Gross substitutes, optimal transport, and matching models

Alfred Galichon (NYU and SciencesPo)
Summer School in Optimal Transport
University of Washington, Seattle

June 2022

Support from European Research Council consolidator grant EQUIPRICE, ERC-CoG No. 866274 is gratefully acknowledged.

Introduction

Practical information

- Material available on:
  https://github.com/alfredgalichon/seattle-2022

- Schedule: Monday 10:30am-12pm, Tuesday 2pm-3:30pm, Wednesday 2pm-3:30pm

About these lectures

These lectures are about *substitutability*. Substitutability is a fundamental property of optimal transport models, although it is less recognized than convexity.

Consider the problem of computing an economic equilibrium

$$Q\left(p\right) = 0$$

where $Q : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is an excess supply function (supply minus demand for each of the n goods), and p is the price vector of each goods.

Grossly speaking, there are only two types of economic models that can be conveniently computed:

- models that rely on convexity, i.e. models where $Q = \nabla U\left(p\right)$ and U is a convex function. The problem reformulated as an optimization problem (minimization of U) and *descent methods*, such as the standard gradient descent, will work:

$$p_z^{t+1} = p_z^t - \epsilon Q_z \left(p^t\right).$$

- models that rely on substitutability – essentially the idea that Q_z is increasing in p_z and decreasing in p_x for $x \neq y$. In this case, *coordinate update methods*, such as nonlinear Jacobi, where we set the price of good z to clear the market for good z will work:

$$Q_z \left(p_z^{t+1}, p_{-z}^t\right) = 0.$$

By an extraordinary coincidence, optimal transport inherits both structures. For this reason, one can compute the OT problem and its regularizations both by descent methods and by coordinate update methods (which is called Sinkhorn's algorithm).

Some matching models can be computed using optimal transport. These are called transferable utility models, and assume that everyone's valuations are expressed in the same monetary unit. In that case, equilibrium in matching problems is the solution to an optimization problem. However, in many other cases, this assumption cannot be made. This includes:

- labor economics: one dollar of the firm not not have the same value as one dollar for the employee (pre-tax or post-tax; decreasing marginal utility, etc)

- family economics: partners may transfer utility by allocating public and private expenditures that can be inefficient

- school choice problems: utility is cardinal and cannot be transfered

We will see mathematically that these models *cannot* be recast as an optimization problem, and so the convex optimization structure is lost. However, we will also see that these models can be reformulated as models with substitutability, and therefore a lot of structure and computational methods can be deduced.

This leads us to start with a rather detailed study of substitutability and its mathematical counterpart, *Z- and M-functions*. There was a vibrant literature on the topic in the 1970s (Birkhoff, Rheinboldt, Ortega, More, Porsching, and some others), see in particular Rheinboldt and Ortega's 1970 book – but this literature has somehow fallen out of attention. In our first lecture, we will unearth some of the main results from that literature, and provide some new ones that we will need.

In the second lecture, we will connect with models of matching with transfers, departing from the assumption of perfectly transferable utility and we will appeal to the machinery developped in the first lecture.

In the third lecture, we will revisit the famous theory by Gale and Shapley of "stable marriages", which is the theory of matchings without transfers, and we will reinterpret the famous deferred acceptance algorithm as a damped version of Jacobi's algorithm. We will revisit results on the lattice structure of stable matchings.

We hope these three lectures will provide a solid mathematical introduction to matching models in economics and econometrics.

A roadmap

Lecture 1. Introduction to gross substitutes

M-matrices and M-maps, nonlinear Perron-Froebenius theory, convergence of Jacobi algorithm. A toy hedonic model.

Lecture 2. Models of matching with transfers

Problem formulation, regularized and unregularized case. IPFP and its convergence. Existence and uniqueness of an equilibrium. Lattice structure.

Lecture 3. Models of matching without transfers

Gale and Shapley's stable matchings. Adachi's formulation. Kelso-Craford. Hatfield-Milgrom.

1 Lecture 1: M-maps

Reference

Reference for this lecture include the following published material:

- Ortega, Rheinboldt (1970). *Iterative Solution of Nonlinear Equations in Several Variables*. SIAM.
- Berry, Gandhi and Haile (2013). Connected substitutes and the invertibility of demand. *Econometrica*.
- Galichon, Kominers, Weber (2019). Costly Concessions: An Empirical Framework for Matching with Imperfectly Transferable Utility. *Journal of Political Economy*.

This lecture also significantly draws on current research with co-authors, in particular the following work in progress:

- [GSV] Galichon, Samuelson and Vernet (2022). Monotone comparative statics for equilibrium problems.

- [GL] Galichon and Léger (2022). Z-mappings.

- [CGLW] Choo, Galichon, Liang, and Weber, Existence and Uniqueness in Matching Function Equilibria with Full Assignment

1.1 Sinkhorn's algorithm as Jacobi's algorithm

Consider the regularized optimal transport problem

$$\begin{aligned} \max_{\mu \geq 0} \quad & \sum_{xy} \mu_{xy} \Phi_{xy} - \sum_{xy} \mu_{xy} \ln \mu_{xy} \\ & \sum_{y \in Y} \mu_{xy} = n_x \\ & \sum_{x \in X} \mu_{xy} = m_y \end{aligned}$$

which has solutions of form

$$\mu_{xy} = \exp\left(\Phi_{xy} - a_x - b_y\right).$$

Note that Sinkhorn's algorithm consists of coming up with an initial guess of a^0 and b^0 and setting

$$\begin{aligned} a_x^{t+1} & : \; n_x = \sum_y \exp\left(\Phi_{xy} - a_x^{t+1} - b_y^t\right) \\ b_y^{t+1} & : \; m_y = \sum_y \exp\left(\Phi_{xy} - a_x^t - b_y^{t+1}\right) \end{aligned}$$

and iterating until approximate convergence is reached. As we have written things, a^{2t} and b^{2t+1} depend on the starting point a^0, while a^{2t+1} and b^{2t} depend on the starting point b^0.

