Chapter 2
Aggregate Models of Economic Dynamics

This chapter explores aggregate optimization models of the neoclassic economic growth theory, which are based on the concept of production functions. The models are described by ordinary differential equations and involve static and dynamic optimization. Section 2.1 analyzes production functions with several inputs, their fundamental characteristics, and major types (Cobb–Douglas, CES, Leontief, and linear). Special attention is given to two-factor production functions and their use in the neoclassic models of economic growth. Sections 2.2 and 2.3 describe and analyze the well-known Solow–Swan and Solow–Ramsey models. Section 2.4 contains maximum principles used to analyze dynamic optimization problems in this and other chapters.

2.1 Production Functions and Their Types

A *production function* describes a relationship

$$y = f(x_1, \ldots, x_n) \tag{2.1}$$

between the aggregate product output y and the *productive inputs* $x_1, \ldots, x_n$ that can include labor, capital, knowledge (human capital), energy consumption, raw materials, natural resources (land, water, minerals), and others. The output y and inputs x_i are assumed to be identical. For example, the labor is the quantity of workers indistinguishable in a productive sense.

Henceforth, we will often use the following definition. The function $r(t) = f'(t)/f(t)$ is the *relative rate* of the function $f(t)$ and is often referred to as the *growth rate* of $f(t)$. If $r \equiv \text{const}$, then $f(t) = C\exp(rt)$.

Economists often use the notation $\dot{f}$ for the derivative of a function f in time. We will keep the standard notation f'.

N. Hritonenko and Y. Yatsenko, *Mathematical Modeling in Economics, Ecology and the Environment*, Springer Optimization and Its Applications 88,
DOI 10.1007/978-1-4614-9311-2_2,

2.1.1 Properties of Production Functions

Commonly accepted properties of production functions are:

1. **Essentiality of inputs**: If at least one $x_i = 0$, then $y = 0$, i.e., production is not possible without any of the inputs.
2. **Positive returns**: $\partial f/\partial x_i > 0$, $i = 1, \ldots, n$, i.e., the output increases if an input increases.
3. **Diminishing returns**: The Hessian matrix

$$H = \begin{bmatrix} \partial^2 f/\partial {x_1}^2 & \ldots & \partial^2 f/\partial x_1 \partial x_n \\ \ldots & \ldots & \ldots \\ \partial^2 f/\partial x_n \partial x_1 & \ldots & \partial^2 f/\partial {x_n}^2 \end{bmatrix} \tag{2.2}$$

   is negatively definite. It means that if only one input x_i increases and the other inputs x_j, $j \neq i$, remain constant, then the efficiency of using the input x_i decreases.
4. **Proportional returns to scale**: $f(\mathbf{x})$ is a *homogeneous function* of degree $\gamma > 0$, i.e.,

$$f(l\mathbf{x}) = l^{\gamma} f(\mathbf{x}), \quad l \in R^1, \quad l > 0, \quad \mathbf{x} = (x_1, \ldots, x_n). \tag{2.3}$$

   The production function $f(\mathbf{x})$ exhibits *increasing returns to scale* at $\gamma > 1$, *decreasing returns to scale* at $\gamma < 1$, and *constant returns to scale* at $\gamma = 1$. The increasing returns mean that a 1 % increase in the levels of all inputs leads to a greater than the 1 % increase of the output y.

   In the case of *constant returns to scale*, the function $f(\mathbf{x})$ is *linearly homogeneous*: $f(l\mathbf{x}) = lf(\mathbf{x})$, and the output increases linearly with respect to a proportional increase of all inputs: a 1 % increase of all inputs produces exactly the 1 % increase of the output. Then, the condition (2.2) is reduced to

$$\partial^2 f/\partial {x_i}^2 < 0, \quad i = 1, \ldots, n. \tag{2.4}$$

2.1.2 Characteristics of Production Functions

The major characteristics of production functions are

- The *average product* $f(x_1, \ldots, x_n)/x_i$ of the i-th input is the output per one unit of the input x_i spent, $i = 1, \ldots, n$.
- The *marginal product* $\partial f/\partial x_i$ of the i-th input describes the additional output obtained due to the increase of the i-th input quantity by one unit.
- The *isoquant* is the set of all possible combinations of inputs $\mathbf{x} = (x_1, \ldots, x_n)$ that yield the same level of the output $y = f(\mathbf{x})$. Along an isoquant, the differential of the function $f(\mathbf{x})$ is zero: $\sum_{i=1}^{n} (\partial f/\partial x_i) dx_i = 0$.

- The *marginal rate of substitution* between the inputs i and j

$$h_{ij} = (\partial f/\partial x_i)/\left(\partial f/\partial x_j\right) \tag{2.5}$$

shows how many units of the j-th input are required to substitute one unit of the i-th input in order to produce the same level of the output y.

- The *partial elasticity of output* with respect to the input i

$$\varepsilon_i(\mathbf{x}) = (\partial f(\mathbf{x})/\partial x_i)/(f(\mathbf{x})/x_i) = \partial \ln f(\mathbf{x})/\partial \ln x_i \tag{2.6}$$

is the ratio between the marginal product and the average product of the i-th input. It describes the increase of the output y when the i-th input increases by 1 %.

- The *total output elasticity* $\varepsilon(\mathbf{x}) = \sum_{i=1}^{n} \varepsilon_i(\mathbf{x})$ describes the output increase under a proportional production scale extension. For a homogeneous production function (2.3), $\varepsilon(\mathbf{x}) = \gamma$.
- The *elasticity of substitution* is a quantitative measure of a possibility of changes in the input combination to produce the same output. It is equal to the relative change in the ratio of the i-th and j-th inputs divided by the relative change in their marginal rate of substitution h_{ij}:

$$\sigma_{ij} = \frac{\mathrm{d}\left(x_i/x_j\right)}{\left(x_i/x_j\right)} \times \frac{h_{ij}}{\mathrm{d}h_{ij}} = \frac{\mathrm{d}\ln\left(x_i/x_j\right)}{\mathrm{d}\ln h_{ij}}. \tag{2.7}$$

This characteristic shows the percentage change of the ratio x_i/x_j of these inputs along an isoquant in order to change their marginal substitution rate by one percent. The larger the σ_{ij}, the greater the *substitutability* between the two inputs. The inputs i and j are *perfect substitutes* at $\sigma_{ij} = \infty$ and they are not substitutable at all at $\sigma_{ij} = 0$. The elasticity of substitution is used for classification of various production functions.

2.1.3 Major Types of Production Functions

2.1.3.1 Linear Production Function

$$y = a_1x_1 + \ldots + a_nx_n, \quad a_i > 0, \quad i = 1, \ldots, n, \tag{2.8}$$

has the following characteristics:

$$\partial f/\partial x_i = a_i, \quad h_{ij} = a_j/a_i = \text{const}, \quad \varepsilon = 1, \quad \sigma_{ij} = \infty, \quad i,j = 1, \ldots, n,$$

i.e., constant returns to scale, constant marginal rates of substitution, all inputs are perfectly substitutable. Despite its mathematical simplicity, the linear production function with several inputs is rarely used in the economic theory because it violates the fundamental economic property of the input essentiality (Sect. 2.1.1). However, the linear production function with one input (capital) known as the *AK production function* has been intensively investigated, see (2.30) below.

2.1.3.2 Cobb–Douglas Production Function

$$y = A{x_1}^{\alpha_1} \times \ldots \times {x_n}^{\alpha_n} \tag{2.9}$$

has the following characteristics:

$$\partial f/\partial x_i = \alpha_i(y/x_i), \quad h_{ij} = \alpha_j x_i/\alpha_i x_j, \quad \varepsilon_i = \alpha_i, \quad \varepsilon = \alpha_1 + \ldots + \alpha_n,$$
$$\sigma_{ij} = 1, \quad i,j = 1, \ldots, n,$$

i.e., the elasticity of substitution for any pair (i, j) of inputs is equal to one. The *returns to scale* are *increasing* in the case $\alpha_1 + \ldots + \alpha_n > 1$, *decreasing* in the case $\alpha_1 + \ldots + \alpha_n < 1$, and *constant* at $\alpha_1 + \ldots + \alpha_n = 1$.

