ISUP — Module d'Optimisation

Transparents de cours

année 2010–2011

Christophe Gonzales

Optimisation

Christophe Gonzales

LIP6 – Université Paris 6, France

Objectifs du cours

Présentation des algorithmes fondamentaux d'optimisation

- optimisation linéaire
  - algorithme simplexe
  - dualité
  - applications
- optimisation non linéaire
  - méthodes de gradient
  - conditions de Kuhn et Tucker
  - programmation quadratique

Plan du cours (1/2)

Partie I : programmation linéaire

1. Forme générale, standard, canonique
2. L'algorithme du simplexe
3. Pièges du simplexe (dégénérescence. . .)
4. Phases I et II
5. Aspect géométrique : polyèdres ; points extrêmes
6. Dualité en programmation linéaire
7. Théorème d'existence et de dualité
8. Lemme de Minkowski-Farkas
9. Applications pratiques

Plan du cours (2/2)

Partie II : optimisation sans contraintes

1. Rappels d'optimisation sans contraintes
2. Optimisation uni-dimensionnelle
3. Méthodes de gradient et de gradient conjugué

Partie III : programmation non linéaire

1. Généralités : programmation convexe ; lagrangien
2. Théorème du col en programmation convexe
3. Conditions de Kuhn et Tucker en programmation convexe
4. Programmes quadratiques
5. Rappels sur les formes quadratiques. Méthode de Wolfe en PQ
6. Programmation convexe à contraintes linéaires

Introduction à la programmation linéaire (1/5)

Problème : combien d'argent doit-on dépenser pour couvrir les besoins énergétiques (2000 Kcal), en protéines (55g) et en calcium (800mg) pour une journée ?

nourriture	taille	Kcal	protéines	calcium	prix
céréales	28g	110	4g	2mg	3
poulet	100g	205	32g	12mg	24
œufs	2	160	13g	54mg	13
lait	237cl	160	8g	285mg	9
clafoutis	170g	420	4g	22mg	20
cassoulet	260g	260	14g	80mg	19

Introduction à la programmation linéaire (2/5)

nourriture	taille	Kcal	protéines	calcium	prix
céréales	28g	110	4g	2mg	3
poulet	100g	205	32g	12mg	24
œufs	2	160	13g	54mg	13
lait	237cl	160	8g	285mg	9
clafoutis	170g	420	4g	22mg	20
cassoulet	260g	260	14g	80mg	19

Exemple : 10 rations de cassoulet $\Longrightarrow$ 2600Kcal, 140g de protéines, 800mg de calcium, prix = 190

Peut-on faire mieux ?

Introduction à la programmation linéaire (3/5)

nourriture	taille	Kcal	protéines	calcium	prix	quantité
céréales	28g	110	4g	2mg	3	x_1
poulet	100g	205	32g	12mg	24	x_2
œufs	2	160	13g	54mg	13	x_3
lait	237cl	160	8g	285mg	9	x_4
clafoutis	170g	420	4g	22mg	20	x_5
cassoulet	260g	260	14g	80mg	19	x_6

Problème : combien d'argent doit-on dépenser pour couvrir les besoins énergétiques (2000 Kcal), en protéines (55g) et en calcium (800mg) pour une journée ?

$$\min 3x_1 + 24x_2 + 13x_3 + 9x_4 + 20x_5 + 19x_6$$
$$s.c. \begin{cases} 110x_1 + 205x_2 + 160x_3 + 160x_4 + 420x_5 + 260x_6 \geq 2000 \\ 4x_1 + 32x_2 + 13x_3 + 8x_4 + 4x_5 + 14x_6 \geq 55 \\ 2x_1 + 12x_2 + 54x_3 + 285x_4 + 22x_5 + 80x_6 \geq 800 \\ x_1 \geq 0,\ x_2 \geq 0,\ x_3 \geq 0,\ x_4 \geq 0,\ x_5 \geq 0,\ x_6 \geq 0 \end{cases}$$

Introduction à la programmation linéaire (4/5)

Contraintes additionnelles pour menus variés :

nourriture	contrainte
céréales	au plus 4 rations par jour
poulet	au plus 3 rations par jour
œufs	au plus 2 rations par jour
lait	au plus 8 rations par jour
clafoutis	au plus 2 rations par jour
cassoulet	au plus 2 rations par jour

Nouveau Problème : combien d'argent doit-on dépenser pour couvrir les besoins énergétiques (2000 Kcal), en protéines (55g) et en calcium (800mg) pour une journée sous les contraintes ci-dessus ?

Introduction à la programmation linéaire (5/5)

nourriture	contrainte
céréales	au plus 4 rations par jour
poulet	au plus 3 rations par jour
œufs	au plus 2 rations par jour
lait	au plus 8 rations par jour
clafoutis	au plus 2 rations par jour
cassoulet	au plus 2 rations par jour

$$\min 3x_1 + 24x_2 + 13x_3 + 9x_4 + 20x_5 + 19x_6$$
$$s.c. \begin{cases} 110x_1 + 205x_2 + 160x_3 + 160x_4 + 420x_5 + 260x_6 \geq 2000 \\ 4x_1 + 32x_2 + 13x_3 + 8x_4 + 4x_5 + 14x_6 \geq 55 \\ 2x_1 + 12x_2 + 54x_3 + 285x_4 + 22x_5 + 80x_6 \geq 800 \\ 0 \leq x_1 \leq 4,\ 0 \leq x_2 \leq 3,\ 0 \leq x_3 \leq 2, \\ 0 \leq x_4 \leq 8,\ 0 \leq x_5 \leq 2,\ 0 \leq x_6 \leq 2 \end{cases}$$

Programmation linéaire

Définition

programme linéaire = $\left\{ \begin{array}{l} \text{fonction objectif linéaire} \\ \text{contraintes linéaires} \end{array} \right.$

$$\max 5x_1 - 2x_2 + 10x_4$$
$$\text{s.c.} \left\{ \begin{array}{rcrcrcrcrcr} x_1 & + & 20x_2 & & & & & + & 20x_5 & \geq & 200 \\ & + & 5x_2 & + & 5x_3 & & & & & \leq & 55 \\ 2x_1 & + & 12x_2 & + & 10x_3 & + & 5x_4 & + & 2x_5 & = & 80 \\ \multicolumn{11}{l}{x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0, \quad x_4 \geq 0, \quad x_5 \geq 0} \end{array} \right.$$

Programme linéaire sous forme standard

Forme standard

$$\max \sum_{j=1}^{n} c_j x_j$$
$$\text{s.c.} \left\{ \begin{array}{rl} \displaystyle\sum_{j=1}^{n} a_{ij} x_j \leq b_i & (i = 1, 2, \ldots, m) \\ x_j \geq 0 & (j = 1, 2, \ldots, n) \end{array} \right.$$

Forme standard en notation matricielle

$$\max c^T x$$
$$\text{s.c.} \left\{ \begin{array}{l} Ax \leq b \\ x \geq 0 \end{array} \right.$$

il existe différentes définitions !

Quelques définitions

$$\max \sum_{j=1}^{n} c_j x_j$$
$$\text{s.c.} \left\{ \begin{array}{rl} \displaystyle\sum_{j=1}^{n} a_{ij} x_j \leq b_i & (i = 1, 2, \ldots, m) \\ x_j \geq 0 & (j = 1, 2, \ldots, n) \end{array} \right.$$

- fonction objectif = $\max \sum_{j=1}^{n} c_j x_j$
- solution = un n-uplet $(x_1, \ldots, x_n)$
- solution réalisable = solution vérifiant toutes les contraintes

Comment trouver la solution réalisable optimale ?

Exemple d'algorithme simplexe (1/7)

Problème à résoudre :

$$
\begin{array}{ll}
\max & 5x_1 + 4x_2 + 3x_3 \\
s.c. & 2x_1 + 3x_2 + x_3 \leq 5 \\
 & 4x_1 + x_2 + 2x_3 \leq 11 \\
 & 3x_1 + 4x_2 + 2x_3 \leq 8 \\
 & x_1 \geq 0,\ x_2 \geq 0,\ x_3 \geq 0
\end{array}
$$

1. introduire des «*variables d'écart*» *pour remplacer les* $\leq$ *par des* $=$
2. appeler z la fonction objectif

$$
\begin{array}{l}
x_4 = 5 - 2x_1 - 3x_2 - x_3 \\
x_5 = 11 - 4x_1 - x_2 - 2x_3 \\
x_6 = 8 - 3x_1 - 4x_2 - 2x_3 \\
z = 5x_1 + 4x_2 + 3x_3 \\
x_1 \geq 0,\ x_2 \geq 0,\ x_3 \geq 0,\ x_4 \geq 0,\ x_5 \geq 0,\ x_6 \geq 0
\end{array}
$$

Exemple d'algorithme simplexe (2/7)

$$
\begin{array}{l}
x_4 = 5 - 2x_1 - 3x_2 - x_3 \\
x_5 = 11 - 4x_1 - x_2 - 2x_3 \\
x_6 = 8 - 3x_1 - 4x_2 - 2x_3 \\
z = 5x_1 + 4x_2 + 3x_3 \\
x_1 \geq 0,\ x_2 \geq 0,\ x_3 \geq 0,\ x_4 \geq 0,\ x_5 \geq 0,\ x_6 \geq 0
\end{array}
$$

résoudre $\max z \quad s.c.\ x_1, x_2, x_3, x_4, x_5, x_6 \geq 0$
$\iff$ résoudre problème d'origine

$(0, 0, 0, 5, 11, 8)$ = solution réalisable.

Idée force du simplexe

1. partir d'une solution réalisable x^0
2. étant donné une solution réalisable x^i, chercher une solution réalisable x^{i+1} «voisine» telle que z augmente
3. Revenir en 2 tant que l'on peut trouver un tel x^{i+1}. Sinon on a trouvé un optimum.

Exemple d'algorithme simplexe (3/7)

$$\begin{aligned} x_4 &= \color{red}{5 - 2x_1} - 3x_2 - x_3 \\ x_5 &= \color{red}{11 - 4x_1} - x_2 - 2x_3 \\ x_6 &= \color{red}{8 - 3x_1} - 4x_2 - 2x_3 \\ z &= \color{red}{5x_1} + 4x_2 + 3x_3 \end{aligned}$$
$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0, \quad x_4 \geq 0, \quad x_5 \geq 0, \quad x_6 \geq 0$$

$(0, 0, 0, 5, 11, 8)$ = solution réalisable, $z = 0$

- si on augmente la valeur de x_1, on augmente z
- si $x_1 = 2$ alors $(2, 0, 0, 1, 3, 2)$ réalisable et $z = 10$
- si $x_1 = 3$ alors $(3, 0, 0, -1, -1, -1)$ non réalisable

$\Longrightarrow$ ne pas trop augmenter x_1

Exemple d'algorithme simplexe (4/7)

Idée force

- choisir d'augmenter une variable ayant un coefficient positif dans z
- augmenter cette variable tant que les autres ne deviennent pas négatives

$$\begin{aligned} x_4 &= \color{red}{5 - 2x_1} - 3x_2 - x_3 \\ x_5 &= \color{red}{11 - 4x_1} - x_2 - 2x_3 \\ x_6 &= \color{red}{8 - 3x_1} - 4x_2 - 2x_3 \\ z &= \color{red}{5x_1} + 4x_2 + 3x_3 \end{aligned}$$

$$\begin{aligned} x_4 &\Longrightarrow x_1 \leq 5/2 = 2,5 \\ x_5 &\Longrightarrow x_1 \leq 11/4 = 2,75 \\ x_6 &\Longrightarrow x_1 \leq 8/3 \approx 2,66 \end{aligned}$$

$\color{red}{\Longrightarrow x_1 = \min\{5/2, 11/4, 8/3\} = 5/2}$

Exemple d'algorithme simplexe (5/7)

augmentation de la valeur d'une variable
$\Longrightarrow$ annulation d'une autre

$$\begin{aligned} x_4 &= 5 - 2x_1 - 3x_2 - x_3 \\ x_5 &= 11 - 4x_1 - x_2 - 2x_3 \\ x_6 &= 8 - 3x_1 - 4x_2 - 2x_3 \\ z &= 5x_1 + 4x_2 + 3x_3 \end{aligned}$$

$x_1 = 5/2 \Longrightarrow x_4 = 0$

Idée force

1. Toujours placer à gauche des signe «=» les variables $\neq 0$
2. Exprimer ces variables en fonction des variables = 0

$\Longrightarrow$ exprimer x_1 en fonction des autres variables

$$x_1 = \frac{5}{2} - \frac{3}{2}x_2 - \frac{1}{2}x_3 - \frac{1}{2}x_4$$

Exemple d'algorithme simplexe (6/7)

$$x_1 = \frac{5}{2} - \frac{3}{2}x_2 - \frac{1}{2}x_3 - \frac{1}{2}x_4$$

$$\begin{aligned} x_5 &= 11 - 4x_1 - x_2 - 2x_3 \\ &= 11 - 4\left(\tfrac{5}{2} - \tfrac{3}{2}x_2 - \tfrac{1}{2}x_3 - \tfrac{1}{2}x_4\right) - x_2 - 2x_3 \\ &= 1 + 5x_2 + 2x_4 \end{aligned}$$

faire le même calcul pour x_6 et z :

$$\begin{aligned} x_1 &= \tfrac{5}{2} - \tfrac{3}{2}x_2 - \tfrac{1}{2}x_3 - \tfrac{1}{2}x_4 \\ x_5 &= 1 + 5x_2 \phantom{- \tfrac{1}{2}x_3} + 2x_4 \\ x_6 &= \tfrac{1}{2} + \tfrac{1}{2}x_2 - \tfrac{1}{2}x_3 + \tfrac{3}{2}x_4 \\ z &= \tfrac{25}{2} - \tfrac{7}{2}x_2 + \tfrac{1}{2}x_3 - \tfrac{5}{2}x_4 \end{aligned}$$

$\Longrightarrow$ recommencer avec x_3

Exemple d'algorithme simplexe (7/7)

Après augmentation de x_3 :

$$\begin{aligned} x_1 &= 2 - 2x_2 - 2x_4 + x_6 \\ x_5 &= 1 + 5x_2 + 2x_4 \\ x_3 &= 1 + x_2 + 3x_4 - 2x_6 \\ z &= 13 - 3x_2 - x_4 - x_6 \end{aligned}$$

⚠ Tous les coefficients de z sont négatifs

$\Longrightarrow$ on est à l'optimum

Solution optimale : $(2, 0, 1, 0, 1, 0)$

L'algorithme du simplexe

- variables à gauche des signe = variables «en base»
- les autres = variables «hors base»
- base réalisable $\Longleftrightarrow$ solution correspondante réalisable

Premier algorithme du simplexe

1. Choisir une solution réalisable x^0
2. Exprimer les variables en base ($\neq 0$) en fonction des variables hors base ($= 0$)
3. s'il existe un coefficient positif dans z, soit x_i la variable correspondante, sinon aller en 7
4. calculer la valeur maximale de x_i de manière à ce que les variables en base restent positives ou nulles. Soit x_j une des variables en base qui s'annule
5. placer x_i dans l'ensemble des variables en base (faire entrer la variable en base) et x_j dans l'ensemble des variables hors base (faire sortir de la base)
6. retourner en 2
7. on est à l'optimum. Les variables en base définissent la solution optimale

Références

Bibliographie

- V. Chvatal (1983) « *Linear Programming* », W.H. Freeman & Company.
- R. J. Vanderbei (1998) « *Linear Programming : Foundations and Extensions* », Kluwer Academic Publishers.
- M. Minoux (1983) « *Programmation mathématique, Théorie et Algorithmes* », Dunod.

Cours 2 : algorithme du simplexe

Christophe Gonzales

LIP6 – Université Paris 6, France

Plan du cours

1. Rappels sur l'algorithme vu la semaine dernière
2. Définition de l'algorithme du simplexe
3. Interprétation géométrique
4. Critères de choix pour les variables entrantes

Rappels sur le cours de la semaine dernière (1/10)

Forme standard

$$\max \sum_{j=1}^{n} c_j x_j$$

$$\text{s.c.} \begin{cases} \displaystyle\sum_{j=1}^{n} a_{ij} x_j \leq b_i & (i = 1, 2, \ldots, m) \\ x_j \geq 0 & (j = 1, 2, \ldots, n) \end{cases}$$

Algorithme de résolution (1/2)

1. Ajouter des variables d'écart $x_{n+1}, \ldots, x_{n+m}$:

$$\begin{aligned} x_{n+i} &= b_i - \sum_{j=1}^{n} a_{ij} x_j \quad (i = 1, 2, \ldots, m) \\ z &= \sum_{j=1}^{n} c_j x_j \end{aligned}$$

Rappels sur le cours de la semaine dernière (2/10)

Algorithme de résolution (2/2)

2. première solution réalisable : $x^0 = (0, \ldots, 0, b_1, b_2, \ldots, b_m)$
   variables en base : $x_{n+1}, \ldots, x_{n+m}$, hors base : $x_1, \ldots, x_n$
3. s'il existe un coefficient positif dans z, soit x_i la variable correspondante, sinon aller en 8
4. calculer la valeur maximale de x_i de manière à ce que les variables en base restent positives ou nulles. Soit x_j une des variables en base qui s'annule
5. faire entrer x_i en base, faire sortir x_j de la base
6. exprimer les variables en base en fonction des variables hors base
7. retourner en 3
8. on est à l'optimum. Les variables en base définissent la solution optimale

Rappels sur le cours de la semaine dernière (3/10)

Problème à résoudre :

$$\begin{array}{lrcl}
\max & x_1 + 5x_2 + x_3 & & \\
s.c. & x_1 + 3x_2 + x_3 & \leq & 3 \\
 & -x_1 + 3x_3 & \leq & 2 \\
 & 2x_1 + 4x_2 - x_3 & \leq & 4 \\
 & x_1 + 3x_2 - x_3 & \leq & 2 \\
 & x_1 \geq 0, \ x_2 \geq 0, \ x_3 \geq 0 & &
\end{array}$$

Première étape : ajouter des variables d'écart :

$$\begin{array}{lrlcl}
\max & x_1 + 5x_2 + x_3 & & & \\
s.c. & x_1 + 3x_2 + x_3 & + \ x_4 & = & 3 \\
 & -x_1 + 3x_3 & + x_5 & = & 2 \\
 & 2x_1 + 4x_2 - x_3 & + x_6 & = & 4 \\
 & x_1 + 3x_2 - x_3 & + x_7 & = & 2 \\
 & \multicolumn{4}{l}{x_1 \geq 0, \ x_2 \geq 0, \ x_3 \geq 0, \ x_4 \geq 0, \ x_5 \geq 0, \ x_6 \geq 0, \ x_7 \geq 0}
\end{array}$$

Rappels sur le cours de la semaine dernière (4/10)

Expression de z et des variables en base en fonction des variables hors base :

$$\begin{aligned}
x_4 &= 3 - x_1 - 3x_2 - x_3 \\
x_5 &= 2 + x_1 - 3x_3 \\
x_6 &= 4 - 2x_1 - 4x_2 + x_3 \\
x_7 &= 2 - x_1 - 3x_2 + x_3 \\
z &= x_1 + 5x_2 + x_3
\end{aligned}$$

variables en base : x_4, x_5, x_6, x_7

$\Longrightarrow$ solution réalisable = $(0, 0, 0, 3, 2, 4, 2)$

3 coefficients positifs dans $z \Longrightarrow x_1$, x_2 et x_3

$\Longrightarrow$ choix (au hasard) de faire rentrer x_1 en base

Rappels sur le cours de la semaine dernière (5/10)

$$\begin{aligned} x_4 &= 3 - x_1 - 3x_2 - x_3 \\ x_5 &= 2 + x_1 - 3x_3 \\ x_6 &= 4 - 2x_1 - 4x_2 + x_3 \\ x_7 &= 2 - x_1 - 3x_2 + x_3 \\ z &= x_1 + 5x_2 + x_3 \end{aligned}$$

4. calcul de la valeur optimale de x_1 :
   - augmenter $x_1 \Longrightarrow$ augmenter z
   - ne pas trop augmenter x_1 afin que x_4, x_5, x_6, x_7, restent ≥ 0

$$\begin{array}{ll} (x_4) & 3 - x_1 \geq 0 \\ (x_5) & 2 + x_1 \geq 0 \\ (x_6) & 4 - 2x_1 \geq 0 \\ (x_7) & 2 - x_1 \geq 0 \end{array}$$

$\Longrightarrow x_1 \leq 3,\ x_1 \geq -2,\ x_1 \leq 2,\ x_1 \leq 2 \Longrightarrow x_1 = 2$

$\Longrightarrow$ variable à sortir de la base : x_6 ou x_7

Rappels sur le cours de la semaine dernière (6/10)

$$\begin{aligned} x_4 &= 3 - x_1 - 3x_2 - x_3 \\ x_5 &= 2 + x_1 - 3x_3 \\ x_6 &= 4 - 2x_1 - 4x_2 + x_3 \\ x_7 &= 2 - x_1 - 3x_2 + x_3 \\ z &= x_1 + 5x_2 + x_3 \end{aligned}$$

5. choix (au hasard) de faire sortir x_7 de la base

$\Longrightarrow x_1 = 2 - 3x_2 + x_3 - x_7$

6. expression des variables en base en fonction des variables hors base :

$$\begin{aligned} x_4 &= 1 - 2x_3 + x_7 \\ x_5 &= 4 - 3x_2 - 2x_3 - x_7 \\ x_6 &= 2x_2 - x_3 + 2x_7 \\ x_1 &= 2 - 3x_2 + x_3 - x_7 \\ z &= 2 + 2x_2 + 2x_3 - x_7 \end{aligned}$$

Rappels sur le cours de la semaine dernière (7/10)

$$\begin{aligned} x_4 &= 1 - 2x_3 + x_7 \\ x_5 &= 4 - 3x_2 - 2x_3 - x_7 \\ x_6 &= 2x_2 - x_3 + 2x_7 \\ x_1 &= 2 - 3x_2 + x_3 - x_7 \\ z &= 2 + 2x_2 + 2x_3 - x_7 \end{aligned}$$

3. coefficients positifs dans z : x_2 et x_3

$\Longrightarrow$ choix (au hasard) : rentrer x_3 en base

$\Longrightarrow$ choix de la variable à sortir de la base :

$$\left.\begin{array}{ll} (x_4) & 1 - 2x_3 \geq 0 \\ (x_5) & 4 - 2x_3 \geq 0 \\ (x_6) & 0 - x_3 \geq 0 \\ (x_1) & 2 + x_3 \geq 0 \end{array}\right\} \Longrightarrow x_3 = \min\{b_i / - \text{coeff de } x_3 : \text{coeff} \neq 0\}$$

⚠ x_3 rentre en base, mais sa valeur est égale à 0

$\Longrightarrow$ la valeur de la fonction objectif ne change pas !

