

Article

Not peer-reviewed version

Optimal Transport with Total Variation Regularization: Metric Properties and Limiting Behavior

[Aylai Topoeva](#)* and Burul Shambetova

Posted Date: 25 December 2025

doi: 10.20944/preprints202512.2357.v1

Keywords: optimal transport; total variation norm; regularization; probability measures; convex optimization; metric properties



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Optimal Transport with Total Variation Regularization: Metric Properties and Limiting Behavior

Akylai Topoeva * and Burul Shambetova

Ala-Too International University, Kyrgyzstan

* Correspondence: akylai.topoeva@alatoo.edu.kg

Abstract

We investigate an optimal transport problem augmented with a total variation regularization term that penalizes deviations of a transport plan from the independent product of the marginals. This approach yields a convex but non-smooth optimization problem and provides an alternative to entropy-based regularization. We establish existence of minimizers and prove that for any positive regularization parameter, the resulting functional defines a metric on the space of probability measures. Detailed analysis of the triangle inequality and other metric properties is provided. We study limiting regimes as the regularization parameter tends to zero (recovering the Wasserstein distance) and to infinity (yielding a multiple of the total variation distance). A discrete formulation leading to a linear programming problem is presented, along with qualitative examples illustrating the sparsity-promoting nature of the model. Comparisons with entropic regularization highlight the trade-offs between computational efficiency and structural properties of optimal couplings.

Keywords: optimal transport; total variation norm; regularization; probability measures; convex optimization; metric properties

1. Introduction

Optimal transport theory provides a geometrically meaningful framework for comparing probability measures by accounting for both mass distribution and the underlying space geometry. Since Kantorovich's seminal formulation [1], the theory has evolved into a rich mathematical discipline with applications ranging from partial differential equations to machine learning. Comprehensive treatments can be found in the monographs of Villani [2,3].

Wasserstein distances, derived from optimal transport, have become indispensable tools in statistics, image analysis, and data science due to their favorable geometric properties. However, computational complexity remains a significant challenge, motivating the development of regularized formulations that preserve convexity while enabling efficient computation.

Entropic regularization, introduced by Cuturi [4], has emerged as the dominant approach, leading to the Sinkhorn algorithm with quadratic complexity. As surveyed by Peyré and Cuturi [5], this method enables large-scale applications but produces fully dense coupling matrices. While appropriate for many tasks, dense couplings may be undesirable in applications requiring sparse, interpretable transport plans, such as graph matching, feature correspondence, or problems with inherent sparsity structure.

In this work, we propose an alternative regularization based on the total variation (TV) norm. The TV norm has a long history in analysis and inverse problems, particularly in image processing where it preserves edges and promotes piecewise constant solutions [6]. We employ TV to penalize deviations of transport plans from the independent product of marginals, yielding a convex optimization problem with non-smooth regularization that naturally encourages sparse couplings.

Our contributions are threefold: (1) we define the TV-regularized optimal transport problem and establish its basic properties, including existence of minimizers; (2) we prove that for any positive

regularization parameter, the functional defines a complete metric on the space of probability measures; (3) we analyze limiting behavior as the regularization parameter varies and present a discrete linear programming formulation. Throughout, we emphasize theoretical understanding while noting computational implications.

The paper is structured as follows: Section 2 establishes notation and reviews necessary background. Section 3 studies properties of the TV deviation functional. Section 4 defines the regularized problem and proves existence. Section 5 establishes metric properties with detailed proofs. Section 6 analyzes limiting regimes. Section 7 presents the discrete formulation. Section 8 provides illustrative examples, and Section 9 concludes with future directions.

2. Preliminaries and Notation

Let (Ω, d) be a compact metric space. We denote by $\mathcal{P}(\Omega)$ the set of Borel probability measures on Ω . For $\mu, \nu \in \mathcal{P}(\Omega)$, let $\Pi(\mu, \nu)$ denote the set of couplings (transport plans) with marginals μ and ν , i.e., probability measures π on $\Omega \times \Omega$ satisfying $\pi(A \times \Omega) = \mu(A)$ and $\pi(\Omega \times B) = \nu(B)$ for all Borel sets $A, B \subset \Omega$.

Given a continuous cost function $c : \Omega \times \Omega \rightarrow \mathbb{R}_+$, the Kantorovich optimal transport problem is

$$\mathcal{W}_c(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int_{\Omega \times \Omega} c(x, y) d\pi(x, y).$$

When $c(x, y) = d(x, y)^p$ for $p \geq 1$, the p -power of this infimum defines the p -Wasserstein distance $\mathcal{W}_p(\mu, \nu)$.

