CHAPTER 4:

THE FRITZ JOHN AND KARUSH-KUHN-TUCKER OPTIMALITY CONDITIONS

4.1 $f(x) = xe^{-2x}$. Then $f'(x) = -2xe^{-2x} + e^{-2x} = 0$ implies that $e^{-2x}(1-2x) = 0 \Rightarrow x = \bar{x} = 1/2$. Also, $f''(x) = 4e^{-2x}(x-1)$. Hence, at $\bar{x} = 1/2$, we have that $f''(\bar{x}) < 0$, and so $\bar{x} = 1/2$ is a strict local max for f. This is also a global max and there does not exist a local/global min since from f'', the function is concave for $x \leq 1$ with $f(x) \to -\infty$ as $x \to -\infty$, and f is convex and monotone decreasing for $x \geq 1$ with $f(x) \to 0$ as $x \to \infty$).

4.4 Let $f(x) = 2x_1^2 - x_1x_2 + x_2^2 - 3x_1 + e^{2x_1+x_2}$.

a. The first-order necessary condition is $\nabla f(x) = 0$, that is:

$$\begin{bmatrix} 4x_1 - x_2 - 3 + 2e^{2x_1+x_2} \\ -x_1 + 2x_2 + e^{2x_1+x_2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

The Hessian $H(x)$ of $f(x)$ is

$$H(x) = \begin{bmatrix} 4 + 4e^{2x_1+x_2} & 2e^{2x_1+x_2} - 1 \\ 2e^{2x_1+x_2} - 1 & 2 + e^{2x_1+x_2} \end{bmatrix},$$ and as can be easily verified,

$H(x)$ is a positive definite matrix for all x. Therefore, the first-order necessary condition is sufficient in this case.

b. $\bar{x} = (0,0)$ is not an optimal solution. $\nabla f(\bar{x}) = [-1 \;\; 1]^t$, and any direction $d = (d_1, d_2)$ such that $-d_1 + d_2 < 0$ (e.g., $d = (1,0)$) is a descent direction of $f(x)$ at $\bar{x}$.

c. Consider $d = (1,0)$. Then $f(\bar{x} + \lambda d) = 2\lambda^2 - 3\lambda + e^{2\lambda}$. The minimum value of $f(\bar{x} + \lambda d)$ over the interval $[0, \infty)$ is 0.94 and is attained at $\lambda^* = 0.1175$.

d. If the last term is dropped, $f(x) = 2x_1^2 - x_1x_2 + x_2^2 - 3x_1$. Then the first-order necessary condition yields a unique solution $\bar{x}_1 = 6/7$ and

$\bar{x}_2 = 3/7$. Again, the Hessian of $f(x)$ is positive definite for all x, and so the foregoing values of x_1 and x_2 are optimal. The minimum value of $f(x)$ is given by –63/49.

4.5 The KKT system is given by:

$$\begin{array}{llllllll}
4x_1^3 & +24x_1 & -x_2 & -u_1 & -2u_2 & -u_3 & & = 1 \\
4x_2^3 & +12x_2 & -x_1 & -u_1 & +u_2 & & -u_4 & = 1 \\
x_1 & +x_2 & & & & & & \geq 6 \\
2x_1 & -x_2 & & & & & & \geq 3
\end{array}$$

$$u_1(6 - x_1 - x_2) = 0, \quad u_2(3 - 2x_1 + x_2) = 0$$

$$u_3 x_1 = 0, \quad u_4 x_2 = 0,$$

$$x_1 \geq 0,\ x_2 \geq 0,\ u_i \geq 0 \text{ for } i = 1, 2, 3, 4.$$

If $\bar{x} = (3, 3)$, then denoting the Lagrange multipliers by $\bar{u}$, we have that $\bar{u}_3 + \bar{u}_4 = 0$. Consequently, the first two equations give $\bar{u}_1 = 152$ and $\bar{u}_2 = 12$. Thus, all the KKT conditions are satisfied at $\bar{x} = (3, 3)$. The Hessian of the objective function is positive definite, and so the problem involves minimizing a strictly convex function over a convex set. Thus, $\bar{x} = (3, 3)$ is the unique global optimum.

4.6 a. In general, the problem seeks a vector y in the column space of A (i.e., $y = Ax$) that is the closest to the given vector b. If b is in the column space of A, then we need to find a solution of the system $Ax = b$. If in addition to this, the rank of A is n, then x is unique. If b is not in the column space of A, then a vector in the column space of A that is the closest to b is the projection of the vector b onto the column space of A. In this case, the problem seeks a solution to the system $Ax = y$, where y is the projection vector of b onto the column space of A. In answers to Parts (b), (c), and (d) below it is assumed that b is not in the column space of A, since otherwise the problem trivially reduces to "find a solution to the system $Ax = b$."

b. Assume that $\|\cdot\|_2$ is used, and let $f(x)$ denote the objective function for this optimization problem. Then, $f(x) = b^t b - 2x^t A^t b + x^t A^t Ax$, and the first-order necessary condition is $A^t Ax = A^t b$. The Hessian matrix of $f(x)$ is $A^t A$, which is positive semidefinite. Therefore,

$f(x)$ is a convex function. By Theorem 4.3.8 it then follows that the necessary condition is also sufficient for optimality.

c. The number of optimal solutions is exactly the same as the number of solutions to the system $A^t Ax = A^t b$.

d. If the rank of A is n, then $A^t A$ is positive definite and thus invertible. In this case, $x = (A^t A)^{-1} A^t b$ is the unique solution. If the rank of A is less than n, then the system $A^t Ax = A^t b$ has infinitely many solutions. In this case, additional criteria can be used to select an appropriate optimal solution as needed. (For details see *Linear Algebra and Its Applications* by Gilbert Strang, Harcourt Brace Jovanovich, Publishers, San Diego, 1988, Third Edition.)

e. The rank of A is 3, therefore, a unique solution exists. $(A^t A) = \begin{bmatrix} 5 & -2 & 1 \\ -2 & 6 & 4 \\ 1 & 4 & 5 \end{bmatrix}$, and $A^t b = [4\ \ 12\ \ 12]^t$. The unique solution is $x^* = [2\ \ \frac{20}{7}\ \ \frac{-2}{7}]^t$.

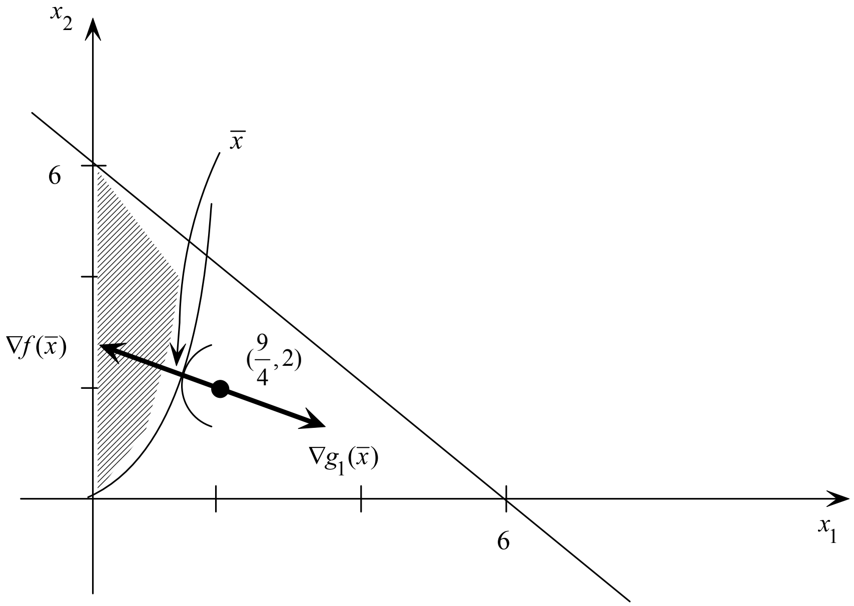
4.7 a. The KKT system for the given problem is:

$$2x_1 + 2u_1x_1 + u_2 - u_3 = 9/2$$
$$2x_2 - u_1 + u_2 - u_4 = 4$$
$$x_1^2 - x_2 \le 0$$
$$x_1 + x_2 \le 6$$
$$u_1(x_1^2 - x_2) = 0, \quad u_2(6 - x_1 - x_2) = 0, \quad x_1u_3 = 0, \quad x_2u_4 = 0$$
$$x_1 \ge 0, \quad x_2 \ge 0, \quad u_i \ge 0 \text{ for } i = 1, 2, 3, 4.$$

At $\bar{x} = (3/2,\ 9/4)^t$, denoting the Lagrange multipliers by $\bar{u}$, we necessarily have $\bar{u}_2 = \bar{u}_3 = \bar{u}_4 = 0$, which yields a unique value for u_1 namely, $\bar{u}_1 = 1/2$. The above values for x_1, x_2, and u_i for $i = 1, 2, 3, 4$ satisfy the KKT system, and therefore $\bar{x}$ is a KKT point.

b. Graphical illustration:

From the graph, it follows that at $\bar{x}$, the gradient of $f(x)$ is a negative multiple of the gradient of $g_1(x) = x_1^2 - x_2$, where $g_1(x) \leq 0$ is the only active constraint at $\bar{x}$.



c. It can be easily verified that the objective function is strictly convex, and that the active constraint function is also convex (in fact, the entire feasible region is convex in this case). Hence, $\bar{x}$ is the unique (global) optimal solution to this problem.

4.8 a. The objective function $f(x_1, x_2) = \dfrac{x_1 + 3x_2 + 3}{2x_1 + x_2 + 6}$ is pesudoconvex over the feasible region (see the proof of Lemma 11.4.1). The constraint functions are linear, and are therefore quasiconvex and quasiconcave. Therefore, by Theorem 4.3.8, if $\bar{x}$ is a KKT point for this problem, then $\bar{x}$ is a global optimal solution.

b. First note that $f(0,0) = f(6,0) = 1/2$, and moreover, $f[\lambda(0,0) + (1-\lambda)(6,0)] = 1/2$ for any $\lambda \in [0,1]$. Since (0, 0) and (6, 0) are feasible solutions, and the feasible region is a polyhedral set, any convex combination of (0, 0) and (6, 0) is also a feasible solution. It is thus sufficient to verify that one of these two points is a KKT point. Consider (6, 0). The KKT system for this problem is as follows:

$$
\begin{aligned}
&\frac{-5x_2}{(2x_1+x_2+6)^2} + 2u_1 - u_2 - u_3 &&= 0\\
&\frac{5x_1+15}{(2x_1+x_2+6)^2} + u_1 + 2u_2 \qquad -u_4 &&= 0\\
&2x_1 + x_2 \le 12\\
&-x_1 + 2x_2 \le 4\\
&x_1 \ge 0,\ x_2 \ge 0,\ u_i \ge 0 \text{ for } i = 1, 2, 3, 4\\
&u_1(2x_1 + x_2 - 12) = 0,\ u_2(-x_1 + 2x_2 - 4) = 0,\ u_3x_1 = 0,\ u_4x_2 = 0.
\end{aligned}
$$

Substituting (6, 0) for (x_1, x_2) into this KKT system yields the following unqiue values for the Lagrangian multipliers: $u_1 = u_2 = u_3 = 0$, and $u_4 = 5/36$. Since $u_i \ge 0,\ \forall i = 1,2,3,4$, we conclude that (6, 0) is indeed a KKT point, and therefore, by Part (a), it solves the given problem. Hence, by the above argument, any point on the line segment joining (0, 0) and (6, 0) is an optimal solution.

