

# Ajustement des paramètres d'un modèle à des données expérimentales

2ème année

E.N.S.T.B.B.

I.P.B.

Année Universitaire 2014-15

# Plan

- 1 Introduction
- 2 Critère des moindres carrés : cas linéaire
- 3 Critère des moindres carrés : cas non linéaire

On dispose d'une série de  $n$  mesures  $\{(x_i, y_i), 1 \leq i \leq n\}$  obtenue en observant une variable  $y$  en fonction d'une variable  $x$ . On dispose aussi d'un modèle de relation entre  $x$  et  $y$  suggéré par les données ou établi à partir de considérations théoriques.

Ce modèle dépend de paramètres  $a_1, a_2, \dots, a_p$ . Le problème consiste à déterminer des valeurs des paramètres  $a_1, a_2, \dots, a_p$  à partir des données expérimentales telles que le graphe de la fonction  $f(x, a_1, a_2, \dots, a_p)$  approche "au mieux" les points  $(x_i, y_i)$  correspondants aux données.

On dispose d'une série de  $n$  mesures  $\{(x_i, y_i), 1 \leq i \leq n\}$  obtenue en observant une variable  $y$  en fonction d'une variable  $x$ . On dispose aussi d'un modèle de relation entre  $x$  et  $y$  suggéré par les données ou établi à partir de considérations théoriques.

Ce modèle dépend de paramètres  $a_1, a_2, \dots, a_p$ . Le problème consiste à déterminer des valeurs des paramètres  $a_1, a_2, \dots, a_p$  à partir des données expérimentales telles que le graphe de la fonction  $f(x, a_1, a_2, \dots, a_p)$  approche "au mieux" les points  $(x_i, y_i)$  correspondants aux données.

On dispose d'une série de  $n$  mesures  $\{(x_i, y_i), 1 \leq i \leq n\}$  obtenue en observant une variable  $y$  en fonction d'une variable  $x$ . On dispose aussi d'un modèle de relation entre  $x$  et  $y$  suggéré par les données ou établi à partir de considérations théoriques.

Ce modèle dépend de paramètres  $a_1, a_2, \dots, a_p$ . Le problème consiste à déterminer des valeurs des paramètres  $a_1, a_2, \dots, a_p$  à partir des données expérimentales telles que le graphe de la fonction  $f(x, a_1, a_2, \dots, a_p)$  approche "au mieux" les points  $(x_i, y_i)$  correspondants aux données.

On dispose d'une série de  $n$  mesures  $\{(x_i, y_i), 1 \leq i \leq n\}$  obtenue en observant une variable  $y$  en fonction d'une variable  $x$ . On dispose aussi d'un modèle de relation entre  $x$  et  $y$  suggéré par les données ou établi à partir de considérations théoriques.

Ce modèle dépend de paramètres  $a_1, a_2, \dots, a_p$ . Le problème consiste à déterminer des valeurs des paramètres  $a_1, a_2, \dots, a_p$  à partir des données expérimentales telles que le graphe de la fonction  $f(x, a_1, a_2, \dots, a_p)$  approche "au mieux" les points  $(x_i, y_i)$  correspondants aux données.

## Exemple

*Lorsqu'on étudie une enzyme, on dispose d'un ensemble de mesures  $(s_i, v_i)$  pour  $i = 1..n$  et d'un modèle théorique par exemple  $v(s) = \frac{V_M s}{K_M + s}$ . A partir de ces mesures, on peut chercher les paramètres  $V_M$  et  $K_M$  qui représentent au mieux les données expérimentales  $(s_i, v_i)$ .*

# Plan

- 1 Introduction
- 2 Critère des moindres carrés : cas linéaire
  - Cas de la droite
  - Généralisation
- 3 Critère des moindres carrés : cas non linéaire
  - introduction
  - Linéarisation de problèmes non linéaires
  - Première méthode itérative: méthode de Newton
  - Autres méthodes itératives
  - Autre méthode: algorithme évolutionnaire
  - Le point de vue du statisticien



## Cas de la droite

On suppose que la liaison entre  $y$  et  $x$  est une droite

$y = a_1 + a_2x$ . On cherche donc l'équation de la droite qui passe "au mieux" entre les points expérimentaux. Un critère pour approcher "au mieux" les points  $(x_i, y_i)$  est le critère des moindres carrés. Il consiste à minimiser la distance entre le point expérimental  $(x_i, y_i)$  et le point  $(x_i, y_i^{th})$  obtenu à partir de  $x_i$  et de l'équation  $y_i^{th} = a_1 + a_2x_i$ . On calcule l'écart entre valeur expérimentale et valeur théorique  $y_i - y_i^{th}$  au carré pour chaque point  $i = 1..n$ .

Puis on fait varier  $(a_1, a_2)$  jusqu'à trouver  $(a_1^*, a_2^*)$  tels que

$S = \sum_{i=1}^n (y_i - y_i^{th})^2$  soit le plus petit possible.

## Cas de la droite

On suppose que la liaison entre  $y$  et  $x$  est une droite  $y = a_1 + a_2x$ . On cherche donc l'équation de la droite qui passe "au mieux" entre les points expérimentaux. Un critère pour approcher "au mieux" les points  $(x_i, y_i)$  est le critère des moindres carrés. Il consiste à minimiser la distance entre le point expérimental  $(x_i, y_i)$  et le point  $(x_i, y_i^{th})$  obtenu à partir de  $x_i$  et de l'équation  $y_i^{th} = a_1 + a_2x_i$ . On calcule l'écart entre valeur expérimentale et valeur théorique  $y_i - y_i^{th}$  au carré pour chaque point  $i = 1..n$ .

Puis on fait varier  $(a_1, a_2)$  jusqu'à trouver  $(a_1^*, a_2^*)$  tels que  $S = \sum_{i=1}^n (y_i - y_i^{th})^2$  soit le plus petit possible.

## Cas de la droite

On suppose que la liaison entre  $y$  et  $x$  est une droite  $y = a_1 + a_2x$ . On cherche donc l'équation de la droite qui passe "au mieux" entre les points expérimentaux. Un critère pour approcher "au mieux" les points  $(x_i, y_i)$  est le critère des moindres carrés. Il consiste à minimiser la distance entre le point expérimental  $(x_i, y_i)$  et le point  $(x_i, y_i^{th})$  obtenu à partir de  $x_i$  et de l'équation  $y_i^{th} = a_1 + a_2x_i$ . On calcule l'écart entre valeur expérimentale et valeur théorique  $y_i - y_i^{th}$  au carré pour chaque point  $i = 1..n$ .

Puis on fait varier  $(a_1, a_2)$  jusqu'à trouver  $(a_1^*, a_2^*)$  tels que

$$S = \sum_{i=1}^n (y_i - y_i^{th})^2 \text{ soit le plus petit possible.}$$

## Cas de la droite

On suppose que la liaison entre  $y$  et  $x$  est une droite  $y = a_1 + a_2x$ . On cherche donc l'équation de la droite qui passe "au mieux" entre les points expérimentaux. Un critère pour approcher "au mieux" les points  $(x_i, y_i)$  est le critère des moindres carrés. Il consiste à minimiser la distance entre le point expérimental  $(x_i, y_i)$  et le point  $(x_i, y_i^{th})$  obtenu à partir de  $x_i$  et de l'équation  $y_i^{th} = a_1 + a_2x_i$ . On calcule l'écart entre valeur expérimentale et valeur théorique  $y_i - y_i^{th}$  au carré pour chaque point  $i = 1..n$ .

Puis on fait varier  $(a_1, a_2)$  jusqu'à trouver  $(a_1^*, a_2^*)$  tels que

$$S = \sum_{i=1}^n (y_i - y_i^{th})^2 \text{ soit le plus petit possible.}$$

## Cas de la droite

On suppose que la liaison entre  $y$  et  $x$  est une droite  $y = a_1 + a_2x$ . On cherche donc l'équation de la droite qui passe "au mieux" entre les points expérimentaux. Un critère pour approcher "au mieux" les points  $(x_i, y_i)$  est le critère des moindres carrés. Il consiste à minimiser la distance entre le point expérimental  $(x_i, y_i)$  et le point  $(x_i, y_i^{th})$  obtenu à partir de  $x_i$  et de l'équation  $y_i^{th} = a_1 + a_2x_i$ . On calcule l'écart entre valeur expérimentale et valeur théorique  $y_i - y_i^{th}$  au carré pour chaque point  $i = 1..n$ .

Puis on fait varier  $(a_1, a_2)$  jusqu'à trouver  $(a_1^*, a_2^*)$  tels que

$$S = \sum_{i=1}^n (y_i - y_i^{th})^2 \text{ soit le plus petit possible.}$$

# Cas de la droite

On cherche donc le minimum de la fonction

$$S(a_1, a_2) = \sum_{i=1}^n (y_i - (a_1 + a_2 x_i))^2$$

Pour cela on cherche les points critiques de cette fonction c'est-à-dire les points où les dérivées partielles par rapport à  $a_i$  avec  $i = 1, 2$  s'annulent simultanément:  $(a_1^*, a_2^*)$  vérifient

$$\frac{\partial S}{\partial a_1}(a_1^*, a_2^*) = 0 \text{ et } \frac{\partial S}{\partial a_2}(a_1^*, a_2^*) = 0,$$

## Cas de la droite

On cherche donc le minimum de la fonction

$$S(a_1, a_2) = \sum_{i=1}^n (y_i - (a_1 + a_2 x_i))^2$$

Pour cela on cherche les points critiques de cette fonction c'est-à-dire les points où les dérivées partielles par rapport à  $a_i$  avec  $i = 1, 2$  s'annulent simultanément:  $(a_1^*, a_2^*)$  vérifient

$$\frac{\partial S}{\partial a_1}(a_1^*, a_2^*) = 0 \text{ et } \frac{\partial S}{\partial a_2}(a_1^*, a_2^*) = 0,$$

## Cas de la droite

On cherche donc le minimum de la fonction

$$S(a_1, a_2) = \sum_{i=1}^n (y_i - (a_1 + a_2 x_i))^2$$

Pour cela on cherche les points critiques de cette fonction c'est-à-dire les points où les dérivées partielles par rapport à  $a_i$  avec  $i = 1, 2$  s'annulent simultanément:  $(a_1^*, a_2^*)$  vérifient

$$\frac{\partial S}{\partial a_1}(a_1^*, a_2^*) = 0 \text{ et } \frac{\partial S}{\partial a_2}(a_1^*, a_2^*) = 0,$$



# Cas de la droite

Les dérivées partielles de  $S$  sont

$$\begin{cases} \frac{\partial S}{\partial a_1}(a_1, a_2) = -2 \sum_{i=1}^n (y_i - (a_1 + a_2 x_i)) \\ \frac{\partial S}{\partial a_2}(a_1, a_2) = -2 \sum_{i=1}^n x_i (y_i - (a_1 + a_2 x_i)). \end{cases}$$

donc le minimum de  $S$  est en  $(a_1^*, a_2^*)$  qui vérifient

$$\begin{cases} \sum_{i=1}^n (y_i - (a_1^* + a_2^* x_i)) = 0 \\ \sum_{i=1}^n x_i (y_i - (a_1^* + a_2^* x_i)) = 0 \end{cases}$$

## Cas de la droite

Les dérivées partielles de  $S$  sont

$$\begin{cases} \frac{\partial S}{\partial a_1}(a_1, a_2) = -2 \sum_{i=1}^n (y_i - (a_1 + a_2 x_i)) \\ \frac{\partial S}{\partial a_2}(a_1, a_2) = -2 \sum_{i=1}^n x_i (y_i - (a_1 + a_2 x_i)). \end{cases}$$

donc le minimum de  $S$  est en  $(a_1^*, a_2^*)$  qui vérifie

$$\begin{cases} \sum_{i=1}^n (y_i - (a_1^* + a_2^* x_i)) = 0 \\ \sum_{i=1}^n x_i (y_i - (a_1^* + a_2^* x_i)) = 0 \end{cases}$$

# Cas de la droite

donc  $(a_1^*, a_2^*)$  vérifient

$$\begin{cases} na_1^* + a_2^* \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \\ a_1^* \sum_{i=1}^n x_i + a_2^* \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i \end{cases}$$

soit encore

$$\begin{cases} a_2^* = \frac{\text{cov}(x, y)}{\text{var}(x)} \\ a_1^* = \bar{y} - a_2^* \bar{x} \end{cases}$$

De plus  $S(a_1, a_2)$  étant strictement convexe, ce point critique  $(a_1^*, a_2^*)$  est bien un minimum pour  $S$ .

