**Note:** For Theorem 4, our lower bound construction does not fully match the claim in the current statement. See Appendix E.2 for the correction (which is still strong enough to support the discussion on reductions which follows the theorem). We will update this in the revised version.

# A Proofs for Section 2

## A.1 Proof of Proposition 1

Clearly, $\mathcal{E}_p(\kappa;\mu,\nu)=0$ if $\kappa$ minimizes (2). On the other hand, if $\mathcal{E}_p(\kappa,\mu,\nu)=0$, then $\kappa_\sharp\mu=\nu$. Thus, $\kappa$ is feasible for (2) with optimal objective value, i.e., it is a minimizer.

Further, if $T^\star$ is an optimal map, then $\mathsf{W}_p(\mu,\nu)=\|T^\star-\mathrm{Id}\,\|_{L^p(\mu)}$ and $T^\star_\sharp\mu=\nu$. We thus bound

$$\begin{aligned}\mathcal{E}_p(T;\mu,\nu)&=\big[\|T-\mathrm{Id}\,\|_{L^p(\mu)}-\mathsf{W}_p(\mu,\nu)\big]_+ +\mathsf{W}_p(T_\sharp\mu,\nu)\\&=\big[\|T-\mathrm{Id}\,\|_{L^p(\mu)}-\|T^\star-\mathrm{Id}\,\|_{L^p(\mu)}\big]_+ +\mathsf{W}_p(T_\sharp\mu,T^\star_\sharp\mu)\\&\le 2\|T-T^\star\|_{L^p(\mu)},\end{aligned}$$

as desired. □

## A.2 Reverse $L^2$ comparison (Remark 2)

Suppose that there exists a unique Brenier map of the form $T^\star=\nabla\varphi$, where $\varphi:\mathbb{R}^d\to\mathbb{R}^d$ is convex and twice differentiable such that $H\varphi\preceq LI_d$ for $L\ge 1$. Fixing any map $T:\mathcal{X}\to\mathcal{Y}$, we abbreviate $\varepsilon=\mathcal{E}_2(T;\mu,\nu)$. By the definition of $\mathcal{E}_2$, we have $\mathsf{W}_2(T_\sharp\mu,\nu)\le\varepsilon$. Let $\lambda\in\mathcal{K}(\mathcal{Y},\mathcal{Y})$ be a kernel which achieves this bound, and take $\kappa=\lambda\circ T$. By construction, we have $\kappa_\sharp\mu=\nu$ and

$$\begin{aligned}\left(\iint\|y-x\|^2\mathrm{d}\kappa(y|x)\mathrm{d}\mu(x)\right)^{\frac12}-\mathsf{W}_2(\mu,\nu)&\le\left(\iint\|T(x)-x\|^2\mathrm{d}\mu(x)\right)^{\frac12}+\mathsf{W}_2(\mu,\nu)+\varepsilon\\&\le 2\varepsilon.\end{aligned}$$

Consequently, we have

$$\begin{aligned}\iint\|y-x\|^2\mathrm{d}\kappa(y|x)\mathrm{d}\mu(x)-\mathsf{W}_2(\mu,\nu)^2&\le 2\varepsilon\left(\left(\iint\|y-x\|^2\mathrm{d}\kappa(y|x)\mathrm{d}\mu(x)\right)^{\frac12}+\mathsf{W}_2(\mu,\nu)\right)\\&\le 2\varepsilon\cdot(2\mathsf{W}_2(\mu,\nu)+2\varepsilon)\end{aligned}$$

Thus, by Proposition 3.1 of Li and Nochetto [2021], we have

$$\begin{aligned}\iint\|y-T^\star(x)\|^2\mathrm{d}\kappa(y|x)\mathrm{d}\mu(x)&\le L\left(\iint\|y-x\|^2\mathrm{d}\kappa(y|x)\mathrm{d}\mu(x)-\mathsf{W}_2(\mu,\nu)^2\right)\\&\le 4L\varepsilon\cdot(\mathsf{W}_2(\mu,\nu)+\varepsilon).\end{aligned}$$

Finally, we bound

$$\begin{aligned}\|T-T_\star\|_{L^2(\mu)}&\le\left(\iint\|y-T^\star(x)\|^2\mathrm{d}\kappa(y|x)\mathrm{d}\mu(x)\right)^{\frac12}+\varepsilon\\&\le\sqrt{4L\varepsilon\cdot(\mathsf{W}_2(\mu,\nu)+\varepsilon)}+\varepsilon\\&\lesssim\sqrt{L\varepsilon\cdot(\mathsf{W}_2(\mu,\nu)+\varepsilon)},\end{aligned}$$

as desired.

# B Proofs for Section 2

## B.1 Proof of Lemma 2

We simply bound

$$\begin{aligned}\big|\mathcal{E}_p(\kappa;\mu,\nu)-\mathcal{E}_p(\kappa;\mu,\nu')\big|&\le|\mathsf{W}_p(\mu,\nu)-\mathsf{W}_p(\mu,\nu')|+|\mathsf{W}_p(\kappa_\sharp\mu,\nu)-\mathsf{W}_p(\kappa_\sharp\mu,\nu')|\\&\le 2\mathsf{W}_p(\nu,\nu')\\&\le 2\,\mathrm{diam}(\mathcal{Y})\|\nu-\nu'\|_{\mathrm{TV}},\end{aligned}$$

where the final inequality uses Fact 1. □

### B.2 Proof of Lemma 3

While the key ideas of this proof are straightforward, measurability issues require some care (we encourage the reader to skip such details on an initial read). In what follows, we equip all spaces of distributions with the weak topology and always employ Borel measurability. By the definition of a Markov kernel, $x \in \mathcal{X} \mapsto \kappa_x(A)$ is a measurable function for each measurable $A \subseteq \mathcal{Y}$. Thus, $(x, x') \in \mathcal{X}^2 \mapsto (\kappa_x(A), \kappa_{x'}(B))$ is measurable for fixed, measurable $A, B \subseteq \mathcal{Y}$, implying that $(x, x') \in \mathcal{X}^2 \mapsto (\kappa_x, \kappa_{x'}) \in \mathcal{P}(\mathcal{Y})^2$ is measurable. Therefore, by Theorem 3.0.8 of Toneian [2019], there exists a measurable map $(x, x') \in \mathcal{X}^2 \mapsto \gamma_{x,x'} \in \Pi(\kappa_x, \kappa_{x'})$ such that $\gamma_{x,x'}$ is an OT plan for $\mathsf{W}_p(\kappa_x, \kappa_{x'})$ for all $x, x' \in \mathcal{X}$.

Now, let $\pi_0 \in \Pi(\mu, \mu')$ be an OT plan for $\mathsf{W}_p(\mu, \mu')$, and define the joint law $\pi$ by $\pi(A \times B \times C \times D) \coloneqq \iint_{A\times B} \gamma_{x,x'}(C \times D)\mathrm{d}\pi_0(x, x')$, which is well-defined due to the measurability argument above. Taking $(X, X', Y, Y') \sim \pi$, our construction ensures the following:

- $X \sim \mu$ and $X' \sim \mu'$ such that $\mathbb{E}[\|X - X'\|^p] = \rho^p$,
- $Y \sim \kappa_X$ and $Y' \sim \kappa_{X'}$ such that $\mathbb{E}[\|Y - Y'\|^p|X, X'] = \mathsf{W}_p(\kappa_X, \kappa_{X'})^p$.

Consequently, we bound

$$\begin{aligned}
\mathbb{E}[\|Y - Y'\|^p] &= \mathbb{E}\big[\mathbb{E}\big[\|Y - Y'\|^p\big|X, X'\big]\big] \\
&= \mathbb{E}[\mathsf{W}_p(\kappa_X, \kappa_{X'})^p] \\
&\leq \mathbb{E}[L^p\|X - X'\|^{\alpha p}] && \text{(Hölder continuity of } \kappa) \\
&\leq L^p\,\mathbb{E}[L\|X - X'\|^p]^\alpha && \text{(Jensen's inequality, } 0 < \alpha \leq 1) \\
&= L^p\rho^{\alpha p}.
\end{aligned}$$

Moreover, using Minkoswki's inequality, we compute

$$\big|\mathbb{E}[\|Y - X\|^p]^{\frac{1}{p}} - \mathbb{E}[\|Y' - X'\|^p]^{\frac{1}{p}}\big| \leq \mathbb{E}[\|X - X'\|^p]^{\frac{1}{p}} + \mathbb{E}[\|Y - Y'\|^p]^{\frac{1}{p}} \leq \rho + L\rho^\alpha.$$

Finally, we bound $|\mathsf{W}_p(\mu, \nu) - \mathsf{W}_p(\mu', \nu)| \leq \mathsf{W}_p(\mu, \mu') \leq \rho$ and

$$|\mathsf{W}_p(\kappa_\sharp\mu, \nu) - \mathsf{W}_p(\kappa_\sharp\mu', \nu)| \leq \mathsf{W}_p(\kappa_\sharp\mu, \kappa_\sharp\mu') \leq \mathbb{E}[\|Y - Y'\|^p]^{\frac{1}{p}} \leq L\rho^\alpha.$$

The definition of $\mathcal{E}_p$ and these bounds give the lemma. □

### B.3 Proof of Lemma 4

To show this result, we prove a slightly more general lemma.

**Lemma 6.** *Fix $\mu, \mu' \in \mathcal{P}(\mathcal{X})$, $\nu \in \mathcal{P}(\mathcal{Y})$, and kernel $\kappa \in \mathcal{K}(\mathcal{X}, \mathcal{Y})$ satisfying $\iint \|y - x\|^p \mathrm{d}\kappa_x(y)\mathrm{d}\mu(x) \leq \mathsf{W}_p(\mu, \kappa_\sharp\mu)^p + \tau^p$ and $\mathsf{W}_p(\kappa_\sharp\mu, \nu) \leq \tau$ for some $\tau \geq 0$. Then, setting $\varepsilon = \|\mu - \mu'\|_{\mathrm{TV}}$, we have $\mathcal{E}_p(\kappa; \mu', \nu) \leq 4\,\mathrm{diam}(\mathcal{Y})\varepsilon^{1/p} + 2\tau$.*

*Proof.* In what follows, we encourage the reader to focus on the $p = 1$ case, where computations are more direct. Write $\varepsilon = \|\mu - \mu'\|_{\mathrm{TV}}$. By the TV bound, there exist $\alpha, \beta, \gamma \in \mathcal{M}_+(\mathcal{X})$ with $\gamma(\mathcal{X}) = 1 - \varepsilon$ and $\alpha(\mathcal{X}) = \beta(\mathcal{X}) = \varepsilon$ such that $\mu = \gamma + \alpha$ and $\mu' = \gamma + \beta$.

