2

Kuhn–Tucker Conditions

In this chapter, necessary conditions for optimality of solution points in mathematical programming problems will be studied. Because of the orientation of this book to present optimization theory as an instrument for qualitative economic analysis, the theory to be described is not immediately concerned with computational aspects of solution techniques, which can be found in many excellent books on mathematical programming, e.g., [11, 12, 27, 23, 3].

The discussion begins with the extension of the Lagrange theory by Kuhn and Tucker [18]—note the contributions by Karush [16] and John [15]—with the derivation of necessary optimality conditions for the optimization problems including inequality constraints.

The rationality of Kuhn–Tucker conditions and their relationship to a saddle point of the Lagrangian function will be explored in Sections 2.2 and 2.3, respectively.

Section 2.4 deals with Kuhn–Tucker conditions for the general mathematical programming problem, including equality and inequality constraints, as well as nonnegative and free variables. Two numerical examples are provided for illustration.

Section 2.5 is devoted to applications of Kuhn–Tucker conditions to a qualitative economic analysis. We will show how to derive general qualitative conclusions, even when the parameters of the involved functions are not numerically specified.

2.1 The Kuhn–Tucker Theorem

The basic mathematical programming problem (1.28), as described in Chapter 1, is that of choosing values of n variables so as to minimize a function of those variables subject to m inequality constraints:

$$\begin{aligned} &\text{minimize} \quad f_0(\mathbf{x}) \\ &\text{subject to} \quad f_i(\mathbf{x}) \leqq 0 \quad (i = 1, 2, \ldots, m). \end{aligned}$$

This problem is a generalization of the classical optimization problem (which uses constraints in equation form), since equality constraints are a special case of inequality

M. Luptáčik, *Mathematical Optimization and Economic Analysis*,
Springer Optimization and Its Applications 36, DOI 10.1007/978-0-387-89552-9_2,

constraints. By introducing m additional variables, called slack variables, $y_i (i = 1, 2, \dots, m)$, the mathematical programming problem (1.28) can be rewritten as a classical optimization problem:

$$\begin{aligned} &\text{minimize} && f_0(\mathbf{x}) \\ &\text{subject to} && f_i(\mathbf{x}) + y_i^2 = 0 \quad (i = 1, 2, \dots, m). \end{aligned}$$

A characterization of the solution to the mathematical programming problem (1.28) is then analogous to the Lagrange theorem for classical optimization problems.

Under the assumption of so-called constraint qualifications (for a detailed discussion, the reader is referred to [1, 26, 37]), which was designed to avoid cusps in the feasible set, the Lagrange theory for a classical optimization problem can be extended to problem (1.28) by the following theorem.

Theorem 2.1 (see [18]). *Assume that $f_k(\mathbf{x})$ $(k = 0, 1, \dots, m)$ are all differentiable. If the function $f_0(\mathbf{x})$ attains at point $\mathbf{x}^0$ a local minimum subject to the set $K = \{\mathbf{x} | f_i(\mathbf{x}) \leqq 0 \ (i = 1, 2, \dots, m)\}$, then there exists a vector of Lagrange multipliers $\mathbf{u}^0$ such that the following conditions are satisfied:*

$$\frac{\partial f_0(\mathbf{x}^0)}{\partial x_j} + \sum_{i=1}^{m} u_i^0 \frac{\partial f_i(\mathbf{x}^0)}{\partial x_j} = 0 \quad (j = 1, 2, \dots, n), \tag{2.1}$$

$$f_i(\mathbf{x}^0) \leqq 0 \quad (i = 1, 2, \dots, m), \tag{2.2}$$

$$u_i^0 f_i(\mathbf{x}^0) = 0 \quad (i = 1, 2, \dots, m), \tag{2.3}$$

$$u_i^0 \geqq 0 \quad (i = 1, 2, \dots, m). \tag{2.4}$$

In other words, the conditions (2.1)–(2.4) *are necessary conditions for a local minimum of problem* (1.28). *For a maximization problem, the nonnegativity condition* (2.4) *is replaced by the nonpositivity condition* $\mathbf{u}^0 \leqq \mathbf{0}$. *Conditions* (2.1)–(2.4) *are called the Kuhn–Tucker conditions.*

Proof. As in the case of the classical optimization problem, the Lagrange function can be defined as a function of the original variables—in our case the variables $\mathbf{x}$ and $\mathbf{y}$—and of the Lagrange multipliers $\mathbf{u}$:

$$L(\mathbf{x}, \mathbf{y}, \mathbf{u}) = f_0(\mathbf{x}) + \sum_{i=1}^{m} u_i (f_i(\mathbf{x}) + y_i^2).$$

The necessary conditions for its local minimum are

$$\frac{\partial L}{\partial x_j} = \frac{\partial f_0(\mathbf{x}^0)}{\partial x_j} + \sum_{i=1}^{m} u_i^0 \frac{\partial (f_i(\mathbf{x}^0) + (y_i^0)^2)}{\partial x_j} = 0 \quad (j = 1, 2, \dots, n), \tag{2.5}$$

$$\frac{\partial L}{\partial y_i} = 2u_i^0 y_i^0 = 0 \quad (i = 1, 2, \dots, m), \tag{2.6}$$

$$\frac{\partial L}{\partial u_i} = f_i(\mathbf{x}^0) + (y_i^0)^2 = 0 \quad (i = 1, 2, \ldots, m). \tag{2.7}$$

Now it can be shown that the conditions in (2.6) correspond to the Kuhn–Tucker conditions (2.3).

Suppose $u_i^0 = 0$. Then $u_i^0 y_i^0 = u_i^0 f_i(\mathbf{x}^0) = 0$, and both conditions (2.6) and (2.3) are satisfied.

If $u_i^0 \neq 0$, then it follows from (2.6) that $y_i^0 = 0$ and therefore $(y_i^0)^2 = -f_i(\mathbf{x}^0) = 0$: Condition (2.3) is satisfied. On the other hand, it follows from (2.3) that $f_i(\mathbf{x}^0) = 0$ and therefore $y_i^0 = 0$: Condition (2.6) is fulfilled.

Since the variables y_i $(i = 1, 2, \ldots, m)$ are auxiliary variables, they can be eliminated from conditions (2.5) and (2.7), and we obtain conditions (2.1) and (2.2).

It remains to show that the Lagrange multipliers must be nonnegative. For this purpose, we consider the classical optimization problem:

$$\begin{aligned} &\text{minimize} \quad f_o(\mathbf{x}) \\ &\text{subject to} \quad f_i(\mathbf{x}) \leqq b_i \quad (i = 1, 2, \ldots, m). \end{aligned} \tag{2.8}$$

For the Lagrange multipliers u_i^0 $(i = 1, 2, \ldots, m)$ of problem (2.8), the following holds (see, e.g., [23, 1st ed., p. 231]):

$$\frac{\partial f_0(\mathbf{x}^0(\mathbf{b}))}{\partial b_i} = -u_i^0 \quad (i = 1, 2, \ldots, m), \tag{2.9}$$

where $\mathbf{x}^0$ denotes the optimal solution of problem (2.8). Hence the Lagrange multipliers u_i^0 $(i = 1, 2, \ldots, m)$ give us the change of the value of the objective function due to a change of the constraint b_i by a small amount. A higher value of the ith component of the vector $\mathbf{b}$ implies an enlargement of the set K. Therefore, the new optimal value of the objective function $f_0(\mathbf{x})$ cannot be worse:

$$\frac{\partial f_0(x^0)}{\partial b_i} \leqq 0 \quad \text{for a minimization problem} \tag{2.10}$$

and

$$\frac{\partial f_0(\mathbf{x}^0)}{\partial b_i} \geqq 0 \quad \text{for a maximization problem.} \tag{2.11}$$

The nonnegativity condition for the Lagrange multipliers (2.4) follows from (2.9) and (2.10). Similarly, conditions in (2.9) and (2.11) imply that the Lagrange multipliers cannot be positive for problem (1.28) with the objective function to be maximized. □

For a geometric interpretation of the Kuhn–Tucker conditions (2.1)–(2.4), we rewrite the conditions in (2.1) as follows:

$$\frac{\partial f_0(\mathbf{x}^0)}{\partial x_j} = -\sum_{i=1}^{m} u_i^0 \frac{\partial f_i(\mathbf{x}^0)}{\partial x_j} \quad (j = 1, 2, \ldots, n),$$

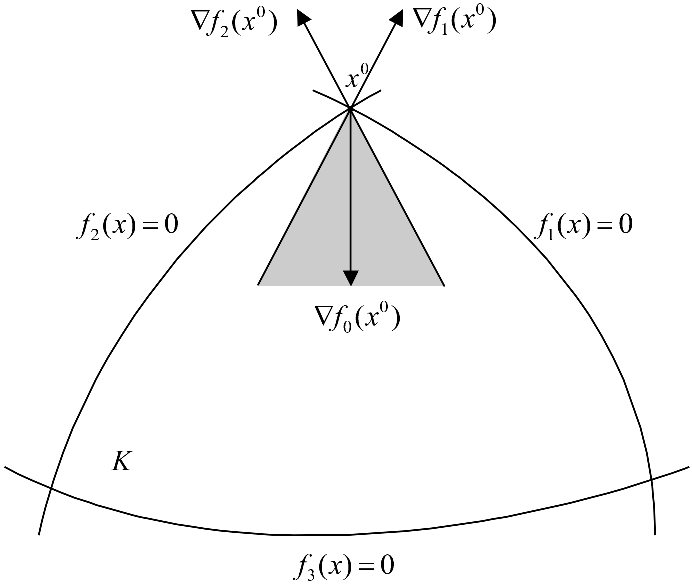

Fig. 2.1. Kuhn–Tucker conditions.

or

$$\nabla f_0(\mathbf{x}^0) = -\sum_{i=1}^{m} u_i^0 \nabla f_i(\mathbf{x}^0),$$

where $\nabla f_0(\mathbf{x})$ denotes the gradient vector (the vector of first-order partial derivatives) of the objective function, and $\nabla f_i(\mathbf{x})$ is the gradient vector of the ith constraint function ($i = 1, 2, \ldots, m$). Thus the gradient of the objective function must, at the optimal solution, be a nonpositive weighted combination of the gradients of the active constraints (the constraints satisfied at the optimal solution as equalities). The gradient vector of the objective function must therefore lie within the cone spanned by the inward-pointing normals to the opportunity set at $\mathbf{x}^0$. This solution is illustrated in Figure 2.1 for the problem in which $n = 2$, $m = 3$.

Using the Lagrange function (without slack variables) for the mathematical programming problem (1.28),

$$\Phi(\mathbf{x}, \mathbf{u}) = f_0(\mathbf{x}) + \sum_{i=1}^{m} u_i f_i(\mathbf{x}), \tag{2.12}$$

the Kuhn–Tucker conditions (2.1)–(2.4) can be rewritten as follows:

$$\frac{\partial \Phi(\mathbf{x}^0, \mathbf{u}^0)}{\partial x_j} = 0 \quad (j = 1, 2, \ldots, n), \tag{2.1$'$}$$

$$\frac{\partial \Phi(\mathbf{x}^0, \mathbf{u}^0)}{\partial u_i} \leqq 0 \quad (i = 1, 2, \ldots, m), \tag{2.2$'$}$$

$$u_i^0 \frac{\partial \Phi(\mathbf{x}^0, \mathbf{u}^0)}{\partial u_i} = 0 \quad (i = 1, 2, \ldots, m), \tag{2.3$'$}$$

$$u_i^0 \geqq 0 \quad (i = 1, 2, \ldots, m). \tag{2.4$'$}$$

The n conditions in (2.1$'$) are the same as in the classical programming case, or in other words, as in the traditional Lagrange theory from classical differential calculus.

