'MATH+ECON+CODE' MASTERCLASS ON MATCHING MODELS, OPTIMAL TRANSPORT AND APPLICATIONS

Alfred Galichon (New York University)

Spring 2018
Day 2, January 16 2018: "Optimal transport I"
Block 4. Discrete matching

Learning objectives: block 4

- Optimal assignment problem
- Pairwise stability, Walrasian equilibrium
- Computation

- [OTME], Ch. 3
- Roth, Sotomayor(1990). *Two-Sided Matching*. Cambridge.
- Koopmans and Beckmann (1957). "Assignment problems and the location of economic activities." *Econometrica*.
- Shapley and Shubik (1972). "The assignment game I: The core." *IJGT*.
- Becker (1993). *A Treatise of the Family*. Harvard.
- Gretsky, Ostroy, and Zame (1992). "The nonatomic assignment model." *Economic Theory*.
- Burkard, Dell'Amico, and Martello (2012). *Assignment Problems*. SIAM.
- Dupuy and Galichon (2014). "Personality traits and the marriage market." *JPE*.

Section 1

Theory

- Consider the problem of assigning a possibly infinite number of workers and firms.
  - Each worker should work for one firm, and each firm should hire one worker.
  - Workers and firms have heterogenous characteristics; let $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ be the characteristics of workers and firms respectively.
  - Workers and firms are in equal mass, which is normalized to one. The distribution of worker's types is P, and the distribution of the firm's types is Q, where P and Q are probability measures on $\mathcal{X}$ and $\mathcal{Y}$.
- It is assumed that if a worker x matches with a firm y, the total output generated is Φ_{xy}. The questions are then:
  - optimality: what is the optimal assignment in the sense that it maximizes the overal output generated?
  - equilibrium: what are the equilibrium assignment and the equilibrium wages
  - efficiency: do these two notions coincide?
- The same tools have been used by Gary Becker to study the heterosexual marriage market, where x is the man's characteristics, and y is the woman's characteristics, and "wages" are replaced by "transfers".

- In this block, we shall take a first look at marriage data (while a worker-firm example will be seen in next block). Dupuy and Galichon (JPE, 2014) study a marriage dataset where, in addition to usual socio-demographic variables (such as education and age), measures of personality traits are reported.
  - The literature on quantitative psychology argues that one can capture relatively well an individual's personality along five dimensions, the "big 5" – consciousness, extraversion, agreableness, emotional stability, autonomy – assessed though a standardized questionaire.
  - In addition to this, we observed a (self-assessed) measure of health, risk-aversion, education, height and body mass index = weight in kg/ (height in m)ˆ2. In total, the available characteristics x_i of man i and y_j of woman j are both 10-dimensional vectors.
  - It is assumed that the surplus of interaction is given by $\Phi\left(x_i, y_j\right) = x_i^\intercal A y_j$, where A is a *given* 10x10 matrix. (later in this course, we'll see how to estimate A based on matched marital data).
- Today, we solve a central planner's problem (a stylized version of the problem OKCupids would solve): given a population of men and a population of women, how do we mutually assign these in order to 1) maximize matching surplus 2) attain a (hopefully) stable assignment.

- The summary statistics are:

TABLE 2
SAMPLE OF YOUNG COUPLES WITH COMPLETE INFORMATION:
SUMMARY STATISTICS BY GENDER ($N = 1{,}158$)

	Husbands		Wives	
	Mean	Standard Error	Mean	Standard Error
Age	35.52	6.01	32.78	4.84
Educational level	2.01	.57	1.87	.57
Height	182.33	7.20	169.35	6.41
BMI	24.53	2.94	23.44	3.83
Health	3.21	.66	3.11	.69
Conscientiousness	−.25	.64	.01	.68
Extraversion	−.12	.69	.16	.60
Agreeableness	−.06	.65	−.04	.64
Emotional stability	.17	.57	−.19	.53
Autonomy	.00	.67	−.01	.69
Risk aversion	.06	.68	−.12	.88

- Assume that the type spaces $\mathcal{X}$ and $\mathcal{Y}$ are finite, so $\mathcal{X} = \{1, ..., N\}$, and $\mathcal{Y} = \{1, ..., M\}$.
- The total mass of workers and jobs is normalized to one. The mass of workers of type x is p_x; the mass of jobs of type y is q_y, with $\sum_x p_x = \sum_y q_y = 1$.
- Let π_{xy} be the mass of workers of type x assigned to jobs of type y. Every worker is busy and every job is filled, thus

$$\sum_{y \in \mathcal{Y}} \pi_{xy} = p_x \text{ and } \sum_{x \in \mathcal{X}} \pi_{xy} = q_y. \tag{1}$$

