

Optimisation

M1 Informatique

Julien Ah-Pine (julien.ah-pine@eric.univ-lyon2.fr)

Université Lyon 2 - ICOM

M1 Informatique 2014-2015

Motivations

Pourquoi un cours sur l'optimisation ?

- Ensemble de techniques centrales dans tout processus de prise de décisions en ingénierie ou en économie
- Beaucoup de branches scientifiques utilisent des méthodes d'optimisation : statistiques, recherche opérationnelle, économie, fouille de données, apprentissage statistique. . .

Motivations

Pourquoi un cours sur l'optimisation ?

- Ensemble de techniques centrales dans tout processus de prise de décisions en ingénierie ou en économie
- Beaucoup de branches scientifiques utilisent des méthodes d'optimisation : statistiques, recherche opérationnelle, économie, fouille de données, apprentissage statistique. . .

Formulation générale d'un problème pratique d'optimisation

- Souvent les tâches se résument à choisir parmi plusieurs alternatives celle(s) qui est(ont) la(les) meilleure(s)
- La mesure de la qualité d'une alternative est donnée par une fonction dite objectif
- Le problème peut ou pas posséder des contraintes que les alternatives doivent satisfaire

Motivations

Pourquoi un cours sur l'optimisation ?

- Ensemble de techniques centrales dans tout processus de prise de décisions en ingénierie ou en économie
- Beaucoup de branches scientifiques utilisent des méthodes d'optimisation : statistiques, recherche opérationnelle, économie, fouille de données, apprentissage statistique. . .

Formulation générale d'un problème pratique d'optimisation

- Souvent les tâches se résument à choisir parmi plusieurs alternatives celle(s) qui est(ont) la(les) meilleure(s)
 - La mesure de la qualité d'une alternative est donnée par une fonction dite objectif
 - Le problème peut ou pas posséder des contraintes que les alternatives doivent satisfaire
- ⇒ Objectif général du cours : présenter les différents types de problèmes classiques d'optimisation et quelques méthodes de résolution

Organisation du cours

- 10 séances de CM et TD en alternance
- 2 séances 1h45 sur machine (introduction au langage AMPL)

Organisation du cours

- 10 séances de CM et TD en alternance
- 2 séances 1h45 sur machine (introduction au langage AMPL)

- 1 évaluation relative aux séances sur machine ($\sim 20\%$)
- 1 épreuve écrite individuelle ($\sim 80\%$)

Organisation du cours

- 10 séances de CM et TD en alternance
- 2 séances 1h45 sur machine (introduction au langage AMPL)

- 1 évaluation relative aux séances sur machine ($\sim 20\%$)
- 1 épreuve écrite individuelle ($\sim 80\%$)

- Pour les supports de cours, rejoignez le groupe M1_INFO_OPT-1415

Références

- Sources principales de ce cours :
 - ▶ E.K. Chong and S.H. Zak (2001). An Introduction to Optimization. Wiley.
 - ▶ R. Bronson and G. Naadimuthu (2007). Operations research. McGraw-Hill.
- Autres références :
 - ▶ J. Nocedal and S. Wright (2006). Numerical Optimization. Springer.
 - ▶ Y. Dodge (2005). Optimisation Appliquée. Springer.
 - ▶ ...

Problèmes d'optimisation

- Formellement nous nous intéressons au problème général suivant :

$$\min f(\mathbf{x})$$

sous la contrainte que $\mathbf{x} \in \mathbb{S}$

où :

- $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est appelé **fonction objectif** ou **fonction de coût**
- \mathbf{x} est un vecteur de \mathbb{R}^n appelé vecteur des **variables de décision**
- \mathbb{S} est un sous-ensemble de \mathbb{R}^n et est appelé **ensemble réalisable**

Problèmes d'optimisation

- Formellement nous nous intéressons au problème général suivant :

$$\min f(\mathbf{x})$$

sous la contrainte que $\mathbf{x} \in \mathbb{S}$

où :

- $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est appelé **fonction objectif** ou **fonction de coût**
- \mathbf{x} est un vecteur de \mathbb{R}^n appelé vecteur des **variables de décision**
- \mathbb{S} est un sous-ensemble de \mathbb{R}^n et est appelé **ensemble réalisable**
- Un problème d'optimisation peut être vu comme un problème de décision où on cherche la meilleure solution \mathbf{x}^* parmi un ensemble de décisions possibles \mathbb{S} . La qualité d'une solution \mathbf{x} est mesurée par la fonction $f(\mathbf{x})$. Dans le cas d'une minimisation, la meilleure solution \mathbf{x}^* est telle que $f(\mathbf{x}^*) \leq f(\mathbf{x}), \forall \mathbf{x} \in \mathbb{S}$

Problèmes d'optimisation

- Formellement nous nous intéressons au problème général suivant :

$$\min f(\mathbf{x})$$

sous la contrainte que $\mathbf{x} \in \mathbb{S}$

où :

- $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est appelé **fonction objectif** ou **fonction de coût**
- \mathbf{x} est un vecteur de \mathbb{R}^n appelé vecteur des **variables de décision**
- \mathbb{S} est un sous-ensemble de \mathbb{R}^n et est appelé **ensemble réalisable**
- Un problème d'optimisation peut être vu comme un problème de décision où on cherche la meilleure solution \mathbf{x}^* parmi un ensemble de décisions possibles \mathbb{S} . La qualité d'une solution \mathbf{x} est mesurée par la fonction $f(\mathbf{x})$. Dans le cas d'une minimisation, la meilleure solution \mathbf{x}^* est telle que $f(\mathbf{x}^*) \leq f(\mathbf{x}), \forall \mathbf{x} \in \mathbb{S}$
- \mathbf{x}^* est alors appelé minimiseur de f sur \mathbb{S}

Problèmes d'optimisation

- Formellement nous nous intéressons au problème général suivant :

$$\min f(\mathbf{x})$$

sous la contrainte que $\mathbf{x} \in \mathbb{S}$

où :

- $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est appelé **fonction objectif** ou **fonction de coût**
- \mathbf{x} est un vecteur de \mathbb{R}^n appelé vecteur des **variables de décision**
- \mathbb{S} est un sous-ensemble de \mathbb{R}^n et est appelé **ensemble réalisable**
- Un problème d'optimisation peut être vu comme un problème de décision où on cherche la meilleure solution \mathbf{x}^* parmi un ensemble de décisions possibles \mathbb{S} . La qualité d'une solution \mathbf{x} est mesurée par la fonction $f(\mathbf{x})$. Dans le cas d'une minimisation, la meilleure solution \mathbf{x}^* est telle que $f(\mathbf{x}^*) \leq f(\mathbf{x}), \forall \mathbf{x} \in \mathbb{S}$
- \mathbf{x}^* est alors appelé minimiseur de f sur \mathbb{S}
- Dans le cas d'une maximisation de la fonction objectif f , le problème revient à résoudre $\min -f(\mathbf{x})$ slc $\mathbf{x} \in \mathbb{S}$. Donc on peut s'intéresser aux problèmes de minimisation uniquement sans perte de généralité

Problèmes d'optimisation (suite)

- Dans le cas particulier où $\mathcal{S} = \mathbb{R}^n$ on dira que le problème est **non contraint**

Problèmes d'optimisation (suite)

- Dans le cas particulier où $\mathcal{S} = \mathbb{R}^n$ on dira que le problème est **non contraint**
- Si $\mathcal{S} \subset \mathbb{R}^n$ on parle d'un problème **contraint** et souvent l'ensemble réalisable est de la forme suivante :
$$\mathcal{S} = \{\mathbf{x} \in \mathbb{R}^n : h(\mathbf{x}) = \mathbf{0}, g(\mathbf{x}) \leq \mathbf{0}\}$$
où $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ et $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$ sont des applications données définissant les contraintes du problème et donc les solutions réalisables

Problèmes d'optimisation (suite)

- Dans le cas particulier où $\mathcal{S} = \mathbb{R}^n$ on dira que le problème est **non contraint**
- Si $\mathcal{S} \subset \mathbb{R}^n$ on parle d'un problème **contraint** et souvent l'ensemble réalisable est de la forme suivante :
$$\mathcal{S} = \{\mathbf{x} \in \mathbb{R}^n : h(\mathbf{x}) = \mathbf{0}, g(\mathbf{x}) \leq \mathbf{0}\}$$
où $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ et $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$ sont des applications données définissant les contraintes du problème et donc les solutions réalisables

- Il existe plusieurs sous-types de problèmes d'optimisation :
 - ▶ Problème non contraint : $\mathcal{S} = \mathbb{R}^n$
 - ▶ Problème avec contraintes d'égalités : $\mathcal{S} = \{\mathbf{x} \in \mathbb{R}^n : h(\mathbf{x}) = \mathbf{0}\}$
 - ▶ Problème avec contraintes d'inégalités : $\mathcal{S} = \{\mathbf{x} \in \mathbb{R}^n : g(\mathbf{x}) \leq \mathbf{0}\}$
 - ▶ Problème contraint : $\mathcal{S} = \{\mathbf{x} \in \mathbb{R}^n : h(\mathbf{x}) = \mathbf{0}, g(\mathbf{x}) \leq \mathbf{0}\}$
 - ▶ Problème linéaire : f, g, h sont des applications linéaires
 - ▶ Problème non linéaire : f, g, h ne sont pas des applications linéaires
 - ▶ Problème discret (optimisation combinatoire) : \mathbf{x} doit être un vecteur d'entiers
 - ▶ ...

Plan du cours

- 1 Rappels de concepts mathématiques
- 2 Bases théoriques de l'optimisation
- 3 Optimisation sans contrainte
- 4 Optimisation linéaire avec contraintes
- 5 Optimisation non linéaire avec contraintes
- 6 Optimisation convexe
- 7 Algorithmes pour problèmes contraints

Rappel du Sommaire

- 1 Rappels de concepts mathématiques
- 2 Bases théoriques de l'optimisation
- 3 Optimisation sans contrainte
- 4 Optimisation linéaire avec contraintes
- 5 Optimisation non linéaire avec contraintes
- 6 Optimisation convexe
- 7 Algorithmes pour problèmes contraints

Espaces vectoriels réels I

- On définit un **vecteur** colonne \mathbf{x} de taille n de la manière suivante :

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

Nous utiliserons également la notation $\mathbf{x} = (x_1, \dots, x_n)$

- Nous travaillerons essentiellement dans \mathbb{R}^n donc les x_i sont des réels
- Nous noterons par \mathbf{x}^t la transposée du vecteur colonne \mathbf{x} :

$$\mathbf{x}^t = (x_1 \quad x_2 \quad \dots \quad x_n)$$

- Sauf mention contraire, un vecteur \mathbf{x} est nécessairement un vecteur colonne de taille $(n \times 1)$
- Deux vecteurs \mathbf{x} et \mathbf{y} sont égaux ssi $\forall i = 1, \dots, n : x_i = y_i$

Rappel du Sommaire

- 1 Rappels de concepts mathématiques
 - Quelques éléments d'algèbre linéaire
 - Quelques éléments de géométrie
 - Quelques éléments d'analyse
- 2 Bases théoriques de l'optimisation
- 3 Optimisation sans contrainte
- 4 Optimisation linéaire avec contraintes
- 5 Optimisation non linéaire avec contraintes
- 6 Optimisation convexe
- 7 Algorithmes pour problèmes contraints

Espaces vectoriels réels II

- La **somme** de deux vecteurs \mathbf{x} et \mathbf{y} de \mathbb{R}^n est un vecteur \mathbf{z} de \mathbb{R}^n défini par :

$$\mathbf{z} = \mathbf{x} + \mathbf{y} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}$$

- Propriétés de la loi interne d'addition entre vecteurs de \mathbb{R}^n
 - ▶ $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n : (\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$ (associativité)
 - ▶ $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n : \mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ (commutativité)
 - ▶ le vecteur $\mathbf{0} = (0, \dots, 0)$, dit **neutre** est tel que $\forall \mathbf{x} \in \mathbb{R}^n : \mathbf{x} + \mathbf{0} = \mathbf{x}$
 - ▶ $\forall \mathbf{x} \in \mathbb{R}^n$, l'élément $-\mathbf{x} = (-x_1, -x_2, \dots, -x_n)$ est appelé **opposé** de \mathbf{x} et est tel que $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$

Espaces vectoriels réels III

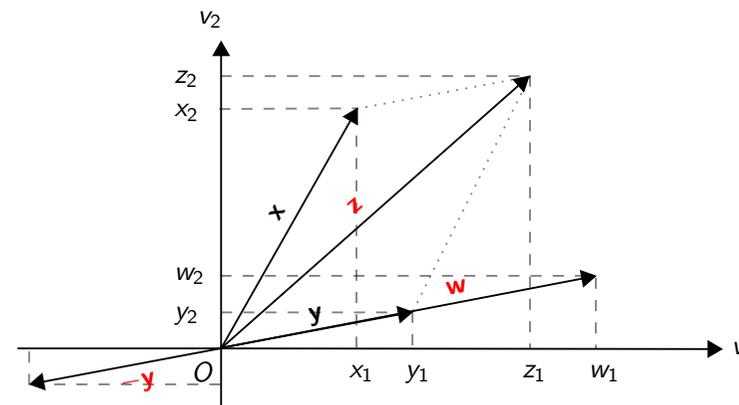
- La **multiplication** d'un vecteur $\mathbf{x} \in \mathbb{R}^n$ par un réel $\lambda \in \mathbb{R}$ est un vecteur \mathbf{y} de \mathbb{R}^n défini par :

$$\mathbf{y} = \lambda \mathbf{x} = \lambda \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \vdots \\ \lambda x_n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

- Propriétés de la loi externe de multiplication d'un vecteurs de \mathbb{R}^n par un réel :
 - $\forall \lambda, \mu \in \mathbb{R}; \forall \mathbf{x} \in \mathbb{R}^n : \lambda(\mu \mathbf{x}) = (\lambda \mu) \mathbf{x}$ (associativité)
 - $\forall \lambda, \mu \in \mathbb{R}; \forall \mathbf{x} \in \mathbb{R}^n : (\lambda + \mu) \mathbf{x} = \lambda \mathbf{x} + \mu \mathbf{x}$ (distributivité)
 - $\forall \lambda \in \mathbb{R}; \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n : \lambda(\mathbf{x} + \mathbf{y}) = \lambda \mathbf{x} + \lambda \mathbf{y}$ (distributivité)
 - $\forall \mathbf{x} \in \mathbb{R}^n : 1 \mathbf{x} = \mathbf{x}$ (1 étant l'élément neutre dans \mathbb{R})

Espaces vectoriels réels IV

- Exemple et représentation géométrique
 - $\mathbf{x} = (x_1, x_2)$ et $\mathbf{y} = (y_1, y_2)$ deux vecteurs de $\mathbb{E} = \mathbb{R}^2$ d'origine $O = (0, 0)$.
 - $\mathbf{z} = \mathbf{x} + \mathbf{y}$, $\mathbf{w} = 2\mathbf{y}$ sont deux autres vecteurs de \mathbb{R}^2 et $\mathbf{y} + (-\mathbf{y}) = \mathbf{0}$.

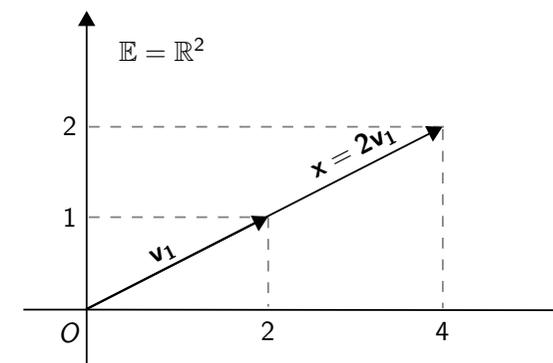


Espaces vectoriels réels V

- L'ensemble des vecteurs \mathbb{R}^n , muni des lois interne (addition entre vecteurs) et externe (multiplication d'un vecteur par un réel) présentées précédemment forme un **espace vectoriel** (par définition)
- Un sous-ensemble $\mathbb{F} \subset \mathbb{R}^n$ est un **sous-espace vectoriel (sev)** de \mathbb{R}^n ssi $\mathbb{F} \neq \emptyset$ et $\forall \mathbf{x}, \mathbf{y} \in \mathbb{F}; \forall \lambda, \mu \in \mathbb{R} \Rightarrow \lambda \mathbf{x} + \mu \mathbf{y} \in \mathbb{F}$
- Exemples fondamentaux de sous-espaces vectoriels
 - Droite vectorielle : Soit $\mathbf{v}_1 \in \mathbb{E}$, $\mathbf{v}_1 \neq \mathbf{0}$ alors : $\mathbb{F} = \{\mathbf{x} \in \mathbb{E} | \exists \lambda \in \mathbb{K} : \mathbf{x} = \lambda \mathbf{v}_1\}$ est un sev de \mathbb{E} dit **droite vectorielle engendrée par \mathbf{v}_1**
 - Plan vectoriel : Soit $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{E}$, alors : $\mathbb{F} = \{\mathbf{x} \in \mathbb{E} | \exists \lambda_1, \lambda_2 \in \mathbb{K} : \mathbf{x} = \lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2\}$ est un sev de \mathbb{E} dit **plan vectoriel engendré par \mathbf{v}_1 et \mathbf{v}_2**
 - Sous-espace engendré** : De manière générale, soient $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p \in \mathbb{E}$ alors $\mathbb{F} = \{\mathbf{x} \in \mathbb{E} | \exists \lambda_1, \dots, \lambda_p \in \mathbb{K} : \mathbf{x} = \lambda_1 \mathbf{v}_1 + \dots + \lambda_p \mathbf{v}_p\}$ est un sev de \mathbb{E} dit **sous-espace engendré par $\mathbf{v}_1, \dots, \mathbf{v}_p$** et est noté $\text{Vec}\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$

Espaces vectoriels réels VI

- Exemple et représentation géométrique
 - $\mathbb{E} = \mathbb{R}^2$ d'origine $O = (0, 0)$.
 - $\mathbf{v}_1 = (2, 1)$ et $\mathbb{F} = \{\mathbf{x} \in \mathbb{E} | \exists \lambda \in \mathbb{K} : \mathbf{x} = \lambda \mathbf{v}_1\}$.



Espaces vectoriels réels VII

- Soit $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ une famille de vecteurs de \mathbb{E} . On dit qu'elle est **libre** si : $\lambda_1 \mathbf{v}_1 + \dots + \lambda_p \mathbf{v}_p = \mathbf{0} \Rightarrow \lambda_1 = 0, \dots, \lambda_p = 0$.
On dit aussi que les vecteurs $\mathbf{v}_1, \dots, \mathbf{v}_p$ sont **linéairement indépendants**.
- Exemple : dans \mathbb{R}^3 , les vecteurs $\mathbf{v}_1 = (1, 2, 1)$, $\mathbf{v}_2 = (-1, 3, 1)$ et $\mathbf{v}_3 = (-1, 13, 5)$ ne sont pas linéairement indépendants (ils sont **liés**) car :

$$2 \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} + 3 \begin{pmatrix} -1 \\ 3 \\ 1 \end{pmatrix} - \begin{pmatrix} -1 \\ 13 \\ 5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

- Un vecteur \mathbf{x} est une **combinaison linéaire** des vecteurs $\mathbf{v}_1, \dots, \mathbf{v}_p$ s'il existe des réels $\lambda_1, \dots, \lambda_p$ tels que : $\mathbf{x} = \lambda_1 \mathbf{v}_1 + \dots + \lambda_p \mathbf{v}_p$

Espaces vectoriels réels VIII

- Soit $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ un ensemble de p vecteurs de \mathbb{E} . L'ensemble des vecteurs s'écrivant comme une combinaison linéaire de ces p vecteurs est appelé **sev engendré** par $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ et est dénoté :
 $Vec\{\mathbf{v}_1, \dots, \mathbf{v}_p\} = \{\sum_{j=1}^p \lambda_j \mathbf{v}_j : \lambda_1, \dots, \lambda_p \in \mathbb{R}\}$
- Etant donné un sev $\mathbb{F} \subset \mathbb{R}^n$, tout ensemble de $p < n$ vecteurs $\mathbf{v}_1, \dots, \mathbf{v}_p$ linéairement indépendants et tel que $\mathbb{F} = Vec\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ forme une **base** de \mathbb{F} . La **dimension** du sous-espace \mathbb{F} notée $dim(\mathbb{F})$ est p et toute base de \mathbb{F} possède ce même nombre de vecteurs.
- Si $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ est une base de \mathbb{F} alors tout $\mathbf{x} \in \mathbb{F}$ peut-être représentée de manière unique de la façon suivante :
 $\mathbf{x} = \lambda_1 \mathbf{v}_1 + \dots + \lambda_p \mathbf{v}_p$

Algèbre matricielle I

- On définit une **matrice** \mathbf{A} de taille $n \times p$ élément de $\mathcal{M}_{n,p}(\mathbb{R})$ (si $n = p$ on notera $\mathcal{M}_n(\mathbb{R})$) de la manière suivante :

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{np} \end{pmatrix}$$

Comme on travaille dans $\mathbb{R}^{n \times p}$ on a $\forall i, \forall j : a_{ij} \in \mathbb{R}$ et on écrira également $\mathcal{M}_{n,p}$ au lieu de $\mathcal{M}_{n,p}(\mathbb{R})$

- Nous noterons par \mathbf{a}_j la j ème colonne de la matrice \mathbf{A} :

$$\mathbf{a}_j = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{nj} \end{pmatrix}$$

On utilisera également la notation $\mathbf{A} = (\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_p)$

Algèbre matricielle II

- Exemple :

$$\mathbf{A} = \begin{pmatrix} 3 & -1 & 1 \\ 0 & 2 & 0 \\ 1 & -1 & 3 \end{pmatrix}$$

On a donc :

$$\mathbf{A} = (\mathbf{a}_1 \ \mathbf{a}_2 \ \mathbf{a}_3) \text{ avec } \mathbf{a}_1 = \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}; \mathbf{a}_2 = \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}; \mathbf{a}_3 = \begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix}$$

- Soit \mathbf{A} une matrice ($n \times p$). La matrice de taille ($p \times n$) dont les lignes sont les colonnes de \mathbf{A} est la **transposée** de \mathbf{A} et est notée \mathbf{A}^t

Algèbre matricielle III

- Exemple :

$$\mathbf{A}^t = \begin{pmatrix} 3 & 0 & 1 \\ -1 & 2 & -1 \\ 1 & 0 & 3 \end{pmatrix}$$

- Si $\mathbf{A} = \mathbf{A}^t$ alors \mathbf{A} est carrée et symétrique
- Le nombre maximum de vecteurs colonnes de \mathbf{A} linéairement indépendants est appelé le **rang** de la matrice et est noté $rg(\mathbf{A})$
- Dans l'exemple précédent $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ sont tous les 3 linéairement indépendants donc $rg(\mathbf{A}) = 3$

Algèbre matricielle V

- Calcul du déterminant de \mathbf{A} :

$$\begin{aligned} \begin{vmatrix} 3 & -1 & 1 \\ 0 & 2 & 0 \\ 1 & -1 & 3 \end{vmatrix} &= 3 \begin{vmatrix} 2 & 0 \\ -1 & 3 \end{vmatrix} - 0 \begin{vmatrix} -1 & 1 \\ -1 & 3 \end{vmatrix} + 1 \begin{vmatrix} -1 & 1 \\ 2 & 0 \end{vmatrix} \\ &= 3((2 \times 3) - (-1 \times 0)) - 0((-1 \times 3) - (-1 \times 1)) \\ &\quad + 1((-1 \times 0) - (2 \times 1)) \\ &= 3 \times 6 - 0 + 1 \times (-2) \\ &= 16 \end{aligned}$$

- Pour toute matrice carrée $\mathbf{A} \in \mathcal{M}_n(\mathbb{K})$, on a : $det(\mathbf{A}^t) = det(\mathbf{A})$
- Un déterminant est nul ssi l'une des colonnes (resp. lignes) est combinaison linéaire des autres colonnes (resp. lignes)
- Exemple : $|\mathbf{B}| = 0$

Algèbre matricielle IV

- Autre exemple :

$$\mathbf{B} = \begin{pmatrix} 3 & -1 & 1 \\ 0 & 2 & 4 \\ 1 & -1 & -1 \end{pmatrix}$$

$rg(\mathbf{B}) \neq 3$ car $\mathbf{b}_3 = \mathbf{b}_1 + 2\mathbf{b}_2$. On a $rg(\mathbf{B}) = 2$.

- Soit \mathbf{A} une matrice carrée d'ordre n . On associe à \mathbf{A} un réel appelé **déterminant** de \mathbf{A} et noté $|\mathbf{A}|$ ou $det(\mathbf{A})$. Il est défini par récurrence de la manière suivante :
 - Si $n = 1$, ie $\mathbf{A} = (a)$ alors $det(\mathbf{A}) = a$
 - Si $n > 1$, notons $\mathbf{A}_{-i-j} \in \mathcal{M}_{n-1}$ la matrice obtenue de \mathbf{A} en supprimant la i ème ligne et la j ème colonne (ie la ligne et la colonne passant par l'élément a_{ij}), on pose alors :

$$\begin{aligned} det(\mathbf{A}) &= (-1)^{i+1} a_{i1} det(\mathbf{A}_{-i-1}) + \dots + (-1)^{i+k} a_{ik} det(\mathbf{A}_{-i-k}) + \dots \\ &\quad \dots + (-1)^{i+n} a_{in} det(\mathbf{A}_{-i-n}) \end{aligned}$$

Algèbre matricielle VI

- Soit \mathbf{A} une matrice $(n \times p)$. On appelle un **mineur** d'ordre $m \leq \min(n, p)$ de \mathbf{A} , le déterminant d'une sous-matrice carrée d'ordre m obtenue à partir de \mathbf{A} en supprimant $n - m$ lignes et $p - m$ colonnes
- Si \mathbf{A} de taille $(n \times p)$ où $n \geq p$, possède un mineur d'ordre p qui est non nul alors les colonnes de \mathbf{A} sont linéairement indépendants et $rg(\mathbf{A}) = p$
- Si \mathbf{A} de taille $(n \times p)$ possède un mineur d'ordre $r \leq p$, $det(\hat{\mathbf{A}})$, tel que (i) $det(\hat{\mathbf{A}}) \neq 0$ (ii) tout mineur de \mathbf{A} obtenu en ajoutant une ligne et une colonne de \mathbf{A} à $\hat{\mathbf{A}}$ devient nul, alors $rg(\mathbf{A}) = r$

Algèbre matricielle VII

- Exemple :

$$\hat{\mathbf{B}} = \begin{pmatrix} 0 & 2 \\ 1 & -1 \end{pmatrix} \text{ est une matrice extraite de } \mathbf{B} = \begin{pmatrix} 3 & -1 & 1 \\ 0 & 2 & 4 \\ 1 & -1 & -1 \end{pmatrix}$$

$\det(\hat{\mathbf{B}})$ est un mineur d'ordre 2 de \mathbf{B} et $\det(\hat{\mathbf{B}}) \neq 0$. Si on ajoute une ligne et une colonne à $\hat{\mathbf{B}}$ on obtient à nouveau \mathbf{B} et on sait que $\det(\mathbf{B}) = 0$. On en conclut que $rg(\mathbf{B}) = 2$.

- Une matrice carrée \mathbf{A} est **non singulière** ou **inversible** si $\det(\mathbf{A}) \neq 0$
- Une matrice carrée \mathbf{A} d'ordre n est **inversible** ssi il existe une matrice carrée \mathbf{B} d'ordre n tel que : $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$ (\mathbf{I} étant la matrice unité)
 \mathbf{B} est appelé inverse de \mathbf{A} et est noté \mathbf{A}^{-1}

Système d'équations linéaires I

- Un **système d'équations linéaires** est un système du type :

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2 \\ \vdots \\ a_{p1}x_1 + a_{p2}x_2 + \dots + a_{pn}x_n = b_p \end{cases}$$

Les a_{ij} et les b_i sont des réels qui sont donnés. Les x_i sont des inconnues de \mathbb{R} et résoudre le système signifie déterminer celles, s'il y en a, qui vérifient toutes les équations.

Soit \mathbf{A} élément de $\mathcal{M}_{p,n}$, de terme général a_{ij} , \mathbf{b} le vecteur $p \times 1$ de terme général b_i et \mathbf{x} le vecteur inconnu de taille $n \times 1$ de terme général x_i . Le système s'écrit alors :

$$\mathbf{Ax} = \mathbf{b}$$

Algèbre matricielle VIII

- Exemple :

$$\mathbf{A}^{-1} = \frac{1}{16} \begin{pmatrix} 6 & 2 & -2 \\ 0 & 8 & 0 \\ -2 & 2 & 6 \end{pmatrix}$$

On peut vérifier que $\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$

Système d'équations linéaires II

- Le système $\mathbf{Ax} = \mathbf{b}$ a une solution ssi $rg(\mathbf{A}) = rg((\mathbf{A} \ \mathbf{b}))$
- Résoudre un système d'équations linéaires c'est étudier l'existence d'un vecteur \mathbf{x} dont les valeurs permettent de vérifier toutes les équations linéaires
- Une méthode simple pour la résolution d'un système d'équations linéaires est la **méthode du pivot (élimination de Gauss)**
- Exemple :

$$\begin{cases} l_1 & 2x + 2y - 2z = 2 \\ l_2 & 3x + 4y - 4z = 5 \\ l_3 & 5x + 10y - 8z = 10 \end{cases}$$

Système d'équations linéaires III

- ▶ 1ère itération :

$$\begin{cases} l'_1 = l_1/2 \\ l'_2 = l_2 - 3l'_1 \\ l'_3 = l_3 - 5l'_1 \end{cases} \quad \begin{cases} x + y - z = 1 \\ y - z = 2 \\ 5y - 3z = 5 \end{cases}$$

- ▶ 2ème itération :

$$\begin{cases} l''_2 = l'_2 \\ l''_3 = l'_3 - 5l''_2 \end{cases} \quad \begin{cases} x + y - z = 1 \\ y - z = 2 \\ 2z = -5 \end{cases}$$

- ▶ 3ème itération : le système est **échelonné** on peut alors résoudre en partant de la dernière équation jusqu'à la première

$$\begin{cases} z = -5/2 \\ y = z + 2 = -1/2 \\ x = 1 - y + z = -1 \end{cases}$$

Produit scalaire euclidien et norme euclidienne I

- Le produit scalaire euclidien entre deux vecteurs $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ est défini de la manière suivante :

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x_i y_i = \mathbf{x}^t \mathbf{y}$$

- Propriétés du produit scalaire :

- ▶ Il est **bilinéaire**, (ie linéaire en chacun des deux arguments), $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n; \forall \lambda \in \mathbb{R}$:

- ★ $\langle \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle$
- ★ $\langle \lambda \mathbf{x}, \mathbf{z} \rangle = \lambda \langle \mathbf{x}, \mathbf{z} \rangle$
- ★ $\langle \mathbf{x}, \mathbf{y} + \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{x}, \mathbf{z} \rangle$
- ★ $\langle \mathbf{x}, \lambda \mathbf{z} \rangle = \lambda \langle \mathbf{x}, \mathbf{z} \rangle$

- ▶ Il est **symétrique**, $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n : \langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$

- ▶ Il est **défini positif**, $\forall \mathbf{x} \in \mathbb{R}^n$:

- ★ $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$
- ★ $\langle \mathbf{x}, \mathbf{x} \rangle = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$

Produit scalaire euclidien et norme euclidienne II

- On appelle **longueur** ou **norme** du vecteur \mathbf{x} , le scalaire :

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \sqrt{\mathbf{x}^t \mathbf{x}} = \sqrt{\sum_{i=1}^n x_i^2}$$

- Propriétés d'une norme $n : \mathbb{R}^n \rightarrow \mathbb{R}^+$ dans un espace vectoriel, $\forall \mathbf{x} \in \mathbb{R}^n; \forall \lambda \in \mathbb{R}$:

- ▶ $n(\lambda \mathbf{x}) = |\lambda| n(\mathbf{x})$
- ▶ $n(\mathbf{x}) = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$
- ▶ Inégalité triangulaire : $n(\mathbf{x} + \mathbf{y}) \leq n(\mathbf{x}) + n(\mathbf{y})$
L'égalité a lieu ssi il existe $\lambda \geq 0$ tel que $\mathbf{y} = \lambda \mathbf{x}$.

- Inégalité de Cauchy-Schwartz :

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \|\mathbf{y}\|$$

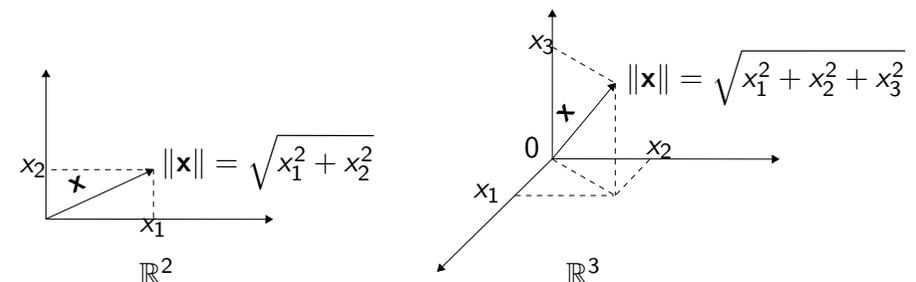
- \mathbf{x} et \mathbf{y} sont orthogonaux ssi $\langle \mathbf{x}, \mathbf{y} \rangle = 0$

- Théorème de Pythagore : si \mathbf{x} et \mathbf{y} sont orthogonaux alors on a :

$$\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$$

Produit scalaire euclidien et norme euclidienne III

- Exemple :

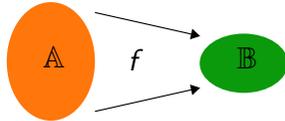


Applications linéaires I

- Soient \mathbb{A} et \mathbb{B} deux ensembles. On appelle f une **application** de \mathbb{A} dans \mathbb{B} , un sous-ensemble du produit cartésien $\mathbb{A} \times \mathbb{B}$ tel que pour tout élément \mathbf{a} de \mathbb{A} correspond un élément \mathbf{b} de \mathbb{B} . On écrira alors :

$$\begin{cases} f : \mathbb{A} \rightarrow \mathbb{B} \\ \mathbf{a} \rightarrow \mathbf{b} \end{cases}$$

- Illustration :



- $f : \mathbb{R}^n \rightarrow \mathbb{R}^p$ est une **application linéaire** de \mathbb{R}^n à \mathbb{R}^p (ce qu'on notera également par $f \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$ si elle vérifie les propriétés suivantes :

- ▶ $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n : f(\mathbf{x} + \mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y})$
- ▶ $\forall \mathbf{x} \in \mathbb{R}^n, \forall \lambda \in \mathbb{R} : f(\lambda \mathbf{x}) = \lambda f(\mathbf{x})$

Applications linéaires II

- Exemple : $\begin{cases} f : \mathbb{R}^3 \rightarrow \mathbb{R}^2 \\ (x_1, x_2, x_3) \rightarrow (2x_1 + x_2, x_2 - x_3) \end{cases}$
- Soit $f \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$ et soient $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ et $\{\epsilon_1, \dots, \epsilon_p\}$ deux bases de \mathbb{R}^n et de \mathbb{R}^p respectivement. Supposons que les images par f des vecteurs $\mathbf{e}_1, \dots, \mathbf{e}_n$ se décomposent sur la base $\{\epsilon_1, \dots, \epsilon_p\}$ de la façon suivante :

$$\begin{array}{rcccccccc} f(\mathbf{e}_1) & = & a_{11}\epsilon_1 & + & a_{21}\epsilon_2 & + & \dots & + & a_{p1}\epsilon_p \\ f(\mathbf{e}_2) & = & a_{12}\epsilon_1 & + & a_{22}\epsilon_2 & + & \dots & + & a_{p2}\epsilon_p \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ f(\mathbf{e}_n) & = & a_{1n}\epsilon_1 & + & a_{2n}\epsilon_2 & + & \dots & + & a_{pn}\epsilon_p \end{array}$$

Applications linéaires III

- On appelle alors **matrice de f** dans les bases $\{\mathbf{e}_i\}, \{\epsilon_j\}$, la matrice notée \mathbf{A} appartenant à $\mathcal{M}_{p,n}$ dont les colonnes sont les composantes des vecteurs $f(\mathbf{e}_1), \dots, f(\mathbf{e}_n)$ dans la base $\{\epsilon_1, \dots, \epsilon_p\}$:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pn} \end{pmatrix}$$

- Pour l'exemple précédent, si on suppose les bases canoniques, on obtient :

$$\mathbf{A} = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & -1 \end{pmatrix}$$

- L'image d'un vecteur $\mathbf{x} \in \mathbb{R}^n$ par $f \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$ est le vecteur $f(\mathbf{x}) \in \mathbb{R}^p$ donné par :

$$f(\mathbf{x}) = \mathbf{Ax}$$

Applications linéaires IV

- Exemple : image par f de $\mathbf{x} = (1, 1, 2)$
 - ▶ Ecriture algébrique : $f(\mathbf{x}) = (2x_1 + x_2, x_2 - x_3) = (3, -1)$
 - ▶ Ecriture matricielle : $f(\mathbf{x}) = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$
- Soit $f \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$ et \mathbf{A} sa représentation matricielle. $f(\mathbb{R}^n)$ est un sous-espace de \mathbb{R}^p appelé **image** de f et sera noté $Im(f)$ ou $Im(\mathbf{A})$. On a : $Im(\mathbf{A}) = \{\mathbf{Ax} : \mathbf{x} \in \mathbb{R}^n\}$
- Soit $f \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$ et \mathbf{A} sa représentation matricielle. Le **noyau** de f est un sous-espace de \mathbb{R}^n noté $Ker(f)$ ou $Ker(\mathbf{A})$. Il est défini par : $Ker(\mathbf{A}) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{Ax} = \mathbf{0}\}$

Valeurs propres et vecteurs propres de matrices carrées I

- Soit \mathbf{A} une matrice carrée d'ordre n . Un réel λ et un vecteur $\mathbf{v} \in \mathbb{R}^n$ sont dits respectivement **valeur propre** et **vecteur propre** de \mathbf{A} s'ils satisfont la relation :

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$$

- Une condition nécessaire et suffisante pour que λ soit valeur propre est $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$ (la matrice $\mathbf{A} - \lambda\mathbf{I}$ est singulière)
- Le développement de $\det(\mathbf{A} - \lambda\mathbf{I})$ aboutit à un polynôme de degré n appelé **polynôme caractéristique** de \mathbf{A} . Les valeurs propres sont donc les racines du polynôme caractéristique
- Si $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$ possède n racines distinctes $\lambda_1, \dots, \lambda_n$ alors il existe n vecteurs $\mathbf{v}_1, \dots, \mathbf{v}_n$ linéairement indépendants tels que $\forall i = 1, \dots, n : \mathbf{A}\mathbf{v}_i = \lambda_i\mathbf{v}_i$

Valeurs propres et vecteurs propres de matrices carrées III

Les valeurs propres sont donc $\lambda_1 = 2$, $\lambda_2 = 2$ (racine d'ordre 2) et $\lambda_3 = 4$. Les vecteurs propres sont par ailleurs :

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} ; \quad \mathbf{v}_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} ; \quad \mathbf{v}_3 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

Ils sont linéairement indépendants et forment donc une base. \mathbf{A} est donc diagonalisable.

- Si \mathbf{A} est une matrice carrée d'ordre n qui est symétrique alors :
 - ▶ toutes les valeurs propres de \mathbf{A} sont réelles
 - ▶ \mathbf{A} possède n vecteurs propres qui sont mutuellement orthogonaux (donc \mathbf{A} est diagonalisable)

Valeurs propres et vecteurs propres de matrices carrées II

- On dit que \mathbf{A} est diagonalisable s'il existe une base de vecteurs propres $\mathbf{P} = (\mathbf{v}_1 \ \dots \ \mathbf{v}_n)$ telle que : $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{A}'$ soit une matrice diagonale
- Exemple :

$$\mathbf{A} = \begin{pmatrix} 3 & -1 & 1 \\ 0 & 2 & 0 \\ 1 & -1 & 3 \end{pmatrix}$$

Son polynôme caractéristique est :

$$\det(\mathbf{A} - \lambda\mathbf{I}) = \begin{vmatrix} 3 - \lambda & -1 & 1 \\ 0 & 2 - \lambda & 0 \\ 1 & -1 & 3 - \lambda \end{vmatrix}$$

Après développement on trouve :

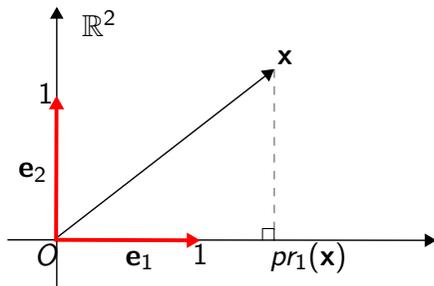
$$\det(\mathbf{A} - \lambda\mathbf{I}) = (2 - \lambda)^2(4 - \lambda)$$

Projections orthogonales I

- Si \mathbb{F} est un sev de \mathbb{R}^n alors on peut lui associer un **sev orthogonal** noté \mathbb{F}^\perp défini par : $\mathbb{F}^\perp = \{\mathbf{x} \in \mathbb{R}^n : \langle \mathbf{x}, \mathbf{y} \rangle = 0, \forall \mathbf{y} \in \mathbb{F}\}$
- Si \mathbb{F} est un sev de \mathbb{R}^n alors $\mathbb{R}^n = \mathbb{F} \oplus \mathbb{F}^\perp$ ie $\forall \mathbf{x} \in \mathbb{R}^n, \exists! \mathbf{x}_1 \in \mathbb{F}, \exists! \mathbf{x}_2 \in \mathbb{F}^\perp : \mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$
- \mathbf{P} est une application linéaire dite **projection orthogonale** sur \mathbb{F} si $\forall \mathbf{x} \in \mathbb{R}^n : (\mathbf{P}\mathbf{x} \in \mathbb{F}) \wedge (\mathbf{x} - \mathbf{P}\mathbf{x} \in \mathbb{F}^\perp)$
- Une matrice \mathbf{P} est un projecteur orthogonal (sur $Im(\mathbf{P})$) ssi $\mathbf{P}^2 = \mathbf{P} = \mathbf{P}^t$

Projections orthogonales II

- Exemple :



- Soit \mathbb{R}^2 muni de sa base canonique et soit :
 $pr_1 : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ avec $pr_1(x_1, x_2) = (x_1, 0)$. pr_1 est une projection orthogonale sur $pr_1(\mathbb{R}^2)$. On a $pr_1(e_1) = e_1$ et $pr_1(e_2) = (0, 0)$ et donc la matrice \mathbf{P} représentant pr_1 est définie par $\mathbf{P} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$

On vérifie bien $\mathbf{P}^2 = \mathbf{P} = \mathbf{P}^t$

Formes quadratiques II

- Les **mineurs principaux dominants** d'une matrice carrée \mathbf{Q} sont les n déterminants des sous-matrices de \mathbf{Q} dénotés $\Delta_1, \dots, \Delta_n$ où Δ_i est le déterminant de la sous-matrice de \mathbf{Q} composée des i premières lignes et colonnes de cette-dernière.
- Une forme quadratique $f(\mathbf{x}) = \mathbf{x}^t \mathbf{Q} \mathbf{x}$ avec $\mathbf{Q} = \mathbf{Q}^t$ est définie positive ssi les mineurs principaux dominants de \mathbf{Q} sont strictement positifs (critère de Sylvester).
- Exemple :

$$\mathbf{Q} = \begin{pmatrix} 3 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 3 \end{pmatrix}$$

Formes quadratiques I

- Une **forme quadratique** $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est une fonction :
 $f(\mathbf{x}) = \mathbf{x}^t \mathbf{Q} \mathbf{x}$ où $\mathbf{Q} \in \mathcal{M}_n$
- On suppose dans la suite que \mathbf{Q} est **symétrique** ($\mathbf{Q} = \mathbf{Q}^t$) sans perte de généralité : si \mathbf{Q} n'était pas symétrique on pourrait toujours la remplacer par une matrice symétrique $\mathbf{Q}_0 = \frac{1}{2}(\mathbf{Q} + \mathbf{Q}^t)$ sans que cela change la forme quadratique.
- Une forme quadratique $f(\mathbf{x}) = \mathbf{x}^t \mathbf{Q} \mathbf{x}$ est :
 - définie positive (**dp**) si : $\forall \mathbf{x} \neq \mathbf{0} : \mathbf{x}^t \mathbf{Q} \mathbf{x} > 0$,
 - semi-définie positive (**sdp**) si : $\forall \mathbf{x} : \mathbf{x}^t \mathbf{Q} \mathbf{x} \geq 0$,
 - définie négative (**dn**) si : $\forall \mathbf{x} \neq \mathbf{0} : \mathbf{x}^t \mathbf{Q} \mathbf{x} < 0$,
 - semi-définie négative (**sdn**) si : $\forall \mathbf{x} : \mathbf{x}^t \mathbf{Q} \mathbf{x} \leq 0$.

Formes quadratiques III

Les mineurs principaux dominants de \mathbf{Q} sont :

$$\Delta_1 = |3| \quad ; \quad \Delta_2 = \begin{vmatrix} 3 & -1 \\ -1 & 2 \end{vmatrix} \quad ; \quad \Delta_3 = \begin{vmatrix} 3 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 3 \end{vmatrix}$$

- Une matrice symétrique \mathbf{Q} est dite **définie positive** si la forme quadratique associée $\mathbf{x}^t \mathbf{Q} \mathbf{x}$ est définie positive. On utilisera alors la notation $\mathbf{Q} > 0$
- Pour l'exemple précédent, $\mathbf{Q} > 0$ car :

$$\Delta_1 = 3 \quad ; \quad \Delta_2 = 5 \quad , \quad \Delta_3 = 12$$

Formes quadratiques IV

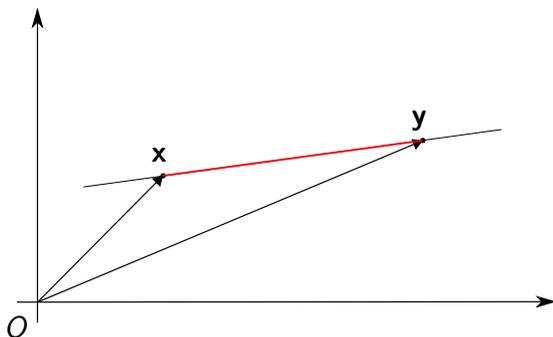
- Similairement on parlera de matrice symétrique \mathbf{Q} semi-définie positive ($\mathbf{Q} \geq 0$), définie négative ($\mathbf{Q} < 0$) et semi-définie négative ($\mathbf{Q} \leq 0$) en raisonnant sur la forme quadratique associée. De plus, \mathbf{Q} est dite **indéfinie** si elle est ni semi-définie positive, ni semi-définie négative
- Une matrice symétrique \mathbf{Q} est définie positive (resp. semi-définie positive) ssi ses valeurs propres sont strictement positives (resp. non négatives)
- Pour l'exemple précédent les valeurs propres de \mathbf{Q} sont $\lambda_1 = 1$, $\lambda_2 = 3$ et $\lambda_3 = 4$

Segment de droite

- Le segment entre deux points \mathbf{x} et \mathbf{y} de \mathbb{R}^n est l'ensemble des points appartenant à la portion de la droite entre \mathbf{x} et \mathbf{y} . Formellement, il est défini de la manière suivante;

$$\{\lambda \mathbf{x} + (1 - \lambda) \mathbf{y} : \lambda \in [0, 1]\}$$

- Illustration dans \mathbb{R}^2 :



Rappel du Sommaire

- 1 Rappels de concepts mathématiques
 - Quelques éléments d'algèbre linéaire
 - Quelques éléments de géométrie
 - Quelques éléments d'analyse
- 2 Bases théoriques de l'optimisation
- 3 Optimisation sans contrainte
- 4 Optimisation linéaire avec contraintes
- 5 Optimisation non linéaire avec contraintes
- 6 Optimisation convexe
- 7 Algorithmes pour problèmes contraints

Algorithmes pour problèmes contraints

Hyperplans et variétés linéaires I

- Soit $u_1, \dots, u_n, v \in \mathbb{R}$ où au moins un des u_i est non nul. L'ensemble des points $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ qui satisfait l'équation linéaire suivante :

$$u_1 x_1 + \dots + u_n x_n = v$$

est appelé un **hyperplan** de \mathbb{R}^n

- On décrit un hyperplan par l'ensemble suivant :

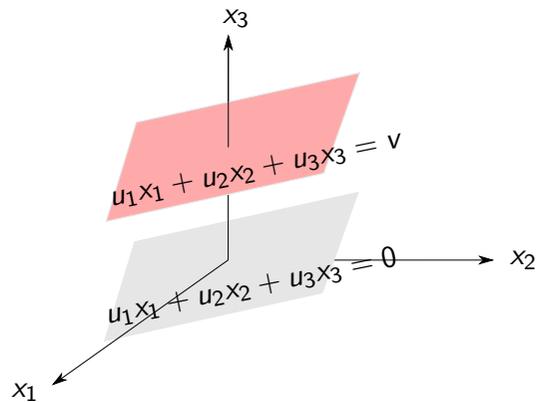
$$\mathbb{H} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{u}^t \mathbf{x} = v\}$$

où $\mathbf{u} = (u_1, \dots, u_n)$

- Les hyperplans étendent à \mathbb{R}^n , le concept de droite dans \mathbb{R}^2 ($ax + by = c$) et celui de plan dans \mathbb{R}^3 ($u_1 x_1 + u_2 x_2 + u_3 x_3 = v$)

Hyperplans et variétés linéaires II

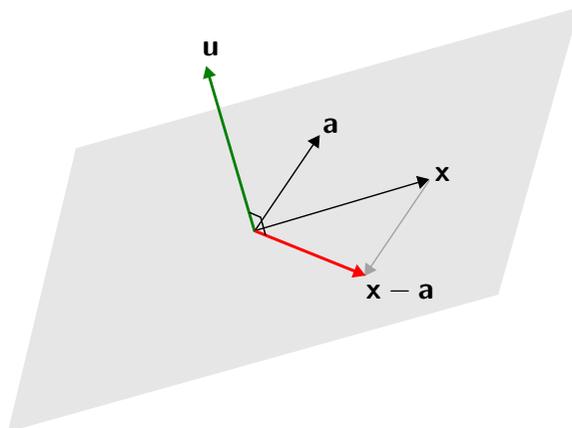
- Illustration dans \mathbb{R}^3 :



- Un hyperplan est un sous-ensemble qui n'est pas nécessairement un sev de \mathbb{R}^n (le vecteur nul n'appartient pas forcément à celui-ci)
- La dimension d'un hyperplan de \mathbb{R}^n est $n - 1$

Hyperplans et variétés linéaires IV

- Illustration dans \mathbb{R}^3 :



Hyperplans et variétés linéaires III

- Un hyperplan sépare \mathbb{R}^n en deux :
 - ▶ Demi-espace "positif" : $\mathbb{H}_+ = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{u}^t \mathbf{x} \geq v\}$
 - ▶ Demi-espace "négatif" : $\mathbb{H}_- = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{u}^t \mathbf{x} \leq v\}$
- Soit $\mathbf{a} \in \mathbb{H}$. Par définition $\mathbf{u}^t \mathbf{a} - v = 0$. On peut alors écrire, $\forall \mathbf{x} \in \mathbb{H}$:

$$\begin{aligned} \mathbf{u}^t \mathbf{x} - v &= \mathbf{u}^t \mathbf{x} - v - (\mathbf{u}^t \mathbf{a} - v) \\ &= \mathbf{u}^t (\mathbf{x} - \mathbf{a}) \\ &= u_1(x_1 - a_1) + \dots + u_n(x_n - a_n) \\ &= 0 \end{aligned}$$

Donc si $\mathbf{a} \in \mathbb{H}$ on peut définir l'hyperplan \mathbb{H} par $\{\mathbf{x} \in \mathbb{R}^n : \langle \mathbf{u}, \mathbf{x} - \mathbf{a} \rangle = 0\}$

- \mathbb{H} est l'ensemble des points \mathbf{x} pour lesquels les vecteurs \mathbf{u} et $\mathbf{x} - \mathbf{a}$ sont orthogonaux. \mathbf{u} est appelé le **vecteur normal** de l'hyperplan \mathbb{H}
- Soit $\mathbf{a} \in \mathbb{H}$ on a alors :
 - ▶ $\mathbb{H}_+ = \{\mathbf{x} \in \mathbb{R}^n : \langle \mathbf{u}, \mathbf{x} - \mathbf{a} \rangle \geq 0\}$
 - ▶ $\mathbb{H}_- = \{\mathbf{x} \in \mathbb{R}^n : \langle \mathbf{u}, \mathbf{x} - \mathbf{a} \rangle \leq 0\}$

Hyperplans et variétés linéaires V

- Une **variété linéaire** de \mathbb{R}^p est un sous-espace de la forme :

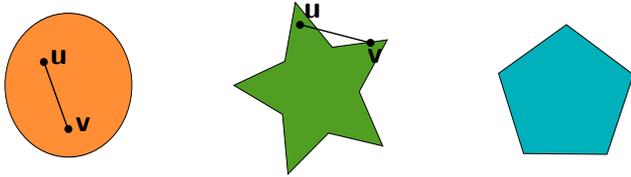
$$\{\mathbf{x} \in \mathbb{R}^p : \mathbf{A}\mathbf{x} = \mathbf{b}\}$$

avec $\mathbf{A} \in \mathcal{M}_{n,p}$ et $\mathbf{b} \in \mathbb{R}^n$

- Si $\dim(\text{Ker}(\mathbf{A})) = r$ on dit que la variété linéaire est de dimension r
- Une variété linéaire est un sev ssi $\mathbf{b} = \mathbf{0}$
- Si la dimension d'une variété linéaire de \mathbb{R}^p est plus petite que p alors celle-ci est l'intersection d'un nombre fini d'hyperplans

Ensembles convexes I

- Un sous-ensemble $\mathbb{O} \subset \mathbb{R}^n$ forme un **ensemble convexe** si pour deux vecteurs $\mathbf{u}, \mathbf{v} \in \mathbb{O}$ leur segment appartient également à \mathbb{O} .
- Exemple et contre-exemple :



- Autres exemples : \emptyset , $\{\mathbf{x}\}$, un segment entre deux points, un sev, un hyperplan, les demi-espaces engendré par un hyperplan, une variété linéaire,...
- Propriétés des ensembles convexes dans \mathbb{R}^n :
 - Si \mathbb{O} est convexe et $\lambda \in \mathbb{R}$ alors $\lambda\mathbb{O} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} = \lambda\mathbf{u}, \mathbf{u} \in \mathbb{O}\}$ est convexe

Ensembles convexes II

- Si \mathbb{O}_1 et \mathbb{O}_2 sont deux ensembles convexes alors
 - $\mathbb{O}_1 + \mathbb{O}_2 = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} = \mathbf{u} + \mathbf{v}, \mathbf{u} \in \mathbb{O}_1, \mathbf{v} \in \mathbb{O}_2\}$ est convexe
 - L'intersection d'une collection d'ensembles convexes est convexe
- \mathbf{x} est une **combinaison convexe** de \mathbf{u} et \mathbf{v} ssi $\mathbf{x} = \lambda\mathbf{u} + (1 - \lambda)\mathbf{v}$ avec $\lambda \in [0, 1]$ (\mathbf{x} appartient au segment entre \mathbf{u} et \mathbf{v})
- Un vecteur \mathbf{x} est un **point extrême** d'un ensemble convexe \mathbb{O} s'il ne peut être exprimé comme une combinaison convexe de deux autres vecteurs de \mathbb{O} . En d'autres termes, un **point extrême** ne peut se trouver sur le segment formé par deux autres vecteurs de l'ensemble.
- Illustration :



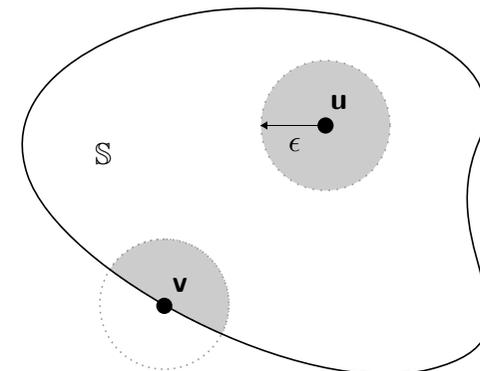
Voisinages I

- La **voisinage** d'un point $\mathbf{x} \in \mathbb{R}^n$ est l'ensemble suivant :

$$\{\mathbf{y} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{y}\| < \epsilon\}$$
 où ϵ est un réel positif. Le voisinage de \mathbf{x} est également appelé la **boule** de centre \mathbf{x} et de rayon ϵ
- Exemples :
 - Dans \mathbb{R}^2 il s'agit des points à l'intérieur du disque centré en \mathbf{x}
 - Dans \mathbb{R}^3 il s'agit des points à l'intérieur de la sphère centrée en \mathbf{x}
- Un point $\mathbf{x} \in \mathbb{S}$ est un **point intérieur** de l'ensemble \mathbb{S} si ce dernier contient le voisinage de \mathbf{x} (ie si tous les points du voisinage de \mathbf{x} sont contenus dans \mathbb{S})
- L'ensemble des points intérieurs de \mathbb{S} est appelé l'**intérieur** de \mathbb{S}
- Un point $\mathbf{x} \in \mathbb{S}$ est un **point frontière** de \mathbb{S} si le voisinage de \mathbf{x} contient au moins un point dans \mathbb{S} et au moins un point hors de \mathbb{S}

Voisinages II

- L'ensemble des points frontières de \mathbb{S} est appelé la **frontière** de \mathbb{S}
- Illustration :



Voisinages III

- Un ensemble \mathbb{S} est un **ouvert** s'il contient le voisinage de chacun de ses points (ie \mathbb{S} n'a pas de points frontières mais que des points intérieurs)
- Un ensemble \mathbb{S} est un **fermé** s'il contient sa frontière (comme dans l'illustration précédente et comme les demi-espaces engendrés par un hyperplan)
- Un ensemble \mathbb{S} qui est contenu dans une boule de rayon fini est dit **borné**
- Un ensemble \mathbb{S} qui est à la fois fermé et borné est dit **compact**
- Les ensembles compacts sont importants en optimisation :
Soit $f : \mathbb{S} \rightarrow \mathbb{R}$ une fonction continue. Si \mathbb{S} est un ensemble compact alors il existe $\mathbf{x}^* \in \mathbb{S}$ tel que $f(\mathbf{x}^*) \leq f(\mathbf{x})$ pour tout \mathbf{x} dans \mathbb{S} . En d'autres termes, f atteint son minimum sur \mathbb{S} (**théorème de Weierstrass**)

Polytopes et polyèdres II

- Pour tout polyèdre convexe $\mathbb{O} \in \mathbb{R}^n$ il existe un entier positif $k \leq n$ tel que \mathbb{O} soit contenu dans une variété linéaire de dimension k mais qui n'est pas entièrement contenue dans aucune variété linéaire de dimension $k - 1$
- De plus il n'existe qu'une seule variété linéaire de dimension k qui contienne \mathbb{O} . Cette variété linéaire est dite **support** du polyèdre et k est la dimension du polyèdre. Exemple : dans \mathbb{R}^2 , un segment est un polyèdre convexe et son support est la droite qui prolonge le segment
- La frontière d'un polyèdre de dimension $k > 0$ est constituée d'un nombre fini de polyèdres de dimension $k - 1$
- Ces polyèdres de dimension $k - 1$ formant la frontière de \mathbb{O} sont appelés les **faces** de \mathbb{O}
- Les faces de dimension $k - 1$, étant elles-mêmes des polyèdres, ont des faces de dimension $k - 2$ et ainsi de suite

Polytopes et polyèdres I

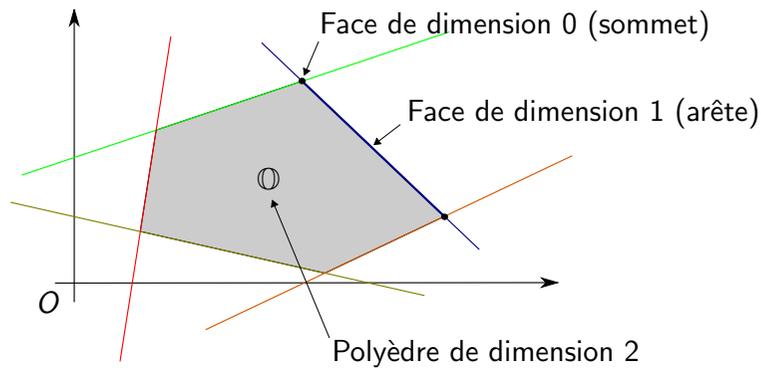
- Soit \mathbb{O} un ensemble compact et soit \mathbf{y} un point frontière de \mathbb{O} . Un hyperplan \mathbb{H} passant par \mathbf{y} est appelé un **hyperplan support** de l'ensemble \mathbb{O} si tout \mathbb{O} est contenu dans l'un des demi-espaces engendré par \mathbb{H}
- Rappelons qu'étant donné un hyperplan \mathbb{H} , les demi-espaces associés \mathbb{H}_+ et \mathbb{H}_- sont des ensembles convexes dans \mathbb{R}^n . Rappelons également que l'intersection d'un ensemble fini d'ensembles convexes est convexe
- Un ensemble qui peut-être exprimé comme étant l'intersection d'un nombre fini de demi-espaces est un **polytope (convexe)**
- Un polytope non vide qui est borné est appelé un **polyèdre** (Remarque : certains auteurs utilisent polytope et polyèdre de manière similaire)

Polytopes et polyèdres III

- Les faces de dimension 1 sont appelées **arêtes** et celles de dimension 0 sont appelées **sommets** du polyèdre

Polytopes et polyèdres IV

- Illustration dans \mathbb{R}^2 :



Suites et limites I

- Une **suite** de nombres réels est une fonction dont le domaine est \mathbb{N} , l'ensemble des entiers naturels $\{1, 2, \dots, k, \dots\}$, et dont l'image est contenu dans \mathbb{R} .
- Une suite peut donc être vue comme étant un ensemble de nombre $\{x_1, x_2, \dots, x_k, \dots\}$ également noté $\{x_k\}$ (ou encore $\{x_k\}_{k=1}^{\infty}$)
- Une suite est dite **strictement croissante** si $x_1 < x_2 < \dots < x_k < \dots$ ie $\forall k : x_k < x_{k+1}$
- Une suite est dite **croissante** si $\forall k : x_k \leq x_{k+1}$
- Exemple :

x_1	x_2	x_3	x_4	x_5	\dots
1	1	2	3	5	\dots
- Une suite est dite **strictement décroissante** si $\forall k : x_k > x_{k+1}$
- Une suite est dite **décroissante** si $\forall k : x_k \geq x_{k+1}$
- Une suite croissante ou décroissante est dite **monotone**

Rappel du Sommaire

- 1 Rappels de concepts mathématiques
 - Quelques éléments d'algèbre linéaire
 - Quelques éléments de géométrie
 - Quelques éléments d'analyse
- 2 Bases théoriques de l'optimisation
- 3 Optimisation sans contrainte
- 4 Optimisation linéaire avec contraintes
- 5 Optimisation non linéaire avec contraintes
- 6 Optimisation convexe
- 7 Algorithmes pour problèmes contraints

Suites et limites II

- Un nombre $x^* \in \mathbb{R}$ est appelé **limite** de la suite $\{x_k\}$ si pour tout $\epsilon > 0$ il existe un nombre K (qui peut dépendre de ϵ) tel que : $\forall k > K : |x_k - x^*| < \epsilon$ ie $\forall k > K : x^* - \epsilon < x_k < x^* + \epsilon$. Dans ce cas, on écrit :

$$x^* = \lim_{k \rightarrow \infty} x_k \text{ ou également } x_k \rightarrow x^*$$

- Une suite possédant une limite est dite **convergente**
- Exemple : $x_k = \frac{1}{k}$ a pour limite $x^* = 0$
- La notion de suite peut être étendue à des suites d'éléments de \mathbb{R}^n : une suite dans \mathbb{R}^n est une fonction dont le domaine est \mathbb{N} mais dont l'image est \mathbb{R}^n
- On utilisera la notation $\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(k)}, \dots\}$ ou $\{\mathbf{x}^{(k)}\}$ pour les suites dans \mathbb{R}^n

Suites et limites III

- Pour les notions de limites dans \mathbb{R}^n on remplace la valeur absolue par la norme euclidienne de vecteurs. Un vecteur $\mathbf{x}^* \in \mathbb{R}^n$ est appelé **limite** de la suite $\{\mathbf{x}^{(k)}\}$ si pour tout $\epsilon > 0$ il existe un nombre K (qui peut dépendre de ϵ) tel que : $\forall k > K : \|\mathbf{x}^{(k)} - \mathbf{x}^*\| < \epsilon$. Dans ce cas, on écrit :

$$\mathbf{x}^* = \lim_{k \rightarrow \infty} \mathbf{x}^{(k)} \text{ ou également } \mathbf{x}^{(k)} \rightarrow \mathbf{x}^*$$

- Toute suite convergente a une et une seule limite
- Exemple dans \mathbb{R}^2 : $\mathbf{x}^{(k)} = (\frac{1}{k}, \frac{k+1}{k})$ a pour limite $\mathbf{x}^* = (0, 1)$
- Une suite $\{\mathbf{x}^{(k)}\}$ de \mathbb{R}^n est **bornée** s'il existe un nombre $B \geq 0$ tel que $\forall k : \|\mathbf{x}^{(k)}\| \leq B$
- Pour l'exemple précédent $\|\mathbf{x}^{(k)}\|^2 = \langle \mathbf{x}^{(k)}, \mathbf{x}^{(k)} \rangle = \frac{1+(k+1)^2}{k^2}$. On voit qu'une borne supérieure pourrait être $B = 5$.
- Toute suite convergente est bornée

Suites et limites V

- Soit une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ et un point $\mathbf{x}_0 \in \mathbb{R}^n$. Supposons qu'il existe $\mathbf{f}^* \in \mathbb{R}^m$ tel que pour toute suite convergente $\{\mathbf{x}^{(k)}\}$ de limite \mathbf{x}_0 , nous avons :

$$\lim_{k \rightarrow \infty} f(\mathbf{x}^{(k)}) = \mathbf{f}^*$$

alors nous utiliserons la notation suivante :

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} f(\mathbf{x})$$

pour désigner la limite \mathbf{f}^*

- La fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ est **continue** en \mathbf{x}_0 ssi :

$$\lim_{k \rightarrow \infty} f(\mathbf{x}^{(k)}) = f(\lim_{k \rightarrow \infty} \mathbf{x}^{(k)}) = f(\mathbf{x}_0)$$

ou encore (en utilisant les notations précédentes) ssi :

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} f(\mathbf{x}) = f(\mathbf{x}_0)$$

