Appendix

PRELIMINARIES

1. THEOREMS OF ALTERNATIVES FOR SYSTEMS OF LINEAR CONSTRAINTS

Here we consider systems of linear constraints, consisting of equations or inequalities or both. A feasible solution of a system is a vector which satisfies all the constraints in the system. If a feasible solution exists, the **system** is said to be **feasible**. The system is said to be **infeasible** if there exists no feasible solution for it. A typical theorem of alternatives shows that corresponding to any given system of linear constraints, system I, there is another associated system of linear constraints, system II, based on the same data, satisfying the property that one of the systems among I, II is feasible iff the other is infeasible. These theorems of alternatives are very useful for deriving optimality conditions for many optimization problems.

First consider systems consisting of linear equations only. The **fundamental inconsistent equation** is

$$0 = 1 \tag{1}$$

consider the following system of equations

$$\begin{aligned} x_1 + x_2 + x_3 &= 2 \\ -x_1 - x_2 - x_3 &= -1. \end{aligned} \tag{2}$$

When we add the two equations in (2), the coefficients of all the variables on the left hand side of the sum are zero, and the right hand side constant is 1. Thus the

fundamental inconsistent equation (1) can be obtained as a linear combination of the two equations in (2). This clearly implies that there exists no feasible solution for (2).
Now consider the general system of linear equations $Ax = b$, written out in full as

$$\sum_{j=1}^{n} a_{ij}x_j = b_i,\ i = 1 \text{ to } m. \tag{3}$$

A linear combination of system (3) with coefficients $\pi = (\pi_1, \ldots, \pi_m)$ is

$$\sum_{j=1}^{n}\Big(\sum_{i=1}^{m} \pi_i a_{ij}\Big)x_j = \Big(\sum_{i=1}^{m} \pi_i b_i\Big) \tag{4}$$

(4) is the same as $(\pi A)x = (\pi b)$. (4) becomes the fundamental inconsistent equation (1) if

$$\begin{aligned} \sum_{i=1}^{m} \pi_i a_{ij} &= 0,\ j = 1 \text{ to } n \\ \sum_{i=1}^{m} \pi_i b_i &= 1 \end{aligned} \tag{5}$$

and in this case, (3) is clearly infeasible. The system of linear equations (3) is said to be **inconsistent** iff the fundamental inconsistent equation (1) can be obtained as a linear combination of the equations in (3), that is, iff there exists $\pi = (\pi_1, \ldots, \pi_m)$ satisfying (5). Clearly an inconsistent system of equations is infeasible. The converse of this statement is also true. So a system of linear equations is infeasible iff it is inconsistent. This is implied by the following theorem of alternatives for systems of linear equations.

Theorem 1 *Let $A = (a_{ij})$, $b = (b_i)$ be given matrices of orders $m \times n$ and $m \times 1$. Let $x = (x_1, \ldots, x_n)^T$ and $\pi = (\pi_1, \ldots, \pi_m)$. Exactly one of the two following systems (I) and (II) has a solution and the other has no solution.*

(I)	(II)
$Ax = b$	$\pi A = 0$
	$\pi b = 1$

Proof. If (I) has a solution $\overline{x}$ and (II) has a solution $\overline{\pi}$, then $A\overline{x} = b$, and so $\overline{\pi}A\overline{x} = \overline{\pi}b$, but $(\overline{\pi}A)\overline{x} = 0$, $\overline{\pi}b = 1$, so this is impossible. So it is impossible for both (I) and (II) to have solutions.

Put (I) in detached coefficient tabular form and introduce the unit matrix of order m on the left hand side of this tableau. The tableau at this stage is

		x	
I		A	b

Perform Gauss-Jordan pivot steps in this tableau to put A into row echelon normal form. For this, perform Gauss-Jordan pivot steps in rows 1 to m, in that order. Consider the step in which the rth row is the pivot row. Let the entries in the current tableau in the rth row at this stage be

$\beta_{r1} \dots \beta_{rm}$		$\bar{a}_{r1} \dots \bar{a}_{rn}$	$\bar{b}_r$

Let $\beta_{r.} = (\beta_{r1}, \dots, \beta_{rm})$. Then, $(\bar{a}_{r1}, \dots, \bar{a}_{rn}) = \beta_{r.}A$ and $\bar{b}_r = \beta_{r.}b$. If $(\bar{a}_{r1}, \dots, \bar{a}_{rn})$ $= 0$ and $\bar{b}_r = 0$, this row at this stage represents a redundant constraint, erase it from the tableau and continue. If $(\bar{a}_{r1}, \dots, \bar{a}_{rn}) = 0$ and $\bar{b}_r \neq 0$, define $\bar{\pi} = \beta_{r.}/\bar{b}_r$. Then we have $\bar{\pi}A = 0$, $\bar{\pi}b = 1$, so $\bar{\pi}$ is a feasible solution of system (II) and (I) has no feasible solution, terminate. If $(\bar{a}_{r1}, \dots, \bar{a}_{rn}) \neq 0$, select a j such that $\bar{a}_{rj} \neq 0$, and perform a pivot step with row r as the pivot row and column j as the pivot column, make x_j the basic variable in the rth row, and continue. If the conclusion that (I) is infeasible is not made at any stage in this process, make the basic variable in each row equal to the final updated right hand side constant in that row, and set all the nonbasic variables equal to zero; this is a solution for system (I). Since (II) cannot have a solution when (I) does, (II) has no solution in this case.

□

Example 1

Let

$$A = \begin{pmatrix} 1 & -2 & 2 & -1 & 1 \\ -1 & 0 & 4 & -7 & 7 \\ 0 & -2 & 6 & -8 & 8 \end{pmatrix}, \quad b = \begin{pmatrix} -8 \\ 16 \\ 6 \end{pmatrix}.$$

So, system (I) in Theorem 1 corresponding to this data is

x_1	x_2	x_3	x_4	x_5	b
1	−2	2	−1	1	− 8
−1	0	4	−7	7	16
0	−2	6	−8	8	6

We introduce the unit matrix of order 3 on the left hand side and apply the Gauss-Jordan method on the resulting tableau. This leads to the following work. Pivot elements are inside a box.

			x_1	x_2	x_3	x_4	x_5	
1	0	0	$\boxed{1}$	-2	2	-1	1	-8
0	1	0	-1	0	4	-7	7	16
0	0	1	0	-2	6	-8	8	6
1	0	0	1	-2	2	-1	1	-8
1	1	0	0	$\boxed{-2}$	6	-8	8	8
0	0	1	0	-2	6	-8	8	6
0	-1	0	1	0	-4	7	-7	-16
$-\frac{1}{2}$	$-\frac{1}{2}$	0	0	1	-3	4	-4	-4
-1	-1	1	0	0	0	0	0	-2

From the last row in the last tableau, we conclude that this system is inconsistent. Defining $\overline{\pi} = (-1, -1, 1)/(-2) = (1/2, 1/2, -1/2)$, we verify that $\overline{\pi}$ is a solution for system (II) in Theorem 1 with data given above.

Now consider a system of linear inequalities. The **fundamental inconsistent inequality** is

$$0 \geqq 1 \tag{6}$$

Consider the following system of inequalities.

$$\begin{aligned} x_1 + x_2 + x_3 &\geqq 2 \\ -x_1 - x_2 - x_3 &\geqq -1. \end{aligned} \tag{7}$$

Adding the two inequalities in (7) yields the fundamental inconsistent inequality (6), this clearly implies that no feasible solution exists for (7).

Given the system of linear inequalities

$$\sum_{j=1}^{n} a_{ij} x_j \geqq b_i, \ i = 1 \text{ to } m \tag{8}$$

a **valid linear combination** of (8) is a linear combination of the constraints in (8) with nonnegative coefficients, that is

$$\sum_{j=1}^{n} \Big(\sum_{i=1}^{m} \pi_i a_{ij} \Big) x_j \geqq \sum_{i=1}^{m} \pi_i b_i \tag{9}$$

where $\pi = (\pi_1, \dots, \pi_m) \geqq 0$. (9) is the fundamental inconsistent equation (6) iff

$$\begin{aligned} \sum_{j=1}^{n} \pi_i a_{ij} &= 0, \ j = 1 \text{ to } m \\ \sum_{i=1}^{m} \pi_i b_i \ &= 1 \end{aligned} \tag{10}$$

and if (10) has a solution $\pi \geqq 0$, (8) is clearly infeasible. The system of linear inequalities (8) is said to be **inconsistent** iff the fundamental inconsistent inequality (6) can be obtained as a valid linear combination of it. We will prove below, that a system of linear inequalities is infeasible iff it is inconsistent. In fact, given any system of linear constraints (consisting of equations and/or inequalities) we will prove that it has no feasible solution iff the fundamental inconsistent inequality (6) can be obtained as a valid linear combination of it. This leads to a theorem of alternatives for that system. These theorems of alternatives can be proven in several ways. One way is by using the duality theorem of linear programming (see [2.26]). Another way is to prove them directly using a lemma proved by A. W. Tucker. We first discuss this Tucker's lemma [see A 10].

Theorem 2 (Tucker's Lemma). *If A is a given $m \times n$ real matrix, the systems*

$$Ax \geqq 0 \tag{11}$$

$$\pi A = 0, \ \pi \geqq 0 \tag{12}$$

where $x = (x_1, \dots, x_n)^T$ and $\pi = (\pi_1, \dots, \pi_m)$, have feasible solutions $\overline{x}$, $\overline{\pi}$ respectively, satisfying

$$(\overline{\pi})^T + A\overline{x} > 0. \tag{13}$$

Proof. We will first prove that there exist feasible solutions x^1, $\pi^1 = (\pi^1_1, \dots, \pi^1_m)$ to (11), (12) respectively, satisfying

$$A_{1.}x^1 + \pi^1_1 > 0. \tag{14}$$

The proof is by induction on m, the number of rows in the matrix A. If $m = 1$ let

$$\begin{aligned} &\pi^1 = (\pi^1_1) = (1), \ x^1 = 0, \ \text{ if } A_{1.} = 0 \\ &\pi^1 = 0, \ x^1 = (A_{1.})^T, \ \text{ if } A_{1.} \neq 0 \end{aligned}$$

and verify that these solutions satisfy (14). So the theorem is true if $m = 1$. We now set up an induction hypothesis.

Induction Hypothesis. If D is any real matrix of order $(m-1) \times n$, there exist vectors $x = (x_j) \in \mathbf{R}^n$, $u = (u_1, \dots, u_{m-1})$ satisfying: $Dx \geqq 0$; $uD = 0$, $u \geqq 0$; $u_1 + D_{1.}x > 0$.

Under the induction hypothesis we will now prove that this result also holds for the matrix A of order $m \times n$. Let A' be the $(m-1) \times n$ matrix obtained by deleting the last row $A_{m.}$ from A. Applying the induction hypothesis on A', we know that there exist $x' \in \mathbf{R}^n$, $u' = (u'_1, \dots, u'_{m-1})$ satisfying

$$A'x' \geqq 0;\ u'A' = 0,\ u' \geqq 0;\ u'_1 + A_{1.}x' > 0. \tag{15}$$

If $A_{m.}x' \geqq 0$, define $x^1 = x'$, $\pi^1 = (u', 0)$, and verify that x^1, π^1 are respectively feasible to (11), (12) and satisfy (14), by (15). On the other hand, suppose $A_{m.}x' < 0$. We now attempt to find a vector $\tilde{x} \in \mathbf{R}^n$ and real number α such that

$$x^1 = \tilde{x} + \alpha x' \text{ and vector } \pi^1$$

together satisfy (14). We have to determine $\tilde{x}$, α, π^1 so that this will be true. For this we require

$$A_{m.}x^1 = A_{m.}\tilde{x} + \alpha A_{m.}x' \geqq 0 \text{ that is}$$
$$\alpha \leqq (A_{m.}\tilde{x})/(-A_{m.}x').$$

So it suffices if we define $\alpha = (A_{m.}\tilde{x})/(-A_{m.}x')$. We still have to determine $\tilde{x}$ and π^1 appropriately. The vector x^1 should also satisfy for $i = 1$ to $m - 1$

$$A_{i.}x^1 = A_{i.}\tilde{x} + \alpha A_{i.}x' = (A_{i.} + \lambda_i A_{m.})\tilde{x} \geqq 0$$

where $\lambda_i = (A_{i.}x')/(-A_{m.}x')$. Now define $B_{i.} = A_{i.} + \lambda_i A_{m.}$, for $i = 1$ to $m - 1$ and let B be the $(m-1) \times n$ matrix whose rows are $B_{i.}$, $i = 1$ to $m - 1$. By applying the induction hypothesis on B, we know that there exists $x'' \in \mathbf{R}^n$, $u'' = (u''_1, \dots, u''_{m-1})$ satisfying

$$Bx'' \geqq 0,\ u''B = 0,\ u'' \geqq 0,\ u''_1 + B_{1.}x'' > 0. \tag{16}$$

We take this vector x'' to be the $\tilde{x}$ we are looking for, and therefore define

$$x^1 = x'' - x'(A_{m.}x'')/(A_{m.}x')$$
$$\pi^1 = \Big(u'', \sum_{i=1}^{m} \lambda_i u''_i\Big).$$

Using (15), (16) and the fact that $A_{m.}x' < 0$ in this case, verify that x^1, π^1 are respectively feasible to (11) and (12) and satisfy (14). So under the induction hypothesis, the result in the induction hypothesis also holds for the matrix A of order $m \times n$. The result in the induction hypothesis has already been verified to be true for matrices with 1 row only. So, by induction, we conclude that there exist feasible solutions x^1, π^1 to (11), (12) respectively, satisfying (14).

For any $i = 1$ to m, the above argument can be used to show that there exist feasible solutions x^i, $\pi^i = (\pi^i_1, \dots, \pi^i_m)$ to (11) and (12) respectively satisfying

$$\pi^i_i + A_{i.}x^i > 0. \tag{17}$$

Define $\overline{x} = \sum_{i=1}^m x^i$, $\overline{\pi} = \sum_{i=1}^m \pi^i$, and verify that $\overline{x}$, $\overline{\pi}$ together satisfy (11) and (12) and (13).

∎

Corollary 1. *Let A, D be matrices of orders $m_1 \times n$ and $m_2 \times n$ respectively with $m_1 \geqq 1$. Then there exist $x = (x_1, \ldots, x_n)^T$, $\pi = (\pi_1, \ldots, \pi_{m_1})$. $\mu = (\mu_1, \ldots, \mu_{m_2})$ satisfying*

$$\begin{array}{ccc} (18) & (19) & (20) \\ \hline Ax \geqq 0 & \pi A + \mu D = 0 & \pi^T + Ax > 0\,. \\ Dx = 0 & \pi \geqq 0 & \end{array}$$

Proof. Applying Tucker's lemma to the systems

$$\begin{array}{rr} (21) & (22) \\ \hline Ax \geqq 0 & \pi A + \gamma D - \nu D = 0 \\ Dx \geqq 0 & \pi, \gamma, \nu \geqq 0 \\ -Dx \geqq 0 & \end{array}$$

we know that there exist $\overline{x}$, $\overline{\pi}$, $\overline{\gamma}$, $\overline{\nu}$ feasible to them, satisfying $\overline{\pi}^T + A\overline{x} > 0$. Verify that $\overline{x}$, $\overline{\pi}$, $\overline{\mu} = \overline{\gamma} - \overline{\nu}$ satisfy (18), (19) and (20).

□

We will now discuss some of the most useful theorems of alternatives for linear systems of constraints.

Theorem 3 (Farkas' Theorem). *Let A, b be given matrices of orders $m \times n$ and $m \times 1$ respectively. Let $x = (x_1, \ldots, x_n)^T$, $\pi = (\pi_1, \ldots, \pi_m)$. Exactly one of the following two systems (I), (II) is feasible.*

$$\begin{array}{rl} \text{(I)} & \text{(II)} \\ \hline Ax = b & \pi A \leqq 0 \\ x \geqq 0 & \pi b \ > 0. \end{array}$$

Proof. Suppose both systems are feasible. Let $\overline{x}$ be feasible to (I) and $\overline{\pi}$ be feasible to (II). Then $(\overline{\pi}A)\overline{x} \leqq 0$ since $\overline{\pi}A \leqq 0$ and $\overline{x} \geqq 0$. Also $\overline{\pi}(A\overline{x}) = \overline{\pi}b > 0$. So there is a contradiction. So it is impossible for both systems (I) and (II) to be feasible.

Suppose (II) is infeasible. Let $y = \pi^T$. So this implies that in every solution of

$$\begin{pmatrix} b^T \\ -A^T \end{pmatrix} \ y \geqq 0 \tag{23}$$

the first constraint always holds as an equation. By Tucker's lemma (Theorem 2) there exists a $\overline{y}$ feasible to (23) and $(\overline{\delta}, \overline{\mu}_1, \ldots, \overline{\mu}_n) \geqq 0$ feasible to

$$(\overline{\delta}, \overline{\mu}_1, \ldots, \overline{\mu}_n) \begin{pmatrix} b^T \\ -A^T \end{pmatrix} = 0 \tag{24}$$

which together satisfy $b^T\overline{y} + \overline{\delta} > 0$. But since $\overline{y}$ is feasible to (23) we must have $b^T\overline{y} = 0$ as discussed above (since (II) is infeasible) and so $\overline{\delta} > 0$. Define $\overline{x}_j = \overline{\mu}_j/\overline{\delta}$ for $j = 1$ to n and let $\overline{x} = (\overline{x}_1, \ldots, \overline{x}_n)^T$. From (24) we verify that $\overline{x}$ is feasible to (I). So if (II) is infeasible, (I) is feasible. Thus exactly one of the two systems (I), (II) is feasible.

□

Note 1. Given A, b, the feasibility of system (I) in Farkas' theorem can be determined using Phase I of the Simplex Method for linear programming problems. If (I) is feasible, Phase I terminates with a feasible solution of (I), in this case system (II) has no feasible solution. If Phase I terminates with the conclusion that (I) is infeasible, the Phase I dual solution at termination provides a vector π which is feasible to system (II).

Note 2. Theorem 3, Farkas' theorem, is often called Farkas' lemma in the literature.

An Application of Farkas' Theorem to Derive Optimality Conditions for LP

To illustrate an application of Farkas' theorem, we will now show how to derive the necessary optimality conditions for a linear program using it. Consider the LP

$$\begin{array}{ll} \text{minimize} & f(x) = cx \\ \text{subject to} & Ax \geqq b \end{array} \tag{25}$$

where A is a matrix of order $m \times n$. The constraints in (25) include all the conditions in the problem, including any bound restrictions, lower or upper, on individual variables. If there are any equality constraints in the problem, each of them can be represented by the corresponding pair of opposing inequality constraints and expressed in the form given in (25) (for example, the equality constraint $x_1 + x_2 - x_3 = 1$ is equivalent to the pair of inequality constraints $x_1 + x_2 - x_3 \geqq 1$, $-x_1 - x_2 + x_3 \geqq -1$). Thus every linear program can be expressed in this form. We now state the necessary optimality conditions for a feasible solution $\overline{x}$ to be optimal to this LP, and prove it using Farkas' theorem.

Theorem 4. *If $\overline{x}$ is a feasible solution for (25), and $\overline{x}$ is optimal to (25), there must exist a vector $\overline{\Pi} = (\overline{\Pi}_1, \ldots, \overline{\Pi}_m)$ which together with $\overline{x}$ satisfies*

$$\begin{aligned} c - \overline{\Pi} A &= 0 \\ \overline{\Pi} &\geqq 0 \\ \overline{\Pi}_i (A_{i.}\overline{x} - b_i) &= 0, \ i = 1 \text{ to } m. \end{aligned} \tag{26}$$

Proof. Consider the case $c = 0$ first. In this case the objective value is a constant, zero, and hence every feasible solution of (25) is optimal to it. It can be verified that $\overline{\Pi} = 0$ satisfies (26) together with any feasible solution $\overline{x}$ for (25).

Now consider the case $c \neq 0$. We claim that the fact that $\overline{x}$ is optimal to (25) implies that $A\overline{x} \not> b$ in this case. To prove this claim, suppose $A\overline{x} > b$. For any $y \in \mathbf{R}^n$, $A(\overline{x} + \alpha y) = A\overline{x} + \alpha A y \geqq b$ as long as α is sufficiently small, since $A\overline{x} > b$.

Take $y = -c^T$. Then, for $\alpha > 0$, $c(\overline{x} + \alpha y) < c\overline{x}$ and $\overline{x} + \alpha y$ is feasible to (25) as long as α is positive and sufficiently small, contradicting the optimality of $\overline{x}$ to (25). So, if $\overline{x}$ is optimal to (25) in this case ($c \neq 0$) at least one of the constraints in (25) must hold as an equation at $\overline{x}$.

Rearrange the rows of A, and let A_1 be the matrix of order $m_1 \times n$ consisting of all the rows in A corresponding to constraints in (25) which hold as equations in (25) when $x = \overline{x}$, and let A_2, of order $m_2 \times n$, be the matrix consisting of all the other rows of A.

By the above argument A_1 is nonempty, that is, $m_1 \geqq 1$. Let b^1, b^2 be the corresponding partition of b. So

$$A = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}, \quad b = \begin{pmatrix} b^1 \\ b^2 \end{pmatrix} \tag{27}$$

and

$$\begin{aligned} A_1\overline{x} &= b^1 \\ A_2\overline{x} &> b^2. \end{aligned} \tag{28}$$

We now show that if $\overline{x}$ is optimal to (25), the system

$$\begin{aligned} A_1 y &\geqq 0 \\ cy &< 0 \end{aligned} \tag{29}$$

cannot have a solution y. Suppose not. Let $\overline{y}$ be a solution for (29). Then for $\alpha > 0$, $A_1(\overline{x} + \alpha\overline{y}) = A_1\overline{x} + \alpha A_1\overline{y} \geqq b^1$; and $A_2(\overline{x} + \alpha\overline{y}) = A_2\overline{x} + \alpha A_2\overline{y} \geqq b^2$ as long as α is sufficiently small, since $A_2\overline{x} > b^2$. So when α is positive but sufficiently small, $\overline{x} + \alpha\overline{y}$ is feasible to (25) and since $c(\overline{x} + \alpha\overline{y}) = c\overline{x} + \alpha c\overline{y} < c\overline{x}$, since $c\overline{y} < 0$, we have a contradiction to the optimality of $\overline{x}$ for (25).

So, (29) has no solution y. By taking transposes, we can put (29) in the form of system (II) under Theorem 3 (Farkas' theorem). Writing the corresponding system (I) and taking transposes again, we conclude that since (29) has no solution, there exists a row vector Π^1 satisfying

$$\begin{aligned} \Pi^1 A_1 &= c \\ \Pi^1 \quad &\geqq 0 \end{aligned} \tag{30}$$

Define $\Pi^2 = 0$ and let $\overline{\Pi} = (\Pi^1, \Pi^2)$. From the fact that A_1, A_2 is a partition of A as in (27), and using (30), (28), we verify that $\overline{\Pi} = (\Pi^1, \Pi^2)$ satisfies (26) together with $\overline{x}$.

□

Example 2

Consider the LP

$$\begin{array}{llrrrrrcr} \text{minimize} & f(x) = & -3x_1 + & x_2 + & 3x_3 & & +5x_5 & & \\ \text{subject to} & & x_1 + & x_2 - & x_3 + & 2x_4 - & x_5 & \geqq & 5 \\ & & -2x_1 & & +2x_3 - & x_4 & +3x_5 & \geqq & -8 \\ & & x_1 & & & & & \geqq & 6 \\ & & & -3x_2 & & +3x_4 & & \geqq & -5 \\ & & & & 5x_3 - & x_4 & +7x_5 & \geqq & 7. \end{array}$$

Let $\overline{x} = (6, 0, -1, 0, 2)^T$. Verify that $\overline{x}$ satisfies constraints 1,2 and 3 in the problem as equations and the remaining as strict inequalities. We have

$$A_1 = \begin{pmatrix} 1 & 1 & -1 & 2 & -1 \\ -2 & 0 & 2 & -1 & 3 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$b^1 = \begin{pmatrix} 5 \\ -8 \\ 6 \end{pmatrix}$$

$$A_2 = \begin{pmatrix} 0 & -3 & 0 & 3 & 0 \\ 0 & 0 & 5 & -1 & 7 \end{pmatrix},$$

$$b^2 = \begin{pmatrix} -5 \\ 7 \end{pmatrix}$$

$$c = (-3, 1, 3, 0, 5), \quad f(x) = cx$$

and $A_1\overline{x} = b^1$, $A_2\overline{x} > b^2$. If we take $\Pi^1 = (1, 2, 0)$ then $\Pi^1 A_1 = c$, $\Pi^1 \geqq 0$. Let $\Pi^2 = (0, 0)$, and $\overline{\Pi} = (\Pi^1, \Pi^2) = (1, 2, 0, 0, 0)$. These facts imply that $\overline{\Pi}$, $\overline{x}$ together satisfy the necessary optimality conditions (26) for this LP.

We leave it to the reader to verify that if $\overline{x}$ is feasible to (25), and there exists a vector $\overline{\Pi}$ such that $\overline{x}$, $\overline{\Pi}$ together satisfy (26), then $\overline{x}$ is in fact optimal to (25), from first principles. Thus the conditions (26) and feasibility are together necessary and sufficient optimality conditions for the LP (25). It can also be verified that any $\overline{\Pi}$ satisfying (26) is an optimum dual solution associated with the LP (25); and that (26) are in fact the dual feasibility and complementary slackness optimality conditions for the LP (25). See [2.26, A10]. Thus Farkas' theorem leads directly to the optimality conditions for the LP (25). Later on, in Appendix 4, we will see that Theorems of alternatives like Farkas' theorem and others discussed below are very useful for deriving optimality conditions in nonlinear programming too. We will now discuss some more theorems of alternatives.

Some Other Theorems of Alternatives

Theorem 5 (Motzkin's Theorem of the Alternatives). *Let $m \geqq 1$, and let A, B, C be given matrices of orders $m \times n$, $m_1 \times n$, $m_2 \times n$. Let $x = (x_1, \ldots, x_n)^T$, $\pi = (\pi_1, \ldots, \pi_m)$, $\mu = (\mu_1, \ldots, \mu_{m_1})$, $\gamma = (\gamma_1, \ldots, \gamma_{m_2})$. Then exactly one of the following two systems (I), (II) is feasible.*

(I)	(II)
$Ax > 0$	$\pi A + \mu B + \gamma C = 0$
$Bx \geqq 0$	$\pi \geq 0,\ \mu \geqq 0$
$Cx = 0$	

Proof. As in the proof of Theorem 3, it can be verified that if both (I), (II) are feasible, there is a contradiction. Suppose system (I) is infeasible. This implies that every feasible solution of

$$\begin{aligned} Ax &\geqq 0 \\ Bx &\geqq 0 \\ Cx &= 0 \end{aligned} \tag{31}$$

satisfies $A_{i.}x = 0$ for at least one $i = 1$ to m. By Corollary 1, there exists $\overline{x}$ feasible to (31) and $\overline{\pi}$, $\overline{\mu}$, $\overline{\gamma}$ feasible to

$$\begin{aligned} \overline{\pi}A + \overline{\mu}B + \overline{\gamma}C &= 0 \\ \overline{\pi} \geqq 0,\ \overline{\mu} &\geqq 0 \end{aligned} \tag{32}$$

satisfying $(\overline{\pi})^T + A\overline{x} > 0$. But since $\overline{x}$ is feasible to (31), $A_{i.}\overline{x} = 0$ for at least one i as discussed above. This implies that for that i, $\overline{\pi}_i > 0$, that is, $\overline{\pi} \geq 0$. So $(\overline{\pi}, \overline{\mu}, \overline{\gamma})$ satisfies (II). So if (I) is infeasible, (II) is feasible. Thus exactly one of the two systems (I), (II) is feasible.

□

Theorem 6 (Gordan's Theorem of the Alternatives). *Give a matrix A of order $m \times n$, exactly one of the following systems (I) and (II) is feasible.*

(I)	(II)
$Ax > 0$	$\pi A = 0$
	$\pi \quad \geq 0$

Proof. Follows from Theorem 5 by selecting B, $C = 0$ there.

□

Theorem 7 (Tucker's Theorem of the Alternatives). *Let $m \geqq 1$, and let A, B, C be given matrices of orders $m \times n$, $m_1 \times n$, $m_2 \times n$ respectively. Let $x = (x_1, \ldots, x_n)^T$, $\pi = (\pi_1, \ldots, \pi_m)$, $\mu = (\mu_1, \ldots, \mu_{m_1})$, $\gamma = (\gamma_1, \ldots, \gamma_{m_2})$. Exactly one of the following systems (I), (II) is feasible.*

(I)	(II)
$Ax \geq 0$	$\pi A + \mu B + \gamma C = 0$
$Bx \geqq 0$	$\pi > 0,\ \mu \geqq 0$
$Cx = 0$	

Proof. As in the proof of Theorem 3, it can be verified that if both (I), (II) are feasible, there is a contradiction. Suppose that (I) is infeasible. This implies that every feasible solution of

$$\begin{aligned} Ax &\geqq 0 \\ Bx &\geqq 0 \\ Cx &= 0 \end{aligned} \tag{33}$$

must satisfy $Ax = 0$. By Corollary 1, there exists $\overline{x}$ feasible to (33) and $\overline{\pi}$, $\overline{\mu}$, $\overline{\gamma}$ feasible to

$$\begin{aligned} \overline{\pi}A + \overline{\mu}B + \overline{\gamma}C &= 0 \\ \pi \geqq 0,\ \mu &\geqq 0 \end{aligned} \tag{34}$$

satisfying $(\overline{\pi})^T + A\overline{x} > 0$. But since $\overline{x}$ is feasible to (33), $A\overline{x} = 0$ as discussed above; so $(\overline{\pi})^T > 0$. So $(\overline{\pi}, \overline{\mu}, \overline{\gamma})$ satisfies (II). So if (I) is infeasible, (II) is feasible. Thus exactly one of the two systems (I), (II) is feasible.

□

Theorem 8 (Gale's Theorem of Alternatives). *Let A, b be given matrices of orders $m \times n$, $m \times 1$ respectively. Let $x = (x_1, \ldots, x_n)^T$, $\pi = (\pi_1, \ldots, \pi_m)$. Exactly one of the following systems (I), (II) is feasible.*

(I)	(II)
$Ax \geqq b$	$\pi A = 0$
	$\pi b = 1$
	$\pi \geqq 0$

Proof. System (I) is equivalent to

$$\begin{aligned} \begin{pmatrix} A & -b \end{pmatrix} \begin{pmatrix} x \\ x_{n+1} \end{pmatrix} &\geqq 0 \\ d \begin{pmatrix} x \\ x_{n+1} \end{pmatrix} &> 0 \end{aligned} \tag{35}$$

where $d = (0, 0, \ldots, 0, 1) \in \mathbf{R}^{n+1}$. (I) is equivalent to (35) in the sense that if a solution of one of these systems is given, then a solution of the other system in the pair can be constructed from it. For example if $\overline{x}$ is a feasible solution of (I), then $(\overline{x}, \overline{x}_{n+1} = 1)$ is a feasible solution of (35). Conversely, if $(\hat{x}, \hat{x}_{n+1})$ is a feasible solution of (35), then $\hat{x}_{n+1} > 0$ and $(1/\hat{x}_{n+1})\hat{x}$ is a feasible solution of (I).

This theorem follows from Theorem 5 applied to (35).

□

For a complete discussion of several other Theorems of alternatives for linear systems and their geometric interpretation, see O. L. Mangasarian's book [A10].

Exercises

1. Let $\mathbf{K}$ be the set of feasible solutions of

$$\sum_{j=1}^{n} a_{ij} x_j \geqq b_i,\ i = 1 \text{ to } m. \tag{36}$$

Assume that $\mathbf{K} \neq \emptyset$. Prove that all $x \in \mathbf{K}$ satisfy

$$\sum_{j=1}^{n} c_j x_j \geqq d \tag{37}$$

iff, for some $\alpha \geqq d$, the inequality $\sum_{j=1}^{n} c_j x_j \geqq \alpha$ is a valid linear combination of the constraints in (36), that is, iff there exists $\overline{\pi} = (\overline{\pi}_1, \dots, \overline{\pi}_m) \geqq 0$, satisfying $c_j = \sum_{i=1}^{m} \overline{\pi}_i a_{ij}$, $j = 1$ to n, and $\alpha = \sum_{i=1}^{m} \overline{\pi}_i b_i$.

2. Let M be a square matrix of order n. Prove that for each $q \in \mathbf{R}^n$, the system "$Mx + q \geqq 0$, $x \geqq 0$" has a solution $x \in \mathbf{R}^n$ iff the system "$My > 0$, $y \geqq 0$" has a solution y. (O. L. Mangasarian [3.42])

3. Let M be a square matrix of order n and $q \in \mathbf{R}^n$. Prove that the following are equivalent

i) the system $Mx + q > 0$, $x \geqq 0$ has a solution $x \in \mathbf{R}^n$,
ii) the system $Mx + q > 0$, $x > 0$ has a solution $x \in \mathbf{R}^n$,
iii) the system $M^T u \leqq 0$, $q^T u \leqq 0$, $0 \leq u$ has no solution $u \in \mathbf{R}^n$.

(O. L. Mangasarian [3.42])

4. Prove that (36) is infeasible iff it is inconsistent (that is, the fundamental inconsistent inequality (6) can be obtained as a valid linear combination of it) as a corollary of the result in Exercise 1.

5. Let A be an $m \times n$ matrix, and suppose the system: $Ax = b$, has at least one solution; and the equation $cx = d$ holds at all solutions of the system $Ax = b$. Then prove that the equation $cx = d$ can be obtained as a linear combination of equations from the system $Ax = b$. That is, there exists $\pi = (\pi_1, \dots, \pi_m)$, such that $c = \pi A$ and $d = \pi b$.

---

2. CONVEX SETS

A subset $\mathbf{K} \subset \mathbf{R}^n$ is said to be convex if $x^1, x^2 \in \mathbf{K}$ implies that $\alpha x^1 + (1-\alpha)x^2 \in \mathbf{K}$ for all $0 \leqq \alpha \leqq 1$. Thus, a subset of $\mathbf{R}^n$ is convex iff given any pair of points in it, the entire line segment connecting these two points is in the set. See Figures 1, 2.

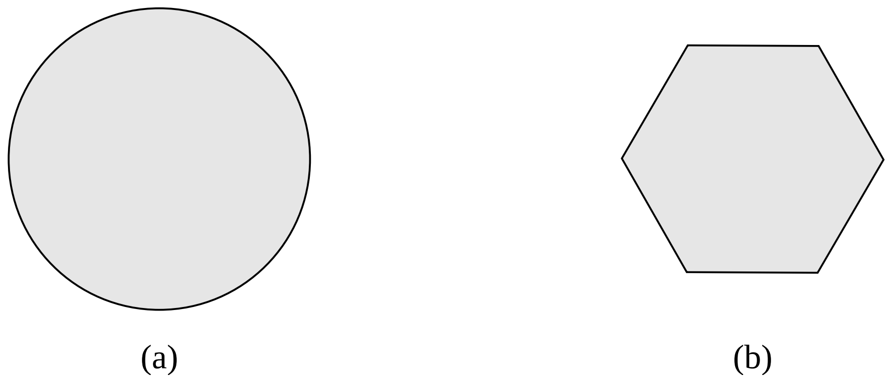

Figure 1 Convex sets. (a) All points inside or on the circle. (b) All points inside or on the polygon.

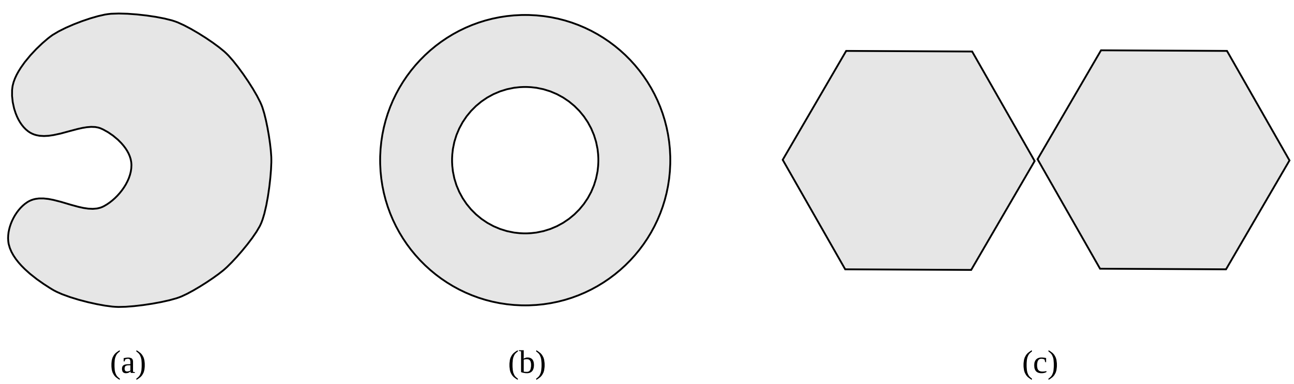

Figure 2 Non-convex sets. (a) All points inside or on the cashew nut. (b) All points on or between two circles. (c) All points on at least one of the two polygons.

2.1 Convex Combinations, Convex Hull

Let $(x^1, \ldots, x^r\}$ be any finite set of points in $\mathbf{R}^n$. A convex combination of this set is a point of the form

$$\alpha_1 x^1 + \ldots + \alpha_r x^r, \text{ where } \alpha_1 + \ldots + \alpha_r = 1 \text{ and } \alpha_1, \ldots, \alpha_r \geqq 0.$$

The set of all convex combinations of $\{x^1, \ldots, x^r\}$ is known as the convex hull of $\{x^1, \ldots, x^r\}$.

Given $\boldsymbol{\Delta} \subset \mathbf{R}^n$, the convex hull of $\boldsymbol{\Delta}$ is the set consisting of all convex combinations of all finite sets of points from $\boldsymbol{\Delta}$.

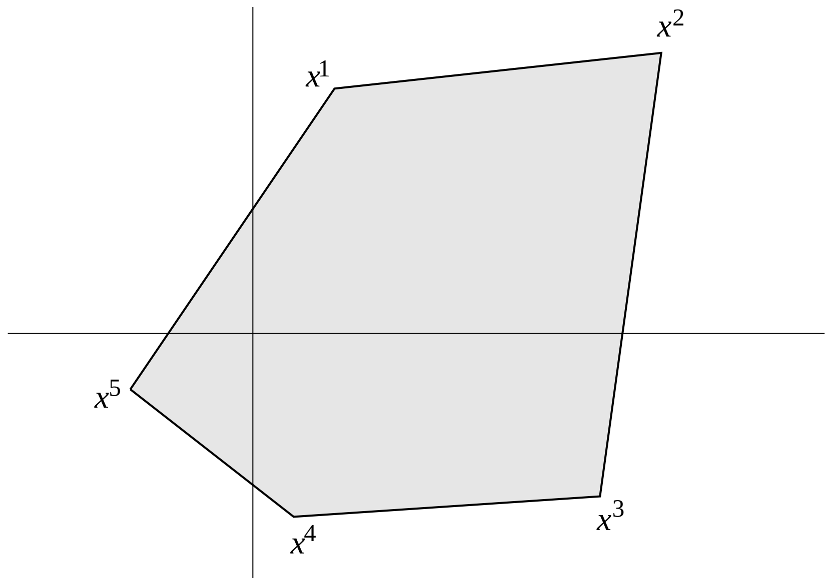

Figure 3 Convex hull of $\{x^1, \dots, x^5\}$ in $\mathbf{R}^2$.