(Note that this formulation allows for mixing, i.e. it allows for $\pi_{xy} > 0$ and $\pi_{xy'} > 0$ to hold simultaneously with $y \neq y'$.) The set of $\pi \geq 0$ satisfying (1) is denoted by

$$\pi \in \mathcal{M}(p, q).$$

- Assume the economic output created when assigning worker x to job y is Φ_{xy}. Hence, under assignment π, the total output is $\sum_{xy} \pi_{xy} \Phi_{xy}$.
- Thus, the optimal assignment is

$$
\begin{aligned}
& \max_{\pi \geq 0} \sum_{xy} \pi_{xy} \Phi_{xy} \qquad (2) \\
& s.t. \ \sum_{y \in \mathcal{Y}} \pi_{xy} = p_x \ \ [u_x] \\
& \qquad \sum_{x \in \mathcal{X}} \pi_{xy} = q_y \ \ [v_y]
\end{aligned}
$$

  and it is now a finite-dimensional linear programming problem.
- Note that it is nothing else than the Monge-Kantorovich problem when P and Q are discrete probability measures on $\mathcal{X} = \{1, ..., N\}$, and $\mathcal{Y} = \{1, ..., M\}$.

THEOREM

(i) The value of the primal problem (2) coincides with the value of the dual problem

$$\min_{u,v} \sum_{x\in\mathcal{X}} p_x u_x + \sum_{y\in\mathcal{Y}} q_y v_y. \tag{3}$$

$$s.t. u_x + v_y \geq \Phi_{xy} \ \ [\pi_{xy} \geq 0]$$

(ii) Both the primal and the dual problems have optimal solutions. If π is a solution to the primal problem and (u, v) a solution to the dual problem, then by complementary slackness,

$$\pi_{xy} > 0 \textit{ implies } u_x + v_y = \Phi_{xy}. \tag{4}$$

- Note that this result is the min-cost flow duality theorem in the bipartite case, as seen in block 2, after setting transportation cost through $xy \in \mathcal{X} \times \mathcal{Y}$ to $c_{xy} = -\Phi_{xy}$, and $n_t = -p_t 1\{t \in \mathcal{X}\} + q_t 1\{t \in \mathcal{Y}\}$. We see various new interpretations of the result.

The proof follows from the min-cost flow duality result, but let us rewrite it anyway. (i) The value of the primal problem (*2*) can be written as $\max_{\pi \geq 0} \min_{u,v} S(\pi, u, v)$, where

$$S(\pi, u, v) := \sum_{xy} \pi_{xy} \Phi_{xy} + \sum_{x \in \mathcal{X}} u_x (p_x - \sum_{y \in \mathcal{Y}} \pi_{xy}) + \sum_{y \in \mathcal{Y}} v_y (q_y - \sum_{x \in \mathcal{X}} \pi_{xy})$$

but by the minmax theorem, this value is equal to $\min_{u,v} \max_{\pi \geq 0} S(\pi, u, v)$, which is the value of the dual problem (3).

(ii) follows by noting that, for a primal solution π and a dual solution (u, v), then $S(\pi, u, v) = \sum_{xy} \pi_{xy} \Phi_{xy}$. ■

- The following statements are equivalent:
  - π is an optimal solution to the primal problem, and (u, v) is an optimal solution to the dual problem, and
  - (i) $\pi \in M(p, q)$
    (ii) $u_x + v_y \geq \Phi_{xy}$
    (iii) $\pi_{xy} > 0$ implies $u_x + v_y \leq \Phi_{xy}$.
- We saw the direct implication. But the converse is easy: take π and (u, v) satisfying (i)–(iii), Then one has

$$dual \leq \sum_x p_x u_x + \sum_y q_y v_y = \sum_{xy} \pi_{xy} (u_x + v_y) \leq \sum_{xy} \pi_{xy} \Phi_{xy} \leq primal$$

but by the MK duality theorem, both ends coincide. Thus π is optimal for the primal and (u, v) for the dual.

