Energy-guided Entropic Neural Optimal Transport

Petr Mokrov[1] Alexander Korotin[1,2] Alexander Kolesov[1]
Nikita Gushchin[1] Evgeny Burnaev[1,2]

[1]Skolkovo Institute of Science and Technology
[2]Artificial Intelligence Research Institute

2024

Entropic Optimal Transport. Theoretical formulation[1]

Let $\mathbb{P}$ and $\mathbb{Q}$ be continuous distributions on $\mathcal{X} \subset \mathbb{R}^{D_x}$ and $\mathcal{Y} \subset \mathbb{R}^{D_y}$. **Entropic Optimal Transport (EOT)** is as follows:

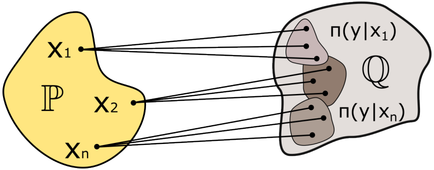
$$\text{EOT}_{c,\epsilon}(\mathbb{P},\mathbb{Q}) = \min_{\pi \in \Pi(\mathbb{P},\mathbb{Q})} \int_{\mathcal{X}} C\big(x, \pi(\cdot|x)\big)\, d\mathbb{P}(x)$$

The minimizer π^* is called the Entropic OT plan.

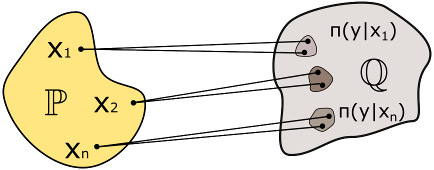
$$C\big(x, \pi(\cdot|x)\big) \overset{def}{=} \int_{\mathcal{Y}} \underbrace{c(x,y)}_{\text{Dissimilarity}} d\pi(y|x) - \epsilon \underbrace{H(\pi(\cdot|x))}_{\text{Diversity}}.$$

Regularization strength ϵ controls the diversity.

- $\Pi(\mathbb{P},\mathbb{Q})$ are distributions on $\mathcal{X} \times \mathcal{Y}$ with marginals $\mathbb{P}$, $\mathbb{Q}$



Stochastic EOT maps for large ϵ.



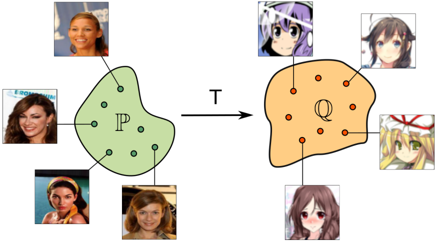
Stochastic EOT maps for small ϵ.

[1] Marco Cuturi (2013). "Sinkhorn distances: Lightspeed computation of optimal transport". In: *Advances in neural information processing systems* 26.

Entropic Optimal Transport. Applications

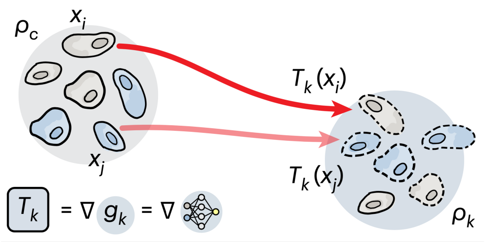
Domain Translation.[2]

By considering two unpaired image datasets as samples from $\mathbb{P}$ and $\mathbb{Q}$, EOT learns a map between datasets that preserves content.



Single-Cell (SC) Biological data.[3]

SC technology determines the gene expression profile of each measured cell, but destroys all measured cells. EOT learns map between cell populations before and after perturbation.



[2] Alexander Korotin, Daniil Selikhanovych, and Evgeny Burnaev (2023). "Neural Optimal Transport". In: *International Conference on Learning Representations*.

[3] Charlotte Bunne et al. (2023). "Learning single-cell perturbation responses using neural optimal transport". In: *Nature Methods*, pp. 1–10.

Solving EOT problem: Lagrangian multipliers → Max-min

♦ Original **EOT** problem:

$$\pi^* = \arg\min_{\pi\in\Pi(\mathbb{P},\mathbb{Q})} \left\{ \int_{\mathcal{X}\times\mathcal{Y}} c(x,y) \underbrace{d\pi(x,y)}_{=d\pi(y|x)d\mathbb{P}(x)} - \epsilon \int_{\mathcal{X}} H(\pi(\cdot|x))d\mathbb{P}(x) \right\}$$

♦ max-min (**dual**) EOT problem, the second marginal distribution constraint is relaxed with **Lagrangian multiplier** $f:\mathcal{Y}\to\mathbb{R}$ (continuous functions):

$$(f^*,\pi^*) = \arg\sup_{f}\min_{\pi\in\Pi(\mathbb{P})} \left\{ \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)d\pi(x,y) - \epsilon\int_{\mathcal{X}} H(\pi(\cdot|x))d\mathbb{P}(x) - \int_{\mathcal{X}\times\mathcal{Y}} f(y)d\pi(x,y) + \int_{\mathcal{Y}} f(y)d\mathbb{Q}(y) \right\}$$

Remark (max-min **reformulation of the EOT problem)**

Original and max-min EOT problems from above are equivalent, i.e., they yield the same optimal value $EOT_{c,\epsilon}$ *and permit the same minimizer* $\pi^* \in \Pi(\mathbb{P},\mathbb{Q})$.

