Week 3

Recall: Given $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, $c : X \times Y \to [0,\infty]$,

$$(MP) = \inf \left\{ \int_X c(x, T(x))\, d\mu(x) \ : \ T_\# \mu = \nu \right\}$$

$$(KP) = \inf \left\{ \int_{X\times Y} c(x,y)\, d\gamma(x,y) \ : \ \gamma \in \underbrace{\Pi(\mu,\nu)}_{\text{set of plans between } \mu \text{ and } \nu} \right\}$$

Narrow convergence : $\mu_k \to \mu \iff \int \varphi\, d\mu_k \to \int \varphi\, d\mu \quad \forall \varphi \in \mathcal{C}_b(X)$

$\iff \liminf_k \int \varphi\, d\mu_k \geq \int \varphi\, d\mu \quad \forall \varphi$ l.s.c., $\varphi \geq -C$. (*)

Remark: $\mu_k \in \mathcal{P}(X)$, $\mu_k \to \mu \Rightarrow \mu \in \mathcal{P}(X)$. Test with $\varphi \equiv 1$.

Theorem (Prokhorov) $A \subseteq \mathcal{P}(X)$ is tight $\iff$ A is relatively compact for narrow convergence

tight: $\forall \varepsilon > 0$, $\exists K_\varepsilon \subset X$ compact s.t. $\mu(X \setminus K_\varepsilon) < \varepsilon$ $\forall \mu \in A$

Theorem : Let $c : X \times Y \longrightarrow \overset{\text{or } [-C,\infty]}{[0,\infty]}$ l.s.c., $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$.

Then, there exists a coupling $\gamma_{opt} \in \Pi(\mu,\nu)$ minimizer in (KP).

Idea: • $\gamma \mapsto \int c\, d\gamma$ is l.s.c. (when $\gamma_k \to \gamma$, $\liminf_k \int c\, d\gamma_k \geq \int c\, d\gamma$) by (*).

• Precompactness of a sequence with bounded cost.

Proof : W.l.o.g. assume $(KP) < +\infty$. Let $(\gamma_k)_k$ a minimizing sequence,

$$\int c\, d\gamma_k \to (KP).$$

Claim: $(\gamma_k)_k \subseteq \Pi(\mu,\nu)$ is tight. $\Pi(\mu,\nu)$ is closed (w.r.t. narrow convergence).

With this claim and Prokhorov theorem, up to a subsequence

$\gamma_{k_j} \to \gamma_{opt} \in \Pi(\mu,\nu)$. Then $(KP) = \lim_{j\to\infty} \int c\, d\gamma_{k_j} \geq \int c\, d\gamma_{opt} \geq (KP)$

$\Longrightarrow$ γ_{opt} is a minimum in (KP).

$\longrightarrow$ Proof of claim :

• Closed: if $\Pi(\mu,\nu) \ni \gamma_k \to \gamma \in \mathcal{P}(X \times Y)$, then

$$\forall \varphi \in \mathcal{C}_b(X), \quad \underbrace{\int \varphi\, d\mu}_{\overset{\gamma_k \in \Pi(\mu,\nu)}{=} \int \varphi(x)\, d\gamma_k(x,y)} \overset{?}{=} \int \varphi(x)\, d\gamma(x,y)$$

as $k \to \infty$ (by narrow convergence).

same for the second marginal.

• <u>Tight</u>: By Lemma ⊙ $\forall \varepsilon > 0$, $\exists K_\varepsilon \subset X$, $K'_\varepsilon \subset Y$ s.t.

$$\mu(X \setminus K_\varepsilon) < \varepsilon, \qquad \nu(Y \setminus K'_\varepsilon) < \varepsilon .$$

($\{\mu\}$ is compact $\Longrightarrow$ it is tight by Prokhorov's).

Take $K_\varepsilon \times K'_\varepsilon$ compact in $X \times Y$.

