Lectures on optimal transport

Alfred Galichon
14th European Summer School in Financial Mathematics
September 1-2, 2021

1 Roadmap

L1. Discrete optimal transport
L2. Inverse optimal transport
L3. Optimal transport and quantile methods

2 Optimal transport in the discrete case

Central planner's solution
Workers characteristics $x \in X$
n_x workers of type x

Firms characteristics $y \in Y$
m_y firms of type y

$\sum_x n_x = \sum_y m_y$

A matching π_{xy} is the mass of workers of type x matched with firms of type y

Constraints:
$\sum_y \pi_{xy} = n_x$
$\sum_x \pi_{xy} = m_y$

Let Φ_{xy} be the productivity (in dollar terms) of a worker of type x matched with a firm of type y.

Central planner's problem:
$\max_{\pi \geq 0} \sum_{xy} \pi_{xy} \Phi_{xy}$
s.t.
$\sum_y \pi_{xy} = n_x$
$\sum_x \pi_{xy} = m_y$

$\max_{\pi \geq 0} \sum_{xy} \pi_{xy} \Phi_{xy} + \min_{u_x, v_y} \sum_x u_x \left(n_x - \sum_y \pi_{xy} \right) + \sum_y v_y \left(m_y - \sum_x \pi_{xy} \right)$
$\max_{\pi \geq 0} \min_{u_x, v_y} \sum_{xy} \pi_{xy} \left(\Phi_{xy} - u_x - v_y \right) + \sum_x u_x n_x + \sum_y v_y m_y$
$\min_{u_x, v_y} \max_{\pi \geq 0} \sum_{xy} \pi_{xy} \left(\Phi_{xy} - u_x - v_y \right) + \sum_x u_x n_x + \sum_y v_y m_y$

$\min_{u_x,v_y} \sum_x u_x n_x + \sum_y v_y m_y + \max_{\pi \geq 0} \sum_{xy} \pi_{xy} \left(\Phi_{xy} - u_x - v_y\right)$
This becomes
$\min_{u_x,v_y} \sum_x u_x n_x + \sum_y v_y m_y + \max_{\pi \geq 0} \sum_{xy} \pi_{xy} \left(\Phi_{xy} - u_x - v_y\right)$

hence we arrive at the **dual to the central planner's problem**
$\min_{u_x,v_y} \sum_x u_x n_x + \sum_y v_y m_y$
s.t. $u_x + v_y \geq \Phi_{xy}$

Complementary slackness
$\pi_{xy} > 0 \implies u_x + v_y = \Phi_{xy}$

The dual problem can be interpreted as the equilibrium in the wage market. Indeed, consider the dual
$\min_{u_x,v_y} \sum_x u_x n_x + \sum_y v_y m_y$
s.t. $u_x + v_y \geq \Phi_{xy}$

Rewrite the constraint into
$v_y \geq \max_x \left\{\Phi_{xy} - u_x\right\}$

Claim: the dual can rewrite as
$\min_{u_x,v_y} \sum_x u_x n_x + \sum_y v_y m_y$
s.t. $v_y = \max_x \left\{\Phi_{xy} - u_x\right\}$

Equation
$v_y = \max_x \left\{\Phi_{xy} - u_x\right\}$
interprets as profit maximization by the firm.

2.1 Becker's theory of marriage

Becker's model of heterosexual marriage market
$x \in X$ man's characteristics in number n_x
$y \in Y$ woman's characteristics in number m_y

Assume that if x matches with y then
- x enjoys utility α_{xy}
- y enjoys utility γ_{xy}

Becker assumed **fully transferable utility**: if x gives up a quantity τ of utility to y, then after transfers
x gets $\alpha_{xy} - \tau$
y gets $\gamma_{xy} + \tau$

Note: we could get $\gamma_{xy} + f(\tau)$ but this would not be optimal transport. This is called matching with imperfectly transferable utility.

The limit without any possible transfers is called non-transferable utility model, this is Gale-Shapley's model.