## Cas de la droite

donc  $(a_1^*, a_2^*)$  vérifient

$$\begin{cases} na_1^* + a_2^* \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \\ a_1^* \sum_{i=1}^n x_i + a_2^* \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i \end{cases}$$

soit encore

$$\begin{cases} a_2^* = \frac{\text{cov}(x, y)}{\text{var}(x)} \\ a_1^* = \bar{y} - a_2^* \bar{x} \end{cases}$$

De plus  $S(a_1, a_2)$  étant strictement convexe, ce point critique  $(a_1^*, a_2^*)$  est bien un minimum pour  $S$ .

## Cas de la droite

donc  $(a_1^*, a_2^*)$  vérifient

$$\begin{cases} na_1^* + a_2^* \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \\ a_1^* \sum_{i=1}^n x_i + a_2^* \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i \end{cases}$$

soit encore

$$\begin{cases} a_2^* = \frac{\text{cov}(x, y)}{\text{var}(x)} \\ a_1^* = \bar{y} - a_2^* \bar{x} \end{cases}$$

De plus  $S(a_1, a_2)$  étant strictement convexe, ce point critique  $(a_1^*, a_2^*)$  est bien un minimum pour  $S$ .

# Cas de la droite

Reprenons le système

$$\begin{cases} na_1^* + a_2^* \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \\ a_1^* \sum_{i=1}^n x_i + a_2^* \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i \end{cases}$$

et écrivons-le sous forme matricielle

$$\begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix} \begin{pmatrix} a_1^* \\ a_2^* \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{pmatrix}$$

# Cas de la droite

Reprenons le système

$$\begin{cases} na_1^* + a_2^* \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \\ a_1^* \sum_{i=1}^n x_i + a_2^* \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i \end{cases}$$

et écrivons-le sous forme matricielle

$$\begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix} \begin{pmatrix} a_1^* \\ a_2^* \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{pmatrix}$$

Remarquons que si l'on pose

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \dots & \\ \dots & \\ 1 & x_n \end{pmatrix} \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ \dots \\ y_n \end{pmatrix}$$

$$\text{alors } {}^tXX = \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix} \quad \text{et } {}^tXY = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{pmatrix}.$$



Notre système

$$\begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix} \begin{pmatrix} a_1^* \\ a_2^* \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{pmatrix}$$

s'écrit alors

$${}^tXX \begin{pmatrix} a_1^* \\ a_2^* \end{pmatrix} = {}^tXY.$$

Par conséquent  $\begin{pmatrix} a_1^* \\ a_2^* \end{pmatrix} = ({}^tXX)^{-1} {}^tXY.$

# Plan

- 1 Introduction
- 2 Critère des moindres carrés : cas linéaire
  - Cas de la droite
  - **Généralisation**
- 3 Critère des moindres carrés : cas non linéaire
  - introduction
  - Linéarisation de problèmes non linéaires
  - Première méthode itérative: méthode de Newton
  - Autres méthodes itératives
  - Autre méthode: algorithme évolutionnaire
  - Le point de vue du statisticien

# Généralisation

Le critère des moindres carrés ne s'applique pas uniquement au cas d'une droite. Maintenant on cherche un vecteur de  $p$

paramètres  $(a_1^*, a_2^*, \dots, a_p^*)$  tels que  $y = \sum_{j=1}^p a_j^* f_j(x)$  où les  $f_j$

sont des fonctions données de  $x$ , seuls les paramètres  $a_j^*$  sont à estimer à partir des données expérimentales. Dans ce cas, on parle de cas linéaire par rapport aux paramètres et on trouve une solution explicite.

# Généralisation

Le critère des moindres carrés ne s'applique pas uniquement au cas d'une droite. Maintenant on cherche un vecteur de  $p$  paramètres  $(a_1^*, a_2^*, \dots, a_p^*)$  tels que  $y = \sum_{j=1}^p a_j^* f_j(x)$  où les  $f_j$  sont des fonctions données de  $x$ , seuls les paramètres  $a_j^*$  sont à estimer à partir des données expérimentales. Dans ce cas, on parle de cas linéaire par rapport aux paramètres et on trouve une solution explicite.

# Généralisation

Le critère des moindres carrés ne s'applique pas uniquement au cas d'une droite. Maintenant on cherche un vecteur de  $p$

paramètres  $(a_1^*, a_2^*, \dots, a_p^*)$  tels que  $y = \sum_{j=1}^p a_j^* f_j(x)$  où les  $f_j$

sont des fonctions données de  $x$ , seuls les paramètres  $a_j^*$  sont à estimer à partir des données expérimentales. Dans ce cas, on parle de cas linéaire par rapport aux paramètres et on trouve une solution explicite.

## Exemple

① *cas linéaires :*

$$y = a_1 + a_2 x, y = a_1 + a_2 x + a_3 x^2,$$

$$y = a_1 \sin(x) + a_2 \cos(2x), y = a_1 \exp(x) + a_2 \exp(2x)$$

② *cas non linéaires :*

$$y = a_1 \exp(-a_2 x), y = \frac{a_3}{1 + a_1 \exp(-a_2 x)}, y = \frac{a_1}{a_2 + x}$$

## Exemple

① *cas linéaires :*

$$y = a_1 + a_2x, y = a_1 + a_2x + a_3x^2,$$

$$y = a_1 \sin(x) + a_2 \cos(2x), y = a_1 \exp(x) + a_2 \exp(2x)$$

② *cas non linéaires :*

$$y = a_1 \exp(-a_2x), y = \frac{a_3}{1 + a_1 \exp(-a_2x)}, y = \frac{a_1}{a_2 + x}$$

Selon le critère des moindres carrés, on cherche  $(a_1^*, a_2^*, \dots, a_p^*)$  qui minimisent

$$S(a_1, a_2, \dots, a_p) = \sum_{i=1}^n \left( y_i - \sum_{j=1}^p a_j f_j(x_i) \right)^2.$$

Ce minimum est à rechercher parmi les points critiques de  $S$  c'est-à-dire les points qui annulent les dérivées partielles

$$\frac{\partial S}{\partial a_k}(a_1^*, a_2^*, \dots, a_p^*) = 0 \quad k = 1, \dots, p.$$



Selon le critère des moindres carrés, on cherche  $(a_1^*, a_2^*, \dots, a_p^*)$  qui minimisent

$$S(a_1, a_2, \dots, a_p) = \sum_{i=1}^n \left( y_i - \sum_{j=1}^p a_j f_j(x_i) \right)^2.$$

Ce minimum est à rechercher parmi les points critiques de  $S$  c'est-à-dire les points qui annulent les dérivées partielles

$$\frac{\partial S}{\partial a_k}(a_1^*, a_2^*, \dots, a_p^*) = 0 \quad k = 1, \dots, p.$$

Selon le critère des moindres carrés, on cherche  $(a_1^*, a_2^*, \dots, a_p^*)$  qui minimisent

$$S(a_1, a_2, \dots, a_p) = \sum_{i=1}^n \left( y_i - \sum_{j=1}^p a_j f_j(x_i) \right)^2.$$

Ce minimum est à rechercher parmi les points critiques de  $S$  c'est-à-dire les points qui annulent les dérivées partielles

$$\frac{\partial S}{\partial a_k}(a_1^*, a_2^*, \dots, a_p^*) = 0 \quad k = 1, \dots, p.$$

Comme

$$\frac{\partial S}{\partial a_k}(a_1, a_2, \dots, a_p) = -2 \sum_{i=1}^n \left( y_i - \sum_{j=1}^p a_j f_j(x_i) \right) f_k(x_i) \quad k = 1, \dots, p,$$

on obtient

$$\sum_{i=1}^n y_i f_k(x_i) = \sum_{i=1}^n \sum_{j=1}^p a_j^* f_j(x_i) f_k(x_i) \quad k = 1, \dots, p.$$

Comme

$$\frac{\partial S}{\partial a_k}(a_1, a_2, \dots, a_p) = -2 \sum_{i=1}^n \left( y_i - \sum_{j=1}^p a_j f_j(x_i) \right) f_k(x_i) \quad k = 1, \dots, p,$$

on obtient

$$\sum_{i=1}^n y_i f_k(x_i) = \sum_{i=1}^n \sum_{j=1}^p a_j^* f_j(x_i) f_k(x_i) \quad k = 1, \dots, p.$$

En posant

$$X = \begin{pmatrix} f_1(x_1) & f_2(x_1) & \cdots & f_p(x_1) \\ f_1(x_2) & f_2(x_2) & \cdots & f_p(x_2) \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ f_1(x_n) & f_2(x_n) & \cdots & f_p(x_n) \end{pmatrix} \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ \cdots \\ y_n \end{pmatrix},$$

sous forme matricielle, le système s'écrit

$${}^t X X \begin{pmatrix} a_1^* \\ a_2^* \\ \cdots \\ \cdots \\ a_p^* \end{pmatrix} = {}^t X Y.$$

En posant

$$X = \begin{pmatrix} f_1(x_1) & f_2(x_1) & \cdots & f_p(x_1) \\ f_1(x_2) & f_2(x_2) & \cdots & f_p(x_2) \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ f_1(x_n) & f_2(x_n) & \cdots & f_p(x_n) \end{pmatrix} \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ \cdots \\ y_n \end{pmatrix},$$

sous forme matricielle, le système s'écrit

$${}^t X X \begin{pmatrix} a_1^* \\ a_2^* \\ \cdots \\ \cdots \\ a_p^* \end{pmatrix} = {}^t X Y.$$

## Exemple

On suppose que l'on dispose de  $n$  mesures  $(x_i, y_i)$ . On cherche un vecteur  $a^*$  de paramètres  $a^* = (a_1^*, a_2^*, a_3^*)$  tels que la parabole  $y = a_1^* + a_2^*x + a_3^*x^2$  passe au mieux parmi les  $n$  mesures  $(x_i, y_i)$ .

On veut donc  $(a_1^*, a_2^*, a_3^*)$  tels que  $y = \sum_{j=1}^3 a_j^* f_j(x)$  avec  $f_1(x) = 1$ ,  $f_2(x) = x$  et  $f_3(x) = x^2$ . Ecrivons le problème sous forme matricielle

$$X = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \dots & & \\ \dots & & \\ 1 & x_n & x_n^2 \end{pmatrix} \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ \dots \\ y_n \end{pmatrix} .$$

## Exemple

On suppose que l'on dispose de  $n$  mesures  $(x_i, y_i)$ . On cherche un vecteur  $a^*$  de paramètres  $a^* = (a_1^*, a_2^*, a_3^*)$  tels que la parabole  $y = a_1^* + a_2^*x + a_3^*x^2$  passe au mieux parmi les  $n$  mesures  $(x_i, y_i)$ .

On veut donc  $(a_1^*, a_2^*, a_3^*)$  tels que  $y = \sum_{j=1}^3 a_j^* f_j(x)$  avec  $f_1(x) = 1$ ,  $f_2(x) = x$  et  $f_3(x) = x^2$ . Ecrivons le problème sous forme matricielle

$$X = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \dots & & \\ \dots & & \\ 1 & x_n & x_n^2 \end{pmatrix} \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ \dots \\ y_n \end{pmatrix} .$$



## Exemple

On suppose que l'on dispose de  $n$  mesures  $(x_i, y_i)$ . On cherche un vecteur  $a^*$  de paramètres  $a^* = (a_1^*, a_2^*, a_3^*)$  tels que la parabole  $y = a_1^* + a_2^*x + a_3^*x^2$  passe au mieux parmi les  $n$  mesures  $(x_i, y_i)$ .

