



Minimizing Optimal Transport for Functions with Fixed-Size Nodal Sets

Qiang Du¹ · Amir Sagiv²

Received: 5 January 2023 / Accepted: 24 July 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Consider the class of zero-mean functions with fixed L^∞ and L^1 norms and exactly $N \in \mathbb{N}$ nodal points. Which functions f minimize $W_p(f_+, f_-)$, the Wasserstein distance between the measures whose densities are the positive and negative parts? We provide a complete solution to this minimization problem on the line and the circle, which provides sharp constants for previously proven “uncertainty principle”-type inequalities, i.e., lower bounds on $N \cdot W_p(f_+, f_-)$. We further show that, while such inequalities hold in many metric measure spaces, they are no longer sharp when the non-branching assumption is violated; indeed, for metric star-graphs, the optimal lower bound on $W_p(f_+, f_-)$ is not inversely proportional to the size of the nodal set, N . Based on similar reductions, we make connections between the analogous problem of minimizing $W_p(f_+, f_-)$ for f defined on $\Omega \subset \mathbb{R}^d$ with an equivalent optimal domain partition problem.

Keywords Wasserstein · Uncertainty principle · Nodal set · Metric graph · Optimal partition

Mathematics Subject Classification 28A75 · 49Q05 · 49Q22 · 52C35

Communicated by Peter Sternberg.

✉ Amir Sagiv
as6011@columbia.edu; amirsagiv@technion.ac.il

Qiang Du
qd2125@columbia.edu

¹ Department of Applied Physics and Applied Mathematics, and Data Science Institute, Columbia University, 500 W120 Street, New York, NY 10027, USA

² Department of Applied Physics and Applied Mathematics, Columbia University, 500 W120 Street, New York, NY 10027, USA

1 Introduction

The study of nodal sets (or zero sets) of functions is a classic topic in analysis. Loosely speaking, the key question is, for a given a function, “*how big*” its nodal set $Z(f) = \{x \mid f(x) = 0\}$ is.¹

In recent years, this question has been connected to the notion of *optimal transport*. The intuition is as follows: given a sufficiently regular metric Borel probability space Ω (e.g., a domain in \mathbb{R}^d or a Riemannian manifold) and a sufficiently regular and bounded real-valued function $f : \Omega \rightarrow \mathbb{R}$ with mean zero over Ω , we consider its positive and negative parts $f_{\pm}(x) \equiv \pm \max(\pm f(x), 0)$ as densities, and the nodal set $Z(f)$ as the interfaces of their supports. If the nodal set is *small*, then there should be some regions in the support of f_+ that cannot be near the support of f_- . Therefore, the cost of transporting f_+ to f_- cannot be arbitrarily small.

A common metric which quantifies the notion of transport cost is the *Wasserstein- p distance* $W_p(f_+, f_-)$. To make the connection between the nodal set and the transport cost more transparent, we ask:

Given a nodal set of a certain size, which functions minimize $W_p(f_+, f_-)$? (1)

The answer to this minimization problem should depend on the norms of f ; first, the larger $\|f\|_1$ is, the more mass there is to transport. Second, the more localized the function is, as measured in this work by $\|f\|_{\infty}$, the more mass can be concentrated near the interface, thus decreasing the transport distance and cost.

Before rigorously formalizing our question (see Question 1 in Sect. 2), we note that the motivation behind it originates from the following type of inequalities: for a domain Ω ,

$$W_1(f_+, f_-) \cdot \text{Surf} \{x \in \Omega : f(x) = 0\} \geq c(\Omega) \left(\frac{\|f\|_{L^1}}{\|f\|_{L^\infty}} \right)^\rho \|f\|_{L^1}, \quad \forall f \in C^0(\bar{\Omega}), \quad (2)$$

where $\rho > 0$, $c(\Omega) > 0$ is a constant that depends only on the domain Ω , and Surf is a surface measure, e.g., the co-dimension 1 Hausdorff measure. The first inequality of this type was proven by Steinerberger (2020) for two-dimensional domains such as the unit square $[0, 1]^2$, with $\rho = 1$. This result was then generalized to certain compact domains in \mathbb{R}^d with arbitrary $d \geq 1$ by Steinerberger and the second author with $\rho = 4 - 1/d$, see Sagiv and Steinerberger (2020). Subsequently, Carroll et al. (2020) improved the exponent to $\rho = 2 - 1/d$, and Cavalletti and Farinelli proved that the optimal $\rho = 1$ is true in all dimensions (Cavalletti and Farinelli 2021). The analysis in Cavalletti and Farinelli (2021) takes place at a much larger class of metric measure spaces that satisfy the Curvature-Dimension (CD(K, N)) condition, where the nodal set can be measured using the notion of Perimeter (Ambrosio and Di Marino 2014; Miranda 2003). That result was then also proven for RCD(K, ∞) metric spaces (Ambrosio et al. 2015, 2014) by De-Ponti and Farinelli for all $p \geq 1$ (2022).

¹ Here we use the term “nodal set” for a broad class of functions (see Sect. 2), whereas in other places it refers to the zero sets of a special class of functions, e.g., Laplacian eigenfunctions.

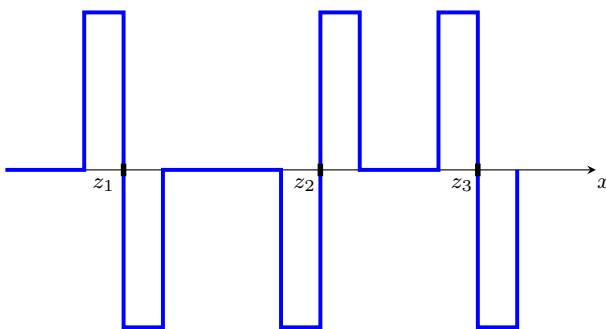


Fig. 1 An exemplary minimizer of $W_p(f_+, f_-)$ with 3 nodal points on an interval

Despite the great level of generality of the above results, some fundamental questions still remain: is the inequality (2) with an exponent $\rho = 1$ sharp? First, we do not know what the optimal constant $c(\Omega)$ is. More fundamentally, one may ask whether a *multiplicative* inequality is the natural one. Indeed, some works studied related additive inequalities (Buttazzo et al. 2020; Candau-Tilh and Goldman 2022; Novack et al. 2023; Xia and Zhou 2021) regarding transport of the Lebesgue measure between disjoint domains.²

But even the most general results regarding inequalities of the form (2) require an essentially non-branching property (Cavalletti and Farinelli 2021; De Ponti and Farinelli 2022): all $\text{RCD}(K, \infty)$ spaces satisfy this property (Rajala and Sturm 2014), whereas the result of Cavalletti and Farinelli (2021) is proven only to $\text{CD}(K, N)$ spaces which satisfy this property. *A primary takeaway of our study is that it is indeed necessary in order for inequalities like (2) to be sharp.*

1.1 Main Results

In this paper, we formalize the minimization problem (1) and establish a strategy to its solution. We find the minimizers on three families of one-dimensional domains: a line (or an interval), a circle, and metric star graphs. The analysis of these domains leads us to the following key conclusions:

- **Generalized nodal sets for $L^1 \cap L^\infty$ functions** Common to all these settings are that the minimizers are *step functions*, and therefore do not have nodal sets in the usual sense. Hence, it is crucial to extend the class of admissible functions and notion of nodal sets of continuous functions to a broader class of $L^1 \cap L^\infty$ functions. Furthermore, it may be expected that this would be the right formulation through which to explore the minimizers in higher dimensions (see Sect. 6).
- **Strategy** Our work establishes a strategy to approach the minimization problem under consideration through a series of reductions of the class of feasible func-

² In particular, Candau-Tilh and Goldman (2022), Proposition 2.5 and Novack et al. (2023), Lemma 2.8 consider the overall transport outside of a bounded set $E \subset \mathbb{R}^d$ and prove that a lower bound on the transport cost must be inversely proportional to the perimeter of E , i.e., if $F \subseteq E^c$ with volume $|E| = |F| = 1$, then $W_p(\mathbb{1}_E, \mathbb{1}_F) > C(d)/\text{Per}(E)$ for some universal $C(d) > 0$.

tions. The reduction strategy consists of three stages: (1) We reduce the functional minimization problem to a geometric one. Fixing the nodal domain, i.e., the interface between the positive and negative parts (see Definition 1), we show that it is always preferable to allocate the L^1 mass near the nodal domain. Thus, given a function f we find a new function g which is (up to a sign and scaling) the indicator function of a subset of the supports of f_{\pm} , and for which the overall transport cost is cheaper. (2) We find that there is always a preferable configuration of the mass and the nodal points such that the optimal transport plan only couples adjacent intervals. (3) Steps (1) and (2) reduce the infinite-dimensional variational problem to a constrained finite dimensional problem, which one can solve by elementary means.

- **The minimizers** By posing the question as a minimization problem, our work goes beyond previous works on the subject since we gain insight into the minimizers themselves. In the simplest case, the interval (or for every smooth non-intersecting curve), the minimizers are step functions whose only possible values are either $\pm \|f\|_{\infty}$ or 0, as can be seen in Fig. 1. Thus, these are functions for which the optimal transport between f_+ and f_- is local, across a single interface. Next, even though the circle introduces a new *global* structure, we show that the nature of the problem (and the minimizers) do not change.
- **Sharp multiplicative inequalities for the interval and the circle** The complete solution of the minimization problem allows us to explore the sharpness of the multiplicative “uncertainty principle” (2). For an interval, we find that multiplicative inequalities are optimal (Sect. 3). The sharp inequality for $p \geq 1$ is

$$W_p(f_+, f_-) \cdot |Z(f)| \geq 2^{-1-\frac{1}{p}} \left(\frac{\|f\|_1}{\|f\|_{\infty}} \right) \|f\|_1^{\frac{1}{p}}.$$

We further show that the same conclusion is not limited to an interval domain but also remains valid for a circle (see Sect. 4). This, in turn, leads us to seek for an example of domains, e.g., a metric star graph, for which the sharp lower bound is not expected to be multiplicative (see Sect. 5).

- **Metric star graphs** We carry our strategy to the more complicated case of metric star graphs (e.g., we need to solve Kontorovich’ problem and not Monge’s); indeed, as before, Theorem 13 reduces the general minimization problem over $L^1 \cap L^{\infty}$ into a finite-dimensional constrained optimization problem. The key difference, however, is that in contrast to the line and the circle, the optimal relation between the minimal transport cost and the other factors under consideration is *not multiplicative* on star graphs. The introduction of a node yields optimal lower bounds which depend on the geometry of the domain. For example, in the case of a star with $D \in \mathbb{N}$ sufficiently long edges,

$$W_1(f_+, f_-)\tilde{N} \geq \frac{1}{4} \frac{\|f\|_1}{\|f\|_{\infty}} \|f\|_1, \quad \tilde{N} = |Z(f)| - 1 + \begin{cases} \frac{D}{2}, & D \text{ even,} \\ \frac{(D+1)(D-1)}{2D} & D \text{ odd.} \end{cases}$$

When some of the edges are not sufficiently long, even more complicated forms of inequalities emerge. Our conclusions for star graphs echo and contrast the analysis of Cavalletti and Farinelli (2021), where the multiplicative (2) lower bound is proved for a broad class of non-branching spaces. Since stars, and metric graphs in general, are branching spaces (see also Erbar et al. (2021)), one might expect a different type of results, as we indeed prove in Sect. 5.

The strategy and issues identified in this work are expected to be generalizable to high dimensional settings. The one-dimensional settings help establish the strategy in its simplest form, and allow us to reach easy-to-compute and precise constants. Furthermore, by considering graphs, we are able to expose the key challenges in going beyond all previous works on non-branching spaces. In Sect. 6, we outline the conjectures and key challenges in generalizing our strategy to multi-dimensional domains, as well as to general metric graphs.

1.2 Structure of the Paper

Preliminaries, definitions, and the formulation of our minimization problem (Question 1) are given in Sect. 2. Then, the problem is solved for the interval (Theorem 5) and the circle (Theorem 7) in Sects. 3 and 4, respectively. For star graphs, in Sect. 5 we re-formulate Question 1 as a finite-dimensional constrained minimization problem (Theorem 13) and solve it for a number of special cases. Finally, an outlook on the problem in multiple dimensions and on general metric graphs is presented in Sect. 6.

2 Settings

2.1 Nodal Sets

Given a one-dimensional domain I , constants $c_\infty, c_1 > 0$, and $N \in \mathbb{N}$, define

$$X = X(c_\infty, c_1, N, I) \equiv \left\{ f : I \rightarrow \mathbb{R} \text{ measurable such that} \begin{array}{l} \|f\|_\infty = c_\infty, \\ \|f\|_1 = c_1, \\ |Z(f)| = N, \\ \int_I f(x) dx = 0. \end{array} \right\}, \quad (3)$$

where $Z(f)$ is the set of points where f changes its sign, defined as follows:

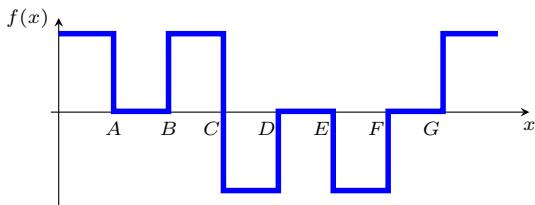
Definition 1 (Effective nodal set) Consider a measurable $f : I \rightarrow \mathbb{R}$ with finite L^1 and L^∞ norms.

- Define

$$Z_1(f) \equiv \partial\{x \mid f(x) > 0\} \cap \partial\{x \mid f(x) < 0\}.$$

- Let $\mathcal{Z}(f) \equiv \pi_0(f^{-1}(\{0\}))$, the set of all connected components of $f^{-1}(\{0\})$.
- Let $\mathcal{Z}'(f) \subseteq \mathcal{Z}(f)$ be the set of all elements of $\mathcal{Z}(f)$ whose closure intersects both $\partial\{x \mid f(x) > 0\}$ and $\partial\{x \mid f(x) < 0\}$.

Fig. 2 Under Definition 1, the portrayed function has $|Z(f)| = 2$



- Let $Z_2(f)$ be a set which consists of a unique point $x \in J$ for every $J \in \mathcal{Z}'(f)$.
- Finally, define the effective nodal set as $Z(f) \equiv Z_1(f) \cup Z_2(f)$.

Our intention is to define $Z(f)$ in a way which exactly captures every sign change of f once. Let us see that this is indeed achieved by the definition: consider, for example, the function in Fig. 2. While $f = 0$ on $[A, B]$, and so $[A, B] \in \mathcal{Z}(f)$, it is not an element of $\mathcal{Z}'(f)$, since it does not intersect with $\partial\{f < 0\}$. Hence, no point in $[A, B]$ is contained in $Z(f)$. The same is true for $[D, E]$. By definition, $C \in Z_1(f)$. Lastly, $[F, G] \in \mathcal{Z}'(f)$, and therefore $F \in Z_2(f)$. Overall, f changes its sign exactly twice and indeed $|Z(f)| = |\{C, F\}| = 2$.