Note that in the first part of the update, a_x is an antitone (=weakly monotone with respect to the coordinatewise order) function of (b_y), while in the second part of the update, b_y is an antitone function of (a_x).

Let us change the sign on one side and define

$$p_x = -a_x, p_y = b_y$$

and define

$$\begin{aligned} Q_x(p) &= \sum_y \exp\left(\Phi_{xy} + p_x - p_y\right) - n_x \\ Q_y(p) &= -\sum_x \exp\left(\Phi_{xy} + p_x - p_y\right) + m_y \end{aligned}$$

so that Sinkhorn's algorithm rewrites as *Jacobi's algorithm*

$$\begin{aligned} p_x^{t+1} &: \; Q_x\left(p_x^{t+1}, p_{-x}^t\right) = 0 \\ p_y^{t+1} &: \; Q_y\left(p_y^{t+1}, p_{-y}^t\right) = 0 \end{aligned}$$

where by p_{-z} we denote the vector of all entries but z, and by (a, p_{-z}) the vector p where the zth entry has been replaced by a.

Our goal in this lecture is the study of Jacobi's algorithm in order to solve

$$Q(p) = 0.$$

Note that the $X + Y$ equations are dependent, so we may drop one – say the one corresponding to y_0, and normalize the corresponding entry of $p_{y_0} = 0$.

1.2 Jacobi's algorithm

We consider quite generally a system of nonlinear equations of the form

$$Q(p) = 0$$

where $Q : \mathbb{R}^Z \to \mathbb{R}^Z$, with $|Z| = n$.

Economic interpretation. There are n goods, and p_z is the price for good z. The demand for good z is a function $D_z(p)$ which depends on all the prices. The supply for good z is similarly a function $S_z(p)$. We have

$$Q_z(p) = S_z(p) - D_z(p)$$

which is interpreted as the excess supply for good z.

Assume p^t is a subsolution of $Q = 0$, where we have defined:

Definition 1.1. *We say* p *is a subsolution of* $Q = 0$ *when*

$$Q(p) \leq 0;$$

we say it is a supersolution when

$$Q(p) \geq 0$$

and a solution when it is both a super and a subsolution, i.e. when $Q(p) = 0$*.*

Broadly speaking, we expect to intepret a subsolution as "prices are too low" and a supersolution as "prices are too high". More on this later.

Recall Jacobi's algorithm: take an initial guess p^0 and set

$$\begin{aligned} p_x^{t+1} &: \; Q_x\left(p_x^{t+1}, p_{-x}^t\right) = 0 \\ p_y^{t+1} &: \; Q_y\left(p_y^{t+1}, p_{-y}^t\right) = 0 \end{aligned}$$

The relationship between p^t and p^{t+1} is called the coordinate update operator and is denoted

$$p^{t+1} = T\left(p^t\right).$$

Economic interpretation. Each good has an auctioneer. At each period, the auctioneer in charge of good z sets the price for good z in order to clear that market as if the other prices did not change. Re-iterate at every period.

Several questions arise about the next iterate p^{t+1} and the sequence $\left(p^t\right)$:

* is p^{t+1} larger than p^t?

* is p^{t+1} a subsolution?

* does the sequence $\left(p^t\right)$ converge towards a solution?

We will introduce the relevant assumptions on Q to answer these questions affirmatively.

Throughout we shall assume:

Continuity assumption: Q is continuous.

1.2.1 Increasing iterates

Assume p^t is a subsolution, that is

$$Q_z\left(p_z^t, p_{-z}^t\right) \leq 0$$

and assume p^{t+1} exists so that

$$Q_z\left(p_z^{t+1}, p_{-z}^t\right) = 0.$$

In order to have $p_z^t \leq p_z^{t+1}$, one needs to have:

Diagonal isotonicity assumption: $Q_z\left(., p_{-z}\right)$ increasing.

This allows us to redefine Jacobi as

$$p_z^t = \inf\left\{\pi : Q_z\left(\pi, p_{-z}^t\right) \geq 0\right\}$$

where the maximum is taken in $\mathbb{R} \cup \{+\infty\}$, with the convention $\inf \emptyset = +\infty$.

Proposition 1.1. *Assume Q is continuous and diagonal isotone, if p^t is a subsolution, then $p^t \leq p^{t+1}$.*

1.2.2 Stable subsolution

We have

$$\begin{aligned} Q_z\left(p_z^t, p_{-z}^t\right) &\leq 0 \\ Q_z\left(p_z^{t+1}, p_{-z}^t\right) &= 0 \end{aligned}$$

therefore we see that if we want that p^{t+1} should be a subsolution, then one should have:

Definition 1.2. *One says that Q is a Z-function when*

$$Q_z\left(p_z, p_{-z}\right) \text{ is antitone with respect to } \left(p_{-z}\right).$$

Alternatively, one says Q has the gross susbstitutes property.

Proposition 1.2. *Assume Q is a continuous and diagonal isotone Z-function. Then if p^t is a subsolution, $p^t \leq p^{t+1}$ and p^{t+1} is also a subsolution.*

Is this sufficient for a solution to exist? No, as the Jacobi sequence may diverge. Consider the following example.

