L3: OT Duality & its geometry

© Jan Obłój

Consider $\varphi \in L^1(\mathcal{X},\mu)$, $\psi \in L^1(\mathcal{Y},\nu)$ s.t. $c(x,y) \geq \varphi(x)+\psi(y)$ $\forall (x,y)\in\mathcal{X}\times\mathcal{Y}$ (3.1)

Then integrating

$$\int c\, d\pi \geq \int(\varphi+\psi)\, d\pi = \int \varphi\, d\mu + \int \psi\, d\nu \quad \forall \pi\in\Pi(\mu,\nu) \quad (3.2)$$

$$\Longrightarrow \quad P(\mu,\nu) = \min_{\pi\in\Pi(\mu,\nu)} \int c\, d\pi \geq \sup_{\substack{\varphi\in L^1(\mu)\\ \psi\in L^1(\nu)}:\ \varphi+\psi\leq c} \int \varphi\, d\mu + \int\psi\, d\nu =: D(\mu,\nu)$$

We will show that in fact equality $P = D$ holds under weak assumptions. First we consider the dual pb in more detail.

While P is about the cost of the allocation given by π, the dual D is about prices.

A different company offers to buy your bread at price $\varphi(x)$ & sell it at price $(-\psi(y))$. To be competitive the P&L has to be better than before: $\varphi(x)+\psi(y)\leq c(x,y)$. But now they want to max

Let us first motivate why $P = D$ via a min-max argument.

Note that $\Pi(\mu,\nu)$ can be described via Lagrange multipliers: $\sup_{\varphi,\psi} \int_{\mathcal{X}}\varphi\, d\mu + \int_{\mathcal{Y}}\psi\, d\nu - \int_{\mathcal{X}\times\mathcal{Y}}(\varphi\oplus\psi)\, d\pi = \begin{cases} 0, & \pi\in\Pi(\mu,\nu)\\ +\infty, & \pi\notin\Pi(\mu,\nu)\end{cases}$

indep of φ,ψ

so

$$\inf_{\pi\in\Pi(\mu,\nu)}\int c\, d\pi = \inf_{\pi\in\mathcal{P}(\mathcal{X}\times\mathcal{Y})}\left(\int c\, d\pi + \sup_{\varphi,\psi} \int\varphi\, d\mu + \int\psi\, d\nu - \int\varphi\oplus\psi\, d\pi\right)$$

min-max thm Rockafellar

requires some compactness + convexity in one & concavity in the other variable.

$$\stackrel{?}{=} \sup_{\varphi,\psi}\inf_{\pi\in\mathcal{P}(\mathcal{X}\times\mathcal{Y})} \int\varphi\, d\mu + \int\psi\, d\nu + \int(c(x,y)-(\varphi+\psi))\, d\pi$$

But $\inf_\pi \int(c-\varphi\oplus\psi)\, d\pi = \begin{cases} 0 & \text{if } \varphi\oplus\psi\leq c \text{ on } \mathcal{X}\times\mathcal{Y}\\ -\infty & \text{otherwise}\end{cases}$ $\Longrightarrow$ $= \sup_{\varphi,\psi:\ \varphi\oplus\psi\leq c}\int\varphi\, d\mu + \int\psi\, d\nu.$

We end up with the same pb as above. We study it in more detail.

Given candidate φ,ψ, $\varphi\oplus\psi\leq c$ we can try to improve them in turn:

- fix φ & replace $\psi \rightsquigarrow \psi_1(y) = \inf_x\left(c(x,y)-\varphi(x)\right) =: \varphi^c(y)$

then
- fix ψ_1 & —"— $\varphi \rightsquigarrow \varphi(x) = \inf_y\left(c(x,y)-\psi_1(y)\right) =: \psi_1^{\,c}(x) = \varphi^{cc}(x)$

etc. — but in fact we stop here since $\forall\varphi$, $\varphi^{ccc} = \varphi^c$.

$\Longrightarrow$ We can restrict to (φ,ψ) of the form (φ^{cc},φ^c).

