CO 255 (Winter 2014 - 1141)

Advanced Introduction to Optimization

Prof. W. Cook
University of Waterloo

LaTeXer: W. Kong
http://wwkong.github.io
Last Revision: April 30, 2014

Table of Contents

These notes are currently a work in progress, and as such may be incomplete or contain errors.

Acknowledgments:

Special thanks to *Michael Baker* and his LaTeX formatted notes. They were the inspiration for the structure of these notes.

Abstract

The purpose of these notes is to provide the reader with a secondary reference to the material covered in CO 255. The formal prerequisite to this course is one of MATH 235, MATH 245, MATH 237, or MATH 247 in which this author recommends either MATH 245 or MATH 247 before enrolling in this course. Readers should be confident in dealing with mathematical rigour as some of the proofs in these notes will need details to be filled in.

Errata

Email: bico@uwaterloo.ca

Office: MC 6314 @ 1-3pm Wednesdays

50% Biweekly Assignments, 50% Final Exam (No midterm) which will be slighted modified from material in class

External sources and help from peers should be cited in all assignments.

1 Introduction

We start with the basic framework. Also, take note that for the rest of this course, only rational numbers will be used.

1.1 Gaussian Elimination

With Gaussian elimination, we either find a solution or we end up with 0 = !0. In general, the system of equations with a vector y being multiplied into the system can be put into the form (a linear combination):

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n &= b_1 | y_1 \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n &= b_2 | y_2 \\ \vdots \quad \vdots &\quad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n &= b_m | y_m \end{aligned}$$

and alternatively

$$\begin{aligned} (a_{11}y_1 + a_{21}y_2 + ... + a_{m1}y_m)x_1 &\quad + \\ (a_{12}y_1 + a_{22}y_2 + ... + a_{m2}y_m)x_2 &\quad + \\ \vdots \quad \vdots &\quad \vdots \\ (a_{1n}y_1 + a_{2n}y_2 + ... + a_{mn}y_m)x_n &\quad = \\ y_1b_1 + y_2b_2 + ... + y_mb_m & \end{aligned}$$

Under matrix notation, though,

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}, x = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}$$

the form of the linear combination is $Ax = b$ and letting $y = \begin{bmatrix} y_1 & \cdots & y_m \end{bmatrix}$, the subsequent form is $(y^T A)x = y^T b$. With Gaussian elimination, we either find

1. $\bar{x}$ such that $A\bar{x} = b$
2. $\bar{y}$ such that $y^T A = 0$ and $y^T b \neq 0$

Lemma 1.1. *Each equation in obtained in Gaussian Elimination is a linear combination of the original equations.*

Proof. Obvious from above. □

Lemma 1.2. *(3) $Ax = b$ has no solution.*

Proof. Suppose that $\bar{x}$ is a solution. Multiplying by $\bar{y}$ we have $0 = \bar{y}^T A\bar{x} = \bar{y}^T b \neq 0$ which is a contradiction. □

Theorem 1.1. *$Ax = b$ has a solution if and only if $y^T b = 0$ for each vector y such that $y^T A = 0$.*

Proof. Apply Gaussian elimination. Exactly one of the two systems has a solution (i) $y^T A = 0$ (ii) $y^T A = 0, y^T b = 1$. □

Problem 1.1. Is Gaussian elimination a polynomial time algorithm?

If elementary operations of adding, comparing, multiplying and dividing are considered, then the algorithm is approximately $O(n^3)$.

Note 1. There are two types of Gaussian elimination considered in class, those with real coefficient multipliers (1) and those with integral multipliers applied not only on the eliminated row but also the base row (2).

Example 1.1. Consider the system:

$$\begin{bmatrix} 2 & 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 \\ 1 & 1 & 2 & 0 & 0 \\ 1 & 1 & 1 & 2 & 0 \\ 1 & 1 & 1 & 1 & 2 \end{bmatrix}$$

To reduce the n^{th} row in the general case we need to multiply by 2^{2^n} and so in the 1000×1000 case, we need $2^{2^{1000}} - 2^{1000}$ binary digits.

So here we make the distinction between the running time number, which is related to the elementary bit operations, and the size of the problem, which is the number of bits needed to write it down.

1.2 Complexity of Algorithms

We define a **decision problem** as a problem with a yes or no answer. In our case, we have the problem "Does $Ax = b$ have a solution?". Formally, if we consider a universe Σ^* with some $L \subseteq \Sigma^*$ then we could rephrase a decision problem as "Does ϕ given $l \in \Sigma^*$ belong to L?". We identify this problem with L. In our case

$$\Sigma^* = \{(A, b) : A \text{ a matrix}, b \text{ a vector}\}$$

and

$$L = \{(A, b) : A\bar{x} = b \text{ for some } \bar{x}\}$$

An **algorithm** is a list of instructions to solve a problem. A **polynomial time algorithm** is an algorithm in which there is some polynomial f such that for a problem of size σ the running time of the algorithm is at most $f(\sigma)$. We will often say the running time is $O(g(\sigma))$ for some function g if there exists a constant k such that the running time is bounded by $kg(\sigma)$.

The class of decision problems solvable in polynomial time is denoted by **P**. The class of decision problems L such that "for any $l \in L$, the fact l is in L has a proof of length polynomially bounded by the size of L" is denoted by **NP**. The class **co-NP** is the set of decision problems in which the reply 'no answer' can be certified in polynomial time.

An **NP-hard** problem is one in which a solution of this problem in polynomial time implies that NP=P. An example is the *traveling salesman problem*. An **NP-complete** problem is one which is both NP-hard and NP.

Note 2. The size of an integer n is $1 + \lceil \log_2 |n| + 1 \rceil$ and the size of a rational $r = p/q$, with p and q relatively prime, is $1 + \lceil \log_2 |p| \rceil + \lceil \log_2 |q| \rceil$. So the size of $Ax = b$ is the sum of the sizes of the rationals in A and b.

Going back, consider if $Ax = b$ has a solution. If yes, then display $\bar{x}$ such that $A\bar{x} = b$ and if no, display $\bar{y}$ such that $\bar{y}^T A = 0, \bar{y}^T b = 1$. Are $\bar{x}, \bar{y}$ of polynomial size?

Lemma 1.3. *Let M be a rational matrix. Then $\det M$ has size at most twice the size of M.*

Proof. Let $M = \left[\frac{p_{ij}}{q_{ij}}\right]$ and M has n rows. Suppose that $|\det M| = p/q$ where p and q are relatively prime. We first know that $|c(q)| \leq |c(M)|$. To see this, note that

$$q = \prod_{i,j} q_{i,j} < 2^{|c(M)|-1} \implies |c(q)| \leq \sum_{i,j} |c(q_{i,j})| < |c(M)|$$

where $c()$ is the encoding function. A similar result holds for p. To see this, note that $\det M$ is an alternating sum over all permutations, so

$$\begin{aligned} |\det M| = \sum_{\pi \in S_n} \text{sgn}(\pi) \cdot \prod_{k=1}^{n} M_{k,\pi(k)} \leq \prod_{i,j}(|p_{ij}|+1|) \quad &\implies \quad |p| = |\det M| \cdot q \leq \prod_{i,j}(|p_{ij}|+1|)q_{ij} < 2^{|c(M)|-1} \\ &\implies \quad |c(p)| < |c(M)| \end{aligned}$$

and hence

$$|c(\det M)| = 1 + |c(p)| + |c(q)| < 2|c(M)|$$

□

Theorem 1.2. *If a rational system $Ax = b$ has a solution then it has one of size polynomially bounded by the size of $A|b$.*

Proof. We may assume rows of A are linearly independent By reordering the columns, we may write $A = [B\ N]$ where B is non-singular and called **basic** and N is **non-basic**. Then $\bar{x} = \begin{pmatrix} B^{-1}b \\ 0 \end{pmatrix}$ is a solution of $Ax = b$. Under Cramer's Rule,

$$B^{-1} = \left[\frac{(-1)^{j+i}\det(B_{ij})}{\det B}\right]$$

and from the above lemma, $\bar{x}$ is of polynomial size. □

Corollary 1.1. *The problem 'Does $Ax = b$ have a solution?' is in NP $\cap$ co-NP.*

Theorem 1.3. *(Edmonds 1967) If A and b are rational then Gaussian elimination is polynomial time.*

Proof. It suffices to show that all numbers that appear are of size polynomially bounded in the size of (A, b). During the execution of the algorithm, we find linear systems $A_k x = b_k$ where $0 \leq k \leq r$ and r is the rank of A. Consider this as working on matrices $E_k = [A_k|b_k]$. We may assume we need not permute any columns. We show all numbers in $(E_k : k = 0, ..., r)$ are of polynomial size by induction on k. The case of $k = 0$ is trivial since $A_0 = A$ and $b_0 = b$ and the result follows from the above theorem. Let $0 < k \leq r$ and suppose the sizes of $E_0, ..., E_{k-1}$ are polynomial in the size of $(A|b)$.

The matrix E_k is of the form $\begin{pmatrix} B & C \\ 0 & D \end{pmatrix}$ where B is non-singular and upper triangular with k rows and k columns. The first k rows of E_k and E_{k-1} are identical. It remains to show the entries in D are small. Consider the entry d_{ij} of D. Let $(E_k)_{ij} = \begin{pmatrix} B & C \\ 0 & d_{ij} \end{pmatrix}$ and note that $|\det((E_k)_{IJ})| = |d_{ij} \det B|$ and hence

$$d_{ij} = \frac{\det(E_k)_{IJ}}{\det B} = \frac{\det(E_k)_{IJ}}{\det(E_k)_{KK}}$$

Now E_k arises from $(A|b)$ by adding multiples of the first k rows to other rows so $\det(E_k)_{IJ} = \det(A|B)_{IJ}$ and $\det(E_k)_{KK} = \det(A|b)_{KK}$ □

1.3 Intro to Integer Programming

Problem 1.2. Does $Ax = b$ with x as an integer have a solution?

To solve this, we first define a few elementary operations:

1. Exchanging two columns
2. Multiply a column by -1 (changing the sign)
3. Adding an integral multiple of one column to another column
   (a) To see this, suppose that a_1 and a_2 are two columns. Replace a_1 by $a_1 + \delta a_2$ and we see that
   $$a_1x_1 + a_2x_2 = (a_1 + \delta a_2)x_1 + a_2(x_2 - \delta x_1)$$
   which is an integer

Example 1.2. Consider the system

$$\begin{pmatrix} 2 & 1 & 4 & 6 \\ 7 & 2 & 5 & 6 \\ 8 & 3 & 10 & 33 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 1 \end{pmatrix}$$

The 'A' matrix is reduced as follows:

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 3 & 2 & -3 & -7 \\ 2 & 3 & -2 & 15 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 2 & -3 & -7 \\ 3 & 2 & -2 & 15 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 2 & 0 & -1 \\ 3 & 2 & 0 & 19 \end{pmatrix}$$

$$\to \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 3 & 0 \\ 3 & -19 & 2 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & -19 & 59 & 0 \end{pmatrix}$$

So we have $x_1 = 1$, $x_2 = 3 - 2x_1 = 1$ and $59x_3 = 1 + 19x_2 - 3x_1 \implies x_3 = 17/59$. To observe why we could not get an integer solution, first compute B^{-1} as the inverse of

$$B = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & -19 & 59 \end{pmatrix}$$

which is

$$B^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -\frac{41}{59} & \frac{19}{59} & \frac{1}{59} \end{pmatrix}$$

If $y = B_3^{-1}$, the third row of B^{-1}, then

$$y^TA = \begin{pmatrix} -\frac{41}{59} & \frac{19}{59} & \frac{1}{59} \end{pmatrix} \begin{pmatrix} 2 & 1 & 4 & 6 \\ 7 & 2 & 5 & 6 \\ 8 & 3 & 10 & 33 \end{pmatrix} = \begin{pmatrix} 1 & 0 & -1 & -2 \end{pmatrix}$$

and

$$y^Tb = \begin{pmatrix} -\frac{41}{59} & \frac{19}{59} & \frac{1}{59} \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ 1 \end{pmatrix} = \frac{17}{59}$$

Similar to Gaussian elimination, we must have $y^TA\bar{x} = y^Tb$ and since the LHS is integral and the RHS is not, the two are not equivalent and the system has no solutions.

Definition 1.1. Let A be a rational m by n matrix of rank m. A is in **Hermite normal form** (HNF) if $A = \begin{bmatrix} B & 0 \end{bmatrix}$ where B is lower triangular and the diagonal entry is the unique maximum in each row and all entries are non-negative.

Theorem 1.4. *Any rational matrix M of full row rank can be converted to Hermite normal formed by sequences of (1),(2),(3), defined at the beginning of this section.*

Proof. If suffices to consider integral matrices M (by scaling). At a general step, we have the matrix in the form

$$\begin{pmatrix} B & 0 \\ C & D \end{pmatrix}$$

where B is lower triangular with a positive diagonal. We modify D such that

[1] $d_{11}, ..., d_{1k}$ are non-negative using operation (2)

[2] As long as possible, use operation (3) to reduce $d_{11} + ... + d_{1k}$

[3] $d_{11} \geq d_{12} \geq ... \geq d_{1k}$ by operation (1)

Since M has full row rank, $d_{11} > 0$. Now by (2), we must have $d_{12} = d_{13} = ... = d_{1k} = 0$. Repeating this, we arrive at $\begin{bmatrix} B & 0 \end{bmatrix}$. We can make B non-negative with unique maximum element on the diagonal. □

Theorem 1.5. *The Hermite normal form of a matrix A is unique.*

Definition 1.2. A non-singular matrix U is called **unimodular** if it is integral and $\det U = \pm 1$.

Note 3. If A is of full row rank, then there exists a unimodular matrix U such that AU is HNF.

