Sommaire

Optimisation linéaire, l'algorithme du simplexe

- Un exemple
- Généralités
- L'algorithme du simplexe sur un exemple
- Définitions et terminologie
- Résumé d'une itération
- La dégénérescence et le cyclage
- Recherche d'un dictionnaire réalisable

Dualité en programmation linéaire

Optimisation non linéaire sans contraintes

Un exemple

Une usine fabrique deux sortes de produits, p_1 et p_2, à l'aide de deux machines m_1 et m_2. La quantité fabriquée de ces produits est un réel positif ou nul. Chaque unité de produit doit passer sur les deux machines dans un ordre indifférent et pendant les temps suivants, exprimés en minutes :

	p_1	p_2
m_1	30 mn	20 mn
m_2	40 mn	10 mn

m_1 : disponible 6000 minutes par mois.
m_2 : disponible 4000 minutes par mois.

Profit réalisé sur une unité du produit p_1 : 400 €.
Profit réalisé sur une unité du produit p_2 : 200 €.

Plan de fabrication mensuel qui maximise le profit ?

Un exemple, suite

	p_1	p_2	disponibilité
m_1	30 mn	20 mn	6000 mn
m_2	40 mn	10 mn	4000 mn
profit	400 €	600 €	

Choix des variables ?

Un exemple, suite

	p_1	p_2	disponibilité
m_1	30 mn	20 mn	6000 mn
m_2	40 mn	10 mn	4000 mn
profit	400 €	600 €	

x : le nombre d'unités du produit p_1 à fabriquer mensuellement.
y : le nombre d'unités du produit p_2 à fabriquer mensuellement.

Mise en équations ?

Un exemple, suite

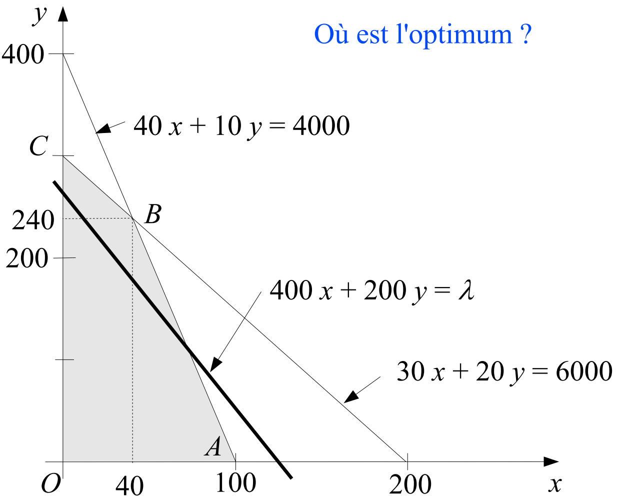
	p_1	p_2	disponibilité
m_1	30 mn	20 mn	6000 mn
m_2	40 mn	10 mn	4000 mn
profit	400 €	600 €	

x : le nombre d'unités du produit p_1 à fabriquer mensuellement.
y : le nombre d'unités du produit p_2 à fabriquer mensuellement.

Le problème peut s'exprimer sous la forme suivante :

$$\text{Maximiser } z = 400x + 200y$$
$$\text{avec les contraintes : } \begin{cases} 30x + 20y \leqslant 6000 \\ 40x + 10y \leqslant 4000 \\ x \geqslant 0, y \geqslant 0. \end{cases}$$

$$\text{Maximiser } z = 400x + 200y \quad \text{avec :} \begin{cases} 30x + 20y \leqslant 6000 \\ 40x + 10y \leqslant 4000 \\ x \geqslant 0, y \geqslant 0. \end{cases}$$



Optimisation linéaire, l'algorithme du simplexe

Un exemple

Généralités

L'algorithme du simplexe sur un exemple

Définitions et terminologie

Résumé d'une itération

La dégénérescence et le cyclage

Recherche d'un dictionnaire réalisable

Généralités

Le problème vu en exemple : problème d'optimisation linéaire car les contraintes et la fonction objectif s'expriment comme fonctions linéaires des variables.

Résolution graphique de l'exemple possible parce qu'il n'y a que deux variables.

Généralités

Plus généralement, un problème de programmation linéaire est un problème qui peut se formuler comme suit :

Forme standard d'un problème de programmation linéaire

Maximiser une forme linéaire de n variables $x_1, \ldots, x_n$: $\sum_{j=1}^{n} c_j x_j$
les variables étant soumises :
- à m contraintes linéaires : pour $i \in \{1, 2, \ldots, m\}, \sum_{j=1}^{n} a_{ij} x_j \leqslant b_i$
- aux n contraintes de positivité : pour $j \in \{1, 2, \ldots, n\}, x_j \geq 0$.

Cette formulation s'appelle la *forme standard* d'un problème de programmation linéaire.

Inégalités toujours larges $\Longrightarrow$ le domaine défini par les contraintes est fermé.

Généralités

Des problèmes où les fonctions sont toutes linéaires peuvent facilement se mettre sous forme standard quand :

• il s'agit de minimisation

• il apparaît des contraintes d'égalité ou d'inégalité large dans l'autre sens

• des variables ont d'autres contraintes que celles d'être positives ou nulles ou sont non contraintes.

Mais on ne sait pas traiter systématiquement des inégalités strictes.

On utilise pour cela les remarques suivantes.

Transformation en un problème mis sous forme standard

• Minimiser une fonction f (linéaire ou non) revient à maximiser $-f$ (puisque minimum de $f = -$ maximum de $(-f)$) ;

• on transforme une inégalité du genre « $\geqslant$ » en une inégalité du genre « $\leqslant$ » en la multipliant par -1 ;

• une égalité :

$$\sum_{j=1}^{n} a_{ij}x_j = b_i$$

revient aux deux inégalités :

$$\sum_{j=1}^{n} a_{ij}x_j \leqslant b_i \text{ et } \sum_{j=1}^{n}(-a_{ij})x_j \leqslant -b_i.$$

Suite et fin dans le prochain transparent.

Transformation en un problème mis sous forme standard

• on remplace une variable contrainte par la double inégalité :

$$\alpha \leqslant x \leqslant \beta$$

par la variable $y = x - \alpha$ et on ajoute les contraintes :

$$y \leqslant \beta - \alpha \text{ et } y \geqslant 0\,;$$

• on remplace une variable x contrainte par l'inégalité :

$$x \geqslant \alpha$$

par la variable

$$y = x - \alpha$$

qui devra être positive ou nulle : $y \geqslant 0$;

• on exprime une variable x qui n'est pas contrainte par la différence de deux variables positives ou nulles :

$$x = x^+ - x^-$$

avec :

$$x^+ \geqslant 0 \text{ et } x^- \geqslant 0.$$

Principe de l'algorithme du simplexe

On considére un problème mis sous forme standard :

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

avec les contraintes : $\begin{cases} \text{pour } i \in \{1, 2, ..., m\}, \sum_{j=1}^{n} a_{ij} x_j \leqslant b_i \\ \text{pour } j \in \{1, 2, ..., n\}, x_j \geqslant 0. \end{cases}$

L'ensemble des points de $\mathbb{R}^n$ de coordonnées $x_1, ..., x_n$ et vérifiant les $m + n$ contraintes précédentes détermine un polyèdre appelé *polyèdre des contraintes*.

Les n-uplets $(x_1, ..., x_n)$ qui satisfont les contraintes s'appellent *solutions réalisables* du problème. Ce sont les coordonnées des points intérieurs (au sens large) du polyèdre des contraintes qui, dans notre exemple, était le quadrilatère $OABC$.

Le polyèdre des solutions réalisables est convexe car intersection de demi-plans, qui sont convexes.

Principe de l'algorithme du simplexe, suite

On peut prouver le théorème suivant :

Théorème

Soit un problème de programmation linéaire dont le polyèdre des contraintes est non vide et dont la fonction à maximiser est majorée sur ce polyèdre. Alors le problème admet un maximum (qui est fini) atteint en au moins un sommet du polyèdre des contraintes.

Idée de l'algorithme du simplexe : passer itérativement d'un sommet du polyèdre des contraintes à un sommet adjacent en suivant des arêtes du polyèdre et de façon à augmenter la valeur de la fonction à optimiser, jusqu'à trouver un sommet où le maximum est atteint.

C'est grâce à la convexité du polyèdre et à la linéarité de la fonction dont on cherche le maximum que l'on peut se contenter de chercher le maximum en un sommet du polyèdre.

Un exemple
Généralités
L'algorithme du simplexe sur un exemple
Définitions et terminologie
Résumé d'une itération
La dégénérescence et le cyclage
Recherche d'un dictionnaire réalisable

L'algorithme du simplexe sur un exemple

Une fabrique de tissus produit quatre types de tissus : T_1, T_2, T_3 et T_4. Ils sont produits en longueur variable, mesurée en kilomètre.

Trois opérations principales : la filature, le tissage, la teinture.

La production d'un kilomètre de tissu nécessite un certain nombre d'heures de filature, de tissage et de teinture. La vente des tissus rapporte un certain profit exprimé en euros.

Les données sont précisées dans le tableau suivant, pour un kilomètre de tissu :

	T_1	T_2	T_3	T_4	disponibilité
filature	2	4	5	7	42
tissage	1	1	2	2	17
teinture	1	2	3	3	24
profit	7	9	18	17	

Plan de fabrication de façon à maximiser le profit ?

L'algorithme du simplexe sur un exemple, suite

x_1, x_2, x_3, x_4 : longueurs respectives de T_1, de T_2, de T_3 et de T_4 produites quotidiennement.

Modélisation du problème :

Maximiser $z = 7x_1 + 9x_2 + 18x_3 + 17x_4$
avec les contraintes :

$$\begin{cases} 2x_1 + 4x_2 + 5x_3 + 7x_4 \leqslant 42 \\ x_1 + x_2 + 2x_3 + 2x_4 \leqslant 17 \\ x_1 + 2x_2 + 3x_3 + 3x_4 \leqslant 24 \\ x_1 \geqslant 0, x_2 \geqslant 0, x_3 \geqslant 0, x_4 \geqslant 0. \end{cases}$$

Problème de programmation linéaire sous forme standard.

L'algorithme du simplexe sur un exemple, suite

Maximiser $z = 7x_1 + 9x_2 + 18x_3 + 17x_4$

$$\text{avec : } \begin{cases} 2x_1 + 4x_2 + 5x_3 + 7x_4 \leqslant 42 \\ x_1 + x_2 + 2x_3 + 2x_4 \leqslant 17 \\ x_1 + 2x_2 + 3x_3 + 3x_4 \leqslant 24 \\ x_1 \geqslant 0, x_2 \geqslant 0, x_3 \geqslant 0, x_4 \geqslant 0. \end{cases}$$

On introduit trois variables dites *variables d'écart* x_5, x_6, x_7, positives ou nulles ; elles mesurent, pour chaque ressource, l'écart entre la quantité initialement disponible et la quantité consommée par le plan de fabrication donné par x_1, x_2, x_3 et x_4.

On obtient ce qui s'appelle un *dictionnaire* :

$$\begin{array}{llrcrcrcrcr} x_5 & = 42 & - & 2x_1 & - & 4x_2 & - & 5x_3 & - & 7x_4 \\ x_6 & = 17 & - & x_1 & - & x_2 & - & 2x_3 & - & 2x_4 \\ x_7 & = 24 & - & x_1 & - & 2x_2 & - & 3x_3 & - & 3x_4 \\ \hline z & = & & 7x_1 & + & 9x_2 & + & 18x_3 & + & 17x_4 \end{array}$$

Le problème : Maximiser z avec, pour $1 \leqslant i \leqslant 7$, $x_i \geqslant 0$.

L'algorithme du simplexe sur un exemple, suite

$$
\begin{array}{rcrcrcrcrcr}
x_5 & = 42 & - & 2x_1 & - & 4x_2 & - & 5x_3 & - & 7x_4 \\
x_6 & = 17 & - & x_1 & - & x_2 & - & 2x_3 & - & 2x_4 \\
x_7 & = 24 & - & x_1 & - & 2x_2 & - & 3x_3 & - & 3x_4 \\
\hline
z & = & & 7x_1 & + & 9x_2 & + & 18x_3 & + & 17x_4
\end{array}
$$

Le polyèdre des contraintes est limité dans $\mathbb{R}^4$ par les hyperplans d'équation $x_1 = 0, x_2 = 0, x_3 = 0, x_4 = 0, x_5 = 0, x_6 = 0, x_7 = 0$.

Dans le dictionnaire ci-dessus :

- x_5, x_6 et x_7 sont les *variables de base* (ou *variables en base*) du dictionnaire ;

- x_1, x_2, x_3 et x_4 sont les *variables hors-base* du dictionnaire.

La *solution basique* associée au dictionnaire est la solution obtenue en attribuant la valeur 0 à toutes les variables hors-base ; les valeurs des variables de base en découlent.

L'algorithme du simplexe sur un exemple, suite

$$
\begin{array}{lcrcrcrcrcr}
x_5 & = 42 & - & 2x_1 & - & 4x_2 & - & 5x_3 & - & 7x_4 \\
x_6 & = 17 & - & x_1 & - & x_2 & - & 2x_3 & - & 2x_4 \\
x_7 & = 24 & - & x_1 & - & 2x_2 & - & 3x_3 & - & 3x_4 \\
\hline
z & = & & 7x_1 & + & 9x_2 & + & 18x_3 & + & 17x_4
\end{array}
$$

On utilisera le signe * lorsqu'il s'agit de valeurs : x^* représentera une valeur prise par la variable x.

Solution basique associée au dictionnaire :
$x_1^* = 0, x_2^* = 0, x_3^* = 0, x_4^* = 0$, ce qui entraîne $x_5^* = 42, x_6^* = 17$ et $x_7^* = 24$.

Les sept variables ayant des valeurs positives ou nulles dans cette solution basique, on dit que ce dictionnaire est *réalisable*.

On peut remarquer que le point de coordonnées (0, 0, 0, 0) est un sommet du polyèdre des contraintes.

La solution basique associée au dictionnaire donne à z la valeur 0.

L'algorithme du simplexe sur un exemple, suite

$$\begin{array}{rcl}
x_5 & = 42 & - \; 2x_1 \; - \; 4x_2 \; - \; 5x_3 \; - \; 7x_4 \\
x_6 & = 17 & - \; x_1 \; - \; x_2 \; - \; 2x_3 \; - \; 2x_4 \\
x_7 & = 24 & - \; x_1 \; - \; 2x_2 \; - \; 3x_3 \; - \; 3x_4 \\
\hline
z & = & 7x_1 + 9x_2 + 18x_3 + 17x_4
\end{array}$$

Base de la méthode : si, choisissant une variable hors-base de coefficient strictement positif, on fait croître celle-ci à partir de 0, les autres variables hors-base restant nulles, la valeur correspondante de la fonction z croît.

Dans notre exemple, choisissons la variable x_3 (on pourrait aussi choisir ici l'une quelconque des trois autres variables hors-base). Ce choix s'exprime par : x_3 *est la variable entrante*.

Gardant x_1, x_2 et x_4 à 0, nous cherchons à augmenter x_3 au maximum, tout en conservant la propriété que le point M de $\mathbb{R}^4$ de coordonnées $(0, 0, x_3, 0)$ reste dans le polyèdre des contraintes (on se déplace sur une arête du polyèdre des contraintes issue du sommet (0, 0, 0, 0)).

L'algorithme du simplexe sur un exemple, suite

$$\begin{array}{lcrcrcrcr} x_5 & = 42 & - & 2x_1 & - & 4x_2 & - & 5x_3 & - & 7x_4 \\ x_6 & = 17 & - & x_1 & - & x_2 & - & 2x_3 & - & 2x_4 \\ x_7 & = 24 & - & x_1 & - & 2x_2 & - & 3x_3 & - & 3x_4 \\ \hline z & = & & 7x_1 & + & 9x_2 & + & 18x_3 & + & 17x_4 \end{array}$$

Les contraintes sur l'augmentation de la variable x_3 sont :
$x_5 \geqslant 0$, ce qui impose $x_3 \leqslant 8,4$;
$x_6 \geqslant 0$, ce qui impose $x_3 \leqslant 8,5$;
$x_7 \geqslant 0$, ce qui impose $x_3 \leqslant 8$.

Le premier hyperplan que rencontre le point M est donc celui d'équation $x_7 = 0$: le point M est alors arrivé à un nouveau sommet du polyèdre des contraintes, à l'intersection des hyperplans d'équations $x_1 = 0$, $x_2 = 0$, $x_4 = 0$, $x_7 = 0$. On dit que x_7 *est la variable sortante*.

Par ce choix : $x_1^* = 0$, $x_2^* = 0$, $x_4^* = 0$, $x_7^* = 0$, les trois autres variables restent positives ou nulles.

