Optimisation non linéaire: Théorie

Plan

1. Introduction et définitions
2. Optimisation sans contraintes
3. Optimisation avec contraintes

1. Introduction et définitions

2. Optimisation sans contraintes

3. Optimisation avec contraintes

4. Extensions

Références

Problème et solutions

- On cherche à résoudre

$$\min_{\mathbf{x}\in\mathbb{R}^n} \{f(\mathbf{x}) : \mathbf{x} \in \Omega\}$$

avec $f : \mathbb{R}^n \to \mathbb{R}$ différentiable et $\Omega \subseteq \mathbb{R}^n$

- Un point réalisable $\mathbf{x}^* \in \Omega$ est un minimum global de la fonction f sur le domaine Ω si

$$f(\mathbf{x}^*) \leq f(\mathbf{x}) \text{ pout tout } \mathbf{x} \in \Omega$$

- Un point réalisable $\mathbf{x}^* \in \Omega$ est un minimum local de f sur Ω s'il existe $\varepsilon > 0$ tel que

$$f(\mathbf{x}^*) \leq f(\mathbf{x}) \text{ pout tout } \mathbf{x} \in \Omega \cap \mathcal{B}_\varepsilon(\mathbf{x}^*)$$

avec $\mathcal{B}_\varepsilon(\mathbf{x}^*) = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{x}^*\| < \varepsilon\}$

Dérivées

- Gradient de f en $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$:

$$\nabla f(\mathbf{x}) = \left(\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \ldots, \frac{\partial f(\mathbf{x})}{\partial x_n} \right) \in \mathbb{R}^n$$

- Dérivée directionnelle de f en $\mathbf{x} \in \mathbb{R}^n$ dans la direction unitaire $\mathbf{d} \in \mathbb{R}^n$:

$$f'_{\mathbf{d}}(\mathbf{x}) = \lim_{t \downarrow 0} \frac{f(\mathbf{x} + t\mathbf{d}) - f(\mathbf{x})}{t} = \mathbf{d}^\top \nabla f(\mathbf{x})$$

- Si les dérivées secondes de f existent et sont continues, alors la matrice hessienne en $\mathbf{x}$ s'écrit

$$\nabla^2 f(\mathbf{x}) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) \right)_{ij}$$

Direction de descente

- $\mathbf{d} \in \mathbb{R}^n$ est une direction (stricte) de descente de f en $\mathbf{x} \in \mathbb{R}^n$ si

$$f'_{\mathbf{d}}(\mathbf{x}) = \mathbf{d}^\top \nabla f(\mathbf{x}) < 0$$

- Pour tout $\alpha \in \mathbb{R}$ petit, on aura $h(\alpha) = f(\mathbf{x} + \alpha \mathbf{d}) < f(\mathbf{x})$ et $h'(0) < 0$

- Principe de la *line search* (recherche linéaire) : Trouver α tel que $h'(\alpha) = 0$

Signe d'une matrice

$A \in \mathbb{R}^{n\times n}$ symétrique est dite

- Semi-définie positive si $\mathbf{x}^\top A\mathbf{x} \geq 0$ pour tout $\mathbf{x} \in \mathbb{R}^n$
- Définie positive si $\mathbf{x}^\top A\mathbf{x} > 0$ pour tout $\mathbf{x} \in \mathbb{R}^n \neq 0$
- Semi-définie négative si $\mathbf{x}^\top A\mathbf{x} \leq 0$ pour tout $\mathbf{x} \in \mathbb{R}^n$
- Définie négative si $\mathbf{x}^\top A\mathbf{x} < 0$ pour tout $\mathbf{x} \in \mathbb{R}^n \neq 0$
- Indéfinie sinon

En pratique, on peut vérifier le signe d'une matrice en examinant ses valeurs propres ou ses mineurs principaux dominants

1. Introduction et définitions

2. Optimisation sans contraintes

3. Optimisation avec contraintes

4. Extensions

Références

Optimisation sans contraintes : CN1

Dans le cas $\Omega = \mathbb{R}^n$:

Condition nécessaire d'optimalité de premier ordre

Si $\mathbf{x}^*$ est un minimum local de f sur $\mathbb{R}^n$, alors $\nabla f(\mathbf{x}^*) = \mathbf{0}$ ($\mathbf{x}^*$ est un point critique)

- Attention : Ce n'est pas une condition suffisante : Un point critique peut être un minimum local, un maximum local, ou un bien un point de selle
- Un point critique $\mathbf{x}$ est un point de selle si pour tout $\varepsilon > 0$ il existe $\mathbf{a}, \mathbf{b} \in \mathcal{B}_\varepsilon(\mathbf{x})$ tels que $f(\mathbf{a}) < f(\mathbf{x}) < f(\mathbf{b})$
- Si $\mathbf{x}$ n'est pas un point critique, il ne peut pas être un minimum ou un maximum

