Submitted to Bernoulli

# Supplement to “Optimal 1-Wasserstein distance for WGANs”

ARTHUR STÉPHANOVITCH[1,a], UGO TANIELIAN[2,b], BENOÎT CADRE[3,c], NICOLAS KLUTCHNIKOFF[3,d] and GÉRARD BIAU[4,e]

[1]Université Paris Cité, CNRS, LPSM, F-75013 Paris, France, [a]stephanovitch@lpsm.paris
[2]Criteo AI Lab, Paris, France, [b]ugo.tanielian@gmail.com
[3]Univ Rennes, CNRS, IRMAR - UMR 6625, F-35000 Rennes, France, [c]benoit.cadre@univ-rennes2.fr, [d]nicolas.klutchnikoff@univ-rennes2.fr
[4]Sorbonne Université, CNRS, LPSM, F-75005 Paris, France, [e]gerard.biau@sorbonne-universite.fr

## 1. Proof of Lemma 2.1

We only focus on the first statement since the proof of the second one is similar. Let $G, G' \in \mathrm{Lip}_K([0,1], \mathbb{R}^d)$. Observe that by the triangle inequality and the primal definition of the 1-Wasserstein distance, we have

$$\begin{aligned}|W_1(G_{\sharp U}, \mu_n) - W_1(G'_{\sharp U}, \mu_n)| &\leqslant W_1(G_{\sharp U}, G'_{\sharp U})\\ &\leqslant \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x-y\| \mathrm{d}\gamma(x,y),\end{aligned}$$

where $\gamma$ is the pushforward distribution of $U$ by the pair $(G, G')$, with marginals $G_{\sharp U}$ and $G'_{\sharp U}$. Thus,

$$\begin{aligned}|W_1(G_{\sharp U}, \mu_n) - W_1(G'_{\sharp U}, \mu_n)| &\leqslant \int_{[0,1]} \|G(u) - G'(u)\| \mathrm{d}u\\ &\leqslant \|G - G'\|_\infty,\end{aligned}$$

where $\|\cdot\|_\infty$ denotes the supremum norm of functions, i.e., for $f : [0,1] \to \mathbb{R}^d$, $\|f\|_\infty = \sup\{\|f(x)\| : x \in [0,1]\}$. Hence the map $\mathrm{Lip}_K([0,1], \mathbb{R}^d) \ni G \mapsto W_1(G_{\sharp U}, \mu_n)$ is continuous with respect to the uniform norm.

Now let $G^0 \equiv X_1$ be a constant function on $[0,1]$. Then, clearly, $W_1(G^0_{\sharp U}, \mu_n) < \infty$. Next, let $G$ be any function in $\mathrm{Lip}_K([0,1], \mathbb{R}^d)$ such that

$$\|G\|_\infty \geqslant W_1(G^0_{\sharp U}, \mu_n) + K + \max_{i=1,\dots,n} \|X_i\|.$$

Then, upon observing that there exists $u_0 \in [0,1]$ such that $\|G(u_0)\| = \|G\|_\infty$ and using the fact that $G$ is $K$-Lipschitz continuous on $[0,1]$, we deduce that for all $u \in [0,1]$ and any $i \in \{1, \dots, n\}$, one has

$$\|G(u) - X_i\| \geqslant \|G\|_\infty - K - \|X_i\| \geqslant \|G\|_\infty - K - \max_{i=1,\dots,n} \|X_i\|.$$

Hence, $\|G(u) - X_i\| \geqslant W_1(G^0_{\sharp U}, \mu_n)$, which implies that $W_1(G_{\sharp U}, \mu_n) \geqslant W_1(G^0_{\sharp U}, \mu_n)$. Therefore, letting

$$\mathscr{H}_K = \{G \in \mathrm{Lip}_K([0,1], \mathbb{R}^d) \ : \ \|G\|_\infty \leqslant W_1(G^0_{\sharp U}, \mu_n) + K + \max_{i=1,\dots,n} \|X_i\|\},$$

we see that

$$\inf_{G\in \mathrm{Lip}_K([0,1],\mathbb{R}^d)} W_1(G_{\sharp U},\mu_n) = \inf_{G\in \mathscr{H}_K} W_1(G_{\sharp U},\mu_n).$$

Endowed with the uniform norm, $\mathscr{H}_K$ is closed and relatively compact by the Arzelà-Ascoli theorem. It is thus a compact subset of $\mathrm{Lip}_K([0,1],\mathbb{R}^d)$. Consequently, by continuity and the above equality, $\mathrm{Lip}_K([0,1],\mathbb{R}^d) \ni G \mapsto W_1(G_{\sharp U},\mu_n)$ attains its minimum on $\mathscr{H}_K$. Therefore, $\widehat{\mathscr{G}}_K$ is not empty.

# 2. Proof of Theorem 2.2

## Proof of $1(i)$

Since $\mu$ is of order 1, one has $\lim_{n\to\infty} W_1(\mu,\mu_n) = 0$ a.s. according to Villani (2008, Theorem 6.8). Hence, by the triangle inequality and because $\widehat{G}_K \in \widehat{\mathscr{G}}_K$, we only need to prove that

$$\lim_{n\to\infty} \inf_{G\in \mathrm{Lip}_K([0,1],\mathbb{R})} W_1(G_{\sharp U},\mu_n) = 0 \text{ a.s.}$$

If $K \geqslant K_0$, then $\mathrm{Lip}_{K_0}([0,1],\mathbb{R}) \subseteq \mathrm{Lip}_K([0,1],\mathbb{R})$. Therefore,

$$0 \leqslant \inf_{G\in \mathrm{Lip}_K([0,1],\mathbb{R})} W_1(G_{\sharp U},\mu_n) \leqslant \inf_{G\in \mathrm{Lip}_{K_0}([0,1],\mathbb{R})} W_1(G_{\sharp U},\mu_n) \leqslant W_1(F^{-1}_{\sharp U},\mu_n),$$

since, by assumption, $F^{-1} \in \mathrm{Lip}_{K_0}([0,1],\mathbb{R})$. But $F^{-1}(U)$ has distribution $\mu$, and thus one has $\lim_{n\to\infty} W_1(F^{-1}_{\sharp U},\mu_n) = 0$. This proves the result.

