- The vectors $V = \{v_1, v_2 \ldots, v_n\}$ are *linearly independent* if $\sum_i \alpha_i v_i = 0$ and $\forall_i \alpha_i \in \mathbb{R}$ imply that $\forall_i \alpha_i = 0$.

  A *vector space* or *linear space* $\mathcal{V}$ in $\mathbb{R}^n$ is a nonempty subset of $\mathbb{R}^n$ closed under vector addition and scalar multiplication.

  A set of vectors in V are said to *span* $\mathcal{V}$ whenever every vector in $\mathcal{V}$ can be expressed as a linear combination of vectors in V and any vector not in $\mathcal{V}$ cannot be expressed as a linear combination of vectors in V.

  The vectors in V form a *basis* of $\mathcal{V}$ iff vectors in V are linearly independent and $span(V) = \mathcal{V}$ iff vectors in V is a minmal spanning set for $\mathcal{V}$ iff vectors in V is a maximal linearly independent subset of $\mathcal{V}$.

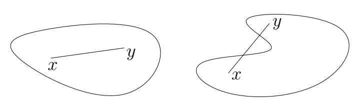
  Though it is not necessary for a vector space to have a unique basis, the number of vectors in any basis of V is the same. The *dimension* of $\mathcal{V}$ is the number of vectors in any basis of $\mathcal{V}$.

  For a matrix A, the vector space spanned by the columns of A is called the *column space, a.k.a., range space,* $C(A)$ of A. For A of order $m \times n$, the $C(A)$ is a subspace of $\mathbb{R}^m$. The *rank* of A is the dimension of $C(A)$. If the rank of A is m, the $C(A)$ is the whole space $\mathbb{R}^m$.

- Given vectors $x_1, \ldots, x_k \in \mathbb{R}^n$ and $\lambda_1, \ldots, \lambda_k \geq 0$ with $\sum_{i=1}^{k} \lambda_i = 1$, we call, $x = \sum_{i=1}^{k} \lambda_i x_i$, a *convex combination* of $x_1, \ldots, x_k$.

  For a point p belonging to line segment $p'p''$, there exists a $\alpha \in [0, 1]$ such that $p = \alpha p' + (1 - \alpha)p''$. That is, p is at distance $\alpha \cdot |p'p''|$ from p'' and $(1 - \alpha) \cdot |p'p''|$ from p'. This denotes p is a convex combination of p' and p''.

- A set S is a *convex set* if for any two points $x, y \in S$, every point on the line segment joining x and y is contained in S.

  Equivalently, a set S is convex if for every set S' of points located in S, the convex combination of points in S' also belongs to S.

  The intersection of any number of convex sets in $\mathbb{R}^n$ is convex.

- Let $S \subseteq \mathbb{R}^n$ be a convex set. The real function $f : S \to \mathbb{R}$ is a *convex function* in S if for any two points $x, y \in S$, $f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y))$ for $\lambda \in \mathbb{R}$ and $0 \leq \lambda \leq 1$. In other words, the real function f is convex whenever $\{(x, y) \in \mathbb{R}^2 : y \geq f(x)\}$ is a convex set. Ex., $x^2, e^x, |x|$.

- The *convex hull* of a set S, denoted by $CH(S)$, is the smallest convex set that contains S. That is, the convex hull of S is the intersection of all convex sets that contain S.

  Let $CC(S)$ be the set that precisely comprises of all the connvex combinations of points in S. Then, $CC(S) = CH(S)$.

  For any set S and $S' \subseteq S$, $CH(S') \subseteq CH(S)$.

- *Caratheodory's theorem*: For a point $x \in \mathbb{R}^d$, suppose x is expressed as a convex combination of $k > d+1$ number of points in $\mathbb{R}^d$. Then, x can be expressed as a convex combination of at most $d+1$ points.

