

# Transport optimal entre GMM pour la synthèse de texture

Julie DELON<sup>1</sup>, Agnès DESOLNEUX<sup>2</sup>, Arthur LECLAIRE<sup>3</sup>

<sup>1</sup>MAP5, CNRS, Université de Paris, France & Institut Universitaire de France (IUF)

<sup>2</sup>Centre Borelli, CNRS, ENS Paris-Saclay, France

<sup>3</sup>IMB, CNRS, Université de Bordeaux, France

julie.delon@u-paris.fr, agnes.desolneux@ens-paris-saclay.fr,  
arthur.leclaire@math.u-bordeaux.fr

**Résumé** – L’utilisation du transport optimal est devenue très populaire en traitement d’images, malgré des temps de calcul en grande dimension souvent prohibitifs. Nous proposons ici d’utiliser une formulation alternative récente consistant à restreindre le problème du transport optimal à l’ensemble des modèles de mélanges gaussiens pour en simplifier le calcul. Nous utilisons cette formulation pour revisiter le modèle de texture Texto basé sur le transport optimal entre distributions de patches d’images. L’utilisation de cette formulation dans ce modèle de texture permet de traiter de plus grands patches, fournissant ainsi des résultats avec de meilleurs détails géométriques.

**Abstract** – Using optimal transport in image processing tasks has become very popular. However, it still faces difficult computational issues when dealing with high dimensional distributions. We propose here to use another formulation that consists in restricting the optimal transport problem to the set of Gaussian mixture models (GMM). As a proof of concept, we use it to revisit the texture model Texto based on optimal transport between distributions of image patches. Using this GMM-based formulation in this texture model allows to deal with larger patches, hence providing results with better geometric details.

## 1 Introduction

Le transport optimal numérique a connu des progrès spectaculaires depuis dix ans et est utilisé aujourd’hui dans des domaines très variés [1, 4, 10, 7]. D’importantes avancées ont notamment été obtenues dans le calcul numérique des solutions du transport, avec l’émergence d’outils tels que le transport régularisé [5] ou le transport par tranches [2]. Il reste cependant complexe de calculer des distances de transport entre distributions empiriques lorsque le nombre d’échantillons ou la dimension de l’espace deviennent trop grands. Ce sont en effet des contextes qui posent de nombreuses questions [12, 9, 3].

Cependant, pour de nombreuses applications, il n’est pas forcément souhaitable de chercher des solutions exactes du transport optimal, et des formulations partageant des propriétés similaires peuvent fournir des solutions plus pertinentes en pratique. Parmi ces formulations alternatives, une distance de transport entre mélanges de gaussiennes (GMM) a été introduite dans [6]. Elle consiste à ne chercher des couplages du problème de transport que parmi les GMM (on la notera GMM-OT dans ce qui suit).

Lorsque GMM-OT est appliqué à des données discrètes, la dimension  $d$  de l’espace et le nombre d’échantillons  $n$  n’apparaissent que dans deux étapes du calcul des solutions : l’étape d’inférence du GMM sur les données, et le calcul de distances de Bures entre les matrices de covariance des composantes des GMM. Cela rend l’approche très polyvalente et robuste à la di-

mension dans la pratique. Dans cet article, pour illustrer l’intérêt de cette approche GMM-OT, nous revisitons le modèle de texture Texto [8], basé sur le transport optimal semi-discret entre distributions de patches. Nous aboutissons à un schéma numérique plus léger et plus simple pour le même problème : le temps de calcul de GMM-OT est au moins un ordre de grandeur plus rapide que celui du transport semi-discret ou régularisé, pour une qualité similaire (voire meilleure) des images synthétisées. Cela permet d’utiliser des patches plus grands, et beaucoup plus de patches que le modèle Texto original.

