

Polyak–Łojasiewicz Inequality for Quadratically Regularized Optimal Transport

Alberto González-Sanz* Marcel Nutz† Andrés Riveros Valdevenito‡

May 26, 2026

Abstract

Quadratically regularized optimal transport (QOT) is an alternative to entropic regularization that yields sparse couplings and avoids numerical instabilities due to exponential scaling. From an optimization viewpoint, the dual QOT objective is concave but features a positive part function which prevents strong concavity and reduces smoothness of optimizers. Consequently, standard arguments for linear convergence of algorithms do not apply. In this paper, we nevertheless establish a quantitative curvature property for the QOT dual. Under mild assumptions covering both continuous and semi-discrete transport problems, we prove a local error bound and a Polyak–Łojasiewicz (PL) inequality, with explicit constants depending only on the problem primitives. These results are obtained by functional-analytic techniques exploiting that near the optimum, the argument of the positive part function is positive on the interior of the support of the optimal coupling. As applications, we derive linear convergence of the gradient ascent, coordinate ascent, and coordinate gradient ascent algorithms on the dual problem, with explicit contraction rates.

Keywords Optimal Transport; Quadratic Regularization; PL Inequality; Linear Convergence
AMS 2020 Subject Classification 49N10; 49N05; 90C25

1 Introduction

Optimal transport provides a principled framework for comparing probability distributions, with ubiquitous applications from machine learning to economics. Given compactly supported probability measures P and Q on \mathbb{R}^d , the optimal transport problem with cost function $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ is

$$\text{OT}(P, Q) = \inf_{\pi \in \Pi(P, Q)} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) d\pi(x, y), \quad (1)$$

where $\Pi(P, Q)$ denotes the set of couplings (or transport plans), i.e., probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals (P, Q) . In particular, the optimal value $\text{OT}(P, Q)$ induces the p -Wasserstein distance when $c(x, y) = \|x - y\|^p$. As (1) is computationally and statistically challenging in high dimensions, the dominant approach in applications is to introduce a regularization term. Popularized by [7], the most widely used is Kullback–Leibler (KL) divergence, leading to the *entropic optimal transport (EOT)* problem

$$\text{EOT}_\varepsilon(P, Q) = \inf_{\pi \in \Pi(P, Q)} \int c(x, y) d\pi(x, y) + \varepsilon \text{KL}(\pi | P \otimes Q). \quad (2)$$

*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.

‡Department of Statistics, Columbia University, ar4151@columbia.edu.

Here $\varepsilon > 0$ is a parameter controlling the strength of the regularization. The associated dual problem is the maximization of

$$\int \left(f(x) + g(y) - \varepsilon e^{\frac{f(x)+g(y)-c(x,y)}{\varepsilon}} \right) d(P \otimes Q)(x, y) \quad (3)$$

over functions $f, g : \mathbb{R}^d \rightarrow \mathbb{R}$. The celebrated Sinkhorn algorithm can be interpreted as the (block) coordinate ascent algorithm on this objective. Two key features underlying the success of EOT and Sinkhorn’s method are the strong concavity of the objective (3) and the fact that its optimizers (f_*, g_*) are automatically as smooth as the cost c . In particular, strong concavity directly yields linear convergence of Sinkhorn’s algorithm [6]. At the same time, entropic regularization introduces structural distortions. The infinite derivative of $t \log t$ at zero entails that the optimal coupling of (2) necessarily has full support (the same support as $P \otimes Q$), even though (1) is typically sparse. This “overspreading” induces blurring or bias in applications such as image processing [2] or manifold learning [33]. Moreover, weak regularization $\varepsilon \ll 1$ leads to numerical instability due to exponentially large and small values in the Sinkhorn updates [29]. See, e.g., [27, 25] for further background.

The most popular alternative, first considered by [24, 2, 10], replaces KL divergence by the squared L^2 -norm (equivalently, χ^2 divergence), leading to the *quadratically regularized optimal transport (QOT)* problem

$$\text{QOT}_\varepsilon(P, Q) = \inf_{\pi \in \Pi(P, Q)} \int c(x, y) d\pi(x, y) + \frac{\varepsilon}{2} \left\| \frac{d\pi}{d(P \otimes Q)} \right\|_{L^2(P \otimes Q)}^2. \quad (4)$$

Its optimal value converges to $\text{OT}(P, Q)$ as $\varepsilon \rightarrow 0$ at the rate $\varepsilon^{2/(d+2)}$ [9]. Unlike entropic regularization, the optimal coupling is sparse for small to moderate ε , as has been observed empirically in many works including [2, 10, 23, 1] and recently established theoretically in [32, 17, 26]. Moreover, exponentially large or small values do not appear in QOT. A variety of computational approaches have been developed, including mirror gradient methods [24], Newton-type algorithms [10], Gauss–Seidel (coordinate ascent) schemes [2, 23], cyclic projections and gradient methods [21], as well as neural network methods parameterizing the dual domain [8, 12, 19, 20, 30].

From an optimization viewpoint, quadratic regularization leads to a markedly different geometry. The dual problem associated with (4) is the maximization of the concave objective

$$\Gamma(f, g) = \int \left(f(x) + g(y) - \frac{1}{2\varepsilon} (f(x) + g(y) - c(x, y))_+^2 \right) d(P \otimes Q)(x, y). \quad (5)$$

Note that the positive part function $(t)_+ = \max(t, 0)$ destroys strong (and even strict) concavity: Γ is linear in $f \oplus g$ as long as $f \oplus g - c \leq 0$, where $(f \oplus g)(x, y) := f(x) + g(y)$. Moreover, optimizers have limited smoothness (not \mathcal{C}^2 even if $c(x, y) = \|x - y\|^2$, cf. [14, 26]). Thus, many of the arguments that are used in EOT do not apply. While well established in computational practice, QOT has long been regarded as difficult to study analytically, and it is fair to say that the theory is still in its infancy. However, two recent results suggest that QOT is better behaved than the aforementioned properties suggest: despite the lack of strong concavity and smoothness, [13] established that QOT has parametric sample complexity (i.e., does not suffer from the same statistical curse of dimensionality as unregularized

optimal transport), and [16] showed that the gradient ascent algorithm for (5) converges linearly. These results hint that the geometry of the QOT dual may retain some features of strong concavity, at least locally around the optimum. The goal of the present work is to crystallize this intuition into a quantitative result.

Indeed, our main result provides an error bound and a Polyak–Łojasiewicz (PL) inequality for the QOT dual (5) under mild assumptions, thus offering the first structural understanding for the optimization and statistical properties of QOT. Denoting by (f_*, g_*) the (essentially unique) dual optimizers, Theorem 2.5 states that

$$\|f \oplus g - f_* \oplus g_*\|_{L^2(P \otimes Q)} \leq \gamma_\varepsilon \max(\|f \oplus g - f_* \oplus g_*\|_\infty, \varepsilon) \|\text{D}\Gamma(f, g)\|_{L^2(P) \times L^2(Q)}$$

and

$$\|\text{D}\Gamma(f, g)\|_{L^2(P) \times L^2(Q)}^2 \geq \frac{1}{\gamma_\varepsilon \max(\|f \oplus g - f_* \oplus g_*\|_\infty, \varepsilon)} (\text{QOT}_\varepsilon(P, Q) - \Gamma(f, g)),$$

where $\text{D}\Gamma$ denotes the gradient and γ_ε is an explicit constant. This error bound and PL inequality are local in the sense that their constants are uniform only for $\|f \oplus g - f_* \oplus g_*\|_\infty \leq \varepsilon$, where ε is the (fixed) regularization parameter. Such a localization is clearly necessary, given the aforementioned linear structure of the dual on part of the domain. Our setting assumes that one marginal has compact and convex support (or more generally, compact connected Lipschitz support) and a density that is bounded from above and below, whereas the second marginal is only assumed to be compactly supported and the transport cost c is any Lipschitz (or uniformly continuous) function. In particular, our setting covers both continuous and semi-discrete transport problems.

It is well known that a PL inequality enables a host of desirable properties for optimization. In this paper, our applications focus on the convergence of several algorithms for the dual QOT problem (and hence also for the primal problem (4), since the dual optimizers yield the optimal coupling via the relation (11) below). Specifically, we show in Corollaries 2.6 to 2.8 that gradient ascent, coordinate ascent, and coordinate gradient ascent all converge linearly (for suitable step sizes), with explicit contraction constants. Of course, the PL inequality may also be useful for analyzing other algorithms. Further important applications, to be developed more fully in forthcoming works, include explicit bounds for the statistical sample complexity of QOT and quantitative stability for the dual (and hence also the primal) QOT optimizers. Regarding the second application, we mention that first Lipschitz-stability results follow immediately along the lines of [3, Section 4.4], whereas stronger results follow with additional work (see [15]). These implications of the PL inequality offer a marked improvement over the Hölder continuity of optimal couplings shown in [1, Theorem 3.3].

The basic intuition for our main result is the following. The dual objective (5) would be strongly concave (in $f \oplus g$) if it weren't for the positive part operator. The latter is active only on the complement of the set $\mathcal{S} = \{(x, y) : f(x) + g(y) \geq c(x, y)\}$. When $(f, g) = (f_*, g_*)$ are dual optimizers, \mathcal{S} is the support of the optimal coupling, by way of (11). While we expect the support to be sparse, the fact that it carries the solution also implies certain lower bounds. Thus, if (f, g) is close to (f_*, g_*) , we know that there is a nontrivial set where the positive part operator is not active, and that set must be exploited as the source of curvature for our result. However, it is not straightforward to turn this intuition into a formal argument (in part, because we have very limited knowledge about the shape

and size of the optimal support). The actual proof follows an original approach leveraging functional-analytic tools: We fix (f, g) and consider the dual objective $\Gamma(f_t, g_t)$ along the interpolation $(f_t, g_t) := t(f, g) + (1 - t)(f_*, g_*)$. The main step is to bound the second derivative $\partial_{tt}^2 \Gamma(f_t, g_t)$ from below, which corresponds to the coercivity of an integral operator whose kernel is $\mathbf{1}_{\{f_t(x) + g_t(y) \geq c(x, y)\}}$. For $t = 0$, this set is the support of the optimal coupling, linking to the intuition sketched above. Establishing the lower bound boils down to bounding the minimum eigenvalue of an associated operator, which requires us to develop a tailored spectral analysis since the operator is not compact.

As the PL inequality is local, drawing the corollaries for the algorithms requires uniform bounds for the iterates. These bounds are algorithm-specific, but once they are obtained, the linear convergences are direct consequences of the PL inequality. Of course, this analysis could be extended to variants of the discussed algorithms, for instance variable step sizes. As mentioned, linear convergence of gradient ascent was previously shown in [16]. However, the contraction constant given there is not explicit (due to the proof containing a qualitative compactness argument). Moreover, the proof is not easily adapted to other algorithms such as coordinate ascent or coordinate gradient ascent since it relies on a symmetry property specific to the gradient ascent algorithm. The derivation through the PL inequality in the present work offers a systematic approach that covers a whole class of algorithms. Apart from [16], we are not aware of previous results on linear convergence for continuous QOT.

Organization Section 2.1 details the assumptions and collects relevant background results. The main result on error bound and PL inequality is Theorem 2.5 in Section 2.2, whereas the linear convergences of the algorithms are stated as Corollaries 2.6 to 2.8 in Section 2.3. The proof of the main result is given in Section 3, and the corollaries for the algorithms are derived in Section 4. Section 5 concludes with a summary and practical remarks. Appendix A extends the results from Lipschitz to uniformly continuous costs, and Appendix B relaxes the convexity assumption on the first marginal support.

2 Main Results

2.1 Problem statement and background

We first detail our assumptions on the marginals P, Q and the cost c . Let \mathcal{L}_d denote the d -dimensional Lebesgue measure and $B_r(x)$ the open ball of center x and radius $r > 0$. We fix probability measures P and Q on \mathbb{R}^d and impose the following conditions throughout.

Assumption 2.1 (Marginals). (a) The first marginal P has convex, compact support Ω and admits a density $\rho := dP/d\mathcal{L}_d$ that is bounded away from zero and infinity on Ω , i.e., there exist constants $0 < \lambda_P < \Lambda_P < \infty$ such that

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

(b) The second marginal Q has compact support Ω' .

The convexity condition can be relaxed to Ω being the closure of a bounded Lipschitz domain. All our results extend to that setting, at the expense of an additional constant that depends on the geometry of Ω and a more complicated proof. This extension is detailed in Section B.

Remark 2.2 (Lower bounds for ball measures). (a) Since Ω is compact and convex, it satisfies a uniform interior cone condition: for every $x \in \Omega$, there exists a convex cone \mathcal{C}_x with vertex x , angle $\theta > 0$ and height $h > 0$ such that $\mathcal{C}_x \subset \Omega$, where θ and h are independent of x (e.g., [18, p. 12]). Hence, Assumption 2.1 implies that there exists $\delta_P \in (0, 1]$, depending only on Ω and λ_P , such that

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

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

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

We note that Q can be continuous or discrete, that is, our setup covers both continuous and semi-discrete optimal transport.

Assumption 2.3 (Cost). The cost $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ is Lipschitz with constant $L > 0$.

Lipschitz-continuity can be relaxed to uniform continuity at the expense of a more complicated expression for the constant in the main results; see Section A.

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

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

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

$$\text{QOT}_\varepsilon^*(P, Q) = \sup_{(f, g) \in L^2(P) \times L^2(Q)} \Gamma(f, g), \quad (7)$$

where the dual objective function $\Gamma : L^2(P) \times L^2(Q) \rightarrow \mathbb{R}$ is defined by

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

The gradient of Γ at $(f, g) \in L^2(P) \times L^2(Q)$ is

$$D\Gamma \begin{pmatrix} f \\ g \end{pmatrix} = \begin{pmatrix} D_1\Gamma \\ D_2\Gamma \end{pmatrix} \begin{pmatrix} f \\ g \end{pmatrix} = \begin{pmatrix} 1 - \frac{1}{\varepsilon} \int (f(\cdot) + g(y) - c(\cdot, y))_+ dQ(y) \\ 1 - \frac{1}{\varepsilon} \int (f(x) + g(\cdot) - c(x, \cdot))_+ dP(x) \end{pmatrix} \quad (9)$$

and its action on $(u, v) \in L^2(P) \times L^2(Q)$ is

$$\langle D\Gamma(f, g), (u, v) \rangle_{L^2(P) \times L^2(Q)} = \int (u \oplus v) \left(1 - \frac{1}{\varepsilon} (f \oplus g - c)_+ \right) d(P \otimes Q).$$

Next, we recall a number of known properties. They do not require the full strength of Assumption 2.1, only that Ω, Ω' are compact and Ω is connected.