First, we note that $\kappa$ must perform well on $\gamma$. Specifically, we have

$$\begin{aligned}
&\iint \|x - y\|^p \mathrm{d}\kappa_x(y)\mathrm{d}\gamma(x) \\
&\quad = \iint \|x - y\|^p \mathrm{d}\kappa_x(y)\mathrm{d}\mu(x) - \iint \|x - y\|^p \mathrm{d}\kappa_x(y)\mathrm{d}\alpha(x) && (\gamma = \mu - \alpha) \\
&\quad \leq \mathsf{W}_p(\mu, \kappa_\sharp\mu)^p + \tau^p - \mathsf{W}_p(\alpha, \kappa_\sharp\alpha)^p && \text{(error bound for } \kappa\text{, def. of } \mathsf{W}_p) \\
&\quad \leq \mathsf{W}_p(\gamma, \kappa_\sharp\gamma)^p + \mathsf{W}_p(\alpha, \kappa_\sharp\alpha)^p + \tau^p - \mathsf{W}_p(\alpha, \kappa_\sharp\alpha)^p && (\mu = \gamma + \alpha) \\
&\quad = \mathsf{W}_p(\gamma, \kappa_\sharp\gamma)^p + \tau^p \\
&\quad \leq (\mathsf{W}_p(\gamma, \kappa_\sharp\gamma) + \tau)^p. && (\ell_p \leq \ell_1)
\end{aligned}$$

Now, letting $\kappa'$ be an optimal kernel for the $\mathsf{W}_p(\mu',\nu)$ problem and writing $D = \mathrm{diam}(\mathcal{Y})$, we have

$$
\begin{aligned}
&\left(\iint \|y-x\|^p \mathrm{d}\kappa_x(y)\mathrm{d}\mu'(x)\right)^{\frac{1}{p}} - \mathsf{W}_p(\mu',\nu) \\
&= \left(\iint \|y-x\|^p \mathrm{d}\kappa_x(y)\mathrm{d}\mu'(x)\right)^{\frac{1}{p}} - \left(\iint \|y-x\|^p \mathrm{d}\kappa'_x(y)\mathrm{d}\mu'(x)\right)^{\frac{1}{p}} && \text{(optimality of } \kappa'\text{)} \\
&= \left(\iint \|y-x\|^p \mathrm{d}\kappa_x(y)\mathrm{d}\gamma(x) + \iint \|y-x\|^p \mathrm{d}\kappa_x(y)\mathrm{d}\beta(x)\right)^{\frac{1}{p}} && (\mu' = \gamma+\beta) \\
&\quad - \left(\iint \|y-x\|^p \mathrm{d}\kappa'_x(y)\mathrm{d}\gamma(x) + \iint \|y-x\|^p \mathrm{d}\kappa'_x(y)\mathrm{d}\beta(x)\right)^{\frac{1}{p}} \\
&\leq \left(\left[\left(\iint \|y-x\|^p \mathrm{d}\kappa_x(y)\mathrm{d}\gamma(x)\right)^{\frac{1}{p}} - \left(\iint \|y-x\|^p \mathrm{d}\kappa'_x(y)\mathrm{d}\gamma(x)\right)^{\frac{1}{p}}\right]_+^p \right.\\
&\quad \left. + \left[\left(\iint \|y-x\|^p \mathrm{d}\kappa_x(y)\mathrm{d}\beta(x)\right)^{\frac{1}{p}} - \left(\iint \|y-x\|^p \mathrm{d}\kappa'_x(y)\mathrm{d}\beta(x)\right)^{\frac{1}{p}}\right]_+^p \right)^{\frac{1}{p}} \\
&\leq \Bigg( \left[\mathsf{W}_p(\gamma, \kappa_\sharp\gamma) + \tau - \mathsf{W}_p(\gamma, \kappa'_\sharp\gamma)\right]_+^p \\
&\quad + \left[\left(\iint \|y-x\|^p \mathrm{d}\kappa_x(y)\mathrm{d}\beta(x)\right)^{\frac{1}{p}} - \left(\iint \|y-x\|^p \mathrm{d}\kappa'_x(y)\mathrm{d}\beta(x)\right)^{\frac{1}{p}}\right]_+^p \Bigg)^{\frac{1}{p}} \\
&\leq \left( \left(\mathsf{W}_p(\kappa_\sharp\gamma, \kappa'_\sharp\gamma) + \tau\right)^p + \iiint \|y-y'\|^p \mathrm{d}\kappa_x(y)\mathrm{d}\kappa'_x(y')\mathrm{d}\beta(x)\right)^{\frac{1}{p}} \\
&\leq \left( \left(\mathsf{W}_p(\kappa_\sharp\gamma, \kappa'_\sharp\gamma) + \tau\right)^p + \varepsilon D^p\right)^{\frac{1}{p}} && \text{(Fact 1)} \\
&\leq \mathsf{W}_p(\kappa_\sharp\gamma, \kappa'_\sharp\gamma) + \tau + D\varepsilon^{1/p}. && (\ell_p \leq \ell_1)
\end{aligned}
$$

The first inequality uses that $(A^p + B^p)^{1/p} - (a^p + b^p)^{1/p} \leq ([A-a]_+^p + [B-b]_+^p)^{1/p}$, which can be obtained by rearranging the $\ell_p$ triangle inequality and using that $A = [A-a]_+ + A \wedge a$. The second inequality uses the previous bound and the fact that $\kappa'$ is feasible for the $\mathsf{W}_p(\gamma, \kappa'_\sharp\gamma)$ problem. The third uses the $\mathsf{W}_p$ triangle inequality and Minkoswki's inequality.

We next bound $\mathsf{W}_p(\kappa_\sharp\gamma, \kappa'_\sharp\gamma)$. Let $\pi \in \Pi(\kappa_\sharp\mu, \nu)$ be an optimal plan for $\mathsf{W}_p(\kappa_\sharp\mu,\nu)$ and define $\lambda \in (1-\varepsilon)\mathcal{P}(\mathcal{Y})$ by $\lambda(\cdot) = \int \pi(\cdot|x)\mathrm{d}\gamma(x)$. By construction, $\mathsf{W}_p(\kappa_\sharp\gamma, \lambda) \leq \mathsf{W}_p(\kappa_\sharp\mu,\nu) \leq \tau$. Moreover, both $\lambda$ and $\kappa'_\sharp\gamma$ are submeasures of $\nu$ with mass $1-\varepsilon$, and so they must share common mass at least $1-2\varepsilon$. This implies that their TV distance is at most $\varepsilon$, and so Fact 1 gives that

$$
\mathsf{W}_p(\kappa_\sharp\gamma, \kappa'_\sharp\gamma) \leq \tau + \mathsf{W}_p(\lambda, \kappa'_\sharp\gamma) \leq \tau + D\varepsilon^{\frac{1}{p}}. \tag{7}
$$

Thus, the previous bound on the optimality gap can be tightened to

$$
\left(\iint \|y-x\|^p \mathrm{d}\kappa_x(y)\mathrm{d}\mu'(x)\right)^{\frac{1}{p}} - \mathsf{W}_p(\mu',\nu) \leq \tau + 2D\varepsilon^{1/p}.
$$

Similarly, we bound the feasibility gap by

$$
\begin{aligned}
\mathsf{W}_p(\kappa_\sharp\mu', \nu) &= \mathsf{W}_p(\kappa_\sharp\gamma + \kappa_\sharp\beta, \kappa'_\sharp\gamma + \kappa'_\sharp\beta) && (\mu' = \gamma+\beta, \kappa'_\sharp\gamma = \nu) \\
&\leq \mathsf{W}_p(\kappa_\sharp\gamma, \kappa'_\sharp\gamma) + \mathsf{W}_p(\kappa_\sharp\beta, \kappa'_\sharp\beta) && \text{(joint convexity of } \mathsf{W}_p\text{)} \\
&\leq \tau + 2D\varepsilon^{\frac{1}{p}}. && \text{(Fact 1 and Eq. 7)}
\end{aligned}
$$

Combining, we have that $\mathcal{E}_p(\kappa; \mu', \nu) \leq 2\tau + 4D\varepsilon^{1/p}$, as desired. □

We now seek to find a suitable error bound $\tau$ in terms of $\mathcal{E}_p(\kappa;\mu,\nu)$. First, we have by definition that $\mathsf{W}_p(\kappa_\sharp\mu,\nu) \le \mathcal{E}_p(\kappa;\mu,\nu)$. Further, $\mathcal{E}_p(\kappa;\mu,\nu) \le 2\operatorname{diam}(\mathcal{X}\cup\mathcal{Y})$. Thus, we can also bound

$$\begin{aligned}\iint \|y-x\|^p \mathrm{d}\kappa_x(y)\mathrm{d}\mu(x) &\le (\mathsf{W}_p(\mu,\nu)+\mathcal{E}_p(\kappa;\mu,\nu))^p\\ &\le (\mathsf{W}_p(\mu,\kappa_\sharp\mu)+2\mathcal{E}_p(\kappa;\mu,\nu))^p\\ &\le \mathsf{W}_p(\mu,\kappa_\sharp\mu)^p + 2p\mathcal{E}_p(\kappa;\mu,\nu)(\mathsf{W}_p(\mu,\kappa_\sharp\mu)\vee 2\mathcal{E}_p(\kappa;\mu,\nu))^{p-1}\\ &\le \mathsf{W}_p(\mu,\kappa_\sharp\mu)^p + 2p\mathcal{E}_p(\kappa;\mu,\nu)(\mathsf{W}_p(\mu,\nu)+3\mathcal{E}_p(\kappa;\mu,\nu))^{p-1}\\ &\le \mathsf{W}_p(\mu,\kappa_\sharp\mu)^p + 2p\mathcal{E}_p(\kappa;\mu,\nu)\mathsf{W}_p(\mu,\nu)^{p-1} + 3^p p\mathcal{E}_p(\kappa;\mu,\nu)^p.\end{aligned}$$

Thus, we can take

$$\begin{aligned}\tau &= \mathcal{E}_p(\kappa;\mu,\nu) \vee \left[2p\mathcal{E}_p(\kappa;\mu,\nu)\mathsf{W}_p(\mu,\nu)^{p-1} + 3^p p\mathcal{E}_p(\kappa;\mu,\nu)^p\right]^{\frac{1}{p}}\\ &\le 3\mathcal{E}_p(\kappa;\mu,\nu)^{\frac{1}{p}}\mathsf{W}_p(\mu,\nu)^{\frac{p-1}{p}} + 5\mathcal{E}_p(\kappa;\mu,\nu)\end{aligned}$$

Plugging into Lemma 6 gives that

$$\mathcal{E}_p(\kappa;\mu',\nu) \le 4D\varepsilon^{1/p} + 6\mathcal{E}_p(\kappa;\mu,\nu)^{\frac{1}{p}}\mathsf{W}_p(\mu,\nu)^{\frac{p-1}{p}} + 10\mathcal{E}_p(\kappa;\mu,\nu),$$

as desired. The $p=1$ result is immediate. □

### B.4 Proof of Lemma 5

First, we note that $\kappa_\sharp(\lambda_\sharp\mu) = (\kappa\circ\lambda)_\sharp\mu$ by the definition of kernel composition. This implies that the two feasibility gaps coincide. Moreover, by Minkowski's inequality, we have

$$\begin{aligned}&\left|\left(\iint \|y-z\|^p \mathrm{d}\kappa_z(y)\mathrm{d}(\lambda_\sharp\mu)(z)\right)^{\frac{1}{p}} - \left(\iint \|y-x\|^p \mathrm{d}(\kappa\circ\lambda)_x(y)\mathrm{d}\mu(x)\right)^{\frac{1}{p}}\right|\\ &\le \left(\iint \|z-x\|^p\mathrm{d}\lambda_x(z)\mathrm{d}\mu(x)\right)^{\frac{1}{p}}\end{aligned}$$

and

$$|\mathsf{W}_p(\lambda_\sharp\mu,\nu) - \mathsf{W}_p(\mu,\nu)| \le \mathsf{W}_p(\mu,\lambda_\sharp\mu) \le \left(\iint \|z-x\|^p \mathrm{d}\lambda_x(z)\mathrm{d}\mu(x)\right)^{\frac{1}{p}}.$$

