Chapter 2
Nonlinear Programming and Discrete-Time Optimal Control

The primary intent of this chapter is to introduce the reader to the theoretical foundations of nonlinear programming as well as the theoretical foundations of deterministic discrete-time optimal control. In fact, deterministic discrete-time optimal control problems, as we shall see, are actually nonlinear mathematical programs with a very particular type of structure. In a later chapter, we will also discover that deterministic continuous-time optimal control problems are specific instances of mathematical programs in topological vector spaces. Consequently, it is imperative for the student of optimal control to have a command of the foundations of nonlinear programming. Particularly important are the notions of local and global optimality in mathematical programming, the Kuhn-Tucker necessary conditions for optimality in nonlinear programming, and the role played by convexity in making necessary conditions sufficient. Readers already comfortable with finite-dimensional nonlinear programming may wish to go immediately to Section 2.9. We do caution, however, that subsequent chapters of this book assume substantial familiarity with finite-dimensional nonlinear programming, so that an overestimate of one's nonlinear programming knowledge can be very detrimental to ultimately obtaining a deep understanding of optimal control theory and differential games.

The following is an outline of the principal topics covered in this chapter:

Section 2.1: Nonlinear Program Defined. A formal definition of a finite-dimensional nonlinear mathematical program, with a single criterion and both equality and inequality constraints, is given.

Section 2.2: Other Types of Mathematical Programs. Definitions of linear, interger and mixed integer mathematical programs are provided.

Section 2.3: Necessary Conditions for an Unconstrained Minimum. We derive necessary conditions for a minimum of a twice continuously differentiable function when there are no constraints.

Section 2.4: Necessary Conditions for a Constrained Minimum. Relying on geometric reasoning, the Kuhn-Tucker conditions, as well as the notion of a constraint qualification, are introduced.

T.L. Friesz, *Dynamic Optimization and Differential Games*, International Series in Operations Research & Management Science 135,
DOI 10.1007/978-0-387-72778-3_2,

Section 2.5: Formal Derivation of the Kuhn-Tucker Conditions. A formal derivation of the Kuhn-Tucker necessary conditions, employing a conic definition of optimality and theorems of the alternative, is provided.

Section 2.6: Sufficiency, Convexity, and Uniqueness. We provide formal definitions of a convex set and a convex function. Then we show formally how those notions influence sufficiency and uniqueness of a global minimum.

Section 2.7: Generalized Convexity and Sufficiency. We extend the notion of convexity to include quasiconvexity and pseudoconvexity; we then show how these extensions may be used to state less restrictive conditions assuring optimality.

Section 2.8: Numerical and Graphical Examples. We provide numerical and graphical examples that illustrate the abstract optimality conditions introduced in previous sections of this chapter.

Section 2.9: Discrete-Time Optimal Control. We use the necessary conditions for nonlinear programs to derive the so-called minimum principle for discrete-time optimal control and associated necessary conditions.

2.1 Nonlinear Program Defined

We are presently interested in a type of optimization problem known as a finite-dimensional mathematical program, namely: find a vector $x \in \Re^n$ that satisfies

$$\left.\begin{array}{rl} \min & f(x) \\ \text{s.t.} & h(x) = 0 \\ & g(x) \leq 0 \end{array}\right\} \tag{2.1}$$

where

$$\begin{aligned} x &= (x_1, \ldots, x_n)^T \in \Re^n \\ f(\cdot) &: \Re^n \to \Re^1 \\ g(x) &= (g_1(x), \ldots, g_m(x))^T : \Re^n \to \Re^m \\ h(x) &= (h_1(x), \ldots, h_q(x))^T : \Re^n \to \Re^q \end{aligned}$$

We call the x_i for $i \in \{1, 2, \ldots, n\}$ decision variables, $f(x)$ the objective function, $h(x) = 0$ the equality constraints and $g(x) \leq 0$ the inequality constraints. Becuase the objective and constraint functions will in general be nonlinear, we shall consider (2.1) to be our canonical form of a nonlinear mathematical program (NLP). The *feasible region* for (2.1) is

$$X \equiv \{x : g(x) \leq 0,\ h(x) = 0\} \subset \Re^n \tag{2.2}$$

which allows us to state (2.1) in the form

$$\left.\begin{array}{l}\min\ f(x)\\ \text{s.t.}\ \ x \in X\end{array}\right\} \tag{2.3}$$

The pertinent definitions of optimality for NLP are:

Definition 2.1. *Global minimum. Suppose $x^* \in X$ and $f(x^*) \le f(x)$ for all $x \in X$. Then $f(x)$ achieves a global minimum on X at x^*, and we say x^* is a global minimizer of $f(x)$ on X.*

Definition 2.2. *Local minimum. Suppose $x^* \in X$ and there exists an $\epsilon > 0$ such that $f(x^*) \le f(x)$ for all $x \in [N_\epsilon(x^*) \cap X]$, where $N_\epsilon(x^*)$ is a ball of radius $\epsilon > 0$ centered at x^*. Then $f(x)$ achieves a local minimum on X at x^*, and we say x^* is a local minimizer of $f(x)$.*

In practice, we will often relax the formal terminology of Definition 2.1 and Definition 2.2 and refer to x^* as a global minimum or a local minimum, respectively.

2.2 Other Types of Mathematical Programs

We note that the general form of a continuous mathematical program (MP) may be specialized to create various types of mathematical programs that have been studied in depth. In particular, if the objective function and all constraint functions are linear, (2.1) is called a *linear program* (LP). In such cases, we normally add slack/surplus variables to the inequality constraints to convert them into equality constraints. That is, if we have the constraint

$$g_i(x) \le 0 \tag{2.4}$$

we convert it into

$$g_i(x) + s_i = 0 \tag{2.5}$$

and solve for both x and s_i. The variable s_i is called a *slack variable* and obeys

$$s_i \ge 0 \tag{2.6}$$

If we have an inequality constraint of the form

$$g_j(x) \ge 0 \tag{2.7}$$

we convert it to the form

$$g_j(x) - s_j = 0 \tag{2.8}$$

where

$$s_j \ge 0 \tag{2.9}$$

is called a *surplus variable*. Thus, we take can convert any problem with inequality constraints into one that has only equality constraints and non-negativity restrictions. So without loss of generality, we take the canonical form of the linear programming problem to be

$$\begin{array}{ll} \min & \sum_{i=1}^{n} c_i x_i \\ \text{s.t.} & \sum_{j=1}^{n} a_{ij} x_j = b_i \quad i = 1, \ldots, m \\ & U_j \geq x_j \geq L_j \qquad j = 1, \ldots, n \\ & x \in \Re^n \end{array} . \tag{2.10}$$

where $n > m$. This problem can be re-stated further, using matrix and vector notation, as

$$\left. \begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b \\ & U \geq x \geq L \\ & x \in \Re^n \end{array} \right\} \text{LP} \tag{2.11}$$

where $c \in \Re^n$, $b \in \Re^n$, and $A \in \Re^{m \times n}$.

If the objective function and/or some of the constraints are nonlinear, (2.1) is called a nonlinear program (NLP) and is written as:

$$\left. \begin{array}{ll} \min & f(x) \\ \text{s.t.} & g_i(x) \leq 0 \quad i = 1, \ldots, m \\ & h_i(x) = 0 \quad i = 1, \ldots, q \\ & x \in \Re^n \end{array} \right\} \text{NLP} \tag{2.12}$$

If all of the elements of x are restricted to be a subset of the integers and I^n denotes the integer real numbers, the resulting program

$$\left. \begin{array}{ll} \min & f(x) \\ \text{s.t.} & g_i(x) \leq 0 \quad i = 1, \ldots, m \\ & h_i(x) = 0 \quad i = 1, \ldots, q \\ & x \in I^n \end{array} \right\} \text{IP} \tag{2.13}$$

is called an integer program (IP). If there are two classes of variables, some that are continuous and some that are integer, as in

$$\left.\begin{array}{lll} \min & f(x,y) & \\ \text{s.t.} & g_i(x,y) \leq 0 & i = 1,\ldots,m \\ & h_i(x,y) = 0 & i = 1,\ldots,q \\ & x \in \Re^n & y \in I^n \end{array}\right\} \text{MIP}, \tag{2.14}$$

the problem is known as a mixed integer program (MIP).

2.3 Necessary Conditions for an Unconstrained Minimum

Necessary conditions for optimality in the mathematical program (2.1) are systems of equalities and inequalities that must hold at an optimal solution $x^* \in X$. Any such condition has the logical structure:

If x^* is optimal, then some property $\mathbf{P}(x^*)$ is true.

Necessary conditions play a central role in the analysis of most mathematical programming models and algorithms. Understanding them is also extremely important to understanding the theory of optimal control, even when considering problems in the infinite-dimensional vector spaces associated with continuous-time optimization. This is because the optimal control necessary condition known as the *minimum principle* requires solution of a finite-dimensional nonlinear program.

We begin our discussion of necessary conditions for mathematical programs by considering a special case of the general finite-dimensional mathematical program introduced in the previous section. In particular, we want to state and prove the following result for mathematical programs without constraints:

Theorem 2.1. *Necessary conditions for an unconstrained minimum. Suppose* $f : \Re^n \longrightarrow \Re^1$ *is twice continuously differentiable for all* $x \in \Re^n$. *Then necessary conditions for* $x^* \in \Re^n$ *to be a local or global minimum of* $\min f(x)$ *s.t.* $x \in \Re^n$ *are*

$$\nabla f\left(x^*\right) = 0 \tag{2.15}$$

$$\nabla^2 f\left(x^*\right) \equiv \left(\frac{\partial^2 f\left(x^*\right)}{\partial x_i \partial x_j}\right) \textit{ must be positive semidefinite} \tag{2.16}$$

That is, the gradient vanishes and the Hessian is positive semidefinite matrix at the minimum of interest.

Proof. Since $f(.)$ is twice continuously differentiable, we may make a Taylor series expansion in the vicinity of $x^* \in \Re^n$, a local minimum:

$$f(x) = f(x^*) + [\nabla f(x^*)]^T (x - x^*) + \frac{1}{2}(x - x^*)^T \nabla^2 f(x^*)(x - x^*) + \|x - x^*\|^2 \mathcal{O}(x - x^*)$$

where $\mathcal{O}(x - x^*) \longrightarrow 0$ as $x \longrightarrow x^*$. If $\nabla f(x^*) \neq 0$, then by picking $x = x^* - \theta \nabla f(x^*)$ we can make $f(x) < f(x^*)$ for sufficiently small $\theta > 0$ and, thereby, directly contradict the fact that x^* is a local minimum. It follows that condition (2.15) is necessary, and we may write

$$f(x) = f(x^*) + \frac{1}{2}(x - x^*)^T \nabla^2 f(x^*)(x - x^*) + \|x - x^*\|^2 \mathcal{O}(x - x^*)$$

If the matrix $\nabla^2 f(x^*)$ is not positive semidefinite, there must exist a direction vector $d \in \Re^n$ such that $d \neq 0$ and $d^T \nabla^2 f(x^*) d < 0$. If we now choose $x = x^* + \theta d$, it is possible for sufficiently small $\theta > 0$ to realize $f(x) < f(x^*)$ in direct contradiction of the fact x^* is a local minimum. ■

2.4 Necessary Conditions for a Constrained Minimum

We comment that necessary conditions for constrained programs have the same logical structure as necessary conditions for unconstrained programs introduced in Section 8.4.4; namely:

If x^* is optimal, then some property $\mathbf{P}(x^*)$ is true.

For constrained programs, we will shortly find that $\mathbf{P}(x^*)$ is either the so-called Fritz John conditions or the Kuhn-Tucker conditions. We now turn to the task of providing an informal motivation of the Fritz John conditions, which are the pertinent necessary conditions for the case when no constraint qualification is imposed.

2.4.1 The Fritz John Conditions

The fundamental theorem on necessary conditions is:

Theorem 2.2. *Fritz John conditions. Let x^* be a (global or local) minimum of*

$$\begin{aligned} &\min\ f(x) \\ &\text{s.t. } x \in \mathcal{F} = \{x \in X_0 : g(x) \leq 0, h(x) = 0\} \subset \Re^n \end{aligned}$$

where X_0 is a nonempty open set, $g : \Re^n \longrightarrow \Re^m$, and $h : \Re^n \longrightarrow \Re^q$. Assume that $f(x)$, $g_i(x)$ for $i \in [1, m]$ and $h_i(x)$ for $i \in [1, q]$ have continuous first *derivatives everywhere on X. Then there must exist multipliers $\mu_0 \in \Re^1_+$, $\mu = (\mu_1, \ldots, \mu_m)^T \in \Re^m_+$, and $\lambda^* = (\lambda_i^*, \ldots ; \lambda_q^*)^T \in \Re^q$ such that*

$$\mu_0 \nabla f(x^*) + \sum_{i=1}^m \mu_i \nabla g_i(x^*) + \sum_{i=1}^q \lambda_i \nabla h_i(x^*) = 0 \tag{2.17}$$

$$\mu_i g_i(x^*) = 0 \quad \forall\ i \in [1, m] \tag{2.18}$$

$$\mu_i \geq 0 \quad \forall\ i \in [0, m] \tag{2.19}$$

$$(\mu_0,\ \mu,\ \lambda) \neq 0 \in \Re^{m+q+1} \tag{2.20}$$

Conditions (2.17), (2.18), (2.19), and (2.20) together with $h(x) = 0$ and $g(x) \leq 0$ are called the Fritz John conditions. We will give a formal proof of their validity in Section 2.5.3. For now our focus is on how the Fritz John conditions are related to the Kuhn-Tucker conditions, which are the chief applied notion of a necessary condition for optimality in mathematical programming.

