Chapter 4

The Fritz John and the Kuhn–Tucker Optimality Conditions

In Chapter 3 we derived an optimality condition for a problem of the form: minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in S$, where f is a convex function and S is a convex set. The necessary and sufficient condition for $\bar{\mathbf{x}}$ to solve the problem was shown to be

$$\nabla f(\bar{\mathbf{x}})^t(\mathbf{x}-\bar{\mathbf{x}}) \geq 0 \qquad \text{for all } \mathbf{x} \in S$$

In this chapter the nature of the set S will be more explicitly specified. In particular we consider problems with inequality and/or equality constraints. The necessary conditions are derived without any convexity assumptions and are sharper than the above in the sense that they explicitly consider the constraint functions and are more easily verifiable, since they deal with a system of equations. Under suitable convexity assumptions, these necessary conditions are also sufficient for optimality.

The following is an outline of the chapter.

SECTION 4.1: **Unconstrained Problems** We briefly consider optimality conditions for unconstrained problems. First-order and second-order conditions are discussed.

SECTION 4.2: **Problems with Inequality Constraints** Both the Fritz John and the Kuhn–Tucker conditions for problems with inequality constraints are derived.

SECTION 4.3: **Problems with Inequality and Equality Constraints** This section extends the results of the previous section to problems with both inequality and equality constraints.

4.1 Unconstrained Problems

An unconstrained problem is a problem of the form: minimize $f(\mathbf{x})$ without any constraints on the vector $\mathbf{x}$. Unconstrained problems seldom arise in practical applications. However, we consider such problems here because optimality conditions for constrained problems become a logical extension of the conditions for unconstrained problems. Furthermore, as shown in Chapter 9, one strategy for solving a constrained problem is to solve a sequence of unconstrained problems.

We define below a local and a global minimum of an unconstrained problem. The definition is a special case of Definition 3.4.1, where the set S is replaced by E_n.

4.1.1 Definition

Consider the problem of minimizing $f(\mathbf{x})$ over E_n, and let $\bar{\mathbf{x}} \in E_n$. If $f(\bar{\mathbf{x}}) \leq f(\mathbf{x})$ for all $\mathbf{x} \in E_n$, then $\bar{\mathbf{x}}$ is called a *global minimum*. If there exists an ε-neighborhood $N_\varepsilon(\bar{\mathbf{x}})$ around $\bar{\mathbf{x}}$ such that $f(\bar{\mathbf{x}}) \leq f(\mathbf{x})$ for each $\mathbf{x} \in N_\varepsilon(\bar{\mathbf{x}})$, then $\bar{\mathbf{x}}$ is called a *local minimum*. Clearly a global minimum is also a local minimum.

Necessary Optimality Conditions

Given a point $\mathbf{x}$ in E_n, we wish to determine, if possible, whether or not the point is a local or a global minimum of a function f. For this purpose, we need to characterize the minimum point. Fortunately, the differentiability assumption of f provides a means of obtaining this characterization. The corollary to Theorem 4.1.2 below gives a first-order necessary condition for $\bar{\mathbf{x}}$ to be a local optimum. Theorem 4.1.3 gives a second-order necessary condition using the Hessian matrix.

4.1.2 Theorem

Suppose that $f: E_n \to E_1$ is differentiable at $\bar{\mathbf{x}}$. If there is a vector $\mathbf{d}$ such that $\nabla f(\bar{\mathbf{x}})^t\mathbf{d} < 0$, then there exists a $\delta > 0$ such that $f(\bar{\mathbf{x}} + \lambda\mathbf{d}) < f(\bar{\mathbf{x}})$ for each $\lambda \in (0, \delta)$, so that $\mathbf{d}$ is a *descent direction* of f at $\bar{\mathbf{x}}$.

Proof

By differentiability of f at $\bar{\mathbf{x}}$, we must have

$$f(\bar{\mathbf{x}} + \lambda\mathbf{d}) = f(\bar{\mathbf{x}}) + \lambda\nabla f(\bar{\mathbf{x}})^t\mathbf{d} + \lambda\|\mathbf{d}\|\alpha(\bar{\mathbf{x}}; \lambda\mathbf{d})$$

where $\alpha(\bar{\mathbf{x}}; \lambda\mathbf{d}) \to 0$ as $\lambda \to 0$. Rearranging the terms and dividing by λ, we get

$$\frac{f(\bar{\mathbf{x}} + \lambda\mathbf{d}) - f(\bar{\mathbf{x}})}{\lambda} = \nabla f(\bar{\mathbf{x}})^t\mathbf{d} + \|\mathbf{d}\|\alpha(\bar{\mathbf{x}}; \lambda\mathbf{d})$$

Since $\nabla f(\bar{\mathbf{x}})^t\mathbf{d} < 0$ and $\alpha(\bar{\mathbf{x}}; \lambda\mathbf{d}) \to 0$ as $\lambda \to 0$, there exists a $\delta > 0$ such that $\nabla f(\bar{\mathbf{x}})^t\mathbf{d} + \|\mathbf{d}\|\alpha(\bar{\mathbf{x}}; \lambda\mathbf{d}) < 0$ for all $\lambda \in (0, \delta)$. The result then follows.

Corollary

Suppose that $f: E_n \to E_1$ is differentiable at $\bar{\mathbf{x}}$. If $\bar{\mathbf{x}}$ is a local minimum, then $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$.

Proof

Suppose $\nabla f(\bar{\mathbf{x}}) \neq \mathbf{0}$ and let $\mathbf{d} = -\nabla f(\bar{\mathbf{x}})$. Then $\nabla f(\bar{\mathbf{x}})^t\mathbf{d} = -\|\nabla f(\bar{\mathbf{x}})\|^2 < 0$, and by Theorem 4.1.2, there is a $\delta > 0$ such that $f(\bar{\mathbf{x}} + \lambda\mathbf{d}) < f(\bar{\mathbf{x}})$ for $\lambda \in (0, \delta)$, contradicting the assumption that $\bar{\mathbf{x}}$ is a local minimum. Hence $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$.

The above condition uses the gradient vector whose components are the first partials of f. Hence, it is called a *first-order condition*. Necessary conditions can also be stated in terms of the Hessian matrix $\mathbf{H}$ whose elements are the second partials of f. These are called *second-order conditions* and are given below.

4.1.3 Theorem

Suppose that $f: E_n \to E_1$ is twice differentiable at $\bar{\mathbf{x}}$. If $\bar{\mathbf{x}}$ is a local minimum, then $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$, and $\mathbf{H}(\bar{\mathbf{x}})$ is positive semidefinite.

Proof

Consider an arbitrary direction $\mathbf{d}$. Then, from differentiability of f at $\bar{\mathbf{x}}$, we have

$$f(\bar{\mathbf{x}} + \lambda\mathbf{d}) = f(\bar{\mathbf{x}}) + \lambda\nabla f(\bar{\mathbf{x}})^t\mathbf{d} + \tfrac{1}{2}\lambda^2\mathbf{d}^t\mathbf{H}(\bar{\mathbf{x}})\mathbf{d} + \lambda^2\|\mathbf{d}\|^2\alpha(\bar{\mathbf{x}}; \lambda\mathbf{d}) \tag{4.1}$$

where $\alpha(\bar{\mathbf{x}}; \lambda\mathbf{d}) \to 0$ as $\lambda \to 0$. Since $\bar{\mathbf{x}}$ is a local minimum, from the corollary to Theorem 4.1.2, we have $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$. Rearranging the terms in (4.1) and dividing by λ^2, we get

$$\frac{f(\bar{\mathbf{x}} + \lambda\mathbf{d}) - f(\bar{\mathbf{x}})}{\lambda^2} = \tfrac{1}{2}\mathbf{d}^t\mathbf{H}(\bar{\mathbf{x}})\mathbf{d} + \|\mathbf{d}\|^2\alpha(\bar{\mathbf{x}}; \lambda\mathbf{d}) \tag{4.2}$$

Since $\bar{\mathbf{x}}$ is a local minimum, $f(\bar{\mathbf{x}} + \lambda\mathbf{d}) \geq f(\bar{\mathbf{x}})$ for λ sufficiently small. From (4.2), it is thus clear that $\frac{1}{2}\mathbf{d}^t\mathbf{H}(\bar{\mathbf{x}})\mathbf{d} + \|\mathbf{d}\|^2\alpha(\bar{\mathbf{x}}; \lambda\mathbf{d}) \geq 0$ for λ sufficiently small. By taking the limit as $\lambda \to 0$, it follows that $\mathbf{d}^t\mathbf{H}(\bar{\mathbf{x}})\mathbf{d} \geq 0$, and hence $\mathbf{H}(\bar{\mathbf{x}})$ is positive semidefinite.

Sufficient Optimality Conditions

The conditions discussed thus far are necessary conditions; that is, they must be true for every local optimal solution. On the other hand, a point satisfying these conditions need not be a local minimum. Theorem 4.1.4 gives a sufficient condition for a local minimum.

4.1.4 Theorem

Suppose that $f: E_n \to E_1$ is twice differentiable at $\bar{\mathbf{x}}$. If $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$ and $\mathbf{H}(\bar{\mathbf{x}})$ is positive definite, then $\bar{\mathbf{x}}$ is a local minimum.

Proof

Since f is twice differentiable at $\bar{\mathbf{x}}$, we must have for each $\mathbf{x} \in E_n$:

$$f(\mathbf{x}) = f(\bar{\mathbf{x}}) + \nabla f(\bar{\mathbf{x}})^t(\mathbf{x}-\bar{\mathbf{x}}) + \tfrac{1}{2}(\mathbf{x}-\bar{\mathbf{x}})^t\mathbf{H}(\bar{\mathbf{x}})(\mathbf{x}-\bar{\mathbf{x}}) + \|\mathbf{x}-\bar{\mathbf{x}}\|^2\alpha(\bar{\mathbf{x}}; \mathbf{x}-\bar{\mathbf{x}}) \tag{4.3}$$

where $\alpha(\bar{\mathbf{x}}; \mathbf{x}-\bar{\mathbf{x}}) \to 0$ as $\mathbf{x} \to \bar{\mathbf{x}}$. Suppose, by contradiction, that $\bar{\mathbf{x}}$ is not a local minimum; that is, suppose there exists a sequence $\{\mathbf{x}_k\}$ converging to $\bar{\mathbf{x}}$ such that $f(\mathbf{x}_k) < f(\bar{\mathbf{x}})$ for each k. Considering this sequence, noting that $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$ and $f(\mathbf{x}_k) < f(\bar{\mathbf{x}})$, and denoting $(\mathbf{x}_k - \bar{\mathbf{x}})/\|\mathbf{x}_k - \bar{\mathbf{x}}\|$ by $\mathbf{d}_k$, (4.3) then implies that

$$\tfrac{1}{2}\mathbf{d}_k^t\mathbf{H}(\bar{\mathbf{x}})\mathbf{d}_k + \alpha(\bar{\mathbf{x}}; \mathbf{x}_k - \bar{\mathbf{x}}) < 0 \qquad \text{for each } k \tag{4.4}$$

But $\|\mathbf{d}_k\| = 1$ for each k, and hence there exists an index set $\mathscr{K}$ such that $\{\mathbf{d}_k\}_{\mathscr{K}}$ converges to $\mathbf{d}$, where $\|\mathbf{d}\| = 1$. Considering this subsequence and the fact that $\alpha(\bar{\mathbf{x}}; \mathbf{x}_k - \bar{\mathbf{x}}) \to 0$ as $k \in \mathscr{K}$ approaches ∞, then (4.4) implies that $\mathbf{d}^t\mathbf{H}(\bar{\mathbf{x}})\mathbf{d} \le 0$. This contradicts the assumption that $\mathbf{H}(\bar{\mathbf{x}})$ is positive definite and the fact that $\|\mathbf{d}\| = 1$. Therefore, $\bar{\mathbf{x}}$ is indeed a local minimum.

In Theorem 4.1.5 below, we show that the necessary condition $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$ is also sufficient for $\bar{\mathbf{x}}$ to be a global minimum if f is pseudoconvex at $\bar{\mathbf{x}}$.

4.1.5 Theorem

Let $f: E_n \to E_1$ be pseudoconvex at $\bar{\mathbf{x}}$. Then $\bar{\mathbf{x}}$ is a global minimum if and only if $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$.

Proof

By the corollary to Theorem 4.1.2, if $\bar{\mathbf{x}}$ is a global minimum, then $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$. Now suppose that $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$, so that $\nabla f(\bar{\mathbf{x}})^t(\mathbf{x}-\bar{\mathbf{x}}) = 0$ for each $\mathbf{x} \in E_n$. By pseudoconvexity of f at $\bar{\mathbf{x}}$, it then follows that $f(\mathbf{x}) \ge f(\bar{\mathbf{x}})$ for each $\mathbf{x} \in E_n$, and the proof is complete.

4.1.6 Example

To illustrate the necessary and sufficient conditions of this section, consider the problem to minimize $f(x) = (x^2 - 1)^3$.

First let us determine the candidate points for optimality satisfying the first-order necessary condition that $\nabla f(x) = 0$. Note that $\nabla f(x) = 6x(x^2-1)^2$, and $\nabla f(-1) = \nabla f(0) = \nabla f(1) = 0$. Now let us examine the second order necessary condition that $\mathbf{H}(x)$ is positive semidefinite. We have $\mathbf{H}(x) = 24x^2(x^2-1) + 6(x^2-1)^2$, and hence $\mathbf{H}(1) = \mathbf{H}(-1) = 0$ and $\mathbf{H}(0) = 6$. In all three cases, the matrix $\mathbf{H}$ is positive semidefinite, and the necessary conditions of Theorem 4.1.3 hold true. This does not imply that each of these points is a local minimum. By sketching the function, the reader could easily verify that the point $x = 0$ is indeed the only local minimum, and also the global minimum. Note also that the points 1 and -1 do not satisfy the sufficient conditions of Theorem 4.1.4, which requires $\mathbf{H}$ to be positive definite. This condition is satisfied at the global optimum $x = 0$.

4.2 Problems with Inequality Constraints

In this section we first develop a necessary optimality condition for the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in S$. Later we let S be the feasible region of a nonlinear programming problem of the form to minimize $f(\mathbf{x})$ subject to $\mathbf{g}(\mathbf{x}) \le \mathbf{0}$ and $\mathbf{x} \in X$.

Geometric Optimality Conditions

In Theorem 4.2.2 below, we develop a necessary optimality condition for the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in S$, using the cone of feasible directions defined below.

4.2.1 Definition

Let S be a nonempty set in E_n, and let $\bar{\mathbf{x}} \in \operatorname{cl} S$. The *cone of feasible directions* of S at $\bar{\mathbf{x}}$, denoted by D, is given by

$$D = \{\mathbf{d} : \mathbf{d} \neq \mathbf{0}, \text{ and } \bar{\mathbf{x}} + \lambda\mathbf{d} \in S \quad \text{for all } \lambda \in (0, \delta) \text{ for some } \delta > 0\}$$

Each nonzero vector $\mathbf{d} \in D$ is called a *feasible direction.*

From the above definition, it is clear that a small movement from $\bar{\mathbf{x}}$ along a vector $\mathbf{d} \in D$ leads to feasible points. Furthermore, from Theorem 4.1.2, if $\nabla f(\bar{\mathbf{x}})^t\mathbf{d} < 0$, then $\mathbf{d}$ is an improving direction; that is, starting from $\bar{\mathbf{x}}$, a small movement along $\mathbf{d}$ will reduce the value of f. As shown in Theorem 4.2.2 below, if $\bar{\mathbf{x}}$ is a local minimum and if $\nabla f(\bar{\mathbf{x}})^t\mathbf{d} < 0$, then $\mathbf{d} \notin D$; that is, a necessary condition for local optimality is that every improving direction is not a feasible direction. This fact is illustrated in Figure 4.1, where the vertices of the cones F_0 and D are translated from the origin to $\bar{\mathbf{x}}$ for convenience.

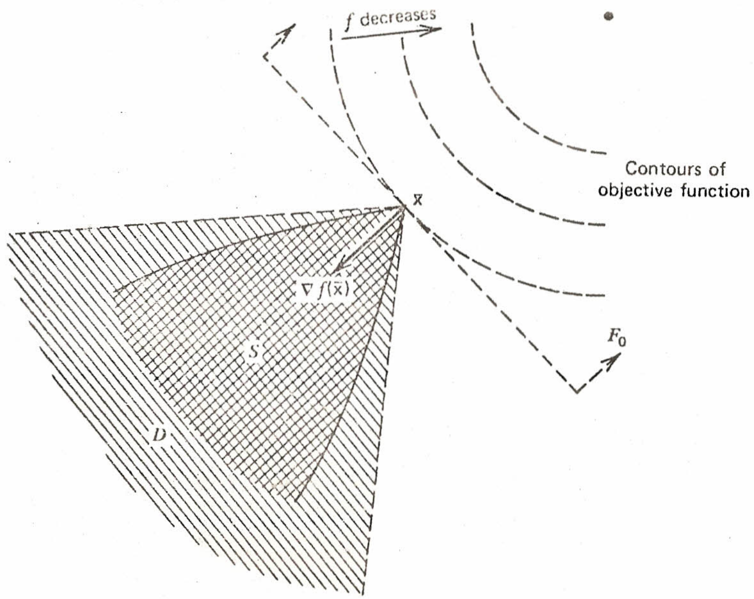


Figure 4.1 Illustration of the necessary condition $F_0 \cap D = \varnothing$.

4.2.2 Theorem

Consider the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in S$, where $f: E_n \to E_1$, and S is a nonempty set in E_n. Suppose that f is differentiable at a point $\bar{\mathbf{x}} \in S$. If $\bar{\mathbf{x}}$ is a local optimal solution, then $F_0 \cap D = \varnothing$, where $F_0 = \{\mathbf{d} : \nabla f(\bar{\mathbf{x}})^t \mathbf{d} < 0\}$ and D is the cone of feasible directions of S at $\bar{\mathbf{x}}$.

Proof

By contradiction, suppose that there exists a vector $\mathbf{d} \in F_0 \cap D$. Then by Theorem 4.1.2, there exists a $\delta_1 > 0$ such that

$$f(\bar{\mathbf{x}} + \lambda \mathbf{d}) < f(\bar{\mathbf{x}}) \qquad \text{for each } \lambda \in (0, \delta_1) \tag{4.5}$$

Furthermore, by Definition 4.2.1, there exists a $\delta_2 > 0$ such that:

$$\bar{\mathbf{x}} + \lambda \mathbf{d} \in S \qquad \text{for each } \lambda \in (0, \delta_2) \tag{4.6}$$

The assumption that $\bar{\mathbf{x}}$ is a local optimal solution to the problem is not compatible with (4.5) and (4.6). Thus $F_0 \cap D = \varnothing$, and the proof is complete.

We specify the feasible region S as follows:

$$S = \{\mathbf{x} \in X : g_i(\mathbf{x}) \le 0 \qquad \text{for } i = 1, \ldots, m\}$$

where $g_i : E_n \to E_1$ for $i = 1, \ldots, m$, and X is a nonempty open set in E_n. This

gives us the following nonlinear programming problem with inequality constraints:

Problem P:

$$\begin{array}{lll} \text{Minimize} & f(\mathbf{x}) & \\ \text{subject to} & g_i(\mathbf{x}) \le 0 & \text{for } i = 1, \ldots, m \\ & \mathbf{x} \in X & \end{array}$$

Recall that a necessary condition for local optimality at $\bar{\mathbf{x}}$ is that $F_0 \cap D = \varnothing$, where F_0 is an open half space defined in terms of the gradient vector $\nabla f(\bar{\mathbf{x}})$, and D is the cone of feasible directions, which is not necessarily defined in terms of the gradients of the functions involved. This precludes us from converting the geometric optimality condition $F_0 \cap D = \varnothing$ into a more usable algebraic statement involving equations. As Theorem 4.2.3 below indicates, we will be able to define an open cone G_0 defined in terms of the gradients of the binding constraints at $\bar{\mathbf{x}}$, such that $G_0 \subset D$. Since $F_0 \cap D = \varnothing$ must hold at $\bar{\mathbf{x}}$, and since $G_0 \subset D$, then $F_0 \cap G_0 = \varnothing$ is also a necessary optimality condition. Since F_0 and G_0 are both defined in terms of the gradient vectors, we will use the condition $F_0 \cap G_0 = \varnothing$ later in the section to develop the optimality conditions credited to Fritz John. With mild additional assumptions, the conditions reduce to the well-known Kuhn–Tucker optimality conditions.

