X2. Linear Programming

We freely use the textbook [Vas00]. Additional references including references to websites with software, can be found in [Vas00] and [Ros00, Section 15.1].

1. What is linear programming?

Definitions:

Optimization is maximization or minimization of a real-valued function, called the *objective function,* on a set, called the *feasible region* or the set of *feasible solutions.*

The values of function on the set are called *feasible values.*

An *optimal solution* (optimizer) is a feasible solution where the objective function reaches an *optimal value* (optimum), i.e., maximal or minimal value respectively.

An optimization problem is called *infeasible*, if there are no feasible solutions, i.e., the feasible region is empty.

It is called *unbounded*, if the feasible values are arbitrary large in the case of maximization problem or arbitrary small in the case of minimization problem.

A *mathematical program* is an optimization problem where the feasible region is a subset of $\mathbf{R}^n$, a finite dimensional real space, i.e., the objective function is a function of one or several real variables.

A *lineal form* in variables $x_1, \ldots, x_n$ is $c_1 x_1 + \cdots + c_n x_n$ where c_i are given numbers.

An *affine function* is a linear form plus a given number.

A *linear constraint* is one of the following three constraints: $f \leq g, f = g, f \geq g$, where f and g are affine functions. In standard form, f is a linear form and g is a given number.

A *linear program* is a mathematical program where the objective function is an affine function and the feasible region is given by a finite system of linear constraints, all of them to be satisfied.

Facts::

1. A linear functions means a linear form in some textbooks and an affine function in others.

2. Linear constraints with equality signs are known as linear equations. The main tool of simplex method, pivot steps, allows us to solve any system of linear equations.

3. In some textbooks, the objective function in a linear program is required to be a linear form. Dropping a constant in the objective function does not change optimal solutions, but the optimal value changes in the obvious way.

4. Solving an optimization problem usually means finding the optimal value and an optimal solution or showing that they do not exist. By comparison, solving a system of linear equations usually means finding all solutions. In both cases, in real life we only find approximate solutions.

5. Linear programs with one and two variables can be solved graphically. In the case of one variable x, the feasible region has one of the following forms: empty; a point $x = a$; a finite interval $a \leq x \leq b$ with $a < b$; a ray $x \geq a$ or $x \leq a$; the whole line. The objective function $f = cx + d$ is represented by a straight line. Depending on the sign of c and

whether we want maximize or minimize f, we move in the feasible region to the right or to the left as far as we can in search for an optimal solution.

In the case of two variables, the feasible region is a closed convex set with finitely many vertices (corners). Together with the feasible region, we can draw in plane levels of the objective function. Unless the objective function is constant, every level (where the function takes a certain value) is a straight line. This picture allows us to see whether the program is feasible and bounded. If it is, the picture allows us to find a vertex which is optimal.

6. Linear Programming is about formulating, collecting data, and solving linear programs, and also about analyzing and implementing solutions in real life.

7. Linear Programming is an important part of Mathematical Programming. In its turn, Mathematical Programming is a part of *Operations Research. Systems Engineering* and *Management Science* are engineering and business versions of Operations Research.

Examples:

1. Here are 3 linear forms in $x, y, z : 2x - 3y + 5z, x + z, y$.

2. Here are 3 affine functions of $x, y, z : 2x - 3y + 5z - 1, x + z + 3, y$.

3. Here are 3 functions of $x, y, z :$ which are not affine: $xy, x^2 + z^3, \sin z$.

4. The constraint $|x| \leq 1$ is not linear, but it is equivalent to a system of two linear constraints, $x \leq 1, x \geq -1$.

5. (An infeasible linear program) Here is a linear program:

Maximize $x + y$ subject to $x \leq -1, x \geq 0$.

It has two variables and two linear constraints. The objective function is a linear form. The program is infeasible.

6. (An unbounded linear program) Here is a linear program:

Maximize $x + y$.

This program has two variables and no constraints. The objective function is a linear form. The program is unbounded.

6. Here is a linear program:

Minimize $x + y$ subject to $x \geq 1, y \geq 2$.

This program has two variables and two linear constraints. The objective function is a linear form. The optimal value (the maximum) is 3. An optimal solution is $x = 1, y = 2$. It is unique.

2. Setting up (formulating) linear programs.

Examples:

1. Finding the maximum of n given numbers $a_1, \ldots, c_n$ does not look like a linear program. However it is equivalent to the following linear program with n variables kx_i and $n + 1$ linear constraints:

$c_1 x_1 + \cdots + c_n x_n \to$ max, all $x_i \geq 0, x_1 + \cdots + x_n = 1$.

An equivalent linear program with one variable y and n linear constraints is

$y \to$ min, $y \geq c_i$ for all i.

2. (Diet problem [V, Example 2.1). The general idea is to select a mix of different foods in such a way that basic nutritional requirements are satisfied at minimum cost. Our example is drastically simplified.

According to the recommendations of a nutritionist, a person's daily requirements for protein, vitamin A, and calcium are as follows: 50 grams of protein, 4000 IUs (international units) of vitamin A, 1000 milligrams of calcium. For illustrative purposes, let us consider a diet consisting only of apples (raw, with skin), bananas (raw), carrots (raw), dates (domestic, natural, pitted, chopped), and eggs (whole, raw, fresh) and let us, if we can, determine the amount of each food to be consumed in order to meet the Recommended Dietary Allowances (RDA) at minimal cost.

Food	Unit	Protein (g)	Vit. A (IU)	Calcium (mg)
apple	1 medium (138 g)	0.3	73	9.6
banana	1 medium (118 g)	1.2	96	7
carrot	1 medium (72 g)	0.7	20253	19
dates	1 cup (178 g)	3.5	890	57
egg	1 medium (44 g)	5.5	279	22

Since our goal is to meet the RDA with minimal cost, we also need to compile the costs of these foods:

Food	Cost (in cents)
1 apple	10
1 banana	15
1 carrot	5
1 cup of dates	60
1 egg	8

Using these data, we can now set up a linear program. Let a, b, c, d, e be variables representing the quantities of the five foods we are going to use in the diet. The objective function to be minimized is the total cost function (in cents),

$$C = 10a + 15b + 5c + 60d + 8e,$$

where the coefficients represent cost per unit of the five items under consideration.

What are the constraints ? Obviously,

$$a, b, c, d, e \geq 0. \qquad (i)$$

These constraints are called *nonnegativity constraints.*

Then, to ensure that the minimum daily requirements of protein, vitamin A, and calcium are satisfied, it is necessary that

$$\begin{cases} 0.3a & + & 1.2b & + & 0.7c & + & 3.5d & + & 5.5e \geq 50 \\ 73a & + & 96b & + & 20253c & + & 890d & + & 279e \geq 4000 \\ 9.6a & + & 7b & + & 19c & + & 57d & + & 22e \geq 1000, \end{cases} \qquad (ii)$$

where, for example, in the first constraint, the term $0.3\,a$ expresses the number of grams of protein in each apple multiplied by the quantity of apples needed in the diet, the second term $1.2\,b$ expresses the number of grams of protein in each banana multiplied by the quantity of bananas needed in the diet, and so forth.

Thus, we have a linear program with 5 variables and 8 linear constraints.

3. (Blending problem [V, Example 2.2]). Many coins in different countries are made from cupronickel (75% copper, 25% nickel). Suppose that the four available alloys (scrap metals), A, B, C, D, to be utilized to produce the coin contain the percentages of copper and nickel shown in the following table:

Alloy	A	B	C	D
% copper	90	80	70	60
% nickel	10	20	30	40
\$/lb	1.2	1.4	1.7	1.9

The cost in dollars per pound of each alloy is given as the last row in the same table.

Notice that none of the four alloys contains the desired percentages of copper and nickel. Our goal is to combine these alloys into a new blend containing the desired percentages of copper and nickel for cupronickel while minimizing the cost. This lends itself to a linear program.

Let a, b, c, d be the amounts of alloys A, B, C, D in pounds to make a pound of the new blend. Thus,

$$a,\ b,\ c,\ d \geq 0. \tag{i}$$

Since the new blend will be composed exclusively from the four alloys, we have

$$a + b + c + d = 1. \tag{ii}$$

The conditions on the composition of the new blend give

$$\begin{cases} .9a &+& .8b &+& .7c &+& .6d &=& .75 \\ .1a &+& .2b &+& .3c &+& .4d &=& .25. \end{cases} \tag{iii}$$

For example, the first equality states that 90% of the amount of alloy A, plus 80% of the amount of alloy B, plus 70% of the amount of alloy C, plus 60% of the amount of alloy D will give the desired 75% of copper in a pound of the new blend. Likewise, the second equality gives the desired amount of nickel in the new blend.

Taking the preceding constraints into account, we minimize the cost function

$$C = 1.2a + 1.4b + 1.7c + 1.9d.$$

In this problem all the constraints, except (i), are *equalities.* In fact, there are three linear equations and four unknowns. However, the three equations are not independent. For example, the sum of the equations in (iii) gives (ii). Thus, (ii) is redundant.

In general, a constraint is said to be *redundant* if it follows from the other constraints of our system. Since it contributes no new information regarding the solutions of the linear program, it can be dropped from consideration without changing the feasible set.

