

Estimation par moindres carrés

David Coulot

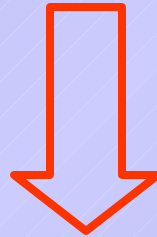
*Cours de 1^{ère} année du cycle
des ingénieurs de l'ENSG*

E-mail : David.Coulot@ensg.ign.fr

Domaines métrologiques



Compréhension des phénomènes physiques



Mesures

Modélisation $\vec{y} \approx f(\vec{x})$

But d'une méthode d'optimisation

Trouver les paramètres \vec{x}

Problèmes

Modèle en général approximatif

Mesures entachées d'erreurs aléatoires

Critères de qualité des estimations des paramètres ?

⇒ **Cadre mathématique**

Estimation statistique

Méthode des moindres carrés

Formulation analytique simple

Critères statistiques

mais ...

Manque de robustesse

Méthode d'estimation en norme L_1 préliminaire

Minima locaux

Ne convient pas aux fonctions modèles non différentiables

Méthodes alternatives (algorithmes génétiques,...)

Plan du cours

I – Rappels mathématiques

A – Rappels d'analyse

B – Rappels de statistiques

II – Méthode des moindres carrés

A – Exemple introductif

B – Formulation générale

C – Cas du modèle linéaire

D – Cas d'école

E – Techniques complémentaires

F – Limites de la méthode

III – Autres méthodes d'optimisation

A – Simplexe modifié

B – Algorithmes génétiques

I – Rappels mathématiques

A - Analyse

d³signing
THE FUTURE

DO NOT FEAR WHAT IS YET TO COME
KID WILL PREPARE YOU

YOU'LL BE ALL RIGHT
WE PROMISS

KULTDESIGN MANIFEST

f fonction numérique $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$

f de classe C^2

Développement de Taylor-Young d'ordre 2

$$f(a+h) = f(a) + f'(a)(h) + \frac{1}{2}Q_{f,a}(h) + o(\|h\|^2)$$

$f'(a)$: application linéaire tangente à f en a

matrice : $\left(\frac{\partial f}{\partial x_1}(a), \dots, \frac{\partial f}{\partial x_n}(a) \right)$

$Q_{f,a}$: forme quadratique fondamentale de f en a

matrice : $\left(\frac{\partial^2 f}{\partial x_i \partial x_j}(a) \right)_{1 \leq i, j \leq n}$

Avec $h = (h_1, \dots, h_n)^T$, le développement donne

$$f(a+h) = f(a) + \sum_{i=1}^n h_i \frac{\partial f}{\partial x_i}(a) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n h_i h_j \frac{\partial^2 f}{\partial x_i \partial x_j}(a) + o(\|h\|^2)$$

Extrema locaux de f

Point critique

$a \in \Omega$ est un point critique de f ssi $f'(a) = 0$

Point critique non dégénéré

Si a est point critique de f , a point critique non dégénéré ssi

$$\det\left(\frac{\partial^2 f}{\partial x_i \partial x_j}(a)\right) \neq 0$$

Extremum local

f admet un maximum (resp. un minimum) local en $a \in \Omega$

s'il existe un voisinage relatif V de a tel que

$$\forall x \in V, f(x) \leq f(a) \text{ (resp. } f(x) \geq f(a) \text{)}$$

f admet un extremum local en a ssi f admet un minimum ou un maximum local en a .

Théorèmes

- 1 - Si f fonction numérique de classe C^1 sur un ouvert Ω de R^n admet un extremum local en $a \in \Omega$ alors a est un point critique de f .
- 2 - Si f fonction numérique de classe C^2 sur un ouvert Ω de R^n admet un maximum (resp. un minimum) local en $a \in \Omega$ alors a est un point critique de f et $Q_{f,a}$ est négative (resp. positive).
- 3 - Soit a un point critique non dégénéré de f fonction numérique de classe C^2 sur Ω ouvert de R^n .
 - (i) Si $Q_{f,a}$ est définie positive, f admet un minimum local strict en a .
 - (ii) Si $Q_{f,a}$ est définie négative, f admet un maximum local strict en a .



I – Rappels mathématiques

Lois de probabilité usuelles

Loi uniforme

Variable aléatoire réelle X suit loi uniforme sur $[a, b[$ ($a < b$):

$$p_X(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in [a, b[\\ 0 & \text{si } x \notin [a, b[\end{cases} \quad E(X) = \frac{a+b}{2} \quad \text{var}(X) = \frac{(b-a)^2}{12}$$

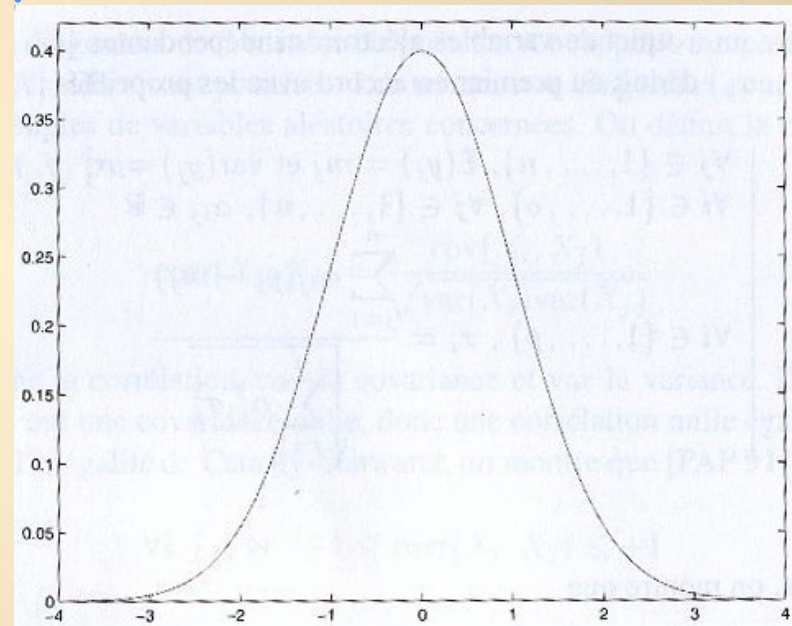
Loi normale (« gaussienne »)

Variable aléatoire réelle X suit loi normale:

$$p_X(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$E(X) = \mu$$

$$\text{var}(X) = \sigma^2$$



Loi normale standard : $\mu=0$ et $\sigma=1$

Fonctions eulériennes

Fonction eulérienne de première espèce : « fonction bêta »

$$B :] 0, \infty[\times] 0, \infty[\rightarrow \mathfrak{R} \quad B : (x, y) \mapsto B(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt$$

Fonction symétrique : $B(x, y) = B(y, x)$

Fonction eulérienne de deuxième espèce : « fonction gamma »

$$\Gamma :] 0, \infty[\rightarrow \mathfrak{R} \quad \Gamma : x \mapsto \Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt$$

Propriétés

$$\Gamma(1) = 1, \quad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

Relation de récurrence : $\Gamma(x+1) = x\Gamma(x)$, $x > 0$

$$\Gamma(n+1) = n!, \quad n \in \mathbb{N}^*$$

Loi du χ^2

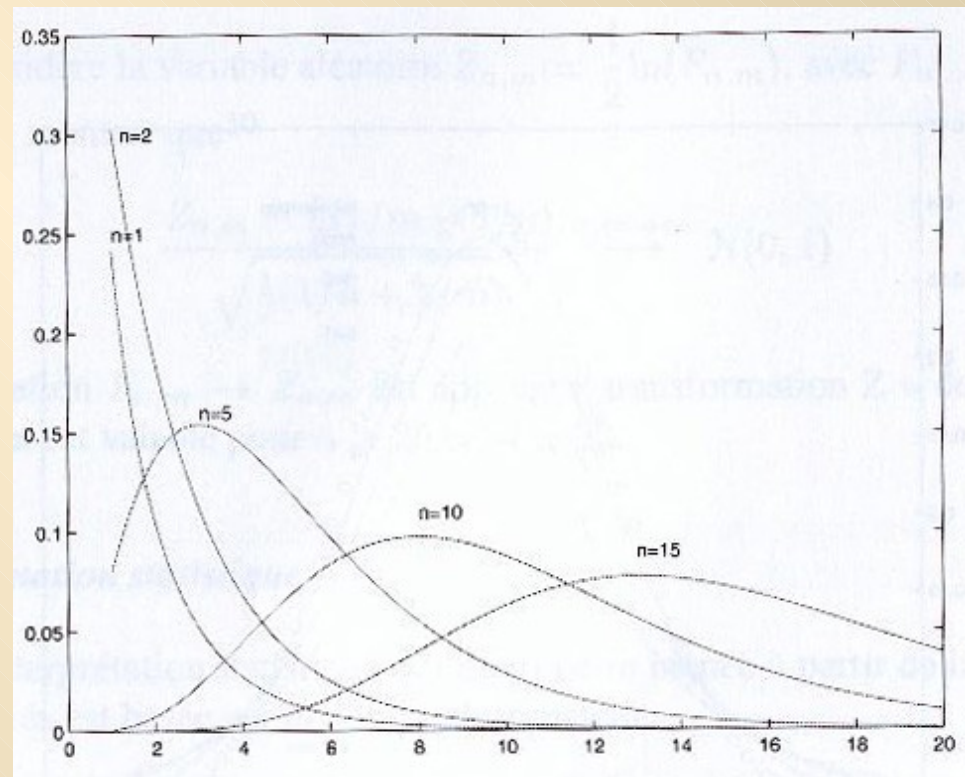
Soient n v. a. r. X_i suivant une loi normale standard, deux à deux indépendantes. La loi suivie par la variable aléatoire

$\chi^2 = \sum_{i=1}^n X_i^2$ est appelée loi du χ^2 à n degrés de liberté.

$$p_{\chi^2}(x) = \begin{cases} \frac{1}{2\Gamma\left(\frac{n}{2}\right)} \left(\frac{x}{2}\right)^{\frac{n}{2}-1} e^{-\frac{x}{2}} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

$$E(\chi^2) = n$$

$$\text{var}(\chi^2) = 2n$$



Lois du χ^2

Loi de Student

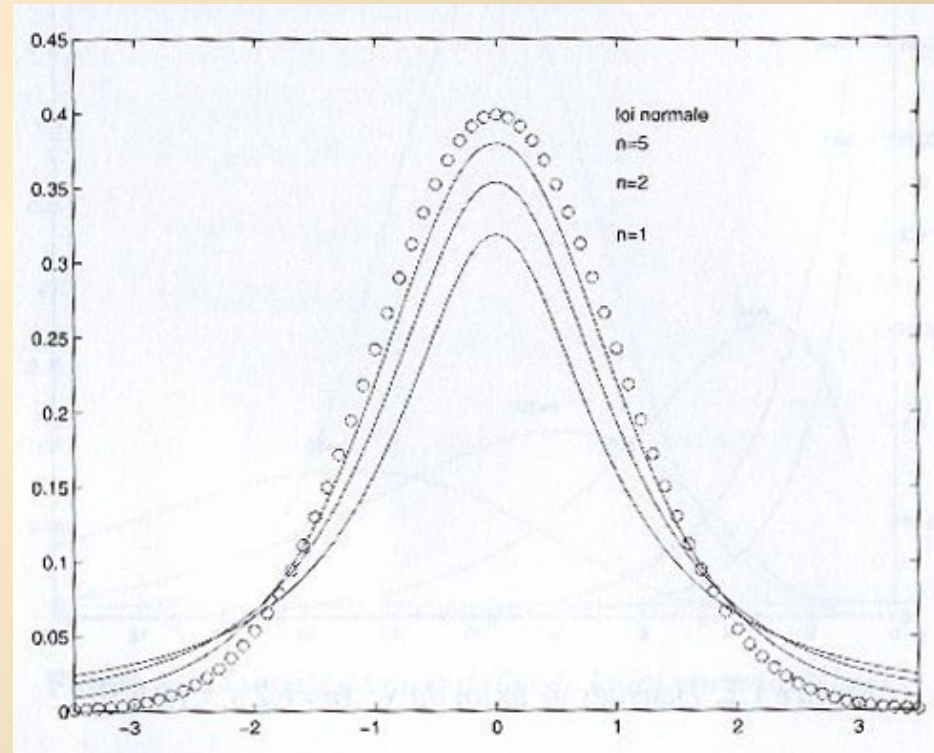
Soient X et χ^2 deux variables aléatoires indépendantes suivant respectivement les lois normale standard et du χ^2 à n degrés de liberté.