On veut donc  $(a_1^*, a_2^*, a_3^*)$  tels que  $y = \sum_{j=1}^3 a_j^* f_j(x)$  avec  $f_1(x) = 1$ ,  $f_2(x) = x$  et  $f_3(x) = x^2$ . Ecrivons le problème sous forme matricielle

$$X = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \dots & & \\ \dots & & \\ 1 & x_n & x_n^2 \end{pmatrix} \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ \dots \\ y_n \end{pmatrix} .$$

$${}^tXX = \begin{pmatrix} n & \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 \\ \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 & \sum_{i=1}^n x_i^4 \end{pmatrix}$$

Sous forme matricielle, on a donc

$${}^tXX \begin{pmatrix} a_1^* \\ a_2^* \\ a_3^* \end{pmatrix} = {}^tXY.$$

$${}^tXX = \begin{pmatrix} n & \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 \\ \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 & \sum_{i=1}^n x_i^4 \end{pmatrix}$$

Sous forme matricielle, on a donc

$${}^tXX \begin{pmatrix} a_1^* \\ a_2^* \\ a_3^* \end{pmatrix} = {}^tXY.$$

## Exemple

$$y = a_1 \exp(x) + a_2 \exp(2x) + a_3 \exp(3x)$$

$${}^tXX = \begin{pmatrix} \sum_{i=1}^n \exp(2x_i) & \sum_{i=1}^n \exp(3x_i) & \sum_{i=1}^n \exp(4x_i) \\ \sum_{i=1}^n \exp(3x_i) & \sum_{i=1}^n \exp(4x_i) & \sum_{i=1}^n \exp(5x_i) \\ \sum_{i=1}^n \exp(4x_i) & \sum_{i=1}^n \exp(5x_i) & \sum_{i=1}^n \exp(6x_i) \end{pmatrix}.$$

## Exemple

On dispose du tableau de mesures  $(x_i, y_i)$  suivant :

$x_i$	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
$y_i$	2.3	0	-1.5	-0.6	1.7	3.1	1.7	-1.6	-4.2	-4

Un modèle de relation entre  $x$  et  $y$  suggéré est  $y = a_1 \sin(x) + a_2 \cos(2x)$ . On veut déterminer  $(a_1^*, a_2^*)$  qui minimisent le critère des moindres carrés.

$$\text{Posons } X = \begin{pmatrix} \sin(x_1) & \cos(2x_1) \\ \sin(x_2) & \cos(2x_2) \\ \dots & \dots \\ \dots & \dots \\ \sin(x_n) & \cos(2x_n) \end{pmatrix} \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ \dots \\ y_n \end{pmatrix} .$$

## Exemple

On dispose du tableau de mesures  $(x_i, y_i)$  suivant :

$x_i$	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
$y_i$	2.3	0	-1.5	-0.6	1.7	3.1	1.7	-1.6	-4.2	-4

Un modèle de relation entre  $x$  et  $y$  suggéré est  $y = a_1 \sin(x) + a_2 \cos(2x)$ . On veut déterminer  $(a_1^*, a_2^*)$  qui minimisent le critère des moindres carrés.

$$\text{Posons } X = \begin{pmatrix} \sin(x_1) & \cos(2x_1) \\ \sin(x_2) & \cos(2x_2) \\ \dots & \dots \\ \sin(x_n) & \cos(2x_n) \end{pmatrix} \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} .$$

De  ${}^tXX \begin{pmatrix} a_1^* \\ a_2^* \end{pmatrix} = {}^tXY$ , on obtient

$$\begin{pmatrix} \sum_{i=1}^n \sin^2(x_i) & \sum_{i=1}^n \sin(x_i) \cos(2x_i) \\ \sum_{i=1}^n \sin(x_i) \cos(2x_i) & \sum_{i=1}^n \cos^2(2x_i) \end{pmatrix} \begin{pmatrix} a_1^* \\ a_2^* \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n \sin(x_i) y_i \\ \sum_{i=1}^n \cos(2x_i) y_i \end{pmatrix}$$

d'où le résultat (à  $10^{-2}$  près)

$$\begin{pmatrix} a_1^* \\ a_2^* \end{pmatrix} = \begin{pmatrix} 1.5 \\ 3 \end{pmatrix}.$$

De  ${}^tXX \begin{pmatrix} a_1^* \\ a_2^* \end{pmatrix} = {}^tXY$ , on obtient

$$\begin{pmatrix} \sum_{i=1}^n \sin^2(x_i) & \sum_{i=1}^n \sin(x_i) \cos(2x_i) \\ \sum_{i=1}^n \sin(x_i) \cos(2x_i) & \sum_{i=1}^n \cos^2(2x_i) \end{pmatrix} \begin{pmatrix} a_1^* \\ a_2^* \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n \sin(x_i) y_i \\ \sum_{i=1}^n \cos(2x_i) y_i \end{pmatrix}$$

d'où le résultat (à  $10^{-2}$  près)

$$\begin{pmatrix} a_1^* \\ a_2^* \end{pmatrix} = \begin{pmatrix} 1.5 \\ 3 \end{pmatrix}.$$



De  ${}^tXX \begin{pmatrix} a_1^* \\ a_2^* \end{pmatrix} = {}^tXY$ , on obtient

$$\begin{pmatrix} \sum_{i=1}^n \sin^2(x_i) & \sum_{i=1}^n \sin(x_i) \cos(2x_i) \\ \sum_{i=1}^n \sin(x_i) \cos(2x_i) & \sum_{i=1}^n \cos^2(2x_i) \end{pmatrix} \begin{pmatrix} a_1^* \\ a_2^* \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n \sin(x_i) y_i \\ \sum_{i=1}^n \cos(2x_i) y_i \end{pmatrix}$$

d'où le résultat (à  $10^{-2}$  près)

$$\begin{pmatrix} a_1^* \\ a_2^* \end{pmatrix} = \begin{pmatrix} 1.5 \\ 3 \end{pmatrix}.$$

# Plan

- 1 Introduction
- 2 Critère des moindres carrés : cas linéaire
  - Cas de la droite
  - Généralisation
- 3 Critère des moindres carrés : cas non linéaire
  - **introduction**
  - Linéarisation de problèmes non linéaires
  - Première méthode itérative: méthode de Newton
  - Autres méthodes itératives
  - Autre méthode: algorithme évolutionnaire
  - Le point de vue du statisticien

Supposons que la relation liant  $x$  et  $y$  soit

$$y = f(x, a_1, a_2, \dots, a_p).$$

Le critère des moindres carrés est alors:

$$S = \sum_{i=1}^n (y_i - f(x_i, a_1, a_2, \dots, a_p))^2. \text{ Le minimum doit vérifié}$$

$\frac{\partial S}{\partial a_k}(a_1, a_2, \dots, a_p) = 0 \quad k = 1, \dots, p.$  Or les dérivées partielles sont

$$\frac{\partial S}{\partial a_k} = -2 \sum_{i=1}^n (y_i - f(x_i, a_1, a_2, \dots, a_p)) \frac{\partial f}{\partial a_k} \quad k = 1, \dots, p,$$

ce qui conduit à un système d'équations non linéaires dont en général on ne connaît pas de solution exacte.

Supposons que la relation liant  $x$  et  $y$  soit

$$y = f(x, a_1, a_2, \dots, a_p).$$

Le critère des moindres carrés est alors:

$$S = \sum_{i=1}^n (y_i - f(x_i, a_1, a_2, \dots, a_p))^2. \text{ Le minimum doit vérifié}$$

$\frac{\partial S}{\partial a_k}(a_1, a_2, \dots, a_p) = 0 \quad k = 1, \dots, p.$  Or les dérivées partielles sont

$$\frac{\partial S}{\partial a_k} = -2 \sum_{i=1}^n (y_i - f(x_i, a_1, a_2, \dots, a_p)) \frac{\partial f}{\partial a_k} \quad k = 1, \dots, p,$$

ce qui conduit à un système d'équations non linéaires dont en général on ne connaît pas de solution exacte.

Supposons que la relation liant  $x$  et  $y$  soit

$$y = f(x, a_1, a_2, \dots, a_p).$$

Le critère des moindres carrés est alors:

$$S = \sum_{i=1}^n (y_i - f(x_i, a_1, a_2, \dots, a_p))^2. \text{ Le minimum doit vérifié}$$

$\frac{\partial S}{\partial a_k}(a_1, a_2, \dots, a_p) = 0 \quad k = 1, \dots, p.$  Or les dérivées partielles sont

$$\frac{\partial S}{\partial a_k} = -2 \sum_{i=1}^n (y_i - f(x_i, a_1, a_2, \dots, a_p)) \frac{\partial f}{\partial a_k} \quad k = 1, \dots, p,$$

ce qui conduit à un système d'équations non linéaires dont en général on ne connaît pas de solution exacte.

Supposons que la relation liant  $x$  et  $y$  soit

$$y = f(x, a_1, a_2, \dots, a_p).$$

Le critère des moindres carrés est alors:

$$S = \sum_{i=1}^n (y_i - f(x_i, a_1, a_2, \dots, a_p))^2. \text{ Le minimum doit vérifié}$$

$\frac{\partial S}{\partial a_k}(a_1, a_2, \dots, a_p) = 0 \quad k = 1, \dots, p.$  Or les dérivées partielles sont

$$\frac{\partial S}{\partial a_k} = -2 \sum_{i=1}^n (y_i - f(x_i, a_1, a_2, \dots, a_p)) \frac{\partial f}{\partial a_k} \quad k = 1, \dots, p,$$

ce qui conduit à un système d'équations non linéaires dont en général on ne connaît pas de solution exacte.

La résolution dans le cas où le problème est non linéaire est donc plus complexe. Par le passé, ne disposant pas de calculateur (calculatrice ou même ordinateur) diverses méthodes ont été développées: en effet certains problèmes non linéaires peuvent être transformés en un problème linéaire par un changement de variables. Cela permet de se ramener à des problèmes déjà traités. Comme ces méthodes demeurent assez utilisées, nous les présentons ici. Néanmoins elles se justifient moins et il faut bien comprendre qu'alors on minimise autre chose que les écarts entre les points expérimentaux et ceux du modèle théorique.

La résolution dans le cas où le problème est non linéaire est donc plus complexe. Par le passé, ne disposant pas de calculateur (calculatrice ou même ordinateur) diverses méthodes ont été développées: en effet certains problèmes non linéaires peuvent être transformés en un problème linéaire par un changement de variables. Cela permet de se ramener à des problèmes déjà traités. Comme ces méthodes demeurent assez utilisées, nous les présentons ici. Néanmoins elles se justifient moins et il faut bien comprendre qu'alors on minimise autre chose que les écarts entre les points expérimentaux et ceux du modèle théorique.



La résolution dans le cas où le problème est non linéaire est donc plus complexe. Par le passé, ne disposant pas de calculateur (calculatrice ou même ordinateur) diverses méthodes ont été développées: en effet certains problèmes non linéaires peuvent être transformés en un problème linéaire par un changement de variables. Cela permet de se ramener à des problèmes déjà traités. Comme ces méthodes demeurent assez utilisées, nous les présentons ici. Néanmoins elles se justifient moins et il faut bien comprendre qu'alors on minimise autre chose que les écarts entre les points expérimentaux et ceux du modèle théorique.

La résolution dans le cas où le problème est non linéaire est donc plus complexe. Par le passé, ne disposant pas de calculateur (calculatrice ou même ordinateur) diverses méthodes ont été développées: en effet certains problèmes non linéaires peuvent être transformés en un problème linéaire par un changement de variables. Cela permet de se ramener à des problèmes déjà traités. Comme ces méthodes demeurent assez utilisées, nous les présentons ici. Néanmoins elles se justifient moins et il faut bien comprendre qu'alors on minimise autre chose que les écarts entre les points expérimentaux et ceux du modèle théorique.

La résolution dans le cas où le problème est non linéaire est donc plus complexe. Par le passé, ne disposant pas de calculateur (calculatrice ou même ordinateur) diverses méthodes ont été développées: en effet certains problèmes non linéaires peuvent être transformés en un problème linéaire par un changement de variables. Cela permet de se ramener à des problèmes déjà traités. Comme ces méthodes demeurent assez utilisées, nous les présentons ici. Néanmoins elles se justifient moins et il faut bien comprendre qu'alors on minimise autre chose que les écarts entre les points expérimentaux et ceux du modèle théorique.