Solving EOT problem: Max-min → Energy-based EOT

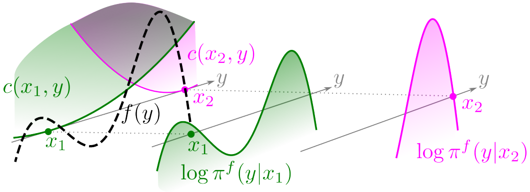
♦ max-min **EOT** problem:

$$\sup_f \underbrace{\min_{\pi\in\Pi(\mathbb{P})}\left\{\int_{\mathcal{X}\times\mathcal{Y}} c(x,y)d\pi(x,y) - \epsilon\int_{\mathcal{X}} H(\pi(\cdot|x))d\mathbb{P}(x) - \int_{\mathcal{X}\times\mathcal{Y}} f(y)d\pi(x,y)}_{\text{has a closed-form solution } \pi^f\in\Pi(\mathbb{P})} + \int_{\mathcal{Y}} f(y)d\mathbb{Q}(y)\right\}$$

Conditional plans $\pi^f(\cdot|x)$ have **known unnormalized density**:

$$\pi^f(y|x) = \frac{1}{Z(f,x)}\exp\left(\frac{f(y)-c(x,y)}{\epsilon}\right)$$

- $Z(f,x) \stackrel{\text{def}}{=} \int_{\mathcal{Y}} \exp\left(\frac{f(y)-c(x,y)}{\epsilon}\right)dy$ is the **Normalization** const.



♦ Energy-based EOT (after substitution π^f):

$$f^* = \arg\sup_f\left\{-\epsilon\int_{\mathcal{X}} \log Z(f,x)d\mathbb{P}(x) + \int_{\mathcal{Y}} f(y)d\mathbb{Q}(y)\right\}, \quad \pi^* = \pi^{f^*}.$$

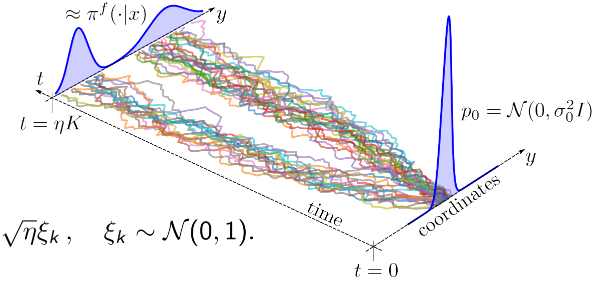
Energy-based EOT: Dealing with π^f

$$d\pi^f(x,y) = d\pi^f(y|x)d\mathbb{P}(x)\,; \quad \pi^f(y|x) = \frac{1}{Z(f,x)}\exp\left(\frac{f(y)-c(x,y)}{\epsilon}\right).$$

Sampling from π^f.

1. Sample $x \sim \mathbb{P}$ (source distribution/dataset).
2. Sample $y \sim \pi^f(\cdot|x)$ (*see below*).

Sampling from $\pi^f(\cdot|x)$. **Euler-Maruyama** simulation (a.k.a. Langevin sampling):



- Take prior p_0 (e.g., Gaussian or Uniform); Sample $y_0 \sim p_0$.
- Fix $\eta > 0$ and for $k = 0, 1, \dots K$ do:

$$y_{k+1} = y_k - \frac{\eta}{2}\nabla_y\left(\frac{c(x,y)-f(y)}{\epsilon}\right)\Bigg|_{y=y_k} + \sqrt{\eta}\xi_k\,, \quad \xi_k \sim \mathcal{N}(0,1).$$

- Output $y = y_K$. It is (approximately) sampled from $\pi^f(\cdot|x)$.

Energy-based EOT: Training aspects

- Energy-based EOT objective w.r.t. (scalar-valued) Neural Network f_θ:

$$\mathcal{L}(\theta) = -\epsilon \int_{\mathcal{X}} \log Z(f_\theta, x) d\mathbb{P}(x) + \int_{\mathcal{Y}} f_\theta(y) d\mathbb{Q}(y) \longrightarrow \max_\theta .$$

Theorem (Gradient of Energy-based EOT loss $\mathcal{L}(\theta)$)

$$\frac{\partial}{\partial\theta}\mathcal{L}(\theta) = -\int_{\mathcal{X}\times\mathcal{Y}} \left[\frac{\partial}{\partial\theta} f_\theta(y)\right] d\pi^{f_\theta}(x, y) + \int_{\mathcal{Y}} \left[\frac{\partial}{\partial\theta} f_\theta(y)\right] d\mathbb{Q}(y).$$

Proxy loss for Energy-based EOT:

$$L(\theta) = -\int_{\mathcal{X}\times\mathcal{Y}} f_\theta(y) d\pi^{f_{\text{stop\_grad}(\theta)}}(x, y) + \int_{\mathcal{Y}} f_\theta(y) d\mathbb{Q}(y)$$

[Sample estimate] $\widehat{L}(\theta)$ = [Langevin Sampling from π^{f_θ}; Monte-Carlo] + [Monte-Carlo]

Loss $L(\theta)$ is **maximized** with help of a stochastic gradient ascent procedure.