Proposition 2.4. (i) *The strong duality $\text{QOT}_\varepsilon(P, Q) = \text{QOT}_\varepsilon^*(P, Q)$ holds.*

- (ii) The dual problem (7) admits an optimizer $(f_*, g_*) \in L^2(P) \times L^2(Q)$. The functions (f_*, g_*) are unique up to translation in the following sense: fixing one optimizer (f_*, g_*) , the set of all optimizers is given by $\{(f_* + a, g_* - a) : a \in \mathbb{R}\}$.
- (iii) We can choose versions $f_* : \Omega \rightarrow \mathbb{R}$ and $g_* : \Omega' \rightarrow \mathbb{R}$ that are L -Lipschitz. Those versions are fixed in all that follows. We call (f_*, g_*) the potentials.
- (iv) The potentials (f_*, g_*) satisfy the first-order condition

$$\begin{cases} \int (f_*(x) + g_*(y) - c(x, y))_+ dQ(y) = \varepsilon & \text{for all } x \in \Omega, \\ \int (f_*(x) + g_*(y) - c(x, y))_+ dP(x) = \varepsilon & \text{for all } y \in \Omega'. \end{cases} \quad (10)$$

- (v) The primal problem (6) has a unique solution $\pi_\varepsilon \in \Pi(P, Q)$. It is related to the potentials by

$$\frac{d\pi_\varepsilon}{d(P \otimes Q)}(x, y) = \frac{1}{\varepsilon} (f_*(x) + g_*(y) - c(x, y))_+. \quad (11)$$

We refer to the proof of [16, Lemma 3.1] for the uniqueness of the potentials and to [26, Section 2] for all other claims.

Given functions $f : \Omega \rightarrow \mathbb{R}$ and $g : \Omega' \rightarrow \mathbb{R}$, we denote $(f \oplus g)(x, y) := f(x) + g(y)$. The uniqueness of the potentials up to translation implies that $f_* \oplus g_*$ is uniquely determined. More generally, we observe that the dual objective (8) can be seen as a function of $f \oplus g$ rather than the two separate functions f and g . This point of view is used for our main proofs, which consider the operator $\Phi(f \oplus g) := \Gamma(f, g)$ on the space $\mathcal{H} = \{f \oplus g : (f, g) \in L^2(P) \times L^2(Q)\} \subset L^2(P \otimes Q)$. This is mathematically convenient and also leads to slightly tighter constants. On the other hand, algorithms (and scientists) often consider f, g as individual functions that are updated separately, hence we take that point of view in the presentation of the main results below.

2.2 PL inequality

Our main result is a PL-type inequality for the dual objective Γ of (8). We recall that $(f_*, g_*) \in \mathcal{C}(\Omega) \times \mathcal{C}(\Omega')$ are fixed (but arbitrary) potentials and the constants $\lambda_P, \Lambda_P, \delta_P, L$ were defined in Assumption 2.1, Remark 2.2 and Assumption 2.3.

Theorem 2.5 (PL inequality for Γ). *For $(f, g) \in L^\infty(\Omega) \times L^\infty(\Omega')$, we have the error bound*

$$\|f \oplus g - f_* \oplus g_*\|_{L^2(P \otimes Q)} \leq \gamma_\varepsilon \max(\|f \oplus g - f_* \oplus g_*\|_\infty, \varepsilon) \|\mathrm{D}\Gamma(f, g)\|_{L^2(P) \times L^2(Q)}$$

and the PL inequality

$$\|\mathrm{D}\Gamma(f, g)\|_{L^2(P) \times L^2(Q)}^2 \geq \frac{1}{\gamma_\varepsilon \max(\|f \oplus g - f_* \oplus g_*\|_\infty, \varepsilon)} (\mathrm{QOT}_\varepsilon(P, Q) - \Gamma(f, g)), \quad (12)$$

where

$$\gamma_\varepsilon = 16 \left(\delta_P^{-1} \max\left(\frac{8L}{\varepsilon}, 1\right) \right)^d \frac{\Lambda_P^2 (\lceil 8L \mathrm{diam}(\Omega)/\varepsilon \rceil)^{d+2}}{\lambda_P^2 \inf_{y \in \Omega'} Q(B_{\frac{\varepsilon}{8L}}(y))}. \quad (13)$$

The proof is given in Section 3.

2.3 Convergence of algorithms

Next, we apply the PL inequality to show that three natural algorithms for the dual problem (7) converge linearly for suitable step sizes: gradient ascent, coordinate ascent, and coordinate gradient ascent. All algorithms have iterates (f_n, g_n) , and linear convergence means that the suboptimality gap $\Delta_n := \text{QOT}_\varepsilon(P, Q) - \Gamma(f_n, g_n) \equiv \Gamma(f_*, g_*) - \Gamma(f_n, g_n)$ satisfies $\Delta_n \leq (1 - q)^n \Delta_0$ for some $q \in (0, 1)$ that is made explicit below. Numerical experiments for these algorithms can be found in [21] (they are omitted in the journal version [22]). The experiments suggest that all three algorithms are efficient and consistent in the examples considered, but no theoretical analysis was given.

The proof of each corollary below is reported in Section 4 and combines three facts: the PL inequality (Theorem 2.5), the Lipschitz property of $\text{D}\Gamma$ (Lemma 4.1), and a uniform bound for $\|f_n \oplus g_n - f_* \oplus g_*\|_\infty$. The latter gives a uniform constant for the PL inequality and has a slightly different proof for each algorithm; after that, the argument for linear convergence is standard. As mentioned in the introduction, linear convergence of gradient ascent was previously shown in [16], but without explicit control of the contraction rate. Apart from that, we are not aware of any previous results on convergence rates of QOT algorithms. For the statements that follow, we recall that $(f_*, g_*) \in \mathcal{C}(\Omega) \times \mathcal{C}(\Omega')$ are fixed (but arbitrary) potentials and the constant γ_ε defined in (13).

2.3.1 Gradient ascent

The gradient ascent algorithm with step size $\eta > 0$ is initialized at $(f_0, g_0) \in L^\infty(\Omega) \times L^\infty(\Omega')$ and defined by

$$\begin{pmatrix} f_{n+1} \\ g_{n+1} \end{pmatrix} = \begin{pmatrix} f_n \\ g_n \end{pmatrix} + \eta \text{D}\Gamma \begin{pmatrix} f_n \\ g_n \end{pmatrix}, \quad n \geq 0, \quad (14)$$

where the gradient $\text{D}\Gamma$ has the explicit form stated in (9). It is well known that a PL inequality implies linear convergence of gradient ascent. Given the specific form of our (local) PL inequality (12), we only need to bound $\|f_n \oplus g_n - f_* \oplus g_*\|_\infty$ uniformly in n .

Corollary 2.6 (Linear convergence of gradient ascent). *Let $(f_0, g_0) \in L^\infty(\Omega) \times L^\infty(\Omega')$ and define $(f_n, g_n)_{n \geq 1}$ by (14). For every $\eta \in (0, \varepsilon)$ and $n \geq 1$,*

$$\|f_n \oplus g_n - f_* \oplus g_*\|_\infty \leq 2\|f_0 \oplus g_0 - f_* \oplus g_*\|_\infty \quad (15)$$

and the suboptimality gap $\Delta_n := \text{QOT}_\varepsilon(P, Q) - \Gamma(f_n, g_n)$ satisfies

$$\Delta_n \leq (1 - q)^n \Delta_0 \quad \text{for} \quad q := \frac{\eta \left(1 - \frac{\eta}{\varepsilon}\right)}{\gamma_\varepsilon \max(2\|f_0 \oplus g_0 - f_* \oplus g_*\|_\infty, \varepsilon)}.$$

Moreover,

$$\|f_n \oplus g_n - f_* \oplus g_*\|_{L^2(P \otimes Q)}^2 \leq \frac{\gamma_\varepsilon^2 \max(2\|f_0 \oplus g_0 - f_* \oplus g_*\|_\infty, \varepsilon)^2}{\eta \left(1 - \frac{\eta}{\varepsilon}\right)} \Delta_0 (1 - q)^n.$$

2.3.2 Coordinate ascent

The coordinate ascent algorithm is an implicit algorithm which is initialized at $g_0 \in L^\infty(\Omega')$ and iteratively optimizes the coordinates of Γ ,

$$f_n := \operatorname{argmax}_f \Gamma(f, g_n), \quad g_{n+1} := \operatorname{argmax}_g \Gamma(f_n, g), \quad n \geq 0. \quad (16)$$

This is equivalent to iteratively solving the first-order conditions for $n \geq 0$:

$$\text{define } f_n(x) \text{ via } \varepsilon = \int (f_n(x) + g_n(y) - c(x, y))_+ dQ(y), \quad (17)$$

$$\text{then } g_{n+1}(y) \text{ via } \varepsilon = \int (f_n(x) + g_{n+1}(y) - c(x, y))_+ dP(x). \quad (18)$$

These equations imply that the argmax is unique and a bounded function (by the same arguments as [26, Lemmas 2.4, 2.5]), so that the iterates are uniquely defined.

We remark that the coordinate ascent algorithm becomes Sinkhorn's algorithm when the quadratic regularization is replaced by KL divergence (i.e., entropic optimal transport). In that case, the above equations are solved explicitly as the properties of the exponential function allow one to take f_n and g_{n+1} out of the integral. For quadratic regularization, the definition remains implicit. In [22], methods such as line search were proposed to solve the implicit problems.

Corollary 2.7 (Linear convergence of coordinate ascent). *Let $g_0 \in L^\infty(\Omega')$ and define $(f_n, g_n)_{n \geq 0}$ by (16). Then*

$$\|g_{n+1} - g_*\|_\infty \leq \|f_n - f_*\|_\infty \leq \|g_n - g_*\|_\infty \leq \dots \leq \|g_0 - g_*\|_\infty \quad (19)$$

and the suboptimality gap $\Delta_n := \operatorname{QOT}_\varepsilon(P, Q) - \Gamma(f_n, g_n)$ satisfies

$$\Delta_n \leq (1 - q)^n \Delta_0 \quad \text{for} \quad q := \frac{\varepsilon}{2\gamma_\varepsilon \max(2\|g_0 - g_*\|_\infty, \varepsilon)}.$$

Moreover,

$$\|f_n \oplus g_n - f_* \oplus g_*\|_{L^2(P \otimes Q)}^2 \leq \frac{2\gamma_\varepsilon^2 \max(2\|g_0 - g_*\|_\infty, \varepsilon)^2}{\varepsilon} \Delta_0 (1 - q)^n.$$

In the definition of q , the constant $\|g_0 - g_*\|_\infty$ can be replaced by $\inf_{a \in \mathbb{R}} \|g_0 - g_* + a\|_\infty$ since the corollary holds for any pair of potentials, including $(f_* + a, g_* - a)$.

2.3.3 Coordinate gradient ascent

The coordinate gradient ascent algorithm with step size $\eta > 0$ is initialized at $(f_0, g_0) \in L^\infty(\Omega) \times L^\infty(\Omega')$ and defined by

$$f_{n+1} := f_n + \eta D_1 \Gamma \begin{pmatrix} f_n \\ g_n \end{pmatrix}, \quad g_{n+1} := g_n + \eta D_2 \Gamma \begin{pmatrix} f_{n+1} \\ g_n \end{pmatrix}, \quad n \geq 0. \quad (20)$$

Corollary 2.8 (Linear convergence of coordinate gradient ascent). *Let $(f_0, g_0) \in L^\infty(\Omega) \times L^\infty(\Omega')$ and define $(f_n, g_n)_{n \geq 1}$ by (20). For every $\eta \in (0, \varepsilon/\sqrt{2})$,*

$$\max(\|f_{n+1} - f_*\|_\infty, \|g_{n+1} - g_*\|_\infty) \leq \max(\|f_n - f_*\|_\infty, \|g_n - g_*\|_\infty) \quad (21)$$

and the suboptimality gap $\Delta_n := \text{QOT}_\varepsilon(P, Q) - \Gamma(f_n, g_n)$ satisfies

$$\Delta_n \leq (1 - q)^n \Delta_0 \quad \text{for} \quad q := \frac{\eta \left(1 - \frac{\eta}{2\varepsilon}\right)}{2\gamma_\varepsilon \max(2\|f_0 \oplus g_0 - f_* \oplus g_*\|_\infty, \varepsilon)}.$$

Moreover,

$$\|f_n \oplus g_n - f_* \oplus g_*\|_{L^2(P \otimes Q)}^2 \leq \frac{2\gamma_\varepsilon^2 \max(2\|f_0 \oplus g_0 - f_* \oplus g_*\|_\infty, \varepsilon)^2}{\eta \left(1 - \frac{\eta}{2\varepsilon}\right)} \Delta_0 (1 - q)^n.$$

3 Proof of the PL inequality

Consider the linear subspace

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

which is closed by the argument in [28, p. 370]. Hence, \mathcal{H} is a Hilbert space with the induced inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}} = \langle \cdot, \cdot \rangle_{L^2(P \otimes Q)}$. We define the operator $\Phi : \mathcal{H} \rightarrow \mathbb{R}$ by $\Phi(f \oplus g) = \Gamma(f, g)$, that is,

$$\Phi(f \oplus g) := \int \left(f(x) + g(y) - \frac{1}{2\varepsilon} (f(x) + g(y) - c(x, y))_+^2 \right) d(P \otimes Q)(x, y). \quad (22)$$

Working with Φ and \mathcal{H} will be convenient as the direct sum removes the non-uniqueness of the potentials. The relation between the gradients of Φ and Γ will be detailed in Lemma 3.8.

We fix $(f, g) \in L^\infty(\Omega) \times L^\infty(\Omega')$ in addition to the direct sum $f_* \oplus g_*$ of the potentials, and denote by ϕ the function

$$[0, 1] \ni t \mapsto \phi(t) = -\Phi(f_t \oplus g_t), \quad \text{where} \quad f_t \oplus g_t := ((1 - t)f_* + tf) \oplus ((1 - t)g_* + tg).$$

Note that ϕ is non-decreasing, convex, $\mathcal{C}^{1,1}$ and attains its minimum at 0. Moreover,

$$\phi'(t) = \int ((f_* - f) \oplus (g_* - g)) \left(1 - \frac{1}{\varepsilon} (f_t \oplus g_t - c)_+ \right) d(P \otimes Q), \quad (23)$$

and for a.e. $t \in [0, 1]$,

$$\phi''(t) = \frac{1}{\varepsilon} \int_{f_t \oplus g_t \geq c} ((f_* - f) \oplus (g_* - g))^2 d(P \otimes Q). \quad (24)$$

Here the differentiation under the integral can be justified by combining the proof of [11, Theorem 2.27] and the a.e. differentiability of $s \mapsto (s)_+$. In particular, we have the expansion

$$\begin{aligned} \phi(t) &= \phi(0) + \int_0^t \phi'(s) ds = \phi(0) + t\phi'(0) + \int_0^t \int_0^s \phi''(r) dr ds \\ &= \phi(0) + \int_0^t \int_0^s \phi''(r) dr ds, \end{aligned} \quad (25)$$

where in the last step, we used (23) and (11) and the fact that $\pi_\varepsilon \in \Pi(P, Q)$ yields

$$\phi'(0) = \int ((f_* - f) \oplus (g_* - g)) [d(P \otimes Q) - d\pi_\varepsilon] = 0.$$

In what follows, we set

$$C_{f,g} := \|f_* \oplus g_* - f \oplus g\|_\infty, \quad r_0 := \min\left(\frac{\varepsilon}{2C_{f,g}}, 1\right) \quad (26)$$

and show in Theorem 3.7 that $\|(f_* - f) \oplus (g_* - g)\|_{L^2(P \otimes Q)}^2 \leq \frac{\varepsilon}{\beta_\varepsilon} \phi''(r)$ for a.e. $r \in [0, r_0]$, for a certain constant β_ε . Note that r_0 depends on f, g only through $C_{f,g}$; this will be important below to obtain an inequality for arbitrary f, g . We mention that the choice of such r_0 is inspired by an argument of [31] in the context of entropic statistical optimal transport. Once Theorem 3.7 is shown, the following theorem will be deduced via (24).

