VẬN CHUYỂN TỐI ƯU MỘT PHẦN ĐA BIÊN: KHẮC PHỤC CÁC CHIẾN LƯỢC MỞ RỘNG BẤT KHẢ THI VÀ CÁC PHƯƠNG PHÁP PRIMAL-DUAL HIỆU QUẢ

ON MULTI-MARGINAL PARTIAL OPTIMAL TRANSPORT: RECTIFYING INFEASIBLE EXTENSION STRATEGIES AND EFFICIENT PRIMAL-DUAL METHODS

Nguyễn Đình Thiên Quang - 23521285

Hoàng Đức Dũng - 23520328

UIT-HCM

Tóm tắt

- Lớp: CS519.Q11.KHTN
- Link Github của nhóm: https://github.com/Binbonk5/CS519.Q11.KHTN
- Link YouTube video: https://www.youtube.com/watch?v=oJYt4nf3nJU

Hoàng Đức Dũng

Nguyễn Đình Thiên Quang

Giới thiệu

UIT-HCM

1. Bài toán Optimal Transport (OT):

- Là bài toán tìm "kế hoạch vận chuyển" tối ưu nhất để dời toàn bộ khối lượng từ **Phân phối Nguồn** sang lấp đầy **Phân phối Đích**.
- *Mục tiêu:* Tối thiểu hóa tổng chi phí (Tổng khối lượng x Chi phí di chuyển).
- *Ví dụ kinh điển:* Bài toán dời đất (Earth Mover's Distance) - chuyển đất từ đống này sang hố kia sao cho tốn ít công sức nhất.

2. Lịch sử phát triển

- **OT Truyền thống [1942]:** "Tiêu chuẩn vàng" để so sánh phân phối. Dùng để so sánh ***sự khác biệt*** giữa các phân phối xác suất.
  - Đòi hỏi bảo toàn khối lượng tuyệt đối (tổng khối lượng nguồn bằng đích) và thất bại trước dữ liệu có nhiễu (outliers).
- **Partial OT (POT) [2010]:** Giải phóng sự gò bó bằng cách chỉ vận chuyển ***một phần*** khối lượng s.
  - Loại bỏ nhiễu hiệu quả nhưng chỉ giải quyết bài toán giữa **hai** phân phối (nguồn và đích) .
- **Multi-marginal POT (MMPOT) [2022]:** Là bản nâng cấp của POT**.** Bước nhảy vọt vào kỷ nguyên **đa phương thức.**
  - Kết nối đa nguồn dữ liệu và xử lý được sự không cân bằng của dữ liệu.

3. Phân tích MMPOT [2022] (SOTA)

Để giải MMPOT, giới nghiên cứu sử dụng **Chiến lược mở rộng [Le et al., 2022]**:

- **Cách làm:** Thêm các "Điểm giả" (Dummy points) để biến bài toán thiếu hụt thành bài toán cân bằng, nhằm tận dụng thuật toán Sinkhorn tốc độ cao.
- **Nghịch lý nảy sinh:** Khi kết hợp với Điều chuẩn Entropy (Entropic Regularization), chiến lược này gặp "*Gót chân Achilles*".
- **Tính bất khả thi:** Tính phi tuyến của Entropy khiến khối lượng vận chuyển thực tế sai lệch so với mục tiêu.
- **Bùng nổ chi phí:** Để giảm sai số, độ phức tạp bị đẩy từ $\mathcal{O}(1/\epsilon^2)$ lên tới $\mathcal{O}(1/\epsilon^4)$.

Mục tiêu

UIT-HCM

Câu hỏi đặt ra:

Làm thế nào để giải quyết bài toán MMPOT với điều chuẩn Entropy mà không cần thông qua bước mở rộng điểm giả, nhằm đảm bảo nghiệm thu được là khả thi (đúng ràng buộc khối lượng) và đạt tốc độ hội tụ tối ưu?

- Nhóm đề xuất **phát triển một cách tiếp cận mới** dựa trên phương pháp **Primal-Dual (Gốc-Đối ngẫu)** trực tiếp trên bài toán MMPOT, thay vì cố gắng sửa chữa các chiến lược mở rộng cũ trên domain mới.

