'MATH+ECON+CODE' MASTERCLASS ON COMPETITIVE EQUILIBRIUM: WALRASIAN EQUILIBRIUM WITH SUBSTITUTES

Alfred Galichon (NYU)

Spring 2018
Day 3, May 23 2018: matching with general transfers (1)
Block 7. Equilibrium transport

Learning objectives: block 7

- Equilibrium flow problem and min-cost flow problem
- Equilibrium transport problem and optimal transport problem
- Perfect matching, min-cut max-flow theorem and Strassen's theorem
- Reduced network and Bellman-Ford algorithm

- Demange and Gale (1985). The Strategy Structure of Two-Sided Matching Markets. *Econometrica*.
- Ahuja, Magnanti and Orlin (1993). *Network Flows: Theory, Algorithms, and Applications*. Pearson.
- Bertsekas (1998). *Network Optimization: Continuous and Discrete Models*. Athena scientific.
- Villani (2003). *Topics in Optimal Transportation*. AMS.
- Villani (2008). *Optimal transport: old and new*. Springer.
- Trudinger (2014). On the local theory of prescribed Jacobian equations. *Discrete and continuous dynamical systems*.
- Nöldeke, Samuelson (2017). The implementation duality.
- Galichon, Vernet (2018). Monotone comparative statics in the equilibrium flow problem.
- Galichon, Samuelson (2018). Multivocal gross substitutes.

Section 1

Equilibrium flows

The reference for the following is Galichon and Vernet (2018).

- Consider a trading network $(\mathcal{Z}, \mathcal{A})$ where $\mathcal{Z}$ is the set of nodes and $\mathcal{A}$ is the set of directed arcs.
- Consider ∇ a $\mathcal{A} \times \mathcal{Z}$ matrix such that $\nabla_{az} = 1$ if z is the endpoint of a, and -1 if z is the starting point of a. For $f \in \mathbb{R}^{\mathcal{Z}}$ and $xy \in \mathcal{A}$ one has $(\nabla f)_{xy} = f_y - f_x$.
- Let $p \in \mathbb{R}^{\mathcal{Z}}$ be the price vector of a commodity, such that p_z is the price at node z.
- Let $R : \mathbb{R}^{\mathcal{Z}} \mapsto \mathbb{R}^{\mathcal{A}}$ be a function $R_{xy}(p)$ be the rent of the strategy that consists in buying at x at price p_x and selling at y at price p_y. $R_{xy}(p)$ is decreasing in p_x, increasing in p_y, and does not depend on the other entries of p. Examples:
  - Additive case: $R_{xy}(p) = p_y - p_x - c_{xy}$ (no tax). Note that in that case, $R(p) = \nabla p - c$.
  - Linear case: $R_{xy}(p) = p_y - (1+\tau) p_x - c_{xy}$ (import tax)
  - More generally, $R_{xy}(p) = p_y - C_{xy}(p_x)$.

- Pairwise stability: Because there is free entry, the prices are such that there is no positive rent on any arc, that is:

$$R_{xy}\left(p\right) \leq 0 \ \forall xy \in \mathcal{A}$$

- Note that the set of p such that $R_{xy}\left(p\right) \leq 0$ for all $xy \in \mathcal{A}$ is a sublattice of $\mathbb{R}^{\mathcal{Z}}$.
- One may want to normalize the prices at some "ground" node. In that case, we will denote the set of nodes by $\mathcal{Z}_0$ instead of $\mathcal{Z}$, where $\mathcal{Z}_0 = \tilde{\mathcal{Z}} \cup \{0\}$ is the full set of nodes, including the ground node which is 0, and $\tilde{\mathcal{Z}}$, the set of non-ground nodes.
- In the additive case, this writes

$$p_y - p_x \leq c_{xy} \ \forall xy \in \mathcal{A}.$$

- Let s_z be the exit flow, i.e. the flow leaving the network at z, and let μ_{xy} be the flow of commodity through arc xy. One has for all nodes $z \in \mathcal{Z}_0$
$$\sum_{x:xz\in\mathcal{A}} \mu_{xz} - \sum_{y:zy\in\mathcal{A}} \mu_{zy} = s_z$$
which can be rewritten
$$\nabla^\intercal \mu = s.$$
- Note that we for a feasible flow to exist, one must have
$$\sum_{z\in\mathcal{Z}_0} s_z = 0$$
for this reason it is enough to specify the exit flow for the non-ground nodes, and deduce $s_0 = -\sum_{z\in\tilde{\mathcal{Z}}} s_z$.