We contrast our definition with those used in previous works:

- In Sagiv and Steinerberger (2020) and Steinerberger (2020), the objects of study are continuous functions, and therefore the interface between the supports of f_+ and f_- is always contained in the set $f^{-1}(\{0\})$. For general measurable functions, therefore, Definition 1 of the effective nodal set *need not* coincide with $f^{-1}(\{0\})$ for $f \in C^0(I)$ (consider e.g., $f(x) = x^2$ on $[-1, 1]$). Moreover, for functions which are merely bounded but not continuous, these interfaces are not necessarily in $f^{-1}(\{0\})$, and hence the somewhat more complicated form of Definition 1.
- The authors of Cavalletti and Farinelli (2021) considered the quantity $\text{Per}(\{x \mid f(x) > 0\})$. First note that, in Euclidean settings, the Perimeter coincides with the Hausdorff measure \mathcal{H}^{d-1} of the reduced boundary of $\{f > 0\}$, and in particular for $d = 1$ this is just the cardinality of the set. Furthermore, the zero set in Cavalletti and Farinelli (2021), $\partial\{x \mid f(x) > 0\}$, includes points where f does not change signs, e.g., A and B in Fig. 2. This is not an issue when seeking lower bounds of the type (2), as adding points which are not interfaces between f_+ and f_- can only *increase* the left-hand side of (2). Since we solve minimization problems on functions with exactly N sign changes, it is easier if we avoid such issues, even at a cost of a slightly more elaborated definition.

2.2 Optimal Transport and the Wasserstein Distance

We briefly recall the definition and certain key properties of the *Wasserstein- p distance*. We refer to Santambrogio (2015) and Villani (2003) for a more comprehensive treatment of this topic. Let $p \geq 1$, $\Omega \subseteq \mathbb{R}^n$ a Borel set, and denote by $\mathbb{P}_p(\Omega)$ the set of all Borel probability measures on Ω with finite p -th moments. Define the Wasserstein- p distance between two measures $\mu_1, \mu_2 \in \mathbb{P}_p(\Omega)$ as

$$W_p(\mu_1, \mu_2) \equiv \inf_{\gamma \in \Gamma(\mu_1, \mu_2)} K_p^{\frac{1}{p}}(\gamma), \quad K_p(\gamma) \equiv \int_{\Omega \times \Omega} |x - y|^p \, d\gamma(x, y), \quad (4)$$

where $|x - y|$ is the geodesic distance on I and $\Gamma(\mu_1, \mu_2)$ is the set of all Borel probability measures on $\Omega \times \Omega$ with marginals μ_1 and μ_2 , i.e.,

$$\mu_1(A) = \gamma(\Omega \times A), \quad \mu_2(A) = \gamma(A \times \Omega), \quad (5)$$

for any $\gamma \in \Gamma(\mu_1, \mu_2)$ and any Borel set $A \subseteq \Omega$.

A measure γ is often called a *transport plan* and γ minimizing K_p is said to solve the *Kantorovich problem*. In some cases, there exists an optimal transport *map*, a function $T : \Omega \rightarrow \Omega$ which solves the so-called *Monge problem*:

$$\inf_{\{T \mid T_{\#}\mu_1 = \mu_2\}} K_p^{\frac{1}{p}}(T), \quad K_p(T) \equiv \int_{\Omega} |x - T(x)|^p \, d\mu_1(x), \quad (6)$$

where by $T_{\#}\mu_1$ we mean the pushforward of μ_1 by T , i.e., the measure which assigns to any Borel set $A \subseteq \Omega$ the measure $T_{\#}\mu_1(A) = \mu_1(T^{-1}(A))$. Any map that pushes μ_1 to μ_2 induces a transport plan $(\text{id}, T)_{\#}\mu_1 \in \Gamma(\mu_1, \mu_2)$, and the transport cost is unambiguously defined, i.e., we can write $K_p(T)$ as a shorthand for $K_p((\text{id}, T)_{\#}\mu_1)$.

On an *interval* $I \subseteq \mathbb{R}$, for any $p \geq 1$ and any two atomless measures, an optimal transport plan is induced by an optimal transport *map* (Santambrogio 2015, Theorem 2.9), i.e., $W_p^p(\mu_1, \mu_2) = K_p(T)$, with a monotonically increasing T defined by

$$T = F_{\mu_1}^{-1} \circ F_{\mu_2}, \quad (7)$$

where $F_{\mu}(y) \equiv \mu(-\infty, y)$ is the cumulative distribution function (CDF), and the inverse is taken in the generalized sense, $F_{\mu}^{-1}(x) \equiv \inf\{t \in \mathbb{R} \mid F_{\mu}(t) \geq x\}$.³ For $p > 1$, this map is also unique (Santambrogio 2015). Hence, the Wasserstein- p distance has the much simpler form

$$W_p(\mu_1, \mu_2) = \|F_{\mu_1}^{-1} - F_{\mu_2}^{-1}\|_{L^p(\mathbb{R})}. \quad (8)$$

In the particular case of $p = 1$, one gets the more straightforward formula with the CDFs (and not their inverses) $W_1(\mu_1, \mu_2) = \|F_{\mu_1} - F_{\mu_2}\|_{L^1(\mathbb{R})}$, see Salvemini (1943) and Vallender (1974).

Our main question can be rigorously formalized as

Question 1 For a one-dimensional domain I with nonzero measure $|I| > 0$, $p \geq 1$, $c_{\infty} > 0$, $c_1 \in (0, c_{\infty}|\Omega|]$, and $N \in \mathbb{N}_+$, what are the minimizers and the minimum value of the minimization problem

$$\min_{f \in X(c_{\infty}, c_1, N, I)} W_p(f_+, f_-). \quad (9)$$

³ The statement holds more generally for the optimal transport with respect to any convex cost function $h(x - y)$ on the line, but we will not pursue this level of generality here.

By $W_p(f_+, f_-)$, we mean the W_p -distance between the measures whose densities are f_+ and f_- . Note that $c_1 \in (0, c_\infty |\Omega|]$ and $N > 0$ are specified in the statement of Question 1 as necessary and sufficient condition for $X(c_\infty, c_1, N, I) \neq \emptyset$.

A crucial ingredient to study the above minimization problem is to establish an equivalence formulation that provides a characterization of the minimizers to the original problem. Connected to this, we present a sub-class of functions, defined below:

Definition 2 Let I be a one-dimensional domain (a curve or a metric graph). Denote by $X_s = X_s(c_\infty, c_1, N, I)$ the set of step-functions $f \in X(c_\infty, c_1, N, I)$ such that $f = \pm c_\infty$ only on intervals adjacent to points in $Z(f)$ and 0 everywhere else; see, e.g., Figs. 1 and 3B.

3 The Interval

In this section, we study the case of a nonempty interval $I = (0, L)$ with $L > 0$. Our strategy to answer Question 1, here and throughout this paper, is that of *optimization*; for every candidate function $f \in X$, we attempt to construct $g \in X$ such that $W_p(f_+, f_-) > W_p(g_+, g_-)$. The minimizers will therefore be the only functions for which further optimization is not possible, and we will show that, by construction, those are also global minimizers.

Lemma 2 Let $I = (0, L)$. For every $f \in X = X(c_\infty, c_1, N, I)$, denote $Z(f) = \{z_1, \dots, z_N\}$ with $z_i < z_{i+1}$ for all $1 \leq i < N$. Then, there exists a function $g \in X_s$ such that

- (i) $Z(f) = Z(g)$
- (ii) For any $z_i \in Z(g)$, $g \geq 0$ on $I_i = (z_i, z_{i+1})$ if and only if $f \geq 0$ there, and

$$\int_{I_i} g(x) \, dx = \int_{I_i} f(x) \, dx.$$

- (iii) $W_p(f_+, f_-) \geq W_p(g_+, g_-)$ for any $p \geq 1$, with strong inequality if $f \notin X_s$.

Remark Even though the optimal transport plan in this case is given by the monotonic map (7), we will work in this proof with a general coupling $\gamma \in \Gamma(f_+, f_-)$. This level of generality shows that at least the geometric nature of the problem extends to the following more general setting: let Ω be a simple curve, $c(x, y) = h(|x - y|)$ where $h : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is a monotonically increasing function in the geodesic distance for two points $x, y \in \Omega$, and define the c -transport cost as

$$K_c(T) = \int_{\Omega \times \Omega} c(x, T(x)) \, d\mu(x). \quad (10)$$

Then, defining the optimal transport with respect to $c(x, y)$ analogously to (4), Lemma 2 still holds. Also, it will allow us to generalize this statement immediately to star graphs in Sect. 5.

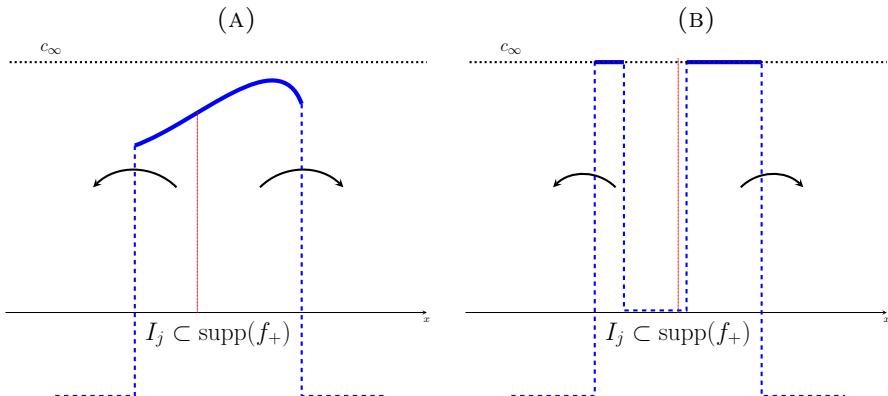


Fig. 3 **A** For an interval I_j where $f(x) \geq 0$ and a given transport plan $\gamma \in \Gamma(f_+, f_-)$, the mass left of the dotted lines will be transported to intervals I_ℓ with $\ell < j$, and the mass to the right of the dotted to intervals I_ℓ with $\ell > j$. **B** The cost of transport for the portrayed re-organization of f on I_j is cheaper

Proof Consider an optimal transport plan between f_+ and f_- , i.e., a measure $\gamma \in \Gamma(f_+, f_-)$ such that $K_p(\gamma) = W_p^p(f_+, f_-)$, see Eq. (5). The intuition is that for each $0 \leq j \leq N$, some of the mass on I_j has to be transported to the left by γ , and some has to be transported to the right, see Fig. 3a.⁴ It is therefore less costly to have those respective masses already concentrated near the endpoints of I_j , see Fig. 3b.

To prove the above intuition rigorously, we will inductively define a sequence of functions in the following way: let $f^0 \equiv f$ and $\gamma^0 \equiv \gamma$. For $1 \leq j \leq N$, we will define a function f^j and a new transport map $\gamma^j \in \Gamma(f_+^j, f_-^j)$ such that they satisfy conditions (i–iii) of Lemma 2 on the intervals I_0, \dots, I_{j-1} and such that $K_p(\gamma_{j-1}) \geq K_p(\gamma_j)$ for any $p \geq 1$.

Since f has a definite sign on each interval, suppose without loss of generality that f is non-negative on I_j (the negative case is completely symmetric). Define $p^j \equiv \gamma^j(I_j \times \cdot)$; this is a Borel measure on I which specifies how much mass from I_j is transported in I under γ^j .

Since the functions f_+ and f_- (thus the functions f_+^j and f_-^j) have disjoint supports, γ^j does not transports from I^j into itself, and so $p^j(I_j) = \gamma^j(I_j \times I_j) = 0$. There exist two constants $\lambda_j, \varrho_j \geq 0$, such that (identifying $z_0 = 0$ and $z_{N+1} = L$)

$$p^j((z_0, z_j)) = \lambda_j, \quad p^j((z_{j+1}, z_{N+1})) = \varrho_j.$$

These are the masses transported from I_j to its left and right, respectively.

⁴ The red dotted line, separating between the left and right transported parts in I_j , is vertical in Fig. 3a, since the optimal transport in this case is given by a map. In the case when the optimal transport is given by a coupling, this separation would be better depicted by a curve $h(x)$ with $0 \leq h(x) \leq f_+(x)$.

We construct a new function $f^{j+1} \in X$ with $Z(f^j) = Z(f^{j+1})$ as follows⁵:

$$f^{j+1}(x) \equiv \text{sign}(f)(x) \cdot \begin{cases} c_\infty, & x \in \left(z_j, z_j + \frac{\lambda_j}{c_\infty}\right), \\ 0, & x \in \left(z_j + \frac{\lambda_j}{c_\infty}, z_{j+1} - \frac{\varrho_j}{c_\infty}\right), \\ c_\infty, & x \in \left(z_{j+1} - \frac{\varrho_j}{c_\infty}, z_j\right) \\ f^j(x), & \text{otherwise.} \end{cases}$$

Clearly f^{j+1} satisfies conditions (i)–(ii) of Lemma 2 on the interval I_j , and by induction on the intervals I_0, \dots, I_{j-1} as well. Since the mass transported from I_j to the left is now concentrated on $(z_j, z_j + \lambda_j/c_\infty)$ as much as possible (with density $= c_\infty$), one can transport this mass to the left at a lower cost, by Definition (4). The same holds for the transport out of I_j to the right. The transport from any other positive interval is defined identically to γ^j . The resulting γ^{j+1} , the optimal transport plan between f_+^{j+1} and f_-^{j+1} , satisfies $K_p(\gamma^j) \geq K_p(\gamma^{j+1})$ for any $p \geq 1$. This is because, under the new optimal transport plan, some of the mass is transported over a shorter distance, and no mass is transported over a longer distance.

Moreover, note that if $f^{j+1} \neq f^j$, i.e., if the construction really did change the function (and so also $\gamma^j \neq \gamma^{j+1}$), then a nonzero mass is now transported over a shorter distance, and so a strict inequality holds $K_p(\gamma^j) > K_p(\gamma^{j+1})$.

Finally, we set $g \equiv f^{N+1}$. □

Lemma 2 implies that for any $p \geq 1$

$$\min_{f \in X(c_\infty, c_1, N, I)} W_p(f_+, f_-) \geq \min_{f \in X_s(c_\infty, c_1, N, I)} W_p(f_+, f_-).$$

And furthermore the minimum on the left hand side is attained only on $X_s(c_\infty, c_1, N, I)$. Hence, Question 1 reduces as follows:

Corollary 3 *For $I = (0, L)$, the two minimization problems*

$$\min_{f \in X(c_\infty, c_1, N, I)} W_p(f_+, f_-) \quad \text{and} \quad \min_{f \in X_s(c_\infty, c_1, N, I)} W_p(f_+, f_-),$$

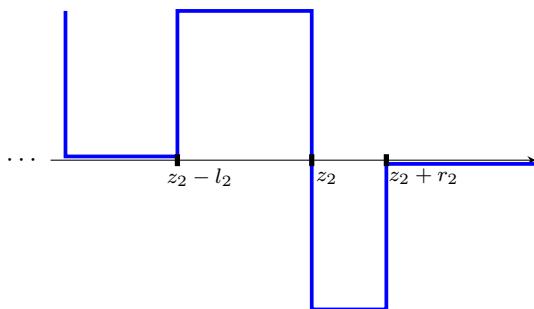
have the same minimizers and the same minimum value for any $p \geq 1$.

Corollary 3 allows us consider the minimizers in Question 1 only from X_s , simply by shifting the mass without changing the nodal set itself. To further reduce the problem, we will now allow for non-local shifts: we will move mass across sub-intervals, which may also change the nodal set (but not its size). To ensure that these non-local shifts reduce the transport cost, we will make use of the monotonicity properties of (7).

Lemma 4 *Let $I = (0, L)$ and $f \in X_s(c_\infty, c_1, N, I)$. There exists a function $g \in X_s$ such that $W_p(f_+, f_-) \geq W_p(g_+, g_-)$, with the following property: denoting the optimal transport map (7) of g by $T = T[g]$, then T only transport mass between*

⁵ Since the nodal set $Z(f^j)$ is independent of j , we write z_j unambiguously.

Fig. 4 A nodal point z_2 and the supports of f_+ and f_- to its left and right, with $z_2 - l_2$ and $z_2 + r_2$ annotated



adjacent intervals, i.e., $T(I_j) \subseteq I_{j-1} \cup I_{j+1}$ for every $I_j = (z_j, z_{j+1})$ where $\{z_1, \dots, z_N\} = Z(g)$, $z_0 = 0$, and $z_{N+1} = L$.