- Đề tài sẽ chứng minh tính hiệu quả của phương pháp mới so với SOTA ([Le et al., 2022]) dựa trên các độ đo cụ thể:

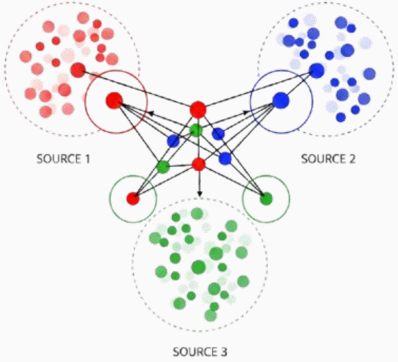
Tiêu chí	MMPOT[2022]	Greenkhorn-MMPOT (Ours)	PDAAM (Ours)	APDAGD (Ours)
Tính khả thi	Không	Có	Có	Có
Số vòng lặp	$\mathcal{O}(\varepsilon^{-4})$	$\mathcal{O}(\varepsilon^{-2})$	$\mathcal{O}(\varepsilon^{-1})$	$\mathcal{O}(\varepsilon^{-1})$
Độ phức tạp	$\mathcal{O}(n^m\varepsilon^{-4})$	$\mathcal{O}(n^m\varepsilon^{-2})$	$\mathcal{O}(n^{m+0.5}\varepsilon^{-1})$	$\mathcal{O}(n^{m+0.5}\varepsilon^{-1})$

Định nghĩa bài toán

UIT-HCM

1.Đầu vào (Input):

- m phân phối rời rạc: $p_1, ..., p_m$ (mỗi phân phối có kích thước n).
- Tensor chi phí: $\mathbf{C}$ (kích thước n^m) biểu thị chi phí kết nối các điểm dữ liệu.
- Khối lượng vận chuyển: s (với $0 < s \leq \min \|\mathbf{p}_i\|_1$).



2.Đầu ra (Output):

- Tensor vận chuyển: $\mathbf{X}$ (cùng kích thước với $\mathbf{C}$) thể hiện kế hoạch ghép nối tối ưu.

3.Mục tiêu:

- Tìm $\mathbf{X}$ để **tổng chi phí là nhỏ nhất**: $$\min_{\mathbf{X}} \langle \mathbf{C}, \mathbf{X} \rangle = \sum_{i_1,\ldots,i_m} \mathbf{C}_{i_1\ldots i_m} \mathbf{X}_{i_1\ldots i_m}$$

4.Ràng buộc:

- Tensor $\mathbf{X}$ phải thỏa mãn tập phương án chấp nhận được $\Pi_s(\mathbf{p}_1, \ldots, \mathbf{p}_m)$, được định nghĩa bởi hệ phương trình sau (bao gồm biến bù $\mathbf{q}_j$):

$$\mathbf{X} \in \Pi_s(\mathbf{p}_1, \ldots, \mathbf{p}_m) \iff \begin{cases} \displaystyle\sum_{\substack{i_1,\ldots,i_{j-1} \\ i_{j+1},\ldots,i_m}} \mathbf{X}(i_1, \ldots, i_m) + \mathbf{q}_j(i_j) = \mathbf{p}_j(i_j), & \forall j \in [m], \forall i_j \in [n], \\ \displaystyle\sum_{i_1,\ldots,i_m} \mathbf{X}(i_1, \ldots, i_m) = s, & \mathbf{X} \geq 0, \mathbf{q}_j \geq \mathbf{0}. \end{cases}$$

Nội dung và Phương pháp

UIT-HCM

Nội dung 1: Phân tích sự bất khả thi của các chiến lược mở rộng (Extension Strategies)
Phương pháp thực hiện:

- **Chứng minh tính bất khả thi:**
  - [Le et al, 2022] chứng minh rằng có thể **giải bài toán MMPOT bằng bài toán MMOT tương ứng**, dựa vào việc 2 bài toán này **có thể viết dưới dạng bài toán quy hoạch tuyến tính (Linear Programming).**
  - *Vấn đề:* Tuy điều này đúng, nhưng bài toán MMOT lại được giải bằng cách dùng **Entropic regularization - một regularizer phi tuyến (non-linear)**
    => Không chắc chắn đảm bảo tính khả thi
  - *Đóng góp:* Chứng minh sự bất khả thi này cho *tất cả các chiến thuật đề xuất trong [Le et al, 2022]*, dẫn đến tất cả thuật toán SOTA trước đó có lý thuyết tệ hơn thực tế.
    => Điều được thật sự kiểm chứng trong thực nghiệm (Fig 1)
    => **Khắc phục bằng cách đề xuất thuật toán Rounding mới**