Theorem 1.6. *(Integer Farkas Lemma) Let $Ax = b$ be a rational system. Then there exists an integral solution x if and only if $y^T b$ is integer for each rational y such that $y^T A$ is integral.*

Proof. ($\Longrightarrow$) If $A\bar{x} = b$, then $y^T A\bar{x} = y^T b$. If $y^T A$ and $\bar{x}$ are integral, then $y^T b$ is an integer.

($\Longleftarrow$) Suppose that $y^T b$ is integer for all y such that $y^T A$ is integral. Then $\nexists y$ such that $y^T A = 0$ and $y^T b$ is non-integer. By the theorem for linear equations, this implies that $Ax = b$ has a solution. Assume that A has full row rank. We may apply (1), (2), (3) to reduce A to HNF with $A = \begin{bmatrix} B & 0 \end{bmatrix}$. Note that $B^{-1} \begin{bmatrix} B & 0 \end{bmatrix} = \begin{bmatrix} I & 0 \end{bmatrix}$ and so if $y = (B^{-1})_j$ for any j^{th} row of B, then $y^T B$ is integral. Thus, $y^T b$ is integer for each row of B^{-1}. Thus, $B^{-1} b$ is integral. But

$$\begin{bmatrix} B & 0 \end{bmatrix} \begin{bmatrix} B^{-1} b \\ 0 \end{bmatrix} = b$$

and so $\begin{bmatrix} B^{-1} b & 0 \end{bmatrix}^T$ is an integral solution to the system. □

We can see that if $A\bar{x} = b$ does have a solution, we can display $\bar{x}$ and verify (NP), if not, we have to show y such that $y^T A$ is integral and $y^T b$ and is non-integer (co-NP). It is possible to show that $\exists \bar{x}, y$ of size polynomial in the size of A and b. There also exists a polynomial time algorithm to compute HNF.

Note 4. The above theorem can be reformulated as there exists x such that $Ax = b$ with x integral $\iff$ there is no y such that $y^T A$ is integral and $y^T b$ is non-integral.

2 Systems of Inequalities

The system of equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n &\leq b_1 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n &\leq b_m \end{aligned}$$

can be written (informally) as $Ax \leq b$.

Problem 2.1. Does $Ax \leq b$ have a solution? (**Linear programming**) [Dantzig]

Note that $Ax \leq b, -Ax \leq -b, -Ix \leq 0 \iff Ax = b, x \geq 0$.

Example 2.1. Consider the system of equations:

$$\begin{array}{cccrr} -x_1 & -x_2 & +x_3 & \leq & -2 \\ 2x_1 & -x_2 & -x_3 & \leq & -4 \\ -x_1 & -x_2 & & \leq & 1 \\ & -x_2 & -x_3 & \leq & -2 \end{array}$$

We then scale to make the coefficients of x_1 either $0, 1,$ or -1 and reorder the inequalities:

$$\begin{array}{cccrr} x_1 & -\frac{1}{2}x_2 & -\frac{1}{2}x_3 & \leq & -2 \\ -x_1 & -x_2 & +x_3 & \leq & -2 \\ -x_1 & -x_2 & & \leq & 1 \\ & -x_2 & -x_3 & \leq & -2 \end{array} \implies \begin{array}{cccccccc} & & & & x_1 & \leq & -2 & +\frac{1}{2}x_2 \quad +\frac{1}{2}x_3 \\ 2 & -x_2 & x_3 & \leq & x_1 & & & \\ -1 & x_2 & & \leq & x_1 & & & \\ & -x_2 & -x_3 & \leq & & & -2 & \end{array}$$

If (x_1, x_2, x_3) is a solution, then (x_2, x_3) satisfies

$$\begin{array}{cccccccc} 2 & -x_2 & +x_3 & \leq & -2 & +\frac{1}{2}x_2 & +\frac{1}{2}x_3 \\ -1 & x_2 & & \leq & -2 & +\frac{1}{2}x_2 & +\frac{1}{2}x_3 \\ & -x_2 & -x_3 & \leq & -2 & & \end{array}$$

Furthermore, if (x_2, x_3) satisfies the reduced system, then we can find a value for x_1 such that (x_1, x_2, x_3) satisfies the original system $\implies$ x_1 has been eliminated (*Fourier-Motzkin Elimination*). The next iteration gives us

$$\begin{array}{cccc} \frac{8}{3} & +\frac{1}{3}x_3 & \leq & x_2 \\ \frac{2}{3} & -\frac{1}{3}x_3 & \leq & x_2 \\ 2 & -x_3 & \leq & x_2 \end{array}$$

and this is trivial solve since eliminating x_2 gives the empty system. In particular, suppose we set $x_3 = 0$. Then we choose x_2 such that

$$\begin{array}{ccc} \frac{8}{3} & \leq & x_2 \\ \frac{2}{3} & \leq & x_2 \\ 2 & \leq & x_2 \end{array}$$

Let $x_2 = 4$ and we must choose x_1 such that $x_1 \leq 0, -2 \leq x_1, -5 \leq x_1$. We choose $x_1 = -1$. Thus, we have a solution, $(-1, 4, 0)$, to the original system.

Remark 2.1. The only inequality that has no solutions is

$$0x_1 + 0x_2 + ... + 0x_n \leq t$$

where $t < 0$. That is, $0^T x \leq t$ which we call an **infeasible inequality**. The only way that Fourier-Motzkin (F-M) elimination fails is if it produces a infeasible inequality.

Note 5. Every inequality we produce in F-M is the sum of positive multiples of inequalities in $Ax \leq b$. That is, $(y^T A)x \leq y^T b$ with $y \geq 0$ which are non-negative linear combinations of $Ax \leq b$. If $\bar{x}$ is a solution to $Ax \leq b$ then $(y^T A)\bar{x} \leq y^T b$.

Also, if an infeasible inequality is a non-negative linear combination of $Ax \leq b$, then $Ax \leq b$ has no solution. If $\exists y \geq 0$ such that $y^T A = 0, y^T b < 0$ then $Ax \leq b$ has no solution.

Theorem 2.1. *(Farkas' Lemma v1) $Ax \leq b$ has a solution if and only if $y^T b \geq 0$ for each vector $y \geq 0$ such that $y^T A = 0$.*

Proof. ($\implies$) Apply F-M. □

The equivalent statement for the above is that exactly one of the two systems has a solution:

1. $Ax \leq b$
2. $\exists y \geq 0$ such that $y^T A = 0, y^T b < 0$

Theorem 2.2. *(Farkas' Lemma v2) Only one of the two systems holds:*

- There exists a solution to the system $Ax = b$ and $x \geq 0$
- There exists a vector y such that $y^T A \geq 0$ and $b^T y < 0$

Summary 1. So does $Ax \leq b$ have a solution?

- Yes - display $\bar{x}$ such that $A\bar{x} \leq b$
- No - display $\bar{y}$ such that $\bar{y}^T A = 0$ and $\bar{y}^T b < 0$

This is a variation of Farkas' Lemma [Farkas, 1894] which is the following.

Theorem 2.3. *The system $Ax = b, x \geq 0$ has a solution if and only if $y^T b \geq 0$ for each vector y such that $y^T A \geq 0$.*

Proof. Write $Ax = b, x \geq 0$ as $Ax \leq b, -Ax \leq -b, -Ix \leq 0$ or

$$\begin{bmatrix} A \\ -A \\ -I \end{bmatrix} X \leq \begin{bmatrix} b \\ -b \\ 0 \end{bmatrix}$$

So $Ax = b, x \geq 0$ has a solution

$$\iff \begin{bmatrix} y' \\ y'' \\ z \end{bmatrix} \begin{bmatrix} b \\ -b \\ 0 \end{bmatrix} \geq 0 \text{ for each } \begin{bmatrix} y' \\ y'' \\ z \end{bmatrix} \geq 0 \text{ such that } \begin{bmatrix} y' \\ y'' \\ z \end{bmatrix}^T \begin{bmatrix} A \\ -A \\ -I \end{bmatrix} = 0 \iff$$

$$\iff (y' - y'')^T b \geq 0 \text{for each } \begin{bmatrix} y' \\ y'' \\ z \end{bmatrix} \geq 0 \text{such that } (y' - y'')^T A - z^T I = 0$$

$$\iff (y' - y'')^T b \geq 0 \text{ for each } y', y'', z \geq 0 \text{ such that } (y' - y'')^T A = z$$

$$\iff y \equiv y' - y'' \text{ and } y^T b \geq 0 \text{ for each } y \text{ such that } y^T A \geq 0$$

□

2.1 Integer Linear Programming

Summary 2. In summary, the previous sections say:

1. $Ax = b$ has a solution $\iff \nexists y$ such that $y^T A = 0$, $y^T b = 1$
2. $Ax = b$ with x integral has a solution $\iff \nexists y$ such that $y^T A$ integral, $y^T b$ non-integral
3. $Ax \leq b$ has a solution $\iff \nexists y$ such that $y^T A = 0$, $y^T b < 0$, $y \geq 0$

Problem 2.2. The next level up is does $Ax \leq b$, x integral have a solution?

To do this, we will study the structure of $\{x : Ax \leq b\}$. Linear algebra studies sets of the form $\{x : Ax = b\}$ where linear spaces are of the form $L = \{x : Ax = 0\}$ and these are finitely generated. This suggests we look at $C = \{x : Ax \leq 0\}$ where

$$\begin{aligned} x \in C, y \in C &\implies x + y \in C \\ x \in C, \lambda \in \mathbb{R}, \lambda \geq 0 &\implies \lambda x \in C \end{aligned}$$

A set satisfying the above is called a **cone**. Suppose we have vectors $x^1, ..., x^M$. Let $D = \{\lambda_1 x^1 + ... + \lambda_M x^M : \lambda_1 \geq 0, ..., \lambda_M \geq 0\}$.

Lemma 2.1. *D is a cone and is denoted by $cone\{x_1, ..., x_n\}$. It is the smallest cone containing $x^1, ..., x^M$. A cone like D is called a* ***finitely-generated cone****. A cone of the form $\{x : Ax \leq 0\}$ is a called a* ***polyhedral*** *cone.*

Theorem 2.4. *(Farkas, Minkowski, Weyl) A cone is polyhedral $\iff$ it is finitely generated.*

(Sketch) The idea behind the proof is that $b \in cone\{a_1, ..., a_m\} \iff \exists a$ solution to $y^T A = b, y \geq 0$, $A = [a_1 ... a_m]^T \iff b^T x \geq 0$ for all solutions to $Ax \geq 0$. Since there are infinitely $x's$, we need to choose a finite subset. So we need a sharper version of Farkas.

Theorem 2.5. *(Fundamental Theorem of Linear Inequalities, [Schrijver, p. 85]) Let $a^1, ..., a^M \in \mathbb{R}^n$ and let $t = rank\{a^1, .., a^M, b\}$ where $b \in \mathbb{R}^n$. Then exactly one of the two statements is true.*

1. b is a non-negative linear combination of linearly independent vectors from $a^1, ..., a^M$

2. There exists a hyperplane $\{x : C^T x = 0\}$ containing $(t-1)$ linearly independent vectors from $a^1, ..., a^M$ such that $C^T b < 0$ and $C^T a^1, ..., C^T a^M \geq 0$.

Proof. We may assume $a^1, ..., a^M$ span $\mathbb{R}^n$. Otherwise, use a transformation to map the space into a subspace with some $x_j = 0$. We first show that we cannot have both (1) and (2). Indeed, let $b = \lambda_1 a^1 + ... + \lambda_M a^M$ for some $\lambda_i \geq 0$ and suppose we have C as in (2). Then

$$\begin{aligned} C^T b < 0 \quad &\implies \quad C^T(\lambda_1 a^1 + ... + \lambda_M a^M) < 0 \\ &\implies \quad \lambda_1 \underbrace{C^T a^1}_{\geq 0} + ... + \lambda_M \underbrace{C^T a^M}_{\geq 0} < 0 \end{aligned}$$

which is impossible and we are done here. We will show that either (i) or (ii) must be true. Choose a linearly independent set of vectors $a_{i_1}, ..., a_{i_n}$ from $a^1, ..., a^M$. Let $B = \{a_{i_1}, ..., a_{i_n}\}$. We apply the following (**simplex**) algorithm.

1. Write $b = \lambda_{i_1} a_{i_1} + ... + \lambda_{i_n} a_{i_n}$. If $\lambda_{i_1}, ..., \lambda_{i_n} \geq 0$ then (1) holds and we *stop*.

2. Choose the smallest index h among $i_1, ..., i_n$ having $\lambda_h < 0$. Let $\{x : C^T x = 0\}$ be the hyperplane spanned by $B \backslash \{a_h\}$. Scale C so that $C^T a_h = 1$. Note that this means

$$\begin{aligned} c^T b = c^T(\lambda_{i_1} a_{i_1} + ... + \lambda_{i_n} a_{i_n}) \quad &= \quad \lambda_{i_1} C^T a_{i_1} + ... + \lambda_{i_n} C^T a_{i_n} \\ &= \quad \lambda_h C^T a_h = \lambda_h < 0 \end{aligned}$$

3. If $C^T a^1 \geq 0, ..., C^T a^M \geq 0$ then (2) holds and we *stop.*

4. Choose the smallest s such that $C^T a_s < 0$. Replace B by removing a_h and adding a_s. That is, $B \mapsto (B \backslash \{a_h\}) \cup \{a_s\}$.

