

Two-mode stability for multi-marginal optimal transport maps

Bang-Xian Han* Zhuo-Nan Zhu †

June 29, 2026

Abstract

We establish a two-mode stability theory for Monge solutions of the multi-marginal optimal transport problem with barycentric quadratic cost. The associated tuple of maps splits into an external barycentric mode and internal relative modes. A quadratic lower bound for the Kantorovich defect controls the internal modes and yields a square-root estimate without invoking any two-marginal stability result. The external mode is the optimal transport map from the fixed source to the Wasserstein barycenter. Combining the resulting two-mode estimate with Mérigot’s sharp stability theorem gives a $\frac{1}{4}$ -Hölder estimate for general perturbations, while barycenter-preserving perturbations satisfy a $\frac{1}{2}$ -Hölder estimate. We prove that both exponents and the dependence on the weights are optimal. We also examine the scope of such a decomposition beyond the barycentric cost: collective-coordinate perturbations and uniformly concave costs of the sum retain the two-mode estimate, whereas graph interactions, hedonic costs and translation-invariant costs reveal two possible obstructions—loss of relative coercivity and lack of stability for the remaining external modes.

Keywords: multi-marginal optimal transport, optimal transport map, quantitative stability, Kantorovich potential, Wasserstein barycenter

MSC 2020: Primary 49Q22; Secondary 49N15, 28A33, 60B05

Contents

1	Introduction	2
2	Mode Decomposition and Stability Estimates	7
2.1	Barycentric and internal coordinates	7
2.2	Separation of barycentric and internal directions	9
2.3	Gluing argument for perturbed marginals	10
2.4	Proof of Theorem 1.2	13
2.5	Proof of Theorem 1.3	14
2.6	Proof of Theorem 1.7	16

*School of Mathematics, Shandong University, Jinan, China. Email: hanbx@sdu.edu.cn

†School of Mathematical Sciences, University of Science and Technology of China, Hefei, China. Email: zhuonanzhu@mail.ustc.edu.cn

3	Optimality and Barycenter-Preserving Examples	17
3.1	Proof of Proposition 1.9	17
3.2	Barycenter-preserving perturbations	21
4	The Scope of the Two-Mode Mechanism	23
4.1	Collective-coordinate perturbations and persistence of the two-mode estimate	23
4.2	Uniformly concave costs of the sum	28
4.3	From cyclic interactions to complete-graph symmetry	28
4.4	Normal modes for pairwise quadratic interaction costs	30
4.5	Hedonic costs and generalized external variables	31
4.6	Failure of relative coercivity: Coulomb and determinant costs	32
	References	32

1 Introduction

Classical roots and the stability gap. Optimal transport, initiated by Monge [30] and relaxed by Kantorovich [21], provides a variational way to compare probability measures. In the Euclidean quadratic case, given two probability measures $\rho, \mu \in \mathcal{P}(\mathbb{R}^n)$, the Monge problem asks for a map $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ with $T_{\#}\rho = \mu$ minimizing

$$\int_{\mathbb{R}^n} \frac{1}{2} |x - T(x)|^2 d\rho(x)$$

among all transport maps from ρ to μ . Such a minimizer is called an optimal transport map. Brenier’s landmark theorem [6] states that, when the source measure is absolutely continuous, this problem has a unique solution and the optimal transport map is the gradient of a convex function.

The multi-marginal analogue was developed by Gangbo and Świech [17]. Their theorem gives existence and uniqueness of Monge solutions for the barycentric quadratic cost considered below, a cost that also provides the multi-marginal formulation of the Wasserstein barycenter problem [1]. These existence theorems leave a quantitative question open: how stable are the maps under perturbations of the marginals? In the two-marginal case, if T_μ and $T_{\tilde{\mu}}$ denote the quadratic optimal transport maps from a fixed source ρ to two targets μ and $\tilde{\mu}$ respectively, one seeks estimates of the form

$$\|T_\mu - T_{\tilde{\mu}}\|_{L^2(\rho)} \leq CW_p^\alpha(\mu, \tilde{\mu}),$$

for suitable constants $C, \alpha > 0$ and $p \geq 1$, usually for compactly supported targets. Existing stability results are typically established for $p = 1$, while stability for other values of p is deduced from the inequality $W_1 \leq W_p$. An early stability result of this form is due to Berman [3], who obtained a dimension-dependent exponent under density and geometric assumptions on the source measure. This line of work was later improved by Delalande and Mérigot [14] with the uniform exponent $\frac{1}{6}$, and further extended by Letrouit and Mérigot [28] using the gluing method to more general source geometries and density profiles. Two further works improved these estimates under bounded-density and compact-support assumptions on the source measure: Cazelles, Pauwels and Portales [10] obtained the exponent $\frac{1}{5}$ in the context of statistical estimation of Monge maps, while Mérigot [29] proved the optimal exponent $\frac{1}{4}$. In dimension $n \geq 2$, this exponent is optimal; see also [27]

for sources of instability. By contrast, quantitative stability for multi-marginal Monge maps is much less developed, although such maps appear in Wasserstein barycenters, density functional theory and economic matching. Letrouit’s lecture notes [26] also mention the extension of quantitative stability estimates to multi-marginal optimal transport as a natural open direction.

Multi-marginal optimal transport. We consider a barycentric multi-marginal analogue of the preceding stability problem. Fix $N \geq 3$ and weights

$$\lambda_i > 0, \quad \sum_{i=1}^N \lambda_i = 1.$$

Let $S, Y \subseteq \mathbb{R}^n$ be compact, let $\rho \in \mathcal{P}(S)$, and let $\mu_2, \dots, \mu_N \in \mathcal{P}(Y)$. In the main part of this paper we study the cost

$$c(x_1, \dots, x_N) := \inf_{y \in \mathbb{R}^n} \sum_{i=1}^N \frac{\lambda_i}{2} |x_i - y|^2. \quad (1.1)$$

This weighted quadratic cost appears in the multi-marginal theory of Gangbo and Świech [17] and in the study of the Wasserstein barycenter problem. We call it the barycentric quadratic cost.

For the barycentric quadratic cost (1.1), the Monge problem is to solve

$$\inf_{T_2, \dots, T_N} \int_S c(x_1, T_2(x_1), \dots, T_N(x_1)) \, d\rho(x_1) \quad (1.2)$$

among maps $T_i : S \rightarrow Y$ satisfying $(T_i)_\# \rho = \mu_i$, $i = 2, \dots, N$. Its Kantorovich relaxation is

$$\inf_{\pi \in \Pi(\rho, \mu_2, \dots, \mu_N)} \int_{S \times Y^{N-1}} c(x_1, \dots, x_N) \, d\pi(x_1, \dots, x_N), \quad (1.3)$$

where $\Pi(\rho, \mu_2, \dots, \mu_N)$ is the set of probability measures with the prescribed marginals. The minimizer in (1.1) is the Euclidean weighted barycenter $\sum_{i=1}^N \lambda_i x_i$, and hence

$$c(x_1, \dots, x_N) = \frac{1}{2} \sum_{i=1}^N \lambda_i |x_i|^2 - \frac{1}{2} \left| \sum_{i=1}^N \lambda_i x_i \right|^2 = \frac{1}{2} \sum_{1 \leq i < j \leq N} \lambda_i \lambda_j |x_i - x_j|^2. \quad (1.4)$$

Gangbo and Świech’s theorem [17] (see also Kim and Pass [24]) extends Brenier’s result to this multi-marginal setting: if ρ is absolutely continuous, then the Monge problem has a unique solution (T_2, \dots, T_N) . We call these maps multi-marginal optimal transport maps.

Multi-marginal optimal transport extends the two-marginal theory and appears in several contexts, including Wasserstein barycenters, statistics and data analysis, multi-agent matching problems in economics, and density functional theory. We refer to Pass [34] for a survey, and to [1, 11, 12, 33, 35] for further discussions.

The two-mode decomposition and Wasserstein barycenters. Let

$$\Omega := \overline{\text{conv}(S \cup Y)} \subseteq \mathbb{R}^n, \quad \mathbb{P} := \lambda_1 \delta_\rho + \sum_{i=2}^N \lambda_i \delta_{\mu_i} \in \mathcal{P}(\mathcal{P}(\Omega)).$$

The Wasserstein barycenter of \mathbb{P} is a minimizer of

$$\mathrm{Var}_{\mathbb{P}}(\nu) := \frac{\lambda_1}{2} W_2^2(\nu, \rho) + \sum_{i=2}^N \frac{\lambda_i}{2} W_2^2(\nu, \mu_i), \quad \nu \in \mathcal{P}_2(\mathbb{R}^n).$$

The following proposition links the Wasserstein barycenter to the barycentric multi-marginal optimal transport problem (see [17], [1, Section 4], and [24, Proposition 5.1]).

Proposition 1.1 (Barycenter map). *Let $\mathbb{P} = \lambda_1 \delta_\rho + \sum_{i=2}^N \lambda_i \delta_{\mu_i} \in \mathcal{P}(\mathcal{P}(\Omega))$ and assume $\rho \ll \mathcal{L}^n$. Then there exists a unique Wasserstein barycenter of \mathbb{P} , denoted by $\mu_{\mathbb{P}} \in \mathcal{P}(\Omega)$. Moreover, if (T_2, \dots, T_N) are the multi-marginal optimal transport maps, then*

$$B := \lambda_1 \mathrm{Id} + \sum_{i=2}^N \lambda_i T_i$$

is the unique optimal transport map from ρ to $\mu_{\mathbb{P}}$, for the quadratic cost. We call B the barycenter map.

This characterization reveals the two-mode structure of the stability problem. The barycenter map B records the averaged displacement and forms the external mode. Each individual map also contains a relative component:

$$V_i := T_i - M, \quad M := \frac{B - \lambda_1 \mathrm{Id}}{1 - \lambda_1}, \quad \sum_{i=2}^N \lambda_i V_i = 0.$$

The variables $(V_i)_{i \geq 2}$ form the internal modes: they describe the relative positions of the target points inside a fiber of the barycenter map. This is the analogue of the center-of-mass and relative-coordinate decomposition in classical mechanics.

The two components have different stability mechanisms. Once the external coordinate is fixed, the Kantorovich defect is quadratically coercive in the internal directions. The external coordinate, by contrast, is itself a two-marginal optimal transport map from ρ to the Wasserstein barycenter. The stability problem therefore separates into a genuinely multi-marginal coercivity estimate on the fibers of the barycenter map and a two-marginal stability estimate for the barycenter map itself.

Main results. We write $\mathcal{P}_{\mathrm{ac}}(S)$ for the set of probability measures on S that are absolutely continuous with respect to the Lebesgue measure \mathcal{L}^n . Given another tuple of target marginals $\tilde{\mu}_2, \dots, \tilde{\mu}_N \in \mathcal{P}(Y)$, let $(\tilde{T}_2, \dots, \tilde{T}_N)$ be the corresponding multi-marginal optimal transport maps and set

$$\tilde{B} := \lambda_1 \mathrm{Id} + \sum_{i=2}^N \lambda_i \tilde{T}_i, \quad \tilde{\mathbb{P}} := \lambda_1 \delta_\rho + \sum_{i=2}^N \lambda_i \delta_{\tilde{\mu}_i}.$$

Our proof follows this separation literally. We first obtain a multi-marginal estimate for the internal modes from a quadratic lower bound for the Kantorovich defect. This step does not use any stability theorem for two-marginal optimal transport maps. Only afterwards do we control the external mode by identifying it with the transport map to the Wasserstein barycenter and applying Mériçot's sharp estimate.

The different stability exponents therefore have different origins. General perturbations inherit the $\frac{1}{4}$ exponent from the external barycenter mode. Barycenter-preserving

perturbations and self-comparisons remove that mode and retain the $\frac{1}{2}$ exponent produced by internal quadratic coercivity. The examples in §3 show that both exponents, as well as the dependence on the weights, are optimal.

Our first result gives a two-mode estimate for the stability of the maps. It does not use any two-marginal stability result.

Theorem 1.2. *Let $N \geq 3$ and let $S, Y \subseteq \mathbb{R}^n$ be compact. Let $\rho \in \mathcal{P}_{\text{ac}}(S)$. Then there exists a constant $C > 0$, depending on N and $\text{diam}(S \cup Y)$, such that for any $\{\mu_i\}_{i=2}^N, \{\tilde{\mu}_i\}_{i=2}^N \subseteq \mathcal{P}(Y)$,*

$$\sum_{i=2}^N \lambda_i \|T_i - \tilde{T}_i\|_{L^2(\rho)} \leq C \left(\left(\sum_{i=2}^N \lambda_i W_1(\mu_i, \tilde{\mu}_i) \right)^{1/2} + \|B - \tilde{B}\|_{L^2(\rho)} \right).$$

Here (T_2, \dots, T_N) and $(\tilde{T}_2, \dots, \tilde{T}_N)$ are the multi-marginal optimal transport maps associated with $(\rho, \mu_2, \dots, \mu_N)$ and $(\rho, \tilde{\mu}_2, \dots, \tilde{\mu}_N)$, respectively.

The first term in Theorem 1.2 corresponds to the *internal multi-marginal mode*. The second term is the *external mode*: by Proposition 1.1, it is the distance between two barycenter maps. Combining this theorem with Mérigot's sharp two-marginal stability estimate [29] yields the following general perturbation estimate.

Theorem 1.3. *Let $N \geq 3$ and let $S, Y \subseteq \mathbb{R}^n$ be compact. Let $\rho \in \mathcal{P}_{\text{ac}}(S) \cap L^\infty(S)$. With the notation of Theorem 1.2, there exists a constant $C > 0$, depending on $\lambda_1, N, \text{diam}(S \cup Y), n$ and $\|\rho\|_{L^\infty}$, such that*

$$\sum_{i=2}^N \lambda_i \|T_i - \tilde{T}_i\|_{L^2(\rho)} \leq C \left(\sum_{i=2}^N \lambda_i W_1(\mu_i, \tilde{\mu}_i) \right)^{1/4}.$$

Moreover, the exponent $\frac{1}{4}$ is optimal in dimension $n \geq 2$.

Remark 1.4. The dimensional restriction $n \geq 2$ in the optimality statement is essential. In dimension $n = 1$, the barycentric multi-marginal optimal transport maps are given by monotone rearrangements on the line (cf. [36, Chapter 1]). Hence

$$\|T_i - \tilde{T}_i\|_{L^2(\rho)} = W_2(\mu_i, \tilde{\mu}_i) \leq \text{diam}(Y)^{1/2} W_1^{1/2}(\mu_i, \tilde{\mu}_i), \quad i = 2, \dots, N.$$

Therefore the general perturbation estimate holds with the optimal exponent $\frac{1}{2}$ in dimension $n = 1$. See Proposition 1.9 for the optimality of this exponent.

The preceding two theorems identify the obstruction to a $\frac{1}{2}$ -Hölder estimate: the external barycenter mode. If $B = \tilde{B}$, then this obstruction disappears and only the internal coercive mode remains, which motivates the following corollary. Notably, in contrast with Theorem 1.3, no L^∞ bound on the density of the source measure is needed in this regime.

Corollary 1.5. *Under the same assumptions and notation as in Theorem 1.2, assume moreover that $B = \tilde{B}$. Then there exists a constant $C > 0$, depending only on N and $\text{diam}(S \cup Y)$, such that*

$$\sum_{i=2}^N \lambda_i \|T_i - \tilde{T}_i\|_{L^2(\rho)} \leq C \left(\sum_{i=2}^N \lambda_i W_1(\mu_i, \tilde{\mu}_i) \right)^{1/2}.$$

Remark 1.6. The exponent $\frac{1}{2}$ in Corollary 1.5 is reminiscent of Ambrosio-type $\frac{1}{2}$ stability estimates for two-marginal optimal transport maps reported in [19]. Here the source of the estimate is different: rather than relying on regularity of the optimal transport map, the estimate follows from the internal quadratic coercivity of the barycentric quadratic cost once the external barycenter mode has been removed.

We say that $\{\tilde{\mu}_i\}_{i=2}^N$ is a *barycenter-preserving perturbation* of $\{\mu_i\}_{i=2}^N$, if $\mu_{\mathbb{P}} = \mu_{\tilde{\mathbb{P}}}$. This condition is equivalent, by uniqueness of the optimal transport map from ρ , to $B = \tilde{B}$. In particular, any weight-preserving permutation of the target marginals preserves the barycenter map. This yields the following self-comparison result. Barycenter-preserving perturbations are not limited to permutations of the marginal measures; in §3.2, we give two non-permutation examples.