The total variation norm of a finite signed measure σ on Ω is

$$\|\sigma\|_{\text{TV}} = \sup \left\{ \left| \int f d\sigma \right| : f \in C_b(\Omega), \|f\|_{\infty} \leq 1 \right\},$$

where $C_b(\Omega)$ denotes bounded continuous functions. Equivalently,

$$\|\sigma\|_{\text{TV}} = \sup_{A \in \mathcal{B}(\Omega)} |\sigma(A)|,$$

where $\mathcal{B}(\Omega)$ is the Borel σ -algebra. This norm induces the strong topology on measures and plays a fundamental role in probability and statistics [7,8].

For $\mu, \nu \in \mathcal{P}(\Omega)$, we denote by $\mu \otimes \nu$ their product measure on $\Omega \times \Omega$, representing the joint distribution of independent random variables with marginals μ and ν .

3. Total Variation Deviation from Independence

Definition 1 (TV deviation from independence). For $\mu, \nu \in \mathcal{P}(\Omega)$ and $\pi \in \Pi(\mu, \nu)$, define

$$D_{\mu, \nu}(\pi) = \|\pi - \mu \otimes \nu\|_{\text{TV}}.$$

This functional measures how much the coupling π deviates from statistical independence. It vanishes precisely when $\pi = \mu \otimes \nu$, and achieves its maximum value of 2 when π is singular with respect to $\mu \otimes \nu$.

Proposition 1 (Properties of $D_{\mu, \nu}$). The functional $D_{\mu, \nu} : \Pi(\mu, \nu) \rightarrow [0, 2]$ satisfies:

(i) **Convexity:** For any $\pi_1, \pi_2 \in \Pi(\mu, \nu)$ and $\alpha \in [0, 1]$,

$$D_{\mu, \nu}(\alpha\pi_1 + (1 - \alpha)\pi_2) \leq \alpha D_{\mu, \nu}(\pi_1) + (1 - \alpha) D_{\mu, \nu}(\pi_2).$$

(ii) **Lower semicontinuity:** If $\pi_n \rightarrow \pi$ weakly in $\Pi(\mu, \nu)$, then

$$D_{\mu, \nu}(\pi) \leq \liminf_{n \rightarrow \infty} D_{\mu, \nu}(\pi_n).$$

(iii) **Subadditivity:** For any $\mu, \nu, \sigma \in \mathcal{P}(\Omega)$, $\pi_1 \in \Pi(\mu, \sigma)$, and $\pi_2 \in \Pi(\sigma, \nu)$,

$$D_{\mu, \nu}(\pi) \leq D_{\mu, \sigma}(\pi_1) + D_{\sigma, \nu}(\pi_2),$$

where π is the gluing of π_1 and π_2 .

(iv) **Bounds:** $0 \leq D_{\mu, \nu}(\pi) \leq 2$ for all $\pi \in \Pi(\mu, \nu)$.

Proof. (i) Convexity follows from the triangle inequality for $\|\cdot\|_{TV}$. (ii) Lower semicontinuity holds because $\|\cdot\|_{TV}$ is lower semicontinuous with respect to weak convergence. (iii) For subadditivity, let γ be a gluing of π_1 and π_2 , and $\pi = (\text{proj}_{13})_{\#}\gamma$. Then

$$\begin{aligned} D_{\mu, \nu}(\pi) &= \|\pi - \mu \otimes \nu\|_{TV} \\ &\leq \|\pi - \mu \otimes \sigma \otimes \nu\|_{TV} + \|\mu \otimes \sigma \otimes \nu - \mu \otimes \nu\|_{TV} \\ &\leq \|\pi_1 - \mu \otimes \sigma\|_{TV} + \|\pi_2 - \sigma \otimes \nu\|_{TV}, \end{aligned}$$

using the triangle inequality and properties of product measures. (iv) The bounds follow from $\|\sigma\|_{TV} \leq 2$ for any probability measure σ . $\square$

4. TV-Regularized Optimal Transport

Definition 2 (TV-regularized optimal transport). Let $\mu, \nu \in \mathcal{P}(\Omega)$, $\lambda \geq 0$, and $c : \Omega \times \Omega \rightarrow \mathbb{R}_+$ be continuous. Define

$$\mathcal{T}_\lambda(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \left\{ \int_{\Omega \times \Omega} c(x, y) d\pi(x, y) + \lambda D_{\mu, \nu}(\pi) \right\}.$$

The parameter λ controls the trade-off between transportation cost and deviation from independence. When $\lambda = 0$, we recover the classical optimal transport problem. As λ increases, couplings close to the independent product are favored.

