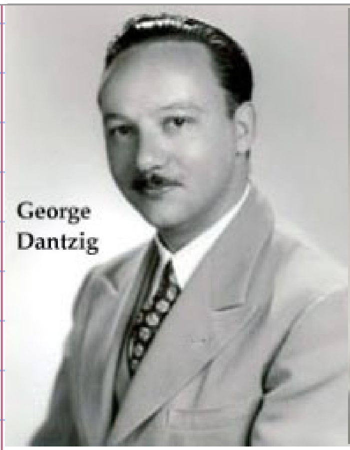
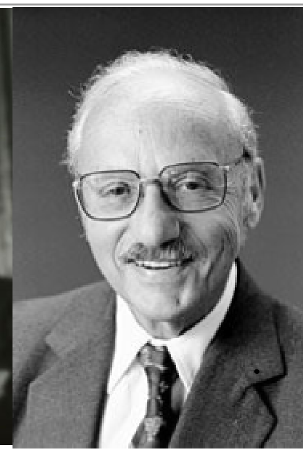
Linear Programming: The Simplex Method I



George Dantzig
1914 – 2005

<u>Def</u>: If f, c_i $(i \in \mathcal{E} \cup \mathcal{I})$ are all affine functions, the constraint optimization problem is called a linear program (LP).

Easy fact: Any LP can be recast as a LP in the standard form:

$$\min c^T x \quad \text{s.t. } Ax = b,\ x \geq 0.$$

Here is how:

Given

$$\min_{x \in \mathbb{R}^n} c^T x + b \quad \text{s.t.} \quad \sum_{j=1}^{n} A_{ij} x_j \leq b_i,\ i \in \mathcal{I}$$
$$\sum_{j=1}^{n} A_{ij} x_j = b_i,\ i \in \mathcal{E}$$

In matrix notations:

$$A^{\mathcal{I}} x \leq b^{\mathcal{I}}$$
$$A^{\mathcal{E}} x = b^{\mathcal{E}}$$

(1) we can always drop the constant b in the objective, as its presence cannot affect the minimizer(s).

(2) For any $i \in \mathcal{I}$, introduce the slack variable z_i into the problem and rewrite each inequatity constraint as

$$\sum_{j=1}^{n} A_{ij} x_j + z_i = b_i \quad , \quad z_i \geq 0$$

So, in matrix notations, the inequality constraints become:

$$A^{\mathcal{I}} x + z = b^{\mathcal{I}} \quad , \quad z \geq 0 \quad \text{OR} \quad \begin{bmatrix} A^{\mathcal{I}} & I \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} = b^{\mathcal{I}}, \quad z \geq 0$$

So far, the general LP is recast as

$$\min_{\begin{bmatrix} x \\ z \end{bmatrix} \in \mathbb{R}^{n+|\mathcal{I}|}} \begin{bmatrix} c \\ 0 \end{bmatrix}^T \begin{bmatrix} x \\ z \end{bmatrix} \quad \text{s.t.} \quad \begin{bmatrix} A^{\mathcal{I}} & I \\ A^{\mathcal{E}} & 0 \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} = \begin{bmatrix} b^{\mathcal{I}} \\ b^{\mathcal{E}} \end{bmatrix}, \quad \begin{matrix} z \geq 0 \\ (x - \text{free}) \end{matrix}$$

This form is still not quite standard, since not all variables are constrained to be non-negative.

Here is the last trick:

(3) write $x = x^+ - x^-$, $x^+ = \max(x, 0) \geq 0$, $x^- = \max(-x, 0) \geq 0$.
We can then further recast the LP as:

$$\min_{\begin{bmatrix} x^+ \\ x^- \\ z \end{bmatrix}} \begin{bmatrix} c \\ -c \\ 0 \end{bmatrix}^T \begin{bmatrix} x^+ \\ x^- \\ z \end{bmatrix} \quad \text{s.t.} \quad \begin{bmatrix} A^{\mathcal{I}} & -A^{\mathcal{I}} & I \\ A^{\mathcal{E}} & -A^{\mathcal{E}} & 0 \end{bmatrix} \begin{bmatrix} x^+ \\ x^- \\ z \end{bmatrix} = \begin{bmatrix} b^{\mathcal{I}} \\ b^{\mathcal{E}} \end{bmatrix}, \quad \begin{bmatrix} x^+ \\ x^- \\ z \end{bmatrix} \geq 0,$$

which is a LP in the standard form.

Terminologies:

We say that a LP is infeasible if the feasible region is empty.

We say that a LP is unbounded if the objective function is unbounded below on the feasible region, i.e. $\exists\, x^k$ s.t. $Ax^k = b$, $x^k \geq 0$ and $c^T x^k \to -\infty$ as $k \to \infty$.

Ex: Give a 2-D example for each. (Hint: it's easy.)

For a standard form LP

$$\min c^T x \quad \text{s.t.} \quad \underset{m\times n}{A} x = b, \; x \geq 0,$$

we may assume

$$\text{rank}(A) = m < n.$$

Otherwise the system contains redundant rows, or is infeasible, or defines a unique point.

<u>Optimality conditions</u> Being a convex problem, KKT guarantees global optimality.

$$\mathcal{L}(x,\lambda,s) = c^T x - \lambda^T(Ax-b) - s^T x$$

KKT : $A^T\lambda + s = c$, $Ax = b$, $x \geq 0$, $s \geq 0$, $\underbrace{x_i s_i = 0, \; i=1,\cdots,n}_{\text{can be rewritten as}}$.

