Calculus for Olympiad Problems?

0 Introduction

To some extent, I regard this note, as should you, as analogous to sex education in school. It doesn't condone or encourage you to use calculus on Olympiad inequalities. However, it rather seeks to ensure that if you insist on using calculus, that you do so properly. =)

Having made that disclaimer, let me turn around and praise calculus as a wonderful discovery that immeasurably improves our understanding of real-valued functions. If you haven't encountered calculus yet in school, this section will be a taste of what awaits you but no more than that. The treatment is far too compressed to give a comprehensive exposition. After all, isn't that what textbooks are for.

1 Continuous Functions

1.1 Continuity

What are continuous functions? You could probably say that if you draw a graph of a continuous function, there will be no 'irregular' breaks, or you could draw it in one smooth motion - all these seem to convey a notion related to the word continuous. Although this is not altogether correct, it puts us on the right track and it is often helpful to visualize continuous functions in this way.

One of the purposes of this note will be to show that continuous functions are not quite as simple as our initial intuitive feelings might lead us to believe; in fact, most continuous functions, have graphs which, though unbroken and connected, are not smooth at all. The depth of this remark can be illuminated by pointing out that of the continuous functions only the simplest and most well-behaved can be graphed at all.

Nevertheless, continuity is an extremely important concept in mathematics and mathematics application and will occupy a central role in the remainder of this note. We begin with the definition of continuity of a function at a point in its domain.

Definition 1.1 *f is said to be continuous at x_0 if $\lim_{x\to x_0} f(x) = f(x_0)$; equivalently, given any $\varepsilon > 0$, there exists a $\delta > 0$ such that $|f(x) - f(x_0)| < \varepsilon$ whenever $|x - x_0| < \delta$.*

Example 1.1

1. $f_1(x) = x^3 + 19x^2 + 99x + 1$ *is continuous on* $\Re$.
2. $f_2(x) = \frac{x^2-4}{x-2}$ *is discontinuous at* $x = 2$ *since* $f(2)$ *does not exists.*
3. $f_3(x) = \begin{cases} sgn|x| & \text{if } x \neq 0 \\ 1 & \text{if } x = 0 \end{cases}$ *is continuous on* $\Re$.

Definition 1.2 *f is said to be right continuous (or continuous from the right) at x_0 if $\lim_{x\to x_0^+} f(x) = f(x_0)$; similarly, f is left continuous (continuous from the left) at x_0 if $\lim_{x\to x_0^-} f(x) = f(x_0)$.*

In order for f to be right-continuous at x_0, it is necessary that f be defined on an interval of the form $[x_0, x_0+\eta)$ for some $\eta > 0$. In f is left-continuous at x_0 then f must be defined on $(x_0 - \eta, x_0]$ for some $\eta > 0$. Also, f is continuous if and only if f is both left- and right-continuous at x_0.

Example 1.2

1. $g_1(x) = |x|$ *is continuous on* $\Re$.
2. $g_2(x) = \lfloor x \rfloor$ *is discontinuous at every integer and is continuous at all other real numbers.*

Theorem 1.1 *If* f *and* g *are each continuous at* $x = x_0$, *then* $f + g, f \cdot g$ *are continuous at* x_0 *and* f/g *is continuous at* x_0 *provided* $g(x_0) \neq 0$.

Theorem 1.2 *If* f *is continuous at* $x = x_0$, *and* g *is continuous at* $f(x_0)$ *then* $(g \circ f)(x) = g(f(x))$ *is continuous at* x_0.

1.2 Properties of Continuous Functions

Definition 1.3 *A function* $f : A \to B$ *is called* **bounded** *if there exists a real number* $M > 0$ *such that* $|f(x)| \leq M$ *for every* $x \in A$.

In this section, we investigate some important properties of continuous functions. The functions $f(x) = x$ on $\Re$ and $f(x) = \frac{1}{x}$ on $(0, 1)$ are not bounded. Yet each of these function is continuous on the given domain. Our first result is that if a function f is continuous on a closed bounded interval $[a, b]$, then f is necessarily bounded on $[a, b]$.

Theorem 1.3 *If* f *is continuous on the closed, bounded interval* $[a, b]$, *then* f *is bounded on* $[a, b]$.

Recall that each polynomial function is bounded on every bounded interval I, and hence continuous functions on $\Re$ are like polynomials in this way. Of course, the polynomials are only a small subset of the set of all continuous functions on $\Re$.

Suppose the function f is bounded on the set A; and define

$$M = \sup_{x \in A} f(x) \text{ and } m = \inf_{x \in A} f(x).$$

Example 1.3

1. Let $f_1(x) = x^2$ *on* $(0, 2)$. f *is continuous and bounded on* $(0, 2)$. $M = 4$ *and* $m = 0$, *but there are no points* $x_1, x_2 \in (0, 2)$ *with* $f(x_1) = 4$ *and* $f(x_2) = 0$.
2. Let $f_2(x) = \frac{x|x|}{(1+x^2)}$ *on* $\Re$. *Note that* $f(x)$ *is an odd function;* $f(x)$ *is continuous and bounded on* $\Re$, $M = 1$ *and* $m = -1$, *but there are no points* $x_1, x_2 \in \Re$ *with* $f(x_1) = 1$ *and* $f(x_2) = -1$.

