3. Linear Programming and Polyhedral Combinatorics

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

Definition 3.1 *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 3.2 *A polyhedron is the intersection of finitely many halfspaces: $P = \{x \in \mathbb{R}^n : Ax \leq b\}$.*

Definition 3.3 *A polytope is a bounded polyhedron.*

Definition 3.4 *If P is a polyhedron in $\mathbb{R}^n$, the projection $P_k \subseteq \mathbb{R}^{n-1}$ of P is defined as $\{y = (x_1, x_2, \cdots, x_{k-1}, x_{k+1}, \cdots, x_n) : x \in P \text{ for some } x_k\}$.*

This is a special case of a projection onto a linear space (here, we consider only coordinate projection). By repeatedly projecting, we can eliminate any subset of coordinates.

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 do not involve x_k). Similarly, we can take a linear combination of an inequality in S_+ and one in S_- to eliminate the coefficient of x_k. This shows that the inequalities:

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

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), and therefore is a polyhedron.

Definition 3.5 *Given points $a^{(1)}, a^{(2)}, \cdots, a^{(k)} \in \mathbb{R}^n$,*

- *a linear combination is $\sum_i \lambda_i a^{(i)}$ where $\lambda_i \in \mathbb{R}$ for all i,*
- *an affine combination is $\sum_i \lambda_i a^{(i)}$ where $\lambda_i \in \mathbb{R}$ and $\sum_i \lambda_i = 1$,*
- *a conical combination is $\sum_i \lambda_i a^{(i)}$ where $\lambda_i \geq 0$ for all i,*
- *a convex combination is $\sum_i \lambda_i a^{(i)}$ where $\lambda_i \geq 0$ for all i and $\sum_i \lambda_i = 1$.*

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

The fact that Definition 3.6 implies Definition 3.3 can be shown by using Fourier-Motzkin elimination repeatedly on

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

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

3.1 Solvability of System of Inequalities

In linear algebra, we saw that, for $A \in R^{m \times n}$, $b \in R^m$, $Ax = b$ 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 3.1 (Theorem of the Alternatives) *$Ax \leq b$ 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 (as Fourier-Motzkin only performs nonnegative combinations of linear inequalities).

Another version of the above theorem is Farkas' lemma:

Lemma 3.2 *$Ax = b$, $x \geq 0$ has no solution if and only if there exists y with $A^T y \geq 0$ and $b^T y < 0$.*

Exercise 3-1. Prove Farkas' lemma from the Theorem of the Alternatives.

3.2 Linear Programming Basics

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

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

where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$ and the variables x are in $\mathbb{R}^n$. Any x satisfying $Ax \leq b$ is said to be *feasible*. If no x satisfies $Ax \leq b$, we say that the linear program is *infeasible*, and its optimum value is $-\infty$ (as we are maximizing over an empty set). If the objective function value of the linear program can be made arbitrarily large, we say that the linear program is *unbounded* and its optimum value is $+\infty$; otherwise it is *bounded*. If it is neither infeasible, not unbounded, then its optimum value is finite.

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:

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

(D) is the dual and (P) is the *primal*. The terminology for the dual is similar. If (D) has no feasible solution, it is said to be *infeasible* and its optimum value is $+\infty$ (as we are minimizing over an empty set). If (D) is unbounded (i.e. its value can be made arbitrarily negative) then its optimum value is $-\infty$.

The primal and dual spaces should not be confused. If A is $m \times n$ then we have n primal variables and m dual variables.

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 $Ax \leq b$. In particular, weak duality implies that if the primal is unbounded then the dual must be infeasible.

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 3.3 (Strong Duality) *Assume that (P) and (D) are feasible, and let z^* be the optimum value of the primal and w^* the optimum value of the dual. Then $z^* = w^*$.*

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) $c^T x \geq b^T y$. 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).