Training Energy-based EOT: Ultimate Algorithm

Algorithm 1: Entropic OT via Energy-Based Modelling

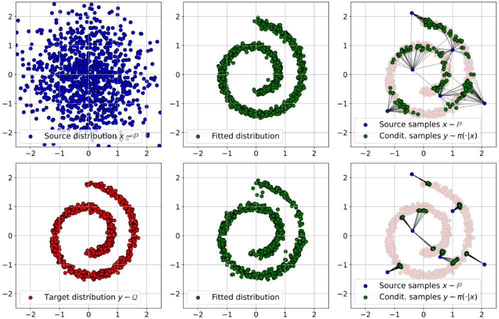
Input : Source and target distributions $\mathbb{P}$ and $\mathbb{Q}$, accessible by samples;
Entropy regularization coefficient $\epsilon > 0$, cost function $c(x, y) : \mathbb{R}^{D_x} \times \mathbb{R}^{D_y} \to \mathbb{R}$;
number of Langevin steps $K > 0$, Langevin discretization step size $\eta > 0$;
basic noise std $\sigma_0 > 0$; potential network $f_\theta : \mathbb{R}^{D_y} \to \mathbb{R}$, batch size $N > 0$.

Output: trained potential network f_{θ^*} recovering optimal conditional EOT plans

for $i = 1, 2, \dots$ **do**

- Derive batches $\{x_n\}_{n=1}^N = X \sim \mathbb{P}$, $\{y_n\}_{n=1}^N = Y \sim \mathbb{Q}$ of sizes N;
- Sample basic noise $Y^{(0)} \sim \mathcal{N}(0, \sigma_0)$ of size N;
- **for** $k = 1, 2, \dots, K$ **do**
  - Sample $Z^{(k)} = \{z_n^{(k)}\}_{n=1}^N$, where $z_n^{(k)} \sim \mathcal{N}(0, 1)$;
  - Obtain $Y^{(k)} = \{y_n^{(k)}\}_{n=1}^N$ with Langevin steps:
  - $$y_n^{(k)} \leftarrow y_n^{(k-1)} + \frac{\eta}{2\epsilon} \cdot \texttt{stop\_grad}\left(\frac{\partial}{\partial y} \left[f_\theta(y) - c(x_n, y) \right] \Big|_{y = y_n^{(k-1)}} \right) + \sqrt{\eta} z_n^{(k)}$$
- $$\widehat{L} \leftarrow -\frac{1}{N} \left[\sum_{y_n^{(K)} \in Y^{(K)}} f_\theta\left(y_n^{(K)}\right) \right] + \frac{1}{N} \left[\sum_{y_n \in Y} f_\theta(y_n) \right];$$
- Perform a gradient step over θ by using $\frac{\partial \widehat{L}}{\partial \theta}$;

Energy-guided EOT: Experiments

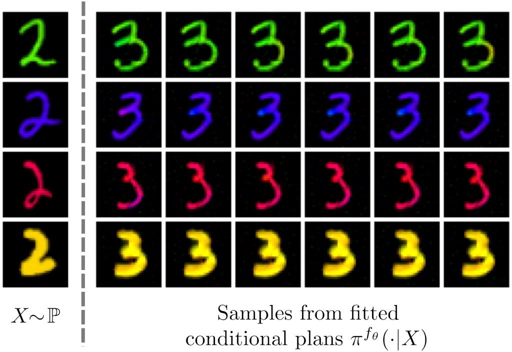


(a) Input and target distributions $\mathbb{P}$ and $\mathbb{Q}$.

(b) **Our** fitted distributions; $\varepsilon = 10^{-1}$ (up), $\varepsilon = 10^{-3}$ (down).

(c) **Our** fitted conditional plans $\pi(\cdot|x)$; $\varepsilon = 10^{-1}$ (up), $\varepsilon = 10^{-3}$ (down)

2D Gaussian to Swissroll translation with different Entropic regularizations.



Unpaired image to image translation on Colored MNIST.

Energy-guided EOT: Latent-space AFHQ'512 experiment

Unpaired I2I by latent-space **Energy-guided EOT**

source 512×512 images

generated 512×512 images by **Energy-guided EOT** working in the latent space of StyleGAN

Thank you!

Energy-guided Entropic Neural Optimal Transport

Towards the Unification of Entropic Optimal Transport and Energy-based Modeling.

GitHub

https://github.com/PetrMokrov/Energy-guided-Entropic-OT

OpenReview.net

https://openreview.net/forum?id=d6tUsZeVs7

ICLR

https://iclr.cc/virtual/2024/poster/18274