## Proof of (2)

The result is proved by contradiction. Fix $K > 0$ and assume that on an event of strictly positive probability

$$\liminf_{n\to\infty} W_1(\widehat{G}_{K\sharp U},\mu) = 0.$$

Since $\lim_{n\to\infty} W_1(\mu,\mu_n) = 0$ a.s. and $\widehat{G}_K \in \widehat{\mathscr{G}}_K$, we see that

$$\inf_{G\in \mathrm{Lip}_K([0,1],\mathbb{R})} W_1(G_{\sharp U},\mu) = 0.$$

Now, by Lemma 2.1, there exists $G_K \in \mathrm{Lip}_K([0,1],\mathbb{R})$ such that

$$W_1(G_{K\sharp U},\mu) = \inf_{G\in \mathrm{Lip}_K([0,1],\mathbb{R})} W_1(G_{\sharp U},\mu).$$

So, $W_1(G_{K\sharp U},\mu) = 0$ and therefore, since $F^{-1}(U)$ has distribution $\mu$, we have

$$G_K(U) \overset{\mathscr{L}}{\sim} F^{-1}(U). \tag{1}$$

Next, by continuity of $G_K$, there exists a compact set $C \subseteq \mathbb{R}$ such that $\mathbb{P}(G_K(U) \in C) = 1$. But, since $S(\mu)$ is unbounded, $\mathbb{P}(F^{-1}(U) \in C) = \mu(C) < 1$, which contradicts (1).

### Proof of $1(ii)$

We show the result by contradiction, assuming as in the proof of statement (2) that for $K < 1/K_1$, on an event of strictly positive probability,

$$\liminf_{n\to\infty} W_1(\widehat{G}_{K\sharp U}, \mu) = 0.$$

Arguing as in the previous proof, we have that $G_K(U) \overset{\mathscr{L}}{\sim} F^{-1}(U)$. Then, it is a classical exercise to deduce from (1), since $F^{-1}(u) > -\infty$ for all $u \in (0,1)$ and $F$ is continuous, that $F \circ G_K(U) \overset{\mathscr{L}}{\sim} U$. Iterating this relation leads to

$$(F \circ G_K)^{\ell}(U) \overset{\mathscr{L}}{\sim} U, \quad \forall \ell \geqslant 0. \tag{2}$$

Moreover, both assumptions $F \in \mathrm{Lip}_{K_1}(\mathbb{R}, [0,1])$ and $G_K \in \mathrm{Lip}_K([0,1], \mathbb{R})$ imply

$$|F \circ G_K(u) - F \circ G_K(v)| \leqslant KK_1|u - v| \leqslant KK_1, \quad \forall (u,v) \in [0,1]^2.$$

Repeating this inequality entails, for all $\ell \geqslant 0$,

$$|(F \circ G_K)^{\ell}(u) - (F \circ G_K)^{\ell}(v)| \leqslant (KK_1)^{\ell}, \quad \forall (u,v) \in [0,1]^2.$$

But, for all $u \in [0,1]$, the sequence $((F \circ G_K)^{\ell}(u))_{\ell \geqslant 1}$ is bounded by 1. In addition, $KK_1 < 1$ by assumption. Thus, there exist $a \in [0,1]$ and a subsequence $(\ell_q)_{q\geqslant 1}$ such that, for all $u \in [0,1]$,

$$\lim_{q\to\infty} (F \circ G_K)^{\ell_q}(u) = a.$$

Hence, as $q \to \infty$, $(F \circ G_K)^{\ell_q}(U)$ almost surely converges to $a$, which contradicts (2).

## 3. Proof of Theorem 2.3

Looking for a contradiction, we start as in the proof of Theorem 2.2, cases $(1ii)$ and (2), by assuming that on an event of strictly positive probability,

$$\liminf_{n\to\infty} W_1(\widehat{G}_{K\sharp U}, \mu) = 0.$$

As we have seen, this implies $W_1(G_{K\sharp U}, \mu) = 0$ and, in turn, since the support of $G_{K\sharp U}$ is included in $G_K([0,1])$, $S(\mu) \subseteq G_K([0,1])$. By our assumption on $S(\mu)$, we therefore conclude that $\lambda_d(G_K([0,1])) > 0$. Moreover, since $G_K \in \mathrm{Lip}_K([0,1], \mathbb{R}^d)$, we have that $0 < \lambda_d(G_K([0,1])) = \mathcal{H}_d(G_K([0,1])) \leqslant K^d \mathcal{H}_d([0,1])$, where $\mathcal{H}_d$ is the $d$-dimensional Hausdorff measure (see, e.g., Evans and Gariepy, 2015, Theorem 2.8). But this is impossible since $\mathcal{H}_d([0,1]) = 0$ as soon as $d > 1$.

## 4. Proof of Proposition 3.1

To lighten the notation, it is assumed throughout the proof that the $X_i$'s are ordered by increasing values, i.e., $X_1 \leqslant X_2 \leqslant \cdots \leqslant X_n$. According to Santambrogio (2015, Proposition 2.17), the 1-Wasserstein distance between two probability measures $\pi_1$ and $\pi_2$ on the real line, with respective generalized inverses $F_1^{-1}$ and $F_2^{-1}$, is such that

$$W_1(\pi_1, \pi_2) = \int_0^1 |F_1^{-1}(u) - F_2^{-1}(u)| \mathrm{d}u.$$

Since $\widehat{G}_K^\star$ is monotone and continuous, the generalized inverse of $\widehat{G}^\star_{K\sharp U}$ is $\widehat{G}_K^\star$. On the other hand, denoting by $F_{\mu_n}^{-1}$ the generalized inverse of $\mu_n$, we have $F_{\mu_n}^{-1}(u) = \sum_{i=1}^n X_i \mathbb{1}\{u \in ((i-1)/n, i/n]\}$. Therefore,

$$\begin{aligned} W_1(\widehat{G}^\star_{K\sharp U}, \mu_n) &= \int_0^1 |\widehat{G}_K^\star(u) - F_{\mu_n}^{-1}(u)| \mathrm{d}u \\ &= \sum_{i=1}^{n-1} \int_{i/n - \frac{X_{i+1}-X_i}{2K}}^{i/n} \Big| X_i + K\big(u - (\frac{i}{n} - \frac{X_{i+1}-X_i}{2K})\big) - X_i \Big| \mathrm{d}u \\ &\quad + \sum_{i=1}^{n-1} \int_{i/n}^{i/n + \frac{X_{i+1}-X_i}{2K}} \Big| \frac{X_{i+1}-X_i}{2K} + K(u - \frac{i}{n}) - X_{i+1} \Big| \mathrm{d}u \\ &= \sum_{i=1}^{n-1} \frac{1}{2} K \big( \frac{(X_{i+1}-X_i)^2}{4K^2} + \frac{(X_{i+1}-X_i)^2}{4K^2} \big) \\ &= \frac{1}{4K} \sum_{i=1}^{n-1} (X_{i+1} - X_i)^2, \end{aligned}$$

as desired.

