Supplementary Material for Distributional Convergence of the Sliced Wasserstein Process

Jiaqi Xi[1] and Jonathan Niles-Weed[1,2]

[1]Courant Institute of Mathematical Sciences, New York University, NY 10012
[2]Center for Data Science, New York University, NY 10011

A Proofs

A.1 Proof of Theorem 2.1

To prove Theorem 2.1, we will first show a uniform central limit theorem for the empirical process indexed by the set $\mathcal{C} \times \mathbb{S}^{d-1}$ where $\mathcal{C} := \{f : [-R, R] \to \mathbb{R},\ \phi(0) = 0,\ \|f\|_{\text{Lip}} \le L\}$ for some positive quantity L. Explicitly, for $f \in \mathcal{C}$, $u \in \mathbb{S}^{d-1}$, write h_u for the function $x \mapsto u^\top x$, let $X_1, \ldots, X_n \sim P, Y_1, \ldots, Y_m \sim Q$ denote n and m i.i.d. samples generated from P and Q respectively, and define $\mathbb{B}_n \in \ell^\infty(\mathcal{C} \times \mathbb{S}^{d-1})$ by

$$\begin{aligned}\mathbb{B}_{nm}(f, u) &:= \sqrt{n}(P_n - P)(f \circ h_u) + \sqrt{m}(Q_m - Q)(f^c \circ h_u) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n f(u^\top X_i) - \mathbb{E} f(u^\top X_i) + \frac{1}{\sqrt{m}} \sum_{j=1}^m f^c(u^\top Y_i) - \mathbb{E} f^c(u^\top Y_i)\,.\end{aligned}$$

We first show that $\mathbb{B}_n$ possesses a weak limit.

Proposition A.1. *As $n, m \to \infty$, the empirical process $\mathbb{B}_{nm}$ satisfies*

$$\mathbb{B}_{nm} \rightsquigarrow \mathbb{B} \quad \textit{in } \ell^\infty(\mathcal{C} \times \mathbb{S}^{d-1}),$$

where $\mathbb{B}$ is the tight Gaussian process with covariance

$$\begin{aligned}\mathbb{E}\mathbb{B}(f, u)\mathbb{B}(g, v) = &\int f(u^\top x) g(v^\top x)\, \mathrm{d}P(x) - \int f(u^\top x)\, \mathrm{d}P(x) \int g(v^\top x)\, \mathrm{d}P(x) \\ &+ \int f^c(u^\top y) g^c(v^\top y)\, \mathrm{d}Q(y) - \int f^c(u^\top y)\, \mathrm{d}Q(y) \int g^c(v^\top y)\, \mathrm{d}Q(y)\,. \end{aligned} \tag{1}$$

Moreover, this process is uniformly continuous with respect to the semimetric

$$\rho((f, u), (g, v)) = \|f \circ h_u - g \circ h_v\|_{L^2(P)} + \|f^c \circ h_u - g^c \circ h_v\|_{L^2(Q)}\,, \tag{2}$$

with respect to which $\mathcal{C} \times \mathbb{S}^{d-1}$ is totally bounded.

Proof. The assertions of this proposition will follow from the fact that the classes of functions $\mathcal{F} := \{f \circ h_u(x) : (f, u) \in \mathcal{C} \times \mathbb{S}^{d-1}\}$ and $\mathcal{F}^C := \{f^c \circ h_u(x) : (f, u) \in \mathcal{C} \times \mathbb{S}^{d-1}\}$ are P and Q-Donsker, respectively. Indeed, if we assume this Donsker property, then we have

$$\begin{aligned}\sqrt{n}(P_n - P) &\rightsquigarrow \mathbb{B}_P \quad \text{in } \ell^\infty(\mathcal{F}) \\ \sqrt{m}(Q_m - Q) &\rightsquigarrow \mathbb{B}_Q \quad \text{in } \ell^\infty(\mathcal{F}^C)\end{aligned}$$

for tight P- and Q-Brownian bridges $\mathbb{B}_P$ and $\mathbb{B}_Q$, respectively [see 11, Section 2.1]. These tight Gaussian processes possess uniformly continuous sample paths with respect to the semi-metrics ρ_P and ρ_Q, respectively, where for $F, G \in \mathcal{F}$,

$$\rho_P^2(F, G) = \int (F - G)^2\, \mathrm{d}P - \left(\int (F - G)\, \mathrm{d}P\right)^2\,,$$

and analogously for Q, and $\mathcal{F}$ and $\mathcal{F}^C$ are totally bounded with respect to these semi-metrics [11, Example 1.5.10]. In particular, since ρ_P is dominated by the $L^2(P)$ norm, and likewise for Q, these processes are also uniformly $L^2(P)$ and $L^2(Q)$ continuous and $\mathcal{F}$ and $\mathcal{F}^C$ are $L^2(P)$ and $L^2(Q)$ totally bounded.

Since P_n and Q_m are independent, the above considerations imply that

$$(\sqrt{n}(P_n - P), \sqrt{m}(Q_m - Q)) \rightsquigarrow (\mathbb{B}_P, \mathbb{B}_Q) \quad \text{in } \ell^\infty(\mathcal{F}) \times \ell^\infty(\mathcal{F}^C),$$

for a tight Gaussian limit $(\mathbb{B}_P, \mathbb{B}_Q)$ with sample paths almost surely continuous with respect to the metric on $\mathcal{F} \times \mathcal{F}^C$ given by the sum of the $L^2(P)$ and $L^2(Q)$ metrics on $\mathcal{F}$ and $\mathcal{F}^C$. Finally, there exists a continuous map from $\ell^\infty(\mathcal{F}) \times \ell^\infty(\mathcal{F}^C)$ to $\ell^\infty(\mathcal{C} \times \mathbb{S}^{d-1})$ given by associating $(S, T) \in \ell^\infty(\mathcal{F}) \times \ell^\infty(\mathcal{F}^C)$ with the element of $\ell^\infty(\mathcal{C} \times \mathbb{S}^{d-1})$ sending (f, u) to $S(f \circ h_u) + T(f^c \circ h_u)$, and the continuous mapping theorem therefore furnishes the desired convergence.

It remains to show that $\mathcal{F}$ and $\mathcal{F}^C$ are P- and Q-Donsker. We first prove that $\mathcal{F}$ is P-Donsker. By [11, Theorem 2.5.6], it suffices to show that

$$\int_0^\infty \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}, L_2(P))}\, \mathrm{d}\epsilon < \infty \,. \tag{3}$$

By Lemma A.4, we may replace the bracketing number $N_{[]}(\varepsilon, \mathcal{F}, L_2(P))$ by the uniform covering number $N(\varepsilon/2, \mathcal{F}, \|\cdot\|_\infty)$.

[11, Theorem 2.7.1] gives an upper bound for the covering entropy of $\mathcal{C}$: there exists some positive constant C_1 depending on R and L such that

$$\log N(\varepsilon, \mathcal{C}, \|\cdot\|_\infty) \le C_1 \varepsilon^{-1}. \tag{4}$$

[12, Lemma 6.2] shows that there exists a positive constant C_2 depending on R such that

$$N(\varepsilon, \{h_u,\, u \in \mathbb{S}^{d-1}\}, \|\cdot\|_\infty) = N(\varepsilon/2, \mathbb{S}^{d-1}, \|\cdot\|_2) \le C_2 \varepsilon^{-d} \,. \tag{5}$$

Consequently, apply Lemma A.5 to $N(\varepsilon, \mathcal{F}, \|\cdot\|_\infty)$ and we get

$$N(\varepsilon, \mathcal{F}, \|\cdot\|_\infty) \le \log N(\varepsilon, \mathcal{C}, \|\cdot\|_\infty) + \log N(\varepsilon/2, \mathbb{S}^{d-1}, \|\cdot\|_2) \le C_1 \varepsilon^{-1} + \log(C_2 \varepsilon^{-d}) \,. \tag{6}$$

We obtain that

$$\begin{aligned}
\int_0^\infty \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}, L_2(P))}\, \mathrm{d}\epsilon &\le \int_0^\infty \sqrt{N(\varepsilon/2, \mathcal{F}, \|\cdot\|_\infty)}\, \mathrm{d}\epsilon \\
&= \int_0^{2\mathrm{diam}(\mathcal{F})} \sqrt{N(\varepsilon/2, \mathcal{F}, \|\cdot\|_\infty)}\, \mathrm{d}\epsilon \\
&\le \int_0^{2\mathrm{diam}(\mathcal{F})} \sqrt{C_1 \varepsilon^{-1} + \log(C_2 \varepsilon^{-d})}\, \mathrm{d}\epsilon \\
&< \infty \,.
\end{aligned}$$

Therefore $\mathcal{F}$ is P-Donsker. The argument for $\mathcal{F}^C$ is identical, since Lemma A.6 shows that the estimate (6) holds for $\mathcal{F}^C$ as well. □

To prove Theorem 2.1, we combine the above result with the functional delta method. Let $\iota : \ell^\infty(\mathcal{C} \times \mathbb{S}^{d-1}) \to \ell^\infty(\mathbb{S}^{d-1})$ be defined by

$$\iota(\Phi)(u) := \sup_{f \in \mathcal{C}} \Phi(f, u) \,. \tag{7}$$

The following proposition shows that ι is Hadamard directionally differentiable tangentially to the set of continuous functions at all functions for which the supremum in (7) is uniquely achieved.

