SUPPLEMENTARY MATERIAL

Proof of Proposition 2.1

If we denote $\gamma := (T_\mu, T_\nu)_\# \rho$, then $\gamma \in \Pi(\mu, \nu)$. The change of variable formula gives

$$W_2^2(\mu,\nu) \leq \int_{\mathcal{Y}\times\mathcal{Y}} ||y - y'||_2^2 d\gamma(y, y')$$
$$= \int_{\mathcal{X}} ||T_\mu(x) - T_\nu(x)||_2^2 \rho(x) \mathrm{d}x = ||T_\mu - T_\nu||^2_{L^2(\rho)}.$$

The continuity of the map $\mu \mapsto T_\mu$ follows from e.g. Exercise 2.17 in (Villani, 2003). To prove (iii), we use the following lemma:

Lemma 5.1. *Let ρ be uniform on the unit disc $\mathcal{X} \subseteq \mathbb{R}^2$. Then, there is a curve $\theta \in [0, 2\pi] \to \mu_\theta \in \mathcal{P}(\mathcal{X})$ and $C > 0$ such that $\|T_{\mu_\theta} - T_{\mu_0}\|_{L^2(\rho)} \geq C W_2(\mu_\theta, \mu_0)^{1/2}$.*

Proof. Given $\theta \in \mathbb{R}$, we denote $x_\theta = (\cos\theta, \sin(\theta))$ and $\mu_\theta = \frac{1}{2}(\delta_{x_\theta} + \delta_{-x_\theta})$. Then, the optimal transport map between ρ and μ_θ is given by

$$T_{\mu_\theta}(x) = \begin{cases} x_\theta & \text{if } \langle x|x_\theta\rangle \geq 0 \\ -x_\theta & \text{if not.} \end{cases}$$

One can easily check that for θ one has $W_2(\mu_0, \mu_\theta) \leq \frac{|\theta|}{\sqrt{\pi}}$. For $\theta > 0$ we set

$$D_\theta = \{x \in \mathbb{R}^2 \mid \langle x|x_0\rangle \geq 0 \text{ and } \langle x|x_\theta\rangle \leq 0\}.$$

Then, on D_θ, $T_{\mu_\theta} \equiv x_{-\theta}$ and $T_{\mu_0} \equiv x_0$, giving

$$\|T_{\mu_\theta} - T_{\mu_0}\|^2_{L^2(\rho)} \geq \int_{D_\theta} \|x_{-\theta} - x_0\|^2 \,\mathrm{d}x = |D_\theta| \, \|x_{-\theta} - x_0\|^2 .$$

Moreover, if $|\theta| \leq \frac{\pi}{2}$ one has $\|x_{-\theta} - x_0\|^2 \geq 2$, thus giving $\|T_{\mu_\theta} - T_{\mu_0}\|^2_{L^2(\rho)} \geq 2\,|D_\theta| \geq \frac{|\theta|}{\pi}$. □

Proof of Theorem 2.2

The proof of Theorem 2.2 is based on the following lemma.

Lemma 5.2. *Under the assumptions of Theorem 2.2,*

$$||T_\mu - T_\nu||^2_{L^2(\rho)} \leq 2K \int_{\mathcal{Y}} (\psi_\nu - \psi_\mu) \mathrm{d}(\mu - \nu).$$

Proof. From convex analysis, the map $T_\mu = \nabla\phi_\mu$ is K-Lipschitz if and only if $\psi_\mu = \phi_\mu^*$ is $\frac{1}{K}$-strongly convex. We denote $A = \int_{\mathcal{Y}} \psi_\nu \mathrm{d}(\mu - \nu)$ and $B = \int_{\mathcal{Y}} \psi_\mu \mathrm{d}(\nu - \mu)$.