4.2.3 Theorem

Let $g_i : E_n \to E_1$ for $i = 1, \ldots, m$, and let X be a nonempty open set in E_n. Consider the *Problem P* to minimize $f(\mathbf{x})$ subject to $g_i(\mathbf{x}) \le 0$ for $i = 1, \ldots, m$, and $\mathbf{x} \in X$. Let $\bar{\mathbf{x}}$ be a feasible point, and let $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$. Furthermore, suppose that f and g_i for $i \in I$ are differentiable at $\bar{\mathbf{x}}$ and that g_i for $i \notin I$ is continuous at $\bar{\mathbf{x}}$. If $\bar{\mathbf{x}}$ is a local optimal solution, then $F_0 \cap G_0 = \varnothing$, where

$$F_0 = \{\mathbf{d} : \nabla f(\bar{\mathbf{x}})^t \mathbf{d} < 0\}$$

$$G_0 = \{\mathbf{d} : \nabla g_i(\bar{\mathbf{x}})^t \mathbf{d} < 0 \qquad \text{for each } i \in I\}$$

Proof

Let $\mathbf{d} \in G_0$. Since $\bar{\mathbf{x}} \in X$, and X is open, there exists a $\delta_1 > 0$ such that

$$\bar{\mathbf{x}} + \lambda \mathbf{d} \in X \qquad \text{for } \lambda \in (0, \delta_1) \tag{4.7}$$

Also, since $g_i(\bar{\mathbf{x}}) < 0$ and since g_i is continuous at $\bar{\mathbf{x}}$ for $i \notin I$, there exists a $\delta_2 > 0$ such that

$$g_i(\bar{\mathbf{x}} + \lambda \mathbf{d}) < 0 \qquad \text{for } \lambda \in (0, \delta_2) \text{ and for } i \notin I \tag{4.8}$$

Finally, since $\mathbf{d} \in G_0$, $\nabla g_i(\bar{\mathbf{x}})^t \mathbf{d} < 0$ for each $i \in I$, and by Theorem 4.1.2, there exists a $\delta_3 > 0$ such that

$$g_i(\bar{\mathbf{x}} + \lambda \mathbf{d}) < g_i(\bar{\mathbf{x}}) = 0 \qquad \text{for } \lambda \in (0, \delta_3) \text{ and for } i \in I \tag{4.9}$$

From (4.7), (4.8), and (4.9), it is clear that points of the form $\bar{\mathbf{x}} + \lambda \mathbf{d}$ are feasible to *Problem P* for each $\lambda \in (0, \delta)$, where $\delta = \text{minimum } (\delta_1, \delta_2, \delta_3)$. Thus $\mathbf{d} \in D$, where D is the cone of feasible directions of the feasible region at $\bar{\mathbf{x}}$. We have shown thus far that $\mathbf{d} \in G_0$ implies that $\mathbf{d} \in D$, and hence $G_0 \subset D$. By Theorem 4.2.2, since $\bar{\mathbf{x}}$ locally solves *Problem P*, $F_0 \cap D = \emptyset$. Since $G_0 \subset D$, it follows that $F_0 \cap G_0 = \emptyset$, and the proof is complete.

4.2.4 Example

$$\begin{array}{ll} \text{Minimize} & (x_1 - 3)^2 + (x_2 - 2)^2 \\ \text{subject to} & x_1^2 + x_2^2 \le 5 \\ & x_1 + x_2 \le 3 \\ & x_1 \ge 0 \\ & x_2 \ge 0 \end{array}$$

In this case, we let $g_1(\mathbf{x}) = x_1^2 + x_2^2 - 5$, $g_2(\mathbf{x}) = x_1 + x_2 - 3$, $g_3(\mathbf{x}) = -x_1$, $g_4(\mathbf{x}) = -x_2$, and $X = E_2$. Consider the point $\bar{\mathbf{x}} = (\frac{9}{5}, \frac{6}{5})^t$, and note that the only binding constraint is $g_2(\mathbf{x}) = x_1 + x_2 - 3$. Also note that

$$\nabla f(\bar{\mathbf{x}}) = \left(\frac{-12}{5}, \frac{-8}{5}\right)^t \qquad \text{and} \qquad \nabla g_2(\bar{\mathbf{x}}) = (1, 1)^t$$

The sets F_0 and G_0, with the origin translated to $(\frac{9}{5}, \frac{6}{5})^t$ for convenience, are shown in Figure 4.2. Since $F_0 \cap G_0 \ne \emptyset$, $\bar{\mathbf{x}} = (\frac{9}{5}, \frac{6}{5})^t$ is not a local optimal solution to the above problem.

Now consider the point $\bar{\mathbf{x}} = (2, 1)^t$, and note that the first two constraints are binding. The corresponding gradients at this point are

$$\nabla f(\bar{\mathbf{x}}) = (-2, -2)^t, \qquad \nabla g_1(\bar{\mathbf{x}}) = (4, 2)^t, \qquad \text{and} \qquad \nabla g_2(\bar{\mathbf{x}}) = (1, 1)^t$$

The sets F_0 and G_0 are shown in Figure 4.3, and indeed $F_0 \cap G_0 = \emptyset$. Note that Theorem 4.2.3 gives a necessary condition and hence $F_0 \cap G_0 = \emptyset$ does not guarantee that $\bar{\mathbf{x}} = (2, 1)^t$ is an optimal point. We can only conclude that $\bar{\mathbf{x}}$ is one of the candidate points that can solve the problem under consideration.

It might be interesting to note that the utility of Theorem 4.2.3 also depends on how the constraint set is expressed. This is illustrated by Example 4.2.5 below.

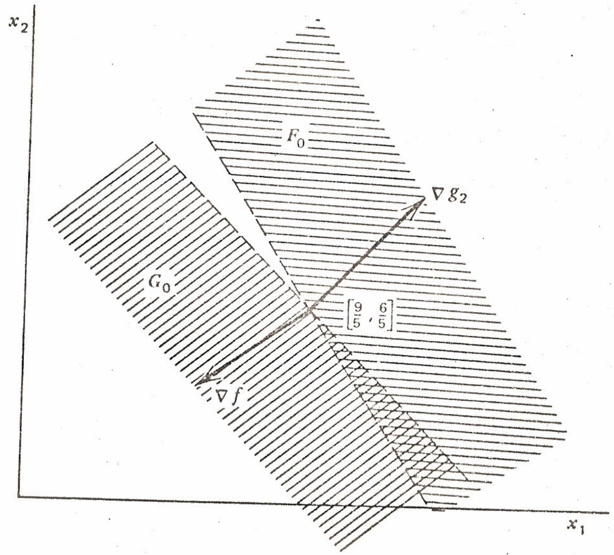


Figure 4.2 Illustration of $F_0 \cap G_0 \ne \emptyset$ at a non-optimal point.

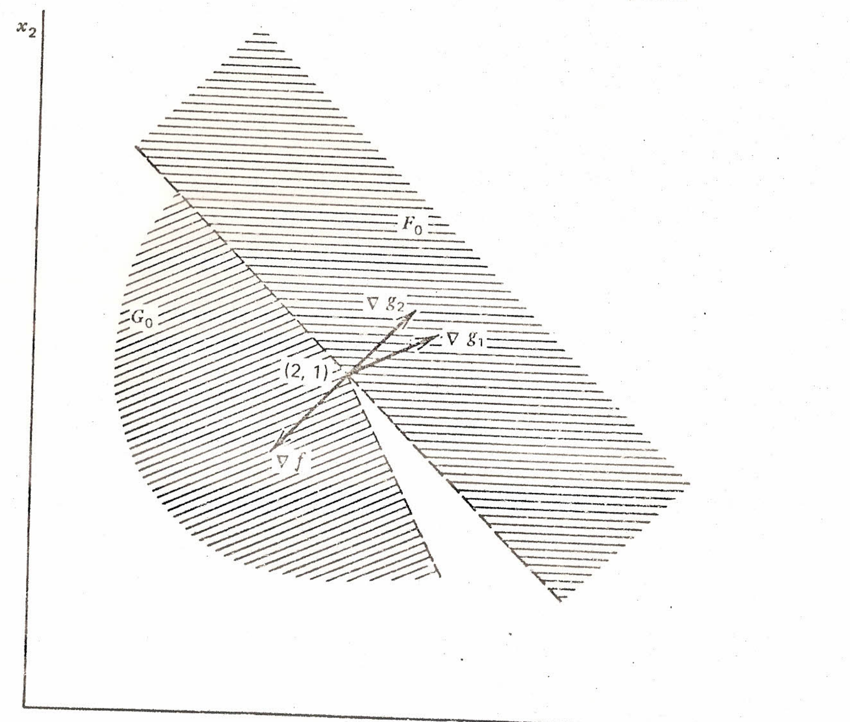


Figure 4.3 Illustration of $F_0 \cap G_0 = \emptyset$ at an optimal point.

4.2.5 Example

$$\begin{aligned} \text{Minimize} \quad & (x_1-1)^2+(x_2-1)^2 \\ \text{subject to} \quad & (x_1+x_2-1)^3 \le 0 \\ & x_1 \ge 0 \\ & x_2 \ge 0 \end{aligned}$$

Note that the necessary condition of Theorem 4.2.3 holds true at each feasible point with $x_1+x_2=1$. However, the constraint set can be represented equivalently by

$$\begin{aligned} x_1+x_2 &\le 1 \\ x_1 &\ge 0 \\ x_2 &\ge 0 \end{aligned}$$

It can be easily verified that $F_0 \cap G_0 = \emptyset$ is satisfied only at the point $(\frac{1}{2}, \frac{1}{2})$.

There are several cases where the necessary conditions of Theorem 4.2.3 are satisfied trivially by possibly nonoptimal points also. Some of these cases are discussed below.

Suppose that $\bar{\mathbf{x}}$ is a feasible point such that $\nabla f(\bar{\mathbf{x}})=\mathbf{0}$. Clearly $F_0=\{\mathbf{d}:\nabla f(\bar{\mathbf{x}})'\mathbf{d}<0\}=\emptyset$, and hence $F_0\cap G_0=\emptyset$. Thus, any point $\bar{\mathbf{x}}$ with $\nabla f(\bar{\mathbf{x}})=\mathbf{0}$ satisfies the necessary optimality conditions. Likewise, any point $\bar{\mathbf{x}}$ with $\nabla g_i(\bar{\mathbf{x}})=\mathbf{0}$ for some $i\in I$ will also satisfy the necessary conditions. Now consider the following example with an equality constraint:

$$\begin{aligned} \text{Minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g(\mathbf{x})=0 \end{aligned}$$

The equality constraint $g(\mathbf{x})=0$ could be replaced by the inequality constraints $g_1(\mathbf{x})=g(\mathbf{x})\le 0$ and $g_2(\mathbf{x})=-g(\mathbf{x})\le 0$. Let $\bar{\mathbf{x}}$ be any feasible point. Then $g_1(\bar{\mathbf{x}})=g_2(\bar{\mathbf{x}})=0$. Note that $\nabla g_1(\bar{\mathbf{x}})=-\nabla g_2(\bar{\mathbf{x}})$, and therefore there could exist no vector $\mathbf{d}$ such that $\nabla g_1(\bar{\mathbf{x}})'\mathbf{d}<0$ and $\nabla g_2(\bar{\mathbf{x}})'\mathbf{d}<0$. Therefore, $G_0=\emptyset$, and hence $F_0\cap G_0=\emptyset$. In other words, the necessary condition of Theorem 4.2.3 is satisfied by all feasible solutions and is hence not usable.

The Fritz John Optimality Conditions

We now reduce the geometric necessary optimality condition $F_0\cap G_0=\emptyset$ to a statement in terms of the gradients of the objective function and of the binding constraints. The resulting optimality conditions, credited to Fritz John [1948], are given below.

4.2.6 Theorem (The Fritz John Conditions)

Let X be a nonempty open set in E_n, and let $f:E_n\to E_1$, and $g_i:E_n\to E_1$ for $i=1,\ldots,m$. Consider *Problem P* to minimize $f(\mathbf{x})$ subject to $\mathbf{x}\in X$ and $g_i(\mathbf{x})\le 0$ for $i=1,\ldots,m$. Let $\bar{\mathbf{x}}$ be a feasible solution, and let $I=\{i:g_i(\bar{\mathbf{x}})=0\}$. Furthermore, suppose that f and g_i for $i\in I$ are differentiable at $\bar{\mathbf{x}}$ and that g_i for $i\notin I$ are continuous at $\bar{\mathbf{x}}$. If $\bar{\mathbf{x}}$ locally solves *Problem P*, then there exists scalars u_0 and u_i for $i\in I$, such that

$$\begin{aligned} u_0\nabla f(\bar{\mathbf{x}})+\sum_{i\in I}u_i\nabla g_i(\bar{\mathbf{x}}) &= \mathbf{0} \\ u_0,\ u_i &\ge 0 \qquad \text{for } i\in I \\ (u_0, \mathbf{u}_I) &\ne (0, \mathbf{0}) \end{aligned}$$

where $\mathbf{u}_I$ is the vector whose components are u_i for $i\in I$. Furthermore, if g_i for $i\notin I$ are also differentiable at $\bar{\mathbf{x}}$, then the Fritz John conditions can be written in the following equivalent form:

$$\begin{aligned} u_0\nabla f(\bar{\mathbf{x}})+\sum_{i=1}^{m}u_i\nabla g_i(\bar{\mathbf{x}}) &= \mathbf{0} \\ u_i g_i(\bar{\mathbf{x}}) &= 0 \qquad \text{for } i=1,\ldots,m \\ u_0,\ u_i &\ge 0 \qquad \text{for } i=1,\ldots,m \\ (u_0, \mathbf{u}) &\ne (0, \mathbf{0}) \end{aligned}$$

where $\mathbf{u}$ is the vector whose components are u_i for $i=1,\ldots,m$.

Proof

Since $\bar{\mathbf{x}}$ locally solves *Problem P*, then by Theorem 4.2.3, there exists no vector $\mathbf{d}$ such that $\nabla f(\bar{\mathbf{x}})'\mathbf{d}<0$ and $\nabla g_i(\bar{\mathbf{x}})'\mathbf{d}<0$ for each $i\in I$. Now, let $\mathbf{A}$ be the matrix whose rows are $\nabla f(\bar{\mathbf{x}})'$ and $\nabla g_i(\bar{\mathbf{x}})'$ for $i\in I$. The optimality condition of Theroem 4.2.3 is then equivalent to the statement that the system $\mathbf{Ad}<\mathbf{0}$ is inconsistent. By Theorem 2.3.9 there exists a nonzero vector $\mathbf{p}\ge\mathbf{0}$ such that $\mathbf{A}'\mathbf{p}=\mathbf{0}$. Denoting the components of $\mathbf{p}$ by u_0 and u_i for $i\in I$, the first part of the result follows. The equivalent form of the necessary conditions is readily obtained by letting $u_i=0$ for $i\notin I$, and the proof is complete.

In the Fritz John conditions, the scalars u_0 and u_i for $i=1,\ldots,m$ are usually called *Lagrangian multipliers.* The condition $u_i g_i(\bar{\mathbf{x}})=0$ for $i=1,\ldots,m$ is called the *complementary slackness* condition. It requires that $u_i=0$ if the corresponding inequality is nonbinding; that is, if $g_i(\bar{\mathbf{x}})<0$. Likewise, it permits $u_i>0$ only for those constraints that are binding. The Fritz John conditions can also be written in vector notation as follows:

$$u_0\nabla f(\bar{\mathbf{x}})+\nabla \mathbf{g}(\bar{\mathbf{x}})\mathbf{u}=\mathbf{0}$$
$$\mathbf{u}'\mathbf{g}(\bar{\mathbf{x}})=0$$
$$(u_0,\mathbf{u})\ge(0,\mathbf{0})$$
$$(u_0,\mathbf{u})\ne(0,\mathbf{0})$$

Here $\nabla\mathbf{g}(\bar{\mathbf{x}})$ is an $n\times m$ matrix whose ith column is $\nabla g_i(\bar{\mathbf{x}})$, and $\mathbf{u}$ is an m vector denoting the Lagrangian multipliers.

4.2.7 Example

$$\begin{array}{ll} \text{Minimize} & (x_1-3)^2+(x_2-2)^2 \\ \text{subject to} & x_1^2+x_2^2\le 5 \\ & x_1+2x_2\le 4 \\ & -x_1\le 0 \\ & -x_2\le 0 \end{array}$$

The feasible region for the above problem is illustrated in Figure 4.4. We now verify that the Fritz John conditions are true at the optimal point (2, 1). First note that the set of binding constraints I at $\bar{\mathbf{x}}=(2,1)'$ is given by $I=\{1,2\}$. Thus the Lagrangian multipliers u_3 and u_4 associated with $-x_1\le 0$ and $-x_2\le 0$, respectively, are equal to zero. Note that

$$\nabla f(\bar{\mathbf{x}})=(-2,-2)' \qquad \nabla g_1(\bar{\mathbf{x}})=(4,2)' \qquad \nabla g_2(\bar{\mathbf{x}})=(1,2)'$$

Thus, $u_0=3$, $u_1=1$, and $u_2=2$ will satisfy the Fritz John conditions, since we now have a nonzero vector $(u_0,u_1,u_2)\ge\mathbf{0}$ satisfying

$$u_0\begin{pmatrix}-2\\-2\end{pmatrix}+u_1\begin{pmatrix}4\\2\end{pmatrix}+u_2\begin{pmatrix}1\\2\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}$$

As another illustration, let us check whether the Fritz John conditions are true at the point $\hat{\mathbf{x}}=(0,0)'$. Here, the set of binding constraints is $I=\{3,4\}$, and thus $u_1=u_2=0$. Note that

$$\nabla f(\hat{\mathbf{x}})=(-6,-4)' \qquad \nabla g_3(\hat{\mathbf{x}})=(-1,0)' \qquad \nabla g_4(\hat{\mathbf{x}})=(0,-1)'$$

Also note that

$$u_0\begin{pmatrix}-6\\-4\end{pmatrix}+u_3\begin{pmatrix}-1\\0\end{pmatrix}+u_4\begin{pmatrix}0\\-1\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}$$

holds true if and only if $u_3=-6u_0$ and $u_4=-4u_0$. If $u_0>0$, then $u_3,u_4<0$, contradicting the nonnegativity restrictions. If, on the other hand, $u_0=0$, then

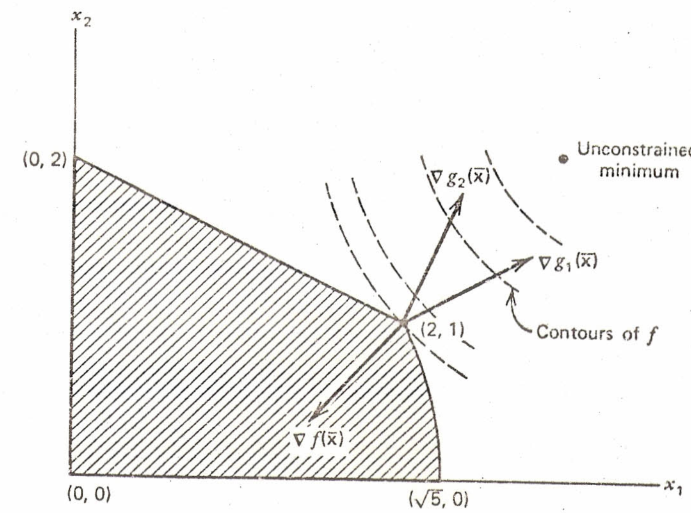


Figure 4.4 Illustration of Example 4.2.7.

$u_3=u_4=0$, which contradicts the stipulation that the vector (u_0,u_3,u_4) is nonzero. Thus the Fritz John conditions do not hold true at $\hat{\mathbf{x}}=(0,0)'$, which also shows that the origin is not a local optimal point.

4.2.8 Example

Consider the following problem from Kuhn and Tucker [1951].

$$\begin{array}{ll} \text{Minimize} & -x_1 \\ \text{subject to} & x_2-(1-x_1)^3\le 0 \\ & -x_2\le 0 \end{array}$$

The feasible region is illustrated in Figure 4.5. We now verify that the Fritz John conditions indeed hold true at the optimal point $\bar{\mathbf{x}}=(1,0)'$. Note that the set of binding constraints at $\bar{\mathbf{x}}$ is given by $I=\{1,2\}$. Also,

$$\nabla f(\bar{\mathbf{x}})=(-1,0)' \qquad \nabla g_1(\bar{\mathbf{x}})=(0,1)' \qquad \nabla g_2(\bar{\mathbf{x}})=(0,-1)'$$

In particular,

$$u_0\begin{pmatrix}-1\\0\end{pmatrix}+u_1\begin{pmatrix}0\\1\end{pmatrix}+u_2\begin{pmatrix}0\\-1\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}$$

is true only if $u_0=0$. Thus, the Fritz John conditions are true at $\bar{\mathbf{x}}$ by letting $u_0=0$ and $u_1=u_2=\alpha$, where α is a positive scalar.

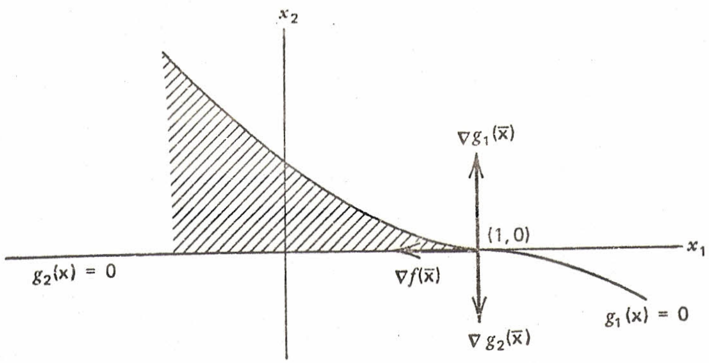


Figure 4.5 Illustration of Example 4.2.8.