Rk By Lemma 2.4, $c = \lim\uparrow c_k$ of Lip-cont functions. Then $\varphi^c_k(y) = \inf_x\left(c_k(x,y)-\varphi(x)\right)$

is Lip-cont so in particular measurable. $\varphi^c = \lim \uparrow \varphi^c_k$ & hence measurable.

Rk Note that (3.1) was a bit too strong: for (3.2) it was enough to ask that (3.1) holds π-a.s. $\forall_{\pi \in \Pi(\mu,\nu)}$. Specifically we can replace (3.1) with

(3.1)' $\quad \varphi(x) + \psi(y) \le c(x,y) \quad \forall_{x \in X \setminus N_\mu,\ y \in Y \setminus N_\nu}$ for some $\mu(N_\mu)=0$, $\nu(N_\nu)=0$

since $\pi\left((X\setminus N_\mu \times Y \setminus N_\nu)^c \right) \le \pi\left(N_\mu \times Y\right) + \pi\left(X \times N_\nu\right) = \mu(N_\mu) + \nu(N_\nu) = 0$

Rk For an optimal $\pi^* \in \Pi(\mu,\nu)$ if $P = D$ we have to have equalities throughout & hence

$$\varphi^{cc} \oplus \varphi^c = c \quad d\pi^*\text{-a.e.}$$

We will see that such a relation in fact characterises c-cyclically monotone sets & allows to construct a proof of the duality along the lines:

+ various cont./bdd assumptions

+ limiting arguments

↳ there exists at least one π^* concentrated on a c-cyclically monotone set Γ

↳ Γ is supported by some φ^c : $\varphi^{cc} + \varphi^c \le c$ with equality on Γ

↳ we get $\int c \, d\pi^* = \int \varphi^{cc} d\mu + \int \varphi^c d\nu \Rightarrow$ duality & optimality

We will come back to some of these ideas but first we render the above min-max argument rigorous.

Ex (Linear programming)

For $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$
$$\sup_{Ax \le b} c \cdot x = \inf_{y \ge 0,\ A^T y = c} b \cdot y$$

Thm (Minmax, von Neumann, Sion). Let K be a compact convex subset of a Hausdorff topological space & $\mathcal{Z}$ be a convex subset of a vector space. Let $v : K \times \mathcal{Z} \to \mathbb{R}$ be s.t.
$x \mapsto v(x,y)$ is lsc & convex $\forall_{y \in \mathcal{Z}}$
$y \mapsto v(x,y)$ is concave $\forall_{x \in K}$

Then
$$\min_{x \in K} \sup_{y \in \mathcal{Z}} v(x,y) = \sup_{y \in \mathcal{Z}} \min_{x \in K} v(x,y)$$

Fenchel - Rockafeller duality

Let E be a normed vector space & $\Theta : E \to \mathbb{R} \cup \{+\infty\}$ convex

$$\Theta^*(z^*) := \sup_{z \in E} \left[\langle z^*, z \rangle - \Theta(z) \right] \quad \text{for } z^* \in E^*$$

the topological dual.

<u>Thm</u> (F-R) Let Θ, Ξ be two convex functions on E s.t. for some $z_0 \in E$, $\Theta(z_0) < \infty$, $\Xi(z_0) < \infty$ & Θ is continuous at z_0.

Then $\inf_{z \in E} (\Theta(z) + \Xi(z)) = \max_{z^* \in E^*} \{ -\Theta^*(-z^*) - \Xi^*(z^*) \}$

(FR)

<u>Rk</u> sup = max on RHS; uses axiom of choice if E is not separable.