The following results can be verified to be true:

1. $\mathbf{K} \subset \mathbf{R}^n$ is convex iff for any finite number r, given $x^1, \dots, x^r \in \mathbf{K}$, $\alpha_1 x^1 + \dots + \alpha_r x^r \in \mathbf{K}$ for all $\alpha_1, \dots, \alpha_r$ satisfying $\alpha_1 + \dots + \alpha_r = 1$, $\alpha_1 \geqq 0, \dots, \alpha_r \geqq 0$.
2. The intersection of any family of convex subsets of $\mathbf{R}^n$ is convex. The union of two convex sets may not be convex.
3. The set of feasible solutions of a system of linear constraints

$$\begin{aligned} A_{i.}x &= b_i, \ i = 1 \text{ to } m \\ &\geqq b_i, \ i = m+1 \text{ to } m+p \end{aligned}$$

is convex. A convex set like this is known as a **convex polyhedron**. A bounded convex polyhedron is called a **convex polytope**.

4. The set of feasible solutions of a homogeneous system of linear inequalities in $x \in \mathbf{R}^n$,

$$Ax \geqq 0 \tag{38}$$

is known as a **convex polyhedral cone**. Given a convex polyhedral cone, there exists a finite number of points $x^1, \dots, x^s$ such that the cone is $\{x : x = \alpha_1 x^1 + \dots + \alpha_s x^s, \alpha_1 \geqq 0, \dots, \alpha_s \geqq 0\} = \text{Pos}\{x^1, \dots, x^s\}$. The polyhedral cone which is the set of feasible solutions of (38) is said to be a **simplicial cone** if A is a nonsingular square matrix. Every simplicial cone of dimension n is of the form $\text{Pos}\{B_{.1}, \dots, B_{.n}\}$ where $\{B_{.1}, \dots, B_{.n}\}$ is a basis for $\mathbf{R}^n$.

5. Given two convex subsets of $\mathbf{R}^n$, $\mathbf{K}_1$, $\mathbf{K}_2$, their sum, denoted by $\mathbf{K}_1 + \mathbf{K}_2 = \{x + y : x \in \mathbf{K}_1, y \in \mathbf{K}_2\}$ is also convex.

Separating Hyperplane Theorems

Given two nonempty subsets $\mathbf{K}_1$, $\mathbf{K}_2$ of $\mathbf{R}^n$, the hyperplane $\mathbf{H} = \{x : cx = \alpha\}$ is said to separate $\mathbf{K}_1$ and $\mathbf{K}_2$ if $cx - \alpha$ has the same sign for all $x \in \mathbf{K}_1$, say $\geqq 0$, and the opposite sign for all $x \in \mathbf{K}_2$, that is, if

$$\begin{aligned} cx &\geqq \alpha \text{ for all } x \in \mathbf{K}_1 \\ &\leqq \alpha \text{ for all } x \in \mathbf{K}_2. \end{aligned}$$

Here we will prove that if two convex subsets of $\mathbf{R}^n$ are disjoint, there exists a separating hyperplane for them. See Figures 4, 5.

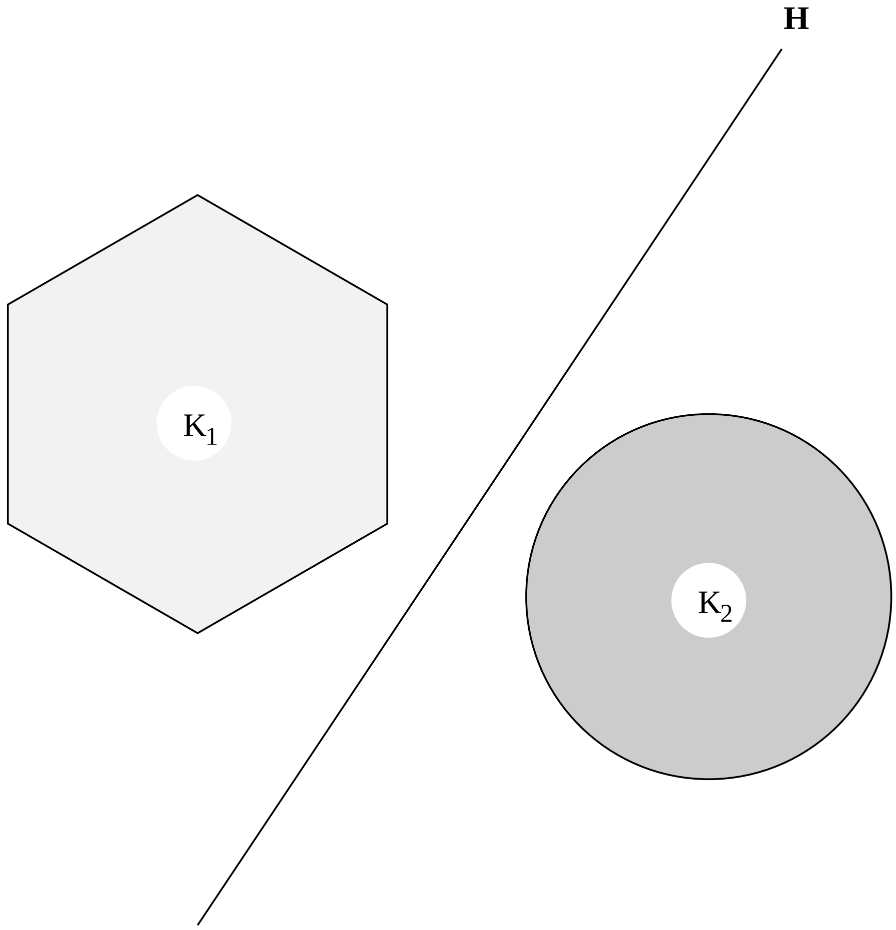

Figure 4 The hyperplane $\mathbf{H}$ separates the two disjoint convex sets $\mathbf{K}_1$ and $\mathbf{K}_2$.

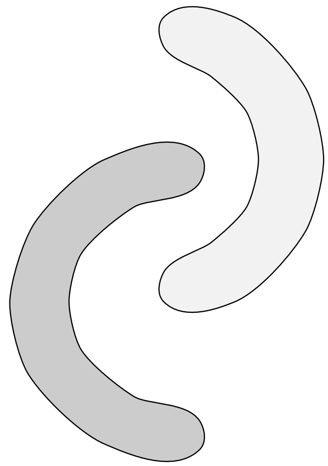

Figure 5 Even though the two cashew nuts (both nonconvex) are disjoint, they cannot be separated by a hyperplane.

Theorem 9. *Let* $\mathbf{K}$ *be a nonempty closed convex subset of* $\mathbf{R}^n$ *and* $0 \notin \mathbf{K}$. *Then there exists a hyperplane containing the origin separating it from* $\mathbf{K}$.

Proof. Take any point $\hat{x} \in \mathbf{K}$, and let $\mathbf{E} = \{x : \|x\| \leqq \|\hat{x}\|\}$. Since $0 \notin \mathbf{K}$, $\hat{x} \neq 0$, and hence $\mathbf{E}$ is a nonempty ball. Let $\mathbf{\Gamma} = \mathbf{E} \cap \mathbf{K}$. $\mathbf{\Gamma}$ is a bounded closed convex subset of $\mathbf{R}^n$, not containing the origin. The problem: minimize $\|x\|$ over $x \in \mathbf{\Gamma}$, has an optimum solution, since a continuous function attains its minimum on a compact set. We will show that this problem has a unique optimum solution. Suppose not. Let $x^1, x^2 \in \mathbf{\Gamma}$, $x^1 \neq x^2$, minimize $\|x\|$ over $x \in \mathbf{\Gamma}$. Let $x^3 = (x^1 + x^2)/2$. By Cauchy-Schwartz inequality $|(x^1)^T x^2| \leqq \|x^1\| \cdot \|x^2\|$ with equality holding iff $x^1 = \lambda x^2$ for some real number λ. So $\|x^3\|^{\mathbf{2}} = (\|x^1\|^{\mathbf{2}} + \|x^2\|^{\mathbf{2}} + 2|(x^1)^T x^2|)/4 \leqq (\|x^1\|^{\mathbf{2}} + \|x^2\|^{\mathbf{2}} + 2\|x^1\| \cdot \|x^2\|)/4$ by Cauchy-Schwartz inequality. Let $\|x^1\| = \gamma$. So $\|x^2\| = \gamma$ also, since both x^1, x^2 minimize $\|x\|$ over $x \in \mathbf{\Gamma}$. So, from the above, we have $\|x^3\|^{\mathbf{2}} \leqq \gamma^{\mathbf{2}}$. Since $x^3 \in \mathbf{\Gamma}$ and $\gamma^{\mathbf{2}}$ is the minimum of $\|x\|^{\mathbf{2}}$ over $x \in \mathbf{\Gamma}$, $\|x^3\|^{\mathbf{2}} \leqq \gamma^{\mathbf{2}}$ implies that $\|x^3\|^{\mathbf{2}} = \gamma^{\mathbf{2}}$. By Cauchy-Schwartz inequality, this equality holds iff $x^1 = \lambda x^2$ for some scalar λ. But since $\|x^1\| = \|x^2\|$, we must have $\lambda = +1$ or -1. If $\lambda = -1$, $x^3 = 0$, and this contradicts the fact that $0 \notin \mathbf{\Gamma}$. So $\lambda = +1$, that is, $x^1 = x^2$. So the problem of minimizing $\|x\|$ over $x \in \mathbf{\Gamma}$, has a unique optimum solution, say $\overline{x}$. We will now prove that

$$(x - \overline{x})^T \overline{x} \geqq 0 \text{ for all } x \in \mathbf{K}. \tag{39}$$

$\overline{x}$ minimizes $\|x\|$ over $x \in \mathbf{\Gamma}$, and from the definition of $\mathbf{\Gamma}$, it is clear that $\overline{x}$ also minimizes $\|x\|$ over $x \in \mathbf{K}$. Let $x \in \mathbf{K}$. By convexity of $\mathbf{K}$, $\overline{x} + \alpha(x - \overline{x}) \in \mathbf{K}$ for all $0 \leqq \alpha \leqq 1$. So $\|\overline{x} + \alpha(x - \overline{x})\|^{\mathbf{2}} \geqq \|\overline{x}\|^{\mathbf{2}}$ for all $0 \leqq \alpha \leqq 1$. That is, $\alpha^{\mathbf{2}}\|x - \overline{x}\|^{\mathbf{2}} + 2\alpha(x - \overline{x})^T \overline{x} \geqq 0$ for all $0 \leqq \alpha \leqq 1$. So for $0 < \alpha \leqq 1$, we have $\alpha\|x - \overline{x}\|^{\mathbf{2}} + 2(x - \overline{x})^T \overline{x} \geqq 0$. Making α approach zero through positive values, this implies (39).

Conversely, if $\overline{x} \in \mathbf{K}$ satisfies (39), then for any $x \in \mathbf{K}$, $\|x\|^{\mathbf{2}} = \|(x-\overline{x})+\overline{x}\|^{\mathbf{2}} = \|x-\overline{x}\|^{\mathbf{2}} + \|\overline{x}\|^{\mathbf{2}} + 2(x-\overline{x})^T\overline{x} \geqq \|\overline{x}\|^{\mathbf{2}}$ (by (39)), and this implies that $\overline{x}$ minimizes $\|x\|$ over $x \in \mathbf{K}$. Thus (39) is a necessary and sufficient optimality condition for the problem of minimizing $\|x\|$ over $x \in \mathbf{K}$.

Since $0 \notin \mathbf{K}$, $\overline{x} \neq 0$. From (39) we have $(\overline{x})^T x \geqq \|\overline{x}\|^{\mathbf{2}} > 0$ for all $x \in \mathbf{K}$. So the hyperplane $\{x : (\overline{x})^T x = 0\}$ through the origin separates $\mathbf{K}$ from 0.

□

Theorem 10. *Let* $\mathbf{K}$ *be a nonempty convex subset of* $\mathbf{R}^n$, $b \notin \mathbf{K}$. *Then* $\mathbf{K}$ *can be separated from* b *by a hyperplane.*

Proof. If $\mathbf{K}$ is a closed convex subset, by translating the origin to b and using Theorem 9 we conclude that $\mathbf{K}$ and b can be separated by a hyperplane.

If $\mathbf{K}$ is not closed, let $\overline{\mathbf{K}}$ be the closure of $\mathbf{K}$. If $b \notin \overline{\mathbf{K}}$, then again by the previous result b and $\overline{\mathbf{K}}$ can be separated by a hyperplane, which also separates b and $\mathbf{K}$.

So assume that $b \in \overline{\mathbf{K}}$. Since $b \in \overline{\mathbf{K}}$ but $\notin \mathbf{K}$, b must be a boundary point of $\mathbf{K}$. So every open neighborhood of b contains a point not in $\overline{\mathbf{K}}$. So we can get a sequence of points $\{b^r : r = 1 \text{ to } \infty\}$ such that $b^r \notin \overline{\mathbf{K}}$ for all r, and b^r converges to b as r tends to ∞. Since $b^r \notin \overline{\mathbf{K}}$, by the previous result, there exists c^r such that $c^r(x - b^r) \geqq 0$ for all $x \in \mathbf{K}$, with $\|c^r\| = 1$. The sequence of row vectors $\{c^r : r = 1, \ldots\}$ all lying on the unit sphere in $\mathbf{R}^n$ (which is a closed bounded set) must have a limit point. Let c be a limit point of $\{c^r : r = 1, 2, \ldots\}$. So $\|c\| = 1$. Let $\mathbf{S}$ be a monotonic increasing sequence of positive integers such that c^r converges to c as r tends to ∞ through $r \in \mathbf{S}$. But $c^r(x - b^r) \geqq 0$ for all $x \in \mathbf{K}$. Taking the limit in this inequality, as r tends to ∞ through $r \in \mathbf{S}$ we conclude that $c(x - b) \geqq 0$ for all $x \in \mathbf{K}$. So the hyperplane $\{x : cx = cb\}$ separates $\mathbf{K}$ from b.

□

Corollary 2. *Let* $\mathbf{K}$ *be a convex subset of* $\mathbf{R}^n$, *and let* b *be a boundary point of* $\mathbf{K}$. *Then there exists a row vector* $c \neq 0$, $c \in \mathbf{R}^n$ *such that* $cx \geqq cb$ *for all* $x \in \mathbf{K}$.

Proof. Follows from the arguments in the proof of Theorem 10.

□

The hyperplane $\{x : cx = cb\}$ in Corollary 2 is known as a **supporting hyperplane** for the convex set $\mathbf{K}$ at its boundary point b.

Theorem 11. *If* $\mathbf{K}_1$, $\mathbf{K}_2$ *are two mutually disjoint convex subsets of* $\mathbf{R}^n$, *there exists a hyperplane separating* $\mathbf{K}_1$ *from* $\mathbf{K}_2$.

Proof. Let $\mathbf{\Gamma} = \mathbf{K}_1 - \mathbf{K}_2 = \{x - y : x \in \mathbf{K}_1, y \in \mathbf{K}_2\}$. Since $\mathbf{K}_1$, $\mathbf{K}_2$ are convex, $\mathbf{\Gamma}$ is a convex subset of $\mathbf{R}^n$. Since $\mathbf{K}_1 \cap \mathbf{K}_2 = \emptyset$, $0 \notin \mathbf{\Gamma}$. So by Theorem 10, there exists a row vector $c \neq 0$, $c \in \mathbf{R}^n$, satisfying

$$cz \geqq 0 \text{ for all } z \in \mathbf{\Gamma}. \tag{40}$$

Let $\alpha =$ Infimum $\{cx : x \in \mathbf{K}_1\}$, $\beta =$ Supremum $\{cx : x \in \mathbf{K}_2\}$. By (40), we must have $\alpha \geqq \beta$. So if $\gamma = (\alpha + \beta)/2$, we have

$$\begin{aligned} cx &\geqq \gamma \text{ for all } x \in \mathbf{K}_1 \\ &\leqq \gamma \text{ for all } x \in \mathbf{K}_2. \end{aligned}$$

So $x : \{cx = \gamma\}$ is a hyperplane that separates $\mathbf{K}_1$ from $\mathbf{K}_2$.

$\square$

The theorems of alternatives discussed in Appendix 1, can be interpreted as separating hyperlane theorems about separating a point from a convex polyhedral cone not containing the point.

Exercises

---

6. Let $\mathbf{K}$ be a closed convex subset of $\mathbf{R}^n$ and $x \in \mathbf{R}^n$ and let y be the nearest point (in terms of the usual Euclidean distance) in $\mathbf{K}$ to x. Prove that $(x-y)^T(y-z) \geqq 0$ for all $z \in \mathbf{K}$. Also prove that $\|y - z\| \leqq \|x - z\|$ for all $z \in \mathbf{K}$.

7. Given sets $\mathbf{\Gamma}$, $\mathbf{\Delta}$ define $\alpha\mathbf{\Gamma} = \{\alpha x : x \in \mathbf{\Gamma}\}$ and $\mathbf{\Gamma} + \mathbf{\Delta} = \{x + y : x \in \mathbf{\Gamma}, y \in \mathbf{\Delta}\}$. Is $\mathbf{\Gamma} + \mathbf{\Gamma} = 2\mathbf{\Gamma}$? Also, when $\mathbf{\Gamma} = \{(x_1, x_2)^T : (x_1 - 1)^{\mathbf{2}} + (x_2 - 1)^{\mathbf{2}} \leqq 1\}$, $\mathbf{\Delta} = \{(x_1, x_2)^T : (x_1 + 4)^{\mathbf{2}} + (x_2 + 4)^{\mathbf{2}} \leqq 4\}$, find $\mathbf{\Gamma} + \mathbf{\Delta}$, $2\mathbf{\Gamma}$, $\mathbf{\Gamma} + \mathbf{\Gamma}$ and draw a figure in $\mathbf{R}^2$ illustrating each of these sets.

8. Prove that a convex cone in $\mathbf{R}^n$ is either equal to $\mathbf{R}^n$ or is contained in a half-space generated by a hyperplane through the origin.

9. Let $\mathbf{\Delta}_1 = \{x^1, \ldots, x^r\} \subset \mathbf{R}^n$. If $y^1, y^2 \in \mathbf{R}^n$, $y^1 \neq y^2$ are such that

$$\begin{aligned} y^1 &\in \text{ convex hull of } \{y^2\} \cup \mathbf{\Delta}_1 \\ y^2 &\in \text{ convex hull of } \{y^1\} \cup \mathbf{\Delta}_1 \end{aligned}$$

prove that both y^1 and y^2 must be in the convex hull of $\mathbf{\Delta}_1$. Using this and an induction argument, prove that if $\{y^1, \ldots, y^m\}$ is a set of distinct points in $\mathbf{R}^n$ and for each $j = 1$ to m

$$y^j \in \text{ convex hull of } \mathbf{\Delta}_1 \cup \{y^1, \ldots, y^{j-1}, y^{j+1}, \ldots, y^m\}$$

then each $y^j \in$ convex hull of $\mathbf{\Delta}_1$.

On Computing a Separating Hyperplane

Given a nonempty convex subset $\mathbf{K} \subset \mathbf{R}^n$, and a point $b \in \mathbf{R}^n$, $b \notin \mathbf{K}$, Theorem 10 guarantees that there exists a hyperplane $\mathbf{H} = \{x : cx = \alpha, c \neq 0\}$ which separates b from $\mathbf{K}$. It is a fundamental result, in mathematics such results are called **existence theorems**. This result can be proved in many different ways, and most books on convexity or optimization would have a proof for it. However, no other book seems to discuss how such a separating hyperplane can be computed, given b and $\mathbf{K}$ in some form (this essentially boils down to determining the vector c in the definition of the separating hyperplane $\mathbf{H}$), or how difficult the problem of computing it may be. For this reason, the following is very important. In preparing this, I benefitted a lot from discussions with R. Chandrasekaran.

However elegant the proof may be, an existence theorem cannot be put to practical use unless an efficient algorithm is known for computing **the thing** whose existence the theorem establishes. In order to use Theorem 10 in practical applications, we should be able to compute the separating hyperplane $\mathbf{H}$ given b and $\mathbf{K}$. Procedures to be used for constructing an algorithm to compute $\mathbf{H}$ depend very critically on the form in which the set $\mathbf{K}$ is made available to us. In practice, $\mathbf{K}$ may be specified either as the set of feasible solutions of a given system of constraints, or as the set of points satisfying a well specified set of properties, or as the convex hull of a set of points satisfying certain specified properties or constraints or those that can be obtained by a well defined constructive procedure. The difficulty of computing a separating hyperplane depends on the form in which $\mathbf{K}$ is specified.

K *Represented by a System of Linear Inequalities*

Consider the case, $\mathbf{K} = \{x : A_{i.}x \geqq d_i, i = 1 \text{ to } m\}$, where $A_{i.}$, d_i are given for all $i = 1$ to m. If $b \notin \mathbf{K}$, there must exist an i between 1 to m satisfying $A_{i.}b < d_i$. Find such an i, suppose it is r. Then the hyperplane $\{x : A_{r.}x = d_r\}$ separates $\mathbf{K}$ from b in this case.

K *Represented by a System of Linear Equations and Inequalities*

Consider the case, $\mathbf{K} = \{x : A_{i.}x = d_i, i = 1 \text{ to } m, \text{ and } A_{i.}x \geqq d_i,\ i = m+1 \text{ to } m+p\}$ where $A_{i.}$, d_i are given for all $i = 1$ to $m + p$. Suppose $b \notin \mathbf{K}$. If one of the inequality constraints $A_{i.}x \geqq d_i$, $i = m + 1$ to $m + p$, is violated by b, a hyperplane separating $\mathbf{K}$ from b, can be obtained from it as discussed above. If b satisfies all the inequality constraints in the definition of $\mathbf{K}$, it must violate one of the equality constraints. In this case, find an i, $1 \leqq i \leqq m$, satisfying $A_{i.}b \neq d_i$, suppose it is r, then the hyperplane $\{x : A_{r.}x = d_r\}$ separates $\mathbf{K}$ from b.

$\mathbf{K}$ *Represented as a Nonnegative Hull of a Specified Set of Points*

Consider the case, $\mathbf{K} =$ nonnegative hull of $\{A_{.j} : j = 1 \text{ to } t\} \subset \mathbf{R}^n$, t finite. Let A be the $n \times t$ matrix consisting of column vectors $A_{.j}$, $j = 1$ to t. Then $\mathbf{K} = \text{Pos}(A)$, a convex polyhedral cone, expressed as the nonnegative hull of a given finite set of points from $\mathbf{R}^n$. In this special case, the separating hyperplane theorem becomes exactly Farkas' theorem (Theorem 3). See Section 4.6.7 of [2.26] or [1.28]. Since $b \notin \text{Pos}(A)$, system (I) of Farkas' theorem, Theorem 3, has no feasible solution, and hence system (II) has a solution π. Then the hyperplane $\{x : \pi x = 0\}$ separates b from $\text{Pos}(A)$. The solution π for system (II) can be computed efficiently using Phase I of the simplex method, as discussed in Note 1 of Appendix 1. Given any point $b \in \mathbf{R}^n$, this provides an efficient method to check whether $b \in \text{Pos}(A)$ (which happens when system (I) of Farkas' theorem, Theorem 3, with this data, has a feasible solution); and if not, to compute a hyperplane separating b from $\text{Pos}(A)$, as long as the number of points in the set $\{A_{.j} : j = 1 \text{ to } t\}$, t is not too large. If t is very large, the method discussed here for computing a separating hyperplane, may not be practically useful, this is discussed below using some actual examples.

$\mathbf{K}$ *Represented as the Convex Hull of a Specified Set of Points*

Consider the case where $\mathbf{K}$ is specified as the convex hull of a given set of points $\{A_{.j} : j = 1 \text{ to } t\} \subset \mathbf{R}^n$. So, in this case, $b \notin \mathbf{K}$, iff the system

$$\begin{aligned} \sum_{j=1}^{t} A_{.j} x_j &= b \\ \sum_{j=1}^{t} x_j &= 1 \\ x_j &\geqq 0, \ j = 1 \text{ to } t \end{aligned}$$

has no feasible solution $x = (x_j)$. This system is exactly in the same form as system (I) of Farkas' theorem, Theorem 3, and a separating hyperplane in this case can be computed using this theorem, as discussed above, as long as t is not too large.

$\mathbf{K}$ *Represented by a System of "$\leqq$" Inequalities Involving Convex Functions*

Now consider the case where $\mathbf{K}$ is represented as the set of feasible solutions of a system of inequalities

$$f_i(x) \leqq 0, \ i = 1 \text{ to } m$$

where each $f_i(x)$ is a differentiable convex function defined on $\mathbf{R}^n$. See the following section, Appendix 3, for definitions of convex functions and their properties. In this case, $b \notin \mathbf{K}$, iff there exists an i satisfying $f_i(b) > 0$. If $b \notin \mathbf{K}$, find such an i, say r. Then the hyperplane $H = \{x : f_r(b) + \nabla f_r(b)(x - b) = 0\}$ separates b from $\mathbf{K}$, by Theorem 15 of Appendix 3 (see Exercise 11 in Appendix 3).

K *Represented by a System of "$\leqq$" Inequalities Involving General Functions*

Now consider the case in which the convex set $\mathbf{K}$ is represented by a system of constraints

$$g_i(x) \leqq 0,\ i = 1 \text{ to } m$$

where the functions $g_i(x)$ are not all convex functions. It is possible for the set of feasible solutions of such a system to be convex set. As an example let $n = 2$, $x = (x_1, x_2)^T$, and consider the system

$$\begin{aligned} -x_1 - x_2 + 2 &\leqq 0 \\ x_1 \qquad - 1 &\leqq 0 \\ x_2 - 1 &\leqq 0 \\ -x_1^4 - x_2^4 + 2 &\leqq 0. \end{aligned}$$

This system has the unique solution $x = (1, 1)^T$, and yet, not all the functions in the system are convex functions. As another example, let M be a P-matrix of order n which is not a PSD matrix, and $q \in \mathbf{R}^n$. Consider the system in variables $z = (z_1, \ldots, z_n)^T$

$$\begin{aligned} -z &\leqq 0 \\ -q - Mz &\leqq 0 \\ z^T(q + Mz) &\leqq 0. \end{aligned}$$

This system has the unique solution $\bar{z}$ ($\bar{z}$ is the point which leads to the unique solution of the LCP (q, M)), so the set of feasible solutions of this system is convex, being a singleton set, and yet the constraint function $z^T(q + Mz)$ is not convex, since M is not PSD.

In general, when the functions $g_i(x)$, $i = 1$ to m are not all convex, even though the set $\mathbf{K} = \{x : g_i(x) \leqq 0,\ i = 1 \text{ to } m\}$ may be convex, and $b \notin \mathbf{K}$, there is no efficient method known for computing a hyperplane separating b from $\mathbf{K}$. See Exercise 40.

Now we consider some cases in which $\mathbf{K}$ is the convex hull of a set of points specified by some properties.

K *Is the Convex Hull of the Tours of a Traveling Salesman Problem*

Consider the famous traveling salesman problem in cities $1, 2, \ldots, n$. See [1.28]. In this problem, a salesman has to start in some city, say city 1, visit each of the other cities exactly once in some order, and in the end return to the starting city, city 1. If he travels to cities in the order i to $i+1$, $i = 1$ to $n-1$ and then from city n to city 1, this route can be represented by the order "$1, 2, \dots, n; 1$". Such an order is known as a **tour**. So, a tour is a circuit spanning all the cities, that leaves each city exactly once. From the starting city, city 1, he can go to any of the other $(n-1)$ cities. So there are $(n-1)$ different ways in which he can pick the city that he travels from the starting city, city 1. From that city he can travel to any of the remaining $(n-2)$ cities, etc. Thus the total number of possible tours in an n city traveling salesman problem is $(n-1)(n-2)\dots 1 = (n-1)!$ Given a tour, define a $0-1$ matrix $x = (x_{ij})$ by

$$x_{ij} = \begin{cases} 1 & \text{if the salesman goes from city } i \text{ to city } j \text{ in the tour} \\ 0 & \text{otherwise.} \end{cases}$$

Such a matrix $x = (x_{ij})$ is called the **tour assignment** corresponding to the tour. An **assignment** (of order n) is any $0-1$ square matrix $x = (x_{ij})$ of order n satisfying

$$\begin{aligned} \sum_{j=1}^{n} x_{ij} &= 1, \ i = 1 \text{ to } n \\ \sum_{i=1}^{n} x_{ij} &= 1, \ j = 1 \text{ to } n \\ x_{ij} &= 0 \text{ or } 1 \text{ for all } i, j. \end{aligned}$$

Every tour assignment is an assignment, however not all assignments may be tour assignments. For example, if $n = 5$

$$x^1 = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

is a tour assignment representing the tour 1, 4, 2, 5, 3; 1 covering all the cities 1 to 5. But the assignment

$$x^2 = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

is not a tour assignment, since it consists of two subtours 1, 2, 3; 1 and 4, 5; 4 each spanning only a proper subset of the original set of cities.

Let $\mathbf{K}_T$ be the convex hull of all the $(n-1)!$ tour assignments of order n. $\mathbf{K}_T$ is well defined, it is the convex hull of a finite set of points in $\mathbf{R}^{n \times n}$. However, if n is large (even $n \geqq 10$), the number of tour assignments, $(n-1)!$ is very large. $\mathbf{K}_T$ is of course a convex polytope. It can be represented as the set of feasible solutions of a system of linear constraints, but that system is known to contain a very large number of constraints. Deriving a linear constraint representation of $\mathbf{K}_T$ remains an unsolved problem. In this case, if $b = (b_{ij})$ is a given square matrix of order n satisfying the conditions

$$\begin{aligned} b_{ii} &= 0, \ i = 1 \text{ to } n \\ \sum_{j=1}^{n} b_{ij} &= 1, \ i = 1 \text{ to } n \\ \sum_{i=1}^{n} b_{ij} &= 1, \ i = 1 \text{ to } n \\ 0 \leqq b_{ij} &\leqq 1, \ \text{for all } i, j = 1 \text{ to } n \end{aligned}$$

even to check whether $b \in \mathbf{K}_T$ is a hard problem for which no efficient algorithm is known. Ideally, given such a b, we would like an algorithm which

either determines that $b \in \mathbf{K}_T$

or determines that $b \notin \mathbf{K}_T$ and produces in this case a hyperplane separating b from $\mathbf{K}_T$

and for which the computational effort in the worst case is bounded above by a polynomial in n. No such algorithm is known, and the problem of constructing such an algorithm, or even establishing whether such an algorithm exists, seems to be a very hard problem. If such an algorithm exists, using it we can construct efficient algorithms for solving the traveling salesman problem, which is the problem of finding a minimum cost tour assignment that minimizes $\sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij}$ for given cost matrix $c = (c_{ij})$.

K *Is the Convex Hull of Feasible Solutions of an Integer Linear System*

Let A, d be given integer matrices of orders $m \times n$ and $m \times 1$ respectively. Consider the following systems: $x = (x_j) \in \mathbf{R}^n$

$$\begin{aligned} Ax &= d \\ x &\geqq 0 \\ x &\text{ an integer vector} \end{aligned}$$

or the system

$$\begin{aligned} Ax &= d \\ x &\geqq 0 \\ 0 \leqq x_j &\leqq 1, \ j = 1 \text{ to } n \\ x_j &\text{ integer for all } j. \end{aligned}$$

Let $\mathbf{K}_I$ denote the convex hull of all feasible solutions of such a system. Again, $\mathbf{K}_I$ is a well defined set, it is the convex hull of integer feasible solutions to a specified system of linear constraints. Given a point $b \in \mathbf{R}^n$, ideally we would like an algorithm which either determines that $b \in \mathbf{K}_I$

or determines that $b \notin \mathbf{K}_I$ and produces in this case a hyperplane separating b from $\mathbf{K}_I$

and for which the computational effort in the worst case is bounded above by a polynomial in the size of (A, b). No such algorithm is known.

K *Is the Convex Hull of Extreme Points of an Unbounded Convex Polyhedron*

Let A, d be given integer matrices of orders $m \times n$ and $m \times 1$ respectively, with rank $(A) = m$. Let $\mathbf{\Gamma}$ be the set of feasible solutions of the system

$$\begin{aligned} Ax &= d \\ x &\geqq 0. \end{aligned}$$

Suppose it is known that $\mathbf{\Gamma}$ is an unbounded convex polyhedron. $\mathbf{\Gamma}$ has a finite set of extreme points, each of these is a BFS of the above system. Let $\mathbf{K}$ be the convex hull of all these extreme points $\mathbf{\Gamma}$. Here again $\mathbf{K}$ is a well defined convex polytope, but it is the convex hull of extreme points of $\mathbf{\Gamma}$, and the number of these extreme points may be very large. See Section 3.7 of [2.26]. In general, given a point $b \in \mathbf{\Gamma}$, the problem of determining whether $b \in \mathbf{K}$, and the problem of determining a separating hyperplane separating b and $\mathbf{K}$ when $b \notin \mathbf{K}$, are very hard problems for which no efficient algorithms are known (the special case when $n = m + 2$ or $m + 1$ are easy, because in this case the dimension of $\mathbf{\Gamma}$ is at most two).

Summary

This discussion clearly illustrates the fact that even though we have proved the existence of separating planes, at the moment algorithms for computing one of them efficiently are only known when $\mathbf{K}$ can be represented in very special forms.

3. CONVEX, CONCAVE FUNCTIONS, THEIR PROPERTIES

Let $\mathbf{\Gamma}$ be a convex subset of $\mathbf{R}^n$ and let $f(x)$ be a real valued function defined on $\mathbf{\Gamma}$. $f(x)$ is said to be a **convex function** iff for any $x^1, x^2 \in \mathbf{\Gamma}$, and $0 \leqq \alpha \leqq 1$, we have

$$f(\alpha x^1 + (1 - \alpha)x^2) \leqq \alpha f(x^1) + (1 - \alpha)f(x^2). \tag{41}$$

This inequality is called **Jensen's inequality** after the Danish mathematician who first discussed it. The important property of convex functions is that when you join two points on the surface of the function by a chord, the function itself lies underneath the chord on the interval joining these points, see Figure 6.

Similarly, if $g(x)$ is a real valued function defined on the convex set $\mathbf{\Gamma} \subset \mathbf{R}^n$, it is said to be a **concave function** iff for any $x^1, x^2 \in \mathbf{\Gamma}$ and $0 \leqq \alpha \leqq 1$, we have

$$g(\alpha x^1 + (1-\alpha)x^2) \geqq \alpha g(x^1) + (1-\alpha)g(x^2). \tag{42}$$

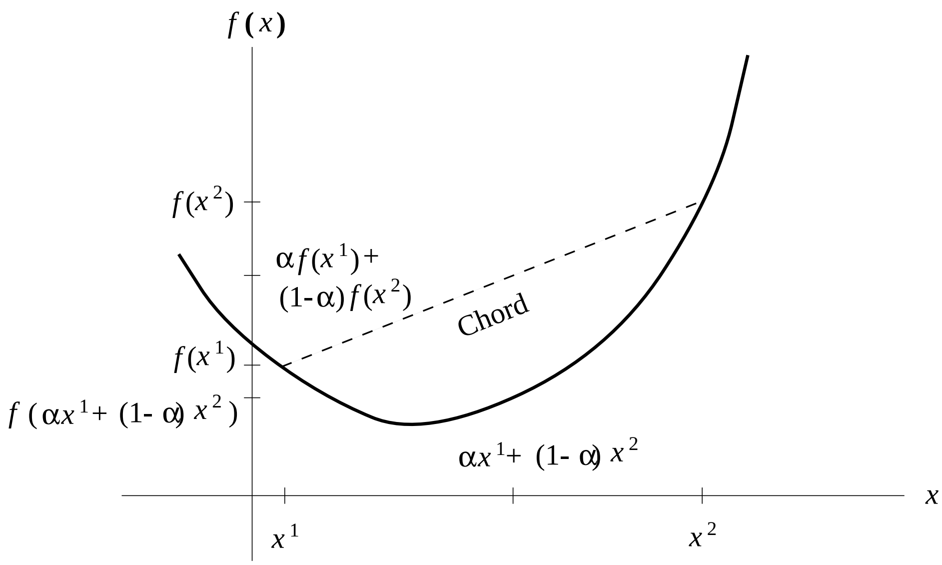

Figure 6 A convex function defined on the real line.

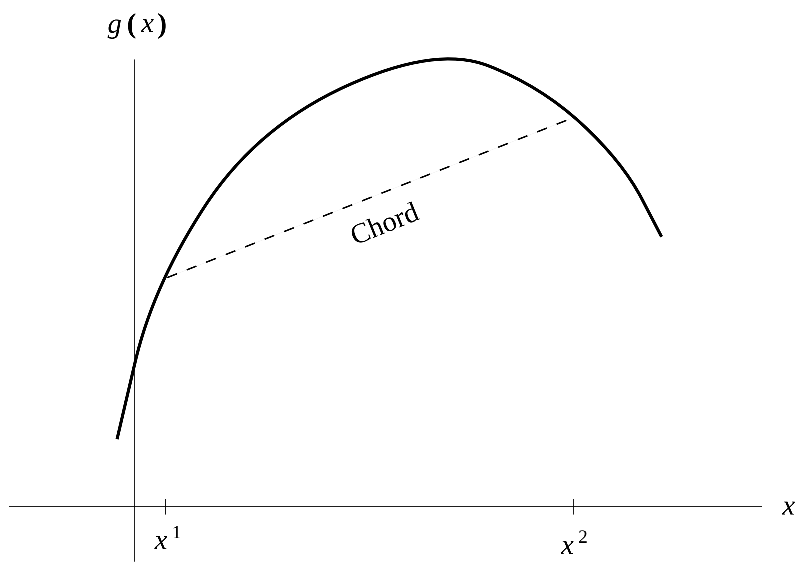

Figure 7 A concave function defined on the real line.

Clearly, a function is concave iff its negative is convex. Also, a concave function lies above the chord on any interval, see Figure 7. Convex and concave functions figure prominently in optimization. In mathematical programming literature, the problem of either minimizing a convex function, or maximizing a concave function, on a convex set, are known as convex programming problems. For a convex programming problem, a local optimum solution is a global optimum solution (see Theorem 12 below) and hence any techniques for finding a local optimum will lead to a global optimum on these problems.