Optimisation sans contraintes : CN2

Condition nécessaire de second ordre

Si $\mathbf{x}^*$ est un minimum local de f sur $\mathbb{R}^n$, alors $\nabla f(\mathbf{x}^*) = \mathbf{0}$ et la matrice hessienne $\nabla^2 f(\mathbf{x}^*)$ est semi-définie positive

Preuve : Soit $\mathbf{x}^*$ un minimum local. Pour toute direction unitaire $\mathbf{d} \in \mathbb{R}^n$ et pour $t \in \mathbb{R}$ suffisamment petit :

$$\begin{aligned} f(\mathbf{x}^*) \leq f(\mathbf{x}^* + t\mathbf{d}) \quad & \simeq f(\mathbf{x}^*) + t\mathbf{d}^\top \nabla f(\mathbf{x}^*) + \tfrac{t^2}{2}\mathbf{d}^\top \nabla^2 f(\mathbf{x}^*)\mathbf{d} \\ & = f(\mathbf{x}^*) + \tfrac{t^2}{2}\mathbf{d}^\top \nabla^2 f(\mathbf{x}^*)\mathbf{d} \end{aligned}$$

$$\Rightarrow \mathbf{d}^\top \nabla^2 f(\mathbf{x}^*)\mathbf{d} \geq 0$$

Si la matrice hessienne en un point critique est indéfinie, alors il s'agit d'un point de selle

Optimisation sans contraintes : CS2

Condition suffisante de second ordre

si $\mathbf{x}^* \in \mathbb{R}^n$ est un point critique et si $\nabla^2 f(\mathbf{x}^*)$ est :

- Définie positive : $\mathbf{x}^*$ est un minimum local
- Définie négative : $\mathbf{x}^*$ est un maximum local
- Indéfinie : $\mathbf{x}^*$ est un point de selle
- Semi-définie positive ou semi-définie négative : on ne peut rien dire

Optimisation sans contraintes : Convexité

$f : \mathbb{R}^n \to \mathbb{R}$ est convexe si :

- pour tout $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ et tout $\lambda \in [0; 1]$, $f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y})$
- ou si sa matrice hessienne $\nabla^2 f(\mathbf{x})$ est semi-définie positive **pour tout $\mathbf{x} \in \mathbb{R}^n$**

- Si f est convexe, la CN1 devient suffisante : Il suffit de trouver un point critique pour minimiser f
- f est concave si $-f$ est convexe

Optimisation sans contraintes : Méthode du gradient

Pour la minimisation d'une fonction $f : \mathbb{R}^n \to \mathbb{R}$ sans contraintes

[0] Initialisation

Point de départ : $\mathbf{x}^0 \in \mathbb{R}^n$
$k \leftarrow 0$

[1] Itération k

Calculer $\mathbf{d}^k = -\nabla f(\mathbf{x}^k)$ (dir. de descente)
Si $(\mathbf{d}^k = \mathbf{0})$: Stop (point critique)
Trouver $\alpha^k \in \arg\min_{\alpha \geq 0} h(\alpha) = f(\mathbf{x}^k + \alpha \mathbf{d}^k)$
$\mathbf{x}^{k+1} \leftarrow \mathbf{x}^k + \alpha^k \mathbf{d}^k$
$k \leftarrow k + 1$
Aller en **[1]**

Méthode de Newton

- Soit le modèle quadratique de f autour de $\mathbf{x}$:

$$m(\mathbf{d}) = f(\mathbf{x}) + \mathbf{d}^\top \nabla f(\mathbf{x}) + \frac{1}{2}\mathbf{d}^\top \nabla^2 f(\mathbf{x})\mathbf{d}$$

- Si $\nabla^2 f(\mathbf{x})$ est définie positive, alors m est une fonction convexe et on peut identifier son minimum global avec

$$\nabla m(\mathbf{d}) = \nabla f(\mathbf{x}) + \nabla^2 f(\mathbf{x})\mathbf{d} = \mathbf{0}$$

- Au lieu de considérer $\mathbf{d}^k = -\nabla f(\mathbf{x}^k)$ comme direction de descente, on prend donc $\mathbf{d}^k = -\left(\nabla^2 f(\mathbf{x}^k)\right)^{-1} \nabla f(\mathbf{x}^k)$ (direction de Newton)