Suites et limites IV

- Pour une suite $\{x_k\}$ de \mathbb{R} , un nombre B est appelé **majorant** si $\forall k : x_k \leq B$
- Pour une suite $\{x_k\}$ de \mathbb{R} , un nombre B est appelé **minorant** si $\forall k : x_k \geq B$
- Clairement une suite $\{x_k\}$ est **bornée** si elle possède à la fois un majorant et un minorant
- Le plus petit des majorants de $\{x_k\}$ est appelé **borne supérieure**
- Le plus grand des minorants de $\{x_k\}$ est appelé **borne inférieure**
- Toute suite de \mathbb{R} qui est monotone et bornée est convergente
- Soit $\{\mathbf{x}^{(k)}\}$ une suite de \mathbb{R}^n et soit $\{m_k\}$ une suite strictement croissante de \mathbb{N} . La suite $\{\mathbf{x}^{(m_1)}, \mathbf{x}^{(m_2)}, \dots, \mathbf{x}^{(m_k)}, \dots\}$ est alors appelé **sous-suite** (ou suite extraite) de $\{\mathbf{x}^{(k)}\}$ (cela revient à négliger certains éléments de la suite)
- Si $\{\mathbf{x}^{(k)}\}$ est convergente et a pour limite \mathbf{x}^* alors toute sous-suite de $\{\mathbf{x}^{(k)}\}$ est également convergente et a pour limite \mathbf{x}^*

Différentiabilité I

- Le calcul différentiel est basé sur l'idée que l'on peut approximer une quelconque fonction par une fonction affine
- Une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}^p$ est dite **affine** s'il existe une fonction linéaire $g \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$ et un vecteur $\mathbf{y} \in \mathbb{R}^p$ tel que :

$$f(\mathbf{x}) = g(\mathbf{x}) + \mathbf{y}$$

On notera $\mathcal{A}(\mathbb{R}^n, \mathbb{R}^p)$ l'ensemble des fonctions affines de \mathbb{R}^n dans \mathbb{R}^p

- Soit une fonction $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$ quelconque et un point $\mathbf{x}_0 \in \mathbb{R}^n$. On souhaite approximer h par une fonction affine $f \in \mathcal{A}(\mathbb{R}^n, \mathbb{R}^p)$ dans le voisinage de \mathbf{x}_0 . Il est alors naturel d'imposer (i) $f(\mathbf{x}_0) = h(\mathbf{x}_0)$
- On impose ensuite que (ii) $f(\mathbf{x})$ approche $h(\mathbf{x})$ plus rapidement que \mathbf{x} approche \mathbf{x}_0 ce qui s'exprime par :

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}_0, \mathbf{x} \in \mathbb{R}^n} \frac{\|h(\mathbf{x}) - f(\mathbf{x})\|}{\|\mathbf{x} - \mathbf{x}_0\|} = 0$$

Différentiabilité II

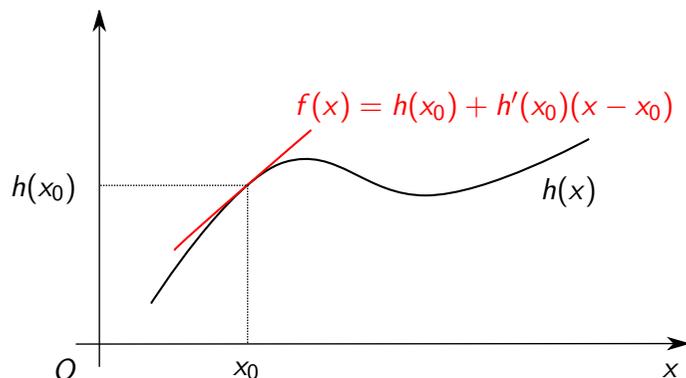
- La condition (ii) assure que f approxime h dans le voisinage de \mathbf{x}_0 dans le sens où l'erreur d'approximation en \mathbf{x} est "petit" en comparaison de la distance entre \mathbf{x} et \mathbf{x}_0
- Une fonction $h : \mathbb{S} \rightarrow \mathbb{R}^p$ est dite **différentiable** en $\mathbf{x}_0 \in \mathbb{S}$ s'il existe une fonction affine qui approxime h dans le voisinage de \mathbf{x}_0 , ie qu'il existe une fonction linéaire $g \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$ telle que :

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}_0, \mathbf{x} \in \mathbb{S}} \frac{\|h(\mathbf{x}) - (g(\mathbf{x} - \mathbf{x}_0) + h(\mathbf{x}_0))\|}{\|\mathbf{x} - \mathbf{x}_0\|} = 0$$

- Ci-dessus, la fonction linéaire g est déterminée de façon unique pour h et \mathbf{x}_0 . Elle est appelée **dérivée** de h en \mathbf{x}_0
- La fonction f est dite **différentiable** sur \mathbb{S} si elle est différentiable en tout point de \mathbb{S}

Différentiabilité IV

- Illustration



Différentiabilité III

- Cas particulier de $h : \mathbb{R} \rightarrow \mathbb{R}$ différentiable en x_0
- Dans \mathbb{R} une fonction affine f est de la forme $f(x) = ax + b$
- Comme on veut $f(x_0) = h(x_0) = ax_0 + b$ on obtient $f(x) = ax + b = ax + h(x_0) - ax_0 = a(x - x_0) + h(x_0)$

- Comme h est différentiable en x_0 on a par définition :

$$\lim_{x \rightarrow x_0} \frac{|h(x) - (a(x - x_0) + h(x_0))|}{|x - x_0|} = 0$$

- L'équation ci-dessus se simplifie pour donner de manière équivalente :

$$\lim_{x \rightarrow x_0} \frac{h(x) - h(x_0)}{x - x_0} = a$$

- Le nombre a est dénoté $h'(x_0)$ et est dénommé **dérivée** (première) de h en x_0
- La fonction affine f est alors donnée par

$$f(x) = h(x_0) + h'(x_0)(x - x_0)$$

- Cette fonction affine est **tangente** à h en x_0

Matrice des dérivées partielles I

- La dérivée en un point donné d'une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$, peut être représentée par une matrice élément de $\mathcal{M}_{m,n}$
- Afin de déterminer la représentation matricielle \mathbf{L} de la dérivée de la fonction différentiable $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$, nous utilisons la base canonique $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ de \mathbb{R}^n et nous considérons les vecteurs suivant :

$$\mathbf{x}_j = \mathbf{x}_0 + t\mathbf{e}_j, \quad j = 1, \dots, n$$

- Par définition de la différentiabilité, nous avons :

$$\lim_{t \rightarrow 0} \frac{f(\mathbf{x}_j) - (t\mathbf{L}\mathbf{e}_j + f(\mathbf{x}_0))}{t} = \mathbf{0}, \quad j = 1, \dots, n$$

où \mathbf{L} est une matrice de $\mathcal{M}_{m,n}$ représentant une application linéaire

- De manière équivalente, nous avons :

$$\lim_{t \rightarrow 0} \frac{f(\mathbf{x}_j) - f(\mathbf{x}_0)}{t} = \mathbf{L}\mathbf{e}_j, \quad j = 1, \dots, n$$

Matrice des dérivées partielles II

- \mathbf{Le}_j est la j ème colonne de la matrice \mathbf{L} . Par ailleurs, le vecteur \mathbf{x}_j ne diffère de \mathbf{x}_0 que vis à vis de la j ème composante. De plus la différence au niveau de cette composante est de t . Ainsi, le membre de gauche de l'équation précédente est appelée la **dérivée partielle** de f par rapport à x_j que l'on note :

$$\frac{\partial f}{\partial x_j}(\mathbf{x}_0)$$

- La limite d'un vecteur est calculée en considérant la limite de chacune de ses composantes. Ainsi si nous avons :

$$f(\mathbf{x}) = \begin{pmatrix} f_1(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x}) \end{pmatrix}$$

alors la dérivée partielle de f par rapport à x_j est donné par le vecteur suivant :

Matrice des dérivées partielles IV

- En résumé, si $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ est une fonction différentiable en \mathbf{x}_0 alors la dérivée de h en \mathbf{x}_0 est déterminée de façon unique et est représentée par la matrice des dérivées partielles de taille $(m \times n)$ $Dh(\mathbf{x}_0)$. La meilleure approximation affine de f dans le voisinage de \mathbf{x}_0 est alors donnée par :

$$f(\mathbf{x}) = h(\mathbf{x}_0) + Dh(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)$$

dans le sens où $h(\mathbf{x}) = f(\mathbf{x}) + r(\mathbf{x})$ et $\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \|r(\mathbf{x})\|/\|\mathbf{x} - \mathbf{x}_0\| = 0$
Les colonnes de $Df(\mathbf{x}_0)$ sont les vecteurs de dérivées partielles et le vecteur $\frac{\partial f}{\partial x_j}(\mathbf{x}_0)$ est un vecteur tangent en \mathbf{x}_0 à la courbe de f obtenue en variant la j ème composante de \mathbf{x}

Matrice des dérivées partielles III

$$\frac{\partial f}{\partial x_j}(\mathbf{x}_0) = \begin{pmatrix} \frac{\partial f_1}{\partial x_j}(\mathbf{x}_0) \\ \vdots \\ \frac{\partial f_m}{\partial x_j}(\mathbf{x}_0) \end{pmatrix}$$

et la matrice \mathbf{L} est de la forme suivante :

$$\begin{aligned} \mathbf{L} &= \begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}_0) & \cdots & \frac{\partial f}{\partial x_n}(\mathbf{x}_0) \\ \frac{\partial f_1}{\partial x_1}(\mathbf{x}_0) & \cdots & \frac{\partial f_1}{\partial x_n}(\mathbf{x}_0) \\ \vdots & \cdots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathbf{x}_0) & \cdots & \frac{\partial f_m}{\partial x_n}(\mathbf{x}_0) \end{pmatrix} \\ &= \begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}_0) & \cdots & \frac{\partial f}{\partial x_n}(\mathbf{x}_0) \\ \frac{\partial f_1}{\partial x_1}(\mathbf{x}_0) & \cdots & \frac{\partial f_1}{\partial x_n}(\mathbf{x}_0) \\ \vdots & \cdots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathbf{x}_0) & \cdots & \frac{\partial f_m}{\partial x_n}(\mathbf{x}_0) \end{pmatrix} \end{aligned}$$

- \mathbf{L} est appelé la **matrice Jacobienne** ou matrice des dérivées partielles (premières) de f en \mathbf{x}_0 et est également dénotée $Df(\mathbf{x}_0)$

Matrice des dérivées partielles V

- Si $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est différentiable, alors la fonction ∇f défini par :

$$\nabla f(\mathbf{x}) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\mathbf{x}) \end{pmatrix} = Df(\mathbf{x})^t$$

est appelé **gradient** de f

- ∇f est une fonction de \mathbb{R}^n dans \mathbb{R}^n et peut être vue comme un **champ de vecteurs** (fonction qui associe à tout point un vecteur)
- Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$, si ∇f est différentiable, on dit que f est deux fois différentiable et on écrit la dérivée de ∇f de la manière suivante :

$$D^2f = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{pmatrix}$$

Matrice des dérivées partielles VI

La matrice $D^2f(\mathbf{x})$ est appelée **matrice hessienne** de f en \mathbf{x}

Règles de dérivation II

- Soit deux fonctions $f, g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ différentiables sur \mathbb{R}^n . Soit la fonction $h : \mathbb{R}^n \rightarrow \mathbb{R}$ définie par $h(\mathbf{x}) = f(\mathbf{x})^t g(\mathbf{x})$. Alors h est différentiable sur \mathbb{R}^n de dérivée :

$$Dh(\mathbf{x}) = f(\mathbf{x})^t Dg(\mathbf{x}) + g(\mathbf{x})^t Df(\mathbf{x})$$

- Ci-dessous quelques formules de dérivations dans le cas multivarié. La dérivée est calculée par rapport à \mathbf{x} . $\mathbf{A} \in \mathcal{M}_{m,n}$ et $\mathbf{y} \in \mathbb{R}^m$ sont donnés.

- ▶ $D(\mathbf{y}^t \mathbf{A} \mathbf{x}) = \mathbf{y}^t \mathbf{A}$

- ▶ Si $m = n$ alors $D(\mathbf{x}^t \mathbf{A} \mathbf{x}) = \mathbf{x}^t (\mathbf{A} + \mathbf{A}^t)$

Si $\mathbf{y} \in \mathbb{R}^n$, on conclut de la première formule :

- ▶ $D(\mathbf{y}^t \mathbf{x}) = \mathbf{y}^t$

Si \mathbf{A} est symétrique, on conclut de la seconde formule :

- ▶ $D(\mathbf{x}^t \mathbf{A} \mathbf{x}) = 2\mathbf{x}^t \mathbf{A}$

- ▶ $D(\mathbf{x}^t \mathbf{x}) = 2\mathbf{x}^t$

Règles de dérivation I

- Une fonction $f : \mathbb{S} \rightarrow \mathbb{R}^m$, avec $\mathbb{S} \subset \mathbb{R}^n$, est dite **continûment différentiable** sur \mathbb{S} si elle est différentiable sur \mathbb{S} et si $Df : \mathbb{S} \rightarrow \mathbb{R}^m$ est continue i.e. les dérivées partielles sont continues. Dans ce cas, on dit que f est de **classe \mathcal{C}^1**
- Si les dérivées partielles d'ordre p de f sont continues alors on dit que f est de **classe \mathcal{C}^p**
- Notons que si $f \in \mathcal{C}^2$ alors la matrice hessienne est symétrique
- Soit $f : \mathbb{S} \rightarrow \mathbb{R}^m$ une fonction différentiable en $\mathbf{x} \in \mathbb{S} \subset \mathbb{R}^n$ et soit $g : \mathbb{T} \rightarrow \mathbb{R}^p$ une fonction différentiable en $f(\mathbf{x}) \in \mathbb{T} \subset \mathbb{R}^m$. Soit la **composition** $h : \mathbb{S} \rightarrow \mathbb{R}^p$ définie par $h(\mathbf{x}) = g \circ f(\mathbf{x}) = g(f(\mathbf{x}))$. Alors h est différentiable sur \mathbb{S} de dérivée :

$$Dh(\mathbf{x}) = Dg(f(\mathbf{x}))Df(\mathbf{x})$$

Lignes de niveau et gradients I

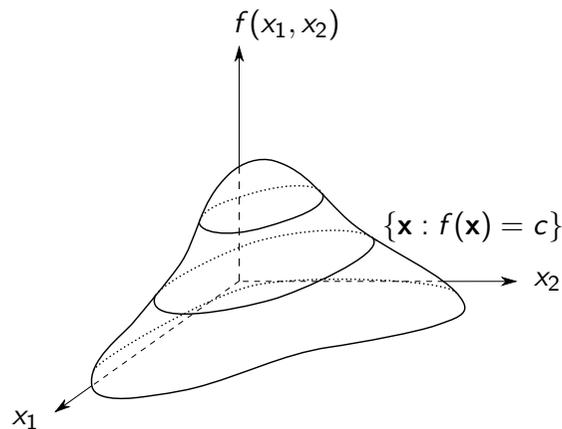
- Soit une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$. On appelle **ligne de niveau** c de la fonction f l'ensemble des points suivants :

$$\mathbb{L} = \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) = c\}$$

- Exemple : pour $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ les lignes de niveau sont en général des courbes tandis que pour $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ les lignes de niveau sont en général des surfaces

Lignes de niveau et gradients II

- Illustration (lignes de niveau)



Lignes de niveau et gradients III

- Prenons \mathbf{x}_0 un point appartenant à \mathbb{L} de $f : f(\mathbf{x}_0) = c$
- Supposons qu'il existe une courbe γ contenue dans \mathbb{L} représentée par une fonction continûment différentiable $g : \mathbb{R} \rightarrow \mathbb{R}^n$. Supposons également que $g(t_0) = \mathbf{x}_0$ et que $Dg(t_0) = \mathbf{v} \neq \mathbf{0}$ de telle sorte que \mathbf{v} est le vecteur tangent à γ en \mathbf{x}_0
- En appliquant la règle de dérivation des fonctions composées à la fonction $h(t) = f(g(t))$ en t_0 on obtient :

$$Dh(t_0) = Df(g(t_0))Dg(t_0) = Df(\mathbf{x}_0)\mathbf{v}$$

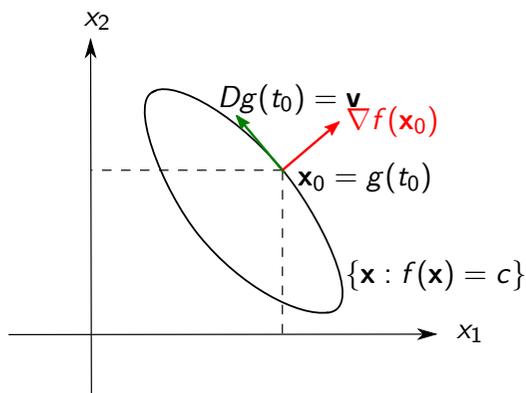
Comme de plus γ est contenue dans \mathbb{L} nous avons :

$h(t) = f(g(t)) = c$ une constante ce qui implique que $Dh(t) = 0$ et donc :

$$Df(\mathbf{x}_0)\mathbf{v} = \nabla f(\mathbf{x}_0)^t \mathbf{v} = 0$$

Lignes de niveau et gradients IV

- Ainsi, le vecteur $\nabla f(\mathbf{x}_0)$ est orthogonal au vecteur tangent à toute courbe passant par \mathbf{x}_0 dans la ligne de niveau \mathbb{L} défini par $f(\mathbf{x}) = f(\mathbf{x}_0)$
- Illustration (orthogonalité entre gradient et ligne de niveau)



Séries de Taylor

- Soit une fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ qui soit m fois continûment différentiable ($f \in \mathcal{C}^m$) sur un interval $[a, b]$. Dénotons $h = b - a$. Alors :

$$f(b) = f(a) + \frac{h}{1!}f^{(1)}(a) + \frac{h^2}{2!}f^{(2)}(a) + \dots + \frac{h^m}{m!}f^{(m)}(a) + o(h^m)$$

où $f^{(i)}$ est la i ème dérivée de f et $o(h^m)$ représente un terme négligeable qui converge vers 0 plus vite que h^m

- Soit maintenant une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Supposons que $f \in \mathcal{C}^2$ et soit \mathbf{x}_0 un point de \mathbb{R}^n . Au voisinage de ce point on a :

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \frac{1}{1!}Df(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + \frac{1}{2!}(\mathbf{x} - \mathbf{x}_0)^t D^2f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + o(\|\mathbf{x} - \mathbf{x}_0\|^2)$$

- Beaucoup de méthodes numériques en optimisation se fondent sur les séries de Taylor

Rappel du Sommaire

- 1 Rappels de concepts mathématiques
- 2 Bases théoriques de l'optimisation
- 3 Optimisation sans contrainte
- 4 Optimisation linéaire avec contraintes
- 5 Optimisation non linéaire avec contraintes
- 6 Optimisation convexe
- 7 Algorithmes pour problèmes contraints

Optimisation

- Rappelons qu'on s'intéresse dans ce cours au problème suivant :

$$\min f(\mathbf{x})$$

sous la contrainte que $\mathbf{x} \in \mathbb{S}$

où :

- ▶ $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est une **fonction objectif** à valeur dans \mathbb{R}
 - ▶ \mathbf{x} est un vecteur de \mathbb{R}^n appelé vecteur des **variables de décision**
 - ▶ \mathbb{S} est un sous-ensemble de \mathbb{R}^n et est appelé **ensemble réalisable**
 - ▶ Si $\mathbb{S} = \mathbb{R}^n$ on dit que le problème est **non contraint**
- Dans cette section, on considère des définitions et des propriétés de base concernant la résolution générale des problèmes d'optimisation (contraint ou non contraint)

Optimisation

- Rappelons qu'on s'intéresse dans ce cours au problème suivant :

$$\min f(\mathbf{x})$$

sous la contrainte que $\mathbf{x} \in \mathbb{S}$

où :

- ▶ $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est une **fonction objectif** à valeur dans \mathbb{R}
- ▶ \mathbf{x} est un vecteur de \mathbb{R}^n appelé vecteur des **variables de décision**
- ▶ \mathbb{S} est un sous-ensemble de \mathbb{R}^n et est appelé **ensemble réalisable**
- ▶ Si $\mathbb{S} = \mathbb{R}^n$ on dit que le problème est **non contraint**

Minimiseur local et global

- On distingue deux types de minimiseurs.

Minimiseur local et global

- On distingue deux types de minimiseurs.

Définition. (Minimiseur local)

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction à valeurs réelles défini sur un ensemble $\mathbb{S} \subset \mathbb{R}^n$. Un point $\mathbf{x}^* \in \mathbb{S}$ est un **minimiseur local** de f sur \mathbb{S} s'il existe $\epsilon > 0$ tel que $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ pour tout $\mathbf{x} \in \mathbb{S} \setminus \{\mathbf{x}^*\}$ et $\|\mathbf{x} - \mathbf{x}^*\| < \epsilon$

Minimiseur local et global

- On distingue deux types de minimiseurs.

Définition. (Minimiseur local)

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction à valeurs réelles défini sur un ensemble $\mathbb{S} \subset \mathbb{R}^n$. Un point $\mathbf{x}^* \in \mathbb{S}$ est un **minimiseur local** de f sur \mathbb{S} s'il existe $\epsilon > 0$ tel que $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ pour tout $\mathbf{x} \in \mathbb{S} \setminus \{\mathbf{x}^*\}$ et $\|\mathbf{x} - \mathbf{x}^*\| < \epsilon$

Définition. (Minimiseur global)

Un point $\mathbf{x}^* \in \mathbb{S}$ est un **minimiseur global** de f sur \mathbb{S} si $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ pour tout $\mathbf{x} \in \mathbb{S} \setminus \{\mathbf{x}^*\}$

Minimiseur local et global

- On distingue deux types de minimiseurs.

Définition. (Minimiseur local)

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction à valeurs réelles défini sur un ensemble $\mathbb{S} \subset \mathbb{R}^n$. Un point $\mathbf{x}^* \in \mathbb{S}$ est un **minimiseur local** de f sur \mathbb{S} s'il existe $\epsilon > 0$ tel que $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ pour tout $\mathbf{x} \in \mathbb{S} \setminus \{\mathbf{x}^*\}$ et $\|\mathbf{x} - \mathbf{x}^*\| < \epsilon$

Définition. (Minimiseur global)

Un point $\mathbf{x}^* \in \mathbb{S}$ est un **minimiseur global** de f sur \mathbb{S} si $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ pour tout $\mathbf{x} \in \mathbb{S} \setminus \{\mathbf{x}^*\}$

- Si dans les définitions précédentes on remplace \geq par $>$ on parle alors de **minimiseur local strict** ou **minimiseur global strict**

Minimiseur local et global

- On distingue deux types de minimiseurs.

Définition. (Minimiseur local)

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction à valeurs réelles défini sur un ensemble $\mathbb{S} \subset \mathbb{R}^n$. Un point $\mathbf{x}^* \in \mathbb{S}$ est un **minimiseur local** de f sur \mathbb{S} s'il existe $\epsilon > 0$ tel que $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ pour tout $\mathbf{x} \in \mathbb{S} \setminus \{\mathbf{x}^*\}$ et $\|\mathbf{x} - \mathbf{x}^*\| < \epsilon$

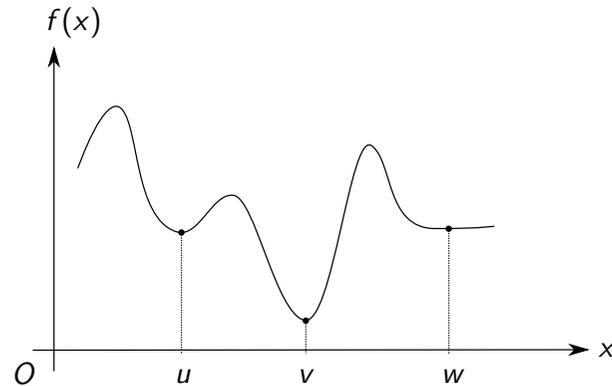
Définition. (Minimiseur global)

Un point $\mathbf{x}^* \in \mathbb{S}$ est un **minimiseur global** de f sur \mathbb{S} si $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ pour tout $\mathbf{x} \in \mathbb{S} \setminus \{\mathbf{x}^*\}$

- Si dans les définitions précédentes on remplace \geq par $>$ on parle alors de **minimiseur local strict** ou **minimiseur global strict**
- La notation $\arg \min_{\mathbf{x} \in \mathbb{S}} f(\mathbf{x})$ représente le **minimiseur global** de f sur \mathbb{S} en supposant que celui-ci existe et est unique

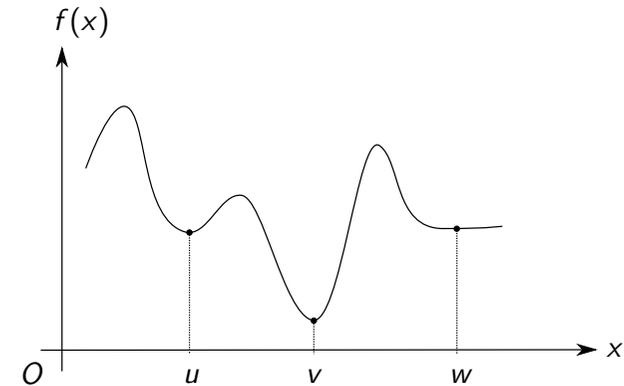
Illustration

Dans le cas $f : \mathbb{R} \rightarrow \mathbb{R}$



Illustration

Dans le cas $f : \mathbb{R} \rightarrow \mathbb{R}$



- u est un minimiseur local strict de f
- v est un minimiseur global strict de f
- w est un minimiseur local de f

Résolution d'un problème d'optimisation

- Dans l'absolu, résoudre un problème d'optimisation signifie déterminer $\arg \min_{x \in S} f(x)$
- Mais en général, il est difficile de déterminer le (ou les) minimiseur global
- En pratique, on se contente de déterminer un (ou des) minimiseur local
- Dans la suite on s'intéresse aux conditions mathématiques nous permettant de caractériser les minimiseurs locaux

Conditions pour un minimiseur local

Les conditions pour un minimiseur local de $f : \mathbb{R}^n \rightarrow \mathbb{R}$ reposent sur la dérivée première et seconde de f :

- La dérivée première de f est dénotée Df et est définie par :

$$Df = \left(\frac{\partial f}{\partial x_1} \quad \cdots \quad \frac{\partial f}{\partial x_n} \right)$$

Conditions pour un minimiseur local

Les conditions pour un minimiseur local de $f : \mathbb{R}^n \rightarrow \mathbb{R}$ reposent sur la dérivée première et seconde de f :

- La dérivée première de f est dénotée Df et est définie par :

$$Df = \left(\frac{\partial f}{\partial x_1} \quad \cdots \quad \frac{\partial f}{\partial x_n} \right)$$

- Le gradient de f est noté ∇f et est égale à Df^t :

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

Exemple

- $f(x_1, x_2) = 5x_1 + 8x_2 + x_1x_2 - x_1^2 - 2x_2^2$

Conditions pour un minimiseur local

Les conditions pour un minimiseur local de $f : \mathbb{R}^n \rightarrow \mathbb{R}$ reposent sur la dérivée première et seconde de f :

- La dérivée première de f est dénotée Df et est définie par :

$$Df = \left(\frac{\partial f}{\partial x_1} \quad \cdots \quad \frac{\partial f}{\partial x_n} \right)$$

- Le gradient de f est noté ∇f et est égale à Df^t :

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

- La dérivée seconde de f est notée D^2f et est définie par :

$$D^2f = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

Exemple

- $f(x_1, x_2) = 5x_1 + 8x_2 + x_1x_2 - x_1^2 - 2x_2^2$

- $Df(\mathbf{x}) = \nabla f(\mathbf{x})^t =$

Exemple

- $f(x_1, x_2) = 5x_1 + 8x_2 + x_1x_2 - x_1^2 - 2x_2^2$
- $Df(\mathbf{x}) = \nabla f(\mathbf{x})^t = \left(\frac{\partial f}{\partial x_1} \quad \frac{\partial f}{\partial x_2} \right) =$

Exemple

- $f(x_1, x_2) = 5x_1 + 8x_2 + x_1x_2 - x_1^2 - 2x_2^2$
- $Df(\mathbf{x}) = \nabla f(\mathbf{x})^t = \left(\frac{\partial f}{\partial x_1} \quad \frac{\partial f}{\partial x_2} \right) = (5 + x_2 - 2x_1 \quad 8 + x_1 - 4x_2)$
- $D^2f(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{x}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_2^2}(\mathbf{x}) \end{pmatrix} =$

Exemple

- $f(x_1, x_2) = 5x_1 + 8x_2 + x_1x_2 - x_1^2 - 2x_2^2$
- $Df(\mathbf{x}) = \nabla f(\mathbf{x})^t = \left(\frac{\partial f}{\partial x_1} \quad \frac{\partial f}{\partial x_2} \right) = (5 + x_2 - 2x_1 \quad 8 + x_1 - 4x_2)$
- $D^2f(\mathbf{x}) =$

Exemple

- $f(x_1, x_2) = 5x_1 + 8x_2 + x_1x_2 - x_1^2 - 2x_2^2$
- $Df(\mathbf{x}) = \nabla f(\mathbf{x})^t = \left(\frac{\partial f}{\partial x_1} \quad \frac{\partial f}{\partial x_2} \right) = (5 + x_2 - 2x_1 \quad 8 + x_1 - 4x_2)$
- $D^2f(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{x}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_2^2}(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} -2 & 1 \\ 1 & -4 \end{pmatrix}$

Direction admissible

- Etant donné un problème d'optimisation sur un ensemble réalisable \mathbb{S} , un minimiseur peut être un point intérieur, ou un point frontière de \mathbb{S}

Direction admissible

- Etant donné un problème d'optimisation sur un ensemble réalisable \mathbb{S} , un minimiseur peut être un point intérieur, ou un point frontière de \mathbb{S}
- Afin d'étudier le cas où le minimiseur se trouve sur la frontière, on a besoin d'introduire le concept de direction admissible :

Direction admissible

- Etant donné un problème d'optimisation sur un ensemble réalisable \mathbb{S} , un minimiseur peut être un point intérieur, ou un point frontière de \mathbb{S}
- Afin d'étudier le cas où le minimiseur se trouve sur la frontière, on a besoin d'introduire le concept de direction admissible :

Définition. (Direction admissible)

Un vecteur $\mathbf{d} \in \mathbb{R}^n$ tel que $\mathbf{d} \neq \mathbf{0}$ est une **direction admissible** en $\mathbf{x} \in \mathbb{S}$ s'il existe $\alpha_0 > 0$ tel que $\mathbf{x} + \alpha \mathbf{d} \in \mathbb{S}$ pour tout $\alpha \in [0, \alpha_0]$

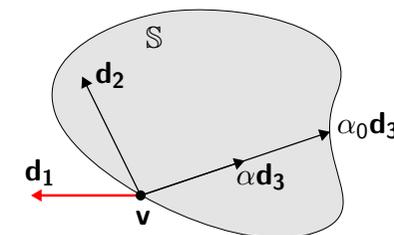
Direction admissible

- Etant donné un problème d'optimisation sur un ensemble réalisable \mathbb{S} , un minimiseur peut être un point intérieur, ou un point frontière de \mathbb{S}
- Afin d'étudier le cas où le minimiseur se trouve sur la frontière, on a besoin d'introduire le concept de direction admissible :

Définition. (Direction admissible)

Un vecteur $\mathbf{d} \in \mathbb{R}^n$ tel que $\mathbf{d} \neq \mathbf{0}$ est une **direction admissible** en $\mathbf{x} \in \mathbb{S}$ s'il existe $\alpha_0 > 0$ tel que $\mathbf{x} + \alpha \mathbf{d} \in \mathbb{S}$ pour tout $\alpha \in [0, \alpha_0]$

- Illustration :



Dérivée directionnelle

- Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ et \mathbf{d} une direction admissible en $\mathbf{x} \in \mathbb{S}$. La **dérivée directionnelle** de f dans la direction de \mathbf{d} , dénotée $\frac{\partial f}{\partial \mathbf{d}}$, est la fonction à valeurs réelles définie par :

$$\frac{\partial f}{\partial \mathbf{d}}(\mathbf{x}) = \lim_{\alpha \rightarrow 0} \frac{f(\mathbf{x} + \alpha \mathbf{d}) - f(\mathbf{x})}{\alpha}$$

Dérivée directionnelle

- Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ et \mathbf{d} une direction admissible en $\mathbf{x} \in \mathbb{S}$. La **dérivée directionnelle** de f dans la direction de \mathbf{d} , dénotée $\frac{\partial f}{\partial \mathbf{d}}$, est la fonction à valeurs réelles définie par :

$$\frac{\partial f}{\partial \mathbf{d}}(\mathbf{x}) = \lim_{\alpha \rightarrow 0} \frac{f(\mathbf{x} + \alpha \mathbf{d}) - f(\mathbf{x})}{\alpha}$$

- Si $\|\mathbf{d}\| = 1$, alors $\frac{\partial f}{\partial \mathbf{d}}(\mathbf{x})$ est le **taux d'accroissement (ou pente)** de f en \mathbf{x} dans la direction \mathbf{d}

Dérivée directionnelle

- Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ et \mathbf{d} une direction admissible en $\mathbf{x} \in \mathbb{S}$. La **dérivée directionnelle** de f dans la direction de \mathbf{d} , dénotée $\frac{\partial f}{\partial \mathbf{d}}$, est la fonction à valeurs réelles définie par :

$$\frac{\partial f}{\partial \mathbf{d}}(\mathbf{x}) = \lim_{\alpha \rightarrow 0} \frac{f(\mathbf{x} + \alpha \mathbf{d}) - f(\mathbf{x})}{\alpha}$$

- Si $\|\mathbf{d}\| = 1$, alors $\frac{\partial f}{\partial \mathbf{d}}(\mathbf{x})$ est le **taux d'accroissement (ou pente)** de f en \mathbf{x} dans la direction \mathbf{d}
- Pour calculer la dérivée directionnelle $\frac{\partial f}{\partial \mathbf{d}}$ supposons que \mathbf{x} et \mathbf{d} sont donnés. Dans ce cas, $f(\mathbf{x} + \alpha \mathbf{d}) = f(g(\alpha))$ est une fonction de α de \mathbb{R} dans \mathbb{R} avec $g : \mathbb{R} \rightarrow \mathbb{R}^n$ et nous avons :

$$\begin{aligned} \frac{\partial f}{\partial \mathbf{d}}(\mathbf{x}) &= \left. \frac{df \circ g}{d\alpha}(\alpha) \right|_{\alpha=0} \\ &= Df(g(0)) dg(0) = Df(\mathbf{x}) \mathbf{d} \\ &= \nabla f(\mathbf{x})^t \mathbf{d} = \langle \nabla f(\mathbf{x}), \mathbf{d} \rangle \end{aligned}$$

Dérivée directionnelle

- Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ et \mathbf{d} une direction admissible en $\mathbf{x} \in \mathbb{S}$. La **dérivée directionnelle** de f dans la direction de \mathbf{d} , dénotée $\frac{\partial f}{\partial \mathbf{d}}$, est la fonction à valeurs réelles définie par :

$$\frac{\partial f}{\partial \mathbf{d}}(\mathbf{x}) = \lim_{\alpha \rightarrow 0} \frac{f(\mathbf{x} + \alpha \mathbf{d}) - f(\mathbf{x})}{\alpha}$$

- Si $\|\mathbf{d}\| = 1$, alors $\frac{\partial f}{\partial \mathbf{d}}(\mathbf{x})$ est le **taux d'accroissement (ou pente)** de f en \mathbf{x} dans la direction \mathbf{d}
- Pour calculer la dérivée directionnelle $\frac{\partial f}{\partial \mathbf{d}}$ supposons que \mathbf{x} et \mathbf{d} sont donnés. Dans ce cas, $f(\mathbf{x} + \alpha \mathbf{d}) = f(g(\alpha))$ est une fonction de α de \mathbb{R} dans \mathbb{R} avec $g : \mathbb{R} \rightarrow \mathbb{R}^n$ et nous avons :

$$\begin{aligned} \frac{\partial f}{\partial \mathbf{d}}(\mathbf{x}) &= \left. \frac{df \circ g}{d\alpha}(\alpha) \right|_{\alpha=0} \\ &= Df(g(0)) dg(0) = Df(\mathbf{x}) \mathbf{d} \\ &= \nabla f(\mathbf{x})^t \mathbf{d} = \langle \nabla f(\mathbf{x}), \mathbf{d} \rangle \end{aligned}$$

- Si $\|\mathbf{d}\| = 1$, alors $\langle \nabla f(\mathbf{x}), \mathbf{d} \rangle = \langle \mathbf{d}, \nabla f(\mathbf{x}) \rangle$ est le taux d'accroissement de f en \mathbf{x} dans la direction \mathbf{d}

Exemple

- Soit $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ tel que $f(x_1, x_2, x_3) = x_1 x_2 x_3$ et soit $\mathbf{d} = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{\sqrt{2}}\right)$

$$\frac{\partial f}{\partial \mathbf{d}}(\mathbf{x}) =$$

Exemple

- Soit $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ tel que $f(x_1, x_2, x_3) = x_1 x_2 x_3$ et soit $\mathbf{d} = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{\sqrt{2}}\right)$
- La dérivée directionnelle de f dans la direction de \mathbf{d} est :

$$\frac{\partial f}{\partial \mathbf{d}}(\mathbf{x}) = \nabla f(\mathbf{x})^t \mathbf{d} =$$

Exemple

- Soit $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ tel que $f(x_1, x_2, x_3) = x_1 x_2 x_3$ et soit $\mathbf{d} = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{\sqrt{2}}\right)$
- La dérivée directionnelle de f dans la direction de \mathbf{d} est :

$$\frac{\partial f}{\partial \mathbf{d}}(\mathbf{x}) =$$

Exemple

- Soit $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ tel que $f(x_1, x_2, x_3) = x_1 x_2 x_3$ et soit $\mathbf{d} = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{\sqrt{2}}\right)$
- La dérivée directionnelle de f dans la direction de \mathbf{d} est :

$$\frac{\partial f}{\partial \mathbf{d}}(\mathbf{x}) = \nabla f(\mathbf{x})^t \mathbf{d} = (x_2 x_3 \quad x_1 x_3 \quad x_1 x_2) \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \frac{x_2 x_3 + x_1 x_3 + \sqrt{2} x_1 x_2}{2}$$

Exemple

- Soit $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ tel que $f(x_1, x_2, x_3) = x_1 x_2 x_3$ et soit $\mathbf{d} = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{\sqrt{2}}\right)$
- La dérivée directionnelle de f dans la direction de \mathbf{d} est :

$$\frac{\partial f}{\partial \mathbf{d}}(\mathbf{x}) = \nabla f(\mathbf{x})^t \mathbf{d} = (x_2 x_3 \quad x_1 x_3 \quad x_1 x_2) \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \frac{x_2 x_3 + x_1 x_3 + \sqrt{2} x_1 x_2}{2}$$

- Remarque : comme $\|\mathbf{d}\| = 1$, l'équation ci-dessus est également le taux d'accroissement de f en \mathbf{x} dans la direction de \mathbf{d}

Condition nécessaire du 1er ordre (CNPO)

Théorème. (Condition nécessaire du 1er ordre (CNPO))

Soit \mathbb{S} un sous-ensemble de \mathbb{R}^n et f une fonction de classe \mathcal{C}^1 de \mathbb{S} dans \mathbb{R} . Si \mathbf{x}^* est un minimiseur local de f sur \mathbb{S} alors pour toute direction admissible \mathbf{d} en \mathbf{x} , nous avons :

$$\mathbf{d}^t \nabla f(\mathbf{x}^*) \geq 0$$

Condition nécessaire du 1er ordre (CNPO)

Théorème. (Condition nécessaire du 1er ordre (CNPO))

Soit \mathbb{S} un sous-ensemble de \mathbb{R}^n et f une fonction de classe \mathcal{C}^1 de \mathbb{S} dans \mathbb{R} . Si \mathbf{x}^* est un minimiseur local de f sur \mathbb{S} alors pour toute direction admissible \mathbf{d} en \mathbf{x} , nous avons :

$$\mathbf{d}^t \nabla f(\mathbf{x}^*) \geq 0$$

- Une autre façon équivalente d'exprimer la CNPO est la suivante :

$$\frac{\partial f}{\partial \mathbf{d}}(\mathbf{x}^*) \geq 0, \text{ pour toute direction admissible en } \mathbf{x}^*$$

Preuve de la CNPO

Démonstration.

- Supposons que \mathbf{x}^* est un minimiseur local. Alors, pour toute direction admissible \mathbf{d} , il existe $\alpha_0 > 0$ tel que pour tout $\alpha \in [0, \alpha_0]$:

$$f(\mathbf{x}^*) \leq f(\mathbf{x}^* + \alpha \mathbf{d})$$

Preuve de la CNPO

Démonstration.

- Supposons que \mathbf{x}^* est un minimiseur local. Alors, pour toute direction admissible \mathbf{d} , il existe $\alpha_0 > 0$ tel que pour tout $\alpha \in [0, \alpha_0]$:

$$f(\mathbf{x}^*) \leq f(\mathbf{x}^* + \alpha \mathbf{d})$$

- Par conséquent, pour tout $\alpha \in [0, \alpha_0]$, nous avons :

$$\frac{f(\mathbf{x}^* + \alpha \mathbf{d}) - f(\mathbf{x}^*)}{\alpha} \geq 0$$

Preuve de la CNPO

Démonstration.

- Supposons que \mathbf{x}^* est un minimiseur local. Alors, pour toute direction admissible \mathbf{d} , il existe $\alpha_0 > 0$ tel que pour tout $\alpha \in [0, \alpha_0]$:

$$f(\mathbf{x}^*) \leq f(\mathbf{x}^* + \alpha \mathbf{d})$$

- Par conséquent, pour tout $\alpha \in [0, \alpha_0]$, nous avons :

$$\frac{f(\mathbf{x}^* + \alpha \mathbf{d}) - f(\mathbf{x}^*)}{\alpha} \geq 0$$

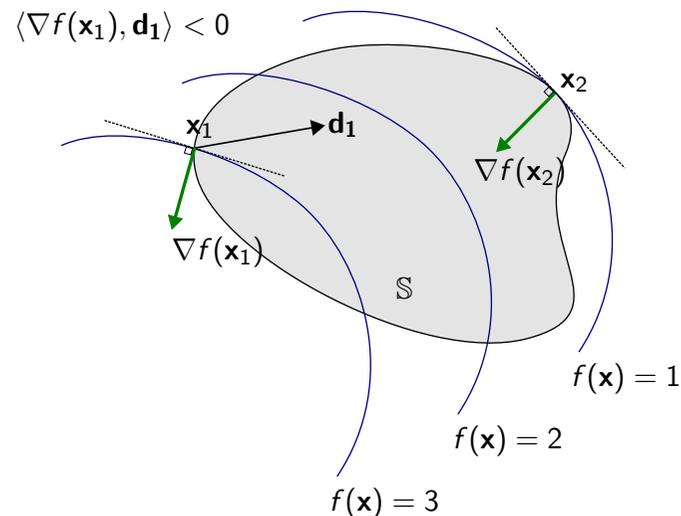
- En considérant la limite quand $\alpha \rightarrow 0$, on conclut donc que si \mathbf{x}^* est un minimiseur local, alors :

$$\frac{\partial f}{\partial \mathbf{d}}(\mathbf{x}^*) \geq 0$$

□

Illustration

- Cas d'un minimiseur sur la frontière de \mathbb{S} :



Condition nécessaire du 1er ordre (CNPO) (point intérieur)

Théorème. (Condition nécessaire du 1er ordre (CNPO) (point intérieur))

Soit \mathbb{S} un sous-ensemble de \mathbb{R}^n et f une fonction de classe \mathcal{C}^1 de \mathbb{S} dans \mathbb{R} . Si \mathbf{x}^* est un minimiseur local de f sur \mathbb{S} et si \mathbf{x}^* est un point intérieur alors toutes les directions en \mathbf{x}^* sont admissibles et dans ce cas la CNPO devient :

$$\nabla f(\mathbf{x}^*) = \mathbf{0}$$

On dit alors que \mathbf{x}^* est un **point critique** (ou stationnaire)

Condition nécessaire du 1er ordre (CNPO) (point intérieur)

Théorème. (Condition nécessaire du 1er ordre (CNPO) (point intérieur))

Soit \mathbb{S} un sous-ensemble de \mathbb{R}^n et f une fonction de classe \mathcal{C}^1 de \mathbb{S} dans \mathbb{R} . Si \mathbf{x}^* est un minimiseur local de f sur \mathbb{S} et si \mathbf{x}^* est un point intérieur alors toutes les directions en \mathbf{x}^* sont admissibles et dans ce cas la CNPO devient :

$$\nabla f(\mathbf{x}^*) = \mathbf{0}$$

On dit alors que \mathbf{x}^* est un **point critique** (ou stationnaire)

Démonstration.

A faire □

Condition nécessaire du 2nd ordre (CNSO)

Théorème. (Condition nécessaire du 2nd ordre (CNSO))

Soit \mathbb{S} un sous-ensemble de \mathbb{R}^n et f une fonction de classe \mathcal{C}^2 de \mathbb{S} dans \mathbb{R} , \mathbf{x}^* est un minimiseur local de f sur \mathbb{S} et \mathbf{d} est une direction admissible en \mathbf{x} . Si $\mathbf{d}^t \nabla f(\mathbf{x}^*) = 0$ alors nous avons :

$$\mathbf{d}^t D^2 f(\mathbf{x}^*) \mathbf{d} \geq 0$$

où $D^2 f$ est la matrice hessienne de f

Preuve de la CNSO

Démonstration.

Preuve par l'absurde

- Supposons qu'il existe une direction admissible \mathbf{d} en \mathbf{x}^* telle que $\mathbf{d}^t \nabla f(\mathbf{x}^*) = 0$ et $\mathbf{d}^t D^2 f(\mathbf{x}^*) \mathbf{d} < 0$

Preuve de la CNSO

Démonstration.

Preuve par l'absurde

- Supposons qu'il existe une direction admissible \mathbf{d} en \mathbf{x}^* telle que $\mathbf{d}^t \nabla f(\mathbf{x}^*) = 0$ et $\mathbf{d}^t D^2 f(\mathbf{x}^*) \mathbf{d} < 0$
- Soit $\mathbf{x}(\alpha) = \mathbf{x}^* + \alpha \mathbf{d}$ et la fonction composée $\phi(\alpha) = f(\mathbf{x}^* + \alpha \mathbf{d}) = f(\mathbf{x}(\alpha))$. On a alors la série de Taylor suivante :

$$\phi(\alpha) = \phi(0) + \phi'(0)\alpha + \phi''(0)\frac{\alpha^2}{2} + o(\alpha^2)$$
 avec $\phi'(0) = \langle \nabla f(\mathbf{x}), \mathbf{d} \rangle = 0$ et $\phi''(0) = \mathbf{d}^t D^2 f(\mathbf{x}^*) \mathbf{d} < 0$

Preuve de la CNSO

Démonstration.

Preuve par l'absurde

- Supposons qu'il existe une direction admissible \mathbf{d} en \mathbf{x}^* telle que $\mathbf{d}^t \nabla f(\mathbf{x}^*) = 0$ et $\mathbf{d}^t D^2 f(\mathbf{x}^*) \mathbf{d} < 0$
- Soit $\mathbf{x}(\alpha) = \mathbf{x}^* + \alpha \mathbf{d}$ et la fonction composée $\phi(\alpha) = f(\mathbf{x}^* + \alpha \mathbf{d}) = f(\mathbf{x}(\alpha))$. On a alors la série de Taylor suivante :

$$\phi(\alpha) = \phi(0) + \phi'(0)\alpha + \phi''(0)\frac{\alpha^2}{2} + o(\alpha^2)$$
avec $\phi'(0) = \langle \nabla f(\mathbf{x}), \mathbf{d} \rangle = 0$ et $\phi''(0) = \mathbf{d}^t D^2 f(\mathbf{x}^*) \mathbf{d} < 0$
- Pour α suffisamment petit on a : $\phi(\alpha) - \phi(0) = \phi''(0)\frac{\alpha^2}{2} + o(\alpha^2) < 0$

Preuve de la CNSO

Démonstration.

Preuve par l'absurde

- Supposons qu'il existe une direction admissible \mathbf{d} en \mathbf{x}^* telle que $\mathbf{d}^t \nabla f(\mathbf{x}^*) = 0$ et $\mathbf{d}^t D^2 f(\mathbf{x}^*) \mathbf{d} < 0$
- Soit $\mathbf{x}(\alpha) = \mathbf{x}^* + \alpha \mathbf{d}$ et la fonction composée $\phi(\alpha) = f(\mathbf{x}^* + \alpha \mathbf{d}) = f(\mathbf{x}(\alpha))$. On a alors la série de Taylor suivante :

$$\phi(\alpha) = \phi(0) + \phi'(0)\alpha + \phi''(0)\frac{\alpha^2}{2} + o(\alpha^2)$$
avec $\phi'(0) = \langle \nabla f(\mathbf{x}), \mathbf{d} \rangle = 0$ et $\phi''(0) = \mathbf{d}^t D^2 f(\mathbf{x}^*) \mathbf{d} < 0$
- Pour α suffisamment petit on a : $\phi(\alpha) - \phi(0) = \phi''(0)\frac{\alpha^2}{2} + o(\alpha^2) < 0$ ce qui implique que $f(\mathbf{x}^* + \alpha \mathbf{d}) < f(\mathbf{x}^*)$

Preuve de la CNSO

Démonstration.

Preuve par l'absurde

- Supposons qu'il existe une direction admissible \mathbf{d} en \mathbf{x}^* telle que $\mathbf{d}^t \nabla f(\mathbf{x}^*) = 0$ et $\mathbf{d}^t D^2 f(\mathbf{x}^*) \mathbf{d} < 0$
- Soit $\mathbf{x}(\alpha) = \mathbf{x}^* + \alpha \mathbf{d}$ et la fonction composée $\phi(\alpha) = f(\mathbf{x}^* + \alpha \mathbf{d}) = f(\mathbf{x}(\alpha))$. On a alors la série de Taylor suivante :

$$\phi(\alpha) = \phi(0) + \phi'(0)\alpha + \phi''(0)\frac{\alpha^2}{2} + o(\alpha^2)$$
avec $\phi'(0) = \langle \nabla f(\mathbf{x}), \mathbf{d} \rangle = 0$ et $\phi''(0) = \mathbf{d}^t D^2 f(\mathbf{x}^*) \mathbf{d} < 0$
- Pour α suffisamment petit on a : $\phi(\alpha) - \phi(0) = \phi''(0)\frac{\alpha^2}{2} + o(\alpha^2) < 0$ ce qui implique que $f(\mathbf{x}^* + \alpha \mathbf{d}) < f(\mathbf{x}^*)$ ce qui contredit le fait que \mathbf{x}^* soit un minimiseur local

□

Condition nécessaire du 2nd ordre (CNSO) (point intérieur)

Théorème. (Condition nécessaire du 2nd ordre (CNSO) (point intérieur))

Soit \mathbb{S} un sous-ensemble de \mathbb{R}^n et f une fonction de classe \mathcal{C}^2 de \mathbb{S} dans \mathbb{R} . Si \mathbf{x}^* est un minimiseur local de f sur \mathbb{S} et si \mathbf{x}^* est un point intérieur alors :

- \mathbf{x}^* est un point critique :

$$\nabla f(\mathbf{x}^*) = 0$$

- et $D^2 f(\mathbf{x}^*)$ est semi-définie positive (ce qu'on notera par $D^2 f(\mathbf{x}^*) \geq 0$), ie $\forall \mathbf{d} \in \mathbb{R}^n$:

$$\mathbf{d}^t D^2 f(\mathbf{x}^*) \mathbf{d} \geq 0$$

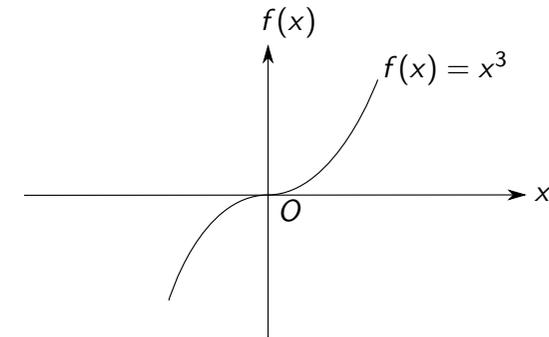
où $D^2 f$ est la matrice hessienne de f

CNPO et CNSO sont nécessaires mais pas suffisantes

- Soit $f(x) = x^3$
- $df(0) = f'(0) = 0$
- $d^2f(0) = f''(0) = 0$
- Mais $x = 0$ n'est pas un minimiseur

CNPO et CNSO sont nécessaires mais pas suffisantes

- Soit $f(x) = x^3$
- $df(0) = f'(0) = 0$
- $d^2f(0) = f''(0) = 0$
- Mais $x = 0$ n'est pas un minimiseur
- Illustration :



Condition suffisante du 2nd ordre (CSSO) (point intérieur)

Théorème. (Condition suffisante du 2nd ordre (CSSO) (point intérieur))

Soit f une fonction de classe \mathcal{C}^2 de $\mathbb{S} \subset \mathbb{R}^n$ dans \mathbb{R} . Soit $\mathbf{x}^* \in \mathbb{S}$ un point intérieur de \mathbb{S} . Si les deux conditions suivantes sont satisfaites :

- 1 $\nabla f(\mathbf{x}^*) = \mathbf{0}$
- 2 $D^2f(\mathbf{x}^*) > 0$ (matrice définie positive)

alors \mathbf{x}^* est un minimiseur local strict de f

Preuve de la CSSO (point intérieur)

Démonstration.

- Comme $f \in \mathcal{C}^2$, la matrice hessienne D^2f est symétrique et donc $D^2f(\mathbf{x}^*) = (D^2f)^t(\mathbf{x}^*)$

Preuve de la CSSO (point intérieur)

Démonstration.

- Comme $f \in \mathcal{C}^2$, la matrice hessienne D^2f est symétrique et donc $D^2f(\mathbf{x}^*) = (D^2f)^t(\mathbf{x}^*)$
- Supposons que $D^2f(\mathbf{x}^*) > 0$ (2ème condition). En utilisant l'inégalité de Rayleigh, nous avons $\forall \mathbf{d} \neq \mathbf{0}$:
 $0 < \lambda_{\min}(D^2f(\mathbf{x}^*))\|\mathbf{d}\|^2 \leq \mathbf{d}^t D^2f(\mathbf{x}^*) \mathbf{d}$

Preuve de la CSSO (point intérieur)

Démonstration.

- Comme $f \in \mathcal{C}^2$, la matrice hessienne D^2f est symétrique et donc $D^2f(\mathbf{x}^*) = (D^2f)^t(\mathbf{x}^*)$
- Supposons que $D^2f(\mathbf{x}^*) > 0$ (2ème condition). En utilisant l'inégalité de Rayleigh, nous avons $\forall \mathbf{d} \neq \mathbf{0}$:
 $0 < \lambda_{\min}(D^2f(\mathbf{x}^*))\|\mathbf{d}\|^2 \leq \mathbf{d}^t D^2f(\mathbf{x}^*) \mathbf{d}$
- En supposant que $\nabla f(\mathbf{x}^*) = \mathbf{0}$ (condition 1) et en utilisant les séries de Taylor, nous obtenons :
 $f(\mathbf{x}^* + \mathbf{d}) - f(\mathbf{x}^*) = \frac{1}{2} \mathbf{d}^t D^2f(\mathbf{x}^*) \mathbf{d} + o(\|\mathbf{d}^2\|)$

Preuve de la CSSO (point intérieur)

Démonstration.

- Comme $f \in \mathcal{C}^2$, la matrice hessienne D^2f est symétrique et donc $D^2f(\mathbf{x}^*) = (D^2f)^t(\mathbf{x}^*)$
- Supposons que $D^2f(\mathbf{x}^*) > 0$ (2ème condition). En utilisant l'inégalité de Rayleigh, nous avons $\forall \mathbf{d} \neq \mathbf{0}$:
 $0 < \lambda_{\min}(D^2f(\mathbf{x}^*))\|\mathbf{d}\|^2 \leq \mathbf{d}^t D^2f(\mathbf{x}^*) \mathbf{d}$
- En supposant que $\nabla f(\mathbf{x}^*) = \mathbf{0}$ (condition 1) et en utilisant les séries de Taylor, nous obtenons :

$$f(\mathbf{x}^* + \mathbf{d}) - f(\mathbf{x}^*) = \frac{1}{2} \mathbf{d}^t D^2f(\mathbf{x}^*) \mathbf{d} + o(\|\mathbf{d}^2\|)$$

$$\geq \frac{\lambda_{\min}(D^2f(\mathbf{x}^*))}{2} \|\mathbf{d}\|^2 + o(\|\mathbf{d}^2\|)$$

Preuve de la CSSO (point intérieur)

Démonstration.