In these two examples, the extreme values of the function f, M and m, are not always assumed by the function. The following important theorem gives us conditions under which the extreme values are necessarily assumed.

Theorem 1.4 *(Extreme Value Theorem) If* f *is continuous on* $[a, b]$, *then there exists points* $x_1, x_2 \in [a, b]$ *such that* $f(x_2) \leq f(x) \leq f(x_1)$ *for all* $x \in [a, b]$.

Proof Suppose that f is continuos on $[a, b]$; then f is bounded on $[a, b]$, and so $M = \sup_{x \in [a,b]} f(x)$ and $m = \inf_{x \in [a,b]} f(x)$ exist as real numbers. Suppose that the value M is not assumed; then $f(x) < M$ for every $x \in [a, b]$. Define $g(x) = \frac{1}{M - f(x)}$ on $[a, b]$; clearly $g(x) > 0$ for every $x \in [a, b]$ and by Theorem 1.1, g is continuous on $[a, b]$. Then by Theorem 1.3, g is bounded on $[a, b]$, and so there is a real number $k > 0$ such that $g(x) \leq k$ for every $x \in [a, b]$. Now, for every $x \in [a, b]$,

$$k \geq g(x) = \frac{1}{M - f(x)}$$

and so $M - f(x) \geq 1/k > 0$. Hence $f(x) \leq M - 1/k$ on $[a, b]$, and this contradicts the definition of M as the least upper bound of f on $[a, b]$. Therefore, the value M must be assumed. To prove the existence of an $x_2 \in [a, b]$ with $f(x_2) = m$, we can apply the same argument to the function $-f$.

Theorem 1.5 *(Intermediate Value Theorem) If f is continuous on $[a, b]$ and k is between $f(a)$ and $f(b)$ then there exists $c \in (a, b)$ such that $f(c) = k$.*

Theorem 1.6 *If f is one-to-one and continuous on $[a, b]$, then f is strictly monotone on $[a, b]$.*

2 Differentiable Functions

2.1 The Derivative

In this section, we study the derivative of a function and the properties of differentiable functions. The reader may recall from elementary calculus that the derivative of a function f is a new function f' which represents the rate of change of f as x changes. It therefore has applications in any discipline where change is measured. We first assume that f is defined in a neighbourhood of x_0.

Definition 2.1 *The* **derivative** *of f at x_0 is*

$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

provided the limit exists. When the limit exists, we say that f is **differentiable** *at x_0.*

The fraction on the right-hand side of the equation is also known as the difference quotient. An equivalent definition for the derivative of $f(x)$ at x_0 is

$$f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}.$$

Theorem 2.1 *If f is differentiable at x_0, then f is continuous at x_0.*

Proof Recall that f is continuous at x_0 if $\lim_{x \to x_0} f(x) = f(x_0)$, or equivalently $\lim_{x \to x_0} f(x) - f(x_0) = 0$. We assume that f is differentiable at x_0, then

$$\begin{aligned}
\lim_{x \to x_0} f(x) - f(x_0) &= \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}(x - x_0) \\
&= \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} \lim_{x \to x_0} (x - x_0) \\
&= f'(x_0) \cdot 0 = 0.
\end{aligned}$$

We assume that the reader is familiar with the geometrical interpretation of the derivative $f'(x_0)$ as the slope of the line tangent to the graph of $y = f(x)$ at the point $(x, f(x_0))$. We also assume familiarity with the differentiation of polynomials and the basic rules of differentiation:

1. $(f + g)'(x) = f'(x) + g(x)$
2. $(kf)'(x) = kf'(x)$
3. $(fg)'(x) = f'(x)g(x) + f(x)g'(x)$
4. $(f/g)'(x) = [g(x)f'(x) - f(x)g'(x)]/[g(x)]^2$

whenever the right-hand sides of the equations exist.

Theorem 2.2 *If f is differentiable at x_0 and g is differentiable at $y_0 = f(x_0)$ then $h = g \circ f$ is differentiable at x_0 and*

$$h'(x_0) = (g \circ f)'(x_0) = g'[f(x_0)] \cdot f'(x_0) = g'(y_0) \cdot f'(x_0).$$

2.2 Properties of Differentiable Functions

In this section, we study some important properties of differentiable functions; in particular, we shall be concerned with the relationship between f and f'. We say that f is differentiable on the set A if f is differentiable at each point in A.

We begin by investigating maxima and minima of a function f and see how they are related to f'. Such considerations are extremely important in the various applications.

Definition 2.2 *$f(x_0)$ is a* **local maximum** *of the function f if for all x in some neighbourhood of x_0 we have $f(x) \leq f(x_0)$. Similarly, $f(x_0)$ is a* **local minimum** *of the function f if for all x in some neighbourhood of x_0 we have $f(x) \geq f(x_0)$.*

Definition 2.3 *Let x_0 be an element in the interval I. $f(x_0)$ is the* **absolute maximum** *of f on I if $f(x_0) \geq f(x)$ for all $x \in I$. Similarly, $f(x_0)$ is the* **absolute minimum** *of f on I if $f(x_0) \leq f(x)$ for all $x \in I$.*

It is clear from the above definitions that if x_0 is an interior point of the interval I, and $f(x_0)$ is the absolute maximum of $f(x)$ on I then $f(x_0)$ is a local maximum of f. A similar statement holds for minima.

