MAX/MIN FOR FUNCTIONS OF SEVERAL VARIABLES

Abstract. These notes supplement the material in §8.2 of our text on quadratic forms and symmetric matrices. They show that the proper way generalization to functions of several variables of the Calculus I second derivative test for local maxima and minima involves a symmetric matrix formed from second partial derivatives.

1. Max/Min for functions of one variable

In this section f will be a function defined and differentiable in an open interval I of the real line.

1.1. **Critical points.** A point t_0 in I is called a *critical point* of f if $f'(t_0) = 0$. Geometrically this says that the graph $y = f(x)$ has a horizontal tangent at the point $(t_0, f(t_0))$ in $\mathbb{R}^2$.

1.2. **Local maxima and minima.** We say f has a (strict) *local maximum* at t_0 if there is some open interval containing t_0 for which $f(t_0) > f(t)$ for each t in that interval. There's a similar definition, which I leave to you, for (strict) *local minimum.*

Recall from Calculus I that:

> *If f (differentiable in an interval containing t_0) has a local maximum or minimum at t_0, then t_0 is a critical point of f, i.e., $f'(t_0) = 0$.*

However *the converse is not true:* there are critical points that are neither maxima nor minima. For example: 0, which is a critical point for $f(t) = t^3$, is neither a local maximum nor a local minimum (sketch the graph!).

So the best we can say under the hypotheses given on f is that *its local maxima and minima lie among the critical points, but they need not exhaust the critical points.*

1.3. **The Second Derivative Test.** *Suppose f is defined in an open interval I and that both the first and second derivatives of f exist and are continuous on I. Suppose t_0 in I is a critical point of f.*

(a) *If $f''(t_0) > 0$ then t_0 is a strict* minimum *of f.*

Date: April 5, 2005.

(b) *If* $f''(t_0) < 0$ *then* t_0 *is a strict* maximum *of* f*.*

(To keep this result straight, think of the examples $f(t) = t^2$ and $f(t) = -t^2$, with $t_0 = 0$.)

1.4. **One way to think of the proof.** Taylor's Theorem tells us that for h sufficiently small, $f(t_0 + h)$ is very nearly $f(t_0) + f'(t_0)h + \frac{1}{2}f''(t_0)h^2$. Since t_0 is a critical point, $f'(t_0) = 0$, so we have for all h sufficiently close to 0:

$$f(t_0 + h) \approx f(t_0) + \frac{1}{2}f''(t_0)h^2.$$

If $f''(t_0) > 0$ then, on the right-hand side of the above "equation", the term we add to $f'(t_0)$ is positive, hence $f(t_0+h) > f(t_0)$ for all t sufficiently close to t_0, i.e., t_0 is a strict local minimum of f. The same idea handles the case $f''(t_0) < 0$.

1.5. **The case** $f''(t_0) = 0$**.** Here we still can't say anything. f may have a maximum at t_0 ($f(t) = t^4$, $t_0 = 0$), a minimum ($f(t) = -t^4$, $t_0 = 0$), or neither ($f(t) = t^3$, $t_0 = 0$).

2. The two-variable case

Now we assume f is defined in an open disc in $\mathbb{R}^2$, centered at a point p_0. Let's assume that all partial derivatives of f of first and second order exist and are continuous on this disc. I leave it to you to formulate carefully the notion of "p_0 is a strict local maximum (resp. minimum) for f."

We call p_0 a *critical point* of f if both first partial derivatives of f at p_0 are zero. Geometrically this means that the graph $z = f(x, y)$ has a horizontal tangent plane at the point $(p_0, f(p_0))$ in $\mathbb{R}^3$.

2.1. **First and second derivative of a function of two variables.** For a function f as above:

- The *first derivative at* p_0 isn't any more a number, it's a *row matrix*[1]
$$f'(p_0) = \left[\frac{\partial f}{\partial x}(p_0), \frac{\partial f}{\partial y}(p_0)\right].$$
- The *second derivative* of f at p_0 is now *a* 2×2 *matrix:*
$$f''(p_0) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2}(p_0) & \frac{\partial^2 f}{\partial x \partial y}(p_0) \\ \frac{\partial^2 f}{\partial y \partial x}(p_0) & \frac{\partial^2 f}{\partial x^2}(p_0) \end{bmatrix}$$

[1]This row matrix is often denoted grad $f(p_0)$, or $\nabla f(p_0)$, and called the *gradient* of f.

