Calculus - Lecture 10

Higher order derivatives. Optimization.

EVA KASLIK

Second order partial derivatives

Let $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ be a partially differentiable function with respect to every variable x_j, $j = \overline{1,n}$ on A.

The function f is two times partially differentiable at a with respect to every variable if all partial derivatives $\frac{\partial f_i}{\partial x_j}$ are partially differentiable at $a \in A$ with respect to every variable x_k.

Notation for the second order partial derivative of f:

$$\frac{\partial}{\partial x_k}\left(\frac{\partial f_i}{\partial x_j}\right)(a) = \frac{\partial^2 f_i}{\partial x_k \partial x_j}(a)$$

.

Second order Fréchet derivative

The function f is two times differentiable at the point $a \in A$ if the partial derivatives $\dfrac{\partial f_i}{\partial x_j}$ are differentiable at a.

The second order Fréchet derivative of f at the point a is the function $d_a^2 f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$ given by the formula

$$d_a^2 f(u)(v) = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} \sum_{k=1}^{n} \frac{\partial^2 f_i}{\partial x_j \partial x_k}(a) u_j v_k \right) e_i$$

where $u, v \in \mathbb{R}^n$, $e_i = (0, ..., 0, 1, 0, ..., 0)$, $i = \overline{1, n}$.

The second order Fréchet derivative of f at a satisfies

$$\lim_{u \to 0} \frac{\|d_{a+u} f(v) - d_a f(v) - d_a^2 f(u)(v)\|}{\|u\|} = 0 \quad , \ \forall v \in \mathbb{R}^n.$$

Second order derivatives for two variable functions

Assume that $f : \mathbb{R}^2 \to \mathbb{R}$.

Second order partial derivatives:

$$\frac{\partial^2 f}{\partial x^2} = f_{xx} = (f_x)_x = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) \qquad \frac{\partial^2 f}{\partial x \partial y} = f_{yx} = (f_y)_x = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right)$$

$$\frac{\partial^2 f}{\partial y \partial x} = f_{xy} = (f_x)_y = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) \qquad \frac{\partial^2 f}{\partial y^2} = f_{yy} = (f_y)_y = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right)$$

Second order Fréchet derivative at $a = (a_1, a_2) \in \mathbb{R}^2$:

the function $d_a^2 f : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ given by

$$d_a^2 f(u)(v) = f_{xx}(a_1, a_2)u_1 v_1 + f_{xy}(a_1, a_2)u_1 v_2 + f_{yx}(a_1, a_2)u_2 v_1 + f_{yy}(a_1, a_2)u_2 v_2$$

for any $u = (u_1, u_2),\ v = (v_1, v_2) \in \mathbb{R}^2$.

Example

Consider the function $f(x,y) = xe^{xy}$.

The first order partial derivatives are:

$$f_x = e^{xy} + xye^{xy} \qquad \text{and} \qquad f_y = x^2 e^{xy}.$$

The second order partial derivatives are:

$$\begin{aligned} f_{xx} &= (f_x)_x = 2ye^{xy} + xy^2 e^{xy} & \qquad f_{xy} &= (f_x)_y = 2xe^{xy} + x^2 ye^{xy} \\ f_{yx} &= (f_y)_x = 2xe^{xy} + x^2 ye^{xy} & \qquad f_{yy} &= (f_y)_y = x^3 e^{xy} \end{aligned}$$

The Second order Fréchet derivative at the point $a = (1,0)$ is the function $d^2_{(1,0)} f : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ given by:

$$\begin{aligned} d^2_{(1,0)} f(u)(v) &= f_{xx}(1,0)u_1 v_1 + f_{xy}(1,0)u_1 v_2 + f_{yx}(1,0)u_2 v_1 + f_{yy}(1,0)u_2 v_2 \\ &= 2(u_1 v_2 + u_2 v_1) + u_2 v_2 \end{aligned}$$

for any $u = (u_1, u_2),\ v = (v_1, v_2) \in \mathbb{R}^2$.

