Second Order Partial Derivatives; the Hessian Matrix; Minima and Maxima

Second Order Partial Derivatives  We have seen that the partial derivatives of a differentiable function $\phi(X) = \phi(x_1, x_2, ..., x_n)$ are again functions of $n$ variables in their own right, denoted by

$$\frac{\partial \phi}{\partial x_k}(x_1, x_2, ..., x_n), \ k = 1, 2, ..., n.$$  

If these functions, in turn, remain differentiable, each one engenders a further set of $n$ functions, the second order partial derivatives of $\phi(x_1, x_2, ..., x_n)$:

$$\frac{\partial \phi}{\partial x_k}(x_1, x_2, ..., x_n) \to \frac{\partial^2 \phi}{\partial x_j \partial x_k}(x_1, x_2, ..., x_n), \ j = 1, 2, ..., n.$$  

The function $\phi(x_1, x_2, ..., x_n)$ is said to be twice continuously differentiable in a region $\mathcal{D}$ if each of these second order partial derivative functions is, in fact, a continuous function in $\mathcal{D}$. The partial derivatives $\frac{\partial^2 \phi}{\partial x_j \partial x_k}$ for which $j \neq k$ are called mixed partial derivatives. For them we have a very important theorem, proved in 1734 by Leonhard Euler.

**Theorem 1 (Equality of Mixed Partial Derivatives)**  If $\phi(X) = \phi(x_1, x_2, ..., x_n)$ is continuously differentiable in a region $\mathcal{D}$ then, in that region,

$$\frac{\partial^2 \phi}{\partial x_j \partial x_k}(x_1, x_2, ..., x_n) \equiv \frac{\partial^2 \phi}{\partial x_k \partial x_j}(x_1, x_2, ..., x_n).$$

**Proof**  We will give the proof only for the case $n = 2$; the proof for $n > 2$ is similar but a little more complicated. For $n = 2$ we can replace $x_1$ by $x$, $x_2$ by $y$. Clearly there is nothing to prove for the “unmixed” partial derivatives $\frac{\partial^2 \phi}{\partial x^2}, \frac{\partial^2 \phi}{\partial y^2}$.  

With \( X_0 = (x_0, y_0) \in \mathcal{D} \) and \( \Delta x \neq 0, \Delta y \neq 0 \), we form the double difference

\[
S (X_0, \Delta x, \Delta y) = \phi (x_0 + \Delta x, y_0 + \Delta y) - \phi (x_0 + \Delta x, y_0) - (\phi (x_0, y_0 + \Delta y) - \phi (x_0, y_0)).
\]

Let \( g(x) = \phi (x, y_0 + \Delta y) - \phi (x, y_0) \). Then

\[
S (X_0, \Delta x, \Delta y) = g(x_0 + \Delta x) - g(x_0).
\]

The differentiability of \( \phi \) implies that of \( g \). Applying the mean value theorem of elementary calculus, there is a value \( \xi \) between \( x_0 \) and \( x_0 + \Delta x \) such that

\[
g(x_0 + \Delta x) - g(x_0) = \frac{dg}{dx} (\xi) \Delta x, \text{ i.e.,}
\]

\[
S (X_0, \Delta x, \Delta y) = \left( \frac{\partial \phi}{\partial x} (\xi, y_0 + \Delta y) - \frac{\partial \phi}{\partial x} (\xi, y_0) \right) \Delta x.
\]

Applying the mean value theorem again, there is a value \( \eta \) between \( y_0 \) and \( y_0 + \Delta y \) such that

\[
S (X_0, \Delta x, \Delta y) = \frac{\partial^2 \phi}{\partial y \partial x} (\xi, \eta) \Delta x \Delta y.
\]

Since the point \( \xi, \eta \) must tend to \( (x_0, y_0) \) as \( \Delta x \) and \( \Delta y \) both tend to zero, the continuity of \( \frac{\partial^2 \phi}{\partial y \partial x} \) implies that

\[
\frac{\partial^2 \phi}{\partial y \partial x} (x_0, y_0) = \lim_{\Delta x, \Delta y \to 0} \frac{S (X_0, \Delta x, \Delta y)}{\Delta x \Delta y}.
\]

