# Convergence of Gaussian-smoothed optimal transport distance with sub-gamma distributions and dependent samples
# Supplementary Materials

## A Upper Bounds on GOT

This section provides proofs of the upper bounds on the GOT provided in Section 3. For the convenience of the reader we repeat some of the necessary definitions. Recall that the feature map $\psi_x \colon \mathbb{R}^d \to [0,\infty)$ is defined by

$$\psi_x(z) := \frac{\sqrt{\omega_d}}{2^{\frac{d+p}{2}}} \frac{\|z\|^{\frac{d-1+2p}{2}}}{\sqrt{f(\|z\|)}} \phi\Big(\frac{z}{\sqrt{2}} - \frac{x}{\sigma}\Big), \tag{S.1}$$

where $\omega_d = 2\pi^{d/2}/\Gamma(d/2)$ is the volume of the unit sphere in $\mathbb{R}^d$, $\phi(u) = (2\pi)^{-d/2}\exp(-\frac{1}{2}\|u\|^2)$ is the standard Gaussian density on $\mathbb{R}^d$, and $f$ is a probability density function on $[0,\infty)$ that satisfies

$$f(x) \geq a x^{d+2p-1} \exp(-bx^2), \tag{S.2}$$

for some $a > 0$ and $b \in (0, 1/2)$.

**Lemma S.1.** *The feature map in* (S.1) *defines a positive semidefinite kernel* $k \colon \mathbb{R}^n \times \mathbb{R}^n \to [0,\infty)$ *according to*

$$k(x,y) := \int_{\mathbb{R}^d} \psi_x(z)\psi_y(z)\,dz. \tag{S.3}$$

*Furthermore, this kernel can also be expressed as*

$$k(x,y) = \exp\Big(-\frac{\|x-y\|^2}{4\sigma^2}\Big) I_f\Big(\frac{\|x+y\|}{\sqrt{2}\sigma}\Big), \tag{S.4}$$

*where*

$$I_f(u) := \frac{\omega_d}{2^{d+p}(2\pi)^{\frac{d}{2}}} \int_0^\infty \frac{x^{d-1+2p}}{f(x)} g_{d,u}(x)\,dx, \tag{S.5}$$

*and* $g_{d,u}(x)$ *is the density of* $\|Z\|$ *when* $Z \sim \mathcal{N}(\mu, I_d)$ *with* $\|\mu\| = u$.

*Proof.* First we establish that $\psi_x$ is square integrable. By the assumed lower bound in (S.2) and the fact that $\phi^2(y/\sqrt{2}) = (2\pi)^{d/2}\phi(y)$, we can write

$$\int_{\mathbb{R}^d} |\psi_x(z)|^2\,dz \leq \frac{C_{d,p}}{a} \int_{\mathbb{R}^d} \exp\big(-b\|z\|^2\big)\phi\Big(z - \frac{\sqrt{2}x}{\sigma}\Big)\,dz. \tag{S.6}$$

This integral is the moment generating function of the non-central chi-square distribution with $d$ degrees of freedom and non-centrality parameter $2\|x\|^2/\sigma^2$ evaluated at $b$. Under the assumption $b < 1/2$, this integral is finite.

To establish the form given in (S.4) we can expand the squares to obtain:

$$\phi\Big(\frac{z}{\sqrt{2}} - \frac{x}{\sigma}\Big)\phi\Big(\frac{z}{\sqrt{2}} - \frac{y}{\sigma}\Big) = (2\pi)^{-\frac{d}{2}} \exp\Big(-\frac{\|x-y\|^2}{4\sigma^2}\Big)\phi\Big(z - \frac{x+y}{\sqrt{2}\sigma}\Big).$$

Since the first factor does not depend on $z$, it follows that

$$k(x,y) = \frac{\omega_d}{2^{d+p}(2\pi)^{\frac{d}{2}}} \exp\Big(-\frac{\|x-y\|^2}{4\sigma^2}\Big) \int_{\mathbb{R}^d} \frac{\|z\|^{d-1+2p}}{f(\|z\|)} \phi\Big(z - \frac{x+y}{\sqrt{2}\sigma}\Big)\,dz.$$

In this case, we recognize the integral as the expectation of $\|\cdot\|^{d-1+2p}/f(\cdot)$ under the chi-distribution with $d$ degrees of freedom and parameter $u = \|x+y\|/(\sqrt{2}\sigma)$. ∎

### A.1 Proof of Theorem 2

The following result is an immediate consequence of (Villani, 2008, Theorem 6.13) adapted to the notation of this paper.