By taking the logarithm of both sides of (2.9), we obtain a linear expression

$$\ln y = \ln A + \sum_{i=1}^{n} \alpha_i \ln x_i,$$

that after differentiation becomes

$$y'/y = \alpha_1\left(x_1'/x_1\right) + \ldots + \alpha_n\left(x_n'/x_n\right), \tag{2.10}$$

i.e., the growth rate of the output in the Cobb–Douglas production function is equal to the weighted sum of the growth rates of the inputs.

2.1.3.3 Production Function with Fixed Proportions

$$y = A\min(x_1, \ldots, x_n), \tag{2.11}$$

has the following characteristics:

$$h_{ij} = \begin{cases} 0, & x_j > x_i \\ \infty, & x_j < x_i \end{cases}, \quad \varepsilon = 1, \quad \sigma_{ij} = 0, \quad i,j = 1, \ldots n,$$

i.e., the elasticity of substitution for all inputs is zero (the inputs are not substitutable). This production function is also known as the *piecewise-linear production function* and the *Leontief production function*.

2.1.3.4 Production Function with Constant Elasticity of Substitution (CES)

$$y = A\,[\beta_1 x_1{}^{\rho} + \ldots + \beta_n x_n{}^{\rho}]^{\gamma/\rho}, \tag{2.12}$$

where $\rho < 1$, $\rho \neq 0$, $\beta_i > 0$, $i = 1, \ldots, n$, $\beta_1 + \ldots + \beta_n = 1$, has the following characteristics:

$$h_{ij} = (\beta_i/\beta_j)(x_j/x_i)^{1-\rho}, \quad \varepsilon = \gamma, \quad \sigma_{ij} = 1/(1-\rho), \quad i,j = 1, \ldots, n,$$

i.e., the elasticity of substitution is a positive constant. The CES production function is also known as the *Solow production function*.

The CES production function is the most general among the production functions considered above: it leads to the linear production function as $\rho \to 1$, to the Cobb–Douglas production function as $\rho \to 0$, and to the production function with fixed proportions as $\rho \to -\infty$.

The Cobb–Douglas and CES production functions are frequently used in low-sector aggregate economic models (see Sects. 2.2–2.4, 3.3, 3.4, 10.2), whereas the production function with fixed proportions is used in the multi-sector input–output models and determines fixed sets of productive technologies in specific industries.

2.1.4 Two-Factor Production Functions

Production functions with two inputs, called *two-factor production functions*, are the most common in economics and are usually written as

$$Q = F(K, L), \tag{2.13}$$

where

Q is the output,

K is the amount of capital used,

L is the amount of labor used.

The capital K reflects the total cost of the equipment, machines, buildings, etc., used in production process. Such production functions are characterized by single values of the marginal rate of substitution h and the elasticity of substitution σ between capital and labor.

A two-factor production function is called the *neoclassical production function*, if it satisfies the following properties:

1. *Essentiality of inputs*:

$$F(K,0) = F(0,L) = 0. \tag{2.14}$$

2. *Positive and diminishing returns*:

$$\partial F/\partial K > 0, \quad \partial F/\partial L > 0, \quad \partial^2 F/\partial K^2 < 0, \quad \partial^2 F/\partial L^2 < 0. \tag{2.15}$$

3. *Constant returns to scale*: $F(K, L)$ is a linearly homogeneous function,

$$F(lK, lL) = lF(K,L) \quad \text{for } l > 0. \tag{2.16}$$

4. *The Inada conditions*: the marginal products of capital and labor satisfy

$$\lim_{K\to 0}\frac{\partial F}{\partial K} = \infty, \quad \lim_{L\to 0}\frac{\partial F}{\partial L} = \infty, \quad \lim_{K\to\infty}\frac{\partial F}{\partial K} = 0, \quad \lim_{L\to\infty}\frac{\partial F}{\partial L} = 0. \tag{2.17}$$

The Inada conditions mean that the production increases very fast if the production input (capital or labor) is low and increases slowly, whereas the production increase is very slow if the production input has been already abundant and more is added. Property 1 holds if the other three properties hold.

Per capita variables. At condition (2.16), the production function (2.13) can be rewritten as $Q = LF(K/L, 1)$ or in the so-called *intensive form* (or *per capita* form) as

$$q = f(k), \quad f(k) = F(k,1) \tag{2.18}$$

where
$q = Q/L$ is the *output per worker* or the *productivity*,
$k = K/L$ is the *capital per worker* or the *capital–labor ratio*.

The intensive form (2.18) of production functions is more convenient for analysis and illustration because it reduces the number of variables. Then, the marginal products of capital and labor are

$$\partial F/\partial K = f'(k), \quad \partial F/\partial L = f(k) - kf'(k), \tag{2.19}$$

the marginal rate of substitution between labor and capital is

$$h = \frac{\partial F/\partial L}{\partial F/\partial K} = \frac{f(k) - kf'(k)}{f'(k)}, \tag{2.20}$$

and *Properties 1–4 of the neoclassical production function* become

$$f(0) = 0, \quad f^{'}(k) > 0, \quad f^{''}(k) < 0, \quad \lim_{k\to 0} f^{'}(k) = \infty, \qquad \lim_{k\to\infty} f^{'}(k) = 0. \qquad (2.21)$$

The two-factor versions of the major production functions are provided below.

2.1.4.1 Two-Factor Cobb–Douglas Production Function

$$Q = AK^{\alpha}L^{\beta}, \quad \alpha > 0, \quad \beta > 0, \qquad (2.22)$$

where the *total factor productivity A* reflects the level of technology. In the general case when $\alpha + \beta \neq 1$, the Cobb–Douglas production is not neoclassical because it does not satisfy Property 3 of constant returns. The Cobb–Douglas production at $\alpha + \beta = 1$ is neoclassical and can be presented in the standard and intensive forms as

$$Q = AK^{\alpha}L^{1-\alpha} \text{ or } q = Ak^{\alpha}, \quad 0 < \alpha < 1. \qquad (2.23)$$

Then, the marginal products of capital and labor of (2.22) are

$$\partial Q/\partial K = \alpha A k^{\alpha-1}, \quad \partial Q/\partial L = (1-\alpha)Ak^{\alpha}, \qquad (2.24)$$

the marginal rate of substitution is $h = k\,(1-\alpha)/\alpha$, the output elasticity of capital is $\varepsilon_K = \alpha$, the total output elasticity is $\varepsilon = 1$, and the elasticity of substitution is $\sigma = 1$. The graph of the Cobb–Douglas production function (2.23) in the intensive form is shown Fig. 2.1 with a black curve and is typical for the neoclassical production functions. The output $f(k)$ increases indefinitely when the capital per capita $k \to \infty$, which reflects the Inada condition (2.17). Some economists consider such an increase to be unrealistic.

