Entropy-Regularized Optimal Transport Generative Models

Dong Liu, Minh Thành Vu, Saikat Chatterjee, and Lars K. Rasmussen

Division of Information Science and Engineering
KTH Royal Institute of Technology, Stockholm, SE-100 44, Sweden.
E-mail: {doli, mtvu, sach, lkra}@kth.se

Problem Essence

Probabilistic Modeling: use a probabilistic model Q to estimate an unknown distribution P. This can be quantified by a distance measure $\mathcal{D}$, such that:

$$\mathcal{D}(P, Q) \to 0, Q \to P$$

Motivations

Cases where KL (Kullback-Leibler) or JS (Jensen-Shannon) divergence may be problematic:

- No explicit density function of probability model
- Support mismatch
- In high-dimensional space, KL and JS are sensitive to perturbations.

What else?

Benefit of employing OT distance:

- Applicable to both implicit and explicit models
- Regardless of match or mismatch of supports
- Bound up for distribution perturbation

Advantages of entropy regulation:

- Smoothed solution
- Avoid Lipschitz constraint to enforce (Kantorovich-Rubinstein duality solution, e.g. WGAN).

Preliminary

- Work space, $(\mathcal{X}, \|\cdot\|_2), \mathcal{X} \subset \mathbb{R}^d$
- Distribution P with finite support $\mathcal{X}_1 \subset \mathcal{X}$
- Distribution Q with finite support $\mathcal{X}_2 \subset \mathcal{X}$

Modeling

$$\underset{g:\mathcal{Z}\to\mathcal{X}}{\text{argmin}}\ W(P, Q) = \underset{\theta}{\text{argmin}}\ W(P, Q)$$

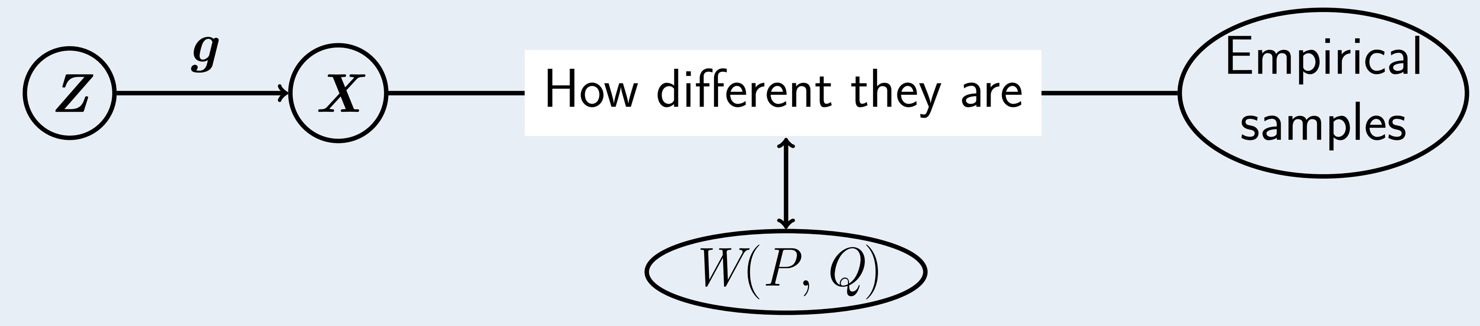
with

$$W(P, Q) = \min_{\pi \in \Pi(P,Q)} \langle \pi, M \rangle - \lambda H(\pi),$$

- Cost matrix $[M]_{i,j} = d(x^{(i)}, y^{(j)}) = \|x^{(i)} - y^{(j)}\|_2^2$
- Optimization domain $\Pi(P, Q) = \{\pi : \pi\mathbb{1} = P, \pi^T\mathbb{1} = Q\}$
- $H(\pi) = \sum_{i,j} -\pi_{i,j} \log(\pi_{i,j})$

