Problème d'optimisation convexe

30 septembre 2024

Formulation

Forme usuelle d'un problème d'optimisation

$$
\begin{array}{ll}
\min_{\mathbf{x}} & f(\mathbf{x}) \\
\text{tel que} & g_i(\mathbf{x}) \leq 0, i = 1, \cdots, k \\
& h_j(\mathbf{x}) = 0, j = 1, \cdots, m \\
& \mathbf{x} \in \mathcal{X}
\end{array}
$$

Problème d'optimisation convexe (P)

$$
\begin{array}{ll}
\min_{\mathbf{x}} & f(\mathbf{x}) \\
\text{tel que} & g_i(\mathbf{x}) \leq 0, i = 1, \cdots, k \\
& \mathbf{A}\mathbf{x} = \mathbf{b}, j = 1, \cdots, m \\
& \mathbf{x} \in \mathcal{X}
\end{array}
$$

où les g_i sont des fonctions convexes, et $\mathcal{X}$ est un ensemble convexe inclut dans l'intersection des domaines de définitions des fonctions f et $\{g_i\}$. On appelle f la fonction objective et g_i les contraintes d'inégalité .

Equivalence

La maximisation d'une fonction concave sous contrainte convexe est un problème d'optimisation convexe.

Définitions et terminologie

Point admissible

On dit que $\mathbf{x}$ est un point admissible ou un point réalisable si $\mathbf{x} \in \mathcal{X}$, $\mathbf{A}\mathbf{x} = \mathbf{b}$, et $\forall i, g_i(\mathbf{x}) \leq 0$.

Valeur optimale

Le minimum de $f\mathbf{x})$ sur l'ensemble des points admissible est appelé la valeur optimale souvent notée $f^\star$.

Optimalité

$\mathbf{x}^\star$ est dit point optimal si $\mathbf{x}^\star$ est admissible et $\forall \mathbf{x}, f(\mathbf{x}^\star) \leq f(\mathbf{x})$. On dit également $\mathbf{x}^\star$ est une solution du problème ou un minimiseur.

sous-optimalité

Si $\mathbf{x}$ est admissible et $f(\mathbf{x}) \leq f(\mathbf{x}^\star) + \epsilon$, on dit alors que $\mathbf{x}$ est ϵ-sous-optimal.

Contrainte active

Si $\mathbf{x}$ est admissible et $g_i(\mathbf{x}) = 0$, on dit alors que g_i est une contrainte active en $\mathbf{x}$.

Solution d'un problème d'optimisation convexe

Soit

$$\begin{array}{rll} \{\mathbf{x}_{opt}\} = & \arg\min_{\mathbf{x}} & f(\mathbf{x}) \\ & \text{tel que} & g_i(\mathbf{x}) \leq 0, i = 1, \cdots, k \\ & & \mathbf{A}\mathbf{x} = \mathbf{b}, j = 1, \cdots, m \\ & \mathbf{x} \in \mathcal{X} & \end{array}$$

Propriétés

- Pour certains problèmes, on peut avoir $\{\mathbf{x}_{opt}\} = \emptyset$. Par exemple, si l'ensemble des points admissibles est vide ou si f n'est pas bornée inférieurement sur l'ensemble des points admissibles.
- L'ensemble $\{\mathbf{x}_{opt}\}$ est un ensemble convexe.
- Si f est strictement convexe, si il existe une solution, alors elle est unique. $\{\mathbf{x}_{opt}\}$ n'est constituée que d'un seul élément.

Exemple : non-negative Lasso

Soit le problème donné par $\mathbf{y} \in \mathbb{R}^n$, $\mathbf{X} \in \mathbb{R}^{n\times d}$ et

$$\begin{array}{ll} \min_{\mathbf{w}} & \frac{1}{2}\|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2^2 \\ \text{tel que} & \|\mathbf{w}\|_1 \leq 1 \\ & w_i \geq 0 \end{array}$$

- Est que ce problème est convexe ? quelle est la fonction objective ? quelles sont les contraintes d'inégalités ? les contraintes linéaires d'égalités ?
- Quel est l'ensemble des points admissibles ?
- Est ce que la solution est unique, quand $n \geq d$ et $n < d$?
- quelles sont les réponses à ces questions si on supprime les contraintes $w_i \geq 0$?

Exemple : Séparateur à vaste marge

Soit le problème donné par $y \in \{+1, -1\}$, $\mathbf{X} \in \mathbb{R}^{n \times d}$, dont les lignes sont les exemples d'apprentissage

$$\begin{array}{ll} \min_{\mathbf{w}, w_0} & \frac{1}{2}\|\mathbf{w}\|_2^2 \\ \text{tel que} & y_i(\mathbf{x}_i^\top \mathbf{w} + w_0) \geq 1 \end{array}$$

- Est que ce problème est convexe ? quelle est la fonction objective ? quelles sont les contraintes d'inégalités ? les contraintes linéaires d'égalités ?
- Sous quelles conditions sur les exemples d'apprentissage le problème admet une solution ?

