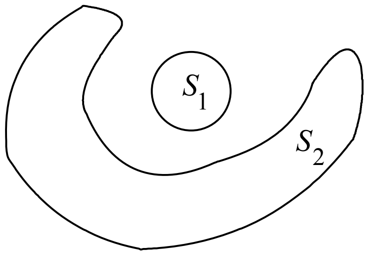
CHAPTER 2:

CONVEX SETS

2.1 Let $x \in conv(S_1 \cap S_2)$. Then there exists $\lambda \in [0,1]$ and $x_1,\ x_2 \in S_1 \cap S_2$ such that $x = \lambda x_1 + (1-\lambda)x_2$. Since x_1 and x_2 are both in S_1, x must be in $conv(S_1)$. Similarly, x must be in $conv(S_2)$. Therefore, $x \in conv(S_1) \cap conv(S_2)$. (Alternatively, since $S_1 \subseteq conv(S_1)$ and $S_2 \subseteq conv(S_2)$, we have $S_1 \cap S_2 \subseteq conv(S_1) \cap conv(S_2)$ or that $conv[S_1 \cap S_2] \subseteq conv(S_1) \cap conv(S_2)$.)

An example in which $conv(S_1 \cap S_2) \neq conv(S_1) \cap conv(S_2)$ is given below:



Here, $conv(S_1 \cap S_2) = \varnothing$, while $conv(S_1) \cap conv(S_2) = S_1$ in this case.

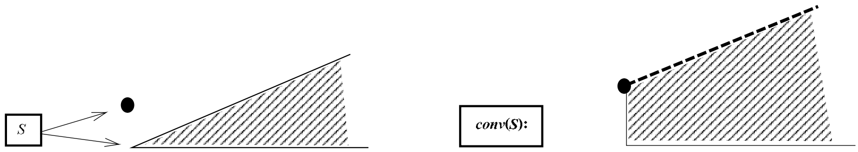
2.2 Let S be of the form $S = \{x : Ax \leq b\}$ in general, where the constraints might include bound restrictions. Since S is a polytope, it is bounded by definition. To show that it is convex, let y and z be any points in S, and let $x = \lambda y + (1-\lambda)z$, for $0 \leq \lambda \leq 1$. Then we have $Ay \leq b$ and $Az \leq b$, which implies that

$$Ax = \lambda Ay + (1-\lambda)Az \leq \lambda b + (1-\lambda)b = b,$$

or that $x \in S$. Hence, S is convex.

Finally, to show that S is closed, consider any sequence $\{x_n\} \to x$ such that $x_n \in S$, $\forall n$. Then we have $Ax_n \leq b$, $\forall n$, or by taking limits as $n \to \infty$, we get $Ax \leq b$, i.e., $x \in S$ as well. Thus S is closed.

2.3 Consider the closed set S shown below along with $conv(S)$, where $conv(S)$ is not closed:



Now, suppose that $S \subseteq \mathbb{R}^p$ is closed. Toward this end, consider any sequence $\{x_n\} \to x$, where $x_n \in conv(S)$, $\forall n$. We must show that $x \in conv(S)$. Since $x_n \in conv(S)$, by definition (using Theorem 2.1.6), we have that we can write $x_n = \sum_{r=1}^{p+1} \lambda_{nr} x_n^r$, where $x_n^r \in S$ for $r = 1,...,p+1$, $\forall n$, and where $\sum_{r=1}^{p+1} \lambda_{nr} = 1$, $\forall n$, with $\lambda_{nr} \geq 0$, $\forall r, n$. Since the λ_{nr}-values as well as the x_n^r-points belong to compact sets, there exists a subsequence K such that $\{\lambda_{nr}\}_K \to \lambda_r$, $\forall r = 1,...,p+1$, and $\{x_n^r\} \to x^r$, $\forall r = 1,...,p+1$. From above, we have taking limits as $n \to \infty$, $n \in K$, that

$$x = \sum_{r=1}^{p+1} \lambda_r x^r \text{, with } \sum_{r=1}^{p+1} \lambda_r = 1,\ \lambda_r \geq 0,\ \forall r = 1,...,p+1,$$

where $x^r \in S$, $\forall r = 1,...,p+1$ since S is closed. Thus by definition, $x \in conv(S)$ and so $conv(S)$ is closed. □

2.7 a. Let y^1 and y^2 belong to *AS*. Thus, $y^1 = Ax^1$ for some $x^1 \in S$ and $y^2 = Ax^2$ for some $x^2 \in S$. Consider $y = \lambda y^1 + (1-\lambda)y^2$, for any $0 \leq \lambda \leq 1$. Then $y = A[\lambda x^1 + (1-\lambda)x^2]$. Thus, letting $x \equiv \lambda x^1 + (1-\lambda)x^2$, we have that $x \in S$ since S is convex and that $y = Ax$. Thus $y \in AS$, and so, *AS* is convex.

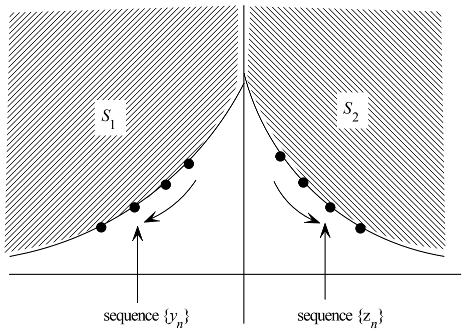
b. If $\alpha \equiv 0$, then $\alpha S \equiv \{0\}$, which is a convex set. Hence, suppose that $\alpha \neq 0$. Let αx^1 and $\alpha x^2 \in \alpha S$, where $x^1 \in S$ and $x^2 \in S$. Consider $\alpha x = \lambda \alpha x^1 + (1-\lambda)\alpha x^2$ for any $0 \leq \lambda \leq 1$. Then, $\alpha x = \alpha[\lambda x^1 + (1-\lambda)x^2]$. Since $\alpha \neq 0$, we have that $x = \lambda x^1 + (1-\lambda)x^2$, or that $x \in S$ since S is convex. Hence $\alpha x \in \alpha S$ for any $0 \leq \lambda \leq 1$, and thus αS is a convex set.

2.8 $S_1 + S_2 = \{(x_1, x_2) : 0 \leq x_1 \leq 1,\ 2 \leq x_2 \leq 3\}$.

$S_1 - S_2 = \{(x_1, x_2) : -1 \le x_1 \le 0,\ -2 \le x_2 \le -1\}$.

2.12 Let $S = S_1 + S_2$. Consider any $y,\ z \in S$, and any $\lambda \in (0,1)$ such that $y = y_1 + y_2$ and $z = z_1 + z_2$, with $\{y_1, z_1\} \subseteq S_1$ and $\{y_2, z_2\} \subseteq S_2$. Then $\lambda y + (1-\lambda)z = \lambda y_1 + \lambda y_2 + (1-\lambda)z_1 + (1-\lambda)z_2$. Since both sets S_1 and S_2 are convex, we have $\lambda y_i + (1-\lambda)z_i \in S_i$, $i = 1, 2$. Therefore, $\lambda y + (1-\lambda)z$ is still a sum of a vector from S_1 and a vector from S_2, and so it is in *S*. Thus *S* is a convex set.

Consider the following example, where S_1 and S_2 are closed, and convex.



Let $x_n = y_n + z_n$, for the sequences $\{y_n\}$ and $\{z_n\}$ shown in the figure, where $\{y_n\} \subseteq S_1$, and $\{z_n\} \subseteq S_2$. Then $\{x_n\} \to 0$ where $x_n \in S$, $\forall n$, but $0 \notin S$. Thus *S* is not closed.

Next, we show that if S_1 is compact and S_2 is closed, then *S* is closed. Consider a convergent sequence $\{x_n\}$ of points from *S*, and let *x* denote its limit. By definition, $x_n = y_n + z_n$, where for each *n*, $y_n \in S_1$ and $z_n \in S_2$. Since $\{y_n\}$ is a sequence of points from a compact set, it must be bounded, and hence it has a convergent subsequence. For notational simplicity and without loss of generality, assume that the sequence $\{y_n\}$ itself is convergent, and let *y* denote its limit. Hence, $y \in S_1$. This result taken together with the convergence of the sequence $\{x_n\}$ implies that $\{z_n\}$ is convergent to *z*, say. The limit, *z*, of $\{z_n\}$ must be in S_2, since S_2 is a closed set. Thus, $x = y + z$, where $y \in S_1$ and $z \in S_2$, and therefore, $x \in S$. This completes the proof. □

2.15 a. First, we show that $conv(S) \subseteq \hat{S}$. For this purpose, let us begin by showing that S_1 and S_2 both belong to $\hat{S}$. Consider the case of S_1 (the case of S_2 is similar). If $x \in S_1$, then $A_1 x \le b_1$, and so, $x \in \hat{S}$ with $y = x$, $z = 0$, $\lambda_1 = 1$, and $\lambda_2 = 0$. Thus $S_1 \cup S_2 \subseteq \hat{S}$, and since $\hat{S}$ is convex, we have that $conv[S_1 \cup S_2] \subseteq \hat{S}$.

