

Quadratic functions



Outline

Quadratic functions

Gradient, Hessian and level sets

Convexity of quadratic functions

Quadratic functions with a linear part

Quadratic functions

Introduction: Quadratic functions

- ▶ In this section we define the **quadratic functions**: functions of several variables built from a **square matrix** Q , written compactly as $\mathbf{x}' Q \mathbf{x}$.
- ▶ After the linear functions, they are the simplest functions of several variables, and they appear in many optimization problems.
- ▶ We will see that the matrix Q can always be assumed **symmetric**: this choice simplifies the gradient, the Hessian and the study of convexity in the next sections.

Quadratic function (1/2)

Definition 1: quadratic function

Given a square symmetric matrix

$$\underbrace{\begin{pmatrix} q_{11} & q_{12} & \dots & q_{1n} \\ q_{21} & q_{22} & \dots & q_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ q_{n1} & q_{n2} & \dots & q_{nn} \end{pmatrix}}_Q \in \mathbb{R}^{n \times n}$$

Quadratic function (2/2)

Definition 1: quadratic function

the associated (scalar) function

$$f : \mathbb{R}^n \rightarrow \mathbb{R}, \quad f : \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}}_{\mathbf{x}} \mapsto \underbrace{\sum_{i=1}^n \sum_{j=1}^n q_{ij} x_i x_j}_{\mathbf{x}' \mathbf{Q} \mathbf{x}}$$

*is called a **quadratic function**.*

The definition asks for a symmetric matrix.

Is this a restriction? We now see that every square matrix can be replaced by a symmetric one, without changing the function.

Symmetric matrix (1/2)

- If the square matrix of a quadratic function is not symmetric, it can be made symmetric by replacing each off-diagonal pair with its average:

$$q_{ij} \leftarrow \frac{q_{ij} + q_{ji}}{2}, \quad q_{ji} \leftarrow \frac{q_{ij} + q_{ji}}{2}, \quad \forall i, j \in \{1, 2, \dots, n\}, \quad i < j$$

Symmetric matrix (2/2)

This is valid because for every $i, j \in \{1, 2, \dots, n\}$ with $i < j$, we have:

$$\begin{aligned}q_{ij} x_i x_j + q_{ji} x_j x_i &= q_{ij} x_i x_j + q_{ji} x_i x_j \\&= (q_{ij} + q_{ji}) x_i x_j \\&= \frac{1}{2}(q_{ij} + q_{ji}) x_i x_j + \frac{1}{2}(q_{ij} + q_{ji}) x_i x_j \\&= \frac{q_{ij} + q_{ji}}{2} x_i x_j + \frac{q_{ij} + q_{ji}}{2} x_j x_i\end{aligned}$$

Therefore, the quadratic functions defined by $\mathbf{Q}$ and $\frac{1}{2}(\mathbf{Q} + \mathbf{Q}')$ are equivalent, since they yield the same value for every $\mathbf{x} \in \mathbb{R}^n$.

No loss of generality

Without loss of generality, we can assume that the matrix Q of a quadratic function is symmetric.

Example: equivalent matrices (1/2)

Consider the following non-symmetric 3×3 matrix:

$$\begin{pmatrix} 2 & -4 & 0 \\ 2 & 2 & 6 \\ 0 & -8 & 2 \end{pmatrix}$$

An equivalent quadratic function is obtained by using the following symmetric matrix:

$$\begin{pmatrix} 2 & \frac{2+(-4)}{2} & 0 \\ \frac{2+(-4)}{2} & 2 & \frac{-8+6}{2} \\ 0 & \frac{-8+6}{2} & 2 \end{pmatrix} = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}$$

Example: equivalent matrices (2/2)

Indeed, both matrices define the same function:

$$\begin{aligned}f\left(\begin{pmatrix}x_1 \\ x_2 \\ x_3\end{pmatrix}\right) &= 2x_1^2 - 4x_1x_2 + 2x_2x_1 + 2x_2^2 + 6x_2x_3 - 8x_3x_2 + 2x_3^2 \\&= 2x_1^2 - x_1x_2 - x_2x_1 + 2x_2^2 - x_2x_3 - x_3x_2 + 2x_3^2 \\&= 2x_1^2 + 2x_2^2 + 2x_3^2 - 2x_1x_2 - 2x_2x_3\end{aligned}$$

Summary

Quadratic function:

- ▶ Given a square matrix $\mathbf{Q} \in \mathbb{R}^{n \times n}$:
$$f(\mathbf{x}) = \sum_{i=1}^n \sum_{j=1}^n q_{ij} x_i x_j = \mathbf{x}' \mathbf{Q} \mathbf{x}$$

Symmetric matrix:

- ▶ $\mathbf{Q}$ and $\frac{1}{2}(\mathbf{Q} + \mathbf{Q}')$ define the same quadratic function
- ▶ Without loss of generality, $\mathbf{Q}$ is symmetric

Gradient, Hessian and level sets



Introduction: Gradient, Hessian and level sets

- ▶ In this section we compute the **gradient** and the **Hessian** of a quadratic function: they have a very simple form, $2 \mathbf{Q} \mathbf{x}$ and $2 \mathbf{Q}$.
- ▶ Then we study the **level sets**, the sets of points where the function is constant, and their shapes (ellipses, hyperbolas, pairs of lines).
- ▶ With examples in two variables we draw level sets, gradients and graphs: they help to see how a quadratic function behaves.