$x^T s = 0$ as $x \geq 0, s \geq 0$

OR " $0 \leq x \perp s \geq 0$ ".

Assume (x^*, λ^*, s^*) satisfies KKT,

$$c^T x^* = (A^T\lambda^* + s^*)^T x^* = (Ax^*)^T \lambda^* = b^T \lambda^*. \quad - (*)$$

$$\underbrace{s^{*T} x^*}_{=0}$$

We know from a general result (on "KKT meets convexity") that the KKT point x^* must be a global minimizer of the LP (a convex problem). But let's prove it from scratch:

Let $\bar{x}$ be any feasible point, so $A\bar{x}=b$, $\bar{x}\geqslant 0$. Then:

$$c^T\bar{x} \underset{\text{KKT}}{\overset{\uparrow}{=}} (A^T\lambda^*+s^*)^T\bar{x} = \lambda^{*T}\underbrace{A\bar{x}}_{=b} + \underset{0}{\overset{\vee\!/}{\bar{x}^T}}\underset{0}{\overset{\vee\!/}{s^*}}{}^{\nwarrow \text{KKT}} \geq b^T\lambda^* \underset{(*)}{\overset{\uparrow}{=}} c^Tx^*$$

(The relevance of convexity is hidden by linearity in this proof!)
Good or bad?

We can say more: $\bar{x}$ is optimal $\Leftrightarrow$ $\bar{x}^Ts^*=0$.

The dual of the standard form LP is:

$$\max_{\lambda,s}\ b^T\lambda \quad \text{s.t. } A^T\lambda+s=c,\ s\geqslant 0.$$

} Ex: Derive it.

dual variables are the Lagrange multipliers of the primal problem

How are the primal and dual LPs related?

Note that the dual can be written as $\boxed{\max_{\lambda} b^T\lambda \text{ st } A^T\lambda\leq c.}$

Let's write down its KKT conditions. $\mathcal{L}(\lambda,x)=-b^T\lambda-x^T(c-A^T\lambda)$, $\nabla_\lambda\mathcal{L}(\lambda,x)=-b+Ax$

KKT conditions: $Ax=b$, $A^T\lambda\leq c$, $x\geqslant 0$, $x_i(c-A^T\lambda)_i=0$, $i=1,\dots,n$

Identical to the KKT conditions of the primal if we define $s=c-A^T\lambda$!

Consequence:

$$\boxed{\begin{array}{l}\text{The optimal Lagrange multipliers of the primal problem are the optimal variables in the dual problems.}\\[1em] \text{The optimal Lagrange multipliers of the dual problem are the optimal variables in the primal problems.}\end{array}}$$

Of course, if λ^*, x^* satisfy the KKT conditions, then λ^* is a solution of the dual problem. (we can use the general "KKT meets convexity" result again, or prove it directly as in the primal case.)

Also, the dual of the dual is the primal. (Check it!)

And never forget that the dual problem, by design, satisfies the weak duality property. But in case you did forget, here is a direct proof specific for LPs:

Let x be primal feasible (ie. $Ax = b$, $x \geq 0$)
λ be dual feasible (ie. $A^T\lambda \geq c$).

Then

$$c^Tx - b^T\lambda = c^Tx - \lambda^T Ax = (c - A^T\lambda)^T x \geq 0. \quad \text{Q.E.D.}$$

Here is a fundamental result for LP. (It does not quite follow directly from the general results on duality in the previous chapter.)

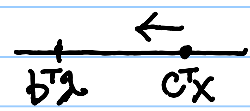
Theorem (Strong Duality)

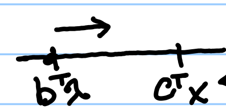
(i) If either the primal or dual LP has a (finite) solution, then so does the other, and the objectives are equal.

(ii) If either the primal or dual is unbounded, then the other problem is infeasible.

We have basically proved (i). Can you put the pieces together? (Hint: it's easy.)

Proof of (ii):





if it exists, c^Tx upper bounds $b^T\lambda$, so $b^T\lambda$ cannot $\uparrow +\infty$.

if it exists, $b^T\lambda$ lower bounds c^Tx, c^Tx cannot $\downarrow -\infty$.

Q.E.D.

Note: The strong duality theorem does not rule out the lurking possibility that both the primal and dual problems are infeasible. And, in fact, it can happen:

$$\min\ \overset{-1}{c^T}x \qquad\qquad \max\ b^T\lambda$$

$$\text{s.t. } \underset{0}{A}x = \underset{1}{b},\ x \geq 0 \qquad\qquad \text{s.t. } \underset{0}{A}\lambda \leq \underset{-1}{c}$$

Geometry of the feasible set $\Omega = \{x \in \mathbb{R}^n : Ax = b,\ x \geq 0\}$.

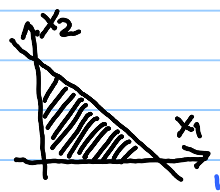
A: $m \times n$, $\operatorname{rank}(A) = m \leq n$

Note: $\Omega = \{x \in \mathbb{R}^n : Ax = b\} \cap \{x \in \mathbb{R}^n : x \geq 0\}$

$$\{x \in \mathbb{R}^n : Ax = b\} = \underbrace{x^0}_{\text{a particular solution of } Ax=b} + \underbrace{\operatorname{null}(A)}_{(n-m)\text{-dim. linear subspace}}$$

$\boxed{\begin{array}{l}\Omega \text{ is a } (n-m)\text{-dimensional plane (parallel to null}(A)) \text{ intersected with the non-negative octant of } \mathbb{R}^n. \\ \text{It is a (convex) polytope.}\end{array}}$

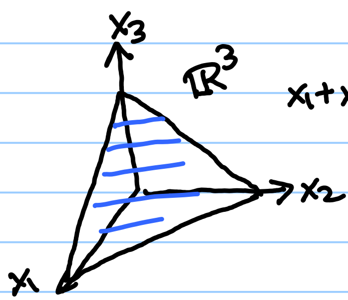
Eg.



