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 $\mathbf{x}_o \in \mathbb{R}^n$ is called a **local minimum** of f if there is a neighborhood U of $\mathbf{x}_o$ such that*

$$f(\mathbf{x}) \geq f(\mathbf{x}_o) \quad \text{for all } \mathbf{x} \in U.$$

*A point $\mathbf{x}_o \in \mathbb{R}^n$ is called a **local maxmum** of f if there is a neighborhood U of $\mathbf{x}_o$ such that*

$$f(\mathbf{x}) \leq f(\mathbf{x}_o) \quad \text{for all } \mathbf{x} \in U \ .$$

*A point $\mathbf{x}_o \in \mathbb{R}^n$ 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 $\mathbf{x}_o$ where*

$$\nabla f(\mathbf{x}_o) = \mathbf{0}.$$

*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 $\mathbf{x}_o$ is a local extremum. Then*

$$\nabla f(\mathbf{x}_o) = \mathbf{0}$$

i.e., $\mathbf{x}_o$ is a critical point of f.

Proof. Suppose $\mathbf{x}_o$ is an extremum of f. Then there exists a neighborhood N of $\mathbf{x}_o$ such that either

$$f(\mathbf{x}) \geq f(\mathbf{x}_o) \quad , \quad \text{for all } \mathbf{x} \in N$$

or

$$f(\mathbf{x}) \leq f(\mathbf{x}_o) \quad , \quad \text{for all } \mathbf{x} \in NR.$$

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

$$\sigma(t) \in N \quad , \quad \text{for all } t \in I_N \ .$$

But then if we define

$$h(t) \equiv f(\sigma(t))$$

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

$$h(t) = f(\sigma(t)) \geq f(\mathbf{x}_o) = h(0) \quad , \quad \text{for all } t \in I_N$$

or

$$h(t) = f(\sigma(t)) \leq f(\mathbf{x}_o) = h(0) \quad , \quad \text{for all } t \in I_N \quad .$$

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

$$0 = \frac{dh}{dt}(0) \quad .$$

Employing the chain rule we get

$$\begin{aligned} 0 &= \tfrac{dh}{dt}(0) \\ &= \nabla f \big|_{\sigma(0)} \cdot \tfrac{d\sigma}{dt}(0) \\ &= \nabla f(\mathbf{x}_o) \cdot \tfrac{d\sigma}{dt}(0) \end{aligned}$$

Thus, $\nabla f(\mathbf{x}_o)$ must be perpendicular to the tangent vector of σ at $t = 0$. But the curve σ is arbitrary (except for the fact that it passes through $\mathbf{x}_o$); hence, $\nabla f(\mathbf{x}_o)$ must be perpendicular the tangent vector of every curve through $\mathbf{x}_o$. But the only vector perpendicular to every other vector is the zero vector. Thus, we must have

$$\nabla f(\mathbf{x}_o) = \mathbf{0} \quad .$$

□

This theorem gives a necessary condition for a point $\mathbf{x}_o \in U$ 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 $\mathbf{x}_o$,

$$f(\mathbf{x}) \approx f(\mathbf{x}_o) + \nabla f(\mathbf{x}_o) \cdot (\mathbf{x} - \mathbf{x}_o) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_o)^T \mathbf{H}(\mathbf{x}_o)(\mathbf{x} - \mathbf{x}_o) \quad .$$

Now if $\mathbf{x}_o$ is a critical point of f then $\nabla f(\mathbf{x}_o) = \mathbf{0}$, so

$$f(\mathbf{x}) \approx f(\mathbf{x}_o) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_o)^T \mathbf{H}f(\mathbf{x}_o)(\mathbf{x} - \mathbf{x}_o) \quad .$$

Thus, to show that, for all points close to $\mathbf{x}_o$, $f(\mathbf{x})$ is always greater than $f(\mathbf{x}_o)$ it would suffice to show that

$$\frac{1}{2}(\mathbf{x} - \mathbf{x}_o)^T \mathbf{H}f(\mathbf{x}_o)(\mathbf{x} - \mathbf{x}_o) \geq 0 \quad \text{for all } \mathbf{x};$$

and to show that, for all points close to $\mathbf{x}_o$, $f(\mathbf{x})$ is always less than $f(\mathbf{x}_o)$ it would suffice to show that

$$\frac{1}{2}(\mathbf{x} - \mathbf{x}_o)^T \mathbf{H}f(\mathbf{x}_o)(\mathbf{x} - \mathbf{x}_o) \leq 0 \quad \text{for all } \mathbf{x}.$$

DEFINITION 11.4. *An $n \times n$ matrix $\mathbf{A}$ is called **positive definite** if*

$$\mathbf{v}^T \mathbf{A} \mathbf{v} > 0$$

*for any non-zero n-dimensional column vector $\mathbf{v}$. An $n \times n$ matrix $\mathbf{A}$ is called **negative definite** if*

$$\mathbf{v}^T \mathbf{A} \mathbf{v} < 0$$

for any non-zero n-dimensional column vector $\mathbf{v}$.

