

Cours 13 : Méthodes de gradient

Christophe Gonzales

LIP6 – Université Paris 6, France

Méthodes primales / duales

- $f : \mathbb{R}^n \mapsto \mathbb{R}$
- $\min f(x)$
s.c. $g_i(x) \leq 0, i = \{1, \dots, m\}, x \in \mathbb{R}^n$

Méthodes primales / duales

- **méthodes primales :**
 - génération d'une séquence de **solutions**, i.e., de points satisfaisant les contraintes
 - séquence \implies fait décroître $f \implies$ propriété anytime
 - méthodes difficiles à implanter et à calibrer
- **méthodes duales :**
 - résolution d'une séquence de problèmes d'optimisation sans contraintes \implies utilisation du lagrangien
 - méthodes plus robustes que les primales et convergence plus facile à assurer
 - pas de propriété anytime

Cours 13 : Méthodes de gradient

2/15

Méthode du gradient projeté

⚠ si $\min_{x \in \mathbb{R}^n} f(x)$ obtenu pour \hat{x} tel que $g_i(\hat{x}) < 0 \forall i$
alors \hat{x} optimum du problème sans contraintes

Sinon le min de $f(x)$ sous contraintes obtenu pour \hat{x} sur la frontière des réalisables

Méthode du gradient projeté = méthode primale

Méthode du gradient projeté

Principe : adapter les méthodes de gradient des problèmes d'optimisation sans contrainte

\implies projeter les déplacements sur la frontière de la région des réalisables

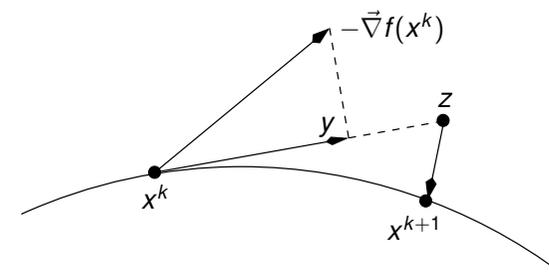
\implies génération d'une séquence de points réalisables

Exemple : méthode de Rosen (1960)

Méthode du gradient projeté de Rosen

Idee : projeter le gradient sur la frontière des réalisables
 \implies chemin sur la frontière dans la direction de la plus grande pente

\implies fonctionne essentiellement avec de contraintes linéaires :



contraintes linéaires \implies le déplacement reste sur l'ensemble des réalisables

Cours 13 : Méthodes de gradient

3/15

Cours 13 : Méthodes de gradient

4/15

$$(P) \quad \begin{array}{l} \min f(x) \\ \text{s.c.} \begin{cases} a_i^T x \leq b_i, & i \in I_1 \\ a_i^T x = b_i, & i \in I_2 \end{cases} \end{array}$$

Remarques :

- pas de contraintes de signes sur les $x_i \implies$ incluses dans I_1
- les contraintes de I_2 peuvent être transformées en I_1 (dédoublément)

Prochains slides : résolution de (P) par la méthode de Rosen

$$(P) \quad \begin{array}{l} \min f(x) \\ \text{s.c.} \begin{cases} a_i^T x \leq b_i, & i \in I_1 \\ a_i^T x = b_i, & i \in I_2 \end{cases} \end{array}$$

- On suppose qu'on connaît une solution x
 \implies au pire, utiliser un simplexe pour trouver un x
- $I^0 = \{i \in I_1 : a_i^T x = b_i\} \cup I_2$
 $=$ ensemble des contraintes serrées
- $A^0 =$ matrice composée des lignes $a_i^T, i \in I^0$
- Rosen \implies déplacements réalisables
 \implies si d est une direction alors $a_i^T d = 0 \forall i \in I^0$
 $\implies A^0 d = 0$

Déplacement $d \implies$ faire décroître f
 \implies minimiser $\vec{\nabla} f(x)^T d$

Principe général de la méthode

- 1 partir d'un point x réalisable
- 2 calculer une direction d solution de :

$$\begin{array}{l} \min \vec{\nabla} f(x)^T d \\ \text{s.c.} \begin{cases} A^0 d = 0 \\ \|d\| = 1 \end{cases} \end{array}$$

- 3 se déplacer selon d pour obtenir un nouvel x
- 4 si l'on n'est pas sur l'optimum, revenir en 1

Problème : peut-on résoudre ce problème rapidement ?

