*The Annals of Probability*
2013, Vol. 41, No. 5, 3140–3156
DOI: 10.1214/12-AOP812


# THE EMPIRICAL COST OF OPTIMAL INCOMPLETE TRANSPORTATION

BY EUSTASIO DEL BARRIO[1] AND CARLOS MATRÁN[1]

*IMUVA, Instituto de Matemáticas, Universidad de Valladolid*

We consider the problem of optimal incomplete transportation between the empirical measure on an i.i.d. uniform sample on the $d$-dimensional unit cube $[0, 1]^d$ and the true measure. This is a family of problems lying in between classical optimal transportation and nearest neighbor problems. We show that the empirical cost of optimal incomplete transportation vanishes at rate $O_P(n^{-1/d})$, where $n$ denotes the sample size. In dimension $d \geq 3$ the rate is the same as in classical optimal transportation, but in low dimension it is (much) higher than the classical rate.

**1. Introduction.** Consider two probability measures on $\mathbb{R}^d$, $P$ and $Q$, and the set $\mathbb{T}(P, Q)$ of maps transporting $P$ to $Q$, that is, the set of all measurable maps $T : \mathbb{R}^d \to \mathbb{R}^d$ such that, if the initial space is endowed with the probability $P$, then the distribution of the random variable $T$ is $Q$. Monge's optimal transportation problem consists of relocating a certain amount of mass from its original distribution to a different target distribution minimizing the transportation cost. In more abstract terms, the problem consists of finding a transportation map $T_0 \in \mathbb{T}(P, Q)$ such that

$$T_0 := \operatorname*{arg\,min}_{T \in \mathbb{T}(P,Q)} \int_{\mathbb{R}^d} \|x - T(x)\|^p P(dx).$$

Here, and throughout the paper, we assume $p \geq 1$. Remarkably, under some smoothness assumptions, Monge's problem is intimately related to the $L_p$-Wasserstein distance by

$$\mathcal{W}_p(P, Q) = \min_{T \in \mathbb{T}(P,Q)} \left( \int_{\mathbb{R}^d} \|x - T(x)\|^p P(dx) \right)^{1/p},$$

where the $L_p$-Wasserstein distance between $P$, $Q$ is defined as

$$\mathcal{W}_p(P, Q) := \left( \inf_{\tau \in \mathcal{M}(P,Q)} \left\{ \int \|x - y\|^p \, d\tau(x, y) \right\} \right)^{1/p},$$

Received July 2011; revised September 2012.

[1]Supported in part by the Spanish Ministerio de Educación y Ciencia and FEDER, Grant MTM2011-28657-C02-01.

*MSC2010 subject classifications.* Primary 60B10; secondary 05C70, 60C05.

*Key words and phrases.* Optimal transportation, optimal matching, optimal incomplete transportation, optimal partial matching, random quantization, rates of convergence.

and $\mathcal{M}(P, Q)$ is the set of probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals $P$ and $Q$. This functional is related to very important problems in mathematics, the study of which has led to deep developments in several fields of research and applications, linked to such important names as Ampère, Kantorovich, Rubinstein, Zolotarev and Dobrushin, among others. To avoid a huge number of references, we refer to the books by Rachev and Rüschendorf [16] and by Villani [20, 21], for an updated account of the interest and implications of the problem. However, we emphasize the importance of the topic in the development of the theory of probability metrics and its implications in statistics, particularly in goodness of fit problems. Focusing on such kind of problems, the functional of interest is $\mathcal{W}_p(P_n, Q)$, where $P_n$ is the empirical measure associated to a sample $X_1, \dots, X_n$, of independent identically distributed (i.i.d.) random vectors with law $P$, and $Q$ is any target probability measure on $\mathbb{R}^d$, or $\mathcal{W}_p(P_n, Q_n)$, where $Q_n$ stands for the empirical measure on a second, independent i.i.d. random sample, $Y_1, \dots, Y_n$. These empirical versions are connected to a combinatorial optimization problem, namely the *optimal matching* problem. In fact, $\mathcal{W}_p^p(P_n, Q_n) = T_p(n)$, where

$$T_p(n) := \min_{\pi} \frac{1}{n} \sum_{i=1}^{n} \|X_i - Y_{\pi(i)}\|^p \tag{1.1}$$

and $\pi$ ranges over the permutations of the set $\{1, \dots, n\}$. A lot of work has been devoted to analyzing the rate and mode of convergence of (1.1) and several variants of it, beginning with the seminal paper by Ajtai, Komlos and Tusnady [1] in the case in which both samples come from the same underlying probability law $P$. The problem can be equivalently formulated in terms of $\mathcal{W}_p(P_n, P)$, the distance between the empirical and true distributions. Further references will be provided later, but now let us mention the series of papers authored by Talagrand [17, 18], Talagrand and Yukich [19] and Dobrić and Yukich [10], which in the case when $P$ is the uniform distribution on the $d$-dimensional unit cube, $[0, 1]^d$, essentially shows that

$$\big(T_p(n)\big)^{1/p} = \begin{cases} O_P\big(n^{-1/2}\big), & \text{if } d = 1, \\ O_P\Big(\Big(\dfrac{\log n}{n}\Big)^{1/2}\Big), & \text{if } d = 2, \\ O_P\big(n^{-1/d}\big), & \text{if } d \geq 3. \end{cases} \tag{1.2}$$

This paper deals with the empirical cost of optimal *incomplete* or *partial* transportation. This is the case in which the amount of mass required in the target distribution is smaller than that in the original one. Then, we do not have to move all the original mass, but we can dismiss a fraction of it. Of course, we would like to complete this task with a minimal cost. A more general version is possible if we admit that we only have to fulfill a fraction of the target distribution. The general formulation of this problem, with quadratic cost, has been introduced by Caffarelli and McCann [7], relating it to a Monge–Ampère double obstacle problem. They obtain remarkable results on the existence, uniqueness and regularity of the optimal solutions in a well-separated situation. Figalli [11] improved the results covering the case of nondisjoint supports for the involved probability measures. Independently, Álvarez-Esteban et al. [2] introduced the problem in the context of similarity of probabilities, obtaining a more general result of existence and uniqueness of the optimal solution. Moreover [2] includes almost sure consistency of sample solutions to the true ones. In a subsequent paper, Álvarez-Esteban et al. [3] noticed the faster rate of decay of the cost of empirical incomplete transportation (in the $L_2$ case) and introduced a procedure for testing similarity of probabilities based on this fact.

A convenient mathematical formulation of this optimal incomplete transportation problem can be done with the help of the concept of *trimmings* of a probability.

DEFINITION 1.1. Given $0 \leq \alpha \leq 1$ and Borel probability measures $P$, $R$ on $\mathbb{R}^d$, we say that $R$ is an $\alpha$-trimming of $P$ if $R$ is absolutely continuous with respect to $P$, and the Radon–Nikodym derivative satisfies $\frac{dR}{dP} \leq \frac{1}{1-\alpha}$. The set of all $\alpha$-trimmings of $P$ will be denoted by $\mathcal{R}_\alpha(P)$.