4.9 Note that $c \neq 0$, as given.

a. Let $f(d) = -c^t d$, and $g(d) = d^t d - 1$. The KKT system for the given problem is as follows:

$$
\begin{aligned}
-c + 2du &= 0\\
d^t d &\le 1\\
u(d^t d - 1) &= 0\\
u &\ge 0.
\end{aligned}
$$

$d = \bar{d} \equiv c/\|c\|$ and $u = \bar{u} \equiv \|c\|/2$ yields a solution to this system. Hence, $\bar{d}$ is a KKT point. Moreover, $\bar{d}$ is an optimal solution, because it is a KKT point and sufficiency conditions for optimality are met since $f(d)$ is a linear function, hence it is pseudoconvex, and $g(d)$ is a convex function, hence it is quasiconvex. Furthermore, $\bar{d}$ is the unique global optimal solution since the KKT system provides necessary and sufficient conditions for optimality in this case, and $\bar{d} = c/\|c\|$, $\bar{u} = \|c\|/2$ is its unique solution. To support this statement, notice that if $u > 0$, then $d^t d = 1$, which together with the first equation results in $d = c/\|c\|$ and $u = \|c\|/2$. If $u = 0$, then the first equation is inconsistent regardless of d since $c \neq 0$.

b. The steepest ascent direction of a differentiable function $f(x)$ at $\bar{x}$ can be found as an optimal solution $\bar{d}$ to the following problem:

$$\text{Maximize } \{\nabla f(\bar{x})^t d : d^t d \leq 1\},$$

which is identical to the problem considered in Part (a) with $c = \nabla f(\bar{x})$. Thus, if $\nabla f(\bar{x}) \neq 0$, then the steepest ascent direction is given by $\bar{d} = \nabla f(\bar{x}) / \|\nabla f(\bar{x})\|$.

4.10 a. In order to determine whether a feasible solution $\bar{x}$ is a KKT point, one needs to examine if there exists a feasible solution to the system:

$$\nabla f(\bar{x}) + \sum_{i \in I} \nabla g_i(\bar{x}) u_i = 0, \; u_i \geq 0 \text{ for } i \in I,$$

where I is the set of indices of constraints that are active at $\bar{x}$.

Let $c = -\nabla f(\bar{x})$ and let $A^t = [\nabla g_i(\bar{x}), i \in I]$. Then the KKT system can be rewritten as follows:

$$A^t u = c, \; u \geq 0. \tag{1}$$

Therefore, $\bar{x}$ is a KKT point if and only if System (1) has a solution. Note that System (1) is linear, and it has a solution if and only if the optimal objective value in the following problem is zero:

$$\begin{aligned} \text{Minimize} \quad & e^t y \\ \text{subject to} \quad & A^t u \pm y = c \\ & u \geq 0, \; y \geq 0, \end{aligned}$$

where e is a column vector of ones, y is a vector of artificial variables, and where $\pm y$ denotes that the components of y are ascribed the same sign as that of the respective components of c. This problem is a Phase I LP for finding a nonnegative solution to $A^t u = c$.

b. In the presence of equality constraints $h_i(x) = 0$, $i = 1, \ldots, \ell$, the KKT system is given by

$$\nabla f(\bar{x}) + \sum_{i \in I} \nabla g_i(\bar{x}) u_i + \sum_{i=1}^{\ell} \nabla h_i(\bar{x}) v_i = 0, \; u_i \geq 0 \text{ for } i \in I,$$

where I is the set of indices of the inequality constraints that are active at $\bar{x}$. Let A^t be as defined in Part (a), and let $B^t = [\nabla h_i(\bar{x}),\ i = 1,\dots,\ell]$ be the $n \times \ell$ Jacobian matrix at $\bar{x}$ for the equality constraints. Then the corresponding Phase I problem is given as follows:

$$\begin{array}{ll} \text{Minimize} & e^t y \\ \text{subject to} & A^t u + B^t v \pm y = c \\ & u \geq 0,\ \ y \geq 0. \end{array}$$

c. In this example, we have $\bar{x} = (1,2,5)^t$, $c = -\nabla f(\bar{x}) = -(8,3,23)^t$. Furthermore, $I = \{1,3\}$, and therefore,

$A^t = [\nabla g_1(\bar{x})\ \ \nabla g_3(\bar{x})] = \begin{bmatrix} 2 & -1 \\ 4 & -1 \\ -1 & 0 \end{bmatrix}$. Thus, $\bar{x} = (1,2,5)^t$ is a KKT point if and only if the optimal objective value of the following problem is zero:

$$\begin{array}{lllllll} \text{Minimize} & y_1 & + & y_2 & +y_3 & & \\ \text{subject to} & -2u_1 & + & u_3 & +y_1 & & = 8 \\ & -4u_1 & + & u_3 & & +y_2 & = 3 \\ & u_1 & & & & +y_3 & = 23 \\ & \multicolumn{6}{l}{u_1 \geq 0,\ u_3 \geq 0,\ y_i \geq 0 \text{ for } i = 1, 2, 3.} \end{array}$$

However, the optimal solution to this problem is given by $\bar{u}_1 = 2.5$, $\bar{u}_3 = 13$, $\bar{y}_1 = \bar{y}_2 = 0$, $\bar{y}_3 = 20.5$, and the optimal objective value is positive (20.5), and so we conclude that $\bar{x} = (1,2,5)^t$ is not a KKT point.

4.12 Let $y_j = \dfrac{a_j}{b} x_j$ and $d_j = \dfrac{c_j a_j}{b}$ for $j = 1,\dots, n$. Then the given optimization problem is equivalent to the following, re-written in a more convenient form:

$$\text{Minimize} \quad \sum_{j=1}^{n} \left(\frac{d_j}{y_j} \right)$$

subject to $\sum_{j=1}^{n} y_j = 1$

$y_j \geq 0$ for $j = 1,\ldots, n$.

The KKT system for the above problem is given as follows:

$$\frac{-d_j}{y_j^2} + v - u_j = 0 \text{ for } j = 1,\ldots, n$$

$$u_j y_j = 0,\ u_j \geq 0, \text{ and } y_j \geq 0 \text{ for } j = 1,\ldots, n.$$

Readily, for each $j = 1,\ldots, n$, y_j must take on a positive value, and hence $u_j = 0$, $\forall j = 1,\ldots,n$. The KKT system thus yields $y_j = \frac{\sqrt{d_j}}{\sqrt{v}}$, $\forall j$, which upon summing and using $\sum_{j=1}^{n} y_j = 1$ gives $v = \left[\sum_{j=1}^{n} \sqrt{d_j}\right]^2$. Thus $(\bar{y}, \bar{v}, \bar{u})$ given by $\bar{y}_j = \frac{\sqrt{d_j}}{\sum_{j=1}^{n} \sqrt{d_j}}$, $\forall j = 1,\ldots,n$, $\bar{v} = \left[\sum_{j=1}^{n} \sqrt{d_j}\right]^2$, and $\bar{u}_j = 0$, $\forall j = 1,\ldots,n$, is the (unique) solution to the above KKT system. The unique KKT point for the original problem is thus given by

$$\bar{x}_j = \frac{b\sqrt{a_j c_j}}{a_j \sum_{j=1}^{n} \sqrt{a_j c_j}}, \quad \forall j = 1,\ldots,n.$$

4.15 Consider the problem

Minimize $\sum_{j=1}^{n} x_j$

subject to $\prod_{j=1}^{n} x_j = b$,

$x_j \geq 0$, $\forall j = 1,\ldots,n$,

where b is a positive constant. Since feasibility requires $x_j > 0$, $\forall j = 1,...,n,$ the only active constraint is the equality restriction, and because of the linear independence constraint qualification, the KKT conditions are necessary for optimality. The KKT system for this problem is thus given as follows:

$$1 + v\prod_{i\neq j}^{n} x_i = 0 \ \text{ for } j = 1,\ldots, n$$

$$\prod_{j=1}^{n} x_j = b.$$

By multiplying the jth equation by x_j for $j = 1,\ldots,n$, and noting that $\prod_{j=1}^{n} x_j = b,$ we obtain

$$x_j + vb = 0 \ \text{ for } j = 1,\ldots, n.$$

Therefore, $\sum_{j=1}^{n} x_j + nbv = 0,$ which gives the unique value for the Lagrange multiplier $v = -\sum_{j=1}^{n} x_j/nb.$ By substituting this expression for v into each of the equations $x_j + vb = 0$ for $j = 1,\ldots, n$, we then obtain $x_j = \frac{1}{n}\sum_{k=1}^{n} x_k$ for $j = 1,\ldots, n$. This necessarily implies that the values of x_j are all identical, and since $\prod_{j=1}^{n} x_j = b,$ we have that $\bar{x}_j = b^{1/n},$ $\forall j = 1,...,n$ yields the unique KKT solution, and since the KKT conditions are necessary for optimality, this gives the unique optimum to the above problem. Therefore, $\frac{1}{n}\sum_{j=1}^{n} \bar{x}_j = b^{1/n}$ is the optimal objective function value. We have thus shown that for any positive vector x such that $\prod_{j=1}^{n} x_j = b,$ we have that

$$\frac{1}{n}\sum_{j=1}^{n} x_j \geq \text{minimum}\left\{\frac{1}{n}\sum_{j=1}^{n} x_j : \prod_{j=1}^{n} x_j = b\right\} = b^{1/n} = \left(\prod_{j=1}^{n} x_j\right)^{1/n}.$$

But, for any given positive vector x, the product of its components is a constant, and so the above inequality implies that $\frac{1}{n}\sum_{j=1}^{n} x_j \geq \left(\prod_{j=1}^{n} x_j\right)^{1/n}$.

