Massachusetts Institute of Technology
18.433: Combinatorial Optimization
Michel X. Goemans

Handout 8
September 29, 2005

LP and Polyhedral Combinatorics

Brief summary of what was seen in the introductory lectures on linear programming and polyhedral combinatorics.

Definition 1 *A halfspace in $\mathbb{R}^n$ is a set of the form $\{x \in \mathbb{R}^n : a^T x \leq b\}$ for some vector $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$.*

Definition 2 *A polyhedron is the intersection of finitely many halfspaces: $P = \{x \in \mathbb{R}^n : Ax \leq b\}$.*

Definition 3 *A polytope is a bounded polyhedron.*

Definition 4 *If P is a polyhedron in $\mathbb{R}^n$, the projection P_k of P is defined as $\{y = (x_1, x_2, \cdots, x_{k-1}, x_{k+1}, \cdots, x_n) : x \in P \text{ for some } x_k\}$.*

We claim that P_k is also a polyhedron and this can be proved by giving an explicit description of P_k in terms of linear inequalities. For this purpose, one uses Fourier-Motzkin elimination. Let $P = \{x : Ax \leq b\}$ and let

- $S_+ = \{i : a_{ik} > 0\}$
- $S_- = \{i : a_{ik} < 0\}$
- $S_0 = \{i : a_{ik} = 0\}$

Clearly, any element in P_k must satisfy the inequality $a_i^T x \leq b_i$ for all $i \in S_0$ (these inequalities not involve x_k). Similarly, we can take a linear combination of an inequality in S_+ and in S_- to eliminate the coefficient of x_k. This shows that the inequalities

$$a_{ik}\left(\sum_j a_{lj}x_j\right) - a_{lk}\left(\sum_j a_{kj}x_j\right) \leq a_{ik}b_l - a_{lk}b_i \tag{1}$$

for $i \in S_+$ and $l \in S_-$ are satisfied by all elements of P_k. Conversely, for any vector $(x_1, x_2, \cdots, x_{k-1}, x_{k+1}, \cdots, x_n)$ satisfying (1) for all $i \in S_+$ and $l \in S_-$ and also

$$a_i^T x \leq b_i \text{ for all } i \in S_0 \tag{2}$$

we can find a value of x_k such that the resulting x belongs to P (by looking at the bounds on x_k that each constraint imposes, and showing that the largest lower bound is smaller than the smallest upper bound). This shows that P_k is described by (1) and (2).

Definition 5 *Given points $a^{(k)} \in \mathbb{R}^n$,*

- *a linear combination is* $\sum_k \lambda_k a^{(k)}$ *where* $\lambda_k \in \mathbb{R}$,
- *an affine combination is* $\sum_k \lambda_k a^{(k)}$ *where* $\lambda_k \in \mathbb{R}$ *and* $\sum_k \lambda_k = 1$,
- *a conical combination is* $\sum_k \lambda_k a^{(k)}$ *where* $\lambda_k \geq 0$,
- *a convex combination is* $\sum_k \lambda_k a^{(k)}$ *where* $\lambda_k \geq 0$ *and* $\sum_k \lambda_k = 1$.

The set of all linear combinations of elements of S is called the linear hull of S and denoted by $lin(S)$. Similarly, by replacing linear by affine, conical or convex, we define the affine hull, $aff(S)$, the conic hull, $cone(S)$ and the convex hull, $conv(S)$. We can give an equivalent definition of a polytope.

Definition 6 *A polytope is the convex hull of a finite set of points.*

The fact that Definition 6 implies Definition 3 can be shown by using Fourier-Motzkin elimination repeatedly on

$$\begin{aligned} x - \sum_k \lambda_k a^{(k)} &= 0 \\ \sum_k \lambda_k &= 1 \\ \lambda_k &\geq 0 \end{aligned}$$

to eliminate all variables λ_k and keep only the variables x. The converse will be discussed later in these notes.