2.4.2 Geometry of the Kuhn-Tucker Conditions

Under certain regularity conditions called constraint qualifications, we may be certain that $\mu_0 \neq 0$. In that case, without loss of generality, we may take $\mu_0 = 1$. When $\mu_0 = 1$, the Fritz John conditions are called the Kuhn-Tucker conditions and (2.17) is called the Kuhn-Tucker identity. In either case, (2.18) and (2.19) together are called the complementary slackness conditions. Sometimes it is convenient to define the Lagrangean function:

$$L(x,\lambda, \mu_0, \mu) \equiv \mu_0 f(x) + \lambda^T h(x) + \mu^T g(x) \tag{2.21}$$

By virtue of this definition, identity (2.17) can be expressed as

$$\nabla_x L(x^*, \lambda, \mu_0, \mu) = 0 \tag{2.22}$$

At the same time (2.18) and (2.19) can be written as

$$\mu^T g(x^*) = 0 \tag{2.23}$$

$$\mu \geq 0 \tag{2.24}$$

Furthermore, we may give a geometrical motivation for the Kuhn-Tucker conditions by considering the following abstract problem with two decision variables and two inequality constraints:

$$\left.\begin{array}{rl} \min & f(x_1, x_2) \\ \text{s.t.} & g_1(x_1, x_2) \leq 0 \\ & g_2(x_1, x_2) \leq 0 \end{array}\right\} \tag{2.25}$$

The functions $f(.)$, $g_1(.)$, and $g_2(.)$ are assumed to be such that the following are true:

1. all functions are differentiable;
2. the feasible region $X \equiv \{(x_1, x_2) : g_1(x_1, x_2) \leq 0,\ g_2(x_1, x_2) \leq 0\}$ is a convex set;
3. all level sets $S_k \equiv \{(x_1, x_2) : f(x_1, x_2) \leq f_k\}$ are convex, where $f_k \in [\alpha, +\infty) \subset \Re^1_+$ is a constant and α is the unconstrained minimum of $f(x_1, x_2)$; and
4. the level curves

$$C_k = \left\{(x_1, x_2) : f(x_1, x_2) = f_k \in \Re^1\right\}$$

for the ordering

$$f_0 < f_1 < f_2 < \cdots < f_k$$

do not cross one another, and C_k is the locus of points for which the objective function has the constant value f_k.

Figure 2.1 is one realization of the above stipulations. Note that there is an uncountable number of level curves and level sets since f_k may be any real number from the interval $[\alpha, +\infty) \subset \Re^1_+$. In Figure 2.1, because the gradient of any function points in the direction of maximal increase of the function, we see there is a $\mu_1 \in \Re^1_{++}$ such that

$$\nabla f(x_1^*, x_2^*) = -\mu_1 \nabla g_1(x_1^*, x_2^*), \tag{2.26}$$

where (x_1^*, x_2^*) is the optimal solution formed by the tangency of $g_1(x_1^*, x_2^*) = 0$ with the level curve $f(x_1^*, x_2^*) = f_2$. Evidently, this observation leads directly to

$$\nabla f(x_1^*, x_2^*) + \mu_1 \nabla g_1(x_1^*, x_2^*) + \mu_2 \nabla g_2(x_1^*, x_2^*) = 0 \tag{2.27}$$

$$\mu_1 g_1(x_1^*, x_2^*) = 0 \tag{2.28}$$

$$\mu_2 g_2(x_1^*, x_2^*) = 0 \tag{2.29}$$

$$\mu_1, \mu_2 \geq 0, \tag{2.30}$$

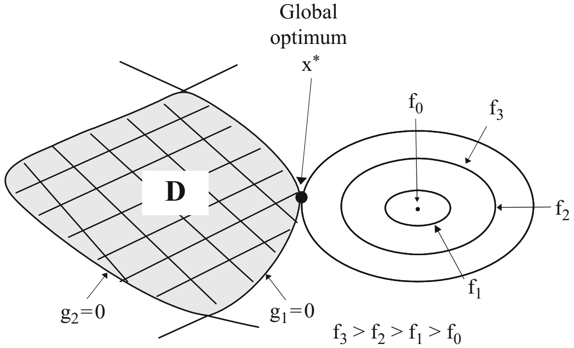

Fig. 2.1 Geometry of an optimal solution

Note that $g_1(x_1^*, x_2^*) = 0$ allows us to conclude that (2.28) holds even though $\mu_1 > 0$. Similarly, (2.26) implies that $\mu_2 = 0$, so (2.29) holds even though $g_2(x_1^*, x_2^*) \neq 0$. Clearly, the nonnegativity conditions (2.30) also hold. By inspection, (2.27), (2.28), (2.29), and (2.30) are the Kuhn-Tucker conditions (Fritz John conditions with $\mu_0 = 1$) for the mathematical program (2.25).

2.4.3 The Lagrange Multiplier Rule

We wish to give a statement of a particular instance of the Kuhn-Tucker theorem on necessary conditions for mathematical programming problems, together with some informal remarks about why that theorem holds when a constraint qualification is satisfied. Since our informal motivation of the Kuhn-Tucker conditions in the next section depends on the Lagrange multiplier rule (LMR) for mathematical programs with equality constraints, we must first state and motivate the LMR. To that end, take x and y to be scalars and $F(x, y)$ and $h(x, y)$ to be scalar functions. Consider the following mathematical program with two decision variables and a single equality constraint:

$$\left.\begin{array}{ll} \min & F(x, y) \\ \text{s.t.} & h(x, y) = 0 \end{array}\right\} \tag{2.31}$$

Assume that $h(x, y) = 0$ may be manipulated to find x in terms of y. That is, we know

$$x = H(y) \tag{2.32}$$

so that

$$F(x, y) = F[H(y), y] \equiv \Phi(y) \tag{2.33}$$

and (2.31) may be thought of as the one-dimensional unconstrained problem

$$\min_y \Phi(y) \tag{2.34}$$

which has the apparent necessary condition

$$\frac{d\Phi(y)}{dy} = 0 \tag{2.35}$$

By the chain rule we have the alternative form

$$\frac{d\Phi(y)}{dy} = \frac{\partial F(H, y)}{\partial y} + \frac{\partial F(H, y)}{\partial H}\frac{\partial H}{\partial y} = 0 \tag{2.36}$$

Applying the chain rule to the equality constraint $h(x, y) = 0$ leads to

$$dh(x, y) = \frac{\partial h}{\partial x}dx + \frac{\partial h}{\partial y}dy = 0 \tag{2.37}$$

from which we obtain

$$\frac{\partial x}{\partial y} = (-1)\frac{\partial h/\partial y}{\partial h/\partial x} \tag{2.38}$$

The necessary condition (2.36), with the help of (2.32) and (2.38), becomes

$$\begin{aligned}\frac{\partial F}{\partial y} + \frac{\partial F}{\partial x}\frac{\partial x}{\partial y} &= \frac{\partial F}{\partial y} + \frac{\partial F}{\partial x}(-1)\frac{\partial h/\partial y}{\partial h/\partial x} \\ &= \frac{\partial F}{\partial y} + (-1)\,\frac{\partial F/\partial x}{\partial h/\partial x}\frac{\partial h}{\partial y} \\ &= \frac{\partial F}{\partial y} + \lambda\frac{\partial h}{\partial y} = 0 \end{aligned} \tag{2.39}$$

where we have defined the Lagrange multiplier to be

$$\lambda = (-1)\,\frac{\partial F/\partial x}{\partial h/\partial x} \tag{2.40}$$

The LMR consists of (2.39) and (2.40), which we restate as

$$\frac{\partial F}{\partial x} + \lambda\frac{\partial h}{\partial x} = 0 \tag{2.41}$$

$$\frac{\partial F}{\partial y} + \lambda\frac{\partial h}{\partial y} = 0 \tag{2.42}$$

Recognizing that the generalization of (2.41) and (2.42) involves Jacobian matrices, we are not surprised to find that, for the equality constrained mathematical program

$$\begin{aligned} &\min \quad f(x) \\ &\text{s.t.} \quad h(x) = 0 \end{aligned}$$

where $x \in \Re^n$ and $h \in \Re^q$, the following result holds:

Theorem 2.3. *Lagrange multiplier rule. Let $x^* \in \Re^n$ be any local maximum or minimum of $f(x)$ subject to the constraints $h_i(x) = 0$ for $i \in [1, q]$, where $x \in \Re^n$ and $q < n$. If it is possible to choose a set of q variables for which the Jacobian*

$$J[h(x^*)] \equiv \begin{bmatrix} \frac{\partial h_1(x^*)}{\partial x_1} & \cdots & \frac{\partial h_1(x^*)}{\partial x_q} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_q(x^*)}{\partial x_1} & \cdots & \frac{\partial h_q(x^*)}{\partial x_q} \end{bmatrix} \tag{2.43}$$

has an inverse, then there exists a unique vector of Lagrange multipliers $\lambda = (\lambda_1, \ldots, \lambda_m)^T$ *satisfying*

$$\frac{\partial f(x^*)}{\partial x_j} + \sum_{i=1}^{q} \lambda_i \frac{\partial h_i(x^*)}{\partial x_j} = 0 \quad j \in [1, n] \tag{2.44}$$

The formal proof of this classical result is contained in most texts on advanced calculus. Note that (2.44) is a necessary condition for optimality.

2.4.4 Motivating the Kuhn-Tucker Conditions

We now wish, using the Lagrange multiplier rule, to establish that the Kuhn-Tucker conditions are valid when an appropriate constraint qualification holds. In fact we wish to consider the following result:

Theorem 2.4. *Kuhn-Tucker conditions. Let* $x^* \in X$ *be a local minimum of*

$$\begin{aligned} &min \quad f(x) \\ &s.t. \quad x \in X = \{x \in X_0 : g(x) \leq 0, h(x) = 0\} \subset \Re^n \end{aligned}$$

where X_0 *is a nonempty open set. Assume that* $f(x)$, $g_i(x)$ *for* $i \in [1, m]$ *and* $h_i(x)$ *for* $i \in [1, q]$ *have continuous first derivatives everywhere on* X *and that a constraint qualification holds. Then there must exist multipliers* $\mu = (\mu_1, \ldots, \mu_m)^T \in \Re^q$ *and* $\lambda^* = (\lambda_i^*, \ldots, \lambda_q^*)^T \in \Re^m$ *such that*

$$\nabla f(x^*) + \textstyle\sum_{i=1}^{m} \mu_i \nabla g_i(x^*) + \sum_{i=1}^{q} \lambda_i \nabla h_i(x^*) = 0 \tag{2.45}$$

$$\mu_i g_i(x^*) = 0 \quad \forall\, i \in [1, m] \tag{2.46}$$

$$\mu_i \geq 0 \quad \forall\, i \in [1, m] \tag{2.47}$$

Expression (2.45) is the Kuhn-Tucker identity and conditions (2.46) and (2.47), as we have indicated previously, are together referred to as the complementary slackness conditions. Do not fail to note that the Kuhn-Tucker conditions are necessary conditions. A solution of the Kuhn-Tucker conditions, without further information, is only a candidate optimal solution, sometimes referred to as a "Kuhn-Tucker point." In fact, it is possible for a particular Kuhn-Tucker point not to be an optimal solution.