The m conditions in (2.2$'$) are the constraints of the mathematical programming problem which permits solution at the boundary of the set of feasible solutions or at an interior point of this set.

The m conditions in (2.3$'$), which are known as the complementary slackness conditions of mathematical programming, serve essentially to determine which of the two regimes will apply: whether the boundary or the interior minimum point occurs. If the ith constraint is not binding (an interior point), then the corresponding Lagrange multiplier will be zero. If the multiplier u_i is positive, then the corresponding ith constraint is binding (boundary solution). The reader should bear in mind that the converse is not true.

The m conditions in (2.4$'$), requiring that the Lagrange multipliers be nonnegative, stem from the fact that the constraints in (2.2$'$) are written as inequalities rather than as equalities; if a constraint is an equality, then the corresponding element of $\mathbf{u}^0$ is unrestricted, as in the classical programming case.

In most of the models of mathematical programming in economics (see Chapter 1), nonnegativity conditions are required. Obviously, it would be possible to include nonnegativity conditions in the set of constraints $f_i(\mathbf{x}) \leqq 0$ $(i = 1, 2, \ldots, m)$. But as we will show now, the Lagrange multipliers corresponding to the nonnegativity conditions can be eliminated. It is therefore useful to consider the nonnegativity conditions separately.

We consider the following mathematical programming problem:

$$\begin{aligned} \text{minimize} \quad & f_0(\mathbf{x}) \\ \text{subject to} \quad & f_i(\mathbf{x}) \leqq 0 \quad (i = 1, 2, \ldots, m), \\ & -x_j \leqq 0 \quad (j = 1, 2, \ldots, n). \end{aligned} \tag{1.28a}$$

First, we write the Lagrange function for problem (1.28a):

$$\Psi(\mathbf{x}, \mathbf{u}, \mathbf{w}) = f_0(\mathbf{x}) + \sum_{i=1}^{m} u_i f_i(\mathbf{x}) + \sum_{j=1}^{n} w_j(-x_j).$$

Then according to Theorem 2.1, the Kuhn–Tucker conditions become

$$\frac{\partial \psi}{\partial x_j} = \frac{\partial f_0(\mathbf{x}^0)}{\partial x_j} + \sum_{i=1}^{m} u_i^0 \frac{\partial f_i(\mathbf{x}^0)}{\partial x_j} - w_j^0 = 0 \qquad (j = 1, 2, \ldots, n),$$

or, equivalently,

$$\frac{\partial f_0(\mathbf{x}^0)}{\partial x_j} + \sum_{i=1}^{m} u_i^0 \frac{\partial f_i(\mathbf{x}^0)}{\partial x_j} = w_j^0 \quad (j = 1, 2, \ldots, n). \tag{2.13}$$

Furthermore,

$$\frac{\partial \psi}{\partial u_i} = f_i(\mathbf{x}^0) \leqq 0 \qquad (i = 1, 2, \ldots, m), \tag{2.14}$$

$$u_i^0 \frac{\partial \psi}{\partial u_i} = u_i^0 f_i(\mathbf{x}^0) = 0 \qquad (i = 1, 2, \ldots, m), \tag{2.15}$$

$$\frac{\partial \psi}{\partial w_j} = -x_j^0 \leqq 0 \qquad (j = 1, 2, \ldots, n), \tag{2.16}$$

$$w_j^0 \frac{\partial \psi}{\partial w_j} = w_j^0(-x_j^0) = 0 \qquad (j = 1, 2, \ldots, n),$$

which, because of (2.13), can be rewritten as

$$x_j^0 \left(\frac{\partial f_0(\mathbf{x}^0)}{\partial x_j} + \sum_{i=1}^{m} u_i^0 \frac{\partial f_i(\mathbf{x}^0)}{\partial x_j} \right) = 0 \qquad (j = 1, 2, \ldots, n), \tag{2.17}$$

$$u_i^0 \geqq 0 \qquad (i = 1, 2, \ldots, m), \tag{2.18}$$

$$w_j^0 \geqq 0 \qquad (j = 1, 2, \ldots, n). \tag{2.19}$$

Using the Lagrange function (2.12), the Kuhn–Tucker conditions (2.13)–(2.19) can be summarized symmetrically with respect to **x** and **u** as

$$\frac{\partial \Phi(\mathbf{x}^0, \mathbf{u}^0)}{\partial \mathbf{x}} \geqq \mathbf{0}, \tag{2.20}$$

$$\mathbf{x}^0 \frac{\partial \Phi(\mathbf{x}^0, \mathbf{u}^0)}{\partial \mathbf{x}} = 0, \tag{2.21}$$

$$\mathbf{x}^0 \geqq \mathbf{0}, \tag{2.22}$$

$$\frac{\partial \Phi(\mathbf{x}^0, \mathbf{u}^0)}{\partial \mathbf{u}} \leqq \mathbf{0}, \tag{2.23}$$

$$\mathbf{u}^0 \frac{\partial \Phi(\mathbf{x}^0, \mathbf{u}^0)}{\partial \mathbf{u}} = 0, \tag{2.24}$$

$$\mathbf{u}^0 \geqq \mathbf{0}. \tag{2.25}$$

The reader will note that in the case of nonnegativity conditions for the variables **x**, condition (2.1) of the Kuhn–Tucker theorem has been replaced by two sets of conditions (2.20)–(2.21). An intuitive explanation of this matter will be given in the next section.

2.2 Rationale of the Kuhn–Tucker Conditions

As already mentioned, the Kuhn–Tucker conditions are the natural generalization of the Lagrange multiplier approach, from classical differential calculus replacing equality constraints by inequality constraints, to take account of the possibility that the maximum or minimum in question can occur not only at a boundary point but also at an interior point. The calculus requirements are generally appropriate only if the extremum (i.e., the maximum or minimum) occurs at a point at which all of the variables (including the slack variables) take nonzero values.

Now we consider—for simplicity, but without loss of generality—the minimization of the function $f(x)$ subject to $x \geqq 0$. In this case, the matter can be illustrated graphically. Suppose first that we are at a point at which the value of x can either be increased or decreased (the interior point A in Figure 2.2). By the usual logic of marginal analysis, we must have $\frac{df}{dx} = 0$, for otherwise either a rise or a fall in the value of x could increase the value of f, and f would not be at its minimum.

On the other hand, suppose we are testing for the possibility of a boundary minimum at which $x = 0$. In Figure 2.2, two possibilities for local minimum of the function $f(x)$ subject to $x \geqq 0$ can be observed. If $\frac{df}{dx} = 0$, the point with $x = 0$ (point B in Figure 2.2) may be a minimum for the usual reasons, and if $\frac{df}{dx} > 0$, it may be a minimum point simply because it is impossible to reduce the value of x any further (point C in Figure 2.2).

Direct generalization for the function with n variables leads to the following conclusions. Given a differentiable function $f(x_1, x_2, \dots, x_n)$,

- for an interior minimum (maximum), it is necessary that $\frac{\partial f}{\partial x_j} = 0 \; (j = 1, 2, \dots, n)$;
- for a boundary minimum, it is necessary that $\frac{\partial f}{\partial x_j} \geqq 0 \; (j = 1, 2, \dots, n)$.

The reader may check that—by the same reasoning—for a boundary maximum it is necessary that $\frac{\partial f}{\partial x_j} \leqq 0$.

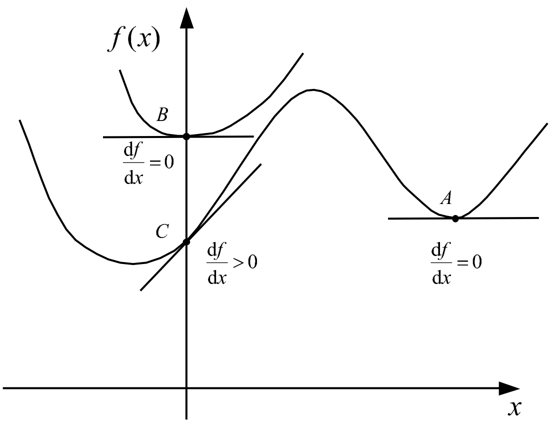

Fig. 2.2. Minimum of the function $f(x)$ subject to $x \geqq 0$.

Similar to the interpretation of the complementary slackness conditions (2.3′) or (2.24), the conditions in (2.21) serve to determine which solution case occurs; if the value of x_j under consideration is positive (interior minimum case), then (2.21) requires $\frac{\partial \Phi}{\partial x_j} = 0$. If $\frac{\partial \Phi}{\partial x_j} > 0$, then we can only have a boundary minimum ($x_j = 0$).

2.3 Kuhn–Tucker Conditions and a Saddle Point of the Lagrange Function

We consider the Lagrange function $\Phi(\mathbf{x}, \mathbf{u})$ as defined in (2.12). The necessary conditions for a local minimum of the Lagrange function (2.12), regarded as a function of $\mathbf{x}$ only, subject only to the nonnegativity conditions $x_j \geqq 0$ $(j = 1, 2, \ldots, n)$ are exactly the Kuhn–Tucker conditions (2.20)–(2.22) for problem (1.28a). At the same time, the Kuhn–Tucker conditions (2.23)–(2.25) provide the necessary conditions for a local maximum of the Lagrange function (2.12), regarded as a function of $\mathbf{u}$ only, subject only to the nonnegativity conditions $u_i \geqq 0$ $(i = 1, 2, \ldots, m)$. A graphical illustration of this property of the point $(\mathbf{x}^0, \mathbf{u}^0)$ from the Kuhn–Tucker conditions (2.20)–(2.25) is depicted in Figure 2.3. This leads to the following concept of a saddle point.

Definition 2.1. A point $(\mathbf{x}^0, \mathbf{u}^0)$ with $\mathbf{x}^0 \geqq \mathbf{0}$ and $\mathbf{u}^0 \geqq \mathbf{0}$ is said to be a *saddle point* of the Lagrange function $\Phi(\mathbf{x}, \mathbf{u})$ if

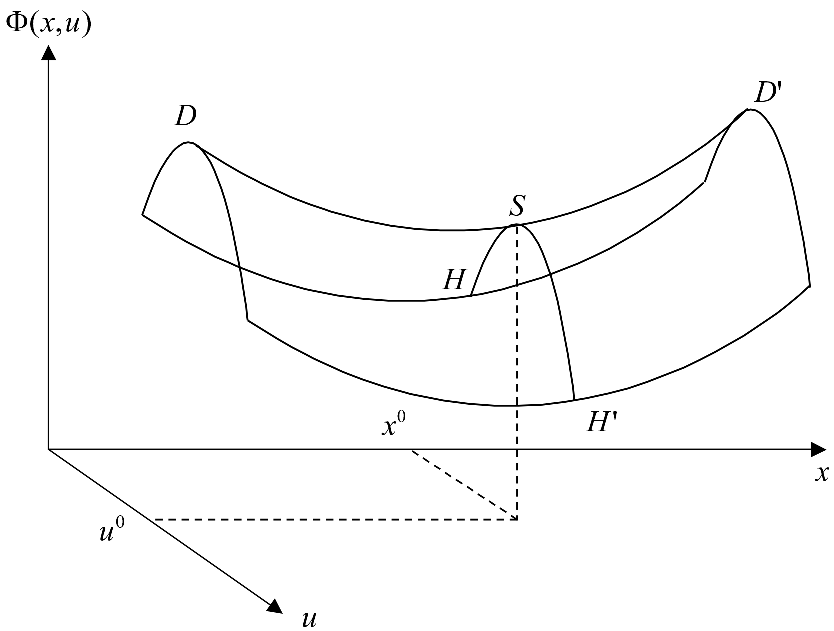

Fig. 2.3. Saddle point of the Lagrange function.

$$\Phi(\mathbf{x}^0, \mathbf{u}) \leqq \Phi(\mathbf{x}^0, \mathbf{u}^0) \leqq \Phi(\mathbf{x}, \mathbf{u}^0)$$

for all $\mathbf{x} \geqq \mathbf{0}$, $\mathbf{u} \geqq \mathbf{0}$.

