Convex functions

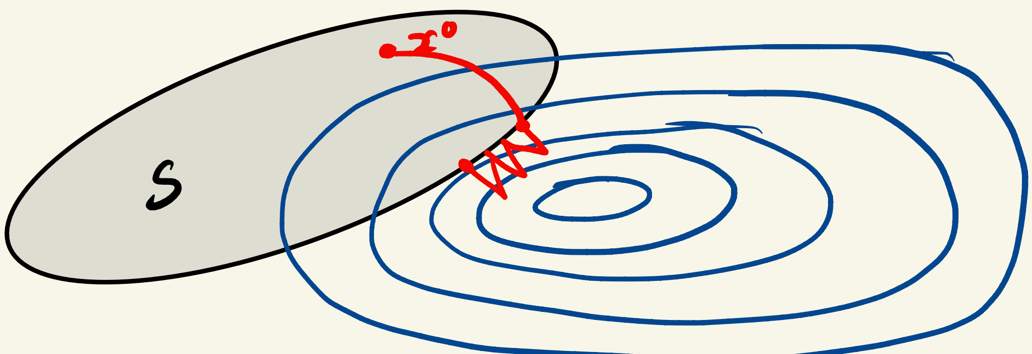
- Projected gradient descent (demo)
- convex function
- 1st/2nd order characterization of convex function

Projected Gradient descent

$$\min_x \quad f(x) \quad \text{subject to} \quad x \in S.$$

Projected gradient descent

$$x^{(k+1)} = \mathrm{proj}_S \left(x^{(k)} - \alpha^{(k)} \nabla f(x^{(k)}) \right)$$



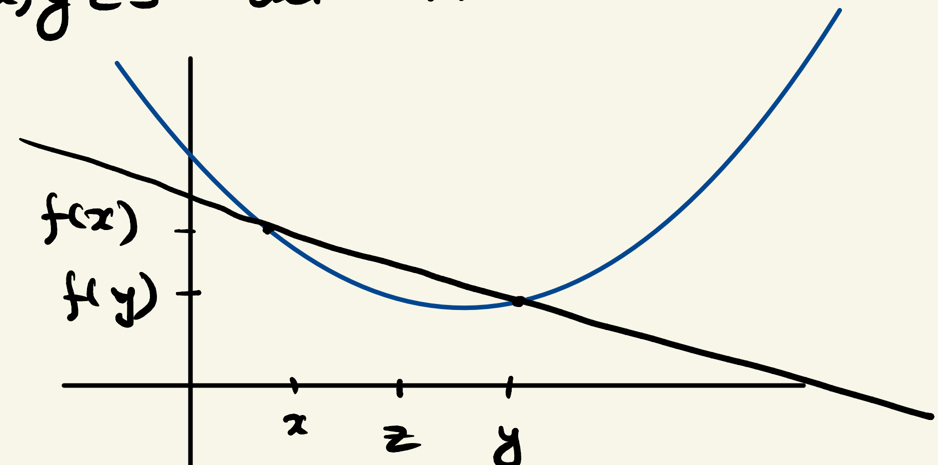
$$f(x^k) - f(x^k + \alpha d^k) \geq -\mu \alpha \| \widetilde{\nabla f(x_k)} \|_2^2$$

Convex Functions

- A function $f: S \to \mathbb{R}$ defined on a convex set $S \subseteq \mathbb{R}^n$ is called convex if

$$f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y)$$

for any $x, y \in S$ and $\lambda \in [0,1]$.



$z = \lambda x + (1-\lambda) y$. $\quad f(z) \leq \lambda f(x) + (1-\lambda) f(y)$.

Convex functions

- f is strictly convex if for all $x, y \in S$

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y)$$

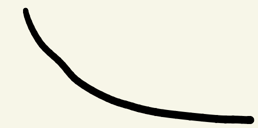
for all $\lambda \in (0,1)$.

The function is not flat.

- f is concave if $-f$ is convex.

Examples

Convex functions



- affine : $a^T x + b$ for any $a \in \mathbb{R}^n$, $b \in \mathbb{R}$.
- exponential function : $e^{\alpha x}$ for any $\alpha \in \mathbb{R}$
- powers : x^α for scalar $x \geq 0$, when $\alpha \leq 0$ or $\alpha \geq 1$.
- absolute value : $|x|^\alpha$ for all $\alpha \geq 1$.
- norm

Concave functions.

- affine functions (both concave and convex)
- power x^α for scalar $x \geq 0$, when $0 \leq \alpha \leq 1$
- logarithm : $\log(x)$ for $x \geq 0$.

Affine functions.

Let $f(x) = a^T x + b$. To show convexity, take $x, y \in \mathbb{R}^n$ and $\lambda \in [0, 1]$. Then

$$f(\lambda x + (1-\lambda) y) = a^T(\lambda x + (1-\lambda) y) + b.$$
$$= \lambda a^T x + (1-\lambda) a^T y + (1-\lambda) b + \lambda b$$
$$= \lambda f(x) + (1-\lambda) f(y).$$

Thus, convexity of f follows.

