Advanced topics in microeconometrics: Matching Models and their Applications

Alfred Galichon (NYU and Scpo)

Spring 2018
Day 1, January 15 2018: Linear programming
Block 1. Linear programming duality

- Schedule: Monday 8am-10am.
- Dates: Jan 29, Feb 5,12,19, Mar 5(+),12,19(+),26, Apr 9,16,23,30
- Location: 28SP, H103.
- Office hours: by apointment (my email: ag133@nyu.edu).
- Course webpage: http://alfredgalichon.com/microeconometrics2018s/
- Material available from: https://github.com/alfredgalichon/microeconometrics2018s
- Text (optional): Galichon (2016). *Optimal Transport Methods in Economics*, Princeton.

- A. Galichon: professor of economics and of mathematics at NYU, visiting Sciences Po (spring 2017-2018).

- This course is taught under the 'math&econ+code' format: 12 "blocks"; each block = 50 minutes of theory (math & econ) + 1 hour of code
  - coding most often based on an empirical application related to the theory just seen
  - students are expected to write their own code; we'll ensure that it is operational at the end of each block
- Programming: our demos will be done in R and the support will be in R only, but you are welcome to use the language of your choice e.g. Matlab, C++, Python, Julia... Solvers used will be Gurobi (for LP) and NLOPT (for nonlinear optimization), so make sure your language of choice has a convenient interface to these.
- Questions?

- This course is focused on models of demand, matching models, and optimal transport methods, with various applications pertaining to labor markets, economics of marriage, industrial organization, matching platforms, networks, and international trade, from the crossed perspectives of theory, empirics and computation.
- It will introduce tools from economic theory, mathematics, econometrics and computing, on a needs basis, without any particular prerequisite other than the equivalent of a first year graduate sequence in econ or in applied math.

- Part I: Tools
  - Jan: linear programming
  - Feb: optimal transport toolbox
- Part II: Models
  - March: multinomial choice and demand models
  - April: estimation of models of matching

- (https://www.r-project.org/), Rstudio (https://www.rstudio.com/), and, for Windows users only, Rtools (https://cran.r-project.org/bin/windows/Rtools/).
- gurobi (www.gurobi.com; you'll need to obtain an academic license, which is free if you're located on an educational domain).

On a sheet of paper, please indicate:

1. Your name and email
2. Your academic background (which program were you enrolled in the previous years)
3. Whether you are taking this course for credit
4. What you are looking for in this course (2 sentences max).
5. From a scale of 1 to 5, whether you are less familiar (1) or more familiar (5) with the concepts below:

1. Linear programming
2. Legendre-Fenchel transforms
3. Min-cost flow problem
4. Becker's theory of marriage
5. Backward induction
6. Compensating wage differentials
7. Envelope theorem
8. Complementary slackness
9. Walrasian equilibrium
10. Hotelling's spatial model
11. Logit choice model
12. Gradient descent
13. Newton descent
14. Coordinate descent

Note: even if you anwer mostly "1", don't worry! These concepts will be explained in due course.

Block 1: Linear programming

- Linear programming duality
- Economic interpretation of the dual
- Numerical computation

- [OTME], App. B
- Stigler (1945), The cost of subsistence. *Journal of Farm Economics*.
- Dantzig (1990), The diet problem. *Interface*.
- Complements:
  - Gale (1960), *The theory of linear economic models*.
  - Vohra (2011), *Mechanism Design: A Linear Programming Approach*.
- www.gurobi.com
- www.gnu.org/software/glpk/

Section 1

Theory

- During World War II, engineers in US Army were wondering how to feed their personnel at minimal cost, leading to what is now called the "optimal diet problem".
  - Nutritionists have identified a number of vital nutrients (calories, protein, calcium, iron, etc.) that matter for a person's health, and have determined the minimum daily intake of each nutrient
  - For each basic food (pasta, butter, bread, etc), nutritionists have characterized the intake in each of the various nutrients
  - Each food has a unit cost, and the problem is to find the optimal diet = combination of foods that meet the minimal intake in each of the nutrients and achieves minimal cost
- The problem was taken on by G. Stigler, who published a paper about it in 1945, giving a first heuristic solution, exhibiting a diet that costs \$39.93 per year in 1939 dollars. Later (in 1947) it was one of the first application of G.B. Dantzig's method (the simplex algorithm), which provided the exact solution (\$39.67). It then took 120 man-day to perform this operation. At the end of this block, the computer will perform it for us in a fraction of second.
- However, don't try this diet at home! Dantzig did so and almost died from it...