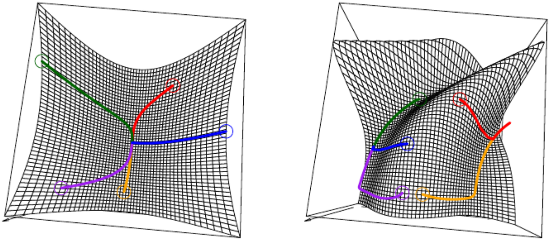
Local = Global

Optimum local

Pour un problème d'optimisation, on dit qu'un point admissible $\mathbf{x}$ est un optimum local si il existe R tel que :

$$f(\mathbf{x}) \leq f(\mathbf{x}') \qquad \text{pour tout } \mathbf{x}' \text{ tel que} \|\mathbf{x} - \mathbf{x}'\|_2 \leq R$$

- Pour un problème convexe, un optimum local est également un optimum global, donc solution du problème.
- cette propriété n'est généralement pas vraie pour des problèmes d'optimisation non-convexes.



Convex   Nonconvex

Reformulation des contraintes

Le problème

$$\begin{array}{ll} \min_{\mathbf{x}} & f(\mathbf{x}) \\ \text{tel que} & g_i(\mathbf{x}) \leq 0, i = 1, \cdots, k \\ & \mathbf{A}\mathbf{x} = \mathbf{b}, j = 1, \cdots, m \\ & \mathbf{x} \in \mathcal{X} \end{array}$$

peut se ré-écrire

$$\min_{\mathbf{x}} f(\mathbf{x}) \text{ sous contraintes } \mathbf{x} \in \mathcal{S}$$

où $\mathcal{S} = \{\mathbf{x} : g_i(\mathbf{x}) \leq 0, \forall i = 1, \cdots, k, \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \in \mathcal{X}\}$.
Une forme non-contrainte de ce problème peut également s'obtenir :

$$\min_{\mathbf{x}} f(\mathbf{x}) + I_{\mathcal{S}}(\mathbf{x})$$

où $I_{\mathcal{S}}(\mathbf{x}) = 0$ si $\mathbf{x} \in \mathcal{S}$ et $I_{\mathcal{S}}(\mathbf{x}) = \infty$ sinon. On appelle $I_{\mathcal{S}}$ la fonction indicatrice de $\mathcal{S}$.

Condition d'optimalité du premier ordre

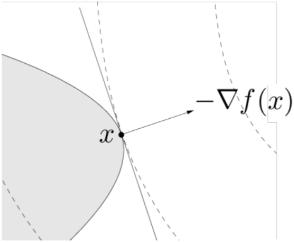
Cas des fonctions convexes et différentiables

Pour un problème convexe

$$\min_{\mathbf{x}} f(\mathbf{x}) \text{ sous contraintes } \mathbf{x} \in \mathcal{S}$$

le point admissible $\mathbf{x}^\star$ est optimal si et seulement si

$$\nabla f(\mathbf{x}^\star)^\top (\mathbf{y} - \mathbf{x}^\star) \geq 0 \quad \forall \mathbf{y} \in \mathcal{S}$$



- Toute direction admissible à partir de $\mathbf{x}^\star$ est aligné avec un gradient croissant.
- si $\mathcal{S} = \mathbb{R}^d$, cette condition équivaut à $\nabla f(\mathbf{x}^\star) = 0$

Exemple : optimisation quadratique

Soit le problème de minimisation d'une fonction quadratique

$$\min_{\mathbf{x}} \frac{1}{2}\mathbf{x}^\top \mathbf{A}\mathbf{x} + \mathbf{b}^\top \mathbf{x} + c$$

où $\mathbf{A}$ est une matrice définie positive (DP). La condition d'optimalité du premier ordre est

$$\nabla f(\mathbf{x}^\star) = \mathbf{A}\mathbf{x}^\star + \mathbf{b} = 0$$

Etude de cas

- si $\mathbf{A}$ est strictement DP alors $\mathbf{A}$ est inversible, on a une solution unique $\mathbf{x}^\star = -\mathbf{A}^{-1}\mathbf{b}$
- si $\mathbf{A}$ n'est pas inversible, on peut avoir soit aucune solution soit une infinité de solution.

Exemple : optimisation quadratique sous constrainte

Soit le problème d'optimisation dans $\mathbb{R}$

$$\min_x \frac{1}{2}x^2 \qquad \text{sc } x \in \mathcal{S}$$

- quel est la solution du problème si $\mathcal{S} = [-1, 1]$?
- quel est la solution du problème si $\mathcal{S} = [1, 2]$?
- quel est la solution du problème si $\mathcal{S} = [-2, -1]$?