4.2.9 Example

$$
\begin{array}{ll}
\text{Minimize} & -x_1 \\
\text{subject to} & x_1 + x_2 - 1 \le 0 \\
 & -x_2 \quad\quad\ \le 0
\end{array}
$$

The feasible region is sketched in Figure 4.6, and the optimal point is $\bar{\mathbf{x}} = (1, 0)^t$. Note that

$$\nabla f(\bar{\mathbf{x}}) = (-1, 0)^t \qquad \nabla g_1(\bar{\mathbf{x}}) = (1, 1)^t \qquad \nabla g_2(\bar{\mathbf{x}}) = (0, -1)^t$$

and the Fritz John conditions are true with $u_0 = u_1 = u_2 = \alpha$ for any positive scalar α.

As in the case of Theorem 4.2.3, there are points that satisfy the Fritz John conditions trivially. If a point $\bar{\mathbf{x}}$ satisfies $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$ or $\nabla g_i(\bar{\mathbf{x}}) = \mathbf{0}$ for some $i \in I$, then clearly we can let the corresponding Lagrangian multiplier be any positive number, set all the other multipliers equal to zero, and satisfy the conditions of Theorem 4.2.6. The Fritz John conditions of Theorem 4.2.6 also hold true trivially at each feasible point for problems with equality constraints if each equality constraint is replaced by two inequalities. Specifically, if $g(\mathbf{x}) = 0$ is replaced by $g(\mathbf{x}) \le 0$ and $-g(\mathbf{x}) \le 0$, then the Fritz John conditions are true with $u_1 = u_2 = \alpha$ and setting all the other multipliers equal to zero, where α is a positive scalar.

The Kuhn–Tucker Conditions

In Examples 4.2.7 and 4.2.9, we observed that the Lagrangian multiplier u_0 was positive at the optimal point $\bar{\mathbf{x}}$, while in Example 4.2.8, u_0 was equal to zero. Also note that Examples 4.2.8 and 4.2.9 differed in that in Example 4.2.8 the

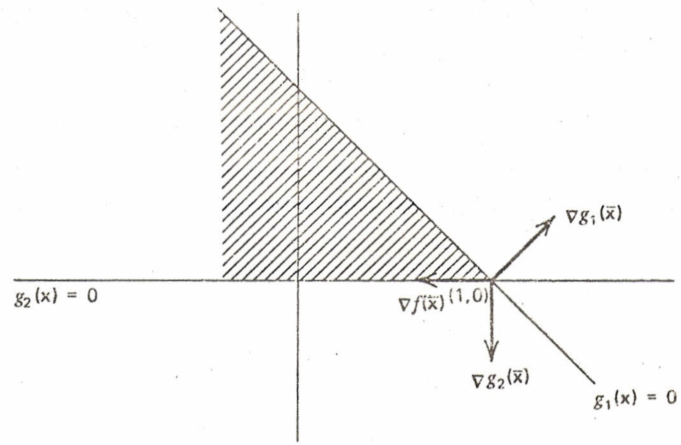


Figure 4.6 Illustration of Example 4.2.9.

gradients of the binding constraints were linearly dependent, whereas in Example 4.2.9, they were not.

If the Lagrangian multiplier u_0 is equal to zero, the Fritz John conditions do not make use of any information pertaining to the gradient of the objective function. They merely state that there exists a nonnegative and nontrivial linear combination of the gradients of the binding constraints that adds up to zero. Thus, when $u_0 = 0$, the Fritz John conditions are of no practical value in locating an optimal point. Hence, we are more interested in the cases where $u_0 > 0$. Kuhn and Tucker [1951] independently developed necessary optimality conditions that are precisely the Fritz John conditions with the added property that $u_0 > 0$. Various conditions could be imposed on the constraints in order to guarantee that $u_0 > 0$. These conditions are usually called *constraint qualifications* and are discussed in more detail in Chapter 5.

In Theorem 4.2.10 below, by imposing the constraint qualification that the gradient vectors of the binding constraints are linearly independent, we obtain the Kuhn–Tucker conditions.

4.2.10 Theorem (Kuhn–Tucker Necessary Conditions)

Let X be a nonempty open set in E_n, and let $f: E_n \to E_1$ and $g_i: E_n \to E_1$ for $i = 1, \ldots, m$. Consider *Problem P* to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in X$ and $g_i(\mathbf{x}) \le 0$ for $i = 1, \ldots, m$. Let $\bar{\mathbf{x}}$ be a feasible solution, and let $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$. Suppose that f and g_i for $i \in I$ are differentiable at $\bar{\mathbf{x}}$ and that g_i for $i \notin I$ are continuous at $\bar{\mathbf{x}}$. Furthermore, suppose that $\nabla g_i(\bar{\mathbf{x}})$ for $i \in I$ are linearly independent. If $\bar{\mathbf{x}}$ locally solves *Problem P*, then there exist scalars u_i for $i \in I$ such that

$$\nabla f(\bar{\mathbf{x}}) + \sum_{i \in I} u_i \nabla g_i(\bar{\mathbf{x}}) = \mathbf{0}$$

$$u_i \geq 0 \qquad \text{for } i \in I$$

In addition to the above assumptions, if g_i for $i \notin I$ is also differentiable at $\bar{\mathbf{x}}$, then the Kuhn–Tucker conditions could be written in the following equivalent form:

$$\nabla f(\bar{\mathbf{x}}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{\mathbf{x}}) = \mathbf{0}$$

$$u_i g_i(\bar{\mathbf{x}}) = 0 \qquad \text{for } i = 1, \ldots, m$$

$$u_i \geq 0 \qquad \text{for } i = 1, \ldots, m$$

Proof

By Theorem 4.2.6, there exist scalars u_0 and $\hat{u}_i$ for $i \in I$, not all equal to zero, such that

$$u_0 \nabla f(\bar{\mathbf{x}}) + \sum_{i \in I} \hat{u}_i \nabla g_i(\bar{\mathbf{x}}) = \mathbf{0} \tag{4.10}$$

$$u_0, \hat{u}_i \geq 0 \qquad \text{for } i \in I$$

Note that $u_0 > 0$, because (4.10) would contradict the assumption of linear independence of $\nabla g_i(\bar{\mathbf{x}})$ for $i \in I$ if $u_0 = 0$. The first part of the theorem then follows by letting $u_i = \hat{u}_i / u_0$. The equivalent form of the necessary conditions follows by letting $u_i = 0$ for $i \notin I$. This completes the proof.

As in the Fritz John conditions, the scalars u_i are called the *Lagrangian multipliers*, and the requirement that $u_i g_i(\bar{\mathbf{x}}) = 0$ for $i = 1, \ldots, m$ is referred to as the *complementary slackness* condition. Note that the Kuhn–Tucker conditions can be written in vector form as

$$\nabla f(\bar{\mathbf{x}}) + \nabla \mathbf{g}(\bar{\mathbf{x}}) \mathbf{u} = \mathbf{0}$$

$$\mathbf{u}' \mathbf{g}(\bar{\mathbf{x}}) = 0$$

$$\mathbf{u} \geq \mathbf{0}$$

Here, $\nabla \mathbf{g}(\bar{\mathbf{x}})$ is an $n \times m$ matrix whose ith column is $\nabla g_i(\bar{\mathbf{x}})$, and $\mathbf{u}$ is an m vector denoting the Lagrangian multipliers.

Now consider Examples 4.2.7, 4.2.8, and 4.2.9 discussed earlier. In Example 4.2.7, at $\bar{\mathbf{x}} = (2, 1)'$, the reader may verify that $u_1 = \frac{1}{3}$, $u_2 = \frac{2}{3}$, and $u_3 = u_4 = 0$ will satisfy the Kuhn–Tucker conditions. Example 4.2.8 does not satisfy the assumptions of Theorem 4.2.10 at $\bar{\mathbf{x}} = (1, 0)'$, since $\nabla g_1(\bar{\mathbf{x}})$ and $\nabla g_2(\bar{\mathbf{x}})$ are linearly dependent. In fact, in this example, we saw that $u_0 = 0$. In Example 4.2.9, $u_1 = u_2 = 1$ will satisfy the Kuhn–Tucker conditions.

Geometric Interpretation of the Kuhn–Tucker Conditions

Note that any vector of the form $\sum_{i \in I} u_i \nabla g_i(\bar{\mathbf{x}})$, where $u_i \geq 0$ for $i \in I$, belongs to the cone spanned by the gradients of the binding constraints. The Kuhn–Tucker conditions $-\nabla f(\bar{\mathbf{x}}) = \sum_{i \in I} u_i \nabla g_i(\bar{\mathbf{x}})$ and $u_i \geq 0$ for $i \in I$ can then be interpreted as $-\nabla f(\bar{\mathbf{x}})$ belonging to the above mentioned cone.

Figure 4.7 illustrates two points $\mathbf{x}_1$ and $\mathbf{x}_2$. Note that $-\nabla f(\mathbf{x}_1)$ belongs to the cone spanned by the gradients of the binding constraints at $\mathbf{x}_1$, and hence $\mathbf{x}_1$ is a Kuhn–Tucker point; that is, $\mathbf{x}_1$ satisfies the Kuhn–Tucker conditions. On the other hand, $-\nabla f(\mathbf{x}_2)$ lies outside the cone spanned by the gradients of the binding constraints at $\mathbf{x}_2$, and thus contradicts the Kuhn–Tucker conditions.

Likewise, in Figures 4.4 and 4.6, for $\bar{\mathbf{x}} = (2, 1)'$ and $\bar{\mathbf{x}} = (1, 0)'$, respectively, $-\nabla f(\bar{\mathbf{x}})$ is in the cone spanned by the gradients of the binding constraints at $\bar{\mathbf{x}}$. On the other hand, in Figure 4.5, for $\bar{\mathbf{x}} = (1, 0)'$, $-\nabla f(\bar{\mathbf{x}})$ lies outside the cone spanned by the gradients of the binding constraints at $\bar{\mathbf{x}}$.

Theorem 4.2.11 below shows that, under moderate convexity assumptions, the Kuhn–Tucker conditions are also sufficient for optimality.

4.2.11 Theorem (Kuhn–Tucker Sufficient Conditions)

Let X be a nonempty open set in E_n, and let $f: E_n \to E_1$ and $g_i: E_n \to E_1$ for $i = 1, \ldots, m$. Consider *Problem P* to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in X$ and $g_i(\mathbf{x}) \leq 0$ for $i = 1, \ldots, m$. Let $\bar{\mathbf{x}}$ be a feasible solution, and let $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$. Suppose that f is pseudoconvex at $\bar{\mathbf{x}}$ and that g_i is quasiconvex and differentiable at $\bar{\mathbf{x}}$ for each $i \in I$. Furthermore, suppose that the Kuhn–Tucker conditions

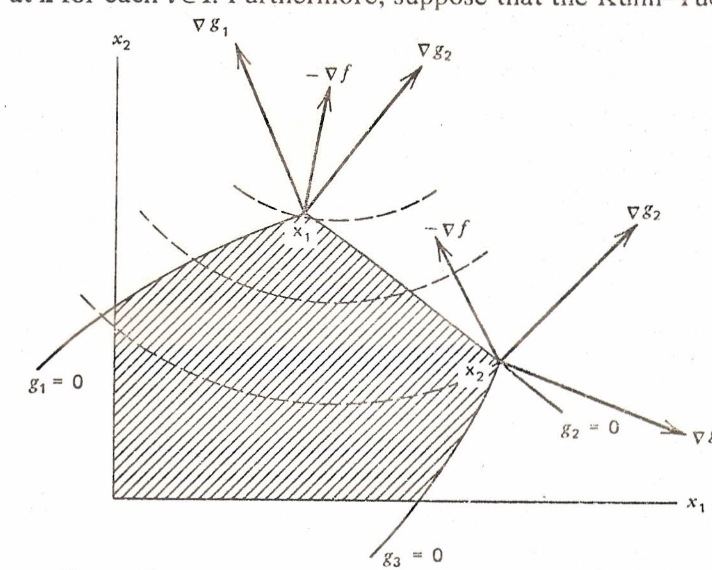


Figure 4.7 Geometric illustration of the Kuhn-Tucker conditions.

hold true at $\bar{\mathbf{x}}$; that is, there exists nonnegative scalars u_i for $i \in I$ such that $\nabla f(\bar{\mathbf{x}}) + \sum_{i \in I} u_i \nabla g_i(\bar{\mathbf{x}}) = \mathbf{0}$. Then $\bar{\mathbf{x}}$ is a global optimal solution to *Problem P*.

Proof

Let $\mathbf{x}$ be a feasible solution to *Problem P*. Then for $i \in I$, $g_i(\mathbf{x}) \le g_i(\bar{\mathbf{x}})$, since $g_i(\mathbf{x}) \le 0$ and $g_i(\bar{\mathbf{x}}) = 0$. By quasiconvexity of g_i at $\bar{\mathbf{x}}$, it follows that

$$g_i[\bar{\mathbf{x}} + \lambda(\mathbf{x} - \bar{\mathbf{x}})] = g_i[\lambda \mathbf{x} + (1-\lambda)\bar{\mathbf{x}}] \le \text{maximum}\,[g_i(\mathbf{x}), g_i(\bar{\mathbf{x}})] = g_i(\bar{\mathbf{x}})$$

for all $\lambda \in (0, 1)$. This implies that g_i does not increase by moving from $\bar{\mathbf{x}}$ along the direction $\mathbf{x} - \bar{\mathbf{x}}$. Thus, by Theorem 4.1.2, we must have $\nabla g_i(\bar{\mathbf{x}})^t(\mathbf{x} - \bar{\mathbf{x}}) \le 0$. Multiplying by u_i and summing over I, we get $[\sum_{i \in I} u_i \nabla g_i(\bar{\mathbf{x}})^t](\mathbf{x} - \bar{\mathbf{x}}) \le 0$. But since $\nabla f(\bar{\mathbf{x}}) + \sum_{i \in I} u_i \nabla g_i(\bar{\mathbf{x}}) = \mathbf{0}$, it follows that $\nabla f(\bar{\mathbf{x}})^t(\mathbf{x} - \bar{\mathbf{x}}) \ge 0$. Then, by pseudoconvexity of f at $\bar{\mathbf{x}}$, we must have $f(\mathbf{x}) \ge f(\bar{\mathbf{x}})$, and the proof is complete.

Needless to say, if f and g_i are convex at $\bar{\mathbf{x}}$ and hence both pseudoconvex and quasiconvex at $\bar{\mathbf{x}}$, then the Kuhn–Tucker conditions are sufficient. Also, if convexity at a point is replaced by the stronger requirement of global convexity, the Kuhn–Tucker conditions are also sufficient.

4.3 Problems with Inequality and Equality Constraints

In this section we generalize the optimality conditions of the previous section to handle equality constraints as well as inequality constraints. Consider the following nonlinear programming *Problem P*.

$$\begin{array}{lll} \text{Minimize} & f(\mathbf{x}) & \\ \text{subject to} & g_i(\mathbf{x}) \le 0 & \text{for } i = 1, \ldots, m \\ & h_i(\mathbf{x}) = 0 & \text{for } i = 1, \ldots, l \\ & \mathbf{x} \in X & \end{array}$$

As a natural extension of Theorem 4.2.3, in Theorem 4.3.1 below, we show that if $\bar{\mathbf{x}}$ is a local optimal solution to *Problem P*, then $F_0 \cap G_0 \cap H_0 = \emptyset$, where $H_0 = \{\mathbf{d} : \nabla h_i(\bar{\mathbf{x}})^t \mathbf{d} = 0 \text{ for } i = 1, \ldots, l\}$. A reader with only a casual interest in the derivation of optimality conditions may skip the proof of Theorem 4.3.1, since it involves the more advanced concepts of solving a system of differential equations.

4.3.1 Theorem

Let X be a nonempty open set in E_n. Let $f: E_n \to E_1$, $g_i: E_n \to E_1$ for $i = 1, \ldots, m$ and $h_i: E_n \to E_1$ for $i = 1, \ldots, l$. Consider *Problem P* to

$$\begin{array}{lll} \text{Minimize} & f(\mathbf{x}) & \\ \text{subject to} & g_i(\mathbf{x}) \le 0 & \text{for } i = 1, \ldots, m \\ & h_i(\mathbf{x}) = 0 & \text{for } i = 1, \ldots, l \\ & \mathbf{x} \in X & \end{array}$$

Suppose that $\bar{\mathbf{x}}$ is a local optimal solution, and let $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$. Furthermore, suppose that g_i for $i \notin I$ is continuous at $\bar{\mathbf{x}}$, that f and g_i for $i \in I$ are differentiable at $\bar{\mathbf{x}}$, and that h_i for $i = 1, \ldots, l$ is continuously differentiable at $\bar{\mathbf{x}}$. If $\nabla h_i(\bar{\mathbf{x}})$ for $i = 1, \ldots, l$ are linearly independent, then $F_0 \cap G_0 \cap H_0 = \emptyset$, where

$$F_0 = \{\mathbf{d} : \nabla f(\bar{\mathbf{x}})^t \mathbf{d} < 0\}$$
$$G_0 = \{\mathbf{d} : \nabla g_i(\bar{\mathbf{x}})^t \mathbf{d} < 0 \text{ for } i \in I\}$$
$$H_0 = \{\mathbf{d} : \nabla h_i(\bar{\mathbf{x}})^t \mathbf{d} = 0 \text{ for } i = 1, \ldots, l\}.$$

Proof

By contradiction suppose there exists a vector $\mathbf{y} \in F_0 \cap G_0 \cap H_0$; that is, $\nabla f(\bar{\mathbf{x}})^t \mathbf{y} < 0$, $\nabla g_i(\bar{\mathbf{x}})^t \mathbf{y} < 0$ for each $i \in I$, and $\nabla \mathbf{h}(\bar{\mathbf{x}})^t \mathbf{y} = \mathbf{0}$, where $\nabla \mathbf{h}(\bar{\mathbf{x}})$ is an $n \times l$ matrix whose ith column is $\nabla h_i(\bar{\mathbf{x}})$. For $\lambda \ge 0$, define $\boldsymbol{\alpha}: E_1 \to E_n$ by the following differential equation and boundary condition:

$$\frac{d\boldsymbol{\alpha}(\lambda)}{d\lambda} = \mathbf{P}(\lambda)\mathbf{y} \qquad \boldsymbol{\alpha}(0) = \bar{\mathbf{x}} \tag{4.11}$$

where $\mathbf{P}(\lambda)$ is the matrix that projects any vector in the null space of $\nabla \mathbf{h}[\boldsymbol{\alpha}(\lambda)]$. For λ sufficiently small, the above equation is well defined and solvable because $\nabla \mathbf{h}(\bar{\mathbf{x}})$ has full rank and $\mathbf{h}$ is continuously differentiable at $\bar{\mathbf{x}}$, so that $\mathbf{P}$ is continuous in λ. Obviously $\boldsymbol{\alpha}(\lambda) \to \bar{\mathbf{x}}$ as $\lambda \to 0^+$.

We now show that for $\lambda > 0$ and sufficiently small, $\boldsymbol{\alpha}(\lambda)$ is feasible and $f[\boldsymbol{\alpha}(\lambda)] < f(\bar{\mathbf{x}})$, thus contradicting local optimality of $\bar{\mathbf{x}}$. By the chain rule of differentiation and from (4.11), we get

$$\frac{d}{d\lambda} g_i[\boldsymbol{\alpha}(\lambda)] = \nabla g_i[\boldsymbol{\alpha}(\lambda)]^t \mathbf{P}(\lambda)\mathbf{y} \tag{4.12}$$

for each $i \in I$. In particular, $\mathbf{y}$ is in the null space of $\nabla \mathbf{h}(\bar{\mathbf{x}})$, and so for $\lambda = 0$, we have $\mathbf{P}(0)\mathbf{y} = \mathbf{y}$. Hence from (4.12) and the fact that $\nabla g_i(\bar{\mathbf{x}})^t \mathbf{y} < 0$, we get

$$\frac{d}{d\lambda} g_i[\boldsymbol{\alpha}(0)] = \nabla g_i(\bar{\mathbf{x}})^t \mathbf{y} < 0 \tag{4.13}$$

for $i \in I$. This further implies that $g_i[\boldsymbol{\alpha}(\lambda)] < 0$ for $\lambda > 0$ and sufficiently small. For $i \notin I$, $g_i(\bar{\mathbf{x}}) < 0$, and g_i is continuous at $\bar{\mathbf{x}}$, and thus $g_i[\boldsymbol{\alpha}(\lambda)] < 0$ for λ sufficiently small. Also, since X is open, $\boldsymbol{\alpha}(\lambda) \in X$ for λ sufficiently small. In order to show feasibility of $\boldsymbol{\alpha}(\lambda)$, we only need to show that $h_i[\boldsymbol{\alpha}(\lambda)] = 0$ for λ sufficiently small. By the mean value theorem, we have

$$\begin{aligned} h_i[\boldsymbol{\alpha}(\lambda)] &= h_i[\boldsymbol{\alpha}(0)] + \lambda \frac{d}{d\lambda} h_i[\boldsymbol{\alpha}(\mu)] \\ &= \lambda \frac{d}{d\lambda} h_i[\boldsymbol{\alpha}(\mu)] \end{aligned} \tag{4.14}$$

for some $\mu \in (0, \lambda)$. But by the chain rule of differentiation and similar to (4.12), we get

$$\frac{d}{d\lambda} h_i[\alpha(\mu)] = \nabla h_i[\alpha(\mu)]' \mathbf{P}(\mu)\mathbf{y}$$

By construction, $\mathbf{P}(\mu)\mathbf{y}$ is in the null space of $\nabla h_i[\alpha(\mu)]$, and hence from the above equation, we get $(d/d\lambda)h_i[\alpha(\mu)] = 0$. Substituting in (4.14), it follows that $h_i[\alpha(\lambda)] = 0$. Since this is true for each i, it then follows that $\alpha(\lambda)$ is a feasible solution to *Problem P* for each $\lambda > 0$ and sufficiently small. By an argument similar to that leading to (4.13), we get

$$\frac{d}{d\lambda} f[\alpha(0)] = \nabla f(\bar{\mathbf{x}})'\mathbf{y} < 0$$

and hence $f[\alpha(\lambda)] < f(\bar{\mathbf{x}})$ for $\lambda > 0$ and sufficiently small. This contradicts local optimality of $\bar{\mathbf{x}}$. Hence, $F_0 \cap G_0 \cap H_0 = \varnothing$, and the proof is complete.

