Learning Geometry Preserving Optimal Transport Plan via Convex Optimization

Junhyoung Chung*
Department of Statistics
Seoul National University
Seoul 08826, Republic of Korea
junhyoung0534@gmail.com

Abstract

The abstract paragraph should be indented ½ inch (3 picas) on both the left- and right-hand margins. Use 10 point type, with a vertical spacing (leading) of 11 points. The word **Abstract** must be centered, bold, and in point size 12. Two line spaces precede the abstract. The abstract must be limited to one paragraph.

1 Introduction

Optimal transport (OT) provides a powerful mathematical framework for comparing probability measures by quantifying the minimal cost of transporting mass from one distribution to another. In recent years, OT has found wide applications in statistics, machine learning, and computer vision, where distributions often lie on non-Euclidean or structured domains. However, in many real-world problems, each observation possesses both spatial and feature information—for example, geometric shapes with embedded descriptors, or spatially indexed random fields with associated features. In such settings, it is desirable to align not only the feature embeddings but also the underlying spatial structures.

To address this, we consider a *fused optimal transport* (FOT) formulation, which simultaneously accounts for feature similarity and spatial coherence through a kernel-weighted coupling cost. This formulation generalizes both the classical quadratic OT and the Gromov–Wasserstein (GW) transport, providing a flexible interpolation between them. The rest of this section introduces the formal setup, notation, and basic existence results for the fused optimal transport plan.

2 Methodology

Notations. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and (S, d_S) a compact metric space. A measurable map $X : \Omega \to S$ is called a random element with distribution $\mathbb{P}_X := \mathbb{P} \circ X^{-1}$. We also introduce a feature space $M \subset \mathbb{R}^d$ which is compact, and call any one-to-one and continuous $f : S \to M$ a feature function. Throughout this study, we assume that $\operatorname{diam}(S) = \operatorname{diam}(M) = 1$, where $\operatorname{diam}(A) := \sup_{x,x' \in A} d_A(x, x')$. For two probability measures μ, ν on S, denote by

$$\Pi(\mu, \nu) := \{\pi \text{ on } S \times S : \text{ the marginals are } \mu \text{ and } \nu\}$$

the set of all couplings between μ and ν. By the disintegration theorem, for $\pi \in \Pi(\mu, \nu)$ there exists a Markov kernel $k(, \cdot, \cdot) : \mathcal{B}(S) \times S \to [0, 1]$ such that $\pi(dx, dy) = k(dy, x)\mu(dx)$, where $\mathcal{B}(S)$ is the Borel σ-algebra of S.

*https://junhyoung-chung.github.io/

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

Fused Gromov-Wasserstein Discrepancy. For $0 \leq \alpha \leq 1$ and a feature function f, Vayer et al. [2020] propose the following optimization problem:

$$\inf_{\pi \in \Pi(\mathbb{P}_X, \mathbb{P}_Y)} (1-\alpha)\mathbb{E}_{(X,Y)\sim\pi}\left[\|f(X) - f(Y)\|_2^2\right] \\ + \alpha \mathbb{E}_{\substack{(X,Y)\sim\pi \\ (X',Y')\sim\pi}}\left[\left|d_S(X,X') - d_S(Y,Y')\right|^2\right]. \tag{1}$$

The first term enforces feature-wise alignment via f, while the second encourages structural consistency under the spatial metric d_S. When $\alpha = 0$, the problem reduces to classical quadratic OT; when $\alpha = 1$, it approaches the Gromov–Wasserstein setting emphasizing relational geometry.

Proposition 1 (Existence of a minimizer)**.** *For each* $0 \leq \alpha \leq 1$*,* (1) *has at least one minimzers; that is,* (1) *is solvable.*

The existence follows from standard weak compactness of the set of couplings $\Pi(\mathbb{P}_X, \mathbb{P}_Y)$ and lower semicontinuity of the objective functional. However, the minimizer of (1) is not necessarily unique, and due to the non-convexity of the second (structural) term, the optimization landscape may contain multiple local minima. Consequently, standard numerical algorithms can only guarantee convergence to stationary or locally optimal solutions, rather than the global optimum. This highlights the importance of developing a convex reformulation or an appropriate convex relaxation of the fused Gromov–Wasserstein problem to ensure computational tractability and theoretical robustness.

Proposed method.

References

T. Vayer, L. Chapel, R. Flamary, R. Tavenard, and N. Courty. Fused gromov-wasserstein distance for structured objects. *Algorithms*, 13(9):212, 2020.

A Appendix