Observation 1: gradient of a quadratic function

The **gradient** (vector of partial derivatives) of a quadratic function is:

$$\nabla f(\mathbf{x}) = \begin{pmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^n (q_{1j} + q_{j1})x_j \\ \sum_{j=1}^n (q_{2j} + q_{j2})x_j \\ \vdots \\ \sum_{j=1}^n (q_{nj} + q_{jn})x_j \end{pmatrix}$$

Gradient (2/2)

Observation 1: gradient of a quadratic function

$$\nabla f(\mathbf{x}) = \begin{pmatrix} \sum_{j=1}^n (q_{1j} + q_{j1})x_j \\ \sum_{j=1}^n (q_{2j} + q_{j2})x_j \\ \vdots \\ \sum_{j=1}^n (q_{nj} + q_{jn})x_j \end{pmatrix} = \underbrace{\begin{pmatrix} 2 \sum_{j=1}^n q_{1j}x_j \\ 2 \sum_{j=1}^n q_{2j}x_j \\ \vdots \\ 2 \sum_{j=1}^n q_{nj}x_j \end{pmatrix}}_{2 \mathbf{Q} \mathbf{x}}$$

Proof

For every $k \in \{1, 2, \dots, n\}$, only the terms of the double sum with $i = k$ or $j = k$ contain the variable x_k :

- ▶ the terms $q_{kj} x_k x_j$ contribute $q_{kj} x_j$ to the derivative;
- ▶ the terms $q_{ik} x_i x_k$ contribute $q_{ik} x_i$;
- ▶ the term $q_{kk} x_k^2$, with $i = j = k$, contributes $2 q_{kk} x_k$, once in each sum.

Therefore:

$$\begin{aligned}\frac{\partial f(\mathbf{x})}{\partial x_k} &= \sum_{j=1}^n q_{kj} x_j + \sum_{i=1}^n q_{ik} x_i \\ &= \sum_{j=1}^n (q_{kj} + q_{jk}) x_j = 2 \sum_{j=1}^n q_{kj} x_j\end{aligned}$$

where the last equality holds since the matrix $\mathbf{Q}$ is symmetric ($q_{jk} = q_{kj}$).



Observation 2: Hessian of a quadratic function

The **Hessian** (matrix of second-order partial derivatives) of a quadratic function is:

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

Hessian (2/2)

Observation 2: Hessian of a quadratic function

$$\nabla^2 f(\mathbf{x}) = 2 \underbrace{\begin{pmatrix} q_{11} & q_{12} & \cdots & q_{1n} \\ q_{12} & q_{22} & \cdots & q_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ q_{1n} & q_{2n} & \cdots & q_{nn} \end{pmatrix}}_{2 \mathbf{Q}}$$

Proof

By the previous observation, for every $k \in \{1, 2, \dots, n\}$ we have

$$\frac{\partial f(\mathbf{x})}{\partial x_k} = 2 \sum_{j=1}^n q_{kj} x_j$$

whose derivative with respect to x_h is $2 q_{kh}$, for every $h \in \{1, 2, \dots, n\}$. □

Gradient and Hessian describe how the function changes.

We now look at the opposite question:
where does the function **not** change?

Definition 2: level sets of a quadratic function

The *level sets* of a quadratic function (the sets of points in which the function is *constant*) are:

$$\{ \mathbf{x} \in \mathbb{R}^n : \mathbf{x}' \mathbf{Q} \mathbf{x} = k \}, \quad \forall k \in \mathbb{R}$$

Shapes of the level sets (1/2)

Depending on the matrix $\mathbf{Q}$, the nonempty level sets can be:

- ▶ **ellipsoids** (ellipses in $\mathbb{R}^2$);
- ▶ **hyperboloids** (hyperbolas in $\mathbb{R}^2$);
- ▶ when some directions do not change the value of the function, **cylinders** (pairs of parallel lines in $\mathbb{R}^2$).

Shapes of the level sets (2/2)

For $k = 0$ there are also degenerate cases:

- ▶ the single point 0 in place of an ellipsoid;
- ▶ two lines crossing at the origin in $\mathbb{R}^2$ (a cone in $\mathbb{R}^n$) in place of a hyperboloid;
- ▶ a single line through the origin in $\mathbb{R}^2$ (a linear subspace in $\mathbb{R}^n$) in place of a cylinder.

The graph of the quadratic functions of the following examples is a **paraboloid**.

Gradient, Hessian and level sets

- ▶ For a quadratic function with symmetric matrix $\mathbf{Q}$, the gradient is $\nabla f(\mathbf{x}) = 2 \mathbf{Q} \mathbf{x}$ and the Hessian is the constant matrix $\nabla^2 f(\mathbf{x}) = 2 \mathbf{Q}$.
- ▶ At every point where the gradient is nonzero, the gradient is perpendicular to the level set through that point.