In other words, for a fixed $\mathbf{u} = \mathbf{u}^0$, the Lagrange function is minimized at $\mathbf{x}^0$ (due to the second inequality of the relationship in Definition 2.1), whereas for a fixed $\mathbf{x} = \mathbf{x}^0$, the Lagrange function is maximized at $\mathbf{u}^0$ (which follows from the first inequality of the relationship in Definition 2.1).

Now the relationship between the saddle point of the Lagrange function and the optimal solution of problem (1.28a) can be established.

Theorem 2.2. *If there exists a saddle point* $(\mathbf{x}^0, \mathbf{u}^0)$ *of the Lagrange function* $\Phi(\mathbf{x}, \mathbf{u})$, *then* $\mathbf{x}^0$ *is an optimal solution for problem* (1.28a).[1]

In order to obtain the converse of Theorem 2.2, we need convexity properties of the functions $f_k(\mathbf{x})$ $(k = 0, 1, 2, \ldots, m)$, which will be discussed in the next chapter.

2.4 Kuhn–Tucker Conditions for the General Mathematical Programming Problem

The real applications of mathematical programming in economics contain both types of constraints: inequalities as well as equalities. Therefore, we define the general mathematical programming problem as follows:

$$\begin{aligned} \text{minimize} \quad & f_0(\mathbf{x}, \mathbf{y}) \\ \text{subject to} \quad & f_i(\mathbf{x}, \mathbf{y}) \leqq 0 \quad (i = 1, 2, \ldots, m), \\ & g_h(\mathbf{x}, \mathbf{y}) = 0 \quad (h = m+1, \ldots, r), \\ & \mathbf{x} \geqq \mathbf{0}, \\ & \mathbf{y} \in \mathbb{R}^l. \end{aligned} \tag{1.28b}$$

Obviously, problems (1.28) and (1.28a) are special cases of problem (1.28b).

Writing problem (1.28b) in the form (1.28) with $\Phi(\mathbf{x}, \mathbf{y}, \mathbf{u}, \mathbf{v}) = f_0(\mathbf{x}, \mathbf{y}) + \sum_{i=1}^{m} u_i f_i(\mathbf{x}, \mathbf{y}) + \sum_{h=m+1}^{r} v_h g_h(\mathbf{x}, \mathbf{y})$, the reader may verify that the Kuhn–Tucker conditions take the symmetric form

$$\frac{\partial \Phi^0}{\partial \mathbf{x}} \geqq \mathbf{0}, \qquad \frac{\partial \Phi^0}{\partial \mathbf{y}} = \mathbf{0}, \qquad \frac{\partial \Phi^0}{\partial \mathbf{u}} \leqq \mathbf{0}, \qquad \frac{\partial \Phi^0}{\partial \mathbf{v}} = \mathbf{0},$$

$$\mathbf{x}^0 \frac{\partial \Phi^0}{\partial \mathbf{x}} = 0, \qquad \mathbf{u}^0 \frac{\partial \Phi}{\partial \mathbf{u}} = 0, \qquad \mathbf{x}^0 \geqq \mathbf{0}, \qquad \mathbf{u}^0 \geqq \mathbf{0},$$

where $\Phi^0 = \Phi(\mathbf{x}^0, \mathbf{y}^0, \mathbf{u}^0, \mathbf{v}^0)$, $(\mathbf{x}^0, \mathbf{y}^0)$ denotes the local minimum of the function $f_0(\mathbf{x}, \mathbf{y})$ under the constraints of problem (1.28b), and $(\mathbf{u}^0, \mathbf{v}^0)$ are the corresponding

[1] For the proof, see, e.g., [24, pp. 215–217] or [10, p. 539].

Lagrange multipliers. It is worth noting that the Lagrange multipliers **v** related to the equalities are not restricted to the nonnegativity (as in the classical Lagrange theory).

A summary of the rules for the formulation of the Kuhn–Tucker conditions for the general mathematical programming problem is as follows:

Rule 1. For a minimization (maximization) problem write all inequality constraints in the form

$$f_i(x) \leqq 0 \quad (f_i(x) \geqq 0).$$

Rule 2. Write the Lagrange function as the sum of the objective function and the weighted constraints.

Rule 3. The partial derivatives of the Lagrange function

(a) with respect to the nonnegative variables are *nonnegative* (*nonpositive*) for a *minimization* (*maximization*) problem and the complementary slackness condition

$$\mathbf{x}\frac{\partial \Phi}{\partial \mathbf{x}} = 0$$

is fulfilled;

(b) with respect to the free variables are equal to zero;

(c) with respect to the Lagrange multipliers corresponding to the inequality constraints are *nonpositive* (*nonnegative*) for a *minimization* (*maximization*) problem and the complementary slackness condition

$$\mathbf{u}\frac{\partial \Phi}{\partial \mathbf{u}} = 0$$

is fulfilled;

(d) with respect to the Lagrange multipliers corresponding to the equality constraints are equal to zero.

For a numerical illustration, we consider the following example:

$$\begin{aligned} \text{minimize} \qquad & f_0(\mathbf{x}) = x_1^2 - 4x_1 + x_2^2 - 6x_2 \\ \text{subject to} \qquad & x_1 + x_2 \leqq 3, \\ & -2x_1 + x_2 \leqq 2. \end{aligned}$$

The Lagrange function is

$$\Phi(\mathbf{x}, \mathbf{u}) = x_1^2 - 4x_1 + x_2^2 - 6x_2 + u_1(x_1 + x_2 - 3) + u_2(-2x_1 + x_2 - 2).$$

Application of the Kuhn–Tucker conditions (2.1′)–(2.4′) gives

$$\frac{\partial \Phi}{\partial x_1} = 2x_1 - 4 + u_1 - 2u_2 = 0, \tag{2.26}$$

$$\frac{\partial \Phi}{\partial x_2} = 2x_2 - 6 + u_1 + u_2 = 0, \tag{2.27}$$

$$\frac{\partial \Phi}{\partial u_1} = x_1 + x_2 - 3 \leqq 0, \tag{2.28}$$

$$u_1 \frac{\partial \Phi}{\partial u_1} = u_1(x_1 + x_2 - 3) = 0, \tag{2.29}$$

$$\frac{\partial \Phi}{\partial u_2} = -2x_1 + x_2 - 2 \leqq 0, \tag{2.30}$$

$$u_2 \frac{\partial \Phi}{\partial u_2} = u_2(-2x_1 + x_2 - 2) = 0, \tag{2.31}$$

$$u_1 \geqq 0, \qquad u_2 \geqq 0. \tag{2.32}$$

There is in general no simple computational procedure for the solution of these conditions. In order to show how to use the Kuhn–Tucker conditions, it is necessary to explore various cases defined principally by reference to whether each u_i is zero.

For the first case, suppose that $u_1 = 0$ and $u_2 = 0$. From conditions (2.26)–(2.27), we get $x_1 = 2$ and $x_2 = 3$. This vector cannot be a solution of our problem because it violates the first constraint $x_1 + x_2 \leqq 3$.

Second, suppose that $u_1 \neq 0$ and $u_2 = 0$. Then equations (2.26) and (2.27) are reduced to

$$2x_1 + u_1 = 4,$$
$$2x_2 + u_1 = 6.$$

Due to the complementary slackness condition (2.29), inequality (2.28) must be fulfilled as the equality

$$x_1 + x_2 = 3.$$

The above system of equations yields the solution $x_1 = 1, x_2 = 2$, and $u_1 = 2$, which also satisfies the remaining conditions (2.30)–(2.31). In other words, all Kuhn–Tucker conditions (2.26)–(2.32) are satisfied.

The third case corresponds to $u_1 = 0, u_2 \neq 0$. The resulting system of equations,

$$2x_1 - 2u_2 = 4,$$
$$2x_2 + u_2 = 6,$$
$$-2x_1 + x_2 = 2,$$

yields the solution $x_1 = \frac{4}{5}$, $x_2 = \frac{18}{5}$, $u_2 = -\frac{6}{5}$, which violates conditions (2.28) and (2.32).

The last possibility is $u_1 \neq 0$ and $u_2 \neq 0$. Because of the complementary slackness conditions (2.29) and (2.31), both inequality constraints (2.28) and (2.30) must be satisfied as equalities:

$$x_1 + x_2 = 3,$$
$$-2x_1 + x_2 = 2.$$

The solution is $x_1 = \frac{1}{3}$ and $x_2 = \frac{8}{3}$. Substituting these values in (2.26)–(2.27), we obtain a negative value for the Lagrange multiplier $u_2 = -\frac{8}{9}$, which is a contradiction to condition (2.32).

Only the values $x_1 = 1$, $x_2 = 2$, $u_1 = 2$, and $u_2 = 0$ satisfy all Kuhn–Tucker conditions and a simple inspection of the graph of the feasible solutions illustrates that this is indeed the optimal solution of our example.

Without further assumption about the functions $f_k(\mathbf{x})$ $(k = 0, 1, 2, \ldots, m)$, Theorem 2.1 provides only necessary conditions for a local optimal solution of problem (1.28).

In order to illustrate that the Kuhn–Tucker conditions are not sufficient conditions for a local minimum (maximum) of mathematical programming problems, we consider the following very simple one-variable example:

$$\text{maximize} \quad f_0(x) = (x-1)^3 \tag{2.33}$$

$$\text{subject to} \quad x \leqq 2, \tag{2.34}$$

$$x \geqq 0. \tag{2.35}$$

According to Rule 1 for the formulation of the Kuhn–Tucker condition, we rewrite the constraint as $2 - x \geqq 0$. Then the Lagrange function is

$$\Phi(x, u) = (x-1)^3 + u(2-x).$$

Application of the Kuhn–Tucker conditions (2.20)–(2.25) for the maximization problem gives

$$\frac{\partial \Phi}{\partial x} = 3(x-1)^2 - u \leqq 0, \tag{2.36}$$

$$x\frac{\partial \Phi}{\partial x} = x[3(x-1)^2 - u] = 0, \tag{2.37}$$

$$\frac{\partial \Phi}{\partial u} = 2 - x \geqq 0, \tag{2.38}$$

$$u\frac{\partial \Phi}{\partial u} = u(2-x) = 0, \tag{2.39}$$

$$u \geqq 0 \quad \text{(because of Rule 1)}. \tag{2.40}$$

The reader may verify that $x^0 = 1$ and $u^0 = 0$ satisfy the Kuhn–Tucker conditions (2.36)–(2.40). By simple inspection, it can be shown that the maximum of function (2.33) under the constraints (2.34)–(2.35) is at the point $x = 2$ and not at the point $x^0 = 1$.

The question of sufficiency of the Kuhn–Tucker conditions or the "second-order conditions" for the optimal solution of mathematical programming problems will be explored in the next chapter.