Proof Suppose without loss of generality that f is nonnegative on $I_0 = (z_0, z_1)$. By definition, f is characterized by $3N$ nonnegative numbers, $\{z_i, l_i, r_i\}_{i=1}^N$, such that $f = (-1)^{i+1}c_\infty$ on $(z_i - l_i, z_i)$, $f = (-1)^i c_\infty$ on $(z_i, z_i + r_i)$, and $f = 0$ everywhere else,⁶ see Fig. 4.

We will again construct a sequence of functions $f^j \in X_s(c_\infty, c_1, N)$, now characterized by $\{z_i^j, r_i^j, l_i^j\}_{i=1}^N$, such that

- (i) $K_p(T^j) \geq K_p(T^{j+1})$, where T^j is the monotone optimal transport map (7) associated with $W_p(f_+^j, f_-^j)$.
- (ii) $T^{j+1}(I_k) \subseteq I_{k+1} \cup I_{k-1}$ for $k = 0, \dots, j$

Set $f^0 = f$. Suppose without loss of generality that f^j is nonnegative on I_j . By Lemma 2, the mass in $(z_j^j, z_j^j + r_j^j)$ is transported to the left, i.e., to I_0, \dots, I_{j-1} . By the induction assumption, in the j -th step all of the mass in intervals with index less than j is transported to adjacent intervals. Hence our inductive construction only needs to consider mass transported from I_j to the right.

Assume $T^j = T$ transports an interval $E \subseteq (z_{j+1} - l_{j+1}^j, z_{j+1}) \subseteq I_j$ to a non-adjacent sub-interval to the right of I_j , i.e., $T(E) \subseteq (z_{j+3}^j, L)$ but $T(E) \cap I_{j+1} = \emptyset$.⁷

For simplicity, suppose further that E maps into a single interval, i.e., $T(E) \subseteq I_i$ with $i > j + 2$; Otherwise, if T only maps a part of E into I_i , one can split E into different sub-intervals each mapping into a distinct interval.

To construct f^{j+1} , we would like to perform a **shift operation**, that is, to *shift* $T(E)$ into I_{j+1} . To make this more precise, consider first the ideal situation where $f^j = 0$ on some subset $\tilde{E} \subseteq I_{j+1}$ of equal length, i.e., $|\tilde{E}| = |E|$. In loose terms, it means that there is space for E to be shifted into I_{j+1} , and so we would set

⁶ Since $f \in X_s$, behaviors such as the interval $[F, G]$ in Fig. 2 are excluded.

⁷ Since f^j is non-negative on I_j , it is also non-negative on I_{j+2} , and therefore the next-nearest interval on which f is non-positive is $I_{j+3} = (z_{j+3}, z_{j+4})$.

$$f^{j+1}(x) = \begin{cases} -c_\infty, & x \in \tilde{E}, \\ 0, & x \in T(E) \\ f^j(x), & \text{otherwise.} \end{cases} \quad (11)$$

In particular, in this case $Z(f^j) = Z(f^{j+1})$.

The above construction, however, might not be possible; it might be that the interval I_{j+1} is already full, by which we mean that if $(z_{j+1}^j + r_{j+1}^j, z_{j+1}^j - l_{j+2}^j) \subset I_{j+1}$, the sub-interval on which $f = 0$ is shorter than $T(E)$. Then the mass of $T(E)$ cannot be shifted there (and f^{j+1} cannot be defined as in (11)), since f is a step function of height $\pm c_\infty$, and so cannot exceed this value and stay in $X_s(c_\infty, c_1, N)$. If this is the case, we need to *push some or all of the points* z_{j+2}^j, \dots, z_i^j to the right by up to $|E|$, and then we can repeat the above construction as in (11), with the shifted intervals and nodal points.

The shift operation is depicted on Fig. 5. Let us consider the effect of the shift operation on the overall transport cost $W_p(f_+^{j+1}, f_-^{j+1})$:

- The mass $|E|c_\infty$, which was previously transported between from $E \subseteq I_j$ to $T^j(E) \subseteq I_i$, is now transported over as shorter distance, to $T^{j+1}(E) \subseteq I_{j+1}$.
- Suppose that for $\ell > j$, the nodal endpoint z_ℓ^j was pushed to the right. By the inductive construction, the transport from/to I_ℓ could not have been from a point *to the left* of z_j . If it was transported *to/from* a point *to the right* of I_i , then the overall transport distance decreased.
- Suppose that for $\ell > j$, the nodal endpoint z_ℓ^j was pushed to the right. Suppose without loss of generality that $I_\ell \subseteq \text{supp}(f_+)$ (the negative case is analogous) and that for some $D \subseteq I_\ell$, we have $T^j(D) \subseteq I_m \subset \text{supp}(f_-)$ with $\ell > m > j$; the proposed shift would then *increase* the transport distance, by potentially pushing D away from $T(D)$. This scenario, however, is *impossible* due to the monotonicity of T^j , see (7): take two points $x \in E$ and $x' \in D$, then $x < x'$ but we $T(x) > T(x')$ (since $i > \ell$), hence a contradiction.

We note here that our construction cannot change the order of points, i.e., $z_k^{j+1} \leq z_{k+1}^{j+1}$. Similarly, the construction keeps the nodal points inside I , i.e., $z_1^{j+1} \geq 0$ and $z_N^{j+1} \leq L$.

We established the first property in the induction, that $K_p(T^j) \geq K_p(T^{j+1})$. Now, note that the new transport map we constructed is monotone and pushes f_+^{j+1} to f_-^{j+1} . Hence, it is an optimal transport map between f_+^{j+1} and f_-^{j+1} (the unique one for $p > 1$), T^{j+1} . By construction, it satisfies the adjacency requirement, that $T^{j+1}(I_k) \subseteq I_{k+1} \cup I_{k-1}$ for all $k \leq j$.

This completes the j -th step. We take $g = f^N$, which completes the construction. Finally, note that unless $f = g$, we strictly reduced the transport cost at some stage of the induction, and so $W_p(f_+, f_-) > W_p(g_+, g_-)$. \square

Remark The proof of Lemma 4 relies on the convexity of the cost function $h(|x - y|) = |x - y|^p$ with $p \in [1, \infty)$. When h is convex, the monotone map (7) is a solution of

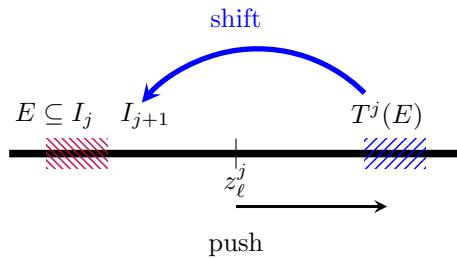


Fig. 5 The shift operation described in the proof of Lemma 4: on the j -th step, some $E \subseteq I_j$ is mapped by T^j to a non-adjacent interval I_i right of E . We shift that mass to I_{j+1} . In the process, some or all of the nodal points in between, e.g., z_ℓ^j with $j+1 < \ell < i$, need to be pushed away (to the right) to “make room” for the shifted mass

Monge’s problem (6), see Santambrogio (2015). This is no longer the case if one considers a concave cost function, e.g., $|x - y|^p$ with $p \in (0, 1)$. For concave costs, it is known that the maps may not be monotone (Gangbo and McCann 1996; McCann 1999), and a different type of analysis will be needed. Another interesting cost function not considered in this work is the L^∞ cost (Barron et al. 2017; Champion et al. 2008).

We conclude that $f \in X_s(c_\infty, c_1, N, I)$ can be a minimizer of the problem (9) for $p \geq 1$ if and only if it has the following structure:

- (i) $Z(f) = \{z_1, \dots, z_N\} \subset (0, L)$.
- (ii) There exist $d_1, \dots, d_N > 0$ such that $z_j + d_j \leq z_{j+1} - d_{j+1}$ for all $j = 1, \dots, N$, $z_1 - d_1 \geq 0$, and $z_N + d_N \leq L$.
- (iii) without loss of generality, assume that $f \geq 0$ on (z_0, z_1) . Then for j odd, $g(x) = c_\infty$ on $(z_j - d_j, z_j)$ and $g(x) = -c_\infty$ on $(z_j, z_j + d_j)$, and vice versa for j even.

In less formal language, a minimizing step-function attains its maximal value c_∞ anti-symmetrically around each nodal point z_j , see Fig. 1. The optimal transport plan $\gamma \in \Gamma(f_+, f_-)$, which is given by the monotone map (7), only transports across each nodal point, and $K_p(\gamma)$ is just the sum of costs accrued at each nodal point. The only questions that now remain concern the distribution of the width parameters d_1, \dots, d_N , and the overall minimal optimal transport cost.

Theorem 5 Let $I = (0, L)$ with $L > 0$, $p \geq 1$, $c_\infty > 0$, $c_1 \in (0, Lc_\infty]$, and $N \in \mathbb{N}_+$. The minimizers of the problem (9) are step functions, anti-symmetric about each nodal point, with value $\pm c_\infty$ and width $c_1/(2c_\infty N)$. These minimizers satisfy

$$\min_{f \in X(c_\infty, c_1, N, I)} W_p(f_+, f_-) = 2^{-\frac{p+1}{p}} \frac{c_1}{Nc_\infty} c_1^{\frac{1}{p}}. \quad (12)$$

Proof By (8), the transport cost across a single nodal point z_j depends on the inverse CDFs.⁸ By Lemmas 2 and 4, it suffices to consider step functions $f \in X_s$ where mass

⁸ Since $F_\mu(y) = \mu(-\infty, y)$ might not be bijective, we define the inverse CDF by $F^{-1}(x) \equiv \inf\{y \in \mathbb{R} \mid F(y) \geq x\}$, see Santambrogio (2015), Section 2.1.

is transported only to adjacent sub-intervals. Suppose without loss of generality that $f = c_\infty$ on $(z_1 - d_1, z_1)$ and $f = -c_\infty$ on $(z_1, z_1 + d_1)$. By definition, $F_+(z_1) = F_-(z_1 + d_1) = c_\infty d_1$. Hence $F_{f_+}^{-1}(c_\infty d_1) - F_{f_-}^{-1}(c_\infty d_1) = d_1$. On the interval (of cumulative probabilities) $(0, c_\infty d_1)$, the inverse CDFs are linear with slope $1/c_\infty$, and so for every $t \in (0, c_\infty d_1)$ the difference between the inverse CDFs is constant, i.e., d_1 . Hence,

$$\left[\int_0^{c_\infty d_1} |F_{f_+}^{-1}(t) - F_{f_-}^{-1}(t)|^p dt \right]^{\frac{1}{p}} = \left[c_\infty d_1^{p+1} \right]^{\frac{1}{p}} = c_\infty^{\frac{1}{p}} d_1^{1+\frac{1}{p}}. \quad (13)$$

Summing up the contributions of all nodal points, then by (8) we have that

$$W_p^p(f_+, f_-) = \sum_{j=1}^N c_\infty d_j^{p+1}. \quad (14a)$$

For simplicity of computations, we will minimize $W_p^p(f_+, f_-)$ (which is equivalent to minimizing $W_p(f_+, f_-)$). This is a constrained minimization problem, under the L^1 constraint

$$\sum_{j=1}^N d_j = \frac{c_1}{2c_\infty}. \quad (14b)$$

By the method of Lagrange multipliers, let

$$\mathcal{L} = c_\infty \sum_j d_j^{p+1} - \lambda \left(\sum_j d_j - c_1/2c_\infty \right).$$

We get from the condition $\partial_{d_j} \mathcal{L} = 0$ that

$$\lambda = (p+1)c_\infty d_j^p, \quad 1 \leq j \leq N. \quad (15)$$

The condition $\partial_\lambda \mathcal{L} = 0$ yields

$$\frac{c_1}{2c_\infty} = \sum_{j=1}^N d_j = N \left[\frac{\lambda}{(p+1)c_\infty} \right]^{\frac{1}{p}},$$

which leads to,

$$\lambda = (p+1)c_\infty \left(\frac{c_1}{2Nc_\infty} \right)^p.$$

Combining the above expressions of λ together, we get $d_j = c_1/2Nc_\infty$, and the overall cost is, by (14a),

$$W_p^p(f_+, f_-) = Nc_\infty \left(\frac{c_1}{2Nc_\infty} \right)^{p+1} = \frac{1}{2^{p+1}} \left(\frac{c_1}{Nc_\infty} \right)^p c_1.$$

□

As a consequence, we have

Corollary 6 *Let $I = (0, L)$ and $p \geq 1$. For any $f \in L^\infty(I)$ we have*

$$W_p(f_+, f_-) \cdot |Z(f)| \geq 2^{-1-\frac{1}{p}} \left(\frac{\|f\|_1}{\|f\|_\infty} \right) \|f\|_1^{\frac{1}{p}}. \quad (16)$$

This inequality is sharp, where equality holds if and only if f is a minimizer of (9) as described in Theorem 5.

We thus see that for the case of $I = (0, L)$, we have a sharp inequality (16) in a multiplicative form. The inequality shows no dependence on the length of L , which can also be seen from the scaling properties of the quantities involved. Moreover, we are able to characterize an explicit constant factor which is also the best possible. Our result establishes that for $p = 1$, the inequality in Cavalletti and Farinelli (2021), Prop. 3.1 is (16), and therefore it is indeed sharp. For $p \geq 1$, the scaling in (De Ponti and Farinelli 2022, Corollary 3.3) of the lower bound is the same as here, $2^{-1-1/p}$, but the overall constant is lower since it is proven in more general settings. We may attribute the proof of the inequality (16) not only to properties of the special geometry in one dimension, but also the way the minimization problem (9) is posed; it is the solution to the latter that leads to, as a by-product, the most natural “uncertainty principle” in a multiplicative form.

4 The Circle

We now consider the minimizers of $W_1(f_+, f_-)$ in $X = X(c_\infty, c_1, N, I)$ for the case where I is a circle. As we see from the earlier discussion on the case of an interval, by scaling of c_1 we can restrict our attention to the unit circle $I = S^1$. The geodesic distance between two points $e^{it}, e^{is} \in S^1$ is defined by

$$d(e^{it}, e^{is}) = \min\{(t - s) \bmod(2\pi), (s - t) \bmod(2\pi)\}.$$

For every $z_j \in Z(f)$ we denote $z_j = e^{is_j}$ such that, without loss of generality

$$0 = s_1 < s_2 < \cdots < s_N < s_{N+1} = 2\pi,$$

i.e., the nodal points are ordered from the x -axis in a counterclockwise direction.

In this section we show that, on a circle, the minimizers and minimal value of (9) are analogous to those on the interval (Theorem 5 and Corollary 6):

Theorem 7 Let $p \geq 1$, $c_\infty > 0$, $c_1 \in (0, 2c_\infty\pi]$, and $N \in \mathbb{N}_+$. The minimizers of $W_p(f_+, f_-)$ over $X(c_\infty, c_1, N, S^1)$ are step functions in $X_s(c_\infty, c_1, N, S^1)$, anti-symmetric about each nodal point, with value $\pm c_\infty$ and width $c_1/2c_\infty N$. Hence

$$\min_{f \in X(c_\infty, c_1, N, S^1)} W_p(f_+, f_-) = 2^{-1-\frac{1}{p}} \frac{c_1}{Nc_\infty} c_1^{\frac{1}{p}}.$$

Proof Throughout this proof, we use the existence of an optimal transport map; for all orders $p \geq 1$ there exists such a map $T : S^1 \rightarrow S^1$, as established for $p > 1$ in see McCann (2001) or Villani (2003), Theorem 2.47, and for $p = 1$ in Feldman and McCann (2002) (see also Santambrogio (2015), Section 3.1 and Caffarelli et al. (2002) and Bianchini and Cavalletti (2013) for general metric settings). A more elaborate analysis of optimal transport maps on the circle appears in Delon et al. (2010).