Théorème

- A^0 matrice $q \times n, q \leq n$, de rang q
 \implies les a_i^T sont linéairement indépendants
(pas de dégénérescence)

- la solution optimale du problème :

$$\begin{array}{l} \min \vec{\nabla} f(x)^T d \\ \text{s.c.} \begin{cases} A^0 d = 0 \\ \|d\| = 1 \end{cases} \end{array}$$

est $d = \bar{y} / \|\bar{y}\|$, où $\bar{y} =$ projection de $-\vec{\nabla} f(x)$ sur $\{y : A^0 y = 0\}$

- $\bar{y} = -P^0 \vec{\nabla} f(x) = -(Id - A^{0T} [A^0 A^{0T}]^{-1} A^0) \vec{\nabla} f(x)$

$P^0 =$ matrice de projection sur $\{y : A^0 y = 0\}$

$Id =$ matrice identité

Résolution de (P) par la méthode de Rosen (4/9)

Démonstration :

soit $S^0 = \{y : A^0 y = 0\}$

soit $S^{0\perp}$ = sous-espace orthogonal à S^0 dans \mathbb{R}^n

$\forall x \in \mathbb{R}^n, \exists \bar{y} \in S^0$ et $\exists \bar{z} \in S^{0\perp}$ tels que $x = \bar{y} + \bar{z}$

\implies en particulier $-\vec{\nabla} f(x) = \bar{y} + \bar{z}$

$\implies -\vec{\nabla} f(x)^T d = \bar{y}^T d + \bar{z}^T d$

Or $d \in S^0 \implies \bar{z}^T d = 0 \implies -\vec{\nabla} f(x)^T d = \bar{y}^T d$

$\implies \vec{\nabla} f(x)^T d$ minimal quand $\bar{y}^T d$ maximal

$d \in S^0, \|d\| = 1$ et $\bar{y} \in S^0 \implies \bar{y}^T d$ maximal pour $d = \bar{y}/\|\bar{y}\|$

\implies il reste maintenant à trouver l'expression de \bar{y}

Résolution de (P) par la méthode de Rosen (5/9)

caractérisation de \bar{y} :

remarque : $S^{0\perp}$ est le sous-espace de \mathbb{R}^n généré par les colonnes de A^{0T}

$\implies \forall \bar{z} \in S^{0\perp}, \exists u$ tel que $\bar{z} = A^{0T} u$

$\implies \exists u$ tel que $-\vec{\nabla} f(x) = \bar{y} + A^{0T} u$

$\bar{y} \in S^0 \implies A^0 \bar{y} = 0 \implies A^0 \bar{y} = -A^0 \vec{\nabla} f(x) - A^0 A^{0T} u = 0$

Or A^0 de rang $q \implies A^0 A^{0T}$ inversible

$\implies u = -[A^0 A^{0T}]^{-1} A^0 \vec{\nabla} f(x)$

$\bar{y} = -\vec{\nabla} f(x) - A^{0T} u = -\vec{\nabla} f(x) + A^{0T} [A^0 A^{0T}]^{-1} A^0 \vec{\nabla} f(x)$
 $= -\left(Id - A^{0T} [A^0 A^{0T}]^{-1} A^0 \right) \vec{\nabla} f(x)$

CQFD

Résolution de (P) par la méthode de Rosen (6/9)

Rappel du théorème

La solution optimale du problème :

$$\begin{aligned} \min \quad & \vec{\nabla} f(x)^T d \\ \text{s.c.} \quad & A^0 d = 0, \quad \|d\| = 1 \end{aligned}$$

est $d = \bar{y}/\|\bar{y}\|$, où \bar{y} = projection de $-\vec{\nabla} f(x)$ sur $\{y : A^0 y = 0\}$

et $\bar{y} = -\left(Id - A^{0T} [A^0 A^{0T}]^{-1} A^0 \right) \vec{\nabla} f(x)$

si $\bar{y} \neq 0$ alors \bar{y} est une direction de descente de gradient :

$$\vec{\nabla} f(x)^T \bar{y} = -(\bar{y}^T + \bar{z}^T) \bar{y} = -\bar{y}^T \bar{y} < 0$$

