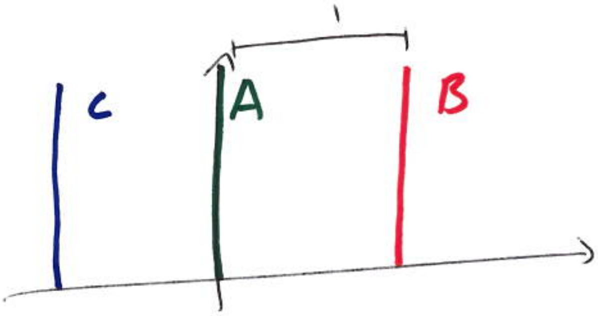
Exercise Sheet 2 Solutions

Lent 2018-2019

Intro to OT

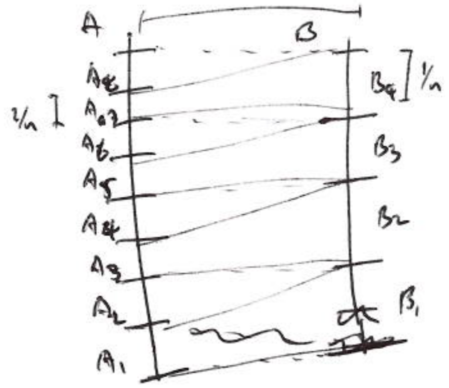
2.1



(a) Each unit of mass from A has to move to either B or C, the distance of which is at least 1. Hence $\inf_{T : T_{\#}\mu = \nu} \|M(T)\| \geq 1$.

In particular if T is any transport map then $x \in A$, $T(x) \in B \cup C$ implies $c(x, T(x)) = |x - T(x)|^2 \geq 1$.

(b)



Consider transport from A_{2i-1} to B_i. Mass is moved at most $\sqrt{1+(1/n)^2}$. Hence the cost is

$$\int_{\frac{2i-2}{2n}}^{\frac{2i-1}{2n}} |x - T_n(x)|^2 \, d\mu(x) \leq \int_{\frac{2i-2}{2n}}^{\frac{2i-1}{2n}} 1 + \left(\tfrac{1}{n}\right)^2 dx$$

A_{2i-1}

$$= \left(1 + \left(\tfrac{1}{n}\right)^2\right) \frac{1}{2n}$$

Similarly, the cost of transport from A_{2i} to C_i is bounded by $\left(1 + \left|\tfrac{1}{n}\right|^2\right) \frac{1}{2n}$. Hence the total cost is bounded by

$$\int_A |x - T_n(x)|^2 \, d\mu(x) = \sum_{i=1}^{n} \left(\int_{A_{2i-1}} |x - T_n(x)|^2 \, d\mu(x) + \int_{A_{2i}} |x - T_n(x)|^2 \, d\mu(x) \right) \leq 1 + \left(\tfrac{1}{n}\right)^2 . \quad (1)$$

(c) The example [illegible] sequence of homeomorphic ~~[illegible]~~ maps shows that $\inf_{T: T_\# \mu = \nu} \mathbb{M}(T) \leq 1$. By part (a) we have $\inf_{T: T_\# \mu = \nu} \mathbb{M}(T) = 1$.

But this means that mass can only move horizontally. This means

mass must be split. In particular, if $T^\dagger$ is optimal, the chosen ~~the $(T^\dagger)^{-1}(U) \subset A$.~~

~~By an open set $U \subset C$, since $T^\dagger$ is measurable~~

~~an open set $U \subset [0,1]$. Then $(T^\dagger)^{-1}(\{-1\} \times U) = \{0\} \times U$~~ and

~~$(T^\dagger)^{-1}(\{1\} \times U) = \{0\} \times U$. Moreover $(T^\dagger)^{-1}(\{1\} \times U)$ This is a contradiction.~~

$T^\dagger(\{0\} \times U) = \{-1\} \times U$. Then $(T^\dagger)^{-1}(\{+1\} \times U)$

$U \subset [0,1]$ s.t. $\mu(\{0\} \times U) > 0$ and $\mu((T^\dagger)^{-1}(\{+1\} \times U)) = 0$, but $\nu(\{-1\} \times U) = \nu(\{+1\} \times U)$

$\cap(\{0\} \times U) = \emptyset$, in particular $\mu((T^\dagger)^{-1}(\{+1\}\times U)) = \nu(\{+1\} \times U)$ so $\nu(\{\pm 1\} \times U) = 0$

$\mu(\{0\} \times U) \Rightarrow \mu(\{0\} \times U) = 0$.

2.2

Assume π' is not optimal. Then $\exists\, \pi'' \in \Pi(\mu', \nu')$ s.t.

$$\int_{X \times Y} c(x,y)\, d\pi''(x,y) < \int_{X \times Y} c(x,y)\, d\pi'(x,y).$$

Define $\hat{\pi} := (\pi - \tilde{\pi}) + \tilde{Z}\pi''$ where $\tilde{Z} = \tilde{\pi}(X \times Y)$. Since $\tilde{\pi} \leq \pi$ then $\hat{\pi}$ is non-negative. And since $\hat{\pi} = \pi - \tilde{Z}\pi' + \tilde{Z}\pi'' = \pi + \tilde{Z}(\pi'' - \pi')$

we have that $\hat{\pi} \in \Pi(\mu, \nu)$. Now

$$\mathbb{K}(\hat{\pi}) = \int_{X \times Y} c(x,y)\, d\hat{\pi}(x,y) = \int_{X \times Y} c(x,y)\, d\pi(x,y) + \tilde{Z}\left(\int_{X \times Y} c(x,y)\, d\pi''(x,y) - \int_{X \times Y} c(x,y)\, d\pi'(x,y) \right)$$

$$< \int_{X \times Y} c(x,y)\, d\pi(x,y) = \mathbb{K}(\pi) = \min_{\Pi(\mu,\nu)} \mathbb{K}$$

This contradicts the fact that π is optimal. It follows that π' is optimal.

