

# CAHIERS DU BURO

J. C. DEVILLE

## **Estimation linéaire et paramètres bornés**

*Cahiers du Bureau universitaire de recherche opérationnelle.*

*Série Recherche*, tome 29 (1978), p. 3-44

[http://www.numdam.org/item?id=BURO\\_1978\\_\\_29\\_\\_3\\_0](http://www.numdam.org/item?id=BURO_1978__29__3_0)

© Institut Henri Poincaré — Institut de statistique de l'université de Paris, 1978, tous droits réservés.

L'accès aux archives de la revue « Cahiers du Bureau universitaire de recherche opérationnelle. Série Recherche » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

# ESTIMATION LINÉAIRE ET PARAMÈTRES BORNÉS

J.C. DEVILLE

Dans ce qui suit, on s'intéresse aux estimateurs linéaires dans le cadre du modèle linéaire habituel. Après des rappels d'algèbre linéaire et de statistique linéaire (théorème de Gauss-Markov et estimation bayésienne), on étudie le cas où certains paramètres sont astreints à rester dans un domaine borné. On montre que les "bons" estimateurs sont alors nécessairement biaisés. On définit une classe "optimale" d'estimateurs dont on montre qu'ils sont minimax. On fournit un algorithme permettant d'obtenir ces estimateurs. On termine sur des exemples qui font rebondir le problème en posant la question de l'estimation affine.

## SOMMAIRE

	Pages
1. Modèle, algèbre et estimation linéaire . . . . .	5
1.1 Le modèle linéaire . . . . .	5
1.2 Rappels et compléments d'algèbre linéaire . . . . .	6
1.3 Le risque dans l'estimation linéaire . . . . .	9
1.4 Propriétés de Gauss-Markov . . . . .	10
1.5 Estimation bayésienne . . . . .	11
1.6 Amélioration locale de l'estimateur de Gauss-Markov . . . . .	12
2 Estimation d'un paramètre borné . . . . .	13
2.1 Paramètre borné . . . . .	13
2.2 Risque . . . . .	13
2.3 Approche bayésienne . . . . .	14
2.4 Estimateurs "optimaux" . . . . .	15
3. Pratique de l'estimation bornée . . . . .	18
3.1 Algorithme de recherche d'un estimateur optimal . . . . .	18
3.2 Opérateurs et matrices . . . . .	24
3.3 Remarques sur l'estimation bornée . . . . .	27

4	Exemples et compléments . . . . .	28
4.1	Deux exemples simples . . . . .	28
4.2	Un exemple tiré d'une étude économétrique . . . . .	30
4.3	Vers une théorie de l'estimation affine . . . . .	33
4.4	Application à l'estimation de la fonction de Cobb-Douglas . . . . .	40
4.5	Remarques et conclusions . . . . .	43

## 1. — MODELE, ALGEBRE ET ESTIMATION LINEAIRES

## 1.1. Le modèle linéaire

Ce qui suit se situe dans le cadre du modèle linéaire habituel :

$$y = Xb + \epsilon \quad \mathbf{E}\epsilon = 0 \quad \mathbf{E}\epsilon\epsilon^t = \sigma^2 \Omega. \quad (1)$$

La matrice  $X$  est supposée de plein rang  $p$ , de même que la matrice  $\Omega$  ( $\text{rg } \Omega = n > p$ ) de façon à ne pas compliquer les choses.

Le vecteur aléatoire  $y$  appartient à un espace  $E$  de dimension finie  $n$  ; la moyenne  $m = Xb$  du vecteur  $y$  peut varier dans un sous-espace  $L$  de  $E$ . Pour nous, la question consistera à estimer  $m$ , l'estimation de  $b$  n'étant en fait que le calcul des coordonnées de  $m$  dans une base particulière fixée par  $X$ . La covariance du vecteur  $y$  ne dépend pas de  $m$ . C'est la forme bilinéaire positive sur le *dual*  $E^{\star}$  de  $E$  définie par :

$$\forall u, v \in E^{\star} : \sigma^2 \Omega(u, v) = \mathbf{E} \langle u, y - m \rangle \langle v, y - m \rangle.$$

L'opérateur symétrique  $\Omega : E^{\star} \rightarrow E$  associé à cette forme bilinéaire positive est supposé de plein rang et donc inversible ; c'est un produit scalaire sur  $E^{\star}$ . On a aussi de ce fait un produit scalaire sur  $E$  défini par :

$$\forall x, y \in E : (x | y) = \langle \Omega^{-1} x, y \rangle = \langle x, \Omega^{-1} y \rangle. \quad (2)$$

Si, comme nous le ferons dans (presque) toute la suite, on identifie le vecteur  $x$  de  $E$  à la forme linéaire  $\Omega^{-1}x$ , l'opérateur de covariance de  $y$  sera tout simplement l'identité dans  $E$  (à  $\sigma^2$  près), en ce sens que :

$$\begin{aligned} \mathbf{E}(x_1 | \epsilon) (\epsilon | x_2) &= \mathbf{E} \langle \Omega^{-1} x_1, \epsilon \rangle \langle \epsilon, \Omega^{-1} x_2 \rangle = \sigma^2 \langle \Omega^{-1} x_1, x_2 \rangle \\ &= \sigma^2 (x_1 | x_2). \end{aligned}$$

*Le modèle linéaire se résumera donc pour nous à la donnée de :*

*.  $E$  espace euclidien*

*.  $L$  sous-espace de  $E$*

*.  $y$  vecteur aléatoire dans  $E$  de moyenne  $m$  dans  $L$  vérifiant :*

$$\mathbf{E}(u | y - m) (y - m | v) = \sigma^2(u | v), (u, v \in E).$$

Cette reformulation à l'avantage de simplifier les notations et d'utiliser des

notions métriques familières. Elle a cependant les inconvénients suivants :

– Le mélange entre vecteur et forme linéaire. Chaque objet de  $E$  possède deux fonctions et il est parfois difficile de démêler ce qui est lié à l'une ou l'autre d'entre elles.

– On ne traite pas le cas où  $\Omega$  est dégénérée, alors qu'on peut très bien le faire sans l'artifice d'une métrique. L'ennui résulte alors d'une complication formelle plus grande et d'un recours systématique aux espaces quotients. Nous avons préféré y renoncer ici, de façon à mieux mettre en relief l'articulation des idées.

## 1.2. Rappels et compléments d'algèbre linéaire

1.2.1. Soit  $T$  un opérateur dans  $E$ . L'adjoint  $T^\star$  de  $T$  est l'unique opérateur qui vérifie identiquement en  $x$  et  $y$  :

$$(x | Ty) = (T^\star x | y).$$

$T$  est hermitien si  $T = T^\star$ ;  $T$ , hermitien, est positif ( $T \geq 0$ ) si  $(x | Tx) \geq 0$  pour tout  $x$ ;  $T$  est strictement positif ( $T > 0$ ) si  $(x | Tx) > 0$  pour tout  $x$  non nul. On écrira  $T > 0$  si  $T \geq 0$  et  $T \neq 0$  et on dira que  $T$  est plus grand que zéro.  $T_1$  et  $T_2$  étant hermitiens, on écrira  $T_1 \geq T_2$ ,  $T_1 > T_2$  ou  $T_1 \geq T_2$  si  $T_1 - T_2$  est positif, strictement positif, ou plus grand que zéro. Il est clair qu'on a :

$$(T^\star)^\star = T$$

$$(T_1 T_2)^\star = T_2^\star T_1^\star$$

etc.

Si  $P$  est un projecteur d'image  $L$  et de noyau  $N$ ,  $P^\star$  est le projecteur d'image  $(N)^\perp$  et de noyau  $L^\perp$ . En effet  $P^\star$  étant idempotent, comme  $P$ , est un projecteur. Maintenant :

$$x \in \text{Ker } P^\star \Leftrightarrow \forall y \in E : (y | P^\star x) = 0 \Leftrightarrow \forall y \in E : (Py | x) = 0 \Leftrightarrow x \in (\text{Im } P)^\perp$$

et en échangeant le rôle de  $P$  et  $P^\star$  on obtient le résultat quant à l'image de  $P^\star$ .

On notera  $x \otimes y$  l'opérateur  $u \mapsto x (y | u)$  et  $x^{\cdot 2} = x \otimes x$ . De ce fait, la covariance d'un vecteur aléatoire  $Z$  n'est autre que  $\mathbf{E} (Z - m)^{\cdot 2}$ .

Notons que  $(y \otimes x)^\star = x \otimes y$ .

1.2.2. Soit  $T$  un opérateur linéaire d'image contenue dans  $L$ . Notons  $S$  la restriction de  $T$  à  $L$  (la semelle de  $T$ ). Si  $\text{rg } S = \text{rg } T$ , on a toujours une factorisation

$T = SP$  où  $P$  est un projecteur dont le noyau est contenu dans celui de  $T$ . Si  $\text{Im } T = L$ , cette factorisation est unique.

*Démonstration :*

$$\cdot \dim \text{Ker } T = \dim E - \dim \text{Im } T \geq n - p.$$

$\cdot L + \text{Ker } T = E$  car on a  $\text{Im } S = \text{Im } T$ . De ce fait, pour tout  $x$  de  $E$  il existe  $x_L \in L$  tel que  $Tx = Tx_L = Sx_L$ . Or  $x = x_i + (x - x_i)$  et  $(x - x_L) \in \text{Ker } T$ .

$\cdot$  Si  $\text{Im } T = L$ ,  $\text{Ker } T$  est un supplémentaire de  $L$ ,

car sa dimension vaut alors  $n - p$ . Si  $P$  projette sur  $L$  le long de  $\text{Ker } T$  on a bien la factorisation annoncée et celle-ci est unique.

$\cdot$  Enfin si  $\text{Im } T \subset L$  strictement, il existe dans  $\text{Ker } T$  un supplémentaire de  $L \cap \text{Ker } T$ . Ce sous-espace est aussi un supplémentaire de  $L$  et si  $P$  projette alors sur  $L$  le long de ce sous-espace, on obtient que :

$$Tx = Tx_L + Tx_S = Tx_L = SPx \text{ (avec } x_L = Px, x_S = x - Px).$$

Inversement il est clair que tout opérateur de la forme  $SP$  est un opérateur de  $E$  dans  $L$ .

1.2.3. *Ensemble polaire :* Le polaire  $A^\circ$  d'un ensemble  $A$  de  $E$  est

$$A^\circ = \{u ; \sup_{x \in A} |(u|x)| \leq 1\}.$$

*Propriétés :*

$$\cdot A \subset B \Rightarrow A^\circ \supset B^\circ, (\cup_i A_i)^\circ = \cap_i A_i^\circ ;$$

$\cdot A^\circ$  est un ensemble fermé absolument convexe (1) ;

$\cdot$  Si  $A$  est un sous-espace,  $A^\circ = A^\perp$ . Si  $A$  est contenu dans un sous-espace,  $A^\circ$  est un cylindre à génératrices parallèles à  $A^\perp$  ;

$\cdot A^{\circ \circ} = A^\circ ; A^{\circ \circ}$  est la fermeture convexe du symétrisé  $A \cup (-A)$  de  $A$  où  $(-A) = \{x | -x \in A\}$ . On appellera cet ensemble *symétrisé convexifié* de  $A$ .

1.2.4 *Forme quadratique positive généralisée* (f.q.p.g.) : c'est une application  $Q : E \rightarrow \bar{\mathbb{R}}^+$  telle que :

$\cdot$  sur un sous-espace  $\text{Dom } Q$  (domaine de  $Q$ )  $Q$  est une forme quadratique positive,

$\cdot$  hors de  $\text{Dom } Q, Q(x) = +\infty$ .

---

(1) C'est-à-dire vérifiant, pour tout  $a$  et  $b$  tels que  $|a| + |b| \leq 1, x \in A^\circ$  et  $y \in A^\circ \Rightarrow ax + by \in A^\circ$ . Dans un espace réel, ceci revient à dire que  $A^\circ$  est convexe et symétrique.

Le noyau,  $\text{Ker}Q$ , de  $Q$  est le sous-espace  $Q(x) = 0$ .

L'inverse  $Q^{-1}$ , de la f.q.p.g  $Q$  se définit de la façon suivante. Soit  $M = \text{Dom}Q \ominus \text{Ker}Q$  le supplémentaire orthogonal de  $\text{Ker}Q$  dans  $\text{Dom}Q$ . L'espace est ainsi décomposé en trois sous-espaces 2 à 2 orthogonaux :  $\text{Ker}Q$ ,  $M$  et  $(\text{Dom}Q)^\perp$ .

La restriction de  $Q$  à  $M$  est une forme quadratique strictement positive, à laquelle est donc associé un opérateur hermitien inversible  $Q_M$ . On posera :

- $Q^{-1}(x) = 0$  si  $x \in (\text{Dom}Q)^\perp$
- $Q^{-1}(x) = (x | Q_M^{-1}x)$  si  $x \in M$
- $Q^{-1}(x) = +\infty$  si  $x \in (\text{Ker}Q)^\perp = M + (\text{Dom}Q)^\perp$ .

