Mathematics 2

Prof. Katia Colaneri

Week 6 - Unconstrained Optimization

University of Rome Tor Vergata

10-15 December 2023

Outline

1. First order partial derivatives
2. Second order partial derivatives
3. Unconstrained optimization

Continuously differentiable $\mathcal{C}^1$ functions

Definition

Let $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$ *be a continuous function of two variables.*
Suppose that f *has both partial derivatives* $f_x(x, y)$ *and* $f_y(x, y)$ *for all* $(x, y) \in D_1 \subseteq D$ *and that the partial derivatives are continuous in* D_1*.*
Then we say that $f(x, y)$ *is continuously differentiable in* D_1, $f \in \mathcal{C}^1(D_1)$

$f \in \mathcal{C}^1(D_1)$ means that:

- $f(x, y)$ is continuous in D_1
- $f_x(x, y)$ and $f_y(x, y)$ exist and are continuous in D_1

Stationary points

Definition (Stationary points)

Let $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$ *be a function of two variables. Suppose that* $f \in \mathcal{C}^1(D)$.

A point $(x_0, y_0) \in D$ *is a stationary point for* f *if*

$$f_x(x_0, y_0) = 0$$

$$f_y(x_0, y_0) = 0$$

that is the gradient vanishes at (x_0, y_0).

Tangent plane

We can define the 1st order Taylor polynomial for functions of two variables:

For any point (x_0, y_0) we get

$$p(x, y) = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)$$

The equation $z = p(x, y)$ is the equation of the tangent plane.

If (x_0, y_0) is a stationary point the tangent plane satisfies:

$$z = f(x_0, y_0)$$

hence, at stationary points the tangent plane is horizontal!

Second order partial derivatives

Let $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$ be a function of two variables and suppose that f has both partial derivatives with respect to x and y, $f_x(x, y)$ and $f_y(x, y)$.

If the functions $f_x(x, y)$ and $f_y(x, y)$ are differentiable with respect to x and y we define the **second order partial derivatives** of f as follows:

- $\dfrac{\partial}{\partial x}\left(\dfrac{\partial f}{\partial x}\right) = \dfrac{\partial^2 f}{\partial x^2} = D_{xx}(f(x, y)) = f_{xx}(x, y)$

Second order partial derivatives

Let $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$ be a function of two variables and suppose that f has both partial derivatives with respect to x and y, $f_x(x, y)$ and $f_y(x, y)$.

If the functions $f_x(x, y)$ and $f_y(x, y)$ are differentiable with respect to x and y we define the **second order partial derivatives** of f as follows:

- $\dfrac{\partial}{\partial x}\left(\dfrac{\partial f}{\partial x}\right) = \dfrac{\partial^2 f}{\partial x^2} = D_{xx}(f(x, y)) = f_{xx}(x, y)$
- $\dfrac{\partial}{\partial y}\left(\dfrac{\partial f}{\partial x}\right) = \dfrac{\partial^2 f}{\partial y \partial x} = D_{xy}(f(x, y)) = f_{xy}(x, y)$

Second order partial derivatives

Let $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$ be a function of two variables and suppose that f has both partial derivatives with respect to x and y, $f_x(x, y)$ and $f_y(x, y)$.

If the functions $f_x(x, y)$ and $f_y(x, y)$ are differentiable with respect to x and y we define the **second order partial derivatives** of f as follows:

- $\dfrac{\partial}{\partial x}\left(\dfrac{\partial f}{\partial x}\right) = \dfrac{\partial^2 f}{\partial x^2} = D_{xx}(f(x, y)) = f_{xx}(x, y)$
- $\dfrac{\partial}{\partial y}\left(\dfrac{\partial f}{\partial x}\right) = \dfrac{\partial^2 f}{\partial y \partial x} = D_{xy}(f(x, y)) = f_{xy}(x, y)$
- $\dfrac{\partial}{\partial x}\left(\dfrac{\partial f}{\partial y}\right) = \dfrac{\partial^2 f}{\partial x \partial y} = D_{yx}(f(x, y)) = f_{yx}(x, y)$

Second order partial derivatives

Let $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$ be a function of two variables and suppose that f has both partial derivatives with respect to x and y, $f_x(x, y)$ and $f_y(x, y)$.

