A spatial Pareto exchange economy problem

Xavier Bacon

Joint work with Guillaume Carlier and Bruno Nazaret

Université Paris 1 Panthéon-Sorbonne

June 17th 2022

Table of Contents

Primal problem

A B C

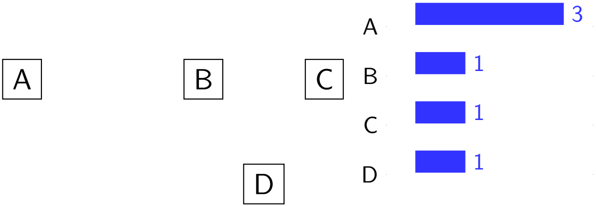
D

- X is a region which we assume to be a compact metric space.

Primal problem

$$(\mathcal{P}) = \quad \sup \quad \left\{ \qquad \right\}$$

Primal problem

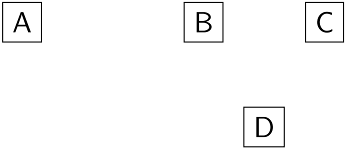


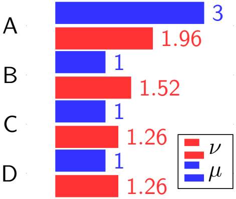
- X is a region which we assume to be a compact metric space.
- $\mu \in \mathcal{M}_+(X)$ is the initial distribution of (say) sugar in region X.

Primal problem

$$(\mathcal{P}) = \quad \sup \quad \left\{ \qquad \qquad \qquad \qquad \right\}$$

Primal problem



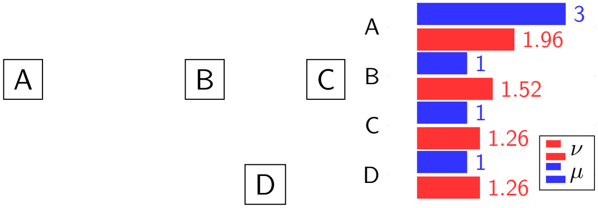


- X is a region which we assume to be a compact metric space.
- $\mu \in \mathcal{M}_+(X)$ is the initial distribution of (say) sugar in region X.
- $\nu \in \mathcal{M}_+(X)$ is the final distribution of sugar in region X.

Primal problem

$$(\mathcal{P}) = \sup_{\nu \in \mathcal{M}_+(X)} \left\{ \qquad \text{subject to } \mu(X) = \nu(X) \right\}$$

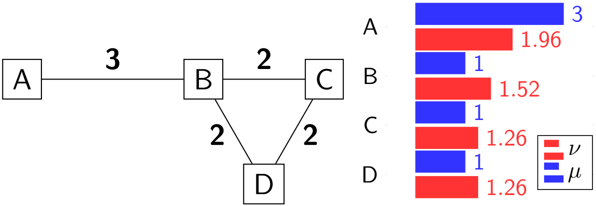
Primal problem



- X is a region which we assume to be a compact metric space.
- $\mu \in \mathcal{M}_+(X)$ is the initial distribution of (say) sugar in region X.
- $\nu \in \mathcal{M}_+(X)$ is the final distribution of sugar in region X.
- $\mathcal{U} = \mathcal{U}(\nu)$ is the average utility.

Primal problem

$$(\mathcal{P}) = \sup_{\nu \in \mathcal{M}_+(X)} \left\{ \mathcal{U}(\nu) \qquad \text{subject to } \mu(X) = \nu(X) \right\}$$



- X is a region which we assume to be a compact metric space.
- $\mu \in \mathcal{M}_+(X)$ is the initial distribution of (say) sugar in region X.
- $\nu \in \mathcal{M}_+(X)$ is the final distribution of sugar in region X.
- $\mathcal{U} = \mathcal{U}(\nu)$ is the average utility.
- $\mathcal{T}(\mu, .) = \mathcal{T}(\mu, \nu)$ is the transport cost between μ and ν.

