ORIE6300 **FA24**

Mathematical Programming I

Notes

Authors: Shaleen Baral

Contents

1. Introduction: LP and Duality

1.1. Why take this class?

Mathematical programming is a very well-studied field. Many optimization problems can be recast as or at least approximated well by mathematical programs. Doing so not only allows us to use algorithms but also use the theory of mathematical programming to uncover structural insights. The course will roughly cover

a. Linear Programming
   - Geometry
   - Duality Theory
   - Algorithms

b. Convex Programming

c. First-Order Methods

1.2. Logistics

Weekly assignments due every Thursday night (11:59pm). Grading scheme is as follows:
- 40% HW assignments (~10)
- 20% in-person final
- 15% take-home midterm
- 15% take-home final
- 10% participation/scribing

1.3. Linear Programming

Let $x \in \mathbb{R}^n$ denote the *decision variables*. We are looking to solve *for* these decision variables. Let $c \in \mathbb{R}^n$ be fixed. Then, $c^T x$ is our *objective function*. We also specify a *constraint* $Ax \leq b$ on our decision variables where $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$ are both fixed. The inequality is read element wise. A linear program optimizes the objective function subject to these given constraints,

$$\max c^T x \qquad \text{s.t.} \qquad Ax \leq b, x \in \mathbb{R}^n.$$

We say $x \in \mathbb{R}^n$ is a *feasible solution* if $Ax \leq b$.

We define the *feasible region*, $Q = \{x \in \mathbb{R}^n \mid Ax \leq b\}$.

A feasible solution $x \in \mathbb{R}^n$ is *optimal* if its *value*, $c^T x$, is at least as large as the value of any other feasible solution.

$$\min b^T y \qquad \text{s.t.} \qquad A^T y = c, y \in \mathbb{R}^m, y \geq 0.$$

Proposition 1.3.1 (Weak Duality): If $x \in \mathbb{R}^n$ feasible in primal and $y \in \mathbb{R}^m$ feasible in dual then $c^T x \leq b^T y$.

Proof: We know that $A^T y = c$. Then,

$$c^T x = y^T A x \leq y^T b.$$

The last inequality is justified by the fact that $y \geq 0$.

□

2. Strong Duality

Recall the primal and dual programs from the previous lecture

$$\begin{aligned} \text{PRIMAL:} \quad & \max c^T x \qquad \text{s.t.} \qquad Ax \leq b. \\ \text{DUAL:} \quad & \min b^T y \qquad \text{s.t.} \qquad A^T y = c, y \geq 0 \end{aligned}$$

First, we interpret our linear program in a “physical” sense. We make three observations,

a. if we place ball at an optimal feasible solution, it doesn’t accelerate

b. if the ball doesn't accelerate, the forces acting on it sum to 0
c. "wall" $a_i x_i \leq b_i$ may exert a force on ball along $-a_i^T$; if $a_i x < b_i$, then this force is 0.

Lemma 2.1: If $x^\star$ is an optimal feasible solution to the primal, then there exists $y \in \mathbb{R}^n$, $y \geq 0$, $c + \sum_{i=1}^m -y_i a_i^T = 0$, and $y_i = 0$ whenever $a_i x^\star < b_i$.

Proposition 2.1 (strict duality): If $x^\star$ is an optimal feasible solution to the primal, there exists an optimal feasible solution of the dual, $y^\star$, such that $c^T x^\star = b^T y^\star$.

Proof: Assuming $c^T x^\star = b^T y^\star$, we are guaranteed the optimality of $y^\star$ by weak duality.

