Chapter One

Fermat: One Variable without Constraints

When a quantity is the greatest or the smallest, at that moment its flow is neither forward nor backward.

I. Newton

- How to find the maxima and minima of a function f of one variable x without constraints?

1.0 SUMMARY

You can never be too rich or too thin.

W. Simpson, wife of Edward VIII

One variable of optimization. The epigraph to this summary describes a view in upper-class circles in England at the beginning of the previous century. It is meant to surprise, going against the usual view that somewhere between too small and too large is the optimum, the "golden mean." Many pragmatic problems lead to the search for *the golden mean* (or *the optimal trade-off* or *the optimal compromise*). For example, suppose you want to play a computer game and your video card does not allow you to have optimal quality ("high resolution screen") as well as optimal performance ("flowing movements"); then you have to make an optimal compromise. This chapter considers many examples where this golden mean is sought. For example, we will be confronted with the problem that a certain type of vase with one long-stemmed rose in it is unstable if there is too little water in it, but as well if there is too much water in it. How much water will give optimal stability? Usually, the reason for such optimization problems is that a trade-off has to be made between two effects. For example, the height of houses in cities like New York or Hong Kong is determined as the result of the following trade-off. On the one hand, you need many people to share the high cost of the land on which the house is built. On the other hand, if you build the house very high, then the specialized costs are forbidding.

Derivative equal to zero. All searches for the "golden mean" can be modeled as problems of optimizing a function f of one variable x, minimization (maximization) if $f(x)$ represents some sort of cost (profit). The following method, due to Fermat, usually gives the correct answer: "put the derivative of f equal to zero," solve the equation, and – if the optimal x has to be an integer – round off to the nearest integer. This is well known from high school, but we try to take a fresh look at this method. For example, we raise the question why this method is so successful. The technical reason for this is of course the great strength of the available calculus for determining derivatives of given functions. We will see that in economic applications a conceptual reason for this success is the *equimarginal rule*. That is, rational decision makers take a marginal action only if the marginal benefit of the action exceeds the marginal cost; they will continue to take action till marginal benefit equals marginal cost.

Snellius's law. The most striking application of the method of Fermat is perhaps the derivation of the law of Snellius on the refraction of light on the boundary between two media—for example, water and air. This law was discovered empirically. The method of Fermat throws a striking light on this technical rule, showing that it is a consequence of the simple principle that light always takes the fastest path (at least for small distances).

1.1 INTRODUCTION

Optimization and the differential calculus. The first general method of solution of extremal problems is due to Pierre de Fermat (1608–1665). In 1638 he presented his idea in a letter to the prominent mathematicians Gilles Persone de Roberval (1602–1675) and Marin Mersenne (1588–1648). Scientific journals did not yet exist, and writing a letter to learned correspondents was a usual way to communicate a new discovery. Intuitively, the idea is that the tangent line at the highest or lowest point of a graph of a function is horizontal. Of course, this tangent line is only defined if the graph has no "kink" at this point.

The exact meaning became clear later when Isaac Newton (1642/43–1727) and Gottfried von Leibniz (1646–1716) invented the elements of classical analysis. One of the motivations for creating analysis was the desire of Newton and Leibniz to find general approaches to the solution of problems of maximum and minimum. This was reflected, in particular, in the title of the first published work devoted to the differential calculus (written by Leibniz, published in 1684). It begins with the words "Nova methodus pro maximis et minimis ...."

The Fermat theorem. In his letter to Roberval and Mersenne, Fermat had—from our modern point of view—the following proposition in mind, now called the (one-variable) Fermat theorem (but he could express his idea only for polynomials): if $\widehat{x}$ is a point of local minimum (or maximum) of f, then *the main linear part of the increment is equal to zero*. The following example illustrates how this idea works.

Example 1.1.1 *Verify the idea of Fermat for the function* $f(x) = x^2$.

Solution. To begin with, we note that the graph of f is a parabola that has its lowest point at $x = 0$.

Now let us see how the idea of Fermat leads to this point. Let $\widehat{x}$ be a point of local minimum of f. Let x be an arbitrary point close to $\widehat{x}$; write h for the increment of the argument, $x - \widehat{x}$, that is, $x = \widehat{x} + h$. The increment of the function,

$$f(x) - f(\widehat{x}) = (\widehat{x} + h)^2 - \widehat{x}^2 = 2\widehat{x}h + h^2,$$

is the sum of the *main linear part* $2\widehat{x}h$ and the *remainder* h^2.

This terminology is reasonable: the graph of the function $h \to 2\widehat{x}h$ is a straight line through the origin and the term h^2, which remains, is negligible—in comparison to h—for h small enough. That h^2 is negligible can be illustrated using decimal notation: if $h \approx 10^{-k}$

("k-th decimal behind point"), then $h^2 = 10^{-2k}$ ("$2k$-th decimal behind point"). For example, if $k = 2$, then $h = 1/100 = .01$, and then the remainder $h^2 = 1/10000 = .0001$ is negligible in comparison to h.

That the main linear part is equal to zero means that $2\widehat{x}h$ is zero for each h, and so $\widehat{x} = 0$.

Royal road. If you want a shortcut to the applications in this chapter, then you can read the statements of the Fermat theorem 1.4, the Weierstrass theorem 1.6, and its corollary 1.7, as well as the solutions of examples 1.3.4, 1.3.5, and 1.3.7 (solution 3); thus prepared, you are ready to enjoy as many of the applications in sections 1.4 and 1.6 as you like. After this, you can turn to the next chapter.

1.2 THE DERIVATIVE FOR ONE VARIABLE

1.2.1 What is the derivative?

To solve problems of interest, one has to combine the idea of Fermat with the differential calculus. We begin by recalling the basic notion of the differential calculus for functions of one variable. It is the notion of *derivative*. The following experiment illustrates the geometrical idea of the derivative.

"Zooming in" approach to the derivative. Choose a point on the graph of a function drawn on a computer screen. Zoom in a couple of times. Then the graph looks like a straight line. Its slope is the derivative. Note that a straight line through a given point is determined by its slope.

Analytical definition of the derivative. For the moment, we restrict our attention to the analytical definition, due to Auguste Cauchy (1789–1857). This is known from high school: it views the derivative $f'(\widehat{x})$ of a function f at a number $\widehat{x}$ as the limit of the quotient of the increments of the function, $f(\widehat{x} + h) - f(\widehat{x})$, and the argument, $h = (\widehat{x} + h) - \widehat{x}$, if h tends to zero, $h \to 0$ (Fig. 1.1).