Theorem 3.1 (PL inequality for Φ). *The dual objective $\Phi : \mathcal{H} \rightarrow \mathbb{R}$ satisfies the error bound*

$$\|f \oplus g - f_* \oplus g_*\|_{L^2(P \otimes Q)} \leq \gamma_\varepsilon \max(\|f \oplus g - f_* \oplus g_*\|_\infty, \varepsilon) \|\mathbf{D}\Phi(f \oplus g)\|_{L^2(P \otimes Q)}$$

and the PL inequality

$$\|\mathbf{D}\Phi(f \oplus g)\|_{L^2(P \otimes Q)}^2 \geq \frac{1}{\gamma_\varepsilon \max(\|f \oplus g - f_* \oplus g_*\|_\infty, \varepsilon)} (\text{QOT}_\varepsilon(P, Q) - \Phi(f \oplus g)) \quad (27)$$

for all $(f, g) \in L^\infty(\Omega) \times L^\infty(\Omega')$, where

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

In this statement, $\mathbf{D}\Phi(f \oplus g) \in \mathcal{H}$ denotes the \mathcal{H} -gradient of Φ . Its explicit form is detailed in Lemma 3.8 below.

3.1 Coordinate-wise coercivity of ϕ''

With a view toward the second derivative (24), the goal of this subsection is the coercivity of the nonnegative bilinear form

$$(u \oplus v, \tilde{u} \oplus \tilde{v}) \mapsto \int_{f_r \oplus g_r \geq c} (u \oplus v)(\tilde{u} \oplus \tilde{v}) d(P \otimes Q)$$

with respect to the first coordinate, for sufficiently small $r \in [0, 1]$.

Lemma 3.2 (Coordinate-wise coercivity). *Let r_0 be as in (26). For all $r \leq r_0$ and $(u, v) \in L^2(P) \times L^2(Q)$,*

$$\int_{f_r \oplus g_r \geq c} (u \oplus v)^2 d(P \otimes Q) \geq \alpha \text{Var}_P(u), \quad \alpha := \frac{\lambda_P^2 \inf_{y \in \Omega'} Q(B_{\frac{\varepsilon}{8L}}(y))}{\Lambda_P^2 (\lceil 8L \text{diam}(\Omega)/\varepsilon \rceil)^{d+2}},$$

where $\text{Var}_P(u) = \int [u(z) - \int u(x) dP(x)]^2 dP(z)$ denotes the variance of u under P .

The proof of Lemma 3.2 uses the following auxiliary result, which is inspired by the Bourgain–Brezis–Mironescu limits for Sobolev energies (e.g., [4, Theorem 1]).

Lemma 3.3. *For any convex and bounded set $\Omega \subset \mathbb{R}^d$, $\varrho > 0$ and $h \in L^2(\Omega)$, the operator*

$$\mathcal{E}_\varrho(h) := \int_{\Omega} \int_{B_\varrho(x) \cap \Omega} (h(x) - h(z))^2 dz dx$$

satisfies

$$\int_{\Omega} \int_{\Omega} (h(x) - h(z))^2 dx dz \leq (\lceil \text{diam}(\Omega)/\varrho \rceil)^{d+2} \mathcal{E}_\varrho(h).$$

Proof. Fix $x, z \in \Omega$ and set $x_k = x + \frac{k}{m}(z - x)$ where $m = \lceil \text{diam}(\Omega)/\varrho \rceil$. Then

$$(h(x) - h(z))^2 = \left(\sum_{k=0}^{m-1} h(x_k) - h(x_{k+1}) \right)^2 \leq m \sum_{k=0}^{m-1} (h(x_{k+1}) - h(x_k))^2$$

and integrating this inequality gives

$$\begin{aligned} & \int_{\Omega} \int_{\Omega} (h(x) - h(z))^2 dz dx \\ & \leq m \sum_{k=0}^{m-1} \underbrace{\int_{\Omega} \int_{\Omega} \left(h \left(x + \frac{k+1}{m}(z-x) \right) - h \left(x + \frac{k}{m}(z-x) \right) \right)^2 dz dx}_{=: I_k}. \end{aligned} \quad (28)$$

For fixed k , we use the change of variables

$$(u, v) = T(x, z) = \left(x + \frac{k+1}{m}(z-x), x + \frac{k}{m}(z-x) \right)$$

with

$$\begin{aligned} DT(x, z) &= A \otimes \mathbb{I}_d, \quad A := \frac{1}{m} \begin{pmatrix} m-k-1 & k+1 \\ m-k & k \end{pmatrix}, \\ \det DT(x, z) &= \det(A \otimes \mathbb{I}_d) = \det(A)^d = \left(-\frac{1}{m} \right)^d \end{aligned}$$

to get

$$I_k = m^d \int_{T(\Omega \times \Omega)} (h(u) - h(v))^2 dudv.$$

We have $T(\Omega \times \Omega) \subset \Omega \times \Omega$ by convexity of Ω . Noting that $(u, v) = T(x, z)$ satisfy $u - v = (z - x)/m$, we also have $\|u - v\| \leq \text{diam}(\Omega)/m \leq \varrho$ for all $x, z \in \Omega$. As a consequence, $I_k \leq m^d \mathcal{E}_\varrho(h)$ and thus (28) yields

$$\int_{\Omega} \int_{\Omega} (h(x) - h(z))^2 dz dx \leq m^{d+2} \mathcal{E}_\varrho(h)$$

as claimed. \square

We can now show Lemma 3.2.

Proof of Lemma 3.2. Fix $r \leq r_0$ and $(u, v) \in L^2(P) \times L^2(Q)$. For each $y \in \Omega'$ we define the section $\mathcal{T}_y^\varepsilon = \{x : f_*(x) + g_*(y) - c(x, y) \geq \frac{\varepsilon}{2}\}$ of the set $\{f_* \oplus g_* - c \geq \frac{\varepsilon}{2}\}$. As $r \leq r_0$ implies $\{f_r \oplus g_r \geq c\} \supset \{f_* \oplus g_* - c \geq \frac{\varepsilon}{2}\}$, we have

$$\begin{aligned} \mathcal{I} &:= \int_{f_r \oplus g_r \geq c} (u \oplus v)^2 d(P \otimes Q) \geq \int_{f_* \oplus g_* - c \geq \frac{\varepsilon}{2}} (u \oplus v)^2 d(P \otimes Q) \\ &= \int_{\Omega'} \int_{\mathcal{T}_y^\varepsilon} (u(x) + v(y))^2 dP(x) dQ(y) \\ &\geq \int_{\Omega'} \min_{s \in \mathbb{R}} \int_{\mathcal{T}_y^\varepsilon} (u(x) + s)^2 dP(x) dQ(y). \end{aligned}$$

For fixed y , the minimality property of the variance of u under the probability measure $P(dx)/P(\mathcal{T}_y^\varepsilon)$ yields

$$\min_{s \in \mathbb{R}} \int_{\mathcal{T}_y^\varepsilon} (u(x) + s)^2 dP(x) = \frac{1}{2P(\mathcal{T}_y^\varepsilon)} \int_{\mathcal{T}_y^\varepsilon} \int_{\mathcal{T}_y^\varepsilon} (u(x) - u(z))^2 dP(x) dP(z).$$

Using also $P(\mathcal{T}_y^\varepsilon) \leq 1$, we conclude that

$$\mathcal{I} \geq \frac{1}{2} \int_{\Omega'} \int_{\mathcal{T}_y^\varepsilon} \int_{\mathcal{T}_y^\varepsilon} (u(x) - u(z))^2 dP(x) dP(z) dQ(y). \quad (29)$$

Choose a measurable selector

$$\Omega \ni x \mapsto y(x) \in \operatorname{argmax}_{y \in \Omega'} (f_*(x) + g_*(y) - c(x, y))_+$$

and set

$$m_x = \max_{y \in \Omega'} (f_*(x) + g_*(y) - c(x, y))_+.$$

Recalling (10), we then have

$$m_x \geq \int_{\Omega'} (f_*(x) + g_*(y) - c(x, y))_+ dQ(y) = \varepsilon.$$

Since f_* , g_* and c are L -Lipschitz, we deduce

$$\begin{aligned} f_*(z) + g_*(y) - c(z, y) &\geq f_*(x) + g_*(y(x)) - c(x, y(x)) - 2L\|x - z\| - 2L\|y(x) - y\| \\ &= m_x - 2L\|x - z\| - 2L\|y(x) - y\| \\ &\geq \varepsilon - 2L\|x - z\| - 2L\|y(x) - y\| \end{aligned}$$

and hence

$$f_*(z) + g_*(y) - c(z, y) - \frac{\varepsilon}{2} \geq \frac{\varepsilon}{2} - 2L\|x - z\| - 2L\|y(x) - y\|.$$

Let $\varrho := \frac{\varepsilon}{8L}$ and $x \in \Omega$. If $z \in B_\varrho(x)$ and $y \in B_\varrho(y(x))$, then

$$f_*(z) + g_*(y) - c(z, y) - \frac{\varepsilon}{2} \geq \frac{\varepsilon}{2} - 2L\varrho - 2L\varrho = \frac{\varepsilon}{2} - \frac{\varepsilon}{4} - \frac{\varepsilon}{4} = 0, \quad (30)$$

meaning that $z \in \mathcal{T}_y^\varepsilon$. This applies in particular to $z = x$, so $x \in \mathcal{T}_y^\varepsilon$. Equivalently, for all $x, z \in \Omega$ and $y \in \Omega'$, we have

$$\mathbf{1}_{B_\varrho(y(x))}(y) \mathbf{1}_{B_\varrho(x)}(z) \leq \mathbf{1}_{\mathcal{T}_y^\varepsilon}(x) \mathbf{1}_{\mathcal{T}_y^\varepsilon}(z) \mathbf{1}_{B_\varrho(y(x))}(y).$$

Applying this in (29) gives

$$\begin{aligned} \mathcal{I} &\geq \frac{1}{2} \int_{\Omega'} \int_{\Omega} \int_{\Omega} \mathbf{1}_{\mathcal{T}_y^\varepsilon}(x) \mathbf{1}_{\mathcal{T}_y^\varepsilon}(z) (u(x) - u(z))^2 dP(x) dP(z) dQ(y) \\ &\geq \frac{1}{2} \int_{\Omega'} \int_{\Omega} \int_{\Omega} \mathbf{1}_{B_\varrho(y(x))}(y) \mathbf{1}_{B_\varrho(x)}(z) (u(x) - u(z))^2 dP(x) dP(z) dQ(y) \\ &= \frac{1}{2} \int_{\Omega} \left[Q(B_\varrho(y(x))) \int_{B_\varrho(x)} (u(x) - u(z))^2 dP(z) \right] dP(x) \\ &\geq \frac{1}{2} \left(\inf_{y \in \Omega'} Q(B_\varrho(y)) \right) \int_{\Omega} \int_{B_\varrho(x)} (u(x) - u(z))^2 dP(z) dP(x). \end{aligned}$$

Recall from Assumption 2.1 that the density ρ of P satisfies $\lambda_P \leq \rho \leq \Lambda_P$ on Ω , so that

$$\int_{\Omega} \int_{B_\varrho(x)} (u(x) - u(z))^2 dP(z) dP(x) \geq \lambda_P^2 \mathcal{E}_\varrho(u).$$

By Lemma 3.3 applied with $\varrho = \frac{\varepsilon}{8L}$,

$$\mathcal{E}_\varrho(u) \geq \frac{1}{(\lceil \text{diam}(\Omega)/\varrho \rceil)^{d+2}} \int_{\Omega} \int_{\Omega} (u(x) - u(z))^2 dx dz.$$

Moreover,

$$\int_{\Omega} \int_{\Omega} (u(x) - u(z))^2 dx dz \geq \frac{1}{\Lambda_P^2} \int_{\Omega} \int_{\Omega} (u(x) - u(z))^2 dP(x) dP(z) = \frac{2}{\Lambda_P^2} \text{Var}_P(u).$$

Combining the last four displays,

$$\mathcal{I} \geq \left(\inf_{y \in \Omega'} Q(B_\varrho(y)) \right) \frac{\lambda_P^2}{\Lambda_P^2} \frac{1}{(\lceil \text{diam}(\Omega)/\varrho \rceil)^{d+2}} \text{Var}_P(u),$$

which gives the claim after recalling that $\varrho = \frac{\varepsilon}{8L}$. \square

Remark 3.4. For a given $y \in \Omega'$, applying the argument around (30) with x replaced by $x(y) \in \arg\max_{x \in \Omega} (f_*(x) + g_*(y) - c(x, y))_+$ and $y(x)$ replaced by y yields that $z \in \mathcal{T}_y^\varepsilon$ for any $z \in B_\varrho(x(y)) \cap \Omega$. That is, $(B_\varrho(x(y)) \cap \Omega) \subset \mathcal{T}_y^\varepsilon$. Symmetrically, we have for any $x \in \Omega$ that $(B_\varrho(y(x)) \cap \Omega') \subset \mathcal{S}_x^\varepsilon$ for the x -section $\mathcal{S}_x^\varepsilon := \{y : f_*(x) + g_*(y) - c(x, y) \geq \frac{\varepsilon}{2}\}$. Recalling that $\varrho = \frac{\varepsilon}{8L}$, we have in particular

$$\inf_{y \in \Omega'} Q(B_{\frac{\varepsilon}{8L}}(y)) \leq Q(\mathcal{S}_x^\varepsilon), \quad \inf_{x \in \Omega} P(B_{\frac{\varepsilon}{8L}}(x)) \leq P(\mathcal{T}_y^\varepsilon). \quad (31)$$

3.2 Uniform coercivity of ϕ''

Fix $r \in [0, 1]$. By the Riesz representation theorem, there exists a unique bounded linear operator $\mathbb{M} : \mathcal{H} \rightarrow \mathcal{H}$ such that

$$\langle \tilde{u} \oplus \tilde{v}, \mathbb{M}(u \oplus v) \rangle_{\mathcal{H}} = \int_{f_r \oplus g_r \geq c} (\tilde{u} \oplus \tilde{v})(u \oplus v) d(P \otimes Q) \quad (32)$$

for all $(u \oplus v), (\tilde{u} \oplus \tilde{v}) \in \mathcal{H}$. Denote by $\mathcal{S}_{r,x}$ and $\mathcal{T}_{r,y}$ the sections of the set $\mathcal{R}_r := \{f_r \oplus g_r \geq c\}$,

$$\mathcal{S}_{r,x} = \{y \in \Omega' : f_r(x) + g_r(y) \geq c(x, y)\}, \quad \mathcal{T}_{r,y} = \{x \in \Omega : f_r(x) + g_r(y) \geq c(x, y)\}.$$