Furthermore, if any component is zero, then this latter inequality holds trivially. □

4.27 a. $d = 0$ is a feasible solution and it gives the objective function value equal to 0. Therefore, $\bar{z} \leq 0$.

b. If $\bar{z} < 0$, then $\nabla f(\bar{x})^t \bar{d} < 0$. By Theorem 4.1.2, $\bar{d}$ is a descent direction. Furthermore, by the concavity of $g_i(x)$ at $\bar{x}$, $i \in I$, since $g_i(\bar{x}) = 0$, there exists a $\delta > 0$ such that $g_i(\bar{x} + \lambda\bar{d}) \leq \lambda\nabla g_i(\bar{x})^t \bar{d}$ for $\lambda \in (0, \delta)$. Since the vector $\bar{d}$ is a feasible solution to the given problem, we necessarily have $\nabla g_i(\bar{x})^t \bar{d} \leq 0$ for $i \in I$, and thus $g_i(\bar{x} + \lambda\bar{d}) \leq 0$ for $\lambda \in (0, \delta)$. All the remaining constraint functions are continuous at $\bar{x}$, and so again there exists a $\delta_1 > 0$ such that $g_i(\bar{x} + \lambda\bar{d}) \leq 0$ for $\lambda \in (0, \delta_1)$, $\forall i = 1, \ldots, m$. This shows that $\bar{d}$ is a feasible descent direction at $\bar{x}$.

c. If $\bar{z} = 0$, then the dual to the given linear program has an optimal solution of objective function value zero. This dual problem can be formulated as follows:

$$\begin{array}{ll}
\text{Maximize} & -v_1^t e - v_2^t e \\
\text{subject to} & -\sum_{i \in I} \nabla g_i(\bar{x}) u_i - v_1 + v_2 = \nabla f(\bar{x}) \\
& u_i \geq 0 \text{ for } i \in I,\ v_1 \geq 0,\ v_2 \geq 0,
\end{array}$$

where $e \in R^n$ is a vector of ones. Thus if $\bar{z} = 0$, then v_1 and v_2 are necessarily equal to 0 at an optimal dual solution, and so there exist nonnegative numbers u_i, $i \in I$, such that $\nabla f(\bar{x}) + \sum_{i \in I} \nabla g_i(\bar{x}) u_i = 0$.

Thus, $\bar{x}$ satisfies the KKT conditions.

4.28 Consider the unit simplex $S = \{y : e^t y = 1,\ y \geq 0\}$, which is essentially an $(n-1)$-dimensional body. Its center is given by $y_0 = (\frac{1}{n}, \ldots, \frac{1}{n})^t$.

Examine a (maximal) sphere with center y_0 and radius r that is inscribed with S. Then, r is the distance from y_0 to the center of the one less dimensional simplex, say, formed in the $(y_1,...,y_{n-1})$-space, where the latter center in the full y-space is thus given by $(\frac{1}{n-1},...,\frac{1}{n-1},0)$. Hence, we get

$$r^2 = (n-1)\left[\frac{1}{(n-1)} - \frac{1}{n}\right]^2 + \frac{1}{n^2} = \frac{1}{n(n-1)}.$$

Therefore, the given problem examines the $(n - 1)$-dimensional sphere formed by the intersection of the sphere given by $\|y - y_0\|^2 \le r^2$ with the hyperplane $e^t y = 1,$ without the nonnegativity restrictions $y \ge 0,$ and seeks the minimal value of any coordinate in this region, say, that of y_1. The KKT conditions for this problem are as follows:

$$\begin{aligned}
&2(y_1 - y_{01})u_0 + v = -1\\
&2(y_i - y_{0i})u_0 + v = 0 \qquad \text{for } i = 2,\ldots, n\\
&\|y - y_0\|^2 \le 1/n(n-1)\\
&e^t y = 1\\
&u_0 \ge 0,\ u_0\left(\|y - y_0\|^2 - \frac{1}{n(n-1)}\right) = 0.
\end{aligned}$$

Let $\bar{y} = [0, \frac{1}{n-1},...,\frac{1}{n-1}]^t$. To show that $\bar{y}$ is a KKT point for this problem, all that one needs to do is to substitute $\bar{y}$ for y in the foregoing KKT system and verify that the resulting system in (u_0, v) has a solution. Readily, $e^t \bar{y} = \frac{n-1}{n-1} = 1,$ and $\bar{y} - y_0$ has $(n - 1)$ coordinates equal to $\frac{1}{n(n-1)},$ and one coordinate (the first one) equal to $-\frac{1}{n},$ so that $\|\bar{y} - y_0\|^2 = \frac{1}{n(n-1)}.$ This means that $\bar{y}$ is a feasible solution. Moreover, the equations for indices 2 through n of the KKT system yield

$v = -\dfrac{2u_0}{n(n-1)}$, which together with the first equation gives $u_0 = \dfrac{n-1}{2} \geq 0$. Thus, $\bar{y}$ is a KKT point for this problem. Since the problem is a convex program, this is an optimal solution. Thus, since this is true for minimizing any coordinate of y, even without the nonnegativity constraints present explicitly, the intersection is embedded in the nonnegative orthant.

4.30 Substitute $y = x - \bar{x}$ to obtain the following equivalent form of Problem $\overline{\mathrm{P}}$:

$$\text{Minimize } \left\{\|y - d\|^2 : Ay = 0\right\}.$$

a. Problem $\overline{\mathrm{P}}$ seeks a vector in the nullspace of A that is closest to the given vector d, i.e., to the vector $-\nabla f(\bar{x})$. Since the rank of A is m, an optimal solution to the problem $\overline{\mathrm{P}}$ is the orthogonal projection of the vector $-\nabla f(\bar{x})$ onto the nullspace of A (i.e., start from $\bar{x}$, take a unit step along $-\nabla f(\bar{x})$, and then project the resulting point orthogonally back onto the constraint surface $Ax = b$).

b. The KKT conditions for Problem $\overline{\mathrm{P}}$ are as follows:

$$\begin{aligned} x + A^t v &= \bar{x} + d \\ Ax &= b. \end{aligned}$$

The objective function of $\overline{\mathrm{P}}$ is strictly convex, and the constraints are linear, and so the KKT conditions for Problem $\overline{\mathrm{P}}$ are both necessary and sufficient for optimality.

c. If $\bar{x}$ is a KKT point for Problem $\overline{\mathrm{P}}$, then there exists a vector $\bar{v}$ of Lagrange multipliers associated with the equations $Ax = b$, such that

$$A^t\bar{v} = d, \text{ that is, } \nabla f(\bar{x}) + A^t\bar{v} = 0.$$

Hence, $\bar{x}$ is a KKT point for Problem P provided $\bar{v} \geq 0$.

d. From the KKT system, we get

$$\hat{x} = \bar{x} + d - A^t v \qquad (1)$$

Multiplying (1) by A and using $A\hat{x} = A\bar{x} = b$, we get

$$AA^t v = Ad.$$

Since A is of full row rank, the $(m \times m)$ matrix AA^t is nonsingular. Thus, $v = (AA^t)^{-1}Ad$. Substituting this into (1), we get $\hat{x} = \bar{x} + d - A^t(AA^t)^{-1}Ad$.

4.31 Let $c^t = \nabla_N f(\bar{x})^t - \nabla_B f(\bar{x})^t B^{-1}N$. The considered direction finding problem is a linear program in which the function $c^t d$ is to be minimized over the region $\{d : 0 \le d_j \le 1,\ j \in J\}$, where J is the set of indices for the nonbasic variables. It is easy to verify that $c^t \bar{d}_N \le 0$ at optimality. In fact, an optimal solution to this problem is given by: $\bar{d}_j = 0$ if $c_j \ge 0$, and $\bar{d}_j = 1$ if $c_j < 0$, $\forall j \in J$. To verify if $\bar{d}$ is an improving direction, we need to examine if $\nabla f(\bar{x})^t\, \bar{d} < 0$, where

$$\nabla f(\bar{x})^t\, \bar{d} = [\nabla_B f(\bar{x})^t \quad \nabla_N f(\bar{x})^t]\begin{bmatrix} \bar{d}_B \\ \bar{d}_N \end{bmatrix} =$$

$$[-\nabla_B f(\bar{x})^t B^{-1}N + \nabla_N f(\bar{x})^t]\bar{d}_N = c^t \bar{d}_N.$$

Therefore, $c^t \bar{d}_N < 0$ implies that $\nabla f(\bar{x})^t\, \bar{d} < 0$. Hence, if $\bar{d} \ne 0$, then we must have $\bar{d}_N \ne 0$ (else $\bar{d}_B = 0$ as well), whence $c^t \bar{d}_N < 0$ from above. This means that $\bar{d}$ is an improving direction at $\bar{x}$. Moreover, to show that $\bar{d}$ is a feasible direction at $\bar{x}$, first, note that

$$A\bar{d} = [B \quad N]\begin{bmatrix} -B^{-1}N\bar{d}_N \\ \bar{d}_N \end{bmatrix} = -N\bar{d}_N + N\bar{d}_N = 0,$$ and therefore, $A(\bar{x} + \lambda\bar{d}) = b$ for all $\lambda \ge 0$. Moreover,

$$\bar{x} + \lambda\bar{d} = \begin{bmatrix} B^{-1}b - \lambda B^{-1}N\bar{d}_N \\ \lambda\bar{d}_N \end{bmatrix} \ge 0$$

for $\lambda > 0$ and sufficiently small since $B^{-1}b > 0$ and $\bar{d}_N \ge 0$, which implies that $\bar{x} + \lambda\bar{d} \ge 0$ for all $0 \le \lambda \le \bar{\lambda}$, where $\bar{\lambda} > 0$. Thus, $\bar{d}$ is a feasible direction at $\bar{x}$. Hence, $\bar{d} \neq 0$ implies that $\bar{d}$ is an improving feasible direction.