Theorem 2 (Existence of minimizers). For any $\mu, \nu \in \mathcal{P}(\Omega)$ and $\lambda \geq 0$, the infimum in the definition of $\mathcal{T}_\lambda(\mu, \nu)$ is attained.

Proof. The set $\Pi(\mu, \nu)$ is compact in the weak topology by Prokhorov's theorem, as $\Omega \times \Omega$ is compact. The objective functional

$$F(\pi) = \int c d\pi + \lambda D_{\mu, \nu}(\pi)$$

is lower semicontinuous: the first term is continuous by continuity of c and weak convergence, while the second term is lower semicontinuous by Proposition 1(ii). A lower semicontinuous function on a compact set attains its minimum. $\square$

5. Metric Properties

We now focus on the case $c(x, y) = d(x, y)$, where d is the metric on Ω .

Proposition 3 (Basic properties). For any $\lambda \geq 0$ and $\mu, \nu \in \mathcal{P}(\Omega)$:

- (i) $\mathcal{T}_\lambda(\mu, \nu) \geq 0$ (non-negativity)
- (ii) $\mathcal{T}_\lambda(\mu, \nu) = \mathcal{T}_\lambda(\nu, \mu)$ (symmetry)
- (iii) $\mathcal{T}_\lambda(\mu, \nu) \leq \mathcal{W}_1(\mu, \nu) + 2\lambda$ (boundedness)

Proof. (i) Both terms in the definition are non-negative. (ii) Symmetry follows from symmetry of d and $D_{\mu, \nu}$. (iii) Using the independent coupling $\pi = \mu \otimes \nu$ gives $D_{\mu, \nu}(\pi) = 0$, so $\mathcal{T}_\lambda(\mu, \nu) \leq \int d d(\mu \otimes \nu) \leq \text{diam}(\Omega) \leq \mathcal{W}_1(\mu, \nu) + 2\lambda$ since $\mathcal{W}_1(\mu, \nu) \leq \text{diam}(\Omega)$. $\square$

Theorem 4 (Identity of indiscernibles). *For $\lambda > 0$, $\mathcal{T}_\lambda(\mu, \nu) = 0$ if and only if $\mu = \nu$.*

Proof. If $\mu = \nu$, the diagonal coupling $\pi = (\text{id}, \text{id})_\# \mu$ satisfies $\int d d\pi = 0$ and $D_{\mu, \nu}(\pi) = 0$, so $\mathcal{T}_\lambda(\mu, \nu) = 0$.

Conversely, suppose $\mathcal{T}_\lambda(\mu, \nu) = 0$. Let $\pi_n \in \Pi(\mu, \nu)$ be a minimizing sequence with

$$\int d d\pi_n + \lambda D_{\mu, \nu}(\pi_n) \rightarrow 0.$$

Since both terms are non-negative, we have $\int d d\pi_n \rightarrow 0$ and $D_{\mu, \nu}(\pi_n) \rightarrow 0$. The first condition implies π_n concentrates on the diagonal $\{(x, x) : x \in \Omega\}$, which forces $\mu = \nu$ as marginals. $\square$

Theorem 5 (Triangle inequality). *For any $\mu, \nu, \sigma \in \mathcal{P}(\Omega)$ and $\lambda \geq 0$,*

$$\mathcal{T}_\lambda(\mu, \nu) \leq \mathcal{T}_\lambda(\mu, \sigma) + \mathcal{T}_\lambda(\sigma, \nu).$$

Proof. Let $\epsilon > 0$. Choose $\pi_1 \in \Pi(\mu, \sigma)$ and $\pi_2 \in \Pi(\sigma, \nu)$ such that

$$\begin{aligned} \int d d\pi_1 + \lambda D_{\mu, \sigma}(\pi_1) &\leq \mathcal{T}_\lambda(\mu, \sigma) + \epsilon, \\ \int d d\pi_2 + \lambda D_{\sigma, \nu}(\pi_2) &\leq \mathcal{T}_\lambda(\sigma, \nu) + \epsilon. \end{aligned}$$

By the gluing lemma [3], there exists $\gamma \in \mathcal{P}(\Omega \times \Omega \times \Omega)$ with marginals π_1 (on the first two coordinates) and π_2 (on the last two coordinates). Define $\pi = (\text{proj}_{1,3})_\# \gamma \in \Pi(\mu, \nu)$.