The Fritz John Conditions

We now express the geometric optimality condition $F_0 \cap G_0 \cap H_0 = \varnothing$ in a more usable algebraic form. This is done in Theorem 4.3.2 below, which is a generalization of the Fritz John conditions of Theorem 4.2.6.

4.3.2 Theorem (The Fritz John Conditions)

Let X be a nonempty open set in E_n, and let $f: E_n \to E_1$, $g_i: E_n \to E_1$ for $i = 1, \ldots, m$, and $h_i: E_n \to E_1$ for $i = 1, \ldots, l$. Consider *Problem P* to

$$\begin{array}{lll} \text{Minimize} & f(\mathbf{x}) & \\ \text{subject to} & g_i(\mathbf{x}) \le 0 & \text{for } i = 1, \ldots, m \\ & h_i(\mathbf{x}) = 0 & \text{for } i = 1, \ldots, l \\ & \mathbf{x} \in X & \end{array}$$

Let $\bar{\mathbf{x}}$ be a feasible solution, and let $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$. Furthermore, suppose that g_i for $i \notin I$ is continuous at $\bar{\mathbf{x}}$, that f and g_i for $i \in I$ are differentiable at $\bar{\mathbf{x}}$, and that h_i for $i = 1, \ldots, l$ is continuously differentiable at $\bar{\mathbf{x}}$. If $\bar{\mathbf{x}}$ locally solves *Problem P*, then there exist scalars u_0, u_i for $i \in I$ and v_i for $i = 1, \ldots, l$ such that

$$u_0 \nabla f(\bar{\mathbf{x}}) + \sum_{i \in I} u_i \nabla g_i(\bar{\mathbf{x}}) + \sum_{i=1}^{l} v_i \nabla h_i(\bar{\mathbf{x}}) = \mathbf{0}$$

$$u_0, u_i \ge 0 \qquad \text{for } i \in I$$

$$(u_0, \mathbf{u}_I, \mathbf{v}) \ne (0, \mathbf{0}, \mathbf{0})$$

where $\mathbf{u}_I$ is the vector whose components are u_i for $i \in I$ and $\mathbf{v} = (v_1, \ldots, v_l)'$. Furthermore, if g_i for $i \notin I$ is also differentiable at $\bar{\mathbf{x}}$, then the Fritz John conditions can be written in the following equivalent form where $\mathbf{u} = (u_1, \ldots, u_m)'$ and $\mathbf{v} = (v_1, \ldots, v_l)'$.

$$u_0 \nabla f(\bar{\mathbf{x}}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{\mathbf{x}}) + \sum_{i=1}^{l} v_i \nabla h_i(\bar{\mathbf{x}}) = \mathbf{0}$$

não é vetor

$$u_i g_i(\bar{\mathbf{x}}) = 0 \qquad \text{for } i = 1, \ldots, m$$

$$u_0, u_i \ge 0 \qquad \text{for } i = 1, \ldots, m$$

$$(u_0, \mathbf{u}, \mathbf{v}) \ne (0, \mathbf{0}, \mathbf{0})$$

Proof

If $\nabla h_i(\bar{\mathbf{x}})$ for $i = 1, \ldots, l$ are linearly dependent, then one can find scalars $v_1, \ldots, v_l$, not all zero, such that $\sum_{i=1}^{l} v_i \nabla h_i(\bar{\mathbf{x}}) = \mathbf{0}$. Letting u_0, u_i for $i \in I$ be equal to zero, the conditions of the first part of the theorem hold trivially.

Now suppose that $\nabla h_i(\bar{\mathbf{x}})$ for $i = 1, \ldots, l$ are linearly independent. Let $\mathbf{A}_1$ be the matrix whose rows are $\nabla f(\bar{\mathbf{x}})'$ and $\nabla g_i(\bar{\mathbf{x}})'$ for $i \in I$, and let $\mathbf{A}_2$ be the matrix whose rows are $\nabla h_i(\bar{\mathbf{x}})'$ for $i = 1, \ldots, l$. Then, from Theorem 4.3.1, local optimality of $\bar{\mathbf{x}}$ implies that the system

$$\mathbf{A}_1\mathbf{d} < \mathbf{0} \qquad \mathbf{A}_2\mathbf{d} = \mathbf{0}$$

is inconsistent. Now consider the following two sets:

$$S_1 = \{(\mathbf{z}_1, \mathbf{z}_2) : \mathbf{z}_1 = \mathbf{A}_1\mathbf{d}, \mathbf{z}_2 = \mathbf{A}_2\mathbf{d}\}$$

$$S_2 = \{(\mathbf{z}_1, \mathbf{z}_2) : \mathbf{z}_1 < \mathbf{0}, \mathbf{z}_2 = \mathbf{0}\}$$

Note that S_1 and S_2 are nonempty convex sets such that $S_1 \cap S_2 = \varnothing$. Then, by Theorem 2.3.8, there exists a nonzero vector $\mathbf{p}' = (\mathbf{p}_1', \mathbf{p}_2')$ such that

$$\mathbf{p}_1'\mathbf{A}_1\mathbf{d} + \mathbf{p}_2'\mathbf{A}_2\mathbf{d} \ge \mathbf{p}_1'\mathbf{z}_1 + \mathbf{p}_2'\mathbf{z}_2 \qquad \text{for each } \mathbf{d} \in E_n \text{ and } (\mathbf{z}_1, \mathbf{z}_2) \in \text{cl } S_2.$$

Letting $\mathbf{z}_2 = \mathbf{0}$ and since each component of $\mathbf{z}_1$ can be made an arbitrarily large negative number, it follows that $\mathbf{p}_1 \ge \mathbf{0}$. Also letting $(\mathbf{z}_1, \mathbf{z}_2) = (\mathbf{0}, \mathbf{0})$, we must have $(\mathbf{p}_1'\mathbf{A}_1 + \mathbf{p}_2'\mathbf{A}_2)\mathbf{d} \ge 0$ for each $\mathbf{d} \in E_n$. Letting $\mathbf{d} = -(\mathbf{A}_1'\mathbf{p}_1 + \mathbf{A}_2'\mathbf{p}_2)$, it follows that $-\|(\mathbf{A}_1'\mathbf{p}_1 + \mathbf{A}_2'\mathbf{p}_2\|^2 \ge 0$, and thus $\mathbf{A}_1'\mathbf{p}_1 + \mathbf{A}_2'\mathbf{p}_2 = \mathbf{0}$.

To summarize, we have shown that there exists a nonzero vector $\mathbf{p}' = (\mathbf{p}_1', \mathbf{p}_2')$ with $\mathbf{p}_1 \ge \mathbf{0}$ such that $\mathbf{A}_1'\mathbf{p}_1 + \mathbf{A}_2'\mathbf{p}_2 = \mathbf{0}$. Denoting the components of $\mathbf{p}_1$ by u_0 and u_i for $i \in I$, and letting $\mathbf{p}_2 = \mathbf{v}$, the first result follows. The equivalent form of the necessary conditions is readily obtained by letting $u_i = 0$ for $i \notin I$, and the proof is complete.

The reader may note that the Lagrangian multiplier v_i associated with the ith equality constraints is unrestricted in sign. The Fritz John conditions could also be written in vector notation as follows:

$$u_0\nabla f(\bar{\mathbf{x}})+\nabla \mathbf{g}(\bar{\mathbf{x}})\mathbf{u}+\nabla \mathbf{h}(\bar{\mathbf{x}})\mathbf{v}=\mathbf{0}$$
$$\mathbf{u}^t\mathbf{g}(\bar{\mathbf{x}})=0$$
$$(u_0, \mathbf{u})\geq(0, \mathbf{0})$$
$$(u_0, \mathbf{u}, \mathbf{v})\neq(0, \mathbf{0}, \mathbf{0})$$

Here, $\nabla \mathbf{g}(\bar{\mathbf{x}})$ is an $n\times m$ matrix whose ith column is $\nabla g_i(\bar{\mathbf{x}})$, and $\nabla \mathbf{h}(\bar{\mathbf{x}})$ is an $n\times l$ matrix whose ith column is $\nabla h_i(\bar{\mathbf{x}})$. Also $\mathbf{u}$ and $\mathbf{v}$ are, respectively, an m vector and an l vector, denoting the Lagrangian multipliers associated with the inequality and equality constraints.

4.3.3 Example

Minimize $x_1^2+x_2^2$

subject to
$$x_1^2+x_2^2\leq 5$$
$$-x_1\leq 0$$
$$-x_2\leq 0$$
$$x_1+2x_2=4$$

Here we have only one equality constraint. We verify below that the Fritz John conditions are true at the optimal point $\bar{\mathbf{x}}=(\frac{4}{5}, \frac{8}{5})^t$. First note that there are no binding inequality constraints at $\bar{\mathbf{x}}$; that is $I=\varnothing$. Hence the multipliers associated with the inequality constraints are equal to zero. Note that

$$\nabla f(\bar{\mathbf{x}})=(\tfrac{8}{5}, \tfrac{16}{5})^t \qquad \text{and} \qquad \nabla h_1(\bar{\mathbf{x}})=(1, 2)^t$$

Thus

$$u_0\begin{pmatrix}\frac{8}{5}\\ \frac{16}{5}\end{pmatrix}+v_1\begin{pmatrix}1\\2\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}$$

is satisfied, for example, by $u_0=5$ and $v_1=-8$.

4.3.4 Example

Minimize $(x_1-3)^2+(x_2-2)^2$

subject to
$$x_1^2+x_2^2\leq 5$$
$$-x_1\leq 0$$
$$-x_2\leq 0$$
$$x_1+2x_2=4$$

This example is the same as Example 4.2.7, with the inequality constraint

$x_1+2x_2\leq 4$ replaced by $x_1+2x_2=4$. At the optimal point $\bar{\mathbf{x}}=(2, 1)^t$, we have only one inequality constraint $x_1^2+x_2^2\leq 5$ binding. The Fritz John condition

$$u_0\begin{pmatrix}-2\\-2\end{pmatrix}+u_1\begin{pmatrix}4\\2\end{pmatrix}+v_1\begin{pmatrix}1\\2\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}$$

is satisfied, for example, by $u_0=3$, $u_1=1$, and $v_1=2$.

4.3.5 Example

Minimize $-x_1$

subject to
$$x_2-(1-x_1)^3=0$$
$$-x_2-(1-x_1)^3=0$$

As shown in Figure 4.8, this problem has only one feasible point, namely $\bar{\mathbf{x}}=(1, 0)^t$. At this point, we have

$$\nabla f(\bar{\mathbf{x}})=(-1, 0)^t \qquad \nabla h_1(\bar{\mathbf{x}})=(0, 1)^t \qquad \nabla h_2(\bar{\mathbf{x}})=(0, -1)^t$$

The condition

$$u_0\begin{pmatrix}-1\\0\end{pmatrix}+v_1\begin{pmatrix}0\\1\end{pmatrix}+v_2\begin{pmatrix}0\\-1\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}$$

is true only if $u_0=0$ and $v_1=v_2=\alpha$, where α is any scalar. Thus the Fritz John necessary conditions are met at the point $\bar{\mathbf{x}}$.

The Kuhn–Tucker Conditions

In the Fritz John conditions, the Lagrangian multiplier associated with the objective function is not necessarily positive. Under further assumptions on the

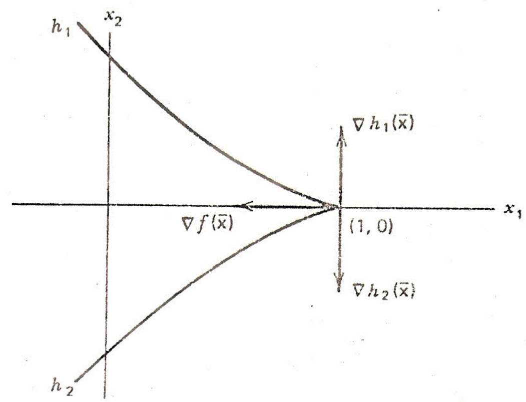


Figure 4.8 Illustration of Example 4.3.5.

constraint set, one can claim that u_0 has to be positive. In Theorem 4.3.6 below we obtain a generalization of the Kuhn–Tucker necessary optimality conditions of Theorem 4.2.10. This is done by imposing a qualification on the gradients of the equality and binding inequality constraints that ensure that $u_0>0$ in the Fritz John conditions. Other qualifications on the constraints to ensure that $u_0>0$ are discussed in Chapter 5.

4.3.6 Theorem (Kuhn–Tucker Necessary Conditions)

Let X be a nonempty open set in E_n, and let $f: E_n \to E_1$, $g_i: E_n \to E_1$ for $i=1,\ldots,m$, and $h_i: E_n \to E_1$ for $i=1,\ldots,l$. Consider *Problem P* to

$$\begin{array}{lll} \text{Minimize} & f(\mathbf{x}) & \\ \text{subject to} & g_i(\mathbf{x}) \le 0 & \text{for } i=1,\ldots,m \\ & h_i(\mathbf{x}) = 0 & \text{for } i=1,\ldots,l \\ & \mathbf{x} \in X & \end{array}$$

Let $\bar{\mathbf{x}}$ be a feasible solution, and let $I=\{i: g_i(\bar{\mathbf{x}})=0\}$. Suppose that f and g_i for $i \in I$ are differentiable at $\bar{\mathbf{x}}$, that g_i for $i \notin I$ is continuous at $\bar{\mathbf{x}}$, and that h_i for $i=1,\ldots,l$ is continuously differentiable at $\bar{\mathbf{x}}$. Further suppose that $\nabla g_i(\bar{\mathbf{x}})$ for $i \in I$ and $\nabla h_i(\bar{\mathbf{x}})$ for $i=1,\ldots,l$ are linearly independent. If $\bar{\mathbf{x}}$ solves *Problem P* locally, then there exist scalars u_i for $i \in I$ and v_i for $i=1,\ldots,l$ such that

$$\nabla f(\bar{\mathbf{x}}) + \sum_{i\in I} u_i \nabla g_i(\bar{\mathbf{x}}) + \sum_{i=1}^{l} v_i \nabla h_i(\bar{\mathbf{x}}) = \mathbf{0}$$

$$u_i \ge 0 \qquad \text{for } i \in I$$

In addition to the above assumptions, if g_i for $i \notin I$ is also differentiable at $\bar{\mathbf{x}}$, then the Kuhn–Tucker conditions could be written in the following equivalent form:

$$\nabla f(\bar{\mathbf{x}}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{\mathbf{x}}) + \sum_{i=1}^{l} v_i \nabla h_i(\bar{\mathbf{x}}) = \mathbf{0}$$

$$u_i g_i(\bar{\mathbf{x}}) = 0 \qquad \text{for } i=1,\ldots,m$$

$$u_i \ge 0 \qquad \text{for } i=1,\ldots,m$$

Proof

By Theorem 4.3.2, there exist scalars u_0 and $\hat{u}_i$ for $i \in I$, and $\hat{v}_i$ for $i=1,\ldots,l$, not all zero, such that

$$u_0 \nabla f(\bar{\mathbf{x}}) + \sum_{i\in I} \hat{u}_i \nabla g_i(\bar{\mathbf{x}}) + \sum_{i=1}^{l} \hat{v}_i \nabla h_i(\bar{\mathbf{x}}) = \mathbf{0} \tag{4.15}$$

$$u_0, \hat{u}_i \ge 0 \qquad \text{for } i \in I$$

Note that $u_0>0$, because if $u_0=0$, then (4.15) would contradict the assumption of linear independence of $\nabla g_i(\bar{\mathbf{x}})$ for $i \in I$ and $\nabla h_i(\bar{\mathbf{x}})$ for $i=1,\ldots,l$. The first result then follows by letting $u_i=\hat{u}_i/u_0$ and $v_i=\hat{v}_i/u_0$. The equivalent form of the necessary conditions follows by letting $u_i=0$ for $i \notin I$. This completes the proof.

Note that the Kuhn–Tucker condition of Theorem 4.3.6 can be written in vector form as follows:

$$\nabla f(\bar{\mathbf{x}}) + \nabla \mathbf{g}(\bar{\mathbf{x}})\mathbf{u} + \nabla \mathbf{h}(\bar{\mathbf{x}})\mathbf{v} = \mathbf{0}$$

$$\mathbf{u}^t \mathbf{g}(\bar{\mathbf{x}}) = 0$$

$$\mathbf{u} \ge \mathbf{0}$$

Here, $\nabla \mathbf{g}(\bar{\mathbf{x}})$ is an $n \times m$ matrix and $\nabla \mathbf{h}(\bar{\mathbf{x}})$ is an $n \times l$ matrix whose ith column, respectively, are $\nabla g_i(\bar{\mathbf{x}})$ and $\nabla h_i(\bar{\mathbf{x}})$. The vectors $\mathbf{u}$ and $\mathbf{v}$ are the Lagrangian multiplier vectors.

Now consider Examples 4.3.3, 4.3.4, and 4.3.5. In Example 4.3.3, the reader can verify that $u_1=u_2=u_3=0$ and $v_1=-\frac{8}{5}$ will satisfy the Kuhn–Tucker conditions at $\bar{\mathbf{x}}=(\frac{4}{5},\frac{8}{5})^t$. In Example 4.3.4, the values of the multipliers satisfying the Kuhn–Tucker conditions at $\bar{\mathbf{x}}=(2,1)^t$ are

$$u_1=\tfrac{1}{3} \qquad u_2=u_3=0 \qquad v_1=\tfrac{2}{3}$$

Finally, Example 4.3.5 does not satisfy the assumptions of Theorem 4.3.6 at $\bar{\mathbf{x}}=(1,0)^t$, since $\nabla h_1(\bar{\mathbf{x}})$ and $\nabla h_2(\bar{\mathbf{x}})$ are linearly dependent.

Theorem 4.3.7 below shows that, under rather mild convexity assumptions on f, g_i, and h_i, the Kuhn–Tucker conditions are also sufficient for optimality.