Next, we show that $\hat{S} \subseteq conv(S)$. Let $x \in \hat{S}$. Then, there exist vectors y and z such that $x = y + z$, and $A_1 y \le b_1\lambda_1$, $A_2 z \le b_2\lambda_2$ for some $(\lambda_1, \lambda_2) \ge 0$ such that $\lambda_1 + \lambda_2 = 1$. If $\lambda_1 = 0$ or $\lambda_2 = 0$, then we readily obtain $y = 0$ or $z = 0$, respectively (by the boundedness of S_1 and S_2), with $x = z \in S_2$ or $x = y \in S_1$, respectively, which yields $x \in S$, and so $x \in conv(S)$. If $\lambda_1 > 0$ and $\lambda_2 > 0$, then $x = \lambda_1 y_1 + \lambda_2 z_2$, where $y_1 = \frac{1}{\lambda_1} y$ and $z_2 = \frac{1}{\lambda_2} z$. It can be easily verified in this case that $y_1 \in S_1$ and $z_2 \in S_2$, which implies that both vectors y_1 and z_2 are in S. Therefore, x is a convex combination of points in S, and so $x \in conv(S)$. This completes the proof □

b. Now, suppose that S_1 and S_2 are not necessarily bounded. As above, it follows that $conv(S) \subseteq \hat{S}$, and since $\hat{S}$ is closed, we have that $c\ell conv(S) \subseteq \hat{S}$. To complete the proof, we need to show that $\hat{S} \subseteq c\ell conv(S)$. Let $x \in \hat{S}$, where $x = y + z$ with $A_1 y \le b_1\lambda_1$, $A_2 z \le b_2\lambda_2$, for some $(\lambda_1, \lambda_2) \ge 0$ such that $\lambda_1 + \lambda_2 = 1$. If $(\lambda_1, \lambda_2) > 0$, then as above we have that $x \in conv(S)$, so that $x \in c\ell conv(S)$. Thus suppose that $\lambda_1 = 0$ so that $\lambda_2 = 1$ (the case of $\lambda_1 = 1$ and $\lambda_2 = 0$ is similar). Hence, we have $A_1 y \le 0$ and $A_2 z \le b_2$, which implies that y is a recession direction of S_1 and $z \in S_2$ (if S_1 is bounded, then $y \equiv 0$ and then $x = z \in S_2$ yields $x \in c\ell conv(S)$). Let $\bar{y} \in S_1$ and consider the sequence

$$x_n = \lambda_n[\bar{y} + \frac{1}{\lambda_n} y] + (1 - \lambda_n)z, \text{ where } 0 < \lambda_n \le 1 \text{ for all } n.$$

Note that $\bar{y} + \frac{1}{\lambda_n} y \in S_1$, $z \in S_2$, and so $x_n \in conv(S)$, $\forall n$. Moreover, letting $\{\lambda_n\} \to 0^+$, we get that $\{x_n\} \to y + z \equiv x$, and so $x \in c\ell conv(S)$ by definition. This completes the proof. $\square$

2.21 a. The extreme points of S are defined by the intersection of the two defining constraints, which yield upon solving for x_1 and x_2 in terms of x_3 that

$$x_1 = -1 \pm \sqrt{5 - 2x_3},\ x_2 = \frac{3 - x_3 \mp \sqrt{5 - 2x_3}}{2}, \text{ where } x_3 \le \frac{5}{2}.$$

For characterizing the extreme directions of S, first note that for any fixed x_3, we have that S is bounded. Thus, any extreme direction must have $d_3 \neq 0$. Moreover, the maximum value of x_3 over S is readily verified to be bounded. Thus, we can set $d_3 = -1$. Furthermore, if $\bar{x} \equiv (0,0,0)$ and $d = (d_1, d_2, -1)$, then $\bar{x} + \lambda d \in S$, $\forall \lambda > 0$, implies that

$$d_1 + 2d_2 \le 1 \qquad (1)$$

and that $4\lambda d_2 \ge \lambda^2 d_1^2$, i.e., $4d_2 \ge \lambda^2 d_1^2$, $\forall \lambda > 0$. Hence, if $d_1 \neq 0$, then we will have $d_2 \to \infty$, and so (for bounded direction components) we must have $d_1 = 0$ and $d_2 \ge 0$. Thus together with (1), for extreme directions, we can take $d_2 = 0$ or $d_2 = 1/2$, yielding $(0,0,-1)$ and $(0, \frac{1}{2}, -1)$ as the extreme directions of S.

b. Since S is a polyhedron in R^3, its extreme points are feasible solutions defined by the intersection of three linearly independent defining hyperplanes, of which one must be the equality restriction $x_1 + x_2 = 1$. Of the six possible choices of selecting two from the remaining four defining constraints, we get extreme points defined by four such choices (easily verified), which yields $(0,1,\frac{3}{2})$, $(1,0,\frac{3}{2})$, $(0,1,0)$, and $(1,0,0)$ as the four extreme points of S. The extreme directions of S are given by extreme points of $D \equiv \{(d_1, d_2, d_3):$ $d_1 + d_2 + 2d_3 \le 0$, $d_1 + d_2 = 0$, $d_1 + d_2 + d_3 = 1$, $d \ge 0\}$, which is empty. Thus, there are no extreme directions of S (i.e., S is bounded).

c. From a plot of S, it is readily seen that the extreme points of S are given by (0, 0), plus all point on the circle boundary $x_1^2 + x_2^2 = 2$ that lie between the points $(-\sqrt{2/5},\ 2\sqrt{2/5})$ and $(\sqrt{2/5},\ 2\sqrt{2/5})$, including the two end-points. Furthermore, since S is bounded, it has no extreme direction.

2.24 By plotting (or examining pairs of linearly independent active constraints), we have that the extreme points of S are given by (0, 0), (3, 0), and (0, 2). Furthermore, the extreme directions of S are given by extreme points of $D = \{(d_1, d_2) : \ -d_1 + 2d_2 \le 0 \quad d_1 - 3d_2 \le 0, \quad d_1 + d_2 = 1, \quad d \ge 0\}$, which are readily obtained as $(\frac{2}{3}, \frac{1}{3})$ and $(\frac{3}{4}, \frac{1}{4})$. Now, let

$$\begin{bmatrix} 4 \\ 1 \end{bmatrix} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix} + \lambda \begin{bmatrix} 3/4 \\ 1/4 \end{bmatrix}, \text{ where } \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix} = \mu \begin{bmatrix} 3 \\ 0 \end{bmatrix} + (1-\mu) \begin{bmatrix} 0 \\ 2 \end{bmatrix},$$

for $(\mu, \lambda) > 0$. Solving, we get $\mu = 7/9$ and $\lambda = 20/9$, which yields

$$\begin{bmatrix} 4 \\ 1 \end{bmatrix} = \frac{7}{9} \begin{bmatrix} 3 \\ 0 \end{bmatrix} + \frac{2}{9} \begin{bmatrix} 0 \\ 2 \end{bmatrix} + \frac{20}{9} \begin{bmatrix} 3/4 \\ 1/4 \end{bmatrix}.$$

2.31 The following result from linear algebra is very useful in this proof:
(∗) An $(m+1) \times (m+1)$ matrix G with a row of ones is invertible if and only if the remaining m rows of G are linearly independent. In other words, if $G = \begin{bmatrix} B & a \\ e^t & 1 \end{bmatrix}$, where B is an $m \times m$ matrix, a is an $m \times 1$ vector, and e is an $m \times 1$ vector of ones, then G is invertible if and only if B is invertible. Moreover, if G is invertible, then

$$G^{-1} = \begin{bmatrix} M & g \\ h^t & f \end{bmatrix}, \text{ where } M = B^{-1}(I + \frac{1}{\alpha} a e^t B^{-1}),\ g = -\frac{1}{\alpha} B^{-1} a,$$

$$h^t = -\frac{1}{\alpha} e^t B^{-1}, \text{ and } f = \frac{1}{\alpha}, \text{ and where } \alpha = 1 - e^t B^{-1} a.$$