Examples

Affine functions are convex and concave

norms are convex.

- $\mathbb{R}^n$
  - affine function: $f(x) = a^T x + b$
  - norms: $\|x\|_p = \left(\sum_{i=1}^{n} x_i^p\right)^{1/p}$ for $p \geq 1$.

    $\|x\|_\infty = \max_i |x_i|$

- $\mathbb{R}^{m \times n}$ (matrices).
  - affine function: $f(X) = Tr(A^T X) + b$

    $= \sum_{j=1}^{n} \sum_{i=1}^{m} a_{ij} x_{ij} + b$.

  - spectral norm: (max singular value).

    $f(X) = \|X\|_2 = \sigma_{max}(X) = \left(\lambda_{max}(X^T X)\right)^{1/2}$

Jensen's inequality

Generalization to convex combination of any numbers of vectors.

Let $f: C \to R$ be convex where $C \subseteq R^n$ is convex.

Then for any $x_1, \dots x_K \in C$:

$$f\left(\sum_{i=1}^{K} \lambda_i x_i\right) \leq \sum_{i=1}^{K} \lambda_i f(x_i)$$

with $\lambda_i \geq 0, \sum_{i=1}^{K} \lambda_i = 1$.

In probability theory: if X is a random variable

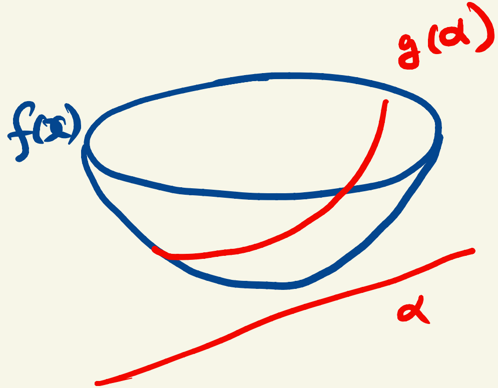
$$f(E(x)) \leq E(f(x))$$

Restriction to lines.

The function $f: S \to \mathbb{R}$ is convex if and only if

$$g(\alpha) = f(x + \alpha d)$$

is convex over $\mathbb{R}$ for all $x \in S$ and $d \in \mathbb{R}^n$.



f convex $\iff$ function in one variable convex.

Example

① $f(x) = \frac{1}{2}x^TAx + b^Tx + c$ is convex $\iff$ $A \succ 0$

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

$$= \frac{1}{2}x^TAx + \alpha x^TAd + \alpha^2 d^TAd + b^Tx + \alpha b^Td + c$$

$$= \alpha^2 d^TAd + \alpha(x^TAd + b^Td) + \left(\frac{1}{2}x^TAx + b^Tx + c\right)$$

② $f(X) = \log\det(X)$. $f: S^n_{++} \to \mathbb{R}$

$$f(X+\alpha D) = \log\det(X+\alpha D)$$

$$= \log\det(X) + \log\det(I + \alpha X^{-1}D)$$

$$= \log\det(X) + \sum_{i=1}^{n} \log(1+\alpha\lambda_i)$$

where λ_i are eigenvalues of $X^{-1}D$.

$\Rightarrow f(X+\alpha D)$ is concave in α $\Rightarrow$ f is concave.

First order characterization.

$f: S \to \mathbb{R}$ is differentiable if the gradient (open)

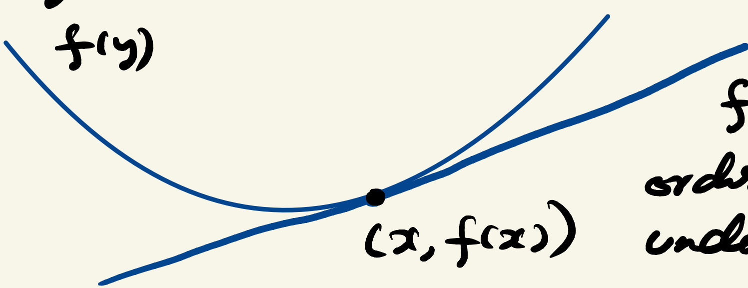
$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_n}(x) \right)$$

exists for all $x \in S$.

1st order condition: Let $f: S \to \mathbb{R}$ be continuously differentiable with $S \subseteq \mathbb{R}^n$ convex. Then f is convex

iff

$$f(x) + \nabla f(x)^T (y-x) \leq f(y), \quad \text{for any } x, y \in S.$$



$f(x) + \nabla f(x)^T (y-x)$ (first order approximation) is a global underestimator.

Second order characterization

$f: S \to \mathbb{R}^n$ ($S \subseteq \mathbb{R}^n$ open) is twice differentiable if $\nabla^2 f(x)$ exists for all $x \in S$.

2nd order characterization: for twice differentiable f with convex S:

① f is convex iff $\nabla^2 f(x) \succeq 0$ for all $x \in S$.

② (sufficient cond for strict convexity). If $\nabla^2 f(x) \succ 0$ for all $x \in S$, then f is strictly convex