

- Maîtrise des concepts mathématiques essentiels à la bonne compréhension des méthodes statistiques et d'analyse de données.
- Pour cela, il faut revoir des éléments d'**analyse** et d'**algèbre linéaire** qui permettent ensuite de comprendre les méthodes d'**optimisation**.
- Car beaucoup de méthodes statistiques sont en fait issus de la solution de problèmes d'**optimisation**. Exemples :
 - ▶ **moindres** carrés ordinaires (régression linéaire),
 - ▶ **maximisation** de la vraisemblance (modèle linéaire gaussien),
 - ▶ **maximisation** de l'inertie projeté (réduction de dimensions -ACP, AFC, ACM-) ...

Analyse et Algèbre linéaire (en vue des applications en optimisation/statistique) L3 MIASH-IDS

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

Université Lyon 2

L3 MIASHS-IDS 2019/2020

- Lorsque le problème d'optimisation n'a pas de solution explicite on a recours à des algorithmes d'**optimisation numérique** (comme pour la régression logistique/algorithmes de Newton-Raphson). Là aussi, les espaces euclidiens jouent un rôle fondamental : on part d'un point quelconque et on se déplace dans l'espace jusqu'à ce qu'on trouve un point pour lequel le gradient de la fonction objectif s'annule.
- De plus, les **formes quadratiques** étudiées également en algèbre linéaire sont fondamentales en optimisation. D'une part, certaines fonctions objectif citées précédemment sont des formes quadratiques. D'autre part, pour caractériser si un point stationnaire est un maximum ou un minimum on étudie le signe de la **matrice hessienne** en ce point. Or la matrice hessienne est une forme quadratique.

- L'**analyse** nous donne des outils mathématiques (dérivées premières et secondes, approximation polynomiale de fonctions, ...) permettant d'étudier les fonctions objectif de ces méthodes statistiques. Exemples de fonctions objectif :
 - ▶ Somme des carrés des résidus,
 - ▶ Vraisemblance,
 - ▶ L'inertie projeté ...
- L'**optimisation** nous donne les conditions permettant de caractériser les solutions optimales du problème. Typiquement il s'agit de points stationnaires c-à-d des "points où la dérivée s'annule".
- Dans le cas général multidimensionnel, les **solutions du problème** sont représentés par des **vecteurs** appartenant à des espaces vectoriels. L'**algèbre linéaire** et plus précisément les **espaces euclidiens**, nous donne donc un cadre permettant de représenter et de caractériser les solutions du problème d'optimisation. Le concept de **gradient** est ici fondamental.

Contenu du cours

- Rappels en analyse pour une fonction numérique unidimensionnelle :
 - ▶ Continuité, limite, différentiabilité,
 - ▶ Monotonie, convexité,
 - ▶ Formule de Taylor.
- Optimisation unidimensionnel (théorique et numérique) :
 - ▶ Condition nécessaire du premier et du second ordre,
 - ▶ Condition suffisante du second ordre.
- Rappels d'algèbre linéaire pour le passage au multidimensionnel :
 - ▶ Vecteurs, espaces vectoriels,
 - ▶ Forme bilinéaire, produit scalaire, distance euclidienne, espaces euclidiens,
 - ▶ Formes quadratiques.
- Rappels d'analyse pour une fonction numérique de plusieurs variables.
- Optimisation multidimensionnelle (théorique et numérique).

Déroulement du cours

- 12 séances de CM (1h45).
- 10 séances de TD (1h45).
- 2 partiels (1 à mi-parcours, 1 à la fin du semestre).
- Les supports de cours sont disponibles à l'adresse suivante eric.univ-lyon2.fr/~jahpine/.

Quelques références faisant partie des sources du cours

- D. Fredon, M. Maumy-Bretrand, F. Bertrand. *Mathématiques - Analyse en 30 fiches*. Dunod. 2009.
- J.P. Lecoutre, Ph. Pilibossian. *Analyse I et II*. Dunod. 2004.
- I. Catto, I. Gentil, G. Pons. *Mathématiques, Elements de calcul différentiel pour l'économie*. Ellipses. 2011
- J. Quinet. *Cours élémentaire de mathématiques supérieures - Tome 2 - Fonctions usuelles*. Dunod. 1976.
- G. Grifone. *Algèbre Linéaire (2ème édition)*. Cepadués. 2002.

Quelques méthodes statistiques du point de vue de l'optimisation

Rappel du Sommaire

- 1 Quelques méthodes statistiques du point de vue de l'optimisation
- 2 Fonctions d'une variable
- 3 Optimisation unidimensionnelle
- 4 Algèbre linéaire
- 5 Fonctions de plusieurs variables
- 6 Optimisation multidimensionnelle

Table de données

- En statistiques, dans le cas général, notre point de départ est une table de données \mathbf{X} de taille $(n \times p)$.

$$\mathbf{X} = \begin{matrix} & \mathbf{x}^1 & \mathbf{x}^2 & \dots & \mathbf{x}^p \\ \mathbf{x}_1 & x_{11} & x_{12} & \dots & x_{1p} \\ \mathbf{x}_2 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{x}_n & x_{n1} & x_{n2} & \dots & x_{np} \end{matrix}$$

- Chaque ligne \mathbf{x}_i avec $i = 1, \dots, n$ correspond à un individu (ou une observation).
- Chaque colonne \mathbf{x}^k avec $k = 1, \dots, p$ correspond à une variable.

Régression linéaire simple

- Supposons que $p = 2$: \mathbf{x} est la variable explicative et \mathbf{y} est la variable à expliquer. Supposons que \mathbf{x} et \mathbf{y} sont quantitatives continues.
- On suppose le modèle suivant, $\forall i$:

$$y_i = a + bx_i + \epsilon_i$$

où ϵ_i est l'erreur pour l'individu i .

- Table de données :

$$\begin{matrix} \mathbf{y} & \mathbf{x} \\ y_1 & x_1 \\ y_2 & x_2 \\ \vdots & \vdots \\ y_n & x_n \end{matrix}$$

- L'estimation de a et b par moindres carrés ordinaires est la solution du problème suivant :

$$\min_{(a,b) \in \mathbb{R}^2} \sum_{i=1}^n \underbrace{(y_i - (a + bx_i))^2}_{\epsilon_i^2}$$

Tendance centrale

- Supposons que $p = 1$ et que l'unique variable notée \mathbf{x} est quantitative.
- On souhaite résumer au mieux \mathbf{x} par un seul nombre. On parle alors de tendance centrale.

- Table de données :

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

- La moyenne arithmétique est la solution du problème suivant :

$$\min_{a \in \mathbb{R}} \sum_{i=1}^n (x_i - a)^2$$

Régression linéaire multiple

- Supposons que $p > 2$: \mathbf{x}^k avec $k = 1, \dots, p$ sont les variables dites explicatives et \mathbf{y} est la 2ème variable dite "à expliquer".
- On suppose le modèle suivant, $\forall i$:

$$y_i = a_0 + \sum_{k=1}^p a_k x_{ik} + \epsilon_i$$

où ϵ_i est l'erreur pour l'individu i .

Régression linéaire multiple (suite)

- Table de données :

$$\begin{matrix} \mathbf{y} & \mathbf{x}^1 & \mathbf{x}^2 & \dots & \mathbf{x}^p \\ \begin{pmatrix} y_1 & x_{11} & x_{12} & \dots & x_{1p} \\ y_2 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ y_n & x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} \end{matrix}$$

- L'estimation de $\mathbf{a} = (a_0, a_1, \dots, a_p)$ par moindres carrés ordinaires est la solution du problème suivant :

$$\min_{\mathbf{a} \in \mathbb{R}^{p+1}} \sum_{i=1}^n \underbrace{(y_i - (a_0 + \sum_{k=1}^p a_k x_{ik}))^2}_{\epsilon_i^2}$$

Régression logistique (suite)

- L'estimation de $\mathbf{a} = (a_0, \dots, a_p)$ se fait par maximisation de la log-vraisemblance qui revient à résoudre le problème suivant :

$$\max_{\mathbf{a} \in \mathbb{R}^{p+1}} \sum_{i=1}^n \left(y_i (a_0 + \sum_{k=1}^p a_k x_{ik}) - \log(1 + \exp(a_0 + \sum_{k=1}^p a_k x_{ik})) \right)$$

Régression logistique

- Supposons que $p > 2$ mais que \mathbf{y} la variable à expliquer est binaire et peut prendre deux valeurs 0 ou 1.
- On suppose que les $y_i | \mathbf{x}_i$ sont des réalisations de variables aléatoires $Y_i | X_i$ i.i.d. selon une loi binomiale. La vraisemblance s'écrit :

$$\prod_{i=1}^n P(y_i | \mathbf{x}_i) = \prod_{i=1}^n P(1 | \mathbf{x}_i)^{y_i} (1 - P(1 | \mathbf{x}_i))^{1-y_i}$$

- On suppose de plus le modèle logit :

$$\ln \frac{P(Y_i = 1 | X_i = \mathbf{x}_i)}{P(Y_i = 0 | X_i = \mathbf{x}_i)} = a_0 + \sum_{k=1}^p a_k x_{ik}$$

Analyse en composantes principales

- Supposons que $p > 2$ et que nous disposons de \mathbf{x}^k avec $k = 1, \dots, p$ variables quantitatives continues :

$$\mathbf{X} = \begin{matrix} & \mathbf{x}^1 & \mathbf{x}^2 & \dots & \mathbf{x}^p \\ \mathbf{x}_1 & \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} \end{matrix}$$

- On cherche un vecteur \mathbf{u} de \mathbb{R}^p tel que la projection orthogonale des \mathbf{x}_i sur \mathbf{u} conserve au mieux l'inertie.
- On montre que \mathbf{u} est la solution du problème suivant :

$$\min_{\mathbf{u} \in \mathbb{R}^p, \|\mathbf{u}\|^2=1} \sum_{i=1}^n \|\mathbf{x}_i - \langle \mathbf{x}_i, \mathbf{u} \rangle \mathbf{u}\|^2$$

Autres exemples de méthodes statistiques utilisant des outils théoriques et/ou pratiques d'optimisation

- Des méthodes statistiques classiques :
 - ▶ L'analyse factorielle des correspondances ...
 - ▶ La méthode des k -means en classification automatique ...
- Des méthodes statistiques plus récentes :
 - ▶ La régression ridge ...
 - ▶ Les SVM (Support Vector Machines) ...
 - ▶ Les réseaux de neurones et le deep learning ...

Rappel du Sommaire

2 Fonctions d'une variable

- Rappels sur les applications, injections, surjections
 - Fonctions numériques
 - Limite et continuité d'une fonction numérique
 - Sens de variation d'une fonction numérique
 - Dérivabilité d'une fonction numérique
 - Différentiabilité d'une fonction numérique
 - Formules de Taylor (approximation d'ordre n d'une fonction)
 - Convexité

Rappel du Sommaire

- 1 Quelques méthodes statistiques du point de vue de l'optimisation
- 2 **Fonctions d'une variable**
- 3 Optimisation unidimensionnelle
- 4 Algèbre linéaire
- 5 Fonctions de plusieurs variables
- 6 Optimisation multidimensionnelle

Applications et fonctions

Définition. (Grphe de \mathbb{A} vers \mathbb{B})

Soient \mathbb{A} et \mathbb{B} deux ensembles. On appelle *graphe de \mathbb{A} vers \mathbb{B}* , un sous-ensemble du produit cartésien $\mathbb{A} \times \mathbb{B}$.

Définition. (Application de \mathbb{A} vers (ou dans) \mathbb{B})

Soient \mathbb{A} et \mathbb{B} deux ensembles. On appelle f une **application univoque (ou fonction)** de \mathbb{A} vers \mathbb{B} , un graphe de \mathbb{A} vers \mathbb{B} tel que pour tout élément a de \mathbb{A} correspond un unique élément b de \mathbb{B} .

On écrira alors :
$$\left\{ \begin{array}{l} f : \mathbb{A} \rightarrow \mathbb{B} \\ a \rightarrow b \end{array} \right. \text{ ou également } f(a) = b.$$

Si $f(a) = b$ alors on dit que :

- b est l'**image** de a .
- a est un **antécédent** de b .

Injections et surjections

Définition. (Surjection (existence d'un antécédent))

Une application $f : \mathbb{A} \rightarrow \mathbb{B}$ est dite **surjective** si $\forall b \in \mathbb{B}, \exists a \in \mathbb{A}, f(a) = b$.

Définition. (Injection (unicité d'un antécédent))

Une application $f : \mathbb{A} \rightarrow \mathbb{B}$ est dite **injective** si $\forall a, a' \in \mathbb{A}, a \neq a' \Rightarrow f(a) \neq f(a')$ ce qui est équivalent à $\forall a, a' \in \mathbb{A}, f(a) = f(a') \Rightarrow a = a'$.

Bijections et applications réciproques

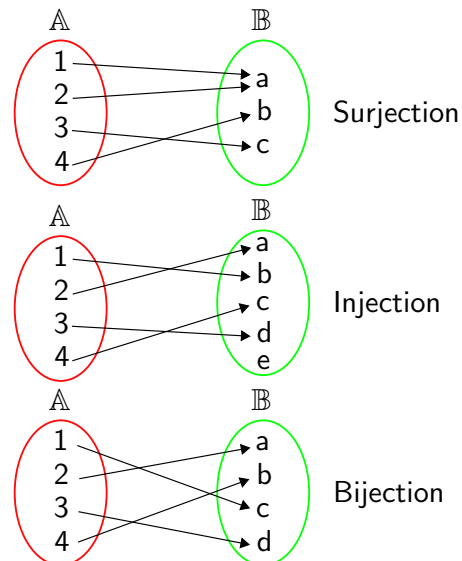
Définition. (Bijection)

Une application $f : \mathbb{A} \rightarrow \mathbb{B}$ est dite **bijection** si elle est surjective et injective. Formellement, f est une bijection si $\forall b \in \mathbb{B}, \exists ! a \in \mathbb{A}, f(a) = b$.

Définition. (Application réciproque)

Une application $f : \mathbb{A} \rightarrow \mathbb{B}$ admet une **application réciproque** notée $f^{-1} : \mathbb{B} \rightarrow \mathbb{A}$ ssi f est une bijection de \mathbb{A} vers \mathbb{B} . Pour tout b dans \mathbb{B} , l'équation $f(a) = b$ admet alors une solution unique a dans \mathbb{A} , notée $a = f^{-1}(b)$. f^{-1} est également appelée **bijection réciproque** de f .

Illustration



Rappel du Sommaire

2 Fonctions d'une variable

- Rappels sur les applications, injections, surjections
- Fonctions numériques
 - Limite et continuité d'une fonction numérique
 - Sens de variation d'une fonction numérique
 - Dérivabilité d'une fonction numérique
 - Différentiabilité d'une fonction numérique
 - Formules de Taylor (approximation d'ordre n d'une fonction)
 - Convexité

Fonction numérique

Définition. (Fonction numérique)

Une fonction numérique (ou plus brièvement **fonction**) f est une application $\mathbb{D}_f \rightarrow \mathbb{R}$ où \mathbb{D}_f , appelé le domaine (de définition), est un sous-ensemble non-vidé de \mathbb{R} ($\mathbb{D}_f \subset \mathbb{R}$). De plus une fonction est telle que pour tout $x \in \mathbb{D}_f$ il existe un unique élément $y \in \mathbb{R}$ tel que $f(x) = y$.

• Quelques exemples classiques :

- ▶ Fonction constante, $f : x \rightarrow a$, où $a \in \mathbb{R}$,
- ▶ Fonction identité, $Id : x \rightarrow x$,
- ▶ Fonctions affines, $f : x \rightarrow a_0 + a_1x$ où $a_0, a_1 \in \mathbb{R}$,
- ▶ Fonctions polynomiales de degré 2, $f : x \rightarrow a_0 + a_1x + a_2x^2$ où $a_0, a_1, a_2 \in \mathbb{R}$,
- ▶ Fonction exponentielle, $f : x \rightarrow \exp(x)$,
- ▶ Fonction logarithme, $f : x \rightarrow \ln(x)$ ($\mathbb{D}_f = \mathbb{R}^{+*}$),
- ▶ Fonction cosinus, $f : x \rightarrow \cos(x)$...

Opérations sur les fonctions numériques

- Soit f et g deux fonctions de domaines \mathbb{D}_f et \mathbb{D}_g et soit λ un réel.
- La fonction $\lambda(f + g)$ est définie sur $\mathbb{D}_f \cap \mathbb{D}_g$ par :

$$(\lambda(f + g))(x) = \lambda f(x) + \lambda g(x)$$

- La fonction fg est définie sur $\mathbb{D}_f \cap \mathbb{D}_g$ par :

$$(fg)(x) = f(x)g(x)$$

- La fonction f/g est définie sur $\mathbb{D}_f \cap \mathbb{D}_g \setminus \{x \in \mathbb{D}_g : g(x) = 0\}$ par :

$$\left(\frac{f}{g}\right)(x) = \frac{f(x)}{g(x)}$$

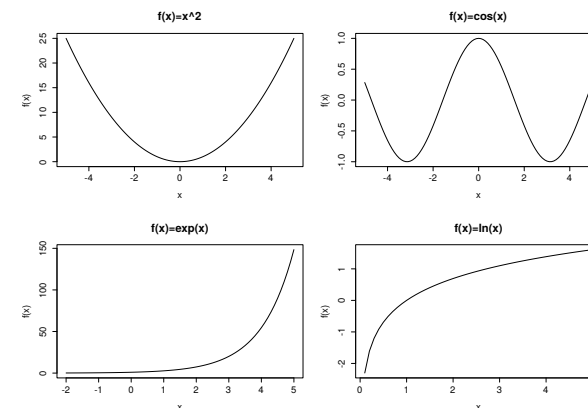
- La fonction composée de f par g notée $g \circ f$, est définie sur $\mathbb{D}_{g \circ f} = \{x \in \mathbb{D}_f : f(x) \in \mathbb{D}_g\}$ et vaut :

$$(g \circ f)(x) = g(f(x))$$

Graphe d'une fonction numérique

Définition. (Graphe d'une fonction)

Le **graphe** de la fonction $f : \mathbb{D}_f \rightarrow \mathbb{R}$ est l'ensemble des points de $\mathbb{D}_f \times \mathbb{R} \subset \mathbb{R}^2$ donné par : $\{(x, f(x)), x \in \mathbb{D}_f\}$.



• Exemples :

Fonctions réciproque

Définition.

Soit f une fonction définie sur \mathbb{D}_f et à image dans $f(\mathbb{D}_f) \subset \mathbb{R}$. Si il existe une fonction $g : f(\mathbb{D}_f) \rightarrow \mathbb{D}_f$ telle que :

$$(g \circ f)(x) = x \text{ et } (f \circ g)(x) = x$$

alors g est la bijection réciproque de f notée f^{-1} .

• Cas usuels :

- ▶ $f(x) = x^2$ sur $\mathbb{D}_f = \mathbb{R}^+$ et $f^{-1}(x) = \sqrt{x}$ sur $\mathbb{D}_{f^{-1}} = \mathbb{R}^+$.
- ▶ $f(x) = \exp(x)$ sur $\mathbb{D}_f = \mathbb{R}$ et $f^{-1}(x) = \ln x$ sur $\mathbb{D}_{f^{-1}} = \mathbb{R}^{+*}$.
- ▶ $f(x) = x^n$ avec $n \in \mathbb{N}^*$ sur $\mathbb{D}_f = \mathbb{R}^+$ et $f^{-1}(x) = x^{1/n}$ sur $\mathbb{D}_{f^{-1}} = \mathbb{R}^+$.

Exemple

- Soit $f(x) = \ln(x)$ et $g(x) = 1 + 2x$.
- $\mathbb{D}_f = \mathbb{R}^{+*}$ et $\mathbb{D}_g = \mathbb{R}$.
- $f^{-1}(x) = \exp(x)$ et $g^{-1}(x) = \frac{1}{2}(x - 1)$.
- $\mathbb{D}_{f^{-1}} = \mathbb{R}$ et $\mathbb{D}_{g^{-1}} = \mathbb{R}$.
- $(g^{-1} \circ g)(x) = \frac{1}{2}((1 + 2x) - 1) = x$.
- Rappelons que $(g \circ f)(x) = g(f(x))$, donc $\mathbb{D}_{g \circ f}$ doit vérifier $x \in \mathbb{D}_f$ et $f(x) \in \mathbb{D}_g$.
- $g \circ f$ est définie sur \mathbb{R}^{+*} et :

$$(g \circ f)(x) = g(f(x)) = 1 + 2 \ln(x)$$

- $f \circ g$ est définie sur $] - 1/2, +\infty[$ et :

$$(f \circ g)(x) = f(g(x)) = \ln(1 + 2x)$$

Rappels sur les suites et leurs limites

- Afin d'introduire les limites des fonctions numériques, il est utile de rappeler au préalable les notions de suites et de limite de suites.
- Une **suite** de nombres réels est une fonction (non numérique) 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 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 $x_1 < x_2 < \dots < x_k < \dots$ c'à d $\forall k, x_k < x_{k+1}$.
- Une suite est dite **croissante** si $\forall k, x_k \leq x_{k+1}$.
 - ▶ Exemple : $\begin{matrix} x_1 & x_2 & x_3 & x_4 & x_5 & \dots \\ 1 & 1 & 2 & 3 & 5 & \dots \end{matrix}$
- 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

2 Fonctions d'une variable

- Rappels sur les applications, injections, surjections
- Fonctions numériques
- Limite et continuité d'une fonction numérique
- Sens de variation d'une fonction numérique
- Dérivabilité d'une fonction numérique
- Différentiabilité d'une fonction numérique
- Formules de Taylor (approximation d'ordre n d'une fonction)
- Convexité

Rappels sur les suites et leurs limites (suite)

- 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 : $\forall k > K, |x_k - x^*| < \epsilon$ c'à d $\forall k > K, x^* - \epsilon < x_k < x^* + \epsilon$. Dans ce cas, on écrit :

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

ou également $x_k \rightarrow x^*$.

- Une suite possédant une limite est dite **convergente**.
- Exemple : $x_k = \frac{1}{k}$ a pour limite $x^* = 0$.
- Toute suite convergente a une et une seule limite.
- $\{x_k\}$ est **bornée** s'il existe un nombre $B \geq 0$ tel que $\forall k, |x_k| \leq B$.
- L'exemple $x_k = \frac{1}{k}$ est une suite bornée ($B = 10$ par exemple).
- De façon générale, toute suite convergente est bornée.

Rappels sur les suites et leurs limites (suite)

- Pour $\{x_k\}$, un nombre B est appelé **majorant** si $\forall k, x_k \leq B$.
- Pour $\{x_k\}$, un nombre B est appelé **minorant** si $\forall k, x_k \geq B$.
- Donc $\{x_k\}$ est **bornée** si elle possède 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.

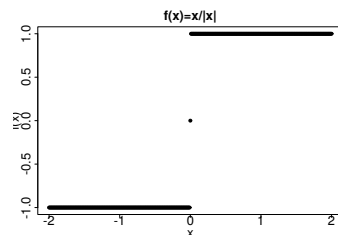
Exemple

- Soit la fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ définie comme suit :

$$f(x) = \begin{cases} \frac{|x|}{x} & \text{si } x \neq 0 \\ 0 & \text{si } x = 0 \end{cases}$$

- Etude de la limite de f en 0.
Deux cas :

- ▶ $x \rightarrow 0$ avec $x > 0$:
 $\lim_{x \rightarrow 0^+} f(x) = 1$.
- ▶ $x \rightarrow 0$ avec $x < 0$:
 $\lim_{x \rightarrow 0^-} f(x) = -1$.



- $\lim_{x \rightarrow 0} f(x) \neq f(0)$ donc f n'est pas continue en 0.

Limite et continuité d'une fonction numérique

- Soit une fonction f et un point $x_0 \in \mathbb{D}_f$. Supposons qu'il existe $f^* \in \mathbb{R}$ tel que **pour toute suite convergente** $\{x_k\}$ **de limite** x_0 :

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

alors nous utiliserons la notation suivante pour désigner la limite f^* :

$$\lim_{x \rightarrow x_0} f(x)$$

Définition. (Fonction continue)

Une fonction $f : \mathbb{D}_f \rightarrow \mathbb{R}$ est dite **continue** en x_0 si elle est définie en ce point (càd $x_0 \in \mathbb{D}_f$) et si :

$$\lim_{x \rightarrow x_0} f(x) = f(x_0)$$

f est dite continue sur \mathbb{D}_f si elle est continue en tout point de \mathbb{D}_f .

Rappel du Sommaire

2 Fonctions d'une variable

- Rappels sur les applications, injections, surjections
- Fonctions numériques
- Limite et continuité d'une fonction numérique
- Sens de variation d'une fonction numérique
- Dérivabilité d'une fonction numérique
- Différentiabilité d'une fonction numérique
- Formules de Taylor (approximation d'ordre n d'une fonction)
- Convexité

Définitions de base

Définition. (Monotonie)

Soit f une fonction définie sur \mathbb{D}_f . On dit que :

- f est **croissante** sur un intervalle \mathbb{I} si $\mathbb{I} \subset \mathbb{D}_f$ et :

$$\forall x_1, x_2 \in \mathbb{I}, x_1 < x_2 \Rightarrow f(x_1) \leq f(x_2)$$

- f est **décroissante** sur un intervalle \mathbb{I} si $\mathbb{I} \subset \mathbb{D}_f$ et :

$$\forall x_1, x_2 \in \mathbb{I}, x_1 < x_2 \Rightarrow f(x_1) \geq f(x_2)$$

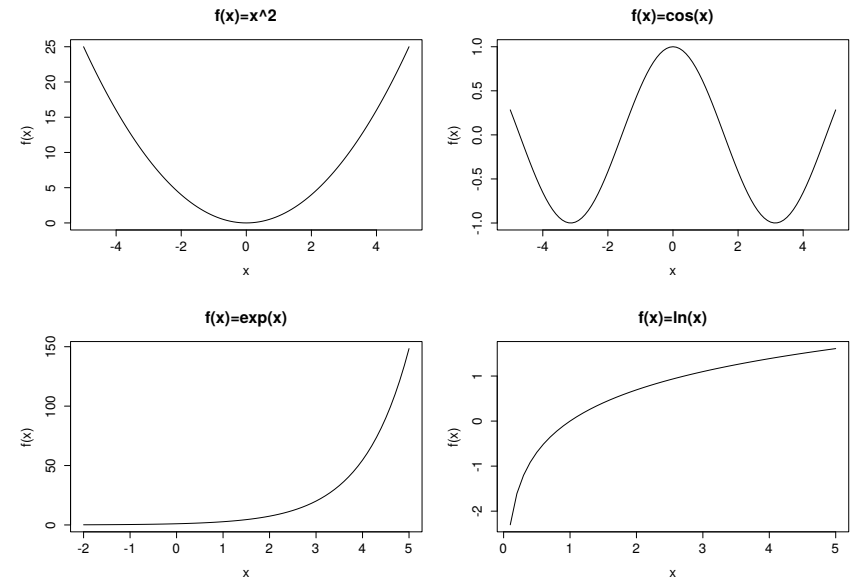
- f est **monotone** si elle est croissante sur \mathbb{I} ou décroissante sur \mathbb{I} .
- Si les inégalités larges précédentes (en rouge) sont remplacées par des inégalités strictes on dira respectivement que : f est strictement croissante, strictement décroissante et strictement monotone sur \mathbb{I} .

Rappel du Sommaire

2 Fonctions d'une variable

- Rappels sur les applications, injections, surjections
- Fonctions numériques
- Limite et continuité d'une fonction numérique
- Sens de variation d'une fonction numérique
- Dérivabilité d'une fonction numérique**
- Différentiabilité d'une fonction numérique
- Formules de Taylor (approximation d'ordre n d'une fonction)
- Convexité

Exemples



Définition de base : nombre dérivé

Définition. (Fonction dérivable)

Soit f une fonction et x_0 un point de \mathbb{D}_f . Soit $f(x_0)$ et $f(x_0 + h)$ les valeurs de f en x_0 et $x_0 + h$. Soit les accroissements absolus suivants :

- $\Delta Id_{x_0}(h) = Id(x_0 + h) - Id(x_0) = h$, accroissement de x à partir de x_0 ,
- $\Delta f_{x_0}(h) = f(x_0 + h) - f(x_0)$, accroissement de $f(x)$ à partir de $f(x_0)$.

Si $\Delta f_{x_0}(h) / \Delta Id_{x_0}(h)$ tend vers une limite finie quand $h \rightarrow 0$ (càd $\Delta Id_{x_0}(h) \rightarrow 0$) alors cette limite est le **nombre dérivé** de f en x_0 noté $f'(x_0)$:

$$f'(x_0) = \lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

Par ailleurs, si f admet un nombre dérivé en tout point $x \in \mathbb{D}_f$ alors on dit que f est **dérivable** sur \mathbb{D}_f .

Illustration et interprétation

- Le rapport $\Delta f_{x_0}(h)/\Delta Id_{x_0}(h)$ représente la tangente (au sens trigonométrique) de l'angle θ (cf ci-dessous à gauche).
- Lorsque $\Delta Id_{x_0}(h) \rightarrow 0$ (càd $h \rightarrow 0$ ou encore $(x_0 + h) \rightarrow x_0$) alors $P' \rightarrow P$ et $\Delta f_{x_0}(h)/\Delta Id_{x_0}(h)$ tend vers la pente de la droite tangente (au sens géométrique) à la courbe en P (cf ci-dessous à droite).
- Illustration avec $f(x) = x^2$:

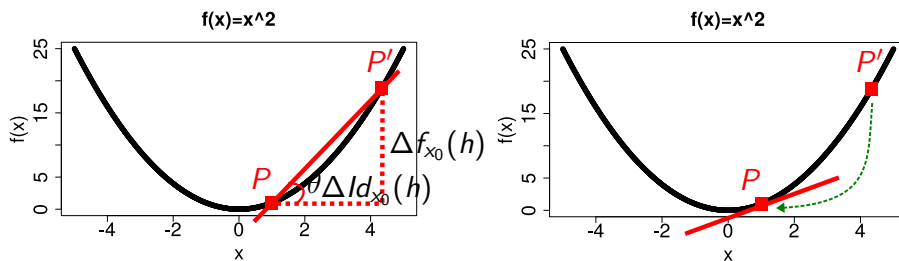


Illustration et interprétation (suite)

- De façon formelle si $f(x) = x^2$ alors $f(x_0) = x_0^2$ et :

$$\begin{aligned} f(x_0 + h) &= (x_0 + h)^2 \\ &= x_0^2 + 2x_0h + h^2 \end{aligned}$$

- La limite de $\Delta f_{x_0}(h)/\Delta Id_{x_0}(h)$ quand $\Delta Id_{x_0}(h) \rightarrow 0$ est telle que :

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{(x_0^2 + 2x_0h + h^2) - (x_0^2)}{h} &= \lim_{h \rightarrow 0} 2x_0 + h \\ &= 2x_0 \end{aligned}$$

- On retrouve le résultat classique du nombre dérivé de x^2 en x_0 :

$$f'(x_0) = 2x_0$$

- La pente de la tangente à la courbe en x_0 vaut $2x_0$.

Définition de la dérivée et des dérivées successives

Définition. (Fonction dérivée)

Si f est dérivable sur \mathbb{D}_f , alors on associe à f sa **dérivée** ou **fonction dérivée**. La dérivée est une fonction numérique notée f' définie par :

$$\begin{aligned} f' : \mathbb{D}_f &\rightarrow \mathbb{R} \\ x &\rightarrow \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \end{aligned}$$

On notera également $f'(x)$ par $\frac{df}{dx}(x)$.

- Si f' , dérivée (première) de f , est dérivable sur \mathbb{D}_f alors on note f'' la dérivée de f' . f'' est appelée **dérivée seconde** de f .
- De façon successive, soit n un entier naturel, on note alors $f^{(n)}$ la **dérivée d'ordre n** de f (on dérive f n fois).
- On notera également $f^{(n)}(x)$ par $\frac{d^n f}{dx^n}(x)$.

Continuité, dérivabilité et classes de fonction dérivables

Théorème.

- Si f est dérivable en $x_0 \in \mathbb{D}_f$ alors f est continue en x_0 .
- Si f est dérivable sur $\mathbb{I} \subset \mathbb{D}_f$ alors f est continue sur \mathbb{I} .

- Attention : la réciproque de ce résultat est fautive (par exemple $f(x) = |x|$ est continue mais non dérivable en 0).

Définition.

Soit n un entier naturel. Soit f une fonction définie sur \mathbb{D}_f . On dit que f est une fonction de classe \mathbb{C}^n sur $\mathbb{I} \subset \mathbb{D}_f$ si elle est n fois dérivable sur \mathbb{I} et si sa dérivée n -ème $f^{(n)}$ est continue sur \mathbb{I} .