L'algorithme du simplexe sur un exemple, suite

$$\begin{array}{rcrcrcrcr}
x_5 & = 42 & - & 2x_1 & - & 4x_2 & - & 5x_3 & - & 7x_4 \\
x_6 & = 17 & - & x_1 & - & x_2 & - & 2x_3 & - & 2x_4 \\
x_7 & = 24 & - & x_1 & - & 2x_2 & - & 3x_3 & - & 3x_4 \\
\hline
z & = & & 7x_1 & + & 9x_2 & + & 18x_3 & + & 17x_4
\end{array}$$

On fait un changement de dictionnaire en échangeant les rôles de x_3 et x_7.

On utilise l'équation du premier dictionnaire qui donne x_7 pour exprimer x_3 en fonction de x_1, x_2, x_4 et x_7 ; on remplace ensuite x_3 par cette expression dans les autres équations du dictionnaire.

On obtient ainsi un deuxième dictionnaire :

$$\begin{array}{rlrcrcrcrcr}
x_3 & = 8 & - & x_1/3 & - & 2x_2/3 & - & x_4 & - & x_7/3 \\
x_5 & = 2 & - & x_1/3 & - & 2x_2/3 & - & 2x_4 & + & 5x_7/3 \\
x_6 & = 1 & - & x_1/3 & + & x_2/3 & & & + & 2x_7/3 \\
\hline
z & = 144 & + & x_1 & - & 3x_2 & - & x_4 & - & 6x_7
\end{array}$$

L'algorithme du simplexe sur un exemple, suite

$$
\begin{array}{llcccccccc}
x_3 & = 8 & - & x_1/3 & - & 2x_2/3 & - & x_4 & - & x_7/3 \\
x_5 & = 2 & - & x_1/3 & - & 2x_2/3 & - & 2x_4 & + & 5x_7/3 \\
x_6 & = 1 & - & x_1/3 & + & x_2/3 & & & + & 2x_7/3 \\
\hline
z & = 144 & + & x_1 & - & 3x_2 & - & x_4 & - & 6x_7
\end{array}
$$

Les variables de base sont maintenant x_3, x_5 et x_6, et les variables hors-base x_1, x_2, x_4 et x_7.

Dans la nouvelle solution basique, la fonction z vaut 144, valeur que l'on obtient en annulant les variables hors-base. On remarque qu'on a ainsi une nouvelle solution réalisable plus intéressante que celle associée au premier dictionnaire.

L'algorithme du simplexe sur un exemple, suite

$$\begin{array}{rcl}
x_3 & = 8 & - \; x_1/3 \; - \; 2x_2/3 \; - \; x_4 \; - \; x_7/3 \\
x_5 & = 2 & - \; x_1/3 \; - \; 2x_2/3 \; - \; 2x_4 \; + \; 5x_7/3 \\
x_6 & = 1 & - \; x_1/3 \; + \; x_2/3 \; + \; 2x_7/3 \\
\hline
z & = 144 & + \; x_1 \; - \; 3x_2 \; - \; x_4 \; - \; 6x_7
\end{array}$$

Dans la nouvelle expression de la fonction z, la variable x_1 a un coefficient strictement positif : on décide de faire entrer x_1 en base pour parcourir une nouvelle arête du polyèdre des contraintes.

Limites sur l'augmentation possible de la valeur de x_1 à partir de la valeur nulle, les autres variables hors base restant à 0 :

$x_3 \geqslant 0$, ce qui impose $x_1 \leqslant 24$;
$x_5 \geqslant 0$, ce qui impose $x_1 \leqslant 6$;
$x_6 \geqslant 0$, ce qui impose $x_1 \leqslant 3$.

C'est la troisième limite qui est la plus contraignante ; x_6 sort de la base.

L'algorithme du simplexe sur un exemple, suite

$$
\begin{array}{lllllllll}
x_3 & = 8 & - & x_1/3 & - & 2x_2/3 & - & x_4 & - & x_7/3 \\
x_5 & = 2 & - & x_1/3 & - & 2x_2/3 & - & 2x_4 & + & 5x_7/3 \\
x_6 & = 1 & - & x_1/3 & + & x_2/3 & & & + & 2x_7/3 \\
\hline
z & = 144 & + & x_1 & - & 3x_2 & - & x_4 & - & 6x_7
\end{array}
$$

En faisant entrer x_1 en base est sortir x_6, on obtient le dictionnaire suivant :

$$
\begin{array}{lllllllll}
x_1 & = 3 & + & x_2 & & & - & 3x_6 & + & 2x_7 \\
x_3 & = 7 & - & x_2 & - & x_4 & + & x_6 & - & x_7 \\
x_5 & = 1 & - & x_2 & - & 2x_4 & + & x_6 & + & x_7 \\
\hline
z & = 147 & - & 2x_2 & - & x_4 & - & 3x_6 & - & 4x_7
\end{array}
$$

Les variables de base sont maintenant x_1, x_3 et x_5, et les variables hors-base x_2, x_4, x_6 et x_7.

Dans la nouvelle solution basique, la fonction z vaut 147, valeur que l'on obtient en annulant les variables hors-base.

L'algorithme du simplexe sur un exemple, suite

$$\begin{array}{lllllllll}
x_1 & = 3 & + & x_2 & & & - & 3x_6 & + & 2x_7 \\
x_3 & = 7 & - & x_2 & - & x_4 & + & x_6 & - & x_7 \\
x_5 & = 1 & - & x_2 & - & 2x_4 & + & x_6 & + & x_7 \\
\hline
z & = 147 & - & 2x_2 & - & x_4 & - & 3x_6 & - & 4x_7
\end{array}$$

Les variables x_2, x_4, x_6, x_7 étant positives ou nulles, l'optimum cherché de z est majoré par 147.

La solution basique actuelle nous fournit donc une solution optimum du problème :

• il faut fabriquer chaque jour trois kilomètres de T_1, zéro de T_2, sept de T_3 et zéro de T_4 ;

• toutes les heures de tissage et de teinture sont utilisées, alors qu'il reste une heure de filage disponible ;

• le bénéfice vaut 147.

L'algorithme du simplexe sur un exemple, remarques

1. La solution obtenue est entière alors que cela n'était pas imposée par la formulation du problème. Ceci n'a rien de général et les problèmes de programmation linéaire en nombres entiers (c'est-à-dire des problèmes de programmation linéaire pour lesquels les variables doivent être entières) peuvent être qualitativement plus compliqués.

2. La méthode consiste, à chaque étape, à faire entrer en base une variable dont le coefficient dans la fonction z à optimiser est strictement positif. Cela ne permet cependant pas toujours d'obtenir une croissance stricte de z. Nous y reviendrons.

3. Nous avons eu la chance de trouver, sans aucune difficulté, un sommet du polyèdre des contraintes, i.e. un dictionnaire réalisable, qui nous sert de point de départ. En effet, « l'origine était réalisable », i.e. l'annulation des variables x_1, x_2, ..., x_n attribue des valeurs positives ou nulles aux variables d'écart, car les b_i étaient tous positifs ou nuls. Nous y reviendrons.

Optimisation linéaire, l'algorithme du simplexe

Un exemple
Généralités
L'algorithme du simplexe sur un exemple
Définitions et terminologie
Résumé d'une itération
La dégénérescence et le cyclage
Recherche d'un dictionnaire réalisable

Définitions et terminologie

Problème de programmation linéaire en les variables $x_1, \ldots, x_n$ mis sous forme standard :

maximiser une forme linéaire $z = \sum_{j=1}^{n} c_j x_j$

les variables étant soumises :

- à m contraintes linéaires : pour $i \in \{1, 2, \ldots, m\}, \sum_{j=1}^{n} a_{ij} x_j \leqslant b_i$
- aux n contraintes de positivité : pour $j \in \{1, 2, \ldots, n\}, x_j \geqslant 0$.

Tout n-uplet de valeurs $(x_1, \ldots, x_n)$ satisfaisant les contraintes constitue une *solution réalisable*.

La fonction z est appelée *fonction objectif*.

Définitions et terminologie, suite

$x_1, \ldots, x_n$: *variables de décision* ou *variables de choix* ou *variables principales* ou *variables initiales*.

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

Une solution $x_1^*, x_2^*, \ldots, x_{n+m}^*$ est réalisable $\Longleftrightarrow$ toutes ses valeurs sont positives ou nulles.

Une solution réalisable qui maximise la fonction objectif : *solution optimale*.

Problème de programmation linéaire n'admetant aucune solution réalisable : *infaisable* ou *non réalisable*.

Problème admetant des solutions réalisables mais n'ayant pas de valeur optimale finie : *non borné*.

Définitions et terminologie, suite

Dictionnaire : système d'équations linéaires liant $x_1, \dots, x_n, x_{n+1}, \dots, x_{n+m}$ et z, et satisfaisant les deux propriétés suivantes :

• les équations constituant un dictionnaire quelconque doivent exprimer z et m des $n+m$ variables $x_1, \dots, x_{n+m}$ en fonction des n autres variables et ce, de manière unique ;

• tout dictionnaire est algébriquement équivalent au dictionnaire définissant les variables d'écart et la fonction objectif, c'est-à-dire au dictionnaire :

$$\begin{cases} x_{n+1} = b_1 - \sum_{j=1}^{n} a_{1j} x_j \\ \dots \\ x_{n+i} = b_i - \sum_{j=1}^{n} a_{ij} x_j \\ \dots \\ x_{n+m} = b_m - \sum_{j=1}^{n} a_{mj} x_j \\ \hline z = \sum_{j=1}^{n} c_j x_j \end{cases}$$

Définitions et terminologie, suite

Base du problème : m des $n + m$ variables $x_1, x_2, \ldots, x_{n+m}$ de sorte que ces m variables s'expriment de manière unique à l'aide des n autres variables, appelées *variables hors-base*.

Une base définit un dictionnaire et réciproquement.

Une base étant fixée, on obtient la *solution basique* associée à cette base (ou au dictionnaire associé à cette base), en attribuant la valeur 0 à toutes les variables hors-base.

Géométriquement, une solution à la fois basique et réalisable correspond à un sommet du polyèdre des contraintes.

L'algorithme du simplexe a pour objectif de déterminer une solution optimale parmi les solutions basiques et réalisables (c'est-à-dire parmi les sommets du polyèdre des contraintes).

Optimisation linéaire, l'algorithme du simplexe

Un exemple
Généralités
L'algorithme du simplexe sur un exemple
Définitions et terminologie
Résumé d'une itération
La dégénérescence et le cyclage
Recherche d'un dictionnaire réalisable

Résumé d'une itération

On suppose que pour $i \in I$, $b_i^{'} \geqslant 0$.

Deux sous-ensembles d'indices :

$J \subset \{1, 2, ..., n + m\}$ avec $|J| = n$, indices des n variables hors-base dans le dictionnaire courant. Initialement, $J = \{1, 2, ..., n\}$;

$I = \{1, 2, ..., n + m\} \setminus J$, indices des m variables de base.

Le dictionnaire courant est décrit par les égalités suivantes :

pour $i \in I$, $x_i = b_i^{'} + \sum_{j \in J} a_{ij}^{'} x_j$

$z = z^* + \sum_{j \in J} c_j^{'} x_j$.

Ce dictionnaire sera par construction toujours réalisable.

Résumé d'une itération, suite

L'itération courante se déroule comme suit :

si tous les coefficients $c_j^{'}$ sont $\leqslant 0$, l'algorithme est terminé : l'annulation des variables hors-base fournit une solution optimale ;

sinon

- on choisit une variable entrante : variable hors-base x_{j_0} pourvue d'un coefficient strictement positif dans z ;
- on détermine la variable sortante x_{i_0} comme étant la variable de base qui restreint le plus la croisssance de x_{j_0} ;
- on extrait x_{j_0} de l'expression courante de x_{i_0} ;
- on remplace x_{j_0} par sa nouvelle expression dans z et dans l'expression des autres variables de base ; on obtient ainsi le nouveau dictionnaire courant à partir duquel on applique l'itération suivante.

Résumé d'une itération, suite

S'il y a plusieurs variables candidates pour entrer en base, on peut :

⋆ prendre la telle première variable rencontrée ;

⋆ ou privilégier la variable ayant le plus grand coefficient dans z (premier critère de Dantzig) ;

⋆ ou privilégier la variable entraînant la plus grande augmentation de z (second critère de Dantzig).

Pour déterminer la variable sortante, on considère, pour $i \in I$ avec $a'_{ij_0} < 0$, les rapports $\frac{-b'_i}{a'_{ij_0}}$: i_0 est l'indice pour lequel ce rapport est le plus petit (s'il y a plusieurs variables candidates pour sortir de la base, on en choisit une arbitrairement).

Remarque. Quand on passe du dictionnaire courant au dictionnaire suivant, on est sûr que celui-ci est réalisable : on passe d'une solution basique réalisable à une autre solution basique réalisable.

Optimisation linéaire, l'algorithme du simplexe

Un exemple
Généralités
L'algorithme du simplexe sur un exemple
Définitions et terminologie
Résumé d'une itération
La dégénérescence et le cyclage
Recherche d'un dictionnaire réalisable

Définition

Les solutions basiques réalisables avec une ou plusieurs variables de base nulles sont dites *dégénérées*.

Exemple

Considérons le dictionnaire (non dégénéré) :

$$
\begin{array}{rlcccccc}
x_4 & = 1 & & & & & - & 2x_3 \\
x_5 & = 3 & - & 2x_1 & + & 4x_2 & - & 6x_3 \\
x_6 & = 2 & + & x_1 & - & 3x_2 & - & 4x_3 \\
\hline
z & = & & 2x_1 & - & x_2 & + & 8x_3
\end{array}
$$

On fait entrer x_3 en base. Les relations $x_4 \geqslant 0$, $x_5 \geqslant 0, x_6 \geqslant 0$ imposent toutes les trois $x_3 \leqslant 0,5$. Chacune des variables x_4, x_5, x_6 est candidate à quitter la base. On choisit x_4, on obtient :

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

Dans la solution basique associée à ce dictionnaire, x_5 et x_6 prennent une valeur nulle : cette solution basique est dégénérée.

Exemple, suite

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

Une itération à partir de ce dictionnaire : x_1 entre en base, $x_5 \geqslant 0$ impose $x_1 \leqslant 0$; z^* n'augmentera pas au cours de cette itération.

Les itérations dégénérées $\rightarrow$ on peut retrouver après un nombre fini d'itérations un dictionnaire déjà rencontré : il y a *cyclage*.

À cause de l'équivalence algébrique des dictionnaires, on retrouve un dictionnaire déjà rencontré dès qu'on retrouve une même partition des $m + n$ variables en variables de base et variables hors-base.

Éviter le cyclage

Règle de Bland

Règle de Bland : lorsque l'on a un choix sur la variable entrante ou sur la variable sortante, on choisit toujours celle de plus petit indice.

Théorème de Bland : Il ne peut y avoir cyclage lorsque, à toute itération effectuée à partir d'un dictionnaire dégénéré, on choisit les variables entrante et sortante comme celles du plus petit indice parmi les candidats possibles.

Preuve dans le polycopié.

Un cas de cyclage

On part du dictionnaire suivant.

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

On fait entrer x_1, on fait sortir x_5.

Après la première itération :

$$
\begin{array}{rcrrrrrrrrr}
x_1 & = & & & 11x_2 & + & 5x_3 & - & 18x_4 & - & 2x_5 \\
x_6 & = & & - & 4x_2 & - & 2x_3 & + & 8x_4 & + & x_5 \\
x_7 & = & 1 & - & 11x_2 & - & 5x_3 & + & 18x_4 & + & 2x_5 \\
\hline
z & = & & & 53x_2 & + & 41x_3 & - & 204x_4 & - & 20x_5
\end{array}
$$

Un cas de cyclage, suite

Rappel :

$$\begin{array}{rcrrrrrrrr}
x_1 & = & & 11x_2 & + & 5x_3 & - & 18x_4 & - & 2x_5 \\
x_6 & = & - & 4x_2 & - & 2x_3 & + & 8x_4 & + & x_5 \\
x_7 & = 1 & - & 11x_2 & - & 5x_3 & + & 18x_4 & + & 2x_5 \\
\hline
z & = & & 53x_2 & + & 41x_3 & - & 204x_4 & - & 20x_5
\end{array}$$

On fait entrer x_2, on fait sortir x_6.

Après la deuxième itération :

$$\begin{array}{rcrrrrrrrr}
x_2 & = & - & 0,5x_3 & + & 2x_4 & + & 0,25x_5 & - & 0,25x_6 \\
x_1 & = & - & 0,5x_3 & + & 4x_4 & + & 0,75x_5 & - & 2,75x_6 \\
x_7 & = 1 & + & 0,5x_3 & - & 4x_4 & - & 0,75x_5 & - & 2,75x_6 \\
\hline
z & = & & 14,5x_3 & - & 98x_4 & - & 6,75x_5 & - & 13,25x_6
\end{array}$$

Un cas de cyclage, suite

Rappel :

$$
\begin{array}{lllrlrlrlr}
x_2 & = & - & 0,5x_3 & + & 2x_4 & + & 0,25x_5 & - & 0,25x_6 \\
x_1 & = & - & 0,5x_3 & + & 4x_4 & + & 0,75x_5 & - & 2,75x_6 \\
x_7 & = 1 & + & 0,5x_3 & - & 4x_4 & - & 0,75x_5 & - & 2,75x_6 \\
\hline
z & = & & 14,5x_3 & - & 98x_4 & - & 6,75x_5 & - & 13,25x_6
\end{array}
$$

On fait entrer x_3, on fait sortir x_1.