2.3

By definition $F^{-1}(F(x)) = \inf\{y \in \mathbb{R} : F(y) \geq F(x)\}$

~~Clearly $x \notin \{y \in \mathbb{R} : F(y) > F(x)\}$ and so by monotonicity of F we have $z \notin \{y \in \mathbb{R} : F(y) > F(x)\}$ $\forall z \geq x$~~

~~$\geq \inf\{y \in \mathbb{R}$~~

By monotonicity of F we have that $F(y) \geq F(x) \Rightarrow y \geq x$.

Hence $\inf\{y \in \mathbb{R} : F(y) \geq F(x)\} \geq x$.

Similarly, $F(F^{-1}(t)) = F(\inf\{y \in \mathbb{R} : F(y) \geq t\})$

consider the sequence $y_n \to \inf\{y \in \mathbb{R} : F(y) \geq t\}$, where $y_n \in \{y \in \mathbb{R} : F(y) \geq t\}$. Hence for each n we have $F(y_n) \geq t$, and by right continuity of F we

Now $\lim_{n\to\infty} F(y_n) \geq t$

have $\lim_{n\to\infty} F(y_n) = F(\lim_{n\to\infty} y_n) = F(\inf\{y \in \mathbb{R} : F(y) \geq t\})$

so $t \leq F(F^{-1}(t))$

If F is invertible then let $F^{\dagger}$ be the inverse of F. It follows that F is strictly increasing. We have

$F^{-1}(t) = \inf\{y \in \mathbb{R} : F(y) \geq t\} = \inf\{y \in \mathbb{R} : y \geq F^{\dagger}(t)\}$ (also by monotonicity)

$= F^{\dagger}(t)$ so $F^{-1} = F^{\dagger}$. Hence $F^{-1}(F(x)) = x$ and

$F(F^{-1}(t)) = t$.

Finally, if $F^{-1}(t) \le x$ then $F(F^{-1}(t)) \le F(x)$ so by the 1^{st} part $t \le F(F^{-1}(t)) \le F(x)$.

And if $F(x) \ge t$ then

$$F^{-1}(t) = \inf \{y \in \mathbb{R} : F(y) \ge t\} \le \inf\{y \in \mathbb{R} : F(y) \ge F(x)\}$$

$$\le x.$$

2.4

By Cor 3.2 $T^{\dagger} = G^{-1} \circ F$ is the optimal transport plan map

when $F(t) = \int_{-\infty}^{t} \chi_{[0,1]}(x)\,dx = \cancel{t\chi_{[0,1]}} \begin{cases} t & \text{if } t \in [0,1] \\ 0 & \text{if } t < 0 \\ 1 & \text{if } t > 1 \end{cases}$

$$G(t) = 2\int_{-\infty}^{t} \chi_{[0,1]}(y)\, y\, dy = \begin{cases} 0 & \text{if } t < 0 \\ 2\int_0^t y\,dy & \text{if } t \in [0,1] \\ 1 & \text{if } t > 1 \end{cases}$$

$$= \begin{cases} 0 & \text{if } t < 0 \\ t^2 & \text{if } t \in [0,1] \\ 1 & \text{if } t > 1 \end{cases}$$

So $G^{-1}(t) = \sqrt{t}$ for $t \in [0,1]$ hence $T^{\dagger}(x) = \sqrt{\cancel{t}x}$ for $x \in [0,1]$.

2.5

Assume $(\varphi,\psi) \in \Phi_c$ maximize $\mathbb{J}$. Let $c \in \mathbb{R}$, then

$$\mathbb{J}(\varphi+c,\psi-c) = \int_X \varphi(x)+c \, d\mu(x) + \int_Y \psi(y)-c \, d\nu(y)$$

$$= \int_X \varphi(x) d\mu + c \int_Y \psi(y) \, d\nu(y) - c$$

$$= \int_X \varphi(x) \, d\mu(x) + \int_Y \psi(y) d\nu(y)$$

$$= \mathbb{J}(\varphi,\psi).$$

And $(\varphi(x)+c) + (\psi(y) - c) = \varphi(x) + \psi(y) \le c(x,y)$

Hence $(\varphi+c, \psi-c) \in \Phi_c$.

It follows that $(\varphi+c, \psi-c)$ maximize $\mathbb{J}$ in Φ_c.

