Chapter 2
Functions Optimization

Natural and exact sciences are based on functional relationships between different variables. If these functional relationships meet certain criteria, they can be studied by using procedures developed and systematized under the framework of ordinary differential and integral calculus. In essence, these procedures allow emphasizing properties common to many types of functions such as continuity, monotony or differentiability.

This chapter reminds a number of concepts and methods of differential calculus which are important by themselves, being, however, applicable both in variational calculus and in optimal control theory (Forray 1975). Only the case of functions of real variables is considered here.

2.1 Weierstrass Theorem

A theoretical result often used in practice is the Weierstrass theorem. This theorem states that if a function $f(x)$ of a single real variable is definite and continuous in every point on a closed finite interval $a \leq x \leq b$, then, on that interval $f(x)$ reaches its *absolute maximum* and *absolute minimum* values.

If an absolute extreme value is reached in an internal point of the interval, it is also a *relative* (or *local*) extreme value.

If a function is continuous on the closed interval $[a, b]$ and differentiable on the open interval (a, b), except for a finite number of points (which may be zero), a simple method for determining the absolute extremes is as follows. The point where the absolute extreme is reached can be one of the following:

1. A point where $f'(x) = 0$. A point where the first derivative is canceled is called *stationary point* or *critical point*. The values of the function $f(x)$ in such points are called *stationary values*.
2. A point at one end of the interval.

V. Badescu, *Optimal Control in Thermal Engineering*, Studies in Systems, Decision and Control 93, DOI 10.1007/978-3-319-52968-4_2

3. A point where the function $f(x)$ is not differentiable.

If a function has a continuous first derivative, the absolute minimum and maximum are found by comparing the points that satisfy the conditions 1–3. As a particular case, if the first derivative of the function $f(x)$ does not cancel on the open interval (a, b), then the extreme values are reached at the ends of the interval.

2.2 Conditions of Extreme

2.2.1 Real Functions of One Variable

If $y = f(x)$ is a function of real variable with a continuous first derivative $f'(x)$ on (a, b), then a *necessary condition* for the existence of a relative extreme in a point x_0, $a < x_0 < b$ is $f'(x_0) = 0$. This is not, however, a *sufficient condition* for an extreme in x_0.

A sufficient condition for the function $f(x)$ to have a relative extreme in point x_0 (i.e. to have maximum or minimum values in the neighboring of x_0) is that, apart from $f'(x_0) = 0$, the second derivative of the function, $f''(x)$, does not cancel in x_0. Thus, if $f''(x_0) < 0$, the function has a relative maximum in x_0 and if $f''(x_0) > 0$ the function has a relative minimum.

If $f''(x_0) = 0$, in deciding whether there is a relative extreme in x_0, one has to study the sign of the successive derivatives (i.e. derivatives of order three, four, etc.) in x_0, so:

1. if the first nonzero derivative is of odd order, then the function does not have an extreme in x_0.
2. if the first nonzero derivative is of even order, then the function has in x_0 a relative extreme, and this extreme is:

- a maximum, if the derivative sign is negative, or
- a minimum, if the derivative is positive.

Example
Find the absolute maximum and minimum of the function $f(x) = x^4 - x^5$ for $-2 \leq x \leq 2$.

Solution

$$f'(x) = 4x^3 - 5x^4$$

To find the *stationary* values, $f'(x) = 0$; therefore $x^3(4 - 5x) = 0$. Stationary values are $x = 0$ (triple roots) and $x = 4/5$.

$f''(x) = 12x^2 - 20x^3$. Therefore, $f''(0) = 0$ and $f''(4/5) < 0$

For $x = 4/5$, the function has a local maximum, which is $f(4/5) = 4^4/5^5$. For $x = 0$, since $f''(0) = 0$, higher order derivatives should be evaluated:

$$f^{(3)}(x) = 24x - 60x^2; \quad f^{(3)}(0) = 0$$

$$f^{(4)}(x) = 24 - 120x; \quad f^{(4)}(0) = 24$$

Since the first non-zero derivative is of even order, and it is positive, the function has a relative minimum in $x = 0$, namely $f(0) = 0$.