# Plan

- 1 Introduction
- 2 Critère des moindres carrés : cas linéaire
  - Cas de la droite
  - Généralisation
- 3 Critère des moindres carrés : cas non linéaire
  - introduction
  - **Linéarisation de problèmes non linéaires**
  - Première méthode itérative: méthode de Newton
  - Autres méthodes itératives
  - Autre méthode: algorithme évolutionnaire
  - Le point de vue du statisticien

## Exemple

*Modèle exponentiel*

*Supposons que la relation théorique entre  $y$  et  $x$  soit*

$$y = a_1 \exp(a_2 x) \quad (2).$$

Posons  $f(x, a_1, a_2) = a_1 \exp(a_2 x)$ . On a

$$\frac{\partial f}{\partial a_1} = \exp(a_2 x)$$

et

$$\frac{\partial f}{\partial a_2} = x a_1 \exp(a_2 x)$$

Définissons

$$S(a_1, a_2) = \sum_{i=1}^n (y_i - a_1 \exp(a_2 x_i))^2$$

## Exemple

*Modèle exponentiel*

*Supposons que la relation théorique entre  $y$  et  $x$  soit*

$$y = a_1 \exp(a_2 x) \quad (2).$$

Posons  $f(x, a_1, a_2) = a_1 \exp(a_2 x)$ . On a

$$\frac{\partial f}{\partial a_1} = \exp(a_2 x)$$

et

$$\frac{\partial f}{\partial a_2} = x a_1 \exp(a_2 x)$$

Définissons

$$S(a_1, a_2) = \sum_{i=1}^n (y_i - a_1 \exp(a_2 x_i))^2$$

## Exemple

*Modèle exponentiel*

*Supposons que la relation théorique entre  $y$  et  $x$  soit*

$$y = a_1 \exp(a_2 x) \quad (2).$$

Posons  $f(x, a_1, a_2) = a_1 \exp(a_2 x)$ . On a

$$\frac{\partial f}{\partial a_1} = \exp(a_2 x)$$

et

$$\frac{\partial f}{\partial a_2} = x a_1 \exp(a_2 x)$$

Définissons

$$S(a_1, a_2) = \sum_{i=1}^n (y_i - a_1 \exp(a_2 x_i))^2$$

## Exemple

*Modèle exponentiel*

*Supposons que la relation théorique entre  $y$  et  $x$  soit*

$$y = a_1 \exp(a_2 x) \quad (2).$$

Posons  $f(x, a_1, a_2) = a_1 \exp(a_2 x)$ . On a

$$\frac{\partial f}{\partial a_1} = \exp(a_2 x)$$

et

$$\frac{\partial f}{\partial a_2} = x a_1 \exp(a_2 x)$$

Définissons

$$S(a_1, a_2) = \sum_{i=1}^n (y_i - a_1 \exp(a_2 x_i))^2$$



On cherche les points critiques de  $S(x, a_1, a_2)$  :

$$\begin{cases} \frac{\partial S}{\partial a_0} = -2 \sum_{i=1}^n (y_i - a_1 \exp(a_2 x_i)) \exp(a_2 x_i) \\ \frac{\partial S}{\partial a_1} = -2 \sum_{i=1}^n (y_i - a_1 \exp(a_2 x_i)) x_i a_1 \exp(a_2 x_i) \end{cases}$$

système non linéaire dont on ne connaît pas de solution exacte.

Cependant ce problème peut être transformé en un problème linéaire par un changement de variable. Posons  $z = \ln(y)$ , en prenant le logarithme de la relation (2), on obtient :

$$z = \ln(a_1) + a_2 x$$

qui est une relation linéaire.

On cherche les points critiques de  $S(x, a_1, a_2)$  :

$$\begin{cases} \frac{\partial S}{\partial a_0} = -2 \sum_{i=1}^n (y_i - a_1 \exp(a_2 x_i)) \exp(a_2 x_i) \\ \frac{\partial S}{\partial a_1} = -2 \sum_{i=1}^n (y_i - a_1 \exp(a_2 x_i)) x_i a_1 \exp(a_2 x_i) \end{cases}$$

système non linéaire dont on ne connaît pas de solution exacte. Cependant ce problème peut être transformé en un problème linéaire par un changement de variable. Posons  $z = \ln(y)$ , en prenant le logarithme de la relation (2), on obtient :

$$z = \ln(a_1) + a_2 x$$

qui est une relation linéaire.

On cherche les points critiques de  $S(x, a_1, a_2)$  :

$$\begin{cases} \frac{\partial S}{\partial a_0} = -2 \sum_{i=1}^n (y_i - a_1 \exp(a_2 x_i)) \exp(a_2 x_i) \\ \frac{\partial S}{\partial a_1} = -2 \sum_{i=1}^n (y_i - a_1 \exp(a_2 x_i)) x_i a_1 \exp(a_2 x_i) \end{cases}$$

système non linéaire dont on ne connaît pas de solution exacte. Cependant ce problème peut être transformé en un problème linéaire par un changement de variable. Posons  $z = \ln(y)$ , en prenant le logarithme de la relation (2), on obtient :

$$z = \ln(a_1) + a_2 x$$

qui est une relation linéaire.

## Exemple

*L'évolution de la concentration d'oxygène dissous dans un fermenteur peut être modélisée par*

$$\frac{dC_{O_2}^{th}}{dt} = K_{la}(C_{O_2}^* - C_{O_2}^{th})$$

*Pour déterminer le paramètre  $K_{la}$ , on peut effectuer le changement de variable  $z = \ln(C_{O_2}^* - C_{O_2}^{th})$ .*

Mais il est préférable de travailler sans effectuer cette transformation...

## Exemple

*L'évolution de la concentration d'oxygène dissous dans un fermenteur peut être modélisée par*

$$\frac{dC_{O_2}^{th}}{dt} = K_{la}(C_{O_2}^* - C_{O_2}^{th})$$

*Pour déterminer le paramètre  $K_{la}$ , on peut effectuer le changement de variable  $z = \ln(C_{O_2}^* - C_{O_2}^{th})$ .*

Mais il est préférable de travailler sans effectuer cette transformation...

## Exemple (Modèle de Mickaëlis-Menten-Henri: détermination des constantes $V_M$ et $K_M$ d'une enzyme E)

La réaction modélisée est  $S + E \rightarrow P$ . On note  $v$  la vitesse d'apparition du produit  $P$  en fonction de la concentration initiale de substrat notée  $S$ . On a

$$v(s) = \frac{V_M s}{K_M + s}$$

Trois linéarisations sont utilisées:

- 1 Lineweaver et Burck (la plus fréquente):  $X = \frac{1}{s}$  et  $Y = \frac{1}{v}$
- 2 Headie-Hofstee:  $X = v$  et  $Y = \frac{v}{s}$
- 3 Hanes-Woolf:  $X = s$  et  $Y = \frac{s}{v}$

Le problème de ces méthodes de changement de variables dans les systèmes non linéaires est que l'on ne minimise plus

la quantité  $S = \sum_{i=1}^n (y_i - f(x_i, a_1, a_2, \dots, a_p))^2$  mais une

quantité "transformée". Aussi ces méthodes qui ont été assez utilisées par le passé (lorsque l'on ne disposait pas d'autres outils) se justifient moins de nos jours. En effet grâce à la puissance des calculateurs et la mise au point de nombreuses méthodes de minimisation, on peut obtenir avec des logiciels de meilleurs résultats.

Le problème de ces méthodes de changement de variables dans les systèmes non linéaires est que l'on ne minimise plus

la quantité  $S = \sum_{i=1}^n (y_i - f(x_i, a_1, a_2, \dots, a_p))^2$  mais une

quantité "transformée". Aussi ces méthodes qui ont été assez utilisées par le passé (lorsque l'on ne disposait pas d'autres outils) se justifient moins de nos jours. En effet grâce à la puissance des calculateurs et la mise au point de nombreuses méthodes de minimisation, on peut obtenir avec des logiciels de meilleurs résultats.



Le problème de ces méthodes de changement de variables dans les systèmes non linéaires est que l'on ne minimise plus la quantité  $S = \sum_{i=1}^n (y_i - f(x_i, a_1, a_2, \dots, a_p))^2$  mais une quantité "transformée". Aussi ces méthodes qui ont été assez utilisées par le passé (lorsque l'on ne disposait pas d'autres outils) se justifient moins de nos jours. En effet grâce à la puissance des calculateurs et la mise au point de nombreuses méthodes de minimisation, on peut obtenir avec des logiciels de meilleurs résultats.

# Plan

- 1 Introduction
- 2 Critère des moindres carrés : cas linéaire
  - Cas de la droite
  - Généralisation
- 3 Critère des moindres carrés : cas non linéaire
  - introduction
  - Linéarisation de problèmes non linéaires
  - **Première méthode itérative: méthode de Newton**
  - Autres méthodes itératives
  - Autre méthode: algorithme évolutionnaire
  - Le point de vue du statisticien

Puisque l'on ne sait pas calculer la solution  $a^*$  de manière explicite, l'idée est de construire une suite qui tend vers  $a^*$ . Ce type de méthodes s'appelle méthodes itératives.

Commençons par la **dimension 1** ie on ne cherche qu'un paramètre  $a^* \in \mathbb{R}$ .

On cherche  $a^* \in \mathbb{R}$  tel que  $S'(a^*) = 0$  (points critiques de  $S$  fonction de  $\mathbb{R} \mapsto \mathbb{R}$ ). Posons  $f = S'$ . On cherche donc les zéros de  $f$  (une fonction assez régulière dont on suppose qu'elle s'annule...). Résoudre  $f(a^*) = 0$  n'est en général pas possible par un calcul explicite, mais il existe plusieurs méthodes pour approcher cette valeur.

Puisque l'on ne sait pas calculer la solution  $a^*$  de manière explicite, l'idée est de construire une suite qui tend vers  $a^*$ . Ce type de méthodes s'appelle méthodes itératives.

Commençons par la **dimension 1** ie on ne cherche qu'un paramètre  $a^* \in \mathbb{R}$ .

On cherche  $a^* \in \mathbb{R}$  tel que  $S'(a^*) = 0$  (points critiques de  $S$  fonction de  $\mathbb{R} \mapsto \mathbb{R}$ ). Posons  $f = S'$ . On cherche donc les zéros de  $f$  (une fonction assez régulière dont on suppose qu'elle s'annule...). Résoudre  $f(a^*) = 0$  n'est en général pas possible par un calcul explicite, mais il existe plusieurs méthodes pour approcher cette valeur.

Puisque l'on ne sait pas calculer la solution  $a^*$  de manière explicite, l'idée est de construire une suite qui tend vers  $a^*$ . Ce type de méthodes s'appelle méthodes itératives.

Commençons par la **dimension 1** ie on ne cherche qu'un paramètre  $a^* \in \mathbb{R}$ .

On cherche  $a^* \in \mathbb{R}$  tel que  $S'(a^*) = 0$  (points critiques de  $S$  fonction de  $\mathbb{R} \mapsto \mathbb{R}$ ). Posons  $f = S'$ . On cherche donc les zéros de  $f$  (une fonction assez régulière dont on suppose qu'elle s'annule...). Résoudre  $f(a^*) = 0$  n'est en général pas possible par un calcul explicite, mais il existe plusieurs méthodes pour approcher cette valeur.

Puisque l'on ne sait pas calculer la solution  $a^*$  de manière explicite, l'idée est de construire une suite qui tend vers  $a^*$ . Ce type de méthodes s'appelle méthodes itératives.

Commençons par la **dimension 1** ie on ne cherche qu'un paramètre  $a^* \in \mathbb{R}$ .

On cherche  $a^* \in \mathbb{R}$  tel que  $S'(a^*) = 0$  (points critiques de  $S$  fonction de  $\mathbb{R} \mapsto \mathbb{R}$ ). Posons  $f = S'$ . On cherche donc les zéros de  $f$  (une fonction assez régulière dont on suppose qu'elle s'annule...). Résoudre  $f(a^*) = 0$  n'est en général pas possible par un calcul explicite, mais il existe plusieurs méthodes pour approcher cette valeur.

Puisque l'on ne sait pas calculer la solution  $a^*$  de manière explicite, l'idée est de construire une suite qui tend vers  $a^*$ . Ce type de méthodes s'appelle méthodes itératives.

Commençons par la **dimension 1** ie on ne cherche qu'un paramètre  $a^* \in \mathbb{R}$ .

On cherche  $a^* \in \mathbb{R}$  tel que  $S'(a^*) = 0$  (points critiques de  $S$  fonction de  $\mathbb{R} \mapsto \mathbb{R}$ ). Posons  $f = S'$ . On cherche donc les zéros de  $f$  (une fonction assez régulière dont on suppose qu'elle s'annule...). Résoudre  $f(a^*) = 0$  n'est en général pas possible par un calcul explicite, mais il existe plusieurs méthodes pour approcher cette valeur.