La v. a. r. $T = \frac{X}{\sqrt{\frac{\chi^2}{n}}}$ suit alors une loi de Student à n degrés de liberté.

$$p_T(x) = \frac{1}{\sqrt{n} B\left(\frac{1}{2}, \frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

$$E(T) = 0, n > 1$$

$$\text{var}(T) = \frac{n}{n-2}, n > 2$$



Lois de Student



I – Rappels mathématiques

Échantillon

Échantillon de taille $n = n$ -uplet de variables aléatoires réelles (X_1, \dots, X_n)

Ensemble (x_1, \dots, x_n) de n réalisations des variables aléatoires X_i
= réalisation ou observation de l'échantillon

Fonction $f(X_1, \dots, X_n)$ = statistique fondée sur l'échantillon

Type particulier : échantillon i.i.d. [*indépendamment et identiquement distribué*]

Fonction de vraisemblance

$$\ell(x_1, \dots, x_n, \theta) = p_{X_1, \dots, X_n}(x_1, \dots, x_n, \theta)$$

Fonction log-vraisemblance

$$L(x_1, \dots, x_n, \theta) = \ln[\ell(x_1, \dots, x_n, \theta)]$$

Fonction score

$$S(x_1, \dots, x_n, \theta) = \frac{\partial}{\partial \theta} [\ell(x_1, \dots, x_n, \theta)]$$

Information de Fisher

$$I(\theta) = E(S(X_1, \dots, X_n, \theta)^2)$$



I – Rappels mathématiques

Estimateur

Échantillon de taille n (X_1, \dots, X_n)

Statistique $f(X_1, \dots, X_n)$ pour estimer des paramètres auxquels est lié l'échantillon = estimateur

Estimateur \hat{T}_n du paramètre θ

Absence de biais

Biais de \hat{T}_n $B(\hat{T}_n) = E(\hat{T}_n) - \theta$

\hat{T}_n non biaisé si $B(\hat{T}_n) = 0$

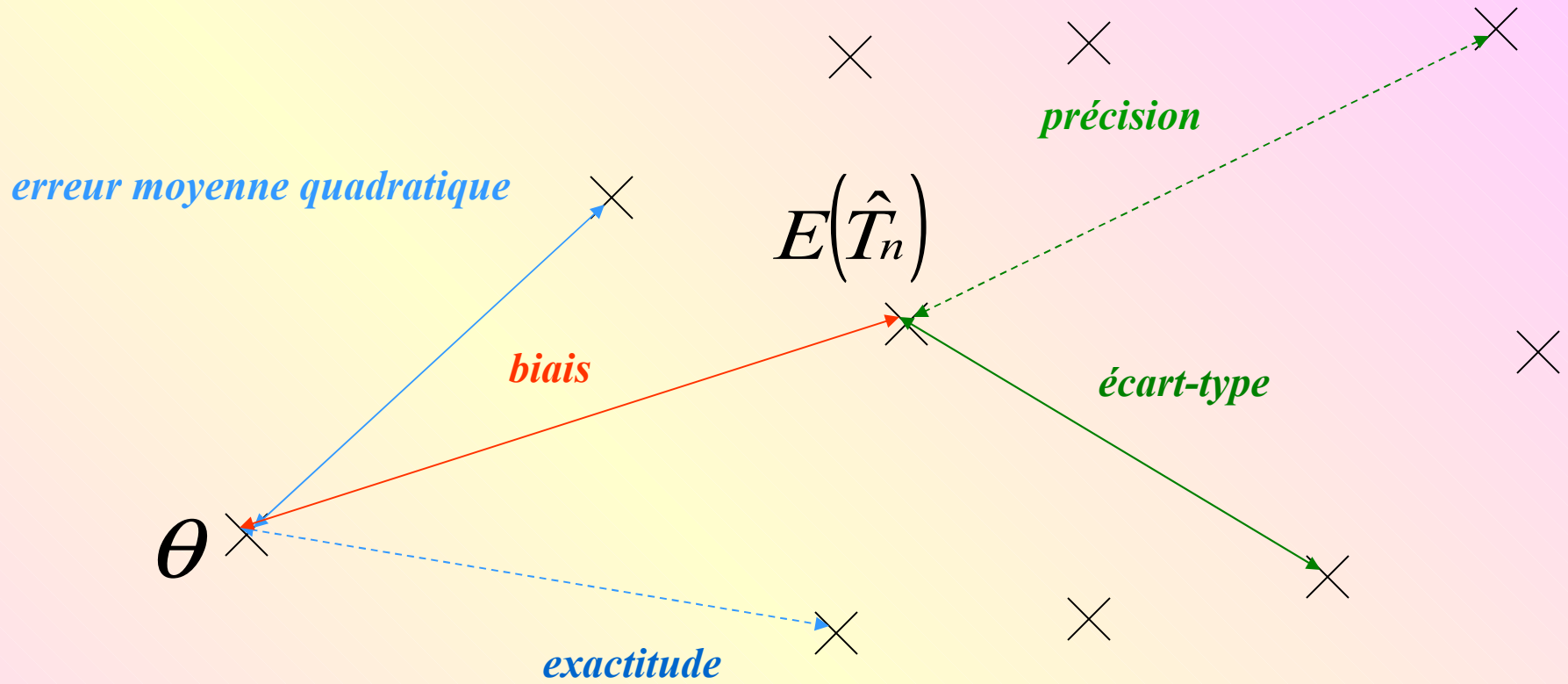
\hat{T}_n asymptotiquement sans biais si

$$\lim_{n \rightarrow \infty} B(\hat{T}_n) = 0$$

$$\sigma(\hat{T}_n) = \sqrt{E\left[(\hat{T}_n - E(\hat{T}_n))^2\right]}$$

$$\text{emq}(\hat{T}_n) = \sqrt{E\left[(\hat{T}_n - \theta)^2\right]}$$

$$\text{emq}(\hat{T}_n)^2 = \sigma(\hat{T}_n)^2 + B(\hat{T}_n)^2$$



Convergence

\hat{T}_n est convergent si

$$\forall \varepsilon > 0, \lim_{n \rightarrow \infty} P\left\{ \left| \hat{T}_n - \theta \right| \geq \varepsilon \right\} = 0$$

➔ Un estimateur convergent est toujours asymptotiquement sans biais

Estimateur optimal

\hat{T}_n est optimal si , quel que soit l'estimateur \hat{R}_n de θ ,

$$\text{var}(\hat{T}_n) \leq \text{var}(\hat{R}_n)$$

Inégalité de Fréchet

\hat{T}_n = estimateur sans biais de f fonction dérivable de θ
 $\text{var}(\hat{T}_n)$ ne peut pas être plus petite que la borne de Fréchet

$$B_F = \frac{f'(\theta)^2}{I(\theta)}$$

Estimateur efficace

Efficacité de \hat{T}_n

$$\text{eff}(\hat{T}_n) = \frac{B_F}{\text{var}(\hat{T}_n)}$$

Si $\text{eff}(\hat{T}_n) = 1$, \hat{T}_n est efficace

Un estimateur efficace est optimal



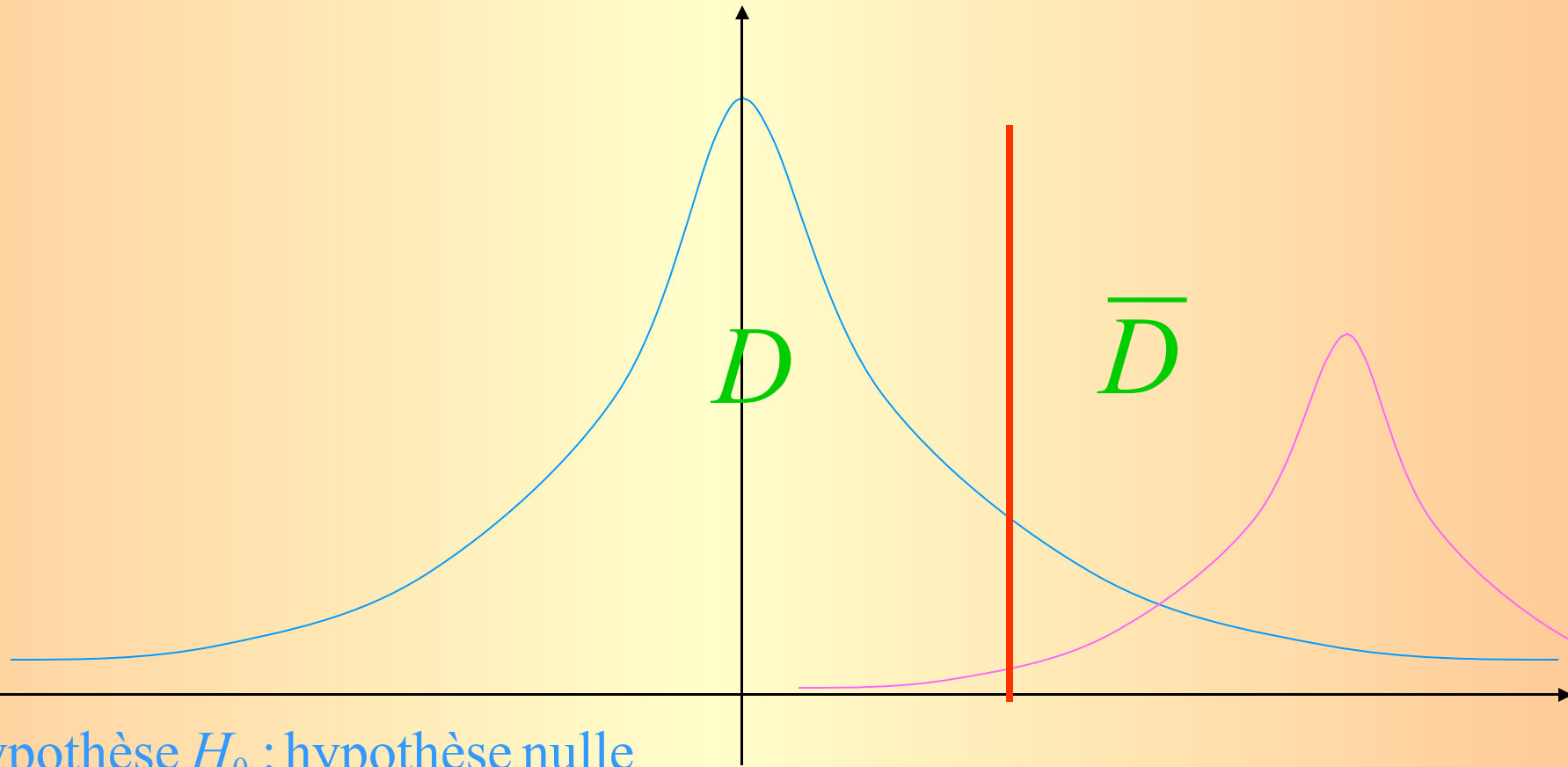
I – Rappels mathématiques

Tests statistiques

2 hypothèses H_0 et H_1 + échantillon $\tilde{X} = (X_1, \dots, X_n)$

Sous hypothèse H_0 : \tilde{X} de densité $p_{\tilde{X}}(\vec{x}, \theta_0)$

Sous hypothèse H_1 : \tilde{X} de densité $p_{\tilde{X}}(\vec{x}, \theta)$, $\theta > \theta_0$