Note that in the extreme case $\alpha = 0$, $\mathcal{R}_0(P)$ is just $P$, while $\mathcal{R}_1(P)$ is the set of all probability measures absolutely continuous with respect to $P$. See [2] for useful alternative characterizations of trimmings of a probability, as well as mathematical properties of the set $\mathcal{R}_\alpha(P)$. Turning back to the partial mass transportation problem, we could represent the target distribution by the probability $Q$ and the initial distribution of mass by $\frac{1}{1-\alpha}P$, $P$ being another probability if the mass required in the target distribution is $1-\alpha$ times the mass in the original locations. An incomplete transportation plan is then a probability measure $\tau$ on $\mathbb{R}^d \times \mathbb{R}^d$ with second marginal equal to $Q$ and first marginal in $\mathcal{R}_\alpha(P)$, and the cost of optimal incomplete transportation is

$$\mathcal{W}_p(\mathcal{R}_\alpha(P), Q) := \min_{R \in \mathcal{R}_\alpha(P)} \mathcal{W}_p(R, Q).$$

In the more general case, with slackness in the target distribution, the optimal incomplete transportation cost would be $\mathcal{W}_p(\mathcal{R}_{\alpha_1}(P), \mathcal{R}_{\alpha_2}(Q))$, the minimal $\mathcal{W}_p$ distance between trimmings of $P$ and $Q$.

This paper gives exact rates of convergence for empirical versions of the optimal incomplete transportation cost. As with classical optimal transportation, the results can be considered in terms of a combinatorial optimization problem, that we call *optimal incomplete matching*. To be precise, assume that we can trim (eliminate) a fixed proportion $\alpha$ of $X$'s points and also of $Y$'s points, and we should only search for the best matching between the nontrimmed samples. Taking for simplicity $m := n - \alpha n$ to be an integer, the new functional of interest is

$$T_{p,\alpha}(n) := \min_{X^*, Y^*} \min_{\pi} \frac{1}{m} \sum_{j=1}^{m} \|X_j^* - Y_{\pi(j)}^*\|^p, \tag{1.3}$$

where $\pi$ varies in the set of permutations of $\{1,\dots,m\}$, $\{X_1^*,\dots,X_m^*\}$ ranges in the subsets of size $m$ of $\{X_1,\dots,X_n\}$, and similarly $\{Y_1^*,\dots,Y_m^*\}$ ranges in the subsets of size $m$ of $\{Y_1,\dots,Y_n\}$. It is easy to check that

$$T_{p,\alpha}(n)=\mathcal{W}_p^p\big(\mathcal{R}_\alpha(P_n),\mathcal{R}_\alpha(Q_n)\big). \tag{1.4}$$

In fact, $\mathcal{W}_p^p(\mathcal{R}_\alpha(P_n),\mathcal{R}_\alpha(Q_n))$ equals the minimum [in $(\pi_{i,j})$] of the linear function $\sum_{i,j=1}^n \|X_i-Y_j\|^p\pi_{i,j}$ subject to the linear constraints $\sum_{i=1}^n \pi_{i,j}\le \frac{1}{n(1-\alpha)}=\frac{1}{m}$, $\sum_{j=1}^n \pi_{i,j}\le\frac{1}{m}$, $\sum_{i,j=1}^n \pi_{i,j}=1$, $\pi_{i,j}\ge 0$ [we are assuming $m=n(1-\alpha)$]. Rescaling we see that $m\mathcal{W}_p^p(\mathcal{R}_\alpha(P_n),\mathcal{R}_\alpha(Q_n))$ equals the minimum [in $(x_{i,j}),(a_i),(b_j)$] of the linear function $\sum_{i,j=1}^n \|X_i-Y_j\|^p x_{i,j}$ subject to the linear constraints $\sum_{i=1}^n x_{i,j}=b_j$, $\sum_{j=1}^n \pi_{i,j}=a_i$, $0\le a_i\le 1$, $0\le b_j\le 1$, $\sum_{i=1}^n a_i=\sum_{j=1}^n b_j=m$, $x_{i,j}\ge 0$. The constraint matrix in this last linear program is totally unimodular (see, e.g., Theorem 13.3 in [15]) and the right-hand side is integer. Hence, the minimum is attained at some integer solution, that is, satisfying $a_i,b_j\in\{0,1\}$, and this implies (1.4).

We will show the somewhat unexpected result that, independently of the value of $\alpha\in(0,1)$, the rates in (1.2) change to

$$\big(T_{p,\alpha}(n)\big)^{1/p}=O_P\big(n^{-1/d}\big) \tag{1.5}$$

for *any* dimension $d\ge 1$. In fact, (1.5) follows from the triangle inequality and

$$\mathcal{W}_p\big(\mathcal{R}_\alpha(P_n),P\big)=O_P\big(n^{-1/d}\big), \tag{1.6}$$

which is the formulation we choose for the results we prove. Our approach relies only on elementary or rather classical tools. In particular, we do not use subadditivity arguments as in [10]. Subadditivity yields a.s. convergence to a constant, rather than just a rate of convergence. On the other hand, the approach in [10] relies on showing subadditivity of a certain Poissonization of the matching functional (subadditivity does not hold for the original matching functional; see Remark 1.1 in [10]). We could also use that approach here, at least for $p=1$ (otherwise duality for optimal matching, which is essential in the cited approach, becomes harder to deal with) but the approximation rate for the Poissonization of the incomplete matching functional would not allow us to recover the present result in dimension $d\le 2$.

The study of the rate of convergence of $D(\mathcal{R}_\alpha(P_n),P)$ for $\alpha\in(0,1)$, for some probability metrics $D$ was started in del Barrio and Matrán [9]. In the case of the Wasserstein metric and dimension $d=1$, the results in [9] already show a very different behavior with respect to the untrimmed case, namely

$$\frac{n}{(\log n)^\nu}\mathcal{W}_p\big(\mathcal{R}_\alpha(P_n),P\big)\to_P 0 \qquad \text{for any } \nu>1 \text{ and every } \alpha>0. \tag{1.7}$$

We close this Introduction mentioning the connection of empirical optimal incomplete tranportation to another important problem in probability, that of *random quantization*. Taking $\alpha=1$ (full trimming) we have that $\mathcal{R}_1(P_n)$ is the set of all probability measures concentrated on the sample points, and $\mathcal{W}_p(\mathcal{R}_1(P_n), P)$ is the minimal $L_p$-cost of relocating a mass distributed according to some probability measure $P$ to a collection of randomly chosen spots $X_1, \dots, X_n$. When $X_1, \dots, X_n$ are $\mathbb{R}^d$-valued i.i.d. random vectors and $P$ is absolutely continuous, the problem can be formulated, in Monge's way, as the minimization of

$$\int_{\mathbb{R}^d} \|x - \varphi(x)\|^p \, dP(x), \tag{1.8}$$

where $\varphi$ varies in the set of all measurable functions with values in $\{X_1, \dots, X_n\}$. Since, for a fixed $x$ in the integrand in (1.8), the distance $\|x - \varphi(x)\|$ is minimized for $\varphi(x) = \arg\min_i \|x - X_i\|$, without any constraint on the capacity to be stored at $X_i$, we obtain that the optimal $\varphi$ is given by this last expression, and hence the optimal cost equals

$$\mathcal{W}_p^p(\mathcal{R}_1(P_n), P) = \int_{\mathbb{R}^d} \min_{1 \le i \le n} \|x - X_i\|^p \, dP(x). \tag{1.9}$$

Random quantization is a well-studied problem; see, for example, the Graf and Luschgy monograph [12] or the more recent paper by Yukich [22]. In particular, the asymptotic behavior of the $L_p$ quantization error is known, hence the rate at which $\mathcal{W}_p(P_{n,1}, P)$ vanishes. A trivial consequence of Definition 1.1 is that $\mathcal{R}_{\alpha_1}(P) \subset \mathcal{R}_{\alpha_2}(P)$ if $0 \le \alpha_1 \le \alpha_2 \le 1$. This implies

$$\mathcal{W}_p(\mathcal{P}_n, P) = \mathcal{W}_p(\mathcal{R}_0(P_n), P) \ge \mathcal{W}_p(\mathcal{R}_\alpha(P_n), P) \ge \mathcal{W}_p(\mathcal{R}_1(P_n), P). \tag{1.10}$$

Hence rates of convergence for the random quantization error are a lower bound for rates of convergence of $\mathcal{W}_p(\mathcal{R}_\alpha(P_n), P)$ for general $\alpha$. In a first look, classical optimal transportation is a *global* problem while random quantization is a *local* one: a point $x$ is mapped through the optimal map $\varphi$ to a sample point which in the case of random quantization, is determined just by sample points which are close to $x$ (the nearest neighbor in fact) while in the case of optimal transportation, two samples with the same sample points in a neighborhood of $x$ may result, however, in very different destinations for $x$ due to capacity constraints. It turns out though, that this different character is only apparent, in terms of rates, in dimensions $d = 1$ or 2. The most relevant fact which we show in this paper is that, again in terms of rates, optimal incomplete transportation shows the same *local* nature as random quantization in *any* dimension.

