Weak Optimal Transport with Unnormalized Kernels

Nathaël Gozlan

Université Paris Cité

New Monge Problems and Applications
14 september 2023

Outline

Work in collaboration with P. Choné and F. Kramarz

I - Weak Optimal Transport with (un)-normalized kernels : motivations and examples

II - General results : primal attainment, duality

III - The particular case of barycentric and conical costs

IV - Perspectives

I - Weak Optimal Transport with (un)-normalized kernels : motivations and examples

Optimal Transport - classical definition

Let $\mathcal{X}, \mathcal{Y}$ be Polish spaces and $\omega : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^+$ a measurable function.

Definition

The optimal transport cost between two probability measures μ and ν is given by

$$\mathrm{T}_\omega(\nu, \mu) = \inf_{\pi \in \Pi(\mu,\nu)} \iint_{\mathcal{X} \times \mathcal{Y}} \omega(x, y) d\pi(x, y),$$

where $\Pi(\mu, \nu)$ denotes the set of probability measures π on $\mathcal{X} \times \mathcal{Y}$ having μ and ν as marginals (called 'transport plans between μ and ν').

Equivalently

$$\mathrm{T}_\omega(\nu, \mu) = \inf_{X \sim \mu, Y \sim \nu} \mathbb{E}[\omega(X, Y)]$$

Classical Examples : Kantorovich distances of order $p \geq 1$

$$W_p^p(\nu, \mu) = \inf_{X \sim \mu, Y \sim \nu} \mathbb{E}[d^p(X, Y)]$$

(if $\mathcal{X} = \mathcal{Y}$ and d metrizes $\mathcal{X}$).

Weak Optimal Transport

G.-Roberto-Samson-Tetali (2017), Alibert-Bouchitté-Champion (2018)

Let $\pi \in \Pi(\mu, \nu)$ be a transport plan between μ and ν written in disintegrated form

$$d\pi(x, y) = d\mu(x)dp_x(y),$$

with $x \mapsto p_x$ a transition kernel (μ a.s unique).

Interpretation :The kernel p_x tells where the mass coming from x is allocated over $\mathcal{Y}$.

If $\omega : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^+$ is a cost function then

$$\iint \omega(x, y)\, d\pi(x, y) = \int \left(\int \omega(x, y)\, dp_x(y) \right) d\mu(x).$$

In other words, transports of mass coming from x are penalized through their mean cost : $\int \omega(x, y)\, dp_x(y)$.

Idea of WOT :introduce more general penalizations.

Weak Optimal Transport (WOT)

Let $\mathcal{P}(\mathcal{Y})$ denote the set of all probability measures on $\mathcal{Y}$.

Definition

Let $c : \mathcal{X} \times \mathcal{P}(\mathcal{Y}) \to \mathbb{R}^+ \cup \{+\infty\}$; the weak optimal transport cost $\mathcal{T}_c(\mu, \nu)$ is defined by

$$\mathcal{T}_c(\mu, \nu) = \inf_{p \in \mathcal{P}(\mu, \nu)} \int c(x, p_x)\, d\mu(x),$$

where $\mathcal{P}(\mu, \nu)$ is the set of all probability kernels p such that $\mu p = \nu$.

Classical transport :

$$c(x, p) = \int \omega(x, y)\, dp(y).$$

In all useful examples, the function c is convex in p.

Comments

- First examples go back to the works of K. Marton (1996) on concentration of measure.
- Further developments by Samson (2000, 2007) : concentration for Markov chains or suprema of empirical processes.
- The framework of weak transport contains many variants of the transport problem : Schrödinger transport problem, martingale transport problem, semi-martingale transport problem,. . .
- General tools (duality, cyclical monotonicity) have been developed to study weak transport problems. See Backhoff-Veraguas, Beiglböck, Pammer (2019).
- Other specific applications : concentration for convex functions (GRST 2017), discrete curvature bounds (GRST 2014, Samson 2022, 2023), model-independent pricing problem (Acciaio-Beiglboeck-Pammer 2021)

  Nice survey paper by Backhoff-Veraguas and Pammer (2020).