Rappels sur le cours de la semaine dernière (8/10)

6 expression des variables en base en fonction des variables hors base :

$$
\begin{aligned}
x_4 &= 1 - 4x_2 + 2x_6 - 3x_7 \\
x_5 &= 4 - 7x_2 + 2x_6 - 5x_7 \\
x_3 &= \phantom{1 -{}} 2x_2 - x_6 + 2x_7 \\
x_1 &= 2 - x_2 - x_6 + x_7 \\
z &= 2 + 6x_2 - 2x_6 + 3x_7
\end{aligned}
$$

3 coefficients positifs dans z : x_2 et x_3

$\Longrightarrow$ choix (au hasard) : rentrer x_2 en base :

$$
\left.
\begin{array}{ll}
(x_4) & 1 - 4x_2 \geq 0 \\
(x_5) & 4 - 7x_2 \geq 0 \\
(x_3) & \phantom{4 -{}} 2x_2 \geq 0 \\
(x_1) & 2 - x_2 \geq 0
\end{array}
\right\} \Longrightarrow x_4 \text{ sort de la base}
$$

Rappels sur le cours de la semaine dernière (9/10)

6 expression des variables en base en fonction des variables hors base :

$$
\begin{aligned}
x_2 &= \tfrac{1}{4} - \tfrac{1}{4}x_4 + \tfrac{1}{2}x_6 - \tfrac{3}{4}x_7 \\
x_5 &= \tfrac{9}{4} + \tfrac{7}{4}x_4 - \tfrac{3}{2}x_6 + \tfrac{1}{4}x_7 \\
x_3 &= \tfrac{1}{2} - \tfrac{1}{2}x_4 \phantom{{}+\tfrac{1}{2}x_6} + \tfrac{1}{2}x_7 \\
x_1 &= \tfrac{7}{4} + \tfrac{1}{4}x_4 - \tfrac{3}{2}x_6 + \tfrac{7}{4}x_7 \\
z &= \tfrac{7}{2} - \tfrac{3}{2}x_4 + \phantom{\tfrac{1}{2}}x_6 - \tfrac{3}{2}x_7
\end{aligned}
$$

3–6 rentrer x_6 en base et sortir x_1 :

$$
\left.
\begin{aligned}
x_2 &= \tfrac{5}{6} - \tfrac{1}{3}x_1 - \tfrac{1}{6}x_4 - \tfrac{1}{6}x_7 \\
x_5 &= \tfrac{1}{2} + \phantom{\tfrac{1}{3}}x_1 + \tfrac{3}{2}x_4 - \tfrac{3}{2}x_7 \\
x_3 &= \tfrac{1}{2} \phantom{{}-\tfrac{1}{3}x_1} - \tfrac{1}{2}x_4 + \tfrac{1}{2}x_7 \\
x_6 &= \tfrac{7}{6} - \tfrac{2}{3}x_1 + \tfrac{1}{6}x_4 + \tfrac{7}{6}x_7 \\
z &= \tfrac{14}{3} - \tfrac{2}{3}x_1 - \tfrac{4}{3}x_4 - \tfrac{1}{3}x_7
\end{aligned}
\right\} \Longrightarrow z : \text{coeffs négatifs} \Longrightarrow \text{optimum}
$$

Rappels sur le cours de la semaine dernière (10/10)

En résumé :

- plusieurs variables peuvent être candidates à entrer en base
  $\Longrightarrow$ critère de choix à définir
- plusieurs variables peuvent être candidates à sortir de la base
  $\Longrightarrow$ critère de choix à définir
- dégénérescence : certaines variables entrant en base peuvent avoir pour valeur 0 $\Longrightarrow$ la fonction objectif n'augmente pas
  $\Longrightarrow$ éviter que l'algorithme ne boucle

Notation en tableau (1/5)

Principe : placer toutes les variables du même côté

$$\begin{array}{rcl}
x_4 & = 3 & - \; x_1 - 3x_2 - \; x_3 \\
x_5 & = 2 & + \; x_1 \qquad\quad - 3x_3 \\
x_6 & = 4 & - 2x_1 - 4x_2 + \; x_3 \\
x_7 & = 2 & - \; x_1 - 3x_2 + \; x_3 \\
z & = & \quad\;\; x_1 + 5x_2 + \; x_3
\end{array}$$

$$\begin{array}{rrrrrrrl}
 & x_1 & + 3x_2 & + x_3 & + x_4 & & & = 3 \\
 & - x_1 & & + 3x_3 & & + x_5 & & = 2 \\
 & 2x_1 & + 4x_2 & - x_3 & & & + x_6 \quad & = 4 \\
 & x_1 & + 3x_2 & - x_3 & & & \quad + x_7 & = 2 \\
-z + & x_1 & + 5x_2 & + x_3 & & & & = 0
\end{array}$$

Notation en tableau (2/5)

La notation en tableau peut s'appliquer à toutes les étapes :

Avant pivot :

dictionnaire	*tableau*
$x_4 = 3 - x_1 - 3x_2 - x_3$	$x_1 + 3x_2 + x_3 + x_4 = 3$
$x_5 = 2 + x_1 - 3x_3$	$-x_1 + 3x_3 + x_5 = 2$
$x_6 = 4 - 2x_1 - 4x_2 + x_3$	$2x_1 + 4x_2 - x_3 + x_6 = 4$
$x_7 = 2 - x_1 - 3x_2 + x_3$	$x_1 + 3x_2 - x_3 + x_7 = 2$
$z = x_1 + 5x_2 + x_3$	$-z + x_1 + 5x_2 + x_3 = 0$

Après pivot : x_1 entre et x_7 sort

dictionnaire	*tableau*
$x_4 = 1 - 2x_3 + x_7$	$2x_3 + x_4 - x_7 = 1$
$x_5 = 4 - 3x_2 - 2x_3 - x_7$	$3x_2 + 2x_3 + x_5 + x_7 = 4$
$x_6 = 2x_2 - x_3 + 2x_7$	$-2x_2 + x_3 + x_6 - 2x_7 = 0$
$x_1 = 2 - 3x_2 + x_3 - x_7$	$x_1 + 3x_2 - x_3 + x_7 = 2$
$z = 2 + 2x_2 + 2x_3 - x_7$	$-z + 2x_2 + 2x_3 - x_7 = -2$

Notation en tableau (3/5)

Pivot en termes de dictionnaires

Faire entrer x_i et sortir x_j :
supp. que x_j est défini à gauche des «=» sur la $k$ème ligne

1. exprimer x_i en fonction des autres variables sur la $k$ème ligne
2. sur toutes les autres lignes, remplacer les x_i par cette expression

Pivot en termes de tableaux

Faire entrer x_i et sortir x_j :

1. diviser la seule ligne dont le coeff de x_j est $\neq 0$ ($k$ème ligne) par le coeff associé à x_i sur cette ligne
   $\Longrightarrow$ le coeff de x_i devient 1
2. pour toute autre ligne $r \neq k$, soustraire a_{ri} fois la $k$ème ligne $\Longrightarrow$ le coeff de x_i sur ces lignes devient 0

Notation en tableau (4/5)

Application du pivot directement sur les tableaux :

$$\begin{array}{rcl} x_1 + 3x_2 + x_3 + x_4 & & = 3 \\ -x_1 + 3x_3 + x_5 & & = 2 \\ 2x_1 + 4x_2 - x_3 + x_6 & & = 4 \\ x_1 + 3x_2 - x_3 + x_7 & & = 2 \\ -z + x_1 + 5x_2 + x_3 & & = 0 \end{array}$$

pivot : x_1 entre et x_7 sort

$$\begin{array}{rcl} 2x_3 + x_4 - x_7 & = & 1 \\ 3x_2 + 2x_3 + x_5 + x_7 & = & 4 \\ -2x_2 + x_3 + x_6 - 2x_7 & = & 0 \\ x_1 + 3x_2 - x_3 + x_7 & = & 2 \\ -z + 2x_2 + 2x_3 - x_7 & = & -2 \end{array}$$

ligne où x_7 est défini : 4ème ligne

Notation en tableau (5/5)

D'un point de vue informatique, stocker uniquement les nombres, pas les chaînes de caractères x_i :

$$\begin{array}{rcl} 2x_3 + x_4 - x_7 & = & 1 \\ 3x_2 + 2x_3 + x_5 + x_7 & = & 4 \\ -2x_2 + x_3 + x_6 - 2x_7 & = & 0 \\ x_1 + 3x_2 - x_3 + x_7 & = & 2 \\ -z + 2x_2 + 2x_3 - x_7 & = & -2 \end{array}$$

$\Longrightarrow$ tableau stocké sous forme informatique :

$$\begin{array}{rrrrrrr|r} 0 & 0 & 2 & 1 & 0 & 0 & -1 & 1 \\ 0 & 3 & 2 & 0 & 1 & 0 & 1 & 4 \\ 0 & -2 & 1 & 0 & 0 & 1 & -2 & 0 \\ 1 & 3 & -1 & 0 & 0 & 0 & 1 & 2 \\ \hline 0 & 2 & 2 & 0 & 0 & 0 & -1 & -2 \end{array}$$

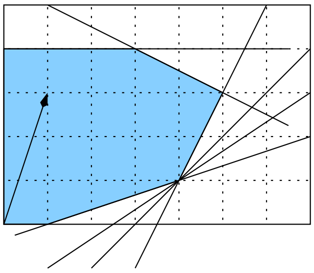
Algorithme du simplexe : 1ère version

1. examiner s'il existe un nombre positif sur la dernière ligne (excepté la dernière colonne qui vaut $-z$). S'il n'y en a pas, aller en 6. sinon, soit j l'index d'une de ces colonnes
2. pour chaque ligne, soit s le nombre dans la colonne la plus à droite et r le nombre dans la colonne j. Déterminer la ligne i ayant le plus petit ratio $s/r \geq 0$. Si les r de toutes les lignes sont négatives ou nulles, aller en 7
3. diviser la ligne i par son coefficient r
4. pour toutes les lignes $\neq i$, soit k le nombre stocké sur cette ligne à la colonne j. soustraire à la ligne k fois la ligne i
5. revenir en 1
6. on est à l'optimum. Les nombres égaux à 0 sur la dernière ligne (excepté la dernière colonne) déterminent la solution optimale.
7. Le problème n'est pas borné, i.e., le max de la fonction objectif est $+\infty$.

Interprétation géométrique de l'algorithme (1/8)

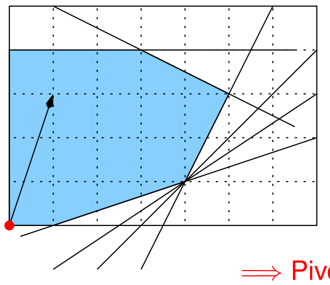
$$\max x_1 + 3x_2$$

$$\text{s.c.} \begin{cases} \qquad\quad x_2 \leq 4 \\ x_1 - 3x_2 \leq 1 \\ 2x_1 - 3x_2 \leq 5 \\ x_1 - x_2 \leq 3 \\ 2x_1 - x_2 \leq 7 \\ x_1 + 2x_2 \leq 11 \\ x_1 \geq 0, \quad x_2 \geq 0 \end{cases}$$



Interprétation géométrique de l'algorithme (2/8)

$$\begin{array}{rrrrrrrrrl}
 & & x_2 + x_3 & & & & & & = & 4 \\
 & x_1 & - 3x_2 & + x_4 & & & & & = & 1 \\
 & 2x_1 & - 3x_2 & & + x_5 & & & & = & 5 \\
 & x_1 & - x_2 & & & + x_6 & & & = & 3 \\
 & 2x_1 & - x_2 & & & & + x_7 & & = & 7 \\
 & x_1 & + 2x_2 & & & & & + x_8 & = & 11 \\
 -z + & x_1 & + 3x_2 & & & & & & = & 0
\end{array}$$

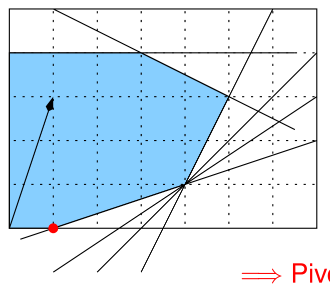


base : $x_3, x_4, x_5, x_6, x_7, x_8$

$\Longrightarrow$ Pivot : on fait entrer x_1 et sortir x_4

Interprétation géométrique de l'algorithme (3/8)

$$\begin{array}{rrrrrrrrrl}
 & & x_2 + x_3 & & & & & & = & 4 \\
 & x_1 & - 3x_2 & + x_4 & & & & & = & 1 \\
 & & 3x_2 & - 2x_4 & + x_5 & & & & = & 3 \\
 & & 2x_2 & - x_4 & & + x_6 & & & = & 2 \\
 & & 5x_2 & - 2x_4 & & & + x_7 & & = & 5 \\
 & & 5x_2 & - x_4 & & & & + x_8 & = & 10 \\
 -z & & + 6x_2 & - x_4 & & & & & = & -1
\end{array}$$

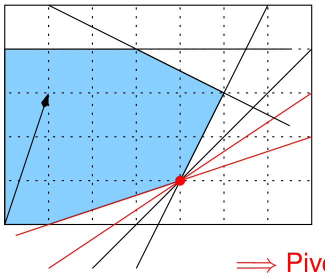


base : $x_1, x_3, x_5, x_6, x_7, x_8$

$\Longrightarrow$ Pivot : on fait entrer x_2 et sortir x_5

Interprétation géométrique de l'algorithme (4/8)

$$
\begin{array}{rcl}
x_3 + \frac{2}{3}x_4 - \frac{1}{3}x_5 & = & 3 \\
x_1 \quad - x_4 + x_5 & = & 4 \\
x_2 \quad - \frac{2}{3}x_4 + \frac{1}{3}x_5 & = & 1 \\
\frac{1}{3}x_4 - \frac{2}{3}x_5 + x_6 & = & 0 \\
\frac{4}{3}x_4 - \frac{5}{3}x_5 \quad + x_7 & = & 0 \\
\frac{7}{3}x_4 - \frac{5}{3}x_5 \quad + x_8 & = & 5 \\
-z \quad + 3x_4 - 2x_5 & = & -7
\end{array}
$$

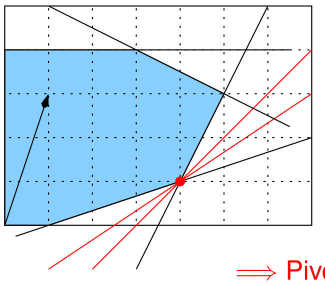


base : $x_1, x_2, x_3, x_6, x_7, x_8$

$\Longrightarrow$ Pivot : on fait entrer x_4 et sortir x_6

Interprétation géométrique de l'algorithme (5/8)

$$
\begin{array}{rcl}
x_3 \quad + x_5 - 2x_6 & = & 3 \\
x_1 \quad - x_5 + 3x_6 & = & 4 \\
x_2 \quad - x_5 + 2x_6 & = & 1 \\
x_4 - 2x_5 + 3x_6 & = & 0 \\
x_5 - 4x_6 + x_7 & = & 0 \\
3x_5 - 7x_6 \quad + x_8 & = & 5 \\
-z \quad + 4x_5 - 9x_6 & = & -7
\end{array}
$$



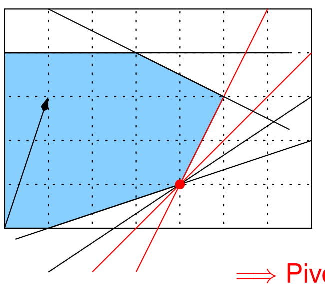
base : $x_1, x_2, x_3, x_4, x_7, x_8$

⚠ dégénérescence !!!!!

$\Longrightarrow$ Pivot : on fait entrer x_5 et sortir x_7

Interprétation géométrique de l'algorithme (6/8)

$$
\begin{array}{rcl}
x_3 \quad - 2x_6 - x_7 & = & 3 \\
x_1 \quad - x_6 + x_7 & = & 4 \\
x_2 \quad - 2x_6 + x_7 & = & 1 \\
x_4 \quad - 5x_6 + 2x_7 & = & 0 \\
x_5 - 4x_6 + x_7 & = & 0 \\
5x_6 - 3x_7 + x_8 & = & 5 \\
-z \quad + 7x_6 - 4x_7 & = & -7
\end{array}
$$



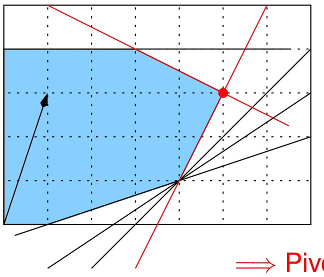
base : $x_1, x_2, x_3, x_4, x_5, x_8$

⚠ dégénérescence !!!!!

$\Longrightarrow$ Pivot : on fait entrer x_6 et sortir x_8

Interprétation géométrique de l'algorithme (7/8)

$$
\begin{array}{rcl}
x_3 + \frac{1}{5}x_7 - \frac{2}{5}x_8 &=& 1 \\
x_1 + \frac{2}{5}x_7 + \frac{1}{5}x_8 &=& 5 \\
x_2 - \frac{1}{5}x_7 + \frac{2}{5}x_8 &=& 3 \\
x_4 - x_7 + x_8 &=& 5 \\
x_5 - \frac{7}{5}x_7 + \frac{4}{5}x_8 &=& 4 \\
x_6 - \frac{3}{5}x_7 + \frac{1}{5}x_8 &=& 1 \\
-z + \frac{1}{5}x_7 - \frac{7}{5}x_8 &=& -14
\end{array}
$$

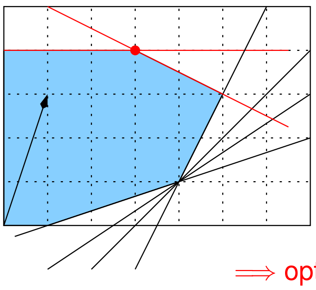


base : $x_1, x_2, x_3, x_4, x_5, x_6$

$\Longrightarrow$ Pivot : on fait entrer x_7 et sortir x_3

Interprétation géométrique de l'algorithme (8/8)

$$
\begin{array}{rcl}
5x_3 + x_7 - 2x_8 &=& 5 \\
x_1 - 2x_3 + x_8 &=& 3 \\
x_2 + x_3 &=& 4 \\
5x_3 + x_4 - x_8 &=& 10 \\
7x_3 + x_5 - 2x_8 &=& 11 \\
3x_3 + x_6 - x_8 &=& 4 \\
-z - x_3 - x_8 &=& -15
\end{array}
$$



base : $x_1, x_2, x_4, x_5, x_6, x_7$

$\Longrightarrow$ optimum !!!!!

Choix des variables entrantes (1/3)

choisir une variable dont le coeff dans la fonction objectif est > 0

$\Longrightarrow$ règle ambiguë : plusieurs variables peuvent être candidates

But : choisir la variable pour minimiser le nombre d'itérations de l'algorithme

Règle du plus grand coefficient

Choisir de faire entrer la variable qui a le plus grand coefficient dans la fonction objectif

grand coeff $\Longrightarrow$ le taux d'augmentation de la fonction objectif est élevé

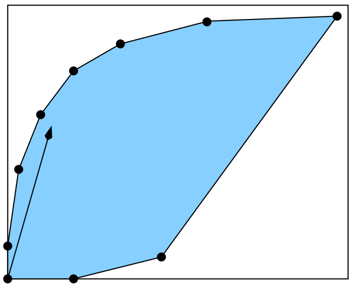
⚠ aucune garantie que ce soit optimal : la variable peut être contrainte à prendre une petite valeur $\Longrightarrow$ peu de variation de la fonction objectif

Choix des variables entrantes (2/3)

Règle du plus grand accroissement de z

Choisir de faire entrer la variable qui fait le plus augmenter la fonction objectif

⚠ aucune garantie que ce soit optimal : on peut faire beaucoup augmenter localement la fonction objectif et rester coincé plus tard :



Choix des variables entrantes (3/3)

La règle du plus grand coefficient est plus souvent utilisée que la règle du plus grand accroissement car elle est calculable plus rapidement

Dégénérescence

Avec les deux règles précédentes, on peut cycler (boucler indéfiniment sur les mêmes itérations)

Cours 3 : Problèmes de dégénérescence

Christophe Gonzales

LIP6 – Université Paris 6, France

Éviter les problèmes de dégénérescence

Dégénérescence $\Longrightarrow$ possibilité de cycles

$\Longrightarrow$ exemple de Beale (1955)

heureusement, les cycles sont rares !

⚠ il existe des méthodes garantissant que l'on ne cycle pas :

1. méthodes des perturbations
2. méthode lexicographique
3. règle du plus petit indice
4. ....

Méthode des perturbations

Idée force

- la dégénérescence est très rare
  $\Longrightarrow$ c'est plutôt un accident
- on peut la supprimer en «perturbant» très légèrement le tableau simplexe $\Longrightarrow$ ajouter des ϵ aux b_i
- $\epsilon \Longrightarrow$ les solutions obtenues par l'algo du simplexe $\approx$ solution du problème d'origine

Fiabilité de la méthode des perturbations (1/2)

ajouter le même ϵ aux b_i $\Longrightarrow$ méthode peu fiable

exemple :

$$
\begin{array}{llllllllll}
\max & 10x_1 - 57x_2 - 9x_3 - 24x_4 + 100x_5 & & & & & \\
s.c. & & x_5 + x_6 & & & & = 1 + \epsilon \\
& 0,5x_1 - 5,5x_2 - 2,5x_3 + 9x_4 + & x_5 & + x_7 & & & = 1 + \epsilon \\
& 0,5x_1 - 1,5x_2 - 0,5x_3 + x_4 + & x_5 & & + x_8 & & = 1 + \epsilon \\
& x_1 \qquad\qquad\qquad\qquad\qquad + & x_5 & & & + x_9 & = 2 + \epsilon \\
& x_1 \geq 0, \ x_2 \geq 0, \ x_3 \geq 0, \ x_4 \geq 0, \ x_5 \geq 0 & & & & &
\end{array}
$$

base réalisable : (x_6, x_7, x_8, x_9)

$$
\begin{array}{ll}
x_6 = 1 + \epsilon & - x_5 \\
x_7 = 1 + \epsilon - 0,5x_1 + 5,5x_2 + 2,5x_3 - 9x_4 & - x_5 \\
x_8 = 1 + \epsilon - 0,5x_1 + 1,5x_2 + 0,5x_3 - x_4 & - x_5 \\
x_9 = 2 + \epsilon - x_1 & - x_5 \\
z = 10x_1 - 57x_2 - 9x_3 - 24x_4 & + 100x_5
\end{array}
$$

$\Longrightarrow$ faire entrer x_5 et sortir (e.g.) x_6

Fiabilité de la méthode des perturbations (2/2)

$$
\begin{array}{ll}
x_6 = 1 + \epsilon & - x_5 \\
x_7 = 1 + \epsilon - 0,5x_1 + 5,5x_2 + 2,5x_3 - 9x_4 & - x_5 \\
x_8 = 1 + \epsilon - 0,5x_1 + 1,5x_2 + 0,5x_3 - x_4 & - x_5 \\
x_9 = 2 + \epsilon - x_1 & - x_5 \\
z = 10x_1 - 57x_2 - 9x_3 - 24x_4 & + 100x_5
\end{array}
$$

faire entrer x_5 et sortir x_6 :

$$
\begin{array}{ll}
x_5 = 1 + \epsilon & - x_6 \\
x_7 = - 0,5x_1 + 5,5x_2 + 2,5x_3 - 9x_4 & + x_6 \\
x_8 = - 0,5x_1 + 1,5x_2 + 0,5x_3 - x_4 & + x_6 \\
x_9 = 1 - x_1 & + x_6 \\
z = 100 + 100\epsilon + 10x_1 - 57x_2 - 9x_3 - 24x_4 & - 100x_6
\end{array}
$$

$\Longrightarrow$ on a perdu les ϵ mais il y a toujours dégénérescence

$\Longrightarrow$ ici, le simplexe cycle au bout de 6 itérations

Solution plus fiable

même ϵ pour chaque b_i $\Longrightarrow$ les ϵ s'éliminent d'une ligne sur l'autre

$\Longrightarrow$ choisir des ϵ_i très différents pour chaque b_i

méthode des perturbations

- choisir $0 < \epsilon_m \ll \epsilon_{m-1} \ll \cdots \ll \epsilon_2 \ll \epsilon_1 \ll 1$
- appliquer l'algorithme du simplexe sur le tableau perturbé

choix possible : $\epsilon_1 = \epsilon$, $\epsilon_2 = \epsilon^2$, $\epsilon_3 = \epsilon^3$, etc

Solution plus fiable (suite)

$$\begin{aligned}
x_6 &= 1 + \epsilon_1 - 0,5x_1 + 5,5x_2 + 2,5x_3 - 9x_4 - x_5\\
x_7 &= 1 + \epsilon_2 - 0,5x_1 + 1,5x_2 + 0,5x_3 - x_4 - x_5\\
x_8 &= 1 + \epsilon_3 - x_5\\
x_9 &= 2 + \epsilon_4 - x_1 - x_5\\
z &= 10x_1 - 57x_2 - 9x_3 - 24x_4 + 100x_5
\end{aligned}$$

contrainte : $0 < \epsilon_4 \ll \epsilon_3 \ll \epsilon_2 \ll \epsilon_1 \ll 1$

traiter les ϵ_i comme des variables :

$$\begin{aligned}
x_6 &= 1 + \epsilon_1 - 0,5x_1 + 5,5x_2 + 2,5x_3 - 9x_4 - x_5\\
x_7 &= 1 + \epsilon_2 - 0,5x_1 + 1,5x_2 + 0,5x_3 - x_4 - x_5\\
x_8 &= 1 + \epsilon_3 - x_5\\
x_9 &= 2 + \epsilon_4 - x_1 - x_5\\
z &= 10x_1 - 57x_2 - 9x_3 - 24x_4 + 100x_5
\end{aligned}$$

$\Longrightarrow$ faire entrer x_5 et sortir x_8

Solution plus fiable (fin)

Après pivotage :

$$\begin{aligned}
x_6 &= \epsilon_1 - \epsilon_3 - 0,5x_1 + 5,5x_2 + 2,5x_3 - 9x_4 + x_8\\
x_7 &= + \epsilon_2 - \epsilon_3 - 0,5x_1 + 1,5x_2 + 0,5x_3 - x_4 + x_8\\
x_5 &= 1 + \epsilon_3 - x_8\\
x_9 &= 1 - \epsilon_3 + \epsilon_4 - x_1 + x_8\\
z &= 100 + 100\epsilon_3 \quad 10x_1 - 57x_2 - 9x_3 - 24x_4 - 100x_8
\end{aligned}$$

pivotages $\Longrightarrow$ les ϵ_i se mélangent sur les m premières colonnes :

m premières colonnes = $r = r_0 + \sum_{j=1}^{m} r_j\epsilon_j$

choix de la variable sortante = la ligne de plus petit r

La méthode lexicographique (1/2)

choix de la variable sortante = la ligne de plus petit $r = r_0 + \sum_{j=1}^{m} r_j\epsilon_j$

or la règle : $0 < \epsilon_4 \ll \epsilon_3 \ll \epsilon_2 \ll \epsilon_1 \ll 1$

$\Longrightarrow$ si ligne i = $r = r_0 + \sum_{j=1}^{m} r_j\epsilon_j$ et ligne j = $s = s_0 + \sum_{j=1}^{m} s_j\epsilon_j$

alors $r < s \Longleftrightarrow r_k < s_k$ pour k le plus petit indice tel que $r_k \neq s_k$.

$\Longrightarrow$ ordre lexicographique

La méthode lexicographique (2/2)

Méthode lexicographique

- créer une colonne par ϵ_i
- choix des variables sortantes = la ligne de plus petit r (au sens lexicographique)

Théorème

L'algorithme du simplexe se termine, i.e., ne cycle pas, dès lors que les variables sortantes sont choisies avec la règle lexicographique

⚠ cette règle n'est à appliquer qu'en cas de dégénérescence

La méthode du plus petit indice

Règle du plus petit indice

Lorsque plusieurs variables sont candidates à entrer en base (selon un certain critère (e.g., les règles ci-dessus)), choisir celle qui a le plus petit indice dans le tableau. Faire de même avec les variables sortant de la base.

Théorème — Bland (1977)

si on applique cette nouvelle règle, l'algorithme du simplexe ne peut cycler.

