Gaussian-Smoothed Sliced Probability Divergences: Supplementary Materials

Mokhtar Z. Alaya *alayaelm@utc.fr*
Université de Technologie de Compiègne,
LMAC (Laboratoire de Mathématiques Appliquées de Compiègne), CS 60 319 - 60 203 Compiègne Cedex

Alain Rakotomamonjy *a.rakotomamonjy@criteo.com*
Criteo AI Lab, Paris, France,

Maxime Berar *maxime.berar@univ-rouen.fr*
Univ Rouen Normandie, INSA Rouen Normandie, Universite Le Havre Normandie
Normandie Univ, LITIS UR4108, Rouen, France

Gilles Gasso *gilles.gasso@insa-rouen.fr*
INSA Rouen Normandie, Univ Rouen Normandie, Universite Le Havre Normandie,
Normandie Univ, LITIS UR4108, Rouen, France

Reviewed on OpenReview: *https://openreview.net/forum?id=weuALLWUV2*

A Proofs

In the following sections, we give the proofs of the theoretical guarantees given in the main of the paper.

A.1 Proof of Theorem 3.1: $\mathrm{G}_\sigma\mathrm{SD}_p$ is a proper metric on $\mathcal{P}_p(\mathbb{R}^d) \times \mathcal{P}_p(\mathbb{R}^d)$

Before starting the proof, we add this notation: the characteristic function of a probability distribution $\mu \in \mathcal{P}(\mathbb{R}^d)$ is $\varphi_\mu(t) = \mathbf{E}_\mu[e^{iX^\top t}]$. Given this definition, similarly to the Fourier transform, the characteristic function of the convolution of two probability distributions readsas $\varphi_{\nu*\mu}(t) = \varphi_\nu(t) \cdot \varphi_\mu(t)$.

• *Non-negativity (or symmetry).* The non-negativity (or symmetry) follows directly from the non-negativity (or symmetry) of D^p, see Definition 2.3.
• *Identity property.* If the base divergence D^p satisfies the identity property in one dimensional measures, then for any $\mu \in \mathcal{P}_p(\mathbb{R}^d)$ and $\mathbf{u} \in \mathbb{S}^{d-1}$, one has that $\mathrm{D}_p(\mathcal{R}_\mathbf{u}\mu * \mathcal{N}_\sigma, \mathcal{R}_\mathbf{u}\mu * \mathcal{N}_\sigma) = 0$, hence, by Definition 2.3, $\mathrm{G}_\sigma\mathrm{SD}_p(\mu, \mu) = 0$. Let us now prove the fact that for any $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d), \mathrm{G}_\sigma\mathrm{SD}^p(\mu, \nu) = 0$ entails $\mu = \nu$ a.s. On one hand, $\mathrm{G}_\sigma\mathrm{SD}_p(\mu, \nu) = 0$ gives the fact that $\mathrm{D}_p(\mathcal{R}_\mathbf{u}\mu * \mathcal{N}_\sigma, \mathcal{R}_\mathbf{u}\nu * \mathcal{N}_\sigma) = 0$ for u_d-almost every $\mathbf{u} \in \mathbb{S}^{d-1}$, hence $\mathcal{R}_\mathbf{u}\mu * \mathcal{N}_\sigma = \mathcal{R}_\mathbf{u}\nu * \mathcal{N}_\sigma$ for u_d-almost every $\mathbf{u} \in \mathbb{S}^{d-1}$. Following the techniques in proof of Proposition 5.1.2 in Bonnotte (2013), for any measure $\eta \in \mathcal{P}(\mathbb{R}^m)$ (with $m \geq 1$), $\mathcal{F}[\eta](\cdot)$ stands for the Fourier transform of η and is given as $\mathcal{F}[\eta](\mathbf{v}) = \int_{\mathbb{R}^m} e^{-i\mathbf{s}^\top\mathbf{v}} \mathrm{d}\eta(\mathbf{s})$ for any $\mathbf{v} \in \mathbb{R}^m$. Then

$$
\begin{aligned}
\mathcal{F}[\mathcal{R}_\mathbf{u}\mu * \mathcal{N}_\sigma](v) &= \int_\mathbb{R} e^{-ivt} \mathrm{d}(\mathcal{R}_\mathbf{u}\mu * \mathcal{N}_\sigma)(t) \\
&= \int_\mathbb{R}\int_\mathbb{R} e^{-i(r+t)v} \mathrm{d}\mathcal{R}_\mathbf{u}\mu(r)\mathrm{d}\mathcal{N}_\sigma(t) \quad \text{(by the definition of the convolution operator)} \\
&= \int_{\mathbb{R}^d}\int_\mathbb{R} e^{-i(\langle \mathbf{u}, \mathbf{s}\rangle + t)v} \mathrm{d}\mu(\mathbf{s})\mathrm{d}\mathcal{N}_\sigma(t) \quad \text{(by the definition of Radon Transform)} \\
&= \int_\mathbb{R} e^{-itv}\mathrm{d}\mathcal{N}_\sigma(t) \int_{\mathbb{R}^d} e^{-i(\langle \mathbf{u}, s\rangle)v} \mathrm{d}\mu(\mathbf{s}) \\
&= \mathcal{F}[\mathcal{N}_\sigma](v)\mathcal{F}[\mu](v\mathbf{u}).
\end{aligned}
$$

Since for u_d-almost every $\mathbf{u} \in \mathbb{S}^{d-1}, \mathcal{R}_\mathbf{u}\mu * \mathcal{N}_\sigma = \mathcal{R}_\mathbf{u}\nu * \mathcal{N}_\sigma$, and hence $\mathcal{F}[\mathcal{R}_\mathbf{u}\mu * \mathcal{N}_\sigma] = \mathcal{F}[\mathcal{R}_\mathbf{u}\nu * \mathcal{N}_\sigma] \Leftrightarrow \mathcal{F}[\mathcal{N}_\sigma]\mathcal{F}[\mu] = \mathcal{F}[\mathcal{N}_\sigma]\mathcal{F}[\nu]$ (by the Fourier transform of the convolution) $\Leftrightarrow \mathcal{F}[\mu] = \mathcal{F}[\nu]$. Since the Fourier

transform is injective, we conclude that $\mu = \nu$.
• *Triangle inequality.* Assume that D is a metric and let $\mu, \nu, \eta \in \mathcal{P}_p(\mathbb{R}^d)$. We then have

$$\begin{aligned}
\mathrm{G}_\sigma\mathrm{SD}_p(\mu,\nu) &= \Big(\int_{\mathbb{S}^{d-1}} \mathrm{D}^p(\mathcal{R}_{\mathbf{u}}\mu * \mathcal{N}_\sigma, \mathcal{R}_{\mathbf{u}}\nu * \mathcal{N}_\sigma) u_d(\mathbf{u})\mathrm{d}\mathbf{u}\Big)^{1/p} \\
&\leq \Big(\int_{\mathbb{S}^{d-1}} \big(\mathrm{D}(\mathcal{R}_{\mathbf{u}}\mu * \mathcal{N}_\sigma, \mathcal{R}_{\mathbf{u}}\eta * \mathcal{N}_\sigma) + \mathrm{D}(\mathcal{R}_{\mathbf{u}}\eta * \mathcal{N}_\sigma, \mathcal{R}_{\mathbf{u}}\nu * \mathcal{N}_\sigma)\big)^p u_d(\mathbf{u})\mathrm{d}\mathbf{u}\Big)^{1/p} \\
&\underbrace{\leq}_{(\star)} \Big(\int_{\mathbb{S}^{d-1}} \big(\mathrm{D}^p(\mathcal{R}_{\mathbf{u}}\mu * \mathcal{N}_\sigma, \mathcal{R}_{\mathbf{u}}\eta * \mathcal{N}_\sigma) u_d(\mathbf{u})\mathrm{d}\mathbf{u}\Big)^{1/p} + \Big(\int_{\mathbb{S}^{d-1}} \mathrm{D}^p(\mathcal{R}_{\mathbf{u}}\eta * \mathcal{N}_\sigma, \mathcal{R}_{\mathbf{u}}\nu * \mathcal{N}_\sigma)\Big)^p u_d(\mathbf{u})\mathrm{d}\mathbf{u}\Big)^{1/p} \\
&= \mathrm{G}_\sigma\mathrm{SD}_p(\mu,\eta) + \mathrm{G}_\sigma\mathrm{SD}_p(\eta,\nu),
\end{aligned}$$

where inequality in $(\star)$ follows from the application of Minkowski inequality.