Après la troisième itération :

$$
\begin{array}{lllrlrlrlr}
x_3 & = & & 8x_4 & + & 1,5x_5 & - & 5,5x_6 & - & 2x_1 \\
x_2 & = & - & 2x_4 & - & 0,5x_5 & + & 2,5x_6 & + & x_1 \\
x_7 & = 1 & & & & & & & - & x_1 \\
\hline
z & = & & 18x_4 & + & 15x_5 & - & 93x_6 & - & 29x_1
\end{array}
$$

Un cas de cyclage, suite

Rappel :

$$
\begin{array}{rcrrrrrrrr}
x_3 & = & & 8x_4 & + & 1,5x_5 & - & 5,5x_6 & - & 2x_1 \\
x_2 & = & - & 2x_4 & - & 0,5x_5 & + & 2,5x_6 & + & x_1 \\
x_7 & = 1 & & & & & & & - & x_1 \\
\hline
z & = & & 18x_4 & + & 15x_5 & - & 93x_6 & - & 29x_1
\end{array}
$$

On fait entrer x_4, on fait sortir x_2.

Après la quatrième itération :

$$
\begin{array}{rcrrrrrrrr}
x_4 & = & - & 0,25x_5 & + & 1,25x_6 & + & 0,5x_1 & - & 0,5x_2 \\
x_3 & = & - & 0,5x_5 & + & 4,5x_6 & + & 2x_1 & - & 4x_2 \\
x_7 & = 1 & & & & & - & x_1 & & \\
\hline
z & = & & 10,5x_5 & - & 70,5x_6 & - & 20x_1 & - & 9x_2
\end{array}
$$

Un cas de cyclage, suite

Rappel :

$$\begin{array}{llrrrrrrrr}
x_4 & = & - & 0,25x_5 & + & 1,25x_6 & + & 0,5x_1 & - & 0,5x_2 \\
x_3 & = & - & 0,5x_5 & + & 4,5x_6 & + & 2x_1 & - & 4x_2 \\
x_7 & = 1 & & & & & - & x_1 & & \\
\hline
z & = & & 10,5x_5 & - & 70,5x_6 & - & 20x_1 & - & 9x_2
\end{array}$$

On fait entrer x_5, on fait sortir x_3.

Après la cinquième itération :

$$\begin{array}{llrrrrrrrr}
x_5 & = & & 9x_6 & + & 4x_1 & - & 8x_2 & - & 2x_3 \\
x_4 & = & - & x_6 & - & 0,5x_1 & + & 1,5x_2 & + & 0,5x_3 \\
x_7 & = 1 & & & - & x_1 & & & & \\
\hline
z & = & & 24x_6 & + & 22x_1 & - & 93x_2 & - & 21x_3
\end{array}$$

Un cas de cyclage, suite

Rappel :

$$
\begin{array}{rcrrrrrrrr}
x_5 & = & & 9x_6 & + & 4x_1 & - & 8x_2 & - & 2x_3 \\
x_4 & = & - & x_6 & - & 0,5x_1 & + & 1,5x_2 & + & 0,5x_3 \\
x_7 & = 1 & & & - & x_1 & & & & \\
\hline
z & = & & 24x_6 & + & 22x_1 & - & 93x_2 & - & 21x_3
\end{array}
$$

On fait entrer x_6, on fait sortir x_4.

Après la sixième itération :

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

On retrouve le dictionnaire de départ : il y a cyclage.

Avec la règle de Bland

Après la cinquième étape :

$$
\begin{array}{llrrcrcrcr}
x_5 & = & & 9x_6 & + & 4x_1 & - & 8x_2 & - & 2x_3 \\
x_4 & = & - & x_6 & - & 0,5x_1 & + & 1,5x_2 & + & 0,5x_3 \\
x_7 & = 1 & & & & & & & - & x_1 \\
\hline
z & = & & 24x_6 & + & 22x_1 & - & 93x_2 & - & 21x_3
\end{array}
$$

La règle change le choix de la sixième étape.

On fait entrer x_1 (et non plus x_6) et on fait sortir x_4.

Après la sixième itération :

$$
\begin{array}{llrrcrcrcr}
x_1 & = & - & 2x_6 & + & 3x_2 & + & x_3 & - & 2x_4 \\
x_5 & = & & x_6 & + & 4x_2 & + & 2x_3 & - & 8x_4 \\
x_7 & = 1 & + & 2x_6 & - & 3x_2 & - & x_3 & + & 2x_4 \\
\hline
z & = & - & 20x_6 & - & 27x_2 & + & x_3 & - & 44x_4
\end{array}
$$

Avec la règle de Bland, suite

Rappel :

$$
\begin{array}{llrrrrrrrr}
x_1 & = & - & 2x_6 & + & 3x_2 & + & x_3 & - & 2x_4 \\
x_5 & = & & x_6 & + & 4x_2 & + & 2x_3 & - & 8x_4 \\
x_7 & = 1 & + & 2x_6 & - & 3x_2 & - & x_3 & + & 2x_4 \\
\hline
z & = & - & 20x_6 & - & 27x_2 & + & x_3 & - & 44x_4
\end{array}
$$

On fait entrer x_3, on fait sortir x_7.

Après la septième itération :

$$
\begin{array}{llrrrrrrrr}
x_3 & = 1 & + & 2x_6 & - & 3x_2 & + & 2x_4 & - & x_7 \\
x_1 & = 1 & & & & & & & - & x_7 \\
x_5 & = 2 & + & 5x_6 & - & 2x_2 & - & 4x_4 & - & 2x_7 \\
\hline
z & = 1 & - & 18x_6 & - & 30x_2 & - & 42x_4 & - & 2x_7
\end{array}
$$

Tous les coefficients dans z sont $\leqslant 0$, la méthode s'arrête.

La règle de Bland a permis d'éviter le cyclage.

Optimisation linéaire, l'algorithme du simplexe

Un exemple
Généralités
L'algorithme du simplexe sur un exemple
Définitions et terminologie
Résumé d'une itération
La dégénérescence et le cyclage
Recherche d'un dictionnaire réalisable

Un exemple où l'origine n'est pas réalisable

On considère le problème suivant, écrit sous forme standard.

Maximiser $z = x_1 - x_2 + x_3$
avec les contraintes :

$$\begin{cases} 2x_1 - x_2 + 2x_3 \leqslant 4 \\ 2x_1 - 3x_2 + x_3 \leqslant -5 \\ -x_1 + x_2 - 2x_3 \leqslant -1 \\ x_1 \geqslant 0, x_2 \geqslant 0, x_3 \geqslant 0. \end{cases}$$

La solution $x_1 = x_2 = x_3 = 0$ n'est pas réalisable. Nous introduisons le *problème auxiliaire* suivant (que nous écrivons sous forme standard également).

Maximiser $w = -x_0$
avec les contraintes :

$$\begin{cases} 2x_1 - x_2 + 2x_3 - x_0 \leqslant 4 \\ 2x_1 - 3x_2 + x_3 - x_0 \leqslant -5 \\ -x_1 + x_2 - 2x_3 - x_0 \leqslant -1 \\ x_0 \geqslant 0, x_1 \geqslant 0, x_2 \geqslant 0, x_3 \geqslant 0. \end{cases}$$

Généralités

D'une façon plus générale, on obtient le problème auxiliaire en ajoutant x_0 aux b_i.

On peut considérer qu'on augmente les ressources d'une même quantité x_0. Si x_0 est assez grand, les nouvelles ressources deviennent toutes positives ou nulles : le problème auxiliaire est réalisable.

Si le problème initial admet une solution réalisable, on peut prendre $x_0 = 0$. La question est de déterminer la plus petite valeur à attribuer à x_0 pour que le problème soit réalisable. On est amené à minimiser x_0 i.e. à maximiser $-x_0$.

Remarque
On peut se contenter d'enlever x_0 aux premiers membres des inégalités correspondant à une valeur négative des seconds membres.

Exemple et généralités

Rappel :

Maximiser $w = -x_0$

avec les contraintes :

$$\begin{cases} 2x_1 - x_2 + 2x_3 - x_0 \leqslant 4 \\ 2x_1 - 3x_2 + x_3 - x_0 \leqslant -5 \\ -x_1 + x_2 - 2x_3 - x_0 \leqslant -1 \\ x_0 \geqslant 0, x_1 \geqslant 0, x_2 \geqslant 0, x_3 \geqslant 0. \end{cases}$$

Le problème auxiliaire admet des solutions réalisables puisque la solution $x_0^* = 5$, $x_1^* = x_2^* = x_3^* = 0$ en est une.

Le problème initial admet une solution réalisable si et seulement si le problème auxiliaire admet 0 pour valeur optimale de la fonction objectif.

Si le problème auxiliaire admet 0 comme valeur optimale, toute solution optimale du problème auxiliaire donne une solution réalisable du problème initial, en « oubliant » x_0.

Si le maximum du problème auxiliaire est strictement négatif, le problème initial n'admet aucune solution réalisable.

Un exemple où l'origine n'est pas réalisable, suite

Rappel :

$$
\begin{array}{rcrcrcrcrcr}
x_4 & = & 4 & - & 2x_1 & + & x_2 & - & 2x_3 & + & x_0 \\
x_5 & = & -5 & - & 2x_1 & + & 3x_2 & - & x_3 & + & x_0 \\
x_6 & = & -1 & + & x_1 & - & x_2 & + & 2x_3 & + & x_0 \\
\hline
w & = & & & & & & & & - & x_0
\end{array}
$$

Ce dictionnaire n'est pas réalisable mais on peut se ramener en une itération à un dictionnaire réalisable : on fait entrer x_0 en base et on fait sortir de la base la variable qui est « la plus négative ».

On obtient :

$$
\begin{array}{rlcrcrcrcr}
x_0 & = 5 & + & 2x_1 & - & 3x_2 & + & x_3 & + & x_5 \\
x_4 & = 9 & & & - & 2x_2 & - & x_3 & + & x_5 \\
x_6 & = 4 & + & 3x_1 & - & 4x_2 & + & 3x_3 & + & x_5 \\
\hline
w & = -5 & - & 2x_1 & + & 3x_2 & - & x_3 & - & x_5
\end{array}
$$

Un exemple où l'origine n'est pas réalisable, suite

$$\text{Rappel :}\quad \begin{array}{rrcrcrcrcr} x_0 & = 5 & + & 2x_1 & - & 3x_2 & + & x_3 & + & x_5 \\ x_4 & = 9 & & & - & 2x_2 & - & x_3 & + & x_5 \\ x_6 & = 4 & + & 3x_1 & - & 4x_2 & + & 3x_3 & + & x_5 \\ \hline w & = -5 & - & 2x_1 & + & 3x_2 & - & x_3 & - & x_5 \end{array}$$

x_2 est variable entrante.

$x_0 \geqslant 0$ implique $x_2 \leqslant 5/3$;
$x_4 \geqslant 0$ implique $x_2 \leqslant 9/2$;
$x_6 \geqslant 0$ implique $x_2 \leqslant 1$.

x_6 qui quitte la base. Le dictionnaire devient :

$$\begin{array}{rcrcrcrcrcr} x_2 & = & 1 & + & 0,75x_1 & + & 0,75x_3 & + & 0,25x_5 & - & 0,25x_6 \\ x_0 & = & 2 & - & 0,25x_1 & - & 1,25x_3 & + & 0,25x_5 & + & 0,75x_6 \\ x_4 & = & 7 & - & 1,5x_1 & - & 2,5x_3 & + & 0,5x_5 & + & 0,5x_6 \\ \hline w & = & -2 & + & 0,25x_1 & + & 1,25x_3 & - & 0,25x_5 & - & 0,75x_6 \end{array}$$

Un exemple où l'origine n'est pas réalisable, suite

Rappel :

$$\begin{array}{rcrcrcrcrcr}
x_2 & = & 1 & + & 0,75x_1 & + & 0,75x_3 & + & 0,25x_5 & - & 0,25x_6 \\
x_0 & = & 2 & - & 0,25x_1 & - & 1,25x_3 & + & 0,25x_5 & + & 0,75x_6 \\
x_4 & = & 7 & - & 1,5x_1 & - & 2,5x_3 & + & 0,5x_5 & + & 0,5x_6 \\
\hline
w & = & -2 & + & 0,25x_1 & + & 1,25x_3 & - & 0,25x_5 & - & 0,75x_6
\end{array}$$

x_3 entre en base : x_0 en sort pour donner le dernier dictionnaire.

$$\begin{array}{rlcrcrcrcr}
x_3 & = 1,6 & - & 0,2x_1 & + & 0,2x_5 & + & 0,6x_6 & - & 0,8x_0 \\
x_2 & = 2,2 & + & 0,6x_1 & + & 0,4x_5 & + & 0,2x_6 & - & 0,6x_0 \\
x_4 & = 3 & - & x_1 & & & - & x_6 & + & 2x_0 \\
\hline
w & = & & & & & & & - & x_0
\end{array}$$

Le problème initial avait une solution réalisable donnée par : $x_1^* = 0,\ x_2^* = 2,2,\ x_3^* = 1,6.$

Un dictionnaire pour le problème initial est obtenu en « oubliant » x_0 et en choisissant comme variables en base x_3, x_2, x_4 exprimées ci-dessus en fonction de x_1, x_5, x_6. Il suffit alors d'exprimer z en fonction des mêmes variables.

Un exemple où l'origine n'est pas réalisable, suite

Rappels :

$z = x_1 - x_2 + x_3$

et on a obtenu :

$$
\begin{array}{lllllllllll}
x_3 & = 1,6 & - & 0,2x_1 & + & 0,2x_5 & + & 0,6x_6 & - & 0,8x_0 \\
x_2 & = 2,2 & + & 0,6x_1 & + & 0,4x_5 & + & 0,2x_6 & - & 0,6x_0 \\
x_4 & = 3 & - & x_1 & & & - & x_6 & + & 2x_0 \\
\hline
w & = & & & & & & & - & x_0
\end{array}
$$

En « oubliant » x_0 et en exprimant z à l'aide de x_1, x_5 et x_6 :

$$
\begin{array}{lllllllll}
x_3 & = 1,6 & - & 0,2x_1 & + & 0,2x_5 & + & 0,6x_6 \\
x_2 & = 2,2 & + & 0,6x_1 & + & 0,4x_5 & + & 0,2x_6 \\
x_4 & = 3 & - & x_1 & & & - & x_6 \\
\hline
z & = -0,6 & + & 0,2x_1 & - & 0,2x_5 & + & 0,4x_6
\end{array}
$$

En partant de ce dictionnaire, réalisable, on peut déterminer le maximum de z en appliquant une nouvelle fois l'algorithme du simplexe. Cette méthode s'appelle *méthode à deux phases*.

Sommaire

Optimisation linéaire, l'algorithme du simplexe

Dualité en programmation linéaire

- Définition du problème dual
- Théorème de la dualité
- Le théorème des écarts complémentaires
- La signification économique du dual
- Problème dual-réalisable

Optimisation non linéaire sans contraintes

Optimisation non linéaire avec contraintes

Dualité en programmation linéaire

Définition du problème dual

Théorème de la dualité

Le théorème des écarts complémentaires

La signification économique du dual

Problème dual-réalisable

Introduction

On revient au problème de tissu.

Maximiser $z = 7x_1 + 9x_2 + 18x_3 + 17x_4$

avec les contraintes :

$$\begin{cases} 2x_1 + 4x_2 + 5x_3 + 7x_4 \leqslant 42 \\ x_1 + x_2 + 2x_3 + 2x_4 \leqslant 17 \\ x_1 + 2x_2 + 3x_3 + 3x_4 \leqslant 24 \\ x_1 \geqslant 0, x_2 \geqslant 0, x_3 \geqslant 0, x_4 \geqslant 0. \end{cases}$$

Si on multiplie l'inégalité de la première contrainte par 4, on obtient : $8x_1 + 16x_2 + 20x_3 + 28x_4 \leqslant 168$.

Les variables de décision étant positives ou nulles, on déduit de la comparaison entre les coefficxients des variables dans z et dans cette dernière inégalité : $z \leqslant 168$.

Si on multiplie la première inégalité par 2, la troisième par 3 et qu'on additionne, on obtient : $7x_1 + 14x_2 + 19x_3 + 23x_4 \leqslant 156$.

On déduit : $z \leqslant 156$.

Introduction, suite

Rappel :
$$\text{Maximiser } z = 7x_1 + 9x_2 + 18x_3 + 17x_4$$
$$\text{avec : } \begin{cases} 2x_1 + 4x_2 + 5x_3 + 7x_4 \leqslant 42 \\ x_1 + x_2 + 2x_3 + 2x_4 \leqslant 17 \\ x_1 + 2x_2 + 3x_3 + 3x_4 \leqslant 24 \\ x_1 \geqslant 0, x_2 \geqslant 0, x_3 \geqslant 0, x_4 \geqslant 0. \end{cases}$$

Peut-on obtenir une meilleure majoration du maximum de z avec des combinaisons linéaires des trois premières contraintes ?

