Stability of Quadratically Regularized Optimal Transport

Alberto González-Sanz* Marcel Nutz†

May 28, 2026

Abstract

Quadratically regularized optimal transport (QOT) is a sparse alternative to entropic optimal transport. We develop a quantitative stability theory for QOT under perturbations of the marginals, the transport cost function, and the regularization parameter. The centerpiece is an L^∞-stability result for the dual potentials. Starting from an error bound in an L^2-space that varies with the marginals, we use a self-bound for the potentials to derive a local L^∞-Lipschitz bound that is uniform over marginals. This bound also yields stability of the optimal coupling and of its support. In particular, we show that for the quadratic transport cost, the support of the optimal coupling is locally Lipschitz in Hausdorff distance under perturbations of the marginals. To the best of our knowledge, this is the first stability result for the optimal support in regularized optimal transport.

Keywords Optimal Transport; Quadratic Regularization; Stability; Support
AMS 2020 Subject Classification 49N10; 49N05; 90C25

Contents

1 Introduction 1
2 Setting and background 5
3 Stability of the potentials in L^2 7
4 Stability of the potentials in L^∞ 9
5 Stability of the optimal coupling 10
6 Stability of the optimal support 12
A Proofs 15

1 Introduction

Optimal transport provides a variational framework for comparing probability measures that has become ubiquitous across optimization, machine learning, statistics, and analysis. In applications, optimal transport is often regularized to reduce computational cost and improve statistical sample complexity. The best-studied regularization is entropic optimal transport (EOT), where the transport coupling is penalized by Kullback–Leibler (KL) divergence; see,

*Department of Statistics, Columbia University, ag4855@columbia.edu.
†Departments of Mathematics and Statistics, Columbia University, mnutz@columbia.edu. Research supported by NSF Grants DMS-2106056, DMS-2407074.

e.g., [7, 29, 27]. EOT leads to Sinkhorn's algorithm and a smooth dual problem. However, the KL penalty forces the optimal coupling to have full support while unregularized optimal transport is typically sparse, and weak regularization can cause numerical instability through exponentially large and small scalings [31]. The most prominent alternative, first considered in [26, 2, 13], is the quadratic regularization, or equivalently χ^2-divergence. It produces sparse couplings, as observed in [2, 13, 24, 1] and recently analyzed theoretically in [32, 18, 28, 17]. It also replaces the exponential structure in the EOT dual by a quadratic, thereby avoiding exponential scaling. Recent work has shown that QOT enjoys parametric sample complexity [16] and linear convergence of several dual algorithms [19, 20], establishing QOT as a viable alternative to EOT.

Given probability measures P, Q on $\mathbb{R}^d$, a cost function $c(x, y)$, and a regularization parameter $\varepsilon > 0$, the QOT problem is

$$\mathrm{QOT}_\varepsilon(P, Q) := \inf_{\pi \in \Pi(P,Q)} \int c(x, y)\, d\pi(x, y) + \frac{\varepsilon}{2} \left\| \frac{d\pi}{d(P \otimes Q)} \right\|^2_{L^2(P \otimes Q)},$$

where $\Pi(P, Q)$ denotes the set of couplings of (P, Q). In practice, the most important example is the quadratic cost $c(x, y) = \frac{1}{2}\|x-y\|^2$, but we shall develop the theory for a generic Lipschitz cost on a subset $\mathcal{X} \times \mathcal{Y} \subset \mathbb{R}^d \times \mathbb{R}^d$. The dual variables of the QOT problem are the so-called potentials f, g solving the dual problem

$$\sup_{(f,g) \in L^2(P) \times L^2(Q)} \int \left(f(x) + g(y) - \frac{1}{2\varepsilon} \big(f(x) + g(y) - c(x, y) \big)^2_+ \right) d(P \otimes Q)(x, y),$$

where $t_+ := \max\{t, 0\}$. Its first-order conditions of optimality are

$$\int \big(f(x) + g(y) - c(x, y) \big)_+ \, dQ(y) = \varepsilon, \qquad \int \big(f(x) + g(y) - c(x, y) \big)_+ \, dP(x) = \varepsilon \tag{1}$$

for $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, respectively. The potentials are unique up to translation under a mild connectedness assumption, so that their direct sum $h = f \oplus g$ is uniquely determined. The primal optimizer π is recovered through

$$\frac{d\pi}{d(P \otimes Q)} = \frac{1}{\varepsilon} \sigma_+ \qquad \text{with the slack variable} \qquad \sigma := h - c, \tag{2}$$

and its support is

$$\mathrm{supp}(\pi) = \overline{\{(x, y) \in \Omega_P \times \Omega_Q : \sigma(x, y) > 0\}}. \tag{3}$$

Thus $h = f \oplus g$ determines the key objects of QOT, motivating us to view

$$(P, Q, c, \varepsilon) \mapsto h_{(P,Q,c,\varepsilon)} \tag{4}$$

as the solution map of our problem.

The present paper provides a quantitative stability theory for this map. Our motivating question is the stability of the optimal support, about which, to the best of our knowledge, there are no prior results. In EOT, the optimal coupling always has full support (i.e., the same support as $P \otimes Q$), and hence support stability is not relevant. Since sparsity is the key difference between QOT and EOT, support stability is arguably the most interesting aspect among the stability properties. We answer this question by showing that the optimal support for quadratic cost is locally Lipschitz continuous (in Hausdorff distance) with respect to the marginals (P, Q), with explicit constants detailed below. In view of (3), our approach is to develop L^∞-stability of $h = f \oplus g$ (and hence of σ), and to further ensure that σ is not flat close to its zero-level set (which marks the boundary of the support).

Our analysis begins with a systematic study of the solution map (4). For simplicity, we focus on perturbations of the marginals (P, Q) in this introduction, and fix c, ε. While our results in the main text cover joint variations of the data (P, Q, c, ε), varying (c, ε) is relatively straightforward because it does not change the function spaces. Our analysis can be contrasted with EOT, where the dual problem is strongly concave and the first-order condition for the potentials is the Schrödinger system

$$\begin{aligned} f(x) &= -\varepsilon \log \int \exp\left(\frac{g(y) - c(x,y)}{\varepsilon}\right) dQ(y), \\ g(y) &= -\varepsilon \log \int \exp\left(\frac{f(x) - c(x,y)}{\varepsilon}\right) dP(x). \end{aligned} \tag{5}$$

As shown by [5], stability results can be derived by applying the implicit function theorem to this system. For QOT, the dual problem is not strongly concave and this approach fails. However, the recent work [20] showed that a local PL inequality/error bound (see Theorem 2.5 below) holds in $L^2(P \otimes Q)$. It is well known (e.g., [3, Section 4.4]) that a PL inequality implies a stability property. However, a direct application would only lead to a limit result: to work with a fixed L^2 space, one would have to restrict to marginals that are absolutely continuous with respect to fixed reference measures P_0, Q_0. By contrast, Theorem 3.1 and Corollary 3.3 below provide local L^2-Lipschitz estimates where the input data are compared in the 1-Wasserstein metric. (Here "local" means that the marginals under consideration must have W_1-distance below an explicit threshold.)

L^2-stability serves as a stepping stone for the L^∞-stability provided in Theorem 4.1 and Corollary 4.2. The upgrade from L^2 to L^∞ is based on a self-bound for the potentials. Such a self-bound is straightforward for EOT, where the Schrödinger system (5) directly expresses f, g as integrals of themselves. For QOT, the self-bound is nontrivial, but can be shown by a monotonicity argument and (1) once the potentials are sufficiently close in L^2 (see Step 3 in the proof of Theorem 4.1).

Stability of the primal objects follows from the dual estimates. Indeed, stability of the optimal density follows directly via (2). This density is with respect to a reference $P \otimes Q$ which is itself varying, but nevertheless allows us to conclude local Lipschitz continuity of the optimal coupling in both total variation and 1-Wasserstein distance. The primal stability results are stated in Theorem 5.1 and Remark 5.2.

We finally return to the optimal support. Potential stability alone cannot control supports: when $c \equiv 0$, the optimal coupling is $P \otimes Q$ and $h \equiv \varepsilon$, so the potentials do not register changes in the marginal supports. Our result therefore takes the Hausdorff displacement of the marginal supports as input, in addition to the quantities controlling the potentials. Under a nondegeneracy condition, Theorem 6.1 then establishes local Lipschitz continuity of the optimal support in Hausdorff distance. The nondegeneracy condition, detailed in (32), states that the slack $\sigma = h - c$ detaches linearly outside the optimal support. Intuitively, this means that the gradient of σ is bounded away from zero at the boundary of the support (which is itself the zero-level set of σ). Whether this holds depends on the geometry of the cost c. For the quadratic cost $c(x, y) = \frac{1}{2}\|x - y\|^2$, the slack σ is concave in each coordinate, and we show in Proposition 6.3 that this implies the nondegeneracy condition with a slope of at least $\varepsilon/\operatorname{diam}(\operatorname{supp} P)$. As a result, the optimal support is indeed stable for the canonical cost (Corollary 6.4), answering our initial question. On the other hand, the nondegeneracy condition is essential for support stability and can fail for certain costs: in Example 6.2, we construct a setting where the marginals converge, the marginal supports do not move, and the potentials converge uniformly, but the optimal supports fail to converge.

Related literature. Classical optimal transport has a substantial stability theory, going back to [6], which proves weak convergence of optimal couplings under weak convergence of the marginal distributions. Quantitative estimates were given by [15], which studies the optimal map from a fixed source to targets moving along a Wasserstein curve. More recently, [8] proves bi-Hölder stability of Brenier maps under Wasserstein perturbations of the target with dimension-free exponents. See, e.g., [21, 23, 25] for more extensive references.

Quadratic and general divergence regularization. The closest stability result for non-entropic regularization is [1], which studies divergence-regularized optimal transport for a broad class of divergences, including the χ^2-divergence underlying QOT. The authors prove $\frac{1}{2p}$-Hölder stability of the optimal couplings in p-Wasserstein distance under perturbations of the marginals, and also derive sample-complexity bounds for empirical marginals. Previously, [11] showed Lipschitz continuity of the optimal value. Both works are based on the shadow technique and the data processing inequality, which do not control the potentials or the supports. More recently, [16] proves parametric sample complexity and central limit theorems for QOT, but does not address deterministic stability or the optimal supports. In a different direction, [12] studies convergence for vanishing regularization parameter $\varepsilon \to 0$, focusing on the optimal value rather than the optimizers.

Several works analyze the structure of QOT and the support of its optimal coupling. Duality is developed in [24] and [28]; the latter also establishes a qualitative sparsity result and uniqueness of the potentials. For quadratic transport cost, sparsity is studied quantitatively in [32, 18, 17]. These works describe the geometry of the support for small ε and fixed marginals; they do not investigate stability. On the algorithmic side, [19] proves linear convergence of gradient descent for the QOT dual. Its proof technique is foundational for the subsequent work [20] proving a local error bound and Polyak–Łojasiewicz inequality for the QOT dual in $L^2(P \otimes Q)$, where the marginals P, Q are fixed. The present paper uses the error bound of [20] as a starting point, but the main developments here are different: the underlying Hilbert space $L^2(P \otimes Q)$ moves with the marginals P, Q, and support stability in the Hausdorff sense requires L^∞-control of the potentials.

Entropic regularization. For entropic optimal transport, stability is much better developed. The differential approach of [5] proves well-posedness and smooth dependence of the Schrödinger system in an L^∞ setting via the implicit function theorem. Refining this approach, [4] show that the potentials are C^k-Lipschitz from the L^2-Wasserstein space of marginals when the transport cost is C^{k+1}. In a more general setting, [14] use a geometric approach based on cyclical invariance to prove qualitative stability of the primal EOT optimizers for weak convergence with respect to perturbations of the marginals and cost. Analyzing Sinkhorn's algorithm in a bounded setting, [9] show that its iterates are L^∞-Lipschitz uniformly over the iteration number, and hence that the potentials are L^∞-Lipschitz, as functions of the marginals in 1-Wasserstein distance. In an unbounded setting, [11] establish Lipschitz continuity of the entropic value and Hölder continuity of the optimal coupling in Wasserstein distance using the shadow construction. Stability of entropic Brenier maps is treated in [10], which obtains sharp Lipschitz constants under target perturbations and improves the dependence on the regularization parameter in the semi-discrete regime. Recent contributions include [33], which proves quantitative stability for many-marginal Schrödinger bridges. For the reasons explained above, the techniques used in EOT do not directly extend to QOT. Moreover, the question of support stability does not arise in EOT.

Organization. Section 2 details the setting and collects basic facts about QOT. Section 3 establishes the L^2-stability of the potentials, whereas L^∞-stability is shown in Section 4. Section 5 derives stability of the optimal densities and couplings. Section 6 concludes with the Hausdorff stability of the optimal support. All proofs are gathered in Appendix A.

2 Setting and background

We fix two subsets $\mathcal{X}, \mathcal{Y} \subset \mathbb{R}^d$ and denote by Ω_ν the topological support of a measure ν. All first marginal probability measures considered in this paper are assumed to have finite second moments and support contained in $\mathcal{X}$, and all second marginal probability measures are assumed to have finite second moments and support contained in $\mathcal{Y}$. The main purpose of the sets $\mathcal{X}, \mathcal{Y}$ is to localize the problem so that, e.g., the quadratic cost becomes Lipschitz. If c is Lipschitz on $\mathbb{R}^d \times \mathbb{R}^d$, one may simply take $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$. Throughout, we use the Euclidean norms on $\mathbb{R}^d$ and $\mathbb{R}^d \times \mathbb{R}^d$ unless otherwise stated.