Now we will give a more precise formulation of this analytical definition. To begin with, note that the derivative $f'(\widehat{x})$ depends only on the behavior of f at a neighborhood of $\widehat{x}$. This means that for a sufficiently small number $\varepsilon > 0$ it suffices to consider $f(x)$ only for x with $|x - \widehat{x}| < \varepsilon$.

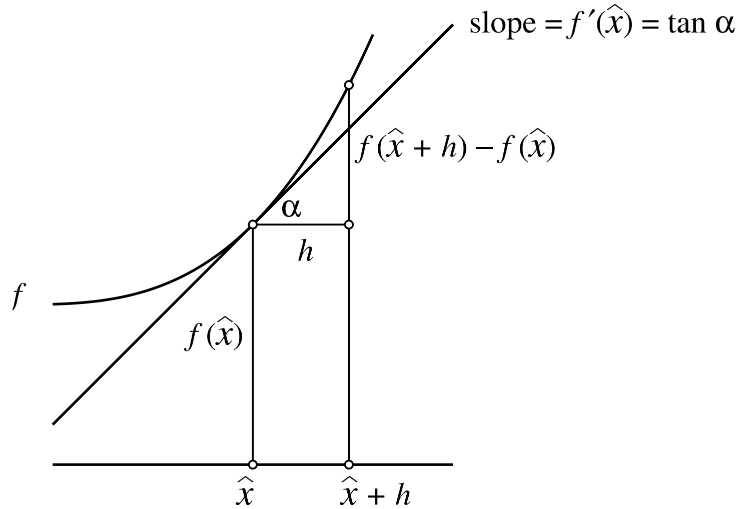

Figure 1.1 Illustrating the definition of differentiability.

We call the set of all numbers x for which

$$|x - \widehat{x}| < \varepsilon$$

the ε*-neighborhood of* $\widehat{x}$. This can also be defined by the inequalities

$$\widehat{x} - \varepsilon < x < \widehat{x} + \varepsilon,$$

or by the inclusion $x \in (\widehat{x}-\varepsilon, \widehat{x}+\varepsilon)$. The notation $f : (\widehat{x}-\varepsilon, \widehat{x}+\varepsilon) \to \mathbb{R}$ denotes that f is defined on the ε-neighborhood of $\widehat{x}$ and takes its values in $\mathbb{R}$, the real numbers.

Definition 1.1 ***Analytical definition of the derivative.*** *Let* $\widehat{x}$ *be a real number,* $\varepsilon > 0$ *a positive number, and* $f : (\widehat{x} - \varepsilon, \widehat{x} + \varepsilon) \to \mathbb{R}$ *a function of one variable* x*. The function* f *is called differentiable at* $\widehat{x}$ *if the limit*

$$\lim_{h \to 0} \frac{f(\widehat{x} + h) - f(\widehat{x})}{h}$$

exists. Then this limit is called the (first) derivative of f *at* $\widehat{x}$ *and it is denoted by* $f'(\widehat{x})$*.*

For a linear function $f(x) = ax$ one has $f'(x) = a$ for all $x \in \mathbb{R}$, that is, for all numbers x of the real line $\mathbb{R}$. The following example is more interesting.

Example 1.2.1 *Compute the derivative of the quadratic function $f(x) = x^2$ at a given number $\widehat{x}$.*

Solution. For each number h one has

$$f(\widehat{x} + h) - f(\widehat{x}) = (\widehat{x} + h)^2 - \widehat{x}^2 = 2\widehat{x}h + h^2$$

and so

$$\frac{f(\widehat{x} + h) - f(\widehat{x})}{h} = 2\widehat{x} + h.$$

Taking the limit $h \to 0$, we get $f'(\widehat{x}) = 2\widehat{x}$.

Analogously, one can show that for the function $f(x) = x^n$ the derivative is $f'(x) = nx^{n-1}$. It is time for a more challenging example.

Example 1.2.2 *Compute the derivative of the absolute value function $f(x) = |x|$.*

Solution. The answer is obvious if you look at the graph of f and observe that it has a kink at $x = 0$ (Fig. 1.2). To begin with, you see

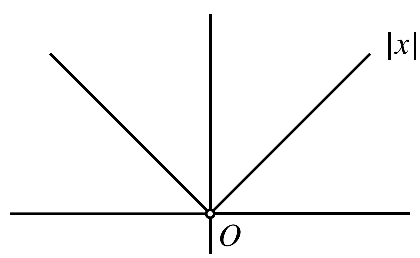

Figure 1.2 The absolute value function.

that f is not differentiable at $x = 0$. For $x > 0$ one has $f(x) = x$ and so $f'(x) = 1$, and for $x < 0$ one has $f(x) = -x$ and so $f'(x) = -1$. These results can be expressed by one formula,

$$|x|' = \frac{x}{|x|} = \mathrm{sgn}(x) \text{ for } x \neq 0.$$

Without seeing the graph, we could also compute the derivative using the definition. For each nonzero number h one has

$$\frac{f(\widehat{x}+h)-f(\widehat{x})}{h}=\frac{|\widehat{x}+h|-|\widehat{x}|}{h}.$$

- For $\widehat{x}<0$ this equals, provided $|h|$ is small enough,

$$\frac{(-\widehat{x}-h)-(-\widehat{x})}{h}=-1,$$

  and this leads to $f'(\widehat{x})=-1$.
- For $\widehat{x}>0$ a similar calculation gives $f'(\widehat{x})=1$.
- For $\widehat{x}=0$ this equals 1 for all $h>0$ and -1 for all $h<0$. This shows that $f'(0)$ does not exist.

The geometrical sense of the differentiability of f at $\widehat{x}$ is that the graph of f makes no "jump" at $\widehat{x}$ and has no "kink" at $\widehat{x}$. The following example illustrates this.

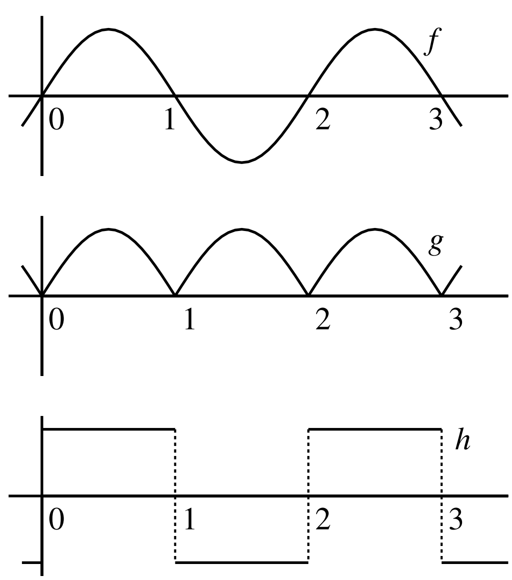

Figure 1.3 Illustrating differentiability.