A.2 Proof of Theorem 3.2: $\mathrm{G}_\sigma\mathrm{SD}_p$ metrizes the weak topology

The proof is done by double implications and the technical material relies on the continuous mapping theorem (Athreya & Lahiri, 2006) and bounded convergence theorem for the first direct implication "$\Rightarrow$". The second one, "$\Leftarrow$", is based on the fact that weak convergence is equivalent to the convergence corresponding to Lévy-Prokhorov distance (Huber, 2011)

"$\Rightarrow$" Assume that $\mu_k \Rightarrow \mu$. Fix $\mathbf{u} \in \mathbb{S}^{d-1}$, the mapping $\mathbf{u} \mapsto \mathcal{R}_{\mathbf{u}}$ is continuous from $\mathbb{R}^d$ to $\mathbb{R}$, then an application of continuous mapping theorem (Athreya & Lahiri, 2006) entails that $\mathcal{R}_{\mathbf{u}}\mu_k \Rightarrow \mathcal{R}_{\mathbf{u}}\mu$. By Lévy's continuity theorem (Athreya & Lahiri, 2006) $\mathcal{R}_{\mathbf{u}}\mu_k * \mathcal{N}_\sigma \Rightarrow \mathcal{R}_{\mathbf{u}}\mu * \mathcal{N}_\sigma$. Therefore, $\lim_{k\to\infty} \mathrm{D}(\mathcal{R}_{\mathbf{u}}\mu_k, \mathcal{R}_{\mathbf{u}}\mu * \mathcal{N}_\sigma) = 0$. Since we suppose that the divergence D is bounded, then there exists $K \geq 0$ such that for any k, $\mathrm{D}^p(\mathcal{R}_{\mathbf{u}}\mu_k, \mathcal{R}_{\mathbf{u}}\mu * \mathcal{N}_\sigma) \leq K$. An application of bounded convergence theorem yields

$$\lim_{k\to\infty} \mathrm{G}_\sigma\mathrm{SD}_p(\mu_k,\mu) = \Big(\int_{\mathbb{S}^{d-1}} \lim_{k\to\infty} \mathrm{D}^p(\mathcal{R}_{\mathbf{u}}\mu_k * \mathcal{N}_\sigma, \mathcal{R}_{\mathbf{u}}\mu * \mathcal{N}_\sigma) u_d(\mathbf{u})\mathrm{d}\mathbf{u}\Big)^{1/p} = 0.$$

"$\Leftarrow$" (By contrapositive). Suppose that μ_k doesn't converge weakly to μ and assume that $\lim_{k\to\infty} \mathrm{G}_\sigma\mathrm{SD}^p(\mu_k,\mu) = 0$. On one hand, since $\mathbb{R}^d$ is a complete separable space then the weak convergence is equivalent to the convergence corresponding to Lévy-Prokhorov distance Λ defined as: The Lévy-Prokhorov distance $\Lambda(\eta,\zeta)$ between $\eta,\zeta \in \mathscr{P}((E,\rho),\mathcal{T})$ (space of probability measures on a measurable metric space) is given by:

$$\Lambda(\eta,\zeta) = \inf_{\varepsilon>0}\{\eta(A) < \zeta(A^\varepsilon) + \varepsilon, \quad \zeta(A) < \eta(A^\varepsilon) + \varepsilon, \quad \text{for all } A \in \mathcal{T}\}, \text{ where } A^\varepsilon = \{x \in E : \rho(x,A) < \varepsilon\}.$$

Hence there exists $\varepsilon > 0$ and a subsequence $\{\mu_{s(k)}\}_{k\in\mathbb{N}}$ such that $\Lambda(\mu_{s(k)},\mu) > \varepsilon$. One the other hand, we have $\lim_{k\to\infty} \mathrm{G}_\sigma\mathrm{SD}^p(\mu_{s(k)},\mu) = 0$, that is equivalent to $\{\mathrm{D}(\mathcal{R}_{\mathbf{u}}\mu_{s(k)} * \mathcal{N}_\sigma, \mathcal{R}_{\mathbf{u}}\nu * \mathcal{N}_\sigma)\}_k$ converges to 0 in $L^p(\mathbb{S}^{d-1}) = \{f : \mathbb{S}^{d-1} \to \mathbb{R} | \int_{\mathbb{S}^{d-1}} f(\mathbf{u})u_d(\mathbf{u})\mathrm{d}u < \infty\}$. Since the L^p-convergence entails the point-wise convergence (Khoshnevisan, 2007), there exists a subsequence $\{\mu_{s(t(k))}\}_k$ such that $\lim_{k\to\infty} \mathrm{D}(\mathcal{R}_{\mathbf{u}}\mu_{s(t(k))} * \mathcal{N}_\sigma, \mathcal{R}_{\mathbf{u}}\mu * \mathcal{N}_\sigma) = 0$ almost everywhere for all $\mathbf{u} \in \mathbb{S}^{d-1}$. Recall that the divergence D metrizes the weak convergence in $\mathcal{P}(\mathbb{R})$ then $\mathcal{R}_{\mathbf{u}}\mu_{s(t(k))} * \mathcal{N}_\sigma \Rightarrow \mathcal{R}_{\mathbf{u}}\mu * \mathcal{N}_\sigma$ almost everywhere for all $\mathbf{u} \in \mathbb{S}^{d-1}$. Therefore, $\mathcal{R}_{\mathbf{u}}\mu_{s(t(k))} \Rightarrow \mathcal{R}_{\mathbf{u}}\mu$ almost everywhere for all $\mathbf{u} \in \mathbb{S}^{d-1}$. Using Cramér-Wold device (Huber, 2011), we get $\mu_{s(t(k))} \Rightarrow \mu$. Since the Lévy-Prokhorov distance metrizes the weak convergence, it entails that $\lim_{k\to\infty} \Lambda(\mu_{s(t(k))},\mu_k) = 0$, that contradicts the fact that $\Lambda(\mu_{s(k)},\mu) > \varepsilon$. We then conclude by contrapositive that $\mu_k \Rightarrow \mu$.

A.3 Proof of Proposition 3.3: $\mathrm{G}_\sigma\mathrm{SD}_p$ is lower semi-continuous

Recall that the base divergence D is lower semi-continuous w.r.t. the weak topology in $\mathcal{P}(\mathbb{R})$, namely for every sequence of measures $\{\mu'_k\}_{k\in\mathbb{N}}$ and $\{\nu'_k\}_{k\in\mathbb{N}}$ in $\mathcal{P}(\mathbb{R})$ such that $\mu'_k \Rightarrow \mu'$ and $\nu'_k \Rightarrow \nu'$, one has $\mathrm{D}(\mu',\nu') \leq \liminf_{k\to\infty} \mathrm{D}(\mu'_k,\nu'_k)$.

Now, let $\{\mu_k\}_{k\in\mathbb{N}}$ and $\{\nu_k\}_{k\in\mathbb{N}}$ are two sequences of measure in $\mathcal{P}_p(\mathbb{R}^d)$ such that $\mu_k \Rightarrow \mu$ and $\nu_k \Rightarrow \nu$.