4. (Manufacturing problem [V, Example 3.3]). We are now going to state a program in which the objective function, a profit function, is to be maximized. A factory produces three products: P1, P2, and P3. The unit of measure for each product is the standard-sized boxes into which the product is placed. The profit per box of P1, P2, and P3 is \$2, \$3 and \$7, respectively. Denote by x_1, x_2, x_3 the number of boxes of P1, P2, and P3, respectively. So the profit function we want to maximize is

$$P = 2x_1 + 3x_2 + 7x_3.$$

The five resources used are raw materials R1 and R2, labor, working area, and time on a machine. There are 1200 lbs of R1 available, 300 lbs of R2, 40 employee-hours of labor, 8000 m^2 of working area, and 8 machine-hours on the machine.

The amount of each resource needed for a box of each of the products is given in the following table (which also includes the aforementioned data):

Resource	Unit	P1	P2	P3	‖	Available
R1	lb	40	20	60	‖	1200
R2	lb	4	1	6	‖	300
Labor	hour	.2	.7	2	‖	40
Area	m^2	100	100	800	‖	8000
Machine	hour	.1	.3	.6	‖	8
Profit	\$	2	3	7	‖	$\rightarrow$ max

As we see from this table, to produce a box of P1 we need 40 pounds of R1, 4 pounds of R2, 0.2 hours of labor, 100 m^2 of working area, and 0.1 hours on the machine. Also, the amount of resources needed to produce a box of P2 and P3 can be deduced from the table.

The constraints are

$$x_1, x_2, x_3 \geq 0, \qquad (i)$$

and

$$
\begin{cases}
40x_1 & + & 20x_2 & + & 60x_3 & \leq & 1200 & \text{(pounds of R1)} \\
4x_1 & + & x_2 & + & 6x_3 & \leq & 300 & \text{(pounds of R2)} \\
.2x_1 & + & .7x_2 & + & 2x_3 & \leq & 40 & \text{(hours of labor)} \\
100x_1 & + & 100x_2 & + & 800x_3 & \leq & 8000 & \text{(area in m}^2) \\
.1x_1 & + & .3x_2 & + & .8x_3 & \leq & 8 & \text{(machine)}.
\end{cases}
\qquad (ii)
$$

5. (Transportation problem [V, Example 2.4]).

Another concern that manufacturers face daily is *transportation costs* for their products. Let us look at the following hypothetical situation and try to set it up as a linear program.

A manufacturer of widgets has warehouses in Atlanta, Baltimore, and Chicago. The warehouse in Atlanta has 50 widgets in stock, the warehouse in Baltimore has 30 widgets

in stock, and the warehouse in Chicago has 50 widgets in stock. There are retail stores in Detroit, Eugene, Fairview, Grove City, and Houston. The retail stores in Detroit, Eugene, Fairview, Grove City, and Houston need at least 25, 10, 20, 30, 15 widgets, respectively. Obviously, the manufacturer needs to ship widgets to all five stores from the three warehouses and he wants to do this in the cheapest possible way. This presents a perfect backdrop for a linear program, to minimize shipping cost. To start, we need to know the cost of shipping one widget from each warehouse to each retail store. This is given by a shipping cost table

	1.D	2.E	3.F	4.G	5.H
1.Atlanta	55	30	40	50	40
2.Baltimore	35	30	100	45	60
3.Chicago	40	60	95	35	30

Thus, it costs \$30 to ship one unit of the product from Baltimore to Eugene (E), \$95 from Chicago to Fairview (F), and so on.

In order to set this up as a linear program, we introduce variables that represent the number of units of product shipped from each warehouse to each store. We have numbered the warehouses according to their alphabetical order and we have enumerated the stores similarly. Let x_{ij}, for all $1 \leq i \leq 3$, $1 \leq j \leq 5$, represent the number of widgets shipped from warehouse $\#i$ to store $\#j$. This gives us 15 unknowns. The objective function (the quantity to be minimized) is the shipping cost given by

$$\begin{aligned} C = {} & 55x_{11} + 30x_{12} + 40x_{13} + 50x_{14} + 40x_{15} \\ & + 35x_{21} + 30x_{22} + 100x_{23} + 45x_{24} + 60x_{25} \\ & + 40x_{31} + 60x_{32} + 95x_{33} + 35x_{34} + 30x_{35} \end{aligned}$$

where $55x_{11}$ represents the cost of shipping one widget from the warehouse in Atlanta to the retail store in Detroit (D) multiplied by the number of widgets that will be shipped, and so forth.

What are the constraints? First, our 15 variables satisfy the condition that

$$x_{ij} \geq 0, \text{ for all } 1 \leq i \leq 3, \ 1 \leq j \leq 5 \tag{i}$$

since shipping a negative amount of widgets makes no sense. Second, since the warehouse $\#i$ cannot ship more widgets than it has in stock, we get

$$\begin{cases} x_{11} + x_{12} + x_{13} + x_{14} + x_{15} \leq 50 \\ x_{21} + x_{22} + x_{23} + x_{24} + x_{25} \leq 30 \\ x_{31} + x_{32} + x_{33} + x_{34} + x_{35} \leq 50 \end{cases} \tag{ii}$$

Next, working with the amount of widgets that each retail store needs, we obtain the following five constraints:

$$\begin{cases} x_{11} & + & x_{21} & + & x_{31} & \geq & 25 \\ x_{12} & + & x_{22} & + & x_{32} & \geq & 10 \\ x_{13} & + & x_{23} & + & x_{33} & \geq & 20 \\ x_{14} & + & x_{24} & + & x_{34} & \geq & 30 \\ x_{15} & + & x_{25} & + & x_{35} & \geq & 15 \end{cases} \qquad (iii)$$

The problem is now set up. It is a linear program with 15 variables and 23 linear constraints.

6. (Job assignment problem [V, Example 2.5]).

Suppose that a production manager must assign n workers to do n jobs. If every worker could perform each job at the same level of skill and efficiency, the job assignments could be issued arbitrarily. However, as we know, this is seldom the case. Thus, each of the n workers is evaluated according to the time he or she takes to perform each job. The time, given in hours, is expressed as a number greater than or equal to zero. Obviously, the goal is to assign workers to jobs in such a way that the total time is as small as possible. In order to set up the notation, we let c_{ij} be the time it takes for worker $\# i$ to perform job $\# j$. Then the times could naturally be written in a table. For example, take $n = 3$ and let the times be given as in the following table:

	a	b	c
A	10	70	40
B	20	60	10
C	10	20	90

We can examine all 6 assignments and find that the minimum value of the total time is 40. So we conclude that the production manager would be wise to assign worker **A** to job **a**, worker **B** to job **c** and worker **C** to job **b**.

In general, this method of selection is not good. The total number of possible ways of assigning jobs is $n! = n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1$. This is an enormous number even for moderate n. For $n = 70$,

$$n! = 11978571669969891796072783721689098736458938142564258575553628646280095827898453196800000000000000000.$$

It has been estimated that if a Sun Workstation computer had started solving this problem at the time of the Big Bang, by looking at all possible job assignments, then by now it would not have finished yet its task.

Although is not obvious, the job assignment problem (with any number n of workers and jobs) can be expressed as a linear program. Namely, we set

$$x_{ij} = 0 \text{ or } 1 \qquad (i)$$

depending on whether the worker i is assigned to do the job j. The total time then is then

$$\sum_i \sum_j c_{ij} x_{ij} (\text{to be minimized}). \qquad (ii)$$

7

The condition that every worker i is assigned to exactly one job is

$$\sum_j x_{ij} = 1 \quad \text{for all } i. \tag{iii}$$

The condition that exactly one worker is assigned to every job j is

$$\sum_i x_{ij} = 1 \quad \text{for all } j. \tag{iv}$$

The constraints (i) are not linear. If we replace them by linear constraints $x_{i,j} \geq 0$, then we obtain a linear program with n^2 variables and $n^2 + 2n$ linear constraints. Mathematically, it is a transportation problem with every demand and supply equal 1. As such, it can be solved by the simplex method (see below). (When $n = 70$ it takes seconds.)

The simplex method for transportation problem does not involve any divisions. Therefore an optimal solution it gives integral values for all components $x_{i,j}$. Thus, the conditions (i) hold, and the simplex method solves the job assignment problem. (Another way to express this is that all vertices of the feasible region (iii)–(iv) satisfy (i).)

3. Standard and canonical forms for linear programs.

Definitions:

LP in *canonical form* [Vas00] is $cx + d \to \min, x \geq 0, Ax \leq b$; where x is a column of distinct (decision) variables, c is a given row, d is a given number, A is a given matrix, b is a given column.

LP in *standard form* [Vas00]) is $cx + d \to \min, x \geq 0, Ax = b$, where x, c, d, A, b are as in the previous paragraph.

