Optimal transport in economics

A short tutorial

Alfred Galichon

(New York University and Sciences Po)

Chamberlain seminar, April 5 2024

Roadmap

I. Basic results

II. Connections and applications

1. Quantiles and quantile regression
2. Conditional choice probability inversion
3. Generalized linear models and inverse optimal transport

III. References

Research supported by the European Research Council under grant ERC-CoG No. 866274.

Code

Code demos will be drawn from the 'math+econ+code' notebooks
`www.math-econ-code.org`.

Part I. Basic results

Theorem (Monge-Kanrotovich)

Let n and m be two probability measures on two finite-dimensional spaces $\mathcal{X}$ and $\mathcal{Y}$ and assume that $\Phi(x, y)$ is upper semicontinuous and that there are $\overline{a} \in L^1(n)$ and $\overline{b} \in L^1(m)$ such that $\Phi(x, y) \leq \overline{a}(x) + \overline{b}(y)$ for all x, y. Let $\mathcal{M}(n, m)$ be the set of distributions of (X, Y) such that $X \sim n, Y \sim n$. Then:

(i) The values of the primal and the dual problems coincide, i.e.

$$\sup_{\mu \in \mathcal{M}(n,m)} \mathbb{E}_\mu[\Phi(X, Y)] = \inf_{u(X)+v(Y) \geq \Phi(X,Y)} \mathbb{E}_n[u(X)] + \mathbb{E}_m[v(Y)]. \tag{1}$$

(ii) The primal problem on the left hand-side of (1) has an optimal solution μ.

(iii) Assume further that there are $\underline{a} \in L^1(n)$ and $\underline{b} \in L^1(m)$ such that $\Phi(x, y) \geq \underline{a}(x) + \underline{b}(y)$ for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. Then the dual problem on the right hand-side of (1) has an optimal solution (u, v) as well.

The Monge-Kantorovich theorem: intuition

- For any (u, v) with $u(x) + v(y) \geq \Phi(x, y)$ and for any $(X, Y) \sim \mu$ such that $X \sim n$ and $Y \sim n$, we have

$$\mathbb{E}_m[u(X)] + \mathbb{E}_n[v(Y)] \geq \mathbb{E}_\mu \Phi(X, Y), \tag{2}$$

  so inf(left-hand side) $\geq$ sup(right-hand side) (weak duality).

- Difficult part: (1) inf $=$ sup (strong duality), and (2) existence of primal and dual solutions.

- If u and v are *feasible* for the dual, then

$$u(x) \geq \sup_y \{\Phi(x, y) - v(y)\} \text{ and } v(y) \geq \sup_x \{\Phi(x, y) - u(x)\} \tag{3}$$

  however, if they are *optimal* for the dual, then these inequalities hold as equalities.

The Monge-Kantorovich theorem: economic interpretation

Consider matching employees and firms. n is distribution of CEO's type x, and m is the distribution of a firm's type y. An (x, y) match produces a monetary output of $\Phi(x, y)$.

Central planner's solution

The central planner picks a distribution $\mu(x, y)$ of matches between CEO x and firm y. The output generated is $\mathbb{E}_\mu[\Phi(X, Y)]$.

Decentralized solution

$u(x)$ is the wage of CEO x, and $v(y)$ the wage of firm y. Incentive compatibility requires that $v(y) = \sup_x\{\Phi(x, y) - u(x)\}$, thus $u(x) + v(y) \geq \Phi(x, y)$. The total output is then $\mathbb{E}_m[u(X)] + \mathbb{E}_n[v(Y)]$.

Equilibrium holds if the two coincide. By Monge-Kantorovich, this is the case!

Discrete marginals

When $\mathcal{X}$ and $\mathcal{Y}$ are finite sets, n and m are two nonnegative vectors of $\mathbb{R}^{\mathcal{X}}$ and $\mathbb{R}^{\mathcal{Y}}$ respectively such that $\sum_x n_x = \sum_y m_y = 1$, and the primal is

$$\begin{aligned} \max_{\mu_{xy} \geq 0} \quad & \sum_{xy \in \mathcal{X} \times \mathcal{Y}} \mu_{xy} \Phi_{xy} \\ s.t. \quad & \sum_{y \in \mathcal{Y}} \mu_{xy} = n_x \text{ and } \sum_{x \in \mathcal{X}} \mu_{xy} = m_y \end{aligned} \tag{4}$$

while the dual is

$$\begin{aligned} \min_{u_x, v_y} \quad & \sum_{x \in \mathcal{X}} n_x u_x + \sum_{y \in \mathcal{Y}} m_y v_y \\ s.t. \quad & u_x + v_y \geq \Phi_{xy}. \end{aligned} \tag{5}$$