- Comme $f \in \mathcal{C}^2$, la matrice hessienne D^2f est symétrique et donc $D^2f(\mathbf{x}^*) = (D^2f)^t(\mathbf{x}^*)$
- Supposons que $D^2f(\mathbf{x}^*) > 0$ (2ème condition). En utilisant l'inégalité de Rayleigh, nous avons $\forall \mathbf{d} \neq \mathbf{0}$:
 $0 < \lambda_{\min}(D^2f(\mathbf{x}^*))\|\mathbf{d}\|^2 \leq \mathbf{d}^t D^2f(\mathbf{x}^*) \mathbf{d}$
- En supposant que $\nabla f(\mathbf{x}^*) = \mathbf{0}$ (condition 1) et en utilisant les séries de Taylor, nous obtenons :

$$f(\mathbf{x}^* + \mathbf{d}) - f(\mathbf{x}^*) = \frac{1}{2} \mathbf{d}^t D^2f(\mathbf{x}^*) \mathbf{d} + o(\|\mathbf{d}^2\|)$$

$$\geq \frac{\lambda_{\min}(D^2f(\mathbf{x}^*))}{2} \|\mathbf{d}\|^2 + o(\|\mathbf{d}^2\|)$$

$$> 0$$
- On déduit que $\forall \mathbf{d}$ de norme suffisamment petite on a :
 $f(\mathbf{x}^* + \mathbf{d}) > f(\mathbf{x}^*)$

Exemple

- Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ tel que $f(\mathbf{x}) = x_1^2 + x_2^2$
- Nous avons :

$$\nabla f(\mathbf{x}) =$$

Exemple

- Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ tel que $f(\mathbf{x}) = x_1^2 + x_2^2$
- Nous avons :

$$\nabla f(\mathbf{x}) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix}$$

- Par ailleurs : $\nabla f(\mathbf{x}) = \mathbf{0} \Rightarrow$

Exemple

- Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ tel que $f(\mathbf{x}) = x_1^2 + x_2^2$
- Nous avons :

$$\nabla f(\mathbf{x}) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix}$$

Exemple

- Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ tel que $f(\mathbf{x}) = x_1^2 + x_2^2$
- Nous avons :

$$\nabla f(\mathbf{x}) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix}$$

- Par ailleurs : $\nabla f(\mathbf{x}) = \mathbf{0} \Rightarrow \mathbf{x} = (0, 0)$

Exemple

- Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ tel que $f(\mathbf{x}) = x_1^2 + x_2^2$
- Nous avons :

$$\nabla f(\mathbf{x}) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix}$$

- Par ailleurs : $\nabla f(\mathbf{x}) = \mathbf{0} \Rightarrow \mathbf{x} = (0, 0)$
- Nous avons de plus :

$$D^2f(\mathbf{x}) =$$

Exemple

- Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ tel que $f(\mathbf{x}) = x_1^2 + x_2^2$
- Nous avons :

$$\nabla f(\mathbf{x}) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix}$$

- Par ailleurs : $\nabla f(\mathbf{x}) = \mathbf{0} \Rightarrow \mathbf{x} = (0, 0)$
- Nous avons de plus :

$$D^2f(\mathbf{x}) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

Exemple

- Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ tel que $f(\mathbf{x}) = x_1^2 + x_2^2$
- Nous avons :

$$\nabla f(\mathbf{x}) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix}$$

- Par ailleurs : $\nabla f(\mathbf{x}) = \mathbf{0} \Rightarrow \mathbf{x} = (0, 0)$
- Nous avons de plus :

$$D^2f(\mathbf{x}) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

qui est définie positive

Exemple

- Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ tel que $f(\mathbf{x}) = x_1^2 + x_2^2$
- Nous avons :

$$\nabla f(\mathbf{x}) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix}$$

- Par ailleurs : $\nabla f(\mathbf{x}) = \mathbf{0} \Rightarrow \mathbf{x} = (0, 0)$
- Nous avons de plus :

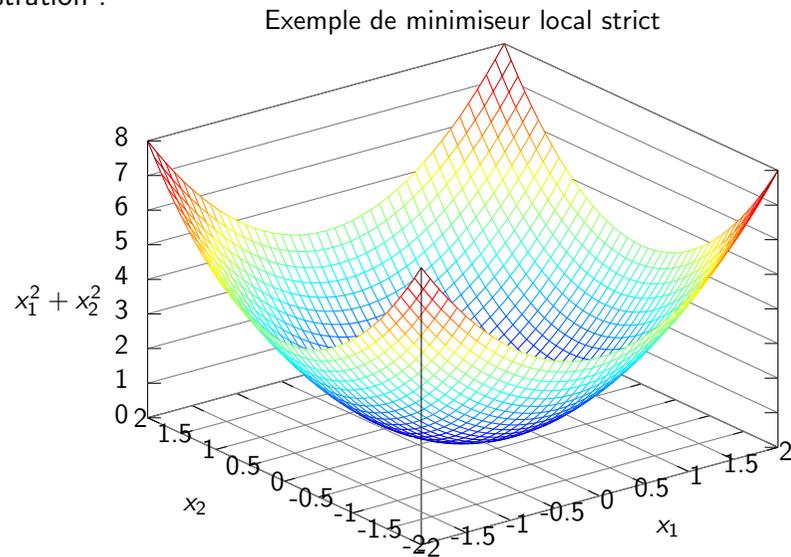
$$D^2f(\mathbf{x}) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

qui est définie positive

- $\mathbf{x} = \mathbf{0} = (0, 0)$ vérifie les CNPO, CNSO et CSSO donc il s'agit d'un minimiseur local strict et $f(\mathbf{0}) = 0$ est le minimum local de f

Exemple (suite)

• Illustration :



Rappel du Sommaire

- 1 Rappels de concepts mathématiques
- 2 Bases théoriques de l'optimisation
- 3 Optimisation sans contrainte
 - Optimisation unidimensionnelle
 - Optimisation multidimensionnelle
- 4 Optimisation linéaire avec contraintes
- 5 Optimisation non linéaire avec contraintes
- 6 Optimisation convexe

Rappel du Sommaire

- 1 Rappels de concepts mathématiques
- 2 Bases théoriques de l'optimisation
- 3 Optimisation sans contrainte
- 4 Optimisation linéaire avec contraintes
- 5 Optimisation non linéaire avec contraintes
- 6 Optimisation convexe
- 7 Algorithmes pour problèmes contraints

Optimisation unidimensionnelle

- Nous nous intéressons exclusivement dans cette sous-section aux fonctions objectif $f : \mathbb{R} \rightarrow \mathbb{R}$
- Nous étudions des méthodes permettant de déterminer un minimiseur de f sur un intervalle $[a_0, b_0]$
- Nous supposons que f est **unimodale** sur $[a_0, b_0]$ ce qui signifie que f a un unique minimiseur local sur cet intervalle

Fonction unimodale

Définition. (Fonction unimodale)

On dit que $f : \mathbb{R} \rightarrow \mathbb{R}$ est unimodale sur un intervalle $[a_0, b_0]$ si elle admet un minimum x^* sur $[a_0, b_0]$ et si $\forall a_1 < b_1$ dans $[a_0, b_0]$:

- $b_1 \leq x^* \Rightarrow f(a_1) > f(b_1)$ (équivalent à $f(a_1) \leq f(b_1) \Rightarrow x^* \leq b_1$)
- $a_1 \geq x^* \Rightarrow f(a_1) < f(b_1)$ (équivalent à $f(a_1) \geq f(b_1) \Rightarrow a_1 \leq x^*$)

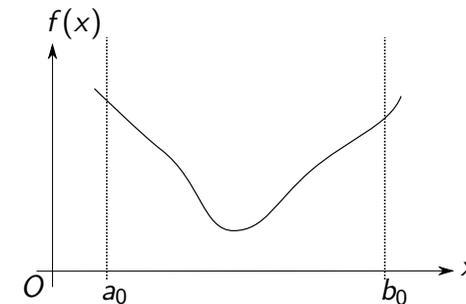
Fonction unimodale

Définition. (Fonction unimodale)

On dit que $f : \mathbb{R} \rightarrow \mathbb{R}$ est unimodale sur un intervalle $[a_0, b_0]$ si elle admet un minimum x^* sur $[a_0, b_0]$ et si $\forall a_1 < b_1$ dans $[a_0, b_0]$:

- $b_1 \leq x^* \Rightarrow f(a_1) > f(b_1)$ (équivalent à $f(a_1) \leq f(b_1) \Rightarrow x^* \leq b_1$)
- $a_1 \geq x^* \Rightarrow f(a_1) < f(b_1)$ (équivalent à $f(a_1) \geq f(b_1) \Rightarrow a_1 \leq x^*$)

Illustration :



Méthodes de type “encadrement de minimiseur”

Principe :

- On évalue la fonction objectif à différents points de $[a_0, b_0]$
- On choisit ces points de sorte à approximer le minimiseur appartenant à un intervalle plus petit
- On rétrécit donc progressivement l'intervalle jusqu'à obtenir une précision suffisamment fine de x^*

Méthodes de type “encadrement de minimiseur”

Principe :

- On évalue la fonction objectif à différents points de $[a_0, b_0]$
- On choisit ces points de sorte à approximer le minimiseur appartenant à un intervalle plus petit
- On rétrécit donc progressivement l'intervalle jusqu'à obtenir une précision suffisamment fine de x^*
- Pseudo-code :

Input : $f, [a_0, b_0], \epsilon$

1 $t \leftarrow 0$

2 **Tant que** $|a_t - b_t| > \epsilon$ **faire**

3 $t \leftarrow t + 1$

4 Trouver $[a_t, b_t]$ tel que $x^* \in [a_t, b_t]$
et $|a_t - b_t| < |a_{t-1} - b_{t-1}|$

5 **Fin Tant que**

6 **Output :** $[a_t, b_t]$ où $x^* = \frac{a_t + b_t}{2}$

Méthodes de type "encadrement de minimiseur"

Principe :

- On évalue la fonction objectif à différents points de $[a_0, b_0]$
- On choisit ces points de sorte à approximer le minimiseur appartenant à un intervalle plus petit
- On rétrécit donc progressivement l'intervalle jusqu'à obtenir une précision suffisamment fine de x^*
- Pseudo-code :

```

Input :  $f, [a_0, b_0], \epsilon$ 
1  $t \leftarrow 0$ 
2 Tant que  $|a_t - b_t| > \epsilon$  faire
3    $t \leftarrow t + 1$ 
4   Trouver  $[a_t, b_t]$  tel que  $x^* \in [a_t, b_t]$ 
   et  $|a_t - b_t| < |a_{t-1} - b_{t-1}|$ 
5 Fin Tant que
6 Output :  $[a_t, b_t]$  où  $x^* = \frac{a_t + b_t}{2}$ 

```

- Quelques méthodes particulières : section d'or, Fibonacci

Méthode de la section ou du nombre d'or

- On détermine deux points intermédiaires a_1 et b_1 de telle sorte à ce que la réduction de l'étendue du nouvel intervalle soit symétrique par rapport aux bornes : $a_1 - a_0 = b_0 - b_1 = \rho(b_0 - a_0)$ avec $\rho < \frac{1}{2}$

Méthode de la section ou du nombre d'or

- On détermine deux points intermédiaires a_1 et b_1 de telle sorte à ce que la réduction de l'étendue du nouvel intervalle soit symétrique par rapport aux bornes : $a_1 - a_0 = b_0 - b_1 = \rho(b_0 - a_0)$ avec $\rho < \frac{1}{2}$
- Illustration :



Méthode de la section ou du nombre d'or

- On détermine deux points intermédiaires a_1 et b_1 de telle sorte à ce que la réduction de l'étendue du nouvel intervalle soit symétrique par rapport aux bornes : $a_1 - a_0 = b_0 - b_1 = \rho(b_0 - a_0)$ avec $\rho < \frac{1}{2}$
- Illustration :



- On évalue $f(a_1)$ et $f(b_1)$:
 - ▶ Si $f(a_1) \leq f(b_1)$ alors $x^* \in [a_0, b_1]$
 - ▶ Si $f(a_1) \geq f(b_1)$ alors $x^* \in [a_1, b_0]$

Méthode de la section ou du nombre d'or

- On détermine deux points intermédiaires a_1 et b_1 de telle sorte à ce que la réduction de l'étendue du nouvel intervalle soit symétrique par rapport aux bornes : $a_1 - a_0 = b_0 - b_1 = \rho(b_0 - a_0)$ avec $\rho < \frac{1}{2}$

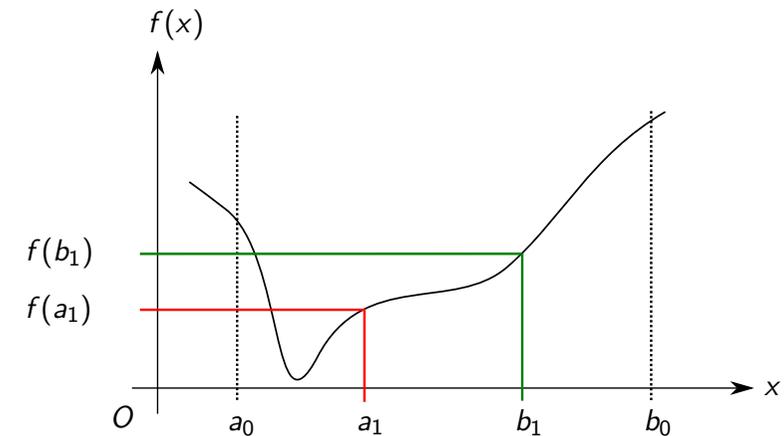
- Illustration :



- On évalue $f(a_1)$ et $f(b_1)$:
 - Si $f(a_1) \leq f(b_1)$ alors $x^* \in [a_0, b_1]$
 - Si $f(a_1) \geq f(b_1)$ alors $x^* \in [a_1, b_0]$
- On peut ensuite réitérer le procédé en gardant la même valeur de ρ

Méthode de la section d'or (suite)

- Illustration :



Méthode de la section d'or (suite)

- Dans l'exemple précédent, on sait que $x^* \in [a_0, b_1]$ mais aussi que $a_1 \in [a_0, b_1]$
- Or on a déjà évalué $f(a_1)$ et on pourrait le faire coïncider avec b_2 . En effet, ainsi, lors de la prochaine itération on n'aurait qu'à évaluer a_2 uniquement

Méthode de la section d'or (suite)

- Dans l'exemple précédent, on sait que $x^* \in [a_0, b_1]$ mais aussi que $a_1 \in [a_0, b_1]$
- Or on a déjà évalué $f(a_1)$ et on pourrait le faire coïncider avec b_2 . En effet, ainsi, lors de la prochaine itération on n'aurait qu'à évaluer a_2 uniquement
- Pour déterminer ρ on suppose, sans perte de généralité, que l'intervalle $[a_0, b_0]$ est de longueur 1 et on pose alors le problème suivant : $\rho(b_1 - a_0) = b_1 - b_2$

Méthode de la section d'or (suite)

- Dans l'exemple précédent, on sait que $x^* \in [a_0, b_1]$ mais aussi que $a_1 \in [a_0, b_1]$
- Or on a déjà évalué $f(a_1)$ et on pourrait le faire coïncider avec b_2 . En effet, ainsi, lors de la prochaine itération on n'aurait qu'à évaluer a_2 uniquement
- Pour déterminer ρ on suppose, sans perte de généralité, que l'intervalle $[a_0, b_0]$ est de longueur 1 et on pose alors le problème suivant : $\rho(b_1 - a_0) = b_1 - b_2$
- Comme $b_1 - a_0 = 1 - \rho$ et $b_1 - b_2 = 1 - 2\rho$ on a $\rho(1 - \rho) = 1 - 2\rho$

Méthode de la section d'or (suite)

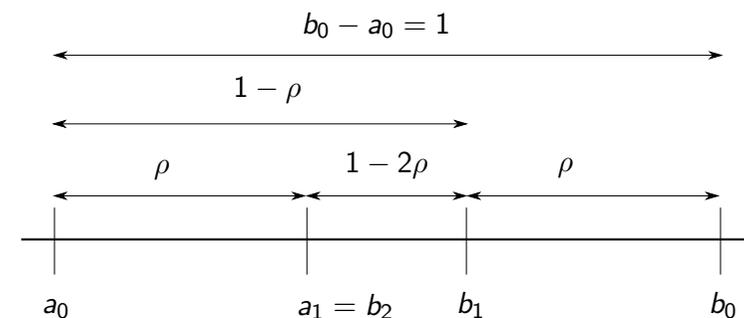
- Dans l'exemple précédent, on sait que $x^* \in [a_0, b_1]$ mais aussi que $a_1 \in [a_0, b_1]$
- Or on a déjà évalué $f(a_1)$ et on pourrait le faire coïncider avec b_2 . En effet, ainsi, lors de la prochaine itération on n'aurait qu'à évaluer a_2 uniquement
- Pour déterminer ρ on suppose, sans perte de généralité, que l'intervalle $[a_0, b_0]$ est de longueur 1 et on pose alors le problème suivant : $\rho(b_1 - a_0) = b_1 - b_2$
- Comme $b_1 - a_0 = 1 - \rho$ et $b_1 - b_2 = 1 - 2\rho$ on a $\rho(1 - \rho) = 1 - 2\rho$
- La solution de l'équation du 2nd degré précédente est : $\rho = \frac{3+\sqrt{5}}{2}$ et comme $\rho < \frac{1}{2}$ on prend la solution $\rho^* = \frac{3-\sqrt{5}}{2} \simeq 0.382$
- Remarque : $2 + \rho^* =$ Nombre d'or (solution de l'équation $x^2 = x + 1$)

Méthode de la section d'or (suite)

- Dans l'exemple précédent, on sait que $x^* \in [a_0, b_1]$ mais aussi que $a_1 \in [a_0, b_1]$
- Or on a déjà évalué $f(a_1)$ et on pourrait le faire coïncider avec b_2 . En effet, ainsi, lors de la prochaine itération on n'aurait qu'à évaluer a_2 uniquement
- Pour déterminer ρ on suppose, sans perte de généralité, que l'intervalle $[a_0, b_0]$ est de longueur 1 et on pose alors le problème suivant : $\rho(b_1 - a_0) = b_1 - b_2$
- Comme $b_1 - a_0 = 1 - \rho$ et $b_1 - b_2 = 1 - 2\rho$ on a $\rho(1 - \rho) = 1 - 2\rho$
- La solution de l'équation du 2nd degré précédente est : $\rho = \frac{3+\sqrt{5}}{2}$ et comme $\rho < \frac{1}{2}$ on prend la solution $\rho^* = \frac{3-\sqrt{5}}{2} \simeq 0.382$

Méthode de la section d'or (suite)

- Illustration :



Algorithme de la méthode de la section d'or

```

Input :  $f, [a_0, b_0], \epsilon$ 
1  $a \leftarrow a_0 + \rho^*(b_0 - a_0); v_a \leftarrow f(a)$ 
2  $b \leftarrow b_0 - \rho^*(b_0 - a_0); v_b \leftarrow f(b)$ 
3 Tant que  $|a_0 - b_0| > \epsilon$  faire
4   Si  $v_a < v_b$  faire
5      $b_0 \leftarrow b$ 
6      $b \leftarrow a; v_b \leftarrow v_a$ 
7      $a \leftarrow a_0 + \rho^*(b_0 - a_0); v_a \leftarrow f(a)$ 
8   Sinon faire
9      $a_0 \leftarrow a$ 
10     $a \leftarrow b; v_a \leftarrow v_b$ 
11     $b \leftarrow b_0 - \rho^*(b_0 - a_0); v_b \leftarrow f(b)$ 
12  Fin Si
13 Fin Tant que
14 Output :  $[a_0, b_0]$  où  $x^* = \frac{a_0 + b_0}{2}$ 

```

Exemple

- Utilisons la méthode de la section d'or pour déterminer un minimiseur local de $f(x) = x^4 - 14x^3 + 60x^2 - 70x$ dans l'intervalle $[0, 2]$ avec une précision $\epsilon = 0.3$

Déroulement de l'algorithme :

Initialisation

$$1 \quad a \leftarrow a_0 + \rho^*(b_0 - a_0) = 0.7639; v_a \leftarrow f(a) = -24.36$$

Exemple

- Utilisons la méthode de la section d'or pour déterminer un minimiseur local de $f(x) = x^4 - 14x^3 + 60x^2 - 70x$ dans l'intervalle $[0, 2]$ avec une précision $\epsilon = 0.3$

Exemple

- Utilisons la méthode de la section d'or pour déterminer un minimiseur local de $f(x) = x^4 - 14x^3 + 60x^2 - 70x$ dans l'intervalle $[0, 2]$ avec une précision $\epsilon = 0.3$

Déroulement de l'algorithme :

Initialisation

$$1 \quad a \leftarrow a_0 + \rho^*(b_0 - a_0) = 0.7639; v_a \leftarrow f(a) = -24.36$$

$$2 \quad b \leftarrow b_0 - \rho^*(b_0 - a_0) = 1.236; v_b \leftarrow f(b) = -18.96$$

Exemple

- Utilisons la méthode de la section d'or pour déterminer un minimiseur local de $f(x) = x^4 - 14x^3 + 60x^2 - 70x$ dans l'intervalle $[0, 2]$ avec une précision $\epsilon = 0.3$

Déroulement de l'algorithme :

Initialisation

- $a \leftarrow a_0 + \rho^*(b_0 - a_0) = 0.7639$; $v_a \leftarrow f(a) = -24.36$
- $b \leftarrow b_0 - \rho^*(b_0 - a_0) = 1.236$; $v_b \leftarrow f(b) = -18.96$

Itération 1

- $|a_0 - b_0| = 2 > 0.3$

Exemple

- Utilisons la méthode de la section d'or pour déterminer un minimiseur local de $f(x) = x^4 - 14x^3 + 60x^2 - 70x$ dans l'intervalle $[0, 2]$ avec une précision $\epsilon = 0.3$

Déroulement de l'algorithme :

Initialisation

- $a \leftarrow a_0 + \rho^*(b_0 - a_0) = 0.7639$; $v_a \leftarrow f(a) = -24.36$
- $b \leftarrow b_0 - \rho^*(b_0 - a_0) = 1.236$; $v_b \leftarrow f(b) = -18.96$

Itération 1

- $|a_0 - b_0| = 2 > 0.3$
- $v_a < v_b$
- $b_0 \leftarrow b = 1.236$

Exemple

- Utilisons la méthode de la section d'or pour déterminer un minimiseur local de $f(x) = x^4 - 14x^3 + 60x^2 - 70x$ dans l'intervalle $[0, 2]$ avec une précision $\epsilon = 0.3$

Déroulement de l'algorithme :

Initialisation

- $a \leftarrow a_0 + \rho^*(b_0 - a_0) = 0.7639$; $v_a \leftarrow f(a) = -24.36$
- $b \leftarrow b_0 - \rho^*(b_0 - a_0) = 1.236$; $v_b \leftarrow f(b) = -18.96$

Itération 1

- $|a_0 - b_0| = 2 > 0.3$
- $v_a < v_b$

Exemple

- Utilisons la méthode de la section d'or pour déterminer un minimiseur local de $f(x) = x^4 - 14x^3 + 60x^2 - 70x$ dans l'intervalle $[0, 2]$ avec une précision $\epsilon = 0.3$

Déroulement de l'algorithme :

Initialisation

- $a \leftarrow a_0 + \rho^*(b_0 - a_0) = 0.7639$; $v_a \leftarrow f(a) = -24.36$
- $b \leftarrow b_0 - \rho^*(b_0 - a_0) = 1.236$; $v_b \leftarrow f(b) = -18.96$

Itération 1

- $|a_0 - b_0| = 2 > 0.3$
- $v_a < v_b$
- $b_0 \leftarrow b = 1.236$
- $b \leftarrow a = 0.7639$; $v_b \leftarrow v_a = -24.36$

Exemple

- Utilisons la méthode de la section d'or pour déterminer un minimiseur local de $f(x) = x^4 - 14x^3 + 60x^2 - 70x$ dans l'intervalle $[0, 2]$ avec une précision $\epsilon = 0.3$

Déroulement de l'algorithme :

Initialisation

- $a \leftarrow a_0 + \rho^*(b_0 - a_0) = 0.7639$; $v_a \leftarrow f(a) = -24.36$
- $b \leftarrow b_0 - \rho^*(b_0 - a_0) = 1.236$; $v_b \leftarrow f(b) = -18.96$

Itération 1

- $|a_0 - b_0| = 2 > 0.3$
- $v_a < v_b$
- $b_0 \leftarrow b = 1.236$
- $b \leftarrow a = 0.7639$; $v_b \leftarrow v_a = -24.36$
- $a \leftarrow a_0 + \rho^*(b_0 - a_0) = 0.4721$; $v_a \leftarrow f(a) = -21.10$

Exemple (suite)

Itération 2

- $|a_0 - b_0| > 0.3$
- $v_b \leq v_a$

Exemple (suite)

Itération 2

- $|a_0 - b_0| > 0.3$

Exemple (suite)

Itération 2

- $|a_0 - b_0| > 0.3$
- $v_b \leq v_a$
- $a_0 \leftarrow a = 0.4721$

Exemple (suite)

Itération 2

- 3 $|a_0 - b_0| > 0.3$
- 8 $v_b \leq v_a$
- 9 $a_0 \leftarrow a = 0.4721$
- 10 $a \leftarrow b = 0.7639$; $v_a \leftarrow v_b = -24.36$

Exemple (suite)

Itération 2

- 3 $|a_0 - b_0| > 0.3$
- 8 $v_b \leq v_a$
- 9 $a_0 \leftarrow a = 0.4721$
- 10 $a \leftarrow b = 0.7639$; $v_a \leftarrow v_b = -24.36$
- 11 $b \leftarrow b_0 - \rho^*(b_0 - a_0) = 0.9443$; $v_b \leftarrow f(b) = -23.59$

Exemple (suite)

Itération 2

- 3 $|a_0 - b_0| > 0.3$
- 8 $v_b \leq v_a$
- 9 $a_0 \leftarrow a = 0.4721$
- 10 $a \leftarrow b = 0.7639$; $v_a \leftarrow v_b = -24.36$
- 11 $b \leftarrow b_0 - \rho^*(b_0 - a_0) = 0.9443$; $v_b \leftarrow f(b) = -23.59$

Itération 3

- 3 $|a_0 - b_0| > 0.3$

Exemple (suite)

Itération 2

- 3 $|a_0 - b_0| > 0.3$
- 8 $v_b \leq v_a$
- 9 $a_0 \leftarrow a = 0.4721$
- 10 $a \leftarrow b = 0.7639$; $v_a \leftarrow v_b = -24.36$
- 11 $b \leftarrow b_0 - \rho^*(b_0 - a_0) = 0.9443$; $v_b \leftarrow f(b) = -23.59$

Itération 3

- 3 $|a_0 - b_0| > 0.3$
- 4 $v_a < v_b$

Exemple (suite)

Itération 2

- 3 $|a_0 - b_0| > 0.3$
- 8 $v_b \leq v_a$
- 9 $a_0 \leftarrow a = 0.4721$
- 10 $a \leftarrow b = 0.7639$; $v_a \leftarrow v_b = -24.36$
- 11 $b \leftarrow b_0 - \rho^*(b_0 - a_0) = 0.9443$; $v_b \leftarrow f(b) = -23.59$

Itération 3

- 3 $|a_0 - b_0| > 0.3$
- 4 $v_a < v_b$
- 5 $b_0 \leftarrow b = 0.9443$

Exemple (suite)

Itération 2

- 3 $|a_0 - b_0| > 0.3$
- 8 $v_b \leq v_a$
- 9 $a_0 \leftarrow a = 0.4721$
- 10 $a \leftarrow b = 0.7639$; $v_a \leftarrow v_b = -24.36$
- 11 $b \leftarrow b_0 - \rho^*(b_0 - a_0) = 0.9443$; $v_b \leftarrow f(b) = -23.59$

Itération 3

- 3 $|a_0 - b_0| > 0.3$
- 4 $v_a < v_b$
- 5 $b_0 \leftarrow b = 0.9443$
- 6 $b \leftarrow a = 0.7639$; $v_b \leftarrow v_a = -24.36$

Exemple (suite)

Itération 2

- 3 $|a_0 - b_0| > 0.3$
- 8 $v_b \leq v_a$
- 9 $a_0 \leftarrow a = 0.4721$
- 10 $a \leftarrow b = 0.7639$; $v_a \leftarrow v_b = -24.36$
- 11 $b \leftarrow b_0 - \rho^*(b_0 - a_0) = 0.9443$; $v_b \leftarrow f(b) = -23.59$

Itération 3

- 3 $|a_0 - b_0| > 0.3$
- 4 $v_a < v_b$
- 5 $b_0 \leftarrow b = 0.9443$
- 6 $b \leftarrow a = 0.7639$; $v_b \leftarrow v_a = -24.36$
- 7 $a \leftarrow a_0 + \rho^*(b_0 - a_0) = 0.6525$; $v_a \leftarrow f(a) = -23.84$

Exemple (suite)

Itération 4

- 3 $|a_0 - b_0| > 0.3$

Exemple (suite)

Itération 4

3 $|a_0 - b_0| > 0.3$

8 $v_b \leq v_a$

Exemple (suite)

Itération 4

3 $|a_0 - b_0| > 0.3$

8 $v_b \leq v_a$

9 $a_0 \leftarrow a = 0.6525$

Exemple (suite)

Itération 4

3 $|a_0 - b_0| > 0.3$

8 $v_b \leq v_a$

9 $a_0 \leftarrow a = 0.6525$

10 $a \leftarrow b = 0.7639$; $v_a \leftarrow v_b = -24.36$

Exemple (suite)

Itération 4

3 $|a_0 - b_0| > 0.3$

8 $v_b \leq v_a$

9 $a_0 \leftarrow a = 0.6525$

10 $a \leftarrow b = 0.7639$; $v_a \leftarrow v_b = -24.36$

11 $b \leftarrow b_0 - \rho^*(b_0 - a_0) = 0.8328$; $v_b \leftarrow f(b) = -24.288$

Exemple (suite)

Itération 4

- 3 $|a_0 - b_0| > 0.3$
- 8 $v_b \leq v_a$
- 9 $a_0 \leftarrow a = 0.6525$
- 10 $a \leftarrow b = 0.7639$; $v_a \leftarrow v_b = -24.36$
- 11 $b \leftarrow b_0 - \rho^*(b_0 - a_0) = 0.8328$; $v_b \leftarrow f(b) = -24.288$

Fin de l'algorithme

- 3 $|a_0 - b_0| < 0.3$

Méthode de Newton

- Dans la méthode de Newton, on suppose que f est de classe \mathcal{C}^2 et que l'on peut donc calculer en tout point $x_k \in \mathbb{R}$: $f(x_k)$, $f'(x_k)$, $f''(x_k)$

Exemple (suite)

Itération 4

- 3 $|a_0 - b_0| > 0.3$
- 8 $v_b \leq v_a$
- 9 $a_0 \leftarrow a = 0.6525$
- 10 $a \leftarrow b = 0.7639$; $v_a \leftarrow v_b = -24.36$
- 11 $b \leftarrow b_0 - \rho^*(b_0 - a_0) = 0.8328$; $v_b \leftarrow f(b) = -24.288$

Fin de l'algorithme

- 3 $|a_0 - b_0| < 0.3$
- 14 $[a_0, b_0] = [0.6525, 0.9443]$

Méthode de Newton

- Dans la méthode de Newton, on suppose que f est de classe \mathcal{C}^2 et que l'on peut donc calculer en tout point $x_k \in \mathbb{R}$: $f(x_k)$, $f'(x_k)$, $f''(x_k)$
- On peut sous ces hypothèses approximer f au voisinage de x_k par une forme quadratique $q : \mathbb{R} \rightarrow \mathbb{R}$ de la forme suivante :

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

Méthode de Newton

- Dans la méthode de Newton, on suppose que f est de classe \mathcal{C}^2 et que l'on peut donc calculer en tout point $x_k \in \mathbb{R}$: $f(x_k), f'(x_k), f''(x_k)$
- On peut sous ces hypothèses approximer f au voisinage de x_k par une forme quadratique $q : \mathbb{R} \rightarrow \mathbb{R}$ de la forme suivante :

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

- Remarque : $q(x_k) = f(x_k), q'(x_k) = f'(x_k), q''(x_k) = f''(x_k)$

Méthode de Newton

- Dans la méthode de Newton, on suppose que f est de classe \mathcal{C}^2 et que l'on peut donc calculer en tout point $x_k \in \mathbb{R}$: $f(x_k), f'(x_k), f''(x_k)$
- On peut sous ces hypothèses approximer f au voisinage de x_k par une forme quadratique $q : \mathbb{R} \rightarrow \mathbb{R}$ de la forme suivante :

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

- Remarque : $q(x_k) = f(x_k), q'(x_k) = f'(x_k), q''(x_k) = f''(x_k)$
- Dans cette approche au lieu de minimiser f , on minimise son approximation q au voisinage d'un point x_k
- Les CNPO pour un minimiseur de q sont alors :

$$q'(x) = 0 \Rightarrow$$

Méthode de Newton

- Dans la méthode de Newton, on suppose que f est de classe \mathcal{C}^2 et que l'on peut donc calculer en tout point $x_k \in \mathbb{R}$: $f(x_k), f'(x_k), f''(x_k)$
- On peut sous ces hypothèses approximer f au voisinage de x_k par une forme quadratique $q : \mathbb{R} \rightarrow \mathbb{R}$ de la forme suivante :

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

- Remarque : $q(x_k) = f(x_k), q'(x_k) = f'(x_k), q''(x_k) = f''(x_k)$
- Dans cette approche au lieu de minimiser f , on minimise son approximation q au voisinage d'un point x_k

Méthode de Newton

- Dans la méthode de Newton, on suppose que f est de classe \mathcal{C}^2 et que l'on peut donc calculer en tout point $x_k \in \mathbb{R}$: $f(x_k), f'(x_k), f''(x_k)$
- On peut sous ces hypothèses approximer f au voisinage de x_k par une forme quadratique $q : \mathbb{R} \rightarrow \mathbb{R}$ de la forme suivante :

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

- Remarque : $q(x_k) = f(x_k), q'(x_k) = f'(x_k), q''(x_k) = f''(x_k)$
- Dans cette approche au lieu de minimiser f , on minimise son approximation q au voisinage d'un point x_k
- Les CNPO pour un minimiseur de q sont alors :

$$q'(x) = 0 \Rightarrow f'(x_k) + f''(x_k)(x - x_k) = 0$$

Méthode de Newton

- Dans la méthode de Newton, on suppose que f est de classe \mathcal{C}^2 et que l'on peut donc calculer en tout point $x_k \in \mathbb{R}$: $f(x_k), f'(x_k), f''(x_k)$
- On peut sous ces hypothèses approximer f au voisinage de x_k par une forme quadratique $q : \mathbb{R} \rightarrow \mathbb{R}$ de la forme suivante :

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

- Remarque : $q(x_k) = f(x_k), q'(x_k) = f'(x_k), q''(x_k) = f''(x_k)$
- Dans cette approche au lieu de minimiser f , on minimise son approximation q au voisinage d'un point x_k
- Les CNPO pour un minimiseur de q sont alors :

$$q'(x) = 0 \Rightarrow f'(x_k) + f''(x_k)(x - x_k) = 0$$

- En posant $x_{k+1} = x$ on obtient une formule itérative permettant de converger vers le minimiseur :

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

Méthode de Newton

- Dans la méthode de Newton, on suppose que f est de classe \mathcal{C}^2 et que l'on peut donc calculer en tout point $x_k \in \mathbb{R}$: $f(x_k), f'(x_k), f''(x_k)$
- On peut sous ces hypothèses approximer f au voisinage de x_k par une forme quadratique $q : \mathbb{R} \rightarrow \mathbb{R}$ de la forme suivante :

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

- Remarque : $q(x_k) = f(x_k), q'(x_k) = f'(x_k), q''(x_k) = f''(x_k)$
- Dans cette approche au lieu de minimiser f , on minimise son approximation q au voisinage d'un point x_k
- Les CNPO pour un minimiseur de q sont alors :

$$q'(x) = 0 \Rightarrow f'(x_k) + f''(x_k)(x - x_k) = 0$$

- En posant $x_{k+1} = x$ on obtient une formule itérative permettant de converger vers le minimiseur :

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

- Méthode basée sur les dérivées premières et secondes de f

Méthode de Newton

- Dans la méthode de Newton, on suppose que f est de classe \mathcal{C}^2 et que l'on peut donc calculer en tout point $x_k \in \mathbb{R}$: $f(x_k), f'(x_k), f''(x_k)$
- On peut sous ces hypothèses approximer f au voisinage de x_k par une forme quadratique $q : \mathbb{R} \rightarrow \mathbb{R}$ de la forme suivante :

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

- Remarque : $q(x_k) = f(x_k), q'(x_k) = f'(x_k), q''(x_k) = f''(x_k)$
- Dans cette approche au lieu de minimiser f , on minimise son approximation q au voisinage d'un point x_k
- Les CNPO pour un minimiseur de q sont alors :

$$q'(x) = 0 \Rightarrow f'(x_k) + f''(x_k)(x - x_k) = 0$$

- En posant $x_{k+1} = x$ on obtient une formule itérative permettant de converger vers le minimiseur :

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

- Méthode basée sur les dérivées premières et secondes de f
- Autres méthodes basées sur les dérivées premières : méthode dichotomique, méthode de la sécante

Algorithme de la méthode de Newton

Input : $f, [a_0, b_0], \epsilon$

- 1 Prendre $x, y \in [a_0, b_0]$ tel que $|x - y| > \epsilon$
- 2 **Tant que** $|x - y| > \epsilon$ **faire**
- 3 $y \leftarrow x$
- 4 $x \leftarrow x - \frac{f'(x)}{f''(x)}$
- 5 **Fin Tant que**
- 6 **Output :** x

Algorithme de la méthode de Newton

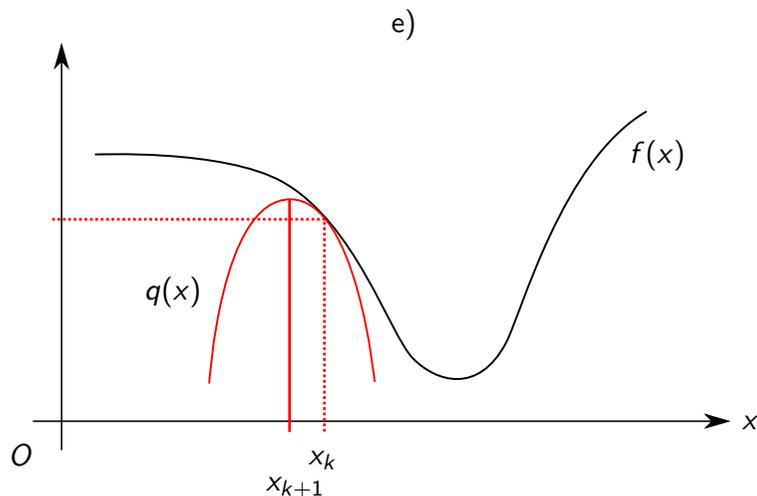
Input : $f, [a_0, b_0], \epsilon$

- 1 Prendre $x, y \in [a_0, b_0]$ tel que $|x - y| > \epsilon$
- 2 **Tant que** $|x - y| > \epsilon$ **faire**
- 3 $y \leftarrow x$
- 4 $x \leftarrow x - \frac{f'(x)}{f''(x)}$
- 5 **Fin Tant que**
- 6 **Output :** x

- La méthode marche bien si $f''(x) \geq 0$ partout sur $[a_0, b_0]$
- La méthode peut ne pas converger s'il existe des $x \in [a_0, b_0]$ tel que $f''(x) < 0$

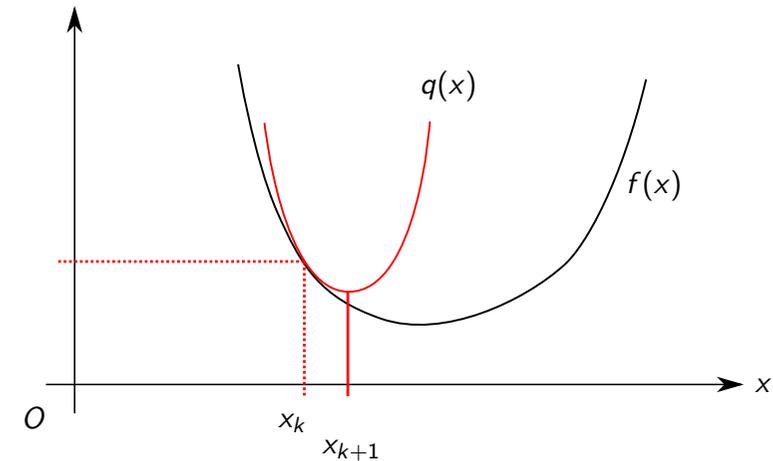
Illustration (suite)

Cas où l'algorithme peut ne pas converger.



Illustration

Cas où l'algorithme converge.



Exemple

- Utilisons la méthode de la section de Newton pour déterminer un minimiseur local de $f(x) = \frac{1}{2}x^2 - \sin(x)$ avec $x_0 = 0.5$ dans l'intervalle $[0, 1]$ avec une précision $\epsilon = 10^{-5}$

Exemple

- Utilisons la méthode de la section de Newton pour déterminer un minimiseur local de $f(x) = \frac{1}{2}x^2 - \sin(x)$ avec $x_0 = 0.5$ dans l'intervalle $[0, 1]$ avec une précision $\epsilon = 10^{-5}$
- Calcul des dérivées :
 - ▶ $f'(x) =$

Exemple

- Utilisons la méthode de la section de Newton pour déterminer un minimiseur local de $f(x) = \frac{1}{2}x^2 - \sin(x)$ avec $x_0 = 0.5$ dans l'intervalle $[0, 1]$ avec une précision $\epsilon = 10^{-5}$
- Calcul des dérivées :
 - ▶ $f'(x) = x - \cos(x)$
 - ▶ $f''(x) = 1 + \sin(x)$

Exemple

- Utilisons la méthode de la section de Newton pour déterminer un minimiseur local de $f(x) = \frac{1}{2}x^2 - \sin(x)$ avec $x_0 = 0.5$ dans l'intervalle $[0, 1]$ avec une précision $\epsilon = 10^{-5}$
- Calcul des dérivées :
 - ▶ $f'(x) = x - \cos(x)$
 - ▶ $f''(x) =$

Exemple

- Utilisons la méthode de la section de Newton pour déterminer un minimiseur local de $f(x) = \frac{1}{2}x^2 - \sin(x)$ avec $x_0 = 0.5$ dans l'intervalle $[0, 1]$ avec une précision $\epsilon = 10^{-5}$
- Calcul des dérivées :
 - ▶ $f'(x) = x - \cos(x)$
 - ▶ $f''(x) = 1 + \sin(x)$
- Déroulement de l'algorithme :

Initialisation

1 $x \leftarrow 0.5; y \leftarrow 0.6$

Exemple

- Utilisons la méthode de la section de Newton pour déterminer un minimiseur local de $f(x) = \frac{1}{2}x^2 - \sin(x)$ avec $x_0 = 0.5$ dans l'intervalle $[0, 1]$ avec une précision $\epsilon = 10^{-5}$
- Calcul des dérivées :
 - ▶ $f'(x) = x - \cos(x)$
 - ▶ $f''(x) = 1 + \sin(x)$
- Déroulement de l'algorithme :

Initialisation

$$1 \quad x \leftarrow 0.5; y \leftarrow 0.6$$

Itération 1

$$2 \quad |x - y| > 10^{-5}$$

Exemple

- Utilisons la méthode de la section de Newton pour déterminer un minimiseur local de $f(x) = \frac{1}{2}x^2 - \sin(x)$ avec $x_0 = 0.5$ dans l'intervalle $[0, 1]$ avec une précision $\epsilon = 10^{-5}$
- Calcul des dérivées :
 - ▶ $f'(x) = x - \cos(x)$
 - ▶ $f''(x) = 1 + \sin(x)$
- Déroulement de l'algorithme :

Initialisation

$$1 \quad x \leftarrow 0.5; y \leftarrow 0.6$$

Itération 1

$$2 \quad |x - y| > 10^{-5}$$

$$3 \quad y \leftarrow x = 0.5$$

Exemple

- Utilisons la méthode de la section de Newton pour déterminer un minimiseur local de $f(x) = \frac{1}{2}x^2 - \sin(x)$ avec $x_0 = 0.5$ dans l'intervalle $[0, 1]$ avec une précision $\epsilon = 10^{-5}$
- Calcul des dérivées :
 - ▶ $f'(x) = x - \cos(x)$
 - ▶ $f''(x) = 1 + \sin(x)$
- Déroulement de l'algorithme :

Initialisation

$$1 \quad x \leftarrow 0.5; y \leftarrow 0.6$$

Itération 1

$$2 \quad |x - y| > 10^{-5}$$

$$3 \quad y \leftarrow x = 0.5$$

$$4 \quad x \leftarrow x - \frac{f'(x)}{f''(x)} = 0.5 - \frac{0.5 - \cos(0.5)}{1 + \sin(0.5)} = 0.7552$$

Exemple (suite)

Itération 2

$$2 \quad |x - y| > 10^{-5}$$

Exemple (suite)

Itération 2

- 2 $|x - y| > 10^{-5}$
- 3 $y \leftarrow x = 0.7552$

Exemple (suite)

Itération 2

- 2 $|x - y| > 10^{-5}$
- 3 $y \leftarrow x = 0.7552$
- 4 $x \leftarrow x - \frac{f'(x)}{f''(x)} = 0.7552 - \frac{0.7552 - \cos(0.7552)}{1 + \sin(0.7552)} = 0.7391$

Exemple (suite)

Itération 2

- 2 $|x - y| > 10^{-5}$
- 3 $y \leftarrow x = 0.7552$
- 4 $x \leftarrow x - \frac{f'(x)}{f''(x)} = 0.7552 - \frac{0.7552 - \cos(0.7552)}{1 + \sin(0.7552)} = 0.7391$

Itération 3

- 2 $|x - y| > 10^{-5}$

Exemple (suite)

Itération 2

- 2 $|x - y| > 10^{-5}$
- 3 $y \leftarrow x = 0.7552$
- 4 $x \leftarrow x - \frac{f'(x)}{f''(x)} = 0.7552 - \frac{0.7552 - \cos(0.7552)}{1 + \sin(0.7552)} = 0.7391$

Itération 3

- 2 $|x - y| > 10^{-5}$
- 3 $y \leftarrow x = 0.7391$

Exemple (suite)

Itération 2

2 $|x - y| > 10^{-5}$

3 $y \leftarrow x = 0.7552$

4 $x \leftarrow x - \frac{f'(x)}{f''(x)} = 0.7552 - \frac{0.7552 - \cos(0.7552)}{1 + \sin(0.7552)} = 0.7391$

Itération 3

2 $|x - y| > 10^{-5}$

3 $y \leftarrow x = 0.7391$

4 $x \leftarrow x - \frac{f'(x)}{f''(x)} = 0.7391 - \frac{0.7391 - \cos(0.7391)}{1 + \sin(0.7391)} = 0.7390$

Exemple (suite)

Itération 2

2 $|x - y| > 10^{-5}$

3 $y \leftarrow x = 0.7552$

4 $x \leftarrow x - \frac{f'(x)}{f''(x)} = 0.7552 - \frac{0.7552 - \cos(0.7552)}{1 + \sin(0.7552)} = 0.7391$

Itération 3

2 $|x - y| > 10^{-5}$

3 $y \leftarrow x = 0.7391$

4 $x \leftarrow x - \frac{f'(x)}{f''(x)} = 0.7391 - \frac{0.7391 - \cos(0.7391)}{1 + \sin(0.7391)} = 0.7390$

Itération 4

2 $|x - y| > 10^{-5}$

3 $y \leftarrow x = 0.7390$

Exemple (suite)

Itération 2

2 $|x - y| > 10^{-5}$

3 $y \leftarrow x = 0.7552$

4 $x \leftarrow x - \frac{f'(x)}{f''(x)} = 0.7552 - \frac{0.7552 - \cos(0.7552)}{1 + \sin(0.7552)} = 0.7391$

Itération 3

2 $|x - y| > 10^{-5}$

3 $y \leftarrow x = 0.7391$

4 $x \leftarrow x - \frac{f'(x)}{f''(x)} = 0.7391 - \frac{0.7391 - \cos(0.7391)}{1 + \sin(0.7391)} = 0.7390$

Itération 4

2 $|x - y| > 10^{-5}$

Exemple (suite)

Itération 2

2 $|x - y| > 10^{-5}$

3 $y \leftarrow x = 0.7552$

4 $x \leftarrow x - \frac{f'(x)}{f''(x)} = 0.7552 - \frac{0.7552 - \cos(0.7552)}{1 + \sin(0.7552)} = 0.7391$

Itération 3

2 $|x - y| > 10^{-5}$

3 $y \leftarrow x = 0.7391$

4 $x \leftarrow x - \frac{f'(x)}{f''(x)} = 0.7391 - \frac{0.7391 - \cos(0.7391)}{1 + \sin(0.7391)} = 0.7390$

Itération 4

2 $|x - y| > 10^{-5}$

3 $y \leftarrow x = 0.7390$

4 $x \leftarrow x - \frac{f'(x)}{f''(x)} = 0.7390 - \frac{0.7390 - \cos(0.7390)}{1 + \sin(0.7390)} = 0.7390$

Exemple (suite)

Fin de l'algorithme

2 $|x - y| < 10^{-5}$

Exemple (suite)

Fin de l'algorithme

2 $|x - y| < 10^{-5}$

6 $x = 0.7390$

Exemple (suite)

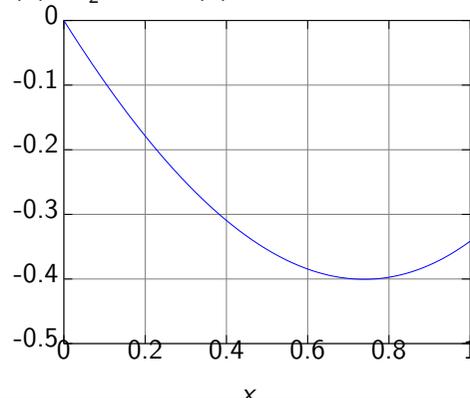
Fin de l'algorithme

2 $|x - y| < 10^{-5}$

6 $x = 0.7390$

Illustration :

$$f(x) = \frac{1}{2}x^2 - \sin(x)$$



Méthode de la sécante

- La méthode de Newton utilise la dérivée seconde mais si celle-ci n'est pas facilement calculable, on peut tenter de l'approximer en utilisant la dérivée première :

$$f''(x_k) \simeq \frac{f'(x_k) - f'(x_{k-1})}{x_k - x_{k-1}}$$

Méthode de la sécante

- La méthode de Newton utilise la dérivée seconde mais si celle-ci n'est pas facilement calculable, on peut tenter de l'approximer en utilisant la dérivée première :

$$f''(x_k) \simeq \frac{f'(x_k) - f'(x_{k-1})}{x_k - x_{k-1}}$$

- On a donc l'algorithme suivant :

Input : $f, [a_0, b_0], \epsilon$

- 1 Prendre $x_0, x_1 \in [a_0, b_0]$ tel que $|x_0 - x_1| > \epsilon$
- 2 $k \leftarrow 1$
- 3 **Tant que** $|x_k - x_{k-1}| > \epsilon$ **faire**
- 4 $x_{k+1} \leftarrow x_k - f'(x_k) \frac{x_k - x_{k-1}}{f'(x_k) - f'(x_{k-1})}$
- 5 $k \leftarrow k + 1$
- 6 **Fin Tant que**
- 7 **Output :** x_k

Optimisation multidimensionnelle

- Nous nous intéressons dans cette sous-section aux fonctions objectif multidimensionnelles (ou multivariables) $f : \mathbb{R}^n \rightarrow \mathbb{R}$

Rappel du Sommaire

- 1 Rappels de concepts mathématiques
- 2 Bases théoriques de l'optimisation
- 3 Optimisation sans contrainte
 - Optimisation unidimensionnelle
 - Optimisation multidimensionnelle
 - Algorithmes utilisant la dérivée première uniquement
 - Algorithmes utilisant la dérivée seconde également
- 4 Optimisation linéaire avec contraintes
- 5 Optimisation non linéaire avec contraintes
- 6 **Optimisation convexe**

Optimisation multidimensionnelle

- Nous nous intéressons dans cette sous-section aux fonctions objectif multidimensionnelles (ou multivariables) $f : \mathbb{R}^n \rightarrow \mathbb{R}$
- Les algorithmes que nous allons étudier reposent tous sur l'idée générale suivante :
 - ▶ On détermine une suite de points dans \mathbb{S} qui converge vers un minimiseur local
 - ▶ A chaque étape on détermine une nouvelle direction et un pas permettant de trouver un nouveau point
 - ▶ La direction est donnée par le gradient
 - ▶ Il existe par contre plusieurs façon de déterminer le pas

Optimisation multidimensionnelle

- Nous nous intéressons dans cette sous-section aux fonctions objectif multidimensionnelles (ou multivariées) $f : \mathbb{R}^n \rightarrow \mathbb{R}$
- Les algorithmes que nous allons étudier reposent tous sur l'idée générale suivante :
 - ▶ On détermine une suite de points dans \mathbb{S} qui converge vers un minimiseur local
 - ▶ A chaque étape on détermine une nouvelle direction et un pas permettant de trouver un nouveau point
 - ▶ La direction est donnée par le gradient
 - ▶ Il existe par contre plusieurs façons de déterminer le pas
- Ces algorithmes sont appelés **algorithmes de descente**

Algorithmes de descente de gradient

- Rappelons que le gradient de f en \mathbf{x}_0 , dénoté $\nabla f(\mathbf{x}_0)$ (s'il n'est pas le vecteur nul) est orthogonal à un vecteur tangent à toute courbe correspondant à une ligne de niveau passant par \mathbf{x}_0

Algorithmes de descente de gradient

- Rappelons que le gradient de f en \mathbf{x}_0 , dénoté $\nabla f(\mathbf{x}_0)$ (s'il n'est pas le vecteur nul) est orthogonal à un vecteur tangent à toute courbe correspondant à une ligne de niveau passant par \mathbf{x}_0
- La direction du taux d'accroissement maximum de f en un point \mathbf{x}_0 est donc orthogonal à la ligne de niveau de f passant par \mathbf{x}_0

Algorithmes de descente de gradient

- Rappelons que le gradient de f en \mathbf{x}_0 , dénoté $\nabla f(\mathbf{x}_0)$ (s'il n'est pas le vecteur nul) est orthogonal à un vecteur tangent à toute courbe correspondant à une ligne de niveau passant par \mathbf{x}_0
- La direction du taux d'accroissement maximum de f en un point \mathbf{x}_0 est donc orthogonal à la ligne de niveau de f passant par \mathbf{x}_0
- Dit autrement, si à partir de \mathbf{x}_0 on devait faire un petit déplacement, alors la fonction f varie davantage en suivant la direction donnée par le gradient de f en \mathbf{x}_0 que toute autre direction

Algorithmes de descente de gradient (suite)

Plus formellement :

- Soit \mathbf{d} une direction telle que $\|\mathbf{d}\| = 1$. Rappelons alors que $\langle \nabla f(\mathbf{x}), \mathbf{d} \rangle$ est le taux d'accroissement de f dans la direction \mathbf{d} au point \mathbf{x} . En utilisant l'inégalité de Cauchy-Schwartz, nous obtenons :

$$\langle \nabla f(\mathbf{x}), \mathbf{d} \rangle \leq \|\nabla f(\mathbf{x})\|$$

Algorithmes de descente de gradient (suite)

Plus formellement :

- Soit \mathbf{d} une direction telle que $\|\mathbf{d}\| = 1$. Rappelons alors que $\langle \nabla f(\mathbf{x}), \mathbf{d} \rangle$ est le taux d'accroissement de f dans la direction \mathbf{d} au point \mathbf{x} . En utilisant l'inégalité de Cauchy-Schwartz, nous obtenons :

$$\langle \nabla f(\mathbf{x}), \mathbf{d} \rangle \leq \|\nabla f(\mathbf{x})\|$$

- Si par contre nous prenons $\mathbf{d} = \frac{\nabla f(\mathbf{x})}{\|\nabla f(\mathbf{x})\|}$ alors nous obtenons :

$$\langle \nabla f(\mathbf{x}), \frac{\nabla f(\mathbf{x})}{\|\nabla f(\mathbf{x})\|} \rangle = \|\nabla f(\mathbf{x})\|$$

- Nous venons de montrer que la direction suivie par $\nabla f(\mathbf{x})$ est celle du taux d'accroissement maximum de f en \mathbf{x}

Algorithmes de descente de gradient (suite)

Plus formellement :

- Soit \mathbf{d} une direction telle que $\|\mathbf{d}\| = 1$. Rappelons alors que $\langle \nabla f(\mathbf{x}), \mathbf{d} \rangle$ est le taux d'accroissement de f dans la direction \mathbf{d} au point \mathbf{x} . En utilisant l'inégalité de Cauchy-Schwartz, nous obtenons :

$$\langle \nabla f(\mathbf{x}), \mathbf{d} \rangle \leq \|\nabla f(\mathbf{x})\|$$

- Si par contre nous prenons $\mathbf{d} = \frac{\nabla f(\mathbf{x})}{\|\nabla f(\mathbf{x})\|}$ alors nous obtenons :

$$\langle \nabla f(\mathbf{x}), \frac{\nabla f(\mathbf{x})}{\|\nabla f(\mathbf{x})\|} \rangle = \|\nabla f(\mathbf{x})\|$$

- On en déduit que la direction $-\nabla f(\mathbf{x})$ est celle conduisant au taux d'accroissement **négatif** maximum (diminution) de f en \mathbf{x}

Algorithmes de descente de gradient (suite)

Algorithmes de descente de gradient (suite)

- On en déduit que la direction $-\nabla f(\mathbf{x})$ est celle conduisant au taux d'accroissement **négatif** maximum (diminution) de f en \mathbf{x}
- Dit autrement, la direction $-\nabla f(\mathbf{x})$ est la direction qu'il faut suivre pour trouver un minimiseur de f en partant de \mathbf{x} : on parle alors d'**algorithmes de descente de gradient**

Algorithmes de descente de gradient (suite)

- On en déduit que la direction $-\nabla f(\mathbf{x})$ est celle conduisant au taux d'accroissement **négatif** maximum (diminution) de f en \mathbf{x}
- Dit autrement, la direction $-\nabla f(\mathbf{x})$ est la direction qu'il faut suivre pour trouver un minimiseur de f en partant de \mathbf{x} : on parle alors d'**algorithmes de descente de gradient**
- Plus formellement :
 - ▶ De $\mathbf{x}^{(0)}$, on considère le point $\mathbf{x}^{(0)} - \alpha \nabla f(\mathbf{x}^{(0)})$ et d'après le développement de Taylor on a :

$$f(\mathbf{x}^{(0)} - \alpha \nabla f(\mathbf{x}^{(0)})) = f(\mathbf{x}^{(0)}) - \alpha \|\nabla f(\mathbf{x}^{(0)})\|^2 + o(\alpha)$$
 - ▶ Ainsi, si $\nabla f(\mathbf{x}^{(0)}) \neq \mathbf{0}$, alors pour $\alpha > 0$ suffisamment petit on a :

$$f(\mathbf{x}^{(0)} - \alpha \nabla f(\mathbf{x}^{(0)})) < f(\mathbf{x}^{(0)})$$

Algorithmes de descente de gradient (suite)

- On en déduit que la direction $-\nabla f(\mathbf{x})$ est celle conduisant au taux d'accroissement **négatif** maximum (diminution) de f en \mathbf{x}
- Dit autrement, la direction $-\nabla f(\mathbf{x})$ est la direction qu'il faut suivre pour trouver un minimiseur de f en partant de \mathbf{x} : on parle alors d'**algorithmes de descente de gradient**
- Plus formellement :
 - ▶ De $\mathbf{x}^{(0)}$, on considère le point $\mathbf{x}^{(0)} - \alpha \nabla f(\mathbf{x}^{(0)})$ et d'après le développement de Taylor on a :

$$f(\mathbf{x}^{(0)} - \alpha \nabla f(\mathbf{x}^{(0)})) = f(\mathbf{x}^{(0)}) - \alpha \|\nabla f(\mathbf{x}^{(0)})\|^2 + o(\alpha)$$

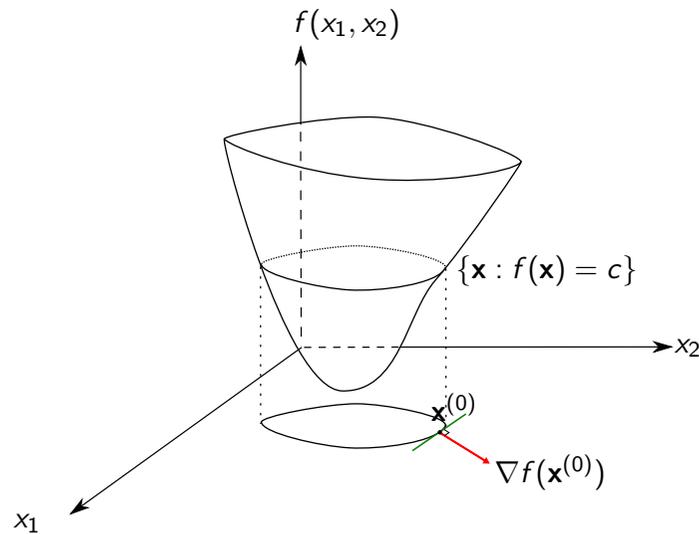
Algorithmes de descente de gradient (suite)

- On en déduit que la direction $-\nabla f(\mathbf{x})$ est celle conduisant au taux d'accroissement **négatif** maximum (diminution) de f en \mathbf{x}
- Dit autrement, la direction $-\nabla f(\mathbf{x})$ est la direction qu'il faut suivre pour trouver un minimiseur de f en partant de \mathbf{x} : on parle alors d'**algorithmes de descente de gradient**
- Plus formellement :
 - ▶ De $\mathbf{x}^{(0)}$, on considère le point $\mathbf{x}^{(0)} - \alpha \nabla f(\mathbf{x}^{(0)})$ et d'après le développement de Taylor on a :

$$f(\mathbf{x}^{(0)} - \alpha \nabla f(\mathbf{x}^{(0)})) = f(\mathbf{x}^{(0)}) - \alpha \|\nabla f(\mathbf{x}^{(0)})\|^2 + o(\alpha)$$
 - ▶ Ainsi, si $\nabla f(\mathbf{x}^{(0)}) \neq \mathbf{0}$, alors pour $\alpha > 0$ suffisamment petit on a :

$$f(\mathbf{x}^{(0)} - \alpha \nabla f(\mathbf{x}^{(0)})) < f(\mathbf{x}^{(0)})$$
 - ▶ Donc, $\mathbf{x}^{(1)} = \mathbf{x}^{(0)} - \alpha \nabla f(\mathbf{x}^{(0)})$ est un point plus proche d'un minimiseur de f que $\mathbf{x}^{(0)}$

Illustration



Pseudo-code de l'algorithme de descente de gradient

Input : $f \in \mathcal{C}^1, \mathbf{x}^0$

- 1 $k \leftarrow 0$
- 2 **Tant que** condition d'arrêt non satisfaite **faire**
- 3 Trouver un pas α_k tel que $f(\mathbf{x}^{(k)} - \alpha_k \nabla f(\mathbf{x}^{(k)})) < f(\mathbf{x}^{(k)})$
- 4 $\mathbf{x}^{(k+1)} \leftarrow \mathbf{x}^{(k)} - \alpha_k \nabla f(\mathbf{x}^{(k)})$
- 5 $k \leftarrow k + 1$
- 6 **Fin Tant que**
- 7 **Output :** $\mathbf{x}^{(k)}$

Algorithmes de descente de gradient (suite)

- Il existe plusieurs types de conditions d'arrêt :
 - ▶ Les conditions nécessaires CNPO, CNSO et CSSO
 - ▶ Précision ϵ atteinte (voir plus loin)

Algorithmes de descente de gradient (suite)

- Il existe plusieurs types de conditions d'arrêt :
 - ▶ Les conditions nécessaires CNPO, CNSO et CSSO
 - ▶ Précision ϵ atteinte (voir plus loin)
- Comment déterminer le pas α_k à chaque itération k ? On distingue :
 - ▶ les méthodes utilisant la dérivée première uniquement
 - ▶ les méthodes utilisant la dérivée seconde également
 - ▶ les méthodes de recherche linéaire exacte et inexacte

Méthode du pas optimal

- On cherche α_k tel que :

$$\alpha_k = \arg \min_{\alpha \in \mathbb{R}^+} \underbrace{f(\mathbf{x}^{(k)} - \alpha \nabla f(\mathbf{x}^{(k)}))}_{\phi(\alpha)}$$

- Dit autrement, α_k est le pas qui permet d'avoir la plus forte diminution de f à chaque étape
- En pratique on utilise une méthode de recherche linéaire ou d'**optimisation unidimensionnelle** pour déterminer α_k

Méthode du pas optimal

- On cherche α_k tel que :

$$\alpha_k = \arg \min_{\alpha \in \mathbb{R}^+} \underbrace{f(\mathbf{x}^{(k)} - \alpha \nabla f(\mathbf{x}^{(k)}))}_{\phi(\alpha)}$$

- Dit autrement, α_k est le pas qui permet d'avoir la plus forte diminution de f à chaque étape

- Dit autrement, α_k est le pas qui permet d'avoir la plus forte diminution de f à chaque étape
- En pratique on utilise une méthode de recherche linéaire ou d'**optimisation unidimensionnelle** pour déterminer α_k

- On a alors la propriété suivante :

Propriété.

Si $\{\mathbf{x}^{(k)}\}$ est une suite de vecteurs issus de la méthode du pas optimal pour une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ alors pour chaque k , le vecteur $\mathbf{x}^{(k+1)} - \mathbf{x}^k$ est orthogonal au vecteur $\mathbf{x}^{(k+2)} - \mathbf{x}^{(k+1)}$

Méthode du pas optimal

- On cherche α_k tel que :

$$\alpha_k = \arg \min_{\alpha \in \mathbb{R}^+} \underbrace{f(\mathbf{x}^{(k)} - \alpha \nabla f(\mathbf{x}^{(k)}))}_{\phi(\alpha)}$$

- Dit autrement, α_k est le pas qui permet d'avoir la plus forte diminution de f à chaque étape
- En pratique on utilise une méthode de recherche linéaire ou d'**optimisation unidimensionnelle** pour déterminer α_k

Méthode du pas optimal

- On cherche α_k tel que :

$$\alpha_k = \arg \min_{\alpha \in \mathbb{R}^+} \underbrace{f(\mathbf{x}^{(k)} - \alpha \nabla f(\mathbf{x}^{(k)}))}_{\phi(\alpha)}$$

- Dit autrement, α_k est le pas qui permet d'avoir la plus forte diminution de f à chaque étape
- En pratique on utilise une méthode de recherche linéaire ou d'**optimisation unidimensionnelle** pour déterminer α_k

- On a alors la propriété suivante :

Propriété.

Si $\{\mathbf{x}^{(k)}\}$ est une suite de vecteurs issus de la méthode du pas optimal pour une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ alors pour chaque k , le vecteur $\mathbf{x}^{(k+1)} - \mathbf{x}^k$ est orthogonal au vecteur $\mathbf{x}^{(k+2)} - \mathbf{x}^{(k+1)}$

Preuve de la propriété précédente

Démonstration.

- Comme $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \alpha_k \nabla f(\mathbf{x}^{(k)})$ on a :

$$\langle \mathbf{x}^{(k+2)} - \mathbf{x}^{(k+1)}, \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)} \rangle = \alpha_{k+1} \alpha_k \langle \nabla f(\mathbf{x}^{(k+1)}), \nabla f(\mathbf{x}^{(k)}) \rangle$$

Preuve de la propriété précédente

Démonstration.

- Comme $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \alpha_k \nabla f(\mathbf{x}^{(k)})$ on a :

$$\langle \mathbf{x}^{(k+2)} - \mathbf{x}^{(k+1)}, \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)} \rangle = \alpha_{k+1} \alpha_k \langle \nabla f(\mathbf{x}^{(k+1)}), \nabla f(\mathbf{x}^{(k)}) \rangle$$
- Il suffit alors de montrer que $\langle \nabla f(\mathbf{x}^{(k+1)}), \nabla f(\mathbf{x}^{(k)}) \rangle = 0$

Preuve de la propriété précédente

Démonstration.

- Comme $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \alpha_k \nabla f(\mathbf{x}^{(k)})$ on a :

$$\langle \mathbf{x}^{(k+2)} - \mathbf{x}^{(k+1)}, \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)} \rangle = \alpha_{k+1} \alpha_k \langle \nabla f(\mathbf{x}^{(k+1)}), \nabla f(\mathbf{x}^{(k)}) \rangle$$
- Il suffit alors de montrer que $\langle \nabla f(\mathbf{x}^{(k+1)}), \nabla f(\mathbf{x}^{(k)}) \rangle = 0$
- Posons $\phi_k(\alpha) = f(\mathbf{x}^{(k)} - \alpha \nabla f(\mathbf{x}^{(k)}))$

Preuve de la propriété précédente

Démonstration.

- Comme $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \alpha_k \nabla f(\mathbf{x}^{(k)})$ on a :

$$\langle \mathbf{x}^{(k+2)} - \mathbf{x}^{(k+1)}, \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)} \rangle = \alpha_{k+1} \alpha_k \langle \nabla f(\mathbf{x}^{(k+1)}), \nabla f(\mathbf{x}^{(k)}) \rangle$$
- Il suffit alors de montrer que $\langle \nabla f(\mathbf{x}^{(k+1)}), \nabla f(\mathbf{x}^{(k)}) \rangle = 0$
- Posons $\phi_k(\alpha) = f(\mathbf{x}^{(k)} - \alpha \nabla f(\mathbf{x}^{(k)}))$
- Par définition $\alpha_k = \arg \min_{\alpha \geq 0} \phi_k(\alpha)$

Preuve de la propriété précédente

Démonstration.

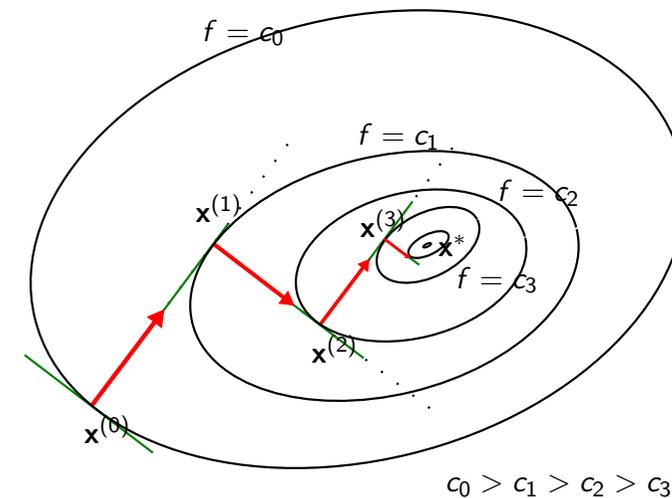
- Comme $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \alpha_k \nabla f(\mathbf{x}^{(k)})$ on a :

$$\langle \mathbf{x}^{(k+2)} - \mathbf{x}^{(k+1)}, \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)} \rangle = \alpha_{k+1} \alpha_k \langle \nabla f(\mathbf{x}^{(k+1)}), \nabla f(\mathbf{x}^{(k)}) \rangle$$
- Il suffit alors de montrer que $\langle \nabla f(\mathbf{x}^{(k+1)}), \nabla f(\mathbf{x}^{(k)}) \rangle = 0$
- Posons $\phi_k(\alpha) = f(\mathbf{x}^{(k)} - \alpha \nabla f(\mathbf{x}^{(k)}))$
- Par définition $\alpha_k = \arg \min_{\alpha \geq 0} \phi_k(\alpha)$
- En utilisant la CNPO et la règle de dérivation on obtient :

$$\begin{aligned} 0 &= \phi_k'(\alpha_k) \\ &= \frac{d\phi_k}{d\alpha}(\alpha_k) \\ &= \nabla f(\mathbf{x}^{(k)} - \alpha_k \nabla f(\mathbf{x}^{(k)}))^t (-\nabla f(\mathbf{x}^{(k)})) \\ &= -\langle \nabla f(\mathbf{x}^{(k+1)}), \nabla f(\mathbf{x}^{(k)}) \rangle \end{aligned}$$

□

Illustration



Méthode du pas optimal (suite)

- A chaque étape, la valeur de la fonction objectif diminue :

Propriété.

Si $\{\mathbf{x}^{(k)}\}$ est une suite de vecteurs issus de la méthode du pas optimal pour une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ et si $\nabla f(\mathbf{x}^{(k)}) \neq \mathbf{0}$ alors $f(\mathbf{x}^{(k+1)}) < f(\mathbf{x}^{(k)})$

Méthode du pas optimal (suite)

- A chaque étape, la valeur de la fonction objectif diminue :

Propriété.

Si $\{\mathbf{x}^{(k)}\}$ est une suite de vecteurs issus de la méthode du pas optimal pour une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ et si $\nabla f(\mathbf{x}^{(k)}) \neq \mathbf{0}$ alors $f(\mathbf{x}^{(k+1)}) < f(\mathbf{x}^{(k)})$

- Si pour k , $\nabla f(\mathbf{x}^{(k)}) = \mathbf{0}$ alors le point $\mathbf{x}^{(k)}$ satisfait la CNPO et on obtient $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)}$

Méthode du pas optimal (suite)

- A chaque étape, la valeur de la fonction objectif diminue :

Propriété.

Si $\{\mathbf{x}^{(k)}\}$ est une suite de vecteurs issus de la méthode du pas optimal pour une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ et si $\nabla f(\mathbf{x}^{(k)}) \neq \mathbf{0}$ alors $f(\mathbf{x}^{(k+1)}) < f(\mathbf{x}^{(k)})$

- Si pour k , $\nabla f(\mathbf{x}^{(k)}) = \mathbf{0}$ alors le point $\mathbf{x}^{(k)}$ satisfait la CNPO et on obtient $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)}$
- Comme précisé précédemment, cette condition peut-être une condition d'arrêt de l'algorithme

Conditions d'arrêt

- En pratique on utilisera des conditions d'arrêt relatives à un paramètre $\epsilon > 0$ spécifié en entrée de l'algorithme

Méthode du pas optimal (suite)

- A chaque étape, la valeur de la fonction objectif diminue :

Propriété.