Facts:

1. Every LP can be written in normal as well as standard form using the following five little tricks:

1. Maximization and minimization problems can be converted to each other. Namely, the problems
   $f(x) \to \min, x \in S$
and
   $-f(x) \to \max, x \in S$
are equivalent in the sense that they have the same optimal solutions and max = -min.

i. The equation $f = g$ is equivalent to the system of two inequalities, $f \geq g, g \leq f$.

ii. The inequality $f \leq g$ is equivalent to the inequality $-f \geq -g$.

iii. The inequality $f \leq g$ is equivalent to $f + x = g, x \geq 0$, where x is a new variable (slack variable).

iv. A variable x unconstrained in sign can be replaced in our program by two new nonnegative variables, $x = x' - x''$, where $x', x'' \geq 0$.

2. The same tricks are sufficient for rewriting any linear program in the standard, canonical, and normal forms, which are different in different textbooks and different software packages. In most cases, all decision variables in these forms are assume to be nonnegative.

8

Examples:

1. The canonical form $cx + d \to \min, x \geq 0, Ax \leq b$
can be rewritten in the following standard form:
$cx + d \to \min, x \geq 0, y \geq 0, Ax + y = b$
2. The standard form $cx + d \to \min, x \geq 0, Ax = b$
can be rewritten in the following canonical form:
$cx + d \to \min, x \geq 0, Ax \leq b, -Ax \leq -b.$

4. Standard row tableaux.

Definitions:

A *standard row tableau* (of [Vas00]) is

$$
\begin{array}{cc}
x & 1 \\
\left[\begin{array}{cc} A & b \\ c & d \end{array}\right] & \begin{array}{l} = u \\ \to \min, \ x \geq 0, u \geq 0. \end{array}
\end{array}
\qquad \text{(SRT)}
$$

with given matrix A, column b, and row c, where all decision variables in x, u are distinct.

This tableau means the following linear program:
$Ax^T + b = u \geq 0, x \geq 0, cx^T + d \to \min.$

The *basic solution* for the standard tableau (SRT) is $x = 0, u = b$. The corresponding value for the objective function is d.

A standard tableau (SRT) is called *row feasible* if $b \geq 0$, i./e., the basic solution is feasible. Graphically, a feasible tableau looks like

$$
\begin{array}{cc}
\oplus & 1 \\
\left[\begin{array}{cc} * & \oplus \\ * & * \end{array}\right] & \begin{array}{l} = \oplus \\ \to \min \end{array}
\end{array}
$$

where $\oplus$ stands for nonnegative entries or variables.

A standard tableau (SRT) is called *optimal* if $b \geq 0$ and $c \geq 0$. Graphically, an optimal tableau looks like

$$
\begin{array}{cc}
\oplus & 1 \\
\left[\begin{array}{cc} * & \oplus \\ \oplus & * \end{array}\right] & \begin{array}{l} = \oplus \\ \to \min. \end{array}
\end{array}
$$

A *bad row* in a standard tableau is a row of the form $[\ominus -] = \oplus$.
where $\ominus$ stands for nonpositive entries, $-$ stands for a negative number, and $\oplus$ stands for a nonnegative variable.

A *bad column* in a standard tableau is a column of the form $\left[\begin{array}{c} \oplus \\ \oplus \\ - \end{array}\right]$, where $\oplus$ stands

for a nonnegative variable and nonnegative numbers and $-$ stands for a negative number.

9

Facts:

1. Standard row tableaux are used in simplex method below.
2. The canonical form above can be written in a standard tableau as follows:

$$\begin{array}{cc} x^T & 1 \end{array}$$
$$\left[\begin{array}{cc} -A & b \\ c & d \end{array} \right] \begin{array}{l} = u \\ \to \min, \ x \geq 0, u \geq 0 \end{array}$$

where $u = b - Ax \geq 0$.

3. The standard form above can be transform into canonical form

$cx + d \to \min, x \geq 0, -Ax \leq -b, Ax \leq b$

which gives the following standard tableau

$$\begin{array}{cc} x^T & 1 \end{array}$$
$$\left[\begin{array}{cc} -A & b \\ A & -b \\ c & d \end{array} \right] \begin{array}{l} = u \\ = v \\ \to \min, \ x \geq 0; u, v \geq 0 \end{array}$$

To get a smaller standard tableau, we can solve the system of linear equations $Ax = b$. If there are no solutions, the linear program is infeasible. Otherwise, we can write the answer in the form $y = Bz + b'$, where the column y contains some variables in x and the column z consists of the rest of variables. This gives the standard tableau

$$\begin{array}{cc} z^T & 1 \end{array}$$
$$\left[\begin{array}{cc} B & b' \\ c' & d' \end{array} \right] \begin{array}{l} = y \\ \to \min, \ x \geq 0, y \geq 0 \end{array}$$

where $c'z + d'$ is the objective function $cx + b$ expressed in the terms of z.

4. The basic solution of an optimal tableau is optimal.

5. An optimal tableau allows us to describe all optimal solutions as follows. The variables on top with nonzero last entries in the corresponding columns must be zeros. Crossing out these columns and the last row, we obtain a system of linear constraints on the remaining variables describing the set of all optimal solutions. In particular, the basic solution is the only optimal solution if all entries in the c-part are positive.

6. A bad row shows that the linear program is infeasible.

7. A bad column in a feasible tableau shows that the linear program is is unbounded.

5. Pivoting.

Definitions:

Given a system of liner equations $Y = AZ + B$ solved for m variables, a *pivot step* solves it for a subset of m variables which differs from Y in one variable.

Variables in Y are called *basic variables*. The variables in in Z are called *nonbasic variables*.

Thus, a pivot step switches a basic and a nonbasic variable. In other words, one variable leaves the basis and another variable enters the basis.

Here is the pivot rule:

$$
\begin{array}{cc}
x & y \\
\left[\begin{array}{cc} \alpha^* & \beta \\ \gamma & \delta \end{array} \right] & \begin{array}{c} = u \\ = v \end{array}
\end{array}
\mapsto
\begin{array}{cc}
u & y \\
\left[\begin{array}{cc} 1/\alpha & -\beta/\alpha \\ \gamma/\alpha & \delta - \beta\gamma/\alpha \end{array} \right] & \begin{array}{c} = x \\ = v. \end{array}
\end{array}
$$

We switch the two variables x and u. The pivot entry α marked by * must be nonzero; β represents any entry in the pivot row which is not the pivot entry; γ represents any entry in the pivot column which is not the pivot entry; δ represents any entry outside the pivot row and column.

Fact:

A way to solve any system $Ax = b$ system of linear equations is to write it in the row tableau
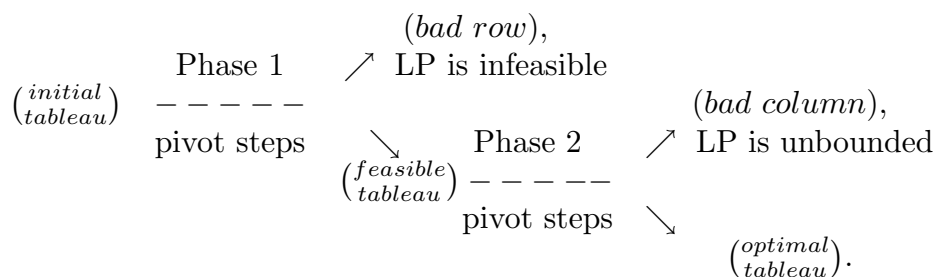
$$
\begin{array}{c}
x^T \\
[A] \quad = b
\end{array}
$$

and moved as many constants from the right margin to the top margin by pivot step. After this we drop the rows which read $c = c$ with a constant c. If one of the rows reads $c_1 = c_2$ with distinct constants c_1, c_2, the system is infeasible. Otherwise, the terminal tableau either has no constants remaining at the right margin, hence we obtain the answer in the form $y = Bz + b'$.

This method requires somewhat more computations that the Gauss elimination, but it solves the system with parametric b.

6. Simplex method.

The simplex method is most common method for solving linear programs. It was suggested by Fourier for linear programs arising from linear approximation (see below). Important early contributions to linear programming were made by L. Walras and L. Kantorovich. The fortuitous synchronization of the advent of the computer and George B. Dantzig's reinvention of the simplex algorithm in 1947 contributed to the explosive development of linear programming with applications to economics, business, industrial engineering, actuarial sciences, operations research, and game theory. For their work in linear programming. P. Samuelson (b. 1915) was awarded the Nobel Prize in Economics in 1970, and L. Kantorovich (1912–1986) and T. C. Koopmans (1910–1985) received the Nobel Prize in Economics in 1975.

Now we give the simplex method in terms of standard tableaux. The method consists of finitely many pivot steps, separated in two phases (stages). In Phase 1, we obtain a feasible tableau or a bad column. In Phase 2 we work with feasible tableaux. We obtain a bad column or an optimal tableau. Here is the scheme of the simplex method where LP stands fot the linear program:

$$\begin{pmatrix} initial \\ tableau \end{pmatrix} \quad \overset{\text{Phase 1}}{\underset{\text{pivot steps}}{- - - - -}} \quad \searrow \quad \overset{(bad\ row),}{\text{LP is infeasible}}$$

$$\begin{pmatrix} feasible \\ tableau \end{pmatrix} \quad \overset{\text{Phase 2}}{\underset{\text{pivot steps}}{- - - - -}} \quad \nearrow \quad \overset{(bad\ column),}{\text{LP is unbounded}} \quad \searrow \quad \begin{pmatrix} optimal \\ tableau \end{pmatrix}.$$

Both phases are similar and can be reduced to each other. Here is Phases 2 and 1 in detail, with a programming loop in 4 substeps for each.

