IB Optimisation: Lecture 6

Mike Tehranchi

University of Cambridge

6 May 2020

A linear program can have

- Functional constraint: equality or inequality or a mixture
- Regional constraint: fixed sign or no constraint or a mixture

For formulationg nice looking theorems on duality, we have seen that it is helpful to pose the linear program with an inequality functional constraint and a fixed sign regional constraint.
But for discussing the simplex method, another formulation is preferred....

Definition

A linear program is in *standard form* if the constraint can be written

$$Ax = b, \; x \geq 0$$

All linear programs can be put in standard form. For instance

- $Ax \le b, x \ \ge 0$ becomes $Ax + z = b, \quad x, z \ge 0$.
- $Ax = b$ becomes $A(x - y) = b, \quad x, y \ge 0$.
- $Ax \le b$ becomes $A(x - y) + z, \quad x, y, z \ge 0$.

Definition

Given an $m \times n$ matrix A with $n > m$, and $b \in \mathbb{R}^m$. A solution $x \in \mathbb{R}^n$ of the equation $Ax = b$ is called *basic* if if at most m entries of x are non-zero, that is, $x_i \neq 0$ for at most m indices $i \in \{1, \ldots, n\}$.

If x is a basic solution and $x \geq 0$, then x is called a *basic feasible solution*, abbreviated *b.f.s.*

Now fix A and b as above, and let $C = \{x \in \mathbb{R}^n : Ax = b, x \geq 0\}$.

Theorem (Extreme points are b.f.s.)

Suppose x is an extreme point of C. Then x is a b.f.s.

Proof. Let x be a point in C that is not basic, i.e. at least $m+1$ indices i are such that $x_i > 0$. Let $x_{i_k} > 0$ for indices $i_1, \ldots, i_r$. The set $\{A_{i_1}, \ldots, A_{i_r}\}$ of columns of the matrix A is linearly dependent, since $r > m$. Hence, there exists a non-trivial linear combination equal to zero

$$w_1 A_{i_1} + \ldots + w_r A_{i_r} = 0.$$

Construct a vector $z \in \mathbb{R}^n$ by

$$z_i = \begin{cases} w_k & \text{if } i = i_k, \text{ for } k = 1, \ldots, r \\ 0 & \text{otherwise.} \end{cases}$$

Then by construction

$$Az = A_1 z_1 + \ldots A_n z_n = w_1 A_{i_1} + \ldots + w_r A_{i_r} = 0$$

Hence $A(x \pm \varepsilon z) = Ax = b$ for any ε. Now, choose $\varepsilon > 0$ small enough that $x \pm \varepsilon z \geq 0$. For such ε we have $x \pm \varepsilon z \in C$. But since

$$x = \tfrac{1}{2}(x + \varepsilon z) + \tfrac{1}{2}(x - \varepsilon z)$$

the point x is not extreme. □

Theorem (b.f.s. are extreme points)

Suppose that every set of m columns of A is linearly independent. Let x be a b.f.s. Then x is an extreme point of C.

Proof. By definition, there are at most m indices i are such that $x_i > 0$. Suppose $x = py + (1-p)z$ for $y, z \in C$ and $0 < p < 1$. Since $y \geq 0$ and $z \geq 0$ we conclude that if $x_i = 0$ for some index i then $y_i = z_i = 0$. Hence, there are at least $n - m$ indices i such that $y_i = z_i = 0$.

The equation $Ay = b = Az$ implies

$$A(y - z) = 0 = \sum_i w_i A_i$$

where $w = y - z$. Since at most m entries of w are non-zero, and any set of m columns of A is linearly independent, we have $w = 0$ or equivalently $x = y = z$. □

In summary, we can find the optimal solution of a linear program by checking each of the basic feasible solutions. Since there are an infinite number of feasible solutions, but a finite number of basic feasible solutions, we have made a lot of progress!