Si $\{\mathbf{x}^{(k)}\}$ est une suite de vecteurs issus de la méthode du pas optimal pour une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ et si $\nabla f(\mathbf{x}^{(k)}) \neq \mathbf{0}$ alors $f(\mathbf{x}^{(k+1)}) < f(\mathbf{x}^{(k)})$

- Si pour k , $\nabla f(\mathbf{x}^{(k)}) = \mathbf{0}$ alors le point $\mathbf{x}^{(k)}$ satisfait la CNPO et on obtient $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)}$
- Comme précisé précédemment, cette condition peut-être une condition d'arrêt de l'algorithme
- Toutefois, en pratique, la valeur numérique du gradient est rarement égale à $\mathbf{0}$

Conditions d'arrêt

- En pratique on utilisera des conditions d'arrêt relatives à un paramètre $\epsilon > 0$ spécifié en entrée de l'algorithme
- Conditions basées sur des mesures "absolues" :
 - ▶ $\|\nabla f(\mathbf{x}^{(k)})\| < \epsilon$
 - ▶ $|f(\mathbf{x}^{(k+1)}) - f(\mathbf{x}^{(k)})| < \epsilon$
 - ▶ $\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\| < \epsilon$

Conditions d'arrêt

- En pratique on utilisera des conditions d'arrêt relatives à un paramètre $\epsilon > 0$ spécifié en entrée de l'algorithme
- Conditions basées sur des mesures "absolues" :
 - ▶ $\|\nabla f(\mathbf{x}^{(k)})\| < \epsilon$
 - ▶ $|f(\mathbf{x}^{(k+1)}) - f(\mathbf{x}^{(k)})| < \epsilon$
 - ▶ $\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\| < \epsilon$
- En raison des unités de mesure, il est préférable d'utiliser des conditions basées sur des mesures "relatives" :
 - ▶ $\frac{|f(\mathbf{x}^{(k+1)}) - f(\mathbf{x}^{(k)})|}{\max\{1, |f(\mathbf{x}^{(k)})|\}} < \epsilon$
 - ▶ $\frac{\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|}{\max\{1, \|\mathbf{x}^{(k)}\|\}} < \epsilon$

Pseudo-code de l'algorithme à pas optimal

Input : $f \in \mathcal{C}^1, \mathbf{x}^0$

- 1 $k \leftarrow 0$
- 2 **Tant que** condition d'arrêt non satisfaite **faire**
- 3 $\alpha_k \leftarrow \arg \min_{\alpha \geq 0} \underbrace{f(\mathbf{x}^{(k)} - \alpha \nabla f(\mathbf{x}^{(k)}))}_{\phi(\alpha)}$
- 4 $\mathbf{x}^{(k+1)} \leftarrow \mathbf{x}^{(k)} - \alpha_k \nabla f(\mathbf{x}^{(k)})$
- 5 $k \leftarrow k + 1$
- 6 **Fin Tant que**
- 7 **Output :** $\mathbf{x}^{(k)}$

Conditions d'arrêt

- En pratique on utilisera des conditions d'arrêt relatives à un paramètre $\epsilon > 0$ spécifié en entrée de l'algorithme
- Conditions basées sur des mesures "absolues" :
 - ▶ $\|\nabla f(\mathbf{x}^{(k)})\| < \epsilon$
 - ▶ $|f(\mathbf{x}^{(k+1)}) - f(\mathbf{x}^{(k)})| < \epsilon$
 - ▶ $\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\| < \epsilon$
- En raison des unités de mesure, il est préférable d'utiliser des conditions basées sur des mesures "relatives" :
 - ▶ $\frac{|f(\mathbf{x}^{(k+1)}) - f(\mathbf{x}^{(k)})|}{\max\{1, |f(\mathbf{x}^{(k)})|\}} < \epsilon$
 - ▶ $\frac{\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|}{\max\{1, \|\mathbf{x}^{(k)}\|\}} < \epsilon$
- On pourra également utiliser un nombre maximal d'itérations comme condition d'arrêt

Exemple

- Utilisons la méthode du pas optimal pour déterminer un minimiseur local de $f(x_1, x_2, x_3) = (x_1 - 4)^4 + (x_2 - 3)^2 + 4(x_3 + 5)^4$ avec $\mathbf{x}^{(0)} = (4, 2, -1)$

Exemple

- Utilisons la méthode du pas optimal pour déterminer un minimiseur local de $f(x_1, x_2, x_3) = (x_1 - 4)^4 + (x_2 - 3)^2 + 4(x_3 + 5)^4$ avec $\mathbf{x}^{(0)} = (4, 2, -1)$

- Calcul du gradient : $\nabla f(\mathbf{x}) = \begin{pmatrix} 4(x_1 - 4)^3 \\ 2(x_2 - 3) \\ 16(x_3 + 5)^3 \end{pmatrix}$

Exemple

- Utilisons la méthode du pas optimal pour déterminer un minimiseur local de $f(x_1, x_2, x_3) = (x_1 - 4)^4 + (x_2 - 3)^2 + 4(x_3 + 5)^4$ avec $\mathbf{x}^{(0)} = (4, 2, -1)$

- Calcul du gradient : $\nabla f(\mathbf{x}) = \begin{pmatrix} 4(x_1 - 4)^3 \\ 2(x_2 - 3) \\ 16(x_3 + 5)^3 \end{pmatrix}$

- Déroulement de l'algorithme :

- $k \leftarrow 0$

Itération 1

- On cherche $\arg \min_{\alpha \geq 0} f(\mathbf{x}^{(0)} - \alpha \nabla f(\mathbf{x}^{(0)}))$

On a : $\nabla f(\mathbf{x}^{(0)}) = (0, -2, 1024)$

$$\begin{aligned} \phi^{(0)}(\alpha) &= f(\mathbf{x}^{(0)} - \alpha \nabla f(\mathbf{x}^{(0)})) \\ &= f\left(\begin{pmatrix} 4 \\ 2 \\ -1 \end{pmatrix} - \alpha \begin{pmatrix} 0 \\ -2 \\ 1024 \end{pmatrix}\right) \end{aligned}$$

Exemple (suite)

Itération 1 (suite)

$$\begin{aligned} \phi^{(0)}(\alpha) &= f(\mathbf{x}^{(0)} - \alpha \nabla f(\mathbf{x}^{(0)})) \\ &= f(4, 2 + 2\alpha, -1 - 1024\alpha) \\ &= (2\alpha - 1)^2 + 4(1024\alpha + 4)^4 \end{aligned}$$

En utilisant la méthode de la sécante on trouve :

$$\alpha_0 \leftarrow \arg \min_{\alpha \geq 0} \phi^{(0)}(\alpha) = 3.967 \times 10^{-3}$$

- $\mathbf{x}^{(1)} \leftarrow \mathbf{x}^{(0)} - \alpha_0 \nabla f(\mathbf{x}^{(0)}) = (4, 2.008, -5.062)$

- $k \leftarrow k + 1 = 1$

Exemple (suite)

Itération 1 (suite)

$$\begin{aligned} \phi^{(0)}(\alpha) &= f(\mathbf{x}^{(0)} - \alpha \nabla f(\mathbf{x}^{(0)})) \\ &= f(4, 2 + 2\alpha, -1 - 1024\alpha) \\ &= (2\alpha - 1)^2 + 4(1024\alpha + 4)^4 \end{aligned}$$

En utilisant la méthode de la sécante on trouve :

$$\alpha_0 \leftarrow \arg \min_{\alpha \geq 0} \phi^{(0)}(\alpha) = 3.967 \times 10^{-3}$$

- $\mathbf{x}^{(1)} \leftarrow \mathbf{x}^{(0)} - \alpha_0 \nabla f(\mathbf{x}^{(0)}) = (4, 2.008, -5.062)$

- $k \leftarrow k + 1 = 1$

Itération 2

- On cherche $\arg \min_{\alpha \geq 0} f(\mathbf{x}^{(1)} - \alpha \nabla f(\mathbf{x}^{(1)}))$

On a : $\nabla f(\mathbf{x}^{(1)}) = (0, -1.984, -0.003875)$

$$\begin{aligned} \phi^{(1)}(\alpha) &= f(\mathbf{x}^{(1)} - \alpha \nabla f(\mathbf{x}^{(1)})) \\ &= (2.008 - 1.984\alpha - 3)^2 + 4(-5.062 + 0.003875\alpha + 5)^4 \end{aligned}$$

Exemple (suite)

Itération 2 (suite)

En utilisant la méthode de la sécante on trouve :

$$\alpha_1 \leftarrow \arg \min_{\alpha \geq 0} \phi^{(1)}(\alpha) = 0.50$$

$$4 \quad \mathbf{x}^{(2)} \leftarrow \mathbf{x}^{(1)} - \alpha_1 \nabla f(\mathbf{x}^{(1)}) = (4, 3, -5.060)$$

$$5 \quad k \leftarrow k + 1 = 2$$

Exemple (suite)

Itération 3 (suite)

$$4 \quad \mathbf{x}^{(3)} \leftarrow \mathbf{x}^{(2)} - \alpha_2 \nabla f(\mathbf{x}^{(2)}) = (4, 3, -5.002)$$

$$5 \quad k \leftarrow k + 1 = 3$$

Exemple (suite)

Itération 2 (suite)

En utilisant la méthode de la sécante on trouve :

$$\alpha_1 \leftarrow \arg \min_{\alpha \geq 0} \phi^{(1)}(\alpha) = 0.50$$

$$4 \quad \mathbf{x}^{(2)} \leftarrow \mathbf{x}^{(1)} - \alpha_1 \nabla f(\mathbf{x}^{(1)}) = (4, 3, -5.060)$$

$$5 \quad k \leftarrow k + 1 = 2$$

Itération 3

$$3 \quad \text{On cherche } \arg \min_{\alpha \geq 0} f(\mathbf{x}^{(2)} - \alpha \nabla f(\mathbf{x}^{(2)}))$$

$$\text{On a : } \nabla f(\mathbf{x}^{(2)}) = (0, 0, -0.003525)$$

$$\begin{aligned} \phi^{(2)}(\alpha) &= f(\mathbf{x}^{(2)} - \alpha \nabla f(\mathbf{x}^{(2)})) \\ &= 4(-5.060 + 0.003525\alpha + 5)^4 \end{aligned}$$

En utilisant la méthode de la sécante on trouve :

$$\alpha_2 \leftarrow \arg \min_{\alpha \geq 0} \phi^{(2)}(\alpha) = 16.29$$

Exemple (suite)

Itération 3 (suite)

$$4 \quad \mathbf{x}^{(3)} \leftarrow \mathbf{x}^{(2)} - \alpha_2 \nabla f(\mathbf{x}^{(2)}) = (4, 3, -5.002)$$

$$5 \quad k \leftarrow k + 1 = 3$$

Fin de l'algorithme

$$2 \quad \mathbf{x}^{(3)} \text{ et } \mathbf{x}^{(2)} \text{ sont très proches : on s'arrête}$$

$$7 \quad \text{Output : } \mathbf{x}^* = (4, 3, -5.002)$$

Remarques sur les méthodes numériques d'optimisation multidimensionnelle

- Dans l'exemple le minimiseur est obtenu en peu d'itérations mais ce n'est pas une règle générale loin de là

Remarques sur les méthodes numériques d'optimisation multidimensionnelle

- Dans l'exemple le minimiseur est obtenu en peu d'itérations mais ce n'est pas une règle générale loin de là
- On a bien sûr recours à des logiciels de calculs numériques pour effectuer les calculs nécessaires à chaque étape

Remarques sur les méthodes numériques d'optimisation multidimensionnelle

- Dans l'exemple le minimiseur est obtenu en peu d'itérations mais ce n'est pas une règle générale loin de là
- On a bien sûr recours à des logiciels de calculs numériques pour effectuer les calculs nécessaires à chaque étape
- On a utilisé la méthode de la sécante pour déterminer le pas optimal. D'autres méthodes numériques d'optimisation unidimensionnelle peuvent être utilisées

Remarques sur les méthodes numériques d'optimisation multidimensionnelle

- Dans l'exemple le minimiseur est obtenu en peu d'itérations mais ce n'est pas une règle générale loin de là
- On a bien sûr recours à des logiciels de calculs numériques pour effectuer les calculs nécessaires à chaque étape
- On a utilisé la méthode de la sécante pour déterminer le pas optimal. D'autres méthodes numériques d'optimisation unidimensionnelle peuvent être utilisées
- En fait les méthodes numériques en optimisation unidimensionnelle appelées également **méthodes de recherche linéaire** sont essentielles aux méthodes numériques en optimisation multidimensionnelle

Remarques sur les méthodes numériques d'optimisation multidimensionnelle (suite)

- Les méthodes d'optimisation unidimensionnelle étant utilisées à chaque étape des méthodes d'optimisation multidimensionnelle, la complexité des premières a un impact sur la complexité des dernières

Remarques sur les méthodes numériques d'optimisation multidimensionnelle (suite)

- Les méthodes d'optimisation unidimensionnelle étant utilisées à chaque étape des méthodes d'optimisation multidimensionnelle, la complexité des premières a un impact sur la complexité des dernières
- Les méthodes d'optimisation unidimensionnelle que nous avons vues, malgré leur efficacité, alourdissent le temps de traitement des méthodes de descente de gradient

Remarques sur les méthodes numériques d'optimisation multidimensionnelle (suite)

- Les méthodes d'optimisation unidimensionnelle étant utilisées à chaque étape des méthodes d'optimisation multidimensionnelle, la complexité des premières a un impact sur la complexité des dernières
- Les méthodes d'optimisation unidimensionnelle que nous avons vues, malgré leur efficacité, alourdissent le temps de traitement des méthodes de descente de gradient
- Il existe en pratique des méthodes de **recherche linéaire** dite **inexacte** dont le but est de déterminer très rapidement un pas apportant un "progrès raisonnable" sur la diminution de f

Remarques sur les méthodes numériques d'optimisation multidimensionnelle (suite)

- Les méthodes d'optimisation unidimensionnelle étant utilisées à chaque étape des méthodes d'optimisation multidimensionnelle, la complexité des premières a un impact sur la complexité des dernières
- Les méthodes d'optimisation unidimensionnelle que nous avons vues, malgré leur efficacité, alourdissent le temps de traitement des méthodes de descente de gradient
- Il existe en pratique des méthodes de **recherche linéaire** dite **inexacte** dont le but est de déterminer très rapidement un pas apportant un "progrès raisonnable" sur la diminution de f
- La notion de "progrès raisonnable" est assimilée à des règles sur α (règle d'Armijo ou de Goldstein par exemple) qui sont vérifiables rapidement

Remarques sur les méthodes numériques d'optimisation multidimensionnelle (suite)

- Les méthodes d'optimisation unidimensionnelle étant utilisées à chaque étape des méthodes d'optimisation multidimensionnelle, la complexité des premières a un impact sur la complexité des dernières
- Les méthodes d'optimisation unidimensionnelle que nous avons vues, malgré leur efficacité, alourdissent le temps de traitement des méthodes de descente de gradient
- Il existe en pratique des méthodes de **recherche linéaire** dite **inexacte** dont le but est de déterminer très rapidement un pas apportant un "progrès raisonnable" sur la diminution de f
- La notion de "progrès raisonnable" est assimilée à des règles sur α (règle d'Armijo ou de Goldstein par exemple) qui sont vérifiables rapidement
- Souvent il est préférable de trouver un pas de "**progrès raisonnable**" rapidement que de déterminer exactement le pas optimal avec un plus **long temps de traitement**

Méthode du pas optimal pour des fonctions quadratiques

- On considère le cas particulier :

$$f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^t \mathbf{Q} \mathbf{x} - \mathbf{b}^t \mathbf{x}$$

où $\mathbf{Q} \in \mathcal{M}_n(\mathbb{R})$ est définie positive, $\mathbf{b} \in \mathbb{R}^n$ et $\mathbf{x} \in \mathbb{R}^n$. On suppose sans perte de généralité que \mathbf{Q} est symétrique

- Dans ce cas le gradient vaut :

$$\nabla f(\mathbf{x}) = \mathbf{Q} \mathbf{x} - \mathbf{b}$$

Méthode du pas optimal pour des fonctions quadratiques

- On considère le cas particulier :

$$f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^t \mathbf{Q} \mathbf{x} - \mathbf{b}^t \mathbf{x}$$

où $\mathbf{Q} \in \mathcal{M}_n(\mathbb{R})$ est définie positive, $\mathbf{b} \in \mathbb{R}^n$ et $\mathbf{x} \in \mathbb{R}^n$. On suppose sans perte de généralité que \mathbf{Q} est symétrique

Méthode du pas optimal pour des fonctions quadratiques

- On considère le cas particulier :

$$f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^t \mathbf{Q} \mathbf{x} - \mathbf{b}^t \mathbf{x}$$

où $\mathbf{Q} \in \mathcal{M}_n(\mathbb{R})$ est définie positive, $\mathbf{b} \in \mathbb{R}^n$ et $\mathbf{x} \in \mathbb{R}^n$. On suppose sans perte de généralité que \mathbf{Q} est symétrique

- Dans ce cas le gradient vaut :

$$\nabla f(\mathbf{x}) = \mathbf{Q} \mathbf{x} - \mathbf{b}$$

- Le hessien vaut :

$$D^2 f(\mathbf{x}) = \mathbf{Q}$$

Méthode du pas optimal pour des fonctions quadratiques

- On considère le cas particulier :

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^t\mathbf{Q}\mathbf{x} - \mathbf{b}^t\mathbf{x}$$

où $\mathbf{Q} \in \mathcal{M}_n(\mathbb{R})$ est définie positive, $\mathbf{b} \in \mathbb{R}^n$ et $\mathbf{x} \in \mathbb{R}^n$. On suppose sans perte de généralité que \mathbf{Q} est symétrique

- Dans ce cas le gradient vaut :

$$\nabla f(\mathbf{x}) = \mathbf{Q}\mathbf{x} - \mathbf{b}$$

- Le hessien vaut :

$$D^2f(\mathbf{x}) = \mathbf{Q}$$

- Pour simplifier les notations on notera : $\mathbf{g}^{(k)} = \nabla f(\mathbf{x}^{(k)})$

Méthode du pas optimal pour des fonctions quadratiques

- On considère le cas particulier :

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^t\mathbf{Q}\mathbf{x} - \mathbf{b}^t\mathbf{x}$$

où $\mathbf{Q} \in \mathcal{M}_n(\mathbb{R})$ est définie positive, $\mathbf{b} \in \mathbb{R}^n$ et $\mathbf{x} \in \mathbb{R}^n$. On suppose sans perte de généralité que \mathbf{Q} est symétrique

- Dans ce cas le gradient vaut :

$$\nabla f(\mathbf{x}) = \mathbf{Q}\mathbf{x} - \mathbf{b}$$

- Le hessien vaut :

$$D^2f(\mathbf{x}) = \mathbf{Q}$$

- Pour simplifier les notations on notera : $\mathbf{g}^{(k)} = \nabla f(\mathbf{x}^{(k)})$

- En appliquant la méthode du pas optimal on cherche donc :

$$\begin{aligned} \alpha_k &= \min_{\alpha \geq 0} f(\mathbf{x}^{(k)} - \alpha \mathbf{g}^{(k)}) \\ &= \arg \min_{\alpha \geq 0} \left(\frac{1}{2}(\mathbf{x}^{(k)} - \alpha \mathbf{g}^{(k)})^t \mathbf{Q} (\mathbf{x}^{(k)} - \alpha \mathbf{g}^{(k)}) - (\mathbf{x}^{(k)} - \alpha \mathbf{g}^{(k)})^t \mathbf{b} \right) \end{aligned}$$

Méthode du pas optimal pour des fonctions quadratiques (suite)

- Dans le cas particulier des fonctions quadratiques, on a une formulation explicite de α_k

Méthode du pas optimal pour des fonctions quadratiques (suite)

- Dans le cas particulier des fonctions quadratiques, on a une formulation explicite de α_k

- α_k est tel que $\frac{d\phi^{(k)}}{d\alpha} = 0$ (CNPO) :

$$\begin{aligned} \frac{d\phi^{(k)}}{d\alpha} = 0 &\Leftrightarrow -(\mathbf{x}^{(k)} - \alpha \mathbf{g}^{(k)})^t \mathbf{Q} \mathbf{g}^{(k)} + \mathbf{b}^t \mathbf{g}^{(k)} = 0 \\ &\Leftrightarrow \alpha (\mathbf{g}^{(k)})^t \mathbf{Q} \mathbf{g}^{(k)} = (\mathbf{x}^{(k)})^t \mathbf{Q} \mathbf{g}^{(k)} - \mathbf{b}^t \mathbf{g}^{(k)} \\ &\Leftrightarrow \alpha (\mathbf{g}^{(k)})^t \mathbf{Q} \mathbf{g}^{(k)} = ((\mathbf{x}^{(k)})^t \mathbf{Q} - \mathbf{b}^t) \mathbf{g}^{(k)} \\ &\Leftrightarrow \alpha (\mathbf{g}^{(k)})^t \mathbf{Q} \mathbf{g}^{(k)} = (\mathbf{g}^{(k)})^t \mathbf{g}^{(k)} \\ &\Leftrightarrow \alpha = \frac{(\mathbf{g}^{(k)})^t \mathbf{g}^{(k)}}{(\mathbf{g}^{(k)})^t \mathbf{Q} \mathbf{g}^{(k)}} \end{aligned}$$

Méthode du pas optimal pour des fonctions quadratiques (suite)

- Dans le cas particulier des fonctions quadratiques, on a une formulation explicite de α_k

- α_k est tel que $\frac{d\phi^{(k)}}{d\alpha} = 0$ (CNPO) :

$$\begin{aligned} \frac{d\phi^{(k)}}{d\alpha} = 0 &\Leftrightarrow -(\mathbf{x}^{(k)} - \alpha \mathbf{g}^{(k)})^t \mathbf{Q} \mathbf{g}^{(k)} + \mathbf{b}^t \mathbf{g}^{(k)} = 0 \\ &\Leftrightarrow \alpha (\mathbf{g}^{(k)})^t \mathbf{Q} \mathbf{g}^{(k)} = (\mathbf{x}^{(k)})^t \mathbf{Q} \mathbf{g}^{(k)} - \mathbf{b}^t \mathbf{g}^{(k)} \\ &\Leftrightarrow \alpha (\mathbf{g}^{(k)})^t \mathbf{Q} \mathbf{g}^{(k)} = ((\mathbf{x}^{(k)})^t \mathbf{Q} - \mathbf{b}^t) \mathbf{g}^{(k)} \\ &\Leftrightarrow \alpha (\mathbf{g}^{(k)})^t \mathbf{Q} \mathbf{g}^{(k)} = (\mathbf{g}^{(k)})^t \mathbf{g}^{(k)} \\ &\Leftrightarrow \alpha = \frac{(\mathbf{g}^{(k)})^t \mathbf{g}^{(k)}}{(\mathbf{g}^{(k)})^t \mathbf{Q} \mathbf{g}^{(k)}} \end{aligned}$$

- On a donc :

$$\alpha_k = \frac{\nabla f(\mathbf{x}^{(k)})^t \nabla f(\mathbf{x}^{(k)})}{\nabla f(\mathbf{x}^{(k)})^t \mathbf{Q} \nabla f(\mathbf{x}^{(k)})}$$

Méthode du pas optimal pour des fonctions quadratiques (suite)

- Dans le cas particulier des fonctions quadratiques, on a une formulation explicite de α_k

- α_k est tel que $\frac{d\phi^{(k)}}{d\alpha} = 0$ (CNPO) :

$$\begin{aligned} \frac{d\phi^{(k)}}{d\alpha} = 0 &\Leftrightarrow -(\mathbf{x}^{(k)} - \alpha \mathbf{g}^{(k)})^t \mathbf{Q} \mathbf{g}^{(k)} + \mathbf{b}^t \mathbf{g}^{(k)} = 0 \\ &\Leftrightarrow \alpha (\mathbf{g}^{(k)})^t \mathbf{Q} \mathbf{g}^{(k)} = (\mathbf{x}^{(k)})^t \mathbf{Q} \mathbf{g}^{(k)} - \mathbf{b}^t \mathbf{g}^{(k)} \\ &\Leftrightarrow \alpha (\mathbf{g}^{(k)})^t \mathbf{Q} \mathbf{g}^{(k)} = ((\mathbf{x}^{(k)})^t \mathbf{Q} - \mathbf{b}^t) \mathbf{g}^{(k)} \\ &\Leftrightarrow \alpha (\mathbf{g}^{(k)})^t \mathbf{Q} \mathbf{g}^{(k)} = (\mathbf{g}^{(k)})^t \mathbf{g}^{(k)} \\ &\Leftrightarrow \alpha = \frac{(\mathbf{g}^{(k)})^t \mathbf{g}^{(k)}}{(\mathbf{g}^{(k)})^t \mathbf{Q} \mathbf{g}^{(k)}} \end{aligned}$$

- On a donc :

$$\alpha_k = \frac{\nabla f(\mathbf{x}^{(k)})^t \nabla f(\mathbf{x}^{(k)})}{\nabla f(\mathbf{x}^{(k)})^t \mathbf{Q} \nabla f(\mathbf{x}^{(k)})}$$

- Finalement si f est une fonction quadratique l'algorithme du pas optimal revient à remplacer la ligne 4 par la formule suivante :

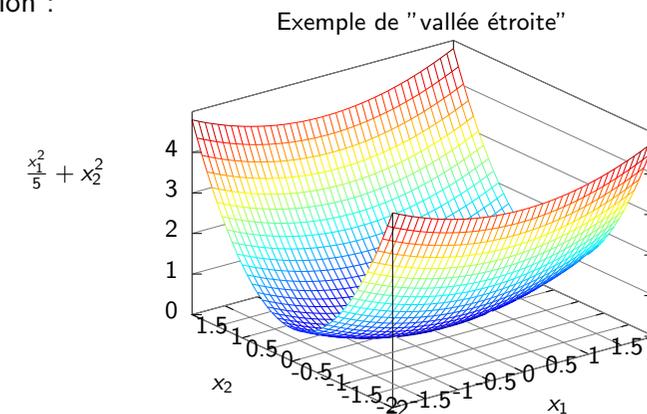
$$4 \quad \mathbf{x}^{(k+1)} \leftarrow \mathbf{x}^{(k)} - \frac{\nabla f(\mathbf{x}^{(k)})^t \nabla f(\mathbf{x}^{(k)})}{\nabla f(\mathbf{x}^{(k)})^t \mathbf{Q} \nabla f(\mathbf{x}^{(k)})} \nabla f(\mathbf{x}^{(k)})$$

Remarques sur la méthode du pas optimal

- Méthode facile à mettre en oeuvre
- La complexité réside notamment en les calculs des α_k et des gradients
- Mais, l'algorithme peut être relativement lent à converger ie beaucoup d'itérations (fonction objectif de type "vallée étroite")

Remarques sur la méthode du pas optimal

- Méthode facile à mettre en oeuvre
- La complexité réside notamment en les calculs des α_k et des gradients
- Mais, l'algorithme peut être relativement lent à converger ie beaucoup d'itérations (fonction objectif de type "vallée étroite")
- Illustration :



Convergence des méthodes numériques itératives

- L'algorithme de descente de gradient est dit **itératif** : on calcule successivement une suite de points

Convergence des méthodes numériques itératives

- L'algorithme de descente de gradient est dit **itératif** : on calcule successivement une suite de points
- Notions de convergence des algorithmes itératifs :

Définition. (Convergence globale)

Un algorithme itératif est dit **globalement convergent** si pour tout point x_0 l'algorithme produit une suite de points qui converge vers un point (le minimiseur) satisfaisant la CNPO (point critique)

Définition. (Convergence locale)

Quand l'algorithme itératif n'est pas globalement convergent, on dit qu'il est **localement convergent** si l'algorithme produit une suite de points qui converge vers un point satisfaisant la CNPO à condition que le point initial x_0 soit **proche du minimiseur**

Convergence des méthodes numériques itératives

- L'algorithme de descente de gradient est dit **itératif** : on calcule successivement une suite de points
- Notions de convergence des algorithmes itératifs :

Définition. (Convergence globale)

Un algorithme itératif est dit **globalement convergent** si pour tout point x_0 l'algorithme produit une suite de points qui converge vers un point (le minimiseur) satisfaisant la CNPO (point critique)

Vitesse de convergence des méthodes numériques itératives

- La notion de **vitesse de convergence** permet de qualifier la plus ou moins grande vitesse avec laquelle un algorithme itératif peut atteindre un minimiseur

Vitesse de convergence des méthodes numériques itératives

- La notion de **vitesse de convergence** permet de qualifier la plus ou moins grande vitesse avec laquelle un algorithme itératif peut atteindre un minimiseur
- Nous avons le résultat suivant sur la méthode du pas optimal :

Théorème.

La suite $\{\mathbf{x}^{(k)}\}$ des points déterminés par la méthode du pas optimal est tel que $\{\mathbf{x}^{(k)}\} \rightarrow \mathbf{x}^*$ pour tout point initial \mathbf{x}_0

Vitesse de convergence des méthodes numériques itératives

- La notion de **vitesse de convergence** permet de qualifier la plus ou moins grande vitesse avec laquelle un algorithme itératif peut atteindre un minimiseur
- Nous avons le résultat suivant sur la méthode du pas optimal :

Théorème.

La suite $\{\mathbf{x}^{(k)}\}$ des points déterminés par la méthode du pas optimal est tel que $\{\mathbf{x}^{(k)}\} \rightarrow \mathbf{x}^*$ pour tout point initial \mathbf{x}_0

- Dans le cas d'une fonction quadratique et pour une méthode à pas fixe (ie $\forall k : \alpha_k = \alpha_0$) on a le résultat suivant :

Théorème.

La suite $\{\mathbf{x}^{(k)}\}$ des points déterminés par la méthode du pas fixe est tel que $\{\mathbf{x}^{(k)}\} \rightarrow \mathbf{x}^*$ pour tout point initial \mathbf{x}_0 ssi :

$$0 < \alpha_0 < \frac{2}{\lambda_{\max}(\mathbf{Q})}$$

Vitesse de convergence des méthodes numériques itératives (suite)

Définition. (Ordre de convergence)

Soit $\{\mathbf{x}^{(k)}\}$ une suite convergeant vers \mathbf{x}^* ie tel que $\lim_{k \rightarrow \infty} \|\mathbf{x}^{(k)} - \mathbf{x}^*\| = 0$.
On dit que l'ordre de convergence est de $p \in \mathbb{R}$ si :

$$0 < \lim_{k \rightarrow \infty} \frac{\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\|}{\|\mathbf{x}^{(k)} - \mathbf{x}^*\|^p} < \infty$$

Vitesse de convergence des méthodes numériques itératives (suite)

Définition. (Ordre de convergence)

Soit $\{\mathbf{x}^{(k)}\}$ une suite convergeant vers \mathbf{x}^* ie tel que $\lim_{k \rightarrow \infty} \|\mathbf{x}^{(k)} - \mathbf{x}^*\| = 0$.
On dit que l'ordre de convergence est de $p \in \mathbb{R}$ si :

$$0 < \lim_{k \rightarrow \infty} \frac{\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\|}{\|\mathbf{x}^{(k)} - \mathbf{x}^*\|^p} < \infty$$

- Etant donné une suite, l'ordre de convergence est une mesure de la vitesse de convergence de la suite : plus p est grand plus l'algorithme converge rapidement

Vitesse de convergence des méthodes numériques itératives (suite)

Définition. (Ordre de convergence)

Soit $\{\mathbf{x}^{(k)}\}$ une suite convergeant vers \mathbf{x}^* ie tel que $\lim_{k \rightarrow \infty} \|\mathbf{x}^{(k)} - \mathbf{x}^*\| = 0$.
On dit que l'ordre de convergence est de $p \in \mathbb{R}$ si :

$$0 < \lim_{k \rightarrow \infty} \frac{\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\|}{\|\mathbf{x}^{(k)} - \mathbf{x}^*\|^p} < \infty$$

- Etant donné une suite, l'ordre de convergence est une mesure de la vitesse de convergence de la suite : plus p est grand plus l'algorithme converge rapidement
- Si $p = 1$ on parle de **convergence linéaire**
- Si $p = 2$ on parle de **convergence quadratique**

Exemple

- Supposons une suite de nombres définie par : $x_k = \frac{1}{k}$
- $x_k \rightarrow 0$

Exemple

- Supposons une suite de nombres définie par : $x_k = \frac{1}{k}$

Exemple

- Supposons une suite de nombres définie par : $x_k = \frac{1}{k}$
- $x_k \rightarrow 0$
- $\frac{|x_{k+1}|}{|x_k|^p} = \frac{k^p}{k+1}$

Exemple

- Supposons une suite de nombres définie par : $x_k = \frac{1}{k}$
- $x_k \rightarrow 0$
- $\frac{|x_{k+1}|}{|x_k|^p} = \frac{k^p}{k+1}$
- On a les différents cas suivants :
 - ▶ Si $p < 1$ la suite converge vers 0

Exemple

- Supposons une suite de nombres définie par : $x_k = \frac{1}{k}$
- $x_k \rightarrow 0$
- $\frac{|x_{k+1}|}{|x_k|^p} = \frac{k^p}{k+1}$
- On a les différents cas suivants :
 - ▶ Si $p < 1$ la suite converge vers 0
 - ▶ Si $p > 1$ la suite diverge vers ∞
 - ▶ Si $p = 1$ la suite converge vers 1

Exemple

- Supposons une suite de nombres définie par : $x_k = \frac{1}{k}$
- $x_k \rightarrow 0$
- $\frac{|x_{k+1}|}{|x_k|^p} = \frac{k^p}{k+1}$
- On a les différents cas suivants :
 - ▶ Si $p < 1$ la suite converge vers 0
 - ▶ Si $p > 1$ la suite diverge vers ∞

Exemple

- Supposons une suite de nombres définie par : $x_k = \frac{1}{k}$
- $x_k \rightarrow 0$
- $\frac{|x_{k+1}|}{|x_k|^p} = \frac{k^p}{k+1}$
- On a les différents cas suivants :
 - ▶ Si $p < 1$ la suite converge vers 0
 - ▶ Si $p > 1$ la suite diverge vers ∞
 - ▶ Si $p = 1$ la suite converge vers 1
- Donc, l'ordre de convergence de la suite est de 1

Vitesse de convergence de la méthode à pas optimal

Théorème.

Soit $\{\mathbf{x}^{(k)}\}$ une suite des points déterminés par la méthode du pas optimal pour la minimisation d'une fonction f . Alors, l'ordre de convergence de $\{\mathbf{x}^{(k)}\}$ est 1 dans le pire des cas (étant donné le point initial \mathbf{x}_0).

Algorithmes de Newton

- Rappelons que la méthode de descente à pas optimal utilise uniquement la dérivée première de f et que l'ordre de convergence est linéaire
- En utilisant les dérivées d'ordre supérieure, on peut avoir un algorithme itératif dont l'ordre de convergence est meilleure : il s'agit de la **méthode de Newton** (dit Newton-Raphson également)

Algorithmes de Newton

- Rappelons que la méthode de descente à pas optimal utilise uniquement la dérivée première de f et que l'ordre de convergence est linéaire

Algorithmes de Newton

- Rappelons que la méthode de descente à pas optimal utilise uniquement la dérivée première de f et que l'ordre de convergence est linéaire
- En utilisant les dérivées d'ordre supérieure, on peut avoir un algorithme itératif dont l'ordre de convergence est meilleure : il s'agit de la **méthode de Newton** (dit Newton-Raphson également)
- Cette méthode utilise de plus la **dérivée seconde** et est plus efficace à condition que le point initial \mathbf{x}_0 est relativement proche du minimiseur

Algorithmes de Newton

- Rappelons que la méthode de descente à pas optimal utilise uniquement la dérivée première de f et que l'ordre de convergence est linéaire
- En utilisant les dérivées d'ordre supérieure, on peut avoir un algorithme itératif dont l'ordre de convergence est meilleure : il s'agit de la **méthode de Newton** (dit Newton-Raphson également)
- Cette méthode utilise de plus la **dérivée seconde** et est plus efficace à condition que le point initial \mathbf{x}_0 est relativement proche du minimiseur
- L'idée est (comme en optimisation unidimensionnelle) d'approximer localement f par une fonction quadratique q et de déterminer le minimiseur de q à la place de f . Ce minimiseur devient alors le point initial de l'itération suivante.

Algorithmes de Newton

- Rappelons que la méthode de descente à pas optimal utilise uniquement la dérivée première de f et que l'ordre de convergence est linéaire
- En utilisant les dérivées d'ordre supérieure, on peut avoir un algorithme itératif dont l'ordre de convergence est meilleure : il s'agit de la **méthode de Newton** (dit Newton-Raphson également)
- Cette méthode utilise de plus la **dérivée seconde** et est plus efficace à condition que le point initial \mathbf{x}_0 est relativement proche du minimiseur
- L'idée est (comme en optimisation unidimensionnelle) d'approximer localement f par une fonction quadratique q et de déterminer le minimiseur de q à la place de f . Ce minimiseur devient alors le point initial de l'itération suivante.
- Si f est quadratique, alors l'approximation est exacte et la méthode permet de déterminer \mathbf{x}^* en une seule itération

Algorithmes de Newton (suite)

- Nous supposons dans cette sous-section que $f \in \mathcal{C}^2$. Rappelons qu'au voisinage de $\mathbf{x}^{(k)}$ la série de Taylor de f est donnée par :

$$f(\mathbf{x}) \simeq f(\mathbf{x}^{(k)}) + \nabla f(\mathbf{x}^{(k)})^t (\mathbf{x} - \mathbf{x}^{(k)}) + \frac{1}{2} (\mathbf{x} - \mathbf{x}^{(k)})^t D^2 f(\mathbf{x}^{(k)}) (\mathbf{x} - \mathbf{x}^{(k)}) = q(\mathbf{x})$$

Algorithmes de Newton (suite)

- Nous supposons dans cette sous-section que $f \in \mathcal{C}^2$. Rappelons qu'au voisinage de $\mathbf{x}^{(k)}$ la série de Taylor de f est donnée par :

$$f(\mathbf{x}) \simeq f(\mathbf{x}^{(k)}) + \nabla f(\mathbf{x}^{(k)})^t (\mathbf{x} - \mathbf{x}^{(k)}) + \frac{1}{2} (\mathbf{x} - \mathbf{x}^{(k)})^t D^2 f(\mathbf{x}^{(k)}) (\mathbf{x} - \mathbf{x}^{(k)}) = q(\mathbf{x})$$

- Pour simplifier les notations nous poserons $\mathbf{g}^{(k)} = \nabla f(\mathbf{x}^{(k)})$

Algorithmes de Newton (suite)

- Nous supposons dans cette sous-section que $f \in \mathcal{C}^2$. Rappelons qu'au voisinage de $\mathbf{x}^{(k)}$ la série de Taylor de f est donnée par :

$$f(\mathbf{x}) \simeq f(\mathbf{x}^{(k)}) + \nabla f(\mathbf{x}^{(k)})^t (\mathbf{x} - \mathbf{x}^{(k)}) + \frac{1}{2} (\mathbf{x} - \mathbf{x}^{(k)})^t D^2 f(\mathbf{x}^{(k)}) (\mathbf{x} - \mathbf{x}^{(k)}) = q(\mathbf{x})$$

- Pour simplifier les notations nous poserons $\mathbf{g}^{(k)} = \nabla f(\mathbf{x}^{(k)})$
- En appliquant les CNPO à q , on obtient :

$$\nabla q(\mathbf{x}) = \mathbf{g}^{(k)} + D^2 f(\mathbf{x}^{(k)}) (\mathbf{x} - \mathbf{x}^{(k)}) = 0$$

Algorithmes de Newton (suite)

- Nous supposons dans cette sous-section que $f \in \mathcal{C}^2$. Rappelons qu'au voisinage de $\mathbf{x}^{(k)}$ la série de Taylor de f est donnée par :

$$f(\mathbf{x}) \simeq f(\mathbf{x}^{(k)}) + \nabla f(\mathbf{x}^{(k)})^t (\mathbf{x} - \mathbf{x}^{(k)}) + \frac{1}{2} (\mathbf{x} - \mathbf{x}^{(k)})^t D^2 f(\mathbf{x}^{(k)}) (\mathbf{x} - \mathbf{x}^{(k)}) = q(\mathbf{x})$$

- Pour simplifier les notations nous poserons $\mathbf{g}^{(k)} = \nabla f(\mathbf{x}^{(k)})$
- En appliquant les CNPO à q , on obtient :

$$\nabla q(\mathbf{x}) = \mathbf{g}^{(k)} + D^2 f(\mathbf{x}^{(k)}) (\mathbf{x} - \mathbf{x}^{(k)}) = 0$$

- Si $D^2 f(\mathbf{x}^{(k)}) > 0$ alors q atteint son minimum en :

$$\mathbf{x}^{(k)} - D^2 f(\mathbf{x}^{(k)})^{-1} \mathbf{g}^{(k)}$$

- L'algorithme itératif de descente de gradient devient :

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - D^2 f(\mathbf{x}^{(k)})^{-1} \mathbf{g}^{(k)}$$

Algorithmes de Newton (suite)

- Nous supposons dans cette sous-section que $f \in \mathcal{C}^2$. Rappelons qu'au voisinage de $\mathbf{x}^{(k)}$ la série de Taylor de f est donnée par :

$$f(\mathbf{x}) \simeq f(\mathbf{x}^{(k)}) + \nabla f(\mathbf{x}^{(k)})^t (\mathbf{x} - \mathbf{x}^{(k)}) + \frac{1}{2} (\mathbf{x} - \mathbf{x}^{(k)})^t D^2 f(\mathbf{x}^{(k)}) (\mathbf{x} - \mathbf{x}^{(k)}) = q(\mathbf{x})$$

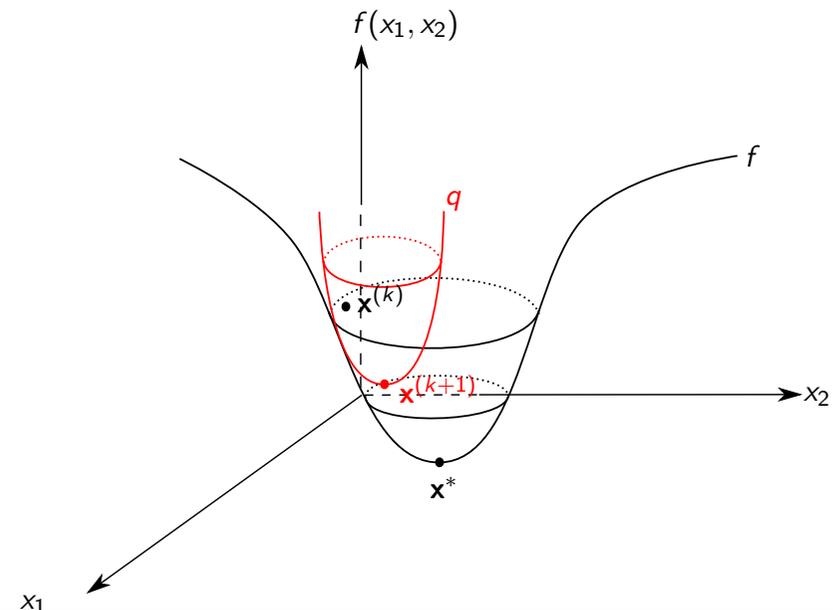
- Pour simplifier les notations nous poserons $\mathbf{g}^{(k)} = \nabla f(\mathbf{x}^{(k)})$
- En appliquant les CNPO à q , on obtient :

$$\nabla q(\mathbf{x}) = \mathbf{g}^{(k)} + D^2 f(\mathbf{x}^{(k)}) (\mathbf{x} - \mathbf{x}^{(k)}) = 0$$

- Si $D^2 f(\mathbf{x}^{(k)}) > 0$ alors q atteint son minimum en :

$$\mathbf{x}^{(k)} - D^2 f(\mathbf{x}^{(k)})^{-1} \mathbf{g}^{(k)}$$

Illustration



Pseudo-code de l'algorithme de Newton

Input : $f \in \mathcal{C}^2, \mathbf{x}^0$

- 1 $k \leftarrow 0$
- 2 **Tant que** condition d'arrêt non satisfaite **faire**
- 3 $\mathbf{x}^{(k+1)} \leftarrow \mathbf{x}^{(k)} - D^2f(\mathbf{x}^{(k)})^{-1}\nabla f(\mathbf{x}^{(k)})$
- 4 $k \leftarrow k + 1$
- 5 **Fin Tant que**
- 6 **Output :** $\mathbf{x}^{(k)}$

Exemple

- Utilisons la méthode de Newton pour déterminer un minimiseur local de
 $f(x_1, x_2, x_3, x_4) = (x_1 - 10x_2)^2 + 5(x_3 - x_4)^2 + (x_2 - 2x_3)^4 + 10(x_1 - x_4)^4$
avec $\mathbf{x}^{(0)} = (3, -1, 0, 1)$
- On a $f(\mathbf{x}_0) = 215$
- Calcul du gradient :

Exemple

- Utilisons la méthode de Newton pour déterminer un minimiseur local de
 $f(x_1, x_2, x_3, x_4) = (x_1 - 10x_2)^2 + 5(x_3 - x_4)^2 + (x_2 - 2x_3)^4 + 10(x_1 - x_4)^4$
avec $\mathbf{x}^{(0)} = (3, -1, 0, 1)$

Exemple

- Utilisons la méthode de Newton pour déterminer un minimiseur local de
 $f(x_1, x_2, x_3, x_4) = (x_1 - 10x_2)^2 + 5(x_3 - x_4)^2 + (x_2 - 2x_3)^4 + 10(x_1 - x_4)^4$
avec $\mathbf{x}^{(0)} = (3, -1, 0, 1)$
- On a $f(\mathbf{x}_0) = 215$
- Calcul du gradient :

$$\nabla f(\mathbf{x}) = \begin{pmatrix} 2(x_1 + 10x_2) + 40(x_1 - x_4)^3 \\ 20(x_1 + 10x_2) + 4(x_2 - 2x_3)^3 \\ 10(x_3 - x_4) - 8(x_2 - 2x_3)^3 \\ -10(x_3 - x_4) - 40(x_1 - x_4)^3 \end{pmatrix}$$

- Calcul du hessien :

Exemple

- Utilisons la méthode de Newton pour déterminer un minimiseur local de
 $f(x_1, x_2, x_3, x_4) = (x_1 - 10x_2)^2 + 5(x_3 - x_4)^2 + (x_2 - 2x_3)^4 + 10(x_1 - x_4)^4$
 avec $\mathbf{x}^{(0)} = (3, -1, 0, 1)$
- On a $f(\mathbf{x}_0) = 215$
- Calcul du gradient :

$$\nabla f(\mathbf{x}) = \begin{pmatrix} 2(x_1 + 10x_2) + 40(x_1 - x_4)^3 \\ 20(x_1 + 10x_2) + 4(x_2 - 2x_3)^3 \\ 10(x_3 - x_4) - 8(x_2 - 2x_3)^3 \\ -10(x_3 - x_4) - 40(x_1 - x_4)^3 \end{pmatrix}$$

- Calcul du hessien :

$$D^2f(\mathbf{x}) = \begin{pmatrix} 2 + 120(x_1 - x_4)^2 & 20 & 0 & -120(x_1 - x_4)^2 \\ 20 & 200 + 12(x_2 - 2x_3)^2 & -24(x_2 - 2x_3)^2 & 0 \\ 0 & -24(x_2 - 2x_3)^2 & 10 + 48(x_2 - 2x_3)^2 & -10 \\ -120(x_1 - x_4)^2 & 0 & -10 & 10 + 120(x_1 - x_4)^2 \end{pmatrix}$$

Exemple (suite)

Déroulement de l'algorithme :

- $k \leftarrow 0$
Itération 1

Exemple (suite)

Déroulement de l'algorithme :

- $k \leftarrow 0$
Itération 1
- $\mathbf{g}^{(0)} = (306, -144, -2, -310)$

Exemple (suite)

Déroulement de l'algorithme :

- $k \leftarrow 0$
Itération 1
- $\mathbf{g}^{(0)} = (306, -144, -2, -310)$

$$D^2f(\mathbf{x}^{(0)}) = \begin{pmatrix} 482 & 20 & 0 & -480 \\ 20 & 212 & -24 & 0 \\ 0 & -24 & 58 & -10 \\ -480 & 0 & -10 & 490 \end{pmatrix}$$

Exemple (suite)

Déroulement de l'algorithme :

1 $k \leftarrow 0$

Itération 1

3 $\mathbf{g}^{(0)} = (306, -144, -2, -310)$

$$D^2f(\mathbf{x}^{(0)}) = \begin{pmatrix} 482 & 20 & 0 & -480 \\ 20 & 212 & -24 & 0 \\ 0 & -24 & 58 & -10 \\ -480 & 0 & -10 & 490 \end{pmatrix}$$

$$D^2f(\mathbf{x}^{(0)})^{-1} = \begin{pmatrix} .1126 & -.0089 & .0154 & .1106 \\ -.0089 & .0057 & .0008 & -.0087 \\ .0154 & .0008 & .0203 & .0155 \\ .1106 & -.0087 & .0155 & .1107 \end{pmatrix}$$

Exemple (suite)

Déroulement de l'algorithme :

1 $k \leftarrow 0$

Itération 1

3 $\mathbf{g}^{(0)} = (306, -144, -2, -310)$

$$D^2f(\mathbf{x}^{(0)}) = \begin{pmatrix} 482 & 20 & 0 & -480 \\ 20 & 212 & -24 & 0 \\ 0 & -24 & 58 & -10 \\ -480 & 0 & -10 & 490 \end{pmatrix}$$

$$D^2f(\mathbf{x}^{(0)})^{-1} = \begin{pmatrix} .1126 & -.0089 & .0154 & .1106 \\ -.0089 & .0057 & .0008 & -.0087 \\ .0154 & .0008 & .0203 & .0155 \\ .1106 & -.0087 & .0155 & .1107 \end{pmatrix}$$

$$D^2f(\mathbf{x}^{(0)})^{-1}\mathbf{g}^{(0)} = (1.4127, -0.8413, -0.2540, 0.7460)$$

Exemple (suite)

Déroulement de l'algorithme :

1 $k \leftarrow 0$

Itération 1

3 $\mathbf{g}^{(0)} = (306, -144, -2, -310)$

$$D^2f(\mathbf{x}^{(0)}) = \begin{pmatrix} 482 & 20 & 0 & -480 \\ 20 & 212 & -24 & 0 \\ 0 & -24 & 58 & -10 \\ -480 & 0 & -10 & 490 \end{pmatrix}$$

$$D^2f(\mathbf{x}^{(0)})^{-1} = \begin{pmatrix} .1126 & -.0089 & .0154 & .1106 \\ -.0089 & .0057 & .0008 & -.0087 \\ .0154 & .0008 & .0203 & .0155 \\ .1106 & -.0087 & .0155 & .1107 \end{pmatrix}$$

$$D^2f(\mathbf{x}^{(0)})^{-1}\mathbf{g}^{(0)} = (1.4127, -0.8413, -0.2540, 0.7460)$$

$$\mathbf{x}^{(1)} \leftarrow \mathbf{x}^{(0)} - D^2f(\mathbf{x}^{(0)})^{-1}\nabla f(\mathbf{x}^{(0)}) = (1.5873, -0, 1587, 0.2540, 0.2540)$$

Exemple (suite)

Déroulement de l'algorithme :

1 $k \leftarrow 0$

Itération 1

3 $\mathbf{g}^{(0)} = (306, -144, -2, -310)$

$$D^2f(\mathbf{x}^{(0)}) = \begin{pmatrix} 482 & 20 & 0 & -480 \\ 20 & 212 & -24 & 0 \\ 0 & -24 & 58 & -10 \\ -480 & 0 & -10 & 490 \end{pmatrix}$$

$$D^2f(\mathbf{x}^{(0)})^{-1} = \begin{pmatrix} .1126 & -.0089 & .0154 & .1106 \\ -.0089 & .0057 & .0008 & -.0087 \\ .0154 & .0008 & .0203 & .0155 \\ .1106 & -.0087 & .0155 & .1107 \end{pmatrix}$$

$$D^2f(\mathbf{x}^{(0)})^{-1}\mathbf{g}^{(0)} = (1.4127, -0.8413, -0.2540, 0.7460)$$

$$\mathbf{x}^{(1)} \leftarrow \mathbf{x}^{(0)} - D^2f(\mathbf{x}^{(0)})^{-1}\nabla f(\mathbf{x}^{(0)}) = (1.5873, -0, 1587, 0.2540, 0.2540)$$

$$f(\mathbf{x}^{(1)}) = 31.8$$

Exemple (suite)

Déroulement de l'algorithme :

$$4 \quad k \leftarrow k + 1 = 1$$

Itération 2

$$3 \quad \mathbf{g}^{(1)} = (94.81, -1.179, 2.371, -94.81)$$

$$D^2f(\mathbf{x}^{(1)}) = \begin{pmatrix} 215.3 & 20 & 0 & -213.3 \\ 20 & 205.3 & -10.67 & 0 \\ 0 & -10.67 & 31.34 & -10 \\ -213.3 & 0 & -10 & 223.3 \end{pmatrix}$$

$$D^2f(\mathbf{x}^{(1)})^{-1}\mathbf{g}^{(1)} = (0.5291, -0.0529, -0.0846, 0.0846)$$

$$\mathbf{x}^{(2)} \leftarrow \mathbf{x}^{(1)} - D^2f(\mathbf{x}^{(1)})^{-1}\nabla f(\mathbf{x}^{(1)}) = (1.0582, -0, 1058, 0.1694, 0.1694)$$

$$f(\mathbf{x}^{(2)}) = 6.28$$

Remarques sur la méthode de Newton

- Remarquons que la k ème itération de l'algorithme peut se résoudre en deux sous-étapes :

- 1 Résoudre $D^2f(\mathbf{x}^{(k)})\mathbf{d}^{(k)} = -\mathbf{g}^{(k)}$

- 2 Calculer $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mathbf{d}^{(k)}$

Exemple (suite)

Déroulement de l'algorithme :

$$4 \quad k \leftarrow k + 1 = 2$$

Itération 3

$$3 \quad \mathbf{g}^{(2)} = (28.09, -0.3475, 0.7031, -28.08)$$

$$D^2f(\mathbf{x}^{(2)}) = \begin{pmatrix} 96.80 & 20 & 0 & -94.80 \\ 20 & 202.4 & -4.744 & 0 \\ 0 & -4.744 & 19.49 & -10 \\ -94.80 & 0 & -10 & 104.80 \end{pmatrix}$$

$$\mathbf{x}^{(3)} \leftarrow \mathbf{x}^{(2)} - D^2f(\mathbf{x}^{(2)})^{-1}\nabla f(\mathbf{x}^{(2)}) = (0.7037, -0, 0704, 0.1121, 0.1111)$$

$$f(\mathbf{x}^{(3)}) = 1.24$$

Remarques sur la méthode de Newton

- Remarquons que la k ème itération de l'algorithme peut se résoudre en deux sous-étapes :

- 1 Résoudre $D^2f(\mathbf{x}^{(k)})\mathbf{d}^{(k)} = -\mathbf{g}^{(k)}$

- 2 Calculer $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mathbf{d}^{(k)}$

- La 1ère sous-étape correspond à la résolution d'un système à n équations à n inconnues ($\mathbf{d}^{(k)}$)

Remarques sur la méthode de Newton

- Remarquons que la k ème itération de l'algorithme peut se résoudre en deux sous-étapes :
 - 1 Résoudre $D^2f(\mathbf{x}^{(k)})\mathbf{d}^{(k)} = -\mathbf{g}^{(k)}$
 - 2 Calculer $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mathbf{d}^{(k)}$
- La 1ère sous-étape correspond à la résolution d'un système à n équations à n inconnues ($\mathbf{d}^{(k)}$)
- Ainsi, un algorithme performant de résolution de systèmes d'équations linéaires est essentiel pour la méthode de Newton

Vitesse de convergence de la méthode de Newton

- Comme pour le cas monovariante, il n'est pas garanti que la méthode de Newton permette de déterminer la bonne direction de descente si $D^2f(\mathbf{x}^{(k)})$ n'est pas définie positive
- Si $\mathbf{x}^{(0)}$ est très loin du minimiseur, même si $D^2f(\mathbf{x}^{(k)}) > 0$ il est possible que la méthode n'aboutisse pas à une descente et dans ce cas, on peut avoir $f(\mathbf{x}^{(k+1)}) > f(\mathbf{x}^{(k)})$

Vitesse de convergence de la méthode de Newton

- Comme pour le cas monovariante, il n'est pas garanti que la méthode de Newton permette de déterminer la bonne direction de descente si $D^2f(\mathbf{x}^{(k)})$ n'est pas définie positive

- Comme pour le cas monovariante, il n'est pas garanti que la méthode de Newton permette de déterminer la bonne direction de descente si $D^2f(\mathbf{x}^{(k)})$ n'est pas définie positive
- Si $\mathbf{x}^{(0)}$ est très loin du minimiseur, même si $D^2f(\mathbf{x}^{(k)}) > 0$ il est possible que la méthode n'aboutisse pas à une descente et dans ce cas, on peut avoir $f(\mathbf{x}^{(k+1)}) > f(\mathbf{x}^{(k)})$
- Malgré ces inconvénients, la méthode de Newton a une meilleure vitesse de convergence que les méthodes utilisant la dérivée première uniquement (à condition que le point initial soit relativement proche du minimiseur)

Vitesse de convergence de la méthode de Newton (suite)

- Dans le cas particulier où $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^t\mathbf{Q}\mathbf{x} - \mathbf{b}^t\mathbf{x}$ est une fonction quadratique, nous avons :

$$\mathbf{g}(\mathbf{x}) = \nabla f(\mathbf{x}) = \mathbf{Q}\mathbf{x} - \mathbf{b}$$

et

$$D^2f(\mathbf{x}) = \mathbf{Q}$$

Vitesse de convergence de la méthode de Newton (suite)

- Dans le cas particulier où $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^t\mathbf{Q}\mathbf{x} - \mathbf{b}^t\mathbf{x}$ est une fonction quadratique, nous avons :

$$\mathbf{g}(\mathbf{x}) = \nabla f(\mathbf{x}) = \mathbf{Q}\mathbf{x} - \mathbf{b}$$

et

$$D^2f(\mathbf{x}) = \mathbf{Q}$$

- Soit $\mathbf{x}^{(0)}$ un point initial, en appliquant la méthode de Newton on obtient :

$$\mathbf{x}^{(1)} =$$

Vitesse de convergence de la méthode de Newton (suite)

- Dans le cas particulier où $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^t\mathbf{Q}\mathbf{x} - \mathbf{b}^t\mathbf{x}$ est une fonction quadratique, nous avons :

$$\mathbf{g}(\mathbf{x}) = \nabla f(\mathbf{x}) = \mathbf{Q}\mathbf{x} - \mathbf{b}$$

et

$$D^2f(\mathbf{x}) = \mathbf{Q}$$

Vitesse de convergence de la méthode de Newton (suite)

- Dans le cas particulier où $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^t\mathbf{Q}\mathbf{x} - \mathbf{b}^t\mathbf{x}$ est une fonction quadratique, nous avons :

$$\mathbf{g}(\mathbf{x}) = \nabla f(\mathbf{x}) = \mathbf{Q}\mathbf{x} - \mathbf{b}$$

et

$$D^2f(\mathbf{x}) = \mathbf{Q}$$

- Soit $\mathbf{x}^{(0)}$ un point initial, en appliquant la méthode de Newton on obtient :

$$\begin{aligned} \mathbf{x}^{(1)} &= \mathbf{x}^{(0)} - D^2f(\mathbf{x}^{(0)})^{-1}\mathbf{g}^{(0)} \\ &= \mathbf{x}^{(0)} - \mathbf{Q}^{-1}(\mathbf{Q}\mathbf{x}^{(0)} - \mathbf{b}) \\ &= \mathbf{Q}^{-1}\mathbf{b} \\ &= \mathbf{x}^* \end{aligned}$$

Vitesse de convergence de la méthode de Newton (suite)

- Dans le cas particulier où $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^t\mathbf{Q}\mathbf{x} - \mathbf{b}^t\mathbf{x}$ est une fonction quadratique, nous avons :

$$\mathbf{g}(\mathbf{x}) = \nabla f(\mathbf{x}) = \mathbf{Q}\mathbf{x} - \mathbf{b}$$

et

$$D^2f(\mathbf{x}) = \mathbf{Q}$$

- Soit $\mathbf{x}^{(0)}$ un point initial, en appliquant la méthode de Newton on obtient :

$$\begin{aligned}\mathbf{x}^{(1)} &= \mathbf{x}^{(0)} - D^2f(\mathbf{x}^{(0)})^{-1}\mathbf{g}^{(0)} \\ &= \mathbf{x}^{(0)} - \mathbf{Q}^{-1}(\mathbf{Q}\mathbf{x}^{(0)} - \mathbf{b}) \\ &= \mathbf{Q}^{-1}\mathbf{b} \\ &= \mathbf{x}^*\end{aligned}$$

- Donc pour une fonction objectif quadratique la méthode trouve le minimiseur en une itération

Autres méthodes de descente

- Il est possible de modifier la méthode de Newton afin de traiter les cas où le point initial est très loin du minimiseur

Vitesse de convergence de la méthode de Newton (suite)

Théorème.

Soit $f \in \mathcal{C}^3$, et $\mathbf{x}^* \in \mathbb{R}^n$ un point tel que $\nabla f(\mathbf{x}^*) = \mathbf{0}$ et $D^2f(\mathbf{x}^*)$ soit inversible. Alors, pour tout point initial \mathbf{x}_0 suffisamment proche de \mathbf{x}^* , la méthode de Newton est bien définie pour tout k et produit une suite $\{\mathbf{x}^{(k)}\}$ convergeant vers \mathbf{x}^* avec un ordre de convergence au moins égale à 2

Autres méthodes de descente

- Il est possible de modifier la méthode de Newton afin de traiter les cas où le point initial est très loin du minimiseur
- La méthode de Newton détermine une direction de descente à condition que $D^2f(\mathbf{x}^{(k)}) > 0$. Si $D^2f(\mathbf{x}^{(k)}) \leq 0$, la méthode de **Levenberg-Marquardt** propose une modification de l'algorithme de Newton permettant de déterminer une direction de descente

Autres méthodes de descente

- Il est possible de modifier la méthode de Newton afin de traiter les cas où le point initial est très loin du minimiseur
- La méthode de Newton détermine une direction de descente à condition que $D^2f(\mathbf{x}^{(k)}) > 0$. Si $D^2f(\mathbf{x}^{(k)}) \leq 0$, la méthode de **Levenberg-Marquardt** propose une modification de l'algorithme de Newton permettant de déterminer une direction de descente
- La méthode dite du **gradient conjugué** est une méthode de descente à pas optimal mais pour laquelle la direction n'est pas exactement le gradient. C'est une méthode utilisant la dérivée première uniquement. Lorsque la fonction est quadratique et que n est grand, la méthode permet de déterminer le minimiseur en au plus n itérations

Rappel du Sommaire

- 1 Rappels de concepts mathématiques
- 2 Bases théoriques de l'optimisation
- 3 Optimisation sans contrainte
- 4 Optimisation linéaire avec contraintes
- 5 Optimisation non linéaire avec contraintes
- 6 Optimisation convexe
- 7 Algorithmes pour problèmes contraints

Autres méthodes de descente

- Il est possible de modifier la méthode de Newton afin de traiter les cas où le point initial est très loin du minimiseur
- La méthode de Newton détermine une direction de descente à condition que $D^2f(\mathbf{x}^{(k)}) > 0$. Si $D^2f(\mathbf{x}^{(k)}) \leq 0$, la méthode de **Levenberg-Marquardt** propose une modification de l'algorithme de Newton permettant de déterminer une direction de descente
- La méthode dite du **gradient conjugué** est une méthode de descente à pas optimal mais pour laquelle la direction n'est pas exactement le gradient. C'est une méthode utilisant la dérivée première uniquement. Lorsque la fonction est quadratique et que n est grand, la méthode permet de déterminer le minimiseur en au plus n itérations
- Pour les problèmes de grande taille, il est très coûteux de calculer et d'inverser le hessien et donc d'utiliser la méthode de Newton. Les méthodes dites de **quasi-Newton** proposent dans ce cas d'approximer $D^2f(\mathbf{x}^{(k)})^{-1}$ (approche DFP ou BFGS). Il s'agit de méthodes "utilisant" des dérivées secondes (stockage d'une matrice $(n \times n)$).

