MAT3100 - Linear Optimization - Lecture 1

- Us:
  > Martin Reimers, Professor computational mathematics
  > Lorenzo Ciardo, PhD student comp. math.

- You? Language?

- Need a student representative ("Tillitsstudent")

- The course (http://www.uio.no/studier/emner/matnat/math/MAT3100):

  > 10 credits/studiepoeng

  > Curriculum:
    - http://www.uio.no/studier/emner/matnat/math/MAT3100/v19/pensumliste/index.html

  > About 16 lectures, Tuesdays 1015-1200
    - http://www.uio.no/studier/emner/matnat/math/MAT3100/v19/timeplan/index.html

  > Lecture notes (Geir Dahl): posted before AND after each lecture
    - Read curriculum/notes before lecture!

  > Weekly exercises/tutorials Wednesdays 1215-14 (no grades)
    - We expect you to do exercises before tutorial

  > 2 mandatory exercises (march and april). Both must be passed (no grades)

  > Written exam May/June

  > Learning outcomes:
    - formulate optimization problems
    - theory of Linear Programming, Duality Theory
    - applications, including resource allocation, game theory and optimization on networks
    - Express optimization problems in OPL-CPLEX and solve them.

LP. Lecture 1.
Chapter 1 and 2: example and the simplex algorithm

This course: introduction to linear optimization and related areas.

- what is LP (lin.opt.=lin.programming)
- more generally: mathematical optimization
- theory, methods, applications
- these notes originally made by Geir Dahl, based on R. Vanderbei: "Linear programming: fundations and extensions". Fourth edition (2014).

- a practical example: production planning
- simplex algorithm, example and some concepts
- simplex algorithm in general

What is linear optimization?

- Linear optimization is to maximize (or minimize) a linear function in several variables subject to constraints that are linear equations and linear inequalities.

Example: production planning

Products:

- door (with glass)
- window

Production facility:

- Factory 1: produces door frames
- Factory 2: produces window frames
- Factory 3: produces glass and mounts doors & windows

Data:

	Hours/series		Hours at disposal
	door	window	
Factory 1	1	0	4
Factory 2	0	2	12
Factory 3	3	2	18
Revenue/series	3	5	

Problem: How many series should be produced of each product in order to maximize the revenue?

The production plan as an LP problem:

decision variables

$$
\begin{array}{lrrl}
\text{maximize} & 3x_1 + & 5x_2 & \\
\text{subject to} & & & \\
& x_1 & & \leq 4 \\
& & 2x_2 & \leq 12 \\
& 3x_1 + & 2x_2 & \leq 18 \\
& \multicolumn{3}{c}{x_1 \geq 0,\ x_2 \geq 0.}
\end{array}
$$

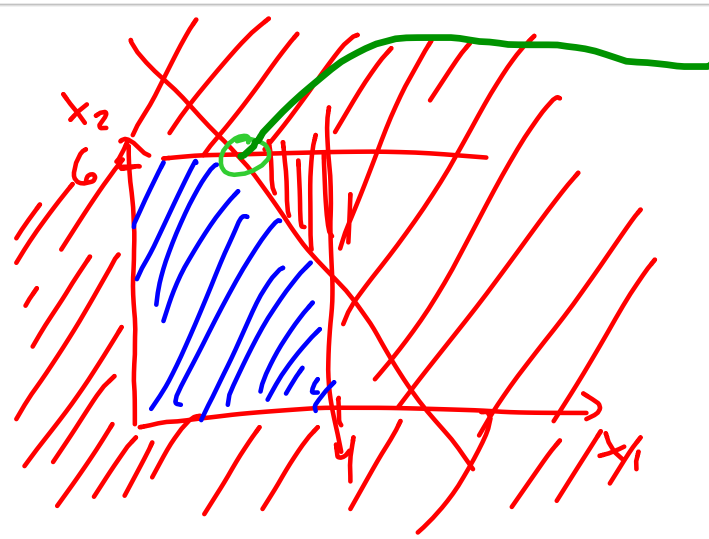
constraints

- We want to find an optimal solution, i.e., a vector (x_1, x_2) which maximize the objective function

$$f(x_1, x_2) = 3x_1 + 5x_2$$

and satisfies all the constraints, i.e. a feasible solution

- Exercise: find the optimal solution



Optimal solution

Feasible solutions

Later we will work with LP problems in matrix form;

$$
\begin{array}{ll}
\text{max} & c^T x \\
\text{subject to} & \\
& Ax \leq b \\
& x \geq O
\end{array}
$$

Standard form

Here c and x are column vectors with n components, A is a $m \times n$ matrix and b is a column vector with m components. O denotes the zero vector (of suitable length). The inequality $Ax \leq b$ is a vector inequality and means that $\leq$ holds componentwise (for every component).