**Lemma S.2** ((Villani, 2008, Theorem 6.13)). *For any $P, Q \in \mathcal{P}_p(\mathbb{R}^d)$,*

$$\mathcal{T}_p(P, Q) \leq 2^{\max(p-1,0)} \int \|x\|^p \, d|P - Q|(x), \tag{S.7}$$

*where $|P - Q|$ denotes the absolute variation the signed measure $P - Q$.*

To proceed, let $p_\sigma(z) = \int_{\mathbb{R}^d} \phi_\sigma(z - x)\, dP(x)$ and $q_\sigma(z) = \int_{\mathbb{R}^d} \phi_\sigma(z - x)\, dQ(x)$ denote the probability density functions of $P * \mathcal{N}_\sigma$ and $Q * \mathcal{N}_\sigma$, respectively. By Lemma S.2, the OT distance between $P * \mathcal{N}_\sigma$ and $Q * \mathcal{N}_\sigma$ is bounded from above by the weighted total variation distance:

$$\mathcal{T}_p^{(\sigma)}(P, Q) \leq 2^{\max(p-1,0)} \int \|z\|^p |p_\sigma(z) - q_\sigma(z)| \, dz. \tag{S.8}$$

In the following we will show that $2^{\max(p-1,0)} \sigma^p \gamma_k(P, Q)$ provides an upper bound on the right-hand side of (S.8). To proceed, recall that the kernel MMD can be expressed as

$$\gamma_k^2(P, Q) = \mathbb{E}[k(X, X')] + \mathbb{E}[k(Y, Y')] - 2\mathbb{E}[k(X, Y)], \tag{S.9}$$

where $X, X'$ are iid $P$ and $Y, Y'$ are iid $Q$. The assumptions $\int \sqrt{k(x,x)}\, P(x) < \infty$ and $\int \sqrt{k(x,x)}\, Q(x) < \infty$ ensure that these expectations are finite, and so, by Fubini's theorem, we can interchange the order of integration:

$$\mathbb{E}[k(X, Y)] = \int k(x, y)\, dP(x)\, dQ(y) = \int \left( \int \psi_x(z)\, dP(x) \right) \left( \int \psi_x(z)\, dQ(x) \right) dz.$$

For each $z \in \mathbb{R}^d$, it follows that

$$\begin{aligned}
\int \psi_x(z)\, dP(x) &= \frac{\sqrt{\omega_d}}{2^{\frac{d+p}{2}}} \frac{\|z\|^{\frac{d-1+2p}{2}}}{\sqrt{f(\|z\|)}} \int_{\mathbb{R}^d} \phi\left( \frac{z}{\sqrt{2}} - \frac{x}{\sigma} \right) dP(x) \\
&= \frac{\sigma^d \sqrt{\omega_d}}{2^{\frac{d+p}{2}}} \frac{\|z\|^{\frac{d-1+2p}{2}}}{\sqrt{f(\|z\|)}} p_\sigma\left( \frac{\sigma z}{\sqrt{2}} \right),
\end{aligned}$$

and this leads to

$$\begin{aligned}
\mathbb{E}[k(X, Y)] &= \frac{\sigma^{2d} \omega_d}{2^{d+p}} \int \frac{\|z\|^{d-1+2p}}{f(\|z\|)} p_\sigma\left( \frac{\sigma z}{\sqrt{2}} \right) q_\sigma\left( \frac{\sigma z}{\sqrt{2}} \right) dz \\
&= \sigma^{-2p} \int \frac{\|z\|^{2p}}{r_\sigma(\|z\|)} p_\sigma(z) q_\sigma(z)\, dz,
\end{aligned}$$

where $r_\sigma(x) := \frac{\sqrt{2}}{\sigma} f(\frac{\sqrt{2}}{\sigma} x) / (\omega_d \|z\|^{d-1})$. Combining this expression with (S.9) leads to

$$\gamma_k^2(P, Q) = \sigma^{-2p} \int \frac{\|z\|^{2p}}{r_\sigma(\|z\|)} (p_\sigma(z) - q_\sigma(z))^2 \, dz.$$

Finally, we note that $z \mapsto r_\sigma(\|z\|)$ is a probability density function on $\mathbb{R}^d$ (it is non-negative and integrates to one) and so by Jensen's inequality and the convexity of the square,

$$\gamma_k^2(P, Q) \geq \sigma^{-2p} \left( \int \|z\|^p |p_\sigma(z) - q_\sigma(z)| \, dz \right)^2.$$

In view of (S.8), this establishes the desired result.

### A.2 Proof of Theorem 3

The fact that the kernel MMD provides an upper bound on $\mathcal{T}_p^{(\sigma)}(P,Q)$ follows directly from Theorem 2. All the remains to be shown is that $\sqrt{k(x,x)}$ is integrable for any probability measure with finite $s$-th moment, where $s = (d+2p+\epsilon)/2$. To this end, we note that by the triangle inequality,

$$M_{d,u}(r) \le 2^{\min(1,r)}(M_d(r) + \|u\|^r),$$

for all $r \ge 0$. Under the assumptions on $\epsilon$, we have $0 \le d+2p-\epsilon < d+2p+\epsilon \le 2s$ and so there exists a constant $C_{d,p,\epsilon,\lambda}$ such that

$$k(x,y) \le C_{d,p,\epsilon,\lambda}\big(1 + \|x\|^{2s} + \|y\|^{2s}\big).$$

Thus, the existence of finite $s$-th moment is sufficient to ensure that $\sqrt{k(x,x)}$ is integrable.