1 Necessary and Sufficient Conditions for the Solvability of System of Inequalities

In linear algebra, we saw that, for $A \in R^{m \times n}$, $b \in R^n$, $Ax = b$ has no solution $x \in \mathbb{R}^n$ if and only if there exists a $y \in R^m$ with $A^T y = 0$ and $b^T y \neq 0$ (in 18.06 notation/terminology, this is equivalent to saying that the column space $C(A)$ is orthogonal to the left null space $N(A^T)$).

One can state a similar *Theorem of the Alternatives* for systems of linear inequalities.

Theorem 1 $Ax \leq b$ *has no solution* $x \in \mathbb{R}^n$ *if and only if there exists* $y \in \mathbb{R}^m$ *such that* $y \geq 0$, $A^T y = 0$ *and* $b^T y < 0$.

One can easily show that both systems indeed cannot have a solution since otherwise $0 > b^T y = y^T b \geq y^T Ax = 0^T x = 0$. For the other direction, one takes the insolvable system $Ax \leq b$ and use Fourier-Motzkin elimination repeatedly to eliminate all variables and thus obtain an inequality like $0^T x \leq c$ where $c < 0$. In the process one has derived a vector y with the desired properties.

Another version of the above theorem is Farkas' lemma:

Theorem 2 $Ax = b$, $x \geq 0$ *has no solution if and only if there exists* y *with* $A^T y \geq 0$ *and* $b^T y < 0$.

2 Linear Programming Basics

A linear program (LP) is the problem of minimizing or maximizing a linear function over a polyhedron:

$$(P)\qquad \begin{array}{ll} \text{Max} & c^T x \qquad\qquad \text{subject to:} \\ & Ax \leq b \end{array}$$

Other equivalent forms involve equalities as well, or nonnegative constraints $x \geq 0$. One version that is often considered when discussing algorithms for linear programming (especially the simplex algorithm) is $\min\{c^T x : Ax = b, x \geq 0\}$.

Another linear program, *dual* to (P), plays a crucial role:

$$(D)\qquad \begin{array}{ll} \text{Min} & b^T y \qquad\qquad \text{subject to:} \\ & A^T y = c \\ & y \geq 0 \end{array}$$

Weak duality is clear: For any feasible solutions x and y to (P) and (D), we have that $c^T x \leq b^T y$. Indeed, $c^T x = y^T Ax \leq b^T y$. The dual was precisely built to get an upper bound on the value of any primal solution. For example, to get the inequality $y^T Ax \leq b^T y$, we need that $y \geq 0$ since we know that $Ax \leq b$.

Strong duality is the most important result in linear programming; it says that we can prove the optimality of a primal solution x by exhibiting an optimum dual solution y.

Theorem 3 *If (P) and (D) are feasible then their optimum values are equal.*

The proof of strong duality is obtained by writing a big system of inequalities in x and y which says that (i) x is primal feasible, (ii) y is dual feasible and (iii) $c^T x = b^T y$. Then use the Theorem of the Alternatives to show that the infeasibility of this system of inequalities would contradict the feasibility of either (P) or (D).

Looking at $c^T x = y^T Ax \leq b^T y$, we observe that to get equality between $c^T x$ and $b^T y$, we need *complementary slackness*:

Theorem 4 *If x is feasible in (P) and y is feasible in (D) then x is optimal in (P) and y is optimal in (D) if and only if for all i either $y_i = 0$ or $\sum_j a_{ij} x_j = b_i$ (or both).*

Linear programs can be solved efficiently either by interior-point algorithms or by the ellipsoid algorithm.

3 Faces of Polyhedra

Definition 7 *$\{a^{(i)} \in \mathbb{R}^n : i \in K\}$ are linearly independent if $\sum_k \lambda_k a^{(k)} = 0$ implies that $\lambda_k = 0$ for all k.*

Definition 8 $\{a^{(i)} \in \mathbb{R}^n : i \in K\}$ *are affinely independent if* $\sum_k \lambda_k a^{(k)} = 0$ *and* $\sum_k \lambda_k = 0$ *together imply that* $\lambda_k = 0$ *for all* k.