Rappel du Sommaire

- 1 Rappels de concepts mathématiques
- 2 Bases théoriques de l'optimisation
- 3 Optimisation sans contrainte
- 4 Optimisation linéaire avec contraintes
 - Programmation linéaire
 - Méthode du simplexe
 - Solution de base réalisable initiale et cas dégénéré
 - Dualité
- 5 Optimisation non linéaire avec contraintes

Introduction

- Un **programme linéaire (PL)** est un cas particulier d'un problème d'optimisation avec contraintes dans lequel il s'agit de déterminer les valeurs de variables de décision ($\mathbf{x} \in \mathbb{R}^n$) qui :
 - ▶ optimisent la valeur d'une **fonction objectif** $f : \mathbb{R}^n \rightarrow \mathbb{R}$ qui est **linéaire**
 - ▶ satisfont à des **contraintes** d'égalités $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$ ou d'inégalités $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ qui sont des **applications linéaires**

Introduction

- Un **programme linéaire (PL)** est un cas particulier d'un problème d'optimisation avec contraintes dans lequel il s'agit de déterminer les valeurs de variables de décision ($\mathbf{x} \in \mathbb{R}^n$) qui :
 - ▶ optimisent la valeur d'une **fonction objectif** $f : \mathbb{R}^n \rightarrow \mathbb{R}$ qui est **linéaire**
 - ▶ satisfont à des **contraintes** d'égalités $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$ ou d'inégalités $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ qui sont des **applications linéaires**
- Un point $\mathbf{x} \in \mathbb{R}^n$ qui satisfait à toutes les contraintes est appelé une **solution réalisable**
- En général le nombre de solutions réalisables est infiniment grand mais on verra que pour le cas des PL on peut chercher la solution dans un ensemble fini de points

Introduction

- Un **programme linéaire (PL)** est un cas particulier d'un problème d'optimisation avec contraintes dans lequel il s'agit de déterminer les valeurs de variables de décision ($\mathbf{x} \in \mathbb{R}^n$) qui :
 - ▶ optimisent la valeur d'une **fonction objectif** $f : \mathbb{R}^n \rightarrow \mathbb{R}$ qui est **linéaire**
 - ▶ satisfont à des **contraintes** d'égalités $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$ ou d'inégalités $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ qui sont des **applications linéaires**
- Un point $\mathbf{x} \in \mathbb{R}^n$ qui satisfait à toutes les contraintes est appelé une **solution réalisable**

Introduction

- Un **programme linéaire (PL)** est un cas particulier d'un problème d'optimisation avec contraintes dans lequel il s'agit de déterminer les valeurs de variables de décision ($\mathbf{x} \in \mathbb{R}^n$) qui :
 - ▶ optimisent la valeur d'une **fonction objectif** $f : \mathbb{R}^n \rightarrow \mathbb{R}$ qui est **linéaire**
 - ▶ satisfont à des **contraintes** d'égalités $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$ ou d'inégalités $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ qui sont des **applications linéaires**
- Un point $\mathbf{x} \in \mathbb{R}^n$ qui satisfait à toutes les contraintes est appelé une **solution réalisable**
- En général le nombre de solutions réalisables est infiniment grand mais on verra que pour le cas des PL on peut chercher la solution dans un ensemble fini de points
- En particulier, nous verrons la **méthode du simplexe** développée par G. Dantzig

Introduction

- Un **programme linéaire (PL)** est un cas particulier d'un problème d'optimisation avec contraintes dans lequel il s'agit de déterminer les valeurs de variables de décision ($\mathbf{x} \in \mathbb{R}^n$) qui :
 - ▶ optimisent la valeur d'une **fonction objectif** $f : \mathbb{R}^n \rightarrow \mathbb{R}$ qui est **linéaire**
 - ▶ satisfont à des **contraintes** d'égalités $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$ ou d'inégalités $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ qui sont des **applications linéaires**
- Un point $\mathbf{x} \in \mathbb{R}^n$ qui satisfait à toutes les contraintes est appelé une **solution réalisable**
- En général le nombre de solutions réalisables est infiniment grand mais on verra que pour le cas des PL on peut chercher la solution dans un ensemble fini de points
- En particulier, nous verrons la **méthode du simplexe** développée par G. Dantzig
- D'autres méthodes existent comme celle des **points intérieurs** proposées par N. Karmarkar

Formalisation d'un PL

- Formellement un PL peut être exprimé de la façon suivante :

$$\begin{array}{l} \min \mathbf{c}^t \mathbf{x} \\ \text{s/c } \mathbf{Ax} = \mathbf{b} \\ \mathbf{x} \geq \mathbf{0} \end{array}$$

où $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{A} \in \mathbb{R}^{m \times n}$

- \mathbf{c} est souvent appelé vecteur des coûts

Formalisation d'un PL

- Formellement un PL peut être exprimé de la façon suivante :

$$\begin{array}{l} \min \mathbf{c}^t \mathbf{x} \\ \text{s/c } \mathbf{Ax} = \mathbf{b} \\ \mathbf{x} \geq \mathbf{0} \end{array}$$

où $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{A} \in \mathbb{R}^{m \times n}$

Formalisation d'un PL

- Formellement un PL peut être exprimé de la façon suivante :

$$\begin{array}{l} \min \mathbf{c}^t \mathbf{x} \\ \text{s/c } \mathbf{Ax} = \mathbf{b} \\ \mathbf{x} \geq \mathbf{0} \end{array}$$

où $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{A} \in \mathbb{R}^{m \times n}$

- \mathbf{c} est souvent appelé vecteur des coûts
- Plusieurs variations sont possibles : problème de maximisation, $\mathbf{Ax} \leq \mathbf{b}$ ou $\mathbf{Ax} \geq \mathbf{b}$

Formalisation d'un PL

- Formellement un PL peut être exprimé de la façon suivante :

$$\begin{aligned} \min \quad & \mathbf{c}^t \mathbf{x} \\ \text{s/c} \quad & \mathbf{Ax} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{aligned}$$

où $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{A} \in \mathbb{R}^{m \times n}$

- \mathbf{c} est souvent appelé vecteur des coûts
- Plusieurs variations sont possibles : problème de maximisation, $\mathbf{Ax} \leq \mathbf{b}$ ou $\mathbf{Ax} \geq \mathbf{b}$
- On montrera que toutes ces variations peuvent se ramener à la forme ci-dessus dite **forme standard**

Exemple d'un PL (suite)

- Dénotons x_1 et x_2 les quantités produites des deux biens.
- Déterminer la fonction objectif :

Exemple d'un PL

- Une entreprise fabrique deux biens (des pièces mécaniques par exemple). Ces fabrications nécessitent l'utilisation de deux ateliers dont les capacités de production exprimées en heures d'usinage sont de 12. Supposons que :
 - Chaque unité du 1er produit nécessite 2h d'usinage dans l'atelier 1 et 1h dans l'atelier 2
 - Chaque unité du 2ème produit nécessite 1h d'usinage dans l'atelier 1 et 2h dans l'atelier 2
- Sachant que la marge sur le 1er produit est $p_1 = 4$ et que celle sur le 2ème produit est de $p_2 = 3$, déterminer un programme mathématique qui modélise le problème de l'optimisation de la marge de l'entreprise sous les contraintes de production décrites précédemment

Exemple d'un PL (suite)

- Dénotons x_1 et x_2 les quantités produites des deux biens.
- Déterminer la fonction objectif :

$$\max f(x_1, x_2) = 4x_1 + 3x_2$$

- Ecrire les contraintes de production :

Exemple d'un PL (suite)

- Dénotez x_1 et x_2 les quantités produites des deux biens.
- Déterminer la fonction objectif :

$$\max f(x_1, x_2) = 4x_1 + 3x_2$$

- Ecrire les contraintes de production :

$$\text{Sous les contraintes : } \begin{cases} 2x_1 + x_2 \leq 12 \\ x_1 + 2x_2 \leq 12 \\ x_1 \geq 0, x_2 \geq 0. \end{cases}$$

PL de dimension 2

- Beaucoup de concepts fondamentaux en PL peuvent être facilement illustrés par des problèmes de dimension 2

Exemple d'un PL (suite)

- Dénotez x_1 et x_2 les quantités produites des deux biens.
- Déterminer la fonction objectif :

$$\max f(x_1, x_2) = 4x_1 + 3x_2$$

- Ecrire les contraintes de production :

$$\text{Sous les contraintes : } \begin{cases} 2x_1 + x_2 \leq 12 \\ x_1 + 2x_2 \leq 12 \\ x_1 \geq 0, x_2 \geq 0. \end{cases}$$

- En reprenant les notations précédentes on a :

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \mathbf{c} = \begin{pmatrix} 4 \\ 3 \end{pmatrix}, \mathbf{A} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 12 \\ 12 \end{pmatrix} \text{ et } \mathbf{Ax} \leq \mathbf{b}$$

PL de dimension 2

- Beaucoup de concepts fondamentaux en PL peuvent être facilement illustrés par des problèmes de dimension 2
- Reprenons l'exemple précédent et considérons les contraintes du problème exprimées par :

$$\mathbf{Ax} \leq \mathbf{b} \Leftrightarrow \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \leq \begin{pmatrix} 12 \\ 12 \end{pmatrix} \Leftrightarrow \begin{cases} 2x_1 + x_2 \leq 12 \\ x_1 + 2x_2 \leq 12 \end{cases}$$

PL de dimension 2

- Beaucoup de concepts fondamentaux en PL peuvent être facilement illustrés par des problèmes de dimension 2
- Reprenons l'exemple précédent et considérons les contraintes du problème exprimées par :

$$\mathbf{Ax} \leq \mathbf{b} \Leftrightarrow \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \leq \begin{pmatrix} 12 \\ 12 \end{pmatrix} \Leftrightarrow \begin{cases} 2x_1 + x_2 \leq 12 \\ x_1 + 2x_2 \leq 12 \end{cases}$$

- Remarquons que chaque inégalité représente un demi-espace et que l'espace réalisable \mathbb{S} qui est donc une intersection de demi-espaces est ainsi un polyèdre noté \mathbb{O} (de dimension 2 ici)

PL de dimension 2 (suite)

- L'ensemble réalisable est donc l'intersection des contraintes suivantes :

$$\begin{cases} x_2 \leq 12 - 2x_1 \\ x_2 \leq \frac{1}{2}(12 - x_1) \\ x_1 \geq 0 \\ x_2 \geq 0 \end{cases}$$

- Exemple : le point $\mathbf{x} = (1, 2)$ est un point réalisable puisqu'il satisfait à toutes les contraintes

PL de dimension 2 (suite)

- L'ensemble réalisable est donc l'intersection des contraintes suivantes :

$$\begin{cases} x_2 \leq 12 - 2x_1 \\ x_2 \leq \frac{1}{2}(12 - x_1) \\ x_1 \geq 0 \\ x_2 \geq 0 \end{cases}$$

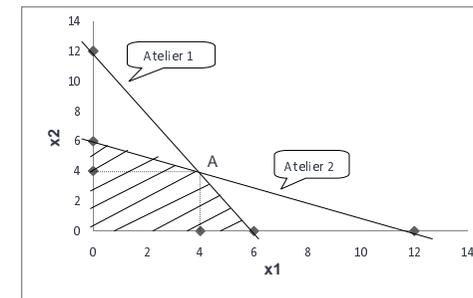
PL de dimension 2 (suite)

- L'ensemble réalisable est donc l'intersection des contraintes suivantes :

$$\begin{cases} x_2 \leq 12 - 2x_1 \\ x_2 \leq \frac{1}{2}(12 - x_1) \\ x_1 \geq 0 \\ x_2 \geq 0 \end{cases}$$

- Exemple : le point $\mathbf{x} = (1, 2)$ est un point réalisable puisqu'il satisfait à toutes les contraintes

- Illustration :



PL de dimension 2 (suite)

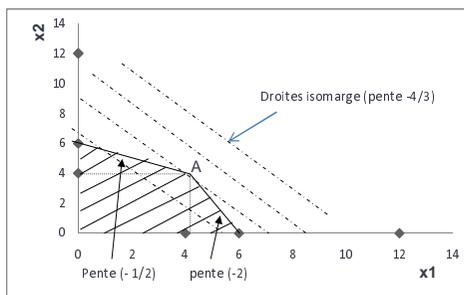
- Considérons à présent la fonction objectif suivante qu'il faut maximiser : $f(\mathbf{x}) = \mathbf{c}^t \mathbf{x} = 4x_1 + 3x_2$

PL de dimension 2 (suite)

- Considérons à présent la fonction objectif suivante qu'il faut maximiser : $f(\mathbf{x}) = \mathbf{c}^t \mathbf{x} = 4x_1 + 3x_2$
- Remarquons que la famille $\{\mathbf{c}^t \mathbf{x} = 4x_1 + 3x_2 = z, z \in \mathbb{R}\}$ forme un ensemble de droites (droites iso-marges en économie) parallèles dans \mathbb{R}^2 . Par exemple :
 - si $z = 1 : 4x_1 + 3x_2 = 1 \Leftrightarrow x_2 = 1 - \frac{4}{3}x_1$
 - si $z = 5 : 4x_1 + 3x_2 = 5 \Leftrightarrow x_2 = 5 - \frac{4}{3}x_1$
 - ...

PL de dimension 2 (suite)

- Considérons à présent la fonction objectif suivante qu'il faut maximiser : $f(\mathbf{x}) = \mathbf{c}^t \mathbf{x} = 4x_1 + 3x_2$
- Remarquons que la famille $\{\mathbf{c}^t \mathbf{x} = 4x_1 + 3x_2 = z, z \in \mathbb{R}\}$ forme un ensemble de droites (droites iso-marges en économie) parallèles dans \mathbb{R}^2 . Par exemple :
 - si $z = 1 : 4x_1 + 3x_2 = 1 \Leftrightarrow x_2 = 1 - \frac{4}{3}x_1$
 - si $z = 5 : 4x_1 + 3x_2 = 5 \Leftrightarrow x_2 = 5 - \frac{4}{3}x_1$
 - ...
- Illustration :



PL de dimension 2 (suite)

- D'un point de vue géométrique, maximiser $\mathbf{c}^t \mathbf{x}$ revient à déterminer la droite iso-marge qui a une intersection non vide avec l'ensemble réalisable (hachurée) et qui conduit à la plus forte valeur $z^* = \mathbf{c}^t \mathbf{x}^*$

PL de dimension 2 (suite)

- D'un point de vue géométrique, maximiser $\mathbf{c}^t \mathbf{x}$ revient à déterminer la droite iso-marge qui a une intersection non vide avec l'ensemble réalisable (hachurée) et qui conduit à la plus forte valeur $z^* = \mathbf{c}^t \mathbf{x}^*$
- Les coordonnées du point \mathbf{x}^* situé à l'intersection entre l'ensemble réalisable et la droite iso-marge de valeur z^* est alors la solution du PL

Polyèdre et PL

- Supposons pour l'instant que les contraintes sont de la forme :
 $\mathbf{Ax} \leq \mathbf{b}$

PL de dimension 2 (suite)

- D'un point de vue géométrique, maximiser $\mathbf{c}^t \mathbf{x}$ revient à déterminer la droite iso-marge qui a une intersection non vide avec l'ensemble réalisable (hachurée) et qui conduit à la plus forte valeur $z^* = \mathbf{c}^t \mathbf{x}^*$
- Les coordonnées du point \mathbf{x}^* situé à l'intersection entre l'ensemble réalisable et la droite iso-marge de valeur z^* est alors la solution du PL
- Dans l'exemple précédent $\mathbf{x}^* = (4, 4)$ est le maximiseur de f et on obtient $f(\mathbf{x}^*) = z^* = 28$

Polyèdre et PL

- Supposons pour l'instant que les contraintes sont de la forme :
 $\mathbf{Ax} \leq \mathbf{b}$
- L'ensemble des points de \mathbb{R}^n satisfaisant $\mathbf{Ax} \leq \mathbf{b}$ est un polyèdre (que l'on suppose non-vidé et borné) noté \mathbb{O}

Polyèdre et PL

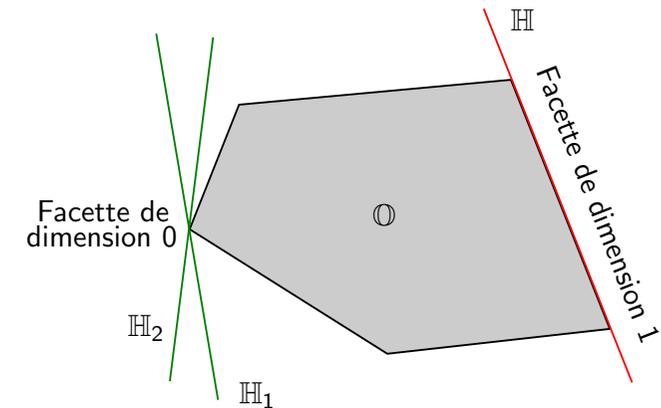
- Supposons pour l'instant que les contraintes sont de la forme :
 $\mathbf{Ax} \leq \mathbf{b}$
- L'ensemble des points de \mathbb{R}^n satisfaisant $\mathbf{Ax} \leq \mathbf{b}$ est un polyèdre (que l'on suppose non-vide et borné) noté \mathbb{O}
- Soit \mathbb{H} un hyperplan de \mathbb{R}^n support de \mathbb{O} :
 - ▶ Si $\dim(\mathbb{O}) < n$ alors l'ensemble des points communs entre \mathbb{O} et \mathbb{H} est \mathbb{O}
 - ▶ Si $\dim(\mathbb{O}) = n$ alors l'ensemble des points communs entre \mathbb{O} et \mathbb{H} est une facette de \mathbb{O} :
 - ★ Si cette facette est de dimension $n - 1$ alors il existe un seul hyperplan support qui est la facette elle-même
 - ★ Si cette facette est de dimension inférieure à $n - 1$ alors il existe une infinité d'hyperplan support

Polyèdre et PL (suite)

- La PL consiste à maximiser une fonction linéaire $\mathbf{c}^t \mathbf{x}$ sur un polyèdre \mathbb{O}

Polyèdre et PL (suite)

- Illustration dans \mathbb{R}^2



Polyèdre et PL (suite)

- La PL consiste à maximiser une fonction linéaire $\mathbf{c}^t \mathbf{x}$ sur un polyèdre \mathbb{O}
- Notons $\tilde{\mathbb{H}}_z$ l'hyperplan d'équation $\mathbf{c}^t \mathbf{x} = z$ et dont le vecteur \mathbf{c} est le vecteur normal

Polyèdre et PL (suite)

- La PL consiste à maximiser une fonction linéaire $\mathbf{c}^t \mathbf{x}$ sur un polyèdre \mathbb{O}
- Notons $\tilde{\mathbb{H}}_z$ l'hyperplan d'équation $\mathbf{c}^t \mathbf{x} = z$ et dont le vecteur \mathbf{c} est le vecteur normal
- $\tilde{\mathbb{H}}_0 = \{\mathbf{x} \in \mathbb{R} : \mathbf{c}^t \mathbf{x} = 0\}$ est ainsi l'hyperplan passant par l'origine

Polyèdre et PL (suite)

- La PL consiste à maximiser une fonction linéaire $\mathbf{c}^t \mathbf{x}$ sur un polyèdre \mathbb{O}
- Notons $\tilde{\mathbb{H}}_z$ l'hyperplan d'équation $\mathbf{c}^t \mathbf{x} = z$ et dont le vecteur \mathbf{c} est le vecteur normal
- $\tilde{\mathbb{H}}_0 = \{\mathbf{x} \in \mathbb{R} : \mathbf{c}^t \mathbf{x} = 0\}$ est ainsi l'hyperplan passant par l'origine
- Soit $\tilde{\mathbb{H}}_{z'}$ l'hyperplan parallèle à $\tilde{\mathbb{H}}_0$ tel que le vecteur normal \mathbf{c} peut être positionné dans le demi-espace ne contenant pas \mathbb{O}
- Nous avons par définition la propriété que $\forall \mathbf{x} \in \mathbb{O} : \mathbf{c}^t \mathbf{x} \leq z'$

Polyèdre et PL (suite)

- La PL consiste à maximiser une fonction linéaire $\mathbf{c}^t \mathbf{x}$ sur un polyèdre \mathbb{O}
- Notons $\tilde{\mathbb{H}}_z$ l'hyperplan d'équation $\mathbf{c}^t \mathbf{x} = z$ et dont le vecteur \mathbf{c} est le vecteur normal
- $\tilde{\mathbb{H}}_0 = \{\mathbf{x} \in \mathbb{R} : \mathbf{c}^t \mathbf{x} = 0\}$ est ainsi l'hyperplan passant par l'origine
- Soit $\tilde{\mathbb{H}}_{z'}$ l'hyperplan parallèle à $\tilde{\mathbb{H}}_0$ tel que le vecteur normal \mathbf{c} peut être positionné dans le demi-espace ne contenant pas \mathbb{O}

Polyèdre et PL (suite)

- La PL consiste à maximiser une fonction linéaire $\mathbf{c}^t \mathbf{x}$ sur un polyèdre \mathbb{O}
- Notons $\tilde{\mathbb{H}}_z$ l'hyperplan d'équation $\mathbf{c}^t \mathbf{x} = z$ et dont le vecteur \mathbf{c} est le vecteur normal
- $\tilde{\mathbb{H}}_0 = \{\mathbf{x} \in \mathbb{R} : \mathbf{c}^t \mathbf{x} = 0\}$ est ainsi l'hyperplan passant par l'origine
- Soit $\tilde{\mathbb{H}}_{z'}$ l'hyperplan parallèle à $\tilde{\mathbb{H}}_0$ tel que le vecteur normal \mathbf{c} peut être positionné dans le demi-espace ne contenant pas \mathbb{O}
- Nous avons par définition la propriété que $\forall \mathbf{x} \in \mathbb{O} : \mathbf{c}^t \mathbf{x} \leq z'$
- Dénotons à présent par $\tilde{\mathbb{O}}_{z'}$ le polyèdre résultant de l'intersection entre \mathbb{O} et $\tilde{\mathbb{H}}_{z'}$. Nous avons alors les propriétés suivantes :

Polyèdre et PL (suite)

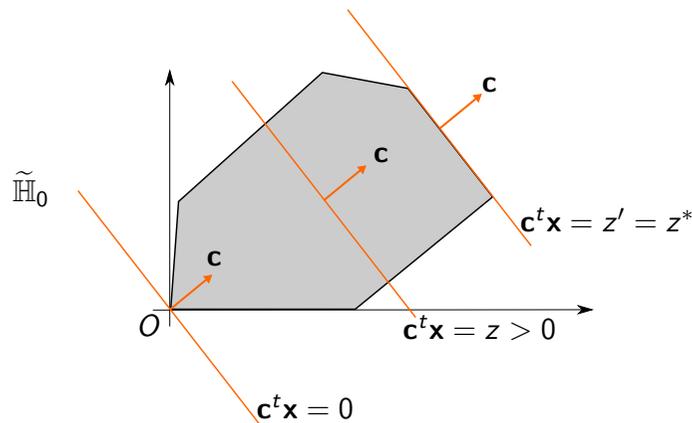
- La PL consiste à maximiser une fonction linéaire $\mathbf{c}^t \mathbf{x}$ sur un polyèdre \mathbb{O}
- Notons $\tilde{\mathbb{H}}_z$ l'hyperplan d'équation $\mathbf{c}^t \mathbf{x} = z$ et dont le vecteur \mathbf{c} est le vecteur normal
- $\tilde{\mathbb{H}}_0 = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{c}^t \mathbf{x} = 0\}$ est ainsi l'hyperplan passant par l'origine
- Soit $\tilde{\mathbb{H}}_{z'}$ l'hyperplan parallèle à $\tilde{\mathbb{H}}_0$ tel que le vecteur normal \mathbf{c} peut être positionné dans le demi-espace ne contenant pas \mathbb{O}
- Nous avons par définition la propriété que $\forall \mathbf{x} \in \mathbb{O} : \mathbf{c}^t \mathbf{x} \leq z'$
- Dénotons à présent par $\tilde{\mathbb{O}}_{z'}$ le polyèdre résultant de l'intersection entre \mathbb{O} et $\tilde{\mathbb{H}}_{z'}$. Nous avons alors les propriétés suivantes :
 - ▶ $\forall \mathbf{x} \in \tilde{\mathbb{O}}_{z'} : \mathbf{c}^t \mathbf{x} = z'$
 - ▶ $z' = z^*$ est la valeur maximale du PL
 - ▶ $\tilde{\mathbb{O}}_{z'}$ est l'ensemble des maximiseurs de f

Polyèdre et PL (suite)

- La PL consiste à maximiser une fonction linéaire $\mathbf{c}^t \mathbf{x}$ sur un polyèdre \mathbb{O}
- Notons $\tilde{\mathbb{H}}_z$ l'hyperplan d'équation $\mathbf{c}^t \mathbf{x} = z$ et dont le vecteur \mathbf{c} est le vecteur normal
- $\tilde{\mathbb{H}}_0 = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{c}^t \mathbf{x} = 0\}$ est ainsi l'hyperplan passant par l'origine
- Soit $\tilde{\mathbb{H}}_{z'}$ l'hyperplan parallèle à $\tilde{\mathbb{H}}_0$ tel que le vecteur normal \mathbf{c} peut être positionné dans le demi-espace ne contenant pas \mathbb{O}
- Nous avons par définition la propriété que $\forall \mathbf{x} \in \mathbb{O} : \mathbf{c}^t \mathbf{x} \leq z'$
- Dénotons à présent par $\tilde{\mathbb{O}}_{z'}$ le polyèdre résultant de l'intersection entre \mathbb{O} et $\tilde{\mathbb{H}}_{z'}$. Nous avons alors les propriétés suivantes :
 - ▶ $\forall \mathbf{x} \in \tilde{\mathbb{O}}_{z'} : \mathbf{c}^t \mathbf{x} = z'$
 - ▶ $z' = z^*$ est la valeur maximale du PL
 - ▶ $\tilde{\mathbb{O}}_{z'}$ est l'ensemble des maximiseurs de f
- $\tilde{\mathbb{O}}_{z'}$ peut être restreint à un point et dans ce cas, le maximiseur est unique et il s'agit d'un point extrême de \mathbb{O}

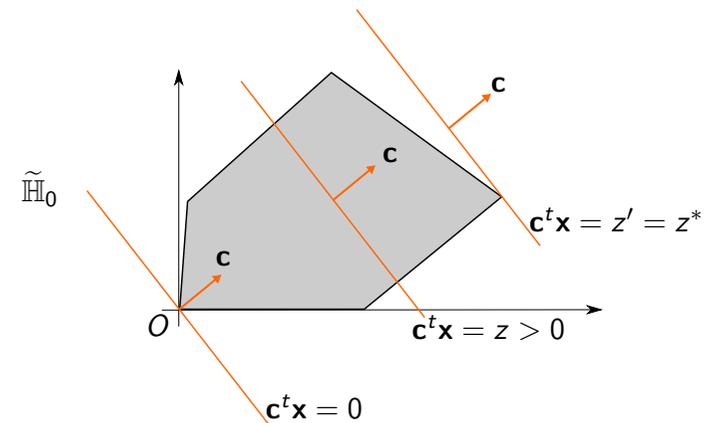
Polyèdre et PL (suite)

- Illustration dans \mathbb{R}^2 (cas général)



Polyèdre et PL (suite)

- Illustration dans \mathbb{R}^2 (cas où $\tilde{\mathbb{O}}_{z^*}$ est une facette de dimension 0)



Forme standard d'un PL

- Nous avons précédemment étudié les structures géométriques sous-jacentes à un ensemble de contraintes linéaires exprimées sous formes d'inégalités \leq

Forme standard d'un PL

- Nous avons précédemment étudié les structures géométriques sous-jacentes à un ensemble de contraintes linéaires exprimées sous formes d'inégalités \leq
- Nous introduisons maintenant une forme canonique dite **forme standard** en fonction de laquelle nous allons exprimer tout PL

Définition. (Forme standard d'un PL)

Un PL est sous forme standard ssi il est exprimé de la façon suivante :

$$\begin{aligned} \min f(\mathbf{x}) &= \mathbf{c}^t \mathbf{x} \\ \text{s/c } \mathbf{Ax} &= \mathbf{b} \\ \mathbf{x} &\geq \mathbf{0} \end{aligned}$$

où \mathbf{A} est une matrice réelle ($m \times n$), $m < n$, $\text{rg}(\mathbf{A}) = m$ et $\mathbf{b} \geq \mathbf{0}$

Forme standard d'un PL

- Nous avons précédemment étudié les structures géométriques sous-jacentes à un ensemble de contraintes linéaires exprimées sous formes d'inégalités \leq
- Nous introduisons maintenant une forme canonique dite **forme standard** en fonction de laquelle nous allons exprimer tout PL

Forme standard d'un PL

- Nous avons précédemment étudié les structures géométriques sous-jacentes à un ensemble de contraintes linéaires exprimées sous formes d'inégalités \leq
- Nous introduisons maintenant une forme canonique dite **forme standard** en fonction de laquelle nous allons exprimer tout PL

Définition. (Forme standard d'un PL)

Un PL est sous forme standard ssi il est exprimé de la façon suivante :

$$\begin{aligned} \min f(\mathbf{x}) &= \mathbf{c}^t \mathbf{x} \\ \text{s/c } \mathbf{Ax} &= \mathbf{b} \\ \mathbf{x} &\geq \mathbf{0} \end{aligned}$$

où \mathbf{A} est une matrice réelle ($m \times n$), $m < n$, $\text{rg}(\mathbf{A}) = m$ et $\mathbf{b} \geq \mathbf{0}$

- Les résultats que nous énonçons par la suite sont valables pour des PL mis sous forme standard. Nous allons d'abord montrer que nous pouvons toujours exprimer un PL quelconque sous forme standard

Forme standard d'un PL (suite)

- A la base les contraintes linéaires sont de la forme $\mathbf{Ax} \sim \mathbf{b}$ où \sim peut être l'un des trois symboles suivants : $\leq, \geq, =$

Forme standard d'un PL (suite)

- A la base les contraintes linéaires sont de la forme $\mathbf{Ax} \sim \mathbf{b}$ où \sim peut être l'un des trois symboles suivants : $\leq, \geq, =$

- De manière plus générale, nous pouvons exprimer chaque contrainte linéaire individuellement :

$$\forall i = 1, \dots, m : \sum_{j=1}^n a_{ij}x_j \sim b_i$$

- Pour mettre un PL sous forme standard il faut respecter le formalisme suivant :

- 1 Problème de minimisation
- 2 Conditions de non-négativité des contraintes :
 $\mathbf{b} \geq \mathbf{0} \Leftrightarrow \forall i = 1, \dots, m : b_i \geq 0$
- 3 Expression des contraintes linéaires par des contraintes d'égalités

Forme standard d'un PL (suite)

- A la base les contraintes linéaires sont de la forme $\mathbf{Ax} \sim \mathbf{b}$ où \sim peut être l'un des trois symboles suivants : $\leq, \geq, =$

- De manière plus générale, nous pouvons exprimer chaque contrainte linéaire individuellement :

$$\forall i = 1, \dots, m : \sum_{j=1}^n a_{ij}x_j \sim b_i$$

Forme standard d'un PL (Conditions de non-négativité des contraintes)

- 1 Le problème à étudier peut-être à la base un problème de maximisation de la fonction linéaire $f(\mathbf{x}) = \mathbf{c}^t \mathbf{x}$

Forme standard d'un PL (Conditions de non-négativité des contraintes)

- 1 Le problème à étudier peut-être à la base un problème de maximisation de la fonction linéaire $f(\mathbf{x}) = \mathbf{c}^t \mathbf{x}$
 \Rightarrow Pour se ramener à un problème de minimisation, il suffira de remplacer $f(\mathbf{x})$ par $-f(\mathbf{x})$ car minimiser $-f(\mathbf{x})$ est équivalent à maximiser $f(\mathbf{x})$

Forme standard d'un PL (Conditions de non-négativité des contraintes)

- 1 Le problème à étudier peut-être à la base un problème de maximisation de la fonction linéaire $f(\mathbf{x}) = \mathbf{c}^t \mathbf{x}$
 \Rightarrow Pour se ramener à un problème de minimisation, il suffira de remplacer $f(\mathbf{x})$ par $-f(\mathbf{x})$ car minimiser $-f(\mathbf{x})$ est équivalent à maximiser $f(\mathbf{x})$

- 2 Les contraintes linéaires sont à la base de la forme suivante :

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

où \sim représente l'un des signes suivants $\leq, \geq, =$.

Nous supposons dans la suite que les constantes b_i sont **non-négatives**, $\forall i = 1, \dots, m$.

Forme standard d'un PL (Conditions de non-négativité des contraintes)

- 1 Le problème à étudier peut-être à la base un problème de maximisation de la fonction linéaire $f(\mathbf{x}) = \mathbf{c}^t \mathbf{x}$
 \Rightarrow Pour se ramener à un problème de minimisation, il suffira de remplacer $f(\mathbf{x})$ par $-f(\mathbf{x})$ car minimiser $-f(\mathbf{x})$ est équivalent à maximiser $f(\mathbf{x})$

- 2 Les contraintes linéaires sont à la base de la forme suivante :

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

où \sim représente l'un des signes suivants $\leq, \geq, =$.

Nous supposons dans la suite que les constantes b_i sont **non-négatives**, $\forall i = 1, \dots, m$.

- \Rightarrow
- Si la
- i
- ème contrainte est telle que
- $b_i < 0$
- , pour se ramener au cas de la forme standard, il suffit de multiplier la
- i
- ème contrainte par
- -1

Forme standard d'un PL (Conditions de non-négativité des contraintes)

- 1 Le problème à étudier peut-être à la base un problème de maximisation de la fonction linéaire $f(\mathbf{x}) = \mathbf{c}^t \mathbf{x}$
 \Rightarrow Pour se ramener à un problème de minimisation, il suffira de remplacer $f(\mathbf{x})$ par $-f(\mathbf{x})$ car minimiser $-f(\mathbf{x})$ est équivalent à maximiser $f(\mathbf{x})$

- 2 Les contraintes linéaires sont à la base de la forme suivante :

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

où \sim représente l'un des signes suivants $\leq, \geq, =$.

Nous supposons dans la suite que les constantes b_i sont **non-négatives**, $\forall i = 1, \dots, m$.

- \Rightarrow
- Si la
- i
- ème contrainte est telle que
- $b_i < 0$
- , pour se ramener au cas de la forme standard, il suffit de multiplier la
- i
- ème contrainte par
- -1
-
- Exemple :
- $2x_1 - 3x_2 \leq -5 \Leftrightarrow -2x_1 + 3x_2 \geq 5$
- .

Forme standard d'un PL (Passage à des contraintes d'égalités)

3 Les contraintes linéaires peuvent être à la base de différentes natures :

$$\forall i = 1, \dots, m : \sum_{j=1}^n a_{ij}x_j \left\{ \begin{array}{l} \leq \\ = \\ \geq \end{array} \right\} b_i$$

⇒ Pour mettre le PL sous forme standard il faut se ramener uniquement à des contraintes d'égalités. On distingue alors deux cas :

a Cas où $\sum_{j=1}^n a_{ij}x_j \leq b_i$

⇒ Une contrainte linéaire de la forme $\sum a_{ij}x_j \leq b_i$ peut être convertie en une contrainte d'égalité en **ajoutant** une nouvelle **variable non-négative** au membre de gauche de l'inégalité. Une telle variable est appelée **variable d'écart** et est égale à la différence entre le membre de droite et celui de gauche

Exemple : $4x_1 + 3x_3 + 5x_4 \leq 300 \Leftrightarrow 4x_1 + 3x_3 + 5x_4 + x_5 = 300$

Forme standard d'un PL (Passage à des contraintes d'égalités - suite)

- Considérons l'inégalité $x_1 \leq 9$ et l'égalité $x_1 + x_2 = 9$ où $x_2 \geq 0$ est une variable d'écart

Forme standard d'un PL (Passage à des contraintes d'égalités - suite)

b Cas où $\sum_{j=1}^n a_{ij}x_j \geq b_i$

⇒ Une contrainte linéaire de la forme $\sum a_{ij}x_j \geq b_i$ peut être convertie en une contrainte d'égalité en **retranchant** une nouvelle variable non-négative au membre de gauche de l'inégalité. Une telle variable est appelée **variable de surplus** et est égale à la différence entre le membre de gauche et celui de droite

Exemple : $4x_1 + 6x_2 + x_3 \geq 54 \Leftrightarrow 4x_1 + 6x_2 + x_3 - x_4 = 54$

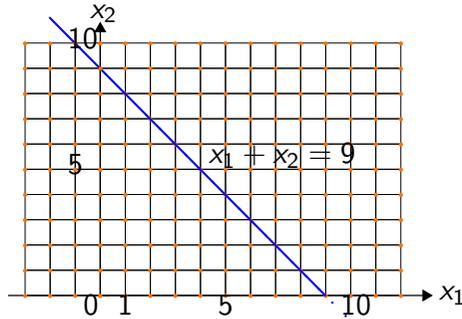
- A première vue, ajouter des variables d'écart ou retrancher des variables de surplus change la nature des contraintes mais en fait il n'en est rien. Nous illustrons cette propriété sur un exemple dans ce qui suit.

Forme standard d'un PL (Passage à des contraintes d'égalités - suite)

- Considérons l'inégalité $x_1 \leq 9$ et l'égalité $x_1 + x_2 = 9$ où $x_2 \geq 0$ est une variable d'écart
- Nous avons les ensembles suivants : $\mathbb{C}_1 = \{x_1 \in \mathbb{R} : x_1 \leq 9\}$ et $\mathbb{C}_2 = \{x_1 \in \mathbb{R} : x_1 + x_2 = 9; x_2 \geq 0\}$

Forme standard d'un PL (Passage à des contraintes d'égalités - suite)

- Considérons l'inégalité $x_1 \leq 9$ et l'égalité $x_1 + x_2 = 9$ où $x_2 \geq 0$ est une variable d'écart
- Nous avons les ensembles suivants : $\mathbb{C}_1 = \{x_1 \in \mathbb{R} : x_1 \leq 9\}$ et $\mathbb{C}_2 = \{x_1 \in \mathbb{R} : x_1 + x_2 = 9; x_2 \geq 0\}$
- Nous voyons en fait que $\mathbb{C}_1 = \mathbb{C}_2$:



Forme standard d'un PL (Exemple)

- Exemple, soit l'ensemble de contraintes suivant :

$$\begin{cases} -x_1 - 2x_2 \geq -3 \\ 4x_1 + 5x_2 \geq 6 \\ 7x_1 + 8x_2 = 15 \\ x_1, x_2 \geq 0 \end{cases}$$

Forme standard d'un PL (Exemple)

- Exemple, soit l'ensemble de contraintes suivant :

$$\begin{cases} -x_1 - 2x_2 \geq -3 \\ 4x_1 + 5x_2 \geq 6 \\ 7x_1 + 8x_2 = 15 \\ x_1, x_2 \geq 0 \end{cases}$$

- Afin d'avoir la condition de non-négativité sur **b** nous avons :

$$\begin{cases} x_1 + 2x_2 \leq 3 \\ 4x_1 + 5x_2 \geq 6 \\ 7x_1 + 8x_2 = 15 \\ x_1, x_2 \geq 0 \end{cases}$$

Forme standard d'un PL (Exemple)

- Exemple, soit l'ensemble de contraintes suivant :

$$\begin{cases} -x_1 - 2x_2 \geq -3 \\ 4x_1 + 5x_2 \geq 6 \\ 7x_1 + 8x_2 = 15 \\ x_1, x_2 \geq 0 \end{cases}$$

- Afin d'avoir la condition de non-négativité sur **b** nous avons :

$$\begin{cases} x_1 + 2x_2 \leq 3 \\ 4x_1 + 5x_2 \geq 6 \\ 7x_1 + 8x_2 = 15 \\ x_1, x_2 \geq 0 \end{cases}$$

- Afin d'avoir des contraintes d'égalités on ajoute les variables d'écart et de surplus pour obtenir :

$$\begin{cases} x_1 + 2x_2 + x_3 = 3 \\ 4x_1 + 5x_2 - x_4 = 6 \\ 7x_1 + 8x_2 = 15 \\ x_1, \dots, x_4 \geq 0 \end{cases}$$

Solutions réalisables

- Nous supposons dans la suite que les PL sont sous forme standard :

$$\begin{array}{l} \min f(\mathbf{x}) = \mathbf{c}^t \mathbf{x} \\ \text{slc } \mathbf{Ax} = \mathbf{b} \\ \mathbf{x} \geq \mathbf{0} \end{array}$$

où \mathbf{A} est une matrice réelle de taille $m \times n$ tel que $m < n$ et $\text{rg}(\mathbf{A}) = m$ et où $\mathbf{b} \geq \mathbf{0}$

Solutions réalisables

- Nous supposons dans la suite que les PL sont sous forme standard :

$$\begin{array}{l} \min f(\mathbf{x}) = \mathbf{c}^t \mathbf{x} \\ \text{slc } \mathbf{Ax} = \mathbf{b} \\ \mathbf{x} \geq \mathbf{0} \end{array}$$

où \mathbf{A} est une matrice réelle de taille $m \times n$ tel que $m < n$ et $\text{rg}(\mathbf{A}) = m$ et où $\mathbf{b} \geq \mathbf{0}$

- Considérons alors le système d'équations $\mathbf{Ax} = \mathbf{b}$ où $\text{rg}(\mathbf{A}) = m$
- Nous noterons la colonne j de \mathbf{A} par $\mathbf{a}_j = (a_{1j}, \dots, a_{mj})$

Solutions réalisables

- Nous supposons dans la suite que les PL sont sous forme standard :

$$\begin{array}{l} \min f(\mathbf{x}) = \mathbf{c}^t \mathbf{x} \\ \text{slc } \mathbf{Ax} = \mathbf{b} \\ \mathbf{x} \geq \mathbf{0} \end{array}$$

où \mathbf{A} est une matrice réelle de taille $m \times n$ tel que $m < n$ et $\text{rg}(\mathbf{A}) = m$ et où $\mathbf{b} \geq \mathbf{0}$

- Considérons alors le système d'équations $\mathbf{Ax} = \mathbf{b}$ où $\text{rg}(\mathbf{A}) = m$

Solutions réalisables

- Nous supposons dans la suite que les PL sont sous forme standard :

$$\begin{array}{l} \min f(\mathbf{x}) = \mathbf{c}^t \mathbf{x} \\ \text{slc } \mathbf{Ax} = \mathbf{b} \\ \mathbf{x} \geq \mathbf{0} \end{array}$$

où \mathbf{A} est une matrice réelle de taille $m \times n$ tel que $m < n$ et $\text{rg}(\mathbf{A}) = m$ et où $\mathbf{b} \geq \mathbf{0}$

- Considérons alors le système d'équations $\mathbf{Ax} = \mathbf{b}$ où $\text{rg}(\mathbf{A}) = m$
- Nous noterons la colonne j de \mathbf{A} par $\mathbf{a}_j = (a_{1j}, \dots, a_{mj})$
- Soit \mathbb{B} un sous-ensemble de m indices de $\{1, \dots, n\}$ et soit $\mathbf{A}_{\mathbb{B}} = (\mathbf{a}_{\mathbb{B}_1}, \dots, \mathbf{a}_{\mathbb{B}_m})$ (où \mathbb{B}_i est le i ème élément de \mathbb{B}) tel que ses colonnes soient linéairement indépendantes. Soit alors $\mathbb{D} = \{1, \dots, n\} \setminus \mathbb{B}$ le complémentaire de \mathbb{B}

Solutions réalisables

- Nous supposons dans la suite que les PL sont sous forme standard :

$$\begin{aligned} \min f(\mathbf{x}) &= \mathbf{c}^t \mathbf{x} \\ \text{s.t. } \mathbf{Ax} &= \mathbf{b} \\ \mathbf{x} &\geq \mathbf{0} \end{aligned}$$

où \mathbf{A} est une matrice réelle de taille $m \times n$ tel que $m < n$ et $\text{rg}(\mathbf{A}) = m$ et où $\mathbf{b} \geq \mathbf{0}$

- Considérons alors le système d'équations $\mathbf{Ax} = \mathbf{b}$ où $\text{rg}(\mathbf{A}) = m$
- Nous noterons la colonne j de \mathbf{A} par $\mathbf{a}_j = (a_{1j}, \dots, a_{mj})$
- Soit \mathbb{B} un sous-ensemble de m indices de $\{1, \dots, n\}$ et soit $\mathbf{A}_{\mathbb{B}} = (\mathbf{a}_{\mathbb{B}_1}, \dots, \mathbf{a}_{\mathbb{B}_m})$ (où \mathbb{B}_i est le i ème élément de \mathbb{B}) tel que ses colonnes soient linéairement indépendantes. Soit alors $\mathbb{D} = \{1, \dots, n\} \setminus \mathbb{B}$ le complémentaire de \mathbb{B}
- On peut réordonner les colonnes de \mathbf{A} de sorte à avoir : $\mathbf{A} = (\mathbf{A}_{\mathbb{B}}, \mathbf{A}_{\mathbb{D}})$

Solutions réalisables (suite)

- La matrice $\mathbf{A}_{\mathbb{B}}$ étant non singulière, nous pouvons donc résoudre le système $\mathbf{A}_{\mathbb{B}} \mathbf{x}_{\mathbb{B}} = \mathbf{b}$ et nous avons :

$$\mathbf{x}_{\mathbb{B}} = \mathbf{A}_{\mathbb{B}}^{-1} \mathbf{b}$$

Solutions réalisables

- Nous supposons dans la suite que les PL sont sous forme standard :

$$\begin{aligned} \min f(\mathbf{x}) &= \mathbf{c}^t \mathbf{x} \\ \text{s.t. } \mathbf{Ax} &= \mathbf{b} \\ \mathbf{x} &\geq \mathbf{0} \end{aligned}$$

où \mathbf{A} est une matrice réelle de taille $m \times n$ tel que $m < n$ et $\text{rg}(\mathbf{A}) = m$ et où $\mathbf{b} \geq \mathbf{0}$

- Considérons alors le système d'équations $\mathbf{Ax} = \mathbf{b}$ où $\text{rg}(\mathbf{A}) = m$
- Nous noterons la colonne j de \mathbf{A} par $\mathbf{a}_j = (a_{1j}, \dots, a_{mj})$
- Soit \mathbb{B} un sous-ensemble de m indices de $\{1, \dots, n\}$ et soit $\mathbf{A}_{\mathbb{B}} = (\mathbf{a}_{\mathbb{B}_1}, \dots, \mathbf{a}_{\mathbb{B}_m})$ (où \mathbb{B}_i est le i ème élément de \mathbb{B}) tel que ses colonnes soient linéairement indépendantes. Soit alors $\mathbb{D} = \{1, \dots, n\} \setminus \mathbb{B}$ le complémentaire de \mathbb{B}
- On peut réordonner les colonnes de \mathbf{A} de sorte à avoir : $\mathbf{A} = (\mathbf{A}_{\mathbb{B}}, \mathbf{A}_{\mathbb{D}})$
- Soit également $\mathbf{x}_{\mathbb{B}}$ le sous-vecteur de \mathbf{x} dont les composantes sont celles dont les indices sont dans \mathbb{B} . On peut donc également réordonner les composantes de sorte que $\mathbf{x} = (\mathbf{x}_{\mathbb{B}}, \mathbf{x}_{\mathbb{D}})$

Solutions réalisables (suite)

- La matrice $\mathbf{A}_{\mathbb{B}}$ étant non singulière, nous pouvons donc résoudre le système $\mathbf{A}_{\mathbb{B}} \mathbf{x}_{\mathbb{B}} = \mathbf{b}$ et nous avons :

$$\mathbf{x}_{\mathbb{B}} = \mathbf{A}_{\mathbb{B}}^{-1} \mathbf{b}$$

- Soit alors $\mathbf{x} = (\mathbf{x}_{\mathbb{B}}, \mathbf{0})$. nous voyons que \mathbf{x} ainsi défini est solution du système $\mathbf{Ax} = \mathbf{b}$

Solutions réalisables (suite)

- La matrice $\mathbf{A}_{\mathbb{B}}$ étant non singulière, nous pouvons donc résoudre le système $\mathbf{A}_{\mathbb{B}}\mathbf{x}_{\mathbb{B}} = \mathbf{b}$ et nous avons :

$$\mathbf{x}_{\mathbb{B}} = \mathbf{A}_{\mathbb{B}}^{-1}\mathbf{b}$$

- Soit alors $\mathbf{x} = (\mathbf{x}_{\mathbb{B}}, \mathbf{0})$. nous voyons que \mathbf{x} ainsi défini est solution du système $\mathbf{Ax} = \mathbf{b}$
- Une telle solution nous conduit à la définition suivante :

Définition. (Solution de base)

$\mathbf{x} = (\mathbf{x}_{\mathbb{B}}, \mathbf{0})$ est appelée **solution de base** de $\mathbf{Ax} = \mathbf{b}$ par rapport à la base $\mathbf{A}_{\mathbb{B}}$. Nous appellerons les composantes de $\mathbf{x}_{\mathbb{B}}$ **variables de base** et les colonnes de $\mathbf{A}_{\mathbb{B}}$ par **colonnes de base**

Solutions réalisables (suite)

- La matrice $\mathbf{A}_{\mathbb{B}}$ étant non singulière, nous pouvons donc résoudre le système $\mathbf{A}_{\mathbb{B}}\mathbf{x}_{\mathbb{B}} = \mathbf{b}$ et nous avons :

$$\mathbf{x}_{\mathbb{B}} = \mathbf{A}_{\mathbb{B}}^{-1}\mathbf{b}$$

- Soit alors $\mathbf{x} = (\mathbf{x}_{\mathbb{B}}, \mathbf{0})$. nous voyons que \mathbf{x} ainsi défini est solution du système $\mathbf{Ax} = \mathbf{b}$
- Une telle solution nous conduit à la définition suivante :

Définition. (Solution de base)

$\mathbf{x} = (\mathbf{x}_{\mathbb{B}}, \mathbf{0})$ est appelée **solution de base** de $\mathbf{Ax} = \mathbf{b}$ par rapport à la base $\mathbf{A}_{\mathbb{B}}$. Nous appellerons les composantes de $\mathbf{x}_{\mathbb{B}}$ **variables de base** et les colonnes de $\mathbf{A}_{\mathbb{B}}$ par **colonnes de base**

- Si une des variables de base de $\mathbf{x}_{\mathbb{B}}$ est nulle alors nous dirons que la solution de base est **dégénérée**

Solutions de base réalisables

- Nous avons également les définitions suivantes quant aux solutions des contraintes linéaires :

Définition. (Solution réalisable)

Un vecteur \mathbf{x} satisfaisant aux contraintes $\mathbf{Ax} = \mathbf{b}$ et $\mathbf{x} \geq \mathbf{0}$ est appelée une **solution réalisable**

Solutions de base réalisables

- Nous avons également les définitions suivantes quant aux solutions des contraintes linéaires :

Définition. (Solution réalisable)

Un vecteur \mathbf{x} satisfaisant aux contraintes $\mathbf{Ax} = \mathbf{b}$ et $\mathbf{x} \geq \mathbf{0}$ est appelée une **solution réalisable**

Définition. (Solution de base réalisable)

Un vecteur \mathbf{x} qui est à la fois une solution de base et une solution réalisable est une **solution de base réalisable**

Solutions de base réalisables

- Nous avons également les définitions suivantes quant aux solutions des contraintes linéaires :

Définition. (Solution réalisable)

Un vecteur \mathbf{x} satisfaisant aux contraintes $\mathbf{Ax} = \mathbf{b}$ et $\mathbf{x} \geq \mathbf{0}$ est appelée une **solution réalisable**

Définition. (Solution de base réalisable)

Un vecteur \mathbf{x} qui est à la fois une solution de base et une solution réalisable est une **solution de base réalisable**

- Si la solution de base réalisable est dégénérée on parle de **solution de base réalisable dégénérée**

Solutions de base réalisables (exemple)

- Exemple :

$$\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4) = \begin{pmatrix} 1 & 1 & -1 & 4 \\ 1 & -2 & -1 & 1 \end{pmatrix} \text{ et } \mathbf{b} = \begin{pmatrix} 8 \\ 2 \end{pmatrix}$$

- Le vecteur $\mathbf{x} = (6, 2, 0, 0)$, par rapport à la base $\mathbf{A}_{\mathbb{B}} = (\mathbf{a}_1, \mathbf{a}_2)$, est une solution de base réalisable

Solutions de base réalisables (exemple)

- Exemple :

$$\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4) = \begin{pmatrix} 1 & 1 & -1 & 4 \\ 1 & -2 & -1 & 1 \end{pmatrix} \text{ et } \mathbf{b} = \begin{pmatrix} 8 \\ 2 \end{pmatrix}$$

- Le vecteur $\mathbf{x} = (6, 2, 0, 0)$, par rapport à la base $\mathbf{A}_{\mathbb{B}} = (\mathbf{a}_1, \mathbf{a}_2)$, est

- Exemple :

$$\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4) = \begin{pmatrix} 1 & 1 & -1 & 4 \\ 1 & -2 & -1 & 1 \end{pmatrix} \text{ et } \mathbf{b} = \begin{pmatrix} 8 \\ 2 \end{pmatrix}$$

- Le vecteur $\mathbf{x} = (6, 2, 0, 0)$, par rapport à la base $\mathbf{A}_{\mathbb{B}} = (\mathbf{a}_1, \mathbf{a}_2)$, est une solution de base réalisable
- Le vecteur $\mathbf{x} = (0, 0, 0, 2)$, par rapport à la base $\mathbf{A}_{\mathbb{B}} = (\mathbf{a}_3, \mathbf{a}_4)$, est

Solutions de base réalisables (exemple)

- Exemple :

$$\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4) = \begin{pmatrix} 1 & 1 & -1 & 4 \\ 1 & -2 & -1 & 1 \end{pmatrix} \text{ et } \mathbf{b} = \begin{pmatrix} 8 \\ 2 \end{pmatrix}$$

- Le vecteur $\mathbf{x} = (6, 2, 0, 0)$, par rapport à la base $\mathbf{A}_{\mathbb{B}} = (\mathbf{a}_1, \mathbf{a}_2)$, est une solution de base réalisable
- Le vecteur $\mathbf{x} = (0, 0, 0, 2)$, par rapport à la base $\mathbf{A}_{\mathbb{B}} = (\mathbf{a}_3, \mathbf{a}_4)$, est une solution de base réalisable dégénérée

Solutions de base réalisables (exemple)

- Exemple :

$$\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4) = \begin{pmatrix} 1 & 1 & -1 & 4 \\ 1 & -2 & -1 & 1 \end{pmatrix} \text{ et } \mathbf{b} = \begin{pmatrix} 8 \\ 2 \end{pmatrix}$$

- Le vecteur $\mathbf{x} = (6, 2, 0, 0)$, par rapport à la base $\mathbf{A}_{\mathbb{B}} = (\mathbf{a}_1, \mathbf{a}_2)$, est une solution de base réalisable
- Le vecteur $\mathbf{x} = (0, 0, 0, 2)$, par rapport à la base $\mathbf{A}_{\mathbb{B}} = (\mathbf{a}_3, \mathbf{a}_4)$, est une solution de base réalisable dégénérée
- Le vecteur $\mathbf{x} = (3, 1, 0, 1)$, par rapport à la base $\mathbf{A}_{\mathbb{B}} = (\mathbf{a}_1, \mathbf{a}_2)$, est une solution réalisable (qui n'est pas de base)

Solutions de base réalisables (exemple)

- Exemple :

$$\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4) = \begin{pmatrix} 1 & 1 & -1 & 4 \\ 1 & -2 & -1 & 1 \end{pmatrix} \text{ et } \mathbf{b} = \begin{pmatrix} 8 \\ 2 \end{pmatrix}$$

- Le vecteur $\mathbf{x} = (6, 2, 0, 0)$, par rapport à la base $\mathbf{A}_{\mathbb{B}} = (\mathbf{a}_1, \mathbf{a}_2)$, est une solution de base réalisable
- Le vecteur $\mathbf{x} = (0, 0, 0, 2)$, par rapport à la base $\mathbf{A}_{\mathbb{B}} = (\mathbf{a}_3, \mathbf{a}_4)$, est une solution de base réalisable dégénérée
- Le vecteur $\mathbf{x} = (3, 1, 0, 1)$, par rapport à la base $\mathbf{A}_{\mathbb{B}} = (\mathbf{a}_1, \mathbf{a}_2)$, est

Solutions de base réalisables (exemple)

- Exemple :

$$\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4) = \begin{pmatrix} 1 & 1 & -1 & 4 \\ 1 & -2 & -1 & 1 \end{pmatrix} \text{ et } \mathbf{b} = \begin{pmatrix} 8 \\ 2 \end{pmatrix}$$

- Le vecteur $\mathbf{x} = (6, 2, 0, 0)$, par rapport à la base $\mathbf{A}_{\mathbb{B}} = (\mathbf{a}_1, \mathbf{a}_2)$, est une solution de base réalisable
- Le vecteur $\mathbf{x} = (0, 0, 0, 2)$, par rapport à la base $\mathbf{A}_{\mathbb{B}} = (\mathbf{a}_3, \mathbf{a}_4)$, est une solution de base réalisable dégénérée
- Le vecteur $\mathbf{x} = (3, 1, 0, 1)$, par rapport à la base $\mathbf{A}_{\mathbb{B}} = (\mathbf{a}_1, \mathbf{a}_2)$, est une solution réalisable (qui n'est pas de base)
- Le vecteur $\mathbf{x} = (0, 2, -6, 0)$, par rapport à la base $\mathbf{A}_{\mathbb{B}} = (\mathbf{a}_2, \mathbf{a}_3)$, est

Solutions de base réalisables (exemple)

- Exemple :

$$\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4) = \begin{pmatrix} 1 & 1 & -1 & 4 \\ 1 & -2 & -1 & 1 \end{pmatrix} \text{ et } \mathbf{b} = \begin{pmatrix} 8 \\ 2 \end{pmatrix}$$

- Le vecteur $\mathbf{x} = (6, 2, 0, 0)$, par rapport à la base $\mathbf{A}_{\mathbb{B}} = (\mathbf{a}_1, \mathbf{a}_2)$, est une solution de base réalisable
- Le vecteur $\mathbf{x} = (0, 0, 0, 2)$, par rapport à la base $\mathbf{A}_{\mathbb{B}} = (\mathbf{a}_3, \mathbf{a}_4)$, est une solution de base réalisable dégénérée
- Le vecteur $\mathbf{x} = (3, 1, 0, 1)$, par rapport à la base $\mathbf{A}_{\mathbb{B}} = (\mathbf{a}_1, \mathbf{a}_2)$, est une solution réalisable (qui n'est pas de base)
- Le vecteur $\mathbf{x} = (0, 2, -6, 0)$, par rapport à la base $\mathbf{A}_{\mathbb{B}} = (\mathbf{a}_2, \mathbf{a}_3)$, est une solution de base (non réalisable)

Propriétés des solutions de base réalisables

- Revenons maintenant sur le lien entre les solutions de base réalisables et notre problème d'optimisation linéaire $f(\mathbf{x}) = \mathbf{c}^t \mathbf{x}$
- Nous avons les définitions et le théorème fondamental suivants :

Propriétés des solutions de base réalisables

- Revenons maintenant sur le lien entre les solutions de base réalisables et notre problème d'optimisation linéaire $f(\mathbf{x}) = \mathbf{c}^t \mathbf{x}$

Propriétés des solutions de base réalisables

- Revenons maintenant sur le lien entre les solutions de base réalisables et notre problème d'optimisation linéaire $f(\mathbf{x}) = \mathbf{c}^t \mathbf{x}$
- Nous avons les définitions et le théorème fondamental suivants :

Définition. (Solutions réalisables optimales)

Un vecteur \mathbf{x} qui permet d'atteindre le minimum de la fonction objectif parmi l'ensemble des vecteurs satisfaisants aux contraintes $\mathbf{Ax} = \mathbf{b}$ et $\mathbf{x} \geq \mathbf{0}$ est appelé une **solution réalisable optimale**. Si de plus \mathbf{x} est une solution de base alors on parle de **solution de base réalisable optimale**

Propriétés des solutions de base réalisables

- Revenons maintenant sur le lien entre les solutions de base réalisables et notre problème d'optimisation linéaire $f(\mathbf{x}) = \mathbf{c}^t \mathbf{x}$
- Nous avons les définitions et le théorème fondamental suivants :

Définition. (Solutions réalisables optimales)

Un vecteur \mathbf{x} qui permet d'atteindre le minimum de la fonction objectif parmi l'ensemble des vecteurs satisfaisants aux contraintes $\mathbf{Ax} = \mathbf{b}$ et $\mathbf{x} \geq \mathbf{0}$ est appelé une **solution réalisable optimale**. Si de plus \mathbf{x} est une solution de base alors on parle de **solution de base réalisable optimale**

Théorème. (Théorème fondamental en PL)

Soit un PL sous forme standard. S'il existe une solution réalisable, alors il existe une solution de base réalisable. S'il existe une solution réalisable optimale, alors il existe une solution de base réalisable optimale.

Propriétés des solutions de base réalisables (suite)

- Remarque sur le théorème fondamental : il permet de réduire la recherche d'une solution optimale d'un PL à la recherche dans un ensemble fini d'éléments. En effet, la solution optimale, si elle existe, peut-être appréhendée comme étant une solution de base réalisable. Comme il y a $\binom{n}{m}$ façon de choisir une base de dimension m dans un ensemble de n vecteurs il s'agit donc de chercher dans un ensemble fini de $\binom{n}{m}$ éléments
- Si l'ensemble est fini, le nombre d'éléments est toutefois très grand

Propriétés des solutions de base réalisables (suite)

- Remarque sur le théorème fondamental : il permet de réduire la recherche d'une solution optimale d'un PL à la recherche dans un ensemble fini d'éléments. En effet, la solution optimale, si elle existe, peut-être appréhendée comme étant une solution de base réalisable. Comme il y a $\binom{n}{m}$ façon de choisir une base de dimension m dans un ensemble de n vecteurs il s'agit donc de chercher dans un ensemble fini de $\binom{n}{m}$ éléments

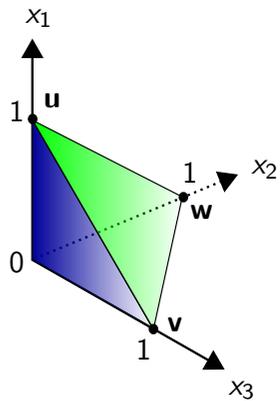
Propriétés des solutions de base réalisables (suite)

- Remarque sur le théorème fondamental : il permet de réduire la recherche d'une solution optimale d'un PL à la recherche dans un ensemble fini d'éléments. En effet, la solution optimale, si elle existe, peut-être appréhendée comme étant une solution de base réalisable. Comme il y a $\binom{n}{m}$ façon de choisir une base de dimension m dans un ensemble de n vecteurs il s'agit donc de chercher dans un ensemble fini de $\binom{n}{m}$ éléments
- Si l'ensemble est fini, le nombre d'éléments est toutefois très grand
- Nous voyons dans ce qui suit une interprétation géométrique de ces propriétés qui permet d'introduire la méthode du simplexe

Interprétation géométrique

Théorème.

L'ensemble des solutions d'un ensemble d'équations linéaires est un ensemble convexe comprenant un ensemble fini de points extrêmes.

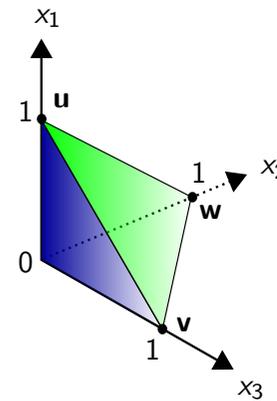


- Soit, dans \mathbb{R}^3 , l'ensemble convexe défini par : $x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_1 + x_2 + x_3 \leq 1$.
- Sur la figure la face en vert est formée des points vérifiant :
- La face en bleue est formée des points vérifiant :
- L'arête **uv** est la frontière formée des points vérifiant à la fois :
- Le sommet **w** est tel que :

Interprétation géométrique

Théorème.

L'ensemble des solutions d'un ensemble d'équations linéaires est un ensemble convexe comprenant un ensemble fini de points extrêmes.



- Soit, dans \mathbb{R}^3 , l'ensemble convexe défini par : $x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_1 + x_2 + x_3 \leq 1$.
- Sur la figure la face en vert est formée des points vérifiant : $x_1 + x_2 + x_3 = 1$
- La face en bleue est formée des points vérifiant : $x_1 + x_2 + x_3 \leq 1$ et $x_2 = 0$
- L'arête **uv** est la frontière formée des points vérifiant à la fois : $x_1 + x_2 + x_3 = 1$ et $x_2 = 0$
- Le sommet **w** est tel que : $x_1 + x_2 + x_3 = 1, x_1 = 0$ et $x_3 = 0$

Interprétation géométrique (suite)

Théorème.

Soit \mathbb{S} l'ensemble des solutions satisfaisant les contraintes $\mathbf{Ax} = \mathbf{b}$ et $\mathbf{x} \geq \mathbf{0}$ où $\mathbf{A} \in \mathcal{M}_{m,n}$ avec $m < n$. Alors \mathbf{x} est un point extrême de \mathbb{S} ssi \mathbf{x} est une solution de base réalisable de $\mathbf{Ax} = \mathbf{b}$ et $\mathbf{x} \geq \mathbf{0}$

Interprétation géométrique (suite)

Théorème.

Soit \mathbb{S} l'ensemble des solutions satisfaisant les contraintes $\mathbf{Ax} = \mathbf{b}$ et $\mathbf{x} \geq \mathbf{0}$ où $\mathbf{A} \in \mathcal{M}_{m,n}$ avec $m < n$. Alors \mathbf{x} est un point extrême de \mathbb{S} ssi \mathbf{x} est une solution de base réalisable de $\mathbf{Ax} = \mathbf{b}$ et $\mathbf{x} \geq \mathbf{0}$

- En d'autres termes, l'ensemble des solutions $\mathbb{S} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{Ax} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$ est identique à l'ensemble des solutions de base réalisable de $\mathbf{Ax} = \mathbf{b}$ et $\mathbf{x} \geq \mathbf{0}$
- Le théorème fondamental permet donc de conclure que pour résoudre un PL, il suffit d'examiner les points extrêmes de \mathbb{S}

Rappel du Sommaire

- 1 Rappels de concepts mathématiques
- 2 Bases théoriques de l'optimisation
- 3 Optimisation sans contrainte
- 4 Optimisation linéaire avec contraintes
 - Programmation linéaire
 - Méthode du simplexe
 - Solution de base réalisable initiale et cas dégénéré
 - Dualité
- 5 Optimisation non linéaire avec contraintes

Rappel des notations

- Rappel des notations :
 - ▶ $\mathbf{x} = (\mathbf{x}_B, \mathbf{x}_D)$ où \mathbf{x}_B (resp \mathbf{x}_D) est le vecteur de taille $(m \times 1)$ (resp $((n - m) \times 1)$) issu de \mathbf{x} dont les composantes sont celles des m variables de bases (resp $n - m$ variables hors base)
 - ▶ $\mathbf{c} = (\mathbf{c}_B, \mathbf{c}_D)$ où \mathbf{c}_B (resp \mathbf{c}_D) est le vecteur de taille $(m \times 1)$ (resp $((n - m) \times 1)$) dont les coûts sont ceux relatifs aux m variables de bases (resp $n - m$ variables hors base)
 - ▶ $\mathbf{A} = (\mathbf{A}_B, \mathbf{A}_D)$ où \mathbf{A}_B (resp \mathbf{A}_D) est la sous-matrice de taille $(m \times m)$ (resp $(m \times (n - m))$) issue de \mathbf{A} dont les composantes sont les colonnes relatives aux m variables de bases (resp $n - m$ variables hors base). \mathbf{A}_B est de rang m .

Introduction

- Il s'agit d'un algorithme permettant de résoudre **efficacement** les PL mis sous leur forme standard :

$$\begin{aligned} \min f(\mathbf{x}) &= \mathbf{c}^t \mathbf{x} \\ \text{s.t. } \mathbf{Ax} &= \mathbf{b} \\ \mathbf{x} &\geq \mathbf{0} \end{aligned}$$

où \mathbf{A} est une matrice réelle de taille $m \times n$ tel que $m < n$ et $\text{rg}(\mathbf{A}) = m$ et où $\mathbf{b} \geq \mathbf{0}$

- Algorithme conçu initialement par G. Dantzig en 1947 : la procédure revient à parcourir les solutions de base réalisables afin d'améliorer la valeur de la fonction objectif jusqu'à obtenir un optimum

Rappel des notations

- Rappel des notations :
 - ▶ $\mathbf{x} = (\mathbf{x}_B, \mathbf{x}_D)$ où \mathbf{x}_B (resp \mathbf{x}_D) est le vecteur de taille $(m \times 1)$ (resp $((n - m) \times 1)$) issu de \mathbf{x} dont les composantes sont celles des m variables de bases (resp $n - m$ variables hors base)
 - ▶ $\mathbf{c} = (\mathbf{c}_B, \mathbf{c}_D)$ où \mathbf{c}_B (resp \mathbf{c}_D) est le vecteur de taille $(m \times 1)$ (resp $((n - m) \times 1)$) dont les coûts sont ceux relatifs aux m variables de bases (resp $n - m$ variables hors base)
 - ▶ $\mathbf{A} = (\mathbf{A}_B, \mathbf{A}_D)$ où \mathbf{A}_B (resp \mathbf{A}_D) est la sous-matrice de taille $(m \times m)$ (resp $(m \times (n - m))$) issue de \mathbf{A} dont les composantes sont les colonnes relatives aux m variables de bases (resp $n - m$ variables hors base). \mathbf{A}_B est de rang m .
- Nous avons : $\mathbf{Ax} = \mathbf{b} \Leftrightarrow \mathbf{A}_B \mathbf{x}_B + \mathbf{A}_D \mathbf{x}_D = \mathbf{b}$

Rappel des notations

- Rappel des notations :

- ▶ $\mathbf{x} = (\mathbf{x}_B, \mathbf{x}_D)$ où \mathbf{x}_B (resp \mathbf{x}_D) est le vecteur de taille $(m \times 1)$ (resp $((n - m) \times 1)$) issu de \mathbf{x} dont les composantes sont celles des m variables de bases (resp $n - m$ variables hors base)
- ▶ $\mathbf{c} = (\mathbf{c}_B, \mathbf{c}_D)$ où \mathbf{c}_B (resp \mathbf{c}_D) est le vecteur de taille $(m \times 1)$ issu de \mathbf{c} (resp $((n - m) \times 1)$) dont les coûts sont ceux relatifs aux m variables de bases (resp $n - m$ variables hors base)
- ▶ $\mathbf{A} = (\mathbf{A}_B, \mathbf{A}_D)$ où \mathbf{A}_B (resp \mathbf{A}_D) est la sous-matrice de taille $(m \times m)$ (resp $(m \times (n - m))$) issue de \mathbf{A} dont les composantes sont les colonnes relatives aux m variables de bases (resp $n - m$ variables hors base). \mathbf{A}_B est de rang m .