5. Go to step 1.

To prove the theorem, we only need to show that the algorithm terminates. Let B_k denote the set B in the k^{th} iteration. If the algorithm does not terminate, then must have $B_k = B_l$ for some $k < l$ (since there are only finitely many choices for the set B). Let r be the highest index for which a_r has been removed from B at the end of one of the iterations $k, ..., l-1$ which we will say, it is p. Since $B_k = B_l$, we must have that a_r is added to B, say in iteration $q < p$. Note that

$$B_p \cap \{a_{r+1}, ..., a_m\} = B_q \cap \{a_{r+1}, ..., a_m\}$$

Let $B_p \equiv \{a_{i_1}, ..., a_{i_n}\}$ and $b = \lambda_{i_1} a_{i_1} + ... + \lambda_{i_n} a_{i_n}$. Let C' be the vector C found in step 2 of iteration q. We have the contradiction

$$(*)\ 0 > C'^T b = C'^T(\lambda_{i_1} a_{i_1} + ... + \lambda_{i_n} a_{i_n}) = \lambda_{i_1} C'^T a_{i_1} + ... + \lambda_{i_n} C'^T a_{i_n} > 0\ (**)$$

where (*) is noted in step (2) of the simplex algorithm and (**) is done as follows. If $i_j > r$ then $C'^T a_{i_j} = 0$ which follows from the choice of C'. If $i_j = r$ then $\lambda i_j < 0$ because r was chosen in step (2) of iteration p and $C' a_{i_j} < 0$ because r was chosen in step (4) of iteration q. If $i_j < r$ then $\lambda_{i_j} \geq 0$ since r was the smallest index with $\lambda_{i_j} < 0$ in iteration p and $C'^T a_{i_j} \geq 0$ since r was the smallest index with $C^{'} a_{i_j} < 0$ in iteration q. □

Summary 3. Given $a_1, ..., a_m \in \mathbb{R}^n$ with rank t, only one of the two must be true [Robert Bland, 1979]:

(1) b is a non-negative combination of linearly independent vectors from $a_1, ..., a_m$

(2) There exists a hyperplane $\{x : C^T x = 0\}$ containing at least $(t-1)$ linear independent vectors from $a_1, ..., a_m$ such that $C^T a_i \geq 0$, $i = 1, ..., m$ and $C^T b < 0$.

Theorem 2.6. *A cone is polyhedral if and only if it is finitely generated (previously stated in a previous lecture).*

Proof. ($\Longleftarrow$) [A] Let $x_1, ..., x_m \in \mathbb{R}^n$ and assume $x_1, ..., x_m$ span $\mathbb{R}^n$. Otherwise, we can work in a subspace of $\mathbb{R}^n$. Consider all linear hyperplanes $\{x : C^T x = 0\}$ that are spanned by $(n-1)$ linearly independent vectors from $x_1, ..., x_m$ and have the property $C^T x_1 \geq 0, ..., C^T x_m \geq 0$. There are only finitely many such C. Call them $C^1, ..., C^l$. If $\bar{x} \in cone\{x_1, ..., x_m\}$, then $C^{iT}\bar{x} \geq 0, \forall i = 1, ..., l$. On the other hand, if $\bar{x} \notin cone\{x_1, ..., x_m\}$, then by the fundamental theorem, there must be some $i \in \{1, ..., l\}$ such that $C^{iT}\bar{x} < 0$. Thus,

$$cone\{x_1, ..., x_m\} = \{x : C^{iT}x \geq 0, ..., C^{lT}x \geq 0\}$$

($\Longrightarrow$) [B] Let $C = \{x : a_1^T x \leq 0, ..., a_m^T x \leq 0\}$. By [A], there exists vectors $b_1, ..., b_t$ such that

$$(*)\ cone\{a_1, ..., a_m\} = \{x : b_1^T x \leq 0, ..., b_t^T x \leq 0\}$$

We will show that $C = cone\{b_1, ..., b_t\}$. To do this, we first show that $cone\{b_1, ..., b_t\} \subseteq C$. This is clear because $b_i \in C$ since $b_i^T a_j \leq 0$ for all $j = 1, ..., m$ by the definition of a cone and (*).

Conversely, to show that $C \subseteq cone\{b_1, ..., b_t\}$, let $\bar{y} \in C$ and suppose $\bar{y} \notin cone\{b_1, ..., b_t\}$. By [A], $cone\{b_1, ..., b_t\}$ is polyhedral. So

$$cone\{b_1, ..., b_t\} = \{y : w^{iT}y \leq 0, ..., w^{kT}y \leq 0\}$$

for some vectors $w^1, ..., w^k$. Thus, for some i, we must have $w^{iT}\bar{y} > 0$. Note that $w^{iT}b_j \leq 0$ for all j. By (*), $w^i \in cone\{a_1, ..., a_m\}$ and thus

$$w^i = \lambda_1 a_1 + ... + \lambda_m a_m$$

where $\lambda_1 \geq 0, ..., \lambda_m \geq 0$. Hence, for each $x \in C$ we have

$$\begin{aligned} w^{iT} x &= (\lambda_1 a_1 + ... + \lambda_m a_m)^T x \\ &= \lambda_1 a_1^T x + ... + \lambda_m a_m^T x \leq 0 \end{aligned}$$

This is a contradiction since $\bar{y} \in C$ and $w^{iT}\bar{y} > 0$. $\square$

Theorem 2.7. *(Caratheodory's Theorem) Let $x_1, ..., x_m \in \mathbb{R}^n$ and suppose $x \in cone\{x_1, ..., x_m\}$. Then, x can be written as a non-negative linear combination of linearly independent vectors from $x_1, ..., x_m$.*

Proof. Fundamental Theorem. (Exercise: Fill in the blanks) $\square$

2.2 Convex Sets

We now study sets of the form $\{x : Ax \leq b\}$ for some matrix A and vector b. Such a set is called a **polyhedron**. Let $x_1, x_2 \in P = \{x : Ax \leq b\}$ and let $0 \leq \lambda \leq 1$ be a a number with

$$\begin{aligned} A(\lambda x_1 + (1-\lambda)x_2) &= \lambda A x_1 + (1-\lambda)Ax_2 \\ &\leq \lambda b + (1-\lambda)b = b \\ &\Longrightarrow \lambda x_1 + (1-\lambda)x_2 \in P \end{aligned}$$

A set S is called **convex** if $x_1, x_2 \in P$ and $0 \leq \lambda \leq 1 \implies \lambda x_1 + (1-\lambda)x_2 \in P$. So polyhedra are convex.

Definition 2.1. The **convex hull** of a set of vectors X is the smallest convex set containing X.

Lemma 2.2. *Let S be a convex set with $x_1, ..., x_m \in S$. Let $\lambda_1, ..., \lambda_m \geq 0$ with $\sum \lambda_i = 1$. Then $\sum_{i=1}^m \lambda_i x_i \in S$.*

Proof. By definition, $1 - \lambda_1 = \sum_{j=2}^m \lambda_j$ and hence

$$v = \frac{1}{1-\lambda_1}\left(\sum_{j=2}^m \lambda_j x_j\right) \in S$$

by induction. This implies $\sum_{i=1}^m \lambda_i x_i = \lambda_1 x_1 + (1-\lambda_1)v \in S$ by convexity. $\square$

Corollary 2.1. *By the lemma above,*

$$Convex\_Hull(X) = \left\{ \sum_{i=1}^{t} \lambda_i x_i, t \geq 0, x_j \in X, \lambda_j \geq 0, j \in \{1, ..., t\}, \sum_{k=1}^{t} \lambda_k = 1 \right\}$$

Definition 2.2. A **polytope** is the convex hull of finitely many vectors.

Theorem 2.8. *(Finite Basis Theorem or FBT) A set P is a polytope iff P is a bounded polyhedron.*

Before proving this theorem, we add a few definitions and sub-theorems.

Definition 2.3. We say that a polyhedron is **bounded** if there exists $l \leq u$ such that $P \subseteq \{x : l \leq x \leq u\}$. For $S, T \subseteq \mathbb{R}^n$, we can define $S + T = \{s + t : s \in S, t \in T\}$, which is called the **Minkowski Sum**.

Theorem 2.9. *A set P is a polyhedron if and only if P is the sum of a polytope and a cone.*

Proof. ($\Longrightarrow$) Suppose that $P = \{x : Ax \leq b\}$. We show $P = Q + C$ where Q is a polytope and C is a cone. Consider the polyhedral cone

$$T = \left\{ \begin{pmatrix} x \\ \lambda \end{pmatrix} : x \in \mathbb{R}^n, \lambda \in \mathbb{R}, \lambda \geq 0, Ax - \lambda b \leq 0 \right\}$$

We know that T is finitely generated by vectors $\begin{pmatrix} x_1 \\ \lambda_1 \end{pmatrix}, ..., \begin{pmatrix} x_2 \\ \lambda_2 \end{pmatrix}$ and we may scale these vectors so that for each i, $\lambda_i = 0$ or $\lambda_i = 1$. Notice that $x \in P \iff \begin{pmatrix} x \\ 1 \end{pmatrix} \in T$. If $\begin{pmatrix} x \\ 1 \end{pmatrix} \in T$ and

$$\begin{pmatrix} x_1 \\ \lambda_1 \end{pmatrix} = \gamma_1 \begin{pmatrix} x_1 \\ \lambda_1 \end{pmatrix} + ... + \gamma_m \begin{pmatrix} x_m \\ \lambda_m \end{pmatrix}, \gamma_1 \geq 0, ..., \gamma_m \geq 0$$

then $\sum(\gamma_i : \lambda_i = 1) = 1$. So $\begin{pmatrix} x \\ 1 \end{pmatrix} \in T \iff x \in \sum(\gamma_i x_i : \lambda = 0) + \sum(\gamma_i x_i : \lambda = 1)$ with $\gamma_1, ..., \gamma_m \geq 0$ and $\sum(\gamma_i : \lambda = 1) = 1$. Thus, letting C be the cone generated by $\{x_i : \lambda_i = 0\}$ and letting Q be the convex hull of $\{x_i : \lambda_i = 1\}$ we have $P = Q + C$. $\square$

($\Longleftarrow$) Now suppose that $P = Q + C$ for some polytope Q and polyhedral cone C. We must show that P is a polyhedron. Let $C = cone(y_1, ..., y_t)$ and $Q = Convex\_Hull(x_1, ..., x_m)$. So $\bar{x} \in P \iff \bar{x}$ can be written as

$$\lambda_1 y_1 + ... + \lambda_t y_t + \gamma_1 x_1 + ... + \gamma_m x_m$$

with $\lambda_i, \gamma_i \geq 0$ and $\sum \gamma_i = 1$. So $\bar{x} \iff$

$$\begin{pmatrix} \bar{x} \\ 1 \end{pmatrix} = \lambda_1 \begin{pmatrix} y_1 \\ 0 \end{pmatrix} + ... + \lambda_t \begin{pmatrix} y_t \\ 0 \end{pmatrix} + \gamma_1 \begin{pmatrix} x_1 \\ 0 \end{pmatrix} + ... + \gamma_m \begin{pmatrix} x_m \\ 1 \end{pmatrix}, \gamma_i \geq 0, \lambda_i \geq 0$$

and $\iff$

$$\begin{pmatrix} \bar{x} \\ 1 \end{pmatrix} = cone\left(\begin{pmatrix} y_1 \\ 0 \end{pmatrix}, ..., \begin{pmatrix} y_t \\ 0 \end{pmatrix}, \begin{pmatrix} x_1 \\ 0 \end{pmatrix}, ..., \begin{pmatrix} x_m \\ 1 \end{pmatrix} \right) = S$$

But S is a polyhedral cone $S = \left\{ \begin{pmatrix} x \\ \lambda \end{pmatrix} : Ax + \lambda b \leq 0 \right\}$ for some A and b. Thus,

$$\bar{x} \in P \iff \begin{pmatrix} \bar{x} \\ 1 \end{pmatrix} \in S \iff A\bar{x} + b \leq 0 \iff A\bar{x} \leq -b$$

and $P = \{x : Ax \leq -b\}$ which is polyhedral.

3 Linear Optimization

A **linear programming** (LP) **problem** is to maximize or minimize a linear function subject to linear equality and inequality constraints. For example,

$$\begin{array}{rl} \max & c_1x_1 + ... + c_nx_n \\ \text{subject to (s.t.)} & Ax \leq b \end{array}$$

Other forms could be

$$\begin{array}{rl} \max & c^Tx \\ \text{subject to (s.t.)} & Ax \leq b \\ & x \geq 0 \end{array}$$

Let $P = \{x : Ax \leq b\}$. Then P is the feasible set of solutions to the LP problem: $\max(c^Tx : Ax \leq b)$.

Example 3.1. (Traveling Salesman Problem or TSP)

Input: Finite number n of cities and the cost (or distance) to travel between each pair.

Output: A minimum cost tour T where T visits all n cities and returns to starting point.

We will be examining the symmetric cost version, where $cost(a, b) = cost(b, a)$. That is, the cost of traveling from a to b is the same as traveling from b to a. For a map of cities connected with edges, we can either specify the paths taken between cities as edges in pairs, or a 0-1 vector in space (particularly in $\mathbb{R}^{\frac{n(n-1)}{2}}$) where this vector defines all pairs and there is a 1 if the path is take and 0 otherwise.

For the TSP, there is a finite set S of of 0-1 vectors representing tours:

$$\min(c^Tx : x \in S), c^T\bar{x} = \sum_{e \in E} c_e\bar{x}$$

where E is the set of all edges (pairs of cities). We can see that $|S| = \frac{(n-1)!}{2}$ and

$$\min(c^Tx : x \in S) = \min(c^Tx : x \in Convex\_Hull(S)) \overset{FBT}{=} \min(c^Tx : Ax \leq b)$$

So the TSP is an LP problem. Alternatively, we can reformulate this as follows. Let $G(V, E)$ where $V = \{0, 1, ..., n-1\}$ and E is the set of all unordered pairs of V. If $C = (C_e : e \in E)$ and C_e is the cost to travel between ends of $e = (u, v)$. The TSP is to find the minimum cost tour.