The remainder of this paper is organized as follows. In Section 2 we give a quick survey on known results about rates of convergence for optimal transportation and random quantization. Section 3 contains new results for optimal incomplete transportation. We consider first the case $d = 1$, and in this case we construct upper and lower envelopes for the optimal solution to the incomplete transportation problem. These are not optimal, but attain the correct rate. Finally, we construct a nearly optimal solution in general dimension starting from the one-dimensional construction.

We will use $EX$ to denote the expected value of a random variable $X$. By $P(\cdot|B)$ [resp., $E(\cdot|B)$, or even $E_B$] we refer to the conditional probability (resp., conditional expectation) given the set $B$. The indicator function of $B$ will be denoted by $I_B$, while the notation $\delta_x$ will be reserved for Dirac's probability measure on the point $x$. Unless otherwise stated, the random vectors will be assumed to be defined on the same probability space $(\Omega, \sigma, \nu)$. We write $\ell^d$ for Lebesgue measure on the space $(\mathbb{R}^d, \beta)$. Finally, convergence in probability (resp., weak convergence of probabilities) will be denoted by $\to_p$ (resp., by $\to_w$), and $\mathcal{L}(X)$ will denote the law of the random vector $X$.

**2. Preliminary results.** The results on the asymptotic behavior of $L_p$-Wasserstein distances between the empirical and parent distributions in the one-dimensional case have been obtained through a quantile representation. If $F$ and $G$ are the distribution functions of $P$ and $Q$ and $F^{-1}$ and $G^{-1}$ are the respective quantile functions, then (see, e.g., Bickel and Freedman [5])

$$\mathcal{W}_p(P, Q) = \left[\int_0^1 \left(F^{-1}(t) - G^{-1}(t)\right)^p dt\right]^{1/p} \tag{2.1}$$

[where $F^{-1}(t) = \inf\{s : F(s) \geq t\}$]. In particular, when $P$ is the uniform distribution on (0, 1), this representation leads to

$$\sqrt{n}\mathcal{W}_p(P_n, P) \to_w \left[\int_0^1 (B(t))^p dt\right]^{1/p} \tag{2.2}$$

with $B(t)$ a Brownian Bridge on [0, 1]; see, for example, [8].

For dimension $d > 1$, there are not explicit expressions for the optimal transportation maps, and limit distribution results as in (2.2) are not available. Rates of convergence to 0 of $\mathcal{W}_p(P_n, P)$ can be given based on different approaches. The case $d = 2$ is the most interesting from the point of view of the mass transportation problem. Ajtai, Komlós and Tusnádi [1] showed that, with probability $1 - o(1)$,

$$C_1\left(\frac{\log n}{n}\right)^{1/2} < \mathcal{W}_1(P_n, Q_n) < C_2\left(\frac{\log n}{n}\right)^{1/2},$$

where $P_n$ and $Q_n$ are the sample distributions corresponding to two independent samples obtained from the uniform distribution on the unit square, $U([0, 1]^2)$. Their combinatorial partition scheme method was refined in Talagrand and Yukich [19] to show [Theorem 1 and Remark (ii) there] that for some constant, $C(p)$,

$$E\left(\mathcal{W}_p\left(P_n, U\left([0, 1]^2\right)\right)\right) \leq C(p)\left(\frac{\log n}{n}\right)^{1/2}. \tag{2.3}$$

The case $d \geq 3$ (and uniform distribution on the $d$-dimensional unit cube) is covered in Talagrand [18]. That paper uses a different approach, based on duality for the optimal transportation problem to give a result (Theorem 1.1), formulated for a very general class of costs functions which includes exponential costs and, as a consequence, implies

$$E\big(\mathcal{W}_p\big(P_n, U\big([0,1]^d\big)\big)\big) \le C(k,p)\frac{1}{n^{1/d}}. \tag{2.4}$$

Further results, dealing with distributions other than the uniform, possibly with unbounded support, are given in Barthe and Bordenave [4].

As we already noted in the Introduction, the so-called random quantizers provide an easy way of giving a lower bound for the rates of convergence of our interest. We give a simple version of the mean asymptotics for the random quantizers, rewritten in terms of the $L_p$-Wasserstein distance between the set of "fully trimmed" sample probabilities and the theoretical distribution, that suffices for our purposes; this is a particular case of Theorem 9.1 in [12].

THEOREM 2.1. *If $X_1,\dots,X_n$ are i.i.d. random vectors uniformly distributed on $[0,1]^d$, then*

$$n^{p/d}E\big(\mathcal{W}_p^p\big(R_1(P_n), U\big([0,1]^d\big)\big)\big) \to \Gamma\bigg(1+\frac{p}{k}\bigg)\omega_d^{-p/d} \qquad \textit{as } n\to\infty,$$

*where* $\omega_d = \frac{\pi^{d/2}}{\Gamma(1+d/2)}$.

For optimal incomplete transportation, a first result on rates of convergence is Theorem 5 in the Appendix of [3], for dimension 1, but it has been largely improved in [9], in the terms expressed in (1.7). For dimension 2 our approach was not successful in going beyond the characteristic "$\log n$" term in the Ajtai–Kómlos–Tusnády result (1.2). This task and the fundamental improvement in dimension 1 are the main goals in this paper. Moreover we notice del Barrio and Matrán [9] also treat the improvement of the "in probability bounds" involved in (1.6) to almost surely bounds. This follows Talagrand's approach [17], continued by Dobrić and Yukich [10], but, in our case, using a powerful concentration inequality of Boucheron et al. [6].

**3. Rates of convergence.** We focus first on the one-dimensional case. Let us consider $n$ distinct points $x_1<\cdots<x_n\in(0,1)$ and set $P_n=\frac{1}{n}\sum_{i=1}^n\delta_{x_i}$, where $\delta_x$ denotes Dirac's measure on $x$. An $\alpha$-trimming of $P_n$ can be written, in terms of a vector $h=(h_1,\dots,h_{n-1})$, as $(P_n)_h=\sum_{i=1}^n b_i\delta_{x_i}$ with $0\le b_i=h_i-h_{i-1}\le\frac{1}{n(1-\alpha)}$ (we set, for convenience, $h_0=0$, $h_n=1$). We therefore write

$$\mathcal{C}_{\alpha,n} := \bigg\{h=(h_1,\dots,h_{n-1})\in\mathbb{R}^{n-1}: 0\le h_i-h_{i-1}\le\frac{1}{n(1-\alpha)}, \\ i=1,\dots,n\bigg\}. \tag{3.1}$$

Our first result is an elementary, but useful, representation of $\mathcal{W}_p(\mathcal{R}_\alpha(P_n),P)$.