- Nous avons : $\mathbf{Ax} = \mathbf{b} \Leftrightarrow \mathbf{A}_B \mathbf{x}_B + \mathbf{A}_D \mathbf{x}_D = \mathbf{b}$

- Expression des variables de base en fonction des variables hors base :

$$\mathbf{x}_B = \mathbf{A}_B^{-1}(\mathbf{b} - \mathbf{A}_D \mathbf{x}_D)$$

Condition d'optimalité d'une solution de base réalisable

- Expression de la fonction objectif :

$$\begin{aligned} f(\mathbf{x}) = \mathbf{c}^t \mathbf{x} &= \mathbf{c}_B^t \mathbf{x}_B + \mathbf{c}_D^t \mathbf{x}_D \\ &= \mathbf{c}_B^t (\mathbf{A}_B^{-1}(\mathbf{b} - \mathbf{A}_D \mathbf{x}_D)) + \mathbf{c}_D^t \mathbf{x}_D \\ &= \mathbf{c}_B^t \mathbf{A}_B^{-1} \mathbf{b} + \underbrace{(\mathbf{c}_D^t - \mathbf{c}_B^t \mathbf{A}_B^{-1} \mathbf{A}_D)}_{\bar{\mathbf{c}}_D} \mathbf{x}_D \end{aligned}$$

- Les quantités $\bar{\mathbf{c}}_D$ sont appelées “coûts” réduits des variables hors base

Condition d'optimalité d'une solution de base réalisable

- Expression de la fonction objectif :

$$\begin{aligned} f(\mathbf{x}) = \mathbf{c}^t \mathbf{x} &= \mathbf{c}_B^t \mathbf{x}_B + \mathbf{c}_D^t \mathbf{x}_D \\ &= \mathbf{c}_B^t (\mathbf{A}_B^{-1}(\mathbf{b} - \mathbf{A}_D \mathbf{x}_D)) + \mathbf{c}_D^t \mathbf{x}_D \\ &= \mathbf{c}_B^t \mathbf{A}_B^{-1} \mathbf{b} + \underbrace{(\mathbf{c}_D^t - \mathbf{c}_B^t \mathbf{A}_B^{-1} \mathbf{A}_D)}_{\bar{\mathbf{c}}_D} \mathbf{x}_D \end{aligned}$$

Condition d'optimalité d'une solution de base réalisable

- Expression de la fonction objectif :

$$\begin{aligned} f(\mathbf{x}) = \mathbf{c}^t \mathbf{x} &= \mathbf{c}_B^t \mathbf{x}_B + \mathbf{c}_D^t \mathbf{x}_D \\ &= \mathbf{c}_B^t (\mathbf{A}_B^{-1}(\mathbf{b} - \mathbf{A}_D \mathbf{x}_D)) + \mathbf{c}_D^t \mathbf{x}_D \\ &= \mathbf{c}_B^t \mathbf{A}_B^{-1} \mathbf{b} + \underbrace{(\mathbf{c}_D^t - \mathbf{c}_B^t \mathbf{A}_B^{-1} \mathbf{A}_D)}_{\bar{\mathbf{c}}_D} \mathbf{x}_D \end{aligned}$$

- Les quantités $\bar{\mathbf{c}}_D$ sont appelées “coûts” réduits des variables hors base

⇒ On peut toujours trouver une base permettant de diminuer la fonction objectif tant qu'il existe une composante de $\bar{\mathbf{c}}_D$ qui est négative. Autrement dit une **condition nécessaire et suffisante d'optimalité** d'une solution de base réalisable est :

$$\mathbf{c}_D^t - \mathbf{c}_B^t \mathbf{A}_B^{-1} \mathbf{A}_D \geq \mathbf{0}$$

Passage d'une solution de base réalisable à une autre

- Pour la solution de base réalisable courante $\mathbf{x} = (\mathbf{x}_B, \mathbf{x}_D)$ et

$$f(\mathbf{x}) = \mathbf{c}_B^t \mathbf{A}_B^{-1} \mathbf{b} + \underbrace{(\mathbf{c}_D^t - \mathbf{c}_B^t \mathbf{A}_B^{-1} \mathbf{A}_D)}_{\bar{\mathbf{c}}_D} \mathbf{x}_D$$

Passage d'une solution de base réalisable à une autre

- Pour la solution de base réalisable courante $\mathbf{x} = (\mathbf{x}_B, \mathbf{x}_D)$ et

$$f(\mathbf{x}) = \mathbf{c}_B^t \mathbf{A}_B^{-1} \mathbf{b} + \underbrace{(\mathbf{c}_D^t - \mathbf{c}_B^t \mathbf{A}_B^{-1} \mathbf{A}_D)}_{\bar{\mathbf{c}}_D} \mathbf{x}_D$$
- Le passage d'une solution de base réalisable à une autre le long d'une arête permettant d'améliorer la fonction objectif s'effectue en faisant entrer une nouvelle variable hors base dans la base et en faisant sortir de la base une autre variable :

Passage d'une solution de base réalisable à une autre

- Pour la solution de base réalisable courante $\mathbf{x} = (\mathbf{x}_B, \mathbf{x}_D)$ et

$$f(\mathbf{x}) = \mathbf{c}_B^t \mathbf{A}_B^{-1} \mathbf{b} + \underbrace{(\mathbf{c}_D^t - \mathbf{c}_B^t \mathbf{A}_B^{-1} \mathbf{A}_D)}_{\bar{\mathbf{c}}_D} \mathbf{x}_D$$
- Le passage d'une solution de base réalisable à une autre le long d'une arête permettant d'améliorer la fonction objectif s'effectue en faisant entrer une nouvelle variable hors base dans la base et en faisant sortir de la base une autre variable :
 - ▶ On cherche dans $\bar{\mathbf{c}}_D$ la variable hors base $s = \arg \min_{i \in D} \{[\bar{\mathbf{c}}_D]_i\}$ où $[\bar{\mathbf{c}}_D]_s$ est la diminution la plus grande que l'on peut avoir pour la fonction objectif. x_s est la variable qui entre dans la base

Passage d'une solution de base réalisable à une autre

- Pour la solution de base réalisable courante $\mathbf{x} = (\mathbf{x}_B, \mathbf{x}_D)$ et

$$f(\mathbf{x}) = \mathbf{c}_B^t \mathbf{A}_B^{-1} \mathbf{b} + \underbrace{(\mathbf{c}_D^t - \mathbf{c}_B^t \mathbf{A}_B^{-1} \mathbf{A}_D)}_{\bar{\mathbf{c}}_D} \mathbf{x}_D$$
- Le passage d'une solution de base réalisable à une autre le long d'une arête permettant d'améliorer la fonction objectif s'effectue en faisant entrer une nouvelle variable hors base dans la base et en faisant sortir de la base une autre variable :
 - ▶ On cherche dans $\bar{\mathbf{c}}_D$ la variable hors base $s = \arg \min_{i \in D} \{[\bar{\mathbf{c}}_D]_i\}$ où $[\bar{\mathbf{c}}_D]_s$ est la diminution la plus grande que l'on peut avoir pour la fonction objectif. x_s est la variable qui entre dans la base
 - ▶ On définit la colonne de travail comme étant \mathbf{a}_s , le vecteur colonne de \mathbf{A}_D relatif à x_s

Passage d'une solution de base réalisable à une autre

- Pour la solution de base réalisable courante $\mathbf{x} = (\mathbf{x}_{\mathbb{B}}, \mathbf{x}_{\mathbb{D}})$ et $f(\mathbf{x}) = \mathbf{c}_{\mathbb{B}}^t \mathbf{A}_{\mathbb{B}}^{-1} \mathbf{b} + \underbrace{(\mathbf{c}_{\mathbb{D}}^t - \mathbf{c}_{\mathbb{B}}^t \mathbf{A}_{\mathbb{B}}^{-1} \mathbf{A}_{\mathbb{D}})}_{\bar{\mathbf{c}}_{\mathbb{D}}} \mathbf{x}_{\mathbb{D}}$
- Le passage d'une solution de base réalisable à une autre le long d'une arête permettant d'améliorer la fonction objectif s'effectue en faisant entrer une nouvelle variable hors base dans la base et en faisant sortir de la base une autre variable :
 - ▶ On cherche dans $\bar{\mathbf{c}}_{\mathbb{D}}$ la variable hors base $s = \arg \min_{i \in \mathbb{D}} \{[\bar{\mathbf{c}}_{\mathbb{D}}]_i\}$ où $[\bar{\mathbf{c}}_{\mathbb{D}}]_s$ est la diminution la plus grande que l'on peut avoir pour la fonction objectif. x_s est la variable qui entre dans la base
 - ▶ On définit la colonne de travail comme étant \mathbf{a}_s , le vecteur colonne de $\mathbf{A}_{\mathbb{D}}$ relatif à x_s
 - ▶ On calcule les ratios entre les membres de droite et les valeurs non nulles de \mathbf{a}_s et on détermine $t = \arg \min_{i \in \{1, \dots, m\} : a_{is} > 0} \left\{ \frac{b_i}{a_{is}} \right\}$. $x_{\mathbb{B}_t}$ est la variable qui sort de la base (ie la t -ème variable dans \mathbb{B})

Passage d'une solution de base réalisable à une autre (suite)

- Le changement de base se traduit par l'expression des variables de base en fonction des variables hors base :
 - ▶ on peut exprimer les variables de base et $f(\mathbf{x})$ en fonction des seules variables hors base
 - ▶ les colonnes de la matrice des contraintes correspondant aux variables de base forment une matrice unité

Passage d'une solution de base réalisable à une autre

- Pour la solution de base réalisable courante $\mathbf{x} = (\mathbf{x}_{\mathbb{B}}, \mathbf{x}_{\mathbb{D}})$ et $f(\mathbf{x}) = \mathbf{c}_{\mathbb{B}}^t \mathbf{A}_{\mathbb{B}}^{-1} \mathbf{b} + \underbrace{(\mathbf{c}_{\mathbb{D}}^t - \mathbf{c}_{\mathbb{B}}^t \mathbf{A}_{\mathbb{B}}^{-1} \mathbf{A}_{\mathbb{D}})}_{\bar{\mathbf{c}}_{\mathbb{D}}} \mathbf{x}_{\mathbb{D}}$
- Le passage d'une solution de base réalisable à une autre le long d'une arête permettant d'améliorer la fonction objectif s'effectue en faisant entrer une nouvelle variable hors base dans la base et en faisant sortir de la base une autre variable :
 - ▶ On cherche dans $\bar{\mathbf{c}}_{\mathbb{D}}$ la variable hors base $s = \arg \min_{i \in \mathbb{D}} \{[\bar{\mathbf{c}}_{\mathbb{D}}]_i\}$ où $[\bar{\mathbf{c}}_{\mathbb{D}}]_s$ est la diminution la plus grande que l'on peut avoir pour la fonction objectif. x_s est la variable qui entre dans la base
 - ▶ On définit la colonne de travail comme étant \mathbf{a}_s , le vecteur colonne de $\mathbf{A}_{\mathbb{D}}$ relatif à x_s
 - ▶ On calcule les ratios entre les membres de droite et les valeurs non nulles de \mathbf{a}_s et on détermine $t = \arg \min_{i \in \{1, \dots, m\} : a_{is} > 0} \left\{ \frac{b_i}{a_{is}} \right\}$. $x_{\mathbb{B}_t}$ est la variable qui sort de la base (ie la t -ème variable dans \mathbb{B})
 - ▶ $\frac{b_t}{a_{ts}}$ est la quantité maximale que l'on peut allouer à x_s sans violer les contraintes de non négativité des variables

Passage d'une solution de base réalisable à une autre (suite)

- Le changement de base se traduit par l'expression des variables de base en fonction des variables hors base :
 - ▶ on peut exprimer les variables de base et $f(\mathbf{x})$ en fonction des seules variables hors base
 - ▶ les colonnes de la matrice des contraintes correspondant aux variables de base forment une matrice unité
- D'un point de vue algèbre cela revient à éliminer la variable entrante du système $\mathbf{Ax} = \mathbf{b}$. Nous pouvons effectuer ceci en appliquant les formules de l'élimination de Gauss-Jordan (méthode du pivot)

Mise en oeuvre de l'algorithme par la méthode des tableaux

- L'algorithme du simplexe consiste à passer d'une solution de base réalisable à une autre solution de base réalisable adjacente (le long d'une arête) en améliorant à chaque itération la valeur de f

Algorithme du simplexe à partir du tableau

Input : Tableau du simplexe relatif à un PL sous forme standard,
 \mathbf{x} une solution de base réalisable initiale

Mise en oeuvre de l'algorithme par la méthode des tableaux

- L'algorithme du simplexe consiste à passer d'une solution de base réalisable à une autre solution de base réalisable adjacente (le long d'une arête) en améliorant à chaque itération la valeur de f
- Pour mettre en pratique l'algorithme, nous utiliserons le "tableau simplexe" qui est une représentation tabulaire regroupant les données calculées à chaque itération de l'algorithme. Il est de la forme suivante où chaque entité est m à j à chaque itération :

		\mathbf{x}^t	
		\mathbf{c}^t	
$\mathbf{x}_{\mathbb{B}}$	$\mathbf{c}_{\mathbb{B}}$	\mathbf{A}	\mathbf{b}
		$\mathbf{c}^t - \mathbf{c}_{\mathbb{B}}^t \mathbf{A}$	$\mathbf{c}_{\mathbb{B}}^t \mathbf{b}$

Algorithme du simplexe à partir du tableau

Input : Tableau du simplexe relatif à un PL sous forme standard,
 \mathbf{x} une solution de base réalisable initiale

- 1 Dans la ligne des coûts réduits calculés par $\mathbf{c}^t - \mathbf{c}_{\mathbb{B}}^t \mathbf{A}$,
 Déterminer $s = \arg \min_{i \in \mathbb{D}} \{[\mathbf{c}^t - \mathbf{c}_{\mathbb{B}}^t \mathbf{A}]_i\}$
 Si $[\mathbf{c}^t - \mathbf{c}_{\mathbb{B}}^t \mathbf{A}]_s \geq 0$ alors optimum atteint et aller en 5

Algorithme du simplexe à partir du tableau

Input : Tableau du simplexe relatif à un PL sous forme standard,
 \mathbf{x} une solution de base réalisable initiale

- 1 Dans la ligne des coûts réduits calculés par $\mathbf{c}^t - \mathbf{c}_{\mathbb{B}}^t \mathbf{A}$,
 Déterminer $s = \arg \min_{i \in \mathbb{D}} \{[\mathbf{c}^t - \mathbf{c}_{\mathbb{B}}^t \mathbf{A}]_i\}$
 Si $[\mathbf{c}^t - \mathbf{c}_{\mathbb{B}}^t \mathbf{A}]_s \geq 0$ alors optimum atteint et aller en 5
- 2 A l'aide de la colonne \mathbf{a}_s de \mathbf{A} et celle de \mathbf{b} ,
 Déterminer $t = \arg \min_{i \in \{1, \dots, m\} : a_{is} > 0} \left\{ \frac{b_i}{a_{is}} \right\}$,
 Si $t = \emptyset$ (ie quand $\forall i : a_{is} \leq 0$) alors problème non borné

Algorithme du simplexe à partir du tableau

Input : Tableau du simplexe relatif à un PL sous forme standard,
 \mathbf{x} une solution de base réalisable initiale

- 1 Dans la ligne des coûts réduits calculés par $\mathbf{c}^t - \mathbf{c}_{\mathbb{B}}^t \mathbf{A}$,
 Déterminer $s = \arg \min_{i \in \mathbb{D}} \{[\mathbf{c}^t - \mathbf{c}_{\mathbb{B}}^t \mathbf{A}]_i\}$
 Si $[\mathbf{c}^t - \mathbf{c}_{\mathbb{B}}^t \mathbf{A}]_s \geq 0$ alors optimum atteint et aller en 5
- 2 A l'aide de la colonne \mathbf{a}_s de \mathbf{A} et celle de \mathbf{b} ,
 Déterminer $t = \arg \min_{i \in \{1, \dots, m\} : a_{is} > 0} \left\{ \frac{b_i}{a_{is}} \right\}$,
 Si $t = \emptyset$ (ie quand $\forall i : a_{is} \leq 0$) alors problème non borné
- 3 Faire entrer x_s dans la base et faire sortir $x_{\mathbb{B}_t}$ de la base,
 Calculer la ligne du pivot t , $\forall j : a_{tj} \leftarrow \frac{a_{tj}}{a_{ts}}$, et $b_t \leftarrow \frac{b_t}{a_{ts}}$,
 Combiner linéairement les lignes $i \neq t$ de \mathbf{A} et \mathbf{b} avec la ligne pivot t
 pour cela on fait $\forall i \neq t, \forall j : a_{ij} \leftarrow a_{ij} - a_{is} a_{tj}$ et $b_i \leftarrow b_i - a_{is} b_t$
- 4 Mettre à jour les 2 premières colonnes du tableau (nouvelle base),
 Calculer les nouveaux coûts réduits $\mathbf{c}^t - \mathbf{c}_{\mathbb{B}}^t \mathbf{A}$,
 Calculer la valeur courante de $f : \mathbf{c}_{\mathbb{B}}^t \mathbf{b}$ et aller en 1

Algorithme du simplexe à partir du tableau

Input : Tableau du simplexe relatif à un PL sous forme standard,
 \mathbf{x} une solution de base réalisable initiale

- 1 Dans la ligne des coûts réduits calculés par $\mathbf{c}^t - \mathbf{c}_{\mathbb{B}}^t \mathbf{A}$,
 Déterminer $s = \arg \min_{i \in \mathbb{D}} \{[\mathbf{c}^t - \mathbf{c}_{\mathbb{B}}^t \mathbf{A}]_i\}$
 Si $[\mathbf{c}^t - \mathbf{c}_{\mathbb{B}}^t \mathbf{A}]_s \geq 0$ alors optimum atteint et aller en 5
- 2 A l'aide de la colonne \mathbf{a}_s de \mathbf{A} et celle de \mathbf{b} ,
 Déterminer $t = \arg \min_{i \in \{1, \dots, m\} : a_{is} > 0} \left\{ \frac{b_i}{a_{is}} \right\}$,
 Si $t = \emptyset$ (ie quand $\forall i : a_{is} \leq 0$) alors problème non borné
- 3 Faire entrer x_s dans la base et faire sortir $x_{\mathbb{B}_t}$ de la base,
 Calculer la ligne du pivot t , $\forall j : a_{tj} \leftarrow \frac{a_{tj}}{a_{ts}}$, et $b_t \leftarrow \frac{b_t}{a_{ts}}$,
 Combiner linéairement les lignes $i \neq t$ de \mathbf{A} et \mathbf{b} avec la ligne pivot t
 pour cela on fait $\forall i \neq t, \forall j : a_{ij} \leftarrow a_{ij} - a_{is} a_{tj}$ et $b_i \leftarrow b_i - a_{is} b_t$

Algorithme du simplexe à partir du tableau

Input : Tableau du simplexe relatif à un PL sous forme standard,
 \mathbf{x} une solution de base réalisable initiale

- 1 Dans la ligne des coûts réduits calculés par $\mathbf{c}^t - \mathbf{c}_{\mathbb{B}}^t \mathbf{A}$,
 Déterminer $s = \arg \min_{i \in \mathbb{D}} \{[\mathbf{c}^t - \mathbf{c}_{\mathbb{B}}^t \mathbf{A}]_i\}$
 Si $[\mathbf{c}^t - \mathbf{c}_{\mathbb{B}}^t \mathbf{A}]_s \geq 0$ alors optimum atteint et aller en 5
- 2 A l'aide de la colonne \mathbf{a}_s de \mathbf{A} et celle de \mathbf{b} ,
 Déterminer $t = \arg \min_{i \in \{1, \dots, m\} : a_{is} > 0} \left\{ \frac{b_i}{a_{is}} \right\}$,
 Si $t = \emptyset$ (ie quand $\forall i : a_{is} \leq 0$) alors problème non borné
- 3 Faire entrer x_s dans la base et faire sortir $x_{\mathbb{B}_t}$ de la base,
 Calculer la ligne du pivot t , $\forall j : a_{tj} \leftarrow \frac{a_{tj}}{a_{ts}}$, et $b_t \leftarrow \frac{b_t}{a_{ts}}$,
 Combiner linéairement les lignes $i \neq t$ de \mathbf{A} et \mathbf{b} avec la ligne pivot t
 pour cela on fait $\forall i \neq t, \forall j : a_{ij} \leftarrow a_{ij} - a_{is} a_{tj}$ et $b_i \leftarrow b_i - a_{is} b_t$
- 4 Mettre à jour les 2 premières colonnes du tableau (nouvelle base),
 Calculer les nouveaux coûts réduits $\mathbf{c}^t - \mathbf{c}_{\mathbb{B}}^t \mathbf{A}$,
 Calculer la valeur courante de $f : \mathbf{c}_{\mathbb{B}}^t \mathbf{b}$ et aller en 1
- 5 **Output :** $\mathbf{x}_{\mathbb{B}} = \mathbf{b}$ et $\mathbf{x}_{\mathbb{D}} = \mathbf{0}$

Bien fondé de l'algorithme du simplexe

Théorème.

Soit \mathbf{x} une solution de base réalisable quelconque et soit le vecteur des "coûts réduits" des variables hors base : $(\mathbf{c}_D^t - \mathbf{c}_B^t \mathbf{A}_B^{-1} \mathbf{A}_D)$

S'il existe une variable hors base x_s telle que son coût réduit soit strictement négatif alors :

- ou bien on peut diminuer indéfiniment la valeur de x_s sans sortir de l'ensemble des solutions réalisables et dans ce cas f est non bornée
- ou bien on peut déterminer une autre solution de base réalisable $\hat{\mathbf{x}}$ tel que $f(\hat{\mathbf{x}})$ soit plus petite que $f(\mathbf{x})$

Exemple d'application

Une entreprise fabrique deux biens (des pièces mécaniques par exemple). Ces fabrications nécessitent l'utilisation de deux ateliers dont les capacités de production exprimées en heures d'usinage sont de 12. Supposons que :

- Chaque unité du 1er produit nécessite 2h d'usinage dans l'atelier 1 et 1h dans l'atelier 2
- Chaque unité du 2ème produit nécessite 1h d'usinage dans l'atelier 1 et 2h dans l'atelier 2

Sachant que la marge sur le 1er produit est $p_1 = 4$ et que celle sur le 2ème produit est de $p_2 = 3$, déterminer un programme mathématique qui modélise le problème de l'optimisation de la marge de l'entreprise sous les contraintes de production décrites précédemment

Bien fondé de l'algorithme du simplexe

Théorème.

Soit \mathbf{x} une solution de base réalisable quelconque et soit le vecteur des "coûts réduits" des variables hors base : $(\mathbf{c}_D^t - \mathbf{c}_B^t \mathbf{A}_B^{-1} \mathbf{A}_D)$

S'il existe une variable hors base x_s telle que son coût réduit soit strictement négatif alors :

- ou bien on peut diminuer indéfiniment la valeur de x_s sans sortir de l'ensemble des solutions réalisables et dans ce cas f est non bornée
- ou bien on peut déterminer une autre solution de base réalisable $\hat{\mathbf{x}}$ tel que $f(\hat{\mathbf{x}})$ soit plus petite que $f(\mathbf{x})$

Théorème.

Sous l'hypothèse de non-dégénérescence, l'algorithme du simplexe converge en un nombre fini d'itérations

Exemple d'application (suite)

- PL de base :

$$\begin{aligned} \max f(x_1, x_2) &= 4x_1 + 3x_2 \\ \text{s.t.} \quad &\begin{cases} 2x_1 + x_2 \leq 12 \\ x_1 + 2x_2 \leq 12 \\ x_1 \geq 0, x_2 \geq 0 \end{cases} \end{aligned}$$

Exemple d'application (suite)

- PL de base :

$$\max f(x_1, x_2) = 4x_1 + 3x_2$$

$$s/c \begin{cases} 2x_1 + x_2 \leq 12 \\ x_1 + 2x_2 \leq 12 \\ x_1 \geq 0, x_2 \geq 0 \end{cases}$$

- PL sous forme standard :

$$\min -f(x_1, x_2) = -4x_1 - 3x_2$$

$$s/c \begin{cases} 2x_1 + x_2 + x_3 = 12 \\ x_1 + 2x_2 + x_4 = 12 \\ x_1, \dots, x_4 \geq 0 \end{cases}$$

$$\text{avec } \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}, \mathbf{c} = \begin{pmatrix} -4 \\ -3 \\ 0 \\ 0 \end{pmatrix}, \mathbf{A} = \begin{pmatrix} 2 & 1 & 1 & 0 \\ 1 & 2 & 0 & 1 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 12 \\ 12 \end{pmatrix}$$

Exemple d'application (méthode du simplexe)

Tableau simplexe initial :

		x_1	x_2	x_3	x_4	
		-4	-3	0	0	
x_3	0	2	1	1	0	12
x_4	0	1	2	0	1	12
		-4	-3	0	0	0

- 1ère itération :

- ▶ Etape 1 (coûts réduits - variable à faire entrer dans la base)
-4 est le plus petit nombre négatif de la dernière ligne (coûts réduits des variables hors base) et correspond à x_1 (variable hors base qui permettrait de diminuer le plus $-f$).

$$\begin{cases} x_3 = 12 - 2x_1 \\ x_4 = 12 - x_1 \end{cases}$$

De combien peut-on augmenter x_1 sans pour autant violer les contraintes ?

Exemple d'application (méthode du simplexe)

Tableau simplexe initial :

		x_1	x_2	x_3	x_4	
		-4	-3	0	0	
x_3	0	2	1	1	0	12
x_4	0	1	2	0	1	12
		-4	-3	0	0	0

Solution de base réalisable initiale :

- la solution de base réalisable initiale $\mathbf{x} = (0, 0, 12, 12)$ où les variables de base sont $\{x_3, x_4\}$.
- l'expression des variables de base en fonction des variables hors-base :

$$\begin{cases} 2x_1 + x_2 + x_3 = 12 \\ x_1 + 2x_2 + x_4 = 12 \end{cases} \Leftrightarrow \begin{cases} x_3 = 12 - 2x_1 - x_2 \\ x_4 = 12 - x_1 - 2x_2 \end{cases}$$
- le vecteur des coûts réduits vaut $\bar{\mathbf{c}}_{\mathbb{D}} = \underbrace{(-4)}_{x_1}, \underbrace{(-3)}_{x_2}$
- la fonction objectif vaut $f(\mathbf{x}) = -4x_1 - 3x_2 = 0$

Exemple d'application (méthode du simplexe) - suite

Tableau simplexe initial :

		x_1	x_2	x_3	x_4	
		-4	-3	0	0	
x_3	0	2	1	1	0	12
x_4	0	1	2	0	1	12
		-4	-3	0	0	0

- 1ère itération :

- ▶ Etape 2 (ratios - variable à faire sortir de la base)

$$\begin{cases} x_3 = 12 - 2x_1 \geq 0 \\ x_4 = 12 - x_1 \geq 0 \end{cases} \Leftrightarrow \begin{cases} x_1 \leq 6 \\ x_1 \leq 12 \end{cases}$$

Ces quantités correspondent aux ratios $12/2 = 6$ (1ère ligne) et $12/1 = 12$ (2ème ligne).

Le plus petit ratio correspond à la 1ère ligne : l'élément pivot est 2.

Exemple d'application (méthode du simplexe) - suite

Tableau simplexe initial :

		x_1	x_2	x_3	x_4	
		-4	-3	0	0	
x_3	0	2	1	1	0	12
x_4	0	1	2	0	1	12
		-4	-3	0	0	0

• 1ère itération :

▶ Etapes 3-4 (changement de base)

On fait entrer x_1 dans la base et on fait sortir x_3 . L'expression des variables de base en fonction des variables hors-base devient :

$$\begin{cases} 2x_1 + x_2 + x_3 = 12 \\ x_1 + 2x_2 + x_4 = 12 \end{cases} \Leftrightarrow \begin{cases} x_1 = 6 - \frac{1}{2}x_2 - \frac{1}{2}x_3 \\ x_4 = 6 - \frac{3}{2}x_2 + \frac{1}{2}x_3 \end{cases}$$

Ceci revient à appliquer les transformations lignessuivantes $l'_1 = l_1/2$ et $l'_2 = l_2 - l'_1$.

Exemple d'application (méthode du simplexe) - suite

Tableau simplexe après la 1ère itération :

		x_1	x_2	x_3	x_4	
		-4	-3	0	0	
x_1	4	1	1/2	1/2	0	6
x_4	0	0	3/2	-1/2	1	6
		0	-1	2	0	-24

• 2ème itération :

▶ Etape 1 (coûts réduits - variable à faire entrer dans la base)

-1 est le plus petit nombre négatif de la dernière ligne (coûts réduits des variables hors base) et correspond à x_2 (variable hors base qui permettrait de diminuer le plus $-f$).

$$\begin{cases} x_1 = 6 - \frac{1}{2}x_2 \\ x_4 = 6 - \frac{3}{2}x_2 \end{cases}$$

De combien peut-on augmenter x_2 sans pour autant violer les contraintes ?

Exemple d'application (méthode du simplexe) - suite

Tableau simplexe après les étapes 3 et 4 de la 1ère itération :

		x_1	x_2	x_3	x_4	
		-4	-3	0	0	
x_1	-4	1	1/2	1/2	0	6
x_4	0	0	3/2	-1/2	1	6
		0	-1	2	0	-24

Solution de base réalisable à l'issue de la 1ère itération :

- la solution de base réalisable $\mathbf{x} = (6, 0, 0, 6)$ où les variables de base sont $\{x_1, x_4\}$.
- l'expression des variables de base en fonction des variables hors-base :

$$\begin{cases} 2x_1 + x_2 + x_3 = 12 \\ x_1 + 2x_2 + x_4 = 12 \end{cases} \Leftrightarrow \begin{cases} x_1 = 6 - \frac{1}{2}x_2 - \frac{1}{2}x_3 \\ x_4 = 6 - \frac{3}{2}x_2 + \frac{1}{2}x_3 \end{cases}$$
- le vecteur des coûts réduits vaut $\bar{\mathbf{c}}_{\mathbb{D}} = \underbrace{(-1)}_{x_2}, \underbrace{(2)}_{x_3}$
- la fonction objectif vaut $f(\mathbf{x}) = -4x_1 - 3x_2 = -24$

Exemple d'application (méthode du simplexe) - suite

Tableau simplexe après la 1ère itération :

		x_1	x_2	x_3	x_4	
		-4	-3	0	0	
x_1	4	1	1/2	1/2	0	6
x_4	0	0	3/2	-1/2	1	6
		0	-1	2	0	-24

• 2ème itération :

▶ Etape 2 (ratios - variable à faire sortir de la base)

$$\begin{cases} x_1 = 6 - \frac{1}{2}x_2 \geq 0 \\ x_4 = 6 - \frac{3}{2}x_2 \geq 0 \end{cases} \Leftrightarrow \begin{cases} x_2 \leq 12 \\ x_2 \leq 4 \end{cases}$$

Ces quantités correspondent aux ratios $6/(1/2) = 12$ (1ère ligne) et $6/(3/2) = 4$ (2ème ligne).

Le plus petit ratio correspond à la 2ème ligne : l'élément pivot est 3/2.

Exemple d'application (méthode du simplexe) - suite

Tableau simplexe après la 1ère itération :

		x_1	x_2	x_3	x_4	
		-4	-3	0	0	
x_1	4	1	1/2	1/2	0	6
x_4	0	0	3/2	-1/2	1	6
		0	-1	2	0	-24

- 2ème itération :

- Etapes 3-4 (changement de base)

On fait entrer x_2 dans la base et on fait sortir x_4 . L'expression des variables de base en fonction des variables hors-base devient :

$$\begin{cases} x_1 = 6 - \frac{1}{2}x_2 - \frac{1}{2}x_3 \\ x_4 = 6 - \frac{3}{2}x_2 + \frac{1}{2}x_3 \end{cases} \Leftrightarrow \begin{cases} x_2 = 4 + \frac{1}{3}x_3 - \frac{2}{3}x_4 \\ x_1 = 4 - \frac{2}{3}x_3 + \frac{1}{3}x_4 \end{cases}$$

Ceci revient à appliquer les transformations lignes suivantes

$$l'_2 = l_2 / (3/2) \text{ et } l'_1 = l_1 - l'_2/2.$$

Exemple d'application (méthode du simplexe) - suite

Tableau simplexe après la 2ème itération :

		x_1	x_2	x_3	x_4	
		-4	-3	0	0	
x_1	-4	1	0	2/3	-1/3	4
x_2	-3	0	1	-1/3	2/3	4
		0	0	5/3	2/3	-28

- Fin de l'algorithme :

- Etape 5 (vérification de la condition d'optimalité)

Il n'y a plus de coût réduit de variables hors base qui soit négatif donc on ne peut plus diminuer la fonction objectif

Exemple d'application (méthode du simplexe) - suite

Tableau simplexe après les étapes 3 et 4 de la 2ème itération :

		x_1	x_2	x_3	x_4	
		-4	-3	0	0	
x_1	-4	1	0	2/3	-1/3	4
x_2	-3	0	1	-1/3	2/3	4
		0	0	5/3	2/3	-28

Solution de base réalisable à l'issue de la 2ème itération :

- la solution de base réalisable $\mathbf{x} = (4, 4, 0, 0)$ où les variables de base sont $\{x_1, x_2\}$.
- l'expression des variables de base en fonction des variables hors-base :

$$\begin{cases} x_2 = 4 + \frac{1}{3}x_3 - \frac{2}{3}x_4 \\ x_1 = 4 - \frac{2}{3}x_3 + \frac{1}{3}x_4 \end{cases}$$
- le vecteur des coûts réduits vaut $\bar{\mathbf{c}}_{\mathbb{D}} = (\underbrace{5/3}_{x_3}, \underbrace{2/3}_{x_4})$
- la fonction objectif vaut $f(\mathbf{x}) = -4x_1 - 3x_2 = -28$

Exemple d'application (méthode du simplexe) - suite

Tableau simplexe après la 2ème itération :

		x_1	x_2	x_3	x_4	
		-4	-3	0	0	
x_1	-4	1	0	2/3	-1/3	4
x_2	-3	0	1	-1/3	2/3	4
		0	0	5/3	2/3	-28

- Fin de l'algorithme :

- Etape 5 (vérification de la condition d'optimalité)

Il n'y a plus de coût réduit de variables hors base qui soit négatif donc on ne peut plus diminuer la fonction objectif

- Etape 6 (détermination de la solution optimale)

La solution de base réalisable $\mathbf{x}^* = (4, 4, 0, 0)$ où les variables de base sont $\{x_1, x_2\}$ et la solution optimale est $-f(\mathbf{x}^*) = -28$

Exemple d'application (méthode du simplexe) - suite

Tableau simplexe après la 2ème itération :

		x_1	x_2	x_3	x_4	
		-4	-3	0	0	
x_1	-4	1	0	$2/3$	$-1/3$	4
x_2	-3	0	1	$-1/3$	$2/3$	4
		0	0	$5/3$	$2/3$	-28

- Fin de l'algorithme :

- ▶ Etape 5 (vérification de la condition d'optimalité)
Il n'y a plus de coût réduit de variables hors base qui soit négatif donc on ne peut plus diminuer la fonction objectif
- ▶ Etape 6 (détermination de la solution optimale)
La solution de base réalisable $\mathbf{x}^* = (4, 4, 0, 0)$ où les variables de base sont $\{x_1, x_2\}$ et la solution optimale est $-f(\mathbf{x}^*) = -28$

Générer une solution de base réalisable initiale

- La méthode du simplexe débute par un tableau valide et cela nécessite donc d'avoir identifié une solution de base réalisable initiale

Rappel du Sommaire

- 1 Rappels de concepts mathématiques
- 2 Bases théoriques de l'optimisation
- 3 Optimisation sans contrainte
- 4 Optimisation linéaire avec contraintes
 - Programmation linéaire
 - Méthode du simplexe
 - Solution de base réalisable initiale et cas dégénéré
 - Dualité
- 5 Optimisation non linéaire avec contraintes

Générer une solution de base réalisable initiale

- La méthode du simplexe débute par un tableau valide et cela nécessite donc d'avoir identifié une solution de base réalisable initiale
- Précédemment, nous avons étudié un exemple où il n'y avait que des contraintes d'inégalités de type \leq

Générer une solution de base réalisable initiale

- La méthode du simplexe débute par un tableau valide et cela nécessite donc d'avoir identifié une solution de base réalisable initiale
- Précédemment, nous avons étudié un exemple où il n'y avait que des contraintes d'inégalités de type \leq
- Dans ce cas précis, une solution de base réalisable initiale est obtenue en attribuant :
 - ▶ à chaque variable d'écart la valeur du membre de droite correspondant
 - ▶ à toute autre variable du problème la valeur nulle

Générer une solution de base réalisable initiale (suite)

- Dans le cas général on a à la fois des contraintes d'inégalités \leq et \geq mais également des contraintes d'égalités

Générer une solution de base réalisable initiale

- La méthode du simplexe débute par un tableau valide et cela nécessite donc d'avoir identifié une solution de base réalisable initiale
- Précédemment, nous avons étudié un exemple où il n'y avait que des contraintes d'inégalités de type \leq
- Dans ce cas précis, une solution de base réalisable initiale est obtenue en attribuant :
 - ▶ à chaque variable d'écart la valeur du membre de droite correspondant
 - ▶ à toute autre variable du problème la valeur nulle
- Rappelons le PL sous forme standard et le tableau du simplexe initial de l'exemple précédent :

$$\begin{array}{l} \min -f(x_1, x_2) = -4x_1 - 3x_2 \\ s/c \begin{cases} 2x_1 + x_2 + x_3 = 12 \\ x_1 + 2x_2 + x_4 = 12 \\ x_1, \dots, x_4 \geq 0 \end{cases} \end{array}$$

		x_1	x_2	x_3	x_4	
		-4	-3	0	0	
x_3	0	2	1	1	0	12
x_4	0	1	2	0	1	12
		-4	-3	0	0	0

Générer une solution de base réalisable initiale (suite)

- Dans le cas général on a à la fois des contraintes d'inégalités \leq et \geq mais également des contraintes d'égalités
- Lorsqu'on est dans cette situation, une méthode permettant d'obtenir une solution de base réalisable initiale est par le biais d'ajouts de **variables artificielles**

Générer une solution de base réalisable initiale (suite)

- Dans le cas général on a à la fois des contraintes d'inégalités \leq et \geq mais également des contraintes d'égalités
- Lorsqu'on est dans cette situation, une méthode permettant d'obtenir une solution de base réalisable initiale est par le biais d'ajouts de **variables artificielles**
- On ajoute des variables artificielles à toutes les contraintes ne comportant pas de variables d'écart et ceci permet d'avoir facilement une solution de base réalisable initiale

Exemple d'un ensemble de contraintes avec différents cas

$$\begin{cases} x_1 + 2x_2 \leq 3 \\ 4x_1 + 5x_2 \geq 6 \\ 7x_1 + 8x_2 = 15 \end{cases}$$

$$\begin{cases} x_1 + 2x_2 + x_3 = 3 \\ 4x_1 + 5x_2 - x_4 = 6 \\ 7x_1 + 8x_2 = 15 \end{cases}$$

Exemple d'un ensemble de contraintes avec différents cas

$$\begin{cases} x_1 + 2x_2 \leq 3 \\ 4x_1 + 5x_2 \geq 6 \\ 7x_1 + 8x_2 = 15 \end{cases}$$

$$\begin{cases} x_1 + 2x_2 \leq 3 \\ 4x_1 + 5x_2 \geq 6 \\ 7x_1 + 8x_2 = 15 \end{cases}$$

$$\begin{cases} x_1 + 2x_2 + x_3 = 3 \\ 4x_1 + 5x_2 - x_4 = 6 \\ 7x_1 + 8x_2 = 15 \end{cases}$$

$$\begin{cases} x_1 + 2x_2 + x_3 = 3 \\ 4x_1 + 5x_2 - x_4 + x_5 = 6 \\ 7x_1 + 8x_2 + x_6 = 15 \end{cases}$$

Exemple d'un ensemble de contraintes avec différents cas

$$\begin{cases} x_1 + 2x_2 \leq 3 \\ 4x_1 + 5x_2 \geq 6 \\ 7x_1 + 8x_2 = 15 \end{cases}$$

$$\begin{cases} x_1 + 2x_2 + x_3 = 3 \\ 4x_1 + 5x_2 - x_4 = 6 \\ 7x_1 + 8x_2 = 15 \end{cases}$$

$$\begin{cases} x_1 + 2x_2 + x_3 = 3 \\ 4x_1 + 5x_2 - x_4 + x_5 = 6 \\ 7x_1 + 8x_2 + x_6 = 15 \end{cases}$$

Une solution de base réalisable est alors $x_3 = 3, x_5 = 6, x_6 = 15$ et $x_1 = x_2 = x_4 = 0$

Coûts de pénalité dans la fonction objectif liés aux variables artificielles

- Les variables d'écart ou de surplus ne changent pas la nature des contraintes ni de la fonction objectif. Par contre, ce n'est pas le cas des variables artificielles. Le nouveau système de contraintes est équivalent à l'initial uniquement si les variables artificielles sont nulles
- Ainsi pour garantir cette solution, les variables artificielles sont ajoutées dans la fonction objectif avec un coefficient très grand lorsqu'il s'agit d'un problème de minimisation. Ce coefficient est dénoté $M > 0$. On appelle cette approche la "**méthode du grand M**"

Coûts de pénalité dans la fonction objectif liés aux variables artificielles

- Les variables d'écart ou de surplus ne changent pas la nature des contraintes ni de la fonction objectif. Par contre, ce n'est pas le cas des variables artificielles. Le nouveau système de contraintes est équivalent à l'initial uniquement si les variables artificielles sont nulles

Exemple d'application

- PL de base : $\max f(x_1, x_2) = x_1 + 2x_2$ s/c $\begin{cases} x_1 + 2x_2 \leq 3 \\ 4x_1 + 5x_2 \geq 6 \\ 7x_1 + 8x_2 = 15 \\ x_1 \geq 0, x_2 \geq 0 \end{cases}$

Exemple d'application

- PL de base : $\max f(x_1, x_2) = x_1 + 2x_2 \text{ s/c } \begin{cases} x_1 + 2x_2 \leq 3 \\ 4x_1 + 5x_2 \geq 6 \\ 7x_1 + 8x_2 = 15 \\ x_1 \geq 0, x_2 \geq 0 \end{cases}$

- PL sous forme standard :

$$\min -x_1 - 2x_2 + Mx_5 + Mx_6 \text{ s/c } \begin{cases} x_1 + 2x_2 + x_3 = 3 \\ 4x_1 + 5x_2 - x_4 + x_5 = 6 \\ 7x_1 + 8x_2 + x_6 = 15 \\ x_1, \dots, x_6 \geq 0 \end{cases}$$

Exemple d'application (suite)

Tableau simplexe initial :

		x_1	x_2	x_3	x_4	x_5	x_6	
		-1	-2	0	0	M	M	
x_3	0	1	2	1	0	0	0	3
x_5	M	4	5	0	-1	1	0	6
x_6	M	7	8	0	0	0	1	15
		-1 - 11M	-2 - 13M	0	M	0	0	21M

- 1ère itération :

- ▶ Base courante : $\mathbb{B} = \{3, 5, 6\}$
- 1 $s = 2$
- 2 Ratios : (1.5, 1.2, 1.875) et donc $t = 2$ et $\mathbb{B}_2 = 5$
- 3 On fait entrer x_2 dans la base et on fait sortir x_5 de la base
- 4 Mise à jour du tableau

Exemple d'application

- PL de base : $\max f(x_1, x_2) = x_1 + 2x_2 \text{ s/c } \begin{cases} x_1 + 2x_2 \leq 3 \\ 4x_1 + 5x_2 \geq 6 \\ 7x_1 + 8x_2 = 15 \\ x_1 \geq 0, x_2 \geq 0 \end{cases}$

- PL sous forme standard :

$$\min -x_1 - 2x_2 + Mx_5 + Mx_6 \text{ s/c } \begin{cases} x_1 + 2x_2 + x_3 = 3 \\ 4x_1 + 5x_2 - x_4 + x_5 = 6 \\ 7x_1 + 8x_2 + x_6 = 15 \\ x_1, \dots, x_6 \geq 0 \end{cases}$$

- Tableau du simplexe initial :

		x_1	x_2	x_3	x_4	x_5	x_6	
		-1	-2	0	0	M	M	
x_3	0	1	2	1	0	0	0	3
x_5	M	4	5	0	-1	1	0	6
x_6	M	7	8	0	0	0	1	15
		-1 - 11M	-2 - 13M	0	M	0	0	21M

Exemple d'application (suite)

Tableau simplexe après la 1ère itération :

		x_1	x_2	x_3	x_4	x_5	x_6	
		-1	-2	0	0	M	M	
x_3	0	-0.6	0.0	1.0	0.4	-0.4	0.0	0.6
x_2	-2	0.8	1.0	0.0	-0.2	0.2	0.0	1.2
x_6	M	0.6	0.0	0.0	1.6	-1.6	1.0	5.4
		0.6 - 0.6M	0	0	-0.4 - 1.6M	2.6M + 0.4	0	5.4M - 2.4

- 2ème itération :

- ▶ Base courante : $\mathbb{B} = \{3, 2, 6\}$
- 1 $s = 4$
- 2 Ratios : (1.5, nan, 3.375) et donc $t = 1$ et $\mathbb{B}_1 = 3$
- 3 On fait entrer x_4 dans la base et on fait sortir x_3 de la base
- 4 Mise à jour du tableau

Exemple d'application (suite)

Tableau simplexe après la 2ème itération :

	x_1	x_2	x_3	x_4	x_5	x_6	
	-1	-2	0	0	M	M	
x_4 0	-1.5	0.0	2.5	1.0	-1.0	0.0	1.5
x_2 -2	0.5	1.0	0.5	0.0	0.0	0.0	1.5
x_6 M	3.0	0.0	-4.0	0.0	0.0	1.0	3.0
	$-3M$	0	$4M + 1$	0	M	0	$3M - 3$

• 3ème itération :

- ▶ Base courante : $\mathbb{B} = \{4, 2, 6\}$
- 1 $s = 1$
- 2 Ratios : $(nan, 3, 1)$ et donc $t = 3$ et $\mathbb{B}_3 = 6$
- 3 On fait entrer x_1 dans la base et on fait sortir x_6 de la base
- 4 Mise à jour du tableau

Dégénérescence

- Au cours du déroulement de l'algorithme du simplexe, il arrive qu'une variable de la base soit nulle. On dit alors que la solution de base réalisable est dégénérée

Exemple d'application (suite)

Tableau simplexe après la 3ème itération :

	x_1	x_2	x_3	x_4	x_5	x_6	
	-1	-2	0	0	M	M	
x_4 0	0.0	0.0	0.5	1.0	-1.0	0.5	3.0
x_2 -2	0.0	1.0	1.167	0.0	0.0	-0.167	1.0
x_1 -1	1.0	0.0	-1.33	0.0	0.0	0.33	1.0
	0	0	1	0	M	M	-3

• 4ème itération :

- ▶ Base courante : $\mathbb{B} = \{4, 2, 1\}$
- 1 Tous les coûts réduits sont positifs ou nuls donc l'optimum est atteint
- 2 $\mathbf{x}^* = (1, 1, 0, 3, 0, 0)$ et $-f(\mathbf{x}^*) = -3$

Dégénérescence

- Au cours du déroulement de l'algorithme du simplexe, il arrive qu'une variable de la base soit nulle. On dit alors que la solution de base réalisable est dégénérée
- Dans ce cas, si on fait sortir de la base cette variable, cela ne permet pas à la variable entrante de diminuer la fonction objectif. Donc lors de cette étape dégénérée la solution n'est pas améliorée

Dégénérescence

- Au cours du déroulement de l'algorithme du simplexe, il arrive qu'une variable de la base soit nulle. On dit alors que la solution de base réalisable est dégénérée
- Dans ce cas, si on fait sortir de la base cette variable, cela ne permet pas à la variable entrante de diminuer la fonction objectif. Donc lors de cette étape dégénérée la solution n'est pas améliorée
- Mais il faut le faire malgré tout car la solution de base réalisable étant différente, la fonction objectif peut à nouveau être diminuée après quelques itérations

Dégénérescence (suite)

- Pour éviter les phénomènes de cycle une méthode simple est la règle de Bland. Il s'agit d'une règle simple qui précise la variable à faire sortir et celle à faire entrer lorsque plusieurs choix sont possibles. Vis à vis de l'algorithme les modifications sont les suivantes (toutes choses étant égales par ailleurs) :

Dégénérescence

- Au cours du déroulement de l'algorithme du simplexe, il arrive qu'une variable de la base soit nulle. On dit alors que la solution de base réalisable est dégénérée
- Dans ce cas, si on fait sortir de la base cette variable, cela ne permet pas à la variable entrante de diminuer la fonction objectif. Donc lors de cette étape dégénérée la solution n'est pas améliorée
- Mais il faut le faire malgré tout car la solution de base réalisable étant différente, la fonction objectif peut à nouveau être diminuée après quelques itérations
- Toutefois il arrive (dans des cas plutôt rares) que l'algorithme passe d'une solution de base réalisable dégénérée à une autre pour finalement revenir sur celle du départ : on est alors bloqué dans un cycle

Dégénérescence (suite)

- Pour éviter les phénomènes de cycle une méthode simple est la règle de Bland. Il s'agit d'une règle simple qui précise la variable à faire sortir et celle à faire entrer lorsque plusieurs choix sont possibles. Vis à vis de l'algorithme les modifications sont les suivantes (toutes choses étant égales par ailleurs) :

- 1 Dans la ligne des coûts réduits calculés par $\mathbf{c}^t - \mathbf{c}_B^t \mathbf{A}$, Déterminer $s = \min\{j \in \mathbb{D} : [\mathbf{c}^t - \mathbf{c}_B^t \mathbf{A}]_j < 0\}$

Dégénérescence (suite)

- Pour éviter les phénomènes de cycle une méthode simple est la règle de Bland. Il s'agit d'une règle simple qui précise la variable à faire sortir et celle à faire entrer lorsque plusieurs choix sont possibles. Vis à vis de l'algorithme les modifications sont les suivantes (toutes choses étant égales par ailleurs) :

- 1 Dans la ligne des coûts réduits calculés par $\mathbf{c}^t - \mathbf{c}_{\mathbb{B}}^t \mathbf{A}$, Déterminer $s = \min\{j \in \mathbb{D} : [\mathbf{c}^t - \mathbf{c}_{\mathbb{B}}^t \mathbf{A}]_j < 0\}$
- 2 A l'aide de la colonne \mathbf{a}_s de \mathbf{A} et celle de \mathbf{b} , Déterminer $t = \min\{j \in \mathbb{B} : \frac{b_j}{a_{js}} = \min_{i \in \{1, \dots, m\} : a_{is} > 0} \{\frac{b_i}{a_{is}}\}\}$,

Rappel du Sommaire

- 1 Rappels de concepts mathématiques
- 2 Bases théoriques de l'optimisation
- 3 Optimisation sans contrainte
- 4 Optimisation linéaire avec contraintes
 - Programmation linéaire
 - Méthode du simplexe
 - Solution de base réalisable initiale et cas dégénéré
 - Dualité
- 5 Optimisation non linéaire avec contraintes

Dégénérescence (suite)

- Pour éviter les phénomènes de cycle une méthode simple est la règle de Bland. Il s'agit d'une règle simple qui précise la variable à faire sortir et celle à faire entrer lorsque plusieurs choix sont possibles. Vis à vis de l'algorithme les modifications sont les suivantes (toutes choses étant égales par ailleurs) :

- 1 Dans la ligne des coûts réduits calculés par $\mathbf{c}^t - \mathbf{c}_{\mathbb{B}}^t \mathbf{A}$, Déterminer $s = \min\{j \in \mathbb{D} : [\mathbf{c}^t - \mathbf{c}_{\mathbb{B}}^t \mathbf{A}]_j < 0\}$
- 2 A l'aide de la colonne \mathbf{a}_s de \mathbf{A} et celle de \mathbf{b} , Déterminer $t = \min\{j \in \mathbb{B} : \frac{b_j}{a_{js}} = \min_{i \in \{1, \dots, m\} : a_{is} > 0} \{\frac{b_i}{a_{is}}\}\}$,

- Autrement dit :

- ▶ Si plusieurs variables hors base ont un coût réduit négatif alors on choisit celle dont l'indice est le plus petit
- ▶ Si plusieurs variables de base peuvent sortir de la base car ayant le même ratio alors on choisit celle dont l'indice est le plus petit

Exemple introductif

- Reprenons le problème de maximisation suivant :

$$\begin{aligned} \max f(x_1, x_2) &= 4x_1 + 3x_2 \\ \text{s.l.c.} \quad &\begin{cases} l_1 : 2x_1 + x_2 \leq 12 \\ l_2 : x_1 + 2x_2 \leq 12 \\ x_1 \geq 0, x_2 \geq 0 \end{cases} \end{aligned}$$

Exemple introductif

- Reprenons le problème de maximisation suivant :

$$\begin{aligned} \max f(x_1, x_2) &= 4x_1 + 3x_2 \\ \text{s.t.} \quad &\begin{cases} l_1 : 2x_1 + x_2 \leq 12 \\ l_2 : x_1 + 2x_2 \leq 12 \\ x_1 \geq 0, x_2 \geq 0 \end{cases} \end{aligned}$$

- Multiplications par 3 la première inégalité :
 $3l_1 : 6x_1 + 3x_2 \leq 36$

Exemple introductif

- Reprenons le problème de maximisation suivant :

$$\begin{aligned} \max f(x_1, x_2) &= 4x_1 + 3x_2 \\ \text{s.t.} \quad &\begin{cases} l_1 : 2x_1 + x_2 \leq 12 \\ l_2 : x_1 + 2x_2 \leq 12 \\ x_1 \geq 0, x_2 \geq 0 \end{cases} \end{aligned}$$

- Multiplications par 3 la première inégalité :
 $3l_1 : 6x_1 + 3x_2 \leq 36$
- Si on compare cette inégalité avec la fonction objectif, on peut voir que : $f(\mathbf{x}) = 4x_1 + 3x_2 \leq 6x_1 + 3x_2 \leq \mathbf{36}$
 On obtient ainsi une **borne supérieure de $f(\mathbf{x}^*)$**

Exemple introductif

- Reprenons le problème de maximisation suivant :

$$\begin{aligned} \max f(x_1, x_2) &= 4x_1 + 3x_2 \\ \text{s.t.} \quad &\begin{cases} l_1 : 2x_1 + x_2 \leq 12 \\ l_2 : x_1 + 2x_2 \leq 12 \\ x_1 \geq 0, x_2 \geq 0 \end{cases} \end{aligned}$$

- Multiplications par 3 la première inégalité :
 $3l_1 : 6x_1 + 3x_2 \leq 36$
- Si on compare cette inégalité avec la fonction objectif, on peut voir que : $f(\mathbf{x}) = 4x_1 + 3x_2 \leq 6x_1 + 3x_2 \leq \mathbf{36}$
 On obtient ainsi une **borne supérieure de $f(\mathbf{x}^*)$**
- On peut faire mieux en prenant :
 $2l_1 + \frac{1}{2}l_2 : 4.5x_1 + 3x_2 \leq 30$

Exemple introductif

- Reprenons le problème de maximisation suivant :

$$\begin{aligned} \max f(x_1, x_2) &= 4x_1 + 3x_2 \\ \text{s.t.} \quad &\begin{cases} l_1 : 2x_1 + x_2 \leq 12 \\ l_2 : x_1 + 2x_2 \leq 12 \\ x_1 \geq 0, x_2 \geq 0 \end{cases} \end{aligned}$$

- Multiplications par 3 la première inégalité :
 $3l_1 : 6x_1 + 3x_2 \leq 36$
- Si on compare cette inégalité avec la fonction objectif, on peut voir que : $f(\mathbf{x}) = 4x_1 + 3x_2 \leq 6x_1 + 3x_2 \leq \mathbf{36}$
 On obtient ainsi une **borne supérieure de $f(\mathbf{x}^*)$**
- On peut faire mieux en prenant :
 $2l_1 + \frac{1}{2}l_2 : 4.5x_1 + 3x_2 \leq 30$
- On a donc :
 $f(\mathbf{x}) = 4x_1 + 3x_2 \leq 4.5x_1 + 3x_2 \leq \mathbf{30}$
 On obtient une meilleure borne supérieure de $f(\mathbf{x}^*)$

Exemple introductif (suite)

- Reprenons le problème de maximisation suivant :

$$\begin{aligned} \max f(x_1, x_2) &= 4x_1 + 3x_2 \\ \text{s.t.} \quad &\begin{cases} l_1 : 2x_1 + x_2 \leq 12 \\ l_2 : x_1 + 2x_2 \leq 12 \\ x_1 \geq 0, x_2 \geq 0 \end{cases} \end{aligned}$$

- On peut faire encore mieux que précédemment :

$$\frac{5}{3}l_1 + \frac{2}{3}l_2 : 4x_1 + 3x_2 \leq 28$$

Généralisation de l'exemple précédent

- Considérons un PL **non standard**, de maximisation et avec des contraintes d'**inégalités \leq uniquement**

$$\begin{aligned} l_1 : & a_{11}x_1 + \dots + a_{1n}x_n \leq b_1 \\ l_2 : & a_{21}x_1 + \dots + a_{2n}x_n \leq b_2 \\ & \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\ l_m : & a_{m1}x_1 + \dots + a_{mn}x_n \leq b_m \end{aligned}$$

Exemple introductif (suite)

- Reprenons le problème de maximisation suivant :

$$\begin{aligned} \max f(x_1, x_2) &= 4x_1 + 3x_2 \\ \text{s.t.} \quad &\begin{cases} l_1 : 2x_1 + x_2 \leq 12 \\ l_2 : x_1 + 2x_2 \leq 12 \\ x_1 \geq 0, x_2 \geq 0 \end{cases} \end{aligned}$$

- On peut faire encore mieux que précédemment :

$$\frac{5}{3}l_1 + \frac{2}{3}l_2 : 4x_1 + 3x_2 \leq 28$$

- On a dans ce cas :

$$f(\mathbf{x}) = 4x_1 + 3x_2 \leq 28.$$

On obtient une meilleure borne supérieure de $f(\mathbf{x}^*)$ (qui s'avère ici être la solution optimale)

Généralisation de l'exemple précédent

- Considérons un PL **non standard**, de maximisation et avec des contraintes d'**inégalités \leq uniquement**

$$\begin{aligned} l_1 : & a_{11}x_1 + \dots + a_{1n}x_n \leq b_1 \quad (\times y_1 \geq 0) \\ l_2 : & a_{21}x_1 + \dots + a_{2n}x_n \leq b_2 \quad (\times y_2 \geq 0) \\ & \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\ l_m : & a_{m1}x_1 + \dots + a_{mn}x_n \leq b_m \quad (\times y_m \geq 0) \end{aligned}$$

Généralisation de l'exemple précédent

- Considérons un PL **non standard**, de maximisation et avec des contraintes d'**inégalités** \leq **uniquement**

$$\begin{aligned}
 l_1 : & \quad a_{11}x_1 + \dots + a_{1n}x_n \leq b_1 \quad (\times y_1 \geq 0) \\
 l_2 : & \quad a_{21}x_1 + \dots + a_{2n}x_n \leq b_2 \quad (\times y_2 \geq 0) \\
 & \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\
 l_m : & \quad a_{m1}x_1 + \dots + a_{mn}x_n \leq b_m \quad (\times y_m \geq 0)
 \end{aligned}$$

$$\sum_i y_i l_i : \quad \sum_i y_i a_{i1} x_1 + \dots + \sum_i y_i a_{in} x_n \leq \sum_i y_i b_i$$

Programme dual du PL de l'exemple (cas particulier)

- Pour déterminer la **plus petite borne supérieure possible** de la fonction objectif il faut résoudre le **PL** suivant :

$$\begin{aligned}
 \min g(\mathbf{y}) &= \sum_{i=1}^m b_i y_i \\
 \text{slc} \quad & \begin{cases} \sum_{i=1}^m y_i a_{i1} \geq c_1 \\ \vdots \\ \sum_{i=1}^m y_i a_{in} \geq c_n \\ y_1 \geq 0, \dots, y_m \geq 0 \end{cases}
 \end{aligned}$$

Généralisation de l'exemple précédent

- Considérons un PL **non standard**, de maximisation et avec des contraintes d'**inégalités** \leq **uniquement**

$$\begin{aligned}
 l_1 : & \quad a_{11}x_1 + \dots + a_{1n}x_n \leq b_1 \quad (\times y_1 \geq 0) \\
 l_2 : & \quad a_{21}x_1 + \dots + a_{2n}x_n \leq b_2 \quad (\times y_2 \geq 0) \\
 & \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\
 l_m : & \quad a_{m1}x_1 + \dots + a_{mn}x_n \leq b_m \quad (\times y_m \geq 0)
 \end{aligned}$$

$$\sum_i y_i l_i : \quad \sum_i y_i a_{i1} x_1 + \dots + \sum_i y_i a_{in} x_n \leq \sum_i y_i b_i$$

- De plus si : $x_1, \dots, x_n \geq 0$; $c_1 \leq \sum_{i=1}^m y_i a_{i1}$; \dots ; $c_n \leq \sum_{i=1}^m y_i a_{in}$ alors :

$$f(\mathbf{x}) = \mathbf{c}^t \mathbf{x} = c_1 x_1 + \dots + c_n x_n \leq \sum_{i=1}^m y_i a_{i1} x_1 + \dots + \sum_{i=1}^m y_i a_{in} x_n \leq \sum_{i=1}^m y_i b_i$$

Programme dual du PL de l'exemple (cas particulier) (suite)

- Dans le cas particulier où les contraintes linéaires sont toutes du type \leq nous avons le formalisme matriciel suivant :

$$\begin{array}{ccc}
 \text{Maximiser} & f(\mathbf{x}) = \mathbf{c}^t \mathbf{x} & \text{Minimiser} & g(\mathbf{y}) = \mathbf{b}^t \mathbf{y} \\
 \text{slc} & \mathbf{A} \mathbf{x} \leq \mathbf{b} & \text{et} & \text{slc} & \mathbf{A}^t \mathbf{y} \geq \mathbf{c} \\
 \text{avec} & \mathbf{x} \geq \mathbf{0} & & \text{avec} & \mathbf{y} \geq \mathbf{0} \\
 \underbrace{\hspace{10em}} & & & \underbrace{\hspace{10em}} & \\
 \text{Primal (PLP)} & & & \text{Dual (PLD)} &
 \end{array}$$

où \mathbf{A}^t est la transposée de \mathbf{A}

Programme dual d'un PL quelconque (cas général)

- A tout **PL quelconque** on peut lui associer un **PL dual, noté PLD**. Le PL initial est alors dénommé le PL primal et sera noté PLP. Le tableau suivant résume les correspondances entre primal et dual.

Primal (PLP)		Dual (PLD)
Problème de maximisation	↔	Problème de minimisation
A matrice des coefficients des contraintes	↔	A^t matrice des coefficients des contraintes
Variable $x_j \geq 0$	↔	$j^{\text{ème}}$ contrainte de type \geq
Variable $x_j \in \mathbb{R}$	↔	$j^{\text{ème}}$ contrainte de type $=$
Variable $x_j \leq 0$	↔	$j^{\text{ème}}$ contrainte de type \leq
$i^{\text{ème}}$ contrainte de type \leq	↔	Variable $y_i \geq 0$
$i^{\text{ème}}$ contrainte de type $=$	↔	Variable $y_i \in \mathbb{R}$
$i^{\text{ème}}$ contrainte de type \geq	↔	Variable $y_i \leq 0$

Programme dual du PL de l'exemple (cas particulier)

$$\text{Minimiser } g(\mathbf{y}) = \mathbf{b}^t \mathbf{y}$$

$$\text{s.l.c. } \mathbf{A}^t \mathbf{y} \geq \mathbf{c} \quad \text{et}$$

$$\text{avec } \mathbf{y} \geq \mathbf{0}$$

Primal (PLP)

$$\text{Maximiser } f(\mathbf{x}) = \mathbf{c}^t \mathbf{x}$$

$$\text{s.l.c. } \mathbf{A} \mathbf{x} \leq \mathbf{b}$$

$$\text{avec : } \mathbf{x} \geq \mathbf{0}$$

Dual (PLD)

- Le dual du dual est le primal (ici on passe d'un problème de minimisation à un problème de maximisation)

Programme dual d'un PL quelconque (cas général)

- A tout **PL quelconque** on peut lui associer un **PL dual, noté PLD**. Le PL initial est alors dénommé le PL primal et sera noté PLP. Le tableau suivant résume les correspondances entre primal et dual.

Primal (PLP)		Dual (PLD)
Problème de maximisation	↔	Problème de minimisation
A matrice des coefficients des contraintes	↔	A^t matrice des coefficients des contraintes
Variable $x_j \geq 0$	↔	$j^{\text{ème}}$ contrainte de type \geq
Variable $x_j \in \mathbb{R}$	↔	$j^{\text{ème}}$ contrainte de type $=$
Variable $x_j \leq 0$	↔	$j^{\text{ème}}$ contrainte de type \leq
$i^{\text{ème}}$ contrainte de type \leq	↔	Variable $y_i \geq 0$
$i^{\text{ème}}$ contrainte de type $=$	↔	Variable $y_i \in \mathbb{R}$
$i^{\text{ème}}$ contrainte de type \geq	↔	Variable $y_i \leq 0$

- On a la propriété que **le dual du dual est le primal**

Exemple de dualisation dans le cas général

$$\begin{array}{l} \text{Minimiser } y_1 + 4y_2 + 3y_3 \\ \text{s.l.c. } \begin{cases} y_1 + 2y_2 + y_3 \geq 2 \\ -y_1 + 3y_2 + y_3 \geq 2 \end{cases} \\ \text{avec } y_1 \geq 0, y_2 \leq 0, y_3 \in \mathbb{R} \end{array}$$

Primal

Exemple de dualisation dans le cas général

$$\underbrace{\begin{array}{l} \text{Minimiser} \quad y_1 + 4y_2 + 3y_3 \\ \text{slc} \quad \begin{cases} y_1 + 2y_2 + y_3 \geq 2 \\ -y_1 + 3y_2 + y_3 \geq 2 \end{cases} \\ \text{avec} \quad y_1 \geq 0, y_2 \leq 0, y_3 \in \mathbb{R} \end{array}}_{\text{Primal}}$$

- Le dual de ce PL est le suivant :

$$\underbrace{\begin{array}{l} \text{Maximiser} \quad 2x_1 + 2x_2 \\ \text{slc} \quad \begin{cases} x_1 - x_2 \leq 1 \\ 2x_1 + 3x_2 \geq 4 \\ x_1 + x_2 = 3 \end{cases} \\ \text{avec} \quad x_1 \geq 0, x_2 \geq 0 \end{array}}_{\text{Dual}}$$

Théorèmes sur la dualité

Lemme.

Soit \hat{x} et \hat{y} deux solutions réalisables du primal (problème de minimisation mis sous forme standard) et du dual associé (problème de maximisation) alors :

$$\underbrace{f(\hat{x}) = \mathbf{c}^t \hat{x}} \geq \underbrace{\mathbf{b}^t \hat{y} = g(\hat{y})}$$

Programme dual d'un PL standard

- On suppose un PL de minimisation mis sous forme standard. Le dual dans ce cas est le suivant :

$$\begin{array}{l} \text{Minimiser} \quad f(\mathbf{x}) = \mathbf{c}^t \mathbf{x} \\ \text{slc} \quad \mathbf{Ax} = \mathbf{b} \\ \text{avec} \quad \mathbf{x} \geq \mathbf{0} \end{array} \quad \text{et} \quad \begin{array}{l} \text{Maximiser} \quad g(\mathbf{y}) = \mathbf{b}^t \mathbf{y} \\ \text{slc} \quad \mathbf{A}^t \mathbf{y} \leq \mathbf{c} \\ \text{avec} \quad \mathbf{y} \in \mathbb{R}^n \end{array}$$

$\underbrace{\hspace{10em}}_{\text{Primal (PLP)}}$
 $\underbrace{\hspace{10em}}_{\text{Dual (PLD)}}$

Théorèmes sur la dualité

Lemme.

