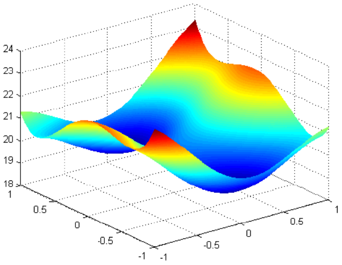
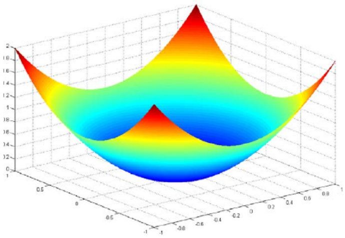
Objectif du cours

> " The great watershed in optimization isn't between linearity and nonlinearity, but convexity and nonconvexity."
>
> — R. Rockafellar, SIAM Review 1993



Non-Convex Optimization



Convex Optimization

- La frontière en optimisation se trouve entre le convexe et le non-convexe. Les problèmes d'optimisation convexe s'étudie de manière plus poussés.

Ensemble et fonctions convexes

24 septembre 2024

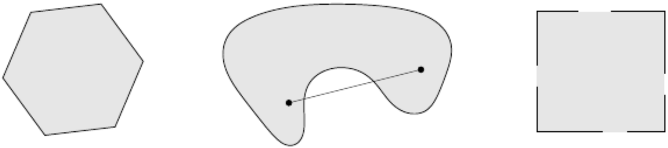
Ensemble convexe

Définition :
soit $\mathcal{X}$ un sous-ensemble de $\mathbb{R}^d$ *ie* $\mathcal{X} \subset \mathbb{R}^d$, on dit que $\mathcal{X}$ est convexe si $\forall \mathbf{x}, \mathbf{y} \in \mathcal{X}$

$$\forall \lambda \in [0, 1] \qquad \lambda \mathbf{x} + (1 - \lambda)\mathbf{y} \in \mathcal{X}$$

Illustration

le segment reliant $\mathbf{x}$ à $\mathbf{y}$ se trouve à l'intérieur de l'ensemble $\mathcal{X}$



le premier ensemble est convexe, les deux autres ne le sont pas.

Méthodologie

Comment prouver la convexité de $\mathcal{X}$

- appliquer la définition

$$\mathbf{x}, \mathbf{y} \in \mathcal{X}, \quad \lambda \in [0,1] \Rightarrow \text{ calculer } \lambda\mathbf{x} + (1-\lambda)\mathbf{y}$$

- simplier le résultat/utiliser les propriétés de $\mathcal{X}$ pour montrer que $\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in \mathcal{X}$

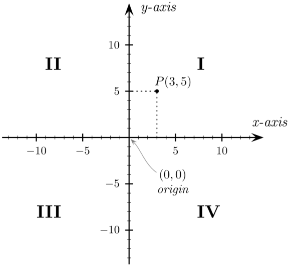
Applications aux ensembles : $\mathbb{R}^d$, $\mathbb{R}^d_+$, hyperplan

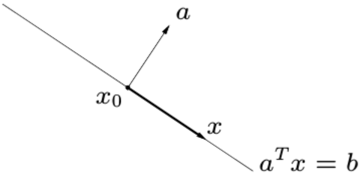
Exemples

Ensembles

- $\mathbb{R}^d$
- Orthant non-negatif de $\mathbb{R}^d$: $\mathbb{R}^d_+$.
- hyperplan :$\{\mathbf{x} \in \mathbb{R}^d : \mathbf{a}^\top \mathbf{x} = b\}$
- demi-espace : :$\{\mathbf{x} \in \mathbb{R}^d : \mathbf{a}^\top \mathbf{x} \leq b\}$

Illustrations



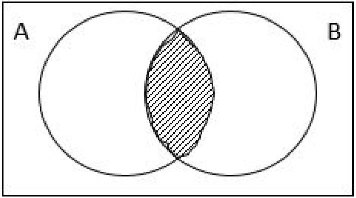


Opérations préservant la convexité

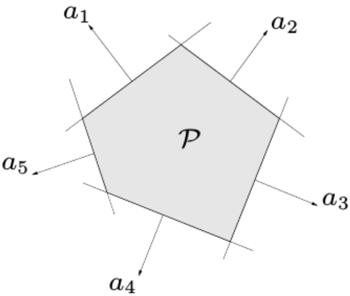
Intersection

si $\mathcal{X}_k$, pour tout k sont des ensembles convexes alors leus intersections $\cap_{k=1}^{K} \mathcal{X}_k$ est convexe