Examples

(1) **Barycentric transport :** $\mathcal{X} = \mathcal{Y} = \mathbb{R}^n$ and

$$c(x,p) = \theta\left(x - \int y\,dp(y)\right),$$

where $\theta : \mathbb{R}^n \to \mathbb{R}^+$ (convex).
We will denote by $\overline{\mathcal{T}}_\theta(\mu,\nu)$ the corresponding weak optimal transport cost.

(2) **Transport with martingale constraints** : $\mathcal{X} = \mathcal{Y} = \mathbb{R}^n$ and

$$c(x,p) = \begin{cases} \int \omega(x,y)\,dp(y) & \text{if } \int y\,dp(y) = x \\ +\infty & \text{otherwise} \end{cases}$$

Beiglboeck-Juillet (2016)

Examples

(3) **Entropic regularized transport / Schrödinger bridges** :
Let R be a reference probability measure on $\mathcal{X} \times \mathcal{X}$

$$\mathcal{T}_H(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} H(\pi|R),$$

where H is the relative entropy defined by

$$H(\pi|R) = \int \log \frac{d\pi}{dR}\, d\pi$$

if $\pi \ll R$ (and $+\infty$ otherwise).

Writing $d\pi(x,y) = d\mu(x) dp_x(y)$ and $dR(x,y) = dm(x) dr_x(y)$, one gets

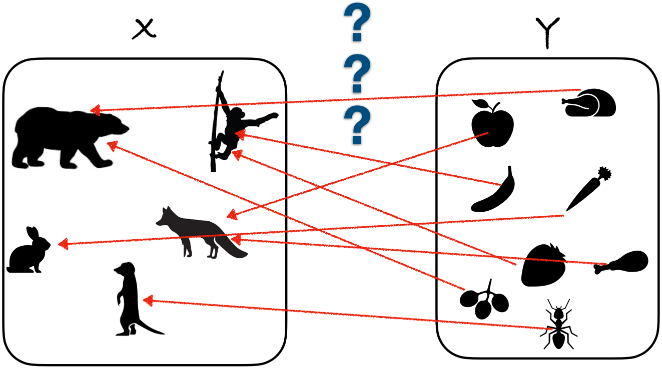
$$H(\pi|R) = H(\mu|m) + \int H(p_x|r_x)\, d\mu(x) := H(\mu|m) + \int c(x, p_x)\, d\mu(x)$$

'Zero noise limit' : Mikami, Thieullen, Léonard, Carlier-Duval-Peyré,. . .
Applications : Cutturi, Peyré,. . .
Functional inequalities : Gentil-Léonard-Ripani, Gigli-Tamanini, Fathi-G.-Prod'homme,. . .

(4) . . .

From WOT to WOTUK : a toy example

A zoo contains several groups of animals. Each week, a certain amount of feed is received by the zoo. What is the best way to feed the animals ?



A first modelization attempt with WOT

- $\mathcal{X}$ is the finite set of all animals of the zoo
- $\mathcal{Y}$ is the set of types of feed received by the zoo
- $\mu = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} \delta_x$ is the distribution of animals
- $\nu \in \mathcal{P}(\mathcal{Y})$ is the distribution of feeds received by the zoo
- For $x \in \mathcal{X}$ and $p \in \mathcal{P}(\mathcal{Y})$,

$$c(x, p)$$

  is some quantity reflecting the health of animal x when it is fed with the distribution of feeds p. By convention, we want to minimize this quantity.
- A coupling

$$d\pi(x, y) = d\mu(x) dp_x(y) \in \Pi(\mu, \nu)$$

  gives a way to allocate the feed among animals. Here p_x represents the feed received by the animal x. The equation

$$\nu = \mu p$$

  means that all the feed has been distributed.
- The best allocation is obtained by solving

$$\mathcal{T}_c(\mu, \nu) = \inf_{p \in \mathcal{P}(\mu, \nu)} \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} c(x, p_x).$$

A first modelization attempt with WOT

Criticism :

- in this model, each animal x receives a fixed portion $1/|\mathcal{X}|$ of the total amount of feed.
- $c(x, p)$ does not depend on the quantity of feed received by animal x but only on the composition of its meal.

$\rightsquigarrow$ We need to relax the WOT framework, in such a way that the quantity of food received by each animal becomes a new optimization parameter.

