THÔNG TIN CHUNG CỦA NHÓM

- Link YouTube video của báo cáo: https://www.youtube.com/watch?v=oJYt4nf3nJU
- Link slides: https://github.com/Binbonk5/CS519.Q11.KHTN

• Họ và Tên: Hoàng Đức Dũng
• MSSV: 23520328

	• Lớp: CS519.Q11.KHTN
	• Tự đánh giá (điểm tổng kết môn): 9.5/10
	• Số buổi vắng: 1
	• Số câu hỏi QT cá nhân: 6
	• Số câu hỏi QT của cả nhóm: 12
	• Link Github: https://github.com/Binbonk5/CS519.Q11.KHTN/
	• Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:
	    ○ Viết poster
	    ○ Làm slide
	    ○ Làm phần các phương pháp thực hiện của bài toán

<table>
<tr>
<td>

- Họ và Tên: Nguyễn Đình Thiên Quang
- MSSV: 23521285

</td>
<td>

- Lớp: CS519.Q11
- Tự đánh giá (điểm tổng kết môn): 9.5/10
- Số buổi vắng: 0
- Số câu hỏi QT cá nhân: 6
- Số câu hỏi QT của cả nhóm: 12
- Link Github: https://github.com/thienquang204/CS519.Q11.
- Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:
    - Lên ý tưởng nghiên cứu
    - Viết phân tích động lực, cơ sở thí nghiệm cho tính bất khả thi của nghiên cứu trước đó.
    - Làm video YouTube

</td>
</tr>
</table>

ĐỀ CƯƠNG NGHIÊN CỨU

TÊN ĐỀ TÀI (IN HOA)

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

TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

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

TÓM TẮT *(Tối đa 400 từ)*

Đề tài nghiên cứu bài toán Vận chuyển Tối ưu Một phần Đa biên (Multi-marginal Partial Optimal Transport - MMPOT), một sự tổng quát hóa của bài toán Vận chuyển tối ưu (OT) rời rạc trên m biên không cân bằng. Các nghiên cứu trước đây thường sử dụng chiến lược **mở rộng** bằng cách thêm các điểm giả (dummy points) để đưa bài toán MMPOT về dạng OT chuẩn, sau đó dùng thuật toán Sinkhorn truyền thống để giải [Le et al 2022].

Tuy nhiên, trong nghiên cứu này, chúng tôi chứng minh rằng chiến lược mở rộng nói trên không chỉ tạo ra các giải pháp **bất khả thi** đối với bài toán MMPOT được điều chỉnh bằng Entropy (Entropic Regularized), mà còn làm giảm hiệu suất tính toán so với các phương pháp OT truyền thống. Để giải quyết vấn đề này, đề tài tiếp cận theo hướng Primal-Dual và đề xuất các thuật toán mới bao gồm: PDAAM, APDAGD và GreenkhornMMPOT. Các thuật toán này không cần sử dụng tensor chi phí mở rộng, đảm bảo tính khả thi của nghiệm và đạt được tốc độ hội tụ tối ưu $O(\frac{1}{\varepsilon})$ hoặc chi phí tính toán tối ưu. Kết quả thực nghiệm trên dữ liệu tổng hợp và bài toán Partial Barycenter trên tập dữ liệu MNIST sẽ chứng minh tính hiệu quả vượt trội của phương pháp đề xuất.

GIỚI THIỆU *(Tối đa 1 trang A4)*

1. Bối cảnh và Lý do chọn đề tài:

Vận chuyển tối ưu (Optimal Transport - OT) là một công cụ mạnh mẽ trong học máy và thị giác máy tính. Tuy nhiên, OT yêu cầu tổng khối lượng giữa các phân phối phải bằng nhau, điều này hạn chế khả năng áp dụng trong thực tế khi dữ liệu bị nhiễu hoặc chỉ cần so khớp một phần. Vận chuyển tối ưu một phần (Partial OT - POT) [Nguyen et al 2024] và phiên bản đa biên của nó (MMPOT) ra đời để giải quyết vấn đề này.