The function $f(x)$ defined above is said to be **strictly convex**, if (41) holds as a strict inequality for $0 < \alpha < 1$ and for all $x^1, x^2 \in \mathbf{\Gamma}$. Likewise $g(x)$ is said to be a **strictly concave function** if (42) holds as a strict inequality for $0 < \alpha < 1$ and for all $x^1, x^2 \in \mathbf{\Gamma}$.

The following results can be verified to be true.

1. A nonnegative combination of convex functions is convex. Likewise a nonnegative combination of concave functions is concave.
2. If $f(x)$ is a convex function defined on the convex set $\mathbf{\Gamma} \subset \mathbf{R}^n$, $\{x : f(x) \leqq \alpha\}$ is a convex set for all real numbers α. Likewise, if $g(x)$ is a concave function defined on the convex set $\mathbf{\Gamma} \subset \mathbf{R}^n$, $\{x : g(x) \geqq \alpha\}$ is a convex set for all real numbers α.
3. If $f_1(x), \dots, f_r(x)$ are all convex functions defined on the convex set $\mathbf{\Gamma} \subset \mathbf{R}^n$, the pointwise supremum function $f(x) = \text{maximum } \{f_1(x), \dots, f_r(x)\}$ is convex.
4. If $g_1(x), \dots, g_r(x)$ are all concave functions defined on the convex set $\mathbf{\Gamma} \subset \mathbf{R}^n$, the pointwise infimum function $g(x) = \text{minimum } \{g_1(x), \dots, g_r(x)\}$ is concave.
5. A convex or concave function defined on an open convex subset of $\mathbf{R}^n$ is continuous (see [A10] for a proof of this).
6. Let $f(x)$ be a real valued function defined on a convex subset $\mathbf{\Gamma} \subset \mathbf{R}^n$. In $\mathbf{R}^{n+1}$, plot the objective value of $f(x)$ along the x_{n+1}-axis. The subset of $\mathbf{R}^{n+1}$, $\mathbf{F} = \{X = (x_1, \dots, x_n, x_{n+1}) : x = (x_1, \dots, x_n) \in \mathbf{\Gamma}, x_{n+1} \geqq f(x)\}$ is known as the **epigraph** of the function $f(x)$. It is the set of all points in $\mathbf{R}^{n+1}$ lying above (along the x_{n+1}-axis) the surface of $f(x)$. See Figure 8 for an illustration of the epigraph of a convex function defined on an interval of the real line $\mathbf{R}^1$. It can be shown that $f(x)$ is convex iff its epigraph is a convex set, from the definitions of convexity of a function and of a set. See Figures 8, 9.
7. Let $g(x)$ be a real valued function defined on a convex subset $\mathbf{\Gamma} \subset \mathbf{R}^n$. In $\mathbf{R}^{n+1}$, plot the objective value of $f(x)$ along the x_{n+1}-axis. The subset of $\mathbf{R}^{n+1}$, $\mathbf{G} = \{X = (x_1, \dots, x_n, x_{n+1}) : x = (x_1, \dots, x_n) \in \mathbf{\Gamma}, x_{n+1} \leqq g(x)\}$ is known as the **hypograph** of the function $g(x)$. It is the set of all points in $\mathbf{R}^{n+1}$ lying below (along the x_{n+1}-axis) the surface of $g(x)$. See Figure 10. It can be shown from the definitions, that $g(x)$ is concave, iff its hypograph is a convex set.

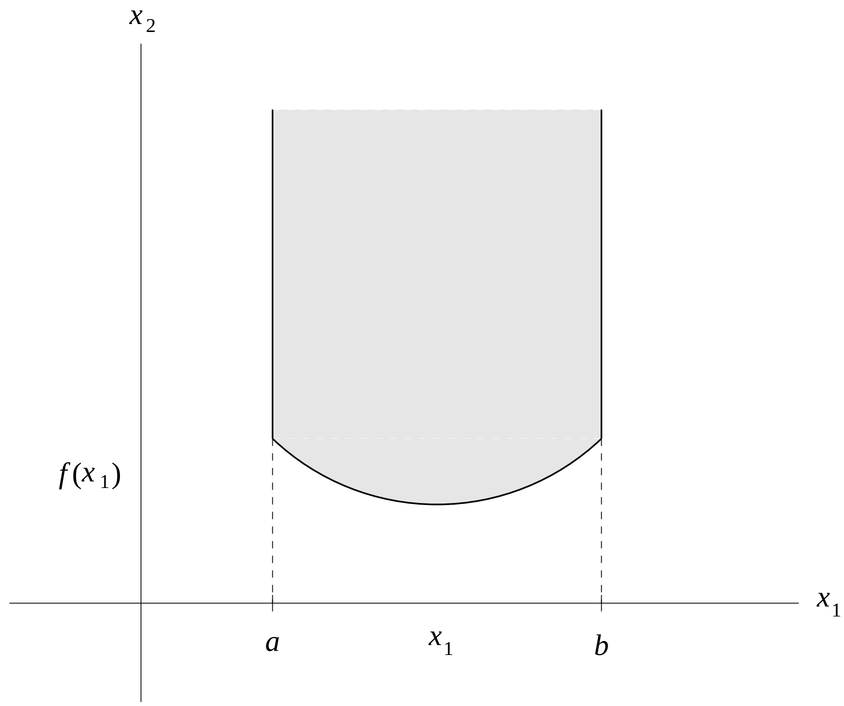

Figure 8 The epigraph of a convex function defined on the interval $a \leqq x_1 \leqq b$ is a convex subset of $\mathbf{R}^2$.

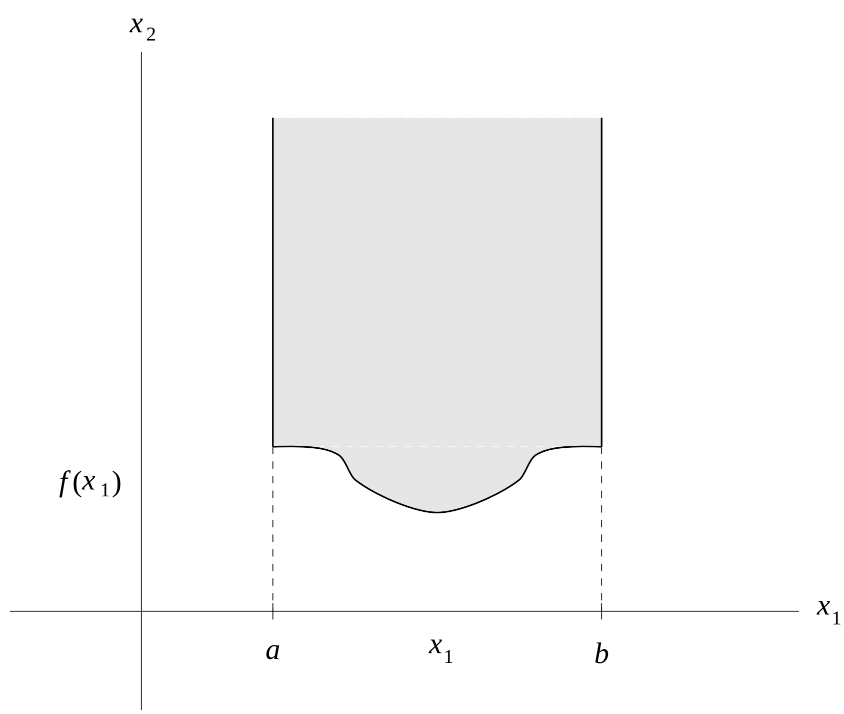

Figure 9 The epigraph of a nonconvex function $f(x_1)$ defined on the interval $a \leqq x_1 \leqq b$, is not a convex subset of $\mathbf{R}^2$.

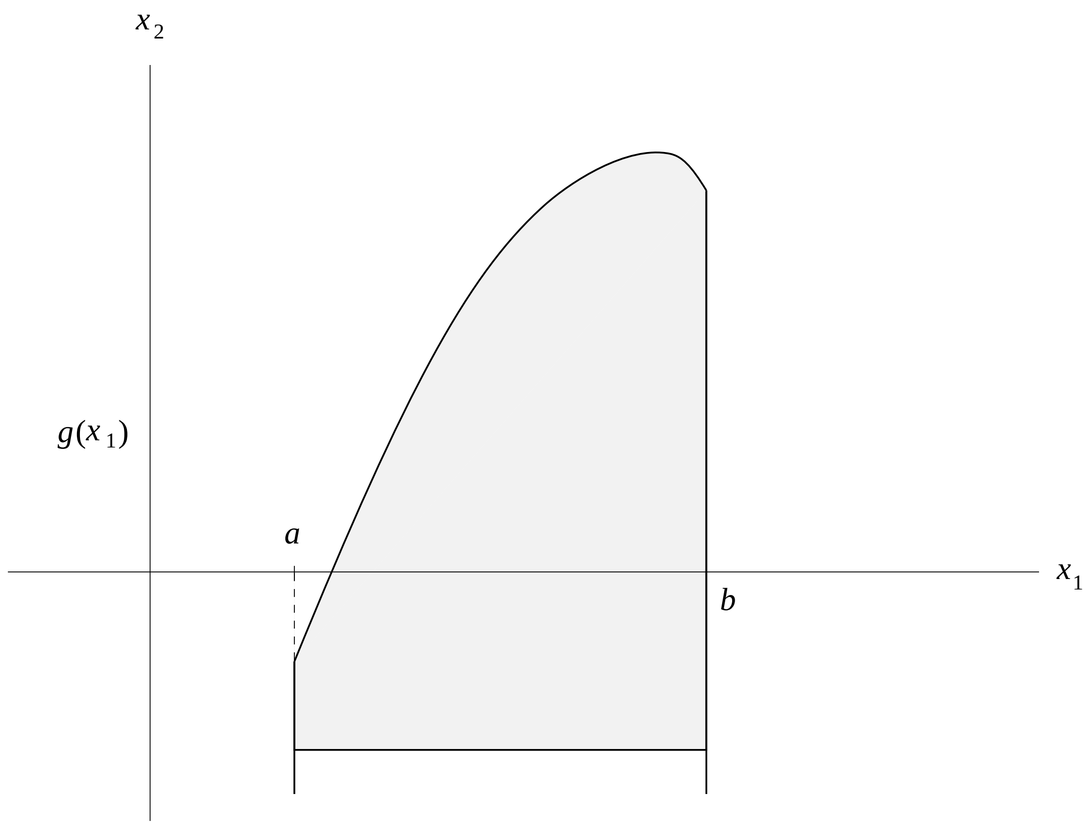

Figure 10 The hypograph of a concave function defined on the interval $a \leqq x_1 \leqq b$ is a convex subset of $\mathbf{R}^2$.

Theorem 12. *For the problem of minimizing a convex function $f(x)$ on a convex set $\mathbf{\Gamma} \subset \mathbf{R}^n$, every local minimum is a global minimum.*

Proof. Let x^1 be a local minimum for this problem. Suppose there exists an $x^2 \in \mathbf{\Gamma}$ such that $f(x^2) < f(x^1)$. Then, by convexity, for $0 < \alpha < 1$,

$$f(x^1 + \alpha(x^2 - x^1)) = f(\alpha x^2 + (1-\alpha)x^1) \leqq (1-\alpha)f(x^1) + \alpha f(x^2) < f(x^1). \quad (43)$$

So when α is positive but sufficiently small, the point $x^1 + \alpha(x^2 - x^1)$ contained in the neighborhood of x^1 satisfies (43), contradicting the local minimum property of x^1. So we cannot have an $x^2 \in \mathbf{\Gamma}$ satisfying $f(x^2) < f(x^1)$, that is, x^1 is in fact the global minimum for $f(x)$ in $\mathbf{\Gamma}$.

□

Theorem 13. *Let f be a real valued convex function defined on the convex set $\mathbf{\Gamma} \subset \mathbf{R}^n$. The set of optimum solutions for the problem of minimizing $f(x)$ over $x \in \mathbf{\Gamma}$ is a convex set.*

Proof. Let $\mathbf{L}$ denote the set of optimum solutions for the problem: minimize $f(x)$ over $x \in \mathbf{\Gamma}$. Let $x^1, x^2 \in \mathbf{L}$. So $f(x^1) = f(x^2) = \lambda$ = minimum value of $f(x)$ over $x \in \mathbf{\Gamma}$. Let $0 \leqq \alpha \leqq 1$. By convexity of $f(x)$, $f(\alpha x^1 + (1-\alpha)x^2) \leqq \alpha f(x^1) + (1-\alpha)f(x^2) = \lambda$

and $\alpha x^1 + (1-\alpha)x^2 \in \mathbf{\Gamma}$, since $\mathbf{\Gamma}$ is convex. Since λ is the minimum value of $f(x)$ over $x \in \mathbf{\Gamma}$, the above inequality must hold as an equation, that is, $f(\alpha x^1 + (1-\alpha)x^2) = \lambda$, which implies that $\alpha x^1 + (1-\alpha)x^2 \in \mathbf{L}$ also. So $\mathbf{L}$ is a convex set.

□

Theorem 14. *For the problem of maximizing a concave function $g(x)$ on a convex set $\mathbf{\Gamma} \subset \mathbf{R}^n$, every local maximum is a global maximum.*

Proof. Similar to Theorem 12.

□

A real valued function $\theta(x)$ defined on an open set $\mathbf{\Gamma} \subset \mathbf{R}^n$ is said to be differentiable at a point $\overline{x} \in \mathbf{\Gamma}$ if the partial derivative vector $\nabla\theta(\overline{x})$ exists, and for each $y \in \mathbf{R}^n$, limit $((\theta(\overline{x} + \lambda y) - \theta(\overline{x}) - \lambda(\nabla\theta(\overline{x}))y)/\lambda)$ as λ tends to zero is zero. $\theta(x)$ is said to be twice differentiable at $\overline{x}$ if the Hessian matrix $H(\theta(\overline{x}))$ exists and for each $y \in \mathbf{R}^n$, limit $[(\theta(\overline{x} + \lambda y) - \theta(\overline{x}) - \lambda(\nabla\theta(\overline{x}))y - (\lambda^{\mathbf{2}}/2)y^T H(\theta(\overline{x}))y]/\lambda^{\mathbf{2}})$ as λ tends to zero is zero.

The real valued function $\theta(x)$ defined on an open set $\mathbf{\Gamma} \subset \mathbf{R}^n$ is said to be continuously differentiable at a point $\overline{x} \in \mathbf{\Gamma}$ if it is differentiable at $\overline{x}$ and the partial derivatives $\frac{\partial\theta(x)}{\partial x_j}$ are all continuous at $\overline{x}$. The function $\theta(x)$ is said to continuously differentiable at a point $\overline{x} \in \mathbf{\Gamma}$ if it is twice differentiable at $\overline{x}$ and the second order partial derivatives $\frac{\partial^2\theta(x)}{\partial x_i \partial x_j}$ are all continuous at $\overline{x}$. The function is said to be differentiable, continuously differentiable, etc., over the set $\mathbf{\Gamma}$, if it satisfies the corresponding property for each point in $\mathbf{\Gamma}$.

Theorem 15 (Gradient Support Inequality): *Let $f(x)$ be a real valued convex function defined on an open convex set $\mathbf{\Gamma} \subset \mathbf{R}^n$. If $f(x)$ is differentiable at $\overline{x} \in \mathbf{\Gamma}$,*

$$f(x) - f(\overline{x}) \geqq (\nabla f(\overline{x}))(x - \overline{x}) \text{ for all } x \in \mathbf{\Gamma}. \tag{44}$$

Conversely, if $f(x)$ is a real valued differentiable function defined on $\mathbf{\Gamma}$ and (44) holds for all $x, \overline{x} \in \mathbf{\Gamma}$, $f(x)$ is convex.

Proof. Suppose $f(x)$ is convex. Let $x \in \mathbf{\Gamma}$. By convexity of $\mathbf{\Gamma}$, $\alpha x + (1-\alpha)\overline{x} = \overline{x} + \alpha(x - \overline{x}) \in \mathbf{\Gamma}$ for all $0 \leqq \alpha \leqq 1$. Since $f(x)$ is convex we have $f(\overline{x} + \alpha(x-\overline{x})) \leqq \alpha f(x) + (1-\alpha)f(\overline{x})$. So for $0 < \alpha \leqq 1$, we have

$$f(x) - f(\overline{x}) \geqq (f(\overline{x} + \alpha(x - \overline{x})) - f(\overline{x}))/\alpha. \tag{45}$$

By definition of differentiability, the right hand side of (45) tends to $\nabla f(\overline{x})(x - \overline{x})$ as α tends to zero through positive values. Since (45) holds for all $0 < \alpha \leqq 1$, this implies (44) as α tends to zero through positive values in (45).

Conversely, suppose $f(x)$ is a real valued differentiable function defined on $\mathbf{\Gamma}$ and suppose (44) holds for all, $x, \overline{x} \in \mathbf{\Gamma}$. Given $x^1, x^2 \in \mathbf{\Gamma}$, from (44) we have, for $0 < \alpha < 1$,

$$\begin{aligned} f(x^1) - f((1-\alpha)x^1 + \alpha x^2) &\geqq \alpha(\nabla f(1-\alpha)x^1 + \alpha x^2)(x^1 - x^2) \\ f(x^2) - f((1-\alpha)x^1 + \alpha x^2) &\geqq -(1-\alpha)(\nabla f((1-\alpha)x^1 + \alpha x^2))(x^1 - x^2). \end{aligned}$$

Multiply the first inequality by $(1-\alpha)$ and the second by α and add. This leads to

$$(1-\alpha)f(x^1) + \alpha f(x^2) - f((1-\alpha)x^1 + \alpha x^2) \geqq 0. \tag{46}$$

Since (46) holds for all $x^1, x^2 \in \mathbf{\Gamma}$ and $0 < \alpha < 1$, $f(x)$ is convex.

□

Theorem 16. *Let $g(x)$ be a concave function defined on an open convex set $\mathbf{\Gamma} \subset \mathbf{R}^n$. If $g(x)$ is differentiable at $\overline{x} \in \mathbf{\Gamma}$,*

$$g(x) \leqq g(\overline{x}) + (\nabla g(\overline{x}))(x - \overline{x}), \text{ for all } x \in \mathbf{\Gamma}. \tag{47}$$

Conversely, if $g(x)$ is a differentiable function defined on $\mathbf{\Gamma}$ and (47) holds for all $x, \overline{x} \in \mathbf{\Gamma}$, $g(x)$ is concave.

Proof. Similar to the proof of Theorem 15.

□

Figures 11, 12 provide illustrations of gradient support inequalities for convex and concave functions.

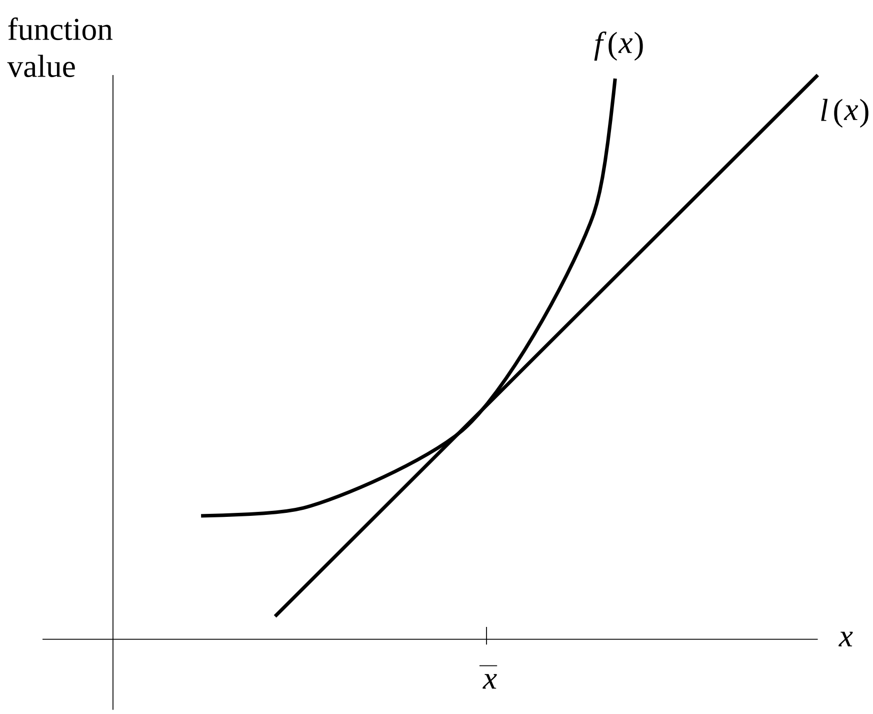

Figure 11 $f(x)$ is a differentiable convex function. $l(x) = f(\overline{x}) + (\nabla f(\overline{x}))(x - \overline{x})$, an affine function (since $\overline{x}$ is a given point), is the first order Taylor series approximation for $f(x)$ around $\overline{x}$. It underestimates $f(x)$ at each point.

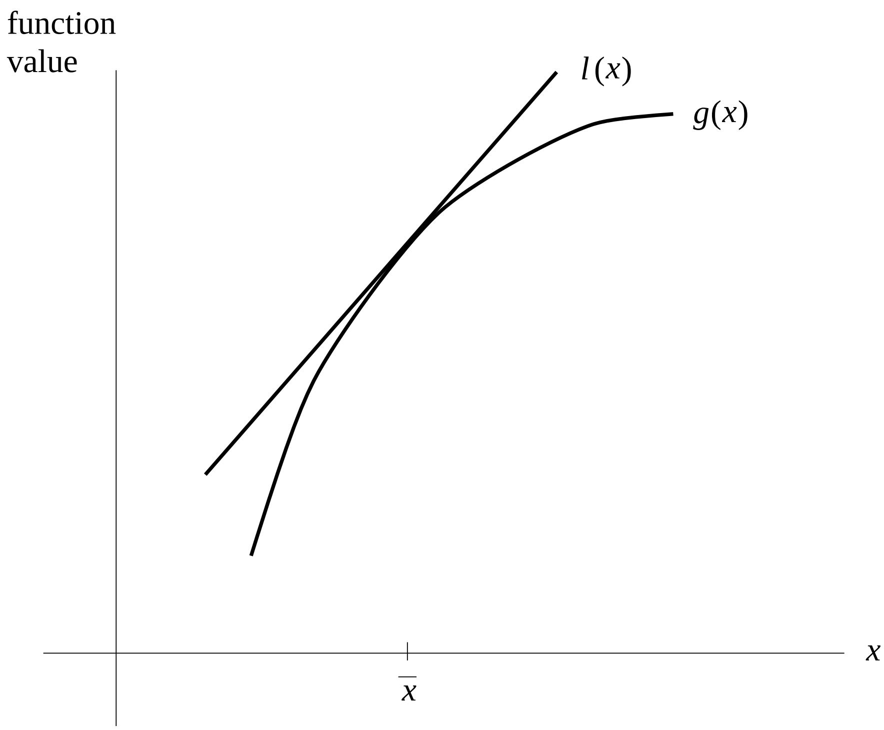

Figure 12 $g(x)$ is a differentiable concave function. $l(x) = g(\overline{x}) + (\nabla g(\overline{x})(x - \overline{x}))$ is the first order Taylor series approximation for $g(x)$ around $\overline{x}$. It overestimates $g(x)$ at each point.

Theorem 17. *Let $f(x)$ be a real valued convex function defined on an open convex subset $\mathbf{\Gamma} \subset \mathbf{R}^n$. If $f(x)$ is twice differentiable at $\overline{x} \in \mathbf{\Gamma}$, $H(f(\overline{x}))$ is* PSD. *Conversely, if $f(x)$ is a twice differentiable real valued function defined on $\mathbf{\Gamma}$ and $H(f(\overline{x}))$ is* PSD *for all $\overline{x} \in \mathbf{\Gamma}$, $f(x)$ is convex.*

Proof. Let $\overline{x} \in \mathbf{\Gamma}$ and $y \in \mathbf{R}^n$. Suppose $f(x)$ is convex. For $\alpha > 0$ and sufficiently small, by Theorem 15 we have

$$(f(\overline{x} + \alpha y) - f(\overline{x}) - \alpha(\nabla f(\overline{x}))y)/\alpha \geqq 0. \tag{48}$$

Taking the limit as α tends to zero through positive values, from (48) we have $y^T H(f(\overline{x}))y \geqq 0$, and since this holds for all $y \in \mathbf{R}^n$, $H(f(\overline{x}))$ is PSD.

Suppose $f(\overline{x})$ is twice differentiable on $\mathbf{\Gamma}$ and $H(f(\overline{x}))$ is PSD for all $\overline{x} \in \mathbf{\Gamma}$. By Taylor's theorem of calculus we have, for $x^1, x^2 \in \mathbf{\Gamma}$, $f(x^2) - f(x^1) - (\nabla f(x^1))(x^2 - x^1) = (x^2 - x^1)^T H(f(x^1 + \alpha(x^2 - x^1)))(x^2 - x^1)/2$ for some $0 < \alpha < 1$. But the latter expression is $\geqq 0$ since $H(f(\overline{x}))$ is PSD for all $\overline{x} \in \mathbf{\Gamma}$. So $f(x^2) - f(x^1) - (\nabla f(x^1))(x^2 - x^1) \geqq 0$ for all $x^1, x^2 \in \mathbf{\Gamma}$. By Theorem 15, this implies that $f(x)$ is convex.

□

Given a general twice continuously differentiable real valued function $f(x)$ defined on $\mathbf{R}^n$, it may be hard to check whether it is convex. For some $\overline{x} \in \mathbf{R}^n$, if $H(f(\overline{x}))$ is PD, we know that in a small convex neighborhood of $\overline{x}$, $H(f(x))$ is PSD, and hence $f(x)$ is locally convex in this neighborhood.

Theorem 18. *Let $g(x)$ be a real valued concave function defined on an open convex subset $\mathbf{\Gamma} \subset \mathbf{R}^n$. If $g(x)$ is twice differentiable at $\overline{x} \in \mathbf{\Gamma}$, $H(g(\overline{x}))$ is* NSD. *Conversely, if $g(x)$ is a twice differentiable real valued function defined on $\mathbf{\Gamma}$ and $H(g(\overline{x}))$ is* NSD *for all $\overline{x} \in \mathbf{\Gamma}$, $g(x)$ is concave.*

Proof. Similar to the proof of Theorem 17.

□

Exercises

---

10. Let $X^r = (x_1^r, \ldots, x_n^r, x_{n+1}^r)^T$ $r = 1$ to m be given points in $\mathbf{R}^{n+1}$. Let $x^r = (x_1^r, \ldots, x_n^r)^T$ $r = 1$ to m. It is required to check whether there exists a convex function $\theta(x)$ defined on $\mathbf{R}^n$ (with the objective value plotted along the x_{n+1}-axis in $\mathbf{R}^{n+1}$) satisfying the property $\theta(x^r) = x_{n+1}^r$ for $r = 1$ to m. Formulate this as a linear programming problem.

11. Let $f(x)$ be a real valued continuously differentiable convex function defined on $\mathbf{R}^n$. Let α be a real number and $\mathbf{K} = \{x : f(x) \leqq \alpha\}$. Given a point $x^0 \notin \mathbf{K}$, develop an efficient method for finding a separating hyperplane separating x^0 from $\mathbf{K}$. Generalize this to the case where $f(x) = (f_1(x), \ldots, f_m(x))$, each $f_i(x)$ being a real valued continuously differentiable function defined on $\mathbf{R}^n$, and $\alpha \in \mathbf{R}^m$.

12. Let $\theta(x)$ be a differentiable convex function defined over a convex set $\mathbf{K} \subset \mathbf{R}^n$. Let $\overline{x}$ be a given point in $\mathbf{K}$. If $\overline{x}$ satisfies the property that it minimizes the linear function $(\nabla\theta(\overline{x}))x$ over $x \in \mathbf{K}$, prove that $\overline{x}$ also minimizes $\theta(x)$ over $x \in \mathbf{K}$.

Convexity, Concavity of a Vector Function

Let $f(x)$ be the vector $(f_i(x))$ where each $f_i(x)$ is a real valued function defined on the convex set $\mathbf{\Gamma} \subset \mathbf{R}^n$. $f(x)$ is said to be convex or concave on $\mathbf{\Gamma}$, iff each $f_i(x)$ has the same property.

Subgradients, and Subdifferential Sets

Let $f(x)$ be a real valued convex function defined on $\mathbf{R}^n$. As defined in Section 2.7.1, the vector $d = (d_1, \dots, d_n)^T$ is said to be a **subgradient of** $f(x)$ at a point $x^0 \in \mathbf{R}^n$, if

$$f(x) \geqq f(x^0) + d^T(x - x^0), \quad \text{for all } x \in \mathbf{R}^n.$$

The set of all such vectors d satisfying this condition is known as the **subdifferential set** for $f(x)$ at x^0, and denoted by the symbol $\partial f(x^0)$.

By Theorem 15, if $f(x)$ is differentiable at x^0, the gradient vector $(\nabla f(x^0))^T \in \partial f(x^0)$, and in fact it can be shown that in this case $\partial f(x^0) = \{\nabla f(x^0)^T\}$. Also, as mentioned in Section 2.7.1, if $f(x) = \max\{f_1(x), \dots, f_r(x)\}$ where each $f_i(x)$ is a differentiable convex function defined on $\mathbf{R}^n$, then for any $\overline{x} \in \mathbf{R}^n$,

$$\partial f(\overline{x}) = \text{ convex hull of } \{\nabla f_i(\overline{x}) : i \text{ such that } f(\overline{x}) = f_i(\overline{x})\}.$$

See Section 2.7.1 for figures illustrating the subgradient property. The definition implies that if $f(x)$ is convex and $d \in \partial f(\overline{x})$, then the affine function $l(x) = f(\overline{x}) + d^T(x - \overline{x})$ is equal to $f(x)$ at $x = \overline{x}$, and is an underestimate for $f(x)$ at all points x.

So the error $f(x) - l(x) = f(x) - (f(\overline{x}) + d^T(x - \overline{x})) \geqq 0$ for all x and $d \in \partial f(\overline{x})$. See Section 2.7.1 for figures illustrating this property. The affine function $l(x)$ defined above is known as a **linearization of** $f(x)$ at $\overline{x}$.

If $h(x)$ is a concave function defined on $\mathbf{R}^n$, the vector d is said to be a subgradient of $h(x)$ at $\overline{x}$ if

$$h(x) \leqq h(\overline{x}) + d^T(x - \overline{x}) \text{ for all } x \in \mathbf{R}^n$$

and the set of all subgradients to $h(x)$ at $\overline{x}$ is denoted by $\partial h(\overline{x})$. With this definition, analogous results to those stated above, can be constructed for concave functions.

Let $g(x)$ be a real valued function defined on $\mathbf{R}^n$ which is neither convex nor concave. If $g(x)$ is differentiable at a point $\overline{x} \in \mathbf{R}^n$, the affine function $l(x) = g(\overline{x}) + \nabla g(\overline{x})(x - \overline{x})$ is known as the **linearization of** $g(x)$ at $\overline{x}$. However, since $g(x)$ is neither convex nor concave, it is possible for the error $g(x) - l(x)$ to take both positive and negative values over $\mathbf{R}^n$. See Figure 13.

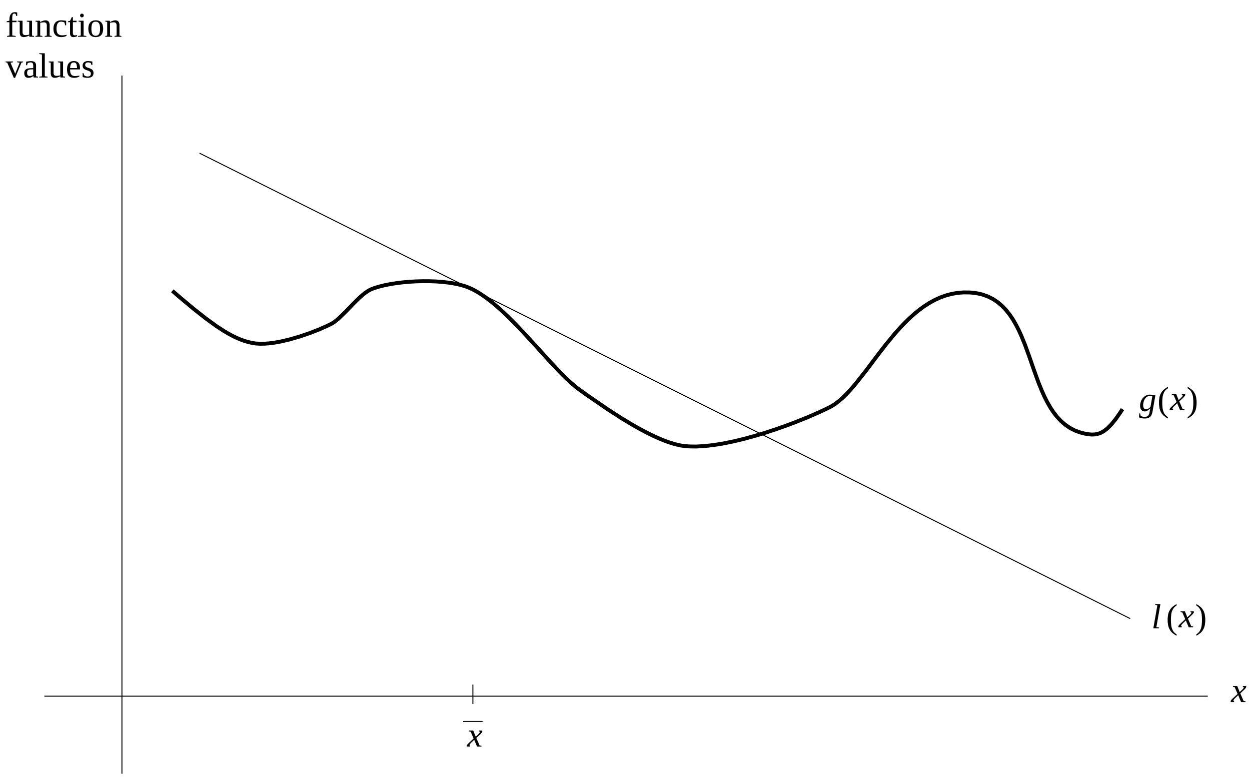

Figure 13 The linearization at $\overline{x}$, $l(x)$, of a differentiable function $g(x)$ which is neither convex nor concave may be $> g(x)$ at some points x, and $< g(x)$ at other points.

For this general function $g(x)$, if it is differentiable at $\overline{x}$, we define $\partial g(\overline{x}) = \{\nabla g(\overline{x})\}$. If $g(x)$ is not differentiable at $\overline{x}$, we let $\partial f(\overline{x})$ denote the convex hull of all limits of sequences of the form $\{\nabla g(x^r) : \{x^r\}$ is a sequence converging to $\overline{x}$, such that $g(x)$ is differentiable at each x^r in the sequence$\}$. In this case, vectors in the set $\partial g(\overline{x})$ are called generalized gradients or subgradients of $g(x)$ at $\overline{x}$. See F. H. Clarke [A1]. With this definition, it can be shown that if $g(x) = \max\{g_1(x), \ldots, g_m(x)\}$, where each $g_i(x)$ is a continuously differentiable function, then $\partial g(\overline{x}) =$ convex hull of $\{\nabla g_i(\overline{x}) : i$ such that $g(\overline{x}) = g_i(\overline{x})\}$. If $g(x)$ is convex, the set $\partial g(x)$ defined here equals the subdifferential set of $g(x)$ at x as defined earlier. Also, it can be shown under fairly general conditions on $g(x)$ (for example, if $g(x)$ is a locally Lipschitz function, that is, if there exists an $\alpha > 0$ such that $|g(x) - g(y)| \leqq \alpha\|x - y\|$ for all x, y) that the following mean value result holds: there exists an $\hat{x}$ on the line segment joining x and y and a $\hat{d} \in \partial g(\hat{x})$, satisfying

$$g(x) - g(y) = \hat{d}^T(x - y).$$

This definition of subgradients or generalized gradients for general functions is used in Section 10.7.9 in constructing an algorithm for constrained line minimization. Also see N. Z. Shor [A13] for a detailed treatment of various types of generalized gradients, and their applications in subgradient algorithms for nondifferentiable minimization.

4. OPTIMALITY CONDITIONS FOR SMOOTH OPTIMIZATION PROBLEMS

Here we briefly survey the known optimality conditions for NLPs in which the objective and constraint functions are continuously differentiable.

The Principles on Which Optimality Conditions are Based

Let $\mathbf{K}$ denote the set of feasible solutions for an optimization problem in which the objective function $\theta(x)$ is to be minimized. Let $\overline{x} \in \mathbf{K}$ be a feasible solution. A feasible direction at $\overline{x}$ for $\mathbf{K}$ is a direction y satisfying the property that beginning at $\overline{x}$, you can move a positive length along a straight line in the direction y, without leaving $\mathbf{K}$. Necessary optimality conditions for this optimization problem are derived, based on two very simple principles. These are the following:

1. If $\overline{x}$ is a local minimum for this optimization problem, then, as you move from $\overline{x}$ straight along any feasible direction at $\overline{x}$ for $\mathbf{K}$, in a small neighborhood of $\overline{x}$, the objective value cannot decrease.

2. Take a one dimensional, nonlinear, differentiable curve in the feasible region $\mathbf{K}$, passing through $\overline{x}$. If $\overline{x}$ is a local minimum for this optimization problem, then, as you move from $\overline{x}$ along this curve, in a small neighborhood of $\overline{x}$, the objective value cannot decrease (in effect this says that if $\overline{x}$ is a local minimum for $\theta(x)$ in $\mathbf{K}$, then $\overline{x}$ must be a local minimum for the one dimensional optimization problem of minimizing $\theta(x)$ on the curve).

Of course 1 is a special case of 2, since a straight line is a differentiable curve. These principles make it possible for us to derive necessary conditions for local minimality in higher dimensional feasible regions using well known necessary conditions for local minimality in one-dimensional optimization problems.

All the necessary optimality conditions are derived using the above principles. Even though the principles are the same, their application leads to optimality conditions which depend on the structure of the problem.

We will now derive optimality conditions for different types of nonlinear programming problems.

Unconstrained Minimization

First consider the unconstrained minimization problem

$$\begin{aligned}&\text{minimize } \theta(x)\\&\text{over } x \in \mathbf{R}^n.\end{aligned} \tag{49}$$

Given $\overline{x} \in \mathbf{R}^n$, $y \in \mathbf{R}^n$, $y \neq 0$, by differentiability of $\theta(x)$, we know that limit of $(\theta(\overline{x}+\alpha y) - \theta(\overline{x}) - \alpha(\nabla\theta(\overline{x}))y)/\alpha$ as α tends to zero is zero. So, if $(\nabla\theta(\overline{x}))y < 0$ by choosing α positive and sufficiently small, we will have $\theta(\overline{x}+\alpha y) < \theta(\overline{x})$. Similarly, if $(\nabla\theta(\overline{x}))y > 0$, by choosing α negative with sufficiently small absolute value we will have again $\theta(\overline{x}+\alpha y) < \theta(\overline{x})$. So if $\overline{x}$ is a local minimum for (49), we must have $(\nabla\theta(\overline{x}))y = 0$ for all $y \in \mathbf{R}^n$, that is

$$\nabla\theta(\overline{x}) = 0 \tag{50}$$

(50) is the **first order necessary condition** for $\overline{x}$ to be a local minimum for (49).