- No trader will operate trades between x and y at a loss. Hence

$$\mu_{xy} > 0 \implies R_{xy}(p) \geq 0$$

which combining with the requirement that p should be stable prices, yields

$$\mu_{xy} > 0 \implies R_{xy}(p) = 0$$

- This is a **complementary slackess** condition, which can be written

$$\sum_{xy \in \mathcal{A}} \mu_{xy} R_{xy}(p) = 0$$

Definition. (μ, p) is a called an equilibrium flow when the following conditions are met:
(i) $\mu \geq 0$ and $\nabla^{\intercal} \mu = s$
(ii) $R(p) \leq 0$
(iii) $\sum_{xy \in \mathcal{A}} \mu_{xy} R_{xy}(p) = 0$

- The problem above is called an equilibrium flow (EQF) problem. As we shall see, μ and p are jointly determined by a pair of coupled problems
  - Given p, μ is the solution to a linear programming problem (max flow problem)
  - Given μ, p is the solution to a dynamic programming problem (generalized shortest path problem)
- However, in the additive case, these two problems become decoupled.

- In the additive case ($R_{xy}(p) = p_y - p_x - c_{xy}$), both μ and p solve linear programming problems that are dual of eachother.
- μ solves the primal problem

$$\min_{\mu \geq 0} \sum_{xy \in \mathcal{A}} \mu_{xy} c_{xy}$$
$$s.t.\ \nabla^{\intercal} \mu = s$$

while p solves the dual problem

$$\max_{s} \sum_{z \in \mathcal{Z}} s_z$$
$$s.t.\ s_y - s_x \leq c_{xy}$$

- Cf. m+e+c_optim lectures (http://alfredgalichon.com/mec_optim/).

- Given a map $f : \mathbb{R}^d \to \mathbb{R}^d$, a nonlinear complementarity problem consists of finding

$$z \geq 0, f(z) \geq 0 \text{ and } z^\intercal f(z) = 0.$$

- This generalizes the complementarity slackness from linear programming, $\max_{x \geq 0} c^\intercal x : Ax \leq d = \min_{y \geq 0} d^\intercal y : A^\intercal y \geq c$, where

$$A^\intercal y - c \geq 0 \;\; [x \geq 0], d - Ax \geq 0 \;\; [y \geq 0]$$

Therefore, in that case, $z = (x, y)$, and $f(z) = \begin{pmatrix} 0 & A^\intercal \\ -A & 0 \end{pmatrix} + \binom{-c}{d}$.

- In the present case, $z = (\mu, p^+, p^-)$ and

$$f(z) = \left(-R\left(p^+ - p^-\right), s - \nabla^\intercal \mu, \nabla^\intercal \mu - s\right).$$

- Let $f:\mathbb{R}\rightarrow\mathbb{R}_{+}$ be a continuous function such that $f\left(-\infty\right)=0$ and $f\left(+\infty\right)=+\infty$, and let $T>0$ be a temperature parameter.
- One can look for $\mu_{xy}^{T}=f\left(R_{xy}\left(p\right)/T\right)$ as an approximation of the solution to the EQF problem. That is, look for p such that

$$\sum_{x:xz\in\mathcal{A}}f\left(R_{xz}\left(p\right)/T\right)-\sum_{y:zy\in\mathcal{A}}f\left(R_{zy}\left(p\right)/T\right)=s_{s}$$

- Therefore the system writes $E^{T}\left(p\right)=s$, where

$$E_{z}^{T}\left(p\right)=\sum_{x:xz\in\mathcal{A}}f\left(R_{xz}\left(p\right)/T\right)-\sum_{y:zy\in\mathcal{A}}f\left(R_{zy}\left(p\right)/T\right)$$

- Note that, $E^T(p)$ satisfies the weak gross substitutes property as $E_z^T(p)$ is weakly decreasing with respect to p_x for $x \neq z$.
- In particular in the differentiable case,

$$\frac{\partial E_z^T(p)}{\partial p_x} = \frac{f'\left(R_{xz}(p) / T\right)}{T} \frac{\partial R_{xz}}{\partial p_x} - \frac{f'\left(R_{zz}(p) / T\right)}{T} \frac{\partial R_{zx}}{\partial p_x} \leq 0.$$

- Consider $\left(\mu^T, p^T\right)$ where $\mu^T = f\left(R_{xy}\left(p^T\right)/T\right)$ and p^T a solution of

$$E^T\left(p^T\right) = 0$$

and assume $\mu^T \to \mu^*$ and $p^T \to p^*$ as $T \to 0$.
- Therefore, μ remains bounded, and we have $f\left(R_{xy}\left(p^T\right)/T\right) \leq K$, thus $R_{xy}\left(p^T\right) \leq Tf^{-1}(K)$, and as a result $R_{xy}(p^*) \leq 0$.
- Further, $\mu_{xy} > 0$ implies $\lim R_{xy}\left(p^T\right) = 0$, thus $R_{xy}(p^*) = 0$.