Starting with \( h(y) = \phi (x_0 + \Delta x, y) - \phi (x_0, y) \) and reversing the order of the above argument, we arrive in much the same way at

\[
\frac{\partial^2 \phi}{\partial x \partial y} (x_0, y_0) = \lim_{\Delta x, \Delta y \to 0} \frac{S (X_0, \Delta x, \Delta y)}{\Delta x \Delta y}.
\]

Since \( X_0 \) is an arbitrary point in \( \mathcal{D} \), the theorem is proved.
Analyzing Stationary Points  

Suppose \( \phi(x, y) \) is twice continuously differentiable and \( X_0 = (x_0, y_0)^* \) is a stationary point for the function. In the section on \textbf{minima and maxima and the gradient method} we began to explore the use of the second order partial derivatives at the stationary point \( X_0 \) as a tool for determining whether \( X_0 \) is a maximum, a minimum, or neither. It is easy to see, in this two dimensional context, that evaluation of \( \frac{\partial^2 \phi}{\partial x^2} \) and \( \frac{\partial^2 \phi}{\partial y^2} \) will not always be decisive this way. The function \( \phi(x, y) = x^2 + 4xy + y^2 \) is easily seen to have a critical point at the origin, \( x = y = 0 \), and the second order partial derivatives \( \frac{\partial^2 \phi}{\partial x^2} \) and \( \frac{\partial^2 \phi}{\partial y^2} \) there are both equal to 2. But \((0, 0)\) is not a minimum because on the diagonal line \( x = t, \ y = -t \) we have \( \phi(t, -t) = 2t^2 - 4t^2 = -2t^2 \), so that \( \phi \) decreases as the point \((x, y)\) moves away from the origin along this line. We need a more systematic analysis to assist us in classifying stationary points.

Such an analysis can be developed in a general context in \( \mathbb{R}^n \) if we assume the function \( \phi(X) \) is twice continuously differentiable; the first and second order partial derivatives are all defined and continuous throughout the region \( \mathcal{D} \) of interest. When this is the case we can define the \textbf{Hessian matrix}

\[
H_\phi(X) = \begin{pmatrix}
\frac{\partial^2 \phi}{\partial x_1^2}(X) & \frac{\partial^2 \phi}{\partial x_1 \partial x_2}(X) & \cdots & \frac{\partial^2 \phi}{\partial x_1 \partial x_n}(X) \\
\frac{\partial^2 \phi}{\partial x_1 \partial x_2}(X) & \frac{\partial^2 \phi}{\partial x_2^2}(X) & \cdots & \frac{\partial^2 \phi}{\partial x_2 \partial x_n}(X) \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial^2 \phi}{\partial x_1 \partial x_n}(X) & \frac{\partial^2 \phi}{\partial x_2 \partial x_n}(X) & \cdots & \frac{\partial^2 \phi}{\partial x_n^2}(X)
\end{pmatrix}
\]

This matrix is always \textit{symmetric}, i.e., \( H_\phi(X)^* = H_\phi(X) \), as a consequence of the equality of mixed second order partial derivatives proved above. In terms of vector differential operators already defined we have

\[
H_\phi(X) = \nabla(\nabla \phi)^*(X);
\]

one forms the (column) gradient vector field \((\nabla \phi)^*(X)\) and then the
Jacobian matrix of that vector field. For this reason we write

\[ H_\phi(X) \equiv \nabla^2 \phi(X). \]

**Proposition** Suppose \( \phi(X) \) is twice continuously differentiable in a region \( D \) which includes the point \( X_0 \). Then, for \( X \) near \( X_0 \) in \( D \),

\[
\phi(X) = \phi(X_0) + \nabla \phi(X_0) (X - X_0) \\
+ \frac{1}{2} (X - X_0)^* \nabla^2 \phi(X_0) (X - X_0) + o(\|X - X_0\|^2),
\]

where \( o(\|X - X_0\|^2) \) indicates the presence of a remainder function \( r(X) \) with the property

\[
\lim_{\|X - X_0\| \to 0} \frac{r(X)}{\|X - X_0\|^2} = 0.
\]

