The Second Derivative Test for Maxima and Minima

Calculus of One Variable

In calculus of one variable, if a smooth function $y = f(x)$ has a critical point at x_0, so $f'(x_0) = 0$, one often uses the second derivative to test if this critical point is a local maximum or minimum. To understand this, one uses a Taylor polynomial centered at $x = x_0$ for $f(x)$, along with the fact that $f'(x_0) = 0$:

$$\begin{aligned} f(x) &= f(x_0) + f'(x_0)(x - x_0) + \tfrac{1}{2} f''(x_0)(x - x_0)^2 + \cdots \text{ higher order terms} \qquad (1) \\ &= f(x_0) + \tfrac{1}{2} f''(x_0)(x - x_0)^2 + \cdots \text{ higher order terms.} \qquad (2) \end{aligned}$$

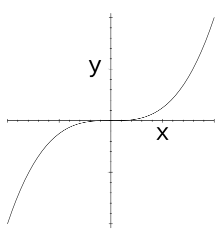
From this one sees that if $f''(x_0) > 0$, then $f(x) > f(x_0)$ for all x sufficiently near x_0 $(x \neq x_0)$. To make this clearer, one can replace (1) by a similar result using The error term in Taylor polynomials:

$$f(x) = f(x_0) + \tfrac{1}{2} f''(z)(x - x_0)^2, \qquad (3)$$

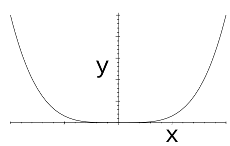
for some value of z between x and x_0. If $f''(x)$ is continuous and positive at x_0, and if x is sufficiently near x_0, then $f''(z) > 0$. It is now obvious from (3) that f has a *strict local minimum at* x_0, that is, $f(x) > f(x_0)$ for all x sufficiently near x_0 $(x \neq x_0)$.

Similarly, if $f''(x_0) < 0$, then f has a strict local minimum at x_0.

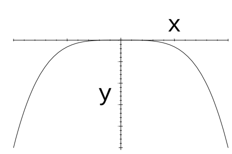
If $f''(x_0) = 0$, then this test fails and one must use higher order terms in the Taylor polynomial. The following graphs illustrate this. For each of these the origin is the only critical point and the second derivative there is zero.



$y = x^3$



$y = x^4$



$y = -x^4$

Calculus of Several Variables

The story for a function of several variables $z = f(\mathbf{X}) = f(x_1, \ldots, x_n)$, is similar. Indeed, we will reduce it to the one variable case. Say we have a function $f(\mathbf{X})$ which we may imagine to be the temperature at $\mathbf{X}$. If $\mathbf{X} = \mathbf{P} = (p_1, \ldots, p_n)$ is a critical point for f then the first derivative test for critical points tells us that $\nabla f(\mathbf{P}) = 0$. To see this, let $\mathbf{V} = (v_1, v_2, \ldots, v_n)$ be a unit vector at $\mathbf{P}$ and consider the function of *one* variable

$$\varphi(t) = f(\mathbf{P} + t\mathbf{V}),$$

which we may thing of as the temperature at the point $\mathbf{P} + t\mathbf{V}$ on the straight line through $\mathbf{P}$ in the direction of the vector $\mathbf{V}$. If f has a local minimum at $\mathbf{P}$, then $\varphi(t)$ has a local minimum at $t = 0$ for every vector $\mathbf{V}$. Thus $\varphi'(t)|_{t=0} = 0$. But using the chain rule we have

$$\begin{aligned}
\varphi'(t) &= \frac{df(\mathbf{P}+t\mathbf{V})}{dt} = \frac{df(p_1+tv_1,\,\ldots,p_n+tv_n)}{dt} \\
&= \frac{\partial f(\mathbf{P}+t\mathbf{V})}{\partial x_1}v_1 + \cdots + \frac{\partial f(\mathbf{P}+t\mathbf{V})}{\partial x_n}v_n \qquad (4)\\
&= \nabla f(\mathbf{P}+t\mathbf{V})\cdot\mathbf{V}.
\end{aligned}$$

This is of course the usual formula for the directional derivative of f in the direction of the vector $\mathbf{V}$. In particular, at $t=0$ we have

$$0 = \varphi'(0) = \nabla f(\mathbf{P})\cdot\mathbf{V}.$$

Since this holds for all possible vectors $\mathbf{V}$ we conclude that $\nabla f(\mathbf{P}) = 0$, as claimed.

