MDI210 : Numerical Analysis and Continuous Optimization

Robert M. Gower
Core Info

- **Where**: Telecom ParisTech
- **Location**: C 48
- **Volume**: 28h
- **When**: 8 weeks
- **Exam**: One exam on 31st of October
- **Online**: Find lecture notes on my homepage
  
  http://www.di.ens.fr/~rgower/teaching.html
- **Exercises**: Do all exercises in the MDI210 lecture notes
Additional References for Numerical Analysis

Matrix Computations: Gene H. Golub and Charles F. Van Loan

Three copies in the library on the 8th floor!
Linear Programming History (1939)

- Assignment 70 people to 70 jobs (more possibilities than particles).
Linear Programming History (1941)
Army Builds Killing Machine (1949)

1949 SCOOP: Scientific Programming Of Optimal Programs

Mathematical Programming: Math used to figured out Flight and logistic programs/schedules
Dantzig, George B. "On the Non-Existence of Tests of 'Student's' Hypothesis Having Power Functions Independent of Sigma." Annals of Mathematical Statistics. No. 11; 1940 (pp. 186-192).

Optimization and Numerical Analysis: Linear Programming

Robert Gower

September 26, 2017
Table of Contents

The Fundamental Theorem

Simple 2D problem

Notation

Simplex Algorithm

Duality
The Problem: Linear Programming

\[
\max_x z \overset{\text{def}}{=} c^\top x
\]

subject to \( Ax \leq b, \)
\[ x \geq 0, \]

where \( c, x \in \mathbb{R}^n, A \in \mathbb{R}^{m \times n}, \) and \( b \in \mathbb{R}^m. \) Equivalently

\[
\max_x z \overset{\text{def}}{=} \sum_{j=1}^n c_j x_j
\]

subject to \( \sum_{j=1}^n a_{ij} x_j \leq b_i, \) for \( i = 1, \ldots, m. \)
\[ x \geq 0. \]
Theorem (Fundamental Theorem of Linear Programming)

Let $P = \{x \mid Ax = b, x \geq 0\}$ then either

1. $P = \{\emptyset\}$

2. $P \neq \{\emptyset\}$ and there exists a vertex $v$ of $P$ such that $v \in \arg\min_{x \in P} c^\top x$

3. There exists $x, d \in \mathbb{R}^n$ such that $x + td \in P$ for all $t \geq 0$ and $\lim_{t \to \infty} c^\top (x + td) = \infty$. 
First example Simplex

The problem

\[
\text{max} \quad 4x_1 + 2x_2 \\
3x_1 + 2x_2 \quad \leq 600 \\
4x_1 + 1x_2 \quad \leq 400 \\
x_1 \geq 0, \quad x_2 \geq 0.
\]

Can be transformed into

\[
\text{max} \quad 4x_1 + 2x_2 \\
x_3 = 600 - 3x_1 - 2x_2 \\
x_4 = 400 - 4x_1 - x_2,
\]

where \(x_3\) and \(x_4\) are slack variables. This is known as the dictionary format and is often written as:

\[
x_3 = 600 - 3x_1 - 2x_2 \\
x_4 = 400 - 4x_1 - x_2 \\
z = 4x_1 + 2x_2
\]
Simple 2D problem

First example Simplex

The dictionary format

\[
\begin{align*}
x_3 &= 600 - 3x_1 - 2x_2 \\
x_4 &= 400 - 4x_1 - x_2 \\
z &= 4x_1 + 2x_2
\end{align*}
\]

admits obvious solution

\[(x_1^*, x_2^*, x_3^*, x_4^*) = (0, 0, 600, 400)\].