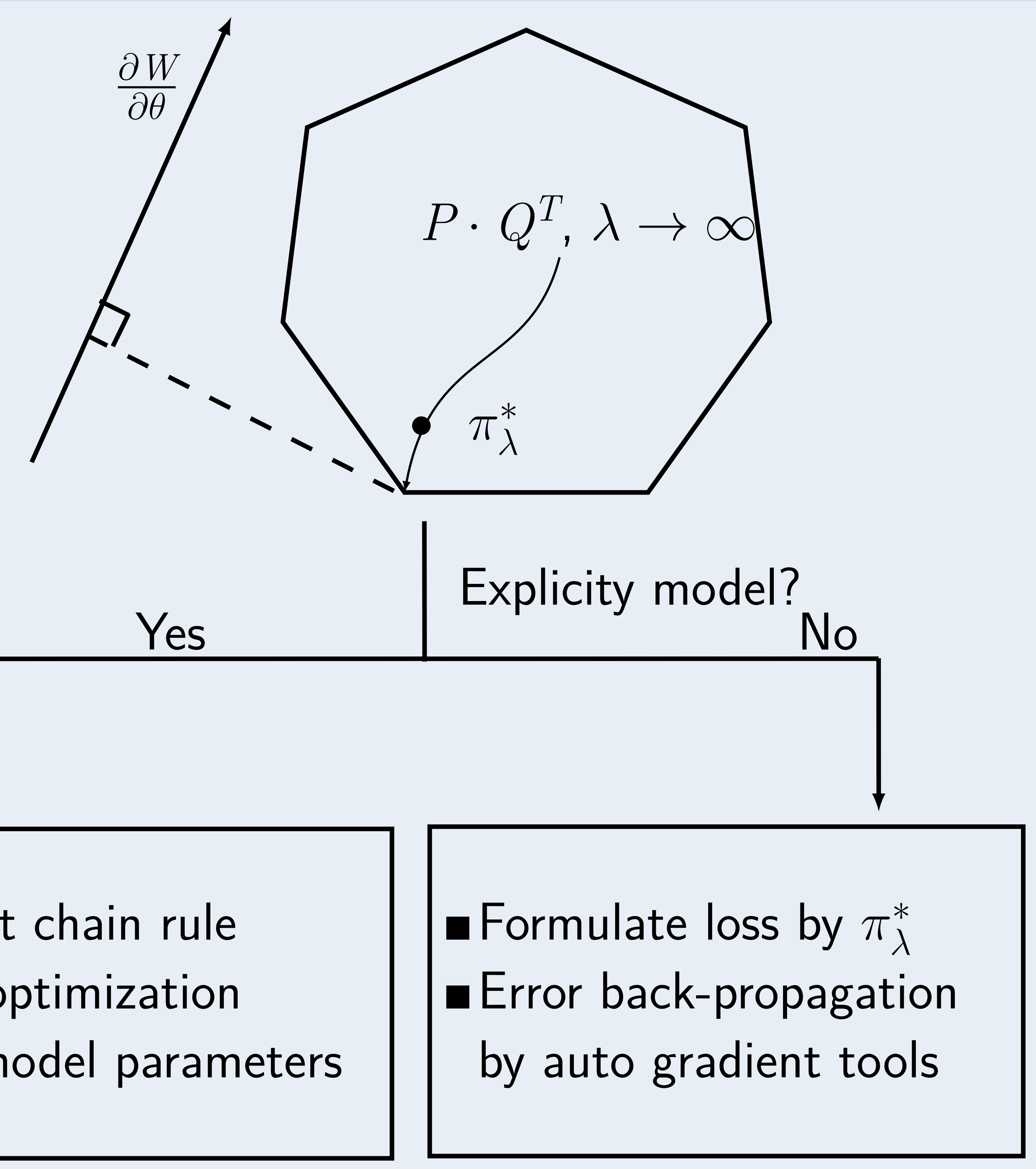
How does it work?

Problem intuition: model learning by sample comparison.