- In the additive case, $R_{xy}(p) = p_y - p_x - c_{xy}$, and

$$E_z^T(p) = \sum_{x:xz\in\mathcal{A}} f\left(\frac{p_y - p_x - c_{xy}}{T}\right) - \sum_{y:zy\in\mathcal{A}} f\left(\frac{p_y - p_x - c_{xy}}{T}\right)$$

- Let $F(z) = \int^z f(t)\,dt$, which is a convex function. We have $E_z(p) = \partial W(p)/\partial p_z$, where

$$W(p) = \sum_{xy\in\mathcal{A}} TF\left(\frac{p_y - p_x - c_{xy}}{T}\right)$$

- In particular, when $f(z) = \exp(z)$, $F(z) = \exp(z)$, and

$$W(p) = \sum_{xy \in \mathcal{A}} T \exp\left(\frac{p_y - p_x - c_{xy}}{T}\right).$$

- Similarly, when $f(z) = z^+$, we get $F(z) = z^2 1\{z \geq 0\}/2$, and

$$W(p) = \sum_{xy \in \mathcal{A}} T \left(\left(\frac{p_y - p_x - c_{xy}}{T}\right)^+\right)^2 /2$$

Section 2

Gross substitutes in equilibrium network flows

Consider a network whose nodes set is $\mathcal{Z}_0 = \mathcal{Z} \cup \{0\}$. One normalizes $p_0 = 0$.
For $p \in \mathbb{R}^{\mathcal{Z}}$, let $\Sigma(p)$ be the set of $s \in \mathbb{R}^{\mathcal{Z}}$ such that letting $\tilde{s} = \left(s, -1_{\mathcal{Z}}^{\intercal} s_z\right)$ and $\tilde{p} = (p, 0)$, there exists $\mu \in \mathbb{R}_{+}^{\mathcal{A}}$ with:
(i) $\mu \geq 0$ and $\nabla^{\intercal} \mu = \tilde{s}$
(ii) $R(\tilde{p}) \leq 0$
(iii) $\sum_{xy \in \mathcal{A}} \mu_{xy} R_{xy}(\tilde{p}) = 0$.

The following theorem is proven in G and Samuelson.
Theorem. The correspondence $p \to \Sigma(p)$ has the Multivocal Gross Substitutes property.

Proof. Let $s \in E(p)$ and $s' \in E(p')$. Then to show that the MGS property is satisfied, we need to show that there exists $s^{\vee} \in E(p \vee p')$ and $s^{\wedge} \in E(p \wedge p')$ such that for all $z \in \mathcal{Z}_0$

$$\begin{cases} 1\left\{z \in \mathcal{Z}_0^{\leq}\right\} s_z + 1\left\{z \in \mathcal{Z}_0^{>}\right\} s'_z \leq s_z^{\wedge} \text{ and} \\ 1\left\{z \in \mathcal{Z}_0^{\leq}\right\} s'_z + 1\left\{z \in \mathcal{Z}_0^{>}\right\} s'_z \geq s_z^{\vee}. \end{cases}$$

where we have defined $\mathcal{Z}_0^{\leq} = \{z \in \mathcal{Z}_0 : p_z \leq p'_z\}$ and $\mathcal{Z}_0^{>} = \{z \in \mathcal{Z}_0 : p_z > p'_z\}$.
Fact (i) $\mu_{xz} > 0$ and $p_z \leq p'_z$ implies $0 = R_{xz}(p_x, p_z) \leq R_{xz}(p_x, p'_z)$ hence $p_x \leq p'_x$.
Thus $\mu_{xz} 1\left\{z \in \mathcal{Z}_0^{\leq}\right\} \leq \mu_{xz} 1\left\{x \in \mathcal{Z}_0^{\leq}\right\}$.
(ii) $\mu'_{xz} > 0$ and $p_z > p'_z$ implies $0 = R_{xz}(p'_x, p'_z) < R_{xz}(p'_x, p_z)$, thus $p_x > p'_x$.
Thus $\mu'_{xz} 1\left\{z \in \mathcal{Z}_0^{>}\right\} \leq \mu'_{xz} 1\left\{x \in \mathcal{Z}_0^{>}\right\}$.