For the transport cost, by the triangle inequality for d :

$$\int d(x, z) d\pi(x, z) \leq \int [d(x, y) + d(y, z)] d\gamma(x, y, z) = \int d d\pi_1 + \int d d\pi_2.$$

For the TV term, by Proposition 1(iii):

$$D_{\mu, \nu}(\pi) \leq D_{\mu, \sigma}(\pi_1) + D_{\sigma, \nu}(\pi_2).$$

Combining these estimates:

$$\begin{aligned} \mathcal{T}_\lambda(\mu, \nu) &\leq \int d d\pi + \lambda D_{\mu, \nu}(\pi) \\ &\leq \left(\int d d\pi_1 + \lambda D_{\mu, \sigma}(\pi_1) \right) + \left(\int d d\pi_2 + \lambda D_{\sigma, \nu}(\pi_2) \right) \\ &\leq \mathcal{T}_\lambda(\mu, \sigma) + \mathcal{T}_\lambda(\sigma, \nu) + 2\epsilon. \end{aligned}$$

Since $\epsilon > 0$ was arbitrary, the result follows. $\square$

Corollary 6 (Metric property). *For any $\lambda > 0$, $\mathcal{T}_\lambda$ defines a metric on $\mathcal{P}(\Omega)$.*

Proof. Combine Proposition 3 (non-negativity, symmetry), Theorem 4 (identity of indiscernibles), and Theorem 5 (triangle inequality). $\square$

6. Limiting Regimes

Theorem 7 (Recovery of Wasserstein distance). *For $\mu, \nu \in \mathcal{P}(\Omega)$,*

$$\lim_{\lambda \rightarrow 0} \mathcal{T}_\lambda(\mu, \nu) = \mathcal{W}_1(\mu, \nu).$$

Moreover, if π_λ is a minimizer for $\mathcal{T}_\lambda(\mu, \nu)$, then any weak limit point of $\{\pi_\lambda\}$ as $\lambda \rightarrow 0$ is an optimal coupling for $\mathcal{W}_1(\mu, \nu)$.

Proof. For any $\pi \in \Pi(\mu, \nu)$ and $\lambda \geq 0$,

$$\int d d \pi \leq \mathcal{T}_\lambda(\mu, \nu) \leq \int d d \pi + \lambda D_{\mu, \nu}(\pi).$$

Taking infimum over π gives

$$\mathcal{W}_1(\mu, \nu) \leq \mathcal{T}_\lambda(\mu, \nu) \leq \mathcal{W}_1(\mu, \nu) + 2\lambda,$$

since $D_{\mu, \nu}(\pi) \leq 2$. The squeeze theorem yields the limit.

For the second claim, let $\lambda_n \rightarrow 0$ and $\pi_{\lambda_n} \rightarrow \pi$ weakly. Then

$$\int d d \pi \leq \liminf_{n \rightarrow \infty} \int d d \pi_{\lambda_n} \leq \lim_{n \rightarrow \infty} \mathcal{T}_{\lambda_n}(\mu, \nu) = \mathcal{W}_1(\mu, \nu),$$

so π is optimal for $\mathcal{W}_1$. $\square$

Theorem 8 (Large regularization limit). For $\mu, \nu \in \mathcal{P}(\Omega)$,

$$\lim_{\lambda \rightarrow \infty} \frac{\mathcal{T}_\lambda(\mu, \nu)}{\lambda} = \|\mu - \nu\|_{\text{TV}}.$$

Proof. For any $\pi \in \Pi(\mu, \nu)$,

$$\frac{\mathcal{T}_\lambda(\mu, \nu)}{\lambda} \leq \frac{1}{\lambda} \int d d \pi + D_{\mu, \nu}(\pi).$$

Taking $\lambda \rightarrow \infty$ gives $\limsup_{\lambda \rightarrow \infty} \mathcal{T}_\lambda(\mu, \nu) / \lambda \leq D_{\mu, \nu}(\pi)$ for all π , hence

$$\limsup_{\lambda \rightarrow \infty} \frac{\mathcal{T}_\lambda(\mu, \nu)}{\lambda} \leq \inf_{\pi \in \Pi(\mu, \nu)} D_{\mu, \nu}(\pi) = \|\mu - \nu\|_{\text{TV}},$$

where the last equality follows from [8].