4.3.7 Theorem (Kuhn–Tucker Sufficient Conditions)

Let X be a nonempty open set in E_n, and consider $f: E_n \to E_1$, $g_i: E_n \to E_1$ for $i=1,\ldots,m$, and $h_i: E_n \to E_1$ for $i=1,\ldots,l$. Consider *Problem P* to

$$\begin{array}{lll} \text{Minimize} & f(\mathbf{x}) & \\ \text{subject to} & g_i(\mathbf{x}) \le 0 & \text{for } i=1,\ldots,m \\ & h_i(\mathbf{x}) = 0 & \text{for } i=1,\ldots,l \\ & \mathbf{x} \in X & \end{array}$$

Let $\bar{\mathbf{x}}$ be a feasible solution, and let $I=\{i: g_i(\bar{\mathbf{x}})=0\}$. Suppose that the Kuhn–Tucker conditions hold at $\bar{\mathbf{x}}$, that is, there exist scalars $\bar{u}_i \ge 0$ for $i \in I$ and $\bar{v}_i$ for

$i=1,\ldots,l$ such that

$$\nabla f(\bar{\mathbf{x}})+\sum_{i\in I}\bar{u}_i\nabla g_i(\bar{\mathbf{x}})+\sum_{i=1}^{l}\bar{v}_i\nabla h_i(\bar{\mathbf{x}})=\mathbf{0} \tag{4.16}$$

Let $J=\{i:\bar{v}_i>0\}$ and $K=\{i:\bar{v}_i<0\}$. Further suppose that f is pseudoconvex at $\bar{\mathbf{x}}$, g_i is quasiconvex at $\bar{\mathbf{x}}$ for $i\in I$, h_i is quasiconvex at $\bar{\mathbf{x}}$ for $i\in J$ and h_i is quasiconcave at $\bar{\mathbf{x}}$ for $i\in K$. Then $\bar{\mathbf{x}}$ is a global optimal solution to *Problem P.*

Proof

Let $\mathbf{x}$ be a feasible solution to *Problem P.* Then for $i\in I$, $g_i(\mathbf{x})\le g_i(\bar{\mathbf{x}})$ since $g_i(\mathbf{x})\le 0$ and $g_i(\bar{\mathbf{x}})=0$. By quasiconvexity of g_i at $\bar{\mathbf{x}}$ it follows that:

$$g_i(\bar{\mathbf{x}}+\lambda(\mathbf{x}-\bar{\mathbf{x}}))=g_i(\lambda\mathbf{x}+(1-\lambda)\bar{\mathbf{x}})\le\text{maximum}\,(g_i(\mathbf{x}),\,g_i(\bar{\mathbf{x}}))=g_i(\bar{\mathbf{x}})$$

for all $\lambda\in(0,1)$. This implies that g_i does not increase by moving from $\bar{\mathbf{x}}$ along the direction $\mathbf{x}-\bar{\mathbf{x}}$. Thus by Theorem 4.1.2, we must have

$$\nabla g_i(\bar{\mathbf{x}})^t(\mathbf{x}-\bar{\mathbf{x}})\le 0 \qquad \text{for } i\in I \tag{4.17}$$

Similarly, since h_i is quasiconvex at $\bar{\mathbf{x}}$ for $i\in J$, and h_i is quasiconcave at $\bar{\mathbf{x}}$ for $i\in K$, we have

$$\nabla h_i(\bar{\mathbf{x}})^t(\mathbf{x}-\bar{\mathbf{x}})\le 0 \qquad \text{for } i\in J \tag{4.18}$$

$$\nabla h_i(\bar{\mathbf{x}})^t(\mathbf{x}-\bar{\mathbf{x}})\ge 0 \qquad \text{for } i\in K \tag{4.19}$$

Multiplying (4.17), (4.18), and (4.19), respectively, by $\bar{u}_i\ge 0$, $\bar{v}_i>0$ and $\bar{v}_i<0$ and adding, we get

$$\left[\sum_{i\in I}\bar{u}_i\nabla g_i(\bar{\mathbf{x}})+\sum_{i\in J\cup K}\bar{v}_i\nabla h_i(\bar{\mathbf{x}})\right]^t(\mathbf{x}-\bar{\mathbf{x}})\le 0 \tag{4.20}$$

Multiplying (4.16) by $\mathbf{x}-\bar{\mathbf{x}}$, and noting that $\bar{v}_i=0$ for $i\notin J\cup K$, then (4.20) implies that

$$\nabla f(\bar{\mathbf{x}})^t(\mathbf{x}-\bar{\mathbf{x}})\ge 0$$

By pseudoconvexity of f at $\bar{\mathbf{x}}$, then $f(\mathbf{x})\ge f(\bar{\mathbf{x}})$, and the proof is complete.

Alternative Forms of the Kuhn–Tucker Conditions for General Problems

Consider the problem to minimize $f(\mathbf{x})$ subject to $g_i(\mathbf{x})\le 0$ for $i=1,\ldots,m$, $h_i(\mathbf{x})=0$ for $i=1,\ldots,l$, and $\mathbf{x}\in X$, where X is an open set in E_n. In this section, we derived the following necessary conditions of optimality at a feasible point $\bar{\mathbf{x}}$.

$$\nabla f(\bar{\mathbf{x}})+\sum_{i=1}^{m}u_i\nabla g_i(\bar{\mathbf{x}})+\sum_{i=1}^{l}v_i\nabla h_i(\bar{\mathbf{x}})=\mathbf{0}$$
$$u_ig_i(\bar{\mathbf{x}})=0 \qquad \text{for } i=1,\ldots,m$$
$$u_i\ge 0 \qquad \text{for } i=1,\ldots,m$$

Some authors prefer to use the multipliers $\lambda_i=-u_i\le 0$ and $\mu_i=-v_i$. In this case, the Kuhn–Tucker conditions could be written as follows:

$$\nabla f(\bar{\mathbf{x}})-\sum_{i=1}^{m}\lambda_i\nabla g_i(\bar{\mathbf{x}})-\sum_{i=1}^{l}\mu_i\nabla h_i(\bar{\mathbf{x}})=\mathbf{0}$$
$$\lambda_ig_i(\bar{\mathbf{x}})=0 \qquad \text{for } i=1,\ldots,m$$
$$\lambda_i\le 0 \qquad \text{for } i=1,\ldots,m$$

Now consider the problem to minimize $f(\mathbf{x})$ subject to $g_i(\mathbf{x})\le 0$ for $i=1,\ldots,m_1$, $g_i(\mathbf{x})\ge 0$ for $i=m_1+1,\ldots,m$, $h_i(\mathbf{x})=0$ for $i=1,\ldots,l$, and $\mathbf{x}\in X$, where X is an open set in E_n. Clearly, one can write $g_i(\mathbf{x})\ge 0$ for $i=m_1+1,\ldots,m$ as $-g_i(\mathbf{x})\le 0$ for $i=m_1+1,\ldots,m$, and use the results of Theorem 4.3.6. It is easy to verify that the necessary conditions can be expressed as follows:

$$\nabla f(\bar{\mathbf{x}})+\sum_{i=1}^{m}u_i\nabla g_i(\bar{\mathbf{x}})+\sum_{i=1}^{l}v_i\nabla h_i(\bar{\mathbf{x}})=\mathbf{0}$$
$$u_ig_i(\bar{\mathbf{x}})=0 \qquad \text{for } i=1,\ldots,m$$
$$u_i\ge 0 \qquad \text{for } i=1,\ldots,m_1$$
$$u_i\le 0 \qquad \text{for } i=m_1+1,\ldots,m$$

We now consider problems of the type to minimize $f(\mathbf{x})$ subject to $g_i(\mathbf{x})\le 0$ for $i=1,\ldots,m$, $h_i(\mathbf{x})=0$ for $i=1,\ldots,l$, and $\mathbf{x}\ge\mathbf{0}$. Such problems with nonnegativity restrictions on the variables frequently arise in practice. Clearly, the Kuhn–Tucker conditions discussed earlier would apply. However, it is sometimes convenient to eliminate the Lagrangian multipliers associated with $\mathbf{x}\ge\mathbf{0}$. The conditions then reduce to

$$\nabla f(\bar{\mathbf{x}})+\sum_{i=1}^{m}u_i\nabla g_i(\bar{\mathbf{x}})+\sum_{i=1}^{l}v_i\nabla h_i(\bar{\mathbf{x}})\ge\mathbf{0}$$
$$\left[\nabla f(\bar{\mathbf{x}})+\sum_{i=1}^{m}u_i\nabla g_i(\bar{\mathbf{x}})+\sum_{i=1}^{l}v_i\nabla h_i(\bar{\mathbf{x}})\right]^t\mathbf{x}=0$$
$$u_ig_i(\bar{\mathbf{x}})=0 \qquad \text{for } i=1,\ldots,m$$
$$u_i\ge 0 \qquad \text{for } i=1,\ldots,m$$

Finally, consider the problem to maximize $f(\mathbf{x})$ subject to $g_i(\mathbf{x})\le 0$ for $i=1,\ldots,m_1$, $g_i(\mathbf{x})\ge 0$ for $i=m_1+1,\ldots,m$, $h_i(\mathbf{x})=0$ for $i=1,\ldots,l$, and

$\mathbf{x} \in X$, where X is an open set in E_n. The necessary conditions for optimality can be written as follows:

$$\nabla f(\bar{\mathbf{x}}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{\mathbf{x}}) + \sum_{i=1}^{l} v_i \nabla h_i(\bar{\mathbf{x}}) = \mathbf{0}$$

$$u_i g_i(\bar{\mathbf{x}}) = 0 \qquad \text{for } i = 1, \ldots, m$$

$$u_i \leq 0 \qquad \text{for } i = 1, \ldots, m_1$$

$$u_i \geq 0 \qquad \text{for } i = m_1 + 1, \ldots, m$$

Exercises

4.1 Consider the following unconstrained problem:

$$\text{minimize } x_1^2 - x_1 x_2 + 2x_2^2 - 2x_1 + e^{x_1 + x_2}$$

a. Write the first-order necessary optimality condition. Is this condition also sufficient for optimality? Why?

b. Is $\bar{\mathbf{x}} = (0, 0)'$ an optimal solution? If not, identify a direction $\mathbf{d}$ along which the function would decrease.

c. Minimize the function starting from $(0, 0)$ along the direction $\mathbf{d}$ obtained in part b above.

4.2 Consider the problem to minimize $\|\mathbf{Ax} - \mathbf{b}\|^2$, where $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{b}$ is an m vector.

a. Give a geometric interpretation of the problem.

b. Write a necessary condition for optimality. Is this also a sufficient condition?

c. Is the optimal solution unique? Why or why not?

d. Can you give a closed-form solution of the optimal solution? Specify any assumptions that you may need.

e. Solve the problem for $\mathbf{A}$ and $\mathbf{b}$ given below

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

4.3 Consider the problem to minimize $f(\mathbf{x})$ subject to $g_i(\mathbf{x}) \leq 0$ for $i = 1, \ldots, m$. Let $\bar{\mathbf{x}}$ be a local minimal point, and let $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$. Suppose that f is differentiable at $\bar{\mathbf{x}}$, that g_i for $i \in I$ is differentiable and concave at $\bar{\mathbf{x}}$, and that g_i for $i \notin I$ is continuous at $\bar{\mathbf{x}}$. Prove that $F_0 \cap G' = \varnothing$, where

$$F_0 = \{\mathbf{d} : \nabla f(\bar{\mathbf{x}})'\mathbf{d} < 0\}, \quad \text{and} \quad G' = \{\mathbf{d} : \nabla g_i(\bar{\mathbf{x}})'\mathbf{d} \leq 0 \quad \text{for } i \in I\}$$

4.4 Consider the following problem:

$$\begin{aligned} \text{Maximize} \quad & 3x_1 - x_2 + x_3^2 \\ \text{subject to} \quad & x_1 + x_2 + x_3 \leq 0 \\ & -x_1 + 2x_2 + x_3^2 = 0 \end{aligned}$$

a. Write the Kuhn–Tucker optimality conditions

b. Using the above conditions, find the optimal solution to the problem.

4.5 Consider the following problem:

$$\begin{aligned} \text{Maximize} \quad & x_1^2 + 4x_1 x_2 + x_2^2 \\ \text{subject to} \quad & x_1^2 + x_2^2 = 1 \end{aligned}$$

a. Using the Kuhn–Tucker conditions, find an optimal solution to the problem.

b. Does the problem have a unique optimal solution?

4.6 Consider the following linear program:

$$\begin{aligned} \text{Maximize} \quad & 2x_1 + 3x_2 \\ \text{subject to} \quad & x_1 + x_2 \leq 8 \\ & -x_1 + 2x_2 \leq 4 \\ & x_1, x_2 \geq 0 \end{aligned}$$

a. Write the Kuhn–Tucker optimality conditions.
b. For each extreme point, verify whether or not the Kuhn–Tucker conditions are true, both algebraically and geometrically. From this, find the optimal solution.

4.7 Consider the following problem:

$$\begin{array}{ll} \text{Minimize} & (x_1-\frac{9}{4})^2+(x_2-2)^2 \\ \text{subject to} & x_2-x_1^2\geq 0 \\ & x_1+x_2\leq 6 \\ & x_1, x_2\geq 0 \end{array}$$

a. Write the Kuhn–Tucker optimality conditions and verify that these conditions are true at the point $\bar{\mathbf{x}}=(\frac{3}{2},\frac{9}{4})^t$.
b. Interpret the Kuhn–Tucker conditions at $\bar{\mathbf{x}}$ graphically.
c. Show that $\bar{\mathbf{x}}$ is indeed the unique global optimal solution.

4.8 Consider the following problem.

$$\begin{array}{ll} \text{Minimize} & x_1^2+2x_2^2 \\ \text{subject to} & x_1 + x_2-1=0 \end{array}$$

Find a point satisfying the Kuhn–Tucker conditions and verify that it is indeed an optimal solution. Resolve the problem if the objective function is replaced by $x_1^3+x_2^3$.

4.9 Write the Kuhn–Tucker necessary optimality conditions for Exercises 1.10 and 1.11. Using these conditions, find the optimal solutions.

4.10 Consider the following one-dimensional minimization problem:

$$\begin{array}{ll} \text{Minimize} & f(\mathbf{x}+\lambda\mathbf{d}) \\ \text{subject to} & \lambda\geq 0 \end{array}$$

where $\mathbf{x}$ is a given vector and $\mathbf{d}$ is a given nonzero direction.
a. Write a necessary condition for a minimum if f is differentiable. Is this condition also sufficient? If not, what assumptions on f would make the necessary condition also sufficient?
b. Suppose that f is convex but not differentiable. Can you develop a necessary optimality condition for the above problem using subgradients of f defined in Section 3.2.

4.11 Consider the following problem.

$$\begin{array}{ll} \text{Minimize} & \dfrac{x_1+3x_2+3}{2x_1+x_2+6} \\ \text{subject to} & 2x_1+ x_2\leq 12 \\ & -x_1+2x_2\leq 4 \\ & x_1, x_2\geq 0 \end{array}$$

a. Show that the Kuhn–Tucker conditions are sufficient for this problem.

b. Show that any point on the line segment joining the points (0, 0) and (6, 0) is an optimal solution.

4.12 Use the Kuhn–Tucker conditions to prove Farkas' theorem discussed in Section 2.3. (*Hint:* Consider the problem maximize $\mathbf{c}^t\mathbf{x}$ subject to $\mathbf{A}\mathbf{x}\leq\mathbf{0}$.)

4.13 Consider the problem to minimize $f(\mathbf{x})$ subject to $g_i(\mathbf{x})\leq 0$ for $i=1,\ldots,m$.
a. Show that verifying whether a point $\bar{\mathbf{x}}$ is a Kuhn–Tucker point is equivalent to finding a vector $\mathbf{u}$ satisfying a system of the form $\mathbf{A}^t\mathbf{u}=\mathbf{c}$, $\mathbf{u}\geq\mathbf{0}$. This can be done using Phase I of linear programming.
b. Indicate the modifications needed in part a if the problem had equality constraints.
c. Illustrate part a by the following problem, where $\bar{\mathbf{x}}=(1, 2, 5)^t$.

$$\begin{array}{ll} \text{Minimize} & 2x_1^2+x_2^2+2x_3^2+x_1x_3-x_1x_2+x_1+2x_3 \\ \text{subject to} & x_1^2+x_2^2- x_3\leq 0 \\ & x_1 +x_2 +2x_3\leq 16 \\ & x_1 +x_2 \geq 3 \\ & x_1, x_2, x_3\geq 0 \end{array}$$

4.14 Consider the problem to minimize $f(\mathbf{x})$ subject to $g_i(\mathbf{x})\leq 0$ for $i=1,\ldots,m$. Let $\bar{\mathbf{x}}$ be a feasible point, and let $I=\{i: g_i(\bar{\mathbf{x}})=0\}$. Suppose that f is differentiable at $\bar{\mathbf{x}}$ and that g_i for $i\in I$ is differentiable and concave at $\bar{\mathbf{x}}$. Furthermore, suppose that g_i for $i\notin I$ is continuous at $\bar{\mathbf{x}}$. Consider the following linear problem:

$$\begin{array}{lll} \text{Minimize} & \nabla f(\bar{\mathbf{x}})^t\mathbf{d} & \\ \text{subject to} & \nabla g_i(\bar{\mathbf{x}})^t\mathbf{d}\leq 0 & \text{for } i\in I \\ & -1\leq d_j\leq 1 & \text{for } j=1,\ldots,n \end{array}$$

Let $\bar{\mathbf{d}}$ be an optimal solution with objective function value $\bar{z}$.
a. Show that $\bar{z}\leq 0$.
b. Show that if $\bar{z}<0$, then there exists a $\delta>0$ such that $\bar{\mathbf{x}}+\lambda\bar{\mathbf{d}}$ is feasible and $f(\bar{\mathbf{x}}+\lambda\bar{\mathbf{d}})<f(\bar{\mathbf{x}})$ for each $\lambda\in(0,\delta)$.
c. Show that if $\bar{z}=0$, then $\bar{\mathbf{x}}$ satisfies the Kuhn–Tucker conditions.

4.15 Let $f: E_n\to E_1$, $g_i: E_n\to E_1$ for $i=1,\ldots,m$ be convex functions. Consider the problem to minimize $f(\mathbf{x})$ subject to $g_i(\mathbf{x})\leq 0$ for $i=1,\ldots,m$. Let M be a proper subset of $\{1,\ldots,m\}$, and suppose that $\hat{\mathbf{x}}$ solves the problem to minimize $f(\mathbf{x})$ subject to $g_i(\mathbf{x})\leq 0$ for $i\in M$. Let $V=\{i: g_i(\hat{\mathbf{x}})>0\}$. If $\bar{\mathbf{x}}$ solves the original problem, show that $g_i(\bar{\mathbf{x}})=0$ for some $i\in V$.
(This exercise also shows that if the unconstrained minimum of f is infeasible, then the constrained minimum lies on the boundary of the feasible region.)

4.16 Consider the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{x}\geq\mathbf{0}$, where f is a differentiable convex function. Let $\bar{\mathbf{x}}$ be a given point and denote $\nabla f(\bar{\mathbf{x}})$ by $(\nabla_1,\ldots,\nabla_n)^t$. Show that $\bar{\mathbf{x}}$ is an optimal solution if and only if $\mathbf{d}=\mathbf{0}$, where $\mathbf{d}$ is defined by

$$d_i=\begin{cases} -\nabla_i & \text{if } x_i>0 \text{ or } \nabla_i<0 \\ 0 & \text{if } x_i=0 \text{ and } \nabla_i\geq 0 \end{cases}$$

4.17 Consider the following problem:

$$\begin{array}{lll} \text{Minimize} & \sum_{j=1}^{n} f_j(x_j) & \\ \text{subject to} & \sum_{j=1}^{n} x_j = 1 & \\ & x_j \geq 0 & \text{for } j=1,2,\ldots,n. \end{array}$$

Suppose that $\bar{\mathbf{x}} = (\bar{x}_1, \ldots, \bar{x}_n)' \geq \mathbf{0}$ solves the above problem. Letting $\delta_j = \dfrac{\partial f_j(\bar{\mathbf{x}})}{\partial x_j}$, show that there exists a scalar k such that

$$\delta_j \geq k \quad \text{and} \quad (\delta_j - k)\bar{x}_j = 0 \quad \text{for } j = 1, 2, \ldots, n$$

4.18 Consider the following problem, where $\mathbf{c}$ is a nonzero vector in E_n.

$$\begin{array}{ll} \text{Maximize} & \mathbf{c}'\mathbf{d} \\ \text{subject to} & \mathbf{d}'\mathbf{d} \leq 1 \end{array}$$

a. Show that $\bar{\mathbf{d}} = \mathbf{c}/\|\mathbf{c}\|$ is a Kuhn–Tucker point. Furthermore, show that $\bar{\mathbf{d}}$ is indeed the unique global optimal solution.
b. Using the result of part a, show that the direction of steepest ascent of f at a point $\mathbf{x}$ is given by $\nabla f(\mathbf{x})/\|\nabla f(\mathbf{x})\|$ provided that $\nabla f(\mathbf{x}) \neq \mathbf{0}$.

4.19 Consider the following problem, where a_j, b, and c_j are positive constants.

$$\begin{array}{lll} \text{Minimize} & \sum_{j=1}^{n} \dfrac{c_j}{x_j} & \\ \text{subject to} & \sum_{j=1}^{n} a_j x_j = b & \\ & x_j \geq 0 & \text{for } j=1,\ldots,n \end{array}$$

Write the Kuhn–Tucker conditions, and solve for the point $\bar{\mathbf{x}}$ satisfying these conditions.

4.20 In geometric programming, the following result is used. If $x_1, \ldots, x_n \geq 0$, then

$$\frac{1}{n}\sum_{j=1}^{n} x_j \geq \left(\prod_{j=1}^{n} x_j\right)^{1/n}$$

Prove the result using the Kuhn–Tucker conditions.
Hint: Consider one of the following problems:

Minimize $\sum_{j=1}^{n} x_j$ subject to $\prod_{j=1}^{n} x_j = 1$, $x_j \geq 0$ for $j = 1, \ldots, n$.

Maximize $\prod_{j=1}^{n} x_j$ subject to $\sum_{j=1}^{n} x_j = 1$, $x_j \geq 0$ for $j = 1, \ldots, n$.

4.21 Let $\mathbf{c}$ be an n vector, $\mathbf{b}$ an m vector, $\mathbf{A}$ an $m \times n$ matrix, and $\mathbf{H}$ a symmetric $n \times n$ positive definite matrix. Consider the following two problems:

(1) Minimize $\mathbf{c}'\mathbf{x} + \frac{1}{2}\mathbf{x}'\mathbf{H}\mathbf{x}$ subject to $\mathbf{A}\mathbf{x} \leq \mathbf{b}$

(2) Minimize $\mathbf{h}'\mathbf{v} + \frac{1}{2}\mathbf{v}'\mathbf{G}\mathbf{v}$ subject to $\mathbf{v} \geq \mathbf{0}$

where $\mathbf{G} = \mathbf{A}\mathbf{H}^{-1}\mathbf{A}'$ and $\mathbf{h} = \mathbf{A}\mathbf{H}^{-1}\mathbf{c} + \mathbf{b}$. Investigate the relationship between the Kuhn–Tucker conditions of Problems 1 and 2.

