Collisional Multi-Marginal Optimal Transport for Generative AI

Mohsen Sadr[1] and M. Hossein Gorji[2]

[1]Department of Mechanical Engineering, MIT, Cambridge, MA 02139, USA
[2]Laboratory for Computational Engineering, Empa, Dübendorf, Switzerland

Massachusetts Institute of Technology

Empa Materials Science and Technology

MMOT Problem

Let $\mathcal{P}(\mathcal{X}_i)$ be the space of non-negative Borel measures over $\mathcal{X}_i \subset \mathbb{R}^n$, and

$$\mathcal{P}_2(\mathcal{X}_i) := \left\{ \mu \in \mathcal{P}(\mathcal{X}_i) \middle| \int_{\mathcal{X}_i} \|x\|_2^2 \mu(dx) < \infty \right\} \tag{1}$$

with $\|.\|_2$ the L^2-Euclidean norm. Consider K probability measures $\mu_i \in \mathcal{P}_2(\mathcal{X}_i)$ with $i \in \{1, ..., K\}$.

We are interested in the Multi-Marginal Optimal Transport problem (MMOT), which seeks the minimization

$$\pi_{\text{opt}} := \underset{\pi \in \Pi(\mu_1,...,\mu_K)}{\arg\min} \int_{\mathcal{X}} c(x_1, ..., x_K)\pi(dx) \ , \tag{2}$$

where $\mathcal{X}$ is the product set $\mathcal{X} := \mathcal{X}_1 \times ... \times \mathcal{X}_K$.

The MMOT optimization is constrained on Π, i.e., the set of coupling measures

$$\Pi(\mu_1, ..., \mu_K) := \left\{ \pi \in \mathcal{P}_2(\mathcal{X}) \middle| \text{proj}_i(\pi) = \mu_i \ \forall i \in \{1, ..., K\} \right\} \tag{3}$$

and $\text{proj}_i : \mathcal{X} \to \mathcal{X}_i$ is the canonical projection. For simplicity, here we consider

$$c(x_1, ..., x_K) = \sum_{i=1}^{K} \sum_{j=i+1}^{K} \frac{1}{2} \|x_i - x_j\|_2^2 \tag{4}$$

as the cost function.

Main Idea

Swapping Algorithm.
For each marginal $i \in \{1, ..., K\}$ and samples $j, k \in \{1, ..., N_p\}$ with $k \geq j$, Iterated Swapping Algorithm (ISA) updates the samples via

$$(X_{j,t+1}^{(i)}, X_{k,t+1}^{(i)})^T = \mathcal{K}_{j,k}(X_{j,t}^{(i)}, X_{k,t}^{(i)})^T. \tag{5}$$

The swaps are guided by the discrete cost

$$m(\tilde{\pi}_t) = \mathbb{E}_{\tilde{\pi}_t}[c] \tag{6}$$

where $\tilde{\pi}_t$ is the empirical measure of X_t. The swapping kernel is given by

$$\mathcal{K}_{j,k} = \begin{cases} I_{2n\times 2n} & \text{if } m(\tilde{\pi}_t^{X_j^{(i)} \leftrightarrow X_k^{(i)}}) \geq m(\tilde{\pi}_t) \\ J_{2n\times 2n} & \text{if } m(\tilde{\pi}_t^{X_j^{(i)} \leftrightarrow X_k^{(i)}}) < m(\tilde{\pi}_t) \end{cases} \tag{7}$$

with $I_{n\times n}$ as the identity matrix and J an exchange matrix of the form

$$J_{2n\times 2n} = \begin{bmatrix} 0_{n\times n} & I_{n\times n} \\ I_{n\times n} & 0_{n\times n} \end{bmatrix} \tag{8}$$

and $0_{n\times n}$ is a $n \times n$ matrix with zero entries.

Collision-based OT.
We propose a collision process that evolves an initial joint measure of $\{\mu_1, \mu_2\}$ in a fashion similar to binary collisions, where collisions refer to swapping the state of two particles.

Let ρ_t be the time dependent density of the joint measure. An equivalent collision operator of the Boltzmann-type can be described as

$$Q[\rho_t, \rho_t] = \int_{\mathbb{R}^{2n}} \rho_t(x_1, y)\rho_t(x, y_1)\Omega(x, x_1, y, y_1)dx_1 dy_1 - \alpha(x, y)\rho_t(x, y), \tag{9}$$

$$\alpha(x, y) = \int_{\mathbb{R}^{2n}} \rho_t(x_1, y_1)\Omega(x, x_1, y, y_1)dx_1 dy_1 \tag{10}$$

and the collision kernel reads

$$\Omega(x, x_1, y, y_1) = H\Big(c(x, y) + c(x_1, y_1) - c(x_1, y) - c(x, y_1)\Big) \tag{11}$$

where $H(.)$ is the Heaviside function.

Heuristically, the kinetic model (9) describes a process where binary collisions are only accepted if the cost c is decreased by the swaps between the two randomly picked sample points.