2.6

$$\Xi(u) = \begin{cases} \int_X \varphi(x) \, d\mu(x) + \int_Y \psi(y) \, d\nu(y) & \text{if } u(x,y) = \varphi(x) + \psi(y) \\ +\infty & \text{else} \end{cases}$$

If $u \neq 0$ then $\exists (x_0,y_0) \in X \times Y$ s.t. $u(x_0,y_0) \neq 0$. WLOG assume $u(x_0,y_0) > 0$. We claim $\nexists \varphi,\psi$ s.t. $u(x,y) = \varphi(x) + \psi(y)$. Suppose $u(x,y) = \varphi(x) + \psi(y)$. Now since $0 = \lim_{|x| \to \infty} u(x,y_0) = \lim_{|x|\to\infty} \varphi(x) + \psi(y_0)$ we have $\lim_{|x|\to\infty} \varphi(x) = -\psi(y_0)$. Similarly $\lim_{|y|\to\infty} \psi(y) = -\varphi(x_0)$. So $\lim_{\substack{|x|\to\infty \\ |y|\to\infty}} u(x,y) = -\psi(y_0) - \varphi(x_0) = -u(x_0,y_0) \neq 0$.

This contradicts $u(x,y) \to 0$ as $|x|+|y| \to +\infty$. So ~~such~~ no such φ and ψ exist. (then $\Xi(u) \neq +\infty$ unless $u \equiv 0$.)

Note: If X & Y are locally compact Hausdorff spaces then, by the Riesz-Markov theorem, $\left[C_0(X \times Y)\right]^* = M(X \times Y)$, hence for non-compact sets we would have to consider $E = C_0(X \times Y)$ in the proof of Lemma 4.8 which, as the example shows, step 1 does not hold in this generality.

2.7

~~By the duality theorem we know that~~

$$\cancel{M(T) \geq J(\varphi,\psi)}$$

~~$\forall\, T: X \to Y$ with $T_\# \mu = \nu$ and $(\varphi,\psi) \in \Phi_c$. So for (φ,ψ) to be optimal it is enough to find a T s.t. $M(T) = J(\varphi,\psi)$.~~

2.7 Let $x_0 \in \operatorname{spt}(\mu)$. Suppose ~~we can find a x_0 s.t.~~

$\phi(x_0) + \psi(T_0(x_0)) < c(x_0, T(x_0))$. By continuity of φ, ψ, T and c

we have that $\varphi(x) + \psi(T(x)) < c(x, T(x))$ in some neighborhood U of x_0

with $\mu(U) > 0$. Since $J(\varphi,\psi) \leq M(T)$ then we must in fact

have

$$J(\varphi,\psi) = \int_X \varphi(x) \, d\mu(x) + \int_Y \psi(y) \, d\nu(y)$$

$$= \int_X \varphi(x) + \psi(T(x)) \, d\mu(x)$$

$$< \int_X c(x, T(x)) \, d\mu(x)$$

$$= M(T)$$

This contradicts optimality of ~~J~~ (φ,ψ) and/or T (by the duality theorem)

Now consider the map $F(x) = c(x, T(x_0)) - \varphi(x)$, $x \in X$.

Since $\varphi(x) + \psi(y) \leq c(x,y)$ we have

$$F(x) \geq \psi(T(x_0)) \qquad \forall x \in X$$

and $F(x_0) = c(x_0, T(x_0)) - \varphi(x_0) = \psi(T(x_0))$ by the 1st part.

Then x_0 minimises F and $\therefore$ $\nabla F(x_0) = 0$

$$\Rightarrow \nabla_x c(x_0, T(x_0)) = \nabla \varphi(x_0)$$ as required.

2.8

By the duality theorem we know that

$$\mathcal{M}(T) \geq \mathbb{J}(\varphi,\psi)$$

$\forall T: X \to Y$ with $T_\# \mu = \nu$ and $(\varphi,\psi) \in \Phi_c$. So for (φ,ψ) to be optimal it is enough to find a T s.t. $\mathcal{M}(T) = \mathbb{J}(\varphi,\psi)$.

Since we are given candidate φ and ψ and we know from Exam 1.9 that $T(x) = x+1$ is optimal we ~~could~~ just check $\mathcal{M}(T) = \mathbb{J}(\varphi,\psi)$ directly.

~~Instead let us derive φ,ψ.~~ Check that $(\varphi,\psi) \in \Phi_c$

$\varphi(x) + \psi(y) = y - x \leq |x-y| = c(x,y)$

~~From 2.7 we have~~ And

~~$\varphi'(x) =$~~

$$\mathcal{M}(T) = \int_{\mathbb{R}} |x - T(x)| \, d\mu(x)$$

$$= \int_0^1 |x - x - 1| \, dx = 1$$

$$\mathbb{J}(\varphi,\psi) = \int \varphi(x) \, d\mu(x) + \int \psi(y) \, d\nu(y)$$

$$= -\int_0^1 x \, dx + \int_1^2 y \, dy = -\left[\frac{x^2}{2}\right]_0^1 + \left[\frac{y^2}{2}\right]_1^2$$

$$= -\tfrac{1}{2} + \tfrac{4}{2} - \tfrac{1}{2} = 1.$$ So $\mathcal{M}(T) = \mathbb{J}(\varphi,\psi)$ as required

2.9

We already know that $T(x) = x+1$ is optimal (and if we didn't we could check after we find (φ, ψ) using duality).