Proof. Now, set:
$\mu^{\wedge}_{xz} = 1\left\{x \in \mathcal{Z}_0^{\leq}\right\}\mu_{xz} + 1\left\{x \in \mathcal{Z}_0^{>}\right\}\mu'_{xz}$, and $s^{\wedge}_z = \sum_x \mu^{\wedge}_{xz} - \sum_y \mu^{\wedge}_{zy}$.
We have $\mu^{\wedge}_{xy} > 0$ implies $R\left(p \wedge p'\right) = 0$, and

$$
\begin{aligned}
s^{\wedge}_z &= \sum_x \mu^{\wedge}_{xz} - \sum_y \mu^{\wedge}_{zy} \\
&= \sum_x (1\left\{x \in \mathcal{Z}_0^{\leq}\right\}\mu_{xz} + 1\left\{x \in \mathcal{Z}_0^{>}\right\}\mu'_{xz}) \\
&- \sum_y (1\left\{z \in \mathcal{Z}_0^{\leq}\right\}\mu_{zy} + 1\left\{x \in \mathcal{Z}_0^{>}\right\}\mu'_{zy}) \\
&\geq \sum_x (1\left\{z \in \mathcal{Z}_0^{\leq}\right\}\mu_{xz} + 1\left\{z \in \mathcal{Z}_0^{>}\right\}\mu'_{xz}) \\
&- \sum_y (1\left\{z \in \mathcal{Z}_0^{\leq}\right\}\mu_{zy} + 1\left\{z \in \mathcal{Z}_0^{>}\right\}\mu'_{zy}) \\
&= 1\left\{z \in \mathcal{Z}_0^{\leq}\right\} s_z + 1\left\{z \in \mathcal{Z}_0^{>}\right\} s'_z, \; QED.
\end{aligned}
$$

A similar argument shows that $1\left\{z \in \mathcal{Z}_0^{\leq}\right\} s'_z + 1\left\{z \in \mathcal{Z}_0^{>}\right\} s'_z \geq s^{\vee}_z$.

Theorem. The correspondence $s \to \Sigma^{-1}(s)$ is isotone in Veinott's strong set order. That is, if $s \in \Sigma(p) \leq s' \in \Sigma(p')$, then $s \in \Sigma(p \wedge p')$ and $s' \in \Sigma(p \vee p')$.

Proof. Directly follows from the inverse isotonicity theorem of G and Samuelson.

Corollary. The set of equilibrium prices $E^{-1}(s)$ is a lattice.

Proof. Take $p \in \Sigma^{-1}(s)$ and $p' \in \Sigma^{-1}(s)$. Then $s \leq s$ yields $s \in \Sigma(p \wedge p')$ and $s \in \Sigma(p \vee p')$.
In the bipartite case, this theorem was first proven by Demange and Gale (1985).

Section 3

From dual to primal and conversely

- Let Γ be a subset of $\mathcal{A}$. A flow $\mu \geq 0$ is a perfect matching along Γ whenever (i) it is a feasible flow, i.e.

$$\nabla^{\intercal} \mu = s,$$

and (ii) there is now flow outisde of Γ, i.e. $\mu_a > 0 \implies a \in \Gamma$.

- Clearly, the problem of recovering the primal solution (i.e. the flow μ) based on the dual solution (i.e. the prices p) is a perfect matching – simply define

$$\Gamma = \{ a \in \mathcal{A} : R_a(p) = 0 \}.$$

- The perfect matching problem is a linear programming problem: indeed, it can be solved using

$$\min_{\mu \geq 0} \sum_a \mu_a 1 \{ a \notin \Gamma \}$$
$$s.t. \ \nabla^{\intercal} \mu = s$$

- Assume that we know $\mu_{xy} > 0$ and we would like to recover the equilibrium prices $p \in \mathbb{R}^{\mathcal{Z}_0}$ such that $p_0 = 0$, $R_{xy}(p) \leq 0$ for all xy, and $\mu_{xy} > 0$ implies $R_{xy}(p) \leq 0$.
- From the lattice representation theorem, we know that this set is a sublattice of $\mathbb{R}^{\mathcal{Z}_0}$. We would like to get the largest element of this set.
- As we shall see, this is a *dynamic programming problem*.

- Extend the set of arcs by adding the reverse of the arcs where there is a positive amount of flow, i.e. $\mathcal{A}^r = \mathcal{A} \cup \{yx : xy \in \mathcal{A},\ \mu_{xy} > 0\}$. For such reverse arcs yx, define $R_{yx}(p) = -R_{xy}(p)$. Such a network is called *reduced network*.
- See textbook treatments in Ahuja, Magnuti and Orlin (1993) and Bertsekas (1998).

- We shall restrict ourselves to the case $R_{xy}(p) = p_y - C_{xy}(p_x)$. In that case, for reverse arcs yx, we define $C_{yx}(p) = C_{xy}^{-1}(p)$.