Theorem 1.7. *Under the same assumptions and notation as in Theorem 1.2, let σ be a permutation of $\{2, \dots, N\}$ such that*

$$\lambda_{\sigma(i)} = \lambda_i, \quad i = 2, \dots, N.$$

Then there exists a constant $C > 0$, depending on N and $\text{diam}(S \cup Y)$, such that

$$\sum_{i=2}^N \lambda_i \|T_i - T_{\sigma(i)}\|_{L^2(\rho)} \leq C \left(\sum_{i=2}^N \lambda_i W_1(\mu_i, \mu_{\sigma(i)}) \right)^{1/2}.$$

In particular, if $\lambda_i = \lambda_j$, then

$$\|T_i - T_j\|_{L^2(\rho)} \leq C(\lambda_i + \lambda_j)^{-1/2} W_1^{1/2}(\mu_i, \mu_j).$$

Moreover, the exponent $\frac{1}{2}$ is optimal in dimension $n \geq 1$. In dimension $n \geq 2$, if one keeps the exponent $\frac{1}{2}$ in $W_1(\mu_i, \mu_j)$, then the power $-\frac{1}{2}$ of $(\lambda_i + \lambda_j)$ is optimal as well.

Combining Theorem 1.7 with Mériçot’s sharp two-marginal stability estimate gives an upper bound with two scales. The first scale is free of the factor $(\lambda_i + \lambda_j)$ but has exponent $\frac{1}{4}$; the second scale has the stronger exponent $\frac{1}{2}$ but necessarily carries the weight factor.

Corollary 1.8. *Under the same assumptions and notation as in Theorem 1.3, assume moreover that $\lambda_i = \lambda_j$. Then there exists a constant $C > 0$, depending on λ_1 , N , $\text{diam}(S \cup Y)$, n and $\|\rho\|_{L^\infty}$, such that*

$$\|T_i - T_j\|_{L^2(\rho)} \leq C \min \left\{ W_1^{1/4}(\mu_i, \mu_j), (\lambda_i + \lambda_j)^{-1/2} W_1^{1/2}(\mu_i, \mu_j) \right\}.$$

In particular, for any $\theta \in [\frac{1}{4}, \frac{1}{2}]$, one has

$$\|T_i - T_j\|_{L^2(\rho)} \leq C(\lambda_i + \lambda_j)^{1/2-2\theta} W_1^\theta(\mu_i, \mu_j).$$

Moreover, the exponents are optimal in the sense of Proposition 1.9.

The following proposition shows that the exponents and the weight dependence in the preceding estimates cannot be improved within the stated scale of inequalities. We prove it by model examples in §3.1.

Proposition 1.9. *The estimates in Theorem 1.7 and Corollary 1.8 are optimal in the following sense:*

- (a) The exponent $\frac{1}{2}$ in Theorem 1.7 is optimal in dimension $n \geq 1$;
- (b) if the constant is independent of $(\lambda_i + \lambda_j)$, then the exponent $\frac{1}{4}$ is optimal in dimension $n \geq 2$;
- (c) if one keeps the exponent $\frac{1}{2}$ in $W_1(\mu_i, \mu_j)$, then the factor $(\lambda_i + \lambda_j)$ has the optimal exponent $-\frac{1}{2}$ in dimension $n \geq 2$;
- (d) more generally, if one keeps the exponent $\theta \in [\frac{1}{4}, \frac{1}{2}]$ in $W_1(\mu_i, \mu_j)$, then the factor $(\lambda_i + \lambda_j)$ has the optimal exponent $(\frac{1}{2} - 2\theta)$ in dimension $n \geq 2$.

Scope of the two-mode mechanism. Section 4 asks which parts of this argument persist beyond the barycentric quadratic cost. The discussion is organized around two structural tests: quadratic coercivity of the Kantorovich defect in relative directions, and quantitative stability of the remaining collective coordinate. For costs of the form $c_0 + F(\sum_i \lambda_i x_i)$, the internal geometry is unchanged, while a transformed collective map is identified with the optimal transport map to a penalized Wasserstein barycenter. This yields the same two-mode estimate and, under bounded-density assumptions, the same $\frac{1}{4}$ exponent. Uniformly concave costs of the sum follow as a direct corollary.

The later subsections clarify the boundary of the method. Full symmetrization turns a cyclic nearest-neighbour interaction into a complete-graph interaction. For general pairwise quadratic costs, the target–target interaction matrix produces a normal-mode decomposition: some modes are coercive in the Kantorovich defect, while the remaining signed modes would require a separate stability theory. Hedonic costs exhibit an analogous collective variable but lack the two estimates needed to close the argument. Coulomb and determinant costs show that dependence on relative coordinates alone is insufficient when the relative Hessian is singular, indefinite, or sign-indefinite.

When the marginals have densities and the optimal maps are regular, the relations $(T_i)_{\#}\rho = \mu_i$, together with first-order optimality and the barycentric balance condition, form a coupled Monge–Ampère system. The mode decomposition separates the scalar equation for the collective map from the equations for the relative coordinates. In this sense, our estimates provide a weak metric coercivity principle for the system.

Organization of the paper. The present Section 1 states the problem, the two-mode decomposition and the main results. Section 2 proves the stability estimates: Subsections 2.1–2.3 develop the mode decomposition, the Kantorovich-defect coercivity estimate, and the gluing argument, while Subsections 2.4–2.6 prove the main theorems and the two-scale corollary. Section 3 proves optimality and constructs non-permutation barycenter-preserving perturbations. Section 4 studies the scope of the mechanism: it first gives rigorous extensions to collective-coordinate perturbations and uniformly concave costs of the sum, and then discusses cyclic symmetrization, normal-mode decompositions for pairwise quadratic interactions, and the obstructions arising for hedonic, Coulomb and determinant costs.

2 Mode Decomposition and Stability Estimates

2.1 Barycentric and internal coordinates

The proof of Theorem 1.2 relies on separating the barycentric direction from the internal multi-marginal directions. We first introduce the corresponding barycentric coordinates.

For $\{\mu_i\}_{i=2}^N, \{\tilde{\mu}_i\}_{i=2}^N \subseteq \mathcal{P}(Y)$ and $x_1 \in S$, define the target-side barycenters

$$M(x_1) := \frac{B(x_1) - \lambda_1 x_1}{1 - \lambda_1} = \frac{1}{1 - \lambda_1} \sum_{i=2}^N \lambda_i T_i(x_1),$$

$$\tilde{M}(x_1) := \frac{\tilde{B}(x_1) - \lambda_1 x_1}{1 - \lambda_1} = \frac{1}{1 - \lambda_1} \sum_{i=2}^N \lambda_i \tilde{T}_i(x_1).$$

For $i = 2, \dots, N$, define the internal directions

$$V_i(x_1) := T_i(x_1) - M(x_1), \quad \tilde{V}_i(x_1) := \tilde{T}_i(x_1) - \tilde{M}(x_1), \quad x_1 \in S.$$

Then we have

$$\sum_{i=2}^N \lambda_i V_i(x_1) = 0, \quad \sum_{i=2}^N \lambda_i \tilde{V}_i(x_1) = 0, \quad x_1 \in S.$$

For $x_1 \in S, x_2, \dots, x_N \in Y$, define

$$B_x := \lambda_1 x_1 + \sum_{i=2}^N \lambda_i x_i, \quad M_x := \frac{B_x - \lambda_1 x_1}{1 - \lambda_1} = \frac{1}{1 - \lambda_1} \sum_{i=2}^N \lambda_i x_i;$$

and

$$V_{i,x} := x_i - M_x, \quad i = 2, \dots, N, \quad \text{which implies} \quad \sum_{i=2}^N \lambda_i V_{i,x} = 0.$$

Here B_x, M_x and $V_{i,x}$ are the pointwise analogues of B, M and V_i associated with the tuple (x_1, \dots, x_N) .

Remark 2.1 (Geometric interpretation of the decomposition). The decomposition $T_i = M + V_i$ with $\sum_{i=2}^N \lambda_i V_i = 0$, is analogous to the center-of-mass and Jacobi-coordinate decomposition in classical mechanics. In the Wasserstein setting, the barycenter map B is the averaged coordinate, while the internal directions $(V_i)_{i \geq 2}$ describe multi-marginal couplings with the same barycenter. The quadratic coercivity in Proposition 2.3 is the positive definiteness of the barycentric quadratic cost in these relative directions. Thus the stability loss comes from the averaged direction, namely the two-marginal barycenter map.

Lemma 2.2. For $x_1 \in S, x_2, \dots, x_N \in Y$, we have

$$c(x_1, \dots, x_N) = \frac{\lambda_1}{2(1 - \lambda_1)} |B_x - x_1|^2 + \frac{1}{2} \sum_{i=2}^N \lambda_i |V_{i,x}|^2. \quad (2.1)$$

Proof. Since $V_{i,x} := x_i - M_x$ for $i = 2, \dots, N$, by (1.4), we have

$$\begin{aligned} c(x_1, \dots, x_N) &= \sum_{i=1}^N \frac{\lambda_i}{2} |x_i - B_x|^2 \\ &= \frac{\lambda_1}{2} |x_1 - B_x|^2 + \sum_{i=2}^N \frac{\lambda_i}{2} |V_{i,x} + M_x - B_x|^2 \\ &\stackrel{*}{=} \frac{\lambda_1}{2} |x_1 - B_x|^2 + \sum_{i=2}^N \frac{\lambda_i}{2} |M_x - B_x|^2 + \sum_{i=2}^N \frac{\lambda_i}{2} |V_{i,x}|^2 \\ &= \frac{\lambda_1}{2(1 - \lambda_1)} |B_x - x_1|^2 + \frac{1}{2} \sum_{i=2}^N \lambda_i |V_{i,x}|^2, \end{aligned} \quad (2.2)$$

where (*) follows from $\sum_{i=2}^N \lambda_i V_{i,x} = 0$. □

2.2 Separation of barycentric and internal directions

Let $(\varphi_1, \dots, \varphi_N)$ be a c -conjugate tuple of Kantorovich potentials for the multi-marginal optimal transport problem $(\rho, \mu_2, \dots, \mu_N)$. For $x_1 \in S$, $x_2, \dots, x_N \in Y$, define

$$R(x_1, \dots, x_N) := c(x_1, \dots, x_N) - \sum_{i=1}^N \varphi_i(x_i).$$

Since $(\varphi_1, \dots, \varphi_N)$ is c -conjugate, we have

$$R(x_1, \dots, x_N) \geq 0, \quad \forall (x_1, \dots, x_N) \in S \times Y^{N-1}.$$

Moreover, by the standard duality theorem for marginal problems (cf. [22]), in the form used for multi-marginal optimal transport (cf. [17], [24]), we have

$$R(x_1, T_2(x_1), \dots, T_N(x_1)) = 0, \quad \rho\text{-a.e. } x_1 \in S. \quad (2.3)$$

The following proposition is the key estimate in the proof of Theorem 1.2.

Proposition 2.3. *For ρ -a.e. $x_1 \in S$ and every $x_2, \dots, x_N \in Y$, we have*

$$R(x_1, x_2, \dots, x_N) \geq \frac{1}{4} \sum_{i=2}^N \lambda_i^2 |V_{i,x} - V_i(x_1)|^2 - \frac{3}{2} |B_x - B(x_1)|^2. \quad (2.4)$$

Proof. Fix $x_1 \in S$ such that

$$R(x_1, T_2(x_1), \dots, T_N(x_1)) = 0. \quad (2.5)$$

Denote

$$\Delta x_i := x_i - T_i(x_1), \quad \Delta V_i := V_{i,x} - V_i(x_1), \quad i = 2, \dots, N.$$

Then we have

$$\Delta B := B_x - B(x_1) = \sum_{i=2}^N \lambda_i \Delta x_i, \quad \sum_{i=2}^N \lambda_i \Delta V_i = 0.$$

Moreover, note that

$$\Delta V_i = \Delta x_i - (M_x - M(x_1)) = \Delta x_i - \frac{\Delta B}{1 - \lambda_1}, \quad i = 2, \dots, N, \quad (2.6)$$

by Lemma 2.2 and the fact that c is quadratic, we have the exact expansion

$$c(x_1, \dots, x_N) = c(x_1, T_2(x_1), \dots, T_N(x_1)) + \sum_{i=2}^N \langle p_i, \Delta x_i \rangle + Q(\Delta B, \Delta V_i), \quad (2.7)$$

where

$$p_i := \nabla_{x_i} c(x_1, T_2(x_1), \dots, T_N(x_1)), \quad Q(\Delta B, \Delta V_i) := \frac{\lambda_1}{2(1 - \lambda_1)} |\Delta B|^2 + \frac{1}{2} \sum_{i=2}^N \lambda_i |\Delta V_i|^2.$$

On the other hand, for $i = 2, \dots, N$,

$$\begin{aligned}
\varphi_i(x_i) &\leq c(x_1, \dots, T_{i-1}(x_1), x_i, T_{i+1}(x_1), \dots, T_N(x_1)) \\
&\quad - \varphi_1(x_1) - \sum_{\substack{2 \leq j \leq N \\ j \neq i}} \varphi_j(T_j(x_1)) \\
&= c(x_1, \dots, T_i(x_1), \dots, T_N(x_1)) + \langle p_i, \Delta x_i \rangle \\
&\quad + \frac{\lambda_i(1-\lambda_i)}{2} |\Delta x_i|^2 - \varphi_1(x_1) - \sum_{\substack{2 \leq j \leq N \\ j \neq i}} \varphi_j(T_j(x_1)) \\
&\stackrel{(2.5)}{=} \varphi_i(T_i(x_1)) + \langle p_i, \Delta x_i \rangle + \frac{\lambda_i(1-\lambda_i)}{2} |\Delta x_i|^2.
\end{aligned} \tag{2.8}$$

Thus,

$$\begin{aligned}
&R(x_1, \dots, x_N) \\
&\stackrel{(2.5)}{=} R(x_1, \dots, x_N) - R(x_1, T_2(x_1), \dots, T_N(x_1)) \\
&\geq \frac{\lambda_1}{2(1-\lambda_1)} |\Delta B|^2 + \frac{1}{2} \sum_{i=2}^N \lambda_i |\Delta V_i|^2 - \sum_{i=2}^N \frac{\lambda_i(1-\lambda_i)}{2} |\Delta x_i|^2 \\
&\stackrel{(2.6)}{=} \frac{1}{2} \sum_{i=2}^N \lambda_i^2 |\Delta V_i|^2 + \left(\frac{\sum_{i=2}^N \lambda_i^2}{2(1-\lambda_1)^2} - \frac{1}{2} \right) |\Delta B|^2 - \frac{1}{1-\lambda_1} \sum_{i=2}^N \lambda_i(1-\lambda_i) \langle \Delta V_i, \Delta B \rangle \\
&\geq \frac{1}{2} \sum_{i=2}^N \lambda_i^2 |\Delta V_i|^2 - \frac{1}{2} |\Delta B|^2 + \frac{1}{1-\lambda_1} \langle \sum_{i=2}^N \lambda_i^2 \Delta V_i, \Delta B \rangle \\
&\geq^* \frac{1}{2} \sum_{i=2}^N \lambda_i^2 |\Delta V_i|^2 - \frac{1}{2} |\Delta B|^2 - \frac{|\Delta B|}{1-\lambda_1} \left(\sum_{i=2}^N \lambda_i^2 \right)^{1/2} \left(\sum_{i=2}^N \lambda_i^2 |\Delta V_i|^2 \right)^{1/2} \\
&\geq^{**} \frac{1}{4} \sum_{i=2}^N \lambda_i^2 |\Delta V_i|^2 - \frac{3}{2} |\Delta B|^2,
\end{aligned} \tag{2.9}$$

where (*) follows from Hölder's inequality

$$\left| \langle \sum_{i=2}^N \lambda_i^2 \Delta V_i, \Delta B \rangle \right| \leq |\Delta B| \left(\sum_{i=2}^N \lambda_i^2 \right)^{1/2} \left(\sum_{i=2}^N \lambda_i^2 |\Delta V_i|^2 \right)^{1/2};$$

and (**) follows from Young's inequality and the fact

$$\left(\sum_{i=2}^N \lambda_i^2 \right)^{1/2} \leq \sum_{i=2}^N \lambda_i = 1 - \lambda_1.$$

□

2.3 Gluing argument for perturbed marginals

Denote

$$D := \text{diam}(S \cup Y), \quad \delta := \sum_{i=2}^N \lambda_i W_1(\mu_i, \tilde{\mu}_i) \leq D;$$

and denote

$$\gamma := (\text{Id}, T_2, \dots, T_N) \# \rho, \quad \tilde{\gamma} := (\text{Id}, \tilde{T}_2, \dots, \tilde{T}_N) \# \rho.$$

Lemma 2.4. *There exists a coupling $\bar{\gamma} \in \Pi(\rho, \mu_2, \dots, \mu_N, \tilde{\mu}_2, \dots, \tilde{\mu}_N)$ with coordinates $(x_1, x_2, \dots, x_N, \tilde{x}_2, \dots, \tilde{x}_N)$, such that*

$$(x_1, \tilde{x}_2, \dots, \tilde{x}_N)_{\#} \bar{\gamma} = \tilde{\gamma}, \quad (x_i, \tilde{x}_i)_{\#} \bar{\gamma} = \pi_i, \quad i = 2, \dots, N, \quad (2.10)$$

where $\pi_i \in \Pi(\mu_i, \tilde{\mu}_i)$ is a W_1 -optimal transport plan between μ_i and $\tilde{\mu}_i$.