By Theorem 2.6.4, an n-dimensional vector d is an extreme point of D if and only if the matrix $\begin{bmatrix} A \\ e^t \end{bmatrix}$ can be decomposed into $[B_D\ N_D]$ such that $\begin{bmatrix} d_B \\ d_N \end{bmatrix}$, where $d_N = 0$ and $d_B = B_D^{-1} b_D \ge 0$, where $b_D = \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix}$. From Property (∗) above, the matrix $\begin{bmatrix} A \\ e^t \end{bmatrix}$ can be decomposed into $[B_D\ N_D]$, where B_D is a nonsingular matrix, if and only if A can be decomposed into $[B\ N]$, where B is an $m \times m$ invertible matrix. Thus, the matrix B_D must necessarily be of the form $\begin{bmatrix} B & a_j \\ e^t & 1 \end{bmatrix}$, where B is an $m \times m$ invertible submatrix of A. By applying the above equation for the inverse of G, we obtain

$$d_B = B_D^{-1} b_D = \begin{bmatrix} -\frac{1}{\alpha} B^{-1} a_j \\ \frac{1}{\alpha} \end{bmatrix} = \frac{1}{\alpha} \begin{bmatrix} -B^{-1} a_j \\ 1 \end{bmatrix},$$

where $\alpha = 1 - e^t B^{-1} a_j$. Notice that $d_B \geq 0$ if and only if $\alpha > 0$ and $B^{-1} a_j \leq 0$. This result, together with Theorem 2.6.6, leads to the conclusion that d is an extreme point of D if and only if d is an extreme direction of S.

Thus, for characterizing the extreme points of D, we can examine bases of $\begin{bmatrix} A \\ e^t \end{bmatrix}$, which are limited by the number of ways we can select $(m+1)$ columns out of n, i.e.,

$$\binom{n}{m+1} = \frac{n!}{(m+1)!(n-m-1)!},$$

which is fewer by a factor of $\frac{1}{(m+1)}$ than that of the Corollary to Theorem 2.6.6.

2.42 Problem P: Minimize $\{c^t x : Ax = b,\ x \geq 0\}$.

(Homogeneous) Problem D: Maximize $\{b^t y : A^t y \leq 0\}$.

Problem P has no feasible solution if and only if the system $Ax = b$, $x \geq 0$, is inconsistent. That is, by Farkas' Theorem (Theorem 2.4.5), this occurs if and only if the system $A^t y \leq 0$, $b^t y > 0$ has a solution, i.e., if and only if the homogeneous version of the dual problem is unbounded. □

2.45 Consider the following pair of primal and dual LPs, where e is a vector of ones in $\mathbb{R}^m$:

P:	Max	$e^t p$	**D**:	Min	$0^t x$
	subject to	$A^t p = 0$			$Ax \geq e$
		$p \geq 0.$			x unres.

Then, System 2 has a solution $\Leftrightarrow P$ is unbounded (take any feasible solution to System 2, multiply it by a scalar λ, and take $\lambda \to \infty$) $\Leftrightarrow D$

is infeasible (since P is homogeneous) $\Leftrightarrow$ $\nexists$ a solution to $Ax > 0$ $\Leftrightarrow$ $\nexists$ a solution to $Ax < 0$. $\square$

2.47 Consider the system $A^t y = c,\ y \geq 0$:

$$\begin{aligned} 2y_1 + 2y_2 &= -3 \\ y_1 + 2y_2 &= 1 \\ -3y_1 &= -2 \\ (y_1, y_2) &\geq 0. \end{aligned}$$

The first equation is in conflict with $(y_1, y_2) \geq 0$. Therefore, this system has no solution. By Farkas' Theorem we then conclude that the system $Ax \leq 0$, $c^t x > 0$ has a solution.

2.49 ($\Rightarrow$) We show that if System 2 has a solution, then System 1 is inconsistent. Suppose that System 2 is consistent and let y_0 be its solution. If System 1 has a solution, x_0, say, then we necessarily have $x_0^t A^t y_0 = 0$. However, since $x_0^t A^t = c^t$, this result leads to $c^t y_0 = 0$, thus contradicting $c^t y_0 = 1$. Therefore, System 1 must be inconsistent.

($\Leftarrow$) In this part we show that if System 2 has no solution, then System 1 has one. Assume that System 2 has no solution, and let $S = \{(z_1, z_0) : z_1 = -A^t y,\ z_0 = c^t y,\ y \in \mathbb{R}^m\}$. Then S is a nonempty convex set, and $(z_1, z_0) = (0,1) \notin S$. Therefore, there exists a nonzero vector (p_1, p_0) and a real number α such that $p_1^t z_1 + p_0 z_0 \leq \alpha < p_1^t 0 + p_0$ for any $(z_1, z_0) \in S$. By the definition of S, this implies that $-p_1^t A^t y + p_0 c^t y \leq \alpha < p_0$ for any $y \in \mathbb{R}^m$. In particular, for $y = 0$, we obtain $0 \leq \alpha < p_0$. Next, observe that since α is nonnegative and $(-p_1^t A^t + p_0 c^t) y \leq \alpha$ for any $y \in \mathbb{R}^m$, then we necessarily have $-p_1^t A^t + p_0 c^t = 0$ (or else y can be readily selected to violate this inequality). We have thus shown that there exists a vector (p_1, p_0) where $p_0 > 0$, such that $Ap_1 - p_0 c = 0$. By letting $x = \dfrac{1}{p_0} p_1$, we concluce that x solves the system $Ax - c = 0$. This shows that System 1 has a solution. $\square$

2.50 Consider the pair of primal and dual LPs below, where e is a vector of ones in $\mathbb{R}^p$:

P: Max $e^t u$
subject to $A^t u + B^t v = 0$
$u \geq 0$, v unres.

D: Min $0^t x$
subject to $Ax \geq e$
$Bx = 0$
x unres.

Hence, System 2 has a solution $\Leftrightarrow$ P is unbounded (take any solution to System 2 and multiply it with a scalar λ and take $\lambda \to \infty$) $\Leftrightarrow$ D is infeasible (since P is homogeneous) $\Leftrightarrow$ there does not exist a solution to $Ax > 0$, $Bx = 0$ $\Leftrightarrow$ System 1 has no solution. □

2.51 Consider the following two systems for each $i \in \{1,...,m\}$:

System I: $Ax \geq 0$ with $A_i x > 0$

System II: $A^t y = 0$, $y \geq 0$, with $y_i > 0$,

where A_i is the ith row of A. Accordingly, consider the following pair of primal and dual LPs:

P: Max $e_i^t y$
subject to $A^t y = 0$
$y \geq 0$

D: Min $0^t x$
subject to $Ax \geq e_i$
x unres,

where e_i is the ith unit vector. Then, we have that System II has a solution $\Leftrightarrow$ P is unbounded $\Leftrightarrow$ D is infeasible $\Leftrightarrow$ System I has no solution. Thus, exactly one of the systems has a solution for each $i \in \{1,...,m\}$. Let $I_1 = \{i \in \{1,...,m\}:$ System I has a solution; say $x^i\}$, and let $I_2 = \{i \in \{1,...,m\}:$ System II has a solution; say, $y^i\}$. Note that $I_1 \cup I_2 = \{1,...,m\}$ with $I_1 \cap I_2 = \varnothing$. Accordingly, let $\overline{x} = \sum_{i \in I_1} x^i$ and $\overline{y} = \sum_{i \in I_2} y^i$, where $\overline{x} \equiv 0$ if $I_1 = \varnothing$ and $\overline{y} \equiv 0$ if $I_2 = \varnothing$. Then it is easily verified that $\overline{x}$ and $\overline{y}$ satisfy Systems 1 and 2, respectively, with $A\overline{x} + \overline{y} = \sum_{i \in I_1} Ax^i + \sum_{i \in I_2} y^i > 0$ since $Ax^i \geq 0$, $\forall i \in I_1$, and $y^i \geq 0$, $\forall i \in I_2$, and moreover, for each row i of this system, if $\forall i \in I_1$ then we have $A_i x^i > 0$ and if $i \in I_2$ then we have $y^i > 0$.