Combining these two error bounds gives the lemma. □

## C Proofs for Section 3

### C.1 Proof of Theorem 1

By the support constraint, our cost $\|x-y\|^p$ is $pd^{(p-1)/2}$-Lipschitz over $\mathcal{X}\times\mathcal{Y}$. Thus, by Theorem 1 of Genevay et al. [2019], we have

$$S_{p,\tau}(\hat{\mu}_n,\hat{\nu}_n) \le \mathsf{W}_p(\hat{\mu}_n,\hat{\nu}_n)^p + 2\tau d\log\left(e^2 pd^{p/2-1}\tau^{-1}\right).$$

Since $\hat{\pi}_{\tau,n}$ achieves the left hand side above and KL divergence is non-negative, we have

$$\iint \|x-y\|^p \mathrm{d}\hat{\kappa}_n(y|x)\mathrm{d}\hat{\mu}_n(x) \le \mathsf{W}_p(\hat{\mu}_n,\hat{\nu}_n)^p + 2\tau d\log\left(e^2pd^{p/2-1}\tau^{-1}\right).$$

Taking $p$th roots and noting that $(\hat{\kappa}_n)_\sharp\hat{\mu}_n = \hat{\nu}_n$, this implies that

$$\mathcal{E}_p(\hat{\kappa}_n;\hat{\mu}_n,\hat{\nu}_n) \le \left[2\tau d\log\left(e^2pd^{p/2-1}\tau^{-1}\right)\right]^{\frac{1}{p}} \le 4(\tau d)^{\frac{1}{p}}\log(e^2 d\tau^{-1}).$$

Now, by (6), $(\hat{\kappa}_n)_x$ is obtained by applying softmax to $v(x) := \left((g_\tau(Y_i) - \|x-Y_i\|^p)/\tau\right)_{i=1}^n \in \mathbb{R}^n$. Since the $\ell_1,\ell_\infty$ Lipschitz constant of the softmax operation is $\le 1$, we have

$$\|(\hat{\kappa}_n)_x - (\hat{\kappa}_n)_{x'}\|_{\mathrm{TV}} \le \frac{1}{2}\|v(x)-v(x')\|_\infty \le \frac{1}{2}pd^{(p-1)/2}\tau^{-1}\|x-x'\|_2$$

for all $x, x' \in [0,1]^d$. Thus, by Fact 1, $\hat{\kappa}_n$ is Hölder continuous under $\mathsf{W}_p$ with exponent $1/p$ and constant $2\sqrt{d}\tau^{-1/p}$. Applying Lemma 3 now gives

$$\begin{aligned}\mathcal{E}_p(\hat{\kappa}_n; \mu, \nu) &\le \mathcal{E}_p(\hat{\kappa}_n; \mu, \hat{\nu}_n) + \mathsf{W}_p(\hat{\nu}_n, \nu)\\ &\le \mathcal{E}_p(\hat{\kappa}_n; \hat{\mu}_n, \hat{\nu}_n) + 2\mathsf{W}_p(\mu, \hat{\mu}_n) + 4\sqrt{d}\,\mathsf{W}_p(\mu, \hat{\mu})^{\frac{1}{p}}\tau^{-\frac{1}{p}} + \mathsf{W}_p(\nu, \hat{\nu}_n)\\ &\le 4(\tau d)^{\frac{1}{p}}\log(e^2 d\tau^{-1}) + 4\sqrt{d}\,\mathsf{W}_p(\mu, \hat{\mu})^{\frac{1}{p}}\tau^{-\frac{1}{p}} + 2\mathsf{W}_p(\mu, \hat{\mu}_n) + \mathsf{W}_p(\nu, \hat{\nu}_n).\end{aligned}$$

Taking expectations, applying Lemma 1, and plugging in $\tau$ gives the theorem. □

### C.2 Minimax Lower Bound under Sampling

Fix $\mu = \delta_0$, so that the constant kernel $\kappa^\star$ defined by $\kappa^\star_x \equiv \nu$ is optimal. Note that the error $\mathcal{E}_p(\kappa; \mu, \nu)$ of any kernel $\kappa$ is thus lower bounded by the feasibility gap $\mathsf{W}_p(\kappa_0, \nu)$. Since we only observe $n$ i.i.d. samples from $\nu \in \mathcal{P}([0,1]^d)$, any upper bound on an estimator for this problem instance also gives an upper bound for $n$-sample distribution estimation of $\nu$ under $\mathsf{W}_p$. However, the minimax lower bound of Singh and Póczos [2018] implies that no distribution estimator can achieve $\mathsf{W}_p$ error less than $n^{-1/(2p \vee d)}$ for all $\nu \in \mathcal{P}([0,1]^d)$.

### C.3 Proof of Theorem 2

We start with some helpful lemmas.

**Lemma 7** (Rigollet [2015], Theorem 1.14)**.** *Let $\mu \in \mathcal{P}(\mathbb{R})$ be 1-sub-Gaussian. Then, for $X_1, \dots, X_n$ sampled i.i.d. from $\mu$, we have $\max_{i=1,\dots,n} X_i \le \sqrt{2\log(n/\delta)}$ with probability at least $1-\delta$.*

**Lemma 8.** *Let $\mu \in \mathcal{P}(\mathbb{R}^d)$ be 1-sub-Gaussian and let $\mathcal{P}$ denote the regular partition of $\mathbb{R}^d$ into cubes of side-length $r > 0$. Then, for any choice of rounding map $r_\mathcal{P}$, we have*

$$\mathbb{E}[\|(r_\mathcal{P})_\sharp(\hat{\mu}_n - \mu)\|_{\mathrm{TV}}] = \widetilde{O}\left(\sqrt{\frac{5^d r^{-d}}{n}}\right).$$

*Proof.* Let $B$ denote a ball of radius $R = \sqrt{2\log(n)}$ centered at the origin, so that $\mu(B) \ge 1 - 1/n$ by Lemma 7. Write $\mathcal{P}_R$ for the subset of partition blocks $P \in \mathcal{P}$ which intersect $B$, and note that $|\mathcal{P}_R| \le \mathrm{vol}(B) r^{-d} \le (3R/r)^d$. We then bound

$$\begin{aligned}\mathbb{E}[\|(r_\mathcal{P})_\sharp(\hat{\mu}_n - \mu)\|_{\mathrm{TV}}] &= \frac{1}{2}\mathbb{E}\left[\sum_{P\in\mathcal{P}} |(\hat{\mu}_n - \mu)(P)|\right]\\ &= \frac{1}{2}\mathbb{E}\left[\sum_{P\in\mathcal{P}_R} |(\hat{\mu}_n - \mu)(P)| + \sum_{P\in\mathcal{P}\setminus\mathcal{P}_R} |(\hat{\mu}_n - \mu)(P)|\right]\\ &\lesssim \sqrt{\frac{|\mathcal{P}_R|}{n}} + \mathbb{E}\left[\sum_{P\in\mathcal{P}\setminus\mathcal{P}_R} \hat{\mu}_n(P) + \mu(P)\right]\\ &\lesssim \sqrt{\frac{|\mathcal{P}_R|}{n}} + \mu(\mathbb{R}^d \setminus B)\\ &\lesssim \sqrt{\frac{(3\sqrt{2\log(n)})^d r^{-d}}{n}} + \frac{1}{n}\\ &= \widetilde{O}\left(\sqrt{\frac{5^d r^{-d}}{n}}\right),\end{aligned}$$

as desired. □

**Lemma 9.** *There exists a partition $\mathcal{P}$ parameterized by $\delta > 0$ such that, for all $\mu \in \mathcal{P}(\mathbb{R}^d)$ with $\mathbb{E}_\mu[\|X\|^{p+1}] \le 1$ and any rounding map $r_\mathcal{P}$, we have $\mathbb{E}[\|(r_\mathcal{P})_\sharp(\hat{\mu}_n - \mu)\|_{\mathrm{TV}}] = \widetilde{O}\big(\sqrt{\delta^{-d}/n}\big)$ and $\|r_\mathcal{P} - \mathrm{Id}\,\|_{L^p(\mu)} \lesssim \delta$.*

*Proof.* Let $X_0$ be a minimal $(3\delta)$-covering of the unit ball, denoted $S_0$. In particular, this implies that $|X_0| \le \delta^{-d}$. Now, take $\mathcal{P}_0$ to be the Voronoi partition of $S_0$ induced by $X_0$, so that $\mathcal{P}_0$ has at most $\delta^{-d}$ cells of diameter at most $6\delta$. Then for each integer $i > 0$, set $S_i \coloneqq 2^i S_0 \setminus 2^{i-1} S_0$, and let $\mathcal{P}_i$ be the dilated partition $\{(2^i P) \cap S_i : P \in \mathcal{P}_0\}$. By construction, $|\mathcal{P}_i| \le \delta^{-d}$ and each $P \in \mathcal{P}_i$ has diameter at most $2^i \cdot 6\delta$, for all $i \ge 0$. Moreover, by Markov's inequality, we have

$$\mu(S_i) \le \Pr_{\mu}\bigl(\|X\| > 2^{i-1}\bigr) = \Pr_{\mu}\bigl(\|X\|^{p+1} > 2^{(p+1)(i-1)}\bigr) \le 2^{(1-i)(p+1)}$$

for each $i > 0$. We thus bound

$$\begin{aligned}
\mathbb{E}[\|(r_{\mathcal{P}})_\sharp(\mu - \hat{\mu}_n)\|_{\mathrm{TV}}] &= \mathbb{E}\left[\sum_{i=0}^{\infty} \sum_{P \in \mathcal{P}_i} |(\mu - \hat{\mu}_n)(P)|\right] \\
&\le \sum_{i=0}^{\infty} \sum_{P \in \mathcal{P}_i} \sqrt{\mathrm{Var}_{\mu^{\otimes n}}[\hat{\mu}_n(P)]} \\
&\le \frac{1}{\sqrt{n}} \cdot \sum_{i=0}^{\infty} \sum_{P \in \mathcal{P}_i} \sqrt{\mu(P)} \\
&\le \frac{1}{\sqrt{n}} \cdot \sum_{i=0}^{\infty} \sqrt{|\mathcal{P}_i| \mu(S_i)} \\
&\le (\delta^d n)^{-\frac{1}{2}} \cdot \left(1 + \sum_{i=1}^{\infty} 2^{\frac{1-i}{2}}\right) \\
&\lesssim (\delta^d n)^{-\frac{1}{2}}.
\end{aligned}$$