## 5. Proof of Theorem 3.2

As in the proof of Proposition 3.1, it is assumed without loss of generality that the $X_i$'s are ordered by increasing values, i.e., $X_1 \leqslant X_2 \leqslant \cdots \leqslant X_n$. Let $G : [0, 1] \to \mathbb{R}$ be an arbitrary $K$-Lipschitz continuous function in $\widehat{\mathscr{G}}_K$, with $K \geqslant n \max_{i=1,\ldots,n-1} (X_{i+1} - X_i)$. According to Proposition 3.1, the first statement will be proven if we show that for such a function $G$,

$$W_1(G_{\sharp U}, \mu_n) \geqslant \sum_{i=1}^{n-1} \frac{(X_{i+1}-X_i)^2}{4K}.$$

Let $\Pi(\pi_1, \pi_2)$ be the set of couplings between two probability measures $\pi_1$ and $\pi_2$. According to Ambrosio and Gigli (2013, Lemma 2.12), for any $\pi \in \Pi(G_{\sharp U}, \mu_n)$, there exists a coupling $\gamma \in \Pi(\lambda_1, \mu_n)$ such that $\pi = (G, \mathrm{Id})_{\#\gamma}$, where $\lambda_1$ stands for the Lebesgue measure on the interval $[0, 1]$ and Id is the identity function. Therefore,

$$\begin{aligned} W_1(G_{\sharp U}, \mu_n) &= \inf_{\pi \in \Pi(G_{\sharp U}, \mu_n)} \int_{\mathbb{R}\times\mathbb{R}} |x - y| \mathrm{d}\pi(x, y) \\ &\geq \inf_{\gamma \in \Pi(\lambda_1, \mu_n)} \int_{[0,1]\times\mathbb{R}} |G(u) - y| \mathrm{d}\gamma(u, y). \end{aligned}$$

Since the function $(u,y) \mapsto |G(u)-y|$ is continuous, then, according to Pratelli (2007, Theorem B), we have

$$\inf_{\gamma\in\Pi(\lambda_1,\mu_n)} \int_{[0,1]\times\mathbb{R}} |G(u)-y|\mathrm{d}\gamma(u,y) = \inf_T \int_0^1 |G(u)-T(u)|\mathrm{d}u,$$

where the infimum is taken over all measurable functions $T:[0,1]\to\{X_1,\dots,X_n\}$ such that $T_{\sharp U}=\mu_n$. Any such transport map $T$ takes the form $T(u)=\sum_{i=1}^n X_i\mathbb{1}\{u\in C_i\}$, where $C_1,\dots,C_n$ are Borel subsets of $[0,1]$ such that $\lambda_1(C_i)=\frac{1}{n}$. We conclude that

$$W_1(G_{\sharp U},\mu_n) \geqslant \inf_{C_1,\dots,C_n} \sum_{i=1}^n \int_{C_i} |G(u)-X_i|\mathrm{d}u, \tag{3}$$

where the infimum is taken over all disjoint Borel sets $C_1,\dots,C_n\subseteq[0,1]$ such that $\lambda_1(C_i)=\frac{1}{n}$. To prove the first statement of the theorem, it is therefore sufficient to lower bound the infimum above.

The case $n=1$ is clear since the function $G(u)\equiv X_1$ satisfies $W_1(G_{\sharp U},\mu_1)=0$. Thus, in the sequel, it is assumed that $n\geqslant 2$. We let $a=\inf_{[0,1]} G$, $b=\sup_{[0,1]} G$, and $\ell_1\leqslant\ell_2$ so that $X_{\ell_1}=\min_{X_i\geqslant a} X_i$ and $X_{\ell_2}=\max_{X_i\leqslant b} X_i$. Note that we can safely assume that $\ell_1$ and $\ell_2$ are well-defined, since for $\hat{G}(u):=G(u)\mathbb{1}\{G(u)\in[X_1,X_n]\}+X_1\mathbb{1}\{G(u)<X_1\}+X_n\mathbb{1}\{G(u)>X_n\}$, we have

$$\inf_{C_1,\dots,C_n} \sum_{i=1}^n \int_{C_i} |G(u)-X_i|\mathrm{d}u \geqslant \inf_{C_1,\dots,C_n} \sum_{i=1}^n \int_{C_i} |\hat{G}(u)-X_i|\mathrm{d}u.$$

We also suppose that $n>\ell_2\geqslant\ell_1+1>1$ and leave the other cases as straightforward adaptations. Since $G$ is continuous, for each $i\in\{\ell_1,\dots,\ell_2-1\}$, there exists $u_i\in[0,1]$ such that $G(u_i)=\frac{X_i+X_{i+1}}{2}$. We let $A_i^-=[u_i-\frac{X_{i+1}-X_i}{2K},u_i]$, $A_i^+=[u_i,u_i+\frac{X_{i+1}-X_i}{2K}]$, and write $T(u)=\sum_{j=1}^n X_j\mathbb{1}\{u\in C_j\}$. With this notation,

$$\begin{aligned}\int_{A_i^-} |G(u)-T(u)|\mathrm{d}u &= \sum_{j=1}^i \int_{A_i^-} (G(u)-X_i+X_i-X_j)\mathbb{1}\{u\in C_j\}\mathrm{d}u \\ &\quad+ \sum_{j=i+1}^n \int_{A_i^-} (X_{i+1}-G(u)+X_j-X_{i+1})\mathbb{1}\{u\in C_j\}\mathrm{d}u \\ &= \sum_{j=1}^i \Big[\int_{A_i^-} (G(u)-X_i)\mathbb{1}\{u\in C_j\}\mathrm{d}u+\lambda_1(C_j\cap A_i^-)(X_i-X_j)\Big] \\ &\quad+ \sum_{j=i+1}^n \Big[\int_{A_i^-} (X_{i+1}-G(u))\mathbb{1}\{u\in C_j\}\mathrm{d}u+\lambda_1(C_j\cap A_i^-)(X_j-X_{i+1})\Big].\end{aligned} \tag{4}$$