Proposition A.2. *Let $\iota : \ell^\infty(\mathcal{C} \times \mathbb{S}^{d-1}) \to \ell^\infty(\mathbb{S}^{d-1})$ be defined as in (7), and denote by $\mathsf{C}_u(\mathcal{C} \times \mathbb{S}^{d-1}, \rho)$ the set of elements of $\ell^\infty(\mathcal{C} \times \mathbb{S}^{d-1})$ which are uniformly continuous with respect to the semi-metric ρ defined in (2). Then for $\Phi \in \mathsf{C}_u(\mathcal{C} \times \mathbb{S}^{d-1}, \rho)$ such that $\Phi(\cdot, u)$ has a unique maximizer for every $u \in \mathbb{S}^{d-1}$, the function ι is Hadamard directionally differentiable at Φ tangentially to $\mathsf{C}_u(\mathcal{C} \times \mathbb{S}^{d-1}, \rho)$, with derivative $\iota'_\Phi : \ell^\infty(\mathcal{C} \times \mathbb{S}^{d-1}) \to \ell^\infty(\mathbb{S}^{d-1})$ given by*

$$\iota'_\Phi(\Psi)(u) = \Psi(f_u, u) \,, \quad u \in \mathbb{S}^{d-1}, \Psi \in \mathsf{C}_u(\mathcal{C} \times \mathbb{S}^{d-1}, \rho) \,, \tag{8}$$

where $\Phi(f_u, u) = \sup_{\mathcal{C}} \Phi(\cdot, u)$.

Proof. Fix an arbitrary $u \in \mathbb{S}^{d-1}$. [2, Theorem 2.1] shows that the function $\iota^u : \ell^\infty(\mathcal{C} \times \mathbb{S}^{d-1}) \to \mathbb{R}$ defined by

$$\iota^u(\Phi) := \sup_{f \in \mathcal{C}} \Phi(f, u)$$

is Hadamard directionally differentiable, with derivative

$$(\iota^u)'_\Phi(\Psi) = \lim_{\epsilon \to 0} \sup_{f \in B_{\epsilon,u}(\Phi)} \Psi(f, u) \tag{9}$$

where $B_{\epsilon,u}(\Phi) := \{f \in \mathcal{C} : \Phi(f,u) \geq \sup_{\mathcal{C}} \Phi(\cdot, u) - \epsilon\}$. Moreover, we claim that if Φ and Ψ are uniformly continuous, then the expression for the derivative simplifies to

$$(\iota^u)'_\Phi(\Psi) = \Psi(f_u, u) = \iota'_\Phi(\Psi)(u)\,. \tag{10}$$

To see this, we define $\phi, \psi \in \ell^\infty(\mathcal{C})$ by $\phi(\cdot) = \Phi(\cdot, u)$ and $\psi(\cdot) = \Psi(\cdot, u)$, so that the right side of (9) reads

$$\lim_{\epsilon \to 0} \sup_{f \in B_\epsilon(\phi)} \psi(f)\,,$$

where $B_\epsilon(\phi) := \{f \in \mathcal{C} : \phi(f) \geq \sup_{\mathcal{C}} \phi - \epsilon\}$. By assumption, the functions ϕ and ψ are uniformly continuous with respect to the semi-metric

$$\rho^u(f, g) := \|f \circ h^u - g \circ h^u\|_{L^2(P)} + \|f^c \circ h^u - g^c \circ h^u\|_{L^2(Q)}\,,$$

and $\mathcal{C}$ is totally bounded with respect to this semi-metric. Following [2, Corollary 2.5], it is enough to show that $\mathcal{C}$ is complete with respect to ρ^u. The completeness of $L^2((h_u)_\sharp P)$ implies that any sequence $f_n \in \mathcal{C}$ which is Cauchy with respect to ρ^u possesses a limit f, and by passing to a subsequence we may assume that $f_n \to f$ pointwise, and, since the elements of $\mathcal{C}$ are bounded and equicontinuous, we may further assume that $f_n \to f$ uniformly on $[-R, R]$ by the Arzelà–Ascoli theorem. Since $\mathcal{C}$ is closed with respect to pointwise convergence, $f \in \mathcal{C}$, and since c-transforms are preserved under uniform convergence, we also have $f^c_{n_k} \to f^c$ uniformly in $[-R, R]$. Therefore $(f_{n_k}, f^c_{n_k}) \to (f, f^c)$ for some $f \in \mathcal{C}$ uniformly, and hence $f_{n_k} \to f$ in ρ^u. This proves (10).

We now turn to the differentiability of ι. To prove the Hadamard differentiability of ι, according to Proposition 3.5 of [10], is equivalent to prove ι is Lipschitz and that the function $\iota_n(\Phi, \Psi) \in \ell^\infty(\mathbb{S}^{d-1})$ defined by

$$\iota_n(\Phi, \Psi)(\cdot) := \sup_{f \in \mathcal{C}} (s_n \Phi(f, \cdot) + \Psi(f, \cdot)) - s_n \sup_{f \in \mathcal{C}} \Phi(f, \cdot)$$

converges uniformly to the limit $\iota'_\Phi(\Psi)(\cdot)$ for any positive increasing sequence $s_n \to \infty$. The Lipschitz property is obvious. Indeed, for any $\Phi_1, \Phi_2 \in \ell^\infty(\mathcal{C} \times \mathbb{S}^{d-1})$,

$$\sup_{u \in \mathbb{S}^{d-1}} \left| \sup_f \Phi_1(f, u) - \sup_f \Phi_2(f, u) \right| \leq \sup_{f,u} |\Phi_1(f, u) - \Phi_2(f, u)| \leq \|\Phi_1 - \Phi_2\|_{\ell^\infty(\mathcal{C} \times \mathbb{S}^{d-1}])}.$$

As for the uniform convergence, we first show that $\iota_n(\Phi, \Psi)(u) \to \iota'_\Phi(\Psi)(u)$ pointwise. This follows directly from the Hadamard differentiability of ι^u, since

$$\iota_n(\Phi, \Psi)(u) = \iota^u(s_n \Phi + \Psi) - s_n \iota^u(\Phi) \to (\iota^u)'_\Phi(\Psi) = \iota'_\Phi(\Psi)(u) \quad \text{as } n \to \infty.$$

Moreover, we show in Lemma A.3 that the functions $\iota_n(\Phi, \Psi)$ and $\iota'(\Phi, \Psi)$ are continuous on $\mathbb{S}^{d-1}$. Therefore, by [8, Theorem 7.13], to show uniform convergence on $\mathbb{S}^{d-1}$ it suffices to show that the sequence $\{\iota_n(\Phi, \Psi)\}_{n \geq 1}$ is monotonically non-increasing for all $u \in \mathbb{S}^{d-1}$. This follows directly from the definition of ι_n. Indeed, for any $u \in \mathbb{S}^{d-1}$, we have

$$\begin{aligned}
&\iota_n(\Phi, \Psi)(u) - \iota_{n+1}(\Phi, \Psi)(u) \\
&\quad = \sup_{f \in \mathcal{C}} (s_n \Phi(f, u) + \Psi(f, u)) - \sup_{f \in \mathcal{C}} (s_{n+1} \Phi(f, u) + \Psi(f, u)) + \sup_{f \in \mathcal{C}} (s_{n+1} - s_n) \Phi(f, u) \\
&\qquad\qquad \geq \sup_{f \in \mathcal{C}} ((s_n + s_{n+1} - s_n) \Phi(f, u) + \Psi(f, u)) - \sup_{f \in \mathcal{C}} (s_{n+1} \Phi(f, u) + \Psi(f, u)) = 0.
\end{aligned}$$