- An *extreme point* of a convex set S is a point p in S that cannot be expressed as a convex combination of points in $S - \{p\}$. In other words, a point $p \in S$ is an extreme point of S if and only if p is a vertex of $CH(S)$.

- The vectors $v_1, \ldots, v_r$ are *affinely independent* if $\sum_i \alpha_i v_i = 0, \forall_i \alpha_i \in \mathbb{R}$, and $\sum_i \alpha_i = 0$ together imply that $\forall_i \alpha_i = 0$.

  An *affine space* A in $\mathbb{R}^n$ is the space resulting from adding a fixed vector t to all the elements of a linear space S, i.e., $A = \{t + y \,|\, y \in S\}$. (For example, a plane in $\mathbb{R}^2$ not necessarily passing through the origin is an affine space.) The dimension of A is the dimension of S. If B is an arbitrary subset of $\mathbb{R}^n$, then the dimension of B is the smallest dimension of any affine space containing B.

  A *hyperplane* in $\mathbb{R}^n$ is an affine subspace of dimension $n-1$; in other words, it is the set of all solutions of a linear equation of the form $a_1x_1 + \ldots + a_nx_n = b$, where $a_1, \ldots, a_n$ are not all 0.

  The hyperplane with equation $a_1x_1 + \ldots + a_nx_n = b$ induces two *closed half-spaces*: $\{x \in \mathbb{R}^n : x_1 + \ldots + a_nx_n \leq b\}$ and $\{x \in \mathbb{R}^n :_1 x_1 + \ldots + a_nx_n \geq b\}$.

- For some matrix $A \in \mathbb{R}^{m \times n}$ and some vector $b \in \mathbb{R}^m$, the set P comprising points in $\{x \in \mathbb{R}^n : Ax \leq b\}$ is a *polyhedron* (a.k.a., *H-polyhedron* or *polyhedral set*). That is, the convex set obtained by the intersection of a finite number of affine halfspaces is a polyhedron. If A and b are rational, then P is a *rational polyhedron*. A bounded H-polyhedron is called a *H-polytope*.

  A subset P of $\mathbb{R}^n$ is called a *polytope* (a.k.a., *V-polytope*) if P is the convex hull of a finite number of vectors.

  Minkowski-Weyl's theorem: A subset P of $\mathbb{R}^n$ is a polytope if and only if it is a bounded polyhedron.

  A polyhedron P is *pointed* if it has at least one vertex.

- The *dimension of a convex polyhedron* $P \subseteq \mathbb{R}^n$ is the smallest dimension of an affine subspace containing P. Equivalently, it is the largest $d \leq n$ for which P contains points $x_0, x_1, \ldots, x_d$ such that the d-tuple of vectors $(x_1 - x_0, \ldots, x_d - x_0)$ is linearly independent. That is, the dimension is the maximum number of affinely independent points in P minus 1. If the dimension of P is n then P is *full-dimensional*.

- For a hyperplane H, if $H \cap P \neq \phi$ and P is entirely contained in one of the closed half-spaces defined by H, then H is called a *supporting hyperplane of* P. Formally, let $P = \{x : Ax \leq b\}$ be a nonempty polyhedron. If c is a nonzero vector for which $\delta = max\{cx : x \in P\}$ is finite, then $\{x : cx = \delta\}$ is called a supporting hyperplane of P.

- A *face* of P is, P itself, or the intersection of P with a supporting hyperplane of P.

  F is a face of P if and only if $F = \{x \in P : A'x = b'\} \neq \phi$ for some subsystem $A'x \leq b'$ of $Ax \leq b$.

  A point x for which $\{x\}$ is a face is called a *vertex* of P. That is, a vertex is a 0-dimensional face. An edge is a 1-dimensional face.

A *facet* of P is any maximal face distinct from P. An inequality $cx \leq \delta$ is *facet-defining* for P if $cx \leq \delta$ for all $x \in P$ and $\{x \in P : cx = \delta\}$ is a facet of P.