⚠ cette règle n'est à appliquer qu'en cas de dégénérescence

Cours 4 : méthode révisée du simplexe

Christophe Gonzales

LIP6 – Université Paris 6, France

En route vers la méthode révisée

Point de départ :

- à chaque itération du simplexe, on recalcule entièrement le tableau
- seule une petite partie du tableau sert pour une itération donnée

$\Longrightarrow$ perte de temps

principe de la méthode révisée du simplexe

Essayer de reconstruire cette petite partie à partir du tableau d'origine

$\Longrightarrow$ a priori, moins de calculs à effectuer

- méthode utilisée pour résoudre les gros problèmes linéaires (milliers de variables et de contraintes), en général peu denses (beaucoup de 0)

Relation tableau à l'itération n – tableau d'origine (1/6)

Problème de départ :

$$\begin{array}{lrrrrl}
\max & 19x_1 + & 13x_2 + & 12x_3 + & 17x_4 & \\
s.c. & 3x_1 + & 2x_2 + & x_3 + & 2x_4 & \leq 255 \\
 & x_1 + & x_2 + & x_3 + & x_4 & \leq 117 \\
 & 4x_1 + & 3x_2 + & 3x_3 + & 4x_4 & \leq 420 \\
 & \multicolumn{5}{l}{x_1 \geq 0,\ \ x_2 \geq 0,\ \ x_3 \geq 0,\ \ x_4 \geq 0}
\end{array}$$

Introduction des variables d'écart :

$$\begin{array}{lrrrrrrrl}
\max & 19x_1 + & 13x_2 + & 12x_3 + & 17x_4 & & & & \\
s.c. & 3x_1 + & 2x_2 + & x_3 + & 2x_4 & + x_5 & & & = 255 \\
 & x_1 + & x_2 + & x_3 + & x_4 & & + x_6 & & = 117 \\
 & 4x_1 + & 3x_2 + & 3x_3 + & 4x_4 & & & + x_7 & = 420 \\
 & \multicolumn{8}{l}{x_1,\ x_2,\ x_3,\ x_4,\ x_5,\ x_6,\ x_7 \geq 0}
\end{array}$$

Relation tableau à l'itération n – tableau d'origine (2/6)

Tableau d'origine :

$$\begin{array}{rrrrrl}
3x_1 + & 2x_2 + & x_3 + & 2x_4 + & \boxed{x_5 \quad\quad} & = 255 \\
x_1 + & x_2 + & x_3 + & x_4 & + x_6 & = 117 \\
4x_1 + & 3x_2 + & 3x_3 + & 4x_4 & + x_7 & = 420 \\
\hline
-z + 19x_1 & + 13x_2 & + 12x_3 & + 17x_4 & & = 0
\end{array}$$

Première itération : faire entrer x_1 et sortir x_5 :

$$\begin{array}{rrrrrrl}
 & \boxed{x_1} + \frac{2}{3}x_2 + & \frac{1}{3}x_3 + & \frac{2}{3}x_4 + & \frac{1}{3}x_5 & & = 85 \\
 & \frac{1}{3}x_2 + & \frac{2}{3}x_3 + & \frac{1}{3}x_4 - & \frac{1}{3}x_5 + & x_6 & = 32 \\
 & \frac{1}{3}x_2 + & \frac{5}{3}x_3 + & \frac{4}{3}x_4 - & \frac{4}{3}x_5 & + x_7 & = 80 \\
\hline
-z & + \frac{1}{3}x_2 + & \frac{17}{3}x_3 + & \frac{13}{3}x_4 - & \frac{19}{3}x_5 & & = -1615
\end{array}$$

Relation tableau à l'itération n – tableau d'origine (3/6)

base $(\hat{x}_1, \hat{x}_6, \hat{x}_7) = (85, 32, 80) \Longrightarrow \hat{x}_2, \hat{x}_3, \hat{x}_4, \hat{x}_5 = 0$

opérations algébriques $\Longrightarrow$ toute solution réalisable d'un tableau est aussi solution réalisable des tableaux précédents

Tableau d'origine :

$$\underbrace{\begin{array}{lll} 3x_1 & & \\ x_1 + x_6 & & \\ 4x_1 & & + x_7 \end{array}}_{B} \begin{array}{l} = \\ = \\ = \end{array} \underbrace{\begin{array}{l} 255 \\ 117 \\ 420 \end{array}}_{b}$$

Première itération :

$$\underbrace{\begin{array}{lll} x_1 & & \\ & x_6 & \\ & & x_7 \end{array}}_{B^{-1}B} \begin{array}{l} = \\ = \\ = \end{array} \underbrace{\begin{array}{l} 85 \\ 32 \\ 80 \end{array}}_{B^{-1}b}$$

$\Longrightarrow$ les tableaux du simplexe s'expriment en fonction de B^{-1}

Relation tableau à l'itération n – tableau d'origine (4/6)

- problème à résoudre :
  $\max c^T x$
  s.c. $\begin{cases} Ax = b \\ x \geq 0 \end{cases}$
- base réalisable $\Longrightarrow$ $x = \begin{bmatrix} x_B \\ x_N \end{bmatrix}$:

  $x_B = (x_1, x_6, x_7)$ $\quad x_N = (x_2, x_3, x_4, x_5)$
- $A = [\, B \;\; N \,]$:

$$\begin{array}{ccccccc} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \end{array}$$
$$\begin{bmatrix} 3 & 2 & 1 & 2 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 & 0 \\ 4 & 3 & 3 & 4 & 0 & 0 & 1 \end{bmatrix} = \Bigg[\underbrace{\begin{matrix} 3 & 0 & 0 \\ 1 & 1 & 0 \\ 4 & 0 & 1 \end{matrix}}_{B} \quad \underbrace{\begin{matrix} 2 & 1 & 2 & 1 \\ 1 & 1 & 1 & 0 \\ 3 & 3 & 4 & 0 \end{matrix}}_{N} \Bigg]$$

- $Ax = Bx_B + Nx_N$
- $Ax = b \Longrightarrow x_B = B^{-1}b - B^{-1}Nx_N$

Relation tableau à l'itération n – tableau d'origine (5/6)

- $x = \begin{bmatrix} x_B \\ x_N \end{bmatrix} \Longrightarrow c^T = [c_B^T \; c_N^T]$

  $z = c^T x = 19x_1 + 13x_2 + 12x_3 + 17x_4$

  $\Longrightarrow c^T = [19 \quad 13 \quad 12 \quad 17 \quad 0 \quad 0 \quad 0]$

  $\Longrightarrow c_B^T = [19 \quad 0 \quad 0]$ et $c_N^T = [13 \quad 12 \quad 17 \quad 0]$
- $c^T x = c_B^T x_B + c_N^T x_N$
- $x_B = B^{-1}b - B^{-1}Nx_N \Longrightarrow \boxed{c_B^T x_B = c_B^T B^{-1}b - c_B^T B^{-1}Nx_N}$
- $z = c_B^T x_B + c_N^T x_N \Longrightarrow z = c_B^T B^{-1}b - c_B^T B^{-1}Nx_N + c_N^T x_N$

  $\Longrightarrow \boxed{z = c_B^T B^{-1}b + (c_N^T - c_B^T B^{-1}N)x_N}$

Relation tableau à l'itération n – tableau d'origine (6/6)

Définition d'un dictionnaire

- B = base, N = hors base
- $x_B = B^{-1}b - B^{-1}Nx_N$
- $z = c_B^T B^{-1}b + (c_N^T - c_B^T B^{-1}N)x_N$

$\Longrightarrow$ méthode révisée du simplexe

Méthode révisée du simplexe (1/8)

À chaque itération de l'algorithme :

1. choisir une variable entrante
2. choisir une variable sortante $\Longrightarrow (x_B, x_N)$
3. faire une mise à jour de la solution réalisable : $\widehat{x_B}$

Exemple :

$$\begin{array}{rcl} 3x_1 + 2x_2 + x_3 + 2x_4 + x_5 & & = 255 \\ x_1 + x_2 + x_3 + x_4 & + x_6 & = 117 \\ 4x_1 + 3x_2 + 3x_3 + 4x_4 & + x_7 & = 420 \\ \hline -z + 19x_1 + 13x_2 + 12x_3 + 17x_4 & & = 0 \end{array}$$

$$\Longrightarrow B = \begin{bmatrix} 3 & 0 & 0 \\ 1 & 1 & 0 \\ 4 & 0 & 1 \end{bmatrix} \qquad \widehat{x_B} = B^{-1}b = \begin{bmatrix} \widehat{x_1} \\ \widehat{x_6} \\ \widehat{x_7} \end{bmatrix} = \begin{bmatrix} 85 \\ 32 \\ 80 \end{bmatrix}$$

Méthode révisée du simplexe (2/8)

prochaine variable entrante : une variable dont le coefficient dans z est positif $\Longrightarrow$ calculer z

calcul de $z = c_B^T B^{-1} b + (c_N^T - c_B^T B^{-1} N) x_N$:

1. calculer $y^T = c_B^T B^{-1} \Longrightarrow$ résoudre le système $y^T B = c_B^T$:

$$[y_1 \ \ y_2 \ \ y_3] \cdot \begin{bmatrix} 3 & 0 & 0 \\ 1 & 1 & 0 \\ 4 & 0 & 1 \end{bmatrix} = c_B^T = [19 \ \ 0 \ \ 0] \Longrightarrow y^T = [\tfrac{19}{3} \ \ 0 \ \ 0]$$

2. calculer $h^T = y^T N$:

$$h^T = y^T N = [\tfrac{19}{3} \ \ 0 \ \ 0] \begin{bmatrix} 2 & 1 & 2 & 1 \\ 1 & 1 & 1 & 0 \\ 3 & 3 & 4 & 0 \end{bmatrix} = [\tfrac{38}{3} \ \ \tfrac{19}{3} \ \ \tfrac{38}{3} \ \ \tfrac{19}{3}]$$

3. calculer $\widehat{c}_N^T = c_N^T - c_B^T B^{-1} N = c_N^T - h^T$:

$$\widehat{c}_N^T = [13 \ \ 12 \ \ 17 \ \ 0] - [\tfrac{38}{3} \ \ \tfrac{19}{3} \ \ \tfrac{38}{3} \ \ \tfrac{19}{3}] = [\tfrac{1}{3} \ \ \tfrac{17}{3} \ \ \tfrac{13}{3} \ \ \tfrac{-19}{3}]$$

Méthode révisée du simplexe (3/8)

$$\Longrightarrow \widehat{c}_N^T = [\tfrac{1}{3} \ \ \tfrac{17}{3} \ \ \tfrac{13}{3} \ \ \tfrac{-19}{3}]$$

Après la première itération du simplexe :

$$\begin{array}{lcccccccl}
\boxed{x_1} & + \frac{2}{3}x_2 + & \frac{1}{3}x_3 + & \frac{2}{3}x_4 + & \frac{1}{3}x_5 & & & = & 85 \\
& \frac{1}{3}x_2 + & \frac{2}{3}x_3 + & \frac{1}{3}x_4 - & \frac{1}{3}x_5 + & x_6 & & = & 32 \\
& \frac{1}{3}x_2 + & \frac{5}{3}x_3 + & \frac{4}{3}x_4 - & \frac{4}{3}x_5 & & + x_7 & = & 80 \\
\hline
-z & + \frac{1}{3}x_2 + & \frac{17}{3}x_3 + & \frac{13}{3}x_4 - & \frac{19}{3}x_5 & & & = & -1615
\end{array}$$

$\widehat{c}_N^T = c_N^T - c_B^T B^{-1} N$ = coeffs hors base de la fonction objectif pour la base (x_1, x_6, x_7)

Méthode révisée du simplexe (4/8)

$$\widehat{c}_N^T = [\tfrac{1}{3} \ \ \tfrac{17}{3} \ \ \tfrac{13}{3} \ \ \tfrac{-19}{3}]$$

choix de la variable entrante : n'importe quelle variable de coeff positif dans $\widehat{c}_N^T$

$$\widehat{c}_N^T = c_N^T - c_B^T B^{-1} N = c_N^T - y^T N$$
$$= [13 \ \ 12 \ \ 17 \ \ 0] - [\tfrac{19}{3} \ \ 0 \ \ 0] \begin{bmatrix} 2 & 1 & 2 & 1 \\ 1 & 1 & 1 & 0 \\ 3 & 3 & 4 & 0 \end{bmatrix}$$
$$= [\tfrac{1}{3} \ \ \tfrac{17}{3} \ \ \tfrac{13}{3} \ \ \tfrac{-19}{3}]$$

$\Longrightarrow$ on n'est pas obligé de calculer tout $\widehat{c}_N^T$:

e.g., calculer $\widehat{c}_N^T$ colonne par colonne et s'arrêter quand on a un nombre positif

Méthode révisée du simplexe (5/8)

$$\widehat{c}_N^T = \begin{bmatrix} \frac{1}{3} & \frac{17}{3} & \frac{13}{3} & \frac{-19}{3} \end{bmatrix} \qquad \widehat{x_B} = B^{-1}b = \begin{bmatrix} \widehat{x}_1 \\ \widehat{x}_6 \\ \widehat{x}_7 \end{bmatrix} = \begin{bmatrix} 85 \\ 32 \\ 80 \end{bmatrix}$$

$$\begin{array}{rrrrrrrr}
 & 3x_1 + & 2x_2 + & x_3 + & 2x_4 + x_5 & & & = 255 \\
 & x_1 + & x_2 + & x_3 + & x_4 & + x_6 & & = 117 \\
 & 4x_1 + & 3x_2 + & 3x_3 + & 4x_4 & & + x_7 & = 420 \\
\hline
-z + & 19x_1 + & 13x_2 + & 12x_3 + & 17x_4 & & & = 0
\end{array}$$

choix de la variable entrante : x_3

$\Longrightarrow$ augmenter la valeur de x_3 tout en assurant que les valeurs de x_B restent positives

$Bx_B + Nx_N = b \Longrightarrow x_B = B^{-1}b - B^{-1}Nx_N$

soit a la colonne de N correspondant à $x_3 \Longrightarrow x_B = B^{-1}b - B^{-1}ax_3$

Méthode révisée du simplexe (6/8)

soit a la colonne de N correspondant à $x_3 \Longrightarrow x_B = B^{-1}b - B^{-1}ax_3$

calcul de $d = B^{-1}a$: résoudre $Bd = a$:

$$\begin{bmatrix} 3 & 0 & 0 \\ 1 & 1 & 0 \\ 4 & 0 & 1 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix} \Longrightarrow d = \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{5}{3} \end{bmatrix}$$

$$\Longrightarrow \begin{bmatrix} x_1 \\ x_6 \\ x_7 \end{bmatrix} = \begin{bmatrix} 85 \\ 32 \\ 80 \end{bmatrix} - x_3 \times \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{5}{3} \end{bmatrix} \geq \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$\Longrightarrow$ valeur max de x_3 = 48 (ligne de x_6) $\Longrightarrow$ x_6 sort de la base

Méthode révisée du simplexe (7/8)

$$d = \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{5}{3} \end{bmatrix}$$

Après la première itération du simplexe :

$$\begin{array}{r|c|rrrr|cc|cr}
 & x_1 & + \frac{2}{3}x_2 + & \frac{1}{3}x_3 + & \frac{2}{3}x_4 + & \frac{1}{3}x_5 & & & = & 85 \\
 & & \frac{1}{3}x_2 + & \frac{2}{3}x_3 + & \frac{1}{3}x_4 - & \frac{1}{3}x_5 + & x_6 & & = & 32 \\
 & & \frac{1}{3}x_2 + & \frac{5}{3}x_3 + & \frac{4}{3}x_4 - & \frac{4}{3}x_5 & & + x_7 & = & 80 \\
\hline
-z & & + \frac{1}{3}x_2 + & \frac{17}{3}x_3 + & \frac{13}{3}x_4 - & \frac{19}{3}x_5 & & & = & -1615
\end{array}$$

$\Longrightarrow$ d = colonne de x_3 après la première itération du simplexe

Méthode révisée du simplexe (8/8)

Après la première itération du simplexe :

$$
\begin{array}{rcrcrcrcrcrcrcr}
 & x_1 & + \frac{1}{2}x_2 & & & + \frac{1}{2}x_4 & + \frac{1}{2}x_5 & - \frac{1}{2}x_6 & & = & 69 \\
 & & \frac{1}{2}x_2 & + & x_3 & + \frac{1}{2}x_4 & - \frac{1}{2}x_5 & + \frac{3}{2}x_6 & & = & 48 \\
 & & -\frac{1}{2}x_2 & & & + \frac{1}{2}x_4 & - \frac{1}{2}x_5 & - \frac{5}{2}x_6 & + x_7 & = & 0 \\
\hline
-z & & -\frac{5}{2}x_2 & & & + \frac{3}{2}x_4 & - \frac{7}{2}x_5 & - \frac{17}{2}x_6 & & = & -1887
\end{array}
$$

$$
d = \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{5}{3} \end{bmatrix} \Longrightarrow \widehat{x_B} - 48 \times d = \begin{bmatrix} 85 \\ 32 \\ 80 \end{bmatrix} - 48 \times \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{5}{3} \end{bmatrix} = \begin{bmatrix} 69 \\ 0 \\ 0 \end{bmatrix}
$$

$\widehat{x_B} - 48d \Longrightarrow$ valeurs des variables en base $\Longrightarrow$ nouvel $\widehat{x_B}$

Synthèse sur la méthode révisée du simplexe

Itération de l'algorithme révisé du simplexe

B = base courante $\quad \widehat{x_B}$ = valeur de la solution courante

1. calculer y^T tel que $y^T B = c_B^T$ ($\Longrightarrow y^T = c_B^T B^{-1}$)
2. choisir une colonne entrante : n'importe quelle colonne a de N telle que $c_a^T > y^T a$, où c_a^T = coeff de la colonne a dans c^T
3. calculer d tel que $Bd = a$ ($\Longrightarrow d \approx$ valeur de la colonne a après itération du simplexe)
4. trouver le plus grand nombre t tel que $\widehat{x_B} - td \geq 0$
   - si $t = +\infty$: problème non borné
   - sinon : au moins 1 ligne = 0 $\Longrightarrow$ variable sortant de la base
5. remplacer $\widehat{x_B}$ par $\widehat{x_B} - td$, puis la ligne correspondant à la variable sortante par t
6. remplacer dans B la colonne de la variable sortante par celle de la variable entrante

Calcul efficace de B^{-1} (1/4)

Efficacité de l'algo révisé : calcul de $y^T B = c_B^T$ et de $Bd = a$

- B_k : base après k itérations
- B_k ne diffère de B_{k-1} que par la colonne a rentrant à l'itération k
- supp que la colonne de a dans B_k soit la $p$ème
- a est la colonne qui rentre $\Longrightarrow$ à la $k$ème itération, $B_{k-1}d = a$

calcul de B_k

- E_k = matrice unité dont la $p$ème colonne est remplacée par d
- $B_k = B_{k-1}E_k$

$$
\begin{bmatrix} 3 & 2 & 4 \\ 2 & 1 & 5 \\ 4 & 2 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 & 3 & 0 \\ 0 & 1 & 0 \\ 0 & 4 & 1 \end{bmatrix}
$$

Calcul efficace de B^{-1} (2/4)

- $B_0 = I$
- $B_1 = E_1$
- $B_2 = E_1 E_2$
- $B_3 = E_1 E_2 E_3$
- ....

- $y^T B_k = c_B^T \Longrightarrow y^T E_1 E_2 E_3 \dots E_k = c_B^T$
- $(\dots(((y^T E_1)E_2)E_3)\dots E_k) = c_B^T$
- calcul de y :
  1. $y_k^T E_k = c_B^T$
  2. $y_{k-1}^T E_{k-1} = y_k^T$
  3. $y_{k-2}^T E_{k-2} = y_{k-1}^T$
  4. ....
  k. $y^T E_1 = y_2^T$

Calcul efficace de B^{-1} (3/4)

résolution pratique de $y_k^T E_k = c_B^T$:

$$[y_k^1 \quad y_k^2 \quad y_k^3 \quad y_k^4] \cdot \begin{bmatrix} 1 & 3 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 4 & 1 & 0 \\ 0 & 5 & 0 & 1 \end{bmatrix} = [3 \quad 32 \quad 4 \quad 1]$$

$$\Longrightarrow \begin{bmatrix} y_k^1 \\ y_k^2 \\ y_k^3 \\ y_k^4 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 4 \\ 1 \end{bmatrix}$$

$\Longrightarrow$ seul y_k^2 nécessite un calcul : $\begin{cases} s-1 & \text{additions} \\ s-1 & \text{multiplications} \\ 1 & \text{division} \end{cases}$

Calcul efficace de B^{-1} (4/4)

En pratique :

- lorsque $B_0 \neq I$: factorisations triangulaires de B_0
- en général : factorisation de B_k à l'aide des E_k plus rapide que le calcul de B_k^{-1}
- si ça n'est plus le cas (trop d'itérations) : refactoriser B_k comme si c'était un nouveau B_0

$\Longrightarrow$ méthode révisée plus rapide que la méthode standard sur de grosses instances

Cours 5 : Phase I – Phase 2

Christophe Gonzales

LIP6 – Université Paris 6, France

Problème d'initialisation de l'algo du simplexe

Problème de départ :

$$\begin{array}{lrcrcr}
\max & -2x_1 & - & x_2 & & \\
s.c. & -2x_1 & - & 3x_2 & \leq & -19 \\
 & 3x_1 & + & 4x_2 & \leq & 32 \\
 & \multicolumn{5}{l}{x_1, x_2 \geq 0}
\end{array}$$

Introduction des variables d'écart :

$$\begin{array}{lrcrcl}
\max & -2x_1 & - & x_2 & & \\
s.c. & -2x_1 & - & 3x_2 & + & \boxed{\begin{array}{ll} x_3 & \\ & + x_4 \end{array}} \begin{array}{l} = -19 \\ = \ \ \ 32 \end{array} \\
 & 3x_1 & + & 4x_2 & & \\
 & \multicolumn{5}{l}{x_1, x_2, x_3, x_4 \geq 0}
\end{array}$$

Base évidente : x_3, x_4 non réalisable ! ! ! ! ! ! ! !

Comment initialiser l'algo du simplexe ? (1/4)

Problème avec variables d'écart :

$$\begin{array}{lrcrcl}
\max & -2x_1 & - & x_2 & & \\
s.c. & -2x_1 & - & 3x_2 & + & \boxed{\begin{array}{ll} x_3 & \\ & + x_4 \end{array}} \begin{array}{l} = -19 \\ = \ \ \ 32 \end{array} \\
 & 3x_1 & + & 4x_2 & & \\
 & \multicolumn{5}{l}{x_1, x_2, x_3, x_4 \geq 0}
\end{array}$$

Introduction de variables artificielles :

$$\begin{array}{lrcrcrl}
\max & -2x_1 & - & x_2 & & & \\
s.c. & -2x_1 & - & 3x_2 & + & x_3 & \boxed{\begin{array}{rr} & -x_5 \\ x_4 & \end{array}} \begin{array}{l} = -19 \\ = \ \ \ 32 \end{array} \\
 & 3x_1 & + & 4x_2 & & + & \\
 & \multicolumn{6}{l}{x_1, x_2, x_3, x_4, x_5 \geq 0}
\end{array}$$

$\Longrightarrow$ nouvelle base réalisable évidente : x_4, x_5

Simplexe $\Longrightarrow$ si $x_5 = 0$ alors optimum du problème de départ

Comment initialiser l'algo du simplexe ? (2/4)

$$\begin{array}{ll} \max & -2x_1 - x_2 \\ s.c. & -2x_1 - 3x_2 + x_3 - x_5 = -19 \\ & 3x_1 + 4x_2 + x_4 = 32 \\ & x_1, x_2, x_3, x_4, x_5 \geq 0 \end{array}$$

Assurer que $x_5 = 0$:

$$\begin{array}{ll} \min & x_5 \\ s.c. & -2x_1 - 3x_2 + x_3 - x_5 = -19 \\ & 3x_1 + 4x_2 + x_4 = 32 \\ & x_1, x_2, x_3, x_4, x_5 \geq 0 \end{array}$$

Problème d'origine réalisable ssi $\min x_5 \Longrightarrow x_5 = 0$

Phase I : résolution du problème $\min x_5$

Comment initialiser l'algo du simplexe ? (3/4)

Résolution du problème :

$$\begin{array}{ll} \min & x_5 \\ s.c. & -2x_1 - 3x_2 + x_3 - x_5 = -19 \\ & 3x_1 + 4x_2 + x_4 = 32 \\ & x_1, x_2, x_3, x_4, x_5 \geq 0 \end{array}$$

Expression en fonction des variables hors base :

$$\begin{array}{ll} \max & 2x_1 + 3x_2 - x_3 - 19 \\ s.c. & -2x_1 - 3x_2 + x_3 - x_5 = -19 \\ & 3x_1 + 4x_2 + x_4 = 32 \\ & x_1, x_2, x_3, x_4, x_5 \geq 0 \end{array}$$

Résolution : faire entrer x_2 et sortir x_5 :

$$\begin{array}{ll} \max & - x_5 \\ s.c. & \frac{2}{3}x_1 + x_2 - \frac{1}{3}x_3 + \frac{1}{3}x_5 = \frac{19}{3} \\ & \frac{1}{3}x_1 + \frac{4}{3}x_3 + x_4 - \frac{4}{3}x_5 = \frac{20}{3} \\ & x_1, x_2, x_3, x_4, x_5 \geq 0 \end{array}$$

Variables en base : x_2, x_4
$\Longrightarrow x_5 = 0$
$(x_1 = 0, x_2 = \frac{19}{3}, x_3 = 0, x_4 = \frac{20}{3})$ = solution réalisable du problème d'origine

Comment initialiser l'algo du simplexe ? (4/4)

Dernière itération du simplexe :

$$\begin{array}{ll} \max & - x_5 \\ s.c. & \frac{2}{3}x_1 + x_2 - \frac{1}{3}x_3 + \frac{1}{3}x_5 = \frac{19}{3} \\ & \frac{1}{3}x_1 + \frac{4}{3}x_3 + x_4 - \frac{4}{3}x_5 = \frac{20}{3} \\ & x_1, x_2, x_3, x_4, x_5 \geq 0 \end{array}$$