Are the maximum $f(4/5)$ and the minimum $f(0)$ absolute or relative? The values of $f(x)$ at the extremities of the interval are compared. Then, $f(2) = -16$ and $f(-2) = 48$. Therefore, in $x = -2$ the function has an absolute maximum and in $x = 2$ the function has an absolute minimum.

2.2.2 Functions of Several Variables

Finding the extreme values of functions of several variables is more complicated. The Weierstrass theorem can be applied in this situation, too; an extended version of this theorem says that a continuous function in a closed domain D of the variables reaches a maximum value and a minimum value within the domain or on the boundary of the domain.

Consider the case of a differentiable function $f(x_1, x_2, \ldots, x_n)$ in a domain D. Then, the necessary condition for an extreme in a point P within the domain is:

$$\frac{\partial f}{\partial x_1} = \frac{\partial f}{\partial x_2} = \cdots = \frac{\partial f}{\partial x_n} = 0 \tag{2.2.1}$$

in P. Consequently, the differential of the function f in point P is canceled: $df \equiv \sum_{i=1}^{n} f_{x_i} dx_i = 0$. Here the common notation $f_{x_i} \equiv \partial f / \partial x_i$ has been used.

Sufficiency conditions for the existence of the extreme are more complicated. They are treated in the following.

2.2.2.1 Functions of Two Variables

First, the case of a function of two variables will be considered. Such a function has a *maximum* in a point P if:

$$f_x = f_y = 0 \quad , \quad f_{xx} < 0 \quad \text{and} \quad f_{xx} f_{yy} - f_{xy}^2 > 0 \tag{2.2.2}$$

The function has a *minimum* if:

$$f_x = f_y = 0 \quad , \quad f_{xx} > 0 \quad \text{and} \quad f_{xx}f_{yy} - f_{xy}^2 > 0 \tag{2.2.3}$$

If $f_{xx}f_{yy} - f_{xy}^2 < 0$, the function does not have a maximum or a minimum. If $f_{xx}f_{yy} - f_{xy}^2 = 0$, another method has to be used to find the extreme (if any).

The previous conditions have a simple geometric interpretation. The necessary conditions for a stationary value ($f_x = f_y = 0$ in point (x_0, y_0)) assume that the tangent plane to the surface $z = f(x, y)$ in that point is horizontal (i.e. parallel to the plane Oxy). If the point is an extreme point (either maximum or minimum), then in its proximity the tangent plane does not intersect the surface. In case of a saddle point (in which, although the first derivative is canceled, there is no minimum or maximum) the plane cuts the surface after a curve which has several branches in that point.

Example
Find the size of a paralellepipedic open box of volume 4 dm^3, whose surface area is minimum.

Solution
Denote by x, y, z the length, width and height of the box, respectively, all of them being positive. The surface area is:

$$A = xy + 2xz + 2yz$$

The volume is $xyz = 4$ so that $z = 4/(xy)$. Therefore:

$$A = xy + \frac{8}{y} + \frac{8}{x}$$

To obtain the stationary values:

$$A_x = y - \frac{8}{x^2} = 0, \quad A_y = x - \frac{8}{y^2} = 0$$

One finds $x = y = 2$ and $A = 12$. Then, $z = 1$.

The derivatives of second order in the point $(2, 2)$ are:

$$A_{xx} = \frac{16}{x^3} = 2; \quad A_{yy} = \frac{16}{y^3} = 2; \quad A_{xy} = 1$$

so that $A_{xx}A_{yy} - A_{xy}^2 = 4 - 1 > 0$. Since $A_{xx} > 0$, the function has a relative minimum in the point $(2, 2)$. It can be shown that it is an absolute minimum.