## 2 Rappels sur le transport optimal entre mélanges de gaussiennes

Cette section rappelle les principaux résultats de [6]. La distance de Wasserstein quadratique entre deux mesures de probabilité  $\mu_0$  et  $\mu_1$  sur  $\mathbb{R}^d$  de seconds moments finis s’écrit

$$W_2^2(\mu_0, \mu_1) := \inf_{\gamma \in \Pi(\mu_0, \mu_1)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|y_0 - y_1\|^2 d\gamma(y_0, y_1), \quad (1)$$

où  $\Pi(\mu_0, \mu_1)$  est l’ensemble des mesures de probabilité sur  $\mathbb{R}^d \times \mathbb{R}^d$  dont les marginales sont  $\mu_0$  et  $\mu_1$ . Une solution  $\gamma^*$  de (1) est appelée plan de transport optimal entre  $\mu_0$  et  $\mu_1$ . Cette distance a été largement utilisée pour différentes applications en science des données, en particulier pour définir des barycentres entre mesures de probabilité.

**Définition de  $MW_2$ .** Notons  $GMM_d$  l'ensemble des mesures de probabilité qui s'écrivent comme des GMM finis sur  $\mathbb{R}^d$ . Les plans de transport et barycentres entre GMM pour  $W_2$  ne sont généralement pas eux-mêmes des GMM, ce qui est limitant si cette représentation est utilisée pour analyser ou générer des données. Pour cette raison, les autrices de [6] proposent de modifier la formulation de  $W_2$  en imposant aux couplages d'être des GMM sur  $\mathbb{R}^d \times \mathbb{R}^d$ . Plus précisément, soient  $\mu_0, \mu_1 \in GMM_d$ , on peut définir

$$MW_2^2(\mu_0, \mu_1) := \inf_{\gamma \in \Pi^{GMM}(\mu_0, \mu_1)} \int_{\mathbb{R}^{2d}} \|y_0 - y_1\|^2 d\gamma(y_0, y_1), \quad (2)$$

où  $\Pi^{GMM}(\mu_0, \mu_1)$  est l'ensemble des mesures de probabilité de  $GMM_{2d}$  de marginales  $\mu_0$  et  $\mu_1$ . Il est montré dans [6] que  $MW_2$  définit une distance entre éléments de  $GMM_d$ . De plus, si  $\mu_0 = \sum_{k=1}^{K_0} \pi_0^k \mu_0^k$  et  $\mu_1 = \sum_{l=1}^{K_1} \pi_1^l \mu_1^l$ , où les  $\pi_0^k, \pi_1^l$  sont des scalaires positifs de somme 1, et les  $\mu_0^k, \mu_1^l$  des mesures gaussiennes, on peut montrer [6] que

$$MW_2^2(\mu_0, \mu_1) = \min_{w \in \Pi(\pi_0, \pi_1)} \sum_{k,l} w_{kl} W_2^2(\mu_0^k, \mu_1^l), \quad (3)$$

où  $\Pi(\pi_0, \pi_1)$  est l'ensemble des matrices  $K_0 \times K_1$  d'entrées positives et de marginales  $\pi_0$  et  $\pi_1$ . Cette expression discrète rend  $MW_2$  très simple à calculer en pratique, même en grande dimension. En effet, la distance  $W_2$  entre deux mesures gaussiennes  $\mu = \mathcal{N}(m, \Sigma)$  et  $\tilde{\mu} = \mathcal{N}(\tilde{m}, \tilde{\Sigma})$  a la forme close

$$W_2^2(\mu, \tilde{\mu}) = \|m - \tilde{m}\|^2 + \text{tr} \left( \Sigma + \tilde{\Sigma} - 2 \left( \Sigma^{\frac{1}{2}} \tilde{\Sigma} \Sigma^{\frac{1}{2}} \right)^{\frac{1}{2}} \right), \quad (4)$$

où  $M^{\frac{1}{2}}$  est l'unique racine semi-définie positive de la matrice semi-définie positive  $M$ . Si les différents paramètres des GMM  $\mu_0$  et  $\mu_1$  sont connus, calculer (3) revient à calculer  $K_0 \times K_1$  distances  $W_2$  entre gaussiennes et à résoudre un problème de transport discret de taille  $K_0 \times K_1$ . Il est de même possible de définir des barycentres pour  $MW_2$ , et cela donne lieu à une formulation discrète du même type [6].