Back to Becker's model

No matter the transfer, x and y if they match get joint surplus $\Phi_{xy} = \alpha_{xy} + \gamma_{xy}$.

2.2 Koopmans-Beckmann's model of transportation

$x \in X$ the locations of plants. Production (of a commodity) by plant x is n_x per day.

$y \in Y$ the locations of cities. Consumption of the same commodity by city y is m_y per day.

Total production = total consumption

(unit) Transportation cost from plant x to city y is c_{xy} per unit of mass.

Tolstoi's problem (1920).

Define π_{xy} the mass of commodity which will route from x to y. The optimal shipping plan is given by

$\min_{\pi \geq 0} \sum_{xy} \pi_{xy} c_{xy}$

s.t.

$\sum_{y} \pi_{xy} = n_x$

$\sum_{x} \pi_{xy} = m_y$

The dual of this problem is

$\max_{p_x, p_y} \sum_y m_y p_y - \sum_x n_x p_x$

$p_y - p_x \leq c_{xy}$

$\sum_y m_y p_y = \sum_x n_x p_x + \sum_{xy} \pi_{xy} c_{xy}$: total price at destination=total price at origin+total shipping cost

The constraint in the dual

$p_y \leq c_{xy} + p_x$

says that there is no pair xy such that

$p_y > c_{xy} + p_x$

this is a no-arbitrage conditions.

3 Computation of the optimal matching

In $\Phi(x, y) = x^\top A y = \sum_{kl} A_{kl} x^k y^l$

Most LP solvers compute

$\max_{z \geq 0} c^\top z$

s.t. $Mz = d$

$\max_{\pi\geq 0}\sum_{xy}\pi_{xy}\Phi_{xy}$
s.t.
$\sum_y \pi_{xy}=n_x$
$\sum_x \pi_{xy}=m_y$

$vec\left(\pi\right)$

$I_X\pi 1_Y=n$
$1_X^\top \pi I_Y=m^\top$

Get $vec\left(\pi 1_Y\right)=(...)vec\left(\pi\right)$ and $vec\left(1_X^\top\pi\right)=(...)\,vec\left(\pi\right)$
We have
$vec\left(AXB^\top\right)=\left(A\otimes B\right)vec\left(X\right)$

THis gives
$vec\left(I_X\pi 1_Y\right)=n$
$vec\left(1_X^\top\pi I_Y\right)=m$
therefore
$\left(I_X\otimes 1_Y^\top\right)vec\left(\pi\right)=n$
$\left(1_X^\top\otimes I_Y\right)vec\left(\pi\right)=m$

The OT problem can be computed as
$\max_{\pi\geq 0} vec\left(\pi\right)^\top vec\left(\Phi\right)$
s.t.
$\left(I_X\otimes 1_Y^\top\right)vec\left(\pi\right)=n$
$\left(1_X^\top\otimes I_Y\right)vec\left(\pi\right)=m$

4 Entropic regularization

$\max_{\pi\geq 0}\sum_{xy}\pi_{xy}\Phi_{xy}-\sigma\sum_{xy}\pi_{xy}\ln\pi_{xy}$
s.t.
$\sum_y \pi_{xy}=n_x\ \ [u_x]$
$\sum_x \pi_{xy}=m_y\ \ [v_y]$

$\max_{\pi\geq 0}\sum_{xy}\pi_{xy}\Phi_{xy}-\sigma\sum_{xy}\pi_{xy}\ln\pi_{xy}+\min_{u_x,v_y}\sum_x u_x\left(n_x-\sum_y\pi_{xy}\right)+\sum_y v_y\left(m_y-\sum_x\pi_{xy}\right)$
$\max_{\pi\geq 0}\min_{u_x,v_y}\sum_{xy}\pi_{xy}\left(\Phi_{xy}-u_x-v_y-\sigma\log\pi_{xy}\right)+\sum_x u_x n_x+\sum_y v_y m_y$
$\min_{u_x,v_y}\max_{\pi\geq 0}\sum_{xy}\pi_{xy}\left(\Phi_{xy}-u_x-v_y-\sigma\log\pi_{xy}\right)+\sum_x u_x n_x+\sum_y v_y m_y$
$\min_{u_x,v_y}\sum_x u_x n_x+\sum_y v_y m_y+\max_{\pi\geq 0}\sum_{xy}\pi_{xy}\left(\Phi_{xy}-u_x-v_y-\sigma\log\pi_{xy}\right)$