2.5 The Kuhn–Tucker Conditions and Economic Analysis

As illustrated in the previous section, the Kuhn–Tucker conditions can be helpful in the solution of specific numerical problems. Many algorithms of quadratic programming are based on these conditions. For economists, the Kuhn–Tucker conditions can be more useful for derivation of qualitative results without the necessity of specifying numerically the parameters of a mathematical programming problem. The primary aim is to characterize the optimal behavior of an economic agent under consideration. "As a result the Kuhn–Tucker conditions may perhaps constitute the most powerful single weapon provided to economic theory by mathematical programming" [4, p. 165].

A few examples will illustrate how the Kuhn–Tucker conditions can be used as an instrument for qualitative economic analysis.

2.5.1 Peak Load Pricing

Many profit-maximizing firms are confronted with the situation that the demand for a given product varies by the hour of the day so that at some times the capacity of the firm is fully utilized (peak periods), while at other times demand is slow so that some capacity remains underutilized (off-peak periods). As shown by Littlechild [20] with the aid of Kuhn–Tucker analysis and previously formulated by Steiner [35] and Williamson [38], in such situations the differential pricing is—in the sense of profit maximization—optimal. According to [4, p. 167], the following proposition can be formulated.

Proposition 2.1. *The profit-maximizing outputs will be such that prices at off-peak periods will merely cover marginal operating costs* (*raw materials, labor, etc.*), *while in peak periods the prices will exceed marginal operating costs. The sum of the excesses of these prices over marginal operating costs for all peak periods will just add up to marginal capital cost, i.e., they will sum to the marginal cost of increasing capacity.*

Proof. We denote the quantity demanded during each of the 24 hours of the day by $x_1, x_2, \ldots, x_{24}$ and the corresponding prices (e.g., telephone rates) by $p_1, p_2, \ldots, p_{24}$. It is assumed that all $x_i > 0$, i.e., that some output is sold during each hour of the day. The hourly output capacity is denoted by y. The function $C(x_1, x_2, \ldots, x_{24})$ describes the daily total operating cost and $g(y)$ the daily cost of capital (capacity). We assume that the marginal operating cost, $\frac{\partial C}{\partial x_i}$, as well as the marginal capacity cost, $\frac{dg}{dy}$, are positive. Furthermore, it is assumed that prices are not affected by the firm's output, i.e., $\frac{\partial p_i}{\partial x_i} = 0$. (Perfect competition prevails. The prices $p_1, \ldots, p_{24}$ can therefore be regarded as given and fixed.)

The firm seeks to maximize the total profit per day,

$$\pi = \sum_{i=1}^{24} p_i x_i - C(x_1, x_2, \ldots, x_{24}) - g(y),$$

subject to the 24 hourly capacity constraints,

$$\begin{aligned} x_i &\leqq y \quad (i = 1, 2, \ldots, 24), \\ x_i &\geqq 0 \quad (i = 1, 2, \ldots, 24), \end{aligned}$$

$$y \geqq 0.$$

The Lagrange function has the form

$$\Phi(\mathbf{x}, y, \mathbf{u}) = \sum_{i=1}^{24} p_i x_i - C(x_1, x_2, \ldots, x_{24}) - g(y) + \sum_{i=1}^{24} u_i (y - x_i).$$

Under the assumption of a perfectly competitive firm, the Kuhn–Tucker conditions are then

$$\frac{\partial \Phi}{\partial x_i} = p_i - \frac{\partial C}{\partial x_i} - u_i \leqq 0 \quad (i = 1, 2, \ldots, 24), \tag{2.41}$$

$$x_i \frac{\partial \Phi}{\partial x_i} = x_i \left(p_i - \frac{\partial C}{\partial x_i} - u_i \right) = 0 \quad (i = 1, 2, \ldots, 24), \tag{2.42}$$

$$\frac{\partial \Phi}{\partial y} = -\frac{dg}{dy} + \sum_{i=1}^{24} u_i \leqq 0, \tag{2.43}$$

$$y \frac{\partial \Phi}{\partial y} = y \left(-\frac{dg}{dy} + \sum_{i=1}^{24} u_i \right) = 0, \tag{2.44}$$

$$\frac{\partial \Phi}{\partial u_i} = y - x_i \geqq 0 \quad (i = 1, 2, \ldots, 24), \tag{2.45}$$

$$u_i \frac{\partial \Phi}{\partial u_i} = u_i (y - x_i) = 0 \quad (i = 1, 2, \ldots, 24), \tag{2.46}$$

$$u_i \geqq 0 \quad (i = 1, 2, \ldots, 24). \tag{2.47}$$

Since we have assumed that $x_i > 0$ $(i = 1, 2, \ldots, 24)$, it follows from (2.45) that $y > 0$ (that is, if capacity, y, were zero, nothing could be produced).

Because we are only interested in solutions in which all x_i and y are positive, (2.41) and (2.43) become the following:

$$p_i - \frac{\partial C}{\partial x_i} - u_i = 0 \quad (i = 1, 2, \ldots, 24), \tag{2.41$'$}$$

$$-\frac{\partial g}{\partial y} + \sum_{i=1}^{24} u_i = 0. \tag{2.43$'$}$$

In any off-peak period t, there is by definition excess capacity $(y > x_t)$. Therefore, by the complementary slackness condition (2.46), we must have $u_t = 0$ for off-peak periods.

Then the first part of Proposition 2.1 follows immediately from (2.41$'$):

$$p_t = \frac{\partial C}{\partial x_t} \quad \text{for any off-peak period } t;$$

that is, for any off-peak period, it is optimal to set the price equal to the marginal operating cost, $\frac{\partial C}{\partial x_t}$. Since there is excess capacity, demand should be encouraged by charging a price as low as possible without incurring a loss on the marginal unit sold.

For any peak period, s, the capacity of the firm is fully utilized ($x_s = y$). Since we have assumed that $\frac{dg}{dy} > 0$ (increasing output capacity requires additional capital), it follows from (2.43′) that

$$\frac{dg}{dy} = \sum_s u_s > 0,$$

that is, at least for some of the peak periods, the Lagrange multipliers must be positive. Then we obtain from (2.41′) that

$$p_s = \frac{\partial C}{\partial x_s} + u_s \quad \text{for any peak period } s.$$

The price will exceed the marginal operating cost by a supplementary amount equal to the value of the Lagrange multiplier u_s. Moreover, by (2.43′) the sum of these supplements for all peak periods together will be exactly equal to the marginal capacity cost, $\frac{\partial g}{\partial y}$. Since peak period demand presses on capacity, any increase in this demand must require additional capital, and it must therefore cover its marginal capital cost, $\frac{dg}{dy}$. □

This completes the proof of Proposition 2.1 as the basic principles for the setting of daytime and evening telephone rates, for the higher accommodation prices in the peak season, etc. This principle can be applied in the recent discussion about the road pricing system as well.

2.5.2 Revenue Maximization under a Profit Constraint[2]

Suppose that a firm produces a single product whose output is q and that its sales are affected by its advertising expenditure a. The firm will maximize its total revenue $R(q, a)$ subject to a profit constraint,

$$\Pi = R(q, a) - C(q) - a \geqq m,$$

where $C(q)$ indicates the cost of production and where the marginal revenue of advertising and the marginal cost of output are both positive ($\frac{\partial R}{\partial a} > 0$, $\frac{dC}{dq} > 0$). Then the behavior of the firm is described by the following.

Proposition 2.2. *The revenue-maximizing output will be such that the profit is equal to the prescribed level m, the marginal revenue* $\frac{\partial R}{\partial q}$ *is positive, and the marginal profit* $\frac{\partial \Pi}{\partial q}$ *is negative.*

Proof. The firm's decision problem is

$$\begin{array}{ll} \text{maximize} & R(q, a) \\ \text{subject to} & R(q, a) - C(q) - a \geqq m, \\ & \qquad q \geqq 0, \qquad a \geqq 0. \end{array}$$

[2] Formulation of the problem by Baumol [4, p. 170].

The Lagrange function becomes

$$\Phi(q,a,u) = R(q,a) + u(R(q,a) - C(q) - a - m),$$

and the Kuhn–Tucker conditions are

$$\frac{\partial \Phi}{\partial q} = \frac{\partial R}{\partial q} + u\left(\frac{\partial R}{\partial q} - \frac{dC}{dq}\right) \leqq 0,$$

or

$$(1+u)\frac{\partial R}{\partial q} - u\frac{dC}{dq} \leqq 0, \tag{2.48}$$

$$q\frac{\partial \Phi}{\partial q} = q\left[(1+u)\frac{\partial R}{\partial q} - u\frac{dC}{dq}\right] = 0, \tag{2.49}$$

$$\frac{\partial \Phi}{\partial a} = \frac{\partial R}{\partial a} + u\frac{\partial R}{\partial a} - u \leqq 0,$$

or

$$(1+u)\frac{\partial R}{\partial a} \leqq u,$$

or

$$\frac{\partial R}{\partial a} \leqq \frac{u}{1+u}, \tag{2.50}$$

$$a\frac{\partial \Phi}{\partial a} = a\left[(1+u)\frac{\partial R}{\partial a} - u\right] = 0, \tag{2.51}$$

$$\frac{\partial \Phi}{\partial u} = R(q,a) - C(q) - a - m \geqq 0, \tag{2.52}$$

$$u\frac{\partial \Phi}{\partial u} = u\left[R(q,a) - C(q) - a - m\right] = 0, \tag{2.53}$$

$$u \geqq 0. \tag{2.54}$$

Assuming $q > 0$ in the solution, condition (2.48) can be written as

$$\frac{\frac{\partial R}{\partial q}}{\frac{dC}{dq}} = \frac{u}{1+u}. \tag{2.48$'$}$$

Since we have assumed $\frac{\partial R}{\partial a} > 0$, it follows from (2.50) and (2.54) that $u > 0$. The complementary slackness condition (2.53) then implies $\Pi = m$. Taking into account our assumption that $\frac{dC}{dq} > 0$, it follows from (2.48′) that the marginal revenue $\frac{\partial R}{\partial q}$ is positive and smaller than the marginal cost $\frac{dC}{dq}$. Therefore, the marginal profit with respect to output, $\frac{\partial \Pi}{\partial q} = \frac{\partial R}{\partial q} - \frac{dC}{dq}$, must be negative. □

From the economic interpretation point of view, it is interesting to compare the obtained results with the results for a profit-maximizing firm. The reader may verify that the necessary condition for profit-maximizing output is that the marginal revenue is equal to the marginal cost. In our model, at the constrained revenue-maximizing output, the marginal revenue is lower than the marginal cost. The implication of this result for the linear revenue function

$$R(q, a) = \alpha_1 q + \alpha_2 a \quad \text{with } \alpha_1 > 0, \quad \alpha_2 > 0$$

and the quadratic cost function

$$C(q) = cq^2 \quad \text{with } c > 0^3$$

is that the constrained revenue-maximizing output from our model is higher than the profit-maximizing output. The optimal solution q_π for the profit-maximizing firm follows directly from the condition

$$\frac{\partial R}{\partial q} = \alpha_1 = \frac{dC}{dq} = 2cq,$$

i.e.,

$$q_\pi = \frac{\alpha_1}{2c}.$$

For the revenue-maximizing firm, the necessary condition for the optimal solution becomes

$$\frac{\partial R}{\partial q} = \alpha_1 = \frac{u}{1+u}\frac{dC}{dq} = \frac{u}{1+u}2cq,$$

and consequently

$$q_R = \frac{\alpha_1}{2c}\left(\frac{1+u}{u}\right) > \frac{\alpha_1}{2c} = q_\pi .$$

2.5.3 Behavior of the Firm under Regulatory Constraint

The regulation of monopolies is an important subject in applied economic analysis. In the sectors with network structure, such as telecommunications, electricity and gas, and railway systems with high fixed and irreversible (sunk) costs, it is cheaper to produce goods by a single firm than by many firms. These situations are called *natural monopolies* and occur whenever the average costs of production for a single firm are declining over a broad range of output levels. The reason lies in the so-called "bundling advantage": With increasing diameter of the pipe, the volume increases

[3] It can be shown that in this case the second-order conditions are fulfilled.

more rapidly than its girth, which is crucial for the costs. The average costs of production fall as the scale of production increases; we say there are economies of scale. A natural monopoly with ireversible costs implies a barrier to market entry and is characterized by sustainable market power. In order to prevent this monopoly power over the customers and to guarantee the reliability and quality of supply at economically or politically desired prices, the regulation of monopolistic firms has been introduced. For the regulation of interstate telephone and telegraph service and of radio and television broadcasting in the United States, the Federal Communications Commission was created in 1934, and the Civil Aeronautics Board, which regulated the prices charged by the interstate scheduled airlines as well as entry into the industry, was established in 1938. The Federal Energy Regulatory Commission was established in the United States in 1977. Independent regulatory agencies operate now in all countries of the European Union.