Consider a linear program in standard form

$$P: \text{ maximise } c^\top x \text{ subject to } Ax = b, \; x \geq 0.$$

where $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ and A is a $m \times n$ matrix.

- Assume assume that $n > m$ and that every set of m columns of A is linearly independent.
- Assume *non-degeneracy*: each b.f.s. x is such that there are *exactly* m indices i such that $x_i \neq 0$.

Our goal is to find a mechanical method of finding the optimal solution of P. We will do this by introducing methods of increasing efficiency.

Method 1. (somewhat naive) We know that if a convex function f takes its maximum on a convex set C, it is maximised at one of the extreme points of C. In our case the function $f(x) = c^\top x$ and the set $C = \{x : Ax = b, x \geq 0\}$ are convex, so it is enough to consider the extreme points of C.

But we also know that the extreme points of the set of a feasible solutions of a linear program are precisely the basic feasible solutions.

- Fix a set $B \subset \{1, \ldots, n\}$ with cardinality $|B| = m$,
- let $N = \{1, \ldots, n\} \setminus B$ with cardinality $|N| = n - m$.
- If $B = \{i_1, \ldots, i_m\}$, we will let

$$A_B = (A_{i_1} \ \ldots \ A_{i_m})$$

be the $m \times m$ matrix formed by taking the columns of A indexed by $i \in B$.
- Note by assumption that A_B is invertible.
- For a point $x \in \mathbb{R}^n$, let

$$x_B = (x_{i_1} \ \ldots \ x_{i_m})^\top.$$

- Define the notation c_B, A_N, x_N, and c_N similarly.

We first aim to find a basic solution of $Ax = b$.

- This equation is $A_B x_B + A_N x_N = b$.
- Setting $x_N = 0$ yields $x_B = A_B^{-1} b$. S
- Rearranging the coordinates if necessary, we may write the basic point as $x = \binom{x_B}{x_N} = \binom{A_B^{-1} b}{0}$.

- Check that this x is feasible. That is, we need $A_B^{-1} b \geq 0$.
- Assuming that x is a basic feasible solution, compute the objective function $c^\top x = c_B^\top x_B$.
- Repeat this procedure for all $\binom{n}{m}$ possible choices of the set B.
- If the problem has a maximum value, it will be among the list of b.f.s. just computed.

The problems

- When n is reasonably large, we need to look at an extremely large number $\binom{n}{m}$ of candidate solutions.
- This method does not detect the possibility that no maximiser exists, i.e. the problem is unbounded

Method 2. (a little better) We take the good idea of computing basic feasible solutions introduced above, but now try to avoid testing all of the large number of candidate solutions.

- The idea now is to use the fundamental theorem of linear programming.
- (example sheet) The dual problem is

$$D: \text{ minimise } b^\top \lambda \text{ subject to } A^\top \lambda \geq c.$$

- The fundamental theorem tells us that a point $x^* \in \mathbb{R}^n$ is optimal for the primal problem P *if and only if* there exists a Lagrange multiplier $\lambda^* \in \mathbb{R}^m$ such that
  1. $Ax^* = b,\ x^* \geq 0$ (primal feasibility)
  2. $A^\top \lambda^* \geq c$ (dual feasiblity)
  3. $(c - A^\top \lambda^*)^\top x^* = 0$ (complementary slackness)

Consider the b.f.s. $x = \binom{A_B^{-1} b}{0}$.

- Associate to this b.f.s. a Lagrange multiplier λ by complementary slackness

$$
\begin{aligned}
0 &= (c - A^\top \lambda)^\top x \\
&= (c_B - A_B^\top \lambda)^\top x_B + (c_N - A_N^\top \lambda)^\top x_N \\
&= (c_B - A_B^\top \lambda)^\top x_B .
\end{aligned}
$$

- So we take $\lambda = (A_B^\top)^{-1} c_B$.