4.22 Consider the following problem:

$$\begin{array}{ll} \text{Minimize} & f(\mathbf{x}) \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{array}$$

Let $\bar{\mathbf{x}}' = (\bar{\mathbf{x}}_B', \bar{\mathbf{x}}_N')$ be an extreme point, where $\bar{\mathbf{x}}_B = \mathbf{B}^{-1}\mathbf{b} > \mathbf{0}$, $\bar{\mathbf{x}}_N = \mathbf{0}$, and $\mathbf{A} = [\mathbf{B}, \mathbf{N}]$ with $\mathbf{B}$ invertible. Now consider the following direction-finding problem:

$$\begin{array}{lll} \text{Minimize} & [\nabla_N f(\bar{\mathbf{x}}) - \nabla_B f(\bar{\mathbf{x}})\mathbf{B}^{-1}\mathbf{N}]'\mathbf{d}_N & \\ \text{subject to} & 0 \leq d_j \leq 1 & \text{for each nonbasic component } j \end{array}$$

where $\nabla_B f(\bar{\mathbf{x}})$ and $\nabla_N f(\bar{\mathbf{x}})$ denote the gradient of f with respect to the basic and nonbasic variables respectively. Let $\bar{\mathbf{d}}_N$ be an optimal solution, and let $\bar{\mathbf{d}}_B = -\mathbf{B}^{-1}\mathbf{N}\bar{\mathbf{d}}_N$. Show that if $\bar{\mathbf{d}}' = (\bar{\mathbf{d}}_B', \bar{\mathbf{d}}_N') \neq (\mathbf{0}, \mathbf{0})$, then it is an improving feasible direction. What are the implications of $\bar{\mathbf{d}} = \mathbf{0}$?

4.23 Consider the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{A}\mathbf{x} \leq \mathbf{b}$. Suppose that $\bar{\mathbf{x}}$ is a feasible solution such that $\mathbf{A}_1\bar{\mathbf{x}} = \mathbf{b}_1$ and $\mathbf{A}_2\bar{\mathbf{x}} < \mathbf{b}_2$, where $\mathbf{A}' = (\mathbf{A}_1', \mathbf{A}_2')$ and $\mathbf{b}' = (\mathbf{b}_1', \mathbf{b}_2')$. Assuming that $\mathbf{A}_1$ has full rank, the matrix $\mathbf{P}$ that projects any vector in the null space of $\mathbf{A}_1$ is given by

$$\mathbf{P} = \mathbf{I} - \mathbf{A}_1'(\mathbf{A}_1\mathbf{A}_1')^{-1}\mathbf{A}_1$$

a. Let $\bar{\mathbf{d}} = -\mathbf{P}\nabla f(\bar{\mathbf{x}})$. Show that if $\bar{\mathbf{d}} \neq \mathbf{0}$, then it is an improving feasible direction; that is, $\bar{\mathbf{x}} + \lambda\bar{\mathbf{d}}$ is feasible and that $f(\bar{\mathbf{x}} + \lambda\bar{\mathbf{d}}) < f(\bar{\mathbf{x}})$ for $\lambda > 0$ and sufficiently small.
b. Suppose that $\bar{\mathbf{d}} = \mathbf{0}$, and that $\mathbf{u} = -(\mathbf{A}_1\mathbf{A}_1')^{-1}\mathbf{A}_1\nabla f(\bar{\mathbf{x}}) \geq \mathbf{0}$. Show that $\bar{\mathbf{x}}$ is a Kuhn–Tucker point.
c. Show that $\bar{\mathbf{d}}$ generated above is of the form $\lambda\hat{\mathbf{d}}$ for some $\lambda > 0$, where $\hat{\mathbf{d}}$ is an optimal solution of the following problem:

$$\begin{array}{ll} \text{Minimize} & \nabla f(\bar{\mathbf{x}})'\mathbf{d} \\ \text{subject to} & \mathbf{A}_1\mathbf{d} = \mathbf{0} \\ & \|\mathbf{d}\|^2 \leq 1 \end{array}$$

d. Make all possible simplifications if $\mathbf{A} = -\mathbf{I}$ and $\mathbf{b} = \mathbf{0}$, that is, if the constraints are of the form $\mathbf{x} \geq \mathbf{0}$.

4.24 Consider the following problem:

$$\begin{array}{ll} \text{Minimize} & x_1^2 - x_1x_2 + 2x_2^2 - 4x_1 - 5x_2 \\ \text{subject to} & x_1 + 2x_2 \leq 6 \\ & x_1 \leq 2 \\ & x_1, x_2 \geq 0 \end{array}$$

a. Solve the problem geometrically, and verify the optimality of your solution by the Kuhn–Tucker conditions.

b. Find the direction $\bar{\mathbf{d}}$ of Exercise 4.23 at the optimal solution. Verify that $\bar{\mathbf{d}}=\mathbf{0}$ and that $\mathbf{u}\geq\mathbf{0}$.
c. Find the direction $\bar{\mathbf{d}}$ of Exercise 4.23 at $\bar{\mathbf{x}}=(1,\frac{5}{3})^t$. Verify that $\bar{\mathbf{d}}$ is an improving feasible direction. Also verify that the optimal solution $\hat{\mathbf{d}}$ of part c of Exercise 4.23 indeed points along $\bar{\mathbf{d}}$.

4.25 Investigate the relationship between the optimal solutions and the Kuhn–Tucker conditions of the following two problems, where $\boldsymbol{\lambda}\geq\mathbf{0}$ is a given fixed vector.

Problem P: Minimize $f(\mathbf{x})$ subject to $\mathbf{x}\in X$, $\mathbf{g}(\mathbf{x})\leq\mathbf{0}$

Problem P': Minimize $f(\mathbf{x})$ subject to $\mathbf{x}\in X$, $\boldsymbol{\lambda}^t\mathbf{g}(\mathbf{x})\leq 0$.

(*Problem P'* has only one constraint and is referred to as the *surrogate problem.*)

4.26 Consider the following problem:

$$\begin{array}{lll} \text{Minimize} & f(\mathbf{x}) & \\ \text{subject to} & g_i(\mathbf{x})\leq 0 & \text{for } i=1,\ldots,m \\ & h_i(\mathbf{x})=0 & \text{for } i=1,\ldots,l \\ & \mathbf{x}\in X & \end{array}$$

Let $\bar{\mathbf{x}}$ be a local optimal solution to the problem, and let $I=\{i: g_i(\bar{\mathbf{x}})=0\}$. Furthermore, suppose that $\nabla g_i(\bar{\mathbf{x}})$ for $i\in I$ and $\nabla h_i(\bar{\mathbf{x}})$ for $i=1,\ldots,l$ are linearly independent.
The second-order necessary conditions for local optimality can be written as follows. There exist a vector $\mathbf{u}\geq\mathbf{0}$ and a vector $\mathbf{v}$ such that

$$\nabla f(\bar{\mathbf{x}})+\sum_{i=1}^{m}u_i\nabla g_i(\bar{\mathbf{x}})+\sum_{i=1}^{l}v_i\nabla h_i(\bar{\mathbf{x}})=\mathbf{0}$$

$$u_i g_i(\bar{\mathbf{x}})=0 \qquad \text{for } i=1,\ldots,m$$

and such that

$$\mathbf{L}(\bar{\mathbf{x}})=\mathbf{F}(\bar{\mathbf{x}})+\sum_{i=1}^{m}u_i\mathbf{G}_i(\bar{\mathbf{x}})+\sum_{i=1}^{l}v_i\mathbf{H}_i(\bar{\mathbf{x}})$$

is positive semidefinite on the linear subspace

$$M=\{\mathbf{y}:\nabla g_i(\bar{\mathbf{x}})\mathbf{y}=0 \quad \text{for } i\in I, \quad \nabla h_i(\bar{\mathbf{x}})\mathbf{y}=0 \quad \text{for } i=1,\ldots,l\}$$

where $\mathbf{F}(\bar{\mathbf{x}})$, $\mathbf{G}_i(\bar{\mathbf{x}})$, and $\mathbf{H}_i(\bar{\mathbf{x}})$ are the Hessian matrices of f, g_i, and h_i, respectively, at $\bar{\mathbf{x}}$.
a. Verify the second-order necessary optimality conditions in Exercise 4.4.
b. The above conditions need not be sufficient for a local minimum. However, if $\mathbf{L}(\bar{\mathbf{x}})$ is positive definite on

$$M'=\{\mathbf{y}:\nabla g_i(\bar{\mathbf{x}})\mathbf{y}=0 \quad \text{if } u_i>0, \quad \nabla h_i(\bar{\mathbf{x}})\mathbf{y}=0 \quad \text{for } i=1,\ldots,l\}$$

then $\bar{\mathbf{x}}$ is indeed a local minimum. Does this second-order sufficient condition hold at the optimal point obtained in part a above?

4.27 Consider the following problem:

$$\begin{array}{ll} \text{Minimize} & \mathbf{c}^t\mathbf{x}+\frac{1}{2}\mathbf{x}^t\mathbf{H}\mathbf{x} \\ \text{subject to} & \mathbf{A}\mathbf{x}\leq\mathbf{b} \end{array}$$

where $\mathbf{c}$ is an n vector, $\mathbf{b}$ is an m vector, $\mathbf{A}$ is an $m\times n$ matrix, and $\mathbf{H}$ is an $n\times n$ symmetric matrix.
a. Write the second-order necessary optimality conditions of Exercise 4.26. Make all possible simplifications.
b. Is it necessarily true that every local minimum to the above problem is also a global minimum? Prove or give a counterexample.
c. Provide the first- and second-order necessary optimality conditions for the special case where $\mathbf{c}=\mathbf{0}$ and $\mathbf{H}=\mathbf{I}$. In this case the problem reduces to finding the point in a polyhedral set closest to the origin.
(The above problem is referred to in the literature as a *least distance programming problem.*)

4.28 Consider the following problem:

$$\begin{array}{ll} \text{Minimize} & -x_1+x_2 \\ \text{subject to} & x_1^2+x_2^2-2x_1=0 \\ & (x_1,x_2)\in X \end{array}$$

where X is the convex combinations of the points $(-1,0)$, $(0,1)$, $(1,0)$, and $(-1,0)$.
a. Find the optimal solution graphically.
b. Do the Fritz John or the Kuhn–Tucker conditions hold at the optimal solution in part a? If not, explain in terms of Theorems 4.3.2 and 4.3.6.
c. Replace the set X by a suitable system of inequalities and answer part b. What are your conclusions?

4.29 Consider the following problem to minimize $f(\mathbf{x})$ subject to $g_i(\mathbf{x})\leq 0$ for $i=1,\ldots,m$ and $h_i(\mathbf{x})=0$ for $i=1,\ldots,l$. Suppose that $\bar{\mathbf{x}}$ solves the problem locally, and let $I=\{i: g_i(\bar{\mathbf{x}})=0\}$. Furthermore, suppose that g_i for $i\in I$ is differentiable at $\bar{\mathbf{x}}$, g_i for $i\notin I$ is continuous at $\bar{\mathbf{x}}$, $h_1,\ldots,h_l$ are affine, that is, h_i is of the form $h_i(\mathbf{x})=\mathbf{a}_i^t\mathbf{x}-b_i$.
a. Show that $F_0\cap G\cap H_0=\emptyset$, where

$$F_0=\{\mathbf{d}:\nabla f(\bar{\mathbf{x}})^t\mathbf{d}<0\}$$
$$H_0=\{\mathbf{d}:\nabla h_i(\bar{\mathbf{x}})^t\mathbf{d}=0 \text{ for } i=1,\ldots,l\}$$
$$G=\{\mathbf{d}:\nabla g_i(\bar{\mathbf{x}})^t\mathbf{d}\leq 0 \text{ for } i\in J \text{ and } \nabla g_i(\bar{\mathbf{x}})^t\mathbf{d}<0 \text{ for } i\in I-J\}$$
$$J=\{i\in I: g_i \text{ is pseudoconcave at } \bar{\mathbf{x}}\}$$

b. Can this condition be verified by using linear programming? If so, illustrate in detail.

4.30 Let X be a nonempty open set in E_n, and consider $f: E_n\to E_1$, $g_i: E_n\to E_1$ for $i=1,\ldots,m$, and $h_i: E_n\to E_1$ for $i=1,\ldots,l$. Consider *Problem P* to

$$\begin{array}{lll} \text{Minimize} & f(\mathbf{x}) & \\ \text{subject to} & g_i(\mathbf{x})\leq 0 & \text{for } i=1,\ldots,m \\ & h_i(\mathbf{x})=0 & \text{for } i=1,\ldots,l \\ & \mathbf{x}\in X & \end{array}$$

Let $\bar{\mathbf{x}}$ be a feasible solution, and let $I=\{i:g_i(\bar{\mathbf{x}})=0\}$. Suppose that the Kuhn–Tucker conditions hold at $\bar{\mathbf{x}}$, that is, there exist scalars $\bar{u}_i \geq 0$ for $i\in I$ and $\bar{v}_i$ for $i=1,\ldots,l$ such that

$$\nabla f(\bar{\mathbf{x}})+\sum_{i\in I}\bar{u}_i\nabla g_i(\bar{\mathbf{x}})+\sum_{i=1}^{l}\bar{v}_i\nabla h_i(\bar{\mathbf{x}})=\mathbf{0}$$

a. Suppose that f is pseudoconvex at $\bar{\mathbf{x}}$ and that ϕ is quasiconvex at $\bar{\mathbf{x}}$, where

$$\phi(\mathbf{x})=\sum_{i\in I}\bar{u}_i g_i(\mathbf{x})+\sum_{i=1}^{l}\bar{v}_i h_i(\mathbf{x})$$

Show that $\bar{\mathbf{x}}$ is a global optimal solution to *Problem P*.

b. Show that if $f+\sum_{i\in I}\bar{u}_i g_i+\sum_{i=1}^{l}\bar{v}_i h_i$ is pseudoconvex, then $\bar{\mathbf{x}}$ is a global optimal solution to *Problem P*.

c. Show by means of examples that the convexity assumptions in parts a and b above and those of Theorem 4.3.7 are not equivalent to each other.

4.31 Consider the *bilinear program* to minimize $\mathbf{c}'\mathbf{x}+\mathbf{d}'\mathbf{y}+\mathbf{x}'\mathbf{H}\mathbf{y}$ subject to $\mathbf{x}\in X$ and $\mathbf{y}\in Y$, where X and Y are bounded polyhedral sets in E_n and E_m, respectively. Let $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ be extreme points of the sets X and Y, respectively.

a. Verify that the objective function is neither quasiconvex nor quasiconcave.

b. Prove that there exists an extreme point $(\bar{\mathbf{x}},\bar{\mathbf{y}})$ that solves the bilinear program.

c. Prove that the point $(\hat{\mathbf{x}},\hat{\mathbf{y}})$ is a local minimum of the bilinear program if and only if the following are true:
 (i) $\mathbf{c}'(\mathbf{x}-\hat{\mathbf{x}})\geq 0$ and $\mathbf{d}'(\mathbf{y}-\hat{\mathbf{y}})\geq 0$ for each $\mathbf{x}\in X$ and $\mathbf{y}\in Y$,
 (ii) $\mathbf{c}'(\mathbf{x}-\hat{\mathbf{x}})+\mathbf{d}'(\mathbf{y}-\hat{\mathbf{y}})>0$ whenever $(\mathbf{x}-\hat{\mathbf{x}})'\mathbf{H}(\mathbf{y}-\hat{\mathbf{y}})<0$.

d. Show that the point $(\hat{\mathbf{x}},\hat{\mathbf{y}})$ is a Kuhn–Tucker point if and only if $(\mathbf{c}'+\hat{\mathbf{y}}'\mathbf{H})(\mathbf{x}-\hat{\mathbf{x}})\geq 0$ for each $\mathbf{x}\in X$ and $(\mathbf{d}'+\hat{\mathbf{x}}'\mathbf{H})(\mathbf{y}-\hat{\mathbf{y}})\geq 0$ for each $\mathbf{y}\in Y$.

e. Consider the problem to minimize $x_2+y_1+x_2y_1-x_1y_2+x_2y_2$ subject to $(x_1,x_2)\in X$ and $(y_1,y_2)\in Y$, where X is the polyhedral set defined by its extreme points $(0,0)$, $(0,1)$, $(1,4)$, $(2,4)$, and $(3,0)$, and Y is the polyhedral set defined by its extreme points $(0,0)$, $(0,1)$, $(1,5)$, $(3,5)$, $(4,4)$, and $(3,0)$. Verify that the point $(x_1,x_2,y_1,y_2)=(0,0,0,0)$ is a Kuhn–Tucker point but not a local minimum. Verify that the point $(x_1,x_2,y_1,y_2)=(3,0,1,5)$ is both a Kuhn–Tucker point and a local minimum. What is the global minimum to the problem?

Notes and References

In this chapter we develop first- and second-order optimality conditions for unconstrained optimization problems. These classical results can be found in most textbooks dealing with real analysis. For more details on this subject and for information regarding the handling of equality constraints via the Lagrangian multiplier rule, refer to Bartle [1976] and Rudin [1964].

In Section 4.2 we treat the problem of minimizing a function in the presence of inequality constraints and develop the Fritz John [1948] necessary optimality conditions. A weaker form of these conditions, in which the nonnegativity of the multipliers was not asserted, was derived by Karush [1939]. Under a suitable constraint qualification, the Lagrangian multiplier associated with the objective function is positive, and the Fritz John conditions reduce to those of Kuhn and Tucker [1951], which were independently derived. Even though the Kuhn–Tucker conditions were originally derived by Karush [1939] using calculus of variations, the work has not received much attention, since it has not been published. An excellent historical review of optimality conditions for nonlinear programming can be found in Kuhn [1976]. The reader may refer to the following references for further study of the Fritz John and Kuhn–Tucker conditions: Abadie [1967b], Avriel [1967], Canon, Cullum, and Polak [1966], Gould and Tolle [1972], Luenberger [1973], Mangasarian [1969a], and Zangwill [1969].

Mangasarian and Fromovitz [1967] generalized the Fritz John conditions to handle both equality and inequality constraints. Their approach used the implicit function theorem. In Section 4.3 we develop the Fritz John conditions for equality and inequality constraints by constructing a feasible arc, as in the work of Fiacco and McCormick [1968].

In Section 4.4 we show that the Kuhn–Tucker conditions are indeed sufficient for optimality under suitable convexity assumptions. This result was proved by Kuhn and Tucker [1951] if the functions f, g_i for $i\in I$ are convex, the functions h_i for all i are affine, and the set X is convex. This result was generalized later, so that weaker convexity assumptions are needed to guarantee optimality, as shown in section 4.4 (see Mangasarian [1969a]). the reader may also refer to Bhatt and Misra [1975], who relaxed the condition that h_i be affine, provided that the associated Lagrangian multiplier has the correct sign.

Other generalizations and extensions of the Fritz John and Kuhn–Tucker conditions were developed by many authors. One such extension is to relax the condition that the set X is open. In this case we obtain necessary optimality conditions of the minimum principle type. For details on this type of optimality conditions, see Bazaraa and Goode [1972], Canon, Cullum, and Polak [1970], and Mangasarian [1969a]. Another extension is to treat the problem in an infinite-dimensional setting. The interested reader may refer to Canon, Cullum, and Polak [1970], Dubovitskii and Milyutin [1965], Guignard [1969], Halkin and Neustadt [1966], Hestenes [1966], Neustadt [1968], and Varaiya [1967]. It is also worth mentioning that several authors have developed second-order optimality conditions for constrained problems. For a thorough study of this topic, see Avriel [1976], Fiacco [1968], Luenberger [1973], McCormick [1967], and Messerli and Polak [1969].

Chapter 5

Constraint Qualifications

In Chapter 4 we considered *Problem P* to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in X$ and $g_i(\mathbf{x}) \leq 0$, $i = 1, \ldots, m$. We obtained the Kuhn–Tucker necessary conditions for optimality by deriving the Fritz John conditions and then asserting that the multiplier associated with the objective function is positive when a constraint qualification is satisfied. In this chapter, we develop the Kuhn–Tucker conditions directly without first deriving the Fritz John conditions. This is done under various constraint qualifications for problems with inequality constraints and for problems with both inequality and equality constraints.

The following is an outline of the chapter.

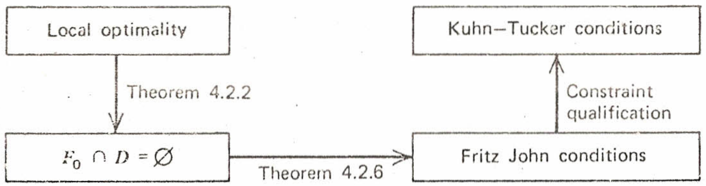
SECTION 5.1: **The Cone of Tangents** We introduce the cone of tangents T and show that $F_0 \cap T = \varnothing$ is a necessary condition for local optimality. Using a constraint qualification, we derive the Kuhn–Tucker conditions directly for problems with inequality constraints.

