Appendix

A.0. Additional Notation

For a given probability measure $\mu \in \mathcal{P}$, let $\Phi_\mu(t) := \mathbb{E}\big[e^{i\langle t,X\rangle}\big]$ with $X \sim \mu$ denote its characteristic function. Let $C^k(\mathbb{R}^d)$ denote the class of k-times continuously differentiable functions on $\mathbb{R}^d$. Let $\mathcal{L}(X)$ denote the law of a random variable X. We write $\lesssim$ for inequalities up to some numerical constant.

A.1. Proofs for Section 3

We first prove the following lemmas.

Lemma 2 (General smooth metrics)**.** *Let $\kappa \in \mathcal{P}$ be a distribution whose characteristic function never vanishes. If d is a metric on $\mathcal{X} \subset \mathcal{P}$ and $\mathcal{X}$ is closed under taking convolutions with κ, then $\mathsf{d}_\kappa : (\mu,\nu) \mapsto \mathsf{d}(\mu * \kappa, \nu * \kappa)$ is also a metric on $\mathcal{X}$.*

Proof. Non-negativity and symmetry follow from definition. The triangle inequality is also straightforward, since for $\mu_1,\mu_2,\mu_3 \in \mathcal{X}$, the triangle inequality for d gives

$$\begin{aligned}\mathsf{d}_\kappa(\mu_1,\mu_2) &= \mathsf{d}(\mu_1 * \kappa, \mu_2 * \kappa)\\ &\le \mathsf{d}(\mu_1 * \kappa, \mu_3 * \kappa) + \mathsf{d}(\mu_3 * \kappa, \mu_2 * \kappa)\\ &= \mathsf{d}_\kappa(\mu_1,\mu_3) + \mathsf{d}_\kappa(\mu_3,\mu_2).\end{aligned}$$

Finally, if $\mathsf{d}_\kappa(\mu,\nu) = 0$, then $\mu * \kappa = \nu * \kappa$. Recalling that the characteristic function of a convolution of measures factors into a product, i.e., $\Phi_{\mu_1 * \mu_2} = \Phi_{\mu_1}\Phi_{\mu_2}$, and since the characteristic function of κ never vanishes, we have $\mu = \nu$. ◻

Lemma 3 (Contractive property of convolution)**.** *For any probability measure $\kappa \in \mathcal{P}$, $\mathsf{W}_p(\mu * \kappa, \nu * \kappa) \le \mathsf{W}_p(\mu,\nu)$. In particular, $\mathsf{W}_p^{(\sigma)}(\mu,\nu) \le \mathsf{W}_p(\mu,\nu)$.*

Proof. Let (X,Y) be an optimal coupling for $\mathsf{W}_p(\mu,\nu)$. Then taking $Z \sim \kappa$ independently,

$$\begin{aligned}\mathsf{W}_p(\mu * \kappa, \nu * \kappa)^p &\le \mathbb{E}\left[|(X+Z) - (Y+Z)|^p\right]\\ &= \mathbb{E}\left[|X-Y|^p\right] = \mathsf{W}_p(\mu,\nu).\end{aligned}$$ ◻

Lemma 4 (Coupling decomposition)**.** *If $\pi \in \Pi(\mu * \mathcal{N}_\sigma, \nu * \mathcal{N}_\sigma)$, then there exists a coupling (X,Y,Z,Z') such that $(X,Z) \sim \mu \otimes \mathcal{N}_\sigma$, $(Y,Z') \sim \nu \otimes \mathcal{N}_\sigma$, and $(X+Z, Y+Z') \sim \pi$.*

Proof. If suffices to find a coupling $(X+Z, Y+Z', Z, Z')$ with the correct marginals. First, note that we already have couplings $(X+Z, Y+Z')$, $(X+Z, Z)$ and $(Y+Z', Z')$, given by π, $(\mu * \mathcal{N}_\sigma) \otimes \mathcal{N}_\sigma$, and $(\nu * \mathcal{N}_\sigma) \otimes \mathcal{N}_\sigma$, respectively. Hence, we can apply the gluing lemma (see, e.g., (Villani, 2003)) between π and $(\mu * \mathcal{N}_\sigma) \otimes \mathcal{N}_\sigma$ to obtain a coupling $(X+Z, Y+Z', Z)$ and then between π, $(\nu * \mathcal{N}_\sigma) \otimes \mathcal{N}_\sigma$ to obtain a coupling $(X+Z, Y+Z', Z')$. We apply the gluing lemma a final time between the outcomes of its previous applications to obtain a coupling $(X+Z, Y+Z', Z, Z')$. ◻

A.1.1. Proof of Proposition 1

Lemma 2 verifies that $\mathsf{W}_p^{(\sigma)}$ is a metric on $\mathcal{P}_p$, since $\Phi_{\mathcal{N}_\sigma}(t) = e^{-\sigma^2|t|^2/2} \neq 0$, for all $t \in \mathbb{R}^d$. To show that $\mathsf{W}_p^{(\sigma)}$ induces the same topology as W_p, it suffices to prove that

$$\mathsf{W}_p(\mu_n,\mu) \to 0 \iff \mathsf{W}_p^{(\sigma)}(\mu_n,\mu) \to 0.$$

The "⇒" direction follows by Lemma 3. For the other direction, suppose that $\mathsf{W}_p^{(\sigma)}(\mu_n,\mu) \to 0$. By Lemma 4, we can find a coupling $\big((X_n,Z_n),(X,Z)\big)$ with $(X_n,Z_n) \sim \mu_n \otimes \mathcal{N}_\sigma$ and $(X,Z) \sim \mu \otimes \mathcal{N}_\sigma$ such that $\mathsf{W}_p^{(\sigma)}(\mu_n,\mu)^p = \mathbb{E}[|X_n + Z_n - (X+Z)|^p]$. We will show that $X_n \xrightarrow{d} X$ and $\mathbb{E}[|X_n|^p] \to \mathbb{E}[|X|^p]$, which yields the desired result.

To that end, it is sufficient (and necessary) to show that $X_n \xrightarrow{d} X$ and that $|X_n|^p$ is uniformly integrable. Since convergence in distribution is equivalent to pointwise convergence of characteristic functions, from $X_n + Z_n \xrightarrow{d} X + Z$, we have for all $t \in \mathbb{R}^d$ that

$$\begin{aligned}\lim_{n\to\infty} \Phi_{\mu_n}(t) e^{-\sigma^2|t|^2/2} &= \lim_{n\to\infty} \Phi_{\mu_n * \mathcal{N}_\sigma}(t)\\ &= \Phi_{\mu * \mathcal{N}_\sigma}(t) = \Phi_\mu(t) e^{-\sigma^2|t|^2/2},\end{aligned}$$

implying that $\lim_{n\to\infty} \Phi_{\mu_n}(t) = \Phi_\mu(t)$, for all $t \in \mathbb{R}^d$, and hence that $X_n \xrightarrow{d} X$. To verify the uniform integrability, observe that $|X_n|^p \le 2^{p-1}(|X_n + Z_n|^p + |Z_n|^p)$. By construction, $|X_n + Z_n|^p$ is uniformly integrable, while $|Z_n|^p \overset{d}{=} |Z|^p$ is trivially uniformly integrable, implying the uniform integrability of their sum and hence $|X_n|^p$. ◻

A.1.2. Proof of Lemma 1

By Lemma 3, we have $\mathsf{W}_p^{(\sigma_2)}(\mu,\nu) \le \mathsf{W}_p^{(\sigma_1)}(\mu,\nu)$. For the other direction, let $X \sim \mu$, $Y \sim \nu$, $Z_X \sim \mathcal{N}_{\sigma_1}$, $Z_Y \sim \mathcal{N}_{\sigma_1}$, $Z'_X \sim \mathcal{N}_{\sqrt{\sigma_2^2 - \sigma_1^2}}$, and $Z'_Y \sim \mathcal{N}_{\sqrt{\sigma_2^2 - \sigma_1^2}}$. The smooth p-Wasserstein distance of parameter σ_2 is given as a minimization over couplings of the aforementioned random variables subject to the mutual independence of (X, Z_X, Z'_X) along with that of (Y, Z_Y, Z'_Y). With this convention, we have

$$\begin{aligned}&\mathsf{W}_p^{(\sigma_2)}(\mu,\nu)\\ &= \inf\left(\mathbb{E}\left[\left|\big((X+Z_X)-(Y+Z_Y)\big)+(Z'_X - Z'_Y)\right|^p\right]\right)^{1/p}.\end{aligned}$$

Now, Minkoski's inequality gives

$$\begin{aligned}\mathsf{W}_p^{(\sigma_2)}(\mu,\nu) &\geq \inf\Big[\Big(\mathbb{E}\Big[\big|(X+Z_X)-(Y+Z_Y)\big|^p\Big]\Big)^{1/p} \\ &\qquad - \Big(\mathbb{E}\Big[\big|Z_X'-Z_Y'\big|^p\Big]\Big)^{1/p}\Big] \\ &\geq \mathsf{W}_p^{(\sigma_1)}(\mu,\nu) - \sup\big(\mathbb{E}\big[|Z_X'-Z_Y'|^p\big]\big)^{1/p} \\ &\geq \mathsf{W}_p^{(\sigma_1)}(\mu,\nu) - 2\big(\mathbb{E}\big[|Z_X'|^p\big]\big)^{1/p}.\end{aligned}$$

Recall that for $Z \sim \mathcal{N}(0, \mathrm{I}_d)$,

$$\mathbb{E}\big[|Z|^p\big] = \frac{2^{p/2}\Gamma((p+d)/2)}{\Gamma(d/2)}.$$

If p is even, then above term is bounded by $(d+2p-2)^{p/2}$. In general, we round p up to the nearest even integer to obtain the bound $(d+2p+2)^{p/2}$, completing the proof. □

A.1.3. Proof of Corollary 1

The proof follows that of Theorem 3 in (Goldfeld & Greenewald, 2020). For Claim (ii), we simply apply Lemma 1, taking $\sigma_1 = 0$ and $\sigma_2 \to 0$. For Claim (i), monotonicity follows directly from the contractive property established in the previous proof. For left continuity of $\mathsf{W}_p^{(\sigma)}$, we apply Lemma 1 with $\sigma_2 = \sigma$ and $\sigma_1 \nearrow \sigma$. For right continuity, take $\sigma_k \searrow \sigma$ and define $\varepsilon_k = \sqrt{\sigma_k^2 - \sigma^2}$. Then,

$$\mathsf{W}_p^{(\sigma_k)}(\mu,\nu) = \mathsf{W}_p^{(\varepsilon_k)}(\mu * \mathcal{N}_\sigma, \nu * \mathcal{N}_\sigma) \to \mathsf{W}_p^{(\sigma)}(\mu,\nu)$$

as $k \to \infty$. Claim (iii) follows from Corollary 2.4 of (Chen & Niles-Weed, 2020). □

A.1.4. Proof of Proposition 2

A close inspection of the proof of Theorem 4 in (Goldfeld & Greenewald, 2020), which covers the $p = 1$ case up to extraction of a subsequence, reveals that the only required properties of $|\cdot|^1$ are its non-negativity and continuity. These also hold for $|\cdot|^p$, so the theorem applies to $\mathsf{W}_p^{(\sigma)}$. Further, the proof implies that any weakly convergent subsequence of couplings converges to an optimal coupling for $\mathsf{W}_p^{(\sigma)}(\mu,\nu)$. Since for $p > 1$ optimal couplings are unique (see, e.g., Theorem 2.44 of (Villani, 2003)), Prokhorov's Theorem implies that extraction of a subsequence is not necessary. □

A.1.5. Proof of Theorem 1

We begin with a useful result bounding unsmoothed W_p by a dual Sobolev norm, adapting a proof from (Dolbeault et al., 2009).