Phase 2 of simplex method. We start with a feasible tableau (SRT).

1. Is the tableau optimal? If yes, we write the answer: $\min = d$ at $x = 0, u = b$.

2. Are there any bad columns? If yes, the linear program is unbounded.

3. (Choosing a pivot entry.) We choose a negative entry in the c-part, say, c_j in column j. Then we consider all negative entries $a_{i,j}$ above to choose our pivot entry. For every $a_{i,j} < 0$ we compute $b_i/a_{i,j}$ where b_i is the last entry in the row i. Then we maximize $b_i/a_{i,j}$ to obtain our pivot entry $a_{i,j}$.

4. Pivot and go to Substep 1.

Phase 1 of simplex method. We start with a standard tableau (SRT).

1. Is the tableau feasible? If yes, go to Phase 2.

2. Are there bad rows? If yes, the LP is infeasible.

3. (Choosing a pivot entry) Let the first negative number in the b-part be in the row i. Consider the subtableau consisting of the first i rows with the i-th row multiplied by -1, and choose the pivot entry as in Substep 3 of Phase 2.

4. Pivot and go to Substep 1.

Definitions:

A pivot step (in Pase 1 or 2) is *degenerate* if the last entry in the pivot row is 0, i.e., the (basic) feasible solution stays the same.

A *cycle* in the simplex method (Phase 1 or 2) is a finite sequence of pivot steps which starts and ends with the same tableau.

Facts:

1. In Phase 2, the current values of the objective function (the last entry in the last row) either improves (decreases) or stays the same (if and only if the pivot step is degenerate).

1. In Phase 1, the first negative entry in the last column increases or stays the same (if and only if the pivot step is degenerate).

3. Following this method, we ether terminate in finitely many pivot steps, or after a while all our steps are degenerate, i.e., the basic solution does not change and we have a cycle.

4. The basic solutions in a cycle are all the same.

5. To prevent cycling we can make a perturbation (small change in the b-part) such that none of the entries in this part of the tableau is ever 0 (see [Vas00] for detail).

6. Another approach was suggested by Bland. We can make an ordered list of our variables, and then whenever there is a choice we choose the variable highest on the list (see [Vas00] for a reference and the proof that this rule prevents cycling). Bland's rule also turns the simplex method into a deterministic algorithm.

7. With Bland's rule, the simplex method (both phases) terminates in at most $\binom{m+n}{n} - 1$ pivot steps where m is the number of basis variables and n is the number of nonbasic variables (so the tableau has $m+1$ rows and $n+1$ columns). The number $\binom{m+n}{n}$ here is an upper bound for the number of all standard tableaux which can be obtained from the initial tableau by pivot steps, up to permutation of the variables on the top (between themselves) and permutation of the variables at the right margin (between themselves).

8. Thus, every linear program is either infeasible, or unbounded, or has an optimal solution.

9. If all date for a LP are rational numbers, then all entries of all standard tableaux are rational numbers. In particular, all basic solutions are rational. If the LP is feasible and bounded, then there is an optimal solution in rational numbers. However, like for system of linear equations, the numerators and denominators could be so large that finding them could be impractical.

7. Geometric interpretation of Phase 2.

Definitions:

1. A subset S of $\mathbf{R}^N$ is *closed* if S contains the limit of any convergent sequence in S.

2. A *convex linear combination* or *mixture* of two points x, y is $\alpha x + (1 - \alpha)y$ where $0 \leq \alpha \leq 1$.

3. A set S is *convex* if it contains all mixtures of its points.

4. An *extreme point* in a convex set is a point which is not the halfsum of any two distinct points in the set, i.e., the set stays convex after removing the point.

5. Two extreme points in a convex set are *adjacent* if the set stays convex after deleting the line segment connecting the points.

Facts:

1. The feasible region for any linear program is a closed convex set with finitely many extreme points (vertices).

2. The vertices of the feasible region are the the basic solutions of the feasible tableaux. Permutation of rows or columns does not chage the basic solution. A degenerate pivot steps do not change the basis solution.

3. Any pivot step takes us from a vertex to an adjacent vertex.

4. The set of optimal solutions for any linear program is a closed convex set with finitely many extreme points (vertices).

5. The number of pivot steps in Phase 2 without cycles (say, with Bland's rule) is less than the number of verices in the feasible region.

6. For (ST) with $m+1$ rows and $n+1$ columns, the number of vertices in the feasible region is $\leq \binom{m+n}{m}$. When $n = 1$ this upper bound can be improved to 2. When $n = 2$,

this upper bound can be improved to $m+2$. It is unknown whether a polynomial in $m+n$ bound exists in general.

8. Duality.

Definitions:

A *standard column tableau* has the form

$$
\begin{array}{c}
-y \\
1
\end{array}
\left[
\begin{array}{cc}
A & b \\
c & d
\end{array}
\right]
\qquad y \geq 0, v \geq 0
\tag{SCT}
$$
$$
\begin{array}{cc}
\| & \downarrow \\
v & \max
\end{array}
$$

The *basic solution* associated with (SCT) is $y = 0, v = c$.

The tableau (SCT) is called *column feasible* if $c \geq 0$, i.e., the basic solution is feasible. (So a tableau is optimal if an only if it is both row and column feasible.)

The linear program in (SCT) is called *dual* to that in (SRT). We can write both as the row and the column problem with the same matrix:

$$
\begin{array}{c}
\\
-y \\
1 \\
\\
\end{array}
\begin{array}{cc}
x & 1 \\
\left[
\begin{array}{cc}
A & b \\
c & d
\end{array}
\right.
\\
= v \quad = w
\end{array}
\begin{array}{l}
\\
= u \\
= z \rightarrow \min \\
\rightarrow \quad \max.
\end{array}
\qquad
\begin{array}{l}
\\
x \geq 0, u \geq 0 \\
y \geq 0, v \geq 0
\end{array}
\tag{ST}
$$

Facts:

1. The linear program in the standard row tableau (SRT) can be written in the following standard column tableau:

$$
\begin{array}{c}
-x^T \\
1
\end{array}
\left[
\begin{array}{cc}
-A^T & c^T \\
b^T & -d
\end{array}
\right]
\qquad x \geq 0, u \geq 0
$$
$$
\begin{array}{cc}
\| & \downarrow \\
u^T & \max.
\end{array}
$$

Then the dual problem becomes the row problem with the same matrix. This shows that the dual of the dual is the primal program.

2. The pivot rule for column and row tableaux is the same inside tableaux:

$$
\begin{array}{c}
-x \\
-u
\end{array}
\left[
\begin{array}{cc}
\alpha^* & \beta \\
\gamma & \delta
\end{array}
\right]
\mapsto
\begin{array}{c}
-u \\
-u
\end{array}
\left[
\begin{array}{cc}
1/\alpha & -\beta/\alpha \\
\gamma/\alpha & \delta - \beta\gamma/\alpha
\end{array}
\right].
$$
$$
\begin{array}{cccc}
\| & \| & & \| & \| \\
u & v & & x & v
\end{array}
$$

3. If a linear program has an optimal solution, then the simplex method produces an optimal tableau. This tableau is also optimal for the dual program, hence programs have the same optimal values (the duality theorem).

4. The duality theorem is a deep fact with several interpretations and applications. Geometrically, the duality theorem means that certain convex sets can be separated from points outside them by hyperplanes. We will see another interpretation of the duality theorem in the theory of matrix games (see below). For problems in economics, the dual

problems and the duality theorem also have important economic interpretations (see examples below).

5. If a linear program is unbounded, then (after writing it in a standard row tableau) the simplex method produces a row feasible tableau with a bad column. The bad column shows that the dual problem is infeasible.

6. There is a standard tableau which have both a bad row and a bad column, hence there is an infeasible linear program such that the dual program is also infeasible.

7. Here is another way to express the duality theorem: given a feasible solution for a linear problem and a feasible solution for the dual problem, they are both optimal if and only if the feasible values are the same.

8. Given a feasible solution for a linear problem and a feasible solution for the dual problem, they are both optimal if an only if for every pair of dual variables at least one value is 0, i.e., the *complementary slackness* condition $vx^T + y^T u = 0$ in terms of (ST) holds.

9. More precisely, given a feasible solution (x, u) for the row program in (ST) and a a feasible solution (y, v) for the column program, the difference $cx^T + d) - (-y^T u + d)$ between the corresponding feasible values is $vx^T + y^T u$.

10. Thus, all pairs $(x, u), (y, v)$ of optimal solutions for the row and column programs in (ST) are described by the following system of linear constraints: $Ax^T + b = u \geq 0, x \geq 0, -y^T A + c = v \geq 0, y \geq 0, vx^T + y^T u = 0$. This is a way to show that solving a liner program can be reduced to finding a feasible solution for a finite system of linear constraints.