By exercise 2.7 we have $\nabla\varphi(x_0) = \nabla_x c(x_0, T(x_0))$

Now $\nabla_x c(x_0, y) = \frac{d}{dx}(x-y)^2 = 2(x-y)$ so

and $\varphi'(x) = 2(x - T(x)) = -2$

Integrating gives us $\varphi(x) = -2x + c$ for any $c \in \mathbb{R}$. Take $c = 0$.

Now $\psi(y) \leq |x-y|^2 - \varphi(x)$ $\forall x$

so $\psi(y) := \inf\limits_{x \in \mathbb{R}} \left(|x-y|^2 + 2x \right)$

$= \inf\limits_{x} \left(x^2 + y^2 - 2xy + 2x \right)$

$= \cancel{\inf} + x$

Let $F(x) = x^2 + y^2 - 2xy + 2x$

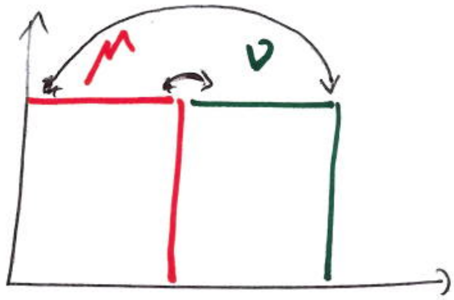
then $F'(x) = 2x - 2y + 2 = 2(x - y + 1)$

so $F'(x) = 0 \Rightarrow x - y + 1 = 0 \Rightarrow x = y - 1$

Hence $F(x) \geq F(y-1) = \cancel{(y-1)^2 + y^2 - 2(y-1)y} \; 1 + 2(y-1) = 2y - 1$

So $\psi(y) = 2y - 1$. One can check $\mathbb{J}(\varphi, \psi) = 1 = M(T)$.

2.10



Note c is concave.

Guess $T(x) = 2-x$ is the optimal map $\left(x \in [0,1] \right)$

Look for a potential ϕ that satisfies $\nabla \phi(x_0) = \nabla_x c(x_0, T(x_0))$

Now $\nabla_x c(x,y) = \frac{1}{2} |x-y|^{-1/2} \operatorname{sgn}(x-y)$

So
$$\phi'(x) = \tfrac{1}{2} |x - T(x)|^{-1/2} \operatorname{sgn}(x - T(x))$$
$$= \tfrac{1}{2} |2x-2|^{-1/2} \operatorname{sgn}(2x-2) \qquad \left[\begin{array}{l} 2x \in [0,2] \\ 2x-2 \in [-2,0] \end{array} \right.$$
$$= -\tfrac{1}{2} (2-2x)^{-1/2}$$
$$= -\frac{1}{2\sqrt{2}(1-x)^{1/2}}$$

So
$$\phi(x) = \frac{1}{\sqrt{2}} \sqrt{1-x} + c \qquad \text{, let } c=0$$
$$= \sqrt{\frac{1-x}{2}}$$

Now
$$\psi(y) = \inf_x \underbrace{\left(\sqrt{y-x} - \sqrt{\frac{1-x}{2}} \right)}_{F(x)}$$

$$F'(x) = \tfrac{1}{2}(y-x)^{-1/2} - \tfrac{1}{2}\left(\frac{1-x}{2}\right)^{-1/2} \times \left(-\tfrac{1}{2}\right)$$
$$= \tfrac{1}{2}(y-x)^{-1/2} + \tfrac{1}{4}\left(\frac{1-x}{2}\right)^{-1/2}$$

One can also use 2.7 to say $\psi(T(x)) = c(x, T(x)) - \varphi(x)$ and rearrange for $y = T(x)$. but then one needs to check $(\varphi, \psi) \in \Phi_c$.

So $F'(x) = 0 \Rightarrow 2(y-x)^{-1/2} = \left(\frac{1-x}{2}\right)^{-1/2}$

$$\Rightarrow \frac{y-x}{4} = \frac{1-x}{2}$$

$$\Rightarrow y-x = 2-2x$$

$$\Rightarrow \cancel{y = 2x} \quad \cancel{x=} \quad y = 2-x$$

$$\Rightarrow x = 2-y$$

So $F(x) \geq F(2-y) = \sqrt{y-2+y} - \sqrt{\frac{1-2+y}{2}}$

$$= \sqrt{2y-2} - \sqrt{\frac{y-1}{2}}$$

$$= \sqrt{2}\sqrt{y-1} - \frac{1}{\sqrt{2}}\sqrt{y-1}$$

$$= \left(\sqrt{2} - \frac{1}{\sqrt{2}}\right)\sqrt{y-1}$$

$$= \sqrt{2}\left(1-\frac{1}{2}\right)\sqrt{y-1}$$

$$= \frac{\sqrt{2}}{2}\sqrt{y-1}$$

Hence $\psi(y) = \frac{\sqrt{2}}{2}\sqrt{y-1}$

One can check $\mathcal{J}(\phi,\psi) = \frac{2\sqrt{2}}{3}$ and $\mathcal{M}(T) = \frac{2\sqrt{2}}{3}$ and (ϕ,ψ) maximize the latter by duality T minimizes MOT and (ϕ,ψ) maximize the dual problem.

2.11

Let $x_1 \sim x_2$ if $Ax_1 = Ax_2$ and $E = \mathbb{R}^n/\sim$. We first check that

$\langle x_1, x_2 \rangle_E = (Ax_1) \cdot (Ax_2)$ is an inner product.

