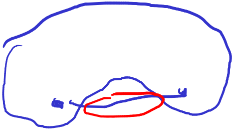
Convexity



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

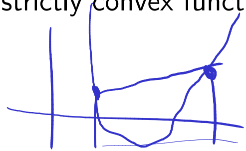
$$\alpha \vec{x} + (1-\alpha) y \in S \quad \leftarrow \text{convex combination}$$

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

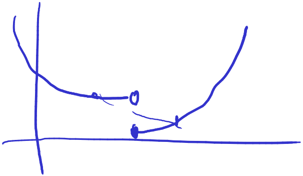
$$< \quad \leftarrow \text{strictly convex}$$

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

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



Convexity: Consequences

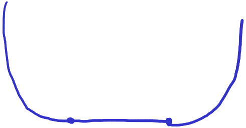


If f is convex, ...

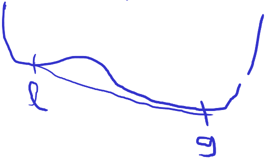
- f is continuous at interior points
- a local min is a global min

If f is strictly convex, ...

a local min is a unique global min



non-unique



Optimality Conditions

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.

In 1D:

$f'(x) = 0$ <u>necessary</u> condition

$f'(x) = 0$ and $f''(x) > 0$ sufficient condition

In nD

$\nabla f(x) = 0$ necessary

H_f positive definite $\quad H_f(x) = \begin{pmatrix} \partial_{x_1 x_1} f & \partial_{x_2 x_1} f \\ \partial_{x_1 x_2} f & \ddots \end{pmatrix}$

Optimization: Observations

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

$\nabla f(x) = 0$

Q: Is the Hessian symmetric?

Yes. (Schwartz thm). $\frac{\partial^2}{\partial x \partial y} f = \frac{\partial^2}{\partial y \partial x}$

Q: How can we practically test for positive definiteness?

Cholesky ($A = LL^T$)

In-Class Activity: Optimization Theory

In-class activity: Optimization Theory

Sensitivity and Conditioning (1D)

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

$$\underbrace{|f(x^*) - f(\tilde{x})|}_{\Delta f} < tol \quad (x^* \text{ is the true min}).$$

$$f(x^*+h) = f(x^*) + \underbrace{f'(x^*)\cdot h}_{=0} + \frac{f''(x^*)}{2}h^2 + \text{HOT}.$$

$$\Rightarrow \quad |\Delta f| = \left|\tfrac{1}{2} f''(x^*) h^2\right| < tol.$$

$$\Rightarrow \quad \underbrace{|\tilde{x} - x^*|}_{8} = |h| \le \sqrt{2 \cdot \underbrace{tol}_{16} / f''(x^*)}$$

half as many digits.

$\nabla f = 0$

Sensitivity and Conditioning (nD)

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

$$f(x^* + hs) = f(x^*) + h \underbrace{\nabla f(x^*)}_{=0} s + \frac{h^2}{2} s^T H_f(x^*) s + \text{HOT}.$$

direction $\|s\| = 1$

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

conditioning depends on H_f.

Unimodality

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

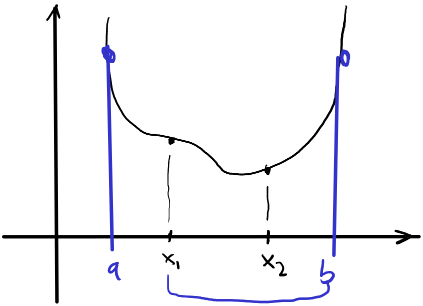
f is called unimodal if $x_1 < x_2$

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

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

Golden Section Search

Suppose we have an interval with f unimodal:



Would like to maintain unimodality.

→ Pick x_1, x_2 inside bracket $(x_1 < x_2)$
If $f(x_1) > f(x_2)$ reduce to (x_1, b)
If $f(x_1) < f(x_2)$ reduce to (a, x_2)

Golden Section Search: Efficiency

Where to put x_1, x_2?

$$x_1 = a + (1-\alpha)(b-a)$$
$$x_2 = a + \alpha\,(b-a)$$
$$\alpha^2 = 1-\alpha$$
$$\alpha = (\sqrt{5}-1)/2 \rightarrow \text{golden section search}$$

Convergence rate?

Linear

Demo: Golden Section Proportions [cleared]

Newton's Method

Reuse the Taylor approximation idea, but for optimization.

$$f(x+h) \approx f(x) + f'(x)h + f''(x)\frac{h^2}{2} := \hat{f}(h)$$

- approx f with $\hat{f}$ at x_k
- min $\hat{f}(h)$ to get x_{k+1}

$$\hat{f}'(h) = 0 = f'(x_k) + f''(x_k)\cdot h$$

$$\Rightarrow h = -\frac{f'(x_k)}{f''(x_k)}$$ (solving $f'(x)=0$ with Newton).

$\Rightarrow$ quadratic.

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

In-Class Activity: Optimization Methods

In-class activity: Optimization Methods

Steepest Descent

Given a scalar function $f : \mathbb{R}^n \to \mathbb{R}$ at a point $\boldsymbol{x}$, which way is down?

Direction of steepest descent: $-\nabla f$

"Line search" e.g. Golden section.

1. x_0 = init guess
2. $s_k = -\nabla f(x_k)$
3. $\min f(x_k + \alpha_k s_k)$
4. $x_{k+1} = x_k + \alpha_k s_k$

Demo: Steepest Descent [cleared]

Steepest Descent: Convergence

Consider quadratic model problem:

$\nabla f = Ax + c$

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T A\boldsymbol{x} + \boldsymbol{c}^T\boldsymbol{x}$$

where A is SPD. (A good model of f near a minimum.)

$e_k = x_k - x^*$, then

$$\|e_{k+1}\|_A = \sqrt{e_{k+1}^T A e_{k+1}} = \underbrace{\frac{\sigma_{max}(A) - \sigma_{min}(A)}{\sigma_{max}(A) + \sigma_{min}(A)}}_{\displaystyle \frac{\kappa(A)-1}{\kappa(A)+1}} \|e_k\|_A$$

$\Rightarrow$ linear convergence.