Primal problem

$$(\mathcal{P}) = \sup_{\nu \in \mathcal{M}_+(X)} \left\{ \mathcal{U}(\nu) - \mathcal{T}(\mu, \nu) \text{ subject to } \mu(X) = \nu(X) \right\}$$

Utility

The average utility $\mathcal{U}$ is given by

$$\mathcal{U}(\nu) = \int_X U(x, \nu(x)) \,\mathrm{d}m(x)$$

where

- m is a reference measure *i.e.* $\mu \ll m$.
- $U : (x, \nu) \in X \times \mathbb{R}_+ \mapsto U(x, \nu) \in \mathbb{R}$ *is the preference of the agent located in* x*.*

Utility

The average utility $\mathcal{U}$ is given by

$$\mathcal{U}(\nu) = \int_X U(x, \nu(x))\,\mathrm{d}m(x)$$

where

- m is a reference measure *i.e.* $\mu \ll m$.
- $U : (x, \nu) \in X \times \mathbb{R}_+ \mapsto U(x, \nu) \in \mathbb{R}$ *is the preference of the agent located in* x*.*

Technical assumptions on U

1. for m-a.e. $x \in X$, $\nu \in \mathbb{R}_+ \mapsto U(x, \nu)$ is upper semi-continuous, concave, nondecreasing.
2. for every $\nu \in \mathbb{R}_+$, $x \in X \mapsto U(x, \nu)$ is m-measurable.
3. $(x, \nu) \mapsto U(x, \nu)$ is sublinear with respect to ν uniformly in $x \in X$.

What is $\mathcal{T}(\mu, \nu)$?

- $\mu \in \mathcal{M}_+(\mathcal{X}), \nu \in \mathcal{M}_+(\mathcal{Y})$ are respectively the source and the target distributions.

What is $\mathcal{T}(\mu, \nu)$?

- $\mu \in \mathcal{M}_+(\mathcal{X}), \nu \in \mathcal{M}_+(\mathcal{Y})$ are respectively the source and the target distributions.
- $c = c(x, y)$ is the transport cost satifying for all $x, y \in X$,

$$c(x, y) \geqslant 0 \text{ and } c(x, x) = 0.$$

What is $\mathcal{T}(\mu, \nu)$?

- $\mu \in \mathcal{M}_+(\mathcal{X}), \nu \in \mathcal{M}_+(\mathcal{Y})$ are respectively the source and the target distributions.
- $c = c(x, y)$ is the transport cost satifying for all $x, y \in X$,

$$c(x, y) \geqslant 0 \text{ and } c(x, x) = 0.$$

Transference plan and related cost

$\gamma \in \mathcal{M}_+(X \times Y)$ is a **transference plan between** μ **and** ν if

$$\int_Y \mathrm{d}\gamma(x, y) = \mu(x) \quad \text{and} \quad \int_X \mathrm{d}\gamma(x, y) = \nu(y)$$

Its related cost with respect to c is given by

$$\langle c \mid \gamma \rangle \overset{\text{def.}}{=} \iint_{X \times Y} c(x, y) \, \mathrm{d}\gamma(x, y) \in [0, \infty].$$

Optimal Transport

$\Pi(\mu, \nu)$ denotes the set of all transference plans between μ and ν.

Optimal Transport

$\Pi(\mu, \nu)$ denotes the set of all transference plans between μ and ν.

Optimal Transport problem

The optimal transport cost between μ and ν is given by

$$\mathcal{T}(\mu, \nu) = \inf \left\{ \langle c \mid \gamma \rangle : \gamma \in \Pi(\mu, \nu) \right\}.$$

$\Pi(\mu, \nu)$ denotes the set of all transference plans between μ and ν.

Optimal Transport problem

The optimal transport cost between μ and ν is given by

$$\mathcal{T}(\mu, \nu) = \inf \{\langle c \mid \gamma \rangle : \gamma \in \Pi(\mu, \nu)\}.$$

Remark : Marginal constraints imply

$$\mu(X) = \int_X \mathrm{d}\mu(x) = \iint_{X \times Y} \mathrm{d}\gamma(x, y) = \int_Y \mathrm{d}\nu(y) = \nu(Y)$$

and then $\mathcal{T}(\mu, \nu) = \infty$ if $\mu(X) \neq \nu(Y)$ ($\inf_{\varnothing} = \infty$).

Summarizing table

	Unregularized	
Primal	$\overbrace{\sup_{\nu} \mathcal{U}(\nu) - \mathcal{T}(\mu,\nu)}^{(\mathcal{P})}$	

Table of Contents

Duality in Optimal Transport

Marginal constraints admit the following *dual* formulation :

$$\int_Y \mathrm{d}\gamma(x,y) = \mu(x) \iff \forall \varphi \in C(X), \iint_{X\times Y} \varphi(x)\,\mathrm{d}\gamma(x,y) = \int_X \varphi(x)\,\mathrm{d}\mu(x),$$

$$\int_X \mathrm{d}\gamma(x,y) = \nu(y) \iff \forall \psi \in C(Y), \iint_{X\times Y} \psi(y)\,\mathrm{d}\gamma(x,y) = \int_X \psi(y)\,\mathrm{d}\nu(y),$$

