A short proof on the rate of convergence of the empirical measure for the Wasserstein distance

Vincent Divol

A short proof on the rate of convergence of the empirical measure for the Wasserstein distance

Vincent Divol[a]

[a] *Université Paris-Saclay and Inria Saclay, France*

Abstract

We provide a short proof that the Wasserstein distance between the empirical measure of a n-sample and the estimated measure is of order $n^{-1/d}$, if the measure has a lower and upper bounded density on the d-dimensional flat torus.

For $1 \le p < \infty$, let W_p be the p-Wasserstein distance between measures, defined for two probability measures μ, ν with finite pth moments supported on a metric space (Ω, ρ) by

$$W_p(\mu, \nu) := \inf_{\pi \in \Pi(\mu,\nu)} C_p(\pi)^{1/p}, \tag{1}$$

where $\Pi(\mu, \nu)$ is the set of transport plans between μ and ν, that is the set of probability measures on $\Omega \times \Omega$, with first marginal μ and second marginal ν, and $C_p(\pi) = \iint \rho(x,y)^p \mathrm{d}\pi(x,y)$ is the cost of the plan π. We define the distance W_∞ by replacing the quantity $C_p(\pi)^{1/p}$ by the π-essential supremum of ρ.

Let μ be a probability measure on some metric space (Ω, ρ), and let μ_n be the empirical measure associated to a n-sample $X_1, \ldots, X_n$ of law μ. The question of studying rates of convergence between μ and μ_n for Wasserstein distances W_p has attracted a lot of attention over recent years (see e.g. [5, 6]). If no bounds on the density are assumed, then the quantity $\mathbb{E} W_p(\mu_n, \mu)$ is known to be bounded by a quantity of order $n^{-\frac{1}{2p}} + n^{-\frac{1}{d}}$ when Ω is a d-dimensional domain, and this bound is tight (see e.g [5]). For $p = \infty$, Nicolás García Trillos and Dejan Slepčev [6] have shown that $\mathbb{E} W_\infty(\mu_n, \mu)$

Email address: `vincent.divol@inria.fr` (Vincent Divol)

is of order $(\log n/n)^{1/d}$ (for $d \geq 3$) in the case where μ has a density f which is lower bounded and upper bounded on some convex domain Ω. As $W_p \leq W_\infty$, the same rate also holds for any $1 \leq p \leq \infty$. This exhibits the following phenomenon: when $2p > d$, the problem of reconstructing μ for the Wasserstein distance is strictly harder if no bounds on the underlying density are assumed.

In this note, we propose to give a short proof of the fact that $\mathbb{E}W_p(\mu_n, \mu) \lesssim n^{-1/d}$ (for $d \geq 3$) for bounded densities. We restrict to the case where Ω is the d-dimensional flat torus Ω in order to avoid complications due to boundary effects. Let $\mathcal{P}_0$ be the set of probability distributions on Ω, having a density f satisfying $f_{\min} \leq f \leq f_{\max}$ for some $f_{\max} \geq f_{\min} > 0$.

Theorem. *Let $\mu \in \mathcal{P}_0$ and $1 \leq p < \infty$. Then, there exists a constant C such that*

$$\mathbb{E}W_p(\mu_n, \mu) \leq C \begin{cases} n^{-1/d} & \text{if } d \geq 3, \\ (\log n)^{1/2} n^{-1/2} & \text{if } d = 2, \\ n^{-1/2} & \text{if } d = 1. \end{cases} \tag{2}$$

The standard approach for bounding the distance $W_p(\mu_n, \mu)$ consists in precisely assessing the masses given by the measures μ_n and μ on dyadic partitions of the domain Ω (see e.g. [6]). We propose to take a different route by relying on a result from [3] which asserts that the Wasserstein distance is controlled by the pointed negative Sobolev distance when comparing measures having lower bounded densities. The proof is then completed by using tools from Fourier analysis.

We also note that minimax results from [7] (proven for measures on the cube) can be straightforwardly adapted to the setting of the flat torus. In particular, those results imply that the rates exhibited in the theorem are optimal on the class $\mathcal{P}_0$ (up to a logarithmic factor for $d = 2$).