Assumption 2.1 (Cost)**.** The cost $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is Lipschitz with constant $L > 0$.

The quadratically regularized optimal transport (QOT) problem with regularization parameter $\varepsilon > 0$ is

$$\mathrm{QOT}_\varepsilon(P, Q) := \inf_{\pi \in \Pi(P,Q)} \int c(x, y)\, d\pi(x, y) + \frac{\varepsilon}{2} \left\| \frac{d\pi}{d(P \otimes Q)} \right\|^2_{L^2(P \otimes Q)} \tag{6}$$

with the convention that the last term is $+\infty$ if $\pi \not\ll P \otimes Q$. The dual problem is

$$\mathrm{QOT}^\star_\varepsilon(P, Q) = \sup_{(f,g) \in L^2(P) \times L^2(Q)} \Phi(f \oplus g), \tag{7}$$

where $f \oplus g(x, y) := f(x) + g(y)$ denotes the direct sum and the dual objective function Φ is

$$\Phi(f \oplus g) := \int \left(f(x) + g(y) - \frac{1}{2\varepsilon} \left(f(x) + g(y) - c(x, y) \right)_+^2 \right) d(P \otimes Q)(x, y). \tag{8}$$

The next proposition summarizes several standard properties. For a function w defined on a set A, we write $\|w\|_{L^\infty(A)} := \sup_{z \in A} |w(z)|$.

Proposition 2.2. *Let P, Q be probability measures on $\mathbb{R}^d$ with finite second moments and let c satisfy Assumption 2.1. Then:*

(i) The strong duality $\mathrm{QOT}_\varepsilon(P, Q) = \mathrm{QOT}^\star_\varepsilon(P, Q)$ holds.

(ii) *The dual problem* (7) *admits an optimizer* $(f, g) \in L^2(P) \times L^2(Q)$. *If* Ω_P *is connected, the functions* (f, g) *are unique in* $L^1(P) \times L^1(Q)$ *up to translation, in the sense that fixing one optimizer* (f, g), *the set of all optimizers is* $\{(f + a, g - a) : a \in \mathbb{R}\}$.

(iii) *We can choose versions* $f : \mathcal{X} \to \mathbb{R}$ *and* $g : \mathcal{Y} \to \mathbb{R}$ *that are L-Lipschitz and satisfy the first-order condition*

$$\begin{cases} \int (f(x) + g(y) - c(x, y))_+ \, dQ(y) = \varepsilon & \textit{for all } x \in \mathcal{X}, \\ \int (f(x) + g(y) - c(x, y))_+ \, dP(x) = \varepsilon & \textit{for all } y \in \mathcal{Y}. \end{cases} \tag{9}$$

In the following, we always choose such versions, and call them potentials. If Ω_P *is connected, they are again unique up to translation.*

(iv) *If* c *is bounded on* $\mathcal{X} \times \mathcal{Y}$, *then so are* (f, g). *More precisely, their oscillations are bounded as* $\operatorname{osc}_{\mathcal{X}}(f), \operatorname{osc}_{\mathcal{Y}}(g) \le \operatorname{osc}_{\mathcal{X}\times\mathcal{Y}}(c) \le 2\|c\|_{L^\infty(\mathcal{X}\times\mathcal{Y})}$, *and*

$$-5\|c\|_{L^\infty(\mathcal{X}\times\mathcal{Y})} + \varepsilon \le f(x) + g(y) \le 5\|c\|_{L^\infty(\mathcal{X}\times\mathcal{Y})} + \varepsilon, \qquad (x, y) \in \mathcal{X} \times \mathcal{Y}. \tag{10}$$

(v) *The primal problem* (6) *has a unique solution* $\pi \in \Pi(P, Q)$. *It is related to the dual by*

$$\frac{d\pi}{d(P \otimes Q)}(x, y) = \frac{1}{\varepsilon}\big(f(x) + g(y) - c(x, y)\big)_+ \qquad P \otimes Q\textit{-a.s.} \tag{11}$$

and its support is

$$\operatorname{supp}(\pi) = \overline{\{(x, y) \in \Omega_P \times \Omega_Q : f(x) + g(y) - c(x, y) > 0\}}. \tag{12}$$

Let $\mathcal{L}_d$ denote the d-dimensional Lebesgue measure and $B_r(x)$ the open ball of center x and radius $r > 0$. The following additional conditions on the reference marginals P, Q are used in the error bound (Theorem 2.5) below.

Assumption 2.3 (Reference marginals)**.** (a) The first marginal P has convex, compact support Ω_P and admits a density $\rho := dP/d\mathcal{L}_d$ that is bounded away from zero and infinity on Ω_P, i.e., there exist constants $0 < \lambda_P \le \Lambda_P < \infty$ such that

$$\lambda_P \le \rho(x) \le \Lambda_P \quad \text{for all } x \in \Omega_P.$$

(b) The second marginal Q has compact support Ω_Q.

Assumption 2.3 implies the following bounds on ball measures.

Remark 2.4 (Lower bounds for ball measures)**.** (a) Since Ω_P is compact and convex, it satisfies a uniform interior cone condition: for every $x \in \Omega_P$, there exists a convex cone $\mathcal{C}_x$ with vertex x, angle $\theta > 0$ and height $h > 0$ such that $\mathcal{C}_x \subset \Omega_P$, where θ and h are independent of x (e.g., [22, p. 12]). Hence, Assumption 2.3 implies that there exists $\delta_P \in (0, 1]$, depending only on Ω_P and λ_P, such that

$$P(B_r(x)) \ge \delta_P \, \min(r^d, 1) \quad \text{for all } r > 0 \text{ and } x \in \Omega_P.$$

(b) Since Ω_Q is compact, lower semicontinuity of $y \mapsto Q(B_r(y))$ implies

$$\inf_{y \in \Omega_Q} Q(B_r(y)) > 0 \quad \text{for all } r > 0.$$

Next, consider the linear subspace

$$\mathcal{H} = \big\{u \oplus v : (u, v) \in L^2(P) \times L^2(Q)\big\} \subset L^2(P \otimes Q),$$

which is closed by the argument in [30, p. 370]. Hence, $\mathcal{H}$ is a Hilbert space with the induced inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}} = \langle \cdot, \cdot \rangle_{L^2(P\otimes Q)}$. In the following, we consider the dual objective Φ of (8) as an operator

$$\Phi : \mathcal{H} \to \mathbb{R}$$

and denote by $\mathrm{D}\Phi(h)$ its $\mathcal{H}$-gradient. For any $h, w \in \mathcal{H}$, the action of $\mathrm{D}\Phi(h)$ on w is

$$\langle \mathrm{D}\Phi(h), w\rangle_{L^2(P\otimes Q)} = \int w\, R_h\, d(P \otimes Q), \tag{13}$$

where

$$R_h := 1 - \frac{1}{\varepsilon}(h - c)_+ \in L^2(P \otimes Q).$$

In particular,

$$\|\mathrm{D}\Phi(h)\|_{L^2(P\otimes Q)} = \sup_{w\in\mathcal{H}:\, \|w\|_{L^2(P\otimes Q)}\le 1} \int w\, R_h\, d(P \otimes Q). \tag{14}$$

This norm governs the right-hand side in the following local error bound for the dual problem, established in [20, Theorem 3.1].

Theorem 2.5 (Error bound)**.** *Let (P, Q) satisfy Assumption 2.3, let c satisfy Assumption 2.1, let $\varepsilon > 0$, and let (f, g) be the associated potentials. Setting $h = f \oplus g \in \mathcal{H}$, the error bound*

$$\|\tilde{h} - h\|_{L^2(P\otimes Q)} \le \gamma_\varepsilon \max\left(\|\tilde{h} - h\|_{L^\infty(P\otimes Q)}, \varepsilon\right) \|\mathrm{D}\Phi(\tilde{h})\|_{L^2(P\otimes Q)}$$

holds for all $\tilde{h} \in \mathcal{H}$, where

$$\gamma_\varepsilon = 16 \left(\delta_P^{-1} \max\left(\frac{8L}{\varepsilon}, 1\right)^d\right) \frac{\Lambda_P^2}{\lambda_P^2} \frac{\big(\lceil 8L \operatorname{diam}(\Omega_P)/\varepsilon\rceil\big)^{d+2}}{\inf_{y\in\Omega_Q} Q\big(B_{\frac{\varepsilon}{8L}}(y)\big)}.$$

3 Stability of the potentials in L^2

Our first result reformulates the local error bound into an L^2-estimate for nearby data. The nontrivial part is (ii), establishing that the locality condition—which is in the stronger L^∞-norm—is automatically satisfied when the data are close enough.

Theorem 3.1 (L^2-Lipschitz stability)**.** *Let $\varepsilon, \varepsilon' > 0$, let c, c' satisfy Assumption 2.1, let (P, Q) satisfy Assumption 2.3, and let (P', Q') be any marginals with finite second moments. Let f, g be the potentials (unique up to translation) associated with (P, Q, c, ε), and let f', g' be (any) potentials associated with $(P', Q', c', \varepsilon')$. Set $h := f \oplus g$ and $h' := f' \oplus g'$, and*

$$\Delta := 2L \left(W_1(P, P')^2 + W_1(Q, Q')^2\right)^{1/2} + \|c - c'\|_{L^2(P\otimes Q)} + |\varepsilon - \varepsilon'|.$$

Then, the following hold.

(i) Local error-bound: We have

$$\|h - h'\|_{L^2(P\otimes Q)} \le \frac{\gamma_\varepsilon}{\varepsilon} \max\left(\|h - h'\|_{L^\infty(\Omega_P\times\Omega_Q)}, \varepsilon\right) \Delta. \tag{15}$$

(ii) Modulus of continuity into L^∞: For $\delta > 0$, define

$$\vartheta_\delta := \delta_P \min\left\{\left(\frac{\delta}{8L}\right)^d, 1\right\} \inf_{y\in\Omega_Q} Q\Big(B_{\delta/(8L)}(y)\Big). \tag{16}$$

Then

$$\Delta < \frac{\min(\delta,\varepsilon)\sqrt{\vartheta_\delta}}{2\gamma_\varepsilon} \qquad \Longrightarrow \qquad \|h - h'\|_{L^\infty(\Omega_P\times\Omega_Q)} < \delta. \tag{17}$$

(iii) Local Lipschitz continuity in L^2: We have

$$\|h - h'\|_{L^2(P\otimes Q)} \le \gamma_\varepsilon \Delta \qquad \textit{whenever } \Delta < \frac{\varepsilon\sqrt{\vartheta_\varepsilon}}{2\gamma_\varepsilon}.$$

The estimate in Theorem 3.1 is asymmetric—both the norm and the constants are tied to the reference datum (P, Q, c, ε). The next definition introduces a class on which the relevant constants are uniform, yielding a symmetric corollary.

Definition 3.2 (Class $\mathfrak{D}$ of data)**.** Fix constants

$$\underline{\varepsilon}, D, L \in (0,\infty), \qquad 0 < \underline{\lambda} \le \overline{\Lambda} < \infty, \qquad 0 < \underline{\delta}, \underline{q} \le 1.$$

Let $\mathfrak{D} = \mathfrak{D}(\underline{\varepsilon}, D, L, \underline{\lambda}, \overline{\Lambda}, \underline{\delta}, \underline{q})$ be the class of all quadruples (P, Q, c, ε) such that:

(a) $\varepsilon \ge \underline{\varepsilon}$;

(b) $c : \mathcal{X}\times\mathcal{Y} \to \mathbb{R}$ is L-Lipschitz;

(c) Ω_P is compact and convex, and $\mathrm{diam}(\Omega_P) \le D$;

(d) P admits a density $\rho = dP/d\mathcal{L}_d$ satisfying $\underline{\lambda} \le \rho(x) \le \overline{\Lambda}$ for all $x \in \Omega_P$;

(e) P satisfies $P(B_r(x)) \ge \underline{\delta} \min(r^d, 1)$ for all $x \in \Omega_P$, $r > 0$;

(f) Ω_Q is compact and $\inf_{y\in\Omega_Q} Q\big(B_{\underline{\varepsilon}/(8L)}(y)\big) \ge \underline{q}$.

The preceding theorem becomes symmetric on this class, after applying it with each datum as the reference.

