Optimal transport mapping via input convex neural networks

Computational Optimal Transport Final Project

Elías Masquil

MVA
ENS Paris-Saclay

December 24, 2021

Overview

1. **Learning the OT from samples**
2. **A novel minmax optimization problem**
3. **Theoretical guarantees**
4. **Numerics**
5. **Critics and conclusions**

Solving Monge problem from samples

- (Monge) Given two probability distributions on $\mathbb{R}^d$: P and Q find the optimal mapping $T : P \to Q$ minimizing the quadratic cost
- Probabilistic interpretation of Kantorovich is natural for stochastic optimization:
  $W_2^2(P, Q) := \inf_{\pi \in \mathbb{M}_+(X,Y). \pi_1 = P, \pi_2 = Q} \frac{1}{2} \mathbb{E}_{(X,Y)\sim\pi} ||X - Y||^2$
- Dual formulation: $W_2^2(P, Q) = \sup_{(f,g)\in\Phi_c} \mathbb{E}_P[f(X)] + \mathbb{E}_Q[g(Y)]$
  $\Phi_C = \{(f, g) \in L^1(P) \times L^1(Q) : f(x) + g(y) \leq \frac{1}{2}||x - y||^2 \ \forall (x, y) \ dP \otimes dQ \ \text{a.e}\}$
- Suitable for both continuous and discrete cases

A primer on Input Convex Neural Networks

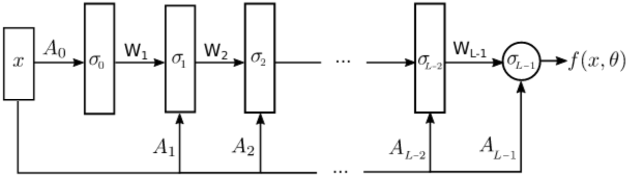


Figure: ICNN architecture (extracted from the article)

- Activations must be convex and non decreasing (first one only convex)
- $W_l \geq 0 \forall l$

A novel minmax optimization problem

Theorem (Alternative dual)

$$\sup_{f \in \mathbf{CVX}(P)} \inf_{g \in \mathbf{CVX}(Q)} -\mathbb{E}_P f(X) - \mathbb{E}_Q[< Y, \nabla g(Y) > -f(\nabla g(Y))] + \frac{1}{2}(\mathbb{E}_P ||X||^2 + E_Q||Y||^2)$$

- f and g will be approximated as ICNNs
- In the above formulation optimal g must be convex ($g^{optimal} = f^*$). That constraint is relaxed and a regularization term is added

$$R(\theta_g) = \lambda \sum ||max(-W_l, 0)||_F^2$$

Algorithm

Input: Source dist. Q, Target dist. P, Batch size M, Generator iterations K. Total iterations T. Evaluation iterations V.

Sample test batches $(X_i^{test})_i^M \sim P$ and $(Y_i^{test})_i^M \sim Q$ if continuous case, else $(X_i^{test})_i = P$ and $(Y_i^{test})_i^M = Q$

for t = 1, ..., T **do**

  Sample batches $(X_i)_i^M \sim P$ and $(Y_i)_i^M \sim Q$

  for k = 1, ..., K **do**

    Update the weights θ_g with Adam, to minimize $\frac{1}{M}\sum_i^M f(\nabla g(Y_i)) - < Y_i, \nabla g(Y_i) > + R(\theta_g)$

  Update the weights θ_f with Adam, to minimize $\frac{1}{M}\sum_i^M -f(\nabla g(Y_i)) + f(X_i)$

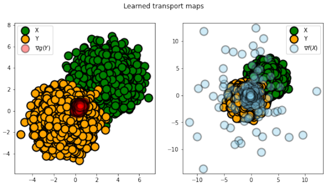
  Clip the weights W_l of θ_f to enforce the concavity constraint

  Every V iterations: evaluate the performance over X^{test} and Y^{test}

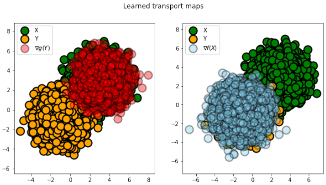
Theoretical guarantees

- Brenier's theorem + relationship between primal and dual
- Any convex function over a compact domain can be approximated in sup-norm by an ICNN to the desired accuracy
- The error between ∇g and the optimal transport map ∇g_0 is bounded with a function of the minimization and maximization gaps

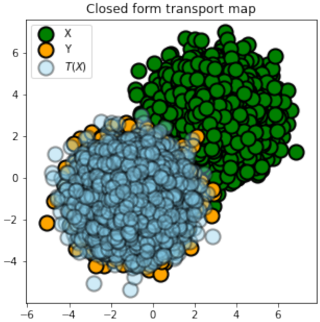
Discrete example



(a) Learned transport at iteration 500 (only first 2 components)