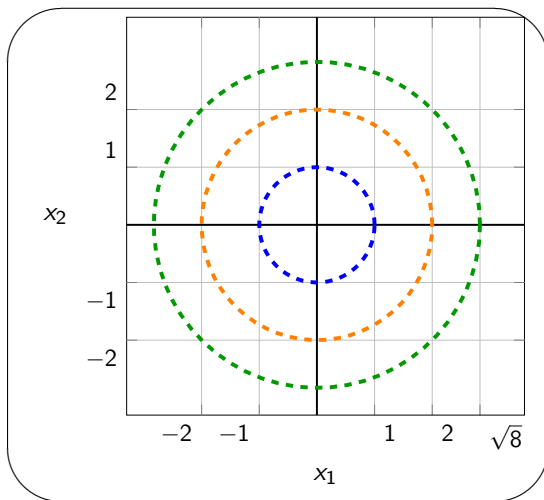
We now see these ideas on two quadratic functions
of two variables, the identity matrix and a general matrix:
level sets, gradients and graphs.

Example: level sets, identity matrix (1/3)

With $\mathbf{Q} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, the associated quadratic function is:

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}, \quad f : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto (x_1 \quad x_2) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1^2 + x_2^2$$

Example: level sets, identity matrix (2/3)



Example: level sets, identity matrix (3/3)

The figure shows:

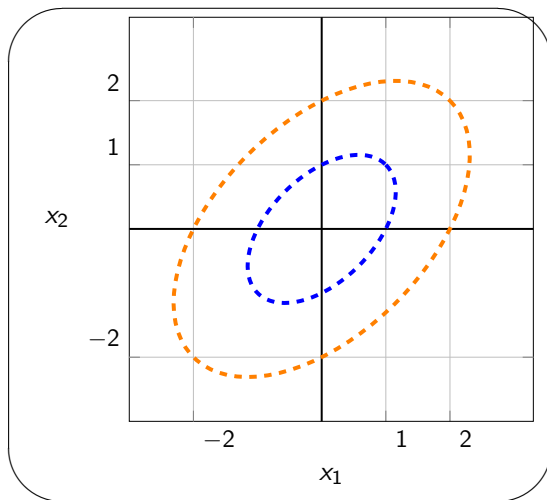
- ▶ Blue circle: level set $x_1^2 + x_2^2 = 1$ (radius 1)
- ▶ Orange circle: level set $x_1^2 + x_2^2 = 4$ (radius 2)
- ▶ Green circle: level set $x_1^2 + x_2^2 = 8$ (radius $\sqrt{8}$)

Example: level sets, general matrix (1/3)

With $\mathbf{Q} = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$, the associated quadratic function is:

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

Example: level sets, general matrix (2/3)



Example: level sets, general matrix (3/3)

The figure shows:

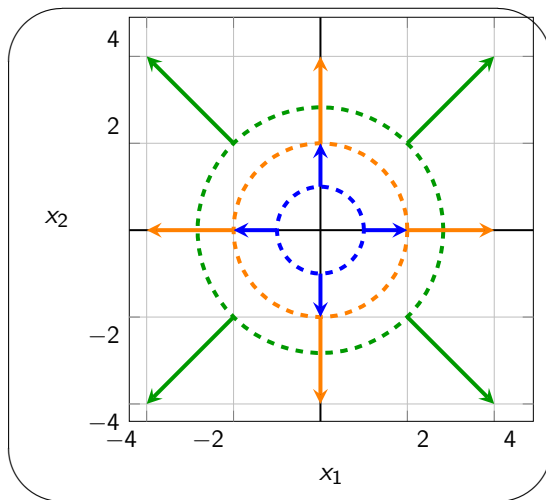
- ▶ Blue ellipse: level set $2 x_1^2 - 2 x_1 x_2 + 2 x_2^2 = 2$
- ▶ Orange ellipse: level set $2 x_1^2 - 2 x_1 x_2 + 2 x_2^2 = 8$

Example: gradients, identity matrix (1/4)

With $\mathbf{Q} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, the gradient of the quadratic function is:

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

Example: gradients, identity matrix (2/4)



Example: gradients, identity matrix (3/4)

The figure shows:

- ▶ Blue vectors (gradients with norm 2 on level set $x_1^2 + x_2^2 = 1$):

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

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

- ▶ Orange vectors (gradients with norm 4 on level set $x_1^2 + x_2^2 = 4$):

$$\nabla f \left(\begin{pmatrix} -2 \\ 0 \end{pmatrix} \right) = \begin{pmatrix} -4 \\ 0 \end{pmatrix}, \quad \nabla f \left(\begin{pmatrix} 0 \\ -2 \end{pmatrix} \right) = \begin{pmatrix} 0 \\ -4 \end{pmatrix},$$

$$\nabla f \left(\begin{pmatrix} 0 \\ 2 \end{pmatrix} \right) = \begin{pmatrix} 0 \\ 4 \end{pmatrix}, \quad \nabla f \left(\begin{pmatrix} 2 \\ 0 \end{pmatrix} \right) = \begin{pmatrix} 4 \\ 0 \end{pmatrix}$$