Corollary 3.3 (Uniform L^2-stability)**.** *Let $\mathfrak{D}$ be as in Definition 3.2, and define*

$$\begin{aligned}
\overline{\gamma} &:= 16\left(\underline{\delta}^{-1}\max\left(\frac{8L}{\underline{\varepsilon}}, 1\right)^d\right)\frac{\overline{\Lambda}^2}{\underline{\lambda}^2}\frac{\left(\lceil 8L\,D/\underline{\varepsilon}\rceil\right)^{d+2}}{\underline{q}} \\
\underline{\vartheta} &:= \underline{\delta}\min\left\{\left(\frac{\underline{\varepsilon}}{8L}\right)^d, 1\right\}\underline{q} \\
\overline{\eta} &:= \frac{\underline{\varepsilon}\sqrt{\underline{\vartheta}}}{2\overline{\gamma}}.
\end{aligned}$$

For $(P,Q,c,\varepsilon) \in \mathfrak{D}$*, let* $h_{(P,Q,c,\varepsilon)} := f \oplus g$*, where* (f,g) *are the associated potentials. Given*

$$(P,Q,c,\varepsilon),\ (P',Q',c',\varepsilon') \in \mathfrak{D},$$

abbreviate $h := h_{(P,Q,c,\varepsilon)}$ *and* $h' := h_{(P',Q',c',\varepsilon')}$*, and set*

$$\begin{aligned}
\Delta_W &:= \Big(W_1(P,P')^2 + W_1(Q,Q')^2\Big)^{1/2} \\
\Delta &:= 2L\,\Delta_W + \|c-c'\|_{L^2(P\otimes Q)} + |\varepsilon-\varepsilon'| \\
\Delta' &:= 2L\,\Delta_W + \|c-c'\|_{L^2(P'\otimes Q')} + |\varepsilon-\varepsilon'| \\
\bar{\mu} &:= \frac{P\otimes Q + P'\otimes Q'}{2} \\
\bar{\Delta} &:= 2L\,\Delta_W + \|c-c'\|_{L^2(\bar{\mu})} + |\varepsilon-\varepsilon'|.
\end{aligned}$$

If

$$\max\{\Delta,\Delta'\} < \overline{\eta},$$

then

$$\|h-h'\|_{L^2(P\otimes Q)} \le \overline{\gamma}\,\Delta \tag{18}$$

$$\|h-h'\|_{L^2(P'\otimes Q')} \le \overline{\gamma}\,\Delta' \tag{19}$$

$$\|h-h'\|_{L^2(\bar{\mu})} \le \overline{\gamma}\,\bar{\Delta}. \tag{20}$$

4 Stability of the potentials in L^∞

Our next result upgrades the $L^2(P\otimes Q)$-estimate of Theorem 3.1 to an $L^\infty(\mathcal{X}\times\mathcal{Y})$-estimate. This is important in two respects. First, it removes the dependence on the marginals in the definition of the norm measuring the error, thus allowing for a statement that is uniform across a class of marginals. Second, L^∞-control will be crucial to control the optimal support in Hausdorff sense in Section 6.

Theorem 4.1 (L^∞-Lipschitz stability). *Let* $\varepsilon,\varepsilon' > 0$*, let* c,c' *satisfy Assumption 2.1, let* (P,Q) *satisfy Assumption 2.3, and let* (P',Q') *be any marginals with finite second moments. Let* f,g *be the potentials (unique up to translation) associated with* (P,Q,c,ε)*, and let* f',g' *be (any) potentials associated with* (P',Q',c',ε')*. Set* $h := f\oplus g$ *and* $h' := f'\oplus g'$*, let* γ_ε *be as in Theorem 2.5, and let* ϑ_ε *be as in (16). Define*

$$\begin{aligned}
\Delta_* &:= d_{\mathfrak{D}}\big((P,Q,c,\varepsilon),(P',Q',c',\varepsilon')\big) \\
&:= 2L\,\Big(W_1(P,P')^2 + W_1(Q,Q')^2\Big)^{1/2} + \|c-c'\|_{L^\infty(\mathcal{X}\times\mathcal{Y})} + |\varepsilon-\varepsilon'| \qquad (21) \\
\widehat{q}_\varepsilon &:= \inf_{y\in\Omega_Q} Q\Big(B_{\varepsilon/(4L)}(y)\Big) \\
\widehat{\kappa}_\varepsilon &:= \delta_P \min\left\{\Big(\frac{\varepsilon}{4L}\Big)^d, 1\right\} \\
\widehat{\eta}_\varepsilon &:= \min\left\{\frac{\varepsilon\sqrt{\vartheta_\varepsilon}}{2\gamma_\varepsilon},\ \frac{\varepsilon\,\widehat{q}_\varepsilon}{2(1+\gamma_\varepsilon)},\ \frac{\varepsilon\,\widehat{\kappa}_\varepsilon}{2(1+\gamma_\varepsilon)}\right\}. \qquad (22)
\end{aligned}$$

Then $\widehat{q}_\varepsilon, \widehat{\kappa}_\varepsilon, \widehat{\eta}_\varepsilon > 0$. *If*

$$\Delta_* < \widehat{\eta}_\varepsilon, \tag{23}$$

then, writing $a := \int (g' - g)\, dQ$, *we have*

$$\|f - f' - a\|_{L^\infty(\mathcal{X})} \le (1+\gamma_\varepsilon)\widehat{q_\varepsilon}^{-1}\,\Delta_*, \tag{24}$$

$$\|g - g' + a\|_{L^\infty(\mathcal{Y})} \le (1+\gamma_\varepsilon)\widehat{\kappa}_\varepsilon^{-1}\,\Delta_*, \tag{25}$$

and consequently

$$\|h - h'\|_{L^\infty(\mathcal{X}\times\mathcal{Y})} \le (1+\gamma_\varepsilon)\big(\widehat{q_\varepsilon}^{-1} + \widehat{\kappa}_\varepsilon^{-1}\big)\Delta_*. \tag{26}$$

By making the constants uniform over a class $\mathfrak{D}$ of data, we immediately obtain a uniform version of Theorem 4.1.

Corollary 4.2 (Uniform L^∞-stability)**.** *Let* $\mathfrak{D} = \mathfrak{D}(\underline{\varepsilon}, D, L, \underline{\lambda}, \overline{\Lambda}, \underline{\delta}, \underline{q})$ *be as in Definition 3.2. For* $(P, Q, c, \varepsilon) \in \mathfrak{D}$, *let* $h_{(P,Q,c,\varepsilon)} = f \oplus g$, *where* (f, g) *are the associated potentials. Then the map*[1]

$$\mathfrak{D} \ni (P, Q, c, \varepsilon) \longmapsto h_{(P,Q,c,\varepsilon)}$$

is uniformly locally Lipschitz with respect to $d_{\mathfrak{D}}$, *defined in* (21)*. More precisely, let*

$$\begin{aligned}
\overline{\gamma} &:= 16\left(\underline{\delta}^{-1}\max\left(\frac{8L}{\underline{\varepsilon}}, 1\right)^{d}\right)\frac{\overline{\Lambda}^2}{\underline{\lambda}^2}\,\frac{\big(\lceil 8L\,D/\underline{\varepsilon}\rceil\big)^{d+2}}{\underline{q}}\\
\underline{\vartheta} &:= \underline{\delta}\min\left\{\left(\frac{\underline{\varepsilon}}{8L}\right)^{d}, 1\right\}\underline{q}\\
\underline{\widehat{\kappa}} &:= \underline{\delta}\min\left\{\left(\frac{\underline{\varepsilon}}{4L}\right)^{d}, 1\right\}\\
\overline{\eta}_* &:= \min\left\{\frac{\underline{\varepsilon}\sqrt{\underline{\vartheta}}}{2\overline{\gamma}}, \frac{\underline{\varepsilon}\,\underline{q}}{2(1+\overline{\gamma})}, \frac{\underline{\varepsilon}\,\underline{\widehat{\kappa}}}{2(1+\overline{\gamma})}\right\},\\
\overline{C} &:= (1+\overline{\gamma})\big(\underline{q}^{-1} + \underline{\widehat{\kappa}}^{-1}\big).
\end{aligned} \tag{27}$$

Then, for any pair $(P, Q, c, \varepsilon), (P', Q', c', \varepsilon') \in \mathfrak{D}$ *satisfying*

$$d_{\mathfrak{D}}\big((P, Q, c, \varepsilon), (P', Q', c', \varepsilon')\big) < \overline{\eta}_*$$

we have

$$\|h_{(P,Q,c,\varepsilon)} - h_{(P',Q',c',\varepsilon')}\|_{L^\infty(\mathcal{X}\times\mathcal{Y})} \le \overline{C}\, d_{\mathfrak{D}}\big((P, Q, c, \varepsilon), (P', Q', c', \varepsilon')\big). \tag{28}$$

5 Stability of the optimal coupling

We now turn to the stability of the primal optimizers. Throughout this section, we work on the admissible class $\mathfrak{D} = \mathfrak{D}(\underline{\varepsilon}, D, L, \underline{\lambda}, \overline{\Lambda}, \underline{\delta}, \underline{q})$ from Definition 3.2. Given probability measures μ, ν on a measurable space, we write

$$\|\mu - \nu\|_{\mathrm{TV}} := \sup_A |\mu(A) - \nu(A)|$$

[1]If c is bounded on $\mathcal{X}\times\mathcal{Y}$, then this is a map into $L^\infty(\mathcal{X}\times\mathcal{Y})$. In general, the difference $h_{(P,Q,c,\varepsilon)} - h_{(P',Q',c',\varepsilon')}$ in (28) is in $L^\infty(\mathcal{X}\times\mathcal{Y})$, even though the two terms need not be bounded individually.

for the total variation distance. The next theorem provides stability estimates for the optimal coupling as well as its density with respect to the product $P \otimes Q$ of the marginals (where $P \otimes Q$ is itself varying).

Theorem 5.1 (Stability of primal optimizers)**.** *Let $\mathfrak{D} = \mathfrak{D}(\underline{\varepsilon}, D, L, \underline{\lambda}, \overline{\Lambda}, \underline{\delta}, \underline{q})$ be as in Definition 3.2. Let (P, Q, c, ε), $(P', Q', c', \varepsilon') \in \mathfrak{D}$, let h, h' be the corresponding direct sums of potentials, let $\pi \in \Pi(P, Q), \pi' \in \Pi(P', Q')$ be the corresponding optimal couplings, and denote their densities as*

$$\zeta := \frac{1}{\varepsilon}(h - c)_+, \qquad \zeta' := \frac{1}{\varepsilon'}(h' - c')_+.$$

Set

$$\mu := P \otimes Q, \qquad \mu' := P' \otimes Q', \qquad \bar{\mu} := \frac{\mu + \mu'}{2},$$

$$\Delta_W := \Big(W_1(P, P')^2 + W_1(Q, Q')^2\Big)^{1/2},$$

$$\Delta_{\mathrm{TV}} := \|P - P'\|_{\mathrm{TV}} + \|Q - Q'\|_{\mathrm{TV}},$$

and

$$\begin{aligned} \Delta &:= 2L\,\Delta_W + \|c - c'\|_{L^2(\mu)} + |\varepsilon - \varepsilon'|, \\ \Delta' &:= 2L\,\Delta_W + \|c - c'\|_{L^2(\mu')} + |\varepsilon - \varepsilon'|, \\ \bar{\Delta} &:= 2L\,\Delta_W + \|c - c'\|_{L^2(\bar{\mu})} + |\varepsilon - \varepsilon'|. \end{aligned}$$

Let $\overline{\gamma}$ and $\overline{\eta}$ be the constants from Corollary 3.3 and define

$$\Gamma := (\Omega_P \cup \Omega_{P'}) \times (\Omega_Q \cup \Omega_{Q'}), \qquad \Theta := (\Omega_P \times \Omega_Q) \cup (\Omega_{P'} \times \Omega_{Q'}),$$

$$D_* := \sup_{z, z' \in \Theta} \|z - z'\|, \qquad A := 1 + \frac{6\|c\|_{L^\infty(\Gamma)}}{\varepsilon}, \qquad A' := 1 + \frac{6\|c'\|_{L^\infty(\Gamma)}}{\varepsilon'},$$

and

$$\hat{\Delta} := \min\left\{ \frac{\overline{\gamma}\,\bar{\Delta} + \|c - c'\|_{L^2(\bar{\mu})} + A'|\varepsilon - \varepsilon'|}{\varepsilon}, \frac{\overline{\gamma}\,\bar{\Delta} + \|c - c'\|_{L^2(\bar{\mu})} + A|\varepsilon - \varepsilon'|}{\varepsilon'} \right\}.$$

Assume that

$$\max\{\Delta, \Delta'\} < \overline{\eta},$$

then the following hold.

(i) L^2-stability of the density:

$$\|\zeta - \zeta'\|_{L^2(\bar{\mu})} \leq \hat{\Delta}. \tag{29}$$

(ii) TV-stability of the optimal coupling:

$$\|\pi - \pi'\|_{\mathrm{TV}} \leq \frac{\hat{\Delta}}{2} + \frac{A + A'}{2}\,\Delta_{\mathrm{TV}}. \tag{30}$$

(iii) W_1-stability of the optimal coupling:

$$W_1(\pi,\pi') \le \frac{D_* \hat{\Delta}}{2} + \sqrt{2}\left[\frac{A+A'}{2} + \frac{(\sqrt{2}+1)L\,D_*}{4}\Big(\frac{1}{\varepsilon}+\frac{1}{\varepsilon'}\Big)\right]\Delta_W. \tag{31}$$

Complementing the L^2-stability of the density in (29), we have the following L^∞-stability if the perturbation of the cost function is quantified in L^∞.