3.1 Duality

Theorem 3.1. *(Weak Duality Theorem) If $\bar{x}$ satisfies $Ax \leq b$ and $\bar{y}$ satisfies $\bar{y}^TA = c^T, y \geq 0$ then $c^T\bar{x} \leq \bar{y}^Tb$.*

Proof. We have $A\bar{x} \leq b$. Multiplying by $\bar{y}$ we have $\bar{y}^TA\bar{x} \leq \bar{y}^Tb$. By $\bar{y}^TA = c^T$ we have

$$c^T\bar{x} = \bar{y}^TA\bar{x} \leq \bar{y}^Tb$$

□

Theorem 3.2. *(Duality Theorem [Von Neumann 1947]) We have*

$$\underbrace{\max(c^Tx : Ax \leq b)}_{\textit{Primal Problem}} = \underbrace{\min(y^Tb : y^TA = c^T, y \geq 0)}_{\textit{Dual Problem}}$$

provided each of the two LP models have feasible solutions.

Proof. By Weak Duality, we need to show there exists $\bar{x}$ and $\bar{y}$ such that $c^T\bar{x} \geq \bar{y}^T b$ (which implies $c^T\bar{x} = \bar{y}^T b$). Thus, we need to show there exists a solution to

$$Ax \leq b, y^T A = c^T, c^T x \geq y^T b, y \geq 0$$

Note that $y^T A = c^T \iff A^T y = c$. Writing as a matrix,

$$\begin{array}{c} u \\ \lambda \\ v \\ w \end{array} \begin{bmatrix} A & 0 \\ -c^T & b^T \\ 0 & A^T \\ 0 & -A^T \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \leq \begin{bmatrix} b \\ 0 \\ c \\ -c \end{bmatrix}, y \geq 0$$

By Farkas, this system has a solution if and only if $u^T b + v^T c - w^T c \geq 0$ for all $u, \lambda, v, w \geq 0$ such that $u^T A - \lambda c^T = 0$ and $\lambda b^T + v^T A^T - w^T A^T \geq 0$. To prove this theorem, we show that this is true via considering cases.

Case I ($\lambda > 0$): We have

$$\begin{aligned} u^T b &= b^T u = \frac{1}{\lambda}\lambda b^T u \\ &\geq \frac{1}{\lambda}(w^T - v^T)A^T u \\ &= \frac{1}{\lambda}(w^T - v^T)\lambda c \\ &= (w^T - v^T)c \end{aligned}$$

and so $u^T b - (w^T - v^T)c \geq 0$ which is what we want.

Case 2 ($\lambda = 0$): Let $\bar{x}, \bar{y}$ satisfy $A\bar{x} \leq b, \bar{y}^T A = c^T, y \geq 0$. Thus, $u^T b \geq u^T Ax = \lambda c^T \bar{x} = 0$ and

$$\begin{aligned} (w^T - v^T)c &= (w^T - v^T)A^T\bar{y} \\ &\leq \lambda b^T \bar{y} = 0 \end{aligned}$$

and hence $u^T b \geq (w^T - v^T)c$ which is what we want. □

Example 3.2. (TSP revisited) The TSP can be reformulated as a relaxed LP problem where in this modified model, we have

$$\begin{aligned} \text{Variables} &= (X_e : e \in E) \\ \min \quad & \sum (c_e X_e : e \in E) \\ \text{sb. to} \quad & 0 \leq X_e \leq 1, \forall e \in E \\ \sum (X_e : e \text{ meets } v) &= 2, \forall v \in V \\ (1) \ \sum (X_e : e \text{ has one end in } S) &\geq 2, \forall S \subseteq V, \emptyset \neq S \neq V \end{aligned}$$

where this whole system is called a **subtour polytope** and (1) is called a **subtour elimination constraint**. By the Finite Basis Theorem, there exists further inequalities satisfied by all tours such that the optimal solution to the LP is a tour. So

$$\text{TSP} \overset{(1)}{=} \min(c^T x : Ax \leq b) \overset{(2)}{=} \max(b^T y : A^T y = c, y \geq 0)$$

where (1) is by FBT and (2) is by LP Duality. By Weak Duality, we know that for any $\bar{y}$ we have TSP $\geq b^T\bar{y}$. So any dual solution $\bar{y}$ gives a lower bound for the TSP. By Caratheodory's Theorem, there exists an optimal dual solution with at most $|E|$ non-zero components.

Define SUB as the optimal value of $c^T x$ over the subtour polytope. The triangle inequality tells us

$$cost(a, b) \leq cost(a, c) + cost(c, b), \forall c$$

It is known that TSP $\leq \frac{3}{2} \cdot$ SUB. The $\frac{4}{3}$'rds conjecture says TSP $\leq \frac{4}{3} \cdot$ SUB.

Theorem 3.3. *If the primal LP* $\max(c^T x : Ax \leq b)$ *has an optimal solution, the dual LP* $\min(y^T b : y^T A = 0, y \geq 0)$ *also has an optimal solution and the Duality Theorem holds.*

Proof. It suffices to show that the dual LP has a feasible solution. Suppose that the dual LP has no solution, where $A^T y = c$ and $y \geq 0$. By Farkas, there exists a solution z such that $z^T c \leq -1$ and $z^T A^T \geq 0$. That is, $Az \geq 0$ and $c^T z \leq -1$. Let x^* be an optimal solution to the primal LP. But

$$A(x^* - z) = Ax^* - Az \leq b$$

$$c^T(x^* - z) = c^T x^* - c^T z > c^T x^*$$

This is a contradiction since x^* is an optimal solution. □

Theorem 3.4. *(Affine Farkas' Lemma) Suppose $c^T x \leq \delta$ for all x such that $Ax \leq b$ and suppose there exists a solution to $Ax \leq b$. Then for some $\delta' \leq \delta$ we have that $c^T x \leq \delta'$ is a non-negative linear combination of $Ax \leq b$.*

Proof. Following the previous argument, there exists a solution to $A^T y = c, y \geq 0$. Thus, by the duality theorem, there is some $\bar{y}$ such that $\bar{y}$ is an optimal solution to

$$\min(y^T b : y^T A = c^T, y \geq 0) = \delta'$$

Thus, $\bar{y}$ gives the non-negative combinations of $Ax \leq b$ where

$$\bar{y}^T Ax \leq \bar{y}^T b \implies c^T x \leq \delta' \leq \delta$$

and $\bar{y}$ gives the non-negative combination of $Ax \leq b$. □

Proposition 3.1. *Suppose that $\bar{x}$ and $\bar{y}$ are feasible solutions to the primal and dual LPs respectively. Then the following are equivalent.*

1) $\bar{x}$ and $\bar{y}$ are optimal solutions

2) $c^T \bar{x} = \bar{y}^T b$

3) If a component of $\bar{y}$ is positive, then the corresponding inequality $Ax \leq b$ is satisfied by $\bar{x}$ as an equation. That is $\bar{y}^T(b - A\bar{x}) = 0$

In (3), we can say that being an optimal solution is equivalent to the ***complementary slackness conditions (CSC)*** *which are for each $j = 1, ..., m$ either $\bar{y}_j = 0$ OR $a_j^T \bar{x} = b_j$.*

Proof. (1) $\iff$ (2) Use the Duality Theorem.

(2) $\implies$ (3) We have

$$\begin{aligned} c^T x = y^T A\bar{x} \leq \bar{y}^T b \quad &\implies \quad c^T \bar{x} = y^T b \iff \bar{y}^T A\bar{x} = \bar{y}^T b \\ &\iff \quad \bar{y}^T A\bar{x} - \bar{y}^T b = 0 \\ &\iff \quad \bar{y}^T(A\bar{x} - b) = 0 \end{aligned}$$

(3) $\implies$ (2) Same proof. □

Theorem 3.5. *For each inequality $a_i^T x \leq b_i$ in $Ax \leq b$, exactly one of the following holds:*

(1) The maximum in the primal LP has an optimal solution $\bar{x}$ with $a_i^T \bar{x} < b_i$

(2) The minimum in the dual LP has an optimal solution $\bar{y}$ with $\bar{y}_i > 0$

Proof. Omitted. See Schrijver. □

Theorem 3.6. *(Motzkin's Transposition Theorem) There exists a vector x with $Ax < b$, $Bx \leq c$ iff for all vectors $y \geq 0, z \geq 0$,*

(i) If $y^T A + z^T B = 0$ then $y^T b + z^T c \geq 0$.

(ii) If $y^T A + z^T B = 0, y \neq 0$, then $y^T b + z^T c > 0$

Proof. It is easy to see that the conditions (i) and (ii) are necessary ($\implies$ is done). Now suppose that (i) and (ii) hold. By Farkas, we know there exists a solution x to $Ax \leq b$ and $Bx \leq c$. Notice that (ii) implies that for each inequality $a_i^T x \leq b_i$ in $Ax \leq b$ there is no solution to

$$y \geq 0, z \geq 0, y^T A + z^T B = -a_i^T, y^T b + z^T c \leq -b_i$$

This implies that there exists a vector x^i with

$$Ax^i \leq b, Bx^i \leq c, a_i^T x^i < b_i$$

(See Assignment 2 for details). The **barycentre** $\bar{x} = \frac{1}{m}(x^1 + ... + x^m)$ satisfies

$$A\bar{x} < b, B\bar{x} \leq c$$

which is what we wanted. □

3.2 Structure of Polyhedra

Definition 3.1. The **characteristic cone** of a polyhedron $P = \{x : Ax \leq b\}$ is defined as

$$Char\_Cone(P) = \{y : x + y \in P, \forall x \in P\}$$

Lemma 3.1. *We have* $y \in Char\_Cone(P) \iff \exists x \in P$ *with* $x = \lambda y \in P$ *for any* $\lambda \geq 0$.

Proof. Let $y \in Char\_Cone(P)$. Let $x \in P$. Thus, $x + ky \in P$ for all $k = 1, 2, ...$. Since P is convex, $x + ky \in P$ for all $\lambda \geq 0$. Let $x \in P$ and let y be a vector such that $x + \lambda y \in P$ for all $\lambda \geq 0$. Let $Ax \leq b$ be a system such that $P = \{x : Ax \leq b\}$. Then we must have $Ay \leq 0$. That is, if $a_i^T y > 0$ then for large enough λ we would have $a_i^T(x + \lambda y) > b_i$. Thus, for any $\bar{x} \in P$ we have $A(\bar{x} + \bar{y}) = A\bar{x} + A\bar{y} \leq b$. □

Lemma 3.2. *If* $P = \{x : Ax \leq b\}$ *then* $Char\_Cone(P) = \{y : Ay \leq 0\}$.

Note 6. If $P = Q + C$ where Q is a polytope and C is a cone, then $C = Char\_Cone(P)$. If P is bounded, then $Char\_Cone(P) = \{0\}$.

Definition 3.2. Vectors $x^1, ..., x^m \in \mathbb{R}^n$ are **affinely independent** if the solution to

$$\begin{cases} \sum_{i=1}^m \lambda_i x^i &= 0 \\ \sum_{i=1}^m \lambda_i &= 0 \end{cases}$$

is $\lambda_1 = ... = \lambda_m = 0$. The **dimension** of a set $K \subseteq \mathbb{R}^n$ is one less than the maximum cardinalities of an affinely independent subset of K.

Lemma 3.3. *Let* $x^1, ..., x^m \in \mathbb{R}^n$ *and let* $w \in \mathbb{R}^n$. *If* $x^1, ..., x^m$ *are affinely independent then* $x^1 - w, ..., x^m - w$ *are affinely independent.*

Proof. Suppose

$$\begin{cases} \sum_{i=1}^m \lambda_i (x^i - w) &= 0 \\ \sum_{i=1}^m \lambda_i &= 0 \end{cases}$$

We have

$$\sum_{i=1}^m \lambda_i (x^i - w) = \sum_{i=1}^m \lambda_i x^i - w \underbrace{\left(\sum_{i=1}^m \lambda_i \right)}_{=0} = \sum_{i=1}^m \lambda_i x^i = 0$$

and hence $\lambda_1 = ... = \lambda_m = 0$. □

Definition 3.3. The **affine hull** of $X \subseteq \mathbb{R}^n$ is

$$Affine\_Hull(P) = \{\lambda_1 x^1 + ... + \lambda_t x^t : t \geq 1,\ x^1, ..., x^t \in X,\ \lambda_1 + ... + \lambda_t = 1\}$$

Definition 3.4. Given $P = \{x : Ax \leq b\}$, if $a_i^T x \leq b_i$ holds as an equation for all $\bar{x} \in P$ (that is, $a_i^T \bar{x} = b_i, \forall \bar{x} \in P$), then $a_i^T x \leq b$ is called an **implicit equation**. If $A^= x \leq b^=$ are the implicit equations in $Ax \leq b$ then we denote the remaining equations as $A^+ x \leq b^+$.

Note 7. There exists $\bar{x} \in P$ such that $A^= \bar{x} = b^=$ and $A^+ \bar{x} < b^+$.

Lemma 3.4. *We have*

$$Affine\_Hull(P) = \{x : A^= x = b^=\} = \{x : A^= x \leq b^=\}$$

Proof. (1) $[Affine\_Hull(P) \subseteq \{x : A^= x = b^=\}]$ By definition $P \subseteq \{x : A^= x = b^=\}$. Suppose that $\bar{x} = \lambda_1 x^1 + ... + \lambda_t x^t$ with $x^1 + ... + x^m \in P$ and $\lambda_1 + ... + \lambda_t = 1$. Then,

$$A^= \bar{x} = \lambda_1 A^= x^1 + ... + \lambda_t A^= x^t = \lambda_1 b^= + ... + \lambda_t b^= = b^=$$

(2) $[\{x : A^= x = b^=\} \subseteq \{x : A^= x \leq b^=\}]$ Trivial by definition.

(3) $[\{x : A^= x \leq b^=\} \subseteq Affine\_Hull(P)\}]$ Let $\bar{x}$ satisfy $A^= \bar{x} \leq b^=$. Let $x' \in P$ be such that $A^= x' = b^=, A^+ x' < b$. If $\bar{x} = x'$ then $\bar{x} \in P \implies \bar{x} \in Affine\_Hull(P)$. If $\bar{x} \neq x'$, then the line segment connecting $\bar{x}$ and x' contains more that one point in P. Therefore, the affine hull of P contains the entire line through x' and $x \implies \bar{x} \in Affne\_Hull(P)$. ☐

Definition 3.5. $P \subseteq \mathbb{R}^n$ has **full dimension** if $dim(P) = n$.