By continuous mapping theorem (Bowers & Kalton, 2014) and Levy's continuity theorem, we obtain $\mathcal{R}_{\mathbf{u}}\mu_k * \mathcal{N}_\sigma \Rightarrow \mathcal{R}_{\mathbf{u}}\mu * \mathcal{N}_\sigma$ and $\mathcal{R}_{\mathbf{u}}\nu_k * \mathcal{N}_\sigma \Rightarrow \mathcal{R}_{\mathbf{u}}\nu * \mathcal{N}_\sigma$ for all $\mathbf{u} \in \mathbb{S}^{d-1}$. Since the base divergence D is a lower semi-continuous with respect to weak topology in $\mathcal{P}(\mathbb{R})$, then

$$\mathrm{D}^p(\mathcal{R}_{\mathbf{u}}\mu * \mathcal{N}_\sigma, \mathcal{R}_{\mathbf{u}}\nu * \mathcal{N}_\sigma) \le \big(\liminf_{k\to\infty} \mathrm{D}(\mathcal{R}_{\mathbf{u}}\mu_k * \mathcal{N}_\sigma, \mathcal{R}_{\mathbf{u}}\nu_k * \mathcal{N}_\sigma)\big)^p \le \liminf_{k\to\infty} \mathrm{D}^p(\mathcal{R}_{\mathbf{u}}\mu_k * \mathcal{N}_\sigma, \mathcal{R}_{\mathbf{u}}\nu_k * \mathcal{N}_\sigma).$$

It gives

$$\mathrm{G}_\sigma\mathrm{SD}_p(\mu,\nu) \le \Big(\int_{\mathbb{S}^{d-1}} \liminf_{k\to\infty} \mathrm{D}^p(\mathcal{R}_{\mathbf{u}}\mu_k * \mathcal{N}_\sigma, \mathcal{R}_{\mathbf{u}}\nu_k * \mathcal{N}_\sigma) u_d(\mathbf{u})\mathrm{d}\mathbf{u}\Big)^{1/p}.$$

Furthermore, by application of Fatou's lemma (Bowers & Kalton, 2014), we get

$$\mathrm{G}_\sigma\mathrm{SD}_p(\mu,\nu) \le \liminf_{k\to\infty} \Big(\int_{\mathbb{S}^{d-1}} \mathrm{D}^p(\mathcal{R}_{\mathbf{u}}\mu_k * \mathcal{N}_\sigma, \mathcal{R}_{\mathbf{u}}\nu_k * \mathcal{N}_\sigma) u_d(\mathbf{u})\mathrm{d}\mathbf{u}\Big)^{1/p} = \liminf_{k\to\infty} \mathrm{G}_\sigma\mathrm{SD}_p(\mu_k,\nu_k),$$

which is the desired result.

A.4 Proofs of statistical properties

A.4.1 Proof of Lemma 3.5: $\mathcal{R}_{\mathbf{u}}\hat{\mu}_n * \mathcal{N}_\sigma$ is an average of Gaussian mixture

Straighforwardly, for every Borelian $I \in \mathcal{B}(\mathbb{R})$, we have

$$\begin{aligned}
\mathcal{R}_{\mathbf{u}}\hat{\mu}_n * \mathcal{N}_\sigma(I) &= \int_r \int_s \mathbf{1}_I(r+s)\mathrm{d}\{\frac{1}{n}\sum_{i=1}^n \delta_{\mathbf{u}^\top X_i}\}(r)\mathrm{d}\mathcal{N}_\sigma(s) \\
&= \frac{1}{n}\sum_{i=1}^n \int_s \mathbf{1}_I(\mathbf{u}^\top X_i + s) f_{\mathcal{N}_\sigma}(s)\mathrm{d}s \\
&= \frac{1}{n}\sum_{i=1}^n \int_{s'} \mathbf{1}_I(s') f_{\mathcal{N}_\sigma}(s' - \mathbf{u}^\top X_i)\mathrm{d}s' \\
&= \frac{1}{n}\sum_{i=1}^n \int_{s'} \mathbf{1}_I(s') f_{\mathcal{N}(\mathbf{u}^\top X_i, \sigma^2)}(s')\mathrm{d}s' \quad (\text{since } f_{\mathcal{N}_\sigma}(s' - \mathbf{u}^\top X_i) = f_{\mathcal{N}(\mathbf{u}^\top X_i,\sigma^2)}(s')) \\
&= \frac{1}{n}\sum_{i=1}^n \mathcal{N}(\mathbf{u}^\top X_i, \sigma^2)(I).
\end{aligned}$$

Thanks to Theorem of Cramér and Wold (Cramér & Wold, 1936), we conclude the equality between the measures $\mathcal{R}_{\mathbf{u}}\hat{\mu}_n * \mathcal{N}_\sigma = \frac{1}{n}\sum_{i=1}^n \mathcal{N}(\mathbf{u}^\top X_i, \sigma^2)$.

A.4.2 Proof of Proposition 3.8

Let us give first the overall structure of the proof. We we use frequently the triangle inequality for Wasserstein distances between the quantities $\hat{\hat{\mu}}_n$, $\frac{1}{n}\mathcal{N}_\sigma(u^\top X_i, \sigma^2)$ and $\mathcal{R}_{\mathbf{u}}\mu * \mathcal{N}_\sigma$. We then obtain two quantities, **I** and **II** (see below for explicit), bounding $\mathbf{E}_{\mu^{\otimes n}|\mathcal{N}_\sigma^{\otimes n}}[\hat{\mathrm{G}}_\sigma\mathrm{SW}_p(\hat{\mu}_n,\mu)]$. To control **I** bound, we use a well known converging bound in Fournier & Guillin (2015) of Wasserstein distance between empirical and true measure. For **II** bound, we consider maximal TV-coupling in Villani (2009)] and use result of the $2p$-moment of absolute Gaussian random variable founded in Winkelbauer (2014).

On one hand, using triangle inequality of Wasserstein distance, we have

$$\begin{aligned}
\mathbf{E}_{\mu^{\otimes n}|\mathcal{N}_\sigma^{\otimes n}}[\hat{\mathrm{G}}_\sigma\mathrm{SW}_p(\hat{\mu}_n,\mu)] &= \mathbf{E}_{\mu^{\otimes n}|\mathcal{N}_\sigma^{\otimes n}}\Big[\Big(\int_{\mathbb{S}^{d-1}} \mathrm{W}_p^p(\hat{\hat{\mu}}_n, R_{\mathbf{u}}\mu * \mathcal{N}_\sigma)u_d(\mathbf{u})\mathrm{d}\mathbf{u}\Big)^{1/p}\Big] \\
&\le \Big(\mathbf{E}_{\mu^{\otimes n}|\mathcal{N}_\sigma^{\otimes n}}\Big[\int_{\mathbb{S}^{d-1}} \mathrm{W}_p^p(\hat{\hat{\mu}}_n, R_{\mathbf{u}}\mu * \mathcal{N}_\sigma)u_d(\mathbf{u})\mathrm{d}\mathbf{u}\Big]\Big)^{1/p} \\
&\le \Big(\int_{\mathbb{S}^{d-1}} \mathbf{E}_{\mu^{\otimes n}|\mathcal{N}_\sigma^{\otimes n}}[\mathrm{W}_p^p(\hat{\hat{\mu}}_n, R_{\mathbf{u}}\mu * \mathcal{N}_\sigma)]u_d(\mathbf{u})\mathrm{d}\mathbf{u}\Big)^{1/p} \\
&\le (\mathbf{I} + \mathbf{II})^{1/p}
\end{aligned}$$

where

$$\mathbf{I} \triangleq 2^{p-1} \int_{\mathbb{S}^{d-1}} \mathbf{E}_{\mu^{\otimes n} | \mathcal{N}_\sigma^{\otimes n}} \Big[\mathrm{W}_p^p \Big(\hat{\hat{\mu}}_n, \frac{1}{n} \sum_{i=1}^n \mathcal{N}(\mathbf{u}^\top X_i, \sigma^2) \Big) \Big] u_d(\mathbf{u}) \mathrm{d}\mathbf{u}$$

and

$$\mathbf{II} \triangleq 2^{p-1} \int_{\mathbb{S}^{d-1}} \mathbf{E}_{\mu^{\otimes n} | \mathcal{N}_\sigma^{\otimes n}} \big[\mathrm{W}_p^p \big(\frac{1}{n} \sum_{i=1}^n \mathcal{N}(\mathbf{u}^\top X_i, \sigma^2), R_{\mathbf{u}}\mu * \mathcal{N}_\sigma) \big) \big] u_d(\mathbf{u}) \mathrm{d}\mathbf{u}$$

The proof is based on two steps to control the quantities **I** and **II**.
Step 1: Control of **I**.
Let us state the following lemma:

Lemma A.1 (See proof of Theorem 1 in Fournier & Guillin (2015)). *Let $\eta \in \mathcal{P}(\mathbb{R})$ and let $p \geq 1$. Assume that $M_q(\eta) < \infty$ for some $q > p$. There exists a constant $C_{p,q}$ depending only on p, q such that, for all $n \geq 1$,*

$$\mathbf{E}[\mathrm{W}_p^p(\hat{\eta}_n, \eta)] \leq C_{p,q} M_q(\eta)^{p/q} \Delta_n(p, q),$$

where

$$\Delta_n(p,q) = \begin{cases} n^{-1/2} \mathbf{1}_{q > 2p}, \\ n^{-1/2} \log(n) \mathbf{1}_{q = 2p} \\ n^{-(q-p)/q} \mathbf{1}_{p < q < 2p}. \end{cases}.$$

