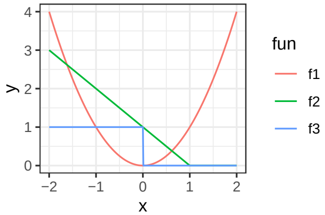
Optimization

Numerical differentiation



Learning goals

- TODO
- TODO

DIFFERENTIATION

We consider a function $f : \mathbb{R}^d \to \mathbb{R}$. We have seen that for proving necessary and sufficient conditions for optima, we need to compute a function's derivatives.

We distinguish between:

- Symbolic differentiation: Exact handling of mathematical expressions
- Numerical differentiation: Approximative calculation of the derivative

Numerical differentiation is often necessary if the derivative is not given, cannot be calculated or if the function itself is only indirectly available (e.g. via measured values).

NUMERICAL DIFFERENTIATION

- Approximation of differentiation $\frac{\partial f}{\partial x_i}$:
  - Newton's difference quotient for $\epsilon > 0$:

$$D_{\mathbf{x}}(\epsilon) = \frac{f(\mathbf{x} + \epsilon \cdot \boldsymbol{e}_i) - f(\mathbf{x})}{\epsilon},$$

  - Symmetric difference quotient for $\epsilon > 0$:

$$D_{\mathbf{x}}(\epsilon) = \frac{f(\mathbf{x} + \epsilon \cdot \boldsymbol{e}_i) - f(\mathbf{x} - \epsilon \cdot \boldsymbol{e}_i)}{2\epsilon}$$

- Symmetric approximation is more accurate, but f must be evaluated twice.
- **Essential question:** How should ϵ be chosen? In any case we require $\epsilon > \epsilon_m$ (ϵ should be larger than the machine epsilon).

NUMERICAL DIFFERENTIATION

From now on, let us consider univariate functions $f : \mathbb{R} \to \mathbb{R}$ only. We calculate the Taylor series at the location $\epsilon = 0$ with

$$f(x + \epsilon) = \sum_{k=0}^{\infty} \frac{\epsilon^k}{k!} \nabla^{(k)} f(x):$$

$$\begin{array}{llcl}
\text{(A)} & f(x+\epsilon) & = & f(x) + f'(x) \cdot \epsilon + \frac{1}{2!} f''(x) \cdot \epsilon^2 + \frac{1}{3!} f^{(3)}(x) \cdot \epsilon^3 \\
 & & + & \frac{1}{4!} f^{(4)}(x) \cdot \epsilon^4 + ... \\
\text{(B)} & f(x-\epsilon) & = & f(x) - f'(x) \cdot \epsilon + \frac{1}{2!} f''(x) \cdot \epsilon^2 - \frac{1}{3!} f^{(3)}(x) \cdot \epsilon^3 \\
 & & + & \frac{1}{4!} f^{(4)}(x) \cdot \epsilon^4 - ...
\end{array}$$

NUMERICAL DIFFERENTIATION

Thus the following applies for Newton's difference quotient

$$\frac{f(x+\epsilon)-f(x)}{\epsilon} \approx f'(x) + \underbrace{f''(x)\frac{\epsilon}{2} + \frac{1}{3!}f^{(3)}(x)\cdot\epsilon^2 + ...}_{\text{Error } \in \mathcal{O}(\epsilon)}$$

and for the symmetric difference quotient

$$\begin{aligned}\frac{f(x+\epsilon)-f(x-\epsilon)}{2\epsilon} &\approx \frac{(\mathrm{A})-(\mathrm{B})}{2\epsilon} = \frac{1}{2\epsilon}\left(2\cdot f'(x)\epsilon + 2\frac{1}{3!}f^{(3)}(x)\cdot\epsilon^3 + 2\frac{1}{5!}f^{(5)}(x)\epsilon^5\right) \\ &= f'(x) + \underbrace{c_1\epsilon^2 + c_2\epsilon^4 + ...}_{\text{Error } \in \mathcal{O}(\epsilon^2)}\end{aligned}$$

with $c_1 = \frac{1}{3!}f^{(3)}(x)$ and $c_2 = \frac{1}{5!}f^{(5)}(x)$.

NUMERICAL DIFFERENTIATION

We observe a **trade off** between the mathematical and a numerical error:

- The **mathematical error** is smaller for smaller ϵ:
  - The error is in $\mathcal{O}(\epsilon)$ for Newton's difference quotient
  - The error is in $\mathcal{O}(\epsilon^2)$ for the Symmetric difference quotient (so symmetrical is much more accurate)
- The **numerical error** error may explode for small ϵ: The problem is extremely bad conditioned for small ϵ (loss of significance and then division by ϵ).

Aim: When choosing ϵ, we have to find a compromise between mathematical and numerical error!

NUMERICAL DIFFERENTIATION

Let δ be a bound for the relative error in the calculation of $f(x)$ and $f(x+\epsilon)$ (that is, we only have access to $\tilde{f}(x), \tilde{f}(x+\epsilon)$ with relative error δ).