To obtain the second derivative test we again use the chain rule and compute $\varphi''(t)$. For simplicity, assume $n=2$; the general case is essentially identical. From the first term in (4) we find

$$\begin{aligned}
\left(\frac{d}{dt}\right)\left(\frac{\partial f(\mathbf{P}+t\mathbf{V})}{\partial x_1}v_1\right) &= \frac{\partial}{\partial x_1}\left(\frac{\partial f(\mathbf{P}+t\mathbf{V})}{\partial x_1}\right)v_1^2 + \frac{\partial}{\partial x_1}\left(\frac{\partial f(\mathbf{P}+t\mathbf{V})}{\partial x_2}\right)v_1v_2,\\
&= f_{11}(\mathbf{P}+t\mathbf{V})v_1^2 + f_{12}(\mathbf{P}+t\mathbf{V})v_1v_2, \qquad (5)
\end{aligned}$$

where we have used the standard notation $f_{11} = \partial^2 f/\partial x_1^2$, etc. There is a similar formula for $(d/dt)(\partial f(\mathbf{P}+t\mathbf{V})/\partial x_2)$. Thus at $t=0$, after we collect terms and use the fact that for a smooth function $f_{12} = f_{21}$, we conclude with the formula

$$\varphi''(0) = f_{11}(\mathbf{P})v_1^2 + 2f_{12}(\mathbf{P}+t\mathbf{V})v_1v_2 + f_{22}(\mathbf{P})v_2^2. \qquad (6)$$

This is a quadratic polynomial in the components of the vector $\mathbf{V}$ and suggests that we write it using the notation developed in the previous section for quadratic polynomials in several variables. Thus we introduce the *second derivative matrix* (this is sometimes called the *hessian matrix*),

$$f''(\mathbf{P}) = \begin{pmatrix} f_{11}(\mathbf{P}) & f_{12}(\mathbf{P}) \\ f_{21}(\mathbf{P}) & f_{22}(\mathbf{P}) \end{pmatrix}.$$

This gives the cleaner expression

$$\varphi''(0) = \mathbf{V}\cdot f''(\mathbf{P})\mathbf{V}. \qquad (7)$$

Armed with this we can write the Taylor polynomial for φ about $t=0$

$$\begin{aligned}
f(\mathbf{P}+t\mathbf{V}) &= \varphi(t) = \varphi(0) + \varphi'(0)t + \tfrac{1}{2}\varphi''(0)t^2 + \text{ higher order terms} \qquad (8)\\
&= f(\mathbf{P}) + \tfrac{1}{2}[\mathbf{V}\cdot f''(\mathbf{P})\mathbf{V}]t^2 + \text{ higher order terms.} \qquad (9)
\end{aligned}$$

It should be clear from this that if $\mathbf{V}\cdot f''(\mathbf{P})\mathbf{V} > 0$ for all directions $\mathbf{V}(\neq 0)$ then for all $\mathbf{X} = \mathbf{P}+t\mathbf{V}$ near $\mathbf{P}$, that is, for t sufficiently small, we have $f(\mathbf{X}) > f(\mathbf{P})$. Thus f has a strict local minimum at $\mathbf{P}$. Similarly if $\mathbf{V}\cdot f''(\mathbf{P})\mathbf{V} < 0$ for all directions $\mathbf{V}(\neq 0)$ then f has a strict local maximum at $\mathbf{P}$, while if $\mathbf{V}\cdot f''(\mathbf{P})\mathbf{V}$ assumes both positive and negative values for various vectors $\mathbf{V}$, then f has a saddle point at $\mathbf{P}$.

We can summarize this using the language of the preceding section.

