Mathematics for Informatics

Multivariate optimization
(lecture 2 of 12)

Francesco Dolce

dolcefra@fit.cvut.cz

Czech Technical University in Prague

Winter 2024/2025

created: September 11, 2024, 14:20

What shall we do today?

- Multivariate optimization:
  - Gradient
  - Tangent plane
  - Critical points on two or more variables
  - Hessian (matrix)

Gradient of a function

The gradient of a function $f(x_1, x_2, \ldots, x_n)$ at the (n-dimensional) point $b \in \mathbb{R}^n$ is the n-dimensional vector function $\nabla f(b)$ defined by

$$\nabla f(b) = \left(\frac{\partial f}{\partial x_1}(b), \frac{\partial f}{\partial x_2}(b), \ldots, \frac{\partial f}{\partial x_n}(b) \right).$$

Gradient of a function

The gradient of a function $f(x_1, x_2, \dots, x_n)$ at the (n-dimensional) point $b \in \mathbb{R}^n$ is the n-dimensional vector function $\nabla f(b)$ defined by

$$\nabla f(b) = \left(\frac{\partial f}{\partial x_1}(b), \frac{\partial f}{\partial x_2}(b), \dots, \frac{\partial f}{\partial x_n}(b) \right).$$

Example

Find the gradient of the function $f(x, y) = x^2 + xy + y^2$ *at the point* $(1, 1)$.

Gradient of a function

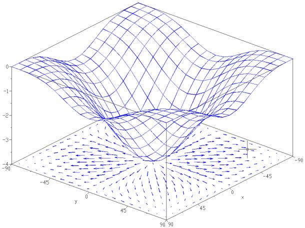
The gradient of a function $f(x_1, x_2, \ldots, x_n)$ at the (n-dimensional) point $b \in \mathbb{R}^n$ is the n-dimensional vector function $\nabla f(b)$ defined by

$$\nabla f(b) = \left(\frac{\partial f}{\partial x_1}(b), \frac{\partial f}{\partial x_2}(b), \ldots, \frac{\partial f}{\partial x_n}(b) \right).$$

Example

Find the gradient of the function $f(x, y) = x^2 + xy + y^2$ *at the point* $(1, 1)$.

Geometrical meaning: the gradient points is the direction of the greatest rate of increase of the function. Its magnitude equals the rate of increase.



Gradient and the directional derivative

We saw that the partial derivative with respect to x at the point a is equal to the slope of tangent line at this point in direction parallel to the x-axis.

Gradient and the directional derivative

We saw that the partial derivative with respect to x at the point a is equal to the slope of tangent line at this point in direction parallel to the x-axis.

Example

If we are on the graph of the fonction $f(x,y) = x^2 + xy + y^2$ *at the point* $(1,1)$ *and we start moving in the direction parallel to the* x*-axis, i.e., in the direction of the vector* $(1,0)$*, we will go "uphill" under the angle* $\arctan 3$ *since*

$$\frac{\partial f}{\partial x}(1,1) = 2 + 1 = 3.$$

Gradient and the directional derivative

We saw that the partial derivative with respect to x at the point a is equal to the slope of tangent line at this point in direction parallel to the x-axis.

Example

If we are on the graph of the fonction $f(x,y) = x^2 + xy + y^2$ *at the point* $(1,1)$ *and we start moving in the direction parallel to the* x*-axis, i.e., in the direction of the vector* $(1,0)$*, we will go "uphill" under the angle* $\arctan 3$ *since*

$$\frac{\partial f}{\partial x}(1,1) = 2 + 1 = 3.$$

What will be the slope if we move in the direction of a general vector $\vec{v}$?

Gradient and the directional derivative

We saw that the partial derivative with respect to x at the point a is equal to the slope of tangent line at this point in direction parallel to the x-axis.

Example

If we are on the graph of the fonction $f(x, y) = x^2 + xy + y^2$ *at the point* $(1, 1)$ *and we start moving in the direction parallel to the* x*-axis, i.e., in the direction of the vector* $(1, 0)$*, we will go "uphill" under the angle* $\arctan 3$ *since*

$$\frac{\partial f}{\partial x}(1, 1) = 2 + 1 = 3.$$

What will be the slope if we move in the direction of a general vector $\vec{v}$?