Proof. This follows by applying the gluing lemma repeatedly (cf. [2, Lemma 2.1]). \square

Corollary 2.5. *We have*

$$\left(\int |B_x - \tilde{B}(x_1)|^2 d\bar{\gamma} \right)^{1/2} \leq D^{1/2} \delta^{1/2}, \quad \left(\int \sum_{i=2}^N \lambda_i^2 |V_{i,x} - \tilde{V}_i(x_1)|^2 d\bar{\gamma} \right)^{1/2} \leq D^{1/2} \delta^{1/2}.$$

Here B_x and $V_{i,x}$ are associated with the coordinates (x_1, x_2, \dots, x_N) .

Proof. Since

$$(x_1, \tilde{x}_2, \dots, \tilde{x}_N)_{\#} \bar{\gamma} = \tilde{\gamma} = (\text{Id}, \tilde{T}_2, \dots, \tilde{T}_N)_{\#} \rho,$$

we have

$$\tilde{x}_i = \tilde{T}_i(x_1), \quad \bar{\gamma}\text{-a.e.}, \quad i = 2, \dots, N.$$

Estimate of the barycentric part: By the definitions of B_x and $\tilde{B}(x_1)$, we have

$$\begin{aligned} \left(\int |B_x - \tilde{B}(x_1)|^2 d\bar{\gamma} \right)^{1/2} &= \left(\int \left| \sum_{i=2}^N \lambda_i (x_i - \tilde{T}_i(x_1)) \right|^2 d\bar{\gamma} \right)^{1/2} \\ &\leq \sum_{i=2}^N \lambda_i \left(\int |x_i - \tilde{x}_i|^2 d\bar{\gamma} \right)^{1/2} \\ &\stackrel{(2.10)}{\leq} D^{1/2} \sum_{i=2}^N \lambda_i W_1^{1/2}(\mu_i, \tilde{\mu}_i) \\ &\leq D^{1/2} \left(\sum_{i=2}^N \lambda_i W_1(\mu_i, \tilde{\mu}_i) \right)^{1/2} = D^{1/2} \delta^{1/2}. \end{aligned} \quad (2.11)$$

Estimate of the internal part: By the definitions of $V_{i,x}$ and $\tilde{V}_i(x_1)$, we have

$$V_{i,x} - \tilde{V}_i(x_1) = (x_i - \tilde{T}_i(x_1)) - \frac{1}{1 - \lambda_1} \sum_{j=2}^N \lambda_j (x_j - \tilde{T}_j(x_1)). \quad (2.12)$$

Using $\lambda_i < 1$ for $i = 2, \dots, N$, we have

$$\begin{aligned} \sum_{i=2}^N \lambda_i^2 |V_{i,x} - \tilde{V}_i(x_1)|^2 &\leq \sum_{i=2}^N \lambda_i |V_{i,x} - \tilde{V}_i(x_1)|^2 \\ &= \sum_{i=2}^N \lambda_i |x_i - \tilde{T}_i(x_1)|^2 - \frac{1}{1 - \lambda_1} \left| \sum_{i=2}^N \lambda_i (x_i - \tilde{T}_i(x_1)) \right|^2 \\ &\leq \sum_{i=2}^N \lambda_i |x_i - \tilde{T}_i(x_1)|^2. \end{aligned} \quad (2.13)$$

Integrating with respect to $\bar{\gamma}$ implies

$$\int \sum_{i=2}^N \lambda_i^2 |V_{i,x} - \tilde{V}_i(x_1)|^2 d\bar{\gamma} \leq \sum_{i=2}^N \lambda_i \int |x_i - \tilde{x}_i|^2 d\bar{\gamma} \stackrel{(2.10)}{\leq} D \sum_{i=2}^N \lambda_i W_1(\mu_i, \tilde{\mu}_i) = D\delta, \quad (2.14)$$

which proves the estimates. \square

Denote

$$\hat{\gamma} := (x_1, x_2, \dots, x_N)_{\#} \bar{\gamma} \in \Pi(\rho, \mu_2, \dots, \mu_N).$$

Lemma 2.6. *We have*

$$\left| \int c(x_1, \dots, x_N) d\hat{\gamma} - \int c(x_1, \dots, x_N) d\gamma \right| \leq 2D\delta. \quad (2.15)$$

Proof. Recall that $\Omega = \overline{\text{conv}(S \cup Y)}$. Since $\text{diam}(\Omega) = D$, for any $x_1 \in S, x_2, \dots, x_N \in Y$, we have

$$|\nabla_{x_i} c(x_1, \dots, x_N)| = |\lambda_i(x_i - B_x)| \leq \lambda_i \text{diam}(\Omega) = \lambda_i D, \quad i = 2, \dots, N. \quad (2.16)$$

Then for $(x_2, \dots, x_N), (\tilde{x}_2, \dots, \tilde{x}_N) \in Y^{N-1}$, by the triangle inequality, we have

$$|c(x_1, x_2, \dots, x_N) - c(x_1, \tilde{x}_2, \dots, \tilde{x}_N)| \leq D \sum_{i=2}^N \lambda_i |x_i - \tilde{x}_i|. \quad (2.17)$$

Thus,

$$\begin{aligned} & \left| \int c(x_1, \dots, x_N) d\hat{\gamma} - \int c(x_1, \dots, x_N) d\bar{\gamma} \right| \\ & \stackrel{(2.10)}{=} \left| \int c(x_1, x_2, \dots, x_N) - c(x_1, \tilde{x}_2, \dots, \tilde{x}_N) d\bar{\gamma} \right| \\ & \stackrel{(2.17)}{\leq} D \sum_{i=2}^N \lambda_i \int |x_i - \tilde{x}_i| d\bar{\gamma} \stackrel{(2.10)}{=} D\delta. \end{aligned} \quad (2.18)$$

Since γ is the optimal transport plan of $(\rho, \mu_2, \dots, \mu_N)$, we have

$$\int c(x_1, \dots, x_N) d\gamma \leq \int c(x_1, \dots, x_N) d\hat{\gamma} \leq \int c(x_1, \dots, x_N) d\bar{\gamma} + D\delta. \quad (2.19)$$

By the same argument with the two target tuples interchanged, there also exists a coupling $\underline{\gamma} \in \Pi(\rho, \mu_2, \dots, \mu_N, \tilde{\mu}_2, \dots, \tilde{\mu}_N)$ with coordinates $(x_1, x_2, \dots, x_N, \tilde{x}_2, \dots, \tilde{x}_N)$, such that

$$(x_1, x_2, \dots, x_N)_{\#} \underline{\gamma} = \gamma, \quad (x_i, \tilde{x}_i)_{\#} \underline{\gamma} = \pi_i, \quad i = 2, \dots, N.$$

Denote

$$\tilde{\gamma} := (x_1, \tilde{x}_2, \dots, \tilde{x}_N)_{\#} \underline{\gamma} \in \Pi(\rho, \tilde{\mu}_2, \dots, \tilde{\mu}_N).$$

Similarly, we have

$$\int c(x_1, \dots, x_N) d\tilde{\gamma} \leq \int c(x_1, \dots, x_N) d\bar{\gamma} \leq \int c(x_1, \dots, x_N) d\gamma + D\delta. \quad (2.20)$$

Combining (2.18), (2.19) and (2.20) completes the proof. \square

2.4 Proof of Theorem 1.2

Proof of Theorem 1.2. Integrating (2.4) with respect to $\hat{\gamma}$ gives

$$\frac{1}{4} \int \sum_{i=2}^N \lambda_i^2 |V_{i,x} - V_i(x_1)|^2 d\hat{\gamma} - \frac{3}{2} \int |B_x - B(x_1)|^2 d\hat{\gamma} \leq \int R(x_1, x_2, \dots, x_N) d\hat{\gamma}. \quad (2.21)$$

Note that

$$\begin{aligned} \int R(x_1, x_2, \dots, x_N) d\hat{\gamma} &= \int c(x_1, \dots, x_N) d\hat{\gamma} - \sum_{i=1}^N \int \varphi_i(x_i) d\hat{\gamma} \\ &= \int c(x_1, \dots, x_N) d\hat{\gamma} - \int \varphi_1(x_1) d\rho - \sum_{i=2}^N \int \varphi_i(x_i) d\mu_i \quad (2.22) \\ &\stackrel{(2.3)}{=} \int c(x_1, \dots, x_N) d\hat{\gamma} - \int c(x_1, \dots, x_N) d\gamma \stackrel{(2.15)}{\leq} 2D\delta. \end{aligned}$$

Moreover, by Corollary 2.5, we have

$$\begin{aligned} \left(\int |B_x - B(x_1)|^2 d\hat{\gamma} \right)^{1/2} &\leq \left(\int |B_x - \tilde{B}(x_1)|^2 d\hat{\gamma} \right)^{1/2} + \left(\int |\tilde{B}(x_1) - B(x_1)|^2 d\hat{\gamma} \right)^{1/2} \\ &\leq D^{1/2} \delta^{1/2} + \|B - \tilde{B}\|_{L^2(\rho)}. \end{aligned} \quad (2.23)$$

Thus, by (2.21), (2.22), (2.23) and Corollary 2.5, we have

$$\begin{aligned} &\left(\int \sum_{i=2}^N \lambda_i^2 |V_i(x_1) - \tilde{V}_i(x_1)|^2 d\rho \right)^{1/2} = \left(\int \sum_{i=2}^N \lambda_i^2 |V_i(x_1) - \tilde{V}_i(x_1)|^2 d\hat{\gamma} \right)^{1/2} \\ &\leq \left(\int \sum_{i=2}^N \lambda_i^2 |V_i(x_1) - V_{i,x}|^2 d\hat{\gamma} \right)^{1/2} + \left(\int \sum_{i=2}^N \lambda_i^2 |V_{i,x} - \tilde{V}_i(x_1)|^2 d\hat{\gamma} \right)^{1/2} \quad (2.24) \\ &\leq \left(6 \left(D^{1/2} \delta^{1/2} + \|B - \tilde{B}\|_{L^2(\rho)} \right)^2 + 8D\delta \right)^{1/2} + D^{1/2} \delta^{1/2} \\ &\leq C_1 \left(\delta^{1/2} + \|B - \tilde{B}\|_{L^2(\rho)} \right), \end{aligned}$$

where C_1 depends only on D .

On the other hand, by the definitions of V_i, \tilde{V}_i, B , and \tilde{B} , we have

$$T_i(x_1) - \tilde{T}_i(x_1) = (V_i(x_1) - \tilde{V}_i(x_1)) + \frac{1}{1 - \lambda_1} (B(x_1) - \tilde{B}(x_1)), \quad i = 2, \dots, N. \quad (2.25)$$

Therefore, by Hölder's inequality,

$$\begin{aligned} \sum_{i=2}^N \lambda_i \|T_i - \tilde{T}_i\|_{L^2(\rho)} &\leq \sum_{i=2}^N \lambda_i \|V_i - \tilde{V}_i\|_{L^2(\rho)} + \frac{1}{1 - \lambda_1} \sum_{i=2}^N \lambda_i \|B - \tilde{B}\|_{L^2(\rho)} \\ &\leq (N-1)^{1/2} \left(\sum_{i=2}^N \lambda_i^2 \|V_i - \tilde{V}_i\|_{L^2(\rho)}^2 \right)^{1/2} + \|B - \tilde{B}\|_{L^2(\rho)} \quad (2.26) \\ &\leq C \left(\left(\sum_{i=2}^N \lambda_i W_1(\mu_i, \tilde{\mu}_i) \right)^{1/2} + \|B - \tilde{B}\|_{L^2(\rho)} \right), \end{aligned}$$

where C depends on $\text{diam}(S \cup Y)$ and N . This completes the proof. \square

2.5 Proof of Theorem 1.3

We combine Theorem 1.2 with the following estimate from [29, Theorem 2.2].

Proposition 2.7. *Let $S, Y \subseteq \mathbb{R}^n$ be compact and let $\Omega = \overline{\text{conv}(S \cup Y)}$. Let $\rho \in \mathcal{P}_{\text{ac}}(S) \cap L^\infty(S)$. Then there exists a constant $C_2 > 0$, depending on $\text{diam}(S \cup Y)$, n , and $\|\rho\|_{L^\infty}$, such that for any $\mu, \tilde{\mu} \in \mathcal{P}(\Omega)$,*

$$\|\nabla\phi_\mu - \nabla\phi_{\tilde{\mu}}\|_{L^2(\rho)}^4 \leq C_2 \int_\Omega (\psi_\mu - \psi_{\tilde{\mu}}) d(\mu - \tilde{\mu}), \quad (2.27)$$

where ψ_μ and $\psi_{\tilde{\mu}}$ are the Kantorovich potentials from μ to ρ and $\tilde{\mu}$ to ρ , respectively, for the quadratic cost; $\phi_\mu := \psi_\mu^c$ and $\phi_{\tilde{\mu}} := \psi_{\tilde{\mu}}^c$ are the corresponding dual Kantorovich potentials.

Proof of Theorem 1.3. By Proposition 1.1, \mathbb{P} and $\tilde{\mathbb{P}}$ admit unique Wasserstein barycenters $\mu_{\mathbb{P}}, \mu_{\tilde{\mathbb{P}}} \in \mathcal{P}(\Omega)$. Moreover, by the global zero-order balance condition (cf. [1], [7, Proposition 1.1]), there exist Kantorovich potentials $\{\psi_{\mu_{\mathbb{P}}}^i\}_{i=1}^N$ such that

$$\sum_{i=1}^N \lambda_i \psi_{\mu_{\mathbb{P}}}^i(y) = 0, \quad \forall y \in \Omega, \quad (2.28)$$

where $\psi_{\mu_{\mathbb{P}}}^i$ is the Kantorovich potential from $\mu_{\mathbb{P}}$ to μ_i for $i = 1, \dots, N$ with $\mu_1 = \rho$. Similarly, there exist Kantorovich potentials $\{\psi_{\mu_{\tilde{\mathbb{P}}}}^i\}_{i=1}^N$ satisfying the analogue of (2.28), where $\psi_{\mu_{\tilde{\mathbb{P}}}}^i$ is the Kantorovich potential from $\mu_{\tilde{\mathbb{P}}}$ to $\tilde{\mu}_i$ with $\tilde{\mu}_1 = \rho$.