It is clearly (i) symmetric $\langle x_1, x_2 \rangle_E = \langle x_2, x_1 \rangle_E$

(ii) linear $\langle \alpha x_1 + \beta x_2, x_3 \rangle_E = \alpha \langle x_1, x_3 \rangle_E + \beta \langle x_2, x_3 \rangle_E$

and (iii) positive $\langle x_1, x_1 \rangle_E = \|Ax_1\|^2 \geq 0$

And if $\|Ax_1\|^2 = 0 \Rightarrow Ax_1 = 0 \Rightarrow Ax_1 = A0 \Rightarrow x_1 \sim 0$ so

$x_1 = 0$ in E.

By the Riesz representation theorem $\forall F \in E^* \ \exists \ y \in \mathbb{R}^m$ s.t. $y = Ax'$, $x' \in \mathbb{R}^n$

$F(x) = \langle y, x \rangle_E = y \cdot (Ax)$, so $E^* = \mathrm{Ran}(A) \subseteq \mathbb{R}^m$

Define
$$\Theta(x) = \begin{cases} 0 & \text{if } Ax \geq -b \\ +\infty & \text{else} \end{cases} \qquad \Theta : E \to [0, +\infty]$$

$$\Xi(x) = c \cdot x \qquad \Xi : E \to \mathbb{R}$$

[To show Ξ is well defined we recall that $\exists y_0$ s.t. $A^T y_0 = c$ i.e. $c \cdot x = (A^T y_0) \cdot x = y_0 \cdot (Ax)$ so $c \cdot x$ only depends on Ax]

Ξ is linear so is convex. To see that Θ is convex pick

$x, y \in E$ and $t \in (0,1)$. We want to show

$$\Theta(tx + (1-t)y) \leq t\Theta(x) + (1-t)\Theta(y)$$

If either $\Theta(x) = +\infty$ or $\Theta(y) = +\infty$ the result is trivial. Assume

$\Theta(x) = \Theta(y) = 0$. It is enough to show $A(tx + (1-t)y) \geq -b$.

Since $Ax \geq -b$ and $Ay \geq -b$ we have $A(tx + (1-t)y) = tAx + (1-t)Ay \geq -b$

as required. So Θ is convex.

Any z_0 st. $Az_0 \geq -b$ satisfies $\Theta(z_0) < \infty$ and $\Xi(z_0) < \infty$, and Θ is cts at z_0.

So (1) $$\inf_{x \in E} \left(\Theta(x) + \Xi(x) \right) = \max_{y \in E^*} \left(-\Theta^*(-y) - \Xi^*(y) \right)$$

by Fenchel Rockafellar duality.

The LHS can be written

(2) $$\inf_{x \in E} \left(\Theta(x) + \Xi(x) \right) = \inf_{Ax \geq -b} c \cdot x = -\sup_{Ax \leq b} c \cdot x$$

For the RHS we compute

$$\Theta^*(y) = \sup_{x \in E} \left(\langle y, x \rangle_E - \Theta(x) \right) \qquad [A^\# y = \bar{y}]$$

$$= \sup_{\substack{x \in E \\ Ax \geq -b}} \bar{y} \cdot (Ax)$$

So $$\Theta^*(-y) = \sup_{\substack{x \in E \\ Ax \geq -b}} (-\bar{y}) \cdot (Ax) = \sup_{\substack{x \in E \\ Ax \leq b}} \bar{y} \cdot (Ax) = \begin{cases} \bar{y} \cdot b & \text{if } \bar{y} \geq 0 \\ +\infty & \text{else} \end{cases}$$

And $$\Xi^*(y) = \sup_{x \in E} \left(\langle y, x \rangle_E - \Xi(x) \right)$$

$$= \sup_{x \in E} \left(\bar{y} \cdot (Ax) - c \cdot x \right)$$

$$= \sup_{x \in E} \left((A^T \bar{y} - c) \cdot x \right)$$

$$= \begin{cases} 0 & \text{if } A^T \bar{y} = c \\ +\infty & \text{else} \end{cases}$$

clearly, if $\bar{y} \geq 0$, $\sup_{\substack{x \in E \\ Ax \leq b}} \bar{y} \cdot (Ax) = \bar{y} \cdot \mathrm{Proj}_{\mathrm{Ran}(A)}(b)$

(let $P = \mathrm{Proj}_{\mathrm{Ran}(A)}$, note $b - P(b) \in (\mathrm{Ran}\,A)^\perp = N(A^T)$)

So $\bar{y} \cdot b = (Ay) \cdot b = y \cdot (A^T b) = y \cdot (A^T P(b)) = (Ay) \cdot P(b) = \bar{y} \cdot P(b)$

since $A^T(b - P(b)) = 0$.