A method

- Check that the b.f.s. $x = \binom{A_B^{-1} b}{0}$ (primal feasibility)
- The multiplier $\lambda = (A_B^\top)^{-1} c_B$ satisfies complementary slackness by construction.
- If $A^\top \lambda \geq c$ then STOP! since x would be optimal.
- Otherwise, choose another basis set $B \subset \{1, \ldots, n\}$ and start again.

This is better than the first method since it is now not always necessary to search through all of the possible basic solutions of the problem. Of course, we might be unlucky and have to search through all of the other solutions until we find one with a corresponding Lagrange multiplier λ satisfying dual feasibility.

Method 3 (The simplex algorithm) Recapping, given a b.f.s. we found a way of determining whether or not is is optimal. Assuming that our b.f.s. is not optimal, the remaining challenge is to find a clever way of another b.f.s. In particular, we would like to find a new b.f.s. that increases the objective function.

- Start with an initial b.f.s. x_0 corresponding to a basis B_0. Let $N_0 = \{1, \dots, n\} \backslash B_0$.
- Associate a Lagrange multiplier λ_0 by complementary slackness $(c - A^\top \lambda_0)_i = 0$ for all $i \in B_0$. In particular, $(c - A^\top \lambda_0)^\top x_0 = 0 \Rightarrow c^\top x_0 = b^\top \lambda_0$.
- Let $\mu_0 = c - A^\top \lambda_0$. If $\mu_0 \leq 0$ then we are done and x_0 is optimal. Otherwise continue.

Claim. Suppose that

- There is a $j \in N_0$ such that $\mu_{j,0} > 0$.
- There is an $i \in B_0$ such that there is a basic feasible solution x_1 with basis

$$B_1 = B_0 \cup \{j\} \setminus \{i\}.$$

Then

$$\begin{aligned} c^\top x_1 &= c^\top x_0 + \mu_{j,0} x_{j,1} \\ &> c^\top x_0. \end{aligned}$$

Proof of claim.

- If x is any feasible solution then

$$\begin{aligned} c^\top x &= c^\top x + \lambda_0^\top (b - Ax) \\ &= b^\top \lambda_0 + (c - A^\top \lambda_0)^\top x \\ &= c^\top x_0 + \sum_{k \in N_0} \mu_{k,0} x_k \end{aligned}$$

- We have summed only over N_0 since $\mu_{k,0} = 0$ for $k \in B_0$ by construction.

- The new basis B_1 is formed from B_0 by replacing previously basic index i and with previously non-basic index j.
- The new set of non-basic indices is

$$N_1 = N_0 \cup \{i\} \backslash \{j\}.$$

-

$$\sum_{k \in N_0} \mu_{k,0} x_{k,1} = \sum_{k \in N_1} \mu_{k,0} \underbrace{x_{k,1}}_{=0} + \underbrace{\mu_{j,0}}_{>0} \underbrace{x_{j,1}}_{>0} - \underbrace{\mu_{i,0}}_{=0} x_{i,1}$$

$$> 0$$

- We have used the assumption of non-degeneracy to assert $x_{j,1} > 0$.

□

Claim. Suppose that

- There is a $j \in N_0$ such that $\mu_{j,0} > 0$.
- For every $i \in B_0$ the basic solution x_1 with basis

$$B_1 = B_0 \cup \{j\} \backslash \{i\}.$$

  is *infeasible* (that is, $x_{k,1} < 0$ for some $k \in B_1$.)

Then the problem is unbounded.

Proof. Next time

If no x_1 feasible exists, then STOP! the problem is unbounded. Otherwise, test the optimality of x_1 and repeat if necessary. Stop when we come to a b.f.s. whose corresponding Lagrange multiplier satisfies dual feasibility (or say the problem is unbounded if no such b.f.s. exists)

We may be unlucky and have to search through all possible b.f.s. However, by choosing the sequence of b.f.s. as described here, we know at least that we are always marching up hill – every iteration of the algorithm strictly increases the value of the objective function.