Note 8. P has full dimension $\iff$ there are no implicit equations.

Note 9. We have $dim(P) = n - rank(A)$.

Definition 3.6. We say that $c^T x \leq \delta$ is called **valid** for P if $\forall \bar{x} \in P$ we have $c^T \bar{x} \leq \delta$.

Definition 3.7. $\{x : c^T x = \delta\}$ is called a **supporting hyperplane** of P if $\delta = \max(c^T x : Ax \leq b)$ and c is not the zero vector. This implies that it is a valid inequality and the hyperplane touches P.

Definition 3.8. F is a **face** of P if either $F = P$ or F is the intersection of P and a supporting hyperplane.

Theorem 3.7. *F is a face of P $\iff$ $F \neq \emptyset$ and $F = \{x \in P : A'x = b'\}$ for some subsystem $Ax' \leq b'$ of $Ax \leq b$.*

Proof. ($\implies$) Suppose $F = P \cap \{x : c^T x = \delta\}$. Consider the LP problem $\max(c^T x : Ax \leq b)$. Since $c^T x \leq \delta$ is valid, this LP has a finite optimal value. By the duality theorem, there exists an optimal solution to $\min(y^T b : y^T A = c^T, y \geq 0)$. Let y^* be an optimal solution. Let $I = \{i : y_i^* > 0\}$. By the CSC, a vector $\bar{x}$ is optimal for the primal LP $\iff$ $a_i^T \bar{x} = b_i$ for all $i \in I$.

But F is the set of optimal solutions to the primal LP. Thus, $F = \{x \in P : A'x = b\}$ where $A'x = b'$ are the equations $a_i^T x = b_i$ for any $i \in I$.

($\Longleftarrow$) Suppose $F = \{x \in P : A'x = b'\}$. We want to construct c such that $\max(c^T x : Ax \leq b) = F$. Let c be the sum of the rows of A'. Then every optimal solution satisfies $A'x = b'$ (since every $x \in P$ satisfies $Ax \leq b$). ☐

Definition 3.9. A **facet** is a maximal face of P that is not P itself.

Theorem 3.8. *Suppose no inequality in $A^+ x \leq b^+$ is redundant in $Ax \leq b$. Then there is a 1-1 correspondence between the facets of P and inequalities in $A^+ x \leq b^+$ and*

$$F_i = \{x \in P : a_i^+ x = b_i^+\}$$

for facets F and inequalities $a_i^+ x \leq b^+$ in $A^+ x \leq b^+$.

Theorem 3.9. *If F is a facet of P, then $\dim(F) = \dim(P) - 1$.*

Corollary 3.1. *If P is full-dimensional and $Ax \leq b$ is irredundant, then $Ax \leq b$ is the unique linear representation of P, up to multiplying the inequalities by positive scalars.*

Definition 3.10. A **minimal** face of P is a face that contains no other face of P.

Proposition 3.2. *A face F is minimal $\iff$ F is an affine subspace (that is $F = \{x : A'x = b'\}$ for some subsystem $A'x \leq b'$).*

Proposition 3.3. *Suppose $F = \{x : A'x = b'\}$ is a minimal face of P with $A'x \leq b'$ a subsystem of $Ax \leq b$. Then $rank(A') = rank(A)$.*

Definition 3.11. A face is called a **vertex** if it consists of a single point.

Definition 3.12. P is called **pointed** if it contains faces that are vertices. If P is pointed, then every minimal face is a vertex $\implies$ every bounded non-empty polyhedron is pointed.

Definition 3.13. A face of dimension 1 is called an **edge**. If the face is a half-line, it is called a **ray**. Two vertices of P are called **adjacent** or **neighbours** if they are contained in an edge.

Remark 3.1. We claim that the cone $C = \{x : Ax \leq 0\}$ is the only minimal face in

$$Lin\_Space(C) = \{x : Ax = 0\}$$

Let $t = \dim(Lin\_Space(C))$. Let $G_1, ..., G_s$ be the faces of dimension $t+1$. If C is pointed, the $G_1, ..., G_s$ are **extreme rays** of C. For each $i \in 1, ..., s$ let $y_i \in G_i \backslash Lin\_Space(C)$. Choose $z_1, ..., z_u$ be such that $Lin\_Space = Cone\{z_1, ..., z_u\}$.

Theorem 3.10. $C = Cone\{y_1, ..., y_s, z_1, ..., z_u\}$

Proof. Induction on $\dim(C) - t$. □

Theorem 3.11. *Let $F_1, ..., F_r$ be the minimal faces of $P = \{x : Ax \leq b\}$ and for each i let $X^i \in F_i$. Then*

$$P = Conv\_Hull\{X^1, ..., X^r\} + Char\_Cone(P)$$

Corollary 3.2. *If P is bounded, then P is the convex hull of its vertices.*

3.3 Polyhedral Combinatorics

If $S \subseteq \mathbb{R}^n, w \in \mathbb{R}^n$ then an optimization problem might be of the form $\min(w^T x : x \in S)$.

Example 3.3. Consider a graph $G = (V, E)$ where V and E are the vertices and edges respectively.

1) X is a **stable set** if there are no edges between vertices in X. Let $S = \{X \subseteq V : X \text{ is a stable set}\}$, $w = \{w_v : v \in V\}$. The **max-weight** stable set problem is finding the set X with largest weight $\sum_{v \in X} w_v$. If X is a stable set, look at its characteristic vector

$$\begin{aligned} y^X &= (y_v^X : v \in V) \\ y_v^X &= \begin{cases} 1 & \text{if } v \in V \\ 0 & \text{if } v \notin V \end{cases} \end{aligned}$$

The weight of stable set $w^T y^X$ is either maximized or minimized.

2) $S \equiv$ Characteristic vectors of tours $(X_e : e \in E)$ (e.g. TSP)

3) $S \equiv$ Cuts in G (max cut vs. min cuts)

4) $S \equiv$ Matchings in G. $M \subseteq E$ is a **matching** if $\forall v \in V$ at most one edge in M has v as one of its ends. The plan is to find

$$\begin{aligned} \min(w^T x : x \in S) &= \min(w^T x : x \in Conv\_Hull(S)) \\ &= \min(w^T x : Ax \leq b) \\ &= \max(y^T b : y^T A = w^T, y \geq 0) \end{aligned}$$

Problem 3.1. For the fourth example, how can we find the matrix and vector in the system $Ax \leq b$?

Solution. (Lovasz' Idea) Suppose $\dim(S) = n$. Then each facet of $P = Conv\_Hull(S)$ has dimension $n-1$ so there is a unique hyperplane that contains each facet. To show $P = \{x : Ax \leq b\}$ we need:

(1) $P \subseteq \{x : Ax \leq b\}$

(2) Each facet of P is induced by some inequality in $Ax \leq b$.

Theorem 3.12. *(Edmonds' Matching Theorem) If $G = (V, E), v \in V$, let $\delta(v) = \{e \in E : v \text{ is an end of } e\}$. For $A \subseteq V$, let $\gamma(A) = \{e \in E : e \text{ has both ends in } A\}$. Edmonds found that the convex hull of matchings is defined by*

$$x_e \geq 0, \forall e \in E$$

and also

$$\begin{aligned} \sum (X_e : e \in \delta(v)) &\leq 1, \forall v \in V \\ \sum (X_e : e \in \gamma(A)) &\leq \frac{|A|-1}{2}, \forall A \subseteq V, |A| \text{ odd} \end{aligned}$$

Proof. Suppose that $w^T w \leq t$ induces a facet of the matching polytope. Let M^* be the set of matchings M such that $w^T x^M = t$ then

$$x_e^M = \begin{cases} 1 & e \in M \\ 0 & e \notin M \end{cases}$$

(Case I) Suppose that $w_e < 0$ for some e. Every matching in M^* must satisfy $x_e^M = 0$. Since $M \in M^*, e \in M$, the matching $M \backslash \{e\}$ would violate $w^T x \leq t$. So $x_e \geq 0$ induces the same facet of $w^T x \leq t$.

(Case II) $w_e \geq 0$ for all e.

[Case II.A] Suppose $\exists v \in V$ such that each M^* meets V. Every $M \in M^*$ satisfies $\sum(X_e : e \in \delta(v)) = 1$ so $\sum(X_e : e \in \delta(v)) \leq 1$ induces the same facet as $w^T x \leq t$.

[Case II.B] Suppose $\nexists v$ such that $M \in M^*$ meets v. Let

$$\begin{aligned} E' &= \{e \in W : w_e > 0\} \\ A &= \{v \in V : V \text{ meets some edge in } E'\} \end{aligned}$$

We claim that each $M \in M^*$ satisfies $\sum(X_e : e \in \delta(v)) = \frac{|A|-1}{2}$ (hard proof). □

Remark 3.2. If $Ax \leq b$ has a solution, let F be a minimal face of $P = \{x : Ax \leq b\}$. Thus, $F = \{x : A'x = b'\}$ for a subsystem $A'x \leq b'$. Use Cramer's rule to obtain an $\bar{x}$ such that $A'\bar{x} = b$.

Definition 3.14. In general, we can represent a rational polyhedron $P \subseteq \mathbb{R}^n$ as

(1) $P = \{x : Ax \leq b\}$

(2) $P = Conv\_Hull\{x^1, ..., x^k\} + Cone\{y^1, ..., y^t\}$

The **vertex complexity** ν is the minimum $\nu \geq n$ such that $\nu \geq$ size of each x^i and y^i. The **facet complexity** σ is the minimum $\sigma \geq 0$ such that $\sigma \geq$ size of each inequality on $Ax \leq b$ (over all possible representations and defining systems for P).

Theorem 3.13. *We have* $\nu \leq 4n^2\sigma^2$ *and* $\sigma \leq 4n^2\nu$

Proof. Prove by using minimal faces and equations through affinely independent sets of vectors. □

4 Algorithms and Complexity

Theorem 4.1. *If any one of the the problems (1), (2), (3) are polynomial time solvable, then all three are polynomial time solvable.*

(1) A, b *rational, does* $Ax \leq b$ *have a solution?*

(2) A, b *rational, find a solution to* $Ax \leq b$ *if one exists.*

(3) A, b, c *rational, solve* $\max(c^T x : Ax \leq b)$. *Give whether or not it is infeasible, optimal with a provided solution, or unbounded with a* z *such that* $Ax \leq 0$ *and* $c^T z > 0$.

Proof. Clearly $(3) \implies (2) \implies (1)$.

($(1) \implies (2)$) Check if $Ax \leq b$ has a solution . If not, stop. If it does, check if the system

$$\begin{aligned} a_1^T x &= b_1 \\ a_2^T x &\leq b_2 \\ \vdots \quad & \vdots \quad \vdots \\ a_m^T x &\leq b_m \end{aligned}$$

has a solution. If no solution then remove $a_1^T x \leq b_1$ since it is redundant. If yes, then replace $a_1^T \leq b_1$ by $a_1^T x = b_1$. Repeat for each of the remaining inequalities. We eventually obtain a system of equation. Check with Gaussian elimination.

((2) $\implies$ (3)) Check if $Ax \leq b$ has a solution. If no, then stop because it is infeasible. Check if $y^T A = c^T$ for $y \geq 0$ has a solution. If no solution then by Farkas, there exists z such that $Az \leq 0, c^T z = 1$. Find the z to prove LP is unbounded. Stop for the unbounded case. Now find a solution to

$$\begin{aligned} Ax &\leq b \\ A^T y &= c \\ y &\geq 0 \\ c^T x &= b^T y \end{aligned}$$

The solution $(\bar{x}, \bar{y})$ is an optimal primal-dual pair of solutions. $\square$

4.1 Simplex Algorithm

A basic overview:

- SIAM/IEEE top 10 algorithms of the century
- To date, no polynomial time variant is known
- Examples showing standard variants require exponential number of steps

Algorithm 1. (Simplex Algorithm) The standard algorithm works with the standard form A (an $m \times n$ matrix) where we are solving the problem

$$\begin{aligned} \min & \quad c^T x \\ Ax &= b \\ x &\geq 0 \end{aligned}$$

We will equivalently denote $x = X$. Let B be an ordered set of indices $\{B_1, ..., B_m\}$from $\{1, ..., n\}$. B is called a **basis header** and determines a basis $\mathbb{B} = A_B$ consisting of columns $A_{B_1}, ..., A_{B_m}$ if $\mathbb{B}$ is non-singular. N denotes the non-basic variables $\{1, ..., n\} \backslash B$. We then have the new algorithm

$$\begin{aligned} \min & \quad C_B^T X_B + C_N^T X_N \\ A_B X_B + A_N X_N &= b \\ X_B \geq 0 & \quad X_N \geq 0 \end{aligned}$$

B is **primal feasible** if $\mathbb{B}^{-1} b \geq 0$. In a general iteration of the (revised) primal simplex algorithm, we have a primal feasible B and vectors

$$X_B = \mathbb{B}^{-1} b \text{ and } D_N = C_N - A_N^T \left(\mathbb{B}^{-1}\right)^T C_B$$

The steps are the following.

(1) [Pricing] If $D_N \geq 0$ then B is optimal and you stop. Otherwise let

$$j = \text{argmin}(D_k : k \in N)$$

where variable X_j is the **entering variable**.

(2) [FTRAN] Solve $\mathbb{B} y = A_j$ (column of A)

(3) [Ratio Test] If $y \leq 0$ then the LP is unbounded and we stop. Otherwise, let

$$i = \text{argmin}(\left[X_B\right]_k / y_k : y_k > 0, k = 1, ..., m)$$

where the variable $\left[X_B\right]_i$ is the **leaving variable**.

(4) [BTRAN] Solve $\mathbb{B}^T z = e_i$ where e_i is the i^{th} unit vector.