Important theorems

Theorem (Mixed derivative theorem of Schwarz)

If the function f is twice differentiable at a, then

$$\frac{\partial^2 f_i}{\partial x_j \partial x_k}(a) = \frac{\partial^2 f_i}{\partial x_k \partial x_j}(a) \quad , \forall i = \overline{1,m},\ j,k = \overline{1,n}.$$

Theorem (Criterion for second order differentiability)

If the second order partial derivatives $\frac{\partial^2 f_i}{\partial x_j \partial x_k}$ exist in a neighborhood of a and they are continuous at a, then f is two times differentiable at a.

Higher order partial derivatives

The function $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ is k-times partially differentiable at $a \in A$ with respect to every variable if

- f is $(k-1)$-times partially differentiable with respect to every variable on an open neighborhood of a
- every $(k-1)$-th order partial derivative $\dfrac{\partial^{k-1} f_i}{\partial x_{j_{k-1}} \cdots \partial x_{j_1}}$ is partially differentiable with respect to every variable x_{j_k} at a.

The k-th order partial derivative of f at a is

$$\frac{\partial^k f_i}{\partial x_{j_k} \partial x_{j_{k-1}} \cdots \partial x_{j_1}}(a) = \frac{\partial}{\partial x_{j_k}} \left(\frac{\partial^{k-1} f_i}{\partial x_{j_{k-1}} \cdots \partial x_{j_1}} \right)(a)$$

Higher order differentiability

The function $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ is k-times differentiable at a if the partial derivatives of order $(k-1)$ are differentiable at a.

The Fréchet derivative of order k of f at a is the function $d_a^k f : \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}^m$ given by

$$d_a^k f(u^1)(u^2)\cdots(u^k) = \sum_{i=1}^{m} \left(\sum_{j_1=1}^{n} \sum_{j_2=1}^{n} \cdots \sum_{j_k=1}^{n} \frac{\partial^k f_i}{\partial x_{j_k} \cdots \partial x_{j_1}}(a) \cdot u^1_{j_1} u^2_{j_2} \cdots u^k_{j_k} \right) e_i$$

The Fréchet derivative of order k of f at a satisfies:

$$\lim_{\|u^k\| \to 0} \frac{\|d_{a+u^k}^{k-1} f(u^1)(u^2)\cdots(u^{k-1}) - d_a^{k-1} f(u^1)(u^2)\cdots(u^{k-1}) - d_a^k f(u^1)(u^2)\cdots(u^k)\|}{\|u^k\|} = 0$$

Important results

Theorem (Mixed derivative theorem)

If the function is k-times differentiable at a, then the following relations hold:

$$\frac{\partial^k f_i}{\partial x_{j_1} \partial x_{j_2} \cdots \partial x_{j_k}}(a) = \frac{\partial^k f_i}{\partial x_{\sigma(j_1)} \partial x_{\sigma(j_2)} \cdots \partial x_{\sigma(j_k)}}(a)$$

Theorem (Criterion for k-times differentiability)

If the partial derivatives of k-th order of the function f exist in a neighborhood of a and they are continuous at a, then f is k-times differentiable at a.

Minimum and maximum values

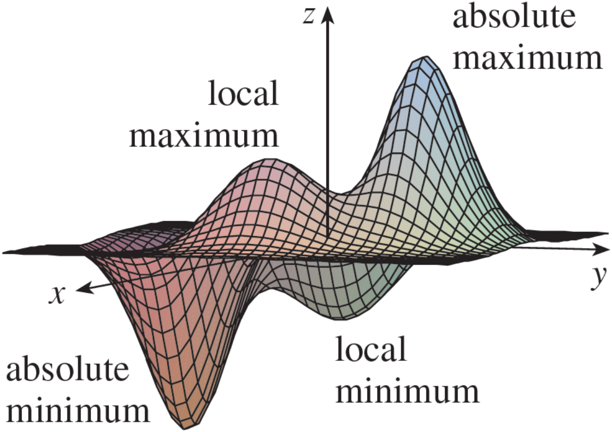
The point $a \in A$ is a local minimum point of the function $f : A \subset \mathbb{R}^n \to \mathbb{R}^1$ if there exists a neighborhood $V \subset A$ of a such that $f(a) \leq f(x)$ for any $x \in V$.