⚠ $x_5 = 0$
$\Longrightarrow$ on peut supprimer x_5 des équations

Retour sur le problème d'origine :

$$\begin{array}{ll} \max & -2x_1 - x_2 \\ s.c. & \frac{2}{3}x_1 + x_2 - \frac{1}{3}x_3 = \frac{19}{3} \\ & \frac{1}{3}x_1 + \frac{4}{3}x_3 + x_4 = \frac{20}{3} \\ & x_1, x_2, x_3, x_4 \geq 0 \end{array}$$

Algo du simplexe sur ce problème : phase II

Phase I – phase 2 (1/5)

Problème d'origine (après introduction des variables d'écart) :

$$\max \sum_{j=1}^{n} c_j x_j$$

$$\text{s.c.} \left\{ \begin{array}{rl} \sum_{j=1}^{n} a_{ij} x_j = b_i & (i = 1, 2, \ldots, m) \\ x_j \geq 0 & (j = 1, 2, \ldots, n) \end{array} \right.$$

Introduction des variables artificielles :

$x_{n+1} \geq 0, \ldots, x_{n+m} \geq 0$ telles que :

$$\sum_{j=1}^{n} a_{ij} x_j + w_i x_{n+i} = b_i, \text{ où } w_i = \left\{ \begin{array}{r} 1 \text{ si } b_i \geq 0 \\ -1 \text{ si } b_i < 0 \end{array} \right.$$

⚠ si $b_i \geq 0$: variables d'écart $\Longrightarrow$ variables artificielles inutiles

Phase I – phase 2 (2/5)

Nouveau simplexe :

$$\left\{ \begin{array}{rl} \sum_{j=1}^{n} a_{ij} x_j + w_i x_{n+i} = b_i & (i = 1, 2, \ldots, m) \\ x_j \geq 0 & (j = 1, 2, \ldots, n+m) \end{array} \right.$$

Solution réalisable :

$$x_j = \left\{ \begin{array}{ll} 0 & \text{si } j \leq n \\ b_i / w_i & \text{si } j > n \end{array} \right.$$

Phase I – phase 2 (3/5)

Détermination d'une solution réalisable du problème d'origine :

$$\min \sum_{i=1}^{m} x_{n+i}$$

$$\text{s.c.} \left\{ \begin{array}{rl} \sum_{j=1}^{n} a_{ij} x_j + w_i x_{n+i} = b_i & (i = 1, 2, \ldots, m) \\ x_j \geq 0 & (j = 1, 2, \ldots, n+m) \end{array} \right.$$

Proposition

Le problème d'origine a une solution réalisable si et seulement si le problème ci-dessus a une solution dont la valeur (fonction objectif) vaut 0

Phase I – phase 2 (4/5)

Début de la phase 2 :

si $\min \sum_{i=1}^{m} x_{n+i} > 0$ alors problème d'origine non réalisable

sinon tableau simplexe :

$\Longrightarrow$ solution réalisable $(x_1^*, \ldots, x_{n+m}^*)$ telle que $x_{n+i}^* = 0 \; \forall i = 1, \ldots, m$

Problème d'origine équivalent à :

$$\max \sum_{j=1}^{n} c_j x_j$$

$$\text{s.c.} \begin{cases} \displaystyle\sum_{j=1}^{n} a_{ij} x_j + w_i x_{n+i} = b_i & (i = 1, 2, \ldots, m) \\ x_j \geq 0 & (j = 1, 2, \ldots, n) \\ x_{n+i} = 0 & (i = 1, 2, \ldots, m) \end{cases}$$

Phase I – phase 2 (5/5)

Résolution du problème d'origine :

1. supprimer toutes les variables artificielles hors base

⚠ il peut rester des variables artificielles en base (présence de contraintes redondantes)

2. résoudre avec l'algo du simplexe et variable artificielle sort de la base $\Longrightarrow$ la supprimer du problème à résoudre
$\Longrightarrow$ élimination progressive des variables artificielles

Variation de la phase I

$$\min \sum_{i=1}^{m} x_{n+i}$$

$$\text{s.c.} \begin{cases} \displaystyle\sum_{j=1}^{n} a_{ij} x_j + w_i x_{n+i} = b_i & (i = 1, 2, \ldots, m) \\ x_j \geq 0 & (j = 1, 2, \ldots, n+m) \end{cases}$$

$\Longrightarrow$ ne tient pas compte de la fonction objectif

$\Longrightarrow$ risque d'obtenir une solution réalisable très éloignée de l'optimum du problème d'origine

la méthode «big M»

1. choisir un M très grand
2. résoudre $\max \sum_{j=1}^{n} c_j x_j - M \sum_{i=1}^{m} x_{n+i}$ au lieu de $\min \sum_{i=1}^{m} x_{n+i}$

Cours 6 : dualité

Christophe Gonzales

LIP6 – Université Paris 6, France

En route vers la dualité (1/5)

$$\begin{array}{llllll} \max & 4x_1 + & x_2 + 5x_3 + 3x_4 & & \\ s.c. & x_1 - & x_2 - & x_3 + 3x_4 & \leq & 1 \\ & 5x_1 + & x_2 + 3x_3 + 8x_4 & & \leq & 55 \\ & -x_1 + & 2x_2 + 3x_3 - 5x_4 & & \leq & 3 \\ & x_1, x_2, x_3, x_4 \geq 0 \end{array}$$

Algo du simplexe $\Longrightarrow$ borne inférieure de la fonction objectif

Et si on voulait une borne supérieure ?

2ème contrainte $\times \frac{5}{3}$: $\frac{25}{3}x_1 + \frac{5}{3}x_2 + 5x_3 + \frac{40}{3}x_4 \leq \frac{275}{3}$

or $z \leq \frac{25}{3}x_1 + \frac{5}{3}x_2 + 5x_3 + \frac{40}{3}x_4 \Longrightarrow \boxed{z \leq \frac{275}{3}}$

En route vers la dualité (2/5)

$$\begin{array}{llllll} \max & 4x_1 + & x_2 + 5x_3 + 3x_4 & & \\ s.c. & x_1 - & x_2 - & x_3 + 3x_4 & \leq & 1 \\ & 5x_1 + & x_2 + 3x_3 + 8x_4 & & \leq & 55 \\ & -x_1 + & 2x_2 + 3x_3 - 5x_4 & & \leq & 3 \\ & x_1, x_2, x_3, x_4 \geq 0 \end{array}$$

Somme des 2ème et 3ème contraintes :
$4x_1 + 3x_2 + 6x_3 + 3x_4 \leq 58$

$\Longrightarrow \boxed{z \leq 58}$

principe valable pour toute combinaison linéaire à coeffs ≥ 0

En route vers la dualité (3/5)

Principe du dual

- faire une combinaison linéaire des contraintes : $\sum_{i=1}^{m} y_i \times$ $i$ème contrainte, avec $y_i \geq 0$
- z inférieur à la combinaison linéaire $\Longrightarrow z \leq \sum_{i=1}^{m} y_i b_i$

En route vers la dualité (4/5)

$$\begin{array}{ll}
\max & 4x_1 + x_2 + 5x_3 + 3x_4 \\
s.c. & x_1 - x_2 - x_3 + 3x_4 \leq 1 \\
 & 5x_1 + x_2 + 3x_3 + 8x_4 \leq 55 \\
 & -x_1 + 2x_2 + 3x_3 - 5x_4 \leq 3 \\
 & x_1, x_2, x_3, x_4 \geq 0
\end{array}$$

$$\begin{array}{l}
y_1 \times (x_1 - x_2 - x_3 + 3x_4 \leq 1) \\
y_2 \times (5x_1 + x_2 + 3x_3 + 8x_4 \leq 55) \\
y_3 \times (-x_1 + 2x_2 + 3x_3 - 5x_4 \leq 3) \\
\hline
(y_1 + 5y_2 - y_3) \times x_1 + \\
(-y_1 + y_2 + 2y_3) \times x_2 + \\
(-y_1 + 3y_2 + 3y_3) \times x_3 + \\
(3y_1 + 8y_2 - 5y_3) \times x_4 \leq (y_1 + 55y_2 + 3y_3)
\end{array}$$

En route vers la dualité (5/5)

$$\begin{array}{l}
(y_1 + 5y_2 - y_3) \times x_1 + \\
(-y_1 + y_2 + 2y_3) \times x_2 + \\
(-y_1 + 3y_2 + 3y_3) \times x_3 + \\
(3y_1 + 8y_2 - 5y_3) \times x_4 \leq (y_1 + 55y_2 + 3y_3)
\end{array}$$

Or fonction objectif = $4x_1 + 1x_2 + 5x_3 + 3x_4$

$$\Longrightarrow \begin{array}{l}
(y_1 + 5y_2 - y_3) \geq 4 \\
(-y_1 + y_2 + 2y_3) \geq 1 \\
(-y_1 + 3y_2 + 3y_3) \geq 5 \\
(3y_1 + 8y_2 - 5y_3) \geq 3
\end{array}$$

$$\Longrightarrow \boxed{z \leq (y_1 + 55y_2 + 3y_3)}$$

meilleure borne $\Longrightarrow$ $\boxed{\min y_1 + 55y_2 + 3y_3}$

Problème dual

Définition du dual

- problème d'origine : le primal :

$$\max \sum_{j=1}^{n} c_j x_j$$

$$s.c. \begin{cases} \displaystyle\sum_{j=1}^{n} a_{ij} x_j \leq b_i & (i = 1, 2, \ldots, m) \\ x_j \geq 0 & (j = 1, 2, \ldots, n) \end{cases}$$

- le problème dual :

$$\min \sum_{i=1}^{m} b_i y_i$$

$$s.c. \begin{cases} \displaystyle\sum_{i=1}^{m} a_{ij} y_i \geq c_j & (j = 1, 2, \ldots, n) \\ y_i \geq 0 & (j = 1, 2, \ldots, m) \end{cases}$$

Comparaison primal – dual (1/2)

$$\min \sum_{i=1}^{m} b_i y_i$$

$$s.c. \begin{cases} \displaystyle\sum_{i=1}^{m} a_{ij} y_i \geq c_j & (j = 1, 2, \ldots, n) \\ y_i \geq 0 & (j = 1, 2, \ldots, m) \end{cases}$$

- $(x_1, \ldots, x_n)$ solution du primal
- $(y_1, \ldots, y_m)$ solution du dual

$$\sum_{j=1}^{n} c_j x_j \leq \sum_{j=1}^{n} \left(\sum_{i=1}^{m} a_{ij} y_i \right) x_j \leq \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} x_j \right) y_i \leq \sum_{i=1}^{m} b_i y_i$$

$$\sum_{j=1}^{n} c_j x_j \leq \sum_{i=1}^{m} b_i y_i$$

Comparaison primal – dual (2/2)

$$\sum_{j=1}^{n} c_j x_j \leq \sum_{i=1}^{m} b_i y_i$$

- $(x_1^*, \ldots, x_n^*)$ solution du primal
- $(y_1^*, \ldots, y_m^*)$ solution du dual

alors $\displaystyle\sum_{j=1}^{n} c_j x_j^* = \sum_{i=1}^{m} b_i y_i^* \Longrightarrow (x_1^*, \ldots, x_n^*)$ et $(y_1^*, \ldots, y_m^*)$ optimaux

Démonstration :

transparent précédent : $\forall\ (x_1, \ldots, x_n),\ \displaystyle\sum_{j=1}^{n} c_j x_j \leq \sum_{i=1}^{m} b_i y_i^* = \sum_{j=1}^{n} c_j x_j^*$

transparent précédent : $\forall\ (y_1, \ldots, y_m),\ \displaystyle\sum_{i=1}^{m} b_i y_i \geq \sum_{j=1}^{n} c_j x_j^* = \sum_{i=1}^{m} b_i y_i^*$

Théorème de la dualité

Théorème de D. Gale, H.W. Kuhn & A.W. Tucker (1951)

Théorème de la dualité

- Si le primal a une solution optimale $(x_1^*, \ldots, x_n^*)$
- Alors le dual a une solution optimale $(y_1^*, \ldots, y_m^*)$ telle que :

$$\sum_{j=1}^{n} c_j x_j^* = \sum_{i=1}^{m} b_i y_i^*$$

Démonstration du théorème de la dualité (1/7)

Démonstration :

supposons que $(x_1^*, \ldots, x_n^*)$ solution optimale du primal

transparents précédents :

$\exists\ (y_1^*, \ldots, y_m^*)$ tel que $\sum_{j=1}^{n} c_j x_j^* = \sum_{i=1}^{m} b_i y_i^* \Longrightarrow (y_1^*, \ldots, y_m^*)$ optimal

$\Longrightarrow$ il suffit de montrer qu'il existe une solution du dual telle que :

$$\sum_{j=1}^{n} c_j x_j^* = \sum_{i=1}^{m} b_i y_i^*$$

Démonstration du théorème de la dualité (2/7)

Problème d'origine :

$$\max \sum_{j=1}^{n} c_j x_j$$

$$s.c. \begin{cases} \sum_{j=1}^{n} a_{ij} x_j \leq b_i & (i = 1, 2, \ldots, m) \\ x_j \geq 0 & (j = 1, 2, \ldots, n) \end{cases}$$

Introduction des variables d'écart :

$$x_{n+i} = b_i - \sum_{j=1}^{n} a_{ij} x_j \qquad (i = 1, 2, \ldots, m)$$

À l'optimum du primal :

$$z = z^* + \sum_{k=1}^{n+m} \widehat{c}_k x_k, \qquad \text{avec les } \widehat{c}_k \leq 0$$

Démonstration du théorème de la dualité (3/7)

Définition des y_i^* : $y_i^* = -\widehat{c_{n+i}}$

⚠ les y_i^* sont bien ≥ 0

y_i^* = -coeff dans z de la variable d'écart de la $i$ème contrainte

Reste de la démo : montrer que $(y_1^*, \ldots, y_m^*)$ est réalisable

À l'optimum du primal :

$$z = z^* + \sum_{k=1}^{n+m} \widehat{c}_k x_k = z^* + \sum_{k=1}^{n} \widehat{c}_k x_k + \sum_{k=n+1}^{n+m} \widehat{c}_k x_k$$

$$= z^* + \sum_{k=1}^{n} \widehat{c}_k x_k - \sum_{i=1}^{m} y_i^* x_{n+i}$$

Démonstration du théorème de la dualité (4/7)

Variables d'écart : $x_{n+i} = b_i - \sum_{j=1}^{n} a_{ij} x_j \qquad (i = 1, 2, \ldots, m)$

$$\Longrightarrow z = z^* + \sum_{k=1}^{n} \widehat{c}_k x_k - \sum_{i=1}^{m} y_i^* x_{n+i}$$

$$= z^* + \sum_{k=1}^{n} \widehat{c}_k x_k - \sum_{i=1}^{m} y_i^* \left(b_i - \sum_{j=1}^{n} a_{ij} x_j \right)$$

$$= \left(z^* - \sum_{i=1}^{m} b_i y_i^* \right) + \sum_{j=1}^{n} \left(\widehat{c}_j + \sum_{i=1}^{m} a_{ij} y_i^* \right) x_j$$

Démonstration du théorème de la dualité (5/7)

À l'origine $z = \sum_{j=1}^{n} c_j x_j$

algo du simplexe : opérations algébriques

$\Longrightarrow \forall$ tableaux, $z = \sum_{j=1}^{n} c_j x_j$

$\Longrightarrow$ d'après le transparent précédent :

$$z = \sum_{j=1}^{n} c_j x_j = \left(z^* - \sum_{i=1}^{m} b_i y_i^* \right) + \sum_{j=1}^{n} \left(\widehat{c}_j + \sum_{i=1}^{m} a_{ij} y_i^* \right) x_j$$

Démonstration du théorème de la dualité (6/7)

$$\sum_{j=1}^{n} c_j x_j = \left(z^* - \sum_{i=1}^{m} b_i y_i^* \right) + \sum_{j=1}^{n} \left(\widehat{c_j} + \sum_{i=1}^{m} a_{ij} y_i^* \right) x_j$$

équation valable pour tout $(x_1, \dots, x_n)$

$$(x_1, \dots, x_n) = (0, \dots, 0) \Longrightarrow z^* = \sum_{i=1}^{m} b_i y_i^*$$

$$(x_j = 1, x_k = 0 \ \forall\ k \neq j) \Longrightarrow c_j = \widehat{c_j} + \sum_{i=1}^{m} a_{ij} y_i^* \qquad (j = 1, \dots, n)$$

Or condition d'arrêt du simplexe : $\widehat{c_k} \leq 0$

$$\Longrightarrow \sum_{i=1}^{m} a_{ij} y_i^* \geq c_j \qquad (j = 1, \dots, n)$$

Démonstration du théorème de la dualité (7/7)

Conclusion :

- Si $y_i^* = -\widehat{c_{n+i}}$ alors :
- $y_i^* \geq 0$
- $\sum_{i=1}^{m} a_{ij} y_i^* \geq c_j \qquad (j = 1, \dots, n)$

$\Longrightarrow (y_1^*, \dots, y_m^*)$ est une solution réalisable du dual

- $z^* = \sum_{j=1}^{n} c_j x_j^* = \sum_{i=1}^{m} b_i y_i^*$

$\Longrightarrow (y_1^*, \dots, y_m^*)$ solution optimale

CQFD

Application (1/2)

Problème d'origine :

$$\begin{array}{lrcl}
\max & 5x_1 + 4x_2 + 3x_3 & & \\
s.c. & 2x_1 + 3x_2 + \ \ x_3 & \leq & 5 \\
 & 4x_1 + \ \ x_2 + 2x_3 & \leq & 11 \\
 & 3x_1 + 4x_2 + 2x_3 & \leq & 8 \\
 & x_1 \geq 0, \ \ x_2 \geq 0, \ \ x_3 & \geq & 0
\end{array}$$

Après introduction des variables d'écart :

$$\begin{array}{lcl}
x_4 & = & 5 - 2x_1 - 3x_2 - \ \ x_3 \\
x_5 & = & 11 - 4x_1 - \ \ x_2 - 2x_3 \\
x_6 & = & 8 - 3x_1 - 4x_2 - 2x_3 \\
z & = & \qquad 5x_1 + 4x_2 + 3x_3
\end{array}$$

$x_1 \geq 0, \ \ x_2 \geq 0, \ \ x_3 \geq 0, \ \ x_4 \geq 0, \ \ x_5 \geq 0, \ \ x_6 \geq 0$

Application (2/2)

Dictionnaire à l'optimum :

$$\begin{aligned} x_1 &= 2 - 2x_2 - 2x_4 + x_6 \\ x_5 &= 1 + 5x_2 + 2x_4 \\ x_3 &= 1 + x_2 + 3x_4 - 2x_6 \\ z &= 13 - 3x_2 - x_4 - x_6 \end{aligned}$$

- solution du primal : $(x_1, x_2, x_3, x_4, x_5, x_6) = (2, 0, 1, 0, 1, 0)$
- solution du dual : $(y_1, y_2, y_3) = (-\widehat{c_4}, -\widehat{c_5}, -\widehat{c_6}) = (1, 0, 1)$

⚠ dans le simplexe sous forme tabulaire, on a $-z$
$\Longrightarrow$ ne pas multiplier les coeffs de la dernière ligne par -1

Relations entre primal et dual (1/4)

Expression d'un problème dual

$$\min \sum_{i=1}^{m} b_i y_i$$
$$\text{s.c.} \begin{cases} \sum_{i=1}^{m} a_{ij} y_i \geq c_j & (j = 1, 2, \ldots, n) \\ y_i \geq 0 & (j = 1, 2, \ldots, m) \end{cases}$$

Or $\min f = -\max -f$

$$-\max \sum_{i=1}^{m} (-b_i) y_i$$
$$\text{s.c.} \begin{cases} \sum_{i=1}^{m} (-a_{ij}) y_i \leq -c_j & (j = 1, 2, \ldots, n) \\ y_i \geq 0 & (j = 1, 2, \ldots, m) \end{cases}$$

le dual est un nouveau primal !

Relations entre primal et dual (2/4)

Dual

$$-\max \sum_{i=1}^{m} (-b_i) y_i$$
$$\text{s.c.} \begin{cases} \sum_{i=1}^{m} (-a_{ij}) y_i \leq -c_j & (j = 1, 2, \ldots, n) \\ y_i \geq 0 & (j = 1, 2, \ldots, m) \end{cases}$$

Dual du dual

le dual du dual = primal

$$\max \sum_{j=1}^{n} c_j x_j$$
$$\text{s.c.} \begin{cases} \sum_{j=1}^{n} a_{ij} x_j \leq b_i & (i = 1, 2, \ldots, m) \\ x_j \geq 0 & (j = 1, 2, \ldots, n) \end{cases}$$

Relations entre primal et dual (3/4)

Relations primal – dual

- le dual du dual = le primal
- primal a un optimum $\Longleftrightarrow$ dual a un optimum
- $\sum_{j=1}^{n} c_j x_j \leq \sum_{i=1}^{m} b_i y_i$

  $\Longrightarrow$ primal non borné $\Longrightarrow$ dual non réalisable
- dual non borné $\Longrightarrow$ primal non réalisable

⚠ primal et dual peuvent être tous deux non réalisables :

$$\begin{array}{lrcl} \max & 2x_1 - x_2 & & \\ s.c. & x_1 - x_2 & \leq & 1 \\ & -x_1 + x_2 & \leq & -2 \\ & x_1, x_2 \geq 0 & & \end{array}$$

Relations entre primal et dual (4/4)

		Dual		
		$\exists$ optimum	non réalisable	non borné
Primal	$\exists$ optimum	✔	✖	✖
	non réalisable	✖	✔	✔
	non borné	✖	✔	✖

$\Longrightarrow$ si le primal et le dual ont des solutions réalisables alors ils ont un optimum

Conséquence pratique

Il peut être avantageux d'appliquer l'algo du simplexe sur le dual plutôt que sur le primal

tableau du dual à l'optimum $\Longrightarrow$ solution optimale du primal

Exemple : problème primal à 9 variables et 99 contraintes

$\Longrightarrow$ 100 lignes dans le primal et 10 lignes dans le dual

nb d'itérations du simplexe $\approx$ proportionnel au nb de lignes

$\Longrightarrow$ moins d'itérations dans le dual

algo révisé du simplexe $\Longrightarrow$ itérations pas plus coûteuses avec le dual

Théorème de complémentarité

Théorème de complémentarité

- $(x_1^*, \dots, x_n^*)$: solution réalisable du primal
- $(y_1^*, \dots, y_m^*)$: solution réalisable du dual

Alors une condition nécessaire et suffisante pour que $(x_1^*, \dots, x_n^*)$ et $(y_1^*, \dots, y_m^*)$ soient optimaux simultanément :

- $\sum_{i=1}^{m} a_{ij} y_i^* = c_j$ ou $x_j^* = 0$ (ou les 2) $\quad \forall\, j = 1, 2, \dots, n$

et

- $\sum_{j=1}^{n} a_{ij} x_j^* = b_i$ ou $y_i^* = 0$ (ou les 2) $\quad \forall\, i = 1, 2, \dots, m$

Démonstration du théorème de complémentarité (1/3)

Démonstration :

$$\text{dual} \Longrightarrow \sum_{i=1}^{m} a_{ij} y_i^* \geq c_j$$

$$\Longrightarrow \left(\sum_{i=1}^{m} a_{ij} y_i^* \right) x_j^* \geq c_j x_j^*$$

$$\text{primal} \Longrightarrow \sum_{j=1}^{n} a_{ij} x_j^* \leq b_i \Longrightarrow \left(\sum_{j=1}^{n} a_{ij} x_j^* \right) y_i^* \leq b_i y_i^*$$

$$\Longrightarrow \sum_{j=1}^{n} c_j x_j^* \leq \sum_{j=1}^{n} \left(\sum_{i=1}^{m} a_{ij} y_i^* \right) x_j^* = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} x_j^* \right) y_i^* \leq \sum_{i=1}^{m} b_i y_i^*$$

Démonstration du théorème de complémentarité (2/3)

$$\sum_{j=1}^{n} c_j x_j^* \leq \sum_{j=1}^{n} \left(\sum_{i=1}^{m} a_{ij} y_i^* \right) x_j^* = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} x_j^* \right) y_i^* \leq \sum_{i=1}^{m} b_i y_i^*$$

Si $(x_1^*, \dots, x_n^*)$ et $(y_1^*, \dots, y_m^*)$ optimaux

alors théorème de la dualité $\Longrightarrow \sum_{j=1}^{n} c_j x_j^* = \sum_{i=1}^{m} b_i y_i^*$

$$\Longrightarrow \sum_{j=1}^{n} c_j x_j^* = \sum_{j=1}^{n} \left(\sum_{i=1}^{m} a_{ij} y_i^* \right) x_j^*$$

$$\text{dual} \Longrightarrow \sum_{i=1}^{m} a_{ij} y_i^* \geq c_j \Longrightarrow \sum_{i=1}^{m} a_{ij} y_i^* x_j^* = c_j x_j^*$$

$$\Longrightarrow x_j^* = 0 \quad \text{ou} \quad c_j = \sum_{i=1}^{m} a_{ij} y_i^*$$

démo similaire pour $\sum_{j=1}^{n} a_{ij} x_j^* = b_i \quad$ ou $\quad y_i^* = 0$

Démonstration du théorème de complémentarité (3/3)