Similarly, we bound

$$\begin{aligned}
\|r_{\mathcal{P}} - \mathrm{Id}\,\|_{L^p(\mu)} &\le \left(\sum_{i=0}^{\infty} \sum_{P \in \mathcal{P}_i} \mu(P) \operatorname{diam}(P)^p\right)^{\frac{1}{p}} \\
&\le \left(\sum_{i=0}^{\infty} (2^i \cdot 6\delta)^p \mu(S_i)\right)^{\frac{1}{p}} \\
&\lesssim \left(\sum_{i=0}^{\infty} (2^i \delta)^p 2^{(1-i)(p+1)}\right)^{\frac{1}{p}} \\
&= \delta \left(\sum_{i=0}^{\infty} 2^{p+1-i}\right)^{\frac{1}{p}} \\
&\lesssim \delta,
\end{aligned}$$

as desired. □

We now prove a slight strengthening of the theorem which does not explicitly trim the target points, i.e., we take $R = \infty$ so that $\hat{\nu}_n = \nu_n'$. We do perform a trimming-type step in our analysis, noting that for $D = \sqrt{4\log(n)}$, we have $\max_{i=1,\dots,n} \|Y_i\| \le DS$ with probability at least $1/n$, by Lemma 7. Now, for a general partition $\mathcal{P}$, we bound

$$\begin{aligned}
\mathcal{E}_p(\hat{\kappa}_n; \mu, \nu) &= \mathcal{E}_p(\bar{\kappa}_n \circ r_{\mathcal{P}}; \mu, \nu) \\
&\le \mathcal{E}_p(\bar{\kappa}_n; \mu', \nu) + 2\,\|r_{\mathcal{P}} - \mathrm{Id}\,\|_{L^p(\mu)} && \text{(Lemma 5)} \\
&\le \mathcal{E}_p(\bar{\kappa}_n; \mu', \hat{\nu}_n) + 2\,\|r_{\mathcal{P}} - \mathrm{Id}\,\|_{L^p(\mu)} + \mathsf{W}_p(\nu, \hat{\nu}_n) && \text{(Lemma 2)} \\
&\lesssim \|(r_{\mathcal{P}})_\sharp(\mu - \hat{\mu}_n)\|_{\mathrm{TV}}^{1/p} \cdot \operatorname{diam}(\operatorname{supp}(\hat{\nu}_n)) + \delta^{\frac{1}{p}} + \|r_{\mathcal{P}} - \mathrm{Id}\,\|_{L^p(\mu)} + \mathsf{W}_p(\nu, \hat{\nu}_n),
\end{aligned}$$

where the last inequality follows by Lemma 6 and our choice of $\bar{\kappa}_n$. Applying this bound for the regular cube partition and taking expectations, we bound

$$\mathbb{E}[\mathcal{E}_p(\hat{\kappa}_n;\mu,\nu)] \lesssim D\,\mathbb{E}[\|(r_{\mathcal{P}})_\sharp(\mu-\hat{\mu}_n)\|_{\mathrm{TV}}]^{1/p} + \frac{1}{n} + \delta^{\frac{1}{p}} + \sqrt{d}r + \mathsf{W}_p(\nu,\hat{\nu}_n)$$
$$\lesssim \widetilde{O}\left(\frac{5^d r^{-d}}{n}\right)^{\frac{1}{2p}} + \delta^{\frac{1}{p}} + \sqrt{d}r + \widetilde{O}_p\left(n^{-\frac{1}{d\vee 2p}}\right). \qquad \text{(Lemmas 1, 7 and 8)}$$

Taking $r = n^{-1/(d+2p)}$, we obtain $\mathbb{E}[\mathcal{E}_p(\hat{\kappa}_n;\mu,\nu)] = \widetilde{O}_{p,d}(n^{-1/(d+2p)}) + \delta^{1/p}$. The same rate is obtained under bounded $2p$th moments by using the alternative partition from Lemma 9. Thus, to achieve the desired rate, it suffices to solve the preliminary OT problem to accuracy $\delta = n^{-p/(d+2p)}$.

Computational complexity is dominated by this OT computation. The source and target distributions are both supported on $n$ points, and we require accuracy $\delta = n^{-p/(d+2p)}$. Computing the relevant cost matrix requires time $O(n^2 d)$. Using a state of the art OT solver based on entropic OT (e.g., Luo et al., 2023) gives a running time of $O(C_\infty n^2/\delta) = O(C_\infty n^{2+p/(d+2p)})$, where $C_\infty$ is the largest distance between a source point and a target point.

### C.4 One-Dimensional Refinements (Remark 3)

In one dimension, OT maps can be expressed concisely in terms of CDFs; in particular, if $\mu$ and $\nu$ have strictly increasing CDFs $F_\mu$ and $F_\nu$, respectively, then the map $T^\star(x) = F_\nu^{-1}(F_\mu(x))$ solves the $\mathsf{W}_p(\mu,\nu)$ problem for all $p \geq 1$. As a result, many OT-based inference tasks become more analytically tractable when $d=1$, including map estimation. In fact, minor adjustments to folklore techniques imply that the optimal risk of $n^{-1/(2p)}$ is achievable when $d=1$. We now provide a clean derivation of this risk bound using the Kolmogorov-Smirnov (KS) distance.

The KS distance is a useful alternative to the TV metric in one dimension, defined via $\|\mu-\nu\|_{\mathrm{KS}} := \sup_{t\in\mathbb{R}} |(\mu-\nu)((-\infty,t])| = \|F_\mu - F_\nu\|_\infty$. We always have $\|\mu-\nu\|_{\mathrm{KS}} \leq \|\mu-\nu\|_{\mathrm{TV}}$, since $\|\mu-\nu\|_{\mathrm{TV}}$ can alternatively be expressed as $\sup_{A \text{ meas.}} |(\mu-\nu)(A)|$. A comparison with $\mathsf{W}_p$ mirroring Fact 1 is direct.

[proof] **Lemma 10** ($\mathsf{W}_p$-KS comparison)**.** *For $\mu,\nu \in \mathcal{P}([0,D])$, we have $\mathsf{W}_p(\mu,\nu) \leq D\|\mu-\nu\|_{\mathrm{KS}}^{1/p}$.*

*Proof.* Writing $F, G$ for the CDFs of $\mu$ and $\nu$, with generalized inverses $F^{-1}$ and $G^{-1}$, we bound

$$\begin{aligned}
\mathsf{W}_p(\mu,\nu)^p &= \int_0^1 |F^{-1}(u) - G^{-1}(u)|^p \mathrm{d}u \\
&\leq D^{p-1} \int_0^1 |F^{-1}(u) - G^{-1}(u)| \mathrm{d}u \\
&= D^{p-1} \int_0^D |F(x) - G(x)| \mathrm{d}x \\
&\leq D^p \|\mu-\nu\|_{\mathrm{KS}}.
\end{aligned}$$

Taking $p$th roots gives the statement. ☐

The KS distance admits useful empirical convergence guarantees not shared by the TV distance.

**Fact 2** (KS empirical convergence, Massart, 1990)**.** For all $\mu \in \mathcal{P}(\mathbb{R})$, $\mathbb{E}[\|\mu-\hat{\mu}_n\|_{\mathrm{KS}}] \leq 1/\sqrt{n}$.

Moreover, for fixed $\mu$ and $\nu$, there exists an optimal kernel for $\mathsf{W}_p(\mu,\nu)$ (namely, based on CDFs as above), which is near-optimal for all $\mu'$ in a KS neighborhood of $\mu$, as shown next.

[proof] **Lemma 11** (KS corruptions in $\mu$)**.** *For $\mathcal{X},\mathcal{Y} \subseteq \mathbb{R}$, fix $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$. There exists an optimal kernel $\kappa^\star \in \mathcal{K}(\mathcal{X},\mathcal{Y})$ for the $\mathsf{W}_p(\mu,\nu)$ problem such that, for all $\mu' \in \mathcal{P}(\mathcal{X})$, we have*

$$\mathcal{E}_p(\kappa^\star;\mu',\nu) \lesssim \mathrm{diam}(\mathcal{Y})\|\mu-\mu'\|_{\mathrm{KS}}^{1/p}.$$

*Proof.* Write $F, F', G$, for the CDFs of $\mu$, $\mu'$, and $\nu$, respectively, and let $\varepsilon = \|\mu-\mu'\|_{\mathrm{KS}} = \|F-F'\|_\infty$. Write $D = \mathrm{diam}(\mathcal{Y})$ and suppose without loss of generality that $\mathcal{Y} = [0,D]$. For now, suppose further that $1/\varepsilon = 3M$ is a multiple of 3 (without loss of generality) and that $F$ is continuous (which will be relaxed). We consider the kernel induced by the map $T^\star = G^{-1} \circ F$, where $G^{-1}$ is the generalized inverse of $G$ with $G^{-1}(q)$ defined as as 0 for $q \leq 0$ and $D$ for $q \geq 1$. In particular, we compare $T^\star$ with the optimal kernel $G^{-1} \circ F'$ for the $\mathsf{W}_p(\mu', \nu)$ problem, bounding

$$\begin{aligned}\int_{\mathcal{X}} \left|G^{-1}(F(x)) - G^{-1}(F'(x))\right|^p \mathrm{d}F'(x) &\leq \int_{\mathcal{X}} \left|G^{-1}(F'(x) \pm \varepsilon) - G^{-1}(F'(x))\right|^p \mathrm{d}F'(x) \\ &= \int_0^1 \left|G^{-1}(u \pm \varepsilon) - G^{-1}(u)\right|^p \mathrm{d}u \\ &= \sum_{i=0}^{3M-1} \int_{i\varepsilon}^{(i+1)\varepsilon} \left|G^{-1}(u \pm \varepsilon) - G^{-1}(u)\right|^p \mathrm{d}u \\ &\leq \varepsilon \sum_{i=0}^{3M-1} \left[G^{-1}((i+2)\varepsilon) - G^{-1}((i-1)\varepsilon)\right]^p.\end{aligned}$$

