LECTURE 11

Maxima and Minima

**Definition 11.1.** Let \( f : \mathbb{R}^n \to \mathbb{R} \) be a real-valued function of several variables. A point \( x_o \in \mathbb{R}^n \) is called a **local minimum** of \( f \) if there is a neighborhood \( U \) of \( x_o \) such that

\[
f(x) \geq f(x_o) \quad \text{for all } x \in U.
\]

A point \( x_o \in \mathbb{R}^n \) is called a **local maximum** of \( f \) if there is a neighborhood \( U \) of \( x_o \) such that

\[
f(x) \leq f(x_o) \quad \text{for all } x \in U.
\]

A point \( x_o \in \mathbb{R}^n \) is called a **local extremum** if it is either a local minimum or a local maximum.

**Definition 11.2.** Let \( f : \mathbb{R}^n \to \mathbb{R} \) be a real-valued function of several variables. A **critical point** of \( f \) is a point \( x_o \) where

\[
\nabla f(x_o) = 0.
\]

If a critical point is not also a local extremum then it is called a **saddle point**.

**Theorem 11.3.** Suppose \( U \) is an open subset of \( \mathbb{R}^n \), \( f : U \to \mathbb{R} \) is differentiable, and \( x_o \) is a local extremum. Then

\[
\nabla f(x_o) = 0
\]

i.e., \( x_o \) is a critical point of \( f \).

**Proof.** Suppose \( x_o \) is an extremum of \( f \). Then there exists a neighborhood \( N \) of \( x_o \) such that either

\[
f(x) \geq f(x_o) \quad \text{for all } x \in N
\]

or

\[
f(x) \leq f(x_o) \quad \text{for all } x \in NR.
\]

Let \( \sigma : I \subset \mathbb{R} \to \mathbb{R}^n \) be any smooth path such that \( \sigma(0) = x_o \). Since \( \sigma \) is in particular continuous, there must be a subinterval \( I_N \) of \( I \) containing 0 such that

\[
\sigma(t) \in N \quad \text{for all } t \in I_N.
\]

But then if we define

\[
h(t) = f(\sigma(t))
\]

we see that since \( \sigma(t) \) lies in \( N \) for all \( t \in I_N \), we must have either

\[
h(t) = f(\sigma(t)) \geq f(x_o) = h(0) \quad \text{for all } t \in I_N
\]

or

\[
h(t) = f(\sigma(t)) \leq f(x_o) = h(0) \quad \text{for all } t \in I_N.
\]

Thus, \( t = 0 \) is an extremum for the smooth function \( h(t) \). From Calc. I, we know that if \( h(t) \) is differentiable and has a critical point at \( t = 0 \), then necessarily

\[
0 = \frac{dh}{dt}(0).
\]
Employing the chain rule we get
\[
0 = \frac{df}{dt}(0)
= \nabla f \big|_{\sigma(0)} \cdot \frac{d\sigma}{dt}(0)
= \nabla f (x_0) \cdot \frac{d\sigma}{dt}(0)
\]
Thus, \( \nabla f (x_0) \) must be perpendicular to the tangent vector of \( \sigma \) at \( t = 0 \). But the curve \( \sigma \) is arbitrary (except for the fact that it passes through \( x_0 \)); hence, \( \nabla f (x_0) \) must be perpendicular the tangent vector of every curve through \( x_0 \). But the only vector perpendicular to every other vector is the zero vector. Thus, we must have
\[
\nabla f (x_0) = 0.
\]

This theorem gives a necessary condition for a point \( x_0 \in U \) to be a local extremum for a \( C^2 \) function on \( U \subset \mathbb{R}^n \). It is not sufficient however. To develop a sufficient condition, we shall look a little more closely at the Taylor expansion of a \( C^2 \) function in the neighborhood of a critical point.

In general, for points sufficiently close to \( x_0 \),
\[
f(x) \approx f(x_0) + \nabla f(x_0) \cdot (x - x_0) + \frac{1}{2} (x - x_0)^T \nabla \nabla f(x_0) (x - x_0)
\]
Now if \( x_0 \) is a critical point of \( f \) then \( \nabla f(x_0) = 0 \), so
\[
f(x) \approx f(x_0) + \frac{1}{2} (x - x_0)^T \nabla \nabla f(x_0) (x - x_0).
\]
Thus, to show that, for all points close to \( x_0 \), \( f(x) \) is always greater than \( f(x_0) \) it would suffice to show that
\[
\frac{1}{2} (x - x_0)^T \nabla \nabla f(x_0) (x - x_0) \geq 0 \quad \text{for all } x;
\]
and to show that, for all points close to \( x_0 \), \( f(x) \) is always less than \( f(x_0) \) it would suffice to show that
\[
\frac{1}{2} (x - x_0)^T \nabla \nabla f(x_0) (x - x_0) \leq 0 \quad \text{for all } x.
\]

Definition 11.4. An \( n \times n \) matrix \( A \) is called positive definite if
\[
v^T Av > 0
\]
for any non-zero \( n \)-dimensional column vector \( v \). An \( n \times n \) matrix \( A \) is called negative definite if
\[
v^T Av < 0
\]
for any non-zero \( n \)-dimensional column vector \( v \).