Réciproque

$$\text{Si } \left[x_j^* = 0 \quad \text{ou} \quad c_j = \sum_{i=1}^{m} a_{ij} y_i^* \right] \quad \text{et} \quad \left[y_i^* = 0 \quad \text{ou} \quad \sum_{j=1}^{n} a_{ij} x_j^* = b_i \right]$$

$$\text{alors } \sum_{j=1}^{n} c_j x_j^* = \sum_{j=1}^{n} \left(\sum_{i=1}^{m} a_{ij} y_i^* \right) x_j^* = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} x_j^* \right) y_i^* = \sum_{i=1}^{m} b_i y_i^*$$

théorème de la dualité $\Longrightarrow (x_1^*, \ldots, x_n^*)$ et $(y_1^*, \ldots, y_m^*)$ optimaux

CQFD

Rappel : Théorème de complémentarité

Théorème de complémentarité

- $(x_1^*, \ldots, x_n^*)$: solution réalisable du primal
- $(y_1^*, \ldots, y_m^*)$: solution réalisable du dual

Alors une condition nécessaire et suffisante pour que $(x_1^*, \ldots, x_n^*)$ et $(y_1^*, \ldots, y_m^*)$ soient optimaux simultanément :

- $\sum_{i=1}^{m} a_{ij} y_i^* = c_j$ ou $x_j^* = 0$ (ou les 2) $\quad \forall\, j = 1, 2, \ldots, n$

et

- $\sum_{j=1}^{n} a_{ij} x_j^* = b_i$ ou $y_i^* = 0$ (ou les 2) $\quad \forall\, i = 1, 2, \ldots, m$

$\Longrightarrow$ si $x_j^* > 0$ alors $\sum_{i=1}^{m} a_{ij} y_i^* = c_j$ $\qquad$ si $y_i^* > 0$ alors $\sum_{j=1}^{n} a_{ij} x_j^* = b_i$

Complémentarité : corollaire

Corollaire du théorème de complémentarité

- $(x_1^*, \ldots, x_n^*)$: solution réalisable du primal
- $(x_1^*, \ldots, x_n^*)$ optimal si et seulement si $\exists\, (y_1^*, \ldots, y_m^*)$ tel que :

$$\sum_{i=1}^{m} a_{ij} y_i^* = c_j \quad \text{dès que} \quad x_j^* > 0$$

$$y_i^* = 0 \quad \text{dès que} \quad \sum_{j=1}^{n} a_{ij} x_j^* < b_i$$

et tel que :

$$\sum_{i=1}^{m} a_{ij} y_i^* \geq c_j \qquad \forall\, j = 1, 2, \ldots, n$$

$$y_i^* \geq 0 \qquad \forall\, i = 1, 2, \ldots, m$$

Interprétation économique des variables duales (1/5)

$$\max \sum_{j=1}^{n} c_j x_j$$
$$\text{s.c.} \begin{cases} \sum_{j=1}^{n} a_{ij} x_j \le b_i & (i = 1, 2, \ldots, m) \\ x_j \ge 0 & (j = 1, 2, \ldots, n) \end{cases}$$

Problème : maximisation du profit d'une fabrique de meubles

$\Longrightarrow$ utilise des matières premières et en produit des meubles

$\Longrightarrow$ x_j = nombre de meubles d'un certain type (chaises, bureaux) fabriqués
c_j = prix en € d'une unité de produit
a_{ij} = quantité de la $i$ème matière première nécessaire à la construction d'une unité du $j$ème type de meuble
b_i = quantité de la $i$ème matière première disponible

Interprétation économique des variables duales (2/5)

x_j = nombre de meubles d'un certain type (chaises, bureaux) fabriqués
c_j = prix en € d'une unité de produit
a_{ij} = quantité de la $i$ème matière première nécessaire à la construction d'une unité du $j$ème type de meuble
b_i = quantité de la $i$ème matière première disponible

---

variable	unité
x_j	unité de produit j
c_j	€ par unité de produit j
a_{ij}	unité de ressource i par unité de produit j
b_i	unité de ressource i

Interprétation économique des variables duales (3/5)

$$\min \sum_{i=1}^{m} b_i y_i$$
$$\text{s.c.} \begin{cases} \sum_{i=1}^{m} a_{ij} y_i \ge c_j & (j = 1, 2, \ldots, n) \\ y_i \ge 0 & (j = 1, 2, \ldots, m) \end{cases}$$

variable	unité
c_j	€ par unité de produit j
a_{ij}	unité de ressource i par unité de produit j

$\sum_{i=1}^{m} a_{ij} y_i \ge c_j \Longrightarrow$ (unité de ressource i par unité de produit j)$\times$ unité de y_i = € par unité de produit j

$\Longrightarrow$ y_i exprimé en € par unité de ressource i

Interprétation économique des variables duales (4/5)

Interprétation économique

y_i mesure l'apport d'une unité de ressource i au profit de l'entreprise

on augmente d'1 le nombre d'unités de ressource $i \Longrightarrow$ le profit augmente de y_i

$\Longrightarrow$ on est prêt à payer cette unité de ressource au maximum un prix de y_i

les y_i sont souvent appelés «*prix marginaux*»

Interprétation économique des variables duales (5/5)

Théorème

Si le primal a au moins une solution optimale non dégénérée, alors $\exists\, \epsilon > 0$ tel que si $|t_i| \leq \epsilon\ \forall i = 1, 2, \ldots, m$ alors :

$$\max \sum_{j=1}^{n} c_j x_j$$

$$s.c. \begin{cases} \displaystyle\sum_{j=1}^{n} a_{ij} x_j \leq b_i + t_i & (i = 1, 2, \ldots, m) \\ x_j \geq 0 & (j = 1, 2, \ldots, n) \end{cases}$$

a une solution optimale et dont la valeur est $z^* + \sum_{i=1}^{m} y_i^* t_i$

où $y_1^*, \ldots, y_m^*$ = solution optimale du dual

Cours 7 : applications théoriques de la programmation linéaire

Christophe Gonzales

LIP6 – Université Paris 6, France

Plan du cours

1. Théorème de l'alternative
2. Lemme de Farkas

Théorème de l'alternative (1/3)

Théorème de l'alternative

- A une matrice $m \times n$
- c un vecteur de taille n
- un et un seul des énoncés suivants est vrai :
  1. il existe v tel que $v^T A \leq c^T$
  2. il existe $x \geq 0$ tel que $Ax = 0$ et $c^T x < 0$

⚠ Rappel : les vecteurs sont tous par défaut en colonne

Théorème de l'alternative (2/3)

1. il existe v tel que $v^T A \leq c^T$
2. il existe $x \geq 0$ tel que $Ax = 0$ et $c^T x < 0$

Démonstration :

démo de 1 $\Longrightarrow$ non 2 :

supposons 1 alors le PL suivant a un optimum :

max $v^T.0$

$$s.c. \begin{cases} v^T A + d^T = c^T \\ d^T \geq 0 \end{cases}$$

$\Longrightarrow$ dual a un optimum :

min $c^T x$

$$s.c. \begin{cases} Ax = 0 \\ x \geq 0 \end{cases}$$

valeur du dual = valeur du primal $\Longrightarrow c^T x \not< 0 \Longrightarrow$ non 2

Théorème de l'alternative (3/3)

1. il existe v tel que $v^T A \leq c^T$
2. il existe $x \geq 0$ tel que $Ax = 0$ et $c^T x < 0$

Démo de non 2 $\Longrightarrow$ 1 :

On sait que $S = \{x : Ax = 0, \ \ x \geq 0\} \neq \emptyset$ car contient 0.

Non 2 $\Longrightarrow c^T x \geq 0$ pour tout $x \in S$

$\Longrightarrow$ min $c^T x$

$$s.c. \begin{cases} Ax = 0 \\ x \geq 0 \end{cases}$$

a une solution, de valeur 0 $\Longrightarrow$ dual a une solution :

max $v^T.0$

$$s.c. \begin{cases} v^T A + d^T = c^T \\ d^T \geq 0 \end{cases}$$

$\Longrightarrow$ il existe v tel que $v^T A \leq c^T \Longrightarrow$ 1 CQFD

Lemme de Farkas (1/5)

Lemme de Farkas

- A une matrice $m \times n$
- c un vecteur de taille n
- les deux énoncés suivants sont équivalents :
  1. $Ax \geq 0 \Longrightarrow c^T x \geq 0$
  2. il existe $v \geq 0$ tel que $v^T A = c^T$

Lemme de Farkas (2/5)

1. $Ax \geq 0 \Longrightarrow c^T x \geq 0$
2. il existe $v \geq 0$ tel que $v^T A = c^T$

Démonstration :

démo de 2 $\Longrightarrow$ 1 :

Soit $v \geq 0$ tel que $v^T A = c^T$. Supposons que $Ax \geq 0$

alors $v \geq 0$ et $Ax \geq 0 \Longrightarrow v^T Ax \geq 0$

Or $v^T A = c^T \Longrightarrow v^T Ax = c^T x \geq 0 \Longrightarrow$ 1

Lemme de Farkas (3/5)

1. $Ax \geq 0 \Longrightarrow c^T x \geq 0$
2. il existe $v \geq 0$ tel que $v^T A = c^T$

---

démo de 1 $\Longrightarrow$ 2 :

Soit le PL :

$\min c^T x$

$s.c.\ Ax \geq 0$

Pas de contraintes sur le signe de $x \Longrightarrow$ équivalent à :

$\min c^T x' - c^T x''$

$$s.c. \begin{cases} Ax' - Ax'' \geq 0 \\ x' \geq 0,\ x'' \geq 0 \end{cases}$$

Lemme de Farkas (4/5)

$\min c^T x' - c^T x''$

$$s.c. \begin{cases} Ax' - Ax'' \geq 0 \\ x' \geq 0,\ x'' \geq 0 \end{cases}$$

a pour dual :

$\max v^T.0$

$$s.c. \begin{cases} v^T A \geq c^T \\ v^T(-A) \geq -c^T \\ v \geq 0 \end{cases}$$

ce qui équivaut à :

$\max v^T.0$

$$s.c. \begin{cases} v^T A = c^T \\ v \geq 0 \end{cases}$$

Lemme de Farkas (5/5)

1. $Ax \geq 0 \Longrightarrow c^T x \geq 0$
2. il existe $v \geq 0$ tel que $v^T A = c^T$

PL1 : $\min c^T x$

$\quad$ *s.c.* $Ax \geq 0$

a donc pour dual :

PL2 : $\max v^T.0$

$$s.c. \begin{cases} v^T A = c^T \\ v \geq 0 \end{cases}$$

Or, d'après 1, PL1 minoré par 0, atteint pour $x = 0$

Théorème de la dualité : PL2 a une solution optimale = 0

$\Longrightarrow$ il existe v_* tel que $v_* \geq 0$ et $v_*^T A = c^T \Longrightarrow$ 2

CQFD

Cours 8 : Applications pratiques de la programmation linéaire

Christophe Gonzales

LIP6 – Université Paris 6, France

Plan du cours

1. Jeux à deux joueurs à somme nulle
2. Théorème du MINIMAX en stratégies pures
3. stratégies mixtes
4. Théorème du MINIMAX en stratégies mixtes

Jeux à deux joueurs

Théorie des jeux

- théorie des jeux = étude des situations (les *jeux*) où des agents (les *joueurs*) ont à choisir des stratégies
- stratégies $\Longrightarrow$ résultat (paiement, gain) pour chaque joueur
- les résultats dépendent des stratégies jouées par tous les joueurs

Jeu à deux joueurs

Un jeu dans lequel il n'y a que 2 agents

Exemple de jeu à deux joueurs

Exemple : le dilemme des prisonniers

- Deux criminels présumés : Bonnie et Clyde
- interrogés séparément par la police $\Longrightarrow$ 3 cas :

1. ils nient tous les deux $\Longrightarrow$ pas de preuve
   $\Longrightarrow$ faible peine (1 an)
2. ils avouent tous les deux
   $\Longrightarrow$ peine plus forte (8 ans)
3. l'un des deux avoue tandis que l'autre nie
   $\Longrightarrow$ peines = 0 an pour l'un et 10 ans pour l'autre

Problème : Que vont faire, que doivent faire, les prisonniers ?

Forme normale d'un jeu à deux joueurs

Définition de la forme normale

- jeu à deux joueurs $\Longrightarrow$ tableau de gains
- matrice de jeu = matrice des gains
- lignes = stratégies du 1er joueur
- colonnes = stratégies du 2ème joueur

Exemple : le dilemme des prisonniers

tous deux nient $\Longrightarrow$ peines d'1 an
tous deux avouent $\Longrightarrow$ peines de 8 ans
l'un avoue, l'autre nie $\Longrightarrow$ peines respectives = 0 an et 10 ans

Bonnie \ Clyde	Nier	Avouer
Nier	(-1,-1)	(-10,0)
Avouer	(0,-10)	(-8,-8)

Jeu à deux joueurs à somme nulle

Définition d'un jeu à deux joueurs à somme nulle

Jeu pour lequel la somme des paiements est toujours égale à 0 $\forall$ stratégies des joueurs

Forme normale d'un jeu à deux joueurs à somme nulle

- matrice des gains comme dans le jeu à 2 joueurs classique
- mais on n'écrit que le gain du 1er joueur (celui des lignes)
- gain du 2ème joueur = $-$gain du 1er joueur

Exemple de jeu à deux joueurs à somme nulle

- 3 boites : Noire, Rouge, Verte
- joueur II : répartit 2 pièces de 1 € entre les 3 boites
- joueur I : choisit 1 boite et gagne son contenu

Matrice de jeu :

joueur I \ joueur II	NN	RR	VV	NR	NV	RV
N	2	0	0	1	1	0
R	0	2	0	1	0	1
V	0	0	2	0	1	1

Jeu à deux joueurs à somme nulle : forme générale

joueur I \ joueur II	1	2	...	j_0	...	n
1	$g_{1,1}$	$g_{1,2}$	...	g_{1,j_0}	...	$g_{1,n}$
2	$g_{2,1}$	$g_{2,2}$	...	g_{2,j_0}	...	$g_{2,n}$
...	...	...	...	...	...	...
i_0	$g_{i_0,1}$	$g_{i_0,2}$	...	g_{i_0,j_0}	...	$g_{i_0,n}$
...	...	...	...	...	...	...
m	$g_{m,1}$	$g_{m,2}$	...	g_{m,j_0}	...	$g_{m,n}$

⚠ les gains $g_{i,j}$ dépendent des stratégies i et j des joueurs

- $g_{i,j}$ = gain du joueur I
- $-g_{i,j}$ = gain du joueur II

Stratégie des joueurs

Les joueurs choisissent leurs stratégies à l'insu l'un de l'autre

joueur I : stratégie $i_0 \Longrightarrow$ gain minimum $= \min_{j=1}^{n} g_{i_0,j}$

joueur prudent $\Longrightarrow$ rendre ce gain le plus élevé possible

critère MAXIMIN

stratégie du joueur I : stratégie i telle que $\max_{i=1}^{m} \min_{j=1}^{n} g_{i,j}$

second joueur : stratégie $j_0 \Longrightarrow$ perte maximale $= \max_{i=1}^{m} g_{i,j_0}$

joueur prudent $\Longrightarrow$ rendre cette perte la plus petite possible

critère MINIMAX

stratégie du joueur II : stratégie j telle que $\min_{j=1}^{n} \max_{i=1}^{m} g_{i,j}$

Théorème du minimax en stratégies pures (1/2)

Théorème du minimax en stratégies pures

$$\max_{i=1}^{m} \min_{j=1}^{n} g_{i,j} \leq \min_{j=1}^{n} \max_{i=1}^{m} g_{i,j}$$

Démonstration :

$\forall i, \forall j, \ \min_{j=1}^{n} g_{i,j} \leq g_{i,j}$

$\forall j, \ \max_{i=1}^{m} \min_{j=1}^{n} g_{i,j} \leq \max_{i=1}^{m} g_{i,j}$

membre de gauche = constante $\Longrightarrow \max_{i=1}^{m} \min_{j=1}^{n} g_{i,j} \leq \min_{j=1}^{n} \max_{i=1}^{m} g_{i,j}$

CQFD

Théorème du minimax en stratégies pures (2/2)

⚠ $\max_{i=1}^{m} \min_{j=1}^{n} g_{i,j} \neq \min_{j=1}^{n} \max_{i=1}^{m} g_{i,j}$:

- 3 boites : Noire, Rouge, Verte
- joueur II : répartit 2 pièces de 1 € entre les 3 boites
- joueur I : choisit 1 boite et gagne son contenu

joueur I \ joueur II	NN	RR	VV	NR	NV	RV	min
N	2	0	0	1	1	0	0
R	0	2	0	1	0	1	0
V	0	0	2	0	1	1	0
max	2	2	2	1	1	1	

Stratégies mixtes / stratégies pure (1/3)

les joueurs ne peuvent être certains d'obtenir plus que ce qu'ils peuvent s'assurer

VON NEUMANN : en moyenne peuvent-ils s'assurer plus ?

$\Longrightarrow$ concept de stratégie mixte

stratégie mixte

- joueur I : $p = \begin{bmatrix} p^1 \\ p^2 \\ \vdots \\ p^m \end{bmatrix}$ joueur II : $q = \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \end{bmatrix}$
- p = stratégie mixte du joueur I, p^i = proba de jouer la stratégie i
- q = stratégie mixte du joueur II, q_j = proba de jouer la stratégie j

Stratégies mixtes / stratégies pure (2/3)

stratégie pure du joueur I : p = vecteur unité, i.e., $p^i = \begin{cases} 1 & \text{si } i = i_0 \\ 0 & \text{si } i \neq i_0 \end{cases}$

stratégie pure du joueur II : q = vecteur unité, i.e., $q_j = \begin{cases} 1 & \text{si } j = j_0 \\ 0 & \text{si } j \neq j_0 \end{cases}$

Ensemble de stratégies mixtes

- $\mathcal{P} = \{$ensemble des stratégies mixtes du 1er joueur$\}$
$$= \left\{ p : p^i \geq 0 \text{ et } \sum_{i=1}^{m} p^i = 1 \right\}$$
- $\mathcal{Q} = \{$ensemble des stratégies mixtes du 2ème joueur$\}$
$$= \left\{ q : q_j \geq 0 \text{ et } \sum_{j=1}^{n} q_j = 1 \right\}$$

Stratégies mixtes / stratégies pure (3/3)

$Pr(\{i,j\})$ = proba que le joueur I joue i et le joueur II joue j

Les joueurs jouent indépendamment $\Longrightarrow$ $\boxed{Pr(\{i,j\}) = p^i q_j}$

espérance mathématique de gain du joueur I

- le joueur I joue la stratégie p
- le joueur II joue la stratégie q
- G = matrice du jeu

Espérance de gain = $\displaystyle\sum_{i=1}^{m}\sum_{j=1}^{n} p^i q_j g_{i,j} = p^T G q$

à partir de maintenant : les joueurs veulent optimiser leur espérance de gain

Minimax et maximin en stratégies mixtes

Minimax et maximin en stratégies mixtes

$$\text{MAXIMIN} = \max_{p \in \mathcal{P}} \min_{q \in \mathcal{Q}} p^T G q$$

$$\text{MINIMAX} = \min_{q \in \mathcal{Q}} \max_{p \in \mathcal{P}} p^T G q$$

- MAXIMIN = espérance de gain minimale du joueur I
- MINIMAX = espérance de perte maximale du joueur II

⚠ En stratégies pures, MAXIMIN $\leq$ MINIMAX

Théorème du Minimax en stratégies mixtes

Théorème du minimax (von Neumann)

Dans tout jeu à deux joueurs à somme nulle, le minimax et le maximin en stratégies mixtes sont égaux, i.e.,

$$\max_{p\in\mathcal{P}} \min_{q\in\mathcal{Q}} p^T Gq = \min_{q\in\mathcal{Q}} \max_{p\in\mathcal{P}} p^T Gq$$

Démonstration du théorème du minimax (1/6)

Démonstration :

- démonstration de $\max_{p\in\mathcal{P}} \min_{q\in\mathcal{Q}} p^T Gq \leq \min_{q\in\mathcal{Q}} \max_{p\in\mathcal{P}} p^T Gq$:

  $\Longrightarrow$ identique à la démo du théorème en stratégies pures

- G_i^T = ligne i de la matrice du jeu
- G^j = colonne j de la matrice du jeu
- Montrons que ce qu'un joueur peut s'assurer en espérance contre les stratégies pures de l'adversaire, il peut aussi l'assurer contre les stratégies mixtes, i.e.,

$$\max_{p\in\mathcal{P}} \min_{q\in\mathcal{Q}} p^T Gq = \max_{p\in\mathcal{P}} \min_{j=1}^{n} p^T G^j = \min_{q\in\mathcal{Q}} \max_{i=1}^{m} G_i^T q$$

Démonstration du théorème du minimax (2/6)

Soit p une stratégie mixte quelconque

Démonstration de $\min_{q\in\mathcal{Q}} p^T Gq = \min_{j=1}^{n} p^T G^j$:

stratégie pure = stratégie mixte particulière

$\Longrightarrow \min_{q\in\mathcal{Q}} p^T Gq \leq \min_{j=1}^{n} p^T G^j$

Réciproquement : si $\forall j$, $p^T G^j \geq \mu$ alors :

$$\forall q,\ p^T Gq = p^T \sum_{j=1}^{n} G^j q_j = \sum_{j=1}^{n} \left[p^T G^j\right] q_j \geq \sum_{j=1}^{n} \mu q_j = \mu$$

de même : $\max_{p\in\mathcal{P}} p^T Gq = \max_{i=1}^{m} G_i^T q$

Démonstration du théorème du minimax (3/6)

Résumé du transparent précédent :

Soit p une stratégie mixte quelconque
alors :

$$\min_{q\in\mathcal{Q}} p^T Gq = \min_{j=1}^{n} p^T G^j \text{ et } \max_{p\in\mathcal{P}} p^T Gq = \max_{i=1}^{m} G_i^T q$$

vrai quel que soit p donc :

$$\max_{p\in\mathcal{P}} \min_{q\in\mathcal{Q}} p^T Gq = \max_{p\in\mathcal{P}} \min_{j=1}^{n} p^T G^j$$

$$\min_{q\in\mathcal{Q}} \max_{p\in\mathcal{P}} p^T Gq = \min_{q\in\mathcal{Q}} \max_{i=1}^{m} G_i^T q$$

Démonstration du théorème du minimax (4/6)

Démonstration par l'absurde de MINIMAX $\leq$ MAXIMIN :

Supposons que MAXIMIN $<$ MINIMAX

Soit α tel que MAXIMIN $< \alpha <$ MINIMAX

Or MAXIMIN $= \max_{p\in\mathcal{P}} \min_{j=1}^{n} p^T G^j$ (cf. transparent précédent)

donc $\forall$ stratégie mixte p, $\min_{j=1}^{n} p^T G^j < \alpha$

donc $\nexists p$ telle que $\forall j$, $p^T G^j \geq \alpha$

donc le système $\begin{cases} p^T G^j \geq \alpha \ \forall j \\ p^T 1_m = 1 \\ p^i \geq 0 \forall i \end{cases}$ est incompatible

où 1_m = vecteur de taille m constitué uniquement de 1

Démonstration du théorème du minimax (5/6)

$$\begin{cases} p^T G^j \geq \alpha \ \forall j \\ p^T 1_m = 1 \\ p^i \geq 0 \forall i \end{cases} \iff \begin{cases} p^T G \geq \alpha_n \\ p^T 1_m \geq 1 \\ -p^T 1_m \geq -1 \\ p^T I_m \geq 0_m \end{cases} \iff \begin{cases} p^T(-G) \leq -\alpha_n \\ p^T(-1_m) \leq -1 \\ p^T 1_m \leq 1 \\ p^T(-I_m) \leq 0_m \end{cases}$$

où I_m = matrice identité et α_n = vecteur constitué de n α

Rappel : théorème de l'alternative

un et un seul des énoncés suivants est vrai :

1. il existe v tel que $v^T A \leq c^T$
2. il existe $x \geq 0$ tel que $Ax = 0$ et $c^T x < 0$

$A = [-G \quad -1_m \quad 1_m \quad -I_m]$ et $c^T = [-\alpha_n^T \quad -1 \quad 1 \quad 0_m^T]$

$\Longrightarrow \nexists p$ tel que $p^T A \leq c^T$

$\Longrightarrow \exists x \geq 0$ tel que $Ax = 0$ et $c^T x < 0$

Démonstration du théorème du minimax (6/6)

$A = [-G \quad -1_m \quad 1_m \quad -I_m]$ et $c^T = [-\alpha_n^T \quad -1 \quad 1 \quad 0_m^T]$

$$\exists x = \begin{bmatrix} y \\ z' \\ z'' \\ d \end{bmatrix} \geq \begin{bmatrix} 0_n \\ 0 \\ 0 \\ 0_m \end{bmatrix} \text{ tel que } Ax = 0 \text{ et } c^T x < 0$$

$$\Longrightarrow \text{système} \begin{cases} c^T x = -\alpha_n^T y - z' + z'' < 0 \\ Ax \ = -Gy - 1_m z' + 1_m z'' - I_m d = 0_m \\ y \geq 0_n, \ z' \geq 0, \ z'' \geq 0, \ d \geq 0_m \end{cases}$$

$$\Longrightarrow \text{système} \begin{cases} \alpha_n^T y + z' - z'' > 0 \\ Gy + 1_m z' - 1_m z'' + I_m d = 0_m \\ y \geq 0_n, \ z' \geq 0, \ z'' \geq 0, \ d \geq 0_m \end{cases}$$

$\Longrightarrow Gy \leq 1_m(z'' - z') \leq 1_m \alpha_n^T y \Longrightarrow \exists q$ t.q. $Gq \leq 1_m \alpha_n^T q = 1_m \alpha = \alpha_m$

$\Longrightarrow G_i q \leq \alpha \ \forall i \Longrightarrow$ contradiction de notre hypothèse de départ

