Introduction to Linear Programming

Thomas Wiemann
University of Chicago

TA Discussion # 4
Econ 31740

February 7, 2022

Recap (Contd.)

We've seen last week that seemingly non-linear problems can be characterized as a linear problem. For instance, piece-wise linear convex (or concave) functions are permissible.

For example, consider the piecewise linear objective function $f(x) = \min_i g_i(x)$ where

$$g_i(x) = \begin{cases} 5 - x, & \text{if } i = 1, \\ 2 - 0.2x, & \text{if } i = 2, \\ -9 + x, & \text{if } i = 3. \end{cases}$$

Notice that for a given x, $f(x)$ is equivalent to the minimum scalar z such that $z \geq g_i(x), \forall i$. We can thus introduce an additional variable z and bound the objective function $z \geq g_i(x), \forall i$ as an additional set of constraints.

Nested problems like $\min_{x \geq 0} \left(\min_i \{d_i + c_i^\top x\} \right) \text{ s.t. } Ax \geq b$ can therefore be restated as linear problems.

Recap (Contd.)

We've also seen last week how general linear programs can be converted into standard form.

For example,

$$
\begin{aligned}
\min_{x} \quad & c'x \\
\text{s.t.} \quad & Ax = b, \\
& Dx \geq e, \\
& x \in \mathbb{R}^n,
\end{aligned}
\qquad \equiv \qquad
\begin{aligned}
\min_{\tilde{x}} \quad & \tilde{c}'\tilde{x} \\
\text{s.t.} \quad & \tilde{A}\tilde{x} = \tilde{b}, \\
& \tilde{x} \geq \mathbf{0}_n.
\end{aligned}
$$

where we have defined

$$
\tilde{A} := \begin{bmatrix} A & -A & \mathbf{0} \\ D & -D & -\mathbf{I} \end{bmatrix}, \quad \tilde{b} := \begin{bmatrix} b \\ e \end{bmatrix}, \quad \tilde{x} := \begin{bmatrix} x_+ \\ x_- \\ s \end{bmatrix}, \quad \text{and} \quad \tilde{c} := \begin{bmatrix} c \\ -c \\ 0 \end{bmatrix}.
$$

Outline*

Why do we care about standard form? Because its optimal solution – if it exists – takes the form of a *basic feasible solution*.

Today's discussion aims to illustrate the notion of basic feasible solutions using the geometry of linear programming. This is important because the simplex algorithm and it's adaptations are based on this key concept.

Key topics of today:

1. Geometry of Linear Programming
   a) Polyhedra and convex sets
   b) Extreme points, vertices, and basic feasible solutions
   c) Existence and optimality of extreme points
2. Brief Review of Duality Theory
3. An Application to Quantile Regression

*This discussion is built on the TA material of Joshua Shea and Chapter 2 of Bertsimas and Tsitsiklis (1997).

To streamline the discussion, we'll need some definitions. All definitions, theorems, and corollaries are taken from Bertsimas and Tsitsiklis (1997).

Definition 4.1

A set $S \subset \mathbb{R}^n$ is *convex* if

$$\lambda x + (1 - \lambda) y \in S, \quad \forall x, y \in S, \lambda \in [0, 1].$$

Definition 4.2 (Polyhedron)

A *polyhedron* is a set that can be described in the form

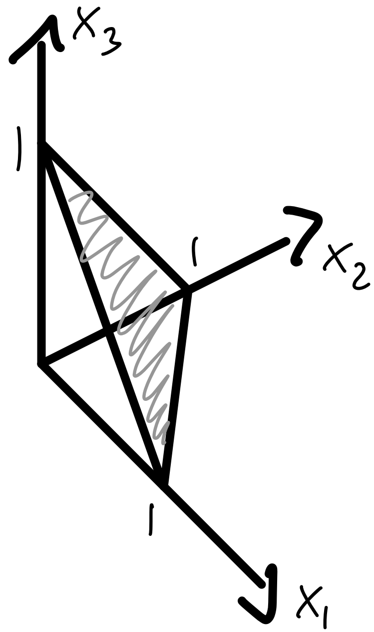
$$\{x \in \mathbb{R}^n | \, Ax \geq b\},$$

where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$.