The observation made above can now be phrased more succinctly.

Theorem 11.5. Let \( f : U \subset \mathbb{R}^n \to \mathbb{R} \) be a \( C^2 \) function. If \( x_0 \) is a critical point of \( f \) and the Hessian matrix \( \nabla \nabla f(x_0) \) is positive definite, then \( x_0 \) is a local minimum. If \( x_0 \) is a critical point of \( f \) and the Hessian matrix \( \nabla \nabla f(x_0) \) is negative definite, then \( x_0 \) is a local maximum.

What remains is to develop some useful techniques for detecting when a given matrix is positive or negative definite.

Theorem 11.6. An \( n \times n \) matrix is positive (respectively, negative) definite if and only if all its eigenvalues are positive (respectively, negative).

For the particular case when \( n = 2 \) we have an easier test for positive and negative definiteness.
Lemma 11.7. Let \( A = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \) be a symmetric \( 2 \times 2 \) matrix with \( a, b, c \in \mathbb{R} \). Then \( A \) is positive definite if and only if

\[
a > 0 \quad \text{and} \quad ac - b^2 > 0 ,
\]

and \( A \) is negative definite if and only if

\[
a < 0 \quad \text{and} \quad ac - b^2 > 0 .
\]

Proof. Let \( v = (x, y) \) be an arbitrary non-zero vector. Then

\[
v^T Av = (x, y) \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = (x, y) \begin{pmatrix} ax + by \\ bx + cy \end{pmatrix} = ax^2 + 2bxy + cy^2 = a(x + \frac{b}{a} y)^2 + \left(c - \frac{b^2}{a}\right)y^2
\]

If \( A \) is positive definite, the expression on the right hand side must positive for all choices of \( x \) and \( y \). In particular, it must be positive when \( y = 0 \), which leads to

\[
0 < ax^2
\]

which implies that

\[
a > 0 .
\]

The right hand side must also be positive when \( x = -\frac{b}{a} y \); this leads to

\[
0 < \left(c - \frac{b^2}{a}\right)y^2
\]

which implies

\[
ac - b^2 > 0 .
\]

If \( A \) is negative definite, then the expression on the right hand side of (1) must be negative for all choices of \( x \) and \( y \) (not both zero). Choosing \( y = 0 \) leads to the condition

\[
0 > ax^2
\]

which implies that

\[
a < 0 .
\]

Setting \( x = -\frac{b}{a} y \) leads to the condition

\[
0 < \left(c - \frac{b^2}{a}\right)y^2
\]

Multiplying both sides of this inequality by \( \frac{a}{y^2} \) we get the following condition

\[
ac - b^2 > 0 .
\]

(Because \( a \) must be negative, when we multiply both sides of the inequality by \( \frac{a}{y^2} \) we must also reverse the direction of the inequality.)

We can now employ the above criterion to determine when the Hessian matrix of a function \( f : U \subset \mathbb{R}^2 \rightarrow \mathbb{R} \) is positive or negative definite. Namely,

\[
Hf(x_o) = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2}(x_o) & \frac{\partial^2 f}{\partial x \partial y}(x_o) \\ \frac{\partial^2 f}{\partial x \partial y}(x_o) & \frac{\partial^2 f}{\partial y^2}(x_o) \end{pmatrix}
\]
is positive definite if

\[(i) \quad \left( \frac{\partial^2 f}{\partial x^2} (x_o) \right) \left( \frac{\partial^2 f}{\partial y^2} (x_o) \right) \left( \frac{\partial^2 f}{\partial x \partial y} (x_o) \right)^2 > 0 \]

and negative definite if

\[(i') \quad \left( \frac{\partial^2 f}{\partial x^2} (x_o) \right) \left( \frac{\partial^2 f}{\partial y^2} (x_o) \right) \left( \frac{\partial^2 f}{\partial x \partial y} (x_o) \right)^2 > 0 \]

\[
\quad \frac{\partial^2 f}{\partial x^2} (x_o) < 0.
\]

Let me now summarize the second derivative test for the case when \(n = 2\).

\textbf{Theorem 11.8. (Second Derivative Test).} Let \(f\) be a \(C^2\) function on open subset \(U \subset \mathbb{R}^2\). If \(x_o \in U\) is a critical point of \(f\) and

\[(i) \quad \left( \frac{\partial^2 f}{\partial x^2} (x_o) \right) \left( \frac{\partial^2 f}{\partial y^2} (x_o) \right) \left( \frac{\partial^2 f}{\partial x \partial y} (x_o) \right)^2 > 0 \]

\[(ii) \quad \frac{\partial^2 f}{\partial x^2} (x_o) > 0 \]

then \(x_o\) is a local minimum. If \(x_o \in U\) is a critical point of \(f\) and

\[(i') \quad \left( \frac{\partial^2 f}{\partial x^2} (x_o) \right) \left( \frac{\partial^2 f}{\partial y^2} (x_o) \right) \left( \frac{\partial^2 f}{\partial x \partial y} (x_o) \right)^2 > 0 \]

\[(ii') \quad \frac{\partial^2 f}{\partial x^2} (x_o) < 0 \]

then \(x_o\) is a local maximum.