FOC in the inner problem
$\left(\Phi_{xy}-u_x-v_y-\sigma\log\pi_{xy}\right)-\sigma=0$
thus
$\pi_{xy}=\exp\left(\frac{\Phi_{xy}-u_x-v_y-\sigma}{\sigma}\right)$

and the value of the inner problem is $\sigma \sum_{xy} \pi_{xy} = \sigma \sum_{xy} \exp\left(\frac{\Phi_{xy} - u_x - v_y - \sigma}{\sigma}\right)$.
The outer problem becomes
$\min_{u_x, v_y} F(u, v) := \sum_x u_x n_x + \sum_y v_y m_y + \sigma \sum_{xy} \exp\left(\frac{\Phi_{xy} - u_x - v_y - \sigma}{\sigma}\right)$

Foc yield
wrt u_x: $n_x = \sum_y \exp\left(\frac{\Phi_{xy} - u_x - v_y - \sigma}{\sigma}\right)$
wrt v_y: $m_y = \sum_x \exp\left(\frac{\Phi_{xy} - u_x - v_y - \sigma}{\sigma}\right)$

(Blockwise) Coordinate descent.
Given v_y's compute u_x by $\min_{u_x} F(u, v)$
$n_x = \sum_y \exp\left(\frac{\Phi_{xy} - u_x - v_y - \sigma}{\sigma}\right)$
yields
$\exp(u_x/\sigma) = \sum_y \exp\left(\frac{\Phi_{xy} - v_y - \sigma}{\sigma}\right)/n_x$
similarly Given u_x's compute v_y by $\min_{v_y} F(u, v)$
yields
$\exp(v_y/\sigma) = \sum_x \exp\left(\frac{\Phi_{xy} - u_x - \sigma}{\sigma}\right)/m_y$

Sinkhorn's algorithm / Itererated Proportional Fitting Procedure
https://arxiv.org/abs/1609.06349

Write $A_x = \exp(-u_x/\sigma)$ and $B_y = \exp(-v_y/\sigma)$
and $K_{xy} = \exp\left(\frac{\Phi_{xy} - \sigma}{\sigma}\right)$, we have
$A_x = n_x / \sum_y K_{xy} B_y$
similarly Given u_x's compute v_y by $\min_{v_y} F(u, v)$
yields
$B_y = m_y / \sum_x K_{xy} A_y$

When σ is small
$u_x = \sigma \log\left(\sum_y \exp\left(\frac{\Phi_{xy} - v_y - \sigma}{\sigma}\right)\right) - \sigma \log n_x$
log-sum-exp trick
When computing numerically
$\sigma \log \sum_y \exp\left(\frac{z_y}{\sigma}\right)$
you should
$m = \max_y \{z_y\}$
and use the fact that for any c
$\sigma \log \sum_y \exp\left(\frac{z_y - c}{\sigma}\right) + c = \sigma \log \sum_y \exp\left(\frac{z_y}{\sigma}\right)$
use this with $c = m$ and get
$\sigma \log \sum_y \exp\left(\frac{z_y}{\sigma}\right) = \sigma \log \sum_y \exp\left(\frac{z_y - m}{\sigma}\right) + m$

5 Inverse optimal transport

Go back to the labor market, but this time introducing unmatched agents

Consider primal problem

$\max_{\pi\geq 0}\sum_{xy}\pi_{xy}\Phi_{xy}$

s.t.

$\sum_y \pi_{xy}\leq n_x$

$\sum_x \pi_{xy}\leq m_y$

which has dual

$\min_{u_x\geq 0, v_y\geq 0}\sum_x n_x u_x+\sum_y m_y v_y$

s.t.