- A important variant of the problem exists with $\sum_{x\in\mathcal{X}} p_x \neq \sum_{y\in\mathcal{Y}} q_y$ and the primal constraints become inequality constraints. The duality then becomes

$$\begin{array}{rcr} \max_{\pi\geq 0} \sum \pi_{xy}\Phi_{xy} & = & \min_{u,v} \sum_{x\in\mathcal{X}} p_x u_x + \sum_{y\in\mathcal{Y}} q_y v_y \\ s.t.\ \sum_{y\in\mathcal{Y}} \pi_{xy} \leq p_x & & u \geq 0,\ v \geq 0 \\ \sum_{x\in\mathcal{X}} \pi_{xy} \leq q_y & & u_x + v_y \geq \Phi_{xy} \end{array}$$

- In a marriage context, an important concept is stability:
  - An outcome is a vector (π, u, v), where u_x and v_y are x's and y's payoffs, and π is a matching that is
$$\pi \in \mathcal{M}(p, q). \tag{5}$$
  - A pair xy is blocking if x and y can find a way of sharing their joint surplus Φ_{xy} in such a way that x gets more than u_x and y gets more than v_y. Hence there is no blocking pair if and only if for every x and y, one has
$$u_x + v_y \geq \Phi_{xy}. \tag{6}$$
  - If x and y are actually matched, their utilities u_x and v_y need to be feasible, i.e. the above inequality should be saturated. Hence
$$\pi_{xy} > 0 \text{ implies } u_x + v_y = \Phi_{xy} \tag{7}$$
- **Definition**: A matching that satisfies (5), (6), and (7) is called a stable matching.
- As it turns out, these conditions are precisely the conditions that express complementarity slackness in the Monge-Kantorovich problem. Therefore, outcome(π, u, v) is stable if and only if π is a solution to the primal problem, and (u, v) is a solution to the dual problem.

- Back to the workers / firms interpretation and assume for now that workers are indifferent between any two firms that offer the same salary. We argue that $u(x)$ can be interpreted as the equilibrium wage of worker x, while $v(y)$ can be interpreted as the equilibrium profit of firm y. Indeed:

PROPOSITION

If (u, v) is a solution to the dual of the Kantorovich problem, then

$$u_x = \sup_{y \in \mathcal{Y}} \left(\Phi_{xy} - v_y \right) \tag{8}$$

$$v_y = \sup_{x \in \mathcal{X}} \left(\Phi_{xy} - u_x \right). \tag{9}$$

- Therefore, u_x can be interpreted as equilibrium wage of worker x, and v_y as equilibrium profit of firm y. In this interpretation, all workers get the same wage at equilibrium.

Equilibrium wages when workers are not indifferent between firms

- Assume now that if a worker of type x works for a firm of type y for wage w_{xy}, then gets $\alpha_{xy} + w_{xy}$, where α_{xy} is the nonmonetary payoff associated with working with a firm of type y. The firm's profit is $\gamma_{xy} - w_{xy}$, where γ_{xy} is the economic output.
- If an employee of type x matches with a firm of type y, they generate joint surplus Φ_{xy}, given by

$$\Phi_{xy} = \underbrace{\alpha_{xy} + w_{xy}}_{\text{employee's payoff}} + \underbrace{\gamma_{xy} - w_{xy}}_{\text{firm's payoff}} = \alpha_{xy} + \gamma_{xy}$$

which is independent from w.
- Workers choose firms which maximize their utility, i.e. solve

$$u_x = \max_y \left\{ \alpha_{xy} + w_{xy} \right\} \tag{10}$$

and $u_x = \alpha_{xy} + w_{xy}$ if x and y are matched. Similarly, the indirect payoff vector of firms is

$$v_y = \max_x \left\{ \gamma_{xy} - w_{xy} \right\} \tag{11}$$

and, again, $v_y = \gamma_{xy} - w_{xy}$ if x and y are matched.

- As a result,
$$u_x + v_y \geq \alpha_{xy} + \gamma_{xy} = \Phi_{xy}$$
and equality holds if x and y are matched. Thus, if w_{xy} is an equilibrium wage, then the triple (π, u, v) where π is the corresponding matching, and u_x and v_y are defined by (10) and (11) defines a stable outcome.
- Conversely, let (π, u, v) be a stable outcome. Then let $\bar{w}_{xx}$ and $\underline{w}_{xy}$ be defined by
$$\bar{w}_{xy} = u_x - \alpha_{xy} \text{ and } \underline{w}_{xy} = \gamma_{xy} - v_y.$$
- One has $\bar{w}_{xy} \geq \underline{w}_{xy}$. Any w_{xy} such that $\bar{w}_{xy} \geq w_{xy} \geq \underline{w}_{xy}$ is an equilibrium wage. Indeed, $\pi_{xy} > 0$ implies $\bar{w}_{xy} = \underline{w}_{xy}$, thus (10) and (11) hold. Given u and v, w_{xy} is uniquely defined on the equilibrium path (ie. when x and y are such that $\pi_{xy} > 0$), but there are multiple choices of w outside the equilibrium path.
- Note that all workers of the same type get the same indirect utility, but not necessarily the same wage.