Now, we actually show that such a feasibly solution must exist. Fix $x^\star$, let $y^\star$ be the dual solution given by the prior lemma. Then $y^\star$ is dual feasible as

$$c^T x^\star = \left(A^T y^\star\right)^T x^\star = {y^\star}^T A x^\star \leq {y^\star}^T b = b^T y^\star.$$

The middle term can be expanded to incorporate how tight the inequality is.

$${y^\star}^T A x^\star = y^\star(b + Ax^\star - b) = y^\star b + {y^\star}^T (Ax^\star - b)$$

Note however that,

$${y^\star}^T (Ax^\star - b) = \sum_{i=1}^m y_i (a_i x^\star - b_i)$$

If $(a_i x^\star - b_i)$ is non-zero for some $i \in [m]$, then by the prior lemma $y_i = 0$. Thus, this term must be zero.

$$c^T x^\star = {y^\star}^T A x^\star = y^\star b$$

$\square$

Now, we focus on taking the dual of the dual. First, there are two forms in which we can specify a linear program

$$\begin{aligned} \text{basic form:} \quad & \max c^T x \quad \text{s.t.} \quad Ax \leq b. \\ \text{standard form:} \quad & \min b^T y \quad \text{s.t.} \quad A^T y = c, y \geq 0 \end{aligned}$$

There are two strategies

a. transform LPs from standard form into basic form, then take dual.
Let $\overline{c} = b, \overline{A} = A^T, \overline{b} = c$. Then, the LP in standard form given above can be represented in basic form as follows

$$\begin{aligned} \max\,(-\overline{c})^T x \quad \text{such that} \quad -\overline{A}x &\leq -\overline{b} \\ \overline{A}x &\leq \overline{b} \\ -Ix &\leq 0 \end{aligned}$$

Now, we take the dual of this program. We have

$$\min \ \begin{pmatrix} -\overline{b} & \overline{b} & 0 & \cdots & 0 \end{pmatrix} y \quad \text{such that} \quad \begin{pmatrix} -\overline{A}^T & \overline{A}^T & -I \end{pmatrix} y = -\overline{c}$$

For convenience, let $y = \begin{pmatrix} s \\ t \\ w \end{pmatrix}$ such that

$$\begin{pmatrix} -\overline{A}^T & \overline{A}^T & -I \end{pmatrix} y = -\overline{A}^T s + \overline{A}^T t - w.$$

This lets us rewrite our program as

$$\max \overline{b}^T (s - t) \quad \text{such that} \quad \overline{A}^T (s - t) + w = \overline{c}$$
$$s, t, w \geq 0.$$

Making the substitution $z = s - t$, we simplify this expression.

$$\max \overline{b}^T z \quad \text{such that} \quad \overline{A}^T z + w = \overline{c}$$
$$w \geq 0.$$

Finally the requirement that $w \geq 0$ is equivalent to simply saying $\overline{A}^T z \leq \overline{c}$. Thus, we have recovered the following program

$$\max \overline{b}^T z \quad \text{such that} \quad \overline{A}^T z \leq \overline{c}.$$

b. directly rederive dual for LPs in standard form.

3. Polyhedron

3.1. Polyhedrons

Definition 3.1.1: A set $S \subseteq \mathbb{R}^n$ such that $S = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ for some $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$ is called a *polehdron.*

Definition 3.1.2: A set $S \subseteq \mathbb{R}^n$ is *convex* if, for all $x, y \in S$ and all $0 \leq \theta \leq 1$, we have $\theta x + (1 - \theta)x \in S$.

Lemma 3.1.1: Any polyhedron P is convex.

Proof: Suppose $P = \{x \mid Ax \leq b\}$ and let $x, y \in P$. Then for $\theta \in [0, 1]$, consider

$$A[\theta x + (1 - \theta)x] = \theta Ax + (1 - \theta)Ax \leq \theta b + (1 - \theta)b = b.$$

Thus, $\theta x + (1 - \theta)x \in P$.

□

3.2. Vertices

For the remaining definitions, let $P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ be a polyhedron.

Definition 3.2.1: A point $x \in P$ is a *vertex* if $\exists c \in \mathbb{R}^n$ such that $c^T x > c^T y, \ \forall y \in P - x$.