Formules de dérivées usuelles

$f(x)$	$f'(x)$	Condition
$a \in \mathbb{R}$	0	
x	1	
$\frac{1}{x}$	$-\frac{1}{x^2}$	$x \neq 0$
$x^n, n \in \mathbb{N}^*$	nx^{n-1}	
$x^{-n}, n \in \mathbb{N}^*$	$-nx^{-n-1}$	$x \neq 0$
\sqrt{x}	$\frac{1}{2\sqrt{x}}$	$x > 0$
$x^{1/n}, n \in \mathbb{N}^*$	$\frac{1}{n}x^{1/n-1}$	$x > 0$ si n est pair
$\sin(x)$	$\cos(x)$	
$\cos(x)$	$-\sin(x)$	
$\ln(x)$	$\frac{1}{x}$	$x > 0$
$\exp(x)$	$\exp(x)$	
$a^x, a \in \mathbb{R}^{+*} \setminus \{1\}$	$a^x \ln(x)$	

Opérations sur les fonctions dérivables

- Soit f et g deux fonctions numériques dérivables en x_0 alors :
 - ▶ La fonction $\lambda(f + g)$ est dérivable en x_0 et :

$$(\lambda(f + g))'(x_0) = \lambda f'(x_0) + \lambda g'(x_0)$$

- ▶ La fonction fg est dérivable en x_0 et :

$$(fg)'(x_0) = f'(x_0)g(x_0) + f(x_0)g'(x_0)$$

- ▶ La fonction f/g est dérivable en x_0 et :

$$\left(\frac{f}{g}\right)'(x_0) = \frac{f'(x_0)g(x_0) - f(x_0)g'(x_0)}{g^2(x_0)}$$

- Si f est dérivable en x_0 et g est dérivable en $f(x_0)$ alors la fonction composée de f par g est dérivable en x_0 et :

$$(g \circ f)'(x_0) = g'(f(x_0))f'(x_0)$$

Règles de dérivation

- Les résultats précédents se généralisent aux dérivées.
- Soit f et g deux fonctions dérivables de dérivées f' et g' .

Fonction	Dérivée	Condition
$f + g$	$f' + g'$	
fg	$f'g + fg'$	
$\frac{f}{g}$	$\frac{f'g - fg'}{g^2}$	$g \neq 0$
$\frac{1}{f}$	$-\frac{f'}{f^2}$	$f \neq 0$
\sqrt{f}	$\frac{f'}{2\sqrt{f}}$	$f > 0$
$f^n, n \in \mathbb{N}^*$	$nf'f^{n-1}$	
$f^{-n}, n \in \mathbb{N}^*$	$-nf'f^{-n-1}$	$f \neq 0$
$\ln(f)$	$\frac{f'}{f}$	$f \neq 0$
$\exp(f)$	$f' \exp(f)$	
$f \circ g$	$(f' \circ g)g'$	

Exemple

- Soit $f(x) = \ln(x)$ et $g(x) = 1 + 2x$.
- $f'(x) = 1/x$ et $g'(x) = 2$
- La dérivée de $g \circ f$ vaut :

$$\begin{aligned} (g \circ f)'(x) &= g'(f(x))f'(x) \\ &= \frac{2}{x} \end{aligned}$$

- La dérivée de $f \circ g$ vaut :

$$\begin{aligned} (f \circ g)'(x) &= f'(g(x))g'(x) \\ &= \frac{2}{1 + 2x} \end{aligned}$$

Application à l'étude du sens de variation

Théorème.

Soit f une fonction définie sur \mathbb{D}_f et dérivable sur \mathbb{I} un intervalle ouvert de \mathbb{D}_f :

- Si $\forall x \in \mathbb{I}, f'(x) = 0$ alors f est constante sur \mathbb{I} .
- Si $\forall x \in \mathbb{I}, f'(x) \geq 0$ alors f est croissante sur \mathbb{I} .
- Si $\forall x \in \mathbb{I}, f'(x) \leq 0$ alors f est décroissante sur \mathbb{I} .

Si les inégalités larges précédentes (en rouge) sont remplacées par des inégalités strictes alors on dit que f est strictement croissante, strictement décroissante sur \mathbb{I} .

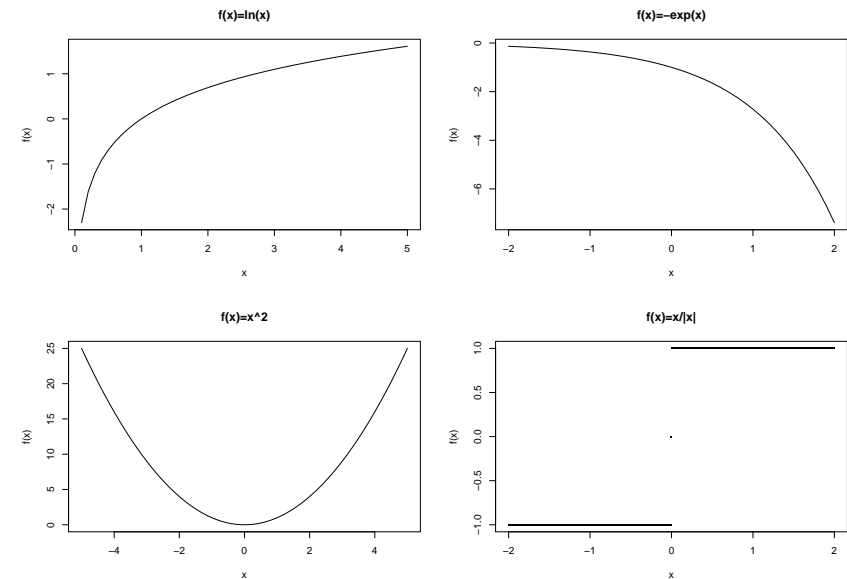
- Donc, le signe de la dérivée de f sur \mathbb{I} permet d'identifier le sens de variation de f sur cet intervalle.

Rappel du Sommaire

2 Fonctions d'une variable

- Rappels sur les applications, injections, surjections
- Fonctions numériques
- Limite et continuité d'une fonction numérique
- Sens de variation d'une fonction numérique
- Dérivabilité d'une fonction numérique
- Différentiabilité d'une fonction numérique
- Formules de Taylor (approximation d'ordre n d'une fonction)
- Convexité

Exemples



Différentiabilité

- La notion de différentiabilité est proche de celle de dérivabilité. Elles sont équivalentes dans le cas des fonctions numériques.
- La différentiabilité met néanmoins l'accent sur l'idée qu'une fonction peut être approximée localement par une fonction affine.
- Cependant, la différentiabilité présente l'avantage de se généraliser à des fonctions de plusieurs variables (contrairement à la dérivabilité).
- On reviendra donc également sur cette notion dans le cas des fonctions de plusieurs variables.
- Rappelons que la dérivée de f en x_0 est définie par :

$$f'(x_0) = \lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

- Rappelons également qu'une fonction affine ϕ s'écrit :

$$\phi(x) = a_0 + a_1 x$$

avec $a_0, a_1 \in \mathbb{R}$.

Différentiabilité (suite)

- On cherche donc à approximer f par une fonction affine ϕ au voisinage de x_0 .
- Pour cela, une première condition est que ces deux fonctions soient identiques en x_0 :

$$\phi(x_0) = f(x_0)$$

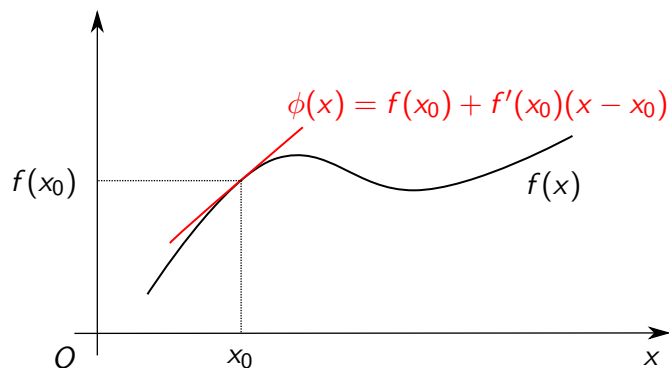
- Une deuxième condition est que $\phi(x)$ approche $f(x)$ plus rapidement que x approche x_0 . Ceci s'exprime par :

$$\lim_{x \rightarrow x_0} \frac{|f(x) - \phi(x)|}{|x - x_0|} = 0$$

Dit autrement, l'erreur d'approximation dans le voisinage de x_0 doit être "petite" en comparaison de la distance entre x et x_0 .

Différentiabilité (suite)

- Au voisinage de x_0 , $\phi(x) = f(x_0) + f'(x_0)(x - x_0)$ est l'équation de la droite tangente à f passant par x_0 .
- Illustration :



Différentiabilité (suite)

Définition. (Fonction différentiable)

Soit f une fonction et x_0 un point de \mathbb{D}_f . On dit que f est **différentiable** en x_0 s'il existe un réel a_1 tel que :

$$\lim_{x \rightarrow x_0} \frac{f(x) - (f(x_0) + a_1(x - x_0))}{x - x_0} = 0$$

- On voit que f est différentiable en x_0 ssi f est dérivable en x_0 . En effet, l'expression précédente peut s'écrire :

$$\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} = a_1$$

a_1 est donc par définition le nombre dérivé $f'(x_0)$.

Définition de la différentielle

Définition. (Différentielle)

Soit f une fonction différentiable en x_0 un point de \mathbb{D}_f . La fonction linéaire suivante est appelée **différentielle de f en x_0** :

$$\begin{aligned} df_{x_0} : \mathbb{R} &\rightarrow \mathbb{R} \\ h &\rightarrow f'(x_0)h \end{aligned}$$

- Reprenons l'exemple de $f(x) = x^2$. Au voisinage de x_0 :
 - Le nombre dérivé est :

$$f'(x_0) = 2x_0$$

- La différentielle est :

$$df_{x_0} \underbrace{(x - x_0)}_h = \underbrace{2x_0}_{f'(x_0)} \underbrace{(x - x_0)}_h$$

- L'équation de la tangente est :

$$\begin{aligned} \phi(x) &= f(x_0) + df_{x_0}(x - x_0) \\ &= x_0^2 + 2x_0(x - x_0) \end{aligned}$$

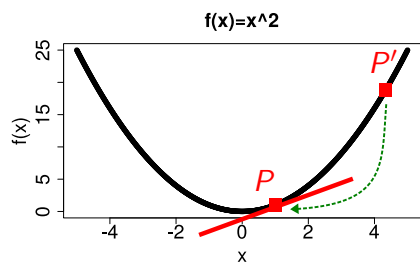
Accroissement absolu et accroissement différentiel

- Rappelons que l'accroissement absolu de f lorsque x passe de x_0 à $x_0 + h$ est noté : $\Delta f_{x_0}(h) = f(x_0 + h) - f(x_0)$.
- Nous venons d'introduire la différentielle de f en x_0 :

$$df_{x_0}(h) = f'(x_0)h$$

- En fait, pour h suffisamment petit nous avons $\Delta f_{x_0}(h) \approx df_{x_0}(h)$:

$$f(x_0 + h) - f(x_0) \approx f'(x_0)h$$



Rappel du Sommaire

2 Fonctions d'une variable

- Rappels sur les applications, injections, surjections
- Fonctions numériques
- Limite et continuité d'une fonction numérique
- Sens de variation d'une fonction numérique
- Dérivabilité d'une fonction numérique
- Différentiabilité d'une fonction numérique
- Formules de Taylor (approximation d'ordre n d'une fonction)
- Convexité

Exemple

- Soit $f(x) = x^3 - 3x^2 + 6x$ définie sur $\mathbb{D}_f = \mathbb{R}$.
- f est dérivable sur \mathbb{R} .
- $f'(x) = 3x^2 - 6x + 6$.
- $df_x(h) = f'(x)h = (3x^2 - 6x + 6)h$.
- $\Delta f_x(h) = f(x+h) - f(x) = (x+h)^3 - 3(x+h)^2 + 6(x+h) - (x^3 - 3x^2 + 6x) = 3x^2h + 3xh^2 + h^3 - 6xh - 3h^2 + 6h$.
- Prenons $x = 1$ on a :
 - ▶ $df_1(h) = 3h$.
 - ▶ $\Delta f_1(h) = h^3 + 3h$.
- L'écart $\Delta f_1(h) - df_1(h) = h^3$. Pour $h = 0.1$ on a donc un écart de 0.0001 entre accroissement absolu et accroissement différentiel
- La différentielle de f est bien une fonction linéaire (de h) permettant d'approximer les variations de f au voisinage d'un point.

Formule de Taylor-Young

- La notion d'approximation d'une fonction f se généralise au-delà du concept de différentielle par la formule de Taylor.
- Ce résultat est par ailleurs fondamental en optimisation.

Théorème.

Soit f une fonction dérivable sur un interval $\mathbb{I} \subset \mathbb{D}_f$ jusqu'à l'ordre n (ou de classe \mathbb{C}^n) alors pour $x_0, x_0 + h \in \mathbb{I}$ on a :

$$f(x_0 + h) = f(x_0) + \frac{h}{1!} f^{(1)}(x_0) + \frac{h^2}{2!} f^{(2)}(x_0) + \dots + \frac{h^n}{(n)!} f^{(n)}(x_0) + o(h^n)$$

où $f^{(n)}$ est la dérivée d'ordre n de f et $o(h^n)$ représente un terme négligeable qui converge vers 0 plus vite que h^n .

Formule de Taylor-Young (suite)

- La formule de Taylor donne une **approximation** de f (dérivable n fois) au voisinage de x_0 , par un **polynôme d'ordre n** dont les coefficients dépendent des valeurs des dérivées successives de f en x_0 .
- Quand $n = 1$, on obtient une **approximation linéaire** (affine) de f au voisinage de x_0 . On retrouve ainsi l'équation de la tangente $\phi(x)$ vues précédemment dans le cadre de la différentiabilité.
- Quand $n = 2$, on obtient une **approximation quadratique** de f au voisinage de x_0 qui est plus fine que l'approximation linéaire.
- Ainsi, plus n est grand, plus le degré du polynôme est grand et plus l'approximation de f au voisinage du point considéré est fine.
- Les approximations linéaires et quadratiques données par la formule de Taylor sont la base de nombreuses méthodes d'optimisation numérique.

Formule de Taylor-Maclaurin (suite)

- Cas usuels de la formule de Taylor-Maclaurin (jusqu'à l'ordre 2 ou 3) :

Fonction	Formule de Taylor-Maclaurin
$\sqrt{1+x}$	$1 + \frac{x}{2} - \frac{x^2}{8} + o(x^2)$
$\frac{1}{1+x}$	$1 - x + x^2 + o(x^2)$
$\exp x$	$1 + x + \frac{x^2}{2} + o(x^2)$
$\ln(1+x)$	$x - \frac{x^2}{2} + o(x^2)$
$\cos x$	$1 - \frac{x^2}{2} + o(x^2)$
$\sin x$	$x - \frac{x^3}{3!} + o(x^3)$

Formule de Taylor-Maclaurin

- En posant $x = x_0 + h$, la formule de Taylor-Young s'écrit également :

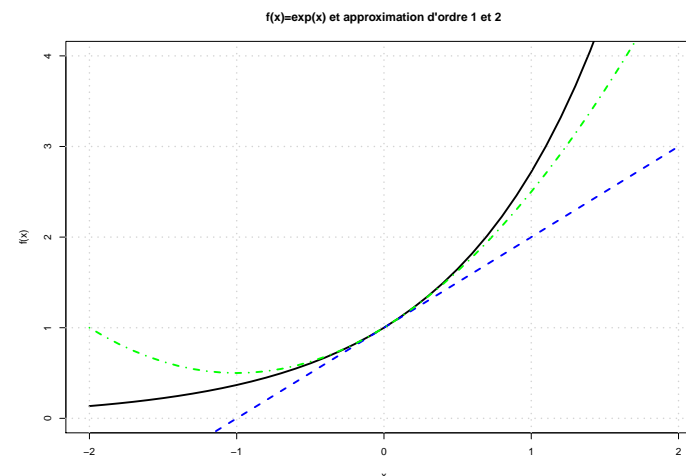
$$f(x) = f(x_0) + \frac{x - x_0}{1!} f^{(1)}(x_0) + \frac{(x - x_0)^2}{2!} f^{(2)}(x_0) + \dots \\ \dots + \frac{(x - x_0)^n}{(n)!} f^{(n)}(x_0) + o((x - x_0)^n)$$

- Le cas particulier $x_0 = 0$ est la **formule de Taylor-Maclaurin** :

$$f(x) = f(0) + \frac{x}{1!} f^{(1)}(0) + \frac{x^2}{2!} f^{(2)}(0) + \dots + \frac{x^n}{(n)!} f^{(n)}(0) + o(x^n)$$

Illustration

- Ci-dessous les approximations affine et quadratique de $\exp(x)$ au voisinage de 0 (en utilisant donc la formule de Taylor-Maclaurin).



Exemple

- Soit la fonction $f(x) = \frac{\ln(x)}{x^2}$.
- $\mathbb{D}_f = \mathbb{R}^{+*}$.
- $f'(x) = \frac{\frac{1}{x}x^2 - \ln(x)(2x)}{x^4} = \frac{1-2\ln(x)}{x^3}$.
- Approximation affine de f au voisinage de 1 :

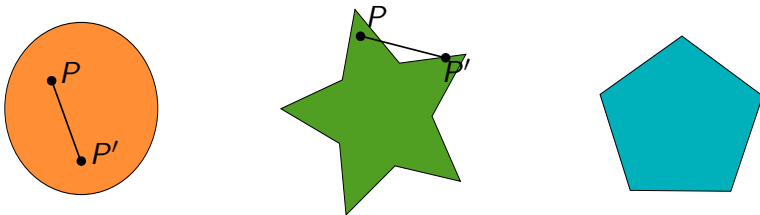
$$\begin{aligned} f(x) &= f(1) + \frac{x-1}{1!} f'(1) + o((x-1)) \\ &\approx x - 1 \end{aligned}$$

- $f''(x) = \frac{-2x^3 - (1-2\ln(x))(3x^2)}{x^6} = \frac{-5+6\ln(x)}{x^4}$.
- Approximation quadratique de f au voisinage de 1 :

$$\begin{aligned} f(x) &= f(1) + \frac{x-1}{1!} f'(1) + \frac{(x-1)^2}{2!} f''(1) + o((x-1)^2) \\ &\approx -1 + x - \frac{5}{2}(x-1)^2 \end{aligned}$$

Ensembles convexes

- Un sous-ensemble \mathbb{D} du plan \mathbb{R}^2 forme un **ensemble convexe** si pour deux points $P, P' \in \mathbb{D}$ leur segment appartient également à \mathbb{D} .
- Exemple et contre-exemple :



Rappel du Sommaire

2 Fonctions d'une variable

- Rappels sur les applications, injections, surjections
- Fonctions numériques
- Limite et continuité d'une fonction numérique
- Sens de variation d'une fonction numérique
- Dérivabilité d'une fonction numérique
- Différentiabilité d'une fonction numérique
- Formules de Taylor (approximation d'ordre n d'une fonction)
- Convexité

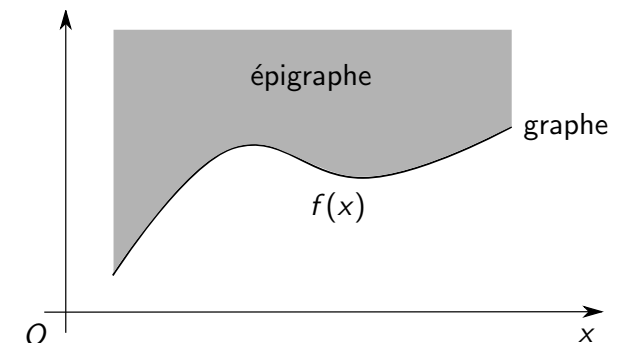
Epigraphe d'une fonction

Définition. (Epigraphe)

L'**épigraphe** de la fonction $f : \mathbb{D}_f \rightarrow \mathbb{R}$, dénoté $\text{epi}(f)$, est l'ensemble des points de $\mathbb{D}_f \times \mathbb{R} \subset \mathbb{R}^2$ donné par :

$$\{(x, y), x \in \mathbb{D}_f, y \in \mathbb{R}, y \geq f(x)\}$$

- Illustration :



Fonction convexe

Définition. (Fonction convexe)

Une fonction f définie sur \mathbb{D}_f est **convexe** si son épigraphe est un ensemble convexe.

- La définition précédente permet d'illustrer géométriquement la notion de convexité. Une définition alternative de la convexité d'une fonction numérique, est la suivante :

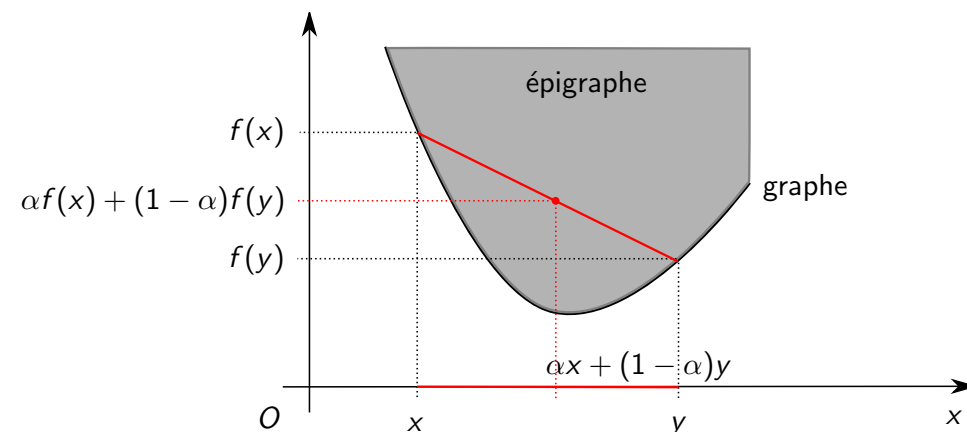
Définition. (Fonction (strictement) convexe)

Soit f une fonction définie sur $\mathbb{D}_f \subset \mathbb{R}$. On dit que f est **convexe** sur un intervalle $\mathbb{I} \subset \mathbb{D}_f$ si $\forall x, y \in \mathbb{I}$ et $\forall \alpha \in [0, 1]$ on a :

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

Si l'inégalité large précédente (en **rouge**) est remplacée par une inégalité stricte on dira que f est **strictement convexe**.

Illustration



- Le segment reliant $(x, f(x))$ et $(y, f(y))$ est au-dessus de f (ou sur le graphe de f si f est convexe au sens large).

Fonction concave

Définition. (Fonction (strictement) concave)

Soit f une fonction définie sur \mathbb{D}_f . On dit que f est **concave** sur un intervalle $\mathbb{I} \subset \mathbb{D}_f$ si $\forall x, y \in \mathbb{I}$ et $\forall \alpha \in [0, 1]$ on a :

$$f(\alpha x + (1 - \alpha)y) \geq \alpha f(x) + (1 - \alpha)f(y)$$

Si l'inégalité large précédente (en **rouge**) est remplacée par une inégalité stricte on dira que f est **strictement concave**.

- Donc f est (strictement) concave si $-f$ est (strictement) convexe.

Exemple

- Soit $f = x^2$ définie sur \mathbb{R} .
- Soit $x, y \in \mathbb{R}, x \neq y$:

$$\begin{aligned} f(\alpha x + (1 - \alpha)y) &= (\alpha x + (1 - \alpha)y)^2 \\ &= \alpha^2 x^2 + (1 - \alpha)^2 y^2 + \alpha(1 - \alpha)2xy \\ &< \alpha^2 x^2 + (1 - \alpha)^2 y^2 + \alpha(1 - \alpha)(x^2 + y^2) \\ &< \underbrace{\alpha x^2 + (1 - \alpha)y^2}_{\alpha f(x) + (1 - \alpha)f(y)} \end{aligned}$$

- Donc $f(x) = x^2$ est strictement convexe.

Approximation affine et convexité

- Ci-dessous une définition alternative de la convexité faisant le lien avec les approximation affines vues précédemment.

Définition.

Soit f une fonction de classe \mathbb{C}^1 définie sur \mathbb{D}_f et \mathbb{I} un intervalle de \mathbb{D}_f alors :

- f est convexe sur \mathbb{I} si son épigraphe sur cet intervalle est au-dessus de toutes ses tangentes :

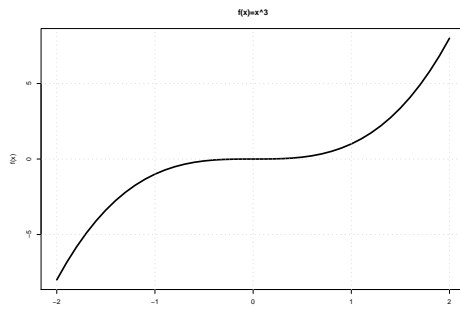
$$\forall x \in \mathbb{I}, \forall y \in \mathbb{I}, f(x) \geq f(y) + (x - y)f'(y)$$

- f est concave sur \mathbb{I} si son épigraphe sur cet intervalle est au-dessous de toutes ses tangentes :

$$\forall x \in \mathbb{I}, \forall y \in \mathbb{I}, f(x) \leq f(y) + (x - y)f'(y)$$

Exemple

- Soit $f(x) = x^3$ définie sur \mathbb{R} .
- $f'(x) = 3x^2$ et $f''(x) = 6x$.
- Deux intervalles :
 - ▶ Sur $\mathbb{I} =]-\infty, 0]$ on a $f''(x) \leq 0$ donc f est concave.
 - ▶ Sur $\mathbb{I} = [0, +\infty[$ on a $f''(x) \geq 0$ donc f est convexe.



Dérivées secondes et convexité

- La dérivée seconde de f (si elle existe) est un outil pratique pour étudier la convexité locale d'une fonction.

Théorème.

Soit f une fonction de classe \mathbb{C}^2 définie sur \mathbb{D}_f et \mathbb{I} un intervalle de \mathbb{D}_f alors :

- f est convexe sur \mathbb{I} si $\forall x \in \mathbb{I}, f''(x) \geq 0$.
- f est concave sur \mathbb{I} si $\forall x \in \mathbb{I}, f''(x) \leq 0$.

Si les inégalités larges précédentes (en rouge) sont remplacées par des inégalités strictes on dira que f est **strictement convexe** et **strictement concave** sur \mathbb{I} .

Rappel du Sommaire

- 1 Quelques méthodes statistiques du point de vue de l'optimisation
- 2 Fonctions d'une variable
- 3 Optimisation unidimensionnelle
- 4 Algèbre linéaire
- 5 Fonctions de plusieurs variables
- 6 Optimisation multidimensionnelle

Rappel du Sommaire

3 Optimisation unidimensionnelle

- Définitions de base
- Existence d'extrema pour une fonction continue
- Caractérisation des extrema d'une fonction continue
- Algorithmes d'optimisation numérique unidimensionnelle

Extremum local

Définition.

Soit f une fonction définie sur \mathbb{D}_f et soit \mathbb{S} un sous-ensemble de \mathbb{D}_f ($\mathbb{S} \subset \mathbb{D}_f$). On dit que :

- f possède un **minimum local** sur \mathbb{S} en $x^* \in \mathbb{S}$ si il existe un voisinage \mathbb{V}_{x^*} de x^* tel que :

$$\forall x \in \mathbb{V}_{x^*} \cap \mathbb{S}, f(x^*) \leq f(x)$$

- f possède un **maximum local** sur \mathbb{S} en $x^* \in \mathbb{S}$ si il existe un voisinage \mathbb{V}_{x^*} de x^* tel que :

$$\forall x \in \mathbb{V}_{x^*} \cap \mathbb{S}, f(x^*) \geq f(x)$$

Si les inégalités larges précédentes (en rouge) sont remplacées par des inégalités strictes on dira que f possède un **minimum local strict** et un **maximum local strict** en $x^* \in \mathbb{S}$ sur \mathbb{S} .

Extremum global

Définition.

Soit f une fonction définie sur \mathbb{D}_f et soit \mathbb{S} un sous-ensemble de \mathbb{D}_f ($\mathbb{S} \subset \mathbb{D}_f$). On dit que :

- f possède un **minimum global** sur \mathbb{S} en $x^* \in \mathbb{S}$ si :

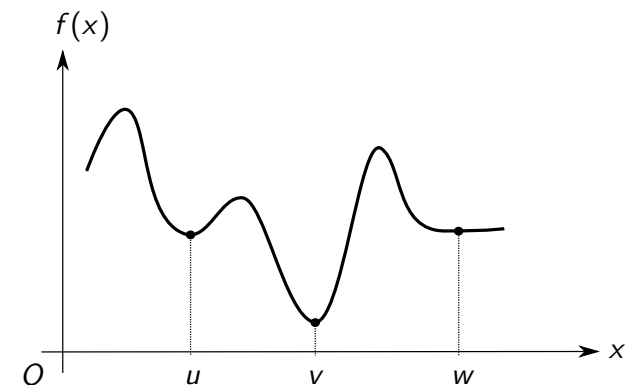
$$\forall x \in \mathbb{S}, f(x^*) \leq f(x)$$

- f possède un **maximum global** sur \mathbb{S} en $x^* \in \mathbb{S}$ si :

$$\forall x \in \mathbb{S}, f(x^*) \geq f(x)$$

Si les inégalités larges précédentes (en rouge) sont remplacées par des inégalités strictes on dira que f possède un **minimum global strict** et un **maximum global strict** en $x^* \in \mathbb{S}$ sur \mathbb{S} .

Illustration



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

Rappel du Sommaire

3 Optimisation unidimensionnelle

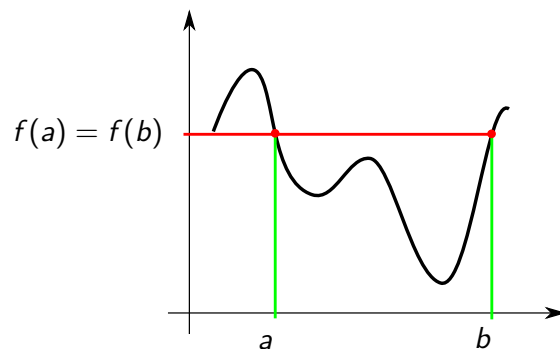
- Définitions de base
- Existence d'extrema pour une fonction continue
- Caractérisation des extrema d'une fonction continue
- Algorithmes d'optimisation numérique unidimensionnelle

Théorème de Rolle

Théorème. (Théorème de Rolle)

Soit f une fonction continue sur $[a, b] \subset \mathbb{I}$, dérivable sur $]a, b[$ et telle que $f(a) = f(b)$. Alors il existe au moins un point $c \in]a, b[$ tel que :

$$f'(c) = 0$$



Théorème des valeurs intermédiaires

Théorème.

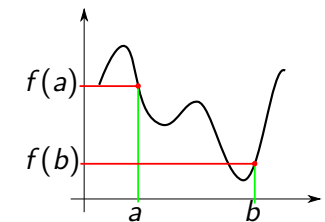
Soit f une fonction continue sur un intervalle $\mathbb{I} \subset \mathbb{D}_f$ et a et b deux réels de \mathbb{I} . Toutes les valeurs comprises entre $f(a)$ et $f(b)$ sont atteintes par f en au moins un point de $]a, b[$.

Ainsi en supposant que $f(a) > f(b)$ alors on a :

$$\forall y \in]f(b), f(a)[, \exists x \in]a, b[, y = f(x)$$

Corollaire.

Si f est continue sur un intervalle \mathbb{I} alors $f(\mathbb{I})$ est un intervalle.



Théorème des bornes ou de Weierstrass

Théorème. (Théorème de Weierstrass)

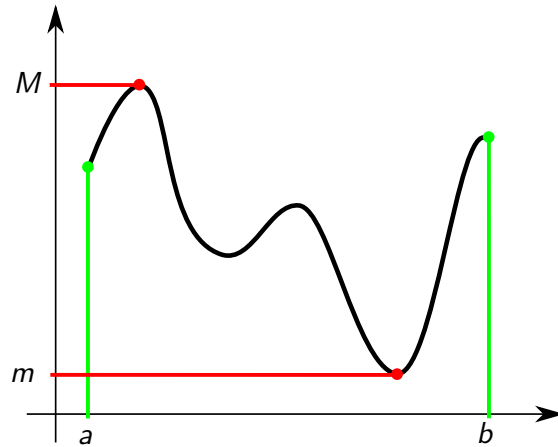
Si f est une fonction continue sur un intervalle fermé $[a, b]$, alors l'image $f([a, b])$ est un intervalle fermé que l'on notera $[m, M]$. On en déduit que f est bornée sur $[a, b]$ et qu'elle atteint ses bornes c-à-d :

$$\exists (c, d) \in [a, b]^2, \forall x \in [a, b], \underbrace{f(c)}_m \leq f(x) \leq \underbrace{f(d)}_M$$

- C'est le premier théorème d'optimisation puisqu'en d'autres termes, il permet de montrer qu'une fonction continue sur un intervalle fermé possède un minimum (m) et un maximum (M) et que ces bornes sont atteintes (respectivement en c et d).

Théorème des bornes ou de Weierstrass (suite)

- Illustration :



Condition nécessaire du premier ordre (CNPO)

Théorème. (CNPO)

Soit f une fonction dérivable sur un intervalle ouvert \mathbb{I} et soit $x^* \in \mathbb{I}$. Si f possède un extremum local ou global sur \mathbb{I} en x^* alors :

$$f'(x^*) = 0$$

Définition.

On appelle **point critique (ou stationnaire)** de f , tout point $x^* \in \mathbb{D}_f$ tel que f est dérivable en x^* et $f'(x^*) = 0$.

Rappel du Sommaire

3 Optimisation unidimensionnelle

- Définitions de base
- Existence d'extrema pour une fonction continue
- Caractérisation des extrema d'une fonction continue
- Algorithmes d'optimisation numérique unidimensionnelle

Condition nécessaire du premier ordre (CNPO) (suite)

- Ainsi si x^* est tel que $f'(x^*) \neq 0$ alors x^* ne peut pas être un extremum. La CNPO nous permet donc de trouver des points pouvant être des extrema. Mais si x^* est tel que $f'(x^*) = 0$ ce n'est pas forcément un extremum (exemple de $f(x) = x^3$ en $x^* = 0$). La condition est nécessaire mais pas suffisante !
- Autrement dit un extremum est forcément un point critique mais tout point critique n'est pas forcément un extremum.