Duality in Optimal Transport

Marginal constraints admit the following *dual* formulation :

$$\int_Y \mathrm{d}\gamma(x,y) = \mu(x) \iff \forall \varphi \in C(X), \iint_{X\times Y} \varphi(x)\,\mathrm{d}\gamma(x,y) = \int_X \varphi(x)\,\mathrm{d}\mu(x),$$

$$\int_X \mathrm{d}\gamma(x,y) = \nu(y) \iff \forall \psi \in C(Y), \iint_{X\times Y} \psi(y)\,\mathrm{d}\gamma(x,y) = \int_X \psi(y)\,\mathrm{d}\nu(y),$$

and then

$$\sup_{\varphi,\psi} \int_X \varphi\,\mathrm{d}\mu + \int_Y \psi\,\mathrm{d}\nu - \iint_{X\times Y} \underbrace{\varphi(x)+\psi(y)}_{\varphi\oplus\psi(x,y)}\,\mathrm{d}\gamma$$

Duality in Optimal Transport

Marginal constraints admit the following *dual* formulation :

$$\int_Y \mathrm{d}\gamma(x,y) = \mu(x) \iff \forall \varphi \in C(X), \iint_{X\times Y} \varphi(x)\,\mathrm{d}\gamma(x,y) = \int_X \varphi(x)\,\mathrm{d}\mu(x),$$

$$\int_X \mathrm{d}\gamma(x,y) = \nu(y) \iff \forall \psi \in C(Y), \iint_{X\times Y} \psi(y)\,\mathrm{d}\gamma(x,y) = \int_X \psi(y)\,\mathrm{d}\nu(y),$$

and then

$$\sup_{\varphi,\psi} \int_X \varphi\,\mathrm{d}\mu + \int_Y \psi\,\mathrm{d}\nu - \iint_{X\times Y} \underbrace{\varphi(x)+\psi(y)}_{\varphi\oplus\psi(x,y)}\,\mathrm{d}\gamma = \begin{cases} 0 \text{ if } \gamma \in \Pi(\mu,\nu) \\ +\infty \text{ otherwise,} \end{cases}$$

Duality in Optimal Transport

Marginal constraints admit the following *dual* formulation :

$$\int_Y \mathrm{d}\gamma(x,y) = \mu(x) \iff \forall \varphi \in C(X), \iint_{X\times Y} \varphi(x)\,\mathrm{d}\gamma(x,y) = \int_X \varphi(x)\,\mathrm{d}\mu(x),$$

$$\int_X \mathrm{d}\gamma(x,y) = \nu(y) \iff \forall \psi \in C(Y), \iint_{X\times Y} \psi(y)\,\mathrm{d}\gamma(x,y) = \int_X \psi(y)\,\mathrm{d}\nu(y),$$

and then

$$\sup_{\varphi,\psi} \int_X \varphi\,\mathrm{d}\mu + \int_Y \psi\,\mathrm{d}\nu - \iint_{X\times Y} \underbrace{\varphi(x)+\psi(y)}_{\varphi\oplus\psi(x,y)}\,\mathrm{d}\gamma = \begin{cases} 0 \text{ if } \gamma \in \Pi(\mu,\nu) \\ +\infty \text{ otherwise,} \end{cases}$$

and the Optimal Transport problem can be rewritten as follows

Duality in Optimal Transport

$$\mathcal{T}(\mu,\nu) = \inf\left\{\langle c \mid \gamma\rangle : \gamma \in \Pi(\mu,\nu)\right\}$$

Duality in Optimal Transport

$$\begin{aligned}\mathcal{T}(\mu,\nu) &= \inf\left\{\langle c \mid \gamma\rangle : \gamma \in \Pi(\mu,\nu)\right\} \\ &= \inf_{\gamma}\,\langle c \mid \gamma\rangle + \sup_{\varphi,\psi}\left(\langle \varphi \mid \mu\rangle + \langle \psi \mid \nu\rangle - \langle \varphi \oplus \psi \mid \gamma\rangle\right)\end{aligned}$$