LEMMA 3.1. *If* $x_1 < \cdots < x_n \in (0,1)$, $P_n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$, $P$ *is the uniform distribution on* $[0,1]$, $\mathcal{C}_{\alpha,n}$ *is defined by* (3.1) *and* $p \geq 1$, *then*

$$\begin{aligned}
&\mathcal{W}_p^p(\mathcal{R}_\alpha(P_n), P)\\
&\quad = \frac{1}{p+1}(x_1^{p+1} + (1-x_n)^{p+1})\\
&\qquad + \frac{1}{2^p(p+1)}\sum_{i=1}^{n-1}(x_{i+1}-x_i)^{p+1}\\
&\qquad + \min_{h\in\mathcal{C}_{\alpha,n}} \frac{1}{p+1}\sum_{i=1}^{n-1}\left(\frac{x_{i+1}-x_i}{2}\right)^{p+1} f_p\left(\frac{h_i-(x_{i+1}+x_i)/2}{(x_{i+1}-x_i)/2}\right),
\end{aligned}$$

*where* $f_p(y) = (1+|y|)^{p+1} + (1-|y|)^{(p+1)} - 2$ *and* $t^{(p)}$ *denotes the odd extension to* $(-\infty,\infty)$ *of the function* $t^p$ *on* $[0,\infty)$.

PROOF. We note first that the quantile function associated to $(P_n)_h$ takes the value $x_i$ in the interval $(h_{i-1}, h_i]$. Hence, using (2.1), we see that $\mathcal{W}_p^p(\mathcal{R}_\alpha(P_n), P) = \min_{h\in\mathcal{C}_{\alpha,n}}\sum_{i=1}^n A_i$, where $A_i = \int_{h_{i-1}}^{h_i}|x_i - t|^p\,dt$. Since $t^{(p+1)}/(p+1)$ is a primitive of $|t|^p$, we can write

$$\begin{aligned}
A_i &= \int_{h_{i-1}}^{x_i}|x_i-t|^p\,dt + \int_{x_i}^{h_i}|x_i-t|^p\,dt\\
&= \frac{1}{p+1}\left[(x_i-h_{i-1})^{(p+1)} + (h_i-x_i)^{(p+1)}\right].
\end{aligned}$$

From this, recalling that $h_0 = 0$, $h_1 = 1$, we get

$$\sum_{i=1}^n A_i = \frac{1}{p+1}\left[(x_1^{p+1} + (1-x_n)^{p+1}) + \frac{1}{2^p}\sum_{i=1}^{n-1}(x_{i+1}-x_i)^{p+1} + \sum_{i=1}^{n-1} B_i\right]$$

with $B_i = (x_{i+1}-h_i)^{(p+1)} + (h_i - x_i)^{(p+1)} - 2(\frac{x_{i+1}-x_i}{2})^{p+1}$. Now it is easy to see that $B_i = (\frac{x_{i+1}-x_i}{2})^{p+1} f_p(\frac{h_i-(x_{i+1}+x_i)/2}{(x_{i+1}-x_i)/2})$, which completes the proof. $\square$

The function $f_p$ in Lemma 3.1 is a piecewise polynomial for integer $p$. For instance $f_1(y) = 2y^2$, $|y| \leq 1$, $f_1(y) = 2(2|y|-1)$, $|y| > 1$; $f_2(y) = 6y^2$, $y \in \mathbb{R}$. For general $p \geq 1$, $f_p$ is a nonnegative, even and convex function, strictly increasing on $[0,\infty)$, which attains its minimum at $y = 0$, with $f_p(0) = 0$. This suggests that a good trimming vector $h = (h_1,\ldots,h_{n-1}) \in \mathcal{C}_{\alpha,n}$ should be as close as possible to the midranks, $\frac{x_i+x_{i+1}}{2}$. With this observation in mind, we denote

$$\hat{h} = \arg\min_{h\in\mathcal{C}_{\alpha,n}}\sum_{i=1}^{n-1}\left(\frac{x_{i+1}-x_i}{2}\right)^{p+1} f_p\left(\frac{h_i-(x_{i+1}+x_i)/2}{(x_{i+1}-x_i)/2}\right)$$

and define

$$u_i = \max_{i \le j \le n-1} \left( \frac{x_j + x_{j+1}}{2} - \frac{1}{1-\alpha} \frac{j}{n} \right) \vee \frac{-\alpha}{1-\alpha}, \qquad i = 1, \dots, n-1,$$

$u_n = -\frac{\alpha}{1-\alpha}$, $\bar{f}_0 = 0$, and $\bar{f}_i = u_i \wedge 0$, $i = 1, \dots, n$. Finally, we set

$$\bar{h}_i = \bar{f}_i + \frac{1}{1-\alpha} \frac{i}{n}, \qquad i = 0, \dots, n.$$

Note that, for any $h = (h_1, \dots, h_{n-1}) \in \mathcal{C}_{\alpha,n}$, $h_i - \frac{i}{n(1-\alpha)}$ is a sequence that decreases from 0 to $-\frac{\alpha}{1-\alpha}$, while $\bar{f}_i$ is the lowest decreasing sequence from 0 to $-\frac{\alpha}{1-\alpha}$ which lies above the sequence $\frac{x_i + x_{i+1}}{2} - \frac{i}{n(1-\alpha)}$. In the next result we see that $\bar{h}_i$ is a feasible trimming and that feasible solutions that exceed this one cannot be optimal.

LEMMA 3.2. *$\bar{h}_0 = 0$, $\bar{h}_n = 1$ and if $\bar{h} = (\bar{h}_1, \dots, \bar{h}_{n-1})$, then $\bar{h} \in \mathcal{C}_{\alpha,n}$. Furthermore,*

$$\hat{h}_i \le \bar{h}_i, \qquad i = 1, \dots, n-1.$$

PROOF. $\bar{h}_0 = 0$ and $\bar{h}_n = 1$ are obvious. To prove $\bar{h} \in \mathcal{C}_{\alpha,n}$ we check, equivalently, that $0 \ge \bar{f}_i - \bar{f}_{i-1} \ge -\frac{1}{1-\alpha} \frac{1}{n}$, $i = 1, \dots, n$. Clearly, $u_1 \ge \cdots \ge u_n$ and $\bar{f}_1 \le 0 = \bar{f}_0$ and, consequently, $\bar{f}_0 \ge \cdots \ge \bar{f}_n$. To see that $\bar{f}_i - \bar{f}_{i-1} \ge -\frac{1}{n(1-\alpha)}$ observe that $u_i = u_{i-1}$ unless $u_{i-1} = \frac{x_{i-1} + x_i}{2} - \frac{i-1}{n(1-\alpha)}$, but then

$$\begin{aligned} u_i - u_{i-1} &\ge \left( \frac{x_i + x_{i+1}}{2} - \frac{i}{n(1-\alpha)} \right) - \left( \frac{x_{i-1} + x_i}{2} - \frac{i-1}{n(1-\alpha)} \right) \\ &= \frac{x_{i+1} - x_{i-1}}{2} - \frac{1}{n(1-\alpha)} \\ &\ge -\frac{1}{n(1-\alpha)} \end{aligned}$$