Here, the first equality uses that $F'_\sharp \mu' = \mathrm{Unif}([0,1])$, and the second inequality uses that $G^{-1}$ is monotonic. If $F'$ is discontinuous, one should replace it with the kernel $\tilde{F}'$ which coincides with $F'$ where continuous and, at any point $x$ where there is a jump from $p_1$ to $p_2$, satisfies $\tilde{F}'_\sharp \delta_x = \mathrm{Unif}([p_1, p_2])$. By this choice, we have $\tilde{F}'_\sharp \mu = \mathrm{Unif}([0,1])$, and one can do the same for $F$ to obtain $\tilde{F}$ such that $\mathsf{W}_\infty(\tilde{F}_\sharp \mu, \tilde{F}'_\sharp \mu) \leq \varepsilon$. At this point, we can derive the same bound as above. Now, writing $\Delta_i = G^{-1}((i+3)\varepsilon) - G^{-1}(i\varepsilon)$, we have

$$\begin{aligned}&\int_{\mathcal{X}} \left|G^{-1}(F(x)) - G^{-1}(F'(x))\right|^p \mathrm{d}F'(x) \\ &\qquad \leq \varepsilon \sum_{i=-1}^{3M-2} \Delta_i^p \\ &\qquad = \varepsilon \left( \sum_{i=0}^{M-1} \Delta_{3i-1}^p + \sum_{i=0}^{M-1} \Delta_{3i}^p + \sum_{i=0}^{M-1} \Delta_{3i+1}^p \right) \\ &\qquad = \varepsilon D^p \left( \sum_{i=0}^{M-1} \left(\frac{\Delta_{3i-1}}{D}\right)^p + \sum_{i=0}^{M-1} \left(\frac{\Delta_{3i}}{D}\right)^p + \sum_{i=0}^{M-1} \left(\frac{\Delta_{3i+1}}{D}\right)^p \right) \\ &\qquad \leq \varepsilon D^p \left( \left(\sum_{i=0}^{M-1} \frac{\Delta_{3i-1}}{D}\right)^p + \left(\sum_{i=0}^{M-1} \frac{\Delta_{3i}}{D}\right)^p + \left(\sum_{i=0}^{M-1} \frac{\Delta_{3i+1}}{D}\right)^p \right) \\ &\qquad = O(D\varepsilon^{1/p})^p.\end{aligned}$$

Thus, we have $\mathcal{E}(G^{-1} \circ F; \mu', \nu) \lesssim \|G^{-1} \circ F - G^{-1} \circ F'\|_{L^p(\mu')} \leq D\varepsilon^{1/p}$, as desired. □

Together, the three results stated above yield our desired risk bound.

**Proposition 2.** *Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mu \in \mathcal{P}(\mathbb{R})$ and $Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} \nu \in \mathcal{P}([0,1])$. Then the estimator $\hat{\kappa}_n$ which, given $\hat{\mu}_n$ and $\hat{\nu}_n$, returns the optimal kernel for $\mathsf{W}_p(\hat{\mu}_n, \hat{\nu}_n)$ given by Lemma 11, achieves risk $\mathbb{E}[\mathcal{E}_p(\hat{\kappa}_n; \mu, \nu)] \lesssim n^{-1/(2p)}$.*

*Proof.* By Fact 2, we have that $\mathbb{E}[\|\mu - \hat{\mu}_n\|_{\mathrm{KS}}] \leq n^{-1/2}$. Consequently, we bound

$$\begin{aligned}\mathcal{E}_p(\hat{\kappa}_n; \mu, \nu) &\leq \mathcal{E}_p(\hat{\kappa}_n; \mu, \hat{\nu}_n) + \mathsf{W}_p(\nu, \hat{\nu}_n) \\ &\leq \|\mu - \hat{\mu}_n\|_{\mathrm{KS}}^{1/p} + \mathsf{W}_p(\nu, \hat{\nu}_n).\end{aligned}$$

Taking expectations and applying Fact 2 and Lemma 1 gives the desired rate. □

Unfortunately, we are unaware of any multivariate extension of the KS distance that obeys a useful comparison inequality with $\mathsf{W}_p$ (like Fact 10) while maintaining strong empirical convergence guarantees (like Fact 2), inhibiting the further development of this approach.

# D Additional Details for Section 4

We note that the minimax lower bounds in Corollaries 1 and 2 follow by combining the reduction to distribution estimation from Appendix C.2 with existing lower bounds for distribution estimation under $\mathsf{W}_p$ from Singh and Póczos [2018] and Weed and Berthet [2019], respectively.

# E Proofs for Section 5

We first recall some basic facts used throughout.

**Fact 3** (TV contraction under Markov kernels)**.** For $\mu, \nu \in \mathcal{P}(\mathcal{X})$ and kernel $\kappa \in \mathcal{K}(\mathcal{X}, \mathcal{Y})$, we have $\|\kappa_\sharp \mu - \kappa_\sharp \nu\|_{\mathrm{TV}} \leq \|\mu - \nu\|_{\mathrm{TV}}$.

This follows by the data processing inequality.

**Fact 4** ($\mathsf{W}_p$ contraction under convolution)**.** For $\mu, \nu, \alpha \in \mathcal{P}(\mathcal{X})$, we have $\mathsf{W}_p(\mu * \alpha, \nu * \alpha) \leq \mathsf{W}_p(\mu, \nu)$, where $*$ denotes convolution between probability measures.

This follows by considering the couplings $(X + Z, Y + Z')$ of $\mu * \alpha$ and $\nu * \alpha$ which set $Z = Z'$.

**Fact 5** (TV discrete empirical convergence)**.** For a finite set $S$ with $|S| = k$, any distribution $\mu \in \Delta(S)$ exhibits empirical convergence in TV at rate $\mathbb{E}[\|\hat{\mu}_n - \mu\|_{\mathrm{TV}}] \lesssim \sqrt{k/n}$.

To simplify discussion of our corruption model, we employ the *$\varepsilon$-outlier-robust $p$-Wasserstein distance*

$$\mathsf{W}_p^\varepsilon(\mu, \nu) := \min_{\substack{\mu' \in \mathcal{P}(\mathbb{R}^d) \\ \|\mu' - \mu\|_{\mathrm{TV}} \leq \varepsilon}} \mathsf{W}_p(\mu', \nu) = \min_{\substack{\nu' \in \mathcal{P}(\mathbb{R}^d) \\ \|\nu' - \nu\|_{\mathrm{TV}} \leq \varepsilon}} \mathsf{W}_p(\mu, \nu'). \tag{8}$$

The second equality follows from the observation that, if $\mathbb{E}[\|X' - Y\|^p] \leq c$ and $X = X'$ with probability at least $1 - \varepsilon$, then the random variable $Y' = Y\mathbb{1}\{X = X'\} + X\mathbb{1}\{X \neq X'\}$ satisfies $\mathbb{E}[\|X - Y'\|^p] \leq c$. See Nietert et al. [2023a] for a thorough examination of $\mathsf{W}_p^\varepsilon$ in the context of robust statistics. Under the setting of Section 5, our corruption model can be equivalently stated as follows: given the standard empirical measures $\hat{\mu}_n \in \mathcal{P}(\mathcal{X})$ and $\hat{\nu}_n \in \mathcal{P}(\mathcal{Y})$, we observe corrupted versions $\tilde{\mu}_n \in \mathcal{P}(\mathcal{X})$ and $\tilde{\mu}_n \in \mathcal{P}(\mathcal{Y})$ such that $\mathsf{W}_p^\varepsilon(\tilde{\mu}_n, \hat{\mu}_n) \vee \mathsf{W}_p^\varepsilon(\tilde{\nu}_n, \hat{\nu}_n) \leq \rho$.

For this setting, we handle sampling error using the following lemma, which mirrors Lemma 8.

**Lemma 12** (Prop. 2 of Goldfeld et al., 2020)**.** *Fix $\sigma > 0$ and $1$-sub-Gaussian $\mu \in \mathcal{P}(\mathbb{R}^d)$. Then, the $n$-sample empirical measure $\hat{\mu}_n$ satisfies $\mathbb{E}\big[\|N_\sharp^\sigma(\mu - \hat{\mu}_n)\|_{\mathrm{TV}}\big] \leq \sqrt{3^d(1 \vee \sigma^{-d})/n}$.*

In order to apply our $\mathsf{W}_p$ stability result, Lemma 3, we use that any kernel become continuous if one first applies Gaussian convolution.

**Lemma 13.** *Fix $\bar{\kappa} \in \mathcal{K}(\mathcal{X}, \mathcal{Y})$, $\sigma > 0$, and let $\kappa = \bar{\kappa} \circ N^\sigma$. Then, for all $x, x' \in \mathcal{X}$, we have $\mathsf{W}_p((\kappa_x, \kappa_{x'}) \leq \mathrm{diam}(\mathcal{Y})[\|x - x'\|/(2\sigma)]^{1/p}$.*

*Proof.* We simply compute

$$\begin{aligned}
\mathsf{W}_p(\kappa_x, \kappa_{x'}) &\leq \mathrm{diam}(\mathcal{Y})\|\kappa_\sharp(N_x^\sigma - N_{x'}^\sigma)\|_{\mathrm{TV}}^{1/p} && \text{(Fact 1)}\\
&\leq \mathrm{diam}(\mathcal{Y})\|N_x^\sigma - N_{x'}^\sigma\|_{\mathrm{TV}}^{1/p} && \text{(data processing ineq.)}\\
&\leq \mathrm{diam}(\mathcal{Y})\|\mathcal{N}(x, \sigma^2 I_d) - \mathcal{N}(x', \sigma^2 I_d)\|_{\mathrm{TV}}^{1/p}\\
&\leq \mathrm{diam}(\mathcal{Y})\|x - x'\|^{1/p}(2\sigma)^{-1/p},
\end{aligned}$$

where the final inequality follows by the closed form of KL divergence between Gaussians, combined with Pinsker's inequality. $\square$

We split the proof of Theorem 4 into the upper bound (Appendix E.1) and lower bound (Appendix E.2).

### E.1 Proof of Theorem 4 (Upper Bound)

To start, we decompose $N^\sigma = N^{\sigma_1+\sigma_2} = N^{\sigma_1} \circ N^{\sigma_2}$, for $\sigma_1, \sigma_2$ to be tuned later. By our corruption model, there exists an intermediate measure $\mu_n' \in \mathcal{P}(\mathbb{R}^d)$ such that $\|\hat{\mu}_n - \mu_n'\|_{\mathrm{TV}} \le \varepsilon$ and $\mathsf{W}_p(\mu_n', \tilde{\mu}_n) \le \rho$. By Facts 3 and 4, these bounds are preserved under convolution, so $\|N_\sharp^{\sigma_1}(\hat{\mu}_n - \mu_n')\|_{\mathrm{TV}} \le \varepsilon$ and $\mathsf{W}_p(N_\sharp^{\sigma_1}\mu_n', N_\sharp^{\sigma_1}\tilde{\mu}_n) \le \rho$. By the TV triangle inequality, we have $\|N_\sharp^{\sigma_1}(\mu - \mu_n')\|_{\mathrm{TV}} \le \tau \coloneqq \varepsilon + \|N_\sharp^{\sigma_1}(\mu - \hat{\mu}_n)\|_{\mathrm{TV}}$. We conclude that $\mathsf{W}_p^\tau(N_\sharp^{\sigma_1}\tilde{\mu}_n, N_\sharp^{\sigma_1}\mu) \le \rho$. By the symmetric nature of $\mathsf{W}_p^\tau$, there must also exist $\alpha \in \mathcal{P}(\mathbb{R}^d)$ such that $\mathsf{W}_p(N_\sharp^{\sigma_1}\mu, \alpha) \le \rho$ and $\|\alpha - N_\sharp^{\sigma_1}\tilde{\mu}_n\|_{\mathrm{TV}} \le \tau$.

