Blackbox optimization

Sébastien Le Digabel

GERAD
GROUPE D'ÉTUDES ET DE RECHERCHE EN ANALYSE DES DÉCISIONS

POLYTECHNIQUE MONTRÉAL
TECHNOLOGICAL UNIVERSITY

ETICS 2023

Structure globale du cours ETICS

- Introduction à l'optimisation $\simeq$ 2h
- Algorithmes pour l'optimisation de boîtes noires $\simeq$ 1h
- Applications de l'optimisation de boîtes noires $\simeq$ 1h
- Partie pratique: Solveur et benchmarking $\simeq$ 2h

Blackbox optimization: Part 1/4: Introduction à l'optimisation

Sébastien Le Digabel

GERAD — GROUPE D'ÉTUDES ET DE RECHERCHE EN ANALYSE DES DÉCISIONS

POLYTECHNIQUE MONTRÉAL — TECHNOLOGICAL UNIVERSITY

ETICS 2023

Plan

Source

Ce cours est un condensé de mes notes du cours
MTH8415: Fondements de recherche opérationnelle

Termes importants du cours

- **Recherche opérationnelle (RO):** *Ensemble de techniques mathématiques appliquées à la modélisation, l'optimisation et l'analyse d'un processus*
- Modélisation
- Optimisation:
  - Continue
  - Linéaire (OL)
  - Non linéaire (ONL)
  - Combinatoire (OC)
  - En nombres entiers (ONE)
  - Sans dérivées (DFO)
  - De boîtes noires (BBO)
- Théorie des Graphes:
  - Cheminements optimaux
  - Flots
  - Problèmes de transport

Problème d'optimisation

L'optimisation est un domaine qui étudie les problèmes de la forme

$$\min_{\mathbf{x} \in \mathcal{X}} \{ f(\mathbf{x}) : \mathbf{x} \in \Omega \}$$

où

- $\mathcal{X}$ est un ensemble de dimension n:
  Les variables d'optimisation
- $\Omega \subseteq \mathcal{X}$ est l'ensemble des solutions réalisables:
  Les contraintes
- La fonction objectif f prend ses valeurs sur $\mathcal{X}$

Problème v.s. Instance

- Un problème correspond au modèle

$$\min_{\mathbf{x}\in\mathcal{X}}\{f_a(\mathbf{x}) : \mathbf{x} \in \Omega_a\}$$

  dans lequel a est un ensemble de paramètres non déterminé (c'est le cas général)

- Une instance du problème est une formulation du modèle dans laquelle a est déterminé (c'est un cas particulier)
- En pratique, on conçoit souvent des algorithmes pour des problèmes, qu'on teste sur plusieurs instances
- On peut aussi s'intéresser à une seule instance ou a une famille d'instances particulières. Dans ce cas on concevra une méthode plus spécialisée

Modèle d'optimisation

- Pour un problème donné, l'expression de f, $\mathcal{X}$ et Ω permet d'obtenir un modèle d'optimisation
- Optimisation continue (OL et ONL): $\mathcal{X} = \mathbb{R}^n$
- OC: $\mathcal{X}$ est un ensemble discret
- ONE: $\mathcal{X} = \mathbb{Z}^n$ ou $\mathbb{N}^n$ ou $\{0, 1\}^n$
- En théorie des graphes, il n'est pas forcément nécessaire d'exprimer un modèle d'optimisation. On se sert directement d'un graphe pour représenter le problème

Modèle d'optimisation (continue) non linéaire

$$\min_{\mathbf{x}\in\mathcal{X}=\mathbb{R}^n} f(\mathbf{x})$$

$$\text{s.c.} \quad \begin{cases} g_1(\mathbf{x}) \leq 0 \\ g_2(\mathbf{x}) \leq 0 \\ \dots \\ g_m(\mathbf{x}) \leq 0 \\ \mathbf{l} \leq \mathbf{x} \leq \mathbf{u} \end{cases}$$

- $f : \mathbb{R}^n \to \mathbb{R}$ différentiable: Fonction objectif
- $g_i(\mathbf{x}) \leq 0$, $i \in \{1, 2, \dots, m\}$: Contraintes
- $g_i : \mathbb{R}^n \to \mathbb{R}$ différentiable, $i \in \{1, 2, \dots, m\}$: Membres de gauche des contraintes
- $\mathbf{l}, \mathbf{u} \in \mathbb{R}^n$: Bornes sur les variables $\mathbf{x}$. Peuvent être $\pm\infty$
- $\Omega = \{\mathbf{x} \in \mathbb{R}^n : g_i(\mathbf{x}) \leq 0 \; \forall i \in \{1, 2, \dots, m\}, \mathbf{l} \leq \mathbf{x} \leq \mathbf{u}\}$

Modèle d'optimisation linéaire (forme standard)

$$\max_{\mathbf{x}\in\mathbb{R}^n} \quad f(\mathbf{x}) = \sum_{j=1}^{n} c_j x_j$$

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

Peut être exprimé de façon matricielle:

$$\max_{\mathbf{x}\in\mathbb{R}^n} \quad f(\mathbf{x}) = \mathbf{c}^\top \mathbf{x}$$

$$\text{s.c.} \quad \begin{cases} A\mathbf{x} \leq \mathbf{b} \\ \mathbf{x} \geq \mathbf{0} \end{cases}$$

avec $\mathbf{c} \in \mathbb{R}^n$ (coûts), $\mathbf{b} \in \mathbb{R}^m$ (membres de droite), et $A \in \mathbb{R}^{m\times n}$

Notes

- Fonction objectif (pas objective)
- Optimisation et pas programmation
- $\min$ et $\max$ sont équivalents
- Contraintes égalité $(=)$ et contraintes inégalité $(\leq$ et $\geq)$. On peut transformer des égalités en inégalités et vice-versa

Optimisation combinatoire (OC)

L'ONE et la théorie des graphes sont de l'OC

- En théorie, une solution optimale peut être obtenue en énumérant toutes les solutions réalisables et en conservant la meilleure. En pratique, ce procédé est trop long
- Pour les problèmes faciles, une résolution exacte en un temps court est envisageable
- Un grand nombre de problèmes sont difficiles. Des solutions exactes sont envisageables, mais dans un délai acceptable, on se contentera de solutions approchées obtenues par des méthodes heuristiques