<u>Proof</u> (Hahn-Banach)

We want $\inf_{z \in E} (\Theta(z) + \Xi(z)) = \sup_{z^* \in E^*} \inf_{x,y \in E} (\Theta(x) + \Xi(y) + \langle z^*, x-y \rangle)$

$x=y$ gives "$\geq$". For the reverse we need a linear form $z^* \in E^*$ s.t.

$$\inf \Theta + \Xi =: m \leq \Theta(x) + \Xi(y) + \langle z^*, x-y \rangle \ \forall x,y \in E$$

Consider two convex sets $C = \{(x,\lambda) \in E \times \mathbb{R} : \lambda > \Theta(x)\}$

$C' = \{(y,\mu) \in E \times \mathbb{R} : \mu \leq m - \Xi(y)\}$

- $(z_0, \Theta(z_0)+1) \in \mathrm{Int}(C) \Rightarrow \bar{C} = \overline{\mathrm{Int}(C)}$

- $C \cap C' = \emptyset$ since if $m - \Xi(x) \geq \lambda > \Theta(x)$ then $m > \Theta(x) + \Xi(x)$ contradiction.

$\overset{\text{H-B}}{\Longrightarrow}$ $\exists \ l \in (E \times \mathbb{R})^*$ satisfying

$$\inf_{c \in C} \langle l, c \rangle = \inf_{c \in \mathrm{Int}(C)} \langle l, c \rangle \geq \sup_{c' \in C'} \langle l, c' \rangle$$

i.e. $\exists\, w^* \in E^*$ & $\alpha \in \mathbb{R}$, $(w^*, \alpha) \neq (0,0)$ s.t.

$$\langle w^*, x\rangle + \alpha\lambda \geq \langle w^*, y\rangle + \alpha\mu \quad \forall\ \lambda > \Theta(x),\ \mu \leq m - \Xi(y)$$

$\Rightarrow \alpha \neq 0 \Rightarrow$ let $z^* = w^*/\alpha \Rightarrow \langle z^*, x\rangle + \Theta(x) \geq \langle z^*, y\rangle + m - \Xi(y)$

$\lambda \searrow \Theta(x)$ etc.

Kantorovich Duality $P = D$

[going back to Kantorovich but with key works by Rachev & Rüschendorf; Brenier; McCann, Gangbo & others]

Thm (duality) Let $\mathcal{X}, \mathcal{Y}$ be Polish spaces, $\mu \in \mathcal{P}(\mathcal{X})$, $\nu \in \mathcal{P}(\mathcal{Y})$ & $c: \mathcal{X}\times\mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ a lsc cost function bounded from below.

Then:
$$P(\mu,\nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int c\, d\pi = D(\mu,\nu) = \sup_{\substack{\varphi \in L^1(\mu) \\ \psi \in L^1(\nu)}\,:\ \varphi\oplus\psi \leq c\ \mu\otimes\nu\text{-a.e.}} \int \varphi\, d\mu + \int \psi\, d\nu$$

Moreover, the LHS is attained & on the RHS one can restrict to φ, ψ bounded & continuous & to $(\varphi,\psi) = (\varphi^{cc}, \varphi^{c})$.

Proof We only prove now the case of $\mathcal{X}, \mathcal{Y}$ compact & c continuous.

Let $E = C_b(\mathcal{X}\times\mathcal{Y})$ with $\|\cdot\|_\infty$ $\overset{\text{Riesz}}{\Longrightarrow}$ $E^* = \mathcal{M}(\mathcal{X}\times\mathcal{Y})$ space of regular (Radon) measures with TV.

A non-negative $z^* \in E^*$ is a finite Borel measure.

Let $\Theta : E \to \{0, +\infty\}$

$$z \mapsto \begin{cases} 0 & \text{if } z(x,y) \geq -c(x,y) \\ +\infty & \text{otherwise} \end{cases}$$

w.l.o.g. let $c \geq 0$.

$$\Xi : E \to \mathbb{R} \cup \{+\infty\} \text{ via } z \mapsto \begin{cases} \int \varphi\, d\mu + \int \psi\, d\nu & \text{if } z(x,y) = \varphi(x) + \psi(y) \\ +\infty & \text{else} \end{cases}$$

we can move a constant from φ to ψ but that does not affect $\Xi(z)$

We can apply F-R (with $z_0 \equiv 1$).