Lemma. The set of equilibirum prices are the fixed points of an isotone map

$$T(p)_y = \min\left\{p_y, \min_{xy\in\mathcal{A}^r} C_{xy}(p_x)\right\}.$$

Proof. $T(p) = p$ if and only if $p_y \leq C_{xy}(p_x)$ for all x such that $xy \in \mathcal{A}^r$, that is

$$\begin{aligned} p_y &\leq C_{xy}(p_x), \ \forall xy \in \mathcal{A} \\ p_y &\leq C_{yx}^{-1}(p_x), \forall yx \in \mathcal{A} : \mu_{xy} > 0 \end{aligned}$$

that is

$$\begin{aligned} p_y &\leq C_{xy}(p_x), \ \forall xy \in \mathcal{A} \\ p_y &\geq C_{xy}(p_x), \forall xy \in \mathcal{A} : \mu_{xy} > 0. \end{aligned}$$

QED.

This suggests to iterate map T in order to converge to the lattice upper bound of the set of fixed points. This method is known as the Bellman-Ford algorithm, and it is an early instance of dynamic programming.

Algorithm (Bellman-Ford).

At period 1, set $p_0^1 = 0$ and $p_z^1 = +\infty$.

At period $t \geq 2$, set $p_y^t = \min \left\{ p_y^{t-1}, \min_{xy \in \mathcal{A}^r} C_{xy}\left(p_x^{t-1}\right) \right\}$

Repeat until convergence.

- In the additive case, recall that $C_{xy}(p_x) = c_{xy} + p_x$. In this case, following the approach above, we construct the reduced network by adding the reverse arcs yx to $\mathcal{A}$ whenever $\mu_{xy} > 0$. One associates these with cost $c_{yx} = -c_{xy}$.
- One seeks the largest element of the set

$$\{p : p_y - p_x \leq c_{xy} \; \forall xy \in \mathcal{A}^r, p_0 = 0\}$$

which formulates as a linear programming problem

$$\begin{aligned} &\max p_y - p_0 \\ &s.t. \; p_y - p_x \leq c_{xy} \end{aligned}$$

- The Bellman-Ford algorithm consists of deducing the optimal solution in t steps from an optimal solution in $t-1$ steps using Bellman's equation $p_y^t = \min\left\{p_y^{t-1}, \min_{xy\in\mathcal{A}^r}\left\{c_{xy} + p_x^{t-1}\right\}\right\}$.

Section 4

Bipartite case: the equilibrium transport problem

- Consider the case where $\mathcal{Z} = \mathcal{X} \cup \mathcal{Y}$, and $\mathcal{A} = \mathcal{X} \times \mathcal{Y}$. $\mathcal{X}$ are the source nodes, $\mathcal{Y}$ are the destination ones, and each source is connected to a destination.
- $n_x \geq 0$ is the mass at source $x \in \mathcal{X}$ and $m_y \geq 0$ is the mass at destination $y \in \mathcal{Y}$. Assume that the total source mass and total destination mass are the same: $\sum_x n_x = \sum_y m_y$. Set $s_z = -n_z 1\{z \in \mathcal{X}\} + m_y 1\{z \in \mathcal{Y}\}$.
- Then (μ, p) is a solution to the equilibrium transport (ET) problem if:

$$\begin{cases} \sum_{y \in \mathcal{Y}} \mu_{xy} = n_x, \sum_{x \in \mathcal{X}} \mu_{xy} = m_y \\ R_{xy}(p) \leq 0 \\ \sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}}} \mu_{xy} R_{xy}(p) = 0 \end{cases}$$

- In the bipartite case, it will often make sense to set $u_x = p_x$ and $v_y = -p_y$, and $\Psi_{xy}(u_x, v_y) = -R_{xy}(u_x, -v_y)$, so that $\Psi_{xy}(u, v)$ is increasing in u and v, and the problem becomes

$$\begin{cases} \sum_{y\in\mathcal{Y}} \mu_{xy} = n_x, \sum_{x\in\mathcal{X}} \mu_{xy} = m_y \\ \Psi_{xy}(u_x, v_y) \geq 0 \\ \sum_{\substack{x\in\mathcal{X} \\ y\in\mathcal{Y}}} \mu_{xy} \Psi_{xy}(u_x, v_y) = 0 \end{cases}$$

- Interpretation: if x and y match, they can bargain over the feasible sets of utilities (u_x, v_y) such that $\Psi_{xy}(u_x, v_y) \leq 0$.

- Note that if $R_{xy}(p) = p_y - C_{xy}(p_x)$, then $\Psi_{xy}(u_x, v_y) = C_{xy}(u_x) + v_y = v_y - \mathbb{V}_{xy}(u_x)$ where $\mathbb{V}_{xy}(u_x) = -C_{xy}(u_x)$ is continuous and decreasing.
- If (μ, u, v) is a solution to the ET problem in the previous formulation, then the following conjugation relation holds

$$\begin{cases} v_y = \max_{x \in \mathcal{X}} \mathbb{V}_{xy}(u_x) \\ u_x = \max_{y \in \mathcal{Y}} \mathbb{U}_{xy}(v_y) \end{cases}$$

- This relation is called a Galois connection, see Noeldeke and Samuelson (2017). In particular, if $\mathbb{V}_{xy}(u_x) = \Phi_{xy} - u_x$, then v is the Φ-convex conjugate of u, as studied in Villani (2008), and if $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$ and $\Phi_{xy} = x^\intercal y$, then v is the Legendre-French transform of u_x.