We may informally motivate Theorem 2.4 using the Lagrange multiplier rule. This is done by first positing the existence of variables s_i, unrestricted in sign, for $i \in [1, m]$ such that

$$g_i(x^*) + (s_i)^2 = 0 \quad \forall\, i \in [1, m] \tag{2.48}$$

so that the mathematical program (2.1) may be viewed as one with only equality constraints, namely

$$\left.\begin{array}{ll} \min & f(x) \\ \text{s.t.} & h(x) = 0 \\ & g(x) + \operatorname{diag}(s) \cdot s = 0 \end{array}\right\} \tag{2.49}$$

where $s \in \Re^m$ and

$$\operatorname{diag}(s) \equiv \begin{pmatrix} s_1 & 0 & \cdots & 0 & 0 \\ 0 & s_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & s_{m-1} & 0 \\ 0 & 0 & \cdots & 0 & s_m \end{pmatrix} \tag{2.50}$$

To form the necessary conditions for this mathematical program, we first construct the Lagrangian

$$\begin{aligned} L(x,s,\lambda,\mu) &= f(x) + \lambda^T h(x) + \mu^T [g(x) + \operatorname{diag}(s) \cdot s] \\ &= f(x) + \sum_{i=1}^{q} \lambda_i h_i(x) + \sum_{i=1}^{m} \mu_i \left[g_i(x) + s_i^2\right] \end{aligned} \tag{2.51}$$

and then state, using the LMR, the first-order conditions

$$\frac{\partial L(x,s,\lambda,\mu)}{\partial x_i} = \frac{\partial f(x)}{\partial x_i} + \sum_{j=1}^{q} \lambda_j \frac{\partial h_j(x)}{\partial x_i} + \sum_{j=1}^{m} \mu_j \frac{\partial g_j(x)}{\partial x_i} = 0 \quad i \in [1,n] \tag{2.52}$$

$$\frac{\partial L(x,s,\lambda,\mu)}{\partial s_i} = 2\mu_i s_i = 0 \quad i \in [1,m] \tag{2.53}$$

Result (2.52) is of course the Kuhn-Tucker identity (2.45). Note further that both sides of (2.53) may be multiplied by $-s_i$ to obtain the equivalent conditions

$$\mu_i\left(-s_i^2\right) = 0 \quad i \in [1,m] \tag{2.54}$$

which can be restated using (2.48) as

$$\mu_i g_i(x) = 0 \quad i \in [1,m] \tag{2.55}$$

Conditions (2.55) are of course the complementary slackness conditions (2.46).

It remains for us to establish that the inequality constraint multipliers μ_i for $i \in [1,m]$ are nonnegative. To that end, we imagine a perturbation of the inequality constraints by the vector

$$\varepsilon = \left(\varepsilon_1 \; \varepsilon_2 \cdots \varepsilon_m\right)^T \in \Re^m_{++},$$

so that the inequality constraints become

$$g(x) + \text{diag}\,(s) \cdot s = \varepsilon$$

or

$$g_i\,(x) + s_i^2 - \varepsilon_i = 0 \quad i \in [1, m] \tag{2.56}$$

There is an optimal solution for each vector of perturbations, which we call $x\,(\varepsilon)$ where $x^* = x\,(0)$ is the unperturbed optimal solution. As a consequence there is an optimal objective function value

$$Z\,(\varepsilon) \equiv f\left[x^*\,(\varepsilon)\right] \tag{2.57}$$

for each $x^*\,(\varepsilon)$. We note that

$$\frac{\partial Z\,(\varepsilon)}{\partial \varepsilon_i} = \sum_{j=1}^{n} \frac{\partial f\,(x)}{\partial x_j} \frac{\partial x_j\,(\varepsilon)}{\partial \varepsilon_i} \tag{2.58}$$

by the chain rule. Similarly for $k \in [1, m]$

$$\frac{\partial g_k\,(x)}{\partial \varepsilon_i} = \frac{\partial\left[\varepsilon_k - s_k^2\right]}{\partial \varepsilon_i} = \begin{cases} 1 & \text{if } i = k \\ 0 & \text{if } i \neq k \end{cases} \tag{2.59}$$

and for $k \in [1, q]$

$$\frac{\partial h_k\,(x)}{\partial \varepsilon_i} = \sum_{j=1}^{n} \frac{\partial h_k\,(x)}{\partial x_j} \frac{\partial x_j\,(\varepsilon)}{\partial \varepsilon_i} \tag{2.60}$$

Furthermore, we may define

$$\Phi_i \equiv \frac{\partial Z\,(\varepsilon)}{\partial \varepsilon_i} + \sum_{k=1}^{q} \lambda_k \frac{\partial h_k\,(x)}{\partial \varepsilon_i} + \sum_{k=1}^{m} \mu_k \frac{\partial g_k\,(x)}{\partial \varepsilon_i} \tag{2.61}$$

and note that

$$\Phi_i = \frac{\partial Z\,(\varepsilon)}{\partial \varepsilon_i} + \mu_i \tag{2.62}$$

With the help of (2.58), (2.59), and (2.60), we have

$$\begin{aligned} \Phi_i &= \sum_{j=1}^{n} \frac{\partial f\,(x)}{\partial x_j} \frac{\partial x_j\,(\varepsilon)}{\partial \varepsilon_i} + \sum_{k=1}^{q} \lambda_k \sum_{j=1}^{n} \frac{\partial h_k\,(x)}{\partial x_j} \frac{\partial x_j\,(\varepsilon)}{\partial \varepsilon_i} + \sum_{k=1}^{m} \mu_k \sum_{j=1}^{n} \frac{\partial g_k\,(x)}{\partial x_j} \frac{\partial x_j\,(\varepsilon)}{\partial \varepsilon_i} \\ &= \sum_{j=1}^{n} \left[\frac{\partial f\,(x)}{\partial x_j} + \sum_{k=1}^{q} \lambda_k \frac{\partial h_k\,(x)}{\partial x_j} + \sum_{k=1}^{m} \mu_k \frac{\partial g_k\,(x)}{\partial x_j} \right] \frac{\partial x_j\,(\varepsilon)}{\partial \varepsilon_i} = 0 \end{aligned} \tag{2.63}$$

by virtue of the Kuhn-Tucker identity (2.52). From (2.62) and (2.63) it is immediate that

$$\mu_i = (-1)\frac{\partial Z(\varepsilon)}{\partial \varepsilon_i} \quad i \in [1, m] \tag{2.64}$$

We now note that, when the unconstrained minimum of $f(x)$ is external to the feasible region

$$X(\varepsilon) = \{x : g(x) \leq \varepsilon, h(x) = 0\},$$

increasing ε_i can never increase, and may potentially lower, the objective function for all $i \in [1, m]$; that is

$$\frac{\partial Z(\varepsilon)}{\partial \varepsilon_i} \leq 0 \quad i \in [1, m] \tag{2.65}$$

From (2.64) and (2.65) we have the desired result

$$\mu_i \geq 0 \quad \forall i \in [1, m] \tag{2.66}$$

ensuring that the multipliers for inequality constraints are nonnegative.

2.5 Formal Derivation of the Kuhn-Tucker Conditions

We are interested in formally proving that, under the linear independence constraint qualification and some other basic assumptions, the Kuhn-Tucker identity and the complementary slackness conditions form, together with the original mathematical program's constraints, a valid set of necessary conditions. For finite-dimensional mathematical programs, the only type we consider in this chapter, such a demonstration is facilitated by Gordon's lemma, which is in effect a corollary of Farkas' lemma of classical analysis. The problem structure needed to apply Gordon's lemma can be most readily created by expressing the notion of optimality in terms of cones and separating hyperplanes. Throughout this section we consider the mathematical program

$$\min f(x) \quad \text{s.t.} \quad x \in \mathcal{F} \tag{2.67}$$

where, depending on context, either $\mathcal{F}$ is a general set or

$$\mathcal{F} \equiv \{x \in X_0 : g(x) \leq 0\} \subset \Re^n \tag{2.68}$$

and

$$f : \Re^n \to \Re^1 \tag{2.69}$$

$$g : \Re^n \to \Re^m \tag{2.70}$$

where X_0 is a nonempty open set in $\Re^n$. Note that we presently consider only inequality constraints, as any equality constraint

$$h_k(x) = 0$$

may be stated as two inequality constraints

$$\begin{aligned} h_k(x) &\leq 0 \\ -1 \cdot h_k(x) &\leq 0 \end{aligned}$$

2.5.1 Cones and Optimality

A cone is a set obeying the following definition:

Definition 2.3. *Cone. A set C in $\Re^n$ is a cone with vertex zero if $x \in C$ implies that $\theta x \in C$ for all $\theta \in \Re^1_+$.*

Now consider the following definitions:

Definition 2.4. *Cone of feasible directions. For the mathematical program (2.67), provided $\mathcal{F}$ is not empty, the cone of feasible directions at $x \in X$ is*

$$D_0(x) = \{d \neq 0 : x + \theta d \in \mathcal{F} \quad \forall \theta \in (0, \delta) \text{ and some } \delta > 0\}$$

Definition 2.5. *Feasible direction. Every nonzero vector $d \in D_0$ is called a feasible direction at $x \in X$ for the mathematical program (2.67).*

Definition 2.6. *Cone of improving directions. For the mathematical program (10.1), if f is differentiable at $x \in \mathcal{F}$, the cone of improving directions at $x \in \mathcal{F}$ is*

$$F_0(x) = \{d : [\nabla f(x)]^T \cdot d < 0\}$$

Definition 2.7. *Feasible direction of descent. Every vector $d \in F_0 \bigcap D_0$ is called a feasible direction of descent at $x \in \mathcal{F}$ for the mathematical program (2.67).*

Definition 2.8. *Cone of interior directions. For the mathematical program (10.1), if g_i is differentiable at $x \in X$ for all $i \in I(x)$, where*

$$I(x) = \{i : g_i(x) = 0\},$$

then, the cone of interior directions at $x \in \mathcal{F}$ is

$$G_0(x) = \{d : [\nabla g_i(x)]^T \cdot d < 0 \quad \forall\}$$

Note that in Definition 2.4 and Definition 2.6, if $\mathcal{F}$ is a convex set, we may set $\delta = 1$ and refer only to $\theta \in [0, 1]$, as will become clear in the next section after we define the notion of a convex set. Furthermore, the definitions immediately above allow one to characterize an optimal solution of (2.67) as a circumstance for which the intersection of the cone of feasible directions and the cone of improving directions is empty. This has great intuitive appeal for it says that there are no feasible directions that allow the objective to be improved. In fact, the following result obtains:

Theorem 2.5. *Optimality in terms of the cones of feasible and improving directions. Consider the mathematical program*

$$\min f(x) \quad s.t. \quad x \in \mathcal{F} \tag{2.71}$$

where $f : \Re^n \longrightarrow \Re^1$, $\mathcal{F} \subseteq \Re^n$ *and* $\mathcal{F}$ *is nonempty. Suppose also that* f *is differentiable at the local minimum* $x^* \in \mathcal{F}$ *of (2.71). Then at* x^* *the intersection of the cone of feasible directions* D_0 *and the cone of improving directions* F_0 *is empty:*

$$F_0(x^*) \cap D_0(x^*) = \emptyset$$

That is, at the local solution $x^* \in \mathcal{F}$, *no improving direction is also a feasible direction.*

Proof. The result is intuitive. For a formal proof see Bazaraa et al. (1993). ■

Theorem 2.6. *Optimality in terms of the cones of interior and improving directions. Let* $x^* \in \mathcal{F}$ *be a local minimum of the mathematical program*

$$min\, f(x) \quad s.t. \quad x \in \mathcal{F} = \{x \in X_0 : g(x) \leq 0\} \subset \Re^n \tag{2.72}$$

where X_0 *is a nonempty open set in* $\Re^n$, $f : \Re^n \longrightarrow \Re^1$, *and* $g : \Re^n \longrightarrow \Re^m$ *are differentiable at* x^*, *while the* g_i *for* $i \in I$ *are continuous at* x^*. *The cone of improving directions and the cone of interior directions satisfy*

$$F_0(x^*) \cap G_0(x^*) = \emptyset$$

Proof. This result is also intuitive. For a formal proof see Bazaraa et al. (1993). ■

2.5.2 *Theorems of the Alternative*

Farka's Lemma is a specific example of a so-called *theorem of the alternative*. Such theorems provide information on whether a given linear system has a solution when a related linear system has or fails to have a solution. Farkas' lemma has the following statement:

Lemma 2.1. *Farkas' lemma. Let* A *be an* $m \times n$ *matrix of real numbers and* $c \in \Re^n$. *Then exactly one of the following systems has a solution: System 1:* $Ax \leq 0$ *and* $c^T x > 0$ *for some* $x \in \Re^n$; *or System 2:* $A^T y = c$ *and* $y \geq 0$ *for some* $y \in \Re^m$.

Proof. Farkas' lemma is proven in most advanced texts on nonlinear programming. See, for example, Mangasarian (1969). ■

Corollary 2.1. *Gordon's corollary. Let* A *be an* $m \times n$ *matrix of real numbers. Then exactly one of the following systems has a solution: System 1:* $Ax < 0$ *for some* $x \in \Re^n$; *or System 2:* $A^T y = 0$ *and* $y \geq 0$ *for some* $y \in \Re^m$.