Conditions du second ordre

- Pour reconnaître si un point critique est un extremum, on peut étudier le signe de l'accroissement absolu $\Delta f_{x^*}(h) = f(x^* + h) - f(x^*)$ pour h dans un voisinage de 0 (càd pour $h \in [-\epsilon, \epsilon]$) :
 - ▶ Si $\Delta f_{x^*}(h) > 0$ alors x^* est un minimum,
 - ▶ Si $\Delta f_{x^*}(h) < 0$ alors x^* est un maximum.
- L'étude du signe de $\Delta f_{x^*}(h)$ est ardue pour la plupart des fonctions.
- Une autre façon de faire est d'étudier la **convexité**. Supposons que f est convexe sur \mathbb{I} donc $f(x) \geq f(x^*) + (x - x^*)f'(x^*)$ (cf slide 73). x^* étant un point critique on a donc $f'(x^*) = 0$ et par conséquent :

$$f(x) \geq f(x^*)$$

- En pratique pour savoir si un point critique est un extremum on utilisera les propriétés qui suivent basées sur les dérivées secondes.

Condition suffisante du second ordre (CSSO)

Théorème.

Soit f définie sur \mathbb{D}_f contenant un intervalle ouvert \mathbb{I} tel que f soit de classe \mathbb{C}^1 sur \mathbb{I} . Soit $x^* \in \mathbb{I}$ un point critique de f (càd $f'(x^*) = 0$) :

- Si f est **convexe** sur \mathbb{I} alors f admet un **minimum global** sur \mathbb{I} en x^* .
- Si f est **concave** sur \mathbb{I} alors f admet un **maximum global** sur \mathbb{I} en x^* .
- A l'aide du tableau de variation de f' , on peut déterminer si un point critique x^* de f est un maximum ou un minimum :
 - ▶ Si au voisinage de x^* , f' change de signe et passe du - au + alors il s'agit d'un minimum.
 - ▶ Si au voisinage de x^* , f' change de signe et passe du + au - alors il s'agit d'un maximum.
- On peut aussi déterminer le signe de $f''(x^*)$ pour conclure sur la convexité/concavité de f au voisinage de x^* (à condition que f est de classe \mathbb{C}^2).

Conditions suffisantes (cas général)

Théorème.

Soit f une fonction définie sur \mathbb{D}_f contenant un intervalle ouvert \mathbb{I} tel que f soit de classe \mathbb{C}^∞ . Soit x^* un point critique de f et soit $p \in \mathbb{N}$ tel que $f^{(p)}(x^*)$ est la première dérivée non nulle au point x^* .

- Si p est impair, alors f n'admet pas d'extremum en x^* mais un point d'inflexion.
- Si p est pair et $f^{(p)}(x^*) < 0$, alors f admet un maximum local en x^* .
- Si p est pair et $f^{(p)}(x^*) > 0$, alors f admet un minimum local en x^* .

- En général ce théorème est appliqué avec $p = 2$ ou $p = 3$.

Démonstration

Démonstration.

Par la formule de Taylor-Young, on a :

$$f(x^* + h) = f(x^*) + \sum_{i=1}^n \frac{h^i}{i!} f^{(i)}(x^*) + o(h^n)$$

Comme p est la première dérivée non nulle en x^* on a :

$$f(x^* + h) = f(x^*) + \frac{h^p}{p!} f^{(p)}(x^*) + o(h^p)$$

On voit alors que :

- Si p est pair alors $\Delta f_{x^*}(h)$ est du signe de $f^{(p)}(x^*)$ et on a un extremum.
- Si p est impair alors $\Delta f_{x^*}(h)$ change de signe comme h^p et on n'a pas d'extremum.

Exemple

- Soit $f(x) = x^3 - 3x + 2$. On pose $\mathbb{I} = \mathbb{D}_f = \mathbb{R}$.
- $f'(x) = 3x^2 - 3$.
- L'ensemble des points critiques est $\{-1, 1\}$.
- $f''(x) = 6x$.
- En $x^* = -1$, $f''(-1) = -6$. La CSSO implique que f a un maximum local en -1 et il vaut $f(-1) = 4$.
- En $x^* = 1$, $f''(1) = 6$. La CSSO implique que f a un minimum local en 1 et il vaut $f(1) = 0$.
- Extrema globaux? Non car f'' change de signe sur \mathbb{R} et que les limites de f en $\pm\infty$ sont $\pm\infty$.
Contre-exemples : $f(3) = 20 > 4$ et $f(-3) = -16 < 0$.

Rappel du Sommaire

- 3 Optimisation unidimensionnelle
 - Définitions de base
 - Existence d'extrema pour une fonction continue
 - Caractérisation des extrema d'une fonction continue
 - Algorithmes d'optimisation numérique unidimensionnelle

Exemple (suite)

- Soit $f(x) = 2x - x \ln(x)$. On pose $\mathbb{I} = \mathbb{D}_f = \mathbb{R}^{+*}$.
- $f'(x) = 2 - \ln(x) - 1$.
- L'ensemble des points critiques est $\{e\}$.
- $f''(x) = -1/x$.
- $\forall x \in \mathbb{I}, f''(x) < 0$ donc f est concave sur \mathbb{I} .
- En $x^* = e$, $f''(e) < 0$. La CSSO implique que f a un maximum local en e et il vaut $f(e) = e$.
- Extremum global? Oui! car f est concave sur tout \mathbb{D}_f .

Optimisation numérique

- Dans les exemples précédents, les points critiques pouvaient être déterminés de façon **explicite** par la résolution d'une équation.
- Dans le cas général, on ne peut pas toujours déterminer ces points car les équations peuvent être plus ardues à résoudre.
- Par exemple, il est difficile en général de résoudre des polynômes de degré 5 (pour les polynômes de degré 3 et 4 des méthodes existent).
- On a alors recours à des **algorithmes de recherche** et on parle d'**optimisation numérique**.
- La procédure générale est la suivante : on commence par un point x_0 et on modifie ce point de façon itérative en s'assurant qu'au bout d'un certain nombre d'itérations, la recherche converge vers un point critique.
- En bref, on cherche à définir une suite $\{x_k\}$ de sorte à ce que $\lim_{k \rightarrow \infty} f'(x_k) = 0$.

Optimisation numérique (suite)

- Nous étudions des méthodes permettant de déterminer un minimiseur de f sur un intervalle $\mathbb{I} = [a_0, b_0]$.
- Nous supposons que f est **unimodale** sur $[a_0, b_0]$ ce qui signifie que f a un unique minimiseur local sur cet intervalle.
- Il existe plusieurs types d'approche :
 - ▶ Méthodes de type "encadrement de minimiseurs" (section d'or, Fibonacci, ...)
 - ▶ Méthodes utilisant la dérivée première (sécante, ...)
 - ▶ Méthodes utilisant la dérivée première et seconde (Newton, ...)
- Nous voyons essentiellement la méthode de Newton dans la suite.

Méthode de Newton

- Dans la méthode de Newton, on suppose que f est de classe \mathbb{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 :

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

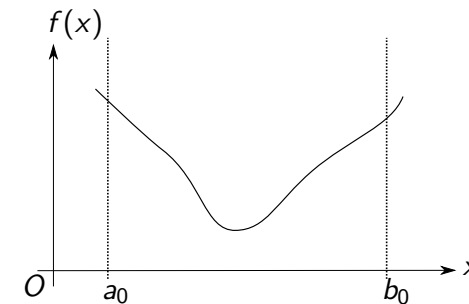
Fonction unimodale

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 :



Algorithme de la méthode de Newton

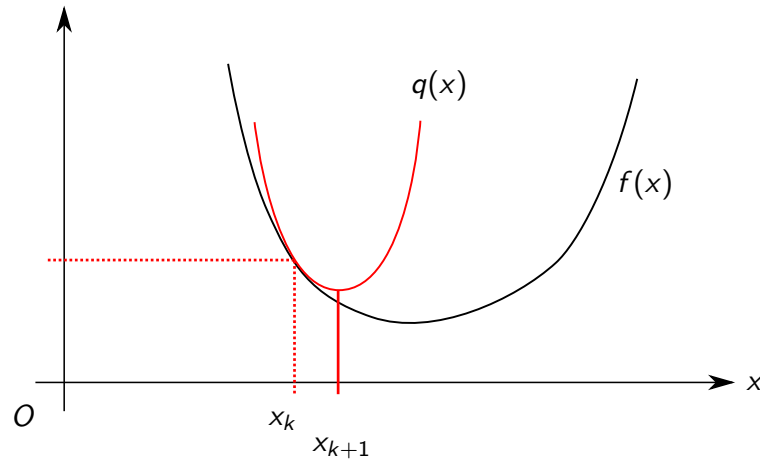
Input : $f, [a_0, b_0], \epsilon$

- 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

Cas où l'algorithme converge.



Exemple

- Utilisons la méthode 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

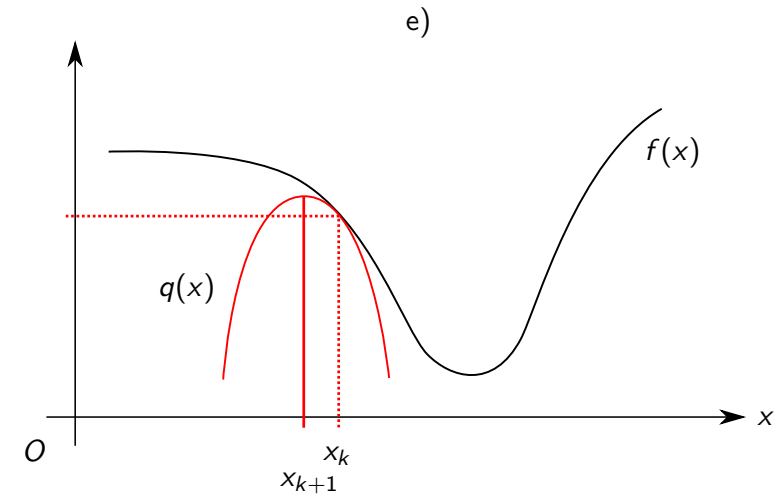
$$2 \quad |x - y| > 10^{-5}$$

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

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

Illustration (suite)

Cas où l'algorithme peut ne pas converger.



Exemple (suite)

Itération 2

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

$$3 \quad y \leftarrow x = 0.7552$$

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

Itération 3

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

$$3 \quad y \leftarrow x = 0.7391$$

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

Itération 4

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

$$3 \quad y \leftarrow x = 0.7390$$

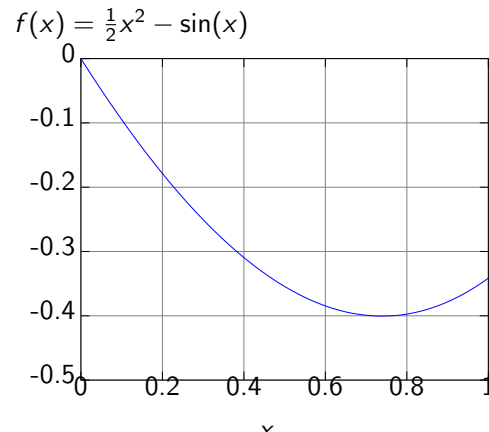
$$4 \quad x \leftarrow x - \frac{f'(x)}{f''(x)} = 0.7390 - \frac{0.7390 - \cos(0.7390)}{1 + \sin(0.7390)} = 0.7390$$

Exemple (suite)

Fin de l'algorithme

- 2 $|x - y| < 10^{-5}$
- 6 $x = 0.7390$

Illustration :



Rappel du Sommaire

- 1 Quelques méthodes statistiques du point de vue de l'optimisation
- 2 Fonctions d'une variable
- 3 Optimisation unidimensionnelle
- 4 Algèbre linéaire
- 5 Fonctions de plusieurs variables
- 6 Optimisation multidimensionnelle

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

Introduction

- En algèbre linéaire on manipule des objets mathématiques \mathbf{x} et \mathbf{y} appartenant à un ensemble \mathbb{E} tel que :
 - ▶ l'addition $\mathbf{x} + \mathbf{y}$ est un objet de \mathbb{E} .
 - ▶ la multiplication, $\lambda \mathbf{x}$ où λ est un réel (ou complexe), est également un objet de \mathbb{E} .
- De nombreux domaines scientifiques (différentes branches des mathématiques, physiques, chimie, économie, ...) présentent des problèmes dont le formalisme mathématique met en oeuvre ce type de structure algébrique.
- Ce cadre mathématique commun est la notion d'**espace vectoriel (ev)**.

Introduction (suite)

- Dans notre cas, nous abordons par la suite les fonctions f de **plusieurs variables réelles** appartenant à \mathbb{R}^n et à valeurs dans \mathbb{R} (ou dans \mathbb{R}^p parfois).
- Le domaine de définition \mathbb{D}_f est désormais de dimensions $n > 1$.
- Les arguments de f sont des vecteurs de l'ev $\mathbb{E} = \mathbb{R}^n$ et on a donc :

$$\begin{cases} f : \mathbb{R}^n \rightarrow \mathbb{R} \\ \mathbf{x} \rightarrow y \end{cases}$$

- Chercher à minimiser f revient donc à déterminer un minimiseur qui est un vecteur \mathbf{x}^* de \mathbb{R}^n .
- Il est donc essentiel de revoir au préalable les fondamentaux des ev afin de savoir/comprendre comment déterminer le minimiseur \mathbf{x}^* .
- La suite de la section comprend donc des rappels sur les ev.

Notations (suite)

- Une matrice peut être vue comme une collection de vecteurs colonnes. Elle sera notée en gras mais en lettre majuscule. On a \mathbf{A} de taille $(n \times p)$ qui s'écrit :

$$\begin{aligned} \mathbf{A} &= (\mathbf{a}^1 \quad \dots \quad \mathbf{a}^p) \\ &= \begin{pmatrix} a_{11} & \dots & a_{1p} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{np} \end{pmatrix} \end{aligned}$$

où pour tout $k = 1, \dots, p$:

$$\mathbf{a}^k = \begin{pmatrix} a_{1k} \\ \vdots \\ a_{nk} \end{pmatrix}$$

et où $a_{11}, \dots, a_{np} \in \mathbb{K}$.

Notations

- Un scalaire est noté en lettre minuscule : $x \in \mathbb{K}$ où \mathbb{K} un corps commutatif tel que \mathbb{C} ou \mathbb{R} (on prendra \mathbb{R}).
- Un vecteur peut être vue comme une collection de réels. Il sera noté en lettre minuscule mais en gras. Par défaut c'est un vecteur colonne. On a donc \mathbf{x} de taille $(n \times 1)$ qui s'écrit :

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

où $x_1, \dots, x_n \in \mathbb{K}$.

- Mais on notera également le vecteur \mathbf{x} par (x_1, \dots, x_n) .
- Un vecteur ligne sera noté comme suit (sans les ,) :

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

où \mathbf{x}^\top est la transposée de \mathbf{x} .

Rappel du Sommaire

4 Algèbre linéaire

- Définitions
 - Bases d'un espace vectoriel
 - Les théorèmes fondamentaux sur la dimension et sommes directes
 - Applications linéaires et matrices
 - Déterminants
 - Formes bilinéaires et formes quadratiques
 - Espaces euclidiens
 - Éléments de topologie dans \mathbb{R}^n

Définition d'un espace vectoriel

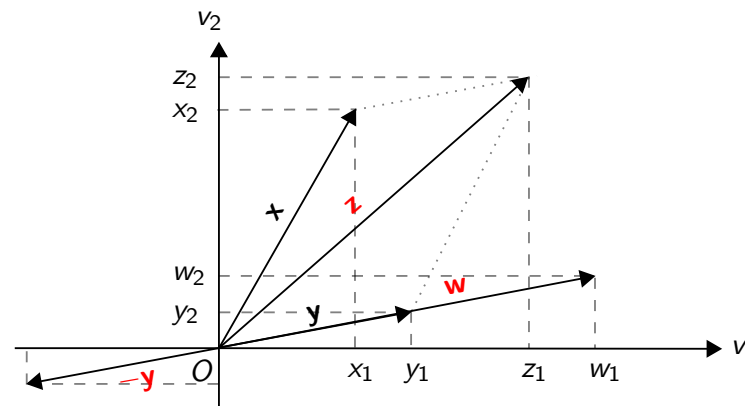
Définition.

Soit \mathbb{K} un corps commutatif (on prendra \mathbb{R}). On appelle **espace vectoriel (ev)** sur \mathbb{K} un ensemble \mathbb{E} sur lequel on a deux lois de composition :

- 1 Une loi interne dite **addition** $+$: $\mathbb{E} \times \mathbb{E} \rightarrow \mathbb{E}$ et vérifiant :
 - ▶ $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{E} : (\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$ (associativité).
 - ▶ $\forall \mathbf{x}, \mathbf{y} \in \mathbb{E} : \mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ (commutativité).
 - ▶ il existe un élément de \mathbb{E} , $\mathbf{0}_{\mathbb{E}}$ ou plus simplement $\mathbf{0}$, dit **neutre** tel que : $\forall \mathbf{x} \in \mathbb{E} : \mathbf{x} + \mathbf{0} = \mathbf{x}$,
 - ▶ pour tout $\mathbf{x} \in \mathbb{E}$, il existe un élément de \mathbb{E} et noté $-\mathbf{x}$, dit **opposé** de \mathbf{x} tel que : $\forall \mathbf{x} \in \mathbb{E} : \mathbf{x} + (-\mathbf{x}) = \mathbf{0}$.
- 2 Une loi externe de domaine \mathbb{K} c-à-d une application $\mathbb{K} \times \mathbb{E} \rightarrow \mathbb{E}$, dite **multiplication**, qui vérifie :
 - ▶ $\forall \lambda, \mu \in \mathbb{K}; \forall \mathbf{x} \in \mathbb{E} : \lambda(\mu\mathbf{x}) = (\lambda\mu)\mathbf{x}$ (associativité).
 - ▶ $\forall \lambda, \mu \in \mathbb{K}; \forall \mathbf{x} \in \mathbb{E} : (\lambda + \mu)\mathbf{x} = \lambda\mathbf{x} + \mu\mathbf{x}$ (distributivité).
 - ▶ $\forall \lambda \in \mathbb{K}; \forall \mathbf{x}, \mathbf{y} \in \mathbb{E} : \lambda(\mathbf{x} + \mathbf{y}) = \lambda\mathbf{x} + \lambda\mathbf{y}$ (distributivité).
 - ▶ $\forall \mathbf{x} \in \mathbb{E} : 1\mathbf{x} = \mathbf{x}$ (1 étant l'élément neutre dans \mathbb{K}).

Illustration

- $\mathbf{x} = (x_1, x_2)$ et $\mathbf{y} = (y_1, y_2)$ deux vecteurs de \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 $\mathbf{y} + (-\mathbf{y}) = \mathbf{0}$.



Autres exemples

- $\mathbb{E} = \mathbb{R}^n$ muni des deux lois suivantes :
 - ▶ $(x_1, \dots, x_n) + (y_1, \dots, y_n) = (x_1 + y_1, \dots, x_n + y_n)$.
 - ▶ $\lambda(x_1, \dots, x_n) = (\lambda x_1, \dots, \lambda x_n)$.
 Ici, $\mathbf{0} = (0, \dots, 0)$ et l'opposé de (x_1, \dots, x_n) est $(-x_1, \dots, -x_n)$.
- $\mathbb{E} = \mathbb{R}^n$ est l'ev que l'on utilisera en optimisation mais le concept d'ev peut être appliqué à d'autres ensembles.

Sous-espaces vectoriels

Définition.

Soit \mathbb{E} un ev et \mathbb{F} une partie non vide de \mathbb{E} . On dit que \mathbb{F} est un **sous-espace vectoriel (sev)** de \mathbb{E} , si la restriction des lois de \mathbb{E} à \mathbb{F} fait de \mathbb{F} un ev.

- En pratique, pour vérifier que \mathbb{F} est un sev, il suffira de montrer les deux conditions données ci-dessous.

Propriété.

Soit \mathbb{E} un ev et $\mathbb{F} \subset \mathbb{E}$. \mathbb{F} est un **sev** de \mathbb{E} ssi :

- $\mathbb{F} \neq \emptyset$
- $\mathbf{x}, \mathbf{y} \in \mathbb{F}; \lambda, \mu \in \mathbb{K} \Rightarrow \lambda\mathbf{x} + \mu\mathbf{y} \in \mathbb{F}$.

Exemples fondamentaux de sous-espaces vectoriels

- Droite vectorielle : Soit $\mathbf{v}_1 \in \mathbb{E}$, $\mathbf{v}_1 \neq \mathbf{0}$ alors :
 $\mathbb{F} = \{\mathbf{x} \in \mathbb{E} \mid \exists \lambda \in \mathbb{K} : \mathbf{x} = \lambda \mathbf{v}_1\}$ 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 :
 $\mathbb{F} = \{\mathbf{x} \in \mathbb{E} \mid \exists \lambda_1, \lambda_2 \in \mathbb{K} : \mathbf{x} = \lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2\}$ 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
 $\mathbb{F} = \{\mathbf{x} \in \mathbb{E} \mid \exists \lambda_1, \dots, \lambda_p \in \mathbb{K} : \mathbf{x} = \lambda_1 \mathbf{v}_1 + \dots + \lambda_p \mathbf{v}_p\}$ est un sev de \mathbb{E} dit **sous-espace engendré par $\mathbf{v}_1, \dots, \mathbf{v}_p$** ce qui sera noté $\text{Vec}\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$.

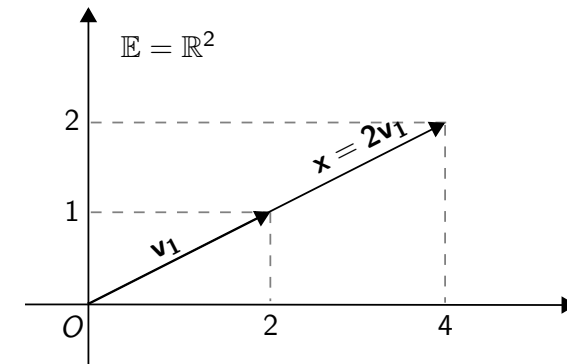
Rappel du Sommaire

4 Algèbre linéaire

- Définitions
- Bases d'un espace vectoriel
- Les théorèmes fondamentaux sur la dimension et sommes directes
- Applications linéaires et matrices
- Déterminants
- Formes bilinéaires et formes quadratiques
- Espaces euclidiens
- Éléments de topologie dans \mathbb{R}^n

Exemple et représentation géométrique

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



- Autre définition de $\mathbb{F} : \mathbb{F} = \{\mathbf{x} \in \mathbb{E} \mid x_1 - 2x_2 = 0\}$.
 - ▶ $\mathbf{0} \in \mathbb{F}$
 - ▶ $\mathbf{x}, \mathbf{y} \in \mathbb{F}; \lambda, \mu \in \mathbb{K}$ sont tels que $\lambda \mathbf{x} + \mu \mathbf{y} \in \mathbb{F}$.

Famille de vecteurs génératrice

Définition.

Une famille de vecteurs $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ d'un ev \mathbb{E} est dite **génératrice**, si $\mathbb{E} = \text{Vec}\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ ce qui veut dire :

$$\forall \mathbf{x} \in \mathbb{E}, \exists \lambda_1, \dots, \lambda_p \in \mathbb{K} \text{ tel que } : \mathbf{x} = \lambda_1 \mathbf{v}_1 + \dots + \lambda_p \mathbf{v}_p$$

- On dit aussi que tout \mathbf{x} de \mathbb{E} est une **combinaison linéaire** des vecteurs \mathbf{v}_i .

Définition.

Un ev est dit de **dimension finie**, s'il existe une famille génératrice finie. Dans le cas contraire on dit qu'il est de dimension infinie.

Exemple

- Dans \mathbb{R}^2 , la famille $\{\mathbf{e}_1 = (1, 0), \mathbf{e}_2 = (0, 1)\}$ est une famille génératrice car tout $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$ s'écrit :

$$(x_1, x_2) = x_1(1, 0) + x_2(0, 1) = x_1\mathbf{e}_1 + x_2\mathbf{e}_2$$

- Dans \mathbb{R}^2 , la famille $\{\mathbf{v}_1 = (1, 1), \mathbf{v}_2 = (1, -1)\}$ est génératrice. En effet, soit $\mathbf{x} = (a, b) \in \mathbb{R}^2$ avec a et b quelconques. On peut montrer qu'il existe $x_1, x_2 \in \mathbb{R}$ tels que $\mathbf{x} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2$:

$$\begin{aligned}(a, b) &= x_1(1, 1) + x_2(1, -1) \\ &= (x_1 + x_2, x_1 - x_2)\end{aligned}$$

On a donc $\begin{cases} a = x_1 + x_2 \\ b = x_1 - x_2 \end{cases}$ et on montre que tout $\mathbf{x} = (a, b)$ peut être combinaison linéaire de $\mathbf{v}_1, \mathbf{v}_2$. Il suffit pour cela de prendre :

$$\begin{cases} x_1 = \frac{a+b}{2} \\ x_2 = \frac{a-b}{2} \end{cases}$$

Famille de vecteurs libres et liés

Définition.

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**. Une famille qui n'est pas libre est dite **liée**.

- Exemple : dans \mathbb{R}^3 , les vecteurs $\mathbf{v}_1 = (1, 2, 1)$, $\mathbf{v}_2 = (-1, 3, 1)$ et $\mathbf{v}_3 = (-1, 13, 5)$ sont liés car :

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

Résultats sur les familles libres

Propriété.

Une famille $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ est **liée** ssi l'un au moins des vecteurs \mathbf{v}_i s'écrit comme une combinaison linéaire des autres vecteurs de la famille (càd si au moins l'un des vecteurs appartient à l'ev engendré par les autres).

Propriété.

Soit $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ une **famille de vecteurs libres** et \mathbf{x} un vecteur quelconque de $\text{Vec}\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ (càd \mathbf{x} est une combinaison linéaire des \mathbf{v}_i). Alors la **décomposition de \mathbf{x} sur les \mathbf{v}_i est unique**.

Démonstration.

Soient $\mathbf{x} = \lambda_1\mathbf{v}_1 + \dots + \lambda_p\mathbf{v}_p$ et $\mathbf{x} = \mu_1\mathbf{v}_1 + \dots + \mu_p\mathbf{v}_p$ deux décompositions de \mathbf{x} . En soustrayant, on obtient

$\mathbf{0} = (\lambda_1 - \mu_1)\mathbf{v}_1 + \dots + (\lambda_p - \mu_p)\mathbf{v}_p$. Les \mathbf{v}_i formant une famille libre on a : $\lambda_1 - \mu_1 = 0, \dots, \lambda_p - \mu_p = 0$ et donc $\lambda_1 = \mu_1, \dots, \lambda_p = \mu_p$. \square

Base d'un espace vectoriel

Définition.

$\mathbb{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ est une **base** si les vecteurs $\mathbf{v}_1, \dots, \mathbf{v}_p$ forment une famille **libre et génératrice**.

Propriété.

Une famille $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ est une **base** de l'ev \mathbb{E} , ssi tout $\mathbf{x} \in \mathbb{E}$ se décompose d'une façon unique sur les \mathbf{v}_i :

$$\forall \mathbf{x} \in \mathbb{E}, \exists!(x_1, \dots, x_p) \in \mathbb{K}^p : \mathbf{x} = x_1\mathbf{v}_1 + \dots + x_p\mathbf{v}_p$$

Définition.

Si $\mathbb{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ est une base de \mathbb{E} et si $\mathbf{x} \in \mathbb{E}$ est défini par $\mathbf{x} = x_1\mathbf{v}_1 + \dots + x_p\mathbf{v}_p$ alors les $x_i, i = 1, \dots, p$ sont appelées les **composantes de \mathbf{x} dans \mathbb{B}** ce que l'on notera par : $\mathbf{x} = (x_1, \dots, x_p)$.

Exemples de base

- Dans \mathbb{R}^2 , la famille $\{\mathbf{e}_1 = (1, 0), \mathbf{e}_2 = (0, 1)\}$ est une base (appelée **base canonique**). Elle est génératrice et libre également :

$$\begin{aligned}\lambda_1 \mathbf{e}_1 + \lambda_2 \mathbf{e}_2 = \mathbf{0} &\Leftrightarrow \lambda_1(1, 0) + \lambda_2(0, 1) = \mathbf{0} \\ &\Leftrightarrow (\lambda_1 + 0, 0 + \lambda_2) = (0, 0) \\ &\Leftrightarrow \lambda_1 = 0 \text{ et } \lambda_2 = 0\end{aligned}$$

- Dans le cas $\mathbb{E} = \mathbb{R}^n$, on prendra par défaut la **base canonique** $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ avec $\forall i = 1, \dots, n$:

$$\mathbf{e}_i = (0, \dots, 0, 1, 0, \dots, 0)$$

Rappel du Sommaire

4 Algèbre linéaire

- Définitions
- Bases d'un espace vectoriel
- Les théorèmes fondamentaux sur la dimension et sommes directes
- Applications linéaires et matrices
- Déterminants
- Formes bilinéaires et formes quadratiques
- Espaces euclidiens
- Éléments de topologie dans \mathbb{R}^n

Théorème sur l'existence d'une base

Théorème.

Soit $\mathbb{E} \neq \emptyset$ un ev de dimension finie alors :

- De toute famille génératrice on peut extraire une base.
- Toute famille libre peut être complétée de manière à former une base (théorème de la base incomplète).

- Exemple : Dans \mathbb{R}^3 , les vecteurs $\mathbf{v}_1 = (1, 3, 0)$ et $\mathbf{v}_2 = (2, 1, 0)$ forme une famille libre mais non génératrice.
- On peut alors ajouter le vecteur canonique $\mathbf{e}_3 = (0, 0, 1)$ ce qui forme une famille génératrice et libre.

Dimension d'un espace vectoriel

Théorème.

Dans un ev \mathbb{E} sur \mathbb{K} de dimension finie, toutes les bases ont le même nombre d'éléments. Ce nombre est appelé **dimension** de \mathbb{E} sur \mathbb{K} et est notée $\dim(\mathbb{E})$.

Corollaire.

- Dans un ev de dimension p , toute famille de vecteurs ayant plus de p éléments est liée.
- Dans un ev de dimension p , toute famille de vecteurs ayant moins de p éléments ne peut être génératrice.

- Exemple : Dans \mathbb{R}^3 , la famille $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, (a, b, c)\}$ est liée puisque $(a, b, c) = a\mathbf{e}_1 + b\mathbf{e}_2 + c\mathbf{e}_3$. Par ailleurs, dans \mathbb{R}^3 , la famille $\{(1, 1, 0), (0, 0, 1)\}$ n'est pas génératrice puisque par exemple on ne peut décomposer le vecteur $(1, 2, 3)$ en fonction de ces derniers.

Dimension d'un ev et d'un sev

Théorème.

Soit \mathbb{E} un ev de dimension p alors :

- Toute famille génératrice ayant p éléments est une base.
- Toute famille libre ayant p éléments est une base.

Théorème.

Soit \mathbb{E} un ev de dimension finie et \mathbb{F} un sev de \mathbb{E} . Alors \mathbb{F} est de dimension finie et :

- $\dim(\mathbb{F}) \leq \dim(\mathbb{E})$.
- $\dim(\mathbb{F}) = \dim(\mathbb{E}) \Leftrightarrow \mathbb{F} = \mathbb{E}$.

Rappel du Sommaire

4 Algèbre linéaire

- Définitions
- Bases d'un espace vectoriel
- Les théorèmes fondamentaux sur la dimension et sommes directes
- Applications linéaires et matrices
- Déterminants
- Formes bilinéaires et formes quadratiques
- Espaces euclidiens
- Éléments de topologie dans \mathbb{R}^n

Somme et Somme directe

Définition.

Soient $\mathbb{E}_1, \dots, \mathbb{E}_n$ des sev d'un même ev \mathbb{E} . On note :

$$\mathbb{E}_1 + \dots + \mathbb{E}_n = \{ \mathbf{x} \in \mathbb{E} \mid \exists \mathbf{x}_1 \in \mathbb{E}_1, \dots, \exists \mathbf{x}_n \in \mathbb{E}_n : \mathbf{x} = \mathbf{x}_1 + \dots + \mathbf{x}_n \}$$

$\mathbb{E}_1 + \dots + \mathbb{E}_n$ est un sev de \mathbb{E} dit **somme des sous-espaces** \mathbb{E}_i .

Définition.

Soient $\mathbb{E}_1, \dots, \mathbb{E}_n$ des sev d'un même ev \mathbb{E} . On note :

$$\mathbb{E}_1 \oplus \dots \oplus \mathbb{E}_n = \{ \mathbf{x} \in \mathbb{E} \mid \exists ! \mathbf{x}_1 \in \mathbb{E}_1, \dots, \exists ! \mathbf{x}_n \in \mathbb{E}_n : \mathbf{x} = \mathbf{x}_1 + \dots + \mathbf{x}_n \}$$

$\mathbb{E}_1 + \dots + \mathbb{E}_n$ est un sev de \mathbb{E} dit **somme directe des sous-espaces** \mathbb{E}_i .

Introduction

- On a rappelé les définitions générales des applications et leurs propriétés (cf slide 20).
- On s'intéresse ici à des applications possédant des propriétés particulières dites de **linéarité**.
- Une **application linéaire** entre deux ev \mathbb{E} et \mathbb{E}' permet de transformer les vecteurs de \mathbb{E} en des vecteurs de \mathbb{E}' .
- Toute application linéaire peut être représentée par une **matrice** et réciproquement toute matrice peut être associée à une application linéaire.
- On étudie donc deux représentations équivalentes des applications linéaires : l'une algébrique (équations linéaires) l'autre matricielle.
- Dans notre cas, nous utiliserons les applications linéaires pour approximer linéairement une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ localement en un vecteur $\mathbf{x} \in \mathbb{R}^n$.