Théorème du Minimax en stratégies mixtes

Théorème du minimax (von Neumann)

Dans tout jeu à deux joueurs à somme nulle, le minimax et le maximin en stratégies mixtes sont égaux, i.e.,

$$\max_{p \in \mathcal{P}} \min_{q \in \mathcal{Q}} p^T G q = \min_{q \in \mathcal{Q}} \max_{p \in \mathcal{P}} p^T G q$$

$\Longrightarrow$ théorème de dualité

Cours 9 : Optimisation non linéaire monodimensionnelle sans contraintes

Christophe Gonzales

LIP6 – Université Paris 6, France

Plan du cours

1. fonctions unimodales
2. méthode de la suite de Fibonacci
3. méthode dichotomique
4. méthode de Newton

Minimum local, global

f : fonction $[a, b] \subset \mathbb{R} \mapsto \mathbb{R}$

minimum global

- f passe par un minimum (global ou absolu) en $\hat{x}$ si :

$$x \in [a, b] \Longrightarrow f(x) \geq f(\widehat{x})$$

- minimum strict : $x \neq \widehat{x} \Longrightarrow f(x) > f(\widehat{x})$

minimum local

- f passe par un minimum local en $\hat{x}$ si :

$$\exists \text{ un voisinage } V \text{ de } \widehat{x} \text{ tel que } x \in V \Longrightarrow f(x) \geq f(\widehat{x})$$

- minimum local strict : $x \in V \backslash \{\widehat{x}\} \Longrightarrow f(x) > f(\widehat{x})$

⚠ Dans $\mathbb{R}$, voisinage V de $\widehat{x}$ = ensemble t.q. :
$V \supseteq$ intervalle ouvert $]\alpha, \beta[\supseteq \{\widehat{x}\}$

Fonctions unimodales

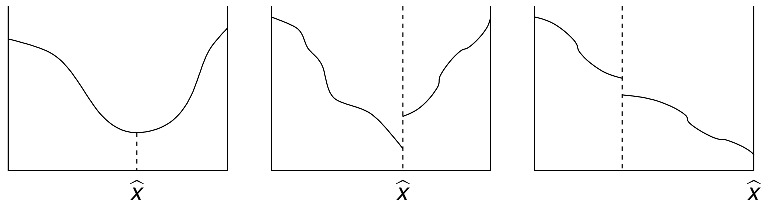
Définition d'une fonction unimodale

f est unimodale lorsqu'il existe $\widehat{x} \in [a, b]$ tel que :

$$x^1 < x^2 \leq \widehat{x} \implies f(x^1) > f(x^2)$$
$$\widehat{x} \leq x^1 < x^2 \implies f(x^1) < f(x^2)$$

$\implies$ une fonction unimodale passe par un minimum strict

⚠ unimodale $\not\Longrightarrow$ dérivable



Fonctions convexes et strictement convexes

Définition d'une fonction convexe

f est convexe si $\forall x, y \in [a, b]$, $\forall \lambda \in [0, 1]$:

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

Définition d'une fonction strictement convexe

f est strictement convexe si $\forall x, y \in [a, b]$, $\forall \lambda \in]0, 1[$:

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

strictement convexe $\implies$ unimodale

convexe $\not\Longrightarrow$ unimodale
unimodale $\not\Longrightarrow$ convexe

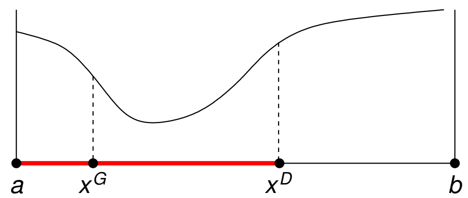
Unimodalité et recherche de minima

Lemme 1

- fonction unimodale
- minimum local $\implies$ minimum global

Lemme 2

- $f : [a, b] \to \mathbb{R}$ unimodale
- x^G, x^D tels que $a < x^G < x^D < b$ alors :
  - $f(x^G) < f(x^D) \implies \widehat{x} \in [a, x^D]$
  - $f(x^G) > f(x^D) \implies \widehat{x} \in [x^G, b]$
  - $f(x^G) = f(x^D) \implies \widehat{x} \in [x^G, x^D]$.



Application : méthode de la suite de Fibonacci (1/6)

Rappel : suite de Fibonacci

- $F_0 = F_1 = 1$
- $F_{k+2} = F_{k+1} + F_k$
- $\dfrac{F_{k+1}}{F_k} \longrightarrow \dfrac{1+\sqrt{5}}{2}$

méthode de la suite de Fibonacci : utiliser Fibonacci pour trouver rapidement le minimum d'une fonction unimodale sur un intervalle $[a, b]$

Application : méthode de la suite de Fibonacci (2/6)

méthode de la suite de Fibonacci

- on se donne N = nombre total de fois où l'on évaluera f en un point
- initialisation : $a_1 = a$, $b_1 = b$
- itérations :

$$x_k^G = a_k + \frac{F_{N-k}}{F_{N+2-k}}(b_k - a_k) \quad x_k^D = a_k + \frac{F_{N+1-k}}{F_{N+2-k}}(b_k - a_k)$$

  - $f(x_k^G) < f(x_k^D) \Longrightarrow a_{k+1} = a_k$ et $b_{k+1} = x_k^D$
  - $f(x_k^G) > f(x_k^D) \Longrightarrow a_{k+1} = x_k^G$ et $b_{k+1} = b_k$
  - $f(x_k^G) = f(x_k^D) \Longrightarrow a_{k+2} = x_k^G$ et $b_{k+2} = x_k^D$
- arrêt : longueur de l'intervalle après N évaluations de f :

$$\leq \prod_{k=1}^{N-1} \frac{F_{N+1-k}}{F_{N+2-k}}(b-a) = \frac{b-a}{F_{N+1}}$$

Application : méthode de la suite de Fibonacci (3/6)

calcul de x_{k+1}^D et x_{k+1}^G : cas $f(x_k^G) < f(x_k^D)$

- $a_{k+1} = a_k$ et $b_{k+1} = x_k^D$
- $$\begin{aligned} x_{k+1}^D &= a_{k+1} + \frac{F_{N+1-(k+1)}}{F_{N+2-(k+1)}}(b_{k+1} - a_{k+1}) \\ &= a_k + \frac{F_{N-k}}{F_{N+1-k}}(x_k^D - a_k) \\ &= a_k + \frac{F_{N-k}}{F_{N+1-k}} \times \frac{F_{N+1-k}}{F_{N+2-k}}(b_k - a_k) \\ &= a_k + \frac{F_{N-k}}{F_{N+2-k}}(b_k - a_k) \\ &= x_k^G \end{aligned}$$

$\Longrightarrow$ seul $f(x_{k+1}^G)$ doit être évalué à l'itération $k+1$

Application : méthode de la suite de Fibonacci (4/6)

taille de l'intervalle $[a_{k+1}, b_{k+1}]$: cas $f(x_k^G) < f(x_k^D)$

- $a_{k+1} = a_k$ et $b_{k+1} = x_k^D$
- $b_{k+1} - a_{k+1} = x_k^D - a_k = \dfrac{F_{N+1-k}}{F_{N+2-k}}(b_k - a_k)$

$\Longrightarrow \dfrac{b_{k+1} - a_{k+1}}{b_k - a_k} = \dfrac{F_{N+1-k}}{F_{N+2-k}}$

Par symétrie :

cas $f(x_k^G) > f(x_k^D)$

- seul $f(x_{k+1}^D)$ doit être évalué à l'itération $k+1$
- $\dfrac{b_{k+1} - a_{k+1}}{b_k - a_k} = \dfrac{F_{N+1-k}}{F_{N+2-k}}$

Application : méthode de la suite de Fibonacci (5/6)

taille de l'intervalle $[a_{k+2}, b_{k+2}]$: cas $f(x_k^G) = f(x_k^D)$

- $a_{k+2} = x_k^G$ et $b_{k+2} = x_k^D$
- $$\begin{aligned} \frac{b_{k+2} - a_{k+2}}{b_k - a_k} &= \frac{F_{N+1-k} - F_{N-k}}{F_{N+2-k}} = \frac{F_{N-1-k}}{F_{N+2-k}} \\ &= \frac{F_{N-1-k}}{F_{N-k}} \left[\frac{F_{N-k}}{F_{N+1-k}} \times \frac{F_{N+1-k}}{F_{N+2-k}} \right] \\ &\leq \frac{F_{N-k}}{F_{N+1-k}} \times \frac{F_{N+1-k}}{F_{N+2-k}} \end{aligned}$$

⚠ évaluation de f en x_{k+2}^G et x_{k+2}^D mais intervalle plus réduit que pour 2 itérations des cas $f(x_k^G) \neq f(x_k^D)$

Application : méthode de la suite de Fibonacci (6/6)

Résumé :

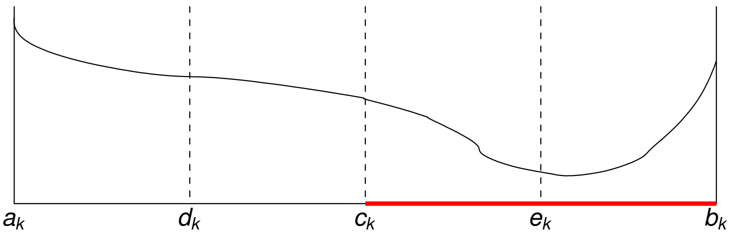
- $f(x_k^G) < f(x_k^D) \Longrightarrow$ seul $f(x_{k+1}^G)$ doit être évalué
$$\frac{b_{k+1} - a_{k+1}}{b_k - a_k} = \frac{F_{N+1-k}}{F_{N+2-k}}$$
- $f(x_k^G) > f(x_k^D) \Longrightarrow$ seul $f(x_{k+1}^D)$ doit être évalué
$$\frac{b_{k+1} - a_{k+1}}{b_k - a_k} = \frac{F_{N+1-k}}{F_{N+2-k}}$$
- $f(x_k^G) = f(x_k^D) \Longrightarrow f(x_{k+2}^G)$ et $f(x_{k+2}^D)$ doivent être évalués
$$\frac{b_{k+2} - a_{k+2}}{b_k - a_k} \leq \frac{F_{N-k}}{F_{N+1-k}} \times \frac{F_{N+1-k}}{F_{N+2-k}}$$

à l'arrêt : $b_N - a_N \leq \prod_{j=1}^{N-1} \dfrac{F_{N+1-j}}{F_{N+2-j}}(b-a) = \dfrac{b-a}{F_{N+1}}$

La méthode dichotomique

L'algorithme par dichotomie

- à l'itération k : intervalle $[a_k, b_k]$
- $d_k = \dfrac{3a_k + b_k}{4}$ $\quad c_k = \dfrac{a_k + b_k}{2}$ $\quad e_k = \dfrac{a_k + 3b_k}{4}$
- $f(c_k) > f(e_k) \Longrightarrow a_{k+1} = c_k$ et $b_{k+1} = b_k$
  $f(d_k) < f(c_k) \Longrightarrow a_{k+1} = a_k$ et $b_{k+1} = c_k$
  sinon $a_{k+1} = d_k$ et $b_{k+1} = e_k$
- arrêt : quand $b_k - a_k \leq \epsilon$



Fonctions dérivables

$f : [a, b] \subset \mathbb{R} \mapsto \mathbb{R}$ dérivable. f' = dérivée de f

Point stationnaire

$\widehat{x} \in]a, b[$: point stationnaire si $f'(\widehat{x}) = 0$

Lemme 3

- f dérivable
- $\widehat{x} \in]a, b[$
- $\widehat{x}$ minimum local de $f \Longrightarrow \widehat{x} =$ point stationnaire

Lemme 4

- f dérivable et convexe
- $\widehat{x} \in]a, b[$
- $\widehat{x}$ minimum local de $f \Longleftrightarrow \widehat{x} =$ point stationnaire

Dichotomie pour fonctions unimodales dérivables

Méthode dichotomique

- f unimodale, dérivable
- $a_1 = a$, $b_1 = b$
- $f'(a_k) < 0$ et $f'(b_k) > 0 \Longrightarrow \widehat{x} \in [a_k, b_k]$
- calcul de $f'\left(\dfrac{a_k + b_k}{2}\right)$
- $f'\left(\dfrac{a_k + b_k}{2}\right) > 0 \Longrightarrow \widehat{x} \in \left[a_k, \dfrac{a_k + b_k}{2}\right]$
- $f'\left(\dfrac{a_k + b_k}{2}\right) < 0 \Longrightarrow \widehat{x} \in \left[\dfrac{a_k + b_k}{2}, b_k\right]$
- $f'\left(\dfrac{a_k + b_k}{2}\right) = 0 \Longrightarrow \widehat{x} = \dfrac{a_k + b_k}{2}$

Méthode de Newton (1/3)

fonction de classe C^2

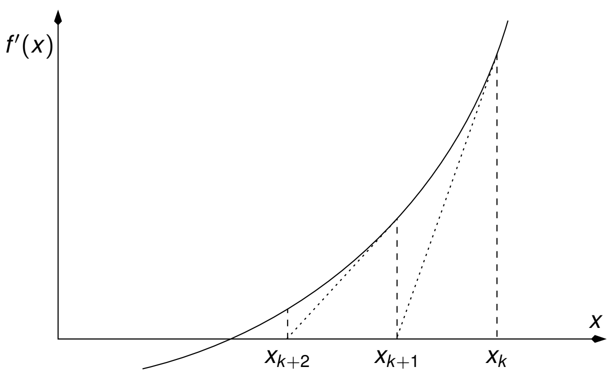
- $f : [a, b] \subset \mathbb{R} \mapsto \mathbb{R}$
- f : 2 fois dérivable
- f'' continue

Méthode de Newton

- *principe* : engendrer une suite de points (x^k) tendant vers un point stationnaire
- point stationnaire : $f'(\widehat{x}) = 0$
- itération k : f' est remplacée par sa linéarisée en x^k : $l(x) = f'(x^k) + [x - x^k]f''(x^k)$
- x^{k+1} déterminé par $l(x^{k+1}) = 0$:

$$\Longrightarrow x^{k+1} = x^k - \frac{f'(x^k)}{f''(x^k)}$$

Méthode de Newton (2/3)



Méthode de Newton (3/3)

Conditions suffisantes de convergence de la méthode

Conditions suffisantes de convergence à partir d'un point de départ x^0 quelconque :

- f de classe C^3 et $f'(a) \times f'(b) < 0$
- $f''(x) > 0 \ \forall x$ ($\Longrightarrow$ stricte convexité)
- $0 \leq 1 - \dfrac{d}{dx}\left[\dfrac{f'(x)}{f''(x)}\right] \leq q < 1 \ \forall x$

$\Longrightarrow$ taux de convergence quadratique :

$$\exists \beta > 0 \text{ tel que } \left\|x^{k+1} - \widehat{x}\right\| \leq \beta \left\|x^k - \widehat{x}\right\|^2$$

⚠ conditions très restrictives

en pratique : on applique Newton même si ces conditions ne sont pas satisfaites

Cours 10 : Optimisation non linéaire multidimensionnelle sans contraintes

Christophe Gonzales

LIP6 – Université Paris 6, France

Plan du cours

Optimisation d'une fonction différentiable à n variables

1. méthode du gradient
2. méthode du gradient conjugué (la semaine prochaine)

Fonction différentiable

$f : C \mapsto \mathbb{R}$: fonction définie dans un convexe ouvert C de $\mathbb{R}^n$:

$$f : x = \begin{bmatrix} x_1 \\ \ldots \\ x_j \\ \ldots \\ x_n \end{bmatrix} \longmapsto f(x) = f(x_1, ..., x_j, ..., x_n)$$

fonction différentiable

f est différentiable en $x \in C$ si $\exists$ dérivées partielles $\dfrac{\partial f(x)}{\partial x_j}$, $j = 1, ..., n$ et admet pour approximation du 1^{er} ordre la forme linéaire qu'elles définissent :

$$f(x+h) = [f(x_1+h_1, \ldots, x_n+h_n)] = f(x) + \sum_{j=1}^{n} \frac{\partial f(x)}{\partial x_j}.h_j + o(\|h\|)$$

Gradient (1/3)

Définition du gradient

$\vec{\nabla} f(x)$: gradient de f en x = le vecteur $\vec{\nabla} f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \cdots \\ \frac{\partial f(x)}{\partial x_j} \\ \cdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}$

$\Longrightarrow f(x+h) = f(x) + \vec{\nabla} f(x)^T.h + o(\|h\|)$

$\Longrightarrow$ La variation de f est du $2^{\text{ème}}$ ordre lorsque $\vec{\nabla} f(x)^T.h = 0$

$\vec{\nabla} f(x) \neq 0 \Longrightarrow \left\{ y : f(y) = f(x) + \vec{\nabla} f(x)^T.[y-x] \right\}$ = l'hyperplan tangent en x à l'hypersurface de niveau $\{z : f(z) = f(x)\}$

normale de l'hyperplan = $\vec{\nabla} f(x)$

Gradient (2/3)

normale de l'hyperplan = $\vec{\nabla} f(x)$

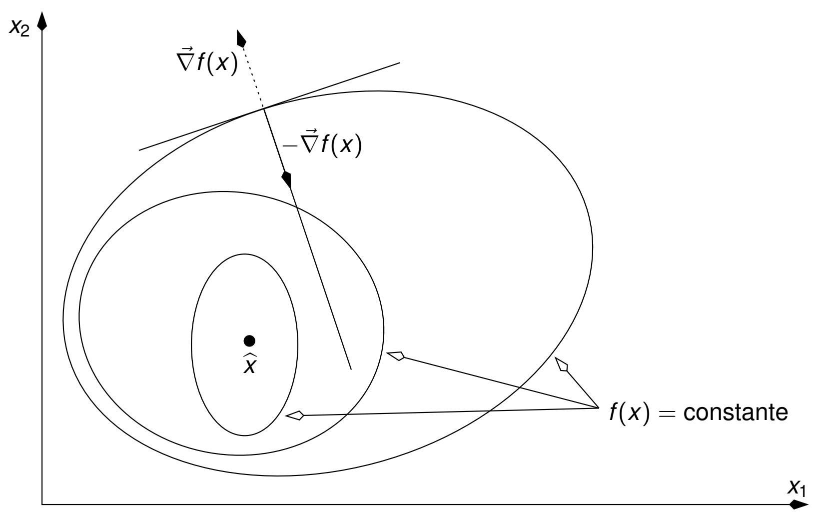
$$\begin{aligned} f(x+\lambda \vec{\nabla} f(x)) &= f(x) + \lambda \vec{\nabla} f(x)^T.\vec{\nabla} f(x) \\ &= f(x) + \lambda \left\| \vec{\nabla} f(x) \right\|^2 \\ &> f(x) \text{ pour } \lambda > 0 \end{aligned}$$

$\Longrightarrow$ normale dirigée du côté des points où f prend des valeurs plus élevées qu'en x

on cherche min $f \Longrightarrow \delta = -\vec{\nabla} f(x)$ = direction intéressante

$\delta = -\vec{\nabla} f(x)$ = direction de l'anti-gradient

Gradient (3/3)



Minima et points stationnaires

- minimum (global ou absolu) de f en $\widehat{x}$: $x \in C \Longrightarrow f(x) \geq f(\widehat{x})$
- minimum local de f en $\widehat{x}$: $\exists$ un voisinage V de $\widehat{x}$ tel que $x \in V \Longrightarrow f(x) \geq f(\widehat{x})$
- voisinage V de $\widehat{x}$ dans C = un sous-ensemble de C contenant une boule ouverte $\{y : \|y - x\| < \epsilon\}$
- $\widehat{x}$ = point stationnaire de f si $\vec{\nabla} f(\widehat{x}) = 0$

Lemme

- f différentiable
- minimum local de f en $\widehat{x} \Longrightarrow \widehat{x} =$ point stationnaire
- de plus, si f convexe :
  $\widehat{x} =$ point stationnaire $\Longrightarrow \widehat{x} =$ minimum local

Direction de l'anti-gradient (1/2)

La méthode du gradient s'appuie sur :

Proposition

La direction de l'anti-gradient est la direction de plus grande pente :

$$\max_{\{h : \|h\| = \|\vec{\nabla} f(x)\|\}} \lim_{\lambda \downarrow 0} \frac{f(x) - f(x + \lambda h)}{\lambda \|h\|}$$

est atteint pour $h = -\vec{\nabla} f(x)$.

Direction de l'anti-gradient (2/2)

Démonstration de la proposition :

$f(x) - f(x + \lambda h) = -\lambda \vec{\nabla} f(x)^T.h + o(\lambda \|h\|) \Longrightarrow$ pour $\|h\| = \left\|\vec{\nabla} f(x)\right\|$:

$$\frac{f(x) - f(x + \lambda h)}{\lambda \|h\|} = \frac{-\lambda \vec{\nabla} f(x)^T.h + o\left(\lambda \left\|\vec{\nabla} f(x)\right\|\right)}{\lambda \left\|\vec{\nabla} f(x)\right\|}$$

$$= \frac{-\vec{\nabla} f(x)^T.h}{\left\|\vec{\nabla} f(x)\right\|} + \frac{o(\lambda)}{\lambda}$$

$$\Longrightarrow \lim_{\lambda \downarrow 0} \frac{f(x) - f(x + \lambda h)}{\lambda \|h\|} = \frac{-\vec{\nabla} f(x)^T.h}{\left\|\vec{\nabla} f(x)\right\|}$$

où $-\vec{\nabla} f(x).h$ est maximum $\left(\text{sous } \|h\| = \left\|\vec{\nabla} f(x)\right\|\right)$ pour $h = -\vec{\nabla} f(x)$

Pas de la méthode du gradient

variations de f lorsque l'on part de x dans la direction de l'anti-gradient = celles de la fonction d'une seule variable $\lambda \geq 0$:

$$\varphi : \lambda \longmapsto \varphi(\lambda) = f(x - \lambda\vec{\nabla} f(x))$$

Or $\varphi'(\lambda) = -\sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(x - \lambda\vec{\nabla} f(x)).\vec{\nabla} f_j(x) = -\vec{\nabla} f(x - \lambda\vec{\nabla} f(x)).\vec{\nabla} f(x)$

en particulier : $\varphi'(0) = -\vec{\nabla} f(x)^T.\vec{\nabla} f(x) = -\left\|\vec{\nabla} f(x)\right\|^2 < 0$

$\Longrightarrow \varphi$ strictement décroissante au voisinage de $\lambda = 0$

> tant que $\varphi'(\lambda) > 0$, on continue à se déplacer et on s'arrête en : $\widetilde{x} = x - \widetilde{\lambda}\vec{\nabla} f(x)$, où $\widetilde{\lambda}$ est la plus petite valeur de λ solution de $\varphi'(\lambda) = 0$ (si elle existe)

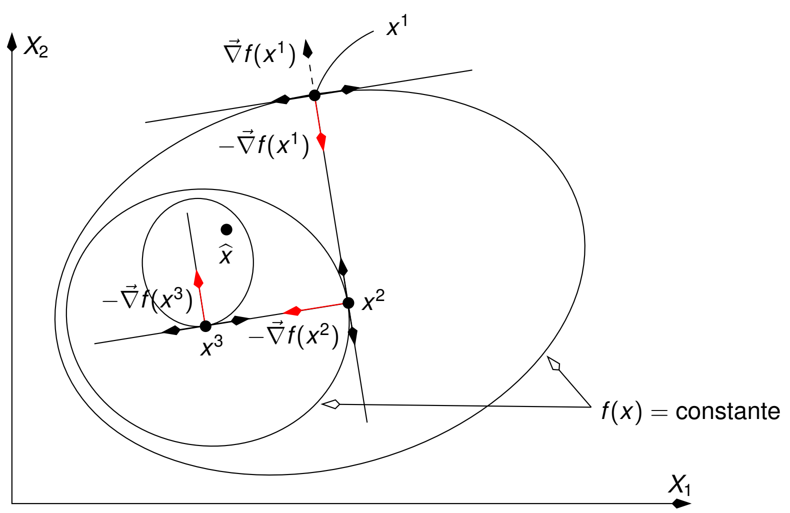
Méthode du gradient (1/2)

Méthode du gradient — Cauchy (1847), Curry (1944)

- partir d'un point initial x^0
- on répète le pas $x \rightarrow \widetilde{x}$ précédent : $x^k \rightarrow x^{k+1} = \widetilde{(x^k)}$
- critères d'arrêts possibles :
  1. $\max_{i=1}^{n} \left|\frac{\partial f}{\partial x_i}(x^k)\right| < \epsilon$
  2. $\left\|\vec{\nabla} f(x^k)\right\| < \epsilon$
  3. $\left|f(x^{k+1}) - f(x^k)\right| < \epsilon$

Les 3 critères d'arrêt indiquent que f est proche d'être stationnaire

Méthode du gradient (2/2)



Point faible de la méthode du gradient (1/2)

$\varphi'(\widetilde{\lambda}) = -\vec{\nabla} f(\widetilde{x})^T . \vec{\nabla} f(x) = 0$

$\Longrightarrow$ les gradients en x et $\widetilde{x}$ sont orthogonaux

$\Longrightarrow$ à chaque pas on prend une direction orthogonale à la direction précédente

cheminement de la méthode du gradient : «en zigzag»

Point faible de la méthode du gradient (2/2)

$\Longrightarrow$ pour éviter les zigzags et accélérer la convergence, on peut avoir recours à l'un des procédés suivants :

- diminuer le pas : ne pas aller jusqu'en $\widetilde{x}$.