- *Asymptotic upper bound theorem*: A d-dimensional polytope with n vertices has $O(n^{\lfloor d/2 \rfloor})$ facets and $O(n^{\lfloor d/2 \rfloor})$ faces.

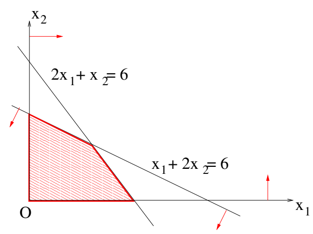
  Since any polytope formed with f half-spaces can have at most f facets, in d-dimensions, the maximum number of vertices a polytope has is $\Theta(f^{\lfloor d/2 \rfloor})$.

- *Separation theorem*: Let $S \subseteq \mathbb{R}^n$ be a finite set and let $p \in \mathbb{R}^n \backslash CH(S)$. Then there exists an inequality $w^T x \leq t$ that separates p from $CH(S)$, that is, $w^T s \leq t$ for all $s \in CH(S)$ but $w^T p > t$.

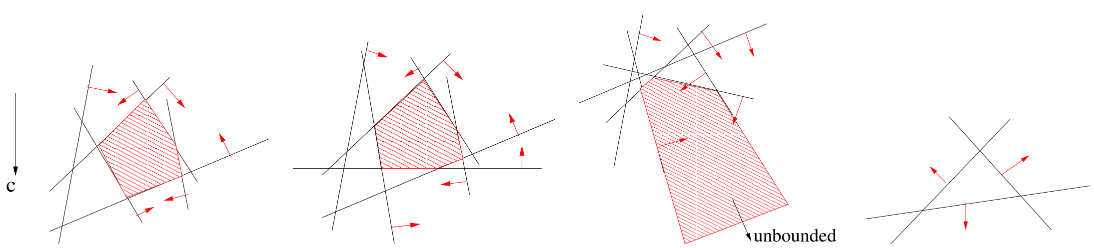
- *Theorem of the alternatives*: The system $A_{m \times n} x \leq b$ has no solution $x \in \mathbb{R}^n$ if and only if there exists a $y \in \mathbb{R}^m$ such that $y \geq 0$, $A^T y = 0$ and $b^T y < 0$.

  Farkas' lemma: The system $Ax = b$ has a nonnegative solution if and only if there is no vector y satisfying $y^T A \geq 0$ and $y^T b < 0$.

- The linear constraints together form a convex feasible region, which is a convex polyhedra.



- The LP $\min\{c^T x : Ax \geq b, x \geq 0\}$ is *unbounded* from below if there exists a feasible x such that $c^T x < \lambda$ for every real number λ. (Analogous definition applies to maximization LP as well.) When the feasible region is empty, then the LP is said to be *infeasible*. Any linear program either (i) has an optimal solution with a finite objective value, (ii) is infeasible, or (iii) is unbounded.



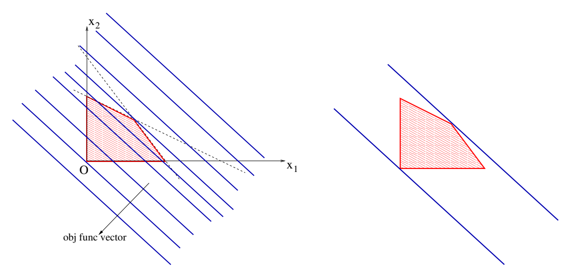
- Consider an LP $\min\{c^T x : Ax = b, x \geq 0\}$. Let B be a basis of A. The *basic solution* corresponding to basis B is a vector $x \in \mathbb{R}^n$ with (i) for every $j \notin B$, $x_j = 0$, and (ii) for every $j \in B$, the j^{th} element of column vector x is equal to the j^{th} element of $B^{-1}b$. If x is both basic and feasible, then x is a *basic feasible solution*. Every basic feasible solution is uniquely determined by the set B. The variables x_j with $j \in B$ are the *basic variables*, and the remaining variables are the *nonbasic variables*.