Exponential Convergence

Let us consider the Cauchy problem

$$\frac{\partial \rho}{\partial t} = P[\rho, \rho] - \hat{\alpha}\rho \tag{12}$$

where $P[\rho, \rho]$ is a bilinear operator, and $\hat{\alpha} \neq 0$ is a constant. The solution to the Cauchy problem can be written as

$$\rho = e^{-\hat{\alpha}t} \sum_{k=0}^{\infty} (1 - e^{-\hat{\alpha}t})^k \rho_k \tag{13}$$

where ρ_k is given by the recurrence formula

$$\rho_k = \frac{1}{k+1} \sum_{h=0}^{k} \frac{1}{\hat{\alpha}} P[\rho_h, \rho_{k-h}]. \tag{14}$$

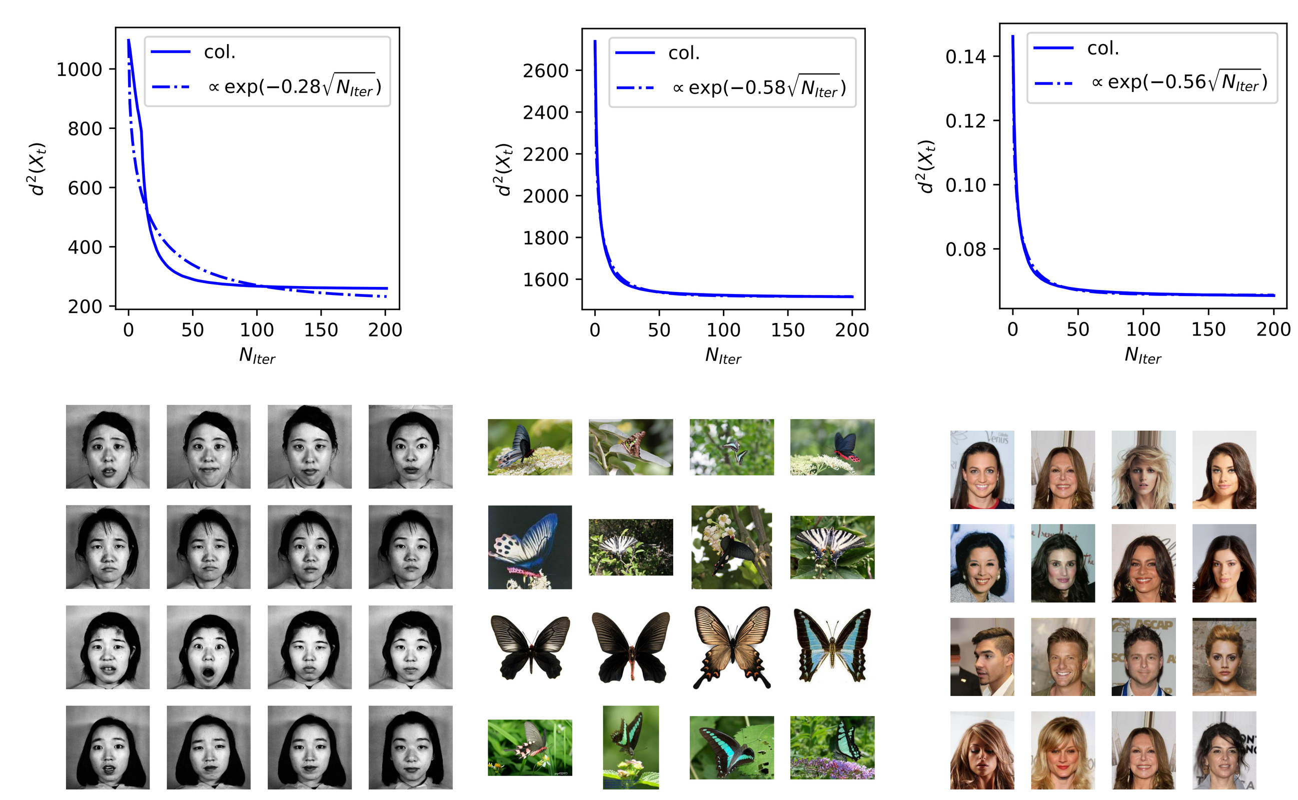
By defining $P[\rho, \rho] := Q[\rho, \rho] + \hat{\alpha}\rho$, formally we have $\lim_{k\to\infty} \rho_k = \lim_{t\to\infty} \rho = \rho^*$, where ρ^* is the equilibrium solution to the Boltzmann equation, i.e. the target sub-optimal joint density in this context. For a given ϵ and $t > t_0$, there exists finite n_0 and K where the Wild expansion is bounded $F^{P_r(n)}(x) < K$, such that

$$|\rho - \rho^*| < K n_0 e^{-\hat{\alpha}t_0} + \frac{2}{3}\epsilon e^{-\hat{\alpha}t} \sum_{n=1}^{\infty} (1 - e^{-\hat{\alpha}t})^{n-1} \ . \tag{15}$$

Randomized Swapping Algorithm

Input: $X := [X^{(1)}, ..., X^{(K)}]$ and tolerance $\hat{\epsilon}$
repeat
  for $i = 1, \ldots, K$ **do**
    Generate an even random list of particle indices R.
    Decompose R into same-size subsets I and J where $I \cap J = \varnothing$ and $|I| = |J| = \lfloor N_p/2 \rfloor$.
    for $k = 1, \ldots, \lfloor N_p/2 \rfloor$ **do**
      if $m(\tilde{\pi}_t^{X_{I_k}^{(i)} \leftrightarrow X_{J_k}^{(i)}}) < m(\tilde{\pi}_t)$ **then**
        $X_{I_k}^{(i)} \leftarrow X_{J_k}^{(i)}$ and $X_{J_k}^{(i)} \leftarrow X_{I_k}^{(i)}$.
      end if
    end for
  end for
until Convergence in $\mathbb{E}_{\tilde{\pi}_t}[c(X_t^{(1)}, ..., X_t^{(K)})]$ with tolerance $\hat{\epsilon}$
Output: X

Wasserstein Distance in Datasets



Training Diffusion Model (U-Net)

DDPM Collisional OT Loss

Preprint at arXiv:2412.16385.