On étudie :
$$\text{Minimiser } w = 42y_1 + 17y_2 + 24y_3$$
$$\text{avec : } \begin{cases} 2y_1 + y_2 + y_3 \geqslant 7 \\ 4y_1 + y_2 + 2y_3 \geqslant 9 \\ 5y_1 + 2y_2 + 3y_3 \geqslant 18 \\ 7y_1 + 2y_2 + 3y_3 \geqslant 17 \\ y_1 \geqslant 0, y_2 \geqslant 0, y_3 \geqslant 0. \end{cases}$$

Toute valeur réalisable de w donne un majorant du maximum de z. Si on prend : $y_1^* = 0$, $y_2^* = 3$, $y_3^* = 4$, on obtient une solution réalisable de ce second problème pour lequel $w = 147$. On déduit : $z \leqslant 147$. Or, on a vu que le maximum de z est 147 (obtenu pour $x_1^* = 3$, $x_2^* = 0$, $x_3^* = 7$, $x_4^* = 0$) : on a résolu le second problème.

Introduction, suite

Le second problème, de minimisation, est appelé *problème dual* du premier.

Les valeurs $y_1^* = 0$, $y_2^* = 3$, $y_3^* = 4$ qui ont donné le minimum du problème dual n'ont pas été choisies au hasard ; on verra qu'elles apparaissent dans le dernier dictionnaire de la résolution du problème de tissus (ce sont les opposés des coefficients des trois variables d'écart dans la fonction z).

Définition

On considère le problème (P) mis sous forme standard :

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

avec les contraintes : $\begin{cases} \text{pour } i \in \{1, 2, ..., m\}, \sum_{j=1}^{n} a_{ij} x_j \leqslant b_i \\ \text{pour } j \in \{1, 2, ..., n\}, x_j \geqslant 0 \end{cases}$

Le *problème dual* (D) du problème (P) s'écrit :

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

avec les contraintes : $\begin{cases} \text{pour } j \in \{1, 2, ..., n\}, \sum_{i=1}^{m} a_{ij} y_i \geqslant c_j \\ \text{pour } i \in \{1, 2, ..., m\}, y_i \geqslant 0 \end{cases}$

Le problème (P) prend alors le nom de *problème primal*.

Remarque

On établit facilement que le problème dual de (D) est (P).

Proposition

Pour toute solution réalisable $(x_1^\star, x_2^\star, \ldots, x_n^\star)$ du problème primal et toute solution réalisable $(y_1^\star, y_2^\star, \ldots, y_m^\star)$ du problème dual, en posant : $z^\star = \sum_{j=1}^n c_j x_j^\star$ et $w^\star = \sum_{i=1}^m a_{ij} y_i^\star$, on a :

$$z^\star \leqslant w^\star.$$

De plus, si $z^\star = w^\star$ alors les solutions $(x_1^\star, x_2^\star, \ldots, x_n^\star)$ et $(y_1^\star, y_2^\star, \ldots, y_m^\star)$ sont optimales respectivement pour le problème primal et le problème dual.

Preuve :

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

Retour sur l'exemple

$$(P)\quad \text{Maximiser } z = 7x_1 + 9x_2 + 18x_3 + 17x_4$$
$$\text{avec : } \begin{cases} 2x_1 + 4x_2 + 5x_3 + 7x_4 \leqslant 42 \\ x_1 + x_2 + 2x_3 + 2x_4 \leqslant 17 \\ x_1 + 2x_2 + 3x_3 + 3x_4 \leqslant 24 \\ x_1 \geqslant 0, x_2 \geqslant 0, x_3 \geqslant 0, x_4 \geqslant 0. \end{cases}$$

$$(D)\quad \text{Minimiser } w = 42y_1 + 17y_2 + 24y_3$$
$$\text{avec : } \begin{cases} 2y_1 + y_2 + y_3 \geqslant 7 \\ 4y_1 + y_2 + 2y_3 \geqslant 9 \\ 5y_1 + 2y_2 + 3y_3 \geqslant 18 \\ 7y_1 + 2y_2 + 3y_3 \geqslant 17 \\ y_1 \geqslant 0, y_2 \geqslant 0, y_3 \geqslant 0. \end{cases}$$

- $y_1^\star = 0,\ y_2^\star = 2,\ y_3^\star = 3$, solution réalisable du dual $\to w^* = 147$,
- $x_1^\star = 3,\ x_2^\star = 0,\ x_3^\star = 7,\ x_4^\star = 0$, solution réalisable du primal $\to z^* = 147$.

La proposition montre que ces deux solutions sont optimales respectivement pour le dual et le primal.

Corollaires

Si le problème primal admet une solution réalisable et est non borné, le problème dual n'admet pas de solution réalisable.

Si le problème dual admet une solution réalisable et est non borné, le problème primal n'admet pas de solution réalisable.

(P) et (D) ne peuvent pas être simultanément non bornés.

Remarque : il existe des cas pour lesquels (P) et (D) sont simultanément non réalisables.

Dualité en programmation linéaire

Définition du problème dual
Théorème de la dualité
Le théorème des écarts complémentaires
La signification économique du dual
Problème dual-réalisable

Théorème de la dualité et proposition

Théorème :
Si le problème primal a une solution optimale $x_1^\star, x_2^\star, \ldots, x_n^\star$, alors le problème dual a une solution optimale $y_1^\star, y_2^\star, \ldots, y_m^\star$ et

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

(autrement dit, le maximum primal est égal au minimum dual).

Proposition :
Si le problème primal admet une solution optimale et si l'expression de la fonction objectif du primal dans le dernier dictionnaire obtenu par la méthode du simplexe s'écrit :

$$z = z^* + \sum_{k=1}^{n+m} d_k x_k$$

(où x_{n+i} représente la i^{e} variable d'écart), alors une solution optimale du problème dual est donnée par $y_i^* = -d_{n+i}$.

Preuve du théorème de la dualité

On suppose le primal résolu par la méthode du simplexe. Aux n variables initiales du problème on a ajouté m variables d'écart $x_{n+1}, \ldots, x_{n+m}$.

À la ie contrainte du primal sont associées la variable d'écart x_{n+i} et la variable y_i du dual, ce qui établit un lien canonique entre x_{n+i} et y_i.

Dans le dernier dictionnaire du simplexe primal :

$$z = z^* + \sum_{k=1}^{n+m} d_k x_k.$$

Les d_k sont tous négatifs ou nuls (puisqu'il s'agit du dernier dictionnaire) et les d_k associés aux variables en base sont nuls.

$x_{n+i} = b_i - \sum_{j=1}^{n} a_{ij} x_j$ par définition des variables d'écart.

Preuve du théorème de la dualité, suite

On rappelle : $z = z^* + \sum_{k=1}^{n+m} d_k x_k$ et $x_{n+i} = b_i - \sum_{j=1}^{n} a_{ij} x_j$.
On pose, pour $i \in \{1, ..., m\}$, $y_i^* = -d_{n+i}$; on a alors : $y_i^* \geqslant 0$.

En distinguant dans z les variables d'écart des autres :
$z = z^* + \sum_{j=1}^{n} d_j x_j - \sum_{i=1}^{m} \left(b_i - \sum_{j=1}^{n} a_{ij} x_j \right) y_i^*$, ou encore :
$z = z^* - \sum_{i=1}^{m} b_i y_i^* + \sum_{j=1}^{n} (d_j + \sum_{i=1}^{m} a_{ij} y_i^*) x_j$.

Par définition de z, on a aussi : $z = \sum_{j=1}^{n} c_j x_j$.

On déduit de ces égalités :

$$\begin{cases} z^* = \sum_{i=1}^{m} b_i y_i^* \\ \text{pour } j \in \{1, ..., n\}, c_j = d_j + \sum_{i=1}^{m} a_{ij} y_i^* \end{cases}$$

Les d_k ($k \in \{1, ..., n+m\}$) étant négatifs ou nuls, on obtient finalement :

$$\begin{cases} \text{pour } j \in \{1, ..., n\}, \sum_{i=1}^{m} a_{ij} y_i^* \geqslant c_j \\ \text{pour } i \in \{1, ..., m\}, y_i^* \geqslant 0. \end{cases}$$

Donc $y_1^*, y_2^*, ..., y_m^*$ forment une solution réalisable du dual qui donne à la fonction objectif du dual la valeur z^*. D'où le théorème.

Dualité en programmation linéaire

Définition du problème dual
Théorème de la dualité
Le théorème des écarts complémentaires
La signification économique du dual
Problème dual-réalisable

Théorème des écarts complémentaires

Ce théorème permet de prouver qu'une solution pour le problème primal donne la solution optimale.

Théorème des écarts complémentaires

Une solution réalisable $x_1^\star, \dots, x_n^\star$ du primal est optimale si et seulement s'il existe des nombres $y_1^\star, \dots, y_m^\star$ vérifiant ce qui suit :

- pour $i \in \{1, \dots, m\}$, si $\sum_{j=1}^{n} a_{ij} x_j^\star < b_i$, alors $y_i^\star = 0$
- pour $j \in \{1, \dots, n\}$, si $x_j^\star > 0$, alors $\sum_{i=1}^{m} a_{ij} y_i^\star = c_j$

et constituant une solution réalisable du problème dual.

Si les conditions sont vérifiées, les nombres $y_1^\star, \dots, y_m^\star$ constituent une solution optimale du problème dual.

La preuve découle de la preuve de la première proposition sur la dualité donnée auparavant.

Application

$$(P)\quad \begin{array}{l} \text{Maximiser } z = 7x_1 + 9x_2 + 18x_3 + 17x_4 \\ \text{avec :} \left\{ \begin{array}{l} 2x_1 + 4x_2 + 5x_3 + 7x_4 \leqslant 42 \\ x_1 + x_2 + 2x_3 + 2x_4 \leqslant 17 \\ x_1 + 2x_2 + 3x_3 + 3x_4 \leqslant 24 \\ x_1 \geqslant 0, x_2 \geqslant 0, x_3 \geqslant 0, x_4 \geqslant 0. \end{array} \right. \end{array}$$

Considérons la déclaration : « $x_1^\star = 3, x_2^\star = 0, x_3^\star = 7, x_4^\star = 0$ donnent une solution optimale de (P) ».
Ces valeurs définissent une solution réalisable du problème primal.

Cherchons s'il existe $y_1^\star, y_2^\star, y_3^\star$ vérifiant :

$$\left\{ \begin{array}{l} y_1^\star = 0 \text{ puisque première contrainte du problème « non saturée »} \\ 2y_1^\star + y_2^\star + y_3^\star = 7 \text{ puisque } x_1^\star > 0 \\ 5y_1^\star + 2y_2^\star + 3y_3^\star = 18 \text{ puisque } x_3^\star > 0. \end{array} \right.$$

On obtient : $\left\{ \begin{array}{l} y_2^\star + y_3^\star = 7 \\ 2y_2^\star + 3y_3^\star = 18. \end{array} \right.$ qui donne $y_2^\star = 3, y_3^\star = 4$.

Ces valeurs satisfont les inégalités du problème dual.
La solution proposée pour le primal est donc bien optimale et $y_1^\star, y_2^\star, y_3^\star$ donnent le maximum primal à la fonction objectif duale.

Dualité en programmation linéaire

Définition du problème dual
Théorème de la dualité
Le théorème des écarts complémentaires
La signification économique du dual
Problème dual-réalisable

Interprétations pour le problème primal

- b_i représente la quantité totale de la ressource i ;
- a_{ij} représente la quantité de la ressource i consommée par la fabrication d'une unité de produit j ;
- x_j représente la quantité fabriquée de produit j ;
- c_j représente la valeur unitaire du produit j.

La relation à l'optimum : $z^* = \sum_{j=1}^{n} c_j x_j^* = \sum_{i=1}^{m} b_i y_i^\star$ induit que y_i doit représenter la « valeur unitaire de la ressource i ». Ces variables duales y_i sont souvent appelées *prix implicites*.

La valeur de y_i donne le montant maximum que l'on serait prêt à payer pour obtenir une unité supplémentaire de la ressource i et le prix minimum auquel on pourrait vendre une unité de la ressource i.

Interprétations pour le problème dual

Comme a_{ij} représente la quantité de la ressource i requise pour fabriquer une unité de produit j, $\sum_{i=1}^{m} a_{ij}y_i$ représente la somme à dépenser pour acquérir les ressources nécessaires à la fabrication d'une unité du produit j.

On suppose qu'une personne étrangère à l'entreprise souhaite acquérir les ressources de l'entreprise :
• les relations $\sum_{i=1}^{m} a_{ij}y_i \geqslant c_j$ $(j \in \{1, ..., n\})$ expriment que cet acheteur doit proposer pour les ressources un prix tel que ce soit plus intéressant pour l'entreprise de lui vendre ses ressources que de fabriquer elle-même les produits
• la minimisation de $\sum_{i=1}^{m} b_i y_i$ exprime que l'acheteur désire faire cet achat des ressources à un prix minimum.

Théorème

On considère le problème (P) :

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

avec les contraintes : $\begin{cases} \text{pour } i \in \{1, ..., m\}, \sum_{j=1}^{n} a_{ij} x_j \leqslant b_i \\ \text{pour } j \in \{1, 2, ..., n\}, x_j \geqslant 0. \end{cases}$

On suppose que la base optimale de (P) est non dégénérée. Pour des variations δb_i des b_i, on considère le problème (P_δ) défini par :

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

avec les contraintes : $\begin{cases} \text{pour } i \in \{1, ..., m\}, \sum_{j=1}^{n} a_{ij} x_j \leqslant b_i + \delta b_i \\ \text{pour } j \in \{1, 2, ..., n\}, x_j \geqslant 0. \end{cases}$

On suppose que les variations δb_i sont suffisamment faibles pour que la base optimale de (P) soit encore réalisable pour (P_δ). La variation de la valeur optimum de la fonction objectif du programme linéaire vaut alors $\sum_{i=1}^{m} \delta b_i y_i^*$ où $(y_1^*, ..., y_m^*)$ est solution optimale du problème dual de (P).

Preuve du théorème

On considère la suite des dictionnaires obtenus lorsqu'on résout (P) par la méthode du simplexe. Quand on change b pour $b + \delta b$, seules les constantes des seconds membres sont changées.

L'hypothèse selon laquelle le dernier dictionnaire reste réalisable entraîne que ce dernier dictionnaire reste optimal.

Les coefficients des variables hors-bases dans la ligne de z étant inchangés, la solution du problème dual est inchangée (et la non dégénérescence de la base optimale de (P) entraîne l'unicité de cette solution optimale du problème dual).

Preuve du théorème, suite

La valeur optimale commune aux nouveaux problèmes primal et dual vaut :

$$\sum_{i=1}^{m}(b_i + \delta b_i)y_i^\star = \sum_{i=1}^{m} b_i y_i^\star + \sum_{i=1}^{m} \delta b_i y_i^\star$$

où $(y_1^\star, ..., y_m^\star)$ est solution optimale du problème dual de P.

La variation de la valeur optimum de la fonction objectif vaut $\sum_{i=1}^{m} \delta b_i y_i^\star$.

Remarque : si la base optimale de (P) est non dégénérée, pour des raisons de continuité, il existe des variations non nulles des δb_i assez petites pour conserver le fait que la base optimale de (P) reste réalisable.

Application

Le fabricant de tissus de notre exemple a la possibilité de faire faire à ses ouvriers spécialisés dans la teinture quelques heures supplémentaires à un prix horaire de t euros. A-t-il ou non intérêt à utiliser cette possibilité ?

La contrainte sur le nombre d'heure de teinture est la troisième.

La solution optimale du problème dual est $(0, 3, 4)$.

u : nombre d'heures supplémentaires pour la teinture (avec u petit).

La variation du second membre est $(0, 0, u)$. La variation de la fonction objectif est égale à $4u$ (euros) mais l'entreprise doit payer $t.u$ euros.

Le fabricant a intérêt à recourir à cette solution dès que $t \leqslant 4$ euros. On retrouve là l'interprétation de $y_3^\star$.

Dualité en programmation linéaire

Définition du problème dual
Théorème de la dualité
Le théorème des écarts complémentaires
La signification économique du dual
Problème dual-réalisable

Utilisation du problème dual

L'utilisation du problème dual permet, sans utiliser l'algorithme à deux phases décrit dans le premier chapitre, de résoudre un problème de programmation linéaire où la solution nulle n'est pas réalisable, pourvu que les coefficients c_j de la fonction objectif du problème écrit sous forme standard soient tous négatifs ou nuls. Un tel problème est dit *dual-réalisable*.