We use that $(\nabla\phi_\mu)_\#\rho = \mu$ (resp. $(\nabla\phi_\nu)_\#\rho = \nu$) and $\nabla\phi_\mu = \nabla\psi_\mu^*$ (resp. $\nabla\phi_\nu = \nabla\psi_\nu^*$) by convexity of ϕ_μ (resp. ϕ_ν) to do the following change of variable:

$$A = \int_{\mathcal{X}} (\psi_\nu(\nabla\psi_\mu^*) - \psi_\nu(\nabla\psi_\nu^*)) \mathrm{d}\rho.$$

We now use the inequality $\psi_\nu(y) - \psi_\nu(z) \geq \langle y - z|v\rangle$, which holds for all v in the subdifferential $\partial\psi_\nu(z)$. The convex functions ψ_ν, ψ_μ are differentiable ρ-almost everywhere. Taking $z = \nabla\psi_\nu^*(x)$ and $y = \nabla\psi_\mu^*(x)$, and using $x \in \partial\psi_\nu(z)$, we obtain

$$A \geq \int_{\mathcal{X}} \langle \mathrm{id}, \nabla\psi_\mu^* - \nabla\psi_\nu^*\rangle \mathrm{d}\rho.$$

Using the strong convexity of ψ_μ, we get a similar lower bound on B, with an extra quadratic term

$$\begin{aligned} B &= \int_{\mathcal{X}} (\psi_\mu(\nabla\psi_\nu^*) - \psi_\mu(\nabla\psi_\mu^*))\mathrm{d}\rho \\ &\geq \int_{\mathcal{X}} \big(\langle \mathrm{id}, \nabla\psi_\nu^* - \nabla\psi_\mu^*\rangle + \frac{1}{2K}||\nabla\psi_\nu^* - \nabla\psi_\mu^*||_2^2\big)\mathrm{d}\rho. \end{aligned}$$

Summing up these inequalities we get:

$$\begin{aligned} \int_{\mathcal{Y}} (\psi_\nu - \psi_\mu) d(\mu - \nu) \geq & \frac{1}{2K}\int_{\mathcal{X}} ||\nabla\psi_\nu^* - \nabla\psi_\mu^*||_2^2 \mathrm{d}\rho \\ &= \frac{1}{2K}||T_\nu - T_\mu||_{L^2(\rho)}^2. \end{aligned}$$

□

Proof of Theorem 2.2. Formula (3) clearly shows that $\mathrm{Lip}(\psi_\mu) \leq M_{\mathcal{X}}$, where $\mathrm{Lip}(f)$ denotes the Lipschitz constant of f. Combining this with Lemma 5.2,

$$\begin{aligned} ||T_\mu - T_\nu||_{L^2(\rho)}^2 &\leq 2K \int_{\mathcal{Y}} (\psi_\nu - \psi_\mu)\mathrm{d}(\mu - \nu) \\ &\leq 2K \max_{\mathrm{Lip}(f)\leq 2M_{\mathcal{X}}} \int_{\mathcal{Y}} f\mathrm{d}(\mu-\nu) \\ &= 4KM_{\mathcal{X}} \max_{\mathrm{Lip}(f)\leq 1} \int_{\mathcal{Y}} f\mathrm{d}(\mu-\nu) \\ &= 4KM_{\mathcal{X}} W_1(\mu,\nu), \end{aligned}$$

where we used Kantorovich-Rubinstein's theorem to get the last equality. □

Proof of Corollary 2.4

We first state a simple lemma that links the uniform norm of a Lipschitz function to its $L^2(\rho)$ norm:

Lemma 5.3. *If f is L-Lipschitz on $\mathcal{X}$, then*

$$||f||_\infty \leq C||f||_{L^2(\mathcal{X})}^{\frac{2}{d+2}},$$

for some C depending on L, d and $\mathcal{X}$ only.

Proof. If $||f||_\infty = \eta$, then there exists $x_0 \in \mathcal{X}$ such that for all $x \in B(x_0, \frac{\eta}{2L}) \cap \mathcal{X}$ we have $|f(x)| \geq \frac{\eta}{2}$. This implies that

$$||f||_{L^2(\mathcal{X})}^2 \geq \int_{B(x_0,\frac{\eta}{2l})\cap\mathcal{X}} f(x)^2 \mathrm{d}y \geq \beta_d \left(\frac{\eta}{2L}\right)^d \eta^2 = \frac{\beta_d}{(2L)^d} \|f\|_\infty^{d+2},$$

where β_d is the volume of the d-dimensional unit ball. □

Proof of Corollary 2.4. Theorem 2.3 implies

$$||\nabla\psi_\mu - \nabla\psi_\nu||_{L^2(\mathcal{Y})}^2 \leq C\Big(\int_{\mathcal{Y}} (\psi_\nu - \psi_\mu)\mathrm{d}(\mu-\nu)\Big)^{\frac{1}{2^{d-1}}},$$