Application linéaire

Définition. (Application linéaire)

Soient \mathbb{E} et \mathbb{E}' deux ev sur le même corps \mathbb{K} et f une **application** de \mathbb{E} dans \mathbb{E}' . On dit que f est **linéaire** si :

- 1 $\forall \mathbf{x}, \mathbf{y} \in \mathbb{E} : f(\mathbf{x} + \mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y})$.
- 2 $\forall \mathbf{x} \in \mathbb{E}, \forall \lambda \in \mathbb{K} : f(\lambda \mathbf{x}) = \lambda f(\mathbf{x})$.

L'ensemble des applications linéaires de \mathbb{E} dans \mathbb{E}' est noté $\mathcal{L}_{\mathbb{K}}(\mathbb{E}, \mathbb{E}')$ ou plus simplement $\mathcal{L}(\mathbb{E}, \mathbb{E}')$.

Définition. (Forme linéaire)

Une application linéaire de \mathbb{E} dans \mathbb{K} (le corps \mathbb{K} étant considéré tel un ev sur lui-même) est appelée plus particulièrement **forme linéaire**.

Remarque sur la linéarité

- Reprenons l'exemple d'application linéaire précédent :

$$\left\{ \begin{array}{l} f : \quad \mathbb{R}^3 \rightarrow \mathbb{R}^2 \\ (x_1, x_2, x_3) \rightarrow (2x_1 + x_2, x_2 - x_3) \end{array} \right.$$

- La **linéarité** de f tient au fait que les composantes x_i dans l'espace d'arrivée (ici \mathbb{R}^2), apparaissent toutes à la puissance 1. En d'autres termes, chaque composante dans l'espace d'arrivée est un **polynôme homogène de degré 1** en les x_i .
- Ainsi l'exemple suivant n'est pas une application linéaire :

$$\left\{ \begin{array}{l} f : \quad \mathbb{R}^3 \rightarrow \mathbb{R} \\ (x_1, x_2, x_3) \rightarrow (x_1 x_2 + x_3) \end{array} \right.$$

On remarquera par exemple que $f(\lambda \mathbf{x}) \neq \lambda f(\mathbf{x})$.

Exemple

$$\bullet \left\{ \begin{array}{l} f : \quad \mathbb{R}^3 \rightarrow \mathbb{R}^2 \\ (x_1, x_2, x_3) \rightarrow (2x_1 + x_2, x_2 - x_3) \end{array} \right. \text{ est linéaire :}$$

$$\begin{aligned} f(\mathbf{x} + \mathbf{y}) &= f((x_1, x_2, x_3) + (y_1, y_2, y_3)) \\ &= f(x_1 + y_1, x_2 + y_2, x_3 + y_3) \\ &= (2(x_1 + y_1) + (x_2 + y_2), (x_2 + y_2) - (x_3 + y_3)) \\ &= ((2x_1 + x_2) + (2y_1 + y_2), (x_2 - x_3) + (y_2 - y_3)) \\ &= f(x_1, x_2, x_3) + f(y_1, y_2, y_3) \\ f(\lambda \mathbf{x}) &= f(\lambda(x_1, x_2, x_3)) \\ &= f(\lambda x_1, \lambda x_2, \lambda x_3) \\ &= (2\lambda x_1 + \lambda x_2, \lambda x_2 - \lambda x_3) \\ &= \lambda(2x_1 + x_2, x_2 - x_3) \\ &= \lambda f(x_1, x_2, x_3) \end{aligned}$$

Endomorphisme, isomorphisme et automorphisme

Définition.

On appelle un **endomorphisme** de \mathbb{E} , une application linéaire de \mathbb{E} dans \mathbb{E} (même espace d'arrivée et de départ). L'ensemble des endomorphismes de \mathbb{E} est noté $\text{End}_{\mathbb{K}}(\mathbb{E})$ ou plus simplement $\text{End}(\mathbb{E})$.

Définition.

On appelle un **isomorphisme** de \mathbb{E} sur \mathbb{E}' , une application linéaire **bijective** de \mathbb{E} dans \mathbb{E}' .

Définition.

On appelle un **automorphisme**, un endomorphisme **bijectif**.

Exemples

- $\begin{cases} id_{\mathbb{E}} : \mathbb{E} \rightarrow \mathbb{E} \\ \mathbf{x} \rightarrow \mathbf{x} \end{cases}$ est un automorphisme.
- $\begin{cases} f : \mathbb{R}^3 \rightarrow \mathbb{R}^3 \\ (x_1, x_2, x_3) \rightarrow (\lambda_1 x_1, \lambda_2 x_2, \lambda_3 x_3) \end{cases}$ où $\lambda_1, \lambda_2, \lambda_3 > 0$ est un automorphisme.

Exemple

- Soit $\mathbb{E} = \mathbb{E}_1 \oplus \mathbb{E}_2$. Tout $\mathbf{x} \in \mathbb{E}$ se décompose donc de manière unique de la manière suivante : $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$ avec $\mathbf{x}_1 \in \mathbb{E}_1$ et $\mathbf{x}_2 \in \mathbb{E}_2$. Alors : $\begin{cases} p_1 : \mathbb{E} \rightarrow \mathbb{E}_1 \\ \mathbf{x}_1 + \mathbf{x}_2 \rightarrow \mathbf{x}_1 \end{cases}$ est une application linéaire dite **projecteur sur \mathbb{E}_1 parallèlement à \mathbb{E}_2** .
Dans ce cas, on a $Im(p_1) = \mathbb{E}_1$ et $Ker(p_1) = \mathbb{E}_2$.

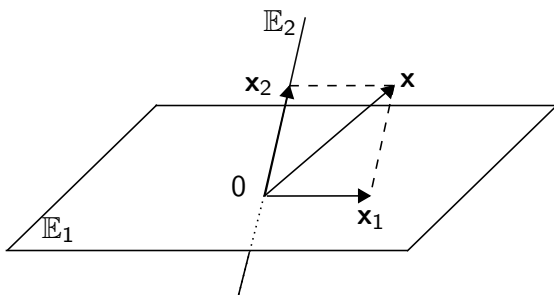


Image et noyau d'une application linéaire

Propriété.

Soit $f : \mathbb{E} \rightarrow \mathbb{E}'$ une application linéaire et \mathbb{F} un sev de \mathbb{E} . Alors $f(\mathbb{F})$ est un sev de \mathbb{E}' .

Définition.

En particulier, $f(\mathbb{E})$ est un sev de \mathbb{E}' appelé **image** de f et noté $Im(f)$. Sa dimension est appelée **rang** de f et est noté $rg(f)$ et on a :

$$rg(f) = dim(Im(f))$$

Définition.

Soit $f \in \mathcal{L}(\mathbb{E}, \mathbb{E}')$. Le **noyau** de f , noté $Ker(f)$, est un sev de \mathbb{E} défini comme suit :

$$Ker(f) = \{\mathbf{x} \in \mathbb{E} : f(\mathbf{x}) = \mathbf{0}\}$$

Propriété du noyau d'une application linéaire

Propriété.

Soit $f \in \mathcal{L}(\mathbb{E}, \mathbb{E}')$. Alors f est **injective** ssi $Ker(f) = \{\mathbf{0}\}$.

Démonstration.

- ⇐ Supposons $Ker(f) = \{\mathbf{0}\}$. Prenons $\mathbf{x}, \mathbf{y} \in \mathbb{E}$ tels que $f(\mathbf{x}) = f(\mathbf{y})$. On a $f(\mathbf{x}) - f(\mathbf{y}) = \mathbf{0}$ d'où $f(\mathbf{x} - \mathbf{y}) = \mathbf{0}$. Ainsi, $\mathbf{x} - \mathbf{y} \in Ker(f)$ et comme $Ker(f) = \{\mathbf{0}\}$, on a $\mathbf{x} = \mathbf{y}$ ce qui montre que f est injective.
- ⇒ Supposons que f est injective. Prenons $\mathbf{x} \in Ker(f)$, on a donc $f(\mathbf{x}) = \mathbf{0}$. On a par ailleurs $f(\mathbf{0}) = \mathbf{0}$ pour toute application linéaire. Donc $f(\mathbf{x}) = f(\mathbf{0})$. Comme f est injective, on a donc $\mathbf{x} = \mathbf{0}$ et donc $Ker(f) = \{\mathbf{0}\}$.

□

Rappels sur les matrices

Définition.

On appelle **matrice** de type (p, n) à coefficients dans \mathbb{K} un tableau \mathbf{A} de pn éléments de \mathbb{K} rangés sur p lignes et n colonnes :

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

ou en abrégé $\mathbf{A} = (a_{ij})$.

L'ensemble des matrices à p lignes et n colonnes est noté $\mathcal{M}_{p,n}(\mathbb{K})$. Si $p = n$ on notera $\mathcal{M}_{n,n}(\mathbb{K})$ par $\mathcal{M}_n(\mathbb{K})$.

- Exemple : $\begin{pmatrix} 2 & 3 & -7 \\ 0 & -6 & 4 \end{pmatrix} \in \mathcal{M}_{2,3}(\mathbb{R})$, $\begin{pmatrix} 1-i & 5 \\ 4i & 3+2i \end{pmatrix} \in \mathcal{M}_2(\mathbb{C})$

Rappels sur les matrices (suite)

- Sur l'ensemble $\mathcal{M}_{p,n}(\mathbb{K})$ on définit les lois suivantes :
 - addition** : si $\mathbf{A} = (a_{ij})$ et $\mathbf{B} = (b_{ij})$ alors on note $\mathbf{C} = \mathbf{A} + \mathbf{B}$ la matrice $\mathbf{C} = (c_{ij})$ tel que $\forall i = 1, \dots, p; \forall j = 1, \dots, n : c_{ij} = a_{ij} + b_{ij}$.
 - produit par un scalaire** : si $\mathbf{A} = (a_{ij})$ et $\lambda \in \mathbb{K}$ alors on note $\lambda\mathbf{A}$ la matrice de terme général (λa_{ij}) .
- Exemple :

$$\begin{pmatrix} 2 & 3 & -7 \\ 0 & -6 & 4 \end{pmatrix} + \begin{pmatrix} 1 & -1 & 6 \\ 0 & -1 & 0 \end{pmatrix} = \begin{pmatrix} 3 & 2 & -1 \\ 0 & -7 & 4 \end{pmatrix}$$

$$2 \begin{pmatrix} 2 & 3 & -7 \\ 0 & -6 & 4 \end{pmatrix} = \begin{pmatrix} 4 & 6 & -14 \\ 0 & -12 & 8 \end{pmatrix}$$

Matrices et applications linéaires

- Soient \mathbb{E} et \mathbb{E}' deux ev sur \mathbb{K} , de dimension n et p respectivement et $f : \mathbb{E} \rightarrow \mathbb{E}'$ une application linéaire. Prenons une base $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ de \mathbb{E} et une base $\{\epsilon_1, \dots, \epsilon_p\}$ de \mathbb{E}' . Les images par f des vecteurs $\mathbf{e}_1, \dots, \mathbf{e}_n$ se décomposent sur la base $\{\epsilon_1, \dots, \epsilon_p\}$:

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

Définition.

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

Matrices et applications linéaires (suite)

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

\updownarrow

$$M(f)_{\mathbf{e}_i, \epsilon_j} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & \dots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pn} \end{pmatrix}$$

- On pourra utiliser également la notation $(f(\mathbf{e}_1) \dots f(\mathbf{e}_n))_{\epsilon_j}$. S'il n'y a pas d'ambiguïté on écrira aussi $M(f)$ mais il est clair que la matrice associée à f dépend du choix des bases de \mathbb{E} et \mathbb{E}' .

Matrices et applications linéaires (suite)

Propriété.

Soient \mathbb{E} et \mathbb{E}' deux ev sur \mathbb{K} de dimension n et p respectivement, $\{e_i\}$ et $\{\epsilon_j\}$ des bases de \mathbb{E} et \mathbb{E}' . Alors l'application :

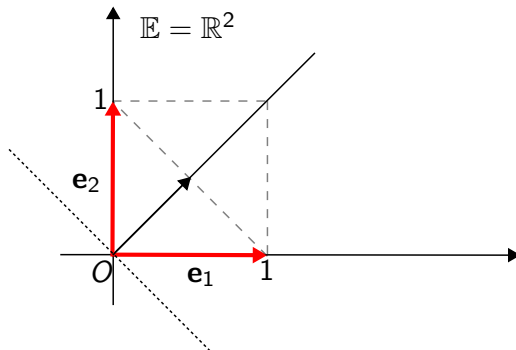
$$\begin{cases} M : \mathcal{L}_{\mathbb{K}}(\mathbb{E}, \mathbb{E}') \rightarrow \mathcal{M}_{p,n}(\mathbb{K}) \\ f \rightarrow M(f)_{e_i, \epsilon_j} \end{cases}$$

est un **isomorphisme** d'espaces vectoriels càd M est bijective, $M(f + g) = M(f) + M(g)$ et $M(\lambda f) = \lambda M(f)$.

En particulier $\dim(\mathcal{L}(\mathbb{E}, \mathbb{E}')) = pn$

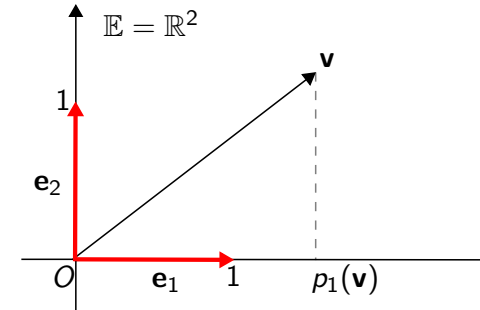
- ▶ On peut représenter l'ensemble des applications linéaires d'un ev dans un autre par le biais de matrices et vice-versa.
- L'ensemble des applications linéaires de l'ev \mathbb{E} de dimension n dans l'ev \mathbb{E}' de dimension p est un ev de dimension pn .

Exemple (suite)



- Soit $\mathbb{E} = \mathbb{R}^2$ et f la projection sur la première bissectrice parallèlement à la deuxième bissectrice.
- Considérons la base canonique de \mathbb{R}^2 .
- On a $f(e_1) = f(e_2) = \epsilon = \frac{1}{2}(e_1 + e_2)$ et donc $M(f)_{e_i} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$.

Exemple



- Soit $\mathbb{E} = \mathbb{R}^2$ et $\begin{cases} p_1 : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \\ (x_1, x_2) \rightarrow (x_1, 0) \end{cases}$.
- Considérons la base canonique de \mathbb{R}^2 .
- On a $p_1(e_1) = e_1$ et $p_1(e_2) = (0, 0)$ et donc :

$$M(p_1)_{e_i} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

Composition d'applications linéaires

Propriété.

- Soit $f \in \mathcal{L}_{\mathbb{K}}(\mathbb{E}, \mathbb{E}')$ et $g \in \mathcal{L}_{\mathbb{K}}(\mathbb{E}', \mathbb{E}'')$ alors la **composition** : $\begin{cases} g \circ f : \mathbb{E} \rightarrow \mathbb{E}'' \\ x \rightarrow g(f(x)) \end{cases}$ est un élément de $\mathcal{L}_{\mathbb{K}}(\mathbb{E}, \mathbb{E}'')$.
- Par ailleurs, $\forall f, h \in \mathcal{L}_{\mathbb{K}}(\mathbb{E}, \mathbb{E}')$; $g, k \in \mathcal{L}_{\mathbb{K}}(\mathbb{E}', \mathbb{E}'')$, $\forall \lambda \in \mathbb{K}$:
 - ▶ $g \circ (f + h) = g \circ f + g \circ h$
 - ▶ $(g + k) \circ f = g \circ f + k \circ f$
 - ▶ $g \circ (\lambda f) = \lambda g \circ f$
- Enfin si f est bijective f^{-1} est linéaire.

- Comme précédemment, nous allons montrer que nous pouvons représenter et déterminer les compositions d'applications linéaires entre ev en utilisant des matrices.

Rappels sur les matrices (suite)

- On appelle **produit** de matrices l'application :

$$\left\{ \begin{array}{l} \mathcal{M}_{p,n}(\mathbb{K}) \times \mathcal{M}_{n,q}(\mathbb{K}) \rightarrow \mathcal{M}_{p,q}(\mathbb{K}) \\ \mathbf{A} \quad , \quad \mathbf{B} \quad \rightarrow \quad \mathbf{C} \end{array} \right.$$

où $\mathbf{C} = (c_{ij})$ est tel que $\forall i = 1, \dots, p, \forall j = 1, \dots, q$:

$$\begin{aligned} c_{ij} &= a_{i1}b_{1j} + a_{i2}b_{2j} \dots + a_{in}b_{nj} \\ &= \sum_{k=1}^n a_{ik}b_{kj} \end{aligned}$$

- Exemple si $\mathbf{A} = \begin{pmatrix} 2 & -1 & 1 \\ 1 & 0 & 2 \end{pmatrix}$ et $\mathbf{B} = \begin{pmatrix} 2 & 0 & 1 & 0 \\ 1 & 1 & 2 & 2 \\ 1 & 2 & 1 & 0 \end{pmatrix}$ alors :

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

Rappels sur les matrices (suite)

- La multiplication matricielle est :

► **Associative** : $\forall \mathbf{A} \in \mathcal{M}_{p,n}, \forall \mathbf{B} \in \mathcal{M}_{n,q}, \forall \mathbf{C} \in \mathcal{M}_{q,m}$,

$$\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$$

► **Distributive** à gauche et à droite par rapport à l'addition : $\forall \mathbf{A}, \mathbf{D} \in \mathcal{M}_{p,n}, \forall \mathbf{B}, \mathbf{C} \in \mathcal{M}_{n,q}$,

$$\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC} \text{ et } (\mathbf{A} + \mathbf{D})\mathbf{B} = \mathbf{AB} + \mathbf{DB}$$

“Matrice” d'un vecteur

Définition.

Soit \mathbb{E} un ev de dimension n , $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ une base de \mathbb{E} et $\mathbf{x} = x_1\mathbf{e}_1 + \dots + x_n\mathbf{e}_n$ un vecteur de \mathbb{E} . On appelle **matrice de \mathbf{x} dans la base $\{\mathbf{e}_i\}$** la **matrice colonne des composantes de \mathbf{x} dans $\{\mathbf{e}_i\}$** :

$$M(\mathbf{x})_{\mathbf{e}_i} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

ce qui sera aussi noté plus simplement par $M(\mathbf{x})$ ou \mathbf{x} si pas d'ambiguïté.

- Un vecteur $\mathbf{x} \in \mathbb{E}$ est formellement défini comme combinaison linéaire des éléments de la base $\{\mathbf{e}_i\}$. La représentation $M(\mathbf{x})_{\mathbf{e}_i}$ est la matrice colonne ($n \times 1$) remplie des composantes de la combinaison linéaire définissant \mathbf{x} . En fait, par abus de langage on utilise le terme **vecteur \mathbf{x}** pour désigner la représentation matricielle de cet élément.

Calcul de l'image d'un vecteur

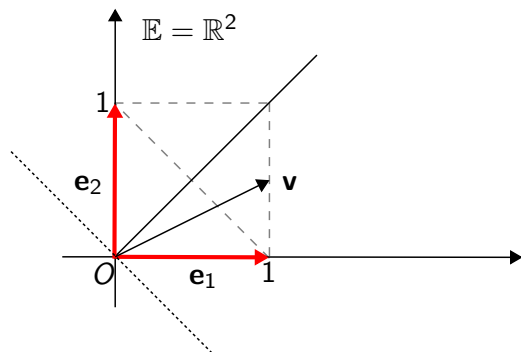
Propriété.

Soient \mathbb{E} et \mathbb{E}' deux ev sur \mathbb{K} , $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ et $\{\epsilon_1, \dots, \epsilon_p\}$ deux bases de \mathbb{E} et \mathbb{E}' respectivement. Pour toute application $f \in \mathcal{L}_{\mathbb{K}}(\mathbb{E}, \mathbb{E}')$ et pour tout vecteur $\mathbf{x} \in \mathbb{E}$, on a :

$$M(f(\mathbf{x}))_{\epsilon_j} = M(f)_{\epsilon_i, \epsilon_j} M(\mathbf{x})_{\mathbf{e}_i}$$

ou plus brièvement $M(f(\mathbf{x})) = M(f)M(\mathbf{x}) = M(f)\mathbf{x}$.

Exemple



- Soit $\mathbb{E} = \mathbb{R}^2$ et f la projection sur la première bissectrice parallèlement à la deuxième bissectrice.
- Considérons la base canonique de \mathbb{R}^2 . L'image du vecteur $\mathbf{v} = (1, \frac{1}{2})$ est donnée par : $M(f)_{\mathbf{e}_i} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{3}{4} \\ \frac{3}{4} \end{pmatrix}$.

Inverse d'une matrice carrée

Définition.

Une matrice carrée $\mathbf{A} \in \mathcal{M}_n(\mathbb{K})$ est dite **inversible** s'il existe une matrice $\mathbf{A}' \in \mathcal{M}_n(\mathbb{K})$ telle que :

$$\mathbf{A}\mathbf{A}' = \mathbf{A}'\mathbf{A} = \mathbf{I}$$

où \mathbf{I} est la matrice diagonale d'ordre n .

\mathbf{A}' est dite **inverse** de \mathbf{A} et est notée \mathbf{A}^{-1} .

- Exemple soit $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix}$ alors :

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

Composition d'applications linéaires et multiplication matricielle

Propriété.

Soient \mathbb{E}, \mathbb{E}' et \mathbb{E}'' trois ev de dimension finie sur \mathbb{K} , $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$, $\{\epsilon_1, \dots, \epsilon_p\}$ et $\{\eta_1, \dots, \eta_q\}$ des bases de \mathbb{E}, \mathbb{E}' et \mathbb{E}'' respectivement. Si $g \in \mathcal{L}_{\mathbb{K}}(\mathbb{E}, \mathbb{E}')$ $f \in \mathcal{L}_{\mathbb{K}}(\mathbb{E}', \mathbb{E}'')$, alors on a :

$$M(f \circ g)_{\mathbf{e}_i, \eta_k} = M(f)_{\epsilon_j, \eta_k} M(g)_{\mathbf{e}_i, \epsilon_j}$$

ou plus brièvement $M(f \circ g) = M(f)M(g)$.

Inverse d'une matrice carrée et isomorphisme

Propriété.

Soient \mathbb{E} et \mathbb{E}' deux ev de même dimension n sur \mathbb{K} , $\{\mathbf{e}_i\}$ une base de \mathbb{E} , $\{\epsilon_j\}$ une base de \mathbb{E}' . Une application linéaire $f : \mathbb{E} \rightarrow \mathbb{E}'$ est **bijective** ssi $M(f)_{\mathbf{e}_i, \epsilon_j}$ est **inversible**. De plus :

$$M(f)^{-1}_{\mathbf{e}_i, \epsilon_j} = M(f^{-1})_{\epsilon_j, \mathbf{e}_i}$$

ou d'une manière plus concise : $M(f^{-1}) = M(f)^{-1}$.

Rang d'une application linéaire et rang d'une matrice

- On a vu (cf slide 138) que le rang d'une application linéaire est la dimension de l'image de celle-ci. De plus, on a établi l'isomorphisme d'ev entre $\mathcal{L}_{\mathbb{K}}(\mathbb{E}, \mathbb{E}')$ et $\mathcal{M}_{p,n}(\mathbb{K})$. Comment se traduit le rang d'une application linéaire par le biais de sa matrice associée ?

Propriété.

Soient \mathbb{E} et \mathbb{E}' deux ev de dimension finie et $f \in \mathcal{L}(\mathbb{E}, \mathbb{E}')$. Soient $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ et $\{\epsilon_1, \dots, \epsilon_p\}$ deux bases quelconques de \mathbb{E} et \mathbb{E}' respectivement et $\mathbf{A} = M(f)_{\mathbf{e}_i, \epsilon_j}$, on a alors :

$$\text{rg}(f) = \text{rg}(\mathbf{A})$$

- Rappel : le rang d'une matrice est le nombre de vecteurs colonnes qui soient linéairement indépendants.

Rappel du Sommaire

4 Algèbre linéaire

- Définitions
- Bases d'un espace vectoriel
- Les théorèmes fondamentaux sur la dimension et sommes directes
- Applications linéaires et matrices
- Déterminants**
- Formes bilinéaires et formes quadratiques
- Espaces euclidiens
- Eléments de topologie dans \mathbb{R}^n

Transposée d'une matrice et rang associé

Définition.

Soit $\mathbf{A} \in \mathcal{M}_{p,n}(\mathbb{K})$ et soit \mathbf{A}^\top la matrice de $\mathcal{M}_{n,p}(\mathbb{K})$ telle que les lignes de \mathbf{A}^\top sont les colonnes de \mathbf{A} . Alors \mathbf{A}^\top est dite la transposée de \mathbf{A} .

- Exemple :

$$\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 0 & -3 \\ 3 & 5 \end{pmatrix} \leftrightarrow \mathbf{A}^\top = \begin{pmatrix} 2 & 0 & 3 \\ 1 & -3 & 5 \end{pmatrix}$$

Propriété.

Pour toute matrice \mathbf{A} , on a :

$$\text{rg}(\mathbf{A}) = \text{rg}(\mathbf{A}^\top)$$

Introduction

- Le concept de déterminant est intéressant à plusieurs égards :
 - Le déterminant d'un ensemble de vecteurs rassemblés au sein d'une matrice permet de savoir si ceux-ci forment un système libre. Le déterminant permet donc de déterminer si n vecteurs d'un ev de dimension n forment une base.
 - Le déterminant d'un endomorphisme permet de savoir si celui-ci est bijectif ou pas. Il est lié au calcul de l'inverse d'une matrice carrée.
 - D'un point de vue purement algébrique, le déterminant est une forme multilinéaire alternée ce qui représente une généralisation des formes linéaires.
 - D'un point de vue purement géométrique, le déterminant correspond au calcul d'une aire ou d'un volume . . .
 - Le déterminant permet également de connaître le signe d'une forme quadratique (que l'on verra ultérieurement) en étudiant les mineurs dominants principaux. Ce type de problème intervient en optimisation multidimensionnelle.

Définition des déterminants par récurrence

Définition.

Soit $\mathbf{A} = (a_{ij}) \in \mathcal{M}_n(\mathbb{K})$. On définit, par récurrence, une application $\det : \mathcal{M}_n(\mathbb{K}) \rightarrow \mathbb{K}$ de la manière suivante :

- Si $n = 1$, càd $\mathbf{A} = (a)$ alors $\det(\mathbf{A}) = a$.
- Si $n > 1$, notons $\mathbf{A}_{ij} \in \mathcal{M}_{n-1}(\mathbb{K})$ la matrice obtenue de \mathbf{A} en supprimant la i ème ligne et la j ème colonne (càd la ligne et la colonne passant par l'élément a_{ij}), on pose alors :

$$\det(\mathbf{A}) = a_{11}\det(\mathbf{A}_{11}) + \dots + (-1)^{k+1}a_{1k}\det(\mathbf{A}_{1k}) + \dots + (-1)^{n+1}a_{1n}\det(\mathbf{A}_{1n})$$

Le scalaire $\det(\mathbf{A})$ est dit **déterminant** de \mathbf{A} et :

Si $\mathbf{A} = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix}$ alors on dénote $\det(\mathbf{A})$ par $\begin{vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{vmatrix}$.

Propriétés du déterminant

Théorème.

- Le déterminant est une application multilinéaire. Si

$\mathbf{A} = (\mathbf{c}_1 \ \dots \ \mathbf{c}_n)$ alors $\forall \lambda \in \mathbb{K}, \forall k = 1, \dots, n :$

- ▶ $\det((\mathbf{c}_1, \dots, \mathbf{a}_k + \mathbf{b}_k, \dots, \mathbf{c}_n)) = \det((\mathbf{c}_1, \dots, \mathbf{a}_k, \dots, \mathbf{c}_n)) + \det((\mathbf{c}_1, \dots, \mathbf{b}_k, \dots, \mathbf{c}_n))$
- ▶ $\det((\mathbf{c}_1, \dots, \lambda \mathbf{c}_k, \dots, \mathbf{c}_n)) = \lambda \det((\mathbf{c}_1, \dots, \mathbf{c}_k, \dots, \mathbf{c}_n))$

- Si deux colonnes sont égales, alors le déterminant est nul.

- Le déterminant est une forme (valeurs dans \mathbb{K}) multilinéaire (linéaire en chaque argument càd pour chaque colonne) alternée (si on permute deux colonnes, le déterminant change de signe).

Théorème.

Soit $\mathbf{A} = (\mathbf{c}_1 \ \dots \ \mathbf{c}_n) \in \mathcal{M}_n(\mathbb{K})$. Alors les vecteurs $\mathbf{c}_1, \dots, \mathbf{c}_n$ forment une base de \mathbb{K}^n ssi $\det(\mathbf{A}) \neq 0$. Autrement dit, un déterminant est nul ssi l'une des colonnes est combinaison linéaire des autres colonnes.

Déterminants de matrice de $\mathcal{M}_2(\mathbb{K})$

- Si $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ alors :

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

- Exemple :

$$\begin{aligned} \det \left(\begin{pmatrix} 4 & -1 \\ 3 & 2 \end{pmatrix} \right) &= \begin{vmatrix} 4 & -1 \\ 3 & 2 \end{vmatrix} \\ &= 4 \times 2 - (-1) \times 3 \end{aligned}$$

Déterminant de la transposée d'une matrice

Théorème.

Pour toute matrice $\mathbf{A} \in \mathcal{M}_n(\mathbb{K})$, on a :

$$\det(\mathbf{A}^\top) = \det(\mathbf{A})$$

Corollaire.

Toute les propriétés des déterminants relatives aux colonnes peuvent être affirmées pour les lignes :

- Le déterminant est une application linéaire par rapport à chaque ligne.
- Si une matrice a deux lignes égales, le déterminant est nul.
- Le déterminant d'une matrice est non nul ssi les vecteurs lignes sont indépendants.

Calcul des déterminants

- Nous avons présenté le calcul du déterminant en considérant le développement selon la première ligne. Compte tenu des résultats précédents, on peut considérer le calcul du déterminant en développant selon une ligne ou une colonne quelconque. Dans le cadre de cette généralisation on introduit :

Définition.

Soit $\mathbf{A} = (a_{ij}) \in \mathcal{M}_n(\mathbb{K})$. On appelle **cofacteur** de l'élément a_{ij} le scalaire :

$$\text{cof}(a_{ij}) = (-1)^{i+j} \det(\mathbf{A}_{ij})$$

- Exemple : $\mathbf{A} = \begin{pmatrix} 1 & 0 & -3 \\ 2 & 4 & -2 \\ 5 & -1 & 3 \end{pmatrix}$. Calculer $\text{cof}(a_{12})$.

$$\text{cof}(a_{12}) = \text{cof}(0) = (-1)^{1+2} \begin{vmatrix} 2 & -2 \\ 5 & 3 \end{vmatrix} = -16.$$

Propriétés

Propriété.

Le déterminant ne change pas si à une ligne (resp à une colonne), on ajoute une combinaison linéaire des autres lignes (resp des autres colonnes).

- En pratique, on fait usage de cette propriété afin de faire apparaître le plus de zéros sur les lignes ou les colonnes d'une matrice ce qui facilite le calcul du déterminant.

Calcul des déterminants (suite)

- En termes de cofacteurs, le développement du déterminant selon la première ligne s'écrit :
 $\det(\mathbf{A}) = a_{11}\text{cof}(a_{11}) + \dots + a_{1k}\text{cof}(a_{1k}) + \dots + a_{1n}\text{cof}(a_{1n})$.
- On peut généraliser le développement à toute ligne ou colonne quelconque de \mathbf{A} :

Théorème.

- ▶ Développement de $\det(\mathbf{A})$ selon la i ème ligne :

$$\det(\mathbf{A}) = a_{i1}\text{cof}(a_{i1}) + \dots + a_{ik}\text{cof}(a_{ik}) + \dots + a_{in}\text{cof}(a_{in})$$

- ▶ Développement de $\det(\mathbf{A})$ selon la j ème colonne :

$$\det(\mathbf{A}) = a_{1j}\text{cof}(a_{1j}) + \dots + a_{kj}\text{cof}(a_{kj}) + \dots + a_{nj}\text{cof}(a_{nj})$$

Exemple

- Calculez $\det(\mathbf{A}) = \begin{vmatrix} 1 & -2 & 1 & 3 & 4 \\ 1 & -1 & 0 & 2 & 4 \\ 2 & 1 & 3 & 1 & 2 \\ -1 & 0 & 1 & 1 & 3 \\ 0 & 1 & -1 & 1 & 3 \end{vmatrix}$.

- On a $\det(\mathbf{A}) = \begin{vmatrix} 1 & -2 & 1 & 3 & 4 \\ 0 & 1 & -1 & -1 & 0 \\ 0 & 5 & 1 & -5 & -6 \\ 0 & -2 & 2 & 4 & 7 \\ 0 & 1 & -1 & 1 & 3 \end{vmatrix}$.

- Si on développe par rapport à la 1ère colonne on obtient :

$$\det(\mathbf{A}) = \begin{vmatrix} 1 & -1 & -1 & 0 \\ 5 & 1 & -5 & -6 \\ -2 & 2 & 4 & 7 \\ 1 & -1 & 1 & 3 \end{vmatrix}$$

Exemple (suite)

- On a ensuite : $\det(\mathbf{A}) = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 5 & 6 & 0 & -6 \\ -2 & 0 & 2 & 7 \\ 1 & 0 & 2 & 3 \end{vmatrix}$.