Duality in Optimal Transport

$$\begin{aligned}\mathcal{T}(\mu,\nu) &= \inf\left\{\langle c \mid \gamma\rangle : \gamma \in \Pi(\mu,\nu)\right\} \\ &= \inf_{\gamma}\, \langle c \mid \gamma\rangle + \sup_{\varphi,\psi}\left(\langle \varphi \mid \mu\rangle + \langle \psi \mid \nu\rangle - \langle \varphi \oplus \psi \mid \gamma\rangle\right) \\ &= \inf_{\gamma}\,\sup_{\varphi,\psi} \underbrace{\langle \varphi \mid \mu\rangle + \langle \psi \mid \nu\rangle + \langle c - \varphi \oplus \psi \mid \gamma\rangle}_{\text{Lagrangian}(\gamma,\varphi,\psi)}\end{aligned}$$

Duality in Optimal Transport

$$\begin{aligned}
\mathcal{T}(\mu,\nu) &= \inf\left\{\langle c \mid \gamma\rangle : \gamma \in \Pi(\mu,\nu)\right\} \\
&= \inf_{\gamma} \langle c \mid \gamma\rangle + \sup_{\varphi,\psi}\left(\langle \varphi \mid \mu\rangle + \langle \psi \mid \nu\rangle - \langle \varphi \oplus \psi \mid \gamma\rangle\right) \\
&= \inf_{\gamma}\sup_{\varphi,\psi} \underbrace{\langle \varphi \mid \mu\rangle + \langle \psi \mid \nu\rangle + \langle c - \varphi \oplus \psi \mid \gamma\rangle}_{\text{Lagrangian}(\gamma,\varphi,\psi)} \\
\text{(no duality gap)} &= \sup_{\varphi,\psi} \langle \varphi \mid \mu\rangle + \langle \psi \mid \nu\rangle + \underbrace{\inf_{\gamma}\langle c - \varphi \oplus \psi \mid \gamma\rangle}_{= \begin{cases} 0 \text{ if } \varphi \oplus \psi \leqslant c \\ -\infty \text{ otherwise}\end{cases}}
\end{aligned}$$

Duality in Optimal Transport

$$\mathcal{T}(\mu,\nu) = \inf\{\langle c \mid \gamma\rangle : \gamma \in \Pi(\mu,\nu)\}$$

$$= \inf_{\gamma} \langle c \mid \gamma\rangle + \sup_{\varphi,\psi} \left(\langle \varphi \mid \mu\rangle + \langle \psi \mid \nu\rangle - \langle \varphi \oplus \psi \mid \gamma\rangle\right)$$

$$= \inf_{\gamma} \sup_{\varphi,\psi} \underbrace{\langle \varphi \mid \mu\rangle + \langle \psi \mid \nu\rangle + \langle c - \varphi \oplus \psi \mid \gamma\rangle}_{\text{Lagrangian}(\gamma,\varphi,\psi)}$$

(no duality gap)
$$= \sup_{\varphi,\psi} \langle \varphi \mid \mu\rangle + \langle \psi \mid \nu\rangle + \underbrace{\inf_{\gamma} \langle c - \varphi \oplus \psi \mid \gamma\rangle}_{= \begin{cases} 0 \text{ if } \varphi \oplus \psi \leqslant c \\ -\infty \text{ otherwise} \end{cases}}$$

$$= \sup_{\varphi \oplus \psi \leqslant c} \langle \varphi \mid \mu\rangle + \langle \psi \mid \nu\rangle$$

$$
\begin{aligned}
\mathcal{T}(\mu,\nu) &= \inf\{\langle c \mid \gamma\rangle : \gamma \in \Pi(\mu,\nu)\} \\
&= \inf_{\gamma}\langle c \mid \gamma\rangle + \sup_{\varphi,\psi}\left(\langle\varphi \mid \mu\rangle + \langle\psi \mid \nu\rangle - \langle\varphi\oplus\psi \mid \gamma\rangle\right) \\
&= \inf_{\gamma}\sup_{\varphi,\psi} \underbrace{\langle\varphi \mid \mu\rangle + \langle\psi \mid \nu\rangle + \langle c - \varphi\oplus\psi \mid \gamma\rangle}_{\text{Lagrangian}(\gamma,\varphi,\psi)} \\
\text{(no duality gap)} &= \sup_{\varphi,\psi}\langle\varphi \mid \mu\rangle + \langle\psi \mid \nu\rangle + \underbrace{\inf_{\gamma}\langle c - \varphi\oplus\psi \mid \gamma\rangle}_{=\begin{cases} 0 \text{ if } \varphi\oplus\psi \leqslant c \\ -\infty \text{ otherwise}\end{cases}} \\
&= \sup_{\varphi\oplus\psi\leqslant c}\langle\varphi \mid \mu\rangle + \langle\psi \mid \nu\rangle \\
&\overset{\text{def.}}{=} \text{Dual of OT}(\mu,\nu)
\end{aligned}
$$