Analysis of this problem and methods for solving it are based on linear algebra.

LP is closely tied to theory/methods for solving systems of linear inequalities. Such systems have the form

$$Hx \leq f$$

where H is a $m \times n$ matrix and f is a column vector (length m).

Example.

$$\begin{array}{rcrcr} 3x_1 & + & x_2 & \leq & 4 \\ x_1 & - & 2x_2 & \leq & 17 \\ -x_1 & & & \leq & 0 \end{array}$$

Central questions;

- existence of solution,
- how to find a solution, possibly *all* solutions.

Such problems may be written as LP problems: let the objective function have all its coefficients equal to 0.

More about linear inequalities later.

The simplex method

The simplex method is a general method for solving LP problems.

- Later we distinguish between the simplex method and the simplex algorithm, but this is not important now.
- The method was developed by George B. Dantzig around 1947 in connection with the investigation of transportation problems for the U.S. Air Force. Published in 1951.
- An interesting article in Washington Post in 2005 may be found on the course web page; Dantzig passed away in 2005.
- Other contributors to the development of linear programming were T.J.Koopmans and L.V.Kantorovich, who were awarded the Nobel prize in economics for this work in 1975.

The simplex algorithm, an example

Example: We want to solve

$$\begin{array}{llllll}
\max & 5x_1+ & 4x_2+ & 3x_3 & \\
\text{subject to} & & & & \\
\text{(i)} & 2x_1+ & 3x_2+ & x_3 & \leq 5 \\
\text{(ii)} & 4x_1+ & x_2+ & 2x_3 & \leq 11 \\
\text{(iii)} & 3x_1+ & 4x_2+ & 2x_3 & \leq 8 \\
& & x_1, x_2, x_3 \geq 0. & &
\end{array}$$

First, we convert to equations by introducing slack variables for every $\leq$-inequality: for instance (i) is replaced by

$$w_1 = 5 - 2x_1 - 3x_2 - x_3 \quad \text{and } w_1 \geq 0.$$

Can write the problem in the following form, called a dictionary:

$$\begin{array}{llcrcrcrcr}
\max & \eta & = & & & 5x_1 & + & 4x_2 & + & 3x_3 \\
\text{subject to} & & & & & & & & & \\
\text{(i)} & w_1 & = 5 & & - & 2x_1 & - & 3x_2 & - & x_3 \\
\text{(ii)} & w_2 & = 11 & & - & 4x_1 & - & x_2 & - & 2x_3 \\
\text{(iii)} & w_3 & = 8 & & - & 3x_1 & - & 4x_2 & - & 2x_3
\end{array}$$

$x_1, x_2, x_3, w_1, w_2, w_3 \geq 0$

- Left-hand side: dependent variables = basic variables.
- Right-hand side: independent variables = nonbasic variables.

Initial solution: Let $x_1 = x_2 = x_3 = 0$ and this gives $w_1 = 5, w_2 = 11, w_3 = 8$.

We always let the nonbasic variables be equal to zero. The basic variables are then *uniquely* determined; they become equal to the constants on the right-hand side.

Is this an optimal solution? No !!
For instance, we can increase x_1 while keeping $x_2 = x_3 = 0$. Then

- η (the value of the objective function) will increase
- we obtain new values for the basic variables; determined by x_1
- watch out: the w_j's may approach 0!

Maximum increase of x_1: want to avoid that the basic variables, one or more, become negative.

$$\begin{array}{lllllllllll}
 & & & & & & \overset{0}{\shortparallel} & & \overset{0}{\shortparallel} & & \\
\text{(i)} & w_1 & = 5 & - & 2x_1 & - & \cancel{3x_2} & - & \cancel{x_3} & & \boxed{x_1 \le \frac{5}{2}} \\
\text{(ii)} & w_2 & = 11 & - & 4x_1 & - & \cancel{x_2} & - & \cancel{2x_3} & & x_1 \le \frac{11}{4} \\
\text{(iii)} & w_3 & = 8 & - & 3x_1 & - & \cancel{4x_2} & - & \cancel{2x_3} & & x_1 \le \frac{8}{3}
\end{array}$$

This gives the new solution $x_1 = 5/2$, $x_2 = x_3 = 0$ and therefore $w_1 = 0$, $w_2 = 1$, $w_3 = 1/2$. And now $\eta = 25/2$.
Thus: a (much) better solution!!

How to proceed? The dictionary is not well suited for testing optimality, so we want to transform to a new dictionary.