Termes importants

- Optimum local vs global
- Algorithme exact vs heuristique

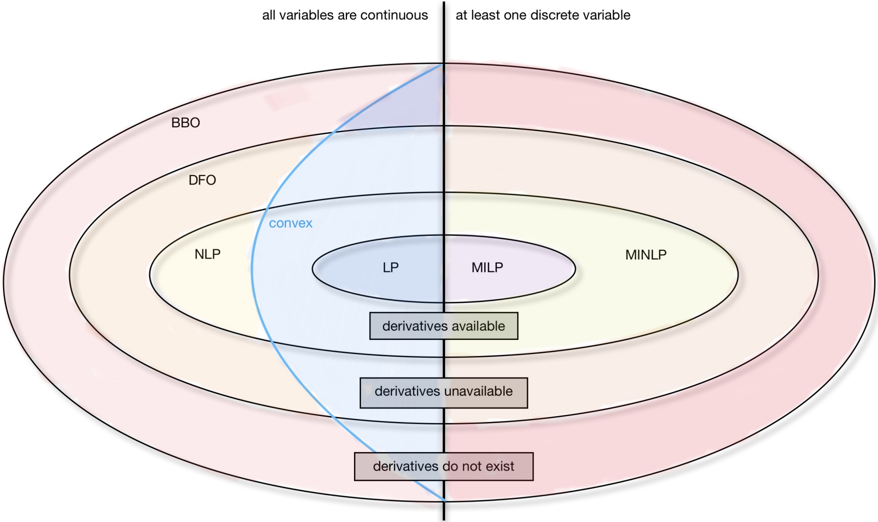
- En OL, on aura un optimum global
- En ONL, la plupart du temps, un optimum local, et si le problème est convexe, on aura un optimum global
- En OC, on aura soit une solution exacte (=un optimum global), soit un optimum local qui dépend d'un voisinage, ou alors une solution heuristique

Extensions

- Optimisation sans dérivées / de boîtes noires
- Optimisation multiobjectifs
- Optimisation multi-niveaux
- Optimisation stochastique
- Optimisation robuste
- Optimisation conique

...

Schéma global



Modèle d'optimisation linéaire (forme standard)

$$\max_{\mathbf{x}\in\mathbb{R}^n} \quad f(x) = \sum_{j=1}^{n} c_j x_j$$

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

Peut être exprimé de façon matricielle:

$$\max_{\mathbf{x}\in\mathbb{R}^n} \quad f(\mathbf{x}) = \mathbf{c}^\top \mathbf{x}$$

$$\text{s.c.} \quad \begin{cases} A\mathbf{x} \leq \mathbf{b} \\ \mathbf{x} \geq \mathbf{0} \end{cases}$$

avec $\mathbf{c} \in \mathbb{R}^n$ (coûts), $\mathbf{b} \in \mathbb{R}^m$ (membres de droite), et $A \in \mathbb{R}^{m\times n}$

Tout modèle d'OL peut être mis sous forme standard

- Si l'objectif est de minimiser f, il suffit de maximiser $-f$
- Une contrainte égalité peut être remplacée par deux contraintes inégalité
- Une contrainte $\geq$ peut être remplacée par une contrainte $\leq$
- Une variable sans bornes peut être remplacée par la différence de deux nouvelles variables non-négatives

Quelques définitions

- Un point $\mathbf{x} \in \mathbb{R}^n$ qui satisfait toutes les contraintes est tel que $A\mathbf{x} \leq \mathbf{b}$ et $\mathbf{x} \geq \mathbf{0}$. Il est appelé un point réalisable (ou admissible)
- La valeur d'un point $\mathbf{x}$ est la valeur $f(\mathbf{x})$
- Une solution (ou solution optimale, ou optimum global) $\mathbf{x}^*$ est un point réalisable dont la valeur (optimale) est la plus grande possible. Autrement dit, il n'existe pas d'autre point réalisable $\mathbf{y}$ tel que $f(\mathbf{y}) > f(\mathbf{x}^*)$
- Il peut y avoir plusieurs solutions optimales. On les trouve dans l'ensemble $\underset{A\mathbf{x}\leq\mathbf{b},\mathbf{x}\geq\mathbf{0}}{\arg\max}\ f(\mathbf{x})$
- On note $\mathbf{x}^* \in \underset{A\mathbf{x}\leq\mathbf{b},\mathbf{x}\geq\mathbf{0}}{\arg\max}\ f(\mathbf{x})$ ("$\in$", pas "$=$")

Trois possibilités

Pour tout modèle d'optimisation linéaire, une seule des trois situations suivantes peut survenir:

1. Il existe au moins une solution optimale. Soit une, soit une infinité
2. Le problème est non réalisable: Il n'est pas possible de satisfaire toutes les contraintes, i.e. le polyèdre formé par les contraintes est vide
3. Le problème est non borné: Il n'a pas de valeur optimale finie

Exemple de problème non borné

$$\max_{x_1,x_2} f(x_1,x_2) = x_1 - x_2$$

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

On peut poser $x_2 = 0$ et x_1 aussi grand que l'on veut pour obtenir $f = \infty$

Illustration sur un exemple 2D

$$\max_{x,y} 350x + 300y$$

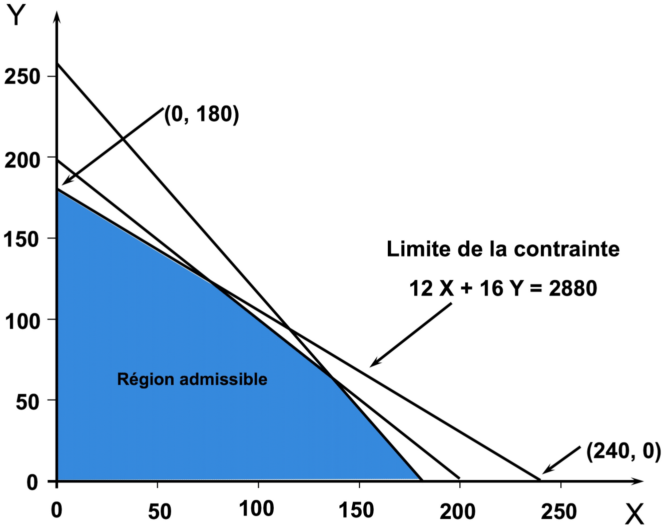
$$\text{s.c.} \begin{cases} x + y & \leq 200 & (1) \\ 9x + 6y & \leq 1566 & (2) \\ 12x + 16y & \leq 2880 & (3) \\ x \geq 0 & & (4) \\ y \geq 0 & & (5) \end{cases}$$

Interprétation graphique d'une contrainte

Considérons la contrainte (1): $x + y \leq 200$

- La droite $x + y = 200$ passe par les points (0,200) et (200,0) et divise le plan en 3 parties:
  - La partie au dessus de la droite correspond à l'ensemble des points tels que $x + y > 200$
  - La partie en dessous de la droite correspond à l'ensemble des points tels que $x + y < 200$
  - La partie sur la droite correspond à l'ensemble des points tels que $x + y = 200$
- La solution du problème sera en dessous ou sur la droite
- Répéter ce raisonnement pour les 5 contraintes donne une région convexe appelée un polyèdre. Cette région correspond à l'ensemble des points réalisables