Soit \hat{x} et \hat{y} deux solutions réalisables du primal (problème de minimisation mis sous forme standard) et du dual associé (problème de maximisation) alors :

$$\underbrace{f(\hat{x}) = \mathbf{c}^t \hat{x}} \geq \underbrace{\mathbf{b}^t \hat{y} = g(\hat{y})}$$

Démonstration.

$$(\mathbf{Ax} = \mathbf{b}) \Rightarrow (\hat{y}^t \mathbf{Ax} = \hat{y}^t \mathbf{b})$$

Comme de plus, $\hat{x} \geq \mathbf{0}, \mathbf{A}^t \hat{y} \leq \mathbf{c}$ (règle de dualité) nous avons alors

$$(\mathbf{A}^t \hat{y})^t \hat{x} \leq \mathbf{c}^t \hat{x} \text{ et enfin } \hat{y}^t \mathbf{b} \leq \mathbf{c}^t \hat{x}. \quad \square$$

Théorèmes sur la dualité

Lemme.

Soit $\hat{\mathbf{x}}$ et $\hat{\mathbf{y}}$ deux solutions réalisables du primal (problème de minimisation mis sous forme standard) et du dual associé (problème de maximisation) alors :

$$\underbrace{f(\hat{\mathbf{x}}) = \mathbf{c}^t \hat{\mathbf{x}}}_{\text{primal}} \geq \underbrace{\mathbf{b}^t \hat{\mathbf{y}} = g(\hat{\mathbf{y}})}_{\text{dual}}$$

Démonstration.

$$(\mathbf{A}\hat{\mathbf{x}} = \mathbf{b}) \Rightarrow (\hat{\mathbf{y}}^t \mathbf{A}\hat{\mathbf{x}} = \hat{\mathbf{y}}^t \mathbf{b})$$

Comme de plus, $\hat{\mathbf{x}} \geq \mathbf{0}$, $\mathbf{A}^t \hat{\mathbf{y}} \leq \mathbf{c}$ (règle de dualité) nous avons alors

$$(\mathbf{A}^t \hat{\mathbf{y}})^t \hat{\mathbf{x}} \leq \mathbf{c}^t \hat{\mathbf{x}} \text{ et enfin } \hat{\mathbf{y}}^t \mathbf{b} \leq \mathbf{c}^t \hat{\mathbf{x}}. \quad \square$$

- Ici, une solution réalisable du primal (resp dual) permet de définir une borne supérieure (resp inférieure) du dual (resp primal)

Applications des propriétés de dualité

- Si le primal a beaucoup de contraintes et moins de variables alors la méthode du simplexe sera plus efficace sur le dual

Théorèmes sur la dualité (suite)

Corollaire.

Si \mathbf{x}^* et \mathbf{y}^* sont deux solutions réalisables du primal et du dual associé tels que $\mathbf{c}^t \mathbf{x}^* = \mathbf{b}^t \mathbf{y}^*$ alors \mathbf{x}^* est optimum du primal et \mathbf{y}^* optimum du dual

Théorème.

Etant donné un PL primal PLP et le PL dual associé PLD :

- Si PLP et PLD ont des solutions réalisables alors chacun d'eux a une solution optimale \mathbf{x}^* et \mathbf{y}^* tel que :

$$f(\mathbf{x}^*) = \mathbf{c}^t \mathbf{x}^* = \mathbf{b}^t \mathbf{y}^* = g(\mathbf{y}^*)$$

- Si l'un d'eux a un optimum non borné, l'autre n'a pas de solution réalisable

Applications des propriétés de dualité

- Si le primal a beaucoup de contraintes et moins de variables alors la méthode du simplexe sera plus efficace sur le dual
- A l'optimum, le tableau du simplexe fournit à la fois la solution optimale pour le primal et pour le dual

Applications des propriétés de dualité

- Si le primal a beaucoup de contraintes et moins de variables alors la méthode du simplexe sera plus efficace sur le dual
- A l'optimum, le tableau du simplexe fournit à la fois la solution optimale pour le primal et pour le dual
- La solution optimale du dual peut s'interpréter d'un point de vue économique (en supposant ici que le primal est un problème de maximisation) :

$$\text{De } f(\mathbf{x}^*) = \underbrace{\sum_j c_j x_j^*}_{\mathbf{c}^t \mathbf{x}^*} = \underbrace{\sum_i b_i y_i^*}_{\mathbf{b}^t \mathbf{y}^*} = g(\mathbf{y}^*) \text{ nous en déduisons :}$$

- $\frac{\partial f}{\partial b_i}(\mathbf{x}^*) = y_i^*$
- y_i^* représente le **prix marginal** de la ressource i à l'optimum
- Autrement dit, y_i^* représente l'augmentation potentielle de la valeur optimale du problème si la ressource i , actuellement limitée à b_i , se voyait augmenter d'une unité.

Exemple

Une entreprise fabrique deux biens (des pièces mécaniques par exemple). Ces fabrications nécessitent l'utilisation de deux ateliers dont les capacités de production exprimées en heures d'usinage sont de 12. Supposons que :

- Chaque unité du 1er produit nécessite 2h d'usinage dans l'atelier 1 et 1h dans l'atelier 2.
- Chaque unité du 2ème produit nécessite 1h d'usinage dans l'atelier 1 et 2h dans l'atelier 2.

Sachant que la marge sur le 1er produit est $p_1 = 4$ et que celle sur le 2ème produit est de $p_2 = 3$, déterminer un programme mathématique qui modélise le problème de l'optimisation de la marge de l'entreprise sous les contraintes de production décrites précédemment.

Exemple (suite)

$$\begin{array}{ll} \text{Maximiser} & f(x_1, x_2) = 4x_1 + 3x_2 \\ \text{slc} & \begin{cases} 2x_1 + x_2 \leq 12 \\ x_1 + 2x_2 \leq 12 \end{cases} \\ \text{avec} & x_1 \geq 0, x_2 \geq 0 \end{array}$$

Primal (PLP)

$$\begin{array}{ll} \text{Maximiser} & f(x_1, x_2) = 4x_1 + 3x_2 \\ \text{slc} & \begin{cases} 2x_1 + x_2 \leq 12 \\ x_1 + 2x_2 \leq 12 \end{cases} \\ \text{avec} & x_1 \geq 0, x_2 \geq 0 \end{array}$$

Primal (PLP)

- Le dual est le suivant :

$$\begin{array}{ll} \text{Minimiser} & g(y_1, y_2) = 12y_1 + 12y_2 \\ \text{slc} & \begin{cases} 2y_1 + y_2 \geq 4 \\ y_1 + 2y_2 \geq 3 \end{cases} \\ \text{avec} & y_1 \geq 0, y_2 \geq 0 \end{array}$$

Dual (PLD)

Exemple (suite)

Le tableau à l'optimum est :

		x_1	x_2	x_3	x_4	
		-4	-3	0	0	
x_1	4	1	0	$2/3$	$-1/3$	4
x_2	3	0	1	$-1/3$	$2/3$	4
		0	0	$5/3$	$2/3$	-28

Exemple (suite)

Le tableau à l'optimum est :

		x_1	x_2	x_3	x_4	
		-4	-3	0	0	
x_1	4	1	0	$2/3$	$-1/3$	4
x_2	3	0	1	$-1/3$	$2/3$	4
		0	0	$5/3$	$2/3$	-28

- Sur sa dernière ligne l'opposé des "coûts" réduits correspond à la solution optimale du dual : $\mathbf{y}^* = (5/3, 2/3)$.
- $y_1^* = 5/3$ est le "prix" marginal de l'atelier 1 tandis que $y_2^* = 2/3$ est le "prix" marginal de l'atelier 2.

Exemple (suite)

Le tableau à l'optimum est :

		x_1	x_2	x_3	x_4	
		-4	-3	0	0	
x_1	4	1	0	$2/3$	$-1/3$	4
x_2	3	0	1	$-1/3$	$2/3$	4
		0	0	$5/3$	$2/3$	-28

- Sur sa dernière ligne l'opposé des "coûts" réduits correspond à la solution optimale du dual : $\mathbf{y}^* = (5/3, 2/3)$.

Rappel du Sommaire

- 1 Rappels de concepts mathématiques
- 2 Bases théoriques de l'optimisation
- 3 Optimisation sans contrainte
- 4 Optimisation linéaire avec contraintes
- 5 **Optimisation non linéaire avec contraintes**
- 6 Optimisation convexe
- 7 Algorithmes pour problèmes contraints

Introduction

- Nous étudions désormais les méthodes permettant de résoudre les problèmes d'optimisation non linéaires avec contraintes qui peuvent être formulés de la manière suivante :

$$\begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.l.c.} & h_i(\mathbf{x}) = 0 \quad i = 1, \dots, m \\ & g_j(\mathbf{x}) \leq 0 \quad j = 1, \dots, p \end{array}$$

où $\mathbf{x} \in \mathbb{R}^n$, $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $h_i : \mathbb{R}^n \rightarrow \mathbb{R}$, $g_j : \mathbb{R}^n \rightarrow \mathbb{R}$ et $m \leq n$

- Lorsqu'on exprime le problème par des notations matricielles on parle alors de **forme standard** :

$$\begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.l.c.} & h(\mathbf{x}) = \mathbf{0} \\ & g(\mathbf{x}) \leq \mathbf{0} \end{array}$$

où $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ et $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$

Exemple de problème non linéaire contraint

- Problème initial :

$$\begin{array}{ll} \min & f(x_1, x_2) = (x_1 - 1)^2 + x_2 - 2 \\ \text{s.l.c.} & \begin{cases} x_2 - x_1 = 1 \\ x_1 + x_2 \leq 2 \end{cases} \end{array}$$

- Forme standard :

$$\begin{array}{ll} \min & f(x_1, x_2) = (x_1 - 1)^2 + x_2 - 2 \\ \text{s.l.c.} & \begin{cases} x_2 - x_1 - 1 = 0 \\ x_1 + x_2 - 2 \leq 0 \end{cases} \end{array}$$

- La solution de ce problème est $\mathbf{x}^* = \left(\frac{1}{2}, \frac{3}{2}\right)$ et $f(\mathbf{x}^*) = -\frac{1}{4}$

Introduction (suite)

- Rappelons les définitions suivantes :

Définition. (Solution et ensemble de solutions réalisables)

Un vecteur \mathbf{x} satisfaisant aux contraintes $h(\mathbf{x}) = \mathbf{0}$ et $g(\mathbf{x}) \leq \mathbf{0}$ est appelée une **solution réalisable**. L'ensemble des solutions réalisables est noté \mathbb{S} et est donc défini par :

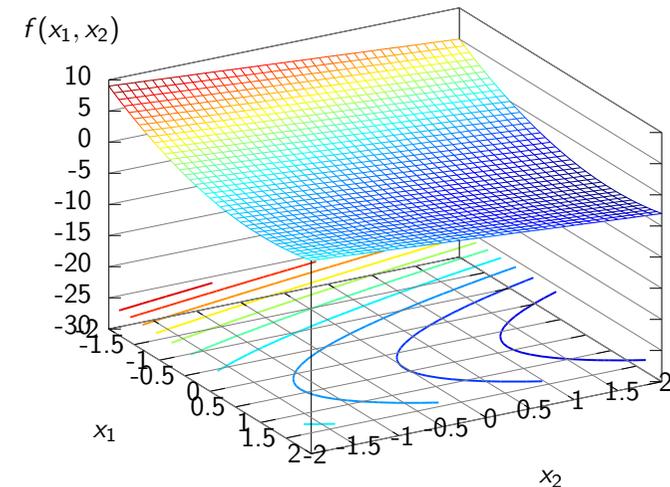
$$\mathbb{S} = \{\mathbf{x} \in \mathbb{R}^n : h(\mathbf{x}) = \mathbf{0}, g(\mathbf{x}) \leq \mathbf{0}\}$$

- Rappelons également que nous avons traité un cas particulier, la programmation linéaire, pour lequel les fonctions f , h_i et g_j sont toutes des formes linéaires ce que nous pouvons exprimer par :

$$\begin{array}{ll} \min & \mathbf{c}^t \mathbf{x} \\ \text{s.l.c.} & \mathbf{A} \mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{array}$$

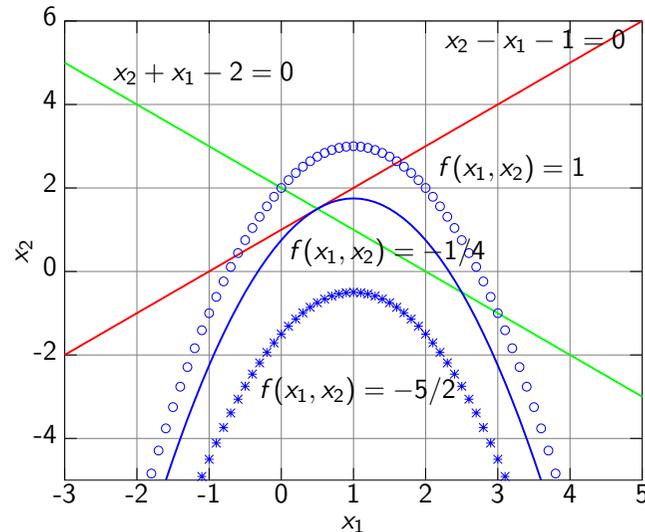
Exemple de problème non linéaire contraint (suite)

- Graphique de la fonction objectif et des lignes de niveaux



Exemple de problème non linéaire contraint (suite)

- Ce problème simple peut être appréhendé du point de vue graphique



Problèmes avec contraintes d'égalités, points réguliers

- Nous étudions dans un premier temps les problèmes non linéaires avec des contraintes d'égalités qui peuvent être formulées de la manière suivante :

$$\begin{aligned} \min f(\mathbf{x}) \\ \text{s.t. } h(\mathbf{x}) = \mathbf{0} \end{aligned}$$

où :

- $\mathbf{x} \in \mathbb{R}^n$
- $f : \mathbb{R}^n \rightarrow \mathbb{R}$
- $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $h = (h_1, \dots, h_m)$ avec $m \leq n$
- h est une fonction (vectorielle) continûment différentiable (classe \mathcal{C}^1)
- Nous avons la définition suivante :

Définition. (Point régulier)

Un point \mathbf{x} satisfaisant les contraintes $h_1(\mathbf{x}) = 0, \dots, h_m(\mathbf{x}) = 0$ est un **point régulier** des contraintes h si les vecteurs gradients $\nabla h_1(\mathbf{x}), \dots, \nabla h_m(\mathbf{x})$ sont linéairement indépendants.

Rappel du Sommaire

- Rappels de concepts mathématiques
- Bases théoriques de l'optimisation
- Optimisation sans contrainte
- Optimisation linéaire avec contraintes
- Optimisation non linéaire avec contraintes
 - Problèmes avec contraintes d'égalités
 - Problèmes avec contraintes d'inégalités
- Optimisation convexe

Points réguliers (suite)

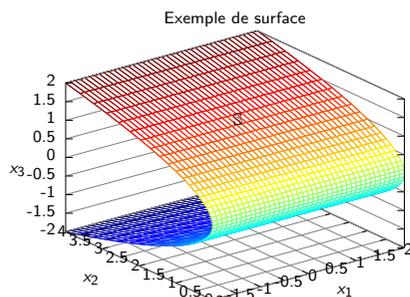
- Soit $Dh(\mathbf{x})$ la matrice Jacobienne de h en \mathbf{x} définie par :

$$\begin{aligned} Dh(\mathbf{x}) &= \begin{pmatrix} \frac{\partial h}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial h}{\partial x_n}(\mathbf{x}) \\ \frac{\partial h_1}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial h_1}{\partial x_n}(\mathbf{x}) \\ \vdots & \cdots & \vdots \\ \frac{\partial h_m}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial h_m}{\partial x_n}(\mathbf{x}) \end{pmatrix} \\ &= \begin{pmatrix} \nabla h_1(\mathbf{x})^t \\ \vdots \\ \nabla h_m(\mathbf{x})^t \end{pmatrix} \end{aligned}$$

- Alors \mathbf{x} est un point régulier si la matrice Jacobienne est de rang m
- L'ensemble des contraintes d'égalité $h_1(\mathbf{x}) = 0, \dots, h_m(\mathbf{x}) = 0$ décrit une **surface** $\mathbb{S} = \{\mathbf{x} \in \mathbb{R}^n : h(\mathbf{x}) = \mathbf{0}\}$. Si on suppose que tous les points de \mathbb{S} sont réguliers alors la dimension de \mathbb{S} est $n - m$

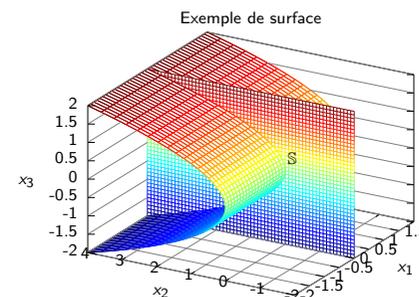
Exemple

- Soit $n = 3$ et $m = 1$ (une contrainte d'égalité)
- Si tout point de \mathbb{S} est régulier alors \mathbb{S} est de dimension $n - m = 3 - 1 = 2$
 - ▶ Par exemple, prenons : $h_1(\mathbf{x}) = x_2 - x_3^2 = 0$
 - ▶ Le vecteur gradient est $\nabla h_1(\mathbf{x}) = (0, 1, -2x_3)$
 - ▶ Donc pour tout $\mathbf{x} \in \mathbb{R}^3$, $\nabla h_1(\mathbf{x}) \neq \mathbf{0}$
 - ▶ Par ailleurs, on voit que $\mathbb{S} = \{\mathbf{x} \in \mathbb{R}^3 : \mathbf{x} = (x_1, x_3^2, x_3), x_1, x_3 \in \mathbb{R}\}$
 - ▶ On en déduit donc que $\dim(\mathbb{S}) = n - m = 3 - 1 = 2$
 - ▶ Illustration :



Autre exemple

- Soit $n = 3$ et $m = 2$ (deux contraintes d'égalité)
- Si tout point de \mathbb{S} est régulier alors \mathbb{S} est de dimension $n - m = 3 - 2 = 1$
 - ▶ Par exemple, prenons : $h_1(\mathbf{x}) = x_2 - x_3^2 = 0$ et $h_2(\mathbf{x}) = x_1 = 0$
 - ▶ Les vecteurs gradients sont $\nabla h_1(\mathbf{x}) = (0, 1, -2x_3)$ et $h_2(\mathbf{x}) = (1, 0, 0)$
 - ▶ Donc pour tout $\mathbf{x} \in \mathbb{R}^3$, $\lambda_1 \nabla h_1(\mathbf{x}) + \lambda_2 \nabla h_2(\mathbf{x}) = \mathbf{0} \Rightarrow \lambda_1 = \lambda_2 = 0$
 - ▶ Par ailleurs, on voit que $\mathbb{S} = \{\mathbf{x} \in \mathbb{R}^3 : \mathbf{x} = (0, x_3^2, x_3), x_3 \in \mathbb{R}\}$
 - ▶ On en déduit donc que $\dim(\mathbb{S}) = n - m = 3 - 2 = 1$
 - ▶ Illustration :

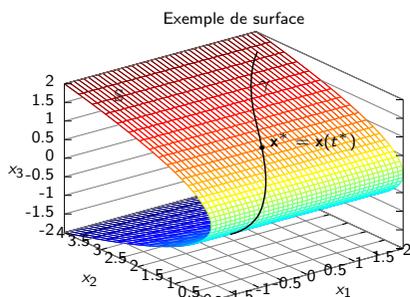


Courbe

Définition. (Courbe)

Une courbe γ sur une surface \mathbb{S} est un ensemble de points $\{\mathbf{x}(t) \in \mathbb{S} : t \in [a, b]\}$ tel que $\mathbf{x} : [a, b] \rightarrow \mathbb{S}$ est une fonction continue

- Illustration :



- Tout point de la courbe γ satisfait aux équations définissant \mathbb{S}
- On dit que γ passe par \mathbf{x}^* s'il existe $t^* \in [a, b]$ tel que $\mathbf{x}(t^*) = \mathbf{x}^*$

Courbe différentiable

- Intuitivement, on peut voir une courbe γ représentée par $\mathbf{x}(t)$ comme étant un chemin sur la surface \mathbb{S} qui est fonction du temps t et qui passe par le point \mathbf{x}^* au temps t^*

Définition. (Courbe différentiable)

- Une courbe $\gamma = \{\mathbf{x}(t) \in \mathbb{S} : t \in [a, b]\}$ est **différentiable** si :

$$\frac{d\mathbf{x}}{dt}(t) = \left(\frac{dx_1}{dt}(t), \dots, \frac{dx_n}{dt}(t) \right)$$

existe pour tout $t \in [a, b]$

- Une courbe γ est **deux fois différentiable** si :

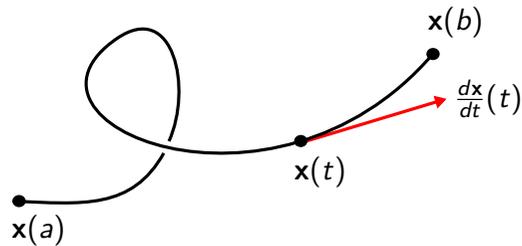
$$\frac{d^2\mathbf{x}}{dt^2}(t) = \left(\frac{d^2x_1}{dt^2}(t), \dots, \frac{d^2x_n}{dt^2}(t) \right)$$

existe pour tout $t \in [a, b]$

- $\frac{d\mathbf{x}}{dt}(t)$ et $\frac{d^2\mathbf{x}}{dt^2}(t)$ sont des vecteurs de \mathbb{R}^n . Le premier peut être vu comme le vecteur vitesse (instantanée) de $\mathbf{x}(t)$ et le second comme le vecteur accélération (taux de changement de la vitesse)

Courbe différentiable (suite)

- Illustration :

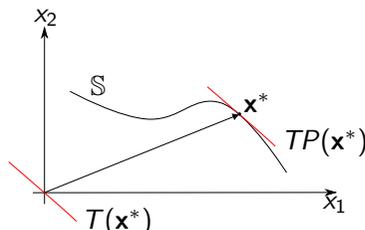


Espace tangent (suite)

- On suppose que \mathbf{x}^* est régulier (ie $Dh(\mathbf{x}^*)$ est de rang m). Dans ce cas $T(\mathbf{x}^*)$ est de dimension $n - m$ où m étant le nombre de contraintes d'égalités.
- $T(\mathbf{x}^*)$ est un sous-espace vectoriel de \mathbb{R}^n donc il passe par l'origine et non nécessairement par \mathbf{x}^* . Il est plus intéressant de visualiser l'espace tangent comme étant un espace passant par \mathbf{x}^* , on introduit pour cela les **espaces affines**. Toutefois, par abus de langage, nous supposons par la suite que l'**espace tangent en \mathbf{x}^*** est l'espace affine défini par :

$$TP(\mathbf{x}^*) = T(\mathbf{x}^*) + \mathbf{x}^* = \{\mathbf{x} + \mathbf{x}^* : \mathbf{x} \in T(\mathbf{x}^*)\}$$

- Illustration :



Espace tangent

- Rappelons que $\mathbb{S} = \{\mathbf{x} \in \mathbb{R}^n : h(\mathbf{x}) = \mathbf{0}\}$ où $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ est de classe \mathcal{C}^1 . \mathbb{S} est une surface de \mathbb{R}^n

Définition. (Espace tangent)

L'**espace tangent** en un point \mathbf{x}^* de la surface $\mathbb{S} = \{\mathbf{x} \in \mathbb{R}^n : h(\mathbf{x}) = \mathbf{0}\}$ est l'ensemble défini par :

$$T(\mathbf{x}^*) = \{\mathbf{y} : Dh(\mathbf{x}^*)\mathbf{y} = \mathbf{0}\}$$

- $T(\mathbf{x}^*)$ est $\text{Ker}(Dh(\mathbf{x}^*))$ ie le noyau de l'application linéaire représentée par la matrice Jacobienne $Dh(\mathbf{x}^*)$. Donc $T(\mathbf{x}^*)$ est un sous-espace de \mathbb{R}^n
- Rappelons la définition de la matrice Jacobienne de h :

$$Dh(\mathbf{x}^*) = \begin{pmatrix} \frac{\partial h_1}{\partial x_1}(\mathbf{x}^*) & \dots & \frac{\partial h_1}{\partial x_n}(\mathbf{x}^*) \\ \vdots & \dots & \vdots \\ \frac{\partial h_m}{\partial x_1}(\mathbf{x}^*) & \dots & \frac{\partial h_m}{\partial x_n}(\mathbf{x}^*) \end{pmatrix}$$

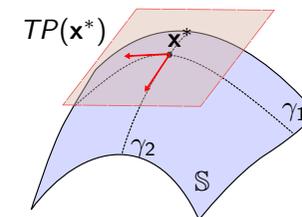
Espace tangent (suite)

- Intuitivement, on peut voir l'espace tangent en un point \mathbf{x}^* d'une surface comme étant l'ensemble des vecteurs tangents à la surface en ce point. Or tout vecteur gradient d'une courbe γ appartenant à la surface et passant par \mathbf{x}^* est tangent à la courbe et donc à la surface

Théorème.

Supposons que $\mathbf{x}^* \in \mathbb{S}$ est un point régulier et que $T(\mathbf{x}^*)$ est l'espace tangent en \mathbf{x}^* . Alors, $\mathbf{y} \in T(\mathbf{x}^*)$ ssi il existe une courbe différentiable de \mathbb{S} passant par \mathbf{x}^* tel que \mathbf{y} est la dérivée de cette courbe en \mathbf{x}^*

- Illustration :



Espace normal

Définition. (Espace normal)

L'espace normal en un point \mathbf{x}^* de la surface $\mathbb{S} = \{\mathbf{x} \in \mathbb{R}^n : h(\mathbf{x}) = 0\}$ est l'ensemble défini par :

$$N(\mathbf{x}^*) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} = Dh(\mathbf{x}^*)^t \mathbf{y}, \mathbf{y} \in \mathbb{R}^m\}$$

- $N(\mathbf{x}^*)$ est $Im(Dh(\mathbf{x}^*)^t)$ ie l'image de l'application linéaire représentée par la matrice $Dh(\mathbf{x}^*)^t$ ou encore l'espace engendré par les vecteurs gradients de h : $N(\mathbf{x}^*) = Vec\{\nabla h_1(\mathbf{x}^*), \dots, \nabla h_m(\mathbf{x}^*)\} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} = \sum_{j=1}^m y_j \nabla h_j(\mathbf{x}^*), y_1, \dots, y_m \in \mathbb{R}\}$
- $N(\mathbf{x}^*)$ étant un sous-espace de \mathbb{R}^n il contient $\mathbf{0}$. Comme pour précédemment, il est plus intéressant de considérer l'espace normal en \mathbf{x}^* comme l'espace affine associé passant par \mathbf{x}^* . On nommera également par espace normal l'espace affine défini par :

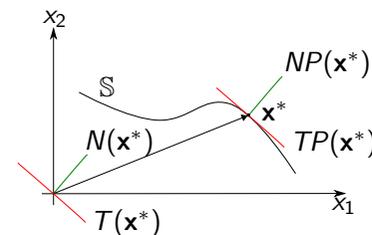
$$NP(\mathbf{x}^*) = N(\mathbf{x}^*) + \mathbf{x}^* = \{\mathbf{x} + \mathbf{x}^* \in \mathbb{R}^n : \mathbf{x} \in N(\mathbf{x}^*)\}$$

Conditions de Lagrange dans le cas \mathbb{R}^2

- Le fameux théorème de Lagrange donne des conditions nécessaires caractérisant les points optimiseurs de problèmes contraints
- Pour mieux comprendre ce théorème on étudie d'abord les problèmes à 2 dimensions avec 1 contrainte d'égalité
- $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ est la seule contrainte
- Rappelons qu'à chaque point \mathbf{x}^* du domaine de h (ici \mathbb{R}^2), le vecteur gradient $\nabla f(\mathbf{x}^*)$ est orthogonal à la ligne de niveau $\mathbb{L} = \{\mathbf{x} \in \mathbb{R}^2 : h(\mathbf{x}) = c\}$ passant par \mathbf{x}^*
- En effet, prenons $\mathbf{x}^* = (x_1^*, x_2^*)$ tel que $h(\mathbf{x}^*) = 0$ et supposons que $\nabla h(\mathbf{x}^*) \neq \mathbf{0}$
- Dans ce cas la ligne de niveau est $\mathbb{L} = \{\mathbf{x} \in \mathbb{R}^2 : h(\mathbf{x}) = 0\}$ et supposons qu'au voisinage de \mathbf{x}^* on a une courbe $\gamma = \{\mathbf{x}(t)\}$ parcourant la ligne de niveau avec $\mathbf{x} : [a, b] \rightarrow \mathbb{R}^2$

Espace normal (suite)

- Illustration :



Propriété. (Relation entre espace tangent et espace normal)

$$T(\mathbf{x}^*) = N(\mathbf{x}^*)^\perp \text{ et } N(\mathbf{x}^*) = T(\mathbf{x}^*)^\perp$$

où $N(\mathbf{x}^*)^\perp$ est le sous-espace orthogonal à $N(\mathbf{x}^*)$

Propriété. (Somme directe de l'espace tangent et de l'espace normal)

$$\mathbb{R}^n = T(\mathbf{x}^*) \oplus N(\mathbf{x}^*)$$

Conditions de Lagrange dans le cas \mathbb{R}^2 (suite)

- On a alors :

$$\begin{aligned} h(\mathbf{x}(t)) = 0 &\Rightarrow \frac{d}{dt} h(\mathbf{x}(t)) = 0 \\ &\Rightarrow Dh(\mathbf{x}(t)) \frac{d\mathbf{x}}{dt}(t) = 0 \\ &\Rightarrow \nabla h(\mathbf{x}(t))^t \frac{d\mathbf{x}}{dt}(t) = 0 \\ &\Rightarrow \nabla h(\mathbf{x}(t)) \text{ est orthogonal à } \frac{d\mathbf{x}}{dt}(t) \end{aligned}$$

\Rightarrow En d'autres termes, pour $\mathbf{x} \in \mathbf{x}(t)$, le gradient de h est orthogonal au vecteur tangent à la ligne de niveau représentée par $\mathbf{x}(t)$

- Supposons à présent que nous cherchions un minimiseur de $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ sur $\mathbb{S} = \{\mathbf{x} : h(\mathbf{x}) = 0\}$

- Si $\mathbf{x}^* = \mathbf{x}(t^*)$ est un minimiseur de f alors on a :

$$\begin{aligned} \frac{d}{dt} f(\mathbf{x}(t^*)) = 0 &\Rightarrow Df(\mathbf{x}(t^*)) \frac{d\mathbf{x}}{dt}(t^*) = 0 \\ &\Rightarrow \nabla f(\mathbf{x}(t^*))^t \frac{d\mathbf{x}}{dt}(t^*) = 0 \\ &\Rightarrow \nabla f(\mathbf{x}(t^*)) \text{ est orthogonal à } \frac{d\mathbf{x}}{dt}(t^*) \end{aligned}$$

\Rightarrow Ainsi, en \mathbf{x}^* , le gradient de f est également orthogonal au vecteur tangent à la ligne de niveau représentée par $\mathbf{x}(t)$

Conditions de Lagrange dans le cas \mathbb{R}^2 (suite)

⇒ Des deux observations précédentes nous concluons que si \mathbf{x}^* est un minimiseur de f sur \mathbb{S} alors $\nabla h(\mathbf{x}^*)$ et $\nabla f(\mathbf{x}^*)$ sont tous deux orthogonaux à $\frac{d\mathbf{x}}{dt}(t)$ ce qui veut dire qu'ils sont "parallèles"

Théorème. (Théorème de Lagrange pour $n = 2$ et $m = 1$)

Soit \mathbf{x}^* un minimiseur de $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ satisfaisant à la contrainte $h(\mathbf{x}) = 0, h : \mathbb{R}^2 \rightarrow \mathbb{R}$. Alors, si $\nabla h(\mathbf{x}^*) \neq \mathbf{0}$, il existe un scalaire λ^* tel que :

$$\nabla f(\mathbf{x}^*) + \lambda^* \nabla h(\mathbf{x}^*) = \mathbf{0}$$

λ^* est appelé multiplicateur de Lagrange

- Autrement dit ce théorème donne des conditions nécessaires pour qu'un point \mathbf{x}^* soit un minimiseur local de f :
 - ▶ $\nabla f(\mathbf{x}^*) + \lambda^* \nabla h(\mathbf{x}^*) = \mathbf{0}$
 - ▶ $h(\mathbf{x}) = 0$

Illustration

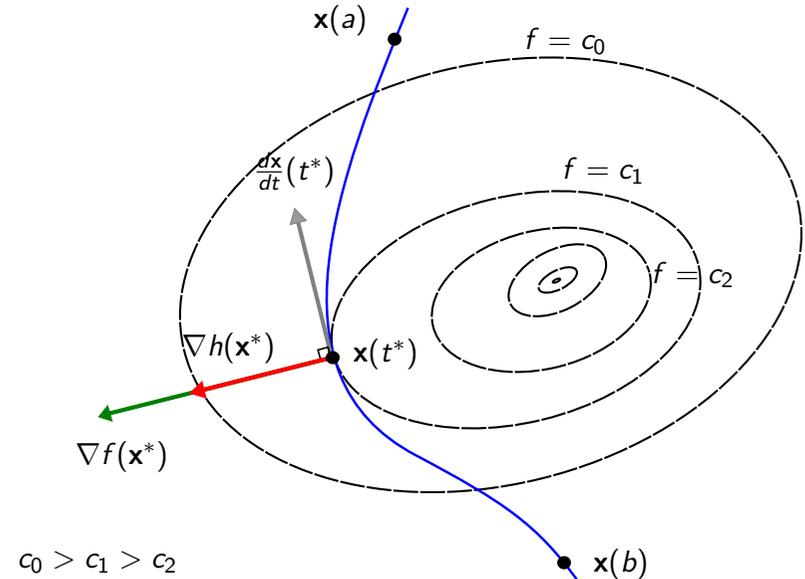


Illustration (suite)

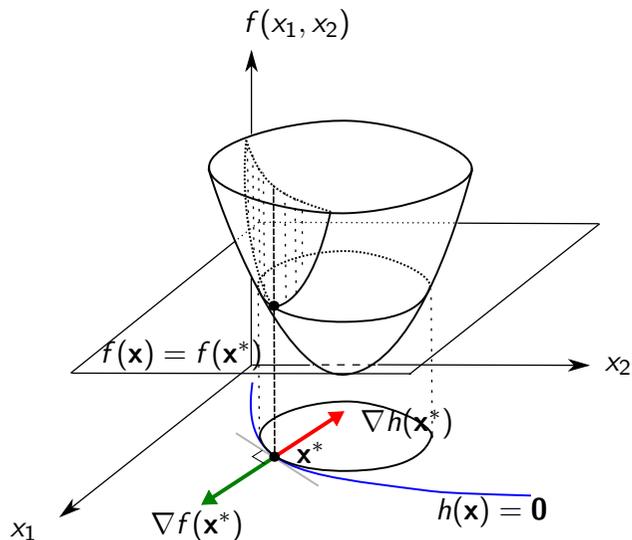
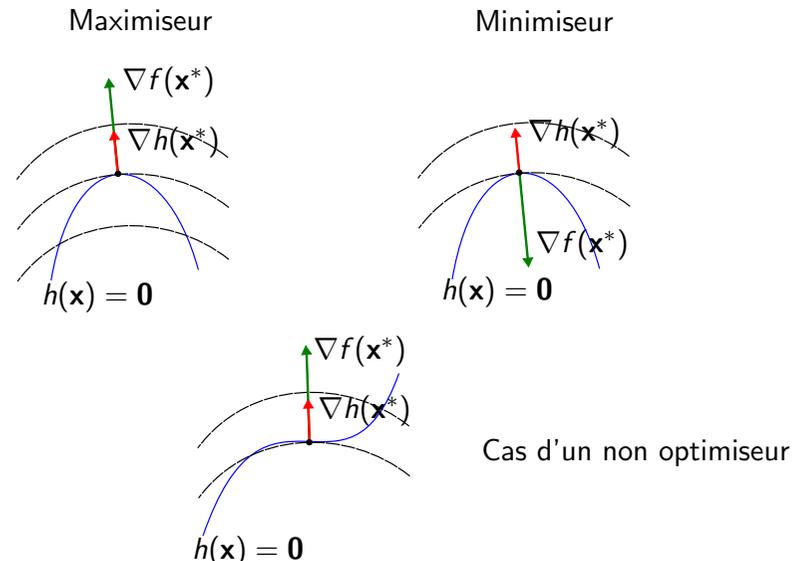


Illustration (suite)

- Attention il s'agit de conditions nécessaires mais non suffisantes



Conditions de Lagrange cas général

Théorème. (Théorème de Lagrange)

Soit \mathbf{x}^* un minimiseur (ou maximiseur) local de $f : \mathbb{R}^n \rightarrow \mathbb{R}$ satisfaisant aux contraintes $h(\mathbf{x}) = 0, h : \mathbb{R}^n \rightarrow \mathbb{R}^m, m \leq n$. Alors, si \mathbf{x}^* est un point régulier, il existe un $\lambda^* \in \mathbb{R}^m$ tel que :

$$Df(\mathbf{x}^*) + \lambda^{*t} Dh(\mathbf{x}^*) = \mathbf{0}^t$$

λ^{*t} est appelé (vecteur de) multiplicateurs de Lagrange

- Le théorème de Lagrange stipule que si \mathbf{x}^* est un optimiseur alors le vecteur gradient de f en \mathbf{x}^* , $\nabla f(\mathbf{x}^*)$, peut s'exprimer comme une combinaison linéaire des m vecteurs gradients de h
- Autrement dit, \mathbf{x}^* ne peut pas être un optimiseur si $\mathbf{x}^* \notin N(\mathbf{x}^*)$

Fonction de Lagrange

- Il est utile d'introduire la **fonction de Lagrange** (ou Lagrangien) notée l et définie par :

$$l(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda^t h(\mathbf{x})$$

- On peut alors exprimer les conditions de Lagrange de la manière suivante : si \mathbf{x}^* est un minimiseur local de f alors

$$Dl(\mathbf{x}^*, \lambda^*) = \mathbf{0}^t$$

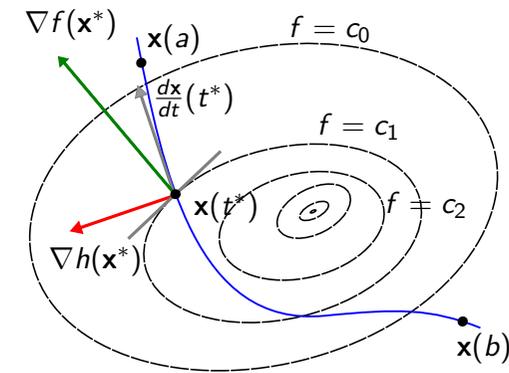
où la dérivée D est par rapport aux deux arguments \mathbf{x} et λ de l

- En effet si on note $D_{\mathbf{x}}$ la dérivée par rapport à \mathbf{x} et D_{λ} la dérivée par rapport à λ alors :

$$\begin{aligned} Dl(\mathbf{x}^*, \lambda^*) = \mathbf{0}^t &\Leftrightarrow (D_{\mathbf{x}}l(\mathbf{x}^*, \lambda^*) \quad D_{\lambda}l(\mathbf{x}^*, \lambda^*)) = \mathbf{0}^t \\ &\Leftrightarrow (Df(\mathbf{x}^*) + \lambda^{*t} Dh(\mathbf{x}^*) \quad h(\mathbf{x}^*)^t) = \mathbf{0}^t \end{aligned}$$

- En d'autres termes, les conditions nécessaires de Lagrange sont équivalentes aux CNPO des problèmes non contraints appliquées à la fonction de Lagrange
- L'équation précédente représente $n + m$ équations à $n + m$ inconnues x_1^*, \dots, x_n^* et $\lambda_1^*, \dots, \lambda_m^*$

Illustration



$$c_0 > c_1 > c_2$$

Cas où les conditions de Lagrange ne sont pas satisfaites

Exemple

- Déterminons les optimiseurs de la fonction $f(\mathbf{x}) = x_1^2 + x_2^2$ sur l'ellipse $\mathbb{S} = \{(x_1, x_2) : h(x_1, x_2) = x_1^2 + 2x_2^2 - 1 = 0\}$
- Nous avons :

$$\nabla f(\mathbf{x}) =$$

Exemple

- Déterminons les optimiseurs de la fonction $f(\mathbf{x}) = x_1^2 + x_2^2$ sur l'ellipse $\mathbb{S} = \{(x_1, x_2) : h(x_1, x_2) = x_1^2 + 2x_2^2 - 1 = 0\}$

- Nous avons :

$$\begin{aligned}\nabla f(\mathbf{x}) &= (2x_1, 2x_2) \\ \nabla h(\mathbf{x}) &= \end{aligned}$$

Exemple

- Déterminons les optimiseurs de la fonction $f(\mathbf{x}) = x_1^2 + x_2^2$ sur l'ellipse $\mathbb{S} = \{(x_1, x_2) : h(x_1, x_2) = x_1^2 + 2x_2^2 - 1 = 0\}$

- Nous avons :

$$\begin{aligned}\nabla f(\mathbf{x}) &= (2x_1, 2x_2) \\ \nabla h(\mathbf{x}) &= (2x_1, 4x_2)\end{aligned}$$

- le Lagrangien vaut :

$$l(\mathbf{x}, \lambda) = x_1^2 + x_2^2 + \lambda(x_1^2 + 2x_2^2 - 1)$$

- On a donc :

$$\begin{aligned}Dl(\mathbf{x}, \lambda) &= (D_{\mathbf{x}}l(\mathbf{x}, \lambda) \quad D_{\lambda}l(\mathbf{x}, \lambda)) \\ &= (\nabla f(\mathbf{x})^t + \lambda \nabla h(\mathbf{x})^t \quad h(\mathbf{x})) \\ &= \end{aligned}$$

Exemple

- Déterminons les optimiseurs de la fonction $f(\mathbf{x}) = x_1^2 + x_2^2$ sur l'ellipse $\mathbb{S} = \{(x_1, x_2) : h(x_1, x_2) = x_1^2 + 2x_2^2 - 1 = 0\}$

- Nous avons :

$$\begin{aligned}\nabla f(\mathbf{x}) &= (2x_1, 2x_2) \\ \nabla h(\mathbf{x}) &= (2x_1, 4x_2)\end{aligned}$$

- le Lagrangien vaut :

$$l(\mathbf{x}, \lambda) =$$

Exemple

- Déterminons les optimiseurs de la fonction $f(\mathbf{x}) = x_1^2 + x_2^2$ sur l'ellipse $\mathbb{S} = \{(x_1, x_2) : h(x_1, x_2) = x_1^2 + 2x_2^2 - 1 = 0\}$

- Nous avons :

$$\begin{aligned}\nabla f(\mathbf{x}) &= (2x_1, 2x_2) \\ \nabla h(\mathbf{x}) &= (2x_1, 4x_2)\end{aligned}$$

- le Lagrangien vaut :

$$l(\mathbf{x}, \lambda) = x_1^2 + x_2^2 + \lambda(x_1^2 + 2x_2^2 - 1)$$

- On a donc :

$$\begin{aligned}Dl(\mathbf{x}, \lambda) &= (D_{\mathbf{x}}l(\mathbf{x}, \lambda) \quad D_{\lambda}l(\mathbf{x}, \lambda)) \\ &= (\nabla f(\mathbf{x})^t + \lambda \nabla h(\mathbf{x})^t \quad h(\mathbf{x})) \\ &= (2(1 + \lambda)x_1 \quad 2(1 + 2\lambda)x_2 \quad x_1^2 + 2x_2^2 - 1)\end{aligned}$$

- Ensuite :

$$Dl(\mathbf{x}, \lambda) = \mathbf{0}^t \Leftrightarrow$$

Exemple

- Déterminons les optimiseurs de la fonction $f(\mathbf{x}) = x_1^2 + x_2^2$ sur l'ellipse $\mathbb{S} = \{(x_1, x_2) : h(x_1, x_2) = x_1^2 + 2x_2^2 - 1 = 0\}$

- Nous avons :

$$\begin{aligned}\nabla f(\mathbf{x}) &= (2x_1, 2x_2) \\ \nabla h(\mathbf{x}) &= (2x_1, 4x_2)\end{aligned}$$

- le Lagrangien vaut :

$$l(\mathbf{x}, \lambda) = x_1^2 + x_2^2 + \lambda(x_1^2 + 2x_2^2 - 1)$$

- On a donc :

$$\begin{aligned}DI(\mathbf{x}, \lambda) &= (D_{\mathbf{x}}l(\mathbf{x}, \lambda) \quad D_{\lambda}l(\mathbf{x}, \lambda)) \\ &= (\nabla f(\mathbf{x})^t + \lambda \nabla h(\mathbf{x})^t \quad h(\mathbf{x})) \\ &= (2(1 + \lambda)x_1 \quad 2(1 + 2\lambda)x_2 \quad x_1^2 + 2x_2^2 - 1)\end{aligned}$$

- Ensuite :

$$DI(\mathbf{x}, \lambda) = \mathbf{0}^t \Leftrightarrow \begin{cases} 2(1 + \lambda)x_1 = 0 \\ 2(1 + 2\lambda)x_2 = 0 \\ x_1^2 + 2x_2^2 = 1 \end{cases}$$

Exemple (suite)

- On a donc :

$$DI(\mathbf{x}, \lambda) = \mathbf{0}^t \Leftrightarrow \begin{cases} 2(1 + \lambda)x_1 = 0 \\ 2(1 + 2\lambda)x_2 = 0 \\ x_1^2 + 2x_2^2 = 1 \end{cases} \Leftrightarrow$$

Exemple (suite)

- On a donc :

$$DI(\mathbf{x}, \lambda) = \mathbf{0}^t \Leftrightarrow \begin{cases} 2(1 + \lambda)x_1 = 0 \\ 2(1 + 2\lambda)x_2 = 0 \\ x_1^2 + 2x_2^2 = 1 \end{cases} \Leftrightarrow \begin{cases} x_1 = 0 \text{ ou } \lambda = -1 \\ x_2 = 0 \text{ ou } \lambda = -\frac{1}{2} \\ x_1^2 + 2x_2^2 = 1 \end{cases}$$

- On en déduit les solutions possibles suivantes :

$$\mathbf{x}^{(1)} = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \mathbf{x}^{(2)} = \begin{pmatrix} 0 \\ -\frac{1}{\sqrt{2}} \end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \mathbf{x}^{(4)} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$$

- On obtient $f(\mathbf{x}^{(1)}) = f(\mathbf{x}^{(2)}) = \frac{1}{2}$ et $f(\mathbf{x}^{(3)}) = f(\mathbf{x}^{(4)}) = 1$

Exemple (suite)

- On a donc :

$$DI(\mathbf{x}, \lambda) = \mathbf{0}^t \Leftrightarrow \begin{cases} 2(1 + \lambda)x_1 = 0 \\ 2(1 + 2\lambda)x_2 = 0 \\ x_1^2 + 2x_2^2 = 1 \end{cases} \Leftrightarrow \begin{cases} x_1 = 0 \text{ ou } \lambda = -1 \\ x_2 = 0 \text{ ou } \lambda = -\frac{1}{2} \\ x_1^2 + 2x_2^2 = 1 \end{cases}$$

- On en déduit les solutions possibles suivantes :

$$\mathbf{x}^{(1)} = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \mathbf{x}^{(2)} = \begin{pmatrix} 0 \\ -\frac{1}{\sqrt{2}} \end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \mathbf{x}^{(4)} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$$

- On obtient $f(\mathbf{x}^{(1)}) = f(\mathbf{x}^{(2)}) = \frac{1}{2}$ et $f(\mathbf{x}^{(3)}) = f(\mathbf{x}^{(4)}) = 1$

Exemple (suite)

- On a donc :

$$DI(\mathbf{x}, \lambda) = \mathbf{0}^t \Leftrightarrow \begin{cases} 2(1 + \lambda)x_1 = 0 \\ 2(1 + 2\lambda)x_2 = 0 \\ x_1^2 + 2x_2^2 = 1 \end{cases} \Leftrightarrow \begin{cases} x_1 = 0 \text{ ou } \lambda = -1 \\ x_2 = 0 \text{ ou } \lambda = -\frac{1}{2} \\ x_1^2 + 2x_2^2 = 1 \end{cases}$$

- On en déduit les solutions possibles suivantes :

$$\mathbf{x}^{(1)} = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \mathbf{x}^{(2)} = \begin{pmatrix} 0 \\ -\frac{1}{\sqrt{2}} \end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \mathbf{x}^{(4)} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$$

- On obtient $f(\mathbf{x}^{(1)}) = f(\mathbf{x}^{(2)}) = \frac{1}{2}$ et $f(\mathbf{x}^{(3)}) = f(\mathbf{x}^{(4)}) = 1$
- On conclut que s'il existe des minimiseurs de f alors ils sont en $\mathbf{x}^{(1)}$ et $\mathbf{x}^{(2)}$ et s'il existe des maximiseurs ils sont en $\mathbf{x}^{(3)}$ et $\mathbf{x}^{(4)}$
- Il s'avère en fait que $\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)}$ et $\mathbf{x}^{(3)}$, $\mathbf{x}^{(4)}$ sont respectivement les minimiseurs et maximiseurs de f sur \mathbb{S}

Conditions nécessaires du second ordre (CNSO)

- Les CNPO de Lagrange permettent de trouver les points critiques parmi lesquels pourraient se trouver les optimiseurs
- Nous avons maintenant besoin de plus d'outils afin de caractériser ces points critiques, nous introduisons donc de nouvelles conditions basées sur la différentielle d'ordre 2 ie la matrice hessienne
- On suppose désormais que $f : \mathbb{R}^n \rightarrow \mathbb{R}$ et $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ sont de classe \mathcal{C}^2 et soit $l(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda^t h(\mathbf{x})$ le Lagrangien
- Soit $D^2 l(\mathbf{x}, \lambda)$ la matrice hessienne ou hessien de l . On note respectivement par D_x^2 et D_λ^2 , la dérivée seconde par rapport à \mathbf{x} et λ
- On a $D_x^2 l(\mathbf{x}, \lambda) = D_x^2 f(\mathbf{x}) + \lambda_1 D_x^2 h_1(\mathbf{x}) + \dots + \lambda_m D_x^2 h_m(\mathbf{x})$ avec :

$$D^2 h_j = \begin{pmatrix} \frac{\partial^2 h_j}{\partial x_1^2} & \frac{\partial^2 h_j}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 h_j}{\partial x_1 \partial x_n} \\ \frac{\partial^2 h_j}{\partial x_2 \partial x_1} & \frac{\partial^2 h_j}{\partial x_2^2} & \dots & \frac{\partial^2 h_j}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \dots & \vdots \\ \frac{\partial^2 h_j}{\partial x_n \partial x_1} & \frac{\partial^2 h_j}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 h_j}{\partial x_n^2} \end{pmatrix}$$

Conditions nécessaires du second ordre (CNSO) (suite)

Théorème. (CNSO)

Soit \mathbf{x}^* un minimiseur local de $f : \mathbb{R}^n \rightarrow \mathbb{R}$ satisfaisant aux contraintes $h(\mathbf{x}) = 0$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $m \leq n$ et $f, h \in \mathcal{C}^2$. Alors, si \mathbf{x}^* est un point régulier, il existe un $\lambda^* \in \mathbb{R}^m$ tel que :

- $Df(\mathbf{x}^*) + \lambda^{*t} Dh(\mathbf{x}^*) = \mathbf{0}^t$
- Pour tout $\mathbf{y} \in T(\mathbf{x}^*)$, nous avons : $\mathbf{y}^t D_x^2 l(\mathbf{x}^*, \lambda^*) \mathbf{y} \geq 0$

- $D_x^2 l$ joue un rôle similaire que $D^2 f$ dans le cas non contraint
- Toutefois, dans le cas contraint, il est seulement requis aux vecteurs appartenant à l'espace tangent en \mathbf{x}^* de satisfaire à la propriété de semi-définie positivité de $D_x^2 l$
- Ces conditions sont uniquement nécessaires, nous présentons des conditions suffisantes dans ce qui suit

Conditions suffisantes du second ordre (CSSO)

Théorème. (CSSO)

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ et $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$, deux fonctions de classe \mathcal{C}^2 . S'il existe un point réalisable \mathbf{x}^* et $\lambda^* \in \mathbb{R}^m$ tels que :

- $Df(\mathbf{x}^*) + \lambda^{*t} Dh(\mathbf{x}^*) = \mathbf{0}^t$
- Pour tout $\mathbf{y} \in T(\mathbf{x}^*)$, $\mathbf{y} \neq \mathbf{0}$, nous avons : $\mathbf{y}^t D_x^2 l(\mathbf{x}^*, \lambda^*) \mathbf{y} > 0$

alors \mathbf{x}^* est un minimiseur local strict de f satisfaisant aux contraintes $h(\mathbf{x}) = \mathbf{0}$

- Autrement dit, si \mathbf{x}^* vérifie les conditions de Lagrange et si $D_x^2 l(\mathbf{x}^*, \lambda^*)$ est définie positive pour tout vecteur de $T(\mathbf{x}^*)$ alors c'est un minimiseur local strict
- Dans le cas d'un maximiseur, la condition 2 est remplacée par "définie négative" (toutes choses étant égales par ailleurs)

Rappel du Sommaire

- 1 Rappels de concepts mathématiques
- 2 Bases théoriques de l'optimisation
- 3 Optimisation sans contrainte
- 4 Optimisation linéaire avec contraintes
- 5 Optimisation non linéaire avec contraintes
 - Problèmes avec contraintes d'égalités
 - Problèmes avec contraintes d'inégalités
- 6 Optimisation convexe

Contraintes actives et points réguliers

Définition. (Contrainte active)

Une contrainte d'inégalité $g_j \leq 0$ est dite **active** en \mathbf{x}^* si $g_j(\mathbf{x}^*) = 0$. Elle est dite **inactive** en \mathbf{x}^* si $g_j(\mathbf{x}^*) < 0$

Définition. (Ensemble d'indices d'inégalités actives)

Soit \mathbf{x}^* un point satisfaisant les contraintes $h(\mathbf{x}^*) = 0$ et $g(\mathbf{x}^*) \leq 0$. On note $\mathbb{J}(\mathbf{x}^*)$ l'ensemble des indices de contraintes d'inégalités qui sont actives : $\mathbb{J}(\mathbf{x}^*) = \{j : g_j(\mathbf{x}^*) = 0\}$

Définition. (Point régulier)

Un point \mathbf{x}^* est dit **point régulier** si les vecteurs :

$$\nabla h_i(\mathbf{x}^*), i = 1, \dots, n \text{ et } \nabla g_j(\mathbf{x}^*), j \in \mathbb{J}(\mathbf{x}^*)$$

sont linéairement indépendants

Problèmes avec contraintes d'égalités et d'inégalités

- Nous étudions maintenant le cas général des problèmes non linéaires avec des contraintes d'égalités et d'inégalités qui peuvent être formulés de la manière suivante :

$$\begin{aligned} \min & f(\mathbf{x}) \\ \text{s/c} & h(\mathbf{x}) = \mathbf{0} \\ & g(\mathbf{x}) \leq \mathbf{0} \end{aligned}$$

où :

- ▶ $\mathbf{x} \in \mathbb{R}^n$
- ▶ $f : \mathbb{R}^n \rightarrow \mathbb{R}$
- ▶ $h : \mathbb{R}^n \rightarrow \mathbb{R}^m, h = (h_1, \dots, h_m)$ avec $m \leq n$
- ▶ $g : \mathbb{R}^n \rightarrow \mathbb{R}^p, g = (g_1, \dots, g_p)$ avec $m \leq n$

Conditions de Karush-Kuhn-Tucker (KKT)

Théorème. (Théorème de KKT)

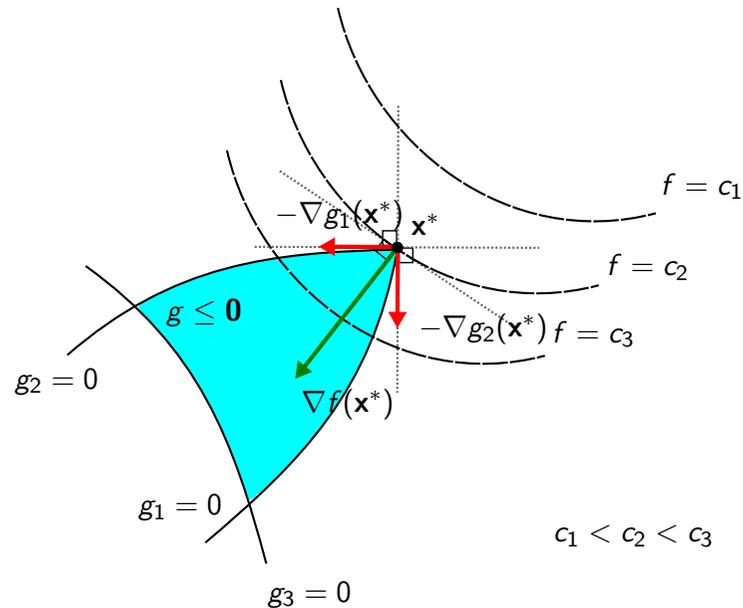
Soit f, h, g des fonctions de classe \mathcal{C}^1 . Soit \mathbf{x}^* un point régulier qui est un minimiseur local de f satisfaisant aux contraintes $h(\mathbf{x}) = \mathbf{0}$ et $g(\mathbf{x}) \leq \mathbf{0}$.