- Si on développe maintenant par rapport à la 1ère ligne on obtient :

$$\det(\mathbf{A}) = \begin{vmatrix} 6 & 0 & -6 \\ 0 & 2 & 7 \\ 0 & 2 & 3 \end{vmatrix}.$$

- On développe par rapport à la 1ère colonne et on obtient :

$$\det(\mathbf{A}) = 6(6 - 14) = -48.$$

Déterminants d'un endomorphisme

Définition.

Soit \mathbb{E} un ev de dimension finie et f un endomorphisme de \mathbb{E} . On appelle **déterminant de f** , le déterminant de la matrice qui représente f dans une base (quelconque) de \mathbb{E} :

$$\det(f) = \det(M(f)_{\mathbf{e}_i})$$

$\{\mathbf{e}_i\}$ étant une base quelconque de \mathbb{E} .

Théorème.

$f \in \text{End}_{\mathbb{K}}(\mathbb{E})$ est un automorphisme ssi $\det(f) \neq 0$.

Déterminants du produit de matrices

Théorème.

Pour deux matrices $\mathbf{A}, \mathbf{B} \in \mathcal{M}_n(\mathbb{K})$, on a :

$$\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$$

Corollaire.

$\mathbf{A} \in \mathcal{M}_n(\mathbb{K})$ est inversible ssi $\det(\mathbf{A}) \neq 0$ et on a alors :

$$\det(\mathbf{A}^{-1}) = \frac{1}{\det(\mathbf{A})}$$

Calcul de l'inverse d'une matrice

Théorème.

Soit $\mathbf{A} \in \mathcal{M}_n(\mathbb{K})$ et $\text{cof}(\mathbf{A})$ la **comatrice** de \mathbf{A} (matrice des cofacteurs) dont le terme est défini par :

$$[\text{cof}(\mathbf{A})]_{ij} = \text{cof}(a_{ij})$$

On a alors le résultat suivant :

$$\mathbf{A}\text{cof}(\mathbf{A})^{\top} = \text{cof}(\mathbf{A})^{\top}\mathbf{A} = \det(\mathbf{A})\mathbf{I}$$

En particulier, si \mathbf{A} est inversible, on a :

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})}\text{cof}(\mathbf{A})^{\top}$$

Exemple

- Soit $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ -1 & 3 \end{pmatrix}$. On a $\det(\mathbf{A}) = 5$. Donc \mathbf{A} est inversible. On a $\text{cof}(\mathbf{A}) = \begin{pmatrix} 3 & 1 \\ -2 & 1 \end{pmatrix}$ et donc $\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \text{cof}(\mathbf{A})^T = \frac{1}{5} \begin{pmatrix} 3 & -2 \\ 1 & 1 \end{pmatrix}$.
- Soit $\mathbf{A} = \begin{pmatrix} 1 & 2 & 0 \\ -1 & 3 & 0 \\ 0 & 1 & -1 \end{pmatrix}$. On a $\det(\mathbf{A}) = -5$. Donc \mathbf{A} est inversible.
On a $\text{cof}(\mathbf{A}) = \begin{pmatrix} -3 & -1 & -1 \\ 2 & -1 & -1 \\ 0 & 0 & 5 \end{pmatrix}$ et donc $\mathbf{A}^{-1} = \frac{1}{-5} \begin{pmatrix} -3 & 2 & 0 \\ -1 & -1 & 0 \\ -1 & -1 & 5 \end{pmatrix}$.

Introduction

- Les formes bilinéaires sont des outils essentiels pour définir les espaces euclidiens. En effet, les notions d'angle, d'orthogonalité, de distances et de normes sont définies à partir du produit scalaire qui est une forme bilinéaire symétrique définie positive.
- Les formes quadratiques sont également importantes puisque certains algorithmes en optimisation multidimensionnelle reposent sur la matrice hessienne qui est une forme quadratique. En particulier le signe d'une matrice hessienne permet de savoir si un point critique est un minimiseur ou un maximiseur.

Rappel du Sommaire

- 4 Algèbre linéaire
 - Définitions
 - Bases d'un espace vectoriel
 - Les théorèmes fondamentaux sur la dimension et sommes directes
 - Applications linéaires et matrices
 - Déterminants
 - Formes bilinéaires et formes quadratiques
 - Espaces euclidiens
 - Éléments de topologie dans \mathbb{R}^n

Forme bilinéaire

Définition. (Forme bilinéaire)

Soit \mathbb{E} un ev sur \mathbb{K} . On appelle une **forme bilinéaire**, une application $s : \mathbb{E} \times \mathbb{E} \rightarrow \mathbb{K}$ qui est linéaire en les deux arguments, c'à d qui vérifie :

- $s(\mathbf{x} + \mathbf{y}, \mathbf{z}) = s(\mathbf{x}, \mathbf{z}) + s(\mathbf{y}, \mathbf{z})$
- $s(\mathbf{x}, \mathbf{y} + \mathbf{z}) = s(\mathbf{x}, \mathbf{y}) + s(\mathbf{x}, \mathbf{z})$
- $s(\lambda \mathbf{x}, \mathbf{y}) = s(\mathbf{x}, \lambda \mathbf{y}) = \lambda s(\mathbf{x}, \mathbf{y})$

Caractéristique d'une forme bilinéaire

- Soit \mathbb{E} un ev de dimension n , $\{\mathbf{e}_i\}$ une base de \mathbb{E} et s une forme bilinéaire de \mathbb{E} . Si $\mathbf{x} = \sum_{i=1}^n x_i \mathbf{e}_i$ et $\mathbf{y} = \sum_{i=1}^n y_i \mathbf{e}_i$, on a :

$$s(\mathbf{x}, \mathbf{y}) = s\left(\sum_{i=1}^n x_i \mathbf{e}_i, \sum_{j=1}^n y_j \mathbf{e}_j\right) = \sum_{i,j=1}^n x_i y_j s(\mathbf{e}_i, \mathbf{e}_j)$$

- Les $s(\mathbf{e}_i, \mathbf{e}_j)$ sont des éléments de \mathbb{K} . Si on pose $a_{ij} = s(\mathbf{e}_i, \mathbf{e}_j)$ alors l'expression de s dans la base $\{\mathbf{e}_i\}$ est :

$$s(\mathbf{x}, \mathbf{y}) = \sum_{i,j=1}^n a_{ij} x_i y_j$$

càd s est du type :

$$s(\mathbf{x}, \mathbf{y}) = a_{11}x_1y_1 + a_{12}x_1y_2 + \dots + a_{ij}x_iy_j + \dots + a_{nn}x_ny_n$$

Caractéristique d'une forme bilinéaire symétrique

- Soit s une forme bilinéaire, elle est symétrique ssi :

$$s(\mathbf{x}, \mathbf{y}) = s(\mathbf{y}, \mathbf{x})$$

- Ainsi, on reconnaît qu'une forme bilinéaire est symétrique par le fait que, en l'écrivant dans une base, on a $\forall i, j = 1, \dots, n : a_{ij} = a_{ji}$.
- Exemples :
 - ▶ $s(\mathbf{x}, \mathbf{y}) = x_1y_1 + 2x_2y_2 - x_3y_1 + x_1y_3$ n'est pas une forme bilinéaire symétrique.
 - ▶ $s(\mathbf{x}, \mathbf{y}) = 2x_1y_1 - 3x_1y_2 - 3x_2y_1 + x_1y_3 + x_3y_1$ est une forme bilinéaire symétrique.

Caractéristique d'une forme bilinéaire

- Ainsi, si s est une forme bilinéaire elle peut être mise sous la forme :

$$s(\mathbf{x}, \mathbf{y}) = a_{11}x_1y_1 + a_{12}x_1y_2 + \dots + a_{ij}x_iy_j + \dots + a_{nn}x_ny_n$$

- Par conséquent, on reconnaît une forme bilinéaire par le fait que, en l'écrivant dans une base, on obtient une **somme de monômes dans lesquels x_i et y_j apparaissent à la puissance 1**.
- Exemples :
 - ▶ $s(\mathbf{x}, \mathbf{y}) = x_1y_1 + 2x_2y_2 - x_3y_1 + x_1y_3$ est une forme bilinéaire.
 - ▶ $s(\mathbf{x}, \mathbf{y}) = 2x_1y_1 + x_1^2y_2 - x_3y_2^2 + x_1y_3$ n'est pas une forme bilinéaire.
 - ▶ $s(\mathbf{x}, \mathbf{y}) = 2x_1y_1 + x_1 - x_3y_2$ n'est pas une forme bilinéaire.

Représentation matricielle d'une forme bilinéaire

- Pour les calculs, il est plus pratique d'utiliser l'**expression matricielle de la forme bilinéaire**.
- Soit \mathbb{E} un ev de dimension finie n sur \mathbb{K} et $s : \mathbb{E} \times \mathbb{E} \rightarrow \mathbb{K}$ une forme bilinéaire.
- Si $\{\mathbf{e}_i\}$ est une base de \mathbb{E} et $\mathbf{x} = \sum_{i=1}^n x_i \mathbf{e}_i$, $\mathbf{y} = \sum_{i=1}^n y_i \mathbf{e}_i$, on a :

$$s(\mathbf{x}, \mathbf{y}) = \sum_{i,j=1}^n x_i y_j s(\mathbf{e}_i, \mathbf{e}_j)$$

- s est donc déterminée par la **connaissance des valeurs $s(\mathbf{e}_i, \mathbf{e}_j)$ sur une base**.

Matrice d'une forme bilinéaire

Définition.

Soient s une forme bilinéaire sur \mathbb{E} et $\{\mathbf{e}_i\}$ une base de \mathbb{E} . On appelle **matrice de s dans la base $\{\mathbf{e}_i\}$** la matrice :

$$M_b(s)_{\mathbf{e}_i} = \begin{pmatrix} s(\mathbf{e}_1, \mathbf{e}_1) & s(\mathbf{e}_1, \mathbf{e}_2) & \dots & s(\mathbf{e}_1, \mathbf{e}_n) \\ s(\mathbf{e}_2, \mathbf{e}_1) & s(\mathbf{e}_2, \mathbf{e}_2) & \dots & s(\mathbf{e}_2, \mathbf{e}_n) \\ \vdots & \vdots & \dots & \vdots \\ s(\mathbf{e}_n, \mathbf{e}_1) & s(\mathbf{e}_n, \mathbf{e}_2) & \dots & s(\mathbf{e}_n, \mathbf{e}_n) \end{pmatrix}$$

On écrira aussi $M_b(s)_{\mathbf{e}_i} = (s(\mathbf{e}_i, \mathbf{e}_j))$.

- Ainsi l'élément à la i ème ligne et la j ème colonne correspond au coefficient de $x_i y_j$.
- Attention : l'application M_b qui à toute forme bilinéaire associe une matrice, ne doit pas être confondue avec l'isomorphisme M qui à toute application linéaire associe une matrice.

Exemples

- Soit $s : \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}$ la forme bilinéaire qui dans la base canonique $\{\mathbf{e}_i\}$ s'écrit :
 $s(\mathbf{x}, \mathbf{y}) = 5x_1 y_1 - 2x_2 y_2 + 4x_3 y_3 + 7x_1 y_2 + 6x_1 y_3 - 4x_3 y_1 + 2x_2 y_3 + 8x_3 y_2$.

- On a alors :

$$M_b(s)_{\mathbf{e}_i} = \begin{pmatrix} 5 & 7 & 6 \\ 0 & -2 & 2 \\ -4 & 8 & 4 \end{pmatrix}$$

- L'image des vecteurs $\mathbf{x} = (1, 1, 0)$ et $\mathbf{y} = (0, 2, 4)$ est :

$$\begin{aligned} s(\mathbf{x}, \mathbf{y}) &= (1 \ 1 \ 0) \begin{pmatrix} 5 & 7 & 6 \\ 0 & -2 & 2 \\ -4 & 8 & 4 \end{pmatrix} \begin{pmatrix} 0 \\ 2 \\ 4 \end{pmatrix} \\ &= 42 \end{aligned}$$

Détermination des images par calcul matriciel

Propriété.

Soit \mathbb{E} un ev sur \mathbb{K} de base $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$. Pour toute forme bilinéaire $s : \mathbb{E} \times \mathbb{E} \rightarrow \mathbb{K}$ représentée par la matrice $M_b(s)_{\mathbf{e}_i} = \mathbf{S}$ et pour tout vecteur $\mathbf{x}, \mathbf{y} \in \mathbb{E}$, on a :

$$\begin{aligned} s(\mathbf{x}, \mathbf{y}) &= M(\mathbf{x})_{\mathbf{e}_i}^\top M_b(s)_{\mathbf{e}_i} M(\mathbf{y})_{\mathbf{e}_i} \\ &= \mathbf{x}^\top \mathbf{S} \mathbf{y} \end{aligned}$$

Exemples (suite)

- Soit $s : \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}$ la forme bilinéaire symétrique qui dans la base canonique $\{\mathbf{e}_i\}$ s'écrit : $s(\mathbf{x}, \mathbf{y}) = 2x_1 y_1 - 3x_1 y_2 - 3x_2 y_1 + x_1 y_3 + x_3 y_1$.
- On a alors :

$$M_b(s)_{\mathbf{e}_i} = \begin{pmatrix} 2 & -3 & 1 \\ -3 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

- L'image des vecteurs $\mathbf{x} = (1, 1, 0)$ et $\mathbf{y} = (0, 2, 4)$ est :

$$\begin{aligned} s(\mathbf{x}, \mathbf{y}) &= (1 \ 1 \ 0) \begin{pmatrix} 2 & -3 & 1 \\ -3 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 2 \\ 4 \end{pmatrix} \\ &= -2 \end{aligned}$$

- On pourra vérifier que $s(\mathbf{x}, \mathbf{y}) = s(\mathbf{y}, \mathbf{x})$.

Forme quadratique

- A toute forme bilinéaire symétrique $s : \mathbb{E} \times \mathbb{E} \rightarrow \mathbb{R}$, on peut lui associer une application $q : \mathbb{E} \rightarrow \mathbb{R}$.

Définition. (Forme quadratique)

Soit \mathbb{E} un ev sur \mathbb{K} et $s : \mathbb{E} \times \mathbb{E} \rightarrow \mathbb{K}$ une forme bilinéaire symétrique. Alors l'application $q : \mathbb{E} \rightarrow \mathbb{K}$ définie par :

$$q(\mathbf{x}) = s(\mathbf{x}, \mathbf{x})$$

est appelée **forme quadratique** sur \mathbb{E} associée à s .

- q n'est pas une forme possédant la propriété de linéarité.

Détermination des images par calcul matriciel

- Comme pour les formes bilinéaires, nous pouvons représenter les formes quadratiques par des matrices (cf slide 180).
- Et donc calculer les images $q(\mathbf{x})$ par calcul matriciel.

Propriété.

Soient \mathbb{E} un ev sur \mathbb{K} de base $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$. Pour toute forme quadratique $q : \mathbb{E} \rightarrow \mathbb{K}$ représentée par la matrice $M_b(q)_{\mathbf{e}_i} = \mathbf{Q}$ et pour tout vecteur $\mathbf{x} \in \mathbb{E}$, on a :

$$\begin{aligned} q(\mathbf{x}) &= M(\mathbf{x})_{\mathbf{e}_i}^\top M_b(q)_{\mathbf{e}_i} M(\mathbf{x})_{\mathbf{e}_i} \\ &= \mathbf{x}^\top \mathbf{Q} \mathbf{x} \end{aligned}$$

- Pour simplifier la lecture on notera la matrice de la forme quadratique q associée à s par \mathbf{Q} et non \mathbf{S} .
- On remarquera que \mathbf{Q} est forcément **symétrique** ($\mathbf{Q} = \mathbf{Q}^\top$).

Caractéristique d'une forme quadratique

- Soit s une forme bilinéaire symétrique et q la forme quadratique associée.
- On a vu comment reconnaître algébriquement une forme bilinéaire symétrique.
- Comme $q(\mathbf{x}) = s(\mathbf{x}, \mathbf{x})$, il est facile de voir que l'écriture algébrique de q dans une base de \mathbb{E} est un **polynôme homogène de degré 2** en les composantes x_i de \mathbf{x} .
- Exemple :
 - ▶ Si $s(\mathbf{x}, \mathbf{y}) = 2x_1y_1 - 3x_1y_2 - 3x_2y_1 + x_1y_3 + x_3y_1$ alors la forme quadratique associée est $q(\mathbf{x}) = s(\mathbf{x}, \mathbf{x}) = 2x_1^2 - 6x_1x_2 + 2x_1x_3$.
 - ▶ Si $s(\mathbf{x}, \mathbf{y}) = 2x_1y_1 + x_1 - x_3y_2$ alors $s(\mathbf{x}, \mathbf{x}) = 2x_1^2 + x_1 - x_2x_3$ n'est pas une forme quadratique (s n'étant pas une forme bilinéaire).

Signe d'une forme quadratique

Définition.

Soit $q : \mathbb{E} \rightarrow \mathbb{K}$ une forme quadratique sur \mathbb{E} et soit \mathbf{Q} sa représentation matricielle. On dit que q est :

- **définie** si : $q(\mathbf{x}) = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$
- **définie positive (dp)** si elle est définie et si :

$$\forall \mathbf{x} \in \mathbb{E}, \mathbf{x} \neq \mathbf{0}, \mathbf{x}^\top \mathbf{Q} \mathbf{x} > 0$$
- **définie négative (dn)** si elle est définie et si :

$$\forall \mathbf{x} \in \mathbb{E}, \mathbf{x} \neq \mathbf{0}, \mathbf{x}^\top \mathbf{Q} \mathbf{x} < 0$$
- **semi-définie positive (sdp)** si :

$$\forall \mathbf{x} \in \mathbb{E}, \mathbf{x} \neq \mathbf{0}, \mathbf{x}^\top \mathbf{Q} \mathbf{x} \geq 0$$
- **semi-définie négative (sdn)** si :

$$\forall \mathbf{x} \in \mathbb{E}, \mathbf{x} \neq \mathbf{0}, \mathbf{x}^\top \mathbf{Q} \mathbf{x} \leq 0$$
- **indéfinie** si elle est ni sdp ni sdn.

Comment déterminer le signe d'une forme quadratique

- Il existe plusieurs méthodes pour étudier le signe d'une forme quadratique.
- Nous étudierons deux approches :
 - ▶ Le critère de Sylvester reposant sur le calcul de déterminants de sous-matrices de \mathbf{Q} .
 - ▶ La détermination du spectre de \mathbf{Q} où le signe des valeurs propres permet de conclure sur le signe de la forme quadratique.
- Dans la suite par abus de langage, nous noterons une forme quadratique soit par q soit par sa matrice représentative \mathbf{Q} .

Exemple

- Soit la forme quadratique représentée par :

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

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

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

- On a $\Delta_1 = 3, \Delta_2 = 5, \Delta_3 = 12$ donc \mathbf{Q} est dp.
- Vérifiez par le critère de Sylvester que $-\mathbf{Q}$ est dn.

Critère de Sylvester

Définition. (Mineurs principaux dominants)

Les **mineurs principaux dominants** d'une matrice carrée \mathbf{A} sont les n valeurs dénotés $\Delta_1, \dots, \Delta_n$ où Δ_i est le déterminant de la sous-matrice de \mathbf{A} composée des i premières lignes et colonnes de cette-dernière.

Théorème. (Critère de Sylvester)

Soit $q : \mathbb{E} \rightarrow \mathbb{K}$ une forme quadratique sur \mathbb{E} et \mathbf{Q} sa matrice carré symétrique associée.

- q est dp ssi les mineurs principaux dominants de \mathbf{Q} sont tels que $\Delta_i > 0$ pour tout i .
- q est dn ssi les mineurs principaux dominants de \mathbf{Q} sont tels que $\Delta_i > 0$ si i est pair et $\Delta_i < 0$ si i est impair.
- Si les mineurs principaux dominants sont non nuls et ne correspondent à aucun des deux cas précédents alors q est indéfinie.

Critère de Sylvester (suite)

Définition. (Mineurs principaux)

Les **mineurs principaux** d'ordre i d'une matrice carrée \mathbf{A} sont l'ensemble des déterminants de toutes les sous-matrices de \mathbf{A} pouvant être composées à partir de i lignes et des i colonnes correspondantes de cette-dernière.

Théorème. (Critère des mineurs principaux)

Soit $q : \mathbb{E} \rightarrow \mathbb{K}$ une forme quadratique sur \mathbb{E} et \mathbf{Q} sa matrice carré symétrique associée.

- q est sdv ssi les mineurs principaux de \mathbf{Q} d'ordre i sont positifs ou nuls pour tout i .
- q est sdn ssi les mineurs principaux de \mathbf{Q} sont positifs ou nuls si i est pair et négatifs ou nuls si i est impair.

Exemple (suite)

- Soit la forme quadratique représentée par :

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

- Les mineurs principaux d'ordre 1 de \mathbf{Q} sont :

$$\Delta_1 = |3|, \text{ mais aussi } |2|, |3|$$

- Les mineurs principaux d'ordre 2 de \mathbf{Q} sont :

$$\Delta_2 = \begin{vmatrix} 3 & -1 \\ -1 & 2 \end{vmatrix}, \text{ mais aussi } \begin{vmatrix} 3 & 0 \\ 0 & 3 \end{vmatrix}, \begin{vmatrix} 2 & -1 \\ -1 & 3 \end{vmatrix}$$

- Le mineur principal d'ordre 3 de \mathbf{Q} est :

$$\Delta_3 = \begin{vmatrix} 3 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 3 \end{vmatrix}$$

Rappels sur les matrices carrées diagonalisables

- Certains endomorphismes peuvent être exprimés dans une base particulière sous forme de matrice diagonale. On parle alors d'**endomorphisme diagonalisable**.

Définition. (Matrice carrée diagonalisable)

On dit que \mathbf{A} , une matrice carrée de taille $(n \times n)$, est diagonalisable sur \mathbb{K} , s'il existe une matrice carrée de taille $(n \times n)$, \mathbf{U} , qui soit inversible et une matrice diagonale \mathbf{D} de taille $(n \times n)$ à coefficients dans \mathbb{K} tel que :

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^{-1}$$

Les colonnes de $\mathbf{U} = (\mathbf{u}^1 \ \dots \ \mathbf{u}^n)$ sont alors des vecteurs propres et les termes de la diagonale de \mathbf{D} , $(\lambda_1, \dots, \lambda_n)$, les valeurs propres associées.

- Une matrice carrée est diagonalisable s'il existe une **base de vecteurs propres** telle que dans celle-ci, la matrice est diagonale.

Vecteurs et valeurs propres d'un endomorphisme

Définition.

Soit $f \in \text{End}_{\mathbb{K}}(\mathbb{E})$ et $\mathbf{A} = M(f)_{\mathbf{e}_i}$, où $\{\mathbf{e}_i\}$ est la base de \mathbb{E} . Un vecteur $\mathbf{u} \in \mathbb{E}$ est dit **vecteur propre** de f si $\mathbf{u} \neq \mathbf{0}$ et s'il existe $\lambda \in \mathbb{K}$ tel que $f(\mathbf{u}) = \lambda\mathbf{u}$, ce qui matriciellement s'écrit :

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

Le scalaire λ est dit **valeur propre** associée à \mathbf{u} . Le **spectre** de f noté $Sp_{\mathbb{K}}(f)$ (ou $Sp_{\mathbb{K}}(\mathbf{A})$) est l'ensemble des valeurs propres de f .

- Les vecteurs propres sont non nuls par définition mais 0 peut être valeur propre. $\text{Ker}(f)$ sont les vecteurs propres associés à 0.
- Si \mathbf{v} est un vecteur propre associé à λ , alors $\forall \mu \in \mathbb{K}, \mu \neq 0, \mu\mathbf{v}$ est aussi vecteur propre associé à λ car : $f(\mu\mathbf{v}) = \mu f(\mathbf{v}) = \lambda(\mu\mathbf{v})$.
- Les vecteurs propres sont soit les vecteurs du noyau, soit les vecteurs qui ne changent pas de direction sous l'action de f .

Valeurs propres et signe d'une forme quadratique

- Dans le cas particulier des **matrices carrées réelles symétriques** pouvant représenter des formes quadratiques, nous avons le résultat suivant.

Théorème.

Les valeurs propres d'une matrice carrée symétrique réelle sont réelles.

Théorème.

Soit $q : \mathbb{E} \rightarrow \mathbb{R}$ une forme quadratique et \mathbf{Q} sa matrice associée.

- q est dp ssi ses valeurs propres sont strictement positives.
- q est dn ssi ses valeurs propres sont strictement négatives.
- q est sdp ssi ses valeurs propres sont positives ou nulles.
- q est sdn ssi ses valeurs propres sont négatives ou nulles.
- q est indéfinie ssi ses valeurs propres sont de signes distincts.

Rappels sur la recherche de valeurs et vecteurs propres

- Soit λ une valeur propre de \mathbf{A} . Il existe donc $\mathbf{x} \neq \mathbf{0}$, $\mathbf{Ax} = \lambda\mathbf{x}$, ce qui est équivalent à $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$. Comme $\mathbf{x} \neq \mathbf{0}$, cela signifie que l'endomorphisme associé à $(\mathbf{A} - \lambda\mathbf{I})$ n'est pas injectif et donc il n'est pas bijectif.
- Pour déterminer les valeurs propres cela revient à résoudre (cf slide 140) :

$$\lambda \text{ est valeur propre de } \mathbf{A} \Leftrightarrow \det(\mathbf{A} - \lambda\mathbf{I}) = 0$$

- Une fois déterminée une valeur propre λ , on trouve son espace propre \mathbb{E}_λ en résolvant un système d'équations linéaires homogène :

$$\mathbf{x} \text{ est vecteur propre de } \mathbf{A} \text{ associé à } \lambda \Leftrightarrow (\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$$

Rappels sur le polynôme caractéristique

- Cette équation est appelée **équation caractéristique** et le membre de gauche **polynôme caractéristique** de \mathbf{A} noté $P_{\mathbf{A}}(\lambda)$.

Propriété.

Les valeurs propres de \mathbf{A} sont les racines du polynôme caractéristique :

$$P_{\mathbf{A}}(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I})$$

- Rappels de propriétés :
 - ▶ $tr(\mathbf{A})^1 = \lambda_1 + \dots + \lambda_n$
 - ▶ $\det(\mathbf{A}) = \lambda_1 \dots \lambda_n$
 - ▶ **Théorème de Cayley-Hamilton** :

$$P_{\mathbf{A}}(\mathbf{A}) = \mathbf{0}$$

où $\mathbf{0}$ est la matrice carrée de taille $(n \times n)$ remplie de 0.

1. tr est l'opérateur trace correspondant à la somme des termes de la diagonale.

Rappels sur la recherche de valeurs propres

- Soit une matrice carrée \mathbf{A} d'ordre n :

$$\mathbf{A} = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \dots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix}$$

- La condition pour que λ soit une valeur propre s'écrit :

$$\begin{vmatrix} a_{11} - \lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - \lambda & \dots & a_{2n} \\ \vdots & \dots & \ddots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} - \lambda \end{vmatrix} = 0$$

- En développant $|\mathbf{A} - \lambda\mathbf{I}|$, on obtient un polynôme de degré n en λ :

$$(-1)^n \lambda^n + b_{n-1} \lambda^{n-1} + \dots + b_1 \lambda + b_0 = 0$$

Exemple

- Prenons \mathbf{A} tel que :

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ -1 & 4 \end{pmatrix}$$

- On a :

$$P_{\mathbf{A}}(\lambda) = |\mathbf{A} - \lambda\mathbf{I}| = \begin{vmatrix} 1 - \lambda & 2 \\ -1 & 4 - \lambda \end{vmatrix} = \lambda^2 - 5\lambda + 6$$

- On remarque que 2 est une racine $P_{\mathbf{A}}(2) = 0$, on en déduit la factorisation suivante : $P_{\mathbf{A}}(\lambda) = (\lambda - 2)(\lambda - 3)$.
- Donc les valeurs propres de \mathbf{A} sont $\lambda_1 = 2$ et $\lambda_2 = 3$.
- Déterminez les vecteurs propres.
- Vérifiez que $\mathbf{A}^2 - 5\mathbf{A} + 6\mathbf{I} = \mathbf{0}$ (\mathbf{I} étant la matrice identité).

Rappel du Sommaire

4 Algèbre linéaire

- Définitions
- Bases d'un espace vectoriel
- Les théorèmes fondamentaux sur la dimension et sommes directes
- Applications linéaires et matrices
- Déterminants
- Formes bilinéaires et formes quadratiques
- **Espaces euclidiens**
- Eléments de topologie dans \mathbb{R}^n

Produit scalaire canonique dans \mathbb{R}^2 ou \mathbb{R}^3

- Soient deux vecteurs de \mathbb{R}^3 , $\mathbf{x} = (x_1, x_2, x_3)$ et $\mathbf{y} = (y_1, y_2, y_3)$. On appelle **produit scalaire** de \mathbf{x} par \mathbf{y} le scalaire noté et défini par :

$$\langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1 + x_2 y_2 + x_3 y_3$$

- Propriétés du produit scalaire :

- ▶ Il est **bilinéaire**, $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^3; \forall \lambda \in \mathbb{R}$:
 - ★ $\langle \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle$
 - ★ $\langle \lambda \mathbf{x}, \mathbf{z} \rangle = \lambda \langle \mathbf{x}, \mathbf{z} \rangle$
 - ★ $\langle \mathbf{x}, \mathbf{y} + \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{x}, \mathbf{z} \rangle$
 - ★ $\langle \mathbf{x}, \lambda \mathbf{z} \rangle = \lambda \langle \mathbf{x}, \mathbf{z} \rangle$
- ▶ Il est **symétrique**, $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^3 : \langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$
- ▶ Il est **défini positif**, $\forall \mathbf{x} \in \mathbb{R}^3$:
 - ★ $\langle \mathbf{x}, \mathbf{x} \rangle > 0$ pour $\mathbf{x} \neq \mathbf{0}$
 - ★ $\langle \mathbf{x}, \mathbf{x} \rangle = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$

Introduction

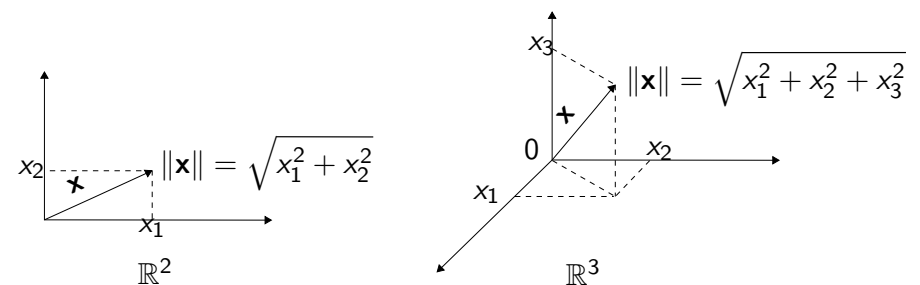
- Les espaces vectoriels (ev) constituent le cadre général de l'algèbre linéaire : il s'agit de la structure minimale permettant de traiter les problèmes linéaires.
- On peut ajouter des concepts supplémentaires aux ev permettant d'enrichir encore plus ces objets mathématiques.
- En particulier, on ajoute aux ev des concepts de **métriques** permettant de définir les notions de normes, angles et distances ... La notion de **produit scalaire** est fondamentale dans cette perspective.
- Nous commencerons par rappeler le produit scalaire dans \mathbb{R}^2 et \mathbb{R}^3 avant de généraliser à \mathbb{E} quelconque.
- En optimisation, les variables inconnues sont regroupées au sein d'un vecteur de \mathbb{R}^n vu comme un espace euclidien. Pour savoir notamment si un algorithme converge on a besoin de notions de distance et de topologie dans \mathbb{R}^n .

Norme dans \mathbb{R}^2 et \mathbb{R}^3

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

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \sqrt{x_1^2 + x_2^2 + x_3^2}$$

- Cette notion correspond à celle déduite du théorème de Pythagore.