Now set $\bar{\kappa} = \kappa_p^\star[N_\sharp^\sigma \tilde{\mu}_n \to \tilde{\nu}_n]$, so that $\mathcal{E}_p(\bar{\kappa}; N_\sharp^\sigma\tilde{\mu}_n, \tilde{\nu}_n) = 0$. Using this, the TV bound above, and the fact that $N_\sharp^\sigma \tilde{\mu}_n = N_\sharp^{\sigma_2}(N_\sharp^{\sigma_1}\tilde{\mu}_n)$, we have $\mathcal{E}_p(\bar{\kappa}; N_\sharp^{\sigma_2}\alpha, \tilde{\nu}_n) \lesssim \mathrm{diam}(\mathcal{Y})\tau^{1/p} \le \sqrt{d}\tau^{1/p}$. Applying Lemma 5, this gives

$$\mathcal{E}_p(\bar{\kappa} \circ N^{\sigma_2}; \alpha, \tilde{\nu}_n) \lesssim \sqrt{d}\tau^{\frac{1}{p}} + \mathbb{E}_{Z\sim\mathcal{N}(0,\sigma_2^2 I_d)}[\|Z\|^p]^{\frac{1}{p}} \lesssim \sqrt{d}\tau^{\frac{1}{p}} + \sqrt{d+p}\,\sigma_2.$$

Consequently, by Lemma 3 and Lemma 13, we have that

$$\begin{aligned}
\mathcal{E}_p(\bar{\kappa} \circ N^{\sigma_2}; N_\sharp^{\sigma_1}\mu, \tilde{\nu}_n) &\lesssim \mathcal{E}_p(\bar{\kappa} \circ N^{\sigma_2}; \alpha, \tilde{\nu}_n) + \rho + (\sqrt{d}/\sigma_2)^{1/p}\rho^{1/p} \\
&\lesssim \mathcal{E}_p(\bar{\kappa} \circ N^{\sigma_2}; \alpha, \tilde{\nu}_n) + \rho + (\sqrt{d}/\sigma_2)^{1/p}\rho^{1/p} \\
&\le \sqrt{d}\tau^{\frac{1}{p}} + \rho + (\sqrt{d}/\sigma_2)^{1/p}\rho^{1/p} + \sqrt{d+p}\,\sigma_2.
\end{aligned}$$

Tuning $\sigma_2$ gives

$$\mathcal{E}_p(\bar{\kappa} \circ N^{\sigma_2}; N_\sharp^{\sigma_1}\mu, \tilde{\nu}_n) \lesssim \sqrt{d}\tau^{\frac{1}{p}} + \sqrt{d}\rho^{\frac{1}{p+1}} + \rho.$$

Apply Lemma 5 once more, we bound

$$\begin{aligned}
\mathcal{E}_p(\bar{\kappa} \circ N^{\sigma}; \mu, \tilde{\nu}_n) &\lesssim \sqrt{d}\tau^{\frac{1}{p}} + \sqrt{d}\rho^{\frac{1}{p+1}} + \rho + \sqrt{d+p}\,\sigma_1 \\
&\lesssim \sqrt{d}\varepsilon^{\frac{1}{p}} + \sqrt{d}\rho^{\frac{1}{p+1}} + \rho + \sqrt{d+p}\,\sigma_1 + \sqrt{d}\,\|N_\sharp^{\sigma_1}(\mu - \hat{\mu}_n)\|_{\mathrm{TV}}^{1/p}
\end{aligned}$$

Taking expectations and applying Lemma 12 yields

$$\begin{aligned}
\mathbb{E}[\mathcal{E}_p(\bar{\kappa} \circ N^{\sigma}; \mu, \tilde{\nu}_n)] &\lesssim \sqrt{d}\varepsilon^{\frac{1}{p}} + \sqrt{d}\rho^{\frac{1}{p+1}} + \rho + \sqrt{d+p}\,\sigma_1 + \mathbb{E}[\|N_\sharp^{\sigma_1}(\mu - \hat{\mu}_n)\|_{\mathrm{TV}}]^{\frac{1}{p}} \\
&\lesssim \sqrt{d}\varepsilon^{\frac{1}{p}} + \sqrt{d}\rho^{\frac{1}{p+1}} + \rho + \sqrt{d+p}\,\sigma_1 + \left(\frac{3^d(1 \vee \sigma^{-d})}{n}\right)^{\frac{1}{2p}}.
\end{aligned}$$

Tuning $\sigma_1$ then gives

$$\mathbb{E}[\mathcal{E}_p(\bar{\kappa} \circ N^{\sigma}; \mu, \tilde{\nu}_n)] \lesssim \sqrt{d}\varepsilon^{\frac{1}{p}} + \sqrt{d}\rho^{\frac{1}{p+1}} + \rho + O_{p,d}(n^{-\frac{1}{d+2p}}).$$

Finally, we note that $\mathsf{W}_p(\tilde{\nu}_n, \hat{\nu}_n) \le \rho + \sqrt{d}\varepsilon^{1/p}$ due to the support bound. Thus, Lemma 2 gives

$$\begin{aligned}
\mathbb{E}[\mathcal{E}_p(\bar{\kappa} \circ N^{\sigma}; \mu, \nu)] &\le \mathbb{E}[\mathcal{E}_p(\bar{\kappa} \circ N^{\sigma}; \mu, \tilde{\nu}_n) + \mathsf{W}_p(\tilde{\nu}_n, \hat{\nu}_n) + \mathsf{W}_p(\hat{\nu}_n, \nu)] \\
&\le \mathbb{E}[\mathcal{E}_p(\bar{\kappa} \circ N^{\sigma}; \mu, \tilde{\nu}_n)] + \rho + \sqrt{d}\varepsilon^{\frac{1}{p}} + \mathbb{E}[\mathsf{W}_p(\hat{\nu}_n, \nu)] \\
&\lesssim \sqrt{d}\varepsilon^{\frac{1}{p}} + \sqrt{d}\rho^{\frac{1}{p+1}} + \rho + O_{p,d}(n^{-\frac{1}{d+2p}}),
\end{aligned}$$

as desired.

For the null estimator, let $\kappa^\star$ be an optimal kernel for the $\mathsf{W}_p(\mu, \nu)$ problem and bound

$$\begin{aligned}
\mathcal{E}(\hat{\kappa}_{\mathrm{null}}; \mu, \nu) &= \left[\left(\int \|x\|^p \mathrm{d}\mu(x)\right)^{\frac{1}{p}} - \left(\iint \|y - x\|^p \mathrm{d}\kappa_x^\star(y)\mathrm{d}\mu(x)\right)^{\frac{1}{p}}\right]_+ + \mathsf{W}_p(\delta_0, \nu) \\
&\le \left[\left(\iint \|y\|^p \mathrm{d}\kappa_x^\star(y)\mathrm{d}\mu(x)\right)^{\frac{1}{p}}\right]_+ + \mathsf{W}_p(\delta_0, \nu) && \text{(Minkowski's inequality)} \\
&\le 2\sqrt{d}, && (\mathcal{Y} \subseteq [0,1]^d)
\end{aligned}$$

as desired.

### E.2 Proof of Theorem 4 (Lower Bound)

**Correction:** Our current lower bound construction does not fully match the claim in the statement of Theorem 4. Rather, we prove the weaker bound of $\sqrt{d}\varepsilon^{1/p} + \rho^{1/2}d^{1/4} \wedge \sqrt{d} + n^{-1/(d \vee 2p)}$. Still, this is sufficient to establish the separation between distribution estimation and kernel estimation discussed after the theorem.

We inherit the $n^{-1/(d\vee 2p)}$ sampling error term of the lower bound from the Dirac mass construction described in Appendix C.2. For the remaining terms, we prove lower bounds which hold even in the infinite-sample population limit, and even when only the source measure is corrupted. Here, an estimator can be viewed as a map $\hat{\kappa}$ from $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y}) \to \mathcal{K}(\mathcal{X}, \mathcal{Y})$, mapping the corrupted source measure $\mu$, guaranteed to satisfy $\mathsf{W}_p^\varepsilon(\tilde{\mu}, \mu) \le \rho$, and the clean target measure $\nu$ to a kernel estimate $\hat{\kappa}[\tilde{\mu}, \nu]$. For $\mathcal{X} = \mathbb{B}^d$ (which, in particular, forces each $\mu \in \mathcal{P}(\mathcal{X})$ to be 1-sub-Gaussian) and $\mathcal{Y} = [-1, 1]^d$, we prove that

$$\sup_{\substack{\mu\in\mathcal{P}(\mathcal{X})\\ \nu\in\mathcal{P}(\mathcal{Y})}} \sup_{\substack{\tilde{\mu}\in\mathcal{P}(\mathcal{X})\\ \mathsf{W}_p^\varepsilon(\tilde{\mu},\mu)\le\rho}} \mathcal{E}_p(\hat{\kappa}[\tilde{\mu},\nu];\mu,\nu) \gtrsim \sqrt{d}\varepsilon^{\frac{1}{p}} + \rho^{\frac{1}{2}} d^{\frac{1}{4}} \wedge \sqrt{d}.$$

The choice of $\mathcal{Y} = [-1, 1]^d$ rather than $[0, 1]^d$ is solely to simplify notation in one of our constructions and can be reverted without loss. Finally, it suffices to lower bound the supremum by $\sqrt{d}\varepsilon^{1/p}$ when $\rho = 0$ and $\sqrt{d\rho} \wedge \sqrt{d}$ when $\varepsilon = 0$, separately, which we do presently.

**TV lower bound.** Fix target measure $\nu = (1-\varepsilon)\delta_0 + \varepsilon\delta_y$, where $y = (1, \dots, 1) \in \mathbb{R}^d$. Consider the candidate clean measures $\mu_1 = \nu$ and $\mu_2 = \delta_0$. Because they are within TV distance $\varepsilon$, the observation $\tilde{\mu} = \nu$ is compatible with both candidates. Abbreviating $\kappa = \hat{\kappa}[\tilde{\mu}, \nu]$, we have

$$\begin{aligned}
\mathcal{E}_p(\kappa;\mu_1,\nu) + \mathcal{E}_p(\kappa;\mu_2,\nu) &\ge \left[\left(\iint \|y-x\|^p \mathrm{d}\kappa_x(y)\mu_1(x)\right)^{\frac{1}{p}} - \mathsf{W}_p(\mu_1,\nu)\right]_+ + \mathsf{W}_p(\kappa_\sharp\mu_2,\nu)\\
&= \left(\iint \|y-x\|^p \mathrm{d}\kappa_x(y)\nu(x)\right)^{\frac{1}{p}} + \mathsf{W}_p(\kappa_\sharp\delta_0,\nu)\\
&\ge (1-\varepsilon)^{\frac{1}{p}}\left(\int \|y\|^p \mathrm{d}\kappa_0(y)\right)^{\frac{1}{p}} + \mathsf{W}_p(\kappa_\sharp\delta_0,\nu)\\
&\ge (1-\varepsilon)^{\frac{1}{p}}\mathsf{W}_p(\kappa_\sharp\delta_0,\delta_0) + (1-\varepsilon)^{\frac{1}{p}}\mathsf{W}_p(\kappa_\sharp\delta_0,\nu)\\
&\ge (1-\varepsilon)^{\frac{1}{p}}\mathsf{W}_p(\delta_0,\nu)\\
&\ge (1-\varepsilon)^{\frac{1}{p}}\varepsilon^{\frac{1}{p}}\sqrt{d}\\
&\ge \frac{1}{2}\varepsilon^{\frac{1}{p}}\sqrt{d}.
\end{aligned}$$

Thus, we must have $\mathcal{E}_p(\kappa;\mu_1,\nu) \vee \mathcal{E}_p(\kappa;\mu_2,\nu) \gtrsim \sqrt{d}\varepsilon^{1/p}$, as desired.