The proof

As $W_p \geq W_q$ if $p \geq q$, we may assume that $p \geq 2$. The proof of the theorem is heavily based on the following result of optimal transport theory, appearing in [3, 2]. Let p^* be the conjugate exponent of p. For $\phi \in L_p$ with $\int \phi = 0$, introduce the pointed negative Sobolev norm

$$\|\phi\|_{\dot{H}_p^{-1}} := \sup \left\{ \int \phi\psi, \ \|\nabla\psi\|_{L_{p^*}} \leq 1 \right\}, \tag{3}$$

2

where the supremum is taken over all smooth functions ψ defined on Ω.

Lemma 1. *Let μ, ν be two measures on Ω having densities f, g. Assume that $f \geq f_{\min}$. Then,*

$$W_p(\mu, \nu) \leq p f_{\min}^{1/p-1} \|f - g\|_{\dot{H}_p^{-1}}. \tag{4}$$

Let K be a smooth radial nonnegative function with $\int K = 1$, supported on the unit ball and, for $h > 0$, let $K_h = h^{-d} K(\cdot/h)$. Let $\mu_{n,h}$ be the measure having density $K_h * \mu_n$ on Ω, i.e. the density at a point $x \in \Omega$ is given by $f_{n,h}(x) := \sum_{j=1}^n K_h(x - X_j)/n$.

Lemma 2. *We have $W_p(\mu_n, \mu_{n,h}) \leq C_0 h$, where $C_0 = \left(\int |x|^p K(x) \mathrm{d}x \right)^{1/p}$.*

Proof. Consider the unique transport plan π_j between $K_h * \delta_{X_j}$ and δ_{X_j}. The cost of π_j is equal to $\int |x - X_j|^p K_h(x - X_j) \mathrm{d}x = h^p \int |x|^p K(x) \mathrm{d}x$. The measure $\frac{1}{n} \sum_{j=1}^n \pi_j$ is a transport plan between $\mu_{n,h}$ and μ_n, with associated cost equal to $h^p \int |x|^p K(x) \mathrm{d}x$. $\square$

By Lemmas 1 and 2,

$$
\begin{aligned}
\mathbb{E} W_p(\mu_n, \mu) &\leq \mathbb{E} W_p(\mu_n, \mu_{n,h}) + \mathbb{E} W_p(\mu_{n,h}, \mu) \\
&\leq C_0 h + p f_{\min}^{1/p-1} \mathbb{E} \|f_{n,h} - f\|_{\dot{H}_p^{-1}}.
\end{aligned}
\tag{5}
$$

To further bound this quantity, we use the following relation between the negative Sobolev norm and the Fourier decomposition of a signal. Given $\phi \in L_p$, we let $\hat{\phi}$ be the sequence of Fourier coefficients of ϕ (indexed by $\mathbb{Z}^d$) and denote by $^\vee$ the inverse Fourier transform. Let $|x| := \sum_{i=1}^d |x_i|$ for $x \in \mathbb{R}^d$. A multiplier s is a bounded sequence indexed by $\mathbb{Z}^d$ such that the operator $\phi \in L_p \mapsto (s\hat{\phi})^\vee \in L_p$ is bounded. A sufficient condition for a sequence to be a multiplier is given by Mikhlin multiplier theorem [1, Theorem 3.6.7, Theorem 5.2.7].

Lemma 3. *Let $s : \mathbb{R}^d \to \mathbb{R}$ be a smooth function such that $|\partial^\alpha s(\xi)| \leq B|\xi|^{-|\alpha|}$ for every multiindex α with $|\alpha| \leq d/2 + 1$. Then, the sequence $(s(m))_{m \in \mathbb{Z}^d}$ is a multiplier with corresponding operator of norm smaller than $C_{p,d} B$.*

Let $a : \mathbb{R}^d \to \mathbb{R}$ be a smooth function with $a(\xi) = 1/|\xi|$ for $|\xi| \geq 1$ and $a(0) = 0$. Let $\mathcal{A}$ be the associated multiplier operator (by Lemma 3) defined by $\mathcal{A}(\phi) = (a\hat{\phi})^\vee$.