Puisque l'on ne sait pas calculer la solution  $a^*$  de manière explicite, l'idée est de construire une suite qui tend vers  $a^*$ . Ce type de méthodes s'appelle méthodes itératives.

Commençons par la **dimension 1** ie on ne cherche qu'un paramètre  $a^* \in \mathbb{R}$ .

On cherche  $a^* \in \mathbb{R}$  tel que  $S'(a^*) = 0$  (points critiques de  $S$  fonction de  $\mathbb{R} \mapsto \mathbb{R}$ ). Posons  $f = S'$ . On cherche donc les zéros de  $f$  (une fonction assez régulière dont on suppose qu'elle s'annule...). Résoudre  $f(a^*) = 0$  n'est en général pas possible par un calcul explicite, mais il existe plusieurs méthodes pour approcher cette valeur.



## Newton en dimension 1

Pour chercher  $a^* \in \mathbb{R}$  tel que  $f(a^*) = 0$ , on va construire une suite  $a_0, a_1, a_2, \dots, a_k, \dots$  tendant vers  $a^*$ . On suppose que l'on connaît  $a^0$  proche de  $a^*$ . On pose  $a_0 = a^0$ . Puis on écrit la formule de Taylor à l'ordre 1 de  $f$  au voisinage de  $a_0$  :

$$f(a) = f(a_0) + (a - a_0)f'(a_0) + (a - a_0)\epsilon(a - a_0).$$

on en déduit l'équation de la tangente en  $a_0$  à la courbe représentative de  $f$ ,

$$y = f(a_0) + (a - a_0)f'(a_0).$$

L'idée est d'approcher  $f$  par sa tangente.

## Newton en dimension 1

Pour chercher  $a^* \in \mathbb{R}$  tel que  $f(a^*) = 0$ , on va construire une suite  $a_0, a_1, a_2, \dots, a_k, \dots$  tendant vers  $a^*$ . On suppose que l'on connaît  $a^0$  proche de  $a^*$ . On pose  $a_0 = a^0$ . Puis on écrit la formule de Taylor à l'ordre 1 de  $f$  au voisinage de  $a_0$  :

$$f(a) = f(a_0) + (a - a_0)f'(a_0) + (a - a_0)\epsilon(a - a_0).$$

on en déduit l'équation de la tangente en  $a_0$  à la courbe représentative de  $f$ ,

$$y = f(a_0) + (a - a_0)f'(a_0).$$

L'idée est d'approcher  $f$  par sa tangente.

## Newton en dimension 1

Pour chercher  $a^* \in \mathbb{R}$  tel que  $f(a^*) = 0$ , on va construire une suite  $a_0, a_1, a_2, \dots, a_k, \dots$  tendant vers  $a^*$ . On suppose que l'on connaît  $a^0$  proche de  $a^*$ . On pose  $a_0 = a^0$ . Puis on écrit la formule de Taylor à l'ordre 1 de  $f$  au voisinage de  $a_0$  :

$$f(a) = f(a_0) + (a - a_0)f'(a_0) + (a - a_0)\epsilon(a - a_0).$$

on en déduit l'équation de la tangente en  $a_0$  à la courbe représentative de  $f$ ,

$$y = f(a_0) + (a - a_0)f'(a_0).$$

L'idée est d'approcher  $f$  par sa tangente.

## Newton en dimension 1

Pour chercher  $a^* \in \mathbb{R}$  tel que  $f(a^*) = 0$ , on va construire une suite  $a_0, a_1, a_2, \dots, a_k, \dots$  tendant vers  $a^*$ . On suppose que l'on connaît  $a^0$  proche de  $a^*$ . On pose  $a_0 = a^0$ . Puis on écrit la formule de Taylor à l'ordre 1 de  $f$  au voisinage de  $a_0$  :

$$f(a) = f(a_0) + (a - a_0)f'(a_0) + (a - a_0)\epsilon(a - a_0).$$

on en déduit l'équation de la tangente en  $a_0$  à la courbe représentative de  $f$ ,

$$y = f(a_0) + (a - a_0)f'(a_0).$$

L'idée est d'approcher  $f$  par sa tangente.

## Newton en dimension 1

Pour chercher  $a^* \in \mathbb{R}$  tel que  $f(a^*) = 0$ , on va construire une suite  $a_0, a_1, a_2, \dots, a_k, \dots$  tendant vers  $a^*$ . On suppose que l'on connaît  $a^0$  proche de  $a^*$ . On pose  $a_0 = a^0$ . Puis on écrit la formule de Taylor à l'ordre 1 de  $f$  au voisinage de  $a_0$  :

$$f(a) = f(a_0) + (a - a_0)f'(a_0) + (a - a_0)\epsilon(a - a_0).$$

on en déduit l'équation de la tangente en  $a_0$  à la courbe représentative de  $f$ ,

$$y = f(a_0) + (a - a_0)f'(a_0).$$

L'idée est d'approcher  $f$  par sa tangente.

L'équation de la tangente en  $a_0$  est  $y = f(a_0) + (a - a_0)f'(a_0)$ .  
On détermine alors le point d'intersection  $(a_1, 0)$  de la tangente  
avec l'axe des abscisses :  $f(a_0) + (a_1 - a_0)f'(a_0) = 0$  soit

$$a_1 = a_0 - \frac{f(a_0)}{f'(a_0)}.$$

L'équation de la tangente en  $a_0$  est  $y = f(a_0) + (a - a_0)f'(a_0)$ .  
On détermine alors le point d'intersection  $(a_1, 0)$  de la tangente  
avec l'axe des abscisses :  $f(a_0) + (a_1 - a_0)f'(a_0) = 0$  soit  
$$a_1 = a_0 - \frac{f(a_0)}{f'(a_0)}.$$

L'équation de la tangente en  $a_0$  est  $y = f(a_0) + (a - a_0)f'(a_0)$ .  
On détermine alors le point d'intersection  $(a_1, 0)$  de la tangente  
avec l'axe des abscisses :  $f(a_0) + (a_1 - a_0)f'(a_0) = 0$  soit

$$a_1 = a_0 - \frac{f(a_0)}{f'(a_0)}.$$



Puis on réitère le processus en approchant  $f$  en  $a_1$  par sa tangente et ainsi de suite. On construit ainsi

$$a_{k+1} = a_k - \frac{f(a_k)}{f'(a_k)}.$$

ce qui est équivalent à

$$a_{k+1} = a_k - \frac{S'(a_k)}{S''(a_k)}.$$

Pour que cela soit satisfaisant, il faut bien sûr que

$\lim_{k \rightarrow +\infty} (a_{k+1} - a_k) = 0$  et  $\lim_{k \rightarrow +\infty} f(a_k) = 0$ , c'est-à-dire que la suite  $(a_k)_{k \in \mathbb{N}}$  converge vers  $a^*$ .

Puis on réitère le processus en approchant  $f$  en  $a_1$  par sa tangente et ainsi de suite. On construit ainsi

$$a_{k+1} = a_k - \frac{f(a_k)}{f'(a_k)}.$$

ce qui est équivalent à

$$a_{k+1} = a_k - \frac{S'(a_k)}{S''(a_k)}.$$

Pour que cela soit satisfaisant, il faut bien sûr que

$\lim_{k \rightarrow +\infty} (a_{k+1} - a_k) = 0$  et  $\lim_{k \rightarrow +\infty} f(a_k) = 0$ , c'est-à-dire que la suite  $(a_k)_{k \in \mathbb{N}}$  converge vers  $a^*$ .

Puis on réitère le processus en approchant  $f$  en  $a_1$  par sa tangente et ainsi de suite. On construit ainsi

$$a_{k+1} = a_k - \frac{f(a_k)}{f'(a_k)}.$$

ce qui est équivalent à

$$a_{k+1} = a_k - \frac{S'(a_k)}{S''(a_k)}.$$

Pour que cela soit satisfaisant, il faut bien sûr que

$\lim_{k \rightarrow +\infty} (a_{k+1} - a_k) = 0$  et  $\lim_{k \rightarrow +\infty} f(a_k) = 0$ , c'est-à-dire que la suite  $(a_k)_{k \in \mathbb{N}}$  converge vers  $a^*$ .

## Newton cas multidimensionnel

Le principe dans le cas multidimensionnel est le même. Mais on cherche un vecteur  $\theta^*$  composé de  $p$  paramètres  $(\theta_1^*, \dots, \theta_p^*)$ . Dans ce qui suit, nous allons donc construire une suite d'éléments de  $\mathbb{R}^p$ . Il ne faut pas confondre les indices de la suite et les indices de composantes,  $\theta_k$  représente un vecteur de  $p$  composantes  $\theta_k = (\theta_{k1}, \theta_{k2}, \dots, \theta_{kp})$  mais ce sera aussi un élément d'une suite  $\theta_0, \theta_1, \dots$  chaque élément étant un vecteur de  $\mathbb{R}^p$ .

## Newton cas multidimensionnel

Le principe dans le cas multidimensionnel est le même. Mais on cherche un vecteur  $\theta^*$  composé de  $p$  paramètres  $(\theta_1^*, \dots, \theta_p^*)$ . Dans ce qui suit, nous allons donc construire une suite d'éléments de  $\mathbb{R}^p$ . Il ne faut pas confondre les indices de la suite et les indices de composantes,  $\theta_k$  représente un vecteur de  $p$  composantes  $\theta_k = (\theta_{k1}, \theta_{k2}, \dots, \theta_{kp})$  mais ce sera aussi un élément d'une suite  $\theta_0, \theta_1, \dots$  chaque élément étant un vecteur de  $\mathbb{R}^p$ .

## Newton cas multidimensionnel

Le principe dans le cas multidimensionnel est le même. Mais on cherche un vecteur  $\theta^*$  composé de  $p$  paramètres  $(\theta_1^*, \dots, \theta_p^*)$ . Dans ce qui suit, nous allons donc construire une suite d'éléments de  $\mathbb{R}^p$ . Il ne faut pas confondre les indices de la suite et les indices de composantes,  $\theta_k$  représente un vecteur de  $p$  composantes  $\theta_k = (\theta_{k1}, \theta_{k2}, \dots, \theta_{kp})$  mais ce sera aussi un élément d'une suite  $\theta_0, \theta_1, \dots$  chaque élément étant un vecteur de  $\mathbb{R}^p$ .

## Newton cas multidimensionnel

Le principe dans le cas multidimensionnel est le même. Mais on cherche un vecteur  $\theta^*$  composé de  $p$  paramètres  $(\theta_1^*, \dots, \theta_p^*)$ . Dans ce qui suit, nous allons donc construire une suite d'éléments de  $\mathbb{R}^p$ . Il ne faut pas confondre les indices de la suite et les indices de composantes,  $\theta_k$  représente un vecteur de  $p$  composantes  $\theta_k = (\theta_{k1}, \theta_{k2}, \dots, \theta_{kp})$  mais ce sera aussi un élément d'une suite  $\theta_0, \theta_1, \dots$  chaque élément étant un vecteur de  $\mathbb{R}^p$ .