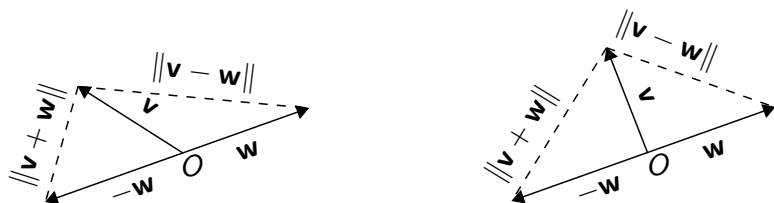
Orthogonalité dans \mathbb{R}^2 et \mathbb{R}^3

- On dit que deux vecteurs \mathbf{v} et \mathbf{w} sont **orthogonaux** si :

$$\langle \mathbf{v}, \mathbf{w} \rangle = 0$$

- On peut exprimer le fait que ces vecteurs soient orthogonaux par la condition suivante :

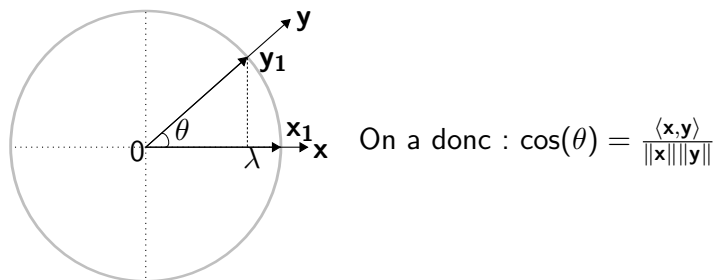
$$\|\mathbf{v} + \mathbf{w}\| = \|\mathbf{v} - \mathbf{w}\|$$



Angle non orienté dans \mathbb{R}^2 et \mathbb{R}^3 (suite)

- Si \mathbf{x} et \mathbf{y} sont deux vecteurs non nuls, l'**angle non orienté** entre ces vecteurs est l'angle compris entre 0 et π que forment les vecteurs de longueur 1 : $\mathbf{x}_1 = \frac{\mathbf{x}}{\|\mathbf{x}\|}$ et $\mathbf{y}_1 = \frac{\mathbf{y}}{\|\mathbf{y}\|}$.
- En trigonométrie, le **cosinus** de θ est défini par : $\cos(\theta) = \lambda$ où λ est le coefficient de la projection orthogonale de \mathbf{y}_1 sur \mathbf{x}_1 :

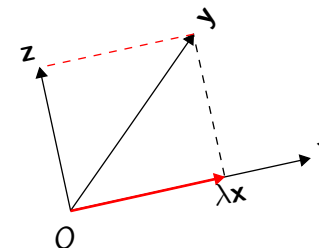
$$\lambda = \frac{\langle \mathbf{x}_1, \mathbf{y}_1 \rangle}{\|\mathbf{x}_1\|^2} = \langle \mathbf{x}_1, \mathbf{y}_1 \rangle = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|}$$



Angle non orienté dans \mathbb{R}^2 et \mathbb{R}^3

- Si \mathbf{x} et \mathbf{y} sont deux vecteurs non nuls, on définit d'abord (le coefficient de) la **projection orthogonale** de \mathbf{y} sur \mathbf{x} , comme étant le scalaire λ que nous noterons également par $Proj_{\mathbf{x}}(\mathbf{y})$, tel que le vecteur $\mathbf{z} = \mathbf{y} - \lambda\mathbf{x}$ soit orthogonal à \mathbf{x} :

$$\langle \mathbf{y} - \lambda\mathbf{x}, \mathbf{x} \rangle = 0 \Rightarrow \lambda = Proj_{\mathbf{x}}(\mathbf{y}) = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\|^2}$$



Espaces euclidiens

Définition.

Soit \mathbb{E} un espace vectoriel réel, on appelle **produit scalaire** sur \mathbb{E} , une application $\langle \cdot, \cdot \rangle : \mathbb{E} \times \mathbb{E} \rightarrow \mathbb{R}$ qui est :

- *Bilinéaire.*
- *Symétrique.*
- *Définie positive.*

Un espace vectoriel réel de dimension finie muni d'un produit scalaire est dit **espace euclidien**.

Exemple

- Soit $\mathbb{E} = \mathbb{R}^n$ associé à la base canonique avec $\langle \cdot, \cdot \rangle$ défini par : $\langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1 + \dots + x_n y_n$ où $\mathbf{x} = (x_1, \dots, x_n)$ et $\mathbf{y} = (y_1, \dots, y_n)$. Le produit scalaire ainsi défini est appelé **produit scalaire canonique**.
- Soit \mathbb{E} un ev réel de dimension n et $\{\mathbf{e}_i\}$ une base de \mathbb{E} . On définit le **produit scalaire associé à la base $\{\mathbf{e}_i\}$** en posant : $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{e}_i} = x_1 y_1 + \dots + x_n y_n$ avec $\mathbf{x} = \sum_i x_i \mathbf{e}_i$ et $\mathbf{y} = \sum_i y_i \mathbf{e}_i$.

Représentation matricielle du produit scalaire (suite)

- Lorsqu'on est dans une base orthonormée $\{\mathbf{e}_i\}$ de \mathbb{E} on a alors :

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^\top \mathbf{y}$$

- Une forme bilinéaire s est symétrique ssi sa matrice est symétrique.
- Un produit scalaire s est tel que sa matrice $\mathbf{S} = M_b(s)_{\mathbf{e}_i}$ vérifie :
 - ▶ $\mathbf{S} = \mathbf{S}^\top$
 - ▶ $\forall \mathbf{x} \in \mathbb{E}, \mathbf{x} \neq 0, \mathbf{x}^\top \mathbf{S} \mathbf{x} > 0$
 - ▶ $\mathbf{x}^\top \mathbf{S} \mathbf{x} = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$

Représentation matricielle du produit scalaire

Propriété.

Soient s une forme bilinéaire sur \mathbb{E} , $\{\mathbf{e}_i\}$ une base, et $\mathbf{S} = M_b(s)_{\mathbf{e}_i}$, $\mathbf{x} = M(\mathbf{x})_{\mathbf{e}_i}$, $\mathbf{y} = M(\mathbf{y})_{\mathbf{e}_i}$ où $\mathbf{x}, \mathbf{y} \in \mathbb{E}$. On a :

$$s(\mathbf{x}, \mathbf{y}) = \mathbf{x}^\top \mathbf{S} \mathbf{y}$$

- Exemple : reprenons la forme bilinéaire symétrique dp du slide 191 et

$$\mathbf{x} = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} \text{ et } \mathbf{y} = \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}, \text{ on a :}$$

$$s(\mathbf{x}, \mathbf{y}) = (1 \ 2 \ -1) \begin{pmatrix} 3 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 3 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix} = 6$$

Diagonalisation de matrices réelles symétriques

Propriété.

La matrice associée à un produit scalaire est telle que ses valeurs propres sont réelles et strictement positives.

Théorème. (Diagonalisation de matrices réelles symétriques)

Si \mathbf{A} est **réelle symétrique** alors elle est diagonalisable (cf slide 195) dans une base orthonormée. Dit autrement, les espaces propres de \mathbf{A} sont deux à deux orthogonaux et si on suppose pour chaque espace propre une base orthonormée^a alors \mathbf{U} est une matrice orthogonale^b et par conséquent :

$$\mathbf{A} = \mathbf{U} \mathbf{D} \mathbf{U}^\top$$

a. Obtenue par orthogonalisation de Schmidt par exemple (cf TD).

b. \mathbf{U} est orthogonale si $\mathbf{U}^\top = \mathbf{U}^{-1}$.

Exemple

- Soit :

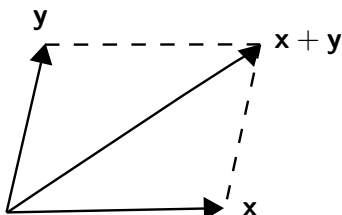
$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

- On a 3 valeurs propres : $Sp_{\mathbb{K}}(\mathbf{A}) = \{0, 1, 2\}$.
- On a 3 vecteurs propres : $\mathbf{v}_1 = (1, -1, 0)$, $\mathbf{v}_2 = (0, 0, 1)$, $\mathbf{v}_3 = (1, 1, 0)$ (donc 3 espaces propres distincts de dimension unitaire).
- Les 3 vecteurs propres sont mutuellement orthogonaux (car \mathbf{A} est réelle symétrique).
- Les 3 vecteurs propres normés sont : $\mathbf{e}_1 = \frac{1}{\sqrt{2}}(1, -1, 0)$, $\mathbf{e}_2 = \mathbf{v}_2 = (0, 0, 1)$, $\mathbf{e}_3 = \frac{1}{\sqrt{2}}(1, 1, 0)$.
- Vérifiez que $\mathbf{A} = \mathbf{UDU}^T$.

Inégalité triangulaire

- Inégalité triangulaire :

$$n(\mathbf{x} + \mathbf{y}) \leq n(\mathbf{x}) + n(\mathbf{y})$$



- L'inégalité triangulaire exprime le fait que dans un triangle la longueur d'un côté est inférieure ou égale à la somme des longueurs des deux autres côtés.

Norme

Définition.

Une norme est une application $n : \mathbb{E} \rightarrow \mathbb{R}^+$ qui vérifie les propriétés suivantes, $\forall \mathbf{x} \in \mathbb{E}; \forall \lambda \in \mathbb{R}$:

- $n(\lambda \mathbf{x}) = |\lambda|n(\mathbf{x})$
- $n(\mathbf{x}) = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$
- 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 $\mathbf{y} = \lambda \mathbf{x}$.

- Soit \mathbb{E} un ev réel muni d'un produit scalaire. L'application $\|\cdot\| : \mathbb{E} \rightarrow \mathbb{R}^+$ définie par, $\forall \mathbf{x} \in \mathbb{E}$:

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$$

est une norme et vérifie donc les propriétés énoncées précédemment.

Inégalités de Cauchy-Schwarz

Propriété. (Inégalité de Cauchy-Schwarz)

Pour tout $\mathbf{x}, \mathbf{y} \in \mathbb{E}$, on a :

$$\langle \mathbf{x}, \mathbf{y} \rangle^2 \leq \|\mathbf{x}\|^2 \|\mathbf{y}\|^2$$

ou encore :

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

Angle non orienté

- Soient \mathbf{x}, \mathbf{y} deux vecteurs non nuls. D'après l'inégalité de Cauchy-Schwarz on a la propriété suivante :

$$\frac{|\langle \mathbf{x}, \mathbf{y} \rangle|}{\|\mathbf{x}\| \|\mathbf{y}\|} \leq 1 \Leftrightarrow -1 \leq \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|} \leq 1$$

- Il existe un et un seul $\theta \in [0, \pi]$ tel que :

$$\cos(\theta) = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|}$$

θ est dit **angle** (non orienté) entre les vecteurs \mathbf{x} et \mathbf{y} .

- Notons enfin la relation qui exprime le **produit scalaire en fonction de la norme** :

$$\langle \mathbf{x}, \mathbf{y} \rangle = \frac{1}{2} (\|\mathbf{x} + \mathbf{y}\|^2 - \|\mathbf{x}\|^2 - \|\mathbf{y}\|^2)$$

Exemple

- Soit dans \mathbb{R}^2 muni du produit scalaire canonique, les vecteurs $\mathbf{u} = (1, 3)$ et $\mathbf{v} = (2, 1)$.
- Les normes sont $\|\mathbf{u}\| = \sqrt{10}$ et $\|\mathbf{v}\| = \sqrt{5}$.
- Le cosinus de l'angle entre \mathbf{u} et \mathbf{v} vaut $\langle \mathbf{u}, \mathbf{v} \rangle / (\|\mathbf{u}\| \|\mathbf{v}\|) = 5 / (\sqrt{10} \sqrt{5}) = 1/\sqrt{2}$.
- La distance entre \mathbf{u} et \mathbf{v} vaut $\|\mathbf{u} - \mathbf{v}\| = \sqrt{\langle \mathbf{u} - \mathbf{v}, \mathbf{u} - \mathbf{v} \rangle} = \sqrt{\langle (-1, 2), (-1, 2) \rangle} = \sqrt{5}$.
- Vérifiez l'inégalité triangulaire dans le cas de cet exemple.

Distance

Définition.

Une distance est une application $d : \mathbb{E} \times \mathbb{E} \rightarrow \mathbb{R}^+$ qui vérifie les propriétés suivantes, $\forall \mathbf{x}, \mathbf{y} \in \mathbb{E}; \forall \lambda \in \mathbb{R}$:

- **Symétrie** : $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$.
- $d(\mathbf{x}, \mathbf{y}) = 0 \Leftrightarrow \mathbf{x} = \mathbf{y}$.
- **Inégalité triangulaire** : $d(\mathbf{x}, \mathbf{z}) \leq d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z})$.

- Soit \mathbb{E} un ev réel muni d'un produit scalaire. L'application $\|\cdot\| : \mathbb{E} \times \mathbb{E} \rightarrow \mathbb{R}^+$ définie par, $\forall \mathbf{x}, \mathbf{y} \in \mathbb{E}$:

$$\|\mathbf{x} - \mathbf{y}\| = \sqrt{\langle \mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle}$$

est une distance et vérifie donc les propriétés énoncées précédemment.

Rappel du Sommaire

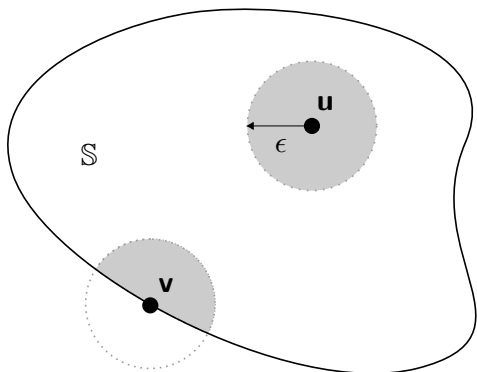
- 4 **Algèbre linéaire**
 - Définitions
 - Bases d'un espace vectoriel
 - Les théorèmes fondamentaux sur la dimension et sommes directes
 - Applications linéaires et matrices
 - Déterminants
 - Formes bilinéaires et formes quadratiques
 - Espaces euclidiens
 - Eléments de topologie dans \mathbb{R}^n

Introduction

- Riche des concepts de métriques vus précédemment, il est également utile de rappeler quelques concepts en topologie qui peuvent servir pour l'étude des fonctions numériques de plusieurs variables et en optimisation.
- Les concepts de voisinage d'un point de \mathbb{R}^n et de parties ouvertes, fermées et compactes de \mathbb{R}^n seront rappelés.

Frontière

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



Voisinage, point intérieur et point frontière

- 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} .
- Dans la suite \mathcal{S} est supposé être un sous-espace de \mathbb{R}^n : $\mathcal{S} \subset \mathbb{R}^n$.
- Un point $\mathbf{x} \in \mathcal{S}$ est un **point intérieur** de l'ensemble \mathcal{S} si ce dernier contient le voisinage de \mathbf{x} (càd si tous les points du voisinage de \mathbf{x} sont contenus dans \mathcal{S})
- L'ensemble des points intérieurs de \mathcal{S} est appelé l'**intérieur** de \mathcal{S} .
- Un point $\mathbf{x} \in \mathcal{S}$ est un **point frontière** de \mathcal{S} si le voisinage de \mathbf{x} contient au moins un point dans \mathcal{S} et au moins un point hors de \mathcal{S} .

Ensembles ouverts, fermés et compacts

- Un ensemble \mathcal{S} est un **ouvert** s'il contient le voisinage de chacun de ses points (càd \mathcal{S} n'a pas de points frontières mais que des points intérieurs).
- Un ensemble \mathcal{S} est un **fermé** s'il contient sa frontière (comme dans l'illustration précédente).
- Un ensemble \mathcal{S} qui est contenu dans une boule de rayon fini est dit **borné**.
- Un ensemble \mathcal{S} qui est à la fois fermé et borné est dit **compact**.
- Les ensembles compacts sont importants en optimisation comme on le verra avec le théorème de Weierstrass étendu au cas multidimensionnel (cf slide 84).

Rappel du Sommaire

- 1 Quelques méthodes statistiques du point de vue de l'optimisation
- 2 Fonctions d'une variable
- 3 Optimisation unidimensionnelle
- 4 Algèbre linéaire
- 5 Fonctions de plusieurs variables**
- 6 Optimisation multidimensionnelle

Rappel du Sommaire

- 5 Fonctions de plusieurs variables**
 - Fonctions de plusieurs variables
 - Limite et continuité d'une fonction de plusieurs variables
 - Dérivées partielles d'ordre 1
 - Différentiabilité
 - Lignes de niveau et gradients
 - Approximation linéaire
 - Dérivées partielles d'ordre 2
 - Convexité
 - Approximation quadratique

Introduction

- Nous avons fait des rappels sur les fonctions numériques réelles à valeurs réelles.
- Nous étendons ci-dessous les différents concepts de limites, continuité, différentiabilité, convexité au cas multidimensionnel.
- Le cas général concerne les fonctions de \mathbb{R}^n à valeurs dans \mathbb{R}^p (fonctions à valeurs vectorielles).
- Les espaces de départ et d'arrivée \mathbb{R}^n et \mathbb{R}^p sont vus tels des espaces euclidiens que nous avons rappelé précédemment. On supposera par défaut la base canonique et le produit scalaire canonique.
- Lors des rappels sur les ev nous avons déjà rencontré des fonctions de \mathbb{R}^n à valeurs dans \mathbb{R}^p particulières : les applications linéaires de \mathbb{R}^n à \mathbb{R}^p qui forment donc un cas particulier (propriété de linéarité qui conduit à l'isomorphisme entre $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$ et $\mathcal{M}_{p,n}$).
- Cependant, nous nous focaliserons plus particulièrement sur le cas $p = 1$ qui sont les fonctions de plusieurs variables à valeurs réelles qui sont les plus rencontrées en statistique.

Fonctions de plusieurs variables

Définition. (Fonction à valeurs vectorielles)

Une fonction de plusieurs variables à valeurs vectorielles f est une application $\mathbb{D}_f \rightarrow \mathbb{R}^p$ où \mathbb{D}_f , appelé le domaine (de définition) de f , est une partie non-vide de \mathbb{R}^n ($\mathbb{D}_f \subset \mathbb{R}^n$).

- Exemple :

$$f(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), f_3(\mathbf{x})) = (\underbrace{x^2 + y^2}_{f_1(\mathbf{x})}, \underbrace{\ln(x^2 + y^2)}_{f_2(\mathbf{x})}, \underbrace{\exp(x^2 + y^2)}_{f_3(\mathbf{x})}).$$

Définition. (Fonction à valeurs réelles)

Une fonction de plusieurs variables à valeurs réelles est une fonction f de plusieurs variables pour laquelle l'espace d'arrivée est \mathbb{R} .

- Exemple : $f(x, y) = x^2 + y^2$.

Graphe d'une fonction de plusieurs variables

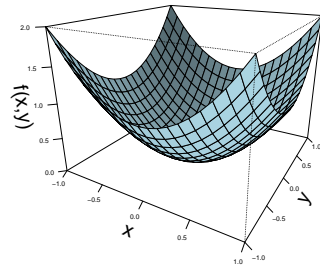
Définition. (Graphe)

Soit f une fonction définie sur $\mathbb{D}_f \subset \mathbb{R}^n$ à valeurs dans \mathbb{R} . On appelle **graphe** de f l'ensemble :

$$\left\{ \underbrace{(x_1, \dots, x_n)}_{\mathbf{x}}, \underbrace{x_{n+1}}_{f(\mathbf{x})} \right\} \in \mathbb{R}^{n+1} : \mathbf{x} \in \mathbb{D}_f \}$$

- Exemple :

$$f(x, y) = x^2 + y^2$$



Fonction affine

Définition.

La fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est une fonction affine de n variables s'il existe $(u_1, \dots, u_n, c) \in \mathbb{R}^{n+1}$ tel que, $\forall (x_1, \dots, x_n) \in \mathbb{R}^n$:

$$f(x_1, \dots, x_n) = u_1 x_1 + \dots + u_n x_n + c$$

Si on dénote $\mathbf{u} = (u_1, \dots, u_n)$ et $\mathbf{x} = (x_1, \dots, x_n)$ alors on a :

$$f(x_1, \dots, x_n) = f(\mathbf{x}) = \langle \mathbf{u}, \mathbf{x} \rangle + c$$

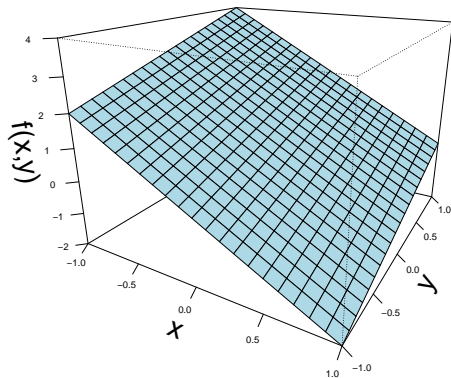
Propriété.

Le graphe de la fonction affine $f : \mathbb{R}^n \rightarrow \mathbb{R}$ avec $f(\mathbf{x}) = \langle \mathbf{u}, \mathbf{x} \rangle + c$, est un **hyperplan^a affine** de \mathbb{R}^{n+1} . Le vecteur \mathbf{u} est alors appelé **vecteur normal** à l'hyperplan.

a. Point dans \mathbb{R} , droite dans \mathbb{R}^2 , plan dans \mathbb{R}^3 .

Exemple

- Exemple : $f(x, y) = -2x + y + 1$ avec $\mathbf{u} = (-2, 1)$ et $c = 1$.

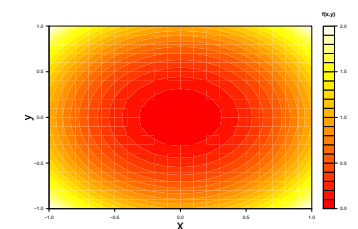
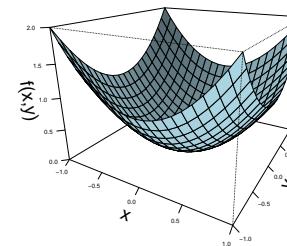


Lignes de niveaux

Définition.

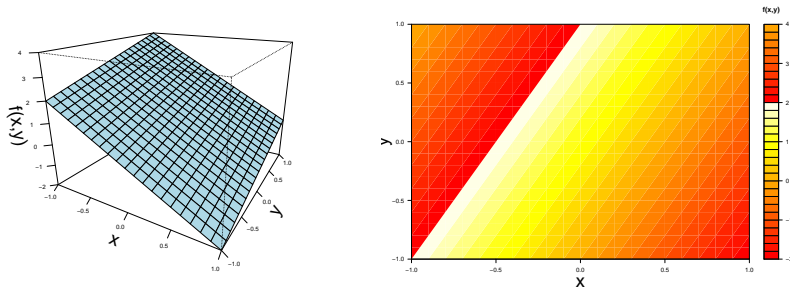
Soit une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$. On appelle **ligne de niveau c** de la fonction f l'ensemble des points suivants :

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



Exemple

- Fonction affine : $f(x, y) = -2x + y + 1$



- $\mathbb{L}_1 = \{\mathbf{x} \in \mathbb{R}^2 : -2x + y + 1 = 1\}$.
- \mathbb{L}_1 est la droite d'équation $y = 2x$.
- Remarque : les lignes de niveaux d'une fonction affine sont des hyperplans affines.

Rappel du Sommaire

5 Fonctions de plusieurs variables

- Fonctions de plusieurs variables
- Limite et continuité d'une fonction de plusieurs variables
- Dérivées partielles d'ordre 1
- Différentiabilité
- Lignes de niveau et gradients
- Approximation linéaire
- Dérivées partielles d'ordre 2
- Convexité
- Approximation quadratique

Opérations sur les fonctions de plusieurs variables

- Les opérations définies pour des fonctions réelles à valeurs réelles (cf slide 27) restent valables pour des fonctions de plusieurs variables.
- Soient f et g deux fonctions définies sur $\mathbb{D}_f, \mathbb{D}_g \subset \mathbb{R}^n$ et soit $\lambda \in \mathbb{R}$.
- La fonction $\lambda(f + g)$ est définie sur $\mathbb{D}_f \cap \mathbb{D}_g$ par :

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

- La fonction fg est définie sur $\mathbb{D}_f \cap \mathbb{D}_g$ par :

$$(fg)(\mathbf{x}) = f(\mathbf{x})g(\mathbf{x})$$

- La fonction f/g est définie sur $\mathbb{D}_f \cap \mathbb{D}_g \setminus \{\mathbf{x} \in \mathbb{D}_f : g(\mathbf{x}) = 0\}$ par :

$$\frac{f}{g}(\mathbf{x}) = \frac{f(\mathbf{x})}{g(\mathbf{x})}$$

Rappels sur les suites et leurs limites dans \mathbb{R}^n

- Afin d'introduire les limites des fonctions de plusieurs variables, il est utile d'étendre les suites et limites de suites de nombres réels au cas des suites et limites de suites de vecteurs 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 .
- 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 :

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

ou également $\mathbf{x}^{(k)} \rightarrow \mathbf{x}^*$.

- Toute suite convergente a une et une seule limite.

Rappels sur les suites et leurs limites dans \mathbb{R}^n (suite)

- 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 $B = 5$ (la suite converge vers 1).
- Toute suite convergente est bornée.
- 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}^* .

Exemples de fonctions continues

Définition.

Soit $\mathbf{x} = (x_1, \dots, x_i, \dots, x_n)$ un vecteur de \mathbb{R}^n . On appelle projections $p_i : \mathbb{R}^n \rightarrow \mathbb{R}$ les fonctions de plusieurs variables définies par, $\forall i = 1, \dots, n$:

$$p_i(\mathbf{x}) = x_i$$

Propriété.

Dans \mathbb{R}^n , les projections $p_i, \forall i = 1, \dots, n$, sont continues.

- Autres exemples de fonctions continues :
 - La fonction constante $f(\mathbf{x}) = c$ avec $c \in \mathbb{R}$.
 - La fonction affine $f(\mathbf{x}) = \langle \mathbf{u}, \mathbf{x} \rangle + c$ avec $\mathbf{u} \in \mathbb{R}^n$ et $c \in \mathbb{R}$.
 - La forme quadratique $q(\mathbf{x}) = \langle \mathbf{x}, \mathbf{x} \rangle$.
 - ...

Limite et continuité d'une fonction de plusieurs variables

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

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

alors pour désigner la limite f^* , nous utiliserons la notation suivante :

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

Définition. (Fonction continue de plusieurs variables)

La fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}^p$ est **continue** en $\mathbf{x}_0 \in \mathbb{R}^n$ 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 $\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} f(\mathbf{x}) = f(\mathbf{x}_0)$.

Composition et continuité

Théorème.

Soient f une fonction de $\mathbb{D}_f \subset \mathbb{R}^n$ dans \mathbb{R} , ϕ une fonction de $\mathbb{I} \subset \mathbb{R}$ dans \mathbb{R} et $\mathbf{x}_0 \in \mathbb{D}_f$. Supposons que $f(\mathbb{D}_f) \subset \mathbb{I}$.

Si f est continue en \mathbf{x}_0 et ϕ est continue en $f(\mathbf{x}_0)$, alors la fonction $\phi \circ f$ est continue en \mathbf{x}_0 .

- Exemple :
 - Soit $q(\mathbf{x}) = \langle \mathbf{x}, \mathbf{x} \rangle$ et $\phi(x) = \sqrt{x}$.
 - $\mathbb{D}_q = \mathbb{R}^n$ et $\mathbb{D}_\phi = \mathbb{R}^+$.
 - $q(\mathbb{D}_q) \subset \mathbb{D}_\phi$.
 - $\phi \circ q$ est donc continue sur \mathbb{R}^n .
 - Que représente $\phi \circ q$?
- On peut utiliser ce théorème afin de montrer qu'une fonction f est continue. On exprime dans ce cas f comme composition de fonctions (simples) continues.

Continuité et topologie

Théorème.

Si $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est une fonction continue alors l'image réciproque par f d'un intervalle ouvert (resp. fermé) est un ouvert (resp. fermé) de \mathbb{R}^n .

Corollaire.

Si $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est une fonction continue alors les ensembles suivants sont des ouverts de \mathbb{R}^n :

$$\{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) < c\}, \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) > c\}$$

Si on remplace ci-dessus les inégalités strictes (en rouge) par des inégalités larges alors les ensembles sont des fermés de \mathbb{R}^n .

Rappel du Sommaire

5 Fonctions de plusieurs variables

- Fonctions de plusieurs variables
- Limite et continuité d'une fonction de plusieurs variables
- Dérivées partielles d'ordre 1
- Différentiabilité
- Lignes de niveau et gradients
- Approximation linéaire
- Dérivées partielles d'ordre 2
- Convexité
- Approximation quadratique

Introduction

- Rappelons que le nombre dérivé de f , une fonction numérique d'une variable, en un point $x_0 \in \mathbb{D}_f$ est basé sur le ratio $\Delta f_{x_0}(h)/\Delta Id_{x_0}(h)$ qui quantifie de combien varie relativement f lorsque l'on passe de x_0 à $x_0 + h$ (cf slide 40).
- Dans \mathbb{R}^n il existe une **infinité de directions** vers lesquelles, nous pouvons faire varier un point \mathbf{x}_0 . Pour modéliser ces changements dans l'espace nous allons nous reposer sur les ev et la notion de base prend ici toute son importance.
- En effet, pour déplacer \mathbf{x}_0 on lui ajoute un vecteur quelconque \mathbf{h} : $\mathbf{x}_0 + \mathbf{h}$. Ces vecteurs sont représentés dans une base $\{\mathbf{e}_i\}$ et donc $\mathbf{h} = h_1\mathbf{e}_1 + \dots + h_n\mathbf{e}_n$. Intuitivement, il suffit d'étudier les variations suivant chaque \mathbf{e}_i pour avoir les variations suivant \mathbf{h} quelconque.
- Pour formaliser cette intuition, nous introduisons dans ce qui suit les n **fonctions partielles** associées à une fonction de n variables. Celles-ci sont des fonctions numériques d'une seule variable : on fait varier une composante et on suppose les autres constantes.

Fonctions partielles

Définition. (Fonctions partielles)

Soit $f : \mathbb{D}_f \rightarrow \mathbb{R}$ où \mathbb{D}_f est un ouvert de \mathbb{R}^n et $\mathbf{a} = (a_1, \dots, a_n) \in \mathbb{D}_f$. Pour tout $i = 1, \dots, n$ on définit les applications partielles de f en \mathbf{a} comme suit :

$$f_{\mathbf{a}}^i : \mathbb{D}_{f_{\mathbf{a}}^i} \rightarrow \mathbb{R} \\ x \rightarrow f(a_1, \dots, a_{i-1}, x, a_{i+1}, \dots, a_n)$$

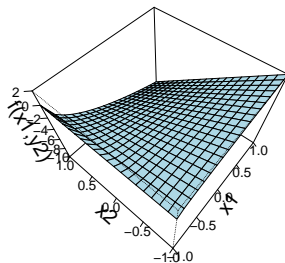
où $\mathbb{D}_{f_{\mathbf{a}}^i} = \{x \in \mathbb{R} : (a_1, \dots, a_{i-1}, x, a_{i+1}, \dots, a_n) \in \mathbb{D}_f\}$.

- Les $f_{\mathbf{a}}^i$ sont des fonctions réelles ($\mathbb{D}_{f_{\mathbf{a}}^i} \subset \mathbb{R}$) à valeurs réelles.
- Exemple :
 - ▶ Soit $f(x_1, x_2) = (\sqrt{2x_1 + 2})(-3x_2 - 2)$ et $\mathbf{a} = (0, 0)$.
 - ▶ Les applications partielles de f en \mathbf{a} sont :

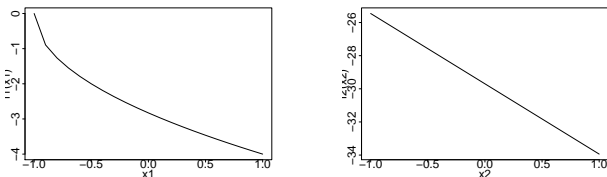
$$f_{\mathbf{a}}^1(x_1) = -2\sqrt{2x_1 + 2} \text{ et } f_{\mathbf{a}}^2(x_2) = \sqrt{2}(-3x_2 - 2)$$

Illustration

- $f(x_1, x_2) = (\sqrt{2x_1 + 2})(-3x_2 - 2)$ sur $[-1, 1]^2$.



- Fonctions partielles de f en $\mathbf{a} = (0, 0)$ sur $[-1, 1]$.



Dérivées partielles d'ordre 1

- Les **dérivées partielles d'ordre 1** d'une fonction de plusieurs variables sont les **dérivées des fonctions partielles** vues précédemment.
- Les fonctions partielles étant des fonctions numériques réelles et à valeurs réelles, les dérivées partielles reviennent à étudier les dérivées des différentes fonctions partielles.

Définition.

Soit $f : \mathbb{D}_f \rightarrow \mathbb{R}$ où \mathbb{D}_f un ouvert de \mathbb{R}^n . Si en tout point $\mathbf{x} = (x_1, \dots, x_i, \dots, x_n)$ de \mathbb{D}_f , f possède des dérivées partielles par rapport à x_i pour tout $i = 1, \dots, n$, on note alors $\frac{\partial f}{\partial x_i}$ l'application :

$$\begin{cases} f : \mathbb{D}_f \rightarrow \mathbb{R} \\ \mathbf{x} \rightarrow \frac{\partial f}{\partial x_i}(\mathbf{x}) \end{cases}$$

Dérivées partielles d'ordre 1 (suite)

- On considère $\mathbb{R}^n \supset \mathbb{D}_f$ tel un ev de base canonique $\{\mathbf{e}_i\}$.
- Soit $\mathbf{a} = (a_1, \dots, a_i, \dots, a_n)$ un point de \mathbb{R}^n .
- On peut faire varier \mathbf{a} en considérant une seule de ses composantes i . On a dans ce cas l'expression suivante :

$$(a_1, \dots, a_{i-1}, \mathbf{a}_i + h, a_{i+1}, \dots, a_n) = \mathbf{a} + h\mathbf{e}_i$$

où $h \in \mathbb{R}$.

- Les dérivées partielles de f en \mathbf{a} peuvent être interprétées de la façon suivante en utilisant la définition des nombres dérivés vue en slide 40 :

$$\frac{\partial f}{\partial x_i}(\mathbf{a}) = \lim_{h \rightarrow 0} \frac{f(\mathbf{a} + h\mathbf{e}_i) - f(\mathbf{a})}{h}$$

- Seule la composante i de \mathbf{a} varie. Les autres composantes sont considérées comme constantes. Ainsi, $\frac{\partial f}{\partial x_i}(\mathbf{a}) = (f_{\mathbf{a}}^i)'(a_i)$.