and the claim follows. We show now that $\hat{f}_i \le \bar{f}_i$, where $\hat{f}_i = \hat{h}_i - \frac{1}{1-\alpha} \frac{i}{n}$. Since $u_i \ge \frac{x_i + x_{i+1}}{2} - \frac{1}{1-\alpha} \frac{i}{n}$, we see that $\bar{f}_i \ge \frac{x_i + x_{i+1}}{2} - \frac{1}{1-\alpha} \frac{i}{n}$, provided $\frac{x_i + x_{i+1}}{2} - \frac{1}{1-\alpha} \frac{i}{n} \le 0$. Now, if $u_i \ge 0$, then $\bar{f}_i = 0 \ge \hat{f}_i$. Let us assume that $u_i < 0$ (hence $\bar{f}_i \ge \frac{x_i + x_{i+1}}{2} - \frac{1}{1-\alpha} \frac{i}{n}$) and $\bar{f}_j \ge \hat{f}_j$, $j < i$, but $\bar{f}_i < \hat{f}_i$. Let us write $k$ for the smallest integer $k > i$ such that $\bar{f}_k \ge \hat{f}_k$ (observe that $k \le n$ since $\bar{f}_n = \hat{f}_n = -\frac{\alpha}{1-\alpha}$). We define $\tilde{f}_j = \hat{f}_j$ if $j < i$ or $j \ge k$ and $\tilde{f}_j = \bar{f}_j$ if $i \le j < k$. Also, write $\tilde{h}_j = \tilde{f}_j - \frac{1}{1-\alpha} \frac{j}{n}$. Clearly, $\tilde{h} \in \mathcal{C}_{\alpha,n}$. But for integer $j \in [i, k)$ we have $\hat{h}_j > \tilde{h}_j \ge \frac{x_j + x_{j+1}}{2}$, which implies $|\hat{h}_j - \frac{x_j+x_{j+1}}{2}| > |\tilde{h}_j - \frac{x_j+x_{j+1}}{2}|$. Consequently,

$$\sum_{j=1}^{n-1}(x_{j+1}-x_j)^{p+1} f_p\left(\frac{\tilde{h}_j-(x_j+x_{j+1})/2}{(x_{j+1}-x_j)/2}\right)$$
$$< \sum_{j=1}^{n-1}(x_{j+1}-x_j)^{p+1} f_p\left(\frac{\hat{h}_j-(x_j+x_{j+1})/2}{(x_{j+1}-x_j)/2}\right),$$

against optimality of $\hat{h}$. Hence, $\bar{f}_i \geq \hat{f}_i$ for all $i$, and the upper bound for $\hat{h}_i$ follows. □

A similar lower bound for $\hat{h}$ can be obtained taking $\underline{f}_i$ to be the greatest decreasing sequence from 0 to $-\frac{\alpha}{1-\alpha}$ which lies below the sequence $\frac{x_i+x_{i+1}}{2} - \frac{i}{n(1-\alpha)}$ and setting $\underline{h}_i = \underline{f}_i + \frac{i}{n(1-\alpha)}$. We note also that Lemma 3.1, combined with Lemma 3.2, gives the following useful lower and upper bounds for the incomplete transportation cost. To be precise,

$$\begin{aligned} V_n(p) &\leq \mathcal{W}_p^p(\mathcal{R}_\alpha(P_n), P) \\ &\leq V_n(p) + \frac{1}{p+1}\sum_{i=1}^{n-1}\left(\frac{x_{i+1}-x_i}{2}\right)^{p+1} f_p\left(\frac{\bar{h}_i-(x_{i+1}+x_i)/2}{(x_{i+1}-x_i)/2}\right), \end{aligned} \tag{3.2}$$

where $V_n(p) = \frac{1}{p+1}(x_1^{p+1} + (1-x_n)^{p+1}) + \frac{1}{2^p(p+1)}\sum_{i=1}^{n-1}(x_{i+1}-x_i)^{p+1}$. We could replace $\bar{h}_i$ with $\underline{h}_i$ or $\tilde{h}_i = (\bar{h}_i + \underline{h}_i)/2$, but in terms of rates, the upper bound above cannot be improved, as we will see later.

Next, we consider the case of a uniform random sample on the unit interval, namely $X_1, \dots, X_n$ are i.i.d. $U(0,1)$ r.v.'s, $(x_1, \dots, x_n) = (X_{(1)}, \dots, X_{(n)})$ is the order statistic and $P_n$ the empirical distribution on the sample. We will use the well-known fact

$$(X_{(1)}, \dots, X_{(n)}) \stackrel{d}{=} \left(\frac{S_1}{S_{n+1}}, \dots, \frac{S_n}{S_{n+1}}\right), \tag{3.3}$$

where, $S_i = \xi_1 + \cdots + \xi_i$ and $\{\xi_i\}_{i=1}^\infty$ are i.i.d. exponentials random variables with unit mean. The following elementary lemma about the concentration of the $S_i$'s around their means will be used repeatedly in the remainder of this section.

LEMMA 3.3. *If $t > 0$, then*

$$P(S_i - i > t) \leq e^{-t}\left(1+\frac{t}{i}\right)^i,$$

*while for* $0 < t < i$

$$P(i - S_i > t) \leq e^{t}\left(1-\frac{t}{i}\right)^i.$$

PROOF. This is just Chernoff's inequality; see, for example, [14], page 16. $\square$

We are ready now to give the rate of convergence of $\mathcal{W}_p(\mathcal{R}_\alpha(P_n), P)$ in the one-dimensional setup.

THEOREM 3.4. *If $P$ is the uniform distribution on $[0,1]$, $X_1, \dots, X_n$ are i.i.d. random variables with common distribution $P$, $P_n$ is the empirical measure on $X_1, \dots, X_n$, $\alpha \in (0,1)$ and $p \geq 1$, then there exist constants, $C_p(\alpha)$, depending only on $p$ and $\alpha$, $c_p > 0$ depending only on $p$, such that for every $n \geq 1$,*

$$\frac{c_p}{n^p} \leq E\big(\mathcal{W}_p^p(\mathcal{R}_\alpha(P_n), P)\big) \leq \frac{C_p(\alpha)}{n^p}.$$

PROOF. For the lower bound simply observe that $E(n^p \mathcal{W}_p^p(\mathcal{R}_\alpha(P_n), P)) \geq n^p E(V_n(p))$, with $V_n(p)$ as in (3.2). The spacing $X_{(i+1)} - X_{(i)}$ follows a beta distribution with parameters 1 and $n+1$, and from this fact it follows that $n^p E(V_n(p)) = \frac{n^p \Gamma(n+1)\Gamma(p+2)}{(p+1)\Gamma(n+p+2)}(2 + \frac{n-1}{2^p})$. It is easy to check (using Stirling's formula, e.g.) that $n^p E(V_n(p)) \to \frac{\Gamma(p+2)}{2^p(p+1)} > 0$ as $n \to \infty$, and hence we can take $c_p = \min_{n \geq 1} n^p E(V_n(p)) > 0$.

For the upper bound we use the representation (3.3), fix $\theta \in (1-\alpha, 1)$ and write $Z = \mathcal{W}_p^p(\mathcal{R}_\alpha(P_n), P))$. Then we split $E(Z)$,

$$E(Z) = E\bigg(ZI\bigg(\frac{S_{n+1}}{n} < \theta\bigg)\bigg) + E\bigg(ZI\bigg(\frac{S_{n+1}}{n} \geq \theta\bigg)\bigg) := E(Z_1) + E(Z_2)$$

and proceed to bound $E(Z_i)$, $i = 1, 2$. To deal with $E(Z_1)$ we note that $Z \leq \mathcal{W}_p^p(P_n, P) = \int_0^1 |G_n^{-1}(t) - t|^p\,dt$, $G_n$ being the distribution function asociated to $P_n$. A simple computation, similar to the proof of Lemma 3.1, shows $\int_0^1 |G_n^{-1}(t) - t|^p\,dt = \int_0^1 |G_n(t) - t|^p\,dt$, both terms equaling, in fact,

$$\frac{1}{p+1}\sum_{i=1}^n \bigg[\bigg|\frac{i}{n} - X_{(i)}\bigg|^{(p+1)} - \bigg|\frac{i-1}{n} - X_{(i)}\bigg|^{(p+1)}\bigg].$$