$\{x \in \mathbb{R}^2 : x_1 + x_2 \leq 1,\ x \geq 0\}$

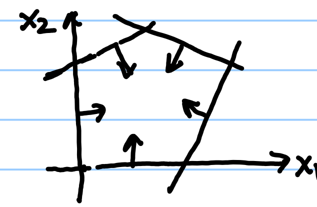
$\Big\downarrow$ put into standard form

$\Omega = \{x \in \mathbb{R}^3 : x_1 + x_2 + \underbrace{x_3}_{\text{slack variable}} = 1,\ x_1, x_2, x_3 \geq 0\}$



$\mathbb{R}^3$

$x_1 + x_2 + x_3 = 1$



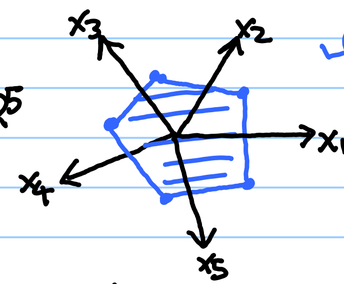
$\{x \in \mathbb{R}^2 : \underbrace{A}_{3\times 2} x \leq b,\ x \geq 0\}$

$\Big\downarrow$ "standardize"

$$\Omega = \left\{x \in \mathbb{R}^5 : [A \;\; I]\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = b,\ x \geq 0\right\}$$

slack vars: x_3, x_4, x_5

$\mathbb{R}^5$



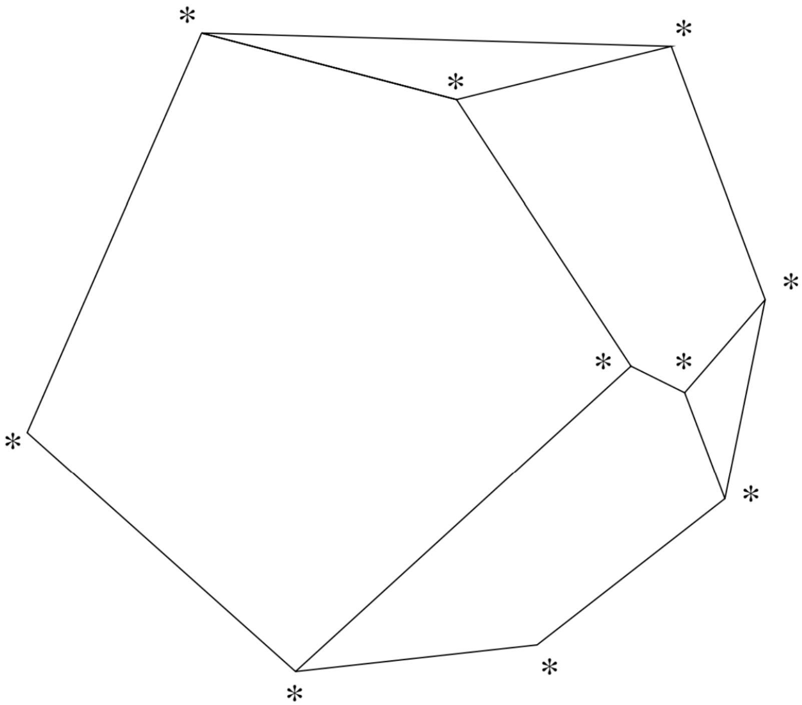
a $(5-3)$-dim. plane in $\mathbb{R}^5$

Note: $[A \;\; I] \in \mathbb{R}^{3\times 5}$ has rank 3 and nullity $5-3=2$.

If the feasible region is in the ("non-standard") form $\Lambda = \{x \in \mathbb{R}^d : \overset{m \times d}{A} x \le b, x \geqslant 0\}$,
then when $d=3$, Λ may look like this:

$\|$ the intersection of $m+d$ half-spaces in $\mathbb{R}^d$

(Every linear inequality $w_1x_1 + \cdots + w_dx_d \leqslant \gamma$ represents a half-space in $\mathbb{R}^d$.)



We can "standardize" Λ to

$$\Omega = \left\{x \in \mathbb{R}^{d+m} : [A, I] \begin{bmatrix} x_1 \\ \vdots \\ x_d \\ x_{d+1} \\ \vdots \\ x_{d+m} \end{bmatrix} = b, x \geqslant 0\right\}$$

slack variables $\rightarrow \{x_{d+1}, \ldots, x_{d+m}\}$

Conversely, given $\Omega = \{x \in \mathbb{R}^n : \overset{\substack{\text{full row} \\ \text{rank } m}}{A} x = b, x \geqslant 0\}$.