Exemples

- Soit la fonction affine $f(\mathbf{x}) = \langle \mathbf{u}, \mathbf{x} \rangle + c$ définie sur \mathbb{R}^n .
 $\forall \mathbf{x} \in \mathbb{R}^n$, f possède les dérivées partielles suivantes, $\forall i = 1, \dots, n$:

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

- Soit la fonction $n(\mathbf{x}) = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$ définie sur \mathbb{R}^n .
 $\forall \mathbf{x} \in \mathbb{R}^n$, n possède les dérivées partielles suivantes, $\forall i = 1, \dots, n$:

$$\frac{\partial n}{\partial x_i}(\mathbf{x}) = \frac{x_i}{\sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}}$$

Classe de fonctions \mathbb{C}^1

Définition.

Soit $f : \mathbb{D}_f \rightarrow \mathbb{R}$ où \mathbb{D}_f est un ouvert de \mathbb{R}^n . On dit que f est de classe \mathbb{C}^1 si elle admet en tout point de \mathbb{D}_f des dérivées partielles d'ordre 1 et si les fonctions $\frac{\partial f}{\partial x_i}, i = 1, \dots, n$ sont continues sur \mathbb{D}_f .

- La définition donnée en slide 44 dans le cas des fonctions réelles à valeurs réelles se généralise naturellement aux fonctions de plusieurs variables pour la classe \mathbb{C}^1 .
- La généralisation de \mathbb{C}^n pour $n > 1$ demande à être précisée (cf plus loin le cas \mathbb{C}^2).

Exemples

- Soit la fonction affine $f(\mathbf{x}) = \langle \mathbf{u}, \mathbf{x} \rangle + c$ définie sur \mathbb{R}^n .

Le vecteur gradient s'écrit :

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

C'est une fonction à valeurs vectorielles qui est constante.

- Soit la fonction $n(\mathbf{x}) = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$ définie sur \mathbb{R}^{n*} .

Le vecteur gradient s'écrit :

$$\nabla n(\mathbf{x}) = \frac{\mathbf{x}}{\sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}}$$

Vecteur gradient

Définition.

Soit $f : \mathbb{D}_f \rightarrow \mathbb{R}$ où \mathbb{D}_f est un ouvert de \mathbb{R}^n et \mathbf{a} un point de \mathbb{D}_f . Si f admet en \mathbf{a} des dérivées partielles d'ordre 1 par rapport aux n variables, on appelle **gradient** de f en \mathbf{a} noté $\nabla f(\mathbf{a})$, le vecteur de \mathbb{R}^n suivant :

$$\nabla f(\mathbf{a}) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{a}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\mathbf{a}) \end{pmatrix}$$

De plus, si f est de classe \mathbb{C}^1 sur \mathbb{D}_f alors le gradient peut être vu comme une fonction à valeurs vectorielles, $\nabla f : \mathbb{D}_f \rightarrow \mathbb{R}^n$ définie par :

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

Matrice Jacobienne

Définition.

Soit $f = (f_1, \dots, f_p)$ une fonction de plusieurs variables à valeurs vectorielles de \mathbb{D}_f vers \mathbb{R}^p où \mathbb{D}_f est un ouvert de \mathbb{R}^n et f_1, \dots, f_p sont des fonctions de plusieurs variables à valeurs réelles. Si f_1, \dots, f_p admettent en $\mathbf{a} \in \mathbb{D}_f$, des dérivées partielles d'ordre 1 alors on appelle **matrice Jacobienne** de f en \mathbf{a} , notée $\mathbf{J}_f(\mathbf{a})$, la matrice de taille $(p \times n)$ définie par :

$$\mathbf{J}_f(\mathbf{a}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{a}) & \dots & \frac{\partial f_1}{\partial x_n}(\mathbf{a}) \\ \vdots & & \vdots \\ \frac{\partial f_p}{\partial x_1}(\mathbf{a}) & \dots & \frac{\partial f_p}{\partial x_n}(\mathbf{a}) \end{pmatrix} = \begin{pmatrix} \nabla f_1(\mathbf{a})^\top \\ \vdots \\ \nabla f_p(\mathbf{a})^\top \end{pmatrix}$$

- Remarque : dans le cas d'une fonction de plusieurs variables à valeurs réelles, $f : \mathbb{D}_f \subset \mathbb{R}^n \rightarrow \mathbb{R}$, on a $\mathbf{J}_f(\mathbf{a})$ de taille $(1 \times n)$ avec :

$$\mathbf{J}_f(\mathbf{a}) = \nabla f(\mathbf{a})^\top$$

Exemples

- Soit :

$$f(\mathbf{x}) = \begin{pmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ f_3(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} x^2 + y^2 \\ \ln(x^2 + y^2) \\ \exp(x^2 + y^2) \end{pmatrix}$$

- On a :

$$\mathbf{J}_f(\mathbf{x}) = \begin{pmatrix} \nabla f_1(\mathbf{x})^\top \\ \nabla f_2(\mathbf{x})^\top \\ \nabla f_3(\mathbf{x})^\top \end{pmatrix} = \begin{pmatrix} 2x & 2y \\ \frac{2x}{x^2+y^2} & \frac{2y}{x^2+y^2} \\ 2x \exp(x^2 + y^2) & 2y \exp(x^2 + y^2) \end{pmatrix}$$

- Vérifiez que si $f(\mathbf{x}) = \mathbf{Ax}$ alors $\mathbf{J}_f(\mathbf{x}) = \mathbf{A}$.

Opérations sur les dérivées partielles

- Ci-dessous on généralise les résultats donnés en slide 46 pour les dérivées partielles des fonctions de plusieurs variables.
- Soit $f : \mathbb{D}_f \rightarrow \mathbb{R}$ et $g : \mathbb{D}_g \rightarrow \mathbb{R}$ où \mathbb{D}_f et \mathbb{D}_g sont des ouverts de \mathbb{R}^n et $\lambda \in \mathbb{R}$. Si f et g possèdent des dérivées partielles d'ordre 1 en $\mathbf{a} \in \mathbb{D}_f \cap \mathbb{D}_g$ alors $\forall i = 1, \dots, n$:

- ▶ La fonction $\lambda(f + g)$ possède des dérivées partielles d'ordre 1 en \mathbf{a} et :

$$\frac{\partial(\lambda(f + g))}{\partial x_i}(\mathbf{a}) = \lambda \frac{\partial f}{\partial x_i}(\mathbf{a}) + \lambda \frac{\partial g}{\partial x_i}(\mathbf{a})$$

- ▶ La fonction fg possède des dérivées partielles d'ordre 1 en \mathbf{a} et :

$$\frac{\partial(fg)}{\partial x_i}(\mathbf{a}) = g(\mathbf{a}) \frac{\partial f}{\partial x_i}(\mathbf{a}) + f(\mathbf{a}) \frac{\partial g}{\partial x_i}(\mathbf{a})$$

- ▶ La fonction f/g possède des dérivées partielles d'ordre 1 en \mathbf{a} à condition que $g(\mathbf{a}) \neq 0$ et :

$$\frac{\partial}{\partial x_i} \left(\frac{f}{g} \right) (\mathbf{a}) = \frac{1}{g^2(\mathbf{a})} \left(g(\mathbf{a}) \frac{\partial f}{\partial x_i}(\mathbf{a}) - f(\mathbf{a}) \frac{\partial g}{\partial x_i}(\mathbf{a}) \right)$$

Composition et dérivation de $\phi \circ f : \mathbb{R}^n \xrightarrow{f} \mathbb{R} \xrightarrow{\phi} \mathbb{R}$

Propriété.

Soit $f : \mathbb{D}_f \rightarrow \mathbb{R}$ où \mathbb{D}_f est un ouvert de \mathbb{R}^n et $\phi : \mathbb{D}_\phi \rightarrow \mathbb{R}$ où \mathbb{D}_ϕ est un intervalle ouvert de \mathbb{R} et \mathbf{a} un point de \mathbb{D}_f . Si f et ϕ sont de classe \mathbb{C}^1 et que $f(\mathbb{D}_f) \subset \mathbb{D}_\phi$, alors la fonction composée $\phi \circ f : \mathbb{D}_f \rightarrow \mathbb{R}$ est de classe \mathbb{C}^1 sur \mathbb{D}_f et $\forall i = 1, \dots, n$:

$$\frac{\partial}{\partial x_i} (\phi \circ f)(\mathbf{a}) = (\phi' \circ f)(\mathbf{a}) \frac{\partial f}{\partial x_i}(\mathbf{a}) = \phi'(f(\mathbf{a})) \frac{\partial f}{\partial x_i}(\mathbf{a})$$

Le gradient de $\phi \circ f$ s'écrit donc :

$$\nabla(\phi \circ f) = (\phi' \circ f) \nabla f$$

- Cette propriété utilise pour chaque i , le résultat vu en slide 46 pour les fonctions réelles, en observant que $\frac{\partial}{\partial x_i} (\phi \circ f)(\mathbf{a}) = (\phi \circ f'_i)'(a_i)$.

Composition et dérivation de $f \circ g : \mathbb{R} \xrightarrow{g} \mathbb{R}^n \xrightarrow{f} \mathbb{R}$

Propriété.

Soit $f : \mathbb{D}_f \rightarrow \mathbb{R}$ une fonction de classe \mathbb{C}^1 sur \mathbb{D}_f un ouvert de \mathbb{R}^n . Soit $g : \mathbb{D}_g \rightarrow \mathbb{R}^n$ une fonction vectorielle sur \mathbb{D}_g un ouvert de \mathbb{R} telle que $g = (g_1, \dots, g_n)$ où chaque fonction $g_i, i = 1, \dots, n$, est dérivable sur \mathbb{D}_g . La fonction composée (réelle et à valeurs réelles) $f \circ g : \mathbb{D}_g \rightarrow \mathbb{R}$ est dérivable sur \mathbb{D}_g et pour tout $t \in \mathbb{D}_g$ on a :

$$\begin{aligned} \frac{d(f \circ g)}{dt}(t) &= \sum_{i=1}^n \frac{\partial f}{\partial x_i}(g(t)) g'_i(t) \\ &= \nabla f(g(t))^\top \mathbf{J}_g(t) \\ &= \langle \nabla f(g(t)), \mathbf{J}_g(t) \rangle \end{aligned}$$

où $\mathbf{J}_g(t) = (g'_1(t), \dots, g'_n(t))$ est $(n \times 1)$ ("matrice" Jacobienne de g).

Règles de dérivation (extension aux cas de plusieurs variables)

- Dans le cas des fonctions de n variables, on se ramène au cas de n fonctions partielles et on raisonne "unitairement" pour tout $i = 1, \dots, n$.
- Les dérivées partielles sont ainsi les fonctions dérivées des fonctions partielles. Dans ce cas, on considère qu'une composante varie et les autres sont constantes.
- Les règles de dérivation vues en slide 47 s'appliquent donc aux dérivées partielles.

Dérivée directionnelle (suite)

Théorème.

Soit $f : \mathbb{D}_f \rightarrow \mathbb{R}$ une fonction de classe \mathbb{C}^1 sur un ouvert de \mathbb{R}^n , \mathbf{a} un point de \mathbb{D}_f et \mathbf{d} un vecteur de norme unitaire de \mathbb{R}^n . Alors, la fonction f possède une dérivée dans la direction \mathbf{d} qui est égale à :

$$D_{\mathbf{d}}f(\mathbf{a}) = \langle \nabla f(\mathbf{a}), \mathbf{d} \rangle$$

Démonstration.

Soit $\phi(h) = f(\mathbf{a} + h\mathbf{d}) = f(g(h))$ avec $g(h) = \mathbf{a} + h\mathbf{d}$. En appliquant la règle de dérivation d'une fonction composée on a

$\phi'(h) = (f \circ g)'(h) = \nabla f(g(h))^{\top} \mathbf{J}_g(h)$. En $h = 0$ on a alors :

$$\phi'(0) = \nabla f(g(0))^{\top} \mathbf{J}_g(0) = \nabla f(\mathbf{a})^{\top} \mathbf{d}$$

□

Dérivée directionnelle

Définition.

Soit $f : \mathbb{D}_f \rightarrow \mathbb{R}$ où \mathbb{D}_f est un ouvert de \mathbb{R}^n , \mathbf{a} un point de \mathbb{D}_f et \mathbf{d} un vecteur de norme unitaire. Si la fonction $\phi : \mathbb{R} \rightarrow \mathbb{R}$ définie par $\phi(h) = f(\mathbf{a} + h\mathbf{d})$ est dérivable en $h = 0$, le nombre dérivé est appelé dérivé de f en \mathbf{a} dans la direction de \mathbf{d} et est noté $D_{\mathbf{d}}f(\mathbf{a})$

$$\begin{aligned} D_{\mathbf{d}}f(\mathbf{a}) &= \phi'(0) \\ &= \lim_{t \rightarrow 0} \frac{\phi(0+t) - \phi(0)}{t} \\ &= \lim_{t \rightarrow 0} \frac{f(\mathbf{a} + t\mathbf{d}) - f(\mathbf{a})}{t} \end{aligned}$$

- Les dérivées partielles vues précédemment sont des cas particuliers des dérivées directionnelles avec $\mathbf{d} = \mathbf{e}_i$. On a en fait $\frac{\partial f}{\partial x_i} = D_{\mathbf{e}_i}f$.

Exemple

- Soit une fonction affine :

$$\begin{aligned} f(\mathbf{x}) &= \langle \mathbf{u}, \mathbf{x} \rangle + c \\ &= u_1x_1 + \dots + u_nx_n + c \end{aligned}$$

- Soit $\mathbf{a} = (a_1, \dots, a_n)$ et $\mathbf{d} = (d_1, \dots, d_n)$ des vecteurs unitaires de \mathbb{R}^n .
- On a $\forall h \in \mathbb{R}$:

$$\begin{aligned} \phi(h) &= f(\mathbf{a} + h\mathbf{d}) \\ &= \langle \mathbf{u}, \mathbf{a} + h\mathbf{d} \rangle + c \\ &= u_1(a_1 + hd_1) + \dots + u_n(a_n + hd_n) + c \end{aligned}$$

- On en déduit :

$$D_{\mathbf{d}}f(\mathbf{a}) = u_1d_1 + \dots + u_nd_n$$

- On voit par ailleurs que :

$$D_{\mathbf{d}}f(\mathbf{a}) = \langle \nabla f(\mathbf{a}), \mathbf{d} \rangle$$

Rappel du Sommaire

5 Fonctions de plusieurs variables

- Fonctions de plusieurs variables
- Limite et continuité d'une fonction de plusieurs variables
- Dérivées partielles d'ordre 1
- Différentiabilité
- Lignes de niveau et gradients
- Approximation linéaire
- Dérivées partielles d'ordre 2
- Convexité
- Approximation quadratique

Différentiabilité (suite)

Définition. (Fonction de plusieurs variables différentiable)

Soit $f : \mathbb{D}_f \rightarrow \mathbb{R}$ où \mathbb{D}_f est un ouvert de \mathbb{R}^n et \mathbf{a} un point de \mathbb{D}_f . On suppose que les vecteurs de \mathbb{R}^n sont représentés dans la base canonique $\{\mathbf{e}_i\}$. On dit que f est **différentiable** en \mathbf{a} s'il existe une fonction linéaire g telle que :

$$\lim_{\mathbf{x} \rightarrow \mathbf{a}} \frac{f(\mathbf{x}) - \overbrace{(f(\mathbf{a}) + g(\mathbf{x} - \mathbf{a}))}^{\phi(\mathbf{x})}}{\|\mathbf{x} - \mathbf{a}\|} = 0$$

- g est linéaire et s'écrit $g(\mathbf{x}) = \langle \mathbf{u}, \mathbf{x} \rangle = u_1 x_1 + \dots + u_n x_n = \mathbf{u}^\top \mathbf{x}$.
- f est donc différentiable s'il existe $\mathbf{u} = (u_1, \dots, u_n) \in \mathbb{R}^n$ tel que :

$$\lim_{\mathbf{x} \rightarrow \mathbf{a}} \frac{f(\mathbf{x}) - (f(\mathbf{a}) + \mathbf{u}^\top (\mathbf{x} - \mathbf{a}))}{\|\mathbf{x} - \mathbf{a}\|} = 0$$

Différentiabilité

- Le calcul différentiel est basé sur l'idée que l'on peut approximer une quelconque fonction par une fonction affine. Ce que nous voyons ci-dessous est la généralisation des résultats vus en slide 52 au cas de fonctions de plusieurs variables.
- Soit $f : \mathbb{D}_f \rightarrow \mathbb{R}$ où \mathbb{D}_f est un ouvert de \mathbb{R}^n . Soit un point $\mathbf{a} \in \mathbb{D}_f$. On souhaite approximer f par une fonction affine ϕ dans le voisinage de \mathbf{a} .
- Une première condition est que f et la fonction affine soient identiques en \mathbf{a} :

$$\phi(\mathbf{a}) = f(\mathbf{a})$$

- Une deuxième condition est que $\phi(\mathbf{x})$ approche $f(\mathbf{x})$ plus rapidement que \mathbf{x} approche \mathbf{a} . Ceci s'exprime par :

$$\lim_{\mathbf{x} \rightarrow \mathbf{a}} \frac{f(\mathbf{x}) - \phi(\mathbf{x})}{\|\mathbf{x} - \mathbf{a}\|} = 0$$

Définition de la différentielle

- On montre en fait que si f est différentiable en \mathbf{a} alors le vecteur \mathbf{u} définissant la fonction affine $\phi(\mathbf{x}) = f(\mathbf{a}) + \mathbf{u}^\top (\mathbf{x} - \mathbf{a})$ est donnée de façon unique par les dérivées partielles, $\forall i$:

$$u_i = \frac{\partial f}{\partial x_i}(\mathbf{a})$$

Définition. (Différentielle)

Soit $f : \mathbb{D}_f \rightarrow \mathbb{R}$ où \mathbb{D}_f est un ouvert de \mathbb{R}^n . Si f est différentiable en $\mathbf{a} \in \mathbb{D}_f$ alors l'application linéaire $df_{\mathbf{a}} : \mathbb{R}^n \rightarrow \mathbb{R}$ définie par :

$$df_{\mathbf{a}}(\mathbf{h}) = \frac{\partial f}{\partial x_1}(\mathbf{a})h_1 + \dots + \frac{\partial f}{\partial x_n}(\mathbf{a})h_n$$

est appelée **différentielle** de f en \mathbf{a} .

Dérivée directionnelle et accroissement différentiel

- Le lien entre différentielle et dérivée directionnelle en $\mathbf{a} \in \mathbb{D}_f$ est clair :

$$df_{\mathbf{a}}(\mathbf{d}) = \nabla f(\mathbf{a})^T \mathbf{d} = \langle \nabla f(\mathbf{a}), \mathbf{d} \rangle = D_{\mathbf{d}}f(\mathbf{a})$$

- Si on passe de \mathbf{a} à $\mathbf{a} + t\mathbf{d}$ pour t suffisamment petit, alors la différentielle donne une approximation de l'accroissement absolu de f (cf slide 57 dans le cas d'une variable).
- Si $\|\mathbf{d}\| = 1$ alors nous pouvons interpréter $\langle \nabla f(\mathbf{a}), \mathbf{d} \rangle$ comme étant une sorte de “**taux**” **d'accroissement différentiel** de la fonction f dans la direction du vecteur \mathbf{d} .

Rappel du Sommaire

5 Fonctions de plusieurs variables

- Fonctions de plusieurs variables
- Limite et continuité d'une fonction de plusieurs variables
- Dérivées partielles d'ordre 1
- Différentiabilité
- Lignes de niveau et gradients**
- Approximation linéaire
- Dérivées partielles d'ordre 2
- Convexité
- Approximation quadratique

Exemple

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

$$\begin{aligned} D_{\mathbf{d}}f(\mathbf{x}) &= df_{\mathbf{x}}(\mathbf{d}) \\ &= \nabla f(\mathbf{x})^T \mathbf{d} \\ &= 2(x_1 \ x_2) \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= \sqrt{2}(x_1 + x_2) \end{aligned}$$

- Vérifiez que $\|\mathbf{d}\| = 1$ puis vérifiez que l'accroissement différentiel de f en $\mathbf{a} = (1, 1)$ dans la direction de \mathbf{d} vaut $2\sqrt{2}$.

Relations entre lignes de niveau et gradient

- Nous étudions les relations entre lignes de niveau et gradients qui nous seront utiles ultérieurement en optimisation.
- Prenons $\mathbf{x}_0 \in \mathbb{D}_f$ un point appartenant à \mathbb{L}_c de $f : f(\mathbf{x}_0) = c$.
- Supposons qu'il existe une courbe γ contenue dans \mathbb{L}_c représentée par une fonction de classe \mathbb{C}^1 , $g : \mathbb{R} \rightarrow \mathbb{R}^n$.
- Supposons que en $t_0 \in \mathbb{R}$, $g(t_0) = \mathbf{x}_0$ et que $\mathbf{J}_g(t_0) = \mathbf{u} \neq \mathbf{0}$ de telle sorte que \mathbf{u} est le **vecteur tangent à γ en \mathbf{x}_0** .
- En appliquant la règle de dérivation des fonctions composées à la fonction $\phi(t) = f(g(t))$ en t_0 , on obtient :

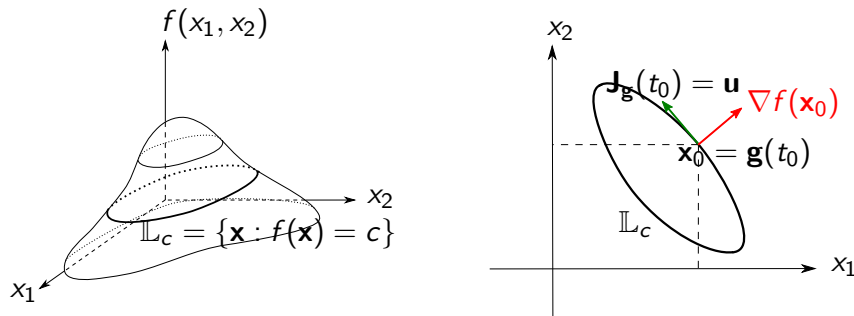
$$\phi'(t_0) = \nabla f(g(t_0))^T \mathbf{J}_g(t_0) = \nabla f(\mathbf{x}_0)^T \mathbf{u}$$

- Comme de plus γ est contenue dans \mathbb{L}_c , $\phi(t) = f(g(t)) = c$ une constante, ce qui implique que $\phi'(t) = 0$ et donc :

$$\nabla f(\mathbf{x}_0)^T \mathbf{u} = \langle \nabla f(\mathbf{x}_0), \mathbf{u} \rangle = 0$$

Relations entre lignes de niveau et gradient (suite)

- $\nabla f(\mathbf{x}_0)$ est le vecteur gradient en \mathbf{x}_0 .
- $\mathbf{u} = \mathbf{J}_g(t_0)$ est le vecteur tangent à γ en \mathbf{x}_0 .
- Comme $\langle \nabla f(\mathbf{x}_0), \mathbf{u} \rangle = 0$ donc **tout vecteur gradient en un point \mathbf{x}_0 est orthogonal au vecteur tangent à la ligne de niveau à laquelle appartient \mathbf{x}_0 !**
- Illustration :



Rappel du Sommaire

5 Fonctions de plusieurs variables

- Fonctions de plusieurs variables
- Limite et continuité d'une fonction de plusieurs variables
- Dérivées partielles d'ordre 1
- Différentiabilité
- Lignes de niveau et gradients
- Approximation linéaire
- Dérivées partielles d'ordre 2
- Convexité
- Approximation quadratique

Approximation affine

Définition. (Approximation affine)

Soit $f : \mathbb{D}_f \rightarrow \mathbb{R}$ où \mathbb{D}_f est un ouvert de \mathbb{R}^n . f possède une approximation affine en \mathbf{a} , s'il existe une forme linéaire $g \in \mathcal{L}(\mathbb{R}^n, \mathbb{R})$ telle que $\forall \mathbf{h}, \mathbf{a} + \mathbf{h} \in \mathbb{D}_f$:

$$f(\mathbf{a} + \mathbf{h}) = f(\mathbf{a}) + g(\mathbf{h}) + o(\|\mathbf{h}\|)$$

où $o(\|\mathbf{h}\|)$ représente un terme négligeable qui converge vers 0 plus vite que $\|\mathbf{h}\|$.

Propriété. (Approximation affine et gradient)

Soit $f : \mathbb{D}_f \rightarrow \mathbb{R}$ où \mathbb{D}_f est un ouvert de \mathbb{R}^n . Si f est de classe \mathcal{C}^1 alors elle possède une approximation affine en $\mathbf{a} \in \mathbb{D}_f$. De plus, on a nécessairement :

$$f(\mathbf{a} + \mathbf{h}) = f(\mathbf{a}) + \langle \nabla f(\mathbf{a}), \mathbf{h} \rangle + o(\|\mathbf{h}\|)$$

Meilleure approximation affine, hyperplan tangent

Définition.

Soit $f : \mathbb{D}_f \rightarrow \mathbb{R}$ où \mathbb{D}_f est un ouvert de \mathbb{R}^n . Si f possède une approximation affine en $\mathbf{a} \in \mathbb{D}_f$ alors la fonction :

$$\phi(\mathbf{x}) = f(\mathbf{a}) + \langle \nabla f(\mathbf{a}), \mathbf{x} - \mathbf{a} \rangle$$

est appelée **meilleure approximation affine** de f au voisinage de \mathbf{a} . L'hyperplan de \mathbb{R}^{n+1} d'équation :

$$\underbrace{x_{n+1}}_{\phi(\mathbf{x})} = f(\mathbf{a}) + \sum_{i=1}^n \frac{\partial f}{\partial x_i}(\mathbf{a})(x_i - a_i)$$

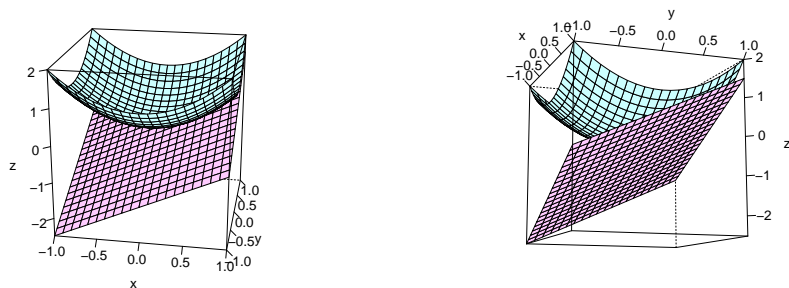
est appelé **hyperplan tangent** au graphe de f en \mathbf{a} .

Exemple

- Soit $f(\mathbf{x}) = x^2 + y^2$ définie sur \mathbb{R}^2 .
- On a $\nabla f(\mathbf{x}) = 2\mathbf{x}$ et en $\mathbf{a} = (1/2, 1/2)$, l'approximation affine est :

$$\phi(\mathbf{x}) = x + y + \frac{1}{2}$$

- Illustration (sous plusieurs angles de vue) :



Rappel du Sommaire

5 Fonctions de plusieurs variables

- Fonctions de plusieurs variables
- Limite et continuité d'une fonction de plusieurs variables
- Dérivées partielles d'ordre 1
- Différentiabilité
- Lignes de niveau et gradients
- Approximation linéaire
- Dérivées partielles d'ordre 2
- Convexité
- Approximation quadratique

Dérivées partielles d'ordre 2

Définition.

Soit $f : \mathbb{D}_f \rightarrow \mathbb{R}$ où \mathbb{D}_f est un ouvert de \mathbb{R}^n et \mathbf{a} un point de \mathbb{D}_f .

Supposons que pour tout $i = 1, \dots, n$, la dérivée partielle d'ordre 1, $\frac{\partial f}{\partial x_i}$, soit définie sur \mathbb{D}_f et qu'elle possède une dérivée partielle par rapport à x_j en \mathbf{a} alors celle-ci est notée $\frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a})$.

Si en tout point de \mathbb{D}_f , les fonctions $\frac{\partial f}{\partial x_i}$, $i = 1, \dots, n$ possèdent une dérivée partielle par rapport à x_j , $j = 1, \dots, n$, on définit les fonctions de plusieurs variables à valeurs réelles suivantes :

$$\begin{cases} \frac{\partial^2 f}{\partial x_j \partial x_i} : \mathbb{D}_f \rightarrow \mathbb{R} \\ \mathbf{x} \rightarrow \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x}) \end{cases}$$

Les fonctions $\frac{\partial^2 f}{\partial x_j \partial x_i}$ si elles existent sont appelées **dérivées partielles d'ordre 2**.

Classe de fonctions \mathbb{C}^2 et Théorème de Schwarz

Définition.

Soit \mathbb{D}_f un ouvert de \mathbb{R}^n , $f : \mathbb{D}_f \rightarrow \mathbb{R}$. On dit que f est de classe \mathbb{C}^2 sur \mathbb{D}_f si elle admet en tout point de \mathbb{D}_f des dérivées partielles d'ordre 2 et si les fonctions $\frac{\partial^2 f}{\partial x_i \partial x_j}$, $\forall i, j = 1, \dots, n$, sont continues sur \mathbb{D}_f .

Théorème. (Théorème de Schwarz)

Soit \mathbb{D}_f un ouvert de \mathbb{R}^n , $f : \mathbb{D}_f \rightarrow \mathbb{R}$ une application de classe \mathbb{C}^2 . Alors, pour tout $i, j = 1, \dots, n$;

$$\frac{\partial^2 f}{\partial x_j \partial x_i} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

Matrice hessienne

Définition.

Soit \mathbb{D}_f un ouvert de \mathbb{R}^n , $f : \mathbb{D}_f \rightarrow \mathbb{R}$ une application de classe \mathbb{C}^2 . On appelle **matrice hessienne** (ou hessien) de f en un point $\mathbf{a} \in \mathbb{D}_f$, la matrice notée $\nabla^2 f(\mathbf{a})$ de terme général :

$$\nabla^2 f(\mathbf{a}) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a}) \right)_{i,j=1,\dots,n}$$

- En raison du théorème de Schwarz on remarquera que la matrice hessienne d'une fonction de classe \mathbb{C}^2 est symétrique.

Rappel du Sommaire

5 Fonctions de plusieurs variables

- Fonctions de plusieurs variables
- Limite et continuité d'une fonction de plusieurs variables
- Dérivées partielles d'ordre 1
- Différentiabilité
- Lignes de niveau et gradients
- Approximation linéaire
- Dérivées partielles d'ordre 2
- Convexité
- Approximation quadratique

Exemple

- Soit $f(x_1, x_2) = 5x_1 + 8x_2 + x_1x_2 - x_1^2 - 2x_2^2$, de classe \mathbb{C}^2 sur \mathbb{R}^2 .
- On a le gradient qui vaut :

$$\nabla f(\mathbf{x}) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}) \\ \frac{\partial f}{\partial x_2}(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} 5 + x_2 - 2x_1 \\ 8 + x_1 - 4x_2 \end{pmatrix}$$

- La matrice hessienne vaut :

$$\nabla^2 f(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{x}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_2^2}(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} -2 & 1 \\ 1 & -4 \end{pmatrix}$$

- Vérifiez que la matrice hessienne est dn pour tout $\mathbf{x} \in \mathbb{R}^2$.

Fonction convexe, concave

- Les définitions données en slide 69 dans le cas d'une fonction réelle à valeurs réelles se généralisent naturellement au cas des fonctions de plusieurs variables.

Définition. (Fonction (strictement) convexe, (strictement) concave)

Soit f une fonction définie sur $\mathbb{D}_f \subset \mathbb{R}^n$. On dit que f est **convexe** sur \mathbb{D}_f si $\forall \mathbf{x}, \mathbf{y} \in \mathbb{D}_f$ et $\forall \alpha \in [0, 1]$ on a :

$$f(\alpha \mathbf{x} + (1 - \alpha)\mathbf{y}) \leq \alpha f(\mathbf{x}) + (1 - \alpha)f(\mathbf{y})$$

Si l'inégalité large précédente (en rouge) est remplacée :

- par une inégalité stricte $<$ alors f est **strictement convexe**,
- par une inégalité large \geq alors f est **concave**,
- par une inégalité stricte $>$ alors f est **strictement concave**.

Point de vue géométrique de la convexité

Définition. (Epigraphe)

L'**épigraphe** de la fonction $f : \mathbb{D}_f \rightarrow \mathbb{R}$, dénoté $\text{epi}(f)$, est l'ensemble des points de $\mathbb{D}_f \times \mathbb{R} \subset \mathbb{R}^{n+1}$ donné par :

$$\{(\mathbf{x}, x_{n+1}), \mathbf{x} \in \mathbb{D}_f, x_{n+1} \in \mathbb{R}, x_{n+1} \geq f(\mathbf{x})\}$$

Définition. (Fonction convexe)

Une fonction f définie sur $\mathbb{D}_f \subset \mathbb{R}^n$ est **convexe** si son épigraphe est un ensemble convexe.

Définition. (Fonction concave)

Une fonction f définie sur $\mathbb{D}_f \subset \mathbb{R}^n$ est **concave** si sa fonction opposée $-f$ est une fonction convexe.

Matrice hessienne et convexité

Propriété.

Soit f une fonction de classe \mathbb{C}^2 définie sur \mathbb{D}_f un ouvert de \mathbb{R}^n alors :

- f est **convexe** sur \mathbb{D}_f si sa matrice hessienne est **sdp** pour tout $\mathbf{x} \in \mathbb{D}_f$:

$$\forall \mathbf{x} \in \mathbb{D}_f, \forall \mathbf{y} \in \mathbb{D}_f, \mathbf{y}^\top \nabla^2 f(\mathbf{x}) \mathbf{y} \geq 0$$

- f est **concave** sur \mathbb{D}_f si sa matrice hessienne est **sdn** pour tout $\mathbf{x} \in \mathbb{D}_f$:

$$\forall \mathbf{x} \in \mathbb{D}_f, \forall \mathbf{y} \in \mathbb{D}_f, \mathbf{y}^\top \nabla^2 f(\mathbf{x}) \mathbf{y} \leq 0$$

Point de vue géométrique de la convexité (suite)

Définition.