Example 1.1. *Consider a linear $Q(p)$ given by*

$$Q(p) = \begin{pmatrix} 1 & -2 \\ -2 & 1 \end{pmatrix} p$$

We see that the Jacobi sequence can be written explicitely as

$$\begin{array}{rcl} p_1^t & = & 2^t p_2 \\ p_2^t & = & 2^t p_1 \end{array}$$

and $p^0 = (1, 1)$ is a subsolution but the Jacobi sequence does not converge to the solution.

In fact, any Jacobi sequence starting from a subsolution other than the solution $(0, 0)$. will diverge – or converge to $(+\infty, +\infty)$.

1.2.3 Existence of solution, part 1

Proposition 1.3. *Assume Q is a continuous and diagonal isotone Z-function. Consider a Jacobi sequence p^t starting from a subsolution p^0. Then if p^t is bounded above,* *it converges toward a solution p^* of $Q = 0$. Moreover, the set of solutions of $Q = 0$ is a lattice and p^* is its lattice lower bound.*

Proof. If p^t is bounded above, then it converges towards p^*. We have $Q_z\left(p_z^{t+1}, p_{-z}^t\right) = 0$ so by continuity, $Q(p^*) = 0$.

Assume $Q(p) = Q(p') = 0$. Note that $Q(p \wedge p') \geq 0 \geq Q(p \vee p')$. We need to show that there is a $p \sqcap p'$ such that $Q(p \sqcap p') = 0$ and for any p'' such that $Q(p'') = 0$, and $p'' \leq p$ and $p'' \leq p'$, we have $p'' \leq p \sqcap p'$. $p \sqcap p'$ is obtained as the limit of the Jacobi sequence starting from $p \wedge p'$, thus $p \sqcap p' \leq p \wedge p'$ in general. Similarly, $p \sqcup p'$ is obtained as the limit of the Jacobi sequence starting from $p \vee p'$. ■

1.2.4 M0- and M-functions

The previous example shows us what went wrong: in that example, the mapping $Q(.)$ allowed for an inversion by allowing the subsolution $(1,1)$ to be above the solution $(0,0)$.Recall, we would expect subsolution to mean “prices are too low” and supersolutions “prices are too high”. Economically, we would not want the substitution effect to be stronger than the price effect. We should thus rule out such inversions. The following notions are defined in [GSV].

Definition 1.3. *Consider* $Q : \mathbb{R}^n \to \mathbb{R}^n$*. We introduce the following notions:*

(i) Q *is weakly point-nonreversing if*

$$\left\{ \begin{array}{c} Q(p) \leq Q(p') \\ p \geq p' \end{array} \right. \quad \text{implies } Q(p) = Q\left(p'\right).$$

(ii) Q *is strongly point-nonreversing if*

$$\left\{ \begin{array}{c} Q(p) \leq Q(p') \\ p \geq p' \end{array} \right. \quad \text{implies } p = p'.$$

Example 1.2. *Consider* $Q(p) = Qp$. *If* Q *is strictly diagonally dominant, then* Q *is strongly point-nonreversing.If it is weakly diagonally dominant, then* Q *is weakly point-nonreversing. Indeed, strict (resp. weak) diagonal dominance implies that* $1^\top Q \geq 0$ *(resp.* $1Q >> 0$*).*

Example 1.3. *Consider* Q *such that* $1^\top Q(p)$ *is weakly isotone in* p. *Then it is weakly point-nonreversing. In particular, this is the case of* $Q : \mathbb{R}^{X \cup Y} \to \mathbb{R}^{X \cup Y}$ *given by*

$$\begin{aligned} Q_x(p) &= \sum_y \exp(\Phi_{xy} + p_x - p_y) - n_x, x \in X \\ Q_y(p) &= -\sum_x \exp(\Phi_{xy} + p_x - p_y) + m_y, y \in Y. \end{aligned}$$

Example 1.4. *Consider* Q *such that* $1^\top Q(p) \leq 1^\top Q(p')$ *and* $p \geq p'$ *implies* $p = p'$. *Then it is strongly point-nonreversing. In particular, this is the case of*

$$Q : \mathbb{R}^{X \cup Y \setminus \{y_0\}} \to \mathbb{R}^{X \cup Y \setminus \{y_0\}}$$

given by

$$\begin{aligned} Q_x(p) &= \sum_y \exp(\Phi_{xy} + p_x - p_y) - n_x, x \in X \\ Q_y(p) &= -\sum_x \exp(\Phi_{xy} + p_x - p_y) + m_y, y \in Y \setminus \{y_0\}. \end{aligned}$$

We next move on to the definition of M- and M0-functions:

Definition 1.4. *A M-function is a Z-function which is strongly point-nonreversing. A M0-function is a Z-function which is weakly point nonreversing.*

We are able to state the following result obtained in [GSV]:

Theorem 1.1 (Inverse isotonicity theorems)**.** *Consider Q a Z-function. Then*

(i) Q is a M-function if and only if it is inverse isotone, i.e.

$$Q\left(p\right) \leq Q\left(p'\right) \text{ implies } p \leq p'.$$

(ii) Q is a M0 function if and only if the set valued map Q^{-1} is isotone in the strong set order, i.e.

$$Q\left(p\right) \leq Q\left(p'\right) \text{ implies } \left\{ \begin{array}{c} Q\left(p\right) = Q\left(p \wedge p'\right) \\ \text{and} \\ Q\left(p'\right) = Q\left(p \vee p'\right) \end{array} \right.$$

Proof. (i) To show the direct implication, assume Q is a M-function and $Q(p) \leq Q(p')$. Then we have

$$\begin{aligned} & Q_z\left(p \vee p'\right) \\ \leq\ & 1\left\{p_z \leq p'_z\right\} Q\left(p'\right) + 1\left\{p_z > p'_z\right\} Q(p) \\ \leq\ & Q\left(p'\right) \end{aligned}$$

and therefore we conclude that $p' = p \vee p'$, and thus $p \leq p'$. Conversely, assume Q is inverse isotone, and show it is strongly point-nonreversing. Assume $Q(p) \leq Q(p')$ and $p \geq p'$. By inverse isotonicity, one has $p \leq p'$, and hence $p = p'$.