For $\mu_i, \tilde{\mu}_i \in \mathcal{P}(\Omega)$, $i = 1, \dots, N$, denote

$$D_{\mu_i}(\mu_{\tilde{\mathbb{P}}}, \mu_{\mathbb{P}}) := \frac{1}{2}W_2^2(\mu_{\tilde{\mathbb{P}}}, \mu_i) - \frac{1}{2}W_2^2(\mu_{\mathbb{P}}, \mu_i) - \int_\Omega \psi_{\mu_{\mathbb{P}}}^i d(\mu_{\tilde{\mathbb{P}}} - \mu_{\mathbb{P}}), \quad (2.29)$$

and

$$D_{\tilde{\mu}_i}(\mu_{\mathbb{P}}, \mu_{\tilde{\mathbb{P}}}) := \frac{1}{2}W_2^2(\mu_{\mathbb{P}}, \tilde{\mu}_i) - \frac{1}{2}W_2^2(\mu_{\tilde{\mathbb{P}}}, \tilde{\mu}_i) - \int_\Omega \psi_{\mu_{\tilde{\mathbb{P}}}}^i d(\mu_{\mathbb{P}} - \mu_{\tilde{\mathbb{P}}}). \quad (2.30)$$

By Kantorovich duality, it holds that $D_{\mu_i}(\mu_{\tilde{\mathbb{P}}}, \mu_{\mathbb{P}}) \geq 0$ and $D_{\tilde{\mu}_i}(\mu_{\mathbb{P}}, \mu_{\tilde{\mathbb{P}}}) \geq 0$. Moreover, since $\mu_1 = \tilde{\mu}_1 = \rho$ and $\rho \in \mathcal{P}_{\text{ac}}(S) \cap L^\infty(S)$, we have

$$\begin{aligned} D_\rho(\mu_{\mathbb{P}}, \mu_{\tilde{\mathbb{P}}}) + D_\rho(\mu_{\tilde{\mathbb{P}}}, \mu_{\mathbb{P}}) &= \int (\psi_{\mu_{\tilde{\mathbb{P}}}}^1 - \psi_{\mu_{\mathbb{P}}}^1) d(\mu_{\tilde{\mathbb{P}}} - \mu_{\mathbb{P}}) \geq C_2^{-1} \|\nabla\phi_{\mu_{\mathbb{P}}}^1 - \nabla\phi_{\mu_{\tilde{\mathbb{P}}}}^1\|_{L^2(\rho)}^4 \\ &\stackrel{*}{=} C_2^{-1} \|B - \tilde{B}\|_{L^2(\rho)}^4, \end{aligned} \quad (2.31)$$

where $\phi_{\mu_{\mathbb{P}}}^1 := (\psi_{\mu_{\mathbb{P}}}^1)^c$ and $\phi_{\mu_{\tilde{\mathbb{P}}}}^1 := (\psi_{\mu_{\tilde{\mathbb{P}}}}^1)^c$, and (*) follows from

$$B(x_1) = x_1 - \nabla\phi_{\mu_{\mathbb{P}}}^1(x_1), \quad \tilde{B}(x_1) = x_1 - \nabla\phi_{\mu_{\tilde{\mathbb{P}}}}^1(x_1), \quad \rho\text{-a.e. } x_1 \in S.$$

Therefore, we have

$$\text{Var}_{\mathbb{P}}(\mu_{\tilde{\mathbb{P}}}) - \text{Var}_{\mathbb{P}}(\mu_{\mathbb{P}}) = \sum_{i=1}^N \lambda_i D_{\mu_i}(\mu_{\tilde{\mathbb{P}}}, \mu_{\mathbb{P}}) \geq \lambda_1 D_\rho(\mu_{\tilde{\mathbb{P}}}, \mu_{\mathbb{P}}), \quad (2.32)$$

and

$$\text{Var}_{\tilde{\mathbb{P}}}(\mu_{\mathbb{P}}) - \text{Var}_{\tilde{\mathbb{P}}}(\mu_{\tilde{\mathbb{P}}}) = \sum_{i=1}^N \lambda_i D_{\tilde{\mu}_i}(\mu_{\mathbb{P}}, \mu_{\tilde{\mathbb{P}}}) \geq \lambda_1 D_\rho(\mu_{\mathbb{P}}, \mu_{\tilde{\mathbb{P}}}), \quad (2.33)$$

which implies

$$\mathrm{Var}_{\mathbb{P}}(\mu_{\tilde{\mathbb{P}}}) - \mathrm{Var}_{\mathbb{P}}(\mu_{\mathbb{P}}) + \mathrm{Var}_{\tilde{\mathbb{P}}}(\mu_{\mathbb{P}}) - \mathrm{Var}_{\tilde{\mathbb{P}}}(\mu_{\tilde{\mathbb{P}}}) \geq \lambda_1 C_2^{-1} \|B - \tilde{B}\|_{L^2(\rho)}^4. \quad (2.34)$$

On the other hand, for $i = 2, \dots, N$, by the gluing lemma, there exists a coupling $\gamma_i \in \Pi(\mu_{\mathbb{P}}, \mu_i, \tilde{\mu}_i)$ with coordinates (y, x_i, \tilde{x}_i) , such that

$$(x_i, \tilde{x}_i)_{\#} \gamma_i = \pi_i, \quad (y, x_i)_{\#} \gamma_i = \pi_i^b,$$

where $\pi_i \in \Pi(\mu_i, \tilde{\mu}_i)$ is a W_1 -optimal transport plan between μ_i and $\tilde{\mu}_i$; $\pi_i^b \in \Pi(\mu_{\mathbb{P}}, \mu_i)$ is the W_2 -optimal transport plan between $\mu_{\mathbb{P}}$ and μ_i . Denote

$$\tilde{\pi}_i^b := (y, \tilde{x}_i)_{\#} \gamma_i \in \Pi(\mu_{\mathbb{P}}, \tilde{\mu}_i).$$

Then we have

$$\begin{aligned} \frac{1}{2} W_2^2(\mu_{\mathbb{P}}, \tilde{\mu}_i) - \frac{1}{2} W_2^2(\mu_{\mathbb{P}}, \mu_i) &\leq \frac{1}{2} \int |y - \tilde{x}_i|^2 d\tilde{\pi}_i^b - \frac{1}{2} \int |y - x_i|^2 d\pi_i^b \\ &= \frac{1}{2} \int |y - \tilde{x}_i|^2 - |y - x_i|^2 d\gamma_i \\ &\leq D \int |x_i - \tilde{x}_i| d\gamma_i = DW_1(\mu_i, \tilde{\mu}_i). \end{aligned} \quad (2.35)$$

Hence,

$$\mathrm{Var}_{\tilde{\mathbb{P}}}(\mu_{\mathbb{P}}) - \mathrm{Var}_{\mathbb{P}}(\mu_{\mathbb{P}}) = \sum_{i=2}^N \frac{\lambda_i}{2} (W_2^2(\mu_{\mathbb{P}}, \tilde{\mu}_i) - W_2^2(\mu_{\mathbb{P}}, \mu_i)) \leq D\delta. \quad (2.36)$$

Similarly, we have

$$\mathrm{Var}_{\mathbb{P}}(\mu_{\tilde{\mathbb{P}}}) - \mathrm{Var}_{\tilde{\mathbb{P}}}(\mu_{\tilde{\mathbb{P}}}) \leq D\delta. \quad (2.37)$$

Thus, by (2.34), (2.36) and (2.37), we have

$$\|B - \tilde{B}\|_{L^2(\rho)} \leq C_3 \lambda_1^{-1/4} \delta^{1/4}, \quad (2.38)$$

where C_3 depends on $\mathrm{diam}(S \cup Y)$, n and $\|\rho\|_{L^\infty}$. Combining this with Theorem 1.2 gives the estimate.

Optimality of the exponent: For $n \geq 2$, let $X, Y \subseteq \mathbb{R}^n$ and ρ be the uniform source measure given by [27, Theorem 1.2] with $p = 1$. Denote $S := \overline{X}$. Then $\rho \in \mathcal{P}_{\mathrm{ac}}(S) \cap L^\infty(S)$ and for any $\alpha > \frac{1}{4}$, there exist $\mu^k, \tilde{\mu}^k \in \mathcal{P}(Y)$ such that, if T^k and \tilde{T}^k denote the unique optimal transport maps from ρ to μ^k and from ρ to $\tilde{\mu}^k$ respectively, then

$$\frac{\|T^k - \tilde{T}^k\|_{L^2(\rho)}}{W_1^\alpha(\mu^k, \tilde{\mu}^k)} \rightarrow +\infty. \quad (2.39)$$

For each $k \in \mathbb{N}$, define $\mu_2^k = \dots = \mu_N^k = \mu^k$, $\tilde{\mu}_2^k = \dots = \tilde{\mu}_N^k = \tilde{\mu}^k$. Then for any $\pi \in \Pi(\rho, \mu_2^k, \dots, \mu_N^k)$, by (1.4), we have

$$\int c(x_1, \dots, x_N) d\pi \geq \frac{1}{2} \sum_{i=2}^N \lambda_1 \lambda_i \int |x_1 - x_i|^2 d\pi \geq \frac{1}{2} \lambda_1 (1 - \lambda_1) W_2^2(\rho, \mu^k). \quad (2.40)$$

By [17] (see also [24]), equality holds if and only if $\pi = (\mathrm{Id}, T^k, \dots, T^k)_{\#} \rho$, which implies the multi-marginal optimal transport maps associated with $(\rho, \mu_2^k, \dots, \mu_N^k)$ and $(\rho, \tilde{\mu}_2^k, \dots, \tilde{\mu}_N^k)$ are (T^k, \dots, T^k) and $(\tilde{T}^k, \dots, \tilde{T}^k)$. Therefore,

$$\sum_{i=2}^N \lambda_i \|T_i^k - \tilde{T}_i^k\|_{L^2(\rho)} = (1 - \lambda_1) \|T^k - \tilde{T}^k\|_{L^2(\rho)}, \quad \sum_{i=2}^N \lambda_i W_1(\mu_i^k, \tilde{\mu}_i^k) = (1 - \lambda_1) W_1(\mu^k, \tilde{\mu}^k). \quad (2.41)$$

Combining this with (2.39) proves the optimality. \square

2.6 Proof of Theorem 1.7

The optimality of the exponents is proved in §3.1.

Proof of Theorem 1.7. Denote

$$\tilde{\mu}_i := \mu_{\sigma(i)}, \quad \tilde{T}_i := T_{\sigma(i)}, \quad i = 2, \dots, N.$$

Since $\lambda_{\sigma(i)} = \lambda_i$, the barycentric quadratic cost $c(x_1, \dots, x_N)$ is invariant under the permutation σ . Hence $(\tilde{T}_2, \dots, \tilde{T}_N)$ are the multi-marginal optimal transport maps associated with $(\rho, \tilde{\mu}_2, \dots, \tilde{\mu}_N)$. Moreover,

$$\tilde{B} = \lambda_1 \text{Id} + \sum_{i=2}^N \lambda_i \tilde{T}_i = \lambda_1 \text{Id} + \sum_{i=2}^N \lambda_i T_{\sigma(i)} = \lambda_1 \text{Id} + \sum_{i=2}^N \lambda_{\sigma(i)} T_{\sigma(i)} = B. \quad (2.42)$$

Corollary 1.5 therefore gives

$$\sum_{i=2}^N \lambda_i \|T_i - T_{\sigma(i)}\|_{L^2(\rho)} = \sum_{i=2}^N \lambda_i \|T_i - \tilde{T}_i\|_{L^2(\rho)} \leq C \left(\sum_{i=2}^N \lambda_i W_1(\mu_i, \mu_{\sigma(i)}) \right)^{1/2}, \quad (2.43)$$

where C depends only on N and $\text{diam}(S \cup Y)$.

If σ is the transposition exchanging i and j , then

$$\sum_{k=2}^N \lambda_k \|T_k - T_{\sigma(k)}\|_{L^2(\rho)} = (\lambda_i + \lambda_j) \|T_i - T_j\|_{L^2(\rho)}, \quad (2.44)$$

and

$$\sum_{k=2}^N \lambda_k W_1(\mu_k, \mu_{\sigma(k)}) = (\lambda_i + \lambda_j) W_1(\mu_i, \mu_j). \quad (2.45)$$

This proves the pairwise estimate. \square

Proof of Corollary 1.8. By Proposition 1.1 and [1, Theorem 5.1] (see also [25, Theorem 5.1]), the unique Wasserstein barycenter $\mu_{\mathbb{P}} \in \mathcal{P}(\Omega)$ satisfies the L^∞ estimate

$$\|\mu_{\mathbb{P}}\|_{L^\infty} \leq \lambda_1^{-n} \|\rho\|_{L^\infty}. \quad (2.46)$$

Moreover, by [24, Proposition 5.1], we have

$$T_i = T_{\mu_{\mathbb{P}}}^i \circ B, \quad T_j = T_{\mu_{\mathbb{P}}}^j \circ B,$$

where $T_{\mu_{\mathbb{P}}}^i$ and $T_{\mu_{\mathbb{P}}}^j$ are the unique optimal transport maps from $\mu_{\mathbb{P}}$ to μ_i and from $\mu_{\mathbb{P}}$ to μ_j , respectively. Then Mériçot's two-marginal stability estimate yields

$$\|T_i - T_j\|_{L^2(\rho)} = \|T_{\mu_{\mathbb{P}}}^i - T_{\mu_{\mathbb{P}}}^j\|_{L^2(\mu_{\mathbb{P}})} \leq C W_1^{1/4}(\mu_i, \mu_j), \quad (2.47)$$

where C depends on λ_1 , $\text{diam}(S \cup Y)$, n and $\|\rho\|_{L^\infty}$. Combining this with Theorem 1.7 gives the stated minimum bounds. \square

3 Optimality and Barycenter-Preserving Examples

3.1 Proof of Proposition 1.9

We prove the optimality of the exponents in Theorem 1.7 and Corollary 1.8 in this subsection.

Proof of Proposition 1.9. Optimality in (a): Fix $n \geq 1$ and $N \geq 3$. Choose $S = [0, 1]^n \subseteq \mathbb{R}^n$ and $\lambda_1 = \lambda_2 = \dots = \lambda_N = \frac{1}{N}$. For any $0 < \epsilon < 1$, define

$$\rho = \mathcal{L}^n|_S \ll \mathcal{L}^n, \quad \mu_2^\epsilon = (1 - \epsilon)\delta_0 + \epsilon\delta_{e_1}, \quad \mu_3 = \dots = \mu_N = \delta_0,$$

where $e_1 = (1, 0, \dots, 0) \in \mathbb{R}^n$. By [17] (see also [24]), there exists a unique tuple of multi-marginal optimal transport maps $(T_2^\epsilon, T_3^\epsilon, \dots, T_N^\epsilon)$ satisfying

$$T_2^\epsilon(x_1) = e_1 \chi_{A_\epsilon}(x_1), \quad T_3^\epsilon(x_1) = \dots = T_N^\epsilon(x_1) = 0, \quad \rho\text{-a.e. } x_1 \in [0, 1]^n,$$

where $A_\epsilon := \{x_1 \in [0, 1]^n : T_2^\epsilon(x_1) = e_1\}$ and $\rho(A_\epsilon) = \epsilon$. Note that by (1.4),

$$\begin{aligned} & \int c(x_1, T_2^\epsilon(x_1), \dots, T_N^\epsilon(x_1)) d\rho(x_1) \\ &= \frac{1}{2} \int \sum_{i=3}^N \lambda_1 \lambda_i |x_1|^2 + \lambda_1 \lambda_2 |x_1 - T_2^\epsilon(x_1)|^2 + \sum_{i=3}^N \lambda_2 \lambda_i |T_2^\epsilon(x_1)|^2 d\rho(x_1) \quad (3.1) \\ &= \frac{(N-1)n}{6N^2} + \frac{N-1}{2N^2} \epsilon - \frac{1}{N^2} \int_{A_\epsilon} \langle x_1, e_1 \rangle d\mathcal{L}^n(x_1). \end{aligned}$$

Hence minimizing the cost is equivalent to maximizing $\int_{A_\epsilon} \langle x_1, e_1 \rangle d\mathcal{L}^n(x_1)$ over all measurable sets $A_\epsilon \subseteq [0, 1]^n$ with $\mathcal{L}^n(A_\epsilon) = \epsilon$. This maximum is attained, up to null sets, by $A_\epsilon = [1 - \epsilon, 1] \times [0, 1]^{n-1}$.

Thus, we have

$$\|T_2^\epsilon - T_3^\epsilon\|_{L^2(\rho)} = \epsilon^{1/2}, \quad W_1(\mu_2^\epsilon, \mu_3) = \epsilon,$$

which proves (a).

Optimality in (b), (c) and (d): This construction is inspired by [27, Theorem 1.2]. We construct infinitely many well-separated cells accumulating at the origin. The argument is divided into six steps.

Step 1: Fix $n \geq 2$ and $N \geq 3$. Denote $\lambda_k := 2^{-k^2}$. After discarding finitely many indices and relabeling, we may assume $0 < \lambda_k \leq \frac{1}{2(N-1)}$ for $k \in \mathbb{N}$. Denote

$$\lambda_1^k := 1 - (N-1)\lambda_k \geq \frac{1}{2}, \quad \lambda_2^k = \dots = \lambda_N^k := \lambda_k.$$

Denote $a_k := \frac{\lambda_k}{\lambda_1^k} \leq 1$ and $K_{n-2} := [-\frac{1}{2}, \frac{1}{2}]^{n-2}$ with the convention that K_0 is a single point. Define two subsets in \mathbb{R}^n by

$$V_+^{a_k} := [3, 5] \times [-a_k, a_k] \times K_{n-2}, \quad V_-^{a_k} := [-5, -3] \times [-a_k, a_k] \times K_{n-2}.$$

Denote $p_k := 2^{-k}e_1$ and $s_k := \eta 2^{-k}$, where $\eta \leq 1$ depends on n and N determined in (3.4) below. Define the k -cell Z_k by

$$Z_k := p_k + s_k(V_+^{a_k} \cup V_-^{a_k}), \quad k \in \mathbb{N},$$

and define

$$V_{k,+} := p_k + s_k V_+^{a_k}, \quad \text{and} \quad V_{k,-} := p_k + s_k V_-^{a_k}.$$

Define the source set S and the source measure ρ by

$$S := \overline{\bigcup_{k=1}^{\infty} Z_k} = \bigcup_{k=1}^{\infty} Z_k \cup \{0\} \subseteq B(0, R), \quad \rho := \frac{\mathcal{L}^n|_S}{\mathcal{L}^n(S)} \in \mathcal{P}_{\text{ac}}(S),$$

where R depends only on n . By the estimate (3.5) below, the sets Z_k are pairwise disjoint, and hence

$$\mathcal{L}^n(S) = \sum_{k=1}^{\infty} \mathcal{L}^n(Z_k) = \sum_{k=1}^{\infty} 8a_k s_k^n < +\infty. \quad (3.2)$$

Thus $\rho \in \mathcal{P}_{\text{ac}}(S) \cap L^\infty(S)$.