Theorem

Given a function $f(x) : \mathbb{R}^n \to \mathbb{R}$*, a point* $a \in \mathbb{R}^n$ *and a* ***unit*** *vector* $\vec{v} \in \mathbb{R}^n$*, the derivative in the direction of the vector* $\vec{v}$ *is the dot product of the gradient and* $\vec{v}$*, i.e,* $\nabla f(a_1, a_2, \ldots, a_n) \cdot \vec{v}$*.*

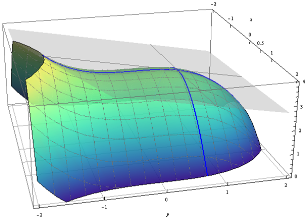
Tangent plane

The tangent plane to a function $f(x, y)$ at the point (x_0, y_0) is a 2-dimensional plane that "touches" the graph of the function at (x_0, y_0).

Tangent plane

The tangent plane to a function $f(x, y)$ at the point (x_0, y_0) is a 2-dimensional plane that "touches" the graph of the function at (x_0, y_0). Its equation is

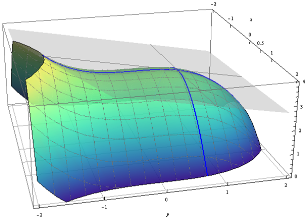
$$z = \frac{\partial f}{\partial x}(x_0, y_0) \cdot (x - x_0) \; + \; \frac{\partial f}{\partial y}(x_0, y_0) \cdot (y - y_0) \; + \; f(x_0, y_0).$$



Tangent plane

The tangent plane to a function $f(x, y)$ at the point (x_0, y_0) is a 2-dimensional plane that "touches" the graph of the function at (x_0, y_0). Its equation is

$$z = \frac{\partial f}{\partial x}(x_0, y_0) \cdot (x - x_0) \; + \; \frac{\partial f}{\partial y}(x_0, y_0) \cdot (y - y_0) \; + \; f(x_0, y_0).$$



Example

Find the tangent plane to $f(x, y) = x^2 + xy + y^2$ *at* $(1, 1)$.

Critical points – two variables

- In the one dimensional case the critical points are those points where the tangent line is parallel to the x-axis, i.e., points where $f'(x) = 0$, or where the derivative does not exist.

Critical points – two variables

- In the one dimensional case the critical points are those points where the tangent line is parallel to the x-axis, i.e., points where $f'(x) = 0$, or where the derivative does not exist.
- The critical points of a two variable function are those points where the tangent plane is parallel to the plane given by the x-axis and the y-axis or where the gradient does not exist.

Critical points – two variables

- In the one dimensional case the critical points are those points where the tangent line is parallel to the x-axis, i.e., points where $f'(x) = 0$, or where the derivative does not exist.
- The critical points of a two variable function are those points where the tangent plane is parallel to the plane given by the x-axis and the y-axis or where the gradient does not exist.

The first class of these points can be found as a solution of

$$\nabla f(x, y) = (0, 0)$$

which leads to the system of two equations for two variables

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

Critical points – more variables

For an n-variable function $f(x_1, x_2, \ldots, x_n)$ the situation is analogous:
The critical points of $f(x_1, x_2, \ldots, x_n)$ are points satisfying the equation

$$\nabla f(x_1, x_2, \ldots, x_n) = 0$$

i.e., points satisfying the system of n equations for n variables

$$\begin{cases} \dfrac{\partial f}{\partial x_1}(x_1, x_2, \ldots, x_n) & = & 0 \\ \dfrac{\partial f}{\partial x_2}(x_1, x_2, \ldots, x_n) & = & 0 \\ & \vdots & \\ \dfrac{\partial f}{\partial x_n}(x_1, x_2, \ldots, x_n) & = & 0 \end{cases},$$

<u>or</u> where the gradient does not exist.