If $\theta(x)$ is twice continuously differential at $\overline{x}$, we know that the limit of $(\theta(\overline{x}+\alpha y) - \theta(\overline{x}) - \alpha(\nabla\theta(\overline{x}))y - (\alpha^{\mathbf{2}}/2)y^T H(\theta(\overline{x}))y)/\alpha^{\mathbf{2}}$ as α tends to zero is zero, where $H(\theta(\overline{x}))$ is the Hessian matrix (the matrix of second order partial derivatives) of $\theta(x)$ at $\overline{x}$. So if $\overline{x}$ is such that (50) is satisfied, and y is such that $y^T H(\theta(\overline{x}))y < 0$ then for $\alpha \neq 0$ and sufficiently small, we will have $\theta(\overline{x}+\alpha y) < \theta(\overline{x})$. So, if $\overline{x}$ is a local minimum for (49) we must have $y^T H(\theta(\overline{x}))y \geqq 0$ for all $y \in \mathbf{R}^n$, when $\overline{x}$ satisfies (50), that is

$$H(\theta(\overline{x})) \text{ must be PSD.} \tag{51}$$

(50) and (51) together are the **second order necessary conditions** for $\overline{x}$ to be a local minimum to (49).

We now state a sufficient optimality condition for (49) in the form of a theorem.

Theorem 19. *Suppose $\theta(x)$ is twice continuously differentiable, and $\overline{x}$ is a point satisfying*

$$\nabla\theta(\overline{x}) = 0, \quad \textit{and } H(\theta(\overline{x})) \text{is PD} \tag{52}$$

then $\overline{x}$ is a local minimum for (49).

Proof. Since $H(\theta(\overline{x}))$ is PD, all its principal subdeterminants are > 0. Since $\theta(x)$ is twice continuously differentiable, all principal subdeterminants of the Hessian matrix $H(\theta(x))$ are continuous functions. These facts imply that there exists an $\varepsilon > 0$, such that if $\mathbf{\Gamma} = \{x : \|x - \overline{x}\| < \varepsilon\}$, all principal subdeterminants of $H(\theta(x))$ are > 0 for all $x \in \mathbf{\Gamma}$. Being a Hessian matrix $H(\theta(x))$ is also symmetric, by Theorem 1.9 of Section 1.3.1, these facts imply that $H(\theta(x))$ is PSD for all $x \in \mathbf{\Gamma}$. By Theorem 17 of Appendix 3, this implies that $\theta(x)$ is convex over $x \in \mathbf{\Gamma}$. So by Theorem 15 of Appendix 3 (the gradient support inequality)

$$\begin{aligned}\theta(x) - \theta(\overline{x}) &\geqq (\nabla\theta(\overline{x}))(x - \overline{x}) \text{ for all } x \in \mathbf{\Gamma}\\&\geqq 0, \text{ since } \nabla\theta(\overline{x}) = 0 \text{ by (52).}\end{aligned}$$

This proves that $\overline{x}$ is a local minimum for $\theta(x)$.

□

Thus a **sufficient condition** for $\overline{x}$ to be a local minimum for (49) is (52).

Example 3

Consider the problem

minimize $\theta(x) = 2x_1^2 + x_2^2 + x_3^2 + x_1x_2 + x_2x_3 + x_3x_1 - 9x_1 - 9x_2 - 8x_3$
over $x \in \mathbf{R}^3$.

From the necessary optimality conditions, we know that every local minimum for this problem must satisfy

$$\begin{aligned}
\frac{\partial\theta(x)}{\partial x_1} &= 4x_1 + x_2 + x_3 - 9 = 0\\
\frac{\partial\theta(x)}{\partial x_2} &= x_1 + 2x_2 + x_3 - 9 = 0\\
\frac{\partial\theta(x)}{\partial x_3} &= x_1 + x_2 + 2x_3 - 8 = 0.
\end{aligned}$$

This system of equations has the unique solution $\overline{x} = (1, 3, 2)^T$. The Hessian matrix is

$$H(\theta(\overline{x})) = \begin{pmatrix} 4 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}.$$

This matrix is PD. So $\overline{x}$ satisfies the sufficient conditions for a local minimum. Clearly, here, $\theta(x)$ is convex and hence $\overline{x}$ is a global minimum for $\theta(x)$.

Example 4

Consider the problem

minimize $\theta(x) = 2x_1^2 + x_3^2 + 2x_1x_2 + 2x_1x_3 + 4x_2x_3 + 4x_1 - 8x_2 + 2x_3$
over $x \in \mathbf{R}^3$.

The first order necessary conditions for a local minimum are

$$\begin{aligned}
\frac{\partial\theta(x)}{\partial x_1} &= 4x_1 + 2x_2 + 2x_3 + 4 = 0\\
\frac{\partial\theta(x)}{\partial x_2} &= 2x_1 \phantom{{}+2x_2} + 4x_3 - 8 = 0\\
\frac{\partial\theta(x)}{\partial x_3} &= 2x_1 + 4x_2 + 2x_3 + 2 = 0.
\end{aligned}$$

This system has the unique solution $\tilde{x} = (-2, -1, 3)^T$. The Hessian matrix is

$$H(\theta(\tilde{x})) = 2 \begin{pmatrix} 2 & 1 & 1 \\ 1 & 0 & 2 \\ 1 & 2 & 1 \end{pmatrix}$$

which is not PSD. So $\tilde{x}$ violates the second order necessary conditions for a local minimum. So the function $\theta(x)$ here does not have a local minimum. It can be verified that in fact $\theta(x)$ is unbounded below on $\mathbf{R}^3$.

Example 5

Let $\theta(x) = -2x_1^{\mathbf{2}} - x_2^{\mathbf{2}} + x_1x_2 - 10x_1 + 6x_2$ and consider the problem of minimizing $\theta(x)$ over $x \in \mathbf{R}^2$. The first order necessary conditions for a local minimum are

$$\begin{aligned} \frac{\partial\theta(x)}{\partial x_1} &= -4x_1 + x_2 - 10 = 0 \\ \frac{\partial\theta(x)}{\partial x_2} &= x_1 - 2x_2 + 6 = 0 \end{aligned}$$

which has the unique solution $\hat{x} = (-2, 2)^T$. The Hessian matrix is

$$H(\theta(\hat{x})) = \begin{pmatrix} -4 & 1 \\ 1 & -2 \end{pmatrix}.$$

Since $H(\theta(\hat{x}))$ is not PSD, $\hat{x}$ violates the second order necessary conditions for being a local minimum of $\theta(x)$. So $\theta(x)$ has no local minimum. In fact, it can be verified that the Hessian matrix is ND, so $\hat{x}$ satisfies the sufficient condition for being a local maximum for $\theta(x)$ (a local maximum for $\theta(x)$ is a local minimum for $-\theta(x)$). Actually, $\theta(x)$ here is concave and $\hat{x}$ is a global maximum point for $\theta(x)$. It can be verified that $\theta(x)$ is unbounded below on $\mathbf{R}^2$.

An Important Caution for NLP Users

These examples point out one important aspect of using nonlinear programming algorithms in practical applications. One should not blindly accept any solution of the first order necessary optimality conditions as a solution to the problem, if it is a nonconvex programming problem (this caution can be ignored if the problem being solved is a linear or other convex programming problem). An effort should be made to check whether the solution is at least a local minimum by using second order necessary optimality conditions, or the sufficient optimality conditions, or at least through a local search in the neighborhood of the point.

Stationary Point Necessary Optimality Conditions for Constrained Minima

Consider the problem

$$\begin{array}{ll} \text{minimize} & \theta(x) \\ \text{subject to} & x \in \mathbf{\Gamma} \end{array} \tag{53}$$

where $\mathbf{\Gamma}$ is a specified subset of $\mathbf{R}^n$, and $\theta(x)$ is a real valued continuously differentiable function defined on $\mathbf{R}^n$.

Given $\overline{x} \in \mathbf{\Gamma}$, $y \neq 0$, $y \in \mathbf{R}^n$ is said to be a **feasible direction** for $\mathbf{\Gamma}$ at $\overline{x}$ if $\overline{x} + \alpha y \in \mathbf{\Gamma}$ for all $0 \leqq \alpha \leqq \lambda$ for some positive λ. As an example, if $\mathbf{\Gamma} = \{x : x = (x_1, x_2)^T, x_1 \geqq 0, x_2 \geqq 0\}$ and $\overline{x} = (1, 0)^T$, then $\{y : y = (y_1, y_2), y_2 \geqq 0\}$ is the set of feasible directions at $\overline{x}$.

Using the definition of differentiability, it follows that if $\overline{x} \in \mathbf{\Gamma}$ is a local minimum for (53), and $\theta(x)$ is continuously differentiable at $\overline{x}$, then we must have

$$(\nabla\theta(\overline{x}))y \geqq 0 \text{ for all feasible directions } y \text{ at } \overline{x} \text{ to } \mathbf{\Gamma}. \tag{54}$$

(54) are the **first order necessary conditions** for $\overline{x}$ to be a local minimum for (53). If $\theta(x)$ is twice continuously differentiable at $\overline{x} \in \mathbf{\Gamma}$, and $\overline{x}$ is a local minimum for (53), we must have

$$(54), \text{ and } y^T H(\theta(\overline{x}))y \geqq 0 \text{ for all feasible directions } y \text{ satisfying } (\nabla\theta(\overline{x}))y = 0. \tag{55}$$

The conditions (54), (55) become simplified if $\mathbf{\Gamma}$ is a convex set. In this case, a feasible direction y at $\overline{x}$ to $\mathbf{\Gamma}$ is $y = x - \overline{x}$ for any $x \in \mathbf{\Gamma}$. See Figure 14. So in case $\mathbf{\Gamma}$ is convex, the necessary conditions for $\overline{x} \in \mathbf{\Gamma}$ to be a local minimum is that (54), (55) hold for all $y = x - \overline{x}$, $x \in \mathbf{\Gamma}$.

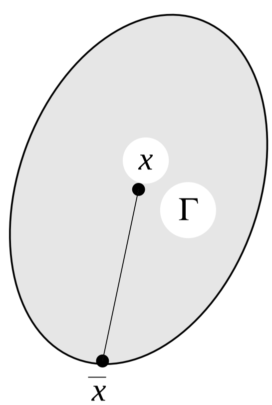

Figure 14 If $\mathbf{\Gamma}$ is a convex set, feasible directions at $\overline{x}$ to $\mathbf{\Gamma}$ are of the form $x - \overline{x}$ for any $x \in \mathbf{\Gamma}$, $x \neq \overline{x}$.

Example 6

Consider the problem

$$\begin{array}{lll} \text{minimize} & \theta(x) = 3x_1x_2 - x_1 - x_2 & \\ \text{subject to} & x_1 & \geqq 1 \\ & x_2 & \geqq 1. \end{array}$$

The set of feasible solutions, **K**, is marked in Figure 15.

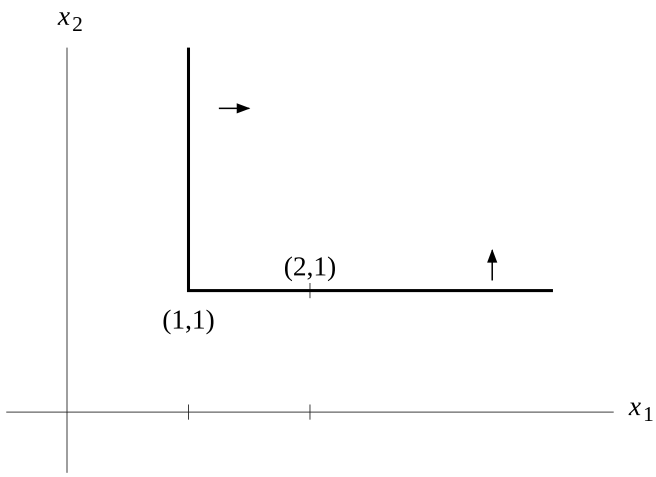

Figure 15

We have

$$\nabla\theta(x) = (3x_2 - 1, 3x_1 - 1)$$
$$H(\theta(x)) = \begin{pmatrix} 0 & 3 \\ 3 & 0 \end{pmatrix}.$$

Let $\overline{x} = (2,1)^T$. The set of feasible directions at $\overline{x}$ to **K** is clearly $\{y : y = (y_1, y_2)^T, y_2 \geqq 0\}$. $\nabla\theta(\overline{x}) = (2,5)$. $\overline{y} = (-1,0)^T$ is a feasible direction to **K** at $\overline{x}$, and yet $(\nabla\theta(\overline{x}))\overline{y} = -2 < 0$ and hence the necessary condition (54) is violated at $\overline{x}$.

Let $\hat{x} = (1,1)^T$. The set of feasible directions to **K** at $\hat{x}$ is clearly $\{y : y \geqq 0\}$. $\nabla\theta(\hat{x}) = (2,2)$ and we verify that both the necessary optimality conditions (54) and (55) are satisfied at $\hat{x}$. Acutally, $\hat{x}$ is the global minimum for this problem.

The conditions (54), (55) are respectively the **first and second order stationary point necessary optimality conditions** for the NLP (53).

Variational Inequality Problem

The stationary point necessary optimality conditions discussed above, lead to a problem commonly known as the variational inequality problem. In this problem we are given a real vector function $f(x) = (f_1(x), \dots, f_n(x))^T$ defined over $\mathbf{R}^n$, and a subset $\mathbf{K} \subset \mathbf{R}^n$. The variational inequality problem with this data, is to find a point $x^* \in \mathbf{K}$ satisfying

$$(x - x^*)^T f(x^*) \geqq 0 \text{ for all } x \in \mathbf{K}.$$

Suppose $\mathbf{K} = \{x : Ax \geqq b, x \geqq 0\}$ where A, b are given matrices of orders $m \times n$ and $m \times 1$, the above variational inequality problem is equivalent to the nonlinear complementarity problem: find $z \in \mathbf{R}^{n+m}$ satisfying

$$z \geqq 0,\ g(z) \geqq 0,\ z^T g(z) = 0$$

where $z = (x_1, \dots, x_n; y_1, \dots, y_m)^T$, $y = (y_1, \dots, y_m)^T$ and

$$g(z) = \begin{pmatrix} f(x) & -A^T y \\ Ax & -b \end{pmatrix}.$$

Optimality Conditions for Equality Constrained Minimization

Consider the NLP

$$\begin{array}{ll} \text{minimize} & \theta(x) \\ \text{subject to} & h_i(x) = 0,\ i = 1 \text{ to } m \end{array} \tag{56}$$

where $\theta(x)$, $h_i(x)$ are all real valued continuously differentiable functions defined on $\mathbf{R}^n$. Let $h(x) = (h_1(x), \dots, h_m(x))^T$. The set of feasible solutions is a surface in $\mathbf{R}^n$, and it is smooth if each $h_i(x)$ is a smooth function (i. e., continuously differentiable). If $\overline{x}$ is a feasible point, when some of the $h_i(x)$ are nonlinear, there may be no feasible direction at $\overline{x}$. In order to retain feasibility while moving from $\overline{x}$, one has to follow a nonlinear curve through $\overline{x}$ which lies on the feasible surface. See Figure 16.

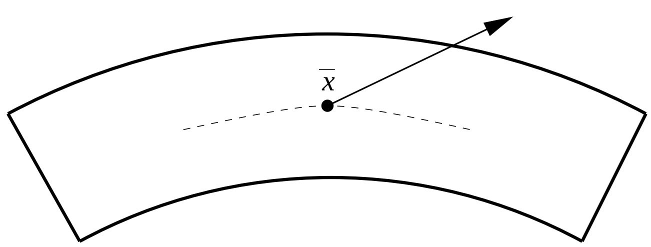

Figure 16 Feasible surface $\mathbf{\Gamma} = \{x : h_1(x) = 0\}$ satisfying a nonlinear equation. At $\overline{x} \in \mathbf{\Gamma}$, the direction marked by the arrow is not a feasible direction, since any move of positive length in that direction takes the point out of $\mathbf{\Gamma}$. To move from $\overline{x}$ and remain inside $\mathbf{\Gamma}$ one has to follow a curve like the dashed curve.

A curve in $\mathbf{R}^n$ is the locus of a point $x(\lambda) = (x_j(\lambda))$, where each $x_j(\lambda)$ is a real valued function of the real parameter λ, as the parameter varies over some interval of the real line. See Figure 17.

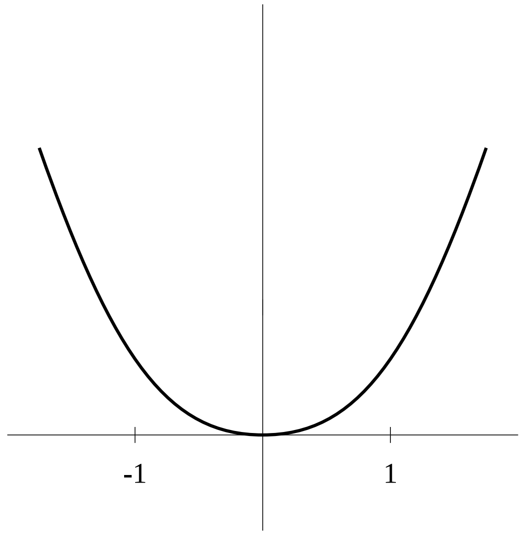

Figure 17 A curve in $\mathbf{R}^2$. $\{x(\lambda) = (\lambda, \lambda^2) : -1 \leqq \lambda \leqq 1\}$ is a piece of a curve (parabola) in $\mathbf{R}^2$ through the origin $x = (0, 0)^T$.

The curve $x(\lambda) = (x_j(\lambda)$ is said to be differentiable at λ if $\frac{dx_j(\lambda)}{d\lambda}$ exists for all j, and twice differentiable if $\frac{d^2x_j(\lambda)}{d\lambda^2}$ exists for all j. The curve $x(\lambda)$ is said to pass through the point $\overline{x}$ if $\overline{x} = x(\overline{\lambda})$ for some $\overline{\lambda}$.

If the curve $x(\lambda)$ defined over $a < \lambda < b$ is differentiable at $\overline{\lambda}$, $a < \overline{\lambda} < b$, then the line $\{x = x(\overline{\lambda}) + \delta \frac{dx}{d\lambda}(\overline{\lambda}) : \delta \text{ real}\}$ is the tangent line to the curve at the point $x(\overline{\lambda})$ on it. See Figure 18.

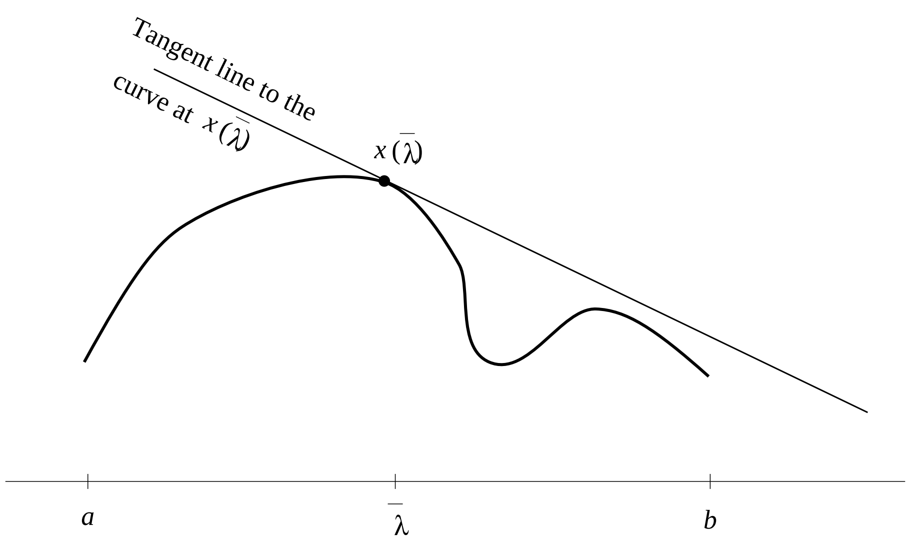

Figure 18

The **tangent plane** at a feasible point $\overline{x}$ to (56) is defined to be the set of all directions $(\frac{dx(\lambda)}{d\lambda})_{\lambda=0}$, where $x(\lambda)$ is a differential curve in the feasible region with $x(0) = \overline{x}$. See Figure 19.

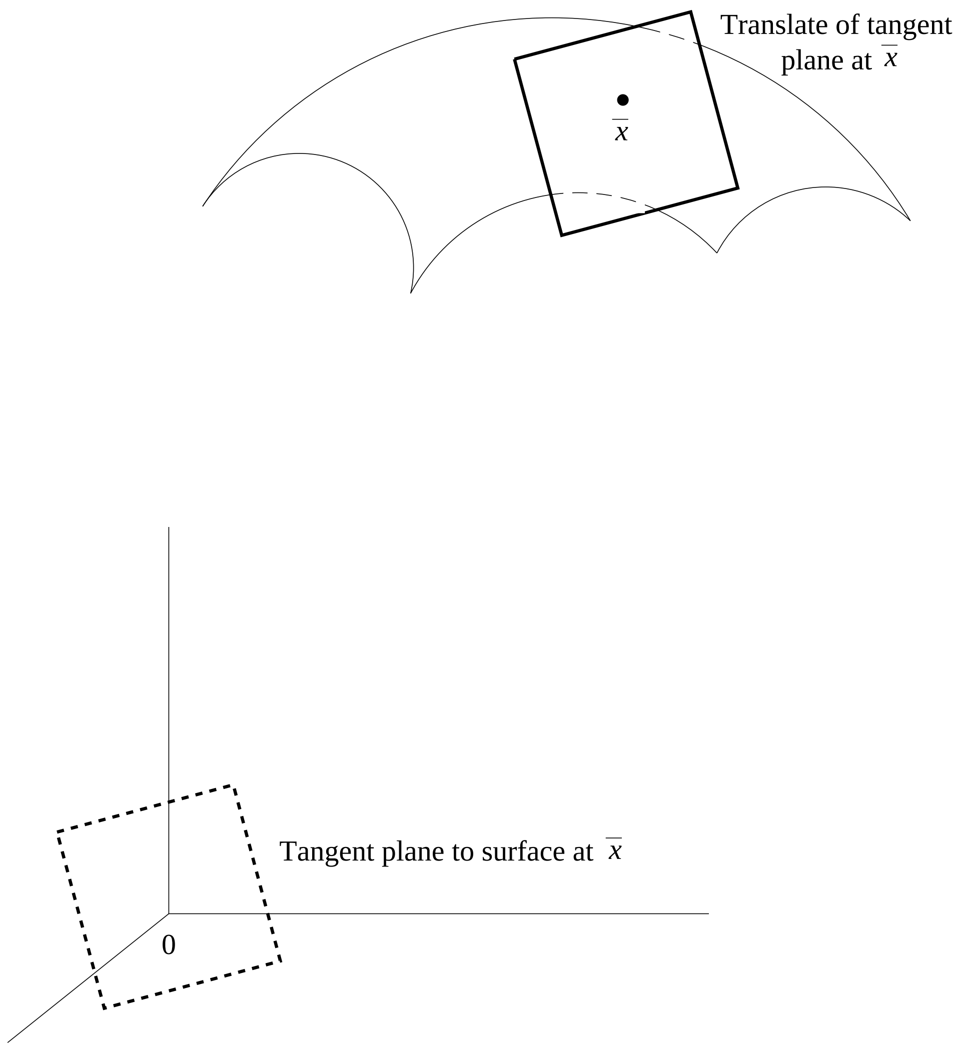

Figure 19 The tangent plane to surface $\{x : h_1(x) = 0\}$ at a point $\overline{x}$ on it is the collection of all directions of tangent lines to differentiable curves lying in surface and passing through $\overline{x}$.

We need the following results to study these tangent planes.

The Implicit Function Theorem

Consider the system of m equations in n variables $x_1, \ldots, x_n$

$$f_i(x_1, \ldots, x_n) = 0, \ i = 1 \text{ to } m \tag{57}$$

where each $f_i(x)$ is continuously differentiable in some open subset $\mathbf{D} \subset \mathbf{R}^n$. Let $\overline{x} \in \mathbf{D}$ be feasible to (57) and let the subset of m variables, $x_1, \ldots, x_m$, say, be such that the $m \times m$ Jacobian $(\frac{\partial f_i(\overline{x})}{\partial x_j} : i = 1 \text{ to } m, \ j = 1 \text{ to } m)$ is nonsingular. Then in a neighborhood of $\overline{x}$, we can use the equations in (57) to express $x_1, \ldots, x_m$ as functions of $x_{m+1}, \ldots, x_n$ on the set of feasible solutions of (57). That is, there exists a neighborhood $\mathcal{D}$ of $(\overline{x}_{m+1}, \ldots, \overline{x}_n)$ in $\mathbf{R}^{n-m}$ and real valued differentiable functions $\psi_i(x_{m+1}, \ldots, x_n)$, $i = 1$ to m; such that for $(x_{m+1}, \ldots, x_n) \in \mathcal{D}$, (57) is equivalent to

$$x_i = \psi_i(x_{m+1}, \ldots, x_n), \ i = 1 \text{ to } m$$

i. e.,

$$f_i(\psi_1(x_{m+1}, \ldots, x_n), \ldots, \psi_m(x_{m+1}, \ldots, x_n), x_{m+1}, \ldots, x_n) = 0, \ i = 1 \text{ to } m \tag{58}$$

holds for all $(x_{m+1}, \ldots, x_n) \in \mathcal{D}$. Further, the partial derivatives $\frac{\partial \psi_i(\overline{x}_{m+1}, \ldots, \overline{x}_n)}{\partial x_j}$, $i = 1$ to m, $j = m + 1$ to n, are obtained by solving the system of equations

$$\sum_{r=1}^{m} \frac{\partial f_i(\overline{x})}{\partial x_r} \frac{\partial \psi_r(\overline{x}_{m+1}, \ldots, \overline{x}_n)}{\partial x_j} + \frac{\partial f_i(\overline{x})}{\partial x_j} = 0, \ j = m + 1 \text{ to } n, \ i = 1 \text{ to } m. \tag{59}$$

It can be verified that (59) is just obtained by setting the derivative of the identity (58) at $\overline{x}$ with respect to x_j to zero for each $j = m + 1$ to n and $i = 1$ to m. See references [10.33] for a proof of the implicit function theorem.

Example 7: An Illustration of the Implicit Function Theorem.

Here we provide a simple example to illustrate the implicit function theorem using a linear system of constraints. Consider the following system in the variables $x = (x_1, x_2, x_3, x_4, x_5)^T$.

$$\begin{aligned} f_1(x) &= x_1 + x_2 + x_3 + x_4 - x_5 - 12 = 0 \\ f_2(x) &= -x_1 + x_2 - 2x_3 - x_4 + 4x_5 - 2 = 0. \end{aligned}$$

Let $\overline{x} = (5, 7, 0, 0, 0)^T$. $\overline{x}$ is a feasible solution, and

$$\begin{bmatrix} \frac{\partial f_1(\overline{x})}{\partial x_1} & \frac{\partial f_1(\overline{x})}{\partial x_2} \\ \frac{\partial f_2(\overline{x})}{\partial x_1} & \frac{\partial f_2(\overline{x})}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

is nonsingular. Therefore, by the implicit function theorem, it is possible to express x_1, x_2 as functions of the remaining variables x_3, x_4, x_5 in a neighborhood of $\overline{x}$ in the feasible region. Since the constraints are linear, we can do this explicitly by solving for x_1, x_2 in terms of x_3, x_4, x_5 using these two equations, and this leads to

$$\begin{aligned} x_1(x_3,x_4,x_5) &= -\frac{3}{2}x_3 - x_4 - \frac{5}{2}x_5 + 10 \\ x_2(x_3,x_4,x_5) &= \ \ \frac{1}{2}x_3 \qquad\quad - \frac{3}{2}x_5 + 7 \end{aligned}$$

where $x_1(x_3,x_4,x_5)$ and $x_2(x_3,x_4,x_5)$ are the expressions for x_1, x_2 as functions of x_3, x_4, x_5, on the feasible region for this system. When the equations are nonlinear, it may not be possible to obtain these expressions explicitly, but the implicit function theorem guarantees the existence of them in a neighborhood of $\overline{x}$ in the feasible region.

We verify that the partial derivatives are

$$\begin{bmatrix} \frac{\partial x_1}{\partial x_3}, & \frac{\partial x_1}{\partial x_4}, & \frac{\partial x_1}{\partial x_5} \\ \frac{\partial x_2}{\partial x_3}, & \frac{\partial x_2}{\partial x_4}, & \frac{\partial x_2}{\partial x_5} \end{bmatrix} = \begin{bmatrix} -\frac{3}{2}, & -1, & -\frac{5}{2} \\ \frac{1}{2}, & 0, & -\frac{3}{2} \end{bmatrix}.$$

The equations corresponding to (59) for this system for $j = 3$ are

$$\begin{aligned} \frac{\partial f_1}{\partial x_1}\frac{\partial x_1}{\partial x_3} + \frac{\partial f_1}{\partial x_2}\frac{\partial x_2}{\partial x_3} + \frac{\partial f_1}{\partial x_3} &= \ \ \frac{\partial x_1}{\partial x_3} + \frac{\partial x_2}{\partial x_3} + 1 = 0 \\ \frac{\partial f_2}{\partial x_1}\frac{\partial x_1}{\partial x_3} + \frac{\partial f_2}{\partial x_2}\frac{\partial x_2}{\partial x_3} + \frac{\partial f_2}{\partial x_3} &= -\frac{\partial x_1}{\partial x_3} + \frac{\partial x_2}{\partial x_3} - 2 = 0 \end{aligned}$$

which together yield $\frac{\partial x_1}{\partial x_3} = -\frac{3}{2}$, $\frac{\partial x_2}{\partial x_3} = \frac{1}{2}$, same as the values obtained above. In a similar manner, writing the equations corresponding to (59) for this system for $j = 4, 5$, we can compute the values $\frac{\partial x_i}{\partial x_j}$ for $i = 1, 2$, $j = 4, 5$, and verify that they are the same as those obtained above.

Constraint Qualifications

In general, determining the tangent plane for (56) at the feasible point $\overline{x}$ is hard. However, if the constraint functions $h_i(x)$ satisfy ceratin conditions at $\overline{x}$, it becomes possible to obtain a simple characterization of the tangent plane for (56) at $\overline{x}$. So these conditions are called **constraint qualifications** because these conditions are specifically on the constraints in (56), not so much on the set of feasible solutions of (56). Several constraint qualifications have been developed, but for most of them, it is very hard to verify whether they hold in any given problem. We will only discuss one constraint qualification, which can be checked efficiently. It is called the **regularity condition**.

The regularity condition is said to hold for (56) at the feasible point $\overline{x}$ if the Jacobian matrix $(\frac{\partial h_i(\overline{x})}{\partial x_j} : i = 1 \text{ to } m, j = 1 \text{ to } n)$ has rank m, in this case the feasible point $\overline{x}$ is called a **regular point** for (56).

Definition. *We denote by* $\nabla h(x) = (\frac{\partial h_i(x)}{\partial x_j} : i = 1 \text{ to } m, j = 1 \text{ to } n)$*, the Jacobian matrix of order* $m \times n$*; the* i*th row vector of* $\nabla h(x)$ *is the gradient vector of* $h_i(x)$ *written as a row vector.*

Tangent Planes at Regular Points

Theorem 20. *If $\overline{x}$ is a regular point for (56), the tangent plane for (56) at $\overline{x}$ is $\{y : (\nabla h(\overline{x}))y = 0\}$.*

Proof. Let $x(\alpha)$ be a differentiable curve lying in the feasible region for α lying in an interval around zero, with $x(0) = \overline{x}$ and $\frac{dx(0)}{d\alpha} = y$. So $h(x(\alpha)) = 0$ for all values of α lying in an interval around zero, and hence $(\frac{dh(x(\alpha))}{d\alpha})_{\alpha=0} = 0$, that is $(\nabla h(\overline{x}))y = 0$. This implies that the tangent plane is a subset of $\{y : (\nabla h(\overline{x}))y\} = 0$.

Suppose $y \in \{y : (\nabla h(\overline{x}))y = 0\}$ and $y \neq 0$. Define new variables $u = (u_1, \dots, u_m)^T$. Consider the following system of m equations in $m+1$ variables $u_1, \dots, u_m$, α.

$$g_i(u, \alpha) = h_i(\overline{x} + \alpha y + (\nabla h(\overline{x}))^T u) = 0, \ i = 1 \text{ to } m. \tag{60}$$

It can be verified that $g(0,0) = 0$ and the Jacobian matrix of $g(u, \alpha)$ with respect to u is nonsingular at $u = 0$, $\alpha = 0$ (since $\overline{x}$ is a regular point of (56)). So by applying the implicit function theorem on (60), we can express u as a differentiable function of α, say $u(\alpha)$, in an interval around $\alpha = 0$, and that (60) holds as an identity in this interval when u in (60) is replaced by $u(\alpha)$, and that $u(0) = 0$, and $\frac{du(0)}{d\alpha}$ is obtained by solving

$$\Big(\frac{d}{d\alpha} h(\overline{x} + \alpha y + (\nabla h(\overline{x}))^T u(\alpha))\Big)_{\alpha=0} = 0$$

which leads to $\frac{d}{d\alpha}u(0) = 0$ since $\nabla h(\overline{x})$ has rank m. So if we define

$$x(\alpha) = \overline{x} + \alpha y + (\nabla h(\overline{x}))^T u(\alpha)$$

this defines a differentiable curve lying in the feasible region for (56) for values of α in an interval around $\alpha = 0$, and that $\frac{dx}{d\alpha}(0) = y$, which implies that y is in the tangent plane for (56) at $\overline{x}$.

$\square$

Example 8

Consider the system

$$h(x_1, x_2) = x_1 = 0$$
$$x = (x_1, x_2)^T \in \mathbf{R}^2.$$

The set of feasible solutions is the x_2-axis in $\mathbf{R}^2$, since $\nabla h(x) = (1, 0)$ every feasible point is a regular point, and the tangent plane at any feasible point x is again the x_2-axis $= \{y : (\nabla h(\overline{x}))y = 0\} = \{y : y = (y_1, y_2), y_1 = 0\}$. On the other hand the system

$$g(x_1, x_2) = x_1^{\mathbf{3}} = 0$$
$$x = (x_1, x_2)^T \in \mathbf{R}^2$$

has the same set of feasible solutions, namely the x_2-axis in $\mathbf{R}^2$. Since $\nabla g(x) = (3x_1^{\mathbf{2}}, 0)$ is zero whenever x is feasible, no feasible point is regular. The tangent plane at every feasible solution is again the x_2-axis in $\mathbf{R}^2$, but $\{y : \nabla g(x)y = 0\} = \mathbf{R}^2$ for every feasible solution x.

Optimality Conditions

Using Theorem 20 we can now derive optimality conditions for (56). If $\overline{x}$ is a feasible regular point for (56), and it is a local minimum, clearly along every differentiable curve $x(\alpha)$ lying in the feasible region for (56) for values of α in an interval around $\alpha = 0$, satisfying $x(0) = \overline{x}$; $\alpha = 0$ must be a local minimum for $\theta(x)$ on this curve. That is, for the problem of minimizing $\theta(x(\alpha))$ over this interval for α, $\alpha = 0$ must be a local minimum. Since $\alpha = 0$ is an interior point of this interval this implies that $\frac{d\theta}{d\alpha}(x(0))$ must be zero. Applying this to all such curves and using Theorem 20 we conclude that $(\nabla\theta(\overline{x}))y = 0$ for all y satisfying $(\nabla h(\overline{x}))y = 0$. By Theorem 1 (see Exercise 5) this implies that there must exist $\overline{\mu} = (\overline{\mu}_1, \ldots, \overline{\mu}_m)$ such that

$$\begin{gathered}\nabla\theta(\overline{x}) - \sum_{i=1}^{m} \overline{\mu}_i \nabla h_i(\overline{x}) = 0 \\ \text{and by feasibility } h(\overline{x}) = 0\end{gathered} \tag{61}$$

the conditions (61) are the **first order necessary optimality conditions** for (56), the vector $\overline{\mu}$ is the vector of **Lagrange multipliers**. (61) is a system of $(n+m)$ equations in $(n+m)$ unknowns (including $\overline{x}$ and $\overline{\mu}$) and it may be possible to solve (61) using algorithms for solving nonlinear equations. If we define the Lagrangian for (56) to be $L(x,\mu) = \theta(x) - \mu h(x)$ where $\mu = (\mu_1, \ldots, \mu_m)$, $h(x) = (h_1(x), \ldots, h_m(x))^T$, (61) becomes: $(\overline{x}, \overline{\mu})$ satisfies

$$\begin{aligned} h(x) &= 0 \\ \nabla_x L(x,\mu) &= 0. \end{aligned} \tag{62}$$

We will now derive the second order necessary optimality conditions for (56). Suppose the functions $\theta(x)$, $h_i(x)$ are all twice continuously differentiable. Let $\overline{x}$ be a feasible solution for (56) which is a regular point. If $\overline{x}$ is a local minimum for (56), by the first order necessary optimality conditions (61), there must exist a row vector of Lagrange multipliers, $\overline{\mu} = (\overline{\mu}_1, \ldots, \overline{\mu}_m)$ such that $\nabla_x(L(\overline{x}, \overline{\mu})) = 0$, where $L(x, \overline{\mu}) = \theta(x) - \overline{\mu}h(x)$ is the Lagrangian. Since $\overline{x}$ is a regular point, the tangent plane to (56) at $\overline{x}$ is $\mathbf{T} = \{y : (\nabla h(\overline{x}))y = 0\}$. Suppose there exists a $\overline{y} \in \mathbf{T}$ satisfying $\overline{y}^T H_x(L(\overline{x}, \overline{\mu}))\overline{y} < 0$. Since $\overline{y} \in \mathbf{T}$, and all the functions are twice continuously differentiable, there exists a twice differentiable curve $x(\lambda)$ through $\overline{x}$ lying in the feasible region (i. e., $x(0) = \overline{x}$, and the curve is defined in an interval of λ with 0 as an interior point, with $h(x(\lambda)) = 0$ for all λ in this interval), such that $(\frac{dx(\lambda)}{d\lambda})_{\lambda=0} = \overline{y}$. Now,

$$\begin{aligned} \frac{d}{d\lambda} L(x(\lambda), \overline{\mu}) &= (\nabla_x L(x(\lambda), \overline{\mu}))\Big(\frac{dx(\lambda)}{d\lambda}\Big) \\ \frac{d^2}{d\lambda^2} L(x(\lambda), \overline{\mu}) &= \Big(\frac{dx(\lambda)}{d\lambda}\Big)^T H_x(L(x(\lambda), \overline{\mu}))\frac{dx(\lambda)}{d\lambda} + (\nabla_x L(x(\lambda), \overline{\mu}))\Big(\frac{d^2 x(\lambda)}{d\lambda^2}\Big) \end{aligned}$$

where $\nabla_x(L(\overline{x}, \overline{\mu}))$, $H_x(L(\overline{x}, \overline{\mu}))$ are the row vector of partial derivatives with respect to x, and the Hessian matrix with respect to x of $L(x, \overline{\mu})$ at $x = \overline{x}$ respectively. At

$\lambda = 0$, we have $\nabla_x L(x(0), \overline{\mu}) = \nabla_x L(\overline{x}, \overline{\mu}) = 0$ by the first order necessary optimality conditions.