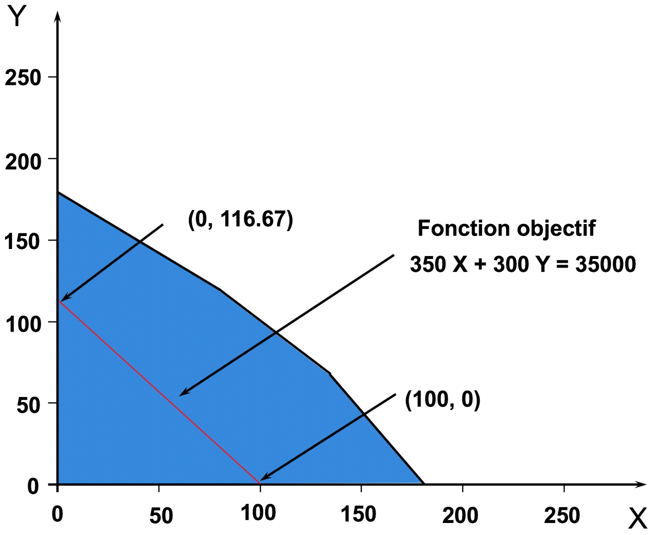
Polyèdre des contraintes



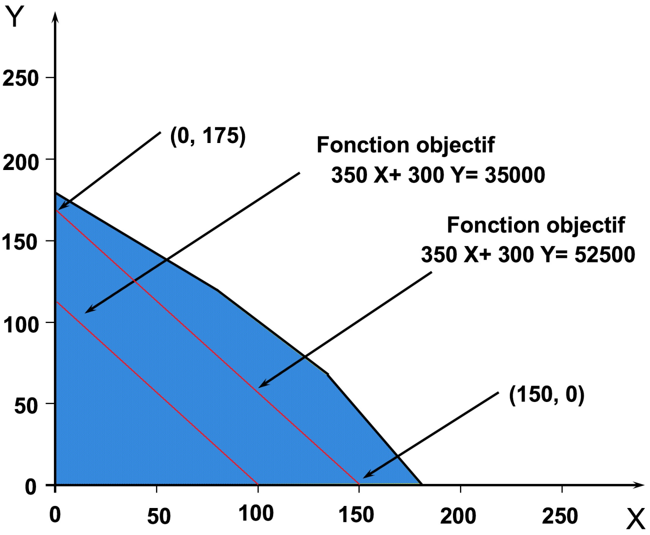
Interprétation graphique de l'objectif

- Considérons l'ensemble des points réalisables tels que:
  - $350x + 300y = 35000$
  - $350x + 300y = 52500$
  - $350x + 300y = 66100$
- Ce sont les courbes de niveau de l'objectif
- Jusqu'à quelle valeur peut-on aller ?

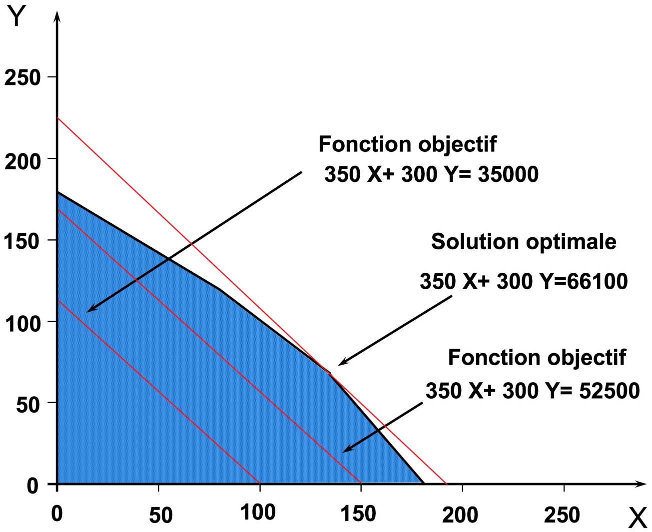
Courbes de niveau de l'objectif



Courbes de niveau de l'objectif



Courbes de niveau de l'objectif



Détermination de la solution

- On déduit des courbes de niveau que la solution optimale est située dans un “coin” de la région réalisable, à l’intersection de deux contraintes. Ce “coin” est appelé point extrême
- Si un problème a au moins une solution optimale, l’une d’entre-elles est un point extrême
- Il est possible de trouver une solution optimale en énumérant tous les points extrêmes

Détermination de la solution (suite)

- Graphiquement, la solution est atteinte lorsque les contraintes (1) et (2) sont actives (i.e. satisfaites à égalité)
- Ceci donne le système à deux équations et deux inconnues suivant:
$$\begin{cases} x + y & = 200 \\ 9x + 6y & = 1566 \end{cases}$$
- Dont la solution est le point extrême $(x^*, y^*) = (122, 78)$. C'est la solution optimale du problème, qui donne la valeur optimale de 66100

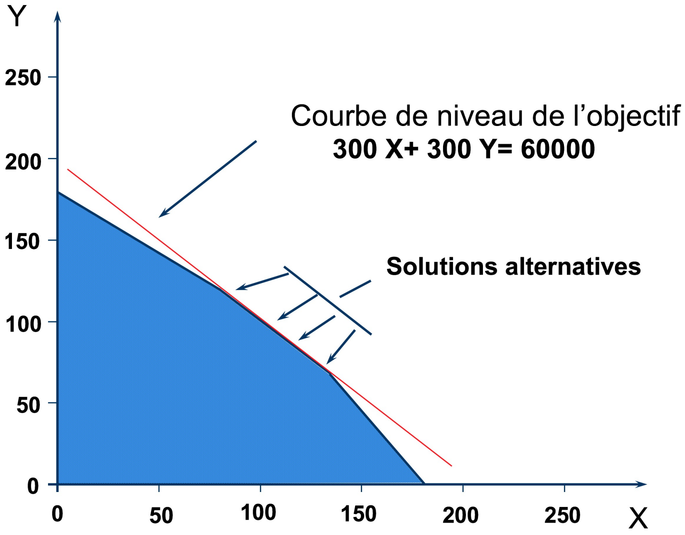
Plus d'une solution optimale

Si la fonction objectif avait été

$$300x + 300y$$

- Le point $(122, 78)$ donne une valeur de 60000
- Le point $(80, 120)$ donne une valeur de 60000
- Il n'y a pas de meilleur point et il y a en fait beaucoup de points équivalents: $(90, 110)$, $(95, 105)$, $(100, 100)$, ... Il y en a une infinité!
- Ceci arrive car les courbes de niveau de l'objectif sont parallèles à la contrainte (1). Toutes les solutions optimales se retrouvent le long de cette contrainte, entre deux points extrêmes

Infinité de solutions



La méthode du simplexe

- Origine: George Dantzig, 1947
- Un des 10 algorithmes du vingtième siècle selon *Computing in Science and Engineering*
- Ne pas confondre avec *l'autre méthode du simplexe* en optimisation sans dérivées [Nelder and Mead, 1965]
- Algorithme itératif qui se déplace d'un point extrême à un autre. Chaque déplacement améliore la qualité de la solution, et si l'algorithme se termine, alors on a la garantie d'avoir un optimum global