Finally, suppose that $\bar{d} = 0$, which means that $\bar{d}_N = 0$. Then $c \geq 0$. The KKT conditions at $\bar{x}$ for the original problem can then be written as follows:

$$\nabla_B f(\bar{x}) - u_B + B^t v = 0$$
$$\nabla_N f(\bar{x}) - u_N + N^t v = 0$$
$$u_B^t \bar{x}_B = 0, \; u_N^t \bar{x}_N = 0, \; u_B \geq 0, \; u_N \geq 0.$$

Let $\bar{u}_B = 0$, $\bar{v}^t = -\nabla_B f(\bar{x})^t B^{-1}$, and $\bar{u}_N^t = \nabla_N f(\bar{x})^t - \nabla_B f(\bar{x})^t B^{-1} N$. Simple algebra shows that $(\bar{u}_B, \bar{u}_N, \bar{v})$ satisfies the above system (solve for v from the first equation and substitute it in the second equation). Therefore, $\bar{x}$ is a KKT point whenever $\bar{d} = 0$ (and is optimal if, for example, f is pseudoconvex). □

4.33 In the first problem, the KKT system is given by:

$$c + Hx + A^t u = 0 \tag{1}$$
$$Ax + y = b \tag{2}$$
$$u^t y = 0$$
$$x \geq 0, \; y \geq 0, \; u \geq 0.$$

Since the matrix H is invertible, Equation (1) yields $H^{-1}c + x + H^{-1}A^t u = 0$. By premultiplying this equation by A, we obtain $AH^{-1}c + Ax + AH^{-1}A^t u = 0$, which can be rewritten as

$$AH^{-1}c + b + Ax - b + AH^{-1}A^t u = 0. \tag{3}$$

Next, note that from Equation (2), we have $y = b - Ax$, so that Equation (3) can be further rewritten as

$$h + Gu - y = 0, \text{ where } u^t y = 0, \; u \geq 0, \; y \geq 0. \tag{4}$$

In the second given problem, the KKT system is given by

$$h + Gv - z = 0, \; v^t z = 0, \; v \geq 0, \; z \geq 0, \tag{5}$$

where z is the vector of Lagrange multipliers. By comparing (4) and (5), we see that the two problems essentially have identical KKT systems, where $u \equiv v$ and $y \equiv z$, that is, the Lagrange multipliers in the first problem are decision variables in the second problem, while the Lagrange multipliers in the second problem are slack variables in the first problem.

4.37 We switch to minimizing the function $f(x_1, x_2) = -x_1^2 - 4x_1x_2 - x_2^2$.

a. The KKT system is as follows:

$$\begin{aligned}
-2x_1 \quad -4x_2 \quad +2vx_1 \quad &= 0 \\
-4x_1 \quad -2x_2 \quad +2vx_2 \quad &= 0 \\
x_1^2 \quad +x_2^2 \quad &= 1.
\end{aligned}$$

There are four solutions to this system:

$$\begin{aligned}
(x_1, x_2) &= (1/\sqrt{2}, 1/\sqrt{2}), \text{ and } v = 3 \\
(x_1, x_2) &= (-1/\sqrt{2}, \ -1/\sqrt{2}), \text{ and } v = 3 \\
(x_1, x_2) &= (1/\sqrt{2}, \ -1/\sqrt{2}), \text{ and } v = -1 \\
(x_1, x_2) &= (-1/\sqrt{2}, \ 1/\sqrt{2}), \text{ and } v = -1.
\end{aligned}$$

The objective function $f(x_1, x_2)$ takes on the value of –3 for the first two points, and the value of 1 at the remaining two. Since the linear independence constraint qualification (CQ) holds, the KKT conditions are necessary for optimality. Hence, there are two optimal solutions: $\bar{x}_1 = (1/\sqrt{2}, \ 1/\sqrt{2})$ and $\bar{x}_2 = (-1/\sqrt{2}, \ -1/\sqrt{2})$. To support this statement, one can use a graphical display, or use the second-order sufficiency condition given in Part (b) below.

b. $L(x) = -x_1^2 - 4x_1x_2 + v(x_1^2 + x_2^2 - 1)$. Therefore,

$$\nabla^2 L(x) = 2\begin{bmatrix} v-1 & -2 \\ -2 & v-1 \end{bmatrix}.$$

For $v = 3$, $\nabla^2 L(x)$ is a positive definite matrix and therefore, $\bar{x}_1 = (\sqrt{2}/2, \ \sqrt{2}/2)$ and $\bar{x}_2 = (-\sqrt{2}/2, \ -\sqrt{2}/2)$ are both strict local optima.

c. See answers to Parts (a) and (b).

4.41 a. See the proof of Lemma 10.5.3.

b. See the proof of Theorem 10.5.4.

c. Let P_d denote the given problem and note that since this problem is convex, the KKT conditions are sufficient for optimality. Hence, it is sufficient to produce a KKT solution to Problem P_d of the given form $\hat{d}$. Toward this end, consider the KKT conditions for Problem P_d:

$$\nabla f(\bar{x}) + A_1^t v + 2du = 0 \tag{1}$$
$$A_1 d = 0 \tag{2}$$
$$\|d\|^2 \le 1,\ u \ge 0,\ u(\|d\|^2 - 1) = 0.$$

Premultiplying (1) by A_1 and using (2), we get

$$A_1 \nabla f(\bar{x}) + A_1 A_1^t v = 0.$$

Since A_1 is of full (row) rank, $A_1 A_1^t$ is nonsingular, and so we get

$$v = -(A_1 A_1^t)^{-1} A_1 \nabla f(\bar{x}). \tag{3}$$

Thus, (1) yields

$$2du = -P\nabla f(\bar{x}) = \bar{d}. \tag{4}$$

Hence, if $\bar{d} = 0$, we can take $d = \hat{d} = 0$ and $u = 0$, which together with (3) yields $\hat{d} = 0$ as a KKT point (hence, an optimum to P_d, with say, $\lambda \equiv 1$). On the other hand, if $\bar{d} \ne 0$, then let $\hat{d} = \dfrac{\bar{d}}{\|\bar{d}\|}$, $u = \dfrac{\|\bar{d}\|}{2}$, and let v be given by (3). Thus, noting that $A_1\hat{d} = 0$ since $A_1 P = 0$, we get that $\hat{d}$ is a KKT point and hence an optimum to P_d (with $\lambda \equiv \|\bar{d}\| > 0$).

d. If $A = -I_n$, then A_1 is an $m \times n$ submatrix of $-I_n$, where m is the number of variables that are equal to zero at the current solution $\bar{x}$. Then $A_1 A_1^t = I_m$, and $A_1^t A_1 = \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix}$. Therefore, $P = \begin{bmatrix} 0 & 0 \\ 0 & I_{n-m} \end{bmatrix}$, and $\bar{d}_j = 0$ if $\bar{x}_j = 0$, and $\bar{d}_j = -\dfrac{\partial f(\bar{x})}{\partial x_j}$ if $\bar{x}_j > 0$. Hence, $\bar{d}$ is the projection of $-\nabla f(\bar{x})$ onto the nullspace of the active (nonnegativity) constraints.

4.43 Note that $C = \{d : Ad = 0\}$ is the nullspace of A, and P is the projection matrix onto the nullspace of A. If $d \in C$, then $Pd = d$, and so, $d = Pw$ with $w \equiv d$. On the other hand, if $d = Pw$ for some $w \in R^n$, we have that $Ad = APw = 0$ since $AP = 0$. Hence, $d \in C$. This shows that $d \in C$ if and only if there exists a $w \in R^n$ such that $Pw = d$. Next, we show that if H is a symmetric matrix, then $d^t Hd \geq 0$ for all $d \in C$ if and only if $P^t HP$ is positive semidefinite.

($\Rightarrow$) Suppose that $d^t Hd \geq 0$ for all $d \in C$. Consider any $w \in R^n$ and let $d = Pw$. Then $Ad = APw = 0$ since $AP = 0$, and so $d \in C$. Thus $d^t Hd \geq 0$, which yields $w^t P^t HPw \geq 0$ for any $w \in R^n$. Hence, the matrix $P^t HP$ is positive semidefinite.

($\Rightarrow$) If $w^t P^t HPw \geq 0$ for all $w \in R^n$, then in particular for any $d \in C$, we have $d^t P^t HPd \geq 0$, which gives $d^t Hd \geq 0$ since for any $d \in C$ we have $Pd = d$. $\square$

CHAPTER 5:

CONSTRAINT QUALIFICATIONS

5.1 Let T denote the cone of tangents of S at $\bar{x}$ as given in Definition 5.1.1.

a. Let W denote the set of directions defined in this part of the exercise. That is, $d \in W$ if there exists a nonzero sequence $\{\beta_k\}$ convergent to zero, and a function $\alpha{:}R \to R^n$ that converges to 0 as $\beta \to 0$, such that $\bar{x} + \beta_k d + \beta_k \alpha(\beta_k) \in S$ for any k. We need to show that $W = T$. First, note that $0 \in W$ and $0 \in T$. Now, let d be a nonzero vector from the set T. Then there exist a positive sequence $\{\lambda_k\}$ and a sequence $\{x_k\}$ of points from S convergent to $\bar{x}$ such that $d = \lim_{k\to\infty} \lambda_k(x_k - \bar{x})$. Without loss of generality, assume that $x_k \neq \bar{x}$, $\forall k$ (since $d \neq 0$). Therefore, for this sequence $\{x_k\}$, consider the nonzero sequence $\{\beta_k\}$ such that $\beta_k d$ is the projection of $x_k - \bar{x}$ onto the vector d. Hence, $\{\beta_k\} \to 0^+$. Furthermore, let $y_k \equiv (x_k - \bar{x}) - \beta_k d$. Because of the projection operation, we have that

$$\|x_k - \bar{x}\|^2 = \beta_k^2 \|d\|^2 + \|y_k\|^2,$$

i.e.,
$$\frac{\|y_k\|^2}{\beta_k^2} = \|d\|^2 \left[\frac{\|x^k - \bar{x}\|^2}{\beta_k^2 \|d\|^2} - 1 \right]. \tag{1}$$

But we have that $\dfrac{\beta_k \|d\|}{\|x^k - \bar{x}\|} = \cos(\gamma_k)$, where γ_k is the angle between $(x_k - \bar{x})$ and d. Since $d \in T$, we have that $\gamma_k \to 0$ and so $\cos(\gamma_k) \to 1$ and thus $\dfrac{\|y_k\|}{\beta_k} \to 0$ from (1). Consequently, we can define $\alpha{:}R \to R^n$ such that $\beta_k \alpha(\beta_k) = y_k$ so that $x_k = \bar{x} + \beta_k d +$

$\beta_k \alpha(\beta_k) \in S$, $\forall k$, with $\alpha(\beta_k) = y_k / \beta_k \to 0$ as $\beta_k \to 0$. Hence, $d \in W$.