So

$$(3)\quad \max_{y\in B^*}\left(-\Theta^*(-y) - \Xi^*(y)\right) = \max_{\substack{\bar y\in \mathbb{R}^m/\ker(A)\\ \bar y\geq 0,\ A^T\bar y = c}} -\bar y\cdot b = -\min_{\substack{y\in \mathbb{R}^m/\ker(A)\\ \bar y\geq 0,\ A^T\bar y=c}} \bar y\cdot b$$

By (1), (2) and (3) we have

$$\min_{\substack{\bar y\in\mathbb{R}^m\\ \bar y\geq 0,\ A^T\bar y = c}} \bar y\cdot b \leq \min_{\substack{\bar y\in \mathbb{R}^m/\ker(A)\\ \bar y\geq 0,\ A^T\bar y=c}} \bar y\cdot b = \sup_{Ax\leq b} c\cdot x \quad (*) \quad ~~as required~~.$$

On the other hand, if $Ax\leq b$ and $\bar y\geq 0$, $A^T\bar y = c$ then

$$c\cdot x = (A^T\bar y)\cdot x = \bar y\cdot(Ax) \leq \bar y\cdot b \quad \text{so} \quad \sup_{Ax\leq b} c\cdot x \leq \min_{\substack{\bar y\in\mathbb{R}^m\\ \bar y\geq 0,\ A^T\bar y=c}} \bar y\cdot b \quad (+)$$

$(*) + (+) \Rightarrow$ result.

2.12

(a) Claim 1: μ_k and ν_k are tight.

Proof of claim 1

~~Let $\varepsilon>0$ and choose compact K_ε s.t. $\mu(K_\varepsilon)\geq 1-\varepsilon$~~

The closures of the sets $\{\mu_k\}_{k\in\mathbb{N}}$, $\{\nu_k\}_{k\in\mathbb{N}}$ are sequentially compact in the weak* topology. Hence by Prokhorov thm they are tight. ∎

Claim 2: $(\pi_k)_{k\in\mathbb{N}}$ is tight.

proof of claim 2

Let $\varepsilon>0$, choose $K_\varepsilon, L_\varepsilon$ s.t. $\mu_k(K_\varepsilon)\geq 1-\varepsilon/2$, $\nu_k(L_\varepsilon)\geq 1-\varepsilon/2$ $\forall k\in\mathbb{N}$. Then

$$\pi_k(K_\varepsilon\times L_\varepsilon) = \pi_k(K_\varepsilon\times Y) - \pi_k(K_\varepsilon\times(Y\setminus L_\varepsilon))$$

$$\geq \pi_k(K_\varepsilon\times Y) - \pi_k(X\times(Y\setminus L_\varepsilon))$$

since $K_\varepsilon\times L_\varepsilon = (K_\varepsilon\times Y)\setminus(K_\varepsilon\times(Y\setminus L_\varepsilon))$ and $K_\varepsilon\times(Y\setminus L_\varepsilon)\subseteq X\times(Y\setminus L_\varepsilon)$

Hence

$$\pi_k (K_\varepsilon \times L_\varepsilon) \geq \mu_k(K_\varepsilon) - \nu_k(Y \setminus L_\varepsilon)$$

$$\geq 1 - \varepsilon/2 - 1 + \nu_k(L_\varepsilon)$$

$$\geq 1 - \varepsilon .$$

Hence $\{\pi_k\}_{k \in \omega}$ is tight ∎

Claim 3: $\{\pi_k\}_{k \in \omega}$ is sequentially compact and any limit is in $\pi \in \Pi(\mu, \nu)$.

proof of claim 3
By Prokhorov theorem $\{\pi_k\}_{k \in \omega}$ is sequentially compact. Let $\pi_k \rightharpoonup \pi$. So $\lim_{k \to \infty} \int_{X \times Y} f(x,y)\, d\pi_k(x,y) = \int_{X \times Y} f(x,y)\, d\pi(x,y)$ $\forall f : X \times Y \to \mathbb{R}$, bounded & cts. Let $f(x,y) = \bar{f}(x)$ then

$$\int_{X \times Y} \bar{f}(x)\, d\pi(x,y) = \lim_{k \to \infty} \int_{X \times Y} \bar{f}(x)\, d\pi_k(x,y) = \lim_{k \to \infty} \int_X \bar{f}(x)\, d\mu_k(x)$$

$$= \int_X \bar{f}(x)\, d\mu(x)$$

And $$\int_{X \times Y} \bar{f}(x)\, d\pi(x,y) = \int_X \int_Y \bar{f}(x)\, d\pi(y|x)\, d(P_\#^X \pi)(x)$$

$$= \int_X \bar{f}(x)\, d(P_\#^X \pi)(x)$$

So $P_\#^X \pi = \mu$. Similarly $P_\#^Y \pi = \nu$. Hence $\pi \in \Pi(\mu, \nu)$ as required.

(b) ~~(ta~~

Let $\overline{\mathbb{K}}_n$ be defined by

$$\overline{\mathbb{K}}_n(\pi) = \begin{cases} \mathbb{K}(\pi) & \text{if } \pi \in \Pi(\mu_n, \nu_n) \\ +\infty & \text{else} \end{cases}$$

claim $\Gamma\text{-}\lim_{n\to\infty} \overline{\mathbb{K}}_n = \overline{\mathbb{K}}_\infty$ where

$$\overline{\mathbb{K}}(\pi) = \begin{cases} \mathbb{K}(\pi) & \text{if } \pi \in \Pi(\mu,\nu) \\ +\infty \end{cases}$$

and Γ-convergence is taken wrt weak* topology.

If so then by standard results in Γ-convergence $\pi_n \xrightarrow{*} \pi$.