Critical points – more variables

For an n-variable function $f(x_1, x_2, \ldots, x_n)$ the situation is analogous:
The critical points of $f(x_1, x_2, \ldots, x_n)$ are points satisfying the equation

$$\nabla f(x_1, x_2, \ldots, x_n) = 0$$

i.e., points satisfying the system of n equations for n variables

$$\begin{cases} \dfrac{\partial f}{\partial x_1}(x_1, x_2, \ldots, x_n) & = & 0 \\ \dfrac{\partial f}{\partial x_2}(x_1, x_2, \ldots, x_n) & = & 0 \\ & \vdots & \\ \dfrac{\partial f}{\partial x_n}(x_1, x_2, \ldots, x_n) & = & 0 \end{cases},$$

or where the gradient does not exist.
(Instead of a tangent plane, we have a tangent hyperplane.)

Critical points – example

Example

Find all critical points of

$$f(x_1, x_2, x_3) = x_1x_3 + x_1^2 - x_2 + x_2x_3 + x_2^2 + 3x_3^2,$$

Critical points – example

Example

Find all critical points of

$$f(x_1, x_2, x_3) = x_1x_3 + x_1^2 - x_2 + x_2x_3 + x_2^2 + 3x_3^2,$$

We get

$$\nabla f(x_1, x_2, x_3) = (x_3 + 2x_1,\ -1 + x_3 + 2x_2,\ x_1 + x_2 + 6x_3)$$

which always exists.

Critical points – example

Example

Find all critical points of

$$f(x_1, x_2, x_3) = x_1x_3 + x_1^2 - x_2 + x_2x_3 + x_2^2 + 3x_3^2,$$

We get

$$\nabla f(x_1, x_2, x_3) = (x_3 + 2x_1,\ -1 + x_3 + 2x_2,\ x_1 + x_2 + 6x_3)$$

which always exists. Thus the critical points are the solution of the system

$$\begin{cases} x_3 + 2x_1 & = & 0 \\ -1 + x_3 + 2x_2 & = & 0 \\ x_1 + x_2 + 6x_3 & = & 0 \end{cases},$$

which, using the standard procedure for a system of linear equations, gives us the only solution $\left(\frac{1}{20}, \frac{11}{20}, \frac{-1}{10}\right)$.

Type of a critical point (1 of 4)

In the one dimensional case, we can use the second derivative to decide the type of the critical point.

Theorem

Let x_0 be a critical point of a function $f(x)$ such that $f'(x_0) = 0$ and $f''(x_0)$ exists.

- *If $f''(x_0) > 0$, then the function is convex at x_0, and x_0 is a point of a minimum.*
- *If $f''(x_0) < 0$, then the function is concave at x_0, and x_0 is a point of a maximum.*
- *If $f''(x_0) = 0$, then x_0 may be a minimum, maximum, inflection point, ...*

Type of a critical point (1 of 4)

In the one dimensional case, we can use the second derivative to decide the type of the critical point.

Theorem

Let x_0 be a critical point of a function $f(x)$ such that $f'(x_0) = 0$ and $f''(x_0)$ exists.

- *If $f''(x_0) > 0$, then the function is convex at x_0, and x_0 is a point of a minimum.*
- *If $f''(x_0) < 0$, then the function is concave at x_0, and x_0 is a point of a maximum.*
- *If $f''(x_0) = 0$, then x_0 may be a minimum, maximum, inflection point, . . .*

Do we have something similar for more variables? What is the second derivative?

Type of a critical point (2 of 4)

The analogue of the second derivative is the following.

Definition

For a function $f(x_1, x_2, \ldots, x_n)$ we define the Hessian matrix as

$$\nabla^2 f(x_1, x_2, \ldots, x_n) = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1^2}(x_1, \ldots, x_n) & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n}(x_1, \ldots, x_n) \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_n \partial x_1}(x_1, \ldots, x_n) & \cdots & \dfrac{\partial^2 f}{\partial x_n^2}(x_1, \ldots, x_n) \end{pmatrix}$$

assuming that all the derivatives exist.

Type of a critical point (3 of 4)

We would like to construct rules like "If $f''(x_0) > 0$, then the critical point x_0 is the point of minimum".

But to say that the matrix is "positive" is problematic ... Let us use a different notion.

Type of a critical point (3 of 4)

We would like to construct rules like "If $f''(x_0) > 0$, then the critical point x_0 is the point of minimum".

But to say that the matrix is "positive" is problematic ... Let us use a different notion.