Next, we show that if $d \in W$, then $d \in T$. For this purpose, let us note that if $d \in W$, then the sequence $\{x_k\} \subseteq S$, where $x_k = \bar{x} + \beta_k d + \beta_k \alpha(\beta_k)$, converges to $\bar{x}$, and moreover, the sequence $\left\{ \frac{1}{\beta_k}(x_k - \bar{x}) - d \right\}$ converges to the zero vector. This shows that there exists a sequence $\{\lambda_k\}$, where $\lambda_k = \frac{1}{\beta_k}$, and a sequence $\{x_k\}$ of points from S convergent to $\bar{x}$ such that $d = \lim_{k\to\infty} \lambda_k (x_k - \bar{x})$. This means that $d \in T$, and so the proof is complete.

b. Again, let W denote the set of directions defined in this part of the exercise. That is, $d \in W$ if there exists a nonnegative scalar λ and a sequence $\{x_k\}$ of points from S convergent to $\bar{x}$, $x_k \neq \bar{x}$ for all k, such that $d = \lim_{k\to\infty} \lambda \frac{x_k - \bar{x}}{\|x_k - \bar{x}\|}$. Again in this case, we have $0 \in W$ and $0 \in T$, and so let d be a nonzero vector in T. Then there exists a sequence $\{x_k\}$ of points from S different from $\bar{x}$ and a positive sequence $\{\lambda_k\}$ such that $x_k \to \bar{x}$, and $d = \lim_{k\to\infty} \lambda_k \|x_k - \bar{x}\| \frac{x_k - \bar{x}}{\|x_k - \bar{x}\|}$. Under the assumption that $d \in T$, the sequence $\left\{ \lambda_k \|x_k - \bar{x}\| \right\}$ is contained in a compact set. Therefore, it must have a convergent subsequence. Without loss of generality, assume that the sequence $\left\{ \frac{x_k - \bar{x}}{\|x_k - \bar{x}\|} \right\}$ itself is convergent. If so, then we conclude that $d = \lim_{k\to\infty} \lambda \frac{x_k - \bar{x}}{\|x_k - \bar{x}\|}$, where $\lambda = \|d\|$. Hence, $d \in W$.

Conversely, let $d \in W$, where again, $d \neq 0$. Then we can simply take $\lambda_k = \dfrac{\lambda}{\|x_k - \bar{x}\|} > 0$ to readily verify that $d \in T$. This completes the proof. $\square$

5.12 a. See the proof of Theorem 10.1.7.

b. By Part (a), $\bar{x}$ is a FJ point. Therefore, there exist scalars u_0 and u_i for $i \in I$, such that

$$u_0 \nabla f(\bar{x}) + \sum_{i \in I} u_i \nabla g_i(\bar{x}) = 0,$$
$$u_0 \geq 0,\ u_i \geq 0 \text{ for } i \in I,\ (u_0, u_i \text{ for } i \in I) \neq 0.$$

If $u_0 = 0$, then the system

$$\sum_{i \in I} u_i \nabla g_i(\bar{x}) = 0,$$
$$u_i \geq 0 \text{ for } i \in I$$

has a nonzero solution. Then, by Gordan's Theorem, no vector d exists such that $\nabla g_i(\bar{x})^t d < 0$ for all $i \in I$. This means that $G_0 = \emptyset$, and so $c\ell(G_0) = \emptyset$, whereas $G' \neq \emptyset$ (since $0 \in G'$). This contradicts Cottle's constraint qualification.

5.13 a. $\bar{x} = [1\ \ 0]^t$, $I = \{1, 2\}$, $\nabla g_1(\bar{x}) = [2\ \ 0]^t$, $\nabla g_2(\bar{x}) = [0\ \ -1]^t$. The gradients of the binding constraints are linearly independent; hence, the linear independence constraint qualification holds. This implies that Kuhn-Tucker's constraint qualification also holds (see Figure 5.2 in the text and its associated comments).

b. If $\bar{x} = [1\ \ 0]^t$, then the KKT conditions yields:

$$\begin{aligned} -1 \quad + 2u_1 \qquad\quad &= 0 \\ - u_2 &= 0, \end{aligned}$$

i.e., $u_1 = \dfrac{1}{2}$ and $u_2 = 0$. Since the Lagrange multipliers are nonnegative, we conclude that $\bar{x}$ is a KKT point.

Note that a feasible solution must be in the unit circle centered at the origin; hence no feasible solution can have its first coordinate greater than 1. Therefore, $\bar{x}$ (which yields the objective value of -1) is the global optimal solution.

5.15 X is an open set, the functions of nonbinding constraints are continuous at $\bar{x}$, and the functions whose indices are in the set J are pseudoconcave at $\bar{x}$. Therefore, by the same arguments as those used in the proof of Lemma 4.2.4, any vector d that satisfies the inequalities $\nabla g_i(\bar{x})^t d \le 0$ for $i \in J$, and $\nabla g_i(\bar{x})^t d < 0$ for $i \in I - J$ is a feasible direction at $\bar{x}$. Hence, if $\bar{x}$ is a local minimum, then the following system has no solution:

$$\nabla f(\bar{x})^t d < 0$$
$$\nabla g_i(\bar{x})^t d < 0 \text{ for } i \in I - J$$
$$\nabla g_i(\bar{x})^t d \le 0 \text{ for } i \in J.$$

Accordingly, consider the following pair of primal and dual programs P and D, where $y_0 \in R$ is a dummy variable:

P: Maximize y_0

subject to
$$\nabla f(\bar{x})^t d + y_0 \le 0$$
$$\nabla g_i(\bar{x})^t d + y_0 \le 0, \quad \forall i \in I - J$$
$$\nabla g_i(\bar{x})^t d \le 0, \quad \forall i \in J.$$

D: Minimize 0

subject to
$$u_0 \nabla f(\bar{x}) + \sum_{i \in I} u_i \nabla g_i(\bar{x}) = 0 \tag{1}$$
$$u_0 + \sum_{i \in I-J} u_i = 1 \tag{2}$$
$$(u_0, u_i \text{ for } i \in I) \ge 0. \tag{3}$$

Then, since the foregoing system has no solution, then we must have that P has an optimal value of zero (since if $y_0 > 0$ for a feasible solution (y_0, d), then P is unbounded), which means that D is feasible, i.e., (1) – (3) has a solution. If $u_0 > 0$ in any such solution, then $\bar{x}$ is a KKT point and we are done. Else, suppose that $u_0 = 0$, which implies by (2) that $I - J \neq \varnothing$. Furthermore, letting d belong to the given nonempty set in the exercise such that $\nabla g_i(\bar{x})^t d \leq 0$ for $i \in J$, and $\nabla g_i(\bar{x})^t d < 0$ for $i \in I - J \neq \varnothing$, we have by taking the inner product of (2) with d that

$$\sum_{i \in J} u_i \nabla g_i(\bar{x})^t d + \sum_{i \in I-J} u_i \nabla g_i(\bar{x})^t d = 0,$$

which yields a contradiction since the first term above is nonpositive and the second term above is strictly negative because $u_i > 0$ for at least one $i \in I - J \neq \varnothing$. Thus $u_0 > 0$, and so $\bar{x}$ is a KKT point. $\square$

5.20 Let $g(d) \equiv d^t d - 1 \leq 0$ be the nonlinear defining constraint. Then $\nabla g(\bar{d})^t d = 2\bar{d}^t d$. Hence G_1 is the set G' defined in the text, and so by Lemma 5.2.1, we have that $T \subseteq G_1$. Therefore, we need to show that $G_1 \subseteq T$. Let d be a nonzero vector from G_1. If $\bar{d}^t d < 0$, i.e., $\nabla g(\bar{d})^t d < 0$, then we readily have that $d \in D$ (see the proof of Lemma 4.2.4 for details), and hence, $d \in T$. Thus, suppose that $\bar{d}^t d = 0$. Then d is tangential to the sphere $d^t d \leq 1$. Hence, since $C_2 \bar{d} < 0$, and $C_1 d \leq 0$ with $d \neq \bar{d}$, there exists a sequence $\{d_k\}$ of feasible points $d_k \to \bar{d}$, $d_k \neq \bar{d}$, and $d_k^t d_k = 1$ such that $\dfrac{d}{\|d\|} = \lim_{k \to \infty} \dfrac{d_k - \bar{d}}{\|d_k - \bar{d}\|}$. Therefore, $d \in T$.

CHAPTER 6:

LAGRANGIAN DUALITY AND SADDLE POINT OPTIMALITY CONDITIONS

6.2 For the problem illustrated in Figure 4.13, a possible sketch of the perturbation function $v(y)$ and the set G are very much similar to that shown in Figure 6.1 (note that the upper envelope of G also increases with y, and only a partial view of G (from above) is shaded in Figure 6.1. Hence, as in Figure 6.1, there is no duality gap for this case.

6.3 Let the left-hand side of the inequality be given by $\phi(\hat{x},\hat{y})$. Hence, we get

$$\sup_{y\in Y}\inf_{x\in X}\phi(x,y) = \phi(\hat{x},\hat{y}) = \inf_{x\in X}\phi(x,\hat{y}) \le \inf_{x\in X}\sup_{y\in Y}\phi(x,y). \quad \square$$

6.4 Let $y_\lambda = \lambda y_1 + (1-\lambda)y_2$, where y_1 and $y_2 \in R^{m+\ell}$ and where $\lambda \in [0,1]$. We need to show that $v(y_\lambda) \le \lambda v(y_1) + (1-\lambda)v(y_2)$. For this purpose, let

$X(y_1) = \{x : g_i(x) \le y_{1i},\ i = 1,\dots,m,\ h_i(x) = y_{1,m+i},\ i = 1,\dots,\ell,\ x \in X\}$
$X(y_2) = \{x : g_i(x) \le y_{2i},\ i = 1,\dots,m,\ h_i(x) = y_{2,m+i},\ i = 1,\dots,\ell,\ x \in X\}$
$X(y_\lambda) = \{x : g_i(x) \le y_{\lambda i},\ i = 1,\dots,m,\ h_i(x) = y_{\lambda,m+i},\ i = 1,\dots,\ell,\ x \in X\}$
$v(y_k) = f(x_k)$, where x_k optimizes (6.9) when $y = y_k$, for $k = 1, 2$, and let
$v(y_\lambda) = f(x^*)$, where x^* optimizes (6.9) when $y = y_\lambda$.

By the definition of the perturbation function $v(y)$, this means that

$x_k \in X(y_k)$ and $f(x_k) = \min\{f(x) : x \in X(y_k)\}$ for $k = 1, 2$, and
$x^* \in X(y_\lambda)$ and $f(x^*) = \min\{f(x) : x \in X(y_\lambda)$.