Remark 5.2 (L^∞-stability of the density)**.** Let $\Delta_* := d_{\mathfrak{D}}\big((P,Q,c,\varepsilon),(P',Q',c',\varepsilon')\big)$ be as in (21) and $\overline{C}$ as in (27). Using the notation of Theorem 5.1, but replacing $\hat{\Delta}$ by

$$\hat{\Delta}_\infty := \min\left\{\frac{\overline{C}\,\Delta_* + \|c-c'\|_{L^\infty(\Gamma)} + A'|\varepsilon-\varepsilon'|}{\varepsilon}, \frac{\overline{C}\,\Delta_* + \|c-c'\|_{L^\infty(\Gamma)} + A|\varepsilon-\varepsilon'|}{\varepsilon'}\right\},$$

we have the following L^∞ bound: if $\Delta_* < \overline{\eta}_*$, where $\overline{\eta}_*$ is as in Corollary 4.2, then

$$\|\zeta-\zeta'\|_{L^\infty(\Gamma)} \le \hat{\Delta}_\infty.$$

The proof is the same as for (29), except that now the application of Corollary 3.3 is replaced by an application of Corollary 4.2.

6 Stability of the optimal support

As the support of the optimal coupling is characterized by (12), we expect that stability of the supports is linked to stability of the potentials. However, the latter alone is not sufficient. This is most evident in the degenerate example $c \equiv 0$, where it is easy to see that the optimal coupling is the product $\pi = P \otimes Q$ of the marginals (for any $\varepsilon > 0$). In particular, the potentials satisfy $f \oplus g \equiv \varepsilon$, independently of (P,Q), showing that control of the potentials alone cannot imply control of the supports. For that reason, the stability result below features a direct control Δ_Ω on the marginal supports on the right-hand side of (33), in addition to the quantities needed to control the potentials.[2]

Given compact sets $\emptyset \ne A, B \subset \mathbb{R}^m$, their Hausdorff distance is denoted

$$d_H(A,B) := \max\Big\{\sup_{z\in A}\operatorname{dist}(z,B), \sup_{z\in B}\operatorname{dist}(z,A)\Big\}.$$

We recall that dist and d_H are computed with respect to the Euclidean norm, i.e., for $z = (x,y) \in \mathbb{R}^d \times \mathbb{R}^d$, we use $\|z\| = (\|x\|^2 + \|y\|^2)^{1/2}$.

To guarantee stability of the optimal supports, the next theorem requires the nondegeneracy condition (32) which will be discussed in detail below. In particular, we will see in Proposition 6.3 that (32) is satisfied with $a = \varepsilon/\operatorname{diam}(\Omega_P)$ for the quadratic cost $c(x,y) = \frac{1}{2}\|x-y\|^2$. On the other hand, Example 6.2 will illustrate that, in general, stability can fail when (32) is not satisfied.

[2]One could, of course, choose a metric such as W_∞ to obtain simultaneous control, but that would yield a very loose bound. E.g., in a case where the marginals vary but their supports are fixed, our control Δ_Ω is vacuous, but W_∞ would give a strong restriction.

Theorem 6.1 (Support stability). *Let $\mathfrak{D} = \mathfrak{D}(\underline{\varepsilon}, D, L, \underline{\lambda}, \overline{\Lambda}, \underline{\delta}, \underline{q})$ be as in Definition 3.2 and $\overline{C}$, $\overline{\eta}_*$ as in Corollary 4.2. For $(P, Q, c, \varepsilon) \in \mathfrak{D}$, let $h_{(P,Q,c,\varepsilon)} = f \oplus g$, where (f, g) are the potentials corresponding to (P, Q, c, ε). Let*

$$(P, Q, c, \varepsilon),\ (P', Q', c', \varepsilon') \in \mathfrak{D}, \qquad \Delta_* := d_{\mathfrak{D}}((P, Q, c, \varepsilon), (P', Q', c', \varepsilon')).$$

Set

$$\Delta_\Omega := \Big(d_H(\Omega_P, \Omega_{P'})^2 + d_H(\Omega_Q, \Omega_{Q'})^2\Big)^{1/2},$$
$$h := h_{(P,Q,c,\varepsilon)}, \qquad h' := h_{(P',Q',c',\varepsilon')},$$
$$\sigma := h - c, \qquad \sigma' := h' - c', \qquad \pi := \pi_{(P,Q,c,\varepsilon)}, \qquad \pi' := \pi_{(P',Q',c',\varepsilon')},$$
$$\Sigma := \operatorname{supp}(\pi), \qquad \Sigma' := \operatorname{supp}(\pi'), \qquad \delta_* := \overline{C}\,\Delta_* + \|c - c'\|_{L^\infty(\mathcal{X}\times\mathcal{Y})}.$$

Assume that there exists $a > 0$ such that

$$\sigma(z) \le -a \operatorname{dist}(z, \Sigma) \qquad \text{for all } z \in \Omega_{P\otimes Q} \setminus \Sigma, \tag{32}$$

and that the same bound holds for $(\sigma', \Sigma', P', Q')$, with the same constant a. If $\Delta_* < \overline{\eta}_*$, the Hausdorff distance of the optimal supports satisfies

$$d_H(\Sigma, \Sigma') \le \left(1 + \frac{(\sqrt{2}+1)L}{a}\right)\Delta_\Omega + \frac{\delta_*}{a}. \tag{33}$$

Under the nondegeneracy assumption (32), the value $\sigma(z)$ is comparable to $-\operatorname{dist}(z, \Sigma)$ outside the support $\Sigma = \overline{\{z \in \Omega_{P\otimes Q} : \sigma(z) > 0\}}$:

$$-\operatorname{Lip}(\sigma)\operatorname{dist}(z, \Sigma) \le \sigma(z) \le -a\operatorname{dist}(z, \Sigma) \qquad \text{for all } z \in \Omega_{P\otimes Q} \setminus \Sigma,$$

as $\sigma \ge 0$ on Σ. This linear detachment prevents instability. In general, if σ is nearly flat at the boundary of the support, a small perturbation of σ can cause a large change of the support. Indeed, the following example, where the marginals vary without changing support and c, ε are fixed, shows that support stability for the optimal coupling can fail if the nondegeneracy condition (32) is removed.

Example 6.2 (Failure of support stability). Let

$$\mathcal{X} = [0, 1], \qquad \mathcal{Y} = \{0, 1\}, \qquad \varepsilon = 1.$$

Define

$$u(x) = \begin{cases} 0, & 0 \le x \le \frac{1}{4}, \\ \frac{32}{5}\left(x - \frac{1}{4}\right), & \frac{1}{4} < x < \frac{1}{2}, \\ \frac{8}{5}, & \frac{1}{2} \le x \le 1. \end{cases}$$

Then u is Lipschitz, $0 \le u \le 8/5 < 2$, and $\int_0^1 u(x)\,dx = 1$. We consider the cost function

$$c(x, 0) = u(x), \qquad c(x, 1) = 2 - u(x),$$

and note that c can be extended to a Lipschitz function on $\mathbb{R} \times \mathbb{R}$. Moreover, we define a family of marginals indexed by $0 \le \eta < 1$ as follows. The second marginal is fixed to be

$$Q^\eta = Q^0 = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1,$$

whereas the first marginal P^η is defined via its density

$$p_\eta(x) := \frac{dP^\eta}{dx}(x) = \begin{cases} 1+\eta, & 0 \le x \le \frac{1}{4}, \\ 1 - \frac{\eta}{3}, & \frac{1}{4} < x \le 1. \end{cases}$$

Then P^η has density bounded above and below on the fixed convex compact support $[0,1]$, P^0 is the uniform measure on $[0,1]$, and $P^\eta \to P^0$ in total variation.

For the marginals P^0, Q^0, the sum of the potentials is given by

$$h^0(x,0) = h^0(x,1) = 2.$$

Indeed, this gives rise to

$$\sigma^0(x,0) = h^0(x,0) - c(x,0) = 2 - u(x), \qquad \sigma^0(x,1) = h^0(x,1) - c(x,1) = u(x).$$

In view of $\frac{1}{2}(\sigma^0(x,0) + \sigma^0(x,1)) = 1$ and $\int_0^1 \sigma^0(x,1)\,dx = \int_0^1 u(x)\,dx = 1$, the plan with density $(\sigma^0)_+$ with respect to $P^0 \otimes Q^0$ satisfies the marginal constraints and is optimal. Its support is

$$\Sigma^0 = \operatorname{supp}(\pi^0) = \big([0,1] \times \{0\}\big) \cup \big([\tfrac{1}{4}, 1] \times \{1\}\big).$$

We note that the nondegeneracy condition (32) fails: as $\sigma^0(0,1) = 0$ and $\operatorname{dist}\big((0,1), \Sigma^0\big) = \frac{1}{4}$, there is no $a > 0$ such that

$$\sigma^0(z) \le -a \operatorname{dist}(z, \Sigma^0) \qquad \text{for all } z \in \Omega_{P^0 \otimes Q^0} \setminus \Sigma^0.$$

Now consider P^η, Q^η. In view of $\int_0^1 u(x) p_\eta(x)\,dx = 1 - \frac{\eta}{3}$, putting

$$h^\eta(x,0) = 2 - \delta_\eta, \qquad h^\eta(x,1) = 2 + \delta_\eta, \qquad \delta_\eta = \frac{\eta}{3}$$

gives

$$\sigma^\eta(x,0) = 2 - u(x) - \delta_\eta, \qquad \sigma^\eta(x,1) = u(x) + \delta_\eta.$$

For $0 < \eta < 1$, both quantities are positive on the whole of $[0,1]$. Moreover, $\frac{1}{2}\big(\sigma^\eta(x,0) + \sigma^\eta(x,1)\big) = 1$ and $\int_0^1 \sigma^\eta(x,1)\,dP^\eta(x) = \int_0^1 \big(u(x) + \delta_\eta\big) p_\eta(x)\,dx = 1$. Thus the optimal plan for $(P^\eta, Q^\eta, c, 1)$ has density $(\sigma^\eta)_+$ with respect to $P^\eta \otimes Q^\eta$, and therefore

$$\Sigma^\eta = \operatorname{supp}(\pi^\eta) = [0,1] \times \{0,1\}.$$

Consequently, Σ^η does not converge to Σ^0:

$$d_H(\Sigma^\eta, \Sigma^0) = \frac{1}{4} \qquad \text{for every } 0 < \eta < 1.$$

On the other hand, $P^\eta \to P^0$ and $Q^\eta = Q^0$, and the marginal supports $\Omega_{P^\eta} = \Omega_{P^0} = [0,1]$ and $\Omega_{Q^\eta} = \Omega_{Q^0} = \{0,1\}$ do not move, so that all conditions of Theorem 6.1 except (32) are satisfied for sufficiently small $\eta > 0$.

To apply Theorem 6.1, we need to verify the nondegeneracy condition (32). The next proposition shows that for the key example of quadratic cost, the condition holds as soon as one marginal support is convex and compact.

Proposition 6.3 (Nondegeneracy)**.** *Let (P,Q,c,ε) satisfy the hypotheses of Proposition 2.2 and assume that Ω_P is convex with $D_P := \operatorname{diam}(\Omega_P) < \infty$. Set*

$$h := h_{(P,Q,c,\varepsilon)}, \qquad \sigma := h - c, \qquad \pi := \pi_{(P,Q,c,\varepsilon)}, \qquad \Sigma := \operatorname{supp}(\pi).$$

Assume that for each $y \in \Omega_Q$, the map

$$x \mapsto \sigma(x,y) \qquad \text{is concave on } \Omega_P. \tag{34}$$

If $D_P = 0$, then $\Omega_{P\otimes Q} \setminus \Sigma = \emptyset$; if $D_P > 0$, then

$$\sigma(z) \leq -\frac{\varepsilon}{D_P} \operatorname{dist}(z,\Sigma) \qquad \text{for all } z \in \Omega_{P\otimes Q} \setminus \Sigma, \tag{35}$$

in particular, the nondegeneracy condition (32) holds with $a = \varepsilon/D_P$.

The assumption (34) holds in particular for the quadratic cost $c(x,y) = \frac{1}{2}\|x-y\|^2$.

Combining Theorem 6.1 and Proposition 6.3 yields the following corollary, where we also specialize to $\varepsilon = \varepsilon'$ to simplify the expressions.

Corollary 6.4 (Support stability for quadratic cost)**.** *Let $\mathfrak{D} = \mathfrak{D}(\varepsilon, D, L, \underline{\lambda}, \overline{\Lambda}, \underline{\delta}, \underline{q})$ be as in Definition 3.2 and $\overline{C}$, $\overline{\eta}_*$ as in Corollary 4.2. Let $\mathcal{X}, \mathcal{Y}$ be such that the quadratic cost $c(x,y) := \frac{1}{2}\|x-y\|^2$ is L-Lipschitz on $\mathcal{X}\times\mathcal{Y}$. Consider marginals (P,Q) and (P',Q') such that (P,Q,c,ε) and (P',Q',c,ε) belong to $\mathfrak{D}$. Set*

$$\Delta_W := \Big(W_1(P,P')^2 + W_1(Q,Q')^2\Big)^{1/2}, \qquad \Delta_\Omega := \Big(d_H(\Omega_P,\Omega_{P'})^2 + d_H(\Omega_Q,\Omega_{Q'})^2\Big)^{1/2}.$$

If $2L\Delta_W < \overline{\eta}_*$, the corresponding optimal supports Σ, Σ' satisfy

$$d_H(\Sigma,\Sigma') \leq \left(1 + \frac{(\sqrt{2}+1)LD}{\varepsilon}\right)\Delta_\Omega + \frac{2LD\,\overline{C}}{\varepsilon}\,\Delta_W.$$

A Proofs

A.1 Proofs for Section 2

The basic structural properties of QOT in Proposition 2.2 are known or follow by known arguments. For completeness, the following gives details and references.