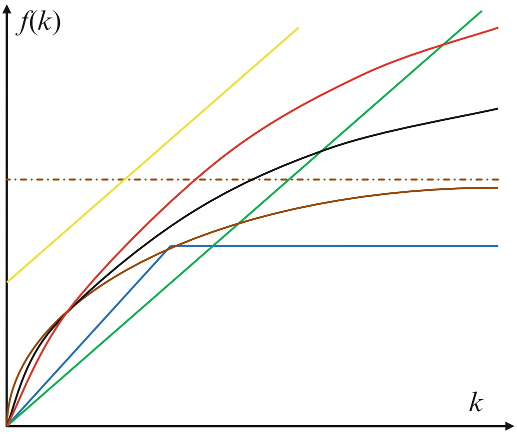


Fig. 2.1 The major types of production functions in per capita form: two-factor Cobb–Douglas (*black curve*), two-factor CES with $\sigma < 1$ (*brown curve*), two-factor CES with $\sigma > 1$ (*red curve*), two-factor Leontief (*blue curve*), two-factor linear (*yellow curve*), and one-factor linear or *AK* production function (*green curve*)

2.1.4.2 Two-Factor CES Production Function

$$Q = A[\alpha(bK)^{\rho} + (1-\alpha)((1-b)L)^{\rho}]^{1/\rho}, \quad \rho < 1 \tag{2.25}$$

$$\text{or} \quad q = A[\alpha(bk)^{\rho} + (1-\alpha)(1-b)^{\rho}]^{1/\rho}. \tag{2.26}$$

Here, the marginal product of capital (2.19) is

$$\partial Q/\partial K = A\alpha b^{\rho}[\alpha b^{\rho} + (1-\alpha)(1-b)^{\rho}k^{-\rho}]^{(1-\rho)/\rho}, \tag{2.27}$$

$$h = (1-\alpha)(1-b)^{\rho}k^{1-\rho}/(\alpha b^{\rho}), \quad \sigma = 1/(1-\rho)$$

The CES production function is not neoclassical because the Inada conditions are violated. It is visible in Fig. 2.1. At a low degree of substitution $\sigma < 1$ ($\rho < 0$), its graph has a horizontal asymptote (see the brown curve in Fig. 2.1). When $\rho \to 0$, the CES production function approaches the Cobb–Douglas production function. At a high degree of substitution $\sigma > 1$ ($0 < \rho < 1$), this function increases faster than the Cobb–Douglas one (see the red curve in Fig. 2.1). At $\sigma = \infty$ ($\rho = 1$), the CES function becomes linear: $Q = A\alpha bK + A(1-\alpha)(1-b)L$. When $\rho \to -\infty$ ($\sigma \to 0$), this production function approaches the Leontief production function $Q = \min[bK, (1-b)L]$ discussed next. There is essential economic evidence that the CES production function better fits many economic processes than the Cobb–Douglas production function. For this reason, the CES production function currently dominates in applied economic research.

We shall notice that some textbooks introduce the CES production function in a slightly different form as $Q = A[\alpha K^{\rho} + (1-\alpha)L^{\rho}]^{1/\rho}$ and/or with the parameter ρ replaced by $-\rho$ (then the new $\rho > -1$).

2.1.4.3 Two-Factor Leontief Production Function (with Fixed Proportions)

$$Q = \min\{aK, bL\} \quad \text{or} \quad q = \min\{ak, b\}. \tag{2.28}$$

Here, the marginal rate of substitution is $h = \infty$ at $k > b/a$ and $h = 0$ at $k < b/a$, and there is no substitution between capital and labor: $\sigma = 0$. This function was introduced by the famous American economist W. Leontief in 1940, well before other types of production functions. It can be obtained from the two-factor function CES (2.25) at $\rho \to -\infty$. This function is built on the suggestion that there is a unique reasonable value $k_0 = b/a$ of the capital labor ratio $k = K/L$ such that all workers and machines are fully employed. An additional capital is useless at $k > b/a$, whereas some part of labor is not used at $k < b/a$. The Leontief function is shown Fig. 2.1 with a blue curve.

2.1.4.4 Two-Factor Linear Production Function

$$Q = AK + BL \quad \text{or} \quad q = Ak + B. \tag{2.29}$$

The marginal and average products of capital and labor of this production function are constant and equal to A, $h = B/A = \text{const}$, and capital and labor are perfect substitutes: $\sigma = \infty$. A primary weakness of the two-factor linear production function (2.29) is that the input essentiality property (2.14) fails, i.e., the production is possible without capital or labor (see the yellow line in Fig. 2.1). This shortcoming disappears in the *one-factor linear production function*

$$Q = AK \quad \text{or} \quad q = Ak, \tag{2.30}$$

also known as the *AK production function* and commonly used in mathematical economics. It is shown Fig. 2.1 with a green line. Both linear production functions (2.29) and (2.30) do not satisfy the property (2.15) of diminishing returns and, thus, do not belong to neoclassical functions. Some modern economists have considered the property (2.15) as obsolete and not applicable to the capital in a broad sense that includes the *human capital* (see Sect. 3.4).

2.2 Solow–Swan Model of Economic Dynamics

The one-sector model explored below is one of the most celebrated models in the economic growth theory [9]. It has become a foundation for further successful studies.

2.2.1 Model Description

Let us consider an economy described by the following dynamic characteristics in the continuous time t:

$Q(t)$—the total *output* produced at time t,
$C(t)$—the amount of *consumption*,
$I(t)$—the amount of gross *investment*,
$L(t)$—the amount of *labor*,
$K(t)$—the amount of *capital*.

The Solow–Swan model is described by the following equations:

$$Q(t) = F(K(t), L(t)), \tag{2.31}$$

i.e., the output Q is determined by a neoclassical production function $F(K, L)$,

$$Q(t) = C(t) + I(t), \tag{2.32}$$

i.e., the output Q is distributed between the consumption C and the investment I,

$$K^{'}(t) = I(t) - \mu K(t), \quad \mu = \text{const} > 0, \tag{2.33}$$

i.e., the capital K depreciates at a constant rate $\mu > 0$ (a constant fraction of the capital leaves a production process at each point of time),

$$L^{'}(t) = \eta L(t), \quad \eta = \text{const} \geq 0 \tag{2.34}$$

i.e., the labor $L(t) = L_0 \exp(\eta t)$ grows at a constant exogenous rate η.

The structure of the Solow–Swan model is shown in Fig. 2.2. The part of the investment in the total product is known as the *saving rate*:

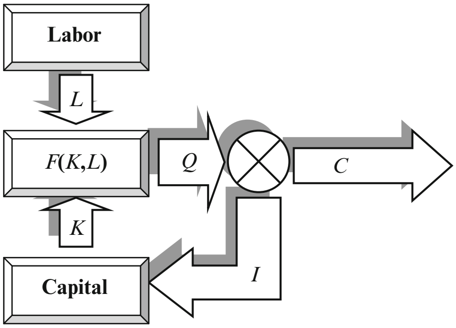
$$s(t) = I(t)/Q(t)$$

The saving rate is assumed to be constant in the classic Solow–Swan model:

$$I(t) = sQ(t), \quad 0 < s < 1, \quad s = \text{const}. \tag{2.35}$$

This assumption simplifies the investigation of the model and leads to a number of essential economic results. More advanced economic models (see next sections) consider the saving rate $s(t)$ as an endogenous control function.

Fig. 2.2 The flow diagram of the Solow–Swan model



2.2.2 Analysis of Model

2.2.2.1 Fundamental Equation of Model (2.31)–(2.35)

Because the production function $F(K, L)$ is neoclassical and, therefore, linearly homogeneous, then $F(K, L) = Lf(k)$ and the equation (2.33) leads to

$$K'(t)/L(t) = f(k(t)) - \mu k(t),$$

where the capital–labor ratio $k = K/L$ is defined as in (2.18). On the other side,

$$k'(t) = K'(t)/L(t) - \eta k(t)$$

by (2.34). Combining the last two equalities, we obtain the *fundamental equation of the Solow–Swan model*

$$k'(t) = sf(k) - (\mu + \eta)k(t). \tag{2.36}$$

Thus, the dynamics of the model (2.31)–(2.35) is reduced to one *autonomous* (not dependent on t explicitly) differential equation (2.36) with respect to k.

2.2.2.2 Steady-State Analysis

The goal of a steady-state analysis is to find possible *steady states*, which can be

- A *stationary trajectory* (unknown variables are constant in time) or
- A *balanced growth path* (all variables grow at the same constant rate).

