MATHÉMATIQUES
VISION
APPRENTISSAGE

Generalized Sliced Distances for Probability Distributions

Mathis Hardion

Introduction

- Optimal transport can be costly in high dimension: slicing to reduce complexity as it has a closed form in $1D$ [12]

- Studied paper [13]: generalize the slicing approach to any distance and more general slices using GRT

- Link a particular case to MMDs, derive gradient flows following [2].

MATHÉMATIQUES VISION APPRENTISSAGE

Generalized Sliced Probability Metrics (GSPMs)

Definition

ξ a distance on $\mathscr{P}(\mathbb{R})$, $\mathcal{F} = \{f_\theta \in C_b(\mathcal{X}), \theta \in \Omega\} \subset C_b(\mathcal{X})$ s.t. the GRT is invertible. (r-)GSPM:

$$\forall \mu, \nu \in \mathscr{P}(\mathbb{R}^d), \zeta_{\mathcal{F}}(\mu, \nu) := \left(\int_\Omega \xi \left(f_\theta \sharp \mu, f_\theta \sharp \nu \right)^r d\theta \right)^{\frac{1}{r}}.$$

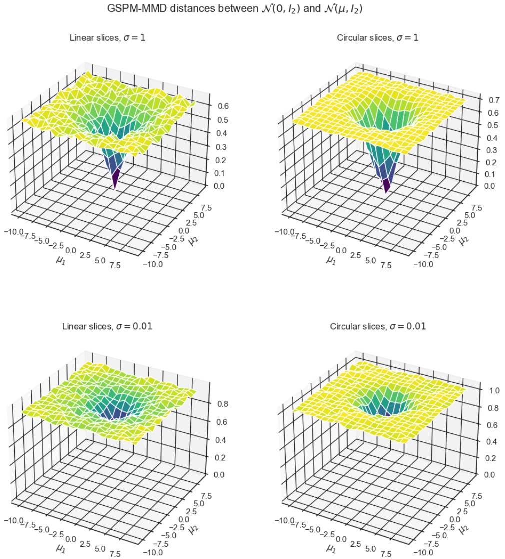
Proposition (link with MMDs)

For pdfs p, q: $\xi(p, q) := \|A(p-q)\|_2$, $\widehat{p}, \widehat{q}$ empirical densities smoothened by RBF ϕ_σ, then empirical GSPM $\leftrightarrow$ empirical MMD with kernel

$$k : (x, y) \mapsto \int_\Omega \langle A\phi_\sigma(\cdot - f_\theta(y)), A\phi_\sigma(\cdot - f_\theta(x)) \rangle \, d\theta$$

MATHÉMATIQUES VISION APPRENTISSAGE

Behaviors



MATHÉMATIQUES VISION APPRENTISSAGE

GSPM-MMD gradient flows

- Follows the work of [2] on MMD flows, noisy Euler-Maruyama scheme:

$$X_{n+1} = X_n + \eta v(X_n + \beta_n U_n, \nu_n), \tag{1}$$

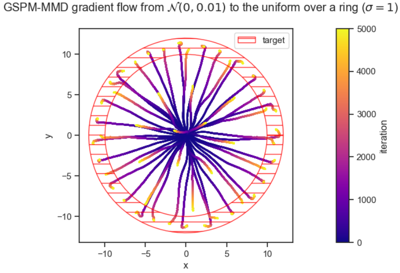
$$v(x, \nu) = -\nabla_x \left(\int k(\cdot, x) d\mu - \int k(\cdot, x) d\nu \right)$$

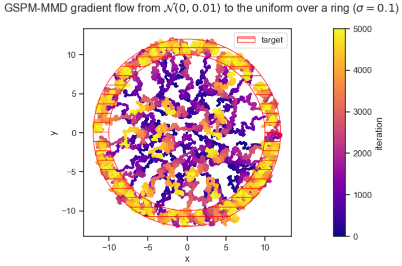
Theorem: Convergence of GSPM-MMD gradient flows

Under regularity assumptions, and if $\sum_{i=0}^{\infty} \beta_i^2 = \infty$, then (1) verifies

$$\zeta(\nu_n, \mu) \leq \zeta(\nu_n, \mu) e^{-2\lambda^2 \eta (1 - 3\eta L) \sum_{i=0}^{n} \beta_i^2}. \tag{2}$$