SECTION 5.2: **Other Constraint Qualifications** We introduce other cones contained in the cone of tangents. Making use of these cones, we present various constraint qualifications that validate the Kuhn–Tucker conditions.

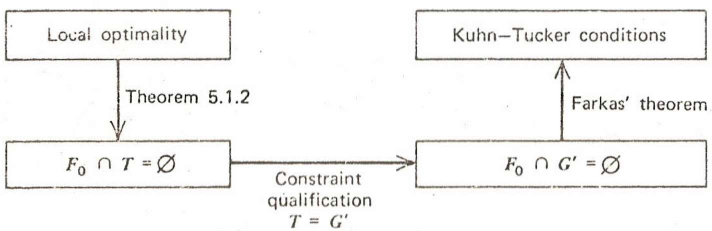
SECTION 5.3: **Problems with Inequality and Equality Constraints** The results of Section 5.2 are extended to problems with equality and inequality constraints.

5.1 The Cone of Tangents

In Section 4.2 we discussed the Kuhn–Tucker necessary optimality conditions for problems with inequality constraints. In particular, we showed that local optimality implies that $F_0 \cap G_0 = \varnothing$, which in turn implies the Fritz John conditions. Under the linear independence constraint qualification, we obtained the Kuhn–Tucker conditions. This process is summarized in the accompanying flowchart. In this section we derive the Kuhn–Tucker conditions



directly without first obtaining the Fritz John conditions. As shown in Theorem 5.1.2 below, a necessary condition for local optimality is that $F_0 \cap T = \varnothing$, where T is the cone of tangents given in Definition 5.1.1 below. Using the constraint qualification $T = G'$, where G' is as defined in Theorem 5.1.3, $F_0 \cap G' = \varnothing$. Using Farkas' theorem, this statement gives the Kuhn–Tucker conditions. This process is summarized in the accompanying flowchart.



5.1.1 Definition

Let S be a nonempty set in E_n, and let $\bar{\mathbf{x}} \in \operatorname{cl} S$. The *cone of tangents* of S at $\bar{\mathbf{x}}$, denoted by T, is the set of all directions $\mathbf{d}$ such that $\mathbf{d} = \lim_{k\to\infty} \lambda_k(\mathbf{x}_k - \bar{\mathbf{x}})$, where $\lambda_k > 0$, $\mathbf{x}_k \in S$ for each k, and $\mathbf{x}_k \to \bar{\mathbf{x}}$.

From the above definition, it is clear that $\mathbf{d}$ belongs to the cone of tangents if there is a feasible sequence $\{\mathbf{x}_k\}$ converging to $\bar{\mathbf{x}}$ such that the directions of the cords $\mathbf{x}_k - \bar{\mathbf{x}}$ converge to $\mathbf{d}$. In Exercise 5.4 we ask the reader to show that the

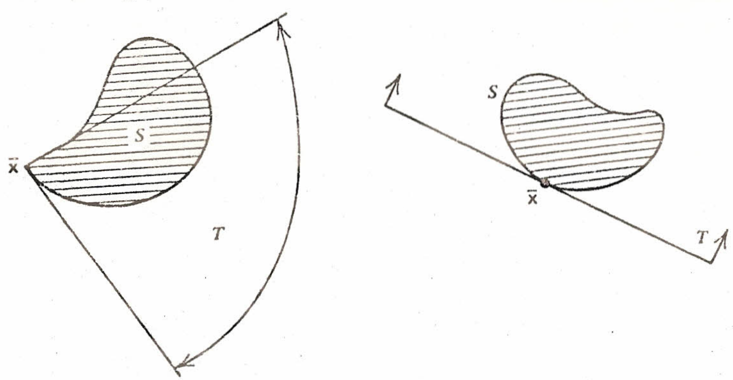


Figure 5.1 Examples of the cone of tangents.

cone of tangents is indeed a closed cone. Figure 5.1 illustrates some examples of the cone of tangents, where the origin is translated to $\bar{\mathbf{x}}$ for convenience.

Theorem 5.1.2 below shows that for a problem of the form: minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in S$, $F_0 \cap T = \varnothing$ is indeed a necessary condition for optimality. Later we specify S to be the set $\{\mathbf{x} \in X : g_i(\mathbf{x}) \le 0 \text{ for } i = 1, \ldots, m\}$

5.1.2 Theorem

Let S be a nonempty set in E_n, and let $\bar{\mathbf{x}} \in S$. Furthermore, suppose that $f: E_n \to E_1$ is differentiable at $\bar{\mathbf{x}}$. If $\bar{\mathbf{x}}$ locally solves the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in S$, then $F_0 \cap T = \varnothing$, where $F_0 = \{\mathbf{d} : \nabla f(\bar{\mathbf{x}})^t\mathbf{d} < 0\}$, and T is the cone of tangents of S at $\bar{\mathbf{x}}$.

Proof

Let $\mathbf{d} \in T$, that is, $\mathbf{d} = \lim_{k\to\infty} \lambda_k(\mathbf{x}_k - \bar{\mathbf{x}})$, where $\lambda_k > 0$, $\mathbf{x}_k \in S$ for each k, and $\mathbf{x}_k \to \bar{\mathbf{x}}$. By differentiability of f at $\bar{\mathbf{x}}$, we get

$$f(\mathbf{x}_k) - f(\bar{\mathbf{x}}) = \nabla f(\bar{\mathbf{x}})^t(\mathbf{x}_k - \bar{\mathbf{x}}) + \|\mathbf{x}_k - \bar{\mathbf{x}}\|\,\alpha(\bar{\mathbf{x}}; \mathbf{x}_k - \bar{\mathbf{x}}) \tag{5.1}$$

where $\alpha(\bar{\mathbf{x}}; \mathbf{x}_k - \bar{\mathbf{x}}) \to 0$ as $\mathbf{x}_k \to \bar{\mathbf{x}}$. Noting the local optimality of $\bar{\mathbf{x}}$, for k large enough, we have $f(\mathbf{x}_k) \ge f(\bar{\mathbf{x}})$, and from (5.1), we get

$$\nabla f(\bar{\mathbf{x}})^t(\mathbf{x}_k - \bar{\mathbf{x}}) + \|\mathbf{x}_k - \bar{\mathbf{x}}\|\,\alpha(\bar{\mathbf{x}}; \mathbf{x}_k - \bar{\mathbf{x}}) \ge 0$$

Multiplying by $\lambda_k > 0$ and taking the limit as $k \to \infty$, the above inequality implies that $\nabla f(\bar{\mathbf{x}})^t\mathbf{d} \ge 0$. So far, we have shown that $\mathbf{d} \in T$ implies that $\nabla f(\bar{\mathbf{x}})^t\mathbf{d} \ge 0$, and hence $F_0 \cap T = \varnothing$, and the proof is complete.

Abadie Constraint Qualification

In Theorem 5.1.3 below we derive the Kuhn–Tucker conditions under the constraint qualification $T = G'$ credited to Abadie.

5.1.3 Theorem (Kuhn–Tucker Necessary Conditions)

Let X be a nonempty set in E_n, and let $f: E_n \to E_1$ and $g_i: E_n \to E_1$ for $i = 1, \ldots, m$. Consider the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in X$, $g_i(\mathbf{x}) \le 0$ for $i = 1, \ldots, m$. Let $\bar{\mathbf{x}}$ be a feasible solution, and let $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$. Suppose that f and g_i for $i \in I$ are differentiable at $\bar{\mathbf{x}}$. Furthermore, suppose that the constraint qualification $T = G'$ is true, where T is the cone of tangents of the feasible region at $\bar{\mathbf{x}}$, and $G' = \{\mathbf{d} : \nabla g_i(\bar{\mathbf{x}})^t\mathbf{d} \le 0 \text{ for } i \in I\}$. If $\bar{\mathbf{x}}$ is a local optimal solution, then there exist nonnegative scalars u_i for $i \in I$ such that

$$\nabla f(\bar{\mathbf{x}}) + \sum_{i\in I} u_i \nabla g_i(\bar{\mathbf{x}}) = \mathbf{0}$$

Proof

By Theorem 5.1.2, $F_0 \cap T = \varnothing$, where $F_0 = \{\mathbf{d} : \nabla f(\bar{\mathbf{x}})^t \mathbf{d} < 0\}$. By assumption, $T = G'$, so that $F_0 \cap G' = \varnothing$. In order words the following system has no

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

Then, by Theorem 2.3.5 (Farkas' theorem), the result follows.

The reader may verify that in Example 4.2.8, the constraint qualification $T = G'$ does not hold true at $\bar{\mathbf{x}} = (1, 0)^t$. Note that the Abadie constraint qualification $T = G'$ could be equivalently stated as $T \supset G'$, since $T \subset G'$ is always true (see Exercise 5.13). Note that openness of the set X and continuity of g_i at $\bar{\mathbf{x}}$ for $i \notin I$ were not explicitly assumed in Theorem 5.1.3. However, without these assumptions, it is unlikely that the constraint qualification $T \supset G'$ would hold true (see Exercise 5.11).

Linearly Constrained Problems

Lemma 5.1.4 below shows that if the constraints are linear, then the Abadie constraint qualification is automatically true. This also implies that the Kuhn–Tucker conditions are always necessary for problems with linear constraints whether the objective function is linear or nonlinear.

5.1.4 Lemma

Let $\mathbf{A}$ be an $m \times n$ matrix, let $\mathbf{b}$ be an m vector, and let $S = \{\mathbf{x} : \mathbf{A}\mathbf{x} \le \mathbf{b}\}$. Suppose $\bar{\mathbf{x}} \in S$ is such that $\mathbf{A}_1\mathbf{x} = \mathbf{b}_1$ and $\mathbf{A}_2\mathbf{x} < \mathbf{b}_2$, where $\mathbf{A}^t = (\mathbf{A}_1^t, \mathbf{A}_2^t)$ and $\mathbf{b}^t = (\mathbf{b}_1^t, \mathbf{b}_2^t)$. Then, $T = G'$, where T is the cone of tangents of S at $\bar{\mathbf{x}}$ and $G' = \{\mathbf{d} : \mathbf{A}_1\mathbf{d} \le \mathbf{0}\}$.

Proof

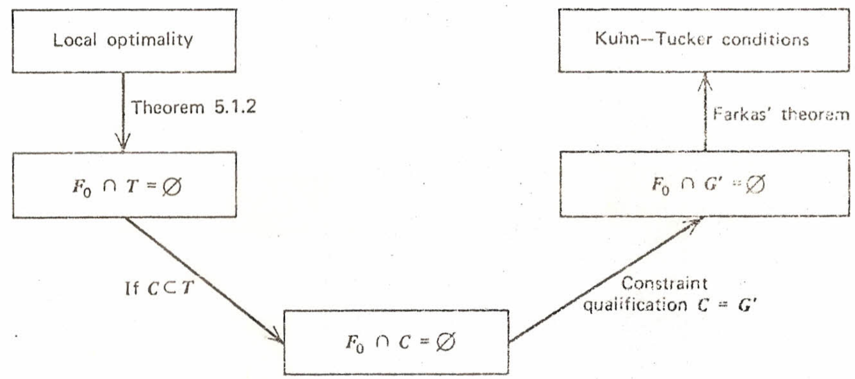
If $\mathbf{A}_1$ is vacuous, $G' = E_n$. Furthermore, $\bar{\mathbf{x}} \in \operatorname{int} S$, and hence $T = E_n$. Thus, $G' = T$. Now suppose that $\mathbf{A}_1$ is not vacuous. Let $\mathbf{d} \in T$, that is, $\mathbf{d} = \lim_{k\to\infty} \lambda_k(\mathbf{x}_k - \bar{\mathbf{x}})$, where $\mathbf{x}_k \in S$ and $\lambda_k > 0$ for each k. Then

$$\mathbf{A}_1(\mathbf{x}_k - \bar{\mathbf{x}}) \le \mathbf{b}_1 - \mathbf{b}_1 = \mathbf{0} \tag{5.2}$$

Multiplying (5.2) by $\lambda_k > 0$ and taking the limit as $k \to \infty$, it follows that $\mathbf{A}_1\mathbf{d} \le \mathbf{0}$. Thus, $\mathbf{d} \in G'$, and $T \subset G'$. Now let $\mathbf{d} \in G'$, that is, $\mathbf{A}_1\mathbf{d} \le \mathbf{0}$. We need to show that $\mathbf{d} \in T$. Since $\mathbf{A}_2\bar{\mathbf{x}} < \mathbf{b}_2$, there is a $\delta > 0$ such that $\mathbf{A}_2(\bar{\mathbf{x}} + \lambda\mathbf{d}) < \mathbf{b}_2$ for all $\lambda \in (0, \delta)$. Furthermore, since $\mathbf{A}_1\bar{\mathbf{x}} = \mathbf{b}_1$ and $\mathbf{A}_1\mathbf{d} \le \mathbf{0}$, then $\mathbf{A}_1(\bar{\mathbf{x}} + \lambda\mathbf{d}) \le \mathbf{b}_1$ for all $\lambda > 0$. Therefore $\bar{\mathbf{x}} + \lambda\mathbf{d} \in S$ for each $\lambda \in (0, \delta)$. This automatically shows that $\mathbf{d} \in T$. Therefore, $T = G'$, and the proof is complete.

5.2 Other Constraint Qualifications

The Kuhn–Tucker conditions have been developed by many authors under various constraint qualifications. In this section we present some of the more important constraint qualifications. In Section 5.1 we learned that local optimality implies that $F_0 \cap T = \varnothing$, and the Kuhn–Tucker conditions follow under the constraint qualification $T = G'$. If we define a cone $C \subset T$, then $F_0 \cap T = \varnothing$ also implies that $F_0 \cap C = \varnothing$. Therefore, any constraint qualification of the form $C = G'$ will lead to the Kuhn–Tucker conditions. this process is illustrated in the accompanying flow chart.



We present below several such cones whose closures are contained in T. Here the feasible region S is given by $\{\mathbf{x} \in X : g_i(\mathbf{x}) \le 0,\ i = 1, 2, \ldots, m\}$. The vector $\bar{\mathbf{x}}$ is a feasible point, and $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$.

The Cone Of Feasible Directions of S at $\bar{\mathbf{x}}$

This cone was introduced earlier in Definition 4.2.1. The cone of feasible directions, denoted by D, is the set of all nonzero vectors $\mathbf{d}$ such that $\bar{\mathbf{x}} + \lambda\mathbf{d} \in S$ for $\lambda \in (0, \delta)$ for some $\delta > 0$.

The Cone of Attainable Directions of S at $\bar{\mathbf{x}}$

A nonzero vector $\mathbf{d}$ belongs to the cone of attainable directions, denoted by A, if there exist a $\delta > 0$ and an $\boldsymbol{\alpha} : E_1 \to E_n$ such that $\boldsymbol{\alpha}(\lambda) \in S$ for $\lambda \in (0, \delta)$, $\boldsymbol{\alpha}(0) = \bar{\mathbf{x}}$, and $\lim_{\lambda\to 0^+} \dfrac{\boldsymbol{\alpha}(\lambda) - \boldsymbol{\alpha}(0)}{\lambda} = \mathbf{d}$. In other words, $\mathbf{d}$ belongs to the cone of attainable directions if there is a feasible arc starting from $\bar{\mathbf{x}}$ that is tangential to $\mathbf{d}$.

The Cone of Interior Directions of S at $\bar{\mathbf{x}}$

This cone, denoted by G_0, was introduced in Section 4.2. More specifically, $G_0 = \{\mathbf{d} : \nabla g_i(\bar{\mathbf{x}})^t\mathbf{d} < 0 \text{ for } i \in I\}$. Note that if X is open and g_i for $i \notin I$ is continuous at $\bar{\mathbf{x}}$, then $\mathbf{d} \in G_0$ implies that $\bar{\mathbf{x}} + \lambda\mathbf{d}$ belongs to the interior of the feasible region for $\lambda > 0$ and sufficiently small.

Lemma 5.2.1 below shows that all the above cones and their closures are contained in T.

5.2.1 Lemma

Let X be a nonempty set in E_n, and let $f: E_n \to E_1$ and $g_i: E_n \to E_1$ for $i = 1, \ldots, m$. Consider the problem to minimize $f(\mathbf{x})$ subject to $g_i(\mathbf{x}) \le 0$ for $i = 1, \ldots, m$ and $\mathbf{x} \in X$. Let $\bar{\mathbf{x}}$ be a feasible point, and let $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$. Suppose that g_i for $i \in I$ is differentiable at $\bar{\mathbf{x}}$, and let $G' = \{\mathbf{d} : \nabla g_i(\bar{\mathbf{x}})^t\mathbf{d} \le 0 \text{ for } i \in I\}$. Then

$$\text{cl } D \subset \text{cl } A \subset T \subset G'$$

where D, A, and T are, respectively, the cone of feasible directions, the cone of attainable directions, and the cone of tangents of the feasible region at $\bar{\mathbf{x}}$. Furthermore, if X is open and g_i for $i \notin I$ is continuous at $\bar{\mathbf{x}}$, then $G_0 \subset D$, so that

$$\text{cl } G_0 \subset \text{cl } D \subset \text{cl } A \subset T \subset G'$$

where G_0 is the cone of interior directions of the feasible region at $\bar{\mathbf{x}}$.

Proof

It can be easily verified that $D \subset A \subset T \subset G'$, and since T is closed (see Exercise 5.4), $\text{cl } D \subset \text{cl } A \subset T \subset G'$. Now note that $G_0 \subset D$, as shown in the proof of Theorem 4.2.3. Hence, the second part of the theorem follows.

We now present some constraint qualifications that validate the Kuhn–Tucker conditions.

Slater Constraint Qualification

The set X is open, g_i for $i \in I$ is pseudoconvex at $\bar{\mathbf{x}}$, g_i for $i \notin I$ is continuous at $\bar{\mathbf{x}}$, and there is an $\mathbf{x} \in X$ such that $g_i(\mathbf{x}) < 0$ for $i \in I$.

Linear Independence Constraint Qualification

The set X is open, g_i for $i \notin I$ is continuous at $\bar{\mathbf{x}}$, and $\nabla g_i(\bar{\mathbf{x}})$ for $i \in I$ are linearly independent.

Cottle Constraint Qualification

The set X is open and g_i for $i \notin I$ is continuous at $\bar{\mathbf{x}}$, and $\text{cl } G_0 = G'$.

Zangwill Constraint Qualification

$$\text{cl } D = G'$$

Kuhn–Tucker Constraint Qualification

$$\text{cl } A = G'$$

The Kuhn–Tucker Conditions

In Theorem 5.1.3 we showed that the Kuhn–Tucker necessary optimality conditions are true under Abadie's constraint qualification $T = G'$. We demonstrate below that all the constraint qualifications discussed above imply that of Abadie, and hence each validate the Kuhn–Tucker necessary conditions. From Lemma 5.2.1, it is clear that Cottle's constraint qualification implies that of Zangwill, which implies that of Kuhn and Tucker, which in turn implies Abadie's qualification. We now show that the first two qualifications imply that of Cottle.

First, suppose that the Slater constraint qualification holds true. Then, there is an $\mathbf{x} \in X$ such that $g_i(\mathbf{x}) < 0$ for $i \in I$. Since $g_i(\mathbf{x}) < 0$ and $g_i(\bar{\mathbf{x}}) = 0$, then by pseudoconvexity of g_i at $\bar{\mathbf{x}}$ it follows that $\nabla g_i(\bar{\mathbf{x}})^t(\mathbf{x} - \bar{\mathbf{x}}) < 0$. Thus $\mathbf{d} = \mathbf{x} - \bar{\mathbf{x}}$ belongs to G_0. Therefore, $G_0 \ne \varnothing$ and the reader can verify that $\text{cl } G_0 = G'$, and hence Cottle's constraint qualification is true. Now suppose that the linear independence constraint qualification is satisfied. Then, $\sum_{i \in I} u_i \nabla g_i(\bar{\mathbf{x}}) = 0$ has no nonzero solution. By Theorem 2.3.9, it follows that there exists a vector $\mathbf{d}$ such that $\nabla g_i(\mathbf{x})^t\mathbf{d} < 0$ for $i \in I$. Thus, $G_0 \ne \varnothing$, and Cottle's qualification holds true.

The relationships among the constraint qualification are illustrated in Figure 5.2.