Example: gradients, identity matrix (4/4)

- ▶ Green vectors (gradients with norm $4\sqrt{2}$ on level set $x_1^2 + x_2^2 = 8$):

$$\nabla f \left(\begin{pmatrix} 2 \\ 2 \end{pmatrix} \right) = \begin{pmatrix} 4 \\ 4 \end{pmatrix}, \quad \nabla f \left(\begin{pmatrix} -2 \\ -2 \end{pmatrix} \right) = \begin{pmatrix} -4 \\ -4 \end{pmatrix},$$

$$\nabla f \left(\begin{pmatrix} -2 \\ 2 \end{pmatrix} \right) = \begin{pmatrix} -4 \\ 4 \end{pmatrix}, \quad \nabla f \left(\begin{pmatrix} 2 \\ -2 \end{pmatrix} \right) = \begin{pmatrix} 4 \\ -4 \end{pmatrix}$$

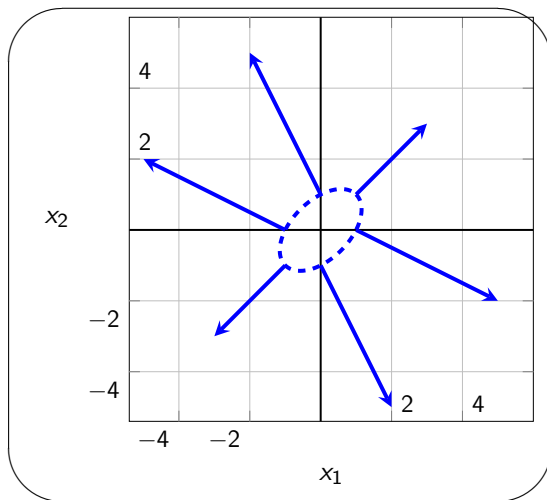
Each gradient is perpendicular to the level set through its point.

Example: gradients, general matrix (1/3)

With $\mathbf{Q} = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$, the gradient of the quadratic function is:

$$\nabla f \left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right) = \begin{pmatrix} 4 x_1 - 2 x_2 \\ -2 x_1 + 4 x_2 \end{pmatrix}$$

Example: gradients, general matrix (2/3)



Example: gradients, general matrix (3/3)

The figure shows:

- ▶ Blue ellipse: level set $2x_1^2 - 2x_1x_2 + 2x_2^2 = 2$
- ▶ Blue vectors (gradients on the level set with norm $2\sqrt{2}$ at diagonal points and norm $2\sqrt{5}$ at axis points):

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

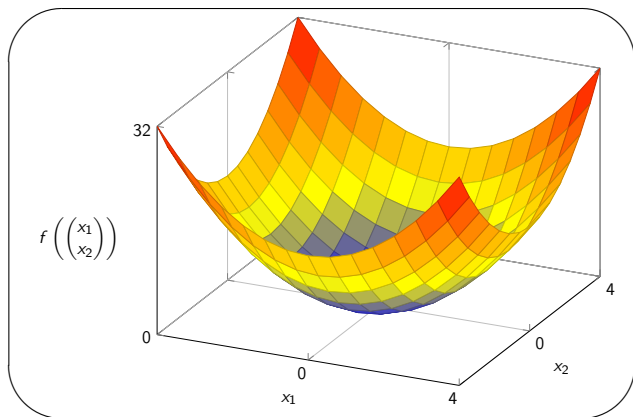
$$\nabla f \left(\begin{pmatrix} 1 \\ 0 \end{pmatrix} \right) = \begin{pmatrix} 4 \\ -2 \end{pmatrix}, \quad \nabla f \left(\begin{pmatrix} -1 \\ 0 \end{pmatrix} \right) = \begin{pmatrix} -4 \\ 2 \end{pmatrix},$$

$$\nabla f \left(\begin{pmatrix} 0 \\ 1 \end{pmatrix} \right) = \begin{pmatrix} -2 \\ 4 \end{pmatrix}, \quad \nabla f \left(\begin{pmatrix} 0 \\ -1 \end{pmatrix} \right) = \begin{pmatrix} 2 \\ -4 \end{pmatrix}$$

Also in this case, each gradient is perpendicular to the level set through its point.

Example: graph, identity matrix

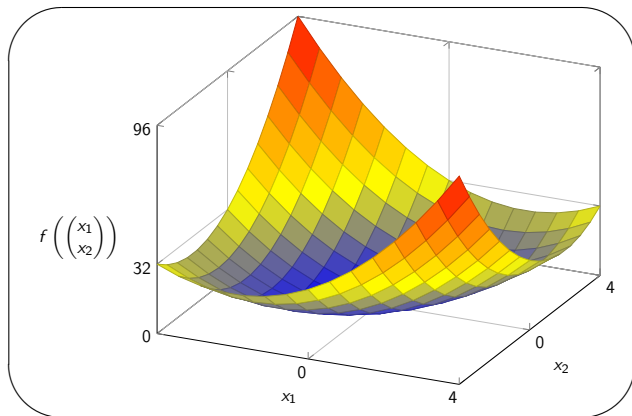
With $\mathbf{Q} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, we have $f\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) = x_1^2 + x_2^2$. Its graph is:



