



**HAL**  
open science

# Stability and upper bounds for statistical estimation of unbalanced transport potentials

Adrien Vacher, François-Xavier Vialard

► **To cite this version:**

Adrien Vacher, François-Xavier Vialard. Stability and upper bounds for statistical estimation of unbalanced transport potentials. 2022. hal-03609629v1

**HAL Id: hal-03609629**

**<https://hal.science/hal-03609629v1>**

Preprint submitted on 17 Mar 2022 (v1), last revised 15 Jun 2022 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Stability and upper bounds for statistical estimation of unbalanced transport potentials.

Adrien Vacher\*<sup>◦</sup>      François-Xavier Vialard\*<sup>◦</sup>

\* Laboratoire Informatique Gaspard Monge Univ Gustave Eiffel, CNRS, LIGM,  
F-77454 Marne-la-Vallée, France

◦ INRIA Paris, 2 rue Simone Iff, 75012, Paris, France

adrien.vacher@u-pem.fr, francois-xavier.vialard@u-pem.fr

## Abstract

In this note, we derive upper-bounds on the statistical estimation rates of unbalanced optimal transport (UOT) maps for the quadratic cost. Our work relies on the stability of the semi-dual formulation of optimal transport (OT) extended to the unbalanced case. Depending on the considered variant of UOT, our stability result interpolates between the OT (balanced) case where the semi-dual is only locally strongly convex with respect to the Sobolev semi-norm  $\dot{H}^1$  and the case where it is locally strongly convex with respect to the  $H^1$  norm. When the optimal potential belongs to a certain class  $\mathcal{C}$  with sufficiently low metric-entropy, local strong convexity enables us to recover super-parametric rates, faster than  $1/\sqrt{n}$ .

## 1 Introduction

In its original formulation, OT is a tool to compare probability distributions: it seeks a map that optimally transports one distribution  $\mu$  to another distribution  $\nu$  with respect to some fixed cost  $c$  and it returns the associated transport cost. This problem was later relaxed into a linear program by Kantorovitch and its primal formulation consists into seeking a coupling instead of a map with minimal cost and whose marginals are constrained to be  $\mu$  and  $\nu$ ; under suitable assumptions on the measures and the cost, this relaxation is tight (Brenier, 1991). Quite recently, OT was extended to arbitrary positive measures (Chizat, 2017), with possibly different masses, thus the name Unbalanced Optimal Transport (UOT). On the primal problem, the hard marginal constraints are relaxed by soft entropic penalties. From an applied point of view, the mass constraint relaxation is indeed a desirable feature: it allows not only displacement of mass but also local growth or shrinkage (Séjourné et al., 2019). In image processing for instance, it can remove or at least decrease blurred areas in favor of sharper contrasts Feydy et al. (2017). From a statistical point of view, UOT

may appear as a more robust version of OT as it is able to cut down the outliers. At the heart of classical OT, rather than the mere OT distance value, the main object of interest is the transport map: in generative imaging, we use the transport map to generate images from noise (Liu et al., 2019), for point cloud approximation, the particle flow is driven by the maps (Mérigot et al., 2021) and in Domain Adaptation, the source distribution is transported on the target using an OT map (Courty et al., 2017). Novel applications include predicting the evolutions of cells from measurements Schiebinger et al. (2019); Yang et al. (2020). Notably in the case of a quadratic cost, Brenier showed that these maps are linked with the dual formulation of Kantorovitch relaxation: the map from  $\mu$  to  $\nu$  corresponds to the the gradients of the first variable of the dual problem that we shall refer to as a *potential*. This potential is the solution of a linear program, yet with an infinite dimensional constraint. Hence, even in the case where the measures  $\mu$  and  $\nu$  are known analytically, there is in general no closed form to compute the OT potentials. In particular, recent methods instead rely on  $n$ -samples empirical counterparts of  $\mu$  and  $\nu$  to statistically estimate the cost and the potentials (Genevay et al., 2016; Seguy et al., 2018). Such methods are thus concerned with the statistical estimation of optimal transport quantities, such as the cost or the potentials, see Panaretos and Zemel (2019, 2020) for an overview of this rapidly growing field. If the distributions  $\mu, \nu$  are only assumed to have a density w.r.t. the Lebesgue measure, the error achieved by the *plug-in* estimator, which is simply the OT between the empirical samples, scales in  $O(n^{-\frac{2}{d}})$  (Chizat et al., 2020). In particular, without further assumptions the OT problem is said to suffer the curse of dimension. However, in the seminal work of Hütter and Rigollet (2021), the authors showed that if the original potential is  $\alpha$ -smooth, then its gradient could be at best estimated in  $n^{-\frac{2(\alpha+1)}{2\alpha+d}}$  with respect to the squared  $L^2$  distance. They also provided an estimator that actually achieved this rate of estimation, hence providing a *minimax* rate under smoothness assumptions. Yet, we emphasize the fact that their estimator as such is infeasible as it requires in particular to project on the space of convex,  $k$ -times differentiable functions. This result has triggered follow-ups on computationally feasible and efficient estimators of the optimal transport maps, for instance Muzellec et al. (2021) leveraging the underlying smoothness of the optimal maps and Pooladian and Niles-Weed (2021); Pooladian et al. (2022); Deb et al. (2021) using entropic regularization.

In this note, we propose to explore the results of statistical estimation of transport potentials to the unbalanced setting and derive upper-bounding rates. In particular, we restrict ourselves to probability measures instead of positive measures, yet it does not affect the relevance of using UOT instead of classical OT as it allows to handle outliers. Instead of making explicit smoothness assumptions on the ground truth, we assume that it belongs to a certain class  $C$  and derive rates of estimation depending on the complexity of  $C$ , namely its metric entropy. In particular, our statistical analysis relies on an unbiased estimation, where our search space for the empirical candidate is the same set  $C$ , that is assumed to contain the ground truth. Using the recent regularity results

on Unbalanced Optimal transport (Gallouët et al., 2021), we shall in particular cover the case where the smoothness assumption is not directly made on the potential  $z_0$  but instead on the measures  $\mu, \nu$ . As in the balanced case, we rely on the semi-dual formulation of UOT for which we derive stability results; interestingly, thanks to the extra convexity brought by the entropic relaxation of the marginal in some cases, we do not need to assume smoothness of the potentials to derive those stability estimates. In the case where the metric entropy of  $C$  slowly diverges, the strong convexity enables us to use localization arguments and derive super-parametric rates. In particular, we obtain two different regimes that depend on the metric entropy of  $C$ ; under smoothness assumptions, our rates closely match those of Hütter and Rigollet (2021) in the highly smooth case.

**Assumptions and notations** In this paper  $X, Y$  are compact subsets of  $\mathbb{R}^d$ ,  $\mu$  and  $\nu$  are positive measures over  $X$  and  $Y$  respectively with their  $n$ -independent samples empirical counterparts  $\hat{\mu}, \hat{\nu}$  when  $\mu$  and  $\nu$  are probability measures. We shall denote by  $\text{supp}(\mu), \text{supp}(\nu)$  the support of  $\mu$  and  $\nu$  respectively. We shall denote by  $\langle \cdot, \cdot \rangle$  the pairing between radon measures and continuous functions and by  $q$  the quadratic function  $q(x) = \frac{1}{2}\|x\|^2$ . The notation  $\|\cdot\|_{L^p_Z}$  for  $p \in [1, +\infty]$  shall refer to the  $L^p$  norm over functions defined on Borel sets  $Z$  as  $\|f\|_{L^p_Z} = (\int_Z |f(x)|^p dx)^{\frac{1}{p}}$ . Conversely, for a probability measure  $\beta$ , we shall denote for  $p \in [1, +\infty]$ ,  $\|g\|_{L^p(\beta)} = (\int_x |g(x)|^p d\beta(x))^{\frac{1}{p}}$ .

## 2 Unbalanced Optimal Transport

In this section, we present unbalanced optimal transport via primal and dual formulations. The latter is used to show stability estimates, generalizing standard strong convexity estimates for the semi-dual in optimal transport. Unbalanced optimal transport (UOT) is a generalization of standard optimal transport which relaxes the marginal constraints using a convex divergence between positive measures. The primal formulation of UOT uses Csiszár divergences which are defined as follows.