We note that $\hat{\hat{\mu}}_n$ is an empirical version of the Gausian mixture $\frac{1}{n} \sum_{i=1}^n \mathcal{N}_\sigma(u^\top X_i, \sigma^2)$. Then, by application of Lemma A.1, we get

$$\mathbf{E}_{\mu^{\otimes n} | \mathcal{N}_\sigma^{\otimes n}} \big[\mathrm{W}_p^p \big(\hat{\hat{\mu}}_n, \frac{1}{n} \sum_{i=1}^n \mathcal{N}(\mathbf{u}^\top X_i, \sigma^2)) \big] \leq C_{p,q} \mathbf{E}_{\mu^{\otimes n}} \Big[M_q^{p/q} \Big(\frac{1}{n} \sum_{i=1}^n \mathcal{N}(\mathbf{u}^\top X_i, \sigma^2) \Big) \Big] \Delta_n(p, q).$$

Let us first upper bound the q-th moment of $M_q\Big(\frac{1}{n}\sum_{i=1}^n \mathcal{N}(\mathbf{u}^\top X_i, \sigma^2)\Big)$, for all $q \geq 1$. For all $\mathbf{u} \in \mathbb{S}^{d-1}$, we have

$$M_q\Big(\frac{1}{n} \sum_{i=1}^n \mathcal{N}(\mathbf{u}^\top X_i, \sigma^2) \Big) = \int_{\mathbb{R}} |t|^q \mathrm{d}\big(\frac{1}{n} \sum_{i=1}^n \mathcal{N}(\mathbf{u}^\top X_i, \sigma^2) \big)(t) = \frac{1}{n} \sum_{i=1}^n M_q(|Z_{i,\mathbf{u}}|^q),$$

where $Z_{i,\mathbf{u}} \sim \mathcal{N}(\mathbf{u}^\top X_i, \sigma^2))$. By Equation (17) in Winkelbauer (2014) we have

$$M_q\Big(\frac{1}{n} \sum_{i=1}^n \mathcal{N}(\mathbf{u}^\top X_i, \sigma^2) \Big) = \frac{1}{n} \frac{2^{q/2}\sigma^q}{\sqrt{\pi}} \Gamma(\frac{q+1}{2}) \sum_{i=1}^n {}_1F_1\big(-\frac{q}{2}, \frac{1}{2}; \frac{-(\mathbf{u}^\top X_i)^2}{2\sigma^2} \big).$$

Since $X_1, \ldots, X_n$ are i.i.d samples from μ, it yields

$$\begin{aligned}
\mathbf{E}_{\mu^{\otimes n}} \Big[M_q^{p/q} \Big(\frac{1}{n} \sum_{i=1}^n \mathcal{N}(\mathbf{u}^\top X_i, \sigma^2) \Big) \Big] &= \frac{2^{q/2}\sigma^q}{\sqrt{\pi}} \Gamma(\frac{q+1}{2}) \mathbf{E}_\mu \big[{}_1F_1\big(-\frac{q}{2}, \frac{1}{2}; \frac{-(\mathbf{u}^\top X)^2}{2\sigma^2} \big) \big] \quad (X \sim \mu) \\
&= \frac{2^{q/2}\sigma^q}{\sqrt{\pi}} \Gamma(\frac{q+1}{2}) \sum_{k=0}^\infty \frac{(-\frac{q}{2})_k}{(\frac{1}{2})_k} \frac{(-1)^k}{(2\sigma^2)^k k!} \mathbf{E}_\mu[(\mathbf{u}^\top X)^{2k}] \\
&\leq \frac{2^{q/2}\sigma^q}{\sqrt{\pi}} \Gamma(\frac{q+1}{2}) \sum_{k=0}^\infty \frac{(-\frac{q}{2})_k}{(\frac{1}{2})_k} \frac{(-1)^k}{(2\sigma^2)^k k!} M_{2k}(\mu).
\end{aligned}$$

Setting $q = 2p$ we have $\Delta_n(p,q) = \frac{\log n}{n}$, then

$$\mathbf{I} \leq 2^{2p-1} C_p \frac{\sigma^{2p}}{\sqrt{\pi}} \Gamma(\frac{2p+1}{2}) \sum_{k=0}^\infty \frac{(-p)_k}{(\frac{1}{2})_k} \frac{(-1)^k}{(2\sigma^2)^k k!} M_{2k}(\mu) \frac{\log(n)}{n}.$$

Step 2: Control of **II**.

We follow the lines of proofs of Proposition 1 in Goldfeld et al. (2020) and Theorem 2 in Nietert et al. (2021). Using a coupling $\hat{\hat{\mu}}_n$ and $\mathcal{R}_{\mathbf{u}}\mu * \mathcal{N}_\sigma)$ via the maximal TV-coupling (see Theorem 6.15 in Villani (2009)]), the control of the total variation of the Wasserstein distance, we get for any fixed $\mathbf{u} \in \mathbb{S}^{d-1}$

$$\mathrm{W}_p^p\big(\frac{1}{n}\sum_{i=1}^n \mathcal{N}(\mathbf{u}^\top X_i, \sigma^2), R_{\mathbf{u}}\mu * \mathcal{N}_\sigma)\big) \le 2^{p-1}\int_{\mathbb{R}} |t|^p |h_{n,\mathbf{u}}(t) - g_{\mathbf{u}}(t)| dt,$$

where $h_{n,\mathbf{u}}$ and $g_{\mathbf{u}}$ are the densities associated with μ_n and $\mathcal{R}_{\mathbf{u}}\mu * \mathcal{N}_\sigma$, respectively. Let $f_{\sigma,\vartheta}$ the probability density function of $\mathcal{N}_{\sigma,\vartheta}$, i.e, $f_{\sigma,\vartheta}(t) = \frac{1}{\sqrt{2\pi(\sigma\vartheta)^2}} e^{-\frac{t^2}{2(\sigma\vartheta)^2}}$ for $\vartheta > 0$ to be specified later. An application of Cauchy-Schwarz inequality gives

$$\begin{aligned}
&\mathbf{E}_{\mu^{\otimes n}|\mathcal{N}_\sigma^{\otimes n}}\Big[\mathrm{W}_p^p\big(\frac{1}{n}\sum_{i=1}^n \mathcal{N}(\mathbf{u}^\top X_i, \sigma^2), R_{\mathbf{u}}\mu * \mathcal{N}_\sigma)\big)\Big] \\
&\le 2^{p-1}\mathbf{E}_{\mu^{\otimes n}|\mathcal{N}_\sigma^{\otimes n}} \int_{\mathbb{R}} |t|^p \sqrt{f_{\sigma,\vartheta}(t)} \frac{|h_{n,\mathbf{u}}(t) - g_{\mathbf{u}}(t)|}{\sqrt{f_{\sigma,\vartheta}(t)}} dt \\
&\le 2^{p-1}\mathbf{E}_{\mu^{\otimes n}|\mathcal{N}_\sigma^{\otimes n}} \Big(\int_{\mathbb{R}} |t|^{2p} f_{\sigma,\vartheta}(t) dt\Big)^{\frac{1}{2}} \Big(\int_{\mathbb{R}} \frac{(h_{n,\mathbf{u}}(t) - g_{\mathbf{u}}(t))^2}{f_{\sigma,\vartheta}(t)} dt\Big)^{\frac{1}{2}} \\
&\le 2^{p-1}\Big(\int_{\mathbb{R}} |t|^{2p} f_{\sigma,\vartheta}(t) dt\Big)^{\frac{1}{2}} \Big(\int_{\mathbb{R}} \mathbf{E}_{\mu^{\otimes n}|\mathcal{N}_\sigma^{\otimes n}} \frac{(h_{n,\mathbf{u}}(t) - g_{\mathbf{u}}(t))^2}{f_{\sigma,\vartheta}(t)} dt\Big)^{\frac{1}{2}}.
\end{aligned}$$

Note that $\int_{\mathbb{R}} |t|^{2p} f_{\sigma,\vartheta}(t) dt$ is the $2p$-th moment of $|\mathcal{N}_{\sigma,\vartheta}(t)|$ equals to (see Equation (18) in Winkelbauer (2014))

$$\int_{\mathbb{R}} |t|^{2p} f_{\sigma,\vartheta}(t) dt = \frac{(\sigma\vartheta)^{2p} 2^p}{\sqrt{\pi}} \Gamma\big(\frac{2p+1}{2}\big).$$