First, we note that the proof of Lemma 2 carries to the circle without change: it is an iterative process done on f in each interval (arch) $\{e^{is} \mid s \in (s_j, s_{j+1})\}_{j=1}^N$. Hence, we can restrict our attention to solving the optimal-transport minimization problem on $X_s(c_\infty, c_1, N, S^1)$, see Definition 2.

Next, we turn to extend Lemma 4 to the circle, i.e., to show that a function $f \in X_s$ has a cheaper transport cost if the optimal transport map associated with it only transports mass between adjacent arcs. Here lies the main new challenge: we cannot simply implement our inductive “shifting” strategy from Lemma 4, since there are no end-points to the circle, and hence no natural candidate arc I_j to start the induction from. To this end, we prove the following lemma:

Lemma 8 Let $f \in X_s$ and let T be the optimal transport map associated with $W_p(f_+, f_-)$ for a fixed $p \geq 1$. Then, there exists a partition $S^1 = J_1 \cup \dots \cup J_K$ into disjoint arcs such that for each arc J_k , either

- (1) All points $x \in J_k \cap \text{supp}(f_+)$ are transported clockwise to $T(x) \in J_k$, or
- (2) all points $x \in J_k \cap \text{supp}(f_+)$ are transported counterclockwise to $T(x) \in J_k$, or
- (3) $f = 0$ on J_k .

To prove Lemma 8, we will rely on the existence of an optimal transport map (in the sense of Monge). For $p > 1$, recall the following theorem due to McCann (2001) (for the Euclidean case, see Brenier (1991)), presented here in a simplified form:

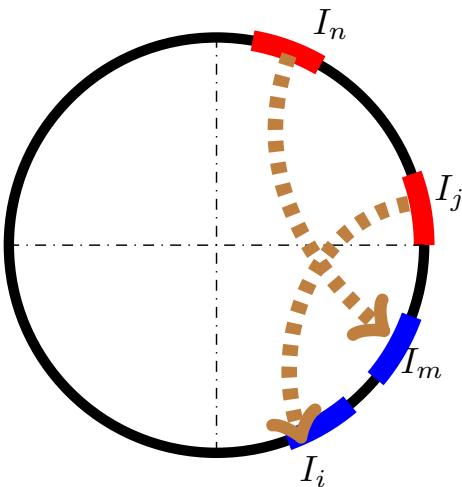
Theorem (McCann (2001)) Let Ω be a C^3 connected, compact, Riemannian manifold without boundaries. Let $p > 1$ and consider two Borel probability measures μ and ν with finite p -moments, such that μ is absolutely continuous with respect to the volume measure of Ω . Then, with respect to the Wasserstein- p distance (4), there exists an optimal transport map T . Moreover, there exists a vector field V such that

$$T(x) = \exp_x[V], \quad (17)$$

where \exp is the exponential map with respect to the geodesic distance.

Remark McCann’s theorem holds for the general class of optimal transport problems with respect to K_c (see (10)) for a strictly convex cost function $c(x, y) = h(|x -$

Fig. 6 The “lagging” scenario as defined in the proof of Lemma 8: red intervals are in the support of f_+ , Blue intervals are in the support of f_- . The two dash-dotted lines are equator lines: They indicate that e.g., I_j would transport to I_i counterclockwise under T



$y|$). Furthermore, the vector-field V is characterized in terms of the gradient of the Kantorovich potential, see McCann (2001), Theorem 13 for details.

The case of $p = 1$ is similar: the circle decomposes into a union of geodesics lines (arcs), known as “transport rays”, which intersect (potentially) only at their endpoints. On each such transport ray, the optimal transport map is monotone; see details in Santambrogio (2015), Section 3.1 and Feldman and McCann (2002).

Proof of Lemma 8 First consider $p > 1$. Choose any $x \in S^1$, and suppose first that $y = T(x)$ is transported clockwise from x , i.e., y is clockwise to x on the shorter arc between the two points. Then, any point $w \in S^1 \cap \text{supp}(f_+)$ lying on that arc is also transported clockwise, since the vector field V points clockwise on that arc. Hence, the set of points $x \in S^1$ for which T transports clockwise is a union of arcs, and so each J_k is a connected component of that set. If $y = T(x)$ is counterclockwise, the proof is analogous, and together these type of arcs cover $\text{supp}(f_+) \cup \text{supp}(f_-)$. The remaining points on S^1 are those where $f = 0$, and by the extension of Lemma 2 to the circle, it too is covered by disjoint arcs. This completes the proof for $p > 1$.

For $p = 1$, the decomposition of the circle into transport rays, on each of which the optimal-transport map is monotone, yields an analogous proof. \square

Proof of Theorem 7 - continued: Suppose first that $J_1 \neq S^1$, i.e., it is not the case that all points $x \in \text{supp}(f_+)$ are transported clockwise (or counterclockwise). Then, on each arc J_k we can apply the analysis from the case of the interval. Suppose points are transported clockwise on J_k . Therefore the counterclockwise end of J_k has to be in $\text{supp}(f_+)$, and we can choose it as the starting point of the inductive process in Lemma 4.

If, on the contrary, $J_1 = S^1$ assume that without loss of generality, all points $x \in \text{supp}(f_+)$ are transported counterclockwise. If each $x \in \text{supp}(f_+)$ is transported to the adjacent interval, the proof is completed. Assume otherwise, i.e., that we are in a “lagging” scenario as in Fig. 6. We claim that this cannot be the optimal transport

map for a minimizer of $W_p(f_+, f_-)$. This can be shown by constructing $g \in X_s$ for which $W_p(g_+, g_-) < W_p(f_+, f_-)$ as follows:

Let $Z(f) = \{z_1, \dots, z_N\}$ be the nodal points arranged in a counterclockwise order, with an arbitrary starting point. As before, denote by I_j the arcs between z_j and z_{j+1} , where I_N is the arc between z_N and z_1 . Assume without loss of generality that $I_1 \subseteq \text{supp}(f_+)$, and set I_1 for g to be the same as for f . Adjacent to I_1 counterclockwise we set a negative arc I_2 precisely of the length $|I_1|$, and we define the new pushforward map $S\#g_+ = g_-$ as the monotonic map from I_1 to I_2 . We do so iteratively - we position positive intervals precisely of the size they had for f , and then a negative interval of the same size. In defining g , the L^1 and L^∞ norms are unchanged, and so is the number of nodal points. While the map S might not be the optimal transport map between g_+ to g_- , the overall transport distances have been reduced and so $W_p(f_+, f_-) = K_p(T) > K_p(S) > W_p(g_+, g_-)$.

Hence, we have extended Lemma 4 to the circle, for all $p \geq 1$. Now, since the transport occurs only across nodal points to adjacent intervals, our Lagrange-multipliers analysis for the case of the interval applies, and we obtain the desired result. \square

Remark The work of Delon et al. (2010) suggests an alternative route to prove Lemma 8 on the circle: “lifting” each measure on the circle to a periodic measure on the line, they study locally optimal transport maps between the “lifted” periodic measure. These measures are similar to $F_\mu^{-1} \circ F_\nu$ up to a shift, and therefore in particular, are monotonic. We do not pursue this strategy further in this work, and refer to Delon et al. (2010) for details.

5 Star Graphs

Given a positive integer D , we define the star graph $S_D = S_D(L_1, \dots, L_D)$ as the quotient space of the disjoint union of D intervals $I_j = [0, L_j]$ where $L_j > 0$ for $1, \dots, D$, under the equivalence relation $0_{I_j} \equiv 0$. For ease of notation, denote by t_j the point $t \in I_j$ for every $1 \leq j \leq D$ and every $t \in (0, L_j]$.

We will call the 0 point the vertex of the star. Definition 1 for $Z(f)$ extends to $I = S_D$. Consequently, the definitions of $X(c_\infty, c_1, N, I)$ (see (3)) and $X_s(c_\infty, c_1, N, I)$ (see Definition 2) naturally extend to the case of star graph $I = S_D$ as well. The distance between any two points $x \in I_j$ and $y \in I_k$ is $|x - y|$ if $j = k$, and $x + y$ otherwise, i.e., if the two points are on different edges of the star, the geodesic distance between them is the length of the path going from x to y through the vertex 0.

Stars are a class of spaces where we might expect the optimal dependence between $W_p(f_+, f_-)$ and the number of nodal points N to be non-multiplicative, for the following reason: In Cavalletti and Farinelli (2021) and De Ponti and Farinelli (2021), the sharp (up to a constant) uncertainty quantification (2) was derived for metric spaces which are essentially non-branching. Intuitively, this means that if $\ell_1(t), \ell_2(t) : [0, 1] \rightarrow \Omega$ are two “generic” geodesic lines of unit length, and $\ell_1(t) = \ell_2(t)$ on an open subset of $[0, 1]$, then $\ell_1 = \ell_2$ everywhere; see Rajala and Sturm (2014) for details. Star graphs, however, are certainly spaces with branching (and so are trees in general). Hence, in search for new types of dependencies

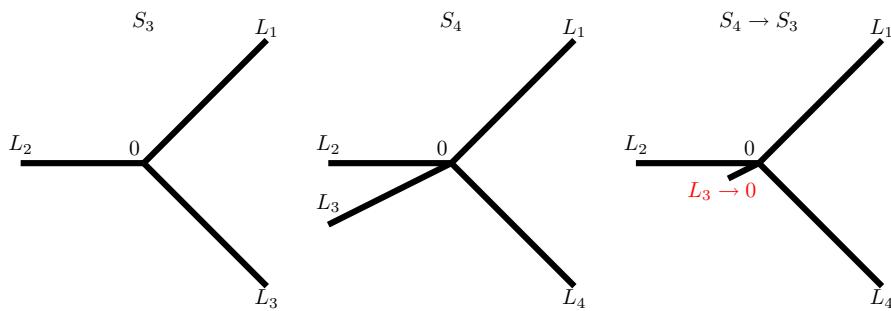


Fig. 7 Left and Center stars have 3 and 4 long edges, respectively. But in a star graph, the lengths of the edges matter: as the length $L_3 \rightarrow 0^+$, S_4 is deformed into S_3 (Right). This metric structure in turn manifests itself in the minimization problem and in its solution, see Theorem 13

between $\min W_p(f_+, f_-)$ and N , stars are excellent candidates over which to study the minimization problem stated in Question 1.

The technical difficulty is that, on the star, we do not have explicit optimal maps such as (7) for the interval, nor do we even expect the existence of optimal transport maps, i.e., we expect a solution to the Kontorovich problem (4), but not Monge's (6). More broadly, stars are an example of metric graphs, on which the study of optimal transport is at a relatively early stage (Erbar et al. 2021; Mazón 2015).

Main results: The key element of our analysis is that it is always “useful”, in the sense of minimizing $W_p(f_+, f_-)$, to position one of the nodal points at the vertex of the star. We approach Question 1 on stars by establishing a correspondence between transport over a star-vertex and transport on the real line (Lemma 9). This equivalence allows us to reduce the minimization problem to a finite-dimensional constrained optimization problem (Theorem 13).

From there, the optimization problem bifurcates into several different cases, depending on both the topology and the lengths of the star's edges. In Sects. 5.2–5.5, we work the details of the following cases:

- For an **even number of sufficiently long edges**, the vertex $0 \in S_D$ is equivalent to $D/2$ nodal points. Hence, we have a multiplicative uncertainty principle of the type (presented here for simplicity with $p = 1$)

$$W_1(g_+, g_-) \geq \frac{c_1^2}{4c_\infty} \frac{1}{\tilde{N}}, \quad \tilde{N} = N - 1 + \frac{D}{2},$$

see Sect. 5.3.

- For an **odd number of sufficiently long edges**, the main complication is that there is an imbalance between the number of positive and negative edges around 0. Nevertheless, we get the same type of inequality, only now with (Sect. 5.4)

$$\tilde{N} \equiv N - 1 + \tilde{D}, \quad \tilde{D} \equiv \frac{(D+1)(D-1)}{2D}.$$

- When **one of the edges is short**, no \tilde{N} or \tilde{D} type inequalities emerge, but the lower bounds we obtain “interpolate” between the case of a star with $D - 1$ long edges (and a “degenerate” D -th edge with length zero) to the case of D long edges (Sect. 5.5).

In summary, for stars with D long edges, the vertex $0 \in S_D$ is effectively equivalent to $D/2$ or $\tilde{D} = \frac{(D+1)(D-1)}{2D}$ nodal points on the line, depending on whether D is even or odd, respectively. Finally, since we can interpolate between D and $D - 1$ edges by shortening/lengthening the edges (see Fig. 7) the general case of a star does not seem to admit such a clean result; indeed, an “uncertainty principle”-type lower bounds on $W_1(f_+, f_-) \cdot |Z(f)|$, such as (16), breaks even in relatively simple metric graphs, thus demonstrating that the non-branching property used in Cavalletti and Farinelli (2021) is indeed necessary.

5.1 Stars – The General Framework

Our strategy to prove the main result, Theorem 13, consists of two parts. In Lemma 9, we analyze the optimal transport problem in the case of a single nodal point on the star vertex. Lemma 11 generalizes Lemma 4 on the optimality of transfer to adjacent sub-interval to the case of the star graph. The main technical difficulty here is that we cannot assume the existence of optimal transport maps (in the sense of Monge’s problem (6), as opposed to Kantorovich (4)), on graphs. Indeed Lemma 9 shows that already in simple settings of $Z(f) = \{0\}$ the optimal transport plan will not be induced by a map. Moreover, we cannot expect to have monotonicity in the strict sense, due to the geometry of the graph. Nonetheless, Lemma 12 shows that the optimal transport plans satisfy a sufficient monotonicity-like property.

We begin with functions satisfying $Z(f) = \{0\}$.

Lemma 9 *Let $D > M \geq 1$ be integers and consider $f \in X(c_\infty, c_1, 1, S_D)$ where for every $x_j \in I_j \subseteq S_D$,*

$$f(x_j) \begin{cases} \geq 0, & 1 \leq j \leq M, \\ \leq 0, & M + 1 \leq j \leq D, \end{cases} \quad (18)$$

Define $g : \mathbb{R} \rightarrow \mathbb{R}$ as⁹

$$g(x) = g(x; f) = g_+(x) + g_-(x) \equiv \sum_{j=1}^M f_j(x) + \sum_{j=M+1}^D f_j(-x) \quad (19)$$

such that g_+ is supported on $(0, \max_{j=1, \dots, M} L_j)$ and g_- is supported on $(-\max_{j=M+1, \dots, D} L_j, 0)$. Then there is a surjective correspondence $\Phi : \Gamma(f_+, f_-) \rightarrow \Gamma(g_+, g_-)$ such that $K_p(\gamma) = K_p(\Phi[\gamma])$ for every $p \geq 1$ and $\gamma \in \Gamma(f_+, f_-)$.

⁹ Here we identify $x \in \mathbb{R}$ with the point $x_j \in I_j$ with the exact same value. This is unambiguous since each f_j is only defined on the respective edge I_j .