Theorem 2.3 *If $f(x_0)$ is a local extremum (maximum or minimum) then either $f'(x_0) = 0$ or $f'(x_0)$ does not exist.*

Proof Suppose $f(x_0)$ is a local maximum (a similar proof holds for the case when $f(x_0)$ is a local minimum). Then there is a $\delta > 0$ such that for every $x \in N_\delta(x_0)$, $f(x) \leq f(x_0)$. Hence

$$\frac{f(x) - f(x_0)}{x - x_0} = \begin{cases} \leq 0 & \text{if } x \text{ satisfies } x_0 < x < x_0 + \delta \\ \geq 0 & \text{if } x \text{ satisfies } x_0 - \delta < x < x_0 \end{cases}$$

Now if $f'(x_0)$ exists, then necessarily,

$$\lim_{x \to x_0^+} \frac{f(x) - f(x_0)}{x - x_0} = f'(x_0) = \lim_{x \to x_0^-} \frac{f(x) - f(x_0)}{x - x_0}$$

But by the above,

$$\lim_{x \to x_0^+} \frac{f(x) - f(x_0)}{x - x_0} \leq 0 \text{ and } \lim_{x \to x_0^-} \frac{f(x) - f(x_0)}{x - x_0} \geq 0$$

It follows that $f'(x_0) = 0$.

The function $f(x) = |x|$ provides an example of a function which has a local extremum at a point where the function fails to be differentiable.

Theorem 2.4 *(Rolle's Theorem) If $f(x)$ is continuous on $[a, b]$, and differentiable on (a, b) and $f(a) = f(b)$ then there exists an $x_0 \in (a, b)$ such that $f'(x_0) = 0$.*

Theorem 2.5 *(Mean-Value Theorem) If f is continuous on $[a, b]$ and differentiable on (a, b), then there exists a point $x_0 \in (a, b)$ such that*

$$f'(x_0) = \frac{f(b) - f(a)}{b - a}.$$

The following theorem, which is our first application of the mean-value theorem, is useful in many ways.

Theorem 2.6 *If f is differentiable on (a, b) and $f'(x) \geq 0$ for every $x \in (a, b)$, then f is monotone increasing on (a, b).*

Proof Suppose $x_1, x_2 \in (a, b)$ with $x_1 < x_2$. Then f is continuous on $[x_1, x_2]$ and is differentiable on (x_1, x_2). By the Mean Value Theorem, there is an x_0, with $x_1 < x_0 < x_2$ such that

$$f'(x_0) = \frac{f(x_2) - f(x_1)}{x_2 - x_1}$$

Since $x_2 - x_1 > 0$ and $f'(x_0) \geq 0$ by hypothesis, it follows that $f(x_1) \leq f(x_2)$, hence f is monotone increasing on (a, b). A similar results holds when $f'(x) \leq 0$ on an interval (a, b). So for example, to find the extrema of a continuous function on a closed interval, it suffices to evaluate it at

1 all points where the derivative vanishes,

2 all points where the derivative is not defined,

3 the endpoints of the interval.

since we know the function has global minima and maxima, and each of these must occur at one of the aforementioned points. If the interval is open or in infinite at either end, one must also check the limiting behavior of the function there.

2.3 Convexity and the Second Derivative

A function f defined on an interval (which may be open, closed or infinite on either end), is said to be convex if the set

$$\{(x, y) \in \Re^2 : y \geq f(x)\}$$

is convex. We say f is concave if -f is convex. This terminology was standard at one time, but today most calculus textbooks use "concave up" and "concave down" for our "convex" and "concave".

A more enlightening way to state the definition might be that f is convex if for any $t \in [0, 1]$ and any x, y in the domain of f, we have

$$tf(x) + (1 - t)f(y) \geq f(tx + (1 - t)y)$$

If f is continuous, it suffices to check this for $t = \frac{1}{2}$. Conversely, a convex function is automatically continuous except possibly at the endpoints of the interval on which it is defined.

Theorem 2.7 *If f is a convex function, then the following statements hold:*
1. If $a \leq b < c \leq d$ then $\frac{f(c)-f(a)}{c-a} \leq \frac{f(d)-f(b)}{d-b}$
(The slope of secant lines through the graph of f increase with either endpoint.)

2. If f is differentiable everywhere, then its derivative (that is, the slope of the tangent line to the graph of f is an increasing function.)

The utility of convexity for proving inequalities comes from two factors. The first factor is Jensen's inequality, which one may regard as a formal statement of the smoothing principle for convex functions.

Theorem 2.8 *(Jensen's Inequality) Let f be a convex function on an interval I and let $w_1, ..., w_n$ be nonnegative real numbers whose sum is 1. Then for all $x_1, ..., x_n \in I$,*

$$w_1 f(x_1) + ... + w_n f(x_n) \geq f(w_1 x_1 + ... + w_n x_n).$$

Proof An easy induction on n. The case $n = 2$ being the second definition above.