The steady-state analysis plays an important role in economics and is mathematically simpler than a complete dynamic analysis.

Let us find and analyze possible balanced growth paths in the model (2.31)–(2.35). It is easy to see that the original variables $Q(t)$, $C(t)$, $I(t)$, and $K(t)$

of the model grow with the same rate only if the capital–labor ratio $k(t)$ is constant. Indeed, substituting $k = \text{const}$ into (2.33), (2.32), and (2.35), we obtain

$$\begin{aligned}&K(t) = kL(t), \quad I(t) = (\mu+\eta)K(t), \quad Q(t) = (\mu+\eta)K(t)/s,\\ &C(t) = Q(t) - I(t),\end{aligned} \tag{2.37}$$

i.e., all these functions increase with the same rate η as the labor $L(t) = L_0\exp(\eta t)$. Therefore, to find steady states, we should assume $k(t) = \text{const}$. Then $k'(t) = 0$ and the equation (2.36) produces the equation

$$sf(k) = (\mu+\eta)k \tag{2.38}$$

for possible steady states $k \equiv \text{const}$. Because $f(0) = 0, f'(k) > 0, \lim_{k\to 0} f'(k) = \infty$, and $\lim_{k\to\infty} f'(k) = 0$, the equation (2.38) has a unique solution $\hat{k} = \hat{k}(s) = \text{const} > 0$ for any given value $s > 0$. The steady-state capital–labor ratio $\hat{k}(s)$ increases when the saving rate s increases.

2.2.2.3 Static Optimization

For a given saving rate s, the steady-state consumption per capita $c = C/L$ is determined by the formula

$$c(s) = f\big(\hat{k}(s)\big) - (\mu+h)\hat{k}(s), \tag{2.39}$$

where the corresponding steady-state capital–labor ratio $\hat{k}(s)$ is determined by (2.38). Because $f(0) = 0$, $f'(k) > 0$ and $f''(k) < 0$, the composite function (2.39) increases for smaller values of s and decreases for larger s.

Then, we can determine the saving rate $s^* = \text{const}$ and the corresponding steady-state $k^* = \hat{k}(s^*)$ that maximizes the consumption per capita (2.39):

$$\max_{0<s\le 1} c(s) = f\big(\hat{k}(s)\big) - (\mu+\eta)\hat{k}(s).$$

The maximization of (2.39) is an optimization problem with one scalar variable s. The necessary extremum condition for (2.39) at an interior $0 < s < 1$ is $c'(s) = 0$ or

$$\mathrm{d}\big[f\big(\hat{k}(s)\big) - (\mu+\eta)\big(\hat{k}(s)\big)/\mathrm{d}s = f'\big(\hat{k}\big)\big) - (\mu+\eta)\big]\mathrm{d}\hat{k}/\mathrm{d}s = 0.$$

Hence, the optimal capital–labor ratio k^* should satisfy

$$f'(k^*) = \mu+\eta. \tag{2.40}$$

The relation (2.40) is known as the *golden rule of capital accumulation*. It implies that the marginal product of capital should be equal to the sum of the depreciation and labor growth rates. After determining the optimal k^* from (2.40), the corresponding *golden-rule saving rate* is found from (2.38) as

$$s^* = (\mu + \eta)k^*/f(k^*) = k^* f'(k^*)/f(k^*), \tag{2.41}$$

i.e., the optimal saving rate s^* is equal to the output elasticity of the capital ε_K (2.6) for the corresponding k^*. The formulas (2.40) and (2.41) for the optimal s^* and k^* are known as the *golden rule of economic growth*.

In the case of the Cobb–Douglas production function (2.22) $F(K, L) = AK^\alpha L^{1-\alpha}$, $0 < \alpha < 1$, the function $f(k) = Ak^\alpha$ and the golden rule is

$$s^* = \alpha, \quad k^* = [As^*/(\mu + \eta)]^{1/(1-\alpha)}. \tag{2.42}$$

At the optimal steady state (s^*, k^*) and the given labor $L(t) = \overline{L}\mathrm{e}^{\eta t}$, the original variables $Q(t)$, $C(t)$, $I(t)$, and $K(t)$ of the model (2.31)–(2.35) grow with the given rate η as

$$K(t) = \overline{K}\mathrm{e}^{\eta t}, \quad I(t) = \overline{I}\mathrm{e}^{\eta t}, \quad Q(t) = \overline{Q}\mathrm{e}^{\eta t}, \quad C(t) = \overline{C}\mathrm{e}^{\eta t}, \tag{2.43}$$

$$\begin{aligned} &\overline{K} = \overline{L}k^*, \overline{I} = (\mu + \eta)\overline{L}k^*, \quad \overline{Q} = \frac{(\mu + \eta)\overline{L}k^*}{s}, \\ &\overline{C} = \frac{(1 - s)(\mu + \eta)\overline{L}k^*}{s}. \end{aligned} \tag{2.44}$$

The constants $\overline{K}$, $\overline{I}$, $\overline{Q}$, $\overline{C}$ in exponential functions of the form (2.43) are often called in economics the *level variables*. In the case of constant labor $L(t) = \overline{L}$ (i.e., $\eta = 0$), the steady state is given by (2.44) and known as a *stationary point*.

Because the aggregate output Q, consumption C, investment I and capital K increase with the same rate η as the exogenous labor L, the Solow–Swan model is classified in the economic theory as the *exogenous growth model*.

2.3 Optimization Versions of Solow–Swan Model

Modern models of economic dynamics often include the dynamic optimization. Mathematically, such problems belong to the optimal control area. This section introduces *the dynamic optimization* into the Solow–Swan model and considers its two optimization versions on finite and infinite planning periods [3].

2.3.1 Optimization over Finite Horizon (Solow–Shell Model)

The Solow–Shell model is the Solow–Swan model (2.31)–(2.34) considered on a finite planning horizon $[0, T]$ in the case when the saving rate $s = I/Q$ depends on the time t and is endogenous [7]. To determine this rate, we consider the following *one-sector optimization* problem:

- Maximize the present value

$$\int_0^T \mathrm{e}^{-rt} c(t)\mathrm{d}t$$

of the *consumption per capita* $c = C/L$ over a given finite *horizon* $[0, T]$, subject to (2.31)–(2.35) and certain initial and terminal conditions.

In this problem, the given *discount rate* $r > 0$ reflects the planner's subjective rate of the decreasing utility of the output produced in more distant future. We still use the same aggregate variables of the Solow–Swan model: the output Q, consumption C, capital K, labor L, and investment I. For simplicity, let the labor $L(t)$ be constant, that is, $\eta = 0$ in (2.34). Switching the model (2.31)–(2.34) to the per capita variables $k = K/L$, $q = Q/L$, $c = C/L$, $i = I/L$, and excluding q, c, and i, the optimization problem under study becomes:

- Find the function $s(t)$, $0 \leq s(t) \leq 1$, and the corresponding $k(t)$, $k(t) \geq 0$, $t \in [0, T]$, which maximize

$$\max_{s,k} \int_0^T \mathrm{e}^{-rt}(1 - s(t))f(k(t))\mathrm{d}t \tag{2.45}$$

under the equality-constraint:

$$k'(t) = s(t)f(k(t)) - \mu k(t), \tag{2.46}$$

and the initial and terminal conditions:

$$k(0) = k_0, \quad k(T) \geq k_T. \tag{2.47}$$

The value of $k(T)$ cannot be arbitrary because the economy will continue after the end of the planning period. The terminal condition $k(T) \geq k_{\mathrm{T}}$ keeps a minimal acceptable level of capital at the end of the finite horizon.