Proof. See Mangasarian (1969). ■

2.5.3 The Fritz John Conditions Again

By using Corollary 2.1 it is quite easy to establish the Fritz John conditions introduced previously and restated here without equality constraints:

Theorem 2.7. *The Fritz John conditions. Let $x^* \in \mathcal{F}$ be a minimum of*

$$\begin{array}{ll} min & f(x) \\ s.t. & x \in \mathcal{F} = \{x \in X_0 : g(x) \leq 0\} \end{array}$$

where X_0 is a nonempty open set in $\Re^n$ and $g : \Re^n \longrightarrow \Re^m$. Assume that $f(x)$ and $g_i(x)$ for $i \in [1, m]$ have continuous first derivatives everywhere on $\mathcal{F}$. Then there must exist multipliers $\mu_0 \in \Re^1_+$ and $\mu = (\mu_1, \ldots, \mu_m)^T \in \Re^m_+$ such that

$$\mu_0 \nabla f(x^*) + \textstyle\sum_{i=1}^m \mu_i \nabla g_i(x^*) = 0 \tag{2.73}$$

$$\mu_i g_i(x^*) = 0 \quad \forall\ i \in [1, m] \tag{2.74}$$

$$\mu_i \geq 0 \quad \forall\ i \in [1, m] \tag{2.75}$$

$$(\mu_0,\ \mu) \neq 0 \in \Re^{m+1} \tag{2.76}$$

Proof. Since $x^* \in \mathcal{F}$ solves the mathematical program of interest, we know from Theorem 2.6 that $F_0(x^*) \cap G_0(x^*) = \emptyset$; that is, there is no vector d satisfying

$$\left[\nabla f(x^*)\right]^T \cdot d < 0 \tag{2.77}$$

$$\left[\nabla g_i(x^*)\right]^T \cdot d < 0 \quad i \in I(x^*) \tag{2.78}$$

where $I(x^*)$ is the set of indices of constraints binding at x^*. Without loss of generality, we may consecutively number the binding constraints from 1 to $|I(x^*)|$ and define

$$A = \begin{pmatrix} [\nabla f(x^*)]^T & 0 & 0 & \cdots & 0 \\ 0 & [\nabla g_1(x^*)]^T & 0 & \cdots & 0 \\ 0 & 0 & [\nabla g_2(x^*)]^T & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & \left[\nabla g_{|I(x^*)|}(x^*)\right]^T \end{pmatrix}$$

As a consequence we may state (2.77) and (2.78) as

$$A\begin{pmatrix} d \\ d \\ \vdots \\ d \end{pmatrix} < 0 \tag{2.79}$$

According to Corollary 2.1, since (2.79) cannot occur, there exists

$$y = \begin{pmatrix} \mu_0 \\ \mu_i : i \in I(x^*) \end{pmatrix} \geq 0$$

such that

$$A^T y = A^T \begin{pmatrix} \mu_0 \\ \mu_i : i \in I(x^*) \end{pmatrix} = 0 \tag{2.80}$$

Expression (2.80) yields

$$\mu_0 \nabla f\left(x^*\right) + \sum_{i=1}^{|I(x^*)|} \mu_i \nabla g_i(x^*) = 0 \tag{2.81}$$

We are free to introduce the additional multipliers

$$\mu_i = 0 \quad i = |I(x^*)| + 1, \ldots, m \tag{2.82}$$

which assure that the complementary slackness conditions (2.74) and (2.75) hold for all multipliers. As a consequence of (2.81) and (2.82), we have (2.73), thereby completing the proof. ■

2.5.4 The Kuhn-Tucker Conditions Again

With the apparatus developed so far, we wish to prove the following restatement of Theorem 2.4 in terms of the linear independence constraint qualification:

Theorem 2.8. *Kuhn-Tucker conditions. Let $x^* \in \mathcal{F}$ be a local minimum of*

$$\begin{aligned} &min\ \ f(x) \\ &s.t.\ \ x \in \mathcal{F} = \{x \in X_0 : g(x) \leq 0, h(x) = 0\} \end{aligned}$$

where X_0 is a nonempty open set in $\Re^n$. Assume that $f(x)$, $g_i(x)$ for $i \in [1, m]$ and $h_i(x)$ for $i \in [1, q]$ have continuous first derivatives everywhere on $\mathcal{F}$ and that the gradients of binding constraint functions are linearly independent. Then there *must exist multipliers* $\mu = (\mu_1, \ldots, \mu_m)^T \in \Re^m$ *and* $\lambda = (\lambda_i, \ldots, \lambda_q)^T \in \Re^q$ *such that*

$$\nabla f(x^*) + \sum_{i=1}^{m} \mu_i \nabla g_i(x^*) + \sum_{i=1}^{q} \lambda_i \nabla h_i(x^*) = 0 \tag{2.83}$$

$$\mu_i g_i(x^*) = 0 \quad \forall\ i \in [1, m] \tag{2.84}$$

$$\mu_i \geq 0 \quad \forall\ i \in [1, m] \tag{2.85}$$

Proof. Recall that a constraint qualification is a condition that guarantees the multiplier μ_0 of the Fritz John conditions is non-zero. We again use the notation

$$I\left(x^*\right) = \left\{i : g_i\left(x^*\right) = 0\right\}, \tag{2.86}$$

for the set of subscripts corresponding to binding inequality constraints. Note also that by their very nature equality constraints are always binding. Linear independence of the gradients of binding constraints means that only zero multipliers

$$\mu_i = 0 \quad \forall\ i \in I\left(x^*\right) \tag{2.87}$$

$$\lambda_i = 0 \quad \forall\ i \in [1, q] \tag{2.88}$$

allow the identity

$$\sum_{i \in I(x^*)} \mu_i \nabla g_i(x^*) + \sum_{i=1}^{q} \lambda_i \nabla h_i(x^*) = 0, \tag{2.89}$$

to hold. We are free to set the multipliers for nonbinding constraints to zero; that is

$$g_i(x^*) < 0 \Longrightarrow \mu_i = 0 \quad \forall\ i \notin I\left(x^*\right)$$

which assures (2.84) and (2.85) hold for $i \in [1, m]$. Consequently, linear independence of the gradients of binding constraints actually means that there are no nonzero multipliers assuring

$$\sum_{i=1}^{m} \mu_i \nabla g_i(x^*) + \sum_{i=1}^{q} \lambda_i \nabla h_i(x^*) = 0 \tag{2.90}$$

That is, either all $\lambda_i = 0$ and all $\mu_i = 0$ or

$$\sum_{i=1}^{m} \mu_i \nabla g_i(x^*) + \sum_{i=1}^{q} \lambda_i \nabla h_i(x^*) \neq 0 \tag{2.91}$$

In the latter case, the Fritz John identity

$$\mu_0 \nabla f\left(x^*\right) + \sum_{i=1}^{m} \mu_i \nabla g_i(x^*) + \sum_{i=1}^{q} \lambda_i \nabla h_i(x^*) = 0 \tag{2.92}$$

immediately forces

$$\mu_0 \neq 0 \tag{2.93}$$

unless $\nabla f(x^*) = 0 \in \Re^n$; in this latter case (2.90) must hold and so we may still enforce (2.93) without contradiction or loss of generality. ■

2.6 Sufficiency, Convexity, and Uniqueness

Sufficient conditions for optimality in a mathematical program are conditions that, if satisfied, ensure optimality. Any such condition has the logical structure:

If property $\mathbf{P}(x^*)$ is true, then x^* is optimal.

It turns out that convexity, a notion that requires careful definition, provides useful sufficient conditions that are relatively easy to check in practice. In particular, we will define a convex mathematical program to be a mathematical program with a convex objective function (when minimizing) and a convex feasible region, and we will show that the Kuhn-Tucker conditions are not only necessary but also sufficient for global optimality in such programs.

2.6.1 Quadratic Forms

A key concept, useful for establishing convexity of functions, is that of a quadratic form, formally defined as follows:

Definition 2.9. *Quadratic form. A quadratic form is a scalar-valued function defined for all $x \in \Re^n$ that takes on the following form:*

$$Q(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j \tag{2.94}$$

where each a_{ij} is a real number.

Note that any quadratic form may be expressed in matrix notation as

$$Q(x) = x^T A x \tag{2.95}$$

where $A = (a_{ij})$ is an $n \times n$ matrix. It is well known that for any given quadratic form there is a symmetric matrix S that allows one to to re-express that quadratic form as

$$Q(x) = x^T S x \tag{2.96}$$

where the elements of $S = (s_{ij})$ are given by $s_{ij} = s_{ji} = (a_{ij} + a_{ji})/2$. Because of this symmetry property, we may assume, without loss of generality, that every quadratic form is already expressed in terms of a symmetric matrix. That is, whenever we encounter a quadratic form such as (2.95) or (2.96), the underlying matrix generating that form may be taken to be symmetric if so doing assists our analysis.

A quadratic form may exhibit various properties, two of which are the subject of the following definition:

Definition 2.10. *Positive definiteness. The quadratic form $Q(x) = x^T S x$ is positive definite on $\Omega \subseteq \Re^n$ if $Q(x) > 0$ for all $x \in \Omega$ and $x \neq 0$. The quadratic form $Q(x) = x^T S x$ is positive semidefinite on $\Omega \subseteq \Re^n$ if $Q(x) \geq 0$ for all $x \in \Omega$.*

Analogous definitions may be made for negative definite and negative semidefinite quadratic forms. An important lemma concerning quadratic forms, which we state without proof, is the following:

Lemma 2.2. *Properties of positive definite matrix. Let the symmetric $n \times n$ matrix S be positive (negative) definite. Then*

1. *The inverse S^{-1} exists;*
2. *S^{-1} is positive (negative) definite; and*
3. *$A^T SA$ is positive (negative) semidefinite for any $m \times n$ matrix A.*

In addition, we will need the following lemma, which we also state without proof:

Lemma 2.3. *Nonnegativity of principal minirs. A quadratic form $Q(x) = x^T S x$, where S is the associated symmetric matrix, is positive semidefinite if and only if it may be ordered so that s_{11} is positive and the following determinants of the principal minors are all nonnegative:*

$$\begin{vmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{vmatrix} \geq 0, \quad \begin{vmatrix} s_{11} & s_{12} & s_{13} \\ s_{21} & s_{22} & s_{23} \\ s_{31} & s_{32} & s_{33} \end{vmatrix} \geq 0, \ \ldots, |S| \geq 0$$

2.6.2 Concave and Convex Functions

This section contains several definitions, lemmas, and theorems related to convex functions and convex sets that we need to fully understand the notion of sufficiency. First, consider the following four definitions:

Definition 2.11. *Convex set. A set $X \subseteq \Re^n$ is convex if for any two vectors $x^1, x^2 \in X$ and any scalar $\lambda \in [0, 1]$ the vector*

$$x = \lambda x^1 + (1 - \lambda)x^2 \tag{2.97}$$

also lies in X.

Definition 2.12. *Strictly convex set. A set $X \subseteq \Re^n$ is strictly convex if for any two vectors x^1 and x^2 in X and any scalar $\lambda \in (0, 1)$ the point*

$$x = \lambda x^1 + (1 - \lambda)x^2 \tag{2.98}$$

lies in the interior of X.

Definition 2.13. *Convex function. A scalar function $f(x)$ is a convex function defined over a convex set $X \subseteq \Re^n$ if for any two vectors $x^1, x^2 \in X$*

$$f(\lambda x^1 + (1 - \lambda)x^2) \leq \lambda f(x^1) + (1 - \lambda) f(x^2) \quad \forall \, \lambda \in [0, 1] \tag{2.99}$$

Definition 2.14. *Strictly convex function. In the above, $f(x)$ is strictly convex if the inequality is a strict inequality ($<$) for all $\lambda \in (0, 1)$.*

Note that concave and strictly concave functions are defined by reversing the inequalities in the preceding definitions.

We are now ready to state the following theorems:

Theorem 2.9. *Sum of convex functions. The sum of any two convex functions is convex.*

Theorem 2.10. *Convexity of linear functions. Any linear function is both convex and concave.*

Theorem 2.11. *Convexity of quadratic form. For any positive semidefinite and symmetric matrix S, the quadratic form $Q(x) = x^T Sx$ is a convex function over all of $\Re^n$.*

The proofs of the preceding results are straightforward and are left to the reader. Another important result is the following that relates convex level sets and convex functions:

Theorem 2.12. *Level sets of convex funtion. If $f(x)$ is a (strictly) convex function over $\Re^n$, then the set of points*

$$S \equiv \{x : f(x) \leq b\}, \tag{2.100}$$

where b is any real number, is a (strictly) convex set.