Illustration



Intersection de disques



Intersection de demi-espaces

Opérations préservant la convexité

Produit cartésien

si $\mathcal{X}_k \subset \mathbb{R}^{n_k}$, pour tout $k = 1, \cdots, M$ sont des ensembles convexes alors

$$\mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \times \mathcal{X}_M = \{(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_M) : \mathbf{x}_k \in \mathcal{X}_k\}$$

est convexe.

Transformation affine

si $\mathcal{X} \subset \mathbb{R}^d$ est convexe et $\mathcal{A}(\mathbf{x}) \mapsto \mathbf{A}\mathbf{x} + \mathbf{b}$ une transformation affine definie par une matrice $\mathbf{A} \in \mathbb{R}^{p \times d}$ et un vecteur $\mathbf{b}$ alors

$$\mathcal{A}(\mathcal{X}) = \{\mathcal{A}(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\}$$

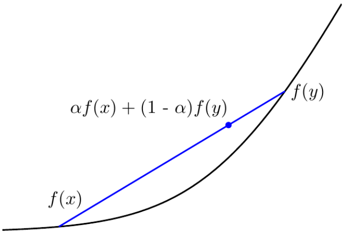
est convexe. Ces transformations incluent par exemples les translations, les rotations, ...

Fonctions convexes

Définition

Une fonction $f : \mathbb{R}^d \mapsto \mathbb{R}$ est convexe si son domaine de définition est convexe et pour tout $\mathbf{x}, \mathbf{y} \in \text{dom} f$ et tout $\lambda \in [0, 1]$

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



autres définitions

- f est concave si $-f$ est convexe
- f est strictement convexe si $f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) < \lambda f(\mathbf{x}) + (1 - \lambda) f(\mathbf{y})$ pour $0 < \lambda < 1$.

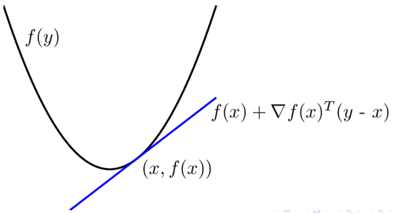
Fonction convexe : caractérisation premier-ordre

Soit la fonction $f : \mathbb{R}^d \mapsto \mathbb{R}$ telle que f soit différentiable *ie*, le gradient de f

$$\nabla f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_d}\right)$$

existe pour tout $\mathbf{x}$ dans l'ensemble de définition de f, on dit alors que f est convexe si son ensemble de définition est convexe et si et seulement si

$$\forall \mathbf{y} \in \mathrm{dom} f, \quad f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x})$$



Fonction convexe : caractérisation second-ordre

- Soit la fonction $f : \mathbb{R}^d \mapsto \mathbb{R}$ telle que f soit deux fois différentiable *ie*, la Hessienne de f
$$\nabla^2 f(\mathbf{x})_{i,j} = \frac{\partial^2 f}{\partial x_i x_j},$$
existe pour tout $\mathbf{x}$ dans l'ensemble de définition de f, on dit alors que f est convexe si son ensemble de définition est convexe et si et seulement si $\nabla^2 f(\mathbf{x})$ est définie positive pour tout $\mathbf{x} \in \mathrm{dom} f$

- si $\nabla^2 f(\mathbf{x})$ est strictement définie positive alors f est strictement convexe.