So, from the above

$$\left(\frac{d}{d\lambda} L(x(\lambda), \overline{\mu})\right)_{\lambda=0} = 0$$
$$\left(\frac{d^2}{d\lambda^2} L(x(\lambda), \overline{\mu})\right)_{\lambda=0} = \overline{y}^T H_x(L(\overline{x}, \overline{\mu}))\overline{y}.$$

Using these in a Taylor series expansion for $f(\lambda) = L(x(\lambda), \overline{\mu})$ up to second order around $\lambda = 0$ leads to

$$f(\lambda) = L(x(\lambda), \overline{\mu}) = L(\overline{x}, \overline{\mu}) + \frac{\lambda^{\mathbf{2}}}{2}\overline{y}^T H_x(L(\overline{x}, \overline{\mu}))\overline{y} + 0(\lambda)$$

where $0(\lambda)$ is a function of λ satisfying the property that limit of $(0(\lambda))/\lambda^{\mathbf{2}}$ as λ tends to zero, is zero. Since $h(x(\lambda)) = 0$ for every point on the curve, we have $f(\lambda) = L(x(\lambda), \overline{\mu}) = \theta(x(\lambda))$ for all λ in the interval of λ on which the curve is defined. So in the neighborhood of $\lambda = 0$ on the curve we have from the above

$$\frac{2(\theta(x(\lambda)) - \theta(\overline{x}))}{\lambda^{\mathbf{2}}} = \frac{2(f(\lambda) - f(0))}{\lambda^{\mathbf{2}}} = \overline{y}^T H_x(L(\overline{x}, \overline{\mu}))\overline{y} + \frac{2(0(\lambda))}{\lambda^{\mathbf{2}}}$$

and since $\overline{y}^T H_x(L(\overline{x}, \overline{\mu}))\overline{y} < 0$ and limit of $(0(\lambda)/\lambda^{\mathbf{2}})$ as λ tends to zero is zero, for all λ sufficiently small $\theta(x(\lambda)) - \theta(\overline{x}) < 0$. For all these λ, $x(\lambda)$ is a point on the curve in the feasible region in the neighborhood of $\overline{x}$, and this is a contradiction to the fact that $\overline{x}$ is a local minimum for (56).

In fact it can be verified that $\overline{y}^T H_x(L(\overline{x}, \overline{\mu}))\overline{y} = (\frac{d^2 f(\lambda)}{d\lambda^2})_{\lambda=0}$, and if this quantity is < 0, $\lambda = 0$ cannot be a local minimum for the one variable minimization problem of minimizing $f(\lambda) = \theta(x(\lambda))$ over λ; or equivalently, that $\overline{x} = x(0)$ is not a local minimum for $\theta(x)$ along the curve $x(\lambda)$.

These facts imply that if $\theta(x)$, $h_i(x)$ are all twice continuously differentiable, and $\overline{x}$ is a regular point which is a feasible solution and a local minimum for (56), there must exist a Lagrange multiplier vector $\overline{\mu}$ such that the following conditions hold.

$$\begin{gathered} h(\overline{x}) = 0 \\ \nabla_x L(\overline{x}, \overline{\mu}) = \nabla\theta(\overline{x}) - \overline{\mu}\nabla h(\overline{x}) = 0 \\ y^T H_x(L(\overline{x}, \overline{\mu}))y \geqq 0 \text{ for all } y \in \mathbf{T} = \{y : (\nabla h(\overline{x}))y = 0\}, \\ \text{that is } H_x(L(\overline{x}, \overline{\mu})) \text{ is PSD on the subspace } \mathbf{T}. \end{gathered} \tag{63}$$

These are the **second order necessary optimality conditions** for a regular feasible point $\overline{x}$ to be a local minimum for (56).

We now state a sufficient optimality condition for (56) in the form of a theorem.

Theorem 21. *Suppose $\theta(x)$, $h_i(x)$, $i = 1$ to m are all twice continuously differentiable functions, and $\overline{x}$ is a feasible point such that there exists a Lagrange multiplier vector $\overline{\mu} = (\overline{\mu}_1, \ldots, \overline{\mu}_m)$ which together satisfy*

$$\begin{aligned} h(\overline{x}) &= 0 \\ \nabla\theta(\overline{x}) - \overline{\mu}\nabla h(\overline{x}) &= 0 \\ y^T H_x(L(\overline{x}, \overline{\mu}))y > 0 \text{ for all } y &\in \{y : (\nabla h(\overline{x}))y = 0\}, y \neq 0 \end{aligned} \tag{64}$$

where $L(x, \overline{\mu}) = \theta(x) - \overline{\mu}h(x)$ is the Lagrangian for (56). Then $\overline{x}$ is a local minimum for (56).

Proof. Suppose $\overline{x}$ is not a local minimum for (56). There must exist a sequence of distinct feasible points $\{x^r : r = 1, 2, \ldots\}$ converging to $\overline{x}$ such that $\theta(x^r) < \theta(\overline{x})$ for all r. Let $\delta_r = \|\overline{x} - x^r\|$, $y^r = (x^r - \overline{x})/\delta_r$. Then $\|y^r\| = 1$ for all r and $x^r = \overline{x} + \delta_r y^r$. Thus $\delta_r \to 0^+$ as $r \to \infty$. Since the sequence of points $\{y^r : r = 1, 2, \ldots)$ all lie on the surface of the unit sphere in $\mathbf{R}^n$, a compact set, the sequence has at least one limit point. Let $\overline{y}$ be a limit point of $\{y^r : r = 1, 2, \ldots\}$. There must exist a subsequence of $\{y^r : r = 1, 2, \ldots\}$ which converges to $\overline{y}$, eliminate all points other than those in this subsequence, and for simplicity call the remaining sequence by the same notation $\{y^r : r = 1, 2, \ldots\}$. So now we have a sequence of points $x^r = \overline{x} + \delta_r y^r$ all of them feasible, such that $\|y^r\| = 1$ for all $r, y^r \to \overline{y}$ and $\delta_r \to 0$ as $r \to \infty$. By feasibility $h(\overline{x} + \delta_r y^r) = 0$ for all r, and by the differentiability of $h(x)$ we have

$$\begin{aligned} 0 = h(\overline{x} + \delta_r y^r) &= h(\overline{x}) + \delta_r \nabla h(\overline{x})y^r + 0(\delta_r) \\ &= \delta_r \nabla h(\overline{x})y^r + 0(\delta_r) \end{aligned}$$

Dividing by $\delta_r > 0$, and taking the limit as $r \to \infty$ we see that $\nabla h(\overline{x})\overline{y} = 0$.

Since $L(x, \overline{\mu})$ is a twice continuously differentiable function in x, applying Taylor's theorem to it, we conclude that for each r, there exists a $0 \leqq \alpha_r \leqq \delta_r$ such that

$$L(\overline{x} + \delta_r y^r, \overline{\mu}) = L(\overline{x}, \overline{\mu}) + \delta_r \nabla_x L(\overline{x}, \overline{\mu})y^r + (1/2)\delta_r^2 (y^r)^T H_x(L(\overline{x} + \alpha_r y^r, \overline{\mu}))y^r.$$

From the fact that $\overline{x} + \delta_r y^r = x^r$ and $\overline{x}$ are feasible, we have $L(x^r, \overline{\mu}) = \theta(x^r)$ and $L(\overline{x}, \overline{\mu}) = \theta(\overline{x})$. Also, from (64), $\nabla_x L(\overline{x}, \overline{\mu}) = 0$. So, from the above equation, we have

$$\theta(x^r) - \theta(\overline{x}) = (1/2)\delta_r^2 (y^r)^T H_x(L(\overline{x} + \alpha_r y^r, \overline{\mu}))y^r. \tag{65}$$

Since $0 \leqq \alpha_r \leqq \delta_r$, and $\delta_r \to 0$ as $r \to \infty$, and by continuity, we know that $H_x(L(\overline{x} + \alpha_r y^r, \overline{\mu}))$ converges to $H_x(L(\overline{x}, \overline{\mu}))$ as $r \to \infty$. Since $y^r \to \overline{y}$ as $r \to \infty$, and $\nabla h(\overline{x})\overline{y} = 0$, from the last condition in (64) and continuity we conclude that when r is sufficiently large, the right-hand side of (65) is ≥ 0, while the left-hand side is < 0, a contradiction. So, $\overline{x}$ must be a local minimum for (56).

$\square$

Thus, (64) provides a **sufficient condition** for a feasible point $\overline{x}$ to be a local minimum for (56).

Example 9

Consider the problem

$$\begin{array}{ll} \text{minimize} & \theta(x) = x_1 x_2 \\ \text{subject to} & x_1 + x_2 = 2. \end{array}$$

The Lagrangian is $L(x, \lambda) = x_1 x_2 - \lambda(x_1 + x_2 - 2)$. So, the first order necessary optimality conditions are

$$\frac{\partial L(x, \lambda)}{\partial x} = (x_2 - \lambda, x_1 - \lambda) = 0$$

which together with the feasibility conditions lead to $\overline{x} = (1, 1)^T$. $\overline{x}$ is the unique solution for the first order necessary optimality conditions. $\overline{x}$, $\overline{\lambda} = 1$ together satisfy the first order necessary conditions for a local minimum. The Hessian of the Lagrangian is

$$H_x(L(\overline{x}, \overline{\lambda})) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The tangent plane at $\overline{x}$ is $\{y : y_1 + y_2 = 0\}$. So on the tangent plane, $y^T H_x((\overline{x}, \overline{\lambda}))y = 2y_1 y_2 = -2y_2^2 < 0$, whenever $y \neq 0$. So the second order necessary optimality conditions for a local minimum are violated at $\overline{x}$. In fact it can be verified that $\overline{x}$ satisfies the sufficient conditions for being a local maximum for $\theta(x)$ in the feasible region. $\theta(x)$ has no local minimum in the feasible region, it is unbounded below in the feasible region.

Example 10

Consider the problem

$$\begin{array}{ll} \text{minimize} & -x_1 - x_2 \\ \text{subject to} & x_1^2 + x_2^2 - 8 = 0. \end{array}$$

The Lagrangian is $L(x, \lambda) = -x_1 - x_2 - \lambda(x_1^2 + x_2^2 - 8)$. The first order necessary optimality conditions are

$$\frac{\partial L(x, \lambda)}{\partial x} = \begin{pmatrix} -1 & -2x_1\lambda \\ -1 & -2x_2\lambda \end{pmatrix}^T = 0$$

together with the constraint on the variables, this leads to the unique solution $\overline{x} = (2, 2)^T$, $\overline{\lambda} = -1/4$. The Hessian of the Lagrangian is

$$H_x(L(\overline{x}, \overline{\lambda})) = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$$

which is PD. Hence the point $\overline{x}$ satisfies the sufficient condition for being a local minimum in this problem.

Example 11

Consider the problem

$$\begin{array}{ll} \text{minimize } \theta(x) & = 2x_1^3 + (1/2)x_2^2 + x_1x_2 + (1/24)x_1 \\ \text{subject to} & \quad x_1 + x_2 = 2. \end{array}$$

The Lagrangian is $L(x,\lambda) = 2x_1^3 + (1/2)x_2^2 + x_1x_2 + (1/24)x_1 - \lambda(x_1 + x_2 - 2)$. The first order necessary optimality conditions are

$$\frac{\partial L(x,\lambda)}{\partial x} = \begin{pmatrix} 6x_1^2 + x_2 + (1/24) - \lambda \\ x_2 + x_1 - \quad \lambda \end{pmatrix}^T = 0.$$

Combining this with the constraints on the variables, we have $\lambda = 2$, $6x_1^2 + x_2 + (1/24) - 2 = 6x_1^2 + (2 - x_1) - 2 + (1/24) = 6x_1^2 - x_1 + (1/24) = 0$. This leads to the unique solution satisfying the first order necessary optimality conditions ($\overline{x} = (1/12, 23/12)^T$, $\overline{\lambda} = 2$). The tangent hyperplane at any feasible solution is $\{y : y_1 + y_2 = 0\}$. The Hessian of the Lagrangian is

$$H_x(L(\overline{x},\overline{\lambda})) = \begin{pmatrix} 12\overline{x}_1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

So, on the tangent hyperplane to the feasible region at $\overline{x}$ we have $y^T H_x(L(\overline{x},\overline{\lambda}))y = (y_1 + y_2)^2 = 0$. Thus the second order necessary conditions for a local minimum are also satisfied. However, the point $\overline{x}$ does not satisfy the sufficient conditions for being a local minimum in this problem, (64), discussed above.

Optimality Conditions for the Inequality Constrained Minimization Problems

Consider the general NLP

$$\begin{array}{ll} \text{minimize} & \theta(x) \\ \text{subject to} & h_i(x) = 0,\ i = 1 \text{ to } m \\ & g_p(x) \geqq 0,\ p = 1 \text{ to } t \end{array} \tag{66}$$

where $\theta(x)$, $h_i(x)$, $g_p(x)$ are all real valued continuously differentiable functions defined on $\mathbf{R}^n$. Let $h(x) = (h_1(x), \ldots, h_m(x))^T$ and $g(x) = (g_1(x), \ldots, g_t(x))^T$.

Let $\overline{x}$ be a feasible solution for (66). The active constraints at $\overline{x}$ are all the equality constraints in (66) and all the inequality constraints which hold as equations at $\overline{x}$ (i. e., $g_p(x)$ for p such that $g_p(\overline{x}) = 0$). Let $\mathbf{P}(\overline{x}) = \{p : p = 1 \text{ to } t, g_p(\overline{x}) = 0\}$. The feasible solution $\overline{x}$ is said to be a regular point for (66) if $\{\nabla h_i(\overline{x}) : i = 1 \text{ to } m\} \cup \{\nabla g_p(\overline{x}) : p \in \mathbf{P}(\overline{x})\}$ is linearly independent. This is a **constraint qualification** known as the **regularity condition** for (66). As mentioned earlier, this is a condition on the active constraints at $\overline{x}$, and not on the set of feasible solutions. As an example, consider the system of constraints

$$\begin{array}{ccccc} (x_1-1)^{\mathbf{2}} & + & (x_2-1)^{\mathbf{2}} & = & 0 \\ x_1^{\mathbf{4}} & + & x_2^{\mathbf{4}} & = & 2 \\ x_1 & & & \leqq & 1 \\ & & x_2 & \leqq & 1 \\ x_1 & + & x_2 & \geqq & 2\,. \end{array}$$

This system has the unique solution $(x_1, x_2)^T = (1, 1)^T$, all the constraints are active and it can be verified that the regularity condition does not hold at this point. On the other hand, if this singleton set is represented by the system of constraints

$$\begin{array}{cccc} x_1 & & = & 1 \\ & x_2 & = & 1 \end{array}$$

then the regularity condition holds at the point. Thus, whether regularity conditions hold or not could depend on the system of constraints chosen to represent the set of feasible solutions. This points out the importance of exercising great care in constructing the model for the problem.

Since the inequality constraints "$g_i(x) \geqq 0$" for $i \notin \mathbf{P}(\overline{x})$ are inactive at $\overline{x}$, the local feasible region around $\overline{x}$ remains unchanged if these inactive inequality constraints are ignored. See Figure 20.

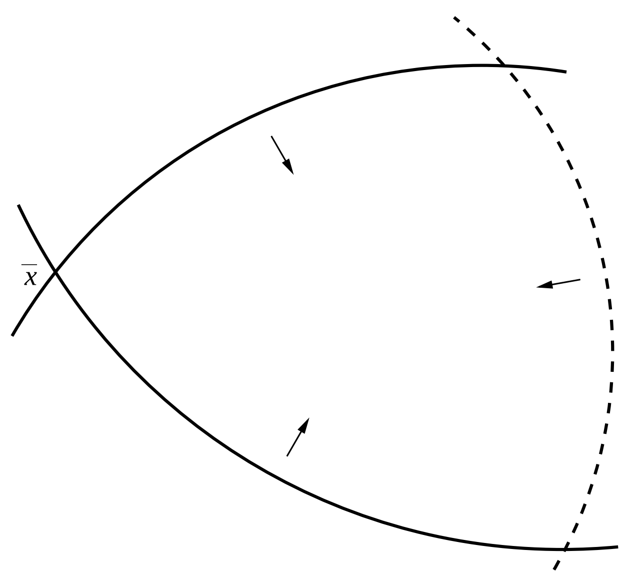

Figure 20 The region which lies on the side of the arrow of each nonlinear surface is the feasible region. The inequality constraint corresponding to the dashed surface is inactive at $\overline{x}$, and it can be ignored for the purpose of deriving optimality conditions for $\overline{x}$ to be a local minimum in the feasible region.

Thus for the purpose of deriving optimality conditions for $\overline{x}$ to be a local minimum for (66), we can ignore the inactive inequality constraints at $\overline{x}$. Also, when all the active constraints at $\overline{x}$ are treated as equality constraints, the local feasible region around $\overline{x}$ becomes smaller, and hence, if $\overline{x}$ is a local minimum for (66), it must be a local minimum for the problem obtained by treating all active constraints at $\overline{x}$ as equality constraints.

Let $\overline{x}$, a feasible regular point for (66), be a local minimum for (66). By the above arguments, it must be a local minimum for the problem,

$$\begin{array}{ll} \text{minimize} & \theta(x) \\ \text{subject to} & h_i(x) = 0,\ i = 1 \text{ to } m \\ & g_p(x) = 0,\ p \in \mathbf{P}(\overline{x}). \end{array} \tag{67}$$

So by previous results, there exists $(\overline{\mu}_1, \ldots, \overline{\mu}_m)$ and $\overline{\pi}_p$ for $p \in \mathbf{P}(\overline{x})$ satisfying

$$\nabla\theta(\overline{x}) - \sum_{i=1}^{m} \overline{\mu}_i \nabla h_i(\overline{x}) - \sum_{p \in \mathbf{P}(\overline{x})} \overline{\pi}_p \nabla g_p(\overline{x}) = 0. \tag{68}$$

We will now prove that if $\overline{x}$ is a local minimum for (66), then $\overline{\pi}_p \geqq 0$ for all $p \in \mathbf{P}(\overline{x})$.

Suppose in (68), $\overline{\pi}_p < 0$ for some $p \in \mathbf{P}(\overline{x})$, say for $p = r$. By the regularity condition, the set $\{\nabla h_i(\overline{x}) : i = 1 \text{ to } m\} \cup \{\nabla g_p(\overline{x}) : p \in \mathbf{P}(\overline{x})\}$ is linearly independent, and by our assumption $r \in \mathbf{P}(\overline{x})$. So there exists a $y \in \mathbf{R}^n$ satisfying

$$\begin{array}{l} (\nabla h_i(\overline{x}))y = 0,\ i = 1 \text{ to } m \\ (\nabla g_p(\overline{x}))y = 0,\ p \in \mathbf{P}(\overline{x}), p \neq r \\ (\nabla g_r(\overline{x}))y = 1. \end{array} \tag{69}$$

By Theorem 20 there exists a differentiable curve $x(\alpha)$ with $x(0) = \overline{x}$, defined for values of α in an interval around $\alpha = 0$, lying on the set of feasible solutions of

$$\begin{array}{l} h_i(x) = 0,\ i = 1 \text{ to } m \\ g_p(x) = 0,\ p \in \mathbf{P}(\overline{x}), p \neq r \end{array} \tag{70}$$

with $\frac{dx(0)}{d\alpha} = y$. Since $(\frac{dg_r(x(\alpha))}{d\alpha})_{\alpha=0} = (\nabla g_r(\overline{x}))y = 1 > 0$, by Taylor's theorem we know that there exists a $\lambda > 0$ such that for all $0 \leqq \alpha \leqq \lambda$, points on the curve $x(\alpha)$ satisfy $g_r(x) \geqq 0$. Using this, it can be verified that when α is positive but sufficiently small, $x(\alpha)$ remains feasible to (66) and since $(\frac{d\theta(x(\alpha))}{d\alpha})_{\alpha=0} = \overline{\pi}_r(\nabla g_r(\overline{x}))y$ (by (68)) < 0, it is a better feasible solution for (66) than $\overline{x}$, contradicting the local minimum property of $\overline{x}$. Thus if $\overline{x}$ is a local minimum for (66) and is a regular point, there must exist $\overline{\mu} = (\overline{\mu}_1, \ldots, \overline{\mu}_m)$, and $\overline{\pi}_p$ for $p \in \mathbf{P}(\overline{x})$ satisfying (68), and $\overline{\pi}_p \geqq 0$ for all $p \in \mathbf{P}(\overline{x})$. Define $\overline{\pi}_p = 0$ for all $p = 1$ to t, $p \notin \mathbf{P}(\overline{x})$ and let $\overline{\pi} = (\overline{\pi}_1, \ldots, \overline{\pi}_t)$. Let $L(x, \mu, \pi) = \theta(x) - \mu h(x) - \pi g(x)$. $L(x, \mu, \pi)$ is the Lagrangian for (66) and $(\overline{\mu}, \overline{\pi})$ are the Lagrange multipliers. These facts imply that if $\overline{x}$ is a regular point local minimum for (66), there exist $\overline{\mu}$, $\overline{\pi}$ satisfying

$$\begin{aligned} \nabla_x L(\overline{x}, \overline{\mu}, \overline{\pi}) &= 0 \\ \overline{\pi} &\geqq 0 \\ \overline{\pi}_p g_p(\overline{x}) &= 0 \text{ for all } p = 1 \text{ to } t \end{aligned} \tag{71}$$

and the feasible conditions

$$h(\overline{x}) = 0, \; g(\overline{x}) \geqq 0.$$

(71) are known as the **first order necessary optimality conditions** for the regular feasible point $\overline{x}$ to be a local minimum for (66). They are also known as the Karush-Kuhn-Tucker (or KKT) necessary conditions for optimality.

Let $\mathbf{T} = \{y : (\nabla h_i(\overline{x}))y = 0, \; i = 1 \text{ to } m, \text{ and } (\nabla g_p(\overline{x}))y = 0, p \in \mathbf{P}(\overline{x})\}$. If all the functions $\theta(x)$, $h_i(x)$, $g_p(x)$ are twice continuously differentiable, and $\overline{x}$ is a regular feasible point for (66), using similar arguments as before, it can be shown that a necessary condition for $\overline{x}$ to be a local minimum for (66) is that there exist Lagrange multiplier vectors $\overline{\mu}$, $\overline{\pi}$ such that

$$\text{(71) holds, and } y^T H_x(L(\overline{x}, \overline{\mu}, \overline{\pi}))y \geqq 0 \text{ for all } y \in \mathbf{T}. \tag{72}$$

(72) are known as **second order necessary conditions** for $\overline{x}$ to be a local minimum for (66).

We now state a sufficient optimality condition for (66) in the form of a theorem.

Theorem 22. *Suppose $\theta(x)$, $h_i(x)$, $g_p(x)$ are all twice continuously differentiable functions, and $\overline{x}$ is a feasible point such that there exists Lagrange multiplier vectors $\overline{\mu} = (\overline{\mu}_1, \ldots, \overline{\mu}_m)$, $\overline{\pi} = (\overline{\pi}_1, \ldots, \overline{\pi}_t)$ which together satisfy*

$$\begin{aligned} &h(\overline{x}) = 0, \; g(\overline{x}) \geqq 0 \\ &\nabla_x L(\overline{x}, \overline{\mu}, \overline{\pi}) = 0 \\ &\overline{\pi} \geqq 0, \; \overline{\pi} g(\overline{x}) = 0 \end{aligned} \tag{73}$$

$$y^T H_x(L(\overline{x}, \overline{\mu}, \overline{\pi}))y > 0, \quad \textit{for all } y \in \mathbf{T}_1, \; y \neq 0$$

where $\mathbf{T}_1 = \{y : (\nabla h_i(\overline{x}))y = 0, \; i = 1$ to m and $(\nabla g_p(\overline{x}))y = 0$ for $p \in \mathbf{P}(\overline{x}) \cap \{p : \overline{\pi}_p > 0\}$, $(\nabla g_p(\overline{x}))y \geqq 0$ for $p \in \mathbf{P}(\overline{x}) \cap \{p : \overline{\pi}_p = 0\}\}$, then $\overline{x}$ is a local minimum for (66).

Proof. Suppose $\overline{x}$ is not a local minimum for (66). As in the proof of Theorem 21, there must exist a sequence of distinct feasible solutions $x^r = \overline{x} + \delta_r y^r$, $r = 1, 2, \ldots$ converging to $\overline{x}$ as $r \to 0^+$, where $\|y^r\| = 1$ for all r; $y^r \to \overline{y}$ and $\delta_r \to 0^+$; such that $\theta(x^r) < \theta(\overline{x})$ for all r. By feasibility, as in the proof of Theorem 21, we have

$$(\nabla h_i(\overline{x}))\overline{y} = 0, \; i = 1 \text{ to } m. \tag{74}$$

For each $p \in \mathbf{P}(\overline{x})$, we have $g_p(\overline{x}) = 0$, and $g_p(x^r) \geqq 0$ by feasibility. So

$$0 \leqq g_p(\overline{x} + \delta_r y^r) - g_p(\overline{x}) = \delta_r(\nabla g_p(\overline{x}))y^r + 0(\delta_r)$$

Dividing by $\delta_r > 0$, and taking the limit as $r \to \infty$, we conclude that

$$(\nabla g_p(\overline{x}))\overline{y} \geqq 0 \text{ for all } p \in \mathbf{P}(\overline{x}). \tag{75}$$

Also, $0 > \theta(\overline{x}+\delta_r y^r) - \theta(\overline{x}) = \delta_r(\nabla\theta(\overline{x}))y^r + 0(\delta_r)$, and again dividing by $\delta_r > 0$, and taking the limit as $r \to \infty$ we conclude that $(\nabla\theta(\overline{x}))\overline{y} \leqq 0$.

Suppose $(\nabla g_p(\overline{x}))\overline{y} > 0$ for some $p \in \mathbf{J} = \{p : \overline{\pi}_p > 0\}$. Then

$$\begin{aligned} 0 \geqq (\nabla\theta(\overline{x}))\overline{y} &= \overline{\mu}(\nabla h(\overline{x}))\overline{y} + \overline{\pi}(\nabla g(\overline{x}))\overline{y}, \text{ by (73)} \\ &= \sum_{p\in\mathbf{J}} \overline{\pi}_p(\nabla g_p(\overline{x}))\overline{y}, \text{ by (73), (74).} \\ &> 0, \text{ by (73), (75) and the assumption that} \\ &\quad (\nabla g_p(\overline{x}))\overline{y} > 0 \text{ for some } p \in \mathbf{J} \end{aligned}$$

a contradiction. So $\overline{y}$ satisfies

$$(\nabla g_p(\overline{x}))\overline{y} = 0 \text{ for all } p \in \mathbf{P}(\overline{x}) \cap \{p : \overline{\pi}_p > 0\}. \tag{76}$$

By (74), (75), (76), we see that $\overline{y} \in \mathbf{T}_1$. From (73) and feasibility we have

$$\begin{aligned} &\theta(x^r) - \theta(\overline{x}) = L(\overline{x}+\delta_r y^r, \overline{\mu}, \overline{\pi}) - L(\overline{x}, \overline{\mu}, \overline{\pi}) = \\ &(1/2)\delta_r^{\mathbf{2}}(y^r)^T H_x(L(\overline{x}+\alpha_r y^r, \overline{\mu}, \overline{\pi}))y^r \end{aligned} \tag{77}$$

where $0 \leqq \alpha_r \leqq \delta_r$, by using (73) on the expression given by Taylor's theorem. When r is sufficiently large, from the continuity, and the conditions satisfied by $\overline{y}$ proved above, and (73), we conclude that the right-hand side of (77) is $\geqq 0$, while $\theta(x^r) - \theta(\overline{x})$ is < 0, a contradiction. So, $\overline{x}$ must be local minimum for (66).

□

Thus (73) provides a **sufficient local minimality condition** for (66). See references [A8, A10, 10.2, 10.12, 10.13, 10.17, 10.27] for a complete discussion of optimality conditions for nonlinear programs.

In inequality constrained problems, we notice that the gap between known second order necessary optimality conditions and sufficient optimality conditions, is quite wide.

The NLP (66) is said to be a **convex programming problem** if $\theta(x)$ is convex, $h_i(x)$ is affine for all i, and $g_p(x)$ is concave for all p. In this case the set of feasible solutions is a convex set. For convex programming problems, we will now show that (71) are both **necessary and sufficient conditions for global optimality**.

Theorem 23. *Suppose (66) is a convex program. The feasible regular point $\overline{x}$ is a global minimum for (66) iff there exists a Lagrange multiplier vector $(\overline{\mu}, \overline{\pi})$ such that $\overline{x}$, $\overline{\mu}$, $\overline{\pi}$ together satisfy (71).*

Proof. The necessity of (71) for optimality has already been established above. We will now prove the sufficiency. Suppose $\overline{x}$ is a feasible solution of (66) satisfying (71).

Let x be any other feasible solution for (66). By Theorem 15

$$\begin{aligned}
\theta(x) - \theta(\overline{x}) &\geqq (\nabla\theta(\overline{x}))(x - \overline{x}) \\
&= \Big(\sum_{i=1}^{m} \overline{\mu}_i \nabla h_i(\overline{x}) + \sum_{p\in\mathbf{P}(\overline{x})} \overline{\pi}_p \nabla g_p(\overline{x})\Big)(x - \overline{x}) \text{ by (71)} \\
&= \sum_{p\in\mathbf{P}(\overline{x})} \overline{\pi}_p \nabla g_p(\overline{x})(x - \overline{x}), \text{ since } h(x) \text{ is affine} \\
&\geqq \sum_{p\in\mathbf{P}(\overline{x})} \overline{\pi}_p (g_p(x) - g_p(\overline{x})) \text{ by Theorem 16, since } g_p(x) \text{ is concave.} \\
&= \sum_{p\in\mathbf{P}(\overline{x})} \overline{\pi}_p g_p(x), \text{ since } g_p(\overline{x}) = 0 \text{ for } p \in \mathbf{P}(\overline{x}). \\
&\geqq 0, \text{ since } \overline{\pi} \geqq 0 \text{ and } g(x) \geqq 0 \text{ for feasibility.}
\end{aligned}$$

So $\overline{x}$ is a global minimum for (66).

∎

Example 12

Consider the problem of determining the electrical current flows in the following electrical network.

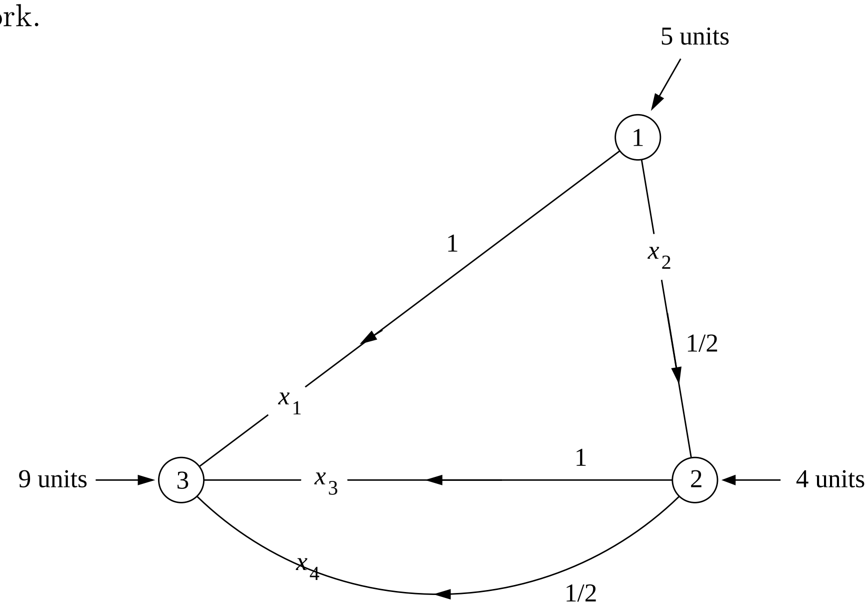

Figure 21

Assume that the current flows on each arc in the direction indicated. A total of 5, 4 units of current enters the system at nodes 1, 2 respectively per unit time. The numbers given on the arcs are the resistences of the arcs. Let x_1, x_2, x_3, x_4 denote the current flows on the arcs as indicated. If r_j denotes the resistence associated with x_j it is known that the power loss is $\sum_{j=1}^{4} r_j x_j^{\mathbf{2}}$. It is required to find out the current flows, under the assumption that the flows would occur so as to minimize the power loss. Hence the x-vector is the optimum solution of the problem

$$\begin{array}{lllllll} \text{minimize} & x_1^2 & +(1/2)x_2^2 & +x_3^2 & +(1/2)x_4^2 & & \\ \text{subject to} & x_1 & +x_2 & & & = & 5 \\ & & -x_2 & +x_3 & +x_4 & = & 4 \\ & & x_j \geqq 0, & j = 1 \text{ to } 4. & & & \end{array} \tag{78}$$

So, the Lagrangian is $L(x, \mu, \pi) = x_1^2 + (1/2)x_2^2 + x_3^2 + (1/2)x_4^2 - \mu_1(x_1 + x_2 - 5) - \mu_2(-x_2 + x_3 + x_4 - 4) - \sum_{j=1}^{4} \pi_j x_j$.

So, the first order necessary optimality conditions are

$$\begin{aligned} \frac{\partial L}{\partial x_1} &= 2x_1 - \mu_1 - \pi_1 = 0 \\ \frac{\partial L}{\partial x_2} &= x_2 - \mu_1 + \mu_2 - \pi_2 = 0 \\ \frac{\partial L}{\partial x_3} &= 2x_3 - \mu_2 - \pi_3 = 0 \\ \frac{\partial L}{\partial x_4} &= x_4 - \mu_2 - \pi_4 = 0 \end{aligned} \tag{79}$$

$$\pi_1, \pi_2, \pi_3, \pi_4 \geqq 0 \tag{80}$$

$$\pi_1 x_1 = \pi_2 x_2 = \pi_3 x_3 = \pi_4 x_4 = 0 \tag{81}$$

and the constraints (78) on the x-variables for feasibility.

The complementary slackness conditions (81) imply that for each j, either the Lagrange multiplier π_j is zero, or the inequality constraint $x_j \geqq 0$ holds as an equality constraint (i. e., it is active) at the optimum. One technique to find a solution to the first order necessary optimality conditions here is to guess the subset of inequality constraints in (78) which will be active at the optimum, called the **active set**. Treat each of the inequality constraints in (78) in this active set as an equation, ignore the inequality constraints in (78) outside the active set (we are assuming that they will be inactive at the optimum). Set the Lagrange multiplier π_j corresponding to each inequality constraint in (78), not in the active set to zero. What remains among (78), (79) is a system of equations, which is solved. If the solution of this system satisfies (80) and the ignored inequality constraints in (78) not in the active set, we are done, this solution solves the first order necessary optimality conditions. If some of these conditions are violated, repeat this process with a different active set. This process, therefore, involves a combinatorial search, which may eventually involve solving 2^t systems where t is the number of inequality constraints in the original NLP ($t = 4$ here), not efficient if t is large. Efficient algorithms for solving NLP's involving inequality constraints either carry out this combinatorial search very efficiently; or do not use it at all, but operate with other efficient methods to find a solution to the first order necessary optimality conditions (see Chapters 2, 10).

We first try treating the inequality constraint $x_3 \geqq 0$ as active, and all the other inequality constraints $x_j \geqq 0$, $j = 1, 2, 4$ as inactive. Ignoring these inactive inequality constraints, and setting $\pi_j = 0$, $j = 1, 2, 4$ leads to the system of equations:

$$
\begin{array}{rrrrrl}
x_1 & +x_2 & & & & = 5 \\
 & -x_2 & +x_4 & & & = 4 \\
2x_1 & & & -\mu_1 & & = 0 \\
 & x_2 & & -\mu_1 & -\mu_2 & = 0 \\
 & & & & -\mu_2 \quad -\pi_3 & = 0 \\
 & & x_4 & & -\mu_2 & = 0.
\end{array}
$$

This system has the unique solution $(x_1, x_2, x_4) = (-2, 7, 11)$, $(\mu_1, \mu_2) = (-4, 11)$, $\pi_3 = -11$. This solution violates the constraints "$x_1 \geqq 0$, $\pi_3 \geqq 0$", so this choice of active set did not lead to a solution of the first order necessary optimality conditions in this problem.

Let us now try treating all the constraints "$x_j \geqq 0$, $j = 1$ to 4" as inactive. Ignoring all these inactive constraints, and setting $\pi_j = 0$, $j = 1$ to 4 leads to the system of equations

$$
\begin{array}{rrrrrrl}
2x_1 & & & & -\mu_1 & & = 0 \\
 & x_2 & & & -\mu_1 & +\mu_2 & = 0 \\
 & & 2x_3 & & & -\mu_2 & = 0 \\
 & & & x_4 & & -\mu_2 & = 0 \\
x_1 & +x_2 & & & & & = 5 \\
 & -x_2 & +x_3 & +x_4 & & & = 4.
\end{array}
$$

This system has the unique solution $\overline{x} = (3, 2, 2, 4)^T$, $\overline{\mu} = (6, 4)$. This solution also satisfies the inequality constraints, on the x_j which were ignored. So $(\overline{x}, \overline{\mu}, \overline{\pi} = 0)$ satisfies the first order necessary optimality conditions for this problem. It can be verified that $\overline{x}$ also satisfies the second order necessary optimality conditions, as well as the sufficient conditions for being a local minimum for this problem. Since $\theta(x)$ is convex here, $\overline{x}$ is in fact a global minimum for this problem.

Optimality Conditions for Linearly Constrained Optimization Problems

Consider the nonlinear program,

$$
\begin{array}{ll}
\text{minimize} & \theta(x) \\
\text{subject to} & Ax = b \\
 & Dx \geqq d
\end{array}
\tag{82}
$$

where A, b, D, d are given matrices of orders $m \times n$, $m \times 1$, $t \times n$ and $t \times 1$ respectively, and $\theta(x)$ is continuously differentiable. Since the constraints are linear, for this problem, we can establish first order necessary optimality conditions of the form in (71) without requiring a regularity type of constraint qualification.