- Assuming everything is smooth, and letting f_P and f_Q be the densities of P and Q we have under some conditions that the equilibrium transportation plan is given by $y = T(x)$, where mass balance yields

$$\left|\det DT(x)\right| = \frac{f_P(x)}{f_Q(T(x))}$$

and optimality in $\max_{x\in\mathcal{X}} \mathbb{V}_{xy}(u(x))$ yieds

$$\partial_x \mathbb{V}_{xT(x)}(u(x)) + \partial_u \mathbb{V}_{xT(x)}(u(x)) \nabla u(x) = 0$$

which thus inverts into

$$T(x) = e(x, u(x), \nabla u(x)).$$

- In the case when $\mathbb{V}_{xy}(u(x)) = x^\intercal y - u(x)$, we get $e(x, u(x), \nabla u(x))$; in the case when $\mathbb{V}_{xy}(u(x)) = \Phi(x,y) - u(x)$, we get $e(x, u(x), \nabla u(x)) = \nabla_x \Phi(x, .)^{-1}(\nabla u(x))$.
- Trudinger (2014) studies Monge-Ampere equations in u of the form

$$\left|\det De(., u, \nabla u)\right| = \frac{f_P}{f_Q(e(., u, \nabla u))}.$$

- When $\Psi_{xy}(u_x, v_y) = u_x + v_y - \Phi_{xy}$, the problem writes

$$\begin{cases} \sum_{y\in\mathcal{Y}} \mu_{xy} = n_x, \sum_{x\in\mathcal{X}} \mu_{xy} = m_y \\ u_x + v_y \geq \Phi_{xy} \\ \mu_{xy} > 0 \implies u_x + v_y = \Phi_{xy} \end{cases}$$

- This are the complementary slackness conditions associated with the optimal transport problem, namely

$$\max_{\mu\geq 0} \sum \mu_{xy}\Phi_{xy}$$
$$s.t. \sum_{y\in\mathcal{Y}} \mu_{xy} = n_x, \sum_{x\in\mathcal{X}} \mu_{xy} = m_y$$

  which has dual

$$\min_{u,v} \sum_{x\in\mathcal{X}} n_x u_x + \sum_{y\in\mathcal{Y}} m_y v_y$$
$$s.t.\ u_x + v_y \geq \Phi_{xy}$$

- Many result extend beyond $\mathcal{X}$ and $\mathcal{Y}$ discrete; the theory is called optimal transport theory.

- Consider now the case when $\sum_x n_x \neq \sum_y m_y$. Then define $\tilde{\mathcal{Z}} = \mathcal{X} \cup \mathcal{Y}$, and add a ground node 0. Let $\mathcal{Z}_0 = \mathcal{X} \cup \mathcal{Y} \cup \{0\}$, and let

$$s_z = -n_z 1\{z \in \mathcal{X}\} + m_y 1\{z \in \mathcal{Y}\} + \left(\sum_{y \in \mathcal{Y}} m_y - \sum_{x \in \mathcal{X}} n_x\right) 1\{z = 0\}.$$

- The set of arcs is now $\mathcal{A} = \mathcal{X} \times \mathcal{Y} \cup \mathcal{X} \times \{0\} \cup \{0\} \times \mathcal{Y}$. We set $p_0 = 0$, so that and the problem becomes

$$\left\{ \begin{array}{l} \sum_{y \in \mathcal{Y}} \mu_{xy} + \mu_{x0} = n_x, \sum_{x \in \mathcal{X}} \mu_{xy} + \mu_{0y} = m_y \\ C_{xy}(p) \leq 0, C_{x0}(p_x, 0) \leq 0, C_{0y}(0, p_y) \leq 0 \\ \sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}}} \mu_{xy} C_{xy}(p) = 0 \end{array} \right.$$

- We can always redefine the problem by setting $u_x = -R_{x0}(p_x, 0)$ and $v_y = -R_{0y}(0, p_y)$, and $\Psi_{xy}(u_x, v_y) = -R_{xy}\left(R_{x0}(.,0)^{-1}(-u_x), R_{0y}(0,.)^{-1}(-v_y)\right)$, which becomes

$$\left\{ \begin{array}{l} \sum_{y \in \mathcal{Y}} \mu_{xy} + \mu_{x0} = n_x, \sum_{x \in \mathcal{X}} \mu_{xy} + \mu_{0y} = m_y \\ \Psi_{xy}(u_x, v_y) \geq 0, u_x \geq 0, v_y \geq 0 \\ \sum_{xy} \mu_{xy} \Psi_{xy}(u_x, v_y) + \sum_x \mu_{x0} u_x + \sum_y \mu_{0y} v_y = 0 \end{array} \right.$$