Lemma 5. *Fix $p > 1$ and suppose that $\mu_0, \mu_1 \in \mathcal{P}_p$ with $\mu_0, \mu_1 \ll \gamma$ for some locally finite Borel measure γ on $\mathbb{R}^d$. Denote their respective densities by $f_i = \mathrm{d}\mu_i/\mathrm{d}\gamma$. If f_0 or f_1 is lower bounded by some $c > 0$, then we have*

$$\mathsf{W}_p(\mu_0,\mu_1) \leq p\, c^{-1/q} \|\mu_0 - \mu_1\|_{\dot{H}^{-1,p}(\gamma)}.$$

Proof. We essentially apply Theorem 5.26 of (Dolbeault et al., 2009), which (for the choice of $\phi(\rho, w) = \rho^{1-p}|w|^p$), bounds W_p from above by the relevant dual Sobolev norm times a constant which depends on a lower bound for *both* f_0 and f_1. The proof exploits the dynamic Benamou-Brenier formulation of optimal transport and the path in $(\mathcal{P}_p, \mathsf{W}_p)$ which interpolates linearly between densities. Before concluding, they show

$$\mathsf{W}_p(\mu_0,\mu_1)^p \leq \int_0^1 \int_{\mathbb{R}^d} ((1-t)f_0 + tf_1)^{1-p} |w|^p \,\mathrm{d}\gamma\,\mathrm{d}t,$$

where $\|w\|_{L_p(\gamma;\mathbb{R}^d)} = \|\mu_0 - \mu_1\|_{\dot{H}^{-1,p}(\gamma)}$ (such w is shown to exist only assuming $\|\mu_0 - \mu_1\|_{\dot{H}^{-1,p}(\gamma)} < \infty$). However, even with the lower bound c on just one of the densities (say f_0 without loss of generality), we have

$$\begin{aligned}&\int_0^1 \int_{\mathbb{R}^d} ((1-t)f_0 + tf_1)^{1-p} |w|^p \,\mathrm{d}\gamma\,\mathrm{d}t \\ &\leq \int_0^1 (tc)^{1-p} \int_{\mathbb{R}^d} |w|^p \,\mathrm{d}\gamma\,\mathrm{d}t \\ &\quad = c^{1-p} \|w\|_{L_p(\gamma;\mathbb{R}^d)}^p \int_0^1 t^{1-p}\,\mathrm{d}t \\ &\quad = p^p c^{1-p} \|\mu_0 - \mu_1\|_{\dot{H}^{-1,p}(\gamma)}^p,\end{aligned}$$

which gives the lemma. □

To prove the theorem, we apply the lemma with $\mu_0 = \mu * \mathcal{N}_\sigma$, $\mu_1 = \nu * \mathcal{N}_\sigma$, and $\gamma = \mathcal{N}_\sigma$. To bound $\mathrm{d}\mu * \mathcal{N}_\sigma / \mathrm{d}\mathcal{N}_\sigma$ from below, let $X \sim \mu$ and compute

$$\begin{aligned}\mu * \varphi_\sigma(y) &= \frac{1}{(2\pi\sigma^2)^{d/2}} \int_{\mathbb{R}^d} e^{-|x-y|^2/(2\sigma^2)} \mathrm{d}\mu(x) \\ &\geq \frac{1}{(2\pi\sigma^2)^{d/2}} e^{-\mathbb{E}[|y-X|^2/(2\sigma^2)]},\end{aligned}$$

where the second step uses Jensen's inequality. The desired conclusion follows because $\mathbb{E}[|y-X|^2] = |y|^2 + \mathbb{E}[|X|^2] - 2\langle y, \mathbb{E}[X]\rangle$ and X has mean zero.

For a related lower bound, we will apply Theorem 5.24 of (Dolbeault et al., 2009) with the choice of $\phi(\rho, w) = |w|^p$ to see that $\mathsf{W}_p(\mu_0,\mu_1) \geq C^{-1} \|\mu_0 - \mu_1\|_{\dot{H}^{-1,p}(\gamma)}$ under the same conditions as Lemma 5 but where C is now an upper bound on the densities. To start, we compute

$$\begin{aligned}\frac{\mu * \varphi_\sigma(y)}{\varphi_{\sqrt{2}\sigma}(y)} &= 2^{d/2} \int_{\mathbb{R}^d} e^{-\frac{|y-x|^2}{2\sigma^2} + \frac{|y|^2}{4\sigma^2}} \mathrm{d}\mu(x) \\ &= 2^{d/2} \int_{\mathbb{R}^d} e^{-\frac{|y-2x|^2}{4\sigma^2} + \frac{|x|^2}{2\sigma^2}} \mathrm{d}\mu(x) \\ &\leq 2^{d/2}\, \mathbb{E}\big[e^{|X|^2/(2\sigma^2)}\big],\end{aligned}$$

where $X \sim \mu$. Hence,

$$\mathsf{W}_p^{(\sigma)}(\mu_0,\mu_1) \geq 2^{-d/2} \left(\mathbb{E}\left[e^{|X_0|^2/(2\sigma^2)}\right] \wedge \mathbb{E}\left[e^{|X_1|^2/(2\sigma^2)}\right]\right)^{-1} \left\|(\mu_0-\mu_1)*\mathcal{N}_\sigma\right\|_{\dot{H}^{-1,p}(\mathcal{N}_{\sqrt{2}\sigma})},$$

where $X_0 \sim \mu_0$ and $X_1 \sim \mu_1$. This bound is only meaningful when μ_0 and μ_1 are sufficiently sub-Gaussian. □

A.1.6. Proof of Proposition 3

For (i), we observe that if $\mu \neq \nu$, then the two measures must share a continuity set A such that $\mu(A) \neq \nu(A)$. We can assume without loss of generality that A does not contain the origin and that $(\mu-\nu)(A) > 0$. Then, for any $C > 0$, there exists sufficiently small σ such that

$$\begin{aligned}\mathsf{d}_p^{(\sigma)}(\mu,\nu) &= \sup_{f:\|\nabla f\|_{L^q(\mathcal{N}_\sigma)}\leq 1} (\mu*\mathcal{N}_\sigma - \nu*\mathcal{N}_\sigma)(f)\\ &\geq (\mu*\mathcal{N}_\sigma - \nu*\mathcal{N}_\sigma)(C\mathbb{1}_A)\\ &= C(\mu*\mathcal{N}_\sigma - \nu*\mathcal{N}_\sigma)(A)\\ &\geq \frac{C}{2}(\mu-\nu)(A).\end{aligned}$$

By taking C arbitrarily large, we see that $\mathsf{d}_p^{(\sigma)}(\mu,\nu)=\infty$, establishing (i). For (ii), we employ Theorem 4 and observe that

$$\kappa^{(\sigma)}(x,y) = \langle x,y\rangle + \frac{1}{4\sigma^2}\langle x,y\rangle^2 + O(\sigma^{-4}).$$

As $\sigma \to \infty$, we obtain the pointwise limit kernel $\kappa^{(\infty)} = \langle x,y\rangle$, which induces the distance given in (ii). Swapping the limit and the expectation in (5) is justified by the Dominated Convergence Theorem given that μ and ν are sub-Gaussian. □

A.2. Proofs for Section 4

A.2.1. Proof of Theorem 2

The argument relies on Proposition 7.10 from (Villani, 2003), which is restated next.

Lemma 6 (Proposition 7.10 in (Villani, 2003))**.** *For any $1 \leq p < \infty$, we have*

$$\mathsf{W}_p(\mu,\nu) \leq 2^{\frac{p-1}{p}} \left(\int_{\mathbb{R}^d} |x|^p \mathrm{d}|\mu-\nu|(x)\right)^{1/p}. \quad (6)$$

This bound follows by coupling μ and ν via the maximal TV-coupling and evaluating the resulting transportation cost.