Méthode quasi-Newton

- La direction de Newton n'est pas définie si la matrice hessienne n'est pas définie positive
- Calculer (et inverser) la matrice hessienne peut aussi être très coûteux
- On peut considérer la direction quasi-Newton

$$\mathbf{d} = -B(\mathbf{x})^{-1}\nabla f(\mathbf{x})$$

  avec $B(\mathbf{x})$ définie positive qui remplace la matrice hessienne
- Une méthode quasi-Newton sera d'autant plus efficace quand elle pourra intégrer l'information de second-ordre dans B

1. Introduction et définitions

2. Optimisation sans contraintes

3. Optimisation avec contraintes

4. Extensions

Références

Optimisation avec contraintes

$$\min_{\mathbf{x}\in\mathbb{R}^n} \{f(\mathbf{x}) : \mathbf{x} \in \Omega\}$$

Théorème

Si Ω est fermé et borné et si f est continue sur Ω, alors il existe un minimum global atteint en un point de Ω et un maximum global atteint en un point de Ω

En pratique, cela signifie que pour résoudre le problème, on peut énumérer tous les candidats (les points critiques) et les comparer afin de trouver les optima

Optimisation avec une contrainte égalité

Avec $\Omega = \{\mathbf{x} \in \mathbb{R}^n : c(\mathbf{x}) = 0\} \subseteq \mathbb{R}^n$:

CN1

Si $\mathbf{x}^*$ est un minimum local de f dans Ω, et si $\nabla c(\mathbf{x}^*) \neq \mathbf{0}$, alors $c(\mathbf{x}^*) = 0$ et il existe $\lambda \in \mathbb{R}$ tel que

$$\nabla f(\mathbf{x}^*) = \lambda \nabla c(\mathbf{x}^*)$$

- Un point $\mathbf{x}^*$ satisfaisant cette condition est appelé un point critique
- **Exemple 1 :** $\min\limits_{x_1,x_2} 3x_1 - 2x_2$ s.c. $x_1^2 + 2x_2^2 = 44$

Optimisation avec une contrainte inégalité

Avec $\Omega = \{\mathbf{x} \in \mathbb{R}^n : c(\mathbf{x}) \geq 0\} \subseteq \mathbb{R}^n$:

CN1

Si $\mathbf{x}^*$ est un minimum local de f dans Ω, alors il existe $\lambda \geq 0$ tel que $\nabla f(\mathbf{x}^*) = \lambda \nabla c(\mathbf{x}^*)$ et $c(\mathbf{x}^*)\lambda = 0$

- Un point $\mathbf{x}^*$ satisfaisant ces conditions est appelé un point critique
- Si $c(\mathbf{x}^*) > 0$, la condition devient $\nabla f(\mathbf{x}^*) = \mathbf{0}$
- **Exemple 2 :** $\min\limits_{x_1,x_2} (x_1 - 1)^2 + (x_2 - 2)^2 \quad \text{s.c.} \quad x_1^2 + x_2^2 \leq 45$

Optimisation avec plusieurs contraintes égalité

Avec $\Omega = \{\mathbf{x} \in \mathbb{R}^n : c_i(\mathbf{x}) = 0, i \in \mathcal{E}\} \subseteq \mathbb{R}^n$ et $|\mathcal{E}| = m$:

CN1

Si $\mathbf{x}^*$ est un minimum local de f dans Ω où $\{\nabla c_i(\mathbf{x}^*) : i \in \mathcal{E}\}$ est un ensemble linéairement indépendant, alors $c_i(\mathbf{x}^*) = 0$ pour tout $i \in \mathcal{E}$ et il existe $\lambda \in \mathbb{R}^m$ tel que

$$\nabla f(\mathbf{x}^*) = \sum_{i \in \mathcal{E}} \lambda_i \nabla c_i(\mathbf{x}^*)$$

- Un point $\mathbf{x}^*$ satisfaisant cette condition est appelé un point critique
- **Exemple 3 :** $\min\limits_{x \in \mathbb{R}^3} x_1 - x_2 + x_3$

$$\text{s.c.} \begin{cases} x_1^2 + x_2^2 + x_3^2 & = 1 \\ x_1^2 + (x_2 - 1)^2 + (x_3 - 2)^2 & = 4 \end{cases}$$

Multiplicateurs de Lagrange

- Les λ des conditions nécessaires sont appelés les multiplicateurs de Lagrange
- Ils peuvent servir à effectuer des analyses de sensibilité sur les membres de droite des contraintes
- En effet, un λ représente la variation de f lorsque le mdd de la contrainte associée augmente d'une unité