Hence,

$$W_p(g_+, g_-) = W_p(f_+, f_-).$$

Proof For every $\gamma \in \Gamma(f_+, f_-)$ and any two measurable sets $E_+, E_- \subseteq (0, \infty)$, define

$$\Phi[\gamma](E_+, -E_-) \equiv \sum_{j=1}^M \sum_{i=M+1}^D \gamma(E_+ \cap I_j, -E_- \cap -I_i), \quad (20)$$

where for every $A \subseteq (0, \infty)$ we define $-A \equiv \{-x \mid x \in A\}$. $\Phi[\gamma]$ is a Borel measure on $\mathbb{R} \times \mathbb{R}$ with marginals g_+ and g_- , simply by additivity. That $K_p(\gamma) = K_p(\Phi[\gamma])$ follows from definition (4): the same amount of mass travels the same distance in both plans. To summarize, we have shown the inclusion

$$\begin{aligned} \{K_p(\gamma) \mid \gamma \in \Gamma(f_+, f_-)\} &= \{K_p(\Phi[\gamma]) \mid \gamma \in \Gamma(f_+, f_-)\} \\ &\subseteq \{K_p(\eta) \mid \eta \in \Gamma(g_+, g_-)\}. \end{aligned}$$

To conclude that $W_p(f_+, f_-) = W_p(g_+, g_-)$ we need to prove inclusion in the other direction, i.e., to find for any $\eta \in \Gamma(g_+, g_-)$ a coupling $\gamma_\eta \in \Gamma(f_+, f_-)$ such that $\Phi[\gamma_\eta] = \eta$. We define γ_η in a symmetric manner: for any two measurable sets $A, B \subseteq (0, \infty)$ and $1 \leq i, j \leq D$,

$$\gamma_\eta(A \cap I_i, B \cap I_j) \equiv \eta(A, -B) \frac{\|f_i\|_{L^1(A \cap I_i)}}{\|g_+\|_{L^1(A)}} \frac{\|f_j\|_{L^1(B \cap I_j)}}{\|g_-\|_{L^1(-B)}},$$

and for any general sets $\mathcal{A}, \mathcal{B} \subseteq S_D$, then

$$\gamma_\eta(\mathcal{A}, \mathcal{B}) \equiv \sum_{i=1}^D \sum_{j=1}^D \gamma_\eta(\mathcal{A} \cap I_i, \mathcal{B} \cap I_j).$$

Again, by definition γ_η is a Borel measure on $S_D \times S_D$. To verify that $\gamma_\eta \in \Gamma(f_+, f_-)$, we compute its marginals. Let $(x, y) \in \text{supp} \gamma_\eta$. Necessarily y is an element in one of the negative intervals I_{M+1}, \dots, I_D , then for any measurable set $A \subseteq (0, \infty)$

$$\begin{aligned} \gamma_\eta(A \cap I_i, S_D) &= \gamma_\eta\left(A \cap I_i, \bigcup_{j=M+1}^D I_j\right) \\ &= \sum_{j=M+1}^D \gamma_\eta(A \cap I_i, I_j) \\ &= \eta(A, (-\infty, 0)) \frac{\|f_i\|_{L^1(A \cap I_i)}}{\|g_+\|_{L^1(A)}} \sum_{j=M+1}^D \frac{\|f_j\|_{L^1(I_j)}}{\|g_-\|_{L^1(-\infty, 0)}} \end{aligned}$$

$$= \eta(A, (-\infty, 0)) \frac{\|f_i\|_{L^1(A \cap I_i)}}{\|g_+\|_{L^1(A)}} \cdot 1 = \|f_i\|_{L^1(A \cap I_i)},$$

where at the last passage we used the fact that g_+ is the density of the second-coordinate marginal of any $\eta \in \Gamma(g_+, g_-)$, i.e., that $\eta(A, (-\infty, 0)) = \|g_+\|_{L^1(A)}$. The calculation of the second-coordinate marginals of γ_η is analogous.

Finally, to prove that $\Phi[\gamma_\eta] = \eta$, consider a pair of open sets $E_+, E_- \subseteq (0, \infty)$. By (20)

$$\begin{aligned} \Phi[\gamma_\eta](E_+, -E_-) &= \sum_{j=1}^M \sum_{i=M+1}^D \eta(E_+, -E_-) \frac{\|f_i\|_{L^1(E_+ \cap I_i)}}{\|g_+\|_{L^1(E_+)} \cdot \|g_-\|_{L^1(-E_-)}} \frac{\|f_j\|_{L^1(E_- \cap I_j)}}{\|g_-\|_{L^1(-E_-)}} \\ &= \frac{\eta(E_+, -E_-)}{\|g_+\|_{L^1(E_+)} \cdot \|g_-\|_{L^1(-E_-)}} \sum_{j=1}^M \|f_j\|_{L^1(E_- \cap I_j)} \sum_{i=M+1}^D \|f_i\|_{L^1(E_+ \cap I_i)} \\ &= \frac{\eta(E_+, -E_-)}{\|g_+\|_{L^1(E_+)} \cdot \|g_-\|_{L^1(-E_-)}} \|f_+\|_{L^1(E_+)} \|f_-\|_{L^1(-E_-)} \\ &= \eta(E_+, -E_-). \end{aligned}$$

□

To solve the minimization problem on S_D , we first state an extension of Lemma 2 to star graphs, where the proof is completely identical to that of Lemma 2:

Lemma 10 *For every $f \in X = X(c_\infty, c_1, N, S_D)$, there exists a function $h \in X_s(c_\infty, c_1, N, S_D)$ such that*

- (i) $Z(f) = Z(h)$
- (ii) *For any sub-interval J between two adjacent nodal points, i.e., $J = (z, z')$ where $z, z' \in Z(f)$ and $(z, z') \cap Z(f) = \emptyset$, then $h \geq 0$ on J if and only if $f \geq 0$ there, and*

$$\int_J h(x) \, dx = \int_J f(x) \, dx. \quad (21)$$

If $0 \notin Z(f)$, (21) also holds when J is a maximal star-like subgraph for which $0 \in J$ but $Z(f)$ is disjoint from the interior of J .

- (iii) $W_p(f_+, f_-) \geq W_p(h_+, h_-)$ for any $p \geq 1$ with equality possible only if $f \in X_s(c_\infty, c_1, N, S_D)$ as well.

We now proceed to prove an analog of Lemma 4 for star graphs; the minimizers of $W_p(f_+, f_-)$ in $X_s(c_\infty, c_1, N, S_D)$ are those where mass is transported only between adjacent sub-intervals. .

Lemma 11 *Let $f \in X_s(c_\infty, c_1, N, S_D)$. There exists $g \in X_s(c_\infty, c_1, N, S_D)$ such that $W_p(f_+, f_-) \geq W_p(g_+, g_-)$ for every $p \geq 1$, with the following property:*

If $\gamma \in \Gamma(g_+, g_-)$ is an optimal transport plan, then γ only transports mass between adjacent intervals. By this, we mean that if J and J' are two closed sub-intervals between two adjacent nodal points (or the smallest star-like subgraph between the nodal points closest to 0), then $\gamma(J \times J') \neq 0$ only if $\partial J \cap \partial J' \neq \emptyset$.

The new challenge in proving Lemma 11 is the absence of a monotonicity property. On the interval, the explicitly optimal transport map (7) is monotone. On the circle, monotonicity is a consequence of the exponential map form of (17) (or of the more specific construction in Delon et al. (2015)). On star graphs, while we cannot hope for monotonicity in the usual sense, we show that a similar property (Lemma 12) is sufficient to demonstrate that the shift operations described in Lemma 4 can be implemented on S_D as well (Lemma 11).

Proof of Lemma 11 We define the following disjoint partition:

$$S_D = J^* \cup \bigcup_{i=1}^D \bigcup_{j=1}^{n(i)} J_{i,j},$$

where on each interval I_i that comprises the star graph S_D , we denote the sub-intervals between adjacent nodal points by $J_{i,1}, \dots, J_{i,n_i} \subset I_i$ where $n_i \geq 0$ for each $1 \leq i \leq D$, and where the subintervals are ordered by *decreasing* distances from 0. Finally, let J^* be the smallest star-like graph between the closest nodal point to 0. If $0 \in Z(f)$, then $J^* = \emptyset$. For ease of notation, it is useful to denote $J^* = J_{i,n(i)+1}$ for all $1 \leq i \leq D$.

We repeat the inductive construction of Lemma 4 (see Fig. 8 for illustration): for each j , we iterate over all $1 \leq i \leq D$ for which $j \leq n(i)$. Suppose without loss of generality that $J_{i,j} \cap \text{supp}(f_+) \neq \emptyset$ and that for some set of nonzero measure $E \subseteq J_{i,j} \cap \text{supp}(f_+)$ we have that $\gamma(E, \cdot)$ is supported on non-adjacent intervals, i.e.,

$$c \equiv \gamma(E, S_D \setminus (J_{i,j-1} \cup J_{i,j+1})) > 0.$$

For each non-adjacent interval where $\gamma(E, \cdot)$ is supported, we will shift that exact same mass to either $J_{i,j-1}$ or $J_{i,j+1}$, depending on the relative order between the intervals.

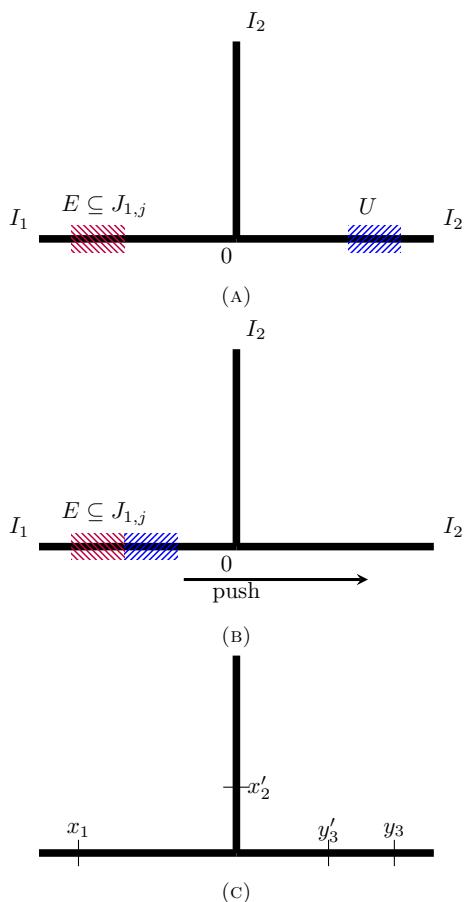
As in Lemma 4, these shifts might require to *push away from* $J_{i,j}$ some nodal points to create room for the negative mass. See Fig. 8 for an illustration. Suppose then that some interval of length A was shifted from $J_{i',j'}$ to $J_{i,j\pm 1}$, and that all or some of the mass between them has been pushed away from $J_{i,j\pm 1}$. To exclude these scenarios, we state a monotonicity lemma (in the spirit of Santambrogio (2015), Theorem 2.9):

Lemma 12 *Let γ be an optimal transport plan with respect to W_p with $p \geq 1$, let $1 \leq i, j, k, l \leq D$, let $x, x', y, y' > 0$, and let $(x_i, y_k), (x'_j, y'_l) \in \text{supp}(\gamma)$. Then the following scenarios are impossible:*

- (1) *i, j, k are distinct, $k = l$, $x' < x$, and $y' < y$.*
- (2) *j, k, l are distinct, $i = j$, $x' < x$, and $y' < y$.*
- (3) *$i = j$, $x' < x$, and $y' < y$ (it may or may not be that $i = k$, i.e., (x_i, y_k) may or may not be part of the same edge).*

Here, as before, for every $1 \leq i \leq D$ and every $x \in (0, L_i)$, we denote by x_i the point on I_i whose distance from the vertex 0 is x . The same notation applies to y_k , x'_j , and y'_l .

Fig. 8 The shifts of Lemma 11 on S_3 . **A** E is being (partially) transported into U under the optimal transport plan γ . **B** After the shift, the relevant mass of U is shifted to the interval adjacent to E , where mass on some of the intervals between E and where U previously was has been shifted to the right. **C** If $x_1 \in E$, $y_3 \in U$, i.e., $(x_1, y_3) \in \text{supp}(\gamma)$, then this “push” could be harmful if $(x'_2, y'_3) \in \text{supp}(\gamma)$, since then y'_3 could be pushed away from x'_2 . We show that this scenario is forbidden when γ is an optimal transport plan with respect to the cost function $h(x - y) = |x - y|^p$ for $p \geq 1$



The proof of Lemma 12 is given after we complete the proof of Lemma 11.

Lemma 12 is a star-graph analog of the monotonicity property (7) of optimal transport on the line: item (3) is the most straightforward, since it implies that all four points lie on the same linear path. Item (1) (and analogously (2)) is depicted in Fig. 8c with i, j , and k being distinct, $k = l$, i.e., x_i and x'_j do not lie on the same edges, and y'_k is on the path between x_i and y_k .

With the help of Lemma 12, we now make the following claim: if a set with nonzero mass between $J_{i,j}$ and $J_{i',j'}$ was pushed away from $J_{i,j}$, then it cannot increase the overall transport cost.

As in the case of the interval or the circle, proving this claim would show that our inductive construction decreases the transport cost, thus proving the lemma.

To proceed with the proof of the above claim, we note that such a “push” away of a set D from $J_{i,j}$ (towards the vertex 0) could increase the overall transport cost in three scenarios: either (i) the optimal-transport plan γ couples D to a set further away from the vertex on I_i , i.e., in $J_{i,1}, \dots, J_{i,j-1}$, or, (ii) γ couples D to an interval between I_i

and D (iii) γ couples D to a *third* interval, i.e., not I_i and not $I_{i'}$. We will now rule out both all these three scenarios.

First, (i) is impossible since there cannot be a point $x'_i \in J_{i,1} \cup \dots \cup J_{i,j-2}$ which was previously transported into a pushed interval, due to the inductive construction. The only other way for the transport distance to increase with the shift is if y'_k is on the path between x_i and y_k and that $(x_i, y_k), (x'_j, y'_k) \in \text{spt}(\gamma)$ with $y' \neq y$. Lemma 12, item (3) similarly rules out scenario (ii). We are left with scenario (iii), depicted in Fig. 8c. We remark that it is really this scenario that distinguishes the star graph from the interval.

Here, $i \neq j$, and so Lemma 12, items (1)–(2), imply that $|x'| \geq |x|$. Hence, up to relabeling of the edges, we have that $(x_1, y_3), (x'_2, y'_3) \in \text{spt}(\gamma)$ with $|x'| \geq |x|$, and $|y| > |y'|$. By a slight abuse of notation, let x and x' be the infimum of all values such that x_1 and x'_2 satisfy these hypotheses, i.e., the closest to the vertex 0.

To rule this scenario out, let us now define an auxiliary function, \bar{f} , which exchanges the values of f on $(0, x)$ between I_1 and I_2 . Formally, for every $j = 1, \dots, D$ and every $t_j \in I_j$

$$\bar{f}(t_j) = \begin{cases} f(t_2) & \text{if } j = 1 \text{ and } t \in (0, x) \\ f(t_1) & \text{if } j = 2 \text{ and } t \in (0, x) \\ f(t_j) & \text{otherwise} \end{cases}$$

Let us also define $\bar{\gamma}$ to be the coupling between \bar{f}_+ and \bar{f}_- which is identical to γ , with the corresponding changes. We now proceed to make one helpful observation that $\bar{\gamma}$ is an optimal transport plan between \bar{f}_+ and \bar{f}_- .

Indeed, since x_1 is transported to I_3 , then no other point on $(0, x_1)$ is transported to (x_1, L_1) . This would contradict monotonicity on the interval I_1 (see item (3) of Lemma 12). The analogous statement holds for x'_2 as well. Hence, the “exchange” of intervals did not change the distances along which mass is transported, and so $K_p(\gamma) = K_p(\bar{\gamma})$, and so $W_p(f_+, f_-) = K_p(\gamma) \geq W_p(\bar{f}_+, \bar{f}_-)$.

