

Functions of several variables



Outline

Functions of several variables

Gradient, Hessian and level sets

Examples of functions that are not quadratic



Functions of several variables



Introduction: Functions of several variables

- ▶ In this section we define the **functions of n real variables**: they take a vector of $\mathbb{R}^n$ and return a real number.
- ▶ In optimization the quantity to minimize or maximize (a cost, a profit, a distance) almost always depends on **many** decisions at the same time.
- ▶ We fix the notation (vectors in bold, column vectors, transpose) that we will use in the whole course.



Function of several real variables

Definition 1: function of several real variables

The function:

$$f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}, \quad f : \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}}_{\mathbf{x}} \mapsto f(\mathbf{x})$$

*is called a (scalar) **function of n real variables** and $D \subseteq \mathbb{R}^n$ is its **domain**.*

Input, output and notation

- ▶ A function, commonly denoted f , of n real variables takes as **input** a vector of n real numbers, commonly represented by the (column) variable vector $\mathbf{x}$.
- ▶ It produces as **output** a real number, the value of the function, commonly denoted $f(\mathbf{x})$.
- ▶ Vectors are written in bold and are column vectors; the transpose of a vector $\mathbf{x}$ is the row vector:

$$\mathbf{x}' = (x_1 \quad x_2 \quad \dots \quad x_n)$$

Example: a function of three variables

Consider the function:

$$f : \mathbb{R}^3 \rightarrow \mathbb{R}, \quad f : \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \mapsto x_1 x_2 + x_3^2$$

At the point $\begin{pmatrix} 2 \\ -1 \\ 3 \end{pmatrix}$ the value of the function is:

$$f \left(\begin{pmatrix} 2 \\ -1 \\ 3 \end{pmatrix} \right) = 2 \cdot (-1) + 3^2 = -2 + 9 = 7$$

Summary

Functions of several variables:

- ▶ $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ maps a vector $\mathbf{x} \in D$ to a real number $f(\mathbf{x})$
- ▶ $D \subseteq \mathbb{R}^n$ is the **domain** of the function

Notation:

- ▶ Vectors are in bold and are column vectors
- ▶ $\mathbf{x}'$ is the transpose of $\mathbf{x}$, a row vector

Gradient, Hessian and level sets



Introduction: Gradient, Hessian and level sets

- ▶ In this section we define three tools to study a function of several variables: the **gradient**, the **Hessian** and the **level sets**.
- ▶ The gradient tells us in which direction the function grows fastest, the Hessian describes its **curvature**, the level sets let us draw the function in the plane.
- ▶ These tools are used throughout the course to recognise convex functions and to find the optimal solutions of optimization problems.



Definition 2: gradient

Given $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$, the **gradient** $\nabla f(\mathbf{x})$ of f at $\mathbf{x}$ is the (column) vector of first-order partial derivatives:

$$\nabla f(\mathbf{x}) := \begin{pmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{pmatrix}$$

Steepest ascent and steepest descent

- ▶ The gradient $\nabla f(\mathbf{x})$ (if it is nonzero) gives the direction of **steepest ascent** (maximum directional derivative among unit directions).
- ▶ $-\nabla f(\mathbf{x})$ gives the direction of **steepest descent**.

From first-order to second-order derivatives

The gradient collects the first-order partial derivatives.
Collecting the second-order partial derivatives we obtain a matrix: the **Hessian**.



Definition 3: Hessian

Given a function $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$, the **Hessian** $\nabla^2 f(\mathbf{x})$ of f at $\mathbf{x}$ is the matrix of second-order partial derivatives:

$$\nabla^2 f(\mathbf{x}) := \begin{pmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_2^2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n^2} \end{pmatrix}$$

Entries of the Hessian and symmetry

- ▶ The entry in row i and column j of the Hessian is obtained by differentiating $\frac{\partial f(\mathbf{x})}{\partial x_j}$ with respect to x_i .
- ▶ If the second-order partial derivatives of f are continuous (as for all the functions of this course), the order of differentiation does not matter:

$$\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} = \frac{\partial^2 f(\mathbf{x})}{\partial x_j \partial x_i}$$

and the Hessian is a **symmetric** matrix.



Local curvature

The Hessian $\nabla^2 f(\mathbf{x})$ describes the **local curvature** of f around $\mathbf{x}$, that is, how the gradient changes when moving away from $\mathbf{x}$.