The second factor is the ease with which convexity can be checked using calculus, namely via the second derivative test.

Theorem 2.9 *Let f be a twice differentiable function on an open interval I. Then f is convex on I if and only if $f''(x) \geq 0$ for all $x \in I$.*

3 Differentiable Functions of Several Variables

3.1 Partial Derivatives and Differentials

The partial derivative of $f(x, y)$ with respect to x at (x_0, y_0) is

$$\begin{aligned} \frac{\partial f}{\partial x}\Big|_{(x_0,y_0)} &= \frac{df(x, y_0)}{dx}\Big|_{x=x_0} \\ &= \lim_{h\to 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h}, \end{aligned}$$

provided the limit exists. Note that $f(x, y_0)$ is a function of x only, since y_0 is a fixed number. Likewise, The partial derivative of $f(x, y)$ with respect to y at (x_0, y_0) is

$$\begin{aligned} \frac{\partial f}{\partial y}\Big|_{(x_0,y_0)} &= \frac{df(x_0, y)}{dy}\Big|_{y=y_0} \\ &= \lim_{k\to 0} \frac{f(x_0, y_0 + k) - f(x_0, y_0)}{k}, \end{aligned}$$

Other notations commonly used (depending on the emphasis) are: f_x, $\frac{\partial f}{\partial x}$, f_y, $\frac{\partial z}{\partial x}$, etc.

Example 3.1 *Given $f(x, y) = x^2 + 3xy + y - 1$, compute f_x and f_y at $(4, -5)$.*

Solution $f_x = 2x + 3y, f_y = 3x + 1$. Thus, at $(4, -5)$, $f_x = 2(4) + 3(-5) = -7$, $f_y = 3(4) + 1 = 13$.

3.2 Higher Derivatives

Since the partial derivatives of $f(x, y)$ are themselves functions of x and y; they may in turn be (partially) differentiated. This produces partial derivatives of **order** 2; (second(-order) partial derivatives.) There are four of them, and the usual notation is:

$$\frac{\partial f^2}{\partial x^2} \text{ or } f_{xx}; \quad \frac{\partial f^2}{\partial x \partial y} \text{ or } f_{yx}; \quad \frac{\partial f^2}{\partial y \partial x} \text{ or } f_{xy}; \quad \frac{\partial f^2}{\partial y^2} \text{ or } f_{yy}.$$

Here, notice the order

$$\frac{\partial f^2}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) \text{ or } (f_y)_x = f_{yx}.$$

Theorem 3.1 *(Euler) If $f(x,y)$ and its partial derivatives f_x, f_y, f_{xy} and f_{yx} are defined throughout an open region containing the point (a,b) and are all continuous at (a,b), then*

$$f_{xy}(a,b) = f_{yx}(a,b).$$

As with second-order derivatives, the order of differentiation is immaterial as long as the function and its derivatives through the order in question are all defined throughout an open region containing the point at which the derivatives are taken and are continuous at that point.

Theorem 3.2 *Suppose that the first partial derivatives of $f(x,y)$ are defined throughout an open region R containing the point (x_0,y_0) and that f_x and f_y are continuous at (x_0,y_0). Then the change*

$$\Delta z \equiv f(x_0+\Delta x, y_0+\Delta y) - f(x_0,y_0)$$

in the value of f that results from moving from (x_0,y_0) to $(x_0+\Delta x, y_0+\Delta y)$ in R satisfies

$$\Delta z = f_x(x_0,y_0)\Delta x + f_y(x_0,y_0)\Delta y + \varepsilon_1 \Delta x + \varepsilon_2 \Delta y \tag{1}$$

where $\varepsilon_1, \varepsilon_2 \to 0$ as $\Delta x, \Delta y \to 0$.

Definition 3.1 *A function $f(x,y)$ is differentiable (totally, for emphasis) at (x_0,y_0) if $f_x(x_0,y_0)$ and $f_y(x_0,y_0)$ exist and the above relation (1) holds for f at (x_0,y_0). We simply call f differentiable if it is differentiable at every point in its domain.*

Theorem 3.3 *(Analogue to Theorem 2.1) If a function $f(x,y)$ is differentiable at (x_0,y_0), then it is continuous at (x_0,y_0).*

4 Extreme Values and Saddle Points

To determine maxima, minima or saddle points of a continuous function $f(x,y)$ on a region R in the xy-plane, we follow the following procedure:

Step 1 Make a list that includes the points where f has its local maxima and minima and evaluate f at all points on the list. The local maxima and minima of f can occur only at
(i) boundary points of R;
(ii) interior points of R where $f_x = 0 = f_y$ and the points where f_x or f_y fail to exist.

Step 2 If R is closed and bounded, look through the list for the maximum and minimum values of f; These will be the absolute maximum and minimum values of f on R.