2.2.2.2 Functions with Arbitrary Finite Number of Variables

The *sufficient condition* for the extreme of a function depending on many variables can be conveniently expressed by using the matrix of the attached *quadratic form*, as shown below. A quadratic form is a homogeneous polynomial of degree two in many variables. Such a form is represented as follows:

$$\begin{gathered} F(x_1, x_2, \ldots, x_n) = a_{11}x_1^2 + a_{12}x_1x_2 + \cdots + a_{1n}x_1x_n + \\ a_{21}x_2x_1 + a_{22}x_2^2 + \cdots + a_{2n}x_2x_n + \\ \cdots \\ a_{n1}x_nx_1 + a_{n2}x_nx_2 + \cdots + a_{nn}x_n^2 \end{gathered} \tag{2.2.4}$$

where the coefficients $a_{ij}(i,j = 1, 2, \ldots, n)$ are arbitrary real numbers. A quadratic form can always be arranged so that $a_{ji} = a_{ij}$. Therefore, the *matrix of the quadratic form* is symmetrical and has the form:

$$A = \begin{vmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ & & \ldots & \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{vmatrix} \tag{2.2.5}$$

Consequently, each quadratic form is associated with a single *symmetric matrix*, and vice versa.

A quadratic form can be written in abbreviated matrix notation as follows:

$$F(x_1, x_2, \ldots, x_n) = x' \cdot A \cdot x \tag{2.2.6}$$

where $x' = |\, x_1 \quad x_2 \quad \ldots \quad x_n \,|$ is a row matrix, the transposed of the column matrix

$$x = \begin{vmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{vmatrix} \tag{2.2.7}$$

A real quadratic form is *positive definite* if, for real values of the variables, it always has a positive value, except when $x_1 = x_2 = \cdots = x_n = 0$. This allows to extend the term "positive definite" for the case of symmetric matrices. Thus, a real symmetric matrix $[A] = [a_{ij}]$ is positive definite if the attached quadratic form

$$F(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}x_ix_j \tag{2.2.8}$$

is positive definite. For example, the unit matrix is positive definite since the attached quadratic form is positive definite.

A quadratic form F is negative definite if $-F$ is positive definite. Similarly, a matrix $[A]$ is negative definite if the matrix $[-A]$ is positive definite. A quadratic form is indefinite if it is neither positive nor negative definite.

The next theorem due to James Joseph Sylvester is useful for defining the necessary and sufficient conditions for the extreme of functions of n variables: A quadratic form $\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j$ is positive definite if and only if all the principal minors of the attached matrix are positive:

$$\Delta_1 \equiv |a_{11}| > 0, \Delta_2 \equiv \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0, \cdots \Delta_n \equiv \begin{vmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ & \ldots & & \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{vmatrix} > 0 \tag{2.2.9}$$

Consider the function $f(x_1, x_2, \ldots, x_n)$ with third-order derivatives continue in the neighboring of the stationary point P specified by the coordinates $x_i = x_i^0 (i = 1, 2, \ldots, n)$, where $f_{x_1} = f_{x_2} = \cdots = f_{x_n} = 0$. For brevity one says that the point P has the coordinate x^0. The total second order differential in x^0 is:

$$d^2 f^0 = \sum_{i=1}^{n}\sum_{j=1}^{n} f^0_{x_ix_j} dx_i dx_j \tag{2.2.10}$$

This differential is a quadratic form in the variables $dx_1, dx_2, \ldots, dx_n$ and, therefore, can be: (i) positive definite, (ii) negative definite or (iii) indefinite. Using Taylor's theorem one can show that in case (i) the function f has a minimum, in case (ii) the function f has a maximum and in case (iii) the function f has no maximum nor minimum, but only if, in addition:

$$D \equiv \begin{vmatrix} f^0_{x_1x_1} & f^0_{x_1x_2} & \ldots & f^0_{x_1x_n} \\ & \ldots & & \\ f^0_{x_nx_1} & f^0_{x_nx_2} & \ldots & f^0_{x_nx_n} \end{vmatrix} \neq 0 \tag{2.2.11}$$

If $D = 0$, another method has to be used for specifying the type of the extreme.

2.2.2.3 Examples

(1) Show that the quadratic form:

$$F = x_1^2 + 2x_2^2 + 5x_3^2 - 2x_1x_2 + 4x_1x_3 - 4x_2x_3$$

is positive definite.