The objective \(z\) will improve if \(x_1 > 0\). Increasing \(x_1\) as much as possible

\[
\begin{align*}
x_3 &\geq 0 \implies 600 - 3x_1 \geq 0 \implies x_1 \leq 200, \\
x_4 &\geq 0 \implies 400 - 4x_1 \geq 0 \implies x_1 \leq 100.
\end{align*}
\]

Thus \(x_1 \leq 100\) to guarantee \(x_4 \geq 0\). This means \(x_4\) will leave the basis and \(x_1\) will enter the basis. Using row operations to isolate \(x_1\) on row 2.

\[
\begin{align*}
x_3 &= 300 - 0 - \frac{5}{4}x_2 \\
x_1 &= 100 - \frac{x_4}{4} - \frac{x_2}{4} \\
z &= 400 - x_4 + x_2
\end{align*}
\]
First example Simplex

From

\[
\begin{align*}
    x_3 &= 300 & 0 & - \frac{5}{4}x_2 \\
    x_1 &= 100 & - \frac{x_4}{4} & - \frac{x_2}{4} \\
    z &= 400 & - x_4 & + x_2
\end{align*}
\]

Now we are at the vertex \((x_1^*, x_2^*) = (100, 0)\). Next we see that increasing \(x_2\) increases the objective value but

\[
\begin{align*}
    x_3 &\geq 0 \Rightarrow 240 \geq x_2, \\
    x_1 &\geq 0 \Rightarrow 400 \geq x_4.
\end{align*}
\]

Increase \(x_2\) upto 240 while respecting the positivity constraints of \(x_3\). Thus \(x_3\) will leave the basis and \(x_2\) will enter the basis. Performing a row elimination again, we have that

\[
\begin{align*}
    x_2 &= 240 & 0 & - \frac{4}{5}x_3 \\
    x_1 &= 40 & - \frac{x_4}{4} & - \frac{1}{5}x_3 \\
    z &= 640 & - x_4 & - \frac{4}{5}x_3
\end{align*}
\]

Now \((x_1^*, x_2^*) = (40, 240)\). Increasing \(x_4\) or \(x_3\) will decrease \(z\).
Problem Notation

We will now formalize the definitions we introduced in the examples.

- There are $n$ variables and $m$ constraints
- The linear objective function $z = \sum_{j=1}^{n} c_j x_j$
- The $m$ inequality constraints in standard form
  \[ \sum_{j=1}^{n} a_{ij} x_j \leq b_i, \text{ for } i \in \{1, \ldots, m\}. \]
- The $n$ positivity constraints $x_j \geq 0$, for $j \in \{1, \ldots, n\}$.
- $x_i^*$ denotes the value of $i$th variable.
- We call $(x_1^*, \ldots, x_n^*) \in \mathbb{R}^n$ a feasible solution if it satisfies the inequality and positivity constraints.
Dictionary Notation

- The slack variables \((x_{n+1}, \ldots, x_{n+m}) \in \mathbb{R}^m\) (variables d’écart)
- The initial dictionary

\[
\begin{align*}
x_{n+1} &= b_1 - \sum_{j=1}^{n} a_{1j}x_j \\
&\vdots \\
x_{n+i} &= b_i - \sum_{j=1}^{n} a_{ij}x_j \\
&\vdots \\
x_{n+m} &= b_m - \sum_{j=1}^{n} a_{mj}x_j \\
\end{align*}
\]

\[
z = \sum_{j=1}^{n} c_jx_j,
\]

- Valid dictionary if \(m\) of the variables \((x_1, \ldots, x_{n+m})\) can be expressed as function of the remaining \(n\) variables.

- The \(m\) variables on the left-hand side are the basic variable (variable de base). The \(n\) variables on the right-hand side are the non-basic (variable hors-base).
Dictionary Notation

After row elimination operations we have a new basis.

- Basic variable set $I \subset \{1, \ldots, n + m\}$ and non-basic set $J = \{1, \ldots, n + m\} \setminus I$

- Current objective value $z^* = \sum_{j=1}^{n} c_j x_j^*$.

- For each basis set $I$ there is a corresponding dictionary

\[
\begin{align*}
    x_i &= b'_i - \sum_{j \in J} a'_{ij} x_j, \text{ for } i \in I \\
    z &= z^* + \sum_{j \in J} c'_j x_j,
\end{align*}
\]

where $a'_{ij}, b'_i, z^* \in \mathbb{R}$ are coefficients resulting from the row operations. For this to a feasible dictionary we require that $b'_i \geq 0$.