Then, for every $\gamma \in \Gamma(\mu,\nu)$, $\gamma((X\setminus K_\varepsilon)\times Y) = \mu(X \setminus K_\varepsilon) < \varepsilon$

$$\gamma(X \times (Y \setminus K'_\varepsilon)) = \nu(Y \setminus K'_\varepsilon) < \varepsilon$$

$$\gamma((X\times Y)\setminus(K_\varepsilon \times K'_\varepsilon)) \le \gamma((X\setminus K_\varepsilon)\times Y) + \gamma(X \times (Y\setminus K'_\varepsilon)) < 2\varepsilon .$$

□

<u>Remark</u>: Same strategy does <u>not</u> work for (MP)!

Indeed, let T_k be a sequence s.t. $\int |x - T_k(x)|^2 \, d\mu(x) \to (MP)$

T_k bounded in $L^2(\mu) \Rightarrow T_k \rightharpoonup T$ weakly in L^2.

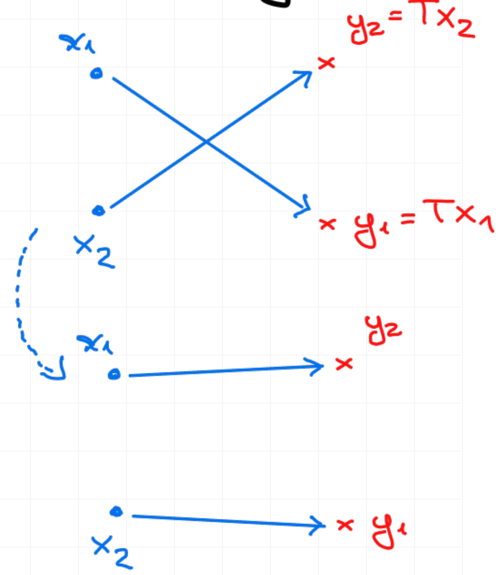
But $T_k \rightharpoonup T$ weakly in L^2, $T_k \# \mu = \nu \not\Rightarrow T \# \mu = \nu$.

$T_k \# \mathcal{L}_{[0,1]} = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}$

$T = 0$, $T_\# \mathcal{L}_{[0,1]} = \delta_0$

□ An optimality condition : c-cyclical monotonicity

Heuristic: (MP) or (KP) with $c(x,y) = |x-y|^2$



I expect, if T optimal,

$$c(x_1,y_1)+c(x_2,y_2) \leq c(x_1,y_2)+c(x_2,y_1)$$

(when $y_i = Tx_i$, (x_i,y_i) are in the support of γ)

Rigorously:

<u>Definition</u>: Given $\mu \in \mathcal{M}(X)$, the support of μ is

$$\operatorname{supp}\mu := \{x \in X : \forall \varepsilon > 0,\ \mu(B_\varepsilon(x)) > 0\}.$$

Equivalently, it is the smallest closed set such that $\mu(X \setminus C) = 0$.

⚠ <u>Exercise</u>: $\{q_i\}_i$ rationals, $\mu = \sum_i \frac{1}{2^i}\delta_{q_i} \in \mathcal{P}(\mathbb{R})$. Then $\operatorname{supp}\mu \neq \mathbb{Q}$, $= \mathbb{R}$,

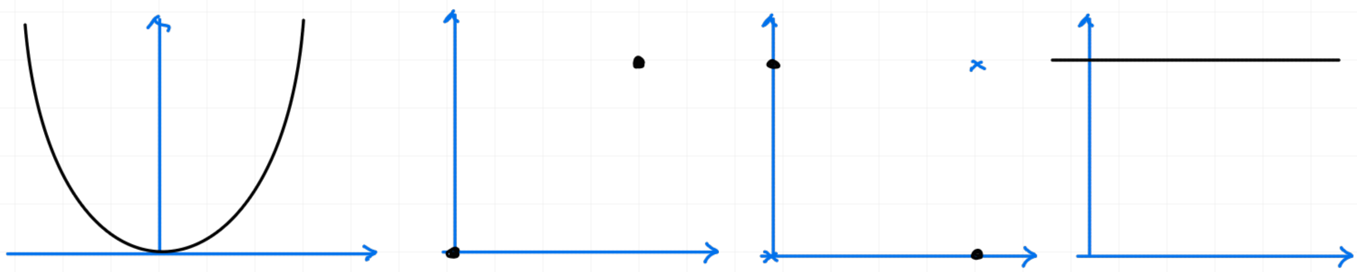
but μ is concentrated on $\mathbb{Q}$. $(\mu(\mathbb{R}\setminus\mathbb{Q}) = 0)$.