SECOND DERIVATIVE TEST *Say a smooth function f has a critical point at* $\mathbf{P}$

if	$f''(\mathbf{P})$	is positive definite	then f	has a strict local minimum at $\mathbf{P}$
if	$f''(\mathbf{P})$	is negative definite	then f	has a strict local maximum at $\mathbf{P}$
if	$f''(\mathbf{P})$	is indefinite	then f	has saddle point at $\mathbf{P}$
if	$f''(\mathbf{P})$	is not invertible	then f	no conclusion unless $f''(\mathbf{P})$ is indefinite

The last item above is because if $f''(\mathbf{P})$ is not invertible, then the matrix $f''(\mathbf{P})$ can be neither positive definite nor negative definite (see OBSERVATION 4 in the preceding section). This is analogous to the situation for a function of one variable when the second derivative is zero. In Examples 3 and 4 below we will see examples where in these circumstances the behavior is uncertain.

EXAMPLE 1. Find and classify all the critical points of $f(x,y) = x^4 + y^4 - 4xy + 1$.

SOLUTION. We first find the critical points by solving

$$0 = f_x = 4x^3 - 4y \qquad \text{and} \qquad 0 = f_y = 4y^3 - 4x.$$

We use the first equation $y = x^3$ in the second to find $(x^3)^3 - x = 0$, that is,

$$0 = x^9 - x = x(x^8 - 1) = x(x^4 - 1)(x^4 + 1) = x(x^2 - 1)(x^2 + 1)(x^4 + 1).$$

Thus there are three real roots, $x = 0, 1, -1$. Using $y = x^3$ again we find there are three critical points: $\mathbf{P}_1 = (0, 0)$, $\mathbf{P}_2 = (1, 1)$, and $\mathbf{P}_3 = (-1, -1)$.

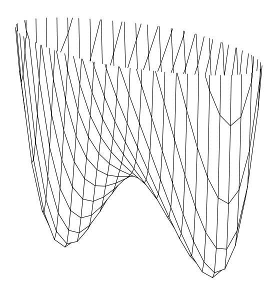
We next calculate the second derivative matrix.

$$f''(\mathbf{P}) = \begin{pmatrix} 12x^2 & -4 \\ -4 & 12y^2 \end{pmatrix},$$

so that at $\mathbf{P}_1$, $\mathbf{P}_2$, $\mathbf{P}_3$

$$f''(\mathbf{P}_1) = \begin{pmatrix} 0 & -4 \\ -4 & 0 \end{pmatrix}, \qquad f''(\mathbf{P}_2) = \begin{pmatrix} 12 & -4 \\ -4 & 12 \end{pmatrix}, \quad \text{and} \quad f''(\mathbf{P}_3) = \begin{pmatrix} 12 & -4 \\ -4 & 12 \end{pmatrix}.$$

Using the determinant test of the previous section, we see that $f''(\mathbf{P}_1)$ is indefinite, while both $f''(\mathbf{P}_2)$ and $f''(\mathbf{P}_3)$ are positive definite. Consequently, f has a saddle point at $\mathbf{P}_1$ and strict local minima at $\mathbf{P}_2$ and $\mathbf{P}_3$. The graph illustrates this clearly.



EXAMPLE 2. Find and classify all the critical points of

$$h(x,y) = \frac{3x^4 + 4x^3 - 12x^2 + 6}{12(1 + y^2)}.$$

SOLUTION. To find the critical points we must solve

$$0 = h_x = \frac{x^3 + x^2 - 2x}{1 + y^2} \quad and \quad 0 = h_y = \frac{-y(3x^4 + 4x^3 - 12x^2 + 6)}{6(1 + y^2)^2}.$$

Since $x^3 + x^2 - 2x = x(x^2 + x - 2) = x(x-1)(x+2)$, the first equation is satisfied only if $x = 0, 1$, or -2. Since none of these satisfy $3x^4 + 4x^3 - 12x^2 + 6 = 0$, the only way the second equation, $h_y = 0$, is satisfied is if $y = 0$. We thus obtain three critical points $\mathbf{P}_1 = (0, 0)$, $\mathbf{P}_2 = (1, 0)$, and $\mathbf{P}_3 = (-2, 0)$.