**Proof** Let \( U \) be an arbitrary unit vector, \( X(t) = X_0 + tU \). Then we set

\[ f(t) = \phi(X(t)). \]

For a twice continuously differentiable function \( f(t) \), *Taylor’s Formula*, with remainder taken after the second order term, says that, as \( t \to 0 \),

\[
f(t) = f(0) + f'(0) t + \frac{1}{2} f''(0) t^2 + o(|t^2|).
\]

Here we have \( f(0) = \phi(X_0) \) and

\[
f'(t) = \frac{d}{dt} \phi(X_0 + tU) = \nabla \phi(X_0 + tU) \frac{d}{dt} (X_0 + tU)
\]

\[ = \nabla \phi(X_0 + tU) U \]

and thus

\[ f'(0) = \nabla \phi(X_0) U. \]
Continuing, we have

\[ f''(t) = \frac{d}{dt} f'(t) = \frac{d}{dt} (\nabla \phi (X_0 + tU) U). \]

Since for real column vectors \( Z^* W = W^* Z \), this is the same as

\[ \frac{d}{dt} (U^* \nabla \phi (X_0 + tU)^*) = U^* \frac{d}{dt} (\nabla \phi (X_0 + tU)^*) \]

\[ = U^* \nabla \nabla^* \phi (X_0 + tU) U. \]

Evaluating at \( t = 0 \) we have

\[ f''(0) = U^* \nabla \nabla^* \phi (X_0) U = U^* \nabla^2 \phi (X_0) U. \]

Thus we have, as \( t \to 0 \),

\[ \phi(X(t)) = f(t) = \phi(X_0) + \nabla \phi(X_0) tU + \frac{1}{2} tU^* \nabla^2 \phi(X_0) tU + o(|t|^2). \]

Since \( X(t) = X_0 + tU \) we have \( \|X(t) - X_0\| = |t| \|U\| = |t| \). Since every choice of \( X \) corresponds to a choice of \( U \) via \( U = \frac{X - X_0}{\|X - X_0\|} \), every \( X \) near \( X_0 \) corresponds to \( X = X(t) = X_0 + tU \) for small \( |t| \). Replacing \( tU \) by \( X - X_0 \) and \( o(|t|^2) \) by \( o(\|X - X_0\|^2) \), we have, as \( \|X - X_0\| \to 0 \),

\[ \phi(X(t)) = \phi(X_0) + \nabla \phi(X_0) (X - X_0) \]

\[ + \frac{1}{2} (X - X_0)^* \nabla^2 \phi(X_0) (X - X_0) + o(\|X - X_0\|^2) \]

as claimed. This completes the proof.

The first three terms in the formula just obtained are referred to as the second order Taylor approximation to \( \phi(X) \) based on the point \( X_0 \).

We intend to put this result to use in analyzing stationary points. Before we can do that, however, we have to introduce some new concepts.
Definition Let $A$ be an $n \times n$ matrix and let $X \in \mathbb{R}^n$. The scalar valued function

$$q(X) = X^* A X = \sum_{j=1}^{n} \sum_{k=1}^{n} a_{jk} x_j x_k$$

is the quadratic form in $X$ associated with the matrix $A$.

Example 1 Let $A$ be the $3 \times 3$ matrix

$$A = \begin{pmatrix} 2 & -2 & 3 \\ 1 & 1 & 2 \\ 0 & -1 & 1 \end{pmatrix}.$$ 

Then, for $X = (x y z)^*$, the quadratic form in $X$ associated with $A$ is

$$q(X) = q(x, y, z) = 2x^2 - 2xy + 3xz + yx + y^2 + 2yz - zy + z^2$$

$$= 2x^2 + y^2 + z^2 - xy + 3xz + yz.$$ 

Thus a quadratic form in $X$ is a function which is a linear combination of products of components of $X$, taken two at a time. The matrix $A$ serves to provide the coefficients accompanying these products in forming $q(X)$.