Hiện nay, các phương pháp giải MMPOT phổ biến dựa trên việc "mở rộng" không gian bằng các điểm giả để quy về bài toán OT. Tuy nhiên, cách tiếp cận này gặp vấn đề lớn về tính khả thi của nghiệm khi áp dụng điều chuẩn Entropy, dẫn đến sai số lớn trong tổng khối lượng vận chuyển và độ phức tạp tính toán cao.

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?

2. Bài toán cụ thể:

Chúng em xét bài toán Vận chuyển tối ưu một phần đa biên (MMPOT) giữa m phân phối rời rạc $\mathbf{p}_1, \dots, \mathbf{p}_m$, với ràng buộc tổng khối lượng vận chuyển là s ($s \leq \min \|\mathbf{p}_i\|_1$). Mục tiêu là tìm phương án vận chuyển $\mathbf{X}$ (tensor) để tối thiểu hóa tổng chi phí:

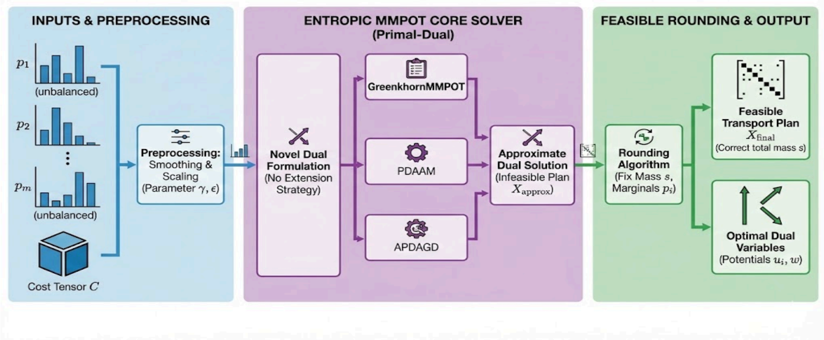
$$\min_{\mathbf{X} \in \Pi_s(\mathbf{p}_1, \dots, \mathbf{p}_m)} \langle \mathbf{C}, \mathbf{X} \rangle$$

Trong đó tập phương án chấp nhận được Π_s được định nghĩa lại chặt chẽ như sau (sử dụng biến bù slack variables $\mathbf{q}_j$):

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

3. Input/Output của đề tài:

- **Input:**
  - m phân phối rời rạc: $\mathbf{p}_1, \dots, \mathbf{p}_m$, trong đó mỗi phân phối có kích thước n
  - Tensor chi phí: $\mathbf{C}$ có kích thước n^m, biểu thị chi phí để kết nối các điểm dữ liệu từ các phân phối khác nhau.
  - Khối lượng vận chuyển (Total Mass): tham số s (với điều kiện $0 < s \leq \min \|\mathbf{p}_i\|_1$).
- **Output**: Tensor vận chuyển (Transport Tensor): $\mathbf{X}$ có cùng kích thước với $\mathbf{C}$ (n^m), thể hiện kế hoạch ghép nối tối ưu giữa các phân phối.



MỤC TIÊU

*1. **Phân tích lý thuyết:*** Chứng minh tính bất khả thi và sự kém hiệu quả của các chiến lược mở rộng (dummy point extension strategies) hiện có trong việc giải bài toán Entropic MMPOT.

*2. **Phát triển thuật toán:*** Xây dựng công thức đối ngẫu (Dual Formulation) mới và thiết kế các thuật toán tối ưu dựa trên khung Primal-Dual (GreenkhornMMPOT, PDAAM, APDAGD) cùng thuật toán làm tròn (Rounding) mới để đảm bảo nghiệm khả thi.