Example: graph, general matrix

With $\mathbf{Q} = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$, we have

$f\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) = 2x_1^2 - 2x_1x_2 + 2x_2^2$. Its graph is:



Summary (1/2)

Gradient and Hessian:

- ▶ Gradient: $\nabla f(\mathbf{x}) = 2 \mathbf{Q} \mathbf{x}$
- ▶ Hessian: $\nabla^2 f(\mathbf{x}) = 2 \mathbf{Q}$, a constant matrix

Level sets:

- ▶ $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{x}' \mathbf{Q} \mathbf{x} = k\}$, for every $k \in \mathbb{R}$
- ▶ Ellipsoids, hyperboloids or cylinders
- ▶ For $k = 0$: degenerate cases (a point, a cone, a linear subspace)

Summary (2/2)

Examples in two variables:

- ▶ Identity matrix: the level sets are circles
- ▶ $\mathbf{Q} = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$: the level sets are ellipses
- ▶ The gradient is perpendicular to the level set through its point
- ▶ The graphs are paraboloids

Convexity of quadratic functions



Introduction: Convexity of quadratic functions

- ▶ In this section we answer the question: **when is a quadratic function convex?**
- ▶ The answer depends only on the matrix $\mathbf{Q}$: the function is convex if and only if $\mathbf{Q}$ is **positive semidefinite**.
- ▶ We prove both implications from the definition of convex function, and we see an example of a convex and of a non-convex quadratic function.

Positive semidefinite matrix

Definition 3: positive semidefinite matrix

A square symmetric matrix $\mathbf{Q} \in \mathbb{R}^{n \times n}$ is **positive semidefinite** if:

$$\mathbf{x}' \mathbf{Q} \mathbf{x} \geq 0, \quad \forall \mathbf{x} \in \mathbb{R}^n$$

We now show that this property of the matrix is exactly what makes the quadratic function convex.

First: positive semidefinite $\Rightarrow$ convex.

PSD matrix implies convex function

Proposition 1

For every square matrix $\mathbf{Q}$, if $\mathbf{Q}$ is positive semidefinite then the associated quadratic function is convex in its domain.

- ▶ The proof starts from $f(\lambda \mathbf{p} + (1 - \lambda) \mathbf{w})$ and arrives at $\lambda f(\mathbf{p}) + (1 - \lambda) f(\mathbf{w})$ in a single chain.
- ▶ The key step: the difference between the two sides is $\lambda(1 - \lambda) (\mathbf{p} - \mathbf{w})' \mathbf{Q} (\mathbf{p} - \mathbf{w})$.

Proof

If the matrix $\mathbf{Q}$ is positive semidefinite, for every pair of points $\mathbf{p}, \mathbf{w} \in \mathbb{R}^n$ and every $\lambda \in [0, 1]$, we have:

$$\begin{aligned} & f(\lambda \mathbf{p} + (1 - \lambda) \mathbf{w}) \\ &= (\lambda \mathbf{p} + (1 - \lambda) \mathbf{w})' \mathbf{Q} (\lambda \mathbf{p} + (1 - \lambda) \mathbf{w}) \\ &= \lambda^2 \mathbf{p}' \mathbf{Q} \mathbf{p} + (1 - \lambda)^2 \mathbf{w}' \mathbf{Q} \mathbf{w} \\ &\quad + \lambda (1 - \lambda) \mathbf{p}' \mathbf{Q} \mathbf{w} + \lambda (1 - \lambda) \mathbf{w}' \mathbf{Q} \mathbf{p} \end{aligned}$$

Proof (continued)

$$\begin{aligned} &= \lambda \mathbf{p}' \mathbf{Q} \mathbf{p} + (1 - \lambda) \mathbf{w}' \mathbf{Q} \mathbf{w} \\ &\quad - \lambda (1 - \lambda) (\mathbf{p}' \mathbf{Q} \mathbf{p} + \mathbf{w}' \mathbf{Q} \mathbf{w} - \mathbf{p}' \mathbf{Q} \mathbf{w} - \mathbf{w}' \mathbf{Q} \mathbf{p}) \end{aligned}$$

$$\begin{aligned} &(\text{since } \lambda^2 = \lambda - \lambda (1 - \lambda) \text{ and} \\ &\quad (1 - \lambda)^2 = (1 - \lambda) - \lambda (1 - \lambda)) \end{aligned}$$

$$\begin{aligned} &= \lambda \mathbf{p}' \mathbf{Q} \mathbf{p} + (1 - \lambda) \mathbf{w}' \mathbf{Q} \mathbf{w} \\ &\quad - \lambda (1 - \lambda) (\mathbf{p} - \mathbf{w})' \mathbf{Q} (\mathbf{p} - \mathbf{w}) \end{aligned}$$

$$\begin{aligned} &\leq \lambda \mathbf{p}' \mathbf{Q} \mathbf{p} + (1 - \lambda) \mathbf{w}' \mathbf{Q} \mathbf{w} \\ &\quad (\text{since } \mathbf{Q} \text{ is positive semidefinite and } \lambda (1 - \lambda) \geq 0) \end{aligned}$$

$$= \lambda f(\mathbf{p}) + (1 - \lambda) f(\mathbf{w})$$

Proof (continued)