The monopoly profit-maximizing level of output is that one for which marginal revenue equals marginal cost. At this output level, price will exceed marginal cost. The profitability of the monopolist will depend on the relationship between price and average cost.

One approach to devising monopoly pricing schemes that is followed in many regulatory situations is to permit the monopoly to charge a price above average cost that is sufficient to earn a "fair" rate of return on investment. From an economic point of view, the interesting question concerns the impact of regulation on the firm's input choices. In the most frequently quoted paper in regulatory economics Averch and Johnson [2] showed that under the rate of return constraint the profit-maximizing firm chooses an inefficient input mix in the sense "that (social) cost" is not minimized at the output it selects [2, p. 1052].

Let us start the analysis considering a basic model of the monopoly firm producing a single output using two inputs, capital and labor, where the respective quantities are denoted q, x_1 and x_2; the production function permits inputs to be employed in any proportion. The unit price of the firm's output is denoted p. Suppose that it can buy as much as it wants of the two inputs at constant unit prices of c_1 and c_2, respectively, so that its profit function is

$$\Pi = pq - c_1 x_1 - c_2 x_2. \tag{2.55}$$

Assuming that $x_1 > 0$ and $x_2 > 0$ (in other words, both production factors are essential), the profit maximization requires that

$$\frac{\partial \Pi}{\partial x_1} = \frac{\partial pq}{\partial x_1} - c_1 = 0, \tag{2.56}$$

$$\frac{\partial \Pi}{\partial x_2} = \frac{\partial pq}{\partial x_2} - c_2 = 0, \tag{2.57}$$

and consequently,

$$\frac{\dfrac{\partial pq}{\partial x_1}}{\dfrac{\partial pq}{\partial x_2}} = \frac{c_1}{c_2}. \tag{2.58}$$

The ratio of marginal revenue products will equal the ratio of the input prices. The marginal revenue product $\frac{\partial pq}{\partial x_i}$ describes the extra revenue that accrues to a firm when it sells the output that is produced by one more unit of input i $(i = 1, 2)$. The marginal revenue product of factor i (MR_i) is given by the multiplication of marginal revenue (MR) by the marginal physical product (MP_i) of factor i: $\mathrm{MR}_i = \mathrm{MR} \cdot \mathrm{MP}_i$. The marginal revenue is the additional revenue obtained by a firm when it is able to sell one more unit of output. The marginal physical product describes the additional output that can be produced by one more unit of a particular input while holding all other inputs constant. According to (2.58), the firm uses an efficient mix of capital and labor in the sense that cost is minimized at the output it selects. Rewriting (2.58) as

$$\frac{\dfrac{\partial pq}{\partial x_1}}{c_1} = \frac{\dfrac{\partial pq}{\partial x_2}}{c_2}, \tag{2.58$'$}$$

every additional Euro given to any input yields the same revenue.

Now, following [2], suppose that the firm is regulated by government, which imposes a constraint on its rate of return. The introduction of such regulatory constraint is motivated by the following argument: "In judging the level of prices charged by firms for services subject to public control, government regulatory agencies commonly employ a 'fair rate of return' criterion: After the firm subtracts its operating expenses from gross revenues, the remaining net revenue should be just sufficient to compensate the firm for its investment in plant and equipment. If the rate of return, computed as the ratio of net revenue to the value of plant and equipment (the rate base), is judged to be excessive, pressure is brought to bear on the firm to reduce prices. If the rate is considered to be too low the firm is permitted to increase prices" [2, p. 1052]. The profit-maximizing behavior of the firm under such a regulatory constraint can then be described by the following.

Proposition 2.3. *The firm does not equate the marginal rate of factor substitution to the ratio of the input prices. The firm has an incentive to increase its investment: The amount of capital used with the regulatory constraint is not less than the amount used without a constraint.*

Proof. We define the firm's production function as

$$q = f(x_1, x_2), \quad \text{where } f_1 = \frac{\partial f}{\partial x_1} > 0, \quad f_2 = \frac{\partial f}{\partial x_2} > 0,$$
$$f(0, x_2) = f(x_1, 0) = 0.$$

That is, marginal products are positive, and production requires both inputs.

The inverse demand function can be written

$$p = p(q), \quad \text{where } p'(q) = \frac{dp}{dq} < 0.$$

The profit Π is defined by (2.55).

Let x_1 denote the physical quantity of plant and equipment in the rate base, b_1 the acquisition cost per unit of plant and equipment in the rate base, β_1 the value of depreciation of plant and equipment during a time period in question, and B_1 the cumulative value of depreciation.

The regulatory constraint of [2] is

$$\frac{pq - c_2x_2 - \beta_1}{b_1x_1 - B_1} \leqq s, \tag{2.59}$$

where the profit net of labor cost and capital depreciation constitutes a percentage of the rate base (net depreciation) no greater than a specified maximum s.

For simplicity, in [2] it was assumed that depreciation (β_1 and B_1) is zero and the acquisition cost b_1 is equal to 1 (i.e., the value of the rate base is equal to the physical quantity of capital). The price, or the "cost of capital," c_1 is the interest cost involved in holding plant and equipment (to be distinguished from the acquisition cost b_1). The regulatory constraint (2.59) can then be rewritten as

$$\frac{pq - c_2x_2}{x_1} \leqq s,$$

or

$$pq - sx_1 - c_2x_2 \leqq 0. \tag{2.60}$$

The "fair rate of return" s is the rate of return allowed by the regulatory agency on plant and equipment in order to compensate the firm for the cost of capital.

If $s < c_1$, the allowable rate of return is less than the actual cost of capital and the firm would withdraw from the market. Therefore, we shall assume that $s \geqq c_1$; the allowable rate of return must at least cover the actual cost of capital.

The problem of the firm is to maximize the profit described by function (2.55) subject to (2.60) and $x_1 \geqq 0$, $x_2 \geqq 0$. The Lagrange function is defined as

$$\Phi(x_1, x_2, u) = p(q)q - c_1x_1 - c_2x_2 - u(p(q)q - sx_1 - c_2x_2),$$

where $q = f(x_1, x_2)$.

The Kuhn–Tucker necessary conditions for a maximum at x_1^0, x_2^0, u^0 are

$$\frac{\partial \Phi}{\partial x_1} = (1-u)[p + p'q]f_1 - c_1 + us \leqq 0, \tag{a}$$

$$x_1\frac{\partial \Phi}{\partial x_1} = x_1\{(1-u)[p + p'q]f_1 - c_1 + us\} = 0, \tag{b}$$

$$\frac{\partial \Phi}{\partial x_2} = (1-u)[p + p'q]f_2 - (1-u)c_2 \leqq 0, \tag{c}$$

$$x_2 \frac{\partial \Phi}{\partial x_2} = x_2\{(1-u)[p+p'q]f_2 - (1-u)c_2\} = 0, \tag{d}$$

$$\frac{\partial \Phi}{\partial u} = -(p(q)q - sx_1 - c_2x_2) \geqq 0, \tag{e}$$

$$u\frac{\partial \Phi}{\partial u} = u(p(q)q - sx_1 - c_2x_2) = 0, \tag{f}$$

$$u \geqq 0. \tag{g}$$

Because the production requires both inputs $x_1^0 > 0$, $x_2^0 > 0$, and assuming $u^0 > 0$ (i.e., the regulatory constraint (2.60) is binding at (x_1^0, x_2^0)), conditions (a), (c), and (e) can be rewritten as the following equalities:

$$(1-u)[p+p'q]f_1 + us = c_1, \tag{2.61}$$

$$[p+p'q]f_2 = c_2, \tag{2.62}$$

$$pq - sx_1 - c_2x_2 = 0. \tag{2.63}$$

The expression $(p + p'q)$ describes the marginal revenue and f_1, f_2 denote the marginal physical products of capital and labor, respectively. Note that (2.61)–(2.63) will determine the values of x_1^0, x_2^0, and u^0.

If there is no regulatory constraint (2.60) so that the constraint is not active ($u = 0$), (2.61) and (2.62) reduce to (2.56)–(2.57) with the familiar rule that the marginal revenue product of each factor is equal to its price.

It follows from (2.61) that $u^0 > 0$ (the binding regulatory constraint (2.60)) will distort the equality of the marginal revenue product of capital $(p + p'q)f_1$ with its actual cost c_1. Consequently, the relative proportions of capital and labor used by the firm will be changed. The marginal rate of factor substitution $\frac{f_1}{f_2}$ is no longer equal to the ratio of the input prices. The first part of Proposition 2.3 is proved.

Assuming that $u > 0$, it is clear from (2.61) that $u = 1$ implies $c_1 = s$. On the other hand, if $c_1 = s$, (2.61) reduces to (2.56), which corresponds to the behavior of unregulated monopoly. Therefore, we sharpen our assumption $s \geqq c_1$ to $s > c_1$ and get $u^0 \neq 1$.

Let the superscript 0 denote the solution of the optimization problem for the regulated monopoly and asterisk the solution for the unregulated monopoly. Furthermore, denote the expression $[p + p'q]f_1$ for the marginal revenue product of capital and the expression $[p + p'q]f_2$ for the marginal revenue of labor by MR_1 and MR_2, respectively.

Adding c_1u^0 to both sides of (2.61) and rearranging terms yields

$$\mathrm{MR}_1^0 = c_1 - \frac{(s - c_1)}{1 - u^0}u^0. \tag{2.64}$$

Under the assumption that $s > c_1$ and $u^0 < 1$ (as claimed in [2]), it follows from (2.64) that $\mathrm{MR}^0 < c_1$.