Theorem 24. *If $\overline{x}$ is a local minimum for (82), there exist Lagrange multiplier vectors $\overline{\mu} = (\overline{\mu}_1, \ldots, \overline{\mu}_m)$, $\overline{\pi} = (\overline{\pi}_1, \ldots, \overline{\pi}_t)$ such that*

$$\begin{aligned} \nabla_x L(\overline{x}, \overline{\mu}, \overline{\pi}) = \nabla\theta(\overline{x}) - \overline{\mu}A - \overline{\pi}D = 0 \\ A\overline{x} = b,\ D\overline{x} \geqq d \\ \overline{\pi} \geqq 0,\ \overline{\pi}(D\overline{x} - d) = 0 \end{aligned} \tag{83}$$

where $L(x, \mu, \pi) = \theta(x) - \mu(Ax - b) - \pi(Dx - d)$ is the Lagrangian for (82).

Proof. Let $\mathbf{P}(\overline{x}) = \{p : 1 \leqq p \leqq t \text{ and } D_{p.}\overline{x} = d_p\}$, it is the index set of active inequality constraints in (82) at the feasible point $\overline{x}$. Since the constraints are linear, the tangent plane to the system determined by the active constraints in (82) at $\overline{x}$ is

$$\mathbf{T} = \{y : A_{i.}y = 0,\ i = 1 \text{ to } m,\ \text{and } D_{p.}y = 0,\ p \in \mathbf{P}(\overline{x})\} \tag{84}$$

whether $\overline{x}$ satisfies the regularity condition for (82) or not. Let

$$\begin{aligned} \mathbf{\Gamma}_1 &= \{y : y \in \mathbf{R}^n,\ A_{i.}y = 0,\ i = 1 \text{ to } m \text{ and } D_{p.}y \geqq 0 \ \text{ for all } p \in \mathbf{P}(\overline{x})\} \\ \mathbf{\Gamma}_2 &= \{y : y \in \mathbf{R}^n,\ (\nabla\theta(\overline{x}))y < 0\}. \end{aligned}$$

We will now show that the fact that $\overline{x}$ is a local minimum for (82) implies that $\mathbf{\Gamma}_1 \cap \mathbf{\Gamma}_2 = \emptyset$. Suppose not. Let $\overline{y} \in \mathbf{\Gamma}_1 \cap \mathbf{\Gamma}_2$. Since $D_{p.}\overline{x} > d_p$ for $p \notin \mathbf{P}(\overline{x})$, and $\overline{y} \in \mathbf{\Gamma}_1$, it can be verified that $\overline{x} + \alpha\overline{y}$ is feasible to (82) when α is positive and sufficiently small, and since $\overline{y} \in \mathbf{\Gamma}_2$, we have $\theta(\overline{x} + \alpha\overline{y}) < \theta(\overline{x})$, contradicting the local minimality of $\overline{x}$ to (82). So $\mathbf{\Gamma}_1 \cap \mathbf{\Gamma}_2 = \emptyset$.

$\mathbf{\Gamma}_1 \cap \mathbf{\Gamma}_2 = \emptyset$ implies by Farkas' theorem (Theorem 3 of Appendix 1) that there exist $\overline{\mu} = (\overline{\mu}_1, \ldots, \overline{\mu}_m)$ and $\overline{\pi}_p$ for $p \in \mathbf{P}(\overline{x})$ satisfying

$$\begin{aligned} \nabla\theta(\overline{x}) = \sum_{i=1}^{m} \overline{\mu}_i A_{i.} + \sum_{p \in \mathbf{P}(\overline{x})} \overline{\pi}_p D_{p.} \\ \overline{\pi}_p \geqq 0 \text{ for all } p \in \mathbf{P}(\overline{x}). \end{aligned}$$

Now define $\overline{\pi}_p = 0$ for $p \notin \mathbf{P}(\overline{x})$, and let $\overline{\pi} = (\overline{\pi}_1, \ldots, \overline{\pi}_t)$. From the above, we verify that $\overline{x}$, $\overline{\mu}$, $\overline{\pi}$ together satisfy (83).

□

The conditions (83) are the **first order necessary optimality conditions** for the linearly constrained optimization problem (82).

If $\theta(x)$ is twice continuously differentiable in (82), since the constraints are linear in (82), it can be verified that the Hessian matrix of the Lagrangian is the same as the Hessian matrix of $\theta(x)$. Using the Taylor series approximation up to the second order, it can be shown that if $\overline{x}$ is a local minimum for (82), there must exist Lagrange multiplier vectors $\overline{\mu}$, $\overline{\pi}$ such that

$$\text{(83) holds and } y^T H(\theta(\overline{x}))y \geqq 0 \text{ for all } y \in \mathbf{T} \text{ of (84).} \tag{85}$$

The conditions (85) correspond to (72), they are the **second order necessary optimality conditions** for (82).

5. Summary of Some Optimality Conditions

All the functions (objective and constraint function) are assumed to be continuously differentiable. They are assumed to be twice continuously differentiable, if the Hessian matrix, appears in the expressions.

Problem	necessary optimality conditions for point $\bar{x}$ to be a local minimum	sufficient optimality conditions for point $\bar{x}$ to be a local miminum
Unconstrained minimization. minimize $\theta(x)$ over $x \in \mathbf{R}^n$	<u>First order conditions</u> $\nabla\theta(\bar{x}) = 0$ <u>Second order conditions</u> $\nabla\theta(\bar{x}) = 0$ and $H(\theta(\bar{x}))$ is PSD.	$\nabla\theta(\bar{x}) = 0$ and $H(\theta(\bar{x}))$ is PD.

Problem	necessary optimality conditions for point $\bar{x}$ to be a local minimum	sufficient optimality conditions for point $\bar{x}$ to be a local miminum
Equality constrained minimization. minimize $\theta(x)$ subject to $h_i(x) = 0$, $i = 1$ to m.	Denote $h(x)=(h_1(x), \dots, h_m(x))^T$. $\mu=(\mu_1, \dots, \mu_m)$ is the row vector of Lagrange multipliers. The Largrangian is $L(x,\mu)=\theta(x)-\mu h(x)=\theta(x)-\sum_{i=1}^{m} \mu_i h_i(x)$.	
	Conditions given here hold under the constraint qualifications that $\bar{x}$ is a regular point (i.e., $\{\nabla h_i(\bar{x})$: $i = 1$ to $m\}$ is a linearly independent set) or under weaker constraint qualifications, for which see references [A8, A10, A12, 10.12, 10.13, 10.26], or if the constraints are all linear. <u>First order conditions</u> Feasibility, $h(\bar{x})=0$ and there exists a Lagrange multiplier vector $\bar{\mu}$, which together with $\bar{x}$ satisfies $\nabla_x L(\bar{x},\bar{\mu}) = \nabla\theta(\bar{x})-\bar{\mu}\nabla h(\bar{x}) = 0$ (i.e., objective gradient is a linear combination of the constraint gradients). <u>Second order conditions</u> $h(\bar{x})=0$ and there exists a Lagrange multiplier vector $\bar{\mu}$, which together with $\bar{x}$ satisfies $\nabla_x(L(\bar{x},\bar{\mu})) = 0$ and $y^T H_x(L(\bar{x},\bar{\mu}))y \geqq 0$	Feasibility, $h(\bar{x})=0$ and there exists a Lagrange multiplier vector $\bar{\mu}$, which together with $\bar{x}$ satisfies $\nabla_x(L(\bar{x},\bar{\mu})) = 0$ and $y^T H_x(L(\bar{x},\bar{\mu}))y > 0$ for all $y \in \{y$: $(\nabla h_i(\bar{x}))y = 0$, $i = 1$ to m, $y \neq 0\}$.
	for all $y \in \{y$: $(\nabla_x h_i(\bar{x}))y = 0$, i=1 to $m\}$.	

Problem	necessary optimality conditions for point $\bar{x}$ to be a local minimum	sufficient optimality conditions for point $\bar{x}$ to be a local miminum
Constrained minimization. minimize $\theta(x)$ subject to $h_i(x)=0$, $i=1$ to m; $g_p(x) \geqq 0$, $p=1$ to t.	Let $h(x)=(h_1(x), \dots, h_m(x))^T$. $g(x) = (g_1(x), \dots, g_t(x))^T$. Let $\mu=(\mu_1, \dots, \mu_m)$, $\pi=(\pi_1, \dots, \pi_t)$ be row vectors of Lagrange multipliers. The Largrangian is $L(x,\mu,\pi)= \theta(x)-\mu h(x)-\pi g(x)=\theta(x) - \sum_{i=1}^{m} \mu_i h_i(x)-\sum_{p=1}^{t} \pi_p g_p(x)$.	
	Conditions given here hold under the constraint qualification that $\bar{x}$ is a regular point (i.e., if $P(\bar{x})=\{p\colon g_p(\bar{x})=0\}$, then $\{\nabla h_i(\bar{x})\colon i=1 \text{ to } m\}\cup\{\nabla g_p(\bar{x})\colon p\in P(\bar{x})\}$ is a linearly independent set), or under weaker constraint qualifications, for which see references cited above, or if the constraints are all linear. <u>First order conditions (KKT conditions)</u> Feasibility, $h(\bar{x})=0$, $g(\bar{x}) \geqq 0$ and there exists Lagrange multiplier vectors $\bar{\mu}$, $\bar{\pi}$ which together with $\bar{x}$ satisfy $\nabla_x L(\bar{x},\bar{\mu},\bar{\pi})=\nabla\theta(\bar{x})-\sum_{i=1}^{m} \bar{\mu}_i \nabla h_i(\bar{x}) - \sum_{p=1}^{t} \bar{\pi}_p \nabla g_p(\bar{x}) = 0$ $\bar{\pi} \geqq 0$ $\bar{\pi}_p g_p(\bar{x}) = 0$ for all p. The last set of conditions here are the **complementary slackness conditions.** It is equivalent to $\bar{\pi} g(\bar{x})=0$. <u>Second order conditions</u> $h(\bar{x})=0$, $g(\bar{x}) \geqq 0$ and there exists Lagrange multiplier vectors $\bar{\mu}$, $\bar{\pi}$, which together with $\bar{x}$ satisfy $\nabla_x L(\bar{x},\bar{\mu},\bar{\pi}) = 0$ $\bar{\pi} \geqq 0$ $\bar{\pi} g(\bar{x}) = 0$ and $y^T H_x(L(\bar{x},\bar{\mu},\bar{\pi}))y \geqq 0$ for all $y \in \{y\colon (\nabla h_i(\bar{x}))y=0,\ i=1 \text{ to } m \text{ and } (\nabla g_p(\bar{x}))y=0,\ p\in P(\bar{x})\}$.	Feasibility, $h(\bar{x}) = 0$ $g(\bar{x}) \geqq 0$ and there exists Lagrange multiplier vectors $\bar{\mu}$, $\bar{\pi}$ which together with $\bar{x}$ satisfy $\nabla_x L(\bar{x},\bar{\mu},\bar{\pi}) = 0$ $\bar{\pi} \geqq 0$ $\bar{\pi} g(\bar{x}) = 0$ and $y^T H_x(L(\bar{x},\bar{\mu},\bar{\pi}))y>0$ for all $y\in T_1=\{y\colon (\nabla h_i(\bar{x}))y=0,\ i=1 \text{ to } m, \text{ and } (\nabla g_p(\bar{x}))y=0 \text{ for } p\in P(\bar{x})\cap\{p\colon \bar{\pi}_p>0\} \text{ and } (\nabla g_p(\bar{x}))y\geqq 0, \text{ for } p\in P(\bar{x})\cap\{p\colon \bar{\pi}_p=0\},\ y\neq 0$. For convex programming problem, that is, when $\theta(x)$ is convex, $h(x)$ is affine and $g(x)$ is concave, the First order conditions (KKT conditions) are both necessary and sufficient for $\bar{x}$ to be a global minimum.

6. Exercises

---

13. Consider the quadratic program

$$\begin{array}{ll} \text{minimize} & cx + (1/2)x^T Dx \\ \text{subject to} & Ax \geqq b \\ & x \geqq 0 \end{array}$$

where D is a general symmetric matrix of order n. Prove that the necessary and sufficient conditions for x^* to be a local minimum to this general quadratic program is that there exist vectors y^*, u^*, v^*, such that

$$\begin{pmatrix} u^* \\ v^* \end{pmatrix} = \begin{pmatrix} D & -A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x^* \\ y^* \end{pmatrix} + \begin{pmatrix} c^T \\ -b \end{pmatrix}$$

$$\begin{pmatrix} u^* \\ v^* \end{pmatrix} \geqq 0, \ \begin{pmatrix} x^* \\ y^* \end{pmatrix} \geqq 0, \ \begin{pmatrix} u^* \\ v^* \end{pmatrix}^T \begin{pmatrix} x^* \\ y^* \end{pmatrix} = 0$$

hold, and for every vector $\xi \in \mathbf{R}^n$ satisfying

$$\begin{aligned} A_{i.}\xi &= 0 \text{ if } y_i^* > 0 \\ A_{i.}\xi &\geqq 0 \text{ if } v_i^* = y_i^* = 0 \\ \xi_j &= 0 \text{ if } u_j^* > 0 \\ \xi_j &\geqq 0 \text{ if } x_j^* = u_j^* = 0 \end{aligned}$$

we have $\xi^T D\xi \geqq 0$. (A. Majthay [A9])

14. Consider the quadratic programming problem

$$\begin{array}{ll} \text{minimize} & cx + (1/2)x^T Dx \\ \text{subject to} & 0 \leqq x \leqq u \end{array}$$

where

$$D = \begin{pmatrix} -2 & -3 & -3 \\ -3 & -5 & -1 \\ -3 & -1 & -4 \end{pmatrix}, \ c = \begin{pmatrix} 4 \\ 3 \\ 5 \end{pmatrix}^T, \ u = \begin{pmatrix} 10 \\ 10 \\ 10 \end{pmatrix}$$

and identify the global optimum solution of this problem. (W. P. Hallman and I. Kaneko [2.15])

15. Let $f(x)$ be a real valued differentiable function defined on $\mathbf{R}^1$. Let $x^0 \in \mathbf{R}^1$. Is the following statement true? "For x^0 to be a local minimum for $f(x)$ in $\mathbf{R}^1$, it is necessary that the derivative $f'(x^0) = 0$; and there must exist an open interval (a, b) around x^0 such that $f'(x) < 0$ for all x in the open interval (a, x^0), and $f'(x) > 0$ for all x in the open interval (x^0, b)". Is this condition sufficient for x^0 to be a local minimum of $f(x)$? Use the function defined by

$$\begin{aligned} f(x) &= x^{\mathbf{2}}(2 + \sin(1/x)), \text{ when } x \neq 0 \\ f(0) &= 0 \end{aligned}$$

and $x^0 = 0$, as an example. (K. Sydsaeter [A14])

16. Let $f(x)$ be a real valued function defined on $\mathbf{R}^1$. Let $x^0 \in \mathbf{R}^1$, and suppose $f(x)$ has continuous nth derivative. A sufficient condition for x^0 to be a strict local minimum for $f(x)$ in $\mathbf{R}^1$, is that $f^{(1)}(x^0) = f^{(2)}(x^0) = \ldots = f^{(n-1)}(x^0) = 0$, and $f^{(n)}(x^0) > 0$ for n even, where, $f^{(r)}(x^0)$ is the rth derivative of $f(x)$ at x^0. Is this condition necessary for x^0 to be a local minimum for $f(x)$? Use the function defined by

$$\begin{aligned} f(x) &= e^{-(\mathbf{1/x^2})}, x \neq 0 \\ f(0) &= 0 \end{aligned}$$

and $x^0 = 0$, as an example. (K. Sydsaeter [A14])

17. It is sometimes stated that minimizing a function subject to constraints is equivalent to finding the unconstrained minimum of the Lagrangian function. Examine whether this statement is true, using the example

$$\begin{array}{ll} \text{minimize} & -x_1 x_2 \\ \text{subject to} & \ x_1 + x_2 = 2 \end{array}$$

and the point $\overline{x} = (1, 1)^T$ which is optimal for it. (K. Sydsaeter [A14])

18. Consider the equality constrained optimization problem (56) and the Lagrangian $L(x, \mu) = \theta(x) - \mu h(x)$ for it. If $(\overline{x}, \overline{\mu})$ is an unconstrained local minimum for $L(x, \mu)$ over $x \in \mathbf{R}^n$, $\mu \in \mathbf{R}^m$, prove that the point $\overline{x}$ must be feasible to (56) and in fact it must be a local minimum for (56). However, show that the converse may not be true, that is, even if $\hat{x}$ is a local minimum for (56), there may not exist a Lagrange multiplier vector $\hat{\mu}$ such that $(\hat{x}, \hat{\mu})$ is an unconstrained local minimum for $L(x, \mu)$. See Exercise 17 above. Develop general conditions on the NLP (56) and the point $\overline{x}$ which can guarantee that if $\overline{x}$ is a local minimum for (56), there exists a Lagrange multiplier vector $\overline{\mu}$ such that $(\overline{x}, \overline{\mu})$ is a local minimum for $L(x, \mu)$ over $x \in \mathbf{R}^n$, $\mu \in \mathbf{R}^m$.

19. Consider the NLP (66) and the Lagrangian $L(x, \mu, \pi) = \theta(x) - \mu h(x) - \pi g(x)$. If $(\overline{x}, \overline{\mu}, \overline{\pi})$ is a local minimum for the problem

$$\begin{array}{ll} \text{minimize} & L(x, \mu, \pi) \\ \text{subject to} & x \in \mathbf{R}^n, \mu \in \mathbf{R}^m \\ \text{and} & \pi \geqq 0 \end{array} \tag{86}$$

prove that $\overline{x}$ must be feasible to (66) and in fact must be a local minimum for (66). However, show that the converse may not be true, that is even if $\hat{x}$ is a local minimum for (66), there may not exist a $\hat{\mu} \in \mathbf{R}^m$ and $\hat{\pi} \in \mathbf{R}^t$, $\hat{\pi} \geqq 0$, such that $(\hat{x}, \hat{\mu}, \hat{\pi})$ is a local minimum for (86).

Develop general conditions on the NLP (66) and the point $\overline{x}$, which can guarantee that if $\overline{x}$ is a local minimum for (66), there exist Lagrange multiplier vectors $\overline{\mu}$, $\overline{\pi}$ such that $(\overline{x}, \overline{\mu}, \overline{\pi})$ is a local minimum for (86).

20. Let $\theta(x)$ be a real valued function defined on $\mathbf{R}^n$ and let $\overline{x} \in \mathbf{R}^n$. Examine the following statement "If $\overline{x}$ is a local minimum along each straight line through $\overline{x}$ in $\mathbf{R}^n$, then $\overline{x}$ is a local minimum for $\theta(x)$ in $\mathbf{R}^n$", and mark whether it is true or false. Use $\theta(x_1, x_2) = (x_2 - x_1^{\mathbf{2}})(x_2 - 2x_1^{\mathbf{2}})$ defined on $\mathbf{R}^2$ and $\overline{x} = (0, 0)^T$ as an example. (K. Sydsaeter [A14])

21. Let A, D be given PD matrices of order n. Solve the following two optimization problems.

$$\begin{array}{lll} \text{(i)} & \text{minimize} & cx \\ & \text{subject to} & (x - \overline{x})^T A (x - \overline{x}) \leqq 1 \\ \text{(ii)} & \text{minimize} & cx + (1/2) x^T D x \\ & \text{subject to} & (x - \overline{x})^T A (x - \overline{x}) \leqq 1. \end{array}$$

Discuss what happens if A is PD but D is either PSD or not even PSD.

22. Consider the following quadratic programming problem

$$\begin{aligned} f(b) &= \text{minimum value of } Q(x) = cx + (1/2) x^T D x \\ &\quad \text{subject to } Ax \geqq b \\ &\quad\ \ x \geqq 0 \end{aligned}$$

where D is a symmetric PSD matrix of order n, $f(b)$ denotes the optimum objective value in this problem as a function of the vector b, and A, b are given matrices of orders $m \times n$ and $m \times 1$ respectively. In this problem, assume that A, c, d remain fixed, but b may vary.

(i) If $f(b)$ is finite for some b, prove that $f(b)$ is finite for all b for which the problem is feasible.

(ii) If $f(b)$ is finite for some b, prove that $f(b)$ is convex over $b \in \mathrm{Pos}(A, -I_m)$.

(iii) What is $\partial f(b)$?

Note: The result in (i) above could be false if D is not PSD. Consider the following example from B. C. Eaves [2.9]

$$\begin{array}{lrl} \text{minimize } Q(x) = & -4x_1 + x_1^2 - x_2^2 & \\ \text{subject to} & -x_1 + x_2 & \geqq b_1 \\ & -x_1 + x_2 & \geqq b_2 \\ & x_1, x_2 & \geqq 0. \end{array}$$

Let $b = (b_1, b_2)^T$. If $b = b^1 = (-2, -4)$, or if $b = b^2 = (-4, -2)$, verify that the problem is feasible and that $Q(x)$ is bounded below on the set of feasible solution. If $b = (b^1 + b^2)/2 = (-3, -3)^T$, verify that $Q(x)$ becomes unbounded below on the set of feasible solutions.

23. Let $\mathbf{K} \subset \mathbf{R}^n$ be a closed convex set. For $x \in \mathbf{R}^n$, define

$$f(x) = \text{ Minimum } \{\|y - x\| : y \in \mathbf{K}\}.$$

Prove that $f(x)$ is convex.

24. Let $\theta(x) = (2x_2 - x_1^2)^2$. Check whether $\theta(x)$ is convex, or concave, or neither, on $-1 \leqq x_1 \leqq 1$, $-1 \leqq x_2 \leqq 1$.

25. Consider the linear program in standard form

$$\begin{array}{ll} \text{minimize} & cx \\ \text{subject to} & Ax = b \\ & x \geqq 0. \end{array}$$

This problem can be written as the following NLP in which the constraints are all equalities, but there are new variables u_j.

$$\begin{array}{ll} \text{minimize} & cx \\ \text{subject to} & Ax = b \\ & u_j^2 - x_j = 0, \text{ for all } j. \end{array}$$

Write down the necessary optimality conditions for this equality constrained NLP, and show that they are equivalent to the duality-complementary slackness conditions for optimality in the above LP.

26. Consider the NLP

$$\begin{array}{ll} \text{minimize} & \theta(x) \\ \text{subject to} & h_i(x) = 0, \ i = 1 \text{ to } m \\ & g_p(x) \geqq 0, \ p = 1 \text{ to } t \end{array}$$

where all the functions $\theta(x)$, $h_i(x)$, $g_p(x)$ are continuously differentiable. If $\overline{x}$ is a local minimum for this problem, and

(a) $\{\nabla h_i(\overline{x}) : i = 1 \text{ to } m\}$ is linearly independent,

(b) there exists a $y \in \mathbf{R}^n$ satisfying

$$\nabla h_i(\overline{x})y = 0,\ i = 1 \text{ to } m$$
$$\nabla g_p(\overline{x})y > 0,\ p \in \mathbf{P}(\overline{x})$$

where $\mathbf{P}(\overline{x}) = \{p : g_p(\overline{x}) = 0\}$.

Prove that there must exist $\overline{\mu} = (\overline{\mu}_1, \ldots, \overline{\mu}_m)$, $\overline{\pi} = (\overline{\pi}_1, \ldots, \overline{\pi}_t)$ such that

$$\nabla\theta(\overline{x}) - \overline{\mu}\nabla h(\overline{x}) - \overline{\pi}\nabla g(\overline{x}) = 0$$
$$\overline{\pi} \geqq 0 \text{ and } \overline{\pi}_p g_p(\overline{x}) = 0 \text{ for all } p = 1 \text{ to } t.$$

27. Consider the NLP

$$\begin{array}{ll} \text{minimize} & \theta(x) = x_1^2 + x_2^2 \\ \text{subject to} & (x_1 - 1)^3 - x_2^2 = 0. \end{array}$$

i) If $\overline{x} = (\overline{x}_1, \overline{x}_2)^T$ is a feasible solution to this problem, prove that $\overline{x}_1$ must be $\geqq 1$. Using this information, prove that $\hat{x} = (1, 0)^T$ is the global minimum for this problem.

ii) Write down the first order necessary optimality conditions for this problem. Does $\hat{x}$ satisfy these conditions? Why? Explain clearly. (R. Fletcher [10.13])

28. Consider the NLP

$$\begin{array}{lrl} \text{minimize} & \theta(x) = x_2 & \\ \text{subject to} & (1 - x_1)^3 - x_2 & \geqq 0 \\ & x_1 \qquad & \geqq 0 \\ & x_2 & \geqq 0. \end{array}$$

Verify that $\overline{x} = (1, 0)^T$ is a global optimum solution to this problem. Is $\overline{x}$ a regular point? Do the first order necessary optimality conditions hold at $\overline{x}$?

If the problem is to minimize: $-x_1$, subject to the constraints given above, verify that $\overline{x}$ is again the global optimum. Do the first order necessary optimality conditions hold at $\overline{x}$ for this problem? Why?

29. In each of the following NLPs, find out the global optimum and check whether the first order necessary optimality conditions hold at it. Explain the reasons for it.

(87)		(88)	
minimize	$-x_1$	minimize	$-x_1$
subject to	$-x_1^2 \geqq 0$	subject to	$-x_1^2 + x_2 \geqq 0$
	$x_1 \geqq 0$		$-x_2 \geqq 0.$

30. Find an optimum solution to the following NLP, using a combinatorial search for the set of active constraints at the optimum

$$\begin{array}{ll} \text{minimize} & \theta(x) = 2x_1^2 + 2x_1x_2 + x_2^2 - 10x_1 - 10x_2 \\ \text{subject to} & x_1^2 + x_2^2 \leqq 5 \\ & 3x_1 + x_2 \leqq 6. \end{array}$$

31. Consider the following NLP

$$\begin{array}{ll} \text{minimize} & -x_1^4 - x_2^4 \\ \text{subject to} & x_1^2 + (x_2 - 1)^2 - 1 = 0. \end{array}$$

Verify that $\bar{x} = (0, 2)^T$ is a global minimum for this problem. Do the first order necessary optimality conditions hold at $\bar{x}$? Is there a $\bar{\mu}$ such that $(\bar{x}, \bar{\mu})$ is a local minimum for the Lagrangian in this problem?

32. Consider the general NLP

$$\begin{array}{ll} \text{minimize} & \theta(x) \\ \text{subject to} & h_i(x) = 0, \ i = 1 \text{ to } m \\ & g_p(x) \geqq 0, \ p = 1 \text{ to } t \end{array}$$

where $\theta(x)$, $h_i(x)$, $g_p(x)$ are all continuously differentiable functions defined on $\mathbf{R}^n$.

One of the hard unsolved problems in NLP is to develop a computationally viable method or characterization to determine whether $\theta(x)$ is bounded below on the feasible solution set for this problem, or diverges $-\infty$ on this set; and when $\theta(x)$ is bounded on the solution set, to determine whether $\theta(x)$ attains its minimum at some finite feasible solution ($\theta(x)$ may only have an infimum in this problem, it may not be an attained minimum).

Another hard problem is to develop optimality conditions for a feasible solution $\bar{x}$ of this problem to be a global minimum for it. In the absence of convexity of $\theta(x)$, concavity of $g(x)$ and affineness of $h(x)$, at present we do not have any conditions for distinguishing the global minimum for this problem, from other local minima that may exist (the only known condition for the global minimum is its definition, that is, $\bar{x}$ is a global minimum iff $\theta(x) \geqq \theta(\bar{x})$ for all feasible solutions x, this condition is not computationally useful, since checking it directly may require computing the function value at uncountably many points).

33. Let A be a given matrix of order $m \times n$. Prove that the following three conditions are equivalent

(i) there exists no $x \in \mathbf{R}^n$ satisfying $Ax \leqq 0$, $x \geq 0$,

(ii) for every $b \in \mathbf{R}^m$, the set $\{x : Ax \leqq b, x \geqq 0\}$ is bounded,

(iii) there exists a $\pi \geqq 0$ satisfying $\pi A > 0$.

34. If $\theta(x)$ is a continuous real valued function defined over $\mathbf{R}^n$ with the monotonicity property (that is for every $0 \leqq x \leqq y$ we have $\theta(x) \leqq \theta(y)$), then prove that the problem

$$\begin{array}{ll} \text{minimize} & \theta(x) \\ \text{subject to} & Ax \geqq b \\ & x \geqq 0 \end{array}$$

has an optimum solution, if it is feasible. (B. C. Eaves [2.8])

35. Let $Q(x) = cx + (\frac{1}{2})x^T Dx$ where D is a symmetric matrix. Let $\alpha > 0$ be given. Prove that the point x^* solves the problem

$$\begin{array}{ll} \text{minimize} & Q(x) \\ \text{subject to} & \|x\| \leqq \alpha \end{array}$$

iff it is feasible and there exists a $\lambda \geqq 0$ satisfying

$$\begin{aligned} &(x^*)^T(D + \lambda I) = -c \\ &\lambda(\alpha - \|x^*\|) = 0 \\ &(D + \lambda I) \text{ is a PSD matrix.} \end{aligned}$$

36. Consider the following NLPs in each of which the variables are $x \in \mathbf{R}^n$.

(89)		(90)	
minimize	cx	minimize	$x^T x$
subject to	$x^T x \leqq 1$	subject to	$-cx \geqq 1$
	$Ax \geqq 0$		$Ax \geqq 0.$

The data in both the problems, the matrices A, c of order $m \times n$ and $1 \times n$ respectively, are the same. Prove that these two problems are equivalent.

37. Let $f(\lambda) : \mathbf{R}^1 \to \mathbf{R}^1$ be a real valued convex function defined on $\mathbf{R}^1$. For any λ, the limit of $\frac{f(\lambda+\varepsilon)-f(\lambda)}{\varepsilon}$ as $\varepsilon \to 0$ through positive values is called the right derivative of $f(\lambda)$ at λ, and denoted by $f'_+(\lambda)$, the limit of the same as $\varepsilon \to 0$ through negative values is called the left derivative of $f(\lambda)$ at λ and denoted by $f'_-(\lambda)$. Prove the following

i) If $\lambda < \gamma$, then $f'_-(\lambda) \leqq f'_+(\lambda) \leqq f'_-(\gamma) \leqq f'_+(\gamma)$.

ii) A necessary and sufficient condition for λ_* to minimize $f(\lambda)$ over $\lambda \in \mathbf{R}^1$ is: $f'_-(\lambda_*) \leqq 0 \leqq f'_+(\lambda_*)$.

iii) The subdifferential $\partial f(\lambda)$ is the line segment $[f'_-(\lambda), f'_+(\lambda)]$.

iv) For each λ, let $g(\lambda) \in \partial f(\lambda)$. Prove that

(a) $P(\lambda, \gamma) = f(\lambda) - [f(\gamma) + g(\gamma)(\lambda - \gamma)] \geqq 0$ for all λ, γ.
(b) If $f(\lambda) \leqq f(\gamma)$, then $P(\lambda, \gamma) \leqq |g(\gamma)| \cdot |\gamma - \lambda|$.
(c) If $g(\lambda)g(\gamma) < 0$, then

$$P(\lambda, \gamma) \geqq |g(\gamma)| \cdot |\lambda - \lambda_*| \text{ where } \lambda_* \text{ is the minimizer of } f(\lambda).$$

(C. Lemarechal and R. Mifflin [10.23])

38. Consider the problem

$$\begin{array}{ll} \text{minimize} & cx \\ \text{subject to} & Ax \leqq 0 \\ & \|x\| = 1 \end{array}$$

where A is of order $m \times n$.

Let $\mathbf{K} = \{y : y = \sum_{i=1}^{m} \mu_i A_{i.}, \mu_i \geqq 0 \text{ for all } i\}$. Prove the following

i) If $c \in \mathbf{K}$, the maximum objective value in this problem, is $\leqq 0$.
ii) If $c \notin \mathbf{K}$, let $b \in \mathbf{K}$ be the point in $\mathbf{K}$ that is closest to c. Then $(c - b)/\|b - c\|$ is the optimum solution of this problem, and the optimum objective value in the problem is $\|c - b\|$.

39. Let $\mathbf{K} \subset \mathbf{R}^n$ be a closed convex polyhedral set partitioned into closed convex polyhedral regions as $\bigcup_{t=1\text{to } r} \mathbf{K}_t$. So if $u \neq v$, the interiors of $\mathbf{K}_u$ and $\mathbf{K}_v$ have an empty intersection, and $\mathbf{K}_u \cap \mathbf{K}_v$ is itself either empty or is either a face of lower dimension or a subset of a face of lower dimension of each of $\mathbf{K}_u$ and $\mathbf{K}_v$. Assume that each $\mathbf{K}_t$ has a nonempty interior. Suppose the real-valued function $f(x)$ is defined on $\mathbf{K}$ by the following

$$f(x) = f_t(x) = c_0^t + \sum_{j=1}^{n} c_j^t x_j, \text{ if } x \in \mathbf{K}_t$$

where c_0^t and c_j^t are all given constants. The definition assumes that if $\mathbf{K}_u \cap \mathbf{K}_v \neq \emptyset$, then $f_u(x) = f_v(x)$ for all $x \in \mathbf{K}_u \cap \mathbf{K}_v$. So $f(x)$ is a continuous piecewise linear function defined on $\mathbf{K}$.

Derive necessary and sufficient conditions for the continuous piecewise linear function $f(x)$ to be convex on $\mathbf{K}$, and develop an efficient algorithm to check whether these conditions hold.

As a numerical example, let $\mathbf{K} = \{x = (x_1, x_2)^T : -1 \leqq x_1 \leqq 1, -1 \leqq x_2 \leqq 1\}$. Consider the partition of $\mathbf{K}$ given in Figure 22. Two piecewise linear functions $f(x)$, $g(x)$ defined on $\mathbf{K}$ are provided in Figure 22. Check whether they are convex on $\mathbf{K}$. (See Section 8.14 in K. G. Murty [2.26].)

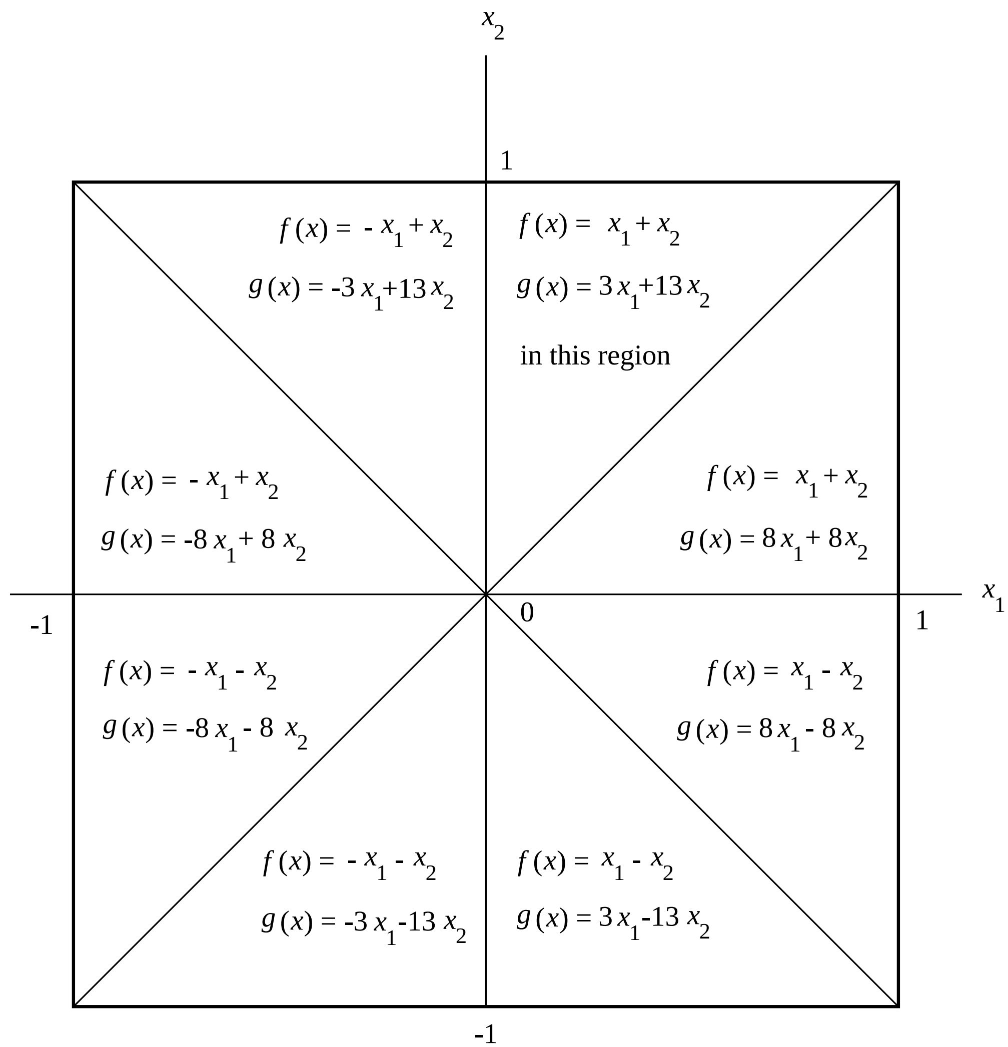

Figure 22

40. (Research Problem) For $i = 1$ to m, $g_i(x)$ is a real valued continuously differentiable function defined on $\mathbf{R}^n$, but $g(x) = (g_1(x), \ldots, g_m(x))^T$ is not convex. Let $\alpha = (\alpha_1, \ldots, \alpha_m)^T \in \mathbf{R}^m$. Let $\mathbf{K}(\alpha) = \{x : g_i(x) \leqq \alpha_i,\ i = 1 \text{ to } m\}$. Let $b = (b_1, \ldots, b_n) \in \mathbf{R}^n$ be a given point.

(i) Assuming that $\mathbf{K}(0) \neq \emptyset$ is a convex set, develop an efficient algorithm to check whether $\mathbf{K}(\alpha)$ is convex for given α. Is this problem easier to solve if either $\alpha \geqq 0$ or $\alpha \leqq 0$?

(ii) Assuming that $\mathbf{K}(0) \neq \emptyset$ is a convex set, and that b is a boundary point of $\mathbf{K}(0)$, (that is, there exists an i such that $g_i(b) = 0$), develop an efficient algorithm to find a $c = (c_1, \ldots, c_n) \neq 0$ satisfying $c(x - b) \geqq 0$ for all $x \in \mathbf{K}(0)$ (then $\mathbf{H} = \{x : c(x - b) = 0\}$ is a supporting hyperplane for the convex set $\mathbf{K}(0)$ at its boundary point b).

(iii) Assuming that $\mathbf{K}(0) \neq \emptyset$ is a convex set and that $b \notin \mathbf{K}(0)$, develop an efficient algorithm to determine a hyperplane separating b from $\mathbf{K}(0)$.

(iv) Consider the special cases of the above problems when all $g_i(x)$ are affine functions, excepting one which is quadratic and nonconvex.

41. Let $\theta(x)$ be a continuously differentiable real valued function defined on $\mathbf{R}^n$. Let $\mathbf{K}$ be a subspace of $\mathbf{R}^n$. If $\overline{x} \in \mathbf{K}$ minimizes $\theta(x)$ over $\mathbf{K}$, prove that $\nabla\theta(\overline{x})$ is orthogonal to every vector in $\mathbf{K}$.