$u_x+v_y\geq \Phi_{xy}$

Introduce unobserved heterogeneity.

Now we assume that n_x and m_y are very large.

Individual worker $i\in I$ has type $x_i\in X$

Individual firm $j\in J$ has type $y_j\in Y$

Assume that the joint output generated by i and j together is

$\Phi_{ij}=\Phi_{x_i y_j}+\varepsilon_{iy_j}+\eta_{x_i j}$. (Separability assumption).

The dual is

$\min\sum_{i\in I}u_i+\sum_{j\in J}v_j$

s.t.

$u_i+v_j\geq \Phi_{ij}:=\Phi_{x_i y_j}+\varepsilon_{iy_j}+\eta_{x_i j}$

$u_i\geq \varepsilon_{i0}$

$v_j\geq \eta_{0j}$

Take a look at the constraint

$u_i+v_j\geq \Phi_{x_i y_j}+\varepsilon_{iy_j}+\eta_{x_i j}$

rewrtie this as

$u_i-\varepsilon_{iy}+v_j-\eta_{xj}\geq \Phi_{xy}$ for all $i:x_i=x$ and for all $j:y_j=y$

$\min_{i:x_i=x}\{u_i-\varepsilon_{iy}\}+\min_{j:y_j=y}\left\{v_j-\eta_{xj}\right\}\geq \Phi_{xy}$

define $U_{xy}=\min_{i:x_i=x}\{u_i-\varepsilon_{iy}\}$ and $V_{xy}=\min_{j:y_j=y}\left\{v_j-\eta_{xj}\right\}$, in which case the constraints become

$U_{xy}+V_{xy}\geq \Phi_{xy}$

We have $U_{xy}=\min_{i:x_i=x}\{u_i-\varepsilon_{iy}\}$, therefore $u_i\geq \max_y\{U_{xy}+\varepsilon_{iy},\varepsilon_{i0}\}$ and similarly $v_j\geq \max_x\left\{V_{xy}+\eta_{xj},\eta_{0j}\right\}$

The micro problem rewrites

$\min\sum_{i\in I}u_i+\sum_{j\in J}v_j$

s.t.

$U_{xy}+V_{xy}\geq \Phi_{xy}$

$u_i\geq \max_y\{U_{xy}+\varepsilon_{iy},\varepsilon_{i0}\}$

$v_j\geq \max_x\left\{V_{xy}+\eta_{xj},\eta_{0j}\right\}$

and in fact
$\min \sum_{i\in I} u_i + \sum_{j\in J} v_j$
s.t.
$U_{xy} + V_{xy} \geq \Phi_{xy}$
$u_i = \max_y \{U_{xy} + \varepsilon_{iy}, \varepsilon_{i0}\}$
$v_j = \max_x \{V_{xy} + \eta_{xj}, \eta_{0j}\}$

which can rewrite
$\min \sum_{i\in I} \max_y \{U_{xy} + \varepsilon_{iy}, \varepsilon_{i0}\} + \sum_{j\in J} \max_x \{V_{xy} + \eta_{xj}, \eta_{0j}\}$
s.t.
$U_{xy} + V_{xy} \geq \Phi_{xy}$

that is
$\min \sum_x n_x \sum_{i:x_i=x} \frac{1}{n_x} \max_y \{U_{xy} + \varepsilon_{iy}, \varepsilon_{i0}\} + \sum_y m_y \sum_{j:y_j=y} \frac{1}{m_y} \max_x \{V_{xy} + \eta_{xj}, \eta_{0j}\}$
s.t.
$U_{xy} + V_{xy} \geq \Phi_{xy}$
We shall assume a statistical behaviour of ε_{iy} and η_{xj}'s. Assume that $(\varepsilon_{iy})_y$ is distributed according to P
and that $(\eta_{xj})_x$ is distributed according to Q.
By the law of large numbers the above expression converges to
$\min_{U,V} \sum_x n_x E_P [\max_y \{U_{xy} + \varepsilon_y, \varepsilon_0\}] + \sum_y m_y E_Q [\max_x \{V_{xy} + \eta_x, \eta_0\}]$
s.t.
$U_{xy} + V_{xy} \geq \Phi_{xy}$