- Our dataset (directory '01-appli-diet') was directly taken from Stigler's article. It is a csv file called 'StiglerData1939.txt':

half are necessary to human beings.[6] The precise determination of our needs for these—and no doubt other yet undiscovered—nutrients lies far in the future.

Nevertheless standards of dietary adequacy have been established, perhaps prematurely and certainly very tentatively. The "allowances" (a term used to indicate their preliminary nature) of the National Research Council embody what is presumably the 1943 consensus of the experts; they are given in Table 1. Other minerals and vitamins are believed to be supplied in adequate quantities if these nutrients are secured from natural foods. The requirements are net of losses in the preparation of food. These standards are met by the minimum cost diets derived subsequently.

TABLE 1. DAILY ALLOWANCES OF NUTRIENTS FOR A MODERATELY ACTIVE MAN (weighing 154 pounds)*

Nutrient	Allowance
Calories	3,000 calories
Protein	70 grams
Calcium	.8 grams
Iron	12 milligrams
Vitamin A	5,000 International Units
Thiamine (B_1)	1.8 milligrams
Riboflavin (B_2 or G)	2.7 milligrams
Niacin (Nicotinic Acid)	18 milligrams
Ascorbic Acid (C)	75 milligrams

* National Research Council, *Recommended Dietary Allowances*, Reprint and Circular Series No. 115, January, 1943.

Nutritive Values of Foods

The minimum cost of an adequate diet is obviously governed by the nutritive values and costs of the foods eligible for inclusion. The very restricted list of foods considered in this study is discussed in Section 3 and the foods are listed in Tables A and B. It may be mentioned here that only natural foods are included; vitamin pills are excluded because they do not contain all of the nutrients (known and unknown) which are necessary to good health.[7]

The nutritive values of common foods are known only roughly,

[6] For summaries of the present knowledge, see *Yearbook of Agriculture*, 1939. Food and Life; H. C. Sherman, *Chemistry of Food and Nutrition* (6th ed., 1941).

[7] Puppies put on a diet in which the vitamins were provided from synthetic sources lived only 100 to 150 days; see J. P. Lambooy and E. S. Nasset, "The Inadequacy of Eight Synthetic B Vitamins for the Nutrition of Puppies," *Journal of Nutrition*, XXVI (1943), 293–302.