Step 3 If R is not closed or not bounded, try the following second derivative test. (The fact that $f_x = 0 = f_y$ at an interior point (a,b) of R does not guarantee that f will have an extreme value there.) However, if f and its first and second partial derivatives are continuous on R; the following test may identify the behaviour of $f(a,b)$:
If $f_x(a,b) = 0 = f_y(a,b)$, then
(i) f has a local maximum at (a,b) if $f_{xx} < 0$ and $f_{xx}f_{yy} - f_{xy}^2 > 0$ at (a,b).
(ii) f has a local minimum at (a,b) if $f_{xx} > 0$ and $f_{xx}f_{yy} - f_{xy}^2 > 0$ at (a,b).
(iii) f has a saddle point at (a,b) if $f_{xx}f_{yy} - f_{xy}^2 < 0$ at (a,b).
The test is inconclusive if if $f_{xx}f_{yy} - f_{xy}^2 = 0$ at (a,b), and we must find some other way to determine the behaviour of f at (a,b).

The expression $f_{xx}f_{yy} - f_{xy}^2$ is called the **discriminant** of f.

Example 4.1 *Find the extreme values of* $f(x, y) = x^2 + y^2$.

Solution The domain of f is the entire plane with no boundary points. The derivatives $f_x = 2x$ and $f_y = 2y$ exist everywhere. Thus local max or min occur only when $f_x = 0 = f_y$; i.e., at the origin (0,0). Since $f(0, 0) = 0$ and $f(x, y) \geq 0$; 0 is the absolute min. (Here we did not need second derivative test, had we used it, we would have identified (0,0) as a local min.)

Example 4.2 *Find the extreme values of* $f(x, y) = xy$.

Solution Since f is differentiable everywhere and its domain has no boundary points, f can assume extreme values only where $f_x = 0 = f_y$. We have $f_x = y$ and $f_y = x$; thus at (0,0) f may assume an extreme value. Now $f_{xx} = 0, f_{yy} = 0$ and $f_{xy} = 1$ which give $f_{xx}f_{yy} - f_{xy}^2 = -1 < 0$ at $(0, 0)$. Hence f has a saddle point at (0,0), it assumes no extreme values.

Example 4.3 *Find the extreme values of* $f(x, y) = xy - x^2 - y^2 - 2x - 2y + 4$.

Solution f is defined and differentiable for all x, y and its domain has no boundary points. The extreme values therefore occur only at points where $f_x = 0 = f_y$.
$f_x = y - 2x - 2$, $f_y = x - 2y - 2$, and $f_x = f_y = 0$ if and only if $y = 2x + 2$ and $x = 2y + 2$, that is if and only if $x = y = -2$. Thus at (-2, -2) f may assume an extreme value. Now $f_{xx} = -2, f_{yy} = -2, f_{xy} = f_{yx} = 1$, therefore $f_{xx}f_{yy} - f_{xy}^2 = 3 > 0$ and $f_{xx} < 0$ implies that (-2, -2) is a maxima.
Hence the maximum value of f is given by $f(-2, -2) = 4 - 4 - 4 + 4 + 4 + 4 = 8$.

5 Lagrange Multipliers

5.1 Tangents and Normals to Surfaces

When a cartesian coordinate system is given, a curve C may be represented by a vector function

$$\mathbf{r}(t) = < x(t), y(t), z(t) > = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k},$$

where t is a real variable. Each value of t corresponds to a point $\mathbf{r}(t_0) = < x(t_0), y(t_0), z(t_0) >$ of C. Such a representation is called a parametric representation of C, and t is called the parameter.

The tangent to a curve C at a point P of C is the limiting position of the line PQ, Q being another point on C, as $Q \to P$ along C. Let C be given by $\mathbf{r}(t)$ with P and Q corresponding to t and $t + \Delta t$ respectively. Then the vector $\frac{1}{\Delta t}(\mathbf{r}(t + \Delta t) - \mathbf{r}(t))$ gives the direction of PQ. Assume that $\mathbf{r}$ is differentiable, then

$$\mathbf{r}' = lim_{\Delta t \to 0} \frac{1}{\Delta t}(\mathbf{r}(t + \Delta t) - \mathbf{r}(t))$$

is a tangent vector of C at P.

Let $f(x, y, z) = c$, c constant, be a surface S in space, and let a curve C given by $\mathbf{r}(t)$ lie on S. Then $f(x(t), y(t), z(t)) = c$. Differentiating this with respect to t we get

$$\frac{df}{dt} = f_x \frac{dx}{dt} + f_y \frac{dy}{dt} + f_z \frac{dz}{dt} = 0$$

or $< f_x, f_y, f_z > \cdot < \frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt} >= \nabla f \cdot \mathbf{r}'(t) = 0$.

This implies that ∇f and $\mathbf{r}'(t)$ are perpendicular to each other. Since $\mathbf{r}'(t)$ is in the tangent plane, ∇f is the **normal** vector of S at a given point P.

Let $P_0(x_0, y_0, z_0)$ be a point on the level surface $f(x, y, z) = c$. Since $\nabla f =< f_x, f_y, f_z > |_{P_0}$ is normal to the surface at P_0, the tangent plane to the surface at P_0 is given by

$$f_x(P_0)(x - x_0) + f_y(P_0)(y - y_0) + f_z(P_0)(z - z_0) = 0,$$

and the normal line of the surface at P_0 is the line

$$x = x_0 + f_x(P_0)t; y = y_0 + f_y(P_0)t; z = z_0 + f_z(P_0)t.$$

5.2 Constrained Maxima and Minima

We sometimes need to find the extreme values of a function whose domain is constrained to lie within some particular subset of the plane. This usually consists mostly of interior points, e.g. a closed disk or closed triangular region. But a function may be subject to other kinds of constraints as well.