Proof: Let x^* be a feasible solution to the primal, and y^* be a feasible solution to the dual. The proof is by contradiction. Because of weak duality, this means that there are no solution $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ such that

$$\begin{cases} Ax & & \leq b \\ & A^T y & = c \\ & Iy & \leq 0 \\ -c^T x & +b^T y & \leq 0 \end{cases}$$

By a variant of the Theorem of the Alternatives or Farkas' lemma (for the case when we have a combination of inequalities and equalities), we derive that there must exist $s \in \mathbb{R}^m$, $t \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, $v \in \mathbb{R}$ such that:

$$\begin{aligned} & s \geq 0 \\ & u \geq 0 \\ & v \geq 0 \\ & A^T s - vc = 0 \\ & At - u + vb = 0 \\ & b^T s + c^T t < 0. \end{aligned}$$

We distinguish two cases.

Case 1: $v = 0$. Then s satisfies $s \geq 0$ and $A^T s = 0$. This means that, for any $\alpha \geq 0$, $y^* + \alpha s$ is feasible for the dual. Similarly, $At = u \geq 0$ and therefore, for any $\alpha \geq 0$, we have that $x^* - \alpha t$ is primal feasible. By weak duality, this means that, for any $\alpha \geq 0$, we have

$$c^T (x^* - \alpha t) \leq b^T (y^* + \alpha s)$$

or

$$c^T x^* - b^T y^* \leq \alpha (b^T s + c^T t).$$

The right-hand-side tend to $-\infty$ as α tends to ∞, and this is a contradiction as the left-hand-side is fixed.

Case 2: $v > 0$. By dividing throughout by v (and renaming all the variables), we get that there exists $s \geq 0$, $u \geq 0$ with

$$\begin{aligned} A^T s &= c \\ At - u &= -b \\ b^T s + c^T t &< 0. \end{aligned}$$

This means that s is dual feasible and $-t$ is primal feasible, and therefore by weak duality $c^T(-t) \leq b^T s$ contradicting $b^T s + c^T t < 0$. △

Exercise 3-2. Show that the dual of the dual is the primal.

Exercise 3-3. Show that we only need either the primal or the dual to be feasible for strong duality to hold. More precisely, if the primal is feasible but the dual is infeasible, prove that the primal will be unbounded, implying that $z^* = w^* = +\infty$.

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

Theorem 3.4 (Complementary Slackness) *If x is feasible in (P) and y is feasible in (D) then x is optimum in (P) and y is optimum 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 using the simplex method; this is not going to be explained in these notes. No variant of the simplex method is known to provably run in polynomial time, but there are other polynomial-time algorithms for linear programming, namely the ellipsoid algorithm and the class of interior-point algorithms.

3.3 Faces of Polyhedra

Definition 3.7 *$\{a^{(i)} \in \mathbb{R}^n : i \in K\}$ are linearly independent if $\sum_i \lambda_i a^{(i)} = 0$ implies that $\lambda_i = 0$ for all $i \in K$.*

Definition 3.8 *$\{a^{(i)} \in \mathbb{R}^n : i \in K\}$ are affinely independent if $\sum_i \lambda_i a^{(i)} = 0$ and $\sum_i \lambda_i = 0$ together imply that $\lambda_i = 0$ for all $i \in K$.*

Observe that $\{a^{(i)} \in \mathbb{R}^n : i \in K\}$ are *affinely* independent if and only if

$$\left\{ \begin{bmatrix} a^{(i)} \\ 1 \end{bmatrix} \in \mathbb{R}^{n+1} : i \in K \right\}$$

are *linearly* independent.

Definition 3.9 *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 affinely spans $\mathbb{R}^n$. In the latter case, we say that P is *full-dimensional*. The dimension of a cube in $\mathbb{R}^3$ is 3, and so is the dimension of $\mathbb{R}^3$ itself (which is a trivial polyhedron).

Definition 3.10 *$\alpha^T x \leq \beta$ 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, as this will imply that the inequality is also satisfied by points in $conv(S) \setminus S$. This observation will be important when dealing with convex hulls of combinatorial objects such as matchings or spanning trees.

Definition 3.11 *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*. Sometimes, one uses *ridges* for faces of dimension $dim(P) - 2$.

Exercise 3-4. List all 28 faces of the cube $P = \{x \in \mathbb{R}^3 : 0 \leq x_i \leq 1 \text{ for } i = 1, 2, 3\}$.