The problem (2.45)–(2.47) is an optimal control problem, in which the function $s(t)$, $t \in [0, T]$, is unknown (rather than the scalar $s =$ const as in the static optimization of Sect. 2.2). In the optimal control terminology, the *independent* unknown function $s(.)$ is referred to as the *control variable* and the corresponding *dependent* unknown $k(.)$ is the *state variable*.

2.3.1.1 Steady-State Analysis

The results of the steady-state analysis of the Solow–Swan model (2.31)–(2.35) remain true for the Solow–Shell model because the fundamental equation (2.36) is the same. In particular, it means that the dynamic optimization problem (2.45)–(2.47) possesses the constant solution $s(t) = s^*$ and $k(t) = k^*$, if the initial condition is $k_0 = k^*$ and the terminal value is $k_T = k^*$ in (2.47). In the general case when $k_0 \neq k^*$ and/ or $k_T \neq k^*$, the solution of the model will have more complicated structure, which is the subject of the dynamic analysis that follows.

2.3.1.2 Dynamic Analysis

A dynamic analysis of such problems is more complex and requires sophisticated mathematical tools. The results provided below are obtained employing the *maximum principle* from Sect. 2.4.

Necessary Condition for an Extremum: If the function $s(t)$, $t \in [0, T]$, is a solution of the optimization problem (2.45)–(2.47), then:

(a) There exists a continuous function $\hat{\lambda}(t)$, $t \in [0, T]$, called the *dual* or *adjoint variable*, that satisfies the *dual equation*

$$\hat{\lambda}'(t) = (\mu + r)\hat{\lambda}(t) - \left[1 - s(t) + \hat{\lambda}(t)s(t)\right]f'(k(t)), \tag{2.48}$$

with the terminal *transversality condition*

$$[k(T) - k_T]e^{-rT}\hat{\lambda}(T) = 0, \tag{2.49}$$

where the corresponding state variable $k(t)$, $t \in [0, T]$, is found from (2.46).

(b) $s(t)$ maximizes $\left[1 - s(t) + \hat{\lambda}(t)s(t)\right]$ at each point $t \in [0, T]$.

The proof of this result follows from Corollary 2.1 of Sect. 2.4. Namely, the current-value Hamiltonian (2.69) for the optimal control problem (2.45)–(2.47) is constructed as

$$\hat{H}\left(s, k, \hat{\lambda}\right) = f(k)(1 - s) + \hat{\lambda}[sf(k) - \mu k], \tag{2.50}$$

and, then, the dual equation (2.48) is obtained from (2.70) as $\hat{\lambda}' = r\hat{\lambda} - \partial H/\partial k$, the state equation (2.46) fits $k' = \partial H/\partial \hat{\lambda}$, and $s(t)$ maximizes $H\left(s, k, \hat{\lambda}\right)$.

Extremum Condition for an Interior Solution. The maximum principle is constructed specifically to handle the case of boundary solutions: $s(t) = 0$ or $s(t) = 1$ in the domain $0 \leq s(t) \leq 1$ at some instants t. The possibility of boundary (or corner) solutions essentially complicates the optimal control dynamics. If a solution is known to be interior in the domain, then the optimality conditions become simpler. Namely, by Corollary 2.2 from Sect. 2.4, if $0 < s(t) < 1$, then the optimal $s(t)$ satisfies $\partial H/\partial s = 0$.

Let us utilize this optimality condition for the optimization problem (2.45)–(2.47). Taking the derivative of (2.50) in s, we obtain $\partial \hat{H}/\partial s = f(k)\left(\hat{\lambda} - 1\right)$. If a priori $0 < s(t) < 1$ for $t \in [0, T]$, then $\partial H/\partial s = 0$ and, therefore, $\hat{\lambda}(t) = 1$. Substituting $\hat{\lambda}$ to (2.48), we obtain

$$0 = \mu + r - f'(k(t)), \tag{2.51}$$

which is the same *golden rule of capital accumulation* (2.40) as obtained during static optimization in the Solow–Swan model of Sect. 2.2.

Structure of Solution: Using the extremum condition (2.48) and (2.49) and rewriting (2.50) as $\hat{H}\left(s, k, \hat{\lambda}\right) = s\left(\hat{\lambda} - 1\right) f(k) - \hat{\lambda}\mu k + f(k)$, we can show that $s(t) = 0$ maximizes $\hat{H}\left(s, k, \hat{\lambda}\right)$ at $\hat{\lambda}(t) < 1$ and $s(t) = 0$ maximizes $\hat{H}\left(s, k, \hat{\lambda}\right)$ at $\hat{\lambda}(t) > 1$. If $\hat{\lambda}(t) = 1$, then $\hat{H}\left(s, k, \hat{\lambda}\right)$ does not depend on s and the optimal k^* is found from (2.48), which is the same as the golden rule of capital accumulation (2.40). Thus, the solution $s(t)$, $t \in [0, T]$, of the optimization problem (2.45)–(2.47) is

$$s(t) = \begin{cases} 0 & \text{when} \quad \hat{\lambda}(t) < 1 \\ s^* & \text{when} \quad \hat{\lambda}(t) = 1 \, , \\ 1 & \text{when} \quad \hat{\lambda}(t) > 1 \end{cases} \tag{2.52}$$

where $0 < s^* < 1$ is the optimal (golden-rule) saving rate (2.41) in the Solow–Swan model. When $s(t) = s^*$, the corresponding trajectory is $k(t) = k^*$, where the unique k^* is found from (2.40).

2.3.1.3 Long-Term and Transition Dynamics

Because of the specifics of economic optimization problems, their dynamic analysis is usually split into two steps: the investigation of a *long-term* dynamics and the investigation of the *transition* dynamics. In many problems, the long-term dynamics is independent of initial conditions of the problem and coincides with the steady state solution of the model. Then, the transition dynamics describes how the optimal trajectory approaches the steady state.

The solution $s(t)$, $k(t)$, $t \in [0, T]$, of the optimization problem (2.45)–(2.47) in the case $k_0 < k^* < k_T$ is illustrated in Fig. 2.3.

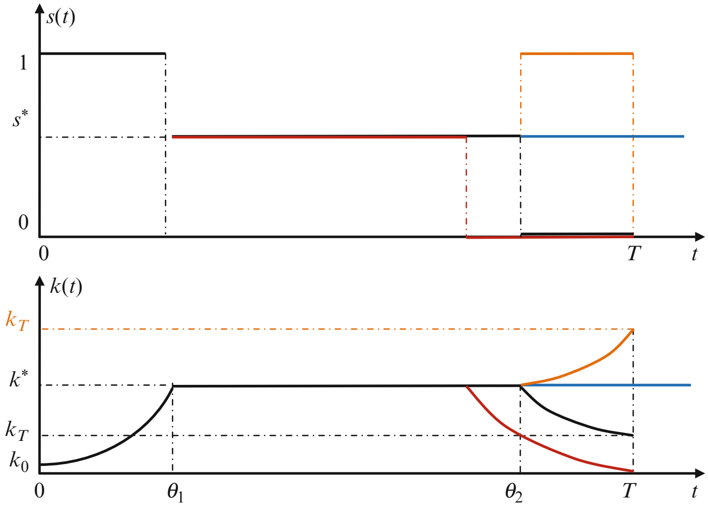


Fig. 2.3 Optimal trajectories in the Solow–Shell model at $k^* > k_0$ in the cases $k^* > k_\mathrm{T}$ (*black curves*), $k^* < k_\mathrm{T}$ (*black–orange curves*), and $k_\mathrm{T} = 0$ (i.e., no terminal condition, *black–red curves*). The lines $s \equiv s^*$ and $k \equiv k^*$ depict the golden rule trajectory. The *blue lines* represent the optimal regime in the infinite-horizon Solow–Ramsey model of Sect. 2.3.2

The transition (short-term) dynamics of the problem (2.45)–(2.47) is common for well-formulated economic problems. The optimal trajectory $s(t)$, $k(t)$ approaches the best steady state solution (s^*, k^*) on the initial interval $[0, \theta_1]$ and the transition dynamics ends at the instant θ_1 such that $k(\theta_1) = k^*$.