Illustration de la méthode du simplexe sur un exemple

$$\max_{x_1,x_2,x_3} 5x_1 + 4x_2 + 3x_3$$

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

Le problème doit être mis sous forme standard

Variables d'écart

Trois nouvelles variables (d'écart), positives ou nulles (une par contrainte) permettent d'obtenir des contraintes égalité:

$$\max_{x_1,x_2,x_3,e_1,e_2,e_3} \quad 5x_1 + 4x_2 + 3x_3$$

$$\text{s.c.} \begin{cases} 2x_1 + 3x_2 + x_3 & +e_1 & = 5 \\ 4x_1 + x_2 + 2x_3 & +e_2 & = 11 \\ 3x_1 + 4x_2 + 2x_3 & +e_3 & = 8 \\ x_1, x_2, x_3, e_1, e_2, e_3 \geq 0 \end{cases}$$

Dictionnaire initial

$$\begin{array}{rrrrr} e_1 = & 5 & -2x_1 & -3x_2 & -x_3 \\ e_2 = & 11 & -4x_1 & -x_2 & -2x_3 \\ e_3 = & 8 & -3x_1 & -4x_2 & -2x_3 \\ \hline z = & 0 & +5x_1 & +4x_2 & +3x_3 \end{array}$$

- Ce dictionnaire est une façon de représenter le point réalisable $\mathbf{x} = (x_1, x_2, x_3) = (0, 0, 0)$ avec $f(\mathbf{x}) = 0$ et $\mathbf{e} = (e_1, e_2, e_3) = (5, 11, 8)$. Il s'agit d'un point extrême
- Les variables en base e_1, e_2, et e_3 sont exprimées en fonction des variables hors-base x_1, x_2, et x_3
- Dans le point courant, les variables hors-base sont toujours nulles

Dictionnaire initial (suite)

$$
\begin{array}{rrrrr}
e_1 = & 5 & -2x_1 & -3x_2 & -x_3 \\
e_2 = & 11 & -4x_1 & -x_2 & -2x_3 \\
e_3 = & 8 & -3x_1 & -4x_2 & -2x_3 \\
\hline
z = & 0 & +5x_1 & +4x_2 & +3x_3
\end{array}
$$

- z représente la valeur courante de l'objectif. Ses coefficients (5,4, et 3) dans le dictionnaire sont appelés les coûts réduits
- S'il y a des coûts réduits positifs, on a intérêts à donner une valeur positive aux variables hors-base associées afin d'augmenter la valeur de z
- Par exemple, ici, si on peut poser $x_1 = 1$, alors z sera augmenté de 5

Première itération

$$
\begin{array}{rrrrr}
e_1 = & 5 & -2x_1 & -3x_2 & -x_3 \\
e_2 = & 11 & -4x_1 & -x_2 & -2x_3 \\
e_3 = & 8 & -3x_1 & -4x_2 & -2x_3 \\
\hline
z = & 0 & +5x_1 & +4x_2 & +3x_3
\end{array}
$$

- On décide d'augmenter la valeur de x_1 et de laisser x_2 et x_3 à zéro. Si on prend x_1 trop grand, les variables en base peuvent devenir négatives. Quelle valeur maximale peut-on choisir ?
- On se sert des contraintes $e_1, e_2, e_3 \geq 0$ où on isole x_1 et on laisse x_2 et x_3 à zéro. Pour e_1, cela donne:

$$5 - 2x_1 \geq 0 \rightarrow 2x_1 \leq 5 \rightarrow x_1 \leq \frac{5}{2}$$

Première itération (suite)

- Pour les trois contraintes, cela donne les inégalités suivantes:

$$x_1 \leq \frac{5}{2} \text{ et } x_1 \leq \frac{11}{4} \text{ et } x_1 \leq \frac{8}{3}$$

- La valeur maximale que peut prendre x_1 est donc

$$x_1 = \min\left\{\frac{5}{2}, \frac{11}{4}, \frac{8}{3}\right\} = \frac{5}{2}$$

  qui correspond à la contrainte $e_1 \geq 0$

- Donc si on pose $x_1 = \frac{5}{2}$, alors on aura $e_1 = 0$. x_1 va entrer en base et e_1 va sortir de base

- Cette opération s'appelle un pivot et va mener au second dictionnaire, et à un nouveau et meilleur point extrême

Première itération: Pivot

- Il faut d'abord effectuer le pivot en exprimer la nouvelle variable de base x_1 en fonction de l'ancienne e_1:

$$e_1 = 5 - 2x_1 - 3x_2 - x_3 \rightarrow x_1 = \frac{5}{2} - \frac{3}{2}x_2 - \frac{1}{2}x_3 - \frac{1}{2}e_1$$

- Il faut ensuite écrire les autres lignes du dictionnaires dans lesquelles on remplace x_1 par sa nouvelle expression. Pour e_2, cela donne: $e_2 = 11 - 4\left(\frac{5}{2} - \frac{3}{2}x_2 - \frac{1}{2}x_3 - \frac{1}{2}e_1\right) - x_2 - 2x_3 = 1 + 5x_2 + 0x_3 + 2e_1$

Première itération: Nouveau dictionnaire

Après avoir fait la même opération pour e_3 et z, le pivot est complété et on obtient le second dictionnaire (première itération):

$$
\begin{array}{rcrrrr}
x_1 & = & \frac{5}{2} & -\frac{3}{2}x_2 & -\frac{1}{2}x_3 & -\frac{1}{2}e_1 \\
e_2 & = & 1 & +5x_2 & +0x_3 & +2e_1 \\
e_3 & = & \frac{1}{2} & +\frac{1}{2}x_2 & -\frac{1}{2}x_3 & +\frac{3}{2}e_1 \\
\hline
z & = & \frac{25}{2} & -\frac{7}{2}x_2 & +\frac{1}{2}x_3 & -\frac{5}{2}e_1
\end{array}
$$

qui correspond au point extrême $\mathbf{x} = (5/2, 0, 0)$ pour la valeur courante de $f(\mathbf{x}) = 25/2$, avec les écarts $\mathbf{e} = (0, 1, 1/2)$ (on a bien un point réalisable)

Première itération: Fin

- Le coût réduit associé à la nouvelle variable hors-base (e_1) est forcément négatif ou nul. Si on refaisait entrer cette variable en base, on obtiendrait le point extrême précédent

- **Critère d'arrêt:** Comme il reste des coûts réduits positifs, cela signifie que l'on peut améliorer la valeur de f

Deuxième itération

$$
\begin{array}{llllll}
x_1 = & \frac{5}{2} & -\frac{3}{2}x_2 & -\frac{1}{2}x_3 & -\frac{1}{2}e_1 \\
e_2 = & 1 & +5x_2 & +0x_3 & +2e_1 \\
e_3 = & \frac{1}{2} & +\frac{1}{2}x_2 & -\frac{1}{2}x_3 & +\frac{3}{2}e_1 \\
\hline
z = & \frac{25}{2} & -\frac{7}{2}x_2 & +\frac{1}{2}x_3 & -\frac{5}{2}e_1
\end{array}
$$