Choisissons, dans  $M$  une base orthonormée de vecteurs propres de  $Q_M$ , dans  $\text{Ker}Q$  et  $(\text{Dom}Q)^\perp$  des bases orthonormées arbitraires. Soit  $(x_i)$  la base de  $E$  obtenue en réunissant ces trois bases. On peut alors écrire :

$$Q = \sum_i \lambda_i x_i \otimes x_i$$

avec  $\lambda_i = 0$  si  $x_i \in \text{Ker}Q$

$\lambda_i = +\infty$  si  $x_i \in (\text{Dom}Q)^\perp = \text{Ker}Q^{-1}$ .

$0 < \lambda_i < +\infty$  si  $x_i \in M$ .

On aura bien :

$$Q(x) = \sum_i \lambda_i (x | x_i)^2.$$

De même :

$$Q^{-1}(x) = \sum_i \lambda_i^{-1} (x | x_i)^2 \quad \text{ou} \quad Q^{-1} = \sum_i \lambda_i^{-1} x_i \otimes x_i$$

avec  $\infty^{-1} = 0$ .

Notons  $(Q \leq 1)$  l'ensemble des  $x$  vérifiant  $Q(x) \leq 1$ .

On a :  $(Q \leq 1)^\circ = (Q^{-1} \leq 1)$ . Cet ensemble n'est autre que l'ellipsoïde indicateur (2) de la forme quadratique  $Q$  dans une structure euclidienne (1).

(1) On peut définir de façon analogue des f.q.p.g sur un espace vectoriel amorphe  $E$ .  $Q^{-1}$  est alors une f.q.p.g. sur le dual  $E^*$  de  $E$ . Le polaire d'une partie de  $E$  est un fermé absolument convexe de  $E^*$ ; il en va de même de l'ellipsoïde indicateur et la relation du texte reste encore vraie dans ce cas, la démonstration étant d'ailleurs la même.

(2) On désignera en fait par ellipsoïde indicateur, soit le "volume" ( $Q^{-1} \leq 1$ ) soit la "surface" ( $Q^{-1} = 1$ ). En principe le contexte lève l'ambiguïté qui peut exister entre ces deux façons de voir.

*Démonstration* : Commençons par le cas où  $Q$  est ordinaire et inversible.  
Alors :

$$(u | x)^2 = (QQ^{-1} u | x)^2 = (Q^{-1} u | Qx)^2 \leq Q^{-1}(u) | Qx \leq Q^{-1}(u) \text{ si } Qx \leq 1.$$

Ce maximum est atteint pour

$$x = Q^{-1} u / \sqrt{Q^{-1}(u)}.$$

Dans le cas général, remarquons d'abord que :  $\text{Ker}Q \subset (Q \leq 1) \subset \text{Dom}Q$ .  
Par suite  $\text{Dom}Q^{\perp} = \text{Ker}Q^{-1} \subset (Q \leq 1)^{\circ} \subset \text{Dom}Q^{-1} = (\text{Ker}Q)^{\perp}$ .

Tout  $u$  de  $(Q \leq 1)^{\circ}$  est donc de la forme  $u = u_M + u'$  avec  $u_M$  dans  $M$  et  $u'$  dans  $\text{Ker}Q^{-1}$ . De même, tout  $x$  de  $(Q \leq 1)$  s'écrit sous la forme  $x = x_M + x'$  avec  $x_M$  dans  $M$  et  $x'$  dans  $\text{Ker}Q$ . Maintenant pour que  $(u | x)^2 \leq 1$  il faut et suffit que  $(u_M | x_M)^2 \leq 1$  ;  $u_M + u'$  sera donc dans  $(Q \leq 1)^{\circ}$  si et seulement si  $u_M \in (Q \leq 1)^{\circ}$ , et pour cela il faut et il suffit que  $u_M \in (Q_M \leq 1)^{\circ} = (Q_M^{-1} \leq 1)$ , ce qui termine la démonstration.

*Remarque* : il est clair que  $Q_1 \leq Q_2 \Leftrightarrow Q_1^{-1} \geq Q_2^{-1}$ .

### 1.3. Le risque dans l'estimation linéaire

*N.B.* : Dans toute la suite (sauf au paragraphe 5.3) on dira estimateur pour application linéaire  $T$  de  $E$  dans  $L$  vérifiant  $\dim T(L) = \text{rg } T$ .

La fonction de perte quadratique utilisée dans le modèle linéaire sera  $\Lambda(x, m) = (x - m) \otimes (x - m)$ . Il s'agit donc d'une fonction définie sur  $L \times L$  à valeurs dans le cône ordonné des opérateurs hermitiens positifs sur  $L$ . Il ne s'agit donc jamais d'une fonction de perte scalaire. Notons cependant que toute fonction de perte quadratique scalaire s'obtient à partir de  $\Lambda$  ; il suffit de faire agir sur  $\Lambda$  une forme linéaire positive dans l'espace des opérateurs hermitiens. On montre (cf. [5] ou [7]) qu'une telle fonction de perte est de la forme  $Q(x - m)$  où  $Q$  est une forme quadratique positive. Or,  $Q(x - m) = \text{Tr}[Q \circ (x - m)^2]$ , ce qui est la forme générale d'une forme linéaire positive sur l'espace des opérateurs hermitiens. Ainsi donc, quand nous obtiendrons un risque minimum unique (Gauss-Markov ou estimation bayésienne), nous obtiendrons d'un même coup un minimum pour toute fonction de perte quadratique scalaire.

Le risque de l'estimateur  $T$  est donc la fonction de  $m$  (et  $\sigma$ ) à valeurs dans le



cône des formes quadratiques positives sur  $L$  définie par :

$$\begin{aligned} R(T, m, \sigma) (\mu) &= \mathbf{E} (\mu | Ty - m)^2 = \mathbf{E} (\mu | T(y - m) + (Tm - m))^2 \\ &= \mathbf{E} (T^{\star} \mu | y - m)^2 + (Tm - m | \mu)^2 \\ &= \sigma^2 \| T^{\star} \mu \|^2 + (Tm - m | \mu)^2. \end{aligned}$$

En utilisant la factorisation  $T = SP$  de 1.2.2, il vient :

$$R(T, m, \sigma) (\mu) = \sigma^2 \| P^{\star} S^{\star} \mu \|^2 + ((S - 1) m | \mu)^2 \quad (3)$$

soit en terme d'opérateur :

$$R(T, m, \sigma) = \sigma^2 SPP^{\star} S^{\star} + (S - 1)m^2 (S^{\star} - 1). \quad (4)$$

#### 1.4. Propriétés de Gauss-Markov

*1.4.1* Le risque ne dépend pas de  $m$  si et seulement si  $S = 1$ , c'est-à-dire si l'estimateur est sans biais. Le risque vaut alors  $\sigma^2 \| P^{\star} \mu \|^2$ . Or  $P^{\star}$  projette sur  $(\text{Ker}P)^{\perp}$  le long de  $L^{\perp}$  ; autrement dit  $P^{\star} \mu - \mu$  est toujours orthogonal à  $L$ . Si  $P_o$  désigne l'orthoprojecteur sur  $L$ ,  $\mu = P_o \mu$  et  $P_o^{\star} = P_o$ . On obtient que :

$$\| P^{\star} \mu \|^2 = \| P_o \mu \|^2 + \| \mu - P_o \mu \|^2.$$

Ceci donne le **théorème de Gauss-Markov** :

*Parmi tous les estimateurs sans biais, il en existe un unique de covariance minimum obtenu par orthoprojection sur  $L$ .*

On remarquera cependant que l'on a  $P^{\star} \mu = \mu$  sur le sous-espace  $L \cap (\text{Ker}P)^{\perp}$  ; ces formes linéaires seront estimées avec une variance minimale par  $P$ .

*1.4.2.* De façon plus générale, à  $m$  fixé, le risque ne dépend de  $P$  que par  $\| P^{\star} S^{\star} u \|^2$ . Le même argument qu'en 1.4.1 ( $\| P^{\star} S^{\star} u \|^2 \geq \| P_o^{\star} S^{\star} u \|^2 = \| S^{\star} u \|^2$ ) montre que tout estimateur de la forme  $SP_o$  est préférable (au sens large) aux estimateurs de forme  $SP$  ( $P \neq P_o$ ). On obtient donc une classe complète d'estimateurs linéaires en prenant des fonctions linéaires de l'estimation de Gauss-Markov, c'est-à-dire en "bricolant" l'espace des paramètres.

Ceci posé on ne s'intéressera plus dans la suite qu'à l'opérateur  $S$  et au risque associé :

$$R(S, m, \sigma) = \sigma^2 SS^{\star} + (S - 1)m^2 (S^{\star} - 1). \quad (5)$$

### 1.5. Estimation bayésienne

On se donne une loi *a priori* du couple  $(m, \sigma)$  ; on pose  $E m^2 = \Gamma_0$  et  $E \sigma^2 = k^2$  ( $k > 0$ ). La forme quadratique  $\Gamma_0$  pourra être une forme quadratique généralisée. Il est facile de voir en effet que, si  $m$  est un vecteur aléatoire mesurable,  $\Gamma_0(u) = E (u | m)^2$  est une f.q.p.g.

Le risque de Bayes est alors l'espérance de  $R$  pour la loi *a priori* soit :

$$R(S, \Gamma_0, k^2) = k^2 S S^{\star} + (S - 1) \Gamma_0 (S^{\star} - 1)$$

où  $A \Gamma A^{\star}$  est la f.q.p.g.  $u \mapsto \Gamma(A^{\star} u)$ .

Si  $\Gamma_0$  est ordinaire, en posant  $\Gamma = \Gamma_0/k^2$ , il vient :

$$k^{-2} R(S, \Gamma_0, k) = (S(1 + \Gamma) - \Gamma) (1 + \Gamma)^{-1} (S(1 + \Gamma) - \Gamma)^{\star} + \\ + \Gamma - \Gamma(1 + \Gamma)^{-1} \Gamma.$$

Le risque admet un minimum unique pour :

$$S_{\Gamma} = \Gamma(1 + \Gamma)^{-1} = (1 + \Gamma)^{-1} \Gamma = (1 + \Gamma^{-1})^{-1}. \quad (6)$$

La valeur de ce minimum est (au facteur  $k^2$  près) :

$$\Gamma - \Gamma(1 + \Gamma)^{-1} \Gamma = 1 - (1 + \Gamma)^{-1} = \Gamma(1 + \Gamma)^{-1} = (1 + \Gamma)^{-1} \Gamma = \\ = (1 + \Gamma^{-1})^{-1} = S_{\Gamma}. \quad (7)$$

Si  $\Gamma = \sum_i \lambda_i x_i \otimes x_i$  pour une base orthonormée  $(x_i)$ , on aura donc :

$$S_{\Gamma} = \sum_i \frac{\lambda_i}{1 + \lambda_i} x_i \otimes x_i. \quad (8)$$

Les formules (6), (7) et (8) restent vraies si  $\Gamma$  est une f.q.p.g. ; de même  $S_{\Gamma}$  fournit toujours l'estimateur de Bayes.

Remarquons :

- $S_{\Gamma}$  est un opérateur hermitien positif de norme inférieure ou égale à 1.
- La correspondance  $\Gamma \Leftrightarrow S_{\Gamma}$  est une bijection entre les opérateurs hermitiens vérifiant  $0 \leq S \leq 1$  et l'ensemble des f.q.p.g. En outre, la formule (6) montre trivialement que cette correspondance est monotone.
- Les deux premières formes du risque (formule (6)) montrent respectivement ce qui est gagné par l'expérimentation par rapport à la loi *a priori* et ce qui est gagné par la loi *a priori* par rapport à l'expérimentation.

— Les estimateurs bayésiens sont admissibles. Supposons en effet que, pour un certain  $S'$ , on ait :  $R(S', m, \sigma^2) \leq R(S_\Gamma, m, \sigma^2)$  pour tout  $m$  et  $\sigma^2$ . En prenant l'espérance par rapport à la loi *a priori*  $(\Gamma_0, k^2)$  on trouve  $E R(S', m, \sigma^2) \leq k^2 S_\Gamma$ . On a donc nécessairement l'égalité et aussi  $S' = S_\Gamma$ . Le problème de savoir s'il existe d'autres estimateurs linéaires admissibles, ou si les hermitiens plus petits que l'unité forment une classe complète, reste ouvert, semble-t-il(1).

### 1.6. Amélioration locale de l'estimateur de Gauss-Markov

Fixons  $m = m_0$  dans  $L$  et  $\sigma = \sigma_0$ . On peut toujours trouver  $S$  tel que :  $R(S, m_0, \sigma_0) < R(1, m_0, \sigma_0) = \sigma_0^2 1$  (risque de Gauss-Markov) sur un voisinage de  $m_0$  et  $\sigma_0$ . Il suffit de prendre  $S = \lambda 1$  avec  $\frac{\|m_0/\sigma_0\|^2}{1 + \|m_0/\sigma_0\|^2} < \lambda < 1$ , ou plus généralement tout  $S$  hermitien vérifiant :

$$S_{m_0} = (m_0/\sigma_0)^2 [1 + (m_0/\sigma_0)^2]^{-1} < S < 1.$$

La vérification de cette propriété est laissée au soin du lecteur. On notera que  $S_{m_0}$  est l'estimateur bayésien obtenu à partir d'une loi qui concentre toute la probabilité sur  $L$  en  $m_0$  et telle que  $E\sigma^2 = \sigma_0^2$ .