Moreover, define

$$m(u \oplus v) := \int_{f_r \oplus g_r \geq c} u \oplus v d(P \otimes Q) \in \mathbb{R}.$$

Lemma 3.5. *The operator $\mathbb{M} : \mathcal{H} \rightarrow \mathcal{H}$ has the explicit representation*

$$\mathbb{M}(u \oplus v) = \mathbb{M}_1(u \oplus v) \oplus \mathbb{M}_2(u \oplus v), \quad (33)$$

where

$$\begin{aligned} \mathbb{M}_1(u \oplus v) &:= uQ(\mathcal{S}_{r,(\cdot)}) + \int_{\mathcal{S}_{r,(\cdot)}} v dQ - \frac{m(u \oplus v)}{2} \in L^2(P), \\ \mathbb{M}_2(u \oplus v) &:= vP(\mathcal{T}_{r,(\cdot)}) + \int_{\mathcal{T}_{r,(\cdot)}} u dP - \frac{m(u \oplus v)}{2} \in L^2(Q). \end{aligned}$$

Proof. The definition (32) can be stated as

$$\langle \tilde{w}, \mathbb{M}w \rangle_{L^2(P \otimes Q)} = \langle \tilde{w}, \mathbf{1}_{\mathcal{R}_r} w \rangle_{L^2(P \otimes Q)} \quad \text{for all } \tilde{w}, w \in \mathcal{H}.$$

This implies that $\mathbb{M}w$ is the orthogonal projection of $\mathbf{1}_{\mathcal{R}_r} w \in L^2(P \otimes Q)$ onto the subspace $\mathcal{H} \subset L^2(P \otimes Q)$. For arbitrary $h \in L^2(P \otimes Q)$, define

$$\begin{aligned} h_P(x) &:= \int h(x, y) dQ(y), \quad x \in \Omega, \\ h_Q(y) &:= \int h(x, y) dP(x), \quad y \in \Omega', \\ \bar{h} &:= \int h(x, y) d(P \otimes Q)(x, y). \end{aligned}$$

We verify below that the orthogonal projection of $h \in L^2(P \otimes Q)$ onto \mathcal{H} is

$$\text{proj } h := h_P \oplus h_Q - \bar{h}.$$

Specializing to $h := \mathbf{1}_{\mathcal{R}_r}(u \oplus v)$, we have

$$\text{proj } h = \left(h_P - \frac{1}{2} \bar{h} \right) \oplus \left(h_Q - \frac{1}{2} \bar{h} \right) = \mathbb{M}_1(u \oplus v) \oplus \mathbb{M}_2(u \oplus v),$$

showing the claim.

It remains to prove that $\text{proj } h$ is indeed the orthogonal projection. Given $h \in L^2(P \otimes Q)$, clearly $h_P \in L^2(P)$, $h_Q \in L^2(Q)$ and $\bar{h} \in \mathbb{R}$, so that $\text{proj } h \in \mathcal{H}$. To complete the proof, we need to show for all $\tilde{u} \oplus \tilde{v} \in \mathcal{H}$ that

$$\begin{aligned} 0 &= \langle \tilde{u} \oplus \tilde{v}, h - \text{proj } h \rangle_{L^2(P \otimes Q)} \\ &= \int_{\Omega \times \Omega'} (\tilde{u}(x) + \tilde{v}(y))(h(x, y) - h_P(x) - h_Q(y) + \bar{h}) dP(x)dQ(y) =: I_1 + I_2, \end{aligned}$$

where

$$I_1 := \int \tilde{u}(x)(h(x, y) - h_P(x) - h_Q(y) + \bar{h}) dP(x)dQ(y)$$

and I_2 is defined analogously with \tilde{v} instead of \tilde{u} . Next, we show that $I_1 = I_2 = 0$. By Fubini's theorem,

$$\int \tilde{u}(x)h(x, y) dP(x)dQ(y) = \int \tilde{u}(x) \left(\int h(x, y) dQ(y) \right) dP(x) = \int \tilde{u}(x)h_P(x) dP(x).$$

Moreover,

$$\int \tilde{u}(x)h_Q(y) dP(x)dQ(y) = \left(\int \tilde{u}(x) dP(x) \right) \left(\int h_Q(y) dQ(y) \right) = \bar{h} \int \tilde{u}(x) dP(x).$$

Expanding I_1 into four integrals, this shows that the first two and the last two integrals cancel, so that $I_1 = 0$. Similarly, $I_2 = 0$, completing the proof. \square

The next lemma is an important technical step for the derivation of our main result. While \mathbb{M} is not compact, the lemma uses the explicit form of \mathbb{M} to establish that the minimum Rayleigh quotient is an eigenvalue. The proof technique is adapted from [16, Proposition 4.1].

Lemma 3.6. *For any $r \in [0, 1]$, the operator $\mathbb{M} : \mathcal{H} \rightarrow \mathcal{H}$ is bounded, self-adjoint and positive. Let*

$$\lambda_0 := \inf_{\|u \oplus v\|_{\mathcal{H}}=1} \langle u \oplus v, \mathbb{M}(u \oplus v) \rangle_{\mathcal{H}} \quad (34)$$

and suppose there exists a constant $\kappa_0 > 0$ such that

$$Q(\mathcal{S}_{r,(\cdot)}) - \lambda_0 \geq \kappa_0 \quad P\text{-a.s.}, \quad P(\mathcal{T}_{r,(\cdot)}) - \lambda_0 \geq \kappa_0 \quad Q\text{-a.s.} \quad (35)$$

Then λ_0 is an eigenvalue of \mathbb{M} . As a consequence, the infimum in (34) is attained and λ_0 is the smallest eigenvalue of \mathbb{M} .

Proof. Step 1. We readily see from the definition (32) that \mathbb{M} is bounded, self-adjoint and positive on the Hilbert space \mathcal{H} . For the remainder of the proof we abbreviate $\lambda := \lambda_0$, $\langle \cdot, \cdot \rangle = \langle \cdot, \cdot \rangle_{\mathcal{H}}$ and $\| \cdot \| = \| \cdot \|_{\mathcal{H}}$, whereas $\| \cdot \|_{\text{op}}$ denotes the operator norm. Noting

$$0 \leq \langle w, \mathbb{M}w \rangle = \int_{\{f_r \oplus g_r \geq c\}} w^2 d(P \otimes Q) \leq \int w^2 d(P \otimes Q) = \|w\|^2,$$

we have $0 \leq \mathbb{M} \leq \mathbb{I}$, where \mathbb{I} denotes the identity. Define the linear operator $\mathbb{L} : \mathcal{H} \rightarrow \mathcal{H}$ by

$$\mathbb{L} := \mathbb{I} - \mathbb{M}.$$

Then \mathbb{L} is also bounded, self-adjoint and positive, so that its operator norm can be expressed via the Rayleigh quotient as

$$\alpha^+ := \sup_{\|w\|=1} \langle \mathbb{L}w, w \rangle = \|\mathbb{L}\|_{\text{op}}. \quad (36)$$

For $w \in \mathcal{H}$ with $\|w\| = 1$, clearly $\langle \mathbb{L}w, w \rangle = 1 - \langle \mathbb{M}w, w \rangle$, so that $\lambda = 1 - \alpha^+$.

Step 2. By the definition of α^+ , there exists a sequence $w_n = u_n \oplus v_n \in \mathcal{H}$ with $\|w_n\| = 1$ and $\langle \mathbb{L}w_n, w_n \rangle \rightarrow \alpha^+$. As $\|\mathbb{L}\|_{\text{op}} = \alpha^+$ implies $\|\mathbb{L}w_n\| \leq \alpha^+$ for all n , we have

$$\begin{aligned} \|\lambda w_n - \mathbb{M}w_n\|^2 &= \|(\mathbb{L} - \alpha^+ \mathbb{I})w_n\|^2 = \|\mathbb{L}w_n\|^2 + (\alpha^+)^2 - 2\alpha^+ \langle \mathbb{L}w_n, w_n \rangle \\ &\leq 2\alpha^+(\alpha^+ - \langle \mathbb{L}w_n, w_n \rangle) \rightarrow 0. \end{aligned} \quad (37)$$

Below, we prove that w_n converges strongly to a limit $w \neq 0$ (along a subsequence). Once that is shown, we have $\lambda w = \mathbb{M}w$ and $\|w\| = 1$, completing the proof.

Step 3. We may choose representatives $(u_n, v_n) \in L^2(P) \times L^2(Q)$ of $w_n = u_n \oplus v_n$ with $\int u_n dP = \int v_n dQ$. Recall the operators \mathbb{M}_1 and \mathbb{M}_2 with $\mathbb{M}(u \oplus v) = \mathbb{M}_1(u, v) \oplus \mathbb{M}_2(u, v)$ from Lemma 3.5. Fubini's theorem shows that this split gives a balanced normalization,

$$\int \mathbb{M}_1(u, v) dP = \int \mathbb{M}_2(u, v) dQ.$$

Note also that if $(f_n, g_n)_{n \geq 0} \subset L^2(P) \times L^2(Q)$ satisfy $\int f_n dP = \int g_n dQ$ for $n \geq 0$, then the convergence $f_n \oplus g_n \rightarrow f_0 \oplus g_0$ in $L^2(P \otimes Q)$ already implies the separate convergences $f_n \rightarrow f_0$ in $L^2(P)$ and $g_n \rightarrow g_0$ in $L^2(Q)$, thanks to the fact that

$$f_n(\cdot) = \int f_n(\cdot) \oplus g_n(y) dQ(y) - \frac{1}{2} \int f_n \oplus g_n d(P \otimes Q)$$

and analogously for g_n . As a consequence, (37) and $\mathbb{M}(u \oplus v) = \mathbb{M}_1(u, v) \oplus \mathbb{M}_2(u, v)$ imply

$$\begin{aligned} \|u_n(\lambda - Q(\mathcal{S}_{r,(\cdot)})) - \mathbb{A}_1 v_n + \frac{1}{2} m(u_n \oplus v_n)\|_{L^2(P)} &\rightarrow 0, \\ \|v_n(\lambda - P(\mathcal{T}_{r,(\cdot)})) - \mathbb{A}_2 u_n + \frac{1}{2} m(u_n \oplus v_n)\|_{L^2(Q)} &\rightarrow 0, \end{aligned} \quad (38)$$

where we have abbreviated the integral terms with the operators

$$\begin{aligned} \mathbb{A}_1 : L^2(Q) &\rightarrow L^2(P), & \mathbb{A}_1 v(x) &:= \int_{\mathcal{S}_{r,x}} v(y) dQ(y), \\ \mathbb{A}_2 : L^2(P) &\rightarrow L^2(Q), & \mathbb{A}_2 u(y) &:= \int_{\mathcal{T}_{r,y}} u(x) dP(x). \end{aligned}$$

These are integral operators with kernels in $L^2(P \otimes Q)$ and hence compact (cf. [5, Theorem 6.12]).

As the sequences $(u_n) \subset L^2(P)$ and $(v_n) \subset L^2(Q)$ are bounded, the Banach–Alaoglu theorem yields (after passing to another subsequence) the weak convergences

$$u_n \rightharpoonup u \text{ in } L^2(P), \quad v_n \rightharpoonup v \text{ in } L^2(Q),$$

for some $u \in L^2(P)$ and $v \in L^2(Q)$. By the compactness of \mathbb{A}_1 and \mathbb{A}_2 , this implies

$$\mathbb{A}_1 v_n \rightarrow \mathbb{A}_1 v \text{ in } L^2(P), \quad \mathbb{A}_2 u_n \rightarrow \mathbb{A}_2 u \text{ in } L^2(Q).$$

Moreover, m is clearly weakly continuous, so that $m(u_n \oplus v_n) \rightarrow m(u \oplus v)$. We can now revisit (38): as $\mathbb{A}_1 v_n \rightarrow \mathbb{A}_1 v$ strongly and $m(u_n \oplus v_n) \rightarrow m(u \oplus v)$ are constants, $u_n(\lambda - Q(\mathcal{S}_{r,(\cdot)}))$ must also converge strongly. Multiplication with $1/(\lambda - Q(\mathcal{S}_{r,(\cdot)}))$ is a bounded operator thanks to the uniform bound (35), so that u_n also converges strongly; the limit must coincide with the weak limit u . Similarly, $v_n \rightarrow v$. As noted after (37), this completes the proof. \square

Now we show the main result of this section.

Theorem 3.7 (Uniform coercivity of \mathbb{M}). *Let r_0 be as in (26) and δ_P as in Remark 2.2. Set*

$$\beta_\varepsilon := \frac{1}{4}\kappa\alpha, \quad \text{where } \kappa := \delta_P \min\left(\frac{\varepsilon}{8L}, 1\right)^d \leq 1, \quad \alpha := \frac{\lambda_P^2}{\Lambda_P^2} \frac{\inf_{y \in \Omega'} Q(B_{\frac{\varepsilon}{8L}}(y))}{(\lceil 8L \text{diam}(\Omega)/\varepsilon \rceil)^{d+2}} \leq 1.$$

Then for all $r \leq r_0$ and $(u, v) \in L^2(P) \times L^2(Q)$, the operator $\mathbb{M} = \mathbb{M}(r)$ of (32) satisfies

$$\langle u \oplus v, \mathbb{M}(u \oplus v) \rangle_{L^2(P \otimes Q)} \geq \beta_\varepsilon \|u \oplus v\|_{L^2(P \otimes Q)}^2, \quad (39)$$

and as a consequence, for a.e. $r \leq r_0$,

$$\|(f_* - f) \oplus (g_* - g)\|_{L^2(P \otimes Q)}^2 \leq \frac{\varepsilon}{\beta_\varepsilon} \phi''(r). \quad (40)$$

Proof. Step 1. We first derive lower bounds on the sections $\mathcal{T}_{r,y}$ and $\mathcal{S}_{r,x}$ of $\{f_r \oplus g_r \geq c\}$. Recall from Remark 3.4 and from the beginning of the proof of Lemma 3.2 that

$$B_{\frac{\varepsilon}{8L}}(x(y)) \cap \Omega \subset \mathcal{T}_y^\varepsilon := \{x : f_*(x) + g_*(y) - c(x, y) \geq \frac{\varepsilon}{2}\} \subset \mathcal{T}_{r,y}.$$

In view of Remark 2.2, it follows that

$$P(\mathcal{T}_{r,y}) \geq \delta_P \min\left(\frac{\varepsilon}{8L}, 1\right)^d = \kappa \quad \text{for all } y \in \Omega'. \quad (41)$$

Analogously, $\mathcal{S}_x^\varepsilon := \{y : f_*(x) + g_*(y) - c(x, y) \geq \frac{\varepsilon}{2}\} \subset \mathcal{S}_{r,x}$, so that (31) yields

$$Q(\mathcal{S}_{r,x}) \geq \inf_{y \in \Omega'} Q(B_{\frac{\varepsilon}{8L}}(y)) =: q_0 \quad \text{for all } x \in \Omega. \quad (42)$$

Step 2. Let λ_0 be as in Lemma 3.6. If $\lambda_0 \geq \beta_\varepsilon$, then (39) holds and we are done. We may thus assume that $\lambda_0 \leq \beta_\varepsilon$. Noting that $\kappa, \alpha \leq 1$ and $\alpha \leq q_0$, the definition of β_ε , (41) and (42) then yield

$$\begin{aligned} P(\mathcal{T}_{r,y}) - \lambda_0 &\geq \frac{3}{4}P(\mathcal{T}_{r,y}) \geq \frac{3}{4}\kappa > 0 \quad \text{for all } y \in \Omega', \\ Q(\mathcal{S}_{r,x}) - \lambda_0 &\geq \frac{3}{4}Q(\mathcal{S}_{r,x}) \geq \frac{3}{4}q_0 > 0 \quad \text{for all } x \in \Omega. \end{aligned} \quad (43)$$

In particular, we may assume that (35) holds.