- On fait entrer x_3 en base. On a $x_3 = \min\{1, 5\} = 1$. C'est e_3 qui sort de base
- Le pivot donne $x_3 = 1 + x_2 + 3e_1 - 2e_3$
- Ce qui permet d'obtenir le nouveau dictionnaire

Deuxième itération: Nouveau dictionnaire et fin

$$
\begin{array}{rrrrr}
x_1 = & 2 & -2x_2 & -2e_1 & +e_3 \\
e_2 = & 1 & +5x_2 & +2e_1 & +0e_3 \\
x_3 = & 1 & +x_2 & +3e_1 & -2e_3 \\
\hline
z = & 13 & -3x_2 & -e_1 & -e_3
\end{array}
$$

- Ce dictionnaire correspond au point extrême $\mathbf{x} = (2, 0, 1)$ avec les écarts $\mathbf{e} = (0, 1, 0)$ et $f(\mathbf{x}) = 13$
- **Tous les coûts réduits sont négatifs**, donc on ne peut plus augmenter la valeur de z. **L'algorithme s'arête** et on a la garantie que $\mathbf{x}^* = (2, 0, 1)$ est la solution optimale, pour une valeur optimale de $f(\mathbf{x}^*) = 13$

Forme tableau

- Dictionnaire:

$$
\begin{array}{rrrrr}
x_1 = & 2 & -2x_2 & -2e_1 & +e_3 \\
e_2 = & 1 & +5x_2 & +2e_1 & +0e_3 \\
x_3 = & 1 & +x_2 & +3e_1 & -2e_3 \\
\hline
z = & 13 & -3x_2 & -e_1 & -e_3
\end{array}
$$

- Tableau:

	Base	5	4	3	0	0	0	
c_j	variable	x_1	x_2	x_3	e_1	e_2	e_3	Valeur
5	x_1	1	2	0	2	0	-1	2
0	e_2	0	-5	0	-2	1	0	1
3	x_3	0	-1	1	-3	0	2	1
coûts réduits		0	-3	0	-1	0	-1	13

Algorithme du simplexe

[0] Initialisation

- Trouver un point extrême réalisable (phase 1)
- Si (échec de la phase 1): Stop (pas de solution)
- Former le dictionnaire initial
- $k \leftarrow 1$

[1] Itération k

- Choix de la variable entrante
- Choix de la variable sortante
- Pivot $\rightarrow$ nouveau dictionnaire
- Si (tous les coûts réduits ≤ 0): Stop (sol. optimale)
- $k \leftarrow k + 1$
- Aller en [1]

Il convient aussi de définir des règles pour les choix des variables entrantes et sortantes

Problème et solutions

- On cerche à résoudre

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

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

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

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

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

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

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

Dérivées

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

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

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

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

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

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

Direction de descente

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

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

- Pour tout $\alpha \in \mathbb{R}$ petit, on aura $h(\alpha) = f(\mathbf{x} + \alpha \mathbf{d}) < f(\mathbf{x})$ et $h'(0) < 0$
- Principe de la *line search* (recherche linéaire): Trouver α tel que $h'(\alpha) = 0$

Signe d'une matrice

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

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

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

Optimisation sans contraintes: CN1

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

Condition nécessaire d'optimalité de premier ordre

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

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

Optimisation sans contraintes: CN2

Condition nécessaire de second ordre

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

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

$$\begin{aligned} f(\mathbf{x}^*) \leq f(\mathbf{x}^* + t\mathbf{d}) \quad &\simeq f(\mathbf{x}^*) + t\mathbf{d}^\top \nabla f(\mathbf{x}^*) + \frac{t^2}{2}\mathbf{d}^\top \nabla^2 f(\mathbf{x}^*)\mathbf{d} \\ &= f(\mathbf{x}^*) + \frac{t^2}{2}\mathbf{d}^\top \nabla^2 f(\mathbf{x}^*)\mathbf{d} \\ \\ &\Rightarrow \mathbf{d}^\top \nabla^2 f(\mathbf{x}^*)\mathbf{d} \geq 0 \end{aligned}$$

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

Optimisation sans contraintes: CS2

Condition suffisante de second ordre

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

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

Optimisation sans contraintes: Convexité

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

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

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

Optimisation sans contraintes: Méthode du gradient

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

[0] Initialisation

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

$k \leftarrow 0$

[1] Itération k

Calculer $\mathbf{d}^k = -\nabla f(\mathbf{x}^k)$ (dir. de descente)

Si $(\mathbf{d}^k = \mathbf{0})$: Stop (point critique)

Trouver $\alpha^k \in \arg\min_{\alpha \geq 0} h(\alpha) = f(\mathbf{x}^k + \alpha \mathbf{d}^k)$

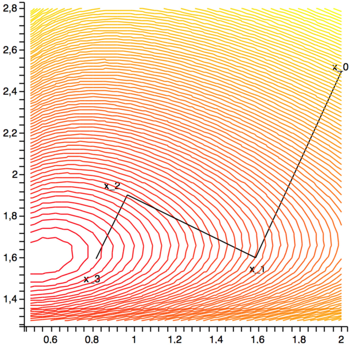
$\mathbf{x}^{k+1} \leftarrow \mathbf{x}^k + \alpha^k \mathbf{d}^k$

$k \leftarrow k + 1$

Aller en **[1]**

Méthode du gradient: Remarques

- Lorsque la minimisation de h est faite de façon exacte, les directions consécutives $\mathbf{d}^k$ et $\mathbf{d}^{k+1}$ sont perpendiculaires: on s'arrête toujours de façon tangente à une courbe de niveau
- La méthode peut prendre un nombre considérable d'itérations avant de converger à un point critique



Méthode de Newton

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

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

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

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

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

Méthode quasi-Newton

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

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

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

Optimisation avec contraintes

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

Théorème

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

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

Optimisation avec une contrainte égalité

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

CN1

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

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

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

Optimisation avec une contrainte inégalité

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

CN1

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

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

Optimisation avec plusieurs contraintes égalité

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

CN1
Si $\mathbf{x}^*$ est un minimum local de f dans Ω où $\{\nabla c_i(\mathbf{x}^*) : i \in \mathcal{E}\}$ est un ensemble linéairement indépendant, alors $c_i(\mathbf{x}^*) = 0$ pour tout $i \in \mathcal{E}$ et il existe $\lambda \in \mathbb{R}^m$ tel que

