Study of *Optimal transport mapping via input convex neural networks*

Lucas Versini

January 16, 2025

MATHÉMATIQUES
VISION
APPRENTISSAGE

Context

μ, ν probability distributions on $\mathbb{R}^d$.

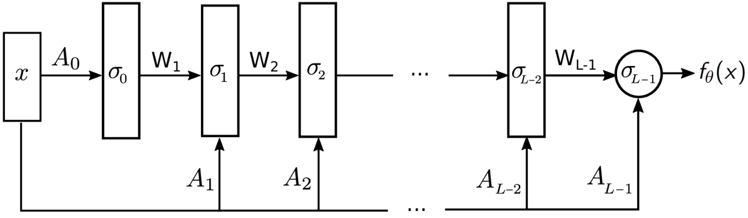
- Monge: $\min_{T \mid T_{\#}\mu = \nu} \int_{\mathbb{R}^d} c(x, T(x)) d\mu(x)$.
- Kantorovitch: $W_2^2(\mu, \nu) = \min_{\gamma \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) d\gamma(x, y)$.
- Dual:

$$W_2^2(\mu, \nu) = \sup_{\substack{f \text{ convex} \\ f^* \in L^1(\nu)}} \inf_{g \text{ convex}} \mathcal{V}_{\mu,\nu}(f, g) + C_{\mu,\nu},$$

with $\mathcal{V}_{\mu,\nu}(f, g) = -\mathbb{E}_\mu [f(X)] - \mathbb{E}_\nu [\langle Y, \nabla g(Y) \rangle - f(\nabla g(Y))]$.

Theorem: if (f, g) is optimal, and μ has a density w.r.t. Lebesgue measure, then ∇f solve Monge problem.

Input Convex Neural Networks (ICNN)



- All coefficients of the weight matrices W_l are non-negative.
- The activation functions σ_l are convex for $l = 0, \ldots, L-1$.
- The activation functions σ_l are non-decreasing for $l = 1, \ldots, L-1$.

Then $x \mapsto f_\theta(x)$ is convex.

MATHÉMATIQUES VISION APPRENTISSAGE

The problem

$$W_2^2(\mu,\nu) = \sup_{\substack{f \text{ convex} \\ f^* \in L^1(\nu)}} \inf_{g \text{ convex}} \left(-\mathbb{E}_\mu\left[f(X)\right] - \mathbb{E}_\nu\left[\langle Y, \nabla g(Y)\rangle - f\left(\nabla g(Y)\right)\right]\right) + C_{\mu,\nu}$$

is thus approximated by

$$\sup_{f \in \mathrm{ICNN}(\mathbb{R}^d)} \inf_{g \in \mathrm{ICNN}(\mathbb{R}^d)} \left(-\mathbb{E}_\mu\left[f(X)\right] - \mathbb{E}_\nu\left[\langle Y, \nabla g(Y)\rangle - f\left(\nabla g(Y)\right)\right]\right) + C_{\mu,\nu},$$

with $C_{\mu,\nu} = \dfrac{1}{2}\mathbb{E}_{X\sim\mu, Y\sim\nu}\left[\|X\|_2^2 + \|Y\|_2^2\right]$.

MATHÉMATIQUES VISION APPRENTISSAGE

Algorithm

Algorithm

Require: Samples from μ and ν, batch size M, generator iterations K, total iterations T

for $t = 1$ to T **do**
  for $k = 1$ to K **do**
    Sample $\{X_i\}_{i=1}^M \sim \mu$, $\{Y_i\}_{i=1}^M \sim \nu$
    Update g to minimize J using Adam
  end for
  Sample $\{X_i\}_{i=1}^M \sim \mu$, $\{Y_i\}_{i=1}^M \sim \nu$
  Update f to maximize J using Adam
  $w \leftarrow \max(w, 0)$ for all coefficients w of W_l in θ_f, $l = 1, \ldots, L-1$
end for

$$J(\theta_f, \theta_g) = \frac{1}{M} \sum_{i=1}^{M} \left(f\left(\nabla g(Y_i)\right) - \langle Y_i, \nabla g(Y_i)\rangle - f(X_i)\right) + \lambda \sum_{W_l \in \theta_g} \| \max(-W_l, 0)\|_F^2. \quad (1)$$

MATHÉMATIQUES VISION APPRENTISSAGE

Theoretical guarantees

$$\varepsilon_1(f,g) = \mathcal{V}_{\mu,\nu}(f,g) - \inf_{\tilde{g} \text{ convex}} \mathcal{V}_{\mu,\nu}(f,\tilde{g})$$

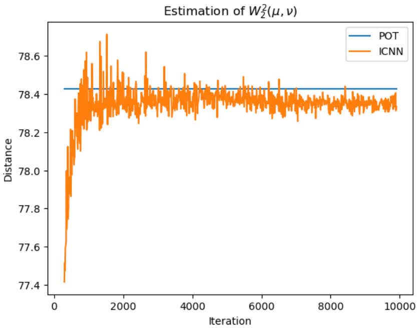
$$\varepsilon_2(f) = \sup_{\tilde{f} \text{ convex}} \inf_{\tilde{g} \text{ convex}} \mathcal{V}_{\mu,\nu}(\tilde{f},\tilde{g}) - \inf_{\tilde{g} \text{ convex}} \mathcal{V}_{\mu,\nu}(f,\tilde{g}).$$