If the revenue function $G \equiv pf(x_1, x_2)$ is concave (this assumption is not mentioned in [2]; it was introduced by Takayama [36]), then the marginal revenue product of capital MR_1 is a nonincreasing function of capital used, and consequently the amount of capital used under the regulatory constraint (x_1^0) is not less than the amount used without a constraint (x_1^*). If G is assumed to be strictly concave, then $\frac{\partial \, \mathrm{MR}_1}{\partial x_1} < 0$; hence $x_1^0 > x_1^*$. Furthermore, it follows from (2.61) and (2.62) that

$$\frac{\mathrm{MR}_1}{\mathrm{MR}_2} = \frac{c_1}{c_2} - \frac{(s - c_1)}{c_2} \frac{u^0}{(1 - u^0)} < \frac{c_1}{c_2}.$$

The marginal rate of substitution between inputs ($\frac{\mathrm{MR}_1}{\mathrm{MR}_2}$) is lower than the ratio of input prices. Each output is produced with more capital and less labor as compared to the unregulated optimum. This effect of overcapitalization contained in the second part of Proposition 2.3 is known as the *Averch–Johnson effect*. In their own words, "If the rate of return allowed by the regulatory agency is greater than the cost of capital but is less than the rate of return that would be enjoyed by the firm were it free to maximize profit without regulatory constraint, then the firm will substitute capital for the other factor of production and operate at an output where cost is not minimized" [2, p. 1053]. □

This inefficiency derives from the fact that the net return of the monopolist on every unit of capital is $s - c_1$, and this creates an incentive to substitute capital for labor. Under regulatory constraint $\Pi^0 + c_1 x_1^0 - s x_1^0 = 0$, and consequently $\Pi^0 = (s - c_1) x_1^0$.

An important question in the Averch–Johnson analysis is whether u^0 is indeed less than one. The argument by Averch–Johnson roughly goes as follows: Since $s > c_1$, u^0 cannot be equal to one, as shown before, for the unconstrained rate of return is $u^0 = 0$. Because of the continuity of u^0 with respect to s, u^0 should always be less than one.

But the continuity of u^0 is not intuitively obvious. The value of the Lagrange multiplier may jump from zero to some nonzero value as the constraint moves from a nonactive to an active stage. Takayama [36] showed that the continuity of u^0 in the Averch–Johnson model depends on the continuity of x_1^0 and x_2^0 with respect to s.

Another way to obtain the condition $u^0 < 1$ uses the optimality conditions (2.61)–(2.63). As already mentioned, under the assumptions $x_1^0 > 0$, $x_2^0 > 0$, and $u^0 > 0$, these equations determine the values of x_1^0, x_2^0, and u^0. The value of u^0 can be obtained explicitly from (2.64), assuming that $s - \mathrm{MR}_1^0 > 0$:

$$u^0 = \frac{c_1 - \mathrm{MR}_1^0}{s - \mathrm{MR}_1^0} > 0. \tag{2.65}$$

Under our assumption that $s > c_1$ and the new condition $s > \mathrm{MR}_1^0$, it follows directly from (2.65) that $u^0 < 1$.

If $u^0 = 1$, then due to (2.61), $s = c_1$, which contradicts the assumption of $s > c_1$.

A condition similar to our condition $s > \mathrm{MR}_1^0$ is used by El-Hodiri and Takayama [9]. Assuming that G is concave, the Averch–Johnson effect, $x_1^0 \geqq x_1^*$, occurs if and only if $\mathrm{MR}_1^0 - c_1 \leqq 0$. They assert that $u^0 < 1$ if and only if $\mathrm{MR}_1^0 \leqq c_1$ (i.e., if and only if $x_1^0 \geqq x_1^*$). They can prove the Averch–Johnson effect without assuming anything about u^0 but with the requirement that $\mathrm{MR}_1^0 \leqq c_1$.

Averch and Johnson applied the model to one particular regulated industry—the domestic telephone and telegraph industry. They found that "the model does raise issues relevant to evaluating market behavior" [2, p. 1052]. The scientific discussion as well as the real applications of the rate-of-return regulation has been initiated.

In the 1980s, the discussion—connected with privatization and deregulation policies in several countries—began to concentrate on the question of how a regulating agency could give the best incentives for efficient production in the regulated firm. The way to increase the efficiency is based on the promotion of the competition. Internal subsidization was increasingly considered undesirable and this view has led to reconsideration of the internal organization of firms which claimed to be natural monopolies. Because in the utilities like telecommunications, electricity, and gas, it is only the distributive grid which has the properties of a natural monopoly, vertical desintegration or unbundling has been proposed. The electricity generation must be separated from the transmission and distribution activities. With respect to these grids, economies of scale are still predominant, maybe even increasing in recent decades. With unbundling a market entry in those parts where no natural monopoly properties prevail can ensure and a presupposition for effective competition is created. Competition is effective when each firm cannot appreciably raise the price above that of its rivals for fear of losing its market share, and can only increase profit by cutting costs. Regulation of networks which remain natural monopolies is needed in order to make the entry possible for different providers and in this way to promote a competition. How to design regulation to fulfill the above requirements and so provide incentives for grid companies to reduce the costs and consequently the prices? The main drawback of the rate of return regulation is the lack of incentives for cost reduction and technological innovation. "A profit maximizing firm subject to a fair return on investment regulation will overcapitalize and select those technical changes which will allow to continue to do so—namely labor augmenting innovations" [34, p. 630].[4] We speak about costs based regulation, in which firms' allowed rate of return is based directly on the reported costs of the individual firm.

During the privatization of British Telecom, Littlechild [21], Director of the Office for Electricity Regulation, proposed a new type of regulation, the so-called *price-cap* regulation.[5] The basic idea is that the price index of the monopolistically supplied goods (or services) must not exceed the retail price index minus an exogenously fixed productivity factor. The customer must be able to buy at the prices of given period the same basket of goods (or services) as in the base period without increasing expenditures. The retail price index (RPI) measuring the inflation rate is the consumer price index, which is a Laspeyers index of the usual type:

$$\mathrm{RPI} = \frac{\sum_i p_i q_i^0}{\sum_i p_i^0 q_i^0}.$$

The superscript 0 defines variables of the base period in which the fixed commodity

---

[4] For further reading, see [13].

[5] For further reading on price-cap regulation, see [7, 19, 17], and for a survey comparing rate of return and price-cap regulation, see [22].

basket of the index was empirically determined. We denote the price of commodity i by p_i and the quantity by q_i $(i = 1, 2, \ldots, n)$.

The profit-maximizing firm is regulated by the following constraint:

$$\sum_{i=1}^{n} p_i q_i^0 \leqq \sum_{i=1}^{n} p_i^0 q_i^0 (1 + \text{RPI} - X), \tag{2.66}$$

where X describes the productivity factor of the sector.

Because both the consumer price index (RPI) as well as the expenditures of the base period $(\sum_{i=1}^{n} p_i^0 q_i^0)$ are for the regulator exogenously given, the only control variable for him remains the productivity factor (X). From the regulatory point of view the relevant question therefore concerns the impact of X on the behavior of the regulated firm.

For this purpose, we consider the following optimization problem:

$$\begin{aligned} &\text{maximize} \quad && \Pi(q) = p(q)q - C(q) \\ &\text{subject to} \quad && p(q)q^0 \leqq b^0, \end{aligned} \tag{2.67}$$

where $\Pi(q)$ describes the profit function, $C(q)$ is the cost function, and $p(q)$ is the inverse demand function. b^0 denotes the right side of the regulatory constraint (2.66):

$$b^0 = p^0 q^0 (1 + \text{RPI} - X).$$

The price cap (b^0) can be faced as a function of the expenditures in the base period $(p^0 q^0)$, of the retail price index (RPI), and of the productivity factor (X):

$$b^0 = b\left(\underset{(+)}{R^0}, \underset{(+)}{\text{RPI}}, \underset{(-)}{X}\right),$$

where $R^0 = p^0 q^0$. With increasing expenditures (R^0) and increasing consumer price index (RPI), the price cap (b^0) rises. The increasing productivity (X) makes the constraint (2.66) tighter. In other words, the increasing productivity implies lower prices for the consumers.

Furthermore, we postulate positive marginal cost and a declining demand function:

$$\frac{dC}{dq} = \text{MC} > 0, \qquad \frac{dp}{dq} < 0.$$

The Lagrange function for the maximization problem (2.67) is

$$\Phi(q, u) = p(q)q - C(q) + u(b^0 - p(q)q^0).$$

Application of the Kuhn–Tucker conditions (2.20)–(2.21) yields

$$p - \frac{dC}{dq} = -\frac{dp}{dq}(q - uq^0). \tag{2.68}$$

Multiplying both sides of (2.68) by $\frac{1}{p}$, we obtain

$$\frac{p - \text{MC}}{p} = -\frac{dp}{dq}\frac{q}{p} + u\frac{q^0}{p}\frac{dp}{dq}.$$

Using the notion of the price elasticity of demand $\varepsilon = \frac{pdq}{qdp} < 0$, the following optimality condition results:

$$\frac{p - \text{MC}}{p} = -\left(1 - u\frac{q^0}{q}\right)\frac{1}{\varepsilon}. \tag{2.69}$$

For the unregulated monopolistic firm with the Lagrange multiplier $u = 0$, the form (2.69) reduces to the well-known Lerner index (see, e.g., [5, p. 26]):

$$\frac{p - \text{MC}}{p} = -\frac{1}{\varepsilon}.$$

The Lerner index measures the market power of monopoly.

For a perfectly competitive firm, price equals marginal cost, so the Lerner index equals 0. The higher the Lerner index is, the higher is the degree of monopoly power. For the profit-maximizing firm, the Lerner index is equal to the reciprocal value of the price elasticity of demand for the firm's product. The lower the price elasticity of demand for the firm's product is, the higher is the degree of monopoly power.

Under the price-cap regulation the Lerner index is modified by the expression $1 - u \cdot \frac{q^0}{q}$.

From the Kuhn–Tucker condition (2.25) follows the nonnegativity of the Lagrange multiplier u. According to (2.9), this multiplier describes the change of the monopoly profit due to a change of the price cap:

$$u = \frac{\partial \Pi}{\partial b^0} \geqq 0.$$

Moreover, it can be shown [33, pp. 166–168] that under the assumption of concavity of Π and convexity of $p(q)q^0$,

$$\frac{\partial u}{\partial b^0} \leqq 0.$$

The lower price cap—due to the stronger regulation by setting the productivity factor (X) higher—implies higher Lagrange multiplier, and according to (2.69) the lower the degree of monopoly power. In this way, price-cap regulation is an appropriate instrument to reduce the market power of natural monopolies. The form (2.69) reveals the central problem of price-cap regulation for the regulatory agency, the determination of the productivity factor (X). "Too high a price ceiling makes the firm an unregulated monopolist, too low cap conflicts with viability, and in between the 'right' price level is difficult to compute" [19, p. 17]. One possibility of how to calculate the X-factor in order to provide incentives for cost reduction and technological innovation and consequently for reduction of network tariffs in the electricity sector is described by [25].

An application of regulatory constraints for environmental economics and policy will be discussed in the next section.

2.5.4 Environmental Regulation: The Effects of Different Restrictions

The forms of standards used in current environmental regulation vary tremendously. The most frequently discussed forms are systems of permits which determine a fixed amount of emission allowed for each emission source independent of the production level of this source. The deficiencies of such a source-based system of permits are investigated and summarized in [31, Chapter VIII] or [32]. Another form of environmental regulation relates to restrictions on pollution per unit of output or input [8]. In an economic sense, restrictions that are based on a unit of output or input are equivalent to a productivity or intensity regulation, well known from the literature, beginning with the Averch–Johnson model [2], and are discussed in the previous section.

In the paper by Helfand [14], the effects of five different forms of pollution standards on input decisions, the level of production, and firm profits are examined using a graphical approach. In this section, we analyze the effects of different kinds of pollution control standards in a more general way using Kuhn–Tucker conditions.