**Definition 1** (Csiszár divergences). *Let  $F : \mathbb{R}_+ \mapsto \mathbb{R}_+ \cup \{+\infty\}$  be a convex lower semicontinuous function such that  $F(1) = 0$ . Its recession constant is  $F'_\infty = \lim_{r \rightarrow \infty} \frac{F(r)}{r}$ . Let  $\mu, \nu$  be non negative Radon measures on a convex domain  $\Omega$  in  $\mathbb{R}^d$ . The Csiszár divergence associated with  $F$  is*

$$D_F(\mu, \nu) = \int_{\Omega} F\left(\frac{d\mu(x)}{d\nu(x)}\right) d\nu(x) + F'_\infty \int_{\Omega} d\mu^\perp, \quad (1)$$

where  $\mu^\perp$  is the orthogonal part of the Lebesgue decomposition of  $\mu$  with respect to  $\nu$ .

The Kullback-Leibler divergence is obtained for  $F(x) = x \log(x) - x + 1$ . The primal formulation of UOT is defined by, for  $\rho_0, \rho_1 \in \mathcal{M}_+(\Omega)$ ,

$$\text{UOT}(\rho_0, \rho_1) = \inf_{\gamma \in \mathcal{M}_+(X \times Y)} D_{F_0}(\gamma_0, \rho_0) + D_{F_1}(\gamma_1, \rho_1) + \int_{X \times Y} c(x, y) d\gamma(x, y),$$

where  $F_0, F_1$  are two possibly different entropy functions. Note that standard OT is recovered for the entropy function  $F(x) = \iota_{\{1\}}(x)$  the convex indicator function of  $\{1\}$ .

The optimization problem associated with UOT is convex and its dual formulation reads, denoting  $F^*$  the Legendre transform of  $F$ ,

$$\sup_{z_0, z_1 \in C_b(X), C_b(Y)} - \int_X F_0^*(-z_0(x)) d\rho_0(x) - \int_Y F_1^*(-z_1(y)) d\rho_1(y) \quad (2)$$

under the constraint

$$z_0(x) + z_1(y) \leq c(x, y). \quad (3)$$

The following proposition shows that at optimality,  $z_0$  is a standard optimal transport potential between modified versions of  $(\mu, \nu)$  for the cost  $c(x, y)$ .

**Proposition 1** (see Lemma 3 in Gallouët et al. (2021)). *Assume that  $\varphi^*$  is differentiable on its domain. At optimality of (2), the pair  $(z_0, z_1)$  reads  $z_1 = z_0^c := \inf_x c(x, \cdot) - z_0(x)$  and  $z_0$  is an optimal transport potential between  $\tilde{\mu} = \partial F_0^*(-z_0)\mu$  and  $\tilde{\nu} = \partial F_1^*(-z_1)\nu$ .*

In the rest of the paper, we shall from now on focus assume that the cost is quadratic  $c(x, y) = q(x - y)$ . Furthermore, to avoid heavy notations, we shall assume  $F_0 = F_1 := \varphi$ . However, similar results can be obtained when the entropy functions are different.

It is possible to optimize Formula (2) with respect to the second variable to obtain the so-called *semi-dual* formulation of UOT. Indeed, the optimal  $z_1$  is the  $c$ -conjugate of  $z_0$ . Using this argument on  $z_0^c$ , one can further assume that  $z_0$  is the  $c$ -conjugate of a function  $z_1$ , which says that  $\tilde{z}_0 = q - z_0$  is convex. In this case, the semi-dual UOT problem can be read as

$$- \inf_{z \in C_b(X), z \text{ convex}} \langle \varphi^*(z - q), \mu \rangle + \langle \varphi^*(z^* - q), \nu \rangle, \quad (4)$$

where we injected  $(q - z)^c = q - z^*$ . Thus, let us introduce

**Definition 2** (Semi-Dual UOT). *Given nonnegative measures  $\mu, \nu$ , the UOT semi-dual is defined by*

$$J_{\mu, \nu}(z) = \langle \varphi^*(z - q), \mu \rangle + \langle \varphi^*(z^* - q), \nu \rangle. \quad (5)$$

When confusion is possible, we shall denote  $J_{\mu, \nu}$  by  $J$ . This semi-dual objective  $J$  remains convex and even gains in convexity with respect to the original objective. This phenomenon is well-known in standard OT and we show how it extends in the unbalanced setting. The important difference with standard OT is that when  $\varphi^*$  is strongly convex, the stability is expressed in an  $H^1$  norm instead of the  $L^2$  norm of the gradient.

**Proposition 2** (Stability estimate). *The semi-dual functional  $J$  is convex. Assume that  $\varphi^*$  is differentiable and that  $\nu$  is absolutely continuous with respect to the Lebesgue measure. Let  $z$  a  $\lambda$ -strongly convex potential and the optimal potential  $z_0$ . Then, it holds*

$$J(z) - J(z_0) \geq \frac{1}{2\lambda} \mathbb{E}_{\tilde{\nu}}[\|\nabla(z_0^* - z^*)\|^2] + C_{z^*} \mathbb{E}_{\nu}[(z_0^* - z^*)^2] + C_z \mathbb{E}_{\mu}[(z_0 - z)^2], \quad (6)$$

with nonnegative constants  $C_z$  and  $C_{z^*}$  depends on  $\varphi^*$  and  $z$ . If  $z$  and  $z^*$  and uniformly bounded on the support of  $\mu$  and  $\nu$  respectively and if  $\varphi^*$  is strongly convex on every compact, then  $C_z$  and  $C_{z^*}$  are uniformly lower bounded by a constant  $C' > 0$ .

*Proof.* We start by applying the convexity inequality

$$\varphi^*(y) - \varphi^*(x) \geq \varphi^{*'}(x)(y - x) + \frac{1}{2} m_{xy} |x - y|^2, \quad (7)$$

where  $m_{xy} \geq 0$ . We apply it in the difference of dual values and we get

$$\begin{aligned} J(z) - J(z_0) &\geq \langle z - z_0, (\varphi^*)'(z_0 - q)\mu \rangle + \langle z^* - z_0^*, (\varphi^*)'(z_0^* - q)\nu \rangle \\ &\quad + C_z \langle (z_0 - z)^2, \mu \rangle + C_{z^*} \langle (z_0^* - z^*)^2, \nu \rangle, \end{aligned}$$

where

$$\begin{cases} C_z = \inf_{x \in \text{supp}(\mu)} m_{z(x)-q(x), z_0(x)-q(x)} \\ C_{z^*} = \inf_{y \in \text{supp}(\nu)} m_{z^*(y)-q(y), z_0^*(y)-q(y)}. \end{cases} \quad (8)$$

Now, recall that  $z_0$  is an optimal potential for the transport of measure  $\tilde{\mu} := \varphi^{*'}(z_0 - q)\mu$  onto the measure  $\tilde{\nu} := \varphi^{*'}(z_0^* - q)\nu$ . Denoting  $\tilde{J}(z) = \langle z, \tilde{\mu} \rangle + \langle z^*, \tilde{\nu} \rangle$ , Equation (9) reads

$$J(z) - J(z_0) \geq \tilde{J}(z) - \tilde{J}(z_0) + C_z \langle (z_0 - z)^2, \mu \rangle + C_{z^*} \langle (z_0^* - z^*)^2, \nu \rangle. \quad (9)$$

We now apply the stability of optimal transport guaranteed by the absolute continuity of  $\nu$  which gives the lower bound  $\tilde{J}(z) - \tilde{J}(z_0) \geq \frac{1}{2\lambda} \mathbb{E}_{\tilde{\nu}}[\|\nabla(z^* - z_0^*)\|^2]$ .  $\square$

Note that this upper-bound encompasses the balanced case and that the entropy can bring extra convexity. In the generic case, we shall write  $J(z) - J(z_0) \geq d_{H^\circ}^\lambda(z, z_0)$  where the pseudo distance  $d_{H^\circ}^\lambda$  is defined as

$$d_{H^\circ}^\lambda(f, g)^2 := \frac{1}{2\lambda} \mathbb{E}_{\tilde{\nu}}[\|\nabla(f^* - g^*)\|^2] + C' (\mathbb{E}_{\nu}[(f^* - g^*)^2] + \mathbb{E}_{\mu}[(f - g)^2]),$$

where  $C' \geq 0$ . When  $C'$  is strictly positive,  $J$  is (locally, at the optimum) strongly convex with respect to an  $H^1$  norm while in the balanced setting, there it is formulated with the semi-norm  $\dot{H}^1$ . Furthermore, when  $C' > 0$ ,  $J(z) - J(z_0)$  not only controls  $z_0^* - z^*$  in an  $L^2$  sense but also  $z_0 - z$  in an  $L^2$  sense. This is a notable improvement with respect to balanced OT: in this case,

in order to upper-bound  $z - z_0$  with  $J(z) - J(z_0)$ , a smoothness assumption must be made on  $z_0$ ; in the UOT case, if the entropy is sufficiently convex, we can obtain an upper-bound *without* this extra smoothness assumption. In the next section, we use this property to apply a localization technique and obtain fast rates without requiring smoothness of the functions in  $C$ .