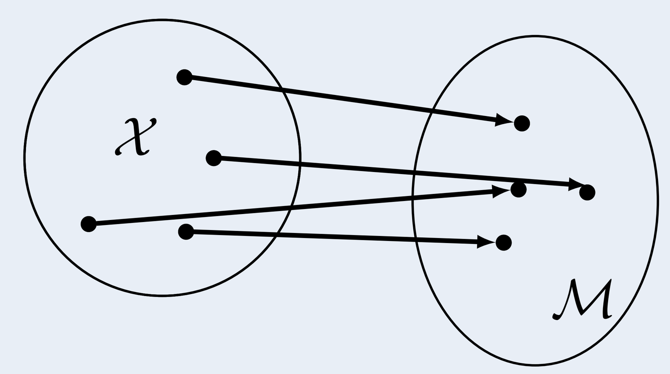
Signal Flow for EOT Generative Model (EOTGM)

Alternative Ways to Use EOTGM



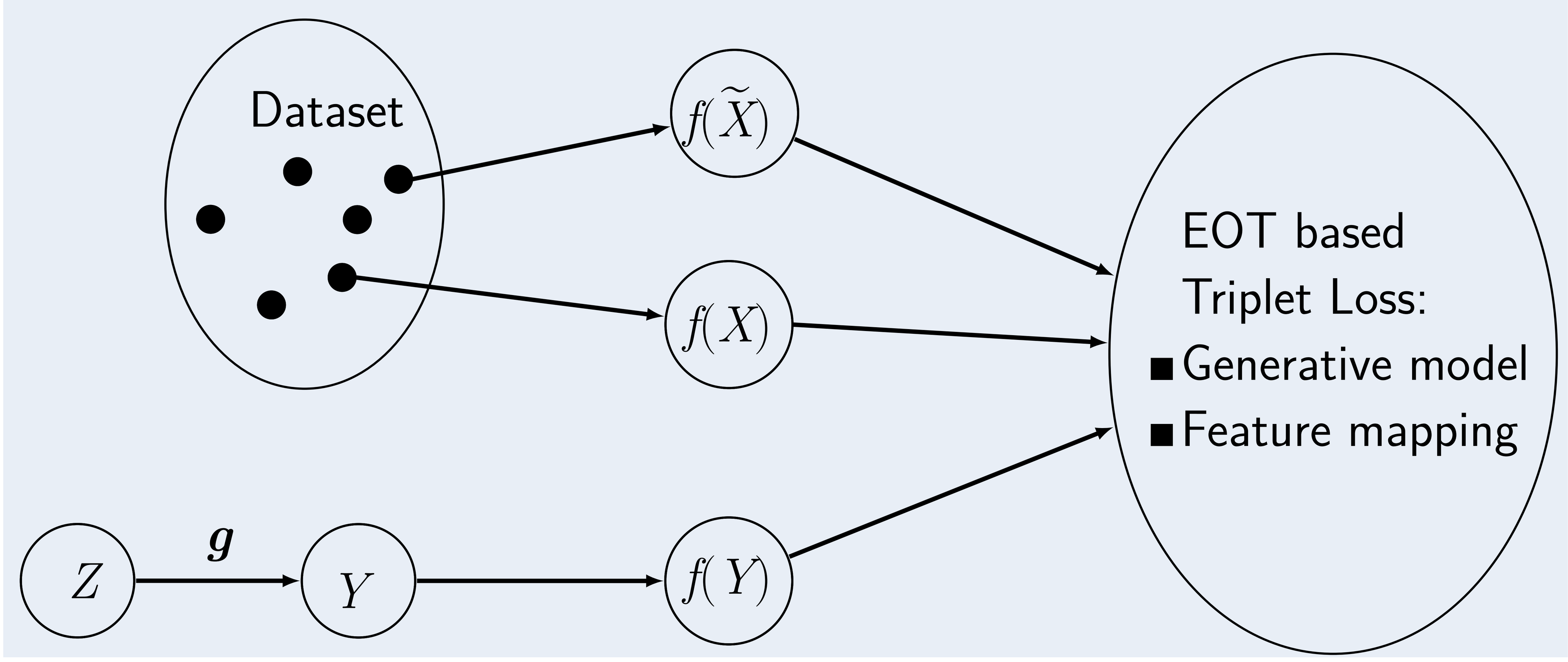
Learning in feature space

- Euclidean distance is not good for multimedia signal comparison
- Do the model learning in feature space $\mathcal{M}$