On cherche les points critiques de  $S$  fonction de  $\mathbb{R}^p \mapsto \mathbb{R}$  ie  $\theta^* \in \mathbb{R}^p$  tel que  $S'(\theta^*) = 0$ . On écrit la formule de Taylor à l'ordre 1 de  $\nabla S$  au voisinage de  $\theta_0$  :

$$\nabla S(\theta) = \nabla S(\theta_0) + H_S(\theta_0)(\theta - \theta_0) + (\theta - \theta_0)\epsilon(\theta - \theta_0).$$

Par conséquent, à partir d'un point  $\theta_0$ , on construit  $\theta_1$  vérifiant  $\nabla S(\theta_1) = 0$  d'où  $0 = \nabla S(\theta_0) + H_S(\theta_0)(\theta_1 - \theta_0)$

$$\theta_1 = \theta_0 - H_S^{-1}(\theta_0)\nabla S(\theta_0).$$

et ainsi on construit une suite à partir de la formule

$$\theta_{k+1} = \theta_k - H_S^{-1}(\theta_k)\nabla S(\theta_k).$$

qui converge vers un minimum de  $S$



On cherche les points critiques de  $S$  fonction de  $\mathbb{R}^p \mapsto \mathbb{R}$  ie  $\theta^* \in \mathbb{R}^p$  tel que  $S'(\theta^*) = 0$ . On écrit la formule de Taylor à l'ordre 1 de  $\nabla S$  au voisinage de  $\theta_0$  :

$$\nabla S(\theta) = \nabla S(\theta_0) + H_S(\theta_0)(\theta - \theta_0) + (\theta - \theta_0)\epsilon(\theta - \theta_0).$$

Par conséquent, à partir d'un point  $\theta_0$ , on construit  $\theta_1$  vérifiant  $\nabla S(\theta_1) = 0$  d'où  $0 = \nabla S(\theta_0) + H_S(\theta_0)(\theta_1 - \theta_0)$

$$\theta_1 = \theta_0 - H_S^{-1}(\theta_0)\nabla S(\theta_0).$$

et ainsi on construit une suite à partir de la formule

$$\theta_{k+1} = \theta_k - H_S^{-1}(\theta_k)\nabla S(\theta_k).$$

qui converge vers un minimum de  $S$

On cherche les points critiques de  $S$  fonction de  $\mathbb{R}^p \mapsto \mathbb{R}$  ie  $\theta^* \in \mathbb{R}^p$  tel que  $S'(\theta^*) = 0$ . On écrit la formule de Taylor à l'ordre 1 de  $\nabla S$  au voisinage de  $\theta_0$  :

$$\nabla S(\theta) = \nabla S(\theta_0) + H_S(\theta_0)(\theta - \theta_0) + (\theta - \theta_0)\epsilon(\theta - \theta_0).$$

Par conséquent, à partir d'un point  $\theta_0$ , on construit  $\theta_1$  vérifiant  $\nabla S(\theta_1) = 0$  d'où  $0 = \nabla S(\theta_0) + H_S(\theta_0)(\theta_1 - \theta_0)$

$$\theta_1 = \theta_0 - H_S^{-1}(\theta_0)\nabla S(\theta_0).$$

et ainsi on construit une suite à partir de la formule

$$\theta_{k+1} = \theta_k - H_S^{-1}(\theta_k)\nabla S(\theta_k).$$

qui converge vers un minimum de  $S$

On cherche les points critiques de  $S$  fonction de  $\mathbb{R}^p \mapsto \mathbb{R}$  ie  $\theta^* \in \mathbb{R}^p$  tel que  $S'(\theta^*) = 0$ . On écrit la formule de Taylor à l'ordre 1 de  $\nabla S$  au voisinage de  $\theta_0$  :

$$\nabla S(\theta) = \nabla S(\theta_0) + H_S(\theta_0)(\theta - \theta_0) + (\theta - \theta_0)\epsilon(\theta - \theta_0).$$

Par conséquent, à partir d'un point  $\theta_0$ , on construit  $\theta_1$  vérifiant  $\nabla S(\theta_1) = 0$  d'où  $0 = \nabla S(\theta_0) + H_S(\theta_0)(\theta_1 - \theta_0)$

$$\theta_1 = \theta_0 - H_S^{-1}(\theta_0)\nabla S(\theta_0).$$

et ainsi on construit une suite à partir de la formule

$$\theta_{k+1} = \theta_k - H_S^{-1}(\theta_k)\nabla S(\theta_k).$$

qui converge vers un minimum de  $S$

On cherche les points critiques de  $S$  fonction de  $\mathbb{R}^p \mapsto \mathbb{R}$  ie  $\theta^* \in \mathbb{R}^p$  tel que  $S'(\theta^*) = 0$ . On écrit la formule de Taylor à l'ordre 1 de  $\nabla S$  au voisinage de  $\theta_0$  :

$$\nabla S(\theta) = \nabla S(\theta_0) + H_S(\theta_0)(\theta - \theta_0) + (\theta - \theta_0)\epsilon(\theta - \theta_0).$$

Par conséquent, à partir d'un point  $\theta_0$ , on construit  $\theta_1$  vérifiant  $\nabla S(\theta_1) = 0$  d'où  $0 = \nabla S(\theta_0) + H_S(\theta_0)(\theta_1 - \theta_0)$

$$\theta_1 = \theta_0 - H_S^{-1}(\theta_0)\nabla S(\theta_0).$$

et ainsi on construit une suite à partir de la formule

$$\theta_{k+1} = \theta_k - H_S^{-1}(\theta_k)\nabla S(\theta_k).$$

qui converge vers un minimum de  $S$

On cherche les points critiques de  $S$  fonction de  $\mathbb{R}^p \mapsto \mathbb{R}$  ie  $\theta^* \in \mathbb{R}^p$  tel que  $S'(\theta^*) = 0$ . On écrit la formule de Taylor à l'ordre 1 de  $\nabla S$  au voisinage de  $\theta_0$  :

$$\nabla S(\theta) = \nabla S(\theta_0) + H_S(\theta_0)(\theta - \theta_0) + (\theta - \theta_0)\epsilon(\theta - \theta_0).$$

Par conséquent, à partir d'un point  $\theta_0$ , on construit  $\theta_1$  vérifiant  $\nabla S(\theta_1) = 0$  d'où  $0 = \nabla S(\theta_0) + H_S(\theta_0)(\theta_1 - \theta_0)$

$$\theta_1 = \theta_0 - H_S^{-1}(\theta_0)\nabla S(\theta_0).$$

et ainsi on construit une suite à partir de la formule

$$\theta_{k+1} = \theta_k - H_S^{-1}(\theta_k)\nabla S(\theta_k).$$

qui converge vers un minimum de  $S$

## Remarque

*Dans la pratique le calcul à chaque étape de l'inverse de la hessienne de  $S$  est coûteux. De plus cette méthode peut diverger. Aussi, on utilise plutôt des méthodes de quasi-Newton, de gradient ou de Levenberg-Marquardt.*

# Plan

- 1 Introduction
- 2 Critère des moindres carrés : cas linéaire
  - Cas de la droite
  - Généralisation
- 3 Critère des moindres carrés : cas non linéaire
  - introduction
  - Linéarisation de problèmes non linéaires
  - Première méthode itérative: méthode de Newton
  - **Autres méthodes itératives**
  - Autre méthode: algorithme évolutionnaire
  - Le point de vue du statisticien

## Définition

Soit  $f : \mathbb{R}^p \rightarrow \mathbb{R}$ . On dit qu'un vecteur  $d \in \mathbb{R}^p$  est une direction de descente pour  $f$  au point  $\theta \in \mathbb{R}^p$  si

$$\forall s > 0 \quad \exists \alpha \in ]0; s[ \quad f(\theta + \alpha d) < f(\theta).$$

## Théorème

Soit  $f : \mathbb{R}^p \rightarrow \mathbb{R}$ . Supposons  $f$  dérivable au point  $\theta$ . Soit  $d \in \mathbb{R}^p$ . Si  $\nabla f(\theta) \cdot d < 0$  alors  $d$  est une direction de descente pour  $f$  en  $\theta$ .



## Définition

Soit  $f : \mathbb{R}^p \rightarrow \mathbb{R}$ . On dit qu'un vecteur  $d \in \mathbb{R}^p$  est une direction de descente pour  $f$  au point  $\theta \in \mathbb{R}^p$  si

$$\forall s > 0 \quad \exists \alpha \in ]0; s[ \quad f(\theta + \alpha d) < f(\theta).$$

## Théorème

Soit  $f : \mathbb{R}^p \rightarrow \mathbb{R}$ . Supposons  $f$  dérivable au point  $\theta$ . Soit  $d \in \mathbb{R}^p$ . Si  $\nabla f(\theta) \cdot d < 0$  alors  $d$  est une direction de descente pour  $f$  en  $\theta$ .

## Remarque

- *Si  $\theta^*$  est un minimum local alors il n'existe aucune direction de descente pour  $f$  en  $\theta^*$ .*
- *Réciproquement s'il n'existe aucune direction de descente pour  $f$  en  $\theta^*$  alors  $\theta^*$  est un point critique. Si de plus  $H_f(\theta^*) \neq 0$  (point critique non dégénéré) alors  $\theta^*$  est un minimum strict.*
- *En un point critique dégénéré, il peut n'y avoir aucune direction de descente sans pour autant que ce point soit un minimum local (ex  $f(x, y) = 2x^4 - 3x^2y + y^2$  n'admet aucune direction de descente en  $(0, 0)$  bien que ce ne soit pas un minimum local).*

## Remarque

- *Si  $\theta^*$  est un minimum local alors il n'existe aucune direction de descente pour  $f$  en  $\theta^*$ .*
- *Réciproquement s'il n'existe aucune direction de descente pour  $f$  en  $\theta^*$  alors  $\theta^*$  est un point critique. Si de plus  $H_f(\theta^*) \neq 0$  (point critique non dégénéré) alors  $\theta^*$  est un minimum strict.*
- *En un point critique dégénéré, il peut n'y avoir aucune direction de descente sans pour autant que ce point soit un minimum local (ex  $f(x, y) = 2x^4 - 3x^2y + y^2$  n'admet aucune direction de descente en  $(0, 0)$  bien que ce ne soit pas un minimum local).*

## Remarque

- Si  $\theta^*$  est un minimum local alors il n'existe aucune direction de descente pour  $f$  en  $\theta^*$ .
- Réciproquement s'il n'existe aucune direction de descente pour  $f$  en  $\theta^*$  alors  $\theta^*$  est un point critique. Si de plus  $H_f(\theta^*) \neq 0$  (point critique non dégénéré) alors  $\theta^*$  est un minimum strict.
- En un point critique dégénéré, il peut n'y avoir aucune direction de descente sans pour autant que ce point soit un minimum local (ex  $f(x, y) = 2x^4 - 3x^2y + y^2$  n'admet aucune direction de descente en  $(0, 0)$  bien que ce ne soit pas un minimum local).

On appelle méthode de descente une méthode générant une suite  $(\theta_k)$

$$\theta_{k+1} = \theta_k + \alpha_k d_k$$

où  $d_k$  est une direction de descente c'est-à-dire vérifiant  $\nabla S(\theta_k) \cdot d_k < 0$  et où  $\alpha_k$  est un réel appelé pas de descente. Les algorithmes des méthodes de descente se généralisent donc de la manière suivante

- initialisation: choix de  $\theta_0$
- étape  $k$ 
  - détermination d'une direction de descente  $d_k \in \mathbb{R}^p$
  - choix d'un pas de descente  $\alpha_k \in \mathbb{R}$
  - $\theta_{k+1} = \theta_k + \alpha_k d_k$
- Test d'arrêt :  $k < k+1$  (et retourner à l'étape  $k$ ) ou fin

On appelle méthode de descente une méthode générant une suite  $(\theta_k)$

$$\theta_{k+1} = \theta_k + \alpha_k d_k$$

où  $d_k$  est une direction de descente c'est-à-dire vérifiant  $\nabla S(\theta_k) \cdot d_k < 0$  et où  $\alpha_k$  est un réel appelé pas de descente. Les algorithmes des méthodes de descente se généralisent donc de la manière suivante

- initialisation: choix de  $\theta_0$
- étape  $k$ 
  - détermination d'une direction de descente  $d_k \in \mathbb{R}^p$
  - choix d'un pas de descente  $\alpha_k \in \mathbb{R}$
  - $\theta_{k+1} = \theta_k + \alpha_k d_k$
- Test d'arrêt :  $k < k+1$  (et retourner à l'étape  $k$ ) ou fin

On appelle méthode de descente une méthode générant une suite  $(\theta_k)$

$$\theta_{k+1} = \theta_k + \alpha_k d_k$$

où  $d_k$  est une direction de descente c'est-à-dire vérifiant  $\nabla S(\theta_k) \cdot d_k < 0$  et où  $\alpha_k$  est un réel appelé pas de descente. Les algorithmes des méthodes de descente se généralisent donc de la manière suivante

- initialisation: choix de  $\theta_0$
- étape  $k$ 
  - détermination d'une direction de descente  $d_k \in \mathbb{R}^p$
  - choix d'un pas de descente  $\alpha_k \in \mathbb{R}$
  - $\theta_{k+1} = \theta_k + \alpha_k d_k$
- Test d'arrêt :  $k < k+1$  (et retourner à l'étape  $k$ ) ou fin