Invoking the lemma and Jensen's inequality, we have

$$\begin{aligned}&\mathbb{E}\left[\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n,\mu)\right]\\ &\leq 2^{\frac{p-1}{p}}\left(\int_{\mathbb{R}^d}|x|^p\,\mathbb{E}\left[\left|\hat{\mu}_n*\varphi_\sigma(x)-\mu*\varphi_\sigma(x)\right|\right]\mathrm{d}x\right)^{1/p}\\ &\leq 2^{\frac{p-1}{p}}n^{-\frac{1}{2p}}\left(\int_{\mathbb{R}^d}|x|^p\sqrt{\mathrm{Var}\left[\varphi_\sigma(x-X)\right]}\mathrm{d}x\right)^{1/p},\end{aligned}$$

where the last inequality follows because $\mathbb{E}\left[\varphi_\sigma(x-X)\right] = \mu*\varphi_\sigma(x)$ for all $x\in\mathbb{R}^d$. Furthermore,

$$\begin{aligned}&\mathrm{Var}\left[\varphi_\sigma(x-X)\right] \leq \mathbb{E}[\varphi_\sigma(x-X)^2]\\ &= \frac{1}{(2\pi\sigma^2)^d}\int_{\mathbb{R}^d} e^{-\frac{|x-y|^2}{\sigma^2}}\mathrm{d}\mu(y)\\ &= \frac{1}{(2\pi\sigma^2)^d}\left(\int_{|y|\leq\frac{|x|}{2}} + \int_{|y|>\frac{|x|}{2}}\right)e^{-\frac{|x-y|^2}{\sigma^2}}\mathrm{d}\mu(y)\\ &\leq \frac{1}{(2\pi\sigma^2)^d}\left(\int_{|y|\leq\frac{|x|}{2}} e^{-\frac{|x-y|^2}{\sigma^2}}\mathrm{d}\mu(y) + \mathbb{P}\left(|X|>\frac{|x|}{2}\right)\right).\end{aligned}$$

If $|y|\leq|x|/2$, then $|x-y|^2\geq|x|^2/4$, which yields

$$\sqrt{\mathrm{Var}(\varphi_\sigma(x-X))} \leq \frac{e^{-\frac{|x|^2}{8\sigma^2}} + \sqrt{\mathbb{P}\left(|X|>\frac{|x|}{2}\right)}}{(2\pi\sigma^2)^{d/2}}.$$

Direct calculations show that

$$\int_{\mathbb{R}^d}|x|^p e^{-\frac{|x|^2}{8\sigma^2}}\mathrm{d}x = \frac{8^{\frac{d+p}{2}}\sigma^{d+p}\pi^{d/2}\Gamma((d+p)/2)}{\Gamma(d/2)}$$

and

$$\begin{aligned}&\int_{\mathbb{R}^d}|x|^p\sqrt{\mathbb{P}\left(|X|>|x|/2\right)}\mathrm{d}x\\ &= \frac{2^{d+p+1}\pi^{d/2}}{\Gamma(d/2)}\int_0^\infty r^{d+p-1}\sqrt{\mathbb{P}(|X|>r)}\mathrm{d}r.\end{aligned}$$

Hence $\mathbb{E}\left[\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n,\mu)\right] = O\big(n^{-1/(2p)}\big)$ if Condition (2) holds. The last assertion follows from Markov's inequality.

To specify the exact constant, we combine the above bounds and simplify to obtain to obtain

$$\mathbb{E}\left[\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n,\mu)\right] \leq 2^{1-1/p}n^{-1/2p} \left(\frac{2^{d+3p/2}\sigma^p\Gamma((d+p)/2)}{\Gamma(d/2)} + \frac{2^{d/2+p+1}I}{\Gamma(d/2)\sigma^d}\right)^{1/p},$$

where I is the integral from Condition (2). By the subadditivity of $t\mapsto t^{1/p}$ and properties of the gamma function, we bound the RHS above by

$$8n^{-1/2p}\left(2^{d/p}\sigma\sqrt{d/2+p+1} + \frac{2^{d/(2p)}I^{1/p}}{\Gamma(d/2)^{1/p}\sigma^{d/p}}\right).$$

If μ is β-sub-Gaussian, then $\mathbb{P}(|X|>r)\leq 2^{d/2}e^{-\frac{r^2}{4\beta^2}}$ and we can bound

$$\begin{aligned}I&=\int_0^\infty r^{d+p-1}\sqrt{\mathbb{P}(|X|>r)}\,\mathrm{d}r\\&\leq 2^{d/4}\int_0^\infty r^{d+p-1}e^{-r^2/(4\beta^2)}\,\mathrm{d}r\\&=2^{d/4-1}(2\beta)^{d+p}\Gamma((d+p)/2)\\&=2^{5d/4+p-1}\beta^{d+p}\Gamma((d+p)/2).\end{aligned}$$

Plugging this into the previous bound, using properties of the gamma function, and simplifying, we obtain

$$\begin{aligned}&\mathbb{E}\left[\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n,\mu)\right]\leq 8n^{-1/2p}\\&\left(2^{d/p}\sigma\sqrt{d+p}+2^{7d/(4p)}\beta^{d/p+1}\sqrt{d+p}\sigma^{-d/p}\right)\\&\leq 8\cdot 4^{d/p}\sqrt{d+p}\left[\sigma+\beta\left(\frac{\beta}{\sigma}\right)^{d/p}\right]\cdot n^{-1/(2p)}.\end{aligned}$$

□

A.2.2. Proof of Theorem 3

For $p\geq 1$, a probability measure $\gamma\in\mathcal{P}$ is said to satisfy the *p-Poincaré inequality* if there exists a finite constant D such that

$$\|f-\gamma(f)\|_{L^p(\gamma)}\leq D\|\nabla f\|_{L^p(\gamma;\mathbb{R}^d)},\quad\forall f\in C_0^\infty.\tag{7}$$

The smallest constant satisfying the above is denoted by $D_p(\gamma)$. We note in particular that $\mathcal{N}_\sigma$ satisfies a p-Poincaré inequality for all $p\geq 1$ (see, e.g., (Boucheron et al., 2013) and Theorem 2.4 of (Milman, 2009)).

Let $\partial_j=\partial/\partial x_j$. For any multi-index $k=(k_1,\ldots,k_d)\in\mathbb{N}_0^d$, define the differential operator

$$\partial^k=\partial_1^{k_1}\cdots\partial_d^{k_d},$$

and let $\bar{k}=\sum_{j=1}^d k_j$. We start by bounding the derivatives of centered functions with bounded homogeneous Sobolev norm after Gaussian smoothing.

Lemma 7. *Fix $\eta>0$. Pick any $f\in C_0^\infty$ such that $\|f\|_{\dot{H}^{1,q}(\mathcal{N}_\sigma)}\leq 1$, and let $f_\sigma=f*\varphi_\sigma-\mathcal{N}_\sigma(f)$. Then for any multi-index $k=(k_1,\ldots,k_d)\in\mathbb{N}_0^d$,*

$$|\partial^k f_\sigma(x)|\lesssim(D_q(\mathcal{N}_\sigma)\vee\sigma^{-\bar{k}+1})\exp\left(\frac{(p-1)(1+\eta)|x|^2}{2\sigma^2}\right)$$

up to constants independent of f,x, and σ.

Proof of Lemma 7. Observe that

$$\begin{aligned}f_\sigma(x)&=\int\varphi_\sigma(x-y)f(y)\mathrm{d}y\\&=\int\frac{\varphi_\sigma(x-y)}{\varphi_\sigma(y)}f(y)\varphi_\sigma(y)\mathrm{d}y.\end{aligned}$$

Applying Hölder's inequality, we have

$$|f_\sigma(x)|\leq\left[\int\frac{\varphi_\sigma^p(x-y)}{\varphi_\sigma^{p-1}(y)}\mathrm{d}y\right]^{1/p}\|f\|_{L^q(\mathcal{N}_\sigma)}.$$

Here, since $\|\nabla f\|_{L^q(\mathcal{N}_\sigma;\mathbb{R}^d)}=\|f\|_{\dot{H}^{1,q}(\mathcal{N}_\sigma)}\leq 1$, we have

$$\|f\|_{L^q(\mathcal{N}_\sigma)}\leq D_q(\mathcal{N}_\sigma)\|\nabla f\|_{L^q(\mathcal{N}_\sigma;\mathbb{R}^d)}\leq D_q(\mathcal{N}_\sigma).$$

Observe that

$$\begin{aligned}&\int\frac{\varphi_\sigma^p(x-y)}{\varphi_\sigma^{p-1}(y)}\mathrm{d}y\\&=\frac{1}{(2\pi\sigma^2)^{d/2}}\int\exp\left[-\frac{p|x-y|^2-(p-1)|y|^2}{2\sigma^2}\right]\mathrm{d}y\\&=e^{-p|x|^2/(2\sigma^2)}\int e^{p\langle x,y\rangle/\sigma^2}\varphi_\sigma(y)\,\mathrm{d}y\\&=\exp\left(\frac{p(p-1)|x|^2}{2\sigma^2}\right).\end{aligned}$$

This yields that

$$|f_\sigma(x)|\leq D_q(\mathcal{N}_\sigma)\exp\left(\frac{(p-1)|x|^2}{2\sigma^2}\right),$$

establishing the claim when $\bar{k}=0$.

Next, we note that

$$\begin{aligned}\nabla f_\sigma(x)&=\int[\nabla_x\varphi_\sigma(x-y)]f(y)\mathrm{d}y\\&=-\int[\nabla_y\varphi_\sigma(x-y)]f(y)\mathrm{d}y\\&=\int\varphi_\sigma(x-y)\nabla_y f(y)\mathrm{d}y.\end{aligned}$$

Since $\|\nabla f\|_{L^q(\mathcal{N}_\sigma;\mathbb{R}^d)}\leq 1$, we can apply the preceding argument to conclude that

$$|\nabla f_\sigma(x)|\leq\exp\left(\frac{(p-1)|x|^2}{2\sigma^2}\right).$$

Finally, we extend to arbitrary derivatives, observing that for any $i=1,\ldots,d$ and $k\in\mathbb{N}_0^d$,

$$\begin{aligned}&\partial^k\partial_i f_\sigma(x)\\&=\int[\partial_i f(y)]\varphi_\sigma(x-y)\prod_{j=1}^d(-1)^{k_j}\sigma^{-k_j}\mathrm{He}_{k_j}\left(\frac{x_j-y_j}{\sigma}\right)\mathrm{d}y.\end{aligned}\tag{8}$$

Here, we use that

$$\partial^k\varphi_\sigma(z)=\varphi_\sigma(z)\left[\prod_{j=1}^d(-1)^{k_j}\sigma^{-k_j}\,\mathrm{He}_{k_j}\left(z_j/\sigma\right)\right],$$

where He_n is the Hermite polynomial of degree n defined by

$$\mathrm{He}_n(x) = (-1)^n e^{x^2/2} \frac{\mathrm{d}^n}{\mathrm{d}x^n} e^{x^2/2}.$$

Return to (8). Pick any $\eta > 0$. Since the product term in (8) can be bounded (up to constants) by $1 + |x-y|^{\bar{k}}$, we have

$$|\partial^k \partial_j f_\sigma(x)| \lesssim \sigma^{-\bar{k}} \int |\partial_j f(y)|\, \varphi_{\sigma(1+\eta)^{-1/2}}(x-y)\mathrm{d}y.$$

up to a constant independent of $f, x,$and σ. The desired bound follows by the same argument we applied to control $|\nabla f_\sigma(x)|$.