- Let A be a $m \times n$ matrix with full row rank. (To remind, the column rank of A is the dimension of the column space of A, while the row rank of A is the dimension of the row space of A. If the rank of A is equal to the number of rows (resp. number of columns), then A is said to have full row rank (resp. full column rank). And, if the rank of A is equal to the minimum of number of rows and the number of columns of A, then A is said to have full rank.) Then, every feasible x to $P = \{x : Ax = b, x \geq 0\}$ is a basic feasible solution if and only if x is an extreme point solution. (Note that after introducing slack variables for each constraint, any LP can be transformed to a linear program in *standard form*: $\{x : Ax = b, x \geq 0\}$.)

  The *rank lemma*: Let $P = \{x \mid Ax \leq b\}$ be a polyhedron in $\mathbb{R}^n$, and let $z \in P$. Also, let A_z be the submatrix of A consisting of those rows a_i of A for which $a_i z = b_i$. Then z is an extreme point of P if and only if $rank(A_z) = n$.

  Every basic feasible solution corresponds to one and only one vertex. And, there exists a basic feasible solution corresponding to every vertex of F. However, the correspondance between bases and basic feasible solutions is not one to one: there can be many bases, each of which corresponds to the same basic feasible solution. If there exists than one basis correspond to the same basic feasible solution x, then x contains more than $n - m$ zeros; and such an x is said to be a *degenerate* solution.

- The objective function is a family of parallel hyperplanes; the objective function value is fixed along any one such hyperplane. Significantly, for the set S of points satisfying the given constraints, $\max\{w^T x : x \in S\} = \max\{w^T x : x \in bd(CH(S))\}$.



  An optimal solution corresponding to a basic feasible solution is called a *basic optimal solution*. As every supporting hyperplane contains at least one vertex, there exists a vertex at which the optimum is guaranteed to occur.

  Let $P = \{x : Ax \geq b, x \geq 0\}$ and assume that $\min\{c^T x : x \in P\}$ is finite. Then, for every $x \in P$, there is an extreme point solution $x' \in P$ such that $c^T x' \leq c^T x$, i.e., there is always an extreme point optimal solution.

  Hence, LP is a combinatorial optimization problem even though it seem to be a continuous optimization problem! (Some say it lies on the boundary between continous and discrete optimization problems.)

- *Local optima is the global optima for any LP*: Consider an optimization problem with $F \subseteq \mathbb{R}^n$ being a convex set and f being a convex objective function defined over F. Then, the global optimality equals the local optimality whenever the locality is defined with respect to an ϵ-radius Euclidean closed ball B, that is, if f achieves a local optima with respect to ball B (say minimum) at $x \in F$ (that is, $\forall_{y \in B} f(x) \leq f(y)$) then $f(x)$ is a global optima as well (that is, $\forall_{y \in F} f(x) \leq f(y)$).

- Dantzig's *simplex algorithm*:

  A variable whose value is currently set to 0 is called a nonbasic variable; and, a variable whose value is currently not equal to 0 is called a basic variable. At any time of execution of the simplex algorithm, every variable is either a basic or a nonbasic variable. In every iteration of simplex algorithm, one nonbasic variable becomes a basic variable and one basic variable becomes a nonbasic variable. The simplex algorithm chooses a nonbasic variable that gives the fastest rate of increase in the objective function value as the *entering* basic variable, and it chooses the basic variable that most limits the increase in the value of the entering basic variable as the *leaving* basic variable. Hence, the updated solution vector in any iteration is guaranteed to improve the objective function value; besides, the new solution vector corresponds to a vertex of the LP polytope. Indeed, the solution vector at the beginning of each iteration corresponds to a vertex of the LP polytope. More surprisingly, any two such successive solutions are adjacent vertices of the LP polytope.