Exemple

Minimiser $x_1 + x_2$

$$\text{avec les contraintes : } \begin{cases} 3x_1 + x_2 \geqslant 4 \\ -7x_1 + x_2 \geqslant -7 \\ x_1 \geqslant 0, x_2 \geqslant 0 \end{cases}$$

dont l'écriture, sous forme standard est :

Maximiser $-x_1 - x_2$

$$\text{avec les contraintes : } \begin{cases} -3x_1 - x_2 \leqslant -4 \\ 7x_1 - x_2 \leqslant 7 \\ x_1 \geqslant 0, x_2 \geqslant 0. \end{cases}$$

Exemple de problème dual-réalisable

Rappel :
$$\text{Maximiser } -x_1 - x_2$$
$$\text{avec les contraintes : } \begin{cases} -3x_1 - x_2 \leqslant -4 \\ 7x_1 - x_2 \leqslant 7 \\ x_1 \geqslant 0, x_2 \geqslant 0. \end{cases}$$

Le problème dual s'écrit :

$$\text{Minimiser } -4y_1 + 7y_2$$
$$\text{avec les contraintes : } \begin{cases} -3y_1 + 7y_2 \geqslant -1 \\ -y_1 - y_2 \geqslant -1 \\ y_1 \geqslant 0, y_2 \geqslant 0. \end{cases}$$

ou encore :
$$\text{Maximiser } 4y_1 - 7y_2$$
$$\text{avec les contraintes : } \begin{cases} 3y_1 - 7y_2 \leqslant 1 \\ y_1 + y_2 \leqslant 1 \\ y_1 \geqslant 0, y_2 \geqslant 0 \end{cases}$$

Les variables d'écarts constituent une base réalisable : la méthode du simplexe ne nécessite qu'une seule phase.

Bases optimales des problèmes primal et dual

Résultats admis.

Une base optimale du problème dual est formée :
• des variables y_i $(1 \leqslant i \leqslant m)$ pour i tel que x_{n+i} soit hors-base
• des variables y_{m+j} $(1 \leqslant j \leqslant n)$ pour j tel que x_j soit hors-base

Une base optimale du problème primal est formée :
• des variables x_j $(1 \leqslant j \leqslant n)$ pour j tel que y_{m+j} soit hors-base
• des variables x_{n+i} $(1 \leqslant i \leqslant m)$ pour i tel que y_i soit hors-base

Sommaire

Optimisation linéaire, l'algorithme du simplexe

Dualité en programmation linéaire

Optimisation non linéaire sans contraintes

- Généralités
- Condition nécessaire et condition suffisante d'optimalité locale
- Fonctions quadratiques
- Fonctions convexes
- Généralités sur les méthodes d'optimisation sans contrainte
- Méthodes de gradient
- Méthode de Newton
- Optimisation unidimensionnelle

Généralités

On considère des fonctions f définies dans $\mathbb{R}^n$ et à valeurs dans $\mathbb{R}$.

On cherche les points où f atteint des extrema locaux ou globaux.

Si x est un élément de $\mathbb{R}^n$, on pose $x = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}$.

Définition du gradient

$f : \mathbb{R}^n \to \mathbb{R}$. On suppose que f admet en un point x des dérivées partielles du premier ordre. Alors :

$$\textit{gradient de } f = \nabla f(x) = \begin{pmatrix} \dfrac{\partial f}{\partial x_1}(x) \\ \dots \\ \dfrac{\partial f}{\partial x_n}(x) \end{pmatrix}.$$

Généralités, suite

Définition du gradient d'un vecteur-ligne de fonctions

$F(x) = (f_1(x), \ldots, f_p(x))$, vecteur-ligne, avec $f_j : \mathbb{R}^n \to \mathbb{R}$ dérivable au point x.
$\nabla F(x)$ est la matrice dont la j^{e} colonne est $\nabla f_j(x)$.

A, matrice carrée constante d'ordre n, $u(x)$ et $v(x)$ deux vecteurs-colonnes :

$$\nabla\left(u^{\mathrm{t}}.A\right) = \nabla\left(u^{\mathrm{t}}\right).A \text{ et } \nabla\left(A.u\right) = A.\nabla\left(u\right).$$

$$\nabla\left(u^{\mathrm{t}}.v\right) = \nabla\left(u^{\mathrm{t}}\right).v + \nabla\left(v^{\mathrm{t}}\right).u$$

Si f admet en x^0 des dérivées partielles continues, formule de Taylor à l'ordre 1 :

$$\boxed{f(x) = f(x^0) + (x - x^0)^{\mathrm{t}}.\nabla f(x^0) + ||x - x^0||.\epsilon(x)}$$

où $\epsilon(x)$ est une fonction qui tend vers 0 quand x tend vers x^0.

Généralités, suite

Remarques

• Si f de classe C^1 :

$x_{n+1} = f(x^0) + (x - x^0)^{\mathrm{t}}.\nabla f(x^0)$ est l'équation de l'hyperplan tangent à la surface S de $\mathbb{R}^{n+1}$ d'équation $x_{n+1} = f(x_1, ..., x_n)$ au point $(x^0, f(x^0))$.

• Variations de f dans une direction d de $\mathbb{R}^n$ donnée, à partir d'un point x^0 de $\mathbb{R}^n$ donné :

pour $s \in \mathbb{R}$, on pose $g(s) = f(x^0 + s.d)$.

On a :
$g'(s) = d^{\mathrm{t}}.\nabla f(x^0 + s.d)$ et $g'(0) = d^{\mathrm{t}}.\nabla f(x^0)$.

Matrice hessienne

Définition de la matrice hessienne

Si f admet des dérivées partielles d'ordre 2 en x :

$$\nabla^2 f(x) = \nabla\left(\nabla f(x)^{\mathrm{t}}\right) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \dots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x) \\ \dots & \dots & \dots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \dots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{pmatrix};$$

$\nabla^2 f$ s'appelle la *matrice hessienne* de f.

Si f est une fonction de classe C^2 (autrement dit, f admet des dérivées partielles à l'ordre 2 continues), la matrice hessienne de f est une matrice symétrique (théorème de Schwarz).

Formule de Taylor d'ordre 2

Si f est de classe C^2 en x^0, formule de Taylor d'ordre 2 :

$$\boxed{\begin{gathered} f(x) = f(x^0) + (x - x^0)^{\mathrm{t}}.\nabla f(x^0) + \frac{1}{2}(x - x^0)^{\mathrm{t}}\nabla^2 f(x^0).(x - x^0) \\ + ||x - x^0||^2.\epsilon(x) \end{gathered}}$$

où $\epsilon(x)$ est une fonction qui tend vers 0 quand x tend vers x^0.

Optimisation non linéaire sans contraintes

Généralités
Condition nécessaire et condition suffisante d'optimalité locale
Fonctions quadratiques
Fonctions convexes
Généralités sur les méthodes d'optimisation sans contrainte
Méthodes de gradient
Méthode de Newton
Optimisation unidimensionnelle

Rappels

Définitions

Soit A une matrice réelle symétrique d'ordre n.

A est positive $\Leftrightarrow \forall h \in \mathbb{R}^n, h^{\mathrm{t}}.A.h \geqslant 0$.

A est définie positive $\Leftrightarrow \forall h \in \mathbb{R}^n \setminus \{0\}, h^{\mathrm{t}}.A.h > 0$.

Une matrice symétrique réelle est diagonalisable sur $\mathbb{R}$.

La matrice A est positive $\Leftrightarrow$ les valeurs propres de A sont positives ou nulles.

La matrice A est définie positive $\Leftrightarrow$ les valeurs propres de A sont strictement positives.

Condition nécessaire d'optimalité

On suppose désormais que f est de classe C^2.

Théorème

Si f admet un minimum local en x^*, alors : 1) $\nabla f(x^*) = 0$
2) $\nabla^2 f(x^*)$ est une matrice positive.

Preuve de 1). $f(x) = f(x^*) + (x - x^*)^t.\nabla f(x^*) + ||x - x^*||.\epsilon(x)$, où $\epsilon(x)$ est une fonction qui tend vers 0 quand x tend vers x^*.

Avec $x = x^* - \theta.\nabla f(x^*)$, avec $\theta \in \mathbb{R}^+$:

$$f(x) - f(x^*) = -\theta.||\nabla f(x^*)||^2 + \theta.\epsilon_1(\theta)$$

$$f(x) - f(x^*) = \theta\left(-||\nabla f(x^*)||^2 + \epsilon_1(\theta)\right),$$

où $\epsilon_1(\theta)$ est une fonction qui tend vers 0 quand θ tend vers 0.

Si $\nabla f(x^*) \neq 0$: contradiction avec la minimalité en x^*.

Preuve de 2)

On suppose qu'il existe $h \in \mathbb{R}^n$ tel qu'on ait : $h^{\mathrm{t}}.\nabla^2 f(x^*).h < 0$.

D'après le développement de Taylor d'ordre 2 :

$$f(x^* + \theta h) - f(x^*) = \theta^2. \left(\frac{1}{2} h^t.\nabla^2 f(x^*).h + \epsilon_2(\theta) \right),$$

où $\epsilon_2(\theta)$ est une fonction qui tend vers 0 quand θ tend vers 0.

Pour θ assez petit, la différence $f(x^* + \theta h) - f(x^*)$ serait négative : contradiction.

Condition suffisante d'optimalité

Théorème

Si une fonction f vérifie en x^* :
- $\nabla f(x^*) = 0$,
- $\nabla^2 f(x^*)$ est une matrice définie positive,

alors f admet un minimum local en x^*.

Preuve.
$S =$ sphère de centre 0 et de rayon 1.
$a = \inf\{h^{\mathrm{t}}.\nabla^2 f(x^*).h \text{ pour } h \in S\}$:

$$\forall h \in \mathbb{R}^n, h^t.\nabla^2 f(x^*).h \geqslant a||h||^2.$$

S est un compact $\Rightarrow$ la valeur a est atteinte et $\nabla^2 f(x^*)$ définie positive $\Rightarrow a > 0$.
Formule de Taylor à l'ordre 2 :

$$f(x^*+h)-f(x^*) = \frac{1}{2}h^t.\nabla^2 f(x^*).h+||h||^2\epsilon(h) \geqslant ||h||^2.\left(\frac{a}{2}+\epsilon(h)\right),$$

où $\epsilon(h)$ tend vers 0 quand h tend vers 0.
Ce qui montre le théorème.

Optimisation non linéaire sans contraintes

Généralités
Condition nécessaire et condition suffisante d'optimalité locale
Fonctions quadratiques
Fonctions convexes
Généralités sur les méthodes d'optimisation sans contrainte
Méthodes de gradient
Méthode de Newton
Optimisation unidimensionnelle

Fonctions quadratiques

A : matrice symétrique d'ordre n, b : vecteur-colonne d'ordre n, $c \in \mathbb{R}$.

Définition

L'application q de $\mathbb{R}^n$ dans $\mathbb{R}$ définie par :

$$q(x) = c + b^{\mathrm{t}}.x + \frac{1}{2}x^{\mathrm{t}}.A.x$$

s'appelle *fonction quadratique*.

Remarque. La partie polynomiale du développement de Taylor d'ordre 2 d'une fonction f est la fonction quadratique q telle que la surface d'équation $x_{n+1} = q(x)$ soit « la plus proche » de la surface d'équation $x_{n+1} = f(x)$ au voisinage du point considéré.

Dérivation

$$q(x) = c + b^{\mathrm{t}}.x + \frac{1}{2}x^{\mathrm{t}}.A.x$$

$$\nabla q(x) = \nabla(x^{\mathrm{t}}).b + \frac{1}{2}[\nabla(x^{\mathrm{t}}).A.x + \nabla((A.x)^{\mathrm{t}}).x].$$

$\nabla(x^{\mathrm{t}})$ est la matrice identité.

$\nabla(A.x)^{\mathrm{t}} = \nabla(x^{\mathrm{t}}.A^{\mathrm{t}}) = \nabla(x^{\mathrm{t}}).A^{\mathrm{t}} = A^{\mathrm{t}} = A.$

D'où : $\boxed{\nabla q(x) = b + A.x}$.

De plus : $\nabla^2 q(x) = \nabla((\nabla q(x))^{\mathrm{t}}) = \nabla(b^{\mathrm{t}} + x^{\mathrm{t}}.A^{\mathrm{t}}) = A^{\mathrm{t}}.$

On a donc : $\boxed{\nabla^2 q(x) = A}$.

Les dérivées d'ordre au moins 3 de q sont nulles.

Une fonction quadratique coïncide avec son développement de Taylor à l'ordre 2.

Optimisation non linéaire sans contraintes

Généralités
Condition nécessaire et condition suffisante d'optimalité locale
Fonctions quadratiques
Fonctions convexes
Généralités sur les méthodes d'optimisation sans contrainte
Méthodes de gradient
Méthode de Newton
Optimisation unidimensionnelle

Fonctions convexes

Définition

On dit qu'une fonction f de $\mathbb{R}^n$ dans $\mathbb{R}$ est convexe si, pour tout x et tout y de $\mathbb{R}^n$ et pour tout λ de $]0,1[$, on a :
$f(\lambda x + (1-\lambda)y) \leqslant \lambda f(x) + (1-\lambda) f(y)$.

Si cette inégalité est stricte, on dit que f est strictement convexe.

Théorème

Théorème

Si f est une fonction convexe et admet des dérivées partielles, alors f admet un minimum global en x^* si et seulement si on a la relation $\nabla f(x^*) = 0$.

Preuve. x^* est un minimum local $\Rightarrow \nabla f(x^*) = 0$.

Réciproque : on suppose $\nabla f(x^*) = 0$.

Soit $x \in \mathbb{R}^n$. On pose, pour $s \in \mathbb{R}$: $g(s) = f(x^* + s(x - x^*))$.

$g(0) = f(x^*)$, $g(1) = f(x)$, $g'(0) = (x - x^*)^t . \nabla f(x^*) = 0$.

g est une fonction convexe de $\mathbb{R}$ dans $\mathbb{R}$; sa dérivée est croissante : pour $s \geqslant 0, g'(s) \geqslant 0$; g est croissante pour $s \geqslant 0$.

$g(1) \geqslant g(0)$: f admet bien un minimum global en x^*.

Théorème

Théorème

Si f est convexe et admet un minimum local en x^*, alors f admet un minimum global en x^*.

Preuve.

Si f admet un minimum local en x^*, alors $\nabla f(x^*) = 0$.

Si de plus f est convexe, le théorème précédent montre qu'elle admet un minimum global en x^*.

Théorème

Théorème admis

Si f est deux fois continûment dérivable, les propositions suivantes sont équivalentes :

a. f est convexe.

b. Pour tout x et tout x^0 de $\mathbb{R}^n$,
$f(x) \geqslant f(x^0) + (\nabla f(x^0))^{\mathrm{t}}.(x - x^0)$ (la surface de $\mathbb{R}^{n+1}$d'équation $x_{n+1} = f(x)$ est au-dessus de ses hyperplans tangents).

c. Pour tout x, $\nabla^2 f(x)$ est positive.

Une fonction quadratique $q(x) = \dfrac{1}{2}x^{\mathrm{t}}.A.x + b^{\mathrm{t}}.x + c$ est convexe si et seulement si A est positive.

Si A est définie positive, alors q est strictement convexe et admet un minimum global unique.

Optimisation non linéaire sans contraintes

Généralités
Condition nécessaire et condition suffisante d'optimalité locale
Fonctions quadratiques
Fonctions convexes
Généralités sur les méthodes d'optimisation sans contrainte
Méthodes de gradient
Méthode de Newton
Optimisation unidimensionnelle

Généralités sur les méthodes d'optimisation sans contrainte

Même si on s'intéresse le plus souvent à des extrema globaux, on cherche en général des extrema locaux.

On indique des méthodes qui concernent des minima.

Pour déterminer un point où une fonction f atteint un minimum local, les méthodes consistent très souvent à construire une suite $x^0, x^1, ..., x^k, ...$ qui doit converger vers un point x^* vérifiant une condition nécessaire d'optimalité, par exemple $\nabla f(x^*) = 0$.

Méthodes de descente

On appelle *méthode de descente* toute méthode où, à chaque étape, on pose $x^{k+1} = x^k + s_k d^k$, où $s_k \in \mathbb{R}^+$ et d^k est une direction de $\mathbb{R}^n$ qui vérifie $(d^k)^{\mathrm{t}}.\nabla f(x^k) < 0$.

$(d^k)^{\mathrm{t}}.\nabla f(x^k) < 0$ signifie que $f(x^k + sd^k)$ a une dérivée négative pour $s = 0$: partant de x^k dans la direction d^k, f décroît (« on descend »).

La différence entre les diverses méthodes de descente porte sur le choix de s_k et de d^k, choix qui doit au minimum assurer $f(x^{k+1}) \leqslant f(x^k)$.

Vitesse de convergence

Lorsque la convergence d'un algorithme a été établie, une qualité importante de cet algorithme est sa *vitesse de convergence*.

• Si on a $\dfrac{||x^{k+1} - x^*||}{||x^k - x^*||} \leqslant \alpha < 1$ pour k assez grand, on dit que la convergence est *linéaire* de taux α.