Optimisation avec contraintes : Cas général

$$\min_{\mathbf{x}\in\mathbb{R}^n} \{f(\mathbf{x}) : \mathbf{x} \in \Omega\}$$

avec

$$\Omega = \left\{ \mathbf{x} \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} c_i(\mathbf{x}) = 0, & i \in \mathcal{E} \\ c_i(\mathbf{x}) \geq 0, & i \in \mathcal{I} \end{array} \right\} \subseteq \mathbb{R}^n$$

et

$$|\mathcal{E}| = m \ , \ |\mathcal{I}| = p$$

- Les fonctions décrivant le problème sont toutes différentiables et Ω est un ensemble fermé et borné
- On va décrire les conditions d'optimalité de ce problème. Trois ingrédients sont nécessaires : Le cône tangent, le cône normal, et la qualification des contraintes

CN1 : Conditions de KKT

La CN1 peut se reformuler comme les conditions de Karush-Kuhn-Tucker (KKT) :

CN1

Si $\mathbf{x}^*$ est un minimum local de f dans Ω, alors il existe $\lambda \in \mathbb{R}^{m+p}$ tel que

$$\begin{aligned}
\nabla f(\mathbf{x}^*) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i \nabla c_i(\mathbf{x}^*) &= \mathbf{0} & \\
\lambda_i c_i(\mathbf{x}^*) &= 0 & i \in \mathcal{I} \\
c_i(\mathbf{x}^*) &= 0 & i \in \mathcal{E} \\
c_i(\mathbf{x}^*) &\geq 0 & i \in \mathcal{I} \\
\lambda_i &\geq 0 & i \in \mathcal{I}
\end{aligned}$$

Comme pour toutes les conditions nécessaires, les conditions de KKT ne sont pas suffisantes et peuvent aussi correspondre à des maximums et des points de selle

Dualité

Fonction Lagrangienne (ou Lagrangien) :

$$L(\mathbf{x}, \lambda) = f(\mathbf{x}) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i c_i(\mathbf{x})$$

(la 1ère équation des conditions KKT peut s'écrire $\nabla_{\mathbf{x}} L(\mathbf{x}, \lambda) = \mathbf{0}$)

Théorème de dualité faible

Soit $\mathbf{x}^* \in \arg\min_{x \in \Omega} f$. Pour tout $\lambda \in \mathbb{R}^{m+p}$ avec $\lambda_i \geq 0$ pour $i \in \mathcal{I}$, on a

$$\overline{L}(\lambda) = \min_{\mathbf{x} \in \Omega} L(\mathbf{x}, \lambda) \leq f(\mathbf{x}^*)$$

La meilleure borne inférieure est donc donnée par le problème dual

$$\max_{\substack{\lambda \in \mathbb{R}^{m+p} \\ \lambda_i \geq 0,\ i \in \mathcal{I}}} \overline{L}(\lambda)$$

Cas de l'optimisation linéaire (1/3)

$$\min_{\mathbf{x}\in\mathbb{R}^n} \mathbf{c}^\top \mathbf{x} \quad \text{s.c.} \begin{cases} A\mathbf{x} & \geq \mathbf{b} \\ \mathbf{x} & \geq \mathbf{0} \end{cases} \quad \text{avec } \mathbf{c}\in\mathbb{R}^n,\ \mathbf{b}\in\mathbb{R}^m,\ A\in\mathbb{R}^{m\times n}$$

- Le Lagrangien est considéré avec les variables $\mathbf{x}\in\mathbb{R}^n$, $\mu\in\mathbb{R}^m$ et $\lambda\in\mathbb{R}^n$:

$$L(\mathbf{x},\lambda,\mu) = \mathbf{c}^\top\mathbf{x} - \mu^\top(A\mathbf{x}-\mathbf{b}) - \lambda^\top\mathbf{x} = \mathbf{b}^\top\mu + \mathbf{x}^\top\left(\mathbf{c}-\lambda-A^\top\mu\right)$$

- Une borne inférieure est donnée par

$$\min_{\mathbf{x}\in\Omega} \mathbf{b}^\top\mu + \mathbf{x}^\top\left(\mathbf{c}-\lambda-A^\top\mu\right)$$

avec $\lambda,\mu\geq\mathbf{0}$

Cas de l'optimisation linéaire (2/3)