Proof. The definition of convexity tells us that

$$\begin{aligned} f(\lambda x^1 + (1 - \lambda)x^2) &\leq \lambda f(x^1) + (1 - \lambda) f(x^2) \\ &\leq \lambda b + (1 - \lambda)b = b \end{aligned}$$

A strict version of this inequality is obtained for strictly convex functions, thereby completing the proof. ■

We will also need the following lemma:

Lemma 2.4. *Intersection of convex sets. The intersection of any two convex sets is itself a convex set.*

Proof. Take x^1 and x^2 within the intersection $X^1 \cap X^2$, where X^1 and X^2 are convex sets. Join these points by a line segment. That line segment and all the points on it are both in X^1 and X^2. ■

It is now trivial to establish the following result:

Theorem 2.13. *Convex feasible region. The feasible region X of the mathematical program (2.3) is a convex set if the following two conditions are met:*

1. *the equality constraint functions $h_i(x) : \Re^n \longrightarrow \Re^1$ for $i \in [1, q]$ are all linear on X; and*
2. *the inequality constraint functions $g_i(x) : \Re^n \longrightarrow \Re^1$ for $i \in [1, m]$ are all convex on X.*

Proof. For the given, the sets

$$S_h = \{x : h(x) = 0\}$$
$$S_g = \{x : g(x) \leq 0\}$$

are convex. The feasible region X obeys

$$X = S_h \cap S_g$$

Hence, X is convex, since the intersection of two convex sets is a convex set. ■

Now we are ready to deal with the following key result:

Theorem 2.14. *Global minimum of a convex program. If the function $f(x) : \Re^n \longrightarrow \Re^1$ is defined and convex on the closed convex set $X \subseteq \Re^n$, then any constrained local minimum of $f(x)$ for $x \in X$ is a global minimum on X. Similarly, if $f(x)$ is concave on the closed convex set X, then any constrained local maximum of $f(x)$ for $x \in X$ is a global maximum on X.*

Proof. Suppose $x^0 \in X$ is a constrained local minimum but not a global minimum, so that there exists some $x^* \in X$ such that $f(x^*) < f(x^0)$. Then for any $\lambda \in [0, 1]$ the convexity of $f(x)$ tells us that

$$\begin{aligned} f(\lambda x^* + (1-\lambda)x^0) &\leq \lambda f(x^*) + (1-\lambda) f(x^0) \\ &< \lambda f(x^0) + (1-\lambda) f(x^0) = f(x^0) \end{aligned} \tag{2.101}$$

Now, consider a straight line segment from x^0 to x^* which must lie entirely in X (by convexity). For any small positive δ (a scalar), there exists $\lambda > 0$ such that

$$x = \lambda x^* + (1 - \lambda)x^0 \tag{2.102}$$

lies in X at a distance δ away from x^0. However, we have already shown in (2.101) that

$$f(x) < f(x^0) \tag{2.103}$$

Since δ may be infinitesimally small, x^0 cannot be a local minimum. Hence, we have a contradiction. ■

Another important result is the following:

Theorem 2.15. *Tangent line property of a convex function. Let $f(x)$ have continuous first partial derivatives. Then $f(x)$ is convex over the convex region $X \subseteq \Re^n$ if and only if*

$$f(x) \geq f(x^*) + \left[\nabla f(x^*)\right]^T (x - x^*) \tag{2.104}$$

for any two vectors x^* and x in X. Moreover, $f(x)$ is concave over the convex region $X \subseteq \Re^n$ if and only if

$$f(x) \leq f(x^*) + \left[\nabla f(x^*)\right]^T (x - x^*) \tag{2.105}$$

for any two vectors x^* and x in X.

This result may be proven by taking a Taylor series expansion of $f(x)$ about the point x^* and arguing that the second order and higher terms sum to a positive number. Theorem 2.15 expresses the geometric property that a tangent to a convex function will underestimates that function. Still another related result is:

Theorem 2.16. *Convexity and positive semidefiniteness of the Hessian. Let $f(x)$ have continuous second partial derivatives. Then $f(x)$ is convex (concave) over some the region $X \subseteq \Re^n$ if and only if its Hessian matrix*

$$H(x) \equiv \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 f}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_n \partial x_1} & \dfrac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{bmatrix} \tag{2.106}$$

is positive (negative) semidefinite.

Proof. We give the proof for concave functions, although the case of convex functions is completely analogous.

(i) [negative semidefiniteness $\Longrightarrow$ concavity] First note that the Hessian H is symmetric by its very nature. We may make a second-order Taylor series expansion of $f(x)$ about a point $x^* \in X$ to obtain

$$f(x) = f(x^*) + \left[\nabla f(x^*)\right]^T (x - x^*) + \frac{1}{2}(x - x^*)^T H[x^* + \theta(x - x^*)](x - x^*) \tag{2.107}$$

for some $\theta \in (0, 1)$. Because X is convex we know that the point

$$x^* + \theta(x - x^*) = \theta x + (1 - \theta)x^* , \tag{2.108}$$

a convex combination of x and x^*, must lie within X. Now suppose that H is negative definite or negative semidefinite throughout X, so that the last term on the righthand side of the Taylor expansion is clearly negative or zero. We get

$$f(x) \leq f(x^*) + \left[\nabla f(x^*)\right]^T (x - x^*) \tag{2.109}$$

It follows from the previous theorem that $f(x)$ is concave.

(ii) [concavity $\Longrightarrow$ negative semidefiniteness] Now assume $f(x)$ is concave throughout X but that the Hessian matrix H is not negative semidefinite at some point $x^* \in X$. Then, of course, there will exist a vector y such that

$$y^T H(x^*) y > 0 \tag{2.110}$$

Now define $x^0 = x^* + y$ and rewrite this last inequality as

$$(x^0 - x^*)^T H(x^*)(x^0 - x^*) > 0 \tag{2.111}$$

Consider another point $x = x^* + \beta(x^0 - x^*)$ where β is a real positive number, so that

$$(x^0 - x^*) = \frac{1}{\beta}(x - x^*) \tag{2.112}$$

It follows that for any such β

$$(x - x^*)^T H(x^*)(x - x^*) > 0 \tag{2.113}$$

Since H is continuous, we may choose x so close to x^* that

$$(x - x^*)^T H[x^* + \theta(x - x^*)](x - x^*) > 0 \tag{2.114}$$

for all $\theta \in [0, 1]$. By hypothesis $f(x)$ is concave over $\Re$ so that

$$f(x) \leq f(x^*) + \left[\nabla f(x^*)\right]^T (x - x^*) \tag{2.115}$$

holds, together with the Taylor series expansion (2.107). Subtracting (2.115) from (2.107) gives

$$0 \geq \frac{1}{2}(x - x^*)^T H[x^* + \theta(x - x^*)](x - x^*) \tag{2.116}$$

for some $\theta \in (0, 1)$. This contradicts (2.114). ■

Note this last theorem cannot be strengthened to say a function is strictly convex if and only if its Hessian is positive definite. Examples may be given of functions that are strictly convex and whose Hessians are not positive definite. However, one can establish that positive definiteness of the Hessian does imply strict convexity by employing some of the arguments from the preceding proof.

Furthermore, the manner of construction of the preceding proofs leads directly to the following corollary:

Corollary 2.2. *Solution set convex. If the constrained global minimum of $f(x)$ for $x \in X \subset \Re^n$ is α when $f(x) : \Re^n \longrightarrow \Re^1$ is convex on X, a convex set, then the set*

$$\Psi = \{x : x \in X \subset \Re^n, \quad f(x) \leq \alpha\} \tag{2.117}$$

is the set of all solutions and is itself convex.

We now turn our attention to the question of additional regularity conditions that will assure that the set Ψ is a singleton. In fact, we will prove the following theorem:

Theorem 2.17. *Unique global minimum. Let $f(\cdot)$ be a strictly convex function defined on a convex set $X \subset \Re^n$. If $f(\cdot)$ attains its global minimum on X, it is attained at a unique point of X.*

Proof. Suppose there are two global minima: $x^1 \in X$ and $x^2 \in X$. Let $f(x^1) = f(x^2) = \alpha$. Then, by the previous corollary the set Ψ is a convex set and is the set of all solutions. Therefore

$$x^1, \ x^2, \ x^3 \in \Psi \tag{2.118}$$

where $x^3 = \lambda x^1 + (1 - \lambda)x^2$, and

$$\alpha = f(x^3) = f(\lambda x^1 + (1 - \lambda)x^2) < \lambda f(x^1) + (1 - \lambda) f(x^2) = \alpha.$$

This is a contradiction and therefore there cannot be two global minima. ■

2.6.3 Kuhn-Tucker Conditions Sufficient

The most significant implication of imposing regularity conditions based on convexity is that they make the Kuhn-Tucker conditions sufficient as well as necessary for global optimality. In fact, we may state and prove the following:

Theorem 2.18. *Kuhn-Tucker conditions sufficient for convex programs. Let*

$$\begin{aligned} f &: X \subset \Re^n \longrightarrow \Re^n \\ g &: X \subset \Re^n \longrightarrow \Re^m \\ h &: X \subset \Re^n \longrightarrow \Re^q \end{aligned}$$

be real-valued, differentiable functions. Suppose X_0 is an open convex set, while f is convex, the g_i are convex for $i \in [1, m]$, and the h_i are linear for $i \in [1, q]$. Take x^* to be a feasible solution of the mathematical program

$$\left.\begin{array}{ll} \min & f(x) \\ \text{s.t.} & h_i(x) = 0 \ (\lambda_i) \quad i \in [1, q] \\ & g_i(x) \leq 0 \ (\mu_i) \quad i \in [1, m] \\ & x \in X_0 \end{array}\right\} \tag{2.119}$$

If there exist multipliers $\mu^* \in \Re^m$ and $\lambda^* \in \Re^q$ satisfying the Kuhn-Tucker conditions

$$\nabla f(x^*) + \sum_{i=1}^{m} \mu_i^* \nabla g_i(x^*) + \sum_{i=1}^{q} \lambda_i^* \nabla h_i(x^*) = 0$$

$$\mu_i^* g_i(x^*) = 0 \quad \mu_i^* \geq 0 \quad i \in [1, m] \quad ,$$

then x^* is a global minimum.

Proof. To simplify the exposition, we shall assume only constraints that are inequalities; this is possible since any linear equality constraint

$$h_k(x) = 0$$

for $k \in [1, m]$ may be restated as two convex inequality constraints in standard form:

$$\begin{aligned} h_k(x) &\leq 0 \\ -h_k(x) &\leq 0 \end{aligned}$$

and absorbed into the definition of $g(x)$. The Kuhn-Tucker identity is then

$$\nabla f(x^*) + \sum_{i=1}^{m} \mu_i^* \nabla g_i(x^*) = 0 \tag{2.120}$$

Post multiplying (2.120) by $(x - x^*)$ gives

$$\left[\nabla f(x^*)\right]^T (x - x^*) + \sum_i \mu_i^* \left[\nabla g_i(x^*)\right]^T (x - x^*) = 0 \tag{2.121}$$

where x^* is a solution of the Kuhn-Tucker conditions and

$$x, x^* \in X = \{x \in X_0 : g(x) \leq 0, h(x) = 0\}$$

We know that for a convex, differentiable function

$$g(x) \geq g(x^*) + \left[\nabla g(x^*)\right]^T (x - x^*) \tag{2.122}$$

From (2.121) and (2.122), we have

$$\begin{aligned}
\left[\nabla f(x^*)\right]^T (x - x^*) &= -\sum \mu_i^* \left[\nabla g_i(x^*)\right]^T (x - x^*) \\
&\geq \sum_i \mu_i^* [g_i(x^*) - g_i(x)] \\
&= \mu_i^* [-g_i(x)] \geq 0
\end{aligned} \tag{2.123}$$

because $\mu_i^* g_i(x^*) = 0$, $\mu_i^* \geq 0$ and $g_i(x) \leq 0$. Hence

$$\left[\nabla f(x^*)\right]^T (x - x^*) \geq 0 \tag{2.124}$$

Because $f(x)$ is convex

$$f(x) \geq f(x^*) + \left[\nabla f(x^*)\right]^T (x - x^*) \tag{2.125}$$

Hence, from (2.124) and (2.125) we get

$$f(x) - f(x^*) \geq \left[\nabla f(x^*)\right]^T (x - x^*) \geq 0 \tag{2.126}$$

That is

$$f(x) \geq f(x^*) \ ,$$

which establishes that any solution of the Kuhn-Tucker conditions is a global minimum for the given. ■

Note that this theorem can be changed to one in which the objective function is strictly convex, thereby assuring that any corresponding solution of the Kuhn-Tucker conditions is an unique global minimum. Its given may also be relaxed if certain results from the theory of generalized convexity are employed.

2.7 Generalized Convexity and Sufficiency

There are certain generalizations of the notion of convexity that allow the sufficiency conditions introduced above to be somewhat weakened. We begin to explore the notion of more general types of convexity by introducing the following definition of a quasiconvex function:

Definition 2.15. *Quasiconvex function. The function $f : X \longrightarrow \Re^n$ is quasiconvex on the set $X \subset \Re^n$ if*

$$f\left(\lambda_1 x^1 + \lambda_2 x^2\right) \leq \max\left[f\left(x^1\right), f\left(x^2\right)\right]$$

for every $x^1, x^2 \in X$ and every $(\lambda_1, \lambda_2) \in \left\{(\lambda_1, \lambda_2) \in \Re^2_+ : \lambda_1 + \lambda_2 = 1\right\}$.