Duality in Optimal Transport

Remark : constraints on φ and ψ are given by

$$\begin{aligned} & \varphi(x) + \psi(y) \leqslant c(x,y) \\ \iff \quad & \psi(y) \leqslant c(x,y) - \varphi(x) \end{aligned}$$

Duality in Optimal Transport

Remark : constraints on φ and ψ are given by

$$\begin{aligned} & \varphi(x) + \psi(y) \leqslant c(x,y) \\ \iff \quad & \psi(y) \leqslant c(x,y) - \varphi(x) \end{aligned}$$

and then substitute ψ with $\min_x \; c(x,y) - \varphi(x) \overset{\text{def.}}{=} \varphi^c(y)$ improves dual cost.

Remark : constraints on φ and ψ are given by

$$\begin{aligned} & \varphi(x) + \psi(y) \leqslant c(x,y) \\ \iff \quad & \psi(y) \leqslant c(x,y) - \varphi(x) \end{aligned}$$

and then substitute ψ with $\min_x \; c(x,y) - \varphi(x) \stackrel{\text{def.}}{=} \varphi^c(y)$ improves dual cost. As a consequence dual problem of Optimal Transport can be rewritten as follows

$$\text{Dual of OT}(\mu,\nu) = \sup_{\varphi} \; \langle \varphi \mid \mu \rangle + \langle \varphi^c \mid \nu \rangle.$$

Duality for Pareto problem

Dual problem of $(\mathcal{P})$

$$(\mathcal{D}) = \inf_{\varphi \in C(X)} \mathcal{K}(\varphi) + \mathcal{V}(\varphi)$$

Duality for Pareto problem

- $\mathcal{K}(\varphi) = -\int_{\mathcal{X}} \varphi^{c} \, \mathrm{d}\mu.$

Dual problem of $(\mathcal{P})$

$$(\mathcal{D}) = \inf_{\varphi \in C(X)} \mathcal{K}(\varphi) + \mathcal{V}(\varphi)$$

Duality for Pareto problem

- $\mathcal{K}(\varphi) = -\int_{\mathcal{X}} \varphi^{c} \, \mathrm{d}\mu$.
- $\mathcal{V}(\varphi) = \int_X V(x, \varphi(x)) \, \mathrm{d}m(x)$ where

$$V(x,\varphi) = \underbrace{\sup_{\nu \in \mathbb{R}_+} U(x,\nu) - \nu\varphi}_{\text{Fenchel conjugate at } -\varphi \text{ of } -\tilde{U}(x,.)}$$

and $\tilde{U}(x,.)$ is the concave $u.s.c.$ extension of $U(x,.)$.

Dual problem of $(\mathcal{P})$ and no duality gap

$$(\mathcal{D}) = \inf_{\varphi \in C(X)} \mathcal{K}(\varphi) + \mathcal{V}(\varphi)$$

Duality for Pareto problem

- $\mathcal{K}(\varphi) = -\int_{\mathcal{X}} \varphi^c \, \mathrm{d}\mu$.
- $\mathcal{V}(\varphi) = \int_X V(x, \varphi(x)) \, \mathrm{d}m(x)$ where

$$V(x,\varphi) = \underbrace{\sup_{\nu \in \mathbb{R}_+} U(x,\nu) - \nu\varphi}_{\text{Fenchel conjugate at } -\varphi \text{ of } -\tilde{U}(x,.)}$$

and $\tilde{U}(x,.)$ is the concave $u.s.c.$ extension of $U(x,.)$.

Dual problem of $(\mathcal{P})$ and no duality gap

$$(\mathcal{D}) = \inf_{\varphi \in C(X)} \mathcal{K}(\varphi) + \mathcal{V}(\varphi) = (\mathcal{P})$$

Summarizing table

	Unregularized	
Primal	$\overbrace{\sup_{\nu} \mathcal{U}(\nu) - \mathcal{T}(\mu, \nu)}^{(\mathcal{P})}$	
Dual	$\overbrace{\inf_{\varphi} \mathcal{K}(\varphi) + \mathcal{V}(\varphi)}^{(\mathcal{D})}$	

Why do we use duality techniques?

Why do we use duality techniques?

ν optimal in $(\mathcal{P})$ and φ optimal in $(\mathcal{D})$

$$\iff$$

$$V(x, \varphi(x)) = U(x, \nu(x)) - \varphi(x)\nu(x) \text{ for } m\text{-a.e. } x.$$

Why do we use duality techniques?