We estimate the error of the numerical differentiation:

$$\Bigg|\underbrace{\frac{\tilde{f}(x+\epsilon)-\tilde{f}(x)}{\epsilon}}_{\text{our approach}} - \underbrace{f'(x)}_{\text{true value}}\Bigg| \overset{(*)}{\leq} \left|f''(\zeta)\right| \frac{\epsilon}{2} + 2\delta \left|f(x)\right| \frac{1}{\epsilon} =: \frac{a \cdot \epsilon}{2} + \frac{2b}{\epsilon},$$

and minimize it by differentiation with respect to ϵ:

$$\tfrac{a\cdot\epsilon}{2} + \tfrac{2b}{\epsilon} \to \min_{\epsilon} \;\Leftrightarrow\; \tfrac{a}{2} - \tfrac{2b}{\epsilon^2} = 0 \;\Leftrightarrow\; \tfrac{a}{2} = \tfrac{2b}{\epsilon^2} \;\Leftrightarrow\; \epsilon^2 = \tfrac{4b}{a}$$

$$\epsilon = 2\sqrt{\tfrac{b}{a}} = 2\sqrt{\tfrac{\delta|f(x)|}{|f''(\zeta)|}}$$

NUMERICAL DIFFERENTIATION

$^{(*)}$ Proof: Since we only have access to approximate values $\tilde{f}(x+\epsilon)$, $\tilde{f}(x)$ with relative error δ, the following applies by definition (relative error):

$$\underbrace{\left|\frac{\tilde{f}(x+\epsilon) - f(x+\epsilon)}{f(x+\epsilon)}\right| \leq \delta,}_{(1)} \qquad \underbrace{\left|\frac{\tilde{f}(x) - f(x)}{f(x)}\right| \leq \delta}_{(2)}$$

Further, by using a Taylor expansion with an exact formulation of the remainder term $R_n(x; x_0)$ ("Lagrange form")

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!} \cdot (x - x_0)^k + \underbrace{\frac{f^{(n+1)}(\zeta)}{(n+1)!}(x - x_0)^{n+1},}_{R_n(x; x_0)} \quad \zeta \in [x_0; x]$$

we get

$$f(x+\epsilon) = f(x) + f'(x) \cdot \epsilon + \frac{f''(\zeta)}{2}\epsilon^2, \qquad \zeta \in [0, \epsilon]$$

$$\Leftrightarrow \quad \underbrace{\frac{f(x+\epsilon) - f(x)}{\epsilon} - f'(x) = f''(\zeta)\frac{\epsilon}{2}}_{(3)}$$

NUMERICAL DIFFERENTIATION

Thus we estimate the total error as follows:

$$
\begin{aligned}
& \left|\frac{\tilde{f}(x+\epsilon)-\tilde{f}(x)}{\epsilon}-f'(x)\right| \\
= \quad & \left|\frac{f(x+\epsilon)-f(x)}{\epsilon}-f'(x)+\frac{\tilde{f}(x+\epsilon)-f(x+\epsilon)}{\epsilon}+\frac{f(x)-\tilde{f}(x)}{\epsilon}\right| \\
\leq \quad & \underbrace{\left|\frac{f(x+\epsilon)-f(x)}{\epsilon}-f'(x)\right|}_{(3)}+\underbrace{\left|\frac{\tilde{f}(x+\epsilon)-f(x+\epsilon)}{f(x+\epsilon)\epsilon}f(x+\epsilon)\right|}_{(1)}+\underbrace{\left|\frac{f(x)-\tilde{f}(x)}{f(x)\epsilon}f(x)\right|}_{(2)} \\
\leq \quad & \left|f''(\zeta)\right|\frac{\epsilon}{2}+\delta\left|f(x+\epsilon)\right|\frac{1}{\epsilon}+\delta\left|f(x)\right|\frac{1}{\epsilon} \\
\approx \quad & \left|f''(\zeta)\right|\frac{\epsilon}{2}+2\delta\left|f(x)\right|\frac{1}{\epsilon}
\end{aligned}
$$

NUMERICAL DIFFERENTIATION

Popular choice of ϵ:

- $\epsilon \approx \sqrt{\delta}$ (if $|f(x)| \approx |f''(\zeta)|$ can be assumed) or
- $\epsilon = |x| \sqrt{\delta}$
  (For partial deriviatives $\frac{\partial f}{\partial x_i}$ we choose $|x_i| \sqrt{\delta}$ analogously.)

Without further knowledge it is often assumed:

$$\delta \approx \epsilon_m$$

More on the choice of ϵ and numerical differentiation: W. Press et al., *Numerical Recipes*, Chapter 5.7

DERIVATIVE-BASED VS. DERIVATIVE-FREE OPTIMIZATION

- If an objective function is assumed to be smooth and (good approximations to) derivatives are easy to compute, derivative-based optimization methods (i.e., methods requiring gradient-information) may be useful and usually yield fast convergence.
- However, if for example
  - smoothness cannot be assumed
  - the problem of computing derivatives is extremely bad conditionned
  - f is time-consuming to evaluate, or
  - f is in some way noisy

  optimization methods that do not require any derivative information have advantages.
- Those methods are referred to as **derivative-free algorithms**, and will be covered in later chapters.