Boundary extrema

If f is continuous on a compact set A, then f has absolute extrema on A. They may occur either at interior or at boundary points of A. If an absolute maximum occurs at an interior point $\mathbf{x}_0$ of A, then $\mathbf{x}_0$ is among the relative maxima in int A. We can try to find it by Theorem 3.4. However, Theorem 3.4 does not apply at boundary points of A.

If $\mathbf{x}_0 \in \text{fr } A$ and $\mathbf{x}_0$ gives an absolute maximum, then $f(\mathbf{x}) \leq f(\mathbf{x}_0)$ for every $\mathbf{x} \in A$, and in particular for every $\mathbf{x} \in \text{fr } A$. Therefore $\mathbf{x}_0$ also gives an absolute maximum among points of fr A. If fr A is sufficiently smooth, the Lagrange multiplier rule (Section 4.8) can be applied.

In Section 3.6 we discuss extrema of linear functions, for which calculus is of no use.

Problems

In Problems 1 through 6 let $A = E^n$ for the indicated n.

1. Find the critical points, relative extrema, and saddle points. Make a sketch indicating the level sets.
   (a) $f(x, y) = x - x^2 - y^2$. (b) $f(x, y) = (x + 1)(y - 2)$.
   (c) $f(x, y) = \sin(xy)$. (d) $f(x, y) = xy(x - 1)$.

2. Find the critical points, relative extrema, and saddle points.
   (a) $f(x, y) = x^3 + x - 4xy - 2y^2$.
   (b) $f(x, y) = x(y + 1) - x^2y$.
   (c) $f(x, y) = \cos x \cosh y$.
   [*Note*: The hyperbolic functions sinh and cosh are defined by
   $$\sinh x = \tfrac{1}{2}[\exp x - \exp(-x)],$$
   $$\cosh x = \tfrac{1}{2}[\exp x + \exp(-x)].$$
   Their derivatives are given by the formulas $\sinh' = \cosh$, $\cosh' = \sinh$.]

3. Let $f(x, y, z) = x^2 + y^2 - z^2$. Show that f has one critical point, which does not give a relative extremum. Describe the level sets.

4. Let $f(x, y, z) = x^2 + 3y^2 + 2z^2 - 2xy + 2xz$. Show that 0 is the minimum value of f.

5. Given $\mathbf{x}_1, \ldots, \mathbf{x}_m$, find the point $\mathbf{x}$ where $\sum_{j=1}^{m} |\mathbf{x} - \mathbf{x}_j|^2$ has an absolute minimum, and find the minimum value.

6. (a) In Problem 1(a) find the (absolute) maximum and minimum values of f on the circular disk $x^2 + y^2 \leq 1$.
   (b) Do the same for 1(c).

7. (a) Let $f(\mathbf{x}) = \psi(\mathbf{a} \cdot \mathbf{x})$, where ψ is of class $C^{(2)}$ and $\mathbf{a} \neq \mathbf{0}$. Find all critical points, and show that every critical point is degenerate.
   (b) Illustrate this result in case $f(x, y) = (x - y)^2$.

8. Let $g(\mathbf{h}) = \sum_{i,j=1}^{n} c_{ij}h^i h^j$. Assume that $g > 0$, that is, that $g(\mathbf{h}) > 0$ for every $\mathbf{h} \neq \mathbf{0}$.
(a) Show that there exists a number $m > 0$ such that $g(\mathbf{h}) \geq m|\mathbf{h}|^2$ for every $\mathbf{h}$. [*Hint*: The polynomial g is continuous and has a positive minimum value m on the unit $(n-1)$-sphere $\{\mathbf{h} : |\mathbf{h}| = 1\}$.]
(b) Suppose that $|C_{ij} - c_{ij}| < \varepsilon n^{-2}$ for each $i,j = 1, \ldots, n$. Let $G(\mathbf{h}) = \sum_{i,j=1}^{n} C_{ij}h^i h^j$. Show that $G(\mathbf{h}) \geq (m - \varepsilon)|\mathbf{h}|^2$ for every $\mathbf{h}$. Hence $G > 0$ if $\varepsilon < m$.