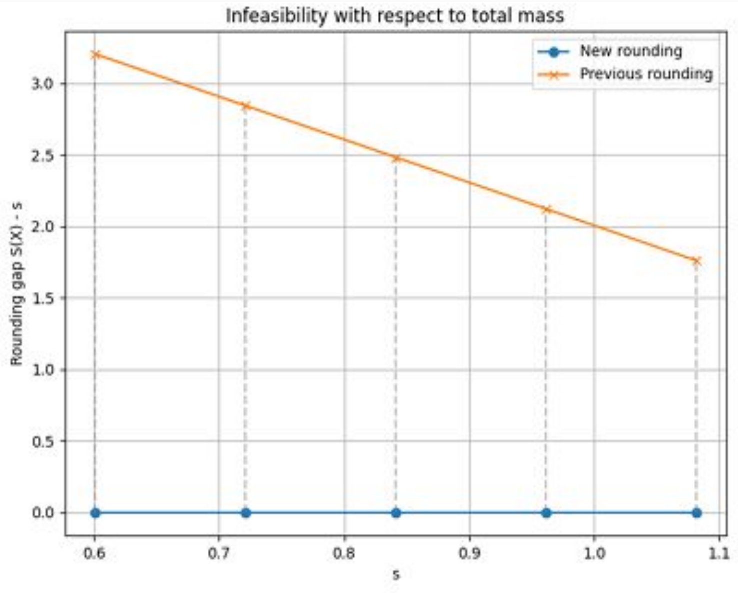


Figure 1 . So sánh tổng khối lượng được vận chuyển thu được bằng phương pháp trước đây (kết hợp chiến lược mở rộng và thuật toán Sinkhorn đa biên, sau đó là bước làm tròn) với khối lượng mục tiêu thực tế.

Nội dung và Phương pháp

UIT-HCM

Nội dung 2: Xây dựng công thức Đối ngẫu mới (Novel Dual Formulation)

Phương pháp thực hiện

- **Thiết lập bài toán LP "tốt" (Primal Problem with Slack Variables):**
  - *SOTA:* Bước đầu tiên của [Le et al,2022] là thay đổi ràng buộc bất đẳng thức của MMPOT bằng cách **thêm hàng + chuẩn hóa**
  - *Đề xuất:* Vì cách này dẫn đến tính bất khả thi, chúng em đề xuất thay đổi ràng buộc này bằng cách **thêm biến nhưng không chuẩn hóa. => Tôn trọng tính đặc thù của MMPOT.**
- **Chuyển đổi sang bài toán Đối ngẫu (Dual Formulation):**
  - Thiết lập bài toán tối ưu đối ngẫu **trực tiếp** mà không cần mở rộng tensor chi phí, giúp giữ nguyên kích thước bài toán.
  - **Tính chất hội tụ:** Chứng minh được hàm mục tiêu đối ngẫu là **lồi mạnh** và **trơn** trên miền xác định

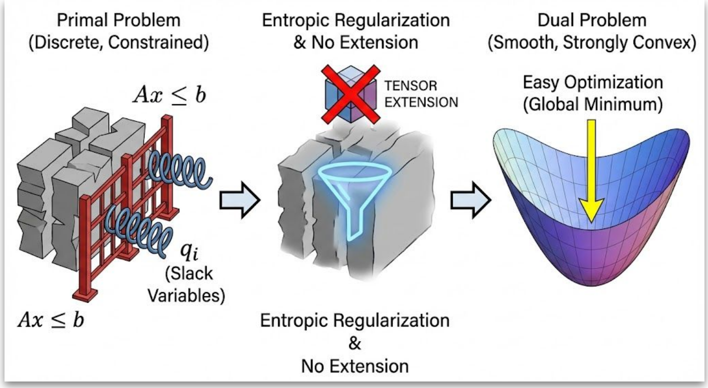


Figure 2. Quy trình giải quyết bài toán

- **Chứng minh đóng góp lý thuyết then chốt:**
  - Chứng minh được **Cận trên mới (Novel Upper Bound)** cho chuẩn L_∞ của nghiệm tối ưu. Đây là cơ sở toán học quan trọng để đảm bảo các thuật toán đề xuất (Greenkhorn, PDAAM) hội tụ nhanh.