Lemma 4. *Let $\phi \in L_p$ with $\int \phi = 0$. Then, $\|\phi\|_{\dot{H}_p^{-1}} \leq C_5 \|\mathcal{A}(\phi)\|_{L_p}$.*

Proof. Let $\psi : \Omega \to \mathbb{R}$ be a smooth function with $\|\nabla\psi\|_{L_{p^*}} \le 1$. As $\hat{\phi}(0) = 0$, we have

$$\int \phi\psi = \sum_{m\in\mathbb{Z}^d} \hat{\phi}(m)\hat{\psi}(m) = \sum_{m\in\mathbb{Z}^d} a(m)\hat{\phi}(m)|m|\hat{\psi}(m) \le \|\mathcal{A}(\phi)\|_{L_p}\|(|\cdot|\hat{\psi})^\vee\|_{L_{p^*}}.$$

Note that $|\cdot| = \sum_{i=1}^d \varepsilon_i e_i$, where $e_i(m) = m_i$ and $\varepsilon_i(m)$ is the sign of m_i. As ε_i is a multiplier (by Lemma 3), we have $\|(|\cdot|\hat{\psi})^\vee\|_{L_{p^*}} \le c\sum_{i=1}^d \|(e_i\hat{\psi})^\vee\|_{L_{p^*}} = c\sum_{i=1}^d \|\partial_i\psi\|_{L_{p^*}} \le C_5$. $\qquad\square$

Hence, to conclude, it suffices to bound $\mathbb{E}\|\mathcal{A}(f_{n,h} - f)\|_{L_p} \le \|\mathcal{A}(f_h - f)\|_{L_p} + \mathbb{E}\|\mathcal{A}(f_{n,h} - f_h)\|_{L_p}$.

Bound of the bias. Let κ be the Fourier transform of K. As K is smooth and compactly supported, κ is a multiplier by Lemma 3. Also, the function $M = a \cdot (\kappa - 1)$ is a multiplier as a product of multiplier. Remark that $\hat{f}_h - \hat{f} = (\kappa(h\cdot) - 1)\hat{f}$, so that $\mathcal{A}(f_h - f) = h(M(h\cdot)\hat{f})^\vee$. As the multiplier norms of M and $M(h\cdot)$ are equal [1, Theorem 3.6.7], we have

$$\|\mathcal{A}(f_h - f)\|_{L_p} \le hC_6\|f\|_{L_p} \le hC_6 f_{\max}. \tag{6}$$

Bound of the fluctuations. Eventually, we bound

$$\mathbb{E}\|\mathcal{A}(f_{n,h} - f_h)\|_{L_p} \le \mathbb{E}\left[\|\mathcal{A}(f_{n,h} - f_h)\|_{L_p}^p\right]^{1/p}. \tag{7}$$

The random variable $\mathcal{A}(f_{n,h})$ is equal to $n^{-1}\sum_{j=1}^n U_j$, where $U_j := \mathcal{A}(K_h * \delta_{X_j}) = \mathcal{A}(K_h)(\cdot - X_j)$ and $\mathbb{E}U_j = \mathcal{A}(f_h)$. We control the expectation of the L_p-norm of the sum of i.i.d. centered functions thanks to the next lemma, which is a direct consequence of Rosenthal inequality [4].

Lemma 5. *Let $U_1, \ldots, U_n$ be i.i.d. functions on L_p. Then, the expectation $\mathbb{E}\left\|\frac{1}{n}\sum_{i=1}^n (U_i - \mathbb{E}U_i)\right\|_{L_p}^p$ is smaller than*

$$C_p n^{-p/2} \int \left(\mathbb{E}|U_1(x)|^2\right)^{p/2} \mathrm{d}x + C_p n^{1-p} \int \mathbb{E}\left[|U_1(x)|^p\right]\mathrm{d}x. \tag{8}$$

Let v_h be the sequence in $\ell_{p^*}(\mathbb{Z}^d)$ defined by $v_h(m) = a(m)\kappa(hm)$ for $m \in \mathbb{Z}^d$. By a change of variable, we obtain

$$\mathbb{E}\left[|U_1(x)|^p\right] = \int f(y)|\mathcal{A}(K_h)(x - y)|^p\mathrm{d}y \le f_{\max}\|\mathcal{A}(K_h)\|_{L_p}^p \le f_{\max}\|v_h\|_{\ell_{p^*}}^p, \tag{9}$$