2.52 Let $f(x) = e^{-x_1} - x_2$. Then $S_1 = \{x : f(x) \le 0\}$. Moreover, the Hessian of *f* is given by $\begin{bmatrix} e^{-x_1} & 0 \\ 0 & 0 \end{bmatrix}$, which is positive semidefinite, and so, *f* is a convex function. Thus, *S* is a convex set since it is a lower-level set of a convex function. Similarly, it is readily verified that S_2 is a convex set. Furthermore, if $\bar{x} \in S_1 \cap S_2$, then we have $-e^{-\bar{x}_1} \ge \bar{x}_2 \ge e^{-\bar{x}_1}$ or $2e^{-\bar{x}_1} \le 0$, which is achieved only in the limit as $\bar{x}_1 \to \infty$. Thus, $S_1 \cap S_2 = \varnothing$. A separating hyperplane is given by $x_2 = 0$, with $S_1 \subseteq \{x : x_2 \ge 0\}$ and $S_2 \subseteq \{x : x_2 \le 0\}$, but there does not exist any strongly separately hyperplane (since from above, both S_1 and S_2 contain points having $x_2 \to 0$).

2.53 Let $f(x) = x_1^2 + x_2^2 - 4$. Let $X = \{\bar{x} : \bar{x}_1^2 + \bar{x}_2^2 = 4\}$. Then, for any $\bar{x} \in X$, the first-order approximation to $f(x)$ is given by

$$f_{FO}(x) = f(\bar{x}) + (x - \bar{x})^t \nabla f(\bar{x}) = (x - \bar{x})^t \begin{bmatrix} 2\bar{x}_1 \\ 2\bar{x}_2 \end{bmatrix} = (2\bar{x}_1)x_1 + (2\bar{x}_2)x_2 - 8.$$

Thus *S* is described by the intersection of infinite halfspaces as follows:

$$(2\bar{x}_1)x_1 + (2\bar{x}_2)x_2 \le 8, \ \forall \bar{x} \in X,$$

which represents replacing the constraint defining *S* by its first-order approximation at all boundary points.

2.57 For the existence and uniqueness proof see, for example, *Linear Algebra and Its Applications* by Gilbert Strang (Harcourt Brace Jovanovich, Inc., 1988).

If $L = \{(x_1, x_2, x_3) : 2x_1 + x_2 - x_3 = 0\}$, then L is the nullspace of $A = [2 \ \ 1 \ -1]$, and its orthogonal complement is given by $\lambda \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$ for any $\lambda \in \mathbb{R}$. Therefore, $\mathbf{x}_1$ and $\mathbf{x}_2$ are orthogonal projections of $\mathbf{x}$ onto *L*, and $L^\perp$, respectively. If $\mathbf{x} = (1 \ \ 2 \ \ 3)$, then $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \mathbf{x}_1 + \mathbf{x}_2$ where $\mathbf{x}_2 = \lambda \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$.

Thus, $\begin{bmatrix}1\\2\\3\end{bmatrix}^t \begin{bmatrix}2\\1\\-1\end{bmatrix} = \lambda \left\| \begin{matrix}2\\1\\-1\end{matrix} \right\|^2 \Rightarrow \lambda = \frac{1}{6}$. Hence, $\mathbf{x}_1 = \frac{1}{6}(4 \ \ 11 \ \ 19)$ and $\mathbf{x}_2 = \frac{1}{6}(2 \ \ 1 \ \ -1)$.

CHAPTER 3:

CONVEX FUNCTIONS AND GENERALIZATIONS

3.1 a. $\begin{bmatrix} 4 & -4 \\ -4 & 0 \end{bmatrix}$ is indefinite. Therefore, $f(x)$ is neither convex nor concave.

b. $H(x) = e^{-(x_1+3x_2)}\begin{bmatrix} x_1 - 2 & 3(x_1 - 1) \\ 3(x_1 - 1) & 9x_1 \end{bmatrix}$. Definiteness of the matrix $H(x)$ depends on x_1. Therefore, $f(x)$ is neither convex nor concave (over R^2).

c. $H = \begin{bmatrix} -2 & 4 \\ 4 & -6 \end{bmatrix}$ is indefinite since the determinant is negative. Therefore, $f(x)$ is neither convex nor concave.

d. $H = \begin{bmatrix} 4 & 2 & -5 \\ 2 & 2 & 0 \\ -5 & 0 & 4 \end{bmatrix}$ is indefinite. Therefore, $f(x)$ is neither convex nor concave.

e. $H = \begin{bmatrix} -4 & 8 & 3 \\ 8 & -6 & 4 \\ 3 & 4 & -4 \end{bmatrix}$ is indefinite. Therefore, $f(x)$ is neither convex nor concave.

3.2 $f''(x) = abx^{b-2}e^{-ax^b}[abx^b - (b-1)]$. Hence, if $b = 1$, then f is convex over $\{x : x > 0\}$. If $b > 1$, then f is convex whenever $abx^b \geq (b-1)$, i.e., $x \geq \left[\frac{(b-1)}{ab}\right]^{1/b}$.

3.3 $f(x) = 10 - 3(x_2 - x_1^2)^2$, and its Hessian matrix is $H(x) = 6\begin{bmatrix} -6x_1^2 + 2x_2 & 2x_1 \\ 2x_1 & -1 \end{bmatrix}$. Thus, f is not convex anywhere and for f to be concave, we need $-6x_1^2 + 2x_2 \leq 0$ and $6x_1^2 - 2x_2 - 4x_1^2 \geq 0$, i.e., $3x_1^2 \geq x_2$ and $x_1^2 \geq x_2$, i.e., $x_1^2 \geq x_2$. Hence, if $S = \{(x_1, x_2) : -1 \leq x_1 \leq 1, -1 \leq x_2 \leq 1\}$, then $f(x)$ is neither convex nor concave on S.

If S is a convex set such that $S \subseteq \{(x_1, x_2) : x_1^2 \ge x_2\}$, then $H(x)$ is negative semidefinite for all $x \in S$. Therefore, $f(x)$ is concave on S.

3.4 $f(x) = x^2(x^2 - 1)$, $f'(x) = 4x^3 - 2x$, and $f''(x) = 12x^2 - 2 \ge 0$ if $x^2 \ge 1/6$. Thus f is convex over $S_1 = \{x : x \ge 1/\sqrt{6}\}$ and over $S_2 = \{x : x \le -1/\sqrt{6}\}$. Moreover, since $f''(x) > 0$ whenever $x > 1/\sqrt{6}$ or $x < -1/\sqrt{6}$, and thus f lies strictly above the tangent plane for all $x \in S_1$ as well as for all $x \in S_2$, f is strictly convex over S_1 and over S_2. For all the remaining values for x, $f(x)$ is strictly concave.

3.9 Consider any x_1, $x_2 \in R^n$, and let $x_\lambda = \lambda x_1 + (1-\lambda)x_2$ for any $0 \le \lambda \le 1$. Then

$f(x_\lambda) = \max\{f_1(x_\lambda),...,f_k(x_\lambda)\} = f_r(x_\lambda)$ for some $r \in \{1,...,k\}$, whence $f_r(x_\lambda) \le \lambda f_r(x_1) + (1-\lambda)f_r(x_2)$ by the convexity of f_r, i.e., $f(x_\lambda) \le \lambda f(x_1) + (1-\lambda)f(x_2)$ since $f(x_1) \ge f_r(x_1)$ and $f(x_2) \ge f_r(x_2)$. Thus f is convex.

If $f_1,...,f_k$ are concave functions, then $-f_1,...,-f_k$ are convex functions $\Rightarrow \max\{-f_1(x),...,-f_k(x)\}$ is convex i.e., $-\min\{f_1(x),...,f_k(x)\}$ is convex, i.e., $f(x) \equiv \min\{f_1(x),...,f_k(x)\}$ is concave.

3.10 Let x_1, $x_2 \in \mathbb{R}^n$, $\lambda \in [0,1]$, and let $x_\lambda = \lambda x_1 + (1-\lambda)x_2$. To establish the convexity of $f(\cdot)$ we need to show that $f(x_\lambda) \le \lambda f(x_1) + (1-\lambda)f(x_2)$. Notice that

$$\begin{aligned} f(x_\lambda) &= g[h(x_\lambda)] \le g[\lambda h(x_1) + (1-\lambda)h(x_2)] \\ &\le \lambda g[h(x_1)] + (1-\lambda)g[h(x_2)] \\ &= \lambda f(x_1) + (1-\lambda)f(x_2). \end{aligned}$$

In this derivation, the first inequality follows since h is convex and g is nondecreasing, and the second inequality follows from the convexity of g. This completes the proof.