Résolution de (P) par la méthode de Rosen (7/9)

Rappel de la méthode de Rosen

- 1 partir d'un point x réalisable
- 2 calculer la direction $d = \bar{y}/\|\bar{y}\|$
- 3 se déplacer selon d pour obtenir un nouvel x réalisable
- 4 si l'on n'est pas sur l'optimum, revenir en 1

on peut calculer la longueur de déplacement max selon d :

$$\alpha_{\max} = \max\{\alpha \geq 0 : x + \alpha \bar{y} \in X\}$$

où $X = \{x \in \mathbb{R}^n : a_i x \leq b_i \forall i \in I_1 \text{ et } a_i x = b_i \forall i \in I_2\}$

prochain point réalisable : $x' = x + \bar{\alpha} \bar{y}$ avec $\bar{\alpha} = \min_{\alpha \in [0, \alpha_{\max}]} f(x + \alpha \bar{y})$

en x' , recalculer la nouvelle matrice de projection, etc

Problème : et si $\bar{y} = 0$?

$$\bar{y} = -\vec{\nabla}f(x) - A^{0T}u$$

$$\Rightarrow \begin{cases} \vec{\nabla}f(x) + A^{0T}u = 0 \\ u = -[A^0 A^{0T}]^{-1} A^0 \vec{\nabla}f(x) \end{cases}$$

Si $u \geq 0$ alors :

damned, mais ce sont les conditions de Kuhn et Tucker!!!!

Rappel : conditions de Kuhn et Tucker

condition nécessaire pour que \hat{x} = minimum local de f :

$\exists \lambda_i \geq 0, i = 1, \dots, m$, tels que :

$$\textcircled{1} \vec{\nabla}f(\hat{x}) + \sum_{i=1}^m \lambda_i \vec{\nabla}g_i(\hat{x}) = 0 \quad \textcircled{2} \lambda_i g_i(\hat{x}) = 0, \quad \forall i \in \{1, \dots, m\}$$

Donc si $\bar{y} = 0$ et $u \geq 0$:

Kuhn et Tucker $\Rightarrow x$ est un optimum local

si $\bar{y} = 0$ et u a des composants strictement négatifs :

$\Rightarrow x$ n'est pas optimal (Kuhn et Tucker = cond nécessaire)

\Rightarrow trouver une nouvelle direction d

supprimer de A^0 une ligne a_i^T pour laquelle $u_i < 0$

\Rightarrow nouvelle matrice de projection

\Rightarrow nouvelle direction \bar{y}' telle que $\bar{y}' \neq 0$ (\Rightarrow descente de gradient)

choisir par exemple la ligne a_i^T pour laquelle u_i min

\Rightarrow complète la méthode du gradient de Rosen

Méthode du gradient projeté de Rosen

Algorithme de Rosen

1 étape $k = 0$: on part du point x^0

2 étape k : on est sur le point x^k

Trouver l'ensemble $I^0(x^k)$ des contraintes serrées

Soit $L^0 = I^0(x^k)$

3 Soit A^0 la matrice des lignes $a_i^T, i \in L^0$

calculer $P^0 = Id - A^{0T} [A^0 A^{0T}]^{-1} A^0$

calculer $y^k = -P^0 \vec{\nabla}f(x^k)$

si $y^k = 0$ alors aller en **5**

4 $y^k \neq 0$: calculer $\alpha_{\max} = \max\{\alpha \geq 0 : x^k + \alpha y^k \in X\}$

calculer x^{k+1} tel que $f(x^{k+1}) = \min_{\alpha \in [0, \alpha_{\max}]} \{f(x^k + \alpha y^k)\}$

$k \leftarrow k + 1$; retourner en **2**

5 calculer $u = -[A^0 A^{0T}]^{-1} A^0 \vec{\nabla}f(x)$

si $u \geq 0$ alors x^k optimum (local)

sinon soit u_i l'élément le plus négatif de u

$L^0 \leftarrow L^0 - \{i\}$; retourner en **3**