On conçoit bien dès lors qu'on puisse améliorer l'estimation de Gauss-Markov à partir du moment où on suppose  $m$  borné. Si  $m$  est contenu dans une boule de rayon  $r$ , on aura  $\|m/\sigma\|^2 / (1 + \|m/\sigma\|^2) \leq \frac{r^2}{\sigma^2 + r^2}$ .

Si  $\sigma \geq k$ , toute homothétie  $S = \lambda 1$  avec  $\frac{r^2}{k^2 + r^2} < \lambda < 1$  améliorera l'estimation de Gauss-Markov. Dans le reste de cet article, on cherche à savoir jusqu'où on peut diminuer le risque sans autres hypothèses. Remarquons que la restriction du champ de  $m$  peut être "naturelle" (on estime une proportion par exemple), ou résulter d'une information *a priori* (si  $m$  est trop grand le modèle est inacceptable). La restriction  $\sigma \geq k$  est moins naturelle. Elle rend compte "physiquement" de l'existence sur la mesure de  $y$  d'erreurs irréductibles, ou d'approximations dans la spécification du modèle.

---

(1) C. GOURIEROUX vient de résoudre le problème en restreignant la classe des estimateurs aux estimateurs hermitiens. Son résultat est le suivant : si  $S$  est hermitien, l'opérateur  $SP_0$  est admissible si et seulement si  $S$  est l'identité ou si  $S$  admet une valeur propre dans  $[0, 1[$ . (note interne de l'E.N.S.A.E., mars 1978).

## 2. ESTIMATION D'UN PARAMETRE BORNE

### 2.1. Paramètre borné.

On suppose désormais que  $m \in D$ , partie arbitraire de  $M$ , à ceci près que nous exigeons, par commodité, que le sous-espace engendré par  $D$  soit  $L$  lui-même. Sinon on pourra dire qu'on a choisi pour sous-espace  $L$ , justement le sous-espace engendré par  $D$ . Nous éliminons ainsi le traitement des estimateurs de Marquardt. (cf. [2]). On suppose aussi, soit que  $\sigma^2$  est fixé, soit que  $\sigma^2 \geq k^2$ .

Posons :

$$F(u) = \sup_{m \in D} (u | m)^2. \quad (9)$$

$F$  est à valeurs dans  $\overline{\mathbf{R}^+}$ ;  $F(u) = +\infty$  signifie que  $m$  peut s'éloigner indéfiniment (dans  $D$ ) de l'hyperplan  $(u | x) = 0$ , c'est-à-dire que la projection de  $m$  sur  $u$  n'est pas bornée. L'ensemble  $(F < \infty)$  est visiblement un sous-espace. On s'en aperçoit en constatant que  $\sqrt{F(u)} = \sup_{m \in D} |(u | m)|$  est une semi-norme, enveloppe

supérieure des semi-normes  $\|u\|_m = |(u | m)|$ . Enfin on a  $F(u) = 0$  si et seulement si  $u = 0$ , sinon  $D$  serait contenu dans l'hyperplan  $(u | m) = 0$  ce qui est contraire à ce que nous avons posé.

Remarquons que le domaine  $(F \leq 1) = D^\circ$  (polaire de  $D$ ). L'hypothèse qui vient de servir équivaut donc à poser que  $(F \leq 1)$  est borné. On peut encore dire, de façon équivalente, que  $D^{\circ\circ}$ , symétrisé fermé convexifié de  $D$ , est un convexe absolument fermé d'intérieur non vide.

### 2.2. Risque

Intéressons-nous au maximum en  $m$  du risque :

$$\overline{R}_\sigma(S)(u) = \text{Max}_{m \in D} R(S, m, \sigma)(u) = \sigma^2 \|S^{\star\star} u\|^2 + F((S^{\star\star} - 1)u), \quad (10)$$

soit aussi :

$$\overline{R}_\sigma(S) = \sigma^2 S S^{\star\star} + F(S^{\star\star} - 1). \quad (11)$$

Si  $F = \Gamma_0$ , forme quadratique généralisée,  $R_\sigma(S)$  n'est autre que le risque de Bayes  $R(S, \Gamma_0, \sigma^2)$ . A  $\sigma$  fixé,  $S_{\Gamma_0/\sigma^2}$  assure un minimum unique de  $R_\sigma(S)$ . Donc :

*Si  $D^{\circ\circ}$  est un ellipsoïde (ou un cylindre elliptique qui est équivalent), et que  $\sigma$  est fixé, il existe un estimateur minimax unique. Le maximum du risque est plus petit que le risque de l'estimation de Gauss-Markov.*

Maintenant si  $\sigma^2 \geq k^2$  n'est pas fixé, on a toujours  $S_{\Gamma_0/k^2} \geq S_{\Gamma_0/\sigma^2}$  et  $k^2 S_{\Gamma_0/k^2} \leq \sigma^2 S_{\Gamma_0/\sigma^2}$ . Il en résulte que  $S_{\Gamma_0/k^2}$  améliore l'estimation de Gauss-Markov indépendamment de  $\sigma$  et est le plus petit estimateur hermitien à réaliser cette performance. De même,  $S_{\Gamma_0/k^2}$  est l'unique estimateur minimisant le maximum du pseudo-risque  $R(S) = \sigma^{-2} R_0(S)$ . Nous dirons que cet estimateur est "optimal".

Que dire dans d'autres cas ? Pour cela nous allons utiliser le détour des techniques bayésiennes.

### 2.3. Approche bayésienne

Soit  $\mathcal{L}(D, k)$  l'ensemble des lois de Bayes en  $(m, \sigma)$  dont le support est contenu dans  $D \times [k, +\infty[$ , c'est-à-dire l'ensemble des lois de Bayes compatibles avec l'information *a priori* dont on dispose. Il est clair que si  $\Gamma$  parcourt l'ensemble des covariances des lois de  $\mathcal{L}$  on a :

$$\sup_{\Gamma \in \{\text{Cov}(\mathcal{L})\}} \Gamma(u) = \sup_{\Lambda \in \mathcal{L}} \sup_{\Lambda} E(u | m)^2 \leq \sup_{m \in D} (u | m)^2 \leq F(u) \quad (12)$$

En fait, on a égalité, comme on le voit facilement en prenant pour  $\Lambda$  une loi dégénérée concentrée, en ce qui concerne  $m$ , au point où le sup est atteint.

Inversement, soit  $\Gamma \leq F$  une forme quadratique. Son ellipsoïde indicateur est contenu dans  $D^{\circ\circ}$ , car :

$$\begin{aligned} \Gamma \leq F &\Leftrightarrow [(F \leq 1) \Rightarrow (\Gamma \leq 1)] \Leftrightarrow [(F \leq 1) \subset (\Gamma \leq 1)] \Leftrightarrow \\ &\Leftrightarrow [(\Gamma \leq 1)^{\circ} \subset (F \leq 1)^{\circ} = D^{\circ\circ}]. \end{aligned}$$

On peut montrer que  $\Gamma$  s'obtient à partir d'une loi de Bayes à support contenu dans  $D \times [k, +\infty[$ , c'est-à-dire l'ensemble des lois de Bayes compatibles avec l'origine (1). Ainsi dire que  $\Gamma \leq F$  équivaut à dire que  $\Gamma$  est la covariance d'une loi de probabilité compatible avec le domaine de variation du paramètre, c'est-à-dire une loi dont le support est contenu dans ce domaine.

---

(1) Il "suffit", si l'ellipsoïde est borné, de prendre la loi concentrée sur l'ellipsoïde et de densité uniforme relativement à la mesure des variétés de dimension  $(n - 1)$  déduite de la mesure des "volumes" dans  $E$  associée au produit scalaire. Dans le cas où l'ellipsoïde est un cylindre, on considère une mesure sur ce cylindre de variance infinie le long des génératrices et de projection uniforme sur la base du cylindre. Pour une mise en place correcte de ces idées, on se reportera aux ouvrages traitant de la mesure sur les variétés dans les espaces de dimension finie, [17] par exemple.

Soit maintenant  $\Gamma \geq F$  une forme quadratique. A  $\sigma$  fixé, intéressons-nous à l'estimateur bayésien correspondant à une loi *a priori* sur  $m$  de covariance  $\Gamma$  (loi "physiquement" impossible). L'opérateur  $S$  est donc :

$$S_\sigma = \frac{\Gamma}{\sigma^2} \left( 1 + \frac{\Gamma}{\sigma^2} \right)^{-1} = (1 + \sigma^2 \Gamma^{-1})^{-1}. \quad (13)$$

Si  $\Gamma = \sum_i \lambda_i (x_i \otimes x_i)$ , on a  $S_\sigma = \sum_i \frac{\lambda_i/\sigma^2}{1 + \lambda_i/\sigma^2} x_i \otimes x_i$ .

On a alors :

$$\bar{R}(S_\sigma) = \sigma^2 S_\sigma S_\sigma^* + F(S_\sigma^* - 1) \leq \sigma^2 S_\sigma S_\sigma^* + (S_\sigma - 1) \Gamma (S_\sigma^* - 1) < \sigma^2 1.$$

Si on suppose donc  $\sigma \geq k$ , l'estimateur  $S_k$  améliorera strictement l'estimation de Gauss-Markov tout en étant amélioré, tant au point de vue du risque bayésien que du point de vue du risque maximum  $\bar{R}$ , par tout estimateur bayésien fourni par une loi de  $\mathcal{L}(D, k)$ .

## 2.4. Estimateurs "optimaux"

2.4.1 Nous allons maintenant caractériser les estimateurs "optimaux" en ce sens que :

- (i) Ils améliorent l'estimation de Gauss-Markov.
- (ii) Ils sont améliorables tant sur le plan du risque bayésien que sur celui du risque maximum par tout estimateur de Bayes fourni par une loi de  $\mathcal{L}(D, k)$ .
- (iii) Si  $S$  est optimal, il est élément minimum dans la classe des opérateurs auto-adjoints vérifiant la propriété (ii). Autrement dit :

$$\left. \begin{array}{l} S' \leq S \\ \text{et } S' \text{ vérifie (ii)} \end{array} \right\} \Rightarrow S' = S.$$

Ces estimateurs sont caractérisés de la façon suivante :

*Définition* : Soit  $\Gamma_0$  une forme quadratique minimale dans l'ensemble  $\Gamma \geq F$ . Un estimateur optimal est donné par

$$T = S_0 P_0 \text{ avec } S_0 = \frac{\Gamma_0}{k^2} \left( 1 + \frac{\Gamma_0}{k^2} \right)^{-1} = (1 + k^2 \Gamma_0^{-1})^{-1}$$

et  $P_0$  estimateur de Gauss-Markov.

Dire que  $\Gamma_0$  est minimale dans la classe  $\Gamma \geq F$ , c'est dire que

$$F \leq \Gamma \leq \Gamma_0 \Rightarrow \Gamma = \Gamma_0.$$

Il est donc clair que les estimateurs optimaux vérifient les trois propriétés qu'on exigeait d'eux et qu'ils sont les seuls à les posséder.

Par définition, les formes quadratiques de la classe  $(\Gamma \geq F)$  sont celles dont l'ellipsoïde indicateur contient le domaine  $D^{\circ\circ}$  (et donc  $D$ ). Les éléments minimaux de cette classe sont donc les formes quadratiques dont l'ellipsoïde est le plus petit tout en contenant le domaine  $D$ .

#### 2.4.2 Existence d'estimateurs optimaux

Elle est assurée par le lemme de Zorn. Soit, en effet,  $A$  un ensemble totalement ordonné et  $\Gamma_\alpha$  une famille de f.q.p.g. telle que  $\Gamma_\alpha \leq \Gamma_\beta$  dès que  $\alpha \leq \beta$ . Posons  $\Gamma(u) = \inf_{\alpha} \Gamma_\alpha(u)$ . Il est clair que :

$$- \forall u : \Gamma(u) \geq 0 \text{ et } \forall \lambda \in \mathbf{R} : \Gamma(\lambda u) = \lambda^2 \Gamma(u).$$

-  $\Gamma(u) < +\infty \Leftrightarrow \exists \alpha : \Gamma_\alpha(u) < +\infty$ .  $\Gamma$  est donc finie sur le sous-espace  $\text{Dom } \Gamma = \bigcup_{\alpha} \text{Dom } \Gamma_\alpha$  (remarquer que  $\text{Dom } \Gamma_\alpha$  est une famille monotone de sous-espaces).

- Comme tous les sous-espaces  $\Gamma_\alpha$  sont de dimension finie, il existe  $\alpha_0$  tel que, si  $\alpha \leq \alpha_0$ ,  $\text{Dom } \Gamma_\alpha = \text{Dom } \Gamma$ . Pour de tels  $\alpha$ ,  $\Gamma_\alpha(u + v) - \Gamma_\alpha(u) - \Gamma_\alpha(v)$  est une forme bilinéaire sur  $\text{Dom } \Gamma$ , et par passage à la limite il en va de même de  $\Gamma(u + v) - \Gamma(u) - \Gamma(v)$ .  $\Gamma$  est donc une forme quadratique sur  $\text{Dom } \Gamma$  et par suite une f.q.p.g..