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

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

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

Multiplicateurs de Lagrange

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

Optimisation avec contraintes: Cas général

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

avec

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

et

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

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

Géométrie de l'ensemble réalisable

- $\mathbf{d} \in \mathbb{R}^n$, $\mathbf{d} \neq 0$, est une direction réalisable à partir de $\mathbf{x} \in \Omega$ s'il existe $\varepsilon > 0$ tel que pour tout $t \in]0; \varepsilon[$, $\mathbf{x} + t\mathbf{d} \in \Omega$
- Un ensemble $\mathcal{K} \subseteq \mathbb{R}^n$ est appelé un cône si pour tout $\mathbf{d} \in \mathcal{K}$ et tout $\lambda \geq 0$, $\lambda\mathbf{d} \in \mathcal{K}$
- L'ensemble de toutes les directions réalisables à partir de $\mathbf{x} \in \Omega$ forme un cône
- Le cône polaire du cône $\mathcal{K} \subseteq \mathbb{R}^n$ est

$$\mathcal{K}^* = \left\{ \mathbf{d} \in \mathbb{R}^n : \mathbf{d}^\top \mathbf{v} \leq 0 \ : \ \mathbf{v} \in \mathcal{K} \right\}$$

- Le polaire du polaire est le cône de départ: $\mathcal{K}^{**} = \mathcal{K}$

Ensemble des contraintes actives

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

- Pour tout $\mathbf{x} \in \mathbb{R}^n$, l'ensemble des contraintes actives $\mathcal{A}(\mathbf{x})$ est l'ensemble des indices des contraintes inégalité satisfaites à égalité en $\mathbf{x}$:

$$\mathcal{A}(\mathbf{x}) = \{i \in \mathcal{I} : c_i(\mathbf{x}) = 0\}$$

- Attention: Même si les contraintes égalité sont toujours actives, elles ne sont pas représentées par $\mathcal{A}(\mathbf{x})$

Qualification de contraintes (1/2)

- Permet d'exprimer le cône tangent en linéarisant les contraintes et d'obtenir des formulations analytiques des cônes tangent et normal
- Condition basique de qualification de contraintes: Avec

$$\Lambda(\mathbf{x}) = \left\{ \lambda \in \mathbb{R}^{m+p} \;\middle|\; \begin{array}{ll} \lambda_i \in \mathbb{R} & i \in \mathcal{E} \\ \lambda_i \geq 0 & i \in \mathcal{A}(\mathbf{x}) \\ \lambda_i = 0 & i \in \mathcal{I} \setminus \mathcal{A}(\mathbf{x}) \end{array} \right\}$$

la condition basique de qualification de contraintes est satisfaite en $\mathbf{x} \in \Omega$ ssi le seul $\lambda \in \Lambda(\mathbf{x})$ tel que

$$\sum_{i \in \mathcal{E}} \lambda_i \nabla c_i(\mathbf{x}) - \sum_{i \in \mathcal{A}(\mathbf{x})} \lambda_i \nabla c_i(\mathbf{x}) = \mathbf{0}$$

est $\mathbf{0}$

Qualification de contraintes (2/2)

- Cette condition n'est pas facile à vérifier et d'autres conditions (plus fortes) peuvent être employées: Par exemple, la LICQ (*Linear Independence Constraint Qualification*) est satisfaite en $\mathbf{x} \in \Omega$ si $\{\nabla c_i(\mathbf{x}) : i \in \mathcal{E} \cup \mathcal{A}(\mathbf{x})\}$ est un ensemble de vecteurs linéairement indépendants

- Dans tout ce qui suit, on suppose que la condition basique est vérifiée

Cône tangent

- Définition géométrique:
  - $\mathbf{d} \in \mathbb{R}^n$ est un vecteur tangent à Ω en $\mathbf{x} \in \mathbb{R}^n$ s'il existe une suite $\{\mathbf{z}_k\}$ de points réalisables avec $\mathbf{z}_k \to \mathbf{x}$ et une suite de réels positifs $\{t_k\}$ avec $t_k \to 0$ tels que

$$\lim_{k\to\infty} \frac{\mathbf{z}_k - \mathbf{x}}{t_k} = \mathbf{d}$$

  - L'ensemble des vecteurs tangents forme le cône tangent $T_\Omega(\mathbf{x})$

- Définition algébrique:

$$T_\Omega(\mathbf{x}) = \left\{ \mathbf{d} \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} \mathbf{d}^\top \nabla c_i(\mathbf{x}) = 0 & i \in \mathcal{E} \\ \mathbf{d}^\top \nabla c_i(\mathbf{x}) \geq 0 & i \in \mathcal{A}(\mathbf{x}) \end{array} \right\}$$

- Le cône tangent correspond aux directions réalisables de premier ordre (i.e. lorsque tout est linéarisé)

Cône normal

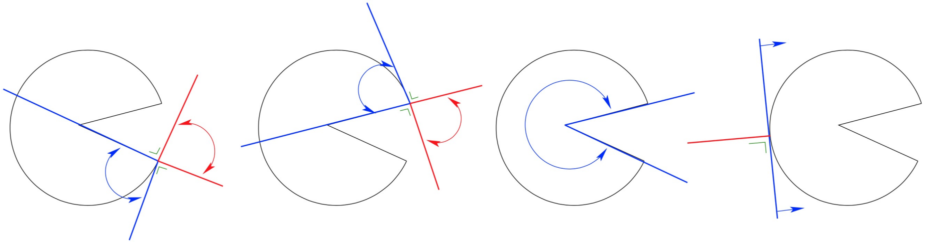
- Définition géométrique: Le cône normal est le polaire du cône tangent:

$$N^*_\Omega(\mathbf{x}) = T_\Omega(\mathbf{x})$$

- Définition algébrique:

$$N_\Omega(\mathbf{x}) = \left\{ \sum_{i\in\mathcal{E}} \lambda_i \nabla c_i(\mathbf{x}) - \sum_{i\in\mathcal{A}(\mathbf{x})} \lambda_i \nabla c_i(\mathbf{x}) \;\middle|\; \begin{array}{ll} \lambda_i \in \mathbb{R} & i \in \mathcal{E} \\ \lambda_i \geq 0 & i \in \mathcal{A}(\mathbf{x}) \end{array} \right\}$$

Exemples de cônes tangent et normal [Orban, 2010]



- Les cônes tangents sont en bleu et les cônes normaux en rouge
- Dans le 3ème cas, le cône normal est réduit à $\{\mathbf{0}\}$
- La pointe de chaque cône devrait correspondre à l'origine, mais ils sont ici translatés