3.11 Let x_1, $x_2 \in S$, $\lambda \in [0,1]$, and let $x_\lambda = \lambda x_1 + (1-\lambda)x_2$. To establish the convexity of f over S we need to show that $f(x_\lambda) - \lambda f(x_1) - (1-\lambda)f(x_2) \le 0$. For notational convenience, let

$D(x) = g(x_1)g(x_2) - \lambda g(x_\lambda)g(x_2) - (1-\lambda)g(x_\lambda)g(x_2)$. Under the assumption that $g(x) > 0$ for all $x \in S$, our task reduces to demonstrating that $D(x) \le 0$ for any $x_1,\ x_2 \in S$, and any $\lambda \in [0,1]$. By the concavity of $g(x)$ we have

$$D(x) \le g(x_1)g(x_2) - \lambda[\lambda g(x_1) + (1-\lambda)g(x_2)]g(x_2) - (1-\lambda)[\lambda g(x_1) + (1-\lambda)g(x_2)]g(x_1).$$

After a rearrangement of terms on the right-hand side of this inequality we obtain

$$\begin{aligned} D(x) &\le -\lambda(1-\lambda)[g(x_1)^2 + g(x_2)^2] + 2\lambda(1-\lambda)g(x_1)g(x_2) \\ &= -\lambda(1-\lambda)[g(x_1)^2 + g(x_2)^2] + 2\lambda(1-\lambda)g(x_1)g(x_2) \\ &= -\lambda(1-\lambda)[g(x_1)^2 + g(x_2)^2 - 2g(x_1)g(x_2)] \\ &= -\lambda(1-\lambda)[g(x_1) - g(x_2)]^2. \end{aligned}$$

Therefore, $D(x) \le 0$ for any $x_1,\ x_2 \in S$, and any $\lambda \in [0,1]$, and thus $f(x)$ is a convex function.

Symmetrically, if g is convex, $S = \{x : g(x) < 0\}$, then from above, $\frac{1}{-g}$ is convex over S, and so $f(x) = 1/g(x)$ is concave over S. □

3.16 Let $x_1,\ x_2$ be any two vectors in R^n, and let $\lambda \in [0,1]$. Then, by the definition of $h(\cdot)$, we obtain $h(\lambda x_1 + (1-\lambda)x_2) = \lambda(Ax_1 + b) + (1-\lambda)(Ax_2 + b) = \lambda h(x_1) + (1-\lambda)h(x_2)$. Therefore,

$$f(\lambda x_1 + (1-\lambda)x_2) = g[h(\lambda x_1 + (1-\lambda)x_2)] = g[\lambda h(x_1) + (1-\lambda)h(x_2)]$$
$$\le \lambda g[h(x_1)] + (1-\lambda)g[h(x_2)] = \lambda f(x_1) + (1-\lambda)f(x_2),$$

where the above inequality follows from the convexity of g. Hence, $f(x)$ is convex. □

By multivariate calculus, we obtain $\nabla f(x) = A^t \nabla g[h(x)]$, and $H_f(x) = A^t H_g[h(x)]A$.

3.18 Assume that $f(x)$ is convex. Consider any $x,\ y \in R^n$, and let $\lambda \in (0,1)$. Then

$$f(x+y) = f\left[\lambda\left(\frac{x}{\lambda}\right) + (1-\lambda)\left(\frac{y}{1-\lambda}\right)\right] \le \lambda f\left(\frac{x}{\lambda}\right) + (1-\lambda)f\left(\frac{y}{1-\lambda}\right)$$

$= f(x) + f(y)$,

and so f is subadditive.

Conversely, let f be a subadditive gauge function. Let $x,\ y \in R^n$ and $\lambda \in [0,1]$. Then

$f(\lambda x + (1-\lambda)y) \le f(\lambda x) + f[(1-\lambda)y] = \lambda f(x) + (1-\lambda)f(y)$,

and so f is convex.

3.21 See the answer to Exercise 6.4.

3.22 a. See the answer to Exercise 6.4.

b. If $y_1 \le y_2$, then $\{x : g(x) \le y_1, x \in S\} \subseteq \{x : g(x) \le y_2, x \in S\}$, and so $\phi(y_1) \ge \phi(y_2)$.

3.26 First assume that $\bar{x} = 0$. Note that then $f(\bar{x}) = 0$ and $\xi^t\bar{x} = 0$ for any vector ξ in R^n.

($\Rightarrow$) If ξ is a subgradient of $f(x) = \|x\|$ at $x = 0$, then by definition we have $\|x\| \ge \xi^t x$ for all $x \in R^n$. Thus in particular for $x = \xi$, we obtain $\|\xi\| \ge \|\xi\|^2$, which yields $\|\xi\| \le 1$.

($\Leftarrow$) Suppose that $\|\xi\| \le 1$. By the Schwarz inequality, we then obtain $\xi^t x \le \|\xi\|\ \|x\| \le \|x\|$, and so ξ is a subgradient of $f(x) = \|x\|$ at $x = 0$.

This completes the proof for the case when $\bar{x} = 0$. Now, consider $\bar{x} \ne 0$.

($\Rightarrow$) Suppose that ξ is a subgradient of $f(x) = \|x\|$ at $\bar{x}$. Then by definition, we have

$$\|x\| - \|\bar{x}\| \ge \xi^t(x - \bar{x}) \text{ for all } x \in R^n. \tag{1}$$

In particular, the above inequality holds for $x = 0$, for $x = \lambda\bar{x}$, where $\lambda > 0$, and for $x = \xi$. If $x = 0$, then $\xi^t\bar{x} \ge \|\bar{x}\|$. Furthermore, by employing the Schwarz inequality we obtain

$$\|\bar{x}\| \le \xi^t\bar{x} \le \|\xi\|\,\|\bar{x}\|. \tag{2}$$

If $x = \lambda\bar{x}$, $\lambda > 0$, then $\|x\| = \lambda\|\bar{x}\|$, and Equation (1) yields $(\lambda - 1)\|\bar{x}\| \ge (\lambda - 1)\xi^t\bar{x}$. If $\lambda > 1$, then $\|\bar{x}\| \ge \xi^t\bar{x}$, and if $\lambda < 1$, then

$\|\bar{x}\| \leq \xi^t \bar{x}$. Therefore, in either case, if ξ is a subgradient at $\bar{x}$, then it must satisfy the equation.

$$\xi^t \bar{x} = \|\bar{x}\|. \tag{3}$$

Finally, if $x = \xi$, then Equation (1) results in $\|\xi\| - \|\bar{x}\| \geq \xi^t \xi - \xi^t \bar{x}$. However, by (2), we have $\xi^t \bar{x} = \|\bar{x}\|$. Therefore, $\|\xi\|(1 - \|\xi\|) \geq 0$. This yields

$$1 - \|\xi\| \geq 0 \tag{4}$$

Combining (2) – (4), we conclude that if ξ is a subgradient of $f(x) = \|x\|$ at $\bar{x} \neq 0$, then $\xi^t \bar{x} = \|\bar{x}\|$ and $\|\xi\| = 1$.

($\Leftarrow$) Consider a vector $\xi \in R^n$ such that $\|\xi\| = 1$ and $\xi^t \bar{x} = \|\bar{x}\|$, where $\bar{x} \neq 0$. Then for any x, we have $f(x) - f(\bar{x}) - \xi^t (x - \bar{x}) = \|x\| - \|\bar{x}\| - \xi^t (x - \bar{x}) = \|x\| - \xi^t x \geq \|x\|(1 - \|\xi\|) = 0$, where we have used the Schwarz inequality $(\xi^t x \leq \|\xi\| \, \|x\|)$ to derive the last inequality. Thus ξ is a subgradient of $f(x) = \|x\|$ at $\bar{x} \neq 0$. This completes the proof. □

In order to derive the gradient of $f(x)$ at $\bar{x} \neq 0$, notice that $\|\xi\| = 1$ and $\xi^t \bar{x} = \|\bar{x}\|$ if and only if $\xi = \frac{1}{\|\bar{x}\|} \bar{x}$. Thus $\nabla f(\bar{x}) = \frac{1}{\|\bar{x}\|} \bar{x}$.