- Enfin si  $\forall \alpha : \Gamma_\alpha \geq F$ , on a aussi  $\Gamma \geq F$ .

L'ensemble  $(\Gamma \geq F)$  est donc inductif et possède (au moins) un élément minimal.

#### 2.4.3 Unicité d'un estimateur optimal

La remarque du paragraphe 2.2 montre que si  $D^{\circ\circ}$  est un ellipsoïde centré à l'origine il existe un estimateur optimal unique. Réciproquement il n'existe qu'un seul estimateur optimal uniquement dans le cas où  $D^{\circ\circ}$  est un ellipsoïde.

Il est clair, en effet, que l'ellipsoïde indicateur d'une f.q.p.g. associée à un estimateur optimal possède des points de contact avec  $D^{\circ\circ}$  (sinon il suffirait de le raccourcir par homothétie). Supposons donc  $\Gamma_0^{\star}$  optimal. Si l'ellipsoïde indicateur de  $\Gamma_0^{\star}$  est  $D^{\circ\circ}$ ,  $\Gamma_0^{\star}$  est l'unique estimateur optimal. Sinon on peut trouver  $x_0^{\star}$  à la

frontière de  $D^{\circ\circ}$  et à l'intérieur de cet ellipsoïde. Soit  $u$  un vecteur tel que  $(u | x) = 1$  soit l'équation d'un hyperplan d'appui en  $x_0^{\star}$  à  $D^{\circ\circ}$  ;  $D^{\circ\circ}$  est tout entier contenu dans "l'ellipsoïde" :

$$\{\Gamma_1^{-1}(x) = (u | x)^2 \leq 1\}.$$

Il est clair que  $\Gamma_1 \geq F$  et que  $\Gamma_0^{\star}$  et  $\Gamma_1$  ne sont pas comparables. On pourra donc trouver un élément minimal  $\Gamma_2$  dans la classe  $F \leq \Gamma \leq \Gamma_1$  qui sera différent de (et non comparable avec)  $\Gamma_0^{\star}$ .

#### 2.4.4 Propriétés minimax

Parmi les estimateurs  $T = SP_0$  avec  $S$  auto-adjoint, les estimateurs optimaux minimisent le maximum du pseudo-risque  $\bar{R}(S) = \sigma^{-2} R_{\sigma}(S)$  ; ceci revient à dire qu'ils sont minimax si on fixe  $\sigma = k$ .

Ce résultat découle en fait de l'algorithme décrit dans les paragraphes suivants. Comme on va le voir, un estimateur optimal s'obtient à partir d'une f.q.p.g.  $\Gamma$  dont l'ellipsoïde indicateur admet un contact maximum avec  $D^{\circ\circ}$ . Ce contact maximum détermine réciproquement et de façon univoque un ellipsoïde.

Soit alors  $S'P_0$  un autre estimateur avec  $S'$  auto-adjoint, soit  $\Gamma'$  la f.q.p.g. associée à  $S'$ , et supposons que l'on ait  $\bar{R}(S') \leq \bar{R}(S)$ . L'ellipsoïde indicateur de  $\Gamma'$  doit avoir avec  $D^{\circ\circ}$  les mêmes éléments de contact que l'ellipsoïde indicateur de  $\Gamma$  (sinon  $\Gamma'$  et  $\Gamma$  ne seraient pas comparables). Ces ellipsoïdes sont donc nécessairement confondus et on aura  $S = S'$ .



### 3. PRATIQUE DE L'ESTIMATION BORNEE

#### 3.1. Algorithme de recherche d'un estimateur optimal

##### 3.1.1 Mise en place

On cherche à déterminer une forme quadratique minimale dans  $\{\Gamma \geq F\}$ . Soit  $Q = \Gamma^{-1}$  ; il est équivalent de chercher un ellipsoïde ( $Q \leq 1$ ) minimum parmi ceux qui contiennent  $D^{\circ\circ}$ . On supposera dans la suite que  $D^{\circ\circ}$  est un polyèdre, enveloppe convexe d'une famille finie de points extrêmes  $\{x_\alpha\}$ . Pour des raisons de symétrie, on pourra se contenter d'une famille ne comportant qu'un point sur deux (si  $x$  est extrême,  $-x$  l'est aussi). Dans le cas d'un domaine quelconque, l'algorithme se généralise facilement à ceci près que :

a)  $\{x_\alpha\}$  n'est pas fini, et l'algorithme est "théorique" en ce sens qu'il ne peut pas être traité par un ordinateur.

b) A chaque étape il faudra examiner la nature des contacts entre  $\{Q = 1\}$  et la frontière de  $D^{\circ\circ}$ . Il faudra en effet déceler les directions  $\nu$  de contact d'ordre supérieur à 2 (id est, telles que :  $\lim_{t \rightarrow 0} \inf \frac{1}{t^2} [\Gamma(u_0 + t\nu) - F(u_0 + t\nu)] = 0$ , avec  $u_0$  vérifiant  $\Gamma(u_0) = F(u_0)$ ).

##### 3.1.2 Principe de l'algorithme

Supposons trouvé un ellipsoïde  $Q = 1$  ( $Q$  inversible) ayant avec  $D^{\circ\circ}$  un contact en  $p$  points  $x_1, \dots, x_p$ , linéairement indépendants et faisant partie de la famille  $\{x_\alpha\}$ . Pour améliorer  $Q$  il faut trouver un ellipsoïde ayant avec  $D^{\circ\circ}$  un contact en  $x_1, \dots, x_p$  et ayant en chacun de ces points le même hyperplan tangent qu'avec ( $Q = 1$ ) (sinon les f.q.p.g. associées ne seraient pas comparables). Le plan tangent en  $x_i$  à ( $Q = 1$ ) a pour équation  $(x | Qx_i) = 1$ . Les vecteurs  $Qx_i$  sont linéairement indépendants, de sorte que, si  $p < n$ , les noyaux des formes linéaires  $Qx_i$  ont une intersection  $N$  de dimension  $n - p$ .

On va maintenant "rétrécir" l'ellipsoïde de la façon suivante :

Le sous-espace  $M$  engendré par les  $x_i$  est supplémentaire de  $N$ . On va "déformer" ( $Q = 1$ ) en lui faisant subir une affinité de base  $M$  le long de  $N$ . Autrement dit, pour  $\lambda$  décroissant à partir de 1, on va transformer ( $Q = 1$ ) continuellement par la famille de transformations :

$$x \mapsto x_M + \lambda x_N$$

où  $(x_M, x_N)$  désigne la décomposition canonique de  $x$  sur les deux sous-espaces. Il est clair qu'on va obtenir un nouveau contact en un point  $x_{p+1}$  pour un nombre  $\lambda^* \leq 1$ .

Si  $p + 1 = n$ , on ne pourra plus rétrécir l'ellipsoïde et on aura obtenu un élément minimum ( $x_{p+1}$  sera en effet indépendant des  $x_1, \dots, x_p$  par construction). Si  $p + 1 < n$  on peut recommencer l'opération (compte tenu des remarques du paragraphe 3.1.4)

*Remarque* : Dans ce qui précède, on suppose le domaine totalement borné. On voit bien qu'on peut se contenter de ce cas. Supposons en effet que  $D$  soit non borné le long d'un sous-espace  $N$  et borné totalement dans un supplémentaire  $B$  de  $N$  (de fait on peut prendre  $B = \{u; F(u) < \infty\}$  et  $N = B^\perp$ ). Il suffira de chercher un ellipsoïde  $E_B$  minimum dans  $B$ ; l'"ellipsoïde" minimum sera le cylindre à génératrices parallèles à  $N$  s'appuyant sur  $E_B$ .

### 3.1.3 Calculs et pratique de l'algorithme

Pour initialiser l'algorithme il suffit de choisir arbitrairement une forme quadratique  $Q$  non dégénérée. Soit  $\rho = \max_{\alpha} Q(x_\alpha)$ , ce maximum étant atteint

en un point  $x_1$ . La forme quadratique  $Q_0 = \frac{1}{\rho} Q$  vérifie  $Q_0(x_1) = 1$  et  $Q(x_\alpha) < 1$  pour tous les autres points (le cas particulier où il y a égalité est traité au paragraphe suivant). L'ellipsoïde  $Q_0 = 1$  contient donc  $D^{\circ\circ}$  (pour des raisons de convexité) et admet un contact en  $x_1$ .

Intéressons-nous maintenant au traitement numérique de l'amélioration décrite en 3.1.2. L'ellipsoïde "aplatis de  $\lambda$ " vérifie l'équation :

$$Q\left(x_M + \frac{1}{\lambda} x_N\right) = 1 \quad (\lambda \in ]0, 1]).$$

Or :

$$\begin{aligned} Q\left(x_M + \frac{1}{\lambda} x_N\right) &= Q\left(\left(1 - \frac{1}{\lambda}\right)x_M + \frac{1}{\lambda} x\right) = \\ &= \left(1 - \frac{1}{\lambda}\right)^2 Q(x_M) + \frac{1}{\lambda^2} Q(x) + 2\left(1 - \frac{1}{\lambda}\right)\frac{1}{\lambda} (Qx_M | x). \end{aligned}$$

Mais  $M$  et  $N$  sont  $Q$ -orthogonaux et  $(Qx_M | x) = (Qx_M | x_M) = Q(x_M)$ .

L'équation de l'ellipsoïde aplati est donc :

$$\mu Q(x) + (1 - \nu) Q(x_M) = 1 \quad \text{avec} \quad \mu = \frac{1}{\lambda^2} \in [1, \infty[ .$$

Reste à évaluer  $Q(x_M)$ , c'est-à-dire, en fait, la projection  $x_M$  de  $x$  le long de  $N$ . Cherchons-la dans la base  $x_1 \dots, x_p$  c'est-à-dire sous la forme :

$$Px = \sum_{i=1}^p (v_i | x) x_i.$$

Dire que  $P$  projette le long de  $N$  implique que :

$$\forall j (x - Px | Qx_j) = 0,$$

ce qui conduit au système d'équations :

$$j = 1, \dots, p : Qx_j = \sum_i (x_i | Qx_j) v_i.$$

Les  $x_i$  étant indépendants, la matrice  $(x_i | Qx_j)$  admet une inverse d'éléments  $\alpha_{ij}$  de sorte que :

$$v_i = \sum_j \alpha_{ij} Qx_j .$$

Remarquons que :  $(v_i | x_j) = \delta_{ij}$  (symbole de Kronecker).

On a alors :

$$Q(x_M) = (Px | QPx) = \sum_{i,j} \alpha_{ij} (Qx_j | x) (Qx_i | x).$$

Autrement dit la forme quadratique  $Q \circ P$  est associée à l'opérateur :

$$R = \sum_{ij} \alpha_{ij} Qx_i \otimes Qx_j . \quad (14)$$

L'amélioration définie en 3.1.2 va donc se faire de la façon suivante :

– Calcul de  $Q(x_\alpha)$  et  $R(x_\alpha)$  pour les  $x_\alpha$  non atteints antérieurement. Tous ces nombres sont inférieurs à un.

– Calcul de  $\mu_\alpha = (1 - R(x_\alpha))/(Q(x_\alpha) - R(x_\alpha))$ .

- On détermine  $\alpha_0$  tel que  $\mu_{\alpha_0}$  soit un minimum et on pose  $x_{p+1} = x_{\alpha_0}$ .
- On remplace  $Q$  par  $Q' = \mu_{\alpha_0} Q + (1 - \mu_{\alpha_0}) R$ .
- On calcule la nouvelle matrice  $(x_i | Q' x_j)$   $i, j = 1, \dots, p + 1$ .

Or :

$$(x_i | Q' x_j) = (x_i | Q x_j) \text{ si } i, j = 1 \dots p \text{ par construction de } R.$$

$(x_i | Q' x_j) = (x_i | Q x_j)$  si  $i$  ou  $j \neq p + 1$  par le fait que  $R$  et  $Q$  définissent la même forme bilinéaire dès qu'un des vecteurs est dans l'espace engendré par  $x_1, \dots, x_p$ .

$$\text{Enfin } (x_{p+1} | Q' x_{p+1}) = 1 \text{ par construction.}$$

### 3.1.4 Problèmes de colinéarité

Dans ce qui précède on a supposé qu'à chaque étape de l'amélioration on travaillait avec des  $(x_i)_{i=1, \dots, p}$  linéairement indépendants. L'algorithme fournit alors un nouveau "point-limite" hors du sous-espace engendré par les  $(x_i)$  puisqu'il y a contraction dans le sous-espace  $N$  qui est supplémentaire de  $M$ . Il ne peut donc se produire de colinéarités que si on atteint plusieurs nouveaux points à la fois, soient  $x_{p+1}, \dots, x_{p+k}$ .