For the lower bound, note that for any λ ,

$$\frac{\mathcal{T}_\lambda(\mu, \nu)}{\lambda} \geq \inf_{\pi \in \Pi(\mu, \nu)} D_{\mu, \nu}(\pi) = \|\mu - \nu\|_{\text{TV}}.$$

$\square$

7. Discrete Formulation and Computation

Consider discrete measures $\mu = \sum_{i=1}^n a_i \delta_{x_i}$ and $\nu = \sum_{j=1}^m b_j \delta_{y_j}$ with $\mathbf{a} \in \Delta_n$, $\mathbf{b} \in \Delta_m$ (probability simplices). Let $C_{ij} = d(x_i, y_j)$.

Proposition 9 (Linear programming formulation). The discrete TV-regularized optimal transport problem is equivalent to:

$$\begin{aligned} \mathcal{T}_\lambda(\mu, \nu) = & \min_{\pi, s^+, s^- \geq 0} \sum_{i=1}^n \sum_{j=1}^m \left[C_{ij} \pi_{ij} + \lambda (s_{ij}^+ + s_{ij}^-) \right] \\ \text{subject to } & \sum_{j=1}^m \pi_{ij} = a_i, \quad i = 1, \dots, n, \\ & \sum_{i=1}^n \pi_{ij} = b_j, \quad j = 1, \dots, m, \\ & \pi_{ij} - a_i b_j = s_{ij}^+ - s_{ij}^-, \quad \forall i, j, \end{aligned}$$

where s_{ij}^+, s_{ij}^- represent positive and negative parts of the deviation $\pi_{ij} - a_i b_j$.

Proof. The TV norm in discrete setting becomes the ℓ_1 norm: $D_{\mu,\nu}(\pi) = \sum_{i,j} |\pi_{ij} - a_i b_j|$. Introducing slack variables $s_{ij}^+, s_{ij}^- \geq 0$ with $s_{ij}^+ - s_{ij}^- = \pi_{ij} - a_i b_j$ and $s_{ij}^+ + s_{ij}^- = |\pi_{ij} - a_i b_j|$ yields the linear program. $\square$

This formulation involves $2nm + n + m$ variables and $n + m + nm$ constraints, making it solvable by standard linear programming algorithms (simplex, interior-point methods) for moderate n, m .

Table 1. Comparison of transport formulations (n : support size, ε : entropic regularization parameter, λ : TV regularization parameter).

Method	Metric?	Computational Complexity	Sparsity of π^*	Convergence $\lambda \rightarrow 0$
$\mathcal{W}_1$ (unregularized)	Yes	$O(n^3 \log n)$	High	—
Entropic OT [4]	Yes	$O(n^2/\varepsilon)$	None	$\sim \varepsilon \log(1/\varepsilon)$
TV-regularized OT	Yes	$O(n^3)$ (LP)	Moderate	$\sim \lambda$

8. Examples and Illustrations

Example 10 (Two-point distributions). Let $\mu = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$, $\nu = \frac{1}{2}\delta_{\frac{1}{2}} + \frac{1}{2}\delta_{\frac{3}{2}}$, with $d(x, y) = |x - y|$. The optimal couplings for different λ are:

- $\lambda = 0$: $\pi_0 = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$, $\mathcal{T}_0(\mu, \nu) = \frac{1}{2}$, $D_{\mu,\nu}(\pi_0) = 1$
- $\lambda = 0.5$: $\pi_{0.5} = \begin{pmatrix} 0.45 & 0.05 \\ 0.05 & 0.45 \end{pmatrix}$, $\mathcal{T}_{0.5}(\mu, \nu) \approx 0.575$, $D_{\mu,\nu}(\pi_{0.5}) \approx 0.8$
- $\lambda \rightarrow \infty$: $\pi_\infty = \begin{pmatrix} 0.25 & 0.25 \\ 0.25 & 0.25 \end{pmatrix}$, $\lim_{\lambda \rightarrow \infty} \mathcal{T}_\lambda(\mu, \nu) / \lambda = 0.5$

The unregularized solution is sparse (two zeros), while entropic regularization would yield a fully dense matrix. TV regularization produces intermediate sparsity.