The last inequality results from the reverse triangle inequality for the supremum. This finishes the proof for Hadamard directional differentiability of ι. $\square$

Lemma A.3. *For $\Phi, \Psi \in \mathsf{C}_u(\mathcal{C} \times \mathbb{S}^{d-1})$, then the functions $\iota_n(\Phi, \Psi)$ are continuous on $\mathbb{S}^{d-1}$. If $\Phi(\cdot, u)$ has a unique maximizer for every $u \in \mathbb{S}^{d-1}$, then $\iota'(\Phi, \Psi)$ is also continuous.*

Proof. For the continuity of ι_n, it suffices to show that if $u \to v$, then

$$\sup_{f \in \mathcal{C}} |\Phi(f, u) - \Phi(f, v)| \to 0\,, \tag{11}$$

and analogously for Ψ. This follows directly from uniform continuity: for any $\epsilon > 0$, there exist a $\delta > 0$ such that

$$\rho((f, u), (g, v)) \leq \delta \implies |\Phi(f, u) - \Phi(g, v)| \leq \epsilon\,.$$

In particular, we have $\sup_{f \in \mathcal{C}} |\Phi(f, u) - \Phi(f, v)| \leq \epsilon$ if $\sup_f \rho((f, u), (f, v)) \leq \delta$. Moreover, since the elements of $\mathcal{C}$ and their c-transforms are uniformly Lipschitz, we have

$$\sup_{f \in \mathcal{C}} \rho((f, u), (f, v)) \leq C \left(\|h_u - h_v\|_{L^2(P)} + \|h_u - h_v\|_{L^2(Q)}\right)$$

for a positive C independent of f, and the right side of the above expression converges to 0 as $u \to v$. Therefore (11) holds, as does the analogous convergence for Ψ. This proves continuity of ι_n.

For ι', we have

$$\iota'_\Phi(\Psi)(u) - \iota_\Phi(\Psi)(v) = \Psi(f_u, u) - \Psi(f_v, v)$$

where f_u, f_v are the maximizers of $\Phi(\cdot, u), \Phi(\cdot, v)$ respectively. Choose any sequence $v_n \to v$ and let f_n denote the unique maximizers of $\Phi(\cdot, v_n)$ correspondingly. Since $\mathcal{C} \times \mathbb{S}^{d-1}$ is totally bounded and complete, we may upon passing to a subsequence assume $(f_n, v_n) \to (f, v) \in \mathcal{C} \times \mathbb{S}^{d-1}$, and by the uniform continuity of Φ we must have $\Phi(f, v) = \sup_{f \in \mathcal{C}} \Phi(f, v)$, and since we have assumed that the supremum is uniquely achieved, $f = f_v$. Since this argument holds on any subsequence, we obtain that the whole sequence converges to (f_v, v), and the uniform continuity of Ψ implies that

$$\lim_{n \to \infty} \Psi(f_n, v_n) = \Psi(f_v, v)\,,$$

as desired. $\square$

We are now in a position to prove the main theorem.

Proof of Theorem 2.1. For the sake of notational simplicity, we prove the special case of (6) when $n = m$ with both sides multiplied by $\sqrt{2}$. Namely, we are going to show

$$\sqrt{n}\left(W_p^p(P_{n\cdot}, Q_{n\cdot}) - W_p^p(P_\cdot, Q_\cdot)\right) \rightsquigarrow \sqrt{2}\mathbb{G} \quad \text{in } \ell^\infty(\mathbb{S}^{d-1}). \tag{12}$$

The general conclusion with $n \neq m$ follows by an analogous argument.

Fix $u \in \mathbb{S}^{d-1}$. By Kantorovich duality, we may write the Wasserstein distance as

$$W_p^p(P_u, Q_u) = \sup_{f \in \mathcal{C}} \int f \circ h_u \, \mathrm{d}P + \int f^c \circ h_u \, \mathrm{d}Q\,.$$

Define $\Phi_{(P,Q)} : \mathcal{C} \times \mathbb{S}^{d-1} \to \mathbb{R}$ by $\Phi_{(P,Q)} = \mathbb{E}_{X \sim P} f \circ h_u(X) + \mathbb{E}_{X \sim Q} f^c \circ h_u(Y)$. Note that $\Phi_{(P,Q)}$ is uniformly continuous with respect to ρ. Indeed, for any $f, g \in \mathcal{C}$ and $u, v \in \mathbb{S}^{d-1}$,

$$\begin{aligned}
&\left|\Phi_{(P,Q)}(f, u) - \Phi_{(P,Q)}(g, v)\right| \\
&\leq \mathbb{E}_{X \sim P} |f \circ h_u(X) - g \circ h_v(X)| + \mathbb{E}_{Y \sim Q} |f^c \circ h_u(Y) - g^c \circ h_v(Y)| \\
&\leq \left(\mathbb{E}_{X \sim P} |f \circ h_u(X) - g \circ h_v(X)|^2\right)^{1/2} + \left(\mathbb{E}_{Y \sim Q} |f^c \circ h_u(Y) - g^c \circ h_v(Y)|^2\right)^{1/2} \\
&= \rho((f, u), (g, v)).
\end{aligned}$$

Moreover, for any $u \in \mathbb{S}^{d-1}$, $\Phi(\cdot, u)$ achieves the maximum over $\mathcal{C}$ at a unique $f_u \in \mathcal{C}$, since under the assumption of (**CC**), the Kantorovich potential corresponding to P_u and Q_u is unique.

We now apply the functional delta method to Proposition A.1 with the supremum function ι. By Kantorovich duality, $\iota(\Phi_{(P,Q)})(u) = W_p^p(P_u, Q_u)$ and $\iota(\Phi_{(P_n,Q_n)})(u) = W_p^p(P_{nu}, Q_{nu})$ Proposition A.2 implies that ι is Hadamard directionally differentiable and the derivative of ι at $\Phi_{(P,Q)}$ in the direction $\mathbb{B}$ is given by

$$\iota'_\Phi(\mathbb{B}_{(P,Q)})(u) = \mathbb{B}_{(P,Q)}(f_u, u).$$

Hence,

$$\sqrt{n}\left(\iota(\Phi_{(P_n,Q_n)}) - \iota(\Phi_{(P,Q)})\right) \rightsquigarrow \iota'_{\Phi_{(P,Q)}}(\mathbb{B})(\cdot) = \mathbb{B}(f_\cdot,\cdot) \quad \text{in } \ell^\infty(\mathbb{S}^{d-1}). \tag{13}$$

This is a centered tight Gaussian process on $\mathbb{S}^{d-1}$, and a direct computation shows that its covariance agrees with that of $\sqrt{2}\mathbb{G}$, as desired. The case where $n \neq m$ follows similarly; the details are omitted. □

A.1.1 Additional Lemmas

The following lemma is included in the proof for Corollary 2.7.2 of [11]. For the sake of clarity, we state it here separately.

Lemma A.4. *Suppose that P is a probability distribution on $\mathbb{R}^d$, then the bracketing number of any class of functions $\mathcal{H}$ with respect to $L_2(P)$ can be bounded above by the covering number with respect to the uniform norm. Explicitly, for $\varepsilon > 0$, we have*

$$N_{[]}(2\varepsilon, \mathcal{H}, L_2(P)) \leq N(\varepsilon, \mathcal{H}, \|\cdot\|_\infty). \tag{14}$$

Proof. The proof is inspired by that of Lemma 6.2 of [12]. Take any $g \in \mathcal{H}$ and suppose that it lies in some ball $B_{\|\cdot\|_\infty}(f, \varepsilon)$. Let $l = f - \varepsilon$, $u = f + \varepsilon$, then

$$l - g = f - g - \varepsilon \leq \varepsilon - \varepsilon = 0,\; u - g = f - g + \varepsilon \geq -\varepsilon + \varepsilon = 0,$$

and $\|l - u\|_{L^2(P)} = 2\varepsilon$. □

The next lemma is included in the proof for Lemma 4.2 of [4]. Again we state and prove it here separately for completeness.