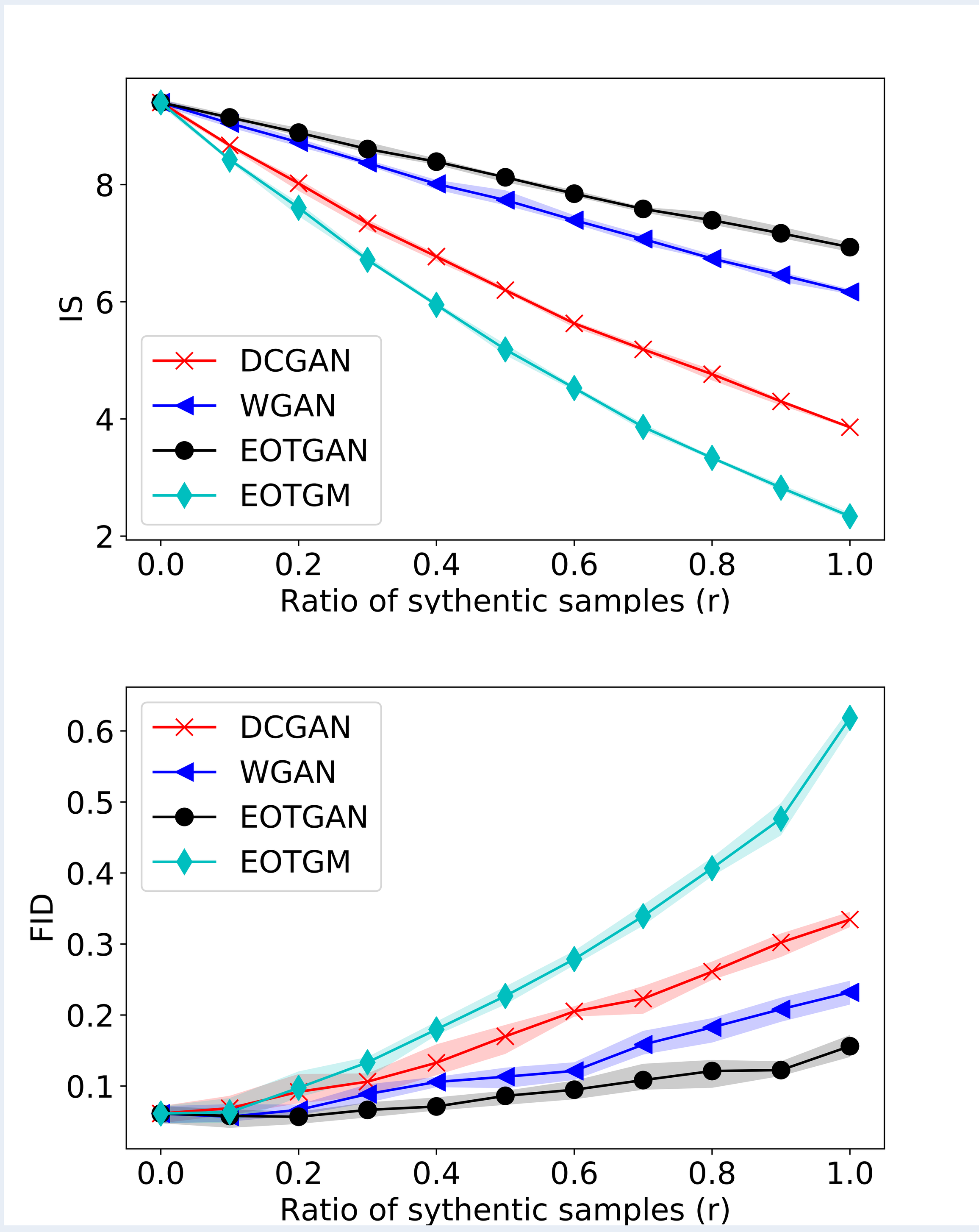


Pushforward mapping, $f : \mathcal{X} \to \mathcal{M}$, and formulate cost matrix in $\mathcal{M}$.

Optimization of g and f: EOTGAN



Numerical Results



Comparison of IS and FID (on MNIST) versus mixing ratio r.

Reference

- M. Cuturi "Sinkhorn distances: Lightspeed computation of optimal transport," in NIPS, 2013.
- M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in ICML, 2017.