Proposition For real vectors $X \in \mathbb{R}^n$ and real matrices $A$ we can assume without loss of generality, in forming quadratic forms $q(X) = X^* A X$, that $A$ is symmetric, i.e., $A^* = A$; $a_{jk} = a_{kj}$, $j, k = 1, 2, \ldots, n$.

Proof Since $q(X)$ is a real scalar, $q(X)^* = q(X)$ and therefore

$$q(X) = \frac{1}{2} (q(X)^* + q(X)) = \frac{1}{2} ((X^* A X)^* + X^* A X)$$

$$= \frac{1}{2} (X^* A^* X + X^* A X) = X^* \left( \frac{1}{2} (A^* + A) \right) X \equiv X^* \tilde{A} X.$$ 

Since we readily see that $(A^* + A)^* = A + A^* = A^* + A$, the matrix $\tilde{A}$ is symmetric and the result follows.
Example 1, Continued  For the $3 \times 3$ matrix $A$ shown earlier and the associated quadratic form $q(X)$ we also have

$$q(X) = X^* \tilde{A} X, \quad \tilde{A} = \begin{pmatrix} 2 & -\frac{1}{2} & \frac{3}{2} \\ -\frac{1}{2} & 1 & \frac{1}{2} \\ \frac{3}{2} & \frac{1}{2} & 1 \end{pmatrix}. $$

From this point on, when discussing quadratic forms, we will assume $A$ is symmetric unless specifically indicated to the contrary.

**Definition**  The quadratic form $q(X) = X^* A \ X$ is **positive (definite)** if and only if $X \neq 0 \Rightarrow q(X) > 0$. It is **non-negative** if $q(X) \geq 0$ for all $X$. The quadratic form is **negative (definite)** if $-q(X)$ is positive definite; **non-positive** if $-q(X)$ is non-negative. If none of these are true then $q(X)$ is **indefinite**. One commonly refers to the matrix $A$ as being positive, non-negative, negative, non-positive or indefinite according as the quadratic form $q(X) = X^* A \ X$ has the property in question.

**Example 2**  Consider the quadratic form in $X = (x, y)$:

$$q(X) = q(x, y) = (x \quad y) \begin{pmatrix} 1 & \alpha \\ \alpha & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = x^2 + 2\alpha xy + y^2.$$

If $|\alpha| < 1$ we can write

$$x^2 + 2\alpha xy + y^2 = x^2 + 2\alpha xy + \alpha^2 y^2 + (1 - \alpha^2)y^2 = (x + \alpha y)^2 + (1 - \alpha^2)y^2$$

and it is easy to see that $q(X)$ is positive. If $|\alpha| = 1$ the above becomes

$$q(X) = (x + \alpha y)^2 \geq 0$$

and we conclude $q(X)$ is non-negative but not positive. If $|\alpha| > 1$ we see that $q(X)$ is positive if $x \neq 0, \ y = 0$ but negative if $y \neq 0$ and $x = -\alpha y$, hence $q(X)$ is indefinite.
The quadratic form of greatest interest for the study of maxima and minima is the third term, $\frac{1}{2} (X - X_0)^* \nabla^2 \phi(X_0) (X - X_0)$ in the second order Taylor approximation developed earlier. When $\phi(X)$ is twice continuously differentiable in some region containing a stationary point $X_0$, so that $\nabla \phi(X_0) = 0$ and the first order term $\nabla \phi(X_0)(X - X_0)$ of the Taylor approximation vanishes, this quadratic form determines whether $X_0$ is a local minimum, a local maximum, or neither.

**Theorem** If $X_0$ is a stationary point for the twice continuously differentiable function $\phi(X)$ then:

i) If $\nabla^2 \phi(X_0)$ is positive, then $X_0$ is a local minimum for $\phi(X)$;

ii) If $\nabla^2 \phi(X_0)$ is negative, then $X_0$ is a local maximum for $\phi(X)$;

iii) If $\nabla^2 \phi(X_0)$ is indefinite, then $X_0$ is neither a local minimum nor a local maximum for $\phi(X)$;

iv) If $\nabla^2 \phi(X_0)$ is none of the above, hence just non-negative or non-positive, then the term $\frac{1}{2} (X - X_0)^* \nabla^2 \phi(X_0)(X - X_0)$ is not decisive in determining the character of the stationary point $X_0$.