Moreover,

$$h_{n,\mathbf{u}}(t) = \frac{1}{n}\sum_{i=1}^n \mathrm{d}\mathcal{N}(\mathbf{u}^\top X_i, \sigma^2)(t) = \frac{1}{n}\sum_{i=1}^n f_{\sigma,\vartheta}(t - \mathbf{u}^\top X_i),$$

It is clear to see that $h_{n,\mathbf{u}}(t)$ is a sum of i.i.d. terms with expectation $g_{\mathbf{u}}(t)$, which implies

$$\begin{aligned}
\mathbf{E}_{\mu^{\otimes n}|\mathcal{N}_\sigma^{\otimes n}}\big[(h_{n,\mathbf{u}}(t) - g_{\mathbf{u}}(t))^2\big] &= \mathbf{V}_{\mu^{\otimes n}}\Big[\frac{1}{n}\sum_{i=1}^n f_{\sigma,\vartheta}(t - \mathbf{u}^\top X_i)\Big] \\
&= \frac{1}{n}\mathbf{V}_\mu[f_{\sigma,\vartheta}(t - \mathbf{u}^\top X] \\
&\le \frac{1}{n}\mathbf{E}_\mu[(f_{\sigma,\vartheta}(t - \mathbf{u}^\top X)^2] \\
&\le \frac{(2\pi\sigma^2)^{-1}}{n}\mathbf{E}_\mu[e^{\frac{-1}{\sigma^2}(t - \mathbf{u}^\top X)^2}].
\end{aligned}$$

Now

$$\mathbf{E}_\mu[e^{\frac{-(t - \mathbf{u}^\top X)^2}{\sigma^2}}] = \int_{\|x\| \le \frac{|t|}{2}} e^{\frac{-1}{\sigma^2}(t - \mathbf{u}^\top x)^2} \mathrm{d}\mu(x) + \int_{\|x\| > \frac{|t|}{2}} e^{\frac{-1}{\sigma^2}(t - \mathbf{u}^\top x)^2} \mathrm{d}\mu(x).$$

Remark that when $\|x\| \le \frac{|t|}{2}$, then $(t - \mathbf{u}^\top X)^2 \ge |t|^2 - |\mathbf{u}^\top x|^2 \ge |t|^2 - \|x\|^2$ (since $\|u\|^2 = 1$). We get $(t - \mathbf{u}^\top X)^2 \ge \frac{|t|^2}{4}$ and hence

$$\int_{\|x\| \le \frac{|t|}{2}} e^{\frac{-1}{\sigma^2}(t - \mathbf{u}^\top x)^2} \mathrm{d}\mu(x) \le e^{\frac{-t^2}{4\sigma^2}} \text{ and } \int_{\|x\| > \frac{|t|}{2}} e^{\frac{-1}{\sigma^2}(t - \mathbf{u}^\top x)^2} \mathrm{d}\mu(x) \le \mathbf{P}\big[\|X\| > \frac{|t|}{2}\big]$$

This gives,

$$\int_{\mathbb{R}} \mathbf{E}_{\mu^{\otimes n}|\mathcal{N}_\sigma^{\otimes n}} \frac{(h_{n,\mathbf{u}}(t) - g_{\mathbf{u}}(t))^2}{f_{\sigma,\vartheta}(t)} \mathrm{d}t \leq \frac{(2\pi\sigma^2)^{-1}(\sqrt{2\pi}\sigma\vartheta)}{n} \Big(\int_{\mathbb{R}} e^{\frac{t^2}{2(\sigma\vartheta)^2}} e^{\frac{-t^2}{4\sigma^2}} \mathrm{d}t + \int_{\mathbb{R}} e^{\frac{t^2}{2(\sigma\vartheta)^2}} \mathbf{P}\big[\|X\| > \frac{|t|}{2}\big] \mathrm{d}t \Big).$$

Note that the integral $\int_{\mathbb{R}} e^{\frac{t^2}{2(\sigma\vartheta)^2}} e^{\frac{-t^2}{4\sigma^2}} \mathrm{d}t = \int_{\mathbb{R}} e^{-\left(\frac{1}{2} - \frac{1}{\vartheta^2}\right)\frac{t^2}{2\sigma^2}} \mathrm{d}t$ is finite if and only if $\frac{1}{2} - \frac{1}{\vartheta^2} > 0$ namely $\vartheta > \sqrt{2}$ and its value is given by

$$\int_{\mathbb{R}} e^{\frac{t^2}{2(\sigma\vartheta)^2}} e^{\frac{-t^2}{4\sigma^2}} \mathrm{d}t = \sqrt{\frac{2\pi\sigma^2}{\frac{1}{2} - \frac{1}{\vartheta^2}}} = \sqrt{\frac{4\pi\sigma^2\vartheta^2}{\vartheta^2 - 2}}.$$

For the second integral

$$\int_{\mathbb{R}} e^{\frac{t^2}{2(\sigma\vartheta)^2}} \mathbf{P}\big[\|X\| > \frac{|t|}{2}\big] \mathrm{d}t = 2\int_0^\infty e^{\frac{t^2}{2(\sigma\vartheta)^2}} \mathbf{P}\big[\|X\| > \frac{t}{2}\big] \mathrm{d}t = 4\int_0^\infty e^{\frac{2\xi^2}{\sigma^2\vartheta^2}} \mathbf{P}\big[\|X\| > \xi\big] \mathrm{d}\xi$$

Then,

$$\mathbf{II} \leq n^{-1/2} 4^{p-1} \Big\{ (2\pi\sigma^2)^{-1}(\sqrt{2\pi}\sigma\vartheta) \frac{(\sigma\vartheta)^{2p} 2^p}{\sqrt{\pi}} \Gamma\big(\frac{2p+1}{2}\big) \Big\}^{\frac{1}{2}} \Big(\sqrt{\frac{4\pi\sigma^2\vartheta^2}{\vartheta^2 - 2}} + 4\int_0^\infty e^{\frac{2\xi^2}{\sigma^2\vartheta^2}} \mathbf{P}\big[\|X\| > \xi\big] \mathrm{d}\xi \Big)^{\frac{1}{2}}.$$

this gives the desired result using the fact that $(a+b)^{1/p} \leq a^{1/p} + b^{1/p}$, for $a, b \geq 0$.

A.4.3 Proof of Proposition 3.11

Using triangle inequality, we have

$$\mathrm{W}_p(\hat{\mu}_n, \hat{\nu}_n) \leq \mathrm{W}_p(\hat{\mu}_n, \mathcal{R}_{\mathbf{u}}\mu * \mathcal{N}_\sigma) + \mathrm{W}_p(\mathcal{R}_{\mathbf{u}}\mu * \mathcal{N}_\sigma, \mathcal{R}_{\mathbf{u}}\nu * \mathcal{N}_\sigma) + \mathrm{W}_p(\mathcal{R}_{\mathbf{u}}\nu * \mathcal{N}_\sigma, \hat{\nu}_n).$$

and then

$$\mathrm{W}_p^p(\hat{\mu}_n, \hat{\nu}_n) \leq 3^{p-1}\big\{ \mathrm{W}_p^p(\hat{\mu}_n, \mathcal{R}_{\mathbf{u}}\mu * \mathcal{N}_\sigma) + \mathrm{W}_p^p(\mathcal{R}_{\mathbf{u}}\mu * \mathcal{N}_\sigma, \mathcal{R}_{\mathbf{u}}\nu * \mathcal{N}_\sigma) + \mathrm{W}_p^p(\mathcal{R}_{\mathbf{u}}\nu * \mathcal{N}_\sigma, \hat{\nu}_n)\big\}.$$

This implies that

$$\begin{aligned} &\mathbf{E}_{\mu^{\otimes n}|\mathcal{N}_\sigma^{\otimes n}} \mathbf{E}_{\nu^{\otimes n}|\mathcal{N}_\sigma^{\otimes n}} [\hat{\mathrm{G}}_\sigma\mathrm{SW}_p(\hat{\mu}_n, \hat{\nu}_n)] \\ &\qquad \leq 3^{1-\frac{1}{p}} \mathrm{G}_\sigma\mathrm{SW}_p(\mu, \nu) + 3^{1-\frac{1}{p}} \mathbf{E}_{\mu^{\otimes n}|\mathcal{N}_\sigma^{\otimes n}} [\hat{\mathrm{G}}_\sigma\mathrm{SW}_p(\hat{\mu}_n, \mu)] + 3^{1-\frac{1}{p}} \mathbf{E}_{\nu^{\otimes n}|\mathcal{N}_\sigma^{\otimes n}} [\hat{\mathrm{G}}_\sigma\mathrm{SW}_p(\hat{\nu}_n, \nu)]. \end{aligned}$$