We may take out an $m \times m$ invertible submatrix from A, reorder the variables accordingly and write

$$Ax = [\underset{m\times(n-m)}{A^{(1)}}, \underset{\substack{m\times m \\ \text{invertible}}}{A^{(2)}}]\begin{bmatrix} x^{(1)} \\ x^{(2)} \end{bmatrix} = b \iff \underbrace{A^{(2)-1}A^{(1)}}_{\substack{\text{call it} \\ \mathcal{A}}} x^{(1)} + \underbrace{x^{(2)}}_{\substack{\geqslant 0 \\ \text{think} \\ \text{of it} \\ \text{as slack} \\ \text{variables}}} = \underbrace{A^{(2)-1} b}_{\substack{\text{call it} \\ \mathscr{b}}}$$

And (up to reordering of coordinates),

$$\Omega = \left\{ \begin{bmatrix} y \\ \mathscr{b} - \mathcal{A}y \end{bmatrix} : \mathcal{A}y \leq \mathscr{b}, \ y \geqslant 0 \right\} \subseteq \mathbb{R}^n$$

which is isomorphic to $\Lambda = \{ y \in \mathbb{R}^{n-m} : \underbrace{\mathcal{A}y \leq \mathscr{b}, \ y \geqslant 0}_{m+(n-m)=n \text{ linear inequalities}} \} \subseteq \mathbb{R}^{n-m}$

maybe a little easier for the theory

$$\Omega \underset{\text{eliminate variables}}{\overset{\text{add (slack) variables}}{\rightleftarrows}} \Lambda$$

Ω: intersection of a $(n-m)$-dim plane with $\mathbb{R}^n_+$

Λ: intersection of n half-spaces in $(n-m)$-dim.

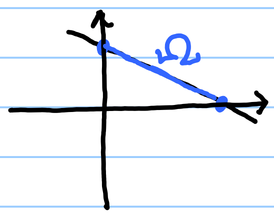
easier to visualize (lives in a lower dim. space)

Vertices of Ω

Def: An extreme point of a convex set Ω is a point in Ω that does not lie on a line segment between two other points in Ω.
(ie. $p \in \Omega$ is an extreme point if $\nexists\, x, y \in \Omega$, $t \in (0,1)$ s.t. $p = (1-t)x + ty$.)

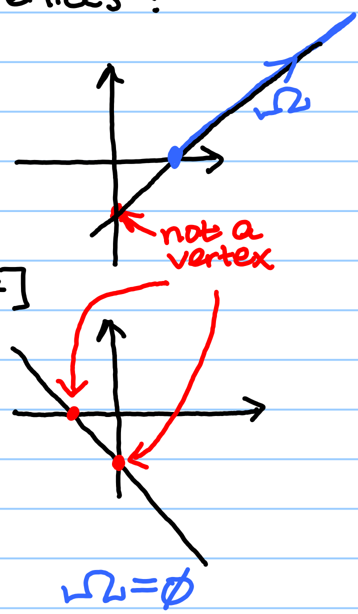
In the case of a convex polytope, we call the extreme points vertices instead.

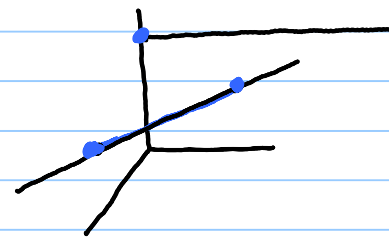
Given $\Omega = \{x : Ax = b,\ x \geq 0\} = \underbrace{\{x : Ax = b\}}_{\text{an } (n-m)\text{-dim plane}} \cap \{x : x \geq 0\}$, how do we characterize and compute its vertices?



$\begin{bmatrix} * & * \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} * \end{bmatrix}$

• — vertex

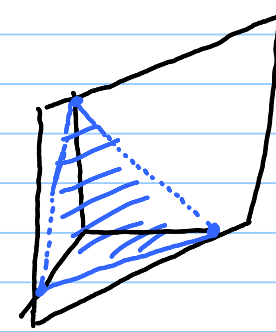




$\begin{bmatrix} * & * & * \\ * & * & * \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} * \\ * \end{bmatrix}$

a vertex typically has 1 zero component

$1 = n - m$, with $n = 3$, $m = 2$



$\begin{bmatrix} * & * & * \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} * \end{bmatrix}$

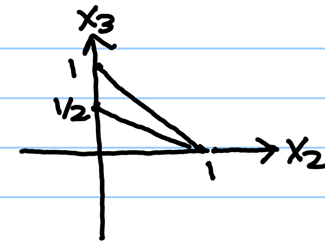
a vertex typically has 2 zero components

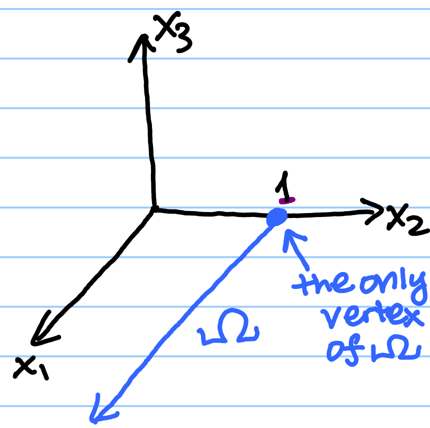
$2 = n - m$

As we shall see, a solution of a LP, if exists, must either be a vertex, or a convex combination of solution vertices.

It seems like the vertices are exactly those solutions of $Ax=b$ with (at least) $n-m$ zero components that also satisfy $x \geq 0$.