ν optimal in $(\mathcal{P})$ and φ optimal in $(\mathcal{D})$

$$\iff$$

$$V(x, \varphi(x)) = U(x, \nu(x)) - \varphi(x)\nu(x) \text{ for } m\text{-a.e. } x.$$

- As a consequence, if we find φ optimal, ν optimal is (the unique) solution of the previous equation.

Table of Contents

Entropic regularization and Sinkhorn descent

- First applied in Optimal Transport by M. Cuturi (2013) :

Entropic regularization and Sinkhorn descent

- First applied in Optimal Transport by M. Cuturi (2013) : for a fixed $\varepsilon > 0$ he introduced the regularized OT problem

$$\mathcal{T}_\varepsilon(\mu, \nu) = \inf_{\gamma \in \Pi(\mu,\nu)} \langle c \mid \gamma \rangle + \varepsilon E(\gamma)$$

Entropic regularization and Sinkhorn descent

- First applied in Optimal Transport by M. Cuturi (2013) : for a fixed $\varepsilon > 0$ he introduced the regularized OT problem

$$\mathcal{T}_\varepsilon(\mu,\nu) = \inf_{\gamma\in\Pi(\mu,\nu)} \langle c \mid \gamma\rangle + \varepsilon E(\gamma)$$

whose unique solution converges to the solution of $\mathcal{T}(\mu,\nu)$ with maximal entropy (Cominetti, San Martin, 1994).

Entropic regularization and Sinkhorn descent

- First applied in Optimal Transport by M. Cuturi (2013) : for a fixed $\varepsilon > 0$ he introduced the regularized OT problem

$$\mathcal{T}_\varepsilon(\mu,\nu) = \inf_{\gamma\in\Pi(\mu,\nu)} \langle c \mid \gamma\rangle + \varepsilon E(\gamma)$$

  whose unique solution converges to the solution of $\mathcal{T}(\mu,\nu)$ with maximal entropy (Cominetti, San Martin, 1994).
- G. Peyré extends it in the context of Wasserstein gradient flows in 2015.

Summarizing table

	Unregularized	ε-regularized
Primal	$\overbrace{\sup_{\nu} \mathcal{U}(\nu) - \mathcal{T}(\mu, \nu)}^{(\mathcal{P})}$	$\overbrace{\sup_{\nu} \mathcal{U}(\nu) - \mathcal{T}_\varepsilon(\mu, \nu)}^{(\mathcal{P}_\varepsilon)}$
Dual	$\overbrace{\inf_{\varphi} \mathcal{K}(\varphi) + \mathcal{V}(\varphi)}^{(\mathcal{D})}$	

Entropic regularization

- X is finite and m is the counting measure on X.

Entropic regularization

- X is finite and m is the counting measure on X.

$$\Rightarrow (\mathcal{D}) = \inf_{\varphi \in \mathbb{R}^X} \mathcal{V}(\varphi) + \sum_{x \in X} \mu(x) \underbrace{\max_{y \in X} [\varphi(y) - c(x,y)]}_{-\varphi^c}$$

Entropic regularization

- X is finite and m is the counting measure on X.

$$\Rightarrow (\mathcal{D}) = \inf_{\varphi \in \mathbb{R}^X} \mathcal{V}(\varphi) + \sum_{x \in X} \mu(x) \underbrace{\max_{y \in X} [\varphi(y) - c(x,y)]}_{-\varphi^c}$$

Let $\varepsilon > 0$ be a regularization parameter and consider the smooth approximation of $(\mathcal{D})$ where max are replaced by soft maxima :

$$(\mathcal{D}_\varepsilon) \stackrel{\text{def.}}{=} \inf_{\varphi \in \mathbb{R}^X} \mathcal{V}(\varphi) + \sum_{x \in X} \mu(x) \operatorname{softmax}_\varepsilon(\varphi - c(x, \cdot))$$

$$= \inf_{\varphi \in \mathbb{R}^X} \mathcal{V}(\varphi) + \varepsilon \sum_{x \in X} \mu(x) \log \sum_{y \in X} \left(\exp \left[\frac{\varphi(y) - c(x,y)}{\varepsilon} \right] \right)$$