Therefore,

$$f(\lambda \mathbf{p} + (1 - \lambda) \mathbf{w}) \leq \lambda f(\mathbf{p}) + (1 - \lambda) f(\mathbf{w})$$

By nature, the domain $\mathbb{R}^n$ is a convex set. Accordingly, the function is convex in its domain. □

Sufficient and necessary condition

- ▶ We have shown that:

Q positive semidefinite $\implies$ convex quadratic function

so “ Q positive semidefinite” is a **sufficient condition** for “convex quadratic function” and “convex quadratic function” is a **necessary condition** for “ Q positive semidefinite”.

- ▶ The implication works also in the other direction, as stated by the following proposition.

Convex function implies PSD matrix

Proposition 2

For every quadratic function, if the function is convex in its domain then the associated square matrix $\mathbf{Q}$ is positive semidefinite.

- ▶ The proof starts from the inequality of the definition of convex function, expands its left-hand side and simplifies.
- ▶ What remains is $(\mathbf{p} - \mathbf{w})' \mathbf{Q} (\mathbf{p} - \mathbf{w}) \geq 0$, for every $\mathbf{p}$ and $\mathbf{w}$.

Proof

If the quadratic function is convex, for every pair of column vectors $\mathbf{p}, \mathbf{w} \in \mathbb{R}^n$ and every scalar $\lambda \in (0, 1)$, we have:

$$\begin{aligned} & (\lambda \mathbf{p} + (1 - \lambda) \mathbf{w})' \mathbf{Q} (\lambda \mathbf{p} + (1 - \lambda) \mathbf{w}) \\ & \leq \lambda \mathbf{p}' \mathbf{Q} \mathbf{p} + (1 - \lambda) \mathbf{w}' \mathbf{Q} \mathbf{w} \end{aligned}$$

where the left-hand side is:

$$\begin{aligned} & (\lambda \mathbf{p} + (1 - \lambda) \mathbf{w})' \mathbf{Q} (\lambda \mathbf{p} + (1 - \lambda) \mathbf{w}) \\ & = \lambda^2 \mathbf{p}' \mathbf{Q} \mathbf{p} + (1 - \lambda)^2 \mathbf{w}' \mathbf{Q} \mathbf{w} \\ & \quad + \lambda (1 - \lambda) \mathbf{p}' \mathbf{Q} \mathbf{w} + \lambda (1 - \lambda) \mathbf{w}' \mathbf{Q} \mathbf{p} \end{aligned}$$

Proof (continued)

This implies, rearranging and simplifying, that:

$$\begin{aligned} & \lambda (1 - \lambda) \mathbf{p}' \mathbf{Q} \mathbf{p} + \lambda (1 - \lambda) \mathbf{w}' \mathbf{Q} \mathbf{w} \\ & - \lambda (1 - \lambda) \mathbf{p}' \mathbf{Q} \mathbf{w} - \lambda (1 - \lambda) \mathbf{w}' \mathbf{Q} \mathbf{p} \geq 0 \end{aligned}$$

which, since $\lambda (1 - \lambda) > 0$, implies:

$$\underbrace{\mathbf{p}' \mathbf{Q} \mathbf{p} + \mathbf{w}' \mathbf{Q} \mathbf{w} - \mathbf{p}' \mathbf{Q} \mathbf{w} - \mathbf{w}' \mathbf{Q} \mathbf{p}}_{=(\mathbf{p}-\mathbf{w})' \mathbf{Q} (\mathbf{p}-\mathbf{w})} \geq 0$$

Proof (continued)

Every vector $\mathbf{x} \in \mathbb{R}^n$ can be written as $\mathbf{x} = \mathbf{p} - \mathbf{w}$ (for instance, with $\mathbf{p} = \mathbf{x}$ and $\mathbf{w} = 0$).

Therefore, $\mathbf{x}' \mathbf{Q} \mathbf{x} \geq 0$ for every $\mathbf{x} \in \mathbb{R}^n$, and the square matrix $\mathbf{Q}$ is positive semidefinite. □

Necessary and sufficient condition (1/2)

- We have shown that:

convex quadratic function $\implies \mathbf{Q}$ positive semidefinite

so “convex quadratic function” is a **sufficient condition** for “ $\mathbf{Q}$ positive semidefinite” and “ $\mathbf{Q}$ positive semidefinite” is a **necessary condition** for “convex quadratic function”.

Necessary and sufficient condition (2/2)

- So we finally have:

Q positive semidefinite $\iff$ convex quadratic function

so “ Q positive semidefinite” is a **necessary and sufficient condition** for “convex quadratic function” and vice versa.

Convexity of a quadratic function

Corollary 1

For every square matrix $\mathbf{Q}$, the associated quadratic function is convex in its domain if and only if $\mathbf{Q}$ is positive semidefinite.