We now want to show an inequality in the other direction. Let γ_* be an optimal transport plan for \bar{f} , then $\bar{\gamma}_*$ is a coupling of f_+ and f_- , and so

$$W_p(f_+, f_-) = K_p(\gamma) = K_p(\bar{\gamma}) \geq W_p(\bar{f}_+, \bar{f}_-) = K_p(\bar{\gamma}_*) = K_p(\bar{\gamma}_*) \geq W_p(f_+, f_-).$$

Hence we have equalities throughout, and in particular $K_p(\bar{\gamma}) = W_p(\bar{f}_+, \bar{f}_-)$, and so $\bar{\gamma}$ is an optimal transport plan from \bar{f}_+ to \bar{f}_- .

By construction $(x'_2, y'_3) \in \text{spt}(\bar{\gamma})$ (as it were with γ), and due to the “exchange” now we have that $(x_2, y_3) \in \text{spt}(\bar{\gamma})$. This violates item (3) in Lemma 12.

Thus we arrive at a contradiction. Hence $x' = x$. But now consider the situation where $(x_1, y_3), (x_2, y'_3) \in \text{spt}(\gamma)$, with $y > y'$. By the exact same argument with the roles of y and x interchanged, we arrive at a contradiction again. We have ruled out scenario (iii), as depicted in Fig. 8c.

To summarize, the proposed shift cannot increase the transport cost, and therefore Lemma 11 is proven. \square

Proof of Lemma 12 Consider first the case of $p > 1$, where $c(x, y) = h(x - y) = |x - y|^p$ is strictly convex. First, recall that the support of γ is c -cyclically monotone (Santambrogio 2015, Theorem 1.38), i.e.,

$$h(x_i - y_k) + h(x'_k - y'_l) \leq h(x_i - y'_l) + h(x'_j - y_k) . \quad (22)$$

We will show in detail how the first of the three scenarios is impossible, as the others follow similarly.

Assume without loss of generality that $x_1 \in I_1$, $y'_3, y_3 \in I_3$, and $x'_2 \in I_2$, as in Fig. 8c. Denote $a = |y| - |y'| > 0$. Then (22) reads as

$$h(|x| + |y'| + a) + h(|x'| + |y'|) \leq h(|x| + |y'|) + h(|x'| + |y'| + a) . \quad (23)$$

Suppose $|x'| \leq |x|$ first. Then

$$\begin{aligned} |x'| + |y'| &< |x'| + |y'| + a \leq |x| + |y'| + a , \\ |x'| + |y'| &\leq |x| + |y'| < |x| + |y'| + a . \end{aligned}$$

We can express the “sandwiched” numbers, $|x| + |y'|$ and $|x'| + |y'| + a$, as linear interpolation between the endpoints $\alpha = |x'| + |y'|$ and $\beta = |x| + |y'| + a$,

$$|x'| + |y'| + a = t\alpha + (1 - t)\beta , \quad |x| + |y'| = t\beta + (1 - t)\alpha ,$$

where

$$t \equiv \frac{a}{a + |x| - |x'|} \in (0, 1) .$$

However, h is strictly convex for $p > 1$, and so combined with (23), then

$$\begin{aligned} h(\alpha) + h(\beta) &\leq h(t\alpha + (1 - t)\beta) + h(t\beta + (1 - t)\alpha) \\ &< th(\alpha) + (1 - t)h(\beta) + th(\beta) + (1 - t)h(\alpha) \\ &= h(\alpha) + h(\beta) , \end{aligned}$$

which is a contradiction.

The case of $p = 1$ is proven by approximating $h(x - y) = |x - y|$ by a sequence of strictly convex functions, see Santambrogio (2015). Crucially, even though $x, y \in S_D$ and not on the real line, $h(x - y)$ is really a shorthand for h operating on the geodesic distance (on the graph) between x and y , and hence the procedure generalizes to the graph.

We have proven item (1) in the lemma. Item (2) is analogous, but with interchanging x and x' with y and y' , respectively. Item (3) reduces to the standard analysis of the interval, since all four points lie on the same line. \square

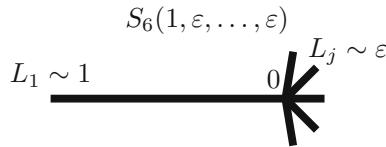


Fig. 9 A star with 6 edges, where $L_2, \dots, L_6 \sim \varepsilon$ with $\varepsilon \ll 1$. However, since most of the edges are very short, this star should be almost equivalent to an interval, in terms of transport and mass allocation. In particular, for a function with a single nodal point ($N = 1$), that point cannot be the vertex 0

Remark While it is preferable to have $0 \in Z(f)$, it might not be possible; A sufficient condition is to have *at least* two sufficiently long edges. Up to a relabeling of the edges, we make the hypothesis that

$$|I_1|, |I_2| > c_1/2c_\infty$$

i.e., all of the positive and negative mass could be allocated on I_1 and I_2 , respectively, in intervals adjacent to 0. This is a sufficient condition that avoids somewhat pathological star graphs for which 0 cannot be efficiently utilized as a nodal point. Such graphs might resemble an interval attached to many short edges at its end, see e.g., Fig. 9.

Since the minimizers of $W_p(f_+, f_-)$ on S_D shift mass only between adjacent sub-intervals (Lemma 11), we can now integrate our analysis of the single-point case (Lemma 9) and conclude that it is always preferable to have a nodal point on the vertex. Hence, minimizers have the following form:

- (1) The vertex is a nodal point, i.e., $0 \in Z(f)$, and there exists $r_1, \dots, r_D \geq 0$ such that f on $[0, r_j] \subseteq I_j$ is non-negative for $1 \leq j \leq M$, and non-positive for $M + 1 \leq j \leq D$. The overall mass concentrated around 0 is therefore

$$\tilde{c}_1 \equiv c_\infty \sum_{j=1}^D r_j.$$

- (2) The other $N - 1$ nodal points are “internal”, i.e., for each $z_\ell \in Z(f)$ there exists $1 \leq j \leq D$ such that $z_\ell \in]0, L_j[$ and f is an anti-symmetric step function around it with length $d_\ell > 0$, as in the interval case of Sect. 3.

Therefore, one needs to optimize only the positive step-widths r_1, \dots, r_D and d_1, \dots, d_{N-1} . We summarize these results in the following theorem:

Theorem 13 Suppose S_D has at least two edges longer than $c_1/2c_\infty$. Question 1, i.e., minimizing $W_p^p(f_+, f_-)$ over $X(c_\infty, c_1, N, S_D)$, is equivalent to finding $r_1, \dots, r_D, d_1, \dots, d_{N-1} > 0$ and a configuration function $q : \{1, \dots, N - 1\} \rightarrow \{1, \dots, D\}$ which minimize

$$\sum_{\ell=1}^{N-1} c_\infty d_\ell^{p+1} + W_p^p(g_+, g_-), \quad (24a)$$

where

$$g(x) \equiv c_\infty \sum_{j=1}^M \mathbb{1}_{[0, r_j]}(x) - c_\infty \sum_{j=M+1}^D \mathbb{1}_{[-r_j, 0]}(x), \quad (24b)$$

subject to the mass conservation conditions

$$\begin{cases} c_\infty \sum_{j=1}^M r_j = c_\infty \sum_{j=M+1}^D r_j \equiv \frac{1}{2} \tilde{c}_1, \\ \tilde{c}_1 + 2c_\infty \sum_{\ell=1}^{N-1} d_\ell = c_1, \end{cases} \quad (24c)$$

and the admissibility constraints

$$r_j + 2 \sum_{q(\ell)=j} d_\ell \leq L_j, \quad 1 \leq j \leq D. \quad (24d)$$

Note that, since g is given explicitly by the equivalence established in Lemma 9, see (19), and since it is a function on the line, $W_p(g_+, g_-)$ is given by the respective inverse CDFs as in (8). Therefore optimizing $W_p^p(g_+, g_-)$ involves a direct, closed form calculation (see below). Moreover, if the edges are long, i.e., $L_j > c_1/c_\infty$ for any $1 \leq j \leq D$, then the first constraint can be satisfied for any assignment of r_j 's and nodal point, i.e., for any choice of q . Hence, (24) becomes computable in closed form *without* having to enumerate over all $(N-1)^D$ configuration functions q .

We can also see what is the complication introduced by short edges - if e.g., $L_D \rightarrow 0$, then effectively the number of effective edges becomes $D-1$ (in the sense that only vanishingly small mass can be assigned to I_D), thus changing the solution. Hence, obtaining a closed-form expression for (24) can be algebraically daunting. In fact, already for $N=1$, where $Z(f) = \{0\}$, there are some complications. We will work out a few cases.

5.2 Long Edges, $N = 1$

We find the minimizers of $W_p(f_+, f_-)$ on $f \in X(c_\infty, \tilde{c}_1, N=1, S_D)$ where all edges are sufficiently long.¹⁰ As discussed above, it is always preferable to have the nodal point at 0. Moreover, it would always be better to distribute the L^1 mass on all edges. This is because the L^1 and L^∞ constraints imply that “wasted” edges (on which $f_j = 0$ everywhere) lead to transport over larger distances, and so to a more expensive optimal transport cost. Hence, $W_p(f_+, f_-) = W_p(g_+, g_-)$, the equivalent function on the line given by (24a). Since the edges are sufficiently long, if f has $1 \leq M < D$ positive edges, then g could be a general function with $\|g\|_\infty = \max(D, D-M)c_\infty$. Hence, by our analysis for the interval (Sect. 3), the optimal g is of the form:

¹⁰ The notation of the L^1 norm by \tilde{c}_1 here will be convenient when we consider $N \geq 1$ in the next section.

$$g(x) = \begin{cases} -(D - M)c_\infty, & x \in (-r_-, 0), \\ Dc_\infty, & x \in (0, r_+), \\ 0, & \text{otherwise.} \end{cases}$$

Crucially, an $f : S_D \rightarrow \mathbb{R}$ to which this g is equivalent is only possible when all of S_D edges (the different I_j 's) are sufficiently long. In that case, the corresponding f will be of the form

$$f_j(x) = \begin{cases} c_\infty, & x \in (0, r_+), 1 \leq j \leq M, \\ -c_\infty, & x \in (0, r_-), M + 1 \leq j \leq D, \\ 0, & \text{otherwise,} \end{cases} \quad (25)$$

where r_+ and r_- are determined by the mass conservation relation (24c), which reduces to

$$Mr_+ = (D - M)r_- . \quad (26)$$

In these settings, minimizing $W_p(f_+, f_-) = W_p(g_+, g_-)$ reduces to finding the optimal parameters $r_+, r_- > 0$ and $1 \leq M \leq D$.

Proposition 14 [N = 1 on a star] Suppose S_D has at least two edges longer than $c_1/2c_\infty$. Then,

$$\arg \min W_p^p(f_+, f_-) \quad \text{over} \quad X_s(c_\infty, \tilde{c}_1, 1, S_D) ,$$

is given by (25) with $M = \lfloor D/2 \rfloor$, the lower integer part of $D/2$.

Proof Since the equivalent g (see Lemma 9) is a function on an interval, we can use the explicit formula in terms of the inverse CDFs, (8). By direct calculation, the inverse CDFs, defined on the interval $[0, \tilde{c}_1/2]$, are $G_+^{-1}(t) = r_+ 2\tilde{c}_1^{-1}(t)$ and $G_-^{-1}(t) = -r_- + r_- 2\tilde{c}_1^{-1}(t)$. Assume without loss of generality that $r_+ \geq r_-$, and so $\alpha \equiv 2\tilde{c}_1^{-1}(r_+ - r_-) \geq 0$. Hence,

$$\begin{aligned} W_p^p(g_+, g_-) &= \int_0^{\tilde{c}_1/2} \left[r_+ 2\tilde{c}_1^{-1}t - (-r_- + r_- 2\tilde{c}_1^{-1}t) \right]^p dt \\ &= \int_0^{\tilde{c}_1/2} [\alpha t + r_-]^p dt \geq \int_0^{\tilde{c}_1/2} [r_-]^p . \end{aligned} \quad (27)$$

Hence, $W_p^p(g_+, g_-)$ is minimized when $\alpha = 0$, i.e., when $r_+ = r_-$, which by the mass conservation relation (26) yields $M = D/2$.

When D is even, substitution of $M = D/2$ into (27) yields

$$D \text{ even} \quad W_p^p(g_+, g_-) = \frac{\tilde{c}_1}{2} r_-^p = \frac{1}{2D^p} \left(\frac{\tilde{c}_1}{c_\infty} \right)^p \tilde{c}_1, \quad r_- = r_+ = \frac{\tilde{c}_1}{c_\infty} \frac{1}{D}, \quad (28)$$

which is the same as in the single step function case for $D = 2$, see (13).

When D is odd, $M = D/2$ is not an integer, and therefore not an admissible choice. We claim that the optimal choices are either $M = \lfloor D/2 \rfloor = (D-1)/2$ or $M = \lceil D/2 \rceil$. To see that, first note that the integrand in (27) is monotonic in α . Moreover, by the mass conservation relation (26), we get that

$$\alpha = \alpha(M) = \frac{1}{c_\infty} \left(\frac{1}{M} - \frac{1}{D-M} \right), \quad (29)$$

for any $1 \leq M \leq D$. Hence, $W_p^p(g_+, g_-)$ is monotonic convex in M near its minimum at $D/2$, and the closest integer values, $\lfloor D/2 \rfloor$ and $\lceil D/2 \rceil$, minimize the cost.

To find the minimal cost, first, by direct computation of the integral in (27), we get

$$W_p^p(g_+, g_-) = \frac{1}{(p+1)\alpha} \left[\left(\frac{\alpha \tilde{c}_1}{2} + r_- \right)^{p+1} - r_-^{p+1} \right]. \quad (30)$$

By substitution of $M = \lfloor D/2 \rfloor = (D-1)/2$ into (29), we get that

$$\alpha \left(\lfloor \frac{D}{2} \rfloor \right) = \frac{1}{c_\infty} \left(\frac{2}{D-1} - \frac{2}{D+1} \right) = \frac{4}{c_\infty(D-1)(D+1)}.$$

Similarly, by the same constraint in (27), we have that

$$r_- c_\infty \frac{D+1}{2} = \frac{\tilde{c}_1}{2} \implies r_- = \frac{\tilde{c}_1}{c_\infty(D+1)}, \quad \text{and} \quad r_+ = \frac{\tilde{c}_1}{c_\infty(D-1)}.$$

Substituting the expressions for α and r_- into (30), we get

$$\begin{aligned} W_p^p(g_+, g_-) &= \frac{c_\infty(D-1)(D+1)}{4(p+1)} \left[\left(\frac{4\tilde{c}_1}{2c_\infty(D-1)(D+1)} + \frac{\tilde{c}_1}{c_\infty(D+1)} \right)^{p+1} \right. \\ &\quad \left. - \left(\frac{\tilde{c}_1}{c_\infty(D+1)} \right)^{p+1} \right] \\ &= \frac{1}{4(p+1)(D+1)^p(D-1)^p} \left(\frac{\tilde{c}_1}{c_\infty} \right)^p \tilde{c}_1 [(D+1)^{p+1} - (D-1)^{p+1}]. \end{aligned}$$

To compare this expression to the even case (28), let us set $p = 1$. Then

$$W_1(g_+, g_-) = \frac{D}{2(D+1)(D-1)} \frac{\tilde{c}_1}{c_\infty} \tilde{c}_1 = \frac{\tilde{c}_1^2}{4\tilde{D}c_\infty}, \quad (31)$$

with $\tilde{D} = \frac{(D+1)(D-1)}{2D}$ as defined before. Since

$$\frac{1}{2(D+1)} \leq \frac{1}{4\tilde{D}} \leq \frac{1}{2(D-1)},$$

we easily see that W_1 lies in between the optimal cost for stars with $D - 1$ and $D + 1$. This is consistent with the intuition that as D increases, the cost decreases.