*3. **Đánh giá thực nghiệm:*** Cài đặt các thuật toán và so sánh hiệu năng (thời gian chạy, sai số hội tụ) với các phương pháp SOTA hiện có trên dữ liệu tổng hợp và ứng dụng Partial Barycenter.

NỘI DUNG VÀ PHƯƠNG PHÁP

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

Nghiên cứu sâu về mặt lý thuyết để chỉ ra những hạn chế cốt lõi của các phương pháp hiện tại (SOTA) khi giải quyết bài toán MMPOT thông qua việc thêm điểm giả (dummy points).

- **Phương pháp thực hiện:**
  - Phân tích cơ sở toán học để làm rõ nguyên nhân khiến các chiến lược mở rộng gây ra sai số về tổng khối lượng vận chuyển, đặc biệt khi áp dụng điều chuẩn Entropy (Entropic Regularization).
  - Sử dụng các chứng minh toán học để khẳng định độ phức tạp tính toán

của các phương pháp này bị tăng lên theo hàm mũ hoặc đòi hỏi các phần tử chi phí cực lớn để đạt được độ chính xác mong muốn.

- **Kết quả dự kiến:**
  - Các định lý và bổ đề chứng minh tính không khả thi (infeasibility) và sự kém hiệu quả của chiến lược mở rộng, tạo tiền đề vững chắc cho việc đề xuất phương pháp mới.

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

Thay vì chỉnh sửa tensor chi phí đầu vào, nghiên cứu sẽ tập trung xây dựng một bài toán tối ưu hóa đối ngẫu mới cho MMPOT.

- **Phương pháp thực hiện:**
  - Thiết lập bài toán quy hoạch tuyến tính cho MMPOT có bổ sung các biến phụ (slack variables) để xử lý ràng buộc bất đẳng thức.
  - Dẫn ra công thức đối ngẫu trơn (smooth dual formulation) trực tiếp từ bài toán gốc mà không cần mở rộng tensor chi phí.
  - Thực hiện chứng minh chặn trên của chuẩn L_∞ cho các biến đối ngẫu để đảm bảo tính hội tụ của thuật toán.
- **Kết quả dự kiến:**
  - Một mô hình toán học đối ngẫu hoàn chỉnh, trơn và lồi, sẵn sàng để áp dụng các thuật toán tối ưu dựa trên gradient hiệu quả.

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

Phát triển bộ công cụ thuật toán để giải quyết bài toán đối ngẫu đã thiết lập và đưa nghiệm về dạng khả thi.

- **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.
  - **GreenkhornMMPOT:** Cải tiến thuật toán Greenkhorn cho bài toán đa biên, sử dụng chiến lược cập nhật tọa độ tham lam (greedy coordinate descent) để tăng tốc độ hội tụ trong các vòng lặp đầu.
  - **PDAAM & APDAGD:** Áp dụng các phương pháp tăng tốc (Accelerated) dựa trên đạo hàm (Gradient Descent) và cực tiểu hóa thay phiên (Alternating Minimization).
- **Kết quả dự kiến:**

- Các thuật toán đạt tốc độ hội tụ tối ưu (ví dụ: $O(\frac{1}{\varepsilon})$ cho phương pháp tăng tốc), vượt trội so với các phương pháp dựa trên Sinkhorn truyền thống.
- Thuật toán Rounding đảm bảo sai số được kiểm soát chặt chẽ.

Nội dung 4: Thực nghiệm và Đánh giá hiệu năng

Kiểm chứng tính đúng đắn và hiệu quả của các phương pháp đề xuất trên dữ liệu mô phỏng và dữ liệu thực tế.

- **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 (marginal error).
- **Kết quả dự kiến:**
  - Bảng số liệu và biểu đồ minh họa sự vượt trội của GreenkhornMMPOT, PDAAM và APDAGD so với các baseline về tốc độ và độ chính xác.
  - Xác nhận tính khả thi của nghiệm sau khi qua bước Rounding mới.

KẾT QUẢ MONG ĐỢI

- 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 THAM 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