This is almost true, but consider $A = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix}$, $b = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$





$$B = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$\Rightarrow x_1 = \text{any } \# \geq 0$

$x_2 = 1$

So $\begin{bmatrix} x_1 \\ 1 \\ 0 \end{bmatrix} \geq 0 \in \Omega$, has $3-2=1$ zero, but it is not a vertex unless $x_1 = 0$

Note that B is singular in this example.

$$\min c^T x \text{ st } Ax=b, x \geq 0$$

$\downarrow$

Def A basic feasible point of (13.1) is a point $x \in \mathbb{R}^n$ that is feasible and $\exists\, \mathcal{B} \subset \{1, \cdots, n\}$ st.

1. $|\mathcal{B}| = m$
2. $i \notin \mathcal{B} \Rightarrow x_i = 0$
3. The $m \times m$ submatrix of A, defined by $B = [A_i]_{i \in \mathcal{B}}$, is non-singular.
   (Here $A = [A_1, \cdots, A_n]$.)

$\mathcal{B}$ is called a basis of the problem. B is called the basis matrix.

The fundamental theorem of LP.

(i) *If (13.1) has a nonempty feasible region, then there is at least one basic feasible point;*

(ii) *If (13.1) has solutions, then at least one such solution is a basic optimal point.*

(iii) *If (13.1) is feasible and bounded, then it has an optimal solution.*

Theorem 13.3.

All basic feasible points for (13.1) are vertices of the feasible polytope $\{x \mid Ax = b, x \geq 0\}$, and vice versa.

Proof of (i): Among all feasible vectors x, choose one with the minimal # of non-zero components, and denote this # by p.

By reordering the columns of A, we may assume WLOG that the non-zeros are $x_1, \cdots, x_p$, so $\underset{>0}{x_1} A_1 + \cdots + \underset{>0}{x_p} A_p = b$.

Suppose that the columns $A_1, \cdots, A_p$ are linearly dependent. Then $\exists\, z_1, \cdots, z_p$, not all 0, s.t.

$$z_1 A_1 + \cdots + z_p A_p = 0.$$

Assume, WLOG, that $z_p = -1$. Note:

$$(x_1 + \varepsilon z_1) A_1 + \cdots + (x_p + \varepsilon \overset{=-1}{z_p}) A_p = b$$

$x(\varepsilon) = x + \varepsilon [z_1, \cdots, z_{p-1}, -1, 0, \cdots, 0]^T$ is feasible for $|\varepsilon|$ small enough.

$\exists\, \bar{\varepsilon} \in (0, x_p]$ st $x_i(\bar{\varepsilon}) = 0$ for some $i = 1, \cdots, p$. [why? consider 2 cases: $z_1, \cdots, z_{p-1} \geq 0$ or $\exists i \in \{1, \cdots, p-1\}$ st. $z_i < 0$.]

This contradicts the minimality of p.

So $A_1, \cdots, A_p$ are linearly independent, which also means $p \leq m$.
If $p = m$, we are done. Otherwise, since A has full rank, we can choose $m-p$ columns from among $A_{p+1}, \cdots, A_n$ to build up a set of m l.i. vectors. We construct $\mathcal{B}$ by adding the corresponding indices to $\{1, \cdots, p\}$. The proof of (i) is complete.

Proof of (ii). (Quite similar.)

Let x^* be a solution with the minimal # (p) of non-zero components.

Assume WLOG $x^*_1, \cdots, x^*_p > 0$. We shall argue $A_1, \cdots, A_p$ must be l.i..

Assume contrary. Then define $x^*(\varepsilon) = x^* + \varepsilon z$, z chosen exactly the same way as in (i).
$x^*(\varepsilon)$ is feasible for sufficiently small ε (positive or negative.)

Since x^* is optimal,

$$c^T(x^* + \varepsilon z) \geq c^T x^* \implies \varepsilon c^T z \geq 0 \text{ for all } \varepsilon \text{ with } |\varepsilon| \text{ small enough}$$
$$\implies c^T z = 0 \quad (\text{why?})$$

But this in turns means $x^* + \varepsilon z$ is optimal for $|\varepsilon|$ small.

Then, using the same logic in (i), $\exists \bar{\varepsilon} > 0$ s.t. $x^*(\bar{\varepsilon})$ is feasible and optimal and has at most $(p-1)$ non-zeros.

This contradicts the minimality of p.

We can now apply the same reasoning as above to conclude that x^* is already a basic feasible point, and therefore a basic optimal point.

(iii) is an existence of minimizer result. We can prove it using a standard analysis argument, based on the closedness of Ω (when $\Omega \neq 0$) and the coercivity of $c^T x$ when $c \neq 0$. (When $c = 0$, every point in Ω is a minimizer.)

But N&W want to think of it as a consequence of the simplex method, we'll come back to it.

The proof of Thm 13.3 is not every hard, one direction uses an argument very similar to that in (i) and (ii) above.

The fundamental theorem of LP suggests an algorithm for solving LP:

1. compute all vertices
2. evaluate the objective function at each vertex
3. pick the vertex that gives the smallest objective value

This would solve any LP that is bounded. (It is not hard to design an extra step to detect unboundedness.)

What's wrong with this algorithm?

Let $m=50$, $n=100$. (We can easily find a meaningful LP way bigger than this.)

Assume your computer can solve a billion 50×50 linear systems per sec. (I doubt if any computer we've seen can do that.)

The algorithm above will take at least : $(\ ?\ )/10^9/60/60/24/365 = \underline{\ ??\ }$ years.

A related question:

If we solve unconstrained (nonlinear) optimization problems of the form

$$\min_{x \in [A,B]^d} f(x), \qquad f: \mathbb{R}^d \to \mathbb{R},$$

why can't we simply:

1. sample the box $[A,B]^d$ with a grid of size $\epsilon = \frac{B-A}{N}$.
2. Evaluate f on each grid point
3. Pick the smallest.