For $k \in \mathbb{N}$, define five points in \mathbb{R}^n by

$$\begin{aligned} A_k &:= p_k + s_k(a_k, 1, 0, \dots, 0), & B_k &:= p_k + s_k(-a_k, -1, 0, \dots, 0), \\ C_k &:= p_k + s_k(-a_k, 1, 0, \dots, 0), & D_k &:= p_k + s_k(a_k, -1, 0, \dots, 0), \end{aligned}$$

and $E_k := p_k + s_k(0, 0, 0, \dots, 0)$. Define the target set by

$$Y_k := \{A_k, B_k, C_k, D_k, E_k\}, \quad k \in \mathbb{N}, \quad Y := \overline{\bigcup_{k=1}^{\infty} Y_k} = \bigcup_{k=1}^{\infty} Y_k \cup \{0\} \subseteq B(0, R).$$

For $k \in \mathbb{N}$, denote

$$v_k := \rho(V_{k,+}) = \rho(V_{k,-}) = \frac{4a_k s_k^n}{\mathcal{L}^n(S)}.$$

Define the target measures $(\mu_2^k, \dots, \mu_N^k)$ by

$$\mu_2^k := v_k \delta_{A_k} + v_k \delta_{B_k} + \sum_{l \neq k} \rho(Z_l) \delta_{E_l}, \quad \mu_3^k := v_k \delta_{D_k} + v_k \delta_{C_k} + \sum_{l \neq k} \rho(Z_l) \delta_{E_l},$$

and

$$\mu_4^k = \dots = \mu_N^k := \sum_{l=1}^{\infty} \rho(Z_l) \delta_{E_l}.$$

By [17] (see also [24]), there exists a unique tuple of multi-marginal optimal transport maps $(T_2^k, T_3^k, \dots, T_N^k)$. We claim that they are given by

$$T_2^k(x_1) = \begin{cases} A_k, & x_1 \in V_{k,+}, \\ B_k, & x_1 \in V_{k,-}, \\ E_l, & x_1 \in Z_l, \quad l \neq k, \end{cases} \quad T_3^k(x_1) = \begin{cases} D_k, & x_1 \in V_{k,+}, \\ C_k, & x_1 \in V_{k,-}, \\ E_l, & x_1 \in Z_l, \quad l \neq k, \end{cases}$$

and

$$T_4^k(x_1) = \dots = T_N^k(x_1) = E_l, \quad x_1 \in Z_l, \quad \rho\text{-a.e. } x_1 \in S.$$

It is enough to show that these maps are pointwise optimal among all admissible target tuples.

Step 2: Let $k \in \mathbb{N}$ and $x_1 \in Z_k$. We claim that any admissible target tuple containing a point from Y_l , $l \neq k$, has strictly larger cost than any admissible target tuple contained

in Y_k^{N-1} . Indeed, by the construction in Step 1, we know $Z_k, Y_k \subseteq B(p_k, R s_k)$. Hence, for any $x_2, \dots, x_N \in Y_k$, by (1.4), we have

$$c(x_1, x_2, \dots, x_N) \leq 2R^2 \sum_{1 \leq i < j \leq N} \lambda_i^k \lambda_j^k s_k^2 \leq C_4 R^2 \lambda_k s_k^2, \quad (3.3)$$

where C_4 depends only on N . We choose

$$\eta = \min \left\{ 1, \frac{1}{8(1 + \sqrt{2C_4})R} \right\} > 0. \quad (3.4)$$

On the other hand, let (y_2, \dots, y_N) be any target tuple with $y_i \in Y_l$ for $l \neq k$. Since $Z_k, Y_k \subseteq B(p_k, R s_k)$ and $Z_l, Y_l \subseteq B(p_l, R s_l)$, we have

$$\text{dist}(Z_k \cup Y_k, Z_l \cup Y_l) \geq |p_k - p_l| - R(s_k + s_l).$$

We now check the last term. If $l > k$, then

$$|p_k - p_l| \geq 2^{-k-1}, \quad s_k + s_l \leq \frac{3}{2} \eta 2^{-k}.$$

Hence, by (3.4),

$$|p_k - p_l| - R(s_k + s_l) \geq \left(\frac{1}{2} - \frac{3}{2} R \eta \right) 2^{-k} \geq 2\sqrt{2C_4} R \eta 2^{-k} = 2\sqrt{2C_4} R s_k.$$

If $l < k$, write $m = k - l \geq 1$. Then

$$|p_k - p_l| = (2^m - 1)2^{-k}, \quad s_k + s_l = (1 + 2^m)\eta 2^{-k}.$$

Again by (3.4),

$$|p_k - p_l| - R(s_k + s_l) \geq (1 - 3R\eta) 2^{-k} \geq 2\sqrt{2C_4} R \eta 2^{-k} = 2\sqrt{2C_4} R s_k.$$

Therefore

$$\text{dist}(Z_k \cup Y_k, Z_l \cup Y_l) \geq |p_k - p_l| - R(s_k + s_l) \geq 2\sqrt{2C_4} R s_k. \quad (3.5)$$

Then by (1.4), we have

$$c(x_1, y_2, \dots, y_N) \geq \frac{\lambda_1^k \lambda_k}{2} |x_1 - y_i|^2 \geq \frac{\lambda_k}{4} \text{dist}^2(Z_k, Y_l) \geq 2C_4 R^2 \lambda_k s_k^2. \quad (3.6)$$

Combining (3.3) and (3.6) proves the claim. Hence, in the pointwise comparison on Z_k , it suffices to restrict to admissible target tuples contained in Y_k^{N-1} . We only need to consider tuples of the form:

$$(x_2, x_3, E_k, \dots, E_k), \quad x_2 \in \{A_k, B_k\}, \quad x_3 \in \{C_k, D_k\}.$$

Step 3: We now compare these finitely many target tuples. First take $x_1 \in V_{k,+}$. We claim that the unique pointwise minimizer is

$$(A_k, D_k, E_k, \dots, E_k).$$

Note that by (1.4), it holds that

$$\begin{aligned} & c(x_1, x_2, x_3, E_k, \dots, E_k) \\ &= \frac{1}{2} \lambda_1^k \lambda_k |x_1 - x_2|^2 + \frac{1}{2} \lambda_1^k \lambda_k |x_1 - x_3|^2 + \frac{1}{2} (\lambda_k)^2 |x_2 - x_3|^2 \\ & \quad + \frac{1}{2} (1 - \lambda_1^k - 2\lambda_k) \left(\lambda_1^k |x_1 - E_k|^2 + \lambda_k |x_2 - E_k|^2 + \lambda_k |x_3 - E_k|^2 \right). \end{aligned} \quad (3.7)$$

Denote $x_1 := p_k + s_k(u, v, z) \in V_{k,+}$ as the coordinates $(u, v, z) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^{n-2}$. Since

$$|A_k - E_k| = |B_k - E_k| = |C_k - E_k| = |D_k - E_k|, \quad (3.8)$$

we have

$$c(x_1, A_k, C_k, \dots, E_k) - c(x_1, A_k, D_k, \dots, E_k) = 2\lambda_k s_k^2 \left(\lambda_1^k (ua_k - v) - \lambda_k (1 - a_k^2) \right) > 0, \quad (3.9)$$

where the inequality follows from

$$\lambda_1^k (ua_k - v) \geq \lambda_1^k (3a_k - a_k) = 2\lambda_k.$$

Similarly, it holds that

$$c(x_1, B_k, D_k, \dots, E_k) - c(x_1, A_k, D_k, \dots, E_k) = 2\lambda_k s_k^2 \left(\lambda_1^k (ua_k + v) - \lambda_k (1 - a_k^2) \right) > 0, \quad (3.10)$$

and

$$c(x_1, B_k, C_k, \dots, E_k) - c(x_1, A_k, D_k, \dots, E_k) = 4\lambda_1^k \lambda_k ua_k s_k^2 > 0. \quad (3.11)$$

Thus, $(A_k, D_k, E_k, \dots, E_k)$ is the unique pointwise minimizer on $V_{k,+}$. By symmetry, $(B_k, C_k, E_k, \dots, E_k)$ is the unique pointwise minimizer on $V_{k,-}$.

Step 4: Let $l \neq k$ and $x_1 \in Z_l$. By the same cell-localization argument as in Step 2, in the pointwise comparison on Z_l , it suffices to restrict to admissible tuples contained in Y_l^{N-1} . For this cell, the only atoms of the target marginals μ_2^k, \dots, μ_N^k in Y_l are all equal to E_l . Thus the only localized admissible tuple is (E_l, E_l, \dots, E_l) , and it is the unique pointwise minimizer on Z_l .

Combining this with the preceding results, the maps given in Step 1 are pointwise optimal among all admissible target tuples for ρ -a.e. $x_1 \in S$. Therefore, these maps are the multi-marginal optimal transport maps associated with $(\rho, \mu_2^k, \dots, \mu_N^k)$.

Step 5: We show that

$$W_1(\mu_2^k, \mu_3^k) = \frac{16a_k^2 s_k^{n+1}}{\mathcal{L}^n(S)}.$$

On the one hand, by matching A_k to C_k and B_k to D_k , we have

$$W_1(\mu_2^k, \mu_3^k) \leq v_k |A_k - C_k| + v_k |B_k - D_k| = \frac{16a_k^2 s_k^{n+1}}{\mathcal{L}^n(S)}. \quad (3.12)$$

On the other hand, define a function f by

$$f(A_k) = f(B_k) := a_k s_k, \quad f(C_k) = f(D_k) := -a_k s_k, \quad f(E_l) := 0, \quad l \in \mathbb{N}.$$

Combining with (3.5), f is 1-Lipschitz on $\{A_k, B_k, C_k, D_k\} \cup \{E_l : l \in \mathbb{N}\}$. By McShane's lemma, f can be extended to \mathbb{R}^n with $\text{Lip}(f) \leq 1$. Thus, by Kantorovich–Rubinstein duality, we have

$$W_1(\mu_2^k, \mu_3^k) \geq \int f d(\mu_2^k - \mu_3^k) = \frac{16a_k^2 s_k^{n+1}}{\mathcal{L}^n(S)}. \quad (3.13)$$

Therefore,

$$W_1(\mu_2^k, \mu_3^k) = \frac{16a_k^2 s_k^{n+1}}{\mathcal{L}^n(S)}, \quad \|T_2^k - T_3^k\|_{L^2(\rho)} = \frac{4\sqrt{2}a_k^{1/2} s_k^{(n+2)/2}}{(\mathcal{L}^n(S))^{1/2}}. \quad (3.14)$$

Step 6: Recall that $a_k = \frac{\lambda_k}{\lambda_1^k} \asymp \lambda_k = 2^{-k^2}$ and $s_k = \eta 2^{-k} \asymp 2^{-k}$ as $k \rightarrow \infty$. Moreover, since $\lambda_1^k = 1 - (N-1)\lambda_k \in [\frac{1}{2}, 1)$, the λ_1 -dependent constant appearing in the proof of Corollary 1.8 remains uniformly bounded along this sequence.

If the exponent $\alpha > \frac{1}{4}$, then

$$\frac{\|T_2^k - T_3^k\|_{L^2(\rho)}}{W_1^\alpha(\mu_2^k, \mu_3^k)} \asymp \lambda_k^{1/2-2\alpha} s_k^{(n+2)/2-(n+1)\alpha} \rightarrow +\infty, \quad (3.15)$$

which proves **(b)**.

Recall that $\lambda_2^k + \lambda_3^k = 2\lambda_k \asymp 2^{-k^2}$ as $k \rightarrow \infty$. If the exponent in $W_1(\mu_2^k, \mu_3^k)$ is $\frac{1}{2}$, then for the exponent $\beta > -\frac{1}{2}$,

$$\frac{\|T_2^k - T_3^k\|_{L^2(\rho)}}{(\lambda_2^k + \lambda_3^k)^\beta W_1^{1/2}(\mu_2^k, \mu_3^k)} \asymp \lambda_k^{-1/2-\beta} s_k^{1/2} \rightarrow +\infty, \quad (3.16)$$

which proves **(c)**.

If the exponent in $W_1(\mu_2^k, \mu_3^k)$ is $\theta \in [\frac{1}{4}, \frac{1}{2}]$, then for the exponent $\gamma > \frac{1}{2} - 2\theta$,

$$\frac{\|T_2^k - T_3^k\|_{L^2(\rho)}}{(\lambda_2^k + \lambda_3^k)^\gamma W_1^\theta(\mu_2^k, \mu_3^k)} \asymp \lambda_k^{1/2-2\theta-\gamma} s_k^{(n+2)/2-(n+1)\theta} \rightarrow +\infty, \quad (3.17)$$

which proves **(d)**. □

3.2 Barycenter-preserving perturbations

We present two non-permutation examples that preserve the barycenter map. The construction relies on the following standard dual characterization of Wasserstein barycenters (cf. [1]); see also [5] for a related averaging-map characterization.

Proposition 3.1. *Let $\mathbb{P} = \lambda_1 \delta_\rho + \sum_{i=2}^N \lambda_i \delta_{\mu_i} \in \mathcal{P}(\mathcal{P}(\Omega))$. Then the following statements are equivalent:*

- (a) $\nu \in \mathcal{P}(\Omega)$ is the Wasserstein barycenter of \mathbb{P} ;
- (b) There exist Kantorovich potentials Ψ_i from ν to μ_i , for $i = 1, \dots, N$ with $\mu_1 = \rho$, after choosing suitable additive constants, such that

$$\sum_{i=1}^N \lambda_i \Psi_i(x) \geq 0, \quad \forall x \in \Omega, \quad \sum_{i=1}^N \lambda_i \Psi_i(x) = 0, \quad \nu\text{-a.e. } x \in \Omega.$$

In the following two examples, fix $\mathbb{P} = \lambda_1 \delta_\rho + \sum_{i=2}^N \lambda_i \delta_{\mu_i} \in \mathcal{P}(\mathcal{P}(\Omega))$ and assume $\rho \in \mathcal{P}_{\text{ac}}(\Omega)$. By Proposition 1.1, there exists a unique Wasserstein barycenter $\mu_{\mathbb{P}}$ of \mathbb{P} with $\mu_{\mathbb{P}} \ll \mathcal{L}^n$. For $i = 1, \dots, N$, let $T_{\mu_{\mathbb{P}}}^i = \text{Id} - \nabla \Psi_{\mu_{\mathbb{P}}}^i$ be the unique optimal transport map from $\mu_{\mathbb{P}}$ to μ_i with $\mu_1 = \rho$, where $\Psi_{\mu_{\mathbb{P}}}^i$ is a Kantorovich potential from $\mu_{\mathbb{P}}$ to μ_i . We choose the additive constants of $\Psi_{\mu_{\mathbb{P}}}^i$ such that Proposition 3.1 **(b)** holds.

Example 3.2 (Balanced translations). Let $a_1 = 0 \in \mathbb{R}^n$, and let $a_2, \dots, a_N \in \mathbb{R}^n$ be such that $\sum_{i=2}^N \lambda_i a_i = 0$. Define

$$\tilde{\Psi}_{\mu_{\mathbb{P}}}^i(x) := \Psi_{\mu_{\mathbb{P}}}^i(x) - \langle a_i, x \rangle, \quad i = 1, \dots, N,$$

and

$$\tilde{T}_{\mu_{\mathbb{P}}}^i := \text{Id} - \nabla \tilde{\Psi}_{\mu_{\mathbb{P}}}^i = T_{\mu_{\mathbb{P}}}^i + a_i, \quad \tilde{\mu}_i := (\tilde{T}_{\mu_{\mathbb{P}}}^i)_{\#} \mu_{\mathbb{P}}.$$

Assume the translations are chosen such that $\text{supp } \tilde{\mu}_i \subseteq \Omega$ for $i = 1, \dots, N$. Consider the perturbed tuple $(\tilde{\mu}_1, \tilde{\mu}_2, \dots, \tilde{\mu}_N)$. Since

$$\frac{1}{2}|x|^2 - \tilde{\Psi}_{\mu_{\mathbb{P}}}^i(x) = \left(\frac{1}{2}|x|^2 - \Psi_{\mu_{\mathbb{P}}}^i(x) \right) + \langle a_i, x \rangle, \quad i = 1, \dots, N \quad (3.18)$$

are convex, the maps $\tilde{T}_{\mu_{\mathbb{P}}}^i$ are the unique optimal transports from $\mu_{\mathbb{P}}$ to $\tilde{\mu}_i$ with $\tilde{\mu}_1 = \rho$. Moreover,

$$\sum_{i=1}^N \lambda_i \tilde{\Psi}_{\mu_{\mathbb{P}}}^i(x) = \sum_{i=1}^N \lambda_i \Psi_{\mu_{\mathbb{P}}}^i(x) - \left\langle \sum_{i=1}^N \lambda_i a_i, x \right\rangle = \sum_{i=1}^N \lambda_i \Psi_{\mu_{\mathbb{P}}}^i(x), \quad (3.19)$$

by Proposition 3.1, $\tilde{\mathbb{P}} = \lambda_1 \delta_{\rho} + \sum_{i=2}^N \lambda_i \delta_{\tilde{\mu}_i} \in \mathcal{P}(\mathcal{P}(\Omega))$ has the unique Wasserstein barycenter $\mu_{\mathbb{P}}$. Therefore, the perturbed tuple $(\rho, \tilde{\mu}_2, \dots, \tilde{\mu}_N)$ preserves the barycenter map.