All polyhedra are convex, and so are their intersections.

Geometry of Linear Programming (Contd.)

For example, suppose $x \in \mathbb{R}^3$, and the constraints are $\mathbf{1}_3^\top x = 1$, $x \geq 0$.



So constraints define the space in which your solution must lie in. The idea of the simplex algorithm is to traverse this space in the most efficient way possible to find the solution.

Geometry of Linear Programming (Contd.)

Basic feasible solutions are the key concept that allow for efficient computation of linear programming solutions.

Definition 4.3 (Basic Feasible Solution)

Consider a polyhedron P defined by linear equality and inequality constraints, and let x^* be an element of $\mathbb{R}^n$.

(a) The vector x^* is a *basic solution* if:

  (i) All equality constraints are active;

  (ii) Out of the constraints that are active at x^*, there are n of them that are linearly independent.

(b) If x^* is a basic solution that satisfies all of the constraints, we say that it is a *basic feasible solution*.

Note that any linear constraints can be written in matrix form. For example $x \geq 0 \Rightarrow \mathbf{I}x \geq 0$. We may thus say "constraints are linearly independent" if the corresponding rows in the constraint matrix are linearly independent.

Geometry of Linear Programming (Contd.)

Henceforth, everything is assumed to be in standard form and assume A is full row rank $A_m, m \leq n$. Assuming A is full rank is harmless, as it means there are no redundant constraints (Theorem 2.5 in Bertsimas and Tsitsiklis (1997)).

Theorem 4.1

Consider the constraints $Ax = b$ and $x \geq 0$ and assume that the $m \times n$ A has linearly independent rows. A vector $x \in \mathbb{R}^n$ is a basic solution if and only if we have $Ax = b$, and there exist indices $B(1), \ldots, B(m)$ such that:

(a) The columns $A_{B(1)}, \ldots, A_{B(1)}$ are linearly independent;

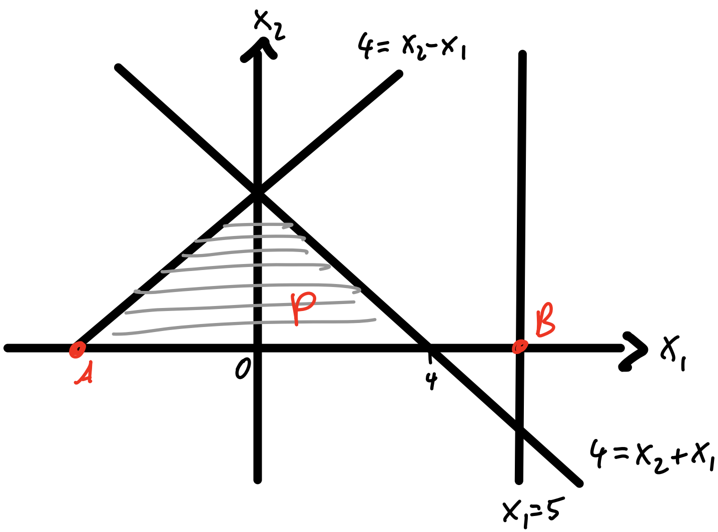
(b) If $i \neq B(1), \ldots, B(m)$, then $x_i = 0$.

The only unknown variables are thus $x_{B(1)}, \ldots, x_{B(m)}$, and they can be solved for using the fact that $Ax = b$.

Geometry of Linear Programming (Contd.)

For example, consider $P \in \mathbb{R}^3$ defined by the constraints

$$4 \geq x_2 - x_1, \quad 5 \geq x_1, \quad 4 \geq x_2 + x_1, \quad x_2 \geq 0, \quad x_3 = 0.$$



At A, binding constraints are:

$x_3 = 0,\ x_2 = 0,\ 4 = x_2 - x_1.$

$$\Rightarrow \begin{bmatrix} -1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \\ 0 \end{bmatrix}$$

Full rank $\Rightarrow$ 3 lin. indep. const.

$\Rightarrow$ A is a basic feasible sol.

Geometry of Linear Programming (Contd.)

As the example suggest, basic feasible solutions have a convenient geometric interpretation.