9. Let $\mathbf{x}_0$ be a nondegenerate critical point of a function f of class $C^{(2)}$. Show that $\mathbf{x}_0$ is isolated, that is, that $\mathbf{x}_0$ has a neighborhood U containing no other critical points of f. [*Hint*: Let $\mathbf{x}$ be another critical point in U. Apply the mean value theorem to each of the functions $f_1, \ldots, f_n$ to find that

$$(*) \qquad 0 = \sum_{j=1}^{n} f_{ij}(\mathbf{y}_i)(x^j - x_0^j), \qquad i = 1, \ldots, n,$$

where each $\mathbf{y}_i \in U$. Show that if U is small enough, $\det(f_{ij}(\mathbf{y}_i)) \neq 0$ and consequently the system of equations (*) has only the solution $\mathbf{x} - \mathbf{x}_0 = \mathbf{0}$, a contradiction.]

*3.6 Convex and concave functions

Functions that are either concave or convex arise naturally in connection with the study of convex sets. They also occur in a wide variety of applications of calculus. We will see that the theory of maxima and minima is much simpler for them than for functions that are neither concave nor convex.

Let f be a real valued function and K a convex subset of the domain of f.

Definition. The function f is *convex on* K if, for every $\mathbf{x}_1, \mathbf{x}_2 \in K$ and $t \in [0, 1]$,

$$(3.21a) \qquad f(t\mathbf{x}_1 + (1-t)\mathbf{x}_2) \leq tf(\mathbf{x}_1) + (1-t)f(\mathbf{x}_2).$$

If strict inequality holds in (3.21a) whenever $\mathbf{x}_1 \neq \mathbf{x}_2$ and $0 < t < 1$, then f is *strictly convex on* K.

The assumption that K is a convex set is needed to ensure that the point $t\mathbf{x}_1 + (1-t)\mathbf{x}_2$ belongs to the domain of f. In order to see the geometric meaning of convexity, let us denote points of E^{n+1} by $(x^1, \ldots, x^n, z)$ or, for short, by $(\mathbf{x}, z)$. Let

$$K^+ = \{(\mathbf{x}, z) : \mathbf{x} \in K, z \geq f(\mathbf{x})\}.$$

If $\mathbf{x}_1 = \mathbf{x}_2$, then (3.21a) holds trivially. Therefore, suppose that $\mathbf{x}_1 \neq \mathbf{x}_2$. Let l denote the line segment in E^{n+1} joining $(\mathbf{x}_1, f(\mathbf{x}_1))$ and $(\mathbf{x}_2, f(\mathbf{x}_2))$. Points of l are of the form

$$(t\mathbf{x}_1 + (1-t)\mathbf{x}_2, tf(\mathbf{x}_1) + (1-t)f(\mathbf{x}_2)),$$

where $t \in [0, 1]$. Inequality (3.21a) says that such points belong to K^+. Therefore, the definition says geometrically that the line segment l is contained in K^+ for every pair of points $\mathbf{x}_1, \mathbf{x}_2 \in K$.

as follows. We have

$$\mathbf{w}^i = \sum_{j=1}^{r} c_j^i \boldsymbol{\varepsilon}^j, \qquad i = 1, \ldots, n,$$

where $\{\boldsymbol{\varepsilon}^1, \ldots, \boldsymbol{\varepsilon}^r\}$ is the standard basis for $(E^r)^*$. If we write $\mathbf{b} = \mathbf{L}^*(\mathbf{a})$, then the components of $\mathbf{b}$ satisfy

$$b_j = \sum_{i=1}^{n} a_i c_j^i, \qquad j = 1, \ldots, r. \tag{4.7}$$

If one regards the covector $\mathbf{a}$ as a row vector, then (4.7) states that $\mathbf{b}$ is obtained by multiplying $\mathbf{a}$ on the right by the matrix (c_j^i).