• Si $\dfrac{||x^{k+1} - x^*||}{||x^k - x^*||}$ tend vers 0 quand k tend vers l'infini, on dit que la *convergence* est *superlinéaire*.

• Si $\dfrac{||x^{k+1} - x^*||}{||x^k - x^*||^\gamma}$ est borné, avec $\gamma > 1$, on dit que la *convergence* est *superlinéaire* d'ordre γ.

Dans le cas $\gamma = 2$, on dit que la convergence est *quadratique*.

Optimisation non linéaire sans contraintes

Généralités
Condition nécessaire et condition suffisante d'optimalité locale
Fonctions quadratiques
Fonctions convexes
Généralités sur les méthodes d'optimisation sans contrainte
Méthodes de gradient
Méthode de Newton
Optimisation unidimensionnelle

Méthodes de gradient, principe

Famille de méthodes itératives pour des fonctions dérivables.

Soient d un vecteur de $\mathbb{R}^n$ et x^k un point de $\mathbb{R}^n$ avec $\nabla f(x^k) \neq 0$.

On pose, pour $s \in \mathbb{R}$: $g(s) = f(x^k + sd)$.

On dit que d est une *direction de descente* si $g'(0) < 0$.

$g'(0) = d^{\mathrm{t}}.\nabla f(x^k)$. En notant θ l'angle entre $\nabla f(x^k)$ et d :

$g'(0) = ||\nabla f(x^k)||.||d||.\cos\theta$.

En supposant d unitaire, $g'(0)$ est minimum si $\cos\theta = -1$, *i.e.* si d est donnée par l'opposé du gradient : $d = -\dfrac{\nabla f(x^k)}{||\nabla f(x^k)||}$.

Cette dernière direction est la *direction de plus grande pente descendante*. C'est ce choix qui est fait dans les méthodes de gradient.

Méthode de la plus forte pente à pas optimal

L'algorithme de la plus forte pente peut s'écrire de la façon suivante :

- Choisir un point de départ x^0 ;
- $k \leftarrow 0]$
- répéter
    - $\star$ $d^k \leftarrow -\nabla f(x^k)$
    - $\star$ déterminer s_k vérifiant $f(x^k + s_k d^k) = \mathrm{Min}_{s \geqslant 0} f(x^k + sd^k)$
    - $\star$ $x^{k+1} \leftarrow x^k + s_k d^k$
    - $\star$ $k \leftarrow k + 1$

tant qu'un test d'arrêt donné n'est pas vérifié.

Test d'arrêt de la méthode de la plus forte pente à pas optimal

Le test d'arrêt peut être par exemple :

• on a épuisé un nombre d'itérations fixé à l'avance ;

• le gradient est très petit : $\sum_{i=1}^{n}(\frac{\partial f}{\partial x_i}(x^k))^2 \leqslant \epsilon$, où ϵ est un paramètre donné ;

• la suite x^k est « presque » stationnaire : $|f(x^{k+1}) - f(x^k)| \leqslant \epsilon$, où ϵ est un paramètre donné.

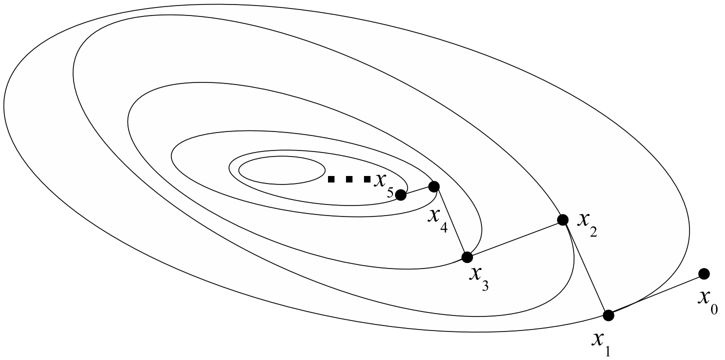
On peut montrer que si $f(x)$ est une fonction de classe C^1 qui tend vers l'infini quand $||x||$ tend vers l'infini, cet algorithme converge vers un point stationnaire (point où le gradient s'annule).

Vitesse de convergence

La vitesse de convergence peut être très faible (linéaire avec un taux proche de 1).

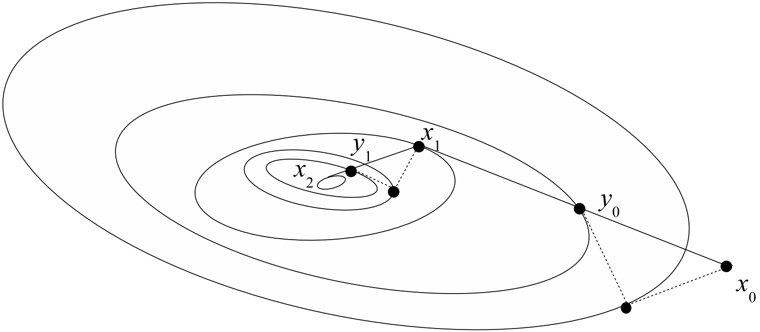
$\frac{d}{ds}[f(x^k - s.\nabla f(x^k))](s_k) = 0$ s'écrit :

$[\nabla f(x^k)]^t.\nabla f(x^{k+1}) = 0$; les directions de déplacement successives sont orthogonales. Il y a convergence en zig-zag.



Méthode de la plus forte pente accélérée

Méthode de descente fondée sur la méthode de la plus forte pente.
Soit p un entier fixé. À partir d'un point x^k, on effectue p itérations de la méthode de la plus forte pente $\rightarrow$ un point y^k.
On pose $d^k = y^k - x^k$.
x^{k+1} = un point où $f(x^k + sd^k)$ admet un minimum pour $s > 0$.
Ci-dessous, $p = 2$.



Optimisation non linéaire sans contraintes

Généralités
Condition nécessaire et condition suffisante d'optimalité locale
Fonctions quadratiques
Fonctions convexes
Généralités sur les méthodes d'optimisation sans contrainte
Méthodes de gradient
Méthode de Newton
Optimisation unidimensionnelle

Méthode de Newton, non prouvée

On suppose f de classe C^2.
Au voisinage d'un point x^k, on approche f par la fonction quadratique donnée par la formule de Taylor d'ordre 2 :
$q(x) = f(x^k) + (x - x^k)^t.\nabla f(x^k) + \dfrac{1}{2}(x - x^k)^t \nabla^2 f(x^k)(x - x^k)$.

Soit x^{k+1} le point, s'il existe, qui minimise q.
Si $\nabla^2 f(x^k)$ est définie positive, on détermine x^{k+1} par l'équation
$\nabla q(x) = 0$ *i.e.* $\nabla f(x^k) + \nabla^2 f(x^k).(x - x^k) = 0$:
$x^{k+1} = x^k - [\nabla^2 f(x^k)]^{-1}.\nabla f(x^k)$.

Proposition

On suppose la fonction f de classe C^3. Si x^0 est choisi suffisamment proche d'un minimum local x^* où la matrice hessienne de f est définie positive, alors la suite (x^k) a une convergence quadratique vers x^*.

Généralités
Condition nécessaire et condition suffisante d'optimalité locale
Fonctions quadratiques
Fonctions convexes
Généralités sur les méthodes d'optimisation sans contrainte
Méthodes de gradient
Méthode de Newton
Optimisation unidimensionnelle

Les méthodes d'optimisation unidimensionnelle ont d'autant plus d'importance qu'elles servent d'outils pour l'optimisation de fonctions de plusieurs variables.

Méthode de Newton, cas unidimensionnel

Soit $f : \mathbb{R} \mapsto \mathbb{R}$.

À partir d'un point x_k, on approche f par :

$$q(x) = f(x_k) + f'(x_k)(x - x_k) + \frac{1}{2} f''(x_k)(x - x_k)^2.$$

On remarque : $q'(x) = f'(x_k) + f''(x_k)(x - x_k)$.

Si $f''(x_k) > 0$ (cas où f est strictement convexe autour de x_k) :

$$\boxed{x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}}$$

qui est le point où q atteint son minimum ($q'(x_{k+1}) = 0$).

Si $f''(x_k) \leqslant 0$, la méthode échoue.

Dichotomie pour une fonction dérivable

Définition

On dit qu'une fonction est *unimodale* s'il existe un réel x^* pour lequel la fonction est strictement décroissante sur $]-\infty, x^*]$ et strictement croissante sur $[x^*, +\infty[$.

x^* est alors minimum global de f.

Soit f dérivable et unimodale. On cherche x^* qui vérifie $f'(x^*) = 0$.

1) Recherche de x_{min} et x_{max} avec $f'(x_{min}) < 0$ et $f'(x_{max}) > 0$.

2)
- $x \leftarrow \frac{1}{2}(x_{min} + x_{max})$.
- Si $f'(x) > 0$, $x_{max} \leftarrow x$, sinon $x_{min} \leftarrow x$.

Répèter l'opération jusqu'à un critère d'arrêt à préciser.

La longueur de l'intervalle est à chaque itération divisée par 2, la convergence est linéaire de taux 0,5.

Détermination de x_{min} et x_{max}

Une bonne méthode est la suivante (on suppose $f'(0) \neq 0$) :

- définir un pas de déplacement $h > 0$.
- si $f'(0) < 0$, faire :
  - ⋆ $x_{min} \leftarrow 0$
  - ⋆ tant que $f'(h) < 0$, faire
    - ▶ $x_{min} \leftarrow h$
    - ▶ $h \leftarrow 2h$
  - ⋆ $x_{max} \leftarrow h$
- sinon si $f'(0) > 0$, faire :
  - ⋆ $h \leftarrow -h$
  - ⋆ $x_{max} \leftarrow 0$
  - ⋆ tant que $f'(h) > 0$, faire
    - ▶ $x_{max} \leftarrow h$
    - ▶ $h \leftarrow 2h$
  - ⋆ $x_{min} \leftarrow h$

Sommaire

Optimisation linéaire, l'algorithme du simplexe

Dualité en programmation linéaire

Optimisation non linéaire sans contraintes

Optimisation non linéaire avec contraintes

- Généralités
- Condition de Lagrange
- Condition de Karush, Kuhn et Tucker
- Méthode de descente
- Cas des fonctions convexes
- Linéarisation

Optimisation non linéaire avec contraintes

Généralités
Condition de Lagrange
Condition de Karush, Kuhn et Tucker
Méthode de descente
Cas des fonctions convexes
Linéarisation

Expression du problème

Dans toute la section, les fonctions f, g_i $(1 \leqslant i \leqslant m)$ et h_j $(1 \leqslant j \leqslant p)$ sont de $\mathbb{R}^n$ dans $\mathbb{R}$ et **de classe $\mathbf{C}^1$**.

$$\begin{array}{ll} & \text{Minimiser } f(x) \\ \text{avec} & \left\{ \begin{array}{ll} \text{pour } i \in \{1, ..., m\}, & g_i(x) \leqslant 0 \\ \text{pour } j \in \{1, ..., p\}, & h_j(x) = 0 \end{array} \right. \\ \text{et} & x \in \mathbb{R}^n. \end{array} \qquad (P)$$

On pose $I = \{1, ..., m\}$ et $J = \{1, ..., p\}$.

Les conditions $g_i(x) \leqslant 0$ et $h_j(x) = 0$ s'appellent les *contraintes*.

Tout x appartenant à $\mathbb{R}^n$ qui vérifie le système des contraintes s'appelle *solution réalisable*.

L'ensemble des solutions réalisables est le *domaine réalisable*. On note X le domaine réalisable.

Si, pour $i \in \{1, 2, ..., m\}$ et pour $x \in X$, on a $g_i(x) = 0$, on dit que la contrainte g_i est *saturée* en x.

Théorème et définition

Le domaine réalisable X est un fermé de $\mathbb{R}^n$.

On supposera X non vide.

Théorème

Si X est borné, alors le problème (P) admet au moins une solution.

Preuve. Le domaine réalisable est un fermé, borné non vide de $\mathbb{R}^n$. Le résultat résulte d'un théorème de topologie : « toute fonction continue sur un fermé borné non vide de $\mathbb{R}^n$ est bornée et atteint ses bornes ».

Définition

Soit f une fonction de $\mathbb{R}^n$ dans $\mathbb{R}$. On dit que f est coercive si elle tend vers $+\infty$ quand $||x||$ tend vers $+\infty$.

Théorème

Théorème

Si la fonction f est coercive, le problème (P) admet au moins une solution.

Preuve. Soit x une solution réalisable.

f coercive $\Rightarrow$ $\exists$ une boule B de $\mathbb{R}^n$ centrée sur l'origine telle que, pour tout x' non dans B, $f(x') > f(x)$.

L'ensemble $B \cap X$ étant un ensemble fermé et borné de $\mathbb{R}^n$, la fonction f atteint son minimum sur $B \cap X$. Ce minimum est aussi un minimum global pour le problème (P).

Définition

Définition

On dit qu'une direction d est admissible en $x^0 \in X$ s'il existe une fonction ϕ de $\mathbb{R}$ dans $\mathbb{R}^n$ telle que :

1. $\phi(0) = x^0$
2. pour tout $t > 0$ assez petit, $\phi(t) \in X$
3. la dérivée à droite de ϕ en 0 est d.

Soit $x^0 \in X$.

On note $A(x^0)$ l'ensemble des directions admissibles en x^0.

On pose $I_0(x^0) = \{i \in I \text{ vérifiant } g_i(x^0) = 0\}$.

Proposition

Si d est une direction admissible en x^0, alors :
(i) pour $i \in I_0(x^0)$, $d^{\mathrm{t}}.\nabla g_i(x^0) \leqslant 0$
(ii) pour $j \in J, d^{\mathrm{t}}.\nabla h_j(x^0) = 0$.

Preuve. Soit ϕ correspondant à la définition de d admissible.

(i) Si $g_i(x^0) = 0$, alors $g_i(\phi(t)) = td^{\mathrm{t}}.\nabla g_i(x^0) + t\epsilon(t)$ où $\epsilon(t) \to 0$ quand $t \to 0^+$ (formule de Taylor pour $t \mapsto g_i(\phi(t))$).

Pour $t > 0$ assez petit, $g_i(\phi(t)) \leqslant 0$: $d^{\mathrm{t}}.\nabla g_i(x^0) + \epsilon(t) \leqslant 0$, ce qui donne (i).
(ii) $h_j(\phi(t)) = h_j(x^0) + td^{\mathrm{t}}.\nabla h_j(x^0) + t\epsilon(t)$ où $\epsilon(t) \to 0$ quand $t \to 0^+$.
Pour $t > 0$ assez petit, $h_j(\phi(t)) = 0$ et $h_j(x^0) = 0$; on a donc pour $t > 0$ assez petit :

$$d^{\mathrm{t}}.\nabla h_j(x^0) + \epsilon(t) = 0, \text{ ce qui donne (ii).}$$

Qualification des contraintes

On note $B(x^0)$ l'ensemble des directions d vérifiant :

- pour $i \in I_0(x^0)$, $d^{\mathrm{t}}.\nabla g_i(x^0) \leqslant 0$

- pour $j \in J, d^{\mathrm{t}}.\nabla h_j(x^0) = 0$.

$A(x^0) \subseteq B(x^0)$.

Définition

On dit que les contraintes sont *qualifiées* en $x^0 \in X$ si toute direction dans $B(x^0)$ est limite d'une suite de directions de $A(x^0)$.

Lemme 1

On suppose que, pour $j \in J$, les fonctions h_j sont affines.

Soient $x^0 \in X$ et d une direction vérifiant :

- pour $i \in I_0(x^0), d^{\mathrm{t}}.\nabla g_i(x^0) < 0$

- pour $j \in J$, $d^{\mathrm{t}}.\nabla h_j(x^0) = 0$.

Alors d est une direction admissible en x^0.

Preuve. Pour $t \geqslant 0$, on pose : $\phi(t) = x^0 + td$.

$\phi(0) = x^0$ et $\phi'(0) = d$.

Pour $j \in J$: $h_j(\phi(t)) = h_j(x^0) + td^{\mathrm{t}}.\nabla h_j(x^0)$ (h_j affine).

$h_j(x^0) = 0$ et $d^{\mathrm{t}}.\nabla h_j(x^0) = 0$. D'où $h_j(\phi(t)) = 0$.

Pour $i \in I_0(x^0)$, $g_i(\phi(t)) = g_i(x^0) + t(d^{\mathrm{t}}.\nabla g_i(x^0) + \epsilon(t))$, où $\epsilon(t) \to 0$ quand $t \to 0$.

$g_i(x^0) = 0$ et $d^{\mathrm{t}}.\nabla g_i(x^0) < 0 \Rightarrow g_i(\phi(t)) \leqslant 0$ pour $t > 0$ petit.

Ainsi, la direction d est admissible.

Lemme 2

On suppose que, pour $j \in J$, les fonctions h_j sont affines.

Soit $x^0 \in X$. S'il existe une direction $\tilde{d}$ telle que :

- pour $i \in I_0(x^0)$, $\tilde{d}^t.\nabla g_i(x^0) < 0$
- pour $j \in J, \tilde{d}^t.\nabla h_j(x^0) = 0$,

alors les contraintes sont qualifiées en x^0.