On appelle méthode de descente une méthode générant une suite  $(\theta_k)$

$$\theta_{k+1} = \theta_k + \alpha_k d_k$$

où  $d_k$  est une direction de descente c'est-à-dire vérifiant  $\nabla S(\theta_k) \cdot d_k < 0$  et où  $\alpha_k$  est un réel appelé pas de descente. Les algorithmes des méthodes de descente se généralisent donc de la manière suivante

- initialisation: choix de  $\theta_0$
- étape  $k$ 
  - détermination d'une direction de descente  $d_k \in \mathbb{R}^p$
  - choix d'un pas de descente  $\alpha_k \in \mathbb{R}$
  - $\theta_{k+1} = \theta_k + \alpha_k d_k$
- Test d'arrêt :  $k < k+1$  (et retourner à l'étape  $k$ ) ou fin



On appelle méthode de descente une méthode générant une suite  $(\theta_k)$

$$\theta_{k+1} = \theta_k + \alpha_k d_k$$

où  $d_k$  est une direction de descente c'est-à-dire vérifiant  $\nabla S(\theta_k) \cdot d_k < 0$  et où  $\alpha_k$  est un réel appelé pas de descente. Les algorithmes des méthodes de descente se généralisent donc de la manière suivante

- initialisation: choix de  $\theta_0$
- étape  $k$ 
  - détermination d'une direction de descente  $d_k \in \mathbb{R}^p$
  - choix d'un pas de descente  $\alpha_k \in \mathbb{R}$
  - $\theta_{k+1} = \theta_k + \alpha_k d_k$
- Test d'arrêt :  $k < k+1$  (et retourner à l'étape  $k$ ) ou fin

On appelle méthode de descente une méthode générant une suite  $(\theta_k)$

$$\theta_{k+1} = \theta_k + \alpha_k d_k$$

où  $d_k$  est une direction de descente c'est-à-dire vérifiant  $\nabla S(\theta_k) \cdot d_k < 0$  et où  $\alpha_k$  est un réel appelé pas de descente. Les algorithmes des méthodes de descente se généralisent donc de la manière suivante

- initialisation: choix de  $\theta_0$
- étape  $k$   
détermination d'une direction de descente  $d_k \in \mathbb{R}^p$   
choix d'un pas de descente  $\alpha_k \in \mathbb{R}$   
 $\theta_{k+1} = \theta_k + \alpha_k d_k$
- Test d'arrêt :  $k < k+1$  (et retourner à l'étape  $k$ ) ou fin

On appelle méthode de descente une méthode générant une suite  $(\theta_k)$

$$\theta_{k+1} = \theta_k + \alpha_k d_k$$

où  $d_k$  est une direction de descente c'est-à-dire vérifiant  $\nabla S(\theta_k) \cdot d_k < 0$  et où  $\alpha_k$  est un réel appelé pas de descente. Les algorithmes des méthodes de descente se généralisent donc de la manière suivante

- initialisation: choix de  $\theta_0$
- étape  $k$   
détermination d'une direction de descente  $d_k \in \mathbb{R}^p$   
choix d'un pas de descente  $\alpha_k \in \mathbb{R}$   
 $\theta_{k+1} = \theta_k + \alpha_k d_k$
- Test d'arrêt :  $k < k+1$  (et retourner à l'étape  $k$ ) ou fin

Il existe deux grandes familles de stratégies de la direction de descente  $d_k$  au point  $\theta_k$ :

- stratégie de Newton  $d_k = -H_S^{-1}(\theta_k)\nabla S(\theta_k)$
- stratégie de Gradient  $d_k = -\nabla S(\theta_k)$

C'est une méthode dérivée du Gradient (Gradient Réduit Généralisé) qu'on retrouvera implémenter dans le solveur d'excel.

Il existe deux grandes familles de stratégies de la direction de descente  $d_k$  au point  $\theta_k$ :

- stratégie de Newton  $d_k = -H_S^{-1}(\theta_k)\nabla S(\theta_k)$
- stratégie de Gradient  $d_k = -\nabla S(\theta_k)$

C'est une méthode dérivée du Gradient (Gradient Réduit Généralisé) qu'on retrouvera implémenter dans le solveur d'excel.

Il existe deux grandes familles de stratégies de la direction de descente  $d_k$  au point  $\theta_k$ :

- stratégie de Newton  $d_k = -H_S^{-1}(\theta_k)\nabla S(\theta_k)$
- stratégie de Gradient  $d_k = -\nabla S(\theta_k)$

C'est une méthode dérivée du Gradient (Gradient Réduit Généralisé) qu'on retrouvera implémenter dans le solveur d'excel.

## Méthode de Gradient

Les méthodes dites de gradient consistent à déterminer  $d_k$  en fonction de  $\nabla S(\theta_k)$  souvent  $d_k = -\nabla S(\theta_k)$  direction opposée au gradient (la plus forte pente). Il s'agit bien d'une direction de descente en effet

$\nabla S(\theta_k) \cdot d_k = \nabla S(\theta_k) \cdot (-\nabla S(\theta_k)) = -\|\nabla S(\theta_k)\|^2 < 0$ . Pour le pas de descente  $\alpha_k$ , il existe diverses stratégies:

- soit fixe  $\alpha_k = \text{constante}$
- soit prédéfini  $\alpha_k = \frac{1}{k}$
- soit calculé en cherchant le minimum d'une fonction d'une variable, le réel  $\alpha_k$  vérifiant

$$S(\theta_k + \alpha_k d_k) = \min_{\alpha \in \mathbb{R}} S(\theta_k + \alpha d_k)$$

## Méthode de Gradient

Les méthodes dites de gradient consistent à déterminer  $d_k$  en fonction de  $\nabla S(\theta_k)$  souvent  $d_k = -\nabla S(\theta_k)$  direction opposée au gradient (la plus forte pente). Il s'agit bien d'une direction de descente en effet

$\nabla S(\theta_k) \cdot d_k = \nabla S(\theta_k) \cdot (-\nabla S(\theta_k)) = -\|\nabla S(\theta_k)\|^2 < 0$ . Pour le pas de descente  $\alpha_k$ , il existe diverses stratégies:

- soit fixe  $\alpha_k = \text{constante}$
- soit prédéfini  $\alpha_k = \frac{1}{k}$
- soit calculé en cherchant le minimum d'une fonction d'une variable, le réel  $\alpha_k$  vérifiant

$$S(\theta_k + \alpha_k d_k) = \min_{\alpha \in \mathbb{R}} S(\theta_k + \alpha d_k)$$



## Méthode de Gradient

Les méthodes dites de gradient consistent à déterminer  $d_k$  en fonction de  $\nabla S(\theta_k)$  souvent  $d_k = -\nabla S(\theta_k)$  direction opposée au gradient (la plus forte pente). Il s'agit bien d'une direction de descente en effet

$\nabla S(\theta_k) \cdot d_k = \nabla S(\theta_k) \cdot (-\nabla S(\theta_k)) = -\|\nabla S(\theta_k)\|^2 < 0$ . Pour le pas de descente  $\alpha_k$ , il existe diverses stratégies:

- soit fixe  $\alpha_k = \text{constante}$
- soit prédéfini  $\alpha_k = \frac{1}{k}$
- soit calculé en cherchant le minimum d'une fonction d'une variable, le réel  $\alpha_k$  vérifiant

$$S(\theta_k + \alpha_k d_k) = \min_{\alpha \in \mathbb{R}} S(\theta_k + \alpha d_k)$$

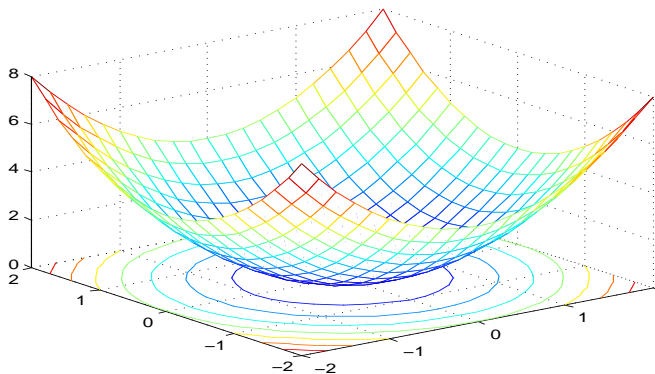
## Méthode de Gradient

Les méthodes dites de gradient consistent à déterminer  $d_k$  en fonction de  $\nabla S(\theta_k)$  souvent  $d_k = -\nabla S(\theta_k)$  direction opposée au gradient (la plus forte pente). Il s'agit bien d'une direction de descente en effet

$\nabla S(\theta_k) \cdot d_k = \nabla S(\theta_k) \cdot (-\nabla S(\theta_k)) = -\|\nabla S(\theta_k)\|^2 < 0$ . Pour le pas de descente  $\alpha_k$ , il existe diverses stratégies:

- soit fixe  $\alpha_k = \text{constante}$
- soit prédéfini  $\alpha_k = \frac{1}{k}$
- soit calculé en cherchant le minimum d'une fonction d'une variable, le réel  $\alpha_k$  vérifiant

$$S(\theta_k + \alpha_k d_k) = \min_{\alpha \in \mathbb{R}} S(\theta_k + \alpha d_k)$$



Introduction

Critère des moindres carrés : cas linéaire

Critère des moindres carrés : cas non linéaire

introduction

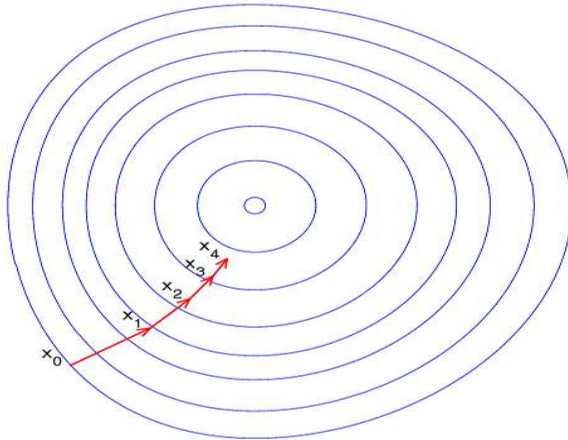
Linéarisation de problèmes non linéaires

Première méthode itérative: méthode de Newton

**Autres méthodes itératives**

Autre méthode: algorithme évolutionnaire

Le point de vue du statisticien



## Exemple

On dispose du tableau de mesures  $(x_i, y_i)$  suivant :

$x_i$	1	2	5	10	15	30
$y_i$	0.95	1.5	1.55	1.75	1.9	1.95

Un modèle de relation entre  $x$  et  $y$  suggéré est  $y = \frac{V_M x}{K_M + x}$ .

On veut déterminer  $(K_M^*, V_M^*)$  qui minimisent le critère des moindres carrés.

Introduction

Critère des moindres carrés : cas linéaire

Critère des moindres carrés : cas non linéaire

introduction

Linéarisation de problèmes non linéaires

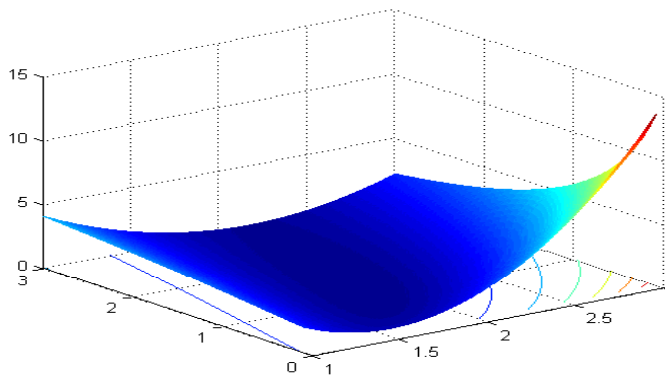
Première méthode itérative: méthode de Newton

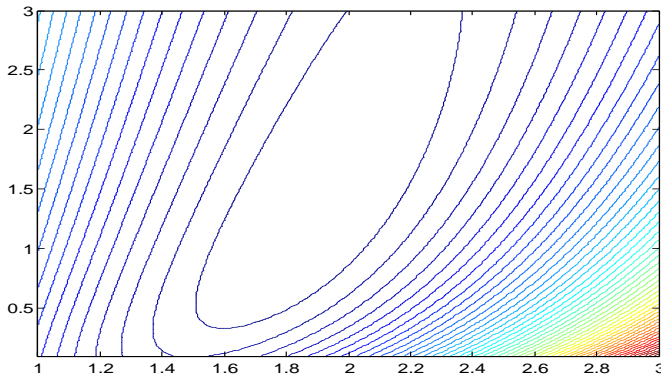
**Autres méthodes itératives**

Autre méthode: algorithme évolutionnaire

Le point de vue du statisticien

avec excel...