5.3 Problems with Inequality and Equality Constraints

In this section we consider problems with both inequality and equality constraints. In particular, consider the following problem:

$$\begin{array}{lll} \text{Minimize} & f(\mathbf{x}) & \\ \text{subject to} & g_i(\mathbf{x}) \le 0 & \text{for } i = 1, \ldots, m \\ & h_i(\mathbf{x}) = 0 & \text{for } i = 1, \ldots, l \\ & \mathbf{x} \in X & \end{array}$$

By Theorem 5.1.2, a necessary optimality condition is $F_0 \cap T = \varnothing$. By imposing the constraint qualification $T = G' \cap H_0$, where $H_0 = \{\mathbf{d} : \nabla h_i(\bar{\mathbf{x}})^t\mathbf{d} = 0$

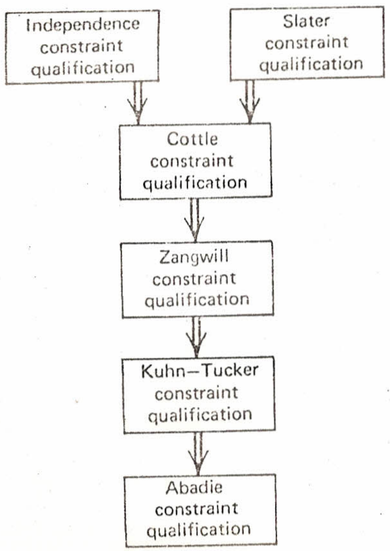
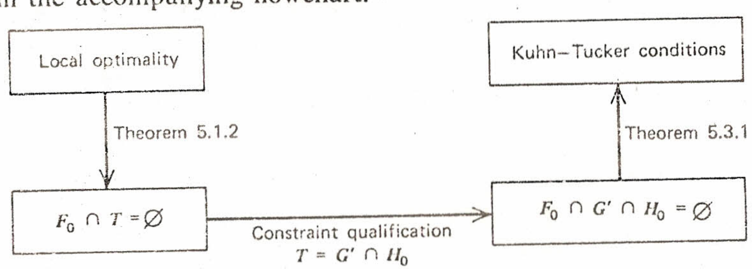


Figure 5.2 Relationships among various constraint qualifications for inequality constrained problems.

for $i=1,\ldots,l\}$, then $F_0 \cap G' \cap H_0 = \varnothing$. By using Farkas' theorem, the Kuhn–Tucker conditions follow from Theorem 5.3.1 below. This process is summarized in the accompanying flowchart.



5.3.1 Theorem (Kuhn–Tucker Conditions)

Let $f: E_n \to E_1$, $g_i: E_n \to E_1$ for $i=1,\ldots,m$ and $h_i: E_n \to E_1$ for $i=1,\ldots,l$, and let X be a nonempty set in E_n. Consider the following problem:

$$\begin{array}{lll} \text{Minimize} & f(\mathbf{x}) & \\ \text{subject to} & g_i(\mathbf{x}) \le 0 & \text{for } i=1,\ldots,m \\ & h_i(\mathbf{x}) = 0 & \text{for } i=1,\ldots,l \\ & \mathbf{x} \in X & \end{array}$$

Let $\bar{\mathbf{x}}$ locally solve the problem, and let $I=\{i: g_i(\bar{\mathbf{x}})=0\}$. Suppose that f, g_i for $i \in I$, and h_i for $i=1,\ldots,l$ are differentiable at $\bar{\mathbf{x}}$. Suppose that the constraint qualification $T = G' \cap H_0$ holds true, where T is the cone of tangents of the feasible region at $\bar{\mathbf{x}}$, and

$$G' = \{\mathbf{d} : \nabla g_i(\bar{\mathbf{x}})' \le 0 \quad \text{for} \quad i \in I\}.$$

$$H_0 = \{\mathbf{d} : \nabla h_i(\bar{\mathbf{x}})'\mathbf{d} = 0 \quad \text{for } i=1,\ldots,l\}$$

Then, $\bar{\mathbf{x}}$ is a Kuhn–Tucker point, that is, there exist scalars $u_i \ge 0$ for $i \in I$ and v_i for $i=1,\ldots,l$ such that

$$\nabla f(\bar{\mathbf{x}}) + \sum_{i \in I} u_i \nabla g_i(\bar{\mathbf{x}}) + \sum_{i=1}^{l} v_i \nabla h_i(\bar{\mathbf{x}}) = \mathbf{0}$$

Proof

Since $\bar{\mathbf{x}}$ solves the problem locally, by Theorem 5.1.2, $F_0 \cap T = \varnothing$. By the constraint qualification, we have $F_0 \cap G' \cap H_0 = \varnothing$, that is, the system $\mathbf{Ad} \le \mathbf{0}$ and $\mathbf{c'd} > 0$ has no solution, where the rows of $\mathbf{A}$ are given by $\nabla g_i(\bar{\mathbf{x}})'$ for $i \in I$, $\nabla h_i(\bar{\mathbf{x}})$ and $-\nabla h_i(\bar{\mathbf{x}})$ for $i=1,\ldots,l$, and $\mathbf{c} = -\nabla f(\bar{\mathbf{x}})$. By Theorem 2.3.5, the system $\mathbf{A'y} = \mathbf{c}$ and $\mathbf{y} \ge \mathbf{0}$ has a solution. This implies that there exist nonnegative scalars u_i for $i \in I$ and α_i, β_i for $i=1,\ldots,l$ such that

$$\nabla f(\bar{\mathbf{x}}) + \sum_{i \in I} u_i \nabla g_i(\bar{\mathbf{x}}) + \sum_{i=1}^{l} \alpha_i \nabla h_i(\bar{\mathbf{x}}) - \sum_{i=1}^{l} \beta_i \nabla h_i(\bar{\mathbf{x}}) = \mathbf{0}$$

Letting $v_i = \alpha_i - \beta_i$ for each i, the result is apparent.

We now present several constraint qualifications that validate the Kuhn–Tucker conditions. These qualifications use several cones that were defined earlier in the chapter. The reader may note that Zangwill's constraint qualification is omitted here, since the cone of feasible directions is usually equal to the zero vector in the presence of nonlinear equality constraints.

Slater Constraint Qualification

The set X is open, g_i for $i \in I$ is pseudoconvex at $\bar{\mathbf{x}}$, g_i for $i \notin I$ is continuous at $\bar{\mathbf{x}}$, h_i for $i=1,\ldots,l$ is quasiconvex, quasiconcave, and continuously differentiable at $\bar{\mathbf{x}}$, and $\nabla h_i(\bar{\mathbf{x}})$ for $i=1,\ldots,l$ are linearly independent. Furthermore, there exists an $\mathbf{x} \in X$ such that $g_i(\mathbf{x}) < 0$ for $i \in I$ and $h_i(\mathbf{x}) = 0$ for $i=1,\ldots,l$.

Linear Independence Constraint Qualification

The set X is open, g_i for $i \notin I$ is continuous at $\bar{\mathbf{x}}$, $\nabla g_i(\bar{\mathbf{x}})$ for $i \in I$ and $\nabla h_i(\bar{\mathbf{x}})$ for $i=1,\ldots,l$ are linearly independent, and h_i for $i=1,\ldots,l$ is continuously differentiable at $\bar{\mathbf{x}}$.

Cottle Constraint Qualification

The set X is open, g_i for $i \notin I$ is continuous at $\bar{\mathbf{x}}$, h_i for $i=1,\ldots,l$ is continuously differentiable at $\bar{\mathbf{x}}$, and $\nabla h_i(\bar{\mathbf{x}})$ for $i=1,\ldots,l$ are linearly independent. Furthermore, $\operatorname{cl}(G_0 \cap H_0) = G' \cap H_0$.

Kuhn–Tucker Constraint Qualification

$$\operatorname{cl} A = G' \cap H_0$$

Abadie Constraint Qualification

$$T = G' \cap H_0$$

The Kuhn–Tucker Conditions

In Theorem 5.3.1, we showed that the Kuhn–Tucker conditions are true if Abadie's constraint qualification $T = G' \cap H_0$ is satisfied. We demonstrate below that all the constraint qualifications given above imply that of Abadie, and hence each validate the Kuhn–Tucker necessary conditions.

As in Lemma 5.2.1, the reader can easily verify that $\operatorname{cl} A \subset T \subset G' \cap H_0$. Now suppose that X is open, g_i for $i \notin I$ is continuous at $\bar{\mathbf{x}}$, h_i for $i=1,\ldots,l$ is continuously differentiable, and $\nabla h_i(\bar{\mathbf{x}})$ for $i=1,\ldots,l$ are linearly independent. From the proof of Theorem 4.3.1, it follows that $G_0 \cap H_0 \subset A$. Thus, $\operatorname{cl}(G_0 \cap H_0) \subset \operatorname{cl} A \subset T \subset G' \cap H_0$. In particular, Cottle's constraint qualification implies that of Kuhn and Tucker, which in turn implies Abadie's constraint qualification.

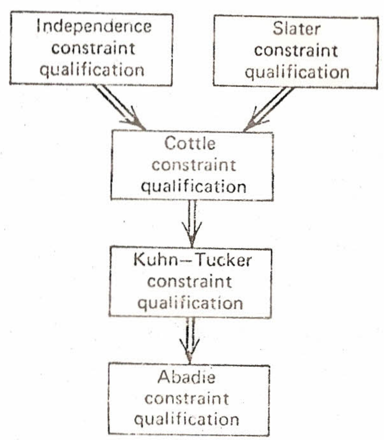


Figure 5.3 Relationships among constraint qualifications for problems with inequality and equality constraints.

We now demonstrate that Slater's constraint qualification and the linear independence constraint qualification imply that of Cottle. Suppose that Slater's qualification is satisfied, so that $g_i(\mathbf{x})<0$ for $i \in I$, $h_i(\mathbf{x})=0$ for $i=1,\ldots,l$ for some $\mathbf{x} \in X$. By pseudoconvexity of g_i at $\bar{\mathbf{x}}$, $\nabla g_i(\bar{\mathbf{x}})^t(\mathbf{x}-\bar{\mathbf{x}})<0$ for $i \in I$. Also since $h_i(\mathbf{x}) = h_i(\bar{\mathbf{x}}) = 0$, quasiconvexity and quasiconcavity of h_i at $\bar{\mathbf{x}}$ imply that $\nabla h_i(\bar{\mathbf{x}})^t(\mathbf{x}-\bar{\mathbf{x}})=0$. Letting $\mathbf{d}=\mathbf{x}-\bar{\mathbf{x}}$, it then follows that $\mathbf{d} \in G_0 \cap H_0$. Thus, $G_0 \cap H_0 \neq \varnothing$, and the reader can verify that $\operatorname{cl}(G_0 \cap H_0) = G' \cap H_0$. Therefore, Cottle's constraint qualification holds true.

Finally, we show that the linear independence qualification implies that of Cottle. By contradiction, suppose that $G_0 \cap H_0 = \varnothing$. Then, using a separation theorem as in the proof of Theorem 4.3.2, it follows that there exists a nonzero vector $(\mathbf{u}_I, \mathbf{v})$ such that $\sum_{i \in I} u_i \nabla g_i(\bar{\mathbf{x}}) + \sum_{i=1}^{l} v_i \nabla h_i(\bar{\mathbf{x}}) = \mathbf{0}$, where $\mathbf{u}_I \geq \mathbf{0}$ is the vector whose ith component is u_i. This contradicts the linear independence assumption. Thus, Cottle constraint qualification holds true.

In Figure 5.3, we summarize the implications of the constraint qualifications discussed above. As mentioned earlier, these implications, together with Theorem 5.3.1, validate the Kuhn–Tucker conditions.

5.1 Find the cone of tangents for each of the following sets at the point $\bar{\mathbf{x}}=(0,0)'$.
a. $S=\{(x_1,x_2): x_2 \geq -x_1^3\}$
b. $S=\{(x_1,x_2): x_1 \text{ is integer}, x_2=0\}$
c. $S=\{(x_1,x_2): x_1 \text{ is rational}, x_2=0\}$

5.2 Let S be a subset of E_n, and let $\bar{\mathbf{x}} \in \text{int } S$. Show that the cone of tangents of S at $\bar{\mathbf{x}}$ is E_n.

5.3 Prove that the cone of tangents of S at $\bar{\mathbf{x}}$ can be equivalently defined as follows:

$$T=\{\mathbf{d}: \mathbf{x}_k=\bar{\mathbf{x}}+\lambda_k\mathbf{d}+\lambda_k\boldsymbol{\alpha}(\lambda_k)\in S \qquad \text{for each } k\}$$

where $\lambda_k>0$ converges to 0, and $\boldsymbol{\alpha}: E_1 \to E_n$ is such that $\boldsymbol{\alpha}(\lambda)\to 0$ as $\lambda \to 0$.

5.4 Prove that the cone of tangents is a closed cone.
Hint: First show that $T=\bigcap_{N\in\mathcal{N}} \text{cl } K(S\cap N, \bar{\mathbf{x}})$, where $K(S\cap N,\bar{\mathbf{x}})=\{\lambda(\mathbf{x}-\bar{\mathbf{x}}): \mathbf{x}\in S\cap N, \lambda>0\}$, and $\mathcal{N}$ is the class of all open neighborhoods about $\bar{\mathbf{x}}$.

5.5 Let $\mathbf{A}$ be an $m\times n$ matrix, and consider the cones $G_0=\{\mathbf{d}: \mathbf{A}\mathbf{d}<\mathbf{0}\}$ and $G'=\{\mathbf{d}: \mathbf{A}\mathbf{d}\leq\mathbf{0}\}$. Prove that
a. G_0 is an open convex cone.
b. G' is a closed convex cone.
c. $G_0=\text{int. } G'$.
d. If $G_0 \neq \emptyset$, then cl $G_0=G'$.

5.6 Consider the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{x}\in X$ and $g_i(\mathbf{x})\leq 0$ for $i=1,\ldots,m$. Let $\bar{\mathbf{x}}$ be a feasible point, and let $I=\{i: g_i(\bar{\mathbf{x}})=0\}$. Suppose that X is open, and g_i for $i\notin I$ is continuous at $\bar{\mathbf{x}}$. Further, suppose that the set

$$\{\mathbf{d}: \nabla g_i(\bar{\mathbf{x}})'\mathbf{d}\leq 0 \quad \text{for } i\in J, \quad \nabla g_i(\bar{\mathbf{x}})'\mathbf{d}<0 \quad \text{for } i\in I-J\}$$

is not empty, where $J=\{i\in I: g_i \text{ is pseudoconcave at } \bar{\mathbf{x}}\}$. Show that this condition is sufficient to validate the Kuhn–Tucker conditions at $\bar{\mathbf{x}}$.
(This is the Arrow–Hurwicz–Uzawa constraint qualification.)

5.7 Consider the problem to minimize $f(\mathbf{x})$ subject to $g_i(\mathbf{x})\leq 0$ for $i=1,\ldots,m$. Let $\bar{\mathbf{x}}$ be feasible, and let $I=\{i: g_i(\bar{\mathbf{x}})=0\}$. Let $(\bar{z},\bar{\mathbf{d}})$ be an optimal solution to the following linear program.

$$\begin{array}{lll} \text{Minimize} & z & \\ \text{subject to} & \nabla f(\bar{\mathbf{x}})'\mathbf{d}-z\leq 0 & \\ & \nabla g_i(\bar{\mathbf{x}})'\mathbf{d}-z\leq 0 & \text{for } i\in I \\ & -1\leq d_j\leq 1 & \text{for } j=1,\ldots,n \end{array}$$

a. Show that the Fritz John conditions hold true if $\bar{z}=0$.
b. Show that if $\bar{z}=0$, then the Kuhn–Tucker conditions hold true under either Slater's or Cottle's constraint qualification.

5.8 For each of the following sets, find the cone of feasible directions and the cone of attainable directions at $\bar{\mathbf{x}}=(0,0)'$.
a. $S=\{(x_1,x_2): -1\leq x_1\leq 1, x_2\geq x_1^{1/3}, x_2\geq x_1\}$
b. $S=\{(x_1,x_2): x_2>x_1^2\}$
c. $S=\{(x_1,x_2): x_2=-x_1^3\}$
d. $S=S_1\cup S_2$, where
$S_1=\{(x_1,x_2): x_1\geq 0, x_2\geq x_1^2\}$, and $S_2=\{(x_1,x_2): x_1\leq 0, -2x_1\leq 3x_2\leq -x_1\}$

5.9 Let $f: E_n\to E_1$ be differentiable at $\bar{\mathbf{x}}$ with a nonzero gradient $\nabla f(\bar{\mathbf{x}})$. Let $S=\{\mathbf{x}: f(\mathbf{x})\geq f(\bar{\mathbf{x}})\}$. Show that the cone of tangents and the cone of attainable directions of S at $\bar{\mathbf{x}}$ are both given by $\{\mathbf{d}: \nabla f(\bar{\mathbf{x}})'\mathbf{d}\geq 0\}$. Does the result hold if $\nabla f(\bar{\mathbf{x}})=\mathbf{0}$? Prove or give a counterexample.

5.10 Consider the following problem:

$$\begin{array}{ll} \text{Minimize} & -x_1 \\ \text{subject to} & x_1^2+x_2^2 \leq 1 \\ & (x_1-1)^3-x_2\leq 0 \end{array}$$

a. Show that the Kuhn–Tucker constraint qualification holds true at $\bar{\mathbf{x}}=(1,0)'$.
b. Show that $\bar{\mathbf{x}}=(1,0)'$ is a Kuhn–Tucker point and also that it is the global optimal solution.

5.11 Consider the problem to minimize $5x-x^2$ subject to $g_1(x)\leq 0$, where $g_1(x)=x$.
a. Verify graphically that $\bar{x}=0$ is the optimal solution.
b. Verify that each of the constraint qualifications discussed in Section 5.2 hold true at $\bar{x}=0$.
c. Verify that the Kuhn–Tucker necessary conditions hold true at $\bar{x}=0$.
Now, suppose that the constraint $g_2(x)\leq 0$ is added to the above problem, where

$$g_2(x)=\begin{cases} -1-x & \text{if } x\geq 0 \\ 1-x & \text{if } x<0 \end{cases}$$

Note that $\bar{x}=0$ is still the optimal solution and that g_2 is discontinuous and nonbinding at $\bar{x}$. Check whether the constraint qualifications discussed in Section 5.2 and whether the Kuhn–Tucker conditions hold true at $\bar{x}$.
(This exercise illustrates the need of the continuity assumption of the nonbinding constraints.)

5.12 Consider the feasible region $S=\{\mathbf{x}\in X: g_1(\mathbf{x})\leq 0\}$, where $g_1(\mathbf{x})=x_1^2+x_2^2-1$, and X is the collection of all convex combinations of the four points $(-1,0)'$, $(0,1)'$, $(1,0)'$, and $(0,-1)'$.
a. Find the cone of tangents T of S at $\bar{\mathbf{x}}=(1,0)'$.
b. Check whether $T\supset G'$, where $G'=\{\mathbf{d}: \nabla g_1(\bar{\mathbf{x}})'\mathbf{d}\leq 0\}$.
c. Replace the set X by four inequality constraints. Repeat parts a and b, where $G'=\{\mathbf{d}: \nabla g_i(\bar{\mathbf{x}})'\mathbf{d}\leq 0 \text{ for } i\in I\}$, and I is the new set of binding constraints at $\bar{\mathbf{x}}=(1,0)'$.

5.13 Let $S=\{\mathbf{x}\in X: g_i(\mathbf{x})\leq 0 \text{ for } i=1,\ldots,m\}$. Let $\bar{\mathbf{x}}\in S$, and let $I=\{i: g_i(\bar{\mathbf{x}})=0\}$. Show that $T\subset G'$, where T is the cone of tangents of S at $\bar{\mathbf{x}}$, and $G'=\{\mathbf{d}: \nabla g_i(\bar{\mathbf{x}})'\mathbf{d}\leq 0 \text{ for } i\in I\}$.

5.14 Let $S=\{\mathbf{x}\in X: g_i(\mathbf{x})\leq 0 \text{ for } i=1,\ldots,m \text{ and } h_i(\mathbf{x})=0 \text{ for } i=1,\ldots,l\}$. Let $\bar{\mathbf{x}}\in S$, and let $I=\{i: g_i(\bar{\mathbf{x}})=0\}$. Show that $T\subset G'\cap H_0$, where T is the cone of tangents of S at $\bar{\mathbf{x}}$, $G'=\{\mathbf{d}: \nabla g_i(\bar{\mathbf{x}})'\mathbf{d}\leq 0 \text{ for } i\in I\}$, and $H_0=\{\mathbf{d}: \nabla h_i(\bar{\mathbf{x}})'\mathbf{d}=0 \text{ for } i=1,\ldots,l\}$.

5.15 Consider the constraints $\mathbf{C}\mathbf{d}\leq\mathbf{0}$ and $\mathbf{d}'\mathbf{d}\leq 1$. Let $\bar{\mathbf{d}}$ be a feasible solution such that $\bar{\mathbf{d}}'\bar{\mathbf{d}}=1$, $\mathbf{C}_1\bar{\mathbf{d}}=\mathbf{0}$, and $\mathbf{C}_2\bar{\mathbf{d}}<\mathbf{0}$, where $\mathbf{C}'=(\mathbf{C}_1', \mathbf{C}_2')$. Show that the constraint qualification $T=G_1=\{\mathbf{d}: \mathbf{C}_1\mathbf{d}\leq\mathbf{0}, \mathbf{d}'\bar{\mathbf{d}}\leq 0\}$ holds true, where T is the cone of tangents of the constraint set at $\bar{\mathbf{d}}$.