(5) [Update] Compute $\alpha_N = -A_N^T z$. Set $B_i = j$. Update X_B (using y) and update D_N (using α_N).

Remark 4.1. At each iteration, the primal solution is $X_B = \mathbb{B}^{-1}b$ and $X_N = 0$. If $X_B \geq 0$ then the primal is feasible. In the dual problem of

$$\begin{array}{lll} \max & & b^T\Pi \\ A^T x & \leq & c \end{array}$$

the dual solution is $\Pi = \left(\mathbb{B}^{-1}\right)^T c_B$ or $\Pi^T = c_B^T\mathbb{B}^{-1}$. The solution is feasible in the dual problem if $A_N^T\Pi \leq C_N$ and $c_N - A_N^T\left(\mathbb{B}^{-1}\right)^T c_B \geq 0$.

The primal objective value is $c_B^T\mathbb{B}^{-1}b$ and the dual objective value is $\Pi^T b = c_B^T\mathbb{B}^{-1}b$.

Problem 4.1. How to get a good implementation?

- How to select an incoming variable?
  - Potential research area
  - No progress since $\sim$ 1990 with the "steepest edge" algorithm
  - Using multicore processors?
- Better Linear Algebra
  - Want # of operations to be proportional to # of non-zeros
  - Mining results from linear algebra
  - Super-sparse LA on GPU?

4.2 Simplex Algorithm Research

Problem 4.2. Here are some interesting research problems:

- How many pivots (iterations) are required to solve an LP?
- Is there a pivot rule where # of pivots is never more that a polynomial in the size of the LP?
- Even stronger: Polynomial in $n + m$?

The following are some noticeable contributions:

1982 Borgwardt

- Polynomial in average case
- 1982 Lanchester Prize

1992 Gil Kalai

- Sub-exponential randomized simplex algorithm
- Expected time of $m^{O(\sqrt{n})}$
- Fulkerson Prize in 1994

2008 Spielman-Teng

- Smoothed analysis of simplex algorithm
- Expected polynomial over small pertubations
- 2009 Fulkerson Prize, 2008 Gödel Prize

2010 Friedmann, Hansen, Zwick

- Sub-exponential *lower* bound for randomized pivot rules
- Random edge: choose incoming variable uniformly at random
  - 2011 STOC Best Paper
  - 2012 Tucker Prize

1972 Klee-Minty Algorithm

- Example where various pivot rules take exponential time
- In n dim, use 2 copies of $(n-1)$-d and takes $2^n - 1$ pivots

Conjecture 4.1. *(Hirsch Conjecture) Bound longest distance between two vertices of a polytope P.*

Algorithm 2. Here is an alternative formulation of the Simplex Algorithm. Using the same notation as in the original section, we do the following:

1. Define the basic solution corresponding to B as $X_N = 0$ and $X_B = \mathbb{B}^{-1}b$
2. Rewrite the restriction as $X_B + \mathbb{B}^{-1}A_N X_N = \mathbb{B}^{-1}b$ (multiply through with $\mathbb{B}^{-1}$) with dual solution $\Pi = (\mathbb{B}^{-1})^T C_B$ or $\Pi^T = C_B^T \mathbb{B}^{-1}$
3. Rewrite the objective by first multiplying $-C_B^T$ to $X_B + \mathbb{B}^{-1}A_N X_N - \mathbb{B}^{-1}b = 0$ and $q = C_B^T X_B + C_N^T X_N$ (the objective value) to get

$$\begin{aligned} 0 &= -C_B^T X_B - C_B^T \mathbb{B}^{-1} A_N X_N + C_B^T \mathbb{B}^{-1} b \\ q &= (C_N^T - C_B^T \mathbb{B}^{-1} A_N) X_N + C_B^T \mathbb{B}^{-1} b \end{aligned}$$

   and get them to the new formulation

$$\begin{aligned} \min \quad & q = (C_N^T - C_B^T \mathbb{B}^{-1} A_N) X_N + C_B^T \mathbb{B}^{-1} b \\ sb. \quad & X_B + \mathbb{B}^{-1} A_N X_N = \mathbb{B}^{-1} b \end{aligned}$$

4. The basic solution $X_N = 0, X_B = \mathbb{B}^{-1}b$ gives the objective $q = C_B^T \mathbb{B}^{-1} b$. If $C_N^T - C_B^T \mathbb{B}^{-1} A_N \geq 0$, then the basic solution is optimal.

Example 4.1. Suppose we have the following LP:

$$\begin{aligned} \max \quad & 5x_1 + 4x_2 + 3x_3 \\ sb. \quad & 2x_1 + 3x_2 + x_3 + x_4 = 5 \\ & 4x_1 + x_2 + 2x_3 + x_5 = 11 \\ & 3x_1 + 4x_2 + x_2 + x_6 = 8 \\ & x_i \geq 0, i = 1, ..., 6 \end{aligned}$$

In the Simplex algorithm, we initialize

$$B = \begin{pmatrix} 4 & 5 & 6 \end{pmatrix}, C_B^T = \begin{pmatrix} 0 & 0 & 0 \end{pmatrix}, \mathbb{B} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

With $\Pi = (\mathbb{B}^{-1})^T C_B = I \cdot 0 = 0$. We then compute

$$C_N^T - C_B^T \mathbb{B}^{-1} A_N = \begin{pmatrix} 2 & 3 & 4 \end{pmatrix}^T$$

Choose X_1 as the entering variable (because we are maximizing here and instead are using argmax) with $q = 5X_1 + 4X_2 + 3X_3 + 0$ and we want to increase the value of X_1. Suppose $X_1 = \delta$. We need (in order to keep all variable non-negative)

$$\begin{aligned} X_4 = 5 - 2\delta \geq 0 &\implies \delta \leq 5/2 \\ X_5 = 11 - 4\delta \geq 0 &\implies \delta \leq 11/4 \\ X_6 = 8 - 3\delta \geq 0 &\implies \delta \leq 8/3 \end{aligned}$$

This is the "Ratio Test" of our original algorithm. So here X_4 is the leaving variable (because 5/2 is the lowest). We then set $X_2 = X_3 = 0$ with

$$X_1 = \frac{5}{2}, X_4 = 0, X_5 = 11 - \frac{20}{2} = 1, X_6 = 8 - \frac{15}{2} = \frac{1}{2}$$

Our new basis is $B = \begin{pmatrix} 1 & 5 & 6 \end{pmatrix}, N = \begin{pmatrix} 2 & 3 & 4 \end{pmatrix}$.

5 Linear Integer Programming

Given rational A, b, c we want to find

$$\max(c^T x : Ax \leq n, x \text{ integer})$$

Some commercial codes include CPlex (IBM), Gurobi, and Xpress (Fico). When we talk about feasibility, we want to know if $Ax \leq b$ with x integral have a solution? Techniques for IP depend heavily on the theory of polyhedra.

Definition 5.1. We define the **integer hull** of P as

$$P_I = Conv\_Hull(P \cap \mathbb{Z}^n)$$

If P is bounded, then P_I is a polyhedron. If C is a rational cone, then $C_I = C$ and we can take all generators to be integer vectors.

Theorem 5.1. *(Meyer's Theorem 1974) If P is a rational polyhedron, then P_I is a polyhedron.*

Proof. Write $P = Q + C$ with Q a polytope and C a cone. We have $C = \{\lambda_1 d_1 + ... + \lambda_s d_s \geq 0\}$ with $d_1, ..., d_s$ integer vectors. Let B be the bounded set

$$B = \{\lambda_1 d_1 + ... + \lambda_s d_s : 0 \leq \lambda_i \leq 1, i = 1, ..., s\}$$

We claim that $P_I = (Q + B)_I + C$. We are done because since $Q + B$ is bounded, $(Q + B)_I$ is a polytope, thus P_I is a polyhedron. To prove this claim, let $p \in P \cap \mathbb{Z}^n$. Then $p = q + c$ for some $q \in Q$ and $c \in C$. It follows that $c = b + c'$ with $b \in B$ and $c' \in C \cap \mathbb{Z}^n$. So $p = q + b + c'$ and $q + b$ is integral. This implies

$$p \in (Q + B)_I + C \implies P_I \subseteq (Q + B)_I + C$$

The other direction is

$$(Q + B)_I + C \subseteq P_I + C = P_I + C_I \subseteq (P + C)_I = P_I$$

□

Problem 5.1. How can we get information about P_I?

An important case is when $P = P_I$ where P is called an integral polyhedral. We would then have

$$\begin{aligned} \text{P is integral} &\iff \text{Each minimal face of } P \text{contains integral vectors} \\ &\iff \max(c^T x : x \in P) \text{ has an integral optimal solution} \\ &\qquad \text{if the maximum is finite} \end{aligned}$$

Note that if $P = \{x : Ax \leq b\}$ is integral, then $\max(c^T x : Ax \leq b, x \text{ integral})$ can be solved in polynomial time. With poly-time LP, we find optimal $\delta = \max(c^T x : Ax \leq b)$ and then a subsystem $A'x \leq b'$ such that each point in the minimal face $F = \{x : A'x = b'\}$ is an optimal solution to the LP. We find the integral solution to $A'x = b'$ using Hermite Normal Form.

Suppose that $\max(w^T x : Ax \leq b) = \delta$ and w is integral. Then every integer solution satisfies $w^T x \leq \lfloor \delta \rfloor$, where we round down to the nearest integer. $w^T x \leq \delta$ is a supporting hyperplane and $w \in \mathbb{Z}^n$. Then every point in $P \cap \mathbb{Z}^n$ satisfies $w^T x \leq \lfloor \delta \rfloor$. (*Chvatal-Gomory Cutting Plane*).

Definition 5.2. If $w^T x \leq \delta$ is valid, then $w^T x \leq \lfloor \delta \rfloor$ is valid for all integer vectors in P. $w^T x \leq \lfloor \delta \rfloor$ is called a Chvatal-Gomory (C-G) **cutting plane** (or cut).

Theorem 5.2. *For $Ax \leq b$, suppose $y \geq 0$ and $y^T A$ is integer valued. Then $(y^T A)x \leq \lfloor y^T b \rfloor$ is a C-G cut.*

Definition 5.3. (Cutting-plane proof) Given $Ax \leq b$, A integer valued, we want to prove that every integer solution satisfies $c^T b \leq t$. The **cutting-plane proof** consists of C-G cuts together with vectors y showing cuts are valid:

Proof. (0) $Ax \leq b$

(1) $\begin{cases} Ax \leq b \\ w^T x \leq \lfloor \delta \rfloor & \text{G-C for (0)} \end{cases}$

(2) $\begin{cases} Ax \leq b \\ w^T x \leq \lfloor \delta \rfloor \\ v^T x \leq \lfloor \gamma \rfloor & \text{G-C for (1)} \end{cases}$

$\vdots$

(k) $c^T x \leq t$ ☐

Theorem 5.3. *(Chvatal (1972)) If $c^T x \leq t$ is valid for all integer solutions to $Ax \leq b$, then it has a cutting-plane proof.*

Definition 5.4. The **Chvatal closure** of a polyhedron P is defined as

$$P' = \{x \in P : x \text{ satisfies all C-G cuts for } P\}$$

Theorem 5.4. *(Schrijver) If P is rational, then P' is a rational polyhedron.*

Proof. (Sketch) Write $P = \{x : Ax \leq b\}$ with A and b integer valued. We obtain a C-G cut for each $y \geq 0$ such that $y^T A$ is integer valued, where

$$\begin{array}{rcl} a_1^T x & \leq & b_1 \\ a_2^T x & \leq & b_2 \\ & \vdots & \\ a_m^T x & \leq & b_m \end{array}$$

and

$$(a_1^T y_1 + a_2^T y_2 + \ldots + a_m^T y_m)x \leq b_1 y_1 + \ldots + b_m y_m$$

The C-G cut is

$$\underbrace{(a_1^T y_1 + a_2^T y_2 + \ldots + a_m^T y_m)}_{w^T} x \leq \underbrace{\lfloor b_1 y_1 + \ldots + b_m y_m \rfloor}_{t}$$

If $y_1 \geq 1$, look at the cut obtained by

$$\begin{array}{rcl} y_1' & = & y_1 - 1 \\ y_2' & = & y_2 \\ & \vdots & \\ y_m' & = & y_m \end{array}$$

The new cut is

$$(w - a_1)^T x \leq t - b_1$$

but every $\bar{x} \in P$ that satisfies the new cut also satisfies $w^T x \leq t$, so we only need C-G cuts such that $0 \leq y \leq 1$ and $y^T A$ integer valued. There are only finitely many such vectors y so we only need finitely many C-G cuts. Hence P' is a polyhedron. ☐

{Freund, Todd, Roundy}

Theorem 5.5. *(Chvatal's Theorem) If P is rational, then there exists k such that $P^{(k)} = P_I$.*

Proof. (Rough Sketch: RE-CHECK FOR FINAL) P_I is a polyhedron defined as $P_I = \{x : Mx \leq d\}$. Let $w^T x \leq t$ be an inequality in $Mx \leq d$. It suffices to show that for some k we have

$$P^{(k)} = (...((P')')...')' \subseteq \{x : w^T x \leq t\}$$

Now let $\delta = \max\{w^T x : x \in P\}$. Thus, $w^T x \leq \lfloor \delta \rfloor$ is a C-G cut. Suppose for large enough k we know that $w^T x \leq q$ is valid for $P^{(k)}$. It suffices to show that for some $k' > k$ we have $w^T x < q$ is valid for $P^{(k')} \implies w^T x \leq q - 1$ is valid for $P^{(k'+1)}$. Let $F = \{x \in P : w^T x = q\}$. By induction on the dimension of the polyhedron, we can assume there exists l such $F^{(l)} = \emptyset$. Applying these cutting planes to the polyhedron $P \cap \{x : w^T x \leq q\}$ we obtain a polyhedron such that $w^T x < q$ is valid. □

Definition 5.5. The smallest k such that $P^{(k)} = P_I$ is called the **Chvatal rank** of P.