- A basic solution: $x_i^* = b'_i$ for $i \in I$ and $x_j^* = 0$ for $j \in J$. 
A Step of the Simplex Method

**Input:** A basic index set $I \subset \{1, \ldots, n + m\}$, $J = I \setminus \{1, \ldots, n + m\}$, constraint coefficients $a'_{ij} \in \mathbb{R}$, $b'_i \geq 0$ and $c'_i \in \mathbb{R}$.

if $c_i \leq 0$ for all $i \in J$ then
   STOP;  # Optimal point found.
Choose a variable $j_0$ to **enter the basis** from the set $j_0 \in \{j \in J : c'_j > 0\}$.

if $a'_{ij} \geq 0$ for all $i \in J$ then
   STOP;  # The problem is unbounded.
Choose a variable $i_0$ to **leave the basis** from the set $i_0 \in \arg\min_{i \in I, a'_{i_0} > 0} \left\{ \frac{b'_i}{a'_{i_0}} \right\}$.

for $i \in I$ do
   \[ a'_i \leftarrow a'_i - \frac{a'_{i_0}}{a'_{i_0 j_0}} a'_{i_0} \]  # Row elimination on pivot $(i_0, j_0)$.
\[ c' \leftarrow c' - \frac{c'_{j_0}}{a'_{i_0 j_0}} a'_{i_0} \]  # Update the cost coefficients.
\[ I \leftarrow (I \setminus \{i_0\}) \cup \{j_0\} \]  # Update basis.

**Output:** $I, a'_{ij}, b'_j, c'_j$. 

How to choose who enters the basis?

1. The mad hatter rule: Choose the first one you see.

2. Dantzig’s 1st rule: $j_0 = \arg \max_{j \in J} c_j$.

3. Dantzig’s 2nd rule: Choose $j_0 \in \{j \in J : c_j > 0\}$ that so that maximizes increase in $z$.

   \[
   j_0 = \arg \max_{j \in J} \left\{ c_j \min_{i : a_{ij} > 0} \left\{ \frac{b_i}{a_{ij}} \right\} \right\} .
   \]

   Effective, but computationally expensive.

4. Bland’s rule: Choose the smallest indices $j_0$ and $i_0$. That is, choose

   \[
   j_0 = \arg \min\{j \in J : c_j > 0\}.
   \]

   \[
   i_0 = \min \left\{ \arg \min_{i : a_{ij_0} > 0} \left\{ \frac{b_i}{a_{ij_0}} \right\} \right\} .
   \]
Degeneracy

If any of the basic variables are zero, then we say that the solution is degenerate. Example on Board and in other lecture slides.
Upper Bounds Using Duality

The LP in standard form

\[
\max_x z \overset{\text{def}}{=} c^\top x \\
\text{subject to } Ax \leq b, \\
x \geq 0,
\]

(LP)

We want to find \( w \in \mathbb{R} \) so that \( z = c^\top x \leq w \) for all \( x \in \mathbb{R}^n \).

Combine rows of constraints?
Look for \( y \geq 0 \in \mathbb{R}^m \) so that \( y^\top A \approx c^\top \), consequently

\[
c^\top x \approx (y^\top A)x \leq y^\top b = w.
\]

Precisely, let \( y \geq 0 \in \mathbb{R}^m \) be such that \( y^\top A \geq c^\top \) or equivalently \( A^\top y \geq c \). Then

\[
c^\top x \leq (y^\top A)x \leq y^\top b.
\]

Can we make this upper bound as tight as possible? Yes, by minimizing \( y^\top b \). That is, we need to the dual linear program.
Duality

**Dual definition**

The LP in standard form

\[
\max_x z \overset{\text{def}}{=} c^\top x
\]

subject to \(Ax \leq b,\)

\(x \geq 0,\)

(LP)

The dual LP:

\[
\max_w w \overset{\text{def}}{=} y^\top b
\]

subject to \(A^\top y \geq c,\)