Definition 3.2.2: A point $x \in P$ is an *extreme point* if there aren't any $z \in P - x, \theta \in [0, 1]$ such that $x = \theta y + (1 - \theta)z$.

Definition 3.2.3: Consider $x \in P$. Then we call $a_i x \leq b_i$ a *binding constraint* if $a_i x = b$. Otherwise, we call it a *non-binding constraint.*

Definition 3.2.4: For any $x \in P$, we may define the following matrices

$$A_= = \text{the submatrix of } A \text{ containing the rows of binding constraints},$$
$$b_= = \text{the subvector of } b \text{ for the binding rows},$$
$$A_< = \text{the submatrix of } A \text{ containing the rows of non-binding constraints},$$
$$b_< = \text{the subvector of } b \text{ for the non-binding rows}.$$

Definition 3.2.5: A point $x \in P$ is a basic feasible solution (BFS) if $\operatorname{rank}(A_=) = n$.

Proposition 3.2.1: Let $P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$. The following are equivalient:

a. x is a vertex,
b. x is an extreme point,
c. x is a BFS.

Proof:

- *(i)* $\Rightarrow$ *(ii)*

We prove this by contradiction. Suppose *(ii)* doesn't hold.

We have $c \in \mathbb{R}^n$ such that $c^T x > c^T y, \forall y \in P - x$. Let $y, z \in P - x$ such that $x = \theta y + (1 - \theta) z$ for some $\theta \in [0, 1]$. Then consider,

$$c^T x = c^T (\theta y + (1 - \theta) z) = \theta c^T y + (1 - \theta) c^T z < \theta c^T x + (1 - \theta) c^T x = c^T x$$

Thus, we have a contradiction.

- *(ii)* $\Rightarrow$ *(iii)*

We prove this by contrapositive, $\neg$ *(iii)* $\Rightarrow \neg$ *(ii)*.

We know that $\operatorname{rank}(A_=) < n$. Then $\exists y \neq 0 \in \mathbb{R}^n$ such that $A_= y = 0$. Note that for any $\varepsilon > 0$ we can conclude that $x \pm \varepsilon y$ satisfies the binding constraints as verified below

$$A_=(x \pm \varepsilon y) = A_= x \pm \varepsilon A y = A_= x = b.$$

Now, we show that for an appropriate choice of $\varepsilon > 0$, these vectors also satisfy the non-binding constraints. Note that $A_< x < b_<$ and hence, $b_< - A_< x > 0$. Thus, for small enough $\varepsilon > 0$ we have

$$\pm \varepsilon A_< y < b_< - A_< x.$$

This immediately gives us what we want,

$$A_<(x \pm \varepsilon y) = A_< x + \varepsilon A_< y < A_< x + b_< - A_< x < b_<.$$

So, x can be written as a convex combination of $x \pm \varepsilon y$.

- *(iii)* $\Rightarrow$ *(i)*

Define $I = \{1 \leq i \leq m \mid a_i x = b_i\}$. Then, let $c = \sum_{i \in I} a_i^T$. Then,

$$c^T x = \left(\sum_{i \in I} a_i \right) x = \sum_{i \in I} a_i x = \sum_{i \in I} b_i.$$

Let $y \in P$. Then,

$$c^T y = \sum_{i \in I} a_i y \leq \sum_{i \in I} b.$$

If this inequality is tight, that is $c^T y = \sum_{i \in I} b$ then, $a_i y = b_i$. As $\operatorname{rank}(A_=) = n$, $A_= y = b$ has a unique solution, $y = x$.

$\square$

4. More Polyhedrons

4.1. Existence of Vertices

Definition 4.1.1: A polyhedron P contains a line if $\exists\, x \in P, y \neq 0 \in \mathbb{R}^n$ such that $\{x + \lambda y \mid y \in \mathbb{R}\} \subseteq P$. Otherwise, P is *pointed*.