Examinons-les séquentiellement. Avec  $x_{p+1}$  on obtient de toute façon une dimension supplémentaire ; si  $x_{p+2}$  fournit lui aussi une dimension supplémentaire, l'algorithme se poursuit comme dans le cas général : on aura seulement  $\mu_{\alpha_0} = 1$  et  $Q^{\alpha} = Q$ . Par contre si  $x_{p+2}$  est colinéaire à  $(x_1, \dots, x_{p+1})$ , on aura :

$$x_{p+2} = \sum_1^{p+1} \alpha_i x_i \quad \text{et} \quad Q x_{p+2} = \sum_1^{p+1} \alpha_i Q x_i ;$$

de ce fait, le plan tangent à l'ellipsoïde en  $x_{p+2}$  sera toujours déterminé par les plans tangents aux autres points. Il suffit alors d'éliminer ce point de l'ensemble des données, de noter qu'on a une solution limite, correspondant à la situation de la figure 1 et de continuer l'algorithme avec  $(p + 1)$  points classés et  $N - p - 2$  points à classer.

On peut ainsi caractériser les solutions obtenues et montrer que, moyennant une initialisation convenable, on les obtient toutes. On peut, en particulier, initialiser l'algorithme avec une forme quadratique du genre  $Q(x) = (u | x)^2$  (dégénérée) qui correspond à  $\Gamma = Q^{-1}$  forme quadratique généralisée. On pourra ainsi obtenir des solutions constituées par des "ellipsoïdes" formés de deux hyperplans contenant une face du polyèdre, ou par des ellipsoïdes cylindriques.

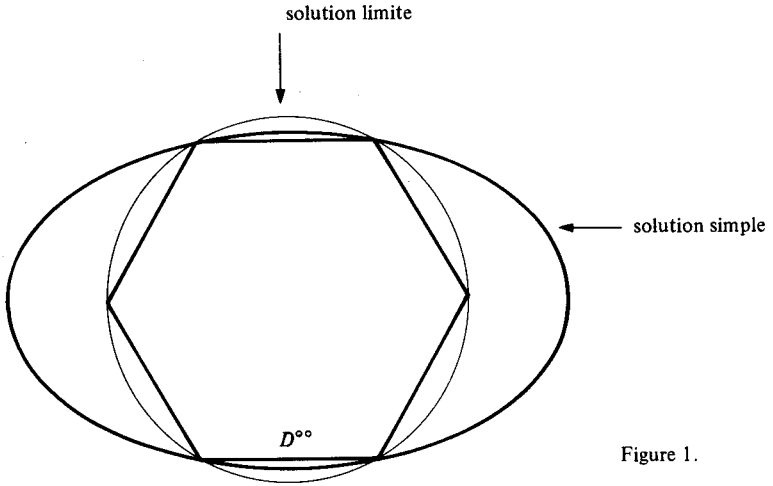


Figure 1.

On pourra aussi, surtout dans les cas de dimension petite, fabriquer directement des  $Q$  optimales en choisissant  $p$  (couples de) points extrêmes de  $D^{\infty}$  ainsi que des plans tangents "compatibles" de façon à ne pas tomber sur une solution impossible comme à la figure 2.

*Remarque* : ces problèmes de colinéarité sont exactement de même nature que ceux évoqués en 3.1. 1 point b) (nature d'un contact). Ceci deviendrait clair par un traitement des problèmes de contact au moyen de l'analyse non standard (cf. [16]).

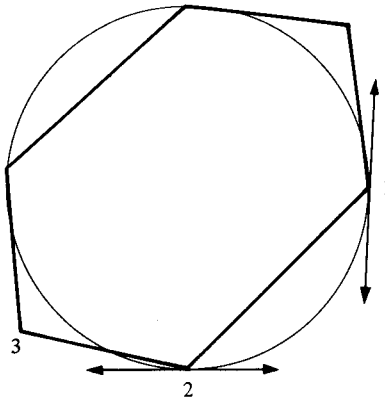
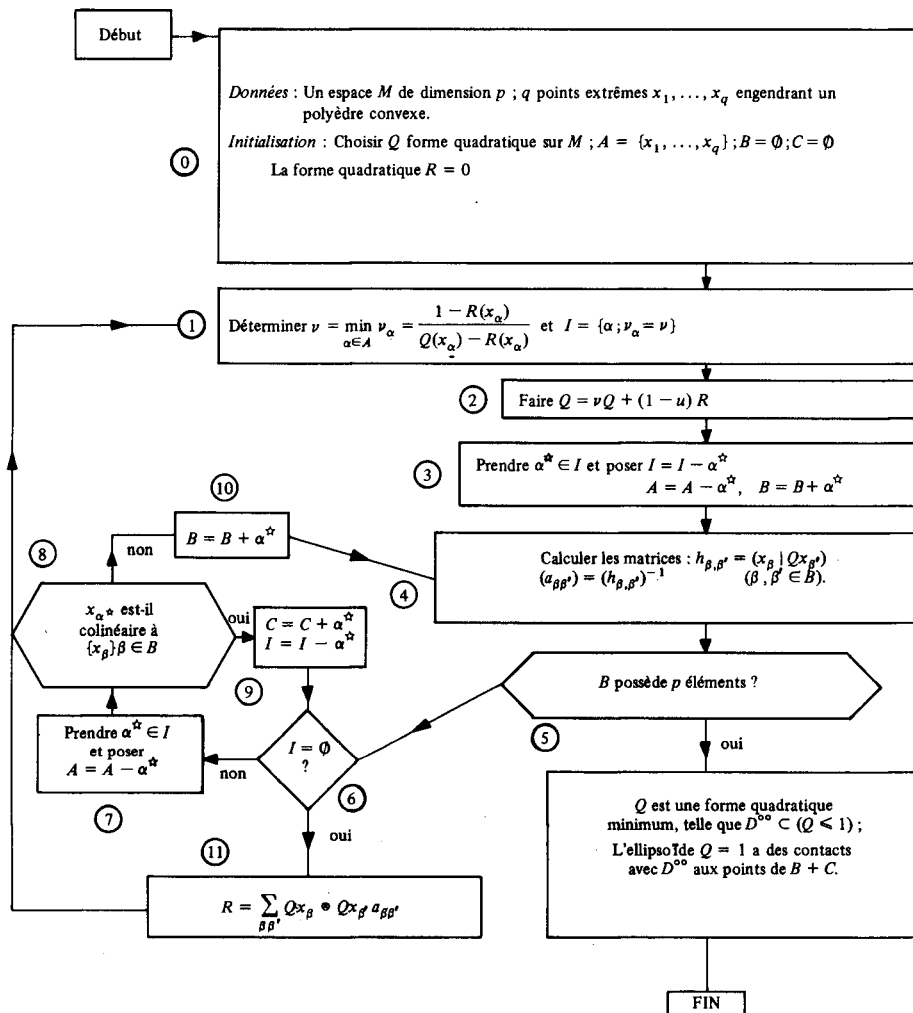


Figure 2.

On détermine un ellipsoïde à partir de deux éléments de contact en 1 et 2. La solution obtenue n'est pas acceptable, car les plans tangents ne sont pas compatibles. L'ellipsoïde obtenu à partir de ces éléments de contact ne contient pas le domaine.

### 3.1.5 Formalisation et organigramme de l'algorithme

L'analyse qui précède se résume dans l'organigramme suivant :



$A$  : ensemble des points qui ne sont pas retenus à une étape donnée.

$B$  : ensemble des points de contact servant de base à une étape donnée.

$C$  : ensemble des points de contact ne servant pas de base (colinéaires à  $B$ )

$I$  : ensemble des points réalisant le minimum.

*Commentaires sur les blocs 4 et 11.*

*Bloc 4.* Le calcul de la matrice  $H = (h_{\beta, \beta'})$  est en fait très simple. Soit  $H^{\star}$  la matrice  $H$  de l'étape précédente,  $B^{\star}$  l'ensemble  $B$  de l'étape précédente. On a  $h_{\beta, \beta'} = h_{\beta, \beta'}^{\star}$  si  $\beta$  et  $\beta'$  sont dans  $B^{\star}$ . Par construction  $h_{\alpha^{\star}, \alpha^{\star}} = 1$ . Reste à calculer  $(x_{\alpha^{\star}} | Qx_{\beta})$  pour  $\beta \in B^{\star}$ .

L'inversion de  $H$  est immédiate du fait qu'on connaît déjà  $H^{\star-1}$ . On doit donc inverser une matrice partitionnée de la forme :

$$H = \left[ \begin{array}{c|c} H^{\star} & V \\ \hline V^t & 1 \end{array} \right] \quad (15)$$

où  $V$  est le vecteur des  $(x_{\alpha^{\star}} | Qx_{\beta})$ ,  $\beta$  dans  $B^{\star}$ . Il est facile de voir que :

$$H^{-1} = \frac{1}{\delta} \left[ \begin{array}{c|c} H^{\star-1} (\delta + VV^t) H^{\star-1} & H^{\star-1} V \\ \hline -V^t H^{\star-1} & 1 \end{array} \right] \quad (16)$$

avec  $\delta = 1 - V^t H^{\star-1} V$ .

*Bloc 11.* Si on travaille dans une base telle que l'on ait :  $(x | y) = x^t M y$  (on assimile un vecteur à la matrice de ses coordonnées), l'opérateur  $Qx_{\beta} \otimes Qx_{\beta'}$  définit la forme quadratique ( $Q$  désignant la matrice de l'application linéaire  $Q$ ) de matrice :

$$MQx_{\beta} x_{\beta'}^t, QM, \quad (17)$$

de sorte que la matrice de la forme quadratique  $R$  sera :

$$MQ \left( \sum_{\beta, \beta'} \alpha_{\beta, \beta'} x_{\beta} x_{\beta'}^t \right) QM. \quad (18)$$

Le paragraphe suivant va préciser les liaisons entre matrices et opérateurs et donner une "traduction" matricielle de ce qui précède.

## 3.2 Opérateurs et matrice

### 3.2.1 Généralités.

L'utilisation d'une métrique (plutôt que d'un schéma de dualité cf. [1]) est cause d'ambiguïtés, de confusions, liées fondamentalement à l'assimilation d'une forme linéaire (élément du *dual*) à un vecteur de l'espace.

Soit  $E$  un espace euclidien muni d'une base. A tout opérateur  $A$  dans  $E$  il correspond deux matrices selon qu'on le fait fonctionner comme application linéaire (a.l.) ou comme forme bilinéaire (f.b.). Si  $A$  est hermitien, la forme bilinéaire associée définit une forme quadratique (f.q.). Si  $A$  est l'identité de  $E$ , sa matrice a.l. est l'identité quelle que soit la base. La f.q.  $(x|x)$  de l'opérateur identité est donnée matriciellement par la matrice métrique  $M$  telle que :  $(x|x) = \underline{x}^t M \underline{x}$  (on note  $\underline{x}$  la matrice colonne de coordonnées du vecteur  $x$  dans la base donnée). Dans les problèmes de régression linéaire, la matrice métrique sur l'espace des paramètres est  $\underline{X}^t \underline{\Omega}^{-1} \underline{X}$

et 
$$(x|y) = \underline{x}^t \underline{X}^t \underline{\Omega}^{-1} \underline{X} y = (\underline{X} \underline{x})^t \underline{\Omega}^{-1} (\underline{X} y).$$

Désignons systématiquement par  $\underline{A}$  la matrice a.l. de  $A$ . La matrice f.q. de  $A$  est  $\underline{M} \underline{A}$ , car  $\underline{x}^t \underline{M} \underline{A} \underline{x} = (x|Ax)$ . La matrice f.q. de l'adjoint  $A^*$  est  $\underline{A}^t \underline{M}$  de sorte que la matrice a.l. de  $A^*$  est  $\underline{M}^{-1} \underline{A}^t \underline{M}$ .  $A$  est auto-adjoint si et seulement si  $\underline{A}^t \underline{M} = \underline{M} \underline{A}$ . Si  $A$  est inversible, la matrice a.l. de  $A^{-1}$  est  $\underline{A}^{-1}$ ; la matrice f.q. de  $A^{-1}$  est  $\underline{M} \underline{A}^{-1}$ .

Inversement, à toute matrice  $\underline{A}$  il correspond deux opérateurs selon que  $A$  fonctionne comme a.l. ou comme f.q. La matrice identité correspond à l'application linéaire identique et à la f.q. :

$$\underline{x}^t \underline{x} = \underline{x}^t \underline{M} \underline{M}^{-1} \underline{x} = (x|M^{-1}x),$$

où  $M^{-1}$  est l'opérateur de matrice a.l.  $\underline{M}^{-1}$ .

De même la matrice  $\underline{M}$  a pour opérateur f.q. l'identité, et une matrice  $\underline{A}$ , définissant une a.l.  $A$ , définit aussi un opérateur par  $\underline{y}^t \underline{A} \underline{x} = (y|Ax)$ ;  $\underline{A}$  est l'opérateur de matrice (a.l.)  $\underline{M}^{-1} \underline{A}$ .

### 3.2.2 Traduction matricielle des résultats précédents

Dans la pratique, le domaine de  $m$  sera donné par des restrictions sur certains paramètres. Le modèle sera donné sous la forme :

$$y = \sum_{i \in I} X_i b_i + \sum_{j \in J} X_j b_j + \epsilon; \quad \mathbf{E} \epsilon = 0; \quad \mathbf{E} \epsilon \epsilon' = \sigma^2 \Omega, \quad \sigma^2 \geq k^2$$

$\{b_i\}$  sans contraintes

$\{b_j\}$  dans un domaine borné.