WOT with Unnormalized Kernels (WOTUK)

Denote by $\mathcal{M}(\mathcal{Y})$ the set of all non-negative finite measures on $\mathcal{Y}$.

Definition

Let $c : \mathcal{X} \times \mathcal{M}(\mathcal{Y}) \to \mathbb{R} \cup \{+\infty\}$; the unnormalized weak transport cost $\mathcal{I}_c(\mu, \nu)$ between $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$ is defined by

$$\mathcal{I}_c(\mu, \nu) = \inf_{q \in \mathcal{Q}(\mu,\nu)} \int c(x, q^x)\, d\mu(x),$$

where $\mathcal{Q}(\mu, \nu)$ is the set of all non-negative kernels q (i.e $q^x(dy) \in \mathcal{M}(\mathcal{Y})$ for all $x \in \mathcal{X}$) such that $\mu q = \nu$.

If $q \in \mathcal{Q}(\mu, \nu)$, then for all $x \in \mathcal{X}$,

$$q_x(dy) = N(x) p_x(dy)$$

with $\int N(x)\, \mu(dx) = 1$ and p a probability kernel transporting $\eta = N\mu$ onto ν.

Interpretation : $N(x)$ represents the quantity of feed received by animal x and p_x is the composition of its meal.

Remark

We deal with measures of probability only by convenience. The definition above also makes sense for positive measures μ, ν with possibly different masses.

Equivalent formulation

Let

$$\Pi(\ll \mu, \nu)$$

be the set of all probability measures π on $\mathcal{X} \times \mathcal{Y}$ such that π_1 is absolutely continuous w.r.t μ.

For any $\pi \in \Pi(\ll \mu, \nu)$, let

$$I_c[\pi] = \int c\left(x, \frac{d\pi_1}{d\mu}(x)\pi_x\right)\, d\mu(x)$$

where $d\pi(x,y) = d\pi_1(x) d\pi_x(y)$.

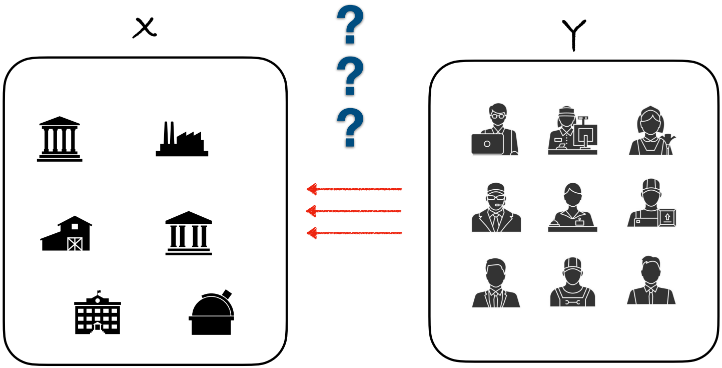
Then, it holds

$$\mathcal{I}_c(\mu,\nu) = \inf_{\pi \in \Pi(\ll \mu,\nu)} I_c[\pi].$$

Indeed,

$$\begin{aligned}
\mathcal{I}_c(\mu,\nu) &= \inf_{\eta \ll \mu} \inf_{p \in \mathcal{P}(\eta,\nu)} \int c\left(x, \frac{d\eta}{d\mu}(x) p_x\right)\, d\mu(x) \\
&= \inf_{\eta \ll \mu} \inf_{\pi \in \Pi(\eta,\nu)} \int c\left(x, \frac{d\pi_1}{d\mu}(x)\pi_x\right)\, d\mu(x) \\
&= \inf_{\pi \in \Pi(\ll \mu,\nu)} I_c[\pi].
\end{aligned}$$

Economic motivation (Choné - Kramarz 2021)



Economic motivation (Choné - Kramarz 2021)

- $\mathcal{X}$ is the space of firms types
- $\mathcal{Y}$ is the space of workers skill's profiles
- μ is the distribution of firms in a given economy (the sizes of the firms are unknown)
- ν is the distribution of workers in a given economy
- q^x represents the workers recruited by firm x.
  For instance $q^x(dy) = \sum_{i=1}^k n_i \delta_{y_i}$ means that firm x has recruited n_i workers with the skill profile y_i. Here the size of firm x is $N(x) = \sum_{i=1}^k n_i$.
- $-c(x, m)$ represents the output of firm x when it recruits a distribution of workers m.