Lemma A.5. *Consider some set of composite functions $\mathcal{A}_\mathcal{H} := \{f \circ h :\ f \in \mathcal{A},\ h \in \mathcal{H}\}$ where the functions in $\mathcal{A}$ are K-Lipschitz. Then*

$$N((K+1)\varepsilon, \mathcal{C}_\mathcal{F}, \|\cdot\|_\infty) \leq N(\varepsilon, \mathcal{C}_R, \|\cdot\|_\infty) \times N(\varepsilon, \mathcal{F}, \|\cdot\|_\infty). \tag{15}$$

Proof. Indeed, for any $g := f \circ h \in \mathcal{A}_\mathcal{H}$, there exists some $f_0 \in \mathcal{A}$ and $h_0 \in \mathcal{H}$ such that $\|f - f_0\|_\infty, \|h - h_0\|_\infty < \varepsilon$, then

$$\begin{aligned}&\|g - f_0 \circ h_0\|_\infty \leq \|f \circ h - f \circ h_0\|_\infty + \|f \circ h_0 - f_0 \circ h_0\|_\infty \\ &\leq K\|h - h_0\|_\infty + \|f - f_0\|_\infty < (K+1)\varepsilon.\end{aligned}$$

□

Lemma A.6. *The covering entropy of $\mathcal{F}^C$ with respect to L^∞ is upper bounded by that of $\mathcal{F}$.*

Proof. Fix any $f^c \in \mathcal{F}^C$. There exists some $f_0 \in \mathcal{F}$ such that $\|f - f_0\|_\infty < \varepsilon$. Then

$$\begin{aligned}\|f^c - f_0^c\|_\infty &= \|\inf_x(|x-y|^p - f(x)) - \inf_x(|x-y|^p - f_0(x))\|_\infty \\ &= \|\sup_x(f_0(x) - |x-y|^p) - \sup_x(|x-y|^p - f(x))\|_\infty \\ &\leq \|\sup_x(|x-y|^p - f(x)) - (|x-y|^p - f_0(x))\|_\infty \\ &= \|\sup_x(f_0(x) - f(x))\|_\infty \leq \|f_0 - f\|_\infty < \varepsilon\,.\end{aligned}$$

The conclusion follows immediately. □

A.2 Proof of Theorem 2.6

Proposition A.7. *Consider two probability distributions P and Q that satisfy (**CC**) and have compact supports contained in the ball $\overline{B(0,R)}$ for some $R > 0$. Let $f_u, f_v \in \mathcal{C}$ be the unique Kantorovich potentials for (P_u, Q_u) and (P_v, Q_v), respectively. Then there exists a constant $C_{R,p}$ depending on R and p such that*

$$\|f_u - f_v\|_\infty \leq C_{R,p}\|u - v\|_2^{p-1}. \tag{16}$$

Moreover, if P and Q are discrete probability distributions on $\{x_1, \ldots, x_N\}, \{y_1, \ldots, y_N\} \subset \overline{B(0,R)}$ respectively such that $P(x_i) = Q(y_i) = 1/N$ for $i = 1, \ldots, N$, then the inequality above also holds.

Proof. By the representation of one-dimensional Wasserstein costs [see 1, Theorem 2.10], we have for any $u \in \mathbb{S}^{d-1}$,

$$W_p^p(P_u, Q_u) = \int_0^1 |P_u^{-1}(t) - Q_u^{-1}(t)|^p \,\mathrm{d}t\,,$$

where, by abuse of notation, P_u^{-1} and Q_u^{-1} denote the inverses of the cumulative distribution functions of P_u and Q_u, i.e.,

$$P_u^{-1}(t) = \inf\{x \in \mathbb{R} : \mathbb{P}_P\{X^\top u \le x\} \ge t\}\,,$$

and analogously for Q_u^{-1}. Under (**CC**), note that these inverses satisfy

$$\begin{aligned} P_u^{-1}(t) \le x &\iff t \le P_u(x) \\ P_u^{-1}(t) \ge x &\iff t \ge P_u(x)\,, \end{aligned}$$

and likewise for Q_u (this follows from the considerations in [9, Section 2.1] combined with the fact that P_u^{-1} is a right inverse for P_u since the support of P_u is connected).

It follows from [9, Theorem 1.17] that the derivative of any optimal Kantorovich potential f_u must satisfy

$$f_u'(x) = p|x - Q_u^{-1} \circ P_u(x)|^{p-2}(x - Q_u^{-1} \circ P_u(x))\,.$$

Note that $|x - Q_u^{-1} \circ P_u(x)| \le 2R$, so that this expression is bounded by $p(2R)^{p-1}$. Therefore, if we define

$$f_u(x) = \int_0^x p|x' - Q_u^{-1} \circ P_u(x')|^{p-2}(x' - Q_u^{-1} \circ P_u(x'))\,\mathrm{d}x'\,,$$

then f_u is $p(2R)^{p-1}$ Lipschitz, satisfies $f_u(0) = 0$, and is a Kantorovich potential, and under (**CC**), it must therefore be the unique optimal potential in $\mathcal{C}$.

If we define $g_u(x') = |x' - Q_u^{-1} \circ P_u(x')|^{p-2}(x' - Q_u^{-1} \circ P_u(x'))$, it follows that

$$\begin{aligned} \|f_u - f_v\|_\infty &= \max_{x \in [-R,R]} \left| \int_0^x p(g_u(x') - g_v(x'))\,\mathrm{d}x' \right| \\ &\le p \int_{-R}^{R} |g_u(x') - g_v(x')|\,\mathrm{d}x'\,. \end{aligned}$$

The function $v \mapsto |v|^{p-2}v$ is $p-1$-Hölder, with norm depending on R and p [see, e.g 5, proof of Corollary 3]. Letting $C_{R,p}$ denote a constant depending on R and p whose value may vary from line to line, we obtain

$$\begin{aligned} \|f_u - f_v\|_\infty &\le C_{R,p} \int_{-R}^{R} |Q_u^{-1} \circ P_u(x') - Q_v^{-1} \circ P_v(x')|^{p-1}\,\mathrm{d}x' \\ &\quad C_{R,p} \int_{-R}^{R} |Q_u^{-1} \circ P_u(x') - Q_v^{-1} \circ P_u(x')|^{p-1} + |Q_v^{-1} \circ P_u(x') - Q_v^{-1} \circ P_v(x')|^{p-1}\,\mathrm{d}x' \\ &\quad \le C_{R,p}(\|Q_u^{-1} \circ P_u - Q_v^{-1} \circ P_u\|_\infty^{p-1} + \int_{-R}^{R} |Q_v^{-1} \circ P_u(x') - Q_v^{-1} \circ P_v(x')|^{p-1}\,\mathrm{d}x') \end{aligned}$$

For the first term, it suffices to note that $\|Q_u^{-1} - Q_v^{-1}\|_\infty$ is bounded. Indeed, [1, equation (2.3)] implies

$$\begin{aligned} \|Q_u^{-1} - Q_v^{-1}\|_\infty &= W_\infty(Q_u, Q_v) \\ &\le \|Y^\top u - Y^\top v\|_{L^\infty(Q)} \\ &\le R\|u - v\|_2\,, \end{aligned}$$

where the first inequality follows from the fact that $(Y^\top u, Y^\top v)$ with $Y \sim Q$ is a valid coupling of Q_u and Q_v. Therefore $\|Q_u^{-1} \circ P_u - Q_v^{-1} \circ P_u\|_\infty \le R\|u - v\|_2$.

For the second term, we first derive an upper bound for the case $p = 2$. The idea is borrowed from the proof of [9, Proposition 2.17]. Through computations, we have

$$\begin{aligned}
&\int_{-R}^{R} |Q_v^{-1} \circ P_u(x') - Q_v^{-1} \circ P_v(x')| \,\mathrm{d}x' \\
&= \mathcal{L}^2 \left(\{ (x', y) \in [-R, R] \times [-R, R] : Q_v^{-1} \circ P_u(x') \le y < Q_v^{-1} \circ P_v(x') \right. \\
&\qquad\qquad \left. \text{or } Q_v^{-1} \circ P_v(x') \le y < Q_v^{-1} \circ P_u(x') \} \right) \\
&= \mathcal{L}^2 \left(\{ (x', y) \in [-R, R] \times [-R, R] : Q_v^{-1} \circ P_u(x') \le y < Q_v^{-1} \circ P_v(x') \} \right) \\
&\quad + \mathcal{L}^2 \left(\{ (x', y) \in [-R, R] \times [-R, R] : Q_v^{-1} \circ P_v(x') \le y < Q_v^{-1} \circ P_u(x') \} \right) .
\end{aligned}$$