Definition 9 *The dimension,* $dim(P)$*, of a polyhedron* P *is the maximum number of affinely independent points in* P *minus 1.*

The dimension can be -1 (if P is empty), 0 (when P consists of a single point), 1 (when P is a line segment), and up to n when P is in $\mathbb{R}^n$. In the latter case, we say that P is *full-dimensional*.

Definition 10 $\alpha^T x \leq \beta$ *is a valid inequality for* P *if* $\alpha^T x \leq \beta$ *for all* $x \in P$.

Observe that for an inequality to be valid for $conv(S)$ we only need to make sure that it is satisfied by all elements of S.

Definition 11 *A face of a polyhedron* P *is* $\{x \in P : \alpha^T x = \beta\}$ *where* $\alpha^T x \leq \beta$ *is some valid inequality of* P.

By definition, all faces are polyhedra. The empty face (of dimension -1) is *trivial*, and so is the entire polyhedron P (which corresponds to the valid inequality $0^T x \leq 0$). Non-trivial are those whose dimension is between 0 and $dim(P) - 1$. Faces of dimension 0 are called *extreme points* or *vertices*, faces of dimension 1 are called *edges*, and faces of dimension $dim(P) - 1$ are called *facets*.

Although there are infinitely many valid inequalities, there are only finitely many faces.

Theorem 5 *Let* $A \in \mathbb{R}^{m \times n}$. *Then any face of* $P = \{x \in \mathbb{R}^n : Ax \leq b\}$ *corresponds to the set*

$$\sum_j a_{ij} x_j = b_i \textit{ for all } i \in I$$

$$\sum_j a_{ij} x_j \leq b_i \textit{ for all } i \notin I,$$

for some set $I \subseteq \{1, \cdots, m\}$. *Therefore, the number of faces of* P *is at most* 2^m.

Proof: Consider any valid inequality $\alpha^T x \leq \beta$. Suppose the corresponding face F is non-empty. Thus F are all optimal solutions to

$$\begin{array}{lll} & \text{Max } \ \alpha^T x & \text{subject to:} \\ (P) & \qquad Ax \leq b. & \end{array}$$

Choose an optimum solution y^* to the dual LP. By complemeentary slackness, the face F is defined by those elements x of P such that $a_i^T x = b_i$ for $i \in I = \{i : y_i^* > 0\}$. Thus F is defined by

$$\sum_j a_{ij} x_j = b_i \text{ for all } i \in I$$

$$\sum_j a_{ij}x_j \leq b_i \text{ for all } i \notin I.$$

As there are 2^m possibilities for F, there are at most 2^m faces. △

For extreme points (faces of dimension 0), the characterization is even stronger (we do not need the inequalities):

Theorem 6 *Let x^* be an extreme point for $P = \{x : Ax \leq b\}$. Then there exists I such that x^* is the unique solution to*

$$\sum_j a_{ij}x_j = b_i \textit{ for all } i \in I.$$

Proof: Given an extreme point x^*, define I by $I = \{i : \sum_j a_{ij}x_j^* = b_i\}$. This means that for $i \notin I$, we have $\sum_j a_{ij}x_j^* < b_i$.

From Theorem 5, we know that x^* is uniquely defined by

$$\sum_j a_{ij}x_j = b_i \text{ for all } i \in I \tag{3}$$

$$\sum_j a_{ij}x_j \leq b_i \text{ for all } i \notin I. \tag{4}$$

Now suppose there exists another solution $\hat{x}$ *when* we consider only the equalities for $i \in I$. Then because of $\sum_j a_{ij}x_j^* < b_i$, we get that $(1-\epsilon)x^* + \epsilon\hat{x}$ also satisfies (3) and (4) for ϵ sufficiently small. A contradiction. △