### 3 Estimation of UOT maps

In this section, we restrict ourselves to the case where  $\mu, \nu$  are probability measures which we only access through their  $n$ -samples stochastic counterparts  $\hat{\mu}, \hat{\nu}$ . Even though this setting is much more restrictive than the original UOT setting where  $\mu$  and  $\nu$  can be arbitrary positive radon measures, it remains nonetheless a relevant setting in ML applications for instance. Indeed, the relaxation of the hard marginal OT constraint by a divergence allow to better handle outliers as shown experimentally by Mukherjee et al. (2021).

In this stochastic setting, a natural way to estimate UOT map is to solve the empirical semi-dual over a given search space  $C$ .

**Definition 3** (Stochastic Semi-Dual Unbalanced OT). *Let  $C$  be a set of real-valued function, we define  $\widehat{UOT}_C$*

$$\widehat{UOT}_C = - \inf_{z \in C} \hat{J}(z), \quad (10)$$

where  $\hat{J} = J_{\hat{\mu}, \hat{\nu}}$ . Conversely, we define the empirical potential

$$\hat{z}_C = \arg \min_{z \in C} \hat{J}(z). \quad (11)$$

When no confusion is possible, we shall simply denote it  $\hat{z}$ .

If the true unbalanced potential  $z_0$  belongs to  $C$ , we can prove that the empirical potential  $\hat{z}$  converges toward  $z_0$  with respect to  $d_{H^\circ}^\lambda$  at a rate that will depend on the complexity of  $C$ .

#### 3.1 Generic case

We show that under suitable assumptions, the solutions of empirical unbalanced semi-dual OT converges toward the ground truth in the  $d_{H^\circ}^\lambda$  sense.

**Assumption 1.** *The measures  $\mu, \nu$  have support included in  $B_R$ , where  $B_r$  is the euclidean ball of  $\mathbb{R}^d$  centered in 0 and of radius  $r$ .*

**Assumption 2.** *The measures  $\mu, \nu$  have densities with respect to the Lebesgue measure on  $B_R$ .*

**Assumption 3.** *There exists  $\tilde{z}_0 \in C$  such that  $\tilde{z}_0$  coincides with  $z_0$  on  $\text{supp}(\mu)$  and with  $\tilde{z}_0^*$  coincides with  $z_0^*$  on  $\text{supp}(\nu)$ .*

**Assumption 4.** *The functions in  $C$  are uniformly bounded by  $M(r)$  over  $B_r$ , uniformly lower bounded by  $l$  and are  $\lambda$ -strongly convex.*

The goal of Assumption 2 is to ensure the existence of the unbalanced transport map between  $\mu, \nu$ . The goal of Assumption 3 is to ensure the absence of bias in the model. We believe that under a finer analysis, Assumption 1 could be replaced with sub-gaussian measures. We show in the following Lemma that Assumption 4 ensures that the conjugate of the functions in  $C$  are both bounded and smooth on every ball.

**Lemma 1.** *For all  $z$  that are  $\lambda$ -strongly convex and such that  $z \geq l$ ,  $\|z\|_{L_{B_r}^\infty} \leq M(r)$ , we have  $\|\nabla z^*\|_{L_{B_r}^\infty} \leq G(r) := \frac{r}{\lambda} + \sqrt{\frac{2(M(0)-l)}{\lambda}}$  and  $\|z^*\|_{L_{B_r}^\infty} \leq M'(r) := rG(r) + M(G(r))$ .*

*Proof.* For  $z \in C$ , we have that  $z^*$  is  $\frac{1}{\lambda}$ -smooth. In particular, for  $x \in B_r$

$$\|\nabla z^*(x)\| = \|\nabla z^*(x) - \nabla z^*(0) + \nabla z^*(0)\| \quad (12)$$

$$\leq \|\nabla z^*(x) - \nabla z^*(0)\| + \|\nabla z^*(0)\| \quad (13)$$

$$\leq \frac{r}{\lambda} + \|\nabla z^*(0)\|. \quad (14)$$

Now recall that  $\nabla z^*(0) = \arg \min_{x \in \mathbb{R}^d} z(x)$ . Since  $z$  is  $\lambda$ -strongly convex, we have the following inequality

$$z(0) \geq z(x_*) + \frac{\lambda}{2} \|x_*\|^2, \quad (15)$$

where  $x_* = \arg \min_{x \in \mathbb{R}^d} z(x)$ . Using that  $z(0) \leq M(0)$  and  $-z \leq -l$ , we recover

$$\|x_*\| \leq \sqrt{\frac{2(M(0) - l)}{\lambda}}. \quad (16)$$

The bound on  $\|z^*\|_{L_{B_r}^\infty}$  follows the definition of the Fenchel-Legendre transform

$$z^*(x) = x^\top \nabla z^*(x) - z(\nabla z^*(x)). \quad (17)$$

□

Finally, combined with the previous Lemma, Assumption 4 also ensures that the conjugate on  $C$  has a Lipschitz behavior with respect to the sup norm.

**Lemma 2.** *Let  $z_1, z_2$  be  $\lambda$ -strongly convex functions such that  $z_1, z_2$  are lower-bounded by  $l$  and bounded by  $M(r)$  on  $B_r$ . We have  $\|z_1^* - z_2^*\|_{L_{B_R}^\infty} \leq \|z_1 - z_2\|_{L_{B_{G(R)}}^\infty}$ , where  $G(r) := \frac{r}{\lambda} + \sqrt{\frac{2(M(0)-l)}{\lambda}}$  as in Lemma 1.*

*Proof.* Let  $x \in B_R$ . By definition of the Fenchel transform, we have for all  $y \in \mathbb{R}^d$

$$z_1^*(x) \geq x^\top y - z_1(y), \quad (18)$$



with equality when  $y = \nabla z_1^*(x)$ . Hence, we have for all  $y$

$$z_1^*(x) - z_2^*(x) \geq x^\top y - z_1(y) + z_2(\nabla z_2^*(x)) - x^\top \nabla z_2^*(x). \quad (19)$$

In particular, for  $y = \nabla z_2^*(x)$ , we obtain

$$z_1^*(x) - z_2^*(x) \geq z_2(\nabla z_2^*(x)) - z_1(\nabla z_2^*(x)), \quad (20)$$

and applying Lemma 1 yields  $z_1^*(x) - z_2^*(x) \geq -\|z_1 - z_2\|_{L^\infty_{B_G(R)}}$ . Conversely, flipping the role of  $z_1, z_2$ , we obtain

$$z_2^*(x) - z_1^*(x) \geq z_1(\nabla z_1^*(x)) - z_2(\nabla z_1^*(x)), \quad (21)$$

which yields  $|z_1^*(x) - z_2^*(x)| \leq \|z_1 - z_2\|_{L^\infty_{B_G(R)}}$ .  $\square$

We have now all the ingredients to derive our result.