**Utiliser  $MW_2$  en pratique.** Les plans de transport optimal pour  $MW_2$  ne fournissent pas directement une application de transport entre les mélanges  $\mu_0$  et  $\mu_1$ . Afin de définir une telle application à partir du plan  $\gamma^*$  on peut utiliser par exemple

$$T(x) = \mathbb{E}_{(X,Y) \sim \gamma^*}(Y | X = x). \quad (5)$$

Comme montré dans [6],  $T$  peut se réécrire

$$T(x) = \frac{\sum_{k,l} w_{k,l}^* g_{m_0^k, \Sigma_0^k}(x) T_{k,l}(x)}{\sum_k \pi_0^k g_{m_0^k, \Sigma_0^k}(x)}, \quad (6)$$

où  $w^*$  est la solution optimale du problème discret (3), les  $T_{k,l}$  sont les cartes affines optimales entre les gaussiennes  $\mu_0^k$  et  $\mu_1^l$ , et  $g_{m, \Sigma}$  est la densité de  $\mathcal{N}(m, \Sigma)$ .

### 3 Le modèle Texto-GMM

Soit  $u : \Omega \rightarrow \mathbb{R}^d$  une texture exemple définie sur un domaine rectangulaire  $\Omega \subset \mathbb{Z}^2$ . Notons  $U$  une initialisation pour

la texture synthétisée. L'idée de Texto-GMM, inspirée de [8], est d'utiliser le transport optimal pour imposer à la distribution des patches de  $U$  de ressembler à celle de  $u$ , à plusieurs échelles. La texture synthétisée est alors recomposée à partir de cette distribution de patches par simple moyenne locale. On notera dans ce qui suit  $\omega = \{0, \dots, w-1\}^2$  le domaine des patches, et  $u_{|a+\omega} \in \mathbb{R}^\omega$  le patch à la position  $a$ .

**Le modèle mono-échelle.** Commençons par décrire la synthèse à une seule échelle. Comme initialisation, on génère un champ aléatoire gaussien stationnaire  $U$  de même moyenne et covariance que la texture exemple  $u$ , défini par

$$\forall a \in \mathbb{Z}^2, U(a) = \bar{u} + \frac{1}{\sqrt{|\Omega|}} \sum_{b \in \Omega} (u(b) - \bar{u}) W(a-b) \quad (7)$$

où  $\bar{u} = \frac{1}{|\Omega|} \sum_{a \in \Omega} u(a)$  et où  $W$  est un champ aléatoire sur  $\mathbb{Z}^2$  dont les valeurs des pixels sont i.i.d. de loi  $\mathcal{N}(0, 1)$ . Le champ  $U$  possède certaines caractéristiques globales de  $u$ , mais pas ses détails. La distribution  $\hat{\mu}$  des patches de  $U$  est alors envoyée par transport optimal sur la distribution  $\hat{\nu}$  des patches de  $u$  afin de réimposer ces détails à l'image synthétisée. Pour que ce calcul de transport soit rapide, on utilise ici GMM-OT comme décrit dans la section précédente, en approchant les distributions discrètes  $\hat{\mu}$  et  $\hat{\nu}$  par des GMM  $\mu$  et  $\nu$  à l'aide de l'algorithme EM<sup>1</sup>. La formule (5) nous permet de déduire une application de transport  $T : \mathbb{R}^\omega \rightarrow \mathbb{R}^\omega$  dans l'espace des patches à partir du plan solution pour  $MW_2(\mu, \nu)$ . Finalement, la nouvelle image synthétisée  $V$  est calculée en moyennant tous les patches transportés, ce qui s'écrit

$$\forall a \in \mathbb{Z}^2, V(a) = \frac{1}{|\omega|} \sum_{h \in \omega} T(U_{|a-h+\omega})(h). \quad (8)$$

Notons qu'à cause de ce moyennage, la distribution de patches de l'image  $V$  n'est pas tout à fait celle des patches transportés. Imposer plus précisément la distribution des patches de l'image synthétisée nécessiterait des techniques d'optimisation du modèle plus sophistiquées.