Two remarks (1/2)

- ▶ With $\mathbf{Q}$ equal to the identity matrix, the quadratic function is

$$f(\mathbf{x}) = \mathbf{x}'\mathbf{x} = \|\mathbf{x}\|_2^2$$

the squared Euclidean norm, whose convexity is proved in the chapter «Convex and concave functions» directly from the definition.

- ▶ The identity matrix is positive semidefinite, since

$$\mathbf{x}'\mathbf{x} = x_1^2 + x_2^2 + \cdots + x_n^2 \geq 0$$

in agreement with the corollary.

Two remarks (2/2)

- ▶ When $\mathbf{x}'\mathbf{Q}\mathbf{x}$ can be written as a sum of squares of linear functions, as in the following example, the matrix $\mathbf{Q}$ is positive semidefinite.
- ▶ The convexity can also be proved directly from the definition, with the inequality

$$(\lambda a + (1 - \lambda)b)^2 \leq \lambda a^2 + (1 - \lambda)b^2$$

of the chapter «Convex and concave functions».

Example: a convex quadratic function (1/3)

With $\mathbf{Q} = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$, for every $\mathbf{x} \in \mathbb{R}^2$ we have:

$$\mathbf{x}' \mathbf{Q} \mathbf{x} = 2 x_1^2 - 2 x_1 x_2 + 2 x_2^2 = x_1^2 + x_2^2 + (x_1 - x_2)^2 \geq 0$$

Therefore, the matrix $\mathbf{Q}$ is positive semidefinite and, by the corollary, the quadratic function of the previous examples is convex in its domain.

Example: a convex quadratic function (2/3)

The same conclusion follows from the definition: for every pair of points $\mathbf{p}, \mathbf{w} \in \mathbb{R}^2$ and every $\lambda \in [0, 1]$, we have:

$$\begin{aligned} & f(\lambda \mathbf{p} + (1 - \lambda) \mathbf{w}) \\ &= (\lambda p_1 + (1 - \lambda) w_1)^2 + (\lambda p_2 + (1 - \lambda) w_2)^2 \\ &\quad + (\lambda(p_1 - p_2) + (1 - \lambda)(w_1 - w_2))^2 \\ &\leq \lambda p_1^2 + (1 - \lambda) w_1^2 + \lambda p_2^2 + (1 - \lambda) w_2^2 \\ &\quad + \lambda(p_1 - p_2)^2 + (1 - \lambda)(w_1 - w_2)^2 \\ &\quad \text{(by the inequality of the chapter} \\ &\quad \text{«Convex and concave functions», three times)} \end{aligned}$$

Example: a convex quadratic function (3/3)

$$\begin{aligned} &= \lambda(p_1^2 + p_2^2 + (p_1 - p_2)^2) \\ &\quad + (1 - \lambda)(w_1^2 + w_2^2 + (w_1 - w_2)^2) \\ &= \lambda f(\mathbf{p}) + (1 - \lambda)f(\mathbf{w}) \end{aligned}$$

Therefore,

$$f(\lambda \mathbf{p} + (1 - \lambda)\mathbf{w}) \leq \lambda f(\mathbf{p}) + (1 - \lambda)f(\mathbf{w})$$

By nature, the domain $\mathbb{R}^2$ is a convex set. Accordingly, the function is convex in its domain.

Example: a quadratic function that is not convex (1/2)

With $\mathbf{Q} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, the quadratic function is

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

and its nonempty level sets with $k \neq 0$ are hyperbolas. The matrix $\mathbf{Q}$ is not positive semidefinite, since:

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



Example: a quadratic function that is not convex (2/2)

By the corollary, the function is not convex in its domain. Indeed, with

$$\mathbf{p} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \mathbf{w} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \quad \lambda = \frac{1}{2}$$

the inequality of the definition of convex function does not hold:

$$\begin{aligned} f\left(\frac{1}{2}\mathbf{p} + \frac{1}{2}\mathbf{w}\right) &= f\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}\right) = 0 \\ &> -1 = \frac{1}{2}f(\mathbf{p}) + \frac{1}{2}f(\mathbf{w}) \end{aligned}$$

Summary

Positive semidefinite matrix:

- ▶ $\mathbf{x}' \mathbf{Q} \mathbf{x} \geq 0$ for every $\mathbf{x} \in \mathbb{R}^n$

Convexity of quadratic functions:

- ▶ $\mathbf{Q}$ positive semidefinite $\iff$ the quadratic function is convex
- ▶ Both implications are proved from the definition of convex function
- ▶ Sum of squares of linear functions: $\mathbf{Q}$ is positive semidefinite
- ▶ $x_1^2 - x_2^2$ is not convex: $\mathbf{Q}$ is not positive semidefinite

Quadratic functions with a linear part



Introduction: Quadratic functions with a linear part

- ▶ In optimization problems the objective is often a quadratic function **plus** a linear function and a constant.
- ▶ In this section we compute the gradient and the Hessian of these functions.
- ▶ We show that the linear part does **not** change convexity: the function is still convex if and only if $\mathbf{Q}$ is positive semidefinite.