The (LHS) of (FR) is

$$\inf_{z \in E} \big(\Theta(z) + \Xi(z)\big) = \inf\Big\{ \int \varphi\, d\mu + \int \psi\, d\nu : \varphi \oplus \psi \geq -c \Big\} = -D(\mu,\nu)$$

Now for Θ^* & Ξ^*. For any $\pi = z^* \in E^* = M(X \times Y)$

$$\Theta^*(-\pi) = \sup_{z : z \geq -c} -\int z \, d\pi = \sup_{z : z \leq c} \int z \, d\pi = \begin{cases} \int c \, d\pi & : \pi \in M_+ \\ +\infty & \text{else} \end{cases}$$

$$\Xi^*(\pi) = \begin{cases} 0 & \text{if } \forall (\varphi, \psi) \in C_b(X) \times C_b(Y) \quad \int \varphi \oplus \psi \, d\pi = \int \varphi \, d\mu + \int \psi \, d\nu \\ +\infty & \text{else} \end{cases}$$

$\pi \in M(\mu, \nu)$

So (PR) reads

$$(LHS) = -D(\mu, \nu) = (RHS) = \max_{\pi \in E^*} \left\{ -\Theta^*(-\pi) - \Xi(\pi) \right\}$$

$$= \max_{\pi \in M(\mu, \nu)} -\int c \, d\pi = -P(\mu, \nu).$$

$\boxtimes$

Step 3 Relax compactness. Keep c bdd & unif cont. Take π^* & use its compactness + restriction property. On the dual use improvements $(\varphi^{cc}, \varphi^c)$. This here shows we can use unif cont dual potentials ($\longrightarrow$ inherit prop/cont. of c).

Step 4 $c = \lim \uparrow c_n$. + compactness of $M(\mu, \nu)$.

$\boxtimes$

Examples / Applications

1. K-R

Consider $X = Y$ & $c(x,y) = d(x,y)$ a lsc metric. Then

Corr (K-R distance) For $\mu, \nu \in P(X)$

$$\inf_{\pi \in M(\mu, \nu)} \int d(x,y) \, \pi(dx, dy) = \sup \left\{ \int \varphi \, d(\mu - \nu) : \varphi \in L^1(d|\mu - \nu|) \text{ \& } \|\varphi\|_{Lip} \leq 1 \right\}$$

$$\text{where } \|\varphi\|_{Lip} = \sup_{x \neq y} \frac{|\varphi(x) - \varphi(y)|}{d(x,y)}$$

Proof

Let $d_n = \dfrac{d}{1 + \frac{1}{n} d} \leq d$ & bounded, $d_n \uparrow d$.

$1\text{-}Lip(d_n) \subseteq 1\text{-}Lip(d)$ & via limits it suffices to prove things for d_n

1-Lip (d_n) $\Rightarrow$ bounded $\Rightarrow$ integrable. $\Rightarrow$ we now assume d is bdd.

We already know that $D(\mu,\nu) = \sup_{\varphi \in C_b(X)} \int \varphi^{dd} d\mu + \int \varphi^{d} d\nu$.

Ex Let (f_α) be a family of functions with common modulus of continuity ω. Then $\inf_\alpha f_\alpha$ & $\sup_\alpha f_\alpha$ also enjoy the same mod. of continuity on their domain.

It follows that $\varphi^d(y) = \inf_x (d(x,y) - \varphi(x))$ is 1-Lip (we had φ bdd)

$\Rightarrow \varphi^d(y) - \varphi^d(x) \leq d(x,y)$ take $y=x$

$$-\varphi^d(x) \leq \inf_y \left(d(x,y) - \varphi^d(y) \right) \leq -\varphi^d(x)$$

Hence $\varphi^{dd} = -\varphi^d$

$$D(\mu,\nu) = \sup_{\varphi \in C_b} \int -\varphi^d d\mu + \int \varphi^d d\nu \leq \sup_{\substack{\varphi : \|\varphi\|_{Lip} \leq 1 \\ \text{bdd}}} \int \varphi \, d(\mu-\nu)$$
$$\leq D(\mu,\nu)$$

Def Let $P_1(X) = \{ \mu \in P(X) : \int d(x_0,x)\mu(dx) < \infty \}$ (def is indep of the choice of x_0).