Solution
The determinant of the form is

$$|a_{ij}| = \begin{vmatrix} 1 & -1 & 2 \\ -1 & 2 & -2 \\ 2 & -2 & 5 \end{vmatrix}$$

and

$$\Delta_1 = |1| = 1; \quad \Delta_2 = \begin{vmatrix} 1 & -1 \\ -1 & 2 \end{vmatrix} = 1; \quad \Delta_3 = \begin{vmatrix} 1 & -1 & 2 \\ -1 & 2 & -2 \\ 2 & -2 & 5 \end{vmatrix} = 1$$

Since all the principal minors are positive, according with Sylvester theorem F is positive definite.

(2) Show that the following quadratic form is not definite:

$$F = x_1^2 + 2x_2^2 + 2x_3^2 - 4x_1x_2 + 2x_1x_3 - 4x_2x_3$$

Solution
The determinant of the form is

$$|a_{ij}| = \begin{vmatrix} 1 & -2 & 1 \\ -2 & 2 & -2 \\ 1 & -2 & 2 \end{vmatrix}$$

The principal minors are:

$$\Delta_1 = |1| = 1; \quad \Delta_2 = \begin{vmatrix} 1 & -2 \\ -2 & 2 \end{vmatrix} = -2; \quad \Delta_3 = \begin{vmatrix} 1 & -2 & 1 \\ -2 & 2 & -2 \\ 1 & -2 & 2 \end{vmatrix} = 2$$

According with Sylvester theorem, F is not definite.

(3) Find the relative maximum and minimum of the function

$$f(x, y, z) = x^2 + y^2 + 3z^2 - xy + 2xz + yz$$

Solution

$$f_x = 2x - y + 2z$$
$$f_y = 2y - x + z$$
$$f_z = 6z + 2x + y$$

To find the stationary values, one computes:

$$f_x = f_y = f_z = 0$$

and the solution is $x = y = z = 0$.

In order to find if in this point the function has a maximum or a minimum, the method of the principal minors is used:

$$f_{xx} = 2 > 0; \begin{vmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{vmatrix} = 3 > 0; \begin{vmatrix} f_{xx} & f_{xy} & f_{xz} \\ f_{yz} & f_{yy} & f_{yz} \\ f_{zx} & f_{zy} & f_{zz} \end{vmatrix} = 4 > 0$$

Since $f(x, y, z) \geq f(0, 0, 0) = 0$ it comes out that $(0, 0, 0)$ is a point of relative minimum.

2.3 Constrained Optimization

Many practical situations, both in life and in engineering applications, require solving optimization problems under additional *constraints* (also called *links* or *restrictions*).

For example, determine the shortest arc of the curve joining two points on the surface of a sphere. That arc of curve is an implicit function of the spatial coordinates x, y, z. In addition, the equation of the sphere is a constraint (link) for those coordinates.

Another example, which is considered to be classic, requires to find among all (plane) closed curves of given length, that curve closing the maximum surface area. The fixed length of the curve is the additional constraint that must be met by all plane curves that constitute the solution of the problem.

2.3.1 Functions of Two Variables

For a gradual introduction, consider first the problem of finding the stationary values of a function of two variables, $f(x, y)$, which are related by the additional constraint $g(x, y) = 0$. From the geometric point of view, $g(x, y) = 0$ is a curve in the plane Oxy (Fig. 2.1).

Consider now the equation $f(x, y) = c$, which is a family of curves of the real parameter c, which, when this parameter varies (taking the values $c_1, c_2, \ldots, c_k, \ldots$), covers part of the plane Oxy.