TABLE A. NUTRITIVE VALUES OF COMMON FOODS PER DOLLAR OF EXPENDITURE, AUGUST 15, 1939

Commodity	Unit	Price Aug. 15, 1939 (cents)	Edible Weight per $1.00 (grams)	Calories (1,000)	Protein (grams)	Calcium (grams)	Iron (mg.)	Vitamin A (1,000 I.U.)	Thiamine (mg.)	Riboflavin (mg.)	Niacin (mg.)	Ascorbic Acid (mg.)
**1. Wheat Flour (Enriched)	10 lb.	36.0	12,600	44.7	1,411	2.0	365		55.4	33.3	441	
2. Macaroni	1 lb.	14.1	3,217	11.6	418	.7	54		3.2	1.9	68	
3. Wheat Cereal (Enriched)	28 oz.	24.2	3,280	11.8	377	14.4	175		14.4	8.8	114	
4. Corn Flakes	8 oz.	7.1	3,194	11.4	252	.1	56		13.5	2.3	68	
5. Corn Meal	1 lb.	4.6	9,861	36.0	897	1.7	99	30.9	17.4	7.9	106	
6. Hominy Grits	24 oz.	8.5	8,005	28.6	680	.8	80		10.6	1.6	110	
7. Rice	1 lb.	7.5	6,048	21.2	460	.6	41		2.0	4.8	60	
8. Rolled Oats	1 lb.	7.1	6,389	25.3	907	5.1	341		37.1	8.9	64	
9. White Bread (Enriched)	1 lb.	7.9	5,742	15.0	488	2.5	115		13.8	8.5	126	
10. Whole Wheat Bread	1 lb.	9.1	4,985	12.2	484	2.7	125		13.9	6.4	160	
11. Rye Bread	1 lb.	9.2	4,930	12.4	439	1.1	82		9.9	3.0	66	
12. Pound Cake	1 lb.	24.8	1,829	8.0	130	.4	31	18.9	2.8	3.0	17	
13. Soda Crackers	1 lb.	15.1	3,004	12.5	288	.5	50					
14. Milk	1 qt.	11.0	8,867	6.1	310	10.5	18	16.8	4.0	16.0	7	177
**15. Evaporated Milk (can)	14½ oz.	6.7	6,035	8.4	422	15.1	9	26.0	3.0	23.5	11	60
16. Butter	1 lb.	30.8	1,473	10.8	9	.2	3	44.2		.2	2	
*17. Oleomargarine	1 lb.	16.1	2,817	20.6	17	.6	6	55.8	.2			
18. Eggs	1 doz.	32.6	1,857	2.9	238	1.0	52	18.6	2.8	6.5	1	
**19. Cheese (Cheddar)	1 lb.	24.2	1,874	7.4	448	16.4	19	28.1	.8	10.3	4	
20. Cream	½ pt.	14.1	1,689	3.5	49	1.7	3	16.9	.6	2.5		17
21. Peanut Butter	1 lb.	17.9	2,534	15.7	661	1.0	48		9.6	8.1	471	
22. Mayonnaise	½ pt.	16.7	1,198	8.6	18	.2	8	2.7	.4	.5		
23. Crisco	1 lb.	20.3	2,234	20.1								
24. Lard	1 lb.	9.8	4,628	41.7				.2		.5	5	
25. Sirloin Steak	1 lb.	39.6	1,145*	2.9	166	.1	34	.2	2.1	2.9	69	
26. Round Steak	1 lb.	36.4	1,246*	2.2	214	.1	32	.4	2.5	2.4	87	
27. Rib Roast	1 lb.	29.2	1,553*	3.4	213	.1	33			2.0		
28. Chuck Roast	1 lb.	22.6	2,007*	3.6	309	.2	46	.4	1.0	4.0	120	
29. Plate	1 lb.	14.6	3,107*	8.5	404	.2	62		.9			
**30. Liver (Beef)	1 lb.	26.8	1,692*	2.2	333	.2	139	169.2	6.4	50.8	316	525
31. Leg of Lamb	1 lb.	27.6	1,643*	3.1	245	.1	20		2.8	3.9	86	
32. Lamb Chops (Rib)	1 lb.	36.6	1,239*	3.3	140	.1	15		1.7	2.7	54	
33. Pork Chops	1 lb.	30.7	1,477*	3.5	196	.2	30		17.4	2.7	60	
34. Pork Loin Roast	1 lb.	24.2	1,874*	4.4	249	.3	37		18.2	3.6	79	
35. Bacon	1 lb.	25.6	1,772*	10.4	152	.2	23		1.8	1.8	71	
36. Ham—smoked	1 lb.	27.4	1,655*	6.7	212	.2	31		9.9	3.3	50	
37. Salt Pork	1 lb.	16.0	2,835*	18.8	164	.1	26		1.4	1.8		
38. Roasting Chicken	1 lb.	30.3	1,497*	1.8	184	.1	30	.1	.9	1.8	68	46
39. Veal Cutlets	1 lb.	42.3	1,072*	1.7	156	.1	24		1.4	2.4	57	
40. Salmon, Pink (can)	16 oz.	13.0	3,489	5.8	705	6.8	45	3.5	1.0	4.9	209	
41. Apples	1 lb.	4.4	9,072	5.8	27	.5	36	7.3	3.6	2.7	5	544
42. Bananas	1 lb.	6.1	4,982	4.9	60	.4	30	17.4	2.5	3.5	28	498
43. Lemons	1 doz.	26.0	2,380	1.0	21	.5	14		.5		4	952
44. Oranges	1 doz.	30.9	4,439	2.2	40	1.1	18	11.1	3.6	1.3	10	1,998
*45. Green Beans	1 lb.	7.1	5,750	2.4	138	3.7	80	69.0	4.3	5.8	37	862
**46. Cabbage	1 lb.	3.7	8,949	2.6	125	4.0	36	7.2	9.0	4.5	26	5,369
47. Carrots	1 bunch	4.7	6,080	2.7	73	2.8	43	188.5	6.1	4.3	89	608
48. Celery	1 stalk	7.3	3,915	.9	51	3.0	23	.9	1.4	1.4	9	313
49. Lettuce	1 head	8.2	2,247	.4	27	1.1	22	112.4	1.8	3.4	11	449
*50. Onions	1 lb.	3.6	11,844	5.8	166	3.8	59	16.6	4.7	5.9	21	1,184