EXAMPLE 2 (*continued*). In this example the row covectors are $\mathbf{w}^1 = \boldsymbol{\varepsilon}^1 - \boldsymbol{\varepsilon}^2 - \boldsymbol{\varepsilon}^3$, $\mathbf{w}^2 = 2\boldsymbol{\varepsilon}^1 + \boldsymbol{\varepsilon}^2 + 4\boldsymbol{\varepsilon}^3$, $\mathbf{w}^3 = -\boldsymbol{\varepsilon}^1 - \boldsymbol{\varepsilon}^3$, and $\mathbf{L}^*(\mathbf{a}) = a_1\mathbf{w}^1 + a_2\mathbf{w}^2 + a_3\mathbf{w}^3$.

Note that Formulas (4.6) and (4.7) are statements about covectors corresponding to Formulas (4.4) and (4.2) about vectors. We say that (4.6) is dual to (4.4), and (4.7) dual to (4.2). Further aspects of the duality between $\mathbf{L}$ and $\mathbf{L}^*$ are mentioned in Problem 5.

PROBLEMS

1. Let $r = 3$, $n = 2$, and $\mathbf{L}$ be the linear transformation such that $\mathbf{L}(\boldsymbol{\varepsilon}_1) = \mathbf{e}_1 - 2\mathbf{e}_2$, $\mathbf{L}(\boldsymbol{\varepsilon}_2) = \mathbf{e}_1$, $\mathbf{L}(\boldsymbol{\varepsilon}_3) = 5\mathbf{e}_1 + \mathbf{e}_2$. Find the matrix of $\mathbf{L}$, the rank, and the kernel.

2. Show that the kernel of a linear transformation is a vector subspace of its domain.

3. Let $r = n$, and let $L^i(\mathbf{t}) = c^i t^i$ for every $\mathbf{t} \in E^n$, where $c^1, \ldots, c^n$ are scalars.
   (a) What is the matrix?
   (b) Find $\mathbf{L}^{-1}$ if it exists.
   (c) If $c^1 = \cdots = c^n > 0$, then $\mathbf{L}$ is called *homothetic* about $\mathbf{0}$. Describe $\mathbf{L}$ geometrically. Show that if $\mathbf{L}$ and $\mathbf{M}$ are homothetic about $\mathbf{0}$, then $\mathbf{L}^{-1}$ and $\mathbf{M} \circ \mathbf{L}$ are also homothetic about $\mathbf{0}$.

4. (a) Show directly from the definitions that the composite of two linear transformations is also linear.
   (b) Let (c_j^i), (d_i^l), and (b_j^l) denote respectively the matrices of $\mathbf{L}$, $\mathbf{M}$, and $\mathbf{M} \circ \mathbf{L}$. Show that

$$b_j^l = \sum_{i=1}^{n} d_i^l c_j^i, \quad \text{for } l = 1, \ldots, p, \qquad j = 1, \ldots, r. \tag{4.8}$$

5. Let $\mathbf{L}^*$ be the dual of the linear transformation $\mathbf{L}$. Show that:
   (a) $\mathbf{L}^*(\mathbf{e}^i) = \mathbf{w}^i$, $i = 1, \ldots, n$, where $\{\mathbf{e}^1, \ldots, \mathbf{e}^n\}$ is the standard basis for $(E^n)^*$, Section 3.2.
   (b) $\mathbf{a} \cdot \mathbf{L}(\mathbf{t}) = \mathbf{L}^*(\mathbf{a}) \cdot \mathbf{t}$, for every covector $\mathbf{a} \in (E^n)^*$ and vector $\mathbf{t} \in E^r$.
   (c) $\mathbf{L}(E^r)$, $\mathbf{L}^*[(E^n)^*]$ have the same dimension ρ, and $\mathbf{L}$, $\mathbf{L}^*$ have the same nullity ν. (You may use the fact that row rank of a matrix equals column rank.)