Proof of Proposition 2.2. By the McShane extension theorem, c admits an L-Lipschitz extension to $\mathbb{R}^d\times\mathbb{R}^d$, still denoted by c. As this extension is of linear growth and P,Q have finite second moments, we are in the setting of [28] with $c \in L^2(P\otimes Q)$. In particular, [28, Section 2] shows (i), (ii) with $(f,g) \in L^1(P)\times L^1(Q)$, and (11). It also yields the existence of L-Lipschitz versions of (f,g) on (Ω_P,Ω_Q), implying that $(f,g) \in L^2(P)\times L^2(Q)$. Moreover, it shows the validity of the first-order condition (9) for all $x \in \Omega_P$ and $y \in \Omega_Q$, respectively.

We can then uniquely extend f to $\mathcal{X}$ by enforcing the first equation of (9), using the monotonicity of $a \mapsto (a + g(y) - c(x,y))_+$ and the intermediate value theorem as in the argument stated below [28, Eq. (2.11)]. Similarly for g on $\mathcal{Y}$. As the proof of the L-Lipschitz property in [28, Section 2] is based solely on the first-order condition, the constructed extensions are also L-Lipschitz.

If one of the two marginal supports is connected, the proof [19, Lemma 3.1] shows that (f,g) are unique on (Ω_P, Ω_Q) up to translation. Uniqueness on $\mathcal{X}$ and $\mathcal{Y}$ then follows by the aforementioned uniqueness of the extensions satisfying (9). The cited lemma is stated for $c(x,y) = \frac{1}{2}\|x-y\|^2$ on a compact domain, but the proof remains valid without changes for uniformly continuous costs on $\mathcal{X} \times \mathcal{Y}$.

Regarding (iv), the oscillation bounds are stated in [28, Lemma 2.5]; again, the proof is based solely on the first-order condition, hence the bounds hold without exceptional set for the chosen versions. To infer (10), set

$$C := \|c\|_{L^\infty(\mathcal{X}\times\mathcal{Y})}, \qquad M := \sup_{\mathcal{X}\times\mathcal{Y}} (f \oplus g), \qquad m := \inf_{\mathcal{X}\times\mathcal{Y}} (f \oplus g).$$

Since

$$\operatorname*{osc}_{\mathcal{X}\times\mathcal{Y}} (f \oplus g) \le \operatorname*{osc}_{\mathcal{X}}(f) + \operatorname*{osc}_{\mathcal{Y}}(g) \le 2 \operatorname*{osc}_{\mathcal{X}\times\mathcal{Y}}(c) \le 4C,$$

we have

$$M - m \le 4C.$$

To prove the upper bound, note that for all $(x,y) \in \mathcal{X} \times \mathcal{Y}$,

$$f(x) + g(y) - c(x,y) \ge M - 4C - C = M - 5C.$$

Taking positive parts, integrating with respect to $P \otimes Q$, using $\Omega_P \times \Omega_Q \subset \mathcal{X} \times \mathcal{Y}$, and (9), we obtain

$$\varepsilon = \int (f \oplus g - c)_+ \, d(P \otimes Q) \ge (M - 5C)_+,$$

which implies $M \le 5C + \varepsilon$. For the lower bound, we similarly have

$$f(x) + g(y) - c(x,y) \le m + 4C + C = m + 5C$$

for all $(x,y) \in \mathcal{X} \times \mathcal{Y}$. Proceeding as above yields $0 < \varepsilon \le (m + 5C)_+ = m + 5C$, that is, $m \ge -5C + \varepsilon$.

Finally, the support (12) readily follows from (11) and the continuity of f, g, c. □

A.2 Proofs for Section 3

For ease of reference, we record the following fact about decompositions of $w \in \mathcal{H}$.

Remark A.1. For $w \in \mathcal{H}$, the decomposition into $w = u \oplus v$ is not unique. If $w = u \oplus v$ is any decomposition with $(u,v) \in L^2(P) \times L^2(Q)$, then

$$\|w\|^2_{L^2(P\otimes Q)} = \|u\|^2_{L^2(P)} + \|v\|^2_{L^2(Q)} + 2\Big(\int u \, dP\Big)\Big(\int v \, dQ\Big),$$

and in particular

$$\|u\|_{L^2(P)}^2 + \|v\|_{L^2(Q)}^2 = \|w\|_{L^2(P\otimes Q)}^2 \qquad \text{if} \qquad \int u\, dP = 0 \text{ or } \int v\, dQ = 0. \tag{36}$$

Writing $\bar{w} := \int w\, d(P\otimes Q)$, a balanced choice of decomposition is

$$u(x) := \int w(x,y)\, dQ(y) - \frac{\bar{w}}{2}, \qquad v(y) := \int w(x,y)\, dP(x) - \frac{\bar{w}}{2}. \tag{37}$$

Then,

$$\int u\, dP = \int v\, dQ = \frac{\bar{w}}{2}, \tag{38}$$

and in particular

$$\|u\|_{L^2(P)}^2 + \|v\|_{L^2(Q)}^2 \leq \|u\|_{L^2(P)}^2 + \|v\|_{L^2(Q)}^2 + \frac{\bar{w}^2}{2} = \|w\|_{L^2(P\otimes Q)}^2. \tag{39}$$

Next, we prove the basic L^2-estimate for perturbations of the data.

Proof of Theorem 3.1. Since Ω_P and Ω_Q are compact, f' and g' are bounded by Proposition 2.2, and Theorem 2.5 applies to h'. Moreover, by Proposition 2.2,

$$\begin{cases} \int (f'(x) + g'(y) - c'(x,y))_+\, dQ'(y) = \varepsilon' & \text{for all } x \in \mathcal{X}, \\ \int (f'(x) + g'(y) - c'(x,y))_+\, dP'(x) = \varepsilon' & \text{for all } y \in \mathcal{Y}. \end{cases} \tag{40}$$

Step 1: We show $\|\mathrm{D}\Phi(h')\|_{L^2(P\otimes Q)} \leq \Delta/\varepsilon$. Let $w \in \mathcal{H}$ satisfy $\|w\|_{L^2(P\otimes Q)} \leq 1$. Let (u,v) be the balanced decomposition $w = u \oplus v$ given by (37), and write $\bar{w} := \int w\, d(P\otimes Q)$. Recalling (13) and using Fubini's theorem, we have

$$\begin{aligned} \langle \mathrm{D}\Phi(h'), w\rangle_{\mathcal{H}} = &\int_{\Omega_P} u(x)\Big(1 - \frac{1}{\varepsilon}\int (h'(x,y) - c(x,y))_+\, dQ(y)\Big)\, dP(x) \\ &+ \int_{\Omega_Q} v(y)\Big(1 - \frac{1}{\varepsilon}\int (h'(x,y) - c(x,y))_+\, dP(x)\Big)\, dQ(y). \end{aligned} \tag{41}$$

For $x \in \Omega_P$, define

$$\xi'_x(y) := (h'(x,y) - c'(x,y))_+, \qquad y \in \mathcal{Y}.$$

Then (40) implies

$$\int \xi'_x(y)\, dQ'(y) = \varepsilon' \qquad \text{for all } x \in \Omega_P,$$

and hence

$$\begin{aligned} 1 - \frac{1}{\varepsilon}\int (h'(x,y) - c(x,y))_+\, dQ(y) = &\frac{\varepsilon - \varepsilon'}{\varepsilon} + \frac{1}{\varepsilon}\int \xi'_x(y)\, d(Q' - Q)(y) \\ &+ \frac{1}{\varepsilon}\int \big[(h'(x,y) - c'(x,y))_+ - (h'(x,y) - c(x,y))_+\big]\, dQ(y). \end{aligned} \tag{42}$$

Similarly, for $y \in \Omega_Q$, defining $\eta'_y(x) := (h'(x,y) - c'(x,y))_+$, we have

$$1 - \frac{1}{\varepsilon}\int (h'(x,y) - c(x,y))_+ \, dP(x) = \frac{\varepsilon - \varepsilon'}{\varepsilon} + \frac{1}{\varepsilon}\int \eta'_y(x)\, d(P' - P)(x) \\ + \frac{1}{\varepsilon}\int \big[(h'(x,y) - c'(x,y))_+ - (h'(x,y) - c(x,y))_+\big]\, dP(x). \quad (43)$$

Substituting (42) and (43) into (41), and using (38), we obtain

$$\begin{aligned}
&\langle \mathrm{D}\Phi(h'), w\rangle_{\mathcal{H}} \\
&= \frac{\varepsilon - \varepsilon'}{\varepsilon}\,\bar{w} && (44)\\
&\quad + \frac{1}{\varepsilon}\int_{\Omega_P \times \Omega_Q} w(x,y)\Big((h'(x,y) - c'(x,y))_+ - (h'(x,y) - c(x,y))_+\Big)\, d(P \otimes Q)(x,y) && (45)\\
&\quad + \frac{1}{\varepsilon}\int_{\Omega_P} u(x)\Big(\int \xi'_x(y)\, d(Q' - Q)(y)\Big)\, dP(x) && (46)\\
&\quad + \frac{1}{\varepsilon}\int_{\Omega_Q} v(y)\Big(\int \eta'_y(x)\, d(P' - P)(x)\Big)\, dQ(y). && (47)
\end{aligned}$$

Next, we estimate each term. For (44), noting $|\bar{w}| \le \|w\|_{L^2(P\otimes Q)} \le 1$, we have

$$\left|\frac{\varepsilon - \varepsilon'}{\varepsilon}\bar{w}\right| \le \frac{|\varepsilon - \varepsilon'|}{\varepsilon}\|w\|_{L^2(P\otimes Q)} \le \frac{|\varepsilon - \varepsilon'|}{\varepsilon}. \quad (48)$$

For (45), using that $t \mapsto t_+$ is 1-Lipschitz and the Cauchy–Schwarz inequality,

$$|(45)| \le \frac{\|c - c'\|_{L^2(P\otimes Q)}}{\varepsilon}\|w\|_{L^2(P\otimes Q)} \le \frac{\|c - c'\|_{L^2(P\otimes Q)}}{\varepsilon}. \quad (49)$$

In (46), the function ξ'_x is $2L$-Lipschitz, uniformly in x, because g' and c' are L-Lipschitz and $t \mapsto t_+$ is 1-Lipschitz. Thus, by Kantorovich–Rubinstein duality,

$$\left|\int \xi'_x(y)\, d(Q' - Q)(y)\right| \le 2L\, W_1(Q, Q') \qquad \text{for all } x \in \Omega_P,$$

showing

$$|(46)| \le \frac{2L}{\varepsilon} W_1(Q, Q')\, \|u\|_{L^2(P)}. \quad (50)$$

Similarly,

$$|(47)| \le \frac{2L}{\varepsilon} W_1(P, P')\, \|v\|_{L^2(Q)}. \quad (51)$$

Write $\Delta_W := [W_1(P,P')^2 + W_1(Q,Q')^2]^{1/2}$. Combining (50), (51), Cauchy–Schwarz, and (39),

$$\begin{aligned}
|(46) + (47)| &\le \frac{2L}{\varepsilon}\Big(W_1(Q,Q')\|u\|_{L^2(P)} + W_1(P,P')\|v\|_{L^2(Q)}\Big)\\
&\le \frac{2L}{\varepsilon}\Delta_W\Big(\|u\|^2_{L^2(P)} + \|v\|^2_{L^2(Q)}\Big)^{1/2}\\
&\le \frac{2L}{\varepsilon}\Delta_W\, \|w\|_{L^2(P\otimes Q)} \le \frac{2L}{\varepsilon}\Delta_W. && (52)
\end{aligned}$$

Collecting (48), (49), and (52), we obtain

$$|\langle \mathrm{D}\Phi(h'), w\rangle_{\mathcal{H}}| \le \frac{1}{\varepsilon}\big(2L\,\Delta_W + \|c-c'\|_{L^2(P\otimes Q)} + |\varepsilon-\varepsilon'|\big) = \frac{\Delta}{\varepsilon}.$$

Since w was arbitrary in the unit ball of $\mathcal{H}$, (14) yields

$$\|\mathrm{D}\Phi(h')\|_{L^2(P\otimes Q)} \le \frac{\Delta}{\varepsilon}. \tag{53}$$