Goal : Find the optimal allocation of workers to optimize the total output in the economy.

II - General results : primal attainment, duality

Primal attainment and duality for WOT

Theorem (Backhoff-Veraguas - Beiglboeck - Pammer (2018))

If $c : \mathcal{X} \times \mathcal{P}(\mathcal{Y}) \to \mathbb{R} \cup \{+\infty\}$ is jointly lower semi-continuous, lower bounded and convex in p, then for all $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$, there exists $p \in \mathcal{P}(\mu, \nu)$ such that

$$\mathcal{T}_c(\mu, \nu) = \int c(x, p_x)\, d\mu(x).$$

Moreover, the following Kantorovich type dual formula holds

$$\mathcal{T}_c(\mu, \nu) = \sup_{f \in \mathcal{C}_b(\mathcal{Y})} \left\{ \int R_c f\, d\mu - \int f\, d\nu \right\}, \qquad \mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$$

with

$$R_c f(x) = \inf_{p \in \mathcal{P}(\mathcal{Y})} \left\{ \int f\, dp + c(x, p) \right\}, \qquad x \in \mathcal{X}.$$

Improves G.-Roberto-Samson-Tetali (2017) and Alibert-Bouchitté-Champion (2018).
Links with backward linear mass transfers Bowles-Ghoussoub (2019).

Duality holds under more general conditions on the cost function : μ, ν have finite k-th moment and c is lower semicontinuous w.r.t W_k topology, $k \geq 1$.

Primal attainment and duality for WOTUK

$\mathcal{X}$ and $\mathcal{Y}$ will always be assumed to be compact.

Usual Assumptions :

(A) c can be written as

$$c(x,m) = \sup_{k\in\mathbb{N}} \left\{ \int a_k(x,y)\, dm(y) + b_k(x) \right\}, \qquad x \in \mathcal{X}, \qquad \forall m \in \mathcal{M}(\mathcal{Y}),$$

where, for all $k \in \mathbb{N}$, $a_k : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ and $b_k : \mathcal{X} \to \mathbb{R}$ are continuous functions.

(B) for all $x \in \mathcal{X}$ and $m \in \mathcal{M}(\mathcal{Y}) \setminus \{0\}$,

$$c'_\infty(x,m) := \lim_{\lambda\to\infty} \frac{c(x,\lambda m)}{\lambda} = +\infty.$$

Remark

(A) implies in particular that c is lower bounded, convex w.r.t its second variable and jointly l.s.c. Is there equivalence ?

Primal attainment and duality for WOTUK

Theorem (CGK, 2023))

If $c : \mathcal{X} \times \mathcal{M}(\mathcal{Y}) \to \mathbb{R} \cup \{+\infty\}$ satisfies Assumptions (A) and (B), then for all $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$, there exists $q \in \mathcal{Q}(\mu, \nu)$ such that

$$\mathcal{I}_c(\mu, \nu) = \int c(x, q_x)\, d\mu(x).$$

Moreover, the following Kantorovich type dual formula holds

$$\mathcal{I}_c(\mu, \nu) = \sup_{f \in \mathcal{C}_b(\mathcal{Y})} \left\{ \int K_c f \, d\mu - \int f \, d\nu \right\}, \qquad \mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$$

with

$$K_c f(x) = \inf_{m \in \mathcal{M}(\mathcal{Y})} \left\{ \int f \, dm + c(x, m) \right\}, \qquad x \in \mathcal{X}.$$

Sketch of proof for primal attainment

As observed earlier,

$$\mathcal{I}_c(\mu,\nu) = \inf_{\pi \in \Pi(\ll \mu,\nu)} I_c[\pi],$$

where

$$\Pi(\ll \mu,\nu)$$

is the set of probability measures on $\mathcal{X} \times \mathcal{Y}$ such that π_1 is absolutely continuous w.r.t μ and

$$I_c[\pi] = \int c\left(x, \frac{d\pi_1}{d\mu}(x)\pi_x\right) d\mu(x)$$

with $d\pi(x,y) = d\pi_1(x) d\pi_x(y)$.

Using Assumption (A), one can show that I_c is lower semicontinuous on $\Pi(\ll \mu,\nu)$.