**$\mathsf{W}_p$ lower bound.** For the remaining bound, we first argue that, for any kernel $\kappa$, its performance for the $\mathsf{W}_p(\mu,\nu)$ problem cannot suffer to much if we compose it with the Euclidean projection onto $\mathrm{supp}(\nu)$, denoted by $\mathrm{proj}_{\mathrm{supp}(\nu)}$.

**Lemma 14.** *For $\mu \in \mathcal{P}(\mathcal{X})$, $\nu \in \mathcal{P}(\mathcal{Y})$, and $\kappa \in \mathcal{K}(\mathcal{X},\mathcal{Y})$, we have*

$$\mathcal{E}_p(\mathrm{proj}_{\mathrm{supp}(\nu)} \circ \kappa;\mu,\nu) \le 4\mathcal{E}_p(\kappa;\mu,\nu).$$

*Proof.* Write $f = \mathrm{proj}_{\mathrm{supp}(\nu)}$ and $\varepsilon = \mathcal{E}_p(\kappa;\mu,\nu)$. Fix a coupling $X, Y, Z$ such that $(X, Z) \sim (\mathrm{Id}, \kappa)_\sharp\mu$, $Y \sim \nu$, and $\mathbb{E}[\|Z - Y\|^p] = \mathsf{W}_p(\kappa_\sharp\mu,\nu)^p \le \varepsilon^p$. Taking $Z' = f(Z)$, we then bound

$$\begin{aligned}
\mathbb{E}[\|X - Z'\|^p]^{1/p} &\le \mathbb{E}[\|X - Z\|^p]^{1/p} + \mathbb{E}[\|Z - Z'\|^p]^{1/p}\\
&= \mathbb{E}[\|X - Z\|^p]^{1/p} + \mathbb{E}[\|Z - f(Z)\|^p]^{1/p}\\
&\le \mathsf{W}_p(\mu,\nu) + \varepsilon + \mathbb{E}[\|Z - Y\|^p]^{1/p}\\
&\le \mathsf{W}_p(\mu,\nu) + 2\varepsilon.
\end{aligned}$$

Similarly, we have

$$\begin{aligned}\mathsf{W}_p(\kappa'_\sharp\mu,\nu) &\le \mathbb{E}[\|Z'-Y\|^p]^{1/p}\\ &\le \mathbb{E}[\|Z-Y\|^p]^{1/p}+\mathbb{E}[\|Z-Z'\|^p]^{1/p}\\ &\le \varepsilon+\varepsilon=2\varepsilon.\end{aligned}$$

Thus, the sum of these two errors is at most $4\varepsilon$, as desired. □

Now, fix target measure $\nu=\frac{1}{2}\delta_{-y}+\frac{1}{2}\delta_y$, where $y=(1,\dots,1)$, and take $c\in[0,1]$ to be tuned later. Then, for each $0\le t\le 1/2$, define measure $\mu_t=(1/2-t)\delta_{-cy}+(1/2+t)\delta_{+cy}$. Now, fix any kernel $\kappa\in\mathcal{K}(\mathcal{X},\{\pm y\})$, where the codomain restriction is without loss of generality due to Lemma 14. Note that its performance on each $\mu_t$ is determined by the two-point distributions $\kappa_\pm:=\kappa_{\pm cy}=(1-\alpha_\pm)\delta_{-cy}+\alpha_\pm\delta_{cy}$. In particular, for $0\le t<1/2$, we compute

$$\begin{aligned}\mathsf{W}_p(\mu_t,\nu)^p &= \big(\tfrac{1}{2}-t\big)(1-c)^p\|y\|^p+t(1+c)^p\|y\|^p+\tfrac{1}{2}(1-c)^p\|y\|^p\\ &= d^{\frac{p}{2}}[(1-t)(1-c)^p+t(1+c)^p],\\ \mathsf{W}_p(\kappa_\sharp\mu_t,\nu)^p &= \mathsf{W}_p\big(\big(\tfrac{1}{2}-t\big)\kappa_-+\big(\tfrac{1}{2}+t\big)\kappa_+,\nu\big)\\ &= \|y-(-y)\|^p\cdot\mathsf{W}_p\big(\big(\tfrac{1}{2}-t\big)\operatorname{Ber}(\alpha_-)+\big(\tfrac{1}{2}+t\big)\operatorname{Ber}(\alpha_+),\operatorname{Ber}\big(\tfrac{1}{2}\big)\big)^p\\ &= (2d)^{\frac{p}{2}}\,\mathsf{W}_p\big(\operatorname{Ber}\big(\big(\tfrac{1}{2}-t\big)\alpha_-+\big(\tfrac{1}{2}+t\big)\alpha_+\big),\operatorname{Ber}\big(\tfrac{1}{2}\big)\big)^p\\ &= (2d)^{\frac{p}{2}}\,\big|\big(\tfrac{1}{2}-t\big)\alpha_-+\big(\tfrac{1}{2}+t\big)\alpha_+-\tfrac{1}{2}\big|,\\ \mathsf{W}_p(\delta_{cy},\kappa_+)^p &= \alpha_+(1-c)^p\|y\|^p+(1-\alpha_+)(1+c)^p\|y\|^p\\ &= d^{\frac{p}{2}}(\alpha_+(1-c)^p+(1-\alpha_+)(1+c)^p),\\ \mathsf{W}_p(\delta_{-cy},\kappa_-)^p &= \alpha_-(1+c)^p\|y\|^p+(1-\alpha_-)(1-c)^p\|y\|^p\\ &= d^{\frac{p}{2}}(\alpha_-(1+c)^p+(1-\alpha_+)(1-c)^p),\\ \iint\|y-x\|^p\mathrm{d}\kappa_x(y)\mathrm{d}\mu_t(x) &= \big(\tfrac{1}{2}-t\big)\mathsf{W}_p(\delta_{-cy},\kappa_-)^p+\big(\tfrac{1}{2}+t\big)\mathsf{W}_p(\delta_{cy},\kappa_+)^p\\ &= d^{\frac{p}{2}}\big[\big(\tfrac{1}{2}-t\big)(\alpha_-(1+c)^p+(1-\alpha_+)(1-c)^p)\\ &\qquad+\big(\tfrac{1}{2}+t\big)(\alpha_+(1-c)^p+(1-\alpha_+)(1+c)^p)\big].\end{aligned}$$

Writing $\Delta=\alpha_--\alpha_+$, we next bound

$$\begin{aligned}&\mathsf{W}_p(\kappa_\sharp\mu_t,\nu)+\mathsf{W}_p(\kappa_\sharp\mu_0,\nu)\\ &=\sqrt{2d}\,\big|\big(\tfrac{1}{2}-t\big)\alpha_-+\big(\tfrac{1}{2}+t\big)\alpha_+-\tfrac{1}{2}\big|^{\frac{1}{p}}+\sqrt{2d}\,\big|\tfrac{1}{2}\alpha_-+\tfrac{1}{2}\alpha_+-\tfrac{1}{2}\big|^{\frac{1}{p}}\\ &\ge\sqrt{2d}\,t^{\frac{1}{p}}|\alpha_+-\alpha_-|^{\frac{1}{p}} \qquad\qquad \text{(subadditivity of } a\mapsto a^{1/p}\text{)}\\ &=\sqrt{2d}\,t^{\frac{1}{p}}|\Delta|^{\frac{1}{p}}\end{aligned}$$

and we simplify

$$\begin{aligned}\iint\|y-x\|^p\mathrm{d}\kappa_x(y)\mathrm{d}\mu_0(x) &= d^{\frac{p}{2}}\left(\frac{1+\alpha_--\alpha_+}{2}(1+c)^p+\frac{1-\alpha_-+\alpha_+}{2}(1-c)^p\right)\\ &= d^{\frac{p}{2}}\left(\frac{1+\Delta}{2}(1+c)^p+\frac{1-\Delta}{2}(1-c)^p\right)\\ \mathsf{W}_p(\mu_0,\nu) &= \sqrt{d}\,(1-c).\end{aligned}$$

Thus, we further bound

$$\begin{aligned}&\left[\left(\iint\|y-x\|^p\mathrm{d}\kappa_x(y)\mathrm{d}\mu_0(x)\right)^{\frac{1}{p}}-\mathsf{W}_p(\mu_0,\nu)\right]_+\\ &=\sqrt{d}\left[\left(\frac{1+\Delta}{2}(1+c)^p+\frac{1-\Delta}{2}(1-c)^p\right)^{\frac{1}{p}}-1+c\right]_+\\ &=\sqrt{d}\left[\left(\frac{1+\Delta}{2}(1+c)^p+\frac{1-\Delta}{2}(1-c)^p\right)^{\frac{1}{p}}-1+c\right].\end{aligned}$$

**Algorithm 1:** Randomized Rounding for Efficient OT Kernel Estimation

**Input:** $n$ corrupted source points $S \subseteq \mathbb{R}^d$ and target points $T \subseteq [0,1]^d$, budgets $\rho \geq 0$, $\varepsilon \in [0,1]$

1: $m \leftarrow n^2$, $\tau \leftarrow n^{-1/(d+2)}$, $\sigma \leftarrow 3^{d/(2+d)}(nd)^{-1/(d+2)} + \rho^{1/2}d^{-1/4}$

2: $S' \leftarrow \{\mathrm{proj}_S(X_i' + Z_i)\}_{i=1}^m$, where each $X_i' \sim S$ and $Z_i \sim \mathcal{N}_\sigma$ are sampled independently

3: Compute kernel $\bar{\kappa} \in \mathcal{K}(S', T)$ s.t. $\bar{\kappa}_\sharp \,\mathrm{Unif}(S') = \mathrm{Unif}(T)$

$$\frac{1}{m}\sum_{x \in S'} \int \|x - y\| \mathrm{d}\bar{\kappa}(y|x) \leq \mathsf{W}_1(S', T) + \tau$$

4: Return $\hat{\kappa} \in \mathcal{K}(\mathbb{R}^d, T)$ defined by $\hat{\kappa} = \bar{\kappa} \circ \mathrm{proj}_S \circ N^\sigma$