Les conditions KKT sont

$$
\begin{aligned}
\nabla_{\mathbf{x}} L(\mathbf{x}, \lambda, \mu) = \mathbf{c} - \lambda - A^\top \mu \quad &= \mathbf{0} \\
\mu_i (A_i^\top \mathbf{x} - \mathbf{b}_i) \quad &= 0 \quad i \in \{1, 2, \dots, m\} \\
\lambda_j x_j \quad &= 0 \quad j \in \{1, 2, \dots, n\} \\
A\mathbf{x} - \mathbf{b} \quad &\geq \mathbf{0} \\
\mathbf{x} \quad &\geq \mathbf{0} \\
\lambda_j \quad &\geq 0 \quad j \in \{1, 2, \dots, n\} \\
\mu_i \quad &\geq 0 \quad i \in \{1, 2, \dots, m\}
\end{aligned}
$$

Obtenir la meilleure borne inférieure revient donc à résoudre

$$
\max_{\lambda, \mu \geq \mathbf{0}} \mathbf{b}^\top \mu \quad \text{s.c.} \quad \left\{ \mathbf{c} - \lambda - A^\top \mu = \mathbf{0} \right.
$$

Cas de l'optimisation linéaire (3/3)

En considérant les λ comme des variables d'écart, on obtient le problème dual

$$\max_{\mu \in \mathbb{R}^n} \mathbf{b}^\top \mu$$

$$\text{s.c.} \begin{cases} A^\top \mu & \leq \mathbf{c} \\ \mu & \geq \mathbf{0} \end{cases}$$

qui correspond à ce qui a été vu en OL

Note : Les conditions KKT redonnent le théorème des écarts complémentaires

Méthodes de pénalités

- Les méthodes de pénalités sont des algorithmes itératifs qui, à chaque itération, considèrent la minimisation sans contraintes d'une fonction critère dans laquelle les violations des contraintes sont associées à des coûts
- Ces coûts vont être augmentés au fil des itérations
- On espère ainsi générer une suite de points tendant à respecter les contraintes

Exemples de pénalités

- Pénalité quadratique :

$$f(\mathbf{x}) + \frac{\mu}{2} \sum_{i \in \mathcal{E}} c_i^2(\mathbf{x}) + \frac{\mu}{2} \sum_{i \in \mathcal{I}} \min\{0, c_i(\mathbf{x})\}^2$$

  - Avantage : Formulation lisse
  - Inconvénients : Non exacte, mal conditionnée

- Pénalité ℓ_1 :

$$f(\mathbf{x}) + \mu \sum_{i \in \mathcal{E}} |c_i(\mathbf{x})| + \mu \sum_{i \in \mathcal{I}} |\min\{0, c_i(\mathbf{x})\}|$$

  - Avantage : Exacte : Il existe μ tel que l'optimum sans contraintes correspond à l'optimum avec contraintes
  - Inconvénient : Formulation non-lisse

Algorithme du Lagrangien augmenté (1/2)

- On considère $\mathcal{I} = \emptyset$ à des fins de simplicité
- Méthode de pénalité basée sur la pénalité quadratique, donc qui donne une formulation lisse, mais qui réduit le mauvais conditionnement grâce à l'emploi des multiplicateurs de Lagrange
- Le Lagrangien augmenté pour contraintes égalité est

$$\begin{aligned} L_a(\mathbf{x}, \lambda, \mu) &= L(\mathbf{x}, \lambda) + \frac{\mu}{2} \sum_{i \in \mathcal{E}} c_i^2(\mathbf{x}) \\ &= f(\mathbf{x}) - \sum_{i \in \mathcal{E}} \lambda_i c_i(\mathbf{x}) + \frac{\mu}{2} \sum_{i \in \mathcal{E}} c_i^2(\mathbf{x}) \end{aligned}$$

- L'algorithme suivant converge vers un point critique (selon les conditions KKT) et fournit également les multiplicateurs de Lagrange

Algorithme du Lagrangien augmenté (2/2)

[0] Initialisation

Point de départ : $\mathbf{x}^0 \in \mathbb{R}^n$, $\lambda^0 \in \mathbb{R}^m$
Précision initiale : $\tau^0 > 0$
Pénalité initiale : $\mu^0 > 0$
$k \leftarrow 1$

[1] Itération k

Trouver approximativement $\mathbf{x}^{k+1} \in \arg\min_{\mathbf{x}\in\mathbb{R}^n} L_a(\mathbf{x}, \lambda^k, \mu^k)$
avec $\|\nabla_{\mathbf{x}} L_a(\mathbf{x}^{k+1}, \lambda^k, \mu^k)\| \leq \tau^k$ comme critère d'arrêt
Si (test de convergence) : Stop (point critique)
$\lambda_i^{k+1} \leftarrow \lambda_i^k - \mu^k c_i(\mathbf{x}^{k+1}) \quad i \in \mathcal{E}$
Choisir $\mu^{k+1} \geq \mu^k$
Choisir τ^{k+1}
$k \leftarrow k + 1$
Aller en [1]