- Problem setup:
  - Assume there are nutrients $i \in \{1, ..., m\}$ (calories, protein, calcium, iron, etc.) that matter for a person's health, in such way that the minimum daily intake of nutrient i should be d_i.
  - Nutrients do not come as standalone elements, but are combined into various foods. Each unit of food $j \in \{1, ..., n\}$ yields a quantity N_{ij} of nutrient $i \in \{1, ..., m\}$. The dollar cost of food j is c_j.
- The problem is to find the diet that achieves the minimal intake of each nutrient at a cheapest price. If $q \in \mathbb{R}^n$ is a vector such that $q_j \geq 0$ is the quantity of food i purchased, the quantity of nutrient i ingested is $\sum_{j=1}^n N_{ij} q_j$, and the cost of the diet is $\sum_{j=1}^n q_j c_j$. The optimal diet is therefore given by

$$\begin{aligned} &\min_{q \geq 0} c^\top q \\ &s.t.\ Nq \geq d. \end{aligned} \tag{1}$$

- Let $c \in \mathbb{R}^n$, $d \in \mathbb{R}^m$, A be a $m \times n$ matrix, and consider the following problem

$$V_P = \max_{x \in \mathbb{R}^n_+} c^\top x \qquad (2)$$
$$s.t. \ Ax = d$$

  This problem is a *linear programming problem*, as the objective function, namely $x \to c^\top x$ is linear, and as the constraint, namely $x \in \mathbb{R}^n_+$ and $Ax = d$ are also linear (or more accurately, affine). Problem (2) is called *primal program*, for reasons to be explained soon. The set of x's that satisfy the constraint are called *feasible solutions*; the set of solutions of problem (2) are called *optimal solutions*.

- Remarks:
  - The previous diet problem can be reformulate into this problem – why?
  - A problem does not necessarily have a feasible solution (e.g. if $A = 0$ and $d \neq 0$), in which case (by convention) $V_P = -\infty$.
  - The whole space may be solution (e.g. if $A = 0$ and $d = 0$), in which case $V_P = +\infty$.

There is a powerful tool called duality which provides much insight into the analysis of problem (2). The idea is to rewrite the problem as

$$V_P = \max_{x \in \mathbb{R}_+^n} \left\{ c^\top x + L_P \left(d - Ax \right) \right\}$$

where $L_P(z)$ is a penalty function whose value is zero if the constraint is met, that is if $z = 0$, and $-\infty$ if it is not, namely if $z \neq 0$. The simplest choice of such penalty function is given by $L_P(z) = \min_{y \in \mathbb{R}^m} \left\{ z^\top y \right\}$. One has

$$V_P = \max_{x \in \mathbb{R}_+^n} \min_{y \in \mathbb{R}^m} \left\{ c^\top x + \left(d - Ax \right)^\top y \right\}.$$

However, the minimax inequality $\max_x \min_y \leq \min_y \max_x$ always holds, thus

$$V_P \leq \min_{y \in \mathbb{R}^m} \max_{x \in \mathbb{R}^n_+} \left\{ c^\top x + (d - Ax)^\top y \right\} = \min_{y \in \mathbb{R}^m} \max_{x \in \mathbb{R}^n_+} \left\{ x^\top \left(c - A^\top y \right) + d^\top y \right\}$$

$$\leq \min_{y \in \mathbb{R}^m} \left\{ d^\top y + L_D \left(c - A^\top y \right) \right\} =: V_D$$

where $L_D(z) = \max_{x \in \mathbb{R}^n_+} \left\{ x^\top z \right\}$ is equal to 0 if $z \in \mathbb{R}^n_-$, and to $+\infty$ if not. Therefore, the value V_D is expressed by the *dual program*

$$V_D = \min_{y \in \mathbb{R}^m} d^\top y, \qquad (3)$$

$$s.t.\ A^\top y \geq c$$

and the weak duality inequality $V_P \leq V_D$ holds. It turns out that as soon as either the primal or dual program has an optimal solution, then both programs have an optimal solution and the values of the two programs coincide, so the weak duality becomes an equality $V_P = V_D$ called strong duality. Further, if $x^* \in \mathbb{R}^n_+$ is an optimal primal solution, and $y^* \in \mathbb{R}^m$ is an optimal dual solution, then complementary slackness holds, that is $x_i^* > 0$ implies $\left(A^\top y^* \right)_i = c_i$.