\(y \geq 0.\)

(DP)

**Lemma (Weak Duality)**

If \(x \in \mathbb{R}^n\) is a feasible point for (LP) and \(y \in \mathbb{R}^m\) is a feasible point for (DP) then

\[
c^\top x \leq y^\top Ax \leq y^\top b.\tag{1}
\]
Duality

Weak Duality

Lemma (Weak Duality)

If \( x \in \mathbb{R}^n \) is a feasible point for (LP) and \( y \in \mathbb{R}^m \) is a feasible point for (DP) then

\[
  c^\top x \leq y^\top A x \leq y^\top b. \tag{2}
\]

Consequently

- If (LP) has an unbounded solution, that is \( c^\top x \to \infty \), then there exists no feasible point \( y \) for (DP)
- If (DP) has an unbounded solution, that is \( y^\top b \to -\infty \), then there exists no feasible point \( x \) for (LP)
- If \( x \) and \( y \) are primal and dual feasible, respectively, and \( c^\top x = y^\top b \), then \( x \) and \( y \) are the primal and dual optimal points, respectively.
Strong Duality

Theorem (Strong Duality)

If (LP) or (DP) is feasible, then \( z^* = w^* \). Moreover, if \( c^* \) is the cost vector of the optimal dictionary of the primal problem (LP), that is, if

\[
z = z^* + \sum_{i=1}^{n+m} c_i^* x_i, \tag{3}
\]

then \( y_i^* = -c_{n+i}^* \) for \( i = 1, \ldots, m \).

First \( c_i^* \leq 0 \) for \( i = 1, \ldots, m + n \) because dictionary is optimal. Consequently \( y_i^* = -c_{n+i}^* \geq 0 \) for \( i = 1, \ldots, m \).
Duality

Strong duality: Proof I

By the definition of the slack variables we have that

$$x_{n+i} = b_i - \sum_{j=1}^{n} a_{ij}x_j, \quad \text{for } i = 1, \ldots, m.$$  \hfill (4)

Consequently, setting $y_i^* = -c_{n+i}^*$, we have that

$$z = z^* + \sum_{j=1}^{n} c_j^* x_j + \sum_{i=n+1}^{n+m} c_i^* x_i$$

$$= z^* + \sum_{j=1}^{n} c_j^* x_j - \sum_{i=1}^{m} y_i^*(b_i - \sum_{j=1}^{n} a_{ij}x_j)$$

$$= z^* - \sum_{i=1}^{m} y_i^* b_i + \sum_{j=1}^{n} \left(c_j^* + \sum_{i=1}^{m} y_i^* a_{ij}\right)x_j$$

$$= \sum_{j=1}^{n} c_jx_j, \quad \forall x_1, \ldots, x_n.$$ \hfill (5)

Last line followed by definition $z = \sum_{j=1}^{n} c_jx_j$. Since the above holds for all $x \in \mathbb{R}^n$, we can match the coefficients.
Strong duality: Proof II

Matching coefficients on $x_j$’s we have

$$z^* = \sum_{i=1}^{m} y_i^* b_i$$  \hspace{1cm} (6)

$$c_j = c_j^* + \sum_{i=1}^{m} y_i^* a_{ij}, \quad \text{for } j = 1, \ldots, n.$$  \hspace{1cm} (7)

Since $c_j^* \leq 0$ for $j = 1, \ldots, n$, the above is equivalent to

$$z^* = \sum_{i=1}^{m} y_i^* b_i$$  \hspace{1cm} (8)

$$\sum_{i=1}^{m} y_i^* a_{ij} \leq c_j, \quad \text{for } j = 1, \ldots, n.$$  \hspace{1cm} (9)

The inequalities (9) prove that $y_i^*$’s satisfies the constraints in (DP), and thus is feasible. The equality (8) shows that $z^* = \sum_{i=1}^{m} y_i^* b_i = w$, a consequently by week duality the $y_i^*$’s are dual optimal. □
G., R & P Richtárik, Randomized Iterative Methods for Linear Systems arXiv:1506.03296