Step 3. Hence, by Lemma 3.6, it suffices to show that the minimal eigenvalue λ_0 of \mathbb{M} satisfies $\lambda_0 \geq \beta_\varepsilon$. Let $w = u \oplus v \neq 0$ be an eigenvector of \mathbb{M} with associated eigenvalue $\lambda_0 \geq 0$ and suppose for contradiction that

$$\lambda_0 < \beta_\varepsilon. \quad (44)$$

We may choose representatives $(u, v) \in L^2(P) \times L^2(Q)$ with the centering $\int u dP = 0$. The eigenvalue equation $\lambda_0(u \oplus v) = \mathbb{M}(u \oplus v)$, the centering of u and Lemma 3.5 imply that

$$\lambda_0 u = uQ(\mathcal{S}_{r,(\cdot)}) + \int_{\mathcal{S}_{r,(\cdot)}} v dQ - \int \left(uQ(\mathcal{S}_{r,(\cdot)}) + \int_{\mathcal{S}_{r,(\cdot)}} v dQ \right) dP$$

and

$$\lambda_0 v = vP(\mathcal{T}_{r,(\cdot)}) + \int_{\mathcal{T}_{r,(\cdot)}} u dP. \quad (45)$$

On the other hand, the eigenvalue equation, Lemma 3.2 and the centering of u yield

$$\lambda_0 \|u \oplus v\|_{L^2(P \otimes Q)}^2 = \langle u \oplus v, \mathbb{M}(u \oplus v) \rangle_{L^2(P \otimes Q)} \geq \alpha \text{Var}_P(u) = \alpha \|u\|_{L^2(P)}^2.$$

The centering also implies $\|u \oplus v\|_{L^2(P \otimes Q)}^2 = \|u\|_{L^2(P)}^2 + \|v\|_{L^2(Q)}^2$, and we conclude that

$$\lambda_0 \geq \alpha \frac{\|u\|_{L^2(P)}^2}{\|u\|_{L^2(P)}^2 + \|v\|_{L^2(Q)}^2}. \quad (46)$$

Next, we bound $\|v\|_{L^2(Q)}^2$. Note that (45) yields the inequality

$$|v| \leq \frac{\int_{\mathcal{T}_{r,(\cdot)}} |u| dP}{|\lambda_0 - P(\mathcal{T}_{r,(\cdot)})|}.$$

Using (43), we deduce

$$|v| \leq \frac{\int_{\mathcal{T}_{r,(\cdot)}} |u| dP}{P(\mathcal{T}_{r,(\cdot)}) - \lambda_0} \leq \frac{\int_{\mathcal{T}_{r,(\cdot)}} |u| dP}{\frac{3}{4}P(\mathcal{T}_{r,(\cdot)})}.$$

Taking squares, applying Jensen's inequality for $P(dx)/P(\mathcal{T}_{r,y})$, integrating with respect to Q , and using (41) again, we get

$$\|v\|_{L^2(Q)}^2 \leq \frac{16}{9} \int \frac{\int_{\mathcal{T}_{r,(\cdot)}} u^2 dP}{P(\mathcal{T}_{r,(\cdot)})} dQ \leq \frac{16}{9} \frac{\|u\|_{L^2(P)}^2}{\kappa} \leq 2 \frac{\|u\|_{L^2(P)}^2}{\kappa}.$$

As $u \oplus v \neq 0$, this shows in particular that $u \neq 0$. Therefore, substituting the last display into (46) and using $\kappa \leq 1$ yields

$$\lambda_0 \geq \alpha \frac{\|u\|_{L^2(P)}^2}{\|u\|_{L^2(P)}^2 + 2\|u\|_{L^2(P)}^2/\kappa} = \alpha \frac{\kappa}{\kappa + 2} \geq \frac{1}{3} \alpha \kappa \geq \beta_\varepsilon,$$

contradicting (44) and completing the proof of (39).

Step 4. Finally, set $w_* := (f_* - f) \oplus (g_* - g)$; then (24) and (39) yield

$$\phi''(r) = \frac{1}{\varepsilon} \int_{\{f_r \oplus g_r \geq c\}} w_*^2 d(P \otimes Q) = \frac{1}{\varepsilon} \langle w_*, \mathbb{M}w_* \rangle_{L^2(P \otimes Q)} \geq \frac{\beta_\varepsilon}{\varepsilon} \|w_*\|_{L^2(P \otimes Q)}^2,$$

which is (40). □

3.3 Conclusion of the proof of Theorem 3.1

We can now complete the proof of the error bound and the PL inequality for Φ .

Proof of Theorem 3.1. Using (40) in (25) yields

$$\phi(r_0) \geq \phi(0) + \frac{r_0^2 \beta_\varepsilon}{2\varepsilon} \|(f_* - f) \oplus (g_* - g)\|_{L^2(P \otimes Q)}^2.$$

On the other hand, the convexity of ϕ and the resulting monotonicity of ϕ' show

$$\phi(r_0) \leq \phi(0) + r_0 \phi'(r_0) \leq \phi(0) + r_0 \phi'(1).$$

Together, we obtain

$$\begin{aligned} \frac{r_0^2 \beta_\varepsilon}{2\varepsilon} \|(f_* - f) \oplus (g_* - g)\|_{L^2(P \otimes Q)}^2 &\leq \phi(r_0) - \phi(0) \leq r_0 \phi'(1) \\ &= r_0 \langle \mathbf{D}\Phi(f \oplus g), f_* \oplus g_* - f \oplus g \rangle_{L^2(P \otimes Q)} \leq r_0 \|\mathbf{D}\Phi(f \oplus g)\|_{L^2(P \otimes Q)} \|(f_* - f) \oplus (g_* - g)\|_{L^2(P \otimes Q)} \end{aligned}$$

and hence

$$\|(f_* - f) \oplus (g_* - g)\|_{L^2(P \otimes Q)} \leq \frac{2\varepsilon}{r_0 \beta_\varepsilon} \|\mathbf{D}\Phi(f \oplus g)\|_{L^2(P \otimes Q)}. \quad (47)$$

Noting that

$$\frac{2\varepsilon}{r_0} = \frac{2\varepsilon}{\min\left(\frac{\varepsilon}{2C_{f,g}}, 1\right)} = \frac{4}{\min(1/C_{f,g}, 2/\varepsilon)} = 4 \max(C_{f,g}, \varepsilon/2) \leq 4 \max(C_{f,g}, \varepsilon)$$

and

$$\beta_\varepsilon^{-1} = 4\kappa^{-1}\alpha^{-1} = 4 \left(\delta_P^{-1} \max\left(\frac{8L}{\varepsilon}, 1\right) \right)^d \frac{\Lambda_P^2 (\lceil 8L \text{diam}(\Omega)/\varepsilon \rceil)^{d+2}}{\lambda_P^2 \inf_{y \in \Omega'} Q(B_{\frac{\varepsilon}{8L}}(y))},$$

this gives the first claim of Theorem 3.1. Moreover, concavity of Φ yields

$$\begin{aligned} \text{QOT}_\varepsilon(P, Q) - \Phi(f \oplus g) &= \Phi(f_* \oplus g_*) - \Phi(f \oplus g) \leq \langle \mathbf{D}\Phi(f \oplus g), f_* \oplus g_* - f \oplus g \rangle_{L^2(P \otimes Q)} \\ &\leq \|\mathbf{D}\Phi(f \oplus g)\|_{L^2(P \otimes Q)} \|(f_* - f) \oplus (g_* - g)\|_{L^2(P \otimes Q)} \end{aligned}$$

and now bounding $\|(f_* - f) \oplus (g_* - g)\|_{L^2(P \otimes Q)}$ by (47) yields the second claim. \square

3.4 Proof of Theorem 2.5

Finally, we translate the result of Theorem 3.1 from Φ to Γ , that is, from the space \mathcal{H} to $L^2(P) \times L^2(Q)$.

Lemma 3.8. *Let $(f, g) \in L^2(P) \times L^2(Q)$ and*

$$I_0 := \int \left(1 - \frac{1}{\varepsilon} (f \oplus g - c)_+ \right) d(P \otimes Q).$$

The gradients of $\Phi : \mathcal{H} \rightarrow \mathbb{R}$ and $\Gamma : L^2(P) \times L^2(Q) \rightarrow \mathbb{R}$ are related by

$$\mathbf{D}\Phi(f \oplus g) = \mathbf{D}_1\Phi(f \oplus g) \oplus \mathbf{D}_2\Phi(f \oplus g), \quad \text{where} \quad \mathbf{D}_i\Phi(f \oplus g) = \mathbf{D}_i\Gamma(f, g) - \frac{1}{2}I_0 \quad (48)$$

and the explicit form of $\mathbf{D}_i\Gamma(f, g)$ is given in (9). As a consequence,

$$\|\mathbf{D}\Phi(f \oplus g)\|_{L^2(P \otimes Q)}^2 = \|\mathbf{D}\Gamma(f, g)\|_{L^2(P) \times L^2(Q)}^2 - I_0^2 \leq \|\mathbf{D}\Gamma(f, g)\|_{L^2(P) \times L^2(Q)}^2.$$

Proof. For all $(f, g), (u, v) \in L^2(P) \times L^2(Q)$, we have

$$\begin{aligned} \langle \mathbf{D}\Phi(f \oplus g), u \oplus v \rangle_{L^2(P \otimes Q)} &= \int (u \oplus v) \left(1 - \frac{1}{\varepsilon}(f \oplus g - c)_+ \right) d(P \otimes Q) \\ &= \langle 1 - \frac{1}{\varepsilon}(f \oplus g - c)_+, u \oplus v \rangle_{L^2(P \otimes Q)} \\ &= \langle \mathbf{D}\Gamma(f, g), (u, v) \rangle_{L^2(P) \times L^2(Q)}. \end{aligned}$$

Now (48) follows as in the proof of Lemma 3.5. Moreover, using the above identity with $u = \mathbf{D}_1\Phi(f \oplus g)$ and $v = \mathbf{D}_2\Phi(f \oplus g)$ yields

$$\begin{aligned} \|\mathbf{D}\Phi(f \oplus g)\|_{L^2(P \otimes Q)}^2 &= \|\mathbf{D}_1\Gamma(f, g)\|_{L^2(P)}^2 + \|\mathbf{D}_2\Gamma(f, g)\|_{L^2(Q)}^2 \\ &\quad - \frac{1}{2}\langle \mathbf{D}_1\Gamma(f, g), I_0 \rangle_{L^2(P)} - \frac{1}{2}\langle \mathbf{D}_2\Gamma(f, g), I_0 \rangle_{L^2(Q)} \\ &= \|\mathbf{D}\Gamma(f, g)\|_{L^2(P) \times L^2(Q)}^2 - I_0^2. \quad \square \end{aligned}$$

Proof of Theorem 2.5. Combine Theorem 3.1 with $\Gamma(f, g) = \Phi(f \oplus g)$ and Lemma 3.8. \square

4 Proofs of linear convergence

We first record several Lipschitz constants associated with $\mathbf{D}\Gamma$. Recall the components $\mathbf{D}_1\Gamma(f, g) \in L^2(P)$ and $\mathbf{D}_2\Gamma(f, g) \in L^2(Q)$ from (9).

Lemma 4.1. *The gradient $\mathbf{D}\Gamma : L^2(P) \times L^2(Q) \rightarrow L^2(P) \times L^2(Q)$ is $\frac{2}{\varepsilon}$ -Lipschitz. Moreover, $\mathbf{D}_i\Gamma(f, \cdot)$ and $\mathbf{D}_i\Gamma(\cdot, g)$ are $\frac{1}{\varepsilon}$ -Lipschitz for $i = 1, 2$, for all $(f, g) \in L^2(P) \times L^2(Q)$.*

Proof. For $(f, g), (u, v) \in L^2(P) \times L^2(Q)$, the inequality $|(t)_+ - (s)_+| \leq |t - s|$ yields

$$\begin{aligned} |\mathbf{D}_1\Gamma(f, g) - \mathbf{D}_1\Gamma(u, v)| &\leq \frac{1}{\varepsilon} \int \left| (u(\cdot) + v(y) - c(\cdot, y))_+ - (f(\cdot) + g(y) - c(\cdot, y))_+ \right| dQ(y) \\ &\leq \frac{1}{\varepsilon} \left\{ |f(\cdot) - u(\cdot)| + \int |g(y) - v(y)| dQ(y) \right\}. \end{aligned}$$

Using $(a + b)^2 \leq 2a^2 + 2b^2$ and Jensen's inequality, we deduce

$$\begin{aligned} \|\mathbf{D}_1\Gamma(f, g) - \mathbf{D}_1\Gamma(u, v)\|_{L^2(P)}^2 &\leq \frac{2}{\varepsilon^2} \left(\|f - u\|_{L^2(P)}^2 + \|g - v\|_{L^2(Q)}^2 \right) \\ &= \frac{2}{\varepsilon^2} \|(f, g) - (u, v)\|_{L^2(P) \times L^2(Q)}^2. \quad (49) \end{aligned}$$

Note that if $f = u$ or $g = v$, the factor 2 is unnecessary, showing the claimed $\frac{1}{\varepsilon}$ -Lipschitz property of $\mathbf{D}_1\Gamma(f, \cdot)$ and $\mathbf{D}_1\Gamma(\cdot, g)$. On the other hand, combining (49) with its analogue for $\mathbf{D}_2\Gamma$ yields

$$\|\mathbf{D}\Gamma(f, g) - \mathbf{D}\Gamma(u, v)\|_{L^2(P) \times L^2(Q)}^2 \leq \frac{4}{\varepsilon^2} \|(f, g) - (u, v)\|_{L^2(P) \times L^2(Q)}^2. \quad \square$$

Remark 4.2. For any bounded functions $f : \Omega \rightarrow \mathbb{R}$ and $g : \Omega' \rightarrow \mathbb{R}$, we note that

$$\|f \oplus g\|_\infty = \inf_{a \in \mathbb{R}} (\|f - a\|_\infty + \|g + a\|_\infty).$$

Recall that if (f_*, g_*) are potentials, then $(f_* - a, g_* + a)$ are also potentials, for any $a \in \mathbb{R}$. Thus, if some inequality

$$A(f, g) \leq B(f, g, \|f - f_*\|_\infty + \|g - g_*\|_\infty)$$

has been established for arbitrary potentials (f_*, g_*) , we can already conclude that

$$A(f, g) \leq B(f, g, \|f \oplus g - f_* \oplus g_*\|_\infty).$$

4.1 Proof of Corollary 2.6 (gradient ascent)

This proof is standard; we state the argument for the sake of completeness.

