Par ailleurs, comme nos données sont réelles, il est possible d'obtenir le barycentre en calculant les fonctions quantiles empiriques de nos histogrammes: celui-ci s'avère moins cohérent que notre barycentre régularisé.

Nous étudions le cas où les mesures ν_i sont empiriques et définies par un petit nombre p_i d'observations. Ainsi, pour chaque $i = 1, \dots, n$, $(X_{ij})_{1 \leq j \leq p_i}$ est une suite de variables aléatoires i.i.d de loi gaussienne $\mathcal{N}(\mu_i, \sigma_i^2)$ et $\nu_i = \frac{1}{p_i} \sum_{j=1}^{p_i} \delta_{X_{ij}}$ où δ_x est la mesure de Dirac en x . Dans un premier temps, nous avons lissé chaque mesure ν_i afin d'obtenir une densité de probabilité par l'estimateur à noyau suivant

$$\hat{f}_{i, h_i}(x) = \frac{1}{p_i h_i} \sum_{j=1}^{p_i} K\left(\frac{x - X_{ij}}{h_i}\right)$$

où $x \in \Omega$ et K est un noyau Gaussien. La fenêtre h_i est choisie par validation croisée. Ensuite, via la formulation en 1D de la distance de Wasserstein, nous avons tracé en bleu dans la Figure 2 le barycentre de Wasserstein non régularisé associé à ces densités de mesures. Dans un second temps, nous avons appliqué notre algorithme directement sur les mesures ν_i : la courbe orange représente leur barycentre régularisé.

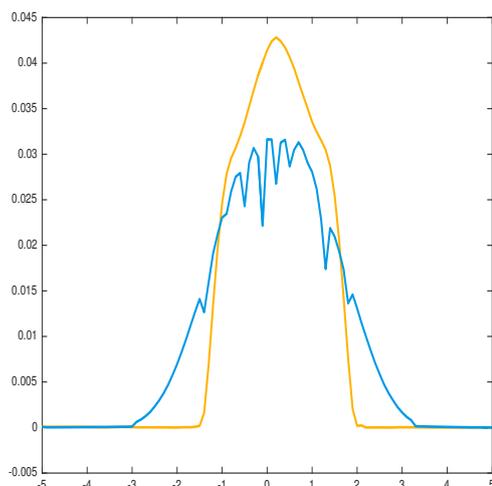


Figure 2: Pour $1 \leq i \leq n = 100$ mesures, nous simulons $(X_{ij})_{1 \leq j \leq p_i}$ variables aléatoires de loi gaussienne $\mathcal{N}(\mu_i, \sigma_i^2)$, où p_i varie entre 5 et 10. A l'aide d'un noyau gaussien et par validation croisée, nous estimons les densités associées aux $(X_{ij})_{1 \leq j \leq p_i}$. La courbe bleu représente le barycentre de Wasserstein de ces densités. La courbe orange représente le barycentre régularisé des ν_i avec $\gamma = 30$ et E une régularisation de type Dirichlet sur les dérivées secondes.

Bibliographie

- [1] Martial Agueh and Guillaume Carlier. Barycenters in the Wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2):904-924, 2011.
- [2] Martin Burger, Marzena Franek and Carola-Bibiane Schönlieb. Regularized regression and density estimation based on optimal transport. *Applied Mathematics Research eXpress*, 2012(2):209-253, 2012.
- [3] Thibault Le Gouic and Jean-Michel Loubes. Existence and consistency of Wasserstein barycenters. *arXiv preprint arXiv:1506.04153*, 2015.
- [4] Amir Beck and Marc Teboulle. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM Journal on Imaging Sciences*, 2(1): 183-202, 2009.