The model used in [14] involves one firm, facing a horizontal output demand curve and using two inputs, x_1 and x_2, with horizontal supply curves. The assumption that there are only two inputs is only for simplicity but without loss of generality. The assumption of a horizontal output demand curve is more limiting and is realistic only for a good whose price is unaffected by production of the firm. In [14], this assumption makes the problem tractable and permits a graphical presentation.

Assume that the firm produces a single output in the quantity q according to the production function $f(x_1, x_2)$ with the usual properties:

$$\begin{aligned} f_1 &= \frac{\partial f}{\partial x_1} > 0, \qquad & f_2 &= \frac{\partial f}{\partial x_2} > 0, \\ f_{11} &= \frac{\partial^2 f}{\partial x_1^2} < 0, \qquad & f_{22} &= \frac{\partial^2 f}{\partial x_2^2} < 0. \end{aligned} \tag{2.70}$$

In other words, the marginal products of both inputs are positive but declining.

The firm also causes pollution, the level of which depends on the level of production and the technology. In order to reduce the level of pollution, the firm can use an abatement activity or invest in new technology. The resulting level of pollution (or net pollution) can be described as follows:

$$P = G(f(x_1, x_2)) - \mathrm{Ab}(x_3),$$

where $\mathrm{Ab}(x_3)$ denotes the abatement activity as a function of abatement expenditure x_3 (or expenditure for development of a new technology). It is assumed that $\frac{d\mathrm{Ab}}{dx_3} > 0$, that is, more abatement equipment (or higher expenditure for development of a new technology) reduces the level of pollution. More generally, we describe the level of net pollution as follows:

$$P = P(x_1, x_2, x_3)$$

with $P_1 = \frac{\partial P}{\partial x_1} > 0$, $P_2 = \frac{\partial P}{\partial x_2} > 0$, and $P_3 = \frac{\partial P}{\partial x_3} < 0$.[6]

[6] In this formulation of the net pollution function P, we differ slightly from the model in [14].

The firm is assumed to maximize profits while facing an output price p and input prices c_1 and c_2 as well as the price of abatement equipment c_3 as given.

The necessary conditions for a profit-maximizing firm without regulatory constraint (i.e., pollution restrictions) are given by (2.56)–(2.57) in the previous section. In economic terms, the value of the marginal product of input i $(i = 1, 2)$ must be equal to its price, i.e., the ratio of marginal revenue products will equal the ratio of the input prices.

Now, similar to the regulatory constraint by [2], pollution restrictions in the form of different kinds of pollution-control standards will be taken into account. What are the effects for the level of production and the firm's profit?

2.5.4.1 Standard as a Set Level of Emissions

Let Z_p be the amount of total pollution permissible in a certain period of time. It can be represented as a constraint in the form $P(\mathbf{x}) \leqq Z_p$. The optimization problem of the profit-maximizing firm is

$$\begin{aligned} \underset{x_1,x_2,x_3}{\text{maximize}} \quad & \Pi = pf(x_1, x_2) - c_1x_1 - c_2x_2 - c_3x_3 \\ \text{subject to} \quad & P(x_1, x_2, x_3) \leqq Z_p, \\ & x_1 \geqq 0, \quad x_2 \geqq 0, \quad x_3 \geqq 0. \end{aligned}$$

We write the Lagrange function

$$\Phi(\mathbf{x}, u) = pf(x_1, x_2) - c_1x_1 - c_2x_2 - c_3x_3 + u(Z_p - P(x_1, x_2, x_3))$$

and the resulting Kuhn–Tucker conditions

$$\frac{\partial \Phi}{\partial x_1} = pf_1 - c_1 - uP_1 \leqq 0, \tag{2.71}$$

or

$$pf_1 \leqq c_1 + uP_1,$$

$$x_1 \frac{\partial \Phi}{\partial x_1} = x_1(pf_1 - c_1 - uP_1) = 0. \tag{2.72}$$

Assuming that $x_1 > 0$, it follows from (2.72) that

$$pf_1 = c_1 + uP_1. \tag{2.73}$$

Furthermore,

$$\frac{\partial \Phi}{\partial x_2} = pf_2 - c_2 - uP_2 \leqq 0, \tag{2.74}$$

or

$$pf_2 \leqq c_2 + uP_2,$$

$$x_2 \frac{\partial \Phi}{\partial x_2} = x_2(pf_2 - c_2 - uP_2) = 0. \tag{2.75}$$

Assuming that $x_2 > 0$, (2.75) implies that

$$pf_2 = c_2 + uP_2. \tag{2.76}$$

We conclude that the value of the marginal product of the input i $(i = 1, 2)$ is equal to the marginal input costs, plus the pollution cost, uP_i, where $u = u^0 = \frac{\partial \Pi(\mathbf{x}^0(Z_p))}{\partial Z_p}$ and $P_i = \frac{\partial P}{\partial x_i}$ $(i = 1, 2)$.

The Lagrange multiplier u^0 describes the effect of a change of the environmental standards for the profit of the firm and P_i expresses the increase of pollution caused by increasing the ith input by a small unit (i.e., the marginal pollution with respect to the input i):

$$\frac{\partial \Phi}{\partial x_3} = -c_3 - uP_3 \leqq 0,$$

$$x_3 \frac{\partial \Phi}{\partial x_3} = x_3(-c_3 - uP_3) = 0.$$

$x_3 > 0$ implies that $c_3 = -uP_3$, where $P_3 < 0$. Therefore, the value of the pollution reduction caused by one additional unit of abatement equipment is equal to its cost.

Finally, we obtain

$$\frac{\partial \Phi}{\partial u} = Z_p - P(x_1, x_2, x_3) \geqq 0,$$

$$u \frac{\partial \Phi}{\partial u} = u(Z_p - P(x_1, x_2, x_3)) = 0,$$

$$u \geqq 0.$$

We conclude that $P(x_1, x_2, x_3) < Z_p$ implies that $u = 0$. In this case, equalities (2.73) and (2.76) reduce to $pf_1 = c_1$, $pf_2 = c_2$, and $c_3 > 0$ implies that $x_3 = 0$. The economic interpretation of this result is straightforward: If the net pollution is below the given level of emissions, no abatement will be necessary, and we get the same solution as in the unregulated case.

For $u > 0$, we obtain $P(x_1, x_2, x_3) = Z_p$, and in the case of essential production factors $(x_1 > 0,\ x_2 > 0)$, conditions (2.73) and (2.76). Because $u > 0$ and $P_1 > 0$, $P_2 > 0$, the value of marginal product of the input i $(i = 1, 2)$ under regulation must be higher than in the unregulated case (see conditions (2.56)–(2.57), (2.73), and (2.76)).

Under the assumption (2.70) on the production function $f(x_1, x_2)$ that the marginal products are decreasing, we can conclude

$$pf_1^{\mathrm{I}} > pf_1^0 \text{ implies } x_1^{\mathrm{I}} < x_1^0,$$

$$pf_2^{\mathrm{I}} > pf_2^0 \text{ implies } x_2^{\mathrm{I}} < x_2^0,$$

where the superscript I denotes the model with environmental constraint expressed as a permissible amount of total pollution and 0 denotes the model without regulation. The effect of this type of environmental regulation is obvious: Both inputs are decreasing, and therefore the level of production also decreases. This is the only way the firm can meet the environmental constraint.

2.5.4.2 Standard as Emissions per Unit of Output

Let Z_{PF} be the emission standard expressed as a set level of pollution per unit of output. This amount of emission may be discharged into the environment at a zero price. The regulatory constraint then becomes

$$\frac{P(x_1, x_2, x_3)}{f(x_1, x_2)} \leqq Z_{\mathrm{PF}},$$

and the objective function of the firm is the profit maximization as in the previous model.

The Lagrange function is

$$\Phi = pf(x_1, x_2) - c_1 x_1 - c_2 x_2 - c_3 x_3 + u(Z_{\mathrm{PF}} f(x_1, x_2) - P(x_1, x_2, x_3)),$$

and the Kuhn–Tucker conditions are

$$\frac{\partial \Phi}{\partial x_1} = pf_1 - c_1 + u(Z_{\mathrm{PF}} f_1 - P_1) \leqq 0, \tag{2.77}$$

$$x_1 \frac{\partial \Phi}{\partial x_1} = x_1[pf_1 - c_1 + u(Z_{\mathrm{PF}} f_1 - P_1)] = 0, \tag{2.78}$$

$$\frac{\partial \Phi}{\partial x_2} = pf_2 - c_2 + u(Z_{\mathrm{PF}} f_2 - P_2) \leqq 0, \tag{2.79}$$

$$x_2 \frac{\partial \Phi}{\partial x_2} = x_2[pf_2 - c_2 + u(Z_{\mathrm{PF}} f_2 - P_2)] = 0, \tag{2.80}$$

$$\frac{\partial \Phi}{\partial x_3} = -c_3 - uP_3 \leqq 0, \tag{2.81}$$

$$x_3 \frac{\partial \Phi}{\partial x_3} = x_3(-c_3 - uP_3) = 0, \tag{2.82}$$

$$\frac{\partial \Phi}{\partial u} = Z_{\mathrm{PF}} f(x_1, x_2) - P(x_1, x_2, x_3) \geqq 0, \tag{2.83}$$

$$u \frac{\partial \Phi}{\partial u} = u[Z_{\mathrm{PF}} f(x_1, x_2) - P(x_1, x_2, x_3)] = 0, \tag{2.84}$$

$$u \geqq 0. \tag{2.85}$$

If $Z_{\mathrm{PF}} f(x_1, x_2) > P(x_1, x_2, x_3)$, then $u = 0$, and because $c_3 > 0$, it follows from (2.82) that $x_3 = 0$.

If the price of abatement equipment c_3 is higher than the value of the pollution reduced by one additional unit of abatement equipment $-uP_3$, then the abatement expenditure x_3 will be zero.

Assuming essential production factors ($x_1 > 0$, $x_2 > 0$), the Kuhn–Tucker conditions (2.77) and (2.79) become equalities:

$$pf_1 - c_1 + u(Z_{PF} f_1 - P_1) = 0,$$
$$pf_2 - c_2 + u(Z_{PF} f_2 - P_2) = 0,$$

or

$$f_1 = \frac{c_1 + uP_1}{p + uZ_{PF}}, \tag{2.86}$$

$$f_2 = \frac{c_2 + uP_2}{p + uZ_{PF}}, \tag{2.87}$$

and therefore

$$\frac{f_1}{f_2} = \frac{c_1 + uP_1}{c_2 + uP_2}. \tag{2.88}$$

We can see that for $u > 0$ (the pollution constraint is binding), the ratio of marginal products cannot equal the ratio of the input prices, as was the case in the absence of the regulatory constraint.

In order to show the effect of the environmental constraint (2.83) for the behavior of the firm, we compare the optimality conditions (2.86)–(2.87) with the optimality conditions without environmental standard (2.56)–(2.57).

Let the superscript II denote the solution of the model with environmental constraint expressed as the maximum amount of emissions per unit of output, i.e., the model in this section.

Recall the first-order optimality conditions for the unregulated firm:

$$f_1^0 = \frac{c_1}{p} \quad \text{and} \quad f_2^0 = \frac{c_2}{p} \tag{2.89}$$

for a given price p of the output.