Exploiting the fact that the function $G$ is $K$-Lipschitz continuous and $G(u_i)=\frac{X_i+X_{i+1}}{2}$, we have that for $u\in A_i^-\cup A_i^+$, $\frac{X_i+X_{i+1}}{2}-K|u_i-u|\leqslant G(u)\leqslant\frac{X_i+X_{i+1}}{2}+K|u_i-u|$. Thus,

$$\sum_{j=1}^i \int_{A_i^-} (G(u)-X_i)\mathbb{1}\{u\in C_j\}\mathrm{d}u + \sum_{j=i+1}^n \int_{A_i^-} (X_{i+1}-G(u))\mathbb{1}\{u\in C_j\}\mathrm{d}u$$

$$
\begin{aligned}
&\geqslant \sum_{j=1}^{i} \int_{A_i^-} \Big( \frac{X_i + X_{i+1}}{2} - K(u_i - u) - X_i \Big) \mathbb{1}\{u \in C_j\} \mathrm{d}u \\
&\quad + \sum_{j=i+1}^{n} \int_{A_i^-} \Big( X_{i+1} - \Big( \frac{X_i + X_{i+1}}{2} + K(u_i - u) \Big) \Big) \mathbb{1}\{u \in C_j\} \mathrm{d}u \\
&= \sum_{j=1}^{n} \int_{A_i^-} \Big( \frac{X_{i+1} - X_i}{2} - K(u_i - u) \Big) \mathbb{1}\{u \in C_j\} \mathrm{d}u \\
&= \int_{A_i^-} \Big( \frac{X_{i+1} - X_i}{2} - K(u_i - u) \Big) \mathrm{d}u \\
&= \frac{(X_{i+1} - X_i)^2}{4K} - \frac{1}{2} \frac{(X_{i+1} - X_i)^2}{4K} \\
&= \frac{(X_{i+1} - X_i)^2}{8K}. \qquad (5)
\end{aligned}
$$

Combining this inequality with (4) yields

$$
\begin{aligned}
\int_{A_i^-} |G(u) - T(u)| \mathrm{d}u \geqslant \frac{(X_{i+1} - X_i)^2}{8K} \\
+ \sum_{j=1}^{i-1} \lambda_1 (C_j \cap A_i^-)(X_i - X_j) + \sum_{j=i+1}^{n} \lambda_1 (C_j \cap A_i^-)(X_j - X_{i+1}).
\end{aligned}
$$

Employing the same technique for $A_i^+$, we obtain

$$
\begin{aligned}
\int_{A_i^+} |G(u) - T(u)| \mathrm{d}u \geqslant \frac{(X_{i+1} - X_i)^2}{8K} \\
+ \sum_{j=1}^{i-1} \lambda_1 (C_j \cap A_i^+)(X_i - X_j) + \sum_{j=i+1}^{n} \lambda_1 (C_j \cap A_i^+)(X_j - X_{i+1}).
\end{aligned}
$$

So, letting $A_i = A_i^- \cup A_i^+$ and using the fact that $X_{\ell+1} \geqslant X_\ell$ for all $\ell \leqslant n-1$, we are led to

$$
\begin{aligned}
\int_{A_i} |G(u) - T(u)| \mathrm{d}u \geqslant \frac{(X_{i+1} - X_i)^2}{4K} \\
+ \sum_{j=1}^{i-1} \lambda_1 (C_j \cap A_i)(X_{j+1} - X_j) + \sum_{j=i+2}^{n} \lambda_1 (C_j \cap A_i)(X_j - X_{j-1}). \qquad (6)
\end{aligned}
$$

Now, let $u_{\ell_1 - 1} \in [0,1]$ be such that $G(u_{\ell_1 - 1}) = \frac{a + X_{\ell_1}}{2}$. With a slight abuse of notation, define $A_{\ell_1 - 1}^- = [u_{\ell_1 - 1} - \frac{X_{\ell_1} - a}{2K}, u_{\ell_1 - 1}]$ and $A_{\ell_1 - 1}^+ = [u_{\ell_1 - 1}, u_{\ell_1 - 1} + \frac{X_{\ell_1} - a}{2K}]$. Then, using the same method as above, one easily shows that, for $A_{\ell_1-1} = A^-_{\ell_1-1} \cup A^+_{\ell_1-1}$,

$$\int_{A_{\ell_1-1}} |G(u) - T(u)| \mathrm{d}u \geqslant \frac{(X_{\ell_1} - a)^2}{4K} + \sum_{j=1}^{\ell_1-1} \lambda_1(C_j \cap A_{\ell_1-1})(a - X_j) + \sum_{j=\ell_1+1}^{n} \lambda_1(C_j \cap A_{\ell_1-1})(X_j - X_{\ell_1}).$$

In a similar fashion, for $u_{\ell_2} \in [0,1]$ such that $G(u_{\ell_2}) = \frac{X_{\ell_2}+b}{2}$ and, with a slight abuse of notation, letting $A_{\ell_2} = [u_{\ell_2} - \frac{b-X_{\ell_2+1}}{2K}, u_{\ell_2} + \frac{b-X_{\ell_2+1}}{2K}]$, we obtain

$$\int_{A_{\ell_2}} |G(u) - T(u)| \mathrm{d}u \geqslant \frac{(b - X_{\ell_2})^2}{4K} + \sum_{j=1}^{\ell_2-1} \lambda_1(C_j \cap A_{\ell_2})(X_{\ell_2} - X_j) + \sum_{j=\ell_2+1}^{n} \lambda_1(C_j \cap A_{\ell_2})(X_j - b).$$

Accordingly,

$$\begin{aligned}\int_{A_{\ell_1-1} \cup A_{\ell_2}} |G(u) - T(u)| \mathrm{d}u \geqslant\ & \frac{(X_{\ell_1} - a)^2}{4K} + \frac{(b - X_{\ell_2})^2}{4K} \\ & + \sum_{j=1}^{\ell_1-2} \lambda_1(C_j \cap A_{\ell_1-1})(X_{j+1} - X_j) \\ & + \lambda_1(C_{\ell_1-1} \cap A_{\ell_1-1})(a - X_{\ell_1-1}) \\ & + \sum_{j=\ell_1+1}^{n} \lambda_1(C_j \cap A_{\ell_1-1})(X_j - X_{j-1}) \\ & + \sum_{j=1}^{\ell_2-1} \lambda_1(C_j \cap A_{\ell_2})(X_{j+1} - X_j) \\ & + \lambda_1(C_{\ell_2+1} \cap A_{\ell_2})(X_{\ell_2+1} - b) \\ & + \sum_{j=\ell_2+2}^{n} \lambda_1(C_j \cap A_{\ell_2})(X_j - X_{j-1}). \end{aligned} \tag{7}$$

Let $B = \bigcup_{i=\ell_1-1}^{\ell_2} A_i$, and observe that the target integral can be decomposed in the following way:

$$\int_0^1 |G(u) - T(u)| \mathrm{d}u = \int_B |G(u) - T(u)| \mathrm{d}u + \int_{B^c} |G(u) - T(u)| \mathrm{d}u. \tag{8}$$