La condition d'optimalité du premier ordre

$$\nabla f(\mathbf{x}^\star)^\top (\mathbf{y} - \mathbf{x}^\star) \geq 0 \quad \forall \mathbf{y} \in \mathcal{S}$$

Transformation monotone

Forme standard

$$\begin{array}{ll} \min_{\mathbf{x}} & f(\mathbf{x}) \\ \text{tel que} & g_i(\mathbf{x}) \leq 0, i = 1, \cdots, k \\ & \mathbf{A}\mathbf{x} = \mathbf{b}, j = 1, \cdots, m \\ & \mathbf{x} \in \mathcal{X} \end{array}$$

- $\psi_0 : \mathbb{R} \mapsto \mathbb{R}$ est une fonction monotone croissante
- $\psi_i : \mathbb{R} \mapsto \mathbb{R}$ est telle que $\psi_i(u) \leq 0$ si et seulement si $u \leq 0$
- $h(\mathbf{x}) = 0$ si et seulement si $\mathbf{x} = 0$.

Forme transformée

$$\begin{array}{ll} \min_{\mathbf{x}} & \psi_0(f(\mathbf{x})) \\ \text{tel que} & \psi_i(g_i(\mathbf{x})) \leq 0, i = 1, \cdots, k \\ & h(\mathbf{A}\mathbf{x} - \mathbf{b}) = 0, j = 1, \cdots, m \\ & \mathbf{x} \in \mathcal{X} \end{array}$$

Exemple de transformation

Exemple

$$\min_{\mathbf{w}} \|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2 \qquad \min_{\mathbf{w}} \frac{1}{2}\|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2^2$$

- les deux problèmes sont équivalents par la transformation $\psi_0(x) = \frac{1}{2}x^2$
- l'ensemble des solutions sont donc identiques
- le premier problème n'est pas différentiable
- le deuxième est différentiable et admet une solution analytique

Transformation : Variable de relachement

Inégalité devient égalité

$$\begin{array}{ll} \min_{\mathbf{x}} & f(\mathbf{x}) \\ \text{tel que} & g_i(\mathbf{x}) \leq 0, i = 1, \cdots, k \\ & \mathbf{A}\mathbf{x} = \mathbf{b}, j = 1, \cdots, m \end{array}$$

Les contraintes d'inégalité peuvent être ré-écrite comme

$$\begin{array}{ll} \min_{\mathbf{x}} & f(\mathbf{x}) \\ \text{tel que} & g_i(\mathbf{x}) + s_i = 0, i = 1, \cdots, k \\ & s_i \geq 0, i = 1, \cdots, k \\ & \mathbf{A}\mathbf{x} = \mathbf{b}, j = 1, \cdots, m \end{array}$$

Perte de convexité

Si $g_i(\mathbf{x})$ n'est pas affine, c-a-d $g_i(\mathbf{x}) \neq \mathbf{a}_i^\top \mathbf{x} + b_i$ alors cette transformation casse la convexité du problème.

Transformation epigraphique

Forme standard : valeur optimale $p^\star$

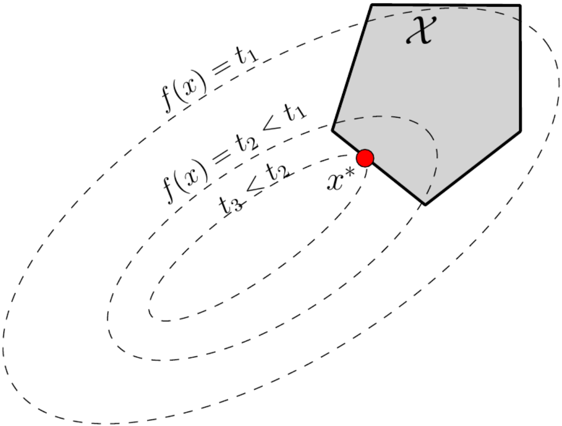
$$\begin{array}{ll} \min_{\mathbf{x}} & f(\mathbf{x}) \\ \text{tel que} & g_i(\mathbf{x}) \leq 0, i = 1, \cdots, k \\ & \mathbf{A}\mathbf{x} = \mathbf{b}, j = 1, \cdots, m \end{array}$$

Forme épigraphique

$$\begin{array}{ll} \min_{\mathbf{x},t} & t \\ & f(\mathbf{x}) - t \leq 0 \\ \text{tel que} & g_i(\mathbf{x}) \leq 0, i = 1, \cdots, k \\ & \mathbf{A}\mathbf{x} = \mathbf{b}, j = 1, \cdots, m \end{array}$$

les solutions $\mathbf{x}^\star$ pour les deux problèmes sont les mêmes et $t^\star = p^\star$.

Transformation epigraphique : intuition



si on definit les courbes de niveau $L_t = \{\mathbf{x} : f(\mathbf{x}) \leq t\}$ $t \in \mathbb{R}$, On cherche le plus petit t pour lequel l'intersection entre L_t et l'ensemble définit par les contraintes est non-vide.