Step 2: Proof of (i)–(iii). Applying the error bound from Theorem 2.5 to h', and using (53), we obtain (15), completing the proof of (i). To show (ii), let $\delta > 0$ and assume the left-hand side of (17) holds. Set

$$\theta := h - h', \qquad m := \|\theta\|_{L^\infty(\Omega_P\times\Omega_Q)},$$

and suppose for contradiction that $m \ge \delta$. The function θ is continuous on the compact set $\Omega_P \times \Omega_Q$, so there exists $(x_0, y_0) \in \Omega_P \times \Omega_Q$ with $|\theta(x_0, y_0)| = m$. Replacing θ by $-\theta$ if necessary, we may assume that

$$\theta(x_0, y_0) = m.$$

Let

$$\varrho := \frac{\delta}{8L}.$$

For all $x \in B_\varrho(x_0) \cap \Omega_P$ and $y \in B_\varrho(y_0) \cap \Omega_Q$, the L-Lipschitz continuity of f, f', g, g' yields

$$\begin{aligned}\theta(x,y) &\ge \theta(x_0,y_0) - |f(x)-f(x_0)| - |f'(x)-f'(x_0)| - |g(y)-g(y_0)| - |g'(y)-g'(y_0)| \\ &\ge m - 2L\|x-x_0\| - 2L\|y-y_0\| \ge m - \frac{\delta}{2} \ge \frac{m}{2}.\end{aligned}$$

Therefore,

$$\begin{aligned}\|\theta\|^2_{L^2(P\otimes Q)} &\ge \frac{m^2}{4}(P\otimes Q)\big((B_\varrho(x_0)\cap\Omega_P)\times(B_\varrho(y_0)\cap\Omega_Q)\big) \\ &= \frac{m^2}{4}P\big(B_\varrho(x_0)\big)Q\big(B_\varrho(y_0)\big) \\ &\ge \frac{m^2}{4}\delta_P \min(\varrho^d, 1)\inf_{y\in\Omega_Q} Q\big(B_\varrho(y)\big) = \frac{\vartheta_\delta}{4}m^2,\end{aligned}$$

or equivalently,

$$m \le \frac{2}{\sqrt{\vartheta_\delta}}\,\|\theta\|_{L^2(P\otimes Q)}. \tag{54}$$

On the other hand, (15) gives

$$\|\theta\|_{L^2(P\otimes Q)} \le \frac{\gamma_\varepsilon}{\varepsilon}\max(m,\varepsilon)\,\Delta. \tag{55}$$

Moreover, distinguishing the cases $\delta \ge \varepsilon$ and $\delta \le \varepsilon$, we see that $m \ge \delta$ implies

$$\max(m,\varepsilon) \le \frac{\varepsilon}{\min(\delta,\varepsilon)}\,m. \tag{56}$$

Combining (54), (55), and (56), we obtain

$$\|\theta\|_{L^2(P\otimes Q)} \leq \frac{2\gamma_\varepsilon}{\min(\delta,\varepsilon)\sqrt{\vartheta_\delta}}\,\Delta\,\|\theta\|_{L^2(P\otimes Q)}.$$

As $\|\theta\|_{L^2(P\otimes Q)} > 0$, this contradicts the left-hand side of (17). Therefore, $m < \delta$, completing the proof of (ii). Finally, (iii) follows by taking $\delta = \varepsilon$ in (ii) and applying (i). $\square$

We easily derive the symmetric L^2-stability statement.

Proof of Corollary 3.3. Applying Theorem 3.1(iii) to (P, Q, c, ε) and $(P', Q', c', \varepsilon')$ yields (18) and (19), respectively. It follows that

$$\|h - h'\|^2_{L^2(\bar{\mu})} \leq \frac{\overline{\gamma}^2}{2}\big(\Delta^2 + (\Delta')^2\big).$$

It is elementary to show that $\frac{1}{2}\big(\Delta^2 + (\Delta')^2\big) \leq \bar{\Delta}^2$, so that (20) follows. $\square$

A.3 Proofs for Section 4

We can now upgrade the L^2-estimate to an L^∞-estimate for the potentials.

Proof of Theorem 4.1. The positivity of $\widehat{q}_\varepsilon, \vartheta_\varepsilon$ is stated in Remark 2.4, and $\widehat{\kappa}_\varepsilon > 0$ is immediate. Hence, $\widehat{\eta}_\varepsilon > 0$. Again, by Proposition 2.2,

$$\begin{cases} \displaystyle\int (f'(x) + g'(y) - c'(x,y))_+ \, dQ'(y) = \varepsilon' & \text{for all } x \in \mathcal{X}, \\ \displaystyle\int (f'(x) + g'(y) - c'(x,y))_+ \, dP'(x) = \varepsilon' & \text{for all } y \in \mathcal{Y}. \end{cases} \tag{57}$$

Clearly, (26) follows from (24) and (25). Steps 1–3 below prove (24), and then Step 4 explains how a variant of those arguments proves (25).

Step 1: L^2-control. As $\|c - c'\|_{L^2(P\otimes Q)} \leq \|c - c'\|_{L^\infty(\mathcal{X}\times\mathcal{Y})}$, the condition (23) implies that the quantity Δ of Theorem 3.1 satisfies $\Delta \leq \Delta_* < \frac{\varepsilon\sqrt{\vartheta_\varepsilon}}{2\gamma_\varepsilon}$, and therefore Theorem 3.1 yields

$$\|h - h'\|_{L^2(P\otimes Q)} \leq \gamma_\varepsilon\,\Delta \leq \gamma_\varepsilon\,\Delta_*. \tag{58}$$

Next, fix the additive constant in (f', g'). By changing (f', g') to $(f' + a, g' - a)$ for $a := \int (g' - g)\, dQ$, we may assume without loss of generality that

$$\int (g' - g)\, dQ = 0.$$

Set

$$u := f - f', \qquad v := g - g'.$$

Then $\int v \, dQ = 0$, and so (36) yields

$$\|h - h'\|^2_{L^2(P\otimes Q)} = \|u\|^2_{L^2(P)} + \|v\|^2_{L^2(Q)}.$$

As a consequence, (58) implies

$$\|u\|_{L^2(P)} \le \gamma_\varepsilon \, \Delta_*, \qquad \|v\|_{L^2(Q)} \le \gamma_\varepsilon \, \Delta_*. \tag{59}$$

Step 2: Slope estimate. Fix $x \in \mathcal{X}$, and define

$$F_x(a) := \int (a + g(y) - c(x,y))_+ \, dQ(y), \qquad a \in \mathbb{R}.$$

By the first identity in the first-order condition (9),

$$F_x(f(x)) = \varepsilon.$$

Since $y \mapsto (f(x) + g(y) - c(x,y))_+$ is continuous on the compact set Ω_Q, there exists

$$y(x) \in \operatorname*{argmax}_{y \in \Omega_Q} (f(x) + g(y) - c(x,y))_+.$$

Using (9) once more, we have

$$(f(x) + g(y(x)) - c(x, y(x)))_+ \ge \int (f(x) + g(y) - c(x,y))_+ \, dQ(y) = \varepsilon,$$

and hence

$$f(x) + g(y(x)) - c(x, y(x)) \ge \varepsilon.$$

Now let $a, b \in [f(x) - \varepsilon/2, f(x) + \varepsilon/2]$ with $a \le b$, and let $y \in B_{\varepsilon/(4L)}(y(x)) \cap \Omega_Q$. Since the map $y \mapsto g(y) - c(x,y)$ is $2L$-Lipschitz, we obtain

$$a + g(y) - c(x,y) \ge f(x) - \frac{\varepsilon}{2} + g(y(x)) - c(x,y(x)) - 2L\|y - y(x)\| \ge \varepsilon - \frac{\varepsilon}{2} - \frac{\varepsilon}{2} = 0.$$

Therefore,

$$(b + g(y) - c(x,y))_+ - (a + g(y) - c(x,y))_+ = b - a$$

for all such y. As the left-hand side is clearly nonnegative for any $y \in \mathcal{Y}$, we deduce that

$$(b + g(y) - c(x,y))_+ - (a + g(y) - c(x,y))_+ \ge (b-a)\mathbf{1}_{B_{\varepsilon/(4L)}(y(x)) \cap \Omega_Q}(y)$$

for all $y \in \mathcal{Y}$. Integrating against $Q(dy)$ and recalling $Q(B_{\varepsilon/(4L)}(y(x))) \ge \widehat{q}_\varepsilon$, we conclude

$$F_x(b) - F_x(a) \ge \widehat{q}_\varepsilon \, (b - a) \qquad \text{for all } a \le b \text{ in } [f(x) - \varepsilon/2, f(x) + \varepsilon/2]. \tag{60}$$

We claim that this implies

$$|F_x(a) - \varepsilon| \ge \widehat{q}_\varepsilon |a - f(x)| \qquad \text{whenever } |a - f(x)| \le \varepsilon/2. \tag{61}$$

Indeed, if $a \le f(x)$, then (60) with $(a,b) = (a, f(x))$ yields

$$\varepsilon - F_x(a) = F_x(f(x)) - F_x(a) \ge \widehat{q}_\varepsilon(f(x) - a),$$

whereas if $a \ge f(x)$, then (60) with $(a,b) = (f(x), a)$ yields

$$F_x(a) - \varepsilon = F_x(a) - F_x(f(x)) \ge \widehat{q}_\varepsilon(a - f(x)).$$

Step 3: The bound for $f - f'$. Fix $x \in \mathcal{X}$. The first identity in (57) yields

$$\begin{aligned}|F_x(f'(x)) - \varepsilon| \le &\left|\int \Big[(f'(x) + g(y) - c(x,y))_+ - (f'(x) + g'(y) - c(x,y))_+\Big]\, dQ(y)\right| \\ &+ \left|\int \Big[(f'(x) + g'(y) - c(x,y))_+ - (f'(x) + g'(y) - c'(x,y))_+\Big]\, dQ(y)\right| \\ &+ \left|\int (f'(x) + g'(y) - c'(x,y))_+ \, d(Q - Q')(y)\right| + |\varepsilon - \varepsilon'|. \end{aligned} \tag{62}$$

Since $t \mapsto t_+$ is 1-Lipschitz, the first term in (62) is bounded by

$$\int |g(y) - g'(y)|\, dQ(y) \le \|v\|_{L^2(Q)}.$$

The second term is bounded by

$$\int |c(x,y) - c'(x,y)|\, dQ(y) \le \|c - c'\|_{L^\infty(\mathcal{X}\times\mathcal{Y})}$$

as $x \in \mathcal{X}$ and $\Omega_Q \subset \mathcal{Y}$. Finally,

$$y \mapsto (f'(x) + g'(y) - c'(x,y))_+$$

is $2L$-Lipschitz on $\mathcal{Y}$, so the third term is bounded by $2L\, W_1(Q, Q')$. Using (59), we obtain

$$|F_x(f'(x)) - \varepsilon| \le (1 + \gamma_\varepsilon)\Delta_* < \frac{\varepsilon\, \widehat{q}_\varepsilon}{2}, \tag{63}$$

where the last inequality is due to (23) and (22).

We claim that (63) implies $|f'(x) - f(x)| < \varepsilon/2$. Indeed, if $f'(x) \ge f(x) + \varepsilon/2$, then the monotonicity of F_x, together with (60), yields

$$F_x(f'(x)) - \varepsilon \ge F_x(f(x) + \varepsilon/2) - F_x(f(x)) \ge \frac{\varepsilon\, \widehat{q}_\varepsilon}{2},$$

contradicting (63). If $f'(x) \le f(x) - \varepsilon/2$, we instead have

$$\varepsilon - F_x(f'(x)) \ge F_x(f(x)) - F_x(f(x) - \varepsilon/2) \ge \frac{\varepsilon\, \widehat{q}_\varepsilon}{2},$$

again contradicting (63). Hence

$$|f'(x) - f(x)| < \frac{\varepsilon}{2}.$$

Now (61) applies with $a = f'(x)$, and using also the first inequality of (63), we have

$$\widehat{q}_\varepsilon\, |f'(x) - f(x)| \le |F_x(f'(x)) - \varepsilon| \le (1 + \gamma_\varepsilon)\Delta_*.$$

Since $x \in \mathcal{X}$ was arbitrary, we conclude that

$$\|u\|_{L^\infty(\mathcal{X})} \le (1 + \gamma_\varepsilon)\widehat{q}_\varepsilon^{-1}\, \Delta_*,$$

completing the proof of (24).

Step 4: The bound for $g - g'$. Fix $y \in \mathcal{Y}$, and define

$$G_y(b) := \int (f(x) + b - c(x,y))_+ \, dP(x), \qquad b \in \mathbb{R}.$$

Arguing as in Step 2, but now integrating against $P(dx)$ and using the constant $\widehat{\kappa}_\varepsilon$, we obtain

$$G_y(b) - G_y(a) \geq \widehat{\kappa}_\varepsilon \, (b - a) \qquad \text{for all } a \leq b \text{ in } [g(y) - \varepsilon/2, g(y) + \varepsilon/2],$$

and therefore

$$|G_y(b) - \varepsilon| \geq \widehat{\kappa}_\varepsilon |b - g(y)| \qquad \text{whenever } |b - g(y)| \leq \varepsilon/2. \tag{64}$$

Arguing as in Step 3, we further see that

$$|G_y(g'(y)) - \varepsilon| \leq (1 + \gamma_\varepsilon)\Delta_* < \frac{\varepsilon \, \widehat{\kappa}_\varepsilon}{2},$$

which is analogous to (63) but has $\widehat{\kappa}_\varepsilon$ on the right-hand side. As before, this implies first that $|g'(y) - g(y)| < \frac{\varepsilon}{2}$, and then, by (64),

$$\widehat{\kappa}_\varepsilon \, |g'(y) - g(y)| \leq |G_y(g'(y)) - \varepsilon| \leq (1 + \gamma_\varepsilon)\Delta_*.$$

Since $y \in \mathcal{Y}$ was arbitrary, we conclude that

$$\|v\|_{L^\infty(\mathcal{Y})} \leq (1 + \gamma_\varepsilon)\widehat{\kappa}_\varepsilon^{-1} \, \Delta_*,$$

proving (25). □

Finally, we make the constants uniform over the admissible class.