Inequalities (6) and (7) provide a lower bound on the first term on the right-hand side of (8). Let us now work out the second term. To this aim, observe that

$$\begin{aligned}
\int_{B^c} |G(u) - T(u)|\mathrm{d}u \geqslant& \sum_{j=1}^{\ell_1-1} \int_{B^c} |G(u) - X_j| \mathbb{1}\{u \in C_j\}\mathrm{d}u \\
&+ \sum_{j=\ell_2+1}^{n} \int_{B^c} |G(u) - X_j| \mathbb{1}\{u \in C_j\}\mathrm{d}u \\
\geqslant& \sum_{j=1}^{\ell_1-2} \int_{B^c} (X_{\ell_1-1} - X_j) \mathbb{1}\{u \in C_j\}\mathrm{d}u \\
&+ \int_{B^c} (a - X_{\ell_1-1}) \mathbb{1}\{u \in C_{\ell_1-1}\}\mathrm{d}u \\
&+ \int_{B^c} (X_{\ell_2+1} - b) \mathbb{1}\{u \in C_{\ell_2+1}\}\mathrm{d}u \\
&+ \sum_{j=\ell_2+2}^{n} \int_{B^c} (X_j - X_{\ell_2+1}) \mathbb{1}\{u \in C_j\}\mathrm{d}u.
\end{aligned}$$

Exploiting $\lambda_1(C_j) = \frac{1}{n}$ for $j \in \{1, \ldots, n\}$, we see that

$$\begin{aligned}
\int_{B^c} |G(u) - T(u)|\mathrm{d}u \geqslant& \sum_{j=1}^{\ell_1-2} \Big(\frac{1}{n} - \sum_{i=\ell_1-1}^{\ell_2} \lambda_1(C_j \cap A_i)\Big)(X_{j+1} - X_j) \\
&+ \Big(\frac{1}{n} - \sum_{i=\ell_1-1}^{\ell_2} \lambda_1(C_{\ell_1-1} \cap A_i)\Big)(a - X_{\ell_1-1}) \\
&+ \Big(\frac{1}{n} - \sum_{i=\ell_1-1}^{\ell_2} \lambda_1(C_{\ell_2+1} \cap A_i)\Big)(X_{\ell_2+1} - b) \\
&+ \sum_{j=\ell_2+2}^{n} \Big(\frac{1}{n} - \sum_{i=\ell_1-1}^{\ell_2} \lambda_1(C_j \cap A_i)\Big)(X_j - X_{j-1}). \qquad (9)
\end{aligned}$$

Thus, using identity (8) together with inequalities (6), (7), and (9), we are led to

$$\begin{aligned}
\int_0^1 |G(u) - T(u)|\mathrm{d}u \geqslant& \frac{(X_{\ell_1} - a)^2}{4K} + \frac{(b - X_{\ell_2})^2}{4K} \\
&+ \sum_{j=1}^{\ell_1-2} \Big(\frac{1}{n} - \sum_{i=\ell_1-1}^{\ell_2} \lambda_1(C_j \cap A_i) + \sum_{i=\ell_1-1}^{\ell_2} \lambda_1(C_j \cap A_i)\Big)(X_{j+1} - X_j) \\
&+ \Big(\frac{1}{n} - \sum_{i=\ell_1-1}^{\ell_2} \lambda_1(C_{\ell_1-1} \cap A_i) + \sum_{i=\ell_1-1}^{\ell_2} \lambda_1(C_{\ell_1-1} \cap A_i)\Big)(a - X_{\ell_1-1})
\end{aligned}$$

$$+\sum_{i=\ell_1}^{\ell_2-1}\frac{(X_{i+1}-X_i)^2}{4K}$$

$$+\Big(\frac{1}{n}-\sum_{i=\ell_1-1}^{\ell_2}\lambda_1(C_{\ell_2+1}\cap A_i)+\sum_{i=\ell_1-1}^{\ell_2}\lambda_1(C_{\ell_2+1}\cap A_i)\Big)(X_{\ell_2+1}-b)$$

$$+\sum_{j=\ell_2+2}^{n}\Big(\frac{1}{n}-\sum_{i=\ell_1-1}^{\ell_2}\lambda_1(C_j\cap A_i)+\sum_{i=\ell_1-1}^{\ell_2}\lambda_1(C_j\cap A_i)\Big)(X_j-X_{j-1}).$$

So,

$$\begin{aligned}\int_0^1|G(u)-T(u)|\mathrm{d}u \geqslant{}& \frac{(X_{\ell_1}-a)^2}{4K}+\sum_{i=\ell_1}^{\ell_2-1}\frac{(X_{i+1}-X_i)^2}{4K}+\frac{(b-X_{\ell_2})^2}{4K}\\ &+\sum_{j\in\{1,\dots,\ell_1-2\}\cup\{\ell_2+1,\dots,n-1\}}\frac{X_{j+1}-X_j}{n}+\frac{1}{n}(a-X_{\ell_1-1})\\ &+\frac{1}{n}(X_{\ell_2+1}-b).\end{aligned}$$

Since $K\geqslant n\max_{i=1,\dots,n-1}(X_{i+1}-X_i)$, we have $\frac{X_{j+1}-X_j}{n}\geqslant\frac{(X_{j+1}-X_j)^2}{K}$, and thus

$$\begin{aligned}\frac{(X_{\ell_1}-a)^2}{4K}+\frac{1}{n}(a-X_{\ell_1-1}) &\geqslant \frac{1}{4K}\big((X_{\ell_1}-a)^2+4(a-X_{\ell_1-1})(X_{\ell_1}-X_{\ell_1-1})\big)\\ &=\frac{1}{4K}\big((X_{\ell_1}-a)^2+4(a-X_{\ell_1-1})(X_{\ell_1}-a)\\ &\quad+4(a-X_{\ell_1-1})^2\big)\\ &\geqslant\frac{(X_{\ell_1}-X_{\ell_1-1})^2}{4K}.\end{aligned}\tag{10}$$

Similarly,

$$\frac{(X_{\ell_2}-b)^2}{4K}+\frac{1}{n}(X_{\ell_2+1}-b)\geqslant\frac{(X_{\ell_2+1}-X_{\ell_2})^2}{4K}.$$

Using once again the assumption on $K$, we conclude that

$$\int_0^1|G(u)-T(u)|\mathrm{d}u\geqslant\sum_{i=1}^{n-1}\frac{(X_{i+1}-X_i)^2}{4K}.$$

To complete the proof, it remains to show that $\widehat{G}_K^\star$ and $\widehat{G}_K^\star\circ S$ are the only minimizers of (1) (Main Document). Returning to inequality (10), we see that if the function $G$ does not visit each data points, then

$$\int_0^1|G(u)-T(u)|\mathrm{d}u>\sum_{i=1}^{n-1}\frac{(X_{i+1}-X_i)^2}{4K}.$$

Also, according to (5), for the function $G$ to be optimal it needs to go at speed $K$ between each observation. Finally, with equation (3), we have that an optimal $G$ must be such that

$$\lambda_1\big(\{u \in [0,1] : |G(u) - X_i| \leqslant |G(u) - X_j|,\ j = 1,\ldots,n\}\big) = \frac{1}{n},$$

a property satisfied by $\widehat{G}_K^\star$ and $\widehat{G}_K^\star \circ S$ according to (4) (Main Document). We conclude that $\widehat{G}_K^\star$ and $\widehat{G}_K^\star \circ S$ are the unique minimizers of Problem (1) (Main Document) as they are the only functions satisfying these three conditions.