Example 1.2.3 *In Figure 1.3, the graphs are drawn of three functions f, g, and h. Determine for each of these functions the points of differentiability.*

Solution. The functions g and h are not differentiable at all integers, as the graph of g has "kinks" at these points and the graph of h makes "jumps" at these points. The functions are differentiable at all other points. Finally, f is differentiable at all points.

Sometimes other notation is used for the derivative, such as $\frac{d}{dx}f(\widehat{x})$ or $\frac{df}{dx}(\widehat{x})$. One writes $D^1(\widehat{x})$ for the set of functions of one variable x for which the first derivative at $\widehat{x}$ exists.

1.2.2 The differential calculus

Newton and Leibniz computed derivatives by using the definition of the derivative. This required great skill, and such computations were beyond the capabilities of their contemporaries. Fortunately, now everyone can learn how to calculate derivatives. *The differential calculus* makes it possible to determine derivatives of functions in a routine way. This consists of a list of basic examples such as

$$(x^r)' = rx^{r-1},\ r \in \mathbb{R};\ (a^x)' = a^x \ln a,\ a > 0;$$

$$(\sin x)' = \cos x;\ (\cos x)' = -\sin x;\ (\ln x)' = 1/x;$$

and rules such as the *product rule*

$$(f(x)g(x))' = f'(x)g(x) + f(x)g'(x)$$

and the *chain rule*

$$(f(g(x))' = f'(g(x))g'(x).$$

In the $\frac{d}{dx}$ notation for the derivative this takes on a form that is easy to memorize,

$$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx},$$

where z is a function of y and y a function of x.

Modest role of definitions. Thus the *definition* of the derivative plays a very modest role: to compute a derivative, you use a user-friendly *calculus*, but not the definition. *Users might be pleasantly surprised to find that a similar observation can be made for most definitions.*

The use of the differential calculus will be illustrated in the analysis of all concrete optimization problems.

1.2.3 Geologists and sin 1°

If you want an immediate impression of the power of the differential calculus, then you can read about the following holiday experience. This example will provide insight, but the details will not be used in the solution of concrete optimization problems.

Once a colleague of ours met in the mountains a group of geologists. They got excited when they heard he was a mathematician, and what was even stranger is the sort of questions they started asking him. They needed the first four or five decimals of $\ln 2$ and of $\sin 1^\circ$. They even wanted him to explain how they could find these decimals by themselves. "Next time, we might need $\sin 2^\circ$ or some cosine and then we won't have you."

Well, not wanting to be unfriendly, our friend agreed to demonstrate this to them on a laptop. However, the best computing power that could be found was a calculator, and this did not even have a sine or logarithm button. Excuse enough not to be able to help, but for some reason he took pity on them and explained how to get these decimals even if you have no computing power. The method is illustrated below for $\sin 1^\circ$.

The geologists explained why they needed these decimals. They possessed some logarithmic lists about certain materials, but these were logarithms to base 2 and the geologists needed to convert these into natural logarithms; therefore, they had to multiply by the constant $\ln 2$. This explains $\ln 2$. They also told a long, enthusiastic story about $\sin 1^\circ$. It involved their wish to make an accurate map of the site of their research, and for this they had to measure angles and do all sorts of calculations.

Example 1.2.4 *How to compute the first five decimals of* $\sin 1^\circ$*?*

Solution. Often the following observation can be used to compare two functions f and g on an interval $[a, b]$ (cf. exercise 1.6.38).

If $f(a) = g(a)$ and $f'(x) \leq g'(x)$ for all x, then $f(x) \leq g(x)$ for all x.

Now we turn to $\sin 1^\circ = \sin \frac{\pi}{180}$, going over to radians.

- **First step.** To begin with, we establish that

$$0 \leq \sin x \leq x$$

for all $x \in [0, \frac{1}{2}\pi]$; that is, we "squash" $\sin x$ in between 0 and x, for all x of interest to the geologists. To this end, we apply the observation above. The three functions 0, $\sin x$, and x have the same value at $x = 0$, and the inequality between their derivatives,

$$0 \leq \cos x \leq 1,$$

holds clearly for all $x \in [0, \frac{1}{2}\pi]$.

- **Second step.** In the same way one can establish that

$$1 - \frac{1}{2}x^2 \leq \cos x \leq 1$$

for all $x \in [0, \frac{1}{2}\pi]$, using the inequality that we have just established. Indeed, the three functions $1 - \frac{1}{2}x^2$, $\cos x$, and 1 have the same value at $x = 0$, and the inequality between their derivatives,

$$-x \leq -\sin x \leq 0,$$

is essentially the inequality that we have just established. This is no coincidence: the expression $1 - \frac{1}{2}x^2$ is chosen in such a way that its derivative is $-x$ and that its value at $x = 0$ equals $\cos 0 = 1$.

- **Third step.** The next step gives

$$x - \frac{1}{6}x^3 \leq \sin x \leq x,$$

again for all $x \in [0, \frac{1}{2}\pi]$. We do not display the verification, but we note that we have chosen the expression $x - \frac{1}{6}x^3$ in such a way that

$$\frac{d}{dx}\left(x - \frac{1}{6}x^3\right) = 1 - \frac{1}{2}x^2$$

and $(x - \frac{1}{6}x^3)_{x=0} = \sin 0$.

- **Approximation is adequate.** Now let us pause to see whether we already have a sufficient grip on the sine to help the geologists. We have

$$\frac{\pi}{180} - \frac{1}{6}\left(\frac{\pi}{180}\right)^3 \leq \sin 1^\circ \leq \frac{\pi}{180}.$$

Is this good enough? A rough calculation on the back of an envelope, using $\pi \approx 3$, gives

$$\frac{1}{6}\left(\frac{\pi}{180}\right)^3 \approx \frac{1}{6}\left(\frac{1}{60}\right)^3 = \frac{1}{6}\frac{1}{216000} \approx 10^{-6}.$$

That is, we have squashed our constant $\sin 1^\circ$ in between two constants that we can compute and that differ by about 10^{-6}. This suffices; we have more than enough precision to get the required five decimals,

$$\sin 1^\circ \approx \frac{\pi}{180} \approx 0.01745,$$

provided we know the first few decimals of $\pi \approx 3.1416$.

The calculation of $\ln 2$ can be done in essentially the same way, but requires an additional idea; we postpone this calculation to exercise 1.6.37.