Hypothèse H_0 : hypothèse nulle

Hypothèse H_1 : hypothèse alternative

D = région d'acceptation du test

\overline{D} = région critique du test

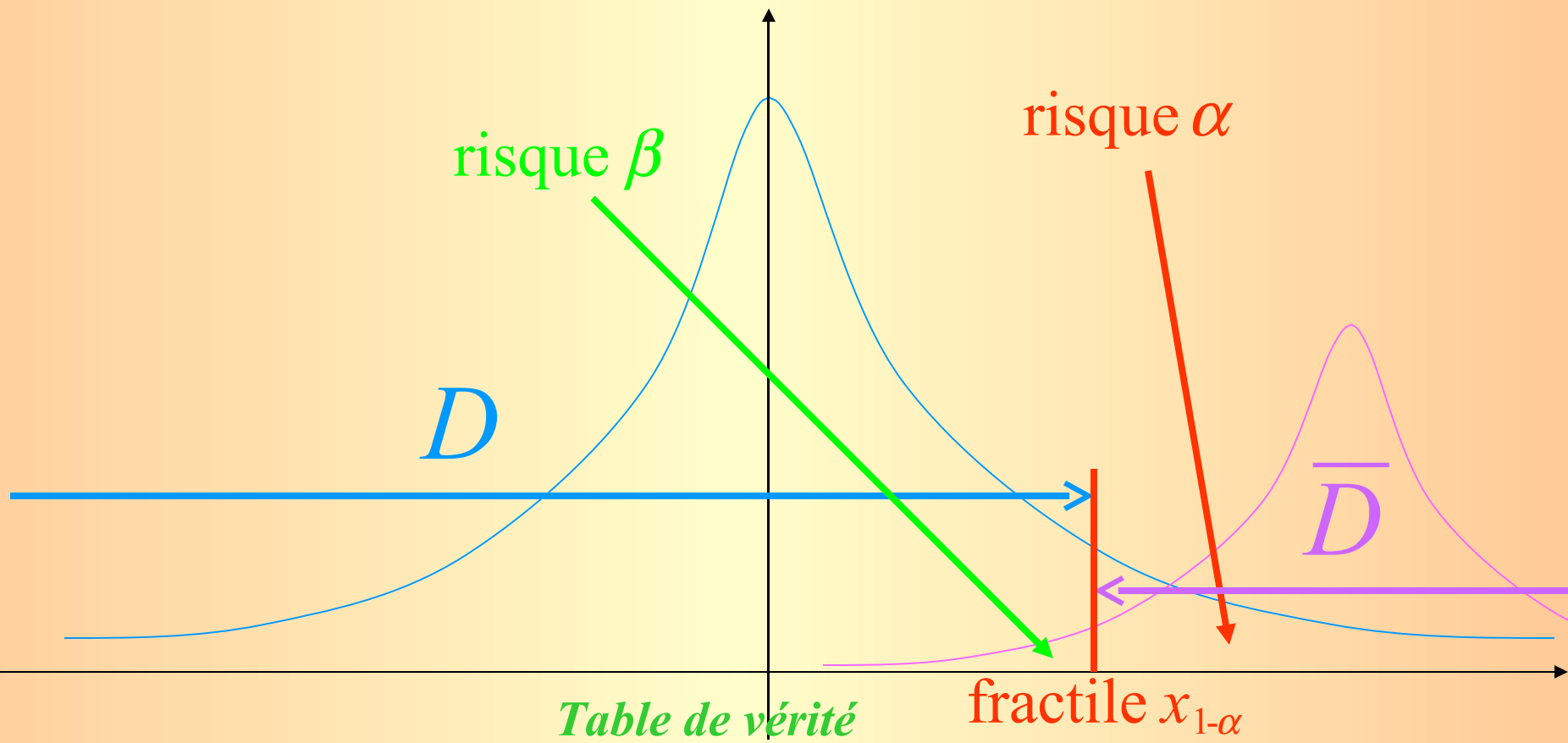


Table de vérité

fractile $x_{1-\alpha}$

$$\alpha = \int_{\bar{D}} p_{\vec{X}}(\vec{x}, \theta_0) d\vec{x}$$

$$\beta = \int_D p_{\vec{X}}(\vec{x}, \theta) d\vec{x}$$

Puissance du test : $\pi = 1 - \beta$

décision/vérité	H_0 vraie	H_1 vraie
accepter H_0	bonne décision	risque β
accepter H_1	risque α	bonne décision



II – Méthode des moindres carrés

A – Exemple introductif

n ($n \geq 2$) sangliers de la forêt gauloise (considérée comme un plan de \mathcal{R}^2)

\Rightarrow Faire passer Obélix près de ces n sangliers en allant "au plus court"

\Leftrightarrow Estimer une droite du plan $y=ax+b$ passant "au mieux" par les sangliers

$n = 2$: facile!

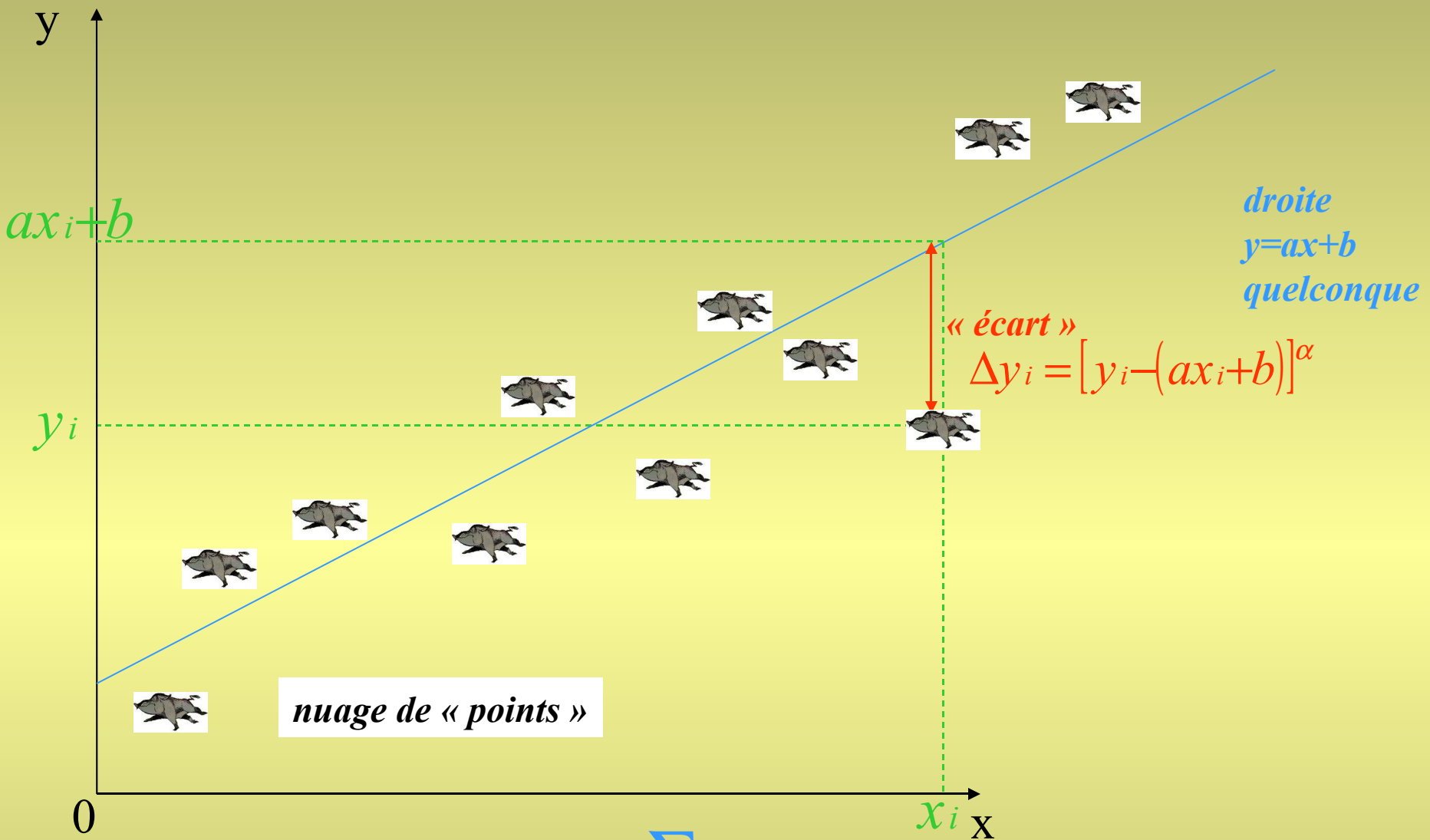
$$\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \end{bmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \Rightarrow \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} \frac{y_1 - y_2}{x_1 - x_2} \\ \frac{x_1 y_2 - x_2 y_1}{x_1 - x_2} \end{pmatrix}$$

$n > 2$?

$$\begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \Leftrightarrow AX = Y$$

\Rightarrow Matrice non inversible

\Rightarrow Redondance d'information



Solution possible = minimiser $\sum_i \Delta y_i$

Moindres carrés $\Rightarrow \alpha = 2 \quad \Leftrightarrow$ Minimiser $\phi(a,b) = \sum_i (ax_i + b - y_i)^2$

Dérivations partielles \Rightarrow matrice
forme quadratique fondamentale $Q_{\varphi,(a,b)}$

$$\begin{bmatrix} 2\sum_i x_i^2 & 2\sum_i x_i \\ . & 2n \end{bmatrix}$$

$\forall (a,b) \in \mathcal{R}^2$, $Q_{\varphi,(a,b)}$

non dégénérée

définie positive

(a,b) = point critique de φ

$$\begin{cases} \sum_i x_i(ax_i + b - y_i) = 0 \\ \sum_i (ax_i + b - y_i) = 0 \end{cases}$$

⇒ Résultats

$$\hat{a} = \frac{\sigma_{xy}}{\sigma_x^2} \quad \hat{b} = \bar{y} - \frac{\sigma_{xy}}{\sigma_x^2} \bar{x}$$

avec $\bar{x} = \frac{1}{n} \sum_i x_i$ $\bar{y} = \frac{1}{n} \sum_i y_i$

$$\sigma_x^2 = \frac{1}{n} \sum_i (x_i - \bar{x})^2$$

$$\sigma_{xy} = \frac{1}{n} \sum_i (x_i - \bar{x}) \cdot (y_i - \bar{y})$$

Exemple linéaire : $\varphi(a,b) = (AX - Y) \cdot (AX - Y)^T$

Si fonction modèle non linéaire ? ...

A large, white iceberg with a jagged, mountain-like peak floats in the middle of a dark blue ocean. The sky is a clear, bright blue with some wispy white clouds. The horizon is visible in the distance. The overall scene is serene and cold.