Examples (generalizations of Examples 1–4 in Section 2 above and their duals):

1. The linear programs $c_1 x_1 + \cdots + c_n x_n \to$ max, all $x_i \geq 0$, $x_1 + \cdots + x_n = 1$ and $y \to$ min, $y \geq c_i$ for all i in Example 1 are dual to each other. To see this, we use the standard tricks: $y = y' - y''$ with $y', y'' \geq 0$; $x_1 + \cdots + x_n = 1$ is equivalent to

$x_1 + \cdots + x_n \leq 1, -x_1 - \cdots - x_n \leq -1$.

We obtain the following standard tableau:

$$
\begin{array}{c}
\begin{array}{ccc} y' & y'' & 1 \end{array} \\
\begin{array}{c} -x^T \\ 1 \end{array}
\left[
\begin{array}{ccc}
J & -J & -c \\
1 & -1 & 0
\end{array}
\right]
\begin{array}{c} = \oplus \\ = y \;\to\; \min \end{array} \\
\begin{array}{ccc} = \oplus & = \oplus & \to \quad \max, \end{array}
\end{array}
$$

where J is the column of n ones and $c = [c_1, \ldots, c_n]^T$.

2. Consider the general diet problem (a generalization of Example 2):

$$Ax \geq b, x \geq 0, C = cx \to \min,$$

where m variables in x represent different foods and n constraints in $Ax \geq b$ represent ingredients. We want to satisfy given requirements b in ingredients using given foods at minimal cost C.

On the other hand, we consider a warehouse that sells the ingredients at prices $y_1, \ldots, y_n \geq 0$. Its objective is to maximize the profit $P = yb$, matching the price for each food: $yA \leq c$.

We can write both problems in a standard tableau using slack variables $u = Ax - b \geq 0$ and $v = c - yA \geq 0$:

$$
\begin{array}{cc}
 & \begin{array}{cc} x^T & 1 \end{array} \\
\begin{array}{c} -y^T \\ 1 \end{array} &
\left[\begin{array}{cc} A & -b \\ c & 0 \end{array} \right]
\begin{array}{l} = u \\ = C \to \min \end{array}
\begin{array}{l} x \geq 0, u \geq 0 \\ y \geq 0, v \geq 0 \end{array} \\
 & \begin{array}{cc} = v & = P \\ & \to \quad \max. \end{array}
\end{array}
$$

So these two problems are dual to each other. In particular, the simplex method solves both problems, and if both problems are feasible, then $\min(C) = \max(P)$. The shadow prices mentioned in §14 turn out to be the optimal prices for the ingredients in the dual problem, so they are called *dual prices* as well.

Another economic interpretation for the same mathematical problem is that the variables in x are intensities of different industrial processes, the constraints correspond to different products, with b being the federal order to be fulfilled (or demand to be satisfied); $C = cx$ is the total cost that you want to minimize using given processes and satisfying given production quotas. With this interpretation for the primal problem, here is an interpretation for the dual problem: A competitor, Ann, who lost the government contract, says that you van you to buy the products from her at her low prices $y \geq 0$, matching unit cost for every process you got (i.e., $yA \leq c$) and maximizing her profit yb.

With this interpretation, the optimal prices y are *marginal costs* for you (i.e., they answer to the question what is the additional cost to produce an additional unit of each product). In §14, we saw that the marginal costs decrease with increase in volume (in linear programming).

3. Consider the general mixing problem (a generalization of Example 3):

$$Ax = b, x \geq 0, C = cx \to \min,$$

where m variables in x represent different alloys and n constraints in $Ax \geq b$ represent elements. We want to satisfy given requirements b in elements using given alloys at minimal cost C.

On the other hand, consider a dealer who buys and sells the elements at prices $y_1, \ldots, y_n$. The positive price means that the dealer sells, and negative price means that the dealer buys. Dealer's objective is to maximize the profit $P = yb$, matching the price for each alloy: $yA \leq c$.

To write the problems in standard tableaux, we use the standard tricks and artificial variables:

$u' = Ax - b \geq 0$, $u'' = -Ax + b \geq 0$;
$v = c - yA \geq 0$; $y = y' - y''$, $y' \geq 0, y'' \geq 0$.

Now we manage to write both problems in the same standard tableau:

$$
\begin{array}{cc}
 & \begin{array}{cc} x^T & 1 \end{array} \\
\begin{array}{c} -y'^T \\ -y''^T \\ 1 \end{array} &
\left[\begin{array}{cc} A & -b \\ -A & b \\ c & 0 \end{array} \right]
\begin{array}{l} = u' \\ = u'' \\ = C \to \min \end{array}
\begin{array}{l} x \geq 0, u \geq 0 \\ y', y'' \geq 0, v \geq 0 \end{array} \\
 & \begin{array}{cc} = v & = P \\ & \to \quad \max. \end{array}
\end{array}
$$

4. Consider a generalization of the manufacturing problem in Example 4:

$$P = cx \to \max, \ Ax \le b, \ x \ge 0,$$

where the variables in x are the amounts of products, P is the profit (or revenue) you want to maximize, constraints $Ax \le b$ correspond to resources (e.g., labor of different types, clean water you use, pollutants you emit, scarce raw materials), and the given column b consists of amounts of resources you have. Then the dual problem

$$yb \to \min, \ yA \ge c, \ y \ge 0$$

admits the following interpretation. Your competitor, Bob offers to buy you out at the following terms: You go out of business, and he buys all resources you have at price $y \ge 0$, matching your profit for every product you may want to produce, and he wants to minimize his cost.

Again Bob's optimal prices are your resource shadow prices by the duality theorem. The shadow price for a resource shows the increase in your profit per unit increase in the quantity b_0 of the resource available or decrease in the profit when the limit b_o decreases by one unit. While changing b_0 we do not change the limits for the other resources and any other data for our program. There are only finitely many values of b_0 for which the downward and upward shadow prices are different. One of these values could be the borderline between the values of b_0 for which the corresponding constraint is binding or nonbinding (in the sense that dropping of this constraint does not change the optimal value).

The shadow price of a resource cannot increase when supply b_0 of this resource increases (the law of diminishing returns, see the next section).

5. (General transportation problem and its dual). We have m warehouses and n retail stores. The warehouse $\# i$ has a_i widgets available and the store $\# j$ needs b_j widgets.

It is assumed that the following *balance condition* holds:

$$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j.$$

If total supply is greater than total demand, the problem can be reduced to the one with the balance condition by introducing a fictitious store where the surplus can be moved at zero cost. If total supply is less than total demand,, the program is infeasible.

The cost of shipping a widgets from warehouse $\# i$ to store $\# j$ is denoted by c_{ij} and the number of widgets shipped from warehouse $\# i$ to store $\# j$ is denoted by x_{ij}. The linear program can be stated as follows:

$$
\begin{cases}
\text{minimize} & C(x_{11}, \ldots, x_{mn}) = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} \, x_{ij} \\
\text{subject to} & \sum_{i=1}^{n} x_{ij} \ge b_j, \quad j = 1, \ldots, m \\
& \sum_{j=1}^{m} x_{ij} \le a_i, \quad i = 1, \ldots, n \\
& x_{ij} \ge 0, \ i = 1, \ldots, n, \ j = 1, \ldots, m.
\end{cases}
$$

The dual program is

$$-\sum_{i=1}^{m} a_i u_i + \sum_{j=1}^{n} b_j v_j \to \max, w_{ij} = c_{ij} + u_i - v_j \geq 0 \text{ for all } i, j; u \geq 0, v \geq 0.$$

So what is a possible meaning of the dual problem? The control variables u_i, v_j of the dual problem are called *potentials* or "zones". While the potentials correspond to the constraints on each retail store and each warehouse (or to the corresponding slack variables), there are other variables w_{ij} in the dual problem that correspond to the decision variable x_{ij} of the primal problem.

Imagine that you want to be a mover and suggest a simplified system of tariffs. Instead mn numbers $c_{i,j}$, we use $m + n$ "zones." Namely, you assign a "zone" $u_i \geq 0$ $i = 1, 2$ to each of the warehouses and a "zone" $v_j \geq 0$ $j = 1, 2$ to each of the retail stores. The price you charge is $v_j - u_i$ instead of $c_{i,j}$. To beat competition, you want $v_j - u_i \leq c_{i,j}$ foe all i, j. Your profit is $-\sum a_i u_i + \sum b_j v_j$ and you want to maximize it.

The simplex method for any transportation problem can be implemented using m by n tables rather than $mn + 1$ by $m + n + 1$ tableaux. Since all pivot entries are ± 1, no division is used. In particular, if all a_i, b_j are integers, we obtain an optimal solution with integral $x_{i,j}$.

Phase 1 is especially simple. If $m, n \geq 2$, we choose any position and write down the maximal possible number (namely, the minimum of supply and demand). Then we cross out the row or column and adjust the demand or supply respectively. If $m = 1 < n$, we cross out the column. If $n = 1 < m$, we cross out the row. If $m = n = 1$ and we cross out both the row and column. Thus, we find a feasible solution in $m + n - 1$ steps, which correspond to pivot steps, and the $m + n - 1$ selected positions correspond to the basic variables.