Of all the curves $f(x, y) = c$ that intersect the curve $g(x, y) = 0$, one searches for that curve whose parameter c takes the smallest or the largest value. In general, the curves $f(x, y) = c$ intersect the curve $g(x, y) = 0$ in two points, in a single point or it does not intersect it. Since the parameter c increases or decreases monotonously, its

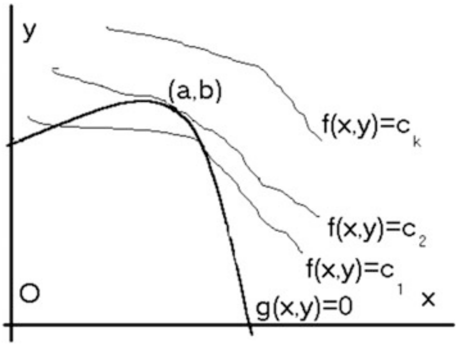

Fig. 2.1 Calculation of the constrained extreme of a function of two variables (adapted from Forray 1975)

lowest (or highest) value is to be found when changing from two points of intersection, to zero points of intersection (i.e. in the point of tangency). This point will be denoted (a, b). In the point of tangency, the slopes of the tangents at the curves $g(x, y) = 0$ and $f(x, y) = c$, respectively, are equal, so that one can write:

$$\frac{f_x}{g_x} = \frac{f_y}{g_y}(\equiv -\lambda) \tag{2.3.1}$$

if $g_x g_y \neq 0$ in (a, b). This yields the following two equations:

$$f_x + \lambda g_x = 0 \qquad f_y + \lambda g_y = 0 \tag{2.3.2}$$

Together with the equation $g(x, y) = 0$, they constitute a system of three equations. Solving this system, one finally find the unknowns (a, b) (i.e. the coordinates of the tangent point) and the parameter λ.

This result becomes invalid if the curve $g(x, y) = 0$ has in $x = a,\ y = b$ a *singular point* (i.e. a point where both partial derivatives g_x and g_y are null).

The above intuitive study constitutes the essential of a classical procedure to finding the extreme of a function under additional constraints, which is called the *Lagrange multipliers method*. In this case, the parameter λ is a multiplier. As already seen, the method does not apply in singular points. The recipe of method application is:

(i) The following function is built:

$$F(x, y, \lambda) = f(x, y) + \lambda g(x, y) \tag{2.3.3}$$

(ii) The following system of three equations is solved for the three unknowns (x, y, λ):

$$\frac{\partial F}{\partial x} = 0 \qquad \frac{\partial F}{\partial y} = 0 \qquad \frac{\partial F}{\partial \lambda} = 0 \tag{2.3.4}$$

It is immediately apparent that this system yields the Eq. (2.3.2) and the constraint $g(x, y) = 0$.

Example
Determine the rectangle of given perimeter length L whose surface area is a maximum.

Solution
Denote by x, y the lengths of the sides of the rectangle $(x > 0, y > 0)$. The surface area of the rectangle is:

$$f(x, y) = xy$$

The constraint is:

$$g(x, y) = 2x + 2y - L = 0$$

The function f should be maximized, by taking into account this constraint. The following system of equations with three unknown is to be solved:

$$\begin{gathered} f_x + \lambda g_x = 0 \\ f_y + \lambda g_y = 0 \\ g(x, y) = 0 \end{gathered}$$

After some computations, one finds:

$$y + 2\lambda = 0; \quad x + 2\lambda = 0; \quad 2x + 2y - L = 0$$

The solution is as follows:

$$x = a = L/4; \quad y = b = L/4 \quad \lambda = -L/4$$

For given perimeter length, the rectangle of maximum surface area is a square.

2.3.2 *Functions with Arbitrary Finite Number of Variables*

Consider the problem of finding the extreme of a function $f(x_1, \ldots, x_n)$ of n variables under k constraints $(k < n)$:

$$g_j(x_1, \ldots, x_n) = 0 \qquad (j = 1, \ldots, k) \tag{2.3.5}$$

In principle, the method of solving this problem is as follows. The equations of the k constraints (2.3.5) can be solved to find k of the n unknown, as functions of the other $n - k$ variables. Then, replacing the expressions just found for these k unknowns, into the expression of the function f, one finds a function of $n - k$ independent variables, whose extremization can be treated as a common problem of maximum or minimum, by using the methods previously presented. This method is typically used when the number of constraints is small, allowing to find the explicit solution of the k unknowns.