	Unregularized	ε-regularized
Primal	$\overbrace{\sup_{\nu} \mathcal{U}(\nu) - \mathcal{T}(\mu, \nu)}^{(\mathcal{P})}$	$\overbrace{\sup_{\nu} \mathcal{U}(\nu) - \mathcal{T}_{\varepsilon}(\mu, \nu)}^{(\mathcal{P}_{\varepsilon})}$
Dual	$\overbrace{\inf_{\varphi} \mathcal{K}(\varphi) + \mathcal{V}(\varphi)}^{(\mathcal{D})}$	$\overbrace{\inf_{\varphi} \mathcal{V}(\varphi) + \sum_{x \in X} \mu(x)\, \mathrm{softmax}_{\varepsilon}(\varphi - c(x, \cdot))}^{(\mathcal{D}_{\varepsilon})}$

$(\mathcal{D}_\varepsilon) \stackrel{\text{def.}}{=} \inf_{\varphi} D_\varepsilon(\varphi)$ can be reformulated by considering the convex function

$$\Phi(\varphi, \psi) = \mathcal{V}(\varphi) - \langle \mu \mid \psi \rangle + \varepsilon \sum_{x \in X} \mu(x) \log \sum_{(x,y) \in X^2} \left(\exp \left[\frac{\psi(x) + \varphi(y) - c(x,y)}{\varepsilon} \right] \right)$$

$(\mathcal{D}_\varepsilon) \stackrel{\text{def.}}{=} \inf_\varphi D_\varepsilon(\varphi)$ can be reformulated by considering the convex function

$$\Phi(\varphi,\psi) = \mathcal{V}(\varphi) - \langle \mu \mid \psi \rangle + \varepsilon \sum_{x\in X} \mu(x) \log \sum_{(x,y)\in X^2} \left(\exp\left[\frac{\psi(x) + \varphi(y) - c(x,y)}{\varepsilon} \right] \right)$$

since for fixed φ, the minimizer of $\psi \mapsto \Phi(\varphi,\psi)$ is explicitly given by

$$\psi(x) = \varepsilon \log(\mu(x)) - \varepsilon \log \left(\sum_{y\in X} e^{\frac{\varphi(y)-c(x,y)}{\varepsilon}} \right)$$

$(\mathcal{D}_\varepsilon) \stackrel{\text{def.}}{=} \inf_{\varphi} D_\varepsilon(\varphi)$ can be reformulated by considering the convex function

$$\Phi(\varphi, \psi) = \mathcal{V}(\varphi) - \langle \mu \mid \psi \rangle + \varepsilon \sum_{x \in X} \mu(x) \log \sum_{(x,y) \in X^2} \left(\exp \left[\frac{\psi(x) + \varphi(y) - c(x,y)}{\varepsilon} \right] \right)$$

since for fixed φ, the minimizer of $\psi \mapsto \Phi(\varphi, \psi)$ is explicitly given by

$$\psi(x) = \varepsilon \log(\mu(x)) - \varepsilon \log \left(\sum_{y \in X} e^{\frac{\varphi(y) - c(x,y)}{\varepsilon}} \right)$$

so replacing in Φ, we get

$$\inf_{\psi} \Phi(\varphi, \psi) = D_\varepsilon(\varphi) + C$$

where $C = \varepsilon \sum_{x \in X} \mu(x)(1 - \log(\mu(x)))$.

As noted above, $(\mathcal{D}_\varepsilon)$ is equivalent to

$$\inf\left\{\Phi(\varphi,\psi) : \varphi,\psi \in \mathbb{R}^X\right\}$$

which can be solved by **coordinate/Sinkhorn descent** : starting from (ψ^0,φ^0), we recursively compute $(\psi^{k+1},\varphi^{k+1})$ by

$$\psi^{k+1} = \underset{\psi\in\mathbb{R}^X}{\text{argmin}}\ \Phi(\varphi^k,\psi)$$

and

$$\varphi^{k+1} = \underset{\varphi\in\mathbb{R}^X}{\text{argmin}}\ \Phi(\varphi,\psi^{k+1}).$$

Notice that first updates are explicitly given by

$$\psi^{k+1}(x) = \varepsilon\log(\mu(x)) - \varepsilon\log\left(\sum_{y\in X}\exp\left[\frac{\varphi^k(y)-c(x,y)}{\varepsilon}\right]\right).$$

Summarizing table

	Unregularized	ε-regularized
Primal	$\overbrace{\sup_{\nu} \mathcal{U}(\nu) - \mathcal{T}(\mu, \nu)}^{(\mathcal{P})}$	$\overbrace{\sup_{\nu} \mathcal{U}(\nu) - \mathcal{T}_{\varepsilon}(\mu, \nu)}^{(\mathcal{P}_{\varepsilon})}$
Dual	$\overbrace{\inf_{\varphi} \mathcal{K}(\varphi) + \mathcal{V}(\varphi)}^{(\mathcal{D})}$	$\overbrace{\inf_{\varphi} \mathcal{V}(\varphi) + \sum_{x \in X} \mu(x)\, \mathrm{softmax}_{\varepsilon}(\varphi - c(x, \cdot))}^{(\mathcal{D}_{\varepsilon})}$