**Sketch of Proof** We use the term “sketch” here because there are some details which should be added to what we give below to make the proof entirely rigorous; these details are not essential to understanding the concepts involved.

We select a unit vector $U$ and consider the straight line $X(t) = X_0 + tU$, which passes through $X_0$ when $t = 0$. With $f(t) = \phi(X(t))$ we have $f'(0) = \nabla \phi(X_0) U = 0$. It will then be familiar from the standard calculus that $f(t)$ has a local minimum at $t = 0$ if $f''(0) > 0$ and a local maximum at $t = 0$ if $f''(0) < 0$. The hypotheses of i) and ii) guarantee $f''(0) = U^* \nabla^2 \phi(X_0) U$ positive in the case of i) and negative in the case of ii), for any choice of the unit vector $U$. Thus
in the case of i), \( \phi(X) \) increases in every direction as \( X \) is displaced away from \( X_0 \), provided that displacement remains small. Similarly, in the case of ii), \( \phi(X) \) decreases in every direction as \( X \) is slightly displaced away from \( X_0 \). We conclude that \( \phi(X) \) has a local minimum at \( X_0 \) in the first case and a local maximum there in the second case.

If iii) applies, we find unit vectors \( U, V \) such that \( U^* \nabla^2 \phi(X_0) U > 0 \) and \( V^* \nabla^2 \phi(X_0) V < 0 \). We let \( X(t) = X_0 + tU, \, \Xi(t) = X_0 + tV \). We then see that \( f(t) = \phi(X(t)) \) has a local minimum at \( t = 0 \) while \( g(t) = \phi(\Xi(t)) \) has a local maximum there. Thus \( \phi(X) \) increases as \( X \) is slightly displaced away from \( X_0 \) in the \( U \) direction and decreases when \( X \) is slightly displaced away from \( X_0 \) in the \( V \) direction. We conclude that \( X_0 \) is neither a local maximum nor a local minimum in this case.

The situation described in iv) is readily illustrated with the example \( \phi(x, y) = x^2 + \alpha y^4 \). Regardless of the value of \( \alpha \), the Hessian at the origin is

\[
\nabla^2 \phi(0, 0) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},
\]

which is easily seen to be non-negative but not positive. For \( \alpha > 0 \) the function \( \phi(x, y) \) clearly has a minimum at \((0, 0)\); for \( \alpha < 0 \), on the other hand, \( \phi(x, 0) > 0 = \phi(0, 0) \) for \( x \neq 0 \) while \( \phi(0, y) < 0 = \phi(0, 0) \) for \( y \neq 0 \). We conclude that \((0, 0)\) is neither a maximum nor a minimum in this case.

All of this, for purposes of application, clearly begs the question: how can we tell, given a real, symmetric matrix \( A \), into which of the categories the quadratic form \( q(X) = X^* A X \) falls? We will indicate some tests below, without giving the proofs - they really are a topic for a course in linear algebra.
Tests for Classification of A  

If $A$ is a real, symmetric $n \times n$ matrix, and $q(X) = X^* A X$, then

i) $q(X)$ is positive if and only if all of the eigenvalues $\lambda$ of the matrix $A$ are positive;

ii) $q(X)$ is negative if and only if all of the eigenvalues $\lambda$ of the matrix $A$ are negative;

iii) $q(X)$ is non-negative if and only if all of the eigenvalues $\lambda$ of the matrix $A$ are non-negative;

iv) $q(X)$ is non-positive if and only if all of the eigenvalues $\lambda$ of the matrix $A$ are non-positive;

i) $q(X)$ is indefinite if some of the eigenvalues $\lambda$ of the matrix $A$ are positive and some are negative.