Another, more general method of solving the problem, relies on the fact that when the extreme of the function f is reached, the differential df is canceled:

$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i = 0 \tag{2.3.6}$$

The n variables are not independent, since they are under the constraints Eq. (2.3.5). It follows that the n differentials dx_i obey k constraints; therefore only $n-k$ differentials among them are independent. The differentials dx_i should fulfill the k constraints, i.e.:

$$\sum_{i=1}^{n} \frac{\partial g_j}{\partial x_i} dx_i = 0 \qquad (j = 1, \cdots, k) \tag{2.3.7}$$

One multiplies each equation j of the equations system (2.3.7) by a parameter λ_j and the next operation is summing up in respect with j, from $j = 1$ to $j = k$. Combining the result obtained with Eq. (2.3.6), one finds:

$$\sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i} + \sum_{j=1}^{k} \lambda_j \frac{\partial g_j}{\partial x_i} \right) dx_i = 0 \tag{2.3.8}$$

The coefficients λ_j are chosen so that k of the n expressions in parentheses are canceled. Then, the rest of $n-k$ expressions in parentheses should always be zero, since the remaining $n-k$ differentials dx_i are arbitrary (being independent). The conclusion is that for a relative extreme, the necessary conditions are:

$$\begin{aligned} &\tfrac{\partial f}{\partial x_i} + \sum_{j=1}^{k} \lambda_j \tfrac{\partial g_j}{\partial x_i} = 0 \qquad (i = 1, \cdots, n) \\ &g_j(x_1, \ldots, x_n) = 0 \qquad (j = 1, \cdots, k) \end{aligned} \tag{2.3.9}$$

The system (2.3.9) of $n+k$ equations must be solved for the same number of unknowns, specifically for $x_1, \ldots, x_n; \quad \lambda_1, \ldots, \lambda_k$.

This optimization method naturally carries again the name *Lagrange multipliers method*, being an extension of the method shown in the case of functions with two variables. The method is more compactly and elegantly formulated as follows:

(i) Build the function

$$F(x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_k) = f(x_1, \ldots, x_n) + \sum_{j=1}^{n} \lambda_j g_j(x_1, \ldots, x_n) \tag{2.3.10}$$

(ii) solve the following system of $n+k$ (generally nonlinear) equations for the unknowns $x_1, \ldots, x_n; \quad \lambda_1, \ldots, \lambda_k$:

$$\begin{aligned} \frac{\partial F}{\partial x_i} &= 0 \qquad (i = 1, \ldots, n) \\ \frac{\partial F}{\partial \lambda_j} &= 0 \qquad (j = 1, \ldots, k) \end{aligned} \tag{2.3.11}$$

Note that this procedure leads to the system of Eq. (2.3.9).

Example

Find the largest and smallest values of z placed on the ellipse formed by the intersection of the plane $x + y + z = 1$ and the ellipsoid $16x^2 + 4y^2 + z^2 = 16$.

Solution

The function to be extremized is:

$$f(x, y, z) = z$$

The constraints are:

$$g_1(x, y, z) = x + y + z - 1 = 0$$

$$g_2(x, y, z) = 16x^2 + 4y^2 + z^2 - 16 = 0$$

The following function is built:

$$F(x, y, z, \lambda_1, \lambda_2) = z + \lambda_1(x + y + z - 1) + \lambda_2(16x^2 + 4y^2 + z^2 - 16)$$

The following system is then solved for the unknowns $(x, y, z, \lambda_1, \lambda_2)$:

$$F_x = F_y = F_z = F_{\lambda_1} = F_{\lambda_2} = 0$$

The maximum and minimum values of z are:

$$z_{\max} = 8/3; \quad z_{\min} = -8/7.$$

Reference

Forray, M.: Calculul variational in stiinta si tehnica. Editura Tehnica, Bucuresti (1975)

http://www.springer.com/978-3-319-52967-7

Optimal Control in Thermal Engineering
Badescu, V.
2017, XV, 588 p. 179 illus., Hardcover
ISBN: 978-3-319-52967-7