Now, to be more precise with constants, we note that since $D_2(\mathcal{N}_\sigma) = \sigma^2$ and $\mathcal{N}_\sigma$ is log-concave, we have by Theorem 2.4 of (Milman, 2009) that $D_q(\mathcal{N}_\sigma) \le C\sigma^2$ for all $q \in [1,\infty]$, for some absolute constant $C > 0$. Next, we recall the explicit formula

$$\mathrm{He}_n(x) = n! \sum_{m=0}^{\lfloor n/2 \rfloor} \frac{(-1)^m}{m!(n-2m)!} \frac{x^{n-2m}}{2^m}.$$

Using $|x|^m \le 1 + |x|^n$ for $m = 1,\dots,n$, we (quite loosely) bound

$$|\mathrm{He}_n(x)| \le n!(1+|x|^n) \sum_{m=0}^{\lfloor n/2 \rfloor} \frac{1}{m!(n-2m)!2^m}.$$

This summand is unimodal and attains its maximum at $m = \left\lceil \frac{n}{2} - \frac{\sqrt{n+2}}{2} \right\rceil$. Using this and Stirling's approximation, we find

$$\begin{aligned}|\mathrm{He}_n(x)| &\le \frac{n!(1+|x|^n)(n+4)}{\Gamma\big(\frac{n}{2} - \frac{\sqrt{n+2}}{2}\big)\Gamma\big(\sqrt{n+2}-1\big)2^{\frac{n}{2}-\frac{\sqrt{n+2}}{2}}} \\ &\le (1+|x|^n)(cn)^{n/2},\end{aligned}$$

for some absolute constant $c > 0$. Now, the product term in (8) is bounded in absolute value by

$$\sigma^{-\bar{k}} \prod_{j=1}^d (1+|z_j|)^n (ck_j)^{k_j/2} \le \sigma^{-\bar{k}} (c\bar{k})^{\bar{k}/2} (1+|z|)^{\bar{k}}.$$

With a bit of calculus, we compute

$$\begin{aligned}|\partial^k \partial_j f_\sigma(x)| \le (c'\bar{k})^{\bar{k}} \sigma^{-\bar{k}} (1+\eta)^{d/2} \\ \int |\partial_j f(y)|\, \varphi_{\sigma(1+\eta)^{-1/2}}(x-y)\mathrm{d}y,\end{aligned}$$

for some second constant $c' > 0$ and any $\eta > 0$, so long as $\sigma \le 1$, say. Applying the same argument used to control $|\nabla f_\sigma(x)|$, we bound

$$\begin{aligned}&\int \frac{\varphi_{\sigma(1+\eta)^{-1/2}}(x-y)^p}{\varphi_\sigma(y)^{p-1}}\,\mathrm{d}y \le \\ &(1+p\eta)^d e^{-\frac{p(1+\eta)|x|^2}{2\sigma^2}} \int e^{-\frac{p(1+\eta)\langle x,y\rangle}{\sigma^2}} \varphi_{\sigma(1+\eta p)^{-1/2}}(y)\,\mathrm{d}y \\ &= (1+p\eta)^d e^{-\frac{p(1+\eta)|x|^2}{2\sigma^2} + \frac{p^2(1+\eta)^2|x|^2}{(1+\eta p)\sigma^2}} \\ &= (1+p\eta)^d e^{-\frac{p(1+\eta)|x|^2}{2\sigma^2} + \frac{p^2(1+\eta)^2|x|^2}{(1+\eta p)\sigma^2}},\end{aligned}$$

which yields

$$\begin{aligned}|\partial^k \partial_j f_\sigma(x)| &\le (c'\bar{k})^{\bar{k}} \eta^{-\bar{k}/2} \sigma^{-\bar{k}} (1+\eta)^{3d/2} \\ &\quad \exp\left(\frac{|x|^2}{2\sigma^2}\left(\frac{p(1+\eta)^2}{(1+\eta p)} - (1+\eta)\right)\right) \\ &\le (c'\bar{k})^{\bar{k}} \eta^{-\bar{k}/2} \sigma^{-\bar{k}} (1+\eta)^{3d/2} \\ &\quad \exp\left(\frac{(p-1)|x|^2}{2\sigma^2}(1+\eta p+\eta)\right).\end{aligned}$$

Substituting η with $\eta/(p+1)$ and combining with the previous results, we establish the bound

$$\begin{aligned}&|\partial^k f_\sigma(x)| \le \\ &(C')^d \bar{k}^{\bar{k}-1} p^{3d/2} \sigma^{1-\bar{k}} \exp\left(\frac{(p-1)|x|^2}{\sigma^2}\right)\end{aligned}$$

for some absolute constant $C' > 0$ and any $k \in \mathbb{N}_0^d$, when $\sigma \le 1$. □

Next, we present a useful lemma concerning empirical approximation for IPMs whose function classes are sufficiently well-behaved.

Lemma 8. *Let $\mathcal{F} \subset C^\alpha(\mathbb{R}^d)$ be a function class where α is a positive integer with $\alpha > d/2$, and let $\{\mathcal{X}_j\}_{j=1}^\infty$ be a cover of $\mathbb{R}^d$ consisting of nonempty bounded convex sets with bounded diameter. Set $M_j = \sup_{f\in\mathcal{F}} \|f\|_{C^\alpha(\mathcal{X}_j)}$ with $\|f\|_{C^\alpha(\mathcal{X}_j)} = \max_{\bar{k}\le\alpha} \sup_{x\in\mathrm{int}(\mathcal{X}_j)} |\partial^k f(x)|$. If $\sum_{j=1}^\infty M_j \mu(\mathcal{X}_j)^{1/2} < \infty$, then $\mathcal{F}$ is μ-Donsker and $\mathbb{E}\left[\|\hat{\mu}_n - \mu\|_{\infty,\mathcal{F}}\right] \lesssim n^{-1/2} \sum_{j=1}^\infty M_j \mu(\mathcal{X}_j)^{1/2}$ up to constants that depend only on $d, \alpha,$ and $\sup_j \mathrm{diam}(\mathcal{X}_j)$.*

Proof of Lemma 8. The lemma follows from Theorem 1.1 in (var der Vaart, 1996). Let $I_1 = \mathcal{X}_1$ and $I_j = \mathcal{X}_j \setminus \bigcup_{k=1}^{j-1} \mathcal{X}_k$ for $j = 2,3,\dots$. The collection $\{I_j\}$ forms a partition of $\mathbb{R}^d$. Define $\mathcal{F}_{\mathcal{X}_j} = \{f \mathbb{1}_{\mathcal{X}_j} : f \in \mathcal{F}\}$ and $\mathcal{F}_{I_j} = \{f\mathbb{1}_{I_j} : f \in \mathcal{F}\}$. Let $F = \sum_j M_j 1_{I_j}$, which gives an envelope for $\mathcal{F}$. Observe that

$$\mu(F^2) = \sum_j M_j^2 \mu(I_j) \le \sum_j M_j^2 \mu(\mathcal{X}_j) < \infty,$$

which also ensures that $\mathcal{F} \subset L^2(\mu)$.

In view of the discussion before Corollary 2.1 in (var der Vaart, 1996), we see that each $\mathcal{F}_{\mathcal{X}_j}$ is μ-Donsker (which implies that $\mathcal{F}_{I_j}$ is μ-Donsker as $\mathcal{F}_{I_j}$ can be viewed as a subset of $\mathcal{F}_{\mathcal{X}_j}$) and

$$\begin{aligned}\mathbb{E}[\|\sqrt{n}(\hat{\mu}_n-\mu)\|_{\infty,\mathcal{F}_{I_j}}] &\le \mathbb{E}[\|\sqrt{n}(\hat{\mu}_n-\mu)\|_{\infty,\mathcal{F}_{\mathcal{X}_j}}] \\ &\lesssim M_j\mu(\mathcal{X}_j)^{1/2}\end{aligned}$$

up to constants that depend only on d,α, and $\sup_j \operatorname{diam}(\mathcal{X}_j)$. The RHS is summable over j so that by Theorem 1.1 in (var der Vaart, 1996), $\mathcal{F}$ is μ-Donsker. The bound on $\mathbb{E}[\|\hat{\mu}_n-\mu\|_{\infty,\mathcal{F}}]$ follows by summing up bounds on $\mathbb{E}[\|\hat{\mu}_n-\mu\|_{\infty,\mathcal{F}_{I_j}}]$. □

We are now in position to prove Theorem 3.

Proof of Theorem 3. Observe that

$$\big((\hat{\mu}_n-\mu)*\mathcal{N}_\sigma\big)(f)=(\hat{\mu}_n-\mu)(f*\varphi_\sigma). \tag{9}$$

and consider the function classes

$$\mathcal{F}=\big\{f\in C_0^\infty:\|f\|_{\dot{H}^{1,q}(\mathcal{N}_\sigma)}\le 1\big\} \tag{10}$$

$$\mathcal{F}*\varphi_\sigma=\big\{f*\varphi_\sigma:f\in\mathcal{F}\big\}. \tag{11}$$

We apply Lemma 8 to show that the function class $\mathcal{F}*\varphi_\sigma$ is μ-Donsker, implying the limit described in the theorem statement. Since for any constant $a\in\mathbb{R}$ and any function $f\in\mathcal{F}$, $(\hat{\mu}_n-\mu)(f*\varphi_\sigma)=(\hat{\mu}_n-\mu)\big((f-a)*\varphi_\sigma\big)$, we only have to verify the conditions of Lemma 8 for $\mathcal{F}*\varphi_\sigma$ with $\mathcal{F}$ replaced by $\{f-\mathcal{N}_\sigma(f):f\in C_0^\infty,\|f\|_{\dot{H}^{1,q}(\mathcal{N}_\sigma)}\le 1\}$.

We first construct a cover $\{\mathcal{X}_j\}_{j=1}^\infty$ as follows. Let $B_r=B(0,r)$. For $\delta>0$ fixed and $r=2,3,\dots$, let $\{x_1^{(r)},\dots,x_{N_r}^{(r)}\}$ be a minimal δ-net of $B_{r\delta}\setminus B_{(r-1)\delta}$. Set $x_1^{(1)}=0$ with $N_1=1$. To bound N_r, we show that the covering number $N(B_{r\delta}\setminus B_{(r-1)\delta},|\cdot|,\epsilon)$, defined as the size of the smallest ϵ-cover of $B_{r\delta}\setminus B_{(r-1)\delta}$, satisfies

$$N(B_{r\delta}\backslash B_{(r-1)\delta},|\cdot|,\epsilon)\le\left(\frac{2r\delta}{\epsilon}+1\right)^d-\left(\frac{2(r-1)\delta}{\epsilon}-1\right)^d \tag{12}$$

for $0<\epsilon\le 2(r-1)\delta$, according to a volumetric argument. Specifically, let $\{x_1,\dots,x_N\}$ be a maximal ϵ-separated subset of $B_{r\delta}\setminus B_{(r-1)\delta}$. By maximality, $\{x_1,\dots,x_N\}$ is an ϵ-net of $B_{r\delta}\setminus B_{(r-1)\delta}$. By construction,

$$\bigcup_{j=1}^N B(x_j,\epsilon/2)\subset B_{r\delta+\epsilon/2}\setminus B_{(r-1)\delta-\epsilon/2}$$

and the balls of the left-hand side (LHS) are disjoint. Comparing volumes, we have

$$N(\epsilon/2)^d\le(r\delta+\epsilon/2)^d-((r-1)\delta-\epsilon/2)^d.$$

This yields the bound on the covering number.

