MA4253 Mathematical Programming *

Defeng Sun

Department of Mathematics
National University of Singapore
Office: S14-0425
Telephone: 6874 3343

Aims/Objectives: The prerequisite for this module is MA2215 Linear Programming and MA3236 Nonlinear Programming. It covers a wide range of topics in continuous optimization.

We first study in a systematic way on the basic definitions of extreme points, vertices, basic feasible solutions, cone, ray, and recession cones of polyhedral sets and present the finite basis theorem for the polyhedral set. Then we study how to use these concepts to handle bounded variables efficiently in the simplex method and to develop decomposition techniques, in particular Dantzig-Wolfe decomposition method, to deal with large-scale optimization problems. Modern interior point methods, in particular, the affine scaling, potential reduction and primal-dual path following algorithms for solving linear programming, are also topics to study. Other topics include: Lemke's pivotal method for the linear complementarity problem (LCP); equivalent equation forms of the LCP; the definition of subgradients of a convex function; subgradients of simple convex functions; reformulation methods for systems of Karush-Kuhn-Tucker conditions, and some advanced topics in Lagrangian duality.

Mode of Evaluation: Class performance (individual) and computer work (group)(20%); Mid-Term test (20%) and Final examination (60%)

Main References

1) D. Bertsimas and J. N. Tsitsiklis, Introduction to Linear Optimization. Athena Scientific, 1997. ISBN: 1-886529-19-1.

2) M. S. Bazaraa, H. D. Sherali and C. M. Shetty, Nonlinear Programming: Theory and Algorithms. John Wiley & Sons, 1993. ISBN: 0-471-55793-5.

Summary of lecture notes will be made available in my webpage at http://www.math.nus.edu.sg/~matsundf/

*This course is taught at Department of Mathematics, National University of Singapore, Semester 1, 2003/2004. E-mail: matsundf@nus.edu.sg

1 The Geometry of Polyhedral Sets

1.1 Convex sets

In mathematical programming, one definitely unavoidable concept is probably the convex set. In linear programming and nonlinear programming, we have already met many convex sets. For examples, the *line segment* between two points in $\Re^n$ is a convex set; a unit ball in $\Re^n$ is a convex set; and more importantly a polyhedral set is a convex set (a formal definition of a polyhedral set is to be given shortly). But, what is a convex set?

Definition 1.1 *A set $S \subseteq \Re^n$ is* **convex** *if for any $x, y \in S$, and any $\lambda \in [0, 1]$, we have $\lambda x + (1 - \lambda)y \in S$, i.e., the whole line segment between x and y is in S.*

Exercise: Give two more sets which are convex and two sets which are not convex.

1.2 Hyperplanes and half spaces

Definition 1.2 *Let a be a nonzero vector in $\Re^n$ and b be a scalar. Then the set*

$$\{x \in \Re^n \mid a^T x = b \}$$

is called a **hyperplane***, where a^T is the transpose of the (column) vector a.*

Geometrically, the hyperplane $\{x \in \Re^n \mid a^T x = b\}$ can be understood by expressing it in the form

$$\{x \in \Re^n \mid a^T (x - x^0) = 0\}$$

where x^0 is any point in the hyperplane, i.e., $a^T x^0 = b$. This representation can then be interpreted as

$$\{x \in \Re^n \mid a^T (x - x^0) = 0\} \; = \; x^0 + a^\perp,$$

where $a^\perp$ denotes the orthogonal complement of a, i.e., the set of all vectors orthogonal to it:

$$a^\perp \; = \; \{d \in \Re^n \mid a^T d = 0\}.$$

This shows that the hyperplane consists an "offset" of the hyperplane from the origin (i.e., x^0), plus all vectors orthogonal to the (normal) vector.

A hyperplane divides $\Re^n$ into two parts, which are called half spaces.

Definition 1.3 *Let a be a nonzero vector in $\Re^n$ and b be a scalar. Then the set*

$$\{x \in \Re^n \mid a^T x \geq b\}$$

is called a **halfspace***.*

Obviously, a halfspace is a convex set and $\{x \in \Re^n \mid a^T x \leq b\}$ is the other halfspace.
Let x^0 be any point on the hyperplane $\{x \in \Re^n \mid a^T(x - x^0) = 0 \}$. Then the halfspace $\{x \in \Re^n \mid a^T x \geq b \}$ can be expressed as

$$\{x \in \Re^n \mid a^T(x - x^0) \geq 0 \}.$$

This suggests a simple geometric interpretation: the half space consists of x^0 plus any vector that makes an acute angle with the normal vector a.