Example 3.3 (Balanced nonlinear perturbations). Let $k, l \in \{2, \dots, N\}$ with $k \neq l$. Let $h \in C_c^2(\mathbb{R}^n)$, and let $\epsilon > 0$ be such that

$$\frac{1}{2}|x|^2 - \Psi_{\mu_{\mathbb{P}}}^k(x) + \epsilon h(x), \quad \frac{1}{2}|x|^2 - \Psi_{\mu_{\mathbb{P}}}^l(x) - \epsilon \frac{\lambda_k}{\lambda_l} h(x) \quad (3.20)$$

remain convex and the perturbed measures $\tilde{\mu}_i$ defined below satisfy $\text{supp } \tilde{\mu}_i \subseteq \Omega$ for $i = 1, \dots, N$. This is an admissibility assumption on the pair (h, ϵ) . Define

$$\begin{aligned} \tilde{\Psi}_{\mu_{\mathbb{P}}}^k(x) &:= \Psi_{\mu_{\mathbb{P}}}^k(x) - \epsilon h(x), & \tilde{\Psi}_{\mu_{\mathbb{P}}}^l(x) &:= \Psi_{\mu_{\mathbb{P}}}^l(x) + \epsilon \frac{\lambda_k}{\lambda_l} h(x), \\ \tilde{\Psi}_{\mu_{\mathbb{P}}}^i(x) &:= \Psi_{\mu_{\mathbb{P}}}^i(x), & i &\neq k, l, \end{aligned}$$

and

$$\tilde{T}_{\mu_{\mathbb{P}}}^k := \text{Id} - \nabla \tilde{\Psi}_{\mu_{\mathbb{P}}}^k = T_{\mu_{\mathbb{P}}}^k + \epsilon \nabla h, \quad \tilde{T}_{\mu_{\mathbb{P}}}^l := \text{Id} - \nabla \tilde{\Psi}_{\mu_{\mathbb{P}}}^l = T_{\mu_{\mathbb{P}}}^l - \epsilon \frac{\lambda_k}{\lambda_l} \nabla h,$$

$$\tilde{T}_{\mu_{\mathbb{P}}}^i := T_{\mu_{\mathbb{P}}}^i, \quad i \neq k, l.$$

Define

$$\tilde{\mu}_i := (\tilde{T}_{\mu_{\mathbb{P}}}^i)_{\#} \mu_{\mathbb{P}}, \quad i = 1, \dots, N.$$

Consider the perturbed tuple $(\tilde{\mu}_1, \tilde{\mu}_2, \dots, \tilde{\mu}_N)$. By the admissibility assumption,

$$\frac{1}{2}|x|^2 - \tilde{\Psi}_{\mu_{\mathbb{P}}}^i(x), \quad i = 1, \dots, N$$

are convex, and the maps $\tilde{T}_{\mu_{\mathbb{P}}}^i$ are the unique optimal transport maps from $\mu_{\mathbb{P}}$ to $\tilde{\mu}_i$ with $\tilde{\mu}_1 = \rho$. Moreover,

$$\sum_{i=1}^N \lambda_i \tilde{\Psi}_{\mu_{\mathbb{P}}}^i(x) = \sum_{i=1}^N \lambda_i \Psi_{\mu_{\mathbb{P}}}^i(x) - \lambda_k \epsilon h(x) + \lambda_l \epsilon \frac{\lambda_k}{\lambda_l} h(x) = \sum_{i=1}^N \lambda_i \Psi_{\mu_{\mathbb{P}}}^i(x). \quad (3.21)$$

By Proposition 3.1, the perturbed tuple $(\rho, \tilde{\mu}_2, \dots, \tilde{\mu}_N)$ preserves the barycenter map.

Remark 3.4. The nonlinear example relies on the admissibility assumption. This condition is needed since the perturbation must preserve the convexity of the functions defined in (3.20). For the diagonal tuple $\mu_2 = \dots = \mu_N = \rho$, this condition holds for any sufficiently small balanced perturbation. Therefore the two examples above use the same idea: the linear balance condition on the Kantorovich potentials preserves the barycenter, while the convexity condition determines the admissible directions.

4 The Scope of the Two-Mode Mechanism

The proofs in Sections 2 use two logically distinct inputs. The first is *relative coercivity*: after a collective coordinate has been fixed, the Kantorovich defect controls the remaining relative variables quadratically. The second is *external-mode stability*: the collective map must admit a variational representation for which a quantitative stability estimate is available. The square-root term in Theorem 1.2 comes from the first input, while the loss to the exponent $\frac{1}{4}$ in Theorem 1.3 comes from the second.

This section tests these two inputs beyond the basic barycentric quadratic cost. Subsection 4.1 proves that perturbations depending only on the collective coordinate preserve the internal coercivity estimate and reduce the external mode to a penalized Wasserstein barycenter. Subsection 4.2 applies this result to uniformly concave costs of the sum. The remaining subsections describe the structural boundary of the argument: cyclic symmetrization explains the complete-graph model, pairwise quadratic interactions lead to a normal-mode decomposition, and hedonic, Coulomb and determinant costs exhibit distinct failures of the two required inputs.

The first two subsections contain rigorous stability results. The remaining subsections are structural analyses and do not assert a general stability theorem.

4.1 Collective-coordinate perturbations and persistence of the two-mode estimate

Let $\Omega = \overline{\text{conv}(S \cup Y)}$. For $(x_1, \dots, x_N) \in S \times Y^{N-1}$, define the linear collective coordinate

$$Z_x := \sum_{i=1}^N \lambda_i x_i.$$

For the quadratic cost (1.1), this coordinate is exactly the pointwise weighted barycenter. We then define

$$M_x := \frac{Z_x - \lambda_1 x_1}{1 - \lambda_1}, \quad V_{i,x} := x_i - M_x, \quad i = 2, \dots, N.$$

Let c_0 denote the weighted quadratic cost in (1.1). Consider the cost

$$c_F(x_1, \dots, x_N) := c_0(x_1, \dots, x_N) + F(Z_x), \quad (4.1)$$

where $F \in C^2(U)$ for some open neighbourhood $U \supseteq \Omega$. Then

$$c_F(x_1, \dots, x_N) = \frac{\lambda_1}{2(1 - \lambda_1)} |Z_x - x_1|^2 + \frac{1}{2} \sum_{i=2}^N \lambda_i |V_{i,x}|^2 + F(Z_x). \quad (4.2)$$

Assume that there exist constants $\kappa_F \in (0, 1]$ and $\tau_F \geq 0$ such that, as quadratic forms on U ,

$$-\tau_F I \leq D^2 F \leq (1 - \kappa_F) I. \quad (4.3)$$

The Kim–Pass twist condition follows from the monotonicity of the transformed collective coordinate. Indeed, the map

$$H(z) := z - \nabla F(z)$$

satisfies

$$\langle H(z) - H(z'), z - z' \rangle \geq \kappa_F |z - z'|^2, \quad z, z' \in U.$$

Thus H is injective. Moreover,

$$D_{x_1} c_F(x_1, \dots, x_N) = \lambda_1(x_1 - H(Z_x)).$$

Hence, for fixed x_1 , the first-order condition determines $H(Z_x)$, and therefore determines the collective coordinate Z_x . This verifies the twist on splitting sets for c_F (cf. [23]). By Kim and Pass [23, Theorem 3.1], for $\rho \in \mathcal{P}_{\text{ac}}(S)$, there exists a unique tuple of multi-marginal optimal transport maps (T_2^F, \dots, T_N^F) associated with $(\rho, \mu_2, \dots, \mu_N)$.

Define the collective map, the associated target mean and the internal directions by

$$Z^F(x_1) := \lambda_1 x_1 + \sum_{i=2}^N \lambda_i T_i^F(x_1),$$

and

$$M^F(x_1) := \frac{Z^F(x_1) - \lambda_1 x_1}{1 - \lambda_1}, \quad V_i^F(x_1) := T_i^F(x_1) - M^F(x_1), \quad i = 2, \dots, N.$$

Let $(\varphi_1^F, \dots, \varphi_N^F)$ be a c_F -conjugate tuple of Kantorovich potentials for the multi-marginal optimal transport problem $(\rho, \mu_2, \dots, \mu_N)$. For $x_1 \in S$, $x_2, \dots, x_N \in Y$, define

$$R_F(x_1, \dots, x_N) := c_F(x_1, \dots, x_N) - \sum_{i=1}^N \varphi_i^F(x_i) \geq 0,$$

and

$$R_F(x_1, T_2^F(x_1), \dots, T_N^F(x_1)) = 0, \quad \rho\text{-a.e. } x_1 \in S.$$

Proposition 4.1. *Let $\rho \in \mathcal{P}_{\text{ac}}(S)$ and let c_F be defined by (4.1). For ρ -a.e. $x_1 \in S$ and every $x_2, \dots, x_N \in Y$, we have*

$$R_F(x_1, \dots, x_N) \geq \frac{\kappa_F}{4} \sum_{i=2}^N \lambda_i^2 |V_{i,x} - V_i^F(x_1)|^2 - \frac{1 + \tau_F + 2\kappa_F}{2} |Z_x - Z^F(x_1)|^2. \quad (4.4)$$

Moreover, let $(\tilde{\mu}_2, \dots, \tilde{\mu}_N) \subseteq \mathcal{P}(Y)$ be another tuple of target marginals, and let $(\tilde{T}_2^F, \dots, \tilde{T}_N^F)$ be the corresponding multi-marginal optimal transport maps. Define

$$\tilde{Z}^F(x_1) := \lambda_1 x_1 + \sum_{i=2}^N \lambda_i \tilde{T}_i^F(x_1).$$

Then

$$\sum_{i=2}^N \lambda_i \|T_i^F - \tilde{T}_i^F\|_{L^2(\rho)} \leq C \left[\left(\sum_{i=2}^N \lambda_i W_1(\mu_i, \tilde{\mu}_i) \right)^{1/2} + \|Z^F - \tilde{Z}^F\|_{L^2(\rho)} \right], \quad (4.5)$$

where C depends on N , $\text{diam}(S \cup Y)$, κ_F and τ_F .

Proof. Fix $x_1 \in S$ such that

$$R_F(x_1, T_2^F(x_1), \dots, T_N^F(x_1)) = 0. \quad (4.6)$$

For $i = 2, \dots, N$, denote

$$\Delta x_i := x_i - T_i^F(x_1), \quad \Delta Z := Z_x - Z^F(x_1) = \sum_{i=2}^N \lambda_i \Delta x_i, \quad \Delta V_i := V_{i,x} - V_i^F(x_1).$$

Then

$$\Delta V_i = \Delta x_i - \frac{\Delta Z}{1 - \lambda_1}, \quad i = 2, \dots, N, \quad \text{and} \quad \sum_{i=2}^N \lambda_i \Delta V_i = 0. \quad (4.7)$$

For $h \in \mathbb{R}^n$, define the Taylor remainder

$$\mathcal{R}_F(h) := F(Z^F(x_1) + h) - F(Z^F(x_1)) - \langle \nabla F(Z^F(x_1)), h \rangle.$$

By Taylor's formula and (4.3), we have

$$\mathcal{R}_F(\Delta Z) \geq -\frac{\tau_F}{2} |\Delta Z|^2, \quad \mathcal{R}_F(\lambda_i \Delta x_i) \leq \frac{1 - \kappa_F}{2} \lambda_i^2 |\Delta x_i|^2, \quad i = 2, \dots, N. \quad (4.8)$$

Therefore, by an argument similar to that in Proposition 2.3, we have

$$\begin{aligned} & R_F(x_1, \dots, x_N) \\ \stackrel{(4.6)}{=} & R_F(x_1, \dots, x_N) - R_F(x_1, T_2^F(x_1), \dots, T_N^F(x_1)) \\ \geq & \frac{\lambda_1}{2(1 - \lambda_1)} |\Delta Z|^2 + \frac{1}{2} \sum_{i=2}^N \lambda_i |\Delta V_i|^2 - \sum_{i=2}^N \frac{\lambda_i(1 - \lambda_i)}{2} |\Delta x_i|^2 + \mathcal{R}_F(\Delta Z) - \sum_{i=2}^N \mathcal{R}_F(\lambda_i \Delta x_i) \\ \stackrel{(4.8)}{\geq} & \frac{\lambda_1}{2(1 - \lambda_1)} |\Delta Z|^2 + \frac{1}{2} \sum_{i=2}^N \lambda_i |\Delta V_i|^2 - \sum_{i=2}^N \frac{\lambda_i(1 - \kappa_F \lambda_i)}{2} |\Delta x_i|^2 - \frac{\tau_F}{2} |\Delta Z|^2 \\ \stackrel{(4.7)}{=} & \frac{\kappa_F}{2} \sum_{i=2}^N \lambda_i^2 |\Delta V_i|^2 + \frac{\kappa_F}{1 - \lambda_1} \left\langle \sum_{i=2}^N \lambda_i^2 \Delta V_i, \Delta Z \right\rangle + \left(\frac{\kappa_F}{2(1 - \lambda_1)^2} \sum_{i=2}^N \lambda_i^2 - \frac{1 + \tau_F}{2} \right) |\Delta Z|^2 \\ \geq & \frac{\kappa_F}{4} \sum_{i=2}^N \lambda_i^2 |\Delta V_i|^2 - \frac{1 + \tau_F + 2\kappa_F}{2} |\Delta Z|^2, \end{aligned} \quad (4.9)$$

where (*) uses the same c -conjugacy argument as in Proposition 2.3, and (**) follows from Hölder's inequality and Young's inequality.

We now derive the stability estimate. By Lemma 2.4, there exists a coupling $\bar{\gamma} \in \Pi(\rho, \mu_2, \dots, \mu_N, \tilde{\mu}_2, \dots, \tilde{\mu}_N)$, such that

$$(x_1, \tilde{x}_2, \dots, \tilde{x}_N)_{\#} \bar{\gamma} = (\text{Id}, \tilde{T}_2^F, \dots, \tilde{T}_N^F)_{\#} \rho, \quad (x_i, \tilde{x}_i)_{\#} \bar{\gamma} = \pi_i, \quad i = 2, \dots, N,$$

where $\pi_i \in \Pi(\mu_i, \tilde{\mu}_i)$ is a W_1 -optimal transport plan between μ_i and $\tilde{\mu}_i$. Recall $\delta = \sum_{i=2}^N \lambda_i W_1(\mu_i, \tilde{\mu}_i)$, $D = \text{diam}(S \cup Y)$. By Corollary 2.5, with B replaced by Z , we have

$$\left(\int |Z_x - \tilde{Z}^F(x_1)|^2 d\bar{\gamma} \right)^{1/2} \leq D^{1/2} \delta^{1/2}, \quad \left(\int \sum_{i=2}^N \lambda_i^2 |V_{i,x} - \tilde{V}_i^F(x_1)|^2 d\bar{\gamma} \right)^{1/2} \leq D^{1/2} \delta^{1/2}. \quad (4.10)$$

Moreover, fix $z_0 \in \Omega$ and define

$$\bar{F}(z) := F(z) - F(z_0) - \langle \nabla F(z_0), z - z_0 \rangle.$$

Since Ω is convex, $D^2 \bar{F} = D^2 F$ and $\nabla \bar{F}(z_0) = 0$. By (4.3), we have

$$\text{Lip}(\bar{F}|_{\Omega}) \leq D \max\{\tau_F, 1 - \kappa_F\}. \quad (4.11)$$

On the other hand, note that

$$c_0(x_1, \dots, x_N) + \bar{F}(Z_x) = c_F(x_1, \dots, x_N) - \sum_{i=1}^N \lambda_i \langle \nabla F(z_0), x_i \rangle + \langle \nabla F(z_0), z_0 \rangle - F(z_0), \quad (4.12)$$

which implies the multi-marginal optimal transport problem is unchanged. Moreover, if the Kantorovich potentials are shifted by the corresponding marginal affine functions, then the residual $c_F - \sum_{i=1}^N \varphi_i^F$ is unchanged as well. Hence the stability argument may be carried out with \bar{F} in place of F . Combining with (4.11), the cost c_F satisfies

$$|c_F(x_1, x_2, \dots, x_N) - c_F(x_1, \tilde{x}_2, \dots, \tilde{x}_N)| \leq C_5 \sum_{i=2}^N \lambda_i |x_i - \tilde{x}_i|, \quad (4.13)$$

where C_5 depends on D, τ_F, κ_F .