Combining, this gives

$$\begin{aligned}
&\mathcal{E}_p(\kappa; \mu_t, \nu) + \mathcal{E}_p(\kappa; \mu_0, \nu)\\
&\geq \left[\sqrt{d}\left(\frac{1+\Delta}{2}(1+c)^p + \frac{1-\Delta}{2}(1-c)^p\right)^{\frac{1}{p}} - 1 + c + \sqrt{2d}\, t^{\frac{1}{p}} |\Delta|^{\frac{1}{p}}\right]\\
&\geq \left[\sqrt{d}\left(\frac{1+\Delta}{2}(1+c) + \frac{1-\Delta}{2}(1-c)\right) - 1 + c + \sqrt{2d}\, t^{\frac{1}{p}} |\Delta|^{\frac{1}{p}}\right]\\
&= \left[\sqrt{d}c(\Delta + 1) + \sqrt{2d}\, t^{\frac{1}{p}} |\Delta|^{1/p}\right]\\
&\geq \sqrt{d}\left[c(1 - |\Delta|) + t^{1/p}|\Delta|\right]\\
&\geq \sqrt{d} \min\{c/2, t^{1/p}/2\}
\end{aligned}$$

Now, supposing that $\rho < \sqrt{d}$, we can safely take $c = t^{1/p} = \rho^{1/2}d^{-1/4}/2$ while ensuring that $c \in [0,1]$ and $t \in [0, 1/2]$, which were the only constraints on our construction. Otherwise, we take $c = t^{1/p} = 1/2$. In either case, we have $\mathsf{W}_p(\mu_0, \mu_t) = t^{1/p} \cdot 2c\sqrt{d} = (\rho \wedge \sqrt{d})/2 \leq \rho$. Thus, the observation $\tilde{\mu} = \mu_0$ is compatible with both $\mu = \mu_0$ and $\mu_t$ under our corruption model. This gives the desired minimax lower bound of $\Omega(\sqrt{d}c \wedge t^{1/p}) = \Omega(d^{1/4}\rho^{1/2} \wedge \sqrt{d})$.

### E.3 Efficient Computation

We now introduce Algorithm 1 to achieve efficient computation, focusing on $p = 1$ where we match the rate of Theorem 4. Here, we identify finite sets with their uniform distributions when convenient.

**Theorem 5** (Efficient implementation)**.** *Under the setting of Section 5 with $p = 1$, the kernel $\hat{\kappa}$ returned by Algorithm 1 matches the risk bound of Theorem 4. Using an entropic OT solver for Step 3, Algorithm 1 runs in time $O((C_\infty + d)n^{2+o_d(1)})$, where $C_\infty = \max_{i,j} \|\tilde{X}_i - \tilde{Y}_j\|$. Moreover, $\hat{\kappa}$ can be evaluated (i.e., given $x \in \mathcal{X}$ we can sample $Y \sim \hat{\kappa}_x$) in time $O(nd)$.*

The proof below employs a similar analysis to that of Theorem 4, with multiple applications of Lemma 4 to account for various sampling errors along with TV contamination. We restrict to $p = 1$ due to the worsened scaling of Lemma 4 for $p > 1$.

*Proof.* Set $\alpha = N^\sigma_\sharp \tilde{\mu}_n$, $\beta = \mathrm{proj}_S \alpha$, and $\beta_m = \mathrm{Unif}(S')$. By construction, $S'$ is sampled i.i.d. from $\beta$, so Fact 5 gives that $\mathbb{E}[\|\beta - \beta_m\|_{\mathrm{TV}}] = \mathbb{E}[\mathbb{E}[\|\beta - \beta_m\|_{\mathrm{TV}} | S']] \lesssim \sqrt{n/m}$. Moreover,

$$\begin{aligned}
\int \|\mathrm{proj}_S(x) - x\| \mathrm{d}\alpha(x) &= \frac{1}{n}\sum_{x \in S}\int \|\mathrm{proj}_S(x+z) - x + z\| \mathrm{d}N^\sigma(z)\\
&\leq \frac{1}{n}\sum_{x \in S}\int \|x - x + z\| \mathrm{d}N^\sigma(z) && (x \in S)\\
&= \int \|z\| \mathrm{d}N^\sigma(z)\\
&\lesssim \sqrt{d}\,\sigma.
\end{aligned}$$

Now, we restate our guarantee for $\bar{\kappa}$; namely, we have:

$$\iint \|x-y\| \mathrm{d}\bar{\kappa}(y|x)\mathrm{d}\beta_m(x) \leq \mathsf{W}_1(\beta_m, \tilde{\nu}_n) + \tau.$$

Thus, by Lemma 6, we have

$$\mathcal{E}_1(\bar{\kappa}; \beta, \tilde{\nu}_n) \lesssim \tau + \sqrt{d}\,\|\beta - \beta_m\|_{\mathrm{TV}},$$

and, applying Lemma 5, we obtain

$$\mathcal{E}_1(\bar{\kappa} \circ \mathrm{proj}_S; \alpha, \tilde{\nu}_n) \lesssim \tau + \sqrt{d}\,\|\beta - \beta_m\|_{\mathrm{TV}} + \sqrt{d}\,\sigma.$$

Now, write $\mu'_n \in \mathcal{P}(\mathbb{R}^d)$ for an intermediate measure such that $\|\mu'_n - \tilde{\mu}_n\|_{\mathrm{TV}} \leq \varepsilon$ and $\mathsf{W}_p(\mu'_n, \hat{\mu}_n) \leq \rho$. Noting that $\alpha = N^\sigma_\sharp \tilde{\mu}_n$, we have by Fact 3 that $\|\alpha - N^\sigma_\sharp \mu'_n\|_{\mathrm{TV}} \leq \varepsilon$. Thus, Lemma 4 gives

$$\mathcal{E}_1(\bar{\kappa} \circ \mathrm{proj}_S; N^\sigma_\sharp \mu'_n, \tilde{\nu}_n) \lesssim \sqrt{d}\varepsilon + \tau + \sqrt{d}\,\|\beta - \beta_m\|_{\mathrm{TV}} + \sqrt{d}\,\sigma.$$

Applying Lemma 5 once more, we obtain

$$\mathcal{E}_1(\bar{\kappa} \circ \mathrm{proj}_S \circ N^{\sigma/2}; N^{\sigma/2}_\sharp \mu'_n, \tilde{\nu}_n) \lesssim \sqrt{d}\varepsilon + \tau + \sqrt{d}\,\|\beta - \beta_m\|_{\mathrm{TV}} + \sqrt{d}\,\sigma.$$

By Lemma 13, the fact that this latest kernel begins with the convolution $N^{\sigma/2}$ ensures that it is $O(\sqrt{d}\sigma^{-1})$-Lipschitz w.r.t. $\mathsf{W}_1$. Moreover, by Fact 4, we have $\mathsf{W}_1(N^{\sigma/2}_\sharp \mu'_n, N^{\sigma/2}_\sharp \hat{\mu}_n) \leq \mathsf{W}_1(\mu'_n, \hat{\mu}_n) \leq \rho$. Thus, Lemma 3 gives

$$\mathcal{E}_1(\bar{\kappa} \circ \mathrm{proj}_S \circ N^{\sigma/2}; N^{\sigma/2}_\sharp \hat{\mu}_n, \tilde{\nu}_n) \lesssim \sqrt{d}\varepsilon + \tau + \sqrt{d}\,\|\beta - \beta_m\|_{\mathrm{TV}} + \sqrt{d}\,\sigma + \rho + \frac{\sqrt{d}\,\rho}{\sigma}.$$

Next, we apply Lemma 4 and Lemma 5 to bound

$$\begin{aligned}
&\mathcal{E}_1(\bar{\kappa} \circ \mathrm{proj}_S \circ N^{\sigma}; \mu, \tilde{\nu}_n)\\
&\lesssim \mathcal{E}_1(\bar{\kappa} \circ \mathrm{proj}_S \circ N^{\sigma/2}; N^{\sigma/2}_\sharp \mu, \tilde{\nu}_n) + \sqrt{d}\,\sigma\\
&\lesssim \mathcal{E}_1(\bar{\kappa} \circ \mathrm{proj}_S \circ N^{\sigma/2}; N^{\sigma/2}_\sharp \hat{\mu}_n, \tilde{\nu}_n) + \sqrt{d}\,\sigma + \sqrt{d}\,\|N^{\sigma/2}(\mu - \hat{\mu}_n)\|_{\mathrm{TV}}\\
&\lesssim \sqrt{d}\,\varepsilon + \tau + \sqrt{d}\,\sigma + \rho + \frac{\sqrt{d}\,\rho}{\sigma} + \sqrt{d}\,\|\beta - \beta_m\|_{\mathrm{TV}} + \sqrt{d}\,\|N^{\sigma/2}(\mu - \hat{\mu}_n)\|_{\mathrm{TV}}.
\end{aligned}$$

Finally, we correct the target measure, using Lemma 2 to bound

$$\begin{aligned}
&\mathcal{E}_1(\bar{\kappa} \circ \mathrm{proj}_S \circ N^{\sigma}; \mu, \nu) \leq \mathcal{E}_1(\bar{\kappa} \circ \mathrm{proj}_S \circ N^{\sigma}; \mu, \tilde{\nu}_n) + 2\mathsf{W}_p(\tilde{\nu}_n, \nu)\\
&\lesssim \sqrt{d}\,\varepsilon + \tau + \sqrt{d}\,\sigma + \rho + \frac{\sqrt{d}\,\rho}{\sigma} + \sqrt{d}\,\|\beta - \beta_m\|_{\mathrm{TV}} + \sqrt{d}\,\|N^{\sigma/2}(\mu - \hat{\mu}_n)\|_{\mathrm{TV}} + \mathsf{W}_p(\hat{\nu}_n, \nu)
\end{aligned}$$

Taking expectations, using our early bound on the first TV distance, and applying Lemma 12 for the second TV distance, and applying Lemma 1 for the Wasserstein distance, we obtain

$$\begin{aligned}
&\mathbb{E}[\mathcal{E}_1(\bar{\kappa} \circ \mathrm{proj}_S \circ N^{\sigma}; \mu, \nu)]\\
&\lesssim \sqrt{d}\,\varepsilon + \tau + \sqrt{d}\,\sigma + \rho + \frac{\sqrt{d}\,\rho}{\sigma} + \sqrt{\frac{dn}{m}} + \sqrt{d\,3^d(1 \vee \sigma^{-d})/n} + c_{p,d} n^{-\frac{1}{p \vee 2d}} \log^2 n
\end{aligned}$$

Our choice of $\sigma$, $m$, and $\tau$ ensure that the desired risk bound holds.

Computational complexity is dominated by the OT computation at Step 3. The source and target distributions are both supported on $n$ points, and we require accuracy $\tau = n^{-1/(d+2)}$. Computing the relevant cost matrix requires time $O(n^2 d)$. Using a state of the art OT solver based on entropic OT (e.g., Luo et al., 2023) gives a running time of $O(C_\infty n^2/\tau) = O(C_\infty n^{2+1/(d+2)})$, where $C_\infty$ is the largest distance between a point in $S$ and a point in $T$. Combining these two gives the first bound. Evaluation complexity is dominated by the projection step, which can be computed in a brute-force manner using $O(nd)$ time. □

# F Extended Experiment Details

We note that all code is provided in the supplement and will be made publicly available on GitHub for the final version. Experiments were completed in less than 2 hours on the specified machine.