## 6. Proof of Proposition 3.3

The first statement is a straightforward consequence of Deheuvels (1984, Theorem 2). Regarding the second statement, we know from Theorem 3.2 that, for all $K \geqslant \underline{K}_1$,

$$W_1(\widehat{G}^\star_{K\sharp U}, \mu_n) = \inf_{G \in \mathrm{Lip}_K([0,1],\mathbb{R})} W_1(G_{\sharp U}, \mu_n) = \frac{1}{4K}\sum_{i=1}^{n-1}(X_{(i+1)} - X_{(i)})^2.$$

Therefore,

$$\begin{aligned} W_1(\widehat{G}^\star_{K\sharp U}, \mu_n) &\leqslant \frac{\sum_{i=1}^{n-1}(X_{(i+1)} - X_{(i)})^2}{n \max_{i=1,\ldots,n-1}(X_{(i+1)} - X_{(i)})} \\ &\leqslant \frac{1}{n}\sum_{i=1}^{n-1}(X_{(i+1)} - X_{(i)}) \\ &= \frac{1}{n}(X_{(n)} - X_{(1)}) \\ &\leqslant \frac{B-A}{n}. \end{aligned}$$

Recalling that $W_1(\mu, \mu_n) = \mathscr{O}(n^{-1/2})$ in probability (Fournier and Guillin, 2015, Theorem 1), the conclusion follows from the triangle inequality.

## 7. Proof of Proposition 4.1

The result is a consequence of the following lemma:

**Lemma 7.1.** *For each $G \in \mathrm{Lip}_K([0,1],\mathbb{R}^d)$, there exists a sequence of functions $(G_m)_{m\in\mathbb{N}}$ in $\mathrm{Lip}_K([0,1],\mathbb{R}^d)$ such that each $G_{m\sharp U}$ is nonatomic and $W_1(G_{m\sharp U}, \mu_n) \to W_1(G_{\sharp U}, \mu_n)$ as $m \to \infty$.*

**Proof.** Let $G \in \mathrm{Lip}_K([0,1],\mathbb{R}^d)$ and $m \in \mathbb{N}$. We define $G_m$ by slightly modifying $G$ on each interval where it is constant. More precisely, let $\mathcal{I}$ be the set of all non degenerated connected components of $G^{-1}(\{y \in \mathbb{R}^d : \lambda_1(G^{-1}(y)) > 0\})$. This set is at most countable and, since $G$ is continuous, it contains only disjoint closed intervals, i.e.,

$$\mathcal{I} = \{[a_\ell, b_\ell] : \ell \in \mathcal{L}\},$$

where $\mathcal{L} \subset \mathbb{N}$ and $0 \leqslant a_\ell < b_\ell \leqslant 1$. Let $K_m = \min(K, 1/m)$, $e_1 = (1, 0, \ldots, 0) \in \mathbb{R}^d$, and

$$G_m(u) = \begin{cases} G(a_\ell) + K_m\big(\frac{b_\ell - a_\ell}{2} - \big|\frac{a_\ell + b_\ell}{2} - u\big|\big)e_1 & \text{if } u \in [a_\ell, b_\ell] \text{ for some } \ell \in \mathcal{L} \\ G(u) & \text{otherwise.} \end{cases}$$

It is easy to see that $G_m \in \mathrm{Lip}_K([0,1], \mathbb{R}^d)$. Moreover, $G_m$ is not constant over any non degenerated interval. Thus, the distribution $G_{m\sharp U}$ is nonatomic. In addition, $\|G_m - G\|_\infty \to 0$ as $m \to \infty$. In particular, for any continuous bounded function $f : \mathbb{R}^d \to \mathbb{R}$, $\|f(G_m) - f(G)\|_\infty \to 0$, so that $G_{m\sharp U} \to G_{\sharp U}$ weakly, as $m$ tends to infinity. As the $G_{m\sharp U}$'s have supports included in the same compact set, we conclude by Villani (2008, Theorem 6.9) that $\lim_{m\to\infty} W_1(G_{m\sharp U}, G_{\sharp U}) = 0$. But, by the triangle inequality,

$$\big|W_1(G_{m\sharp U}, \mu_n) - W_1(G_{\sharp U}, \mu_n)\big| \leqslant W_1(G_{m\sharp U}, G_{\sharp U}),$$

from which $\lim_{m\to\infty} W_1(G_{m\sharp U}, \mu_n) = W_1(G_{\sharp U}, \mu_n)$ follows, as desired. □

## 8. Proof of Proposition 4.2

Assuming that such a transport map $T^\star \in \mathscr{H}^{w^\star}$ exists, we write $w^\star_{T^\star(x)}$ instead of $w^\star_i$ whenever $T^\star(x) = X_i$, $i \in \{1, \ldots, n\}$. Let $\varphi : \mathbb{R}^d \to \mathbb{R}$ be the 1-Lipschitz map defined by

$$\varphi(x) = \|x - T^\star(x)\| - w^\star_{T^\star(x)}.$$

Since $T^\star(X_i) = X_i$ for all $i \in \{1, \ldots, n\}$, we have in particular that $\varphi(x) - \varphi(T^\star(x)) = \|x - T^\star(x)\|$. Then, denoting by

$$\partial\varphi := \{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d : \varphi(x) - \varphi(y) = \|x - y\|\}$$

the superdifferential of $\varphi$ (Villani, 2008, Definition 5.7), the graph of $T^\star$ is included in $\partial\varphi$. Therefore,

$$\begin{aligned}\int_{\mathbb{R}^d\times\mathbb{R}^d} \|x - T^\star(x)\| \mathrm{d}\nu(x) &= \int_{\mathbb{R}^d\times\mathbb{R}^d} (\varphi(x) - \varphi(T^\star(x)))\mathrm{d}\nu(x) \\ &= \int_{\mathbb{R}^d} \varphi(x)\mathrm{d}\nu(x) - \int_{\mathbb{R}^d} \varphi(y)\mathrm{d}\mu_n(y) \\ &\leq W_1(\nu, \mu_n).\end{aligned}$$

We conclude that $T^\star$ is an optimal transport map.