Using the analogue of Lemma 2.6, with (4.13) in place of (2.17), and repeating the proof of Theorem 1.2, we obtain the stability estimate (4.5). \square

The estimate in (4.5) leaves the collective term $\|Z^F - \tilde{Z}^F\|_{L^2(\rho)}$ to be controlled. For $F = 0$, the map Z^F is the Wasserstein barycenter map used in Theorem 1.3. For a general F , Z^F need not be an optimal transport map. The useful external mode is the transformed collective map $H^F := Z^F - \nabla F(Z^F)$, which can be identified with an optimal transport map associated with a penalized Wasserstein barycenter problem. The following proposition makes this reduction precise.

Proposition 4.2. *Let $\rho \in \mathcal{P}_{\text{ac}}(S) \cap L^\infty(S)$ and let c_F be defined by (4.1). With the notation of Proposition 4.1, there exists a constant $C > 0$, depending on $\lambda_1, N, \text{diam}(S \cup Y), \kappa_F, \tau_F, n$ and $\|\rho\|_{L^\infty}$, such that*

$$\sum_{i=2}^N \lambda_i \|T_i^F - \tilde{T}_i^F\|_{L^2(\rho)} \leq C \left(\sum_{i=2}^N \lambda_i W_1(\mu_i, \tilde{\mu}_i) \right)^{1/4}.$$

Proof. As in the proof of Proposition 4.1, we fix $z_0 \in \Omega$ and replace F by the normalized function

$$\bar{F}(z) := F(z) - F(z_0) - \langle \nabla F(z_0), z - z_0 \rangle.$$

This does not change the multi-marginal optimal transport maps. After relabeling \bar{F} as F , we may assume that $\nabla F(z_0) = 0$.

Let K be a compact convex set such that $\Omega \Subset K \Subset U$. Define

$$G(z) := \frac{1}{2}|z|^2 - F(z), \quad G_K^*(y) := \sup_{z \in K} \{\langle z, y \rangle - G(z)\}, \quad V(y) := G_K^*(y) - \frac{1}{2}|y|^2.$$

Here G_K^* is the restricted Legendre transform of G and by (4.3), $D^2G \geq \kappa_F I$.

Since G is convex and $Z_x = \sum_{i=1}^N \lambda_i x_i \in \Omega \Subset K$, the restricted transform satisfies

$$G(Z_x) = \sup_{y \in \mathbb{R}^n} \{\langle Z_x, y \rangle - G_K^*(y)\}. \quad (4.14)$$

Then we have

$$\begin{aligned} c_F(x_1, \dots, x_N) &= \sum_{i=1}^N \frac{\lambda_i}{2} |x_i|^2 - \sup_{y \in \mathbb{R}^n} \{\langle Z_x, y \rangle - G_K^*(y)\} \\ &= \inf_{y \in \mathbb{R}^n} \left\{ \sum_{i=1}^N \frac{\lambda_i}{2} |x_i - y|^2 + V(y) \right\}. \end{aligned} \quad (4.15)$$

Moreover, by Fenchel-Young equality, the unique minimizer is $\nabla G(Z_x)$.

Consequently, define

$$H^F := \nabla G(Z^F) = Z^F - \nabla F(Z^F), \quad \mu_{\mathbb{P}}^F := H_{\#}^F \rho.$$

Then Agueh–Carlier’s barycenter–multi-marginal correspondence [1] applies after the change of collective coordinate $Z^F \mapsto H^F = \nabla G(Z^F)$. It follows that $\mu_{\mathbb{P}}^F$ minimizes the penalized barycenter functional

$$\text{Var}_{\mathbb{P}}(\nu) + \int V \, d\nu, \quad \mathbb{P} = \lambda_1 \delta_{\rho} + \sum_{i=2}^N \lambda_i \delta_{\mu_i},$$

and H^F is the unique optimal transport map from ρ to $\mu_{\mathbb{P}}^F$, for the quadratic cost. Similarly, $\tilde{H}^F := \tilde{Z}^F - \nabla F(\tilde{Z}^F)$ is the unique optimal transport map from ρ to $\tilde{\mu}_{\mathbb{P}}^F := \tilde{H}_{\#}^F \rho$.

Note that the preceding affine normalization only translates both H^F and \tilde{H}^F by the same vector. Hence $H^F - \tilde{H}^F$ and $Z^F - \tilde{Z}^F$ are unchanged. Moreover, after this normalization, the measures $\mu_{\mathbb{P}}^F$ and $\tilde{\mu}_{\mathbb{P}}^F$ are supported in a common compact set Ω_F , whose diameter is controlled by $\text{diam}(S \cup Y)$ and τ_F .

We use the standard Kantorovich-duality/subdifferential formulation for penalized Wasserstein barycenters, as in Bigot–Cazelles–Papadakis [4, Theorem 2.11]. In the present notation, it gives Kantorovich potentials $\{\psi_{\mu_{\mathbb{P}}^F}^{F,i}\}_{i=1}^N$ from $\mu_{\mathbb{P}}^F$ to μ_i with $\mu_1 = \rho$, such that, after adding constants,

$$\sum_{i=1}^N \lambda_i \psi_{\mu_{\mathbb{P}}^F}^{F,i}(y) + V(y) \geq 0 \quad \text{on } \Omega_F, \quad \sum_{i=1}^N \lambda_i \psi_{\mu_{\mathbb{P}}^F}^{F,i} + V = 0 \quad \mu_{\mathbb{P}}^F\text{-a.e.} \quad (4.16)$$

Indeed, (4.16) is just the first-order variational inequality for the functional $\text{Var}_{\mathbb{P}}(\nu) + \int V \, d\nu$, written using Kantorovich potentials as subgradients of the maps $\nu \mapsto \frac{1}{2} W_2^2(\nu, \mu_i)$. Similarly, there exist Kantorovich potentials $\{\psi_{\tilde{\mu}_{\mathbb{P}}^F}^{F,i}\}_{i=1}^N$ satisfying the analogue of (4.16) for $\tilde{\mathbb{P}}$ and $\tilde{\mu}_{\mathbb{P}}^F$.

Thus, with notation analogous to (2.29) and (2.30), we have

$$\text{Var}_{\mathbb{P}}(\mu_{\mathbb{P}}^F) + \int V \, d\mu_{\mathbb{P}}^F - \text{Var}_{\mathbb{P}}(\mu_{\mathbb{P}}^F) - \int V \, d\mu_{\mathbb{P}}^F \geq \lambda_1 D_{\rho}(\mu_{\mathbb{P}}^F, \mu_{\mathbb{P}}^F), \quad (4.17)$$

and

$$\text{Var}_{\tilde{\mathbb{P}}}(\mu_{\mathbb{P}}^F) + \int V \, d\mu_{\mathbb{P}}^F - \text{Var}_{\tilde{\mathbb{P}}}(\mu_{\mathbb{P}}^F) - \int V \, d\mu_{\mathbb{P}}^F \geq \lambda_1 D_{\rho}(\mu_{\mathbb{P}}^F, \mu_{\mathbb{P}}^F), \quad (4.18)$$

which implies

$$\text{Var}_{\mathbb{P}}(\mu_{\mathbb{P}}^F) - \text{Var}_{\tilde{\mathbb{P}}}(\mu_{\mathbb{P}}^F) + \text{Var}_{\tilde{\mathbb{P}}}(\mu_{\mathbb{P}}^F) - \text{Var}_{\tilde{\mathbb{P}}}(\mu_{\mathbb{P}}^F) \geq \lambda_1 \left(D_{\rho}(\mu_{\mathbb{P}}^F, \mu_{\mathbb{P}}^F) + D_{\rho}(\mu_{\mathbb{P}}^F, \mu_{\mathbb{P}}^F) \right). \quad (4.19)$$

Here the two V -terms cancel when the two preceding variational inequalities are added.

Note that by (4.3), it holds that

$$\|Z^F - \tilde{Z}^F\|_{L^2(\rho)} \leq \kappa_F^{-1} \|H^F - \tilde{H}^F\|_{L^2(\rho)}. \quad (4.20)$$

Combining (4.19), (4.20) and Proposition 4.1 with Mériqot’s estimate [29, Theorem 2.2], and repeating the proof of Theorem 1.3, completes the proof. \square

Propositions 4.1 and 4.2 reproduce the two stages of the main argument. The affine collective coordinate Z^F retains the same internal variables and a quadratic relative-coercivity estimate of the same form as for the unperturbed cost. The transformed map $H^F = Z^F - \nabla F(Z^F)$ supplies the external mode: it is an optimal transport map to a penalized Wasserstein barycenter, and the strong monotonicity of $z \mapsto z - \nabla F(z)$ transfers its stability back to Z^F .

4.2 Uniformly concave costs of the sum

Consider costs of the sum, in the form studied by Gangbo–Świech and Heinich [17, 20]:

$$c_h(x_1, \dots, x_N) = h \left(\sum_{i=1}^N \lambda_i x_i \right), \quad (4.21)$$

where $h \in C^2(U)$ for an open neighbourhood $U \supseteq \Omega$. Since adding functions of the individual variables does not change the optimal transport plans, the cost (4.21) is equivalent to the cost function

$$c_0(x_1, \dots, x_N) + F_h(Z_x), \quad F_h(z) := h(z) + \frac{1}{2}|z|^2. \quad (4.22)$$

The quadratic cost (1.1) corresponds to $h(z) = -\frac{1}{2}|z|^2$, and then $F_h = 0$. The following result shows that the uniformly concave costs of the sum are covered by Proposition 4.2.

Corollary 4.3. *Assume that $h \in C^2(U)$ with $U \supseteq \Omega$, satisfies, as quadratic forms on U ,*

$$-\tau_h I \leq D^2 h \leq -\kappa_h I, \quad (4.23)$$

for some constants $0 < \kappa_h \leq \tau_h < \infty$. Let $\rho \in \mathcal{P}_{\text{ac}}(S) \cap L^\infty(S)$ and let c_h be defined by (4.21). Then there exists a constant $C > 0$, depending only on $\lambda_1, N, \text{diam}(S \cup Y), \kappa_h, \tau_h, n$, and $\|\rho\|_{L^\infty}$, such that

$$\sum_{i=2}^N \lambda_i \|T_i^h - \tilde{T}_i^h\|_{L^2(\rho)} \leq C \left(\sum_{i=2}^N \lambda_i W_1(\mu_i, \tilde{\mu}_i) \right)^{1/4},$$

where (T_2^h, \dots, T_N^h) and $(\tilde{T}_2^h, \dots, \tilde{T}_N^h)$ are the unique multi-marginal optimal transport maps associated with $(\rho, \mu_2, \dots, \mu_N)$ and $(\rho, \tilde{\mu}_2, \dots, \tilde{\mu}_N)$, respectively.

Proof. The existence and uniqueness of multi-marginal optimal transport maps follow from [23, Theorem 3.1].

Denote $\kappa_* := \min\{\kappa_h, 1\}$, and $\tau_* := \max\{\tau_h - 1, 0\}$. Then by (4.22) and (4.23),

$$D^2 F_h \geq (1 - \tau_h)I \geq -\tau_* I, \quad D^2 F_h \leq (1 - \kappa_h)I \leq (1 - \kappa_*)I,$$

which means that (4.3) holds with $\kappa_F = \kappa_*$ and $\tau_F = \tau_*$. Hence Proposition 4.2 proves the corollary. \square

4.3 From cyclic interactions to complete-graph symmetry

We next isolate a simple symmetry principle behind complete-graph interaction costs. This discussion is not used in the proof of the stability estimates above, but it explains why the fully symmetric quadratic cost studied in the main part of the paper is naturally related to cyclic nearest-neighbour interactions. The observation is purely combinatorial and therefore does not depend on the Euclidean structure of the ambient space.

Let X be a set and let $k : X \times X \rightarrow \mathbb{R}$ be a function. Define

$$G_{\text{cyc}}(x_1, \dots, x_N) := \frac{1}{N} \sum_{i=1}^N k(x_i, x_{i+1}), \quad x_{N+1} := x_1. \quad (4.24)$$

Then G_{cyc} is invariant under the cyclic group $C_N := \langle (12 \cdots N) \rangle$, but it need not be invariant under the full symmetric group S_N . Its full symmetrization is the complete-graph average.

Proposition 4.4 (Full symmetrization of a cyclic cost). *Define G_{cyc} by (4.24). Then*

$$\frac{1}{N!} \sum_{\sigma \in S_N} G_{\text{cyc}}(x_{\sigma(1)}, \dots, x_{\sigma(N)}) = \frac{1}{N(N-1)} \sum_{i \neq j} k(x_i, x_j). \quad (4.25)$$

If k is symmetric, then

$$\frac{1}{N!} \sum_{\sigma \in S_N} G_{\text{cyc}}(x_{\sigma(1)}, \dots, x_{\sigma(N)}) = \frac{2}{N(N-1)} \sum_{1 \leq i < j \leq N} k(x_i, x_j). \quad (4.26)$$

Proof. Fix an ordered pair (i, j) with $i \neq j$. In the sum over $\sigma \in S_N$ and over cyclic edges $\ell = 1, \dots, N$, the ordered pair (x_i, x_j) appears as $(x_{\sigma(\ell)}, x_{\sigma(\ell+1)})$ exactly $(N(N-2)!)^2$ times: after choosing the cyclic edge ℓ , the two values i, j are fixed at two neighbouring cyclic positions and the remaining $N-2$ entries are arbitrary. Therefore

$$\begin{aligned} \frac{1}{N!} \sum_{\sigma \in S_N} G_{\text{cyc}}(x_{\sigma(1)}, \dots, x_{\sigma(N)}) &= \frac{1}{N!} \frac{1}{N} N(N-2)! \sum_{i \neq j} k(x_i, x_j) \\ &= \frac{1}{N(N-1)} \sum_{i \neq j} k(x_i, x_j). \end{aligned}$$

If k is symmetric, the ordered-pair sum equals twice the sum over unordered pairs. \square

For the quadratic kernel $k(x, y) = \frac{1}{2}|x - y|^2$, Proposition 4.4 gives

$$\frac{1}{N!} \sum_{\sigma \in S_N} G_{\text{cyc}}(x_{\sigma(1)}, \dots, x_{\sigma(N)}) = \frac{1}{N(N-1)} \sum_{1 \leq i < j \leq N} |x_i - x_j|^2. \quad (4.27)$$

In the equal-weight case $\lambda_i = 1/N$, the barycentric quadratic cost (1.1) is

$$c(x_1, \dots, x_N) = \frac{1}{2N^2} \sum_{1 \leq i < j \leq N} |x_i - x_j|^2. \quad (4.28)$$

Thus

$$\frac{1}{N!} \sum_{\sigma \in S_N} G_{\text{cyc}}(x_{\sigma(1)}, \dots, x_{\sigma(N)}) = \frac{2N}{N-1} c(x_1, \dots, x_N). \quad (4.29)$$

Hence the equal-weight barycentric quadratic cost can be viewed, up to a constant factor, as the full S_N -symmetrization of a cyclic quadratic nearest-neighbour cost.

Remark 4.5 (Quotient interpretation). The cyclic cost G_{cyc} descends to the quotient X^N/C_N . Its full symmetrization descends further to the unordered quotient X^N/S_N . When X is a metric space, X^N/S_N can be identified with the space of equally weighted empirical measures endowed with the corresponding Wasserstein matching distance. This quotient-to-empirical-measure viewpoint appears, for example, in Feng's work on Hamilton–Jacobi equations for hydrodynamic limits of collective dynamics [16]. Proposition 4.4 describes the same passage at the level of graph interaction energies: averaging a cycle over all labellings turns it into a complete graph.

The preceding observation motivates the following discussion of general pairwise graph costs, where the complete graph is replaced by an arbitrary weighted graph.