Proof of Corollary 4.2. This follows directly from Theorem 4.1 and the definitions of the constants in Corollary 4.2. □

A.4 Proofs for Section 5

Next, we transfer the stability of potentials to stability of the primal optimizers.

Proof of Theorem 5.1. By Proposition 2.2(v),

$$d\pi = \zeta \, d\mu, \qquad d\pi' = \zeta' \, d\mu'. \tag{65}$$

Step 1: Stability of the density. By Corollary 3.3,

$$\|h - h'\|_{L^2(\bar{\mu})} \leq \overline{\gamma} \, \bar{\Delta}. \tag{66}$$

Moreover, Proposition 2.2(iv), applied to (P, Q, c, ε) and $(P', Q', c', \varepsilon')$ with localization sets $\Omega_P \cup \Omega_{P'}$ and $\Omega_Q \cup \Omega_{Q'}$, yields

$$-5\|c\|_{L^\infty(\Gamma)} + \varepsilon \leq h \leq 5\|c\|_{L^\infty(\Gamma)} + \varepsilon, \qquad -5\|c'\|_{L^\infty(\Gamma)} + \varepsilon' \leq h' \leq 5\|c'\|_{L^\infty(\Gamma)} + \varepsilon'$$

on Γ. Hence

$$0 \leq (h - c)_+ \leq \varepsilon A, \qquad 0 \leq (h' - c')_+ \leq \varepsilon' A' \qquad \text{on } \Gamma.$$

Using

$$\zeta - \zeta' = \frac{(h-c)_+ - (h'-c')_+}{\varepsilon} + \Big(\frac{1}{\varepsilon} - \frac{1}{\varepsilon'}\Big)(h'-c')_+,$$

the fact that $t \mapsto t_+$ is 1-Lipschitz, and (66), we obtain

$$\|\zeta - \zeta'\|_{L^2(\bar\mu)} \le \frac{\|h-h'\|_{L^2(\bar\mu)} + \|c-c'\|_{L^2(\bar\mu)}}{\varepsilon} + \frac{A'}{\varepsilon}|\varepsilon - \varepsilon'| \le \frac{\overline{\gamma}\,\bar\Delta + \|c-c'\|_{L^2(\bar\mu)} + A'|\varepsilon-\varepsilon'|}{\varepsilon}.$$

Similarly,

$$\zeta - \zeta' = \frac{(h-c)_+ - (h'-c')_+}{\varepsilon'} + \Big(\frac{1}{\varepsilon} - \frac{1}{\varepsilon'}\Big)(h-c)_+,$$

and therefore

$$\|\zeta - \zeta'\|_{L^2(\bar\mu)} \le \frac{\overline{\gamma}\,\bar\Delta + \|c-c'\|_{L^2(\bar\mu)} + A|\varepsilon-\varepsilon'|}{\varepsilon'}.$$

Taking the minimum of the two bounds yields (29).

Step 2: Control of the reference measures. Writing

$$\mu - \mu' = (P-P')\otimes Q + P'\otimes(Q-Q'),$$

one sees that

$$\|\mu-\mu'\|_{\mathrm{TV}} \le \|P-P'\|_{\mathrm{TV}} + \|Q-Q'\|_{\mathrm{TV}} = \Delta_{\mathrm{TV}}. \tag{67}$$

Moreover, if $\gamma_P \in \Pi(P,P')$ and $\gamma_Q \in \Pi(Q,Q')$ are optimal couplings for $W_1(P,P')$ and $W_1(Q,Q')$, then $\gamma_P \otimes \gamma_Q$ is a coupling of μ and μ'. Therefore,

$$\begin{aligned} W_1(\mu,\mu') &\le \int \|(x,y)-(x',y')\|\, d(\gamma_P\otimes\gamma_Q)(x,x',y,y') \\ &\le \int \|x-x'\|\, d\gamma_P(x,x') + \int \|y-y'\|\, d\gamma_Q(y,y') \\ &= W_1(P,P') + W_1(Q,Q') \le \sqrt{2}\,\Delta_W. \end{aligned} \tag{68}$$

Step 3: TV-stability. Let

$$u := \frac{d\mu}{d\bar\mu}, \qquad u' := \frac{d\mu'}{d\bar\mu}.$$

Then $u + u' = 2$ $\bar\mu$-a.s., and by (65),

$$\frac{d\pi}{d\bar\mu} = \zeta\, u, \qquad \frac{d\pi'}{d\bar\mu} = \zeta'\, u'.$$

Using $u+u'=2$, we obtain

$$\frac{d\pi}{d\bar\mu} - \frac{d\pi'}{d\bar\mu} = \zeta u - \zeta' u' = (\zeta-\zeta') + \frac{\zeta+\zeta'}{2}(u-u'). \tag{69}$$

Hence,

$$2\|\pi-\pi'\|_{\mathrm{TV}} = \Big\|\frac{d\pi}{d\bar\mu} - \frac{d\pi'}{d\bar\mu}\Big\|_{L^1(\bar\mu)} \le \|\zeta-\zeta'\|_{L^1(\bar\mu)} + \frac{1}{2}\int (\zeta+\zeta')\, d|\mu-\mu'|.$$

Since $\bar{\mu}$ is a probability measure, (29) yields

$$\|\zeta - \zeta'\|_{L^1(\bar{\mu})} \leq \hat{\Delta}.$$

Moreover, by the definitions of A and A',

$$0 \leq \zeta \leq A, \qquad 0 \leq \zeta' \leq A' \qquad \text{on } \Gamma,$$

and $\operatorname{supp}|\mu - \mu'| \subset \Theta \subset \Gamma$. Therefore,

$$\frac{1}{2}\int (\zeta + \zeta')\, d|\mu - \mu'| \leq \frac{A + A'}{2}\, |\mu - \mu'|(\mathbb{R}^d \times \mathbb{R}^d) = (A + A')\|\mu - \mu'\|_{\mathrm{TV}}.$$

Combining the preceding estimates with (67), we obtain

$$2\|\pi - \pi'\|_{\mathrm{TV}} \leq \hat{\Delta} + (A + A')\Delta_{\mathrm{TV}},$$

which is (30).

Step 4: W_1-stability. Let $\varphi : \Theta \to \mathbb{R}$ be 1-Lipschitz. As subtracting a constant from φ does not change $\int \varphi\, d(\pi - \pi')$, we may assume that

$$\|\varphi\|_{L^\infty(\Theta)} \leq \frac{D_*}{2}.$$

Using (69), we find

$$\int \varphi\, d(\pi - \pi') = \int \varphi\, (\zeta - \zeta')\, d\bar{\mu} + \frac{1}{2}\int \varphi\, (\zeta + \zeta')\, d(\mu - \mu').$$

For the first term, (29) gives

$$\begin{aligned} \left|\int \varphi\, (\zeta - \zeta')\, d\bar{\mu}\right| &\leq \frac{D_*}{2}\int |\zeta - \zeta'|\, d\bar{\mu} \\ &\leq \frac{D_*}{2}\|\zeta - \zeta'\|_{L^2(\bar{\mu})} \leq \frac{D_*\,\hat{\Delta}}{2}. \end{aligned}$$

For the second term, define

$$\psi := \frac{1}{2}\,\varphi\,(\zeta + \zeta') \qquad \text{on } \Theta.$$

Since $\|\zeta + \zeta'\|_{L^\infty(\Theta)} \leq A + A'$, and since h, h' are $\sqrt{2}L$-Lipschitz while c, c' are L-Lipschitz, we have

$$\mathrm{Lip}(\zeta) \leq \frac{(\sqrt{2} + 1)L}{\varepsilon}, \qquad \mathrm{Lip}(\zeta') \leq \frac{(\sqrt{2} + 1)L}{\varepsilon'}.$$

Consequently,

$$\begin{aligned} \mathrm{Lip}(\psi) &\leq \frac{1}{2}\,\mathrm{Lip}(\varphi)\,\|\zeta + \zeta'\|_{L^\infty(\Theta)} + \frac{1}{2}\|\varphi\|_{L^\infty(\Theta)}\,\mathrm{Lip}(\zeta + \zeta') \\ &\leq \frac{A + A'}{2} + \frac{(\sqrt{2} + 1)L\,D_*}{4}\Big(\frac{1}{\varepsilon} + \frac{1}{\varepsilon'}\Big). \end{aligned}$$

By Kantorovich–Rubinstein duality on the compact set Θ, and using (68),

$$\begin{aligned}\left|\frac{1}{2}\int \varphi\,(\zeta+\zeta')\,d(\mu-\mu')\right| &= \left|\int \psi\,d(\mu-\mu')\right| \\ &\le \mathrm{Lip}(\psi)\,W_1(\mu,\mu') \\ &\le \sqrt{2}\left[\frac{A+A'}{2}+\frac{(\sqrt{2}+1)L\,D_*}{4}\Big(\frac{1}{\varepsilon}+\frac{1}{\varepsilon'}\Big)\right]\Delta_W.\end{aligned}$$

Collecting the bounds for the two terms, we conclude that

$$\left|\int \varphi\,d(\pi-\pi')\right| \le \frac{D_*\,\hat{\Delta}}{2}+\sqrt{2}\left[\frac{A+A'}{2}+\frac{(\sqrt{2}+1)L\,D_*}{4}\Big(\frac{1}{\varepsilon}+\frac{1}{\varepsilon'}\Big)\right]\Delta_W.$$

Taking the supremum over all 1-Lipschitz φ proves (31). $\square$

A.5 Proofs for Section 6

Next, we prove the Hausdorff stability bound for the optimal support.

Proof of Theorem 6.1. We have $\|h-h'\|_{L^\infty(\mathcal{X}\times\mathcal{Y})} \le \overline{C}\,\Delta_*$ by Corollary 4.2, so that

$$\|\sigma-\sigma'\|_{L^\infty(\mathcal{X}\times\mathcal{Y})} \le \|h-h'\|_{L^\infty(\mathcal{X}\times\mathcal{Y})}+\|c-c'\|_{L^\infty(\mathcal{X}\times\mathcal{Y})} \le \delta_*. \tag{70}$$

Step 1: Level set stability. We first show, without requiring (32), that

$$\overline{\{z\in\Omega_{P\otimes Q}\cap\Omega_{P'\otimes Q'}:\sigma(z)>\delta_*\}} \subset \Sigma' \subset \{z\in\Omega_{P'\otimes Q'}:\sigma(z)\ge -\delta_*\} \tag{71}$$

and

$$\Sigma\triangle\Sigma' \subset \Big(\{z\in\Omega_{P\otimes Q}\cap\Omega_{P'\otimes Q'}:|\sigma(z)|\le\delta_*,\ |\sigma'(z)|\le\delta_*\}\Big)\cup(\Omega_{P\otimes Q}\triangle\Omega_{P'\otimes Q'}). \tag{72}$$

Indeed, by (12),

$$\Sigma=\overline{\{z\in\Omega_{P\otimes Q}:\sigma(z)>0\}},\qquad \Sigma'=\overline{\{z\in\Omega_{P'\otimes Q'}:\sigma'(z)>0\}}.$$

If $z\in\Omega_{P\otimes Q}\cap\Omega_{P'\otimes Q'}$ satisfies $\sigma(z)>\delta_*$, then (70) yields

$$\sigma'(z)\ge\sigma(z)-\delta_*>0,$$

so $z\in\Sigma'$. Passing to closures gives the first inclusion in (71). Conversely, let $z\in\Sigma'$. If $\sigma(z)<-\delta_*$, then (70) implies $\sigma'(z)<0$, which by the continuity of σ' contradicts $z\in\Sigma'$. Therefore $\sigma(z)\ge-\delta_*$, which proves the second inclusion in (71).

Now let $z\in\Sigma\triangle\Sigma'$. If $z\in\Omega_{P\otimes Q}\triangle\Omega_{P'\otimes Q'}$, then z already belongs to the right-hand side of (72). If $z\notin\Omega_{P\otimes Q}\triangle\Omega_{P'\otimes Q'}$, then $z\in\Omega_{P\otimes Q}\cap\Omega_{P'\otimes Q'}$. In that case, $z\in\Sigma\triangle\Sigma'$ and a similar argument as above shows that $|\sigma(z)|\le\delta_*$ and $|\sigma'(z)|\le\delta_*$, proving (72).

Step 2: Hausdorff stability. First note that

$$d_H(\Omega_{P\otimes Q},\Omega_{P'\otimes Q'})\le\Delta_\Omega.$$

Indeed, if $z' = (x', y') \in \Omega_{P'\otimes Q'}$, we can choose $x \in \Omega_P$ and $y \in \Omega_Q$ with $\|x - x'\| \leq d_H(\Omega_P, \Omega_{P'})$ and $\|y - y'\| \leq d_H(\Omega_Q, \Omega_{Q'})$. Then

$$\|(x, y) - z'\| \leq \left(d_H(\Omega_P, \Omega_{P'})^2 + d_H(\Omega_Q, \Omega_{Q'})^2\right)^{1/2} = \Delta_\Omega,$$

and the reverse inclusion is analogous.