This algorithm guarantees an approximate minimizer with ϵ-accuracy.

What's wrong with this algorithm?

Assuming f is smooth, optimality condition is $\nabla f(x^*) = 0$

Idea of the gradient descent method:

if $\nabla f(\bar{x})$ is not yet 0, this is how you get closer!

- If $\bar{x}$ is s.t. $\nabla f(\bar{x}) \neq 0$, then $f(\overbrace{\bar{x} - t\nabla f(\bar{x})})\downarrow$ for at least small $t > 0$.
- choose an appropriate t (step size / "learning rate"), move to the next point. Iterate.

The Simplex Method

Recall the optimality conditions of $\min c^T x$ st $Ax=b$, $x \geq 0$

KKT : $A^T\lambda + s = c$, $Ax=b$, $x \geq 0$, $s \geq 0$, $x_i s_i = 0$, $i=1,\cdots,n$.

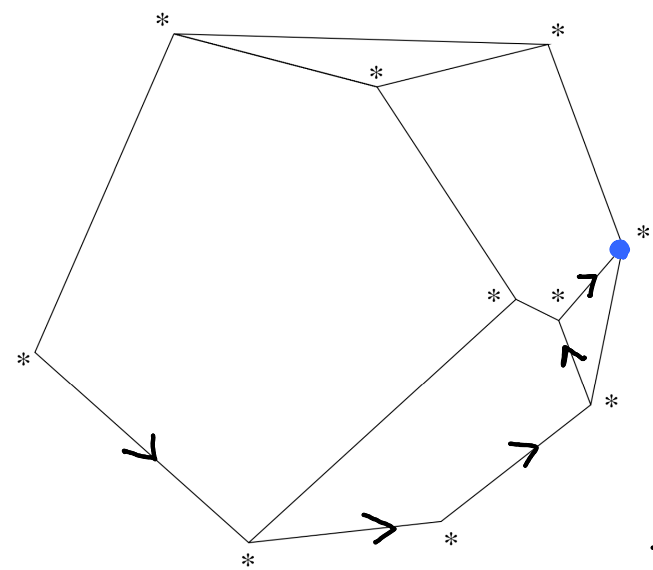
Idea of simplex method :

B' , differ from B in one index

B

- Start from a vertex, move to an adjacent vertex at which the objective $c^T x$ is decreased. Iterate.

- when the problem is unbounded, the final step will move infinitely far without ever reaching a vertex.



Assume we have a vertex, specified by B.

Write : $\mathcal{N} = \{1,\cdots,n\} \setminus \mathcal{B}$

B = the basis matrix $= [A_i]_{i\in\mathcal{B}}$, $N = [A_i]_{i\in\mathcal{N}}$

For any $y \in \mathbb{R}^n$, $y_B := [y_i]_{i\in\mathcal{B}}$, $y_N := [y_i]_{i\in\mathcal{N}}$

$Ax = [B \ \ N]\begin{bmatrix}x_B\\x_N\end{bmatrix} = Bx_B + Nx_N$. If $Ax=b$, then $Bx_B + Nx_N = b$,

$$x_B = -B^{-1}N x_N + B^{-1}b.$$

Since x is a vertex with basis B, $\underline{x_N = 0, \ x_B = B^{-1}b} \geq 0$

If the vertex defined by B does not yet satisfy KKT, how do we "get closer"?

$$c^Tx = c_B^T x_B + c_N^T x_N = c_B^T(-B^{-1}N x_N + B^{-1}b) + c_N^T x_N$$
$$= [c_N - N^T B^{-T} c_B]^T x_N + c_B^T B^{-1} b.$$

	x_N	1
$x_B=$	$-B^{-1}N$	$B^{-1}b \geq 0$
$c^Tx=$	$(c_N - N^T B^{-T} c_B)^T$	$c_B^T B^{-1} b$

≥ 0 if the vertex is optimal
otherwise it has a negative entry

Note: If $c_N - N^T B^{-T} c_B \geq 0$ (*), then the vertex solves the LP.
(HW: This can be proved directly, without using KKT.)

But since KKT are the necc. and suff. conditions for optimality, there must exist s, λ that satisfy KKT.

Choose s st $s_B = 0$, so that the complimentarity condition is satisfied.
Then $A^T\lambda + s = c$ reduces to $(A^T\lambda)_B + s_B^{=0} = c_B \Rightarrow B^T\lambda = c_B \Rightarrow \lambda = B^{-T}c_B$

$$(A^T\lambda)_N + s_N = c_N \Rightarrow N^T\lambda + s_N = c_N$$
$$\Rightarrow s_N = \boxed{c_N - N^T \underbrace{B^{-T} c_B}_{\lambda}}$$

same vector in (*)!

By construction, all the KKT conditions are satisfied if $s_N \geq 0$. But this is exactly $(*)$!

To get closer, we should choose an index $q \in \mathcal{N}$ for which $s_q < 0$. Then, since $x_q \uparrow \Rightarrow c^T x \downarrow$, our procedure for altering $\mathcal{B}$ is based on the following considerations:

- allow x_q to increase from zero during the next step;
- fix all other components of x_N at zero, and figure out the effect of increasing x_q on the current basic vector x_B, given that we want to stay feasible with respect to the equality constraints $Ax = b$;
- keep increasing x_q until one of the components of x_B (x_p, say) is driven to zero, or determining that no such component exists (the unbounded case);
- remove index p (known as the *leaving index*) from $\mathcal{B}$ and replace it with the entering index q.