Given (12), we obtain $N_r\le(2r+1)^d-(2r-3)^d=O(r^{d-1})$. Set

$$\mathcal{X}_j=B(x_j^{(r)},\delta),\ j=\sum_{k=1}^{r-1}N_k+1,\dots,\sum_{k=1}^{r}N_k.$$

By construction, $\{\mathcal{X}_j\}_{j=1}^\infty$ forms a cover of $\mathbb{R}^d$ with diameter 2δ. Set $\alpha=\lfloor d/2\rfloor+1$ and $M_j=\sup_{f\in\mathcal{F}:\mathcal{N}_\sigma(f)=0}\|f*\varphi_\sigma\|_{C^\alpha(\mathcal{X}_j)}$. Fix any $\eta>0$. By Lemma 7,

$$\begin{aligned}&\max_{\sum_{k=1}^{r-1}N_k+1\le j\le\sum_{k=1}^{r}N_j} M_j\\ &\lesssim\sigma^{-\lfloor d/2\rfloor}\exp\left(\frac{(1+\eta)(p-1)r^2\delta^2}{2\sigma^2}\right)\end{aligned}$$

up to constants independent of r and σ. Hence, in view of Lemma 8, the μ-Donsker property of $\mathcal{F}*\varphi_\sigma$ holds if

$$\sum_{r=1}^\infty r^{d-1}\exp\left(\frac{(1+\eta)(p-1)r^2\delta^2}{2\sigma^2}\right)\sqrt{\mathbb{P}(|X|>(r-1)\delta)}$$

is finite. By Riemann approximation, the sum above can be bounded by δ^{-d-1} times

$$\int_1^\infty t^{d-1}\exp\left(\frac{(1+\eta)(p-1)t^2}{2\sigma^2}\right)\sqrt{\mathbb{P}(|X|>t-2\delta)}\mathrm{d}t$$

which is finite under our assumption by choosing η and δ sufficiently small, and absorbing t^{d-1} by the exponential term.

For more precise constants, we assume that μ is contained in a ball of radius R centered at the origin. Then, using the constants from the proof of Lemma 7 with $\eta=1$ and taking $\delta\le R/2$, we find that the $\sqrt{n}\,\mathbb{E}\left[\mathsf{d}_p^{(\sigma)}(\hat{\mu}_n,\mu)\right]$ is bounded by

$$\begin{aligned}&(C')^d d^{d/2}p^{3d/2}\sigma^{1-\bar{k}}4^{d-1}\exp\left(\frac{4(p-1)R^2}{\sigma^2}\right)\\ &\le(cdp^3\sigma^{-1})^{d/2}e^{pR^2\sigma^{-2}},\end{aligned}$$

for some absolute constant $c>0$, so long as $\sigma\le 1$, say. □

A.2.3. Proof of Corollary 2

The moment convergence of $\sqrt{n}\mathsf{d}_p^{(\sigma)}(\hat{\mu}_n,\mu)$ follows from Lemma 2.3.11 in (van der Vaart & Wellner, 1996). Finiteness of $\mathbb{E}[\|G\|_{\dot{H}^{-1,p}(\mathcal{N}_\sigma)}]$ follows from Proposition A.2.3 in (van der Vaart & Wellner, 1996). The second result follows from Theorem 1 after centering μ and $\hat{\mu}_n$ by the mean of μ. Plugging in the constant from the previous proof, we find that

$$\sqrt{n}\,\mathbb{E}\left[\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n,\mu)\right]\le(cdp^3\sigma^{-1})^{d/2}e^{pR^2\sigma^{-2}}$$

when μ is contained in a ball of radius R and $\sigma\le 1$, for some (different) constant $c>0$.

□

A.2.4. Proof of Proposition 4

Without loss of generality, we may assume that X has mean zero. If X is β-sub-Gaussian, then

$$\mathbb{E}[e^{\eta|X|^2}] \le \underbrace{(1-2\beta^2\eta)^{-d/2}}_{=C_\eta} \quad \text{if } \eta < 1/(2\beta^2).$$

By Markov's inequality, we have

$$\mathbb{P}(|X|>r) \le C_\eta e^{-\eta r^2}.$$

Thus,

$$\int_0^\infty e^{\frac{\theta r^2}{2\sigma^2}}\sqrt{\mathbb{P}(|X|>r)}\mathrm{d}r \le C_\eta^{1/2}\int_0^\infty e^{-\left(\eta-\frac{\theta}{\sigma^2}\right)\frac{r^2}{2}}\mathrm{d}r.$$

The right hand side is finite if and only if $\eta > \frac{\theta}{\sigma^2}$. Such η exists if and only if

$$\frac{1}{2\beta^2} > \frac{\theta}{\sigma^2}, \quad \text{i.e.,} \quad \beta < \frac{\sigma}{\sqrt{2\theta}}.$$

Sine $\theta > p-1$ is arbitrary, we obtain the desired conclusion. □

A.2.5. Proof of Proposition 5

Given the comparison result of Theorem 1 and our characterization of $\mathbb{E}[\mathsf{d}_p^{(\sigma)}(\hat{\mu}_n,\mu)]$ in the proof of Theorem 3, it suffices to prove

$$\Pr\Big(\mathsf{d}_p^{(\sigma)}(\hat{\mu}_n,\mu) \ge \mathbb{E}[\mathsf{d}_p^{(\sigma)}(\hat{\mu}_n,\mu)]+t\Big) \le e^{c'nt^2} \qquad (13)$$

for some constant $c'>0$ independent of n and t. We apply Corollary 1 of (Goldfeld & Greenewald, 2020), where the 1-Lipschitz function class Lip_1 is substituted with $\mathcal{F}_0 = \{f-\mathcal{N}_\sigma(f) : f\in C_0^\infty, \|f\|_{\dot{H}^{1,q}(\mathcal{N}_\sigma)} \le 1\}$. The desired conclusion follows according to the same argument, using McDiarmid's inequality, upon observing that for $x,x' \in \mathsf{supp}(\mu)$,

$$\begin{aligned}&\sup_{f\in\mathcal{F}_0}(f*\varphi_\sigma)(x)-(f*\varphi_\sigma)(x')\\ &\le 2\sup_{f\in\mathcal{F}_0, y\in\mathsf{supp}(\mu)}(f*\varphi_\sigma)(y)\\ &\le 2D_q(\mathcal{N}_\sigma)\exp\left(\frac{(p-1)R^2}{2\sigma^2}\right),\end{aligned}$$

where the final inequality uses Lemma 7. □

A.3. Proofs for Section 5

First, we comment on a subtle detail regarding the construction of the homogeneous Sobolev space.

Remark 4. For $\gamma\in\mathcal{P}$ dominating the Lebesgue measure and satisfying the p-Poincaré inequality, the homogeneous Sobolev space $\dot{H}^{1,p}(\gamma)$ can be constructed as a function space over $\mathbb{R}^d$ that contains $\dot{C}_0^\infty$ as a dense subset in an explicit manner (without relying on the completion, which is an abstract metric-topological operation). See Appendix A.6 for details of the construction.

Next, we observe that the inner product on $\dot{H}_0^{1,2}(\mathcal{N}_\sigma)*\varphi_\sigma$ is well-defined. That is, for $f,g\in\dot{H}_0^{1,2}(\mathcal{N}_\sigma)$, we show that $f*\varphi_\sigma = g*\varphi_\sigma$ if and only if $f=g$ almost everywhere. This requires an application of Wiener's Tauberian theorem for $L^2(\mathbb{R}^d)$, with a proof provided for completeness.

Theorem 5 (Wiener's Tauberian theorem for L^2)**.** *If the Fourier transform $\mathsf{F}[f]$ of $f\in L^2(\mathbb{R}^d)$ never vanishes, then the span of the set of translates $\{f_a : f_a(x)=f(a+x), a\in\mathbb{R}^d\}$ is dense in $L^2(\mathbb{R}^d)$.*

Proof. Suppose that $g\in L^2(\mathbb{R}^d)$ is orthogonal to all translates of f. Then, because F is a unitary operator on $L^2(\mathbb{R}^d)$,

$$\begin{aligned}0 &= \int_{\mathbb{R}^d} g(x)f_a(x)\mathrm{d}x\\ &= \int_{\mathbb{R}^d}\mathsf{F}[g](p)\mathsf{F}[f_a](p)\,\mathrm{d}p\\ &= \int_{\mathbb{R}^d} e^{iap}\mathsf{F}[g](p)\mathsf{F}[f](p)\,\mathrm{d}p\end{aligned}$$

for all $a\in\mathbb{R}^d$. Equivalently, we have

$$\mathsf{F}[\mathsf{F}[g]\cdot\mathsf{F}[f]](-a)=0$$

for all $a\in\mathbb{R}^d$. That is, $\mathsf{F}[\mathsf{F}[g]\cdot\mathsf{F}[f]]=0$. Since F is injective, and $\mathsf{F}[f]$ never vanishes, we have $g=0$, implying the desired density result. □

Lemma 9 (Well-definedness of inner product)**.** *For $f\in\dot{H}_0^{1,2}(\mathcal{N}_\sigma)$, $f*\varphi_\sigma=0$ if and only if $f=0$ almost everywhere.*

Proof. By the previous remark, we can consider f as an element of $L^2(\mathcal{N}_\sigma)$. The "if" direction is trivial. For the other direction, recall that f can be realized as the limit in $L^2(\mathcal{N}_\sigma)$ of a sequence $\{f_n\}_{n\in\mathbb{N}}$ of simple functions with compact support. If $f*\varphi_\sigma=0$, we have for any $y\in\mathbb{R}^d$ that

$$\begin{aligned}&\left|\int_{\mathbb{R}^d} f_n(x)\varphi_\sigma(y-x)\,\mathrm{d}x\right|\\ =&\left|\int_{\mathbb{R}^d}(f_n-f)(x)\varphi_\sigma(y-x)\,\mathrm{d}x\right|\\ \le&\|f-f_n\|_{L^2(\mathcal{N}_\sigma)}\int_{\mathbb{R}^d}\frac{\varphi_\sigma(y-x)^2}{\varphi_\sigma(x)}\mathrm{d}x\to 0\end{aligned}$$

as $n \to \infty$. Because the Fourier transform of φ_σ never vanishes, Theorem 5 implies that the span of the functions $\varphi_\sigma(y-\cdot)$ is dense in $L^2(\mathbb{R}^d)$; thus, $\langle f_n, g\rangle_{L^2(\mathbb{R}^d)} \to 0$ for any $g \in L^2(\mathbb{R}^d)$. That is, the sequence f_n converges weakly to 0 in $L^2(\mathbb{R}^d)$. Hence, f_n must converge weakly to 0 in $L^2(\mathcal{N}_\sigma)$ as well (since the density φ_σ is bounded). Seeing as f is the ordinary limit of f_n in $L^2(\mathcal{N}_\sigma)$, it must therefore coincide with the weak limit of 0. □

Now, since functions which are equal almost everywhere have the same convolution with φ_σ, this implies that $\dot{H}^{1,2}_0(\mathcal{N}_\sigma) * \varphi_\sigma$ is realizable as a Hilbert space of functions (not equivalence classes of functions), the most basic requirement for the RKHS property.