The point $a \in A$ is a global minimum point of the function $f : A \subset \mathbb{R}^n \to \mathbb{R}^1$ if $f(a) \leq f(x)$ for any $x \in A$.

The point $a \in A$ is a local maximum point of the function $f : A \subset \mathbb{R}^n \to \mathbb{R}^1$ if there exists a neighborhood $V \subset A$ of a such that $f(a) \geq f(x)$ for any $x \in V$.

The point $a \in A$ is a global maximum point of the function $f : A \subset \mathbb{R}^n \to \mathbb{R}^1$ if $f(a) \geq f(x)$ for any $x \in A$.

Minimum and maximum values



Conditions for local extreme values

Necessary condition for local extrema:
If the function $f : A \subset \mathbb{R}^n \to \mathbb{R}^1$ attains a local minimum or maximum value at the point $a \in A$ and all partial derivatives of f exist at a, then

$$\nabla f(a) = 0,$$

i.e. a is a critical point (stationary point) of f.

Sufficient condition for local extrema:
Assume that $f : A \subset \mathbb{R}^n \to \mathbb{R}^1$ has continuous second order partial derivatives on A and a is a critical point of f.

i) If $d_a^2 f(h)(h) \geq 0$ for $h \in \mathbb{R}^n$ and $\det\left(\frac{\partial^2 f}{\partial x_i \partial x_j}(a)\right) \neq 0$, then a is a local minimum point of f;

ii) If $d_a^2 f(h)(h) \leq 0$ for $h \in \mathbb{R}^n$ and $\det\left(\frac{\partial^2 f}{\partial x_i \partial x_j}(a)\right) \neq 0$, then a is a local maximum point of f.

Second derivative test for two variable functions

Assume that $a = (a_1, a_2) \in A$ is a critical point of the function $f : A \subset \mathbb{R}^2 \to \mathbb{R}$. Consider the Hessian matrix:

$$H_{(a_1,a_2)}f = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x^2}(a_1, a_2) & \dfrac{\partial^2 f}{\partial x \partial y}(a_1, a_2) \\ \dfrac{\partial^2 f}{\partial y \partial x}(a_1, a_2) & \dfrac{\partial^2 f}{\partial y^2}(a_1, a_2) \end{pmatrix}$$

Consider the principal minors of the Hessian matrix:

$$\Delta_1 = \frac{\partial^2 f}{\partial x^2}(a_1, a_2) \qquad \text{and} \qquad \Delta_2 = \det\left(H_{(a_1,a_2)}f\right)$$

- if $\Delta_1 > 0$ and $\Delta_2 > 0$ then $a = (a_1, a_2)$ is a local minimum point of f;
- if $\Delta_1 < 0$ and $\Delta_2 > 0$ then $a = (a_1, a_2)$ is a local maximum point of f;
- if $\Delta_2 < 0$ then $a = (a_1, a_2)$ is a saddle point of f;
- if $\Delta_2 = 0$ then this test is inconclusive.

Examples

Example 1.

$$f(x,y) = x^2 + y^2 - 2x - 6y + 14$$

Partial derivatives:

$$f_x = 2x - 2 \qquad f_y = 2y - 6$$

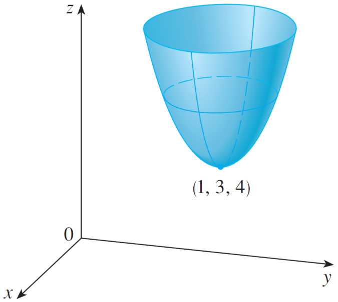
$\Longrightarrow$ critical point: $(1,3)$.

Hessian Matrix at $(1,3)$:

$$H_{(1,3)}f = \begin{pmatrix} f_{xx}(1,3) & f_{xy}(1,3) \\ f_{yx}(1,3) & f_{yy}(1,3) \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

As $\Delta_1 = 2 > 0$ and $\Delta_2 = 4 > 0$ we deduce that $(1,3)$ is a minimum point of the f.

Minimum value: $f(1,3) = 4$



Examples

Example 2.

$$f(x,y) = y^2 - x^2$$

Partial derivatives:

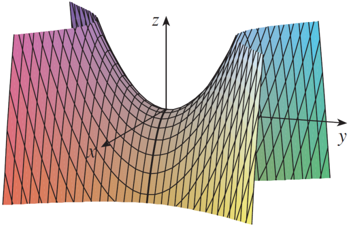
$$f_x = -2x \qquad f_y = 2y$$

$\implies$ critical point: $(0,0)$.