3.27 Since f_1 and f_2 are convex and differentiable, we have

$f_1(x) \geq f_1(\bar{x}) + (x - \bar{x})^t \nabla f_1(\bar{x}), \quad \forall x.$

$f_2(x) \geq f_2(\bar{x}) + (x - \bar{x})^t \nabla f_2(\bar{x}), \quad \forall x.$

Hence, $f(x) = \max\{f_1(x), f_2(x)\}$ and $f(\bar{x}) = f_1(\bar{x}) = f_2(\bar{x})$ give

$$f(x) \geq f(\bar{x}) + (x - \bar{x})^t \nabla f_1(\bar{x}), \quad \forall x \tag{1}$$

$$f(x) \geq f(\bar{x}) + (x - \bar{x})^t \nabla f_2(\bar{x}), \quad \forall x. \tag{2}$$

Multiplying (1) and (2) by λ and $(1 - \lambda)$, respectively, where $0 \leq \lambda \leq 1$, yields upon summing:

$$f(x) \geq f(\overline{x}) + (x - \overline{x})^t[\lambda \nabla f_1(\overline{x}) + (1-\lambda)\nabla f_2(\overline{x})], \quad \forall x,$$

$$\Rightarrow \quad \xi = \lambda \nabla f_1(\overline{x}) + (1-\lambda)\nabla f_2(\overline{x}), \; 0 \leq \lambda \leq 1, \text{ is a subgradient of } f \text{ at } \overline{x}.$$

($\Rightarrow$) Let ξ be a subgradient of *f* at $\overline{x}$. Then, we have,

$$f(x) \geq f(\overline{x}) + (x - \overline{x})^t \xi, \quad \forall x. \tag{3}$$

But $f(x) = \max\{f_1(x), f_2(x)\} =$

$$\max\{f_1(\overline{x}) + (x-\overline{x})^t \nabla f_1(\overline{x}) + \|x - \overline{x}\| 0_1(x \to \overline{x}),$$
$$f_2(\overline{x}) + (x-\overline{x})^t \nabla f_2(\overline{x}) + \|x - \overline{x}\| 0_2(x \to \overline{x})\}, \tag{4}$$

where $0_1(x \to \overline{x})$ and $0_2(x \to \overline{x})$ are functions that approach zero as $x \to \overline{x}$. Since $f_1(\overline{x}) = f_2(\overline{x}) = f(\overline{x})$, putting (3) and (4) together yields

$$\max\{(x-\overline{x})^t[\nabla f_1(\overline{x}) - \xi] + \|x - \overline{x}\| 0_1(x \to \overline{x}),$$
$$(x-\overline{x})^t[\nabla f_2(\overline{x}) - \xi] + \|x - \overline{x}\| 0_2(x \to \overline{x})\} \geq 0, \quad \forall x. \tag{5}$$

Now, on the contrary, suppose that $\xi \notin conv\{\nabla f_1(\overline{x}), \nabla f_2(\overline{x})\}$. Then, there exists a strictly separating hyperplane $\alpha x = \beta$ such that $\|\alpha\| = 1$ and $\alpha^t \xi > \beta$ and $\{\alpha^t \nabla f_1(\overline{x}) < \beta, \; \alpha^t \nabla f_2(\overline{x}) < \beta\}$, i.e.,

$$\alpha^t[\xi - \nabla f_1(\overline{x})] > 0 \text{ and } \alpha^t[\xi - \nabla f_2(\overline{x})] > 0. \tag{6}$$

Letting $(x - \overline{x}) = \varepsilon\alpha$ in (5), with $\varepsilon \to 0^+$, we get upon dividing with $\varepsilon > 0$:

$$\max\{\alpha^t[\nabla f_1(\overline{x}) - \xi] + 0_1(\varepsilon \to 0),$$
$$\alpha^t[\nabla f_2(\overline{x}) - \xi] + 0_2(\varepsilon \to 0)\} \geq 0, \; \forall \varepsilon > 0. \tag{7}$$

But the first terms in both maxands in (7) are negative by (6), while the second terms $\to 0$. Hence we get a contradiction. Thus $\xi \in conv\{\nabla f_1(\overline{x}), \nabla f_2(\overline{x})\}$, i.e., it is of the given form.

Similarly, if $f(x) = \max\{f_1(x),...,f_m(x)\}$, where $f_1,...,f_m$ are differentiable convex functions and $\bar{x}$ is such that $f(\bar{x}) = f_i(\bar{x})$, $\forall i \in I \subseteq \{1,...,m\}$, then ξ is a subgradient of f at $\bar{x} \Leftrightarrow \xi \in conv\{\nabla f_i(\bar{x}),\ i \in I\}$. A likewise result holds for the minimum of differentiable concave functions.

3.28 a. See Theorem 6.3.1 and its proof. (Alternatively, since θ is the minimum of several affine functions, one for each extreme point of X, we have that θ is a piecewise linear and concave.)

b. See Theorem 6.3.7. In particular, for a given vector $\bar{u}$, let $X(\bar{u}) = \{x_1,...,x_k\}$ denote the set of all extreme points of the set X that are optimal solutions for the problem to minimize $\{c^t x + \bar{u}^t(Ax - b) :\ x \in X\}$. Then $\xi(\bar{u})$ is a subgradient of $\theta(u)$ at $\bar{u}$ if and only if $\xi(\bar{u})$ is in the convex hull of $Ax_1 - b,...,Ax_k - b$, where $x_i \in X(\bar{u})$ for $i = 1,...,k$. That is, $\xi(\bar{u})$ is a subgradient of $\theta(u)$ at $\bar{u}$ if and only if $\xi(\bar{u}) = A\sum_{i=1}^{k} \lambda_i x_i - b$ for some nonnegative $\lambda_1,...,\lambda_k$, such that $\sum_{i=1}^{k} \lambda_i = 1$.

3.31 Let $\mathbf{P}_1 : \min\{f(x) : x \in S\}$ and $\mathbf{P}_2 : \min\{f_s(x) : x \in S\}$, and let $S_1 = \{x^* \in S : f(x^*) \le f(x), \forall x \in S\}$ and $S_2 = \{x^* \in S : f_s(x^*) \le f_s(x), \forall x \in S\}$. Consider any $x^* \in S_1$. Hence, x^* solves Problem P_1. Define $h(x) = f(x^*), \forall x \in S$. Thus, the constant function h is a convex underestimating function for f over S, and so by the definition of f_s, we have that

$$f_s(x) \ge h(x) = f(x^*), \forall x \in S. \qquad (1)$$

But $f_s(x^*) \le f(x^*)$ since $f_s(x) \le f(x), \forall x \in S$. This, together with (1), thus yields $f_s(x^*) = f(x^*)$ and that x^* solves Problem P_2 (since (1) asserts that $f(x^*)$ is a lower bound on Problem P_2). Therefore, $x^* \in S_2$. Thus, we have shown that the optimal values of Problems P_1 and P_2 match, and that $S_1 \subseteq S_2$. □

3.37 $\nabla f(x) = \begin{bmatrix} 4x_1 e^{2x_1^2 - x_2^2} & -3 \\ -2x_2 e^{2x_1^2 - x_2^2} & +5 \end{bmatrix}$, $\nabla f\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4e - 3 \\ -2e + 5 \end{bmatrix}$

$$H(x) = 2e^{2x_1^2 - x_2^2}\begin{bmatrix} 8x_1^2 + 2 & -4x_1x_2 \\ -4x_1x_2 & 2x_2^2 - 1 \end{bmatrix}, \quad H\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 2e\begin{bmatrix} 10 & -4 \\ -4 & 1 \end{bmatrix},$$

with $f\begin{bmatrix} 1 \\ 1 \end{bmatrix} = e + 2.$

Thus, the linear (first-order) approximation of f at $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is given by

$$f_1(x) \equiv (e+2) + (x_1 - 1)(4e - 3) + (x_2 - 1)(-2e + 5),$$

and the second-order approximation of f at $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is given by

$$f_2(x) \equiv (e+2) + (x_1 - 1)(4e - 3) + (x_2 - 1)(-2e + 5) + e\left[10(x_1 - 1)^2 - 8(x_1 - 1)(x_2 - 1) + (x_2 - 1)^2\right].$$

f_1 is both convex and concave (since it is affine). The Hessian of f_2 is given by $H\begin{bmatrix} 1 \\ 1 \end{bmatrix}$, which is indefinite, and so f_2 is neither convex nor concave.