In this section, we explore a powerful method for finding extreme values of **constrained** functions: the method of Lagrange multipliers, developed by Lagrange in 1755 to solve maximum-minimum problems in geometry.

Attempts to solve a constrained maximum or minimum problem by substitution, do not always go smoothly. This is one of the reasons for learning the new method of this section.

Example 5.1 *Find the points closest to the origin on the hyperbolic cylinder $x^2 - z^2 - 1 = 0$.*

Solution 1 Substitute $z^2 = x^2 - 1$ in the function (distance squared)

$$F(x, y, z) = x^2 + y^2 + z^2,$$

we get the function $h(x, y) = x^2 + y^2 + (x^2 - 1) = 2x^2 + y^2 - 1$. This has minimum -1 (!) at $(x, y) = (0, 0)$. What has gone wrong?
Ans: The constraint of points on the surface restricts the values of (x, y), which cannot be (0, 0).
Solution 2 Another way to find the points on the constraint surface (cylinder) closest to the origin is to imagine a small sphere centered at the origin expanding like a soap bubble until it just touches the cylinder. At each point of contact, the cylinder and sphere have the same tangent plane and normal line. Therefore, if the sphere and cylinder are represented as the level surfaces obtained by setting

$$f(x, y, z) = x^2 + y^2 + z^2 - a^2$$

and

$$g(x, y, z) = x^2 - z^2 - 1$$

equal to 0. Then the gradients ∇f and ∇g will be parallel where the surfaces touch. (The use of the $-a^2$ term is not exactly necessary; we may use the level curve $F(x, y, z) = a^2$) At any point of contact we should therefore be able to find a scalar λ such that

$$\nabla f = \lambda \nabla g$$

or

$$2x\mathbf{i} + 2y\mathbf{j} + 2z\mathbf{k} = \lambda(2x\mathbf{i} - 2z\mathbf{k}).$$

Thus, the coordinates x, y and z of any point of tangency will have to satisfy the three scalar equations:

$$2x = 2\lambda x; 2y = 0; 2z = -2\lambda z.$$

Now $x^2 - z^2 = 1$ implies $x \neq 0$, and this would imply that $\lambda = 1$, and hence $z = 0$. Since $y = 0$, the points we seek all have coordinates of the form $(x, 0, 0)$. But from $x^2 - z^2 = 1$, we have $x^2 = 1$ or $x = \pm 1$.
The points on the cylinder closest to the origin are the points $(\pm 1, 0, 0)$.

6 The Method of Lagrange Multipliers

In solution 2 of the previous Example, we solved the problem by the method of Lagrange multipliers. In general, the method says that the extreme values of a function $f(x, y, z)$ whose variables are subject to a constraint $g(x, y, z) = 0$ are to be found on the surface $g = 0$ at the points where $\nabla f = \lambda \nabla g$ for some scalar λ (called a Lagrange Multiplier). To explore the method further and see why it works, we first make the following observation, which we state as a theorem.

Theorem 6.1 *(Orthogonal Gradient Theorem) Suppose that $f(x, y, z)$ is differentiable in a region whose interior contains a smooth curve $C : \mathbf{r} = g(t)\mathbf{i} + h(t)\mathbf{j} + k(t)\mathbf{k}$. If P_0 is a point on C where f has a local maximum or minimum relative to its values on C; then ∇f is orthogonal to C at P_0.*

The Theorem above is the key to the method of Lagrange multipliers. Suppose that $f(x, y, z)$ and $g(x, y, z)$ are differentiable and that P_0 is a point on the surface $g(x, y, z) = 0$ where f has a local maximum or minimum value relative to its other values on the surface. Then f takes on a local maximum or minimum at P_0 relative to its values on every differentiable curve through P_0 on the surface $g(x, y, z) = 0$. Therefore, ∇f is orthogonal to the velocity vector of every such differentiable curve through P_0. But so is ∇g (since ∇g is orthogonal to the level surface $g = 0$). Therefore, at P_0, ∇f is some scalar multiple λ of ∇g.

The Procedure in The Method

Suppose that $f(x, y, z)$ and $g(x, y, z)$ are differentiable. To find the local maximum and minimum values of f subject to the constraint $g(x, y, z) = 0$, find the values of x, y, z and λ that simultaneously satisfy the equations $\nabla f = \lambda \nabla g$ and $g(x, y, z) = 0$. For functions of two independent variables, the appropriate equations are $\nabla f = \lambda \nabla g$ and $g(x, y) = 0$.