Next, we prove a lemma which allows us to concentrate on $\sigma = 1$ without loss of generality.

Lemma 10 (Unit smoothing parameter)**.** *For $\mu, \nu \in \mathcal{P}_p$, let $X \sim \mu$ and $Y \sim \nu$. Then,*

$$\mathsf{d}_p^{(\sigma)}(\mu,\nu) = \sigma\,\mathsf{d}_p^{(1)}(\mu',\nu'),$$

where μ' and ν' are the distributions of X/σ and Y/σ, respectively.

Proof of Lemma 10. First, define the isometric isomorphism $T : \dot{H}^{1,q}(\mathcal{N}_\sigma) \to \dot{H}^{1,q}(\mathcal{N}_1)$ by $(Tf)(x) = \sigma^{-1} f(\sigma x)$. We verify

$$\begin{aligned}\int_{\mathbb{R}^d} |\nabla_x (Tf)(x)|^q \,\mathrm{d}\mathcal{N}_1(x) &= \int_{\mathbb{R}^d} |\sigma^{-1}\nabla_x f(\sigma x)|^q \varphi_1(x)\,\mathrm{d}x \\ &= \int_{\mathbb{R}^d} |\nabla f(\sigma x)|^q \varphi_1(x)\,\mathrm{d}x \\ &= \int_{\mathbb{R}^d} |\nabla f(u)|^q \,\mathrm{d}\mathcal{N}_\sigma(u).\end{aligned}$$

Taking independent $X \sim \mu, Y \sim \nu$, and $Z \sim \mathcal{N}_1$ and noting that $f(x) = \sigma\, Tf(x/\sigma)$, we have

$$\begin{aligned}(\mu-\nu)(f * \varphi_\sigma) &= \mathbb{E}\left[f(X+\sigma Z) - f(Y+\sigma Z)\right] \\ &= \sigma \cdot \mathbb{E}\left[Tf(X/\sigma + Z) - Tf(Y/\sigma + Z)\right] \\ &= \sigma \cdot (\mu'-\nu')(Tf * \varphi_1),\end{aligned}$$

where μ' and ν' are the distributions of X/σ and Y/σ, respectively. Thus,

$$\begin{aligned}\mathsf{d}_p^{(\sigma)}(\mu,\nu) &= \sup_{\substack{f \in \dot{H}^{1,q}(\mathcal{N}_\sigma) \\ \|f\|_{\dot{H}^{1,q}(\mathcal{N}_\sigma)} \le 1}} (\mu-\nu)(f*\varphi_\sigma) \\ &= \sigma \sup_{\substack{Tf \in \dot{H}^{1,q}(\mathcal{N}_1) \\ \|Tf\|_{\dot{H}^{1,q}(\mathcal{N}_1)} \le 1}} (\mu'-\nu')(Tf*\varphi_1) \\ &= \sigma\,\mathsf{d}_p^{(1)}(\mu',\nu').\end{aligned}$$

This completes the proof. □

Next, we identify an orthonormal basis of $\dot{H}^{1,2}_0(\mathcal{N}_1) * \varphi_1$. We first prove that Hermite polynomials form an orthonormal basis of $\dot{H}^{1,2}_0(\mathcal{N}_1)$, and then translate this to an orthonormal basis of $\dot{H}^{1,2}_0(\mathcal{N}_1) * \varphi_1$. Here, for $k \in \mathbb{N}_0^d$, we write $x^k := \prod_{i=1}^d x_i$ and $\bar{k} := \sum_{i=1}^d k_i$.

Lemma 11 (Orthonormal basis of $\dot{H}^{1,2}_0(\mathcal{N}_1) * \varphi_1$)**.** *The monomials $\phi_k(x) = (\bar{k}\prod_{i=1}^d k_i)^{-1/2} x^k$, $0 \neq k \in \mathbb{N}_0^d$, comprise an orthonormal basis of $\dot{H}^{1,2}_0(\mathcal{N}_1) * \varphi_1$.*

Proof of Lemma 11. Recall that the Hermite polynomials defined as

$$\mathrm{He}_n(x) = (-1)^n e^{x^2/2} \frac{\mathrm{d}^n}{\mathrm{d}x^n} e^{x^2/2}$$

satisfy $\mathrm{He}_n'(x) = n\,\mathrm{He}_{n-1}(x)$ and $\int \mathrm{He}_n \mathrm{He}_m \,\mathrm{d}\mathcal{N}_1 = n!\,\delta_{n,m}$ (Bogachev, 1998). They admit a natural multivariate extension

$$\mathrm{He}_k(x) = \prod_{i=1}^d \mathrm{He}_{k_i}(x_i), \quad k \in \mathbb{N}_0^d,$$

which satisfies

$$\begin{aligned}\langle \mathrm{He}_k, \mathrm{He}_{k'}\rangle_{\dot{H}^{1,2}(\mathcal{N}_1)} &= \int \langle \nabla \mathrm{He}_k, \nabla \mathrm{He}_{k'}\rangle \mathrm{d}\mathcal{N}_1 \\ &= \sum_{i=1}^d \int \frac{\partial \mathrm{He}_k}{\partial x_i}\frac{\partial \mathrm{He}_{k'}}{\partial x_i}\mathrm{d}\mathcal{N}_1 \\ &= \delta_{k,k'}\bar{k}\prod_{i=1}^d k_i!.\end{aligned}$$

Thus, the normalized polynomials $\widetilde{\mathrm{He}}_k := \left(\bar{k}\prod k_i!\right)^{-1/2}\mathrm{He}_k$, $0 \neq k \in \mathbb{N}_0^d$, form an orthonormal set, and it is easy to check that they span the space of d-variate polynomials Q with $\mathcal{N}_1(Q) = 0$. By Proposition 1.3 of (Schmuland, 1992), polynomials are dense in the inhomogeneous Gaussian Sobolev space $H^{1,2}(\mathcal{N}_1)$, and hence $\dot{H}^{1,2}(\mathcal{N}_1)$, so it follows that the $\widetilde{\mathrm{He}}_k$ polynomials form an orthonormal basis for $\dot{H}^{1,2}_0(\mathcal{N}_1)$.

Next, we observe that, in one dimension, $(\mathrm{He}_n * \varphi_1)(x) = x^n$. To see this, we use the Rodrigues formula for the Hermite polynomials (Rasala, 1981), which states that $\mathrm{He}_n(x) = e^{-D^2/2}[x^n]$. Here, D is the differentiation operator and $\exp$ is defined on operators via its formal power series (working with polynomials, there are no issues of convergence). We can express convolution with a standard Gaussian in a similar way, with $f * \varphi_1 = e^{D^2/2} f$ (where it suffices to consider only f that are polynomials) (Bilodeau, 1962). Together, these reveal that $(\mathrm{He}_n * \varphi_1)(x) = x^n$. Thus, for $0 \neq k \in \mathbb{N}_0^d$, we obtain

$$(\widetilde{\mathrm{He}}_k * \varphi_1)(x) = \left(\bar{k}\prod k_i!\right)^{-1/2} x^k =: \phi_k(x).$$

Since the $\widetilde{\mathrm{He}}_k$ polynomials form an orthonormal basis for $\dot{H}_0^{1,2}(\mathcal{N}_1)$, the ϕ_k monomials form an orthonormal basis for $\dot{H}_0^{1,2}(\mathcal{N}_1) * \varphi_1$, as claimed. □

Now, the theorem follows via routine calculations.

A.3.1. Proof of Theorem 4

By Lemma 7, we have that for any $f \in \dot{H}_0^{1,2}(\mathcal{N}_\sigma)$,

$$\begin{aligned} |(f * \varphi_\sigma)(x)| &\le D_2(\mathcal{N}_\sigma) e^{|x|^2/(2\sigma^2)} \|\nabla f\|_{L^2(\mathcal{N}_\sigma)} \\ &= e^{|x|^2/(2\sigma^2)} \|f * \varphi_\sigma\|_{\dot{H}^{1,2}(\mathcal{N}_\sigma)*\varphi_\sigma}, \end{aligned}$$

so pointwise evaluation at x is a bounded linear operator on $\dot{H}_0^{1,2}(\mathcal{N}_\sigma) * \varphi_\sigma$ for each $x \in \mathbb{R}^d$. This implies that $\dot{H}_0^{1,2}(\mathcal{N}_\sigma) * \varphi_\sigma$ is an RKHS over $\mathbb{R}^d$. For $\sigma = 1$, we can compute the reproducing kernel from the orthonormal basis above (see Theorem 4.20 of (Steinwart & Christmann, 2008)) as

$$\begin{aligned} \kappa^{(1)}(x, y) &= \sum_{0 \ne k \in \mathbb{N}_0^d} \phi_k(x)\phi_k(y) \\ &= \sum_{0 \ne k \in \mathbb{N}_0^d} \Big(|k| \prod k_i!\Big)^{-1} x^k y^k \\ &= \sum_{n=1}^{\infty} \frac{1}{n \cdot n!} \sum_{|k|=n} \frac{n!}{\prod k_i!} x^k y^k \\ &= \sum_{n=1}^{\infty} \frac{1}{n \cdot n!} \langle x, y \rangle^n = -\operatorname{Ein}(-\langle x, y \rangle). \end{aligned}$$