Example 5.1. (Matchings in graphs) Consider $w = \{w_e : e \in E(G)\}$,

$$\delta(v) = \{e \in E(G) : e \text{ has exactly one end meeting } v\}$$

and the **integer program** (IP)

$$\begin{aligned} \max \quad & \sum(w_e X_e : e \in E(G)) \\ \text{sb. to} \quad & \sum(X_e : e \in \delta(v)) \leq 1, \forall v \in V(G) \\ & X_e \geq 0, X_e \text{ integer}, \forall e \in E(G) \end{aligned}$$

Note that if $S \subseteq V$ and

$$\gamma(S) \equiv \{e : e \text{ has both ends in } S\}$$

Then

$$\sum(X_e : e \in \gamma(S)) \leq \frac{|S| - 1}{2}, \forall S \subseteq V, |S| \text{ odd}$$

So let $S \subseteq V$, $|S|$ odd and add the following inequalities

$$\begin{aligned} \sum(X_e, e \in \delta(v)) \leq 1 \qquad & \forall v \in S \\ -X_e \leq 0 \qquad & \forall e \in \delta(S) \end{aligned}$$

So

$$\begin{aligned} \sum(2X_e : e \in \gamma(S)) \leq |S| \quad & \implies \quad \sum(X_e : e \in \gamma(S)) \leq \frac{|S|}{2} \\ & \implies \quad \text{C-G cut is } \sum(X_e : e \in \gamma(S)) \leq \frac{|S| - 1}{2} \end{aligned}$$

Thus, $P' = P_I$ for matchings where

$$P = \begin{cases} \sum(X_e : e \in \delta(v)) \leq 1, & \forall v \in V(G) \\ X_e \geq 0, & \forall e \in E(G) \end{cases}$$

Remark 5.1. Consider the previous example. In the dual problem of the primal problem (P) with objective function $\max \sum(w_e X_e : e \in E)$, the **dual variables** are

$$(y_v : v \in V), (y_S : S \subseteq V, |S| \text{ is odd})$$

and the dual problem would be

$$\begin{aligned} \min \qquad & \sum(1 y_V : v \in V) + \sum\left(\left(\frac{|S| - 1}{2}\right) y_S : S \subseteq V, |S| \text{ odd}\right) \\ y_V \geq 0, y_S \geq 0 \qquad & \forall v \in V, \forall S \\ y_u + y_v + \sum(y_S : e \in \gamma(S), X \subseteq V, |S| \text{ odd}) \qquad & \forall e = (u, v) \in E \end{aligned}$$

Definition 5.6. $Ax \le b$ is called **totally dual integral** (TDI) if $\min(y^T b : y^T A = w^T, y \ge 0)$ has an integral optimal solution for every integral w such that the $\min$ exists.

Aside. Total dual integrality is *not* a property of polyhedron. For any rational $Ax \le b$, there is a large enough k such that $\frac{Ax}{k} \le \frac{b}{k}$ is TDI.

Proposition 5.1. (*Giles & Pulleyblank*) For any rational P, there exists a TDI system $Ax \le b$ such that A is integer valued and $P = \{x : Ax \le b\}$

Note 10. A integer, $Ax \le b$ TDI, then $P' = \{x : Ax \le \lfloor b \rfloor\}$.

Definition 5.7. A rational polyhedron $P = \{x : Ax \le b\}$ is an **integer polyhedron** if the primal LP $\max(w^T x : Ax \le b)$ always has an integral solution $x^* \iff$ every minimal face of P contains integral vectors.

Theorem 5.6. *(Edmonds & Giles) Rational P is an integer polyhedron $\iff$ every supporting hyperplane of P contains integral vectors.*

Proof. ($\implies$) Easy, since intersection of a supporting hyperplane of P contains integral vectors.

($\impliedby$) Follows from Integer Farkas Lemma ☐

Theorem 5.7. *Rational (polyhedron) P is an integer polyhedron $\iff$ for each integral w such that $\max(w^T x : Ax \le b)$ exists, the value $\max(w^T x : Ax \le b)$ is an integer.*

Proof. ($\implies$) Easy, since x^* is integer and so $w^T x^*$ is integer.

($\impliedby$) Follows from above theorem and the fact that if w has relatively prime integer components, then $w^T x = \delta$ has an integer solution for any integer δ. (Hint for A4 Q4) ☐

Aside. $w = \begin{pmatrix} w_1 & w_2 & \cdots & w_n \end{pmatrix}^T$ with $w_1, w_2, ..., w_n$ relatively prime integers $\iff \exists$ integers $\lambda_1, ..., \lambda_n$ such that $\lambda_1 w_1 + ... + \lambda_2 w_2 = 1$ (Hint for A4 Q4).

Remark 5.2. (TDI $Ax \le b$) For every integral w such that the dual LP $\min(y^T b : y^T A = w^T, y \ge 0)$ has an optimal solution, the dual has an integral optimal solution.

Theorem 5.8. *(Edmonds & Giles) $Ax \le b$ TDI and b integral, then $P = \{x : Ax \le b\}$ is an integer polyhedron.*

Problem 5.2. What do TDI systems look like? When is $Ax \le 0$ TDI?

Solution. This is where $\max(w^T x : Ax \le 0) = \min(y^T 0 : y^T A = w, y \ge 0)$ and hence $Ax \le 0$ is TDI $\iff \forall$ integral $w \in \{w : y^T A = w, y \ge 0\}$ there exists integral $y \ge 0$ with $y^T A = w$. Now $A^T y = w^T \iff w \in Cone(a_1, ..., a_m)$. So $Ax \le 0$ is TDI $\iff$ for each integral $w \in Cone(a_1, ..., a_m)$ there exists integers $y_1 \ge 0, ..., y_m \ge 0$ with

$$w = y_1 a_1 + ... + y_m a_m$$

Definition 5.8. A set of vectors $\{a_1, ..., a_m\}$ is called a **Hilbert basis** if every integer vector in $Cone(a_1, ..., a_m)$ is a non-negative integer combination of $a_1, ..., a_m$. That is, $Ax \le 0$ is TDI $\iff$ rows of A are a Hilbert basis.

Theorem 5.9. *$Ax \le b$ is TDI $\iff \forall$ faces $F = \{x : A^0 x = b^0, A' x \le b'\}$ the rows of A^0 form a Hilbert basis (HB).*

Proof. Follows from complementary slackness conditions (CSS). ☐

Theorem 5.10. *If C is a rational cone, then $\exists$ an integral H.B. that generates C.*

Proof. Consider $C = Cone(d_1, ..., d_k)$ with $d_1, ..., d_k$ integral vectors. Let $H = \{a_1, ..., a_t\}$ be the set of integral vectors in the bounded set

$$\{\lambda_1 d_1 + ... + \lambda_k d_k : 0 \le \lambda_i \le 1, i = 1, ..., k\}$$

Note $H \subseteq C$ and $d_1, ..., d_k \in H$. So H generates C. Let $b \in C \cap \mathbb{Z}^n$. Then $b = \mu_1 d_1 + ... + \mu_k d_k$ for some $\mu_i \ge 0$. Write this as

$$\underbrace{b}_{\in \mathbb{Z}^n} = \underbrace{\lfloor \mu_1 \rfloor d_1 + ... + \lfloor \mu_k \rfloor d_k}_{\in \mathbb{Z}^n} + \underbrace{(\mu_1 - \lfloor \mu_1 \rfloor) d_1 + ... + (\mu_k - \lfloor \mu_k \rfloor) d_k}_{\in H}$$

Since b is a non-negative combination of vectors in H, H is a Hilbert basis. ☐

Theorem 5.11. *(Giles + Pulleblank) P is rational $\implies \exists$ TDI $Ax \leq b$ with A integral and $P = \{x : Ax \leq b\}$.*

Remark 5.3. In a min-max, then dual LP has many variables. Maybe exponentially many. How many must be non-zero? Recall that Caratheodory's Theorem says that if $b \in Cone(a_1, ..., a_m)$ then b is a non-negative combination of linearly independent vectors from $a_1, ..., a_m$. Is there an integer Caratheodory's Theorem?

Theorem 5.12. *If C is a pointed cone and $H = \{a_1, ..., a_m\}$ is an integer Hilbert basis, then if $b \in C \cap \mathbb{Z}^n$ then b is a non-negative combination of $2n - 1$ vectors from $\{a_1, ..., a_m\}$. Sebo showed $2n - 1$ can be reduced to $2n - 2$.*

Oddly enough, there exists an example $n = 6$ that requires 7 vectors.

5.1 Ellipsoid Method

Attributed to N. Shor (1970), Judin-Nemirovski (1976). Began with the development of the poly-time LP algorithm by L. Khachian (1979).

Problem 5.3. The main problem is: Does $Ax \leq b$ have a solution?

If $P = \{x : Ax \leq b\} \neq \emptyset$ then we may assume P is of **full dimension** and **bounded** which we define below:

Bounded: Add a box $-2^D \leq x \leq 2^D$ and let ν = vertex complexity

Full dimension: Perturb the initial system to get $P^\epsilon = \{x : Ax \leq b + \epsilon 1\}$ where we have small enough ϵ (but of poly-size) such that $P = \emptyset \iff P^\epsilon = \emptyset$. If $P \neq \emptyset$ and of full dimension, $\exists$ vertices of P say $v_0, ..., v_n$ affinely independent and of size at most 2^D. This implies that

$$\begin{aligned} Vol(P) = Vol(Conv\_Hull\{v_0, ..., v_n\}) &= \frac{1}{n!} \det \begin{bmatrix} 1 & 1 & \cdots & 1 \\ v_0 & v_1 & \cdots & v_n \end{bmatrix} \\ &\geq n^{-n} 2^{-2\nu} \geq 2^{-2n\nu} \end{aligned}$$

Algorithm 3. *(Ellipsoid Method)*

1) Construct a sequence of ellipsoids $E_0, E_1, ..., E_k, ...$ where E_0 is a ball of radius 2^ν centered at 0 and $P \subseteq E_0$.

2) Suppose we have E_k. If its center is $c_k \in P$, we are finished. Otherwise, there is an inequality $a_i^T x \leq b_i$ in $Ax \leq b$ with $a_i^T x_k > b_i$. Consider the half-ellipsoid

$$E_k \cap \{x : a_i^T x \leq b_i\}$$

Let E_{k+1} be the ellipsoid with minimum volume that contains the half-ellipsoid. We can show $Vol(E_{k+1}) \leq e^{-\frac{1}{2n}} Vol(E_k)$. After polynomially many steps, we have

$$Vol(P) \leq Vol(E_k) \leq 2^{-2n\nu} \implies P \neq \emptyset$$

Note 11. Consider an ellipsoid $E \subseteq \mathbb{R}^n$. It has the form

$$E = \{x : (x - c)^T M^T M (x - c) \leq 1\}$$

where M is positive semi-definite.

Remark 5.4. (VERY Important Observation) The method does not need to know $Ax \leq b$ explicitly. We only need to find an inequality $a_i^T x \leq b_i$ with $a_i^T c_k$ and $a_i^T x \leq b_i$ for any $x \in P$. In fact, there are many ways to be given a polyhedron:

- $\{x : Ax \leq b\}$
- $Conv\_Hull\{a_1, ..., a_m\} + Cone\{b_1, ..., b_k\}$
- P_I for some polyhedron $\{x : Ax \leq b\}$
- Graph $\implies$ Polyhedron is convex hull of matchings (stable sets, cuts, tours)

Summary 4. What are some problems given $P \subseteq \mathbb{R}^n$?

- Membership: Given $y \in \mathbb{R}^n$, is $y \in P$?

- Separation: Given $y \in \mathbb{R}^n$, either assert $y \in P$ or provide an inequality $c^T x \leq \delta$ valid for P with $c^T x > \delta$
- Optimization: Given $c \in \mathbb{R}^n$, solve $\max\{c^T x : x \in P\}$

Theorem 5.13. *(Groetchel-Loviesv & Schrijver) For rational polyhedra optimization $\iff$ separation. That is, if we have a polynomial time algorithm for separation then we have a polynomial time algorithm for optimization.*

Example 5.2. (Subtour Polytope for TSP) Given $G = (V, E)$ and $X = (X_e : e \in E)$, then the polytope is parametrized by

$$\begin{aligned} 0 \leq X_e \leq 1 \qquad & \forall e \in E \\ \sum (X_e : e \in \delta(v)) = 2 \qquad & \forall v \in V \\ \sum (X_e : e \in \delta(S)) \geq 2 \qquad & \forall X \subseteq V, \emptyset \neq S \neq V \end{aligned}$$

The separation problem is: Given x^*, is x^* in the subtour polytope? You have to merely check if

$$\begin{aligned} 0 \leq X_e \leq 1 \qquad & \forall e \in E \\ \sum (X_e : e \in \delta(v)) = 2 \qquad & \forall v \in V \end{aligned}$$

The edge capacities are $(X_e^* : e \in E)$. A set $\delta(S)$ is called a cut. We generally want to solve the minimum capacity cut.

(Hint for A3Q6) If $C \subseteq \mathbb{R}^n$ and C is pointed with $H \subseteq C$ a Hilbert basis, let $w \in C \cap \mathbb{Z}^n$ which implies that w can be written as a non-negative integer combination of at most $2n - 1$ vectors from H. Suppose that we want to solve the program

$$\begin{aligned} \max \qquad & \sum (y_i : i = 1, ..., n) \\ \text{sb. to} \qquad & w = y_1 a_1 + ... + y_m a_m \\ & y_i \geq 0, y \text{ int.} \end{aligned}$$

Since C is pointed, we have an optimal y^* and the basic solution has at most n of the y_i^*'s are positive. Let

$$\begin{aligned} w' = (y_1 - \lfloor y_1 \rfloor) a_1 + ... + (y_m - \lfloor y_m \rfloor) a_m \quad & \implies \quad w' = w - \lfloor y_1 \rfloor a_1 + ... + \lfloor y_m \rfloor a_m \\ & \implies \quad w' \in \mathbb{Z}^n \cap C \\ & \implies \quad \exists \lambda_1, ..., \lambda_m \geq 0 \text{ int. with} \\ & \qquad\quad w' = \lambda_1 a_1 + ... + \lambda_m a_m \\ & \implies \quad w = \lfloor y_1 \rfloor a_1 + ... + \lfloor y_m \rfloor a_m \\ & \qquad\quad + \lambda_1 a_1 + ... + \lambda_m a_m \end{aligned}$$

At most n of the $\lfloor y_i \rfloor' s$ are positive and at most $n - 1$ of the $\lambda_i' s$ are positive since $\lambda_i + \lfloor y_i \rfloor$ gives an LP solution.

