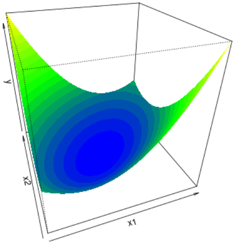
Optimization

Quadratic forms I



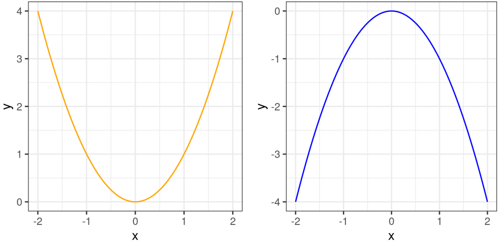
Learning goals

- Definition of quadratic forms
- Gradient, Hessian
- Convexity, concavity

UNIVARIATE QUADRATIC FUNCTIONS

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

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

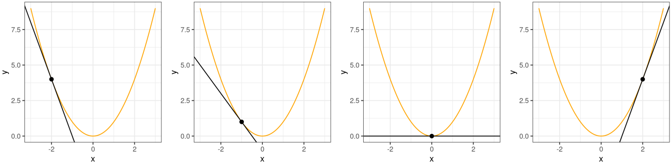


A quadratic function $q_1(x) = x^2$ (left) and $q_2(x) = -x^2$ (right).

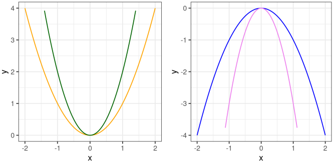
UNIVARIATE QUADRATIC FUNCTIONS

Basic properties:

- **Slope** of tangent at 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 $q''(x) = 2 \cdot a$.

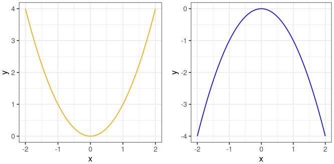


$q_1 = x^2$ (orange) $q_2 = 2x^2$ (green), $q_3(x) = -x^2$ (blue), $q_4 = -3x^2$ (magenta)

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

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



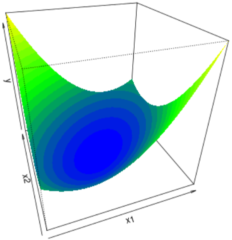
Left: $q_1(x) = x^2$ (convex). Right: $q_2(x) = -x^2$ (concave).

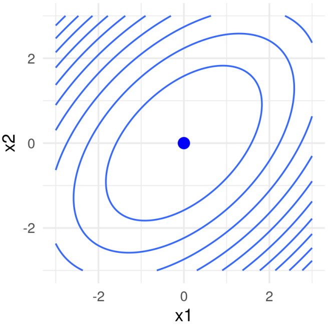
MULTIVARIATE QUADRATIC FUNCTIONS

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

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

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





MULTIVARIATE QUADRATIC FUNCTIONS

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

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

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

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

with $\tilde{\mathbf{A}}_1$ symmetric, $\tilde{\mathbf{A}}_2$ 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:

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

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

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

(using **A** symmetric).

Derivative in direction $\boldsymbol{v} \in \mathbb{R}^d$ is by chain rule

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

MULTIVARIATE QUADRATIC FUNCTIONS

- The Hessian is

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

  (using $\mathbf{A}$ symmetric).

  The curvature in the direction of $\boldsymbol{v} \in \mathbb{R}^d$ is

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

MULTIVARIATE QUADRATIC FUNCTIONS

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

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