## B Convergence Rate

This section provides proofs for the results in Section 4 of the main text as well as Theorem 1. To simplify the notation, we define $r = d+2p$ and let $Y = (\sqrt{2}/\sigma)X + Z$ where $Z \sim \mathcal{N}(0, I_d)$.

Let us first consider some properties of $\mathbb{E}[k(X,X)]$. Since the two moment kernel satisfies $k(x,x) = \alpha_{d,p}\, J((\sqrt{2}/\sigma)\|x\|)$, it follows from the definition of $J(\cdot)$ that

$$\mathbb{E}[k(X,X)] = \frac{\alpha_{d,p}}{2\epsilon}\big(\lambda^{\epsilon}\,\mathbb{E}\big[\|Y\|^{r-\epsilon}\big] + \lambda^{-\epsilon}\,\mathbb{E}\big[\|Y\|^{r+\epsilon}\big]\big). \tag{S.10}$$

Suppose that there exists numbers $M_-$ and $M_+$ such that

$$\mathbb{E}\big[\|Y\|^{r-\epsilon}\big] \le M_-, \quad \mathbb{E}\big[\|Y\|^{r+\epsilon}\big] \le M_+. \tag{S.11}$$

Choosing

$$\lambda = (M_+/M_-)^{1/(2\epsilon)}, \tag{S.12}$$

leads to

$$\mathbb{E}[k(X,X)] \le \frac{\alpha_{d,p}}{\epsilon}\sqrt{M_- M_+}. \tag{S.13}$$

In other words, optimizing the choice of $\lambda$ results in an upper bound on $\mathbb{E}[k(X,X)]$ that depends on only the geometric mean of the upper bounds on $\mathbb{E}[\|Y\|^{r\pm\epsilon}]$.

**Lemma S.3.** *Let $X \in \mathbb{R}^d$ be a random vector satisfying*

$$(\mathbb{E}[\|X\|^s])^{\frac{1}{s}} \le m(s), \tag{S.14}$$

*for some function $m(s)$ for $s \ge 1$. Then, if $r-\epsilon \ge 1$,* (S.11) *holds with*

$$M_{\pm} = \left(\big(\overline{M}_d(r\pm\epsilon)\big)^{\frac{1}{r\pm\epsilon}} + \frac{\sqrt{2}}{\sigma} m(r\pm\epsilon)\right)^{r\pm\epsilon}, \tag{S.15}$$

*where $\overline{M}_d(s) = (d+s)^{(d+s-1)/2} d^{-(d-1)/2} e^{-s/2}$.*

*Proof.* This result follows from Minkowski's inequality, which gives

$$(\mathbb{E}[\|Y\|^s])^{\frac{1}{s}} \le (\mathbb{E}[\|Z\|^s])^{\frac{1}{s}} + \frac{\sqrt{2}}{\sigma}(\mathbb{E}[\|X\|^s])^{\frac{1}{s}}$$

for all $s \ge 1$ and the upper bound on $M_d(s) = \mathbb{E}[\|Z\|^s]$ in Theorem 5. ☐

### B.1 Proof of Theorem 1

The result follows immediately by combining Theorem 4 and Equation (11) in the main text.

### B.2 Proof of Theorem 4

By Lyapunov's inequality and Minkowski's inequality, it follows that for $t \in \{r \pm \epsilon\}$,

$$\begin{aligned}\left(\mathbb{E}\left[\|Y\|^t\right]\right)^{\frac{1}{t}} &\leq \left(\mathbb{E}\left[\|Y\|^{r+\epsilon}\right]\right)^{\frac{1}{r+\epsilon}} \\ &\leq \left(\mathbb{E}\left[\|Z\|^{r+\epsilon}\right]\right)^{\frac{1}{r+\epsilon}} + \frac{\sqrt{2}}{\sigma}\left(\mathbb{E}\left[\|X\|^{r+\epsilon}\right]\right)^{\frac{1}{r+\epsilon}} \\ &\leq \sqrt{d+r+\epsilon} + \frac{\sqrt{2}m}{\sigma},\end{aligned}$$

where the last step holds because $M_d(q) \leq (d+q)^{q/2}$ and the assumption $(\mathbb{E}[\|X\|^s])^{\frac{1}{s}} \leq m$. Thus, for $\lambda = \sqrt{r+\epsilon} + m$, the bound in (S.13) becomes

$$\mathbb{E}[k(X,X)] \leq \frac{\alpha_{d,p}}{\epsilon}\left(\sqrt{d+r+\epsilon} + \frac{\sqrt{2}m}{\sigma}\right)^r.$$

Recalling that $r = d + 2p$ gives the stated result.