By application of Proposition 3.8, it yields This gives that

$$\mathbf{E}_{\mu^{\otimes n}|\mathcal{N}_\sigma^{\otimes n}} \mathbf{E}_{\nu^{\otimes n}|\mathcal{N}_\sigma^{\otimes n}} [\hat{\mathrm{G}}_\sigma\mathrm{SW}_p(\hat{\mu}_n, \hat{\nu}_n)] \leq 3^{1-\frac{1}{p}} \mathrm{G}_\sigma\mathrm{SW}_p(\mu, \nu) + 3\Xi_{p,\sigma,\vartheta} \frac{1}{n^{1/2p}} + 3^{1-\frac{1}{p}} (\Upsilon_{p,\sigma,\mu} + \Upsilon_{p,\sigma,\nu}) \frac{(\log n)^{1/p}}{n^{1/p}}$$

This ends the proof of the first statement in Proposition 3.11. For the second one, we also use a triangle inequality

$$\mathrm{W}_p^p(\mathcal{R}_{\mathbf{u}}\mu * \mathcal{N}_\sigma, \mathcal{R}_{\mathbf{u}}\nu * \mathcal{N}_\sigma) \leq 3^{p-1}\big\{ \mathrm{W}_p^p(\mathcal{R}_{\mathbf{u}}\mu * \mathcal{N}_\sigma, \hat{\mu}_n) + \mathrm{W}_p^p(\hat{\mu}_n, \hat{\nu}_n) + \mathrm{W}_p^p(\hat{\nu}_n), \mathcal{R}_{\mathbf{u}}\nu * \mathcal{N}_\sigma\big\}.$$

Then we control each term as we did before.

A.5 Proof of Proposition 3.12: projection complexity

Using Holder's inequality, we have

$$\begin{aligned} \mathbf{E}_{u_d^{\otimes L}} \big[\big|\widehat{\mathrm{G}_\sigma\mathrm{SD}_p}^p(\mu, \nu) - \mathrm{G}_\sigma\mathrm{SD}_p{}^p(\mu, \nu)\big|\big] &\leq \Big(\mathbf{E}_{u_d^{\otimes L}} \big[\big|\widehat{\mathrm{G}_\sigma\mathrm{SD}_p}^p(\mu, \nu) - \mathrm{G}_\sigma\mathrm{SD}_p^p(\mu, \nu)\big|^2\big] \Big)^{1/2} \\ &= \Big(\mathbf{V}_{u_d^{\otimes L}} \big[\big[\widehat{\mathrm{G}_\sigma\mathrm{SD}_p}^p(\mu, \nu)\big]\Big)^{1/2} \\ &= \frac{A(p, \sigma)}{L^{1/2}}. \end{aligned}$$

A.6 Proof of Corollary 3.13: overall complexity $(p=1)$

By application of triangle inequality, one has

$$|\widehat{\hat{\mathrm{G}}_\sigma\mathrm{SW}}(\hat{\mu}_n,\hat{\nu}_n)-\mathrm{G}_\sigma\mathrm{SW}(\mu,\nu)|\leq|\widehat{\hat{\mathrm{G}}_\sigma\mathrm{SW}}(\hat{\mu}_n,\hat{\nu}_n)-\hat{\mathrm{G}}_\sigma\mathrm{SW}(\hat{\mu}_n,\hat{\nu}_n)|+|\,\hat{\mathrm{G}}_\sigma\mathrm{SW}(\hat{\mu}_n,\hat{\nu}_n)-\mathrm{G}_\sigma\mathrm{SW}(\mu,\nu)|$$

Using Proposition 3.12, we have

$$\mathbf{E}_{u_d^{\otimes L}}\big[|\widehat{\hat{\mathrm{G}}_\sigma\mathrm{SW}}(\hat{\mu}_n,\hat{\nu}_n)-\hat{\mathrm{G}}_\sigma\mathrm{SW}(\hat{\mu}_n,\hat{\nu}_n)|\big]\leq\frac{\hat{A}_\sigma}{\sqrt{L}}:=\frac{\{\mathbf{V}_{\mathbf{u}\sim u_d}[\mathrm{W}(\hat{\hat{\mu}}_n,\hat{\hat{\nu}}_n)]\}^{1/2}}{\sqrt{L}}.$$

Using Proposition 3.11 for $p=1$ we get,

$$\mathbf{E}_{\mu^{\otimes n}|\mathcal{N}_\sigma^{\otimes n}}\mathbf{E}_{\nu^{\otimes n}|\mathcal{N}_\sigma^{\otimes n}}[|\hat{\mathrm{G}}_\sigma\mathrm{SW}(\hat{\mu}_n,\hat{\nu}_n)-\mathrm{G}_\sigma\mathrm{SW}(\mu,\nu)|]\leq 3\Xi_{1,\sigma,\vartheta}\frac{1}{\sqrt{n}}+(\Upsilon_{1,\sigma,\mu}+\Upsilon_{1,\sigma,\nu})\frac{\log n}{n}.$$

Therefore, by applying the expectations with respect to the projection and sampling we obtain

$$\begin{aligned}&\mathbf{E}_{u_d^{\otimes L}}\mathbf{E}_{\mu^{\otimes n}|\mathcal{N}_\sigma^{\otimes n}}\mathbf{E}_{\nu^{\otimes n}|\mathcal{N}_\sigma^{\otimes n}}\big[|\widehat{\hat{\mathrm{G}}_\sigma\mathrm{SW}}(\hat{\mu}_n,\hat{\nu}_n)-\mathrm{G}_\sigma\mathrm{SW}(\mu,\nu)|\big]\\&\qquad\leq\frac{1}{\sqrt{L}}\mathbf{E}_{\mu^{\otimes n}|\mathcal{N}_\sigma^{\otimes n}}\mathbf{E}_{\nu^{\otimes n}|\mathcal{N}_\sigma^{\otimes n}}[\hat{A}_\sigma]+3\Xi_{1,\sigma,\vartheta}\frac{1}{\sqrt{n}}+(\Upsilon_{1,\sigma,\mu}+\Upsilon_{1,\sigma,\nu})\frac{\log n}{n}.\end{aligned}$$

By Jensen inequality, we have

$$\mathbf{E}_{\mu^{\otimes n}|\mathcal{N}_\sigma^{\otimes n}}\mathbf{E}_{\nu^{\otimes n}|\mathcal{N}_\sigma^{\otimes n}}[\hat{A}_\sigma]\leq\big\{\mathbf{E}_{\mu^{\otimes n}|\mathcal{N}_\sigma^{\otimes n}}\mathbf{E}_{\nu^{\otimes n}|\mathcal{N}_\sigma^{\otimes n}}[\mathbf{V}_{\mathbf{u}\sim u_d}[\mathrm{W}(\hat{\hat{\mu}}_n,\hat{\hat{\nu}}_n)]]\big\}^{1/2}.$$

A.7 Proof of Proposition 3.14

For all $\mathbf{u}\in\mathbb{S}^{d-1}$ we have $\mathcal{R}_\mathbf{u}\mu,\mathcal{R}_\mathbf{u}\nu\in\mathcal{P}(\mathbb{R})$. By application of the inequality of noise level satisfied by D in one dimension we get

$$\mathrm{D}^p(\mathcal{R}_\mathbf{u}\mu*\mathcal{N}_{\sigma_2},\mathcal{R}_\mathbf{u}\nu*\mathcal{N}_{\sigma_2})\leq\mathrm{D}^p(\mathcal{R}_\mathbf{u}\mu*\mathcal{N}_{\sigma_1},\mathcal{R}_\mathbf{u}\nu*\mathcal{N}_{\sigma_1}).$$

Then, computing the expectation over the projections $\mathbf{u}$ since the divergence is non-negative concludes the proof.