**Proposition 3.** *Denoting  $\hat{z}_C$  the solution of problem (10), we have under Assumptions 1-4, if the unbalanced optimal transport potential  $z_0$  between  $\mu$  and  $\nu$  belongs to  $C$ , then we have for all  $\delta \leq \frac{M'}{L}$*

$$\mathbb{E}[d_{H^\circ}^\lambda(\hat{z}_C, z_0)^2] \lesssim \delta + \frac{1}{\sqrt{n}} \int_{\frac{\delta}{4}}^{\frac{M'}{L}} \sqrt{n(C, L_{B_{R'}}^\infty, Lu)} du$$

where  $n(C, \|\cdot\|, u)$  is the logarithm of the covering number, also called the metric entropy, of  $C$  with respect to the  $\|\cdot\|$  (semi)-norm at scale  $u$ ,  $M' = (M, R, \lambda, l, \varphi)$ ,  $R' = (M, R, \lambda, l)$ ,  $L = (M, R, \lambda, l, \varphi)$  and  $\lesssim$  hides a factor 64.

*Proof.* We start by applying the strong convexity inequality of the semi-dual and the optimality conditions

$$d_{H^\circ}^\lambda(\hat{z}, z_0)^2 \leq J(\hat{z}) - J(z_0) \quad (22)$$

$$= J(\hat{z}) - \hat{J}(\hat{z}) + \hat{J}(\hat{z}) - \hat{J}(z_0) + \hat{J}(z_0) - J(z_0). \quad (23)$$

Using Assumption 3, the term  $\hat{J}(\hat{z}) - \hat{J}(z_0)$  is negative hence we have

$$d_{H^\circ}^\lambda(\hat{z}, z_0)^2 \leq J(\hat{z}) - \hat{J}(\hat{z}) + \hat{J}(z_0) - J(z_0) \quad (24)$$

$$\leq \sup_{z \in C} \langle \phi^*(z - q), \mu - \hat{\mu} \rangle \quad (25)$$

$$+ \sup_{z \in C^*} \langle \phi^*(z - q), \nu - \hat{\nu} \rangle \quad (26)$$

$$+ \hat{J}(z_0) - J(z_0), \quad (27)$$

where we denoted  $C^* = \{z^*, z \in C\}$ .

**Bound on term (25)** Denoting  $C_0 = \{\phi^*(g - q), g \in C\}$ , we apply Luxburg and Bousquet (2004, Theorem 16) to bound our empirical process

$$W := \sup_{z \in C} \langle \phi^*(z - q), \mu - \hat{\mu} \rangle,$$

and we obtain for all  $\delta > 0$

$$\mathbb{E}[W] \leq 2\delta + \frac{4\sqrt{2}}{\sqrt{n}} \int_{\frac{\delta}{4}}^{\infty} \sqrt{n(C_0, L^2(\hat{\mu}), u)} \, du. \quad (28)$$

Noting that  $\|g\|_{L^2(\hat{\mu})} \leq \|g\|_{L^\infty(\mu)}$  almost surely, we recover the upper bound

$$\mathbb{E}[W] \leq 2\delta + \frac{4\sqrt{2}}{\sqrt{n}} \int_{\frac{\delta}{4}}^{\infty} \sqrt{n(C_0, L^\infty(\mu), u)} \, du. \quad (29)$$

Since the functions in  $C$  are uniformly bounded by  $M(R)$  on  $B_R$  and that  $\mu$  is supported on  $B_R$ , we have  $\forall (g_1, g_2) \in C^2$ ,

$$\|\phi^*(g_1 - q) - \phi^*(g_2 - q)\|_{L^\infty(\mu)} \leq L_{\phi^*}^1 \|g_1 - g_2\|_{L^\infty(\mu)}, \quad (30)$$

where  $L_{\phi^*}^1$  is defined as

$$L_{\phi^*}^1 := \sup_{x \in [-M_1, M_1]} |\partial \phi^*(x)|, \quad (31)$$

and  $M_1 = 2M(R) + R^2$ . In particular, we get the new upper-bound for all  $\frac{\delta}{4} \leq \frac{2M(R)}{L_{\phi^*}^1}$

$$\begin{aligned} \mathbb{E}[W] &\leq 2\delta + \frac{4\sqrt{2}}{\sqrt{n}} \int_{\frac{\delta}{4}}^{\frac{2M(R)}{L_{\phi^*}^1}} \sqrt{n(C, L^\infty(\mu), L_{\phi^*}^1 u)} \, du \\ &\leq 2\delta + \frac{4\sqrt{2}}{\sqrt{n}} \int_{\frac{\delta}{4}}^{\frac{2M(R)}{L_{\phi^*}^1}} \sqrt{n(C, L_{B_R}^\infty, L_{\phi^*}^1 u)} \, du. \end{aligned}$$

**Bound on term (26)** Lemma 1 ensures that the functions in  $C^*$  are uniformly bounded on every ball  $B_r$  by some constant  $M'(r)$ . In particular, we can proceed as in the last paragraph and obtain

$$\mathbb{E}[W^*] \leq 2\delta + \frac{4\sqrt{2}}{\sqrt{n}} \int_{\frac{\delta}{4}}^{\frac{2M'(R)}{L_{\phi^*}^2}} \sqrt{n(C^*, L_{B_R}^\infty, L_{\phi^*}^2 u)} \, du,$$

where  $W^* := \sup_{z \in C^*} \langle z, \nu - \hat{\nu} \rangle$  and  $L_{\phi^*}^2$  is defined as

$$L_{\phi^*}^2 := \sup_{x \in [-M_2, M_2]} |\partial \phi^*(x)|, \quad (32)$$

with  $M_2 = 2M'(R) + R^2$ . Using Lemma 2 that states

$$\|z_1^* - z_2^*\|_{L_{B_R}^\infty} \leq \|z_1 - z_2\|_{L_{B_{G(R)}}^\infty}, \quad (33)$$

for some constant  $G(R)$ , we can control the covering number of  $C^*$  with respect to the  $L_{B_R}^\infty$  and we have the upper-bound for  $\frac{\delta}{4} \leq \frac{2M'(R)}{L_{\phi^*}^2}$

$$\mathbb{E}[W^*] \leq 2\delta + \frac{4\sqrt{2}}{\sqrt{n}} \int_{\frac{\delta}{4}}^{\frac{2M'(R)}{L_{\phi^*}^2}} \sqrt{n(C, L_{B_{G(R)}}^\infty, L_{\phi^*}^2 u)} du.$$

**Final upper bound** Since the term (27) is zero in average, we obtain our final bound

$$d_{H^c}^\lambda(\hat{z}, z_0)^2 \leq 4\delta + \frac{8\sqrt{2}}{\sqrt{n}} \int_{\frac{\delta}{4}}^{\frac{M'}{L}} \sqrt{n(C, L_{B_{R'}}^\infty, Lu)} du,$$

where  $M' = 2 \max(M(R), M'(R))$  and  $L = \max(L_{\phi^*}^1, L_{\phi^*}^2)$   $\square$

Leveraging the recent regularity results on UOT derived in Gallouët et al. (2021), we can deduce from Proposition 3 an upper-bound for the statistical estimation of UOT potentials.