42. Let $\theta(x)$; $g_i(x)$, $i = 1$ to m, be continuously differentiable convex functions defined on $\mathbf{R}^n$. Let $\overline{\theta}$ be the optimum objective value; and $\overline{\pi}$, an optimum Lagrange multiplier vector associated with the NLP

$$\begin{array}{ll} \text{minimize} & \theta(x) \\ \text{subject to} & -g_i(x) \geqq 0, \; i = 1 \text{ to } m. \end{array}$$

Then prove that $\overline{\theta} = \text{Infimum } \{\theta(x) + \overline{\pi} g(x) : x \in \mathbf{R}^n\}$.

43. Arithmetic Mean — Geometric Mean Inequality:

Let $x_1, \ldots, x_n$ be positive real numbers. Let $\delta_1, \ldots, \delta_n$ be positive real numbers satisfying $\delta_1 + \ldots + \delta_n = 1$. Prove that

$$\prod_{i=1}^{n} (x_i)^{\delta_i} \leqq \sum_{i=1}^{n} \delta_i x_i$$

with equality holding iff $x_1 = x_2 = \ldots = x_n$, where "$\prod$" indicates the product sign.

44. Young's Inequality:

Let x, y, p, q be all positive real numbers, and $p > 1$, $q > 1$ satisfying $\frac{1}{p} + \frac{1}{q} = 1$. Prove that

$$xy \leqq \frac{x^{\mathbf{p}}}{p} + \frac{y^{\mathbf{q}}}{q}$$

with equality holding only when $x^{\mathbf{p}} = y^{\mathbf{q}}$.

45. Holder's Inequality:

Let p, q, be positive real numbers > 1 satisfying $\frac{1}{p} + \frac{1}{q} = 1$. Let $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n)$ be real vectors. Prove that

$$\sum_{i=1}^{n} x_i y_i \leqq \left(\sum_{i=1}^{n} |x_i|^{\mathbf{p}}\right)^{\mathbf{1/p}} \left(\sum_{i=1}^{n} |y_i|^{\mathbf{q}}\right)^{\mathbf{1/q}}.$$

46. Minkowski's Inequality:

Let $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n)$ be real vectors and $p \geqq 1$. Prove that

$$\left(\sum_{i=1}^{n} |x_i + y_i|^{p}\right)^{\mathbf{1/p}} \leqq \left(\sum_{i=1}^{n} |x_i|^{\mathbf{p}}\right)^{\mathbf{1/p}} + \left(\sum_{i=1}^{n} |y_i|^{\mathbf{p}}\right)^{\mathbf{1/p}}.$$

47. Let $f(\lambda)$ denote a smooth real valued function defined over $\mathbf{R}^1$. A classical sufficient condition for $\overline{\lambda} \in \mathbf{R}^1$ to be a local minimum for $f(\lambda)$ over $\mathbf{R}^1$ states "$\overline{\lambda}$ is a local minimum for $f(\lambda)$ over $\mathbf{R}^1$ if the first nonzero derivative of $f(\lambda)$ at $\overline{\lambda}$ is of even order, and this derivative is > 0". Develop a generalization of this result to $\mathbf{R}^n$, $n > 1$.

48. Given a vector $y = (y_j) \in \mathbf{R}^n$, define

$$\begin{aligned}
\|y\|_1 &= \sum_{j=1}^{n} |y_j| \\
\|y\|_\infty &= \text{ maximum } \{|y_j| : j = 1 \text{ to } n\} \\
\|y\|_2 &= \sqrt{\sum_{j=1}^{n} y_j^{\mathbf{2}}} \\
y^+ &= (y_j^+) \text{ where } y_j^+ = \text{ maximum } \{0, y_j\}
\end{aligned}$$

$\|y\|_1$, $\|y\|_\infty$, $\|y\|_2$ are called the 1-norm, ∞-norm, 2-norm, respectively, of the vector y.

Consider the system

$$\begin{aligned} Ax &\leqq b \\ Bx &= d \end{aligned} \tag{91}$$

where A, B are fixed matrices of orders $m \times n$, $p \times n$ respectively; and b, d are column vectors of appropriate dimensions. Assume that each row vector of A contains at least one nonzero entry, and if equality constraints do exist, then B is of full row rank (B

could be vacuous, that is, there may be no equality constraints in (91)). Let $\mathbf{K}(b,d)$ denote the set of feasible solutions of (91). Define

$$\begin{aligned}\mu(A,B) = &\ \text{supremum } \|u,v\|_2\\ &\ \text{subject to, } u,\ v \text{ are row vectors in } \mathbf{R}^m,\ \mathbf{R}^p,\\ \|uA+vB\|_1 = &\ 1\\ u \geqq &\ 0\\ &\ \text{and the set of rows of } \begin{pmatrix} A\\ B\end{pmatrix} \text{ corresponding to}\\ &\ \text{nonzero elements of } (u,v) \text{ is linearly independent.}\end{aligned} \tag{92}$$

(i) Prove that $\mu(A,B)$ is finite.

(ii) If $\begin{pmatrix} b^1\\ d^1\end{pmatrix}$, $\begin{pmatrix} b^2\\ d^2\end{pmatrix}$ are such, that $\mathbf{K}(b^1,d^1)$ and $\mathbf{K}(b^2,d^2)$ are both nonempty; for each $x^1 \in \mathbf{K}(b^1,d^1)$, prove that there exists an $x^2 \in \mathbf{K}(b^2,d^2)$ satisfying

$$\|x^1 - x^2\|_\infty \leqq \mu(A,B)\left\|\begin{pmatrix} b^1\\ d^1\end{pmatrix} - \begin{pmatrix} b^2\\ d^2\end{pmatrix}\right\|_2.$$

This result can be interpreted as implying that feasible solutions of (91) are Lipschitz continuous with respect to right hand side constants vector perturbations, with Lipschitz constant $\mu(A,B)$ depending only on the coefficient matrix $\begin{pmatrix} A\\ B\end{pmatrix}$.

(iii) In (91), if B is of full row rank and the system "$Ay < 0$, $By = 0$" has a solution y, prove that $\mathbf{K}(b,d) \neq \emptyset$ for all $\begin{pmatrix} b\\ d\end{pmatrix} \in \mathbf{R}^{m+p}$, and that for any $\begin{pmatrix} b^1\\ d^1\end{pmatrix}$, $\begin{pmatrix} b^2\\ d^2\end{pmatrix} \in \mathbf{R}^{m+p}$, and $x^1 \in \mathbf{K}(b^1,d^1)$, there exists an $x^2 \in \mathbf{K}(b^2,d^2)$ satisfying

$$\|x^1 - x^2\|_\infty \leqq \overline{\mu}(A,B)\left\|\begin{pmatrix} b^1\\ d^1\end{pmatrix} - \begin{pmatrix} b^2\\ d^2\end{pmatrix}\right\|_2$$

where

$$\overline{\mu}(A,B) = \text{ maximum } \|u,v\|_2$$

subject to, u, v are row vectors in $\mathbf{R}^m$, $\mathbf{R}^p$,

$$\begin{aligned}\|uA+vB\|_1 &= 1\\ u &\geqq 0.\end{aligned} \tag{93}$$

(iv) Suppose $\begin{pmatrix} b^1\\ d^1\end{pmatrix}$ is such that $\mathbf{K}(b^1,d^1) \neq \emptyset$. For any $x \in \mathbf{R}^n$, prove that there exists an $x^1 \in \mathbf{K}(b^1,d^1)$ satisfying

$$\|x - x^1\|_\infty \leqq \mu(A,B)\left\|\begin{matrix} (Ax-b^1)^+\\ (Bx-d^1)\end{matrix}\right\|_2.$$

If the Lipschitz constant $\mu(A,B)$ is available, this inequality provides an error bound on how far x is from a feasible solution of (91).

(v) Consider the LP

$$\begin{array}{ll} \text{minimize} & cx \\ \text{subject to} & Ax \leqq b \\ & Bx = d \end{array} \tag{94}$$

with A, B fixed, let $\mathbf{\Gamma}(b,d)$ denote the set of optimum solutions of (94). If it is known that $\mathbf{\Gamma}(b^1,d^1)$ and $\mathbf{\Gamma}(b^2,d^2)$ are both nonempty; for any $x^1 \in \mathbf{\Gamma}(b^1,d^1)$ prove that there exists an $x^2 \in \mathbf{\Gamma}(b^2,d^2)$ satisfying

$$\|x^1 - x^2\|_\infty \leqq \mu(A,B) \left\| \begin{pmatrix} b^1 \\ d^1 \end{pmatrix} - \begin{pmatrix} b^2 \\ d^2 \end{pmatrix} \right\|_2$$

where $\mu(A,B)$ is the Lipschitz constant defined in (92). This result can be interpreted as implying that optimum solutions of LPs are Lipschitz continuous with respect to right hand side constants vector perturbations.

(vi) Consider the LP

$$\begin{array}{ll} \text{minimize} & -(1+\delta)x_1 - x_2 \\ \text{subject to} & x_1 + x_2 \leqq 1 \\ & x_1, x_2 \geqq 0 \end{array} \tag{95}$$

where δ is a real parameter. Show that when $\delta > -1$ and $\delta \neq 0$, this problem has a unique optimum solution $x(\delta)$ given by

$$x(\delta) = \begin{cases} (1,0)^T, & \text{if } \delta > 0 \\ (0,1)^T, & \text{if } -1 < \delta < 0. \end{cases}$$

By showing that

$$\operatorname*{limit}_{\delta \to 0^+} \frac{\|x(\delta) - x(-\delta)\|}{2\delta} = +\infty$$

prove that $x(\delta)$ is not Lipschitzian with respect to δ.

This shows that in general, optimum solutions of linear programs are not Lipschitzian with respect to perturbations in the objective function coefficients.

(vii) Consider the LCP (q,M) of order n. Let $\mathbf{J} \subset \{1,\dots,n\}$. Consider the system

$$\begin{array}{ll} M_{i.}z + q_i \geqq 0,\ z_i = 0, & \text{for all } i \in \mathbf{J} \\ M_{i.}z + q_i = 0,\ z_i \geqq 0, & \text{for all } i \notin \mathbf{J}. \end{array} \tag{96}$$

If z is any solution of (96) then z leads to a solution of the LCP (q,M) (that is, $(w = Mz+q, z)$ is a solution of the LCP (q,M)). Using this fact, Lipschitz continuity of solutions with respect to the right hand side constants vector perturbations, can be established for certain classes of LCPs.

For any $\mathbf{J} \subset \{1,\dots,n\}$, define $A(\mathbf{J})$ to be the square matrix of order n such that

$$(A(\mathbf{J}))_{i.} = \begin{cases} -M_{i.} & \text{for } i \in \mathbf{J} \\ I_{i.} & \text{for } i \notin \mathbf{J}. \end{cases}$$

Similarly, define the square matrix $B(\mathbf{J})$ of order n by

$$(B(\mathbf{J}))_{i.} = \begin{cases} I_{i.} & \text{for } i \in \mathbf{J}. \\ -M_{i.} & \text{for } i \notin \mathbf{J}. \end{cases}$$

Now define

$$\sigma(M) = \text{ maximum } \{\mu(A(\mathbf{J}), B(\mathbf{J})) : \mathbf{J} \subset \{1, \ldots, n\}\}$$

where $\mu(A(\mathbf{J}), B(\mathbf{J}))$ is $\mu(A, B)$ of (92) with $A = A(\mathbf{J})$, $B = B(\mathbf{J})$.

Suppose M is a P-matrix and (w^r, z^r) is the unique solution of the LCP (q, M) when $q = q^r$, $r = 1, 2$. Prove that

$$\|z^1 - z^2\|_\infty \leqq \sigma(M) \|q^1 - q^2\|_2.$$

This establishes that when M is a P-matrix and fixed, the solution of the LCP (q, M) is Lipschitz continuous in q with Lipschitz constant $\sigma(M)$ defined above.

(viii) Let $M = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, $q^1 = \begin{pmatrix} -\varepsilon \\ 1 \end{pmatrix}$, $q^2 = \begin{pmatrix} \varepsilon \\ 1 \end{pmatrix}$ where $\varepsilon > 0$. Show that if $z^1 = (1, \varepsilon)^T$, $z^2 = (0, 0)^T$, the solution of the LCP (q^r, M) is $(w^r = Mz^r + q, z^r)$, $r = 1, 2$. Verify that

$$\operatorname*{limit}_{\varepsilon \to 0^+} \frac{\|z^1 - z^2\|_\infty}{2\varepsilon} = +\infty.$$

This shows that the solution of the LCP (q, M) may not be Lipschitzian in q for fixed M, when M is positive semidefinite but not a P-matrix. (O. L. Mangasarian and T. H. Shiau [A11])

49. Let A, b be given real matrices of orders $m \times n$ and $m \times 1$ respectively. Consider the system of equations

$$Ax = b. \tag{97}$$

This system may or may not have a solution. It is required to find a vector x that satisfies (97) as closely as possible using the least squares measure of deviation. Formulate this as a nonlinear program and write down the optimality conditions for it. Prove that this system of optimality conditions always has a solution.

Now consider the problem of finding a vector x satisfying (97) as closely as possible, subject to the additional constraints $\|x\| = 1$, which is required to be satisfied. This leads to the nonlinear program

$$\begin{array}{ll} \text{minimize} & \|Ax - b\|^{\mathbf{2}} \\ \text{subject to} & \|x\|^2 = 1. \end{array} \tag{98}$$

Discuss how (98) can be solved to optimality efficiently.

50. Let $f(x)$ be a real valued function defined on $\mathbf{R}^n$ which is thrice continuously differentiable. Consider the NLP

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \geqq 0. \end{array}$$

i) Prove that the first order necessary optimality conditions for this NLP can be posed as a nonlinear complementarity problem.
ii) Let $f_i(x) = \frac{\partial f(x)}{\partial x_i}$, $i = 1$ to n. Define $g(x) = (g_i(x) : i = 1 \text{ to } n)^T$ where

$$g_i(x) = -|x_i - f_i(x)|^{\mathbf{3}} + x_i^{\mathbf{3}} + (f_i(x))^{\mathbf{3}}.$$

Prove that solving he NLCP described in (i) above, is equivalent to solving the system of n equations in n unknowns

$$g(x) = 0.$$

Show that $g(x)$ is twice continuously differentiable. (L. Watson [A15])

51. We have received a large shipment of an engineering item. A random sample of 10 items selected from this lot had the following lifetimes in time units.

1.600	1.506	0.501	1.118
0.295	0.070	1.821	
0.055	0.499	3.102	

Assume that the lifetime, x, of items from the lot follows a Weibull distribution with the following probability density function

$$f(x) = \alpha\theta x^{\theta-1} e^{(-\alpha x^{\theta})}, \ x \geqq 0.$$

Formulate the problem of obtaining the maximum liklihood estimators for the parameters α, θ as an NLP. Write down the optimality conditions for this NLP, and solve this NLP using them.

52. Consider the convex polyhedra $\mathbf{K}_1$, $\mathbf{K}_2$, which are the sets of feasible solutions of the systems given below

$$\begin{array}{ccc} \mathbf{K}_1 & \qquad & \mathbf{K}_2 \\ \hline Ax = b & & Dy = d \\ x \geqq 0 & & y \geqq 0. \end{array}$$

It is required to find a pair of points $(x; y)$, $x \in \mathbf{K}_1$, $y \in \mathbf{K}_2$, which are closest in terms of the Euclidean distance, among all such pairs. Does this problem have a unique optimum solution? Why?

Formulate this problem as an NLP and write down the necessary optimality conditions for it. Are these conditions also sufficient for optimality for this problem?

53. Write down the necessary optimality conditions for Sylvester's problem, Exercise 1.25, and determine whether these conditions are also sufficient for optimality.

54. We are given smooth real valued functions $\theta_1(x), \dots, \theta_r(x), g_1(x), \dots, g_m(x)$, all defined over $\mathbf{R}^n$. Consider the following optimization problem:

$$\begin{array}{ll} \text{minimize} & v(x) \\ \text{subject to} & g_i(x) \geqq 0, \ i = 1 \text{ to } m \end{array}$$

where for each $x \in \mathbf{R}^n$, $v(x) = \text{maximum } \{\theta_1(x), \dots, \theta_r(x)\}$. Transform this problem into a smooth NLP with a linear objective function, but with additional constraints than those in this problem. Write down the necessary optimality conditions for the transformed problem and simplify them. State some general conditions on the data in the problem under which these conditions are also sufficient for optimality. Show that this technique can be used to transform any NLP into an NLP in which the objective function is linear.

55. The army has n types of weapons available. Using them, they want to destroy m targets. The following data is given:

$$\begin{aligned} p_{ij} &= \text{probability that a weapon of type } j \text{ shot at target type } i \text{ will destroy it,} \\ v_i &= \text{value in \$ of target } i, \\ b_j &= \text{number of weapons of type } j \text{ available.} \end{aligned}$$

Assume that a weapon shot at a target either destroys it, or leaves it absolutely unaffected.

Formulate the problem of determining the number of weapons of each type to be shot at each of the targets, so as to maximize the expected value destroyed, as an NLP. Neglecting the integer requirements on the decision variables in this problem, write down the necessary optimality conditions for it. Specialize these for the numerical example with the following data.

$$n = 2, \ m = 3$$

$$p = (p_{ij}) = \begin{pmatrix} .25 & .05 \\ .35 & .08 \\ .15 & .17 \end{pmatrix}, \ v = (v_i) = \begin{pmatrix} 150 \\ 95 \\ 375 \end{pmatrix}, \ b = (b_i) = \begin{pmatrix} 6 \\ 10 \end{pmatrix}.$$

56. Let B, A be matrices of order $n \times n$ and $m \times n$ respectively. Suppose $\text{rank}(A) = m$ and B is symmetric and PD on the subspace $\{x : Ax = 0\}$. Then prove that the matrix $\begin{pmatrix} B & A^T \\ A & O \end{pmatrix}$ is nonsigular.

57. Let $f(x)$ be a real valued convex function defined on $\mathbf{R}^n$. Assume that $f(x)$ is twice continuously differentiable at a given point $\overline{x} \in \mathbf{R}^n$. Define

$$\begin{aligned} l(x) &= f(\overline{x}) + \nabla f(\overline{x})(x - \overline{x}) \\ Q(x) &= f(\overline{x}) + \nabla f(\overline{x})(x - \overline{x}) + \tfrac{1}{2}(x - \overline{x})^T H(f(\overline{x}))(x - \overline{x}). \end{aligned}$$

The functions $l(x)$, $Q(x)$ are respectively the first and second order Taylor approximations for $f(x)$ around $\overline{x}$. In Theorem 15 we established that $f(x) - l(x)$ always has the same sign ($\geqq 0$) for all $x \in \mathbf{R}^n$. Discuss whether $f(x) - Q(x)$ always has the same sign for all $x \in \mathbf{R}^n$. If so, what is that sign? Why? (Richard Hughes)

58. Let $\mathbf{K}$ denote the set of feasible solutions of

$$Ax \geqq b \tag{99}$$

where A is an $m \times n$ matrix. We know that $\mathbf{K} \neq \emptyset$ and dimension of $\mathbf{K}$ is n. $\theta(x)$ is a strictly convex function defined on $\mathbf{R}^n$, with a unique unconstrained minimum in $\mathbf{R}^n$, $\overline{x}$. We know that $\overline{x}$ satisfies all but one constraint in (99). Suppose $A_{i.}\overline{x} \geqq b_i$ for $i = 2$ to m, but $A_{1.}\overline{x} < b_1$. Prove that if the problem

$$\begin{array}{ll} \text{minimize} & \theta(x) \\ & x \in \mathbf{K} \end{array}$$

has an optimum solution, the first constraint in (99) must be active at it. What is the appropriate generalization of this result when $\overline{x}$ violates more than one constraint in (99)? (M. Q. Zaman, S. U. Khan, and A. Bari, private communication).

59. Let $f(\lambda) : \mathbf{R}^1 \to \mathbf{R}^1$ be a continuously differentiable strictly increasing function of the real parameter λ.

Let $\theta(x) : \mathbf{R}^n \to \mathbf{R}^1$, $g(x) : \mathbf{R}^n \to \mathbf{R}^m$, $h(x) : \mathbf{R}^n \to \mathbf{R}^t$ be continuously differentiable functions.

Consider the constraint system

$$\begin{aligned} g(x) &\geqq 0 \\ h(x) &= 0 \end{aligned} \tag{100}$$

and the two optimization problems

Problem 1: Minimize $\theta(x)$, subject to (100)

Problem 2: Minimize $f(\theta(x))$, subject to (100).

Rigorously prove that both the problems have the same set of stationary points. (H. L. Li, private communication.)

60. Consider the following separable NLP

$$\begin{aligned} \text{minimize} \quad & \sum_{j=1}^{n} f_j(x_j) \\ \text{subject to} \quad & \sum_{j=1}^{n} x_j \geqq 1 \\ & x_j \geqq 0,\ j = 1 \text{ to } n \end{aligned}$$

where $f_j(x_j)$ is a differentiable function for all j. If $\overline{x} = (\overline{x}_j)$ solves this problem, prove that there must exist a nonnegative scalar k such that

$$\frac{df_j(\overline{x}_j)}{dx_j} \geqq k \text{ for all } j,$$

$$\text{and for all } j \text{ such that } \overline{x}_j > 0\,, \ \frac{df_j(\overline{x}_j)}{dx_j} = k.$$

61. Let A be a given matrix of order $m \times n$. Prove that at least one of the following systems

(I)	(II)
$Ax \geqq 0$	$\pi A \leqq 0$
$x \geqq 0$	$\pi \geqq 0$

has a nonzero feasible solution.

62. Consider the following LP

$$\begin{aligned} \text{minimize} \quad & cx \\ \text{subject to} \quad & Ax \geqq b \\ & x \geqq 0. \end{aligned}$$

Let $\mathbf{K}$, $\mathbf{\Gamma}$ denote the set of feasible solutions of the LP, and its dual respectivley. Prove that either both $\mathbf{K}$ and $\mathbf{\Gamma}$ are empty, or at least one of $\mathbf{K}$, $\mathbf{\Gamma}$ is an unbounded set.

63. Let D be the diagonal matrix $\text{diag}(\lambda_1, \ldots, \lambda_n)$, where $\lambda_1 \geqq \lambda_2 \geqq \ldots \geqq \lambda_n > 0$. Consider the following NLP

$$\begin{aligned} \text{minimize} \quad & (y^T D y)(y^T D^{-1} y) \\ \text{subject to} \quad & y^T y = 1. \end{aligned}$$

(i) Transform this NLP into another problem in new variables $x_1, \ldots, x_n$ in which the objective function to be optimized is a product of two linear functions, $g(x)$ and $h(x)$, say, and the constraints are all linear. Call this transformed problem (P).

(ii) Show that (P) must have a global optimum solution.

Assume that x^* is an optimum solution of (P). Let $h(x^*) = \delta$. Show that x^* must also be an optimum solution of the LP in the variables $x = (x_1, \dots, x_n)$ in which the objective function to be optimized is $g(x)$, and the constraints are the same as those in (P) plus the additional constraint $h(x) = \delta$. Conversely, show that every optimum solution of this LP must also be optimal to (P). Using this, show that (P) has an optimum solution in which two variables among $x_1, \dots, x_n$ are positive, and the others are all zero.

(iii) Consider the problem (P) again. In this problem, substitute $x_i = 0$ for all $i \neq p, q$, for some selected p, q between 1 to n. Show that in the optimum solution of this reduced problem, both x_p and x_q are equal.

(iv) Use the above results to prove Kantorovich's inequality which states the following: Let A be a symmetric PD matrix of order n with eigen values $\lambda_1 \geqq \lambda_2 \geqq \dots \geqq \lambda_n > 0$. Then

$$(y^T A y)(y^T A^{-1} y) \leqq \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n} \quad \text{for all } y \text{ such that } \|y\| = 1.$$

(M. Raghavachari, "A linear programming proof of Kantorovich's inequality", The American Statistician, 40 (1986) 136–137.)

64. $f(x)$ is a real valued differentiable function defined on $\mathbf{R}^n$. Prove that $f(x)$ is a convex function iff

$$(\nabla f(x^2) - \nabla f(x^1))(x^2 - x^1) \geqq 0,$$

for all $x^1, x^2 \in \mathbf{R}^n$. Similarly, prove that a real valued differentiable function $g(x)$ defined on $\mathbf{R}^n$ is concave iff

$$(\nabla g(x^2) - \nabla g(x^1))(x^2 - x^1) \leqq 0,$$

for all $x^1, x^2 \in \mathbf{R}^n$.

65. Let $f(x)$ be a real valued convex function defined on $\mathbf{R}^n$. For each $x \in \mathbf{R}^n$ let $f^+(x)$ and $f^-(x)$ denote the positive and negative parts of $f(x)$, that is, $f^+(x)$ and $f^-(x)$ satisfy for all $x \in \mathbf{R}^n$, $f^+(x) \geqq 0$, $f^-(x) \geqq 0$, $f(x) = f^+(x) - f^-(x)$, $(f^+(x))(f^-(x)) = 0$. Are $f^+(x)$ and $f^-(x)$ both convex functions over $\mathbf{R}^n$? Why?

66. Consider the linearly constrained NLP

$$\begin{array}{ll} \text{minimize} & \theta(x) \\ \text{subject to} & Ax = b \end{array}$$

where $\theta(x)$ is a real valued continuously differentiable convex function defined on $\mathbf{R}^n$, and A is an $m \times n$ matrix of rank m. If x^* is a feasible solution for this problem satisfying

$$\nabla\theta(x^*)(I - A^T(AA^T)^{-1}A) = 0$$

prove that x^* is an optimum solution of the NLP.

67. Consider the NLP

$$\begin{array}{ll} \text{minimize} & cx \\ \text{subject to} & g_i(x) \geqq 0, \ i = 1 \text{ to } m \end{array}$$

where $c = (c_1, \ldots, c_n) \neq 0$, and $g_i(x)$ is a continuously differentiable real valued function defined over $\mathbf{R}^n$ for each $i = 1$ to m. Suppose x^* is a local minimum for this problem, and is a regular point. Prove that there exists at least one i such that $g_i(x^*) = 0$.

68. i) On the x_1, x_2-Cartesian plane, find the nearest point on the parabola $\{x = (x_1, x_2)^T : x_2^2 = 4x_1\}$ to $(1, 0)^T$ in terms of the Euclidean distance.
ii) For the following NLP, check whether either of $x^1 = (1, \frac{1}{2})^T$ or $x^2 = (\frac{1}{3}, -\frac{1}{6})^T$ are optimum solutions

$$\begin{array}{ll} \text{minimize} & x_1^2 + 2x_2^2 \\ \text{subject to} & x_1^2 + \ x_2^2 \leqq 5 \\ & 2x_1 - 2x_2 = 1. \end{array}$$

69. Let $f(x)$ be a real valued continuously differentiable convex function defined over $\mathbf{R}^n$ and let $\mathbf{K}$ be a closed convex subset of $\mathbf{R}^n$. Suppose $\overline{x} \in \mathbf{K}$ is such that it is the nearest point (in terms of the Euclidean distance) in $\mathbf{K}$ to $\overline{x} - \lambda\nabla f(\overline{x})$ for some $\lambda > 0$. Prove that $\overline{x}$ minimizes $f(x)$ over $x \in \mathbf{K}$. Construct the converse of this statement and prove it too.

70. Consider the following NLP

$$\begin{array}{ll} \text{minimize} & \theta(x) \\ \text{subject to} & l \leqq x \leqq k \end{array}$$

where $\theta(x)$ is a real valued twice continuously differentiable function defined on $\mathbf{R}^n$, and l, k are two bound vectors in $\mathbf{R}^n$ satisfying $l < k$. Develop an algorithm for solving this problem, which takes advantage of the special structure of the problem. Write down the termination criteria that you would use, and provide a justification for them. Also, mention what type of a solution the algorithm is guaranteed to obtain at termination.

71. Consider the following NLP

$$\begin{array}{ll} \text{minimize} & \theta(x) \\ \text{subject to} & a_i \leqq g_i(x) \leqq b_i(x), \ i = 1 \text{ to } m \\ \text{and} & l \leqq x \leqq k \end{array}$$

where $\theta(x)$, $g_i(x), \ldots, g_n(x)$ are all real valued twice continuously differentiable functions defined over $\mathbf{R}^n$, and $a = (a_i)$, $b = (b_i)$, l, k satisfy $a \leqq b$, $l < k$. Discuss how you can solve this problem using the algorithm developed in Exercise 70.

72. $\theta(x)$ is a real valued continuously differentiable convex function defined over $\mathbf{R}^n$. $\mathbf{K}$ is a closed convex subset of $\mathbf{R}^n$. If $\overline{x} \in \mathbf{K}$ is the global maximum for $\theta(x)$ over $x \in \mathbf{K}$, prove that

$$\nabla\theta(\overline{x})x \leqq \nabla\theta(\overline{x})\overline{x}, \quad \text{for all } x \in \mathbf{K}.$$

Is the converse of this statement also true? Why?

Would the above inequality hold for all $x \in \mathbf{K}$ if x is only a local maximum for $\theta(x)$ over $\mathbf{K}$ and not a global maximum? Why?

73. If M is a P-matrix of order n (not necessarily PD) prove that the system

$$\begin{aligned} \pi M &> 0 \\ \pi \quad &\geqq 0 \end{aligned}$$

has a solution π.

74. Write down the first order necessary optimality conditions for the following NLP, and find an optimum solution for it.

$$\begin{array}{ll} \text{minimize} & (x_1 - 4)^{\mathbf{2}} + (x_2 + 1)^{\mathbf{2}} \\ \text{subject to} & 7 \leqq x_1 \leqq 14 \\ & 10 \leqq x_2 \leqq 22. \end{array}$$

75. Consider the following linear program

$$\begin{array}{ll} \text{minimize } z(x) = & = cx \\ \text{subject to} & Ax = b \\ & Dx \geqq d. \end{array}$$

Let $\mathbf{K}$ denote the set of feasible solutions for this problem. Show that the primal simplex algorithm for this problem, is exactly the gradient projection method (Section 10.10.5) applied on this problem, beginning with a feasible point x^0 which is an extreme point of $\mathbf{K}$.

76. $\theta(x)$; $h_i(x)$, $i = 1$ to m; $g_p(x)$, $p = 1$ to t are all real valued twice continuously differentiable functions defined over $\mathbf{R}^n$.

i) Consider the NLP

$$\begin{array}{ll} \text{minimize} & \theta(x) \\ \text{subject to} & h_i(x) = 0, \ i = 1 \text{ to } m, \\ & g_p(x) \geqq 0, \ p = 1 \text{ to } t. \end{array}$$

Let $L(x,\mu,\pi) = \theta(x) - \sum_{i=1}^{m} \mu_i h_i(x) - \sum_{p=1}^{t} \pi_p g_p(x)$ be the Lagrangian. Suppose we have a feasible solution $\overline{x}$ to this NLP and Lagrange multiplier vectors $\overline{\mu}$, $\overline{\pi}$ such that $(\overline{x},\overline{\mu},\overline{\pi})$ satisfy the first order necessary optimality conditions for this NLP, and the additional condition that $L(x,\overline{\mu},\overline{\pi})$ is a convex function in x over $\mathbf{R}^n$ (notice that $L(x,\overline{\mu},\overline{\pi})$ could be a convex function, even though $\theta(x)$, $-g_p(x)$, $h_i(x)$ and $-h_i(x)$ are not all convex functions). Then prove that $\overline{x}$ must be a global minimum for this NLP.

ii) Consider the numerical example

$$\begin{array}{ll} \text{minimize} & (x_1-\alpha)^{\mathbf{2}} + (x_2-\alpha)^{\mathbf{2}} \\ \text{subject to} & x_1^2 - 1 = 0 \\ & 1 - x_2^2 \leqq 0 \end{array}$$

where α is any real number satisfying $\|\alpha\| \leqq 1$. Let $\overline{x} = (\overline{x}_1,\overline{x}_2)^T = (1,1)^T$, $\overline{\mu} = (\alpha-1)$, $\overline{\pi} = (1-\alpha)$. Verify that $\overline{x}$ is a global minimum for this problem using the result in (i). (P. Mereau and J. A. Paquet, “A sufficient condition for global constrained extrema”, *Int. J. Control*, 17 (1973) 1065–1071).

77. $\theta(x)$; $g_i(x)$, $i = 1$ to m are all real valued convex functions defined over $\mathbf{R}^n$. Consider the NLP

$$\begin{array}{ll} \text{minimize} & \theta(x) \\ \text{subject to} & g_i(x) \leqq 0,\ i = 1 \text{ to } m. \end{array}$$

i) Prove that the set of all optimum solutions of this problem is a convex set.

ii) A real valued function defined on $\mathbf{R}^n$ is said to be a symmetric function if $f(x) = f(Px)$, for all $x \in \mathbf{R}^n$, and P any permutation matrix of order n. If all the functions $\theta(x)$, $g_i(x)$, $i = 1$ to m, are symmetric functions, and the above problem has an optimum solution, prove that it has one in which all the variables are equal.

Exercises 78 to 98 have been suggested to me by Vasant A. Ubhaya.

---

78. Let $\mathbf{J}$ be an interval of the real line. $f(x)$ is a real valued function defined on $\mathbf{J}$. Prove that $f(x)$ is convex iff for any three point x, y, z in $\mathbf{J}$ with $x < y < z$,

$$\text{determinant} \begin{vmatrix} x & f(x) & 1 \\ y & f(y) & 1 \\ z & f(z) & 1 \end{vmatrix} \geqq 0.$$

79. Let $a_1 \geqq a_2 \geqq \ldots \geqq a_n \geqq 0$ and let $f(x)$ be a real valued convex function defined on the interval $[0, a_1]$ with $f(0) = 0$. Show that

$$\sum_{k=1}^{n} (-1)^{\mathbf{k-1}} f(a_k) \geqq f\Big(\sum_{k=1}^{n} (-1)^{\mathbf{k-1}} a_k\Big).$$

(contd.)

(E. F. Beckenbach and R. Bellman, *"Inequalities"*, Springer-Verlag, New York, 1983, and E. M. Wright, "An inequality for convex functions", *American Mathematical Monthly*, 61 (1984) 620-622.)

80. Let $\mathbf{J}$ be a closed interval of the real line. A real valued function $f(x)$ defined on $\mathbf{J}$ is said to be midconvex or Jensen-convex if

$$f\Big(\frac{x+y}{2}\Big) \leqq \frac{1}{2}(f(x)+f(y))$$

for all $x, y \in \mathbf{J}$. Prove that if $f(x)$ is midconvex, then

$$f(\lambda x + (1-\lambda)y) \leqq \lambda f(x) + (1-\lambda) f(y)$$

for all $x, y \in \mathbf{J}$ and all rational numbers λ between 0 and 1. Hence conclude that a continuous function is midconvex iff it is convex. (A. W. Roberts and D. E. Varberg, *"Convex Functions"*, Academic Press, New York, 1973.)

81. Let $\mathbf{\Gamma} \subset \mathbf{R}^n$ be a convex set, and let $f(x)$ be a real valued convex function defined on $\mathbf{\Gamma}$. Let $g(\lambda)$ be a nondecreasing convex function defined on a real interval $\mathbf{J}$ where the range of $f(x)$ is contained in $\mathbf{J}$. Prove that $h(x) = g(f(x))$ is convex. Use this to show the following:

a) If $f(x)$ is a positive concave function defined on $\mathbf{\Gamma}$, then $1/f(x)$ is convex.

b) If $f(x)$ is a nonnegative convex function defined on $\mathbf{\Gamma}$, then $(f(x))^{\mathbf{r}}$ is convex for $r \geqq 1$.

c) If $f(x)$ is a convex function defined on $\mathbf{\Gamma}$, then $\exp(f(x))$ is convex.

82. Let $y = (y_1, \dots, y_{n-1})^T$, and $\mathbf{\Gamma} = \{y : 0 = y_0 < y_1 < \dots < y_{n-1} < y_n = 1\}$. Define, for $j = 0, 1, \dots, n$

$$\begin{aligned} D_j(y) &= \prod(|y_i - y_j| : \text{ over } 0 \leqq i \leqq n,\ i \neq j) \\ &= (-1)^j \prod((y_i - y_j) : \text{ over } 0 \leqq i \leqq n, i \neq j) \end{aligned}$$

where $\prod$ denotes the product sign. Define

$$F(y) = \sum((D_{n-1-2j}(y))^{-1} : \text{ over } 0 \leqq j \leqq \lfloor (n-1)/2 \rfloor).$$

Show that $F(y)$ is a strictly convex function of y over $\mathbf{\Gamma}$. Prove that $y^* = ((\sin(\pi/2n)^{\mathbf{2}}, (\sin(2\pi/2n))^{\mathbf{2}}, \dots, (\sin((n-1)\pi/2n))^{\mathbf{2}})^T$ is the unique optimum solution for the problem of minimizing $F(y)$ over $\mathbf{\Gamma}$, and that $F(y^*) = 2^{\mathbf{2n-2}}$. Prove the following inequalities for all $y \in \mathbf{\Gamma}$.

(i) $\sum((D_j(y))^{-1} : \text{ over } j \text{ odd}, \quad 1 \leqq j \leqq n-1) \geqq 2^{\mathbf{2n-2}}$, if n is even.

(ii) $\sum((D_j(y))^{-1} : \text{ over } j \text{ even}, \quad 0 \leqq j \leqq n-1) \geqq 2^{\mathbf{2n-2}}$, if n is odd.

(iii) $\sum((D_j(y))^{-1} : \text{ over } j \text{ odd}, \quad 1 \leqq j \leqq n) \geqq 2^{\mathbf{2n-2}}$, if n is odd.

(iv) $\sum((D_j(y))^{-1} : \text{ over } \quad 0 \leqq j \leqq n) \geqq 2^{\mathbf{2n-1}}$, if n is odd.

Furthermore prove that each of the above inequalities holds as an equation iff $y = y^*$ defined above. (V. A. Ubhaya "Nonlinear programming, approximation and optimization on infinitely differentiable functions", *Journal of Optimization Theory and Applications*, 29 (1979), 199-213.)

83. Let $\mathbf{S} \subset \mathbf{R}^{n+1}$ be a convex set. Define a set $\mathbf{\Gamma} \subset \mathbf{R}^n$ and a real valued function $f(x)$ on $\mathbf{\Gamma}$ as follows.

$$\mathbf{\Gamma} = \{x \in \mathbf{R}^n : u \in \mathbf{R}^1, (x, u) \in \mathbf{S}\}.$$
$$f(x) = \inf\{u : x \in \mathbf{\Gamma}, (x, u) \in \mathbf{S}\}.$$

Show that $\mathbf{\Gamma}$ is convex and $f(x)$ is a convex function on $\mathbf{\Gamma}$.