Preuve. Soient $d \in B(x^0)$ et $\tilde{d}$ vérifiant les hypothèses du lemme.

Pour $\lambda \in [0, 1[$, soit $d_\lambda = \lambda d + (1 - \lambda)\tilde{d}$.

Si $i \in I_0(x^0)$: $d_\lambda^t.\nabla g_i(x^0) = \lambda d^t.\nabla g_i(x^0) + (1 - \lambda)\tilde{d}^t.\nabla g_i(x^0) < 0$.

Si $j \in J$: $d_\lambda^t.\nabla h_j(x^0) = \lambda d^t.\nabla h_j(x^0) + (1 - \lambda)\tilde{d}^t.\nabla h_j(x^0) = 0$.

Lemme 1 $\Rightarrow$ d_λ est une direction admissible.

λ_n : suite de nombres tendant vers 1 par valeurs inférieures $\rightarrow$ suite d_{λ_n} de directions admissibles qui tend vers d.

Les contraintes sont qualifiées en x^0.

Proposition

Si : - les fonctions g_i sont convexes,
- les fonctions h_j sont affines,
- et il existe $\tilde{x} \in X$ avec,

pour tout $i \in I, g_i(\tilde{x}) < 0$

pour tout $j \in J, h_j(\tilde{x}) = 0$

alors les contraintes sont qualifiées en tout point de X.

Preuve. Soit $x^0 \in X$ et soit $\tilde{x}$ vérifiant les hypothèses.

$i \in I_0(x_0)$: $0 > g_i(\tilde{x}) \geqslant g_i(x^0) + (\tilde{x} - x^0)^{\mathrm{t}}.\nabla g_i(x^0)$ (g_i convexe).
Comme $g_i(x^0) = 0$: $(\tilde{x} - x^0)^{\mathrm{t}}.\nabla g_i(x^0) < 0$.
On pose $\tilde{d} = \tilde{x} - x^0$; on a $\tilde{d}^{\mathrm{t}}.\nabla g_i(x^0) < 0$.
Si $j \in \{1, ..., p\}$: $0 = h_j(\tilde{x}) = h_j(x^0) + \tilde{d}^{\mathrm{t}}.\nabla h_j(x^0)$ (h_j affines)
d'où : $\tilde{d}^{\mathrm{t}}.\nabla h_j(x^0) = 0$.

Lemme 2 : les contraintes sont qualifiées en x^0.

Proposition

On suppose que, pour $j \in J$, les fonctions h_j sont affines.

Si, en $x^0 \in X$, l'ensemble des gradients

- $\nabla g_i(x^0)$ pour $i \in I_0(x^0)$

- $\nabla h_j(x^0)$ pour $j \in J$

sont linéairement indépendants,

alors les contraintes sont qualifiées en x^0.

Début de la preuve

Preuve. On considère le problème (Q) d'optimisation linéaire ci-dessous :

Maximiser $z = \sum_{i \in I_0(x^0)} \lambda_i$

$$\text{avec} \begin{cases} \sum_{j \in J} \mu_j \nabla h_j(x^0) - \sum_{i \in I_0(x^0)} \lambda_i \nabla g_i(x^0) = 0 \quad (Q) \\ \text{pour } i \in I_0(x^0), \lambda_i \geqslant 0 \end{cases}$$

Le problème (Q) est réalisable puisque la solution nulle l'est.

Supposons qu'il puisse prendre une valeur strictement positive.

Dans cette solution, au moins un λ_i $(i \in I_0(x^0))$ est non nul : les vecteurs $\nabla g_i(x^0)$ $(i \in I_0(x^0))$ et $\nabla h_j(x^0)$ $(j \in J)$ sont linéairement dépendants, contrairement à l'hypothèse.

Le maximum du problème (Q) vaut donc 0 et le problème (Q) admet une solution optimale.

Suite de la preuve

On considère le problème (R) d'optimisation linéaire ci-dessous :

Minimiser $w = 0$

$$\text{avec} \begin{cases} \text{pour } i \in I_0(x^0), d^t.\nabla g_i(x^0) \leqslant -1 \\ \text{pour } j \in J, d^t.\nabla h_j(x^0) = 0. \end{cases} \qquad (R)$$

On peut facilement vérifier que les problèmes (Q) et (R) sont duaux l'un de l'autre.

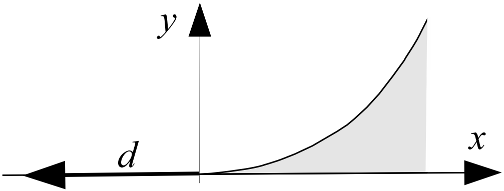
On utilise le théorème de la dualité pour l'optimisation linéaire : le problème (R) est réalisable. On note $\tilde{d}$ une solution réalisable de (R). Le lemme 2 permet de conclure.

Remarque. On peut montrer que cette proposition est encore exacte si on retire l'hypothèse que les fonctions h_j sont affines.

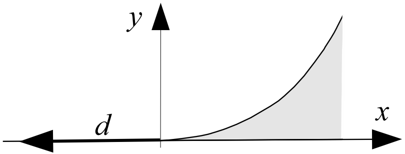
Un exemple de points où les contraintes ne sont pas qualifiées

On considère dans $\mathbb{R}^2$ le domaine défini par :

$$\begin{cases} y \leqslant x^3 \\ y \geqslant 0. \end{cases}$$



Exemple, suite



$$g_1(x,y) = y - x^3 \qquad g_2(x,y) = -y \quad , \quad \begin{cases} g_1(x,y) \leqslant 0 \\ g_2(x,y) \leqslant 0 \end{cases} .$$

Les contraintes g_1 et g_2 sont saturées en (0, 0).

$\nabla g_1(0,0) = (0,1)^{\mathrm{t}}$ et $\nabla g_2(0,0) = (0,-1)^{\mathrm{t}}$.

Pour $d = (-1,0)^{\mathrm{t}}$: $d^{\mathrm{t}}.\nabla g_1(0,0) = 0$ et $d^{\mathrm{t}}.\nabla g_2(0,0) = 0$

La direction d appartient à $B(0,0)$.

Or, la seule direction admissible en (0, 0) est la direction $(1,0)^{\mathrm{t}}$.
d n'est pas limite d'une suite de directions admissibles.

Théorème

On suppose que le problème admet un minimum local en un point x^* où les contraintes sont qualifiées. Alors, si $d \in B(x^*)$: $d^{\mathrm{t}}.\nabla f(x^*) \geqslant 0$ (en particulier, aucune direction admissible en x^* n'est de descente).

Preuve. Soit (d_k) une suite de directions admissibles tendant vers d et soit ϕ_k la fonction associée à d_k.

Soit $t > 0$.

$f[\phi_k(t)] = f(x^*) + td_k^{\mathrm{t}}.\nabla f(x^*) + t\epsilon(t)$ où $\epsilon(t) \to 0$ quand $t \to 0^+$.

Si t est assez petit pour que $\phi_k(t)$ appartienne à X : $f[\phi_k(t)] \geqslant f(x^*)$.

On a alors : $t[d_k^{\mathrm{t}}.\nabla f(x^*) + \epsilon(t)] \geqslant 0$ et donc $d_k^{\mathrm{t}}.\nabla f(x^*) + \epsilon(t) \geqslant 0$.

Par passage à la limite quand t tend vers 0, on obtient $d_k^{\mathrm{t}}.\nabla f(x^*) \geqslant 0$.

Optimisation non linéaire avec contraintes

Généralités

Condition de Lagrange

Condition de Karush, Kuhn et Tucker

Méthode de descente

Cas des fonctions convexes

Linéarisation

Condition nécessaire de Lagrange

On suppose qu'il n'y a **que des contraintes d'égalité** :

$$\begin{aligned}
&\text{Minimiser } f(x) \\
&\text{avec } x \in \mathbb{R}^n \\
&\text{et, pour } 1 \leqslant j \leqslant p, \quad h_j(x) = 0
\end{aligned}$$

où les fonctions f et h_j $(1 \leqslant j \leqslant p)$ sont de classe C^1.
La condition de Lagrange fournit une condition nécessaire pour qu'un élément de $\mathbb{R}^n$ soit un minimum local de (P).

Condition de Lagrange

Soit x^* un minimum local du problème. On suppose que les contraintes sont qualifiées en x^*. Alors il existe des réels $\mu_1, \mu_2, \ldots, \mu_p$ tels que :

$$\nabla f(x^*) = \sum_{j=1}^{p} \mu_j \nabla h_j(x^*)$$

Preuve de la condition de Lagrange

Condition de Lagrange : si x^* est un minimum local du problème où les contraintes sont qualifiées, il existe des réels $\mu_1, \mu_2, \ldots, \mu_p$ tels que : $\nabla f(x^*) = \sum_{j=1}^{p} \mu_j \nabla h_j(x^*)$.

Preuve.
E : le sous-espace de $\mathbb{R}^n$ engendré par les vecteurs $\nabla h_j(x^*)$.

$E^\perp$: le sous-espace orthogonal à E.

$\nabla f(x^*) = y + z$ avec $y \in E$ et $z \in E^\perp$.

Pour $1 \leqslant i \leqslant p$, $(-z)^t.\nabla h_j(x^*) = 0$ puisque $-z$ appartient à $E^\perp$; d'où $-z \in B(x^*)$.

D'après le théorème précédent : $(-z)^t.\nabla f(x^*) \geqslant 0$.

Or : $(-z)^t.\nabla f(x^*) = (-z)^t.y + (-z)^t.z = (-z)^t.z = -||z||^2$.

$-||z||^2 \geqslant 0 \Rightarrow z = 0$. D'où le théorème.

Condition de Lagrange suffisante ?

Le théorème suivant, démontré plus loin, donne des hypothèses pour lesquelles la condition de Lagrange est suffisante.

Théorème

La condition de Lagrange est suffisante lorsque f est convexe dans un ouvert contenant X et que les h_j $(1 \leqslant i \leqslant p)$ sont affines.

Optimisation non linéaire avec contraintes

Généralités
Condition de Lagrange
Condition de Karush, Kuhn et Tucker
Méthode de descente
Cas des fonctions convexes
Linéarisation

Condition nécessaire de Karush, Kuhn et Tucker

On reprend le problème (P) initial :

$$
(P) \quad \text{avec} \quad \begin{array}{l} \text{Minimiser } f(x) \\ \left\{ \begin{array}{ll} \text{pour } 1 \leqslant i \leqslant m, & g_i(x) \leqslant 0 \\ \text{pour } 1 \leqslant j \leqslant p, & h_j(x) = 0 \end{array} \right. \end{array}
$$

$$
\text{et} \quad x \in \mathbb{R}^n
$$

f, g_i $(1 \leqslant i \leqslant m)$ et h_j $(1 \leqslant j \leqslant p)$ de classe C^1.

Théorème (Condition de Karush, Kuhn et Tucker)

On suppose que les contraintes sont qualifiées en x^* et que x^* est un minimum local du problème ; alors il existe :

- m nombres réels positifs ou nuls $\lambda_1, \lambda_2, ..., \lambda_m$
- p nombres réels $\mu_1, \mu_2, ..., \mu_p$

tels que : $\nabla f(x^*) = \sum_{j \in J} \mu_j \nabla h_j(x^*) - \sum_{i \in I_0(x^*)} \lambda_i \nabla g_i(x^*)$.

Première illustration, $n = 2, p = 0$, une seule contrainte d'inégalité saturée

x et y : les coordonnées d'un point.

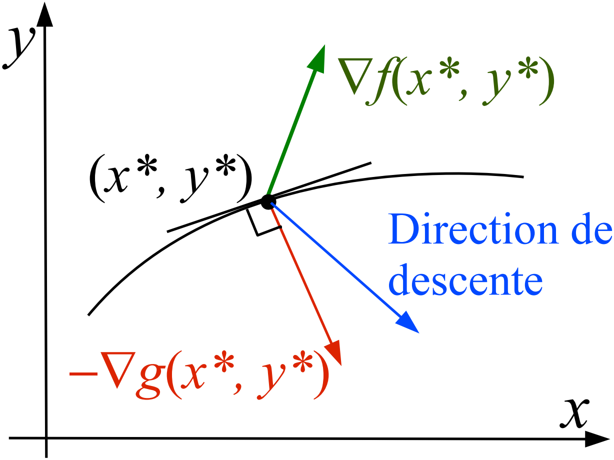
On suppose que seule la contrainte $g(x, y) \leqslant 0$ est saturée en (x^*, y^*) : $g(x^*, y^*) = 0$.

Le vecteur $-\nabla g(x^*, y^*)$ est perpendiculaire à la courbe d'équation $g(x, y) = 0$ et dirigée vers l'intérieur du domaine.

On rappelle que :
• une direction est une direction de descente si elle fait un angle obtus avec $\nabla f(x^*, y^*)$

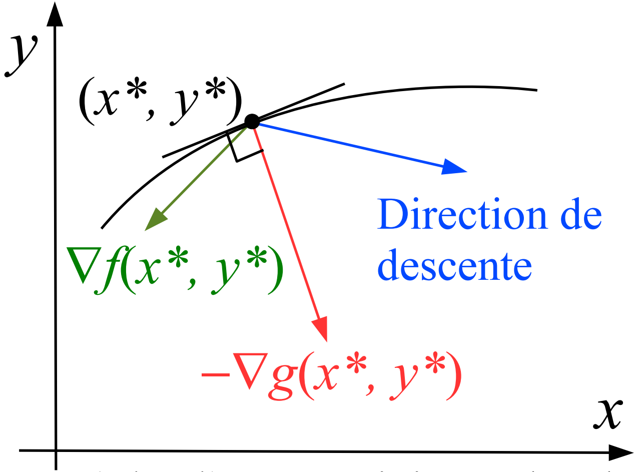
• une direction admissible fait un angle aigu avec $-\nabla g(x^*, y^*)$.

Première illustration, cas 1



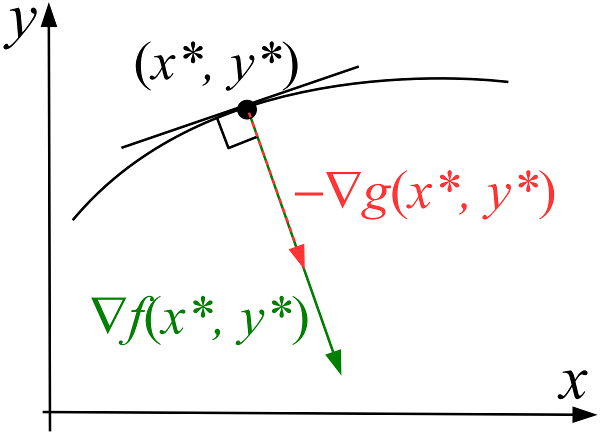
(x^*, y^*) : non minimum local

Première illustration, cas 2



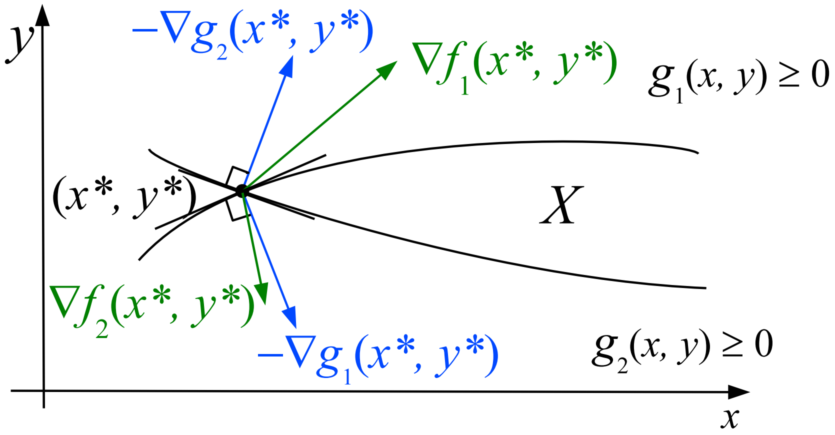
(x^*, y^*) : non minimum local

Première illustration, cas 3



(x^*, y^*) : condition satisfaite

Seconde illustration, $n = 2, p = 0$, deux contraintes d'inégalité saturées



Pour f_1, (x^*, y^*) peut être un minimum local mais pas pour f_2.

Condition de Karush, Kuhn et Tucker suffisante ?

Théorème

On suppose que les contraintes sont qualifiées en un point x^*. La condition de Karush, Kuhn et Tucker en x^* est suffisante pour avoir un minimum local lorsque simultanément f est convexe dans un voisinage de x^*, les g_i $(i \in I_0(x^*))$ sont convexes dans un voisinage de x^* et les h_j $(1 \leqslant j \leqslant p)$ sont affines dans un voisinage de x^*.

Preuve. On suppose qu'il existe des nombres réels positifs ou nuls λ_i $(i \in I_0(x^*))$ et des nombres réels μ_j $(1 \leqslant j \leqslant p)$ tels que :

$\nabla f(x^*) = \sum_{j \in J} \mu_j \nabla h_j(x^*) - \sum_{i \in I_0(x^*)} \lambda_i \nabla g_i(x^*)$.