Hessian Matrix at $(0,0)$:

$$H_{(0,0)}f = \begin{pmatrix} f_{xx}(0,0) & f_{xy}(0,0) \\ f_{yx}(0,0) & f_{yy}(0,0) \end{pmatrix} = \begin{pmatrix} -2 & 0 \\ 0 & 2 \end{pmatrix}$$

$\Delta_2 = -4 < 0 \implies (0,0)$ is a saddle point.



Examples

Example 3.

$$f(x,y) = x^4 + y^4 - 4xy + 1$$

Partial derivatives:

$$f_x = 4x^3 - 4y \qquad f_y = 4y^3 - 4x$$

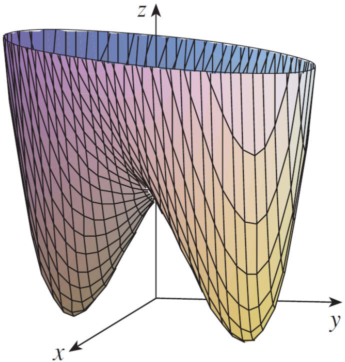
$\Longrightarrow$ critical points: $(0,0)$, $(1,1)$, $(-1,-1)$.

Hessian Matrix:

$$H_{(x,y)}f = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix} = \begin{pmatrix} 12x^2 & -4 \\ -4 & 12y^2 \end{pmatrix}$$

$\Longrightarrow$ $\Delta_1 = 12x^2$ and $\Delta_2 = 144x^2y^2 - 16$

- $(0,0)$ is a saddle point ($\Delta_2 = -16 < 0$)
- $(1,1)$ and $(-1,-1)$ are local minimum points ($\Delta_1 = 12 > 0$ and $\Delta_2 = 128 > 0$)



Lagrange multipliers and constrained optimization

Consider a function $f : A \subset \mathbb{R}^n \to \mathbb{R}^1$, where A is an open set and the set $\Gamma \subset A$, defined by:

$$\Gamma = \{x \in A \ : \ g_i(x) = 0, \ i = \overline{1,p}\} \qquad \text{where } g_i : A \to \mathbb{R}^1 \text{ and } p < n$$

The equations $g_i(x) = 0$ are called constraints.

If the restriction of the function f to the set Γ, i.e. $f|_\Gamma$, has an extreme point $a \in \Gamma$, then this is called conditional extreme point.

Method of Lagrange Multipliers:

Assume that f and g_i, $i = \overline{1,p}$ are continuously differentiable near the conditional extreme point $a \in \Gamma$ and the gradient vectors $\nabla g_i(a)$, $i = \overline{1,p}$ are linearly independent vectors of $\mathbb{R}^n$.

Then there exist some constants $\lambda_1, \lambda_2, ..., \lambda_p$ such that

$$\nabla f(a) = \sum_{i=1}^{p} \lambda_i \nabla g_i(a)$$

Special case: two variables and one constraint

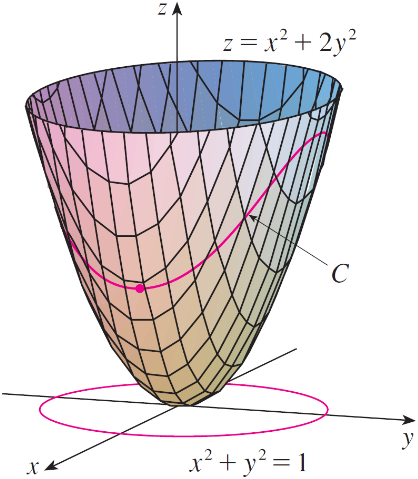
If we want to maximize (minimize) the function $f : A \subset \mathbb{R}^2 \to \mathbb{R}^1$ subject to the constraint $g(x, y) = 0$, we first need to solve the system of three equations