By Fubini's theorem along with the monotonicity of cumulative distribution functions, we have

$$\begin{aligned}
&\mathcal{L}^2 \left(\{ (x', y) \in [-R, R] \times [-R, R] : Q_v^{-1} \circ P_u(x') \le y < Q_v^{-1} \circ P_v(x') \} \right) \\
&= \int_{-R}^{R} \mathcal{L}^1 \left(\{ x' \in [-R, R] : P_v^{-1} \circ Q_v(y) < x' \le P_u^{-1} \circ Q_v(y) \} \right) \,\mathrm{d}y .
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\mathcal{L}^2 \left(\{ (x', y) \in [-R, R] \times [-R, R] : Q_v^{-1} \circ P_v(x') \le y < Q_v^{-1} \circ P_u(x') \} \right) \\
&= \int_{-R}^{R} \mathcal{L}^1 \left(\{ x' \in [-R, R] : P_u^{-1} \circ Q_v(y) < x' \le P_v^{-1} \circ Q_v(y) \} \right) \,\mathrm{d}y .
\end{aligned}$$

Summing up the integrals, we obtain

$$\begin{aligned}
\int_{-R}^{R} |Q_v^{-1} \circ P_u(x') - Q_v^{-1} \circ P_v(x')| \,\mathrm{d}x' &= \int_{-R}^{R} \left| P_u^{-1} \circ Q_v(y) - P_v^{-1} \circ Q_v(y) \right| \,\mathrm{d}y \\
&\le 2R \| P_u^{-1} - P_v^{-1} \|_\infty \le 2R^2 \|u - v\|_2 .
\end{aligned}$$

When $p > 2$, we have

$$\begin{aligned}
\int_{-R}^{R} |Q_v^{-1} \circ P_u(x') - Q_v^{-1} \circ P_v(x')|^{p-1} \,\mathrm{d}x' &\le (2R)^{p-2} \int_{-R}^{R} |Q_v^{-1} \circ P_u(x') - Q_v^{-1} \circ P_v(x')| \,\mathrm{d}x' \\
&\le (2R)^p \|u - v\|_2 .
\end{aligned}$$

Finally, when $1 < p < 2$, again we consider x' as a random variable of the uniform distribution U' on $[-R, R]$, and then by Jensen's inequality

$$\begin{aligned}
\int_{-R}^{R} |Q_v^{-1} \circ P_u(x') - Q_v^{-1} \circ P_v(x')|^{p-1} \,\mathrm{d}x' &= 2R \frac{1}{2R} \int_{-R}^{R} |Q_v^{-1} \circ P_u(x') - Q_v^{-1} \circ P_v(x')|^{p-1} \,\mathrm{d}x' \\
&\le 2R \left(\frac{1}{2R} \int_{-R}^{R} |Q_v^{-1} \circ P_u(x') - Q_v^{-1} \circ P_v(x')| \,\mathrm{d}x' \right)^{p-1} \\
&\le 2R (R \|u - v\|_2)^{p-1} \le 2R^p \|u - v\|_2^{p-1} .
\end{aligned}$$

Hence, combining the upper bounds for the first and second term, we obtain for $p > 1$,

$$\|f_u - f_v\|_\infty \le C_{R,p} (\|u - v\|_2^{p-1} + \|u - v\|_2^{(p-1) \vee 1}) \le C_{R,p} \|u - v\|_2^{p-1}.$$

Now we turn to consider the discrete distributions. In this case, the inverse P_u^{-1} only satisfies

$$\begin{aligned}
P_u^{-1}(t) \le x &\iff t \le P_u(x) \\
P_u^{-1}(t) > x &\iff t > P_u(x) ,
\end{aligned}$$

and likewise for Q_u^{-1}.

Fix any $y \in [-R, R]$, we have

$$\begin{aligned}
&\mathcal{L}^1 \left(\{ x' \in [-R, R] : P_u(x') \le Q_v(y) \text{ and } P_v(x') > Q_v(y) \} \right) \\
&= \mathcal{L}^1 \left(\{ x' \in [-R, R] : P_u(x') < Q_v(y) \text{ and } P_v(x') > Q_v(y) \} \right) \\
&\quad + \mathcal{L}^1 \left(\{ x' \in [-R, R] : P_u(x') = Q_v(y) \text{ and } P_v(x') > Q_v(y) \} \right) \\
&\le \mathcal{L}^1 \left(\{ x' \in [-R, R] : P_v^{-1} \circ Q_v(y) < x' < P_u^{-1} \circ Q_v(y) \} \right) \\
&\quad + \mathcal{L}^1 \left(\{ x' \in [-R, R] : P_u^{-1} \circ Q_v(y) \le x < P_u^{-1} \circ (Q_v(y) + 1/N) \} \right) \\
&= \mathcal{L}^1 \left(\{ x' \in [-R, R] : P_v^{-1} \circ Q_v(y) < x' < P_u^{-1} \circ (Q_v(y) + 1/N) \} \right) .
\end{aligned}$$

Analogously,

$$\begin{aligned}
&\mathcal{L}^1\left(\{x' \in [-R,R] : P_v(x') \le Q_v(y) \text{ and } P_u(x') > Q_v(y)\}\right) \\
&\le \mathcal{L}^1\left(\left\{x' \in [-R,R] : P_u^{-1} \circ Q_v(y) < x' < P_v^{-1} \circ (Q_v(y) + 1/N)\right\}\right) .
\end{aligned}$$

In addition, the two sets have no intersection. Therefore,

$$\begin{aligned}
&\int_{-R}^{R} \left|Q_v^{-1} \circ P_u(x') - Q_v^{-1} \circ P_v(x')\right| \mathrm{d}x' \\
&\le \int_{-R}^{R} \left|P_u^{-1} \circ (Q_v(y) + 1/N) - P_v^{-1} \circ Q_v(y) + P_u^{-1} \circ Q_v(y) - P_v^{-1} \circ (Q_v(y) + 1/N)\right| \mathrm{d}y \\
&\le \int_{-R}^{R} \left|P_u^{-1} \circ (Q_v(y) + 1/N) - P_v^{-1} \circ (Q_v(y) + 1/N)\right| + \left|P_u^{-1} \circ Q_v(y) - P_v^{-1} \circ Q_v(y)\right| \mathrm{d}y \\
&\le 4R\|P_u^{-1} - P_v^{-1}\|_\infty \le 4R^2\|u - v\|_2 .
\end{aligned}$$

□

Proof of Theorem 2.6. We may assume without loss of generality that $m = n$ by discarding addition samples from either P or Q, if necessary. We define an estimator for $u, v \in \mathbb{S}^{d-1}$ by setting

$$\begin{aligned}
\hat{\Sigma}_{u,v} =&(1-\lambda) \int f_{nu}(u^\top x) f_{nv}(v^\top x) \,\mathrm{d}P_n(x) \\
&- (1-\lambda)\left(\int f_{nu}(u^\top x)\,\mathrm{d}P_n(x)\right)\left(\int f_{nv}(v^\top x)\,\mathrm{d}P_n(x)\right) \\
&+ \lambda \int f_{nu}^c(u^\top y) f_{nv}^c(v^\top y)\,\mathrm{d}Q_n(y) \\
&- \lambda\left(\int f_{nu}^c(u^\top y)\,\mathrm{d}Q_n(y)\right)\left(\int f_{nv}^c(v^\top y)\,\mathrm{d}Q_n(y)\right) ,
\end{aligned} \tag{17}$$

where $f_{n\cdot} \in \mathcal{C}$ denotes a Kantorovich potential for $P_{n\cdot}$ and $Q_{n\cdot}$. We are done if we can show the convergence of the first term to the corresponding one in (7). The proof for the other three terms follows similar routine.

We first split the absolute difference between the objective quantities into two terms:

$$\begin{aligned}
&\sup_{u,v \in \mathbb{S}^{d-1}} \left|\int f_{nu}(u^\top x) f_{nv}(v^\top x)\,\mathrm{d}P_n(x) - \int f_u(u^\top x) f_v(v^\top x)\,\mathrm{d}P(x)\right| \\
&\le \frac{1}{n} \sup_{u,v \in \mathbb{S}^{d-1}} \left|\sum_{i=1}^n f_{nu}(u^\top X_i) f_{nv}(v^\top X_i) - f_u(u^\top X_i) f_v(v^\top X_i)\right| \\
&\quad + \sup_{u,v \in \mathbb{S}^{d-1}} \left|\frac{1}{n}\sum_{i=1}^n \left(f_u(u^\top X_i) f_v(v^\top X_i) - \mathbb{E} f_u(u^\top X_i) f_v(v^\top X_i)\right)\right| .
\end{aligned}$$

For the second term, we notice that $P_n - P \rightsquigarrow 0$ in $\ell^\infty(\mathcal{C} \times \mathbb{S}^{d-1})$ and $H : \ell^\infty(\mathcal{C} \times \mathbb{S}^{d-1}) \to \mathbb{R}$ defined by $H(g) = \sup_{u,v} |g(f_u, u) g(f_v, v)|$ is continuous. Therefore, the second term converges to zero in distribution and consequently in probability. Since all f_{nu}, f_u, $u \in \mathbb{S}^{d-1}$ are uniformly bounded by some constant depending only on R and p, dominated convergence theorem yields convergence in mean.