1.3 Polyhedra

Definition 1.4 *A* **polyhedron** *is a set that can be described in the form $\{x \in \Re^n | \ Ax \geq b\}$, where A is an $m \times n$ matrix and b is a vector in $\Re^m$.*

Let $A \in \Re^{m \times n}$ and $b \in \Re^m$ be defined as follows

$$A = \begin{bmatrix} a_1^T \\ \vdots \\ a_m^T \end{bmatrix}, \qquad b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}.$$

Then the polyhedron defined in Definition 1.4 is the intersection of the following halfspaces

$$\{x \in \Re^n \mid a_i^T x \geq b_i \}, \quad i = 1, \ldots, m.$$

It is noted that these halfspaces are finite in number. The intersection of inifitely many half spaces is not necessarily a polyhedron. To see this, we consider two examples. The first example is

$$S_1 := \{x \in \Re^2 \,|\, x^T y \ \leq \ 1 \ : \ \forall \,|y_1| + |y_2| = 1 \,\}.$$

Certainly, S_1 is the intersection of the half spaces

$$h_y \ = \ \{x \in \Re^2 \,|\, x^T y \ \leq \ 1 \,\}, \qquad \forall \,|y_1| + |y_2| = 1 \,.$$

First, let us consider the half spaces of h_y for $y = (1,0)^T, (0,1)^T, (-1,0)^T$ and $y = (0,-1)^T$, respectively.

Second, consider the half spaces of h_y for $y = (\sqrt{2}/2, \sqrt{2}/2)^T, (\sqrt{2}/2, -\sqrt{2}/2)^T, (-\sqrt{2}/2, \sqrt{2}/2)^T$ and $(-\sqrt{2}/2, -\sqrt{2}/2)^T$, respectively.

Clearly, the above half spaces across $x_y = (sign(y_1), sign(y_2))^T$. This actually can be proven rigorously. Let y be such that $|y_1| + |y_2| = 1$. Then

$$x_y^T y \;=\; y_1 \times sign(y_1) + y_2 \times sign(y_2) \;=\; |y_1| + |y_2| = 1\,.$$

Finally, by considering all possible y such that $|y_1| + |y_2| = 1$ we can see that

$$S_1 = \{x \in \Re^2 \mid -1 \le x_1 \le 1\,,\, -1 \le x_2 \le 1\,\}\,,$$

which is a polyhedral set.

Next, let us consider another example

$$S_2 := \{x \in \Re^2 \,|\, x^T y \;\le\; 1 \;:\; \forall\, \|y\| = 1\,\}\,.$$

Then S_2 is the intersection of the half spaces

$$h_y \;=\; \{x \in \Re^2 \,|\, x^T y \;\le\; 1\}\,, \qquad \forall\, \|y\| = 1\,.$$

Evidently, for any y such that $\|y\| = 1$, $x_y = y$ is a point of h_y because

$$x_y^T y = y^T y = 1\,.$$

It is then an easy excercise to see that

$$S_2 = \{x \in \Re^2 \mid \|x\| \leq 1\},$$

which is not a polyhedral set.

The intersection of two polyhedrons is again a polyhedron. So $\{x \in \Re^n \mid Cx \geq d,\ Ax = b\}$ is also a polyhedron, where $C \in \Re^{p\times n}, d \in \Re^p$.

A polyhedron may have different representations. For example

$$\{x \in \Re^2 \mid x_1 + x_2 = 0,\ x_1 \geq 0\} = \{x \in \Re^2 \mid 2x_1 + 2x_2 \geq 0,\ x_1 + x_2 \leq 0, x_1 \geq 0\}.$$

A bounded polyhedron is sometimes called a **polytope**.

Let e_i be the ith unit vector in $\Re^n$. Then by noting that $x_i = e_i^T x$ we know that the **nonnegative orthant**

$$\Re^n_+ = \{x \in \Re^n \mid x_i \geq 0,\ i = 1, \ldots, n\}.$$

is a polyhedron.