Under the given assumptions (the functions $g_i(x)$ are convex, the functions $h_i(x)$ are affine, and the set X is convex) we have from the definition of convexity that $x_\lambda = \lambda x_1 + (1-\lambda)x_2 \in X(y_\lambda)$ for any $\lambda \in [0,1]$. But $f(x^*) = \min\{f(x) : x \in X(y_\lambda)\}$, and so $f(x^*) \le f(x_\lambda)$, which together with the convexity of $f(x)$ implies that $v(y_\lambda) = f(x^*) \le f(x_\lambda) \le \lambda f(x_1) + (1-\lambda)f(x_2) = \lambda v(y_1) + (1-\lambda)v(y_2)$. This completes the proof. $\square$

6.5 The perturbation function from Equation (6.9) for Example 6.3.5 is given by

$$v(y) = \min\{-x_1 - x_2 : x_1 + 2x_2 \le 3 + y, \text{ with } x_1, x_2 \in \{0,1,2,3\}\}.$$

Hence, by examining the different combinations of discrete solutions in the (x_1, x_2)-space, we get

$$v(y) = \begin{cases} \infty & \text{if} \quad y < -3 \\ 0 & \text{if} \quad -3 \le y < -2 \quad \text{[evaluated at } (x_1,x_2) = (0,0)] \\ -1 & \text{if} \quad -2 \le y < -1 \quad \text{[evaluated at } (x_1,x_2) = (1,0)] \\ -2 & \text{if} \quad -1 \le y < 0 \quad \text{[evaluated at } (x_1,x_2) = (2,0)] \\ -3 & \text{if} \quad 0 \le y < 2 \quad \text{[evaluated at } (x_1,x_2) = (3,0)] \\ -4 & \text{if} \quad 2 \le y < 4 \quad \text{[evaluated at } (x_1,x_2) = (3,1)] \\ -5 & \text{if} \quad 4 \le y < 6 \quad \text{[evaluated at } (x_1,x_2) = (3,2)] \\ -6 & \text{if} \quad y \ge 6 \quad \text{[evaluated at } (x_1,x_2) = (3,3)] \end{cases}$$

Note that the optimal primal solution is given by $(\bar{x}_1, \bar{x}_2) = (3,0)$ of objective value –3, which also happens to be the optimum to the underlying linear programming relaxation in which we restrict x_1 and x_2 to lie in [0, 3], thus portending the existence of a saddle point solution. Indeed, for $\bar{u} = 1$, we see from Example 6.3.5 that $\theta(\bar{u}) = -3$, and so $(\bar{x}_1, \bar{x}_2, \bar{u})$ is a saddle point solution and there does not exist a duality gap in this example. Moreover, we see that

$$v(y) \ge 3 - y, \ \forall y$$

as in Equation (6.10) of Theorem 6.2.7, thus verifying the necessary and sufficient condition for the absence of a duality gap.

6.7 Denote $S \equiv conv\{x \in X : Dx = d\}$, and note that since X is a compact discrete set, we have that S is a polytope. Hence, for any linear function $f(x)$, we have $\min f(x) : Dx = d, \ x \in X\} = \min\{f(x) : x \in S\}$. Therefore, for each fixed $\pi \in R^m$, we get

$$\theta(\pi) = \min\{c^t x + \pi^t (Ax - b) : Dx = d, x \in X\}$$
$$= \min\{c^t x + \pi^t (Ax - b) : x \in S\}.$$

Now, consider the LP : $\min\{c^t x : Ax = b,\ x \in S\}$. Then, by strong duality for LPs, we get

$$\min\{c^t x : Ax = b,\ x \in S\} = \max_{\pi \in R^m} \min_{x \in S} \{c^t x + \pi^t (Ax - b)\} = \max_{\pi \in R^m} \theta(\pi). \quad (1)$$

This establishes the required result. Moreover, the optimal value of Problem DP is given by $v^* = \min\{c^t x : x \in S^*\}$, where

$$S^* \equiv conv\{x : Ax = b, Dx = d, x \in X\} \subseteq \{x : Ax = b, x \in S\}. \quad (2)$$

Thus, we get $v^* \geq \min\{c^t x : Ax = b, x \in S\}$, which yields from (1) that $v^* \geq \max_{\pi \in R^m} \theta(\pi)$, where a duality gap exists if this inequality is strict. Therefore, the disparity in (2) potentially causes such a duality gap. □

6.8 Interchanging the role of x and y as stated in the exercise for convenience, and noting Exercise 6.7 and Section 6.4, we have

$$v_1 = \max_{\mu} \bar{\theta}(\mu) = \max_{\mu} \min_{(x,y)} \{c^t x + \mu^t (x - y) : Ay = b,\ y \in Y,\ Dx = d,\ x \in X\}$$

and

$$v_2 = \max_{\pi} \theta(\pi) = \max_{\pi} \min_{x} \{c^t x + \pi^t (Ax - b) : Dx = d,\ x \in X\}.$$

Let $S_1 \equiv conv\{x \in X : Dx = d\}$ and $S_2 \equiv conv\{y \in Y : Ay = b\}$.
Then from Exercise 6.7 (see also Section 6.4), we have that

$$v_1 = \min\{c^t x : x \in S_1,\ y \in S_2,\ x = y\} \quad (1)$$

and

$$v_2 = \min\{c^t x : Ax = b,\ x \in S_1\}. \quad (2)$$

Hence, we get

$$v_2 = \min\{c^t x : Ay = b,\ x = y,\ x \in S_1\}$$

$\leq \min\{c^t x : x \in S_1, y \in S_2, x = y\} = v_1$, where the inequality follows since $S_2 \subseteq \{y : Ay = b\}$. This proves the stated result, where (1) and (2) provide the required partial convex hull relationships.

6.9 First, we show that if $(d_u, d_v) = (0,0)$, then $(\bar{u}, \bar{v})$ solves Problem (D). Problem (D) seeks the maximum of a concave function $\theta(u,v)$ over $\{(u,v) : u \geq 0\}$, and so the KKT conditions are sufficient for optimality. To show that $(\bar{u}, \bar{v})$ is a KKT point for (D), we need to demonstrate that there exists a vector z_1 such that

$$
\begin{aligned}
&-\nabla_u \theta(\bar{u}, \bar{v}) - z_1 = 0 \\
&-\nabla_v \theta(\bar{u}, \bar{v}) = 0 \\
&\quad z_1^t \bar{u} = 0, \; z_1 \geq 0.
\end{aligned}
$$

By assumption, we have $\nabla_v \theta(\bar{u}, \bar{v}) = h(\bar{x}) = 0$, and $\nabla_u \theta(\bar{u}, \bar{v}) = g(\bar{x})$. Moreover, since $d_u = 0$, we necessarily have $g(\bar{x}) \leq 0$ and $g(\bar{x})^t \bar{u} = 0$. Thus, $z_1 = -g(\bar{x})$ solves the KKT system, which implies that $(\bar{u}, \bar{v})$ solves (D). (Alternatively, note from above that if $\bar{x}$ evaluates $\theta(\bar{u}, \bar{v})$, then the given condition implies that $\bar{x}$ is feasible to P with $\bar{u}^t g(\bar{x}) = 0$, and hence $(\bar{x}, \bar{u}, \bar{v})$ is a saddle point, and so by Theorem 6.2.5, $\bar{x}$ and $(\bar{u}, \bar{v})$ respectively solve P and D with no duality gap.)

Next, we need to show that if $(d_u, d_v) \neq (0,0)$, then (d_u, d_v) is a feasible ascent direction of $\theta(u,v)$ at $(\bar{u}, \bar{v})$. Notice that v is a vector of unrestricted variables, and by construction $d_{ui} \geq 0$ whenever $\bar{u}_i = 0$. Hence, (d_u, d_v) is a feasible direction at $(\bar{u}, \bar{v})$. To show that it is also an ascent direction, let us consider $\nabla \theta(\bar{u}, \bar{v})^t d$:

$$
\begin{aligned}
\nabla \theta(\bar{u}, \bar{v})^t d &= \nabla_u \theta(\bar{u}, \bar{v})^t d_u + \nabla_v \theta(\bar{u}, \bar{v})^t d_v = g(\bar{x})^t \hat{g}(\bar{x}) + h(\bar{x})^t h(\bar{x}) \\
&= h(\bar{x})^t h(\bar{x}) + \sum_{i:\bar{u}_i > 0} g_i^2(\bar{x}) + \sum_{i:\bar{u}_i = 0} g_i(\bar{x}) \max\{0, \; g_i(\bar{x})\}.
\end{aligned}
$$

All the foregoing terms are nonnegative and at least one of these is positive, for otherwise, we would have $(d_u, d_v) = (0,0)$.

Thus, $\nabla\theta(\bar{u},\bar{v})^t d > 0$. This demonstrates that (d_u, d_v) is an ascent direction of $\theta(u,v)$ at $(\bar{u},\bar{v})$. $\square$

In the given numerical example,

$$\theta(u_1,u_2) = \min\{x_1^2 + x_2^2 + u_1(-x_1 - x_2 + 4) + u_2(x_1 + 2x_2 - 8) : \\ (x_1,x_2) \in R^2\}.$$

Iteration 1: $(u_1, u_2) = (0, 0)$.

At $(u_1, u_2) = (0, 0)$ we have $\theta(0, 0) = 0$, with $\bar{x}_1 = \bar{x}_2 = 0$. Thus, $d_1 = \max\{0, 4\} = 4$, and $d_2 = \max\{0, -8\} = 0$. Next, we need to maximize the function $\theta(u_1,u_2)$ from (0, 0) along the direction (4, 0). Notice that

$$\begin{aligned}
&\theta[(0, 0) + \lambda(4, 0)] = \theta(4\lambda, 0) \\
&= \min\{x_1^2 + x_2^2 + 4\lambda(-x_1 - x_2 + 4) : (x_1, x_2) \in R^2\} \\
&= \min\{x_1^2 - 4\lambda x_1 : x_1 \in R\} \\
&+ \min\{x_2^2 - 4\lambda x_2 : x_2 \in R\} + 16\lambda \\
&= -8\lambda^2 + 16\lambda,
\end{aligned}$$

and $\max\{\theta(4\lambda, 0) : \lambda \geq 0\}$ is achieved at $\lambda^* = 1$. Hence, the new iterate is (4, 0).

Iteration 2: $(u_1, u_2) = (4, 0)$.

