Announcements

4.1 → Wed

exam 3

→ content cutoff today

→ still dropping lowest score

$f: \mathbb{R}^n \to \mathbb{R}$

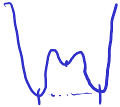
↳ continuous

↳ smooth

Goals

- Existence, uniqueness
- sensitivity
- methods 1D → nD

Existence/Uniqueness

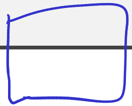


Terminology: global minimum / local minimum

Under what conditions on f can we say something about existence/uniqueness?

If $f : S \to \mathbb{R}$ is continuous on a closed and bounded set $S \subseteq \mathbb{R}^n$, then

a min exists



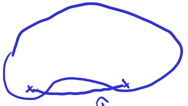
$f : S \to \mathbb{R}$ is called *coercive* on $S \subseteq \mathbb{R}^n$ if

$$\lim_{\|x\| \to \infty} f(x) = \infty$$

If f is coercive and continuous and S is closed, ...

a global min exists.

Convexity



$S \subseteq \mathbb{R}^n$ is called convex if for all $\boldsymbol{x}, \boldsymbol{y} \in S$ and all $0 \leq \alpha \leq 1$

$$\alpha \vec{x} + (1-\alpha)\vec{y}$$

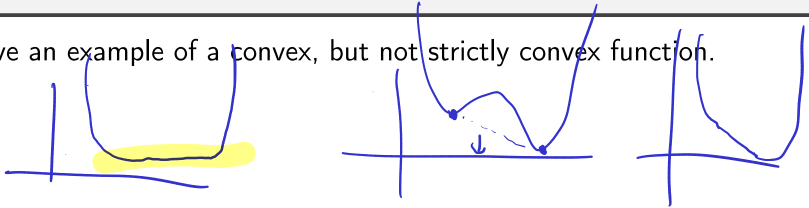
"convex combinations"

$f : S \to \mathbb{R}$ is called convex on $S \subseteq \mathbb{R}^n$ if for $\boldsymbol{x}, \boldsymbol{y} \in S$ and all $0 \leq \alpha \leq 1$

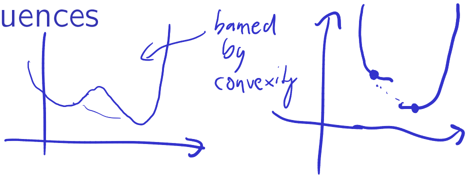
$$f(\alpha \vec{x} + (1-\alpha)\vec{y}) \leq \alpha f(\vec{x}) + (1-\alpha) f(\vec{y})$$

$<$ ← "strict"

Q: Give an example of a convex, but not strictly convex function.



Convexity: Consequences

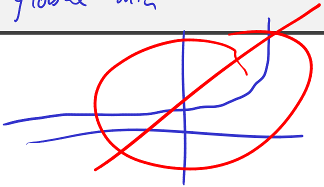


If f is convex, ...

- f is continuous
- x^* is local min $\Rightarrow$ x^* is global min

If f is strictly convex, ...

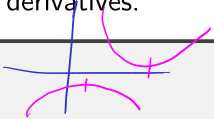
- x^* is a local min $\Rightarrow x^*$ a unique global min



Optimality Conditions

$x^T A x$

If we have found a candidate $\boldsymbol{x}^*$ for a minimum, how do we know it actually is one? Assume f is smooth, i.e. has all needed derivatives.



- 1D:

  necessary: $f'(x^*) = 0$

  sufficient: $f'(x^*) = 0$ and $f''(x^*) > 0$

  if f min then this

  if this then f min

  post-lecture fixes

- nD:

  necessary: $\nabla f = \vec{0}$

  sufficient: $\nabla f = \vec{0}$ and H_f pos. semi def.

  $\vec{x}^T H_f \vec{x} > 0$ ($\vec{x} \neq 0$)

  $H_f(x) = \begin{pmatrix} \partial_{11} & \cdots & \partial_{1n} \\ \partial_{n1} & & \partial_{nn} \end{pmatrix} f$

$$f(\vec{x} + h\vec{s}) = f(\vec{x}) + \cancel{\underbrace{\nabla f(x)}_{=0} \cdot (h\vec{s})} + \frac{h^2}{2} \vec{s}^T H_f(x) \vec{s} + O(h^3)$$

Optimization: Observations

Q: Come up with a hypothetical approach for finding minima.