8+12+

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

Theorem 3.5 *Let $A \in \mathbb{R}^{m \times n}$. Then any non-empty face of $P = \{x \in \mathbb{R}^n : Ax \leq b\}$ corresponds to the set of solutions to*

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

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

for some set $I \subseteq \{1, \cdots, m\}$. Therefore, the number of non-empty 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 optimum solutions to

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

Choose an optimum solution y^* to the dual LP. By complementary 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 non-empty faces. △

The number of faces given in Theorem 3.5 is tight for polyhedra (see exercise below), but can be considerably improved for polytopes in the so-called *upper bound theorem*.

Exercise 3-5. Let $P = \{x \in \mathbb{R}^n : x_i \geq 0 \text{ for } i = 1, \cdots, n\}$. Show that P has $2^n + 1$ faces. How many faces of dimension k does P have?

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

Theorem 3.6 *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 3.5, 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 (as the face was supposed to contain a single point). △

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 $A_B x_B + A_N x_N = b$, 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 $A_B x_B + A_N x_N = 0$ and $x_N = 0$, which means $x^*_N = 0$ and $A_B x^*_B = b$. 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 include 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 $A_B x_B = b$). 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).

One consequence of Theorem 3.5 is:

Corollary 3.7 *The maximal (inclusion-wise) non-trivial faces of a non-empty polyhedron P are the facets.*

For the vertices, one needs one additional condition:

Corollary 3.8 *If $rank(A) = n$ (full column rank) then the minimal (inclusion-wise) non-trivial faces of a non-empty polyhedron $P = \{x \in \mathbb{R}^n : Ax \leq b\}$ are the vertices.*

Exercise 3-7 shows that the rank condition is necessary.

This means that, if a linear program $\max\{c^T x : x \in P\}$ with $P = \{x : Ax \leq b\}$ is feasible, bounded and $rank(A) = n$, then there exists an optimal solution which is a vertex of P (indeed, the set of all optimal solutions define a face — the optimal face — and if this face is not itself a vertex of P, it must contain vertices of P).

We now prove Corollary 3.8.

Proof: Let F be a minimal (inclusion-wise) non-trivial face of F. This means that we have a set I such that

$$F = \{x : \begin{array}{ll} a_i^T x = b_i & i \in I \\ a_j^T x \leq b_j & j \notin I\} \end{array}$$

and adding any element to I makes this set empty. Consider two cases. Either $P = \{x \in \mathbb{R}^n : a_i^T x = b_i \text{ for } i \in I\}$ or not. In the first case, it means that for every $j \notin I$ we have $a_j \in lin(\{a_i : i \in I\})$ and therefore since $rank(A) = n$ we have that the system $a_i^T x = b_i$ for $i \in I$ has a unique solution and thus F is a vertex.

On the other hand, if $P \neq \{x \in \mathbb{R}^n : a_i^T x = b_i \text{ for } i \in I\}$ then let $j \notin I$ such that there exists $\tilde{x}$ with

$$\begin{array}{ll} a_i^T \tilde{x} = b_i & i \in I \\ a_j^T \tilde{x} > b_j. & \end{array}$$

Since F is not trivial, there exists $\hat{x} \in F$, and in particular, $\hat{x}$ satisfies

$$\begin{array}{ll} a_i^T \hat{x} = b_i & i \in I \\ a_j^T \hat{x} \leq b_j. & \end{array}$$

Thus for a suitable convex combination x' of $\hat{x}$ and $\tilde{x}$, we have $a_i^T x' = b_i$ for $i \in I \cup \{j\}$, contrdicting the maximality of I. $\triangle$

We now go back to the equivalence between Definitions 3.3 and 3.6 and claim that we can show that Definition 3.3 implies Definition 3.6.

