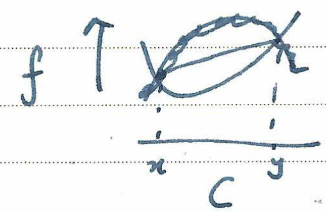
Deriving a dual problem

Kantorovich problem is actually convex.

Convex function: A function $f : \mathbb{R}^d \to \mathbb{R}$ is a convex function if on a convex set C, and $\forall\ x, y \in C$, $\lambda \in [0,1]$,

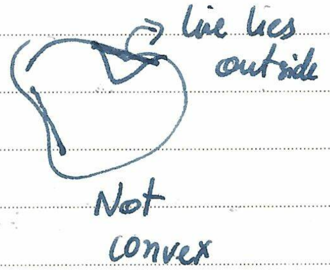
$$f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y)$$

That is,



the line segment joining $f(x)$ and $f(y)$ lies above the graph of the function

A convex set is a set that contains all the line segments joining any pair of points in it.



ie if $x, y \in C$ then

$$\lambda x + (1-\lambda) y \in C$$

for $\lambda \in [0,1]$.

Convex functions on convex feasibility sets form a class of well-studied optimization problems.

Convex optimization typically looks like this

$$\min_{x \in C} f(x)$$

where C is a convex set formed by convex constraints (e.g. $h(x) \leq 0$ for a convex h)

Importantly, a local minimum of f is also a global minimum

This holds even when C is infinite-dimensional.

In the KP $C = \Gamma$, the set of all joint distributions of X & Y with marginals μ & ν.

Ex: Show that Γ is a convex set

The objective function $\int c(x,y)\, d\gamma(x,y)$

is a linear function of γ and hence convex.

In order to prove that a solution to KP exists however, convexity alone is not enough! (in finite dimensions, it is enough)

but the cost function c being continuous in both arguments, $\mathbb{X}$ and $\mathbb{Y}$ being compact in $\mathbb{R}^d$ & μ being absolutely cont. are sufficient.

Typically, we write convex problems in their dual form when they may be mathematically or computationally easier. That is, instead of

$$\min_{x \in C} f(x) \qquad \text{(primal problem)}$$

we ~~write~~ consider a dual problem of the form

$$\max_{v \in \text{[illegible]}^*} -f^*(v) - \sup_{x \in C} \langle -v, x \rangle$$

Here, v is a vector in the dual space (also $\mathbb{R}^d$ in finite dimensions) and
f^* is a convex conjugate function

$$f^*(v) = \sup_y \langle v, y \rangle - f(y)$$

Since $f(x) + f^*(v) \geq \langle x, v \rangle$

(Fenchel - Young inequality)

You can think of f^* as an alternative characterization of the function f on the dual space of vectors.

This alternative parameterization is such that $\min_{x \in C} f(x) \geq \max_v -f^*(v) - \sup_{x \in C} \langle -v, x \rangle$

This is called weak duality: the primal problem has the greater minimizer or they are both equal

Now onto the dual of KP. The dual of the space of probability distributions is some space of functions

The pairing $\langle f, \mu \rangle = \int f \, d\mu$ $=: \mathop{\mathbb{E}}_{x \sim \mu} f(x)$ "expected value of $f(X)$ when $X \sim \mu$".

Now if you consider joint distributions $\gamma \in \Gamma$, then for any $\phi \in C(X)$ (cont functions on X)

$$\int \phi \, d\gamma = \int \phi \, d\mu$$

Similarly, $\int \psi \, d\gamma = \int \psi \, d\nu$ for any $\psi \in C(Y)$

So,
$$\inf_{\gamma \in \Gamma} \int c(x,y) \, d\gamma(x,y) + \sup_{\phi, \psi} \int \phi \, d\mu + \int \psi \, d\nu - \int \phi(x) + \psi(y) \, d\gamma_{(x,y)}$$

$$= \inf_{\gamma \in \Gamma} \int c(x,y) \, d\gamma(x,y)$$

Now onto the dual of KP. The dual of the space of probability distributions is some space of functions

The pairing $\langle f, \mu \rangle = \int f \, d\mu$
$=: \mathbb{E}_{x \sim \mu} f(x)$ "expected value of $f(X)$ when $X \sim \mu$".

Now if you consider joint distributions $\gamma \in \Gamma$, then for any $\phi \in C(X)$ ← cont functions on X

$$\int \phi \, d\gamma = \int \phi \, d\mu$$

Similarly, $\int \psi \, d\gamma = \int \psi \, d\nu$ for any $\psi \in C(Y)$

So,
$$\inf_{\gamma \in \Gamma} \int c(x,y) \, d\gamma(x,y) + \sup_{\phi, \psi} \int \phi \, d\mu + \int \psi \, d\nu - \int \phi(x) + \psi(y) \, d\gamma(x,y)$$

$$= \inf_{\gamma \in \Gamma} \int c(x,y) \, d\gamma(x,y)$$

Rearranging the objective function

$$\sup_{\phi, \psi} \left[\int \phi \, d\mu + \int \psi \, d\nu + \inf_{\gamma \in \Gamma} \int c - \phi \oplus \psi \, d\gamma \right]$$

where $\phi \oplus \psi (x, y) = \phi(x) + \psi(y)$

Note that $\phi \oplus \psi$ is a function on $\mathcal{X} \times \mathcal{Y}$ to $\mathbb{R}$

for any $\gamma \in \Gamma$, if $\phi \oplus \psi \leq c$ pointwise, ~~then~~ on $\mathcal{X} \times \mathcal{Y}$, then

$$\inf_{\gamma \in \Gamma} \int c - \phi \oplus \psi \, d\gamma = 0$$

So, the constraint ~~set~~ Γ can be written alternatively as a constraint set on dual space of Cont functions ϕ, ψ.

Thus the dual problem is:

$$DP^* = \sup_{\substack{\phi \in C(\mathcal{X}) \\ \psi \in C(\mathcal{Y})}} \int \phi \, d\mu + \int \psi \, d\nu$$

$$\text{s.t.} \quad \phi \oplus \psi \leq c \quad \text{on } \mathcal{X} \times \mathcal{Y}.$$

Weak duality applies here, and we get $DP^* \leq KP^*$

(From Luca Nenna's notes)