Summarizing table

	Unregularized	ε-regularized
Primal	$\overbrace{\sup_{\nu} \mathcal{U}(\nu) - \mathcal{T}(\mu, \nu)}^{(\mathcal{P})}$	$\overbrace{\sup_{\nu} \mathcal{U}(\nu) - \mathcal{T}_\varepsilon(\mu, \nu)}^{(\mathcal{P}_\varepsilon)}$
Dual	$\overbrace{\inf_{\varphi} \mathcal{K}(\varphi) + \mathcal{V}(\varphi)}^{(\mathcal{D})}$	$\overbrace{\inf_{\varphi} \mathcal{V}(\varphi) + \sum_{x \in X} \mu(x)\, \mathrm{softmax}_\varepsilon(\varphi - c(x, \cdot))}^{(\mathcal{D}_\varepsilon)}$

Solving $\mathcal{D}_\varepsilon \iff$ Solving $\inf \Phi(\varphi, \psi)$

Summarizing table

	Unregularized	ε-regularized
Primal	$\overbrace{\sup_{\nu} \mathcal{U}(\nu) - \mathcal{T}(\mu,\nu)}^{(\mathcal{P})}$	$\overbrace{\sup_{\nu} \mathcal{U}(\nu) - \mathcal{T}_\varepsilon(\mu,\nu)}^{(\mathcal{P}_\varepsilon)}$
Dual	$\overbrace{\inf_{\varphi} \mathcal{K}(\varphi) + \mathcal{V}(\varphi)}^{(\mathcal{D})}$	$\overbrace{\inf_{\varphi} \mathcal{V}(\varphi) + \sum_{x\in X} \mu(x)\,\mathrm{softmax}_\varepsilon(\varphi - c(x,\cdot))}^{(\mathcal{D}_\varepsilon)}$

Solving $\mathcal{D}_\varepsilon \iff$ Solving $\inf \Phi(\varphi,\psi)$

$\rightarrow$ we solve it by coordinate descent

Summarizing table

	Unregularized	ε-regularized
Primal	$\overbrace{\sup_{\nu} \mathcal{U}(\nu) - \mathcal{T}(\mu,\nu)}^{(\mathcal{P})}$	$\overbrace{\sup_{\nu} \mathcal{U}(\nu) - \mathcal{T}_\varepsilon(\mu,\nu)}^{(\mathcal{P}_\varepsilon)}$
Dual	$\overbrace{\inf_{\varphi} \mathcal{K}(\varphi) + \mathcal{V}(\varphi)}^{(\mathcal{D})}$	$\overbrace{\inf_{\varphi} \mathcal{V}(\varphi) + \sum_{x \in X} \mu(x)\, \mathrm{softmax}_\varepsilon(\varphi - c(x,\cdot))}^{(\mathcal{D}_\varepsilon)}$

Solving $\mathcal{D}_\varepsilon \iff$ Solving $\inf \Phi(\varphi, \psi)$

$\rightarrow$ we solve it by coordinate descent

$\rightarrow$ it gives us the optimal regularized φ

Summarizing table

	Unregularized	ε-regularized
Primal	$\overbrace{\sup_{\nu} \mathcal{U}(\nu) - \mathcal{T}(\mu,\nu)}^{(\mathcal{P})}$	$\overbrace{\sup_{\nu} \mathcal{U}(\nu) - \mathcal{T}_\varepsilon(\mu,\nu)}^{(\mathcal{P}_\varepsilon)}$
Dual	$\overbrace{\inf_{\varphi} \mathcal{K}(\varphi) + \mathcal{V}(\varphi)}^{(\mathcal{D})}$	$\overbrace{\inf_{\varphi} \mathcal{V}(\varphi) + \sum_{x\in X} \mu(x)\, \mathrm{softmax}_\varepsilon(\varphi - c(x,\cdot))}^{(\mathcal{D}_\varepsilon)}$

Solving $\mathcal{D}_\varepsilon \iff$ Solving $\inf \Phi(\varphi, \psi)$

$\rightarrow$ we solve it by coordinate descent

$\rightarrow$ it gives us the optimal regularized φ

$\rightarrow$ we deduce the optimal regularized ν

Merci de votre attention.