Entropic regularization

For $T > 0$, consider the entropic regularized primal

$$\max_{\mu \in \mathcal{M}(n,m)} \sum_{xy \in \mathcal{X} \times \mathcal{Y}} \mu_{xy} \Phi_{xy} - T \sum_{xy \in \mathcal{X} \times \mathcal{Y}} \mu_{xy} \log \mu_{xy}, \tag{6}$$

whose dual is

$$\min_{u,v} \sum_{x \in \mathcal{X}} n_x u_x + \sum_{y \in \mathcal{Y}} m_y v_y + T \sum_{xy \in \mathcal{X} \times \mathcal{Y}} \exp\left(\frac{\Phi_{xy} - u_x - v_y}{T} \right) \tag{7}$$

whose first order conditions determine u and v by what is called a *Bernstein–Schrödinger system* of equations:

$$\begin{cases} n_x = \sum_{y \in \mathcal{Y}} \exp\left((\Phi_{xy} - u_x - v_y)/T \right) \\ m_y = \sum_{x \in \mathcal{X}} \exp\left((\Phi_{xy} - u_x - v_y)/T \right). \end{cases} \tag{8}$$

Entropic regularization (continued)

Theorem

Consider the Bernstein–Schrödinger system (8). Then:

(1) The system has a unique solution (u, v) up to an additive constant.

(2) The set of solutions to the system coincides with the set of minimizers of the dual problem (7).

(3) If (u, v) is a solution to the system, then μ_{xy} defined by

$$\mu_{xy} := \exp\left(\frac{\Phi_{xy} - u_x - v_y}{T}\right) \tag{9}$$

is a solution to the primal regularized problem (6). Conversely, every primal solution arises in that way.

Entropic regularization (continued)

The proof is constructive, and involves the description of the *iterated proportional fitting procedure* (IPFP), also called *Sinkhorn's algorithm*. Let $a_x = \exp(-u_x/T)$ and $b_y = \exp(-v_y/T)$ and define:

$$a_x^{t+1} = \frac{1}{n_x} \sum_{y \in \mathcal{Y}} \exp\left(\frac{\Phi_{xy}}{T}\right) \frac{1}{b_y^t} \text{ and } b_y^{t+1} = \frac{1}{m_y} \sum_{x \in \mathcal{X}} \exp\left(\frac{\Phi_{xy}}{T}\right) \frac{1}{a_x^{t+1}}. \tag{10}$$

We can show that (10) is a strict contraction with respect to *Hilbert's projective metric*

$$d_H(a, b) = \log \max_{x,x'} \left\{ \frac{a_x b_{x'}}{a_{x'} b_x} \right\}. \tag{11}$$

Theorem (Franklin and Lorenz 1989)

The solution $(u^\star, v^\star)$ to (8) exists and is unique up to a constant. For some $K \in (0,1)$, $d_H(a^t, a^\star) \leq K^t d_H(a^0, a^\star)$, and $d_H(b^t, b^\star) \leq K^t d_H(b^0, b^\star)$ $\forall t \geq 0$.

Part II. Connections and applications

1. Quantiles and quantile regression

Quantiles: Consider Y a r.v. with a continuous distribution m, and c.d.f. $F_Y(.)$.

The quantile of Y at $x \in [0,1]$ is the inverse of the c.d.f. at x, $F_Y^{-1}(x)$.

It is the only increasing map T such that $T(X) \sim m$ for $X \sim \mathcal{U}([0,1])$.

It is the minimizer of the loss function $q \to \mathbb{E}\left[x(Y-q)^+ + (1-x)\,(Y-q)^-\right]$.

Quantile regression: If Z is a vector of covariates, then following Koenker we can assume that the conditional quantile of Y given $Z=z$ at $x \in [0,1]$ belongs to a parametric family: $F_{Y|Z}^{-1}(x|z) = z^\top \beta_x$.

The parameter β_x is obtained as the quantile regression estimator of Y on $Z=z$ which is

$$\hat{\beta}_x = \arg\min_{\tilde{\beta}} \mathbb{E}\left[x(Y - Z^\top \beta)^+ + (1-x)\,(Y - Z^\top \beta)^-\right].$$

Quantiles and quantile regression as OT problems

The MK theorem with $\mathcal{X} = \mathcal{Y} = \mathbb{R}$, $n = \mathcal{U}([0,1])$ and m, and $\Phi(x, y) = xy$ leads to:

Theorem (Hardy-Littlewood)

The dual solution OT problem has $u'(x) = F_Y^{-1}(x)$, and $v'(y) = F_Y(y)$. The primal solution is the distribution μ of (X, Y) such that $Y = F_Y^{-1}(X)$.

