Lecture 9: Max-Min and Least Squares

1 Application of Partial Derivatives: Optimization Problem

1.1 Max-Min

Optimization problems refer to the tasks of finding the maximum or minimum point of a function. Here we focus on finding the maximum or minimum point of multivariable functions, such as $f(x, y)$.

Another concept is local max/min point, which means the function value of this point is greater/less than the function values of adjacent points, but it is not necessarily the greatest or leaast function value in the function domain.

Theorem. *If a point in a function f is a local max/min point, then all of the partial derivatives at this point are equal to 0.*

Proof:

We will prove it by contradiction. Suppose there is a local max/min point in a function $f(x_1, x_2, ..., x_n)$, and one of its partial derivative $\frac{\partial f}{\partial x_i}$ is not 0. According to the definition of partial derivatives, in that direction the function values of the two adjacent point are

$$f(x_1, x_2, ..., x_i + \Delta x_i, ..., x_n) = f(x_1, x_2, ..., x_n) + \frac{\partial f}{\partial x_i} \Delta x_i$$

$$f(x_1, x_2, ..., x_i - \Delta x_i, ..., x_n) = f(x_1, x_2, ..., x_n) - \frac{\partial f}{\partial x_i} \Delta x_i$$

Therefore, one of them is less than the local max/min point and the other is greater than the local max/min point. Hence, it is not a local max/min point. There is a contradiction.

Therefore, it is proved that if a point in a function f is a local max/min point, then all of the partial derivatives at this point are equal to 0.

Therefore, A point is a local max/min point
$\Rightarrow$ All partial derivatives at that point are 0
$\Longleftrightarrow$ The tangent plane at that point is horizontal
$\Longleftrightarrow$ The point is a critical point.

1.2 Critical Points

Definition. *A point is a critical point of a function f if and only if all the partial derivatives on this point are equal to 0.*

For a multivariable function with two independent variables $f(x, y)$, a point (x_0, y_0) is a critical point if $\frac{\partial f}{\partial x}(x_0, y_0) = 0$ and $\frac{\partial f}{\partial y}(x_0, y_0) = 0$.

Example 1. *Find all the critical points of the function $f(x, y) = x^2 - 2xy + 3y^2 + 2x - 2y$.*
Solution:

$$\frac{\partial f}{\partial x} = 2x - 2y + 2$$
$$\frac{\partial f}{\partial y} = -2x + 6y - 2$$

For critical points, they should satisfy

$$\begin{cases} \frac{\partial f}{\partial x} = 2x - 2y + 2 = 0 \\ \frac{\partial f}{\partial y} = -2x + 6y - 2 = 0 \end{cases}$$

Solving this linear system, the result is $(-1, 0)$. Therefore, there are only one critical point for this function, which is $(-1, 0)$.

1.3 Types of Critical Points

A critical point of a function f can be

- local maximum point

- local minimum point

- saddle point: a point on the surface of the graph of a function where the slopes in orthogonal directions are all zero, but not a local extremum of the function.

Question 1. *Is there any other possibilities for a critical point of a multivariable function? If not, how to prove it?*
Answer: No, there is not. According to the definition of saddle points, all critical points that are not local max or min points are classified as saddle points.
The point here is that saddle points can be many different forms. Examples:

- *The point $(0, 0)$ on the graph of the function $z = x^2 - y^2$.*

- *The point $(0, 0)$ on the graph of the function $z = x^2 + y^3$.*

How to determine the type of a critical point?

- Test the second derivative (cover in the next lecture).

- Find the codomain of the function and compare the function value on the critical point with the codomain.

Example 2. *Find the type of the critical point $(-1, 0)$ of the function $f(x, y) = x^2 - 2xy + 3y^2 + 2x - 2y$.*
Solution:
The function value of the critical point $(-1, 0)$ is

$$f(-1, 0) = (-1)^2 - 2 \times (-1) \times 0 + 3 \times 0^2 + 2 \times (-1) - 2 \times 0$$
$$= 1 - 2$$
$$= -1$$

To find the codomain of the function, we can transform its formula

$$f(x, y) = x^2 - 2xy + 3y^2 + 2x - 2y$$
$$= (x - y)^2 + 2y^2 + 2x - 2y$$
$$= ((x - y)^2 + 2(x - y) + 1) + 2y^2 - 1$$
$$= (x - y + 1)^2 + 2y^2 - 1$$

Therefore, it is apparent that the codomain of the function is $[-1, \infty]$. Hence, the critical point $(-1, 0)$ is the minimum point of the function.

2 Example of Optimization Problems: Least Squares

Given a series of discrete points (x_1, y_1), (x_2, y_2), ..., (x_n, y_n), find the "best fit" line $y = ax + b$.
The task is to find the best values of a and b in the line equation $y = ax + b$.
We can define a distance function about the line equation $y = ax + b$, and find proper values of a and b to minimize the distance function, in other words, find the minimum point (a_0, b_0).

One way to define the distance function about the line equation $y = ax + b$ is to use the total square deviation, where deviation means the vertical distance between an actual point and the line $y = ax + b$, described by the formula $y_i - (ax_i + b)$. This method is called **least squares**.

Therefore, the distance function is

$$D = (y_1 - ax_1 - b)^2 + (y_2 - ax_2 - b)^2 + ... + (y_n - ax_n - b)^2$$

$$= \sum_{i=1}^{n} (y_i - ax_i - b)^2$$

$$\frac{\partial D}{\partial a} = \sum_{i=1}^{n} -2x_i(y_i - ax_i - b)$$

$$\frac{\partial D}{\partial b} = \sum_{i=1}^{n} -2(y_i - ax_i - b)$$

To find the minimum point, we need to solve the following system of equations:

$$\begin{cases} \frac{\partial D}{\partial a} = \sum_{i=1}^{n} -2x_i(y_i - ax_i - b) = 0 \\ \frac{\partial D}{\partial b} = \sum_{i=1}^{n} -2(y_i - ax_i - b) = 0 \end{cases}$$

which is a linear system of a and b. After simplified, it becomes

$$\begin{cases} (\sum_{i=1}^{n} x_i^2)a + (\sum_{i=1}^{n} x_i)b = \sum_{i=1}^{n} x_i y_i \\ (\sum_{i=1}^{n} x_i)a + nb = \sum_{i=1}^{n} y_i \end{cases}$$

Therefore, we can solve the linear system to get the critical points of the distance function, and hence get the minimum point (a_0, b_0).

The least squares method can be more general. First of all, it can be used to approximate any polynomial functions, since the deviation is a linear function of unknown parameters. For example,

$$y = ax^2 + bx + c$$

$$d_i = y_i - ax_i^2 - bx_i - c$$

Second, for other functions, we can try to transform it into polynomial functions. A typical example is exponential functions:

$$y = ce^{ax}$$

$$\ln y = ax + \ln c$$

$$d_i = \ln y_i - ax_i - \ln c$$