L'espace  $L$  de  $m = \sum_{I \cup J} b_i X_i$  est décomposé en deux supplémentaires,  $N$

engendré par les  $(X_i)_{i \in I}$  et  $M$  engendré par les  $(X_j)_{j \in J}$ . Notons  $P$  le projecteur sur  $M$  le long de  $N$ , et pour tout  $x$  de  $L$  :  $x_M = Px$  et  $x_N = (1 - P)x$ .



Les "ellipsoïdes optimaux" vont être des cylindres à génératrices parallèles à  $N$  et à base elliptique qu'il est naturel et commode de prendre dans  $M$ . L'équation d'un tel cylindre sera  $Q(x_M) = 1$ ,  $Q$  étant une f.q. ordinaire inversible sur  $M$  seulement. On restreindra donc l'algorithme à la recherche d'une matrice  $Q$  sur l'espace  $M$  rapporté à la base  $X_j$ . La forme quadratique sur  $L$  aura alors l'expression matricielle :

$$(b_I^t, b_J^t) \begin{bmatrix} 0 & 0 \\ 0 & Q \end{bmatrix} \begin{bmatrix} b_I \\ b_J \end{bmatrix},$$

où  $b_I$  et  $b_J$  désignent des vecteurs colonnes de coefficients.

Reste à traduire matriciellement l'opérateur  $S$  et l'estimateur optimal  $SP_0$ .

On sait déjà que :

$$P_0 y = (X^t \Omega^{-1} X)^{-1} X^t \Omega^{-1} y \quad (\text{estimation des m.c.g.})$$

Pour  $S$  on a :

$S = (1 + \Gamma^{-1})^{-1} = (1 + k^2 \hat{Q})^{-1}$  (puisque  $\Gamma^{-1} = k^2 Q$ ), où  $\hat{Q}$  est l'opérateur application linéaire tel que :

$$(b \mid \hat{Q}b) = b^t Q b.$$

$\hat{Q}$  a donc pour matrice a.l.  $M^{-1}Q$  de sorte que la matrice a.l. de  $(1 + k^2 \hat{Q})^{-1}$  est :

$$(1 + k^2 M^{-1}Q)^{-1} = (M^{-1}M + k^2 M^{-1}Q)^{-1} = (M + k^2 Q)^{-1} M.$$

L'opérateur  $S$  a donc la matrice (avec  $M = X^t \Omega^{-1} X$ ) :

$$(M + k^2 Q)^{-1} M = (X^t \Omega^{-1} X + k^2 Q)^{-1} X^t \Omega^{-1} X = 1 - k^2 (X^t \Omega X + k^2 Q)^{-1} Q.$$

Entre l'estimation  $\hat{b}$  de Gauss-Markov et l'estimation améliorée  $\hat{\hat{b}}$  on a donc la relation :

$$\hat{\hat{b}} = \hat{b} - k^2 (X^t \Omega^{-1} X + k^2 Q)^{-1} Q \hat{b}. \quad (19)$$

L'estimateur optimal lui-même a l'expression :

$$T = SP_0 = (X^t \Omega^{-1} X + k^2 Q)^{-1} X^t \Omega^{-1}.$$

On reconnaît donc un estimateur de type ridge généralisé. Les "ridge-ordinaire" apparaissent ici comme très particuliers ; dans le cas où  $\Omega = I$ , on aura un estimateur optimal du type ridge si et seulement si l'algorithme du paragraphe 3.2 donne une solution optimale dès la première itération en utilisant l'initialisation

$Q = I$ . Il est clair que cela correspond à un domaine de variation du paramètre très particulier (dans la métrique ordinaire, un tel domaine comporte autant de points extrêmes que de paramètres à la même distance de l'origine).

### 3.3. Remarques sur l'estimation bornée

Pratiquement, l'estimation bornée peut s'appliquer au moins dans trois types de situation :

a) les paramètres sont par essence bornés : ce sont des proportions, des cosinus, des corrélations par exemple.

b) On sait d'avance (information *a priori* utilisée de façon non bayésienne) que les estimations seraient inacceptables si elles dépassaient certaines limites, qu'il faudrait alors revoir le modèle ou la théorie du modèle.

c) L'ensemble des lois de Bayes qu'un utilisateur peut admettre est limité : on est près à dire qu'*a priori*, la variance de tel paramètre ne dépasse pas telle valeur.

Les deux derniers cas consistent à prendre en compte une information *a priori* de façon moins brutale que par les méthodes bayésiennes habituelles.

## 4 EXEMPLES ET COMPLEMENTS

## 4.1 Deux exemples simples

*Premier exemple* : Supposons que  $m$  soit astreint à vérifier  $Q(m) \leq 1$  où  $Q$  est une forme quadratique (éventuellement généralisée ou dégénérée). On a alors un estimateur optimal unique correspondant à une loi *a priori* de covariance  $\Gamma = Q^{-1}$ . L'estimateur optimal s'obtient alors à partir de l'estimation de Gauss-Markov  $P_0 y$  par la transformation :

$$\hat{m} = P_0 y \mapsto \hat{m} = (1 + k^2 Q)^{-1} \hat{m}.$$

*Deuxième exemple* : Supposons que le domaine de  $m$  soit de la forme  $|\langle u_\alpha, m \rangle| \leq 1$  où les  $u_\alpha$  sont  $q$  formes linéaires indépendantes. Quitte à se restreindre à un sous-espace,  $m$  est donc enfermé dans un "parallélogramme", et les  $u_\alpha$  forment une base du dual de l'espace de  $m$ . Soit  $m_\alpha$  la base duale, vérifiant :

$$\begin{aligned} \langle u_\alpha, m_\beta \rangle &= 0 & \text{si } \alpha \neq \beta \\ &= 1 & \text{si } \alpha = \beta. \end{aligned}$$

Les points extrêmes du domaine seront donc les  $2^q$  points de la forme :

$$x = \sum_{\alpha} \epsilon_{\alpha} m_{\alpha} \text{ où } \epsilon_{\alpha} = \pm 1.$$

Par symétrie il suffira de s'intéresser aux  $2^{q-1}$  points de la forme :

$$x_r = m_1 + \sum_{\alpha \neq 1} \epsilon_{\alpha} m_{\alpha}.$$

(Pour préciser le paramétrage on pourra dire que  $\alpha \neq 1$  varie de 2 à  $q$  et que :

$$r = 1 + \sum_{\alpha=2}^q \frac{1}{2} (\epsilon_{\alpha} + 1) 2^{\alpha-2}.$$

Les formes quadratiques "optimales" à trouver vérifieront donc  $Q(x_r) = 1$  pour  $q$  points,  $Q(x_r) \leq 1$  pour les autres, et seront de la forme :

$$Q = \sum_{\alpha \beta=1}^q \alpha_{\alpha, \beta} u_{\alpha} u_{\beta}$$

où les  $\alpha_{\alpha\beta}$  forment une matrice symétrique définie positive (comme on dit). Quitte

à changer la numérotation des axes, on peut supposer que les  $q$  points correspondant à une égalité sont :

$$x_1 = \sum_{\beta=1}^q m_\beta \quad \text{et} \quad x_r = \left( \sum_{\beta \neq r} m_\beta \right) - m_r \quad \text{pour} \quad r = 2, \dots, q.$$

On obtient alors les relations sur les  $\alpha_{\alpha\beta}$  en écrivant  $Q(x_1) = 1$  pour  $r = 1, \dots, q$  :

$$\sum_{\alpha, \beta=1}^q \alpha_{\alpha\beta} = 1$$

et pour  $r = 2, \dots, q$

$$a_{r,r} + \sum_{\substack{\alpha \neq r \\ \beta \neq r}} a_{\alpha\beta} - 2 \sum_{\beta \neq r} a_{\beta,r} = 1,$$

soit :

$$\sum_{\alpha, \beta} a_{\alpha\beta} = 1 \quad \text{et} \quad \sum_{\beta \neq r} a_{\beta,r} = 0 \quad \text{pour} \quad r = 2, \dots, q. \quad (20-a)$$

Une famille des solutions particulières, correspondant à un cas limite, est obtenue par l'ensemble de toutes les formes quadratiques valant 1 en chacun des points  $x_r$  ( $r = 1, 2^{q-1}$ ). Cette famille admet évidemment la forme générale :

$$Q(x) = \sum_{\alpha=1}^q c_\alpha \langle u_\alpha, x \rangle^2 \quad \text{avec} \quad c_\alpha \geq 0 \quad \text{et} \quad \sum_{\alpha} c_\alpha = 1. \quad (20-b)$$

Nous allons traiter le problème de façon complète dans le cas où  $q = 1, 2, 3$  et esquisser la solution dans le cas  $q = 4$ .

• Si  $q = 1$ , on est en fait dans le cas de l'exemple 1 et  $u_\alpha^2$  fournit la solution unique.

• Si  $q = 2$ , les relations (20) conduisent à  $a_{12} = a_{21} = 0$  et  $a_{11} + a_{22} = 1$  ( $a_{ii} \geq 0$ ). Les "ellipsoïdes" optimaux sont les ellipses passant par les 4 sommets du parallélogramme.

Les solutions trouvées sont uniquement du type de la solution particulière donnée en (20-b), mais ceci n'est vrai que pour la dimension 2. En effet :

• Si  $q = 3$  on trouve les relations :  $a_{12} + a_{23} = 0$  et  $a_{13} + a_{23} = 0$  ; posant  $t = a_{12}$  on obtient la famille de matrices de la forme :

$$\begin{bmatrix} a & t & t \\ t & b & -t \\ t & -t & c \end{bmatrix}$$

On sait que  $a, b, c, \geq 0$  et on écrit que  $Q(1, -1, -1) \leq 1$  ce qui donne  $a + b + c - 2t \leq 1$ . Mais comme  $a + b + c + 2t = 1$  on obtient nécessairement que  $t$  est positif. Enfin l'écriture de toutes les relations permettant d'affirmer que la matrice est définie positive donne que :

$$t^2 \leq \frac{abc}{a + b + c} \quad \text{et} \quad t^2 \leq bc + ca + ab.$$

• Si  $p = 4$  on obtient facilement un paramétrage du type :

$$\begin{bmatrix} a & u + v & u + w & w + v \\ & b & u & -v \\ & & c & -w \\ & & & d \end{bmatrix} \quad a + b + c + d + 2(u + v + w) = 1$$

avec  $u, v, w$  positifs. Par contre, les conditions pour que cette matrice soit définie positive sont assez inextricables.

## 4.2 Un exemple tiré d'une étude économétrique

Dans ses remarquables travaux sur le capital, J. MAIRESSE utilise une fonction de production de Cobb-Douglas. A partir de données individuelles d'entreprises tirées du recensement industriel de 1970, comportant la valeur ajoutée  $V_i$ , le capital  $K_i$  et le nombre de salariés  $L_i$ , il écrit le modèle :

$$V_i = k L_i^\alpha K_i^\beta e^{\epsilon_i} ; \mathbf{E} \epsilon_i = 0, \mathbf{E} \epsilon_i \epsilon_j = \sigma^2 \delta_{ij} .$$

Pour éviter les colinéarités on réécrit le modèle :

$$\frac{V_i}{L_i} = k \left( \frac{K_i}{L_i} \right)^\beta L_i^{\alpha + \beta - 1} e^{\epsilon_i} .$$

(On utilise les variables production et capital par tête)

Les contraintes sont  $\alpha \geq 0$ ,  $\beta \geq 0$  et, dans l'hypothèse d'une fonction de production à rendements décroissants,  $\alpha + \beta \leq 1$ .

Avec  $\gamma = \text{Log } k$ ,  $y = \text{Log } \frac{V}{L}$ ,  $x_1 = \text{Log } L$ ,  $x_2 = \text{Log } \frac{K}{L}$  et  $\mu = \alpha + \beta - 1$ ,

on obtient en passant aux logarithmes la régression :

$$y = \gamma + \mu x_1 + \beta x_2 + \epsilon$$

$$\beta \geq 0, \mu \leq 0, \mu \geq \beta - 1$$

Le domaine ainsi déterminé dans l'espace  $(\mu, \beta, \gamma)$  est un cylindre à base triangulaire.