If the functions $f_x(x, y)$ and $f_y(x, y)$ are differentiable with respect to x and y we define the **second order partial derivatives** of f as follows:

- $\dfrac{\partial}{\partial x}\left(\dfrac{\partial f}{\partial x}\right) = \dfrac{\partial^2 f}{\partial x^2} = D_{xx}(f(x, y)) = f_{xx}(x, y)$
- $\dfrac{\partial}{\partial y}\left(\dfrac{\partial f}{\partial x}\right) = \dfrac{\partial^2 f}{\partial y \partial x} = D_{xy}(f(x, y)) = f_{xy}(x, y)$
- $\dfrac{\partial}{\partial x}\left(\dfrac{\partial f}{\partial y}\right) = \dfrac{\partial^2 f}{\partial x \partial y} = D_{yx}(f(x, y)) = f_{yx}(x, y)$
- $\dfrac{\partial}{\partial y}\left(\dfrac{\partial f}{\partial y}\right) = \dfrac{\partial^2 f}{\partial y^2} = D_{yy}(f(x, y)) = f_{yy}(x, y)$

Theorem (Schwarz Theorem)

Let $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$. *If all the functions* $f(x, y)$, $f_x(x, y)$, $f_y(x, y)$, $f_{xy}(x, y)$ *and* $f_{yx}(x, y)$ *are continuous then*

$$f_{xy}(x, y) = f_{yx}(x, y)$$

Theorem (Schwarz Theorem)

Let $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$. *If all the functions* $f(x, y)$, $f_x(x, y)$, $f_y(x, y)$, $f_{xy}(x, y)$ *and* $f_{yx}(x, y)$ *are continuous then*

$$f_{xy}(x, y) = f_{yx}(x, y)$$

Definition

Let $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$ *be a function of two variables. Suppose that*

- $f(x, y)$ *is continuous in* $D_2 \subseteq D$,
- $f_x(x, y)$ *and* $f_y(x, y)$ *are continuous in* $D_2 \subseteq D$,
- $f_{xx}(x, y)$, $f_{xy}(x, y)$, $f_{yx}(x, y)$, $f_{yy}(x, y)$ *are continuous in* $D_2 \subseteq D$.

Then we say that the function f *is* ***twice continuously differentiable*** *or equivalently* $f \in \mathcal{C}^2(D_2)$

Definition

Let $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$ *be a function of two variables. If* f *has all second derivatives with respect to* x *and* y *we can define the* ***Hessian matrix*** *of* f *as the matrix*

$$H_f(x, y) \begin{pmatrix} f_{xx}(x, y) & f_{xy}(x, y) \\ f_{yx}(x, y) & f_{yy}(x, y) \end{pmatrix}$$

Definition

Let $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$ *be a function of two variables. If* f *has all second derivatives with respect to* x *and* y *we can define the* ***Hessian matrix*** *of* f *as the matrix*

$$H_f(x,y) \begin{pmatrix} f_{xx}(x,y) & f_{xy}(x,y) \\ f_{yx}(x,y) & f_{yy}(x,y) \end{pmatrix}$$

If all the functions $f(x,y)$, $f_x(x,y)$, $f_y(x,y)$, $f_{xy}(x,y)$ and $f_{yx}(x,y)$ are continuous by the Schwarz Theorem we get that the Hessian matrix is **symmetric**.

Let $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$. We want to solve problems

$$\min_{(x,y)\in D} f(x,y) \quad \text{or} \quad \max_{(x,y)\in D} f(x,y)$$

Reminder: For functions of one variable we had necessary and sufficient conditions to say if a point in the domain is a maximum or a minimum:

- Necessary condition: $f'(x) = 0$
- Sufficient condition: sign of $f''(x)$. If positive (convex) we get a minimum, if negative (concave) we get a maximum, if zero we get an inflection point with horizontal tangent

Is there a way to generalize this to functions of several variables ?

Maxima and Minima

Let $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$, let (x^*, y^*) and let $B_r \subseteq D$ be a disk centered at (x^*, y^*) with radius r. We say that (x^*, y^*) is

- a **maximum** if $f(x, y) \leqslant f(x^*, y^*)$, for all $(x, y) \in D$
- a local **maximum** if $f(x, y) \leqslant f(x^*, y^*)$, for all $(x, y) \in B_r$
- a **minimum** if $f(x, y) \geqslant f(x^*, y^*)$, for all $(x, y) \in D$
- a local **minimum** if $f(x, y) \geqslant f(x^*, y^*)$, for all $(x, y) \in B_r$

Local Maxima and Minima of general functions

The intuition is : for a point $(x^*, y^*) \in D$ to be a local maximum (or a local minimum)

- the gradient at $\mathbf{x}^*$ vanishes
- the Hessian matrix at $\mathbf{x}^*$ is negative definite (positive definite)

Let's make the intuition concrete.

Theorem (First order necessary condition)

Let $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$ *be a function of two variables. Assume that* f *is differentiable at* $(x^*, y^*) \in D$.

If (x^*, y^*) *is a local maximum or a local minimum for the function* f *then* (x^*, y^*) *is a stationary pont, that is:*

$$f_x(x^*, y^*) = 0$$
$$f_y(x^*, y^*) = 0$$

In other words $\nabla f(x^*, y^*) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

Reminder on the definition of Stationary points

Definition (Stationary points)

Let $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$ *be a function of two variables. Suppose that* $f \in \mathcal{C}^1(D)$.

A point $(x_0, y_0) \in D$ *is a stationary point for* f *if*

$$f_x(x_0, y_0) = 0$$

$$f_y(x_0, y_0) = 0$$

that is the gradient vanishes at (x_0, y_0).