Define $G_x(U) = E_P [\max_y \{U_{xy} + \varepsilon_y, \varepsilon_0\}]$ and $H_y(V) = E_Q [\max_x \{V_{xy} + \eta_x, \eta_0\}]$
and $G(U) = \sum_x n_x G_x(U)$ and $H(V) = \sum_y m_y H_y(V)$
the problem becomes

$\min_{U,V} G(U) + H(V)$
s.t.
$U_{xy} + V_{xy} = \Phi_{xy}$

$\min_U G(U) + H(\Phi - U)$
FOC
$\frac{\partial G(U)}{\partial U_{xy}} = \frac{\partial H}{\partial V_{xy}} (\Phi - U)$

$\frac{\partial G(U)}{\partial U_{xy}} = n_x \frac{\partial G_x(U)}{\partial U_{xy}} = n_x \Pr(y \in \arg\max \{U_{xy} + \varepsilon_y, \varepsilon_0\})$
$\frac{\partial H(V)}{\partial V_{xy}} = m_y \Pr(x \in \arg\max \{V_{xy} + \eta_x, \eta_0\})$

Dual problem:
Start from the penalized problem
$\min_{U,V} G(U) + H(V) + \max_{\pi\geq 0} \sum \pi_{xy} (\Phi_{xy} - U_{xy} - V_{xy})$
$\max_{\pi\geq 0} \sum \pi_{xy}\Phi_{xy} + \min_{U,V} \{G(U) + H(V) - \sum \pi_{xy} (U_{xy} + V_{xy})\}$
$\max_{\pi\geq 0} \sum \pi_{xy}\Phi_{xy} - \max_U \{\sum \pi_{xy} U_{xy} - G(U)\} - \max_V \{\sum \pi_{xy} V_{xy} - H(V)\}$

Define the convex conjugates / Legendre-Fenchel transforms of G and H as
$G^*(\pi) = \max_U \{\sum \pi_{xy} U_{xy} - G(U)\}$
$H^*(\pi) = \max_V \{\sum \pi_{xy} V_{xy} - H(V)\}$
and the problem becomes
$\max_{\pi_{xy} \geq 0} \{\sum \pi_{xy} \Phi_{xy} - G^*(\pi) - H^*(\pi)\}$

FOC yield
$\Phi_{xy} = \frac{\partial G^*(\pi)}{\partial \pi_{xy}} + \frac{\partial H^*(\pi)}{\partial \pi_{xy}}$

When P (the distribution of $(\varepsilon_{iy})_y$) is the Gumbel distribution (extreme-value type 1 – distribution with cdf $\exp(-\exp(-z))$)
https://en.wikipedia.org/wiki/Gumbel_distribution

then $G_x(U) = \log\left(1 + \sum_y \exp U_{xy}\right)$, and therefore $G(U) = \sum_x n_x \log\left(1 + \sum_y \exp U_{xy}\right)$.
We have
$G_x^*(\pi) = \sum_{y \in Y} \frac{\pi_{xy}}{n_x} \log \frac{\pi_{xy}}{n_x} + \frac{\pi_{x0}}{n_x} \log \frac{\pi_{x0}}{n_x}$ where $\pi_{x0} = n_x - \sum_y \pi_{xy}$