II – Méthode des moindres carrés
B – Formulation générale

$E = \mathfrak{R}$ espace vectoriel normé de dimension p

$F = \mathfrak{R}$ espace euclidien de dimension n ($n > p$) de produit scalaire \langle , \rangle

$S =$ matrice de \langle , \rangle relativement à une base de F

$\| \cdot \| =$ norme induite par \langle , \rangle

Problème :

Modéliser vecteur de mesures $y \in F$
avec vecteur de paramètres $x \in E$

par l'intermédiaire de

$$f : \Omega \subset E \rightarrow F$$

Utilisation de la fonction

$$\begin{cases} \varphi : \Omega \subset E \rightarrow \mathfrak{R} \\ x \mapsto \varphi(x) = \|f(x) - y\|^2 \end{cases}$$

Critère moindres carrés \Leftrightarrow minimiser φ

Solution approchée x_0 de $\varphi'(x) = 0$ f linéarisable au voisinage de x_0

Matrice jacobienne A de f en x_0 de plein rang

Développement de Taylor de f au voisinage de x_0

$$f(x) = f(x_0) + f'(x_0) \cdot (x - x_0) + \|x - x_0\| \mathcal{E}(x - x_0)$$

$$\varphi(x) \cong \|f(x_0) + f'(x_0) \cdot (x - x_0) - y\|^2$$

Formulation matricielle

$$\varphi(x) \cong (AX - Y)^T \cdot S \cdot (AX - Y) = \phi(X)$$

$$\text{où } X = x - x_0 \text{ et } Y = y - f(x_0)$$

Minimiser φ au voisinage de x_0



Ordre 1 en $x - x_0$

Minimiser $\phi(X) = (AX - Y)^T \cdot S \cdot (AX - Y)$

Linéarisation de $f \Rightarrow$ modèle linéaire

ϕ fonction numérique de classe C^∞ sur \mathfrak{R}^n

$$\phi(X) = (AX - Y)^T \cdot S \cdot (AX - Y)$$

$$\phi(X) = X^T A^T S A X - X^T A^T S Y - Y^T S A X + Y^T S Y$$

Application linéaire tangente en $X : \phi'(X)$

En différentiant, $\forall dX \in \mathfrak{R}^2$,

$$\phi'(X) \cdot dX = dX^T A^T S A X + X^T A^T S A dX - dX^T A^T S Y - Y^T S A dX$$

$$\phi'(X) \cdot dX = 2(X^T A^T S A - Y^T S A) \cdot dX$$

$$\phi'(X) = 2(X^T A^T S A - Y^T S A)$$

\hat{X} point critique de ϕ ssi $A^T S A \hat{X} = A^T S Y$

Si $A^T S A$ non singulière

$$\hat{X} = (A^T S A)^{-1} A^T S Y$$

En différentiant de nouveau $\phi'(X)$:
forme quadratique fondamentale associée à ϕ en X

Forme quadratique non dégénérée

$$\phi''(X) = 2A^T S A$$

Forme quadratique positive

$$\forall U \in \mathbb{R}^2, U^T \cdot \phi''(X) \cdot U = 2U^T A^T \cdot S \cdot A U = 2(AU)^T \cdot S \cdot (AU) \geq 0$$

Forme quadratique définie

$$U^T \cdot \phi''(X) \cdot U = 0 \Rightarrow (AU)^T \cdot S \cdot (AU) = 0 \Rightarrow AU = 0 \Rightarrow U = 0$$

ϕ admet un minimum local strict en \hat{x} :

$$\hat{x} = x_0 + (A^T S A)^{-1} A^T S (y - f(x_0))$$

Processus de résolution itératif

Solution $\hat{x} = x_0 + (A^T S A)^{-1} A^T S (y - f(x_0))$
invariante par homothétie sur S

$S =$ *matrice de poids*

$X = x - x_0 =$ *vecteur des paramètres*

$Y = y - f(x_0) =$ *vecteur des (pseudo-) observations*

$A =$ *matrice modèle (des dérivées partielles)*

$A^T S A =$ *matrice normale*

$y = f(x) \cong f(x_0) + f'(x_0) \cdot (x - x_0) \Leftrightarrow Y \cong AX$

$\times A^T S =$ *normalisation*

\Rightarrow *équation normale* $A^T S A X = A^T S Y$

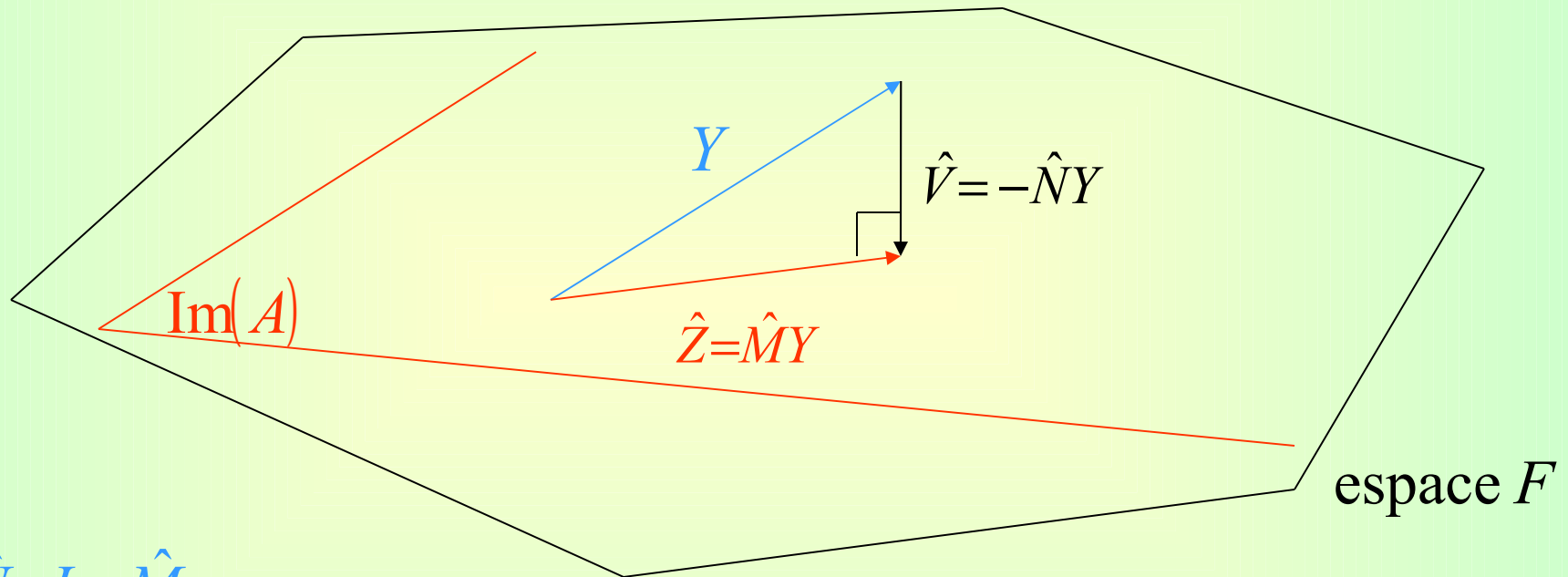
$\hat{V} = A \hat{X} - Y =$ *vecteur des résidus*

Interprétation géométrique

$$\hat{Z} = A\hat{X} = \left[A^T(A^TSA)^{-1}A^T SY \right] = \hat{M}Y$$

\hat{M} = *matrice chapeau* (« hat matrix »)

Matrice d'un projecteur sur $\text{Im}(A)$



$$\hat{N} = I_n - \hat{M}$$

$$\hat{V} = -\hat{N}Y$$

\hat{Z} = projection orthogonale de Y sur $\text{Im}(A)$ $V^T \hat{Z} = 0$



II – Méthode des moindres carrés

C – Modèle linéaire

Systeme linéaire $Y=A\dot{X}+\varepsilon$

Y et ε = variables aléatoires vectorielles

\dot{X} = vecteur des paramètres constants \leftarrow Valeurs vraies

A = matrice déterministe

$$E(Y)=\dot{Y}=A\dot{X}, E(\varepsilon)=0$$

$$E(\varepsilon\varepsilon^T)=\text{var}(Y)=\sigma^2\Lambda, \sigma^2 \text{ inconnu } \sigma^2 = \textit{facteur unitaire de variance}$$

Estimateur des moindres carrés \hat{X}

$$\hat{X}=(A^T S A)^{-1} A^T S Y$$

Estimateur linéaire en Y sans biais

Théorème de Gauss-Markov

Si $S=\frac{1}{\sigma^2}\Lambda^{-1}$, estimateur des moindres carrés optimal dans la classe des estimateurs sans biais et linéaires en Y . Et $\text{var}(\hat{X})=\sigma^2(A^T\Lambda^{-1}A)^{-1}$

Exemple : Variables aléatoires ε_i deux à deux indépendantes

$$\Lambda^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\sigma_n^2} \end{bmatrix} \quad \text{et} \quad \Lambda^{-1}Y = \begin{pmatrix} \frac{y_1}{\sigma_1^2} \\ \vdots \\ \frac{y_n}{\sigma_n^2} \end{pmatrix} \quad \text{Si } A = \begin{bmatrix} a_{11} & \dots & a_{1p} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{np} \end{bmatrix}$$

$$A^T \Lambda^{-1} A = \begin{bmatrix} \sum_{k=1}^n \frac{a_{k1}^2}{\sigma_k^2} & \dots & \sum_{k=1}^n \frac{a_{k1} a_{kj}}{\sigma_k^2} & \dots & \sum_{k=1}^n \frac{a_{k1} a_{kp}}{\sigma_k^2} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \sum_{k=1}^n \frac{a_{ki} a_{k1}}{\sigma_k^2} & \dots & \sum_{k=1}^n \frac{a_{ki} a_{kj}}{\sigma_k^2} & \dots & \sum_{k=1}^n \frac{a_{ki} a_{kp}}{\sigma_k^2} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \sum_{k=1}^n \frac{a_{kp} a_{k1}}{\sigma_k^2} & \dots & \sum_{k=1}^n \frac{a_{kp} a_{kj}}{\sigma_k^2} & \dots & \sum_{k=1}^n \frac{a_{kp}^2}{\sigma_k^2} \end{bmatrix} \quad \text{et} \quad A^T \Lambda^{-1} Y = \begin{pmatrix} \sum_{k=1}^n \frac{a_{k1} y_k}{\sigma_k^2} \\ \vdots \\ \sum_{k=1}^n \frac{a_{ki} y_k}{\sigma_k^2} \\ \vdots \\ \sum_{k=1}^n \frac{a_{kp} y_k}{\sigma_k^2} \end{pmatrix}$$

Résultats Système linéaire $Y=A\dot{X}+\varepsilon \Rightarrow$ estimateur \hat{X}

<i>grandeur</i>	<i>expression mathématique</i>	<i>espérance</i>	<i>variance</i>
<i>erreurs aléatoires (n)</i> ε	<i>inconnue mais on suppose en général erreurs gaussiennes</i>	0	$\sigma^2 \Lambda$
<i>observations (n)</i> Y	/	\dot{Y}	$\sigma^2 \Lambda$
<i>paramètres (p<n)</i> \hat{X}	$(A^T \Lambda^{-1} A)^{-1} A^T \Lambda^{-1} Y$	\dot{X}	$\sigma^2 (A^T \Lambda^{-1} A)^{-1}$
<i>résidus (n)</i> \hat{V}	$A\hat{X}-Y$	0	$\sigma^2 (\Lambda - A(A^T \Lambda^{-1} A)^{-1} A^T)$
<i>facteur unitaire de variance</i> $\hat{\sigma}^2$	$\frac{\hat{V}^T \Lambda^{-1} \hat{V}}{n-p}$	σ^2	/

$\text{cov}(\hat{X}, \hat{V}) = 0$

Observations gaussiennes

$$Y \rightarrow N(A\dot{X}, \sigma^2 \Lambda)$$

$$\hat{V} \rightarrow N\left(0, \sigma^2 \left[\Lambda - A(A^T \Lambda^{-1} A)^{-1} A^T \right]\right)$$

$$\hat{X} \rightarrow N\left(\dot{X}, \sigma^2 \left(A^T \Lambda^{-1} A\right)^{-1}\right)$$