3.39 The function $f(x) = x^t Ax$ can be represented in a more convenient form as $f(x) = \frac{1}{2}x^t(A + A^t)x$, where $(A + A^t)$ is symmetric. Hence, the Hessian matrix of $f(x)$ is $H = A + A^t$. By the superdiagonalization procedure, we can readily verify that $H = \begin{bmatrix} 4 & 3 & 4 \\ 3 & 6 & 3 \\ 4 & 3 & 2\theta \end{bmatrix}$. H is positive semidefinite if and only if $\theta \geq 2$, and is positive definite for $\theta > 2$. Therefore, if $\theta > 2$, then $f(x)$ is strictly convex. To examine the case when $\theta = 2$, consider the following three points: $x_1 = (1, 0, 0)$, $x_2 = (0, 0, 1)$, and $\bar{x} = \frac{1}{2}x_1 + \frac{1}{2}x_2$. As a result of direct substitution, we obtain $f(x_1) = f(x_2) = 2$, and $f(\bar{x}) = 2$. This shows that $f(x)$ is not strictly convex (although it is still convex) when $\theta = 2$.

3.40 $f(x) = x^3 \Rightarrow f'(x) = 3x^2$ and $f''(x) = 6x \geq 0,\ \forall x \in S$. Hence f is convex on S. Moreover, $f''(x) > 0,\ \forall x \in \text{int}(S)$, and so f is strictly convex on int(S). To show that f is strictly convex on S, note that $f''(x) = 0$ only for $x = 0 \in S$, and so following the argument given after Theorem 3.3.8, any supporting hyperplane to the epigraph of f over S at any point $\bar{x}$ must touch it only at $[\bar{x}, f(\bar{x})]$, or else this would contradict the strict convexity of f over int(S). Note that the first nonzero derivative of order greater than or equal to 2 at $\bar{x} = 0$ is $f'''(\bar{x}) = 6$, but Theorem 3.3.9 does not apply here since $\bar{x} = 0 \in \partial(S)$. Indeed, this shows that $f(x) = x^3$ is neither convex nor concave over R. But Theorem 3.3.9 applies (and holds) over int(S) in this case.

3.41 The matrix H is symmetric, and therefore, it is diagonalizable. That is, there exists an orthogonal $n \times n$ matrix Q, and a diagonal $n \times n$ matrix D such that $H = QDQ^t$. The columns of the matrix Q are simply normalized eigenvectors of the matrix H, and the diagonal elements of the matrix D are the eigenvalues of H. By the positive semidefiniteness of H, we have $diag\{D\} \geq 0$, and hence there exists a square root matrix $D^{1/2}$ of D (that is $D = D^{1/2}D^{1/2}$).

If $x = 0$, then readily $Hx = 0$. Suppose that $x^t Hx = 0$ for some $x \neq 0$. Below we show that then Hx is necessarily 0. For notational convenience let $z = D^{1/2}Q^t x$. Then the following equations are equivalent to $x^t Hx = 0$:

$$
\begin{aligned}
& x^t Q D^{1/2} D^{1/2} Q^t x = 0 \\
\Leftrightarrow \quad & z^t z = 0, \text{ i.e., } \|z\|^2 = 0 \\
\Leftrightarrow \quad & z = 0.
\end{aligned}
$$

By premultiplying the last equation by $QD^{1/2}$, we obtain $QD^{1/2}z = 0$, which by the definition of z gives $QDQ^t x = 0$. Thus $Hx = 0$, which completes the proof. $\square$

3.45 Consider the problem

$$
\begin{aligned}
\textbf{P:} \quad & \text{Minimize} \quad (x_1 - 4)^2 + (x_2 - 6)^2 \\
& \text{subject to} \quad x_2 \geq x_1^2 \\
& \qquad\qquad\quad\;\; x_2 \leq 4.
\end{aligned}
$$

Note that the feasible region (denote this by X) of Problem P is convex. Hence, a necessary condition for $\bar{x} \in X$ to be an optimal solution for Problem P is that

$$\nabla f(\bar{x})^t (x - \bar{x}) \geq 0, \ \forall x \in X, \tag{1}$$

because if there exists an $\hat{x} \in X$ such that $\nabla f(\bar{x})^t (\hat{x} - \bar{x}) < 0$, then $d \equiv (\hat{x} - \bar{x})$ would be an improving (since f is differentiable) and feasible (since X is convex) direction.

For $\bar{x} = (2,4)^t$, we have $\nabla f(\bar{x}) = \begin{bmatrix} 2(2-4) \\ 2(4-6) \end{bmatrix} = \begin{bmatrix} -4 \\ -4 \end{bmatrix}$.

Hence,

$$\nabla f(\bar{x})^t (x - \bar{x}) = [-4, -4] = \begin{bmatrix} x_1 - 2 \\ x_2 - 4 \end{bmatrix} = -4x_1 - 4x_2 + 24. \tag{2}$$

But $x_1^2 \leq x_2 \leq 4$, $\forall x \in X \Rightarrow x_2 \leq 4$ and $-2 \leq x_1 \leq 2$, and so $-4x_1 \geq -8$ and $-4x_2 \geq -16$. Hence, $\nabla f(\bar{x})^t (x - \bar{x}) \geq 0$ from (2).

Furthermore, observe that the objective function of Problem P (denoted by $f(x)$) is (strictly) convex since its Hessian is given by $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$, which is positive definite. Hence, by Corollary 2 to Theorem 3.4.3, we have that (1) is also sufficient for optimality to P, and so $\bar{x} = (2,4)^t$ (uniquely) solves Problem P.

3.48 Suppose that λ_1 and λ_2 are in the interval $(0, \delta)$, and such that $\lambda_2 > \lambda_1$. We need to show that $f(x + \lambda_2 d) \geq f(x + \lambda_1 d)$.

Let $\alpha = \lambda_1 / \lambda_2$. Note that $\alpha \in (0,1)$, and $x + \lambda_1 d = \alpha(x + \lambda_2 d) + (1-\alpha)x$. Therefore, by the convexity of f, we obtain $f(x + \lambda_1 d) \leq \alpha f(x + \lambda_2 d) + (1-\alpha) f(x)$, which leads to $f(x + \lambda_1 d) \leq f(x + \lambda_2 d)$ since, by assumption, $f(x) \leq f(x + \lambda d)$ for any $\lambda \in (0, \delta)$.

When f is strictly convex, we can simply replace the weak inequalities above with strict inequalities to conclude that $f(x+\lambda d)$ is strictly increasing over the interval $(0, \delta)$.

3.51 ($\Leftrightarrow$) If the vector d is a descent direction of f at $\bar{x}$, then $f(\bar{x}+\lambda d) - f(\bar{x}) < 0$ for all $\lambda \in (0, \delta)$. Moreover, since f is a convex and differentiable function, we have that $f(\bar{x}+\lambda d) - f(\bar{x}) \geq \lambda \nabla f(\bar{x})^t d$. Therefore, $\nabla f(\bar{x})^t d < 0$.

($\Leftrightarrow$) See the proof of Theorem 4.1.2. □

Note: If the function $f(x)$ is not convex, then it is not true that $\nabla f(\bar{x})^t d < 0$ whenever d is a descent direction of $f(x)$ at $\bar{x}$. For example, if $f(x) = x^3$, then $d = -1$ is a descent direction of f at $\bar{x} = 0$, but $f'(\bar{x})d = 0$.

3.54 ($\Rightarrow$) If $\bar{x}$ is an optimal solution, then we must have $f'(\bar{x};d) \geq 0$, $\forall d \in D$, since $f'(\bar{x};d) < 0$ for any $d \in D$ implies the existence of improving feasible solutions by Exercise 3.5.1.

($\Leftarrow$) Suppose $f'(\bar{x};d) \geq 0$, $\forall d \in D$, but on the contrary, $\bar{x}$ is not an optimal solution, i.e., there exists $\hat{x} \in S$ with $f(\hat{x}) < f(\bar{x})$. Consider $d = (\hat{x} - \bar{x})$. Then $d \in D$ since S is convex. Moreover, $f(\bar{x}+\lambda d) = f(\lambda\hat{x} + (1-\lambda)\bar{x}) \leq \lambda f(\hat{x}) + (1-\lambda) f(\bar{x}) < f(\bar{x})$, $\forall 0 < \lambda \leq 1$. Thus d is a feasible, descent direction, and so $f'(\bar{x};d) < 0$ by Exercise 3.51, a contradiction.

Theorem 3.4.3 similarly deals with nondifferentiable convex functions.

If $S = R^n$, then $\bar{x}$ is optimal $\Leftrightarrow$ $\nabla f(\bar{x})^t d \geq 0$, $\forall d \in R^n$
$\Leftrightarrow$ $\nabla f(\bar{x}) = 0$ (else, pick $d = -\nabla f(\bar{x})$ to get a contradiction).