Theorem 3.9 *If $P = \{x : Ax \leq b\}$ is bounded then $P = conv(X)$ where X is the set of extreme points of P.*

This is a nice exercise using the Theorem of the Alternatives.
Proof: Since $X \subseteq P$, we have $conv(X) \subseteq P$. Assume, by contradiction, that we do not have equality. Then there must exist $\tilde{x} \in P \setminus conv(X)$. The fact that $\tilde{x} \notin conv(X)$ means that there is no solution to:

$$\begin{cases} \sum_{v \in X} \lambda_v v = \tilde{x} \\ \sum_{v \in X} \lambda_v = 1 \\ \lambda_v \geq 0 \qquad\qquad v \in X. \end{cases}$$

By the Theorem of the alternatives, this implies that $\exists c \in \mathbb{R}^n, t \in \mathbb{R}$:

$$\begin{cases} t + \sum_{j=1}^n c_j v_j \geq 0 \quad \forall v \in X \\ t + \sum_{j=1}^n c_j \tilde{x}_j < 0. \end{cases}$$

Since P is bounded, $\min\{c^T x : x \in P\}$ is finite (say equal to z^*), and the face induced by $c^T x \geq z^*$ is non-empty but does not contain any vertex (as all vertices are dominated by $\tilde{x}$ by the above inequalities). This is a contradiction with Corollary 3.8. $\triangle$

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 3.10 *If the 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 3.11 *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.

Exercises

Exercise 3-6. Prove Corollary 3.7.

Exercise 3-7. Show that if $rank(A) < n$ then $P = \{x \in \mathbb{R}^n : Ax \leq b\}$ has no vertices.

Exercise 3-8. Suppose $P = \{x \in \mathbb{R}^n : Ax \leq b, Cx \leq d\}$. Show that the set of vertices of $Q = \{x \in \mathbb{R}^n : Ax \leq b, Cx = d\}$ is a subset of the set of vertices of P.

(In particular, this means that if the vertices of P all belong to $\{0,1\}^n$, then so do the vertices of Q.)

Exercise 3-9. Given two extreme points a and b of a polyhedron P, we say that they are *adjacent* if the line segment between them forms an edge (i.e. a face of dimension 1) of the polyhedron P. This can be rephrased by saying that a and b are adjacent on P if and only if there exists a cost function c such that a and b are the only two extreme points of P minimizing $c^T x$ over P.

Consider the polyhedron (polytope) P defined as the convex hull of all perfect matchings in a (not necessarily bipartite) graph G. Give a necessary and sufficient condition for two matchings M_1 and M_2 to be adjacent on this polyhedron (hint: think about $M_1 \triangle M_2 = (M_1 \setminus M_2) \cup (M_2 \setminus M_1)$) and prove that your condition is necessary and sufficient.)

Exercise 3-10. Show that two vertices u and v of a polyhedron P are adjacent if and only there is a unique way to express their midpoint ($\frac{1}{2}(u+v)$) as a convex combination of vertices of P.

3.4 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 spanning trees of a graph, or the set of incidence vectors of *stable* sets[1] 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 (for which the corresponding optimization problem is polynomially solvable), we will be able to describe their convex hull in terms of linear inequalities.

Given a set X and a proposed system of inequalities $P = \{x : Ax \leq b\}$, 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.

[1] A set S of vertices in a graph $G = (V, E)$ is stable if there are no edges between any two vertices of S.

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 notes about 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:

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

   $\because$ $conv(X) \subseteq P$

   the last equality coming from strong duality. If we can exhibit a solution $x \in X$ (say a perfect matching in the assignment problem) and a dual feasible 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 have 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 first that $P = \{x : Ax \leq b\}$ 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 (see Theorem 3.9). 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 next section.

3. **Focusing on the facets of $conv(X)$.** This leads usually to the shortest and cleanest proofs. Suppose that our proposed P is of the form $\{x \in \mathbb{R}^n : Ax \leq b, Cx = d\}$. We have already argued that $conv(X) \subseteq P$ and we want to show that $P \subseteq conv(X)$.

   First we need to show that we are not missing any equality. This can be done for example by showing that $dim(conv(X)) = dim(P)$ (i.e. showing that if there are $n - d$ linearly independent rows in C we can find $d + 1$ affinely independent points in X).

   Then we need to show that we are not missing a valid inequality that induces a *facet* of $conv(X)$. Consider any valid inequality $\alpha^T x \leq \beta$ for $conv(X)$ with $\alpha \neq 0$. 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 $F = conv(\{x \in X : \alpha^T x = \beta\})$. Assume that this is a non-trivial face; this will happen precisely when α is not in the row space of C. We need to make sure that if F is facet then we have in our description of P an inequality representing it. What we will show is that if F is non-trivial then we can find an inequality $a_i^T x \leq b_i$ 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.