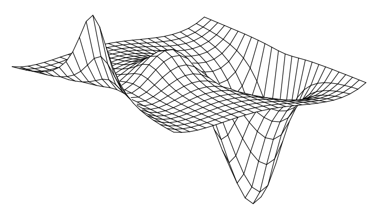
To classify these critical points we need the second derivative matrix. It is

$$f''(\mathbf{P}) = \begin{pmatrix} \frac{3x^2+2x-2}{1+y^2} & \frac{2y(x^3+x^2-2x)}{(1+y^2)^2} \\ \frac{2y(x^3+x^2-2x)}{(1+y^2)^2} & \frac{-(1-3y^2)(3x^4+4x^3-12x^2+6)}{6(1+y^2)^3}. \end{pmatrix}.$$

Thus

$$f''(\mathbf{P}_1) = \begin{pmatrix} -2 & 0 \\ 0 & -1 \end{pmatrix}, \qquad f''(\mathbf{P}_2) = \begin{pmatrix} 3 & 0 \\ 0 & -\frac{1}{6} \end{pmatrix}, \quad \text{and} \quad f''(\mathbf{P}_3) = \begin{pmatrix} 6 & 0 \\ 0 & \frac{13}{3} \end{pmatrix}.$$

Using OBSERVATIONS 1 and 3 of the preceding section (or the determinant test) we see that $f''(\mathbf{P}_1)$ is negative definite, $f''(\mathbf{P}_2)$ is indefinite, while $f''(\mathbf{P}_3)$ is positive definite. Thus $\mathbf{P}_1$ is a strict local maximum, $\mathbf{P}_2$ a saddle point, and $\mathbf{P}_3$ a strict local minimum. The graph clarifies this.



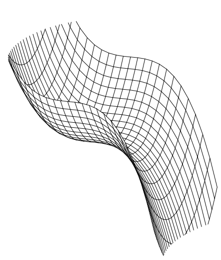
EXAMPLE 3. The functions

$$f(x, y) = x^2 - y^3 \qquad g(x, y) = x^2 + y^4 \qquad h(x, y) = x^2 - y^4$$

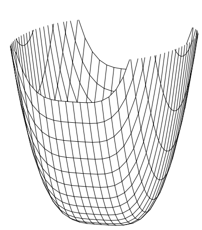
each have only one critical point, located at the origin. For each of them the second derivative matrix there is

$$\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}.$$

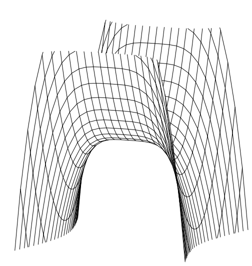
This matrix is not invertible and is not indefinite, so the second derivative test gives no information. To analyze the nature of these critical points one must either investigate the higher order terms in the Taylor polynomial at the origin, or else look at a graph. Either way it is clear that the origin is a more exotic saddle point for the first example, a strict local minimum for the second, and a saddle for the third.



$z = x^2 - y^3$

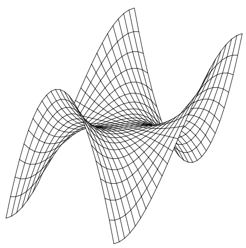


$z = x^2 + y^4$

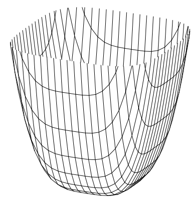


$z = x^2 + y^4$

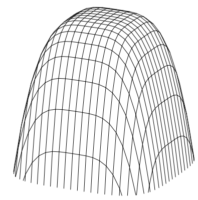
EXAMPLE 4. The examples $p(x,y) = x^3 - 3xy^2$ (a *monkey saddle*, since there is also a place for its tail), $q(x,y) = x^4 + y^4$, and $r(x,y) = -x^4 - y^4$, all of which have the origin as their only critical point and have second derivative matrix 0, further show that if at a critical point the second derivative matrix is not invertible, the nature of the critical point is determined only by a more thorough study.



$z = x^3 - 3xy^2$



$z = x^4 + y^4$



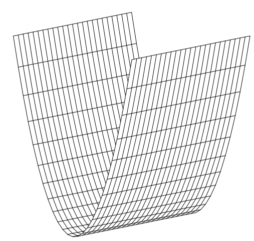
$z = -x^4 - y^4$

EXAMPLE 5. The function $F(x,y) = x^2 + 2xy + y^2$ has critical point where both

$$0 = F_x = 2x + 2y \quad \text{and} \quad 0 = F_y = 2x + 2y.$$

that is, along the whole line $y = -x$. The second derivative matrix there is

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$



which is not invertible, so the second derivative test fails. However, either from the observation that we can write $F(x,y) = (x+y)^2$ or else from the graphs, we conclude that these critical points are all local minima – but not strict local minima.