**Corollary 1.** *Assume that  $\mu$  and  $\nu$  have compact and convex support with densities  $(f, g)$  bounded away from zero and infinity and assume that  $\varphi$  is strictly convex with infinite slope at 0. If  $(f, g)$  are  $k$ -times continuously differentiable with  $k \in \mathbb{N}^*$  then, denoting  $z_0$  an optimal unbalanced OT potential, there exists  $C$  such that the empirical potential  $\hat{z}_C$  verifies*

$$\begin{cases} \mathbb{E}[d_{H^c}^\lambda(\hat{z}_C, z_0)^2] \lesssim n^{-\frac{k+2}{d}} & \text{if } k+2 < d/2, \\ \mathbb{E}[d_{H^c}^\lambda(\hat{z}_C, z_0)^2] \lesssim \frac{\log(n)}{\sqrt{n}} & \text{if } k+2 = d/2, \\ \mathbb{E}[d_{H^c}^\lambda(\hat{z}_C, z_0)^2] \lesssim \frac{1}{\sqrt{n}} & \text{if } k+2 > d/2. \end{cases} \quad (34)$$

*Proof.* Using the Corollary 9 of Gallouët et al. (2021), we can ensure that  $z_0, z_0^*$  are  $(k+2)$ -times continuously differentiable over the support of  $\mu$  and  $\nu$  respectively. Recalling that for all  $x \in \text{supp}(\nu)$

$$\nabla^2 z_0(x) = [\nabla^2 z_0^*(\nabla z_0(x))]^{-1}, \quad (35)$$

and using the fact that  $\nabla z_0$  is a diffeomorphism between the support  $\mu$  and  $\nu$ , we recover that  $z_0$  is  $\lambda$ -strongly convex over  $\text{supp}(\mu)$  where we defined

$$\frac{1}{\lambda} := \sup_{y \in \text{supp}(\nu)} \|\nabla^2 z_0^*(y)\|. \quad (36)$$

Now, recall that in order to apply our previous result, we need to globally bound the strong-convexity constant as well as controlling the sup norm over every ball. To achieve this, we can extend these potentials to the whole domain. Proposition 1.5 in Azagra and Mudarra (2019) provides a  $(k+2)$ -times continuously differentiable convex extension  $\tilde{g}_0$  of  $z_0 - \lambda q$  on the whole domain  $\mathbb{R}^d$ . Defining  $\tilde{z}_0 = \tilde{g}_0 + \lambda q$ , we have that  $\tilde{z}_0$  coincides with  $z_0$  on  $\text{supp}(\mu)$ . Using again the

diffeomorphism property of  $\nabla z_0$  between  $\text{supp}(\mu)$  and  $\text{supp}(\nu)$ , we have that  $\tilde{z}_0^*$  coincides with  $z_0^*$  on  $\text{supp}(\nu)$ . Now let us define

$$C = \{z \mid \|z\|_{L_{B_r}^\infty} \leq \|\tilde{z}_0\|_{L_{B_r}^\infty}, \|\nabla^{k+2} z\|_{L_{B_r}^\infty} \leq \|\nabla^{k+2} \tilde{z}_0\|_{L_{B_r}^\infty}, z \geq l, \\ z \text{ is } \lambda\text{-strongly convex}\},$$

where  $l$  is the minimum of  $\tilde{z}_0$ . The set  $C$  indeed meets Assumption 4 and Assumption 3 hence we can apply Prop. 3 which yields

$$\mathbb{E}[d_{H^\circ}^\lambda(\hat{z}_C, z_0)^2] \lesssim \delta + \frac{1}{\sqrt{n}} \int_{\frac{\delta}{4}}^{\frac{M'}{L}} \sqrt{n(C, L_{B_{R'}}^\infty, Lu)} du. \quad (37)$$

Finally, using van der Vaart and Wellner (1996, Theorem 2.7), we have  $n(C, L_{B_{R'}}^\infty, Lu) \lesssim u^{-\frac{d}{k+2}}$ . If  $\frac{k+2}{d} < 1/2$ , take  $\delta = n^{-\frac{k+2}{d}}$ . For this choice of  $\delta$ ,

$$\frac{1}{\sqrt{n}} \int_{\frac{\delta}{4}}^{\frac{M'}{L}} \sqrt{n(C, L_{B_{R'}}^\infty, Lu)} du \lesssim \frac{1}{\sqrt{n}} (n^{-\frac{k+2}{d}})^{1-\frac{d}{2(k+2)}} \quad (38)$$

$$\lesssim \frac{1}{\sqrt{n}} n^{-\frac{2(k+2)-d}{2d}} \quad (39)$$

$$= n^{-\frac{k+2}{d}}. \quad (40)$$

If  $\frac{k+2}{d} = 1/2$ , take  $\delta = \frac{1}{\sqrt{n}}$ . For this choice of  $\delta$ , the integral is of order  $\log(n)$  which yields the upper-bound

$$\mathbb{E}[d_{H^\circ}^\lambda(\hat{z}_C, z_0)^2] \lesssim \frac{\log(n)}{\sqrt{n}}. \quad (41)$$

Finally, if  $\frac{k+2}{d} > 1/2$ , taking  $\delta = 0$  yields

$$\mathbb{E}[d_{H^\circ}^\lambda(\hat{z}_C, z_0)^2] \lesssim \frac{1}{\sqrt{n}}. \quad (42)$$

□

We see that we have two distinct regimes: a regime where the extra smoothness directly improves the rate of estimation and a highly smooth regime where the rate saturates at  $1/\sqrt{n}$ . In particular, we do not recover the asymptotic rate (with respect to the smoothness)  $1/n$  which is known to be minimax in the case of balanced OT (see Hütter and Rigollet (2021)). We show in the next paragraph that this  $1/\sqrt{n}$  rate can be improved under suitable assumptions on the metric entropy of  $C$ .

### 3.2 Low metric entropy case

When the metric entropy of  $C$  slowly diverges, we can obtain faster rates than  $1/\sqrt{n}$ . The central argument is the localization: thanks to the strong convexity of the semi-dual, we can localize the empirical potential  $\hat{z}$  in a certain

neighborhood of the ground truth  $z_0$ . The next assumption allows to apply the stability result of the semi-dual with a  $L^2$  control on  $\hat{z} - z_0$  and  $\hat{z}^* - z_0^*$ ; From the statistical point of view, it allows to employ localization arguments.

**Assumption 5.** *The conjugate of the entropy  $\varphi^*$  is strongly convex on every compact.*

Hence, instead of controlling the global empirical processes

$$W = \sup_{z \in C} \langle z, \mu - \hat{\mu} \rangle, \quad (43)$$

and

$$W^* = \sup_{z \in C} \langle z^*, \nu - \hat{\nu} \rangle, \quad (44)$$

we simply must control the localized empirical processes

$$W(\tau) := \sup_{z \in C \cap B^\circ(z_0, \tau)} \langle \phi^*(z - q) - \phi^*(z_0 - q), \mu - \hat{\mu} \rangle, \quad (45)$$

and

$$W^*(\tau) := \sup_{z \in C \cap B^\circ(z_0, \tau)} \langle \phi^*(z^* - q) - \phi^*(z_0^* - q), \nu - \hat{\nu} \rangle, \quad (46)$$

where  $\tau$  is a suitable radius and  $B^\circ(z_0, \tau)$  is the ball centered in  $z_0$  of radius  $\tau$  with respect to the  $d_{H^\circ}^\lambda$  pseudo-distance. In the case where the metric entropy of  $C$  grows too fast, the localized processes (45) and (46) behave like the global processes (43) and (44) and we cannot apply localization. However, when the metric entropy of  $C$  is sufficiently low, the following lemma shows that  $W(\tau)$  and  $W^*(\tau)$  are upper bounded with high probability by  $\frac{\tau^{1-\alpha/2}}{\sqrt{n}}$  with  $2 > \alpha > 0$ .

**Lemma 3.** *Under Assumptions 4-5, if we assume that there exists  $(P_\mu, P_\nu)$  and  $\alpha < 2$  such that for every  $u \in \mathbb{R}_{\geq 0}$ ,  $n(C, L^2(\mu), u) \leq P_\mu u^{-\alpha}$  and  $n(C, L^2(\nu), u) \leq P_\nu u^{-\alpha}$ , it holds with probability at least  $1 - e^{-t}$*

$$\begin{cases} W(\tau) \leq \frac{8\sqrt{2P_\mu}}{(1-\frac{\alpha}{2})\sqrt{n(L_{\phi^*}^1)^\alpha}} (K\tau)^{1-\alpha/2} + K\tau\sqrt{\frac{2t}{n}} + \frac{2M(R)L_{\phi^*}^1}{n} \\ W^*(\tau) \leq \frac{8\sqrt{2P_\nu}}{(1-\frac{\alpha}{2})\sqrt{n(L_{\phi^*}^2)^\alpha}} (K'\tau)^{1-\alpha/2} + K'\tau\sqrt{\frac{2t}{n}} + \frac{2M'(R)L_{\phi^*}^2}{n}, \end{cases} \quad (47)$$

where  $L_{\phi^*}^1, L_{\phi^*}^2$  are defined in Equations (31) and (32) respectively and measure local lipschitz behaviors of  $\varphi^*$ ,  $M(R)$  is defined in Assumption 2 and is a uniform bound over  $B_R$  of the potentials in  $C$ ,  $M'(R)$  is defined in Lemma 1 and is a uniform bound over  $B_R$  of the conjugate of the potentials in  $C$ , and  $K = K(R, M, \phi^*)$ ,  $K' = K'(R, M, \phi^*, \lambda, l)$  are the embedding constants of  $(C, L^2(\mu))$  and  $(C^*, L^2(\nu))$  in  $H^\circ$ .

