

Optimization

Quadratic forms I



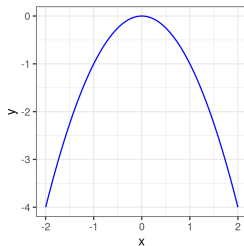
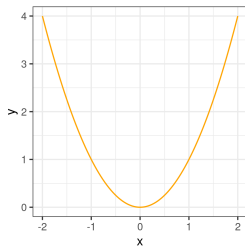
Learning goals

- TODO
- TODO

UNIVARIATE QUADRATIC FUNCTIONS

Consider a quadratic function $q : \mathbb{R} \rightarrow \mathbb{R}$

$$q(x) = a \cdot x^2 + b \cdot x + c, \quad a \neq 0.$$

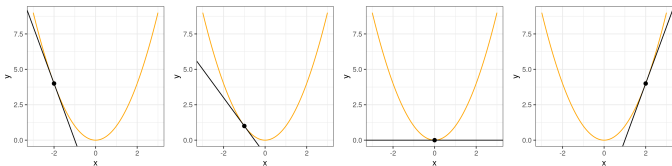


A quadratic function $q_1(x) = x^2$ (left) and $q_2(x) = -x^2$ (right).

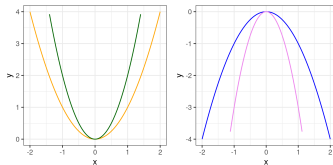
UNIVARIATE QUADRATIC FUNCTIONS

Basic properties can be read off easily:

- The **slope** of a tangent at a point $(\tilde{x}, q(\tilde{x}))$ is given by the first derivative $q'(\tilde{x}) = 2 \cdot a \cdot \tilde{x} + b$



- The **curvature** of q is given by $q''(x) = 2 \cdot a$.



$$q_1 = x^2 \text{ (orange)} \quad q_2 = 2x^2 \text{ (green)}, \quad q_3(x) = -x^2 \text{ (blue)}, \quad q_4 = -3x^2 \text{ (magenta)}$$

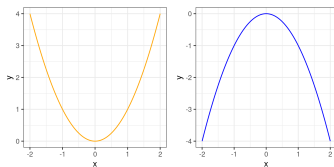
UNIVARIATE QUADRATIC FUNCTIONS

- **Convexity / Concavity:**

- If $a > 0$: q is convex, bounded from below and has a unique global **minimum**
- If $a < 0$: q is concave, bounded from above and has a unique global **maximum**

- The optimum x^* is

$$q'(x) = 0 \Leftrightarrow 2ax + b = 0 \Leftrightarrow x^* = \frac{-b}{2a}$$



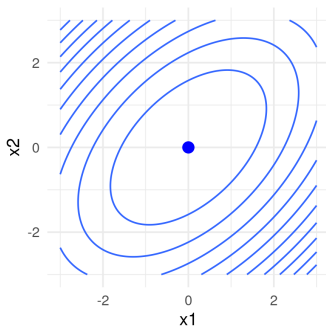
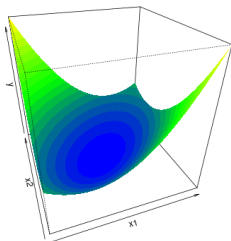
Left: $q_1(x) = x^2$ (convex). Right: $q_2(x) = -x^2$ (concave).

MULTIVARIATE QUADRATIC FUNCTIONS

A quadratic function $q : \mathbb{R}^d \rightarrow \mathbb{R}$ has the following form:

$$q(\mathbf{x}) = \mathbf{x}^\top \mathbf{A} \mathbf{x} + \mathbf{b}^\top \mathbf{x} + c,$$

with $\mathbf{A} \in \mathbb{R}^{d \times d}$ being a full-rank matrix, $\mathbf{b} \in \mathbb{R}^d$, and $c \in \mathbb{R}$.