Soit f une fonction de classe \mathbb{C}^1 définie sur \mathbb{D}_f un ouvert de \mathbb{R}^n alors :

- f est convexe sur \mathbb{D}_f si son épigraphe sur \mathbb{D}_f est au-dessus de tous ses hyperplans :

$$\forall \mathbf{x} \in \mathbb{D}_f, \forall \mathbf{y} \in \mathbb{D}_f, f(\mathbf{x}) \geq f(\mathbf{y}) + (\mathbf{x} - \mathbf{y})^\top \nabla f(\mathbf{y})$$

- f est concave sur \mathbb{D}_f si son épigraphe sur \mathbb{D}_f est au-dessous de tous ses hyperplans :

$$\forall \mathbf{x} \in \mathbb{D}_f, \forall \mathbf{y} \in \mathbb{D}_f, f(\mathbf{x}) \leq f(\mathbf{y}) + (\mathbf{x} - \mathbf{y})^\top \nabla f(\mathbf{y})$$

- On remarquera en effet que :

$$\phi(\mathbf{y}) = f(\mathbf{y}) + (\mathbf{x} - \mathbf{y})^\top \nabla f(\mathbf{y}) = f(\mathbf{y}) + \langle \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle$$

est l'équation de l'hyperplan tangent au graphe de f en \mathbf{x} .

Exemples

- Soit $f(\mathbf{x}) = \frac{1}{6}x^3 + \frac{1}{6}y^3 - xy$ définie sur \mathbb{R}^2 .
- On a $\nabla f(\mathbf{x}) = (\frac{1}{2}x^2 - y, \frac{1}{2}y^2 - x)$ et le hessien vaut :

$$\nabla^2 f(\mathbf{x}) = \begin{pmatrix} x & -1 \\ -1 & y \end{pmatrix}$$

- Les mineurs principaux en fonction de (x, y) sont :

$$x, y, xy - 1$$

- Vérifiez qu'en $\mathbf{a} = (1, 1)$, f est convexe.
- Vérifiez qu'en $\mathbf{b} = (-1, -1)$, f est concave.

Rappel du Sommaire

5 Fonctions de plusieurs variables

- Fonctions de plusieurs variables
- Limite et continuité d'une fonction de plusieurs variables
- Dérivées partielles d'ordre 1
- Différentiabilité
- Lignes de niveau et gradients
- Approximation linéaire
- Dérivées partielles d'ordre 2
- Convexité
- Approximation quadratique

Approximation quadratique (suite)

- On posant $\mathbf{x} = \mathbf{a} + \mathbf{h}$ on pourra également écrire que l'approximation quadratique de f au voisinage de \mathbf{a} vaut :

$$\phi(\mathbf{x}) = f(\mathbf{a}) + \langle \nabla f(\mathbf{a}), \mathbf{x} - \mathbf{a} \rangle + \frac{1}{2}(\mathbf{x} - \mathbf{a})^\top \nabla^2 f(\mathbf{a})(\mathbf{x} - \mathbf{a})$$

- Exemple :

- ▶ Soit $f(\mathbf{x}) = \frac{1}{6}x^3 + \frac{1}{6}y^3 - xy$ définie sur \mathbb{R}^2 .
- ▶ L'approximation affine en $\mathbf{a} = (1, 1)$ vaut :

$$\phi(\mathbf{x}) = \frac{1}{3} - \frac{1}{2}(x + y)$$

- ▶ L'approximation quadratique en $\mathbf{a} = (1, 1)$ vaut :

$$\phi(\mathbf{x}) = \frac{1}{3} - \frac{1}{2}(x + y) + \frac{1}{2}((x - 1)^2 + (y - 1)^2 - 2(x - 1)(y - 1))$$

Approximation quadratique

Propriété. (Approximation quadratique)

Soit $f : \mathbb{D}_f \rightarrow \mathbb{R}$ où \mathbb{D}_f est un ouvert de \mathbb{R}^n . Si f est de classe \mathcal{C}^2 alors elle possède une approximation quadratique en $\mathbf{a} \in \mathbb{D}_f$. De plus, on a nécessairement :

$$f(\mathbf{a} + \mathbf{h}) = \underbrace{f(\mathbf{a})}_{\text{constant}} + \underbrace{\langle \nabla f(\mathbf{a}), \mathbf{h} \rangle}_{\text{linéaire}} + \underbrace{\frac{1}{2} \mathbf{h}^\top \nabla^2 f(\mathbf{a}) \mathbf{h}}_{\text{quadratique}} + \underbrace{o(\|\mathbf{h}\|^2)}_{\text{reste}}$$

où $o(\|\mathbf{h}\|^2)$ représente un terme négligeable qui converge vers 0 plus vite que $\|\mathbf{h}\|^2$.

- Ce résultat correspond à la formule de Taylor-Young développée à l'ordre 2 pour des fonctions de plusieurs variables à valeurs réelles.

Rappel du Sommaire

- 1 Quelques méthodes statistiques du point de vue de l'optimisation
- 2 Fonctions d'une variable
- 3 Optimisation unidimensionnelle
- 4 Algèbre linéaire
- 5 Fonctions de plusieurs variables
- 6 Optimisation multidimensionnelle

Rappel du Sommaire

6 Optimisation multidimensionnelle

- Définitions de base
- Existence et recherche d'extrema
- Caractérisation des extrema d'une fonction
- Algorithmes d'optimisation numérique multidimensionnelle

Minimiseur local et global

- On distingue deux types de minimiseurs.

Définition. (Minimiseur local)

Soit $f : \mathbb{D}_f \rightarrow \mathbb{R}$ où $\mathbb{D}_f \subset \mathbb{R}^n$ et soit $\mathbb{S} \subset \mathbb{D}_f$. Un point $\mathbf{x}^* \in \mathbb{S}$ est un **minimiseur local** de f sur \mathbb{S} s'il existe $\epsilon > 0$ tel que $f(\mathbf{x}^*) \leq f(\mathbf{x})$ pour tout $\mathbf{x} \in \mathbb{S} \setminus \{\mathbf{x}^*\}$, $\|\mathbf{x} - \mathbf{x}^*\| < \epsilon$.

Définition. (Minimiseur global)

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

- Si dans les définitions précédentes on remplace les inégalités larges (en rouge) par des inégalités strictes 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.

Optimisation multidimensionnelle

- De façon générale un problème d'optimisation s'exprime comme suit :

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

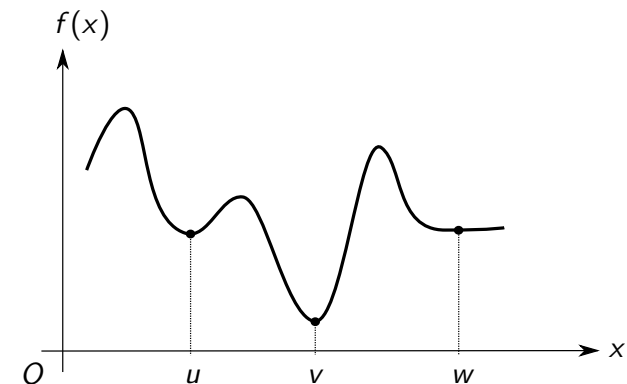
sous la contrainte que $\mathbf{x} \in \mathbb{S}$

où :

- ▶ f est une **fonction objectif** de \mathbb{R}^n 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**.
- 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, on cherche \mathbf{x}^* telle que $f(\mathbf{x}^*) \leq f(\mathbf{x})$, $\forall \mathbf{x} \in \mathbb{S}$.
- S'il s'agit d'un problème de maximisation alors il est équivalent de résoudre $\min -f(\mathbf{x})$ slc $\mathbf{x} \in \mathbb{S}$. Dans ce qui suit, on s'intéresse uniquement aux problèmes de minimisation sans perte de généralité.

Illustration (rappel)

- 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

Rappel du Sommaire

6 Optimisation multidimensionnelle

- Définitions de base
- Existence et recherche d'extrema
- Caractérisation des extrema d'une fonction
- Algorithmes d'optimisation numérique multidimensionnelle

Résolution d'un problème d'optimisation

- Résoudre un problème d'optimisation signifie déterminer $\arg \min_{\mathbf{x} \in \mathbb{S}} f(\mathbf{x})$.
- Mais il est souvent difficile de déterminer le (ou les) minimiseur global.
- On se contente alors de déterminer un (ou des) minimiseur local.
- Dans la suite on s'intéresse aux conditions mathématiques nous permettant de caractériser les minimiseurs locaux.
- Les conditions pour un minimiseur local de $f : \mathbb{R}^n \rightarrow \mathbb{R}$ reposent sur les dérivées partielles d'ordre 1 et 2 de f .
- Ces conditions généralisent celles vues dans le cas de fonctions réelles à valeurs réelles.
- Nous introduisons d'autres définitions permettant d'établir ces conditions dans le cas multidimensionnel.

Théorème de Weierstrass (cas multidimensionnel)

Théorème.

Soit $f : \mathbb{S} \rightarrow \mathbb{R}$ où $\mathbb{S} \subset \mathbb{R}^n$. Si f est continue sur \mathbb{S} et si \mathbb{S} est compact alors $f(\mathbb{S})$ est aussi compact. On en déduit que f est bornée sur \mathbb{S} et qu'elle atteint ses bornes c-à-d :

$$\exists(\mathbf{c}, \mathbf{d}) \in \mathbb{S}^2, \forall \mathbf{x} \in \mathbb{S}, \underbrace{f(\mathbf{c})}_m \leq f(\mathbf{x}) \leq \underbrace{f(\mathbf{d})}_M$$

- C'est le cas général du résultat présenté en slide 84. Il montre qu'une fonction continue sur un compact possède des extrema.

Rappel du Sommaire

6 Optimisation multidimensionnelle

- Définitions de base
- Existence et recherche d'extrema
- Caractérisation des extrema d'une fonction
- Algorithmes d'optimisation numérique multidimensionnelle

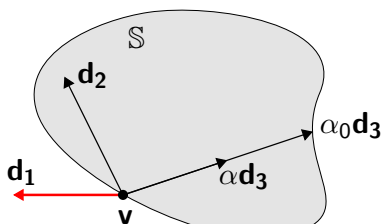
Direction admissible

- 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 :



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 :

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

- Intuitivement, \mathbf{x}^* est un minimiseur local si toutes les directions admissibles en \mathbf{x} conduisent à un accroissement différentiel positif.
- Rappel : si θ est l'angle non-orienté entre \mathbf{x} et \mathbf{y} , on a $\cos(\theta) = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|}$ et donc :

$$\langle \mathbf{x}, \mathbf{y} \rangle \leq 0 \Leftrightarrow \cos(\theta) \leq 0$$

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 :

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

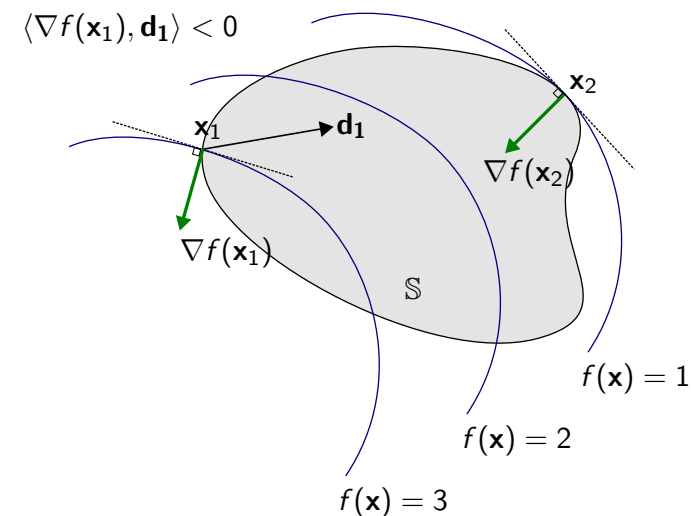
- En considérant la limite quand $\alpha \rightarrow 0$, on conclut donc que si \mathbf{x}^* est un minimiseur local, alors :

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

□

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 on a :

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

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 $\mathbf{d}^\top \nabla f(\mathbf{x}^*) = 0$ alors nous avons :

$$\mathbf{d}^\top \nabla^2 f(\mathbf{x}^*) \mathbf{d} \geq 0$$

où $\nabla^2 f$ 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 $\mathbf{d}^\top \nabla f(\mathbf{x}^*) = 0$ et $\mathbf{d}^\top \nabla^2 f(\mathbf{x}^*) \mathbf{d} < 0$
- 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 l'approximation 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 $\phi''(0) = \mathbf{d}^\top \nabla^2 f(\mathbf{x}^*) \mathbf{d} < 0$

- 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 :

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

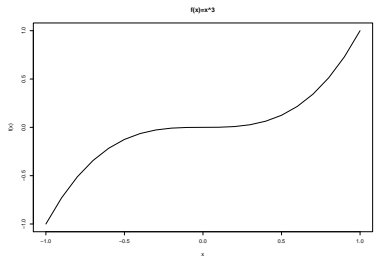
- $\nabla^2 f(\mathbf{x}^*)$ est sdp, $\forall \mathbf{d} \in \mathbb{R}^n$:

$$\mathbf{d}^\top \nabla^2 f(\mathbf{x}^*) \mathbf{d} \geq 0$$

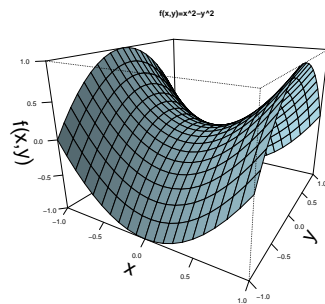
CNPO et CNSO sont nécessaires mais pas suffisantes

- Exemples de **point selle**, càd des points critiques mais qui ne sont pas des optimiseurs.

- $f(x) = x^3$ en 0.



- $f(x, y) = x^2 - y^2$ en $(0, 0)$.



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

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

Soit f une fonction de classe \mathbb{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 :

- $\nabla f(\mathbf{x}^*) = \mathbf{0}$,
- $\nabla^2 f(\mathbf{x}^*)$ est dp.

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

Preuve de la CSSO (point intérieur)

Démonstration.

- Comme $f \in \mathbb{C}^2$, la matrice hessienne $\nabla^2 f$ est symétrique et donc $\nabla^2 f(\mathbf{x}^*) = (\nabla^2 f(\mathbf{x}^*))^\top$
- Supposons que $\nabla^2 f(\mathbf{x}^*)$ est dp (2ème condition). En utilisant l'inégalité de Rayleigh, nous avons $\forall \mathbf{d} \neq \mathbf{0}$: $0 < \lambda_{\min}(\nabla^2 f(\mathbf{x}^*)) \|\mathbf{d}\|^2 \leq \mathbf{d}^\top \nabla^2 f(\mathbf{x}^*) \mathbf{d}$
- En supposant que $\nabla f(\mathbf{x}^*) = \mathbf{0}$ (condition 1) et en utilisant les séries de Taylor, nous obtenons :
$$\begin{aligned} f(\mathbf{x}^* + \mathbf{d}) - f(\mathbf{x}^*) &= \frac{1}{2} \mathbf{d}^\top \nabla^2 f(\mathbf{x}^*) \mathbf{d} + o(\|\mathbf{d}\|^2) \\ &\geq \frac{\lambda_{\min}(\nabla^2 f(\mathbf{x}^*))}{2} \|\mathbf{d}\|^2 + o(\|\mathbf{d}\|^2) \\ &> 0 \end{aligned}$$
- 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 :

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

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

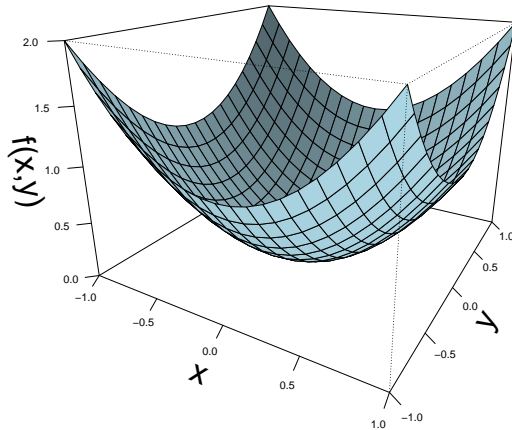
$$\nabla^2 f(\mathbf{x}) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

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)

- Graphe de $f(\mathbf{x}) = x_1^2 + x_2^2$.



Optimisation multidimensionnelle

- Nous nous intéressons dans cette sous-section aux algorithmes d'optimisation numérique pour des fonctions objectif de plusieurs variables et à valeurs réelles, $f : \mathbb{R}^n \rightarrow \mathbb{R}$.
- 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 (cf slide suivant).
 - ▶ Il existe par contre plusieurs façons de déterminer le pas.
- Ces algorithmes sont appelés **algorithmes de descente**.

Rappel du Sommaire

- 6 Optimisation multidimensionnelle
 - Définitions de base
 - Existence et recherche d'extrema
 - Caractérisation des extrema d'une fonction
 - Algorithmes d'optimisation numérique multidimensionnelle
 - Algorithmes utilisant la dérivée première uniquement
 - Algorithmes utilisant la dérivée seconde également

Algorithmes de descente de gradient

- Soit \mathbf{d} une direction telle que $\|\mathbf{d}\| = 1$. Rappelons alors que $\langle \nabla f(\mathbf{x}), \mathbf{d} \rangle$ est l'accroissement différentiel 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 :

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

- Nous venons de montrer que la direction suivie par $\nabla f(\mathbf{x})$ est celle de l'**accroissement différentiel maximum** de f en \mathbf{x} !

Algorithmes de descente de gradient (suite)

- On en déduit que la direction $-\nabla f(\mathbf{x})$ est celle conduisant à l'accroissement différentiel **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 :

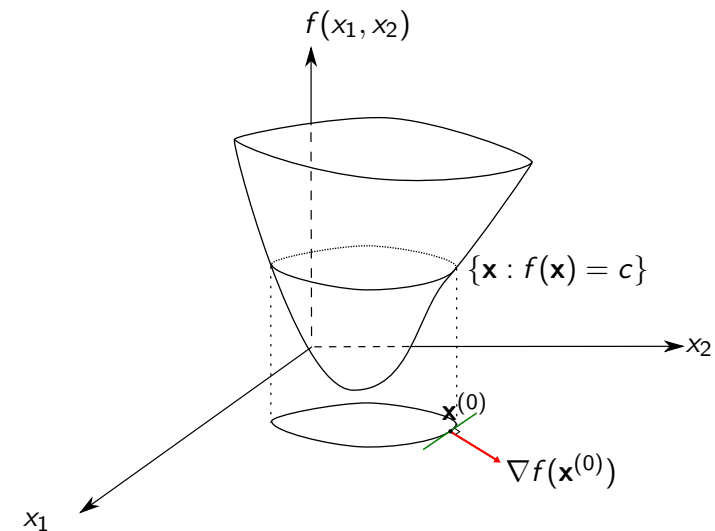
$$f(\mathbf{x}^{(0)} - \alpha \nabla f(\mathbf{x}^{(0)})) = f(\mathbf{x}^{(0)}) - \alpha \|\nabla f(\mathbf{x}^{(0)})\|^2 + o(\alpha)$$

- ▶ 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 :  $f \in \mathbb{C}^1, \mathbf{x}^0$ 
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    $\mathbf{x}^{(k+1)} \leftarrow \mathbf{x}^{(k)} - \alpha_k \nabla f(\mathbf{x}^{(k)})$ 
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 (cf plus loin).
- Comment déterminer le pas α_k à chaque itération k ? On distingue :
 - ▶ la méthode la plus simple consiste à donner une valeur constante à α_k . On parle de **méthode à pas fixe**. Le problème est qu'on n'est pas sûr de converger.
 - ▶ On a donc recouru à d'autres méthodes approchées pour déterminer α_k et on fait la distinction entre :
 - ★ les méthodes utilisant la dérivée première uniquement,
 - ★ les méthodes utilisant la dérivée seconde également,
 - ★ les méthodes de recherche linéaire exacte et inexacte.
- On voit ci-dessous la **méthode du pas optimal** qui possède des propriétés remarquables (contrairement à la méthode du pas fixe).

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.

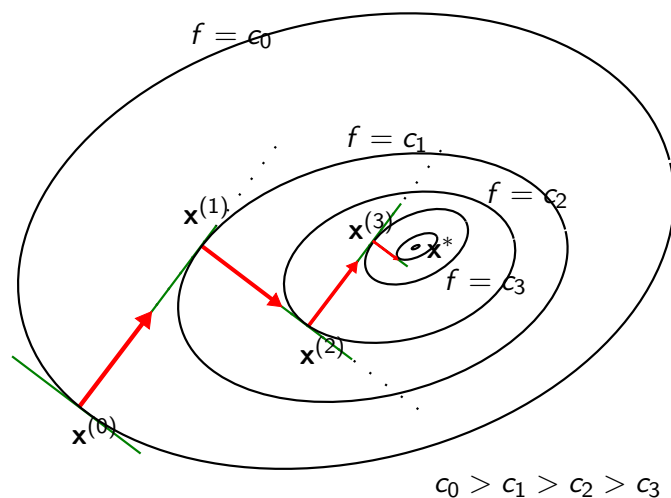
- Comme $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \alpha_k \nabla f(\mathbf{x}^{(k)})$ on a :

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

$$\begin{aligned} 0 &= \phi_k'(\alpha_k) \\ &= \frac{d\phi_k}{d\alpha}(\alpha_k) \\ &= \nabla f(\mathbf{x}^{(k)} - \alpha_k \nabla f(\mathbf{x}^{(k)}))^\top (-\nabla f(\mathbf{x}^{(k)})) \\ &= -\langle \nabla f(\mathbf{x}^{(k+1)}), \nabla f(\mathbf{x}^{(k)}) \rangle \end{aligned}$$

□

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 $f(\mathbf{x}^{(k+1)}) < f(\mathbf{x}^{(k)})$

- 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 vu précédemment, $\nabla f(\mathbf{x}^{(k)}) = \mathbf{0}$ est une condition d'arrêt de l'algorithme.
- Toutefois, en pratique, la gradient vaut rarement exactement $\mathbf{0}$.

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" :
 - $\|\nabla f(\mathbf{x}^{(k)})\| < \epsilon$
 - $|f(\mathbf{x}^{(k+1)}) - f(\mathbf{x}^{(k)})| < \epsilon$
 - $\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\| < \epsilon$
- 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 $\mathbf{x}^{(0)} = (4, 2, -1)$

- 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 :

1 $k \leftarrow 0$

Itération 1

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

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

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

$$= f\left(\begin{pmatrix} 4 \\ 2 \\ -1 \end{pmatrix} - \alpha \begin{pmatrix} 0 \\ -2 \\ 1024 \end{pmatrix}\right)$$

Pseudo-code de l'algorithme à pas optimal

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

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 :** $\mathbf{x}^{(k)}$

Exemple (suite)

Itération 1 (suite)

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

$$= f(4, 2 + 2\alpha, -1 - 1024\alpha)$$

$$= (2\alpha - 1)^2 + 4(1024\alpha + 4)^4$$

En utilisant la méthode de la sécante on trouve :

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

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

5 $k \leftarrow k + 1 = 1$

Itération 2

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

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

$$\phi^{(1)}(\alpha) = f(\mathbf{x}^{(1)} - \alpha \nabla f(\mathbf{x}^{(1)}))$$

$$= (2.008 - 1.984\alpha - 3)^2 + 4(-5.062 + 0.003875\alpha + 5)^4$$

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

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

$$\text{On a : } \nabla f(\mathbf{x}^{(1)}) = (0, 0, -0.003525)$$

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

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.
- 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.
- 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...) 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 :

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

où $\mathbf{Q} \in \mathbb{M}_n(\mathbb{R})$ 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 :

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

- La matrice hessienne vaut :

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

- Pour simplifier les notations on notera : $\mathbf{g}^{(k)} = \nabla f(\mathbf{x}^{(k)})$
- En appliquant la méthode du pas optimal on cherche donc :

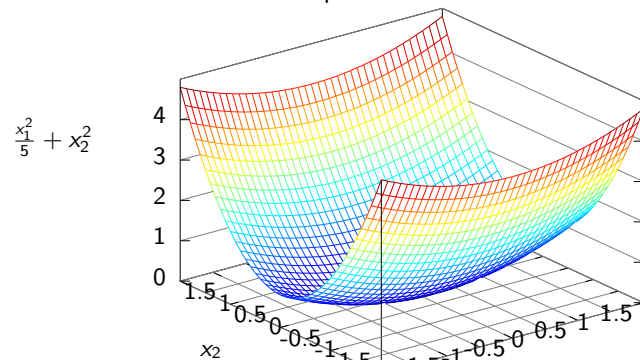
$$\alpha_k = \min_{\alpha \geq 0} f(\mathbf{x}^{(k)} - \alpha \mathbf{g}^{(k)})$$

$$= \arg \min_{\alpha \geq 0} \left(\frac{1}{2}(\mathbf{x}^{(k)} - \alpha \mathbf{g}^{(k)})^\top \mathbf{Q}(\mathbf{x}^{(k)} - \alpha \mathbf{g}^{(k)}) - (\mathbf{x}^{(k)} - \alpha \mathbf{g}^{(k)})^\top \mathbf{b} \right)$$

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 càd beaucoup d'itérations (fonction objectif de type "vallée étroite").
- Illustration :

Exemple de "vallée étroite"



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) :

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

- On a donc :

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

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

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

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 $\mathbf{x}^{(0)}$ 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 $\mathbf{x}^{(0)}$ soit **proche du 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 $\mathbf{x}^{(0)}$.

- 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}^* c'ad 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 : $x_k = \frac{1}{k}$
- $x_k \rightarrow 0$
- $\frac{|x_{k+1}|}{|x_k|^p} = \frac{k^p}{k+1}$
- 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

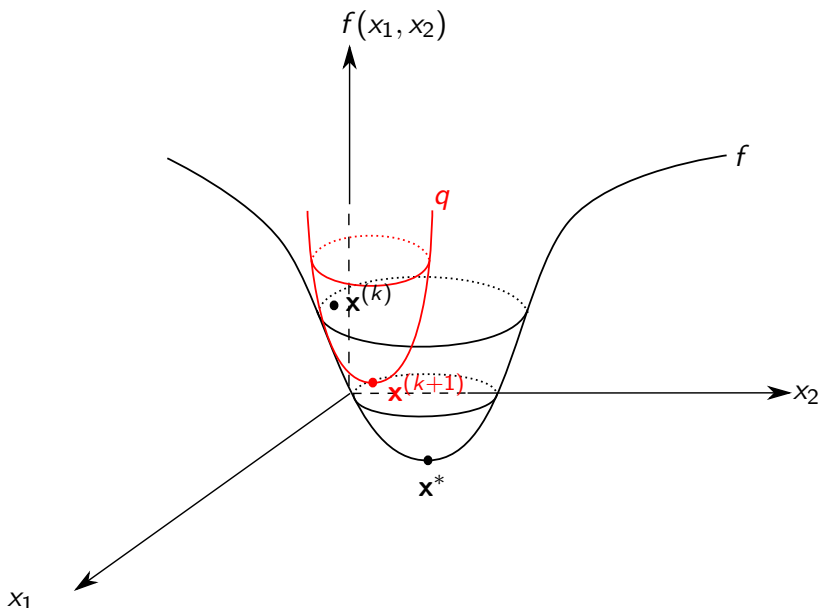
Théorème.

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 est basée sur 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).
- Celle-ci utilise de plus la **dérivée seconde** et est plus efficace à condition que le point initial $\mathbf{x}^{(0)}$ soit assez proche du minimiseur.
- L'idée est 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.

Illustration



Algorithmes de Newton (suite)

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

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

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

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

- Si $\nabla^2 f(\mathbf{x}^{(k)})$ est dp alors q atteint son minimum en :

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

- L'algorithme itératif de descente de gradient devient :

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

Pseudo-code de l'algorithme de Newton

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

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

Exemple

- 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 de la matrice hessienne :

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

Exemple (suite)

Déroulement de l'algorithme :

$$1 \quad k \leftarrow 0$$

Itération 1

$$3 \quad \mathbf{g}^{(0)} = (306, -144, -2, -310)$$

$$\nabla^2 f(\mathbf{x}^{(0)}) = \begin{pmatrix} 482 & 20 & 0 & -480 \\ 20 & 212 & -24 & 0 \\ 0 & -24 & 58 & -10 \\ -480 & 0 & -10 & 490 \end{pmatrix}$$

$$\nabla^2 f(\mathbf{x}^{(0)})^{-1} = \begin{pmatrix} .1126 & -.0089 & .0154 & .1106 \\ -.0089 & .0057 & .0008 & -.0087 \\ .0154 & .0008 & .0203 & .0155 \\ .1106 & -.0087 & .0155 & .1107 \end{pmatrix}$$

$$\nabla^2 f(\mathbf{x}^{(0)})^{-1} \mathbf{g}^{(0)} = (1.4127, -0.8413, -0.2540, 0.7460)$$

$$\mathbf{x}^{(1)} \leftarrow \mathbf{x}^{(0)} - \nabla^2 f(\mathbf{x}^{(0)})^{-1} \nabla f(\mathbf{x}^{(0)}) = (1.5873, -0, 1587, 0.2540, 0.2540)$$

$$f(\mathbf{x}^{(1)}) = 31.8$$

Exemple (suite)

Déroulement de l'algorithme :

$$4 \quad k \leftarrow k + 1 = 1$$

Itération 2

$$3 \quad \mathbf{g}^{(1)} = (94.81, -1.179, 2.371, -94.81)$$

$$\nabla^2 f(\mathbf{x}^{(1)}) = \begin{pmatrix} 215.3 & 20 & 0 & -213.3 \\ 20 & 205.3 & -10.67 & 0 \\ 0 & -10.67 & 31.34 & -10 \\ -213.3 & 0 & -10 & 223.3 \end{pmatrix}$$

$$\nabla^2 f(\mathbf{x}^{(1)})^{-1} \mathbf{g}^{(1)} = (0.5291, -0.0529, -0.0846, 0.0846)$$

$$\mathbf{x}^{(2)} \leftarrow \mathbf{x}^{(1)} - \nabla^2 f(\mathbf{x}^{(1)})^{-1} \nabla f(\mathbf{x}^{(1)}) = (1.0582, -0, 1058, 0.1694, 0.1694)$$

$$f(\mathbf{x}^{(2)}) = 6.28$$

Exemple (suite)

Déroulement de l'algorithme :

$$4 \quad k \leftarrow k + 1 = 2$$

Itération 3

$$3 \quad \mathbf{g}^{(2)} = (28.09, -0.3475, 0.7031, -28.08)$$

$$\nabla^2 f(\mathbf{x}^{(2)}) = \begin{pmatrix} 96.80 & 20 & 0 & -94.80 \\ 20 & 202.4 & -4.744 & 0 \\ 0 & -4.744 & 19.49 & -10 \\ -94.80 & 0 & -10 & 104.80 \end{pmatrix}$$

$$\mathbf{x}^{(3)} \leftarrow \mathbf{x}^{(2)} - \nabla^2 f(\mathbf{x}^{(2)})^{-1} \nabla f(\mathbf{x}^{(2)}) = (0.7037, -0, 0704, 0.1121, 0.1121)$$

$$f(\mathbf{x}^{(3)}) = 1.24$$

Remarques sur la méthode de Newton

- Remarquons que la k ème itération de l'algorithme peut se résoudre en deux sous-étapes :
 - 1 Résoudre $\nabla^2 f(\mathbf{x}^{(k)})\mathbf{d}^{(k)} = -g^{(k)}$.
 - 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 (suite)

- Dans le cas particulier où $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top \mathbf{Q}\mathbf{x} - \mathbf{b}^\top \mathbf{x}$ est une fonction quadratique, nous avons :

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

et

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

- Soit $\mathbf{x}^{(0)}$ un point initial, en appliquant la méthode de Newton on obtient :

$$\begin{aligned} \mathbf{x}^{(1)} &= \mathbf{x}^{(0)} - \nabla^2 f(\mathbf{x}^{(0)})^{-1}g^{(0)} \\ &= \mathbf{x}^{(0)} - \mathbf{Q}^{-1}(\mathbf{Q}\mathbf{x}^{(0)} - \mathbf{b}) \\ &= \mathbf{Q}^{-1}\mathbf{b} \\ &= \mathbf{x}^* \end{aligned}$$

- Donc pour une fonction objectif quadratique la méthode trouve le minimiseur en une itération.

Vitesse de convergence de la méthode de Newton

- Comme pour le cas d'une variable, il n'est pas garanti que la méthode de Newton détermine la bonne direction de descente si $\nabla^2 f(\mathbf{x}^{(k)})$ n'est pas définie positive.
- Si $\mathbf{x}^{(0)}$ est très loin du minimiseur, même si $\nabla^2 f(\mathbf{x}^{(k)}) > 0$ 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.

Vitesse de convergence de la méthode de Newton (suite)

Théorème.

Soit $f \in \mathbb{C}^3$, et $\mathbf{x}^* \in \mathbb{R}^n$ un point tel que $\nabla f(\mathbf{x}^*) = \mathbf{0}$ et $\nabla^2 f(\mathbf{x}^*)$ 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 $\nabla^2 f(\mathbf{x}^{(k)}) > 0$. Si $\nabla^2 f(\mathbf{x}^{(k)}) \leq 0$, 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 $\nabla^2 f(\mathbf{x}^{(k)})^{-1}$ (approche DFP ou BFGS). Il s'agit de méthodes "utilisant" des dérivées secondes (stockage d'une matrice $(n \times n)$).