*Proof.* The proof relies on the Lipschitz behavior of the Legendre transform that preserves the metric entropy of  $C$  and on the Bousquet concentration inequality. We start by analyzing the term  $W(\tau)$ .

**Term  $W(\tau)$**  Let us denote  $C_0 = \{\phi^*(z - q) - \phi^*(z_0 - q), z \in C \cap B^\circ(z_0, \tau)\}$ . For  $g \in C_0$  of the form  $g = \phi^*(z - q) - \phi^*(z_0 - q)$  with  $z \in C \cap B^\circ(z_0, \tau)$ , we have the pointwise bound for all  $x \in B_R$ ,

$$|g(x)| \leq L_{\phi^*}^1 |z(x) - z_0(x)|, \quad (48)$$

where  $L_{\phi^*}^1 := \sup_{x \in [-M_1, M_1]} |\partial \phi^*(x)|$  with  $M_1 = 2M(R) + R^2$  as in the previous proof. This implies  $\|g\|_{L^2(\mu)} \leq L_{\phi^*}^1 \|z - z_0\|_{L^2(\mu)}$ . Since we assumed  $\phi^*$  strongly convex on every compact, there exists  $K = K(R, M, \phi^*) > 0$  such that  $\|z - z_0\|_{L^2(\mu)} \leq K d_{H^\circ}^\lambda(z, z_0)$  and in particular, all  $g \in C_0$  verifies  $\|g\|_{L^2(\mu)} \leq K\tau$ . Hence, applying Luxburg and Bousquet (2004, Theorem 16), we obtain for all  $\frac{\delta}{4} \leq K\tau$

$$\mathbb{E}[W(\tau)] \leq 2\delta + \frac{4\sqrt{2}}{\sqrt{n}} \int_{\frac{\delta}{4}}^{K\tau} \sqrt{n(C_0, L^2(\mu), u)} \, du. \quad (49)$$

Again, taking  $(g_1, g_2) \in C_0^2$  of the form  $g_1 = \phi^*(z_1 - q) - \phi^*(z_0 - q)$  and  $g_2 = \phi^*(z_2 - q) - \phi^*(z_0 - q)$  with  $(z_1, z_2) \in (C \cap B^\circ(z_0, \tau))^2$ , we have

$$\|g_1 - g_2\|_{L^2(\mu)} \leq L_{\phi^*}^1 \|z_1 - z_2\|_{L^2(\mu)}, \quad (50)$$

and in particular, we recover the upper-bound

$$\mathbb{E}[W(\tau)] \leq 2\delta + \frac{4\sqrt{2}}{\sqrt{n}} \int_{\frac{\delta}{4}}^{K\tau} \sqrt{n(C, L^2(\mu), L_{\phi^*}^1 u)} \, du. \quad (51)$$

Now, we assumed that for all  $u \in \mathbb{R}^+$  we had the upper-bound,  $n(C, L^2(\mu), u) \leq P_\mu u^{-\alpha}$  with  $\alpha < 2$ , we obtain taking  $\delta = 0$  our final upper bound

$$\mathbb{E}[W(\tau)] \leq \frac{4\sqrt{2P_\mu}}{(1 - \frac{\alpha}{2})\sqrt{n(L_{\phi^*}^1)^\alpha}} (K\tau)^{1-\alpha/2}. \quad (52)$$

There remains to bound the process  $W(\tau)$  with high probability. We use for this the Bousquet concentration inequality.

**Lemma 4** (Bousquet, see Theorem 26 in Hütter and Rigollet (2021)). *Let  $\mathcal{F}$  be a class of functions such that for every  $f \in \mathcal{F}$ ,  $\|f\|_{L^2(\mu)}^2 \leq \sigma^2$  and  $\|f\|_{L^\infty(\mu)} \leq M$ , then for all  $t > 0$ , we have with probability at least  $1 - e^{-t}$*

$$\sup_{f \in \mathcal{F}} \sqrt{n} |\langle f, \mu - \hat{\mu} \rangle| \leq 2\mathbb{E}[\sup_{f \in \mathcal{F}} \sqrt{n} |\langle f, \mu - \hat{\mu} \rangle|] + \sigma\sqrt{2t} + \frac{M}{\sqrt{n}}t. \quad (53)$$

Applying this result to  $W(\tau)$  yields that with probability at least  $1 - e^{-t}$ ,

$$W(\tau) \leq \frac{8\sqrt{2P_\mu}}{(1 - \frac{\alpha}{2})\sqrt{n(L_{\phi^*}^1)^\alpha}} (K\tau)^{1-\alpha/2} + K\tau\sqrt{\frac{2t}{n}} + \frac{2tM(R)L_{\phi^*}^1}{n}, \quad (54)$$

where we used the pointwise upper-bound (48) and where  $M(R)$  is the constant such that  $\forall z \in C, \|z\|_{L_{B_R}^\infty} \leq M(R)$ .

**Term  $W^*(\tau)$**  We can apply the same reasoning as previously. Indeed, as shown in Lemma 1, there exists a constant  $M'(R)$  such that for all  $z \in C$ ,  $\|z^*\|_{L^\infty_{\tilde{B}_R}} \leq M'(R)$ . In particular, since the potentials  $z^*$  are bounded, we can also leverage the local strong convexity of  $\phi^*$  that yields a constant  $K' = K'(R, M, \phi^*, \lambda, l) > 0$  such that for every  $z \in C$ ,  $\|(z - z_0)^*\|_{L^2(\nu)} \leq K' d_{H^\circ}^\lambda(z, z_0)$ . Hence we recover that with probability at least  $1 - e^{-t}$ ,

$$W^*(\tau) \leq \frac{8\sqrt{2P_\nu}}{(1 - \frac{\alpha}{2})\sqrt{n(L_{\phi^*}^2)^\alpha}} (K'\tau)^{1-\alpha/2} + K'\tau\sqrt{\frac{2t}{n}} + \frac{2tM'(R)L_{\phi^*}^2}{n}. \quad (55)$$

□

Then, the informal reasoning is as follows: if we take  $\tau = d_{H^\circ}^\lambda(\hat{z}, z_0)$ , we have the upper bound:  $\tau^2 \leq W(\tau) + W^*(\tau) \leq \frac{\tau^{1-\alpha/2}}{\sqrt{n}}$  which constrains  $\tau$  as  $\tau^2 \leq n^{-\frac{1}{1+\alpha/2}}$ .

**Proposition 4.** *Under Assumptions 1-5, if we assume that there exists  $(P_\mu, P_\nu)$  and  $\alpha < 2$  such that for every  $u \in \mathbb{R}_{\geq 0}$ ,  $n(C, L^2(\mu), u) \leq P_\mu u^{-\alpha}$  and  $n(C, L^2(\nu), u) \leq P_\nu u^{-\alpha}$  then  $\forall n \geq 1$ ,*

$$\mathbb{E}[d_{H^\circ}^\lambda(\hat{z}, z_0)^2] \lesssim n^{-\frac{1}{1+\alpha/2}}, \quad (56)$$

where  $\lesssim$  hides constants that do not depend on  $n$ .