EXACTLY HOW IS IT DONE?

This process of selecting entering and leaving indices, and performing the algebraic operations necessary to keep track of the values of the variables x, λ, and s, is sometimes known as *pivoting*.

Pivoting

$$c=\begin{bmatrix}-4\\-2\\0\\0\end{bmatrix} \qquad A=\begin{bmatrix}1&1&1&0\\2&\frac{1}{2}&0&1\end{bmatrix},\ b=\begin{bmatrix}5\\8\end{bmatrix}$$

Consider $\min\ -4x_1-2x_2$

$$\text{s.t.}\quad x_1+x_2+x_3=5 \iff x_3:=5-x_1-x_2\geq 0$$
$$2x_1+\tfrac{1}{2}x_2+x_4=8 \iff x_4:=8-2x_1-\tfrac{1}{2}x_2\geq 0$$

Start at $\mathcal{B}=\{3,4\}$. $\mathcal{N}=\{1,2\}$. This corresponds to $\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix}=\begin{bmatrix}0\\0\\5\\8\end{bmatrix}$.

1.

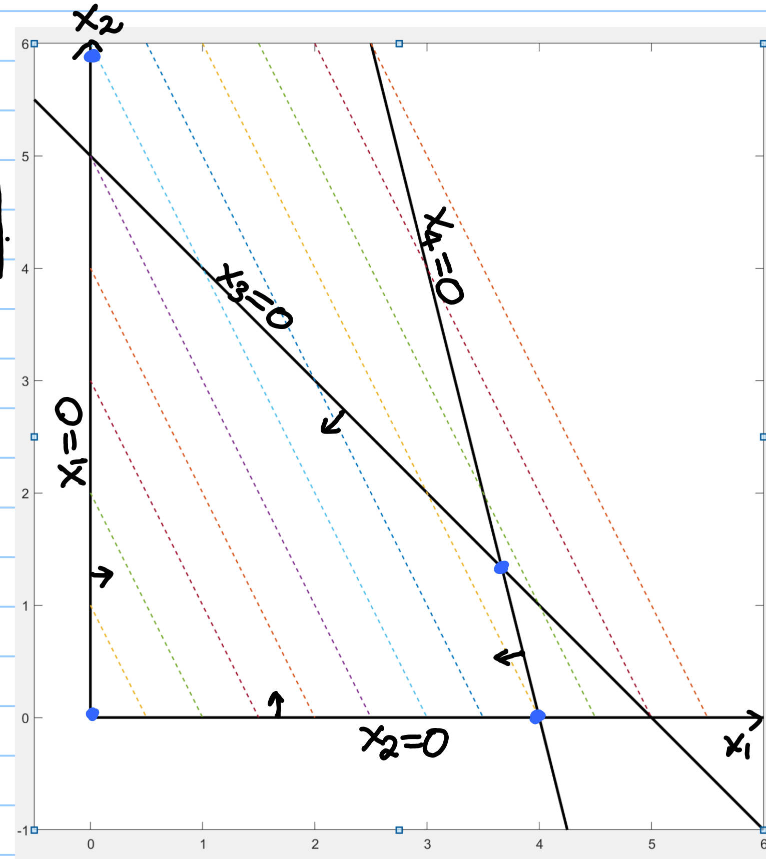
	x_N	1
$x_B=$	$-B^{-1}N$	$B^{-1}b$
$c^Tx=$	$(c_N-N^TB^{-T}c_B)^T$	$c_B^TB^{-1}b$

is now

	x_1	x_2	1
$x_3=$	-1	-1	5
$x_4=$	-2	$-\frac{1}{2}$	8
$c^Tx=$	-4	-2	0

2. Choose $q=1$

If x_2 remains 0, $x_1\uparrow$ from 0 to $\lambda>0$, then
$$x_3=-\lambda+5\geq 0 \iff \lambda\leq 5/1 = 5$$
$$x_4=-2\lambda+8\geq 0 \iff \lambda\leq 8/2 = 4$$
So we can increase x_2 to the value 4, and that is when $x_4=0$.



Pivot: $\mathcal{B}=\{3,4\}$, $\mathcal{N}=\{1,2\}$ becomes
$\mathcal{B}=\{3,1\}$, $\mathcal{N}=\{4,2\}$.

This is a picture in $\mathbb{R}^2$, not $\mathbb{R}^4$.

	x_N	1
$x_B =$	$-B^{-1}N$	$B^{-1}b$
$c^Tx =$	$(c_N - N^T B^{-T} c_B)^T$	$c_B^T B^{-1} b$

is now

	x_4	x_2	1
$x_3 =$		$-3/4$	1
$x_1 =$		$-1/4$	4
$c^Tx =$	$3/2$	$-5/2$	-16

This corresponds to $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \\ 1 \\ 0 \end{bmatrix}$.

Choose $q = 2$.

3. If x_4 remains 0, $x_2 \uparrow$ from 0 to $\lambda > 0$, then
$x_3 = -3/4\,\lambda + 1 \geq 0 \Rightarrow \lambda \leq 1/\tfrac{3}{4} = 4/3$
$x_1 = -1/4\,\lambda + 4 \geq 0 \Rightarrow \lambda \leq 4/\tfrac{1}{4} = 16$

So we can increase x_2 to $4/3$, and that is when $x_3 = 0$.