(ii) To show the direct implication, assume Q is a M0-function and $Q(p) \leq Q(p')$. Then we have

$$\begin{aligned} & Q_z\left(p \vee p'\right) \\ \leq\ & 1\left\{p_z \leq p'_z\right\} Q\left(p'\right) + 1\left\{p_z > p'_z\right\} Q(p) \\ \leq\ & Q\left(p'\right) \end{aligned}$$

and therefore we conclude that $Q(p \vee p') = Q(p')$. Similarly, we have

$$\begin{aligned} & Q(p) \\ \leq \;& 1\left\{p_z \leq p'_z\right\} Q(p) + 1\left\{p_z > p'_z\right\} Q\left(p'\right) \\ \leq \;& Q\left(p'\right) \leq Q_z\left(p \wedge p'\right) \end{aligned}$$

and thus $Q(p \wedge p') = Q(p)$. Conversely, assume Q^{-1} is isotone in the strong set order, and show it is weakly point-nonreversing. Assume $Q(p) \leq Q(p')$ and $p \geq p'$. By inverse isotonicity, one has $Q(p) = Q(p \wedge p') = Q(p')$. ■

1.2.5 Zeros of M0-functions

The following result is obtained in [GL]:

Theorem 1.2. *Assume that Q is a continuous M0-map, that a subsolution exists and that a supersolution exists. The set of solutions is a nonempty sublattice of $\mathbb{R}^n$.*

First, let's show that if the set of solutions exists, then it is a sublattice of $\mathbb{R}^n$. We have $Q(p) = Q(p') = 0$, which implies $Q(p \wedge p') \geq 0$ and $Q(p \vee p') \leq 0$, but by weak point-nonreversingness one has $Q(p \wedge p') = Q(p \vee p') = 0$.

The proof of existence of solutions is constructive and relies on the censored Jacobi algorithm defined below:

Censored Jacobi algorithm. Start from a subsolution p^0 and consider $\bar{p}$ a supersolution. Assume that the sequence p^t has been constructed up to t, and define

$$\begin{aligned} \tilde{p}_z^{t+1} &: \quad Q_z\left(\tilde{p}_z^{t+1}, p_{-z}\right) = 0 \\ p^{t+1} &= \tilde{p}^{t+1} \wedge \bar{p}. \end{aligned}$$

where $\tilde{p}_z^{t+1}$ exists because $Q_z\left(p_z^t, p_{-z}\right) \leq 0$, and

$$Q_z(\bar{p}_z, p_{-z}) \geq Q_z(\bar{p}) \geq 0.$$

In other words, the censored Jacobi algorithm is given by

$$p^{t+1} = T(p) \wedge \bar{p}.$$

Then we have:

Proposition 1.4. *Assume Q is a continuous M0-map. Then the censored Jacobi algorithm converges to a vector p^* which is solution of $Q = 0$. Moreover, p^* is the lattice lower bound of the latter set.*

Proof. We show by induction that p^t is an monotone subsolution which is $\leq \bar{p}$. We have by induction $Q\left(p^t\right) \leq 0$, and therefore $Q\left(\tilde{p}^{t+1}\right) \leq 0$ and $p^t \leq \tilde{p}^{t+1}$. As we have $Q\left(\bar{p}\right) \geq 0$ and because Q is a M0-map, it follows that $Q\left(\tilde{p}^{t+1} \wedge \bar{p}\right) = Q\left(\tilde{p}^{t+1}\right) \leq 0$. We next show that $p^t \leq \tilde{p}^{t+1} \wedge \bar{p}$. Indeed, if $p_z^t < \bar{p}_z$, we have $p_z^t \leq \tilde{p}_z^{t+1}$ and the inequality follows. If $p_z^t = \bar{p}_z$, we have $p_z^t = \bar{p}_z \leq \tilde{p}_z^{t+1}$ and thus $\left(\tilde{p}^{t+1} \wedge \bar{p}\right)_z = \bar{p}_z$, and the inequality also follows.

Let B be the set of z such that there exists t for which $p_z^t = \bar{p}_z$. Let T be the smallest t for which $p_z^t = \bar{p}_z$ for all $z \in B$. We have for $t \geq T$ that $p_z^t = \tilde{p}_z^t < \bar{p}_z$ for $z \notin B$ and $p_z^t = \bar{p}_z$ for $z \in B$, and as a result,

$p^{t+1} = \tilde{p}^{t+1}$ for $t \geq T$. Hence the censored Jacobi algorithm coincides with the standard Jacobi algorithm starting from p^T. As a result, it converges toward p^* the lattice lower bound of $Q = 0$. ■

1.2.6 Zeros of M-functions

The following result appears in Ortega and Rheinboldt.

