

Sequential transport for density estimation and its applications

B. Zanger^{†,1}, O. Zahm¹, T. Cui^{2,1}, M. Schreiber^{§,1}

[†] PhD student (presenting author). [§] PhD supervisor

PhD expected duration: **Oct. 2022 – Sep. 2025**

¹ Université Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, France
 {benjamin.zanger, olivier.zahm}@inria.fr
 {martin.schreiber}@univ-grenoble-alpes.fr

² School of Mathematics and Statistics, University of Sydney, Australia
 tiangang.cui@sydney.edu.au

Abstract

Transport-based methods are receiving growing interest because of their ability to sample easily from the approximated density. These methods aim at building a deterministic diffeomorphism \mathcal{T} , also called a transport map, which pushes forward an arbitrary reference probability density ρ_{ref} to a given target probability density π to be approximated. Typically, the transport map \mathcal{T} is parameterized *e.g.* by invertible neural networks and fitted using variational methods of the form

$$\min_{\mathcal{T} \in \mathcal{M}} D(\pi \| \mathcal{T}_{\#} \rho_{\text{ref}}) \quad (1)$$

with a statistical divergence $D(\cdot \| \cdot)$, typically the (reversed) KL-divergence. An emerging strategy for this problem is to first estimate π by $\tilde{\pi}$ and then to compute a map \mathcal{T} which exactly pushes forward ρ_{ref} to $\tilde{\pi}$, see [2, 1]. Among the infinitely many maps \mathcal{T} which satisfy $\mathcal{T}_{\#} \rho_{\text{ref}} = \tilde{\pi}$, the Knothe–Rosenblatt (KR) map is rather simple to evaluate since it requires only computing the cumulative distribution functions (CDFs) of the conditional marginals of $\tilde{\pi}$. In general, problem (1) is difficult to solve when π is multimodal or when it concentrates on a low-dimensional manifold. The solution proposed in [1] consists in introducing an arbitrary sequence of bridging densities

$$\pi^{(1)}, \pi^{(2)}, \dots, \pi^{(L)} = \pi, \quad (2)$$

with increasing complexity. The sequential strategy consists in building L transport maps $\mathcal{Q}_1, \dots, \mathcal{Q}_L$ one after the other by solving

$$\min_{\mathcal{Q}_\ell \in \mathcal{M}} D(\pi^{(\ell)} \| (\mathcal{T}_{\ell-1} \circ \mathcal{Q}_\ell)_{\#} \rho_{\text{ref}}), \quad \text{where } \mathcal{T}_{\ell-1} = \mathcal{Q}_1 \circ \dots \circ \mathcal{Q}_{\ell-1}. \quad (3)$$

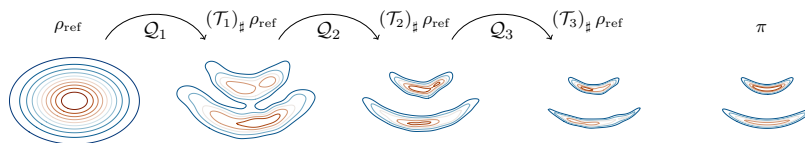


Figure 1: Visualization of the approximation of a bimodal density π (right) using $L = 3$ intermediate tempered densities estimated using SoS (4) and a Gaussian reference density ρ_{ref} .

For suitable statistical distances, so that $D(\pi||\mathcal{T}_\# \rho) = D(\mathcal{T}_\# \pi||\rho)$, these problems are equivalent to estimating the pullback density $(\mathcal{T}_{\ell-1})_\# \pi^{(\ell)}$ with an intermediate approximation $\rho^{(\ell)} = (\mathcal{Q}_\ell)_\# \rho_{\text{ref}}$.

In our work, we contribute to this methodology as follows.

First, we employ Sum-of-Squares (SoS) densities to approximate the intermediate densities $\rho^{(\ell)}$ using α -divergences $D_\alpha(\cdot||\cdot)$. We sequentially solve the variational density approximation problem as in Equation (3) with D_α as the statistical divergence and where

$$\rho^{(\ell)}(\mathbf{x}) = (\Phi(\mathbf{x})^\top A_\ell \Phi(\mathbf{x})) \rho_{\text{ref}}(\mathbf{x}), \quad (4)$$

for some arbitrary orthonormal basis function Φ in $L^2(\rho_{\text{ref}})$. Here, the positivity of the matrix $A_\ell \succeq 0$ ensures the density $\rho^{(\ell)}$ to be positive. Since the α -divergence is defined for general *unnormalized* densities, it is not necessary to know the normalizing constant of π beforehand. α -divergences $D_\alpha(\cdot||\cdot)$ with parameter $\alpha \in \mathbb{R}$ include the Hellinger distance and KL-divergence, which have been used in previous works. The proposed SoS densities permit to efficiently normalize the estimated unnormalized density and to compute the KR map \mathcal{Q}_ℓ such that $(\mathcal{Q}_\ell)_\# \rho_{\text{ref}} = \rho^{(\ell)}$. This combined use of α -divergence for performing SoS density estimation results in a *convex* optimization problems which can be efficiently solved using off-the-shelf toolboxes.

Second, we extend the methodology to the scenario where *only samples* $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(N)}$ from π are available, as opposed to *point-evaluations* of the target density π . In this scenario, we propose to use diffusion-based bridging densities $\pi^{(\ell)}(\mathbf{x})$ where the distribution follows a time-inversed diffusion process such as the Ornstein-Uhlenbeck process with time parameters $t_{\ell-1} \leq t_\ell$ and $t_L = 0$. This idea is at the root of diffusion models.

Third, we present a novel convergence analysis using the geometric properties of α -divergences. This analysis unifies and extends previous analyses proposed in [3, 1] and, more interestingly, it guides the choice of bridging densities. In particular, we show that a smart choice of β_ℓ for tempered densities or of t_ℓ for diffusion-based densities yield a convergence rate of $\mathcal{O}(1/L^2)$ with respect to the number of layer L .

Last, we give an outlook for using sequential measure transport to solve optimal transport problems, where we mitigate the difficulty of estimating the optimal coupling by a sequence of entropic regularized problems.

We demonstrate the capability of our proposed method in unsupervised learning, Bayesian inverse, and optimal transport problems in moderate dimensions.

Short biography (PhD student)

I am an electrical engineer, now pursuing a PhD in applied mathematics at the Université Grenoble Alpes while working at INRIA in the AIRSEA team. In my thesis, I work on estimation of densities using compositional methods and its applications.

References

- [1] Tiangang Cui and Sergey Dolgov. Deep composition of tensor-trains using squared inverse Rosenblatt transports. *Foundations of Computational Mathematics*, September 2021. arXiv:2007.06968 [cs, math, stat].
- [2] Sergey Dolgov, Karim Anaya-Izquierdo, Colin Fox, and Robert Scheichl. Approximation and sampling of multivariate probability distributions in the tensor train decomposition. *Statistics and Computing*, 30(3):603–625, May 2020.
- [3] Josephine Westermann and Jakob Zech. Measure transport via polynomial density surrogates, November 2023. arXiv:2311.04172 [cs, math, stat].