MULTIVARIATE QUADRATIC FUNCTIONS

W.l.o.g. we can assume that \mathbf{A} is a **symmetric** matrix, i.e. $\mathbf{A}^\top = \mathbf{A}$, because there always exists a symmetric matrix $\tilde{\mathbf{A}}$ such that

$$q(\mathbf{x}) = \mathbf{x}^\top \mathbf{A} \mathbf{x} = \mathbf{x}^\top \tilde{\mathbf{A}} \mathbf{x} = \tilde{q}(\mathbf{x}) \quad \forall \mathbf{x}.$$

Justification: We can write $q(\mathbf{x})$ as

$$q(\mathbf{x}) = \mathbf{x}^\top \mathbf{A} \mathbf{x} = \frac{1}{2} \mathbf{x}^\top \underbrace{(\mathbf{A} + \mathbf{A}^\top)}_{\tilde{\mathbf{A}}_1} \mathbf{x} + \frac{1}{2} \mathbf{x}^\top \underbrace{(\mathbf{A} - \mathbf{A}^\top)}_{\tilde{\mathbf{A}}_2} \mathbf{x}$$

with $\tilde{\mathbf{A}}_1$ symmetric, and $\tilde{\mathbf{A}}_2$ anti-symmetric (i.e., $\tilde{\mathbf{A}}_2^\top = -\tilde{\mathbf{A}}_2$). Since $\mathbf{x}^\top \mathbf{A}^\top \mathbf{x}$ is a scalar, it is equal to its transposed and we get:

$$\begin{aligned} \mathbf{x}^\top (\mathbf{A} - \mathbf{A}^\top) \mathbf{x} &= \mathbf{x}^\top \mathbf{A} \mathbf{x} - \mathbf{x}^\top \mathbf{A}^\top \mathbf{x} = \mathbf{x}^\top \mathbf{A} \mathbf{x} - \left(\mathbf{x}^\top \mathbf{A}^\top \mathbf{x} \right)^\top \\ &= \mathbf{x}^\top \mathbf{A} \mathbf{x} - \mathbf{x}^\top \mathbf{A} \mathbf{x} = 0. \end{aligned}$$

Therefore, $q(\mathbf{x}) = \tilde{q}(\mathbf{x})$ with $\tilde{q}(\mathbf{x}) = \mathbf{x}^\top \tilde{\mathbf{A}} \mathbf{x}$ with $\tilde{\mathbf{A}} = \tilde{\mathbf{A}}_1$.

MULTIVARIATE QUADRATIC FUNCTIONS

- The gradient of q is given by

$$\nabla q(\mathbf{x}) = (\mathbf{A}^\top + \mathbf{A}) \mathbf{x} + \mathbf{b} = \nabla q(\mathbf{x}) = 2 \cdot \mathbf{A} \mathbf{x} + \mathbf{b} \in \mathbb{R}^d$$

(the last step follows from assuming \mathbf{A} to be symmetric).

The directional derivative in the direction of $\mathbf{v} \in \mathbb{R}^d$, $\|\mathbf{v}\| = 1$, is

$$\left. \frac{\partial q(\mathbf{x} + h \cdot \mathbf{v})}{\partial h} \right|_{h=0} = \left. \nabla q(\mathbf{x} + h\mathbf{v})^\top \mathbf{v} \right|_{h=0} = \nabla q(\mathbf{x})^\top \mathbf{v}$$

by using the chain rule.

MULTIVARIATE QUADRATIC FUNCTIONS

- The Hessian is given by

$$\nabla^2 q(\mathbf{x}) = (\mathbf{A}^\top + \mathbf{A}) = 2\mathbf{A} := \mathbf{H} \in \mathbb{R}^{d \times d},$$

(again assuming that \mathbf{A} is symmetric).

The curvature in the direction of $\mathbf{v} \in \mathbb{R}^d, \|\mathbf{v}\| = 1$, is

$$\begin{aligned} \left. \frac{\partial^2 q(\mathbf{x} + h \cdot \mathbf{v})}{\partial h^2} \right|_{h=0} &= \left. \frac{\partial [\nabla q(\mathbf{x} + h\mathbf{v})^\top \mathbf{v}]}{\partial h} \right|_{h=0} \\ &= \mathbf{v}^\top \nabla^2 q(\mathbf{x} + h\mathbf{v}) \mathbf{v} \Big|_{h=0} = \mathbf{v}^\top \mathbf{H} \mathbf{v}. \end{aligned}$$

MULTIVARIATE QUADRATIC FUNCTIONS

- If \mathbf{A} has full rank, there exists one unique stationary point (which may be a minimum, maximum, or a saddle point)

$$\begin{aligned}\nabla q(\mathbf{x}) &= 0 \\ 2 \cdot \mathbf{A}\mathbf{x} + \mathbf{b} &= 0 \\ \mathbf{x}^* &= -\frac{1}{2}\mathbf{A}^{-1}\mathbf{b}.\end{aligned}$$