At $(u_1, u_2) = (4, 0)$ we readily obtain that

$$\theta(4, 0) = \min\{x_1^2 + x_2^2 + 4(-x_1 - x_2 + 4) : (x_1, x_2) \in R^2\} = 8,$$

with $\bar{x}_1 = \bar{x}_2 = 2$. Thus, $d_1 = g_1(2, 2) = 0$, and $d_2 = \max\{0, -2\} = 0$. Based on the property of the dual problem, we conclude that at $(u_1, u_2) = (4, 0)$ the Lagrangian dual function $\theta(u_1,u_2)$ attains its maximum value. Thus $(\bar{u}_1, \bar{u}_2) = (4, 0)$ is an optimal solution to Problem D.

6.14 Let $\theta_1(v_0, v)$ be the Lagrangian dual function for the transformed problem.

That is, $\theta_1(v_0, v) = \inf\{f(x) + v_0^t(g(x) + s) + v^t h(x) : (x, s) \in X'\}$.
The above formulation is separable in the variables x and s, which yields
$\theta_1(v_0, v) = \inf\{f(x) + v_0^t g(x) + v^t h(x) : x \in X\} + \inf\{v_0^t s : s \geq 0\}$.

Note that if $v_0 \geq 0$, then $\inf\{v_0^t s : s \geq 0\} = 0$, and otherwise, we get $\inf\{v_0^t s : s \geq 0\} = -\infty$. Therefore, the dual problem seeks the unconstrained maximum of $\theta_1(v_0, v)$, where

$$\theta_1(v_0, v) = \begin{cases} \inf\{f(x) + v_0^t g(x) + v^t h(x) : x \in X\} & \text{if } v_0 \geq 0 \\ -\infty & \text{otherwise.} \end{cases}$$

This representation of $\theta_1(v_0, v)$ shows that the two dual problems are equivalent (with $v_0 = u$).

6.15 For simplicity, we switch to the minimization of $f(x) = -3x_1 - 2x_2 - x_3$.

a. $$\theta(u) = -4u_1 - 3u_2 + \min\{(-3 + u_1)x_1 + (-2 + 2u_1)x_2 + (-1 + u_2)x_3 : x \in X\}. \quad (1)$$

The set X has three extreme points $x_1 = (0, 0, 0)$, $x_2 = (1, 0, 0)$, and $x_3 = (0, 2, 0)$, and three extreme directions $d_1 = (0, 0, 1)$, $d_2 = (0, \frac{1}{2}, \frac{1}{2})$, and $d_3 = (\frac{1}{3}, 0, \frac{2}{3})$. Hence, for $\theta(u) > -\infty$, we must have (examining the extreme directions) that

$$u \in U \equiv \{(u_1, u_2) : u_2 \geq 1,\ 2u_1 + u_2 \geq 3,\ u_1 + 2u_2 \geq 5\}. \quad (2)$$

Hence, any $u \geq 0$ such that $u \in U$ will achieve the minimum in (1) at an extreme point, whence,

$$\theta(u) = \min\{-4u_1 - 3u_2,\ -3u_1 - 3u_2 - 3,\ -3u_2 - 4\}. \quad (3)$$

Putting (2) and (3) together and simplifying the conditions, we get

$$\theta(u) = \begin{cases} -4u_1 - 3u_2 \text{ if } u_1 \geq 3 \text{ and } u_2 \geq 1 \\ -3u_1 - 3u_2 - 3 \text{ if } \frac{1}{3} \leq u_1 \leq 3,\ u_2 \geq 1,\ 2u_1 + u_2 \geq 3, \text{ and } u_1 + 2u_2 \geq 5 \\ -3u_2 - 4 \text{ if } u_1 \leq \frac{1}{3} \text{ and } 2u_1 + u_2 \geq 3 \\ -\infty \text{ otherwise.} \end{cases}$$

b. In this case, we get

$$\theta(u) = -2u_1 - 3u_2 + \min_{x \in X}\{(-3 + 2u_1)x_1 + (-2 + u_1)x_2 + (-1 - u_1 + u_2)x_3\}$$

i.e.,

$$\theta(u) = -2u_1 - 3u_2 + \min\{(-3 + 2u_1)x_1 + (-2 + u_1)x_2 : x_1 + 2x_2 \leq 4, (x_1, x_2) \geq 0\} + \min\{(-1 - u_1 + u_2)x_3 : x_3 \geq 0\}.$$

Noting that the extreme points of the polytope in the first minimization problem in (x_1, x_2) are (0, 0), (4, 0), and (0, 2), and that the second minimization problem yields an optimal objective function value of zero if $-u_1 + u_2 \geq 1$ and goes to $-\infty$ otherwise, we get that

$$\theta(u) = -2u_1 - 3u_2 + \min\{0,\ -12 + 8u_1,\ -4 + 2u_1\} \text{ if } -u_1 + u_2 \geq 1,$$

and is $-\infty$ otherwise. Thus,

$$\theta(u) = \begin{cases} -2u_1 - 3u_2 \text{ if } u_1 \geq 2 \text{ and } -u_1 + u_2 \geq 1 \\ -12 + 6u_1 - 3u_2 \text{ if } u_1 \leq 4/3 \text{ and } -u_1 + u_2 \geq 1 \\ -4 - 3u_2 \text{ if } 4/3 \leq u_1 \leq 2 \text{ and } -u_1 + u_2 \geq 1 \\ -\infty \text{ otherwise.} \end{cases}$$

c. We can select those constraints to define X that will make the minimization over this set relatively easy, e.g., when the minimization problem decomposes into a finite number of simpler, lower dimensional, independent problems.

6.21 Let $\gamma = \inf\{f(x) : g(x) \leq 0, \; h(x) = 0, \; x \in X\}$. Readily, γ is a finite number, since $\bar{x}$ solves Problem P: minimize $f(x)$ subject to $g(x) \leq 0$, $h(x) = 0$, $x \in X$. Moreover, the system

$$f(x) - \gamma < 0, \; g(x) \leq 0, \; h(x) = 0, \; x \in X$$

has no solution. By Lemma 6.2.3, it then follows that there exists a nonzero vector $(\bar{u}_0, \bar{u}, \bar{v})$, such that $(\bar{u}_0, \bar{u}) \geq 0$, and

$$\bar{u}_0(f(x) - \gamma) + \bar{u}^t g(x) + \bar{v}^t h(x) \geq 0 \text{ for all } x \in X. \tag{1}$$

That is, $\phi(\bar{u}_0, \bar{u}, \bar{v}, x) \geq \bar{u}_0\gamma$ for all $x \in X$. But, since $\bar{x}$ solves Problem P, we have $\gamma = f(\bar{x})$. Moreover, $h(\bar{x}) = 0$ and $g(\bar{x}) \leq 0$, so that $\bar{v}^t h(\bar{x}) = 0$ and $\bar{u}^t g(\bar{x}) \leq 0$. Therefore, for any x in X

$$\phi(\bar{u}_0, \bar{u}, \bar{v}, x) \geq \bar{u}_0 f(\bar{x}) + \bar{u}^t g(\bar{x}) + \bar{v}^t h(\bar{x}) = \phi(\bar{u}_0, \bar{u}, \bar{v}, \bar{x}).$$

This establishes the second inequality. To prove the first inequality, note that for any $u \geq 0$, we have

$$\phi(\bar{u}_0, \bar{u}, \bar{v}, \bar{x}) - \phi(\bar{u}_0, u, v, \bar{x}) = (\bar{u} - u)^t g(\bar{x}) +$$
$$(\bar{v} - v)^t h(\bar{x}) = (\bar{u} - u)^t g(\bar{x}) \geq \bar{u}^t g(\bar{x}). \tag{2}$$

Now, from (1) for $x = \bar{x}$, since $f(\bar{x}) = \gamma$, we get $\bar{u}^t g(\bar{x}) + \bar{v}^t h(\bar{x}) \geq 0$, i.e., $\bar{u}^t g(\bar{x}) \geq 0$. But $g(\bar{x}) \leq 0$ since $\bar{x}$ is a feasible solution, and $\bar{u} \geq 0$, which necessarily implies that $\bar{u}^t g(\bar{x}) = 0$. Thus, (2) implies that for any $u \geq 0$ and $v \in R^\ell$, we have that $\phi(\bar{u}_0, \bar{u}, \bar{v}, \bar{x}) - \phi(\bar{u}_0, u, v, \bar{x}) \geq 0$. □

6.23 a. $\theta(u) =$

$$\begin{aligned}
&\min\{-2x_1 + 2x_2 + x_3 - 3x_4 + u_1(x_1 + x_2 + x_3 + x_4 - 8) + \\
&\qquad u_2(x_1 - 2x_3 + 4x_4 - 2) : x \in X\} \\
&= \min\{-2 + u_1 + u_2)x_1 + (2 + u_1)x_2 : \\
&\qquad x_1 + x_2 \leq 8, x_1 \geq 0, x_2 \geq 0\} \\
&+ \min\{1 + u_1 - 2u_2)x_3 + (-3 + u_1 + 4u_2)x_4 : \\
&\qquad x_3 + 2x_4 \leq 6, x_3 \geq 0, x_4 \geq 0\} - 8u_1 - 2u_2.
\end{aligned}$$

The extreme points of $\{(x_1,x_2) \geq 0 : x_1 + x_2 \leq 8\}$ are (0, 0), (8, 0), and (0, 8) in the (x_1,x_2)-space, and the extreme points of $\{(x_3,x_4) \geq 0 : x_3 + 2x_4 \leq 6\}$ are (0, 0), (6, 0), and (0, 3) in the (x_3,x_4)-space. Thus,

$$\begin{aligned}\theta(u) = -8u_1 - 2u_2 &+ \min\{0,-16 + 8u_1 + 8u_2, 16 + 8u_1\} \\ &+ \min\{0, 6 + 6u_1 - 12u_2, -9 + 3u_1 + 12u_2\}. \qquad (1)\end{aligned}$$

Noting that $(u_1,u_2) \geq 0,$ we get that

$$\min\{0,-16 + 8u_1 + 8u_2, 16 + 8u_1\} = \begin{cases} 0 \text{ if } u_1 + u_2 \geq 2 \\ -16 + 8u_1 + 8u_2 \text{ if } u_1 + u_2 \leq 2. \end{cases} \qquad (2)$$

Similarly,

$$\min\{0, 6 + 6u_1 - 12u_2, 9 + 3u_1 + 12u_2\} = \begin{cases} 0 \text{ if } -u_1 + 2u_2 \leq 1 \text{ and } u_1 + 4u_2 \geq 3 \\ 6 + 6u_1 - 12u_2 \text{ if } -u_1 + 2u_2 \geq 1 \text{ and} \\ \qquad\qquad -u_1 + 8u_2 \geq 5 \\ -9 + 3u_1 + 12u_2 \text{ if } u_1 + 4u_2 \leq 3 \text{ and} \\ \qquad\qquad -u_1 + 8u_2 \leq 5. \end{cases} \qquad (3)$$