The optimal trajectory $s(t)$, $k(t)$ leaves the steady state solution (s^*, k^*) at some instant $\theta_2 < T$ near the right end of the planning horizon $[0, T]$. This behavior illustrates the so-called *end-of-horizon effect* and is also common in economic problems. Even if the terminal condition is absent, such effects still take place and even become more substantial. In particular, if $k_T = 0$, then there is no investments at the end $[\theta_2, T]$ of planning horizon.

Mathematically, this end-of-horizon effect appears because the optimal trajectory $k(t)$ must satisfy the transversality condition (2.49). This condition becomes less restrictive at $T = \infty$. Non-importance of the transversality condition for the infinite-horizon problem (2.31)–(2.35) was pointed out by *K*. Shell in [8]. It will be shown in the next section that the end-of-horizon effect is absent in the infinite-horizon problem.

The trajectory $s(t) \equiv s^*$, $k(t) \equiv k^*$ over $[\theta_1, \theta_2]$ represents the *long-term* dynamics of the optimization problem. The optimal saving rate $s(t)$ coincides with the constant golden-rule saving rate s^* in the Solow–Swan model on a certain interior part $[\theta_1, \theta_2]$ of the planning period $[0,T]$. The length of $[\theta_1, \theta_2]$ becomes larger when T increases. It means that *a turnpike property* holds for the optimization problem (2.45)–(2.47), where the *turnpike* trajectory is $s_T \equiv s^*$.

2.3.1.4 Turnpike Properties

The described structure of solutions to the Solow and Solow–Shell models provides a typical example of *turnpike properties*. A turnpike property states that the optimal trajectory on long planning horizons approaches in certain sense a *turnpike trajectory*, which is independent of the length of planning horizon and the initial state of an economic model. The turnpike trajectory usually has a simpler structure than the optimal trajectory; for example, the turnpike trajectory is simply constant $s_T(t) = s^*$ in the model (2.31)–(2.35). There are several categories of turnpike properties known as turnpike theorems in the *weak, normal, strong, and strongest forms*. A turnpike theorem in the normal form will appear during the optimization analysis of vintage capital models in Chap. 6.

The turnpike analysis is an important tool of the theory of linear multi-sector economic models, such as the Neumann–Gale model. Turnpike properties also appear in the nonlinear economic-mathematical models. The solution structure (2.52) demonstrates *the turnpike theorem in the strongest form*: the optimal trajectory $s(t)$ coincides with a unique turnpike trajectory $s_T(t) = s^*$, except for certain initial and final intervals of the fixed length.

Mathematically speaking, the turnpike properties reflect the stability and robustness of optimal regimes. The turnpike properties are not a universal feature of economic models. They appear only when a certain balance exists among various controls in models. It is often easier to find turnpikes and analyze their properties than to solve an optimization problem directly. In general, the turnpike theorems reflect some fundamental tendencies and laws of economic dynamics.

2.3.2 *Infinite-Horizon Optimization (Solow–Ramsey Model)*

The Solow–Shell model (2.31)–(2.35) with the optimization over the infinite planning horizon $[0,\infty)$ is known as the *Solow–Ramsey model with linear utility*. Namely, we consider the following optimization problem:

- A benevolent central planner determines the optimal saving rate $s(t)$, $t \in [0,\infty)$, to maximize the present value of the consumption per capita over the infinite planning horizon $[0,\infty)$:

$$\max_s \int_0^\infty \mathrm{e}^{-rt} c(t)\mathrm{d}t \tag{2.53}$$

subject to the state equation (2.46) with the initial condition $k(0) = k_0$. No terminal conditions are imposed at ∞.

2.3.2.1 Steady-State Analysis

The results of the steady-state analysis of the Solow–Ramsey model (2.53) remain the same as those in the Solow–Swan model (2.45)–(2.47) with the exogenous constant saving rate.

2.3.2.2 Dynamic Analysis

The dynamic analysis of the Solow–Ramsey model includes new mathematical challenges such as the convergence of the improper integral (2.53) along the optimal trajectory c. The condition for this convergence in our model (2.31)–(2.34), (2.53) is simply

$$r > \eta, \tag{2.54}$$

where η is the given growth rate of labor in (2.35). However, finding such conditions becomes more complicated in more advanced models (see for example Chap. 3).

Under (2.54), the extremum conditions remain the same, (2.48)–(2.52), as in the Solow–Shell model. Using (2.48) and (2.52), we can show that the solution $s(t)$, $t \in [0,\infty)$, of the problem (2.53) in the model (2.31)–(2.34) is

$$s(t) = \begin{cases} 1 & \text{at} \quad 0 \le t < \theta_1 \\ s^* & \text{at} \quad \theta_1 \le t < \infty \end{cases}, \tag{2.55}$$

$$k(t) = \begin{cases} k_{tr}(t) & \text{at} \quad 0 \le t < \theta_1 \\ k^* & \text{at} \quad \theta_1 \le t < \infty \end{cases}, \tag{2.56}$$

with the golden-rule saving rate s^* and capital per capita k^* in the Solow–Swan model. Also, it can be shown that the transversality condition (2.49) is reduced to the inequality (2.54). So the transversality condition is less important in the infinite-horizon problem in the sense that it does not directly affect the solution dynamics.

On the qualitative side, the behavior of the optimal trajectories appears to be simpler than in the finite-horizon Solow–Shell model (2.45)–(2.47). The solution $(s(t), k(t))$, $t \in [0, \infty)$, of the optimization problem (2.53) in the case $k_0 < k^*$ is illustrated in Fig. 2.3 by blue curves.

The transition dynamics of the problem (2.53) over the interval $[0, \theta_1]$ is the same as for the Solow–Shell model. The optimal trajectory $(s(t), k(t))$ approaches the steady state (s^*, k^*) and the transition dynamics ends at the instant θ_1 such that $k(\theta_1) = k^*$.

The long-term dynamics is $s(t) \equiv s^*$, $k(t) \equiv k^*$ over $[\theta_1, \infty)$, i.e., the optimal saving rate $s(t)$ coincides with the constant golden-rule saving rate s^* in the Solow–Swan model starting with the time θ_1. As shown in Fig. 2.3, the optimal trajectory $s(t)$, $k(t)$ does not leave the steady state (s^*, k^*) because the *end-of-horizon effects* are absent in infinite-horizon problems.

The considered optimization versions of the Solow–Swan, Solow–Shell, and Solow–Ramsey models, are classified by the economic theory as the *models of exogenous growth* because they cannot generate an endogenous growth when the labor $L(t)$ is constant. However, even small modifications of these models can lead to the *endogenous growth*. For instance, if we replace the neoclassic production function in the model equation (2.31) with the CES or AK production function, then the corresponding models are able to generate an endogenous growth. Models with endogenous growth are discussed in Sect. 3.4.

2.3.3 Central Planner, General Equilibrium, and Nonlinear Utility

The optimization problem (2.53), (2.31)–(2.34) describes the Solow–Ramsey model with linear utility in the *central planner setup*. The alternative economic environment is the *general equilibrium setup*. It assumes a decentralized economy with competitive firms and households with optimizing behavior, which interact on competitive markets. The households own capital assets, provide labor, receive wages, and choose their consumption over saving ratio to maximize their overall utility. The firms hire capital and labor and use them to produce goods to sell in order to maximize their profit. The perfect market *equilibrium* equalizes the supply and demand and determines the relative wages and prices of the capital and produced goods.