Section 5

Strassen's theorem

- Consider $\mathcal{X}$ and $\mathcal{Y}$ two open subsets of respectively $\mathbb{R}^d$ and $\mathbb{R}^{d'}$. Let Γ be a closed subset of $\mathcal{X} \times \mathcal{Y}$, which stand for the set of pairs (x, y) that are compatible.
- For $x \in \mathcal{X}$, denote $\Gamma(x) == \{y \in \mathcal{Y} : (x, y) \in \Gamma\}$ the subset of receivers $y \in \mathcal{Y}$ that are compatible with donor x. Γ is a *set-valued function*, or *correspondence*. For $B \subseteq \mathcal{X}$, denote

$$\Gamma(B) = \{y \in \mathcal{Y} : \exists x \in B, (x, y) \in \Gamma\}.$$

- The problem of maximizing the number of compatible pairs is given by

$$\max_{\pi \in \mathcal{M}(P,Q)} \Pr_\pi ((X, Y) \in \Gamma)$$

or equivalently

$$\min_{\pi \in \mathcal{M}(P,Q)} \Pr_\pi ((X, Y) \notin \Gamma).$$

This is an optimal transport problem with with $0 - 1$ cost (or $0 - 1$ surplus).

- By the Monge-Kantorovich theorem, the previous problem coincides with

$$= \sup \int a(x)\, dP(x) - \int b(y)\, dQ(y)$$
$$s.t.\ a(x) - b(y) \leq 1\{(x,y) \notin \Gamma\}$$

- We will see that we can take a and b valued in $\{0,1\}$. Then $a(x) = 1\{x \in A\}$ and $b(y) = 1\{y \in B\}$, so that the constraint rewrites
$$1\{y \notin B\} \leq 1\{(x,y) \notin \Gamma\} + 1\{x \notin A\}$$
which means that if $y \in \Gamma(x)$ and $x \in A$ implies $y \in B$, that is $\Gamma(A) \subseteq B$. Therefore,
$$= \sup_{A,B} \{P(A) - Q(B) : \Gamma(A) \subseteq B\},$$
hence:
- **Theorem** (Strassen). *Let P and Q be two probability measures on $\mathcal{X}$ and $\mathcal{Y}$, and let $\Gamma : \mathcal{X} \rightrightarrows \mathcal{Y}$ be a closed correspondence. Then*
$$\min_{\pi \in \mathcal{M}(P,Q)} \Pr_\pi((X,Y) \notin \Gamma) = \sup_{A \subseteq \mathcal{X}} \{P(A) - Q(\Gamma(A))\}. \quad (1)$$
where the sup on the right hand side extends to all closed subsets of $\mathcal{X}$

- Let a and b a pair of solutions to the dual problem. Then

$$a(x) = \min_{y \in \mathcal{Y}} \left\{ 1\left\{ (x, y) \notin \Gamma \right\} + b(y) \right\}$$
$$b(y) = \max_{x \in \mathcal{X}} \left\{ a(x) - 1\left\{ (x, y) \notin \Gamma \right\} \right\}$$

- Step 1: a and b valued in $[0, 1]$. One can take wlog $\min_y b(y) = 0$. It follows from $0 \leq 1\{(x, y) \notin \Gamma\} \leq 1$ and the first equality that

$$0 \leq \min_y \left\{ 1\left\{ (x, y) \notin \Gamma \right\} \right\} \leq a(x) \leq 1 + \min_y b(y) = 1$$

Simlarily, it follows from $a(x) \leq 1$ and the second inequality that

$$b(y) \leq 1.$$

- Step 2: a and b can be taken valued in $\{0,1\}$. Indeed, $a(x) = \int_0^1 1\{t \leq a(x)\}\,dt$ and $b(y) = \int_0^1 1\{t \leq b(y)\}\,dt$. Let us show that $1\{t \leq a(x)\} - 1\{t \leq b(y)\} \leq 1\{(x,y) \notin \Gamma\}$. By contradiction, if not, then $1\{(x,y) \notin \Gamma\} = 0$, $b(y) > t$ and $t \leq a(x)$. But this implies $a(x) - b(y) > 0$, yet $a(x) - b(y) \leq 1\{(x,y) \notin \Gamma\} = 0$, a contradiction.
- Next, each of $a_t(x) = 1\{t \leq a(x)\}$ and $b_t(y) = 1\{t \leq b(y)\}$ are feasible, and their convex combination is optimal for the dual; thus each of them is optimal. QED.