Theorem 1.3. *Assume that Q is a continuous M-map, that a subsolution exists and that a supersolution exists. A solution exists and is unique.*

Proof. The proof consists in considering two Jacobi sequences $\hat{p}^t$ and $\check{p}^t$ starting from respectively a sub and a supersolution. $\hat{p}^t$ is a increasing subsolution, $\check{p}^t$ is a decreasing supersolution, and by inverse isotonicty, one has $\hat{p}^t \leq \check{p}^t$, and therefore they both converge to respectively p and p'. $Q(p) = Q(p') = 0$ and inverse isotonicity applied twice yields $p = p'$. ■

The following result appears in Ortega and Rheinboldt.

Theorem 1.4. *Assume that Q is a surjective continuous M-map. Then any Jacobi sequence converges to the unique solution to $Q = 0$.*

Proof. Let p^t be the Jacobi sequence starting from p^0 and consider $\hat{p}^0 = Q^{-1}\left(Q\left(p^0\right) \wedge 0\right)$ and $\check{p}^0 = Q^{-1}\left(Q\left(p^0\right) \vee 0\right)$. Then $\hat{p}^0$ is a subsolution, $\check{p}^0$ is a supersolution, and the Jacobi sequences $\hat{p}^t$ and $\check{p}^t$ starting from these respective vectors are such that

$$\hat{p}^t \leq p^t \leq \check{p}^t$$

and as a result, the limit of these three sequences coincide. ■

1.3 Constant aggregate

Definition 1.5. *Q has constant aggregates if $1^\top Q(p) = \sum_z Q_z(p)$ is constant for all p.*

Note that this holds in the case of regularized optimal transport. Indeed, in that

$$\begin{aligned} Q_x(p) &= \sum_y \exp\left(\Phi_{xy} + p_x - p_y\right) - n_x \\ Q_y(p) &= -\sum_x \exp\left(\Phi_{xy} + p_x - p_y\right) + m_y \end{aligned}$$

and $\sum_z Q_z(p) = \sum_y m_y - \sum_x n_x$.

Proposition 1.5. *Assume Q is a Z-function such that $1^\top Q(p) = 0$, and that there is a $0 \in Z$ such that the restriction and corestriction of Q to $\mathbb{R}^{Z\backslash\{0\}}$ with the normalization $p_0 = \pi$ is a surjective M-function. Then, denoting $p(\pi)$ the solution to $Q = 0$, one has $\pi \leq \pi'$ implies $p(\pi) \leq p(\pi')$.*

Consider $Q(p) = \Delta p - Ap$ where Δ is a diagonal matrix with positive coefficients A has a zero diagonal and nonnegative coefficients, and

$$1^\top Q = 0.$$

Letting $\delta = \mathbf{\Delta 1}$, this implies $\delta^\top = \delta^\top \mathbf{\Delta}^{-1} A$, that is δ is the left Perron eigenvector associated to Perron eigenvalue 1. Thus there is a right Perron eigenvector $\lambda \geq 0$ such that $Q\lambda = 0$.

1.4 Existence of the regularized optimal transport problem

The reference for this section is [CGLW]. Consider solving

$$Q(p) = 0$$

where

$$Q : \mathbb{R}^{X \cup Y \backslash \{y_0\}} \to \mathbb{R}^{X \cup Y \backslash \{y_0\}}$$

given by

$$\begin{aligned} Q_x(p) &= \sum_y \exp(\Phi_{xy} + p_x - p_y) - n_x, x \in X \\ Q_y(p) &= -\sum_x \exp(\Phi_{xy} + p_x - p_y) + m_y, y \in Y \backslash \{y_0\}, \end{aligned}$$

and where we have normalized

$$p_{y_0} = \pi.$$

Step 1. Look for a supersolution. We have

$$\begin{aligned} Q_x(p) &= \sum_y \exp\left(\Phi_{xy} + p_x - p_y\right) - n_x \\ &\geq \exp\left(\Phi_{x0} + p_x - \pi\right) - n_x \end{aligned}$$

so for $\bar{p}_x$ such that

$$\exp\left(\Phi_{x0} + \bar{p}_x - \pi\right) = n_x$$

we have $Q_x(\bar{p}_x, p_y) \geq 0$ for all (p_y). Next, set $\bar{p}_y$ such that

$$\sum_x \exp\left(\Phi_{xy} + \bar{p}_x - \bar{p}_y\right) = m_y$$

and therefore we have

$$\begin{aligned} Q_x(\bar{p}) &\geq 0 \\ Q_y(\bar{p}) &= 0 \end{aligned}$$

Step 2. Q is a M-map, and thus the Jacobi sequence starting from $p^0 = \bar{p}$ is a decreasing supersolution. Hence, if is either convergent to a solution, or unbounded. However, because p^t is a supersolution, one has

$$-Q_{y_0}(p) = \sum_z Q\left(p^t\right) \geq 0$$

and thus

$$\sum_x \exp\left(\Phi_{xy} + p_x^t - \pi\right) \geq m_y$$

and thus all the p_x^t cannot go to $-\infty$. Denote x^* one such element such that

$$p_{x^*}^t \to p_{x^*}^\infty > -\infty.$$

We have

$$\exp\left(\Phi_{x^*y} + p_{x^*} - p_y\right) \leq \sum_x \exp\left(\Phi_{xy} + p_x - p_y\right) \leq m_y$$

thus all p_y remain bounded. Finally, we have

$$\sum_y \exp\left(\Phi_{xy} + p_x - p_y\right) \geq n_x$$

and thus all the p_x remain bounded too.