Condition nécessaire d'optimalité de premier ordre

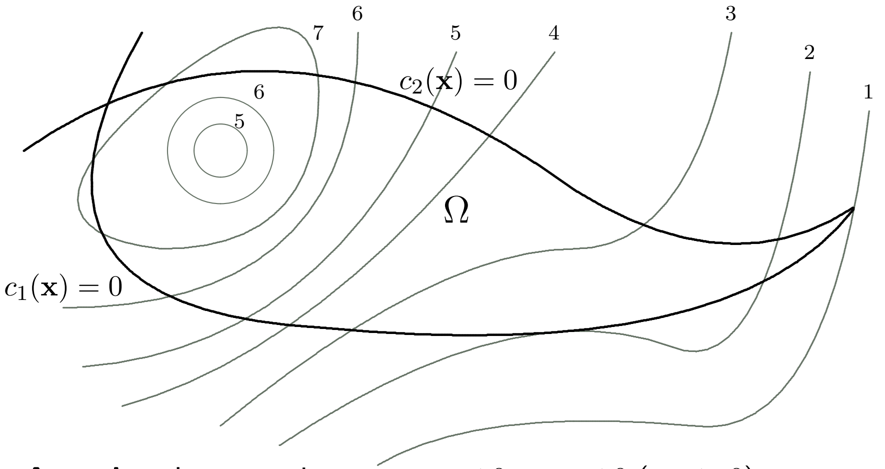
Idée: Si $\mathbf{x}^*$ est un minimum local, alors il n'existe pas de direction $\mathbf{d} \in T_\Omega(\mathbf{x}^*)$ ($\simeq$ dir. réalisable) qui soit une direction de descente, i.e. telle que $\mathbf{d}^\top \nabla f(\mathbf{x}^*) < 0$

CN1

Si $\mathbf{x}^*$ est un minimum local de f dans Ω, alors

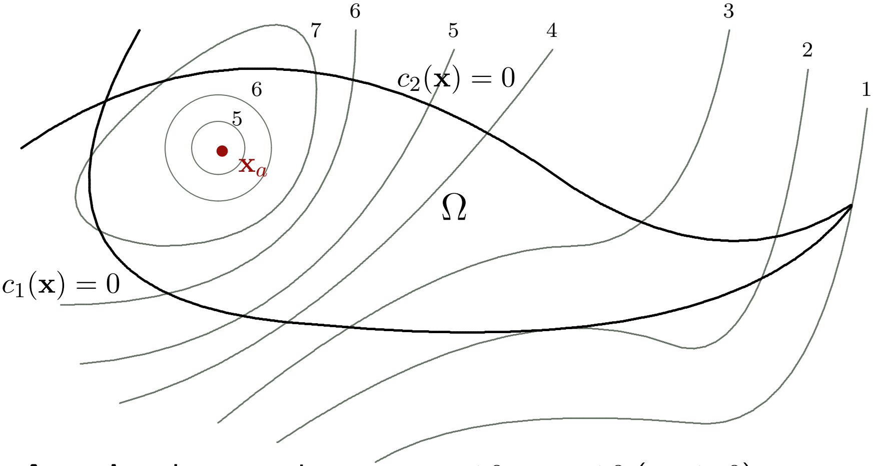
$$-\nabla f(\mathbf{x}^*) \in N_\Omega(\mathbf{x}^*)$$

Interprétation graphique (1/2)



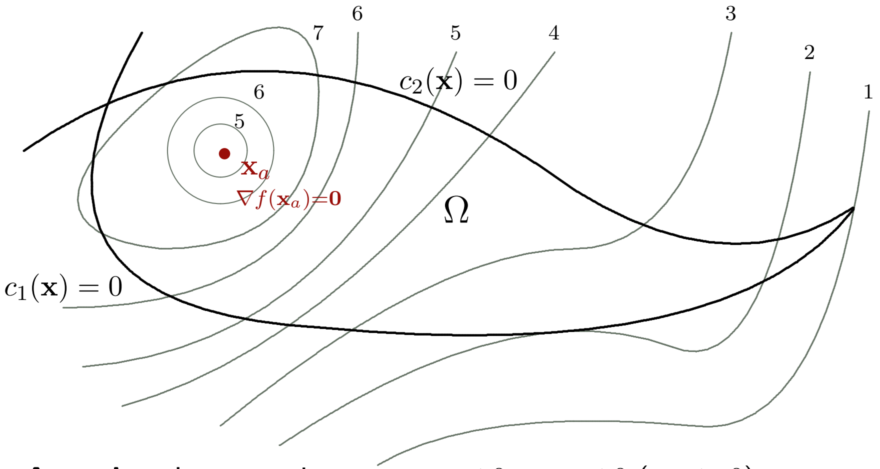
Attention: Les contraintes sont $c_1 \leq 0$ et $c_2 \leq 0$ (pas ≥ 0)

Interprétation graphique (1/2)



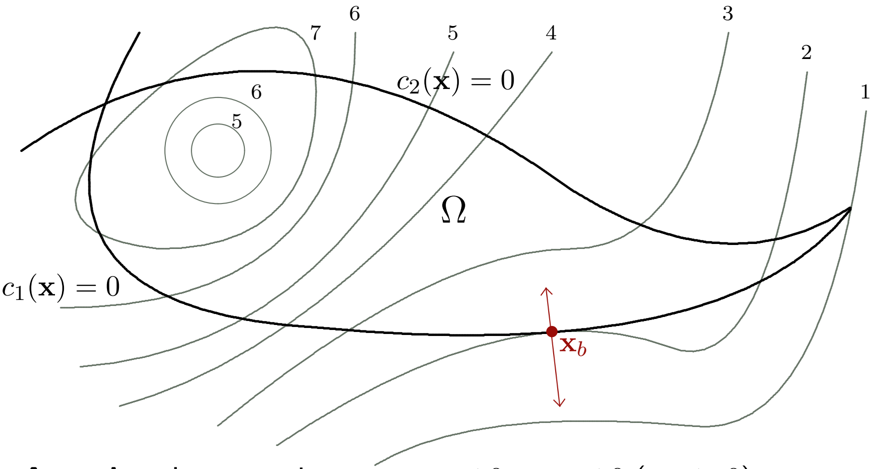
Attention: Les contraintes sont $c_1 \leq 0$ et $c_2 \leq 0$ (pas ≥ 0)

Interprétation graphique (1/2)



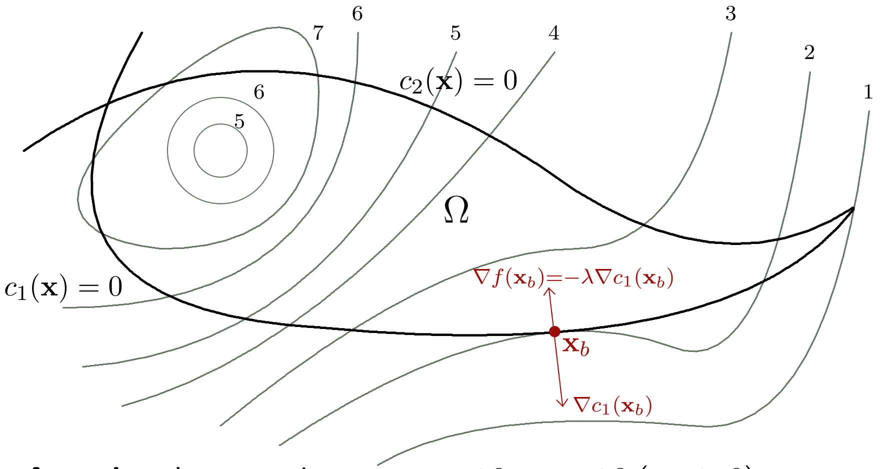
Attention: Les contraintes sont $c_1 \leq 0$ et $c_2 \leq 0$ (pas ≥ 0)