Fonctions de $\mathbb{R}$

Exemples de fonctions convexes

- fonctions affines : $x \mapsto ax + b$ pour tout $a, b \in \mathbb{R}$.
- fonctions exponentielles : $x \mapsto e^{ax}$ pour tout $a \in \mathbb{R}$.
- puissance de la valeur absolue : $x \mapsto |x|^p$, pour tout $p \geq 1$.
- neg-entropie : $x \mapsto x \log x$ pour $x > 0$

Exemples de fonctions concave

- fonctions affines : $x \mapsto ax + b$ pour tout $a, b \in \mathbb{R}$.
- puissance : $x \mapsto x^p$, pour $x > 0$ et pour tout $0 \leq p \leq 1$.
- logarithme : $x \mapsto \log x$ pour $x > 0$

Exemples de fonctions convexes de $\mathbb{R}^d$

Exemples dans $\mathbb{R}^d$

- une fonction affine $f(\mathbf{x}) = \mathbf{a}^\top \mathbf{x} + b$
- norme : $\|\mathbf{x}\|_p = (\sum_{i=1}^d |x_i|^p)^{1p}$; $\|\mathbf{x}\|_\infty = \max_i |x_i|$

Exemple de fonctions convexes sur les matrices $m \times n$ ($\mathbf{X} \in \mathbb{R}^{m\times n}$)

- fonction affine

$$f(\mathbf{X}) = \mathbf{tr}(\mathbf{A}^\top \mathbf{X}) + b = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{i,j} X_{i,j} + b$$

- norme spectrale d'une matrice (plus grande valeur singulière)

$$f(\mathbf{X}) = \|\mathbf{X}\|_2 = \sigma_{max}(\mathbf{X}) = \lambda_{max}(\mathbf{X}^\top \mathbf{X})$$

Exemples de caractérisation

Premier ordre

- fonction de $\mathbb{R}$ quadratique $f(x) = x^2$

$$y^2 \geq x^2 + 2x(y - x)$$

second ordre

- fonction quadratique : $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top \mathbf{P}\mathbf{x} + \mathbf{q}^\top \mathbf{x} + r$

$$\nabla f(\mathbf{x}) = \mathbf{P}\mathbf{x} + q \qquad \nabla^2 f(\mathbf{x}) = P$$

  est convexe si $\mathbf{P} \succeq 0$

- Fonction objective d'une regression lineaire : $f(\mathbf{x}) = \|\mathbf{A}\mathbf{x} + \mathbf{b}\|_2^2$

$$\nabla f(\mathbf{x}) = 2\mathbf{A}^\top(\mathbf{A}\mathbf{x} + b) \qquad \nabla^2 f(\mathbf{x}) = 2\mathbf{A}^\top \mathbf{A}$$

  convexe pour tout $\mathbf{A}$.

- d'autres fonctions $f(x, y) = \frac{x^2}{y}$, log-sum-exp $f(\mathbf{x}) = \log \sum_{i=1}^{d} e^{x_i}$, $\cdots$

Opération préservant la convexité des fonctions

Somme positive

- soit $\lambda_1, \lambda_2 \geq 0$ et $f_1,\ f_2$ deux fonctions convexes, alors $\lambda_1 f_1 + \lambda_2 f_2$ est convexe.

Composition avec une fonction affine

- soit $\mathbf{A} \in \mathbb{R}^{p \times d}$ et $b \in \mathbb{R}^p$ et $f : \mathbb{R}^p \mapsto \mathbb{R}$ une fonction convexe alors $f(\mathbf{A}\mathbf{x} + b)$ est convexe

Exemple

- log barrière : $f(\mathbf{x}) = -\sum_{i=1}^{m} \log(b_i - \mathbf{a}_i^\top \mathbf{x})$ avec $\text{dom} f = \{\mathbf{x} : \mathbf{a}_i^\top \mathbf{x} \leq b_i\}$
- norme de fonction affine : $f(\mathbf{x}) = \|\mathbf{A}\mathbf{x} + b\|$

Opération préservant la convexité des fonctions

Composition

- soit $g : \mathbb{R}^d \mapsto \mathbb{R}$ une fonction convexe et $h : \mathbb{R} \mapsto \mathbb{R}$ une fonction convexe et croissante, alors
$$f(\mathbf{x}) = h(g(\mathbf{x}))$$
est convexe.

Maximum

- si $f_1, \cdots, f_m$ sont des fonctions convexes alors
$$f(\mathbf{x}) = \max_i \{f_1(\mathbf{x}), \cdots, f_m(\mathbf{x})\}$$
est convexe

Exemple

- fonction continue par morceau $f(\mathbf{x}) = \max_{i=1,\cdots,m}(\mathbf{a}_i^\top \mathbf{x} + b)$

Il existe plusieurs autres opérations conservant la convexité.