*Proof.* For  $\tau > 0$ , define  $s = \frac{\tau}{\tau + d_{H^\circ}^\lambda(\hat{z}, z_0)}$  and  $\hat{z}_s = (1-s)z_0 + s\hat{z}$ . By local strong convexity of  $J$ , we have

$$d_{H^\circ}^\lambda(\hat{z}_s, z_0)^2 \leq J(\hat{z}_s) - J(z_0). \quad (57)$$

Let us decompose the right hand side as  $J(\hat{z}_s) - \hat{J}(\hat{z}_s) - (J(z_0) - \hat{J}(z_0)) + \hat{J}(\hat{z}_s) - \hat{J}(z_0)$ . By convexity of  $\hat{J}$ , the last term can be upper-bounded by  $s\hat{J}(\hat{z}) + (1-s)\hat{J}(z_0) - \hat{J}(z_0) = s(\hat{J}(\hat{z}) - \hat{J}(z_0))$ . Since  $\hat{z}$  is the minimizer of the empirical semi-dual, we have in particular that  $s(\hat{J}(\hat{z}) - \hat{J}(z_0)) \leq 0$  which gives

$$\begin{aligned} d_{H^\circ}^\lambda(\hat{z}_s, z_0)^2 &\leq J(\hat{z}_s) - \hat{J}(\hat{z}_s) - (J(z_0) - \hat{J}(z_0)) \\ &= \langle \phi^*(\hat{z}_s - q) - \phi^*(z_0 - q), \mu - \hat{\mu} \rangle + \langle \phi^*(\hat{z}_s^* - q) - \phi^*(z_0^* - q), \nu - \hat{\nu} \rangle. \end{aligned}$$

Now, since  $d_{H^\circ}^\lambda(\hat{z}_s, z_0) = \frac{\tau d_{H^\circ}^\lambda(\hat{z}, z_0)}{\tau + d_{H^\circ}^\lambda(\hat{z}, z_0)} \leq \tau$ , we recover in the end  $d_{H^\circ}^\lambda(\hat{z}_s, z_0)^2 \leq W(\tau) + W^*(\tau)$ .

Let us now consider  $A = \{\tau, d_{H^\circ}^\lambda(\hat{z}, z_0) \geq \tau\}$ . We wish to recover an upper-bound on  $A$ . Remark that  $A = \{\tau, d_{H^\circ}^\lambda(\hat{z}_s, z_0) \geq \frac{\tau}{2}\}$ . In particular, every  $\tau \in A$  verifies with probability at least  $1 - e^{-t}$

$$\frac{\tau^2}{4} \leq \kappa \frac{\tau^{1-\alpha/2}}{\sqrt{n}} + (K + K')\tau\sqrt{\frac{2t}{n}} + \frac{t\kappa'}{n}, \quad (58)$$

where  $\kappa$  and  $\kappa'$  are given in Lemma 3 defined as

$$\begin{cases} \kappa = \frac{8\sqrt{2}}{(1-\frac{\alpha}{2})} \left[ \frac{\sqrt{P_\mu} K^{1-\alpha/2}}{(L_{\varphi^*}^1)^{\frac{\alpha}{2}}} + \frac{\sqrt{P_\nu} (K')^{1-\alpha/2}}{(L_{\varphi^*}^2)^{\frac{\alpha}{2}}} \right] \\ \kappa' = 2(M(R)L_{\varphi^*}^1 + M'(R)L_{\varphi^*}^2). \end{cases} \quad (59)$$

. Let  $A_n = \{\tau \in A, \tau \geq \frac{1}{\sqrt{n}}\}$ . For  $\tau \in A_n$ , we have

$$\frac{\tau^2}{4} \leq \kappa \frac{\tau^{1-\alpha/2}}{\sqrt{n}} + (K + K')\tau \sqrt{\frac{2t}{n}} + \frac{t\kappa'\tau}{\sqrt{n}}. \quad (60)$$

Assuming that  $t \geq 1$ , we have two cases

**Case 1** If  $\tau \leq 1$ , we have

$$\frac{\tau^2}{4} \leq \frac{t\eta\tau^{1-\alpha/2}}{\sqrt{n}}, \quad (61)$$

where  $\eta = (\kappa + \kappa' + \sqrt{2}(K + K'))$  and we recover  $\tau \leq \frac{(4\eta t)^{\frac{1}{1+\alpha/2}}}{n^{\frac{1}{2+\alpha}}}$ .

**Case 2** If  $\tau \geq 1$ , we have  $\frac{\tau^2}{4} \leq \frac{t\eta\tau}{\sqrt{n}}$  i.e.  $\tau \leq \frac{4t\eta}{\sqrt{n}}$ .

In any case, for  $t \geq 1$ , we have with probability at least  $1 - e^{-t}$

$$\sup(A) \leq \frac{(4\eta't)^{\frac{1}{1+\alpha/2}} + (4\eta't)}{n^{\frac{1}{2+\alpha}}}, \quad (62)$$

where we defined  $\eta' = \max(\eta, 1)$ . Now, by definition of  $A$ , we have for all  $\epsilon > 0$ ,  $d_{H^\circ}^\lambda(\hat{z}, z_0) \leq \sup(A) + \epsilon$ . Taking  $\epsilon \rightarrow 0$  gives that with probability at least  $1 - e^{-t}$ , for  $t \geq 1$

$$d_{H^\circ}^\lambda(\hat{z}, z_0) \leq \frac{(4\eta't)^{\frac{1}{1+\alpha/2}} + (4\eta't)}{n^{\frac{1}{2+\alpha}}} \quad (63)$$

$$\leq \frac{8\eta't}{n^{\frac{1}{2+\alpha}}}. \quad (64)$$

And in particular,  $d_{H^\circ}^\lambda(\hat{z}, z_0)^2 \leq \frac{64(\eta')^2 t^2}{n^{\frac{1}{1+\alpha/2}}}$  with probability at least  $1 - e^{-t}$  for  $t \geq 1$ . We denote  $X$  the random variable  $d_{H^\circ}^\lambda(\hat{z}, z_0)^2$ . Since  $X$  is nonnegative almost surely, we can apply Fubini's formula

$$\mathbb{E}[X] = \int_0^\infty P(X > u) du. \quad (65)$$

Let us make the change of variable  $u = \frac{64(\eta')^2 t^2}{n^{\frac{1}{1+\alpha/2}}}$ ,

$$\mathbb{E}[X] = \frac{128(\eta')^2}{n^{\frac{1}{1+\alpha/2}}} \left( \int_0^1 tP(X > \frac{64(\eta')^2 t^2}{n^{\frac{1}{1+\alpha/2}}}) dt + \int_1^\infty tP(X > \frac{64(\eta')^2 t^2}{n^{\frac{1}{1+\alpha/2}}}) dt \right).$$



The integrand in the first term is upper-bounded by 1 and the integrand on the second term is upper bounded by  $te^{-t}$ . Hence we obtain

$$\begin{aligned}\mathbb{E}[d_{H^0}^\lambda(\hat{z}, z_0)^2] &\leq \frac{128(\eta')^2}{n^{\frac{1}{1+\alpha/2}}} \left(1 + \int_1^\infty te^{-t} dt\right) \\ &= \frac{128(1+2e^{-1})(\eta')^2}{n^{\frac{1}{1+\alpha/2}}}.\end{aligned}$$

□

An immediate consequence of this result is that we can improve the rate derived in Corollary 1 and recover the asymptotic behavior in  $1/n$  when the smoothness grows.