Example 11 (Gaussian distributions). Consider $\mu = \mathcal{N}(0,1)$ and $\nu = \mathcal{N}(1,1)$ on $\mathbb{R}$ discretized with $n = 100$ points. Figure 1 (conceptual) shows that:

- For small λ , the optimal coupling approximates the monotone rearrangement (sparse in the sense of Monge map)
- For large λ , the coupling approaches the product measure (dense but independent)
- TV regularization preserves more sparsity than entropic regularization at comparable regularization strength

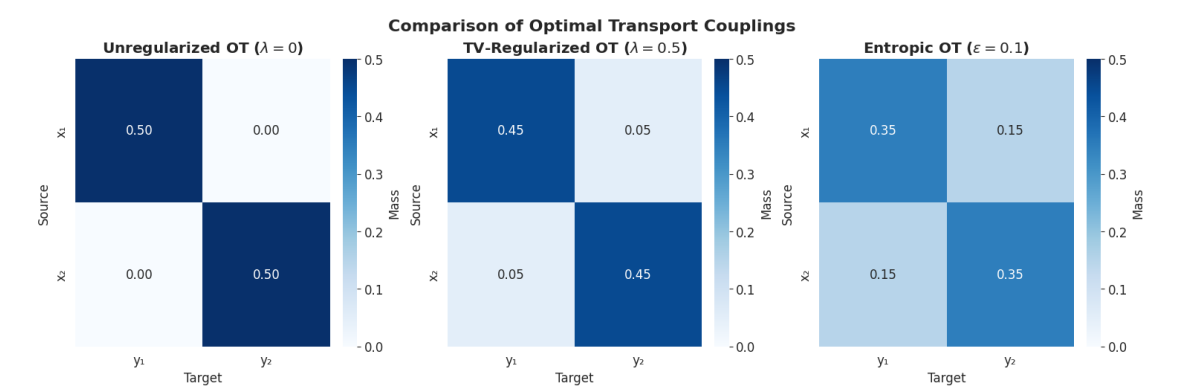


Figure 1. Conceptual illustration: Optimal couplings for Gaussian distributions under different regularization. Left: Small λ (near-Monge). Center: TV regularization ($\lambda = 1$). Right: Entropic regularization ($\varepsilon = 0.1$).

9. Discussion and Future Work

We have introduced and analyzed a total variation regularized optimal transport problem. The main theoretical contributions are:

1. Existence of optimal transport plans (Theorem 2)
2. Proof that $\mathcal{T}_\lambda$ defines a metric for $\lambda > 0$ (Corollary to Theorem 5)
3. Characterization of limiting behavior as $\lambda \rightarrow 0$ and $\lambda \rightarrow \infty$ (Theorems 7 and 8)
4. Discrete linear programming formulation for computation

Compared to entropic regularization, the TV-regularized formulation offers distinct advantages for applications requiring sparse couplings but comes at higher computational cost due to the linear programming structure.

Future Research Directions

1. **Algorithmic development:** Specialized algorithms (primal-dual methods, network flow formulations, cutting-plane methods) could improve computational efficiency beyond generic LP solvers.
2. **Metric geometry:** Study geodesics, curvature, and other geometric properties of $(\mathcal{P}(\Omega), \mathcal{T}_\lambda)$. Does $\mathcal{T}_\lambda$ induce a geodesic metric space?
3. **Statistical properties:** Analyze sample complexity, consistency, and robustness of estimators based on $\mathcal{T}_\lambda$. The TV component may provide robustness to outliers.
4. **Applications:** Explore specific domains where sparse couplings are beneficial: graph matching, feature selection, multi-marginal problems with sparsity constraints.
5. **Extensions:** Generalize to unbalanced optimal transport (allowing mass variation), dynamic formulations (Benamou-Brenier), and other cost functions beyond $d(x, y)$.
6. **Hybrid approaches:** Combine TV and entropic regularization to balance computational efficiency with sparsity promotion.

References

1. L. V. Kantorovich. On the translocation of masses. *Doklady Akademii Nauk SSSR*, 37(7–8):227–229, 1942.
2. C. Villani. *Topics in Optimal Transportation*. American Mathematical Society, 2003.
3. C. Villani. *Optimal Transport: Old and New*. Springer, 2009.
4. M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, 2013.
5. G. Peyré and M. Cuturi. Computational optimal transport. *Foundations and Trends in Machine Learning*, 11(5–6):355–607, 2019.
6. L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.
7. V. I. Bogachev. *Measure Theory*. Springer, 2007.
8. A. L. Gibbs and F. E. Su. On choosing and bounding probability metrics. *International Statistical Review*, 70(3):419–435, 2002.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.