- Hall's marriage lemma: assume there are n donors $i \in \{1, ..., n\}$ and receivers $j \in \{1, ..., n\}$. Let $\Gamma(i) \subseteq \{1, ..., n\}$ be the set of receivers which are compatible with donors i, and for $A \subseteq \{1, ..., n\}$, define $\Gamma(A) = \cup_{i \in A} \Gamma(i)$. A (pure) matching is a permutation σ such that $j = \sigma(i)$ means that i donates to j. A matching is perfect if $\sigma(i) \in \Gamma(ij)$ for all $i \in \{1, ..., n\}$. Hall's theorem says that there is a perfect matching if and only if

$$\forall A \subseteq \{1, ..., n\}, \quad |A| \leq |\Gamma(A)|.$$

- Follows from the previous result by taking $\mathcal{X} = \mathcal{Y} = \{1, ..., n\}$ and P and Q the uniform distributions on these sets. To do this, note that the value of the dual is zero if and only if $P(A) \leq Q(\Gamma(A))$ for all $A \subseteq \mathcal{X}$.
- As for the primal, we'll need to show it has a Monge solution.

- There is a perfect matching iff the value of the (primal) problem is zero:
    - $\Longrightarrow$ is obvious.
    - For $\Leftarrow$, if the value of the problem is zero, there exists $\pi \in \mathcal{M}(P, Q)$ such that $\sum \pi_{ij} 1\{i \notin \Gamma(j)\} = 0$. One can show that w.l.o.g. π can be taken such that $\pi_{ij} = 1\{i = \sigma(j)\}/n$.
- To show the latter, consider among the matrices $\pi \in \mathcal{M}(P, Q)$ with $\sum \pi_{ij} 1\{i \notin \Gamma(j)\} = 0$ the one such that $n\pi$ has the smallest number of noninteger cells.
    - Assume that this number is > 0. Then start with one noninteger cell. There is another noninteger cell on the same line; on the same column of that cell, there is another one; on the line of the latter, another one; etc. At some point, we'll get a cycle. It's possible to strictly decrease the number of noninteger entries of $n\pi$ by removing enough mass on that cycle.
- The previous argument is (in disguise) the Birkhoff-von Neumann theorem: any coupling between the uniform distribution over $\{1, ..., n\}$ and itself can be written as a convex combination of Monge couplings between these distributions.

- Assume that we observe the marginal tax rate of invididuals, which allow us to deduce the income bracket of the individual. This is a random set $\mathbb{X}$. A model predicts the distribution of the income of individuals. If θ is the models' parameter, then $X \sim P_\theta$ is the predicted distribution of income.
- The identified parameter set is the set of θ such that there exists a joint distribution of $\mathbb{X}$ and of $X \sim P_\theta$ such that $X \in \mathbb{X}$ holds almost surely.
- When does this happen? answer using Arstein's theorem. To do this, we need first to describe the distribution of a random set.

- Assume for a minute $\mathcal{X}$ is finite. Then the distribution of $\mathbb{X}$ is characterised by $\pi_{\mathbb{X}}(A) = \Pr(\mathbb{X} = A)$ for each $A \subseteq \mathcal{X}$. This is OK when $\mathcal{X}$ is finite; however, it does not extend well beyond that case. Instead, define the *capacity* of $\mathbb{X}$ as $c_{\mathbb{X}}(A) = \Pr(\mathbb{X} \cap A \neq \emptyset)$.
- The capacity characterizes the distribution of $\mathbb{X}$; we denote $\mathbb{X} \approx c$. In the finite case, one can recover $\pi_{\mathbb{X}}(A)$ from $c_{\mathbb{X}}$ by the Möbius inversion formula

$$\pi_{\mathbb{X}}(A) = \sum_{S \subseteq A} (-1)^{|A \backslash S|} \left(1 - c\left(\mathcal{X} \backslash A\right)\right)$$

- Therefore, it is equivalent to describe the distribution of $\mathbb{X}$ in terms of $\pi_{\mathbb{X}}$ or in terms of $c_{\mathbb{X}}$.

- Arstein's theorem says that one can find a coupling such that $X \in \mathbb{X}$ if and only if
$$\forall A \text{ closed subset of } \mathcal{X}, P(A) \leq c(A).$$
- To do this, assume (actually w.l.o.g.) that $\mathbb{X} = {}^{-1}(Y)$, where $Y \sim Q$ and Γ is a correspondence. Then
$$c(A) = \Pr\left({}^{-1}(Y) \cap A \neq \emptyset\right) = \Pr(Y \in \Gamma(A)) = Q(\Gamma(A)),$$
so that Strassen's theorem readily applies.