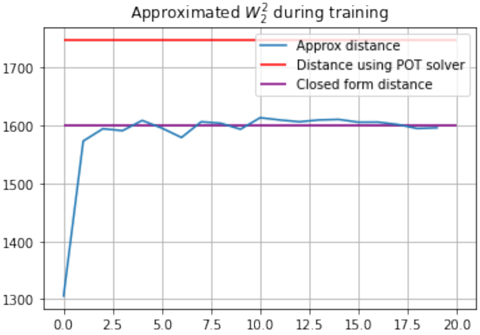
(b) Learned transport at iteration 10000 (only first 2 components)



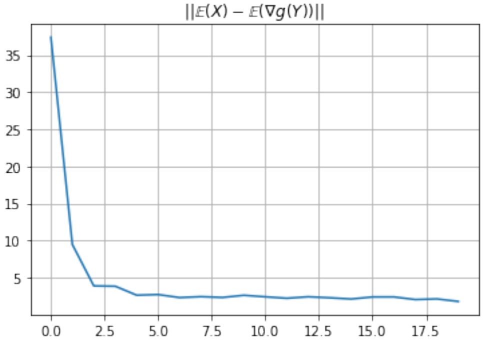
(c) Closed form optimal solution

Figure: 200-D discrete Gaussian distributions (only first 2 components)

Discrete example



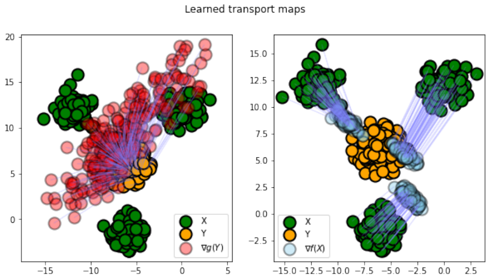
(a) W_2^2 distance



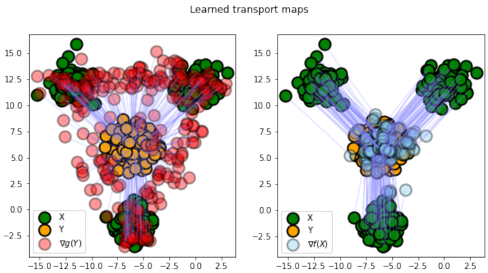
(b) Norm of the difference between the mean of X and $\nabla g(Y)$

Figure: 200-D discrete Gaussian distributions training curves

Continuous example



(a) Learned transport at iteration 500



(b) Learned transport at iteration 10000

Figure: (Continuous) Triangle of Gaussians distributions

Continuous example

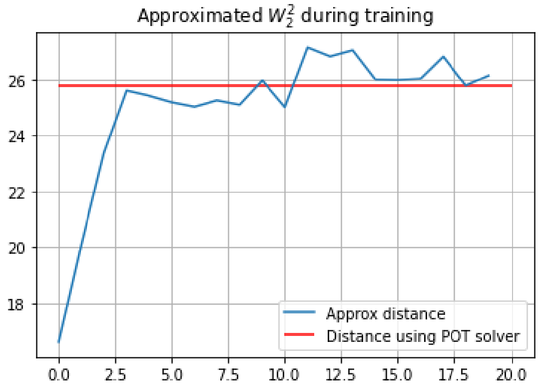


Figure: W_2^2 distance

Critics

- The current implementation is slow (e.g 1 hour for solving triangle of Gaussians). This may depend on many practical details
- Inconsistencies between proposed and actual implementation
- Difficulty to reproduce some results (discontinuity of the maps)

Conclusions and perspectives

- A novel method for finding the OT between samples is introduced. It converges to the accurate solution in many examples.
- Scaling the method is an open challenge
- Implementation details should be analysed
- Can be used for generative models and has some theoretical guarantees

References

Ashok Vardhan Makkuva and Amirhossein Taghvaei and Sewoong Oh and Jason D. Lee (2020)
Optimal transport mapping via input convex neural networks
Proceedings of the 37th International Conference on Machine Learning

Brandon Amos and Lei Xu and J. Zico Kolter (2016)
Input Convex Neural Networks
Proceedings of the 34th International Conference on Machine Learning

Yize Chen and Yuanyuan Shi and Baosen Zhang (2019)
Optimal Control Via Neural Networks: A Convex Approach
Proceedings of the 7th International Conference on Learning Representations (ICLR 2019).

The End