# 1 Smoothed Barycenter Stuff

Let $\lambda$ denote the Lebesgue measure on $\mathbb{R}^d$.

**Proposition 1.** *$W_p(\mu,\nu)^p$ is convex in $(\mu,\nu)$ for $p \in [1,\infty)$.*

*Proof.* Fix $\mu_1,\nu_1,\mu_2,\nu_2 \in \mathcal{P}_p(\mathbb{R}^d)$ and $t \in [0,1]$. Let $\pi_1 \in \Pi(\mu_1,\nu_1)$ be a transport plan achieving $W_p(\mu_1,\nu_1)$, and take $\pi_2 \in \Pi(\mu_2,\nu_2)$ to be a transport achieving $W_p(\mu_2,\nu_2)$. Next, define $\pi = t\pi_1 + (1-t)\pi_2$. Checking marginals, we find $\pi \in \Pi(t\mu_1 + (1-t)\mu_2, t\nu_1 + (1-t)\nu_2)$, so

$$\begin{aligned} W_p(t\mu_1 + (1-t)\mu_2, t\nu_1 + (1-t)\nu_2)^p &\le \mathbb{E}_{X,Y\sim\pi}\|X-Y\|^p \\ &= t\mathbb{E}_{X,Y\sim\pi_1}\|X-Y\|^p + (1-t)\mathbb{E}_{X,Y\sim\pi_2}\|X-Y\|^p \\ &= tW_p(\mu_1,\nu_1)^p + (1-t)W_p(\mu_2,\nu_2)^p. \end{aligned}$$

□

**Proposition 2.** *Let $\mu,\nu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\mu \ll \lambda$, and let $X,Y$ be an optimal coupling of the pair w.r.t. $W_2$. Then,*

1. *a regular conditional distribution of $Y$ given $X = x$ is supported on a single point, $\mu$-a.s.*
2. *there exists a measurable mapping $T : D \subset \mathbb{R}^d \to \mathbb{R}^d$ with $\mu(D) = 1$ such that $Y = T(X)$ a.s.*
3. *$T$ is the gradient of a convex function (implies next)*
4. *$\langle T(x) - T(x'), x - x'\rangle \ge 0$ (can be strengthened to cyclically monotone);*
5. *$W_2(\mu,\nu)^2 = \mathbb{E}[\|X - T(X)\|^2]$.*

TODO: Does 2 not imply 1? 2,5 hold for $p > 1$.

**Proposition 3.** *$W_p(\mu,\cdot)^p$ is strictly convex if $\mu \ll \lambda$ for $p \in (0,\infty)$.*

*Proof.* Fix $\mu,\nu_1,\nu_2 \in \mathcal{P}_p(\mathbb{R}^d)$ with $\mu \ll \lambda$ and let $t \in [0,1]$. Let $T_1,T_2$ denote the optimal transport maps from $\mu$ to $\nu_1$ and $\nu_2$ with respect to $W_p$ (existence guaranteed since $p > 1$ and $\mu \ll \lambda$). Let $\nu_t := t\nu_1 + (1-t)\nu_2$ and define on its support the random function

$$T(x) = \begin{cases} T_1(x) & \text{with probability } t \\ T_2(x) & \text{with probability } 1-t. \end{cases}$$

If $X$ is distributed according to $\mu$, then

$$\begin{aligned} \Pr[T(X) \in A] &= t \cdot T_{1\#}\mu(A) + (1-t)\cdot T_{2\#}\mu(A) \\ &= t\nu_1(A) + (1-t)\nu_2(A) = \nu_t(A). \end{aligned}$$

Hence,

$$\begin{aligned} W_p(\mu, \nu_t)^p &\leq \mathbb{E}\|X - T(X)\|^p \\ &= t\mathbb{E}\|X - T_1(X)\|^p + (1-t)\mathbb{E}\|X - T_2(X)\|^p \\ &= tW_p(\nu_1)^p + (1-t)W_p(\nu_2)^p, \end{aligned}$$

with strict inequality unless $T$ is an optimal transport plan between $\mu$ and $\nu_t$. However, this would force $T$ to be deterministic (almost everywhere), requiring that $T_1 = T_2$ (almost everywhere). In this case, $\nu_1 = T_{1\#}\mu = T_{2\#}\mu = \nu_2$. □

**Corollary 4.** *$W_p^\sigma(\mu, \cdot)^p$ is strictly convex for $p \in (0, \infty)$.*

*Proof.* Convolution with $\mathcal{N}_\sigma$ is linear and non-singular. (Just requires noise model with non-vanishing characteristic function.) □

**Proposition 5.**