### B.3 Proof of Theorem 5

**Lemma S.4.** *Let $X \in \mathbb{R}^d$ be a sub-gamma random vector with parameters $(v, b)$. For all $s \in [0, \infty)$ and $\lambda \in (0, 1/b)$,*

$$\mathbb{E}[\|X\|^s] \leq \frac{2\sqrt{\pi}}{2^{\frac{s}{2}}\Gamma(\frac{s+1}{2})}\left(\frac{s}{\lambda e}\right)^s \exp\left(\frac{\lambda^2 v}{2(1-\lambda b)}\right) M_d(s), \tag{S.16}$$

*where $M_d(s) := 2^{\frac{s}{2}}\Gamma(\frac{d+s}{2})/\Gamma(\frac{d}{2})$. In particular, if $\lambda = (\sqrt{(sb)^2 + 4vp} - sb)/(2v)$, then*

$$\mathbb{E}[\|X\|^s] \leq \left(\sqrt{v + \left(\frac{\sqrt{s}b}{2}\right)^2} + \frac{\sqrt{s}b}{2}\right)^s \frac{2\sqrt{\pi}}{2^{\frac{s}{2}}\Gamma(\frac{s+1}{2})}\left(\frac{s}{e}\right)^{\frac{s}{2}} M_d(s). \tag{S.17}$$

*Proof.* Let $Y = Z^\top X$ where $Z = (Z_1, \ldots, Z_d)$ is independent of $X$ and distributed uniformly on the unit sphere in $\mathbb{R}^d$. Since $Z$ is orthogonally invariant, it may be assumed that $X = (\|X\|, 0, \ldots, 0)$ and thus $Y$ is equal in distribution to $Z_1\|X\|$. Therefore,

$$\mathbb{E}[|Y|^s] = \mathbb{E}[|Z_1|^s]\,\mathbb{E}[\|X\|^s].$$

The variable $Z_1$ has density function

$$f(z) = \frac{\Gamma(\frac{d}{2})}{\sqrt{\pi}\Gamma(\frac{d-1}{2})}(1-z^2)^{(d-3)/2}, \quad z \in [-1, 1],$$

and so the moments are given by

$$\mathbb{E}[|Z_1|^s] = \frac{2\Gamma(\frac{d}{2})}{\sqrt{\pi}\Gamma(\frac{d-1}{2})}\int_0^1 z^s(1-z^2)^{(d-3)/2}\,dz = \frac{\Gamma(\frac{d}{2})\Gamma(\frac{s+1}{2})}{\sqrt{\pi}\Gamma(\frac{d+s}{2})}.$$

To bound the absolute moments of $Y$ we use the basic inequality $u \leq \exp(u-1)$ with $u = \lambda|y|/s$, which leads to

$$|y|^s \leq \left(\frac{s}{\lambda e}\right)^s \exp(\lambda|y|) \leq \left(\frac{s}{\lambda e}\right)^s (e^{\lambda y} + e^{-\lambda y}),$$

for all $s, \lambda \in (0, \infty)$. Noting that $Y$ is equal in distribution to $-Y$ and then using the sub-gamma assumption along with the fact that $Z$ is a unit vector yields

$$\begin{aligned} \mathbb{E}[|Y|^s] &\le 2\Big(\frac{s}{\lambda e}\Big)^s \mathbb{E}[\exp(\lambda Y)] \\ &= 2\Big(\frac{s}{\lambda e}\Big)^s \mathbb{E}\big[\exp(\lambda Z^\top X)\big] \\ &\le 2\Big(\frac{s}{\lambda e}\Big)^s \exp\Big(\frac{\lambda^2 v}{2(1-\lambda b)}\Big). \end{aligned}$$

Combining the above displays yields (S.16).