A.8 Proof of Proposition 3.16: relation between $\mathrm{G}_\sigma\mathrm{SW}^p(\mu,\nu)$ under two noise levels

First, using the contractive property of convolution (see Lemma 3 in Nietert et al. (2021)), stating that for any probability measure $\alpha\in\mathcal{P}(\mathbb{R})$, $\mathrm{W}_p(\mu*\alpha,\nu*\alpha)\leq\mathrm{W}_p(\mu,\nu)$. Hence $\mathrm{W}_p^p(\mu*\mathcal{N}_{\sigma_2},\nu*\mathcal{N}_{\sigma_2})\leq\mathrm{W}_p^p(\mu*\mathcal{N}_{\sigma_1},\nu*\mathcal{N}_{\sigma_1})$. Now using Proposition 3.14 of the oreder relation satisfied by $\mathrm{G}_\sigma\mathrm{SW}^p$ yields

$$\mathrm{G}_{\sigma_2}\mathrm{SW}_p(\mu,\nu)\leq\mathrm{G}_{\sigma_1}\mathrm{SW}_p(\mu,\nu).$$

In the other direction, we have that $\mathcal{N}_{\sigma_2}=\mathcal{N}_{\sigma_1}*\mathcal{N}_{\sqrt{\sigma_2^2-\sigma_1^2}}$ (similarly for $\mathcal{N}_{\sigma_1}$). Setting the following random variables: $X_\mathbf{u}\sim\mathcal{R}_\mathbf{u}\mu, Y_\mathbf{u}\sim\mathcal{R}_\mathbf{u}\nu, Z_X\sim\mathcal{N}_{\sigma_1}, Z_Y\sim\mathcal{N}_{\sigma_1}, Z'_X\sim\mathcal{N}_{\sqrt{\sigma_2^2-\sigma_1^2}}, Z'_Y\sim\mathcal{N}_{\sqrt{\sigma_2^2-\sigma_1^2}}$. The sliced Wasserstein distance $\mathrm{W}_p^p(\mathcal{R}_\mathbf{u}\mu*\mathcal{N}_{\sigma_2},\mathcal{R}_\mathbf{u}\nu*\mathcal{N}_{\sigma_2})$ is given as a minimization over couplings $(X_\mathbf{u},Z_X,Z'_X)$ and $(Y_\mathbf{u},Z_Y,Z'_Y)$, namely

$$\mathrm{W}_p^p(\mathcal{R}_\mathbf{u}\mu*\mathcal{N}_{\sigma_2},\mathcal{R}_\mathbf{u}\nu*\mathcal{N}_{\sigma_2})=\inf_{\substack{X_\mathbf{u},Z_X,Z'_X\\Y_\mathbf{u},Z_Y,Z'_Y}}\mathbf{E}\big[\big|\big((X_\mathbf{u}+Z_X)-(Y_\mathbf{u}+Z_Y)\big)+(Z'_X-Z'_Y)\big|^p\big]$$

Using the inequality $\mathbf{E}[|U+V|^p]-2^{p-1}\mathbf{E}[|W|^p]\leq2^{p-1}\mathbf{E}[|U+V+W|^p]$ for any random variables $U,V,W\in\mathbb{L}_p$ integrable, we obtain,

$$2^{p-1}\mathbf{E}\big[|(X_\mathbf{u}+Z_X)-(Y_\mathbf{u}+Z_Y)+(Z'_X+Z'_Y)|^p\big]\geq\mathbf{E}\big[|(X_\mathbf{u}+Z_X)-(Y_\mathbf{u}+Z_Y)|^p\big]-2^{p-1}\mathbf{E}\big[|(Z'_X-Z'_Y)|^p\big]\big).$$

Hence,

$$\begin{aligned}
2^{p-1}\mathrm{W}_p^p(\mathcal{R}_{\mathbf{u}}\mu * \mathcal{N}_{\sigma_2}, \mathcal{R}_{\mathbf{u}}\nu * \mathcal{N}_{\sigma_2}) &\geq \inf \Big(\mathbf{E}\big[|(X_{\mathbf{u}} + Z_X) - (Y_{\mathbf{u}} + Z_Y)|^p\big] - 2^{p-1}\mathbf{E}\big[|(Z'_X - Z'_Y)|^p\big]\Big)\Big) \\
&\geq \mathrm{W}_p^p(\mathcal{R}_{\mathbf{u}}\mu * \mathcal{N}_{\sigma_1}, \mathcal{R}_{\mathbf{u}}\nu * \mathcal{N}_{\sigma_1}) - 2^{p-1}\sup \mathbf{E}\big[|(Z'_X - Z'_Y)|^p\big] \\
&\geq \mathrm{W}_p^p(\mathcal{R}_{\mathbf{u}}\mu * \mathcal{N}_{\sigma_1}, \mathcal{R}_{\mathbf{u}}\nu * \mathcal{N}_{\sigma_1}) - 2^{2p}\sup \mathbf{E}\big[|(Z'_X)|^p\big].
\end{aligned}$$

Hence,

$$\mathrm{G}_{\sigma_1}\mathrm{SW}_p(\mu, \nu) \leq 2^{1-\frac{1}{p}}\,\mathrm{G}_{\sigma_2}\mathrm{SW}_p(\mu, \nu) + 4\big(\sup \mathbf{E}\big[|(Z'_X)|^p\big]\big)\big)^{1/p}.$$

Finally, for any $p \geq 1$ the p-th moment of $|\mathcal{N}_\sigma|$ satisfies $\mathbf{E}[|\mathcal{N}_\sigma|^p] = \frac{2^p\Gamma((p+1)/2)}{\Gamma(1/2)}\sigma^{2p} \leq 2^{p/2}\sigma^{2p}$, then

$$\mathrm{G}_{\sigma_1}\mathrm{SW}_p(\mu, \nu) \leq 2^{1-\frac{1}{p}}\,\mathrm{G}_{\sigma_2}\mathrm{SW}_p(\mu, \nu) + 2^{\frac{5}{2}}(\sigma_2^2 - \sigma_1^2),$$

and concludes the proof.

A.9 Proof of Proposition 3.17: continuity of the smoothed Gaussian sliced Wasserstein w.r.t. σ

From Lemma 1 in (Nietert et al., 2021), we know that the Gaussian-smoothed Wasserstein is continuous with respect to σ, for any distribution $\mathcal{R}_{\mathbf{u}}\nu$ and $\mathcal{R}_{\mathbf{u}}\mu$. In addition, for any $\mathbf{u}$, we have $\mathrm{W}_p(\mathcal{R}_{\mathbf{u}}\nu * \mathcal{N}_\sigma, \mathcal{R}_{\mathbf{u}}\mu * \mathcal{N}_\sigma) \leq \mathrm{W}_p(\mathcal{R}_{\mathbf{u}}\nu, \mathcal{R}_{\mathbf{u}}\mu)$. Then by applying Lebesgue's dominated convergence theorem (Bowers & Kalton, 2014) to the above inequality with $\mathrm{W}_p(\mathcal{R}_{\mathbf{u}}\nu, \mathcal{R}_{\mathbf{u}}\mu)$ as a dominating function, that is u_d-almost everywhere integrable because both measures are in $\mathcal{P}_p(\mathbb{R}^d)$, we then conclude that the Gaussian-smoothed SWD is continuous w.r.t. σ.