Carlier, Chernozhukov and Galichon extend Monge-Kantorovich to quantile regression:

Theorem (CCG)

Let m be the distribution of (Z, Y). Then $\beta_x = b'(x)$, where b is from

$$\begin{aligned} \min_{b,\psi} \quad & \mathbb{E}_m[Z]^\top \mathbb{E}_n[b(X)] + \mathbb{E}_m[\psi(Z, Y)] \\ s.t. \quad & z^\top b(x) + \psi(z, y) \geq xy \end{aligned}$$

Taking quantiles to higher dimensions

Application to **multivariate quantiles**: when Y is multivariate, replace the product by the scalar product to get a notion of multivariate quantiles. We use the following fundamental result.

Theorem (Brenier 1987 and McCann 1995)

Let n and m be two probability measures on respectively $\mathcal{X}$ and $\mathcal{Y}$ such that n has a nonvanishing density on $\mathcal{X}$. Then there is a unique (up to an additive constant) convex function u such that if $X \sim n$, then $\nabla u\left(X\right) \sim m$.

Taking $n = \mathcal{U}([0,1]^d)$ allows us to define $\nabla u(x)$ as a multivariate quantile, or "vector quantile." This found applications to risk measures by Ekeland, Galichon and Henry (2012), decision theory in Galichon and Henry (2012), multivariate depth in Hallin, Chernozhukov, Galichon and Henry (2017).

2. CCP inversion

Consider the discrete choice problem with $Y+1$ options (default+Y alternatives):

$$u(\varepsilon) = \max_{y\in[Y]} \{U_y + \varepsilon_y, \varepsilon_0\}, \tag{12}$$

where $U_0 = 0$ is the normalization of the default option, and $\varepsilon \sim \mathcal{P}$ is a $Y+1$ random vector of random utility.

If $\mathcal{P}$ has a density, the expression of the conditional choice probability (CCP) of option y as a function of U is given by

$$\pi_y(U) := \mathcal{P}(U_y + \varepsilon_y \geq U_z + \varepsilon_z, \forall z \in [0{:}Y]), \tag{13}$$

This defines the *CCP map*, $\pi(U)$. Inverting the CCP map is important in structural econometrics (McFadden, Berry, Berry-Levinsohn-Pakes, Hotz-Miller).

CCP inversion as an OT problem

Galichon and Salanié (2022) prove the following:

CCP inversion theorem

If $\mathcal{P}$ is continuous, then $U \in \pi^{-1}(\pi)$ if and only if (u, U) is the unique solution such that $U_0 = 0$ of the dual M-K problem

$$\min_{u,U} \int u(\varepsilon)\, d\mathcal{P}(\varepsilon) - \sum_{y \in [0:Y]} \pi_y U_y \tag{14}$$

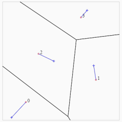
$$s.t.\ u(\varepsilon) - U_y \geq \varepsilon_y\ \forall \varepsilon \in \mathbb{R}^{Y+1}, \forall y \in [0:Y].$$

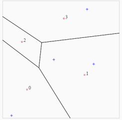
The value of problem (14) is called the *entropy of choice* associated with the discrete choice problem. In the logit case, it is (minus) the usual Gibbs entropy.

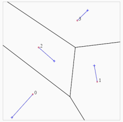
Any discrete choice model can be inverted by simulation+linear programming.

Using OT tools for CCP inversion: characteristics model

The "pure characteristics model" of Berry and Pakes is a Lancasterian model where an agent $i \in [I]$ draws a utility shock ε_{iy} associated with option y given by $\varepsilon_{iy} = \sum_{k \in [K]} \epsilon_{ik} \xi_{ky}$, where ξ_y is a K -dimensional vector of characteristics and ϵ_i, also a K-dimensional vector of valuation of these respective characteristics by agent i, such that $\epsilon_i \sim \mathcal{P}_\epsilon$. See Bonnet, Galichon, Hsieh, O'Hara, and Shum (BGHOS 2022).







Inverting the pure characteristics model is a semi-discrete optimal transport problem, which can be solved with the use of power, a.k.a. Laguerre diagrams.