## 9. Proof of Proposition 5.1

Let us first show that, for all $i \in \{1, \ldots, n + k - 1\}$ and $j \notin \{\sigma(i), \sigma(i + 1)\}$,

$$[V_i + \varphi(\sigma(i)), V_{i+1}] \cap \widehat{G}_K^{\star-1}(\mathrm{Vor}(j)^\circ) = \emptyset.$$

Suppose on the contrary that there exists $t \in (0, 1)$ such that $Y_i := X_{\sigma(i)} + t(X_{\sigma(i+1)} - X_{\sigma(i)}) \in \mathrm{Vor}(j)^\circ$. Then

$$X_j \in B^\circ(Y_i, \|X_{\sigma(i)} - Y_i\|) \cap B^\circ(Y_i, \|X_{\sigma(i+1)} - Y_i\|),$$

where $B^\circ(x,\varepsilon)$ stands for the open ball centered at $x$ of radius $\varepsilon$. Observe that for $t \leqslant 1/2$,

$$B^\circ(Y_i, \|X_{\sigma(i)} - Y_i\|) \subseteq B^\circ\Big(\frac{X_{\sigma(i)} + X_{\sigma(i+1)}}{2}, \frac{\|X_{\sigma(i+1)} - X_{\sigma(i)}\|}{2}\Big),$$

whereas for $t \geqslant 1/2$,

$$B^\circ(Y_i, \|X_{\sigma(i+1)} - Y_i\|) \subseteq B^\circ\Big(\frac{X_{\sigma(i)} + X_{\sigma(i+1)}}{2}, \frac{\|X_{\sigma(i+1)} - X_{\sigma(i)}\|}{2}\Big).$$

Consequently,

$$X_j \in B^\circ\Big(\frac{X_{\sigma(i)} + X_{\sigma(i+1)}}{2}, \frac{\|X_{\sigma(i+1)} - X_{\sigma(i)}\|}{2}\Big).$$

We deduce that $\langle X_{\sigma(i)} - X_j, X_{\sigma(i+1)} - X_j\rangle < 0$ (notation $\langle \cdot,\cdot\rangle$ means the scalar product), and so

$$\|X_{\sigma(i+1)} - X_{\sigma(i)}\|^2 > \|X_{\sigma(i+1)} - X_j\|^2 + \|X_{\sigma(i)} - X_j\|^2.$$

However, such an inequality is impossible by definition of $\sigma$. We conclude that, for all $t \in [0,1/2]$,

$$X_{\sigma(i)} + t(X_{\sigma(i+1)} - X_{\sigma(i)}) \in \mathrm{Vor}(\sigma(i))$$

and, for all $t \in [1/2,1]$,

$$X_{\sigma(i)} + t(X_{\sigma(i+1)} - X_{\sigma(i)}) \in \mathrm{Vor}(\sigma(i+1)).$$

Let us now turn to the computation of $W_1(\widehat{G}^\star_{K\sharp U}, \mu_n)$. First, by definition of $\varphi(i)$, for $i \in \{1,\dots,n\}$, we have

$$\begin{aligned}
&\sum_{j\in\sigma^{-1}(i)} \lambda_1\Big(\big[V_j, V_j + \varphi(i) + \frac{\|X_{\sigma(j+1)} - X_i\|}{2K}\big]\Big)\\
&+ \lambda_1\Big(\big[V_{j-1} + \varphi(\sigma(j-1)) + \frac{\|X_{\sigma(j-1)} - X_i\|}{2K}, V_{j-1} + \varphi(\sigma(j-1)) + \|X_{\sigma(j-1)} - X_i\|\big]\Big)\\
&= \sum_{j\in\sigma^{-1}(i)} \Big(\varphi(i) + \frac{\|X_{\sigma(j+1)} - X_i\|}{2K} + \frac{\|X_{\sigma(j-1)} - X_i\|}{2K}\Big)\\
&= \frac{1}{n}.
\end{aligned}$$

This shows that $\lambda_1(\widehat{G}^{\star -1}_K(\mathrm{Vor}(i))) = \frac{1}{n}, i \in \{1,\dots,n\}$—or, said differently, that the function $\widehat{G}^\star_K$ spends a total time $1/n$ in each Voronoi cell. Now, introduce $T^\star : \mathbb{R}^d \to \{X_1,\dots,X_n\}$ defined $\widehat{G}_{K\sharp U}$-almost everywhere by $T^\star(x) = X_i$ if $x \in \mathrm{Vor}(i)$. Then, clearly, $T^\star \in \mathscr{H}^0$, where we recall that

$$\begin{aligned}
\mathscr{H}^0 = \big\{T : \mathbb{R}^d \to \{X_1,\dots,X_n\} \ :\ &\forall x \in \mathrm{Vor}(i), T(x) = X_i\\
&\text{and } \forall x \in \Gamma^0_{j_1\dots j_p}, T(x) \in \{X_{j_1},\dots,X_{j_p}\}\big\}.
\end{aligned}$$

Arguing as in the proof of Lemma 7.1, one shows that there exists a sequence of functions $(G^\star_m)_{m\in\mathbb{N}} \subset \mathrm{Lip}_K([0,1],\mathbb{R}^d)$ such that each $G^\star_{m\sharp U}$ is nonatomic, $W_1(G^\star_{m\sharp U}, \mu_n) \to W_1(\widehat{G}^\star_{K\sharp U}, \mu_n)$ as $m \to \infty$, and, for all $m$ large enough, $\lambda_1(G_m^{\star-1}(\mathrm{Vor}(i))) = \frac{1}{n}$, $i \in \{1, \ldots, n\}$. According to Proposition 4.2, we have

$$W_1(G^\star_{m\sharp U}, \mu_n) = \int_0^1 \|G^\star_m(u) - T^\star(G^\star_m(u))\| \mathrm{d}u.$$

By dominated convergence, we obtain $W_1(\widehat{G}^\star_{K\sharp U}, \mu_n) = \int_0^1 \|\widehat{G}^\star_K(u) - T^\star(\widehat{G}^\star_K(u))\|\mathrm{d}u$, so that $T^\star$ is an optimal transport map from $\widehat{G}^\star_K$ to $\mu_n$. Finally,