Either D is even or odd, the overall transport cost is cheaper than the case of an internal point. This makes intuitive sense, because a nodal point on the vertex allows for more mass to be transported over a shorter distance, due to Lemma 9. \square

5.3 Long Edges, $N \geq 1$, D is Even

In principle, now one can go to the general case of $N \geq 1$ and, using the minimization formulation (24), find the minimizers. Comparing the even and odd cases in (28) and (31), respectively, it seems that the case of D even is more tractable.

The total cost of transport through the 0-vertex depends on a single parameter $r_- = r_+$, which is not yet determined. Equivalently, it depends on \tilde{c}_1 , the amount of mass concentrated around the vertex, with $0 \leq \tilde{c}_1 \leq c_1$. Since we assume that the edges are sufficiently long, then (based on our analysis for the interval in Sect. 3) the minimization problem (24) simplifies by $d_1 = \dots = d_{N-1} = d_{\text{int}}$. We therefore wish to minimize

$$W_p^p(f_+, f_-) = (N - 1)c_\infty d_{\text{int}}^{p+1} + \frac{D}{2}c_\infty r_+^{p+1}, \quad (32)$$

subject to

$$2(N - 1)d_{\text{int}}c_\infty + Dr_+c_\infty = c_1.$$

As in the case of the interval (Sect. 3), we use the method of Lagrange multipliers

$$\mathcal{L} = (N - 1)c_\infty d_{\text{int}}^{p+1} + \frac{D}{2}c_\infty r_+^{p+1} - \lambda(2(N - 1)d_{\text{int}}c_\infty + Dr_+c_\infty - c_1).$$

Then

$$0 = \partial_{d_{\text{int}}}\mathcal{L} = (p + 1)(N - 1)c_\infty d_{\text{int}}^p - 2\lambda(N - 1)c_\infty \implies d_{\text{int}}^p = \frac{2\lambda}{p + 1},$$

and

$$0 = \partial_{r_+}\mathcal{L} = (p + 1)\frac{D}{2}c_\infty r_+^p - \lambda Dc_\infty \implies r_+^p = \frac{2\lambda}{p + 1}.$$

In particular, $r_+ = d_{\text{int}}$. The constraint associated with c_1 then yields

$$d_{\text{int}} = r_+ = \frac{1}{2} \frac{1}{(N - 1) + \frac{D}{2}} \frac{c_1}{c_\infty}$$

and the minimal optimal transport cost is

$$W_p(f_+, f_-) = \frac{c_1}{c_\infty} \frac{c_1^{\frac{1}{p}}}{2^{1+\frac{1}{p}}} \frac{1}{\tilde{N}}, \quad \tilde{N} \equiv N - 1 + \frac{D}{2}.$$

In comparison with Theorem 5 for the interval, we see that for a star with an *even number of edges* that are sufficiently long, N is replaced by \tilde{N} , i.e., the vertex as a nodal point is equivalent to $D/2$ nodal points on a line.

5.4 Long Edges, $N \geq 1, D$ is Odd

In this case, the same argument yields a much more cumbersome expression, and so we resolve it only for $p = 1$. Lagrange multipliers method now yields

$$\begin{aligned} \mathcal{L} &= (N-1)c_\infty d_{\text{int}}^2 + \frac{1}{4\tilde{D}c_\infty} \tilde{c}_1^2 - \lambda (2(N-1)d_{\text{int}}c_\infty + \tilde{c}_1 - c_1), \\ \text{for } \tilde{D} &\equiv \frac{(D+1)(D-1)}{2D}. \end{aligned}$$

Here, since $r_+ \neq r_-$, it is easier to minimize the cost with respect to \tilde{c}_1 , the total L^1 mass around the vertex. Then

$$0 = \partial_{\tilde{c}_1} \mathcal{L} = \frac{1}{2\tilde{D}c_\infty} \tilde{c}_1 - \lambda,$$

and so

$$0 = \partial_{d_{\text{int}}} \mathcal{L} = 2(N-1)c_\infty d_{\text{int}} - 2\lambda(N-1)c_\infty \implies d_{\text{int}} = \lambda = \frac{1}{2\tilde{D}c_\infty} \tilde{c}_1.$$

The L^1 constraint now reads as

$$\frac{N-1}{\tilde{D}} \frac{\tilde{c}_1}{c_\infty} c_\infty + \tilde{c}_1 = c_1 \implies \tilde{c}_1 = \frac{\tilde{D}}{N-1+\tilde{D}} c_1.$$

As a sanity check, we see that $\tilde{c}_1 \rightarrow c_1$ as $D \rightarrow \infty$. We note that, in this case, the sharp lower bound for $f \in X(c_\infty, c_1, N, S_D)$ reads like

$$\begin{aligned} W_1(f_+, f_-) &\geq (N-1)c_\infty d_{\text{int}}^2 + \frac{1}{4\tilde{D}c_\infty} \tilde{c}_1^2 \\ &\geq \frac{c_1^2}{4(N-1+\tilde{D})c_\infty} \\ &\geq \frac{1}{4\tilde{N}} \frac{c_1^2}{c_\infty}, \quad \text{for } \tilde{N} \equiv N - 1 + \tilde{D}. \end{aligned} \tag{33}$$

Compared with the case where the number of edges is even (see Sec. 5.3), \tilde{D} may be viewed as the effective multiplicity of the nodal point at the vertex.

The key difference between the lower bounds on W_1 for stars and the lower bounds obtained for non-branching spaces (the interval, the circle, and general d -dimensional surfaces in (2), see Carroll et al. (2020), Cavalletti and Farinelli (2021) and Sagiv and Steinerberger (2020)), is that here the lower bound *does not admit a multiplicative ‘uncertainty principle’-like structure* of $4W_1 \cdot N \geq \frac{c_1}{c_\infty} c_1$, where A depends on the domain and norms of f . Nevertheless, for the cases in Sects. 5.3 and 5.4, we may still say that the structure is preserved with an *effective* N that is replaced by \tilde{N} to account for the particular geometry of the nodal point at the vertex of the star. However, discussions below imply that the precise characterization of \tilde{N} could be much more involved than only the degree D with short edges present.

5.5 $D = 3$ with One Short Edge, $N \geq 1$ and $p = 1$

In the long-edge cases discussed above, the minimization problem given in Theorem 13 is assumed to have a solution in the interior of the constraint set (24d). To see further how the geometry affects the conclusion on the minimal optimal transport cost, we consider the special case of $D = 3$, with two long edges and a third short one as an illustrative example. Our goal is to see how the limit $L_3 \rightarrow 0^+$ interpolates between S_3 with three long edges, and $S_2 \cong [-L_1, L_2]$.

First, note that to optimize the transport cost in the present case, it remains advantageous to include the node of the star in $Z(f)$. Moreover, the maximal amount of mass should be assigned to the short edge if its length is sufficiently small. Meanwhile, away from the node, we retain the symmetric structure around the other points in $Z(f)$ that are located on the long edges. Near those points, we utilize earlier discussion and use d_{int} to represent the length of the subintervals on which the functions are taken to be $\pm c_\infty$ with the total optimal transport cost given by $(N - 1)c_\infty d_{\text{int}}^2$.

As before, to get an explicit expression for (24a), we need to work out $W_1(g_+, g_-)$ as a function of the overall mass assigned to the vertex, \tilde{c}_1 . Assume without loss of generality that $f \geq 0$ on I_1 and $f \leq 0$ on I_2 and I_3 near the node. For $L_3 \ll 1$, then the maximal amount of mass should be assigned to I_3 , i.e., $r_3 = L_3$ and $f(x) = -c_\infty$ on I_3 . By the consideration above, we have that the equivalent $g : \mathbb{R} \rightarrow \mathbb{R}$ is of the form

$$g(x) = \begin{cases} c_\infty & x \in (-r_1, 0), \\ -2c_\infty & x \in (0, L_3), \\ -c_\infty & x \in (L_3, r_2), \\ 0 & \text{otherwise.} \end{cases}$$

for two undetermined constants $r_1, r_2 > 0$. For $p = 1$, we can use the CDF counterpart of (8), $W_1(g_+, g_-) = \|G_+ - G_-\|_{L^1(\mathbb{R})}$, where G_\pm are the CDFs of g_\pm . By direct computation

$$G_+(t) - G_-(t) = \begin{cases} c_\infty(t + r_1) & t \in (-r_1, 0), \\ c_\infty r_1 - 2c_\infty t & t \in (0, L_3), \\ c_\infty r_1 - 2c_\infty L_3 - c_\infty t & t \in (L_3, r_2). \end{cases}$$

Since the overall mass conservation (24c) reduces to $r_1 = r_2 + L_3 = \tilde{c}_1/2c_\infty$, we have that

$$\begin{aligned} W_1(g_+, g_-) &= \|G_+ - G_-\|_{L^1(\mathbb{R})} \\ &= \frac{1}{4}\tilde{c}_1r_1 + \frac{1}{4}\tilde{c}_1r_2 + \frac{1}{4}\tilde{c}_1L_3 - r_2c_\infty L_3 \\ &= \frac{\tilde{c}_1^2}{4c_\infty} - r_2c_\infty L_3 \\ &= \frac{\tilde{c}_1^2}{4c_\infty} - \left(\frac{\tilde{c}_1}{2c_\infty} - L_3\right)c_\infty L_3. \end{aligned}$$

Putting these considerations together, the minimization problem in (24) reduces to

$$\begin{aligned} W_1(f_+, f_-) &= (N-1)c_\infty d_{\text{int}}^2 + \frac{\tilde{c}_1^2}{4c_\infty} - \left(\frac{\tilde{c}_1}{2c_\infty} - L_3\right)c_\infty L_3 \\ &= c_\infty \left[(N-1)d_{\text{int}}^2 + \left(\frac{\tilde{c}_1}{2c_\infty} - \frac{L_3}{2}\right)^2 + \frac{3}{4}L_3^2 \right], \end{aligned}$$

subject to the L^1 mass conservation constraint

$$2(N-1)d_{\text{int}} + \frac{\tilde{c}_1}{c_\infty} = \frac{c_1}{c_\infty}.$$

Thus, for $\beta \equiv L_3c_\infty/c_1$, we get the optimal transport cost

$$W_1(f_+, f_-) \geq c_\infty \left[\frac{1}{4N} \left(\frac{c_1}{c_\infty} - L_3 \right)^2 + \frac{3}{4}L_3^2 \right] = \frac{c_1^2}{4c_\infty} \frac{(1-\beta)^2 + 3N\beta^2}{N} \quad (34)$$

with equality for

$$d_{\text{int}} = \frac{1}{2N} \left(\frac{c_1}{c_\infty} - L_3 \right), \quad \text{and} \quad \frac{\tilde{c}_1}{c_\infty} = \frac{c_1}{Nc_\infty} + \frac{(N-1)L_3}{N}.$$

β as an Interpolation Parameter Between S_3 and an Interval

We can understand these expressions between the two limits of either shortening or lengthening I_3 , i.e., as the star graph deforms into an interval with $L_3 \rightarrow 0^+$ on one hand and as

third edge becomes long enough for us to recover the case of three long edges, as described in Sects. 5.2 and 5.4. Note that our analysis involving the short edge holds precisely until the case where there is no distinction between I_2 and I_3 in terms of mass allocation. By the results in Sect. 5.4, the optimal mass around the origin for $D = 3$ is $\tilde{c}_1 = 4c_1/(3N + 1)$. Since near the vertex, L_3 and L_2 both contain the

support of f_- , the threshold length for L_3 to enable the optimal \tilde{c}_1 thus corresponds to $\beta^* = c_\infty L_3/c_1 = \tilde{c}_1/4c_1 = 1/(3N + 1)$.

Indeed, we may rewrite the right-hand side of (34) to get an equivalent form given by

$$W_1(f_+, f_-) \geq \frac{c_1^2}{4c_\infty} \frac{3N + 1}{N} \left(\beta - \frac{1}{3N + 1} \right)^2 + \frac{c_1^2}{4c_\infty \left(N + \frac{1}{3} \right)}.$$

The lower bound is clearly a decreasing function of β for $\beta \in [0, 1/(3N + 1)]$, hence *interpolating*, as expected, between the lower bound derived in (16) for $p = 1$ and $\beta = 0$ (i.e., $D = 2$ for the case of an interval) and that given in (33) with $D = 3$ and $\tilde{D} = \frac{4}{3}$.

Meanwhile, for small but nonzero L_3 , the non-multiplicative nature of the lower bound on the right hand side of (34) is evident. Compared to the cases in Sects. 5.3 and 5.4, we see that the multiplicative form is lost not only in terms of N but also with respect to $\frac{c_1}{c_\infty}$.

Remark What can be said on stars such as in Fig. 9, for which Theorem 13 does not apply? We outline the strategy for their analysis: first, the equivalence relation in Lemma 9 can be extended to a single nodal point *near* the vertex, on the long edge. The adjacency and monotonicity arguments (Lemmas 11 and 12) seem to go through as well. Then, the optimization problem in Theorem 13 should be extended by adding another parameter to represent the location of that “special” nodal point, which is near the vertex. Finally, we remark that the condition for a long edge being $\geq c_1/2c_\infty$ is sufficient, but not necessarily sharp. The sharpness of this lower bound would become one interesting element in the study of Question 1 on general metric graphs.

6 Outlook

An important message of this work is that the original problem of minimizing the transport cost over the specified function class is equivalent to a generalized minimal surface problem associated with *optimal domain partition*. For one-dimensional domains, minimizing the cost reduces to optimally positioning the nodal points and locating masses around them. We believe that this approach can be generalized to more challenging cases, some of which are discussed below.

Multidimensional Domains

In an analogous way to the one-dimensional settings of this study, it seems that on bounded and regular domain $\Omega \subseteq \mathbb{R}^d$, the minimizers will be step functions concentrated near the interfaces between f_+ and f_- . If this is indeed the case, the key remains to be the solution to the minimization problem

$$\min_{f \in X_s(1, c_1/c_\infty, N, \Omega)} W_p(f_+, f_-). \quad (35)$$

It should be commented that, for step functions, the existence of a lower bound ((2) Carroll et al. (2020), Cavalletti and Farinelli (2021) and Sagiv and Steinerberger (2020)) combined with nucleation arguments such as Novack et al. (2023), Lemma 2.8, should yield the existence of minimizers. An element in $X_s(1, c_1/c_\infty, N, \Omega)$ corresponds to a unique partition of Ω into three disjoint subsets $\Omega_+, \Omega_-, \Omega_0 \subset \Omega$, where $\Omega_0 = \Omega \setminus (\Omega_+ \cup \Omega_-)$, such that

$$|\Omega_-| = |\Omega_+| = \frac{c_1}{2}, \quad \mathcal{H}^{d-1}(\Gamma) = N, \quad \Gamma \equiv \partial\Omega_+ \cap \partial\Omega_-,$$

assuming that the partitions are sufficiently regular such that $\mathcal{H}^{d-1}(\Gamma)$ is well defined. Denoting the set of such partitions by $Y(c_1, N, \Omega)$, one may attempt a similar approach to the one presented in this work: reducing the functional minimization problem (35) to a purely geometric one:

$$\min_{f \in X(1, c_1, N, \Omega)} W_p(f_+, f_-) = \min_{(\Omega_-, \Omega_+, \Omega_0) \in Y(c_1, N, \Omega)} W_p(\mathbb{1}_{\Omega_+}, \mathbb{1}_{\Omega_-}) \quad (36)$$

Clearly, the latter becomes an optimal partition problem or a minimal surface problem. That is, the problem under consideration may be formulated as an equivalent geometric question.