Historical comments. What we have just done amounts to rediscovering the *Taylor polynomials* or *Taylor approximations* and the *Taylor series*, the method of their computation, and their use—for the functions sine and cosine at $x = 0$. Taylor series were discovered in 1712 by Brook Taylor (1685–1731), inspired by a conversation in a coffeehouse. However, Taylor's formula was already known to Newton before 1712, and it gave him great pleasure that he could use it to calculate everything. In 1676 Newton wrote to Oldenburg about his feelings regarding his calculations:

"it is my shame to confess with what accuracy I calculated the elementary functions $\sin, \cos, \log$ *etc."*

In some of the exercises you will be invited to a further exploration of this method of approximating functions (exercises 1.6.35–1.6.39).

Behind the sine button. You can check the answer of the example above by pushing the sine button of your calculator. There is no need to know how the calculator computes the sine of angles like 1°. However, for the sake of curiosity, what is behind this button? Maybe tables of values of the sine are stored in the memory? It could also be that there is some device inside that can construct rectangular triangles with given angles and measure the sides? In reality, it is neither. What is behind the sine button (and some of the other buttons) is the method of the Taylor polynomials presented above.

Conclusion of section 1.2. The derivative is the basic notion of differential calculus. It can defined as the limit of the quotient of increments. However, to *calculate* it one should use lists of basic examples, and rules such as the product rule and the chain rule.

1.3 MAIN RESULT: FERMAT THEOREM FOR ONE VARIABLE

Let $\widehat{x}$ be a number and f a function of one variable, defined on an open interval containing $\widehat{x}$, say, the interval of all x with $|x-\widehat{x}| < \varepsilon$ for some positive number $\varepsilon > 0$. The function $f : (\widehat{x}-\varepsilon, \widehat{x}+\varepsilon) \to \mathbb{R}$ might also be defined outside this interval, but this is not relevant for the present purpose of formulating the Fermat theorem. It is sometimes convenient to consider minima and maxima simultaneously. Then we write extr (*extremum*). This sense of the word "extremum" did not exist until the nineteenth century, when it was invented by Paul David Gustav Du Bois-Reymond (1831–1889).

Definition 1.2 ***Main problem: unconstrained one-variable optimization.*** *The problem*

$$f(x) \to \text{extr} \qquad (P_{1.1})$$

is called a one-variable optimization problem without constraints.

The main task of problem $(P_{1.1})$ is to find the points of global extremum, but we will also find the points of local extremum, as a byproduct of the method. We will now be more precise about this. The concept of local extremum is useful, as the available method for finding the extrema makes use of differentiation and this is a *local* operation, using only the behavior near the solution. A minimum (resp. maximum) is from now on often called a *global* minimum (resp. maximum) to emphasize the contrast with a local minimum (resp. maximum). A point is called a point of local minimum (resp. maximum) if it is is a point of global minimum (resp. maximum) on a sufficiently small neighborhood. Now we display the formal definition of local extremum.

Definition 1.3 ***Local minimum and maximum.*** *Let $f : A \to \mathbb{R}$ be a function on some set $A \subseteq \mathbb{R}$. Then $\widehat{x} \in A$ is called a point of local minimum (resp. maximum) if there is a number $\varepsilon > 0$ such that $\widehat{x}$ is* *a global minimum (resp. maximum) of the restriction of $f : A \to \mathbb{R}$ to the set of $x \in A$ that lie in the ε-neighborhood of $\widehat{x}$, that is, for which $|x - \widehat{x}| < \varepsilon$ (Fig. 1.4).*

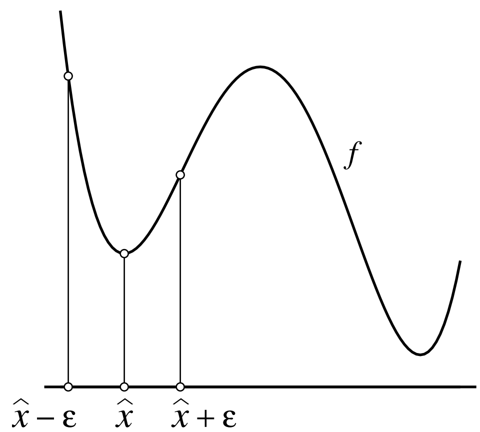

Figure 1.4 Illustrating the definition of local minimum.

An alternative term for global (local) minimum/maximum is *absolute* (*relative*) minimum/maximum.

Now we are ready to formulate the Fermat theorem.

Theorem 1.4 ***Fermat theorem—necessary condition for one-variable problems without constraints.*** *Consider a problem of type $(P_{1.1})$. Assume that the function $f : (\widehat{x} - \varepsilon, \widehat{x} + \varepsilon) \to \mathbb{R}$ is differentiable at $\widehat{x}$. If $\widehat{x}$ is a local extremum of $(P_{1.1})$, then*

$$f'(\widehat{x}) = 0. \tag{1.1}$$

The Fermat theorem follows readily from the definitions. We display the verification for the case of a minimum; then the case of a maximum will follow on replacing f by $-f$.

Proof. We have $f(\widehat{x} + h) - f(\widehat{x}) \geq 0$ for all numbers h with $|h|$ sufficiently small, by the local minimality of $\widehat{x}$.

- If $h > 0$, division by h gives $(f(\widehat{x} + h) - f(\widehat{x}))/h \geq 0$; letting h tend to zero gives $f'(\widehat{x}) \geq 0$.

- If $h < 0$, division by h gives $(f(\widehat{x}+h) - f(\widehat{x}))/h \le 0$ ("dividing a nonnegative number by a negative number gives a nonpositive number"); letting h tend to zero gives $f'(\widehat{x}) \le 0$.

Therefore, $f'(\widehat{x}) = 0$. □

Remark. This proof can be reformulated as follows. If the derivative of f at a point $\bar{x}$ is positive (resp. negative), then f is increasing (resp. decreasing) at $\bar{x}$; therefore, it cannot have a local extremum at $\bar{x}$. This formulation of the proof shows its algorithmic meaning. For example, if we want to minimize f, and have found $\bar{x}$ with $f'(\bar{x}) > 0$, then there exists $\bar{\bar{x}}$ slightly to the left of $\bar{x}$ that is "better than $\bar{x}$" in the sense that $f(\bar{\bar{x}}) < f(\bar{x})$.

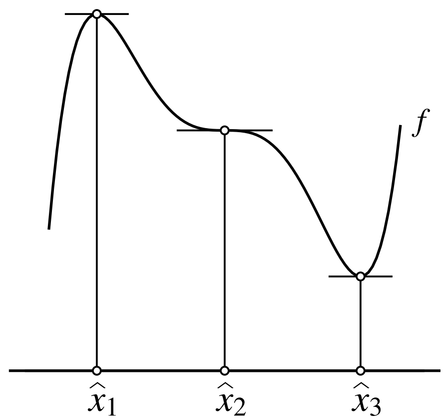

Figure 1.5 Illustrating the definition of stationary point.