Definition 4: level set

Given a function $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ and a constant $k \in \mathbb{R}$, the level set $L(k)$ of f at level k is:

$$L(k) := \{ \mathbf{x} \in D : f(\mathbf{x}) = k \}$$

Level sets and gradient

- ▶ A level set $L(k)$ is the set of points where the function has a constant value k .
- ▶ The gradient $\nabla f(\mathbf{x})$ (if nonzero) is **perpendicular** to the level set passing through $\mathbf{x}$.
- ▶ The gradient points toward **increasing** values of f .

Putting the three tools together

We now compute the gradient, the Hessian and the level sets of a simple quadratic function of two variables, and we draw them.



Example: a quadratic function (1/4)

Consider the quadratic function:

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}, \quad f : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto (x_1 - 2)^2 + (x_2 - 2)^2 + 1$$

The **gradient** of the function is:

$$\nabla f \left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 2(x_1 - 2) \\ 2(x_2 - 2) \end{pmatrix}$$

Example: a quadratic function (2/4)

For example:

- ▶ at the point $\begin{pmatrix} 3 \\ 2 \end{pmatrix}$, we have

$$\nabla f \left(\begin{pmatrix} 3 \\ 2 \end{pmatrix} \right) = \begin{pmatrix} 2(3-2) \\ 2(2-2) \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$$

- ▶ at the point $\begin{pmatrix} 2 \\ 3 \end{pmatrix}$, we have

$$\nabla f \left(\begin{pmatrix} 2 \\ 3 \end{pmatrix} \right) = \begin{pmatrix} 2(2-2) \\ 2(3-2) \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$$

Example: a quadratic function (3/4)

The **Hessian** of the function is:

$$\nabla^2 f \left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

Note that the Hessian is **constant** (it does not depend on the point).

Example: a quadratic function (4/4)

The **level sets** of the function are:

$$L(k) = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2 : (x_1 - 2)^2 + (x_2 - 2)^2 + 1 = k \right\}$$

with $k \in \mathbb{R}$, and therefore:

- ▶ for $k > 1$ they are circles centered at $\begin{pmatrix} 2 \\ 2 \end{pmatrix}$ with radius $\sqrt{k - 1}$;
- ▶ the level set $L(1)$ is the single point $\begin{pmatrix} 2 \\ 2 \end{pmatrix}$;
- ▶ the level sets with $k < 1$ are empty.

Drawing the function

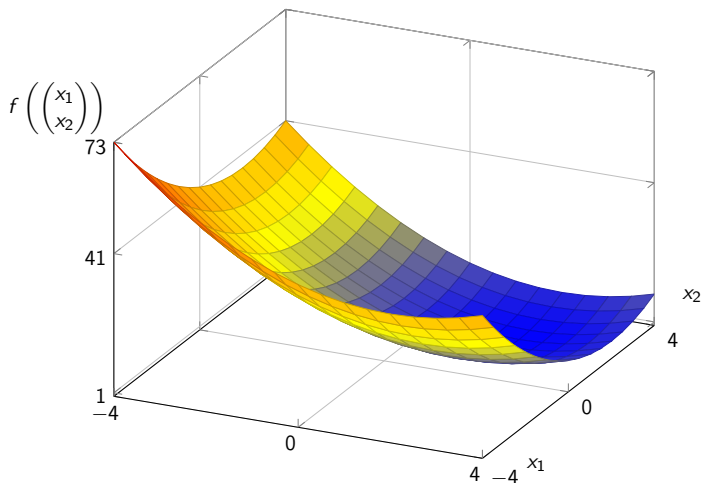
- ▶ We now draw the same function in two ways.
- ▶ First its **graph**, a surface in three dimensions.
- ▶ Then two of its **level sets** in the plane, together with the two gradients we have just computed.

Example: graph of the quadratic function (1/2)

The next figure shows the graph of the function in the domain $[-4, 4] \times [-4, 4]$:

- ▶ its lowest point is at $\begin{pmatrix} 2 \\ 2 \end{pmatrix}$, where $f = 1$;
- ▶ its highest point is at $\begin{pmatrix} -4 \\ -4 \end{pmatrix}$, where $f = 73$.