Let $z' = (x', y') \in \Sigma'$. Since $z' \in \Omega_{P'\otimes Q'}$, we can choose $z = (x, y) \in \Omega_{P\otimes Q}$ such that

$$\|z - z'\| \leq \Delta_\Omega. \tag{73}$$

By the second inclusion in (71),

$$\sigma(z') \geq -\delta_*.$$

Moreover, since $h = f \oplus g$ is $\sqrt{2}L$-Lipschitz and c is L-Lipschitz, $\sigma = h - c$ is $(\sqrt{2} + 1)L$-Lipschitz. Hence

$$\sigma(z) \geq \sigma(z') - (\sqrt{2} + 1)L\|z - z'\| \geq -\delta_* - (\sqrt{2} + 1)L\Delta_\Omega.$$

If $z \notin \Sigma$, then (32) yields

$$-a \operatorname{dist}(z, \Sigma) \geq \sigma(z) \geq -\delta_* - (\sqrt{2} + 1)L\Delta_\Omega,$$

that is,

$$\operatorname{dist}(z, \Sigma) \leq \frac{\delta_* + (\sqrt{2} + 1)L\Delta_\Omega}{a}.$$

If $z \in \Sigma$, the same estimate is trivial. Combining this with (73), we obtain

$$\operatorname{dist}(z', \Sigma) \leq \|z' - z\| + \operatorname{dist}(z, \Sigma) \leq \Delta_\Omega + \frac{\delta_* + (\sqrt{2} + 1)L\Delta_\Omega}{a}.$$

Therefore

$$\sup_{z' \in \Sigma'} \operatorname{dist}(z', \Sigma) \leq \left(1 + \frac{(\sqrt{2} + 1)L}{a}\right)\Delta_\Omega + \frac{\delta_*}{a}.$$

The analogue holds when primed and unprimed quantities are exchanged, and (33) follows. □

Finally, we verify the nondegeneracy condition under the concavity hypothesis, which holds in particular for the quadratic cost.

Proof of Proposition 6.3. If $D_P = 0$, then $\Omega_P = \{x_0\}$. The second identity in (9) gives $\sigma(x_0, y)_+ = \varepsilon$ for all $y \in \Omega_Q$, hence $\sigma(x_0, y) = \varepsilon$ for all $y \in \Omega_Q$. By (12), $\Omega_{P\otimes Q} \subset \Sigma$, so $\Omega_{P\otimes Q} \setminus \Sigma = \emptyset$. We may therefore assume that $D_P > 0$.

Fix $z = (x, y) \in \Omega_{P\otimes Q} \setminus \Sigma$. By the first-order condition (9), there exists $x_* \in \Omega_P$ such that

$$\sigma(x_*, y) \geq \varepsilon.$$

Define

$$S_y := \{\tilde{x} \in \Omega_P : \sigma(\tilde{x}, y) \geq 0\}.$$

Since $x \mapsto \sigma(x, y)$ is continuous and concave on the convex set Ω_P, the set S_y is nonempty, closed, and convex. Moreover,

$$S_y \times \{y\} \subset \Sigma. \tag{74}$$

Indeed, if $\tilde{x} \in S_y$ and $\sigma(\tilde{x}, y) > 0$, then $(\tilde{x}, y) \in \Sigma$ by (12). If $\sigma(\tilde{x}, y) = 0$, then for every $t \in (0, 1)$,

$$\sigma\big((1-t)\tilde{x} + tx_*, y\big) \geq (1-t)\sigma(\tilde{x}, y) + t\sigma(x_*, y) \geq t\varepsilon > 0,$$

where we used concavity of $\sigma(\cdot, y)$ and convexity of Ω_P. Hence

$$\big((1-t)\tilde{x} + tx_*, y\big) \in \Sigma \qquad \text{for all } t \in (0, 1),$$

and letting $t \downarrow 0$ yields $(\tilde{x}, y) \in \Sigma$, showing (74).

Let $z = (x, y) \in \Omega_{P \otimes Q} \setminus \Sigma$. Then (74) implies that $x \notin S_y$. By continuity of $\tilde{x} \mapsto \sigma(\tilde{x}, y)$ along the segment from x to x_*, there exists $\lambda \in (0, 1)$ such that

$$x_0 := (1-\lambda)x + \lambda x_* \in S_y \qquad \text{and} \qquad \sigma(x_0, y) = 0.$$

Using concavity once more, we obtain

$$0 = \sigma(x_0, y) \geq (1-\lambda)\sigma(x, y) + \lambda\sigma(x_*, y) \geq (1-\lambda)\sigma(x, y) + \lambda\varepsilon,$$

and therefore

$$-\sigma(x, y) \geq \frac{\lambda}{1-\lambda}\varepsilon \geq \lambda\varepsilon.$$

On the other hand,

$$\operatorname{dist}(x, S_y) \leq \|x - x_0\| = \lambda\|x - x_*\| \leq \lambda D_P.$$

Combining the last two displays yields

$$-\sigma(x, y) \geq \frac{\varepsilon}{D_P} \operatorname{dist}(x, S_y).$$

Noting that $\operatorname{dist}((x, y), \Sigma) \leq \operatorname{dist}(x, S_y)$ by (74), the claim (35) follows.

For the quadratic cost $c(x, y) = \frac{1}{2}\|x - y\|^2$, it is known that $\sigma(x, y)$ is concave in each component, so that (34) holds. Specifically, $\sigma(x, y) = \langle x, y\rangle - \varphi(x) - \psi(y)$ where φ, ψ are convex functions; see [32] or [18]. $\square$

References

[1] Bayraktar, E., Eckstein, S., and Zhang, X. Stability and sample complexity of divergence regularized optimal transport. *Bernoulli* 31.1 (2025), pp. 213–239. DOI: 10.3150/24-BEJ1725.

[2] Blondel, M., Seguy, V., and Rolet, A. Smooth and Sparse Optimal Transport. *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*. Ed. by A. Storkey and F. Perez-Cruz. Vol. 84. Proceedings of Machine Learning Research. PMLR, 2018, pp. 880–889.

[3] Bonnans, J. F. and Shapiro, A. *Perturbation Analysis of Optimization Problems*. Springer Series in Operations Research. New York: Springer-Verlag, 2000, pp. xviii+601. DOI: 10.1007/978-1-4612-1394-9.

[4] Carlier, G., Chizat, L., and Laborde, M. Displacement smoothness of entropic optimal transport. *ESAIM Control Optim. Calc. Var.* 30 (2024), Paper No. 25, 24. DOI: 10.1051/cocv/2024013.

[5] Carlier, G. and Laborde, M. A differential approach to the multi-marginal Schrödinger system. *SIAM J. Math. Anal.* 52.1 (2020), pp. 709–717. DOI: 10.1137/19M1253800.

[6] Cuesta-Albertos, J. A., Matrán, C., and Tuero-Díaz, A. Optimal transportation plans and convergence in distribution. *J. Multivariate Anal.* 60.1 (1997), pp. 72–83. DOI: 10.1006/jmva.1996.1651.

[7] Cuturi, M. Sinkhorn Distances: Lightspeed Computation of Optimal Transport. *Advances in Neural Information Processing Systems 26*. Ed. by C. J. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger. Curran Associates, Inc., 2013, pp. 2292–2300. DOI: 10.5555/2999792.2999868.

[8] Delalande, A. and Mérigot, Q. Quantitative stability of optimal transport maps under variations of the target measure. *Duke Math. J.* 172.17 (2023), pp. 3321–3357. DOI: 10.1215/00127094-2022-0106.

[9] Deligiannidis, G., Bortoli, V. de, and Doucet, A. Quantitative uniform stability of the iterative proportional fitting procedure. *Ann. Appl. Probab.* 34.1A (2024), pp. 501–516. DOI: 10.1214/23-AAP1970.

[10] Divol, V., Niles-Weed, J., and Pooladian, A.-A. Tight stability bounds for entropic Brenier maps. *Int. Math. Res. Not. IMRN* 2025.7 (2025), rnaf078. DOI: 10.1093/imrn/rnaf078. arXiv: 2404.02855.

[11] Eckstein, S. and Nutz, M. Quantitative stability of regularized optimal transport and convergence of Sinkhorn's algorithm. *SIAM J. Math. Anal.* 54.6 (2022), pp. 5922–5948. DOI: 10.1137/21M145505X.

[12] Eckstein, S. and Nutz, M. Convergence rates for regularized optimal transport via quantization. *Math. Oper. Res.* 49.2 (2024), pp. 1223–1240. DOI: 10.1287/moor.2022.0245.

[13] Essid, M. and Solomon, J. Quadratically regularized optimal transport on graphs. *SIAM J. Sci. Comput.* 40.4 (2018), A1961–A1986. DOI: 10.1137/17M1132665.

[14] Ghosal, P., Nutz, M., and Bernton, E. Stability of entropic optimal transport and Schrödinger bridges. *J. Funct. Anal.* 283.9 (2022), Paper No. 109622. DOI: 10.1016/j.jfa.2022.109622.

[15] Gigli, N. On Hölder continuity-in-time of the optimal transport map towards measures along a curve. *Proc. Edinb. Math. Soc. (2)* 54.2 (2011), pp. 401–409. DOI: 10.1017/S001309150800117X.

[16] González-Sanz, A., Barrio, E. del, and Nutz, M. *Sample Complexity of Quadratically Regularized Optimal Transport*. 2025. arXiv: 2511.09807.

[17] González-Sanz, A., Gvalani, R. S., and Koch, L. *Sharp local sparsity of regularized optimal transport*. 2026. arXiv: 2604.00843.

[18] González-Sanz, A. and Nutz, M. *Sparsity of Quadratically Regularized Optimal Transport: Scalar Case*. SIAM J. Math. Anal., forthcoming. 2024. arXiv: 2410.03353.

[19] González-Sanz, A., Nutz, M., and Riveros Valdevenito, A. *Linear Convergence of Gradient Descent for Quadratically Regularized Optimal Transport*. 2025. arXiv: 2509.08547.

[20] González-Sanz, A., Nutz, M., and Riveros Valdevenito, A. Polyak–Łojasiewicz Inequality for Quadratically Regularized Optimal Transport. *SIAM J. Optim.* (2026). Forthcoming. arXiv: 2605.27175.

[21] González-Sanz, A. and Sheng, S. *Linearization of Monge–Ampère Equations and Statistical Applications*. 2024. arXiv: 2408.06534.

[22] Grisvard, P. *Elliptic Problems in Nonsmooth Domains*. Vol. 24. Monographs and Studies in Mathematics. Boston, MA: Pitman Advanced Publishing Program, 1985, pp. xiv+410.

[23] Letrouit, C. and Mérigot, Q. *Gluing methods for quantitative stability of optimal transport maps*. 2024. arXiv: 2411.04908.

[24] Lorenz, D. A., Manns, P., and Meyer, C. Quadratically regularized optimal transport. *Appl. Math. Optim.* 83.3 (2021), pp. 1919–1949. DOI: 10.1007/s00245-019-09614-w.

[25] Mérigot, Q. *Sharp Stability of Brenier Maps via Quantitative Regularity of Potentials*. HAL preprint, hal-05616391. 2026.

[26] Muzellec, B., Nock, R., Patrini, G., and Nielsen, F. Tsallis Regularized Optimal Transport and Ecological Inference. *Proceedings of the AAAI Conference on Artificial Intelligence* 31.1 (2017). DOI: 10.1609/aaai.v31i1.10854.

[27] Nutz, M. *Introduction to Entropic Optimal Transport*. Lecture notes, Columbia University, https://www.math.columbia.edu/~mnutz/docs/EOT_lecture_notes.pdf. 2021.

[28] Nutz, M. Quadratically regularized optimal transport: existence and multiplicity of potentials. *SIAM J. Math. Anal.* 57.3 (2025), pp. 2622–2649. DOI: 10.1137/24M1638136.

[29] Peyré, G. and Cuturi, M. Computational Optimal Transport: With Applications to Data Science. *Foundations and Trends in Machine Learning* 11.5–6 (2019), pp. 355–607. DOI: 10.1561/2200000073.

[30] Rüschendorf, L. and Thomsen, W. Note on the Schrödinger equation and I-projections. *Statist. Probab. Lett.* 17.5 (1993), pp. 369–375. DOI: 10.1016/0167-7152(93)90257-J.

[31] Schmitzer, B. Stabilized sparse scaling algorithms for entropy regularized transport problems. *SIAM J. Sci. Comput.* 41.3 (2019), A1443–A1481. DOI: 10.1137/16M1106018.

[32] Wiesel, J. and Xu, X. Sparsity of Quadratically Regularized Optimal Transport: Bounds on Concentration and Bias. *SIAM J. Math. Anal.* 57.6 (2025), pp. 6498–6521. DOI: 10.1137/25M1723633.

[33] Yao, R., Kim, Y.-H., and Schiebinger, G. *Quantitative Stability of Many-Marginal Schrödinger Bridge*. 2026. arXiv: 2604.15191.