4.4 Normal modes for pairwise quadratic interaction costs

The preceding symmetrization result singles out the complete graph as the fully symmetric pairwise model. We now examine what remains of the two-mode mechanism for an arbitrary weighted graph. The conclusion is a normal-mode decomposition of the Kantorovich defect; no general stability theorem is claimed.

Consider the pairwise quadratic interaction costs

$$c_A(x_1, \dots, x_N) := \frac{1}{2} \sum_{1 \leq i < j \leq N} a_{ij} |x_i - x_j|^2, \quad a_{ij} = a_{ji} \geq 0. \quad (4.30)$$

Such costs are common in multi-marginal optimal transport and in models of interacting systems (cf. [31, 34]). In the factorized complete-graph case they reduce to the cost studied in the main part of this paper. Indeed, if

$$a_{ij} = m_i m_j, \quad m_i > 0, \quad M := \sum_{i=1}^N m_i, \quad \lambda_i := \frac{m_i}{M},$$

then

$$c_A(x_1, \dots, x_N) = M^2 \left(\frac{1}{2} \sum_{i=1}^N \lambda_i |x_i|^2 - \frac{1}{2} \left| \sum_{i=1}^N \lambda_i x_i \right|^2 \right). \quad (4.31)$$

Thus our main results above apply to this class after changing the constants.

For general coefficients a_{ij} , the relevant object for our method is the target–target quadratic form appearing in the Kantorovich defect. More precisely, let $(\varphi_1^A, \dots, \varphi_N^A)$ be a c_A -conjugate tuple of Kantorovich potentials. Assume that the c_A -cost multi-marginal problem admits a tuple of multi-marginal optimal transport maps (T_2, \dots, T_N) . Define

$$R_A(x_1, \dots, x_N) := c_A(x_1, \dots, x_N) - \sum_{i=1}^N \varphi_i^A(x_i).$$

By the same c -conjugacy argument as in Proposition 2.3, we have

$$R_A(x_1, \dots, x_N) \geq - \sum_{2 \leq i < j \leq N} a_{ij} \langle x_i - T_i(x_1), x_j - T_j(x_1) \rangle, \quad \rho\text{-a.e. } x_1. \quad (4.32)$$

Define the symmetric target–target matrix A_I by

$$(A_I)_{ii} := 0, \quad (A_I)_{ij} := a_{ij} \quad (i \neq j), \quad I := \{2, \dots, N\}.$$

Let $A_I e^k = \alpha_k e^k$, $k = 1, \dots, N-1$, be an orthonormal spectral decomposition. Writing $W_{k,x} := \sum_{i=2}^N e_i^k x_i$ and $W_k(x_1) := \sum_{i=2}^N e_i^k T_i(x_1)$, we have

$$\begin{aligned} R_A(x_1, \dots, x_N) &\geq - \frac{1}{2} \sum_{k=1}^{N-1} \alpha_k \left| \sum_{i=2}^N e_i^k (x_i - T_i(x_1)) \right|^2 \\ &= \frac{1}{2} \sum_{\alpha_k < 0} |\alpha_k| |W_{k,x} - W_k(x_1)|^2 - \frac{1}{2} \sum_{\alpha_k > 0} \alpha_k |W_{k,x} - W_k(x_1)|^2. \end{aligned} \quad (4.33)$$

The distinction between the interaction energy and the Kantorovich defect is important here. As a quadratic form, c_A is represented by the graph Laplacian associated with

(a_{ij}). After the one-variable quadratic terms have been absorbed by the c_A -conjugate potentials, however, the defect estimate is governed by the off-diagonal target–target matrix A_I . Consequently, connectivity of the graph, or a spectral gap for its Laplacian, does not by itself provide the relative coercivity needed for stability.

Thus the modes with $\alpha_k < 0$ are coercive in the Kantorovich defect, whereas the modes with $\alpha_k \geq 0$ have to be treated as external modes. For a general graph these external modes are signed linear combinations of the multi-marginal optimal transport maps. Therefore the argument here suggests a normal mode extension of the internal and external decomposition, but a closed stability theorem for arbitrary graph costs would require additional stability estimates for these signed modes. We do not pursue this here.

In particular, for the unweighted complete graph, changing the sign leads to two different problems. The attractive cost is

$$c_+(x_1, \dots, x_N) = \frac{1}{2} \sum_{1 \leq i < j \leq N} |x_i - x_j|^2, \quad (4.34)$$

and the repulsive harmonic cost is

$$c_-(x_1, \dots, x_N) = -\frac{1}{2} \sum_{1 \leq i < j \leq N} |x_i - x_j|^2. \quad (4.35)$$

For the attractive cost, our main estimate gives positive coercivity in the internal directions. However, for the repulsive harmonic cost, the sign is reversed; the same decomposition gives anti-coercivity rather than stability. In particular, if the constraint $\sum_{i=1}^N x_i = 0$ can be achieved, all admissible couplings supported on that hyperplane have the same cost. This is consistent with the non-uniqueness and high-dimensional optimal supports known for this cost (cf. [15, 18]). Thus the repulsive harmonic cost is not covered by the stability theorem proved above unless additional structure or regularization is imposed.

4.5 Hedonic costs and generalized external variables

Consider the hedonic or multi-agent matching costs

$$c_H(x_1, \dots, x_N) = \inf_{z \in \mathcal{Z}} \sum_{i=1}^N f_i(x_i, z), \quad (4.36)$$

where z is an auxiliary variable, representing for instance a quality, a contract, a task, or a common matching outcome. Such costs appear in hedonic models and matching for teams problems (cf. [8, 11, 33]).

Under suitable compactness and continuity assumptions, the value of the multi-marginal optimal transport problem with cost (4.36) agrees with the generalized barycenter problem

$$\inf_{\nu \in \mathcal{P}(\mathcal{Z})} \sum_{i=1}^N T_{f_i}(\mu_i, \nu), \quad (4.37)$$

where T_{f_i} is the two-marginal optimal transport with cost f_i .

In this sense, the auxiliary variable z serves as a collective variable for the problem. Hence a stability theorem for (4.36) would require two estimates: an internal coercivity estimate for the Kantorovich defect after eliminating z , and a stability estimate for the generalized barycenter in (4.37). These estimates are not consequences of the quadratic proof above. The discussion here only indicates the analogous decomposition.

4.6 Failure of relative coercivity: Coulomb and determinant costs

Consider the translation-invariant costs, such as the Coulomb interaction

$$c_{\text{coul}}(x_1, \dots, x_N) = \sum_{1 \leq i < j \leq N} \frac{1}{|x_i - x_j|}, \quad (4.38)$$

for which common translations are invisible:

$$c_{\text{coul}}(x_1 + \xi, \dots, x_N + \xi) = c_{\text{coul}}(x_1, \dots, x_N).$$

Thus the cost depends only on relative coordinates. However, this translation-invariant structure is not enough to recover the coercivity used in this paper.

There is a simple obstruction. The Coulomb cost is singular near the collision set and, in dimensions at least two, its Hessian is indefinite away from collisions. Hence it does not admit a global quadratic lower bound in the relative coordinates comparable to our Kantorovich defect estimates. Existing results establish finiteness, continuity, dual potentials, and separation of optimal plans from the collision set under non-concentration assumptions [12, 13, 15]; see also [35] for a recent dynamical formulation of multi-marginal optimal transport including translation-invariant costs. These results address structural aspects of the Coulomb problem, but they do not yield the internal coercivity estimate required by our argument. Thus we do not claim a Coulomb stability result here.

A similar obstruction appears for determinant-type costs (cf. [9, 15, 32])

$$c_{\text{det}}(x_1, \dots, x_{n+1}) = \det(x_2 - x_1, \dots, x_{n+1} - x_1), \quad (4.39)$$

which is invariant under common translations and depends only on the relative edge vectors. However, the determinant is multilinear and sign-indefinite. Its second-order structure is governed by mixed derivatives rather than by a positive quadratic internal energy. Therefore, one cannot expect a global coercivity estimate comparable to our Kantorovich defect estimates. A local result may be possible near configurations satisfying a suitable second-variation condition, but this is outside the scope of the present paper.

The examples above isolate the two possible failure mechanisms of the method. A cost may possess a natural collective variable but fail to be coercive on its fibers, as for Coulomb and determinant interactions; or a collective/normal-mode decomposition may exist while one of the two required estimates remains unavailable, as for general graph interactions and hedonic costs. A broader two-mode theory would therefore require both a uniform lower bound for the Kantorovich defect on the fibers of a finite-dimensional collective coordinate and a stable variational representation of the complementary modes. The barycentric and collective-coordinate costs treated above are precisely cases in which both requirements can be verified.

References

- [1] Martial Agueh and Guillaume Carlier, *Barycenters in the Wasserstein space*, SIAM J. Math. Anal. **43** (2011), no. 2, 904–924. [2](#), [3](#), [4](#), [14](#), [16](#), [21](#), [27](#)
- [2] Luigi Ambrosio and Nicola Gigli, *A user’s guide to optimal transport*, in: Modelling and optimisation of flows on networks, Lecture Notes in Math., vol. 2062, Springer, Heidelberg, 2013, pp. 1–155. [11](#)

- [3] Robert J. Berman, *Convergence rates for discretized Monge–Ampère equations and quantitative stability of optimal transport*, *Found. Comput. Math.* **21** (2021), no. 4, 1099–1140. [2](#)
- [4] Jérémie Bigot, Elsa Cazelles and Nicolas Papadakis, *Penalization of barycenters in the Wasserstein space*, *SIAM J. Math. Anal.* **51** (2019), no. 3, 2261–2285, doi:10.1137/18M1185065. [27](#)
- [5] Jérémie Bigot and Thierry Klein, *Characterization of barycenters in the Wasserstein space by averaging optimal transport maps*, *ESAIM Probab. Stat.* **22** (2018), 35–57. [21](#)
- [6] Yann Brenier, *Polar factorization and monotone rearrangement of vector-valued functions*, *Comm. Pure Appl. Math.* **44** (1991), no. 4, 375–417. [2](#)
- [7] Guillaume Carlier, Alex Delalande, and Quentin Mérigot, *Quantitative stability of barycenters in the Wasserstein space*, *Probab. Theory Related Fields* **188** (2024), no. 3–4, 1257–1286. [14](#)
- [8] Guillaume Carlier and Ivar Ekeland, *Matching for teams*, *Econom. Theory* **42** (2010), no. 2, 397–418. [31](#)
- [9] Guillaume Carlier and Bruno Nazaret, *Optimal transportation for the determinant*, *ESAIM Control Optim. Calc. Var.* **14** (2008), no. 4, 678–698. [32](#)
- [10] Elsa Cazelles, Édouard Pauwels, and Léo Portales, *Statistical estimation of Monge transport maps via Brenier potentials*, arXiv:2604.22366 [math.OC], 2026. [2](#)
- [11] Pierre-André Chiappori, Robert J. McCann, and Lars P. Nesheim, *Hedonic price equilibria, stable matching, and optimal transport: equivalence, topology, and uniqueness*, *Econom. Theory* **42** (2010), no. 2, 317–354. [3](#), [31](#)
- [12] Maria Colombo, Simone Di Marino, and Federico Stra, *Continuity of multimarginal optimal transport with repulsive cost*, *SIAM J. Math. Anal.* **51** (2019), no. 4, 2903–2926. [3](#), [32](#)
- [13] Codina Cotar, Gero Friesecke, and Claudia Klüppelberg, *Density functional theory and optimal transportation with Coulomb cost*, *Comm. Pure Appl. Math.* **66** (2013), no. 4, 548–599. [32](#)
- [14] Alex Delalande and Quentin Mérigot, *Quantitative stability of optimal transport maps under variations of the target measure*, *Duke Math. J.* **172** (2023), no. 17, 3321–3357. [2](#)
- [15] Simone Di Marino, Augusto Gerolin, and Luca Nenna, *Optimal transportation theory with repulsive costs*, in: *Topological optimization and optimal transport*, Radon Ser. Comput. Appl. Math., vol. 17, De Gruyter, Berlin, 2017, pp. 204–256, doi:10.1515/9783110430417-010. [31](#), [32](#)
- [16] Jin Feng, *On a Hamilton–Jacobi PDE theory for hydrodynamic limit of action minimizing collective dynamics*, arXiv:2512.20809v1 [math.AP], 2025. [29](#)
- [17] Wilfrid Gangbo and Andrzej Świech, *Optimal maps for the multidimensional Monge–Kantorovich problem*, *Comm. Pure Appl. Math.* **51** (1998), no. 1, 23–45. [2](#), [3](#), [4](#), [9](#), [15](#), [17](#), [18](#), [28](#)

- [18] Augusto Gerolin, Anna Kausamo, and Tapio Rajala, *Non-existence of optimal transport maps for the multi-marginal repulsive harmonic cost*, SIAM J. Math. Anal. **51** (2019), no. 3, 2359–2371. [31](#)
- [19] Nicola Gigli, *On Hölder continuity-in-time of the optimal transport map towards measures along a curve*, Proc. Edinb. Math. Soc. (2) **54** (2011), no. 2, 401–409. [6](#)
- [20] Henri Heinich, *Problème de Monge pour n probabilités*, C. R. Math. Acad. Sci. Paris **334** (2002), no. 9, 793–795. [28](#)
- [21] L. Kantorovitch, *On the translocation of masses*, C. R. (Dokl.) Acad. Sci. URSS **37** (1942), 199–201. [2](#)
- [22] Hans G. Kellerer, *Duality theorems for marginal problems*, Z. Wahrscheinlichkeitstheorie verw. Gebiete **67** (1984), no. 4, 399–432. [9](#)
- [23] Young-Heon Kim and Brendan Pass, *A general condition for Monge solutions in the multi-marginal optimal transport problem*, SIAM J. Math. Anal. **46** (2014), no. 2, 1538–1550. [24](#), [28](#)
- [24] Young-Heon Kim and Brendan Pass, *Multi-marginal optimal transport on Riemannian manifolds*, Amer. J. Math. **137** (2015), no. 4, 1045–1060. [3](#), [4](#), [9](#), [15](#), [16](#), [17](#), [18](#)
- [25] Young-Heon Kim and Brendan Pass, *Wasserstein barycenters over Riemannian manifolds*, Adv. Math. **307** (2017), 640–683. [16](#)
- [26] Cyril Letrouit, *Quantitative stability in optimal transport*, lecture notes for the Cours Peccot, Collège de France, 2025. [3](#)
- [27] Cyril Letrouit, *Unstable optimal transport maps*, Comptes Rendus Math. **364** (2026), no. G2, 333–344, doi:10.5802/crmath.834. [2](#), [15](#), [17](#)
- [28] Cyril Letrouit and Quentin Mérigot, *Gluing methods for quantitative stability of optimal transport maps*, arXiv:2411.04908v2 [math.AP], 2025. [2](#)
- [29] Quentin Mérigot, *Sharp stability of Brenier maps via quantitative regularity of potentials*, HAL preprint hal-05616391, 2026. [2](#), [5](#), [14](#), [27](#)
- [30] Gaspard Monge, *Mémoire sur la théorie des déblais et des remblais*, Mém. Math. Phys. Acad. Royale Sci. (1781), 666–704. [2](#)
- [31] Luca Nenna and Brendan Pass, *An ODE characterization of multi-marginal optimal transport with pairwise cost functions*, IMA J. Numer. Anal. (2025), doi:10.1093/imanum/draf067. [30](#)
- [32] Brendan Pass, *On the local structure of optimal measures in the multi-marginal optimal transportation problem*, Calc. Var. Partial Differential Equations **43** (2012), no. 3–4, 529–536. [32](#)
- [33] Brendan Pass, *Multi-marginal optimal transport and multi-agent matching problems: uniqueness and structure of solutions*, Discrete Contin. Dyn. Syst. **34** (2014), no. 4, 1623–1639. [3](#), [31](#)
- [34] Brendan Pass, *Multi-marginal optimal transport: theory and applications*, ESAIM Math. Model. Numer. Anal. **49** (2015), no. 6, 1771–1790. [3](#), [30](#)

- [35] Brendan Pass and Yair Shenfeld, *A dynamical formulation of multi-marginal optimal transport*, arXiv:2509.22494v2 [math.OC], 2025. [3](#), [32](#)
- [36] Cédric Villani, *Optimal transport. Old and new*, Grundlehren der mathematischen Wissenschaften, vol. 338, Springer, Berlin, 2009. [5](#)

Declaration. The authors declare that they have no conflict of interest and that the manuscript has no associated data.

Funding: This work was supported in part by the National Key R&D Program of China (2021YFA1000900, 2021YFA1002200), the National Natural Science Foundation of China (12201596), the Shandong Provincial Natural Science Foundation (ZR2025QB05) and Tais-han Scholars Program of Shandong Province (tsqn202408059).