Example: graph of the quadratic function (2/2)

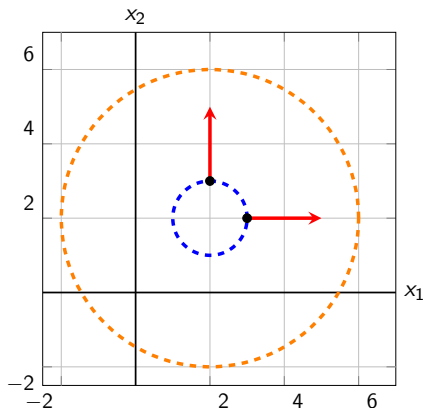


Example: level sets and gradients (1/2)

The next figure shows:

- ▶ the level set $L(2)$, circle of radius 1, in blue;
- ▶ the level set $L(17)$, circle of radius 4, in orange;
- ▶ the two gradients of the previous example, in red, drawn from the points $\begin{pmatrix} 3 \\ 2 \end{pmatrix}$ and $\begin{pmatrix} 2 \\ 3 \end{pmatrix}$ of the level set $L(2)$.

Example: level sets and gradients (2/2)



The gradient vectors are perpendicular to the level sets and point in the direction of maximum increase of the function (away from the center).

Summary (1/2)

Gradient:

- ▶ $\nabla f(\mathbf{x})$ is the column vector of the first-order partial derivatives
- ▶ It gives the direction of steepest ascent; $-\nabla f(\mathbf{x})$ that of steepest descent

Hessian:

- ▶ $\nabla^2 f(\mathbf{x})$ is the matrix of the second-order partial derivatives
- ▶ It is symmetric and describes the local curvature of f

Summary (2/2)

Level sets:

- ▶ $L(k) = \{ \mathbf{x} \in D : f(\mathbf{x}) = k \}$: the points where f takes the value k
- ▶ The gradient is perpendicular to the level set through $\mathbf{x}$ and points toward increasing values of f

Quadratic example:

- ▶ For $(x_1 - 2)^2 + (x_2 - 2)^2 + 1$ the level sets are circles centered at $\begin{pmatrix} 2 \\ 2 \end{pmatrix}$ and the Hessian is constant

Examples of functions that are not quadratic



Introduction: Examples of functions that are not quadratic

- ▶ For a quadratic function the Hessian is constant (chapter «Quadratic functions»).
- ▶ In general, the Hessian **depends on the point**: in this section we see it on two examples, a cubic function and an exponential function.
- ▶ We also check that the gradient is perpendicular to the level set even when the level set is not a circle.

Example: a cubic function (1/3)

Consider the function:

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}, \quad f : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto x_1^2 x_2$$

The **gradient** and the **Hessian** of the function are:

$$\nabla f \left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right) = \begin{pmatrix} 2 x_1 x_2 \\ x_1^2 \end{pmatrix}, \quad \nabla^2 f \left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right) = \begin{pmatrix} 2 x_2 & 2 x_1 \\ 2 x_1 & 0 \end{pmatrix}$$

Example: a cubic function (2/3)

At the point $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$, we have:

$$\nabla f \left(\begin{pmatrix} 1 \\ 2 \end{pmatrix} \right) = \begin{pmatrix} 2 \cdot 1 \cdot 2 \\ 1^2 \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$$

and

$$\nabla^2 f \left(\begin{pmatrix} 1 \\ 2 \end{pmatrix} \right) = \begin{pmatrix} 4 & 2 \\ 2 & 0 \end{pmatrix}$$

Example: a cubic function (3/3)

At the point $\begin{pmatrix} -1 \\ 3 \end{pmatrix}$, we have:

$$\nabla f \left(\begin{pmatrix} -1 \\ 3 \end{pmatrix} \right) = \begin{pmatrix} 2 \cdot (-1) \cdot 3 \\ (-1)^2 \end{pmatrix} = \begin{pmatrix} -6 \\ 1 \end{pmatrix}$$

and

$$\nabla^2 f \left(\begin{pmatrix} -1 \\ 3 \end{pmatrix} \right) = \begin{pmatrix} 6 & -2 \\ -2 & 0 \end{pmatrix}$$

The Hessian changes from point to point and, as expected, it is symmetric at every point.

From the cubic to the exponential function

In the next example we also draw a level set and check, with a computation, that the gradient is perpendicular to it.