Definition 1.5 *Let $x^1, \ldots, x^k$ be vectors in $\Re^n$ and let $\lambda_1, \ldots, \lambda_k$ be nonnegative scalars whose sum is one.*

(a) The vector $\sum_{i=1}^{n} \lambda_i x^i$ is said to be a **convex combination** *of the vectors $x^1, \ldots, x^k$.*

(b) The **convex hull** *(*conv *in short) of the vectors $x^1, \ldots, x^k$ is the set of all convex combinations of these vectors.*

It is easy to see by Definition 1.5 that

$$\operatorname{conv}\{e_1, \ldots, e_n\} = \left\{ x \in \Re^n \mid \sum_{i=1}^{n} x_i = 1,\ x_i \geq 0,\ i = 1, \ldots, n \right\}$$

is a polytope.

1.4 Extreme points, vertices and basic feasible solutions

We have already known that an optimal solution to a linear programming (assume the existence of an optimal solution) can be found at a "corner" of the polyhedron over which we are optimizing. There quite a number of different but equivalent ways to define the concept of a "corner". Here we introduce three of them.

Our first definition defines an *extreme point* of a polyhedron as a point that can not be expressed as a convex combination of two other points of the polyhedron.

Definition 1.6 *Let $P \subseteq \Re^n$ be a polyhedron. A vector $x \in P$ is an* **extreme point** *of P if we cannot find two vectors $y, z \in P$, both different from x, and a scalar $\lambda \in [0, 1]$, such that $x = \lambda y + (1 - \lambda) z$.*

It can be checked easily that the extreme points of

$$P = \operatorname{conv}\{e_1, e_2, e_3\}$$

are e_1, e_2 and e_3.

Clearly, Definition 1.6 is entirely geometric and does not refer to a specific representation of a polyhedron in terms of linear constraints. An alternative geometric definition defines a *vertex* of a polyhedron P as the unique optimal solution to some linear programming problem with feasible set P.

Definition 1.7 *Let $P \subseteq \Re^n$ be a polyhedron. A vector $x \in P$ is a* **vertex** *of P if there exists some $c \in \Re^n$ such that $c^T x < c^T y$ for all y satisfying $y \in P$ and $y \neq x$*

In other words, x is a vertex of P if and only if P is one side of a hyperplane ($\{y \mid c^T y = c^T x\}$) which meets P only at point x.

The above two geometric definitions are easy to accept, but not easy to work with from an algorithmic point of view (think about the simplex method). We need a definition that relies on a representation of a polyhedron in terms of linear constraints. Some terminology is necessary for this purpose.

Consider a polyhedron $P \subseteq \Re^n$, defined in terms of the linear equality and inequality constraints

$$\begin{aligned} a_i^T x &\geq b_i, \;\; i \in M_1 \\ a_i^T x &\leq b_i, \;\; i \in M_2 \\ a_i^T x &= b_i, \;\; i \in M_3 \end{aligned}$$

where M_1, M_2 and M_3 are finite index sets, each a_i is a vector in $\Re^n$ and each b_i is a scalar.

For example, let

$$P = \{x \in \Re^3 \mid a_1^T x \geq 1, \;\; a_2^T x \leq 3, \;\; a_3^T x = 1, \;\; x \geq 0\} \tag{1.1}$$

where $a_1 = (0, 0, 2)^T, a_2 = (4, 0, 0)^T$ and $a_3 = (1, 1, 1)^T$. Let $a_4 = e_1, a_5 = e_2$ and $a_6 = e_3$. Then

$$M_1 = \{1, 4, 5, 6\}, \;\; M_2 = \{2\}, \;\; M_3 = \{3\}.$$

Definition 1.8 *If a vector x^* satisfies $a_i^T x^* = b_i$ for some $i \in M_1, M_2$ or M_3, we say that the corresponding constraint is* **active** *or* **binding** *at x^*. The* **active set** *of P at x^* is defined as*

$$I(x^*) = \{i \in M_1 \cup M_2 \cup M_3 \mid a_i^T x^* = b_i\},$$

i.e., $I(x^*)$ is the set of indices of constraints that are active at x^*.