Every transportation problem with the balance condition has an optimal solution. See [Vas00] for details.

9. Sensitivity Analysis and Parametric Programming.

Sensitivity analysis is concerned with how small changes in data affect the optimal value and optimal solutions, while large changes are studied in parametric programming.

Consider the linear program given by (SRT) with the last column being an affine function of a parameter t, i.e., we replace b, d by affine functions $b + b_1 t, d + d_1 t$ of a parameter t. Then the optimal value becomes a function $f(t)$ of t.

Facts:

1. The function $f(t)$ is defined on a closed convex set S on the line, i.e., S is one of the following "intervals": empty set, a point, an interval $a \leq t \leq b$ with $a < b$, a ray $t \geq a$, a ray $t \leq a$, the whole line.

2. The function $f(t)$ is piece-wise linear, i.e., the "intervals" S is a finite union of "sub-intervals" S_k with $f(t)$ being an affine function on each.

3. The function $f(t)$ is *convex* in the sense that the set of points in the plane below the plot is a convex set. In particular, the function $f(t)$ is continuous.

4. In Parametric Programming there are methods for computing $f(t)$. The set S is covered by a finite set of tableaux optimal for various values of t. The number is these tableaux is at most $\binom{m+n}{m}$.

5. Suppose that the LP with $t = t_0$ (i.e., the LP given by (SRT)) has an optimal tableau T_0. Let $x = x^(0), u = u^(0)$ be the corresponding optimal solution (the basic solution), and let $y = y(0), v = v^(0)$ be the corresponding optimal solution for the dual problem (see (ST) for notation). Assume that the b-part of T_0 has no zero entries. Then the faction $f(t)$ is affine in an interval containing 0. Its slope, i.e., derivative $f'(0)$ at $t = 0$ is $f'(0) = d_1 + b_1 v^(0)$. Thus, we can easily compute $f'(0)$ from T_0. In other words, the optimal tableau give the partial derivatives of the optimal value with respect to the components of given b, d.

6. If we pivot the tableau with parameter, the last column stays an affine function of parameter and the rest of the tableau stays independent of parameter.

6. Similar facts are true if we introduce a parameter into the last row rather than into last column.

7. If both the last row and the last column are affine functions of parameter t, then after pivot steps, the A-part stays independent of t,, b-part and c-part stay affine functions of t, but the d-part becomes a quadratic polynomial in t. So the optimal value is a piece-wise quadratic function of t. It is still convex. "d"-part

8. If we want maximize, say, profit (rather than minimize, say, cost), then the optimal value is concave (convex upward), i.e., the set of points below the graph is convex. The fact that the slope is non-increasing is referred to as the *law of diminishing returns.*

10. Matrix games.

Matrix games are very closely related with linear programming.

Definitions:

A matrix game is given by a matrix A, the *payoff matrix,* of real numbers.

There are two players. The players could be humans, teams, computers, animals. We call them He and She.

He chooses a row, and she chooses a column. The corresponding entry in the matrix represents what she pays to him (the payoff of the row player). Games like chess, football, and blackjack can be thought as (very large) matrix games. Every row represents a strategy, i.e., his decision what to do in every possible situation. Similarly, every column corresponds to a strategy for her. In general, we speak about rows as his (pure) strategies and columns as her (pure) strategies.

A *mixed strategy* is a mixture of pure strategies. In other words, a mixed strategy for him is a probability distribution on the rows and a mixed strategy for here is a probability distribution on the columns.

We write his mixed strategy as columns $p = (p_i)$ with $p \geq 0, \sum p_i = 1$. We write her mixed strategy as rows $q = (q_j)$ with $q \geq 0, \sum q_j = 1$.

The corresponding payoff is $p^T A q^T$, the mathematical expectation.

A pair of strategies (his strategy, her strategy) is called an *equilibrium* or a *saddle point* if neither player can gain by changing his or her strategy. In other words, no player can do better by an unilateral change.

In other words, at an equilibrium, the payoff $p^T A q^T$ has maximum as function of p and minimum as function of q.

A pair (i, j) of pure strategies is an equilibrium if the entry $a_{i,j}$ of the payoff matrix $A = a_{i,j}$ is both largest in its column and the smallest in its row.

His mixed strategy p is called *optimal*, if the his worst case payoff $\min(p^T A)$ is maximal. The maximum is called the *values of game* (for him). Her mixed strategy q is called *optimal*, if the her worst case payoff $\min(-Aq^T)$ is maximal. The maximum is called the value of game for her.

Facts:

1. Every matrix game has an equilibrium

2. A pair (his strategy, here strategy) is an equilibrium if and only if both strategies are optimal.

3. His payoff $p^T A q^T$ at any equilibrium (p, q) equals to his value of game and equals to the negative of her value of game (the *minimax theorem)*.

4. To solve a matrix game means to find an equilibrium and the value of game.

5. Solving a a matrix game can be reduces to solving the following dual pair of linear programs written in a standard tableau:

$$
\begin{array}{cc}
& \begin{array}{cccc} q & \mu' & \mu'' & 1 \end{array} \\
\begin{array}{c} -p \\ -\lambda' \\ -\lambda'' \\ 1 \end{array} &
\left[\begin{array}{cccc}
-A & J' & -J' & 0 \\
J & 0 & 0 & -1 \\
-J & 0 & 0 & 1 \\
0 & 1 & -1 & 0
\end{array}\right]
\begin{array}{l}
= * \geq 0 \\
= * \geq 0 \\
= * \geq 0 \\
= \mu \to \min
\end{array}
\end{array}
$$

$$
\begin{array}{cccc}
\| & \| & \| & \| \\
* & * & * & \lambda \to \quad \max.
\end{array}
$$

Here A is the m by n payoff matrix, J is the row of n ones, J' is the column of m ones, p is his mixed strategy, q is here mixed strategy. Note that her problem is the row problem, and his problem is the column problem. Their problems are dual to each other. Since both problems have feasible solutions (take, for example, $p = J'/m, q = J/n$), the duality theorem says that $\min(f) = \max(g)$. That is, his value of game $= \max(\lambda) = \min(\mu) = -$her value of game. Thus, the minimax theorem follows from the duality theorem.

6. We can save two rows and two columns and get a row feasible tableau as follows:

$$
\begin{array}{cc}
& \begin{array}{cc} q/(\mu + c) & 1 \end{array} \\
\begin{array}{c} -p/(\lambda + c) \\ 1 \end{array} &
\left[\begin{array}{cc}
-A - cJ'J & J' \\
-J & 0
\end{array}\right]
\begin{array}{l}
= * \geq 0 \\
= -1/(\mu + c) \to \min
\end{array}
\end{array}
$$

$$
\begin{array}{cc}
\| & \| \\
* & -1/(\lambda + c) \quad \to \max
\end{array}
$$

Here we made sure that the value of game is positive by adding a number c to all entries of A, i.e., replacing A by $A + cJ'J$.E.g., $c = 1 - \min(A)$. Again his and her problems are dual to each other, since they share the same standard tableau. Since the tableau is feasible, we bypass Phase 1.

7. An arbitrary dual pair (ST) of linear programs can reduced to a matrix game. At first glance it seems impossible, because the linear programs might not be feasible, while every matrix game has optimal strategies. However, we go ahead and consider the following matrix:

$$M = \begin{bmatrix} 0 & -A & -b \\ A^T & 0 & -c^T \\ b^T & c & 0 \end{bmatrix}.$$

Its size is $(m+n+1) \times (m+n+1)$, where $m \times n$ is the size of the matrix A. Since $M^T = -M$. The corresponding matrix game is symmetric, i.e., $M^T = -M$. . No player has an advantage. So the value of the game is 0. Indeed, suppose that he can win a number $\varepsilon > 0$ (no matter what she does); i.e., $p^T M \geq \varepsilon > 0$ for a column $p \geq 0$. Then she can use the same strategy $q = p^T$ against him and she wins at least the same number ε no matter what he does. When both use this strategy, his payoff $p^T A q^T$ should be both positive and negative, which leads to a contradiction.

Suppose $x = \bar{x}, u = \bar{u}$ is an optimal solution for the row problem and $y = \bar{y}, v = \bar{v}$ is an optimal solution for the column problem. Set e to be 1 plus the sum of all entries in $\bar{x}$ and $\bar{y}$.

Then $\bar{p} = [\bar{y}^T, \bar{x}, 1]^T/e$ is a mixed strategy for him. It gives the following payoff:

$$\bar{p}^T M = [\bar{y}^T, \bar{x}, 1]M/e$$

$$= [\bar{x}A^T + b^T, -\bar{y}^T A + c, -\bar{y}^T b - \bar{x}c^T)/e = [\bar{u}^T, \bar{v}, 0]/e \geq 0$$

(we have used that $z = w$ for optimal solutions). Thus, $\bar{p}$ is an optimal mixed strategy for him. Note that its last entry $1/e$ of the column $\bar{p}$ is not 0.