Tests sur les résidus

Loi normale

$$\frac{\Gamma_i(\hat{V})}{\sqrt{\hat{\sigma}^2} \sqrt{\Gamma_{ii} \left(\Lambda - A(A^T \Lambda^{-1} A)^{-1} A^T \right)}} : \text{Student } (n-p)$$

$$\begin{cases} H_0 : \Gamma_i(\hat{V}) = 0 \\ H_1 : \Gamma_i(\hat{V}) \neq 0 \end{cases}$$

Risque de 1^{ière} espèce α , $t_{n-p}\left(1-\frac{\alpha}{2}\right)$ = fractile $1-\frac{\alpha}{2}$ Densité de probabilité symétrique

H_0 acceptée au seuil α si

$$\frac{|\Gamma_i(\hat{V})|}{\sqrt{\hat{\sigma}^2} \sqrt{\Gamma_{ii} \left(\Lambda - A(A^T \Lambda^{-1} A)^{-1} A^T \right)}} \leq t_{n-p}\left(1-\frac{\alpha}{2}\right)$$

Tests sur les paramètres compensés

$$\frac{\Gamma_i(\hat{X}) - \Gamma_i(X_0)}{\sqrt{\hat{\sigma}^2 \Gamma_{ii} \left((A^T \Lambda^{-1} A)^{-1} \right)}} : \text{Student } (n-p)$$

$$\begin{cases} H_0 : \Gamma_i(\hat{X}) = \Gamma_i(X_0) \\ H_1 : \Gamma_i(\hat{X}) \neq \Gamma_i(X_0) \end{cases}$$

Risque de 1^{ière} espèce α , $t_{n-p} \left(1 - \frac{\alpha}{2} \right) = \text{fractile } 1 - \frac{\alpha}{2}$

H_0 acceptée au seuil α si

$$\frac{|\Gamma_i(\hat{X}) - \Gamma_i(X_0)|}{\sqrt{\hat{\sigma}^2 \Gamma_{ii} \left((A^T \Lambda^{-1} A)^{-1} \right)}} \leq t_{n-p} \left(1 - \frac{\alpha}{2} \right)$$



II – Méthode des moindres carrés

D – Cas d'école

On mesure la largeur a , la longueur b et la diagonale d d'un rectangle. De précédentes mesures avaient donné les valeurs $a=1,60$ m et $b=2,08$ m. Le rectangle étant en bois, on sait qu'il peut se déformer sous l'effet des contraintes thermiques. Les mesures effectuées doivent permettre de vérifier si les dimensions du rectangle ont varié de façon significative depuis les dernières mesures. Les mesures sont fournies dans le tableau suivant.

a	b	$\sqrt{a^2+b^2}$
1,634	2,078	2,609
1,643	2,065	2,619
1,659	2,044	2,631
1,656	2,036	2,642
1,621	2,018	2,606

Linéarisation du modèle

$$\begin{cases} a = a_0 + \delta a \\ b = b_0 + \delta b \\ d \equiv d_0 + \frac{a_0}{d_0} \delta a + \frac{b_0}{d_0} \delta b \end{cases}$$

Équations d'observations

$$\begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \\ \frac{a_0}{d_0} & \frac{b_0}{d_0} \\ \vdots & \vdots \\ \frac{a_0}{d_0} & \frac{b_0}{d_0} \end{bmatrix} \cdot \begin{bmatrix} \delta a \\ \delta b \end{bmatrix} = \begin{bmatrix} a_1 - a_0 \\ \vdots \\ a_5 - a_0 \\ b_1 - b_0 \\ \vdots \\ b_5 - b_0 \\ d_1 - d_0 \\ \vdots \\ d_5 - d_0 \end{bmatrix}$$

Normalisation

$$\begin{bmatrix} 6,860 & 2,419 \\ . & 8,144 \end{bmatrix} \cdot \begin{pmatrix} \delta \hat{a} \\ \delta \hat{b} \end{pmatrix} = \begin{pmatrix} 0,163 \\ -0,169 \end{pmatrix}$$

Inversion

$$\begin{cases} \delta \hat{a} = 0,035 \\ \delta \hat{b} = -0,031 \end{cases} \text{ soit } \begin{cases} \hat{a} = 1,635 \\ \hat{b} = 2,050 \end{cases}$$

Résidus

a	b	$\sqrt{a^2+b^2}$
0,001	-0,029	0,011
-0,008	-0,016	0,001
-0,024	0,005	-0,010
-0,021	0,013	-0,021
0,056	0,031	0,014

Facteur unitaire de variance

$$\hat{\sigma}^2 = \frac{\hat{V}^T \cdot \hat{V}}{n-p} \cong \frac{7,38 \cdot 10^{-3}}{13} \cong 5,68 \cdot 10^{-4}$$

Statistique des estimations

Matrice normale inverse

$$\begin{bmatrix} 0,163 & -0,048 \\ \cdot & 0,137 \end{bmatrix}$$

$$\begin{cases} \text{Précision sur } \hat{a} : \sigma_{\hat{a}} = \sqrt{0,163 \times 5,68 \cdot 10^{-4}} \\ \text{Précision sur } \hat{b} : \sigma_{\hat{b}} = \sqrt{0,137 \times 5,68 \cdot 10^{-4}} \end{cases} \text{ soit } \begin{cases} \hat{a} = 1,635 \pm 0,0096 \\ \hat{b} = 2,050 \pm 0,0088 \end{cases}$$

$$\text{Corrélation entre } \hat{a} \text{ et } \hat{b} : \text{corr}(\hat{a}, \hat{b}) = \frac{-0,048}{\sqrt{0,163 \times 0,137}} \cong -0,32$$

Tests statistiques

$$\alpha = 4\%, t_{13}(0,98) = 2,650$$

$$\text{Valeurs tests pour a et b : } \frac{0,035}{0,0096} \cong 3,65 \text{ et } \frac{0,031}{0,0088} \cong 3,52$$

Variations significatives au regard du test effectué

⇒ Itération éventuelle

II – Méthode des moindres carrés

E – Techniques complémentaires



Cumul d'équations normales

$$AX = \begin{bmatrix} A_1 \\ \vdots \\ A_m \end{bmatrix} X \cong Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_m \end{pmatrix}$$

$$\text{var}(Y) = \begin{bmatrix} \Lambda_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \Lambda_m \end{bmatrix}$$



Systeme normal global

$$\left(\sum_{i=1}^m A_i^T \Lambda_i^{-1} A_i \right) X = \sum_{i=1}^m A_i^T \Lambda_i^{-1} Y_i$$

⇒ "Cumul" des sous-systèmes normaux

Exemple : Calculs de géodésie spatiale ; positionnement GPS

Équations de contraintes

Systeme linéaire $AX \cong Y$ avec $\text{var}(Y) = \Lambda$

Information *a priori* sur le vecteur des paramètres X

$$A_0 X = X_0 \text{ où } \text{var}(X_0) = \Lambda_0$$

$$\begin{bmatrix} A_0 \\ A \end{bmatrix} X \cong \begin{pmatrix} Y \\ X_0 \end{pmatrix} \text{ de matrice de poids } \begin{bmatrix} \Lambda^{-1} & 0 \\ 0 & \Lambda_0^{-1} \end{bmatrix}$$

Cumul des équations normales \Rightarrow système normalisé

$$(A^T \Lambda^{-1} A + A_0^T \Lambda_0^{-1} A_0) X = (A^T \Lambda^{-1} Y + A_0^T \Lambda_0^{-1} X_0)$$

" $A_0^T \Lambda_0^{-1} A_0 X = A_0^T \Lambda_0^{-1} X_0$ " = équation de contrainte

Exemple : Positionnement

$$\begin{cases} x = x_0 \pm \sigma_x \\ y = y_0 \pm \sigma_y \\ z = z_0 \pm \sigma_z \end{cases}$$

Réduction de paramètres

Système d'observations :

$$\begin{bmatrix} A_1 & A_2 \end{bmatrix} \cdot \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \equiv Y \text{ avec } \text{var}(Y) = \Lambda$$

⇒ Résoudre ce système selon X_2

Système normal associé

$$\begin{bmatrix} A_1^T \Lambda^{-1} A_1 & A_1^T \Lambda^{-1} A_2 \\ A_2^T \Lambda^{-1} A_1 & A_2^T \Lambda^{-1} A_2 \end{bmatrix} \cdot \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} A_1^T \Lambda^{-1} Y \\ A_2^T \Lambda^{-1} Y \end{pmatrix}$$

Première équation donne

$$X_1 = (A_1^T \Lambda^{-1} A_1)^{-1} A_1^T \Lambda^{-1} Y - (A_1^T \Lambda^{-1} A_1)^{-1} A_1^T \Lambda^{-1} A_2 X_2$$

Seconde équation donne

$$A_2^T \Lambda^{-1} (I - A_1 (A_1^T \Lambda^{-1} A_1)^{-1} A_1^T \Lambda^{-1}) A_2 X_2 = A_2^T \Lambda^{-1} (I - A_1 (A_1^T \Lambda^{-1} A_1)^{-1} A_1^T \Lambda^{-1}) Y$$

Exemple : Calculs de géodésie spatiale

Analyse des composantes de la variance

$$AX = \begin{bmatrix} A_1 \\ \vdots \\ A_m \end{bmatrix} X \cong Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_m \end{pmatrix} \quad \text{var}(Y) = \begin{bmatrix} \sigma_1^2 \Lambda_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_m^2 \Lambda_m \end{bmatrix}$$

⇒ Trouver des estimations des σ_i^2

⇒ *Cumul des équations normales*

Matrice normale: $N = \sum_{i=1}^m A_i^T \Lambda_i^{-1} A_i$ Second membre $K = \sum_{i=1}^m A_i^T \Lambda_i^{-1} Y_i$
On pose $\Omega = N^{-1}$

$\dim(Y_i) = n_i$

Vecteur des résidus du groupe i : \hat{V}_i

Matrice normale du groupe i : $N_i = A_i^T \Lambda_i^{-1} A_i$

Exemple : *Calculs de géodésie spatiale avec différents satellites*

Deux méthodes très proches

Degré de liberté

Hypothèse : $\sigma_i^2 \approx \sigma^2$

$$\hat{\sigma}_i^2 = \frac{\hat{V}_i^T \Lambda_i^{-1} \hat{V}_i}{n_i - \text{tr}(\Omega N_i)}$$

Similaire à $\hat{\sigma}_0^2 = \frac{\hat{V}^T \Lambda^{-1} \hat{V}}{n-p}$

Tous les $\hat{\sigma}_i^2$ sont positifs

Helmert

Pas d'hypothèse supplémentaire

Résolution de $H. \begin{pmatrix} \hat{\sigma}_1^2 \\ \vdots \\ \hat{\sigma}_m^2 \end{pmatrix} = \begin{pmatrix} \hat{c}_1 \\ \vdots \\ \hat{c}_m \end{pmatrix}$

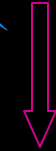


matrice symétrique

$$H_{i,j} = \delta_i^j (n_i - 2 \text{tr}(\Omega N_i)) + \text{tr}(\Omega N_i \Omega N_j)$$



$\hat{\sigma}_i^2$



$$\hat{c}_i = \hat{V}_i^T \Lambda_i^{-1} \hat{V}_i$$

Les $\hat{\sigma}_i^2$ ne sont pas forcément positifs

The background of the slide is a photograph of the International Space Station (ISS) in orbit above Earth. The station's complex structure, including multiple modules and large solar panel arrays, is clearly visible against the dark blue of the planet's atmosphere. The Earth's surface below shows a mix of green landmasses and white cloud cover.