Sketch of proof for primal attainment

For simplicity, suppose that there exists a convex function $\phi : \mathbb{R}_+ \to \mathbb{R}$ such that $\phi(x)/x \to +\infty$ and

$$(B') \qquad c(x,m) \geq \phi(m(\mathcal{Y})), \qquad \forall x \in \mathcal{X}, \forall m \in \mathcal{M}(\mathcal{Y}).$$

Then Assumption (B) holds.

Take π_n a sequence such that $I_c[\pi_n] \to \mathcal{I}_c(\mu,\nu)$ and denote $\eta_n = (\pi_n)_1$. By compactness, one can assume that π_n converges to some probability measure π on $\mathcal{X} \times \mathcal{Y}$ with marginals η and ν. If $\eta \ll \mu$, then since I_c is lsc, it holds

$$I_c[\pi] \leq \liminf_{n\to\infty} I_c[\pi_n] = \mathcal{I}_c(\mu,\nu).$$

So, there is attainment.

Let us show that $\eta \ll \mu$. Using (B'), one sees that

$$\sup_{n\in\mathbb{N}} \int \phi\left(\frac{d\eta_n}{d\mu}(x)\right) d\mu(x) < +\infty.$$

Therefore, by Dunford-Pettis theorem, the sequence $(\frac{d\eta_n}{d\mu}(x))_n$ admits a converging subsequence for the topology $\sigma(L_1(\mu), L_\infty(\mu))$. So $\eta \ll \mu$.

There is not always primal attainment

Suppose that μ is the uniform measure on $\mathcal{X} = [0,1]$ and ν is an arbitrary probability measure on $\mathcal{Y} = [2,3]$ and define

$$c(x,m) = \int |x-y|^2 \, m(dy), \quad x \in [0,1], \qquad m \in \mathcal{M}(\mathcal{Y}).$$

Then,

$$\mathcal{I}_c(\mu,\nu) = \inf_{\mu q = \nu} \iint |y-x|^2 \mu(dx) q^x(dy) = \inf_{\eta \ll \mu} W_2^2(\eta,\nu) = W_2^2(\delta_1,\nu).$$

This lower bound is not reached.

There is not always primal attainment

Suppose that μ is the uniform measure on $\mathcal{X} = [0,1]$ and ν is an arbitrary probability measure on $\mathcal{Y} = [2,3]$ and define

$$c(x,m) = \int |x-y|^2 \, m(dy), \quad x \in [0,1], \qquad m \in \mathcal{M}(\mathcal{Y}).$$

Then,

$$\mathcal{I}_c(\mu,\nu) = \inf_{\mu q = \nu} \iint |y-x|^2 \mu(dx) q^x(dy) = \inf_{\eta \ll \mu} W_2^2(\eta,\nu) = W_2^2(\delta_1,\nu).$$

This lower bound is not reached.

This motivates the following

Definition

A coupling π with second marginal ν is called a weak solution for $\mathcal{I}_c(\mu,\nu)$ if there is a sequence $\pi_n \in \Pi(\ll \mu, \nu)$ such that $I_c[\pi_n] \to \mathcal{I}_c(\mu,\nu)$ and $\pi_n \to \pi$.

Proposition

Weak solutions always exist. Under Assumption (B), any weak solution is a (strong) solution.

Technical Issue : Condition (B) is not meaningful in an economic context ...

Let $o(x,q)$ be the output when a firm x hire a worker with skills profile q.
Relation :

$$c(x,q) = -o(x,q).$$

Natural condition : $o(x,q) \geq 0$ and concave in q. So

$$c'_\infty(x,q) \leq 0$$

and (B) is never satisfied ...

Attainment and duality under weaker conditions

Theorem (CGK, 2023)

Suppose that $c : \mathcal{X} \times \mathcal{M}(\mathcal{Y}) \to \mathbb{R}$ satisfies assumption (A) and is such that

$$\begin{cases} - \text{ for all } m \in \mathcal{M}(\mathcal{Y}), \text{ the functions } c(\cdot, m) \text{ and } c'_\infty(\cdot, m) \text{ are continuous on } \mathcal{X}, \\ \text{and} \\ - \text{ there exists } a \geq 0 \text{ such that } c'_\infty(x,p) \leq a \text{ for all } x \in \mathcal{X} \text{ and } p \in \mathcal{P}(\mathcal{Y}). \end{cases} \quad \text{(C)}$$

Let $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$.