Hence, from Schwarz's inequality we get

$$\begin{aligned}
E(Z_1) &\leq \big(E(Z^2)\big)^{1/2} P\bigg(\frac{S_{n+1}}{n} < \theta\bigg)^{1/2} \\
&\leq \bigg(E\bigg(\bigg(\int_0^1 |G_n(t) - t|^p\,dt\bigg)\bigg)^2\bigg)^{1/2} P\bigg(\frac{S_{n+1}}{n} < \theta\bigg)^{1/2} \\
&\leq \bigg(\int_0^1 E|G_n(t) - t|^{2p}\,dt\bigg)^{1/2} P\bigg(\frac{S_{n+1}}{n} < \theta\bigg)^{1/2}.
\end{aligned}$$

Using the fact that $P(|G_n(t)-t|>\varepsilon)\le 2e^{-2n\varepsilon^2}$ (this follows from Hoeffding's inequality applied to Bernoulli random variables), we see that $E(|G_n(t)-t|^{2p})\le p2^{1-p}\Gamma(p)n^{-p}$. Also, from Lemma 3.3 we get $P(\frac{S_{n+1}}{n}<\theta)=P((n+1)-S_{n+1}>n(1-\theta)+1)\le e^{n(1-\theta)+1}(\frac{n\theta}{n+1})^{n+1}$. Combining these two estimates we get

$$E\big(n^p Z_1\big)\le \big(p2^{1-p}\Gamma(p)\theta e\big(\theta e^{1-\theta}\big)^n n^p\big)^{1/2}. \tag{3.4}$$

The last upper bound is a vanishing sequence (hence, bounded) since $\theta e^{1-\theta}<1$. Note that the bound depends on $\alpha$ through the choice of $\theta\in(1-\alpha,1)$.

We consider now $E(Z_2)$ and recall (3.2). We have $E(n^pV_n(p)I(\frac{S_{n+1}}{n}\ge\theta))\le E(n^pV_n(p))\le\sup_{m\ge1}E(m^pV_m(p))<\infty$ since, as noted above, $E(n^pV_n(p))$ is a convergent sequence as $n\to\infty$. Observe now that $f_p(y)\le 2^{p+1}-2$ if $|y|\le 1$, while $f_p(y)\le 2^{p+1}(p+1)|y|^p-2$ if $|y|\ge 1$. Therefore $f_p(y)\le 2^{p+1}(1+(p+1)|y|^p)$, and it suffices to give an upper bound for

$$\begin{aligned}&E\left(n^pI\left(\frac{S_{n+1}}{n}\ge\theta\right)\sum_{i=1}^{n-1}(X_{(i+1)}-X_{(i)})\left|\bar h_i-\frac{X_{(i)}+X_{(i+1)}}{2}\right|^p\right)\\&\qquad\le\frac{1}{\theta^{p+1}}\frac1n\sum_{i=1}^{n-1}E\left(\xi_{i+1}\left|F_i-\left(S_i+\frac{\xi_{i+1}}{2}-\frac{1}{1-\alpha}\frac{S_{n+1}}{n}i\right)\right|^pI\left(\frac{S_{n+1}}{n}\ge\theta\right)\right),\end{aligned}$$

where we are using representation (3.3) and

$$F_i=\left(\left(\max_{i\le j\le n-1}\left(S_j+\frac{\xi_{j+1}}{2}-\frac{1}{1-\alpha}\frac{S_{n+1}}{n}j\right)\right)\vee\left(-\frac{\alpha}{1-\alpha}S_{n+1}\right)\right)\wedge 0.$$

It only remains to find an upper bound for $E(U_n)$ with

$$U_n:=\frac1n\sum_{i=1}^{n-1}\xi_{i+1}\left|F_i-\left(S_i+\frac{\xi_{i+1}}{2}-\frac{1}{1-\alpha}\frac{S_{n+1}}{n}i\right)\right|^pI\left(\frac{S_{n+1}}{n}\ge\theta\right).$$

We split the sum in $U_n$ into three terms, $U_n=U_n^{(1)}+U_n^{(2)}+U_n^{(3)}$, $U_n^{(1)}$ collecting the summands with $F_i=0$, $U_n^{(3)}$ those with $F_i=-\frac{\alpha}{1-\alpha}S_{n+1}$ and $U_n^{(2)}$ the others. We bound first $E(U_n^{(1)})$. We write $K=\frac{\theta}{1-\alpha}$ and note that $K>1$. Now,

$$U_n^{(1)}\le\frac1n\sum_{i=1}^{n-1}\xi_{i+1}\left|S_i+\frac{\xi_{i+1}}{2}-\frac{1}{1-\alpha}\frac{S_{n+1}}{n}i\right|^pI\Big(\max_{i\le j\le n-1}(S_{j+1}-Kj)\ge 0\Big).$$

Convexity implies that $E(\frac{S_i}{i})^s\le E\xi_1^s$ for $s\ge1$. From the Schwarz inequality and the moment inequality $E|X+Y|^p\le 2^{p-1}(E|X|^p+E|Y|^p)$, $p\ge1$, we get that $(E(\xi_{i+1}|S_i+\frac{\xi_{i+1}}{2}-\frac{1}{1-\alpha}\frac{S_{n+1}}{n}i|^p)^2)^{1/2}\le C_1i^p$ for some absolute constant $C_1$ (not depending on $i$ or $n$). On the other hand, using again Lemma 3.3 with $i=j+1$ and $t=Kj-(j+1)$ [which is positive for $j>(1-\alpha)/(\theta-1+\alpha)$] we have $P(S_{j+1}-$

$Kj \geq 0) \leq e^{-(K-1)j+1}(\frac{j(K-1)}{j+1})^{j+1} \leq Keq^j$, where $q = Ke^{-(K-1)} < 1$. Since $P(\max_{i \leq j \leq n-1}(S_{j+1} - Kj) \geq 0) \leq \sum_{j \geq i} P(S_{j+1} - Kj \geq 0)$, we get, for some constant, $C_2$,

$$P\Bigl(\max_{i \leq j \leq n-1}(S_{j+1} - Kj) \geq 0\Bigr) \leq C_2 \frac{q^i}{1-q}$$

$[C_2 = Ke$ suffices for $i \geq (1-\alpha)/(\theta - 1 + \alpha)$; with a larger constant, if necessary, the bound is true for all $i]$. Combining the last bounds and Schwarz's inequality we obtain, with a new constant $C_3$,

$$E\bigl(U_n^{(1)}\bigr) \leq \frac{C_3}{\sqrt{1-q}} \frac{1}{n} \sum_{i=1}^{n-1} i^p q^{i/2}$$

and, again, the right-hand side is a vanishing sequence.

To deal with $U_n^{(3)}$ we define $\xi_i' = \xi_{n+2-i}$, $S_i' = \xi_1' + \cdots + \xi_i'$, $i = 1, \ldots, n+1$. Observe that $S_i' = S_{n+1} - S_{n+1-i}$, $i = 1, \ldots, n$, and $S_{n+1}' = S_{n+1}$. We also write $A_i = (\frac{S_{n+1}}{n} \geq \theta, \max_{i \leq j \leq n-1}(S_j + \frac{\xi_{j+1}}{2} - \frac{1}{1-\alpha}\frac{S_{n+1}}{n} j) \leq -\frac{\alpha}{1-\alpha} S_{n+1})$. Then

$$\begin{aligned} U_n^{(3)} &= \frac{1}{n} \sum_{i=1}^{n-1} \xi_{i+1} \biggl| -\frac{\alpha}{1-\alpha} S_{n+1} - \biggl(S_i + \frac{\xi_{i+1}}{2} - \frac{1}{1-\alpha} \frac{S_{n+1}}{n} i\biggr)\biggr|^p I_{A_i} \\ &= \frac{1}{n} \sum_{j=1}^{n-1} \xi_{n+1-j} \biggl| -\frac{\alpha}{1-\alpha} S_{n+1} \\ &\qquad\qquad - \biggl(S_{n-j} + \frac{\xi_{n+1-j}}{2} - \frac{1}{1-\alpha} \frac{S_{n+1}}{n}(n-j)\biggr)\biggr|^p I_{A_{n-j}}. \end{aligned}$$