II – Méthode des moindres carrés

F – Limites de la méthode

Linéarisation et minima locaux

⇒ Hypothèses rigoureuses

Manque de robustesse

Observations biaisées : erreur systématique

Faute de mesure : erreur aléatoire

⇒ Tests sur les résidus

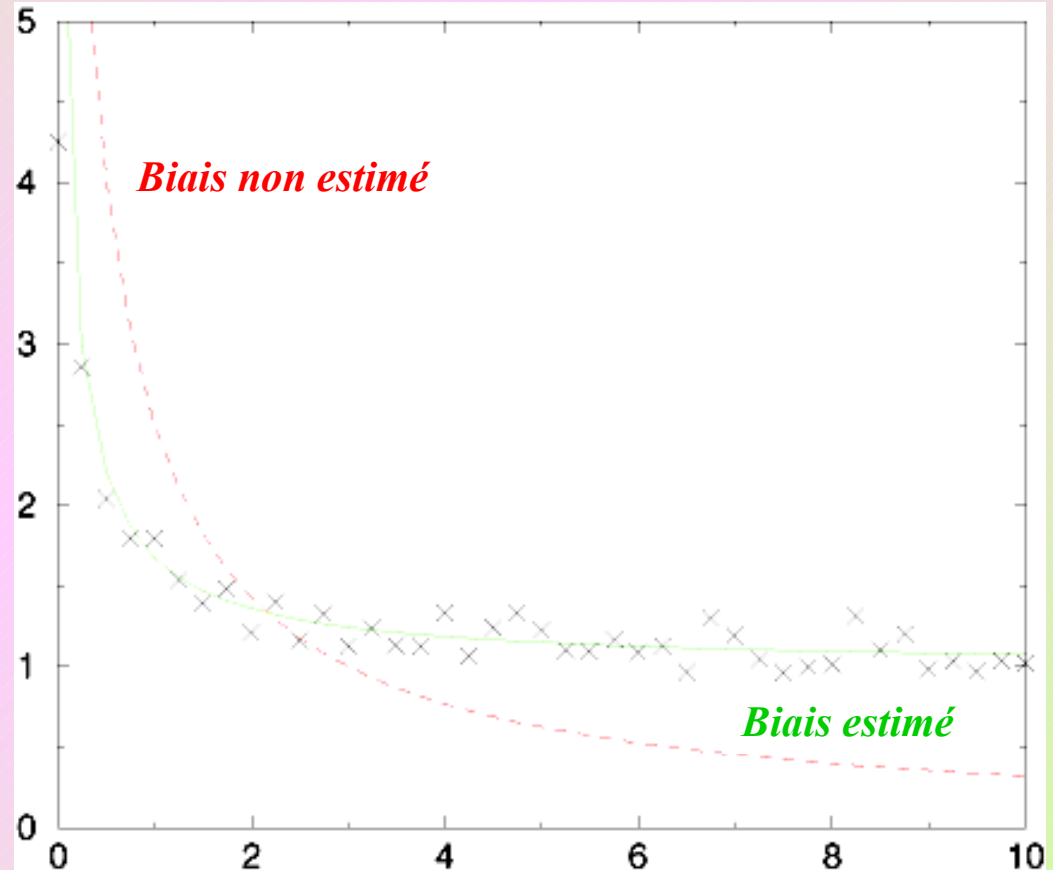
Modèle faux

⇒ Modèle physiquement correct

Manque de robustesse

Observations biaisées : erreur systématique

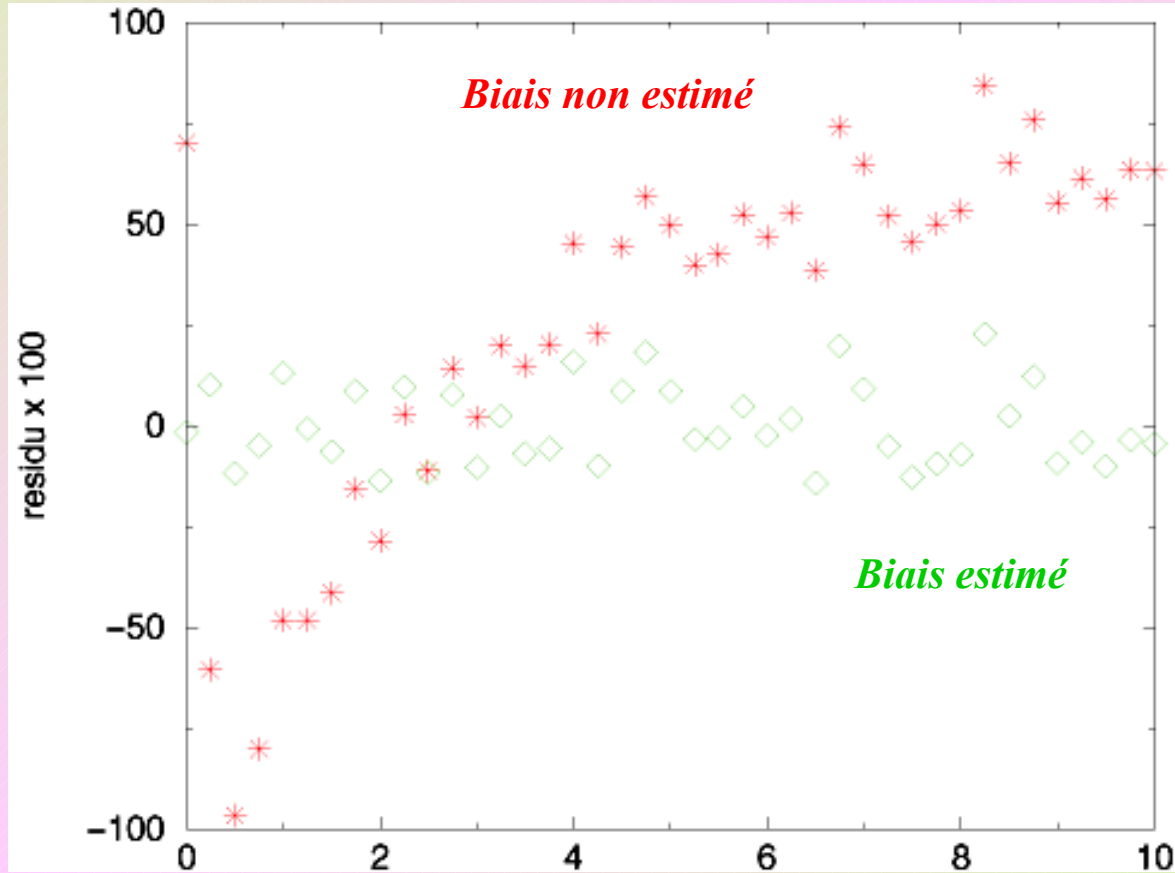
Fonction $x \mapsto \frac{1}{ax+b}$



Manque de robustesse

Observations biaisées : erreur systématique

Résidus de l'estimation

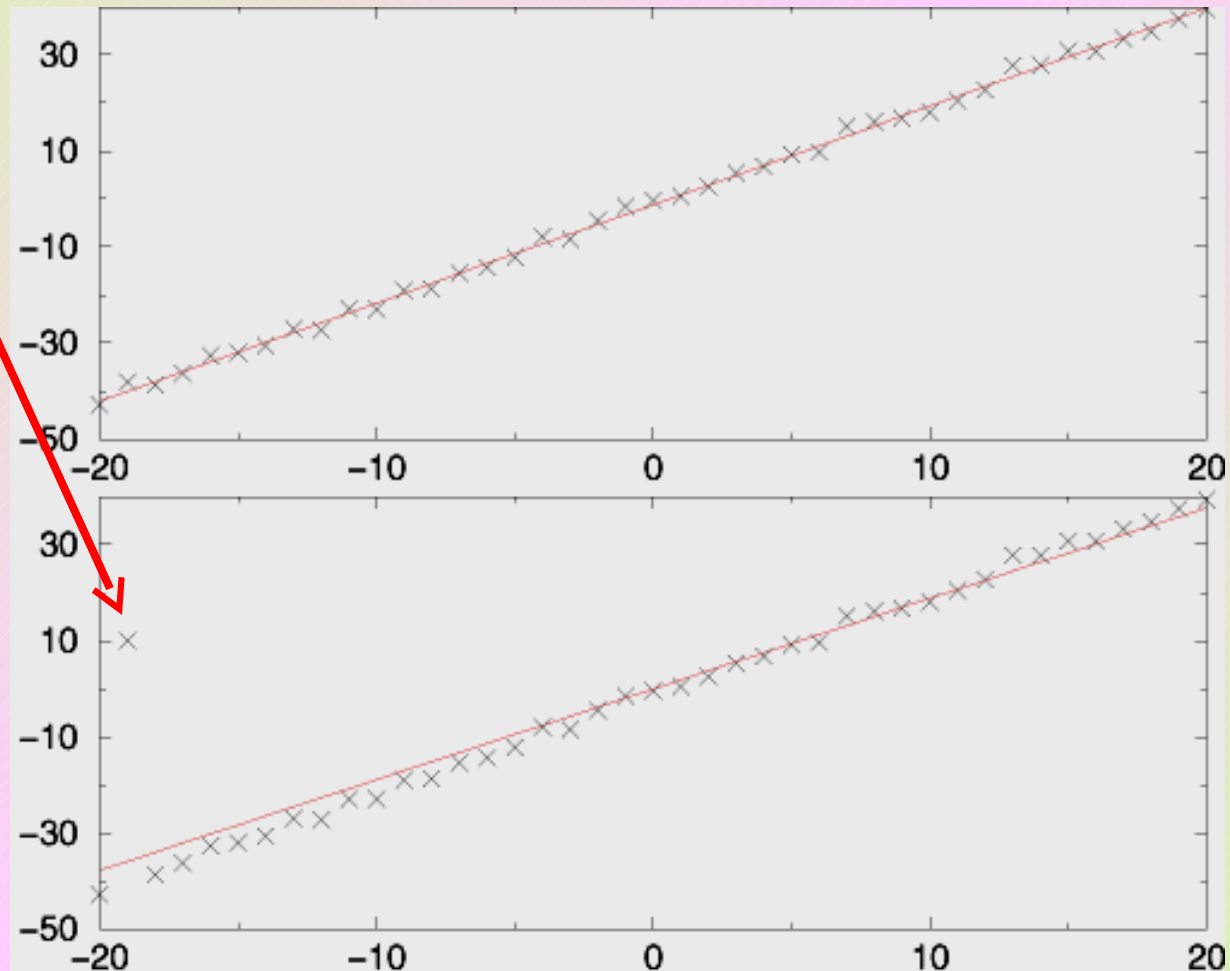


Manque de robustesse

Faute de mesure : erreur aléatoire