A coupling π^* with second marginal ν is a weak solution for $\mathcal{I}_c(\mu,\nu)$ if and only if it minimizes the l.s.c functional

$$\bar{I}_c[\pi] = \int c\left(x, \frac{d\pi_1^{ac}}{d\mu}(x)\pi_x\right) d\mu(x) + \int c'_\infty(x,\pi_x)\, d\pi_1^s(x)$$

among couplings with second marginal ν.

Moreover, the following Kantorovich type dual formula holds

$$\mathcal{I}_c(\mu,\nu) = \sup_{f \in \mathcal{C}_b(\mathcal{Y})} \left\{ \int K_c f\, d\mu - \int f\, d\nu \right\}, \qquad \mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$$

with

$$K_c f(x) = \inf_{m \in \mathcal{M}(\mathcal{Y})} \left\{ \int f\, dm + c(x,m) \right\}, \qquad x \in \mathcal{X}.$$

III - The particular case of barycentric and conical costs

Duality for barycentric transport costs (WOT)

Here $\mathcal{X} = \mathcal{Y} = \mathbb{R}^n$.

Theorem (G.-Roberto-Samson-Tetali, 2017)

Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^n)$ with finite first moments. Let $\theta : \mathbb{R}^n \to \mathbb{R}$ be a convex function and consider

$$\overline{\mathcal{T}}_\theta(\mu,\nu) = \inf_{p \in \mathcal{P}(\mu,\nu)} \int \theta\left(x - \int y \, p_x(dy)\right) \mu(dx)$$

Then,

$$\overline{\mathcal{T}}_\theta(\mu,\nu) = \sup_{\varphi} \left\{ \int Q_\theta \varphi \, d\mu - \int \varphi \, d\nu \right\},$$

where the supremum runs over the set of all convex functions bounded from below and

$$Q_\theta \varphi(x) = \inf_{y \in \mathbb{R}^n} \{\varphi(y) + \theta(x-y)\}, \qquad x \in \mathbb{R}^n.$$

These barycentric cost functions found several applications, in particular for obtaining dimension free concentration inequalities for convex functions. Also involved in a proof of the Caffarelli contraction theorem.

A proof of Strassen theorem using barycentric L_1 cost

Notation : $\mathcal{P}_1(\mathbb{R}^n)$ the set of probability measures with a finite first moment.

Definition

Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^n)$; μ is dominated by ν in the convex order, denoted by $\mu \leq_c \nu$, if

$$\int f \, d\mu \leq \int f \, d\nu, \qquad \text{for all convex function } f : \mathbb{R}^n \to \mathbb{R}.$$

Theorem (Strassen (1965))

Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^n)$; the following propositions are equivalent

(1) $\mu \leq_c \nu$,

(2) there exists a martingale (X_0, X_1) such that $X_0 \sim \mu$ and $X_1 \sim \nu$.

The implication (2) $\Rightarrow$ (1) comes from Jensen inequality.

A proof of Strassen theorem using barycentric L_1 cost

Let $\|\cdot\|$ be some norm on $\mathbb{R}^n$; consider

$$\overline{\mathcal{T}}_1(\mu,\nu) = \inf_{p \in \mathcal{P}(\mu,\nu)} \int \left\| x - \int y \, p_x(dy) \right\| \mu(dx)$$

A proof of Strassen theorem using barycentric L_1 cost

Let $\|\cdot\|$ be some norm on $\mathbb{R}^n$; consider

$$\begin{aligned}\overline{\mathcal{T}}_1(\mu,\nu) &= \inf_{p\in\mathcal{P}(\mu,\nu)} \int \left\| x - \int y\, p_x(dy) \right\| \mu(dx) \\ &= \inf_{(X_0,X_1),X_0\sim\mu,X_1\sim\nu} \mathbb{E}\left[\|X_0 - \mathbb{E}[X_1|X_0]\|\right].\end{aligned}$$

A proof of Strassen theorem using barycentric L_1 cost

Let $\|\cdot\|$ be some norm on $\mathbb{R}^n$; consider

$$\overline{\mathcal{T}}_1(\mu,\nu) = \inf_{p\in\mathcal{P}(\mu,\nu)} \int \left\| x - \int y\, p_x(dy) \right\| \mu(dx)$$
$$= \inf_{(X_0,X_1),X_0\sim\mu,X_1\sim\nu} \mathbb{E}\left[\|X_0 - \mathbb{E}[X_1|X_0]\|\right].$$

Therefore, $\overline{\mathcal{T}}_1(\mu,\nu) = 0$ if and only if there exists a martingale $(X_i)_{i\in\{0,1\}}$ with marginals μ and ν.