Theorem: Assume ν has a density with respect to Lebesgue measure, and let ∇g^* be the optimal transport from ν to μ.
Then for any (f,g) such that f is α-strongly convex, the following holds:

$$\|\nabla g - \nabla g^*\|^2_{L^2(\nu)} \leq \frac{2}{\alpha}\left(\varepsilon_1(f,g) + \varepsilon_2(f)\right).$$

MATHÉMATIQUES
VISION
APPRENTISSAGE

Experiments



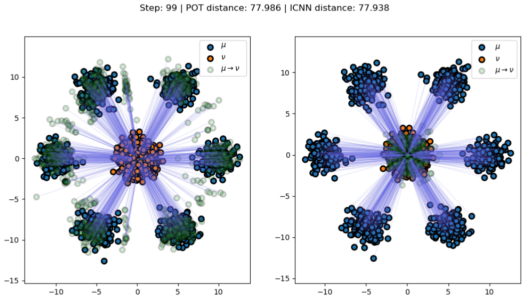
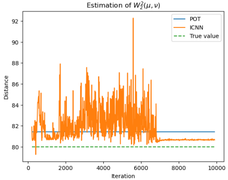


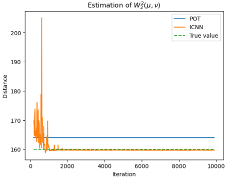
Figure: Gaussian mixture with 6 components, and Gaussian distribution. Estimated distance and optimal transport.

MATHÉMATIQUES
VISION
APPRENTISSAGE

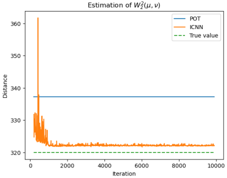
Experiments



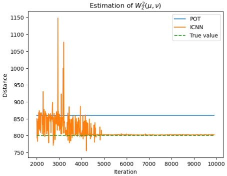
(a) Gaussian in $\mathbb{R}^5$



(b) Gaussian in $\mathbb{R}^{10}$



(c) Gaussian in $\mathbb{R}^{20}$

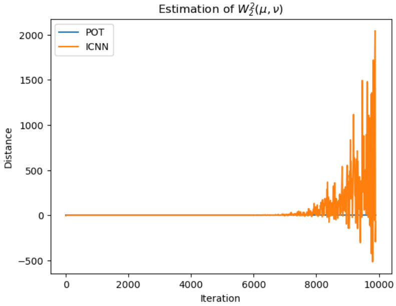


(d) Gaussian in $\mathbb{R}^{50}$

Figure: Translated Gaussian distributions in higher dimensions.

MATHÉMATIQUES VISION APPRENTISSAGE

Experiments



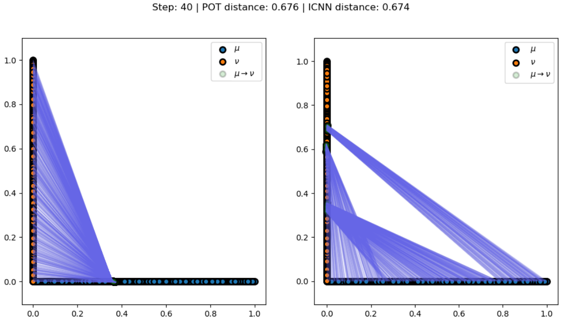
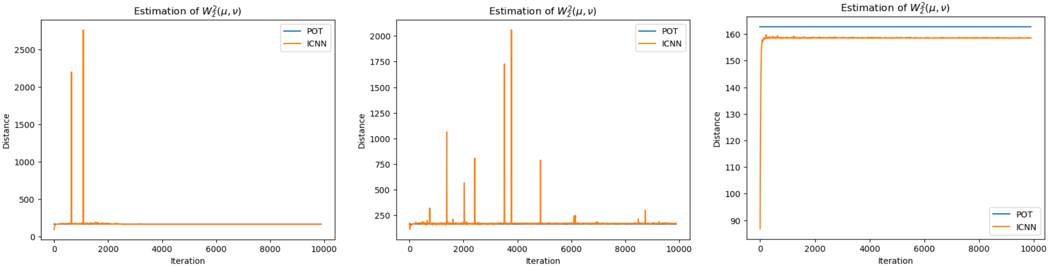


Figure: μ and ν without density (true value: $\frac{2}{3}$)

Experiments



(a) Penalty for g's coefficients (b) Positive coefficients for g (c) No constraints on g

Figure: Impact of enforcing positivity of g's coefficients

MATHÉMATIQUES VISION APPRENTISSAGE

Conclusion

- Can be slow
- Not always stable: need to carefully tune hyperparameters
- The choice of hyperparameters is not discussed in the paper

- Scalable to high dimensions and large datasets
- Easy to implement
- Some theoretical guarantees

MATHÉMATIQUES
VISION
APPRENTISSAGE

References

A. Makkuva, A. Taghvaei, S. Oh, J. Lee
Optimal transport mapping via input convex neural networks
International Conference on Machine Learning, 2020.

B. Amos, L. Xu, J. Z. Kolter
Input convex neural networks
International conference on machine learning, 2017.

R. Flamary et al.
POT: Python Optimal Transport
Journal of Machine Learning Research, 2021.

MATHÉMATIQUES
VISION
APPRENTISSAGE