**Le modèle multi-échelle.** Étendons maintenant le modèle précédent à plusieurs échelles. Pour  $0 \leq s \leq S-1$ , considérons une version sous-échantillonnée  $u_s$  de  $u$  définie sur le sous-domaine  $\Omega_s \subset 2^s \mathbb{Z}^2$ . À l'échelle grossière,  $U_{S-1}$  est initialisé comme le champ gaussien (7) estimé à partir de  $u_{S-1}$ . Supposons maintenant que  $U_s$  à l'échelle  $s \in \{1, \dots, S-1\}$  est donné. À nouveau, on estime des GMM  $\mu_s$  et  $\nu_s$  à partir des distributions de patches de  $U_s$  et  $u_s$ , et on en déduit une application de transport  $T_s$  par GMM-OT entre  $\mu_s$  et  $\nu_s$ , puis on recalcule une image synthétisée  $V_s$  à l'échelle  $s$  à partir des patches transportés. On utilise pour cela une variante de la formule (8), dans laquelle chaque patch transporté  $T_s(U_{s|a+2^s\omega})$  est remplacé par son plus proche voisin  $C_s(a)$  (au sens  $L^2$ ) dans  $u_s$ ,

1. Remarquons qu'il serait intéressant ici de disposer d'une autre méthode d'estimation du GMM minimisant directement une distance de transport entre le GMM et la distribution discrète des patches.

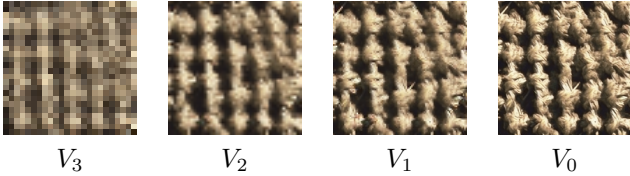


FIGURE 1 – Processus de synthèse *coarse-to-fine*.

soit

$$V_s(a) = \frac{1}{|\omega|} \sum_{b \in 2^s \omega} u_s(C_s(a-b) + b), \quad a \in 2^s \mathbb{Z}^2. \quad (9)$$

Ce mécanisme permet de facilement initialiser la synthèse à l'échelle supérieure, en prenant des patches deux fois plus grands aux mêmes positions. Plus précisément,  $U_{s-1}$  est initialisée en posant pour tout  $a \in 2^s \mathbb{Z}^2$  et tout  $k \in \{0, 2^{s-1}\}^2$ ,

$$U_{s-1}(a+k) = \frac{1}{|\omega|} \sum_{b \in 2^s \omega} u_{s-1}(C_s(a-b) + b + k). \quad (10)$$

À la fin du processus, on obtient l'image synthétisée  $V_0$ . Ce processus de synthèse *coarse-to-fine* est illustré sur la Fig. 1. Mentionnons finalement qu'une fois le modèle estimé (c'est-à-dire tous les GMM et tous les plans de transport calculés à partir d'une synthèse), il est possible de l'utiliser hors-ligne pour faire de la synthèse d'images à la demande (et de taille arbitraire).

## 4 Résultats expérimentaux

Plusieurs exemples de synthèse sont donnés en Fig. 2. Dans les légendes, on note *Texto-GMM* la synthèse obtenue après le processus multi-échelle de la Section 3. Pour toutes les expériences montrées, les GMM estimés ont 4 composantes. On a observé empiriquement que prendre plus de composantes n'améliore que marginalement les résultats visuels (tout en ralentissant l'estimation du modèle).