- We want x_1 and w_1 to "switch sides". So: x_1 should go *into* the basis, while w_1 goes *out* of the basis. This can be done by using the w_1-equation in order to eliminate x_1 from all other equations.
- Equivalent: we may use elementary row operations on the system in order to eliminate x_1: (i) solve for x_1: $x_1 = 5/2 - (1/2)w_1 - (3/2)x_2 - (1/2)x_3$, and (ii) add a suitable multiple of this equation to the other equations so that x_1 disappears and is replaced by terms with w_1.

Remember: elementary row operations do not change the solution set of the linear system of equations.

From previous dictionary

$$
\begin{array}{rlrlrlrl}
\eta & = & & 5x_1 & + & 4x_2 & + & 3x_3 \\
\hline
w_1 & = 5 & - & 2x_1 & - & 3x_2 & - & x_3 \\
w_2 & = 11 & - & 4x_1 & - & x_2 & - & 2x_3 \\
w_3 & = 8 & - & 3x_1 & - & 4x_2 & - & 2x_3
\end{array}
$$

To new dictionary (homework!)

$$
\begin{array}{rlrlrlrl}
\eta & = 25/2 & - & 5/2w_1 & - & 7/2x_2 & + & 1/2x_3 \\
\hline
x_1 & = 5/2 & - & 1/2w_1 & - & 3/2x_2 & - & 1/2x_3 \\
w_2 & = 1 & + & 2w_1 & + & 5x_2 & & \\
w_3 & = 1/2 & + & 3/2w_1 & + & 1/2x_2 & - & 1/2x_3
\end{array}
$$

We have now performed a pivot: the use of elementary row operations (or elimination) to switch two variables (one into and one out of the basis).

Repeat the process!

Optimal? No: we can increase η by increasing x_3 from zero! May increase to $x_3 = 1$ for then $w_3 = 0$ (while the other basic variables are nonnegative).

Then we do another pivot: x_3 goes into the basis, and w_3 leaves the basis. This gives the new dictionary:

$$
\begin{array}{lllllll}
\eta & = 13 & - & w_1 & - & 3x_2 & - & w_3 \\
\hline
x_1 & = 2 & - & 2w_1 & - & 2x_2 & + & w_3 \\
w_2 & = 1 & + & 2w_1 & + & 5x_2 & & \\
x_3 & = 1 & + & 3w_1 & + & x_2 & - & 2w_3
\end{array}
$$

Here we see that all coefficients of the nonbasic variables are nonpositive (in fact negative) in the equation for η. Then every increase of one or more nonbasic variables will result in a soluton where $\eta \leq 13$.

But: *any* feasible solution is obtained by a suitable choice of the nonbasic variables! Why?

Conclusion: we have found an optimal solution! It is
$w_1 = x_2 = w_3 = 0$ and $x_1 = 2, w_2 = 1, x_3 = 1$.
The corresponding value of η is 13, and this is called the optimal value.

The simplex method, in general

Consider a general LP problem

Standard form

$$\begin{array}{lll} \max & \sum_{j=1}^{n} c_j x_j & \\ \text{subject to} & & \\ & \sum_{j=1}^{n} a_{ij} x_j \leq b_i & \text{for } i = 1, \ldots, m \\ & x_j \geq 0 & \text{for } j = 1, \ldots, n. \end{array}$$

where we (now) assume that $b_i \geq 0$ for all $i \leq m$.

Introduce slack variables

$$\begin{aligned} \eta &= \textstyle\sum_{j=1}^{n} c_j x_j \\ w_i &= b_i - \textstyle\sum_{j=1}^{n} a_{ij} x_j \quad \text{for } i = 1, \ldots, m \end{aligned}$$

We do not need to distinguish between slack variables and the original variabels so we get the following dictionary:

$$\begin{aligned} \eta &= \textstyle\sum_{j=1}^{n} c_j x_j \\ x_{n+i} &= b_i - \textstyle\sum_{j=1}^{n} a_{ij} x_j \quad \text{for } i = 1, \ldots, m \end{aligned}$$

Basic

Non - basic

where we have replaced w_i by x_{n+i}. So $x \in \mathbb{R}^{n+m}$.

The algorithm starts with this dictionary where $x_{n+1}, \ldots, x_{n+m}$ are basic variables and $x_1, \ldots, x_n$ are nonbasic variables. Let

- B be the index set of the basic variables.
- N be the index set of the nonbasic variables.

So, initially $B = \{n+1, \ldots, n+m\}$, $N = \{1, \ldots, n\}$. The initial solution is

$$\begin{aligned} x_j &= 0 \quad \text{for } j = 1, \ldots, n \\ x_{n+i} &= b_i \quad \text{for } i = 1, \ldots, m \end{aligned}$$

and the corresponding value of η is $\eta = 0$. Such a solution is called a basic solution.