  Polyak (66) : effectuer des pas prédéterminés en imposant une suite (λ^k) telle que $\lambda^k \downarrow 0$ et $\sum_k \lambda^k = +\infty$
  $\Longrightarrow (x^k)$ tend vers $\widehat{x}$

- utiliser d'autres directions que l'anti-gradient :

  Forsythe (1968), Luenberger (1973) :
  toutes les m itérations, au lieu de partir dans la direction de l'anti-gradient en x^k, «couper» en partant dans la direction $\delta = x^k - x^{k-m}$

- utiliser des directions «conjuguées»

Cours 11 : Gradient conjugué

Christophe Gonzales

LIP6 – Université Paris 6, France

Point faible de la méthode du gradient (1/2)

$\varphi'(\widetilde{\lambda}) = -\vec{\nabla} f(\widetilde{x})^T.\vec{\nabla} f(x) = 0$

$\Longrightarrow$ les gradients en x et $\widetilde{x}$ sont orthogonaux

$\Longrightarrow$ à chaque pas on prend une direction orthogonale à la direction précédente

cheminement de la méthode du gradient : «en zigzag»

Point faible de la méthode du gradient (2/2)

$\Longrightarrow$ pour éviter les zigzags et accélérer la convergence, on peut avoir recours à l'un des procédés suivants :

- diminuer le pas : ne pas aller jusqu'en $\widetilde{x}$.

  Polyak (66) : effectuer des pas prédéterminés en imposant une suite (λ^k) telle que $\lambda^k \downarrow 0$ et $\sum_k \lambda^k = +\infty$
  $\Longrightarrow (x^k)$ tend vers $\widehat{x}$

- utiliser d'autres directions que l'anti-gradient :

  Forsythe (1968), Luenberger (1973) :
  toutes les m itérations, au lieu de partir dans la direction de l'anti-gradient en x^k, «couper» en partant dans la direction $\delta = x^k - x^{k-m}$

- utiliser des directions «conjuguées»

Matrices définies positives

Définition d'une matrice définie positive

- Matrice M carrée $n \times n$
- M est symétrique : $M = M^T$, i.e., $M_{jk} = M_{kj}\ \forall j, k$
- $x^T M x > 0\ \forall x \neq 0$

$\Longrightarrow$ M correspond à une matrice de changement de base dans un espace vectoriel de dimension n :

Propriété

- si M = matrice définie positive
- alors $\exists Q$ matrice carrée
- $\exists P$ matrice diagonale à coeffs > 0 telles que : $M = Q^T P^T P Q$
  $\Longrightarrow x^T M x = x^T Q^T P^T P Q x = (PQx)^T (PQx)$

Forme quadratique définie positive (QDP)

Définition d'une forme quadratique définie positive

- $f : \mathbb{R}^n \mapsto \mathbb{R}$ forme quadratique définie positive
- $f(x) = \frac{1}{2} x^T C x + p^T x \quad \forall x \in \mathbb{R}^n$
- C : matrice définie positive ; p : vecteur de taille n

$\Longrightarrow$ forme développée de f :

$$f(x_1, \ldots, x_n) = \frac{1}{2} \left[\sum_{j=1}^{n} x_j \left[\sum_{k=1}^{n} c_{jk} x_k \right] \right] + \sum_{j=1}^{n} p_j x_j$$

$$= \frac{1}{2} \sum_{j=1}^{n} c_{jj} x_j^2 + \sum_{j=1}^{n} \sum_{k=1}^{j-1} c_{jk} x_j x_k + \sum_{j=1}^{n} p_j x_j$$

Propriétés d'une forme QDP (1/3)

Une forme QDP est strictement convexe

Démonstration : Soit $x, y \in \mathbb{R}^n$, $y \neq x$, et $\lambda \in]0, 1[$

$$\begin{aligned}
f(\lambda x + (1-\lambda) y) \\
&= \tfrac{1}{2} (\lambda x + (1-\lambda) y)^T C (\lambda x + (1-\lambda) y) + p^T (\lambda x + (1-\lambda) y) \\
&= \tfrac{1}{2} \left[\lambda^2 x^T C x + (1-\lambda)^2 y^T C y + \lambda (1-\lambda) \left[x^T C y + y^T C x \right] \right] \\
&\quad + \lambda p^T x + (1-\lambda) p^T y \\
&= \lambda f(x) + (1-\lambda) f(y) \\
&\quad + \tfrac{1}{2} \lambda (1-\lambda) \left[-x^T C x + x^T C y + y^T C x - y^T C y \right] \\
&= \lambda f(x) + (1-\lambda) f(y) - \tfrac{1}{2} \lambda (1-\lambda) (y-x)^T C (y-x)
\end{aligned}$$

Or C définie positive $\Longrightarrow (y-x)^T C (y-x) > 0$

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

Propriétés d'une forme QDP (2/3)

toute forme QDP est strictement convexe $\Longrightarrow$ cours de la semaine dernière :

> Une forme QDP a un unique point stationnaire
> = minimum global

point stationnaire $\Longrightarrow \vec{\nabla} f(x) = 0$

$$\vec{\nabla} f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \cdots \\ \frac{\partial f(x)}{\partial x_j} \\ \cdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix} = \begin{bmatrix} c_{11}x_1 + \sum_{k \neq 1} c_{1k}x_k + p_1 \\ \cdots \\ c_{jj}x_j + \sum_{k \neq j} c_{jk}x_k + p_j \\ \cdots \\ c_{nn}x_n + \sum_{k \neq n} c_{nk}x_k + p_n \end{bmatrix} = \begin{bmatrix} C_1 x + p_1 \\ \cdots \\ C_j x + p_j \\ \cdots \\ C_n x + p_n \end{bmatrix}$$

Propriétés d'une forme QDP (3/3)

> $\vec{\nabla} f(x) = Cx + p$

$\Longrightarrow$ f atteint son minimum en $\widehat{x}$:

unique solution du système linéaire $Cx + p = 0$:

> f minimal en $\widehat{x} = -C^{-1}p$

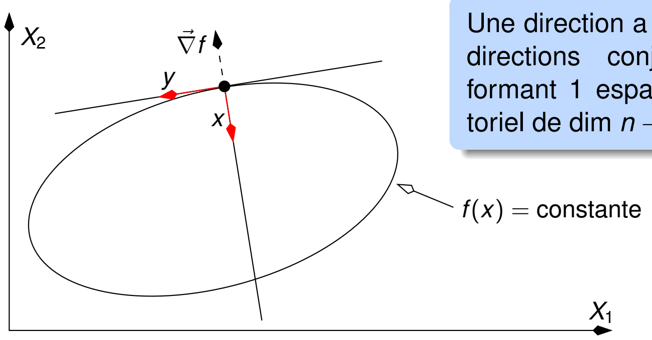
méthode des directions conjuguées

appliquée à une fonction quadratique, c'est une méthode itérative de calcul de la solution $\widehat{x}$ du système $Cx + p = 0$

Directions conjuguées (1/2)

Directions conjuguées

- 2 vecteurs $x, y \in \mathbb{R}^n \setminus \{0\}$
- f : forme quadratique (C matrice définie positive)
- x et y définissent des directions conjuguées si $x^T C y = 0$



> Une direction a 1 ∞ de directions conjuguées formant 1 espace vectoriel de dim $n - 1$

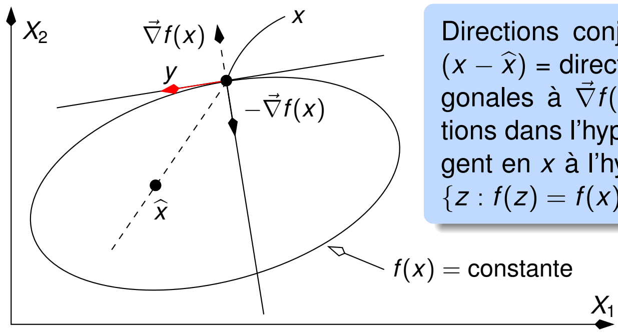
Directions conjuguées (2/2)

Cas particulier de direction conjuguée :

direction $\vec{\nabla} f(x) = Cx + p$

⚠ minimum de f atteint en $\widehat{x}$ tel que $p = -C\widehat{x}$

$\Longrightarrow y^T\vec{\nabla} f(x) = y^T(Cx + p) = y^T Cx + y^T p = y^T C(x - \widehat{x})$



Directions conjuguées de $(x - \widehat{x})$ = directions orthogonales à $\vec{\nabla} f(x)$ = directions dans l'hyperplan tangent en x à l'hypersurface $\{z : f(z) = f(x)\}$

Changement de repère pour gradient conjugué

Lemme 1

- C une matrice définie positive
- $y^1, \dots, y^k, \dots, y^n$ = n vecteurs linéairement indépendants

Alors, on peut construire n vecteurs δ^1, ..., δ^k, ..., δ^n tels que :

- $\delta^1, \dots, \delta^n$ linéairement indépendants
- les δ^i conjugués deux à deux
- $\delta^1 = y^1$

  $\delta^2 = y^2 - \left(\frac{{y^2}^T C\delta^1}{{\delta^1}^T C\delta^1}\right)\delta^1$

  $\dots$

  $\delta^k = y^k - \sum_{l=1}^{k-1}\left(\frac{{y^k}^T C\delta^l}{{\delta^l}^T C\delta^l}\right)\delta^l$

  $\dots$

$\Longrightarrow$ tout vecteur = combinaison linéaire des directions conjuguées fonction de $\vec{\nabla} f$

Expression de l'optimum de f

Lemme 2

- f = forme quadratique définie positive
  $f(x) = \frac{1}{2}x^T Cx + p^T x$
- x^0 = un point quelconque
- n vecteurs $\delta^1, \dots, \delta^n$ linéairement indépendants et conjugués deux à deux

Alors le minimum de f est atteint en $\widehat{x} = x^0 + \sum_{k=1}^{n} \widehat{\lambda}_k \delta^k$ où :

- $\forall k,\ \widehat{\lambda}_k = -\frac{\vec{\nabla} f(x^0)^T \delta^k}{{\delta^k}^T C\delta^k}$
- $\widehat{\lambda}_k$ minimise $\varphi_k(\lambda_k) = f(x^0 + \lambda_k\delta^k)$
- $\widehat{\lambda}_k$ minimise $\psi_k(\lambda_k) = f(x^{k-1} + \lambda_k\delta^k)$ où $x^{k-1} = x^0 + \sum_{l=1}^{k-1} \lambda_l\delta^l$
- en fait on a aussi $\widehat{\lambda}_k = -\frac{\vec{\nabla} f(x^{k-1})^T \delta^k}{{\delta^k}^T C\delta^k}$

Méthode du gradient conjugué pour les formes QDP

Méthode du gradient conjugué

1. *Initialisation :*
   Choix de x^0 point initial et d'une direction $\delta^1 = -\vec{\nabla} f(x^0)$
2. *$k^{ème}$ étape :*
   1. Calcul de $\widehat{\lambda}_k$ minimisant $\varphi_k(\lambda_k) = f(x^{k-1} + \lambda_k \delta^k)$:
   $$\widehat{\lambda}_k = -\frac{\vec{\nabla} f(x^{k-1})^T \delta^k}{\delta^{k^T} C \delta^k} = \frac{\vec{\nabla} f(x^{k-1})^T \vec{\nabla} f(x^{k-1})}{\delta^{k^T} C \delta^k}$$
   2. $x^k = x^{k-1} + \widehat{\lambda}_k \delta^k$
   3. calcul de $\vec{\nabla} f(x^k)$
   4. si $\vec{\nabla} f(x^k) = 0$ alors x^k est optimal fin ; sinon :
   $\delta^{k+1} = -\vec{\nabla} f(x^k) + \beta^k \delta^k$, avec :
   $$\beta^k = \frac{\vec{\nabla} f(x^k)^T C \delta^k}{\delta^{k^T} C \delta^k} = \frac{\|\vec{\nabla} f(x^k)\|^2}{\|\vec{\nabla} f(x^{k-1})\|^2}$$
3. $k \longleftarrow k+1$

Efficacité de la méthode du gradient conjugué

proposition

La méthode du gradient conjugué pour les fonctions quadratiques atteint le minimum en n itérations.

Extension aux fonctions quelconques

Fletcher et Reeves (1964)

Adaptation l'algorithme du gradient conjugué à une fonction différentiable quelconque

Méthode du gradient conjugué pour les fonctions quelconques

1. *Initialisation :*
   Choix de x^0 point initial et d'une direction $\delta^1 = -\vec{\nabla} f(x^0)$
2. *$k^{ème}$ étape :*
   1. Calcul de $\widehat{\lambda}_k$ minimisant $\varphi_k(\lambda_k) = f(x^{k-1} + \lambda_k \delta^k)$
   2. $x^k = x^{k-1} + \widehat{\lambda}_k \delta^k$
   3. calcul de $\vec{\nabla} f(x^k)$
   4. si $\|\vec{\nabla} f(x^k)\| < \varepsilon$ alors fin ; sinon :
   $\delta^{k+1} = -\vec{\nabla} f(x^k) + \beta^k \delta^k$ avec $\beta^k = \dfrac{\|\vec{\nabla} f(x^k)\|^2}{\|\vec{\nabla} f(x^{k-1})\|^2}$
3. $k \longleftarrow k+1$

Convergence de l'algo de Fletcher et Reeves

Convergence de l'algorithme de Fletcher et Reeves

La convergence de l'algorithme de Fletcher et Reeves n'est assurée que si on le rapplique à partir du point 1 périodiquement en prenant comme nouveau point de départ le dernier x^k calculé, e.g., toutes les n itérations.

Cours 12 : Optimisation non linéaire sous contraintes

Christophe Gonzales

LIP6 – Université Paris 6, France

Problème à résoudre

Optimisation sous contraintes

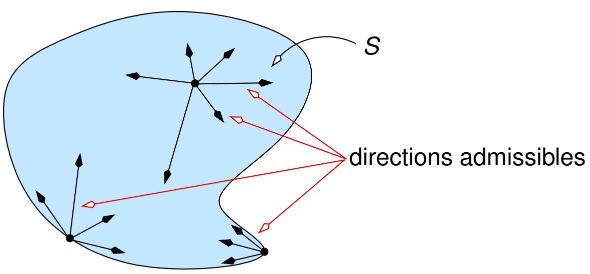
- $f : \mathbb{R}^n \mapsto \mathbb{R}$ fonction à minimiser
- $S \subseteq \mathbb{R}^n$
- Problème d'optimisation : $\min_{x \in S} f(x)$

	f	S
prog linéaire	linéaire	polyèdre convexe
prog quadratique	forme quadratique semi définie positive	polyèdre convexe
prog convexe	convexe	ensemble convexe
prog non linéaire	quelconque	ensemble quelconque

Directions admissibles

Définition d'une direction admissible

- Soit $f : \mathbb{R}^n \mapsto \mathbb{R}, \quad S \subseteq \mathbb{R}^n, \quad x \in S$
- d = direction admissible si $\exists \overline{\lambda} > 0$ t.q. : $x + \lambda d \in S \quad \forall 0 \leq \lambda \leq \overline{\lambda}$



Directions admissibles à l'optimum (1/2)

Proposition 1

- $f : \mathbb{R}^n \mapsto \mathbb{R}$ de classe C^1
- $\widehat{x}$ minimum local de f sur S

$\Longrightarrow \forall$ direction admissible d en $\widehat{x}$, on a $\vec{\nabla} f(\widehat{x})^T d \geq 0$

Corollaire

- $f : \mathbb{R}^n \mapsto \mathbb{R}$ de classe C^1
- une condition nécessaire pour que $\widehat{x} \in S$ soit minimum local de f est que $\vec{\nabla} f(\widehat{x})^T d \geq 0$, $\forall$ direction admissible d en $\widehat{x}$

Directions admissibles à l'optimum (2/2)

Démonstration de la proposition :

d direction admissible en $\widehat{x} \Longrightarrow \exists \overline{\lambda} > 0$ t.q. $\widehat{x} + \lambda d \in S$, $\forall 0 \leq \lambda \leq \overline{\lambda}$

Soit $g(\lambda) = f(\widehat{x} + \lambda d)$

hypothèses $\Longrightarrow g$: minimum local en $\lambda = 0$

développement limité : $g(\lambda) = g(0) + \lambda g'(0) + o(\lambda)$

si $g'(0) < 0$ alors, pour λ petit, $\lambda g'(0) + o(\lambda) < 0$

$\Longrightarrow g(\lambda) < g(0) \Longrightarrow$ impossible : $g(0)$ minimum local

$\Longrightarrow g'(0) \geq 0$. Or $g'(\lambda) = \vec{\nabla} f(\widehat{x} + \lambda d)^T d$

et donc $\vec{\nabla} f(\widehat{x})^T d \geq 0$

Conditions de Kuhn et Tucker (1/4)

Cas particulier d'ensemble S

- $S = \{x : g_i(x) \leq 0,\ i = 1, \ldots, m\}$
- $g_i : \mathbb{R}^n \mapsto \mathbb{R}$ de classe C^1

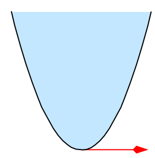
- x^* : point quelconque de S
- contrainte serrée = contrainte telle que $g_i(x^*) = 0$
- $I(x^*) = \{$indices i t.q. $g_i(x^*) = 0\}$

  $= \{$indices des contraintes serrées$\}$
- $D(x^*) = \{d : \vec{\nabla} g_i(x^*)^T d \leq 0, i \in I(x^*)\}$

  $\Longrightarrow$ direction admissible en $x^* \in D(x^*)$

  ⚠ l'inverse n'est pas forcément vrai :

$S = \{x : x_1^2 - x_2 \leq 0\}$



$D(0) = \{(d_1, d_2) : d_2 \geq 0\}$

$\Longrightarrow d = (1, 0) \in D(0)$

Conditions de Kuhn et Tucker (2/4)

$\hat{x}$ = minimum local de f sur S

$I(\hat{x}) = \{i_1, \ldots, i_k\}$

Supp que $D(\hat{x}) = \{$toutes les directions admissibles en $\hat{x}\}$

Soit $\overline{A}$ = matrice des $-\vec{\nabla} g_i(\hat{x})$, $i \in I(\hat{x})$

$$\overline{A} = \begin{bmatrix} -\frac{\partial g_{i_1}}{\partial x_1}(\hat{x}) & -\frac{\partial g_{i_2}}{\partial x_1}(\hat{x}) & \cdots & -\frac{\partial g_{i_k}}{\partial x_1}(\hat{x}) \\ \ldots & \ldots & \ldots & \ldots \\ -\frac{\partial g_{i_1}}{\partial x_n}(\hat{x}) & -\frac{\partial g_{i_2}}{\partial x_n}(\hat{x}) & \cdots & -\frac{\partial g_{i_k}}{\partial x_n}(\hat{x}) \end{bmatrix}$$

Rappel Proposition 1 :

$\hat{x}$ minimum local $\Longrightarrow \forall$ direction admissible d en $\hat{x}$, $\vec{\nabla} f(\hat{x})^T d \geq 0$

$\Longrightarrow$ le système $d^T \overline{A} \geq 0, \vec{\nabla} f(\hat{x})^T d < 0$ n'a pas de solution

Conditions de Kuhn et Tucker (3/4)

Rappel : Lemme de Farkas

- A une matrice $m \times n$
- c un vecteur de taille n
- les deux énoncés suivants sont équivalents :
  1. $Ax \geq 0 \Longrightarrow c^T x \geq 0$
  2. il existe $v \geq 0$ tel que $v^T A = c^T$

le système $d^T \overline{A} \geq 0, \vec{\nabla} f(\hat{x})^T d < 0$ n'a pas de solution

$\Longrightarrow (d^T \overline{A} \geq 0 \Longrightarrow \vec{\nabla} f(\hat{x})^T d \geq 0)$

$\Longrightarrow \exists$ vecteur $\lambda \geq 0$ tel que $\overline{A}\lambda = \vec{\nabla} f(\hat{x})$

$$\Longrightarrow \vec{\nabla} f(\hat{x}) = \sum_{i \in I(\hat{x})} -\lambda_i \vec{\nabla} g_i(\hat{x}) \Longrightarrow \vec{\nabla} f(\hat{x}) + \sum_{i \in I(\hat{x})} \lambda_i \vec{\nabla} g_i(\hat{x}) = 0$$

Conditions de Kuhn et Tucker (4/4)

$\Longrightarrow$ conditions de Kuhn et Tucker (1951) :

Conditions nécessaires d'optimalité de Kuhn et Tucker

- f de classe C^1
- hypothèse de qualification des contraintes :

  $\hat{x}$ = point de S tel que $D(\hat{x}) = \{$directions admissibles en $\hat{x}\}$
- alors condition nécessaire pour que $\hat{x}$ = minimum local de f :

  $\exists \lambda_i \geq 0$, $i = 1, \ldots, m$, tels que :
  1. $\vec{\nabla} f(\hat{x}) + \sum_{i=1}^{m} \lambda_i \vec{\nabla} g_i(\hat{x}) = 0$
  2. $\lambda_i g_i(\hat{x}) = 0, \quad \forall i \in \{1, \ldots, m\}$

lorsque $i \notin I(\hat{x})$, il suffit de fixer $\lambda_i = 0$

Remarque sur la qualification des contraintes

⚠ l'hypothèse de qualification des contraintes est primordiale pour le théorème de Kuhn et Tucker

Exemple :

$S = \{x \in \mathbb{R}^2 : g_i(x) \leq 0, i = 1, 2, 3\}$

$g_1(x_1, x_2) = -x_1 \quad g_2(x_1, x_2) = -x_2 \quad g_3(x_1, x_2) = x_2 - (1 - x_1)^3$

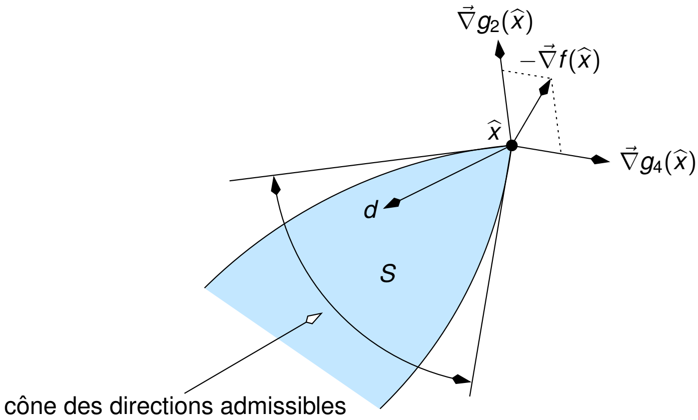
si $f(x_1, x_2) = -x_1$ alors $\widehat{x} = (1, 0)$, $I(\widehat{x}) = \{2, 3\}$

$\vec{\nabla} f(\widehat{x}) = (-1, 0)$

$\lambda_2 \vec{\nabla} g_2(\widehat{x}) + \lambda_3 \vec{\nabla} g_3(\widehat{x}) = \lambda_2(0, 1) + \lambda_3(0, -1) = (0, \lambda_2 - \lambda_3)$

$\Longrightarrow \vec{\nabla} f(\widehat{x}) \neq \lambda_2 \vec{\nabla} g_2(\widehat{x}) + \lambda_3 \vec{\nabla} g_3(\widehat{x})$

Interprétation géométrique de Kuhn et Tucker



Kuhn et Tucker : $-\vec{\nabla} f(\widehat{x})$ = combinaison linéaire (à coeffs positifs) des $\vec{\nabla} g_i(\widehat{x})$

$\Longrightarrow$ $-\vec{\nabla} f(\widehat{x})$ forme un angle obtus avec toute direction admissible

Considérations sur les conditions de Kuhn et Tucker

transparents précédents :

conditions de Kuhn et Tucker = conditions nécessaires

sous certaines hypothèses de convexité : conditions suffisantes

$\Longrightarrow$ on va s'intéresser maintenant à des propriétés de convexité

Convexité (1/3)

Proposition 2

- f de classe C^1
- $S \subseteq \mathbb{R}^n$: ensemble convexe
- f convexe sur $S \iff \forall x, y \in S$:

$$f(y) \geq f(x) + \vec{\nabla} f(x)^T (y - x)$$

Convexité (2/3)

Démonstration :

Supposons f convexe sur S

$\Longrightarrow f(\lambda y + (1-\lambda)x) \leq \lambda f(y) + (1-\lambda) f(x) \quad \forall 0 \leq \lambda \leq 1$

$\Longrightarrow \forall 0 < \lambda \leq 1, \quad \dfrac{f(\lambda y + (1-\lambda)x) - f(x)}{\lambda} \leq f(y) - f(x)$

$$\text{Or } \lim_{\lambda \downarrow 0} \frac{f(\lambda y + (1-\lambda)x) - f(x)}{\lambda} = \lim_{\lambda \downarrow 0} \frac{f(x + \lambda(y - x)) - f(x)}{\lambda} = \vec{\nabla} f(x)^T (y - x)$$

$\Longrightarrow \vec{\nabla} f(x)^T (y - x) \leq f(y) - f(x)$

$\Longrightarrow f(y) \geq f(x) + \vec{\nabla} f(x)^T (y - x)$

Convexité (3/3)

Réciproque :

supp $f(y) \geq f(x) + \vec{\nabla} f(x)^T (y - x) \ \forall x, y \in S$

Soit 2 points $x_1, x_2 \in S$ et $x = \lambda x_1 + (1-\lambda) x_2, \ \lambda \in [0, 1]$

$$\Longrightarrow \begin{cases} f(x_1) \geq f(x) + \vec{\nabla} f(x)^T (x_1 - x) \\ f(x_2) \geq f(x) + \vec{\nabla} f(x)^T (x_2 - x) \end{cases}$$

$$\begin{aligned} \Longrightarrow \lambda f(x_1) + (1-\lambda) f(x_2) &\geq f(x) + \vec{\nabla} f(x)^T [\lambda(x_1 - x) + (1-\lambda)(x_2 - x)] \\ &\geq f(x) + \vec{\nabla} f(x)^T [\lambda x_1 + (1-\lambda) x_2 - x] \\ &\geq f(x) = f(\lambda x_1 + (1-\lambda) x_2) \end{aligned}$$