  If the objective function value cannot be improved by swapping the staus of any of the variables in x from nonbasic to basic (and vice versa), then simplex algorithm is said to have arrived to a local optima at x. From that stage, no adjacent vertex to x on the LP polytope yields a better objective function value. Since both the LP polytope and the objective function are convex, the objective function value at x is a global optima.

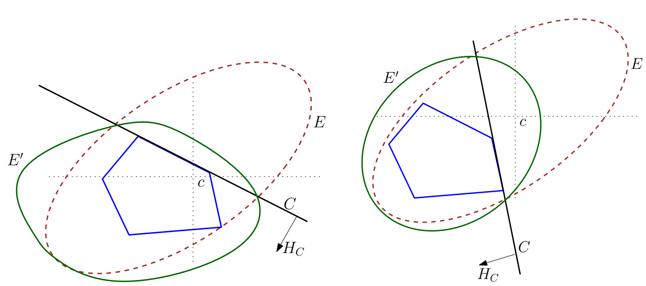
  (i) initialize the basic and nonbasic variables corresponding to a vertex of the LP polytope

  (ii) while (the current solution is not optimal)

  - (a) choose entering and leaving basic variables
  - (b) update the current solution vector

  Using the asymptotic upper bound theorem, the worst-case time of simplex algorithm was proven to be exponential in the number of variables. Klee & Minty provided an instance on which the algorithm indeed visits all the vertices of the polytope (known as Minty's cube). However, the simplex algorithm does well in practice. Specifically, under various probability distributions over the input, the simplex algorithm is proven to be taking polynomial time in the average-case.

- Kachiyan's *Ellipsoid algorithm*:



  (1) start with an ellpsoid E that contains all the basic feasible solutions of the given LP

  (2) invoke a *separation oracle* O with center c of the ellipsoid E as O's input:

  if c is feasible, O returns an arbitrary constraint C that bounds the LP polytope

  else O returns a *violated constraint* C, saying, the LP polytope and the c are on different sides of hyperplane defined by C

(3) computes a minimum volume elliposid E' that contains the region common to both E and H_C such that the $\frac{vol(E')}{vol E} < 2^{\frac{-1}{2(n+1)}}$ where n is the number of variables in the LP

(4) set $E \leftarrow E'$; and go to step (2) unless the ellipsoid is sufficiently small so that it contains at most one basic feasible solution, if any; if one exists, then this must be a basic optimal solution

This algorithm takes $O(n^6 Lt)$ time, where n is the number of variables, L is the number of bits to encode any constaint, and t is the time complexity of the separation oracle. Surprisingly, the time complexity does not depend on the number of constraints! For some LPs, separation oracles that do better than explicitly checking p versus each constraint are possible, i.e., a polynomial-time oracle exists even when the number of constraints is exponential! In practice, the Ellipsoid algorithm does not perform as good as the simplex algorithm. And, due to square-root operations involved, rounding errors need to be handled carefully.

Ex. relaxed LP for finding a minimum-cost arborescence rooted at r:

min $\sum_{(i,j)\in E} c_{ij}x_{ij}$
$\sum_{(i,j)\in E, i\in S, j\in V-S} x_{ij} \geq 1$ for every $S \subseteq V, r \in S$ ——— (1)
$0 \leq x_{ij} \leq 1$ for every $(i,j) \in E$

given a vector x, the separation oracle checks the exponential number of constraints in (1) by invoking a min-cut finding algorithm for at most $n-1$ times: it associates capacity x_{ij} to arc (i,j) in G, for every vertex $t \in V - \{r\}$, if the r-t min-cut capacity is strictly less than 1, then that min-cut $(S, V-S)$ corresponds to a violated constraint; the rest of the constraints can be checked explicitly

- Karmakar's *interior point algorithm* takes $O(n^{3.5}L^2(\lg L)(\lg\lg L))$ time. Here, n is the number of variables and L is the number of bits in encoding the input instance. In practice, this algorithm does not perform as good as the simplex algorithm. And, like in Ellipsoid algorithm, there are rounding errors due to square-root operations.