Stationary points. Solutions of the equation $f'(x) = 0$ are called *stationary points* of the problem $(P_{1.1})$ (Fig. 1.5).

Trade-off. Pragmatic optimization problems always represent a trade-off between two opposite effects. Usually such a problem can be modeled as a problem of optimizing a function of one variable. We give a numerical example of such a trade-off, where one has to minimize the sum ("representing total cost") of two functions of a positive variable x ("representing the two effects"); one of the terms is an increasing and the other one a decreasing function of x. We will give an application of this type of problem to skyscrapers (exercise 1.6.14).

Example 1.3.1 *Find the solution* $\widehat{x}$ *of the following problem:*

$$f(x) = x + 5x^{-1} \to \min, \ x > 0.$$

Solution. By the Fermat theorem, $\widehat{x}$ is a solution of the equation $f'(x) = 0$. Computation of the derivative of f gives $f'(x) = 1 - 5x^{-2}$. The resulting equation $1 - 5x^{-2} = 0$ has a unique positive solution $x = \sqrt{5}$. Therefore, the choice $\widehat{x} = \sqrt{5}$ gives the optimal trade-off.

Remark. Unconstrained one-variable problems can be defined in the following more general and more elegant setting as follows. We need the concept of an open set.

A set V in $\mathbb{R}$, that is, a set V of real numbers, is called *open* if there exists for each number $v \in V$ a sufficiently small number $\varepsilon > 0$ such that the open interval $(v - \varepsilon, v + \varepsilon)$ is contained in V. For example, each open interval (a, b) is an open set. $V \in \mathcal{O}(\mathbb{R})$ means that V is an open set in $\mathbb{R}$; $V \in \mathcal{O}(\widehat{x}, \mathbb{R})$ means that V is an open set in $\mathbb{R}$ containing $\widehat{x}$, and such a set is called a *neighborhood* of $\widehat{x}$ in $\mathbb{R}$.

Let $V \in \mathcal{O}(\mathbb{R})$ and $f : V \to \mathbb{R}$. Then the problem

$$f(x) \to \min \qquad (P_{1.1})$$

is called a *problem without constraints*. A point $\widehat{x} \in V$ is called a point of *local minimum* if there exists $U \subset V$, $U \in \mathcal{O}(\widehat{x}, \mathbb{R})$ such that $\widehat{x}$ is a global minimum of the restriction of f to U.

1.3.1 Existence: the Weierstrass theorem

The Fermat theorem does not give a *criterion* for optimality. The following example illustrates this.

Example 1.3.2 *The function* $f(x) = x^3$ *has a stationary point. This is not a point of local extremum. Show this.*

Solution. Calculate the derivative $f'(x) = 3x^2$ and put it equal to zero, $3x^2 = 0$. This stationarity equation has a unique solution, $x = 0$. However, this is not a point of local extremum, as

$$f(x) = x^3 > 0 = f(0) \quad \text{for all } x > 0$$

and

$$f(x) = x^3 < 0 = f(0) \quad \text{for all } x < 0.$$

Therefore, we need a theorem to complement the Fermat theorem. Its formulation requires the concept of *continuity* of a function.

Definition 1.5 ***Continuous function.*** *A function $f : [a, b] \to \mathbb{R}$ on a closed interval $[a, b]$ is called continuous if*

$$\lim_{x \to \bar{x}} f(x) = f(\bar{x})$$

for all $\bar{x} \in [a, b]$ (that is, for $a \leq \bar{x} \leq b$).

The geometric sense of the concept of continuity is that the graph of the function $f : [a, b] \to \mathbb{R}$ makes no "jumps." That is, "it can be drawn without lifting pencil from paper."

How to verify continuity. To verify that a function is continuous on a closed interval $[a, b]$, one does not use the definition of continuity. Instead, this is done by using the fact that the basic functions such as

$$x^r, \ c^x, \ \sin x, \ \cos x, \ \ln x$$

are continuous if they are defined on $[a, b]$, and by using certain rules, for example, $f + g$, $f - g$ and fg are continuous on $[a, b]$ if f and g are continuous on $[a, b]$. In appendix C two additional rules are given.

We formulate the promised theorem for the case of minimization (of course it holds for maximization as well).

Theorem 1.6 ***Weierstrass theorem.*** *If a function $f : [a, b] \to \mathbb{R}$ on a closed interval is continuous, then the problem*

$$f(x) \to \min, \ a \leq x \leq b,$$

has a point of global minimum.

This fact is intuitively clear, but if you want to prove it, you are forced to delve into the foundations of the real numbers. For this we refer to appendix C.

In many applications, the variable does not run over a closed interval. Usually you can reduce the problem to the theorem above. We give an example of such a reduction. A function $f : \mathbb{R} \to \mathbb{R}$ on $\mathbb{R}$ is called *coercive (for minimization)* if

$$\lim_{|x| \to +\infty} f(x) = +\infty.$$

Corollary 1.7 *If a function $f : \mathbb{R} \to \mathbb{R}$ is continuous and coercive for minimization, then the problem $f(x) \to \min,\ x \in \mathbb{R}$, has a point of global minimum.*

Proof. We choose a number $M > 0$ large enough that $f(x) > f(0)$ for all x with $|x| > M$, as we may, by the coercivity of f. Then we can add to the given problem the constraints

$$-M \leq x \leq M$$

without changing the—possibly empty—solution set. In particular, this does not change the solvability status. It remains to note that the resulting problem,

$$f(x) \to \min,\ -M \leq x \leq M,$$

has a point of global minimum, by the Weierstrass theorem. □

Often you need variants of this result, as the following completion of the analysis of example 1.3.1 illustrates.

Example 1.3.3 *Show that the problem*

$$f(x) = x + 5x^{-1} \to \min,\ x > 0,$$

has a solution.

Solution. Observe that

$$\lim_{x \to +\infty} f(x) = +\infty \quad \text{and} \quad \lim_{x \downarrow 0} f(x) = +\infty.$$

Therefore, we can choose $M > 0$ large enough that

$$f(x) > f(1) \quad \text{if } x < 1/M \text{ or } x > M.$$

Then we can add to the given problem the constraints

$$1/M \leq x \leq M$$

without changing the—possibly empty—solution set. In particular, this does not change the solvability status. It remains to note that the resulting problem,

$$f(x) \to \min,\ 1/M \leq x \leq M,$$

has a point of global minimum, by the Weierstrass theorem.