A proof of Strassen theorem using barycentric L_1 cost

Let $\|\cdot\|$ be some norm on $\mathbb{R}^n$; consider

$$\overline{\mathcal{T}}_1(\mu,\nu) = \inf_{p \in \mathcal{P}(\mu,\nu)} \int \left\| x - \int y \, p_x(dy) \right\| \mu(dx)$$
$$= \inf_{(X_0,X_1), X_0 \sim \mu, X_1 \sim \nu} \mathbb{E}\left[\| X_0 - \mathbb{E}[X_1 | X_0] \| \right].$$

Therefore, $\overline{\mathcal{T}}_1(\mu,\nu) = 0$ if and only if there exists a martingale $(X_i)_{i \in \{0,1\}}$ with marginals μ and ν.

For the cost $\overline{\mathcal{T}}_1$ the duality specializes to

$$\overline{\mathcal{T}}_1(\mu,\nu) = \sup_{\varphi} \left\{ \int \varphi \, d\mu - \int \varphi \, d\nu \right\},$$

where the supremum runs over the set of all 1-Lipschitz and convex functions.

A proof of Strassen theorem using barycentric L_1 cost

Let $\|\cdot\|$ be some norm on $\mathbb{R}^n$; consider

$$\overline{\mathcal{T}}_1(\mu,\nu) = \inf_{p\in\mathcal{P}(\mu,\nu)} \int \left\| x - \int y \, p_x(dy) \right\| \mu(dx)$$

$$= \inf_{(X_0,X_1), X_0\sim\mu, X_1\sim\nu} \mathbb{E}\left[\|X_0 - \mathbb{E}[X_1|X_0]\|\right].$$

Therefore, $\overline{\mathcal{T}}_1(\mu,\nu) = 0$ if and only if there exists a martingale $(X_i)_{i\in\{0,1\}}$ with marginals μ and ν.

For the cost $\overline{\mathcal{T}}_1$ the duality specializes to

$$\overline{\mathcal{T}}_1(\mu,\nu) = \sup_{\varphi}\left\{\int \varphi \, d\mu - \int \varphi \, d\nu\right\},$$

where the supremum runs over the set of all 1-Lipschitz and convex functions.

Thus, if $\mu \leq_c \nu$, then

$$\overline{\mathcal{T}}_1(\mu,\nu) = \sup_{\varphi}\left\{\int \varphi \, d\mu - \int \varphi \, d\nu\right\} = 0$$

and so there exists a martingale (X_0, X_1) with marginals μ and ν.

Duality for conical cost functions (WOTUK)

Here $\mathcal{X}$ is a compact metric space, $\mathcal{Y}$ is a compact subset of $\mathbb{R}^n$ and $\mathcal{Z}$ is the conical convex hull of $\mathcal{Y}$. Consider a cost function of the form

$$c(x,m) = F\left(x, \int y\, dm(y)\right), \qquad x \in \mathcal{X}, \qquad m \in \mathcal{M}(\mathcal{Y}),$$

with $F : \mathcal{X} \times \mathcal{Z} \to \mathbb{R}$.

Theorem (CGK, 2023)

Let $\mu \in \mathcal{P}(\mathcal{X})$ and assume that c satisfies Assumption (A).
If c also satisfies assumption (C) or if the convex hull of $\mathcal{Y}$ does not contain 0, then for any probability measure $\nu \in \mathcal{P}(\mathcal{Y})$, it holds

$$\mathcal{I}_c(\mu,\nu) = \sup_{\varphi \in \Phi(\mathcal{Z}) \cap L^1(\nu)} \left\{ \int Q_F\varphi(x)\, \mu(dx) - \int \varphi(y)\, \nu(dy) \right\},$$

where $\Phi(\mathcal{Z})$ is the set of all lower semicontinuous, convex positively 1-homogenous functions $\varphi : \mathcal{Z} \to \mathbb{R} \cup \{+\infty\}$ and where

$$Q_F\varphi(x) = \inf_{z \in \mathcal{Z}} \{\varphi(z) + F(x,z)\}, \qquad x \in \mathcal{X}.$$

Moreover, there is dual attainment.