$$\begin{aligned}
W_1(\widehat{G}^\star_{K\sharp U}, \mu_n) &= \int_0^1 \|\widehat{G}^\star_K(u) - T^\star(\widehat{G}^\star_K(u))\|\mathrm{d}u \\
&= \sum_{j=1}^{n+k-1} \int_{V_j}^{V_j+\varphi(\sigma(j))+\frac{\|X_{\sigma(j+1)}-X_{\sigma(j)}\|}{2K}} \|X_{\sigma(j)} - \widehat{G}^\star_K(u)\|\mathrm{d}u \\
&\quad + \int_{V_j+\varphi(\sigma(j))+\frac{\|X_{\sigma(j+1)}-X_{\sigma(j)}\|}{2K}}^{V_j+\varphi(\sigma(j))+\|X_{\sigma(j+1)}-X_{\sigma(j)}\|} \|X_{\sigma(j+1)} - \widehat{G}^\star_K(u)\|\mathrm{d}u \\
&= \sum_{j=1}^{n+k-1} \int_{V_j+\varphi(\sigma(j))}^{V_j+\varphi(\sigma(j))+\frac{\|X_{\sigma(j+1)}-X_{\sigma(j)}\|}{2K}} K\big(u - (V_j + \varphi(\sigma(j)))\big)\mathrm{d}u \\
&+ \int_{V_j+\varphi(\sigma(j))+\frac{\|X_{\sigma(j+1)}-X_{\sigma(j)}\|}{2K}}^{V_j+\varphi(\sigma(j))+\|X_{\sigma(j+1)}-X_{\sigma(j)}\|} K(V_j + \varphi(\sigma(j)) + \|X_{\sigma(j+1)} - X_{\sigma(j)}\| - u)\mathrm{d}u \\
&= \sum_{j=1}^{n+k-1} \frac{1}{8K}\|X_{\sigma(j+1)} - X_{\sigma(j)}\|^2 + \frac{1}{8K}\|X_{\sigma(j+1)} - X_{\sigma(j)}\|^2 \\
&= \frac{1}{4K}\sum_{j=1}^{n+k-1} \|X_{\sigma(j+1)} - X_{\sigma(j)}\|^2.
\end{aligned}$$

## 10. Proof of Proposition 5.4

First note, since $\sigma$ is a path with points that may be visited several times, that

$$\begin{aligned}
\underline{K}_2 &\geqslant \sum_{i=1}^{n} \sum_{j\in\sigma^{-1}(i)} \frac{1}{2}(\|X_{\sigma(j-1)} - X_i\| + \|X_{\sigma(j+1)} - X_i\|) \\
&\geqslant \inf_{\tau\in\mathscr{P}_n} \sum_{j=1}^{n-1} \|X_{\tau(j)} - X_{\tau(j+1)}\|, \qquad (11)
\end{aligned}$$

where $\mathscr{P}_n$ stands for the set of permutations of $\{1, \ldots, n\}$. But, according to Steele (1988), under the conditions of the theorem, there exists a constant $C > 0$ satisfying

$$\lim_{n\to\infty} n^{-1+1/d} \inf_{\tau\in\mathscr{P}_n} \sum_{j=1}^{n-1} \|X_{\tau(j)} - X_{\tau(j+1)}\| = C \text{ a.s.}$$

This shows the first statement of the proposition.

We start the proof of the second statement by recalling that, according to Fournier and Guillin (2015, Theorem 1), one has, in probability,

$$W_1(\mu,\mu_n) = \begin{cases} \mathscr{O}(\frac{\log n}{\sqrt{n}}) & \text{for } d=2 \\ \mathscr{O}(n^{-1/d}) & \text{for } d \geqslant 3. \end{cases}$$

Therefore, by the triangle inequality, it is enough to show that, for $d \geqslant 2$, in probability,

$$W_1(\widehat{G}^{\star}_{K\sharp U},\mu_n) = \mathscr{O}(n^{-1/d}).$$

According to Theorem 5.3, we only need to show that, in probability,

$$\frac{1}{4K}\sum_{j=1}^{n+k-1} \|X_{\sigma(j+1)} - X_{\sigma(j)}\|^2 = \mathscr{O}(n^{-1/d}),$$

whenever $K \geqslant \underline{K}_2$. But, by the very definition (12) (Main Document) of the pair $(k,\sigma)$, we have

$$\sum_{j=1}^{n+k-1} \|X_{\sigma(j+1)} - X_{\sigma(j)}\|^2 \leqslant \sum_{j=1}^{n-1} \|X_{\tau(j+1)} - X_{\tau(j)}\|^2,$$

where $\tau \in \mathscr{P}_n$ is a permutation that minimizes the length among the whole set of paths that visit only once each data, i.e.,

$$\sum_{j=1}^{n-1} \|X_{\tau(j+1)} - X_{\tau(j)}\| \leqslant \sum_{j=1}^{n-1} \|X_{\tau'(j+1)} - X_{\tau'(j)}\|, \text{ for all } \tau' \in \mathscr{P}_n.$$

Therefore, since $K \geqslant \underline{K}_2$, we have by inequality (11) ,

$$\frac{1}{K}\sum_{j=1}^{n+k-1} \|X_{\sigma(j+1)} - X_{\sigma(j)}\|^2 \leqslant \frac{\sum_{j=1}^{n-1} \|X_{\tau(j+1)} - X_{\tau(j)}\|^2}{\sum_{j=1}^{n-1} \|X_{\tau(j+1)} - X_{\tau(j)}\|}.$$

Now, under the additional condition on the density of $\mu$, we know by Yukich (2000, Theorem 1.3) that, for each $0 \leqslant \ell \leqslant d$, there exists $C(\ell) > 0$ such that

$$\lim_{n\to\infty} n^{-1+\ell/d} \sum_{j=1}^{n-1} \|X_{\tau(j+1)} - X_{\tau(j)}\|^{\ell} = C(\ell) \text{ a.s.}$$

By the above, we conclude that

$$\frac{1}{4K}\sum_{j=1}^{n+k-1} \|X_{\sigma(j+1)} - X_{\sigma(j)}\|^2 = \mathscr{O}(n^{-1/d}) \text{ a.s.}$$

# References

AMBROSIO, L. and GIGLI, N. (2013). A User's Guide to Optimal Transport. In *Modelling and Optimisation of Flows on Networks: Cetraro, Italy 2009* (B. PICCOLI and M. RASCLE, eds.) 1–155. Springer, Berlin.

DEHEUVELS, P. (1984). Strong limit theorems for maximal spacings from a general univariate distribution. *Ann. Probab.* **12** 1181–1193.

EVANS, L. C. and GARIEPY, R. F. (2015). *Measure Theory and Fine Properties of Functions*. CRC Press, Boca Raton.

FOURNIER, N. and GUILLIN, A. (2015). On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields* **162** 707–738.

PRATELLI, A. (2007). On the equality between Monge's infimum and Kantorovich's minimum in optimal mass transportation. *Ann. Inst. Henri Poincaré Probab. Stat.* **43** 1–13.

SANTAMBROGIO, F. (2015). *Optimal Transport for Applied Mathematicians*. Birkäuser, Cham.

STEELE, J. M. (1988). Growth rates of Euclidean minimal spanning trees with power weighted edges. *Ann. Probab.* **16** 1767–1787.

VILLANI, C. (2008). *Optimal Transport: Old and New*. Springer, Berlin.

YUKICH, J. E. (2000). Asymptotics for weighted minimal spanning trees on random points. *Stochastic Process. Appl.* **85** 123–138.