For example, suppose that P is defined by (1.1). Let $x^* = (0.5, 0, 0.5)^T$. All active constraints at x^* are

$$a_1^T x \geq 1,\ a_3^T x = 1,\ a_5^T x (= x_2) \geq 0$$

and

$$I(x^*) = \{1, 3, 5\}.$$

Recall that vectors $x^1, \ldots, x^k \in \Re^n$ are said to be linearly independent if

$$\alpha_1 x^1 + \ldots + \alpha_k x^k \ = 0 \quad \Longrightarrow \quad \alpha_1 \ = \ \ldots \ = \alpha_k \ = 0.$$

The maximal number of linearly independent vectors in $\Re^n$ is exactly n. Thus $k \leq n$ if $x^1, \ldots, x^k \in \Re^n$ are linearly independent. Note that $x^1, \ldots, x^n \in \Re^n$ are linearly independent if and only if the matrix $M = [x^1 \ \ldots \ x^n]$ is nonsingular, i.e., the determinant of M is not zero.

If there are n constraints of $P \subseteq \Re^n$ that are active at a vector x^*, then x^* satisfies a certain system of n linear equations in n unknowns. This system has a unique solution if any only if the n vectors a_i of these n equations are linear independent. This is stated precisely in the following proposition.

Proposition 1.1 *Let $x^* \in \Re^n$. The following are equivalent.*

(a) There exist n vectors in the set $\{a_i \mid i \in I(x^*)\}$, which are linearly independent.

(b) The span of the vectors a_i, $i \in I(x^*)$, is all of $\Re^n$, that is, every element of $\Re^n$ can be expressed as a linear combination of the vectors $a_i, i \in I(x^*)$.

(c) The system of equations $a_i^T x = b_i, i \in I(x^*)$, has a unique solution.

[Observations: If $I(x^*)$ contains exactly n elements, the proof of the proposition is trivial. $I(x^*)$ may contain more than n elements.]

Proof. (a) $\Longleftrightarrow$ (b)

Suppose that the vectors $a_i, i \in I(x^*)$, span $\Re^n$. Then, the span of these vectors has dimension n. This clearly implies that exist n vectors in the set $\{a_i \mid i \in I(x^*)\}$, which are linearly independent because otherwise if the maximal number of linearly independent vectors is $k \leq n-1$ the the span of these vectors would have dimension k.

Conversely, suppose that n of the vectors $a_i, i \in I(x^*)$, are linearly independent. Then, the subspace spanned by these n vectors is n-dimensional and must be equal to $\Re^n$. Hence, every element of $\Re^n$ is a linear combination of the vectors $a_i, i \in I(x^*)$.

(b) $\Longleftrightarrow$ (c)

If the system of equations $a_i^T x^* = b_i, i \in I(x^*)$, has multiple solutions, say x^1 and x^2, then the nonzero vector $d = x^1 - x^2$ satisfies $a_i^T d = 0, i \in I(x^*)$. Then for any linear combination $\sum_{i \in I(x^*)} \alpha_i a_i$ of vectors $a_i, i \in I(x^*)$, one has

$$d^T \left(\sum_{i \in I(x^*)} \alpha_i a_i \right) = \sum_{i \in I(x^*)} \alpha_i d^T a_i = 0.$$

This, together with the fact that $d^T d > 0$, shows that d is not a linear combination of these vectors. Thus, $a_i, i \in I(x^*)$ do not span $\Re^n$.

Conversely, if the vectors $a_i, i \in I(x^*)$, do not span $\Re^n$, choose a nonzero vector d which is orthogonal to the subspace spanned by these vectors. If x satisfies $a_i^T x = b_i$ for all $i \in I(x^*)$, we also have $a_i^T (x + d) = b_i$ for all $i \in I(x^*)$, thus obtaining multiple solutions. We have therefore established that (b) and (c) are equivalent. QED.

With a slight abuse of language, we will often say that certain *constraints* are *linearly independent*, meaning that the corresponding vectors a_i are linearly independent. We are now ready to provide an algebraic definition of a corner point of the polyhedron P.

Definition 1.9 *Let $x^* \in \Re^n$.*

(a) The vector x^* is called a **basic solution** *if*

(i) $a_i^T x^* = b_i$, $i \in M_3$.