For the first term, we have

$$\begin{aligned}
&\mathbb{E} \sup_{u,v \in \mathbb{S}^{d-1}} \left|P_n\left((f_{nu} \circ h_u)(f_{nv} \circ h_v) - (f_u \circ h_u)(f_v \circ h_v)\right)\right| \\
&\le \mathbb{E} \sup_{u,v \in \mathbb{S}^{d-1}} \left|f_{nu}(u^\top X) f_{nv}(v^\top X) - f_u(u^\top X) f_v(v^\top X)\right| \\
&\le C_{R,p} \mathbb{E} \sup_{u \in \mathbb{S}^{d-1}} \left|f_{nu}(u^\top X) - f_u(u^\top X)\right| .
\end{aligned}$$

Proposition A.7 implies that

$$\begin{aligned}
&\left\|\left|f_{nu}(u^\top x) - f_u(u^\top x)\right| - \left|f_{nv}(v^\top x) - f_v(v^\top x)\right|\right\|_\infty \\
&\le \|(f_{nu}(u^\top x) - f_u(u^\top x)) - (f_{nv}(v^\top x) - f_v(v^\top x))\|_\infty \\
&\le C_{R,p}\|u - v\|_2^{p-1}. \qquad (18)
\end{aligned}$$

Therefore, for any $\varepsilon > 0$, there exists some partition of $\mathbb{S}^{d-1}$ such that for $n \in \mathbb{N}^*$ large enough,

$$\begin{aligned}
&\mathbb{S}^{d-1} \subseteq \bigcup_{i=1}^{N_\varepsilon} B(u_i, \delta_\varepsilon), \\
&\left\|\left|f_{nu}(u^\top x) - f_u(u^\top x)\right| - \left|f_{nu_i}({u_i}^\top x) - f_{u_i}({u_i}^\top x)\right|\right\|_\infty < \frac{\varepsilon}{2},\ i = 1, \dots, N_\varepsilon, \\
&\mathbb{E} \sup_{i=1,\dots,N_\varepsilon} \left|f_{nu}(u^\top X) - f_u(u^\top X)\right| < \frac{\varepsilon}{2}.
\end{aligned}$$

The last inequality follows from the the P-a.s. convergence of Kantorovich potentials by [3, Theorem 2.8] with dominated convergence theorem applied to it and and the finiteness of terms taken over supremum.

Altogether, the first term may be bounded by arbitrarily small numbers when $n \to \infty$:

$$\mathbb{E} \sup_{u\in\mathbb{S}^{d-1}} \left|f_{nu}(u^\top X) - f_u(u^\top X)\right| \le \mathbb{E} \sup_{i=1,\dots,N_\varepsilon} \left|f_{nu}(u^\top X) - f_u(u^\top X)\right| + \frac{\varepsilon}{2} < \varepsilon.$$

This completes the deduction of convergence in mean of the estimator (17). $\square$

A.3 Proof of Theorem 3.1

Proof. We define $H : \ell^\infty(\mathbb{S}^{d-1}) \to \mathbb{R}$ as $H(f) := \int_{\mathbb{S}^{d-1}} f(u)\, d\sigma(u)$. $W_2^2(\mu_\cdot, \nu_\cdot)$ indeed belongs to $\ell^\infty(\mathbb{S}^{d-1})$ since the Wasserstein distance between any one-dimensional projections of probability distributions μ and ν is bounded above by the one between μ and ν themselves. Besides, the integral over unit sphere with respect to uniform measure preserves the sup norm of the functions in $\ell^\infty(\mathbb{S}^{d-1})$.

By definition of weak convergence in $\ell^\infty(\mathbb{S}^{d-1})$, the uniform CLT implies that

$$\sqrt{n}(SW_p^p(P_n, Q_n) - SW_p^p(P, Q)) = \int_{\mathbb{S}^{d-1}} \sqrt{n}(W_p^p(P_{nu}, Q_{nu}) - W_p^p(P_u, Q_u))\, d\sigma(u) \rightsquigarrow S.$$

For any ω in the probability space Ω, $\int_{\mathbb{S}^{d-1}} |\mathbb{G}_u(\omega)|\, d\sigma(u) < \infty$. This can be easily deduced from the fact that $\mathbb{G}$ has continuous sample paths a.s. In addition, $\mathbb{G} : \mathbb{S}^{d-1} \times \Omega \to \mathbb{R}$ is jointly measurable and thus $\omega \mapsto S(\omega)$ is a random variable. Finally, S is Gaussian due to the Riemann integrability of $u \mapsto \mathbb{G}_u(\omega)$.

Finally we compute the mean and variance of S. Trivially, $\mathbb{E}S = 0$. In terms of the variance, we have

$$\begin{aligned}
\mathrm{Var}(S) &= \mathbb{E}\left(\int_{\mathbb{S}^{d-1}} \mathbb{G}(u)\,\mathrm{d}\sigma(u)\right)^2 = \int_{\mathbb{S}^{d-1}}\int_{\mathbb{S}^{d-1}} \mathbb{E}(\mathbb{G}(u)\mathbb{G}(v))\,\mathrm{d}\sigma(u)\,\mathrm{d}\sigma(v) \\
&= \int \left(\int_{\mathbb{S}^{d-1}} f_u(u^\top x)\,\mathrm{d}\sigma(u)\right)^2 \mathrm{d}P(x) - \left(\int (\int_{\mathbb{S}^{d-1}} f_u(u^\top x)\,\mathrm{d}\sigma(u))\,\mathrm{d}P(x)\right)^2 \\
&+ \int \left(\int_{\mathbb{S}^{d-1}} f_u^c(u^\top y)\,\mathrm{d}\sigma(u)\right)^2 \mathrm{d}Q(x) - \left(\int (\int_{\mathbb{S}^{d-1}} f_u^c(u^\top y)\,\mathrm{d}\sigma(u))\,\mathrm{d}Q(y)\right)^2 \\
&= \mathrm{Var}_{X\sim P}\left(\int_{\mathbb{S}^{d-1}} f_u(u^\top X)\,\mathrm{d}\sigma(u)\right) + \mathrm{Var}_{Y\sim Q}\left(\int_{\mathbb{S}^{d-1}} f_u^c(u^\top Y)\,\mathrm{d}\sigma(u)\right). \qquad (19)
\end{aligned}$$

$\square$

Remark A.8. *The variance of S is identical with that of $Z_{(P,Q)}$ derived in Theorem 3 and Lemma 8 of [6] with $\delta = 0$. The variance of $Z_{(P,Q)}$ can be reduced to*

$$\int_{\mathbb{S}^{d-1}}\int_{\mathbb{S}^{d-1}} \left(\int_0^1 |P_u^{-1}(t) - Q_u^{-1}(t)|^p |P_v^{-1}(t) - Q_v^{-1}(t)|^p\,\mathrm{d}t - W_p^p(P_u, Q_u) W_p^p(P_v, Q_v)\right) \mathrm{d}\sigma(u)\,\mathrm{d}\sigma(v).$$

Let π_u and π_v denote the optimal transport plans between P_u, Q_u and P_v, Q_v respectively. Letting $(X_u, Y_u) \sim \pi_u$ and $(X_v, Y_v) \sim \pi_v$, it follows that the expression above is equal to

$$\begin{aligned}
&\int_{\mathbb{S}^{d-1}} \int_{\mathbb{S}^{d-1}} \mathit{cov}\ \left(|X_u - Y_u|^p, |X_v - Y_v|^2\right) \mathrm{d}\sigma(u)\,\mathrm{d}\sigma(v) \\
&= \mathit{Var}\left(\int_{\mathbb{S}^{d-1}} |X_u - Y_u|^p \,\mathrm{d}\sigma(\theta)\right) \\
&= \mathit{Var}\left(\int_{\mathbb{S}^{d-1}} f_u(u^\top X) + f_u^c(u^\top Y)\,\mathrm{d}\sigma(u)\right) \\
&= \mathit{Var}_{X\sim P}\left(\int_{\mathbb{S}^{d-1}} f_u(u^\top X)\,\mathrm{d}\sigma(u)\right) + \mathit{Var}_{Y\sim Q}\left(\int_{\mathbb{S}^{d-1}} f_u^c(u^\top Y)\,\mathrm{d}\sigma(u)\right).
\end{aligned}$$

B Additional Experiments

We present more simulation results in this section, including those with $p = 1$. As mentioned in the main text, the only obstacle to including $p = 1$ is that there do not seem to be general conditions under which the Kantorovich potentials under $p = 1$ are unique. We will show; however, if this fact can be verified by other means in specific cases, then the central limit theorem still holds.