En pratique, lorsque l'on travaille avec des patches plus grands que  $5 \times 5$ , on restreint l'algorithme EM et les projections au plus proche voisin à un sous-ensemble de  $N_p = 10^4$  patches choisis aléatoirement. Ce sous-échantillonnage ne dégrade pas la qualité de la synthèse, et accélère ces deux étapes. Ceci constitue une amélioration notable par rapport à la méthode *Texto* [8] qui était limitée à des distributions de  $10^3$  patches de taille  $3 \times 3$  (à cause de l'algorithme stochastique semi-discret utilisé [8]).

L'estimation du modèle *Texto-GMM* est bien plus rapide qu'avec *Texto* [8], même avec dix fois plus de patches. En effet, l'algorithme GMM-OT est bien plus rapide que l'algorithme stochastique utilisé dans [8] ou même plus rapide que l'algorithme Sinkhorn [5]. Par exemple, pour un problème de transport avec  $10^4$  points à la source et à la cible en dimension 147 (pour des patches couleurs de taille  $7 \times 7$ ), avec une architecture récente, résoudre GMM-OT prend environ 1', alors qu'il faut 2.5h pour  $10^5$  itérations de l'algorithme stochastique semi-discret, ou 40' pour faire  $10^3$  itérations de l'algorithme

Sinkhorn. Avec des patches  $5 \times 5$ , pour une image  $256 \times 256$ , l'algorithme de synthèse pour *Texto-GMM* (avec  $N_p = 10^4$ ) prend environ 1' alors que l'algorithme de synthèse *Texto* (incluant l'estimation) (avec  $N_p = 10^3$ ) prend plus d'1h. L'étape la plus coûteuse de *Texto-GMM* est en fait la projection au plus proche voisin utilisée pour sur-échantillonner, qui pourrait être accélérée avec des algorithmes dédiés [11].

En termes de temps de calcul, la formulation GMM-OT est moins sensible à la dimension et permet donc d'utiliser des patches plus grands, et en plus grand nombre. Cela permet d'obtenir des synthèses plus précises pour les textures structurées, avec une meilleure préservation des détails saillants (Fig. 2). La deuxième colonne de la Fig. 2 (NN) confirme l'importance d'utiliser le transport optimal pour transformer les patches, et pas seulement des projections au plus proche voisin. Les troisième et quatrième colonnes confirment que le modèle *Texto* est limité aux petits patches. À l'inverse, *Texto-GMM* produit de très bons résultats sur ces textures avec le bon choix de paramètres  $w$  et  $S$ . On ne prétend pas ici dépasser l'état de l'art en synthèse de textures, mais à montrer comment un modèle de synthèse basé sur le transport optimal peut être amélioré en utilisant GMM-OT au lieu du transport semi-discret.

Le modèle *Texto-GMM* peut également être adapté pour le mélange de textures en utilisant la formule explicite de [6] pour les barycentres de GMM-OT. Pour un paramètre de mélange  $\alpha \in [0, 1]$ , on part du modèle gaussien mélangé  $U_{S-1}$  de [13], et à chaque échelle  $s$  on applique une transformation par patch qui cible la distribution  $\nu_s^\alpha$  obtenue en mélangeant les GMM  $\nu_s^0, \nu_s^1$  provenant des patches des images  $u^0, u^1$  respectivement. Pour le sur-échantillonnage, on utilise une projection au plus proche voisin sur des patches  $(1-\alpha)p + \alpha T(p)$  où  $p$  est un patch de  $\nu_s^0$  et  $T$  l'application (6) provenant du GMM-OT entre  $\nu_s^0, \nu_s^1$ . Sur la Fig. 3, on peut voir que *Texto-GMM* produit des résultats pertinents pour cette application. Il serait néanmoins intéressant de pouvoir exprimer plus directement le modèle de texture mélangé (en évitant un mélange ad-hoc des patches entre source et cible).