Alors il existe $\lambda^* \in \mathbb{R}^m$ et $\mu^* \in \mathbb{R}^p$ tel que :

- 1 $\mu^* \geq \mathbf{0}$
- 2 $Df(\mathbf{x}^*) + \lambda^{*t} Dh(\mathbf{x}^*) + \mu^{*t} Dg(\mathbf{x}^*) = \mathbf{0}^t$
- 3 $\mu^{*t} g(\mathbf{x}^*) = 0$

où λ^* sont appelés (vecteur des) multiplicateurs de Lagrange et μ^* (vecteur des) multiplicateurs de Karush-Kuhn-Tucker

- Comme $\mu_j^* \geq 0$ et $g_j(\mathbf{x}^*) \leq 0$, la condition 3 implique que si $g_j(\mathbf{x}^*) < 0$ alors $\mu_j^* = 0$. Dit autrement, $\forall j \notin \mathbb{J}(\mathbf{x}^*), \mu_j^* = 0$

Illustration dans \mathbb{R}^2 

Cas où on vérifie les conditions de KKT

Interprétation du théorème de KKT (suite)

- Les conditions de KKT comme les conditions de Lagrange sont nécessaires et on les applique afin de déterminer des points critiques candidats à être des points minimiseurs
- Les conditions de KKT pour un problème de minimisation consistent en 3 équations et 2 inégalités :
 - 1 $\mu^* \geq \mathbf{0}$
 - 2 $Df(\mathbf{x}^*) + \lambda^{*t} Dh(\mathbf{x}^*) + \mu^{*t} Dg(\mathbf{x}^*) = \mathbf{0}^t$
 - 3 $\mu^{*t} \mathbf{g}(\mathbf{x}^*) = 0$
 - 4 $h(\mathbf{x}^*) = \mathbf{0}$
 - 5 $\mathbf{g}(\mathbf{x}^*) \leq \mathbf{0}$

Interprétation du théorème de KKT

- Nous interprétons les conditions de KKT sur l'exemple graphique précédent
- Nous n'avons que des contraintes inégalités $g_j(\mathbf{x}) \leq 0, j = 1, 2, 3$
- \mathbf{x}^* est un minimiseur et satisfait donc les conditions de KKT
- $\mathbb{J}(\mathbf{x}^*) = \{1, 2\}$ donc $g_3(\mathbf{x}^*) < 0$ est inactive par conséquent $\mu_3^* = 0$ (condition 3)
- Selon la condition 2 de KKT nous avons :

$$\nabla f(\mathbf{x}^*) + \mu_1^* \nabla g_1(\mathbf{x}^*) + \mu_2^* \nabla g_2(\mathbf{x}^*) = \mathbf{0}^t$$

ce qui est équivalent à :

$$\nabla f(\mathbf{x}^*) = -\mu_1^* \nabla g_1(\mathbf{x}^*) - \mu_2^* \nabla g_2(\mathbf{x}^*)$$

où $\mu_1^*, \mu_2^* > 0$ (condition 1)

- Ainsi, $\nabla f(\mathbf{x}^*)$ doit être une combinaison linéaire des vecteurs $-\nabla g_1(\mathbf{x}^*)$ et $-\nabla g_2(\mathbf{x}^*)$ avec des coefficients positifs

Conditions de KKT pour un problème de maximisation

- On considère le problème de maximisation suivant :

$$\begin{aligned} \max f(\mathbf{x}) \\ \text{s.t. } h(\mathbf{x}) = \mathbf{0} \\ g(\mathbf{x}) \leq \mathbf{0} \end{aligned}$$

- Dans ce cas, les conditions de KKT sont :

- 1 $\mu^* \geq \mathbf{0}$
- 2 $-Df(\mathbf{x}^*) + \lambda^{*t} Dh(\mathbf{x}^*) + \mu^{*t} Dg(\mathbf{x}^*) = \mathbf{0}^t$
- 3 $\mu^{*t} \mathbf{g}(\mathbf{x}^*) = 0$
- 4 $h(\mathbf{x}^*) = \mathbf{0}$
- 5 $\mathbf{g}(\mathbf{x}^*) \leq \mathbf{0}$

- En multipliant la condition 2 par -1 et en changeant les signes des multiplicateurs, on a les conditions équivalentes suivantes :

- 1 $\mu^* \leq \mathbf{0}$
- 2 $Df(\mathbf{x}^*) + \lambda^{*t} Dh(\mathbf{x}^*) + \mu^{*t} Dg(\mathbf{x}^*) = \mathbf{0}^t$
- 3 $\mu^{*t} \mathbf{g}(\mathbf{x}^*) = 0$
- 4 $h(\mathbf{x}^*) = \mathbf{0}$
- 5 $\mathbf{g}(\mathbf{x}^*) \leq \mathbf{0}$

Conditions de KKT avec contraintes \geq

- On considère le problème de minimisation suivant :

$$\begin{aligned} \min f(\mathbf{x}) \\ \text{s.l.c. } h(\mathbf{x}) = \mathbf{0} \\ g(\mathbf{x}) \geq \mathbf{0} \end{aligned}$$

- Dans ce cas, les conditions de KKT sont :

- $\mu^* \geq \mathbf{0}$
- $Df(\mathbf{x}^*) + \lambda^{*t} Dh(\mathbf{x}^*) - \mu^{*t} Dg(\mathbf{x}^*) = \mathbf{0}^t$
- $\mu^{*t} g(\mathbf{x}^*) = 0$
- $h(\mathbf{x}^*) = \mathbf{0}$
- $g(\mathbf{x}^*) \geq \mathbf{0}$

- En multipliant la condition 5 par -1 et en changeant les signes des multiplicateurs, on a les conditions équivalentes suivantes :

- $\mu^* \leq \mathbf{0}$
- $Df(\mathbf{x}^*) + \lambda^{*t} Dh(\mathbf{x}^*) + \mu^{*t} Dg(\mathbf{x}^*) = \mathbf{0}^t$
- $\mu^{*t} g(\mathbf{x}^*) = 0$
- $h(\mathbf{x}^*) = \mathbf{0}$
- $g(\mathbf{x}^*) \geq \mathbf{0}$

Exemple

- Déterminons les optimiseurs de la fonction

$$f(x_1, x_2) = x_1^2 + x_2^2 + x_1 x_2 - 3x_1 \text{ sous les contraintes que } x_1 \geq 0 \text{ et } x_2 \geq 0$$

Conditions de KKT avec contraintes \geq

- On considère le problème de maximisation suivant :

$$\begin{aligned} \max f(\mathbf{x}) \\ \text{s.l.c. } h(\mathbf{x}) = \mathbf{0} \\ g(\mathbf{x}) \geq \mathbf{0} \end{aligned}$$

- Dans ce cas, les conditions de KKT sont :

- $\mu^* \geq \mathbf{0}$
- $Df(\mathbf{x}^*) + \lambda^{*t} Dh(\mathbf{x}^*) + \mu^{*t} Dg(\mathbf{x}^*) = \mathbf{0}^t$
- $\mu^{*t} g(\mathbf{x}^*) = 0$
- $h(\mathbf{x}^*) = \mathbf{0}$
- $g(\mathbf{x}^*) \geq \mathbf{0}$

Exemple

- Déterminons les optimiseurs de la fonction

$$f(x_1, x_2) = x_1^2 + x_2^2 + x_1 x_2 - 3x_1 \text{ sous les contraintes que } x_1 \geq 0 \text{ et } x_2 \geq 0$$

- On cherche les points vérifiant les conditions de KKT suivantes :

Exemple

- Déterminons les optimiseurs de la fonction
 $f(x_1, x_2) = x_1^2 + x_2^2 + x_1x_2 - 3x_1$ sous les contraintes que $x_1 \geq 0$ et $x_2 \geq 0$

- On cherche les points vérifiant les conditions de KKT suivantes :

- 1 $\mu \leq \mathbf{0}$
- 2 $Df(\mathbf{x}) + \mu^t Dg(\mathbf{x}) = \mathbf{0}^t$
- 3 $\mu^t g(\mathbf{x}) = 0$
- 5 $\mathbf{x} \geq \mathbf{0}$

- Nous avons :

$$Df(\mathbf{x}) =$$

Exemple

- Déterminons les optimiseurs de la fonction
 $f(x_1, x_2) = x_1^2 + x_2^2 + x_1x_2 - 3x_1$ sous les contraintes que $x_1 \geq 0$ et $x_2 \geq 0$

- On cherche les points vérifiant les conditions de KKT suivantes :

- 1 $\mu \leq \mathbf{0}$
- 2 $Df(\mathbf{x}) + \mu^t Dg(\mathbf{x}) = \mathbf{0}^t$
- 3 $\mu^t g(\mathbf{x}) = 0$
- 5 $\mathbf{x} \geq \mathbf{0}$

- Nous avons :

$$Df(\mathbf{x}) = (2x_1 + x_2 - 3 \quad x_1 + 2x_2)$$

- Les conditions 2 et 3 sont équivalentes au système suivant :

$$\begin{cases} 2x_1 + x_2 + \mu_1 = 3 \\ x_1 + 2x_2 + \mu_2 = 0 \\ \mu_1 x_1 + \mu_2 x_2 = 0 \end{cases}$$

Exemple

- Déterminons les optimiseurs de la fonction
 $f(x_1, x_2) = x_1^2 + x_2^2 + x_1x_2 - 3x_1$ sous les contraintes que $x_1 \geq 0$ et $x_2 \geq 0$

- On cherche les points vérifiant les conditions de KKT suivantes :

- 1 $\mu \leq \mathbf{0}$
- 2 $Df(\mathbf{x}) + \mu^t Dg(\mathbf{x}) = \mathbf{0}^t$
- 3 $\mu^t g(\mathbf{x}) = 0$
- 5 $\mathbf{x} \geq \mathbf{0}$

- Nous avons :

$$Df(\mathbf{x}) = (2x_1 + x_2 - 3 \quad x_1 + 2x_2)$$

- Les conditions 2 et 3 sont équivalentes au système suivant :

- a On pose $\mu_1^* = 0$ et $\mu_2^* < 0$ (donc $x_2^* = 0$ car condition 3)

Exemple

Exemple

- a On pose $\mu_1^* = 0$ et $\mu_2^* < 0$ (donc $x_2^* = 0$ car condition 3)
 En remplaçant dans le système on obtient $x_1^* = \frac{3}{2}$ et $\mu_2^* = -\frac{3}{2}$
 Ce point satisfait aux conditions de KKT
- b On pose maintenant $\mu_2^* = 0$ et $\mu_1^* < 0$ (donc $x_1^* = 0$)

Exemple

- a On pose $\mu_1^* = 0$ et $\mu_2^* < 0$ (donc $x_2^* = 0$ car condition 3)
 En remplaçant dans le système on obtient $x_1^* = \frac{3}{2}$ et $\mu_2^* = -\frac{3}{2}$
 Ce point satisfait aux conditions de KKT
- b On pose maintenant $\mu_2^* = 0$ et $\mu_1^* < 0$ (donc $x_1^* = 0$)
 En remplaçant dans le système on obtient $x_2^* = 0$ et $\mu_1^* = 3$
 $\mu_1^* > 0$ donc ce point ne satisfait pas aux conditions de KKT

⇒ Le point $\mathbf{x}^* = (\frac{3}{2}, 0)$, $\mu^* = (0, -\frac{3}{2})$ est le seul point candidat mais on n'est pas sûr qu'il s'agisse d'un minimiseur

- Comme pour précédemment, nous étudions dans la suite des conditions de second ordre permettant de caractériser les points critiques

Exemple

- a On pose $\mu_1^* = 0$ et $\mu_2^* < 0$ (donc $x_2^* = 0$ car condition 3)
 En remplaçant dans le système on obtient $x_1^* = \frac{3}{2}$ et $\mu_2^* = -\frac{3}{2}$
 Ce point satisfait aux conditions de KKT
- b On pose maintenant $\mu_2^* = 0$ et $\mu_1^* < 0$ (donc $x_1^* = 0$)
 En remplaçant dans le système on obtient $x_2^* = 0$ et $\mu_1^* = 3$
 $\mu_1^* > 0$ donc ce point ne satisfait pas aux conditions de KKT

Conditions nécessaires du second ordre (CNSO)

- On suppose désormais que $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ et $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$ sont de classe \mathcal{C}^2 et soit $l(\mathbf{x}, \lambda, \mu)$ le Lagrangien défini par :

$$l(\mathbf{x}, \lambda, \mu) = f(\mathbf{x}) + \lambda^t h(\mathbf{x}) + \mu^t g(\mathbf{x})$$
- Soit $D_x^2 l(\mathbf{x}, \lambda, \mu)$ la matrice hessienne ou hessien de l et soit D_x^2 la dérivée seconde par rapport à \mathbf{x}
- On a $D_x^2 l(\mathbf{x}, \lambda, \mu) = D_x^2 f(\mathbf{x}) + \lambda_1 D_x^2 h_1(\mathbf{x}) + \dots + \lambda_m D_x^2 h_m(\mathbf{x}) + \mu_1 D_x^2 g_1(\mathbf{x}) + \dots + \mu_p D_x^2 g_p(\mathbf{x})$ avec :

$$D_x^2 g_k = \begin{pmatrix} \frac{\partial^2 g_k}{\partial x_1^2} & \frac{\partial^2 g_k}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 g_k}{\partial x_1 \partial x_n} \\ \frac{\partial^2 g_k}{\partial x_2 \partial x_1} & \frac{\partial^2 g_k}{\partial x_2^2} & \dots & \frac{\partial^2 g_k}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \dots & \vdots \\ \frac{\partial^2 g_k}{\partial x_n \partial x_1} & \frac{\partial^2 g_k}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 g_k}{\partial x_n^2} \end{pmatrix}$$

- Dans ce qui suit, l'espace tangent en \mathbf{x}^* est défini par :

$$T(\mathbf{x}^*) = \{\mathbf{y} \in \mathbb{R}^n : Dh(\mathbf{x}^*)\mathbf{y} = \mathbf{0}, Dg_j(\mathbf{x}^*)\mathbf{y} = \mathbf{0}, j \in \mathbb{J}(\mathbf{x}^*)\}$$

Conditions nécessaires du second ordre (CNSO) (suite)

Théorème. (CNSO)

Soit \mathbf{x}^* un minimiseur local de $f : \mathbb{R}^n \rightarrow \mathbb{R}$ satisfaisant aux contraintes $h(\mathbf{x}) = 0, h : \mathbb{R}^n \rightarrow \mathbb{R}^m, m \leq n, g(\mathbf{x}) \leq 0, g : \mathbb{R}^n \rightarrow \mathbb{R}^p$, et $f, h, g \in \mathcal{C}^2$. Alors, si \mathbf{x}^* est un point régulier, il existe $\lambda^* \in \mathbb{R}^m$ et $\mu^* \in \mathbb{R}^p$ tels que :

- 1 $\mu^* \geq 0, Df(\mathbf{x}^*) + \lambda^{*t} Dh(\mathbf{x}^*) + \mu^{*t} Dg(\mathbf{x}^*) = \mathbf{0}^t, \mu^{*t} g(\mathbf{x}^*) = 0$
- 2 Pour tout $\mathbf{y} \in T(\mathbf{x}^*)$, nous avons : $\mathbf{y}^t D_x^2 l(\mathbf{x}^*, \lambda^*, \mu^*) \mathbf{y} \geq 0$

Exemple

- On cherche à minimiser la fonction $f(\mathbf{x}) = (x_1 - 1)^2 + x_2 - 2$ sous les contraintes : $h(\mathbf{x}) = x_2 - x_1 - 1 = 0$ et $g(\mathbf{x}) = x_1 + x_2 - 2 \leq 0$

Conditions suffisantes du second ordre (CSSO)

- Dans ce qui suit, l'espace tangent en \mathbf{x}^* est défini par :
 $\bar{T}(\mathbf{x}^*, \mu^*) = \{\mathbf{y} \in \mathbb{R}^n : Dh(\mathbf{x}^*)\mathbf{y} = \mathbf{0}, Dg_j(\mathbf{x}^*)\mathbf{y} = 0, j \in \bar{\mathbb{J}}(\mathbf{x}^*, \mu^*)\}$
 où $\bar{\mathbb{J}}(\mathbf{x}^*, \mu^*) = \{j : g_j(\mathbf{x}^*) = 0, \mu_j^* > 0\}$
- Remarquons que $\bar{\mathbb{J}}(\mathbf{x}^*, \mu^*) \subset \mathbb{J}(\mathbf{x}^*)$ ce qui en revanche implique $T(\mathbf{x}^*) \subset \bar{T}(\mathbf{x}^*, \mu^*)$

Théorème. (CSSO)

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}, h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ et $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$, trois fonctions de classe \mathcal{C}^2 . S'il existe un point réalisable $\mathbf{x}^* \in \mathbb{R}^n$ et $\lambda^* \in \mathbb{R}^m, \mu^* \in \mathbb{R}^p$ tels que :

- 1 $\mu^* \geq 0, Df(\mathbf{x}^*) + \lambda^{*t} Dh(\mathbf{x}^*) + \mu^{*t} Dg(\mathbf{x}^*) = \mathbf{0}^t, \mu^{*t} g(\mathbf{x}^*) = 0$
 - 2 Pour tout $\mathbf{y} \in \bar{T}(\mathbf{x}^*, \mu^*), \mathbf{y} \neq \mathbf{0}$, nous avons : $\mathbf{y}^t D_x^2 l(\mathbf{x}^*, \lambda^*, \mu^*) \mathbf{y} > 0$
- alors \mathbf{x}^* est un minimiseur local strict de f satisfaisant aux contraintes $h(\mathbf{x}) = \mathbf{0}$ et $g(\mathbf{x}) \leq \mathbf{0}$

- Dans le cas d'un maximiseur, $\mu^* \leq 0$ et la condition 2 est remplacée

Exemple

- On cherche à minimiser la fonction $f(\mathbf{x}) = (x_1 - 1)^2 + x_2 - 2$ sous les contraintes : $h(\mathbf{x}) = x_2 - x_1 - 1 = 0$ et $g(\mathbf{x}) = x_1 + x_2 - 2 \leq 0$
- Pour tout $\mathbf{x} \in \mathbb{R}^2$ on a : $Df(\mathbf{x}) =$

Exemple

- On cherche à minimiser la fonction $f(\mathbf{x}) = (x_1 - 1)^2 + x_2 - 2$ sous les contraintes : $h(\mathbf{x}) = x_2 - x_1 - 1 = 0$ et $g(\mathbf{x}) = x_1 + x_2 - 2 \leq 0$
- Pour tout $\mathbf{x} \in \mathbb{R}^2$ on a : $Df(\mathbf{x}) = (2x_1 - 2 \ 1)$, $Dh(\mathbf{x}) =$

Exemple

- On cherche à minimiser la fonction $f(\mathbf{x}) = (x_1 - 1)^2 + x_2 - 2$ sous les contraintes : $h(\mathbf{x}) = x_2 - x_1 - 1 = 0$ et $g(\mathbf{x}) = x_1 + x_2 - 2 \leq 0$
- Pour tout $\mathbf{x} \in \mathbb{R}^2$ on a : $Df(\mathbf{x}) = (2x_1 - 2 \ 1)$, $Dh(\mathbf{x}) = (-1 \ 1)$ et $Dg(\mathbf{x}) = (1 \ 1)$
- On voit que $\nabla h(\mathbf{x}) = Dh(\mathbf{x})^t$ et $\nabla g(\mathbf{x}) = Dg(\mathbf{x})^t$ sont linéairement indépendants et donc tout point \mathbf{x} est régulier

Exemple

- On cherche à minimiser la fonction $f(\mathbf{x}) = (x_1 - 1)^2 + x_2 - 2$ sous les contraintes : $h(\mathbf{x}) = x_2 - x_1 - 1 = 0$ et $g(\mathbf{x}) = x_1 + x_2 - 2 \leq 0$
- Pour tout $\mathbf{x} \in \mathbb{R}^2$ on a : $Df(\mathbf{x}) = (2x_1 - 2 \ 1)$, $Dh(\mathbf{x}) = (-1 \ 1)$ et $Dg(\mathbf{x}) =$

Exemple

- On cherche à minimiser la fonction $f(\mathbf{x}) = (x_1 - 1)^2 + x_2 - 2$ sous les contraintes : $h(\mathbf{x}) = x_2 - x_1 - 1 = 0$ et $g(\mathbf{x}) = x_1 + x_2 - 2 \leq 0$
- Pour tout $\mathbf{x} \in \mathbb{R}^2$ on a : $Df(\mathbf{x}) = (2x_1 - 2 \ 1)$, $Dh(\mathbf{x}) = (-1 \ 1)$ et $Dg(\mathbf{x}) = (1 \ 1)$
- On voit que $\nabla h(\mathbf{x}) = Dh(\mathbf{x})^t$ et $\nabla g(\mathbf{x}) = Dg(\mathbf{x})^t$ sont linéairement indépendants et donc tout point \mathbf{x} est régulier
- On écrit les conditions de KKT :

$$Df(\mathbf{x}) + \lambda^t Dh(\mathbf{x}) + \mu^t Dg(\mathbf{x}) = \mathbf{0}^t \Leftrightarrow \begin{matrix} \mu \geq 0 \\ \begin{pmatrix} 2x_1 - 2 - \lambda + \mu \\ 1 + \lambda + \mu \end{pmatrix}^t = \begin{pmatrix} 0 \\ 0 \end{pmatrix}^t \\ \mu^t g(\mathbf{x}) = 0 \Leftrightarrow \mu(x_1 + x_2 - 2) = 0 \\ h(\mathbf{x}) = 0 \Leftrightarrow x_2 - x_1 - 1 = 0 \\ g(\mathbf{x}) \leq 0 \Leftrightarrow x_1 + x_2 - 2 \leq 0 \end{matrix}$$

Exemple (suite)

- On cherche les points critiques. On considère dans un premier temps $\mu > 0$ ce qui implique $g(\mathbf{x}) = x_1 + x_2 - 1 = 0$. On obtient alors le système d'équations suivant :

$$\begin{cases} 2x_1 - 2 - \lambda + \mu & = 0 \\ 1 + \lambda + \mu & = 0 \\ x_2 - x_1 - 1 & = 0 \\ x_1 + x_2 - 1 & = 0 \end{cases}$$

Exemple (suite)

- On cherche les points critiques. On considère dans un premier temps $\mu > 0$ ce qui implique $g(\mathbf{x}) = x_1 + x_2 - 1 = 0$. On obtient alors le système d'équations suivant :

$$\begin{cases} 2x_1 - 2 - \lambda + \mu & = 0 \\ 1 + \lambda + \mu & = 0 \\ x_2 - x_1 - 1 & = 0 \\ x_1 + x_2 - 1 & = 0 \end{cases}$$

- La résolution du système donne : $x_1 = \frac{1}{2}, x_2 = \frac{3}{2}, \lambda = -1, \mu = 0$

Exemple (suite)

- On cherche les points critiques. On considère dans un premier temps $\mu > 0$ ce qui implique $g(\mathbf{x}) = x_1 + x_2 - 1 = 0$. On obtient alors le système d'équations suivant :

$$\begin{cases} 2x_1 - 2 - \lambda + \mu & = 0 \\ 1 + \lambda + \mu & = 0 \\ x_2 - x_1 - 1 & = 0 \\ x_1 + x_2 - 1 & = 0 \end{cases}$$

- La résolution du système donne : $x_1 = \frac{1}{2}, x_2 = \frac{3}{2}, \lambda = -1, \mu = 0$
- Toutefois $\mu = 0$ contredit l'hypothèse $\mu > 0$ donc le point n'est pas candidat

Exemple (suite)

- On considère alors $\mu = 0$ ce qui implique $g(\mathbf{x}) = x_1 + x_2 - 1 \leq 0$. On obtient le système d'équation suivant :

$$\begin{cases} 2x_1 - 2 - \lambda & = 0 \\ 1 + \lambda & = 0 \\ x_2 - x_1 - 1 & = 0 \end{cases}$$

et la solution du système doit également vérifier $x_1 + x_2 - 1 \leq 0$

Exemple (suite)

- On considère alors $\mu = 0$ ce qui implique $g(\mathbf{x}) = x_1 + x_2 - 1 \leq 0$. On obtient le système d'équation suivant :

$$\begin{cases} 2x_1 - 2 - \lambda = 0 \\ 1 + \lambda = 0 \\ x_2 - x_1 - 1 = 0 \end{cases}$$

et la solution du système doit également vérifier $x_1 + x_2 - 1 \leq 0$

- On trouve comme solution : $x_1 = \frac{1}{2}, x_2 = \frac{3}{2}, \lambda = -1$ et on voit que le vecteur $\mathbf{x} = (\frac{1}{2}, \frac{3}{2})$ vérifie la contrainte $g(\mathbf{x}) \leq 0$. Donc $\mathbf{x} = (\frac{1}{2}, \frac{3}{2})$ est un point candidat à être un minimiseur

Exemple (suite)

- Pour savoir si $\mathbf{x} = (\frac{1}{2}, \frac{3}{2})$ est un minimiseur on utilise les conditions suffisantes du second ordre : on forme le Lagrangien $l(\mathbf{x}, \lambda, \mu)$ et on étudie le signe de $D_{\mathbf{x}}^2 l$
- On a : $l(\mathbf{x}, \lambda, \mu) = f(\mathbf{x}) + \lambda^t h(\mathbf{x}) + \mu^t g(\mathbf{x}) = (x_1 - 1)^2 + x_2 - 2 + \lambda(x_2 - x_1 - 1) + \mu(x_1 + x_2 - 1)$

Exemple (suite)

- Pour savoir si $\mathbf{x} = (\frac{1}{2}, \frac{3}{2})$ est un minimiseur on utilise les conditions suffisantes du second ordre : on forme le Lagrangien $l(\mathbf{x}, \lambda, \mu)$ et on étudie le signe de $D_{\mathbf{x}}^2 l$

- Pour savoir si $\mathbf{x} = (\frac{1}{2}, \frac{3}{2})$ est un minimiseur on utilise les conditions suffisantes du second ordre : on forme le Lagrangien $l(\mathbf{x}, \lambda, \mu)$ et on étudie le signe de $D_{\mathbf{x}}^2 l$
- On a : $l(\mathbf{x}, \lambda, \mu) = f(\mathbf{x}) + \lambda^t h(\mathbf{x}) + \mu^t g(\mathbf{x}) = (x_1 - 1)^2 + x_2 - 2 + \lambda(x_2 - x_1 - 1) + \mu(x_1 + x_2 - 1)$
- La matrice hessienne est donnée par : $D_{\mathbf{x}}^2 l(\mathbf{x}, \lambda, \mu) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$

Exemple (suite)

- Pour savoir si $\mathbf{x} = (\frac{1}{2}, \frac{3}{2})$ est un minimiseur on utilise les conditions suffisantes du second ordre : on forme le Lagrangien $l(\mathbf{x}, \lambda, \mu)$ et on étudie le signe de $D_x^2 l$
- On a : $l(\mathbf{x}, \lambda, \mu) = f(\mathbf{x}) + \lambda^t h(\mathbf{x}) + \mu^t g(\mathbf{x}) = (x_1 - 1)^2 + x_2 - 2 + \lambda(x_2 - x_1 - 1) + \mu(x_1 + x_2 - 1)$
- La matrice hessienne est donnée par : $D_x^2 l(\mathbf{x}, \lambda, \mu) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$
- Il faut ensuite déterminer $\bar{T}(\mathbf{x}, \mu)$. Notons que comme $\mu = 0$ la contrainte g n'est pas active et dans ce cas : $\bar{T}(\mathbf{x}, \mu) = \{\mathbf{y} \in \mathbb{R}^2 : Dh(\mathbf{x})\mathbf{y} = \mathbf{0}\} =$

Exemple (suite)

- Pour savoir si $\mathbf{x} = (\frac{1}{2}, \frac{3}{2})$ est un minimiseur on utilise les conditions suffisantes du second ordre : on forme le Lagrangien $l(\mathbf{x}, \lambda, \mu)$ et on étudie le signe de $D_x^2 l$
- On a : $l(\mathbf{x}, \lambda, \mu) = f(\mathbf{x}) + \lambda^t h(\mathbf{x}) + \mu^t g(\mathbf{x}) = (x_1 - 1)^2 + x_2 - 2 + \lambda(x_2 - x_1 - 1) + \mu(x_1 + x_2 - 1)$
- La matrice hessienne est donnée par : $D_x^2 l(\mathbf{x}, \lambda, \mu) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$
- Il faut ensuite déterminer $\bar{T}(\mathbf{x}, \mu)$. Notons que comme $\mu = 0$ la contrainte g n'est pas active et dans ce cas : $\bar{T}(\mathbf{x}, \mu) = \{\mathbf{y} \in \mathbb{R}^2 : Dh(\mathbf{x})\mathbf{y} = \mathbf{0}\} = \{\mathbf{y} \in \mathbb{R}^2 : (-1 \ 1) \mathbf{y} = \mathbf{0}\} = \{\mathbf{y} \in \mathbb{R}^2 : y_1 = y_2 = y, y \in \mathbb{R}\}$

Exemple (suite)

- Pour savoir si $\mathbf{x} = (\frac{1}{2}, \frac{3}{2})$ est un minimiseur on utilise les conditions suffisantes du second ordre : on forme le Lagrangien $l(\mathbf{x}, \lambda, \mu)$ et on étudie le signe de $D_x^2 l$
- On a : $l(\mathbf{x}, \lambda, \mu) = f(\mathbf{x}) + \lambda^t h(\mathbf{x}) + \mu^t g(\mathbf{x}) = (x_1 - 1)^2 + x_2 - 2 + \lambda(x_2 - x_1 - 1) + \mu(x_1 + x_2 - 1)$
- La matrice hessienne est donnée par : $D_x^2 l(\mathbf{x}, \lambda, \mu) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$
- Il faut ensuite déterminer $\bar{T}(\mathbf{x}, \mu)$. Notons que comme $\mu = 0$ la contrainte g n'est pas active et dans ce cas : $\bar{T}(\mathbf{x}, \mu) = \{\mathbf{y} \in \mathbb{R}^2 : Dh(\mathbf{x})\mathbf{y} = \mathbf{0}\} = \{\mathbf{y} \in \mathbb{R}^2 : (-1 \ 1) \mathbf{y} = \mathbf{0}\} = \{\mathbf{y} \in \mathbb{R}^2 : y_1 = y_2 = y, y \in \mathbb{R}\}$
- On a alors $\mathbf{y}^t D_x^2 l(\mathbf{x}, \lambda, \mu) \mathbf{y} = (y \ y) \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} y \\ y \end{pmatrix} = 2y^2$ et on vérifie que $\forall \mathbf{y} \in \bar{T}(\mathbf{x}, \mu), \mathbf{y} \neq \mathbf{0} : \mathbf{y}^t D_x^2 l(\mathbf{x}, \lambda, \mu) \mathbf{y} > 0$

Exemple (suite)

- Pour savoir si $\mathbf{x} = (\frac{1}{2}, \frac{3}{2})$ est un minimiseur on utilise les conditions suffisantes du second ordre : on forme le Lagrangien $l(\mathbf{x}, \lambda, \mu)$ et on étudie le signe de $D_x^2 l$
- On a : $l(\mathbf{x}, \lambda, \mu) = f(\mathbf{x}) + \lambda^t h(\mathbf{x}) + \mu^t g(\mathbf{x}) = (x_1 - 1)^2 + x_2 - 2 + \lambda(x_2 - x_1 - 1) + \mu(x_1 + x_2 - 1)$
- La matrice hessienne est donnée par : $D_x^2 l(\mathbf{x}, \lambda, \mu) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$
- Il faut ensuite déterminer $\bar{T}(\mathbf{x}, \mu)$. Notons que comme $\mu = 0$ la contrainte g n'est pas active et dans ce cas : $\bar{T}(\mathbf{x}, \mu) = \{\mathbf{y} \in \mathbb{R}^2 : Dh(\mathbf{x})\mathbf{y} = \mathbf{0}\} = \{\mathbf{y} \in \mathbb{R}^2 : (-1 \ 1) \mathbf{y} = \mathbf{0}\} = \{\mathbf{y} \in \mathbb{R}^2 : y_1 = y_2 = y, y \in \mathbb{R}\}$
- On a alors $\mathbf{y}^t D_x^2 l(\mathbf{x}, \lambda, \mu) \mathbf{y} = (y \ y) \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} y \\ y \end{pmatrix} = 2y^2$ et on vérifie que $\forall \mathbf{y} \in \bar{T}(\mathbf{x}, \mu), \mathbf{y} \neq \mathbf{0} : \mathbf{y}^t D_x^2 l(\mathbf{x}, \lambda, \mu) \mathbf{y} > 0$
- En appliquant les CSSO, on conclut que $\mathbf{x} = (\frac{1}{2}, \frac{3}{2})$ est un minimiseur local strict de f satisfaisant les contraintes

Rappel du Sommaire

- 1 Rappels de concepts mathématiques
- 2 Bases théoriques de l'optimisation
- 3 Optimisation sans contrainte
- 4 Optimisation linéaire avec contraintes
- 5 Optimisation non linéaire avec contraintes
- 6 Optimisation convexe
- 7 Algorithmes pour problèmes contraints

Introduction

- Les problèmes que nous avons étudiés dans la section précédente sont en général très difficiles à résoudre
- Ces difficultés proviennent de la nature de la fonction objectif ou des contraintes. Par exemple, même si la fonction objectif est relativement simple à optimiser, la nature des contraintes peut rendre difficile la résolution du problème

Rappel du Sommaire

- 1 Rappels de concepts mathématiques
- 2 Bases théoriques de l'optimisation
- 3 Optimisation sans contrainte
- 4 Optimisation linéaire avec contraintes
- 5 Optimisation non linéaire avec contraintes
- 6 Optimisation convexe
 - Fonctions et ensembles convexes
 - Conditions nécessaires et suffisantes du 1er ordre

Introduction

- Les problèmes que nous avons étudiés dans la section précédente sont en général très difficiles à résoudre
- Ces difficultés proviennent de la nature de la fonction objectif ou des contraintes. Par exemple, même si la fonction objectif est relativement simple à optimiser, la nature des contraintes peut rendre difficile la résolution du problème
- Nous nous intéressons ici à une famille de problèmes, les **problèmes convexes** (on parle alors d'**optimisation convexe**) où la fonction objectif est convexe et les contraintes définissent un ensemble convexe. Les problèmes convexes ont des propriétés particulières que nous pouvons plus facilement appréhender. En d'autres termes les problèmes convexes sont plus simples à résoudre que les problèmes non convexes

Introduction

- Les problèmes que nous avons étudiés dans la section précédente sont en général très difficiles à résoudre
- Ces difficultés proviennent de la nature de la fonction objectif ou des contraintes. Par exemple, même si la fonction objectif est relativement simple à optimiser, la nature des contraintes peut rendre difficile la résolution du problème
- Nous nous intéressons ici à une famille de problèmes, les **problèmes convexes** (on parle alors d'**optimisation convexe**) où la fonction objectif est convexe et les contraintes définissent un ensemble convexe. Les problèmes convexes ont des propriétés particulières que nous pouvons plus facilement appréhender. En d'autres termes les problèmes convexes sont plus simples à résoudre que les problèmes non convexes
- Les problèmes linéaires sont par exemple un sous-cas de problèmes convexes

Graphe et épigraphe d'une fonction

Définition. (Graphe d'une fonction)

Le **graphe** de la fonction $f : \mathbb{S} \rightarrow \mathbb{R}$, $\mathbb{S} \subset \mathbb{R}^n$, est l'ensemble des points de $\mathbb{S} \times \mathbb{R} \subset \mathbb{R}^{n+1}$ donné par :

$$\{(\mathbf{x}, f(\mathbf{x})), \mathbf{x} \in \mathbb{S}\}$$

Définition. (Epigraphe)

L'**épigraphe** de la fonction $f : \mathbb{S} \rightarrow \mathbb{R}$, $\mathbb{S} \subset \mathbb{R}^n$, dénoté $\text{epi}(f)$, est l'ensemble des points de $\mathbb{S} \times \mathbb{R} \subset \mathbb{R}^{n+1}$ donné par :

$$\{(\mathbf{x}, y), \mathbf{x} \in \mathbb{S}, y \in \mathbb{R}, y \geq f(\mathbf{x})\}$$

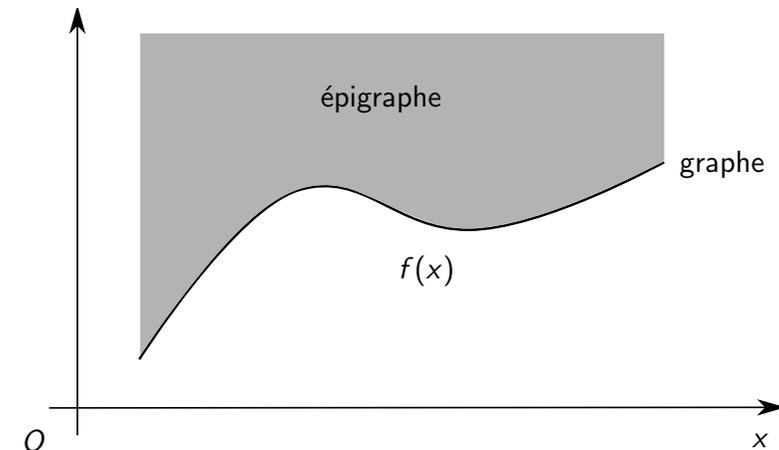
Graphe et épigraphe d'une fonction

Définition. (Graphe d'une fonction)

Le **graphe** de la fonction $f : \mathbb{S} \rightarrow \mathbb{R}$, $\mathbb{S} \subset \mathbb{R}^n$, est l'ensemble des points de $\mathbb{S} \times \mathbb{R} \subset \mathbb{R}^{n+1}$ donné par :

$$\{(\mathbf{x}, f(\mathbf{x})), \mathbf{x} \in \mathbb{S}\}$$

Illustration dans \mathbb{R}



Fonction convexe

Définition. (Fonction convexe)

Une fonction $f : \mathbb{S} \rightarrow \mathbb{R}$, $\mathbb{S} \subset \mathbb{R}^n$, est **convexe** si son épigraphe est un ensemble convexe

Fonction convexe

Définition. (Fonction convexe)

Une fonction $f : \mathbb{S} \rightarrow \mathbb{R}$, $\mathbb{S} \subset \mathbb{R}^n$, est **convexe** si son épigraphe est un ensemble convexe

Théorème. (Fonction convexe)

Si une fonction $f : \mathbb{S} \rightarrow \mathbb{R}$, $\mathbb{S} \subset \mathbb{R}^n$, est convexe alors \mathbb{S} , est un ensemble convexe

Théorème. (Fonction convexe)

Une fonction $f : \mathbb{S} \rightarrow \mathbb{R}$, défini sur un ensemble convexe $\mathbb{S} \subset \mathbb{R}^n$, est **convexe** ssi pour tout $\mathbf{x}, \mathbf{y} \in \mathbb{S}$ et pour tout $\alpha \in [0, 1]$:

$$f(\alpha \mathbf{x} + (1 - \alpha) \mathbf{y}) \leq \alpha f(\mathbf{x}) + (1 - \alpha) f(\mathbf{y})$$

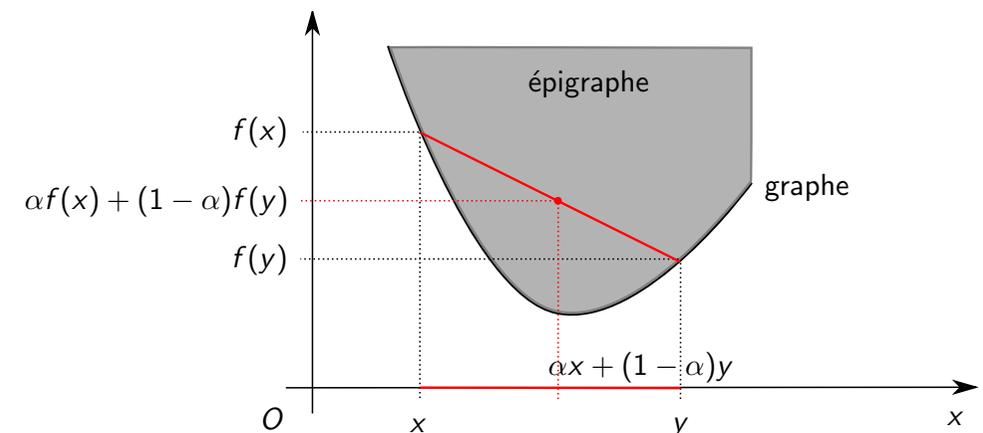
Fonction convexe

Définition. (Fonction convexe)

Une fonction $f : \mathbb{S} \rightarrow \mathbb{R}$, $\mathbb{S} \subset \mathbb{R}^n$, est **convexe** si son épigraphe est un ensemble convexe

Théorème. (Fonction convexe)

Si une fonction $f : \mathbb{S} \rightarrow \mathbb{R}$, $\mathbb{S} \subset \mathbb{R}^n$, est convexe alors \mathbb{S} , est un ensemble convexe

Illustration dans \mathbb{R} 

Le segment reliant $(x, f(x))$ et $(y, f(y))$ est au-dessus ou sur le graphe de f

Fonction strictement convexe et concave

Définition. (Fonction strictement convexe)

Une fonction $f : \mathbb{S} \rightarrow \mathbb{R}$, défini sur un ensemble convexe $\mathbb{S} \subset \mathbb{R}^n$, est **strictement convexe** ssi pour tout $\mathbf{x}, \mathbf{y} \in \mathbb{S}, \mathbf{x} \neq \mathbf{y}$ et pour tout $\alpha \in [0, 1]$:

$$f(\alpha \mathbf{x} + (1 - \alpha)\mathbf{y}) < \alpha f(\mathbf{x}) + (1 - \alpha)f(\mathbf{y})$$

(contre) Exemple

- Soit $f(\mathbf{x}) = x_1 x_2$. Est-ce que f est convexe sur $\mathbb{S} = \{\mathbf{x} \in \mathbb{R}^2 : x_1 \geq 0, x_2 \geq 0\}$?
- Prenons un exemple : $\mathbf{x} = (1, 2)$ et $\mathbf{y} = (2, 1)$

Fonction strictement convexe et concave

Définition. (Fonction strictement convexe)

Une fonction $f : \mathbb{S} \rightarrow \mathbb{R}$, défini sur un ensemble convexe $\mathbb{S} \subset \mathbb{R}^n$, est **strictement convexe** ssi pour tout $\mathbf{x}, \mathbf{y} \in \mathbb{S}, \mathbf{x} \neq \mathbf{y}$ et pour tout $\alpha \in [0, 1]$:

$$f(\alpha \mathbf{x} + (1 - \alpha)\mathbf{y}) < \alpha f(\mathbf{x}) + (1 - \alpha)f(\mathbf{y})$$

Définition. (Fonction (strictement) concave)

Une fonction $f : \mathbb{S} \rightarrow \mathbb{R}$, défini sur un ensemble convexe $\mathbb{S} \subset \mathbb{R}^n$, est **(strictement) concave** ssi $-f$ est (strictement) convexe

(contre) Exemple

- Soit $f(\mathbf{x}) = x_1 x_2$. Est-ce que f est convexe sur $\mathbb{S} = \{\mathbf{x} \in \mathbb{R}^2 : x_1 \geq 0, x_2 \geq 0\}$?
- Prenons un exemple : $\mathbf{x} = (1, 2)$ et $\mathbf{y} = (2, 1)$
- On a : $\alpha \mathbf{x} + (1 - \alpha)\mathbf{y} =$

(contre) Exemple

- Soit $f(\mathbf{x}) = x_1 x_2$. Est-ce que f est convexe sur $\mathbb{S} = \{\mathbf{x} \in \mathbb{R}^2 : x_1 \geq 0, x_2 \geq 0\}$?
- Prenons un exemple : $\mathbf{x} = (1, 2)$ et $\mathbf{y} = (2, 1)$
- On a : $\alpha \mathbf{x} + (1 - \alpha) \mathbf{y} = (2 - \alpha, 1 + \alpha)$
- $f(\alpha \mathbf{x} + (1 - \alpha) \mathbf{y}) =$

(contre) Exemple

- Soit $f(\mathbf{x}) = x_1 x_2$. Est-ce que f est convexe sur $\mathbb{S} = \{\mathbf{x} \in \mathbb{R}^2 : x_1 \geq 0, x_2 \geq 0\}$?
- Prenons un exemple : $\mathbf{x} = (1, 2)$ et $\mathbf{y} = (2, 1)$
- On a : $\alpha \mathbf{x} + (1 - \alpha) \mathbf{y} = (2 - \alpha, 1 + \alpha)$
- $f(\alpha \mathbf{x} + (1 - \alpha) \mathbf{y}) = (2 - \alpha)(1 + \alpha) = 2 + \alpha - \alpha^2$
- $\alpha f(\mathbf{x}) + (1 - \alpha) f(\mathbf{y}) = 2$
- Prenons maintenant $\alpha = \frac{1}{2} \in [0, 1]$, on voit que :

$$f\left(\frac{1}{2} \mathbf{x} + \frac{1}{2} \mathbf{y}\right) =$$

(contre) Exemple

- Soit $f(\mathbf{x}) = x_1 x_2$. Est-ce que f est convexe sur $\mathbb{S} = \{\mathbf{x} \in \mathbb{R}^2 : x_1 \geq 0, x_2 \geq 0\}$?
- Prenons un exemple : $\mathbf{x} = (1, 2)$ et $\mathbf{y} = (2, 1)$
- On a : $\alpha \mathbf{x} + (1 - \alpha) \mathbf{y} = (2 - \alpha, 1 + \alpha)$
- $f(\alpha \mathbf{x} + (1 - \alpha) \mathbf{y}) = (2 - \alpha)(1 + \alpha) = 2 + \alpha - \alpha^2$
- $\alpha f(\mathbf{x}) + (1 - \alpha) f(\mathbf{y}) =$

(contre) Exemple

- Soit $f(\mathbf{x}) = x_1 x_2$. Est-ce que f est convexe sur $\mathbb{S} = \{\mathbf{x} \in \mathbb{R}^2 : x_1 \geq 0, x_2 \geq 0\}$?
- Prenons un exemple : $\mathbf{x} = (1, 2)$ et $\mathbf{y} = (2, 1)$
- On a : $\alpha \mathbf{x} + (1 - \alpha) \mathbf{y} = (2 - \alpha, 1 + \alpha)$
- $f(\alpha \mathbf{x} + (1 - \alpha) \mathbf{y}) = (2 - \alpha)(1 + \alpha) = 2 + \alpha - \alpha^2$
- $\alpha f(\mathbf{x}) + (1 - \alpha) f(\mathbf{y}) = 2$
- Prenons maintenant $\alpha = \frac{1}{2} \in [0, 1]$, on voit que :

$$f\left(\frac{1}{2} \mathbf{x} + \frac{1}{2} \mathbf{y}\right) = \frac{9}{4}$$

(contre) Exemple

- Soit $f(\mathbf{x}) = x_1 x_2$. Est-ce que f est convexe sur $\mathbb{S} = \{\mathbf{x} \in \mathbb{R}^2 : x_1 \geq 0, x_2 \geq 0\}$?
- Prenons un exemple : $\mathbf{x} = (1, 2)$ et $\mathbf{y} = (2, 1)$
- On a : $\alpha \mathbf{x} + (1 - \alpha) \mathbf{y} = (2 - \alpha, 1 + \alpha)$
- $f(\alpha \mathbf{x} + (1 - \alpha) \mathbf{y}) = (2 - \alpha)(1 + \alpha) = 2 + \alpha - \alpha^2$
- $\alpha f(\mathbf{x}) + (1 - \alpha) f(\mathbf{y}) = 2$
- Prenons maintenant $\alpha = \frac{1}{2} \in [0, 1]$, on voit que :

$$f\left(\frac{1}{2} \mathbf{x} + \frac{1}{2} \mathbf{y}\right) = \frac{9}{4} > \frac{1}{2} f(\mathbf{x}) + \frac{1}{2} f(\mathbf{y}) = 2$$
- Donc f n'est pas convexe

Forme quadratique convexe

Propriété. (Forme quadratique convexe)

Une forme quadratique $f : \mathbb{S} \rightarrow \mathbb{R}$, $\mathbb{S} \subset \mathbb{R}^n$, donnée par $f(\mathbf{x}) = \mathbf{x}^t \mathbf{Q} \mathbf{x}$, avec $\mathbf{Q} \in \mathcal{M}_n$, $\mathbf{Q} = \mathbf{Q}^t$, est **convexe** sur \mathbb{S} ssi pour tout $\mathbf{x}, \mathbf{y} \in \mathbb{S}$:

$$(\mathbf{x} - \mathbf{y})^t \mathbf{Q} (\mathbf{x} - \mathbf{y}) \geq 0$$

Forme quadratique convexe

Propriété. (Forme quadratique convexe)

Une forme quadratique $f : \mathbb{S} \rightarrow \mathbb{R}$, $\mathbb{S} \subset \mathbb{R}^n$, donnée par $f(\mathbf{x}) = \mathbf{x}^t \mathbf{Q} \mathbf{x}$, avec $\mathbf{Q} \in \mathcal{M}_n$, $\mathbf{Q} = \mathbf{Q}^t$, est **convexe** sur \mathbb{S} ssi pour tout $\mathbf{x}, \mathbf{y} \in \mathbb{S}$:

$$(\mathbf{x} - \mathbf{y})^t \mathbf{Q} (\mathbf{x} - \mathbf{y}) \geq 0$$

- Remarque : l'exemple précédent $f(\mathbf{x}) = x_1 x_2$ est une forme quadratique avec $\mathbf{Q} = \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

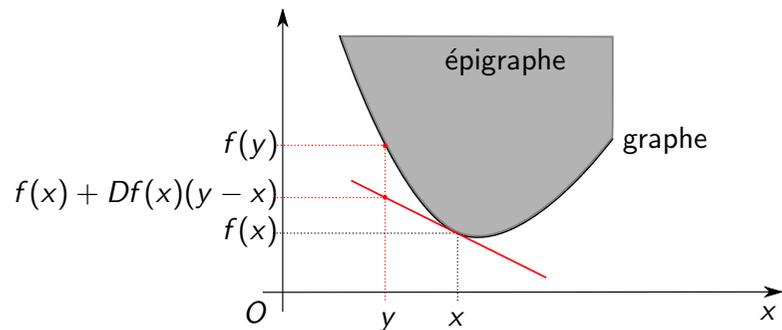
Convexité et fonction de classe \mathcal{C}^1

- La convexité des fonctions différentiables peut être caractérisée par le théorème suivant :

Propriété.

Soit $f : \mathbb{S} \rightarrow \mathbb{R}$, $f \in \mathcal{C}^1$, une fonction définie sur un ensemble ouvert convexe $\mathbb{S} \subset \mathbb{R}^n$. Alors f est convexe sur \mathbb{S} ssi pour tout $\mathbf{x}, \mathbf{y} \in \mathbb{S}$:

$$f(\mathbf{y}) \geq f(\mathbf{x}) + Df(\mathbf{x})(\mathbf{y} - \mathbf{x})$$

Illustration dans \mathbb{R} 

La fonction $g(\mathbf{y}) = f(\mathbf{x}) + Df(\mathbf{x})(\mathbf{y} - \mathbf{x})$ est l'approximation affine de f en \mathbf{x} . f est convexe si $g(\mathbf{y})$ est au-dessous du graphe de f

Exemple

- Déterminer si la fonction $f(\mathbf{x}) = 2x_1x_2 - x_1^2 - x_2^2$ est convexe ou concave ou ni l'une ni l'autre

Convexité et fonction de classe \mathcal{C}^2

- La convexité des fonctions différentiables deux fois peut être caractérisée par le théorème suivant :

Propriété.

Soit $f : \mathbb{S} \rightarrow \mathbb{R}$, $f \in \mathcal{C}^2$, une fonction définie sur un ensemble ouvert convexe $\mathbb{S} \subset \mathbb{R}^n$. Alors f est convexe sur \mathbb{S} ssi pour tout $\mathbf{x} \in \mathbb{S}$, la matrice hessienne de f en \mathbf{x} , $D^2f(\mathbf{x})$, est semi-définie positive

- Notons que suivant la définition de la concavité, une fonction $f : \mathbb{S} \rightarrow \mathbb{R}$, $f \in \mathcal{C}^2$, est concave sur \mathbb{S} , ssi pour tout $\mathbf{x} \in \mathbb{S}$, la matrice hessienne $D^2f(\mathbf{x})$ est semi-définie négative

Exemple

- Déterminer si la fonction $f(\mathbf{x}) = 2x_1x_2 - x_1^2 - x_2^2$ est convexe ou concave ou ni l'une ni l'autre
- Dans un premier temps, la matrice jacobienne vaut $Df(\mathbf{x}) =$

Exemple

- Déterminer si la fonction $f(\mathbf{x}) = 2x_1x_2 - x_1^2 - x_2^2$ est convexe ou concave ou ni l'une ni l'autre
- Dans un premier temps, la matrice jacobienne vaut $Df(\mathbf{x}) = (2x_2 - 2x_1 \quad 2x_1 - 2x_2)$
- Dans un second temps, la matrice hessienne vaut $D^2f(\mathbf{x}) =$

Exemple

- Déterminer si la fonction $f(\mathbf{x}) = 2x_1x_2 - x_1^2 - x_2^2$ est convexe ou concave ou ni l'une ni l'autre
- Dans un premier temps, la matrice jacobienne vaut $Df(\mathbf{x}) = (2x_2 - 2x_1 \quad 2x_1 - 2x_2)$
- Dans un second temps, la matrice hessienne vaut $D^2f(\mathbf{x}) = \begin{pmatrix} -2 & 2 \\ 2 & -2 \end{pmatrix}$
- Les mineurs principaux dominants valent :
 - ▶ $\Delta_1 = -2$
 - ▶ $\Delta_2 =$

Exemple

- Déterminer si la fonction $f(\mathbf{x}) = 2x_1x_2 - x_1^2 - x_2^2$ est convexe ou concave ou ni l'une ni l'autre
- Dans un premier temps, la matrice jacobienne vaut $Df(\mathbf{x}) = (2x_2 - 2x_1 \quad 2x_1 - 2x_2)$
- Dans un second temps, la matrice hessienne vaut $D^2f(\mathbf{x}) = \begin{pmatrix} -2 & 2 \\ 2 & -2 \end{pmatrix}$
- Les mineurs principaux dominants valent :
 - ▶ $\Delta_1 =$

Exemple

- Déterminer si la fonction $f(\mathbf{x}) = 2x_1x_2 - x_1^2 - x_2^2$ est convexe ou concave ou ni l'une ni l'autre
- Dans un premier temps, la matrice jacobienne vaut $Df(\mathbf{x}) = (2x_2 - 2x_1 \quad 2x_1 - 2x_2)$
- Dans un second temps, la matrice hessienne vaut $D^2f(\mathbf{x}) = \begin{pmatrix} -2 & 2 \\ 2 & -2 \end{pmatrix}$
- Les mineurs principaux dominants valent :
 - ▶ $\Delta_1 = -2$
 - ▶ $\Delta_2 = 0$
- $D^2f(\mathbf{x})$ est semi-définie négative pour tout $\mathbf{x} \in \mathbb{R}^2$ donc f est concave

Rappel du Sommaire

- 1 Rappels de concepts mathématiques
- 2 Bases théoriques de l'optimisation
- 3 Optimisation sans contrainte
- 4 Optimisation linéaire avec contraintes
- 5 Optimisation non linéaire avec contraintes
- 6 Optimisation convexe
 - Fonctions et ensembles convexes
 - Conditions nécessaires et suffisantes du 1er ordre

Caractéristiques des problèmes convexes

Théorème.

Soit $f : \mathcal{S} \rightarrow \mathbb{R}$ une fonction convexe définie sur un ensemble convexe $\mathcal{S} \subset \mathbb{R}^n$. Alors un point est un minimiseur global de f sur \mathcal{S} ssi il est un minimiseur local de f

Optimisation convexe

- On considère désormais les problèmes pour lesquels la fonction objectif est convexe et le domaine des solutions réalisables est un ensemble convexe : on parle alors d'**optimisation convexe** ou de **problèmes convexes**
- Deux cas particuliers d'optimisation convexe sont :
 - ▶ la programmation linéaire (fonction objectif linéaire et contraintes linéaires)
 - ▶ la programmation quadratique (fonction objectif quadratique et contraintes linéaires)
- Les problèmes convexes sont une famille de problèmes d'optimisation intéressante car comme nous le verrons :
 - ▶ Si un optimiseur existe alors il est global
 - ▶ Les conditions nécessaires de premier ordre pour un optimiseur sont également suffisantes

Caractéristiques des problèmes convexes

Théorème.

Soit $f : \mathcal{S} \rightarrow \mathbb{R}$ une fonction convexe définie sur un ensemble convexe $\mathcal{S} \subset \mathbb{R}^n$. Alors un point est un minimiseur global de f sur \mathcal{S} ssi il est un minimiseur local de f

- En d'autres termes, tout minimiseur local est minimiseur global

Caractéristiques des problèmes convexes

Théorème.

Soit $f : \mathbb{S} \rightarrow \mathbb{R}$ une fonction convexe définie sur un ensemble convexe $\mathbb{S} \subset \mathbb{R}^n$. Alors un point est un minimiseur global de f sur \mathbb{S} ssi il est un minimiseur local de f

- En d'autres termes, tout minimiseur local est minimiseur global

Lemme.

Soit $g : \mathbb{S} \rightarrow \mathbb{R}$ une fonction convexe définie sur un ensemble convexe $\mathbb{S} \subset \mathbb{R}^n$. Alors pour tout $c \in \mathbb{R}$, l'ensemble $\Gamma_c = \{\mathbf{x} \in \mathbb{S} : g(\mathbf{x}) \leq c\}$ est un ensemble convexe

Caractéristiques des problèmes convexes

Théorème.

Soit $f : \mathbb{S} \rightarrow \mathbb{R}$ une fonction convexe définie sur un ensemble convexe $\mathbb{S} \subset \mathbb{R}^n$. Alors un point est un minimiseur global de f sur \mathbb{S} ssi il est un minimiseur local de f

- En d'autres termes, tout minimiseur local est minimiseur global

Lemme.

Soit $g : \mathbb{S} \rightarrow \mathbb{R}$ une fonction convexe définie sur un ensemble convexe $\mathbb{S} \subset \mathbb{R}^n$. Alors pour tout $c \in \mathbb{R}$, l'ensemble $\Gamma_c = \{\mathbf{x} \in \mathbb{S} : g(\mathbf{x}) \leq c\}$ est un ensemble convexe

Corollaire.

Soit $f : \mathbb{S} \rightarrow \mathbb{R}$ une fonction convexe définie sur un ensemble convexe $\mathbb{S} \subset \mathbb{R}^n$. Alors l'ensemble des minimiseurs globaux de f sur \mathbb{S} est un ensemble convexe

Conditions nécessaires **et suffisantes** du 1er ordre (CNSPO)

Lemme.

Soit $f : \mathbb{S} \rightarrow \mathbb{R}$ une fonction convexe définie sur un ensemble convexe $\mathbb{S} \subset \mathbb{R}^n$ telle que f est de classe \mathcal{C}^1 sur \mathbb{S} . Supposons que $\mathbf{x}^* \in \mathbb{S}$ est tel que pour tout $\mathbf{x} \in \mathbb{S}, \mathbf{x} \neq \mathbf{x}^*$ nous ayons :

$$Df(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) \geq 0$$

alors \mathbf{x}^* est un minimiseur global de f sur \mathbb{S}

Conditions nécessaires **et suffisantes** du 1er ordre (CNSPO)

Lemme.

Soit $f : \mathbb{S} \rightarrow \mathbb{R}$ une fonction convexe définie sur un ensemble convexe $\mathbb{S} \subset \mathbb{R}^n$ telle que f est de classe \mathcal{C}^1 sur \mathbb{S} . Supposons que $\mathbf{x}^* \in \mathbb{S}$ est tel que pour tout $\mathbf{x} \in \mathbb{S}, \mathbf{x} \neq \mathbf{x}^*$ nous ayons :

$$Df(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) \geq 0$$

alors \mathbf{x}^* est un minimiseur global de f sur \mathbb{S}

Démonstration.

- Comme f est convexe sur \mathbb{S} on a :
 $\forall \mathbf{x} \in \mathbb{S} : f(\mathbf{x}) \geq f(\mathbf{x}^*) + Df(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*)$

Conditions nécessaires **et suffisantes** du 1er ordre (CNSPO)

Lemme.

Soit $f : \mathbb{S} \rightarrow \mathbb{R}$ une fonction convexe définie sur un ensemble convexe $\mathbb{S} \subset \mathbb{R}^n$ telle que f est de classe \mathcal{C}^1 sur \mathbb{S} . Supposons que $\mathbf{x}^* \in \mathbb{S}$ est tel que pour tout $\mathbf{x} \in \mathbb{S}$, $\mathbf{x} \neq \mathbf{x}^*$ nous ayons :

$$Df(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) \geq 0$$

alors \mathbf{x}^* est un minimiseur global de f sur \mathbb{S}

Démonstration.

- Comme f est convexe sur \mathbb{S} on a :
 $\forall \mathbf{x} \in \mathbb{S} : f(\mathbf{x}) \geq f(\mathbf{x}^*) + Df(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*)$
- Par conséquent :
 $\forall \mathbf{x} \in \mathbb{S} : Df(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) \geq 0 \Rightarrow \forall \mathbf{x} \in \mathbb{S} : f(\mathbf{x}) \geq f(\mathbf{x}^*)$

Conditions nécessaires **et suffisantes** du 1er ordre (CNSPO)

Lemme.

Soit $f : \mathbb{S} \rightarrow \mathbb{R}$ une fonction convexe définie sur un ensemble convexe $\mathbb{S} \subset \mathbb{R}^n$ telle que f est de classe \mathcal{C}^1 sur \mathbb{S} . Supposons que $\mathbf{x}^* \in \mathbb{S}$ est tel que pour tout $\mathbf{x} \in \mathbb{S}$, $\mathbf{x} \neq \mathbf{x}^*$ nous ayons :

$$Df(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) \geq 0$$

alors \mathbf{x}^* est un minimiseur global de f sur \mathbb{S}

Démonstration.

- Comme f est convexe sur \mathbb{S} on a :
 $\forall \mathbf{x} \in \mathbb{S} : f(\mathbf{x}) \geq f(\mathbf{x}^*) + Df(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*)$
- Par conséquent :
 $\forall \mathbf{x} \in \mathbb{S} : Df(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) \geq 0 \Rightarrow \forall \mathbf{x} \in \mathbb{S} : f(\mathbf{x}) \geq f(\mathbf{x}^*)$
- Ce qui montre que \mathbf{x}^* est un minimiseur global

Conditions nécessaires **et suffisantes** du 1er ordre (CNSPO) (suite)

- On remarquera dans le lemme précédent $\mathbf{x} - \mathbf{x}^*$ peut être interprété comme étant une direction admissible en \mathbf{x}^*

Conditions nécessaires **et suffisantes** du 1er ordre (CNSPO) (suite)

- On remarquera dans le lemme précédent $\mathbf{x} - \mathbf{x}^*$ peut être interprété comme étant une direction admissible en \mathbf{x}^*
- Le lemme précédent peut donc être interprété de la manière suivante :

Conditions nécessaires et suffisantes du 1er ordre (CNSPO) (suite)

- On remarquera dans le lemme précédent $\mathbf{x} - \mathbf{x}^*$ peut être interprété comme étant une direction admissible en \mathbf{x}^*
- Le lemme précédent peut donc être interprété de la manière suivante :

Théorème. (CNSPO)

Soit $f : \mathbb{S} \rightarrow \mathbb{R}$ une fonction convexe définie sur un ensemble convexe $\mathbb{S} \subset \mathbb{R}^n$ et telle que f est de classe \mathcal{C}^1 sur \mathbb{S} . Supposons que $\mathbf{x}^* \in \mathbb{S}$ est tel que pour toute direction admissible \mathbf{d} en \mathbf{x}^* , nous ayons :

$$\mathbf{d}^t \nabla f(\mathbf{x}^*) \geq 0$$

alors \mathbf{x}^* est un minimiseur global de f sur \mathbb{S}

Conditions nécessaires et suffisantes du 1er ordre (CNSPO) (suite)

- Dans le cas de points intérieurs nous avons le résultat suivant :

- Nous étudions les problèmes convexes avec des contraintes d'égalités :

$$\begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s/c} & h(\mathbf{x}) = \mathbf{0} \end{array}$$

où :

- ▶ $f : \mathbb{S} \rightarrow \mathbb{R}$ est une fonction convexe de classe \mathcal{C}^1 sur $\mathbb{S} \subset \mathbb{R}^n$
- ▶ $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ est une fonction vectorielle de classe \mathcal{C}^1 , et $\mathbb{S} = \{\mathbf{x} \in \mathbb{R}^n : h(\mathbf{x}) = \mathbf{0}\}$ est un ensemble convexe

Conditions nécessaires et suffisantes du 1er ordre (CNSPO) (suite)

- Dans le cas de points intérieurs nous avons le résultat suivant :

Corollaire.

Soit $f : \mathbb{S} \rightarrow \mathbb{R}$, $f \in \mathcal{C}^1$, une fonction convexe définie sur un ensemble convexe $\mathbb{S} \subset \mathbb{R}^n$. Supposons que $\mathbf{x}^* \in \mathbb{S}$ est tel que nous ayons :

$$\nabla f(\mathbf{x}^*) = \mathbf{0}$$

alors \mathbf{x}^* est un minimiseur global de f sur \mathbb{S}

Problèmes convexes avec contraintes d'égalités

Problèmes convexes avec contraintes d'égalités

- Nous étudions les problèmes convexes avec des contraintes d'égalités :

$$\begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.l.c.} & h(\mathbf{x}) = \mathbf{0} \end{array}$$

où :

- $f : \mathbb{S} \rightarrow \mathbb{R}$ est une fonction convexe de classe \mathcal{C}^1 sur $\mathbb{S} \subset \mathbb{R}^n$
- $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ est une fonction vectorielle de classe \mathcal{C}^1 , et $\mathbb{S} = \{\mathbf{x} \in \mathbb{R}^n : h(\mathbf{x}) = \mathbf{0}\}$ est un ensemble convexe

Théorème. (CNSPO (Pb convexe avec contraintes d'égalités))

Soit $f : \mathbb{S} \rightarrow \mathbb{R}$, $f \in \mathcal{C}^1$, une fonction convexe sur l'ensemble $\mathbb{S} = \{\mathbf{x} \in \mathbb{R}^n : h(\mathbf{x}) = \mathbf{0}\}$ où $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $h \in \mathcal{C}^1$ et \mathbb{S} est convexe. Supposons qu'il existe $\mathbf{x}^* \in \mathbb{S}$ et $\lambda^* \in \mathbb{R}^m$ tels que :

$$Df(\mathbf{x}^*) + \lambda^{*t} Dh(\mathbf{x}^*) = \mathbf{0}^t$$

alors \mathbf{x}^* est un minimiseur global de f sur \mathbb{S}

Problèmes convexes avec contraintes d'inégalités

- Nous étudions les problèmes convexes avec des contraintes d'inégalités :

$$\begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.l.c.} & h(\mathbf{x}) = \mathbf{0} \\ & g(\mathbf{x}) \leq \mathbf{0} \end{array}$$

où :

- $f : \mathbb{S} \rightarrow \mathbb{R}$ est une fonction convexe de classe \mathcal{C}^1 sur $\mathbb{S} \subset \mathbb{R}^n$
- $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$ sont des fonctions vectorielles de classe \mathcal{C}^1 , et $\mathbb{S} = \{\mathbf{x} \in \mathbb{R}^n : h(\mathbf{x}) = \mathbf{0}, g(\mathbf{x}) \leq \mathbf{0}\}$ est un ensemble convexe

Problèmes convexes avec contraintes d'inégalités (suite)

Théorème. (CNSPO (Pb convexe avec contraintes d'inégalités))

Soit $f : \mathbb{S} \rightarrow \mathbb{R}$, $f \in \mathcal{C}^1$, une fonction convexe sur l'ensemble $\mathbb{S} = \{\mathbf{x} \in \mathbb{R}^n : h(\mathbf{x}) = \mathbf{0}, g(\mathbf{x}) \leq \mathbf{0}\}$ où $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$; $h, g \in \mathcal{C}^1$ et \mathbb{S} est convexe. Supposons qu'il existe $\mathbf{x}^* \in \mathbb{S}$, $\lambda^* \in \mathbb{R}^m$ et $\mu^* \in \mathbb{R}^p$ tels que :

- $\mu^* \geq \mathbf{0}$
- $Df(\mathbf{x}^*) + \lambda^{*t} Dh(\mathbf{x}^*) + \mu^{*t} Dg(\mathbf{x}^*) = \mathbf{0}^t$
- $\mu^{*t} g(\mathbf{x}^*) = 0$

alors \mathbf{x}^* est un minimiseur global de f sur \mathbb{S}

Rappel du Sommaire

- Rappels de concepts mathématiques
- Bases théoriques de l'optimisation
- Optimisation sans contrainte
- Optimisation linéaire avec contraintes
- Optimisation non linéaire avec contraintes
- Optimisation convexe
- Algorithmes pour problèmes contraints

Introduction

- Nous avons vu quelques algorithmes pour des problèmes non contraints
- Nous donnons brièvement quelques indications sur des algorithmes simples pour des problèmes contraints
- Ces algorithmes sont notamment basés sur ceux vus précédemment. il s'agit des méthodes suivantes :
 - ▶ Méthodes des gradients projetés
 - ▶ Méthodes basées sur les pénalisations
- D'autres techniques existent il s'agit notamment des méthodes suivantes :
 - ▶ Méthode du Lagrangien augmenté
 - ▶ Programmation quadratique successive
 - ▶ Méthode des points intérieurs
 - ▶ ...

Pseudo-code de l'algorithme de descente de gradient projeté

Input : $f \in \mathcal{C}^1, \mathbb{S}$ (un ensemble convexe), $\mathbf{x}^0 \in \mathbb{S}$

```

1   $k \leftarrow 0$ 
2  Tant que condition d'arrêt non satisfaite faire
3      Trouver un pas  $\alpha_k$  tel que  $f(\mathbf{x}^{(k)} - \alpha_k \nabla f(\mathbf{x}^{(k)})) < f(\mathbf{x}^{(k)})$ 
4       $\mathbf{x}^{(k+1)} \leftarrow \pi_{\mathbb{S}}(\mathbf{x}^{(k)} - \alpha_k \nabla f(\mathbf{x}^{(k)}))$ 
5       $k \leftarrow k + 1$ 
6  Fin Tant que
7  Output :  $\mathbf{x}^{(k)}$ 

```

Méthode des gradients projetés

- Principe :
 - 1 Utiliser une méthode de descente de gradient en supposant qu'il n'y a pas de contraintes
 - 2 Projeté le point trouvé à chaque itération sur \mathbb{S} l'ensemble des solutions réalisables
- Notons $\pi_{\mathbb{S}}(\mathbf{x})$ la projection de \mathbf{x} sur \mathbb{S} qui peut-être défini de la façon suivante :

$$\pi_{\mathbb{S}}(\mathbf{x}) = \arg \min_{\mathbf{y} \in \mathbb{S}} \|\mathbf{x} - \mathbf{y}\|$$

- Pour appliquer cette stratégie, il faut que $\pi_{\mathbb{S}}$ soit bien définie ce qui n'est pas le cas pour \mathbb{S} quelconque
- Le cas où \mathbb{S} est un ensemble convexe fermé est un cas bien défini. Toutefois, l'approche peut-être très coûteuse en temps de traitement puisque déterminer le projeté est en soi un problème d'optimisation
- Dans des cas spécifiques l'opérateur $\pi_{\mathbb{S}}$ est facile à déterminer. C'est le cas des contraintes linéaires $\mathbf{Ax} = \mathbf{b}$ où $rg(\mathbf{A}) = m$ puisque dans ce cas, $\pi_{\mathbb{S}}(\mathbf{x}) = (\mathbf{I} - \mathbf{A}^t(\mathbf{AA}^t)^{-1}\mathbf{A})\mathbf{x}$

Méthode basée sur la pénalisation

- Principe :
 - 1 Transformer un problème avec contraintes par un problème sans contrainte dont la fonction objectif pénalise les points non réalisables
 - 2 Utiliser un algorithme pour problème non contraint afin d'approximer une solution du problème original
- Supposons sans perte de généralité que nous avons le problème contraint :

$$\begin{aligned} & \min f(\mathbf{x}) \\ \text{s/c } & g_1(\mathbf{x}) \leq 0 \\ & \vdots \\ & g_p(\mathbf{x}) \leq 0 \end{aligned}$$

où :

- ▶ $f : \mathbb{S} \rightarrow \mathbb{R}$
- ▶ $g_j : \mathbb{R}^n \rightarrow \mathbb{R}, j = 1, \dots, p$
- ▶ $\mathbb{S} = \{\mathbf{x} \in \mathbb{R}^n : g_j(\mathbf{x}) \leq 0, j = 1, \dots, p\}$

Méthode basée sur la pénalisation (suite)

- On transforme alors le problème précédent en un problème pénalisé :

$$\min f(\mathbf{x}) + \gamma P(\mathbf{x})$$

où :

- $\gamma \in \mathbb{R}^+$ est appelé paramètre de pénalité
 - $P : \mathbb{R}^n \rightarrow \mathbb{R}$ est appelée fonction de pénalité qui vérifie les propriétés suivantes :
 - P est continue
 - $\forall \mathbf{x} \in \mathbb{R}^n : P(\mathbf{x}) \geq 0$
 - $P(\mathbf{x}) = 0$ ssi \mathbf{x} est réalisable ie $g_1(\mathbf{x}) \leq 0, \dots, g_p(\mathbf{x}) \leq 0$
- Exemples de fonctions de pénalité :
 - $P(\mathbf{x}) = \sum_{j=1}^p g_j^+(\mathbf{x})$ où $g_j^+(\mathbf{x}) = \max\{0, g_j(\mathbf{x})\}$
 - $P(\mathbf{x}) = \sum_{j=1}^p (g_j^+(\mathbf{x}))^2$

Pseudo-code de l'algorithme des méthodes avec pénalisation

- Soit une suite strictement croissante de réels $\{\gamma_k\}$ (telle que $\gamma_k < \gamma_{k+1}$)
- Notons $q(\gamma_k, \mathbf{x}) = f(\mathbf{x}) + \gamma_k P(\mathbf{x})$ et $\mathbf{x}^{(k)} = \arg \min_q(\gamma_k, \mathbf{x})$

Pseudo-code de l'algorithme des méthodes avec pénalisation

- Soit une suite strictement croissante de réels $\{\gamma_k\}$ (telle que $\gamma_k < \gamma_{k+1}$)
- Notons $q(\gamma_k, \mathbf{x}) = f(\mathbf{x}) + \gamma_k P(\mathbf{x})$ et $\mathbf{x}^{(k)} = \arg \min_q(\gamma_k, \mathbf{x})$

Input : q, \mathbf{x}^0

- $k \leftarrow 1$
- Tant que** condition d'arrêt non satisfaite **faire**
- Résoudre le problème non contraint $q(\gamma_k, \mathbf{x})$
en prenant comme solution initiale $\mathbf{x}^{(k-1)}$
- $k \leftarrow k + 1$
- Fin Tant que**
- Output :** $\mathbf{x}^{(k)}$

Bien fondé de l'algorithme des méthodes avec pénalisation

Théorème.

Supposons que la fonction objectif f soit continue et que $\gamma_k \rightarrow \infty$ quand $k \rightarrow \infty$. Alors, la limite de toute sous-suite convergente de $\{\mathbf{x}^{(k)}\}$ est une solution au problème contraint initial