This is a very simple and powerful technique, and this is best illustrated on an example.

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

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

This is known as the *permutahedron*.

Here $conv(X)$ is not full-dimensional; 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 as it is easy to exhibit n affinely independent permutations in X. 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 $S \neq \{1, 2, \cdots, n\}$ (otherwise, we would have the equality $\sum_{i=1}^n x_i = \binom{n+1}{2}$). Moreover, it is easy to see (by an exchange argument) that any permutationσ whose incidence vector x maximizes $\alpha^T x$ will need to satisfy $\sigma(i) \in \{1, 2, \cdots, |S|\}$ for $i \in S$, in other words, it will satisfy the inequality $\sum_{i \in S} x_i \geq \binom{|S|+1}{2}$ at equality. Hence, F is contained in the face corresponding to an inequality in our description, and hence our description contains inequalities for all facets. This is what we needed to prove. That's it!

Exercises

Exercise 3-11. Consider the set $X = \{(\sigma(1), \sigma(2), \cdots, \sigma(n)) : \sigma$ is a permutation of $\{1, 2 \cdots, n\}\}$. Show that $\dim(conv(X)) = n - 1$. (To show that $\dim(conv(X)) \geq n - 1$, exhibit n affinely independent permutations σ (and prove that they are affinely independent).)

Exercise 3-12. A *stable set* S (sometimes, it is called also an independent set) in a graph $G = (V, E)$ is a set of vertices such that there are no edges between any two vertices in S. If we let P denote the convex hull of all (incidence vectors of) stable sets of $G = (V, E)$, it is clear that $x_i + x_j \leq 1$ for any edge $(i, j) \in E$ is a valid inequality for P.

1.Give a graph G for which P is *not* equal to

$$\{x \in \mathbb{R}^{|V|} : \begin{array}{ll} x_i + x_j \leq 1 & \text{for all } (i, j) \in E \\ x_i \geq 0 & \text{for all } i \in V\} \end{array}$$

2.Show that if the graph G is bipartite then P equals

$$\{x \in \mathbb{R}^{|V|} : \begin{array}{ll} x_i + x_j \leq 1 & \text{for all } (i, j) \in E \\ x_i \geq 0 & \text{for all } i \in V\}. \end{array}$$

Exercise 3-13. Let $e_k \in \mathbb{R}^n$ $(k = 0, \ldots, n-1)$ be a vector with the first k entries being 1, and the following $n-k$ entries being -1. Let $S = \{e_0, e_1, \ldots, e_{n-1}, -e_0, -e_1, \ldots, -e_{n-1}\}$, i.e. S consists of all vectors consisting of $+1$ followed by -1 or vice versa. In this problem set, you will study *conv*(S).

1. Consider any vector $a \in \{-1, 0, 1\}^n$ such that (i) $\sum_{i=1}^n a_i = 1$ and (ii) for all $j = 1, \ldots, n-1$, we have $0 \leq \sum_{i=1}^j a_i \leq 1$. (For example, for $n = 5$, the vector $(1, 0, -1, 1, 0)$ satisfies these conditions.) Show that $\sum_{i=1}^n a_i x_i \leq 1$ and $\sum_{i=1}^n a_i x_i \geq -1$ are valid inequalities for *conv*(S).
2. How many such inequalities are there?
3. Show that any such inequality defines a facet of *conv*(S).

   (This can be done in several ways. Here is one approach, but you are welcome to use any other one as well. First show that either e_k or $-e_k$ satisfies this inequality at equality, for any k. Then show that the resulting set of vectors on the hyperplane are affinely independent (or uniquely identifies it).)
4. Show that the above inequalities define the entire convex hull of S.

   (Again this can be done in several ways. One possibility is to consider the 3rd technique described above.)

3.5 Total unimodularity

Definition 3.12 *A matrix A is totally unimodular if every square submatrix of A has determinant $-1, 0$ or $+1$.*

The importance of total unimodularity stems from the following theorem. This theorem gives a subclass of integer programs which are easily solved. A polyhedron P is said to be *integral* if all its vertices or extreme points are integral (belong to $\mathbb{Z}^n$).