With the above notation we see that $-\frac{\alpha}{1-\alpha} S_{n+1} - (S_{n-j} + \frac{\xi_{n+1-j}}{2} - \frac{1}{1-\alpha}\frac{S_{n+1}}{n}(n - j)) = S_j' + \frac{\xi_{j+1}'}{2} - \frac{1}{1-\alpha}\frac{S_{n+1}'}{n} j$, while $A_{n-j} = (\frac{S_{n+1}'}{n} \geq \theta, \min_{1 \leq k \leq j}(S_k' + \frac{\xi_{k+1}'}{2} - \frac{1}{1-\alpha}\frac{S_{n+1}'}{n} k) \geq 0) \subset (\frac{S_{n+1}'}{n} \geq \theta, \max_{j \leq k \leq n-1}(S_k' + \frac{\xi_{k+1}'}{2} - \frac{1}{1-\alpha}\frac{S_{n+1}'}{n} k) \geq 0) := B_j$. These observations imply that

$$\begin{aligned} U_n^{(3)} &= \frac{1}{n} \sum_{j=1}^{n-1} \xi_{j+1}' \biggl| S_j' + \frac{\xi_{j+1}'}{2} - \frac{1}{1-\alpha} \frac{S_{n+1}'}{n} j \biggr|^p I_{A_{n-j}} \\ &\leq \frac{1}{n} \sum_{j=1}^{n-1} \xi_{j+1}' \biggl| S_j' + \frac{\xi_{j+1}'}{2} - \frac{1}{1-\alpha} \frac{S_{n+1}'}{n} j \biggr|^p I_{B_j}. \end{aligned}$$

The last upper bound and $U_n^{(1)}$ are equally distributed. Hence $E(U_n^{(3)}) \leq E(U_n^{(1)}) \to 0$.

We turn now to the central part, $U_n^{(2)}$. Obviously

$$U_n^{(2)} \leq \frac{1}{n}\sum_{i=1}^{n-1} \xi_{i+1} Z_i^p, \tag{3.5}$$

where $Z_i = \sup_{j \geq i}((S_{j+1} - S_i) - K(j-i))_+$. Once more we use Schwarz's inequality to get $E(\xi_{i+1} Z_i^p) \leq (E\xi_{i+1}^2)^{1/2}(EZ_i^{2p})^{1/2} \leq \sqrt{2}(EZ_0^{2p})^{1/2}$. Thus, it only remains to show $EZ_0^{2p} < \infty$. Chernoff's inequality yields $P(S_i - Ki > t) \leq e^{-(t+(K-1)i)}(K + \frac{t}{i})^i$. From this and the fact $\int_0^\infty e^{-t} t^l \, dt = l!$, $l \in \mathbb{N}$, we get, for integer $k \geq 1$,

$$\begin{aligned}
\int_0^\infty t^k P(S_i - Ki > t)\, dt &\leq \int_0^\infty e^{-t} e^{-(K-1)i} \sum_{j=0}^{i} \binom{i}{j} K^j \frac{t^{i-j+k}}{i^{i-j}}\, dt \\
&\leq \frac{e^i i!}{i^i} \sum_{j=0}^{i} \frac{e^{-Ki}(Ki)^j}{j!} (i+k-j)^k \\
&\leq (i+k)^k \frac{e^i i!}{i^i} \sum_{j=0}^{i} \frac{e^{-Ki}(Ki)^j}{j!} \\
&= (i+k)^k \frac{e^i i!}{i^i} P(N_{Ki} \leq i),
\end{aligned} \tag{3.6}$$

where $N_\lambda$ denotes a random variable having Poisson distribution with mean $\lambda$. Chernoff's inequality (for the left tail) gives

$$P(N_{Ki} \leq i) = P\bigl(Ki - N_{Ki} \geq (K-1)i\bigr) \leq \exp\biggl(-iKh\biggl(-\frac{K-1}{K}\biggr)\biggr),$$

where $h(u) = (1+u)\log(1+u) - u$, $u \geq -1$; see, for example, [14], page 19. This, (3.6) and the fact $\frac{e^i i!}{i^i} \leq C\sqrt{i}$ for some constant $C$, imply

$$\begin{aligned}
E(Z_0^{k+1}) &= (k+1) \int_0^\infty t^k P\Bigl(\sup_{i \geq 1}(S_i - Ki)_+ > t\Bigr)\, dt \\
&\leq (k+1) \sum_{i=1}^{\infty} \int_0^\infty t^k P(S_i - Ki > t)\, dt \\
&\leq C' \sum_{i=1}^{\infty} i^{k+1/2} \exp\biggl(-iKh\biggl(-\frac{K-1}{K}\biggr)\biggr) < \infty
\end{aligned}$$

for some constant $C'$ (which depends on $k$), where we have used that $h(-\frac{K-1}{K}) > 0$. This completes the proof. $\square$

Finally, we turn to general dimension. In our last result we combine the upper bound in Theorem 3.4 with a combinatorial approach to give the exact rate of convergence of the empirical cost of optimal incomplete transportation to the uniform distribution on the $d$-dimensional unit cube. The result is not given in terms of expectations as in Theorem 3.4. Our approach allows also to get that type of result, but we refrain from adding more technicalities.

THEOREM 3.5. *If $P$ is the uniform distribution on the unit cube $[0,1]^d$, $X_1,\dots,X_n$ are i.i.d. $P$, $P_n$ is the empirical measure on $X_1,\dots,X_n$ and $\alpha\in(0,1)$, then*

$$\mathcal{W}_p\bigl(\mathcal{R}_\alpha(P_n),P\bigr)=O_P\bigl(n^{-1/d}\bigr).$$

PROOF. For the sake of simplicity we consider the case $d=2$. The idea carries over smoothly to higher dimension. On the other hand, the case $d\ge 3$ follows from known results for the usual transport; recall (2.4). We write $N=[\sqrt{n}]$ and $X_j=[X_{j,1},X_{j,2}]^T$, $j=1,\dots,n$. We denote also $B_i=\sharp\{j\in\{1,\dots,n\}:X_{j,1}\in(\frac{i-1}{N},\frac{i}{N}]\}$, $i=1,\dots,N$. The random vector $(B_1,\dots,B_N)$ follows a multinomial distribution with parameters $n$ and $(\frac{1}{N},\dots,\frac{1}{N})$. Given $B_i=n_i>0$, we denote by $j_1^i,\dots,j_{n_i}^i$ the indices $k$ such that $X_{k,1}\in(\frac{i-1}{N},\frac{i}{N}]$. Then $X_{j_1^i,2},\dots,X_{j_{n_i}^i,2}$ are an i.i.d. $U(0,1)$ sample. We write $P(\frac{\alpha}{2},i)$ for the $\frac{\alpha}{2}$-trimming of the empirical distribution on $X_{j_1^i,2},\dots,X_{j_{n_i}^i,2}$ considered in the proof of Theorem 3.4. Then we have $E(n_i^p\mathcal{W}_p^p(P(\frac{\alpha}{2},i),U(0,1))|B_i=n_i)\le C_p(\alpha/2)$. We write also $\varphi_i$ for the optimal transportation map from $U(0,1)$ to $P(\frac{\alpha}{2},i)$. Then $\mathcal{W}_p^p(P(\frac{\alpha}{2},i),U(0,1))=\int_0^1|x_2-\varphi_i(x_2)|^p\,dx_2$. We recall that $\varphi$ takes values on the set $\{X_{j_1},\dots,X_{j_{n_i}}\}$ and, with $\ell_d$ denoting $d$-dimensional Lebesgue measure, $\ell_1(x:\varphi_i(x)=X_{j_l,2})\le\frac{1}{n_i(1-\alpha/2)}$.