Comparison of (2.89) with (2.86)–(2.87) reveals that the effect of the environmental regulatory constraint (2.83) on the production of the firm is ambiguous. It depends on the relation between the expressions on the right side of (2.86)–(2.87) and (2.89), respectively. If $P_i^{II} < \frac{c_i Z_{PF}}{p}$, then $\frac{c_i + uP_i^{II}}{p + uZ_{PF}} < \frac{c_i}{p}$; therefore, $f_i^{II} < f_i^0$, and consequently, due to the assumption (2.70), $x_i^{II} > x_i^0$ $(i = 1, 2)$. If the marginal pollution with respect to the input i is lower than the exogenously given constant $k_i = \frac{c_i Z_{PF}}{p}$, then the amount of input i used in production will—compared with the basic model—increase.

In the opposite case, if the marginal pollution with respect to the input i is relatively high (higher than the parameter k_i), the amount of input i used in production will—in order to fulfill the environmental standard—decrease.

The effects on production, and therefore (taking into account the possible abatement activity) on the level of pollution, remain ambiguous. To summarize, the effect of the standard defined as emissions per unit of output can lead to similar results as in the Averch–Johnson model: Pollution increases with the imposition of an environmental regulatory constraint. If production increases more rapidly than pollution, the environmental standard can be achieved in spite of increasing pollution.

2.5.4.3 Standard as Emissions per Unit of a Specified Input

Another way in which individual stack policy can be effected is to fix an upper bound for the emissions per unit of specified input, such as restricting the amount of sulfur dioxide emissions per ton of coal used for electricity. Such a type of limitation is referred to in [8] as *intensity regulation* and can be formalized as

$$\frac{P(x_1, x_2, x_3)}{x_i} \leqq Z_{Pi}, \qquad \text{for } i = 1, 2.$$

Without loss of generality we suppose that the intensity regulation is imposed for the second production factor. Then the firm will face the following optimization problem:

$$\begin{array}{ll} \text{maximize} & \Pi = pf(x_1, x_2) - c_1, x_1 - c_2 x_2 - c_3 x_3 \\ \text{subject to} & P(x_1, x_2, x_3) \leqq Z_{P_2} x_2, \\ & x_1 \geqq 0, \quad x_2 \geqq 0, \quad x_3 \geqq 0. \end{array}$$

Using the Lagrange function,

$$\Phi(\mathbf{x}, u) = pf(x_1, x_2) - c_1 x_1 - c_2 x_2 - c_3 x_3 + u(Z_{P_2} x_2 - P(x_1, x_2, x_3)),$$

the Kuhn–Tucker conditions are

$$\frac{\partial \Phi}{\partial x_1} = pf_1 - c_1 - uP_1 \leqq 0,$$

$$x_1 \frac{\partial \Phi}{\partial x_1} = x_1(pf_1 - c_1 - uP_1) = 0, \tag{2.90}$$

$$\frac{\partial \Phi}{\partial x_2} = pf_2 - c_2 + u(Z_{P_2} - P_2) \leqq 0,$$

$$x_2 \frac{\partial \Phi}{\partial x_2} = x_2[pf_2 - c_2 + u(Z_{P_2} - P_2)] = 0, \tag{2.91}$$

$$\frac{\partial \Phi}{\partial x_3} = -c_3 - uP_3 \leqq 0,$$

$$x_3 \frac{\partial \Phi}{\partial x_3} = x_3(-c_3 - uP_3) = 0,$$

$$\frac{\partial \Phi}{\partial u} = Z_{P_2} x_2 - P(x_1, x_2, x_3) \geqq 0,$$

$$u \frac{\partial \Phi}{\partial u} = u[Z_{P_2} x_2 - P(x_1, x_2, x_3)] = 0,$$

$$u \geqq 0.$$

Let the superscript III denote the solution of the model with intensity regulation. Again assuming essential production factors ($x_1 > 0$, $x_2 > 0$) and $u > 0$ (the environmental standard is binding), the Kuhn–Tucker condition (2.90) yields

$$f_1^{\mathrm{III}} = \frac{c_1}{p} + \frac{u}{p}P_1 > \frac{c_1}{p} = f_1^0.$$

Due to the assumption (2.70), we have $x_1^{\mathrm{III}} < x_1^0$; the firm decreases the amount of the first input used.

For the reaction with respect to the regulated input, we look at the Kuhn–Tucker condition (2.91). It provides

$$f_2^{\mathrm{III}} = \frac{c_2}{p} + \frac{u}{p}(P_2 - Z_{P2}).$$

If the marginal pollution with respect to the second input (P_2) is higher than the allowable amount of emissions per unit of this input, then the marginal product f_2^{III} is higher than the marginal product f_2^0 ($= \frac{c_2}{p}$) in the absence of the regulatory constraint. Therefore, due to the declining marginal product, the amount of the regulated input used under the intensity regulation x_2^{III} is lower than without such regulation x_2^0. The firm will decrease the level of production. If the marginal pollution P_2 is lower than the tolerated amount of emissions per unit of the second input, we get the opposite result. Because in this case the marginal product f_2^{III} is lower than the marginal product f_2^0, the amount of the regulated input used in the optimal solution x_2^{III} is higher than the amount x_2^0 used without the regulatory constraint. We have the Averch–Johnson effect with respect to the second input; the substitution of the first input by the second one.

More than fifty years after their formulation, the Kuhn–Tucker conditions became a standard instrument of the analysis used in the textbooks of microeconomic theory (e.g., [28, 33]) and in the monographs devoted to various fields of economics like the theory of money [29], public economics [7], or industrial economics [6, 13]. Moreover, they provide the foundation for the development of more complex optimization models dealing with multiple objectives or with dynamical economic systems.

References and Further Reading

[1] K. J. Arrow, L. Hurwicz, and H. Uzawa, Constraint qualifications in maximization problems, *Naval Res. Logist. Q.*, **8** (1961), 175–190.

[2] H. Averch and L. L. Johnson, Behavior of the firm under regulatory constraint, *Amer. Econ. Rev.*, **52** (1962), 1052–1069.

[3] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming*, 2nd ed., Wiley, New York, 1993.

[4] W. J. Baumol, *Economic Theory and Operations Analysis*, 3rd ed, Prentice–Hall, Englewood Cliffs, NJ, 1972.

[5] H. Bester, *Theorie der Industrieökonomik*, Springer-Verlag, Berlin, 2000.

[6] J. Bormann and J. Finsinger, *Markt und Regulierung*, Verlag Vahlen, Munich, 1999.

[7] D. Bös, *Pricing and Price Regulation: An Economic Theory for Public Enterprises and Public Utilities*, North-Holland, Amsterdam, 1994.

[8] F. Dudenhöffer, The regulation of intensities and productivities: Concepts in environmental policy: Zeitschrift für die gesamte Staatswirtschaft, *J. Institutional Theoret. Econ.*, **140** (1984), 276–287.

[9] M. A. El-Hodiri and A. Takayama, Behavior of the firm under regulatory constraints: Clasifications, *Amer. Econ. Rev.*, **63** (1973), 235–237.

[10] H. A. Eiselt, G. Pederzoli, and C. L. Sandblom, *Continuous Optimization Models*, de Gruyter, Berlin, 1987.

[11] A. Fiacco and G. P. McCormick, *Nonlinear Programming*, Wiley, New York, 1968.

[12] R. Fletcher, Penalty functions, in A. Bachem, M. Grötschel, and B. Korte, eds., *Mathematical Programming: The State of the Art: Bonn* 1982, Springer-Verlag, Berlin, 1983, 87–114.

[13] M. W. Frank, *The Impact of Rate-of-Return Regulation on Technological Innovation*, Ashgate, Aldershot, UK, 2001.

[14] G. E. Helfand, Standards versus standards: The effects of different pollution restrictions, *Amer. Econ. Rev.*, **81**-3 (1991), 622–634.

[15] F. John, Extremum problems with inequalities as side conditions, in K. O. Fridrichs, O. E. Neugebauer, and J. J. Stoker, eds., *Studies and Essays: Courant Anniversary Volume*, Wiley, New York, 1948.

[16] W. Karush, *Minima of Functions of Several Variables with Inequalities as Side Conditions*, M. S. thesis, Department of Mathematics, University of Chicago, Chicago, 1939.

[17] G. Knieps, *Wettbewerbsökonomie: Regulierungstheorie, Industrieökonomie, Wettbewerbspolitik*, Springer-Verlag, Berlin, 2001.

[18] H. W. Kuhn and A. W. Tucker, Nonlinear programming, in J. Neyman, ed., *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, CA, 1951, 481–492.

[19] J. J. Laffont and J. Tirole, *A Theory of Incentives in Procurement and Regulation*, MIT Press, Cambridge, MA, 2000.

[20] S. C. Littlechild, Peak-load pricing of telephone calls, *Bell J. Econ. Manage. Sci.*, **1**-2 (1970), 191–210.

[21] S. C. Littlechild, *Regulation of British Telecommunications' Profitability*, Her Majesty's Stationery Office, London, 1983.

[22] C. Liston, Price-cap versus rate-of-return regulation, *J. Regul. Econ.*, **5** (1993), 25–48.

[23] D. L. Luenberger, *Linear and Nonlinear Programming*, 1st and 2nd eds., Addison–Wesley, Reading, MA, 1973, 1984.

[24] M. Luptáčik, *Nichtlineare Programierung mit ökonomischen Anwendungen*, Athenäum, Königstein, Germany, 1981.

[25] M. Luptáčik, Price-Cap-Regulierung und Benchmarking der österreichischen Netzbetrieber (Price-cap regulation and benchmarking of the Austrian electricity grid companies), in T. Kostal, M. Leibrecht, and A. Pasterniak, eds., *Organisation und Funktionen des Staates im Wandel: Festschrift für Gabriel Obermann*, Facultas.wuv Universitätsverlag, Vienna, 2008, 203–216.

[26] O. L. Mangasarian, *Nonlinear Programming*, McGraw–Hill, New York, 1969.

[27] G. P. McCormick, *Nonlinear Programming: Theory, Algorithms, and Applications*, Wiley, New York, 1983.

[28] W. Nicholson, *Microeconomic Theory: Basic Principles and Extensions*, 5th ed., Dryden Press/Harcourt–Brace–Jovanovich, New York, 1992.

[29] J. Niehans, *The Theory of Money*, Johns Hopkins University Press, Baltimore, 1978.

[30] H. Nikaido, *Convex Structure and Economic Theory*, Academic Press, New York, 1968.

[31] H. Siebert, *Economics of the Environment*, Lexington Books, New York, 1981.

[32] H. Siebert, Some open questions, in H. Siebert, ed., *Umweltallokation im Raum*, Lang, Frankfurt am Main, Germany, 1982, 1–30.
[33] E. Silberberg and W. Suen, *The Structure of Economics: A Mathematical Analysis*, 3rd ed., McGraw–Hill, New York, 2001.
[34] V. K. Smith, The implications of regulation for induced technical change, *Bell J. Econ. Manage. Sci.*, **5**-2 (1974), 623–636.
[35] P. O. Steiner, Peak loads and efficient pricing, *Q. J. Econ.*, **54** (1964), 64–76.
[36] A. Takayama, Behavior of the firm under regulatory constraint, *Amer. Econ. Rev.*, **59** (1969), 255–260.
[37] A. Takayama, *Mathematical Economics*, Cambridge University Press, Cambridge, UK, 1985.
[38] O. E. Williamson, Peak load pricing and optimal capacity under indivisibility constraints, *Amer. Econ. Rev.*, **56** (1966), 810–827.

http://www.springer.com/978-0-387-89551-2

Mathematical Optimization and Economic Analysis
Luptácik, M.
2010, XIV, 294 p. 53 illus., Hardcover
ISBN: 978-0-387-89551-2