5.2 Cutting Plane Method

This applies the separation method. That is for a polyhedron P, and $x^* \in \mathbb{R}^n$, we either:

- Assert $x^* \in P$
- Demonstrate an inequality $w^T x \leq \delta$ valid for P but $w^T x^* > \delta$

In the cutting plane method, we begin with a "simple" polyhedron Q such that $Q \supseteq P$. We then use the simplex method to find the optimal x^* for

$$\max(w^T x : x \in Q)$$

If $x^* \in P$ then x^* is optimal for $\max(w^T x : x \in P)$ and we stop. Otherwise the separation algorithm returns a separating hyperplane $d^T x \leq t$ that is violated by x^* but valid for Q. Replace Q by $Q \cap \{x : d^T x \leq t\}$ and repeat.

The practical idea is to seek separating hyperplanes that define facets of P. When possible, we also try to find many separating hyperplanes and add them all at once to Q.

Example 5.3. (Traveling Salesman Problem... again) Consider n cities represented by vertices V. Let $E = \{(i,j) : 1 \leq i,j \leq n\}$ be the edges. A TSP tour is a cycle that meets over cities $T \subseteq E$ and

$$X_e^{(T)} = \begin{cases} 1 & e \in T \\ 0 & \text{otherwise} \end{cases}$$

The TSP polytope is the convex hull of all tour vectors. We do not know a linear description of TSP polytopes, but there are known classes of facets. Some subtour elimination inequalities are:

- $\sum(X_e : e \in \delta(S)) \geq 2$
- Define facets of the TSP polytope

An initial Q could be

$$\begin{aligned} 0 \leq \quad X_e \quad &\leq 1 \forall e \in E \\ \sum(X_e : e \in \delta(v)) \quad &= 2 \forall v \in V \end{aligned}$$

and we want to minimize $\sum(c_e X_e : e \in E)$. We use the simplex algorithm to obtain x^* and from x^* we obtain a graph G^* with vertices V and edges $(e \in E : X_e^* > 0)$. If G^* is not connected, we use S_i to define subtour elimination inequalities.

The general step is as follows: Compute optimal solution x^* to the LP. If x^* is a tour vector, stop and x^* is an optimal TSP tour. Otherwise, find inequalities satisfied by all tours, but violated by x^*. Add the inequalities to the contraints of the LP and repeat.

In subtour separation, we create a graph G^* with edges $(e \in E : X_e^* > 0)$. In G^*, we find a minimal cut that is a set that minimizes

$$\sum(X_e^* : e \in \delta(s)) = Z^*$$

If $Z^* < Z$ add

$$\sum(X_e : e \in \delta(s)) \geq Z$$

to the LP. Otherwise, x^* satisfies all subtour elimination constraints.

Next step: Combinatorial inequalities [Chvatal (1972), Grotschel-Padberg (1979)].

5.3 Column Generation Method

Example 5.4. (Cutting-Stock Problem) Imagine you have some roll of width r made or paper, metal, etc. The customer widths of b_k rolls of width w_k, where $1 \leq k \leq m$. We define a cutting pattern

$$\begin{aligned} a_p \quad \equiv \quad & a_{p_1} \text{ rolls of width } w_1 \\ & a_{p_2} \text{ rolls of width } w_2 \\ & \vdots \\ & a_{p_m} \text{ rolls of width } w_m \end{aligned}$$

where to be feasible, we need

$$a_{p_1} w_1 + a_{p_2} w_2 + ... + a_{p_m} w_m \leq r$$

The optimization problem is to meet customer demands using as few rolls as possible. If $(a_p : p \in P)$ is the set of all feasible solutions, and $X_p \equiv$ # of rolls of pattern a_p that we cut. We want to minimize

$$\begin{aligned} \min \quad & \sum(X_p : p \in P) \\ \text{sb. to} \quad & \sum(a_p X_p : p \in P) = b \\ & X_p \geq 0 \forall p \in P \end{aligned}$$

Begin with $P^* \subseteq P$ such that the LP is feasible. That is, solve the LP*

$$\begin{array}{ll} \min & \sum(X_p : p \in P^*) \\ \text{sb. to} & \sum(a_p X_p : p \in P^*) = b \\ & X_p \geq 0 \forall p \in P^* \end{array}$$

If we optimize LP* to obtain x^*, we consider the dual LP with variables $y_1, ..., y_m$ and LP

$$\begin{array}{ll} \max & b_1 y_1 + ... + b_m y_m \\ \text{sb. to} & a_{p_1} y_1 + ... + a_{p_m} y_m \leq 1, \forall p \in P \end{array}$$

When we find x^* we also find the optimal dual solution y^* BUT y^* might not satisfy the dual constraints for $p \in P \backslash P^*$.

The **Knapsack problem** is the LP

$$\begin{array}{ll} \max & y_1^* z_1 + ... + y_m^* z_n \\ \text{sb. to} & w_1 z_1 + ... + w_m z_m \leq r \\ & z_i \geq 0, \forall i, z_i \in \mathbb{Z} \end{array}$$

If the object value is > 1, add pattern to LP. If ≤ 1, then x^* and y^* are optimal.

6 Non-Linear Optimization

The general model for non-linear programming (NLP) is we have functions $f, g_1, g_2, ..., g_m : \mathbb{R}^n \mapsto \mathbb{R}$. We want to solve

$$\begin{array}{ll} \min & f(x) \\ \text{sb. to} & g_i(x) \leq 0, i = 1, ..., m \end{array}$$

Any LP model can be written in this form:

$$\begin{array}{rcl} \max\ c^T x & & \min\ -c^T x \\ Ax = b & \rightarrow & Ax - b \leq 0 \\ x \geq 0 & & -Ax + b \leq 0 \\ & & -x \leq 0 \end{array}$$

and alternatively,

$$\begin{array}{rcl} \min\ -c^T x & & f(x) = -c^T x \\ Ax - b \leq 0 & \rightarrow & g_1(x) = a_1^T x - b_1 \\ -Ax + b \leq 0 & & g_2(x) = a_2^T x - b_2 \\ -x \leq 0 & & \vdots \end{array}$$

NLP is hard. For example, if we had $x_1^2 = x_1 \iff x_1 \in \{0, 1\}$ and

$$\begin{array}{rcl} \max\ c^T x & & \min\ -c^T x \\ Ax = b & \rightarrow & Ax \leq b \\ x \in \{0, 1\} & & x_i^2 - x_i \leq 0 \\ & & -x_i^2 + x_i \leq 0 \end{array}$$

6.1 Fermat's Last Theorem

NLP is "wild" (cf. Bill Cook). Here's an example:

Example 6.1. (Fermat's Last Theorem) The statement of **Fermat's Last Theorem** is that there does not exist integers $a, b, c \geq 1$ and integer $k \geq 3$ such that $a^k + b^k = c^k$. Consider the NLP model with $n = 4, m = 4$. Specifically, we have variables $x = (x_1, x_2, x_3, x_4)^T \in \mathbb{R}^n$,

$$f(x) \equiv (x_1^{x_4} + x_2^{x_4} - x_3^{x_4})^2 + (\sin(\pi x_1))^2 + (\sin(\pi x_2))^2 + (\sin(\pi x_3))^2 + (\sin(\pi x_4))^2$$

This is a sum of squares, and so $f(x) \geq 0$. We also have $f(x) = 0 \iff$ all terms are equal to 0 and

$$x_1^{x_4} + x_2^{x_4} = x_3^{x_4}, \sin(\pi x_i) = 0, i = 1, ..., 4$$

Note that $\sin(\pi x_i) = 0 \implies x_i$ is an integer. We have

$$\begin{aligned} g_1(x) &= 1 - x_1 \\ g_2(x) &= 1 - x_2 \\ g_3(x) &= 1 - x_3 \\ g_4(x) &= 3 - x_4 \end{aligned}$$

The NLP model has a solution with $f(x) = 0 \iff$ Fermat's Last Theorem is not true.

6.2 Convex Optimization

Definition 6.1. A function $f : \mathbb{R}^n \mapsto \mathbb{R}$ is called **convex** if $\forall x^1, x^2 \in \mathbb{R}^n$ and $0 \leq \lambda \leq 1$ we have

$$f(\lambda x^1 + (1 - \lambda)x^2) \leq \lambda f(x^1) + (1 - \lambda)f(x^2)$$

The **epigraph** of f is defined as

$$epi(f) = \{(x, y) : y \geq f(x)\}$$

Note 12. f is convex $\iff$ $epi(f)$ is a convex set.

Remark 6.1. IF $g : \mathbb{R}^n \mapsto \mathbb{R}$ is convex, then $\{x : g(x) \leq 0\}$ is convex.

In the best of cases, the g_i's in an NLP are convex. An attack on NLP is to try to exploit LP theory.

Definition 6.2. $g_i(x) \geq 0$ is called **tight** at $\bar{x}$ if $g_i(x) = 0$.

The goal is to find condition that allow us to replace tight constraints by a linear constraint to obtain a relaxation of the NLP.

Definition 6.3. Let $g : \mathbb{R}^n \mapsto \mathbb{R}$. Then $s \in \mathbb{R}^n$ is a **subgradient** of g at $\bar{x}$ if $\forall x \in \mathbb{R}^n$ we have

$$g(\bar{x}) + s^T(x - \bar{x}) \leq g(x)$$

Let $h(x) \equiv g(\bar{x}) + s^T(x - \bar{x})$. Since $\bar{x}$ is a fixed vector, then $h(x) = s^T x + \beta$ for some constant β, where h is some affine function. We want to focus on the linear objective

$$\begin{array}{ll} \min & c^T x \\ \text{sb. to} & g_i(x) \leq 0, i = 1, ..., M \end{array}$$

Given feasible $\bar{x}$, is $\bar{x}$ optimal? We first make the assumptions:

(h_1) g_i convex for all i.

(h_2) g_i differentiable.

(h_3) $\exists$ a feasible $\bar{x}$ such that $g_i(\bar{x}) < 0$ for all i. Such an $\bar{x}$ is called a **Slater point**.

Theorem 6.1. *(Karush-Kuhn-Tucker) Suppose that (h_1), (h_2) and (h_3) hold for the NLP model. Let $\bar{x} \in \mathbb{R}^n$ be feasible for the NLP. Let I be the indices of the tight constraints for $\bar{x}$. Then $\bar{x}$ is an optimal solution to the NLP if an only if $-c \in Cone(\nabla g_i(\bar{x}) : i \in I)$.*

7 Convex Optimization

See: http://www2.isye.gatech.edu/~nemirovs/Lect_ModConvOpt.pdf

Example 7.1. We want to solve

$$\begin{array}{ll} \text{min} & -x_1 - x_2 \\ \text{sb. to} & g_1(x) = -x_1 + x_2^2 \leq 0 \\ & g_2(x) = -x_2 + x_1^2 \leq 0 \\ & g_3(x) - -x_1 + \dfrac{1}{2} \leq 0 \end{array}$$

A Slater point in this convex set is $\hat{x} = (\frac{3}{4}, \frac{3}{4})$ and one of the corners is at $\bar{x} = (1, 1)$. Now

$$\nabla g_1(x) = \left(\frac{\partial g_1(x)}{\partial x_1}, \frac{\partial g_1(x)}{\partial x_2} \right) = (-1, 2x_2) \implies \nabla g_1(\bar{x}) = (-1, 2)$$

and similarly $\nabla g_2(\bar{x}) = (2, -1)$. Our subgradient is

$$g_1(\bar{x}) + s^T(x - \bar{x}) = s^T x - s^T \bar{x} = -x_1 + 2x_2 - 1$$

So we can replace $g_1(x) \leq 0$ by $-x_1 + 2x_2 \leq 1$.

Summary 5. Optimization in General

- Courses @ UW
  - CO 450 (Combinatorial Optimization) Fall 2014; taught by Bill Cook
    - Combinatorial Optimization
    - Networks, Matchings, Matroids
  - CO 452 (Integer Programming) Winter 2015; taught by Ricardo Fukasawa
  - CO 463 (Convex Optimization) Fall 2014
  - CO 466 (Continuous Optimization) Winter 2015
  - CO 471 (Semi-definite Programming) Spring 2014; taught by L. Tuncel
  - Topics Courses in Optimization (Computational Course) Winter 2015; taught by Bill Cook
- Graduate School
  - Operations Research / Industrial Engineering
  - Business Schools
  - Applied Maths
  - ACO - Algorithms, Combinatorics, and Optimization
- CO in UW
  - NLP - Tuncel, Vauasis, Coleman, Wolkowicz
  - IP - Fukasawa, [Sanita]
  - Combinatorial, Algorithmic - Sanita, J. Koenemann, C. Swamy, J. Cheriyan
  - Min-Max Theorems - B. Guenin, [J. Geelan]
- Canadian Graduate Studies in CO
  - UW > McGill (Bruce Shepherd) ∼ UBC (Tom McCormick)
- US Graduate Studies in CO

- MIT OR Center (Sloan Business School)
- Georgia Tech ISysE
- Cornell, Stanford, Berkeley AC
- Wisconsin-Madison, Michigan, Northwestern, Columbia, Carnegie-Mellon (Business / ACO)

Index