Droite de régression linéaire

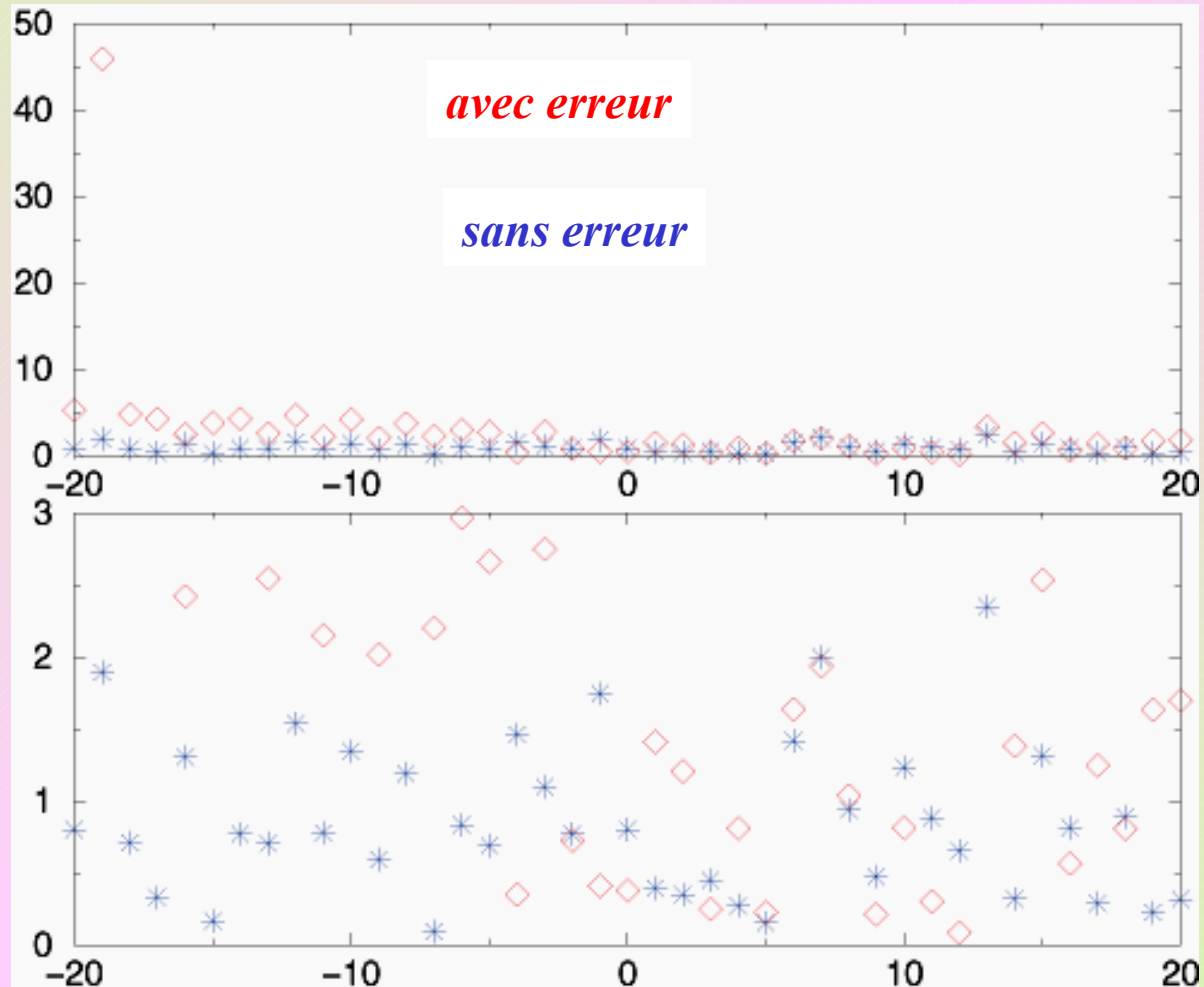
Erreur



Manque de robustesse

Faute de mesure : erreur aléatoire

Résidus de l'estimation





**III – Autres méthodes
d'optimisation**

A – Simplexe modifié

Résoudre le système $AX \cong Y$

Critère moindres carrés : minimiser $\sum_i \left(\sum_j a_{ij} x_j - b_i \right)^2$

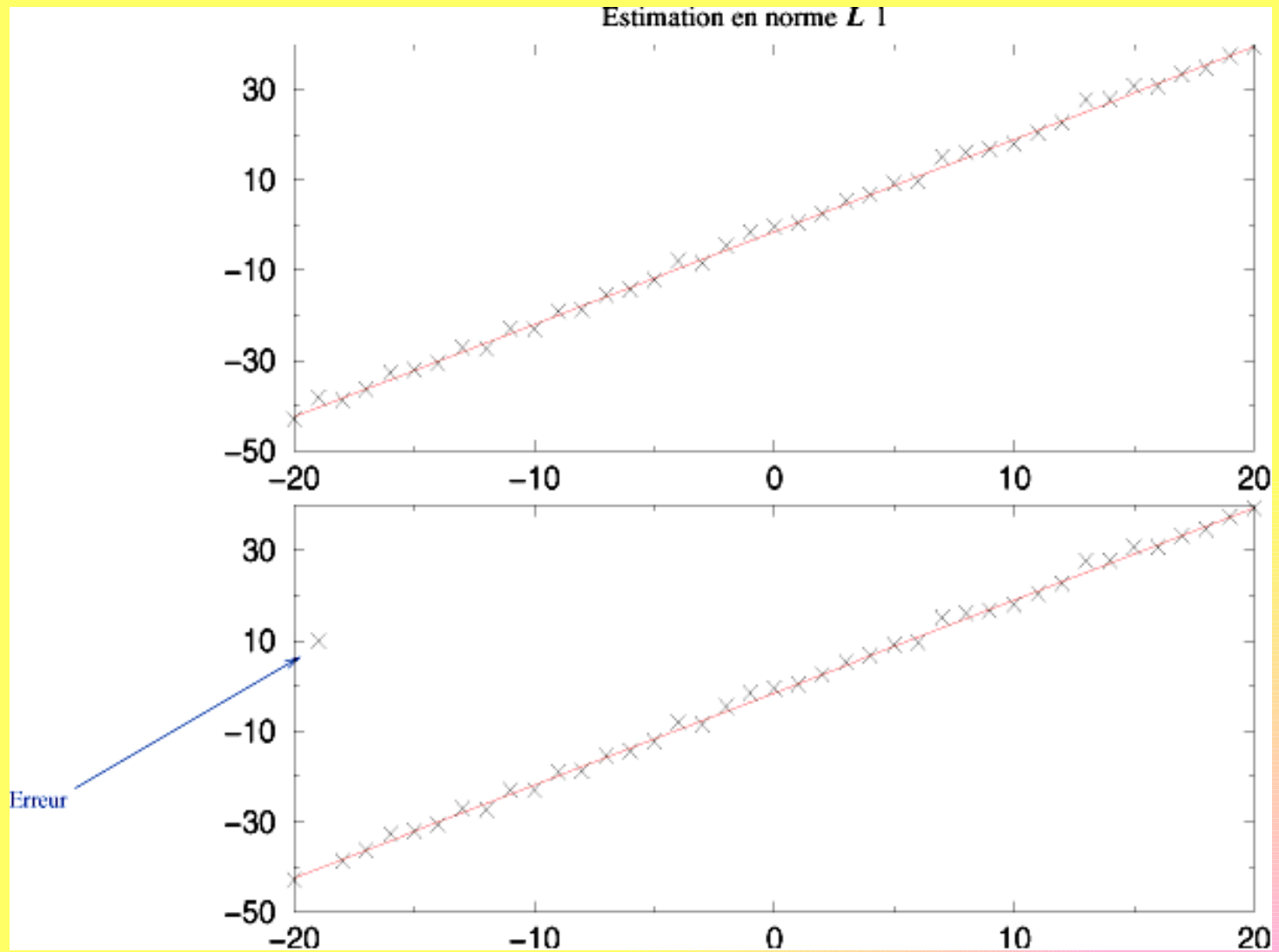
Critère norme L_1 : minimiser $\sum_i \sum_j |a_{ij} x_j - b_i|$

\Rightarrow Algorithme du simplexe modifié

Minimiser $\|AX - B\|_1$ sous les contraintes $CX = D$ et $EX \leq F$

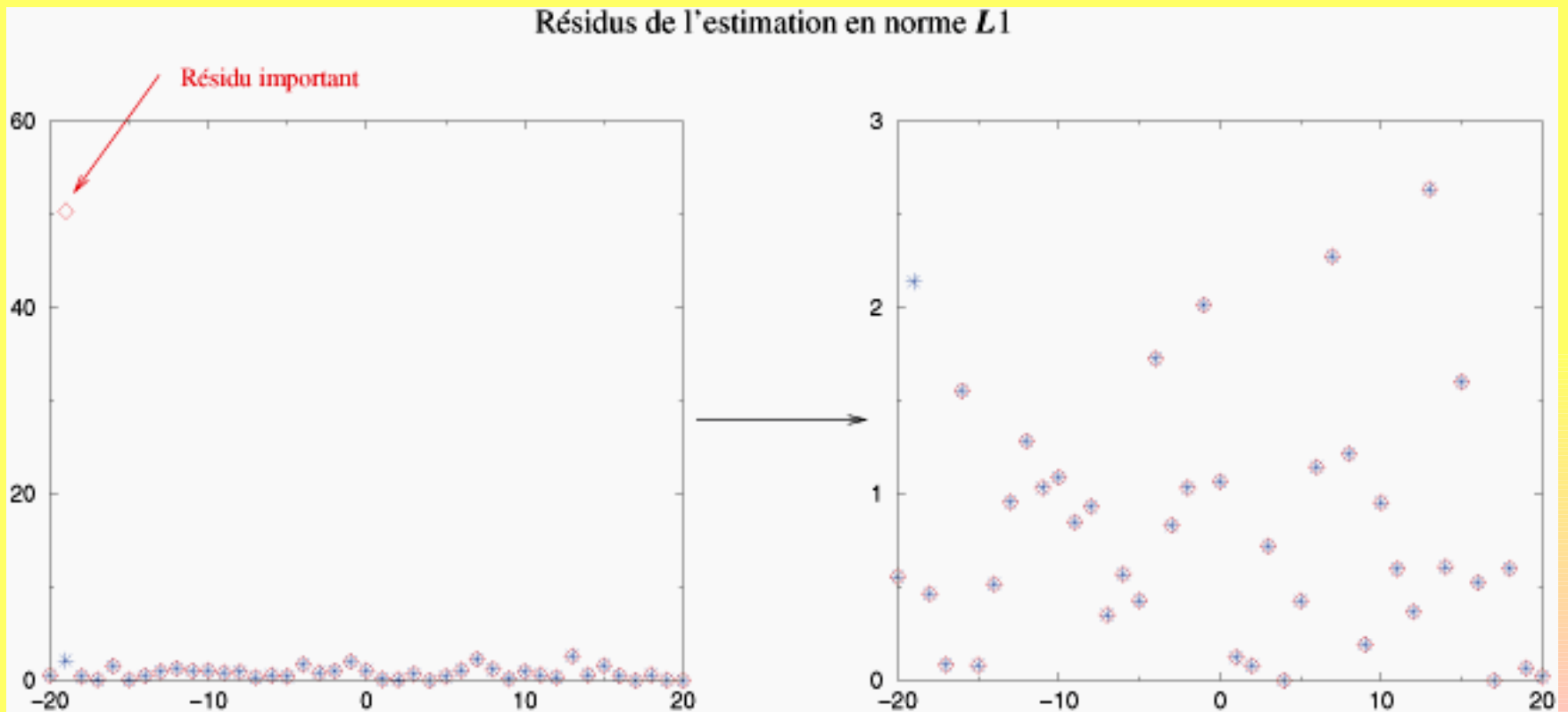
Parcours de tous les sommets du polyèdre convexe des contraintes