We next introduce the notion of a pseudoconvex function:

Definition 2.16. *Pseudoconvex function. The function $f : X \longrightarrow \Re^n$, differentiable on the open convex set $X \subset \Re^n$, is pseudoconvex on X if*

$$\left(x^1 - x^2\right)^T \nabla f\left(x^2\right) \geq 0$$

implies that

$$f(x^1) \geq f(x^2)$$

for every $x^1, x^2 \in X$.

Pseudoconcavity of f occurs of course when $-f$ is pseudoconvex. Furthermore, we shall say a function is pseudolinear (quasilinear) if it is both pseudoconvex (quasiconvex) and pseudoconcave (quasiconcave).

The notions of generalized convexity we have given allow the following theorem to be stated and proven:

Theorem 2.19. *Kuhn-Tucker conditions sufficient for generalized convex programs. Let*

$$\begin{aligned} f &: X \subset \Re^n \longrightarrow \Re^n \\ h &: X \subset \Re^n \longrightarrow \Re^m \\ g &: X \subset \Re^n \longrightarrow \Re^q \end{aligned}$$

be real-valued, differentiable functions. Suppose X_0 is an open convex set, while f is pseudoconvex, the g_i are quasiconvex for $i \in [1, m]$, and the h_i are quasilinear for $i \in [1, q]$. Take x^* to be a feasible solution of the mathematical program

$$\begin{array}{lllll} \textit{min} & f(x) & & \\ \textit{s.t.} & h_i(x) = 0 & (\eta_i) & i \in [1, q] \\ & g_i(x) \leq 0 & (\lambda_i) & i \in [1, m] \\ & x \in X_0 & & \end{array}$$

If there exist multipliers $\mu^* \in \Re^m$ and $\lambda^* \in \Re^q$ satisfying the Kuhn-Tucker conditions

$$\nabla f(x^*) + \sum_{i=1}^{m} \mu_i^* \nabla g_i(x^*) + \sum_{i=1}^{q} \lambda_i^* \nabla h_i(x^*) = 0$$

$$\mu_i^* g_i(x^*) = 0 \quad \mu_i^* \geq 0 \quad i \in [1, m],$$

then x^* is a global minimum.

Proof. The proof is left as an exercise for the reader. ■

We close this section by noting that if in addition to the given of Theorem 2.19 an appropriate notion of strict pseudoconvexity is introduced for the objective function f, then the Kuhn-Tucker conditions become sufficient for a unique global minimizer.

2.8 Numerical and Graphical Examples

In this section we provide several numerical and graphical examples meant to test and refine the reader's knowledge of the material on nonlinear programming presented above. We will need the notions of a level curve C_k and a level set S_k of the objective function $f(x)$ of a mathematical program:

$$C_k = \{x : f(x) = f_k\} \tag{2.127}$$

$$S_k = \{x : f(x) \le f_k\} \tag{2.128}$$

where f_k signifies a numerical value of the objective function of interest. Solving any mathematical program graphically involves four steps:

1. Draw the feasible region.
2. Draw level curves of the objective function.
3. Choose the optimal level curve by selecting, from the points of tangency of level curves and constraint boundaries, the fesible point or points giving the best objective function value.
4. Identify the optimal solution as the point of tangency between the optimal level curve and the feasible region

2.8.1 LP Graphical Solution

Consider the following linear program:

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

subject to

$$3x + 2y \le 6 \tag{2.129}$$

$$\frac{1}{2}x + y \le 2 \tag{2.130}$$

For the present example the optimal solution is, by inspection of Figure 2.2, the point

$$x^* = \begin{pmatrix} x_1^* \\ x_2^* \end{pmatrix} = \begin{pmatrix} 1 \\ \frac{3}{2} \end{pmatrix} \tag{2.131}$$

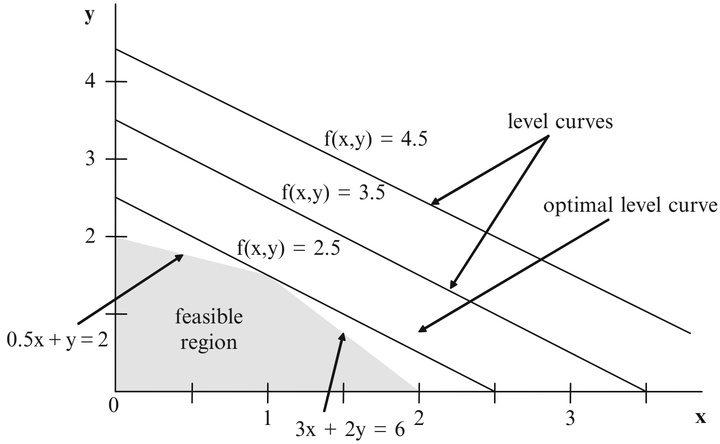

Fig. 2.2 LP graphical solution

One can easily verify the Kuhn-Tucker conditions hold at this point. To do so, it is helpful to restate the problem as follows:

$$\min f(x, y) = -x - y \tag{2.132}$$

$$g_1(x, y) = 3x + 2y - 6 \leq 0 \tag{2.133}$$

$$g_2(x, y) = \frac{1}{2}x + y - 2 \leq 0 \tag{2.134}$$

We note that

$$\nabla f(x, y) = \begin{pmatrix} -1 \\ -1 \end{pmatrix} \tag{2.135}$$

$$\nabla g_1(x, y) = \begin{pmatrix} 3 \\ 2 \end{pmatrix} \tag{2.136}$$

$$\nabla g_2(x, y) = \begin{pmatrix} \frac{1}{2} \\ 1 \end{pmatrix} \tag{2.137}$$

The Kuhn-Tucker identity is

$$\nabla f\,(x_1, x_2) + \lambda_1 \nabla g_1\,(x_1, x_2) + \lambda_2 \nabla g_2\,(x_1, x_2) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{2.138}$$

That is

$$\begin{pmatrix} -1 \\ -1 \end{pmatrix} + \lambda_1 \begin{pmatrix} 3 \\ 2 \end{pmatrix} + \lambda_2 \begin{pmatrix} \frac{1}{2} \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{2.139}$$

The complementary slackness conditions are

$$\lambda_1 g_1(x_1, x_2) = 0 \quad \lambda_1 \geq 0 \tag{2.140}$$

$$\lambda_2 g_2(x_1, x_2) = 0 \quad \lambda_2 \geq 0 \tag{2.141}$$

Note that

$$I = \{i : g_i\left(1, \frac{3}{2}\right) = 0\} = \{1, 2\} \tag{2.142}$$

and we must find multipliers that obey

$$\lambda_1, \lambda_2 \geq 0 \tag{2.143}$$

It is easy to solve the above system and show

$$\lambda_1 = \frac{1}{4} > 0, \lambda_2 = \frac{1}{2} > 0 \tag{2.144}$$

Hence x^* satisfies the Kuhn-Tucker conditions. Because the problem is a linear program, it is a convex program. Therefore, the Kuhn-Tucker conditions are not only necessary but also sufficient, making x^* a global solution.

2.8.2 NLP Graphical Example

Consider the following nonlinear program

$$\min \ f(x_1, x_2) = (x_1 - 5)^2 + (x_2 - 6)^2 \tag{2.145}$$

subject to

$$g_1(x_1, x_2) = \frac{1}{2}x_1 + x_2 - 3 \leq 0 \tag{2.146}$$

$$g_2(x_1, x_2) = x_1 - 2 \leq 0 \tag{2.147}$$

By inspection of Figure 2.3, the point (2, 2) is the globally optimal solution with a corresponding objective function value of 25. Note that

$$\nabla f(2, 2) = \begin{pmatrix} -8 \\ -6 \end{pmatrix} \tag{2.148}$$

$$\nabla g_1(2, 2) = \begin{pmatrix} \frac{1}{2} \\ 1 \end{pmatrix} \tag{2.149}$$

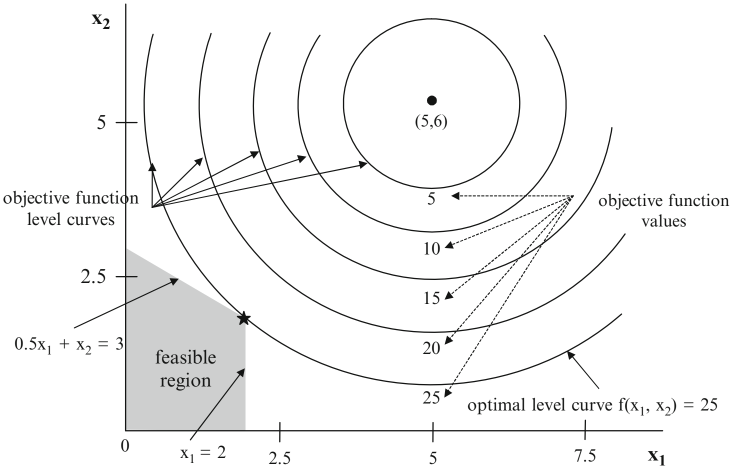

Fig. 2.3 NLP graphical solution

$$\nabla g_2(2,2) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{2.150}$$

The Kuhn-Tucker identity is

$$\begin{pmatrix} -8 \\ -6 \end{pmatrix} + \lambda_1 \begin{pmatrix} \frac{1}{2} \\ 1 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{2.151}$$

The complementary slackness conditions are

$$\lambda_1 g_1(x_1, x_2) = 0 \quad \lambda_1 \geq 0 \tag{2.152}$$

$$\lambda_2 g_2(x_1, x_2) = 0 \quad \lambda_2 \geq 0 \tag{2.153}$$

and

$$I = \{i : g_i(1,2) = 0\} = \{1,2\} \Longrightarrow \lambda_1, \lambda_2 \geq 0 \tag{2.154}$$

Solving the above linear system (2.151) yields multipliers of the correct sign:

$$\lambda_1 = 6 > 0 \tag{2.155}$$

$$\lambda_2 = 5 > 0 \tag{2.156}$$

Consequently, the Kuhn-Tucker conditions are satisfied. Because the program is convex with a strictly convex objective function, we know that the Kuhn-Tucker conditions are both necessary and sufficient for an unique global optimum. So, even without further analysis, we know $(2,2)$ is the unique global optimum.

2.8.3 Nonconvex, Nongraphical Example

Consider the nonlinear program

$$\min f(x_1, x_2) = -x_1 + 0x_2 \tag{2.157}$$

subject to

$$g_1(x_1, x_2) = (x_1)^2 + (x_2)^2 - 2 \le 0 \tag{2.158}$$

$$g_2(x_1, x_2) = x_1 - (x_2)^2 \le 0 \tag{2.159}$$

Note that the feasible region of this mathematical program is not convex; hence, we will have to enumerate all the combinations of binding and nonbinding constraints in order to solve it using the Kuhn-Tucker conditions alone. We begin by observing that

$$\nabla f(x_1, x_2) = \begin{pmatrix} -1 \\ 0 \end{pmatrix} \tag{2.160}$$

$$\nabla g_1(x_1, x_2) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} \tag{2.161}$$

$$\nabla g_2(x_1, x_2) = \begin{pmatrix} 1 \\ -2x_2 \end{pmatrix} \tag{2.162}$$

The Kuhn-Tucker identity is

$$\begin{pmatrix} -1 \\ 0 \end{pmatrix} + \lambda_1 \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ -2x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{2.163}$$

from which we obtain the equations

$$kti1 : -1 + 2\lambda_1 x_1 + \lambda_2 = 0 \tag{2.164}$$

$$kti2 : (\lambda_1 - \lambda_2)\, x_2 = 0 \tag{2.165}$$

The complementary slackness conditions are

$$csc1 : \lambda_1 g_1(x_1, x_2) = 0 \quad \lambda_1 \ge 0 \tag{2.166}$$

$$csc2 : \lambda_2 g_2(x_1, x_2) = 0 \quad \lambda_2 \ge 0 \tag{2.167}$$

Because there are $N = 2$ inequality constraints, there are $2^N = 2^2 = 4$ possible cases of binding and nonbinding constraints:

$$\left.\begin{array}{|l|l|l|}\hline \text{Case} & g_1 & g_2 \\ \hline \text{I} & <0 & <0 \\ \hline \text{II} & <0 & =0 \\ \hline \text{III} & =0 & <0 \\ \hline \text{IV} & =0 & =0 \\ \hline \end{array}\right\} \tag{2.168}$$

It is convenient to use the following symbols and operators for analyzing each of the four cases:

$$\left.\begin{array}{|l|l|}\hline \text{Symbol/Operator} & \text{Meaning} \\ \hline \oplus & \text{consider two statements} \\ \hline \Longrightarrow & \text{the implication of such a consideration} \\ \hline \pitchfork & \text{a contradiction has occurred} \\ \hline dno & \text{does not occur} \\ \hline \end{array}\right\} \tag{2.169}$$

Remembering that we must show each case to either involve a contradiction, thereby indicating that case does not occur, or derive non-negative multipliers which satisfy the Kuhn-Tucker conditions, we present the following analysis:

$\boxed{\text{Case I}}$: $[csc1 \oplus csc2] \Longrightarrow [\lambda_1 = \lambda_2 = 0] \oplus [kti1] \Longrightarrow \left[-1 = 0 \pitchfork\right] \Longrightarrow dno$

$\boxed{\text{Case II}}$: $csc1 \Longrightarrow [\lambda_1 = 0] \oplus [kti1, kti2] \Longrightarrow [\lambda_2 = 1 > 0, \lambda_2 x_2 = 0] \Longrightarrow$

$[x_2 = 0] \oplus \left[g_2 = x_1 - (0)^2 = 0\right] \Longrightarrow csc2$ satisfied $\Longrightarrow$
$\left[x^A = (0,0)^T \text{ is a valid Kuhn-Tucker point}\right]$

$\boxed{\text{Case III}}$: $csc2 \Longrightarrow [\lambda_2 = 0] \oplus [kti1, kti2] \Longrightarrow [-1 + 2\lambda_1 x_1 = 0, \lambda_1 x_2 = 0] \Longrightarrow$

$\boxed{\text{Subcase IIIA}}$: $[\lambda_1 = 0] \oplus [-1 + 2\lambda_1 x_1 = 0] \Longrightarrow \left[-1 = 0 \pitchfork\right] \Longrightarrow dno$

$\boxed{\text{Subcase IIIB}}$: $[\lambda_1 > 0] \oplus [\lambda_1 x_2 = 0] \Longrightarrow [x_2 = 0] \oplus$

$$\left[g_1 = (x_1)^2 + (0)^2 - 2 = 0\right] \Longrightarrow$$

$$\left[x_1 = \pm\sqrt{2}\right] \oplus \left[g_2 = x_1 - (0)^2 \le 0\right] \Longrightarrow \left[x_1 = -\sqrt{2}\right] \oplus [-1 + 2\lambda_1 x_1 = 0] \Longrightarrow$$

$$\left[0 \le \lambda_1 = (2x_1)^{-1} = \left(-2\sqrt{2}\right)^{-1} < 0 \pitchfork\right] \Longrightarrow dno$$

$\boxed{\text{Case IV}}$: $\left[g_1 = (x_1)^2 + (x_2)^2 - 2 = 0\right] \oplus \left[g_2 = x_1 - (x_2)^2 = 0\right] \Longrightarrow$

$[x_1 = 1, x_2 = \pm 1] \oplus [kti1, kti2] \Longrightarrow [-1 + 2\lambda_1 + \lambda_2 = 0, \lambda_1 - \lambda_2 = 0] \Longrightarrow$

$[\lambda_1 = 1/3 > 0, \lambda_2 = 1/3 > 0] \Longrightarrow csc1$ and $csc2$ satisfied $\Longrightarrow$

$\left[x^B = (1,1)^T, x^C = (1,-1)^T \text{ are valid Kuhn-Tucker points}\right]$.

The global optimum is found by noting

$$\begin{aligned} f\left(x^A\right) &= 0 \\ f\left(x^B\right) = f\left(x^C\right) &= -1 < f\left(x^A\right) \end{aligned} \tag{2.170}$$

which means x^B, x^C are alternative global minimizers. Note also that x^A is not a local minimizer.

2.8.4 A Convex, Nongraphical Example

Let us now consider the mathematical program

$$\min f(x_1, x_2) = 0x_1 - x_2 \tag{2.171}$$

subject to

$$g_1(x_1, x_2) = (x_1)^2 + (x_2)^2 - 2 \leq 0 \tag{2.172}$$

$$g_2(x_1, x_2) = -x_1 + x_2 \leq 0 \tag{2.173}$$

Note that this problem is a convex mathematical program since the objective function is linear and the inequality constraint functions are convex. We know the Kuhn-Tucker conditions will be both necessary and sufficient for a nonunique global minimum. This means that we need only find one case of binding and nonbinding constraints that leads to nonnegative inequality constraint multipliers in order to solve (2.171), (2.172), and (2.173) to global optimality. We begin by observing that

$$\nabla f(x_1, x_2) = \begin{pmatrix} 0 \\ -1 \end{pmatrix} \tag{2.174}$$

$$\nabla g_1(x_1, x_2) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} \tag{2.175}$$

$$\nabla g_2(x_1, x_2) = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \tag{2.176}$$

The Kuhn-Tucker identity is

$$\begin{pmatrix} 0 \\ -1 \end{pmatrix} + \lambda_1 \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} + \lambda_2 \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{2.177}$$

from which we obtain the equations

$$kti1 : 2\lambda_1 x_1 - \lambda_2 = 0 \tag{2.178}$$

$$kti2 : -1 + 2\lambda_1 x_2 + \lambda_2 = 0 \tag{2.179}$$

The complementary slackness conditions are

$$csc1 : \lambda_1 g_1 (x_1, x_2) = 0 \quad \lambda_1 \geq 0 \tag{2.180}$$

$$csc2 : \lambda_2 g_2 (x_1, x_2) = 0 \quad \lambda_2 \geq 0 \tag{2.181}$$

Since the present mathematical program has two constraints, the table (2.168) still applies. Let us posit that both constraints are binding, so that the following analysis applies:

$\boxed{\text{Case IV}}$: $\left[g_1 = (x_1)^2 + (x_2)^2 - 2 = 0\right] \oplus [g_2 = -x_1 + x_2 = 0] \Longrightarrow$

$\left[x_1^* = x_2^* = 1\right] \oplus [kti1, kti2] \Longrightarrow [2\lambda_1 - \lambda_2 = 0, -1 + 2\lambda_1 + \lambda_2 = 0] \Longrightarrow$

$\left[\lambda_1 = \dfrac{1}{4} > 0, \lambda_2 = \dfrac{1}{2} > 0\right] \Longrightarrow [csc1 \text{ and } csc2 \text{ are satisfied}] \Longrightarrow$

$\left[x^* = \left(x_1^*, x_2^*\right)^T = (1,1)^T \text{ is a global minimizer}\right]$

However, since the objective function is only convex and not strictly convex, we cannot ascertain without analyzing the three remaining cases whether this global minimizer is unique. The reader may verify that the other three cases lead to contradictions, and thereby determine that $x^* = (1,1)^T$ is a unique global solution.

2.9 Discrete-Time Optimal Control

We are now ready to formulate a fairly general version of the discrete-time optimal control problem. Because time is treated discretely, we avoid in this initial foray into optimal control theory the complications and nuances of infinite-dimensional vector spaces. In particular, we will show that the discrete-time optimal control problem can be restated as a nonlinear mathematical program in standard form. We then show that application of the Kuhn-Tucker conditions leads us directly to a discrete version of Pontryagin's minimum (maximum) principle and the other necessary conditions of discrete-time optimal control.

The *equations of motion*, also called the *dynamics*, that we consider take the form of the following difference equations:

$$x_{t+1} = x_t + f_t(x_t, u_t) \quad t = 0, 1, \ldots, q-1 \tag{2.182}$$

where t is a discrete time index (a nonnegative integer) and q is the number of time steps that constitute our *planning or analysis horizon*. Note further that $x_t \in \Re^n$ and $u_t \in \Re^r$ are vectors, as is $f_t : \Re^n \times \Re^r \longrightarrow \Re^n$. We refer to the x_t as state variables and the u_t as control variables. We assume f_t is continuously differentiable on $\Re^n \times \Re^r$. The initial and terminal conditions for these dynamics are taken, respectively, to be

$$\Phi_0(x_0) = 0 \tag{2.183}$$

$$\Phi_q(x_q) = 0 \tag{2.184}$$

where $\Phi_0 : \Re^n \longrightarrow \Re^{m_0}$ and $\Phi_q : \Re^n \longrightarrow \Re^{m_q}$, while Φ_0 and Φ_q are both $\mathcal{C}^1$ on $\Re^n$. The *control constraints* are stated in abstract form as

$$u_t \in \mathcal{U}_t \equiv \{u : g_t(u_t) \leq 0\} \subseteq \Re^r \quad t = 0, 1, \ldots, q-1 \tag{2.185}$$

where $g_t : \Re^r \longrightarrow \Re^s$ and g_t is $\mathcal{C}^1$ on $\Re^r$. As stressed in our development of the Kuhn-Tucker conditions, there is no loss of generality arising from the fact that we have only explicitly considered inequality constraints on the controls, as any equality constraint may be represented by two appropriately defined inequalities. The final piece of the discrete-time optimal control problem is its cost function defined by

$$J = \Psi(x_q) + \sum_{t=0}^{q-1} F_t(x_t, u_t) \tag{2.186}$$

where $\Psi : \Re^n \longrightarrow \Re^1$ is $\mathcal{C}^1$ on $\Re^n$, while $F_t : \Re^n \times \Re^r \longrightarrow \Re^1$ is $\mathcal{C}^1$ on $\Re^n \times \Re^r$. We assume that J is meant to be minimized.

Assembling the individual pieces presented above, we have the following canonical form of the discrete-time optimal control problem:

$$\min J = \Psi(x_q) + \sum_{t=0}^{q-1} F_t(x_t, u_t) \tag{2.187}$$

subject to

$$x_{t+1} = x_t + f_t(x_t, u_t) \quad t = 0, 1, ..., q-1 \tag{2.188}$$

$$u_t \in \mathcal{U}_t = \{u : g_t(u_t) \leq 0\} \subseteq \Re^r \quad t = 0, 1, \ldots, q-1 \tag{2.189}$$

$$\Phi_0(x_0) = 0 \quad \Phi_q\left(x_q\right) = 0 \tag{2.190}$$

Note that we have included no constraints involving the state variables.

2.9.1 Necessary Conditions

It should be apparent that the discrete-time optimal control problem given by (2.187), (2.188), (2.189), and (2.190) is a finite-dimensional nonlinear mathematical program. Let us put it in the following form:

$$\min Z(x, u) = \Psi\left(x_q\right) + \sum_{t=0}^{q-1} F_t(x_t, u_t) \tag{2.191}$$

subject to

$$h_t(x_t, x_{t+1}) = -x_{t+1} + x_t + f_t(x_t, u_t) = 0 \quad (\tau_{t+1}) \qquad t = 0, 1, ..., q-1 \tag{2.192}$$

$$g_t(u_t) \leq 0 \quad (\lambda_t) \quad t = 0, 1, ..., q-1 \tag{2.193}$$

$$\Phi_0(x_0) = 0 \quad (\rho_0) \tag{2.194}$$

$$\Phi_q\left(x_q\right) = 0 \quad \left(\rho_q\right) \tag{2.195}$$

where for convenience we employ the following notation

$$x = \begin{pmatrix} x_0 \\ \vdots \\ x_q \end{pmatrix} \in \Re^{n(q+1)} \tag{2.196}$$

$$u = \begin{pmatrix} u_0 \\ \vdots \\ u_{q-1} \end{pmatrix} \in \Re^{rq} \tag{2.197}$$

for the vectors of decision variables for our mathematical program; as we have mentioned, in the parlance of optimal control theory, these vectors are vectors of state variables and control variables, respectively.

We assume that a relevant constraint qualification is in force so that the Kuhn-Tucker conditions for the mathematical program (2.191), (2.192), (2.193), (2.194), and (2.195) are a valid characterization of optimality. The names of multipliers for the constraints of (2.192), (2.193), (2.194), and (2.195) are indicated in parentheses next to each constraint. We state the Kuhn-Tucker conditions by first forming the Lagrangean; that is, we price out all constraints and adjoin them to the original objective function to obtain

$$\begin{aligned}\mathcal{L}(x,u,\rho,\tau) = \Psi\left(x_q\right) + \sum_{t=0}^{q-1} F_t(x_t,u_t) + \rho_0\Phi_0(x_0) + \rho_q\Phi_q\left(x_q\right) \\ + \sum_{t=0}^{q-1} \tau_{t+1}^T\left(-x_{t+1} + x_t + f_t(x_t,u_t)\right) + \sum_{t=0}^{q-1}\lambda_t^T g_t(u_t)\end{aligned}$$

where the symbol T denotes the transpose operation and

$$\rho = \begin{pmatrix}\rho_0 \\ \rho_q\end{pmatrix} \in \Re^{m_0+m_q}$$

$$\tau = \begin{pmatrix}\tau_1 \\ \cdot \\ \cdot \\ \cdot \\ \tau_q\end{pmatrix} \in \Re^{nq}$$

are vectors of dual variables (ρ) and *adjoint variables*[1] (τ), respectively.