## Références

- [1] N. Bonneel, M. Van De Panne, S. Paris, and W. Heidrich. Displacement interpolation using Lagrangian mass transport. In *Proc. of SIGGRAPH'Asia*, pages 1–12, 2011.
- [2] N. Bonnotte. *Unidimensional and evolution methods for optimal transportation*. PhD thesis, Paris 11, 2013.
- [3] L. Chizat, P. Roussillon, F. Léger, F.X. Vialard, and G. Peyré. Faster Wasserstein distance estimation with the Sinkhorn divergence. *Adv. Neural Inf. Process. Syst.*, 33 :2257–2269, 2020.
- [4] N. Courty, R. Flamary, A. Habrard, and A. Rakotomamonjy. Joint distribution optimal transportation for domain adaptation. *Adv. Neural Inf. Process. Syst.*, 30, 2017.

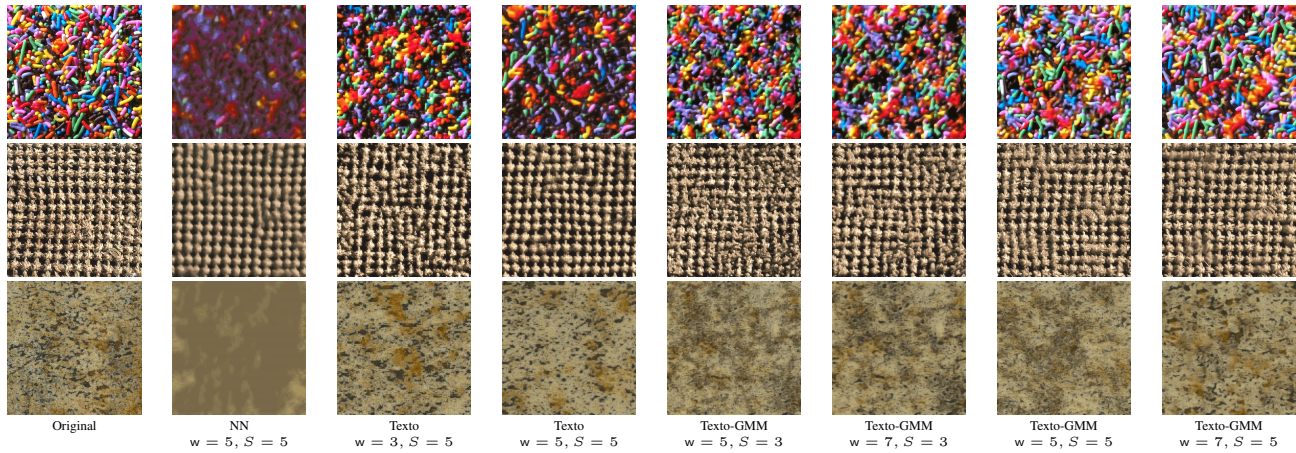


FIGURE 2 – Textures originales  $512 \times 512$  (colonne 1) et synthèses obtenues avec différents modèles expliqués dans l’article, avec différentes valeurs des paramètres  $w$  (taille de patch) et  $S$  (nombre d’échelles). Dans la colonne 2 (NN), la synthèse est obtenue en faisant des projections au plus proche voisin à chaque échelle. Enfin, on montre les résultats obtenus avec le modèle précédent Texto [8] (colonnes 3, 4) et le modèle Texto-GMM model proposé ici (colonnes 5–8).