Learning by doing. Such variants will be used without comment in the solution of concrete problems. Here, and in similar situations, we have kept the following words of Gauss in mind: Sed haec aliaque artificia practica, quae ex usu multo facilius quam ex praeceptis ediscuntur, hic tradere non est instituti nostri. (*But it is not our intention to treat of these details or of other practical artifices that can be learned more easily by usage than by precept.*)

Conclusion. Using only the Fermat theorem does not lead to certainty that an optimum has been found. The recommended method of completing the analysis of a problem is by using the Weierstrass theorem. This allows us to establish the existence of solutions of optimization problems.

1.3.2 Four-step method

In order to be brief and transparent, we will use a four-step method to write down the solutions of all problems in this book:

1. Model the problem and establish existence of global solutions.

2. Write down the equation(s) of the first order necessary conditions.

3. Investigate these equations.

4. Write down the conclusion.

What (not) to write down. We will use shorthand notation, without giving formal definitions for the notation, if the meaning is clear from the context. Moreover, we will always be very selective in the choice of what to display. For example, to establish existence of global solutions, one needs to verify the continuity of the objective function (cf. theorem 1.6). Usually, a quick visual inspection suffices to make sure that a given function is continuous. How to do this is explained above, and more fully in appendix C. We recommend, *not to write down* the verification. In particular, in solving each optimization problem in this book, we have checked the continuity of the objective function, but in the text there is no trace of this verification.

Now we illustrate the four-step method for example 1.3.3 (=example 1.3.1).

Example 1.3.4 *Solve the problem*

$$f(x) = x + 5x^{-1} \to \min, \ x > 0$$

by the four-step method.

Solution

1. $f(0+) = +\infty$ and $f(+\infty) = +\infty$, and so existence of a global solution $\widehat{x}$ follows, using the Weierstrass theorem.

2. Fermat: $f'(x) = 0 \Rightarrow 1 - 5x^{-2} = 0$.

3. $x = \sqrt{5}$.

4. $\widehat{x} = \sqrt{5}$.

Logic of the four-step method. The Fermat theorem (that is, steps 2 and 3 of the four-step method) leads to the following conclusion: *if* the problem has a solution, then it must be $x = \sqrt{5}$. However, it is not possible to conclude from this implication alone that $x = \sqrt{5}$ is the solution of the problem. Indeed, consider again example 1.3.2: the Fermat method leads to the conclusion that if the problem $f(x) = x^3 \to \min$ has a solution, then it must be $x = 0$. However, it is clear that this problem has no solution. This is the reason that the four-step method includes establishing the existence of a global solution (in step 1). By combining the two statements "there exists a solution" and "if a solution exists, then it must be $x = \sqrt{5}$," one gets the required statement: the unique solution is $x = \sqrt{5}$.

Why the four-step method? This problem could have been solved in one of the following two alternative ways. They are more popular, and at first sight they might seem simpler.

1. If we use the simple and convenient technique of the sign of the derivative, known from high school, then the Weierstrass theorem is not needed.

2. Another easy possibility is to use the second order test.

The great thing about the four-step method is that it is *universal*. It works for all optimization problems that can be solved analytically.

1. The sign of derivative test breaks down as soon as we consider optimization problems with two or more variables, as we will from chapter two onward.

2. The second order test can be extended to tests for other types of problems than unconstrained one-variable problems, but these tests are rather complicated. Moreover, these tests allow us to check *local* optimality only, whereas the real aim is to check

global optimality. The true significance of the second order tests is that they give a valuable insight, as we will make clear in section 1.5.1, and more fully in chapter five. We recommend not using them in problem solving.

The need for comparing candidate solutions. If step 3 of the four-step method leads to more than one candidate solution, then the solution(s) can be found by a comparison of values. The following example illustrates this.

Example 1.3.5 *Solve the problem*

$$f(x) = x/(x^2+1) \to \max$$

by the four-step method.

Solution

1. $\lim_{|x|\to+\infty} f(x) = 0$ and there exists a number $\bar{x}$ with $f(\bar{x}) > 0$, in fact $f(x) > 0$ for all $x > 0$, and so existence of a global solution $\widehat{x}$ follows, using the Weierstrass theorem (after restriction to the closed interval $[-N, N]$ for a sufficiently large N).

2. Fermat: $f'(x) = 0 \Rightarrow ((x^2+1) - x(2x))/(x^2+1)^2 = 0$.

3. $x^2 + 1 - 2x^2 = 0 \Rightarrow x^2 = 1 \Rightarrow x = \pm 1$. Compare values of f at the candidate solutions 1 and -1:

$$f(-1) = -\frac{1}{2} \text{ and } f(1) = \frac{1}{2}.$$

4. $\widehat{x} = 1$.

Variant of the coercivity result. In this example we see an illustration of one of the announced variants of corollary 1.7. It can be derived from the Weierstrass theorem in a similar way as corollary 1.7 has been derived from it.

Logic of the four-step method. The Fermat theorem (that is, steps 2 and 3 of the four-step method) leads to the following conclusion: the set of solutions of the problem is included in the set $\{1, -1\}$, to be called the set of candidate solutions. However, this statement does not exclude the possibility that the problem has no solution. By comparing the f-values of the candidates, one is led to the following more precise statement: if there exists a solution, then it must be $x = 1$. By combining this with the statement that a solution exists, which is established in step 1, one gets the required statement: there is a unique solution $x = 1$.

Four-step method is universal. We will see that the four-step method is the universal method for solving analytically all concrete optimization problems of all types. A more basic view on this method is sketched in the remark following the proof of the fundamental theorem of algebra (theorem 2.8).

1.3.3 Uniqueness; value of a problem

The following observation is useful in concrete applications. It can be derived from the Weierstrass theorem.

Corollary 1.8 ***Uniqueness.*** *Let f be a function of one variable, for which the second derivative $f'' = (f')'$ is defined on an interval $[a, b]$. If*

$$f'(a) < 0,\ f'(b) > 0,\ \text{and}\ f''(x) > 0,\ a < x < b,$$

then the problem

$$f(x) \to \min,\ a < x < b,$$

has a unique point of global minimum.

The following example illustrates how this observation can be used.

Example 1.3.6 *Solve the problem $f(x) = x^2 - 2\sin x \to \min$.*

Solution

1. The existence and uniqueness of a global solution $\widehat{x}$ follow from the remark above. Indeed,

$$f''(x) = 2(1 + \sin x) > 0$$

for all x except at points of the form $\frac{3}{2}\pi + 2k\pi,\ k \in \mathbb{Z}$, where it is 0. Moreover,

$$f'(x) = 2x - 2\cos x,$$

and so $f'(a) < 0$ and $f'(b) > 0$ for a sufficiently small and b sufficiently large.