B : boule de centre x^* dans laquelle f et les g_i sont convexes et les h_j affines.

$x \in B \cap X$; on va montrer que $f(x) \geqslant f(x^*)$, ce qui prouvera le théorème.

Preuve suite

Rappel : $\nabla f(x^*) = \sum_{j\in J} \mu_j \nabla h_j(x^*) - \sum_{i\in I_0(x^*)} \lambda_i \nabla g_i(x^*)$.

La convexité de f dans B induit : $f(x) \geqslant f(x^*) + (x - x^*)^{\mathrm{t}}.\nabla f(x^*)$.

$$KKT \Rightarrow f(x) \geqslant f(x^*) + \sum_{j\in J} \mu_j (x - x^*)^{\mathrm{t}}.\nabla h_j(x^*) \\ - \sum_{i\in I_0(x^*)} \lambda_i (x - x^*)^{\mathrm{t}}.\nabla g_i(x^*).$$

Pour $j \in J, (x - x^*)^{\mathrm{t}}.\nabla h_j(x^*) = h_j(x) - h_j(x^*) = 0$.

Soit $i \in I_0(x^*)$.

$g_i(x) \geqslant g_i(x^*) + (x - x^*)^{\mathrm{t}}.\nabla g_i(x^*)$ (convexité de g_i).

$(x - x^*)^{\mathrm{t}}.\nabla g_i(x^*) \leqslant g_i(x) - g_i(x^*)$.

Or, on a $\lambda_i \geqslant 0$; de plus, par hypothèse, $g_i(x^*) = 0$ et $g_i(x) \leqslant 0$.

$\lambda_i (x - x^*)^{\mathrm{t}}.\nabla g_i(x^*) \leqslant 0$.

On obtient $f(x) \geqslant f(x^*)$: f admet un minimum local en x^*.

Généralités
Condition de Lagrange
Condition de Karush, Kuhn et Tucker
Méthode de descente
Cas des fonctions convexes
Linéarisation

Méthode de descente

On considère le problème suivant :

$$\begin{array}{rl} & \text{Minimiser } f(x) \\ \text{avec,} & \text{pour } 1 \leqslant i \leqslant m,\ g_i(x) \leqslant 0. \end{array}$$

On choisit un point de départ $x^0 \in X$ et on construit de façon itérative une suite x^k de X vérifiant $f(x^{k+1}) < f(x^k)$, jusqu'à ce qu'on estime avoir obtenu une approximation satisfaisante.

À partir de x^k, on recherche une direction de descente d qui ne fasse pas sortir « immédiatement » de X.

On cherche alors, en se déplaçant dans la direction d, un point x^{k+1} de X meilleur que x^k.

On recommence à partir de x^{k+1} tant qu'un certain critère d'arrêt n'est pas vérifié.

Méthode de descente, choix de d

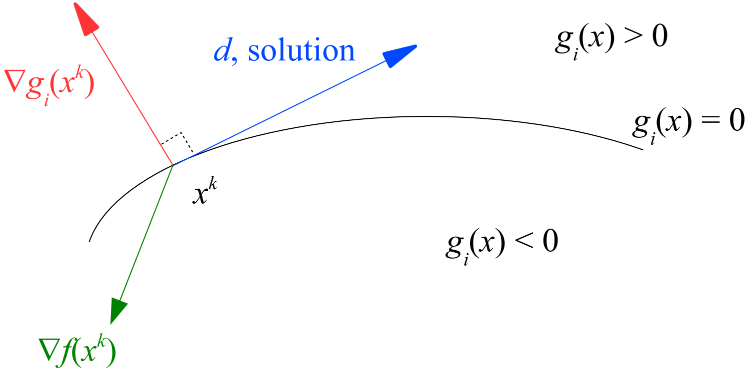
On peut résoudre le problème :

$$\begin{array}{ll} & \text{Minimiser } d^{\mathrm{t}}.f(x) \\ \text{avec} & d^{\mathrm{t}}.\nabla g_i(x^k) \leqslant 0 \text{ pour tout } i \text{ tel que } g_i(x^k) = 0 \\ \text{et} & d^{\mathrm{t}}.d = 1. \end{array}$$

On obtient alors pour d la direction dans $B(x^k)$ de plus grande descente.
On peut choisir de remplacer la condition $d^{\mathrm{t}}.d = 1$ par la condition : $-1 \leqslant d_i \leqslant 1$ $(1 \leqslant i \leqslant n)$ pour avoir un problème linéaire (la direction retenue ne sera pas exactement la direction de plus grande descente).

Méthode de descente, un problème

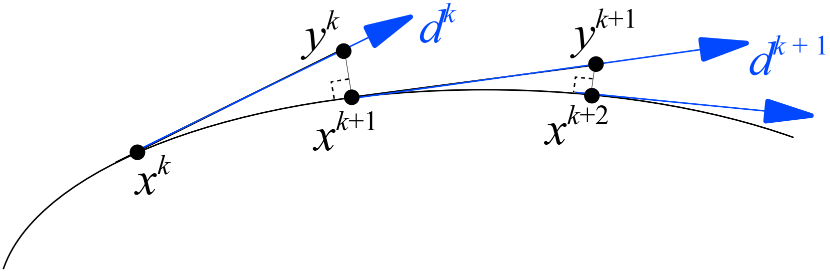
On considére l'exemple représenté ci-dessous.



Tout déplacement dans la direction d fait sortir de X.

Méthode de descente, projection

Il faut une procédure de projection pour que x^{k+1} soit dans X.



Optimisation non linéaire avec contraintes

Généralités
Condition de Lagrange
Condition de Karush, Kuhn et Tucker
Méthode de descente
Cas des fonctions convexes
Linéarisation

Généralités

On suppose dans toute cette partie que :

- les fonctions g_i $(1 \leqslant i \leqslant m)$ sont convexes,
- les fonctions h_j $(1 \leqslant j \leqslant p)$ sont affines,
- la fonction f est convexe.

Définition

Soit C une partie de $\mathbb{R}^n$. On dit que C est convexe si, pour tout couple (x, x') de points de C, le segment d'extrémités x et x' est contenu dans C.

Remarque

Le domaine réalisable du problème (P), c'est-à-dire :

$$\{x \in \mathbb{R}^n | g_i(x) \leqslant 0 \text{ pour } 1 \leqslant i \leqslant m \text{ et } h_j(x) = 0 \text{ pour } 1 \leqslant j \leqslant p\}$$

est convexe.

Théorème 1

Si f est strictement convexe, le problème (P) admet au plus une solution optimale.

Preuve

On suppose qu'il existe dans C deux solutions optimales x et y.

On pose : $z = \dfrac{x+y}{2}$.

Convexité du domaine $\Rightarrow z \in C$

Stricte convexité de $f \Rightarrow f(z) < \dfrac{f(x)+f(y)}{2}$.

On a donc soit $f(z) < f(x)$, soit $f(z) < f(y)$, contradiction avec l'optimalité supposée de x et y.

Conséquences

Rappel : domaine borné $\Rightarrow$ au moins une solution optimale.

Théorème 2

Si le domaine réalisable est borné et si f est strictement convexe, le problème (P) admet une unique solution.

Rappel : f coercive $\Rightarrow$ au moins une solution optimale.

Théorème 3

Si f est strictement convexe et coercive, le problème (P) admet une unique solution.

Théorème 4

On suppose qu'on a un minimum local en un point x^* où les contraintes sont qualifiées. Alors le problème (P) admet un minimum global en x^*.

Preuve. Soit x une solution réalisable du problème (P).

Soit ψ la fonction définie sur l'intervalle $[0, 1]$ par :
$\psi(t) = f[x^* + t(x - x^*)]$.

On a : $\psi(0) = f(x^*)$ et $\psi(1) = f(x)$.

$\psi'(0) = (x - x^*)^{\mathrm{t}}.\nabla f(x^*)$.

La direction $d = x - x^*$ est admissible (car X est convexe).

Donc $\psi'(0) \geqslant 0$ (aucune direction admissible n'est de descente).

Convexité de $f \Rightarrow$ convexité de ψ sur $[0, 1]$.

$\psi'(0) \geqslant 0 \Rightarrow \psi(1) \geqslant \psi(0) \Rightarrow f(x) \geqslant f(x^*)$.

Théorème 5

On suppose que la condition de Karush, Kuhn et Tucker est vérifiée en un point x^* où les contraintes sont qualifiées. Avec les hypothèses de cette partie, le point x^* est un minimum global de (P). De plus, si f est strictement convexe, x^* est l'unique point où (P) atteint le minimum global.

Preuve. D'après le théorème donnant les hypothèses pour que la condition KKT soit suffisante, le problème (P) atteint un minimum local en x^*.

Théorème 4 $\Rightarrow$ le problème (P) admet un minimum global en x^*.

Si, de plus, f est strictement convexe, le théorème 1 $\Rightarrow$ (P) atteint son minimum global uniquement en x^*.

Optimisation non linéaire avec contraintes

Généralités
Condition de Lagrange
Condition de Karush, Kuhn et Tucker
Méthode de descente
Cas des fonctions convexes
Linéarisation

Algorithme basique

Recherche du minimum d'une fonction $f : \mathbb{R}^n \mapsto \mathbb{R}$ de classe C^1 sur un domaine fermé convexe X.

On va « linéariser » f, en l'approchant par son développement de Taylor à l'ordre 1. Algorithme :

- $x^0 \leftarrow$ un point quelconque
- $k \leftarrow 0$
- répéter

  $x^{k+1} \leftarrow$ point qui minimise $f(x^k) + (x - x^k)^{\mathrm{t}}.\nabla f(x^k)$ sur X.
  $k \leftarrow k + 1$

jusqu'à ce qu'un test d'arrêt à préciser soit vérifié.

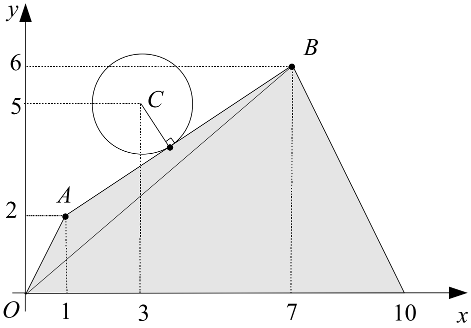
Remarques.

1) x^{k+1} est aussi un point qui minimise $x^{\mathrm{t}}.\nabla f(x^k)$ sur X.

2) Si le domaine est un polyèdre, la détermination de x^{k+1} est un problème d'optimisation linéaire.

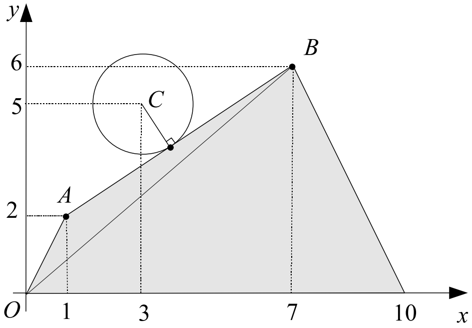
Algorithme basique : exemple et difficulté

$$(P_0) \qquad \text{Minimiser } f(x,y) = (x-3)^2 + (y-5)^2$$
$$\text{avec les contraintes : } \begin{cases} y - 2x \leqslant 0 \\ 2x + y - 20 \leqslant 0 \\ -2x + 3y - 4 \leqslant 0 \\ y \geqslant 0. \end{cases}$$



f est constante sur des cercles centrés sur le point C =(3, 5).
f est minimum pour le point le plus proche de C : $(\frac{49}{13}, \frac{50}{13})$.

Algorithme basique : exemple et difficulté, suite



$\nabla f(x, y) = (2x - 6, 2y - 10)^t$.

On démarre à partir de O, $\nabla f(0, 0) = (-6, -10)^t$.

Le minimum de $-6x - 10y$ est atteint en $B = (7, 6)$.

À partir du sommet B : $\nabla f(7, 6) = (8, 2)^t$. Le minimum de $8x + 2y$ est atteint en $A = (1, 2)$.

À partir de A : $\nabla f(1, 2) = (-4, -6)$. Le minimum de $-4x - 6y$ est atteint en B.

La méthode ne converge pas.

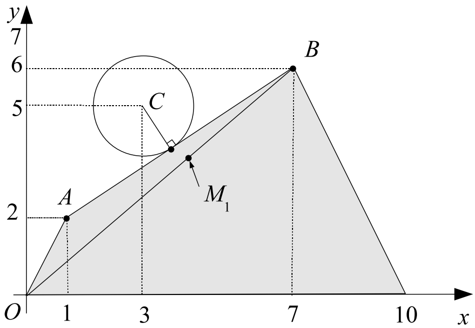
Linéarisation : méthode de Frank et Wolfe

La méthode de Frank et Wolfe s'applique à la minimisation d'**une fonction de classe C^1 convexe sur un domaine X convexe et compact**.

- $x^0 \leftarrow$ un point quelconque
- $k \leftarrow 0$
- répéter
  - $\star$ $\tilde{x^k} \leftarrow$ un point qui minimise $x^{t}.\nabla f(x^k)$ sur X
  - $\star$ $x^{k+1} \leftarrow$ un point qui minimise f sur le segment $[x^k, \tilde{x^k}]$
  - $\star$ $k \leftarrow k+1$

jusqu'à ce qu'un test d'arrêt à préciser soit vérifié.

Frank et Wolfe pour le problème précédent, étape 1



$\nabla f(x, y) = (2x - 6, 2y - 10)^t$.

Étape 1 : à partir du point (0, 0).

Minimum de $-6x - 10y$ sur $X \to (\tilde{x}_0, \tilde{y}_0) = B = (7, 6)$.

Minimum de f sur le segment $[O, B]$?

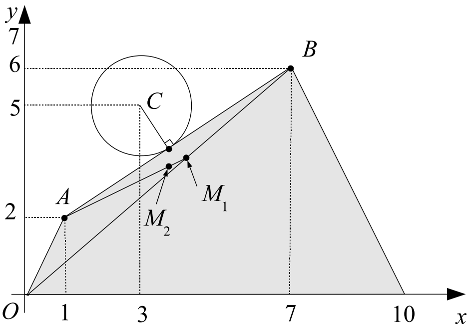
$x = 7t, y = 6t$ $(0 \leqslant t \leqslant 1)$, $\phi_1(t) = f(7t, 6t)$.

Après calcul : le minimum de ϕ_1 est obtenu pour $t_1 = \frac{3}{5} = 0,6$.

La fonction f atteint son minimum sur le segment $[O, B]$ au point :

$M_1 = (x_1, y_1) = (7 \times 0,6\ ; 6 \times 0,6) = (4,2\ ; 3,6)$.

Frank et Wolfe pour le problème précédent, étape 2



$\nabla f(x, y) = (2x - 6, 2y - 10)^{t}$.

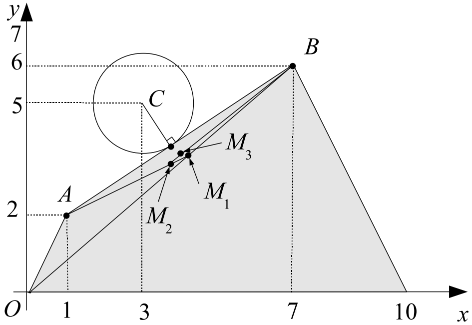
Étape 2 : à partir du point $M_1 = (4,2\ ; 3,6)$.

Minimum de $(2 \times 4,2 - 6)x + (2 \times 3,6 - 10)y = 2,4x - 2,8y$ sur $X \to (\tilde{x}_1, \tilde{y}_1) = A = (1, 2)$.

Minimum de f sur le segment $[M_1, A]$?

Après paramétrisation et calcul : la fonction f atteint son minimum sur le segment $[M_1, A]$ au point : $M_2 = (x_2, y_2) = (3,8\ ; 3,4)$.

Frank et Wolfe pour le problème précédent, étape 3



$\nabla f(x, y) = (2x - 6, 2y - 10)^{\mathrm{t}}$.

Étape 3 : à partir du point $M_2 = (3,8\ ; 3,4)$.

Minimum de $(2 \times 3,8 - 6)x + (2 \times 3,4 - 10)y = 1,6x - 3,2y$ sur $X \to (\tilde{x_2}, \tilde{y_2}) = B = (7, 6)$.

Minimum de f sur le segment $[M_2, B]$?

Après paramétrisation et calcul : la fonction f atteint son minimum sur le segment $[M_2, B]$ au point : $M_3 = (x_3, y_3) \approx (4,101\ ; 3,64)$.

Rappel : point optimal $= (49/13; 50/13) \approx (3,769; 3,846)$.

Convergence de la méthode de Frank et Wolfe

Théorème

Soit f une fonction de $\mathbb{R}^n$ dans $\mathbb{R}$, de classe C^1, strictement convexe. Soit X un polyèdre convexe compact de $\mathbb{R}^n$. La méthode de Frank et Wolfe appliquée au problème (P) de la minimisation de f sur X converge vers le minimum global.