Definition 4.4 (Extreme Point)

Let P be a polyhedron. A vector $x \in P$ is an *extreme point* of P if

$$\nexists y, z \in (P \setminus x), \lambda \in [0,1] : \quad x = \lambda y + (1-\lambda) z.$$

Definition 4.5 (Vertex)

Let P be a polyhedron. A vector $x \in P$ is a *vertex* of P if

$$\exists c : \quad c'x < c'y, \quad \forall y \in (P \setminus x).$$

In other words, x is a vertex of P if and only if P is on one side of the hyperplane $\{y |\, c'y = c'x\}$ which meets P only at the point x.

The following theorem then gives equivalence between basic feasible solutions of a linear program, and vertices and extreme points of a polyhedron.

Theorem 4.2

Let P be a nonempty polyhedron and let $x^* \in P$. Then, the following are equivalent:

(a) x^* is a vertex;

(b) x^* is an extreme point;

(c) x^* is a basic feasible solution.

Corollary 4.2.1

Given a finite number of linear inequality constraints, there can only be a finite number of basic or basic feasible solutions.

This follows from the fact that at any basic feasible solutions, there are n linearly independent active constraints. Since n linearly independent constraints defines a unique point, it follows that different basic solutions correspond to different sets of n active constraints. We can thus bound the number of basic feasible solutions by $\binom{m}{n} < \infty$ (i.e., the unique combinations of choosing n out of m constraints).

It is reassuring that the number of basic feasible solutions is guaranteed to be finite, however, it can still be very large. For example, the unit cube $\{x \in \mathbb{R}^n | \, 0 \leq x_i \leq 1, \, i = 1, \ldots, n\}$ is defined in terms of $2n$ constraints, but has 2^n basic feasible solutions.

Geometry of Linear Programming (Contd.)

We have now characterized a set of solutions to general linear programs in standard form: the set of basic feasible solutions which was shown to be finite. The next key results establish 1) existence of such solutions, and – crucially – 2) that the optimal solution is a basic feasible solution.

We begin with another definition:

Definition 4.6

A polyhedron $P \subset \mathbb{R}^n$ contains a line if there exists a vector $x \in P$ and a nonzero vector $d \in \mathbb{R}^n$ such that $x + \lambda d \in P$, $\forall \lambda \in \mathbb{R}^n$.

Notice that the positive orthant $\{x | \, x \geq 0\}$ does not contain a line. It then follows that a polyhedron in standard form, which is contained in the positive orthant, does not contain a line either.

Existence of basic feasible solutions for polyhedra in standard form is given by the next theorem:

Theorem 4.3

Suppose that the polyhedron $P = \{x \in \mathbb{R}^n | \, a_i' x \geq b_i, \; i = 1, \ldots, m\}$ is nonempty. Then, the following are equivalent:

(a) The polyhedron P has at least one extreme point.

(b) The polyhedron P does not contain a line.

(c) There exist n vectors out of the family $a_1, \ldots, a_n$, which are linearly independent.

Corollary 4.3.1

Every nonempty bounded polyhedron and every nonempty polyhedron in standard form has at least one basic feasible solution.

Geometry of Linear Programming (Contd.)

And finally, the next theorem ensures that the optimal solution (if it exists) is a basic feasible solution:

Theorem 4.4

Consider the linear programming problem of minimizing $c'x$ over a polyhedron P. Suppose that P has at least one extreme point and that there exists an optimal solution. Then, there exists an optimal solution which is an extreme point of P.

Theorem 4.5

Consider the linear programming problem of minimizing $c'x$ over a polyhedron P. Suppose that P has at least one extreme point. Then, either the optimal cost is equal to $-\infty$, or there exists an extreme point which is optimal.

Geometry of Linear Programming (Contd.)

To summarize: The provided results show that the solution to a linear program is an extreme point in the polyhedron defined by the constraints. We've shown how these points/vertices can be solved for as they pertain to n linearly independent constraints, and that there exists only finitely many of such extreme points. Finally, we've concluded that if their exists an optimal solution to the linear program, it must be one of these points.

These results are key for understanding why linear programs can be readily solved. Crucially, the key idea of the simplex algorithm (not discussed today) is based on traversing between basic feasible solutions, and it's convergence is guaranteed given that there are only finitely many such solutions.