<u>liminf</u>

Let $\pi_k \xrightarrow{*} \pi$. WTS $\liminf_{k\to\infty} \overline{\mathbb{K}}_k(\pi_k) \geq \overline{\mathbb{K}}(\pi)$

WLOG assume $\liminf_{k\to\infty} \overline{\mathbb{K}}_k(\pi_k) = \lim_{k\to\infty} \overline{\mathbb{K}}_k(\pi_k) < \infty$.

Then $\pi_k \in \Pi(\mu_k, \nu_k)$ so by part (a) $\pi \in \Pi(\mu,\nu)$.

Since c is cts and bounded we have

$$\lim_{k\to\infty} \mathbb{K}(\pi_k) = \lim_{k\to\infty} \int_{X\times Y} c(x,y)\, d\pi_k(x,y) = \int_{X\times Y} c(x,y)\, d\pi(x,y) = \mathbb{K}(\pi)$$

<u>limsup</u> Let $\pi \in \Pi(\mu,\nu)$ (result is trivial if $\pi \notin \Pi(\mu,\nu)$)
If we find a sequence $\pi_k \in \Pi(\mu_k,\nu_k)$ st. $\pi_k \xrightarrow{*} \pi$ then we are done.

Let $T_k^\mu : X \to X$, $T_k^\nu : Y \to Y$ be with the

$$(T_k^\mu)_\# \mu = \mu_k \qquad , \qquad (T_k^\nu)_\# \nu = \nu_k$$

$$\| T_k^\mu - \mathrm{Id} \|_{L^2(\mu)} \to 0 \qquad \| T_k^\nu - \mathrm{Id} \|_{L^2(\nu)} \to 0$$

So T_k^μ, T_k^ν exist because sol to the Monge problem exist and

$\mu_k \to \mu$, $\nu_k \to \nu$ in the Wasserstein distance.

Define $\quad \pi_k (A \times B) = \pi\left((T_k^\mu)^{-1}(A) \times (T_k^\nu)^{-1} (B) \right) = (T_k^\mu, T_k^\nu)_\# \pi (A \times B)$

so $\quad \pi_k (A \times Y) = \pi\left((T_k^\mu)^{-1}(A) \times (T_k^\nu)^{-1}(Y) \right)$

$$= \pi\left((T_k^\mu)^{-1}(A) \times Y \right)$$

$$= \mu\left((T_k^\mu)^{-1}(A) \right)$$

$$= (T_k^\mu)_\# \mu (A)$$

$$= \mu_k (A)$$

and similarly $\pi_k (X \times B) = \nu_k (B)$. So $\pi_k \in \Pi(\mu, \nu)$.

To show $\pi_k \to \pi$,

$$d_{W^2}^2(\pi_k, \pi) \le \int_{X \times Y} |(x,y) - (T_k^\mu(x), T_k^\nu(y))|^2 \, d\pi(x,y)$$

$$= \int_{X \times Y} |x - T_k^\mu(x)|^2 + |y - T_k^\nu(y)|^2 \, d\pi(x,y)$$

$$= \int_X |x - T_k^\mu(x)|^2 \, d\mu(x) + \int_Y |y - T_k^\nu(y)|^2 \, d\nu(y)$$

$$\to 0 .$$

Ex 2.12 (b) Question credit: Huy Pham

$\ast$ This solution uses results from the Wasserstein distance

~~Assume $c(x,y) = |x-y|^p$~~

Assume π is not optimal then $\exists \pi' \in \Pi(\mu,\nu)$ s.t.

$$\int_{X\times Y} c(x,y)\, d\pi'(x,y) < \int_{X\times Y} c(x,y)\, d\pi(x,y)$$

Since $\mu_k \rightharpoonup \mu$ and $\nu_k \rightharpoonup \nu$ then $\exists$

$$\pi_k^{(1)} \in \Pi(\mu, \mu_k) \quad \text{and} \quad \pi_k^{(2)} \in \Pi(\nu, \nu_k)$$

s.t.

$$\int_{X\times X} |x-y| \, d\pi_k^{(1)}(x,y) \to 0$$

$$\int_{Y\times Y} |x-y| \, d\pi_k^{(2)}(x,y) \to 0$$

Define the probability measure γ_k on $(\mathbb{R}^d)^4$ by

$$\gamma_k(A\times B\times C\times D) = \pi'(A\times B) \cdot \frac{\pi_k^{(1)}(A\times C)}{\mu(A)} \cdot \frac{\pi_k^{(2)}(B\times D)}{\nu(B)}$$

One can check γ_k has marginals μ, ν, μ_k, ν_k respectively

Let $P^{(i)}$ be the projection onto the ith coordinate, i.e.

$$P^{(1)}(x,y,w,z) = x, \quad P^{(2)}(x,y,w,z) = y, \ldots$$

And $\rho^{(i,j)}$ be the projection onto the i^{th}, j^{th} coordinates, i.e.

$$\rho^{(1,2)}(x,y,w,z) = (x,y).$$

Then notice that

$$\rho^{(1,2)}_{\#} \gamma_k = \pi' \in \Pi(\mu,\nu)$$

$$\rho^{(1,3)}_{\#} \gamma_k = \pi_k^{(1)} \in \Pi(\mu,\mu_k)$$

$$\rho^{(2,4)}_{\#} \gamma_k = \pi_k^{(2)} \in \Pi(\nu,\nu_k)$$

$$\rho^{(3,4)}_{\#} \gamma_k =: \hat{\pi}_k \in \Pi(\mu_k,\nu_k)$$

$\forall \varepsilon > 0$,

Choose $K \subset (\mathbb{R}^d)^2$ compact s.t.