Quadratic function with a linear part

In optimization problems we often find functions that are the sum of a quadratic function, a linear function and a constant: given a square symmetric matrix $\mathbf{Q} \in \mathbb{R}^{n \times n}$, a vector $\mathbf{c} \in \mathbb{R}^n$ and a scalar $d \in \mathbb{R}$,

$$g : \mathbb{R}^n \rightarrow \mathbb{R}, \quad g : \mathbf{x} \mapsto \mathbf{x}' \mathbf{Q} \mathbf{x} + \mathbf{c}' \mathbf{x} + d$$

$$= \sum_{i=1}^n \sum_{j=1}^n q_{ij} x_i x_j + \sum_{j=1}^n c_j x_j + d$$

Gradient and Hessian

Since the derivative of $\mathbf{c}'\mathbf{x} + d$ with respect to x_k is c_k , the gradient and the Hessian are:

$$\nabla g(\mathbf{x}) = 2 \mathbf{Q} \mathbf{x} + \mathbf{c}, \quad \nabla^2 g(\mathbf{x}) = 2 \mathbf{Q}$$

The Hessian does not depend on the linear part: it is the same as for $\mathbf{x}'\mathbf{Q}\mathbf{x}$.

The linear part is convex and concave

- ▶ The function $h(\mathbf{x}) = \mathbf{c}'\mathbf{x} + d$ is convex (and concave) in its domain, since the definition holds with equality.
- ▶ The proof starts from $h(\lambda\mathbf{p} + (1 - \lambda)\mathbf{w})$ and arrives at $\lambda h(\mathbf{p}) + (1 - \lambda)h(\mathbf{w})$ using only equalities.

Proof

For every pair of points $\mathbf{p}, \mathbf{w} \in \mathbb{R}^n$ and every $\lambda \in [0, 1]$, we have:

$$\begin{aligned} & h(\lambda \mathbf{p} + (1 - \lambda) \mathbf{w}) \\ &= \mathbf{c}'(\lambda \mathbf{p} + (1 - \lambda) \mathbf{w}) + d \\ &= \lambda \mathbf{c}' \mathbf{p} + (1 - \lambda) \mathbf{c}' \mathbf{w} + \lambda d + (1 - \lambda) d \\ &\quad (\text{since } \lambda + (1 - \lambda) = 1) \\ &= \lambda h(\mathbf{p}) + (1 - \lambda) h(\mathbf{w}) \end{aligned}$$

Therefore, both inequalities of the definitions of convex and concave function hold with equality. By nature, the domain $\mathbb{R}^n$ is a convex set. Accordingly, the function h is convex and concave in its domain. □

Convexity of g

- ▶ Since the sum of convex functions is convex (chapter «Convex and concave functions»), if the matrix $\mathbf{Q}$ is positive semidefinite, the function g is convex in its domain.
- ▶ Conversely, if g is convex, the quadratic function

$$\mathbf{x}'\mathbf{Q}\mathbf{x} = g(\mathbf{x}) + (-\mathbf{c}'\mathbf{x} - d)$$

is convex, as the sum of two convex functions, and the matrix $\mathbf{Q}$ is positive semidefinite by the corollary.

The linear part does not change convexity

A quadratic function $\mathbf{x}'\mathbf{Q}\mathbf{x}$ is convex in its domain if and only if the matrix $\mathbf{Q}$ is positive semidefinite.

Adding a linear function and a constant does not change its convexity.

Summary

Quadratic function with a linear part:

- ▶ $g(\mathbf{x}) = \mathbf{x}' \mathbf{Q} \mathbf{x} + \mathbf{c}' \mathbf{x} + d$
- ▶ $\nabla g(\mathbf{x}) = 2 \mathbf{Q} \mathbf{x} + \mathbf{c}$ and $\nabla^2 g(\mathbf{x}) = 2 \mathbf{Q}$

Convexity:

- ▶ $\mathbf{c}' \mathbf{x} + d$ is convex and concave
- ▶ g is convex if and only if $\mathbf{Q}$ is positive semidefinite

Key messages

Summary (1/2)

1. Quadratic functions:

- ▶ $f(\mathbf{x}) = \mathbf{x}' \mathbf{Q} \mathbf{x}$, with $\mathbf{Q}$ symmetric without loss of generality

2. Gradient, Hessian and level sets:

- ▶ $\nabla f(\mathbf{x}) = 2 \mathbf{Q} \mathbf{x}$ and $\nabla^2 f(\mathbf{x}) = 2 \mathbf{Q}$
- ▶ Level sets: ellipsoids, hyperboloids, cylinders (and degenerate cases for $k = 0$)
- ▶ The gradient is perpendicular to the level sets

Summary (2/2)

3. Convexity:

- ▶ The quadratic function is convex $\iff \mathbf{Q}$ is positive semidefinite

4. Linear part:

- ▶ $\mathbf{x}'\mathbf{Q}\mathbf{x} + \mathbf{c}'\mathbf{x} + d$ is convex $\iff \mathbf{Q}$ is positive semidefinite