Definition

A matrix $A \in \mathbb{R}^{n,n}$ is

(i) *positively definite if for all non-zero vectors $a \in \mathbb{R}^n$ it holds that $aAa^T > 0$;*

Type of a critical point (3 of 4)

We would like to construct rules like "If $f''(x_0) > 0$, then the critical point x_0 is the point of minimum".

But to say that the matrix is "positive" is problematic ... Let us use a different notion.

Definition

A matrix $A \in \mathbb{R}^{n,n}$ is

1. *positively definite if for all non-zero vectors $a \in \mathbb{R}^n$ it holds that $aAa^T > 0$;*
2. *positively semidefinite if for all vectors $a \in \mathbb{R}^n$ it holds that $aAa^T \geq 0$ and the equality is true for at least one non-zero vector $b \in \mathbb{R}^n$;*

Type of a critical point (3 of 4)

We would like to construct rules like "If $f''(x_0) > 0$, then the critical point x_0 is the point of minimum".

But to say that the matrix is "positive" is problematic ... Let us use a different notion.

Definition

A matrix $A \in \mathbb{R}^{n,n}$ is

1. *positively definite if for all non-zero vectors $a \in \mathbb{R}^n$ it holds that $aAa^T > 0$;*
2. *positively semidefinite if for all vectors $a \in \mathbb{R}^n$ it holds that $aAa^T \geq 0$ and the equality is true for at least one non-zero vector $b \in \mathbb{R}^n$;*
3. *negatively definite if for all non-zero vectors $a \in \mathbb{R}^n$ it holds that $aAa^T < 0$;*
4. *negatively semidefinite if for all vectors $a \in \mathbb{R}^n$ it holds that $aAa^T \leq 0$ and the equality is true for at least one non-zero vector $b \in \mathbb{R}^n$;*

Type of a critical point (3 of 4)

We would like to construct rules like "If $f''(x_0) > 0$, then the critical point x_0 is the point of minimum".

But to say that the matrix is "positive" is problematic ... Let us use a different notion.

Definition

A matrix $A \in \mathbb{R}^{n,n}$ is

1. *positively definite if for all non-zero vectors $a \in \mathbb{R}^n$ it holds that $aAa^T > 0$;*
2. *positively semidefinite if for all vectors $a \in \mathbb{R}^n$ it holds that $aAa^T \geq 0$ and the equality is true for at least one non-zero vector $b \in \mathbb{R}^n$;*
3. *negatively definite if for all non-zero vectors $a \in \mathbb{R}^n$ it holds that $aAa^T < 0$;*
4. *negatively semidefinite if for all vectors $a \in \mathbb{R}^n$ it holds that $aAa^T \leq 0$ and the equality is true for at least one non-zero vector $b \in \mathbb{R}^n$;*
5. *indefinite otherwise.*

Type of a critical point (4 of 4)

Theorem

If $f : \mathbb{R}^n \to \mathbb{R}$ has all second partial derivative continuous at a critical point $b \in \mathbb{R}^n$, then

(i) *if $\nabla^2 f(b)$ is positively definite, then b is a point of local minimum;*

(ii) *if $\nabla^2 f(b)$ is negatively definite, then b is a point of local maximum;*

(iii) *if $\nabla^2 f(b)$ is indefinite, then b is a saddle point.*

Sylvester's criterion on definiteness

For an $n \times n$ dimensional **symmetric** matrix A we define the principal minors:

- M_1 is the upper left 1-by-1 corner of A,
- M_2 is the upper left 2-by-2 corner of A,
- ...
- M_n is the upper left n-by-n corner of A.

Theorem

Let $A \in \mathbb{R}^{n,n}$ be a symmetric matrix.

- *A is positively definite if and only if the determinants of all principal minors are positive.*
- *A is negatively definite if and only if the determinant of M_i is negative for odd i and positive for even i.*

Example

Example

Find all minima and maxima of the function

$$f(x, y) = \frac{3x^4 - 4x^3 - 12x^2 + 18}{12(1 + 4y^2)}.$$

Example

Example

Find all minima and maxima of the function

$$f(x,y) = \frac{3x^4 - 4x^3 - 12x^2 + 18}{12(1+4y^2)}.$$

Solution: The critical points are $(-1,0)$, $(0,0)$ and $(2,0)$; they are a saddle point, a point of maximum and a point of minimum, respectively.