→ like supp γ

<u>Definition</u>: A set $\Lambda \subseteq X \times Y$ is <u>c-cyclically monotone</u> if

$$\forall (x_1,y_1),\dots,(x_N,y_N) \in \Lambda, \quad \sum_{i=1}^N c(x_i,y_i) \leq \sum_{i=1}^N c(x_{i+1},y_i), \quad x_{N+1} = x_1.$$

<u>Example</u>: which are $|x-y|^2 = c(x,y)$ - cyclically monotone in $\mathbb{R}^2$?



Theorem: Let $\bar{\gamma}$ optimal in (KP) with $c : X \times Y \to \mathbb{R}$ continuous.

Then, $\operatorname{supp} \bar{\gamma}$ is c-cyclically monotone.

Proof: Assume by contradiction that $\exists (\bar{x}_1, \bar{y}_1), \dots, (\bar{x}_N, \bar{y}_N) \in \operatorname{supp} \bar{\gamma}$

s.t. $\sum c(\bar{x}_i, \bar{y}_i) \geq \sum c(\bar{x}_{i+1}, \bar{y}_i) + \eta, \quad \eta > 0.$

Since c is continuous, $\exists\, \mathcal{U}_i, V_i$ neighbourhoods of $\bar{x}_i, \bar{y}_i$ respect. such that:

$$|c(x,y) - c(\bar{x}_i, \bar{y}_i)| < \frac{\eta}{2N} \qquad \forall (x,y) \in \mathcal{U}_i \times V_i$$

$$|c(x,y) - c(\bar{x}_{i+1}, \bar{y}_i)| < \frac{\eta}{2N} \qquad \forall (x,y) \in \mathcal{U}_{i+1} \times V_i .$$

Let $\varepsilon_i := \bar{\gamma}(\mathcal{U}_i \times V_i) > 0, \quad \varepsilon := \min\{\varepsilon_i : i = 1, \dots, N\},$

$$\gamma_i := \frac{1}{\varepsilon_i} \bar{\gamma}\Big|_{\mathcal{U}_i \times V_i} \in \mathcal{P}(\mathcal{U}_i \times V_i), \quad \mu_i = \pi_X \# \gamma_i, \quad \nu_i = \pi_Y \# \gamma_i .$$

Our competitor in (KP) will be:

$$\gamma' = \underbrace{\bar{\gamma} - \frac{\varepsilon}{N} \sum_{i=1}^{N} \gamma_i}_{\geq \bar{\gamma} - \frac{1}{N} \sum_{i=1}^{N} \underbrace{\frac{\varepsilon}{\varepsilon_i}}_{\leq 1} \bar{\gamma}|_{\mathcal{U}_i \times V_i} \geq \bar{\gamma} - \frac{1}{N} \overbrace{\sum_{i=1}^{N} \bar{\gamma}|_{\mathcal{U}_i \times V_i}}^{\leq \bar{\gamma}} \geq 0} + \frac{\varepsilon}{N} \sum_{i=1}^{N} \mu_{i+1} \otimes \nu_i \quad \in \mathcal{P}(X \times Y)$$

- Verify $\gamma' \in \Gamma(\mu, \nu)$: first marginal (same for second):

$$\pi_X \# \gamma' = \mu - \frac{\varepsilon}{N} \sum_{i=1}^{N} \mu_i + \frac{\varepsilon}{N} \sum_{i=1}^{N} \mu_{i+1} = \mu \qquad (\mu_{N+1} := \mu_1).$$