Example 6.1 *Find the greatest and smallest values that the function $f(x, y) = xy$ takes on the ellipse*

$$\frac{1}{8}x^2 + \frac{1}{2}y^2 = 1.$$

The Geometry of the Solution: The level curves of the function $f(x, y) = xy$ are the hyperbolas $xy = c$. The farther the hyperbolas lie from the origin, the larger the absolute value of f. We want to find the extreme values of $f(x, y)$, given that the point (x, y) also lies on the ellipse $x^2 + 4y^2 = 8$. Which hyperbolas intersecting the ellipse lie farthest from the origin? The hyperbolas that just graze the ellipse, the ones that are tangent to it. At these points, any vector normal to the hyperbola is normal to the ellipse, so $\nabla f = y\mathbf{i} + x\mathbf{j}$

is a multiple ($\lambda = \pm 2$) of $\nabla g = \frac{1}{4}x\mathbf{i} + y\mathbf{j}$.

Therefore, $2y = x$ or $2y = -x$, and the smallest and greatest values are obtained at (2, 1), (-2, 1), (-2, -1), (2, -1), i.e. $M = 2, m = -2$.

6.1 Lagrange Multipliers with Two Constraints

Many problems require us to find the extreme values of a differentiable function $f(x, y, z)$ whose variables are subject to two constraints. If the constraints are $g_1(x, y, z) = 0$ and $g_2(x, y, z) = 0$ and g_1 and g_2 are differentiable, with ∇g_1 not parallel to ∇g_2, we find the constrained local maxima and minima of f by introducing two Lagrange multipliers λ and μ. That is, we locate the points $P(x, y, z)$ where f takes on its constrained extreme values by finding the values of x, y, z, λ and μ that simultaneously satisfy the equations

$$\nabla f = \lambda \nabla g_1 + \mu \nabla g_2,\ g_1(x, y, z) = 0,\ g_2(x, y, z) = 0. \tag{2}$$

The equations in (2) have a nice geometric interpretation. The surfaces $g_1 = 0$ and $g_2 = 0$ (usually) intersect in a smooth curve, say C, and along this curve we seek points where f has local maximum and minimum values relative to its other values on the curve. These are the points where ∇f is normal to C, as we saw in the Orthogonal Gradient Theorem.

But ∇g_1 and ∇g_2 are also normal to C at these points because C lies in the surfaces $g_1 = 0$ and $g_2 = 0$. Therefore ∇f lies in the plane determined by ∇g_1 and ∇g_2, which means that $\nabla f = \lambda \nabla g_1 + \mu \nabla g_2$ for some λ and μ. Since the points we seek also lie in both surfaces, their coordinates must satisfy the equations $g_1(x, y, z) = 0$ and $g_2(x, y, z) = 0$, which are the remaining requirements in Eqs. (2).

7 Using Calculus in the IMO

Which finally brings us to the crux of the entire note. At this point, I would like to say again what is said at the Introduction.

"To some extent, I regard this note, as should you, as analogous to sex education in school. It doesn't condone or encourage you to use calculus on Olympiad inequalities. However, it rather seeks to ensure that if you insist on using calculus, that you do so properly. =)

Having made that disclaimer, let me turn around and praise calculus as a wonderful discovery that immeasurably improves our understanding of real-valued functions. If you haven't encountered calculus yet in school, this section will be a taste of what awaits you but no more than that. The treatment is far too compressed to give a comprehensive exposition. After all, isn't that what textbooks are for."

With this in mind, I shall illustrate how some olympiad-style questions can be solved using the techniques illustrated thus far.
Convexity

Example 7.1 *(USAMO, 1980) Prove that for numbers a, b, c in the interval [0, 1],*

$$\frac{a}{b+c+1} + \frac{b}{c+a+1} + \frac{c}{a+b+1} + (1-a)(1-b)(1-c) \leq 1.$$

Solution For any nonnegative numbers α and β, the function $f(x) = \frac{\alpha}{x+\beta}$ is convex for $x \geq 0$. Viewed as a function in any of the three variables, the given expression is a sum of two convex functions and two linear functions, so it is convex. Thus when two of the variables are fixed, the maximum is attained when the third is at one of the endpoints of the interval, so the values of the expression are always less than the largest value obtained by choosing $a, b, c \in [0, 1]$. An easy check of the eight possible cases shows that the value of the expression cannot exceed 1.

Example 7.2 *(USAMO, 1977) If $a, b, c, d, e \in [p, q]$ with $p > 0$, prove that*

$$(a+b+c+d+e)(\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}+\frac{1}{e}) \leq 25 + 6(\sqrt{\frac{p}{q}} - \sqrt{\frac{q}{p}})^2.$$

Solution If we fix four of the numbers and regard the fifth as a variable x, then the left side becomes a function of the form $\alpha x + \frac{\beta}{x} + \gamma$, where α, β, γ are positive and x ranges over the interval $[p, q]$. This function is convex on the interval $[p, q]$, being the sum of a linear and a convex function, so it attains its maximum at one (or possibly both) of the endpoints of the interval of definition. As before, this shows that if we are trying to maximise the value of the expression, it is enough to let a, b, c, d, e take the values p, q.

If n of the numbers are equal to p, then $5 - n$ are equal to q, then the left hand side is equal to

$$n^2 + (5-n)^2 + n(5-n)(\frac{p}{q} + \frac{q}{p}) = 25 + n(5-n)(\sqrt{\frac{p}{q}} - \sqrt{\frac{q}{p}})^2.$$

The maximal value of $n(5-n)$ is attained when $n = 2$ or 3, in which case $n(5-n) = 6$, and the inequality is proved.