The Kuhn-Tucker identity is, of course, obtained by setting the partial derivatives of $\mathcal{L}(x,u,\rho,\tau)$ equal to zero; let us begin with the following:

$$\nabla_x\mathcal{L}(x,u,\rho,\tau) = 0 \tag{2.198}$$

It follows that

$$\frac{\partial\mathcal{L}}{\partial x_0} = \frac{\partial F_0}{\partial x_0} + \rho_0\frac{\partial\Phi_0}{\partial x_0} + \tau_1^T + \tau_1^T\frac{\partial f_0}{\partial x_0} = 0 \tag{2.199}$$

$$\frac{\partial\mathcal{L}}{\partial x_t} = \frac{\partial F_t}{\partial x_t} - \tau_t^T + \tau_{t+1}^T + \tau_{t+1}^T\frac{\partial f_t}{\partial x_t} = 0 \tag{2.200}$$

If we agree to define

$$\tau_0 = -\left[\frac{\partial\Phi_0}{\partial x_0}\right]^T\rho_0 \tag{2.201}$$

[1] These discrete-time adjoint variables are clearly mathematical programming dual variables; in optimal control theory, we refer to them as adjoint variables by tradition.

then (2.199) and (2.200) can be written as

$$\tau_0 = \tau_1 + \left[\frac{\partial f_0}{\partial x_0}\right]^T \tau_1 + \nabla_{x_0} F_0 \tag{2.202}$$

$$\tau_t = \tau_{t+1} + \left[\frac{\partial f_t}{\partial x_t}\right]^T \tau_{t+1} + \nabla_{x_t} F_t \quad t = 1, \ldots, q-1 \tag{2.203}$$

We note that (2.203) and (2.204) have the same form as one another, so they may be conveniently represented by the single statement

$$\tau_t = \tau_{t+1} + \left[\frac{\partial f_t}{\partial x_t}\right]^T \tau_{t+1} + \nabla_{x_t} F_t \quad t = 0, \ldots, q-1 \tag{2.204}$$

Next note that

$$\frac{\partial \mathcal{L}}{\partial x_q} = \frac{\partial \Psi}{\partial x_q} + \rho_q^T \frac{\partial \Phi_q}{\partial x_q} - \tau_q^T = 0 \tag{2.205}$$

which can be rewritten as

$$\tau_q = \nabla_{x_q} \Psi + \left[\frac{\partial \Phi_q}{\partial x_q}\right]^T \rho_q \tag{2.206}$$

The remaining partial derivatives of interest are those of the Lagrangean with respect to the control variables, which are set to zero:

$$\nabla_u \mathcal{L}(x, u, \rho, \tau) = 0 \tag{2.207}$$

It follows that

$$\frac{\partial \mathcal{L}}{\partial u_t} = \lambda_t^T \frac{\partial g_t}{\partial u_t} + \tau_{t+1}^T \frac{\partial f_t}{\partial u_t} + \frac{\partial F_t}{\partial u_t} = 0 \quad t = 0, \ldots, q-1 \tag{2.208}$$

which can be rewritten as

$$\nabla_{u_t} \left[\sum_{t=0}^{q-1} F_t(x_t, u_t) + \sum_{t=0}^{q-1} \tau_{t+1}^T \left(-x_{t+1} + x_t + f_t(x_t, u_t)\right) \right.$$
$$\left. + \sum_{t=0}^{q-1} \lambda_t^T g_t(u_t) \right] = 0 \tag{2.209}$$

The final conditions for us to mention are

$$\lambda_t^T g_t = 0 \quad \lambda_t \geq 0 \quad t = 1, \ldots, q-1 \tag{2.210}$$

which are recognized as the complementary slackness conditions associated with the control inequality constraints and their multipliers.

In deriving the equations and inequalities of this section that express the necessary conditions, the arguments of all functions and their derivatives have been purposely omitted in order to simplify the notation. The complete set of necessary conditions for the discrete-time optimal control problem consist of the original problem constraints together with the conditions we have derived. That is to say, the necessary conditions are

$$\begin{aligned}
\text{equations of motion} &: \quad (2.182)\\
\text{initial conditions} &: \quad (2.183)\\
\text{terminal conditions} &: \quad (2.184)\\
\text{control constraints} &: \quad (2.185)\\
\text{adjoint equations} &: \quad (2.202)\\
\text{transversality conditions} &: \quad (2.206)\\
\text{stationarity conditions for the controls} &: \quad (2.209)
\end{aligned}$$

Note that these conditions constitute a so-called *two-point boundary-value problem*.

2.9.2 The Minimum Principle

In this section we wish to manipulate the necessary conditions for the discrete-time optimal control problem developed from application of the Kuhn-Tucker conditions into the traditional form used to study and analyze optimal control problems; in the process we will articulate Pontryagin's *minimum principle*. The mathematics of this section are essentially algebra and some simple differentiation; the substantive aspect of the discrete-time optimal control problem analysis has already been completed in the previous section. However, the success of modern optimal control theory is in no small part due to the elegant, concise statement of the necessary conditions that we are about to give (and which is usually attributed to Pontryagin and his colleagues); packaging is important!

We begin the task of reformulating the necessary conditions by defining the *Hamiltonian*:

$$H_t\left(x_t, \tau_{t+1}, u_t\right) \equiv F_t\left(x_t, u_t\right) + \tau_{t+1}^T f_t\left(x_t, u_t\right) \quad t = 0, \ldots, q-1 \tag{2.211}$$

where $x_t \in \Re^n$ will be called the *state variable vector* while τ_t and u_t were named, in Section 2.9.1, the *adjoint vector* and *control vector*, respectively; furthermore $H_t : \Re^n \times \Re^n \times \Re^r \longrightarrow \Re^1$. It is immediate that the equations of motion may be stated as

$$x_{t+1} - x_t = \nabla_{\tau_{t+1}} H_t\left(x_t, \tau_{t+1}, u_t\right) \quad t = 0, \ldots, q-1 \tag{2.212}$$

and the adjoint equations as

$$\tau_{t+1} - \tau_t = -\nabla_{x_t} H_t (x_t, \tau_{t+1}, u_t) \quad t = 0, \ldots, q-1 \tag{2.213}$$

Results (2.212) and (2.213) are completely analogous to Hamilton's equations of classical mechanics that describe conservative Newtonian systems in terms of generalized coordinates (position and momentum). For this reason, these equations are sometimes still called Hamilton's equations, although there is no implication that (2.212) and (2.213) carry with them any of the assumptions or implications of classical mechanics.

We may also, in light of the definition of the Hamiltonian (2.211), restate the stationarity conditions for the optimal controls as

$$\nabla_{u_t} \left[H_t (x_t, \tau_{t+1}, u_t) + \sum_{t=0}^{q-1} \lambda_t^T g_t (u_t) \right] = 0 \quad t = 0, \ldots, q-1 \tag{2.214}$$

$$\lambda_t^T g_t (u_t) = 0 \quad \lambda_t \geq 0 \quad t = 0, \ldots, q-1 \tag{2.215}$$

The system (2.214) and (2.215) is immediately recognized as the necessary conditions for statically minimizing the Hamiltonian with respect to the controls under the assumption that all other variables are held fixed. We restate this observation as

$$H_t (x_t, \tau_{t+1}, u_t) \leq H_t (x_t, \tau_{t+1}, u) \quad \forall u \in \mathcal{U}_t \quad t = 0, \ldots, q-1 \tag{2.216}$$

Expression (2.216) is Pontryagin's minimum principle.

2.9.3 Discrete Optimal Control Example

Consider the following discrete-time optimal control problem:

$$\min \ J = \sum_{t=0}^{5} \frac{1}{2} (x_t)^2 \tag{2.217}$$

subject to

$$x_{t+1} - x_t = u_t \qquad t = 0, 1, 2, 3, 4 \tag{2.218}$$

$$x_0 = 3 \tag{2.219}$$

$$-1 \leq u_t \leq 1 \quad t = 0, 1, 2, 3, 4 \tag{2.220}$$

The Hamiltonian is

$$H_t = \frac{1}{2} (x_t)^2 + \lambda_{t+1} (u_t) \quad t = 0, 1, 2, 3, 4$$

The minimum principle is

$$u_t = \begin{cases} +1 & \lambda_{t+1} < 0 \\ u_t^s & \lambda_{t+1} = 0 \\ -1 & \lambda_{t+1} > 0 \end{cases} \qquad t = 0, 1, 2, 3, 4$$

The adjoint equations are

$$\begin{aligned} \lambda_{t+1} - \lambda_t &= -\nabla_{x_t} H_t(x_t, \lambda_{t+1}, u_t) \\ &= -x_t \qquad\qquad\qquad\qquad t = 0, 1, 2, 3, 4 \end{aligned}$$

Inspection indicates that the objective function will be minimized by the application of the control $u_t = -1$ until the state variable reaches zero at an unknown time t_1; thereafter a so-called singular control $u_t^s = 0$ is applied, until the end of time horizon. Then

$$x_{t+1} - x_t = u_t = -1 \qquad t = 0, 1, \ldots, t_1$$

Since $x_0 = 3$ is given, we have

$$\begin{aligned} x_1 &= x_0 - 1 = 2 \\ x_2 &= x_1 - 1 = 1 \\ x_3 &= x_2 - 1 = 0 \end{aligned}$$

Consequently it is discovered that

$$t_1 = 2$$

Following the prior assumption, we find that

$$x_{t+1} - x_t = u_t = 0 \qquad t = 3, 4$$

which yields

$$x_4 = x_5 = 0$$

Now let us consider the conditions for adjoint variables. According to the minimum principle, we should have $\lambda_t > 0$ for $t = 0, 1, 2$ in order that $u_t = -1$ for the same time intervals. From the transversality conditions and the adjoint equations, we have

$$\begin{aligned} \lambda_5 &= 0 \\ \lambda_4 &= \lambda_5 + x_4 = 0 \\ \lambda_3 &= \lambda_4 + x_3 = 0 \\ \lambda_2 &= \lambda_3 + x_2 = 1 \\ \lambda_1 &= \lambda_2 + x_1 = 3 \\ \lambda_0 &= \lambda_1 + x_0 = 6 \end{aligned}$$

which satisfies the minimum principle. In summary, the solution is

t	0	1	2	3	4	5
x_t	$+3$	$+2$	$+1$	0	0	0
u_t	-1	-1	-1	0	0	0
λ_t	$+6$	$+3$	$+1$	0	0	0

It is instructive to approach the same problem from a purely mathematical programming perspective. In fact off-the-shelf finite-dimensional mathematical programming software or the Kuhn-Tucker conditions (without invoking the notion of the Hamiltonian and the minimum principle) may be applied directly to the nonlinear program (2.217), (2.218), (2.219), and (2.220). We leave the demonstration that the mathematical programming approach yields an identical result as an exercise for the reader.

2.10 Exercises

1. Create an example of a mathematical program with two decision variables for which no constraint qualification exists.
2. Prove or disprove: a nonlinear program with a strictly convex objective function and a non-convex feasible region arising from constraints satisfying the linear independence constraint qualification may never have a unique global optimum.
3. Solve the following nonconvex, nonlinear program graphically:

$$\min f(x_1, x_2) = -x_1 + 0x_2$$

   subject to

$$g_1(x_1, x_2) = (x_1)^2 + (x_2)^2 - 2 \leq 0$$
$$g_2(x_1, x_2) = x_1 - (x_2)^2 \leq 0$$

4. Solve the nonconvex, nonlinear program of Exercise 3 above using the Kuhn-Tucker conditions without appeal to graphical information.
5. The example of Section 2.9.3 suggests that a *singular control* arises when it appears linearly in the Hamiltonian and has a coefficient that vanishes. Propose an alternative definition that relies on the language and optimality conditions of nonlinear programming.
6. Use the minimum principle to solve the following discrete-time optimal control problem:

$$\min J = \sum_{t=0}^{5} \left[\frac{1}{2}(x_t)^2 + u_t \right]$$

subject to

$$
\begin{aligned}
x_{t+1} - x_t &= u_t \qquad t = 0, 1, 2, 3, 4 \\
x_0 &= 3 \\
-1 \le u_t &\le 1 \quad t = 0, 1, 2, 3, 4
\end{aligned}
$$

7. Use the minimum principle to solve the following discrete-time optimal control problem:

$$
\min J = \sum_{t=0}^{5} \left[\frac{1}{2}(x_t)^2 + \frac{1}{2}(u_t)^2 \right]
$$

subject to

$$
\begin{aligned}
x_{t+1} - x_t &= u_t \qquad t = 0, 1, 2, 3, 4 \\
x_0 &= 3 \\
-1 \le u_t &\le 1 \quad t = 0, 1, 2, 3, 4
\end{aligned}
$$

List of References Cited and Additional Reading

Bazaraa, M., H. Sherali, and C. Shetty (1993). *Nonlinear Programming: Theory and Algorithms*. New York: John Wiley.

Canon, M., C. Cullum, and E. Polak (1970). *Theory of Optimal Control and Mathematical Programming*. New York: McGraw-Hill.

Luenberger, D. G. (1984). *Linear and Nonlinear Programming*. Reading, MA: Addison-Wesley.

Mangasarian, O. (1969). *Nonlinear Programming*. New York: McGraw-Hill.

http://www.springer.com/978-0-387-72777-6

Dynamic Optimization and Differential Games
Friesz, T.L.
2010, XIV, 502 p., Hardcover
ISBN: 978-0-387-72777-6