Proof of Corollary 2.6. The bound (15) for the iterates is shown in [16, Lemma 5.2]. (The norm in that reference is defined differently but has the same value by Remark 4.2. Moreover, the reference assumes that c is the quadratic cost, but the proof applies to any bounded c .)

By Lemma 4.1, $\text{D}\Gamma$ is $\frac{2}{\varepsilon}$ -Lipschitz on $L^2(P) \times L^2(Q)$, hence Γ is $\frac{2}{\varepsilon}$ -smooth. In particular,

$$\Gamma(f + u, g + v) \geq \Gamma(f, g) + \langle \text{D}\Gamma(f, g), (u, v) \rangle_{L^2(P) \times L^2(Q)} - \frac{1}{\varepsilon} \|(u, v)\|_{L^2(P) \times L^2(Q)}^2. \quad (50)$$

We apply (50) with $(f, g) = (f_n, g_n)$ and $(u, v) = \eta \text{D}\Gamma(f_n, g_n)$. Using the update rule (14) and $\eta \in (0, \varepsilon)$, we obtain

$$\begin{aligned} \Gamma(f_{n+1}, g_{n+1}) &\geq \Gamma(f_n, g_n) + \eta \|\text{D}\Gamma(f_n, g_n)\|_{L^2(P) \times L^2(Q)}^2 - \frac{\eta^2}{\varepsilon} \|\text{D}\Gamma(f_n, g_n)\|_{L^2(P) \times L^2(Q)}^2 \\ &= \Gamma(f_n, g_n) + \eta \left(1 - \frac{\eta}{\varepsilon}\right) \|\text{D}\Gamma(f_n, g_n)\|_{L^2(P) \times L^2(Q)}^2. \end{aligned}$$

Subtracting this from the optimal value $\text{QOT}_\varepsilon(P, Q)$ gives

$$\Delta_{n+1} \leq \Delta_n - \eta \left(1 - \frac{\eta}{\varepsilon}\right) \|\text{D}\Gamma(f_n, g_n)\|_{L^2(P) \times L^2(Q)}^2. \quad (51)$$

Using the PL inequality (12) at (f_n, g_n) and the bound (15),

$$\|\text{D}\Gamma(f_n, g_n)\|_{L^2(P) \times L^2(Q)}^2 \geq \frac{1}{\gamma_\varepsilon M} \Delta_n, \quad M := \max(2\|f_0 \oplus g_0 - f_* \oplus g_*\|_\infty, \varepsilon),$$

hence we conclude that

$$\Delta_{n+1} \leq \Delta_n - \eta \left(1 - \frac{\eta}{\varepsilon}\right) \frac{1}{\gamma_\varepsilon M} \Delta_n = (1 - q) \Delta_n, \quad q := \frac{1}{\gamma_\varepsilon M} \eta \left(1 - \frac{\eta}{\varepsilon}\right).$$

Iterating this estimate yields $\Delta_n \leq (1 - q)^n \Delta_0$.

It remains to prove the bound for the iterates in $L^2(P \otimes Q)$. By the error bound in Theorem 2.5 and the definition of M ,

$$\|f_n \oplus g_n - f_* \oplus g_*\|_{L^2(P \otimes Q)}^2 \leq \gamma_\varepsilon^2 M^2 \|\text{D}\Gamma(f_n, g_n)\|_{L^2(P) \times L^2(Q)}^2. \quad (52)$$

On the other hand, (51) implies

$$\eta\left(1 - \frac{\eta}{\varepsilon}\right) \|\mathrm{D}\Gamma(f_n, g_n)\|_{L^2(P) \times L^2(Q)}^2 \leq \Delta_n - \Delta_{n+1} \leq \Delta_n \leq (1 - q)^n \Delta_0.$$

Substituting this estimate into (52), we conclude that

$$\|f_n \oplus g_n - f_* \oplus g_*\|_{L^2(P \otimes Q)}^2 \leq \frac{\gamma_\varepsilon^2 M^2}{\eta(1 - \eta/\varepsilon)} \Delta_0 (1 - q)^n. \quad \square$$

4.2 Proof of Corollary 2.7 (coordinate ascent)

Again, the proof has two steps. The first step is to bound the iterates. Once that is achieved, the second step follows the standard argument for deriving linear convergence of coordinate ascent from a PL inequality.

Proof of Corollary 2.7. Step 1. Set

$$\mathcal{S}_x := \{y \in \Omega' : f_*(x) + g_*(y) \geq c(x, y)\}, \quad \mathcal{S}_x^n := \{y \in \Omega' : f_n(x) + g_n(y) \geq c(x, y)\}. \quad (53)$$

Using (10) and the definition (17) of f_n , as well as the inequality $(a)_+ - (b)_+ \leq (a - b)\mathbf{1}_{a \geq 0}$, we have for every $x \in \Omega$ that

$$\begin{aligned} 0 &= \int (f_*(x) + g_*(y) - c(x, y))_+ - (f_n(x) + g_n(y) - c(x, y))_+ dQ(y) \\ &\leq \int_{\mathcal{S}_x} f_*(x) + g_*(y) - f_n(x) - g_n(y) dQ(y) \end{aligned}$$

and hence

$$(f_n(x) - f_*(x))Q(\mathcal{S}_x) \leq \int_{\mathcal{S}_x} g_*(y) - g_n(y) dQ(y) \leq \|g_n - g_*\|_\infty Q(\mathcal{S}_x).$$

Symmetrically, $(f_*(x) - f_n(x))Q(\mathcal{S}_x^n) \leq \|g_n - g_*\|_\infty Q(\mathcal{S}_x^n)$. Noting that (10) and (17) imply $Q(\mathcal{S}_x) > 0$ and $Q(\mathcal{S}_x^n) > 0$, we conclude

$$\|f_n - f_*\|_\infty \leq \|g_n - g_*\|_\infty.$$

Arguing analogously for g_{n+1} yields

$$\|g_{n+1} - g_*\|_\infty \leq \|f_n - f_*\|_\infty.$$

Combining these two inequalities yields the claim (19).

Step 2. Recall the expressions of $\mathrm{D}_1\Gamma$ and $\mathrm{D}_2\Gamma$ from (9). By Lemma 4.1, $\mathrm{D}_i\Gamma$ is $\frac{1}{\varepsilon}$ -Lipschitz in each of its coordinates. In particular, the descent lemma yields

$$\Gamma(f, g_n) \geq \Gamma(f_{n-1}, g_n) + \langle \mathrm{D}_1\Gamma(f_{n-1}, g_n), f - f_{n-1} \rangle_{L^2(P)} - \frac{1}{2\varepsilon} \|f - f_{n-1}\|_{L^2(P)}^2$$

for every $f \in L^2(P)$. Taking supremum over f and recalling the definition of f_n , we obtain

$$\begin{aligned} \Gamma(f_n, g_n) &\geq \Gamma(f_{n-1}, g_n) + \sup_f \left\{ \langle \mathrm{D}_1\Gamma(f_{n-1}, g_n), f - f_{n-1} \rangle_{L^2(P)} - \frac{1}{2\varepsilon} \|f - f_{n-1}\|_{L^2(P)}^2 \right\} \\ &= \Gamma(f_{n-1}, g_n) + \frac{\varepsilon}{2} \|\mathrm{D}_1\Gamma(f_{n-1}, g_n)\|_{L^2(P)}^2. \end{aligned}$$

Analogously,

$$\Gamma(f_{n-1}, g_n) \geq \Gamma(f_{n-1}, g_{n-1}) + \frac{\varepsilon}{2} \|\mathbf{D}_2\Gamma(f_{n-1}, g_{n-1})\|_{L^2(Q)}^2.$$

Combining the two inequalities yields

$$\Gamma(f_n, g_n) \geq \Gamma(f_{n-1}, g_{n-1}) + \frac{\varepsilon}{2} \left(\|\mathbf{D}_1\Gamma(f_{n-1}, g_n)\|_{L^2(P)}^2 + \|\mathbf{D}_2\Gamma(f_{n-1}, g_{n-1})\|_{L^2(Q)}^2 \right),$$

but as $\mathbf{D}_1\Gamma(f_{n-1}, g_{n-1}) = 0$ and $\mathbf{D}_2\Gamma(f_{n-1}, g_n) = 0$ by the first-order conditions (17) and (18) for f_{n-1} and g_n , respectively, this even gives

$$\begin{aligned} \Gamma(f_n, g_n) &\geq \Gamma(f_{n-1}, g_{n-1}) + \frac{\varepsilon}{2} \left(\|\mathbf{D}\Gamma(f_{n-1}, g_n)\|_{L^2(P) \times L^2(Q)}^2 + \|\mathbf{D}\Gamma(f_{n-1}, g_{n-1})\|_{L^2(P) \times L^2(Q)}^2 \right) \\ &\geq \Gamma(f_{n-1}, g_{n-1}) + \frac{\varepsilon}{2} \|\mathbf{D}\Gamma(f_{n-1}, g_{n-1})\|_{L^2(P) \times L^2(Q)}^2. \end{aligned} \quad (54)$$

Next, we apply the PL inequality (12) to the last term. Using that

$$\|f_{n-1} \oplus g_{n-1} - f_* \oplus g_*\|_\infty \leq \|f_{n-1} - f_*\|_\infty + \|g_{n-1} - g_*\|_\infty \leq 2\|g_0 - g_*\|_\infty$$

by (19) and abbreviating $M := \max(2\|g_0 - g_*\|_\infty, \varepsilon)$, we obtain

$$\Gamma(f_n, g_n) \geq \Gamma(f_{n-1}, g_{n-1}) + \frac{\varepsilon}{2} \frac{1}{\gamma_\varepsilon M} (\text{QOT}_\varepsilon(P, Q) - \Gamma(f_{n-1}, g_{n-1})).$$

Subtracting this from $\text{QOT}_\varepsilon(P, Q)$ gives $\Delta_n \leq \Delta_{n-1}(1 - \frac{\varepsilon}{2\gamma_\varepsilon M})$ and hence the claim for the suboptimality gap. The claim for the iterates follows via the error bound, as in the proof of Corollary 2.6. \square

4.3 Proof of Corollary 2.8 (coordinate gradient ascent)

Once again, the argument has two steps.

Proof of Corollary 2.8. Step 1. Recall the notation (53). The bound (21) follows by a similar argument as, e.g., [16, Lemma 5.2]; we give the details for completeness. From (20) we obtain that for all $x \in \Omega$,

$$\begin{aligned} f_{n+1}(x) - f_*(x) &= f_n(x) - f_*(x) + \frac{\eta}{\varepsilon} \left(\varepsilon - \int (f_n(x) + g_n(y) - c(x, y))_+ dQ(y) \right) \\ &= f_n(x) - f_*(x) \\ &\quad + \frac{\eta}{\varepsilon} \left(\int (f_*(x) + g_*(y) - c(x, y))_+ - (f_n(x) + g_n(y) - c(x, y))_+ dQ(y) \right) \\ &\leq f_n(x) - f_*(x) + \frac{\eta}{\varepsilon} \left(\int_{\mathcal{S}_x} f_*(x) + g_*(y) - (f_n(x) + g_n(y)) dQ(y) \right) \\ &= \left(1 - \frac{\eta}{\varepsilon} Q(\mathcal{S}_x) \right) (f_n(x) - f_*(x)) - \frac{\eta}{\varepsilon} \left(\int_{\mathcal{S}_x} g_n(y) - g_*(y) dQ(y) \right) \\ &\leq \left(1 - \frac{\eta}{\varepsilon} Q(\mathcal{S}_x) \right) \|f_n - f_*\|_\infty + \frac{\eta}{\varepsilon} Q(\mathcal{S}_x) \|g_n - g_*\|_\infty, \end{aligned}$$

which is a convex combination of $\|f_n - f_*\|_\infty$ and $\|g_n - g_*\|_\infty$. Analogously,

$$f_*(x) - f_{n+1}(x) \leq \left(1 - \frac{\eta}{\varepsilon} Q(\mathcal{S}_x^n)\right) \|f_n - f_*\|_\infty + \frac{\eta}{\varepsilon} Q(\mathcal{S}_x^n) \|g_n - g_*\|_\infty.$$

Together, we have

$$\|f_{n+1} - f_*\|_\infty \leq \max(\|f_n - f_*\|_\infty, \|g_n - g_*\|_\infty).$$

Repeating the argument for g_{n+1} gives

$$\|g_{n+1} - g_*\|_\infty \leq \max(\|f_{n+1} - f_*\|_\infty, \|g_n - g_*\|_\infty) \leq \max(\|f_n - f_*\|_\infty, \|g_n - g_*\|_\infty)$$

and combining the two bounds yields (21).