L'estimation des moindres carrés (Gauss-Markov) donne les résultats suivants :

$$\hat{\beta} = 0,195 \quad \hat{\mu} = -0,013 \quad \hat{\gamma} = 2,286 \quad \hat{\sigma}^2 = 0,419.$$

Pour faciliter les calculs, on se place dans la base déterminée par le vecteur  $(1, \dots, 1)$ ,  $(x_1 - \bar{x}_1)$ ,  $(x_2 - \bar{x}_2)$ . Les deux derniers axes sont donc orthogonaux au premier, de sorte que la matrice métrique vaut :

$$\begin{bmatrix} 0,865 & 0,218 & 0 \\ & 1,040 & 0 \\ & & 1 \end{bmatrix}.$$

Le domaine symétrisé convexe est alors le cylindre dont la base dans le plan  $(\mu, \beta)$  est le carré  $(1,0)$ ,  $(0,1)$ ,  $(-1,0)$ ,  $(0,1)$ . Les formes quadratiques minimales sont alors, d'après (20-b), de la forme :

$$a \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} [1,1,0] + b \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} [1,-1,0], \quad \begin{array}{l} a + b = 1 \\ a \geq 0, b \geq 0 \end{array}$$

soit encore

$$\begin{bmatrix} 1 & \theta & 0 \\ \theta & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (\theta = a - b, \quad \text{soit } |\theta| \leq 1).$$

L'estimation de  $\sigma^2$  étant de 0,419 on va admettre que  $\sigma^2$  ne peut pas être inférieur à  $k^2 = 0,1$  ou 0,2. En appliquant la formule (19) on est conduit aux estimations "améliorées" suivantes :

$\theta$	$k^2 = 0,1$			$k^2 = 0,2$		
	$\hat{\beta}$	$\hat{\mu}$	$\hat{\gamma}$	$\hat{\beta}$	$\hat{\mu}$	$\hat{\gamma}$
- 1	0,171	+ 0,008	2,253	0,155	+ 0,021	2,236
- 0,5	0,172	0,000	2,285	0,155	+ 0,009	2,287
0	0,174	- 0,008	2,315	0,157	- 0,004	2,339
0,5	0,177	- 0,015	2,338	0,161	- 0,018	2,390
1	0,180	- 0,025	2,374	0,169	- 0,034	2,440

La contrainte de négativité de  $\mu$ , correspondant à l'hypothèse de rendement décroissant, n'est pas toujours satisfaite. Ceci n'est pas étonnant ; en effet, à cause de la symétrisation du domaine, on a en fait utilisé des contraintes (cf. Fig. 3) :

$$|\mu| \leq 1 - |\beta|.$$

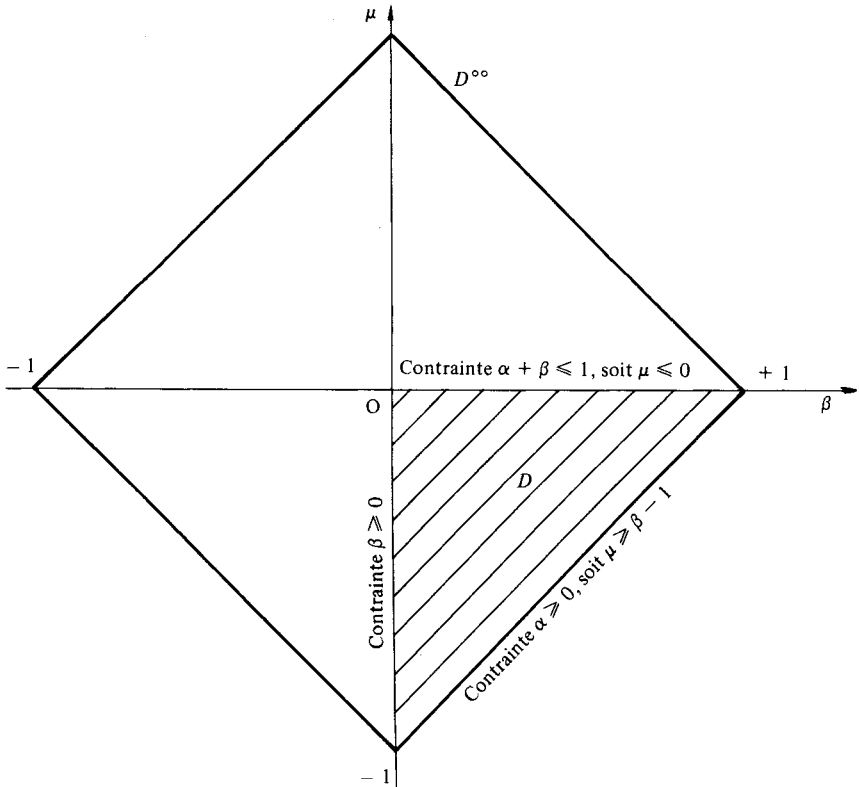


Figure 3 : Domaine des paramètres  $\beta$  et  $\mu$  dans l'estimation de la fonction de Cobb-Douglas.

De ce fait,  $\mu$  étant proche de zéro, on voit apparaître l'instabilité de l'estimation des moindres carrés dès que l'écart-type du modèle est un peu grand.

### 4.3 Vers une théorie de l'estimation affine

#### 4.3.1 Généralités

Dans l'exemple précédent, on a vu le domaine  $D^{\circ\circ}$  devenir sensiblement plus gros que  $D$  (cf. figure 3). Il n'en aurait pas été de même si l'origine avait été placée ailleurs, au centre de gravité du triangle par exemple. On aurait alors obtenu une forme quadratique  $Q$  plus grande et donc une estimation plus raccourcie, mieux stabilisée, des paramètres. On est donc naturellement conduit à se poser le problème de l'estimation affine :

$$Ay = a + Ty, a \in L, T \text{ linéaire vérifiant } \dim T(L) = \operatorname{rg} T. \quad (21)$$

En utilisant toujours la factorisation  $T = SP$  du paragraphe 1.2.2, on voit facilement que le risque (forme quadratique) de l'estimateur  $A$  prend la forme :

$$R_A(m) = \mathbf{E}(Ay - m)^2 = \sigma^2 SPP^{\star\star} S^{\star\star} + (S - 1)(m - a)^2 (S^{\star\star} - 1).$$

Il est clair qu'on peut réduire le problème aux estimateurs  $A = a + SP_0$  où  $P_0$  est le projecteur de Gauss-Markov et que le risque est alors :

$$R_{a,s}(m) = \sigma^2 SS^{\star\star} + (S - 1)(m - a)^2 (S^{\star\star} - 1).$$

On a alors la **proposition** suivante :

*Si  $m \in D$ ,  $\sigma^2 \geq k^2$ , les estimateurs "optimaux" qui minimisent  $\frac{1}{\sigma^2} R_{a,s}(m)$*

*sont obtenus nécessairement :*

1. *Pour les points  $a$  rendant minimum  $(D - a)^{\circ\circ}$  pour la relation d'ordre d'inclusion des ensembles de  $L$ .*

2.  *$a$  étant ainsi choisi, en prenant  $S$  optimal au sens du paragraphe 2.4.*

La démonstration est très facile : si  $D_a = (D - a)^{\circ\circ} \subset D_{a'}$ , tout estimateur optimal pour l'origine placé en  $a'$  peut être amélioré en un estimateur optimal en  $a$  de risque inférieur. Si l'inclusion est stricte, il existe des estimateurs optimaux en  $a'$  qui peuvent être strictement améliorés.

On remarquera que les  $a$  "optimaux" correspondent aux estimateurs constants de risque minimum !



Inversement, si  $a$  est optimal, certains estimateurs centrés en  $a$  peuvent être améliorés. Autrement dit, un estimateur affine optimal n'est pas n'importe quel couple ( $a$  constante optimale,  $S$  linéaire optimal l'origine étant en  $a$ ). Supposons en effet que  $S$  soit linéaire optimal en  $a$  ; soient  $(x_i)$  les points de contact de l'ellipse  $\Sigma$  associé à  $S$  avec  $D_a$ . S'il existe  $a'$  tel que  $D_{a'}$  soit contenu strictement dans  $\Sigma$  (et non pas  $D_a$  !), le couple  $(a, S)$  est améliorable.

On verra un exemple de ce phénomène au sous-paragraphe 4.3.3.

#### 4.3.2 Estimateur affine d'une proportion

On désire estimer  $m \in [0,1]$ , sachant que l'estimation de Gauss-Markov tombe en  $x \in \mathbf{R}$  (dans un modèle où il n'y a qu'un seul paramètre borné).

Le modèle est donc pour nous :

$x$  v.a.r. de moyenne  $m \in [0,1]$ , d'écart-type  $\sigma \geq k$ .

Planter l'origine en  $a \in \mathbf{R}$ , revient à estimer  $m$  par :

$$\hat{x} = a + s(x - a).$$

Le risque

$$\mathbf{E}(\hat{x} - m)^2 = \mathbf{E}(a + s(x - a) - m)^2 = \mathbf{E}[s(x - m) + (s - 1)(m - a)]^2$$

vaut :

$$R_{a,s}(m) = \sigma^2 s^2 + (s - 1)^2 (m - a)^2.$$

$$\text{D'où : } \max_m \frac{1}{\sigma^2} R = s^2 + (s - 1)^2 \frac{a^2}{k^2}, \text{ si } a \geq \frac{1}{2}$$

$$= s^2 + (s - 1)^2 \frac{(1 - a)^2}{k^2}, \text{ si } a \leq \frac{1}{2}.$$

$$\text{Pour } a \geq \frac{1}{2}, \text{ on a } s_a = \frac{a^2}{a^2 + k^2}$$

$$\text{et } \hat{x} = \frac{a(k^2 + ax)}{a^2 + k^2}.$$

La figure 3 montre le réseau des courbes  $\hat{x}(a)$  indicées par l'estimateur de G.M.  $x$ , pour  $k^2 = 0,3$  de façon à accentuer un peu le mouvement des courbes. Pour les construire, on a remarqué que :

- Le réseau est symétrique par rapport au point  $(\frac{1}{2}, \frac{1}{2})$ .
- Le lieu des extremums de l'ensemble des courbes est la droite  $\hat{x} = \frac{1}{2}a$ , atteint pour  $a = x \pm \sqrt{x^2 + k^2}$  si  $a \geq \frac{1}{2}$ .
- Chaque courbe a une asymptote  $\hat{x} = x$ .
- On a le tableau de valeurs suivant :

$x \backslash a$	- 1	- 0,5	0	0,25	0,5	0,75	1	1,5	2
-0,60	-0,628	-0,588	-0,462	-0,304	0,000	-0,130	-0,231	-0,353	-0,419
-0,30	-0,349	-0,323	-0,231	-0,109	0,136	0,065	0,000	-0,088	-0,140
-0,05	-0,116	-0,103	-0,038	0,054	0,250	0,228	0,192	0,132	0,093
0	-0,070	-0,059	0,000	0,087	0,273	0,261	0,231	0,176	0,140
0,35	0,256	0,250	0,269	0,315	0,432	0,489	0,500	0,485	0,465
0,50	0,395	0,382	0,385	0,413	0,500	0,587	0,615	0,618	0,605
0,65	0,535	0,515	0,500	0,515	0,568	0,685	0,731	0,750	0,744
1,00	0,860	0,824	0,769	0,739	0,727	0,913	1,000	1,059	1,070
1,05	0,907	0,868	0,808	0,772	0,750	0,946	1,038	1,103	1,116
1,3	1,140	1,088	1,000	0,935	0,864	1,109	1,231	1,323	1,349
1,6	1,419	1,353	1,231	1,130	1,000	1,304	1,462	1,588	1,628

#### 4.3.3 Recherche d'une origine optimale

On ne peut guère donner d'idées très générales. Il est clair par exemple, qu'à deux dimensions, il faudra toujours prendre l'origine dans la fermeture convexe de  $D$  (sinon on améliorera les choses en s'en "rapprochant"). Or, ce principe n'est même pas toujours vrai à trois dimensions. Le cas où  $D$  est un tétraèdre (fig. 9) le montre.