Nội dung và Phương pháp

Nội dung 3: Thiết kế và Cài đặt các thuật toán tối ưu (Solvers) và Làm tròn (Rounding)

Phương pháp thực hiện:

- **Thuật toán Rounding:** Cài đặt quy trình **Enforcing Procedure (EP)** để chiếu nghiệm gần đúng (từ bài toán entropy) về nghiệm khả thi thỏa mãn các ràng buộc biên và tổng khối lượng.
- **Greenkhorn-MMPOT:** Thiết kế thuật toán Greenkhorn dựa trên nguyên lí **coordinate minimization**, từ đó giải quyết vấn đề về việc thiếu thuật toán giải MMPOT với độ chính xác thấp trong thời gian rất nhanh.
- **PDAAM & APDAGD:** Áp dụng các phương pháp **Accelerated** dựa trên **Gradient Descent** và **Alternating Minimization** để thiết kế thuật giải cho lời giải với độ chính xác cao hơn, tuy nhiên trong thời gian lâu hơn.

Nội dung 4: Thực nghiệm và Đánh giá hiệu năng

Phương pháp thực hiện:

- Sử dụng bộ dữ liệu Gaussian tổng hợp (Synthetic Gaussian) để kiểm tra tính chất hội tụ và độ chính xác lý thuyết.
- Thử nghiệm trên bộ dữ liệu ảnh chữ viết tay MNIST (bài toán Partial Barycenter) để đánh giá khả năng ứng dụng thực tế.
- Đo lường và so sánh dựa trên các metrics: khoảng cách hàm mục tiêu (Objective value gap), thời gian thực thi (runtime), và sai số ràng buộc.

Kết quả dự kiến

UIT-HCM

- Báo cáo chứng minh toán học về giới hạn của các phương pháp cũ và tính đúng đắn của phương pháp mới.
- Mã nguồn Python cài đặt hoàn chỉnh 03 thuật toán: GreenkhornMMPOT, PDAAM, APDAGD và thuật toán Rounding.
- Các biểu đồ so sánh cho thấy các thuật toán cho ra lời giải thỏa ràng buộc, xấp xỉ gần đúng với lời giải tối ưu trong thời gian nhanh hơn so với thuật toán SOTA là dummy point strategy.
- Ứng dụng: Giải nhanh bài toán tìm Wasserstein Barycenter, ứng dụng trong domain adaptation, dictionary learning, representation learning,..., .

Tài liệu thảo khảo

[1]. Anh Duc Nguyen, Tuan Dung Nguyen, Quang Minh Nguyen, Hoang H. Nguyen, Lam M. Nguyen, Kim-Chuan Toh. On partial optimal transport: Revising the infeasibility of sinkhorn and efficient gradient methods. AAAI 2024: 14387-14395.

[2]. Tianyi Lin, Nhat Ho, Marco Cuturi, Michael I. Jordan. On the Complexity of Approximating Multimarginal Optimal Transport. J. Mach. Learn. Res. 23: 65:1-65:43 (2022).

[3]. Khang Le, Huy Nguyen, Khai Nguyen, Tung Pham, Nhat Ho. On Multimarginal Partial Optimal Transport: Equivalent Forms and Computational Complexity. AISTATS 2022: 4397-4413.

[4]. Marco Cuturi. Sinkhorn Distances: Lightspeed Computation of Optimal Transport. NIPS 2013: 2292-2300.

[5]. Pavel Dvurechensky, Alexander Gasnikov, Alexey Kroshnin. Computational Optimal Transport: Complexity by Accelerated Gradient Descent Is Better Than by Sinkhorn's Algorithm. ICML 2018: 1366-1375