Step 2. Lemma 4.1 implies

$$\Gamma(f_{n+1}, g_{n+1}) \geq \Gamma(f_{n+1}, g_n) + \langle \mathbf{D}_2\Gamma(f_{n+1}, g_n), g_{n+1} - g_n \rangle_{L^2(Q)} - \frac{1}{2\varepsilon} \|g_{n+1} - g_n\|_{L^2(Q)}^2.$$

The definition of g_{n+1} shows that $g_{n+1} - g_n = \eta \mathbf{D}_2\Gamma(f_{n+1}, g_n)$. Substituting this gives

$$\Gamma(f_{n+1}, g_{n+1}) \geq \Gamma(f_{n+1}, g_n) + \eta \left(1 - \frac{\eta}{2\varepsilon}\right) \|\mathbf{D}_2\Gamma(f_{n+1}, g_n)\|_{L^2(Q)}^2. \quad (55)$$

Analogously, we have

$$\Gamma(f_{n+1}, g_n) \geq \Gamma(f_n, g_n) + \eta \left(1 - \frac{\eta}{2\varepsilon}\right) \|\mathbf{D}_1\Gamma(f_n, g_n)\|_{L^2(P)}^2. \quad (56)$$

Plugging (56) into (55) yields

$$\Gamma(f_{n+1}, g_{n+1}) \geq \Gamma(f_n, g_n) + \eta \left(1 - \frac{\eta}{2\varepsilon}\right) \left(\|\mathbf{D}_1\Gamma(f_n, g_n)\|_{L^2(P)}^2 + \|\mathbf{D}_2\Gamma(f_{n+1}, g_n)\|_{L^2(Q)}^2 \right). \quad (57)$$

Note that the inequality $\|u\|^2 \geq \frac{1}{2}\|v\|^2 - \|u - v\|^2$ holds in any normed space. Applying this with $u := \mathbf{D}_2\Gamma(f_{n+1}, g_n) \in L^2(Q)$ and $v := \mathbf{D}_2\Gamma(f_n, g_n) \in L^2(Q)$ yields

$$\|\mathbf{D}_2\Gamma(f_{n+1}, g_n)\|_{L^2(Q)}^2 \geq \frac{1}{2} \|\mathbf{D}_2\Gamma(f_n, g_n)\|_{L^2(Q)}^2 - \|\mathbf{D}_2\Gamma(f_{n+1}, g_n) - \mathbf{D}_2\Gamma(f_n, g_n)\|_{L^2(Q)}^2.$$

By Lemma 4.1 and the definition of f_{n+1} , the last term satisfies

$$\|\mathbf{D}_2\Gamma(f_{n+1}, g_n) - \mathbf{D}_2\Gamma(f_n, g_n)\|_{L^2(Q)} \leq \frac{1}{\varepsilon} \|f_{n+1} - f_n\|_{L^2(P)} = \frac{\eta}{\varepsilon} \|\mathbf{D}_1\Gamma(f_n, g_n)\|_{L^2(P)},$$

so that

$$\|\mathbf{D}_2\Gamma(f_{n+1}, g_n)\|_{L^2(Q)}^2 \geq \frac{1}{2} \|\mathbf{D}_2\Gamma(f_n, g_n)\|_{L^2(Q)}^2 - \frac{\eta^2}{\varepsilon^2} \|\mathbf{D}_1\Gamma(f_n, g_n)\|_{L^2(P)}^2.$$

Plugging this into (57) and recalling that $\eta \in (0, \varepsilon/\sqrt{2})$, we deduce

$$\Gamma(f_{n+1}, g_{n+1}) \geq \Gamma(f_n, g_n) + \frac{\eta}{2} \left(1 - \frac{\eta}{2\varepsilon}\right) \|\mathbf{D}\Gamma(f_n, g_n)\|_{L^2(P) \times L^2(Q)}^2.$$

This bound has the same form as (54) except for the different constant. Moreover, (21) implies that $\|f_n - f_*\|_\infty + \|g_n - g_*\|_\infty \leq 2(\|f_0 - f_*\|_\infty + \|g_0 - g_*\|_\infty)$. We can now proceed exactly as after (54) to deduce the final result, except that we also use Remark 4.2 to replace $\|f_0 - f_*\|_\infty + \|g_0 - g_*\|_\infty$ by $\|f_0 \oplus g_0 - f_* \oplus g_*\|_\infty$. Once again, the claim for the iterates follows via the error bound, as in the proof of Corollary 2.6. \square

5 Conclusion

In this paper we established a local Polyak–Łojasiewicz inequality for the dual objective of quadratically regularized optimal transport. The proof proceeds by a uniform coercivity estimate for the second-order form obtained from the dual objective along the segment joining an arbitrary pair (f, g) to an optimizer (f_*, g_*) . This yields the error bound and PL inequality of Theorem 2.5; the dependence of the constant on the L^∞ distance to the optimizer reflects the local nature of the result which is unavoidable given the shape of the dual objective.

As an application, we derived linear convergence rates for three natural dual algorithms: gradient ascent (14), coordinate ascent (16), and coordinate gradient ascent (20). Coordinate ascent is implicit for QOT, since each coordinate update requires solving (17)–(18). By contrast, gradient ascent and coordinate gradient ascent are explicit schemes, but they require a choice of step size. The sufficient step-size assumptions in Corollaries 2.6 and 2.8 ensure monotonicity, uniform L^∞ control of the iterates, and hence a uniform PL constant along the trajectory.

Qualitatively, the theoretical results predict the behavior observed numerically. In practice (see, e.g., the plots in [16] for gradient ascent), we observe a finite burn-in phase followed by a clear linear convergence regime. The constants in Section 2.3 are necessarily conservative: in practice the burn-in is typically much shorter, and the asymptotic contraction is often much faster than what is guaranteed by the explicit bounds. While the constants in Theorem 2.5 have a pessimistic explicit dependence on ε , numerical experiments often show that the number of iterations required to reach a fixed accuracy scales closer to $1/\varepsilon$ over moderate ranges.

The step-size restrictions are likewise conservative. Our theory requires $\eta < \varepsilon$ for gradient ascent and $\eta < \varepsilon/\sqrt{2}$ for coordinate gradient ascent. These assumptions are sufficient for the estimates in (15) and (21), but they are not sharp as stability thresholds. Numerically, convergence does fail when the step size is taken too large, but the breakdown often occurs at values larger than those imposed by the proof (see also [16]). In particular, the optimal step size (obtained a posteriori by grid search) is often relatively large.

Lastly, let us briefly comment on some practical aspects that are not captured by our theory. The comparison between the two explicit schemes is nuanced. Coordinate gradient ascent typically requires fewer outer iterations than gradient ascent, because each iteration uses the updated first coordinate before updating the second. However, each such iteration is more expensive. At the conservative step sizes guaranteed by Corollaries 2.6 and 2.8, the wall-time performance of the two schemes is therefore often comparable. If, instead, the best stable step size is supplied by an oracle, coordinate gradient ascent is frequently faster; its advantage then stems from its ability to use a larger effective step size while retaining a faster contraction per iteration. In practice, an adaptive choice of step size (such as monotone Armijo backtracking with Barzilai–Borwein initialization) substantially accelerates both schemes. Then, we often observe that coordinate gradient ascent converges faster (certainly in iterations, but even when measured in wall-time), especially when ε is small. Gradient ascent is sometimes faster for large ε , thanks to the simpler iteration. In both cases, the final step size is often a multiple of ε , and these large steps seem to be responsible for the improvement. The implicit coordinate ascent scheme does not require a step size, but instead requires solving the first-order equations. Implemented with an efficient solver (such as an active-set Newton method), coordinate ascent is competitive with the explicit

methods, and often the fastest when ε is moderate—while the implicit nature makes each iteration costly, a better contraction constant more than compensates for that. Broadly speaking, the three methods seem comparable, with wall-time ratios rarely exceeding 3 for moderate experiments. However, extensive experiments and sophisticated implementations are left for future work.

A Extension to uniformly continuous costs

This appendix shows how to generalize our results from Lipschitz costs c to costs that are merely uniformly continuous. Set

$$R := 1 \vee \text{diam}(\Omega) \vee \text{diam}(\Omega').$$

Assumption A.1 (Cost with modulus of continuity). There exists a modulus of continuity $\omega : [0, \infty) \rightarrow [0, \infty)$, that is, a nondecreasing continuous function with $\omega(0) = 0$, such that

$$|c(x, y) - c(x', y')| \leq \omega(\|x - x'\| + \|y - y'\|) \quad \text{for all } x, x' \in \Omega, y, y' \in \Omega'. \quad (58)$$

Under Assumption A.1, Proposition 2.4 remains valid, except that item (iii) is replaced by

$$|f_*(x) - f_*(x')| \leq \omega(\|x - x'\|), \quad |g_*(y) - g_*(y')| \leq \omega(\|y - y'\|); \quad (59)$$

cf. [26]. We fix these versions of the potentials throughout the remainder of the appendix.

To state the results, define the radius

$$\varrho_{\varepsilon, \omega} := \sup \left\{ r \in [0, R] : 2\omega(r) + \omega(2r) \leq \frac{\varepsilon}{2} \right\} > 0 \quad (60)$$

and the constants

$$\alpha_{\varepsilon, \omega} := \frac{\lambda_P^2 \inf_{y \in \Omega'} Q(B_{\varrho_{\varepsilon, \omega}}(y))}{\Lambda_P^2 (\lceil \text{diam}(\Omega) / \varrho_{\varepsilon, \omega} \rceil)^{d+2}}, \quad (61)$$

$$\kappa_{\varepsilon, \omega} := \delta_P \min(\varrho_{\varepsilon, \omega}, 1)^d, \quad \beta_{\varepsilon, \omega} := \frac{1}{4} \kappa_{\varepsilon, \omega} \alpha_{\varepsilon, \omega}, \quad (62)$$

$$\gamma_{\varepsilon, \omega} := 4\beta_{\varepsilon, \omega}^{-1} = 16 \delta_P^{-1} \min(\varrho_{\varepsilon, \omega}, 1)^{-d} \frac{\Lambda_P^2 (\lceil \text{diam}(\Omega) / \varrho_{\varepsilon, \omega} \rceil)^{d+2}}{\lambda_P^2 \inf_{y \in \Omega'} Q(B_{\varrho_{\varepsilon, \omega}}(y))}. \quad (63)$$

Remark A.2. If one strengthens (58) to the coordinatewise estimate

$$|c(x, y) - c(x', y')| \leq \omega(\|x - x'\|) + \omega(\|y - y'\|),$$

then one can replace (60) with any radius satisfying $\omega(\varrho_{\varepsilon, \omega}) \leq \varepsilon/8$. In particular, denoting by ω^{-1} the (generalized) inverse of ω , one may take the more explicit radius

$$\varrho_{\varepsilon, \omega} := \omega^{-1}(\varepsilon/8) \wedge R.$$

Lemma A.3 (Coordinate-wise coercivity with a modulus of continuity). *Let r_0 be as in (26). For all $r \leq r_0$ and $(u, v) \in L^2(P) \times L^2(Q)$,*

$$\int_{f_r \oplus g_r \geq c} (u \oplus v)^2 d(P \otimes Q) \geq \alpha_{\varepsilon, \omega} \text{Var}_P(u).$$

Proof. The proof is identical to that of Lemma 3.2 up to (29). It remains to replace the estimate following the definition of m_x .

Choose a measurable selector

$$\Omega \ni x \mapsto y(x) \in \operatorname{argmax}_{y \in \Omega'} (f_*(x) + g_*(y) - c(x, y))_+$$

and define

$$m_x := \max_{y \in \Omega'} (f_*(x) + g_*(y) - c(x, y))_+.$$

As before, (10) implies $m_x \geq \varepsilon$ for all $x \in \Omega$. Let $x, z \in \Omega$ and $y \in \Omega'$. Using (58) and (59), we obtain

$$\begin{aligned} f_*(z) + g_*(y) - c(z, y) &\geq f_*(x) + g_*(y(x)) - c(x, y(x)) \\ &\quad - \omega(\|x - z\|) - \omega(\|y(x) - y\|) - \omega(\|x - z\| + \|y(x) - y\|) \\ &= m_x - \omega(\|x - z\|) - \omega(\|y(x) - y\|) - \omega(\|x - z\| + \|y(x) - y\|). \end{aligned}$$

Hence, if $z \in B_{\varrho_{\varepsilon, \omega}}(x)$ and $y \in B_{\varrho_{\varepsilon, \omega}}(y(x))$, then by (60),

$$f_*(z) + g_*(y) - c(z, y) \geq \varepsilon - 2\omega(\varrho_{\varepsilon, \omega}) - \omega(2\varrho_{\varepsilon, \omega}) \geq \frac{\varepsilon}{2},$$

so that $z \in \mathcal{T}_y^\varepsilon$. Thus the inclusion used in the proof of Lemma 3.2 remains valid with $\varepsilon/(8L)$ replaced by $\varrho_{\varepsilon, \omega}$. Repeating the remainder of that proof with $\varrho = \varrho_{\varepsilon, \omega}$ yields the claim. \square

Remark A.4. The proof of Remark 3.4 is unchanged with $\varepsilon/(8L)$ replaced by $\varrho_{\varepsilon, \omega}$. In particular,

$$\inf_{y \in \Omega'} Q(B_{\varrho_{\varepsilon, \omega}}(y)) \leq Q(\mathcal{S}_x^\varepsilon), \quad \inf_{x \in \Omega} P(B_{\varrho_{\varepsilon, \omega}}(x)) \leq P(\mathcal{T}_y^\varepsilon). \quad (64)$$

Theorem A.5 (Uniform coercivity of \mathbb{M} under a modulus of continuity). *Let r_0 be as in (26). Then for all $r \leq r_0$ and $(u, v) \in L^2(P) \times L^2(Q)$,*

$$\langle u \oplus v, \mathbb{M}(u \oplus v) \rangle_{L^2(P \otimes Q)} \geq \beta_{\varepsilon, \omega} \|u \oplus v\|_{L^2(P \otimes Q)}^2, \quad (65)$$

and consequently, for a.e. $r \in [0, r_0]$,

$$\|(f_* - f) \oplus (g_* - g)\|_{L^2(P \otimes Q)}^2 \leq \frac{\varepsilon}{\beta_{\varepsilon, \omega}} \phi''(r). \quad (66)$$

Proof. The proof of Theorem 3.7 carries over once (41) and (42) are replaced by

$$P(\mathcal{T}_{r, y}) \geq \delta_P \min(\varrho_{\varepsilon, \omega}, 1)^d = \kappa_{\varepsilon, \omega}, \quad Q(\mathcal{S}_{r, x}) \geq \inf_{y \in \Omega'} Q(B_{\varrho_{\varepsilon, \omega}}(y)).$$

These follow from Remark A.4 together with Remark 2.2. Note also that

$$\alpha_{\varepsilon, \omega} \leq \inf_{y \in \Omega'} Q(B_{\varrho_{\varepsilon, \omega}}(y)),$$

so Step 2 of the proof of Theorem 3.7 is unchanged as well. Therefore the same spectral argument yields (65), and (66) then follows from (24) exactly as before. \square

Corollary A.6 (PL inequality for Φ with a modulus of continuity). *Theorem 3.1 remains valid under Assumption A.1, with γ_ε replaced throughout by $\gamma_{\varepsilon,\omega}$.*

Proof. Replace Theorem 3.7 by Theorem A.5 in the proof of Theorem 3.1. Since the proof depends on the cost regularity only through the constant β_ε , the same argument yields the conclusion with $\beta_{\varepsilon,\omega}$ in place of β_ε , hence with $\gamma_{\varepsilon,\omega} = 4\beta_{\varepsilon,\omega}^{-1}$ in place of γ_ε . \square

Corollary A.7 (PL inequality for Γ with a modulus of continuity). *Assume Assumptions 2.1 and A.1. Then for every $(f, g) \in L^\infty(\Omega) \times L^\infty(\Omega')$,*

$$\|f \oplus g - f_* \oplus g_*\|_{L^2(P \otimes Q)} \leq \gamma_{\varepsilon,\omega} \max(\|f \oplus g - f_* \oplus g_*\|_\infty, \varepsilon) \|\mathrm{D}\Gamma(f, g)\|_{L^2(P) \times L^2(Q)},$$

and

$$\|\mathrm{D}\Gamma(f, g)\|_{L^2(P) \times L^2(Q)}^2 \geq \frac{\mathrm{QOT}_\varepsilon(P, Q) - \Gamma(f, g)}{\gamma_{\varepsilon,\omega} \max(\|f \oplus g - f_* \oplus g_*\|_\infty, \varepsilon)}.$$

Proof. Lemma 3.8 is unchanged. Therefore, Corollary A.6 implies the stated bounds exactly as in the proof of Theorem 2.5. \square

Remark A.8. If $\omega(r) = Lr$, the explicit radius from Remark A.2 becomes

$$\varrho_{\varepsilon,\omega} = \frac{\varepsilon}{8L} \wedge R$$

and (63) reduces to (13) (up to the truncation at R). That is, the results in the present appendix recover the results in the main text when specialized to a Lipschitz cost c .