**Corollary 2.** *Assume that  $\mu$  and  $\nu$  have compact and convex support with densities  $(f, g)$  bounded away from zero and infinity and assume that  $\varphi^*$  is strongly convex on every compact. If  $(f, g)$  are  $k$ -times continuously differentiable with  $k+2 > d/2$  then, denoting  $z_0$  an optimal unbalanced OT potential, there exists  $C$  such that the empirical potential  $\hat{z}_C$  verifies*

$$\mathbb{E}[d_{H^0}^\lambda(\hat{z}_C, z_0)^2] \lesssim n^{-\frac{1}{1+\frac{d}{2(k+2)}}}. \quad (66)$$

*Proof.* As in the proof of Corollary 1, the original potential  $z_0$  is  $\lambda$ -strongly convex and  $(k+2)$ -times continuously differentiable over  $\text{supp}(\mu)$ . It can be extended to  $\tilde{z}_0$  that is also  $\lambda$ -strongly convex and  $(k+2)$ -times continuously differentiable but over the whole domain  $\mathbb{R}^d$ . The function  $\tilde{z}_0$  is such that it coincides with  $z_0$  on  $\text{supp}(\mu)$  and its conjugate  $\tilde{z}_0^*$  coincides with  $z_0^*$  on  $\text{supp}(\nu)$ . We define the set

$$\begin{aligned}C = \{z \mid \|z\|_{L_{B_r}^\infty} \leq \|\tilde{z}_0\|_{L_{B_r}^\infty}, \|\nabla^{k+2}z\|_{L_{B_r}^\infty} \leq \|\nabla^{k+2}\tilde{z}_0\|_{L_{B_r}^\infty}, z \geq l, \\ z \text{ is } \lambda\text{-strongly convex}\},\end{aligned}$$

where  $l$  is the minimum of  $\tilde{z}_0$ . Theorem 2.7 of van der Vaart and Wellner (1996) ensures that the metric entropy at scale  $u$  of  $C$  with respect to  $L^2(\mu)$  and  $L^2(\nu)$  is upper bounded by  $u^{-\frac{k+2}{d}}$ . Hence we can apply Prop. 4 and obtain

$$\mathbb{E}[d_{H^0}^\lambda(\hat{z}_C, z_0)^2] \lesssim n^{-\frac{1}{1+\frac{d}{2(k+2)}}}. \quad (67)$$

□

Hence, using Corollary 1 and Corollary 2, we obtain a rate of  $n^{-\frac{\alpha+2}{d}}$  when  $\alpha+2 < d/2$  and  $n^{-\frac{1}{1+\frac{d}{2(\alpha+2)}}}$  when  $\alpha+2 > d/2$ ; note that we continuously transition from one rate to another when  $\alpha+2 = d/2$  where we recover (up to log factor) the parametric rate  $1/\sqrt{n}$ . Even though the fast rates do not encompass the balanced case as they require  $\varphi^*$  to be at least locally strongly convex, we still compare our rates to the ones of Hütter and Rigollet (2021).

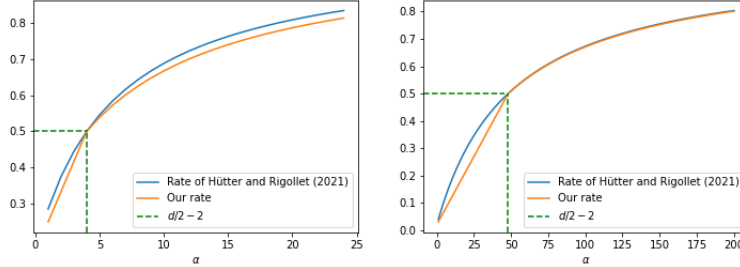


Figure 1: Comparison of our rates against the rates of Hutter and Rigollet (2021): on the left for  $d = 12$  and on the right for  $d = 100$ .

Under the same assumptions, they propose estimators that achieve a rate in  $n^{-\frac{\alpha+1}{\alpha+d/2}}$ . As shown in Fig. 1 for the case  $d = 100$ , their rate is faster for any  $\alpha > 0$  yet when we transition in the highly smooth regime  $\alpha + 2 > d/2$ , our rate closely matches theirs. This discrepancy is due to the fact that we have no bias in our model *i.e.* we assumed  $z_0 \in C$ . On their side, Hütter and Rigollet (2021) fixed  $C$  to be a finite wavelet basis that does not necessarily contain  $z_0$ . In particular, they improve the bias variance trade-off on two levels: in the low smooth regime, they can benefit the acceleration given by the localization as they choose a small class of functions and for the same reason, in the highly smooth regime, they better leverage the localization.

## References

- Daniel Azagra and Carlos Mudarra. Smooth convex extensions of convex functions. *Calculus of Variations and Partial Differential Equations*, 2019.
- Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 1991.
- Lenaïc Chizat. *Unbalanced optimal transport: Models, numerical methods, applications*. PhD thesis, Université Paris sciences et lettres, 2017.
- Lenaïc Chizat, Pierre Roussillon, Flavien Léger, François-Xavier Vialard, and Gabriel Peyré. Faster wasserstein distance estimation with the sinkhorn divergence. In *NeurIPS*, 2020.
- Nicolas Courty, Rémi Flamary, Devis Tuia, and Alain Rakotomamonjy. Optimal transport for domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.
- Nabarun Deb, Promit Ghosal, and Bodhisattva Sen. Rates of Estimation of Optimal Transport Maps using Plug-in Estimators via Barycentric Projections. *arXiv e-prints*, art. arXiv:2107.01718, July 2021.

- Jean Feydy, Benjamin Charlier, François-Xavier Vialard, and Gabriel Peyré. Optimal transport for diffeomorphic registration. In *MICCAI*, 2017.
- Thomas Gallouët, Roberta Ghezzi, and François-Xavier Vialard. Regularity theory and geometry of unbalanced optimal transport, 2021.
- Aude Genevay, Marco Cuturi, Gabriel Peyré, and Francis Bach. Stochastic Optimization for Large-scale Optimal Transport. In *NeurIPS*, 2016.
- Jan-Christian Hütter and Philippe Rigollet. Minimax estimation of smooth optimal transport maps. *The Annals of Statistics*, 2021.
- Huidong Liu, Xianfeng Gu, and Dimitris Samaras. Wasserstein gan with quadratic transport cost. In *Proceedings of the IEEE/CVF international conference on computer vision*, 2019.
- Ulrike von Luxburg and Olivier Bousquet. Distance-based classification with lipschitz functions. *JMLR*, 2004.
- Quentin Mérigot, Filippo Santambrogio, and Clément Sarrazin. Non-asymptotic convergence bounds for wasserstein approximation using point clouds. *NeurIPS*, 2021.
- Debarghya Mukherjee, Aritra Guha, Justin M Solomon, Yuekai Sun, and Mikhail Yurochkin. Outlier-robust optimal transport. In *ICML*, 2021.
- Boris Muzellec, Adrien Vacher, Francis Bach, François-Xavier Vialard, and Alessandro Rudi. Near-optimal estimation of smooth transport maps with kernel sums-of-squares, 2021.
- Victor Panaretos and Yoav Zemel. *An Invitation to Statistics in Wasserstein Space*. 01 2020. ISBN 978-3-030-38437-1. doi: 10.1007/978-3-030-38438-8.
- Victor M. Panaretos and Yoav Zemel. Statistical aspects of wasserstein distances. *Annual Review of Statistics and Its Application*, 6(1):405–431, Mar 2019. ISSN 2326-831X. doi: 10.1146/annurev-statistics-030718-104938. URL <http://dx.doi.org/10.1146/annurev-statistics-030718-104938>.
- Aram-Alexandre Pooladian and Jonathan Niles-Weed. Entropic estimation of optimal transport maps, 2021.
- Aram-Alexandre Pooladian, Marco Cuturi, and Jonathan Niles-Weed. Debiasser Beware: Pitfalls of Centering Regularized Transport Maps. *arXiv e-prints*, art. arXiv:2202.08919, February 2022.
- Geoffrey Schiebinger, Jian Shu, Marcin Tabaka, Brian Cleary, Vidya Subramanian, Aryeh Solomon, Joshua Gould, Siyan Liu, Stacie Lin, Peter Berube, et al. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 2019.

- Vivien Seguy, Bharath B. Damodaran, Remi Flamary, Nicolas Courty, Antoine Rolet, and Mathieu Blondel. Large-scale optimal transport and mapping estimation. In *ICLR*, 2018.
- Thibault Séjourné, Jean Feydy, François-Xavier Vialard, Alain Trounev, and Gabriel Peyré. Sinkhorn divergences for unbalanced optimal transport. *arXiv preprint arXiv:1910.12958*, 2019.
- Aad W van der Vaart and Jon A Wellner. *Weak convergence and empirical processes*. Springer, 1996.
- Karren D. Yang, Karthik Damodaran, Saradha Venkatachalapathy, Ali C Soylemezoglu, G. V. Shivashankar, and Caroline Uhler. Predicting cell lineages using autoencoders and optimal transport. *PLoS Computational Biology*, 16, 2020.