Therefore the primal problem in this case rewrites
$\max_{\pi_{xy} \geq 0} \left\{\sum \pi_{xy} \Phi_{xy} - \sum_x n_x \sum_{y \in Y} \frac{\pi_{xy}}{n_x} \log \frac{\pi_{xy}}{n_x} - \pi_{x0} \log \frac{\pi_{x0}}{n_x} - \sum_y m_y \sum_{x \in X} \frac{\pi_{xy}}{m_y} \log \frac{\pi_{xy}}{m_y} - \pi_{0y} \log \frac{\pi_{0y}}{m_y}\right\}$
which yields
$\max_{\pi_{xy} \geq 0} \left\{\sum \pi_{xy} \Phi_{xy} - \sum_{xy} \pi_{xy} \log \frac{\pi_{xy}^2}{n_x m_y} - \pi_{x0} \log \frac{\pi_{x0}}{n_x} - \pi_{0y} \log \frac{\pi_{0y}}{m_y}\right\}$

First order conditions wrt π_{xy}:
$\Phi_{xy} = 2 + 2\log \pi_{xy} - \log(n_x m_y) - (1 + \log \pi_{x0} - \log n_x) - (1 + \log \pi_{0y} - \log m_y)$.
This is Choo and Siow's formula (Journal of Political Economy, 2006)

$$\Phi_{xy} = \log \frac{\pi_{xy}^2}{\pi_{x0}\pi_{0y}}.$$

6 Continued, 9/2/2021

Today:
* structural estimation of matching models
* quantile methods

Parameterize $\Phi_{xy}^\lambda = \sum_k \lambda_k \phi_{xy}^k$

Use of generalized linear models.

Start from the primal problem we wrote yesterday
$W(\lambda) = \max_{\pi_{xy} \geq 0} \left\{\sum_{xy} \pi_{xy} \Phi_{xy}^\lambda - \sum_{xy} \pi_{xy} \log \frac{\pi_{xy}^2}{n_x m_y} - \pi_{x0} \log \frac{\pi_{x0}}{n_x} - \pi_{0y} \log \frac{\pi_{0y}}{m_y}\right\}$
$\frac{\partial W}{\partial \lambda_k} = \sum_{xy} \pi_{xy}^\lambda \phi_{xy}^k$=predicted moment of order k
Look for λ such that predicted moments = observed moments

Observed moment of order k: $\sum_{xy} \hat{\mu}_{xy} \phi^k_{xy}$ where $\hat{\mu}_{xy}$ is the ** observed ** number of pairs xy.

Look for λ such that $\frac{\partial W}{\partial \lambda_k} = \sum_{xy} \hat{\mu}_{xy} \phi^k_{xy}$. This can be interpreted as the first order conditions of

$$\min_{\lambda} \left\{ W(\lambda) - \sum_k \sum_{xy} \lambda_k \hat{\mu}_{xy} \phi^k_{xy} \right\}$$

Link with generalized linear models. https://en.wikipedia.org/wiki/Generalized_linear_model
The convex dual problem above is
$W(\lambda) = \min_{u_x, v_y} \left\{ \sum_x n_x u_x + \sum_y m_y v_y + 2 \sum_{xy} \exp\left(\frac{\Phi_{xy} - u_x - v_y}{2} \right) + \sum_x \exp(-u_x) + \sum_y \exp(-v_y) \right\}$

The estimation problem then becomes

$$\min_{\lambda, u, v} \left\{ \sum_x n_x u_x + \sum_y m_y v_y + 2 \sum_{xy} \exp\left(\frac{\Phi_{xy} - u_x - v_y}{2} \right) + \sum_x \exp(-u_x) + \sum_y \exp(-v_y) - \sum_k \sum_{xy} \lambda_k \hat{\mu}_{xy} \phi^k_{xy} \right\}$$

$$\min_{\lambda, u, v} \left\{ \sum_x n_x u_x + \sum_y m_y v_y + 2 \sum_{xy} \exp\left(\frac{\Phi_{xy} - u_x - v_y}{2} \right) + \sum_x \exp(-u_x) + \sum_y \exp(-v_y) - \sum_k \sum_{xy} \lambda_k \hat{\mu}_{xy} \phi^k_{xy} + \right.$$

$\pi_{xy} = \exp\left(\frac{\Phi^{\lambda}_{xy} - u_x - v_y}{2} \right)$
$\pi_{x0} = \exp(-u_x)$
$\pi_{0y} = \exp(-v_y)$

This model is a generalized model with 2-way fixed effects.