Using OT tools for CCP inversion: random coefficient logit

In the random coefficient logit model of Berry, Levinsohn and Pakes, the random utility is the sum of a probit and logit term:

$$\varepsilon = \xi^\top \epsilon + \sigma \eta$$

where ϵ (Gaussian) is independent from η (i.i.d. Gumbel). One can simulate I draws of ϵ_i and one has:

Theorem (BGHOS)

Given market shares π_y, the systematic utility U_y in the random coefficient logit model is obtained as the limit for $I \to +\infty$ of the solution to the entropic-regularized optimal transport problem

$$\min_{u_i, U_y} \left\{ \frac{1}{I} \sum_{i \in [I]} u_i - \sum_{y \in [Y]} \pi_y U_y + \sigma \sum_{iy \in [I \times Y]} \exp\left(\frac{\varepsilon_{iy} - u_i + U_y}{\sigma}\right) \right\} \tag{15}$$

3. Generalized linear models and fixed effects

If $\tilde{\mu}$ is a count random variable, we may assume $\tilde{\mu}$ given regressors $\tilde{\rho} = \rho$ is Poisson, so

$$\Pr\left(\tilde{\mu} = \mu | \tilde{\rho} = \rho\right) = \exp(-\exp(\rho_0 + \rho^\top \theta)) \frac{\exp(\mu(\rho_0 + \rho^\top \theta))}{\mu!}, \tag{16}$$

where ρ_0 is an intercept term. Stacking the regressors ρ_ω into a design matrix R, the Poisson regression estimator $\hat{\theta}$ is thus the parameter vector θ solution to

$$\max_{\theta \in \mathbb{R}^P} \{\mu^\top R\theta - \sum_{\omega \in \Omega} \exp\left(\rho_{0\omega} + (R\theta)_\omega\right)\}. \tag{17}$$

We can generalize the Poisson regression to generalize linear models by replacing exp by l^{-1}, where l is the *link function*.

Entropic OT as a generalized linear model

Proposition

The entropic regularized optimal transport problem can be computed as a Poisson regression of $\hat{\mu}_{xy} = n_x m_y$ on regressors solely composed of x- and y-fixed effects, with intercept ϕ_{xy}/T.

This is a generalization of the well-known "Poisson trick" which asserts that logistic regression = Poisson + x-fixed effects. Here, optimal transport = Poisson + x-and y-fixed effects.

Entropic OT problems can easily be computed that way in Stata!

Gravity equation as regularized OT

If we parameterize $\phi^{\lambda}_{xy} = \sum_k \lambda_k \phi_{kxy}$, we get the PPML estimator of the gravity equation (Santos Silva and Tenreyro), when we observe trade flows $\hat{\mu}_{xy}$

$$\min_{\lambda_k, u_x, v_y} \sum_{xy} \hat{\mu}_{xy}(u_x + v_y - \phi^{\lambda}_{xy}) + \sum_{xy} \exp(u_x + v_y - \phi^{\lambda}_{xy}).$$

Carlier, Dupuy, Galichon and Sun propose the SISTA algorithm for efficient computation of λ, with a possible Lasso-type L1 penalization. Idea: alternate Sinkhorn steps on u and v, and proximal gradient steps on λ.

Estimation of matching models

In the model of matching of Choo and Siow, if ϕ^{λ}_{xy} is the matching surplus, and u_x, v_y are the average utilities of x and y, one has at equilibrium

$$\mu_{xy} = \sqrt{n_x m_y} e^{\frac{\phi^{\lambda}_{xy} - u_x - v_y}{2}} \text{ and } \mu_{x0} = n_x e^{-u_x} \text{ and } \mu_{0y} = m_y e^{-v_y}. \tag{18}$$

Galichon and Salanié show that λ can be estimated using:

Theorem (GS)

The moment equations (18) are the optimality conditions associated with

$$\min_{\lambda,u,v} \left\{ \begin{array}{c} \sum_{x\in[X]} n_x u_x + \sum_{y\in[Y]} m_y v_y - \sum_{xyk\in[X\times Y\times K]} \hat{\mu}_{xy} \phi_{xyk} \lambda_k \\ +2 \sum_{xy\in[X\times Y]} \sqrt{n_x m_y} e^{\frac{(\Phi\lambda)_{xy} - u_x - v_y}{2}} + \sum_{x\in[X]} n_x e^{-u_x} + \sum_{y\in[Y]} m_y e^{-v_y} \end{array} \right\}.$$

which can be computed as a weighted generalized linear model.

Part III. References

Online resources

Gabriel Peyré's blog `https://twitter.com/gabrielpeyre`

Sargent and Stachurski's quantecon lectures `https://quantecon.org/`.

My math+econ+code lectures `www.math-econ-code.org`

Texts

Peyré, Cuturi (2019).*Computational Optimal Transport: With Applications to Data Science*. Foundations and trends in machine learning.

Galichon (2016). *Optimal Transport Methods in Economics*. Princeton.

Galichon. *Mathematical methods of discrete choice*. In preparation.

Santambrogio (2015). *Optimal Transport for Applied Mathematicians*. Birkhaüser.