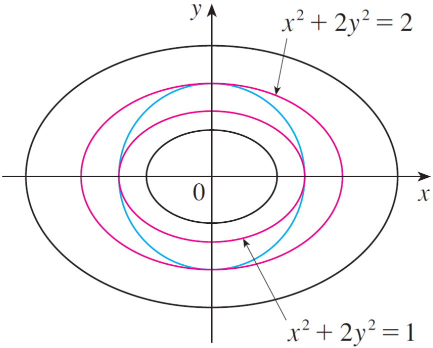
$$\begin{cases} g(x,y) = 0 \\ \dfrac{\partial f}{\partial x}(x,y) = \lambda \dfrac{\partial g}{\partial x}(x,y) \\ \dfrac{\partial f}{\partial y}(x,y) = \lambda \dfrac{\partial g}{\partial y}(x,y) \end{cases}$$

with respect to the variables x, y, λ. The points (x, y) that we find are the only possible locations of the extreme values of f subject to the constraint $g(x, y) = 0$.

Example

Find the extreme values of $f(x, y) = x^2 + 2y^2$ on the circle $x^2 + y^2 = 1$.





Example

Find the extreme values of $f(x,y) = x^2 + 2y^2$ on the circle $x^2 + y^2 = 1$.

constraint: $g(x,y) = x^2 + y^2 - 1 = 0$.

We have to solve the system:

$$\begin{cases} g(x,y) = 0 \\ f_x = \lambda g_x \\ f_y = \lambda g_y \end{cases} \implies \begin{cases} x^2 + y^2 = 1 \\ 2x = \lambda \cdot 2x \\ 4y = \lambda \cdot 2y \end{cases}$$

- if $x = 0$, then $y = \pm 1$;
- if $\lambda = 1$, then $y = 0$ and $x = \pm 1$.

$\implies$ possible extreme points: $(1,0)$, $(-1,0)$, $(0,1)$ and $(0,-1)$.

Evaluating f at each of these points gives the minimum and maximum value of the function on the circle $x^2 + y^2 = 1$:

$$f(\pm 1, 0) = \underbrace{1}_{\text{min}} \quad \text{and} \quad f(0, \pm 1) = \underbrace{2}_{\text{max}}.$$

Special case: three variables and two constraints

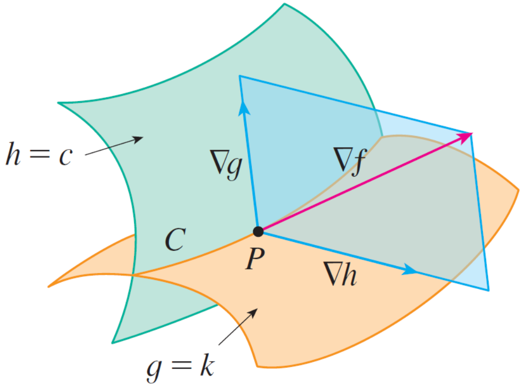
If we want to maximize (minimize) the function $f : A \subset \mathbb{R}^3 \to \mathbb{R}^1$ subject to constraints $g(x, y, z) = 0$ and $h(x, y, z) = 0$, we first need to solve the system of five equations

$$\begin{cases} g(x,y,z) = 0 \\ h(x,y,z) = 0 \\ \dfrac{\partial f}{\partial x}(x,y,z) = \lambda_1 \dfrac{\partial g}{\partial x}(x,y,z) + \lambda_2 \dfrac{\partial h}{\partial x}(x,y,z) \\ \dfrac{\partial f}{\partial y}(x,y,z) = \lambda_1 \dfrac{\partial g}{\partial y}(x,y,z) + \lambda_2 \dfrac{\partial h}{\partial y}(x,y,z) \\ \dfrac{\partial f}{\partial z}(x,y,z) = \lambda_1 \dfrac{\partial g}{\partial z}(x,y,z) + \lambda_2 \dfrac{\partial h}{\partial z}(x,y,z) \end{cases}$$

with respect to the variables $x, y, z, \lambda_1, \lambda_2$. The points (x, y, z) that we find are the only possible locations of the extreme values of f subject to the two constraints.

Special case: three variables and two constraints

∇f is in the plane determined by ∇g and ∇f:



Exercise. Find the maximum possible area of a right triangle of fixed perimeter P.