Every iteration is a pivot (or change of basis) where

- one index k is moved from N to B (x_k is entering (ingoing) variable; it is a new basic variable because it results in an increase of η),
- another index l is moved from B to N (x_l is leaving (outgoing) variable; this variable leaves the basis because it becomes 0 as the first one, and
- we find the new dictionary from the old by performing row operations (or elimination)
- the basic solution that corresponds to the new dictionary is feasible.

At the start of every pivot we have the dictionary (with $\bar{b}_i \geq 0$):

$$\begin{aligned} \eta &= \bar{\eta} + \sum_{j \in N} \bar{c}_j x_j \\ x_i &= \bar{b}_i - \sum_{j \in N} \bar{a}_{ij} x_j \quad \text{for } i \in B. \end{aligned}$$

Selection of entering variable: choose a $k \in N$ with $\bar{c}_k > 0$. If no such index exists, the current solution is optimal and we terminate. Often several $\bar{c}_j$'s are positive. There are several principles for selection of the entering variable, but a simple, and often used, principle is to choose $k = j$ with $\bar{c}_j$ largest possible. Why?

Selection of leaving variable: Also here we may have several choices. First we have to determine the maximum increase of the entering variable x_k. From

$$x_i = b_i - \bar{a}_{ik} x_k \quad \text{for } i \in B$$

we see that

- if $\bar{a}_{ik} \leq 0$, x_i will increase when x_k is increased. Such basic variables will not become zero when x_k is increased (we assume now that $\bar{b}_i > 0$)
- if, however, $\bar{a}_{ik} > 0$, then x_i will decrease and it becomes zero when

$$x_k = b_i / \bar{a}_{ik}.$$

Selection of leaving variable, cont.

So, we can increase x_k to the value

$$\theta := \min\{b_i/\bar{a}_{ik} : \bar{a}_{ik} > 0\}.$$

What happens when $x_k = \theta$?

Well, all variables are still nonnegative. Good! And at least one basic variable has become zero, in fact $x_i = 0$ for all $i \in B$ satisfying

$$b_i/\bar{a}_{ik} = \theta.$$

Conclusion: Leaving variable x_l is selected so that

$$b_l/\bar{a}_{l,k} = \min\{b_i/\bar{a}_{ik} : \bar{a}_{ik} > 0\}.$$

Pivot rule: a rule which tells us which entering variable to choose and which leaving variable to choose.

Several pivot rules are around, so we get several variants of the simplex algorithm.

The pivot is terminated by the row operations:
assume x_k is entering variable and x_l is leaving variable. Then x_l is on the left-hand side in "equation number l":

$$x_l = \bar{b}_l - \sum_{j \in N} \bar{a}_{lj} x_j$$

For every $i \neq l$ we add $\bar{a}_{ik}/\bar{a}_{lk}$ times equation l to equation i.

Furthermore, we use equation l to solve for x_k and therefore express x_k as a function of the other variables.

Result: an equivalent system of equations is obtained (e.g. the same solutions) with coefficient 0 associated with each x_k in every equation $i \neq l$. Further, the new basis variable x_k is on the left-hand side of the equations! This is the new dictionary.

Comments

Some questions remain:

- how to find an initial basic feasible solution if there are negative b_i's?
- in a dictionary: what happens if some $\bar{b}_i$'s are 0? Does this cause problems for the pivot rule?
- will the algorithm terminate?
- and, if not, can we repair this somehow?

We will start working on these questions in Lecture 2!

Finally, let us consider an important application of LP in connection with linear l_1-approximation.

Application: linear approximation

Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$, and let a_i be the i'th row in A considered as a column vector. Recall that the l_1-norm of a vector $y \in \mathbb{R}^n$ is $\|y\|_1 = \sum_{i=1}^n |y_i|$.
The linear approximation problem

$$\min\{\|Ax - b\|_1 : x \in \mathbb{R}^n\}$$

may be solved as the following LP problem

$$\begin{array}{lrll} \text{min} & \sum_{i=1}^m z_i & & \\ \text{subject to} & & & \\ & a_i^T x - b_i & \leq z_i & (i \leq m) \\ & -(a_i^T x - b_i) & \leq z_i & (i \leq m) \end{array}$$

Proof: Because every optimal solution in this LP satisfies $z_i = |a_i^T x - b_i|$ for every $i \leq m$. $\square$

This means that one has an alternative method to the traditional least squares method based on solving $\min\{\|Ax - b\|_2 : x \in \mathbb{R}^n\}$ and this problem has lots of important applications (see any linear algebra textbook).

A similar LP approach works for the linear approximation problem

$$\min\{\|Ax - b\|_\infty : x \in \mathbb{R}^n\}$$

in the ℓ_∞-norm given by $\|z\|_\infty = \max_i |z_i|$.