$\|\mu\|_{KR} := \sup \{ \int_X \varphi \, d\mu : \varphi \in L^1(\mu) \text{ \& } \|\varphi\|_{Lip} \leq 1 \}$ is a norm on $P_1(X)$.

The associated distance: $W_1^d(\mu,\nu) := \|\mu - \nu\|_{KR}$ is our first example of a Wasserstein-distance.

② Bdd c, $c \geq 0$ $\Longrightarrow$ in $D(\mu,\nu)$ it is enough to consider $0 \le \varphi \le \|c\|_\infty$

$-\|c\|_\infty \le \psi \le 0$

$$-\sup\varphi \le \varphi^c(y) = \inf_x \left(c(x,y) - \varphi(x)\right) \le \|c\|_\infty - \sup\varphi$$

$$-\sup\varphi^c \le \varphi^{cc} = \|c\|_\infty - \sup\varphi^c$$

also $(\varphi + \text{const})^c = \varphi^c - \text{const}$ so we can assume $\sup\varphi = \|c\|_\infty$

$\Longrightarrow -\|c\|_\infty \le \varphi^c \le 0 \Longrightarrow 0 \le \varphi^{cc} \le \|c\|_\infty$

3. TV Let $c(x,y) = 1\!\!1_{x \neq y}$. Then $\inf_{\pi \in \Pi(\mu,\nu)} \pi(X \neq Y) = \sup \int \varphi^{cc} d\mu + \int \varphi^c d\nu$

But $-1 \le \varphi^c \le 0$ & $\varphi^{cc}(x) = \inf_y \left(1\!\!1_{y \neq x} - \varphi^c(y)\right) \le -\varphi^c(x)$

$\quad\quad\quad\quad\quad\quad \geq (-\varphi^c(x)) \wedge 1$

$\Longrightarrow \varphi^{cc} = -\varphi^c$.

$$\inf_{\pi \in \Pi(\mu,\nu)} \pi(X \neq Y) = \sup_{0 \le \varphi \le 1} \int_{\mathcal{X}} \varphi \, d(\mu - \nu) = (\mu - \nu)_+(\mathcal{X}) \underset{\text{parties}}{=} (\mu - \nu)_-(\mathcal{X}) = \frac{1}{2}\|\mu - \nu\|_{TV}$$

More on c-concavity & c-cyclical monotonicity

Lemma 3.1 Let $\varphi : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$. Then $\varphi^{ccc} = \varphi^c$. [Exercise]

Proof $\varphi^c(y) = \inf_x (c(x,y) - \varphi(x))$; $\varphi^{cc}(x) = \inf_y (c(x,y) - \varphi^c(y)) = \inf_y \left(c(x,y) - \inf_{\tilde{x}}(c(\tilde{x},y) - \varphi(\tilde{x}))\right)$

$$= \inf_y \sup_{\tilde{x}} \left(c(x,y) - c(\tilde{x},y) + \varphi(\tilde{x})\right)$$

$$\varphi^{ccc}(\tilde{y}) = \inf_x \sup_y \inf_{\tilde{x}} \left(c(x,\tilde{y}) - c(x,y) + c(\tilde{x},y) - \varphi(\tilde{x})\right)$$

$$\overset{\tilde{x}=x}{\geq} \inf_x \sup_y \left(c(x,\tilde{y}) - \cancel{c(x,y)} + \cancel{c(x,y)} - \varphi(x)\right) = \inf_x (c(x,\tilde{y}) - \varphi(x)) = \varphi^c(\tilde{y})$$

$$\underset{y=\tilde{y}}{\leq} \inf_x \inf_{\tilde{x}} \left(\cancel{c(x,\tilde{y})} - \cancel{c(x,\tilde{y})} + c(\tilde{x},\tilde{y}) - \varphi(\tilde{x})\right)$$

$$= \varphi^c(\tilde{y}).$$

∎

Def We say that φ is c-concave if $\varphi \not\equiv -\infty$ and $\exists\, \psi : \mathcal{Y} \to \mathbb{R} \cup \{\pm\infty\}$ s.t. $\varphi = \psi^c$.