Proposition 4.1.1: Let $P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ be a non-empty polyhedron. The following three are equivalent

a. P has at least one vertex
b. P is pointed
c. $\operatorname{rank}(A) = n$

Proof:

- *(ii) ⇒ (i)*

Consider any point $x \in P$. If $\operatorname{rank}(A_=) = n$, then x is a vertex and we are done. Otherwise, we have $\operatorname{rank}(A_=) < b$. Then $\exists y \neq 0 \in \mathbb{R}^n$ such that $A_= y = 0$. We know that the line $\{x + \lambda y \mid \lambda \in \mathbb{R}\}$ isn't entirely contained in P as P is pointed. That is to say, $\exists \lambda^\star$ and $1 \leq j \leq m$ such that $a_j x < b_j$ however $a_j(x + \lambda^\star y) = b_j$. Let $x' = x + \lambda^\star y$. Equivalently, we have $x' \in P$. Then,

$$A_= x' = A(x + \lambda^\star y) = \underbrace{A_= x}_{b_=} + \lambda^\star A_= y = b_=$$

$A'_=$ is a superset of the rows, then $\exists M \in \mathbb{R}$ of $A_=$.

If a is a row of $A_=$, $ay = 0$. Note that $a_j y \neq 0$. That is a_j is not a linear combination of the rows of $A_=$. Thus, $\operatorname{rank}(A_{='}) > \operatorname{rank}(A_=)$.

- *(i) ⇒ (iii)*

Suppose $x \in P$ is a vertex. Then

$$n = \operatorname{rank}(A_=) \leq \operatorname{rank}(A) \leq n.$$

Thus, $\operatorname{rank}(A) = n$.

- *(iii) ⇒ (ii)*

Assume $\operatorname{rank}(A) = n$. For the sake of contradiction, suppose P contains a line $\{x + \lambda y \mid \lambda \in \mathbb{R}\} \subseteq P$ for $y \neq 0$. Let $1 \leq j \leq m$. Then we know that $\forall \lambda \in \mathbb{R}$, $a_j(x + \lambda y) \leq b_j$ holds. Taking the limit as $\lambda \to \operatorname{sgn}(a_j y) \cdot \infty$, we get that $a_j(x + \lambda y) \to \infty$ if $a_j y \neq 0$. Thus, it must be the case that $a_j y = 0$. By injectivity of A, we note that $y = 0$. Thus, it must be the case that P is pointed.

□

Definition 4.1.2: A set $X \subseteq B$ where B is some normed space is called bounded whenever $\exists M \in \mathbb{R}$, such that $\|x\| \leq M$ for all $x \in P$.

Corollary 4.1.1.1: If P is nonempty and bounded then P has at least one vertex.

Corollary 4.1.1.2: The feasible region of an LP in standard form, if it is nonempty, has at least one vertex.

Proposition 4.1.2: Given an LP maximizes $c^T x$ over $x \in P$, P is nonempty, pointed polyhedron. If there is an optimal feasible solution $x^\star$, there is an optimal feasible solution that is a vertex of P.

Proof: Define $Q = P \cap \{x \in \mathbb{R}^n \mid c^T x = c^T x^\star\}$. Q is a nonempty polyhedron as $x^\star \in Q$. Note that Q is pointed becaue $Q \subseteq P$ is pointed. By the prior proposition, $\exists$ vertex v of Q. Note that v is feasible and $c^T v = c^T x^\star$, so v is optimal. Suppose that v is not an extreme point of P, i.e. $y, z \in P - \{v\}, 0 \leq \theta \leq 1$ such that $v = \theta y + (1 - \theta) z$. Then,

$$c^T x^\star = c^T v = c^T(\theta y) + (1-\theta)z = \theta c^T y + (1-\theta)c^T z \le c^T x^\star$$

Note that $c^T y = c^T z = c^T x^\star$ as otherwise $c^T x^\star < c^T x^\star$, a clear contradiction. Note that as $y, z \in Q - \{v\}$, this would imply that v is not an extreme point of Q.