1.5 A toy hedonic model

Consider a surge pricing priblem in an uber-like environment. We have partitioned the city in a finite number of locations (say, blocks).

$x \in X =$ location of the driver

$y \in Y =$location of the passenger

Assume $z \in Z$ is the pickup location.

Assume that for a drive at x, the cost of picking up at z is c_{xz}

if the price of the ride at z is p_z the utility if the driver is $p_z - c_{xz} + \sigma\varepsilon_z$, where the vector (ε_z) is random.

if the driver does not pickup anyone, the utility is normalized to ε_0.

Assume that $(\varepsilon_z) \sim Gumbel$ and is iid. Then the probability that a driver at x will demand a ride z is

$$\frac{\exp\left(\frac{p_z - c_{xz}}{\sigma}\right)}{1 + \sum_{z'} \exp\left(\frac{p_{z'} - c_{xz'}}{\sigma}\right)}$$

Now assume that there are n_x drivers in area x, and therefore the supply for rides z is

$$S_z(p) = \sum_{x \in X} n_x \frac{\exp\left(\frac{p_z - c_{xz}}{\sigma}\right)}{1 + \sum_{z' \in Z} \exp\left(\frac{p_{z'} - c_{xz'}}{\sigma}\right)}$$

Let's study the properties of $S(p)$. Do we have gross substitutes? Well, $S_z(p) =$

$$\sum_{x \in X} \frac{n_x}{\exp\left(\frac{-p_z + c_{xz}}{\sigma}\right) + \sum_{z' \neq z} \exp\left(\frac{p_{z'} - p_z + c_{xz} - c_{xz'}}{\sigma}\right) + 1}$$

and we immediately see that $S_z(p)$ is decreasing w.r.t. $p_{z'}$ (for $z' \neq z$), and increasing with respect to p_z.

Now. let's focus on demand. This is the same as before, except for the fact that utility of a passenger at y seeking a ride in a cell z is now

$$u_{yz} - p_z + \eta_z$$

where η is iid Gumbel. The induced demand is

$$D_z(p) = \sum_{y \in Y} \frac{m_y \exp(u_{yz} - p_z)}{1 + \sum_{z'} \exp\left(u_{yz'} - p_{z'}\right)}$$

and we see that $-D_z(p)$ is a Z-matrix, and therefore

$$Q_z(p) = S_z(p) - D_z(p)$$

also has the gross substitute property.

Note that when $p_z = c$ and $c \to +\infty$, $Q_z(p) \geq 0$; while when $p_z = c$, $c \to -\infty$ we have $Q_z(p) \leq 0$. As a result, there is a unique solution.

2 Lecture 2: models of matching with transfers

Reference

Reference for today's lecture include:

- Galichon, Kominers, Weber (2019). Costly Concessions: An Empirical Framework for Matching with Imperfectly Transferable Utility. *Journal of Political Economy*.

2.1 Microfoundation of the matching model

Today we shall see models that formulate as

$$\begin{cases} \sum_{y\in\mathcal{Y}} M_{xy}(a_x, b_y) = n_x \\ \sum_{y\in\mathcal{Y}} M_{xy}(a_x, b_y) = m_y \end{cases}$$

where $M_{xy}(a_x, b_y)$ is continuous and increasing in a_x and b_y and will stand for the equilibrium number of matches between types x and y. Obviously, optimal transport falls into this category. We are actually going to consider a slight variant of the problem where we allow agents to remain unmatched, and consider

$$\begin{cases} \sum_{y\in\mathcal{Y}} M_{xy}(a_x, b_y) + M_{x0}(a_x) = n_x \\ \sum_{y\in\mathcal{Y}} M_{xy}(a_x, b_y) + M_{y0}(b_y) = m_y \end{cases}$$

where $M_{x0}(a_x)$ and $M_{y0}(b_y)$ will stand for the number of unmatched agents of types x and y.

Consider a matching model between workers (CEOs) and firms with (possibly) taxes. There are n_x workers of type

$x \in X$ and m_y firms of type $y \in Y$. Assume w_{xy} is now the gross wage paid by the firm to the worker.

If matched with firm y, a worker x gets utility

$$\alpha_{xy} + N(w_{xy}) + \sigma\varepsilon_y$$

where ε_y is Gumbel distributed and the ε_y's are independent, and if unmatched, a worker x gets utility

$$\sigma\varepsilon_0$$

As a result, the expected indirect utility of a worker of type x is

$$u_x = E\left[\max_{y \in Y} \left\{\alpha_{xy} + N(w_{xy}) + \sigma\varepsilon_y, \sigma\varepsilon_0\right\}\right]$$

where $N(w)$ is the net wage assuming the gross wage is w. Typically, N is

- increasing

- concave

- piecewise affine

In that case, we can represent $N(w)$ as

$$N(w) = \min_{k=1,\ldots,K} (1-\tau_k)(w - w_k)$$

where τ_k is the kth tax bracket: $\tau_1 = 0 < \tau_2 < ... < \tau_K$.

If there are no taxes, then we have simply $N(w) = w$.