Conversely, given any optimal strategy $p = \bar{p}$ for him (that is, $\bar{p}$ is an optimal strategy for her) with a nonzero last entry, say $1/e$, we can write $\bar{p}^T = [\bar{y}^T, \bar{x}, 1]/e$ with nonnegative rows $\bar{y}^T, \bar{x}$ of appropriate sizes. Since $\bar{p}$ is optimal, $\bar{p}^T M \geq 0$. So

$$\bar{x}A^T + b^T \geq 0, -\bar{y}^T A + c \geq 0, -\bar{y}^T b - \bar{x}c^T \geq 0.$$

This shows that $x = \bar{x}, u = A\bar{x}^T + b$ is an optimal solution for the preceding row problem and $y = \bar{y}, v = c - \bar{y}^T A$ is an optimal solution for the column programs [recall again that $\min(z) = \max(w))$.]

Thus, there is a 1-1 correspondence between the optimal strategies of the game with nonzero last entries and the optimal solutions of the two linear programs. If there is no optimal strategy with a nonzero last entry, then there are no optimal solutions for the linear programs.

8. The definition of equilibria makes sense for any game (not only for two-player zero-sum games). However finding equilibria is not the same as solving game when there are equilibria with different payoffs or when cooperation between players is possible and makes sense.

Examples:

21

1. (Example 19.3 of [Vas03]) Solve the matrix game

$$A = \begin{bmatrix} 5 & 0 & 6 & 1 & -2 \\ 2 & 1 & 2 & 1 & 2 \\ -9 & 0 & 5 & 2 & -9 \\ -9 & -8 & 0 & 4 & 2 \end{bmatrix}.$$

We mark the maximal entries in every column by *. Then we mark the minimal entries in each row by '. The positions marked by both * and ' are exactly the saddle points:

$$A = \begin{bmatrix} 5^* & 0 & 6^* & 1 & -2' \\ 2 & 1^{*'} & 2 & 1' & 2^* \\ -9' & 0 & 5 & 2 & -9' \\ -9' & -8 & 0 & 4^* & 2^* \end{bmatrix}.$$

In this example, the position $(i, j) = (2, 2)$ is the only saddle point. The corresponding payoff (the value of game) is 1.

2. This small matrix game is known as Heads and Tails or Matching Pennies. We will call the players *He* and *She*. He chooses: heads (H) or tails (T). Independently, she chooses: H or T. If they choose the same, he pays her a penny. Otherwise, she pays him a penny. Here is his payoff in cents:

$$\begin{array}{cc} & \textit{She} \\ & \begin{array}{cc} H & T \end{array} \\ \textit{He} \begin{array}{c} H \\ T \end{array} & \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}. \end{array}$$

There is no equilibrium in pure strategies. The only equilibrium in mixed strategies is $((H+T)/2, (H+T)/2)$. The value of game is 0. The game is not symmetric in the usual sense. However the game is symmetric in the following sense: if we switch the players and also switch H and T for a player, then we get the same game.

3. Another game is Rock, Scissors, Paper. In this game two players simultaneously choose *Rock*, *Scissors*, or *Paper*, usually by a show of hand signals on the count of three, and the payoff function is defined by the rules *Rock breaks Scissors, Scissors cuts Paper, Paper covers Rock*, and every strategy ties against itself. Valuing a win at 1, a tie at 0, and a loss at -1, we can represent the game with the following matrix, where, for both players, strategy 1 is *Rock*, strategy 2 is *Scissors*, and strategy 3 is *Paper*:

$$A = \begin{bmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix}.$$

The only optimal strategy for the column player is $q = [1/3, 1/3, 1/3]$. Since the game is symmetric, the value is 0, and q^T is the only optimal strategy for the row player.

11. Linear approximation.

Definition:

Given real number $p \geq 1$ and an integer $n \geq 1$, and a column $e = [e_i] \in \mathbf{R}^n$, the l^p-*norm* $\|e\|_p$ of e is

$\|e\|_p = (\sum |e_i^p|)^{1/p}.$

We also define

$\|e\|_\infty = \lim_{p\to\infty} \|e\|_p = \max(|e_i|).$

A l^p-best linear approximation (fit) of a given column w with n entries by the columns of a given m by n matrix A is AX where X is an optimal solution for $\|w - AX\| \to \min.$

In other word, we want the vector $w - AX$ of of *residuals* (offsets, errors) to be smallest in a certain sense.

Facts:

1. Most common values for p are $2, 1, \infty$.

1. In statistics, usually $p = 2$ and the first column of the matrix A is the column of ones. In simple regression analysis, $n = 2$. In multiple regression, $n \geq 3$. In time series analysis, the second column of A is an arithmetic progression representing time; typically, this column is $[1, 2, ..., m]^T$.

2. If $n = 1$ and the matrix A is a column of ones, we want to approximate given numbers w_i by one number Y. When $p = 2$, the best fit is the arithmetic mean $(\sum w_i)/m$. When $p = 1$, the best fits are the medians. When $p = \infty$, the best fit is the midrange $(\min(w_i) + \max(w_i))/2$.

3. When $p = 2$, the l^2-norm is the usual Euclidean norm, the most common way to measure the size of a vector and this norm is used in Euclidean geometry. The best fit is known as the least squares fit. To find it, we drop a perpendicular from w onto the column space of A. In other words, we want the vector $w - AX$ to be orthogonal to all columns of A—that is, $A^T(w - AX) = 0$. This gives a system of n linear equations $A^T AX = A^T w$ for n unknowns in the column X. The system always has a solution. Moreover, the best fit AX is the same for all solutions X. In the case when w belongs to the column space, the best fit is w (this is true for all p). Otherwise, X is unique.

4. The best l^1-fits appeared first in connection with data on star movements. Bosovitch (about 1756), Laplace (1789), Gauss (1809), and Fourier (about 1822) proposed methods of solving those problems. In fact, Fourier considered also l^∞-approxima-tion, and he suggested a method of finding feasible solutions for an arbitrary system of linear constraints. Strangely enough, works on l^2-approximation, where finding the best fit reduces to solving a system of linear equations, appeared only in the nineteenth century (Legendre, Gauss).

5. The best l^∞-fit is also know as the least-absolute-deviation fit and the Chebyshev approximation.

6. When $p = 1$, finding the best fit can be reduced to a linear program. Namely, we reduce the optimization problem with the objective function $\|e\|_1 \to \min$, where $e = (e_i) = w - AX$, to a linear program using m additional variables u_i such that $|e_i| \leq u_i$ for all i. We obtain the following linear program with $m + n$ variables a_j, u_i and $2m$ linear constraints:

$$\sum u_i \to \min, \ -u_i \leq w_i - A_i X \leq u_i \text{ for } i = 1, \ldots, m,$$

where A_i is the ith row of the given matrix A.

7. When $p = \infty$, finding the best fit can be also reduced to a linear program. Namely reduce the optimization problem with the objective function $\|e\|_\infty \to \min$, where $e = (e_i) = w - AX$ to a linear program using an additional variable u such that $e_i| \leq u$ for all i. A similar trick was used when we reduced solving matrix games to linear programming. We obtain the following linear program with $n+1$ variables a_j, u and $2m$ linear constraints:

$$t \to \min, \ -u \leq w_i - A_i X \leq u \text{ for } i = 1, \ldots, m,$$

where A_i is the ith row of the given matrix A.

Examples:

1. [Vas03, Problem 22.7] Find the best l^p-fit $w = ch^2$ for $p = 1, 2, \infty$ given the following data:

i	1	2	3
Height h in m	1.6	1.5	1.7
Weight w in kg	65	60	70

Compare the optimal values for c with those for the best fits of the form $w/h^2 = c$ with the same p (the number w/h^2 in kg/m^2 is known as BMI, the body mass index). Compare the minimums with those for the best fits of the form $w = b$ with the same p.

Solution.

Case $p = 1$. We could convert this problem to a linear program with four variables and then solve it by the simplex method (see Fact 6 above). But we can just consider the nonlinear problem with the objective function

$$f(c) = |65 - 1.6^2 c| + |60 - 1.5^2 c| + |70 - 1.7^2 c|$$

to be minimized and no constraints. The function $f(c)$ is piecewise affine and convex, with the slope changing at

$$c = 70/1.7^2 \approx 24, c = 65/1.6^2 \approx 25, \text{ and } c = 60/1.5^2 \approx 27.$$

The slopes of $f(c)$ are

$-1.6^2 - 1.5^2 - 1.7^2 < 0$ for $c \leq 70/1.7^2$,
$-1.6^2 - 1.5^2 + 1.7^2 \approx -2$ for $70/1.7^2 \leq c \leq 65/1.6^2$,
$1.6^2 - 1.5^2 + 1.7^2 \approx 3$ for $65/1.6^2 \leq c \leq 60/1.5^2$,

and

$1.6^2 + 1.5^2 + 1.7^2 > 0$ for $c \geq 60/1.5^2$.