□

4.2. Convex Hull

Definition 4.2.1: Given $v_1, ..., v_k \in \mathbb{R}^n$, A *convex combination* is any $\sum_{i=1}^k \lambda_i v_i$ where $\lambda_i \ge 0$ for all $i \in [k]$ and $\sum_{i=1}^k \lambda_i = 1$.

Definition 4.2.2: Given $v_1, ..., v_k \in \mathbb{R}^n$, the *convex hull*, $\operatorname{conv}(\{v_1, ..., v_k\})$ is the set of all their convex combinations.

Definition 4.2.3: Any $S \subseteq \mathbb{R}^n$ that can be written as $S = \operatorname{conv}(\{v_1, ..., v_k\})$ for some $v_1, ..., v_k$ is called a *polytope*.

Proposition 4.2.1 (Carathéodory's Theorem): Given $v_1, ..., v_k \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$, such that $y \in \operatorname{conv}(\{v_1, ..., v_k\})$. Then, there exists $S \subseteq \{v_1, ..., v_k\}$, $|S| \le n + 1$ such that $y \in \operatorname{conv}(S)$.

Proof: Let $A = \begin{pmatrix} v_1 & v_2 & \cdots & v_k \end{pmatrix} \in \mathbb{R}^{n \times k}$ and $\lambda = (\lambda_1, \lambda_2, ..., \lambda_k) \in \mathbb{R}^k$. Then, $y \in \operatorname{conv}(\{v_1, ..., v_k\})$ translates to the condition that $A\lambda = y$ for $\lambda \ge 0$ such that $\mathbf{1} \cdot \lambda = 1$. Note that his polyhedron is nonempty. By the prior proposition, there exists som vertex $\lambda^\star$ of this polyhedron. Furthermore, $\lambda^\star$ is a BFS. Thus, $A_=$ has rank k. Thus, at least $k - (n+1)$ entries of $\lambda^\star$ are 0.

□

5. Bounded Polyhedra = Polytopes

5.1. Bounded Polyhedra ⊆ Polytope

5.2. Separating Hyperplane Theorem

Proposition 5.2.1 (Separating Hyperplane Theorem): If $c \in \mathbb{R}^n$ is nonempty, closed and convex, $y \notin C$, then there exist $a \in \mathbb{R}^n$, $b \in \mathbb{R}$ such that $a^T y > b > a^T x$.

6. Polytopes $\subseteq$ Bounded Polyhedra

We shall prove the following result in this lecture.

Proposition 6.1: Any polytope that con tains 0 in its interior is a bounded polyhedron.

Proof: Let P be our polytope. It is enough, by Lemma 6.1, to show that $(P^o)^o$ is a bounded polyhedron. Then, by Lemma 6.2, we note that P^o is a polyhedron. Since P^o contains $\mathbf{0}$, it is a bounded polyhedron. Thus, P^o is a polytope and $P = (P^o)^o$ is a polyhedron. Finally, P is bounded by the convexity of the Euclidean norm.

□

Definition 6.1: Let $S \subseteq \mathbb{R}^n$. Then, the *polar of* S is the set

$$S^o = \{z \in \mathbb{R}^n \mid z^T x \leq 1, \ \forall x \in S\}.$$

Lemma 6.1: If $C \subseteq \mathbb{R}^n$ is closed, convex and contains 0, then $(C^o)^o = C$.

Proof: Fix $x \in C$. We need to show that $x^T z \leq 1$ for all $z \in C^o$. But since $z \in C^o$, $x^T z = z^T x \leq 1$. Thus, $C \subseteq (C^o)^o$.

Fix $x \in (C^o)^o$. For contradiction, assume $x \notin C$. □

Lemma 6.2: The polar of a polytope is a polyhedron.

Lemma 6.3: If $\mathbf{0}$ is in the interior of S, then S^o is bounded.