Section 2

CODING

- The application consists of data on a population of heterosexual men and women, which includes education, height, BMI, health, the 'big 5' personality traits, and a measure of risk aversion. These data are stored in the files '`Xvals.csv`' (men) and '`Yvals.csv`' (women) on 1158 men and women. (The data comes from married households, which is why there is the same number of men and women).
- We postulate that the form of the surplus function is

$$\Phi_{ij} = x_i^\intercal A y_j$$

where x_i and y_j are the 10-dimensional characteristics of man i and woman j, and the form of A, a 10x10 matrix, is given (it is stored in the file '`affinitymatrix.csv`'). Again, we'll see later how to solve the econometrics problem of estimating A.

- This problem of computation of the Optimal Assignment Problem, more specifically of (π, u, v), is arguably the most studied problem in Computer Science, and dozens, if not hundreds of algorithms exist, whose running time is polynomial in $\max(n, m)$, typically a power less than three of the latter.
- Famous algorithms include: the Hungarian algorithm (Kuhn-Munkres); Bertsekas' auction algorithm; Goldberg and Kennedy's pseudoflow algorithm. For more on these, see the book by Burkard, Dell'Amico, and Martello, and a introductory presentation in http://www.assignmentproblems.com/doc/LSAPIntroduction.pdf.
- Here, we will show how to solve the problem with the help of a Linear Programming solver used as a black box; our challenge here will be to carefully set up the constraint matrix as a sparse matrix in order to let a large scale Linear Programming solvers such as Gurobi recognize and exploit the sparsity of the problem.

- Let Π and Φ be the matrices with typical elements (π_{xy}) and (Φ_{xy}). We let p, q, u, v, and 1 the column vectors with entries (p_x), (q_y), (u_x), (v_y), and 1, respectively. Problem (2) rewrites using matrix algebra as

$$\max_{\Pi \geq 0} Tr\left(\Pi'\Phi\right) \tag{12}$$
$$\Pi 1_M = p$$
$$1_N' \Pi = q'.$$

- We need to convert matrices into vectors; this can be done for instance by stacking the columns of Π into a single column vector (typical in R or Matlab). This operation is called *vectorization*, which we will denote

$$vec\left(A\right),$$

which reshapes a $N \times M$ matrix into a $nm \times 1$ vector. In R, this is implemented by `c(A)`; in Matlab, by `reshape(A,[n*m,1])`.

- The objective function rewrites as

$$vec\left(\Phi\right)' vec\left(\Pi\right).$$

- Recall that if A is a $M \times p$ matrix and B a $N \times q$ matrix, then the Kronecker product $A \otimes B$ of A and B is a $mn \times pq$ matrix such that
$$vec\left(BXA'\right) = (A \otimes B)\, vec\,(X)\,. \tag{13}$$
In R, $A \otimes B$ is implemented by `kronecker(A,B)`; in Matlab, by `kron(A,B)`.
- The first constraint $I_N \Pi 1_M = p$, vectorizes therefore as
$$\left(1'_M \otimes I_N\right) vec\,(\Pi) = vec\,(p)\,,$$
and similarly, the second constraint $1'_N \Pi I_M = q'$, vectorizes as
$$\left(I_M \otimes 1'_N\right) vec\,(\Pi) = vec\,(q)\,.$$
- Note that the matrix $1'_M \otimes I_N$ is of size $N \times NM$, and the matrix $I_M \otimes 1'_N$ is of size $M \times NM$; hence the full matrix involved in the left-hand side of the constraints is of size $(N+M) \times NM$. In spite of its large size, this matrix is *sparse*. In R, the identity matrix I_N is coded as `sparseMatrix(1:N,1:N)`, in Matlab as `Speye(N)`.

- Setting $z = vec\left(\Pi\right)$, the Linear Programming problem then becomes

$$\max_{z \geq 0} vec\left(\Phi\right)' z \tag{14}$$
$$s.t.\ \left(1_M' \otimes I_N\right) z = vec\left(p\right)$$
$$\left(I_M \otimes 1_N'\right) z = vec\left(q'\right)$$

  which is ready to be passed on to a linear programming solver such as Gurobi.
- A LP solver typically computes programs of the form

$$\max_{z \geq 0} c' z \tag{15}$$
$$s.t.\ Az = d.$$

  In R, Gurobi is called to compute program (15) by
  `gurobi(list(A=A,obj=c,modelsense="max",rhs=d,sense="="))`.