Local maxima and local minima are stationary points of the function f.

The converse implication is not true!

Example

Consider the function

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

We have that

$$\nabla f(x,y) = \begin{pmatrix} 2x \\ -2y \end{pmatrix}$$

Hence

$$\nabla f(0,0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

However $(0,0)$ *is not a maximum nor a minimum. The point* $(0,0)$ *is called a saddle point.*

Saddle point

Let $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$, let (x^*, y^*) and let $B_r \subseteq D$ be a disk centered at (x^*, y^*) with radius r. We say that (x^*, y^*) is a **saddle point** if it is a stationary point and there exist two points $(x_1, y_1) \in B_r$ and $(x_2, y_2) \in B_r$ such that

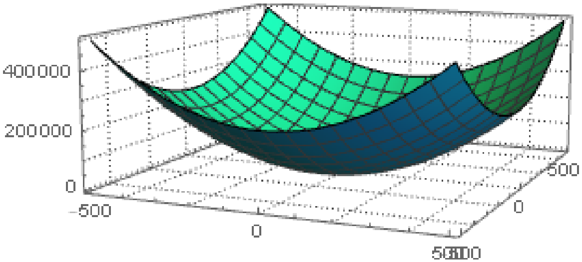
$$f(x_1, y_1) < f(x^*, y^*) \qquad f(x_2, y_2) > f(x^*, y^*)$$

Theorem (Second order sufficient conditions)

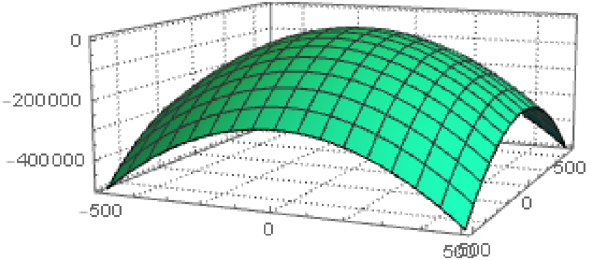
Let $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$ *be a function of two variables. Suppose that* $f \in \mathcal{C}^2(D)$ *and let* $H_f(x, y)$ *be the Hessian matrix of* f*. Let* $(x^*, y^*) \in D$ *be a stationary point for* f*.*

- *If* $H_f(x^*, y^*)$ *is negative definite then* (x^*, y^*) *is a local maximum*
- *If* $H_f(x^*, y^*)$ *is positive definite then* (x^*, y^*) *is a local minimum*
- *If* $H_f(x^*, y^*)$ *is indefinite then* (x^*, y^*) *is a saddle point*

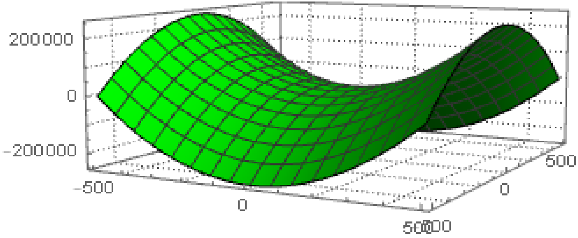
Positive definite: local minimum



Negative definite: local minimum



Indefinite: saddle point



What if $H_f(x^*, y^*)$ is only semidefinite?

In this case we have only partial information about the critical points:

Theorem

Let $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$ be a function of two variables. Suppose that $f \in \mathcal{C}^2(D)$ and let $H_f(x, y)$ be the Hessian matrix of f. Let $(x^*, y^*) \in D$ be a stationary point for f.

- *If $H_f(x^*, y^*)$ is negative semidefinite but not the null matrix, then (x^*, y^*) is either a local maximum or a saddle point.*
- *If $H_f(x^*, y^*)$ is positive semidefinite but not the null matrix then (x^*, y^*) is either a local minimum*
- *If $H_f(x^*, y^*)$ is the null matrix (x^*, y^*) the stationary point could be of any nature (local maximum, local minimum or saddle point)*

Be Careful:

This criterion cannot provide a precise information about the nature of the critical points, then if H is only semidefinite we need more investigation.

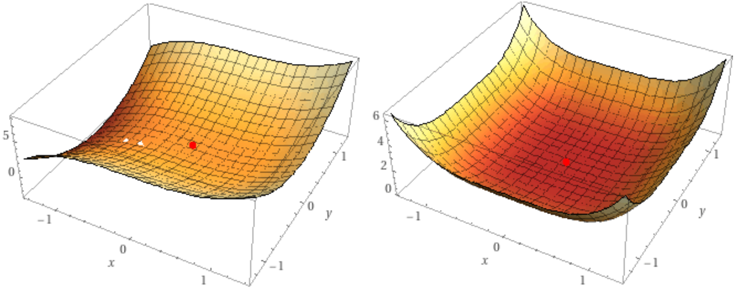


Figure: Left: $x^3 + y^2$; Right: $x^4 + y^4$