B.1 Sliced Wasserstein Distance.

Consider the example in section 4.1. Instead of $p = 2$, we investigate the asymptotic behavior of the case $p = 1$. We first give an explicit representation of the theoretical limit of the example given in section 4.1. Then the unique 1-Lipschitz function that achieves the 1-Wasserstein distance between P_θ and Q_θ is $\phi_0^\theta(x) = -\mathrm{sign}(a_\theta)x$. Hence, we have

$$\sqrt{n}\,(W_1(P_{n}.,Q_{n}.) - W_1(P.,Q.)) \rightsquigarrow \mathbb{G},$$

where $\mathbb{G}$ is the mean-zero Gaussian process indexed by $\mathbb{S}^2$ with covariance functions

$$\mathbb{E}\mathbb{G}(u)\mathbb{G}(v) = \frac{2}{3}\mathrm{sign}(a_u)\mathrm{sign}(a_v)\langle u, v\rangle.$$

It follows from Theorem 3.1 that the limiting distribution of the empirical 1-Wasserstein distance is the centered Gaussian S with variance

$$\mathrm{Var}(S) = \frac{2}{3}\int_{\mathbb{S}^2}\int_{\mathbb{S}^2} \mathrm{sign}(a_u)\mathrm{sign}(a_v)\langle u, v\rangle\, d\sigma^3(u) d\sigma^3(v) \approx 0.164.$$

We sample i.i.d. observations $X_1, \ldots, X_n \sim P$ and $Y_1, \ldots, Y_n \sim Q$ with size $n = 50, 100, 500$. This process is repeated 500 times. We then compare the finite distributions of 1-Wasserstein distance with the theoretical limit given in section 3.1. We demonstrate the results using kernel density estimators in Figure 1. We see that the finite-sample empirical distribution gets closer to the limiting Gaussian distribution in 9 as the sample size n increases.

In addition, we simulate the re-scaled plug-in bootstrap approximations by sampling $n = 1000$ observations of P and Q. Fix some empirical SW $\sqrt{n}SW_2^2(P_n, Q_n)$, we generate $B = 500$ replications of $\sqrt{l}(SW_1(\hat{P}_n^*, \hat{Q}_n^*) - SW_1(P_n, Q_n))$. The distributions of the replications with various replacement numbers l, compared with the finite-sample empirical distribution and the theoretical limit, are shown in Figure 2. We observe that the naive bootstrap ($l = n$) better approximates the finite sample distribution compared to fewer replacements ($l = n^{1/2}, n^{3/4}$). This is consistent with the observation of inference on finite spaces. [7]

B.2 Max-Sliced Wasserstein Distance

Consider the example in section 4.2. Again we estimate the distributional limit of the empirical distributions of 1-Wasserstein distance but with $a = 2$. The unique 1-Lipschitz function that achieves 1-Wasserstein distance between P_{e_1} and Q_{e_1} or equivalently P_{-e_1} and Q_{-e_1} is $\phi_0^{e_1}(x) = -|x|$. Consequently, the theoretical limit in this case is the mean-zero Gaussian with variance

$$\mathrm{Var}(\mathbb{G}_{\pm e_1}) = \frac{1}{2}\int_{-1}^{1} x^2\,dx - \left(\frac{1}{2}\int_{-1}^{1} -|x|\,dx\right)^2 + \frac{1}{4}\int_{-2}^{2} y^2\,dy - \left(\frac{1}{4}\int_{-2}^{2} |y|\,dy\right)^2 = \frac{5}{12}.$$

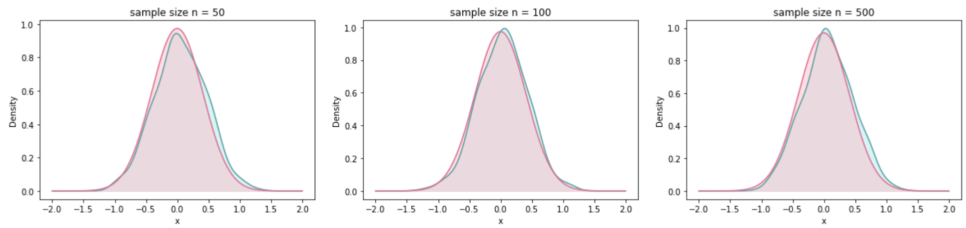

Figure 1: Comparison of the finite sample density (pale turquoise) and the limit distribution of the empirical sliced distance (pink).

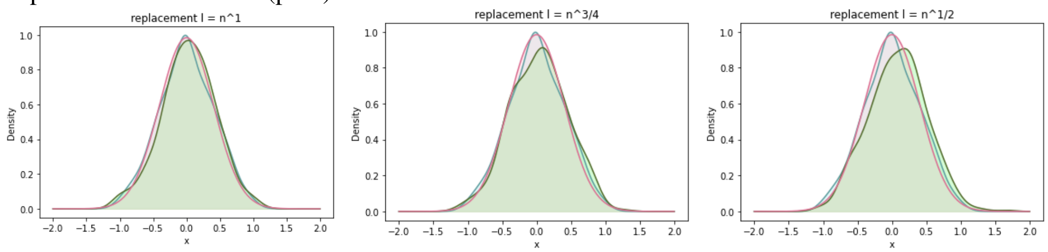

Figure 2: Bootstrap for the empirical sliced distance. Illustration of the re-scaled plug-in bootstrap approximation ($n = 1000$) with replacement $l \in \{n, n^{3/4}, n^{1/2}\}$. Finite bootstrap densities (pale green) are compared to the corresponding finite sample density (pale turquoise) and the limit distribution (pink).

The plots of comparison between the theoretical limit and the finite sample distributions of $n = 100, 500, 1000$ each of which is repeated 1000 times are given in Figure 3. The simulation of bootstrap is plotted in Figure 4. The naive bootstrap ($l = n$) better approximates the finite sample distribution compared to fewer replacements ($l = n^{1/2}, n^{3/4}$).

In order to give some sense of non-Gaussian limiting distributions, i.e. when the directions that maximize the Wasserstein distance of 1-dimensional projections are not unique, we give an example of such cases. Let P be the uniform distribution over $\mathbb{S}^2$ and Q uniform over $2\mathbb{S}^2$. The plots of comparison between the theoretical limit and the finite sample distributions of $n = 1000, 5000, 10000$ each of which is repeated 5000 times are given in Figure 5. The simulation of bootstrap with $B = 500$ replications is plotted in Figure 6. We see that the overall performance is worse than the cases when the limits are Gaussian. The replacement $l = 3/4$ has the closest approximation comparatively.

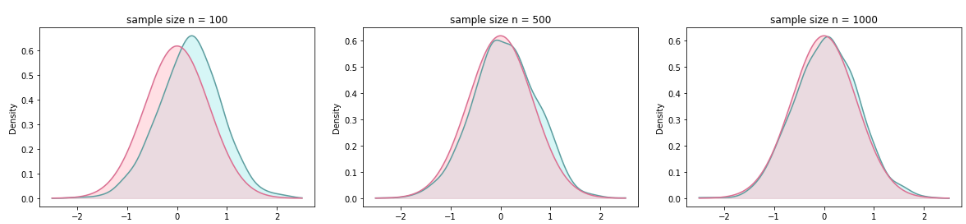

Figure 3: Comparison of the finite sample density (pale turquoise) and the limit distribution of the empirical max-sliced Wasserstein distance (pink).