This *general equilibrium setup* has many modifications and simplifications. In particular, more players can be added, such as government, resource extraction firms, R&D firms, and others. On the other side, the separation of firms and households is not mandatory because households can perform the functions of firms. In the central planner setup, an economy is managed by a benevolent central planner that maximizes the utility of households. In many economic models, fundamental equations obtained in the central planner problem will be the same as in the general equilibrium framework. This textbook focuses on the productive side of the economy, so general equilibrium models are omitted. We refer the interested reader to Chaps. 1 and 2 of [1].

2.3.3.1 Utility Functions

Optimization problems in the central planner and general equilibrium frameworks often maximize a so-called *individual or social utility* that nonlinearly depends on consumption, rather than the direct amount of consumption. For example, the

Solow–Ramsey model with nonlinear utility maximizes the present value of the consumer's utility over $[0,\infty)$:

$$\max_{s} \int_{0}^{\infty} \mathrm{e}^{-rt} u(c(t))\mathrm{d}t \tag{2.57}$$

instead of (2.53). The nonlinear function $u(c)$ in (2.57) is called the *utility function* and describes the value of the future consumption product c for consumers. The function $u(c)$ is smooth, increasing and concave downward: $u'(c) > 0$, $u''(c) < 0$. Its concavity reflects the property of *diminishing marginal utility*: the product is more valuable for households when its amount is small. The utility function is said to satisfy the *Inada conditions* if $\lim_{c\to 0} u'(c) = \infty$ and $\lim_{c\to\infty} u'(c) = 0$.

Two most common utility functions are:

- The *isoelastic* (or *power*) utility function $u(c) = c^{1-\gamma}/(1-\gamma)$, where $0 < \gamma < 1$.
- The *logarithmic* utility function $u(c) = \ln c$.

2.4 Appendix: Maximum Principle

A maximum principle is the most popular type of extremum conditions for the optimal control of differential and integral equations. Below we provide the standard maximum principle for the optimal control of a scalar nonlinear ordinary differential equation, which is used for analyzing dynamic optimization models in Sect. 2.3 and other chapters.

Let us consider the following *optimal control* problem:

- Find the control function $u(t)\in\mathbf{R}^1$ and the corresponding $x(t)\in\mathbf{R}^1$, $t\in[0, T)$, which maximize

$$\max_{u,x} \int_{0}^{T} f(x(t), u(t), t)\mathrm{d}t, \tag{2.58}$$

subjected to the *state equation*

$$x'(t) = g(x(t), u(t), t), \tag{2.59}$$

the inequality-constraint

$$u_{\min}(t) \le u(t) \le u_{\max}(t), \tag{2.60}$$

and the initial and terminal conditions

$$x(0) = x_0, \quad x(T) \geq x_T. \tag{2.61}$$

The functions $f(x, u, t)$ and $g(x, u, t)$ are differentiable in x and u and continuous in t. The presence of the inequality-constraint (2.60) is a distinguished feature of the optimal control problems as opposed to the calculus of variations. The closed interval $U(t) = [u_{\min}(t), u_{\max}(t)] \subset \mathbf{R}^1$ is called the *control region* of the problem (2.58)–(2.61).

Definition. The *Hamiltonian* of the optimal control problem (2.58)–(2.61) is the function

$$H(x, u, \lambda, t) = f(x, u, t) + \lambda g(x, u, t), \tag{2.62}$$

where

λ is the *dual* (*costate*, *adjoint*) *variable*.

The dual variable reflects the change in the objective function due to changes in the constraints.

Then, the state equation (2.59) can be rewritten as

$$x'(t) = \frac{\partial H(x(t), u(t), \lambda, t)}{\partial \lambda}. \tag{2.63}$$

Statement 2.1 (*Pontryagin maximum principle*): If the function $u^*(t)$, $t \in [0, T]$, is a solution of the optimal control problem (2.58)–(2.61), then:

(a) The dual variable $\lambda(t)$, $t \in [0, T]$, exists and satisfies the *dual equation*

$$\lambda'(t) = -\frac{\partial H(x(t), u(t), \lambda(t), t)}{\partial x} \tag{2.64}$$

with the *transversality conditions* at the right end $t = T$:

$$\lambda(T) \geq 0, [x(T) - x_T]\lambda(T) = 0; \tag{2.65}$$

(b) The corresponding state variable $x(t)$, $t \in [0, T]$, is found from (2.59) for the given $u^*(t)$.

(c) For each $t \in [0, T]$, $u^*(t)$ maximizes $H(x, u, \lambda, t)$:

$$H(x(t), u(t), \lambda(t), t) = \max_{v \in U(t)} H(x(t), v(t), \lambda(t), t). \tag{2.66}$$

The proof of this statement is out of the scope of this textbook. It is available in textbooks on the optimal control and mathematical economics, e.g., [2, 4, 5, 6]. The maximum principle delivers only necessary condition for an extremum.

Necessary and sufficient condition for an extremum: If the $H(x,u,\lambda,t)$ is concave in u and x for each $t\in[0, T]$, then the conditions (2.64)–(2.66) are also sufficient for the function u to be a solution of the optimal control problem (2.58)–(2.61).

Extremum conditions of the form (2.64)–(2.66) are known as the *maximum principle* because they reduce an optimal control problem to maximization of the function H of one or several variables. There are numerous modifications of the Pontryagin maximum principle for various extensions of the problem (2.58)–(2.61). Below we discuss some of them that are applicable to specific models in the textbook.

2.4.1 Scalar Controls

The maximum principle is powerful when the control variable is a vector function. In our problem (2.58)–(2.61) with one scalar control u, the maximum condition (2.66) can be easily resolved and leads to the following structure of the optimal control:

$$u(t) = \begin{cases} u_{\min}(t) & \text{if } \dfrac{\partial H(x(t), u(t), \lambda(t), t)}{\partial u} < 0 \\ u_{\min}(t) < \tilde{u}(t) < u_{\max}(t) & \text{if } \dfrac{\partial H(x(t), u(t), \lambda(t), t)}{\partial u} = 0 \\ u_{\max}(t) & \text{if } \dfrac{\partial H(x(t), u(t), \lambda(t), t)}{\partial u} > 0 \end{cases}. \tag{2.67}$$

The formula (2.67) demonstrates that the optimal control $u(t)$ can be piecewise continuous at natural conditions. In more complicated problems, optimal controls possess many (even, indefinitely many) jumps and, therefore, are supposed to be measurable functions.

2.4.2 Discounted Optimization

In many economic and environmental models, optimization problems appear in a special form (2.58)–(2.61), where the state equation (2.59) is autonomous, the control region $[u_{\min}, u_{\max}]$ does not depend on t, and the time t explicitly appears only in the function f in (2.58) as the multiplier e^{-rt}:

$$g(x, u, t) = \hat{g}(x, u), \quad f(x, u, t) = e^{-rt}\hat{f}(x, u). \tag{2.68}$$

Then, the maximum principle can be simplified by introducing the so-called *current-value dual variable* $\hat{\lambda}(t) = \lambda(t)e^{rt}$ and the *current-value Hamiltonian* of the problem (2.58)–(2.61), (2.68)

$$\hat{H}\left(x, u, \hat{\lambda}\right) = \hat{f}(x, u, t) + \hat{\lambda}\, g(x, u). \tag{2.69}$$

Then, the dual variable λ in (2.62) is called the *present-value* dual variable.

Corollary 2.1 (*Current-value Maximum Principle*): If the optimal control problem (2.58)–(2.61) is of the form (2.68) and u is its solution, then:

(a) The dual variable $\hat{\lambda}(t)$, $t \in [0, T]$, exists and satisfies the *dual equation*

$$\hat{\lambda}' - r\hat{\lambda} = -\frac{\partial \hat{H}\left(x, u, \hat{\lambda}\right)}{\partial x}, \tag{2.70}$$

with the *transversality conditions* at the right end $t = T$

$$\hat{\lambda}(T) \geq 0, \quad [x(T) - x_T]e^{-rT}\hat{\lambda}(T) = 0; \tag{2.71}$$

(b) The corresponding $x(t)$, $t \in [0, T]$, is found from (2.59).