Solve $\nabla f = 0$

Q: Is the Hessian symmetric?

yes, Schwarz's theorem

Q: How can we practically test for positive definiteness?

Cholesky

Sensitivity and Conditioning (1D)

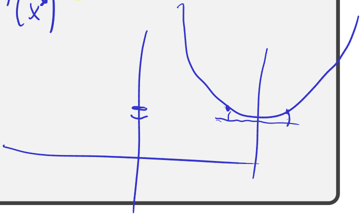
How does optimization react to a slight perturbation of the minimum?

Suppose $|f(\tilde{x}) - f(x^*)| < \text{tol}$ $\qquad \tilde{x} = x + h$

$$f(x^* + h) = f(x^*) + \underbrace{f'(x^*)}_{=0} h + f''(x^*)\frac{h^2}{2} + O(h^3)$$

$$\text{tol} > |f(x^* + h) - f(x^*)| \approx \left| f''(x^*) \frac{h^2}{2} \right|$$

$$h = |\tilde{x} - x^*| \leq \sqrt{\frac{2\,\text{tol}}{f''(x^*)}}$$



Sensitivity and Conditioning (nD)

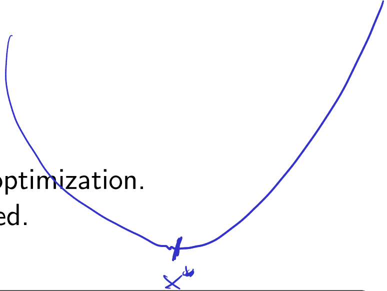
How does optimization react to a slight perturbation of the minimum?

$$f(\vec{x}+h\vec{s}) = f(\vec{x}) \quad + \frac{h^2}{2} \quad \vec{s}^T H_f(x)\vec{s} + O(h^3)$$

$$|h^2| \leq \frac{2\,tol}{\lambda_{min}(H_f(x^*))}$$

sensitivity in nD setting depends on conditioning of H_f

Unimodality



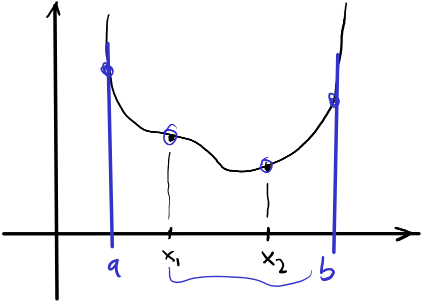
Would like a method like bisection, but for optimization.
In general: No invariant that can be preserved.
Need *extra assumption.*

f is called unimodal on an open interval, if there exists an x^* such that for all $x_1 < x_2$ in the interval

$$x_2 < x^* \Rightarrow f(x_1) > f(x_2)$$
$$x^* < x_1 \Rightarrow f(x_1) < f(x_2)$$

Golden Section Search

Suppose we have an interval with f unimodal:



Would like to maintain unimodality.

$f(x_1) > f(x_2) \Rightarrow$ reduce to $[x_1, b]$

$f(x_1) < f(x_2) \Rightarrow$ reduce to $[a, x_2]$

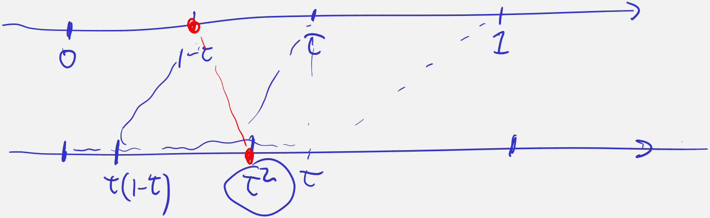
Golden Section Search: Efficiency

Where to put x_1, x_2?

Symmetry:

$$x_1 = a + (1-\tau)(b-a)$$
$$x_2 = a + \tau(b-a)$$

$$\tau^2 = 1-\tau$$
$$\tau = (\sqrt{5}-1)/2$$



Convergence rate?

linear

Newton's Method

Reuse the Taylor approximation idea, but for optimization.

Demo: Newton's Method in 1D [cleared]