Now it is clear that $f(c)$ is minimized at

$$c = x_1 = 65/1.6^2 \approx 25.39.$$

This value equals the median of the three observed BMIs

$$65/1.6^2, 60/1.5^2, 70/1.7^2.$$

The optimal value is

$$\min = |65 - c_1 1.6^2| + |60 - c_1 1.5^2| + |70 - c_1 1.7^2| = 6.25.$$

To compare this with the best l^1-fit for the model $w = b$, we compute the median $b = x_1 = 65$ and the corresponding optimal values:

$$\min = |65 - x_1| + |60 - x_1| + |70 - x_1| = 10.$$

So the model $w = ch^2$ is better than $w = b$ for our data with the l^1-approach.

Case $p = 2$. Our optimization problem can be reduced to solving a linear equation for c (see Fact 3 above). Here we solve the problem using calculus, taking the advantage of the fact that our objective function

$$f(c) = (65 - 1.6^2 c)^2 + (60 - 1.5^2 c)^2 + (70 - 1.7^2 c)^2$$

is differentiable. We set $f'(c) = 0$, which gives, after division by -2,

$$1.6^2(65 - 1.6^2 c) + 1.5^2(60 - 1.5^2 c) + 1.7^2(70 - 1.7^2 c) = 0,$$

hence the optimal solution is

$$c = x_2 = 2518500/99841 \approx 25.225.$$

This x_2 is not the mean of the observed BMIs, which is about 25.426. The optimal value is $\min \approx 18$.

The mean of w_i is 65, and the corresponding minimal value is $5^2 + 0^2 + 5^2 = 50$. So again the model $w = ch^2$ is better than $w = b$.

Case $p = \infty$. We could reduce this problem to a linear program with two variables and then solve it by graphical method or simplex method (see Favt 7 above). But we can do a graphical method with one variable. The objective function to minimize now is

$$f(c) = \max(|65 - 1.6^2 c|, |60 - 1.5^2 c|, |70 - 1.7^2 c|).$$

This objective function $f(c)$ is piecewise affine and convex. We can plot the function $f(c)$ and see that the optimal solution $a \approx 25$. Around this point,

$$65 - 1.6^2 c \approx 1, 60 - 1.5^2 c \approx 3.75, 70 - 1.7^2 c \approx -2.25;$$

hence
$$f(c) = \max(60 - 1.5^2 c, -70 + 1.7^2 c).$$

So the exact optimal solution satisfies $60 - 1.5^2 c = -70 + 1.7^2 c$; hence the optimal solution is
$$c_\infty = 6500/257 \approx 25.29.$$

It differs from the midrange of the BMIs, which is about 25.44. The optimal value is ≈ 3.

On the other hand, the midrange of the weights w_i is 65, which gives min $= 5$ for the model $w = b$ with the best l^∞-fit. So again the model $w = ch^2$ is better than $w = b$.

2. [Vas03, Exercise 2 on p. 255] A student is interested in the number w of integer points $[x, y]$ in the disc $x^2 + y^2 \le r^2$ of radius r. He computed w for some r:

r	1	2	3	4	5	6	7	8	9
w	5	13	29	45	81	113	149	197	253

The student wants to approximate w by a simple formula $w = ar + b$ with constants a, b. But you feel that the area of the disc, πr^2 would be a better approximation, and hence the best l^2-fit of the form $w = ar^2$ should work even better for the numbers above.

Compute the best l^2-fit (the least squares fit) for both models, $w = ar + b$ and $w = ar^2$ and find which is better. Also compare both optimal values with

$$\sum_{i=1}^{9} (w_i - \pi i^2)^2.$$

Solution.

For our data, the equation $w = ar + b$ is the system of linear equations $AX = w$, where

$w = [5, 13, 29, 45, 81, 113, 149, 197, 253]^T$,

$A = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}^T$, and $X = \begin{bmatrix} a \\ b \end{bmatrix}$.

The least-squares solutions are solutions to $A^T A X = A^T w$, i.e.,

$$\begin{bmatrix} 285 & 45 \\ 45 & 9 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 6277 \\ 885 \end{bmatrix}.$$

The solution is $a = 463/15, b = -56$. The error

$$\sum_{i=1}^{9} (w_i - 463i/15 + 56)^2 = 48284/15 \approx 3218.93.$$

For $w = ar^2$, the system is $Ba = w$, where w is as above and
$B = [1^2, 3^2, 3^2, 4^2, 5^2, 6^2, 7^2, 8^2, 9^2]^T$.
The equation $B^T B a = B^T w$ is $415398a = 47598$, hence $a = 23799/7699 \approx 3.09118$.
The error

$$\sum_{i=1}^{9} (w_i - 23799i^2/7699)^2 = 2117089/7699 \approx 274.982.$$

So the model $w = ar^2$ gives a much better least-squares fit the the model $w = ar + b$ although it has one parameter instead of two.

The model $w = \pi r^2$ without parameters gives the error

$$\sum_{i=1}^{9} (w_i - \pi i)^2 = 147409 - 95196\pi + 15398\pi^2 \approx 314.114.$$

It is known that $w_i/h_i^2 \to \pi$ as $i \to \infty$. Moreover, $|w_i - \pi r_i^2|/r_i$ is bounded as $i \to \infty$. It follows that $a \to \pi$ for the the best L_p-fit $w = ar^2$ for any $p \geq 1$ as we use data for $r = 1, 2, \ldots n$ with $n \to \infty$.

12. Interior point methods.

See [Vas03] for details and references.

Definitions:

A point x in a subset $X \subset \mathbf{R}^n$ is called *interior* if

$$\{x + (y - x)\delta/\|y - x\|_2 : y \in X, y \neq x\} \subset X$$

for some $\delta > 0$.

The *boundary* of X are the points in X which are not interior.

An *interior solution* for a linear program is an interior ponts of the feasible region.

An interior point method for solving bounded linear programs starts with with given interior solutions produces a sequence of interior solutions which converges to an optimal solution.

An exterior point method for linear programs strts with a point outside the feasible region and produces a sequence of points which converges to feasible solution (in the case when the LP is feasible).

Facts:

1. In linear programming, the problem of finding a feasible solution (Phase 1) and the problem of finding of an optimal solution starting from a feasible solution (Pase 2) can be reduced to each other. So the difference between interior point methods and exterior point methods is not so sharp.

2. The simplex method, Phase 1 can be consider as an exterior point method which (theoretically) converges in finitely many steps. The ellipsoid method is truly exterior point method. Khachian proved an exponential convergence for the method.

3. In simplex method, Phase 2, we travel from a vertex to an adjacent vertex and reach an optimal vertex (if the LP is bounded) in finitely many steps. The vertices are not interior solutions unless the feasible region consists of single point. If an optimal solution for a linear program is interior, then all feasible solutions are optimal.

3. The first interior point method was given by Brown in the context of matrix games. The convergence was proved by J.Robinson. J. von Neumann suggested a similar method with better convergence.

4. Karmerkar suggested an interior point method with exponential convergence. After him, many similar method were suggested. They can be interpreted as follows. We modify

our objective function by adding penalty for approaching the boundary. Then we use the Newton method. The recent progress is related with understanding of properties of convex functions which make minimization by Newton method efficient. Recent interior methods beat simplex methods for very large problems.

5. [Vas03, Section A8] On the other hand, it is known that for LPs with small number of variables (or, by duality, with a small number of constraints), there are faster methods than the simplex method.

For example, consider (SRT) with only 2 variables on top and m variables at the right margin (basis variables). The feasible region S has at most $m + 2$ vertices. Phase 2, starting with any vertex, terminates in at most $m + 1$ pivot steps.

At each pivot step, it takes at most two comparisons to check whether the tableau is optimal or to find a pivot column. Then in m sign checking, at most m divisions, and at most $m - 1$ comparisons we find a pivot entry or a bad column.

Next we pivot to compute the new $3m + 3$ entries of the tableau. In one division we find the new entry in the pivot row that is not the last entry (the last entry was computed before) and in $2m$ multiplications and $2m$ additions we find the new entries outside the pivot row and column. Finally, we find the new entries in the pivot column in $m + 1$ divisions. So a pivot step, including finding a pivot entry and pivoting, can be done in $8m + 3$ *operations*—arithmetic operations and comparisons.

Thus, Phase 2 can be done in $(m + 1)(8m + 3)$ operations, While small savings in this number are possible [e.g., at the $(m + 1)$-th pivot step we need to compute only the last column of the tableau] it is unlikely that any substantial reduction of this number for any modification of the simplex method can be achieved (in the worst case). Concerning the number of pivot steps, for any $m \geq 1$ it is clearly possible for S to be a bounded convex $(m + 2)$-gon, in which case for any vertex there is a linear objective function such that the simplex method requires exactly $\lfloor 1 + n/2 \rfloor$ pivot steps with only one choice of pivot entry at each step. It is also possible to construct an $(m + 2)$-gon, an objective point, and an initial vertex in a way that m pivot steps with unique choice are required (or with two choices at the first step such that the first choice leads to the optimal solution while the second choice leads to m additional pivot steps with unique choice).

But it is possible to do Phase 2 into $\leq 100m + 100$ operations.

References

[Vas00] L.N. Vaserstein (in collaboration with C.C. Byrne), *Introduction to linear programming,* Prentice Hall, 2003.

[Ros00] K. H. Rosen, ed., *Handbook of discrete and combinatorial mathematics,* CRC Press, 2000.