Theorem 3.12 *Let A be a totally unimodular matrix. Then, for any integral right-hand-side b, the polyhedron*

$$P = \{x : Ax \leq b, x \geq 0\}$$

is integral.

Before we prove this result, two remarks can be made. First, the proof below will in fact show that the same result holds for the polyhedrons $\{x : Ax \geq b, x \geq 0\}$ or $\{x : Ax = b, x \geq 0\}$. In the latter case, though, a slightly weaker condition than totally unimodularity is sufficient to prove the result. Secondly, in the above theorem, one can prove the converse as well: If $P = \{x : Ax \leq b, x \geq 0\}$ is integral for all integral b then A must be totally unimodular (this is not true though, if we consider for example $\{x : Ax = b, x \geq 0\}$).

Proof: Adding slacks, we get the polyhedron $Q = \{(x, s) : Ax + Is = b, x \geq 0, s \geq 0\}$. One can easily show (see exercise below) that P is integral iff Q is integral.

Consider now any bfs of Q. The basis B consists of some columns of A as well as some columns of the identity matrix I. Since the columns of I have only one nonzero entry per column, namely a one, we can expand the determinant of A_B along these entries and derive that, in absolute values, the determinant of A_B is equal to the determinant of some square submatrix of A. By definition of totally unimodularity, this implies that the determinant of A_B must belong to $\{-1, 0, 1\}$. By definition of a basis, it cannot be equal to 0. Hence, it must be equal to ± 1.

We now prove that the bfs must be integral. The non-basic variables, by definition, must have value zero. The vector of basic variables, on the other hand, is equal to $A_B^{-1}b$. From linear algebra, A_B^{-1} can be expressed as

$$\frac{1}{\det A_B} A_B^{adj}$$

where A_B^{adj} is the adjoint (or adjugate) matrix of A_B and consists of subdeterminants of A_B. Hence, both b and A_B^{adj} are integral which implies that $A_B^{-1}b$ is integral since $|\det A_B| = 1$. This proves the integrality of the bfs. △

Exercise 3-14. Let $P = \{x : Ax \leq b, x \geq 0\}$ and let $Q = \{(x, s) : Ax + Is = b, x \geq 0, s \geq 0\}$. Show that x is an extreme point of P iff $(x, b - Ax)$ is an extreme point of Q. Conclude that whenever A and b have only integral entries, P is integral iff Q is integral.

In the case of the bipartite matching problem, the constraint matrix A has a very special structure and we show below that it is totally unimodular. This alongs with Theorem 3.12 proves Theorem 1.6 from the notes on the bipartite matching problem. First, let us restate the setting. Suppose that the bipartition of our bipartite graph is (U, V) (to avoid any confusion with the matrix A or the basis B). Consider

$$\begin{aligned}
P \quad &= \quad \{x: \quad \sum_j x_{ij} = 1 && i \in U \\
& \qquad\qquad\quad \sum_i x_{ij} = 1 && j \in V \\
& \qquad\qquad\quad x_{ij} \geq 0 && i \in U, j \in V\} \\
&= \quad \{x: \quad Ax = b, x \geq 0\}.
\end{aligned}$$

Theorem 3.13 *The matrix A is totally unimodular.*

The way we defined the matrix A corresponds to a *complete* bipartite graph. If we were to consider any bipartite graph then we would simply consider a submatrix of A, which is also totally unimodular by definition.

Proof: Consider any square submatrix T of A. We consider three cases. First, if T has a column or a row with all entries equal to zero then the determinant is zero. Secondly, if there exists a column or a row of T with only one $+1$ then by expanding the determinant along that $+1$, we can consider a smaller sized matrix T. The last case is when T has at least two nonzero entries per column (and per row). Given the special structure of A, there must in fact be *exactly* 2 nonzero entries per column. By adding up the rows of T corresponding to the vertices of U and adding up the rows of T corresponding to the vertices of V, one therefore obtains the same vector which proves that the rows of T are linearly dependent, implying that its determinant is zero. This proves the totally unimodularity of A. $\triangle$

We conclude with a technical remark. One should first remove one of the rows of A before applying Theorem 3.12 since, as such, it does not have full row rank and this fact was implicitly used in the definition of a bfs. However, deleting a row of A still preserves its totally unimodularity.