We note that $\kappa^{(1)}$ is positive semi-definite by this construction. The MMD formulation (5) follows because

$$\begin{aligned} &\mathsf{d}_2^{(1)}(\mu, \nu) \\ &= \sup\Big\{\mu(f * \varphi_1) - \nu(f * \varphi_1) : \\ &\qquad f \in \dot{H}_0^{1,2}(\mathcal{N}_1), \|f\|_{\dot{H}^{1,2}(\mathcal{N}_1)} \le 1\Big\} \\ &= \sup\Big\{\mu(g) - \nu(g) : \\ &\qquad g \in \dot{H}_0^{1,2}(\mathcal{N}_1) * \varphi_1, \|g\|_{\dot{H}^{1,2}(\mathcal{N}_1)*\varphi_1} \le 1\Big\} \\ &= \mathsf{MMD}_{\dot{H}_0^{1,2}(\mathcal{N}_1)*\varphi_1}(\mu, \nu) \end{aligned}$$

and the RHS of (5) is the standard kernel formulation of an MMD (Gretton et al., 2012). The extension to general σ follows from Lemma 10 and the uniqueness of the reproducing kernel. □

A.4. Proofs for Section 6

A.4.1. Proof of Proposition 7

We first consider the size control. Suppose that $\mu = \nu$. Without loss of generality, we may assume that μ is not a point mass. To handle shifts of distributions, for any $a \in \mathbb{R}^d$, we represent

$$\begin{aligned} &\sqrt{\frac{mn}{N}} \mathsf{d}_p^{(\sigma)}(\hat{\mu}_m * \delta_{-a}, \hat{\nu}_n * \delta_{-a}) \\ &= \Bigg\| \sqrt{\frac{n}{N}} \sqrt{m}(\hat{\mu}_m - \mu)(f(\cdot - a) * \varphi_\sigma) \\ &\qquad - \sqrt{\frac{m}{N}} \sqrt{n}(\hat{\nu}_n - \mu)(f(\cdot - a) * \varphi_\sigma) \Bigg\|_{\infty, \mathcal{F}}, \end{aligned}$$

where the function class $\mathcal{F} = \{f \in C_0^\infty : \|f\|_{\dot{H}^{1,q}(\mathcal{N}_\sigma)} \le 1\}$ is the one from the proof of Theorem 3. Consider another function class

$$\mathcal{F}_{\mathsf{shift}} = \{f(\cdot - a) : f \in C_0^\infty, \|f\|_{\dot{H}^{1,q}(\mathcal{N}_\sigma)} \le 1, |a| \le C\}$$

for some large enough constant $C < \infty$ such that the mean a_μ of μ satisfies $|a_\mu| < C$. It is not difficult to see from the proof of Theorem 3 that the function class $\mathcal{F}_{\mathsf{shift}} * \varphi_\sigma$ is μ-Donsker, which implies that (cf. Theorem 1.5.7 in (van der Vaart & Wellner, 1996))

$$\begin{aligned} \limsup_{m \to \infty} \mathbb{P}\Bigg(&\sup_{\substack{f \in C_0^\infty \\ \|f\|_{\dot{H}^{1,q}(\mathcal{N}_\sigma)} \le 1 \\ |a-b| < \delta}} \big|\sqrt{m}(\hat{\mu}_m - \mu)\big((f(\cdot - a) \\ &\qquad - f(\cdot - b)) * \varphi_\sigma\big)\big| > \epsilon \Bigg) \to 0 \end{aligned}$$

as $\delta \to 0$, for all $\epsilon > 0$. Here we used the fact that $|a-b| \to 0$ implies that $\operatorname{Var}\big((f(X-a) - f(X-b)) * \varphi_\sigma\big) \to 0$. Since $\bar{X}_m \to a_\mu$ a.s. by the law of large numbers, we have

$$\begin{aligned} \sup_{\substack{f \in C_0^\infty \\ \|f\|_{\dot{H}^{1,q}(\mathcal{N}_\sigma)} \le 1}} &\big|\sqrt{m}(\hat{\mu}_m - \mu)\big((f(\cdot - \bar{X}_m) \\ &\qquad - f(\cdot - a_\mu)) * \varphi_\sigma\big)\big| \xrightarrow{\mathbb{P}} 0. \end{aligned}$$

A similar result holds for $\hat{\nu}_n$. Now, by Theorem 3, we have

$$\begin{aligned} W_{m,n} \le p &\sqrt{\frac{mn}{N}} \\ &\min\Big\{e^{\operatorname{tr}\hat{\Sigma}_X/(2q\sigma^2)} \mathsf{d}_p^{(\sigma)}(\hat{\mu}_m * \delta_{-\bar{X}_m}, \hat{\nu}_n * \delta_{-\bar{X}_m}), \\ &\qquad e^{\operatorname{tr}\hat{\Sigma}_Y/(2q\sigma^2)} \mathsf{d}_p^{(\sigma)}(\hat{\mu}_m * \delta_{-\bar{Y}_n}, \hat{\nu}_n * \delta_{-\bar{Y}_n})\Big\}, \end{aligned} \tag{14}$$

In view of this inequality, together with the fact that $\hat{\Sigma}_X \to \Sigma_\mu$ and $\hat{\Sigma}_Y \to \Sigma_\mu$ a.s., where Σ_μ is the covariance matrix of μ, we conclude that $W_{m,n}$ is at most

$$p e^{\operatorname{tr}\Sigma_\mu/(2q\sigma^2)} \sqrt{\frac{mn}{N}} \mathsf{d}_p^{(\sigma)}(\hat{\mu}_m * \delta_{-a_\mu}, \hat{\nu}_n * \delta_{-a_\mu}) + o_{\mathbb{P}}(1).$$

Now, the function class $\mathcal{F} * \varphi_\sigma$ is Donsker w.r.t. $\mu * \delta_{-a_\mu}$, so that from p. 361 of (van der Vaart & Wellner, 1996), we have

$$\sqrt{\frac{mn}{N}}\mathsf{d}_p^{(\sigma)}(\hat{\mu}_m * \delta_{-a_\mu}, \hat{\nu}_n * \delta_{-a_\mu}) \stackrel{d}{\to} \|G\|_{\mathcal{F},\infty},$$

where G is the Gaussian process that appears in Theorem 3 with μ replaced by $\mu * \delta_{-a_\mu}$.

Is is easy to show that the distribution function of $\|G\|_{\dot{H}^{-1,p}(\mathcal{N}_\sigma)}$ is continuous (cf. the proof of Lemma 3 in (Goldfeld et al., 2020a)), so long as μ is not a point mass (in which case the proposition is trivially true). To show that the test has asymptotic level α, it then suffices to show that (cf. Lemma 23.3 in (van der Vaart, 1998))

$$\begin{aligned}\mathbb{P}^B\left(\sqrt{\frac{mn}{N}}\mathsf{d}_p^{(\sigma)}\left(\hat{\mu}_m^B * \delta_{-\bar{Z}_N}, \hat{\nu}_n^B * \delta_{-\bar{Z}_N}\right) \le t\right)&\\ \stackrel{\mathbb{P}}{\to} \mathbb{P}\left(\|G\|_{\dot{H}^{-1,p}(\mathcal{N}_\sigma)} \le t\right), \quad \forall t \ge 0.&\end{aligned} \tag{15}$$

Observe that

$$\begin{aligned}&\sqrt{\frac{mn}{N}}\mathsf{d}_p^{(\sigma)}\left(\hat{\mu}_m^B * \delta_{-\bar{Z}_N}, \hat{\nu}_n^B * \delta_{-\bar{Z}_N}\right)\\ &= \Bigg\| \sqrt{\frac{n}{N}}\sqrt{m}\big(\hat{\mu}_m^B - \hat{\gamma}_N\big)\big(f(\cdot - \bar{Z}_N) * \varphi_\sigma\big)\\ &\qquad - \sqrt{\frac{m}{N}}\sqrt{n}\big(\hat{\nu}_n^B - \hat{\gamma}_N\big)\big(f(\cdot - \bar{Z}_N) * \varphi_\sigma\big)\Bigg\|_{\infty,\mathcal{F}}.\end{aligned}$$

Since the function class $\mathcal{F}_{\mathsf{shift}} * \varphi_\sigma$ is μ-Donsker, by Theorem 3.6.1 in (van der Vaart & Wellner, 1996), the bootstrap process $\sqrt{m}(\hat{\mu}_m^B - \hat{\gamma}_N)$ indexed by $\mathcal{F}_{\mathsf{shift}} * \varphi_\sigma$ converges in distribution in $\ell^\infty(\mathcal{F}_{\mathsf{shift}} * \varphi_\sigma)$ *unconditionally*, which implies that

$$\begin{aligned}\limsup_{m,n\to\infty}\mathbb{P}\Bigg(\sup_{\substack{f\in C_0^\infty\\ \|f\|_{\dot{H}^{1,q}(\mathcal{N}_\sigma)}\le 1\\ |a-b|<\delta}} \big|\sqrt{m}\big(\hat{\mu}_m^B - \hat{\gamma}_N\big)\big((f(\cdot - a)\\ - f(\cdot - b)) * \varphi_\sigma\big)\big| > \epsilon\Bigg) \to 0\end{aligned}$$

as $\delta \to 0$, for all $\epsilon > 0$. Since $\bar{Z}_N \to a_\mu$ a.s. by the law of large numbers, we have

$$\begin{aligned}\sup_{f\in C_0^\infty, \|f\|_{\dot{H}^{1,q}(\mathcal{N}_\sigma)}\le 1} \big|\sqrt{m}\big(\hat{\mu}_m^B - \hat{\gamma}_N\big)\big((f(\cdot - \bar{Z}_N)\\ - f(\cdot - a_\mu)) * \varphi_\sigma\big)\big| \stackrel{\mathbb{P}}{\to} 0.\end{aligned}$$

An analogous result holds for $\hat{\nu}_n^B$. Thus, we have

$$\begin{aligned}&\sqrt{\frac{mn}{N}}\mathsf{d}_p^{(\sigma)}(\hat{\mu}_m^B * \delta_{-\bar{Z}_N}, \hat{\nu}_n^B * \delta_{-\bar{Z}_N})\\ &= \sqrt{\frac{mn}{N}}\mathsf{d}_p^{(\sigma)}(\hat{\mu}_m^B * \delta_{-a_\mu}, \hat{\nu}_n^B * \delta_{-a_\mu}) + o_{\mathbb{P}}(1).\end{aligned}$$

The desired conclusion (15) follows from Theorem 3.7.6 in (van der Vaart & Wellner, 1996) combined with the fact that the function class $\mathcal{F} * \varphi_\sigma$ is $\mu * \delta_{-a_\mu}$-Donsker.