(c) For each $t \in [0, T]$, $u(t)$ maximizes $\hat{H}\left(x, u, \hat{\lambda}\right)$:

$$\hat{H}\left(x(t), u(t), \hat{\lambda}(t)\right) = \max_{v \in U} \hat{H}\left(x(t), v(t), \hat{\lambda}(t)\right). \tag{2.72}$$

In contrast to (2.64), the dual differential equation (2.70) is autonomous and much easier to solve. The maximization problem (2.72) is the same for all t.

2.4.3 Interior Controls

In the case of *interior controls*, the maximum principle leads to the following simpler optimality condition for *interior controls*.

Corollary 2.2. Let Statement 1.1 hold. If a priori $u_{\min}(t) < u(t) < u_{\max}(t)$ for $t \in [0, T]$, then the optimal $u(t)$ satisfies

$$\partial H(x(t), u(t), \lambda(t), t)/\partial u = 0, \quad t \in [0, T]. \tag{2.73}$$

In the case of the discounted problem (2.58)–(2.61), (2.68), the condition (2.73) is

$$\partial \hat{H}(x(t), u(t), \lambda(t))/\partial u = 0, \quad t \in [0, T]. \tag{2.74}$$

The condition (2.73) is simpler to deal with than (2.66) or (2.72). Corollary 2.2 is often used in the steady-state analysis of economic optimization problems, where steady-state trajectories are naturally interior in the domain of admissible controls. Economists often call the formulas (2.64), (2.65), (2.73) the *first-order optimality conditions*.

2.4.4 Transversality Conditions

The transversality conditions represent a necessary part of the optimality conditions, which is often overlooked. Their relevance depends on the specifics of the problem under study. We provide several transversality conditions for different versions of the optimal control problem (2.58)–(2.61).

Problems with a fixed right end: If the terminal condition in (2.61) is strengthened to the equality $x(T) = x_{\mathrm{T}}$, then the corresponding maximum principle does not involve the transversality condition (2.65) on $\lambda(T)$.

Problems with free right end: If the optimal control problem (2.58)–(2.61) does not have a terminal condition on $x(T)$ at all ("the right end of x is free"), then the transversality condition (2.65) becomes

$$\lambda(T) = 0. \tag{2.75}$$

Free terminal-time problems: If the right boundary point T in the optimal control problem (2.58)–(2.61) is not specified, then the corresponding transversality condition is

$$H(x(T), u(T), \lambda(T), T) = 0. \tag{2.76}$$

The infinite planning horizon $[0,\infty)$: If $T = \infty$ in the problem (2.58)–(2.61), then the corresponding transversality condition (2.65) becomes

$$\lim_{T\to\infty} \lambda(T) \geq 0, \ \lim_{T\to\infty} [x(T) - x_T]\lambda(T) = 0. \tag{2.77}$$

2.4.5 Maximum Principle and Dynamic Programming

An alternative approach for solving optimization problems is the *dynamic programming* method developed by R. Bellman. Although the maximum principle and dynamic programming are related, they have their essential differences, strengths and shortages. In a certain sense, the maximum principle is more practical for simple deterministic problems explored in this textbook. In particular, the maximum principle does not imply Bellman's equation while the dynamic programming conditions imply the maximum principle. In mathematical economics, the maximum principle is often considered to be a more powerful method for analytic solution [5].

Exercises

1. Fill in the table below:

Two-factor production function (PF)					
Properties	Linear PF	Leontief PF	Cobb–Douglas PF $\alpha > 0, \beta > 0$	Cobb–Douglas PF $\alpha + \beta = 1$	CES PF
Essentiality of inputs					
Positive returns					
Diminishing returns					
Homogeneity					
Returns to scale					
Marginal rate of substitution					
Total output elasticity					
Elasticity of substitution					
Inada conditions					
Neoclassical PF (yes, no)					

2. Prove that the total output elasticity of a homogeneous production function is equal to the degree of homogeneity: $\varepsilon(\mathbf{x}) = \gamma$.
3. Justify that the CES production function is more general as compared to the linear, Leontief, and Cobb–Douglas production functions.
4. Is the three-factor production function $F(K, L, N) = AK^{\alpha}L^{\beta}N^{\gamma}$, $\alpha > 0, \beta > 0, \gamma > 0$, homogeneous? If yes, then what is its degree of homogeneity? If no, then what condition should be implemented for the homogeneity?
5. The two-factor CES production function can be presented as $F(K,L) = A[\alpha K^{\rho} + (1 - \alpha)L^{\rho}]^{1/\rho}$. Is this CES function a neo-classical production function?
6. Prove that the first property *Essentiality of inputs* of the neoclassical production functions holds if the other three properties (*Positive returns*, *Diminishing returns*, and *Proportional returns to scale*) are valid.
7. Derive the formula (2.19): $\partial F/\partial K = f'(k)$, $\partial F/\partial L = f(k) - kf'(k)$, where $f(k) = F(k, 1)$, and $k = K/L$.

8. Find $\partial Q/\partial K$ and $\partial Q/\partial L$ for the CES production function (2.23) and prove that its marginal rate of substitution is $h = (1-\alpha)(1-b)^{\rho}\ k^{1-\rho}/(\alpha\ b^{\rho})$ and the elasticity of substitution is $\sigma = 1/(1-\rho)$.
9. Show that the variables $Q(t)$, $C(t)$, $I(t)$, and $K(t)$ of the Solow–Swan model (2.31)–(2.35) at the optimal steady state (s^*, k^*) are given by formulas (2.43) and (2.44).
10. Provide the steady-state analysis of the Solow–Shell model (2.45)–(2.47) and show that its golden rule of capital accumulations (2.51) and optimal steady state (s^*, k^*) are the same as in the Solow–Swan model (2.31)–(2.35).

References[1]

1. 🕮 Barro, R.J., Sala-i-Martin, X.: Economic Growth. The MIT Press, Cambridge, MA (2003)
2. 🕮 Caputo, M.R.: Foundations of Dynamic Economic Analysis: Optimal Control Theory and Applications. Cambridge University Press, New York (2005)
3. Cass, D.: Optimum growth in an aggregative model of capital accumulation. Rev. Econ. Stud. **32**, 233–240 (1965)
4. 🕮 Chiang, A.C., Wainwright, K.: Fundamental Methods of Mathematical Economics, 4th edn. McGraw-Hill, New York (2005)
5. 🕮 Intriligator, M.D.: Mathematical Optimization and Economic Theory. Society for Industrial and Applied Mathematics, Philadelphia (2002)
6. 🕮 Leonard, D., Van Long, N.: Optimal Control Theory and Static Optimization in Economics. Cambridge University Press, New York (1992)
7. Shell, K.: Optimal programs of capital accumulation for an economy in which there is exogenous technical change. In: Shell, K. (ed.) Essays on the Theory of Optimal Economic Growth, Chapter I, pp. 1–30. MIT Press, Cambridge (1967)
8. Shell, K.: Applications of Pontryagin's maximum principle to economics. In: Kuhn, H.W., Szegö, G.P. (eds.) Mathematical Systems Theory and Economics, pp. 241–292. Springer, Berlin (1969)
9. Solow, R.: A contribution to the theory of economic growth. Quart. J. Econ. **70**, 65–94 (1956)

[1] The book symbol 🕮 means that the reference is a textbook recommended for further student reading.

http://www.springer.com/978-1-4614-9310-5

Mathematical Modeling in Economics, Ecology and the Environment
Hritonenko, N.; Yatsenko, Y.
2013, XVII, 296 p. 34 illus., 8 illus. in color., Hardcover
ISBN: 978-1-4614-9310-5