$$\left.\begin{aligned} \mu_k((\rho^{(1)}K)^c) &< \varepsilon \\ \nu_k((\rho^{(2)}K)^c) &< \varepsilon \end{aligned}\right\} \forall k \in \mathbb{N}$$

$$\mu((\rho^{(1)}K)^c) < \varepsilon$$

$$\nu((\rho^{(2)}K)^c) < \varepsilon$$

$$\pi(K^c) < \varepsilon$$

and $\pi'(K^c) < \varepsilon$

$$\hat{\pi}_k(K^c) < \varepsilon$$

existence of such K follow from tightness of μ_k and ν_k, and the last follows from tightness of $\hat{\pi}_k$ which follows from (B).

The latter 2 conditions imply

$$\int_{K^c} c(x,y)\, d\pi(x,y) \le \varepsilon C$$

$$\int_{K^c} c(x,y)\, d\pi'(x,y) \le \varepsilon C \quad \text{when } c(x,y) \le C \ \forall x,y.$$

$(x,y),(w,z) \in K$
~~$(x,w),(y,z) \in K$~~

Since c is uniformly cts on K $\exists\, \delta$ s.t. if ~~$x,y,w,z \in K$~~

$|x-w| \le \delta$ and $|y-z| \le \delta$ then $|c(x,y) - c(w,z)| \le \varepsilon$.

Thus

$$\int_{K \times K} |c(x,y) - c(w,z)| \, d\gamma_k(x,y,w,z)$$

$$= \int_{\substack{|x-w| \le \delta \\ |y-z| \le \delta \\ (x,y),(w,z) \in K}} |c(x,y) - c(w,z)| \, d\gamma_k(x,y,w,z)$$

$$+ \int_{\substack{|x-w| > \delta \\ (x,y),(w,z) \in K}} |c(x,y) - c(w,z)| \, d\gamma_k(x,y,w,z)$$

$$+ \int_{\substack{|y-z| > \delta \\ (x,y),(w,z) \in K}} |c(x,y) - c(w,z)| \, d\gamma_k(x,y,w,z)$$

$$\le \varepsilon + 2C \, \pi_k^{(1)}(\{|x-w| > \delta\})$$

$$+ 2C \, \pi_k^{(2)}(\{|y-z| > \delta\})$$

Now,

$$\pi_n^{(1)}\left(\{|x-w|>\delta\}\right) = \int_{|x-w|>\delta} d\pi_n^{(1)}(x,w)$$

$$\leq \frac{1}{\delta}\int_{X\times X} |x-w| \, d\pi_n^{(1)}(x,w) \qquad \text{(Markov's inequality)}$$

$$\longrightarrow 0$$

And similarly

$$\pi_n^{(2)}\left(\{|y-z|>\delta\}\right) \longrightarrow 0$$

Hence $\int_{K\times K} |c(x,y)-c(w,z)| \, d\gamma_n(x,y,w,z) \longrightarrow 0.$

~~Now,~~ $\cancel{\int_{X\times Y} c(x,y) \, d\pi'(x,y) = \int_{X\times Y\times X\times Y} c(x,y) \, d\gamma_n(x,y,w,z)}$

$\cancel{\geq \int_{X\times Y\times X\times Y} c(w,z)}$

Now

(1) $\int_{X\times Y\times X\times Y} c(w,z)\, d\gamma_k(x,y,w,z) = \int_{X\times Y} c(w,z)\, d\hat{\pi}_k(w,z) \geq \int_{X\times Y} c(x,y)\, d\pi_k(x,y)$

where $\hat{\pi}_k = P^{(3,4)}_{\#}\gamma \in \Pi(\mu_k,\nu_k)$

and $\pi_k \in \Pi(\mu_k,\nu_k)$ is chosen optimally.

But $\int_{X\times Y\times X\times Y} c(w,z)\, d\gamma_k(x,y,w,z) \leq \int_{X\times Y\times X\times Y} c(x,y)\, d\gamma_k(x,y,w,z)$

$+ \int_{X\times Y\times X\times Y} |c(x,y) - c(w,z)|\, d\gamma_k(x,y,w,z)$

$\leq \int_{X\times Y} c(x,y)\, d\pi'(x,y) + o(1)\,\varepsilon + 2C\int_{K^c\times K} d\gamma_k(x,y,w,z)$

$+ 2C\int_{K\times K^c} d\gamma_k(x,y,w,z)$

$\leq \int_{X\times Y} c(x,y)\, d\pi'(x,y) + \varepsilon + o(1) + 2C\,\pi'(K^c) + 2C\hat{\pi}_k(K^c)$

(2) $\leq \int_{X\times Y} c(x,y)\, d\pi'(x,y) + (4C+1)\,\varepsilon + o(1)$

By (1) and (2) and letting $\varepsilon \to 0$ and $k\to\infty$ we have

$\int_{X\times Y} c(x,y)\, d\pi(x,y) \leq \int_{X\times Y} c(x,y)\, d\pi'(x,y)$ a contradiction.