Droite de régression linéaire



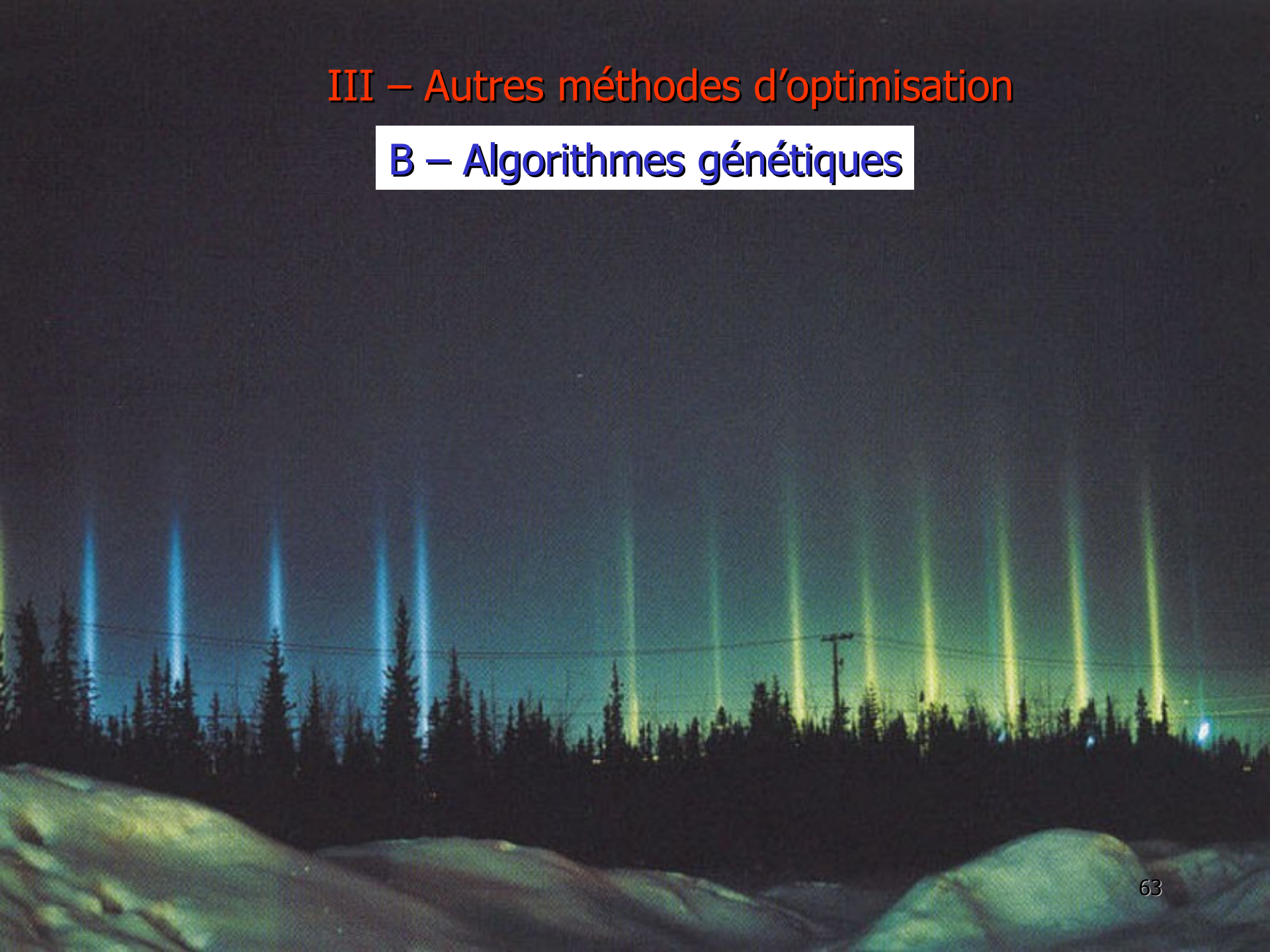
Droite de régression linéaire

Résidus de l'estimation



III – Autres méthodes d'optimisation

B – Algorithmes génétiques



Déroulement de l'algorithme

Population initiale

*Initialisation
Population initiale
= population courante*

Étape 1 : Évaluation
Population courante

Conditions vérifiées

Solutions

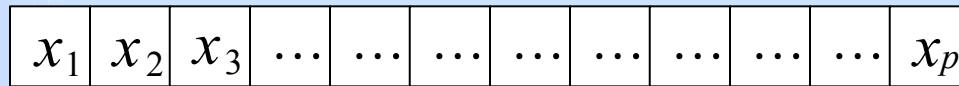
*Aptitude
à la reproduction*

Étape 2 : Sélection et duplication
Population intermédiaire

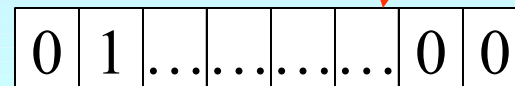
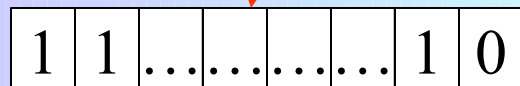
*Saut d'une
génération*

Étape 3 : Croisement et mutation
Population suivante

Codage des paramètres

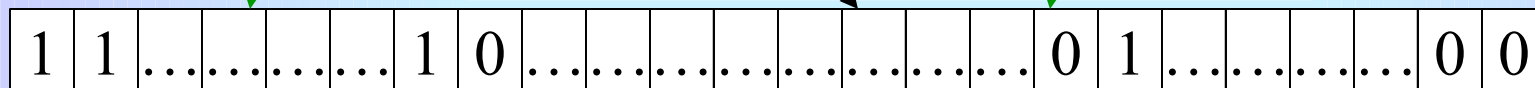


Vecteur x des paramètres
(p composantes)



Codage de chaque composante
sur l bits

gène (bit)



Vecteur chromosome (pl composantes) (individu)

Population = plusieurs chromosomes

Évaluation

Fonction d'évaluation : performance des individus d'une population

« Moindres carrés » : $e(x) = \sum_i (f_i(x) - y_i)^2$

« Norme L1 » : $e(x) = \sum_i |f_i(x) - y_i|$

« Médiane » : $e(x) = C - \text{médiane} \left((f_i(x) - y_i)^2 \right)$

Sélection et duplication

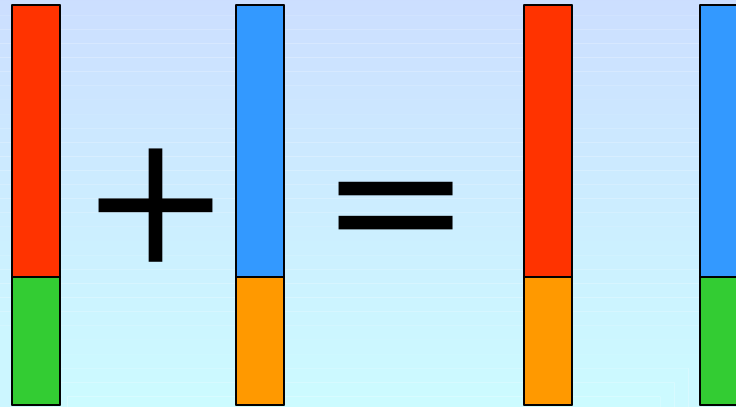
Fonction d'évaluation e \longrightarrow Fonction d'aptitude : $a(x) = \frac{e(x)}{\bar{e}}$

Population courante \longrightarrow Population intermédiaire

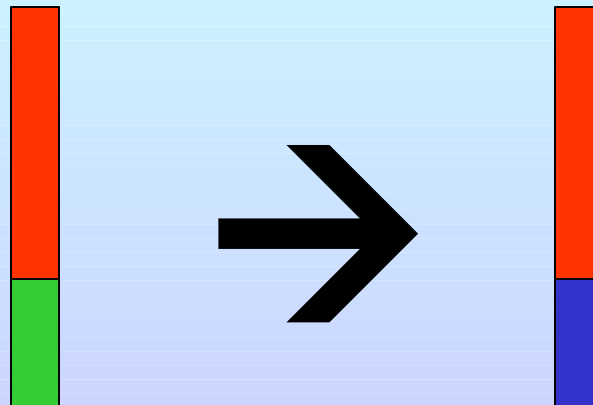
Sélection orientée vers l'exploitation

Croisement

Population intermédiaire : 2 individus , probabilité p_c

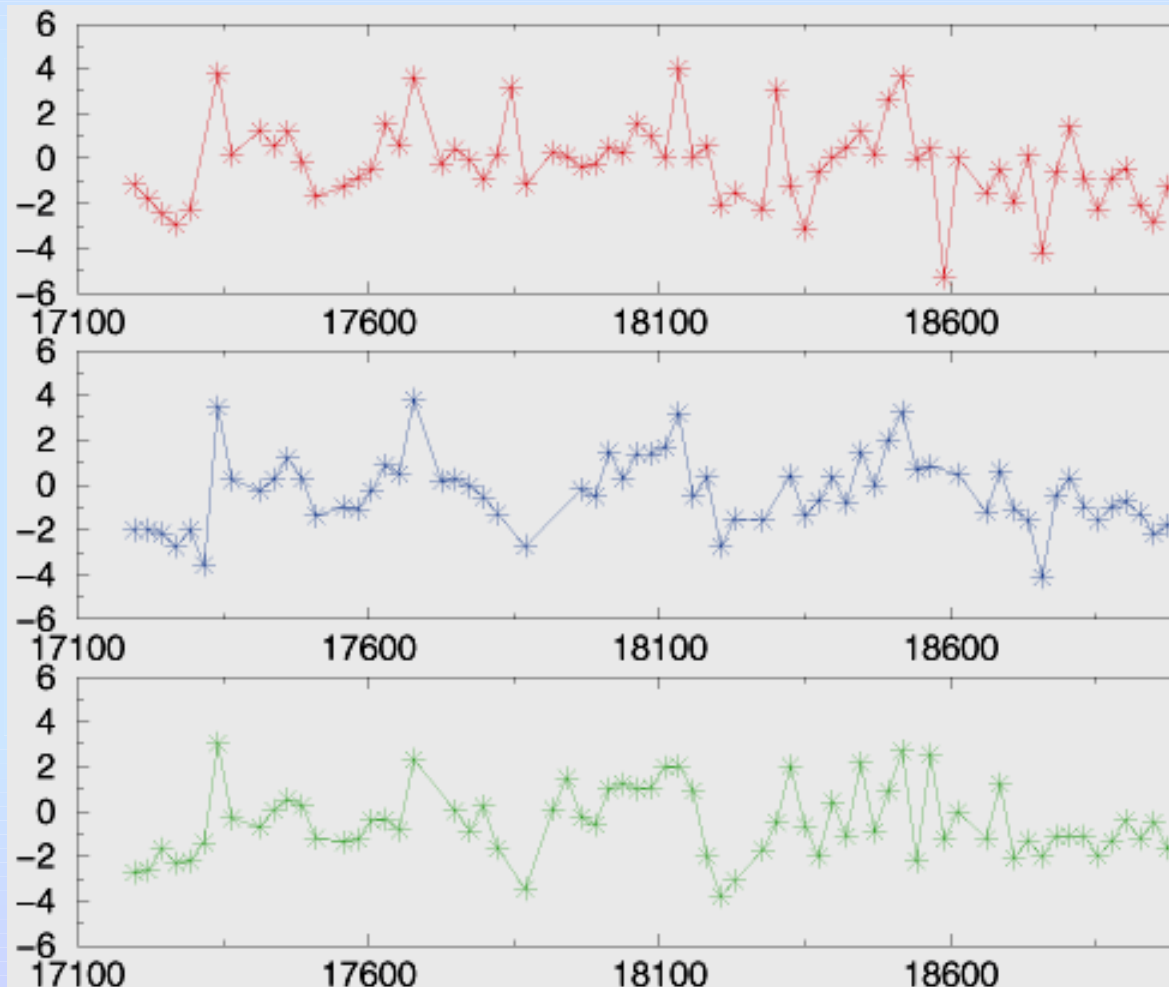


Mutation



Exemple numérique

Séries temporelles pour les trois fonctions « coût » :
« moindres carrés », « norme L1 » et « médiane »



Conclusion



Déroulement d'une estimation par moindres carrés

Acquisition des observations

Choix d'un modèle + paramètres

Restitution par moindres carrés

Solutions

Méthode de repondération des groupes d'observations

Poids optimaux

Test sur les résidus

Observation(s) rejetée(s)

Test sur les paramètres

Paramètre(s) rejeté(s)

Convergence

Solutions finales

$\forall i, \sigma_i^2 \leq 1$