84. Let $\mathbf{\Gamma} \subset \mathbf{R}^n$ and $f(x)$ be any real valued function defined on $\mathbf{\Gamma}$. The epigraph $\mathbf{E}(f)$ of $f(x)$ is a subset of $\mathbf{R}^{n+1}$ defined as in Appendix 3. Assume that $\mathbf{\Gamma}$ is closed, and show that $\mathbf{E}(f)$ is closed iff $f(x)$ is lower semi-continuous. In particular, $\mathbf{E}(f)$ is closed if $\mathbf{\Gamma}$ is closed and $f(x)$ is continuous.

85. Let $\mathbf{\Gamma} \subset \mathbf{R}^n$ be a convex set and $f(x)$ be a real valued bounded function defined on $\mathbf{\Gamma}$. The greatest convex minorant $\overline{f}(x)$ of $f(x)$ is the largest convex function which does not exceed $f(x)$ at any point in $\mathbf{\Gamma}$, viz.,

$$\overline{f}(x) = \sup\{h(x) : h(y) \text{ is convex and } h(y) \leqq f(y) \text{ for all } y \text{ in } \mathbf{\Gamma}\}, x \in \mathbf{\Gamma}.$$

Show that $\overline{f}(x)$ defined in this way is, indeed, convex. If $\mathbf{E}(f)$ is the epigraph of $f(x)$ then show that

$$\overline{f}(x) = \inf\{u : x \in \mathbf{\Gamma}, (x, u) \in \text{co}(\mathbf{E}(f))\},$$

where $\text{co}(\mathbf{E}(f))$ is the convex hull of $\mathbf{E}(f)$, i. e., the smallest convex subset of $\mathbf{R}^{n+1}$ containing $\mathbf{E}(f)$.

86. Let $\mathbf{\Gamma} \subset \mathbf{R}^n$ be convex and $f(x)$ be a real valued convex function defined on $\mathbf{\Gamma}$. Assume $0 \leqq f(x) < 1$. Show that $(1+f(x))^{\mathbf{1/2}}$ and $(1-f(x))^{-\mathbf{1/2}}$ are convex functions on $\mathbf{\Gamma}$. Is $((1 + f(x))/(1 - f(x)))^{\mathbf{1/2}}$ convex?

87. Let $f(x)$ be a real homogeneous polynomial of degree 2 defined on $\mathbf{R}^n$, i. e.,

$$f(x) = \sum_i a_i x_i^{\mathbf{2}} + \sum_{i<j} b_{ij} x_i x_j,$$

where $x = (x_1, x_2, \ldots, x_n)$; and a_i, b_{ij} are given numbers. Show that $f(x)$ is convex iff $f(x)$ is nonnegative on $\mathbf{R}^n$.

88. Let $f(x)$ be a real valued function defined on the interval $\mathbf{J} = [0,1]$. The nth ($n \geqq 1$) Bernstein polynomial for $f(x)$ is defined by

$$B_n(f,x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} x^{\mathbf{k}}(1-x)^{\mathbf{n-k}}.$$

Note that $B_n(f,0) = f(0)$ and $B_n(f,1) = f(1)$. Show the following:

(a) If $f(x)$ is nondecreasing on $\mathbf{J}$, then $B_n(f,x)$ is nondecreasing on $\mathbf{J}$.

(b) If $f(x)$ is convex on $\mathbf{J}$, then $B_n(f,x)$ is convex on $\mathbf{J}$. In this case, $B_{n-1}(f,x) \geqq B_n(f,x)$ for $0 < x < 1$ and $n \geqq 2$.

(c) If $f(x)$ is bounded on $\mathbf{J}$, then $B_n(f,x) \to f(x)$ as $n \to \infty$ at any point x in $\mathbf{J}$ at which $f(x)$ is continuous. Furthermore, if $f(x)$ is continuous on $\mathbf{J}$, then this convergence is uniform on $\mathbf{J}$. Hence conclude that the class of nondecreasing (convex) polynomials on $\mathbf{J}$ are dense in the class of continuous nondecreasing (convex) functions on $\mathbf{J}$ when the uniform norm $\|f\| = \max\{|f(x)| : x \in \mathbf{J}\}$ is used to generate a metric for the set of continuous functions $f(x)$.

(P. J. Davis, *"Interpolation and Approximation"*, Dover, New York, 1975).

89. Let $\mathbf{\Gamma}$ be a convex subset of $\mathbf{R}^n$ and $f(x)$ a real valued function defined on $\mathbf{\Gamma}$. $f(x)$ is said to be a quasiconvex function if $\{x \in \mathbf{\Gamma} : f(x) \leqq \alpha\}$ is a convex set for all real α.

A real valued function $g(x)$, defined on a convex set is said to be quasiconcave, if $-g(x)$ is quasiconvex.

Show that $f(x)$ is quasi-convex on $\mathbf{\Gamma}$ iff

$$f(\lambda x + (1-\lambda)y) \leqq \max\{f(x), f(y)\}$$

holds for all $x, y \in \mathbf{\Gamma}$, all $0 \leqq \lambda \leqq 1$.

90. The following result is well known:

Let $\mathbf{\Gamma} \subset \mathbf{R}^n$ and $\mathbf{\Delta} \subset \mathbf{R}^m$ be compact convex subsets. Let $h(x,y)$ be a continuous real valued function defined on $\mathbf{\Gamma} \times \mathbf{\Delta}$ be such that, for each $y \in \mathbf{\Delta}$, $h(x,y)$ is a quasiconcave function of x; and for each $x \in \mathbf{\Gamma}$, $h(x,y)$ is a quasiconvex function of y. Then,

$$\min_{y\in\Delta}\max_{x\in\Gamma} h(x,y) = \max_{x\in\Gamma}\min_{y\in\Delta} h(x,y).$$

(See, e. g., H. Nikaidô, "On Von Neumann's minimax theorem", *Pacific Journal of Mathematics*, 4 (1954), 65–72, for the above result and M. Sion, "On general minimax theorems", *Pacific Journal of Mathematics*, 8 (1958), 171–176, for more general versions.) Using the above result, derive the following:

Let $\mathbf{K}$, $\mathbf{P}$ be bounded subsets of $\mathbf{R}^2$ with the property that there exists a $\delta > 0$ such that $u_1 \geqq \delta$ for all $u = (u_1, u_2)^T \in \mathbf{K}$ and $v_1 \geqq \delta$ for all $v = (v_1, v_2)^T \in \mathbf{P}$. Then,

$$\inf_{v\in\mathbf{P}}\left\{\sup_{u\in\mathbf{K}}\left\{\frac{u_2+v_2}{u_1+v_1}\right\}\right\} = \sup_{u\in\mathbf{K}}\left\{\inf_{v\in\mathbf{P}}\left\{\frac{u_2+v_2}{u_1+v_1}\right\}\right\} \qquad \text{(contd.)}$$

(V. A. Ubhaya, "Almost monotone approximation in L_∞", *Journal of Mathematical Analysis and Applications*, 49 (1975), 659–679).

91. A metric on $\mathbf{R}^n$ is a real valued function $d(x, y)$ defined over ordered pairs of points in $\mathbf{R}^n$ satisfying the following properties.

$$\begin{aligned} d(x,y) &\geqq 0, \quad \text{for all } x, y \in \mathbf{R}^n \\ d(x,y) &= 0, \quad \text{iff } x = y \\ d(x,y) &> 0, \quad \text{iff } x \neq y \\ d(x,y) &= d(y,x), \quad \text{for all } x, y \in \mathbf{R}^n \\ d(x,y) + d(y,z) &\geqq d(x,z), \quad \text{for all } x, y, z \in \mathbf{R}^n. \end{aligned}$$

Let $d(x, y)$ be a metric on $\mathbf{R}^n$ and $\mathbf{F}$ be a nonempty subset of $\mathbf{R}^n$. For each x in $\mathbf{R}^n$, let $f(x)$ denote the minimum distance between x and $\mathbf{F}$, viz.,

$$f(x) = \inf\{d(x, u) : u \in \mathbf{F}\}.$$

Show that

$$|f(x) - f(y)| \leqq d(x, y)$$

for all x, y in $\mathbf{R}^n$. Thus f is nonexpansive.

92. Let

$$d'(f, g) = \max\{w_i |f_i - g_i| : 1 \leqq i \leqq n\},$$

denote the distance between two vectors $f = (f_1, f_2, \ldots, f_n)$ and $g = (g_1, g_2, \ldots, g_n)$, where $w = (w_1, w_2, \ldots, w_n) > 0$ is a given weight vector. A vector g is called isotonic if $g_i \leqq g_{i+1}$, $1 \leqq i < n$. Given a vector f, the problem is to find an isotonic vector g which minimizes $d'(f, g)$. Such a g, called an optimal vector, is not unique in general. Denote the minimum of $d'(f, g)$ over isotonic vectors g, for given f, by Δ.

Define the following quantities:

$$\begin{aligned} \theta &= \max\Big\{\frac{w_i w_j}{w_i + w_j}(f_i - f_j) : 1 \leqq i \leqq j \leqq n\Big\}, \\ \underline{g}_i &= \max\{f_j - \theta/w_j : 1 \leqq j \leqq i\}, 1 \leqq i \leqq n, \\ \overline{g}_i &= \min\{f_j + \theta/w_j : 1 \leqq j \leqq n\}, 1 \leqq i \leqq n. \end{aligned}$$

Show the following: (a) Duality: $\Delta = \theta$, (b) Optimality: $\underline{g}$ and $\overline{g}$ are optimal vectors with $\underline{g} \leqq \overline{g}$. Furthermore, an isotonic g is an optimal vector iff $\underline{g} \leqq g \leqq \overline{g}$. (V. A. Ubhaya, "Isotone optimization, I, II", *Journal of Approximation Theory*, 12 (1974), 146–159 and 315–331).

93. Consider Exercise 92 with $w_i = 1$ for all i and define

$$d(f, g) = \max\{|f_i - g_i| : 1 \leqq i \leqq n\}.$$

$$\underline{h}_i = \max\{f_j : 1 \leqq j \leqq i\},\ 1 \leqq i \leqq n,$$
$$\overline{h}_i = \min\{f_j : i \leqq j \leqq n\},\ 1 \leqq i \leqq n.$$

Show the following:

$$\theta = \max\{(\underline{h}_i - f_i) : 1 \leqq i \leqq n\} = \max\{(f_i - \overline{h}_i) : 1 \leqq i \leqq n\}$$

and

$$\underline{g}_i = \underline{h}_i - \theta,\ \overline{g}_i = \overline{h}_i + \theta.$$

Construct an $O(n)$ algorithm for computing optimal vectors $\underline{g}$ and $\overline{g}$.

94. Let $d(f, g)$, as defined in Exercise 93, denote the distance between two vectors f and g. A vector g is called **convex** if it satisfies

$$g_{i-1} - 2g_i + g_{i+1} \geqq 0,\ 1 < i < n,$$

or more generally,

$$a_{i-1}g_{i-1} - (a_{i-1} + a_i)g_i + a_i g_{i+1} \geqq 0,\ 1 < i < n,$$

where a_i, $1 \leqq i < n$, are given positive numbers. Given a vector f, the problem is to find a convex vector g, called an optimal vector, which minimizes $d(f, g)$. Let Δ denote the minimum of $d(f, g)$ over convex vectors g, for given f.

The greatest convex minorant $\overline{h} = (\overline{h}_1, \overline{h}_2, \ldots, \overline{h}_n)$ of f is the largest convex vector (i. e. satisfying the above condition) which does not exceed f. (See Exercise 85). Show the following: $\Delta = (1/2)d(f, \overline{h})$ and $\overline{g} = \overline{h} + e\Delta$ is the maximal optimal vector, i. e., for all optimal vectors g it is true that $\overline{g} \geqq g$. Construct an $O(n)$ algorithm for computing $\overline{h}$ and then $\overline{g}$. (V. A. Ubhaya, "An $O(n)$ algorithm for discrete n-point convex approximation with applications to continuous case", *Journal of Mathematical Analysis and Applications*, 72 (1979), 338–354.)

95. In connection with Exercise 94 consider the following LP.

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{n} x_i \\ \text{subject to} & -x_{i-1} + 2x_i - x_{i+1} \geqq -f_{i-1} + 2f_i - f_{i+1},\ 1 < i < n \\ & x_i \geqq 0,\ 1 \leqq i \leqq n. \end{array}$$

Show that the LP has a unique optimal solution x^* and the quantities defined in Exercise 94 for the first convexity constraint are given by

$$\Delta = (1/2)\max\{x_i^* : 1 \leqq i \leqq n\},$$
$$\overline{g}_i = \Delta - x_i^* + f_i,\ 1 \leqq i \leqq n.$$

Devise a special pivoting strategy in conjunction with the Dual Simplex Algorithm of linear programming to solve the above LP in $O(n)$ computing time. (V. A. Ubhaya, "Linear time algorithms for convex and monotone approximation", *Computers and Mathematics with Applications*, An International Journal, 9 (1983), 633–643.)

96. A vector $g = (g_1, g_2, \ldots, g_n)$ is called **quasiconvex** if

$$g_j \leqq \max\{g_p, g_q\},$$

for all j with $p \leqq j \leqq q$ and for all $1 \leqq p \leqq q \leqq n$. Show that g is quasiconvex iff there exists $1 \leqq r \leqq n$ such that $g_i \geqq g_{i+1}$ for $1 \leqq i < r$ and $g_i \leqq g_{i+1}$ for $r \leqq i < n$. Show that the set of all quasiconvex vectors is a closed nonconvex cone, but the set of all isotonic or convex vectors is a closed convex cone.

Let $d(f, g)$ be as defined in Exercise 93. Given a vector f, consider the problem of finding a quasiconvex vector g, called an optimal vector, which minimizes $d(f, g)$. Show that there exist two optimal vectors $\underline{g}$ and $\overline{g}$ with $\underline{g} \leqq \overline{g}$ so that any quasiconvex vector g with $\underline{g} \leqq g \leqq \overline{g}$ is also an optimal vector. Furthermore, $\overline{g}$ is the maximal optimal vector, i. e., for all optimal vectors g it is true that $\overline{g} \geqq g$. Construct an $O(n)$ algorithm to compute $\underline{g}$ and $\overline{g}$. (V. A. Ubhaya, "Quasi-convex optimization", *Journal of Mathematical Analysis and Applications*, 116 (1986), 439–449.)

97. Exercise 93 to 96 involved finding an isotonic, convex or quasiconvex vector g minimizing $d(f, g)$ given the vector f. Such an optimal vector g is not unique in general. For each f it is of interest to select an optimal vector f' (in each of three cases) so that f' is least sensitive to perturbations in f. Specifically, the following two conditions may be imposed on the selection f' for f.

(i) $d(f', h') \leqq C\ d(f, h)$ holds for all vectors f, h for some least number C. This makes the mapping T defined by $T(f) = f'$ Lipschitzian with constant C.

(ii) The selection f' is such that the number C is smallest among all selections of optimal vectors for f. This makes T optimal.

Thus a mapping T satisfying (i) and (ii) may be called an optimal Lipschitzian selection operator.

Show that optimal Lipschitzian selections are possible for the three problems as shown below. Here $\underline{g}$ and $\overline{g}$ are as defined in Exercises 93, 94 and 96.

(a) Isotonic problem: $T(f) = f' = (1/2)(\underline{g} + \overline{g})$ and $C = 1$.
(b) Convex problem: $T(f) = f' = \overline{g}$ and $C = 2$.
(c) Quasiconvex problem: $T(f) = f' = \overline{g}$ and $C = 2$.

(V. A. Ubhaya, "Lipschitz condition on minimum norm problems on bounded functions", *Journal of Approximation Theory*, 45 (1985), 201–218, also "Optimal Lipschitzian selection operator in quasi-convex optimization", *Journal of Mathematical Analysis and Applications*, to appear).

98. Prove that the functions $\log x$ and $x \log x$ are respectively concave and convex on the interval $0 < x < \infty$. Using this, eastablish the following inequality: if $x > 0$, $y > 0$, both $x, y \in \mathbf{R}^1$, then

$$\log\frac{x+y}{2} \leqq \frac{x\log x + y\log y}{x+y} \leqq \log\frac{\mathbf{x^2}+y^\mathbf{2}}{x+y}.$$

---

99. (i) Let $\theta(x)$, $h_i(x)$, $i = 1$ to m be continuously differentiable real valued functions defined over $\mathbf{R}^n$. Consider the nonlinear program.

$$\begin{array}{ll} \text{minimize} & \theta(x) \\ \text{subject to} & h_i(x) = 0,\ i = 1 \text{ to } m. \end{array}$$

Prove that if $\hat{x}$ is a feasible solution to this nonlinear program which is a local minimum for this NLP, then the set of vectors $\{\nabla\theta(\hat{x}); \nabla h_i(\hat{x}),\ i = 1 \text{ to } m\}$ must be a linearly dependent set.

(ii) Consider the following NLP

$$\begin{array}{ll} \text{minimize} & \theta(x) \\ \text{subject to} & h_i(x) = 0,\ i = 1 \text{ to } m \\ & g_p(x) \geqq 0,\ p = 1 \text{ to } t \end{array}$$

where $\theta(x)$, $h_i(x)$, $g_p(x)$ are all continuously differentiable real valued functions defined over $\mathbf{R}^n$. Let $\overline{x}$ be a feasible solution to this NLP. Define $\mathbf{P}(\overline{x}) = \{p : p = 1 \text{ to } t, g_p(\overline{x}) = 0\}$. If $\overline{x}$ is a local minimum for this NLP, prove that the set of vectors

$$\{\nabla\theta(\overline{x})\} \cup \{\nabla h_i(\overline{x}) : i = 1 \text{ to } m\} \cup \{\nabla g_p(\overline{x}) : p \in \mathbf{P}(\overline{x})\}$$

must be a linearly dependent set. In addition, prove that there must exist a linear dependence relation for this set of vectors of the form

$$\delta_0\nabla\theta(\overline{x}) - \sum_{i=1}^{m} \mu_i \nabla h_i(\overline{x}) - \sum_{p\in\mathbf{P}(\overline{x})} \pi_p g_p(\overline{x}) = 0$$

where $(\delta_0,\ \mu_i$ for $i = 1$ to m; π_p to $p \in \mathbf{P}(\overline{x})) \neq 0$ and $(\delta_0, \pi_p : p \in \mathbf{P}(\overline{x})) \geqq 0$.

100. Consider the following general QP

$$\begin{array}{lrl} \text{minimize} & Q(x) & = cx + (1/2)x^T Dx \\ \text{subject to} & Ax & \geqq b \\ & x & \geqq 0. \end{array}$$

Define the following:

$\mathbf{K} =$ Set of feasible solutions of this problem.

$\mathbf{L} =$ Set of all local minima for this problem.

$\mathbf{G} =$ Set of all global minima for this problem.

If **K** is bounded, prove that each of the sets **L** and **G**, is a union of a finite number of convex polyhedra. Is this result also true when **K** is not bounded?

101. Maximum Area Hexagon of Diameter One: A problem which has long intrigued mathematicians if finding the maximum area convex polygon in $\mathbf{R}^2$ with an even number of sides, and an upper bound on its diameter. The diameter of a convex polygon is defined to be the maximum distance between any pair of points in it. When the number of sides is odd, the regular polygon has the maximum area; but this may not be true when the number of sides is even.

Consider the special case of this problem, of finding the maximum area hexagon of diameter one. Clearly, without any loss of generality, one can assume that two of the vertices of the hexagon are $(0, 0)$ and $(0, x_1)$; and that the other vertices have coordinates and positions as entered in the following figure,

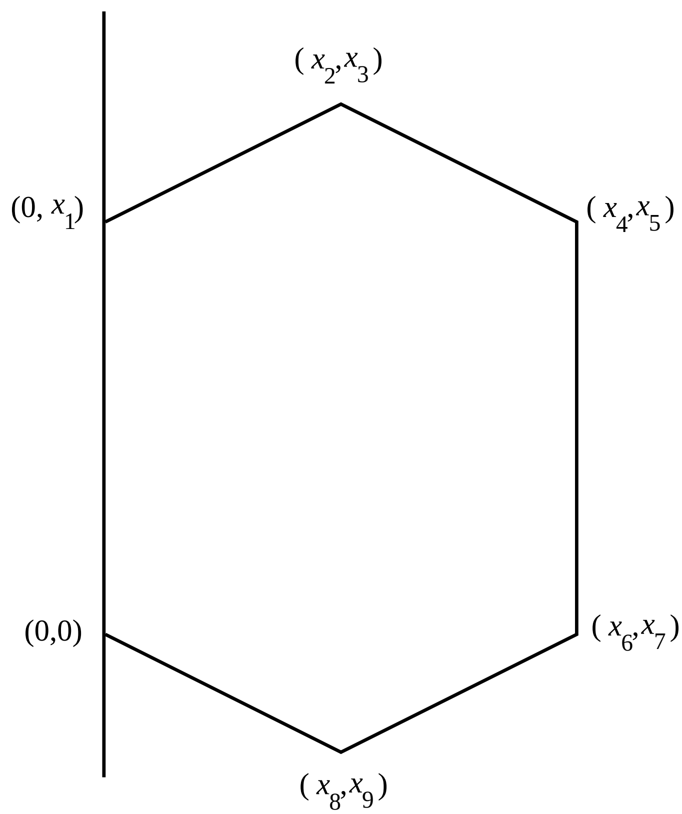

Figure 23

where x_2, x_4, x_6, x_8 are all $\geqq 0$. Formulate the problem of finding the maximum area hexagon of diameter one, as a nonlinear program in terms of the variables x_1 to x_9. Check whether your model is a convex or nonconvex NLP. Write down the necessary optimality conditions for your problem. Solve it on a computer using one of the algorithms discussed in this text.

102. Let $g_i(x)$ be a differentiable convex function defined on $\mathbf{R}^n$ for $i = 1$ to m. Let $\overline{x}$ be a feasible solution of the system

$$g_i(x) \leqq 0, \;\; i = 1 \text{ to } m$$

and let $\mathbf{J}(\overline{x}) = \{i : g_i(\overline{x}) = 0\}$. Prove that the system: $g_i(x) < 0$, $i = 1$ to m, has a feasible solution iff the objective value in the following LP, in which the variables are λ, $d = (d_1, \ldots, d_n)^T$, is unbounded above.

$$\begin{array}{ll} \text{minimize} & \lambda \\ \text{subject to} & \nabla g_i(\overline{x}) d + \lambda \leqq 0, \;\; i \in \mathbf{J}(\overline{x}). \end{array}$$

103. Let $g_i(x)$ be a differentiable convex function defined on $\mathbf{R}^n$ for $i = 1$ to m. Let $\overline{x}$ be a feasible solution of the system

$$g_i(x) \leqq 0, \;\; i = 1 \text{ to } m.$$

Prove that the system: $g_i(x) < 0$, $i = 1$ to m, has a feasible solution, iff the following system has no feasible solution $\pi = (\pi_1, \ldots, \pi_m)$.

$$\begin{aligned} \sum_{i=1}^{m} \pi_i \nabla g_i(\overline{x}) &= 0 \\ \pi &\geq 0. \end{aligned}$$

104. Let $A(m \times n)$, $B(m \times p)$, $a(1 \times n)$, $b(1 \times p)$ be given matrices. Prove that exactly one of the following two systems (I), (II) has a feasible solution, and the other is infeasible.

$$\begin{array}{ccc} \text{(I)} & \qquad & \text{(II)} \\ \hline Ax + By = 0 & & \pi A \leqq a \\ ax + by < 0 & & \pi B = b \\ x \geqq 0 & & \end{array}$$

105. Let $A(m \times n)$, $B(m \times p)$, $a(1 \times n)$, $b(1 \times p)$ be given matrices where b is in the linear hull of the row vectors of B. Prove that exactly one of the following systems (I), (II) has a feasible solution, and the other is infeasible.

$$\begin{array}{ccc} \text{(I)} & \qquad & \text{(II)} \\ \hline Ax + By = 0 & & \pi A < a \\ ax + by \leqq 0 & & \pi B = b \\ x \geq 0 & & \end{array}$$

106. Consider the following NLP involving the vectors of decision variables $x \in \mathbf{R}^{n_1}$, and $y \in \mathbf{R}^{n_2}$

$$\begin{array}{lll} \text{minimize} & h(x,y) = \theta(x) + cy & \\ \text{subject to} & g(x) + ay = b^1 & (m_1 \text{ constraints}) \\ & Bx + Dy = b^2 & (m_2 \text{ constraints}) \\ & l^1 \leqq x \leqq u^1 & \\ & l^2 \leqq y \leqq u^2 & \end{array}$$

where $\theta(x)$, $g(x)$ are continuously differentiable functions. Given $\overline{x} \in \mathbf{R}^{n_1}$, $\overline{y} \in \mathbf{R}^{n_2}$ satisfying the bound constraints in the NLP, define the following LP, which comes from a linearization of the NLP around $(\overline{x}, \overline{y})$.

$$\begin{array}{ll} \text{minimize} & \nabla\theta(\overline{x})d + cy \\ \text{subject to} & \nabla g(\overline{x})d + Ay = b^1 - g(\overline{x}) \\ & Bd + Dy = b^2 - B\overline{x} \end{array}$$

$$\max\{l^1_j - \overline{x}_j, s_j\} \leqq d_j \leqq \min\{u^1_j - \overline{x}_j, s_j\}, \;\; j = 1 \text{ to } n_1$$
$$l^2 \leqq y \leqq u^2$$

where $d = x - \overline{x}$, $s = (s_j) \in \mathbf{R}^{n_1}$, $s > 0$ is a vector of small positive numbers used to bound d in the LP to keep the linearization reasonably accurate. Prove the following

i) If $(\overline{x}, \overline{y})$ is feasible to the NLP, $(d, y) = (0, \overline{y})$ is feasible to the above LP for any $s > 0$.

ii) If the constraint matrix of the LP has full row rank, and $(\overline{x}, \overline{y})$ is a feasible solution of the NLP, then $(0, \overline{y})$ is an optimum solution of the above LP iff $(\overline{x}, \overline{y})$ is a KKT point for the NLP.

iii) Let $(\overline{x}, \overline{y})$ be a feasible solution for the NLP, and suppose $(0, \overline{y})$ is not an optimum solution for the above LP. If (d^0, y^0) is an optimum solution for the LP, then $\nabla\theta(x)d^0 + c(y^0 - \overline{y}) < 0$, that is, $(d^0, y^0 - \overline{y})$ is a descent direction for the NLP at the point $(\overline{x}, \overline{y})$. (See F. Palacios-Gomez, L. Lasdon and M. Engquist, "Nonlinear optimization by successive linear programming", *Management Science*, 28, 10 (October 1982) 1106–1120.)

107. Consider the following NLP

$$\begin{array}{ll} \text{minimize} & Q(x) = cx + \frac{1}{2}x^T Dx \\ \text{subject to} & \|x\| \leqq \delta \end{array}$$

where D is a PD symmetric matrix of order n and $\delta > 0$. Write down the KKT optimality conditions for this problem. Prove that the optimum solution of this problem is $x(\lambda) = -(D + \lambda l)^{-1}c^T$ for the unique $\lambda \geqq 0$ such that $\|x(\lambda)\| = \delta$; unless $\|x(0)\| \leqq \delta$, in which case, $x(0)$ is the optimum solution.

108. Let $f(x) = f_1(x), \ldots, f_n(x))^T$ where each $f_j(x)$ is a continuous function defined ofer $\mathbf{R}^n$. Let $\mathbf{K}$ be a closed convex cone in $\mathbf{R}^n$. Define the polar cone of $\mathbf{K}$ to be $\mathbf{K}^* = \{y : y \in \mathbf{R}^n, y^T x \geqq 0 \text{ for all } x \in \mathbf{K}\}$. (For example, if $\mathbf{K}$ is the nonnegative orthant, $\mathbf{K}^*$ is tha same. Let $\mathbf{J} \subset \{1, \ldots, n\}$. If $\mathbf{K}$ is the orthant $\{x : x = (x_j) \in \mathbf{R}^n, x_j \geqq 0 \text{ for } j \notin \mathbf{J}, x_j \leqq 0 \text{ for } j \in \mathbf{J}\}$, then $\mathbf{K}^*$ is again $\mathbf{K}$ itself.)

The generalized complementary problem corresponding to $f(x)$ and $\mathbf{K}$ is to find x satisfying

$$x \in \mathbf{K},\ f(x) \in \mathbf{K}^*,\ x^T f(x) = 0 \tag{101}$$

using the hypothesis that $\mathbf{K}$ is a closed convex cone, prove that the generalized complementarity problem (101) is equivalent to the variational inequality problem: find $x^* \in \mathbf{K}$ satisfying

$$(x - x^*)^T f(x^*) \geqq 0 \text{ for all } x \in \mathbf{K} \tag{102}$$

(see Karamardian [1.14]).

109. Let $\mathbf{K}$, $\mathbf{K}^*$, $f(x)$ be defined as in the previous Exercise 108. For any $x \in \mathbf{R}^n$ define $P_{\mathbf{K}}(x)$ to be the projection of x into $\mathbf{K}$ (i. e., the nearest point in $\mathbf{K}$ to x, in terms of the usual Euclidean distance). Prove that a solution $x^* \in \mathbf{K}$ to the variational inequality problem (102), can be characterized by the relation

$$x^* = P_{\mathbf{K}}(x^* - \rho f(x^*))$$

where ρ is a positive constant. Using this, show that the generalized complementarity problem (101) can be formulated as the fixed point problem of finding $x \in \mathbf{K}$ satisfying

$$x = g(x) \tag{103}$$

where $g(x) = \lambda P_{\mathbf{K}}(x - \rho f(x)) + (1 - \lambda)x$, with a constant $\rho > 0$ and $0 < \lambda \leqq 1$. Here λ is known as the relaxation factor used after the projection.

Study the application of the successive substitution method for solving (103). This method will begin with a given $x^0 \in \mathbf{K}$, and generate the sequence of points $\{x^r : r = 0, 1, \ldots\}$ using the iteration, $x^{r+1} = g(x^r)$. The iterative methods discussed in Sections 9.3, 9.4, 9.5 are special cases of this general approach. Study the convergence properties of the sequence of points generated under this method (M. Aslam Noor, and K. Inayat Noor, "Iterative methods for variational inequalities and nonlinear programming", *Operations Research Verf.*, 31 (1979) 455–463).

110. Let $\mathbf{K} \subset \mathbf{R}^n$ be convex and let $f(x) = (f_1(x), \ldots, f_n(x))^T$, where each $f_i(x)$ is a continuous real valued function defined over $\mathbf{K}$. Define a point $\overline{x} \in \mathbf{K}$ to be a **critical point** for the pair $(f, \mathbf{K})$ if $y = \overline{x}$ minimizes $(f(\overline{x}))^T y$ over $y \in \mathbf{K}$. Let $\mathbf{\Gamma}(f, \mathbf{K})$ denote the set of all critical points for the pair $(f, \mathbf{K})$.

Let $\mathbf{\Delta}(f, \mathbf{K})$ denote the set of all points $\overline{x} \in \mathbf{K}$ such that $y = \overline{x}$ minimizes $\|y - \overline{x} + f(\overline{x})\|$ over $y \in \mathbf{K}$. Prove that $\mathbf{\Delta}(f, \mathbf{K}) \subset \mathbf{\Gamma}(f, \mathbf{K})$.

Let $\theta(x)$ be a real valued continuously differentiable function defined over $\mathbf{K}$. Consider the NLP

$$\begin{array}{ll} \text{minimize} & \theta(x) \\ \text{over} & x \in \mathbf{K}. \end{array}$$

Prove that every stationary point for this NLP is a critical point for the pair $(\nabla\theta(x), \mathbf{K})$. If $\mathbf{K} = \mathbf{R}^n_+ = \{x : x \in \mathbf{R}^n, x \geqq 0\}$, prove that the problem of finding a critical point for the pair $(f, \mathbf{R}^n_+)$ is equivalent to the nonlinear complementarity problem (NLCP): find $x \in \mathbf{R}^n$ satisfying

$$x \geqq 0,\ f(x) \geqq 0,\ x^T f(x) = 0.$$

Let $d \in \mathbf{R}^n$, $d > 0$ be a given vector. Let $\mathbf{D}(\alpha) = \{x : x \in \mathbf{R}^n, x \geqq 0, \text{ and } d^T x \leqq \alpha\}$, for each $\alpha \geqq 0$. If $\mathbf{K} = \mathbf{D}(\alpha)$ for some $\alpha \geqq 0$, prove that $\overline{x} \in \mathbf{D}(\alpha)$ is a critical point for the pair $(f, \mathbf{D}(\alpha))$, iff there is a $w \in \mathbf{R}^n_+$ and $z_0 \geqq 0$ such that,

$$\begin{aligned} f(\overline{x}) = w - dz_0,\ \overline{x}^T w = 0 \\ z_0(\alpha - d^T\overline{x}) = 0. \end{aligned}$$

Also, prove that if $\overline{x}$ is a critical point of $(f, \mathbf{D}(\alpha))$ and $d^T\overline{x} < \alpha$, then $\overline{x}$ is a critical point of $(f, \mathbf{R}^n_+)$. Conversely if $\overline{x} \in \mathbf{\Gamma}(f, \mathbf{R}^n_+)$ and $d^T\overline{x} \leqq \alpha$, then $\overline{x} \in \mathbf{\Gamma}(f, \mathbf{D}(\alpha))$.

Consider the case where $\mathbf{K}$ is nonempty, compact and convex. In this case, for each $x \in \mathbf{K}$, define $h(x)$ to be the y that minimizes $\|y - x + f(x)\|$ over $y \in \mathbf{K}$. Using $h(x)$ and Brower's fixed point theorem show that $(f, \mathbf{K})$ has a critical point.
(B. C. Eaves [3.20])

111. Let $f(x) = (f_1(x), \dots, f_n(x))^T$ where each $f_i(x)$ is a continuous real valued function defined over $\mathbf{R}^n$. Consider the NLCP: find x satisfying

$$x \geqq 0,\ f(x) \geqq 0,\ x^T f(x) = 0.$$

For each $x \geqq 0$, define $h(x)$ to be the y that minimizes $\|y - x + f(x)\|$ over $y \geqq 0$. If $h(x) = (h_i(x))$, show that

$$h_i(x) = \begin{cases} 0, & \text{if } f_i(x) - x_i \geqq 0 \\ x_i - f_i(x), & \text{if } f_i(x) - x_i \leqq 0 \end{cases}$$

Prove that the following conditions are equivalent

i) x solves the NLCP given above,
ii) $x \geqq 0$ and $h(x) = x$,
iii) $x \geqq 0$ and $x \in \mathbf{\Gamma}(f, \mathbf{R}^n_+)$

where $\mathbf{\Gamma}(f, \mathbf{R}^n_+)$ is defined in the previous Exercise 110.

Suppose there is a compact convex set $\mathbf{S} \subset \mathbf{R}^n$ such that for each $x \in \mathbf{R}^n_+ \setminus \mathbf{S}$, there is a $y \in \mathbf{S}$ satisfying $(y - x)^T f(x) < 0$. Under this condition, prove that every fixed point of $h(x)$ lies in the set $\mathbf{S}$.
(R. Saigal and C. Simon [3.67], B. C. Eaves [3.20])

112. (Research Problem): In Section 11.4.1, subsection 5, we described a special direct procedure for obtaining a true optimum solution for an LP, from a given near optimum solution for it. Consider the QP (1.11). Assuming that D is PSD, and that a near optimum feasible solution, $\overline{x}$, is given for it, develop a special direct procedure to obtain a true optimum solution for the QP (1.11), from $\overline{x}$.

113. An economic model leads to the following optimization problem. The decision variables in this problem are $x \in \mathbf{R}^n$, $y \in \mathbf{R}^n$ and $z \in \mathbf{R}^p$. The problem is

$$\begin{array}{ll} \text{minimize} & cx + dy + az \\ \text{subject to} & A_1x + A_2y + A_3z = b \\ & x, y, z \geqq 0 \\ & \text{and } x^Ty = 0 \end{array}$$

where A_1, A_2, A_3 are given matrices of order $m \times n$, $m \times n$, $m \times p$ respectively, and c, d, a, b are given vectors of appropriate dimensions. Formulate this as a mixed integer linear programming problem.

114. Let $x \in \mathbf{R}^1$. Define $\mathbf{F}(x) = \{x^3 + 3x^2 - 9x - 24\}$. Compute a Kakutani fixed point of $\mathbf{F}(x)$ using the algorithm discussed in Section 2.7.8.

7. References

A1 F. H. Clarke, "Generalized Gradients and Applications", *Transactions of the American Mathematical Society*, 205 (1975) 247–262.

A2 R. W. Cottle, F. Giannessi and J. L. Lions, Editors, "Variational Inequalities and Complementarity Problems", *Proceedings of International Summer School held in Sicily,* John Wiley & Sons, Inc., 1978.

A3 J. E. Dennis, Jr. and J. J. More', "Quasi-Newton Methods, Motivation and Theory", *SIAM Review*, 19 (1977) 46–89.

A4 W. Karush, "Minima of Functions of Several Variables with Inequalities as Side Conditions", M. S. Dissertation, Department of Mathematics, University of Chicago, Chicago, Illinois (1939).

A5 H. W. Kuhn, "Solvability and Consistency for Linear Equations and Inequalities", *American Mathematical Monthly*, 63 (1956) 217–232.

A6 H. W. Kuhn and A. W. Tucker, "Nonlinear Programming", *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, J. Neyman (editor), University of California Press, Berkeley, California (1951) 481–492.

A7 H. W. Kuhn, "Nonlinear Programming: A Historical View", in: R. W. Cottle and C. E. Lemke (editors), *Nonlinear Programming: Proceedings of the Joint SIAM-AMS Symposium on Applied Mathematics*, American Mathematical Society, Providence, Rhode Island (1976).

A8 D. G. Luenberger, *Introduction to Linear and Nonlinear Programming,* Second ed., Addison-Wesley, Reading, USA (1984).

A9 A. Majthay, "Optimality Conditions for Quadratic Programming", *Mathematical Programming,* 1 (1971) 359–365.

A10 O. L. Mangasarian, *Nonlinear Programming,* McGraw-Hill, New York (1969).

A11 O. L. Mangasarian and T. H. Shiau, "Lipschitz Continuity of Solutions of Linear Inequalities, Programs and Complementarity Problems", Computer Science Technical Report # 599, University of Wisconsin, Madison, June 1985.

A12 G. P. McCormick, "Second Order Conditions for Constrained Minima", *SIAM Journal on Applied Mathematics*, 15 (1967) 641–652.

A13 N. Z. Shor, *Minimization Methods for Nondifferentiable Functions,* Springer Series in Computational Mathematics 3, Springer-Verlag, New York, 1985.

A14 K. Sydsaeter, "Letter to the Editor on Some Frequently Occuring Errors in the Economic Literature Concerning Problems of Maxima and Minima", *Journal of Economic Theory*, 9 (1974) 464–466.

A15 L. Watson, "Solving the Nonlinear Complementarity Problem by a Homotopy Method", *SIAM Journal on Control and Optimization*, 17 (1979) 36–46.