$\Longrightarrow f$ convexe sur S

Caractérisation d'un minimum global (1/2)

Rappel : Proposition 1 : minimum **local**

- $f : \mathbb{R}^n \mapsto \mathbb{R}$ de classe C^1
- $\widehat{x}$ minimum local de f sur S

$\Longrightarrow \forall$ direction admissible d en $\widehat{x}$, on a $\vec{\nabla} f(\widehat{x})^T d \geq 0$

Proposition 3 : minimum **global**

- $S \subseteq \mathbb{R}^n$: ensemble convexe
- $f : \mathbb{R}^n \mapsto \mathbb{R}$ de classe C^1 convexe sur S
- $\widehat{x}$ minimum global de f sur $S \Longleftrightarrow \forall y \in S, \vec{\nabla} f(\widehat{x})^T (y - \widehat{x}) \geq 0$

Caractérisation d'un minimum global (2/2)

Démonstration :

$\widehat{x}$ minimum global de $f \Longrightarrow \widehat{x}$ minimum local de f

Proposition 1 $\Longrightarrow \vec{\nabla} f(\widehat{x})^T d \geq 0 \quad \forall$ direction d admissible en $\widehat{x}$

S convexe $\Longrightarrow (y \in S \Longrightarrow (y - \widehat{x})$ admissible)

$\Longrightarrow \forall y \in S, \vec{\nabla} f(\widehat{x})^T (y - \widehat{x}) \geq 0$

Réciproque : supp que $\vec{\nabla} f(\widehat{x})^T (y - \widehat{x}) \geq 0 \quad \forall y \in S$

f convexe $\Longrightarrow f(y) \geq f(\widehat{x}) + \vec{\nabla} f(\widehat{x})^T (y - \widehat{x})$ d'après proposition 2

$\Longrightarrow f(y) \geq f(\widehat{x})$

$\Longrightarrow \widehat{x} =$ optimum global de f sur S

Rappel : Conditions nécessaire d'optimalité

Rappel : Conditions de Kuhn et Tucker

- f de classe C^1
- $I(x^*) = \{\text{indices } i \text{ t.q. } g_i(x^*) = 0\}$
- $D(x^*) = \{d : \vec{\nabla} g_i(x^*)^T d \leq 0, i \in I(x^*)\}$
- $\widehat{x}$ = point de S tel que $D(\widehat{x}) = \{\text{directions admissibles en } \widehat{x}\}$
- alors condition nécessaire pour que $\widehat{x} =$ minimum local de f :

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

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

Conditions suffisantes d'optimalité de Kuhn et Tucker

Conditions suffisantes d'optimalité de Kuhn et Tucker

- g_i convexes de classe C^1, $i = 1, \ldots, m$
- $S = \{x : g_i(x) \leq 0,\ i = 1, \ldots, m\}$
- f de classe C^1 convexe sur S
- $\widehat{x}$ = point de S tel que $\exists \lambda_i \geq 0$, $i = 1, \ldots, m$, tels que :
  1. $\vec{\nabla} f(\widehat{x}) + \sum_{i=1}^{m} \lambda_i \vec{\nabla} g_i(\widehat{x}) = 0$
  2. $\lambda_i g_i(\widehat{x}) = 0, \quad \forall i \in \{1, \ldots, m\}$

$\Longrightarrow \widehat{x} =$ minimum global de f sur S

Conditions suffisantes d'optimalité (2/2)

Démonstration :

Soit $y \in S$, $y \neq \widehat{x}$. On veut montrer que $f(y) \geq f(\widehat{x})$

les g_i convexes $\Longrightarrow$ S convexe

$\Longrightarrow \widehat{d} = y - \widehat{x}$ direction admissible en $\widehat{x}$

$\Longrightarrow \vec{\nabla} g_i(\widehat{x})^T \widehat{d} \leq 0 \ \ \forall i \in I(\widehat{x})$

Or conditions 1 et 2 de Kuhn et Tucker équivalentes à :

$$\vec{\nabla} f(\widehat{x})^T d \geq 0 \ \forall \text{ direction } d \text{ telle que } \vec{\nabla} g_i(\widehat{x})^T d \leq 0 \ \forall i \in I(\widehat{x})$$

On sait que $\widehat{d}$ vérifie $\vec{\nabla} g_i(\widehat{x})^T \widehat{d} \leq 0 \ \ \forall i \in I(\widehat{x})$

Hypothèse de la proposition $\Longrightarrow \vec{\nabla} f(\widehat{x})^T \widehat{d} = f(\widehat{x})^T (y - \widehat{x}) \geq 0$

Or Prop. 3 : $\widehat{x}$ minimum global de f $\Longleftrightarrow \forall y \in S$, $\vec{\nabla} f(\widehat{x})^T (y - \widehat{x}) \geq 0$

$\Longrightarrow \widehat{x}$ minimum global de f

Fonction de Lagrange

Définition de la fonction de Lagrange

- *Problème d'origine* :

  min $f(x)$

  s.c. $g_i(x) \leq 0, \quad i = 1, \ldots, m$
- *Fonction de Lagrange* :

  $$L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)$$
- $\lambda_i \geq 0$: *multiplicateur de Lagrange*

Fonctions primales / duales

Fonction de Lagrange : $L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)$

Fonction primale / duale

- Fonction primale : $L^*(x) = \max_{\lambda \geq 0} L(x, \lambda)$
- Fonction duale : $L_*(\lambda) = \min_{x} L(x, \lambda)$

Problème primal / dual

- Problème primal : $L^*(\widehat{x}) = \min_{x} L^*(x) = \min_{x} \max_{\lambda \geq 0} L(x, \lambda)$
- Problème dual : $L_*(\widehat{\lambda}) = \max_{\lambda \geq 0} L_*(\lambda) = \max_{\lambda \geq 0} \min_{x} L(x, \lambda)$

Interprétation du problème primal / dual

$$L^*(x) = \max_{\lambda \geq 0} L(x, \lambda)$$
$$= \max_{\lambda \geq 0} \left(f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) \right)$$
$$= \begin{cases} f(x) \text{ si } g_i(x) \leq 0 \ \forall i \\ +\infty \text{ si } g_i(x) > 0 \text{ pour un } i \end{cases}$$

Problème primal : $L^*(\widehat{x}) = \min_x L^*(x) \Longrightarrow$ ne s'intéresser qu'aux x tels que $g_i(x) \leq 0$

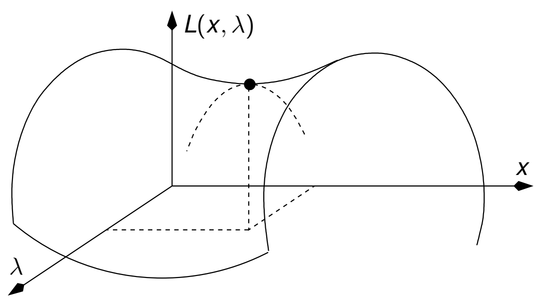
$\Longrightarrow$ le problème devient alors : $\begin{cases} \min f(x) \\ s.c. \;\; g_i(x) \leq 0, \quad i = 1, \ldots, m \end{cases}$

Point Selle

Définition d'un point selle

- Soit $\overline{x} \in S, \quad \overline{\lambda} \geq 0$
- $(\overline{x}, \overline{\lambda})$ est un point selle de $L(x, \lambda)$ si :

$$L(\overline{x}, \overline{\lambda}) \leq L(x, \overline{\lambda}) \qquad \forall x \in S$$
$$L(\overline{x}, \lambda) \leq L(\overline{x}, \overline{\lambda}) \qquad \forall \lambda \geq 0$$



Caractérisation d'un point Selle (1/4)

Proposition 4 : Caractérisation d'un point Selle

- Soit $\overline{x} \in S, \quad \overline{\lambda} \geq 0$
- $(\overline{x}, \overline{\lambda})$ est un point selle de $L(x, \lambda)$ si et seulement si :
  1. $L(\overline{x}, \overline{\lambda}) = \min_{x \in S} L(x, \overline{\lambda})$
  2. $g_i(\overline{x}) \leq 0 \; \forall i \in \{1, \ldots, m\}$
  3. $\overline{\lambda}_i g_i(\overline{x}) = 0 \; \forall i \in \{1, \ldots, m\}$

Caractérisation d'un point Selle (2/4)

Démonstration :

supp $(\overline{x}, \overline{\lambda})$ point selle $\Longrightarrow L(\overline{x}, \lambda) \leq L(\overline{x}, \overline{\lambda}) \leq L(x, \overline{\lambda}) \quad \forall x, \forall \lambda \geq 0$

$L(\overline{x}, \overline{\lambda}) \leq L(x, \overline{\lambda}) \; \forall x \Longrightarrow L(\overline{x}, \overline{\lambda}) = \min_x L(x, \overline{\lambda}) \Longrightarrow$ ❶

$$L(\overline{x}, \lambda) \leq L(\overline{x}, \overline{\lambda}) \Longleftrightarrow f(\overline{x}) + \sum_{i=1}^{m} \lambda_i g_i(\overline{x}) \leq f(\overline{x}) + \sum_{i=1}^{m} \overline{\lambda}_i g_i(\overline{x})$$

$$\Longleftrightarrow \sum_{i=1}^{m} (\lambda_i - \overline{\lambda}_i) g_i(\overline{x}) \leq 0 \quad \forall \lambda \geq 0$$

Si $\exists i$ t.q. $g_i(\overline{x}) > 0$, alors λ_i suffisamment grand

$$\Longrightarrow \sum_{i=1}^{m} (\lambda_i - \overline{\lambda}_i) g_i(\overline{x}) > 0 \Longrightarrow \text{impossible} \Longrightarrow ❷ \text{ est vrai}$$

Caractérisation d'un point Selle (3/4)

$$L(\overline{x}, \lambda) \leq L(\overline{x}, \overline{\lambda}) \Longleftrightarrow \sum_{i=1}^{m} (\lambda_i - \overline{\lambda}_i) g_i(\overline{x}) \leq 0 \quad \forall \lambda \geq 0$$

$$\lambda = 0 \Longrightarrow \sum_{i=1}^{m} \overline{\lambda}_i g_i(\overline{x}) \geq 0$$

Or $\overline{\lambda} \geq 0$ et $g_i(\overline{x}) \leq 0 \Longrightarrow \sum_{i=1}^{m} \overline{\lambda}_i g_i(\overline{x}) \leq 0$

$$\Longrightarrow \sum_{i=1}^{m} \overline{\lambda}_i g_i(\overline{x}) = 0 \Longrightarrow \overline{\lambda}_i g_i(\overline{x}) = 0 \quad \forall i \Longrightarrow ❸$$

Caractérisation d'un point Selle (4/4)

Réciproque :

Supposons ❶, ❷ et ❸

❶ $\Longrightarrow L(\overline{x}, \overline{\lambda}) \leq L(x, \overline{\lambda}) \quad \forall x \in S$

❸ $\Longrightarrow L(\overline{x}, \overline{\lambda}) = f(\overline{x})$

$$L(\overline{x}, \lambda) = f(\overline{x}) + \sum_{i=1}^{m} \lambda_i g_i(\overline{x}) \leq f(\overline{x}) = L(\overline{x}, \overline{\lambda}) \;\; \forall \lambda \geq 0$$

$\Longrightarrow L(\overline{x}, \lambda) \leq L(\overline{x}, \overline{\lambda}) \leq L(x, \overline{\lambda}) \Longrightarrow (\overline{x}, \overline{\lambda})$ point selle

Condition du point selle (1/2)

Condition d'optimalité du point selle

Si $(\widehat{x}, \widehat{\lambda})$ point selle de la fonction de Lagrange, alors :

- $\min_x L(x, \widehat{\lambda}) = L(\widehat{x}, \widehat{\lambda}) = \max_{\lambda \geq 0} L(\widehat{x}, \lambda)$
- $\widehat{x}$ = solution de $\begin{cases} \min f(x) \\ s.c. \;\; g_i(x) \leq 0, \quad i = 1, \ldots, m \end{cases}$

Condition du point selle (2/2)

Démonstration :

$(\widehat{x}, \widehat{\lambda})$ point selle $\Longrightarrow L(\widehat{x}, \lambda) \leq L(\widehat{x}, \widehat{\lambda}) \leq L(x, \widehat{\lambda}) \quad \forall x, \forall \lambda \geq 0$

$L(\widehat{x}, \lambda) \leq L(\widehat{x}, \widehat{\lambda}) \; \forall \lambda \geq 0 \Longrightarrow L(\widehat{x}, \widehat{\lambda}) = \max_{\lambda \geq 0} L(\widehat{x}, \lambda)$

$L(\widehat{x}, \widehat{\lambda}) \leq L(x, \widehat{\lambda}) \; \forall x \Longrightarrow L(\widehat{x}, \widehat{\lambda}) = L(\widehat{x}, \widehat{\lambda}) = \min_x L(x, \widehat{\lambda})$

$$L(\widehat{x}, \widehat{\lambda}) \leq L(x, \widehat{\lambda}) \; \forall x \Longleftrightarrow f(\widehat{x}) + \sum_{i=1}^{m} \widehat{\lambda}_i g_i(\widehat{x}) \leq f(x) + \sum_{i=1}^{m} \widehat{\lambda}_i g_i(x)$$

Or ❸ de la proposition 4 : $\widehat{\lambda}_i g_i(\widehat{x}) = 0 \; \forall i \in \{1, \ldots, m\}$

$$\Longrightarrow f(\widehat{x}) \leq f(x) + \sum_{i=1}^{m} \widehat{\lambda}_i g_i(x) \leq f(x) \;\; \forall x \in S$$

$\Longrightarrow \widehat{x} =$ optimum global de f sur S

Cours 13 : Méthodes de gradient

Christophe Gonzales

LIP6 – Université Paris 6, France

Méthodes primales / duales

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

Méthodes primales / duales

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

Méthode du gradient projeté

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

Sinon le min de $f(x)$ sous contraintes obtenu pour $\widehat{x}$ 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

$\Longrightarrow$ projeter les déplacements sur la frontière de la région des réalisables

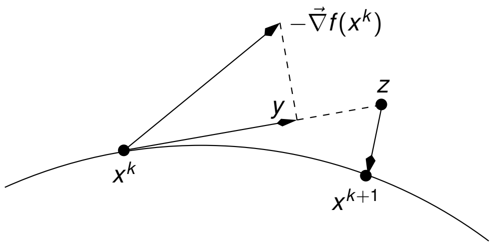
$\Longrightarrow$ génération d'une séquence de points réalisables

Exemple : méthode de Rosen (1960)

Méthode du gradient projeté de Rosen

Idée : projeter le gradient sur la frontière des réalisables
$\Longrightarrow$ chemin sur la frontière dans la direction de la plus grande pente

$\Longrightarrow$ fonctionne essentiellement avec de contraintes linéaires :



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

Problème non linéaire avec contraintes linéaires

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

Remarques :

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

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

Résolution de (P) par la méthode de Rosen (1/9)

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

- On suppose qu'on connaît une solution x
  $\Longrightarrow$ 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 $\Longrightarrow$ déplacements réalisables
  $\Longrightarrow$ si d est une direction alors $a_i^T d = 0 \; \forall i \in I^0$
  $\Longrightarrow A^0 d = 0$

Résolution de (P) par la méthode de Rosen (2/9)

Déplacement $d \Longrightarrow$ faire décroître f

$\Longrightarrow$ 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 :

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

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 ?

Résolution de (P) par la méthode de Rosen (3/9)

Théorème

- A^0 matrice $q \times n$, $q \leq n$, de rang q
  $\Longrightarrow$ les a_i^T sont linéairement indépendants
  (pas de dégénérescence)
- la solution optimale du problème :

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

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

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

  $P^0 =$ 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 $\mathbb{R}^n$

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

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

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

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

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

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

$\Longrightarrow$ il reste maintenant à trouver l'expression de $\overline{y}$

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

caractérisation de $\overline{y}$:

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

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

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

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

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

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

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

CQFD

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

Rappel du théorème

La solution optimale du problème :

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

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

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

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

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

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 $d = \overline{y}/\|\overline{y}\|$
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 :

$$\alpha_{\max} = \max\{\alpha \geq 0 : x + \alpha\overline{y} \in X\}$$
$$\text{où } X = \{x \in \mathbb{R}^n : a_i x \leq b_i \ \forall i \in I_1 \text{ et } a_i x = b_i \ \forall i \in I_2\}$$

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

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

Résolution de (P) par la méthode de Rosen (8/9)

Problème : et si $\overline{y} = 0$?

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

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

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 $\widehat{x}$ = minimum local de f :

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

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

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

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

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

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

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

$\Longrightarrow$ trouver une nouvelle direction d

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

$\Longrightarrow$ nouvelle matrice de projection

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

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

$\Longrightarrow$ 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 $P^0 = Id - A^{0T}\left[A^0 A^{0T}\right]^{-1} A^0$
   calculer $y^k = -P^0 \vec{\nabla} f(x^k)$
   si $y^k = 0$ alors aller en 5
4. $y^k \neq 0$: calculer $\alpha_{\text{max}} = \max\{\alpha \geq 0 : x^k + \alpha y^k \in X\}$
   calculer x^{k+1} tel que $f(x^{k+1}) = \min_{\alpha \in [0, \alpha_{\text{max}}]}\{f(x^k + \alpha y^k)\}$
   $k \leftarrow k + 1$; retourner en 2
5. calculer $u = -\left[A^0 A^{0T}\right]^{-1} A^0 \vec{\nabla} f(x)$
   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

Cours 14 : Optimisation quadratique

Christophe Gonzales

LIP6 – Université Paris 6, France

Programmation quadratique

Définition d'un programme quadratique

$$\min Q(x) = \frac{1}{2}x^T Cx + p^T x$$
$$s.c. \begin{cases} Ax = b \\ x \geq 0 \end{cases}$$

avec C symétrique semi-définie positive

Rappel : matrice semi-définie positive

- Matrice C carrée $n \times n$
- C est symétrique : $C = C^T$, i.e., $C_{jk} = C_{kj}\ \forall j, k$
- $x^T Cx \geq 0\ \forall x$

C semi-définie positive $\Longrightarrow$ Q convexe

Q convexe $\Longrightarrow$ optimum local = optimum global

méthode de Rosen $\Longrightarrow$ optimum global

Conditions de Kuhn et Tucker pour la prog quadratique

$$\min Q(x) = \frac{1}{2}x^T Cx + p^T x$$
$$s.c. \begin{cases} Ax = b \\ x \geq 0 \end{cases}$$

Conditions de Kuhn et Tucker pour ce problème

- $Ax = b$
- $Cx - v + A^T u = -p$
- $x \geq 0, \quad v \geq 0$
- $x^T v = 0$

⚠ Les conditions de Kuhn et Tucker diffèrent du cours 9 car contraintes $Ax = b$ et non($Ax \leq b$)

⚠ à part $x^T v = 0$, les autres équations sont linéaires
résolution par simplexe ? $\Longrightarrow$ méthode de Wolfe (1959)

Méthode de Wolfe

$$\min Q(x) = \frac{1}{2}x^T Cx + p^T x$$
$$s.c. \begin{cases} Ax = b \\ x \geq 0 \end{cases}$$

2 formes pour la méthode de Wolfe : la courte et la longue

formes de la méthode de Wolfe

- *forme courte* :
  Suppose que $p = 0$ ou C définie positive
- *forme longue* :
  Pas de contrainte sur p ou C
  Revient à appliquer 2 fois la forme courte

⚠ dans la suite du cours, étude de la forme courte

Méthode de Wolfe – forme courte (1/8)

$$\min Q(x) = \frac{1}{2}x^T Cx + p^T x$$
$$s.c. \begin{cases} Ax = b \\ x \geq 0 \end{cases}$$

$$\begin{aligned} & Ax = b \\ & Cx - v + A^T u = -p \\ & x \geq 0, \quad v \geq 0 \\ & x^T v = 0 \end{aligned}$$

Idée de la méthode de Wolfe

- ajout de variables artificielles $\Longrightarrow$ solution réalisable des conditions de Kuhn et Tucker
- suppression de ces variables via l'algorithme du simplexe $\Longrightarrow$ similaire à une Phase I du simplexe excepté qu'on rajoute une règle pour assurer que $x^T v = 0$

Méthode de Wolfe – forme courte (2/8)

Kuhn et Tucker : $\begin{cases} Ax = b \\ Cx - v + A^T u = -p \\ x \geq 0, \quad v \geq 0 \\ x^T v = 0 \end{cases}$

Vers une solution réalisable :

introduire des variables artificielles :

$w^T = (w_1, \ldots, w_m) \qquad z^{1T} = (z_1^1, \ldots, z_n^1) \qquad z^{2T} = (z_1^2, \ldots, z_n^2)$

$\Longrightarrow$ nouveau système élargi :

$$\begin{cases} Ax + w = b \\ Cx - v + A^T u + z^1 - z^2 = -p \\ x \geq 0, \quad v \geq 0, \quad z^1 \geq 0, \quad z^2 \geq 0, \quad w \geq 0 \\ x^T v = 0 \end{cases}$$

Méthode de Wolfe – forme courte (3/8)

$$\begin{cases} Ax + w = b \\ Cx - v + A^T u + z^1 - z^2 = -p \\ x \geq 0, \quad v \geq 0, \quad z^1 \geq 0, \quad z^2 \geq 0, \quad w \geq 0 \\ x^T v = 0 \end{cases}$$

Solution évidente du système :

$$\begin{cases} x = 0, \quad v = 0, \quad u = 0 \\ w = b \\ z_i^1 = -p_i \text{ si } p_i \text{ négatif} \\ z_i^2 = \ \ p_i \text{ si } p_i \text{ positif} \end{cases}$$

Méthode de Wolfe :
élimination des z_i^j et des w_i de la base

base réalisable : les w_i, les z_i^1 ou z_i^2 non nuls

⚠ si $p_i = 0$, z_i^1 ou z_i^2 entre en base (peu importe lequel)

Méthode de Wolfe – forme courte (4/8)

élimination des w_i :

Partir de la solution du transparent précédent et résoudre :

$$\min \sum_{i=1}^{m} w_i$$

$$s.c. \quad \begin{cases} Ax + w = b \\ Cx - v + A^T u + z^1 - z^2 = -p, \\ x \geq 0, \quad v \geq 0, \quad z^1 \geq 0, \quad z^2 \geq 0, \quad w \geq 0 \\ x^T v = 0 \end{cases}$$

⚠ Dans cette phase de la méthode, on conserve toujours $u = 0$ et $v = 0$, i.e., les u_i et les v_i restent hors base

⚠ problème d'origine = contraintes incompatibles $\Longrightarrow \hat{w} \neq 0$

⚠ si dégénérescence, faire attention à sortir les $w_i = 0$ de la base

Méthode de Wolfe – forme courte (5/8)

fin de la phase précédente :

base = m variables x_i et n variables z_i^1 ou z_i^2

2ème phase : élimination des z_i^j

on supprime du système les colonnes relatives aux w_i et aux z_i^j qui ne sont pas en base $\Longrightarrow$ nouveau système :

$$\begin{cases} Ax = b \\ Cx - v + A^T u + Dz = -p \\ x \geq 0, \quad v \geq 0, \quad z \geq 0 \\ x^T v = 0 \end{cases}$$

où $z =$ vecteur des z_i^j encore en base

$D =$ matrice diagonale : 1 pour les z_i^1 et des -1 pour les z_i^2

Méthode de Wolfe – forme courte (6/8)

Actuellement : solution réalisable du système avec $u = 0$ et $v = 0$

réalisation de la 2ème phase :

minimiser $\sum_i z_i$ tout en préservant $x^T v = 0$

$\Longrightarrow$ utilisation du simplexe en ajoutant la règle :

Règle supplémentaire de pivotage du simplexe

- Si x_i se trouve en base, v_i ne peut rentrer en base
- Si v_i se trouve en base, x_i ne peut rentrer en base

Si on trouve $\sum_i \hat{z}_i = 0$ alors programme quadratique résolu

Problème : se peut-il que la nouvelle règle de pivotage empêche tout pivotage alors que $\sum_i \hat{z}_i > 0$?

Méthode de Wolfe – forme courte (7/8)

Lemme

- Soit q un vecteur de taille h
- Soit (G, H) une partition de $\{1, \ldots, n\}$
- Soit le programme linéaire en w :

$$\min q^T w$$
$$s.c. \begin{cases} Ax = b \\ Cx - v + A^T u + Rw = f \\ x \geq 0, \quad v \geq 0, \quad w \geq 0, \quad v_G = 0, \quad x_H = 0 \end{cases}$$

- $\hat{w}$: solution optimale du programme linéaire
- alors il existe un vecteur r tel que :

$$Cr = 0 \qquad Ar = 0 \qquad q^T \hat{w} = f^T r$$

Application à notre problème :

$$\begin{cases} w = z \\ q^T = (1, 1, \ldots, 1) \\ q^T w = \sum_i z_i, \qquad R = D, \qquad f = -p \end{cases}$$

Méthode de Wolfe – forme courte (8/8)

lemme précédent $\Longrightarrow \exists r$ tel que :

$$Cr = 0 \qquad \text{et} \qquad min \sum_i z_i = -p^T r$$

Or hypothèse de la méthode courte : $p = 0$ ou C définie positive

$$p = 0 \Longrightarrow \sum_i z_i = 0$$

C définie positive $\Longrightarrow r = 0 \Longrightarrow \sum_i z_i = 0$

Donc, dans tous les cas, $\sum_i z_i = 0$

La méthode de Wolfe (forme courte) résout le problème quadratique d'origine