Finally, under the specified value of $\lambda$ it follows that

$$\frac{\lambda^2 v}{1-\lambda b} = p, \qquad \frac{\sqrt{s}}{\lambda} = \sqrt{v + \left(\frac{\sqrt{s}v}{2}\right)^2} + \frac{\sqrt{s}b}{2}$$

and plugging this expression back into the bound gives (S.17). □

Theorem 5 now follows as a corollary of Lemma S.4. Starting with (S.17) and using the basic inequality $\sqrt{a^2+b^2} \le a+b$ leads to

$$\mathbb{E}[\|X\|^s] \le \big(\sqrt{v} + \sqrt{s}b\big)^s \frac{2\sqrt{\pi}}{2^{\frac{s}{2}}\Gamma(\frac{s+1}{2})}\Big(\frac{s}{e}\Big)^{\frac{s}{2}} M_d(s).$$

To simplify the expressions involving the Gamma functions we use the lower bound $\log\Gamma(z) \ge (z-\frac{1}{2})\log z - z + \frac{1}{2}\log(2\pi)$ for $z > 0$, which leads to

$$\frac{2\sqrt{\pi}}{2^{\frac{s}{2}}\Gamma(\frac{s+1}{2})}\Big(\frac{s}{e}\Big)^{\frac{s}{2}} \le \sqrt{2e}\Big(\frac{s}{s+1}\Big)^{\frac{s}{2}}.$$

Combining this bound with the expression above yields

$$\begin{aligned} \mathbb{E}[\|X\|^s] &\le \sqrt{2e}\big(\sqrt{v} + \sqrt{s}b\big)^s \left(\frac{s}{s+1}\right)^{\frac{s}{2}} M_d(s) \\ &\le \sqrt{2e}\big(\sqrt{v} + \sqrt{s}b\big)^s M_d(s). \end{aligned}$$

This completes the proof of Theorem 5.

### B.4 Proof of Theorem 6

Since $Z \sim \mathcal{N}(0, I_d)$ is sub-gamma with parameters $(1, 0)$ it follows that $Y = (\sqrt{2}/\sigma)X + Z$ is sub-gamma with parameters $(1 + 2v/\sigma^2, \sqrt{2}b/\sigma)$. For $t > -r$ we can apply Theorem 5 to obtain

$$\mathbb{E}\big[\|Y\|^{r+t}\big] \le \sqrt{2e}\Big(\sqrt{1+2v/\sigma^2} + \sqrt{r+t}\sqrt{2}b/\sigma\Big)^{r+t}\overline{M}_d(r+t) = \frac{\sqrt{2e}}{\sigma^{r+t}} m(r+t).$$

Under the specified value of $\lambda = (m(\epsilon)/m(-\epsilon))^{1/(2\epsilon)}$, it then follows from (S.13) that

$$\mathbb{E}[k(X,X)] \le \frac{\sqrt{2e}\alpha_{k,p}}{\sigma^r \epsilon}\sqrt{m(-\epsilon)m(\epsilon)}. \tag{S.18}$$

To proceed, let $(v', b') = (\sigma^2 + 2v, \sqrt{2}b)$ and consider the decomposition

$$\log(m(-\epsilon)m(\epsilon)) = 2\log m(0) + A + B,$$

where

$$\begin{aligned} A &:= (r-\epsilon)\log\Big(\sqrt{v'} + \sqrt{r-\epsilon}\,b'\Big) + (r+\epsilon)\log\Big(\sqrt{v'} + \sqrt{r+\epsilon}\,b'\Big) - 2r\log(\sqrt{v'} + \sqrt{r}b') \\ B &:= \log\overline{M}_d(r-\epsilon) + \log\overline{M}_d(r+\epsilon) - 2\log\overline{M}_d(r). \end{aligned}$$

Using the basic inequalities $\sqrt{1+x}-1 \le x/2$ and $\log(1+x) \le x$, the term $A$ can be bounded from above as follows:

$$\begin{aligned} A &= (r-\epsilon)\log\left(1+\frac{(\sqrt{r-\epsilon}-\sqrt{r})b'}{\sqrt{v'}+\sqrt{r}b'}\right)+(r+\epsilon)\log\left(1+\frac{(\sqrt{r+\epsilon}-\sqrt{r})b'}{\sqrt{v'}+\sqrt{r}b'}\right) \\ &\le (r-\epsilon)\log\left(1+\frac{(\sqrt{r-\epsilon}-\sqrt{r})}{\sqrt{r}}\right)+(r+\epsilon)\log\left(1+\frac{(\sqrt{r+\epsilon}-\sqrt{r}}{\sqrt{r}}\right) \\ &\le (r-\epsilon)\log\left(1-\frac{\epsilon}{2r}\right)+(r+\epsilon)\log\left(1+\frac{\epsilon}{2r}\right) \\ &\le -(r-\epsilon)\frac{\epsilon}{2r}+(r+\epsilon)\frac{\epsilon}{2r} \\ &= \frac{\epsilon^2}{r}. \end{aligned}$$

Similarly, one finds that

$$B = \frac{d+r-\epsilon-1}{2}\log\left(1-\frac{\epsilon}{d+r}\right)+\frac{d+r+\epsilon-1}{2}\log\left(1+\frac{\epsilon}{d+r}\right) \le \frac{\epsilon^2}{d+r}.$$

Combining these bounds with the fact that $r \ge d$ leads to

$$\sqrt{m(-\epsilon)m(\epsilon)} \le (\sqrt{v'}+\sqrt{r}b')^r \overline{M}_d(r)\exp\left(\frac{3\epsilon^2}{4d}\right).$$