Pivot: $B = \{3, 1\}$, $N = \{4, 2\}$ becomes $B = \{2, 1\}$, $N = \{4, 3\}$.

	x_N	1
$x_B =$	$-B^{-1}N$	$B^{-1}b$
$c^Tx =$	$(c_N - N^T B^{-T} c_B)^T$	$c_B^T B^{-1} b$

is now

	x_4	x_3	1
$x_2 =$			$1\tfrac{1}{3} \geq 0$
$x_1 =$			$3\tfrac{2}{3} \geq 0$
$c^Tx =$	$4/3 \geq 0$	$4/3 \geq 0$	$-17\tfrac{1}{3}$

min. value $= -17\tfrac{1}{3}$

minimizer $= \begin{bmatrix} 3\tfrac{2}{3} \\ 1\tfrac{1}{3} \\ 0 \\ 0 \end{bmatrix}$.

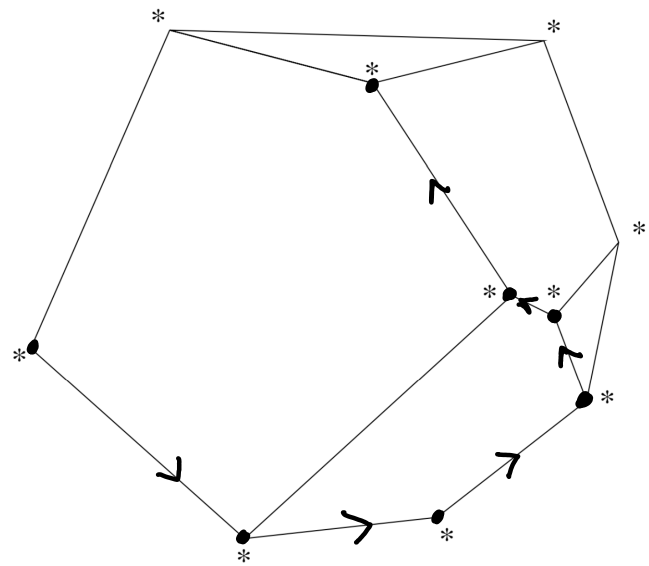
[Apparently, the computation in N&W's Example 13.1 is erroneous.
Does anyone have the audacity to write the authors an email?]

HW: Work out the simplex steps if you choose $q = 2$ in step 2.

The example above tells us how to pivot in each step.

Procedure 13.1 (One Step of Simplex).

Given $\mathcal{B}, \mathcal{N}$, $x_B = B^{-1}b \geq 0$, $x_N = 0$;
Solve $B^T\lambda = c_B$ for λ,
Compute $s_N = c_N - N^T\lambda$; (* pricing *)
if $s_N \geq 0$
  stop; (* optimal point found *)
Select $q \in \mathcal{N}$ with $s_q < 0$ as the entering index;
Solve $Bd = A_q$ for d;
if $d \leq 0$
  stop; (* problem is unbounded *)
Calculate $x_q^+ = \min_{i \,|\, d_i > 0} (x_B)_i / d_i$, and use p to denote the minimizing i; (* ratio test *)
Update $x_B^+ = x_B - dx_q^+$, $x_N^+ = (0, \ldots, 0, x_q^+, 0, \ldots, 0)^T$;
Change $\mathcal{B}$ by adding q and removing the basic variable corresponding to column p of B.



Let's be more formal. From this we will see that things can potentially go wrong in the presence of degenerate vertices.

When we identify an index $q \in N$ s.t. $s_q < 0$ from the current iterate x (with basis B).

Assume that we find a $p \in B$ in the ratio test above. Call the new iterate x^+.

$$Ax^+ = Bx_B^+ + A_q x_q^+ = Bx_B \; (= Ax = b)$$

$$[A_1 \cdots A_p \cdots A_q \cdots A_n] \begin{bmatrix} x_1 \\ \vdots \\ x_p \\ \vdots \\ x_q \\ \vdots \\ x_n \end{bmatrix} = b$$

so $\quad x_B^+ = x_B - B^{-1} A_q x_q^+$

Let's see how this pivot affects the value of $c^T x$:

$$c^T x^+ = c_B^T x_B^+ + c_q x_q^+ = \underbrace{c_B^T x_B}_{= c^T x} - c_B^T B^{-1} A_q x_q^+ + c_q x_q^+ = c^T x + \underbrace{(c_q - c_B^T B^{-1} A_q)}_{s_q} x_q^+$$

Since $c^T x^+ = c^T x + \underset{\overset{\vee}{0}}{s_q} \underset{\overset{\wedge|}{0}}{x_q^+}$, the simplex step produces a (strict) decrease in the objective value if $x_q^+ > 0$.

This is where the concept of degeneracy comes in:

Definition 13.1 (Degeneracy).

A basis $\mathcal{B}$ is said to be degenerate *if $x_i = 0$ for some $i \in \mathcal{B}$, where x is the basic feasible solution corresponding to $\mathcal{B}$. A linear program (13.1) is said to be* degenerate *if it has at least one degenerate basis.*

In above, if $\mathcal{B}$ is non-degenerate, then we are guaranteed that $x_q^+ > 0$, hence a strict ↓ in the objective value at this step.

If the LP is non-degenerate, then we can ensure a decrease in $c^T x$ at every step. We have:

Theorem 13.4.

Provided that the linear program (13.1) is nondegenerate and bounded, the simplex method terminates at a basic optimal point.

The proof is simply based on the fact that there are only a finite # of vertices.