Lemma 3.1 shows that φ is c-concave iff $\varphi = \varphi^{cc}$.

Thm (dual existence) In the setting of Thm [Duality], assume that $c \leq c_X \oplus c_Y$ for some $c_X \in L^1(\mu)$, $c_Y \in L^1(\nu)$. Then $\mathcal{D}(\mu,\nu)$ admits maximiser $(\varphi^{cc}, \varphi^c)$.

Proof Assume $\mathcal{X}, \mathcal{Y}$ are compact & c is cont. $\Rightarrow$ c is unif cont with modulus ω. Both φ^{cc}, φ^c inherit this modulus. So if $(\varphi_n^{cc}, \varphi_n^c)$ is an optimizing sequence it is equicontinuous & w.l.o.g. (by adding a constant & subtracting) we can take $\varphi_n^{cc} \geq 0$ & then also $\varphi_n^{cc} \leq \omega(\text{diam}(\mathcal{X}))$ $\Rightarrow$ $\varphi_n^c = \varphi_n^{ccc} \in [\min c - \omega(\text{diam}(\mathcal{X})), \max c]$ so equibounded $\xrightarrow{A-A}$ a converging subsequence $\to$ (φ, ψ) which attains $\mathcal{D}(\mu,\nu)$. This will already be of the form $(\varphi^{cc}, \varphi^c)$ but we can always improve (φ,ψ) by taking $(\varphi^{cc}, \varphi^c)$ so OK. ∎

Take such optimizers π^*, $(\varphi^{cc}, \varphi^c)$. Then

$$\mathcal{P}(\mu,\nu) = \int c \, d\pi^* \geq \int (\varphi^{cc} \oplus \varphi^c) \, d\pi^* = \int \varphi^{cc} d\mu + \int \varphi^c d\nu = \mathcal{D}(\mu,\nu) = \mathcal{P}(\mu,\nu)$$

$$\Rightarrow \varphi^{cc} \oplus \varphi^c = c \quad d\pi^*\text{-a.e.}$$

Conversely, if for $\pi \in \Pi(\mu,\nu)$ we have $\varphi^{cc} \oplus \varphi^c = c$ $d\pi$ - a.e. $\Rightarrow$ π is an optimiser.

Lemma 3.2 Suppose $(\varphi^{cc}, \varphi^c)$ is an optimiser in $D(\mu,\nu)$. Then

$\pi \in \Pi(\mu,\nu)$ is a optimiser in $P(\mu,\nu)$ iff $\varphi^{cc} \oplus \varphi^c = c$ $d\pi$ - a.e.

Lemma 3.3 $\pi^* \in \Pi(\mu,\nu)$ — " — " — $\Rightarrow$ $\pi^*(\Gamma) = 1$ for some Γ - c-cyclically monotone.

Proof Write $\varphi := \varphi^{cc}$ & $\psi = \varphi^c$. "$\Rightarrow$" Take π^*.

$\forall$ pts $(x_1,y_1), \dots, (x_n,y_n)$ $\forall$ σ - permutation

$$\sum_i c(x_i, y_{\sigma(i)}) \geq \sum \varphi(x_i) + \psi(y_{\sigma(i)}) = \sum_i \varphi(x_i) + \psi(y_i)$$

$$\underset{d\pi^*\text{-a.e.}}{=} \sum_i c(x_i, y_i) \Rightarrow \pi^* \text{ concentrated on c-c.m.}$$

If no optimiser for $D(\mu,\nu)$ then take a sequence, pass to a sub-sequence converging $d\pi^*$ - a.e., let the set of points where conv holds be Γ, & use l.s.c.

"$\Leftarrow$" Fix $(x_0,y_0) \in \Gamma$.