Plugging this inequality back into (S.18) yields

$$\mathbb{E}[k(X,X)] \le \frac{\sqrt{2e}\alpha_{k,p}}{\epsilon}(\sqrt{1+2v/\sigma^2}+\sqrt{2r}b)^r \overline{M}_d(r)\exp\left(\frac{3\epsilon^2}{4d}\right). \tag{S.19}$$

Finally, by the lower bound $\log\Gamma(z) \ge (z-\frac{1}{2})\log z - z + \frac{1}{2}\log(2\pi)$ and the basic inequality $(1+p/d)^d \le e^p$ for $p,d,\ge 0$ we can write

$$\alpha_{d,p}\overline{M}_d(r) \le \frac{\sqrt{\pi}(d+p)^{d+p}}{e^p d^d}\frac{d}{\sqrt{d+p}} \le \sqrt{\pi}(d+p)^p\sqrt{d}.$$

Hence,

$$\mathbb{E}[k(X,X)] \le \sqrt{2\pi e}(d+p)^p\sqrt{d}\left(\sqrt{1+2v/\sigma}+\sqrt{2r}b/\sigma\right)^r\frac{\exp\left(\frac{3\epsilon^2}{4d}\right)}{\epsilon}.$$

This bound holds for all $\epsilon \in [0,r]$. Evaluating with $\epsilon = \sqrt{d}$ and recalling that $r = d+2p$ gives the stated result.

### B.5 Proof of Lemma 7

Note that $Q_{ij}$ is absolutely continuous with respect to $P \otimes P$ and let $\lambda_{ij} = dQ_{ij}/d(P\otimes P)$ denote the Radon-Nikodym derivative. Then

$$\begin{aligned} r_{ij} &= \int k(\lambda_{ij}-1)d(P\otimes P) \\ &= \int k(\sqrt{\lambda_{ij}}+1)(\sqrt{\lambda_{ij}}-1)d(P\otimes P) \\ &\le \sqrt{2}\left(\sqrt{\int k^2 dQ_{ij}}+\sqrt{\int k^2 d(P\otimes P)}\right)d_H(Q_{ij},P\otimes P), \end{aligned}$$

where the last step is by the Cauchy-Scharz inequality and we have used that fact that $d_H^2(Q_{ij},P\otimes P) = \frac{1}{2}\int(\sqrt{\lambda}-1)^2 d(P\otimes P)$.

Next, since $k^2(x,y) \le k(x,x)k(y,y)$ for any positive semidefinite kernel, it follows that

$$\int k^2(x,y)dQ_{ij}(x,y) \le \int k(x,x)k(y,y)dQ_{ij}(x,y) = \int k^2(x,x)dP(x),$$

and thus the stated result follows from the assumption $\mathbb{E}_P\left[k^2(X,X)\right] \le C_{k,P}^2$.

# C Experimental Details and Additional Results

In this section, we provide details of the experiments in Section 5 of the main text and additional numerical results. Our experiments are based on the two-moment kernel given in Definition 1.

## C.1 Numerical Computation of the Two-Moment Kernel

To evaluate the two-moment kernel given in Definition 1 we need to numerically compute the function $M_{d,u}(s)$, which is the $s$-th moment of the non-central chi-distribution with $d$ degrees of freedom and parameter $u$. For all $s \geq 0$, this function can be written as a Poisson mixture of the (central) moments according to

$$M_{d,u}(s) = \sum_{k=0}^{\infty} \frac{u^{2k} \exp(-\frac{1}{2}u^2)}{2^k k!} M_{d+2k,0}(s). \tag{S.20}$$

This series can be approximated efficiently by retaining only the terms with $k \approx u^2/2$.

Alternatively, if $s = 2\ell$ where $\ell$ is an integer, then $M_{d,u}(2\ell)$ is the $\ell$-th moment of the chi-square distribution with $d$ degrees of freedom and non-centrality parameter $u^2$. The integer moments of this distribution can be obtained by differentiating the moment generating function. An explicit formula is given by Johnson et al. (1995, pg. 448)

$$M_{d,u}(2\ell) = 2^\ell \Gamma(\ell + d/2) \sum_{j=0}^{\ell} \binom{k}{j} \frac{(u^2/2)^j}{\Gamma(j + d/2)}. \tag{S.21}$$

Here we see that $M_{d,u}(2\ell)$ is a degree $\ell$ polynomial in $u^2$.