Example: an exponential function (1/5)

Consider the function:

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}, \quad f : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto e^{x_1} + x_2^2$$

The **gradient** and the **Hessian** of the function are:

$$\nabla f \left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right) = \begin{pmatrix} e^{x_1} \\ 2x_2 \end{pmatrix}, \quad \nabla^2 f \left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right) = \begin{pmatrix} e^{x_1} & 0 \\ 0 & 2 \end{pmatrix}$$

Example: an exponential function (2/5)

- ▶ The level set $L(2)$ is the curve $e^{x_1} + x_2^2 = 2$, that is:

$$x_1 = \ln(2 - x_2^2) \quad \text{for } |x_2| < \sqrt{2}$$

- ▶ It contains the points $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} \ln 2 \\ 0 \end{pmatrix}$, since:

$$e^0 + 1^2 = 2 \quad \text{and} \quad e^{\ln 2} + 0^2 = 2$$

Example: an exponential function (3/5)

At these points:

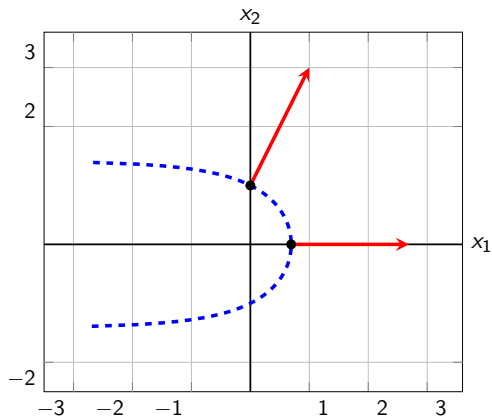
$$\nabla f \left(\begin{pmatrix} 0 \\ 1 \end{pmatrix} \right) = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \nabla f \left(\begin{pmatrix} \ln 2 \\ 0 \end{pmatrix} \right) = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$$

and

$$\nabla^2 f \left(\begin{pmatrix} 0 \\ 1 \end{pmatrix} \right) = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \quad \nabla^2 f \left(\begin{pmatrix} \ln 2 \\ 0 \end{pmatrix} \right) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

Example: an exponential function (4/5)

The level set $L(2)$ (in blue) and the two gradient vectors (in red):



Example: an exponential function (5/5)

- ▶ The tangent direction of the level set at $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ is $\begin{pmatrix} -2 \\ 1 \end{pmatrix}$: the derivative of $x_1 = \ln(2 - x_2^2)$ with respect to x_2 is $\frac{-2x_2}{2-x_2^2}$, equal to -2 at $x_2 = 1$.

- ▶ And indeed:

$$\begin{pmatrix} 1 & 2 \end{pmatrix} \begin{pmatrix} -2 \\ 1 \end{pmatrix} = -2 + 2 = 0$$

- ▶ So the gradient is perpendicular to the level set also when the level set is not a circle.

Gradient and Hessian: what to remember

- ▶ The gradient $\nabla f(\mathbf{x})$ is the column vector of first-order partial derivatives: it points in the direction of steepest ascent and is perpendicular to the level set through $\mathbf{x}$.
- ▶ The Hessian $\nabla^2 f(\mathbf{x})$ is the (symmetric) matrix of second-order partial derivatives and describes the local curvature of f .

Summary

Functions that are not quadratic:

- ▶ The Hessian depends on the point: for $x_1^2 x_2$ it is

$$\nabla^2 f \left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right) = \begin{pmatrix} 2x_2 & 2x_1 \\ 2x_1 & 0 \end{pmatrix}$$

- ▶ The Hessian is symmetric at every point

Gradient and level sets:

- ▶ For $e^{x_1} + x_2^2$ the level set $L(2)$ is not a circle, but the gradient is still perpendicular to it

Key messages

Summary (1/2)

1. Functions of several variables:

- ▶ $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ maps a column vector $\mathbf{x}$ to a real number $f(\mathbf{x})$

2. Gradient:

- ▶ $\nabla f(\mathbf{x})$: column vector of the first-order partial derivatives
- ▶ Direction of steepest ascent, perpendicular to the level sets



Summary (2/2)

3. Hessian:

- ▶ $\nabla^2 f(\mathbf{x})$: symmetric matrix of the second-order partial derivatives
- ▶ It describes the local curvature; it is constant for quadratic functions and depends on the point in general

4. Level sets:

- ▶ $L(k) = \{ \mathbf{x} \in D : f(\mathbf{x}) = k \}$, the points where f is equal to k