Next we define the map $\varphi$ on $(0,1]\times(0,1]$ as follows. If $x=[x_1,x_2]^T$ is such that $x_1\in(\frac{i-1}{N},\frac{i}{N}]$ and $\varphi_i(x_2)=X_{j_l,2}$, then $\varphi(x)=X_{j_l}$. In other words, points on the stripe $(\frac{i-1}{N},\frac{i}{N}]\times(0,1]$ are mapped to one of the observations on that stripe, the precise one being determined by the $\alpha/2$ trimming function on the second coordinate. Clearly, for $x\in(\frac{i-1}{N},\frac{i}{N}]$,

$$\|x-\varphi(x)\|\le\frac{1}{N}+|x_2-\varphi_i(x_2)|.$$

From this we get

$$\begin{aligned}\int_{(0,1]\times(0,1]}\|x-\varphi(x)\|^p\,dx&=\sum_{i=1}^N\int_{((i-1)/N,i/N]\times(0,1]}\|x-\varphi(x)\|^p\,dx\\&\le\frac{2^{p-1}}{N^p}+\frac{2^{p-1}}{N}\sum_{i=1}^N\int_0^1|x_2-\varphi_i(x_2)|^p\,dx_2.\end{aligned}$$

Furthermore, $\ell_2(x:\varphi(x)=X_j)\le \frac{1}{N}\frac{1}{n_i(1-\frac{\alpha}{2})}$ if $X_j\in(\frac{i-1}{N},\frac{i}{N}]\times(0,1]$. Thus, $\varphi$ maps $P$ into an $\alpha$-trimming of $P_n$ if

$$\frac{1}{N}\frac{1}{n_i(1-\alpha/2)}\le\frac{1}{n(1-\alpha)},\qquad i=1,\dots,N,$$

or, equivalently, if $\min_{1\le i\le N} n_i\ge \frac{n}{N}A$, with $A=\frac{1-\alpha}{1-\alpha/2}<1$. As a consequence, on the set $B=(\min_{1\le i\le N} n_i\ge\frac{n}{N}A)$,

$$\mathcal{W}_p^p(\mathcal{R}_\alpha(P_n),P)\le\frac{2^{p-1}}{N^p}+\frac{2^{p-1}}{N}\sum_{i=1}^N\int_0^1|x_2-\varphi_i(x_2)|^p\,dx_2.$$

Now, we note that

$$E\left(I_B\sum_{i=1}^N\int_0^1|x_2-\varphi_i(x_2)|^p\,dx_2\Big|B_1=n_1,\dots,B_N=n_N\right)$$
$$\le C_p(\alpha/2)I_B\sum_{i=1}^N\frac{1}{n_i^p}\le C(\alpha/2)\frac{1}{A^p}\sum_{i=1}^N\frac{N^p}{n^p}.$$

But the last two displays imply that

$$E(\mathcal{W}_p^p(\mathcal{R}_\alpha(P_n),P)I_B)\le\frac{2^{p-1}}{N^p}+C_p(\alpha/2)\frac{2^{p-1}}{A^p}\frac{N^p}{n^p}.$$

This, toghether with the fact that $P(B^C)\to 0$ as $n\to\infty$ (see Theorem 7, page 112, in [13]) implies that $\mathcal{W}_p^p(\mathcal{R}_\alpha(P_n),P)=O_P(n^{-p/2})$ and completes the proof. $\square$

## REFERENCES


[1] AJTAI, M., KOMLÓS, J. and TUSNÁDY, G. (1984). On optimal matchings. *Combinatorica* **4** 259–264. MR0779885

[2] ÁLVAREZ-ESTEBAN, P. C., DEL BARRIO, E., CUESTA-ALBERTOS, J. A. and MATRÁN, C. (2011). Uniqueness and approximate computation of optimal incomplete transportation plans. *Ann. Inst. Henri Poincaré Probab. Stat.* **47** 358–375. MR2814414

[3] ÁLVAREZ-ESTEBAN, P. C., DEL BARRIO, E., CUESTA-ALBERTOS, J. A. and MATRÁN, C. (2012). Similarity of samples and trimming. *Bernoulli* **18** 606–634. MR2922463

[4] BARTHE, F. and BORDENAVE, C. (2011). Combinatorial optimization over two random point sets. Available at arXiv:1103.2734v1.

[5] BICKEL, P. J. and FREEDMAN, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217. MR0630103

[6] BOUCHERON, S., LUGOSI, G. and MASSART, P. (2003). Concentration inequalities using the entropy method. *Ann. Probab.* **31** 1583–1614. MR1989444

[7] CAFFARELLI, L. A. and MCCANN, R. J. (2010). Free boundaries in optimal transport and Monge–Ampère obstacle problems. *Ann. of Math.* (2) **171** 673–730. MR2630054

[8] CSÖRGŐ, M. and HORVÁTH, L. (1993). *Weighted Approximations in Probability and Statistics*. Wiley, Chichester. MR1215046

[9] DEL BARRIO, E. and MATRÁN, C. (2013). Rates of convergence for partial mass problems. *Probab. Theory Related Fields* **155** 521–542. MR3034786
[10] DOBRIĆ, V. and YUKICH, J. E. (1995). Asymptotics for transportation cost in high dimensions. *J. Theoret. Probab.* **8** 97–118. MR1308672
[11] FIGALLI, A. (2010). The optimal partial transport problem. *Arch. Ration. Mech. Anal.* **195** 533–560. MR2592287
[12] GRAF, S. and LUSCHGY, H. (2000). *Foundations of Quantization for Probability Distributions. Lecture Notes in Math.* **1730**. Springer, Berlin. MR1764176
[13] KOLCHIN, V. F., SEVAST'YANOV, B. A. and CHISTYAKOV, V. P. (1978). *Random Allocations*. V. H. Winston & Sons, Washington, DC. MR0471016
[14] MASSART, P. (2007). *Concentration Inequalities and Model Selection. Lecture Notes in Math.* **1896**. Springer, Berlin. MR2319879
[15] PAPADIMITRIOU, C. H. and STEIGLITZ, K. (1998). *Combinatorial Optimization: Algorithms and Complexity*. Dover, Mineola, NY. MR1637890
[16] RACHEV, S. T. and RÜSCHENDORF, L. (1998). *Mass Transportation Problems*. (2 *Vols*). Springer, New York.
[17] TALAGRAND, M. (1992). Matching random samples in many dimensions. *Ann. Appl. Probab.* **2** 846–856. MR1189420
[18] TALAGRAND, M. (1994). The transportation cost from the uniform measure to the empirical measure in dimension $\geq 3$. *Ann. Probab.* **22** 919–959. MR1288137
[19] TALAGRAND, M. and YUKICH, J. E. (1993). The integrability of the square exponential transportation cost. *Ann. Appl. Probab.* **3** 1100–1111. MR1241036
[20] VILLANI, C. (2003). *Topics in Optimal Transportation. Graduate Studies in Mathematics* **58**. Amer. Math. Soc., Providence, RI. MR1964483
[21] VILLANI, C. (2009). *Optimal Transport: Old and New. Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **338**. Springer, Berlin. MR2459454
[22] YUKICH, J. E. (2008). Limit theorems for multi-dimensional random quantizers. *Electron. Commun. Probab.* **13** 507–517. MR2447837

FACULTAD DE CIENCIAS
UNIVERSIDAD DE VALLADOLID
PASEO DE BELÉN 9
47011 VALLADOLID
SPAIN
E-MAIL: tasio@eio.uva.es
matran@eio.uva.es