and as in Theorem 2.2, the quantity in the parenthesis can be upper bounded by $2M_{\mathcal{X}}W_1(\mu,\nu)$. Adding a constant to ψ_μ if necessary, we can assume that $\int_{\mathcal{Y}} \psi_\mu(y)\mathrm{d}y = \int_{\mathcal{Y}} \psi_\nu(y)\mathrm{d}y$. The Poincaré-Wirtinger inequality on $\mathcal{Y}$ then implies

$$||\psi_\mu - \psi_\nu||_{L^2(\mathcal{Y})}^2 \leq C' W_1(\mu,\nu)^{\frac{1}{2^{d-1}}},$$

for some C' depending only on ρ, $\mathcal{X}$ and $\mathcal{Y}$.

We reuse the fact that $\psi_\mu - \psi_\nu$ is Lipschitz with a constant at most $2M_{\mathcal{X}}$ to use Lemma 5.3:

$$||\psi_\mu - \psi_\nu||_\infty \leq C'' W_1(\mu,\nu)^{\frac{2}{2^{d-1}(d+2)}}.$$

Since $\phi_\mu = \psi_\mu^*$ and $\phi_\nu = \psi_\nu^*$, the definition of the Legendre transform yields

$$||\phi_\mu - \phi_\nu||_\infty \leq C'' W_1(\mu,\nu)^{\frac{2}{2^{d-1}(d+2)}}.$$

We conclude using Proposition 3.6 and the fact that ϕ_μ is diam($\mathcal{Y}$)-Lipschitz (as the Legendre transform of the function ψ_μ on $\mathcal{Y}$): there exists a constant C depending only on ρ, $\mathcal{X}$ and $\mathcal{Y}$ such that

$$||T_\mu - T_\nu||_{L^2(\rho)} \leq C W_1(\mu,\nu)^{\frac{1}{2^{(d-1)}(d+2)}}. \qquad \square$$

Proof of Lemma 3.2

For any $N > 0$, we consider a finite partition $\mathcal{Y} = \sqcup_{1\leq i\leq N} \mathcal{Y}_i^N$, we let $\varepsilon_N = \max_i \operatorname{diam}(\mathcal{Y}_i^N)$ and we assume that $\lim_{N\to+\infty} \varepsilon_N = 0$. Then, we define

$$\mu_N^k = \sum_{1\leq i\leq N} \left[\left(1-\frac{1}{N}\right)\mu^k(\mathcal{Y}_i^N) + \frac{1}{N^2}\right]\delta_{y_i^N},$$

where $y_i^N \in \mathcal{Y}_i^N$. Then, it is easy to check that the support of the measures μ_N^0 and μ_N^1 is the set $\{y_1^N,\dots,y_N^N\}$. Moreover,

$$\left\|\mu_N^1 - \mu_N^0\right\|_{\mathrm{TV}} \leq \left\|\mu^1 - \mu^0\right\|_{\mathrm{TV}}.$$

In addition, $W_1(\mu_N^k,\mu^k) \leq \varepsilon_N \xrightarrow{N\to+\infty} 0$. Combined with the triangle inequality, we deduce

$$\begin{aligned}|W_1(\mu_N^0,\mu_N^1) - W_1(\mu^0,\mu^1)| &= |W_1(\mu_N^0,\mu_N^1) - W_1(\mu_N^0,\mu^1) + W_1(\mu_N^0,\mu^1) - W_1(\mu^0,\mu^1)| \\ &\leq |W_1(\mu_N^0,\mu_N^1) - W_1(\mu_N^0,\mu^1)| + |W_1(\mu_N^0,\mu^1) - W_1(\mu^0,\mu^1)| \\ &\leq W_1(\mu_N^1,\mu^1) + W_1(\mu_N^0,\mu^0) \\ &\leq 2\varepsilon_N \xrightarrow{N\to+\infty} W_1(\mu^0,\mu^1).\end{aligned}$$

Using the stability of optimal transport maps (Proposition 2.1), we finally deduce that

$$\lim_{N\to+\infty} ||T_{\mu_N^1} - T_{\mu_N^0}||_{L^2(\rho)} = ||T_{\mu^1} - T_{\mu^0}||_{L^2(\rho)}.$$