If P is given as $\{x : Ax = b, x \geq 0\}$ (as is often the case), the theorem still applies (as we still have a system of inequalities). In this case, the theorem says that every extreme point x^* can be obtained by setting some of the variables to 0, and solving for the unique solution to the resulting system of equalities. Without loss of generality, we can remove from $Ax = b$ equalities that are redundant; this means that we can assume that A has full row rank ($rank(A) = m$ for $A \in \mathbb{R}^{m \times n}$). Letting N denote the indices of the *non-basic* variables that we set of 0 and B denote the remaining indices (of the so-called *basic* variables), we can partition x^* into x_B^* and x_N^* (corresponding to these two sets of variables) and rewrite $Ax = b$ as $A_Bx_B + A_Nx_N = b$, where A_B and A_N are the restrictions of A to the indices in B and N respectively. The theorem says that x^* is the unique solution to $A_Bx_B + A_Nx_N = 0$ and $x_N = 0$, which means $x_N^* = 0$ and $A_Bx_B^* = b$. This latter system must have a unique solution, which means that A_B must have full column rank ($rank(A_B) = |B|$). As A itself has rank m, we have that $|B| \leq m$ and we can augment B to includee indices of N such that the resulting B satisfies (i) $|B| = m$ and (ii) A_B is a $m \times m$ invertible matrix (and thus there is still a unique solution to $A_Bx_B = b$). In linear programming terminology, a *basic feasible solution* or *bfs* of $\{x : Ax = b, x \geq 0\}$ is obtained by choosing a set $|B| = m$ of indices with A_B invertible and letting $x_B = A_B^{-1}b$ and $x_N = 0$ where N are the indices not in B. All extreme points are bfs and vice versa (although two different bases B may lead to the same extreme point, as there might be many ways of extending A_B into a $m \times m$ invertible matrix in the discussion above).

We now go back to the equivalence between Definitions 3 and 6 and claim that we can show that Definition 3 implies Definition 6 (this is a nice exercise using the Theorem of the Alternatives).

Theorem 7 *If $P = \{x : Ax \leq b\}$ is bounded then $P = conv(X)$ where X is the set of extreme points of P.*

One easy consequence of Theorem 5 is:

Corollary 8 *The maximal (inclusion-wise) non-trivial faces of a polyhedron P are the facets.*

When describing a polyhedron P in terms of linear inequalities, the only inequalities that are needed are the ones that define facets of P. This is stated in the next few theorems. We say that an inequality in the system $Ax \leq b$ is *redundant* if the corresponding polyhedron is unchanged by removing the inequality. For $P = \{x : Ax \leq b\}$, we let $I_=$ denote the indices i such that $a_i^T x = b_i$ for all $x \in P$, and $I_<$ the remaining ones (i.e. those for which there exists $x \in P$ with $a_i^T x < b_i$).

This theorem shows that facets are sufficient:

Theorem 9 *If face associated with $a_i^T x \leq b_i$ for $i \in I_<$ is not a facet then the inequality is redundant.*

And this one shows that facets are necessary:

Theorem 10 *If F is a facet of P then there must exists $i \in I_<$ such that the face induced by $a_i^T x \leq b_i$ is precisely F.*

In a *minimal* description of P, we must have a set of *linearly independent equalities* together with precisely one inequality for each facet of P.

4 Polyhedral Combinatorics

In one sentence, polyhedral combinatorics deals with the study of polyhedra or polytopes associated with discrete sets arising from combinatorial optimization problems (such as matchings for example). If we have a discrete set X (say the incidence vectors of matchings in a graph, or the set of incidence vectors of *stable* sets in a graph), we can consider $conv(X)$ and attempt to describe it in terms of linear inequalities. This is useful in order to apply the machinery of linear programming. However, in some (most) cases, it is actually hard to describe the set of all inequalities defining $conv(X)$; this occurs whenever optimizing over X is hard and this statement can be made precise in the setting of computational complexity. For matchings, or spanning trees, or several other structures, we will be able to describe their convex hull in terms of linear inequalities.

Given a set X and a proposed system of inequalities $P = \{x : Ax \le b\}$, it is usually easy to check whether $conv(X) \subseteq P$. Indeed, for this, we only need to check that every member of X satisfies every inequality in the description of P. The reverse inclusion is more difficult. Here are 3 general techniques to prove that $P \subseteq conv(X)$ (if it is true!) (once we know that $conv(X) \subseteq P$).