7 Quantiles

7.1 A model of CEOs and firms

Now workers = CEOs are measured by their talent $x \in R$=extra return on capital. Assume that x is distributed according P (continous)

Firms are captured by their asset size $y \in R$, drawn from a distribution Q (continuous)

$\Phi(x, y) = xy$.

Optimal matching of CEOs and firms:
$\max_{\pi(x,y) \geq 0} \int \Phi(x, y) \pi(x, y) \, dx dy$
s.t. $\int \pi(x, y) \, dy = p(x)$

and $\int \pi(x,y)\,dx = q(y)$

the dual
$\min \int u(x)\,p(x)\,dx + \int v(y)\,q(y)\,dy$
s.t. $u(x) + v(y) \geq \Phi(x,y) = xy$

Agenda:
1. Use common sense to guess a solution to the primal
2. Use the primal solution to find the dual solution
3. Predict wages accordingly

Intuitively, try to assign the top CEOs to the largest firms, ie. $y = T(x)$ where T is increasing.

Take (X,Y) be the optimal coupling. We would like to have $Y = T(X)$ and $X \sim P$ and $Y \sim Q$.
$\Pr(X \leq x) = F_P(x)$
$\Pr(Y \leq y) = F_Q(y)$
$\Pr(T(X) \leq y) = F_Q(y)$
$\Pr\left(X \leq T^{-1}(y)\right) = F_Q(y)$
$F_P\left(T^{-1}(y)\right) = F_Q(y)$
$T(x) = F_Q^{-1}(F_P(x))$.

This is called the comonotone solution.

Remark: this solution implies that $F_Q(Y) = F_P(X)$. This is called positive assortative matching.

Theorem. When Φ is supermodular, ie when $\partial^2\Phi(x,y)/\partial x \partial y \geq 0$, the the comonotone solution is optimal.

Definition: The quantile is the inverse of the cdf.
F_Q^{-1} is called the quantile function associated with the distribution Q.
$F_Q^{-1}(1/2)$ is the median.

2. Dual solution.
$\min \int u(x)\,p(x)\,dx + \int v(y)\,q(y)\,dy$
s.t. $u(x) + v(y) \geq xy$

Assume u, v is optimal.
Yesterday we saw that if u, v is optimal then
$v(y) = \max_x \{xy - u(x)\}$
and
$u(x) = \max_y \{xy - v(y)\}$
that implies that u and v are convex.
Assume u and v are differentiable.
Firm y matches with CEO x which satisfies FOC that is
$y = u'(x)$.

Therefore $u'(x) = F_Q^{-1}(F_P(x))$ and thus we can construct
$u(x) = \int_0^x F_Q^{-1}(F_P(z))\,dz + c$
and
$v(y) = \int_0^y F_P^{-1}(F_Q(w))\,dw - c'$

Consider in particular the case $P = \mathcal{U}([0,1])$. Then
$T(x) = F_Q^{-1}(F_P(x)) = F_Q^{-1}(x)$
and we have that
$T(x) = u'(x)$ where u is part of the dual solution to
$\min \int_0^1 u(x)\,dx + \int v(y)\,dQ(y)$
s.t. $u(x) + v(y) \geq xy$.

Assume the distribution Q is not continuous. Eg assume that $Q = \frac{1}{N}\sum_{j=1}^N \delta_{y_j}$.
$\min \int_0^1 u(x)\,dx + \frac{1}{N}\sum_{j=1}^N v(y_j)$
s.t. $u(x) + v(y_j) \geq xy_j$

$\min_{v \in R^N} \left\{ \int_0^1 \max_j \{xy_j - v_j\}\,dx + \frac{1}{N}\sum_{j=1}^N v_j \right\}$
Problem has a unique (up to a constant solution) $(v_j) \in R^N$
Define
$u_N(x) = \max_j \{xy_j - v_j\}$

Classical definition:
$F_Q^{-1}(t) = \inf\{x \in \mathbb{R} : F_Q(x) > t\}$
which is nondecreasing and cadlag (right-continuous, left-limit)
we also could have taken a different definition and pose
$\sup\{x \in \mathbb{R} : F_Q(x) < t\}$
which is nondecreasing and caglad (left-continuous, right-limit)
The two definitions coincide on the points of continuity (where there is no jump).