Problems

1. Find and classify the critical points of the function in Problem 2 of the preceding section.

2. Find and classify the critical points of the following functions – even if the second derivative test is not applicable.

(a) $f(x_1,x_2) = x_1^2 + 4x_1 + 2x_2^2 + 10$

(b) $\varphi(x,y) = 3 - 2x + 2y + x^2y^2$

(c) $f(x,y) = xy - x + y - 2$

(d) $u(x_1,x_2) = x_1^2 - x_2^2$

(e) $g(x,y) = y^3 - 2x^2 - 2y^2 + y$

(f) $v(x,y) = (x^2+y^2)^2 - 8y^2$

(g) $\psi(x_1,x_2) = \dfrac{\cos x_2}{1+x_1^2}$

(h) $h(x,y) = \dfrac{\cosh(2x)}{1+2y^2}$

(i) $w(x,y) = x^2 - 2xy + \frac{1}{3}y^3 - 3y$

(j) $\psi(y,z) = \dfrac{z^3 - 3z}{1+y^2}$

(k) $k(x_1,x_2) = (x_1^2 - 1)e^{x_2}$

(l) $g(x,y,z) = \dfrac{xye^y}{1+x^2} + z^2 + 2z$

(m) $h(x,y,z) = \dfrac{3x^4 - 6x^2 + 1}{1+z^2} - y^2$

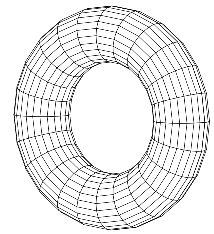
(n) $\varphi(x,y,z) = x^3 - 3x + y^2 + z^2$

(o) $r(x,y,z) = (1 + 2x + 3y - z)^2$.

3. Find and classify the critical points of the function

$$z = z(\phi,\, \theta) = (2 + \cos\phi)\sin\theta, \qquad 0 \le \theta, \phi < 2\pi.$$

This is the z coordinate of a torus standing on its edge; its other coordinates are $y = (2 + \cos\phi)\cos\theta$ and $x = \sin\phi$ (see fig).



4. Find and classify the critical points of the following functions. For the first two, also draw graphs showing the critical points.
   (a) $h(x, y) = (x^2 + 2y^2)e^{1-(x^2+y^2)}$.
   (b) $k(x, y) = (x^2 - 2y^2)e^{1-(x^2+y^2)}$.
   (c) $q(x, y, z) = (x^2 + 2y^2 + 3z^2)e^{1-(x^2+y^2+z^2)}$

5. Find and classify the critical points of both of the functions
   (a) $q(x, y, r, s) = 5x^2 + 2xy - 6xs + 5y^2 - 6ys - 9s^2 - 18xr - 18yr + 3r^2 + 36rs$,
   (b) $h(x, y, r, s) = 5x^2 + 2xy - 6xs + 5y^2 - 6ys + 9s^2 - 18xr - 18yr + 3r^2 + 36rs$.

6. This exercises concerns three games.
   (a) Susan picks any real number x that she wants. Then, Gwen, after hearing Susan's choice, picks a real number y of her choice and pays Susan $xy - 3x + 2y$ dollars. What number should Susan pick? (A negative value of $xy - 3x + 2y$ indicates a payment from Susan to Gwen.) A graph of the "payoff function" $xy - 3x + 2y$ may help understanding.
   (b) Discuss the modified game in which Gwen chooses first and, then, Susan makes her choice after hearing Gwen's choice.
   (c) Discuss the modification in which each player writes her choice secretly on a piece of paper, before hearing the choice of the other player.
   (d) Use Maple to simulate these games, with the computer playing the role of Susan and you playing the role of Gwen. Have the computer make random choices and you make whatever choice you find appropriate. Tabulate the "payoff" to understand which strategies are optimal.