1. **Algorithmically.** This involves linear programming duality. This is what we did in the lecture on the assignment problem (minimum weight matchings in bipartite graphs). In general, consider any cost function c and consider the combinatorial optimization problem of maximizing $c^T x$ over $x \in X$. We know that:

$$\begin{aligned} \max\{c^T x : x \in X\} &= \max\{c^T x : x \in conv(X)\} \\ &\le \max\{c^T x : Ax \le b\} \\ &= \min\{b^T y : A^T y = c, y \ge 0\}, \end{aligned}$$

   the last equality coming from strong duality. If we know can exhibit a solution $x \in X$ (say a perfect matching in the assignment problem) and a dual feasibl solution y (values u_i, v_j in the assignment problem) such that $c^T x = b^T y$ we will have shown that we hav equality throughout, and if this is true for any cost function, this implies that $P = conv(X)$.

   This is usually the most involved approach but also the one that works most often.

2. **Focusing on extreme points.** Show that $P = \{x : Ax \le b\}$ is bounded (thus a polytope) and then study its extreme points. If we can show that every extreme point of P is in X then we would be done since $P = conv(ext(P)) \subseteq conv(X)$, where $ext(P)$ denotes the extreme points of P. The assumption that P is bounded is needed to show that indeed $P = conv(ext(P))$ (not true if P is unbounded).

   In the case of the convex hull of bipartite matchings, this can be done easily and this leads to the notion of Totally Unimodular Matrices (TUM), see the lecture notes on bipartite matchings.

3. **Focusing on the facets of $conv(X)$.** This leads usually to the shortest and cleanest proofs. Let us describe it just in the case in which $conv(X)$ is full-dimensional. Consider any valid inequality $\alpha^T x \le \beta$ for $conv(X)$ with $\alpha \ne 0$. We can assume that α is any vector in $\mathbb{R}^n \setminus \{0\}$ and that $\beta = \max\{\alpha^T x : x \in conv(X)\}$. The face of $conv(X)$ this inequality defines is $F = conv(\{x \in X : \alpha^T x = \beta\})$; it is a non-trivial face because of our assumptions on β and $\alpha \ne 0$. We need to make sure that if F is facet then we have in our description of P an inequality representing it (and it will thus have to be $\alpha^T x \le \beta$ or a multiple of it). What we will show is that, for any $\alpha \ne 0$ and the corresponding face F, we can find an inequality $a_i^T x \le b_i$ in our description of P such that $F \subseteq \{x : a_i^T x = b_i\}$, or simply that every optimum solution to $\max\{\alpha^T x : x \in X\}$ satisfies $a_i^T x = b_i$. This means that if F was a facet, by maximality, we have a representative of F in our description.

Example. Let $X = \{(\sigma(1), \sigma(2), \cdots, \sigma(n)) : \sigma$ is a permutation of $\{1, 2, \cdots, n\}\}$. We claim that

$$conv(X) = \{x \in \mathbb{R}^n : \quad \begin{array}{l} \sum_{i=1}^n x_i = \binom{n+1}{2} \\ \sum_{i \in S} x_i \geq \binom{|S|+1}{2} \quad S \subset \{1, \cdots, n\}\}. \end{array}$$

Here $conv(X)$ is not full-dimensional but the proof is similar; we only need to show that we are not missing any facets and any equality in the description of $conv(P)$. For the equalities, this can be seen easily. For the facets, suppose that $\alpha^T x \leq \beta$ defines a non-trivial facet F of $conv(X)$. Consider maximizing $\alpha^T x$ over all permutations x. Let $S = \arg\min\{\alpha_i\}$; by our assumption that F is non-trivial we have that $S \neq \{1, 2, \cdots, n\}$. Moreover, it is easy to see that any permutation that maximizes $\alpha^T x$ will need to satisfy $\alpha(i) \in \{1, 2, \cdots, |S|\}$ for $i \in S$, in other words, it will satisfy the inequality $\sum_{i \in S} x_i \leq \binom{|S|+1}{2}$ at equality, which was what we needed to prove. That's it!