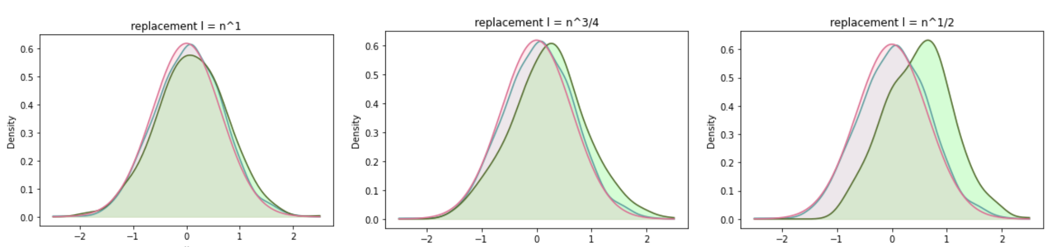

Figure 4: Bootstrap for the empirical max-sliced Wasserstein distance. Illustration of the re-scaled plug-in bootstrap approximation ($n = 1000$) with replacement $l \in \{n, n^{3/4}, n^{1/2}\}$. Finite bootstrap densities (pale green) are compared to the corresponding finite sample density (pale turquoise) and the limit distribution (pink).

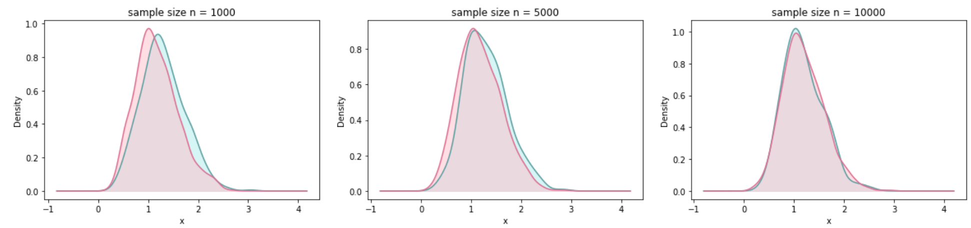

Figure 5: Comparison of the finite sample density (pale turquoise) and the limit distribution of the empirical max-sliced Wasserstein distance (pink).

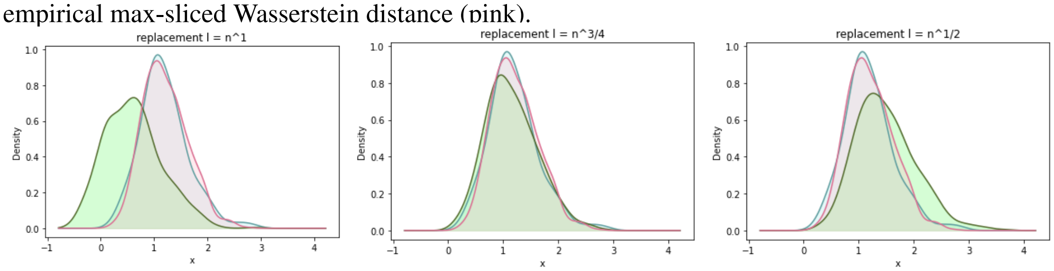

Figure 6: Bootstrap for the empirical one-dimensional WPP. Illustration of the re-scaled plug-in bootstrap approximation ($n = 10000$) with replacement $l \in \{n, n^{3/4}, n^{1/2}\}$. Finite bootstrap densities (pale green) are compared to the corresponding finite sample density (pale turquoise) and the limit distribution (pink).

B.3 Distributional Sliced Wasserstein Distance

In this section, we present a simple example for DSW. Consider two distributions P which is uniform on the surface of the ellipsoid $\{x^2/4 + 4y^2 + z^2 = 1\}$ and Q the uniform distribution on $\mathbb{S}^2$. Let $\mathcal{P}_C$ be a set of 10 two-point probability measures on $\mathbb{S}^2$. Explicitly, for each $\tau \in \mathcal{P}_C$, $\tau \sim \frac{1}{3}\delta_u + \frac{2}{3}\delta_v$ for some $u, v \in \mathbb{S}^2$. One of the measures takes $u = (1, 0, 0)$ and $v = (-1, 0, 0)$. The discussion in Section 3.3 yields that

$$\sqrt{n}\left(DSW_2^2(P_n, Q_n) - DSW_2^2(P, Q)\right) \to \frac{1}{3}\mathbb{G}((1,0,0)) + \frac{2}{3}\mathbb{G}((-1,0,0)) = \mathbb{G}((1,0,0)).$$

We sample i.i.d. observations $X_1, \ldots, X_n \sim P$ and $Y_1, \ldots, Y_n \sim Q$ with size $n = 100, 500, 1000$. This process is repeated 1000 times. The plots of comparison between the theoretical limit and the finite sample distributions are given in Figure 7.

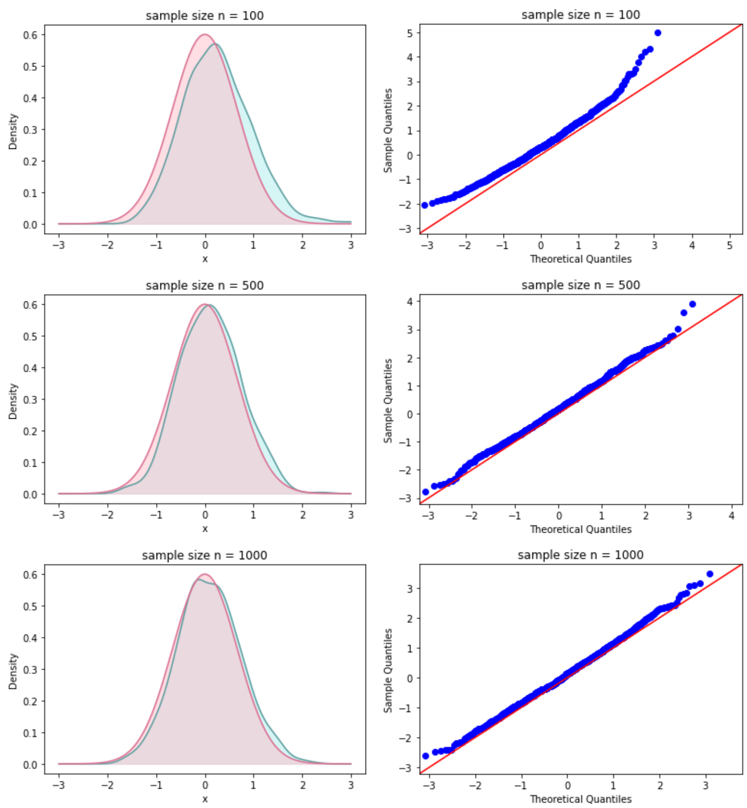

Figure 7: Comparison of the finite sample density (pale turquoise) and the limit distribution of the empirical distributional sliced Wasserstein distance (pink).

References

[1] S. G. Bobkov and M. Ledoux. One-dimensional empirical measures, order statistics, and kantorovich transport distances. *Memoirs of the American Mathematical Society*, 2019.

[2] J. C'arcamo, L.-A. Rodr'iguez, and A. Cuevas. Directional differentiability for supremum-type functionals: Statistical applications. *Bernoulli*, 2020.

[3] E. Del Barrio and J.-M. Loubes. Central limit theorems for empirical transportation cost in general dimension. *The Annals of Probability*, 47(2):926–951, 2019.

[4] O. Dovgoshey and O. Martio. Products of metric spaces, covering numbers, packing numbers and characterizations of ultrametric spaces. *arXiv: Metric Geometry*, 2009.

[5] T. Manole and J. Niles-Weed. Sharp convergence rates for empirical optimal transport with smooth costs. *arXiv preprint arXiv:2106.13181*, 2021.

[6] T. Manole, S. Balakrishnan, and L. Wasserman. Minimax confidence intervals for the sliced wasserstein distance. *arXiv: Statistics Theory*, 2019.

[7] R. Okano and M. Imaizumi. Inference for projection-based wasserstein distances on finite spaces. *arXiv preprint arXiv:2202.05495*, 2022.

[8] W. Rudin. *Principles of Mathematical Analysis*. McGraw - Hill Book C., 1986. URL `https://books.google.com/books?id=frdNAQAACAAJ`.

[9] F. Santambrogio. *Optimal transport for applied mathematicians*, volume 87 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser/Springer, Cham, 2015. ISBN 978-3-319-20827-5; 978-3-319-20828-2. doi: 10.1007/978-3-319-20828-2. URL `https://doi.org/10.1007/978-3-319-20828-2`. Calculus of variations, PDEs, and modeling.

[10] A. S. Shapiro. On concepts of directional differentiability. *Journal of optimization theory and applications*, 66(3):477–487, 1990.

[11] A. van der Vaart and J. A. Wellner. Weak convergence and empirical processes: With applications to statistics. 1996.

[12] J. A. Wellner. Empirical processes: Theory and applications. 2005.