4

where, at the last line, we applied Hausdorff-Young inequality [8, Section XII.2]. The last step consists in bounding $\|v_h\|_{\ell_{p^*}}^{p^*}$. We separate this quantity into two parts: $S_0 = \sum_{|hm|\le 1} |v_h(m)|^{p^*}$ and $S_1 = \sum_{|hm|>1} |v_h(m)|^{p^*}$. To bound S_0, we use that κ is bounded on the unit ball, so that S_0 is of the order

$$\sum_{|hm|\le 1} |m|^{-p^*} \lesssim \begin{cases} h^{p^*-d} & \text{if } d \ge 3 \text{ or } (d=2 \text{ and } p>2) \\ -\log h & \text{if } p=d=2 \\ 1 & \text{if } d=1. \end{cases} \tag{10}$$

To bound S_1, we use that $|\kappa(hm)| \le C_\gamma |hm|^{-\gamma}$ for any $\gamma > 0$. Choosing γ such that $\gamma p^* + p^* > d$, we obtain that S_1 is of the order

$$h^{-\gamma p^*} \sum_{|hm|>1} |m|^{-\gamma p^*-p^*} \lesssim h^{p^*-d}. \tag{11}$$

Putting together inequalities (8), (10) and (11) yields that, for h of the order $n^{-1/d}$, the expectation $\mathbb{E}\|\mathcal{A}(f_{n,h} - f_h)\|_{L_p}$ is of the order

$$\begin{cases} h/\sqrt{nh^d} \lesssim n^{-1/d} & \text{if } d \ge 3, \\ \sqrt{(-\log h)/n} \lesssim (\log n)^{1/2} n^{-1/2} & \text{if } d=2, \\ n^{-1/2} & \text{if } d=1. \end{cases} \tag{12}$$

We conclude the proof by putting together the estimates (5), (6) and (12).

Remark 1. For $p = 2$, Mikhlin multiplier theorem can be replaced by Parseval's theorem, further simplifying the proof.

Remark 2. A similar proof shows that the risk of the measure $\mu_{n,h}$ satisfies $\mathbb{E}W_p(\mu_{n,h}, \mu) \lesssim n^{-(s+1)/(2s+d)}$ if f is assumed to be of regularity s. Indeed, we can exploit the regularity of s to show that, if κ has sufficiently many zero derivatives at 0, then the bias term is of order h^{s+1}, while the fluctuation terms is bounded in the same way. We then obtain the desired rate by choosing h of the order $n^{-1/(2s+d)}$. This rate is in accordance with the minimax result of [7], where a modified wavelet density estimator is shown to attain the same rate of convergence.

References

[1] Loukas Grafakos. *Classical Fourier analysis*, volume 2. Springer.

[2] Sloan Nietert, Ziv Goldfeld, and Kengo Kato. From smooth Wasserstein distance to dual Sobolev norm: Empirical approximation and statistical applications, 2021.

[3] Rémi Peyre. Comparison between W_2 distance and $\dot{H}_{-1}$ norm, and localization of Wasserstein distance. *ESAIM. Control, Optimisation and Calculus of Variations*, 24(4), 2018.

[4] Haskell P Rosenthal. On the subspaces of L_p ($p > 2$) spanned by sequences of independent random variables. *Israel Journal of Mathematics*, 8(3):273–303, 1970.

[5] Shashank Singh and Barnabás Póczos. Minimax distribution estimation in Wasserstein distance. *arXiv preprint arXiv:1802.08855*, 2018.

[6] Nicolás Garcia Trillos and Dejan Slepčev. On the rate of convergence of empirical measures in ∞-transportation distance. *Canadian Journal of Mathematics*, 67(6):1358–1383, 2015.

[7] Jonathan Weed and Quentin Berthet. Estimation of smooth densities in Wasserstein distance. In *Conference on Learning Theory*, pages 3118–3119, 2019.

[8] A. Zygmund and R. Fefferman. *Trigonometric Series*. Cambridge Mathematical Library. Cambridge University Press, 2003.