3.56 Let x_1, $x_2 \in R^n$. Without loss of generality assume that $h(x_1) \geq h(x_2)$. Since the function g is nondecreasing, the foregoing assumption implies that $g[h(x_1)] \geq g[h(x_2)]$, or equivalently, that $f(x_1) \geq f(x_2)$. By the quasiconvexity of h, we have $h(\alpha x_1 + (1-\alpha)x_2) \leq h(x_1)$ for any $\alpha \in [0,1]$. Since the function g is nondecreasing, we therefore have, $f(\alpha x_1 + (1-\alpha)x_2) = g[h(\alpha x_1 + (1-\alpha)x_2)] \leq g[h(x_1)] = f(x_1)$. This shows that $f(x)$ is quasiconvex. □

3.61 Let α be an arbitrary real number, and let $S = \{x : f(x) \le \alpha\}$. Furthermore, let x_1 and x_2 be any two elements of *S*. By Theorem 3.5.2, we need to show that *S* is a convex set, that is, $f(\lambda x_1 + (1-\lambda)x_2) \le \alpha$ for any $\lambda \in [0,1]$. By the definition of $f(x)$, we have

$$f(\lambda x_1 + (1-\lambda)x_2) = \frac{g(\lambda x_1 + (1-\lambda)x_2)}{h(\lambda x_1 + (1-\lambda)x_2)} \le \frac{\lambda g(x_1) + (1-\lambda)g(x_2)}{\lambda h(x_1) + (1-\lambda)h(x_2)}, \tag{1}$$

where the inequality follows from the assumed properties of the functions *g* and *h*. Furthermore, since $f(x_1) \le \alpha$ and $f(x_2) \le \alpha$, we obtain

$$\lambda g(x_1) \le \lambda \alpha h(x_1) \text{ and } (1-\lambda)g(x_2) \le (1-\lambda)\alpha h(x_2).$$

By adding these two inequalities, we obtain $\lambda g(x_1) + (1-\lambda)g(x_2) \le \alpha[\lambda h(x_1) + (1-\lambda)h(x_2)]$. Since *h* is assumed to be a positive-valued function, the last inequality yields

$$\frac{\lambda g(x_1) + (1-\lambda)g(x_2)}{\lambda h(x_1) + (1-\lambda)h(x_2)} \le \alpha,$$

or by (1), $f(\lambda x_1 + (1-\lambda)x_2) \le \alpha$. Thus, *S* is a convex set, and therefore, $f(x)$ is a quasiconvex function. □

Alternative proof: For any $\alpha \in R$, let $S_\alpha = \{x \in S : g(x)/h(x) \le \alpha\}$. We need to show that S_α is a convex set. If $\alpha < 0$, then $S_\alpha = \emptyset$ since $g(x) \ge 0$ and $h(x) \ge 0$, $\forall x \in S$, and so S_α is convex. If $\alpha \ge 0$, then $S_\alpha = \{x \in S : g(x) - \alpha h(x) \le 0\}$ is convex since $g(x) - \alpha h(x)$ is a convex function, and S_α is a lower level set of this function. □

3.62 We need to prove that if $g(x)$ is a convex nonpositive-valued function on *S* and $h(x)$ is a convex and positive-valued function on *S*, then $f(x) = g(x)/h(x)$ is a quasiconvex function on *S*. For this purpose we show that for any $x_1, x_2 \in S$, if $f(x_1) \ge f(x_2)$, then $f(x_\lambda) \le f(x_1)$, where $x_\lambda = \lambda x_1 + (1-\lambda)x_2$, and $\lambda \in [0,1]$. Note that by the definition of *f* and the assumption that $h(x) > 0$ for all $x \in S$, it suffices to show that $g(x_\lambda)h(x_1) - g(x_1)h(x_\lambda) \le 0$. Towards this end, observe that

$g(x_\lambda)h(x_1) \le [\lambda g(x_1) + (1-\lambda)g(x_2)]h(x_1)$ since $g(x)$ is convex and $h(x) > 0$ on S;

$g(x_1)h(x_\lambda) \ge g(x_1)[\lambda h(x_1) + (1-\lambda)h(x_2)]$ since $h(x)$ is convex and $g(x) \le 0$ on S;

$g(x_2)h(x_1) - g(x_1)h(x_2) \le 0$, since $f(x_1) \ge f(x_2)$ and $h(x) > 0$ on S.

From the foregoing inequalities we obtain

$$
\begin{aligned}
& g(x_\lambda)h(x_1) - g(x_1)h(x_\lambda) \\
& \le [\lambda g(x_1) + (1-\lambda)g(x_2)]h(x_1) - g(x_1)[\lambda h(x_1) + (1-\lambda)h(x_2)] \\
& = (1-\lambda)[g(x_2)h(x_1) - g(x_1)h(x_2)] \le 0,
\end{aligned}
$$

which implies that $f(x_\lambda) \le \max\{f(x_1), f(x_2)\} = f(x_1)$. □

Note: See also the alternative proof technique for Exercise 3.61 for a similar simpler proof of this result.

3.63 By assumption, $h(x) \ne 0$, and so the function $f(x)$ can be rewritten as $f(x) = g(x)/p(x)$, where $p(x) \equiv 1/h(x)$. Furthermore, since $h(x)$ is a concave and positive-valued function, we conclude that $p(x)$ is convex and positive-valued on S (see Exercise 3.11). Therefore, the result given in Exercise 3.62 applies. This completes the proof. □

3.64 Let us show that if $g(x)$ and $h(x)$ are differentiable, then the function defined in Exercise 3.61 is pseudoconvex. (The cases of Exercises 3.62 and 3.63 are similar.) To prove this, we show that for any $x_1, x_2 \in S$, if $\nabla f(x_1)^t(x_2 - x_1) \ge 0$, then $f(x_2) \ge f(x_1)$. From the assumption that $h(x) > 0$, it follows that $\nabla f(x_1)^t(x_2 - x_1) \ge 0$ if and only if $[h(x_1)\nabla g(x_1) - g(x_1)\nabla h(x_1)]^t(x_2 - x_1) \ge 0$. Furthermore, note that $\nabla g(x_1)^t(x_2 - x_1) \le g(x_2) - g(x_1)$, since $g(x)$ is a convex and differentiable function on S, and $\nabla h(x_1)^t(x_2 - x_1) \ge h(x_2) - h(x_1)$, since $h(x)$ is a concave and differentiable function on S. By multiplying the latter inequality by $-g(x_1) \le 0$, and the former one by $h(x_1) > 0$, and adding the resulting inequalities, we obtain (after rearrangement of terms):

$$[h(x_1)\nabla g(x_1) - g(x_1)\nabla h(x_1)]^t(x_2 - x_1) \le h(x_1)g(x_2) - g(x_1)h(x_2).$$

The left-hand side expression is nonegative by our assumption, and therefore, $h(x_1)g(x_2) - g(x_1)h(x_2) \geq 0,$ which implies that $f(x_2) \geq f(x_1).$ This completes the proof. □

3.65 For notational convenience let $g(x) = c_1^t x + \alpha_1,$ and let $h(x) = c_2^t x + \alpha_2.$ In order to prove pseudoconvexity of $f(x) = \dfrac{g(x)}{h(x)}$ on the set $S = \{x : h(x) > 0\}$ we need to show that for any $x_1,\ x_2 \in S,$ if $\nabla f(x_1)^t (x_2 - x_1) \geq 0,$ then $f(x_2) \geq f(x_1).$

Assume that $\nabla f(x_1)^t (x_2 - x_1) \geq 0$ for some $x_1,\ x_2 \in S.$ By the definition of *f*, we have $\nabla f(x) = \dfrac{1}{[h(x)]^2}[h(x)c_1 - g(x)c_2].$ Therefore, our assumption yields $[h(x_1)c_1 - g(x_1)c_2]^t (x_2 - x_1) \geq 0.$ Furthermore, by adding and subtracting $\alpha_1 h(x_1) + \alpha_2 g(x_1)$ we obtain $g(x_2)h(x_1) - h(x_2)g(x_1) \geq 0.$ Finally, by dividing this inequality by $h(x_1)h(x_2)$ (> 0), we obtain $f(x_2) \geq f(x_1),$ which completes the proof of pseudoconvexity of $f(x).$ The psueoconcavity of $f(x)$ on *S* can be shown in a similar way. Thus, *f* is pseudolinear. □