Exercise 3-15. If A is totally unimodular then A^T is totally unimodular.

Exercise 3-16. Use total unimodularity to prove König's theorem.

The following theorem gives a necessary and sufficient condition for a matrix to be totally unimodular.

Theorem 3.14 *Let A be a $m \times n$ matrix with entries in $\{-1, 0, 1\}$. Then A is TUM if and only if for all subsets $R \subseteq \{1, 2, \cdots, n\}$ of rows, there exists a partition of R into R_1 and R_2 such that for all $j \in \{1, 2, \cdots, m\}$:*

$$\sum_{i \in R_1} a_{ij} - \sum_{i \in R_2} a_{ij} \in \{0, 1, -1\}.$$

We will prove only the *if* direction.

Proof: Assume that, for every R, the desired partition exists. We need to prove that the determinant of any $k \times k$ submatrix of A is in $\{-1, 0, 1\}$, and this must be true for any k. Let us prove it by induction on k. It is trivially true for $k = 1$. Assume it is true for $k - 1$ and we will prove it for k.

Let B be a $k \times k$ submatrix of A, and we can assume that B is invertible (otherwise the determinant is 0 and there is nothing to prove). The inverse B^{-1} can be written as $\frac{1}{\det(B)} B^*$, where all entries of B^* correspond to $(k-1) \times (k-1)$ submatrices of A. By our inductive hypothesis, all entries of B^* are in $\{-1, 0, 1\}$. Let b_1^* be the first row of B and e_1 be the k-dimensional row vector $[1\ 0\ 0 \cdots 0]$, thus $b_1^* = e_1 B^*$. By the relationship between B and B^*, we have that

$$b_1^* B = e_1 B^* B = \det(B) e_1 B^{-1} B = \det(B) e_1. \tag{5}$$

Let $R = \{i : b_{1i}^* \in \{-1, 1\}\}$. By assumption, we know that there exists a partition of R into R_1 and R_2 such that for all j:

$$\sum_{i \in R_1} b_{ij} - \sum_{i \in R_2} b_{ij} \in \{-1, 0, 1\}. \tag{6}$$

From (5), we have that

$$\sum_{i \in R} b^*_{1i} b_{ij} = \begin{cases} \det(B) & j = 1 \\ 0 & j \neq 1 \end{cases} \tag{7}$$

Since the left-hand-sides of equations (6) and (7) differ by a multiple of 2 for each j (since $b^*_{1i} \in \{-1, 1\}$), this implies that

$$\sum_{i \in R_1} b_{ij} - \sum_{i \in R_2} b_{ij} \begin{cases} = 0 & j \neq 1 \\ \in \{-1, 1\} & j = 1 \end{cases} \tag{8}$$

The fact that we could not get 0 for $j = 1$ follows from the fact that otherwise B would be singular (we would get exactly the 0 vector by adding and subtracting rows of B). If we define $y \in \mathbb{R}^k$ by

$$y_i = \begin{cases} 1 & i \in R_1 \\ -1 & i \in R_2 \\ 0 & \textit{otherwise} \end{cases}$$

we get that $yB = \pm e_1$. Thus

$$y = \pm e_1 B^{-1} = \pm \frac{1}{\det B} e_1 B^* = \pm \frac{1}{\det B} b^*_1,$$

which implies that $\det B$ must be either 1 or -1. △

Exercise 3-17. Suppose we have n activities to choose from. Activity i starts at time t_i and ends at time u_i (or more precisely just before u_i); if chosen, activity i gives us a profit of p_i units. Our goal is to choose a subset of the activities which do not overlap (nevertheless, we can choose an activity that ends at t and one that starts at the same time t) and such that the total profit (i.e. sum of profits) of the selected activities is maximum.

1. Defining x_i as a variable that represents whether activity i is selected ($x_i = 1$) or not ($x_i = 0$), write an integer program of the form $\max\{p^T x : Ax \leq b, x \in \{0,1\}^n\}$ that would solve this problem.
2. Show that the matrix A is totally unimodular, implying that one can solve this problem by solving the linear program $\max\{p^T x : Ax \leq b, 0 \leq x_i \leq 1 \text{ for every } i\}$.