A.10 Proof of Proposition 3.18: continuity of the smoothed sliced squared-MMD w.r.t. σ

Let us first recall the definition of the MMD divergence. Let $k : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a measurable bounded kernel on $\mathbb{R}$ and consider the reproducing kernel Hilbert space (RKHS) $\mathcal{H}_k$ associated with k and equipped with inner product $< \cdot, \cdot >_{\mathcal{H}_k}$ and norm $\|\cdot\|_{\mathcal{H}_k}$. Let $\mathcal{P}_{\mathcal{H}_k}(\mathbb{R})$ be the set of probability measures η such that $\int_{\mathbb{R}} \sqrt{k(t,t)}d\eta(x) < \infty$. The kernel mean embedding is defined as $\Phi_k(\eta) = \int_{\mathbb{R}} k(\cdot, t)d\eta(t)$. The squared-maximum mean discrepancy between $\eta, \zeta \in \mathcal{P}(\mathbb{R})$ denoted as $\mathrm{MMD} : \mathcal{P}_{\mathcal{H}_k}(\mathbb{R}) \times \mathcal{P}_{\mathcal{H}_k}(\mathbb{R}) \to \mathbb{R}_+$ is expressed as the distance between two such kernel mean embeddings. It is defined as Gretton et al. (2012)

$$\mathrm{MMD}^2(\eta, \zeta) = \|\Phi_k(\eta) - \Phi_k(\zeta)\|^2_{\mathcal{H}_k} = \mathbf{E}_{T,T'\sim\eta}[k(T, T')] - 2\mathbf{E}_{T\sim\eta, R\sim\zeta}[k(T, R)] + \mathbf{E}_{R,R'\sim\zeta}[k(R, R')]$$

where T and T' are independent random variables drawn according to η, R and R' are independent random variables drawn according to ζ, and T is independent of R. We define the Gaussian Smoothed Sliced squared-MMD as follows:

$$\begin{aligned}
\mathrm{G}_\sigma\mathrm{MMD}^2(\mu, \nu) &= \int_{\mathbb{S}^{d-1}} \|\Phi_k(\mathcal{R}_{\mathbf{u}}\mu * \mathcal{N}_\sigma) - \Phi_k(\mathcal{R}_{\mathbf{u}}\nu * \mathcal{N}_\sigma)\|^2_{\mathcal{H}_k}\, u_d(\mathbf{u})\mathrm{d}\mathbf{u} \\
&= \int_{\mathbb{S}^{d-1}} \big(\mathbf{E}_{T,T'\sim\mathcal{R}_{\mathbf{u}}\mu*\mathcal{N}_\sigma}[k(T, T')] - 2\mathbf{E}_{T\sim\mathcal{R}_{\mathbf{u}}\mu*\mathcal{N}_\sigma, R\sim\mathcal{R}_{\mathbf{u}}\nu*\mathcal{N}_\sigma}[k(T, R)] \\
&\qquad + \mathbf{E}_{R,R'\sim\mathcal{R}_{\mathbf{u}}\nu*\mathcal{N}_\sigma}[k(R, R')]\big)u_d(\mathbf{u})\mathrm{d}\mathbf{u}.
\end{aligned}$$

From the definition of the smoothed sliced squared-MMD, we have

$$\begin{aligned}
\mathbf{E}_{T,T'\sim\mathcal{R}_{\mathbf{u}}\mu*\mathcal{N}_\sigma}[k(T, T')] &= \iint_{\mathbb{R}\times\mathbb{R}} k(t, t')\mathrm{d}\mathcal{R}_{\mathbf{u}}\mu * \mathcal{N}_\sigma(t)\mathrm{d}\mathcal{R}_{\mathbf{u}}\mu * \mathcal{N}_\sigma(t') \\
&= \iint_{\mathbb{R}\times\mathbb{R}} \Big(\int_{\mathbb{R}} k(t + z, t')\mathrm{d}\mathcal{R}_{\mathbf{u}}\mu(z)\mathcal{N}_\sigma(t)\Big)\mathrm{d}\mathcal{R}_{\mathbf{u}}\mu * \mathcal{N}_\sigma(t') \\
&= \iint_{\mathbb{R}\times\mathbb{R}} \Big(\int_{\mathbb{R}^d} k(t + \mathbf{u}^\top x, t')\mathrm{d}\mu(x)\mathcal{N}_\sigma(t)\Big)\mathrm{d}\mathcal{R}_{\mathbf{u}}\mu * \mathcal{N}_\sigma(t') \\
&= \iint_{\mathbb{R}\times\mathbb{R}} \iint_{\mathbb{R}^d\times\mathbb{R}^d} k(t + \mathbf{u}^\top x, t' + \mathbf{u}^\top x')\mathrm{d}\mu(x)\mathrm{d}\mu(x')\mathrm{d}\mathcal{N}_\sigma(t)\mathrm{d}\mathcal{N}_\sigma(t').
\end{aligned}$$

Similarly,

$$\mathbf{E}_{R,R'\sim\mathcal{R}_{\mathbf{u}}\nu*\mathcal{N}_\sigma}[k(R,R')] = \iint_{\mathbb{R}\times\mathbb{R}}\iint_{\mathbb{R}^d\times\mathbb{R}^d} k(r+\mathbf{u}^\top y, r'+\mathbf{u}^\top y')\mathrm{d}\nu(y)\mathrm{d}\nu(y')\mathrm{d}\mathcal{N}_\sigma(r)\mathrm{d}\mathcal{N}_\sigma(r')$$

and

$$\mathbf{E}_{T\sim\mathcal{R}_{\mathbf{u}}\mu*\mathcal{N}_\sigma, R\sim\mathcal{R}_{\mathbf{u}}\nu*\mathcal{N}_\sigma}[k(T,R)] = \iint_{\mathbb{R}\times\mathbb{R}}\iint_{\mathbb{R}^d\times\mathbb{R}^d} k(t+\mathbf{u}^\top x, r+\mathbf{u}^\top y)\mathrm{d}\mu(x)\mathrm{d}\nu(y)\mathrm{d}\mathcal{N}_\sigma(t)\mathrm{d}\mathcal{N}_\sigma(r).$$

Together the assumption of boundness of the kernel function k and the continuity of integrals, the three latter terms are continuous functions w.r.t. $\sigma \in (0,\infty)$. Again by the boundness of the kernel function k, there exists a positive finite constant C_k such that

$$\left|\mathbf{E}_{T,T'\sim\mathcal{R}_{\mathbf{u}}\mu*\mathcal{N}_\sigma}[k(T,T')] - 2\mathbf{E}_{T\sim\mathcal{R}_{\mathbf{u}}\mu*\mathcal{N}_\sigma, R\sim\mathcal{R}_{\mathbf{u}}\nu*\mathcal{N}_\sigma}[k(T,R)] + \mathbf{E}_{R,R'\sim\mathcal{R}_{\mathbf{u}}\nu*\mathcal{N}_\sigma}[k(R,R')]\right| \leq 4C_k.$$

We conclude the continuity of $\sigma \mapsto \mathrm{G}_\sigma\mathrm{MMD}^2(\mu,\nu)$ by an application of the continuity of integrals.

B Additional experiments

B.1 Sample complexity on CIFAR dataset

We have also evaluated the sample complexity for the CIFAR dataset by sampling sets of increasing size. Results reported in Figure 1 confirms the findings obtained from the toy dataset.

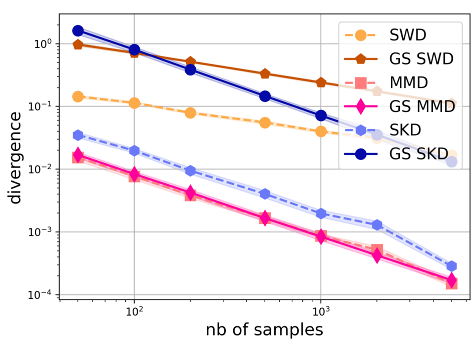

Figure 1: Measuring the divergence between two sets of samples drawn iid from the CIFAR10 dataset. We compare three sliced divergences and their Gaussian smoothed versions with a $\sigma = 3$.

B.2 Identity of indiscernibles

The second experiment aims at checking whether our divergences converge towards a small value when the distributions to be compared are the same. For this, we consider samples from distributions μ and ν chosen as normal distributions with respectively mean $2 \times \mathbf{1}_d$ and $s\mathbf{1}_d$ with varying s (noted as the displacement). Results are depicted in Figure 2. We can see that all methods are able to attain their minimum when $s = 2$. Interestingly, the gap between the Gaussian smoothed and non-smoothed divergences for Wasserstein and Sinkhorn is almost indiscernible as the distance between distribution increases.

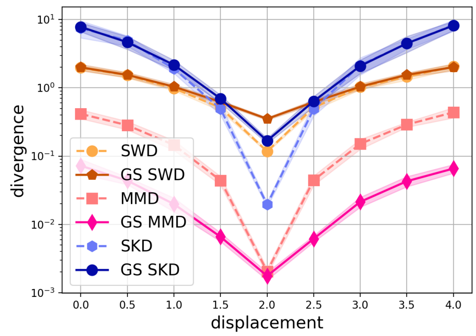

Figure 2: Measuring the divergence between two sets of samples in $\mathbb{R}^{50}$, one with mean $2\mathbf{1}_d$ and the other with mean $s\mathbf{1}_d$ with increasing s. We compare three sliced divergences and their Gaussian smoothed version with a $\sigma = 3$.