Question 15 *For a bounded and regular domain $\Omega \subset \mathbb{R}^d$, $p \geq 1$, $c_1 > 0$, and $N > 0$, what are the optimal partitions $(\Omega_-, \Omega_+, \Omega_0)$ of Ω and the corresponding minimum value associated with the problem*

$$\min_{(\Omega_-, \Omega_+, \Omega_0) \in Y(c_1, N, \Omega)} W_p(\mathbb{1}_{\Omega_+}, \mathbb{1}_{\Omega_-}) \quad (37)$$

This draws interesting connections to possibly other optimal domain partition problems such as those related to different optimal transport minimization problems, e.g., partitions defined by the optimal quantization (centroidal Voronoi tessellations) (Qiang et al. 1999; Mérigot et al. 2021). One may find more discussions on other optimal domain partition problems in Bucur et al. (1998). A variation of Question 15 considered in Buttazzo et al. (2020), Novack et al. (2023), Peletier and Röger (2009) and Xia and Zhou (2021) is to modify the set of feasible partitions by removing the constraint on the prescribed measure of the nodal, and consider the minimization of $|\partial\Omega_+| + \lambda W_1(\mathbb{1}_{\Omega_+}, \mathbb{1}_{\Omega_-})$ where $|\partial\Omega_+|$ is the Hausdorff measure of the boundary set of Ω_+ and $\lambda > 0$ is a penalty constant.

The solution to the purely geometric Question 15 can be used to offer a new perspectives on the existing studies concerning (2) in multiple-dimensions (Steinerberger 2020; Sagiv and Steinerberger 2020; Carroll et al. 2020; Cavalletti and Farinelli 2021) and to make further connections with the bilayer membrane limits (Peletier and Röger 2009; Lussardi 2014). In particular, the optimal transport problem admits a decomposition into a continuous family of one-dimensional optimal-transport problems. One of the main challenges is that, as we seek to solve (37), this decomposition changes as well. Based on the discussion on the one-dimensional examples, it is obvious that the optimal partitions may not be unique in general, even though they yield the same

optimal transport cost. While a complete solution is beyond the scope of this work, one might conjecture that Γ is made of piecewise planar surfaces, which are the conventional minimal surfaces in the Euclidean space.

Metric Graphs and Laplacian Eigenfunctions

This work is a starting point for the study of the Wasserstein minimization problems on metric graphs, for which the star S_D (Sect. 5) is perhaps the most basic nontrivial example. Our analysis for S_D suggests what might be the key challenges for general metric graphs - first, that the associated minimization problems might be algebraically complicated, and second, that the equivalence principles (Lemma 9) are not easily extended to general graphs. The second key challenge is to find a monotonicity-like property which generalizes Lemma 12 to non-star graphs.

While the study of optimal transport on metric graphs is relatively new Erbar et al. (2021), the study of nodal sets of special functions on metric graphs is a well-established field, see e.g., Alon et al. (2018), Band (2014), Band et al. (2012), Berkolaiko and Weyand (2014), Colin de Verdière (2013), Gutzmann et al. (2003) and Hofmann (2021) as well as (Berkolaiko and Kuchment 2013) and the references therein. In its heart, the question is a natural extension of Sturm's Oscillations Theory: given the N -th Laplacian eigenfunction on a certain metric graph, how many zeroes does it have?

Independently, uncertainty principles for the Wasserstein distance such as (16) were applied to generalize Sturm and Sturm-Hurwitz (sums of Laplacian eigenfunctions) type results in dimensions $d \geq 1$ (Carroll et al. 2020; Cavalletti and Farinelli 2021; Sagiv and Steinerberger 2020; Steinerberger 2020). Generally speaking, the approach proceeds in two steps: first, prove a lower bound of the type (16), and then, obtain an upper bound on $W(f_+, f_-)$ for the specialized type of functions under consideration, e.g., Laplacian eigenfunction. The current work might therefore open the way of connecting these two lines of works, though many open challenges remain before such a connection is established.

Minimization over More Specialized Function Classes

Concerning the possible connection to properties of Laplacian eigenfunctions, related studies on lower bounds and uncertainty-principles of Laplacian eigenfunctions on Riemannian manifolds (and RCD spaces in general) can be found in De Ponti and Farinelli (2022), Mukherjee (2021) and Steinerberger (2021). One may note that the key to get the new lower bounds like (16) presented in this work is through the study of the minimization problem over the function class defined by, e.g., (3). The latter class is large enough to allow minimizers taking on the form of step functions. On the other hand, eigenfunctions of elliptic operators are smooth. Thus, another natural research direction to explore is the study of possibly different bounds, associated with similar minimization problems, but over classes of functions that are more regular or of special forms. In particular, the special forms may correspond to discrete function spaces, and the problem of passing from discretizations of L^∞ functions, where the

minimization problem is finite-dimensional, to the continuum limit (studied in this paper), might be of independent interest.

Functional Analytic Perspective

Inequalities such as (2) and (16) can be thought of as interpolation inequalities. On the lower-bound side, there are the L^1 and L^∞ norms involving no derivatives. The upper bound consists of two terms, which can be connected to derivative norms of different orders, respectively: first, $|Z(f)|$ is smaller than $\|D\mathbb{1}_{\text{supp}(f_+)}\|_1$, where the derivative D is taken in the sense of functions of bounded variations (Evans and Garzepy 2018; Giusti and Williams 1984); then, $W_p(f_+, f_-)$ can be viewed as norms of derivatives of negative order, e.g., the W_2 distance is related to the standard Sobolev \dot{H}^{-1} norm; See Loeper (2006) and Peyre (2018) for details.

While this functional-analytic perspective is not used in this paper, it can lead to the exploration of more general variational problems of the type Question 1, e.g., by allowing for constraints not only in L^1 and L^∞ , but in other norms such as L^p norms for $1 < p < \infty$. Moreover, understanding (2) as an interpolation inequality, (16) can be thought of as a sharp interpolation inequality with optimal constants, of which an extensive literature exists (Del Pino and Dolbeault 2002, 2003).

Acknowledgements The authors would like to thank Ram Band for many invigorating discussions and help provided along the way, as well as for suggesting to study this minimization problem on metric graph and highlighting some of the challenges in that area. The authors would also like to thank Raghavendra Venkatraman for bringing to our attention an issue with one of the definitions in a previous version of this manuscript. Many thanks to Fabio Cavalletti and Tim Laux for useful comments and references. Thanks also to Zirui Xu for bringing references Lussardi (2014) and Peletier and Röger (2009) to our attention. The research of Q.D. is supported in part by NSF DMS-1937254, DMS-2012562, DMS 2309245, and DOE DE-SC0022317. A.S. acknowledges the support of the AMS-Simons Travel Grant, and is supported in part by Simons Foundation Math + X Investigator Award #376319 (Michael I. Weinstein).

Data availability Data sharing not applicable to this article as no datasets were generated or analysed during the current study

Declarations

Conflict of interest The authors have no conflict of interest to declare that are relevant to the content of this article.

References

- Alon, L., Band, R., Berkolaiko, G.: Nodal statistics on quantum graphs. *Commun. Math. Phys.* **362**(3), 909–948 (2018)
- Ambrosio, L., Di Marino, S.: Equivalent definitions of bv space and of total variation on metric measure spaces. *J. Funct. Anal.* **266**(7), 4150–4188 (2014)
- Ambrosio, L., Gigli, N., Mondino, A., Rajala, T.: Riemannian RICCI curvature lower bounds in metric measure spaces with σ -finite measure. *Trans. Am. Math. Soc.* **367**(7), 4661–4701 (2015)
- Ambrosio, L., Gigli, N., Savaré, G.: Metric measure spaces with Riemannian Ricci curvature bounded from below. *Duke Math. J.* **163**(7), 1405–1490 (2014)
- Band, R.: The nodal count $\{0, 1, 2, 3, \dots\}$ implies the graph is a tree. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **372**, 20120504 (2014)

Band, R., Berkolaiko, G., Raz, H., Smilansky, U.: The number of nodal domains on quantum graphs as a stability index of graph partitions. *Commun. Math. Phys.* **311**(3), 815–838 (2012)

Barron, E., Bocea, M., Jensen, R.: Duality for the l^∞ optimal transport problem. *Trans. Am. Math. Soc.* **369**(5), 3289–3323 (2017)

Berkolaiko, G., Kuchment, P.: *Introduction to Quantum Graphs*. American Mathematical Society (2013)

Berkolaiko, G., Weyand, T.: Stability of eigenvalues of quantum graphs with respect to magnetic perturbation and the nodal count of the eigenfunctions. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **372**, 20120522 (2014)

Bianchini, S., Cavalletti, F.: The Monge problem for distance cost in geodesic spaces. *Commun. Math. Phys.* **318**(3), 615–673 (2013)

Brenier, Y.: Polar factorization and monotone rearrangement of vector-valued functions. *Commun. Pure Appl. Math.* **44**(4), 375–417 (1991)

Bucur, D., Buttazzo, G., Henrot, A.: Existence results for some optimal partition problems. *Adv. Math. Sci. Appl.* **8**, 571–579 (1998)

Buttazzo, G., Carlier, G., Laborde, M.: On the Wasserstein distance between mutually singular measures. *Adv. Calc. Var.* **13**(2), 141–154 (2020)

Caffarelli, L., Feldman, M., McCann, R.: Constructing optimal maps for Monge's transport problem as a limit of strictly convex costs. *J. Am. Math. Soc.* **15**(1), 1–26 (2002)

Candau-Tilh, J., Goldman, M.: Existence and stability results for an isoperimetric problem with a non-local interaction of Wasserstein type. *ESAIM: control. Optim. Calc. Var.* **28**, 37 (2022)

Carroll, T., Massaneda, X., Ortega-Cerdà, J.: An enhanced uncertainty principle for the Vaserstein distance. *Bull. Lond. Math. Soc.* **52**(6), 1158–1173 (2020)

Cavalletti, F., Farinelli, S.: Indeterminacy estimates and the size of nodal sets in singular spaces. *Adv. Math.* **389**, 107919 (2021)

Champion, T., De Pascale, L., Juutinen, P.: The ∞ -wasserstein distance: Local solutions and existence of optimal transport maps. *SIAM J. Math. Anal.* **40**(1), 1–20 (2008)

de Verdierre, Y.C.: Magnetic interpretation of the nodal defect on graphs. *Anal. PDE* **6**(5), 1235–1242 (2013)

De Ponti, N., Farinelli, S.: Eigenfunctions and a lower bound on the Wasserstein distance. arXiv preprint [arXiv:2104.12097](https://arxiv.org/abs/2104.12097) (2021)

De Ponti, N., Farinelli, S.: Indeterminacy estimates, eigenfunctions and lower bounds on Wasserstein distances. *Calc. Var. Partial. Differ. Equ.* **61**(4), 1–17 (2022)

Del Pino, M., Dolbeault, J.: Best constants for Gagliardo–Nirenberg inequalities and applications to nonlinear diffusions. *J. Math. Pures Appl.ées* **81**(9), 847–875 (2002)

Del Pino, M., Dolbeault, J.: The optimal euclidean lp -sobolev logarithmic inequality. *J. Funct. Anal.* **197**(1), 151–161 (2003)

Delon, J., Salomon, J., Sobolevski, A.: Fast transport optimization for monge costs on the circle. *SIAM J. Appl. Math.* **70**(7), 2239–2258 (2010)

Du, Q., Faber, V., Gunzburger, M.: Centroidal Voronoi tessellations: applications and algorithms. *SIAM Rev.* **41**(4), 637–676 (1999)

Erbar, M., Forkert, D., Maas, J., Mugnolo, D.: Gradient flow formulation of diffusion equations in the Wasserstein space over a metric graph. arXiv preprint [arXiv:2105.05677](https://arxiv.org/abs/2105.05677) (2021)

Evans, L.C., Garzepi, R.F.: *Measure Theory and Fine Properties of Functions*. Routledge (2018)

Feldman, M., McCann, R.: Monge's transport problem on a Riemannian manifold. *Trans. Am. Math. Soc.* **354**(4), 1667–1697 (2002)

Gangbo, W., McCann, R.J.: The geometry of optimal transportation. *Acta Math.* **177**(2), 113–161 (1996)

Giusti, E., Williams, G.H.: *Minimal Surfaces and Functions of Bounded Variation*. Springer (1984)

Gnutzmann, S., Smilansky, U., Weber, J.: Nodal counting on quantum graphs. *Waves Random Med.* **14**(1), S61 (2003)

Hofmann, M., Kennedy, J.B., Mugnolo D., Plümer, M. On pleijel's nodal domain theorem for quantum graphs. In *Annales Henri Poincaré*, pp. 1–30. Springer (2021)

Loeper, G.: Uniqueness of the solution to the Vlasov-Poisson system with bounded density. *J. Math. Pures Appl.* **86**(1), 68–79 (2006)

Lussardi, L., Peletier, M.A., Röger, M.: Variational analysis of a mesoscale model for bilayer membranes. *J. Fixed Point Theory Appl.* **15**(1), 217–240 (2014)

Mazón, J.M., Rossi, J.D., Toledo, J.: Optimal mass transport on metric graphs. *SIAM J. Optim.* **25**(3), 1609–1632 (2015)

McCann, R.J.: Exact solutions to the transportation problem on the line. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **455**, 1341–1380 (1999)

McCann, R.J.: Polar factorization of maps on Riemannian manifolds. *Geometr. Funct. Anal. GAFA* **11**(3), 589–608 (2001)

Mérigot, Q., Santambrogio, F., Sarrazin, C.: Non-asymptotic convergence bounds for Wasserstein approximation using point clouds. *Adv. Neural. Inf. Process. Syst.* **34**, 12810–12821 (2021)

Miranda, M., Jr.: Functions of bounded variation on “good” metric spaces. *J. Math. Pures Appl.* **82**(8), 975–1004 (2003)

Mukherjee, M.: A sharp Wasserstein uncertainty principle for laplace eigenfunctions, arXiv preprint [arXiv:2103.11633](https://arxiv.org/abs/2103.11633) (2021)

Novack, M., Topaloglu, I., Venkatraman, R.: Least Wasserstein distance between disjoint shapes with perimeter regularization. *J. Funct. Anal.* **284**(1), 109732 (2023)

Peletier, M.A., Röger, M.: Partial localization, lipid bilayers, and the elastica functional. *Arch. Ration. Mech. Anal.* **193**(3), 475–537 (2009)

Peyre, R.: Comparison between w_2 distance and \dot{H}^{-1} norm, and localization of Wasserstein distance, *ESAIM: control. Optim. Calc. Var.* **24**(4), 1489–1501 (2018)

Rajala, T., Sturm, K.-T.: Non-branching geodesics and optimal maps in strong $cd(k, \infty)$ -spaces. *Calc. Var. Partial. Differ. Equ.* **50**(3), 831–846 (2014)

Sagiv, A., Steinerberger, S.: Transport and interface: an uncertainty principle for the Wasserstein distance. *SIAM J. Math. Anal.* **52**(3), 3039–3051 (2020)

Salvemini, T.: Sul calcolo degli indici di concordanza tra due caratteri quantitativi. In *Atti della VI Riunione della Soc. Ital. di Statistica* (1943)

Santambrogio, F.: Optimal Transport for Applied Mathematicians. Springer (2015)

Steinerberger, S.: A metric Sturm–Liouville theory in two dimensions. *Calc. Var. Partial. Differ. Equ.* **59**(1), 1–14 (2020)

Steinerberger, S.: Wasserstein distance, fourier series and applications. *Mon. Math.* **194**(2), 305–338 (2021)

Vallender, S.S.: Calculation of the Wasserstein distance between probability distributions on the line. *Theory Prob. Appl.* **18**(4), 784–786 (1974)

Villani, C.: Topics in Optimal Transportation. American Mathematical Society (2003)

Xia, Q., Zhou, B.: The existence of minimizers for an isoperimetric problem with Wasserstein penalty term in unbounded domains. *Adv. Calc. Var.* **16**(1), 1–15 (2021)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.