Interprétation graphique (1/2)



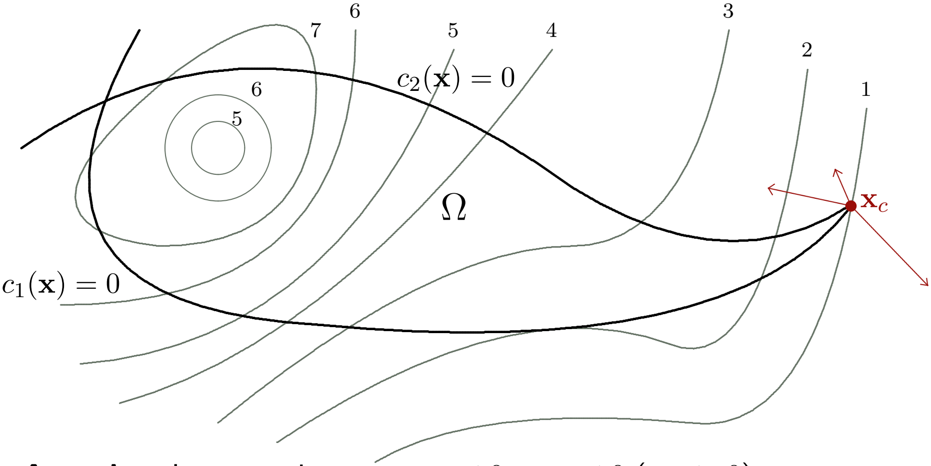
Attention: Les contraintes sont $c_1 \leq 0$ et $c_2 \leq 0$ (pas ≥ 0)

Interprétation graphique (1/2)



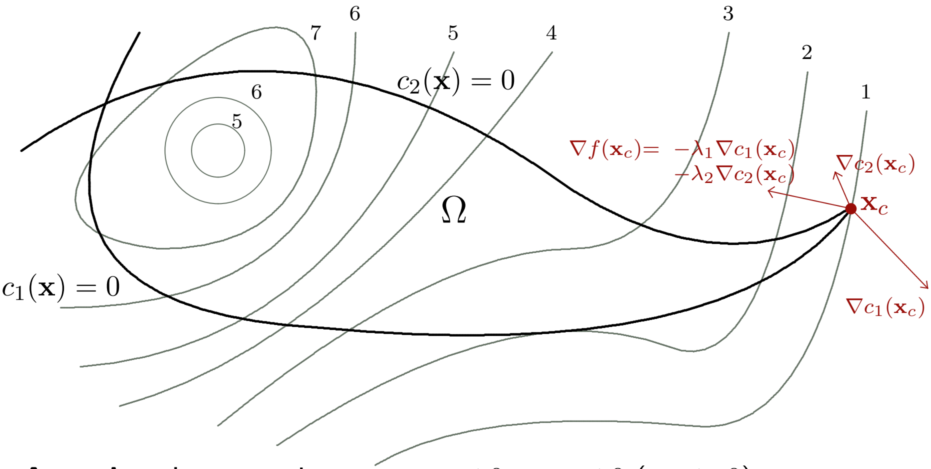
Attention: Les contraintes sont $c_1 \leq 0$ et $c_2 \leq 0$ (pas ≥ 0)

Interprétation graphique (1/2)



Attention: Les contraintes sont $c_1 \leq 0$ et $c_2 \leq 0$ (pas ≥ 0)

Interprétation graphique (1/2)



Attention: Les contraintes sont $c_1 \leq 0$ et $c_2 \leq 0$ (pas ≥ 0)

Interprétation graphique (2/2)

- $\nabla f(\mathbf{x}_a) = \mathbf{0}$, $\mathcal{A}(\mathbf{x}_a) = \emptyset$, $N_\Omega(\mathbf{x}_a) = \{0\}$, $T_\Omega(\mathbf{x}_a) = \mathbb{R}^2$
- $\mathcal{A}(\mathbf{x}_b) = \{1\}$, $N_\Omega(\mathbf{x}_b)$ est réduit à la demi-droite dans la direction $\nabla c_1(\mathbf{x}_b)$, $T_\Omega(\mathbf{x}_b)$ est l'union de $\{\mathbf{0}\}$ et du demi-espace orthogonal à $\nabla c_1(\mathbf{x}_b)$
- $\mathcal{A}(\mathbf{x}_c) = \{1, 2\}$, $N_\Omega(\mathbf{x}_c) = \{\lambda_1 \nabla c_1(\mathbf{x}_c) + \lambda_2 \nabla c_2(\mathbf{x}_c) : \lambda_1, \lambda_2 \geq 0\}$

CN1: Conditions de KKT

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

CN1

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

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

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

Dualité

Fonction Lagrangienne (ou Lagrangien):

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

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

Théorème de dualité faible

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

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

La meilleure borne inférieure est donc donnée par le problème dual

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

Cas de l'optimisation linéaire (1/3)

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

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

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

- Une borne inférieure est donnée par

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

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

Cas de l'optimisation linéaire (2/3)

Les conditions KKT sont

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

Obtenir la meilleure borne inférieure revient donc à résoudre

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

Cas de l'optimisation linéaire (3/3)

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

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

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

qui correspond à ce qui est vu (normalement) en OL

Méthodes de pénalités

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

Exemples de pénalités

- Pénalité quadratique:

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

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

- Pénalité ℓ_1:

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

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

Algorithme du Lagrangien augmenté (1/2)

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

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

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

Algorithme du Lagrangien augmenté (2/2)

[0] Initialisation

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

[1] Itération k

- Trouver approximativement $\mathbf{x}^{k+1} \in \underset{\mathbf{x} \in \mathbb{R}^n}{\arg\min}\, L_a(\mathbf{x}, \lambda^k, \mu^k)$

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

Extensions

- Cas sans contrainte: Méthodes de régions de confiance
- Avec contraintes: Conditions de second ordre (CN2 et CS2)
- Méthode de point intérieurs
- Optimisation globale
- Optimisation sans dérivées (DFO): Que faire quand f n'est pas différentiable?
- Optimisation de boîtes noires (BBO): Que faire quand f est donnée par une boîte noire ?

Introduction de cette introduction

Optimisation linéaire

Optimisation non linéaire

References

References I

Audet, C. (2021).
Optimisation continue.
Presses internationales Polytechnique, Montréal, Canada.

Gauvin, J. (1995).
Leçons de programmation mathématique.
Éditions de l'École Polytechnique de Montréal.

Nelder, J. and Mead, R. (1965).
A Simplex Method for Function Minimization.
The Computer Journal, 7(4):308–313.

Nocedal, J. and Wright, S. (2006).
Numerical Optimization.
Springer Series in Operations Research and Financial Engineering. Springer, Berlin, second edition.

Orban, D. (2010).
Numerical Methods for Nonlinear Optimization and Optimal Control, notes du cours MTH8408.