The observation made above can now be phrased more succinctly.

THEOREM 11.5. *Let $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}$ be a C^2 function. If $\mathbf{x}_o$ is a critical point of f and the Hessian matrix $\mathbf{H}f(\mathbf{x}_o)$ is positive definite, then $\mathbf{x}_o$ is a local minimum. If $\mathbf{x}_o$ is a critical point of f and the Hessian matrix $\mathbf{H}f(\mathbf{x}_o)$ is negative definite, then $\mathbf{x}_o$ 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* $\mathbf{A} = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$ *be a symmetric* 2×2 *matrix with* $a, b, c \in \mathbb{R}$. *Then* $\mathbf{A}$ *is positive definite if and only if*

$$a > 0 \quad and \quad ac - b^2 > 0 \quad ,$$

and $\mathbf{A}$ *is negative definite if and only if*

$$a < 0 \quad and \quad ac - b^2 > 0 \quad .$$

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

$$\begin{aligned}
\mathbf{v}^T \mathbf{A} \mathbf{v} &= (x, y) \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\
&= (x, y) \begin{pmatrix} ax + by \\ bx + cy \end{pmatrix} \\
&= ax^2 + 2bxy + cy^2 \\
&= ax^2 + 2bxy + \tfrac{b^2 y^2}{a} - \tfrac{b^2 y^2}{a} + cy^2 \\
&= a \left(x + \tfrac{by}{a} \right)^2 + \left(c - \tfrac{b^2}{a} \right) y^2
\end{aligned}$$

If $\mathbf{A}$ 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

$$a > 0 \quad .$$

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

$$ac - b^2 > 0 \quad .$$

If $\mathbf{A}$ 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

$$a < 0 \quad .$$

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

$$ac - b^2 > 0 \quad .$$

(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 $f : U \subset \mathbb{R}^2 \to \mathbb{R}$ is positive or negative definite. Namely,

$$\mathbf{H} f(\mathbf{x}_o) = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2}(\mathbf{x}_o) & \frac{\partial^2 f}{\partial x \partial y}(\mathbf{x}_o) \\ \frac{\partial^2 f}{\partial x \partial y}(\mathbf{x}_o) & \frac{\partial^2 f}{\partial y^2}(\mathbf{x}_o) \end{pmatrix}$$

is positive definite if

$$\begin{array}{ll} (i) & \left(\frac{\partial^2 f}{\partial x^2}(\mathbf{x}_o)\right)\left(\frac{\partial^2 f}{\partial y^2}(\mathbf{x}_o)\right) - \left(\frac{\partial^2 f}{\partial x \partial y}(\mathbf{x}_o)\right)^2 > 0 \\ (ii) & \frac{\partial^2 f}{\partial x^2}(\mathbf{x}_o) > 0 \quad . \end{array}$$

and negative definite if

$$\begin{array}{ll} (i') & \left(\frac{\partial^2 f}{\partial x^2}(\mathbf{x}_o)\right)\left(\frac{\partial^2 f}{\partial y^2}(\mathbf{x}_o)\right) - \left(\frac{\partial^2 f}{\partial x \partial y}(\mathbf{x}_o)\right)^2 > 0 \\ (ii') & \frac{\partial^2 f}{\partial x^2}(\mathbf{x}_o) < 0 \quad . \end{array}$$

Let me now summarize the second derivative test for the case when $n = 2$.

THEOREM 11.8. *(Second Derivative Test). Let f be a C^2 function on open subset $U \subset \mathbb{R}^2$. If $\mathbf{x}_o \in U$ is a critical point of f and*

$$\begin{array}{ll} (i) & \left(\frac{\partial^2 f}{\partial x^2}(\mathbf{x}_o)\right)\left(\frac{\partial^2 f}{\partial y^2}(\mathbf{x}_o)\right) - \left(\frac{\partial^2 f}{\partial x \partial y}(\mathbf{x}_o)\right)^2 > 0 \\ (ii) & \frac{\partial^2 f}{\partial x^2}(\mathbf{x}_o) > 0 \end{array}$$

then $\mathbf{x}_o$ is a local minimum. If $\mathbf{x}_o \in U$ is a critical point of f and

$$\begin{array}{ll} (i') & \left(\frac{\partial^2 f}{\partial x^2}(\mathbf{x}_o)\right)\left(\frac{\partial^2 f}{\partial y^2}(\mathbf{x}_o)\right) - \left(\frac{\partial^2 f}{\partial x \partial y}(\mathbf{x}_o)\right)^2 > 0 \\ (ii') & \frac{\partial^2 f}{\partial x^2}(\mathbf{x}_o) < 0 \end{array}$$

then $\mathbf{x}_o$ is a local maximum.