Let $$\varphi(x) = \inf \{ c(x,y_n) - c(x_n,y_n) + c(x_n,y_{n-1}) - c(x_{n-1},y_{n-1}) + c(x_1,y_0) - c(x_0,y_0) : n \in \mathbb{N}, (x_1,y_1), \dots, (x_n,y_n) \in \Gamma \}$$

& $$-\psi(y) = \inf \{ c(x_n,y) - c(x_n,y_{n-1}) + \dots + c(x_1,y_0) - c(x_0,y_0) \quad n \in \mathbb{N}, (x_1,y_1), \dots, (x_n,y_n) \in \Gamma, y_n = y \}$$

$\Rightarrow \varphi < +\infty$; $\psi(y) > -\infty$ iff $y \in \mathrm{proj}_Y(\Gamma)$

$\psi^c(x) = \inf_y \{ c(x,y) - \psi(y) \} = \varphi(x)$ so φ is c - concave.

Also $\varphi \oplus \varphi^c \leq c$ by def & we show $\varphi \oplus \varphi^c = c$ on Γ. ∎

Def $\partial^c \varphi$, a c-superdifferential of a c-concave φ is the set of all $(x,y) \in X \times Y$ s.t.

$$\forall_{z \in X} \quad \varphi(z) \leq \varphi(x) + (c(z,y) - c(x,y))$$

Thm [Rüschendorf] Any c-cyclically monotone set Γ can be included in $\partial^c \varphi$ of a c-concave φ.

Rk This really is the same. $(x,y) \in \partial^c \varphi$ iff

$$c(z,y) - \varphi(z) \geq c(x,y) - \varphi(x) \qquad \forall z \in \mathcal{X}$$

iff $$c(x,y) - \varphi(x) = \inf_z \left(c(z,y) - \varphi(z) \right) = \varphi^c(y)$$

i.e. $$c(x,y) = \varphi(x) + \varphi^c(y) .$$

We briefly review the case $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$ & $c(x,y) = |x-y|^2$ to highlight links with objects known from the classical convex analysis.

Convex functions

- $f: \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is convex if $f(\lambda x + (1-\lambda) y) \le \lambda f(x) + (1-\lambda) f(y)$ $\forall x, y \in \mathbb{R}^d$, $\lambda \in [0,1]$
- sup of convex functions is convex
- f is continuous & loc. Lipschitz on the interior of $\{f < +\infty\}$
- f lsc. $\Rightarrow f(x) = \sup \{ax + b : a, b \text{ s.t. } f(y) \ge ay + b \ \forall y \in \mathbb{R}^d\}$ (f convex l.s.c. is sup of affine functions)
- Legendre - Fenchel: $f^*(y) = \sup_x (xy - f(x))$

  f convex & lsc iff $f^{**} = f$. (Legendre duality)

  Sub differential $\partial f(x) = \{p \in \mathbb{R}^d : f(y) \ge f(x) + p \cdot (y - x) \ \forall y \in \mathbb{R}^d\}$

  - is non-empty for $f \in \mathrm{Int}\{f < +\infty\}$
  - f diff at $x \Rightarrow \partial f(x) = \nabla f(x)$.
  - $p \in \partial f(x) \Leftrightarrow x \in \partial f^*(p) \Leftrightarrow f(x) + f^*(p) = xy$.
  - monotonicity: $p_1 \in \partial f(x_1)$, $p_2 \in \partial f(x_2) \Rightarrow (x_1 - x_2)(p_1 - p_2) \ge 0$. (i.e. f' is increasing)

    $\Leftrightarrow x_1 p_1 + x_2 p_2 \ge x_1 p_2 + x_2 p_1$

  this is of course implied by cyclical monotonicity: $\sum x_i p_i \ge \sum x_i p_{\sigma(i)}$

  Thm [Rockafeller] Every cyclically monotone set Γ is contained in the graph of the subdifferential of a convex function f: $\Gamma \subseteq \{(x,p) : p \in \partial f(x)\}$