7. Let A be a symmetric matrix and let f be the quadratic polynomial $f(\mathbf{X}) = \mathbf{X} \cdot A\mathbf{X}$.
   (a) Show that $\operatorname{grad} f(\mathbf{X}) = A\mathbf{X}$ and that $f''(\mathbf{X}) = 2A$. Note the similarity between this and the special case of a function of one variable case $f(x) = ax^2$ where $f''(x) = 2a$.
   (b) If A is invertible, show that the only critical point of f is at the origin, and that this is a local minimum if and only if A is positive definite.
   (c) Extend parts a) and b) to the more general case where the quadratic polynomial has lower order terms, $f(\mathbf{X}) = \mathbf{X} \cdot A\mathbf{X} + 2\mathbf{b} \cdot \mathbf{X} + c$, where $\mathbf{b}$ is a vector and c a scalar. As a check, compare your result with that obtained in the special case of one variable, $f(x) = ax^2 + 2bx + c$.

8. Least Squares. Let A be a matrix, not necessarily square, b a given vector, and let f be the quadratic polynomial $f(\mathbf{X}) = \|A\mathbf{X} - \mathbf{b}\|^2$. The problem is to minimize f. If A is invertible, just let $\mathbf{X}$ be the solution of $A\mathbf{X} = \mathbf{b}$. The difficulty is if A is not invertible, say because there are more equations than unknowns. Then $f(\mathbf{X})$ measures the discrepancy

between $A\mathbf{X}$ and $\mathbf{b}$. The optimal $\mathbf{X}$ is the value that minimizes this discrepancy. This problem arises, for instance, when one tried to fit experimental data to a straight line that one does not expect to pass through all of the points.

(a) Show that $\text{grad}\, f(\mathbf{X}) = 2A^T(A\mathbf{X} - \mathbf{b})$.

(b) Compute the second derivative matrix, f'', and show that if the only solution of the homogeneous equation $A\mathbf{Z} = 0$ is $Z = 0$, then f has exactly one critical point and it is a local minimum.

(c) Use this to find the "optimal" solution of the following set of four equations in two unknowns:

$$\begin{aligned} 2x + 3y &= 1 \\ 3x + 3y &= 2 \\ 2x + 2y &= 0 \\ x &= -1 \end{aligned}$$

9. If $f(x, y)$ has a local minimum at the point (a, b), show that $f''(a, b)$ must be positive definite or semi-definite. [Suggestion: First state and prove the version for a function of one variable].

10. If $u(x, y)$ has the property that $2u_{xx} + 3u_{yy} = -1$, show that u cannot have a local minimum *anywhere*.

11. If the function $f(x, y)$ has the property that $f''(x, y)$ is positive definite or semi-definite at every point (x, y), show that the surface $z = f(x, y)$ lies above its tangent plane at every point, that is, the surface is *convex*. [Suggestion: First try the analogous assertion for a function of one variable.]

12. Let $k(x, y) = (y - x^2)(y - 2x^2)$.

(a) Show that the origin is the only critical point.

(b) Show that $k(x, y) > 0$ for all points (x, y) except those in the "horn" region $x^2 \le y \le 2x^2$, where $k(x, y) \le 0$. Thus the origin is a saddle point.

(c) On the other hand, note that moving from the origin along straight lines the function is positive for at least a short distance. Since $k(0, 0) = 0$, as one approaches the origin along straight lines, the origin appears to have a local maximum. The point is that because the second derivative matrix $k''(0, 0)$ is not invertible, the behavior of k near the origin is too complicated to be resolved by just considering the behavior of k along straight through the origin. [However, had $k''(0, 0)$ been invertible, then straight lines would have been adequate].

13. Let A be an invertible symmetric matrix and let $f(\mathbf{X}) = (\mathbf{X} \cdot A\mathbf{X})\, e^{-\|X\|^2}$.

(a) Show that the critical points of f are exactly the origin and all the unit vectors that are eigenvectors of A.

(b) In addition assume that A is a positive definite $n \times n$ matrix with *distinct* eigenvalues $0 < \lambda_1 < \lambda_2 < \ldots \lambda_n$ and corresponding orthonormal eigenvectors $\mathbf{V}_1, \ldots, \mathbf{V}_n$. Show that f has a strict local maximum at the two points $\pm\mathbf{V}_n$, a strict local minimum at the origin, while the remaining $2(n-1)$ critical points are saddle points. [Suggestion: first understand the case when A is a diagonal matrix].