2. Fermat: $f'(x) = 0 \Rightarrow 2(x - \cos x) = 0$.

3. The stationarity equation cannot be solved analytically.

4. There is a unique point of global minimum. It can be characterized as the unique solution of the equation $x = \cos x$.

Push the cosine button. Here is a method to find the solution of this problem numerically: if you repeatedly push on the cosine button of a calculator (having switched from degrees to radians to begin with), then the numbers on the display stabilize at 0.73908. This is the unique solution of the equation $x = \cos x$ up to five decimals.

Value problem. Finally, we mention briefly the concept of *value* of an optimization problem, which is useful in the analysis of some concrete problems. For a solvable optimization problem it can be defined as the value of the objective function at a point of global solution. The concept of value can be defined, more generally, for all optimization problems. The precise definition and the properties of the value of a problem are given in appendix B. We will write $S_{\min}$ (resp. $S_{\max}$) for the value of a minimization (resp. maximization) problem.

1.3.4 Illustrations of the Fermat theorem

What explains the success of the Fermat theorem? It is the combination with the power of the differential calculus, which makes it easy to calculate derivatives of functions. Thus, unconstrained optimization problems are reduced to the easier task of solving equations. We give two illustrations.

Example 1.3.7 *Consider the quadratic problem*

$$f(x) = ax^2 + 2bx + c \rightarrow \min$$

with $a \neq 0$. *If* $a > 0$ *it has a unique solution* $\widehat{x} = -b/a$ *and minimum value*

$$(ac - b^2)/a.$$

If $a < 0$ *it has no solution. Verify these statements.*

For those who know matrices and determinants: one can view the minimal value as the quotient of determinants of two symmetric matrices of size 2×2 and 1×1:

$$\begin{pmatrix} a & b \\ b & c \end{pmatrix} \text{ and } \begin{pmatrix} a \end{pmatrix}.$$

The resulting formula can be generalized to quadratic polynomials $f(x_1, \ldots, x_n)$ of n variables, giving the quotient of the determinants of two symmetric matrices of size $(n+1) \times (n+1)$ and $n \times n$, as we will see in theorem 9.4

Solution 1. One does not need the Fermat theorem to solve this problem: one can use one's knowledge of parabolas. It is well known that for $a > 0$ the graph of f is a *valley parabola*, which has a unique point of minimum

$$\widehat{x} = -b/a.$$

Therefore, its minimal value is

$$f(\widehat{x}) = (ac - b^2)/a.$$

For $a < 0$, the graph of f is a *mountain parabola*, and so there is no point of minimum.

Solution 2. The solution of the problem follows also immediately from the formula that you get for $a \neq 0$ by "completing the square,"

$$ax^2 + 2bx + c = a(x + b/a)^2 + (ac - b^2)/a.$$

Solution 3. However, it is our aim to illustrate the Fermat theorem.

1. To begin with,

   $$f(x) = x^2(a + 2b/x + c/x^2);$$

   the first factor of the right-hand side tends to $+\infty$ and the second one to a for $|x| \to +\infty$. If $a < 0$, it follows that the problem has value $-\infty$ and so it has no point of global minimum. If $a > 0$, it follows that f is coercive and so a solution $\widehat{x}$ exists.

2. Fermat: $f'(x) = 0 \Rightarrow 2(ax + b) = 0$.

3. $x = -b/a$.

4. If $a > 0$, then there is a unique minimum, $\widehat{x} = -b/a$; if $a < 0$, then $S_{\min} = -\infty$.

1.3.5 Reflection

The real fun. Four types of things have come up in our story: concrete problems worked out by some calculations, theorems, definitions, and proofs. Each of these has its own flavor and plays its own role. To be honest, the real fun is working out a concrete problem. However, we use a tool to solve the problem: the Fermat theorem. In the formulation of this result, the concepts "derivative" and "local extremum" occur.

Need for a formal approach. Therefore, we feel obliged to give precise definitions of these concepts and a precise proof of this result. Reading and digesting the proof might be a challenge (although it must be said that the proof of the Fermat theorem does not require any ideas; it is essentially a mechanical verification using the definitions). It might also require some effort to get used to the definitions.

Modest role of the formal approach. You might of course choose not to look very seriously at this definition and this proof. We have already emphasized the modest role of the *definition* of the derivative for practical purposes. In the same spirit, reading the proof gives you insight into the Fermat theorem, but this insight is not required for the solution of concrete problems.

These remarks apply to all worked-out problems, theorems, definitions, and proofs.

1.3.6 Comments: physical, geometrical, analytical, and approximation approach to the derivative and the Fermat theorem

Initially, there were three—equivalent—approaches to the notion of derivative and the Fermat theorem: *physical, geometrical*, and *analytical.* To these a fourth one was added later: *by means of approximation.* The analytical approach, due to Cauchy, is the one we have used to define the derivative.

Physical approach. The *physical* approach is due to Newton. Here is how he formulated the two basic problems that led to the differential and integral calculus.

" To clarify the art of analysis, one has to identify some types of problems. [. . .]

1. Let the covered distance be given; it is demanded to find the velocity of movement at a given moment.

2. Let the velocity of movement be given; it is demanded to find the length of the covered distance."

That is, for Newton the derivative is a *velocity*: the first problem asks to find the derivative of a given function. To be more precise, let a function $s(t)$ represent the position of an object that moves along a straight line and let the variable t represent time. Then the derivative $s'(t)$ is the *velocity* $v(t)$ at time t. We give a classical illustration.

Example 1.3.8 *Let us relate two facts that are well known from high school, but are often learned separately. Galileo Galilei (1564–1642), the astronomer who defended the heliocentric theory, is also known for his experiments with stones, which he dropped from the tower of Pisa. His experiments led to two formulas,*

$$s(t) = \frac{1}{2}gt^2$$

for the distance covered in the vertical direction after t seconds, and

$$v(t) = gt$$

for the velocity after t seconds. Here g is a constant ($g \approx 10$). Show that these two formulas are related.

Solution. This was one of the examples of Newton for illustrating his insight that velocity is always the derivative of the covered distance. This insight threw a new light on the formulas of Galileo:

the derivative of the expression $\frac{1}{2}gt^2$ is precisely gt.

Thus the formulas of Galileo are seen to be related.

Newton formulated the Fermat theorem in terms of velocity by the following phrase, which we used as the epigraph to this chapter:

"when a quantity is the greatest or the smallest, at that moment its flow is neither forward nor backward."

Geometrical approach. The *geometrical* approach is due to Leibniz and well known from high school. He interpreted the derivative as the *slope of the tangent line* (Fig. 1.1). This is $\tan\alpha$, the *tangent* of the angle α between the tangent line and the horizontal axis. Then the Fermat theorem means that the tangent line at a point of local extremum is horizontal, as we have already mentioned.