We summarize these results into the following statement.
Theorem. In the setting described above:
(i) The weak duality inequality holds:

$$V_P \leq V_D.$$

(ii) As soon as the primal or the dual program have an optimal solution, then both programs have an optimal solution, and strong duality holds:

$$V_P = V_D.$$

(iii) If $x^* \in \mathbb{R}^n_+$ is an optimal primal solution, and $y^* \in \mathbb{R}^m$ is an optimal dual solution, then complementary slackness holds:

$$x_i^* > 0 \text{ implies } \left(A^\top y^*\right)_i = c_i.$$

- Recall the optimal diet problem

$$\min_{q \geq 0} c^\top q$$
$$s.t.\ Nq \geq d.$$

which has minimax formulation $\min_{q \geq 0} \max_{\pi \geq 0} c^\top q + d^\top \pi - q^\top N^\top \pi$, so the dual is

$$\max_{\pi \geq 0} d^\top \pi$$
$$s.t.\ N^\top \pi \leq c$$

- Interpretation: imagine that there is a new firm called Nutrient Shoppe, who sells raw nutrients. Let π_i be the price of nutrient i. The cost of the diet is $d^\top \pi$. Consumer purchase raw nutrients and can generate "synthetic" foods. The cost of the synthetic version of food j is $\sum_{i=1}^{m} N_{ij} \pi_i = \left(N^\intercal \pi \right)_j$. The constraint thus means that each "synthetic" food is more affordable than its natural counterpart.

- The duality means that it is possible to price the nutrients so that the synthetic foods are cheaper than the natural ones, in such a way that the price of the synthetic diet equals the price of the natural diet.
- Complementary slackness yields:
  - $q_j > 0$ implies $\left(N^{\intercal}\pi\right)_j = c_j$; that is, if natural food j is actually purchased, then the prices of its synthetic and natural versions coincide
  - $\pi_i > 0$ implies $\left(Nq\right)_i = d_i$; that is, if nutrient i has a positive price, then the natural diet has the "just right" amount.

Section 2

Coding

- Install R (https://www.r-project.org/), a free statistical environment
- Install Rstudio (https://www.rstudio.com/), a free IDE for R
- [Windows users only] Install Rtools (https://cran.r-project.org/bin/windows/Rtools/), a toolchain for Windows needed to install certain R packages
- Install Gurobi (www.gurobi.com) – a state-of-the-art commercial lp solver
  - Install the program
  - Obtain a license for Gurobi
  - Install R package for Gurobi
- Install Rglpk (from Rstudio) – an R interface to the GLPK open-source lp solver

- Today we shall retrieve Stigler's diet data and compute the optimal diet in order to compare with Stigler's computations
- We shall do so from R, using in turn Gurobi and GLPK.

```
library(gurobi)
thepath = getwd()
filename="/StiglerData1939.txt"
thedata =
as.matrix(read.csv(paste0(thepath,filename),sep="\t",
header=T))
nbCommodities=length(which(thedata[,1] != "" ))-1
names = thedata[1:nbCommodities,1]
themat = matrix( as.numeric(thedata[,3:13]), ncol = 11)
themat[is.na(themat)] = 0
```

```
N = t(themat[1:nbCommodities,3:11])
d = themat[(nbCommodities+1),3:11]
c = rep(1,nbCommodities)
result = gurobi (
list(A=N,obj=c,modelsense="min",rhs=d,sense=">"),
params=list(OutputFlag=0) )
q_yearly = result$x * 365 # convert into yearly cost
pi = result$pi
cost_daily = result$objval
```

Remark: by default, Gurobi assumes the constraint $x \geq 0$. To remove this constraint, include `ub = -Inf` in the list passed to Gurobi.

```
toKeep = which(q_yearly !=0 )
foods = q_yearly[toKeep]
names(foods) = names[toKeep]
print(foods)
print(paste0("Total cost (optimal)= ", sum(q_yearly*c) ))
```

```
toKeepStigler = c(1,15,46,52,69)
foods_stigler = c(13.33, 3.84,4.11,1.85,16.80)
names(foods_stigler) = names[toKeepStigler]
print(foods_stigler)
print(paste0("Total cost (Stigler)= ",
sum(foods_stigler*c[toKeepStigler]) ))
```

```
library(Rglpk)
resGlpk = Rglpk_solve_LP(obj=c, mat=N,
dir=rep(">",length(d)), rhs=d, bounds = NULL, max = FALSE,
control = list())
print(resGlpk$optimum*365)
```