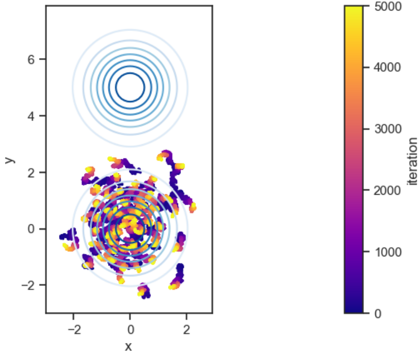
Flow numerics





Flow numerics

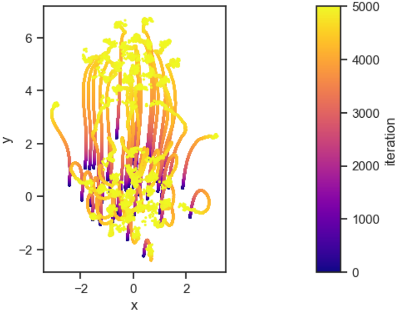
GSPM-MMD flow from one gaussian to a gaussian mixture including initial gaussian



Extensions

- Idea: modify the RBF radius σ over iterations

GSPM-MMD flow from one gaussian to a gaussian mixture including initial gaussian
(linear slices, $\sigma_n = 10 - 9.9\frac{4}{3}\mathbb{1}_{\{n \geq n_{iter}/4\}}\left(\frac{n}{n_{iter}} - \frac{1}{4}\right)$)



Criticisms

- Convergence bound becomes untractable in general settings
- Initial goal of reducing complexity somewhat forgotten: MMD with kernel that must be estimated via MC
- Hyperparameter tuning may be complex

Conclusion

- The presented method allows to lift a 1D probability distance to an arbitrary dimension, gradient flows in MMD case
- Varying RBF radius over iterations seems useful, convergence may be provable
- A link with dual norms could be investigated since many popular metrics can be understood under that framework [9]

References

[1] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Birkhäuser, 2008.

[2] Michael Arbel et al. "Maximum Mean Discrepancy Gradient Flow". In: *Advances in Neural Information Processing Systems*. Ed. by H. Wallach et al. Vol. 32. Curran Associates, Inc., 2019.

[3] Martin Arjovsky, Soumith Chintala, and Léon Bottou. *Wasserstein GAN*. 2017. arXiv: 1701.07875 [stat.ML].

[4] Nicolas Bonneel et al. "Sliced and Radon Wasserstein Barycenters of Measures". In: *Journal of Mathematical Imaging and Vision* 1.51 (2015), pp. 22–45.

[5] Emmanuel Candes. *MATH 262/CME 372: Applied Fourier Analysis and Winter 2021 - Lecture 10*. Ed. by E. Bates. Elements of Modern Signal Processing. Stanford University. Feb. 2021. URL: https://candes.su.domains/teaching/math262/Lectures/Lecture10.pdf.

[6] Gintare Karolina Dziugaite, Daniel M. Roy, and Zoubin Ghahramani. *Training generative neural networks via Maximum Mean Discrepancy optimization*. 2015. arXiv: 1505.03906 [stat.ML].

[7] Peter E. Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations*. Springer Berlin, Heidelberg, Aug. 1992.

[8] L. Ehrenpreis. "The universality of the radon transform". In: *Oxford University Press on Demand* (2003).

[9] Jean Feydy. "Geometric Data Analysis, Beyond Convolutions". PhD thesis. Université Paris-Saclay, 2020.

[10] Arthur Gretton et al. "A Kernel Method for the Two-Sample-Problem". In: *Advances in Neural Information Processing Systems*. Ed. by B. Schölkopf, J. Platt, and T. Hoffman. Vol. 19. MIT Press, 2006.

[11] Andrew Homan and Hanming Zhou. "Injectivity and Stability for a Generic Class of Generalized Radon Transforms". In: *The Journal of Geometric Analysis* 27.2 (Apr. 2017), pp. 1515–1529.

[12] Rabin Julien et al. "Wasserstein Barycenter and its Application to Texture Mixing". In: *SSVM'11*. Israel: Springer, 2011, pp. 435–446.

[13] Soheil Kolouri et al. *Generalized Sliced Distances for Probability Distributions*. 2020. arXiv: 2002.12537 [stat.ML].

[14] Soheil Kolouri et al. "Generalized Sliced Wasserstein Distances". In: *Advances in Neural Information Processing Systems*. Ed. by H. Wallach et al. Vol. 32. Curran Associates, Inc., 2019.

References

[15] Soheil Kolouri et al. “Sliced Wasserstein Auto-Encoders”. In: *International Conference on Learning Representations*. 2019.