Interpretation in economy

Under the assumptions of the preceding theorem, and assumting that $F(x,z) = -G(x,z)$ with $G(x,z) \geq 0$ and concave in z, then

$$-\mathcal{I}_c(\mu,\nu) = \inf_{\varphi \in \Phi^+(\mathcal{Z}) \cap L^1(\nu)} \left\{ \int \sup_{z \in \mathcal{Z}} \{G(x,z) - \varphi(z)\} \, \mu(dx) + \int \varphi(y) \, \nu(dy) \right\},$$

where $\Phi^+(\mathcal{Z})$ is the set of all lower semicontinuous, convex positively 1-homogenous functions $\varphi : \mathcal{Z} \to \mathbb{R}^+ \cup \{+\infty\}$.

Functions $\varphi \in \Phi^+(\mathcal{Z})$ are interpreted as wages.

For a given wage φ,

$$\sup_{z \in \mathcal{Z}} \{G(x,z) - \varphi(z)\}$$

is the maximal output net of wage a firm of type x can obtain.

A variant of Strassen Theorem

Definition

If μ, ν are two probability measures with a finite moment of order 1, we will say that μ is dominated by ν for the positively 1-homogenous convex order if for all $\varphi : \mathbb{R}^d \to \mathbb{R}$ convex and positively 1-homogenous, one has $\int \varphi \, d\mu \leq \int \varphi \, d\nu$.

We will use the notation $\mu \leq_{phc} \nu$ to denote this order.

Theorem (CGK, 2023)

Let μ, ν be two compactly supported probability measures on $\mathbb{R}^d$.
Then the following are equivalent :

(i) $\mu \leq_{phc} \nu$,

(ii) There exists a nonnegative kernel q such that $\mu q = \nu$ and

$$\int y \, q^x(dy) = x$$

for μ almost every x.

Structure of solutions for conical costs

Theorem (CGK, 2023)

Let $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$ be such that $\mathcal{I}_c(\mu,\nu) < +\infty$, and assume that the convex hull of the support of ν does not contain 0. Then, the following identity holds

$$\mathcal{I}_c(\mu,\nu) = \inf_{\gamma \leq_{phc} \nu} \mathcal{T}_F(\mu,\gamma), \tag{1}$$

where and $\mathcal{T}_F$ denotes the classical transport cost associated to the cost function F :

$$\mathcal{T}_F(\mu,\gamma) = \inf_{\pi \in \Pi(\mu,\gamma)} \iint F(x,z)\,\pi(dxdz), \qquad \forall \mu \in \mathcal{P}(\mathcal{X}), \forall \gamma \in \mathcal{P}(\mathcal{Z}).$$

Moreover, suppose that $\bar{q} \in \mathcal{Q}(\mu,\nu)$ is a solution for $\mathcal{I}_c(\mu,\nu)$, consider the map $\bar{S}$ defined by

$$\bar{S}(x) = \int y\,\bar{q}^x(dy), \qquad x \in \mathcal{X},$$

and denote by $\bar{\nu}$ the image of μ under the map $\bar{S}$. Then the following holds :

- the probability measure $\bar{\nu}$ is dominated by ν in the positively 1-homogenous convex order,
- one has that

$$\mathcal{I}_c(\mu,\nu) = \int F(x,\bar{S}(x))\,\mu(dx) = \inf_{\gamma \leq_{phc} \nu} \mathcal{T}_F(\mu,\gamma)$$

IV - Perspectives

Some open questions

- Understand better cyclical monotonicity.
- When is the matching deterministic ?
- What is the characterization of stochastic orders associated to convex positively homogeneous functions of degree k ? Is there a general version of Strassen theorem applying for such classes of functions ?
- Are there applications of transport cost $\mathcal{I}_c$ in functional/concentration inequalities ?

Thank you for your attention !