To show asymptotic consistency, suppose that $\mu \ne \nu$ and note that the preceding argument and Theorem 3.7.6 in (van der Vaart & Wellner, 1996) imply that

$$\mathbb{P}^B\left(W_{m,n}^B \le t\right) \stackrel{\mathbb{P}}{\to} \mathbb{P}\left(p e^{\operatorname{tr}\Sigma_\gamma/(2q\sigma^2)}\|G_\gamma\|_{\dot{H}^{-1,p}(\mathcal{N}_\sigma)} \le t\right)$$

for all $t \ge 0$, where Σ_γ is the covariance matrix of the measure $\gamma = \tau\mu + (1-\tau)\nu$ and G_γ is the Gaussian process from Theorem 3 with μ replaced by $\gamma * \delta_{-a_\gamma}$ (a_γ is the mean vector of γ). Furthermore, it is not difficult to see that $W_{m,n} \stackrel{\mathbb{P}}{\to} \infty$ under the alternative, which implies that $\mathbb{P}\left(W_{m,n} > w_{m,n}^B(1-\alpha)\right) \to 1$ whenever $\mu \ne \nu$. □

Propositions 7,8, and 9 follow from essentially similar proofs to those in (Goldfeld et al., 2020a), which build on (Bernton et al., 2019) and (Pollard, 1980), with arbitrary $p \ge 1$ instead of the $p = 1$ considered therein (indeed, the needed results from (Villani, 2008) hold for all $1 \le p < \infty$), so we omit their proofs for brevity.

A.4.2. Proof of Corollary 3

First, we state a simple lemma to bound generalization error of minimum distance estimation w.r.t. an IPM in terms of the empirical approximation error.

Lemma 12 (Generalization error for GANs)**.** *For an IPM* d *and an estimator* $\hat{\theta}_n \in \Theta$ *with* $\mathsf{d}(\hat{\mu}_n, \nu_{\hat{\theta}_n}) \le \inf_{\theta\in\Theta}\mathsf{d}(\hat{\mu}_n, \nu_\theta) + \epsilon$*, we have*

$$\mathsf{d}(\mu, \nu_{\hat{\theta}_n}) - \inf_{\theta\in\Theta}\mathsf{d}(\mu, \nu_\theta) \le 2\,\mathsf{d}(\mu, \hat{\mu}_n) + \epsilon.$$

This is a consequence of the triangle inequality, see (Zhang et al., 2018) for example. Hence, our conclusion follows upon noting that

$$\begin{aligned}\mathbb{P}\left(2\mathsf{W}_p^{(\sigma)}(\mu, \hat{\mu}_n) > t\right) &\le \mathbb{P}\left(\mathsf{d}_p^{(\sigma)}(\mu, \hat{\mu}_n) > Ct\right)\\ &\le \exp\Big(-n(Ct - C'n^{-1/2})^2\Big)\\ &\le C_1\exp\big(-C_2 n t^2\big),\end{aligned}$$

where constants C, C', C_1, C_2 are independent of n and t. Here, we have combined the concentration result (13), the comparison from Theorem 1, and the fast rate from Corollary 2. □

A.5. Additional Details for Experiments

In Figure 5, we present additional S-MWE experiments for a single Gaussian parameterized by mean and variance, demonstrating similar limiting behavior to the mixture results provided in the main text.

We note that experiments for Figures 1, 3, and 5 were performed on a Dell OptiPlex 7050 PC with 32GB RAM and an 8 core 2.80GHz Intel Core i7 CPU, running in approximately 3 hours, 30 minutes, and 30 minutes, respectively. Computations for Figure 3 were performed on a cluster instance with 14 vCPUs and 112 GB RAM over several hours. Those for Figure 4 were performed on a cluster machine with 14 vCPUs, 60 GB RAM, and a NVIDIA Tesla V100 over nearly 12 hours (hence the restriction to low dimensions).

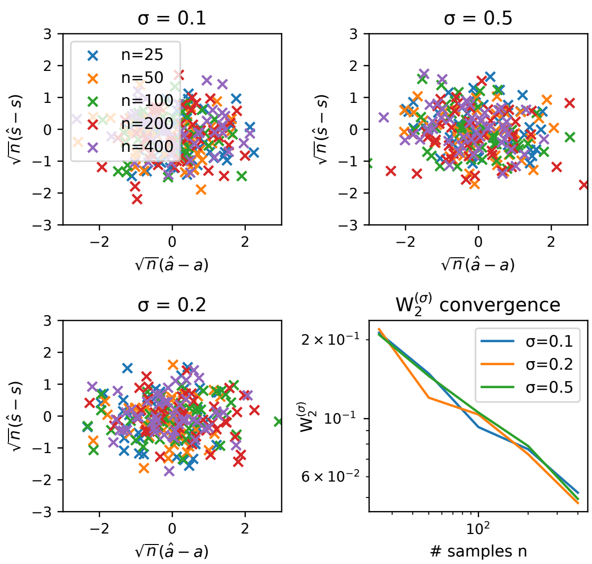

Figure 5. One-dimensional limiting behavior of M-SWE estimates for the mean and standard deviation parameters of $\mu = \mathcal{N}(a, s)$ with $a = 0$ and $s = 1$. Also shown is a log-log plot of $W_2^{(\sigma)}$ convergence in n.

Finally, we describe how the upper bound on $W_2^{(\sigma)}$ was computed for the rightmost plot of Figure 1.

A.5.1. Upper Bound on $\mathbb{E}\left[W_2^{(\sigma)}(\hat{\mu}_n, \mu)\right]$ using $d_2^{(\sigma)}$

By Theorem 4, we have

$$d_2^{(\sigma)}(\hat{\mu}_n, \mu)^2 = \mathbb{E}\left[\kappa^{(\sigma)}(X, X')\right] + \frac{1}{n^2}\sum_{i,j=1}^{n} \kappa^{(\sigma)}(X_i, X_j) - \frac{2}{n}\sum_{i=1}^{n} \mathbb{E}\left[\kappa^{(\sigma)}(X, X_i)\right],$$

where $X, X' \sim \mu$ are independent. Taking expectations, we obtain

$$\mathbb{E}\left[d_2^{(\sigma)}(\hat{\mu}_n, \mu)^2\right] = \mathbb{E}\left[\kappa^{(\sigma)}(X, X')\right] + \frac{1}{n}\mathbb{E}\left[\kappa^{(\sigma)}(X, X)\right] + \left(1 - \frac{1}{n}\right)\mathbb{E}\left[\kappa^{(\sigma)}(X, X')\right] - \frac{2}{n}\sum_{i=1}^{n}\mathbb{E}\left[\kappa^{(\sigma)}(X, X_i)\right] = \frac{1}{n}\left(\mathbb{E}\left[\kappa^{(\sigma)}(X, X)\right] - \mathbb{E}\left[\kappa^{(\sigma)}(X, X')\right]\right).$$

Combining this with Theorem 1, we reach the upper bound

$$\mathbb{E}\left[W_2^{(\sigma)}(\hat{\mu}_n, \mu)\right] \le 2e^{\mathbb{E}[|X|^2]/(4\sigma^2)} \left(\mathbb{E}\left[\kappa^{(\sigma)}(X, X)\right] - \mathbb{E}\left[\kappa^{(\sigma)}(X, X')\right]\right)^{1/2} n^{-1/2}.$$

For Figure 1, we estimate the kernel expectations via Monte Carlo integration with 1,000,000 samples. The kernel itself is computed via standard series-based methods for exponential integrals.

A.6. Explicit Construction of the Homogeneous Sobolev space

Let $\gamma \in \mathcal{P}$ be dominating the Lebesgue measure and satisfying the p-Poincaré inequality. Consider the homogeneous Sobolev space $\dot{H}^{1,p}(\gamma)$, which is constructed in Section 2 as the completion of $\dot{C}_0^\infty$ w.r.t. $\|\cdot\|_{\dot{H}^{1,p}(\gamma)}$. As such, it is not clear that the obtained space is a function space over $\mathbb{R}^d$. To show this is nevertheless the case, we present an explicit construction of $\dot{H}^{1,p}(\gamma)$ that does not rely on the completion.

Let $\mathcal{C} = \{f \in \dot{C}_0^\infty : \gamma(f) = 0\}$. Then, $\|\cdot\|_{\dot{H}^{1,p}(\gamma)}$ is a proper norm on $\mathcal{C}$, and the map $\iota : f \mapsto \nabla f$ is an isometry from $(\mathcal{C}, \|\cdot\|_{\dot{H}^{1,p}(\gamma)})$ into $(L^p(\gamma;\mathbb{R}^d), \|\cdot\|_{L^p(\gamma;\mathbb{R}^d)})$. Let V be the closure of $\iota\mathcal{C}$ in $L^p(\gamma;\mathbb{R}^d)$ under $\|\cdot\|_{L^p(\gamma;\mathbb{R}^d)}$. The inverse map $\iota^{-1} : \iota\mathcal{C} \to \mathcal{C}$ can be extended to V. Indeed, for any $g \in V$, choose $f_n \in \mathcal{C}$ such that $\|\nabla f_n - g\|_{L^p(\gamma;\mathbb{R}^d)} \to 0$. Since ∇f_n is Cauchy in $L^p(\gamma;\mathbb{R}^d)$, f_n is Cauchy in $L^p(\gamma)$ by the p-Poincaré inequality, so $\|f_n - f\|_{L^p(\gamma)} \to 0$ for some $f \in L^p(\gamma)$. Set $\iota^{-1}g = f$ and extend $\|\cdot\|_{\dot{H}^{1,p}(\gamma)}$ by $\|f\|_{\dot{H}^{1,p}(\gamma)} = \lim_{n\to\infty}\|f_n\|_{\dot{H}^{1,p}(\gamma)}$. The space $(\iota^{-1}V, \|\cdot\|_{\dot{H}^{1,p}(\gamma)})$ is a Banach space of functions over $\mathbb{R}^d$.

The homogeneous Sobolev space $\dot{H}^{1,p}(\gamma)$ is now constructed as $\dot{H}^{1,p}(\gamma) = \left\{f + a : a \in \mathbb{R}, f \in \iota^{-1}V\right\}$ with $\|f + a\|_{\dot{H}^{1,p}(\gamma)} = \|f\|_{\dot{H}^{1,p}(\gamma)}$.