B Extension to connected Lipschitz supports

This appendix shows how the convexity assumption on the support Ω of P can be relaxed to

$$\Omega = \bar{U}, \quad U \subset \mathbb{R}^d \text{ is a bounded connected Lipschitz domain.} \quad (67)$$

Apart from this relaxation, we keep the rest of Assumption 2.1. Since U is Lipschitz, $|\partial U| = 0$, and we freely identify Lebesgue integrals over U and over Ω .

Remark B.1 (Lower bounds for ball measures). There exists a constant $\delta_\Omega \in (0, 1]$, depending only on Ω , such that

$$|B_r(x) \cap \Omega| \geq \delta_\Omega r^d, \quad x \in \Omega, \quad 0 < r \leq \mathrm{diam}(\Omega). \quad (68)$$

Indeed, any bounded Lipschitz domain satisfies a uniform interior cone condition; see [18, Theorem 1.2.2]. Consequently,

$$P(B_r(x)) \geq \tilde{\delta}_P \min(r^d, 1), \quad x \in \Omega, \quad r > 0, \quad \text{where } \tilde{\delta}_P := \min(1, \lambda_P \delta_\Omega). \quad (69)$$

The other geometric input needed in the proof is the following nonlocal Poincaré estimate, whose proof reduces locally to the convex case. It replaces Lemma 3.3.

Lemma B.2. *There exists a constant $C_\Omega \geq 1$, depending only on Ω , such that for every $\varrho > 0$ and every $h \in L^2(\Omega)$,*

$$\int_\Omega \int_\Omega (h(x) - h(z))^2 dz dx \leq C_\Omega \left[\frac{\text{diam}(\Omega)}{\varrho} \right]^{d+2} \mathcal{E}_\varrho(h), \quad (70)$$

where

$$\mathcal{E}_\varrho(h) := \int_\Omega \int_{B_\varrho(x) \cap \Omega} (h(x) - h(z))^2 dz dx.$$

Proof. If $\varrho > \text{diam}(\Omega)$, then $B_\varrho(x) \supset \Omega$ for every $x \in \Omega$. If $\varrho = \text{diam}(\Omega)$, this remains true up to a \mathcal{L}_{2d} -null subset of $\Omega \times \Omega$. Hence $\mathcal{E}_\varrho(h)$ equals the left-hand side of (70), and there is nothing to prove. We assume below that $0 < \varrho < \text{diam}(\Omega)$.

Choose a finite collection of open sets U_1, \dots, U_M covering Ω and set

$$\Omega_i := U_i \cap \Omega.$$

We choose this Lipschitz atlas so that each Ω_i is, up to a null set, the image of a bounded convex reference set V_i under a bi-Lipschitz map

$$\Phi_i : V_i \rightarrow \Omega_i.$$

Such an atlas follows from the standard finite Lipschitz atlas for \bar{U} and compactness of \bar{U} . After a refinement, connectedness of U allows us to choose the atlas so that the overlap graph, with vertices $1, \dots, M$ and edge relation

$$i \sim j \iff |\Omega_i \cap \Omega_j| > 0,$$

is connected. Let all constants denoted by C_Ω depend only on this fixed atlas, including the bi-Lipschitz constants, the number of charts, the multiplicity of the cover, and the overlap measures $|\Omega_i \cap \Omega_j|$ along the chosen connected graph. We allow C_Ω to change from line to line.

Fix i and set $g_i := h \circ \Phi_i$ on V_i . Since V_i is convex, Lemma 3.3 applied to V_i at scale comparable to ϱ gives

$$\int_{V_i} \int_{V_i} (g_i(u) - g_i(v))^2 dv du \leq C_\Omega \left[\frac{\text{diam}(\Omega)}{\varrho} \right]^{d+2} \int_{V_i} \int_{V_i \cap B_{\varrho/C_\Omega}(u)} (g_i(u) - g_i(v))^2 dv du.$$

Using the identity

$$\int_{V_i} \int_{V_i} (g_i(u) - g_i(v))^2 dv du = 2|V_i| \inf_{a \in \mathbb{R}} \int_{V_i} (g_i - a)^2 du,$$

we obtain

$$\inf_{a \in \mathbb{R}} \int_{V_i} (g_i - a)^2 du \leq C_\Omega \left[\frac{\text{diam}(\Omega)}{\varrho} \right]^{d+2} \int_{V_i} \int_{V_i \cap B_{\varrho/C_\Omega}(u)} (g_i(u) - g_i(v))^2 dv du.$$

Changing variables $x = \Phi_i(u)$ and $z = \Phi_i(v)$, and using the bi-Lipschitz bounds, yields

$$\inf_{a \in \mathbb{R}} \int_{\Omega_i} (h - a)^2 dx \leq C_\Omega \left[\frac{\text{diam}(\Omega)}{\varrho} \right]^{d+2} \int_{\Omega_i} \int_{\Omega_i \cap B_\varrho(x)} (h(x) - h(z))^2 dz dx. \quad (71)$$

Define the local means and local oscillations

$$m_i := \frac{1}{|\Omega_i|} \int_{\Omega_i} h(x) dx, \quad E_i := \int_{\Omega_i} (h - m_i)^2 dx.$$

By (71),

$$E_i \leq C_\Omega \left[\frac{\text{diam}(\Omega)}{\varrho} \right]^{d+2} \int_{\Omega_i} \int_{\Omega_i \cap B_\varrho(x)} (h(x) - h(z))^2 dz dx. \quad (72)$$

If $i \sim j$, then Jensen's inequality gives

$$|m_i - m_j|^2 \leq \frac{2}{|\Omega_i \cap \Omega_j|} \int_{\Omega_i \cap \Omega_j} (|h - m_i|^2 + |h - m_j|^2) dx \leq C_\Omega (E_i + E_j).$$

Since the overlap graph is finite and connected,

$$|m_i - m_1|^2 \leq C_\Omega \sum_{j=1}^M E_j, \quad i = 1, \dots, M.$$

Therefore,

$$\int_{\Omega} (h - m_1)^2 dx \leq \sum_{i=1}^M \int_{\Omega_i} (h - m_1)^2 dx \leq C_\Omega \sum_{i=1}^M E_i.$$

Combining this with (72) and using the finite multiplicity of the atlas,

$$\inf_{a \in \mathbb{R}} \int_{\Omega} (h - a)^2 dx \leq C_\Omega \left[\frac{\text{diam}(\Omega)}{\varrho} \right]^{d+2} \mathcal{E}_\varrho(h).$$

Finally,

$$\int_{\Omega} \int_{\Omega} (h(x) - h(z))^2 dz dx = 2|\Omega| \inf_{a \in \mathbb{R}} \int_{\Omega} (h - a)^2 dx,$$

and the factor $2|\Omega|$ is absorbed into C_Ω . \square

We can now repeat the proof of Lemma 3.2 using Lemma B.2 instead of Lemma 3.3. This gives the same assertion as in Lemma 3.2 except that its constant α is replaced by

$$\tilde{\alpha}_\varepsilon := \frac{\lambda_P^2}{\Lambda_P^2} \frac{q_0}{C_\Omega \left[\frac{8L \text{diam}(\Omega)}{\varepsilon} \right]^{d+2}}, \quad q_0 := \inf_{y \in \Omega'} Q\left(B_{\frac{\varepsilon}{8L}}(y)\right) > 0. \quad (73)$$

The proof of Theorem 3.7 then carries over with modified constants. Specifically, the assertion now holds with β_ε replaced by

$$\tilde{\beta}_\varepsilon := \frac{1}{4} \tilde{\kappa}_\varepsilon \tilde{\alpha}_\varepsilon, \quad \tilde{\kappa}_\varepsilon := \tilde{\delta}_P \min\left(\frac{\varepsilon}{8L}, 1\right)^d,$$

with $\tilde{\delta}_P$ from (69) and $\tilde{\alpha}_\varepsilon$ from (73). Finally, we obtain the corresponding PL inequality and error bound.

Theorem B.3 (PL inequality on connected Lipschitz supports). *Assume Assumption 2.1, except that convexity of Ω is replaced by*

$$\Omega = \bar{U} \quad \text{for a bounded connected Lipschitz domain } U \subset \mathbb{R}^d,$$

and Assumption 2.3. Then the assertions of Theorems 2.5 and 3.1 remain valid with γ_ε replaced by

$$\tilde{\gamma}_\varepsilon = 16 \left(\tilde{\delta}_P^{-1} \max\left(\frac{8L}{\varepsilon}, 1\right) \right)^d \frac{\Lambda_P^2 C_\Omega \left\lceil \frac{8L \operatorname{diam}(\Omega)}{\varepsilon} \right\rceil^{d+2}}{q_0}, \quad (74)$$

where $\tilde{\delta}_P$ is defined in (69) and $q_0 = \inf_{y \in \Omega'} Q(B_{\varepsilon/(8L)}(y))$.

Corollary B.4 (Linear convergence). *If γ_ε is replaced by $\tilde{\gamma}_\varepsilon$ of (74), all assertions of Corollaries 2.6 to 2.8 carry over to the setting of Theorem B.3.*

References

- [1] E. Bayraktar, S. Eckstein, and X. Zhang. Stability and sample complexity of divergence regularized optimal transport. *Bernoulli*, 31(1):213–239, 2025.
- [2] M. Blondel, V. Seguy, and A. Rolet. Smooth and sparse optimal transport. volume 84 of *Proceedings of Machine Learning Research*, pages 880–889, 2018.
- [3] J. F. Bonnans and A. Shapiro. *Perturbation analysis of optimization problems*. Springer Series in Operations Research. Springer-Verlag, New York, 2000.
- [4] J. Bourgain, H. Brezis, and P. Mironescu. Another look at Sobolev spaces. In *Optimal Control and Partial Differential Equations*, pages 439–455. IOS, Amsterdam, 2001.
- [5] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Universitext. Springer, New York, 2011.
- [6] G. Carlier. On the linear convergence of the multimarginal Sinkhorn algorithm. *SIAM J. Optim.*, 32(2):786–794, 2022.
- [7] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013.
- [8] S. Eckstein and M. Kupper. Computation of optimal transport and related hedging problems via penalization and neural networks. *Appl. Math. Optim.*, 83(2):639–667, 2021.
- [9] S. Eckstein and M. Nutz. Convergence rates for regularized optimal transport via quantization. *Math. Oper. Res.*, 49(2):1223–1240, 2024.
- [10] M. Essid and J. Solomon. Quadratically regularized optimal transport on graphs. *SIAM J. Sci. Comput.*, 40(4):A1961–A1986, 2018.
- [11] G. B. Folland. *Real analysis*. Pure and Applied Mathematics. John Wiley & Sons, New York, second edition, 1999.
- [12] A. Genevay, M. Cuturi, G. Peyré, and F. Bach. Stochastic optimization for large-scale optimal transport. In *Advances in Neural Information Processing Systems 29*, pages 3440–3448, 2016.
- [13] A. González-Sanz, E. del Barrio, and M. Nutz. Sample complexity of quadratically regularized optimal transport. *arXiv:2511.09807*, 2025.

- [14] A. González-Sanz, S. Eckstein, and M. Nutz. Sparse regularized optimal transport without curse of dimensionality. *arXiv:2505.04721*, 2025.
- [15] A. González-Sanz and M. Nutz. Stability of quadratically regularized optimal transport. *In preparation*, 2026.
- [16] A. González-Sanz, M. Nutz, and A. Riveros Valdevenito. Linear convergence of gradient descent for quadratically regularized optimal transport. *arXiv:2509.08547*, 2025.
- [17] A. González-Sanz and M. Nutz. Sparsity of quadratically regularized optimal transport: Scalar case. *SIAM J. Math. Anal.*, *forthcoming*, 2024. Preprint arXiv:2410.03353.
- [18] P. Grisvard. *Elliptic Problems in Nonsmooth Domains*, volume 24 of *Monographs and Studies in Mathematics*. Pitman, Boston, 1985.
- [19] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville. Improved training of Wasserstein GANs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 5769–5779, 2017.
- [20] L. Li, A. Genevay, M. Yurochkin, and J. Solomon. Continuous regularized Wasserstein barycenters. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 17755–17765. Curran Associates, Inc., 2020.
- [21] D. A. Lorenz and H. Mahler. Orlicz-space regularization for optimal transport and algorithms for quadratic regularization. *arXiv:1909.06082*, 2019.
- [22] D. A. Lorenz and H. Mahler. Orlicz space regularization of continuous optimal transport problems. *Appl. Math. Optim.*, 85(2):Paper No. 14, 33, 2022.
- [23] D. A. Lorenz, P. Manns, and C. Meyer. Quadratically regularized optimal transport. *Appl. Math. Optim.*, 83(3):1919–1949, 2021.
- [24] B. Muzellec, R. Nock, G. Patrini, and F. Nielsen. Tsallis regularized optimal transport and ecological inference. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1), 2017.
- [25] M. Nutz. Introduction to entropic optimal transport, 2021. Lecture notes, Columbia University, https://www.math.columbia.edu/~mnutz/docs/EOT_lecture_notes.pdf.
- [26] M. Nutz. Quadratically regularized optimal transport: existence and multiplicity of potentials. *SIAM J. Math. Anal.*, 57(3):2622–2649, 2025.
- [27] G. Peyré and M. Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5–6):355–607, 2019.
- [28] L. Rüschemdorf and W. Thomsen. Note on the Schrödinger equation and I -projections. *Statist. Probab. Lett.*, 17(5):369–375, 1993.
- [29] B. Schmitzer. Stabilized sparse scaling algorithms for entropy regularized transport problems. *SIAM J. Sci. Comput.*, 41(3):A1443–A1481, 2019.
- [30] V. Seguy, B. B. Damodaran, R. Flamary, N. Courty, A. Rolet, and M. Blondel. Large scale optimal transport and mapping estimation. In *International Conference on Learning Representations*, 2018.
- [31] A. J. Stromme. Minimum intrinsic dimension scaling for entropic optimal transport. *arXiv:2306.03398*, 2023.
- [32] J. Wiesel and X. Xu. Sparsity of quadratically regularized optimal transport: Bounds on concentration and bias. *SIAM J. Math. Anal.*, 57(6):6498–6521, 2025.
- [33] S. Zhang, G. Mordant, T. Matsumoto, and G. Schiebinger. Manifold learning with sparse regularised optimal transport. *arXiv:2307.09816*, 2023.