Lagrange Multipliers

Example 7.3 *(From IMO2001 Mailing List) Let a, b, c are real numbers satisfying $a^2 + b^2 + c^2 = 9$. Prove that $2(a+b+c) - abc \leq 10$.*

Solution Set $f = 2(a+b+c) - abc$, $9 = g = a^2 + b^2 + c^2$ (constraint).
Then $\nabla f = \lambda \nabla g$ gives

$$2 - bc = \lambda 2a,\ 2 - ac = \lambda 2b,\ 2 - ab = \lambda 2c$$

Solving, we get the following cases:
(1) $a = b = c$
(2) $a = b = 2\lambda, c = 2\lambda + \frac{1}{\lambda}$, where $\lambda = \pm 1, \pm 1/sqrt(12)$ (or any cyclic permutations of a, b, c).
Checking through all the cases, we find that the maximum is obtained when a = b = 2, c = -1 (or any cyclic permutations of a, b, c), yielding $f = 2(a+b+c) - abc \leq 10$, as desired.

Example 7.4 *(National Team Selection Test 2002, Day 1 Q3) Suppose the sum of m pairwise distinct positive even numbers and n pairwise distinct positive odd numbers is 2002. What is the maximum value of $3m + 4n$?*

Solution Let $f = 3m + 4n$, $g = m(m+1) + n^2$(why?) Lagrange Multipliers imply that $\nabla f = \lambda \nabla g$, and we get

$$3 = \lambda(2m+1),$$

$$4 = \lambda(2n),$$

Substituting this into the inequality $2002 \geq m(m+1)+n^2$, we get

$$\frac{3-\lambda}{2\lambda}\frac{3+\lambda}{2\lambda}+\frac{4}{\lambda^2} \leq 2002.$$

But $\lambda > 0$, thus $\lambda \geq \frac{5}{\sqrt{8009}}$. Substituting the expressions of m and n in terms of λ into g, we get

$$g = \frac{25}{2\lambda} - \frac{3}{2} \leq 222.$$

Such a set of integers satisfy the conditions: $\{1,3,5,...,71\},\{2,4,...,48,50,56\}$. Therefore, the maximum is indeed 222.

Important! In the multivariable realm, a new phenomenon emerges that we did not have to consider in the one-dimensional case. Sometimes we are asked to prove an inequality in the case where the variables satisfy some constraint. The Lagrange multiplier criterion for an **interior** local extremum of the function $f(x_1,...,x_n)$ under the constraint $g(x_1,...,x_n) = c$ is the existence of λ such that

$$\frac{\partial f}{\partial x_i}(x_1,...,x_n) = \lambda \frac{\partial g}{\partial x_i}(x_1,...,x_n).$$

Putting these conditions together with the constraint on g, one may be able to solve and thus put restrictions on the locations of the extrema. Notice that the duality of constrained optimization shows up in the symmetry between f and g in the criterion. It is even more critical here than in the one-variable case that the Lagrange multiplier condition is a necessary one only for an interior extremum.

Unless one can prove that the given function is convex, and thus that an interior extremum must be a global one, one must also check all boundary situations, which is far from easy to do when (as often happens), these extend to infinity in some directions.

I will illustrate this with a final example.

Example 7.5 *(USAMO 2001, Q3) Let a, b and c be nonnegative real numbers such that*

$$a^2+b^2+c^2+abc = 4.$$

Prove that

$$0 \leq ab+bc+ca-abc \leq 2$$

.

Solution Let $f = ab+bc+ca-abc$ and $g = a^2+b^2+c^2+abc = 4$. For the lower bound, the set [0,2] x [0,2] x [0,2] is convex, and since the function is continuous, its maxima/minima is located at the endpoints. So we check all the 8 possible endpoints and we find that (0, 0, 2) and its two other permutations give the lower bound. In addition, these 3 vertices satisfy the constraint. The other 5 vertices don't satisfy the constraint.

Which leads us to finding the upper bound. Using Lagrange Multipliers, we have $\nabla f = \lambda \nabla g$, and thus

$$b+c-bc = \lambda(2a+bc)$$

Now we note that f and g are symmetric in its 3 variables, a, b and c. So we also have

$$a+c-ac = \lambda(2b+ac)$$

and

$$a + b - ab = \lambda(2c + ab)$$

Then we have our constraint

$$a^2 + b^2 + c^2 + abc = 4.$$

Solving, we get 2 critical points in the interior $(1,1,1)$ and $(0, \sqrt{2}, \sqrt{2})$ (and its two other permutations).

Therefore, we get M=2 , at (1,1,1), $(\sqrt{2}, \sqrt{2}, 0)$ and its two permutations, and m=0 at (2,0,0) and its two permutations.

8 Conclusion

It is hoped that this note has helped in one way or another in gifting the reader another tool when solving problems. As this note was constructed very much in a rush, there might well be mistakes and areas for improvement. Thus, all comments and suggestions for improvement are welcomed at asteea@singnet.com.sg

THE END

Version 1.0