The hunt is on for a strongly polynomial time algorithm.

- *Primal-dual LP pairs* -

primal LP:	dual LP:
min $\sum_{i=1}^{n} c_i x_i$	max $\sum_{j=1}^{m} b_j y_j$
$a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n = b_i \quad i \in M$	$y_i <> 0 \quad i \in M$
$a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n \geq b_i \quad i \in \overline{M}$	$y_i \geq 0 \quad i \in \overline{M}$
$x_j \geq 0 \quad j \in N$	$a_{1j}y_1 + a_{2j}y_2 + \ldots + a_{mj}y_m \leq c_j \quad j \in N$
$x_j <> 0 \quad j \in \overline{N}$	$a_{1j}y_1 + a_{2j}y_2 + \ldots + a_{mj}y_m = c_j \quad j \in \overline{N}$

For any primal and dual LP pairs P and D, one of the following four statements must hold: (i) both P and D are feasible, (ii) P is infeasible and D is unbounded, (iii) P is unbounded and D is infeasible, and (iv) both P and D are infeasible.

- *Weak duality theorem*: Let x^*, y^* be feasible solutions of primal and dual LPs respectively. Assuming the primal LP is a minimization LP, $\sum_{i=1}^{m} b_i y_i^* \leq \sum_{j=1}^{n} c_j x_j^*$.

  A consequnce of weak duality theorem: If $\sum_{i=1}^{m} b_i y_i^* = \sum_{j=1}^{n} c_j x_j^*$, then x^* and y^* are optimal solutions to the primal and dual LPs respectively.

- *Strong duality theorem*: The primal LP has an optimal solution if and only if the dual LP has an optimal solution. At optimality, the cost of primal LP is equal to the cost of dual LP.

- *Complementary slackness conditions*: Let x^*, y^* respectively be feasible in a primal LP and its corresponding dual LP. Then, that pair x^*, y^* is optimal if and only if

  $\forall_i \; y_i((a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n) - b_i) = 0$, and
  $\forall_j \; (c_j - (a_{1j}y_1 + a_{2j}y_2 + \ldots + a_{mj}y_m))x_j = 0$.

  The former conditions are known as the pimal complementary slackness conditions whereas the latter are the dual complementary slackness conditions.

- If the coeffiecients in the constraint matrix, objective function, and the RHS vector are all non-negative in the LP, then that LP is called a *covering LP*. The dual of a covering LP is called a *packing LP*.

- Let P be a polytope. Also, let x be an extreme point solution of P. Then, x is said to be *integral* if each coordinate of x is an integer. The polytope P is called *integral* if every extreme point of P is integral.

- The integer linear programming (ILP) is a linear program in which the variables are restricted to be integers. The decision version of the ILP problem is NP-complete.

- A square, integer matrix M is called *unimodular (UM)* if $det(M) = \pm 1$. An integer matrix M is called *totally unimodular (TUM)* if every square, nonsingular submatrix of M is unimodular.
  Examples for TUM matrices include node-arc incidence matrix of a directed graph (ex., ILPs of shortest path, maximum flow, and Hitchcock's transportation problems) and the node-edge incidence matrix of an undirected bipartite graph (ex., ILP of maximum weighted bipartite perfect matching).

  If A is TUM and b is integral, then the basic solutions corresponding to both $\{x : Ax = b, x \geq 0,\, x$ integer$\}$ and $\{x : Ax = b, x \geq 0\}$ are the same.
  Analogously, if A is TUM and b is integral, then the basic solutions corresponding to both $\{x : Ax \leq b, x \geq 0,\, x$ integer$\}$ and $\{x : Ax \leq b, x \geq 0\}$ are the same.

References:

* Combinatorial Optimization by C. H. Papadimitriou and K. Steiglitz.

* Lectures on Discrete Geometry by J. Matousek.