The eigenvalues of $A$ are the roots of the scalar $n$-th degree polynomial equation

$$\det (\lambda I - A) = 0,$$

where $I$ is the $n \times n$ identity matrix

$$I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

A further test for positivity (or negativity if we apply it to $-A$, is the following: the quadratic form $q(X) = X^* A X$ is positive if and only if $\det B > 0$ for all $m \times m$ matrices $B$, $0 < m \leq n$, formed in the following way. Distinct integers $k_1, k_2, \ldots, k_m$, listed in increasing order, are selected from the integers 1 through $n$. Then $B$ is formed from the entries of $A$ which lie in the rows and columns numbered
This test quickly becomes unwieldy and, indeed, unusable for even moderately large values of $n$; the eigenvalue test is preferable.

If $A$ has the form $A = C^*C$ for some $m \times n$ dimensional matrix then $q(X) = X^*AX$ is automatically non-negative because

$$ q(X) = X^*AX = X^*C^*CX = \|CX\|^2. $$

If, in addition, $m \geq n$ and there is no $X \neq 0$ such that $CX = 0$, then $q(X)$ is positive.

**Example 3** Consider the function $\phi(x, y) = \sin(x+y) + \cos(x-y)$. We compute

$$ \frac{\partial \phi}{\partial x}(x, y) = \cos(x+y) - \sin(x-y), \quad \frac{\partial \phi}{\partial y}(x, y) = \cos(x+y) + \sin(x-y). $$

It is easy to see that both of these partial derivatives vanish if and only if both $\cos(x+y)$ and $\sin(x-y)$ are zero, and this is true if and only if

$$ x + y = \frac{(2k+1)\pi}{2}, \quad x - y = j\pi $$

for some integers $k$ and $j$. Solving for $x$ and $y$ we have

$$ x = \frac{(2k+1)\pi}{4} + \frac{j\pi}{2}, \quad y = \frac{(2k+1)\pi}{4} - \frac{j\pi}{2}. $$

There are infinitely many stationary points. Taking $k = 1$, $j = 1$ we obtain one of these, namely

$$ x = \frac{3\pi}{4} + \frac{\pi}{2} = \frac{5\pi}{4}, \quad y = \frac{3\pi}{4} - \frac{\pi}{2} = \frac{\pi}{4}. $$

For general $x$ and $y$ the Hessian matrix is

$$ \nabla^2 \phi(x, y) = \begin{pmatrix} -\sin(x+y) - \cos(x-y) & -\sin(x+y) + \cos(x-y) \\ -\sin(x+y) + \cos(x-y) & -\sin(x+y) - \cos(x-y) \end{pmatrix}. $$
For the point we selected we have $x + y = \frac{3\pi}{2}$, $x - y = \pi$, so the Hessian matrix there is

$$
\nabla^2 \phi \left( \frac{5\pi}{4}, \frac{\pi}{4} \right) = \begin{pmatrix}
-\sin \left( \frac{3\pi}{2} \right) - \cos(\pi) & -\sin \left( \frac{3\pi}{2} \right) + \cos(\pi) \\
-\sin \left( \frac{3\pi}{2} \right) + \cos(\pi) & -\sin \left( \frac{3\pi}{2} \right) - \cos(\pi)
\end{pmatrix} = \begin{pmatrix}
2 & 0 \\
0 & 2
\end{pmatrix}.
$$

We readily conclude in this case that the Hessian matrix is positive and the point $x = \frac{5\pi}{2}$, $y = \frac{\pi}{4}$ is a local minimum for $\phi(x, y)$.

On the other hand, if we take $k = 1$ and $j = 2$, corresponding to the point $x = \frac{7\pi}{4}$, $y = -\frac{\pi}{4}$, we find in the same way that

$$
\nabla^2 \phi \left( \frac{7\pi}{4}, -\frac{\pi}{4} \right) = \begin{pmatrix}
0 & 2 \\
2 & 0
\end{pmatrix}.
$$

The eigenvalues of this matrix are the roots of

$$
\det \begin{pmatrix}
\lambda & -2 \\
-2 & \lambda
\end{pmatrix} = \lambda^2 - 4 = 0,
$$

which are $\pm 2$. The Hessian in this case is indefinite; we have neither a local maximum nor a local minimum.