Proof of Lemma 3.3

Let $x^0 \in V_i(\psi^0)$ and $x^1 \in V_i(\psi^1)$. Then, for all $j \in \{1,\dots,N\}$,

$$\begin{cases}\psi^0(y_j) \geq \psi^0(y_i) + \langle y_j - y_i | x^0\rangle \\ \psi^1(y_j) \geq \psi^1(y_i) + \langle y_j - y_i | x^1\rangle\end{cases}.$$

Taking the convex combination of these inequalities we get for all $j \in \{1,\dots,N\}$,

$$\psi^t(y_j) \geq \psi^t(y_i) + \langle y_j - y_i | (1-t)x^0 + tx^1\rangle.$$

This shows that $(1-t)x^0 + tx^1 \in V_i(\psi^t)$ (note that we use the convexity of $\mathcal{X}$ here). Thus,

$$(1-t)V_i(\psi^0) + tV_i(\psi^1) \subseteq V_i(\psi^t).$$

Taking the Lebesgue measure on both sides and applying Brunn-Minkowski's inequality we get

$$\begin{aligned}G_i(\psi^t)^{1/d} = \rho(V_i(\psi^t))^{1/d} &\geq \rho((1-t)V_i(\psi^0) + tV_i(\psi^1))^{1/d} \\ &\geq (1-t)\rho(V_i(\psi^0))^{1/d} + t\rho(V_i(\psi^1))^{1/d} \\ &\geq (1-t)G_i(\psi^0)^{1/d} + tG_i(\psi^1))^{1/d}.\end{aligned}$$

This inequality directly implies

$$G_i(\psi^t) \geq \min(G_i(\psi^0), G_i(\psi^1)),$$

$$\text{i.e.} \min(G_i(\psi^t), G_i(\psi^0)) \geq \min(G_i(\psi^0), G_i(\psi^1)).$$

Using the following equivalent formulation of the TV distance between probability measures we get (8):

$$\begin{aligned}\frac{1}{2}\left\|G(\psi^t) - G(\psi^0)\right\|_1 &= 1 - \sum_i \min(G_i(\psi^t), G_i(\psi^0)) \\ &\leq 1 - \sum_i \min(G_i(\psi^0), G_i(\psi^1)) = \frac{1}{2}\left\|G(\psi^t) - G(\psi^0)\right\|_1.\end{aligned}$$

To prove (9), we first remark that by (7),

$$G_i(\psi^t) \geq (1-t)^d G_i(\psi^0),$$
$$\text{i.e.} \min(G_i(\psi^t), G_i(\psi^0)) \geq (1-t)^d G_i(\psi^0).$$

We conclude using the same formula as above:

$$\begin{aligned}\frac{1}{2}\left\|G(\psi^t) - G(\psi^0)\right\|_1 &= 1 - \sum_i \min(G_i(\psi^t), G_i(\psi^0)) \\ &\leq 1 - \sum_i (1-t)^d G_i(\psi^0) = 1 - (1-t)^d.\end{aligned}$$

Proof of Proposition 3.4

This proof is a straightforward adaptation of Lemma 3.7 in (Eymard et al., 2000), but we include it for completeness. We consider the function u on $\mathcal{X}$ defined a.e. by $u|_{V_i(\psi)} = v_i$. Then,

$$\langle v^2 - \langle v|G(\psi)\rangle^2 | G(\psi)\rangle = \int_{\mathcal{X}} u^2 - \left(\int_{\mathcal{X}} u\right)^2 = \frac{1}{2}\int_{\mathcal{X}\times\mathcal{X}} (u(x) - u(y))^2 \mathrm{d}y\mathrm{d}x,$$

so it suffices to control the right hand side of this equality. Given (i, j) and $(x, y) \in \mathcal{X}^2$, we denote

$$\chi_{ij}(x, y) = \begin{cases} 1 & \text{if } V_i(\psi) \cap V_j(\psi) \cap [x, y] \neq \emptyset \text{ and } \langle y_j - y_i | y - x\rangle \geq 0 \\ 0 & \text{if not.}\end{cases}$$