2.2. **Example.** $f(x,y) = ax^2 + bxy + cy^2$, where a, b, c are real constants. Then $(0,0)$ is a critical point and

$$f''(0,0) = \begin{bmatrix} 2a & b \\ b & 2c \end{bmatrix}$$

2.3. **The Second Derivative Test.** *Suppose f is a real-valued function defined on an open disc in $\mathbb{R}^2$ centered at the point p_0, and that all partial derivatives through order two exist in this disc, and are continuous there. Suppose that p_0 is a critical point of f. Then:*

(a) *If all the eigenvalues of $f''(p_0)$ are > 0, then p_0 is a* strict local minimum *for f.*
(b) *If all the eigenvalues of $f''(p_0)$ are < 0, then p_0 is a* strict local maximum *for f.*
(c) *If one eigenvalue of $f''(p_0)$ is strictly positive and the other is strictly negative, then p_0 is a* saddle point *of f (a strict local maximum in the "positive eigen-direction" and a strict local minimum in the negative one).*

2.4. **Remark.** The "max-min" part of the second derivative test can be rephrased in the language of positive definiteness:[2]

If p_0 is a critical point of f, then
(a) *$f''(p_0)$ positive definite $\Rightarrow p_0$ a strict local minimum of f.*
(b) *$f''(p_0)$ negative definite $\Rightarrow p_0$ a strict local maximum of f.*

2.5. **Exercises.**

(1) Check that the origin is a critical point for each of the following functions, and use the second derivative test to see if it's a local maximum, local minimum, or a saddle point.
   (a) $f(x,y) = x^2 + xy + y^2 + x^3 + y^3$.
   (b) $f(x,y) = x^2 + 3xy + y^2 + \sin^3 x$
   (c) $f(x,y) = \sin(xy)$
   (d) $f(x,y) = \sin(x^2 + y^2)$
(2) Find all critical points of the function

$$f(x,y) = x^2 + xy + y^2 - 4x - 5y + 5$$

and check each to see if it's a max, min, or saddle point.

[2] See Defn. 8.2.3 and Fact 8.2.4 on p. 375 of our textbook.

2.6. **Idea of proof of two-variable second derivative test.** In Calculus III you proved a two-variable version of Taylor's Theorem whose conclusion, when recast in vector/matrix language, looks like this:

$$f(p_0 + h) = f(p_0) + f'(p_0)h + \frac{1}{2}(h^T f''(p_0)h) + \epsilon(h),$$

where h is a vector in $\mathbb{R}^2$ of small length, and

$$\frac{\|\epsilon(h)\|}{\|h\|} \to 0 \text{ as } \|h\| \to 0.$$

In other words, for p_0 a critical point and h small,

$$f(p_0 + h) \approx f(p_0) + \frac{1}{2}(h^T f''(p_0)h)$$

Now if all the eigenvalues of $f''(p_0)$ are strictly positive, then—as we know from Fact 8.2.4 on page 375 of our textbook—the quadratic form $h^T f''(p_0)h$ is strictly positive for all non-zero $h \in \mathbb{R}^2$. Thus $f(p_0 + h) > f(p_0)$ for all sufficiently small non-zero h, which means that p_0 is a local minimum for f.

The corresponding "proofs" for local maximum and saddle point are similar, and I leave the arguments to you.

3. The Second Derivative Test for functions of n variables

Here we assume that $f(x_1, \ldots, x_n)$ is a function defined at least in an open ball B in $\mathbb{R}^n$ that is centered at a point p_0:

$$B = \{p \in \mathbb{R}^n : \|p - p_0\| < r\},$$

where r is a positive number that is the *radius* of the ball. We assume f is continuous on B and that each of its partial derivatives through second order with, respect to each of the variables $x_1, \ldots, x_n$, exists and is continuous in B. For convenience, let $D_j f = \frac{\partial f}{\partial x_j}$, and $D_{i,j} f = \frac{\partial^2 f}{\partial x_i \partial x_j}$.

Just as in the $n = 2$ case, the *derivative* of f at p_0 is the row-matrix

$$f'(p_0) = [D_1 f(p_0), D_2 f(p_0), \ldots, D_n f(p_0)],$$

and the *second derivative* of f at p_0, which we denote $f''(p_0)$ is the $n \times n$ matrix whose i, j-entry is $D_{ij} f(p_0)$. As in the $n = 2$ case, the continuity assumed for all second partial derivatives of f in B insures that $D_{ij} f(p_0) = D_{j,i} f(p_0)$ for all i and j between 1 and n, i.e., that the matrix $f''(p_0)$ is *symmetric.*

We call p_0 a *critical point* of f if $f'(p_0) = 0$, i.e., if all the first partial derivatives of f at p_0 are zero.

With all of this in hand, the second derivative test looks the same as the one stated above for $n = 2$; I state it here just for local maxima and minima:

> *Under the hypotheses above on f, suppose p_0 is a critical point.*
>
> (a) *If all the eigenvalues of $f''(p_0)$ are > 0 (i.e. if $f''(p_0)$ is positive definite) then p_0 is a strict local minimum of f.*
>
> (b) *If all the eigenvalues of $f''(p_0)$ are < 0 (i.e. if $f''(p_0)$ is negative definite) then p_0 is a strict local maximum of f.*