(ii) Out of $\{a_i\}_{i \in I(x^*)}$, there are n of them that are linearly independent.

(b) If x^* is a basic solution that satisfies all of the constraints, we say that it is a **basic feasible solution**.

Let P be defined by (1.1). Then $x^* = (0.5, 0, 0.5)^T$ is a basic feasible solution because $x^* \in P$ and a_1, a_3, a_5 are linearly independent.

Let us take another example by assuming that $P = \{y \in \Re^m \mid A^T y \leq c\}$, where $A \in \Re^{m \times n}$ and $c \in \Re^n$ (surprisingly familiar? Think about the dual form of a linear programming problem). Then $y \in \Re^m$ is a basic solution if $A_i^T y = c_i$, $i \in J \subseteq \{1, \ldots, n\}$ and there exist m vectors in $\{A_i\}_{i \in J}$ such that they are linearly independent. Here A_i denotes the ith column of A. y is a basic feasible solution if it is a basic solution and $A^T y \leq c$.

Exercise: What is a basic (feasible) solution to $P = \{x \in \Re^n |\ Ax = b,\ x \geq 0\}$?

Note that if the number m of constraints used to define a polyhedron $P \subseteq \Re^n$ is less than n, the number of active constraints at any given point must also be less than n, and there are no basic or basic feasible solutions.

In the next theorem we prove that all these three different definitions to define a "corner" of the polyhedron are equivalent. This is very useful because we can use them interchangeably.

Theorem 1.1 *Let P be a nonempty polyhedron and $x^* \in P$. Then, the following are equivalent:*

(a) x^* *is a vertex;*

(b) x^* *is an extreme point;*

(c) x^* *is a basic feasible solution.*

Proof. Without loss of generality, we assume that P is represented in terms of constraints of the form $a_i^T x \geq b_i$ and $a_i^T x = b_i$.

Vertex $\Longrightarrow$ Extreme point

Suppose that $x^* \in P$ is a vertex. Then, by the definition of vertex, there exists some $c \in \Re^n$ such that

$$c^T x^* \ < \ c^T y, \quad \forall y \in P, \ y \neq x^*.$$

If $y \in P, z \in P$, $y \neq x^*, z \neq x^*$, and $0 \leq \lambda \leq 1$, then

$$c^T x^* \ < \ c^T y \quad \& \quad c^T x^* \ < \ c^T z,$$

which implies that

$$c^T x^* \ < \ c^T(\lambda y + (1 - \lambda) z)$$

and, therefore $x^* \neq \lambda y + (1-\lambda) z$. Thus, x^* cannot be expressed as a convex combination of two other elements of P and is, therefore, an extreme point.

Extreme point $\Longrightarrow$ Basic feasible solution

Suppose that $x^* \in P$ is not a basic feasible solution. We will show that x^* is not an extreme point of P. Let $I = \{i \mid a_i^T x^* = b_i\}$. Since x^* is not a basic feasible solution, there do not exist n linearly independent vectors in the family $a_i, i \in I$. Thus, the vectors $a_i, i \in I$, lie in a proper subspace of $\Re^n$, and there exists some nonzero vector $d \in \Re^n$ such that $a_i^T d = 0$, for all $i \in I$. Let ε be a small positive number and consider the vectors $y = x^* + \varepsilon d$ and $z = x^* - \varepsilon d$. Notice that $a_i^T y = a_i^T x^* = b_i$, for $i \in I$. Furthermore, for $i \notin I$, we have $a_i^T x^* > b_i$ and, provided that ε is small, we will also have $a_i^T y > b_i$. Thus, when ε is small enough, $y \in P$ and, by a similar argument, $z \in P$. We finally notice that $x^* = (y+z)/2$, which implies that x^* is not an extreme point.