- γ' costs less that $\bar{\gamma}$:

$$\text{Cost}(\gamma') = \int c \, d\gamma' = \int c(x,y)\, d\bar{\gamma}(x,y) + \frac{\varepsilon}{N} \sum_{i=1}^{N} \overbrace{\left[-\int c \, d\gamma_i + \int c \, d(\mu_{i+1} \otimes \nu_i) \right]}^{\leq -c(\bar{x}_i, \bar{y}_i) + \frac{\eta}{4N} + c(\bar{x}_{i+1}, \bar{y}_i) + \frac{\eta}{4N}}$$

$$= \int c \, d\bar{\gamma} + \frac{\varepsilon}{N} \overbrace{\sum_{i=1}^{N} \left(c(\bar{x}_{i+1}, \bar{y}_i) - c(\bar{x}_i, \bar{y}_i) \right)}^{< -\eta} + \frac{\varepsilon \eta}{2N} < \int c \, d\bar{\gamma} - \frac{\varepsilon \eta}{2N} < \int c \, d\bar{\gamma},$$

which contradicts the optimality of $\bar{\gamma}$. $\square$

c-cyclically monotone sets with $c(x,y) = \frac{|x-y|^2}{2}$, $X = Y = \mathbb{R}^d$.

Remark: $(KP) = \min_{\gamma \in \Gamma(\mu,\nu)} \int \frac{|x-y|^2}{2} d\gamma = \min_{\gamma \in \Gamma(\mu,\nu)} \int \frac{|x|^2}{2} d\mu + \int \frac{|y|^2}{2} d\nu - \int x\cdot y \, d\gamma$

$$= \int \frac{|x|^2}{2} d\mu + \int \frac{|y|^2}{2} d\nu + \min_{\gamma \in \Gamma(\mu,\nu)} \int -x\cdot y \, d\gamma$$

minimizing costs with $|x-y|^2$ is equivalent to minimizing costs with $-x\cdot y$.

Analogously, c-cyclical monotonicity of Λ means:

$$\sum \frac{|x_i - y_i|^2}{2} \leq \sum \frac{|x_{i+1} - y_i|^2}{2} \qquad \forall \, (x_i, y_i)_{i=1,\dots,N} \in \Lambda$$

$$\sum \left(\frac{|x_i|^2}{2} + \frac{|y_i|^2}{2} - x_i\cdot y_i \right) \leq \sum \left(\frac{|x_{i+1}|^2}{2} + \frac{|y_i|^2}{2} - x_{i+1}\cdot y_i \right)$$

$$\Leftrightarrow \quad \sum y_i \cdot (x_{i+1} - x_i) \leq 0 .$$

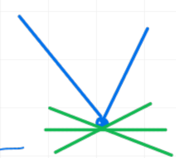
done in exercise class

Exercise: take $\varphi \in C^1(\mathbb{R}^d)$ convex. Graph of $\nabla\varphi$ is c-cyclically monotone $\left(c(x,y) = \frac{|x-y|^2}{2}\right)$.

Definition: Given $\varphi : \mathbb{R}^d \to \mathbb{R}$ convex, the subdifferential of φ at $x \in \mathbb{R}^d$ is:

$$\partial\varphi(x) := \left\{ y \in \mathbb{R}^d : \forall z \in \mathbb{R}^d, \ \varphi(z) \geq \varphi(x) + \langle y, z-x \rangle \right\}$$

all slopes inside



Properties (as exercises):

- φ convex $\Leftrightarrow$ $\partial\varphi(x) \neq \emptyset \ \forall x$
- φ convex, differentiable at x $\Leftrightarrow$ $\partial\varphi(x) = \{p\}$ (only one element), in which case, $p = \nabla\varphi(x)$
- φ convex, x_0 global minimum $\Leftrightarrow$ $0 \in \partial\varphi(x_0)$.

Theorem (Rockafeller): $\left(c(x,y) = \frac{|x-y|^2}{2}\right)$

$S \subseteq \mathbb{R}^d \times \mathbb{R}^d$ is c-cyclically monotone $\Leftrightarrow$ $\exists\, \varphi$ convex : $\mathbb{R}^d \to \mathbb{R}$ such that $S \subseteq \partial\varphi$

$$\partial\varphi = \bigcup_x \{x\} \times \partial\varphi(x)$$