Accordingly, for any tuple $(d, p, \sigma, \lambda, \epsilon)$ such that $d + 2p \pm \epsilon$ are even integers, the two-moment kernel defined in (11) can be expressed as

$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{4\sigma^2}\right) \sum_{\ell=0}^{L} c_\ell \left(\frac{\|x + y\|}{\sqrt{2}\sigma}\right)^{2\ell}, \tag{S.22}$$

where $L = (d + 2p + \epsilon)/2$ and the coefficients $c_0, \dots, c_L$ are given by

$$c_\ell \coloneqq \frac{\alpha_{d,p}}{\epsilon 2^\ell \Gamma(\ell + d/2)} \left[\lambda^\epsilon 2^{L-\epsilon} \Gamma(L - \epsilon + d/2) \binom{L - \epsilon}{\ell} \mathbf{1}_{\{\ell \leq L-\epsilon\}} + \lambda^\epsilon 2^L \Gamma(L + d/2) \binom{L}{\ell}\right], \tag{S.23}$$

with $\alpha_{d,p} \coloneqq (2\pi) 2^{-(p+d)} 2^{-d/2} / \Gamma(d/2)$.

## C.2 Details for Example 1

We now consider Example 1, a specific example of a sub-gamma distribution which shows that the upper bound in Theorem 6 is tight with respect to the scaling of the dimension $d$ and the scale parameter $b$. Specifically, let $X = \sqrt{U} Z$ where $Z \sim \mathcal{N}(0, I_d)$ is a standard Gaussian vector and $U$ is an independent Gamma random variable with shape parameter $1/(2b^2)$ and scale parameter $2b^2$.

**Lemma S.5.** *For $\alpha \in \mathbb{R}^d$ such that $\|\alpha\| \leq 1/b$, it holds that*

$$\mathbb{E}\big[\exp(\alpha^\top X)\big] = -\frac{1}{2b^2} \log\big(1 - \|\alpha\|^2 b^2\big). \tag{S.24}$$

*In particular, this means that $X$ is a sub-gamma random vector with parameters $(1, b)$. Furthermore, for $s > \max\{-b^{-2}, -d\}$,*

$$\mathbb{E}[\|X\|^s] = b^s M_{b^{-2}}(s) M_d(s). \tag{S.25}$$

*Proof.* Observe that $\alpha^\top X = \sqrt{U} \alpha^\top Z$ where $\alpha^\top Z \sim \mathcal{N}(0, \|\alpha\|^2)$. Hence

$$\mathbb{E}\big[\exp(\alpha^\top X)\big] = \mathbb{E}\left[\exp\left(\frac{\|\alpha\|^2}{2} U\right)\right].$$

Recognizing the right-hand side as the moment generating function of the Gamma distribution evaluated at $\|\alpha\|^2/2$ yields (S.24). To see that this distribution satisfies the sub-gamma condition, we use the basic inequality $-\log(1-x) \le x/(1-x) \le x/(1-\sqrt{x})$ for all $x \in (0,1)$.

The expression for the moments follows immediately from the independence of $U$ and $Z$ and the fact that $U^2/b^2$ has Gamma distribution with shape parameter $b^{-2}/2$ and scale parameter 2, which implies that $U/b$ has a chi distribution with $b^{-2}$ degrees of freedom. □

Since $X$ satisfies the sub-gamma condition with parameters $(1, b)$ the upper bound in Theorem 6 applies. Alternatively, for each pair $(\epsilon, \lambda)$ we can consider the exact expression for $\mathbb{E}[k(X,X)]$ given in (S.10) where $r = d + 2p$ and

$$Y = \left(\frac{2}{\sigma^2}U + 1\right)^{1/2} Z.$$

Minimizing this expression with respect to $\lambda$ yields

$$\mathbb{E}[k(X,X)] \ge \frac{\alpha_{d,p}}{\epsilon}\left(\mathbb{E}\left[\|Y\|^{r-\epsilon}\right]\mathbb{E}\left[\|Y\|^{r+\epsilon}\right]\right)^{1/2}. \tag{S.26}$$

To get a lower bound on the moments, we use

$$\mathbb{E}[\|Y\|^s] \ge \left(\frac{\sqrt{2}}{\sigma}\right)^s \mathbb{E}[\|X\|^s]. \tag{S.27}$$

Combining the above displays leads to (22). Using Stirling's approximation $\log\Gamma(z) = (z-\frac{1}{2})\log z - z + \frac{1}{2}\log(2\pi) + o(1)$ as $z \to \infty$ it can be verified that the minimum of this lower bound with respect to $\epsilon$ satisfies the same scaling behavior with respect to $d$ as the upper bound in Theorem 6. Namely, the bound exponential in $d$ if $\delta \ge 1/2$ and superexponential in $d$ if $\delta < 1/2$.

### C.3 Experiments in Section 5.1

In this experiment, $p = 1$, the random variable $X \in \mathbb{R}^d$ is generated according to the distribution in Example 1, and the kernel bandwidth $\sigma$ takes values 1 and 4. The parameters $(\lambda, \epsilon)$ of the two-moment kernel are specified as in Theorem 6 with parameters $(1, b)$, and $k(x, y)$ can be computed as in Appendix C.1.