Worker x's problem: choose a firm y such that

$$\max_{y \in Y} \{\alpha_{xy} + N(w_{xy}) + \sigma\varepsilon_y, \sigma\varepsilon_0\}$$

where α_{xy} is the monetary valuation of the job y's amenities. w_{xy} is the wage of worker x working for firm y;

determined at equilibrium. ε_y is the random utilitiy, assumed logit (Gumbel distributed). As a consequence, the average indirect utility of worker x is

$$u_x = \sigma \log\left(1 + \sum_y \exp\left(\frac{\alpha_{xy} + N\left(w_{xy}\right)}{\sigma}\right)\right)$$

and the probability that worker x picks firm y is

$$\begin{aligned}\frac{\mu_{xy}}{n_x} &= \frac{\exp\left(\frac{\alpha_{xy}+N\left(w_{xy}\right)}{\sigma}\right)}{1+\sum_{y'}\exp\left(\frac{\alpha_{xy'}+N\left(w'_{xy}\right)}{\sigma}\right)} \\ &= \exp\left(\frac{\alpha_{xy} + N\left(w_{xy}\right) - u_x}{\sigma}\right).\end{aligned}$$

Firm's side:

$$\max_{x\in X}\left\{\gamma_{xy} - w_{xy} + \sigma\eta_x\right\}$$

and

$$v_y = \sigma \log\left(1 + \sum_x \exp\left(\frac{\gamma_{xy} - w_{xy}}{\sigma}\right)\right)$$

and the probability that firm y picks worker x is

$$\frac{\mu_{xy}}{m_y} = \mathsf{exp}\left(\frac{\gamma_{xy} - w_{xy} - v_y}{\sigma}\right)$$

The wage w_{xy} is a way to transfer systematic utility from one partner to the next. After transfer, the systematic utilities of both partners are

$$\begin{aligned} \text{for } x,\ U_{xy} &\leq \alpha_{xy} + N\left(w_{xy}\right) \\ \text{for } y,\ V_{xy} &\leq \gamma_{xy} - w_{xy} \end{aligned}$$

More generally, we will consider models in which $(U, V) \in \mathcal{F}_{xy}$ belong to a feasible set $\mathcal{F}_{xy}$ – we shall make assumptions on $\mathcal{F}_{xy}$.

We'll assume free disposal: if $(U, V) \in \mathcal{F}_{xy}$ then $(U', V') \in \mathcal{F}_{xy}$ as soon as $U' \leq U$ and $V' \leq V$.

We describe $\mathcal{F}_{xy}$ by the *distance-to-frontier function*. Introduce

$$D_{xy}(U,V) = \min_{t\in\mathbb{R}} \{t : (U-t, V-t) \in \mathcal{F}_{xy}\}$$

Therefore $D_{xy}(U,V) > 0$ iff $(U,V) \notin \mathcal{F}_{xy}$.

$D_{xy}(U,V) \leq 0$ iff $(U,V) \in \mathcal{F}_{xy}$

$D_{xy}(U,V) < 0$ iff (U,V) is not on the frontier of $\mathcal{F}_{xy}$.

Example 1. Transferable utility case.

$\mathcal{F}_{xy} = \{(U,V) : U + V = \Phi_{xy}\}$ we have

$$D_{xy}(u,v) = \frac{u + v - \Phi_{xy}}{2}$$

indeed, the minimum t such that $(u-t, v-t) \in \mathcal{F}_{xy}$ is such that $(u-t)+(v-t) = \Phi_{xy}$ that is $u+v-2t = \Phi_{xy}$ and thus $t = \frac{u+v-\Phi_{xy}}{2}$.

Example 2. Nonlinear taxes.

In this case $U_{xy} \leq \alpha_{xy} + N(w_{xy})$ and $V_{xy} \leq \gamma_{xy} - w_{xy}$. We have

$$(U, V) \in \mathcal{F}_{xy} \text{ if and only } N\left(\gamma_{xy} - V\right) \geq U - \alpha_{xy}$$

that is in this case

$$\mathcal{F}_{xy} = \left\{(U, V) \in \mathbb{R}^2 : N\left(\gamma_{xy} - V\right) \geq U - \alpha_{xy}\right\}.$$

We have

$$\mathcal{F}_{xy} = \left\{(u, v) : N\left(\gamma_{xy} - v\right) \geq u - \alpha_{xy}\right\}.$$

where

$$N(w) = \min_{k=1,\ldots,K} (1-\tau_k)(w - w_k)$$

We have

$$\begin{aligned}
\mathcal{F}_{xy} &= \left\{ (u,v) : \min_{k=1,\ldots,K} (1-\tau_k)\left(\gamma_{xy} - v - w_k\right) \geq u - \alpha_{x}\right. \\
&= \cap_k \left\{ (u,v) : (1-\tau_k)\left(\gamma_{xy} - v - w_k\right) \geq u - \alpha_{xy} \right\}.
\end{aligned}$$

As the feasible set is an intersection, we need to understand how to compute the distance to the intersection of elementary sets. It turns out that we have

$$D_{F_1 \cap F_2} = \max\left\{ D_{F_1}, D_{F_2} \right\}$$

and

$$D_{F_1 \cup F_2} = \min\left\{ D_{F_1}, D_{F_2} \right\}.$$

Back to our taxation problem

$$D_{F_{xy}}(u,v) = \max_k D^k_{xy}(u,v)$$

where

$$D^k_{xy}(u,v) = \frac{u - \alpha_{xy} + (1-\tau_k)\left(v - \gamma_{xy} + w_k\right)}{2-\tau_k}.$$

Therefore

$$D_{xy} = \max_{k=1..K} \left\{ \frac{u - \alpha_{xy} + (1 - \tau_k)\left(v - \gamma_{xy} + w_k\right)}{2 - \tau_k} \right\}.$$