## Méthode de Levenberg-Marquardt

Levenberg et Marquardt ont proposé une méthode pour passer du schéma avec inversion de la hessienne à celui des plus fortes pentes. Ce dernier est utilisé loin du minimum puis il est remplacé par le schéma d'inversion de la hessienne à mesure qu'on s'approche du minimum. Il existe aussi d'autres méthodes dérivant de Newton et du Gradient: quasi-newton, Gauss-Newton, gradient conjugué,...

## Méthode de Levenberg-Marquardt

Levenberg et Marquardt ont proposé une méthode pour passer du schéma avec inversion de la hessienne à celui des plus fortes pentes. Ce dernier est utilisé loin du minimum puis il est remplacé par le schéma d'inversion de la hessienne à mesure qu'on s'approche du minimum. Il existe aussi d'autres méthodes dérivant de Newton et du Gradient: quasi-newton, Gauss-Newton, gradient conjugué,...

## Comparaison des différentes méthodes

Les méthodes de Gradient sont des méthodes robustes: tant que l'algorithme n'a pas trouvé un point critique, la valeur du critère décroît strictement à chaque itération mais elles sont lentes. De plus l'algorithme peut rencontrer un certain nombre de problèmes de convergence si le minimum est "au fond d'une vallée étroite": la suite peut osciller de part et d'autre de la vallée et progresse laborieusement (beaucoup d'itérations et pas de descente long à déterminer à chaque étape).

## Comparaison des différentes méthodes

Les méthodes de Newton peuvent diverger. En effet  $d_k = -H_S^{-1}(\theta_k)\nabla S(\theta_k)$  est une direction de descente dès que la hessienne  $H_S(\theta_k)$  est définie positive ce qui est toujours vérifiée si  $\theta_k$  est suffisamment proche d'un minimum local (non dégénéré) de  $f$ . Lorsqu'ils convergent ces algorithmes de type Newton sont plus rapides que les algorithmes de Gradient mais plus coûteux (calcul de l'inverse de la hessienne) mais moins robustes (divergence possible). Loin du minimum  $d_k$  n'est plus nécessairement une direction de descente.

# Plan

- 1 Introduction
- 2 Critère des moindres carrés : cas linéaire
  - Cas de la droite
  - Généralisation
- 3 Critère des moindres carrés : cas non linéaire
  - introduction
  - Linéarisation de problèmes non linéaires
  - Première méthode itérative: méthode de Newton
  - Autres méthodes itératives
  - **Autre méthode: algorithme évolutionnaire**
  - Le point de vue du statisticien

Les phénomènes physiques ou biologiques ont été à la source de nombreux algorithmes s'en inspirant plus ou moins librement. Ainsi les réseaux de neurones artificiels s'inspirent du fonctionnement du cerveau humain, l'algorithme de recuit simulé de la thermodynamique, et les algorithmes évolutionnaires (AEs) (dont les plus connus sont les algorithmes génétiques) de l'évolution darwinienne des populations biologiques.

Soit à optimiser une fonction  $S$  à valeurs réelles définie sur un espace  $W$ . Le parallèle avec l'évolution naturelle a entraîné l'apparition d'un vocabulaire spécifique :

- La fonction objectif  $S$  est appelée fonction performance, ou fonction d'adaptation (fitness en anglais)
- Les points de l'espace de recherche  $W$  sont appelés des individus
- Les P-uplets d'individus sont appelés des populations
- On parlera d'une génération pour la boucle principale de l'algorithme

Soit à optimiser une fonction  $S$  à valeurs réelles définie sur un espace  $W$ . Le parallèle avec l'évolution naturelle a entraîné l'apparition d'un vocabulaire spécifique :

- La fonction objectif  $S$  est appelée fonction performance, ou fonction d'adaptation (fitness en anglais)
- Les points de l'espace de recherche  $W$  sont appelés des individus
- Les  $P$ -uplets d'individus sont appelés des populations
- On parlera d'une génération pour la boucle principale de l'algorithme



Soit à optimiser une fonction  $S$  à valeurs réelles définie sur un espace  $W$ . Le parallèle avec l'évolution naturelle a entraîné l'apparition d'un vocabulaire spécifique :

- La fonction objectif  $S$  est appelée fonction performance, ou fonction d'adaptation (fitness en anglais)
- Les points de l'espace de recherche  $W$  sont appelés des individus
- Les P-uplets d'individus sont appelés des populations
- On parlera d'une génération pour la boucle principale de l'algorithme

Soit à optimiser une fonction  $S$  à valeurs réelles définie sur un espace  $W$ . Le parallèle avec l'évolution naturelle a entraîné l'apparition d'un vocabulaire spécifique :

- La fonction objectif  $S$  est appelée fonction performance, ou fonction d'adaptation (fitness en anglais)
- Les points de l'espace de recherche  $W$  sont appelés des individus
- Les P-uplets d'individus sont appelés des populations
- On parlera d'une génération pour la boucle principale de l'algorithme

L'algorithme: La pression de l'environnement est simulée à l'aide de la fonction  $S$ , et le principe darwinien, "Les plus adaptés survivent et se reproduisent", est implanté de la manière suivante :

- Initialisation de la population  $P_0$  (tirage aléatoire)
- Evaluation des individus de  $P_0$  (i.e. calcul de  $S$ ) ;
- Construction de la population  $P_i$  à partir de  $P_{i-1}$  :
  - Sélection des individus les plus performants (au sens de  $S$ ) de  $P_{i-1}$  ;
  - Génération de nouveaux individus, les enfants (croisement et/ou mutation)
  - Evaluation des enfants ;
  - Remplacement des parents au moyen d'une sélection parmi les enfants

Arrêt si niveau souhaité atteint ou nombre fixé de générations

L'algorithme: La pression de l'environnement est simulée à l'aide de la fonction  $S$ , et le principe darwinien, "Les plus adaptés survivent et se reproduisent", est implanté de la manière suivante :

- Initialisation de la population  $P_0$  (tirage aléatoire)
- Evaluation des individus de  $P_0$  (i.e. calcul de  $S$ ) ;
- Construction de la population  $P_i$  à partir de  $P_{i-1}$  :
  - Sélection des individus les plus performants (au sens de  $S$ ) de  $P_{i-1}$  ;
  - Génération de nouveaux individus, les enfants (croisement et/ou mutation)
  - Evaluation des enfants ;
  - Remplacement des parents au moyen d'une sélection parmi les enfants

Arrêt si niveau souhaité atteint ou nombre fixé de générations

L'algorithme: La pression de l'environnement est simulée à l'aide de la fonction  $S$ , et le principe darwinien, "Les plus adaptés survivent et se reproduisent", est implanté de la manière suivante :

- Initialisation de la population  $P_0$  (tirage aléatoire)
- Evaluation des individus de  $P_0$  (i.e. calcul de  $S$ ) ;
- Construction de la population  $P_i$  à partir de  $P_{i-1}$  :
  - Sélection des individus les plus performants (au sens de  $S$ ) de  $P_{i-1}$  ;
  - Génération de nouveaux individus, les enfants (croisement et/ou mutation)
  - Evaluation des enfants ;
  - Remplacement des parents au moyen d'une sélection parmi les enfants

Arrêt si niveau souhaité atteint ou nombre fixé de générations

L'algorithme: La pression de l'environnement est simulée à l'aide de la fonction  $S$ , et le principe darwinien, "Les plus adaptés survivent et se reproduisent", est implanté de la manière suivante :

- Initialisation de la population  $P_0$  (tirage aléatoire)
- Evaluation des individus de  $P_0$  (i.e. calcul de  $S$ ) ;
- Construction de la population  $P_i$  à partir de  $P_{i-1}$  :
  - Sélection des individus les plus performants (au sens de  $S$ ) de  $P_{i-1}$  ;
  - Génération de nouveaux individus, les enfants (croisement et/ou mutation)
  - Evaluation des enfants ;
  - Remplacement des parents au moyen d'une sélection parmi les enfants

Arrêt si niveau souhaité atteint ou nombre fixé de générations

L'algorithme: La pression de l'environnement est simulée à l'aide de la fonction  $S$ , et le principe darwinien, "Les plus adaptés survivent et se reproduisent", est implanté de la manière suivante :

- Initialisation de la population  $P_0$  (tirage aléatoire)
- Evaluation des individus de  $P_0$  (i.e. calcul de  $S$ ) ;
- Construction de la population  $P_i$  à partir de  $P_{i-1}$  :
  - Sélection des individus les plus performants (au sens de  $S$ ) de  $P_{i-1}$  ;
  - Génération de nouveaux individus, les enfants (croisement et/ou mutation)
  - Evaluation des enfants ;
  - Remplacement des parents au moyen d'une sélection parmi les enfants

Arrêt si niveau souhaité atteint ou nombre fixé de générations

L'algorithme: La pression de l'environnement est simulée à l'aide de la fonction  $S$ , et le principe darwinien, "Les plus adaptés survivent et se reproduisent", est implanté de la manière suivante :

- Initialisation de la population  $P_0$  (tirage aléatoire)
- Evaluation des individus de  $P_0$  (i.e. calcul de  $S$ ) ;
- Construction de la population  $P_i$  à partir de  $P_{i-1}$  :
  - Sélection des individus les plus performants (au sens de  $S$ ) de  $P_{i-1}$  ;
  - Génération de nouveaux individus, les enfants (croisement et/ou mutation)
  - Evaluation des enfants ;
  - Remplacement des parents au moyen d'une sélection parmi les enfants

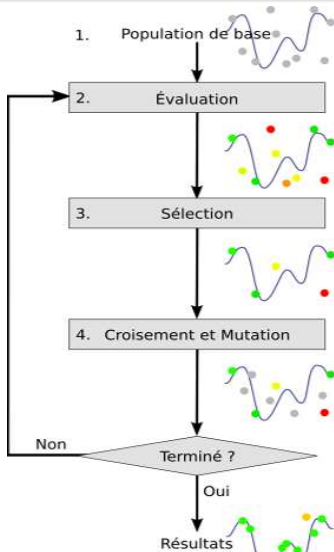
Arrêt si niveau souhaité atteint ou nombre fixé de générations



L'algorithme: La pression de l'environnement est simulée à l'aide de la fonction  $S$ , et le principe darwinien, "Les plus adaptés survivent et se reproduisent", est implanté de la manière suivante :

- Initialisation de la population  $P_0$  (tirage aléatoire)
- Evaluation des individus de  $P_0$  (i.e. calcul de  $S$ ) ;
- Construction de la population  $P_i$  à partir de  $P_{i-1}$  :
  - Sélection des individus les plus performants (au sens de  $S$ ) de  $P_{i-1}$  ;
  - Génération de nouveaux individus, les enfants (croisement et/ou mutation)
  - Evaluation des enfants ;
  - Remplacement des parents au moyen d'une sélection parmi les enfants

Arrêt si niveau souhaité atteint ou nombre fixé de générations



# Plan

- 1 Introduction
- 2 Critère des moindres carrés : cas linéaire
  - Cas de la droite
  - Généralisation
- 3 Critère des moindres carrés : cas non linéaire
  - introduction
  - Linéarisation de problèmes non linéaires
  - Première méthode itérative: méthode de Newton
  - Autres méthodes itératives
  - Autre méthode: algorithme évolutionnaire
  - Le point de vue du statisticien

On sait maintenant obtenir des paramètres en minimisant le critère des moindres carrés. On vérifiera l'adéquation du modèle en étudiant les résidus (où  $r_i = y_i - y_{th(x_i)}$ ) en particulier leur répartition.

Si le modèle est jugé pertinent, l'objectif du statisticien sera de préciser la qualité de ces valeurs.

On sait maintenant obtenir des paramètres en minimisant le critère des moindres carrés. On vérifiera l'adéquation du modèle en étudiant les résidus (où  $r_i = y_i - y_{th(x_i)}$ ) en particulier leur répartition.

Si le modèle est jugé pertinent, l'objectif du statisticien sera de préciser la qualité de ces valeurs.