Weierstrass. The *approximation* approach was given much later, by Karl Weierstrass (1815–1897). He was one of the leaders in rigor in analysis and is known as the "father of modern analysis." In addition he is considered to be one of the greatest mathematics teachers of all time. He began his career as a teacher of mathematics at a secondary school. This involved many obligations: he was also required to teach physics, botany, geography, history, German, calligraphy, and even gymnastics. During a part of his life he led a triple life, teaching during the day, socializing at the local beer hall during the evening, and doing mathematical research at night. One of the examples of his rigorous approach was the discovery of a function that, although continuous, has no derivative at any point. This counter-intuitive function caused dismay among analysts, who depended heavily on their intuition for their discoveries.

The reputation of his university teaching meant that his classes grew to 250 pupils from all around the world. His favorite student was Sofia Kovalevskaya (1850–1891), whom he taught privately since women were not allowed admission to the university. He was a confirmed bachelor and kept his thoughts and feelings always to himself. To this rule, Sofia was the only exception. Here is a quotation from a letter to her from their lifelong correspondence:

"dreamed and been enraptured of so many riddles that remain for us to solve, on finite and infinite spaces, on the stability of the world system, and on all the other major problems of the mathematics and the physics of the future [...] you have been close [...] throughout my life [...] and never have I found anyone who could bring me such understanding of the highest aims of science and such joyful accord with my intentions and basic principles as you."

Approximation approach. Another fruit of the work of Weierstrass on the foundations of analysis is his *approximation* approach to the derivative. One can split the increment $f(\widehat{x}+h)-f(\widehat{x})$ in a unique way as a sum of a term that depends on h in a linear way, called the *main linear part*, and a term that is negligible in comparison to h if h tends to zero, called the *remainder*. Then the main linear part equals the product of a constant and h. This constant is defined to be the derivative $f'(\widehat{x})$.

This is *the universal approach*, as we will see later (definition 2.2 and section 12.4.1), whenever we have to extend the notion of derivative. Moreover, it is the most fruitful approach to the derivative for economic applications. We will give the formal definition of the approximation definition of the derivative. If r is a function of one variable, then we write $r(h) = o(h)$ ("small Landau o") if

$$\lim_{h\to 0} \frac{|r(h)|}{|h|} = 0.$$

For example, $r(h) = h^2$ is $o(h)$, and we write $h^2 = o(h)$. The sense of this concept is that $|r(h)|$ is negligible in comparison to $|h|$, for $|h|$

sufficiently small—to be more precise, $|r(h)|/|h|$ is arbitrary small for $|h|$ small enough.

Definition 1.9 ***Approximation definition of the derivative.*** *Let $\widehat{x}$ be a real number and f a function of one variable x. The function f is called differentiable at $\widehat{x}$ if it is defined, for some $\varepsilon > 0$, on the ε-neighborhood of $\widehat{x}$, that is,*

$$f : (\widehat{x} - \varepsilon, \widehat{x} + \varepsilon) \to \mathbb{R},$$

and if, moreover, there exist a number $a \in \mathbb{R}$ and a function $r : (-\varepsilon, +\varepsilon) \to \mathbb{R}$, for which $r(h) = o(h)$, such that

$$f(\widehat{x} + h) - f(\widehat{x}) = ah + r(h)$$

for all $h \in (-\varepsilon, +\varepsilon)$. Then a is called the (first) derivative of f at $\widehat{x}$ and it is denoted by $f'(\widehat{x})$.

We illustrate this definition for example 1.1.1.

Example 1.3.9 *Compute the derivative of the quadratic function $f(x) = x^2$ at a given number $\widehat{x}$, using the approximation definition.*

Solution. For each number h, one has

$$f(\widehat{x} + h) - f(\widehat{x}) = (\widehat{x} + h)^2 - \widehat{x}^2 = 2\widehat{x}h + h^2,$$

so this can be split up as the sum of a linear term $2\widehat{x}h$ and a small Landau o term h^2. The coefficient of $2\widehat{x}h$ is $2\widehat{x}$; therefore,

$$f'(\widehat{x}) = 2\widehat{x}.$$

As a further illustration of the approximation definition, we use it to write out the proof of the Fermat theorem 1.4 (again for the case of minimization):

Proof. $0 \leq f(\widehat{x} + h) - f(\widehat{x}) = f'(\widehat{x})h + o(h)$ and so, dividing by h and letting h tend to zero from the right (resp. left) gives $0 \leq f'(\widehat{x})$ (resp. $0 \geq f'(\widehat{x})$). Therefore, $f'(\widehat{x}) = 0$. □

Conclusion. There are four—equivalent—approaches to the notion of derivative and the Fermat theorem, each shedding additional light on these central issues.

1.3.7 Economic approach to the Fermat theorem: equimarginal rule

Let us formulate the economic approach to the Fermat theorem. A rational decision maker takes a marginal action only if the marginal benefit of the action exceeds the marginal cost. Therefore,

in the optimal situation the marginal benefit and the marginal cost are equal.

This is called the *equimarginal rule*. To see the connection with the Fermat theorem, note that if we apply the Fermat theorem to the maximization of profit, that is, of the difference between benefit and cost, then we get the *equimarginal* rule ("marginal" is another word for derivative).

The tradition in economics of carrying out a *marginal analysis* was initiated by Alfred Marshall (1842–1924). He was the leading figure in British economics (itself dominant in world economics) from 1890 until his death in 1924. His mastery of mathematical methods never made him lose sight that the name of the game in economics is to understand economic phenomena, whereas in mathematics the depth of the result is the crucial point.

Conclusion section. All optimization problems that can be solved analytically can be solved by one and the same four-step method. For problems of one variable without constraints, this method is based on a combination of the Fermat theorem with the Weierstrass theorem.

1.4 APPLICATIONS TO CONCRETE PROBLEMS

1.4.1 Original illustration of Fermat

Fermat illustrated his method by solving the geometrical optimization problem on the largest area of a right triangle with given sum a of the two sides that make a right angle.

Problem 1.4.1 *Solve the problem of Fermat if the given sum a is 10.*

Solution. The problem can be modeled as follows

$$f(x) = \frac{1}{2}x(10 - x) = -\frac{1}{2}x^2 + 5x \rightarrow \max, \ 0 < x < 10.$$

This is a special case of example 1.3.7, if we rewrite our maximization problem as a minimization problem by replacing $f(x)$ by $-f(x)$. It follows that it has a unique solution, $\widehat{x} = 5$. That is, the right triangle for which the sides that make a right angle both have length 5 is the unique solution of this problem.

(continued)