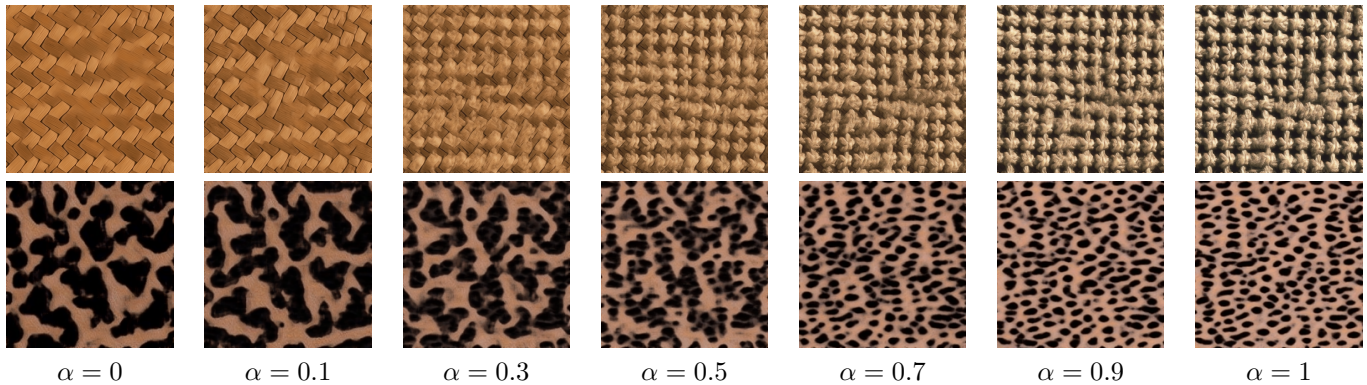


FIGURE 3 – Mélanges de textures à partir de deux images sources  $u_0, u_1$  (non montrées). De gauche à droite, on montre des échantillons du modèle mélangé pour différentes valeurs du paramètre de mélange  $\alpha$  dans  $[0, 1]$ . Pour  $\alpha = 0$  (resp. 1), on retrouve un échantillon du modèle Texto-GMM associé à  $u_0$  (resp.  $u_1$ ). On constate que le modèle mélangé parvient à combiner les structures géométriques de  $u_0$  et  $u_1$ . Les paramètres sont  $(w, S) = (7, 5)$  pour la première ligne et  $(w, S) = (5, 4)$  pour la deuxième.

- [5] M. Cuturi. Sinkhorn distances : Lightspeed computation of optimal transport. In *Adv. Neural Inf. Process. Syst.*, pages 2292–2300, 2013.
- [6] J. Delon and A. Desolneux. A Wasserstein-type distance in the space of Gaussian mixture models. *SIAM Journal on Imaging Sciences*, 13(2) :936–970, 2020.
- [7] J. Feydy, P. Roussillon, A. Trounev, and P. Gori. Fast and scalable optimal transport for brain tractograms. In *Proceedings of MICCAI*, pages 636–644. Springer, 2019.
- [8] B. Galerne, A. Leclaire, and J. Rabin. A texture synthesis model based on semi-discrete optimal transport in patch space. *SIAM J. Imaging Sci.*, 11(4) :2456–2493, 2018.
- [9] A. Genevay, L. Chizat, F. Bach, M. Cuturi, and G. Peyré. Sample complexity of Sinkhorn divergences. In *Proceedings of AISTAT*, pages 1574–1583. PMLR, 2019.
- [10] T. Karras, S. Laine, and T. Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of IEEE CVPR*, pages 4401–4410, 2019.
- [11] L. Liang, C. Liu, Y.Q. Xu, B. Guo, and H.Y. Shum. Real-time texture synthesis by patch-based sampling. *ACM Trans. Graph.*, 20(3) :127–150, 2001.
- [12] J. Weed and F. Bach. Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. *Bernoulli*, 25(4A) :2620–2648, 2019.
- [13] G. Xia, S. Ferradans, G. Peyré, and J. Aujol. Synthesizing and Mixing Stationary Gaussian Texture Models. *SIAM J. Imaging Sci.*, 7(1) :476–508, 2014.
- Remerciements** Les auteurs remercient l’ANR et le GdR Isis pour leur soutien au travers des projets PostProdLEAP (ANR-19-CE23-0027-01), MISTIC (ANR-19-CE40-005), et Rémoga.