Duality Theory

In the final part of the TA session, we briefly review key results from duality theory in the context of linear programming, and then conclude with an application to the quantile regression.

Recall the following relationship between a linear program in standard from (LHS) and its dual (RHS):

$$
\begin{aligned}
\min_{x} \quad & c'x \\
\text{s.t.} \quad & Ax = b, \\
& x \geq \mathbf{0}_n.
\end{aligned}
\qquad \Leftrightarrow \qquad
\begin{aligned}
\min_{p} \quad & p'b \\
\text{s.t.} \quad & p'A \leq c', \\
& p \in \mathbb{R}^n.
\end{aligned}
$$

Notice that the dual problem associates a price variable with each constraint and searches for prices under which the presence (or absence) of the constraints does not affect the optimal cost.

Duality Theory (Contd.)

The following two results are helpful in characterizing solutions to the primal and dual problems of a linear program.

Theorem 4.6 (Strong Duality)

If a linear programming problem has an optimal solution, so does its dual, and the respective optimal costs are equal.

Theorem 4.7 (Complementary Slackness)

Let x and p be feasible solutions to the primal and the dual problem, respectively. The vectors x and p are optimal solutions for the two respective problems if and only if

$$p_i(a_i'x - b_i) = 0, \quad \forall i$$
$$(c_j - p'A_j)x_j = 0, \quad \forall j.$$

If the constraints don't bind at the optimum, then they don't constrain the problem: relaxing them has no impact on the objective value.

Application to Quantile Regression

Recall that quantile regression $\arg\min_{\beta_\tau} \sum_{i=1}^{n} \rho_\tau(y_i - x_i^\top \beta_\tau)$ is equivalent to

$$\min_{u,v,\beta_{\tau+},\beta_{\tau-}} \quad \begin{bmatrix} \tau \mathbf{1}_n & (1-\tau)\mathbf{1}_n & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} u \\ v \\ \beta_{\tau+} \\ \beta_{\tau-} \end{bmatrix}$$

$$\text{s.t.} \quad \begin{bmatrix} \mathbf{I} & -\mathbf{I} & X & -X \end{bmatrix} \begin{bmatrix} u \\ v \\ \beta_{\tau+} \\ \beta_{\tau-} \end{bmatrix} = Y,$$

$$\begin{bmatrix} u \\ v \\ \beta_{\tau+} \\ \beta_{\tau-} \end{bmatrix} \geq \mathbf{0}.$$

Application to Quantile Regression (Contd.)

The corresponding dual problem is

$$\max_{p} \quad p'Y$$
$$\text{s.t.} \quad \begin{bmatrix} \mathbf{I} \\ -\mathbf{I} \\ X' \\ -X' \end{bmatrix} p \leq \begin{bmatrix} \tau \mathbf{1}_n \\ (1-\tau)\mathbf{1}_n \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix},$$
$$p \in \mathbb{R}^n.$$

Notice that this places box constraints on p where $(\tau - 1) \leq p_i \leq \tau, \forall i$, and that the last two sets of constraints imply $X'p = 0$.

Let $X \in \mathbb{R}^{n \times k}$, then there are k equality constraints and n box constraints. Assuming that X has full column rank, it follows from $p \in \mathbb{R}^n$ that $n - k$ of the box constraints must bind at the optimum.

Application to Quantile Regression (Contd.)

Consider now the scenario without degeneracy. Then there is a subset of $i_1, \ldots, i_k : \tau - 1 < p_{i_l} < \tau, l = 1, \ldots, k$ where the box constraints don't bind.

The box constraint were derived from the constraints on u and v, so they are the corresponding Lagrange multipliers. By complementary slackness, it follows that $u_{i_k} = v_{i_k} = 0$, so that for k observations, the residuals from quantile regression are 0.

We can now use the k observations to recover β_τ: $\hat{\beta}_\tau = \tilde{X}^{-1} y$ where $\tilde{X} = [X_{i_1}^\top, \ldots, X_{i_k}^\top]^\top$.

References

Bertsimas, D. and Tsitsiklis, J. N. (1997). *Introduction to linear optimization*, volume 6. Athena Scientific Belmont, MA.