Basic feasible solution $\Longrightarrow$ Vertex

Let x^* be a basic feasible solution and let $I = \{i \mid a_i^T x^* = b_i\}$. Let $c = \sum_{i \in I} a_i$. We then have

$$c^T x^* \ = \ \sum_{i \in I} a_i^T x^* \ = \ \sum_{i \in I} b_i.$$

Furthermore, for any $x \in P$ and any i, we have $a_i^T x \geq b_i$, and

$$c^T x \ = \ \sum_{i \in I} a_i^T x \ \geq \ \sum_{i \in I} b_i. \tag{1.2}$$

This shows that x^* is an optimal solution to the problem of minimizing c^Tx over the set P. Further more, equality holds in (1.2) if and only if $a_i^Tx = b_i$ for all $i \in I$. Since x^* is a basic feasible solution, there are n linearly independent constraints that are active at x^*, and x^* is the unique solution to the system of equations $a_i^Tx = b_i, i \in I$ (Proposition 1.1). It follows that x^* is the unique minimizer of c^Tx over the set P and therefore, x^* is a vertex of P. QED

The above theorem shows that a basic feasible solution is independent of the representation used in defining P. This is in contrast to the definition of a basic solution, which is representation dependent.

1.5 Finite basis theorem for polyhedra

Definition 1.10 *A set $C \subseteq \Re^n$ is a* **cone** *if $\lambda x \in C$ for all $\lambda \geq 0$ and all $x \in C$.*

From the definition we can see that $0 \in C$. For vectors $x^1, \ldots, x^k \in \Re^n$, let

$$\text{cone}\{x^1, \ldots, x^k\} = \{x \in \Re^n \mid x = \sum_{i=1}^{k} \lambda_i x^i, \ \lambda_i \geq 0, \ i = 1, \ldots, k\}.$$

Then, $\text{cone}\{x^1, \ldots, x^k\}$ is a cone and convex set, which is called the convex cone generated by $x^1, \ldots, x^k$.

The set $P = \{x \in \Re^n \mid Ax \geq 0\}$ is called a **polyhedral cone**.

Excercise: Is $\text{cone}\{x^1, \ldots, x^k\}$ a polyhedral cone?

Given $A \in \Re^{m\times n}$ and $b \in \Re^m$, consider

$$P = \{x \in \Re^n \mid Ax \geq b\}$$

and $y \in P$.

Definition 1.11 *The* **recession cone** *of P at y is defined as the set*

$$\{d \in \Re^n |\ A(y + \lambda d) \geq b,\ \text{for all } \lambda \geq 0\}.$$

Roughly speaking, the recession cone of P at y is the set of all directions d along which we can move indefinitely away from y, without leaving the set P. It can be easily seen that the recession cone of P at y is the same as

$$\{d \in \Re^n |\ Ad \geq 0\},$$

and is the polyhedral cone. This means that the recession cone is independent of the starting point y. For $P = \{x \in \Re^n | Ax = b, x \geq 0\}$, where $A \in \Re^{m\times n}$ and $b \in \Re^m$, the recession cone is

$$\{d \in \Re^n |\ Ad = 0, d \geq 0\}.$$

Definition 1.12

(a) A nonzero element d of a polyhedral cone $C \subseteq \Re^n$ is called an **extreme ray** *if there are $n-1$ linearly independent constraints that are active at d.*

(b) An extreme ray of the recession cone associated with a nonempty polyhedron P is also called an extreme ray of P.

For example, consider the simple polyhedral cone $\Re^n_+$. Then the extreme rays of $\Re^n_+$ are

$$\{e_1, \ldots, e_n\}.$$

Since there are n linearly independent constraints that are active at the zero vector, by the definition the zero vector is an extreme point of $\Re^n_+$ (actually the only extreme point).

The following theorem is called the finite basis theorem for polyhedra, obtained by Minkowski.

Theorem 1.2 *If $P = \{x \in \Re^n \mid Ax = b, x \geq 0\}$, then P is finitely generated. In particular, there exist $x^1, \ldots, x^q, d^1, \ldots, \ldots d^r$ in $\Re^n$ such that*

$$P = \text{conv}\{x^1, \ldots, x^q\} + \text{cone}\{d^1, \ldots, d^r\}.$$

Furthermore, the extreme points (basic feasible solutions) of P are contained in the set $\{x^1, \ldots, x^q\}$ and the extreme rays of the recession cone $\{d \in \Re^n \mid Ad = 0, d \geq 0\}$ are contained in the set $\{d^1, \ldots, d^r\}$.

Note that Theorem 1.2 holds for any polyhedral set and it implies that P has only finitely many extreme points. It can also be seen easily that P is bounded if and only if the recession cone of P contains the zero vector only.