- $\dfrac{|x-y|^2}{2} = \dfrac{1}{2} \sum (x_i - y_i)^2 = \dfrac{1}{2}\left(\sum x_i^2 + \sum y_i^2 - 2 \sum x_i y_i\right) = \dfrac{|x|^2}{2} + \dfrac{|y|^2}{2} - x \cdot y$

  $\varphi(x) + \psi(x) \le \dfrac{|x-y|^2}{2} \Leftrightarrow x \cdot y \le \left(\dfrac{|x|^2}{2} - \varphi(x)\right) + \left(\dfrac{|y|^2}{2} - \psi(y)\right)$

Thm (Brenier) Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\mu(A) = 0$ $\forall A \subseteq \mathbb{R}^d$ of Hausdorff dim $(d-1)$. Then there exists a ($d\mu$ - unique) convex u s.t. $\nu = \nabla u_\# \mu$ & $\pi^* = (\mathrm{Id}, \nabla u)_\# \mu \in \Pi(\mu, \nu)$ is optimal for $c = |x-y|^2$.
($\nabla u(x) = x - \nabla \varphi(x)$).

Rk If also ν gives no mass to small sets then

$D\varphi^* \circ D\varphi(x) = x$; $D\varphi \circ D\varphi^*(y) = y$

& $D\varphi^*_{\#}\nu = \mu$ (& it unique $d\nu$ - a.e.).

Rk (φ, φ^*) are the Kantorovich potentials for $\sup_{\pi \in \Pi(\mu,\nu)} \int xy \, d\pi$

& $\left(\frac{|x|^2}{2} - \varphi, \frac{|y|^2}{2} - \varphi^*\right)$ are for $\inf_{\pi} \int |x-y|^2 d\pi$.

Rk If $\mu(dx) = f(x)dx$, $\nu(dx) = g(x)dx$ on $\mathbb{R}^d$ then

$T_{\#}\mu = \nu$ is, via change of variables, $\det(DT) = \frac{f}{g \circ T}$

For the case $c = |x-y|^2$ & $T = D\varphi$ we get the

Monge-Ampère equation: $\det(D^2\varphi(x)) = \frac{f(x)}{g(D\varphi(x))}$
(non-linear, elliptic)

More generally: Spence-Mirrlees condition $\det\left(\frac{\partial^2 c}{\partial y_i \partial x_j}\right) \neq 0$.

Thm Given μ, ν on a compact domain $\Omega \subseteq \mathbb{R}^d$ & $c(x,y) = h(x-y)$ for a strictly convex h, $\exists!\ \pi^* = (id, T)_\# \mu$ (assuming $\mu \ll Leb$ & $\partial\Omega$ negl.) & $T(x) = x - (\nabla h)^{-1}(\nabla\varphi(x))$ for the Kant. potential φ.

Pf

$\varphi(x_0) + \varphi^c(y_0) = c(x_0, y_0)$ i.e. $x \mapsto \varphi(x) - c(x, y_0)$ is minimised at x_0

so $\nabla\varphi(x_0) \in \partial h(x_0 - y_0)$ but since h is str. convex

its gradient is invertible so $x_0 - y_0 = \partial h^{-1}(\nabla\varphi(x_0))$

i.e. $y_0 = T(x_0) = x_0 - \partial h^{-1}(\nabla\varphi(x_0))$.

∎

Stability of opt. π^*

Thm Let X, Y be compact & c cont. If π_n^* are optimal couplings & $\pi_n^* \to \pi$ then π is optimal for $\mu := proj_X \# \pi$ & $\nu = proj_Y \# \pi$.

Pf

Let $\Gamma_n = supp(\pi_n)$. Then $\Gamma_n \to \Gamma$ in Hausdorff & Γ is c-c.m. simply because for $(x_i, y_i) \in \Gamma$ we find $(x_i^n, y_i^n) \in \Gamma_n$ & use cont. of c. & c-CM of Γ_n.

& since $supp(\pi) \subseteq \Gamma \Rightarrow \pi$ is optimal.

∎