Then, $u(y) - u(x) = \sum_{i\neq j}(v(y_j) - v(y_i))\chi_{ij}(x, y)$. Denoting $d_{ij} = \|y_j - y_i\|$, $c_{ij,z} = \left|\langle \frac{z}{\|z\|} | \frac{y_j - y_i}{\|y_j - y_i\|}\rangle\right|$ and applying Cauchy-Schwarz's inequality we get

$$\begin{aligned}(u(y) - u(x))^2 &= \left(\sum_{i\neq j}(v(y_j) - v(y_i))\chi_{ij}(x, y)\right)^2 \\ &\leq \sum_{i\neq j} \frac{(v(y_j) - v(y_i))^2}{d_{ij}c_{ij,y-x}}\chi_{ij}(x, y) \sum_{i\neq j} d_{ij}c_{ij,y-x}\chi_{ij}(x, y).\end{aligned}$$

In addition, when $\chi_{ij}(x, y) = 1$, we have $\langle y - x | y_j - y_i\rangle \geq 0$ so that

$$d_{ij}c_{ij,y-x} = \|y_j - y_i\| \langle \frac{y - x}{\|y - x\|} | \frac{y_j - y_i}{\|y_j - y_i\|}\rangle \geq 0,$$

and

$$\sum_{i\neq j} d_{ij}c_{ij,y-x}\chi_{ij}(x, y) = \sum_{i\neq j}\langle \frac{y - x}{\|y - x\|} | y_j - y_i\rangle \chi_{ij}(x, y) \leq \operatorname{diam}(\mathcal{Y}).$$

Therefore,

$$\begin{aligned}\int_{\mathcal{X}\times\mathcal{X}} (u(y) - u(x))^2 \mathrm{d}x\mathrm{d}y &\leq \operatorname{diam}(\mathcal{Y}) \int_{\mathcal{X}\times\mathcal{X}} \sum_{i\neq j} \frac{(v(y_j) - v(y_i))^2}{d_{ij}c_{ij,y-x}} \chi_{ij}(x, y)\mathrm{d}x\mathrm{d}y \\ &= \operatorname{diam}(\mathcal{Y}) \int_{B(0,\operatorname{diam}(\mathcal{X}))} \sum_{i\neq j} \frac{(v(y_j) - v(y_i))^2}{d_{ij}c_{ij,z}} \left(\int_{\mathcal{X}} \chi_{ij}(x, x + z)\mathrm{d}x\right) \mathrm{d}z.\end{aligned}$$

Moreover, denoting $m_{ij} = \mathcal{H}^{d-1}(V_i(\psi) \cap V_j(\psi))$ we get

$$\int_{\mathcal{X}} \chi_{ij}(x, x+z)\mathrm{d}x \leq m_{ij} \|z\| \, c_{ij,z},$$

thus giving

$$\int_{\mathcal{X}\times\mathcal{X}} (u(y) - u(x))^2 \mathrm{d}x\mathrm{d}y \leq C(d) \operatorname{diam}(\mathcal{Y}) \operatorname{diam}(\mathcal{X})^{d+1} \sum_{i \neq j} \frac{m_{ij}}{d_{ij}} (v(y_j) - v(y_i))^2.$$

Define $H_{ij} = \frac{m_{ij}}{d_{ij}}$, $H_{ii} = -\sum_{j\neq i} H_{ij}$. Then, $\mathrm{D}G(\psi) = H$, and

$$\begin{aligned}
\langle \mathrm{D}G(\psi) v | v \rangle &= \sum_{i,j} H_{ij} v_i v_j \\
&= \sum_i \left(H_{ii} v_i v_i + \sum_{j \neq i} H_{ij} v_i v_j \right) \\
&= \sum_i \sum_{j \neq i} H_{ij} v_i (v_j - v_i) \\
&= \sum_{j \neq i} H_{ij} v_i (v_j - v_i) := A.
\end{aligned}$$

And

$$\sum_{i \neq j} H_{ij} (v(y_j) - v(y_i))^2 = \sum_{i \neq j} H_{ij} v_j (v_j - v_i) - \sum_{i \neq j} H_{ij} v_i (v_j - v_i) = -2A.$$

We finally obtain

$$\iint (u(y) - u(x))^2 \mathrm{d}x\mathrm{d}y \leq -C_{d,\mathcal{X},\mathcal{Y}} \langle \mathrm{DG}(\psi) v | v \rangle.$$