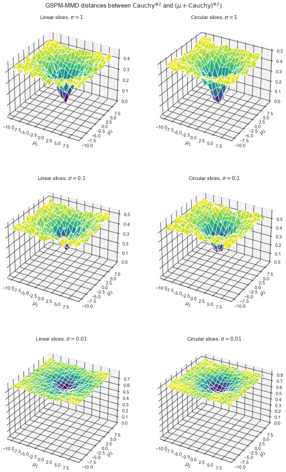
[16] Peter Kuchment. “Generalized transforms of Radon type and their applications”. In: *Proceedings of Symposia in Applied Mathematics* 63 (Jan. 2006).

[17] Antoine Liutkus et al. “Sliced-Wasserstein Flows: Nonparametric Generative Modeling via Optimal Transport and Diffusions”. In: *Proceedings of the 36th International Conference on Machine Learning*. Ed. by Kamalika Chaudhuri and Ruslan Salakhutdinov. Vol. 97. Proceedings of Machine Learning Research. PMLR, June 2019, pp. 4104–4113.

[18] G. I. Papayiannis et al. “On clustering uncertain and structured data with Wasserstein barycenters and a geodesic criterion for the number of clusters”. In: *Journal of Statistical Computation and Simulation* 91.13 (Mar. 2021), pp. 2569–2594.

Appendix

GSPM behavior on heavy tails



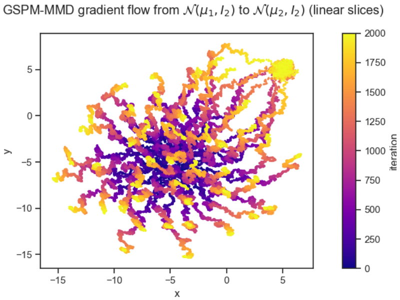
Plateau: explanation

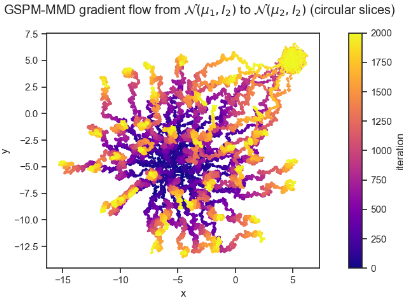
Proposition

$A = \mathrm{id}$, μ, ν with densities bounded above by respective constants B_μ, B_ν. Then,

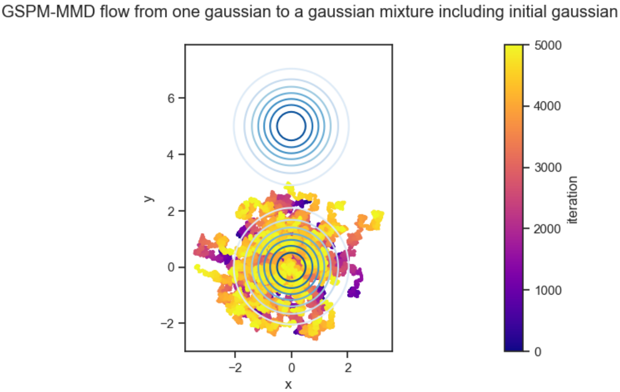
$$\zeta^2(\mu,\nu) \leq \mathrm{Leb}\,(\Omega)\left(B_\mu^2 + B_\nu^2\right). \quad (3)$$

Flow numerics



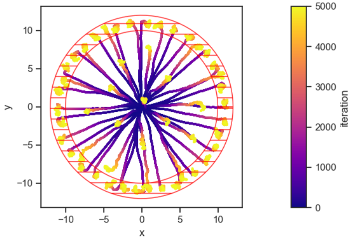


Flow numerics

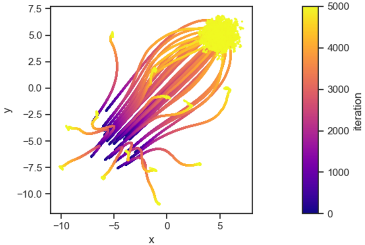


Flow numerics

GSPM-MMD gradient flow from $\mathcal{N}(0, 0.01)$ to the uniform over a ring
(linear slices, $\sigma_n = 1 - 0.9\frac{n}{n_{iter}}$)



GSPM-MMD gradient flow from $\mathcal{N}(\mu_1, 1)$ to $\mathcal{N}(\mu_2, 1)$
(linear slices, $\sigma_n = 10 - 9.9\frac{4}{3}\mathbb{1}_{\{n \geq n_{iter}/4\}}\left(\frac{n}{n_{iter}} - \frac{1}{4}\right))$



GSPM-MMD gradient flow from $\mathcal{N}(0, 0.01)$ to the uniform over a ring
(linear slices, $\sigma_n = 0.1^{\frac{n}{n_{iter}}}$)