7.2 Multivariate quantiles

Fundamental property of (univariate) quantiles.
Take Q a continuous distribution on R.
The quantile function F_Q^{-1} is the only function T such that:
* T is nondecreasing, and
* if $U \sim \mathcal{U}([0,1])$, then $T(U) \sim Q$.

We would like to generalize the result in higher dimension.
Take Q a continuous distribution on R^d.
Is there a function T such that:
* T is cyclically monotone ie this is the gradient of a convex funcion, and
* if $U \sim \mathcal{U}^d = \mathcal{U}\left([0,1]^d\right)$, then $T(U) \sim Q$.

Consider
$\min \int u(x)\, d\mathcal{U}^d + \int v(y)\, dQ(y)$
s.t. $u(x) + v(y) \geq x^\top y$.

$v(y) = \max_x \left\{ x^\top y - u(x) \right\}$
and
$u(x) = \max_y \left\{ x^\top y - v(y) \right\}$

And by foc, y is matched with x such that $y = \nabla u(x)$.
If $X \sim \mathcal{U}^d$, then $\nabla u(X) \sim Q$

When we have a sample y_j, we can define
$\min_{v \in R^N} \left\{ \int_{[0,1]^d} \max_j \left\{ x^\top y_j - v_j \right\} dx + \frac{1}{N} \sum_{j=1}^{N} v_j \right\}$
and the multivariate quantile is such that
$T(x) = \nabla u(x)$ where $u(x) = \max_j \left\{ x^\top y_j - v_j \right\}$.

7.3 Quantile regression

The problem. Consider the t-quantile of Y conditional on X. This is the t-th quantile of the conditional distribution $F^{-1}_{Y|X=x}(t)$.

Quantile regression is about fitting a parametric form for this object, namely

$$F^{-1}_{Y|X=x}(t) = x^\top \beta_t$$

How do we compute β_t?

Idea. Consider for now the unconditional quantile of Y, $q = F_Y^{-1}(t)$.
For $t = 1/2$, we know that q solves

$$\min_q E\left[|Y - q|\right]$$

indeed,

$$\min_q E\left[\tau (Y - q)^+ + (1 - \tau)(Y - q)^-\right]$$

and by FOC

$$\tau E\left[1\left\{Y \geq q\right\}\right] = (1 - \tau) E\left[1\left\{Y \leq q\right\}\right]$$

thus $\Pr(Y \leq q) = \tau$.

Taking loss function $\rho_\tau(z) = \tau z^+ + (1 - \tau) z^-$, then $q = F_Y^{-1}(t)$ is the minimizer of

$$\min_q E\left[\rho_\tau (Y - q)\right].$$

In particular, $x^\top \beta_t$ will minimize the condititional expectation of the loss function

$$\min_q E\left[\rho_\tau (Y - q) | X = x\right]$$

and therefore, β_t will minimize

$$\min_{\beta} E\left[\rho_\tau\left(Y - X^\top \beta\right)|X = x\right]$$

By integrating over x, we get that β_t minimizes

$$\min_{\beta} E\left[\rho_\tau\left(Y - X^\top \beta\right)\right].$$

This is actually a linear programming problem. Indeed, introduce $P = \left(Y - X^\top \beta\right)^+$ and $N = \left(Y - X^\top \beta\right)^-$, the problem becomes

$$\begin{aligned} \min_{\beta, P \geq 0, N \geq 0} \quad & E\left[\tau P + (1-\tau) N\right] \\ s.t. \quad & P - N = \left(Y - X^\top \beta\right) \end{aligned}$$