Examining the six possible combinations given by (2) and (3), and incorporating these within (1), we get that (upon eliminating redundant conditions on (u_1,u_2)), $\theta(u) = \theta_i(u)$ if $u \in U_i$, $i = 1,\ldots,6$, where

$\theta_1(u) = -8u_1 - 2u_2$ and

$U_1 = \{(u_1,u_2) \geq 0 : -u_1 + 2u_2 \leq 1, u_1 + u_2 \geq 2, u_1 + 4u_2 \geq 3\}$

$\theta_2(u) = 6 - 2u_1 - 14u_2$ and

$U_2 = \{(u_1,u_2) \geq 0 : u_1 + u_2 \geq 2, -u_1 + 2u_2 \geq 1\}$

$\theta_3(u) = -9 - 5u_1 + 10u_2$ and

$U_3 = \{(u_1,u_2) \geq 0 : u_1 + 4u_2 \leq 3, u_1 + u_2 \geq 2\}$

$\theta_4(u) = -16 + 6u_2$ and

$U_4 = \{(u_1,u_2) \geq 0 : -u_1 + 2u_2 \leq 1, u_1 + u_2 \leq 2, u_1 + 4u_2 \geq 3\}$

$\theta_5(u) = -10 + 6u_1 - 6u_2$ and

$$U_5 = \{(u_1, u_2) \geq 0 : u_1 + u_2 \leq 2, -u_1 + 2u_2 \geq 1, -u_1 + 8u_2 \geq 5\}$$

$\theta_6(u) = -25 + 3u_1 + 18u_2$ and

$$U_6 = \{(u_1, u_2) \geq 0 : -u_1 + 8u_2 \leq 5, u_1 + 4u_2 \leq 3, u_1 + u_2 \leq 2\}$$

b. Note that $u = (4, 0)$ belongs to U_1 alone. Thus, θ is differentiable at (4, 0), with $\nabla\theta(4,0) = (-8,-2)$.

c. When $u = (4, 0)$ and $d = (-8,-2)$, the second coordinate of $u + \lambda d = (4,0) + \lambda(-8,-2)$ is -2λ, which is negative for all $\lambda > 0$. Since $u_2 = 0$, we have that the gradient of $\theta(u)$ at (4, 0) is not a feasible direction at (4, 0). However, projecting d onto $(d = d_2 \geq 0)$, we get that $d' = (-8,0)$ is a feasible direction of $\theta(u)$ at (4, 0). Moreoever, $\nabla\theta(4,0)^t d' = 64 > 0$. Thus, d' is an improving, feasible direction.

d. To maintain feasibility, we must have $4 - 8\lambda \geq 0$, i.e., λ should be restricted to values in the interval [0, 1/2]. Moreoever,

$$\begin{aligned}
&\theta(u + \lambda d') = \theta[(4,0) + \lambda(-8,0)] = \theta[(4 - 8\lambda, 0)] \\
&= \min\{(2 - 8\lambda)x_1 + (6 - 8\lambda)x_2 : x_1 + x_2 \leq 8, x_1 \geq 0, x_2 \geq 0\} \\
&+ \min\{(5 - 8\lambda)x_3 + (1 - 8\lambda)x_4 : x_3 + 2x_4 \leq 6, x_3 \geq 0, x_4 \geq 0\} \\
&-32(1 - 2\lambda) \\
&= -32(1 - 2\lambda) + \min\{0, 16(1 - 4\lambda), 16(3 - 4\lambda)\} \\
&+ \min\{0, 6(5 - 8\lambda), 3(1 - 8\lambda)\} \\
&= -32(1 - 2\lambda) + \min\{0, 16(1 - 4\lambda)\} \\
&+ \min\{0, 3(1 - 8\lambda)\} \text{ when } 0 \leq \lambda \leq 1/2.
\end{aligned}$$

Thus, we get

$$\theta(u + \lambda d') = \begin{cases} -32 + 64\lambda & \text{for } 0 \leq \lambda < 1/8 \\ -29 + 40\lambda & \text{for } 1/8 \leq \lambda < 1/4 \\ -13 - 24\lambda & \text{for } 1/4 \leq \lambda \leq 1/2. \end{cases}$$

The maximum of $\theta(u + \lambda d') = \theta[(4 - 8\lambda, 0)]$ over $\lambda \in [0,\ 1/2]$ is -19, and is attained at $\lambda = 1/4$.

6.27 For any $u \geq 0$, we have

$$\theta(u) = \min_{x \geq 0}\{x + ug(x)\}.$$

a. For this case, we have (over $x \geq 0$):

$$x + ug(x) = \begin{cases} x - \dfrac{2u}{x} & \text{for } x > 0 \\ 0 & \text{for } x > 0. \end{cases}$$

When $u = 0$, we get $\theta(u) = 0$ (achieved uniquely at $x = 0$).

When $u > 0$, we get $\theta(u) = -\infty$ (as $x \to 0^+$).

Moreover, ξ is a subgradient of θ at $u = 0$ if and only if

$\theta(u) \leq \theta(0) + u\xi,\ \forall u \geq 0$

i.e., $\theta(u) \leq u\xi,\ \forall u \geq 0$.

Noting the form of $\theta,$ we get that any $\xi \in R$ is a subgradient. (Note that at $u = 0$, we get $\theta(u)$ is evaluated by only $x = 0$, where $g(0) = 0,$ which is a subgradient, but Theorem 6.3.3 does not apply since g is not continuous at $x = 0$. Furthermore, if we consider all $u \in R,$ then any $\theta(u) = 0$ for $u < 0$, and any $\xi \leq 0$ is a subgradient of θ at $u = 0$.)

b. For this case, we have (over $x \geq 0$) :

$$x + ug(x) = \begin{cases} x - \dfrac{2u}{x} & \text{for } x > 0 \\ -u & \text{for } x > 0. \end{cases}$$

When $u = 0$, we get $\theta(u) = 0$ evaluated (uniquely) at $x = 0$.

When $u > 0$, we get $\theta(u) = -\infty$ (as $x \to 0^+$).

As above, any $\xi \in R$ is a subgradient of θ at $u = 0$. Moreover, if we consider all of $u \in R,$ then for $-8 \leq u \leq 0,$ it can be verified that $\theta(u) = -u$ (see Part (c) below), so then any $\xi \leq -1$ is a subgradient of θ at $u = 0$.

c. In this case, we get

$$x + ug(x) = \begin{cases} x + \dfrac{2u}{x} & \text{for } x > 0 \\ u & \text{for } x = 0. \end{cases}$$

When $u = 0$, we get $\theta(u) = 0$, evaluated uniquely at $x = 0$.

When $u > 0$, we get $\theta(u) = \min\{u, \min_{x>0}\{x + \frac{2u}{x}\}\}$.

The convex function $x + \dfrac{2u}{x}$ over $x > 0$ achieves a minimum at $x = \sqrt{2u}$ of value $2\sqrt{2u}$. Hence, when $u > 0$, we get $\theta(u) = \min\{u, 2\sqrt{2u}\}$, i.e.,

$$\theta(u) = \begin{cases} 0 & \text{if } u = 0 \\ u & \text{if } u \le 8 \\ 2\sqrt{2u} & \text{if } u > 8. \end{cases}$$

Moreover, any $\xi \ge 1$ is a subgradient, considering either just $u \ge 0$ or all of $u \in R$, since in this case, $\theta(u) = -\infty$ when $u < 0$.

6.29 Assume that $X \neq \varnothing$.

a. The dual problem is: maximize $\theta(v)$, where $\theta(v) = \min\{f(x) + v^t(Ax - b) : x \in X\}$.

b. The proof of concavity of $\theta(v)$ is identical to that of Theorem 6.3.1. Alternatively, since X is a nonempty compact polyhedral set, and for each fixed v, since the function $f(x) + v^t(Ax - b)$ is concave, we have by Theorem 3.4.7 that there exists an extreme point of X that evaluates $\theta(v)$. Thus, if $vert(X)$ denotes the finite set of extreme points of X, we have that $\theta(v) = \min_{\hat{x} \in vert(X)} \{f(\hat{x}) + v^t(A\hat{x} - b)\}$. Thus, the dual function $\theta(v)$ is given by the minimum of a finite number of affine functions, and so is piecewise linear and concave [see also Exercise 3.9].

c. For a given $\hat{v}$, let $X(\hat{v})$ denote the set of optimal extreme point solutions to the problem of minimizing $f(x) + \hat{v}^t(Ax - b)$ over X.

Then, by Theorem 6.3.7, $\xi(\hat{v})$ is a subgradient of $\theta(v)$ at $\hat{v}$ if and only if $\xi(\hat{v}) = Ax - b$ for some x in the convex hull of $X(\hat{v})$. Moreoever, denoting $\partial\theta(\hat{v})$ as the subdifferential (set of subgradients) of θ at $\hat{v}$, we have that d is an ascent direction for θ at $\hat{v}$ if and only if $\inf\{\xi^t d : \xi \in \partial\theta(\hat{v})\} > 0$, i.e., if and only if $\min\{d^t(A\hat{x} - b) : \hat{x} \in X(\hat{v})\} > 0$. Hence, if $Ax = b$ for some $x \in X(\hat{v})$, then the set of ascent directions of $\theta(v)$ at v is empty. Otherwise, an ascent direction exists. In this case, the steepest ascent direction, $\hat{d}$, can be found by employing Theorem 6.3.11. Namely, $\hat{d} = \hat{\xi}/\|\hat{\xi}\|$, where $\hat{\xi}$ is a subgradient of $\theta(v)$ at $\hat{v}$ with the smallest Euclidean norm. To find $\hat{\xi}$, we can solve the following problem: minimize $\|Ax - b\|$ subject to $x \in conv[X(\hat{v})]$. If $\hat{x}$ is an optimal solution for this problem, then $\hat{\xi} = A\hat{x} - b$.

d. If X is not bounded, then it is not necessarily true that for each v there exists an optimal solution for the problem to minimize $f(x) + v^t(Ax - b)$ subject to $x \in X$. For all such vectors the dual function value $\theta(v)$ is $-\infty$. However, $\theta(v)$ is still concave and piecewise linear over the set of all vectors v for which $\min\{f(x) + v^t(Ax - b) : x \in X\}$ exists.