In the Monte-Carlo computation of the average of $\hat{\Delta}^2_\gamma$ (the right column of Figure 2), $2n$ samples of $X$ are partitioned into two independent datasets $\{X_i\}_{i=1}^n$ and $\{X'_i\}_{i=1}^n$, each having $n$ samples. The kernel MMD (squared) distance has the empirical estimator (Gretton et al., 2012)

$$\gamma_k^2(P_n, P'_n) = \frac{1}{n^2}\sum_{i,j=1}^n k(X_i, X_j) + \frac{1}{n^2}\sum_{i,j=1}^n k(X'_i, X'_j) - \frac{2}{n^2}\sum_{i=1,j}^n k(X_i, X'_j),$$

and then, by definition,

$$\begin{aligned}\mathbb{E}\left[\gamma_k^2(P_n, P'_n)\right] &= 2(\frac{1}{n}\mathbb{E}[k(X,X)] + (1-\frac{1}{n})\mathbb{E}[k(X,X')] - 2\mathbb{E}[k(X,X')] \\ &= \frac{2}{n}(\mathbb{E}[k(X,X)] - \mathbb{E}[k(X,X')]).\end{aligned}$$

Recall that

$$\gamma_k^2(P, P_n) = \int\int k(x,x')dP(x)dP(x') + \frac{1}{n^2}\sum_{i,j=1}^n k(X_i, X_j) - \frac{2}{n}\sum_{i=1}^n \int k(x, X_i)dP(x),$$

and then

$$\begin{aligned}\mathbb{E}\left[\gamma_k^2(P, P_n)\right] &= \mathbb{E}[k(X,X')] + \frac{1}{n}\mathbb{E}[k(X,X)] + (1-\frac{1}{n})\mathbb{E}[k(X,X')] - 2\mathbb{E}[k(X,X')] \\ &= \frac{1}{n}(\mathbb{E}[k(X,X)] - \mathbb{E}[k(X,X')]).\end{aligned}$$

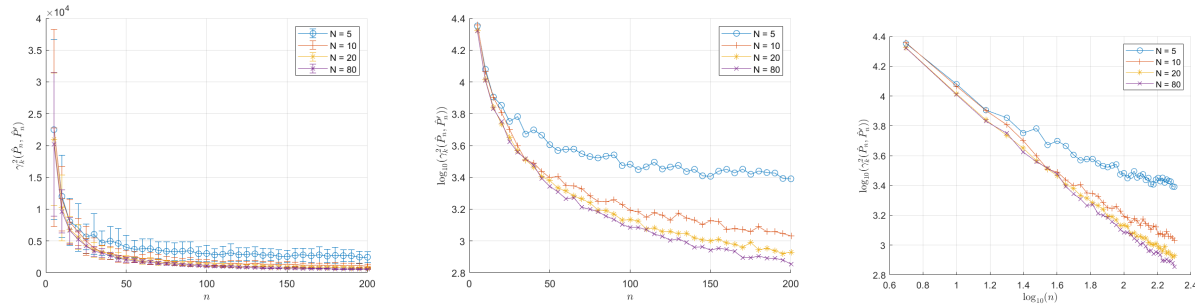


Figure S.1: Empirical values of $\gamma_k^2(P_n, P_n')$ as a function of $n$ for various values of $N$ for dependent samples in Example 2, that is, the same experiment as in Figure 3. (Left) Mean and standard deviation averaged over 100 realizations. (Middle) The $\log_{10}$ of the mean value in the left plot. (Right) The log-log plot of the mean value in the left plot.

Thus, if we define

$$\hat{\Delta}_\gamma^2 \coloneqq \frac{n}{2}\gamma_k^2(P_n, P_n'),$$

the expectation of $\hat{\Delta}_\gamma^2$ equals $\mathbb{E}[k(X,X)] - \mathbb{E}[k(X,X')] = \mathbb{E}\big[n\gamma_k^2(P,P_n)\big]$.

### C.4 Experiments in Section 5.2

In this experiment, $d = 5$, $p = 1$, $\sigma = 1/2$ and the parameters $(\epsilon, \lambda)$ of the two-moment kernel are specified as in Theorem 6 with parameters $(1, 0)$. Figure 3 in the main text plots the values of $\gamma_k^2(P_n, P_n')$ as a function of increasing $N$ and for various values of $n$. Figure S.1 plots $\gamma_k^2(P_n, P_n')$ as a function of increasing $n$ and for various values of $N$. Note that in this setting, the typical correlation between samples is of magnitude $1/\sqrt{N}$, and thus the overall dependence is not negligible when $N$ is relatively small compared to $n$.