Note that any distance-to-frontier function D is translation-invariant, that is

$$D(u + t, v + t) = t + D(u, v).$$

Back to the matching model. Recall that we had

$$u_x = \log\left(1 + \sum_y \exp \frac{U_{xy}}{\sigma}\right)$$

and

$$v_y = \log\left(1 + \sum_x \exp \frac{V_{xy}}{\sigma}\right)$$

and

$$\begin{aligned} \frac{\mu_{xy}}{n_x} &= \exp\left(\frac{U_{xy} - u_x}{\sigma}\right) \\ \frac{\mu_{xy}}{m_y} &= \exp\left(\frac{V_{xy} - v_y}{\sigma}\right) \end{aligned}$$

therefore

$$\begin{aligned} U_{xy} &= u_x + \sigma \ln \frac{\mu_{xy}}{n_x} \\ V_{xy} &= v_y + \sigma \ln \frac{\mu_{xy}}{m_y} \end{aligned}$$

and we have

$$D_{xy}\left(U_{xy}, V_{xy}\right) = 0$$

Replacing, we get

$$D_{xy}\left(u_x + \sigma \ln \frac{\mu_{xy}}{n_x}, v_y + \sigma \ln \frac{\mu_{xy}}{m_y}\right) = 0$$

Recalling that

$$D\left(u + t, v + t\right) = t + D\left(u, v\right),$$

this yields

$$\sigma \ln \mu_{xy} + D_{xy}\left(u_x - \sigma \ln n_x, v_y - \sigma \ln m_y\right) = 0$$

that is

$$\begin{aligned}
\mu_{xy} &= \exp\left(-\frac{D_{xy}\left(u_x - \sigma \ln n_x, v_y - \sigma \ln m_y\right)}{\sigma}\right), \\
\mu_{x0} &= \exp\left(-\frac{u_x - \sigma \ln n_x}{\sigma}\right) \\
\mu_{0y} &= \exp\left(-\frac{v_y - \sigma \ln m_y}{\sigma}\right)
\end{aligned}$$

and denoting $a_x = u_x - \sigma \ln n_x$ and $b_y = v_y - \sigma \ln m_y$ we obtain

$$\begin{aligned}
\mu_{xy} &= \exp\left(-\frac{D_{xy}\left(a_x, b_y\right)}{\sigma}\right) \\
\mu_{x0} &= \exp\left(-\frac{a_x}{\sigma}\right) \\
\mu_{0y} &= \exp\left(-\frac{b_y}{\sigma}\right)
\end{aligned}$$

Recall

$$\frac{\mu_{x0}}{n_x} = \frac{\exp(0)}{\exp(0) + \sum_y \exp\frac{U_{xy}}{\sigma}} = \frac{1}{\exp\left(\frac{u_x}{\sigma}\right)} = \exp\left(-\frac{u_x}{\sigma}\right)$$

Now we just need to solve for (a_x, b_y), which we do using

$$\begin{aligned} \sum_y \mathsf{exp}\left(-\frac{D_{xy}(a_x, b_y)}{\sigma}\right) + \mathsf{exp}\left(-\frac{a_x}{\sigma}\right) &= n_x \\ \sum_x \mathsf{exp}\left(-\frac{D_{xy}(a_x, b_y)}{\sigma}\right) + \mathsf{exp}\left(-\frac{b_y}{\sigma}\right) &= m_y \end{aligned}$$

Introduce $p_x = a_x$, $p_y = -b_y$ and

$$\begin{aligned} Q_x(p) &= -\sum_y \mathsf{exp}\left(-\frac{D_{xy}(p_x, -p_y)}{\sigma}\right) + \mathsf{exp}\left(-\frac{p_x}{\sigma}\right) + n_x \\ Q_y(p) &= \sum_x \mathsf{exp}\left(-\frac{D_{xy}(p_x, -p_y)}{\sigma}\right) + \mathsf{exp}\left(\frac{p_y}{\sigma}\right) - m_y \end{aligned}$$

A comment about the optimzation structure. If this were to a gradient, one would have $Q_x(p) = \partial V(p) / \partial p_x$ and thus

$$\frac{\partial Q_x(p)}{\partial p_y} = \frac{\partial^2 V}{\partial p_x \partial p_y} = \frac{\partial Q_y(p)}{\partial p_x}$$

But we have

$$\begin{aligned}
\frac{\partial Q_x(p)}{\partial p_y} &= \exp\left(-\frac{D_{xy}(p_x,-p_y)}{\sigma}\right)\frac{\partial_{p_y} D_{xy}(p_x,-p_y)}{\sigma} \\
\frac{\partial Q_y(p)}{\partial p_x} &= \exp\left(-\frac{D_{xy}(p_x,-p_y)}{\sigma}\right)\frac{\partial_{p_x} D_{xy}(p_x,-p_y)}{\sigma}
\end{aligned}$$

Remark. Indirect utility of x is

$$\begin{aligned}
u_x &= E\left[\max_y \{U_{xy}+\varepsilon_y\}\right] \\
v_y &= E\left[\max_x \left\{V_{xy}+\eta_y\right\}\right]
\end{aligned}$$

this contrasts with a situation where J is the set of individual firms and

$$u_i = \max_{j\in J}\left\{U_{ij}+\varepsilon_{ij}\right\}$$

even if $U_{ij} = 0$, we have $u_i = \max_{j\in J}\left\{\varepsilon_{ij}\right\}$ thus $E[u_i] = \log\sum_{j=1}^{J}\exp 0 = \log J$.

Dagsvik (IER 2000) and Menzel (Econometrica 2015) have increasing returns to scale.

2.2 Connection with prescribed Jacobian equations

See discussion.