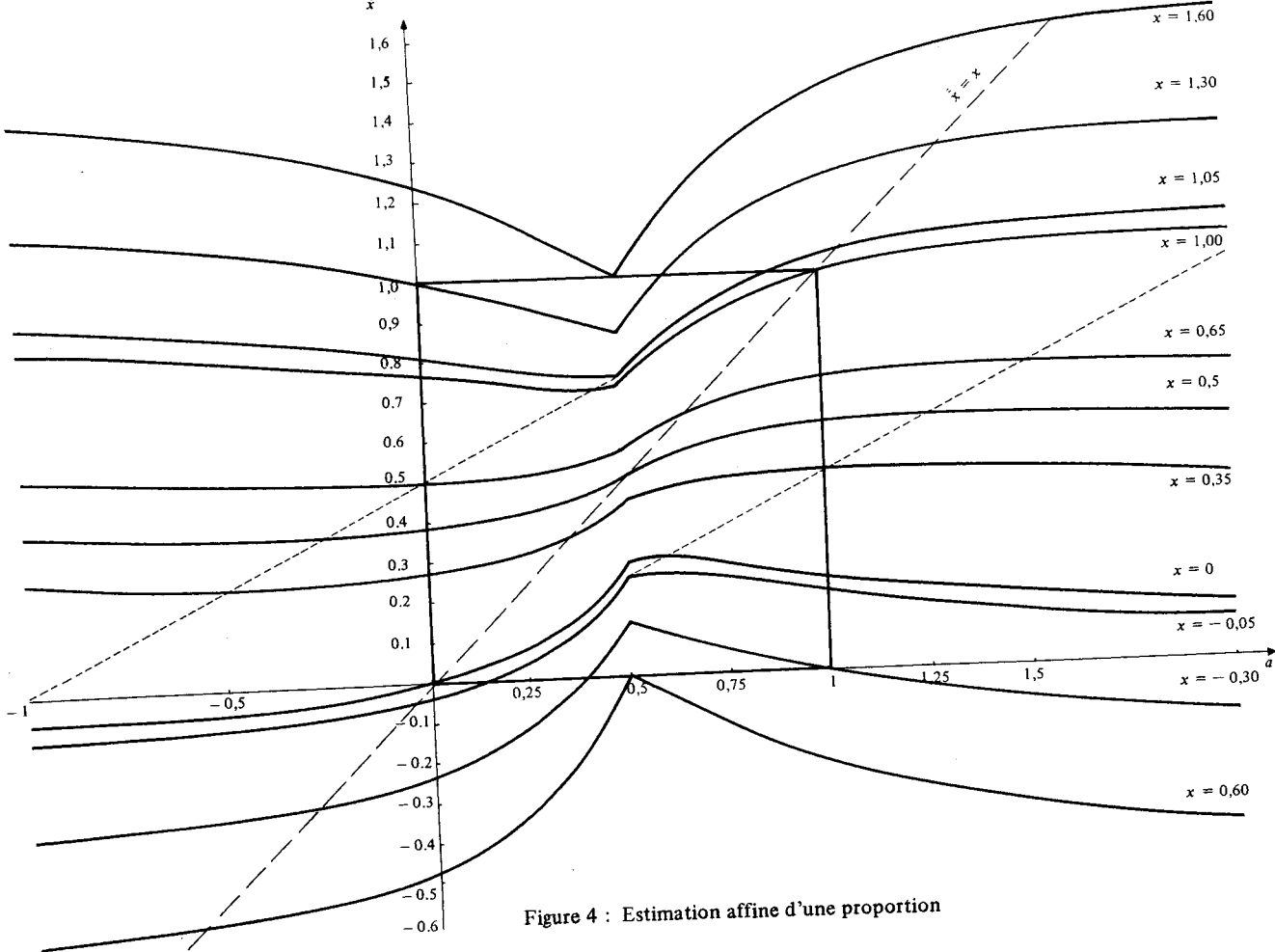


Figure 4 : Estimation affine d'une proportion

En revanche, si le convexe de  $D$  possède un centre de symétrie, ce point est évidemment une origine optimale, et les seules origines optimales possibles sont les centres de symétrie. Si  $D$  est complètement borné et symétrique, il n'y a d'ailleurs qu'un seul centre et donc une seule origine optimale. En utilisant le résultat du 2.4.3 on obtient donc l'existence d'un estimateur affine optimal unique dans le cas, et uniquement dans le cas, où la fermeture convexe de  $D$  est un ellipsoïde.

De façon générale, la recherche des  $a$  rendant  $D_a$  minimal semble difficile. On peut toutefois résoudre complètement le problème dans le cas de l'exemple développé au paragraphe 4.2, c'est-à-dire, en fait, dans le cas où  $D$  est un triangle de  $\mathbb{R}^2$ . Soient  $O'$ ,  $B'$ ,  $C'$  les milieux des segments  $BC$ ,  $CO$  et  $OC$  respectivement. Prenons  $a$  à l'intérieur du triangle  $OB'C'$  par exemple (figure 5).

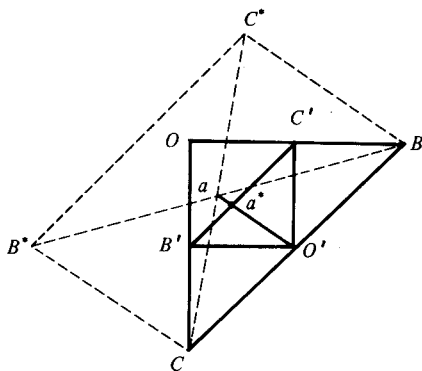


Figure 5

Le domaine  $D_a$  est (à une translation près), le parallélogramme  $BCB^{\star}C^{\star}$ ; comme  $aO'$  est parallèle à  $B^{\star}C$  et  $C^{\star}B$ ,  $D_a$  va "rétrécir" si  $a$  se rapproche de  $O'$  et cela jusqu'au moment où  $a$  va se trouver en  $a^{\star}$ , intersection de  $aO'$  et  $BC'$ . A ce moment, en effet, le point  $O$  va "traverser"  $BC$ , et  $D_a$  deviendra un hexagone. Les origines optimales se trouvent donc nécessairement dans le triangle  $O'B'C'$ . Inversement on se rend compte facilement que tout point de ce triangle est une origine optimale.

On arrive ainsi à 3 formes extrêmes qui sont des parallélogrammes et à une forme intérieure, hexagonale (fig. 6).

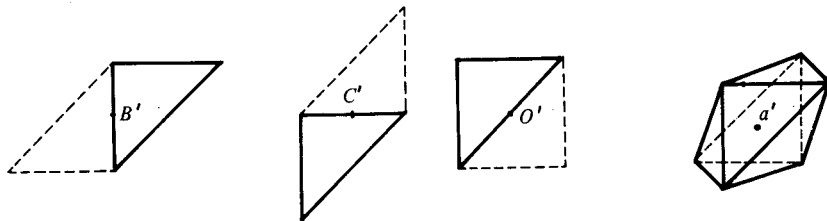


Figure 6

Donnons maintenant l'exemple d'un couple ( $a = \text{optimal}$ ,  $S = \text{optimal}$ ) qui ne donne pas une estimation affine optimale. Dans l'exemple précédent, plantons l'origine en  $a = (1/3, -1/3)$  (fig. 7).

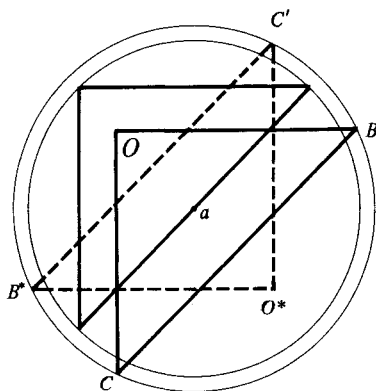


Figure 7

Les points  $BCB^*C^*$  sont sur un même cercle de rayon  $\sqrt{5}/3 = 0,745$ . Si l'origine est placée en  $O'$ , le domaine est un carré de côté unité. Le cercle de rayon  $\sqrt{2}/2 = 0,708$  fournit une estimation meilleure que la précédente. Par contre, dans le premier cas, l'ellipse passant par les 6 points  $C^*BO^*CB^*O$  fournira un estimateur non améliorable.

Dans des cas plus compliqués, la recherche d'estimateurs affines optimaux s'avère difficile. Après les triangles de  $\mathbb{R}^2$ , les domaines les plus simples sont les quadrilatères de  $\mathbb{R}^2$ , correspondant donc à 4 contraintes d'inégalités linéaires sur 2 paramètres. La figure 8 représente le cas général (c'est-à-dire sans particularités de parallélisme des côtés). Les origines optimales sont à l'intérieur du quadrilatère (non convexe !)  $acbd$ . Pour obtenir ce résultat on procède de la façon suivante : le quadrilatère  $bcBd$  (par exemple) correspond à l'ensemble des points  $\alpha$  tels que  $D_\alpha$  soit le symétrisé convexifié (par rapport à  $\alpha$ ) du triangle  $ACD$ . On peut, à l'intérieur de ce quadrilatère, améliorer  $D_\alpha$  en déplaçant  $\alpha$  dans la direction de  $a$ , milieu de  $AC$ . On fait le même raisonnement pour les zones :  $Cdac$ , le triangle est  $ABD$ , on améliore vers  $b$ , et  $Acaf$ , le triangle est  $BCD$ , on améliore vers  $b$ .

La zone  $Dfbe$  correspond au cas où  $D_\alpha$  est déterminé par le triangle  $ABC$ . Tant que  $\alpha$  n'est pas dans le triangle  $adc$  (lieu des optimaux pour le triangle  $ABC$ ) on peut toujours améliorer en allant soit vers  $a$ , soit vers  $d$ , soit vers  $c$  selon le cas.

Les origines optimales sont donc nécessairement dans la zone  $acbd$  ; un raisonnement variationnel direct montre que tous les points de cette zone fournissent des  $D_\alpha$  minimaux.

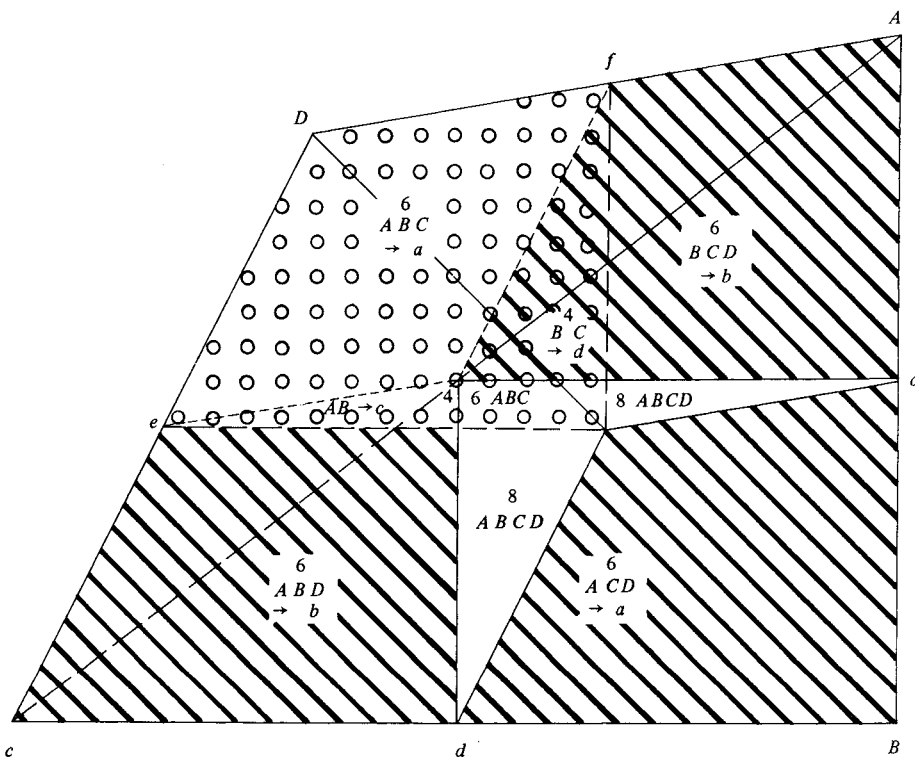


Figure 8

On a porté sur la figure, pour chaque zone, le nombre de côtés du polygone  $D_\alpha$  ainsi que les sommets du quadrilatère sur lesquels ils sont construits. On donne enfin, s'il y a lieu, la direction dans laquelle on peut améliorer  $D_\alpha$ .

La résolution de ce cas est donc relativement compliquée. Le cas d'un pentagone, ou d'un tétraèdre de  $\mathbf{R}^3$ , qui même à l'examen de polyèdres du type de celui

de la figure 9, sont d'une complexité suffisante pour qu'on n'insiste pas trop sur la recherche d'une origine optimale.

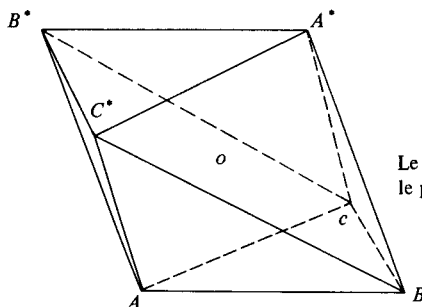


Figure 9  
Le modèle tridimensionnel  
le plus simple !

Dans la pratique, il faudra choisir une origine à l'intérieur du domaine, plutôt "au milieu" et appliquer l'algorithme d'amélioration du paragraphe 3.1.

#### 4.4 Application à l'estimation de la fonction de Cobb-Douglas

Reprenons l'exemple du numéro 4.2 et plaçons l'origine en  $O' = (0,5; -0,5)$  (figures 5 et 11). Les formes quadratiques optimales sont maintenant, du fait que le domaine est un carré de côté un, données par les matrices :

$$Q = \begin{bmatrix} 4\delta & 0 & 0 \\ 0 & 4(1 - \delta) & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{avec } 0 \leq \delta \leq 1.$$

On obtient les estimations suivantes :

$k^2 = 0,1$			$k^2 = 0,2$	
$\delta$	$\hat{\beta}$	$\hat{\mu}$	$\hat{\beta}$	$\hat{\mu}$
0	0,230	-0,154	0,250	-0,231
0,25	0,253	-0,131	0,291	-0,204
0,5	0,271	-0,105	0,318	-0,167
0,75	0,284	-0,073	0,335	-0,116
1	0,295	-0,034	0,345	-0,045

Les variations sont ici plus importantes qu'au paragraphe 4.2.

Cela tient à deux raisons :

– le point  $(\hat{\beta}, \hat{\mu})$  est maintenant plus loin de l'origine  $O'$  qu'il n'était de  $O$ . Une même "réduction" des longueurs va donc se traduire par des variations plus importantes de l'estimation.

– l'origine  $O'$  étant optimale, on utilise des réductions plus importantes que celles autorisées par l'origine "naturelle".

Dans ce cas comme dans celui de l'estimation d'une proportion, il apparaît qu'une origine optimale fournit des "réductions" nettement plus forte qu'une origine non optimale.

Consulté sur ces résultats, J. MAIRESSE considère comme inadmissibles les estimations obtenues en ce qui concerne le paramètre  $\mu$ . Sa compétence d'économiste l'amène à formuler sur  $\mu$  une restriction *a priori* : le paramètre doit être compris entre 0,9 et 1,1. Autrement dit, le domaine de variation du couple  $(\beta, \mu)$  est maintenant le quadrilatère :  $\beta \geq 0, |\mu| \leq 0,1, \mu \geq \beta - 1$  (fig. 10).

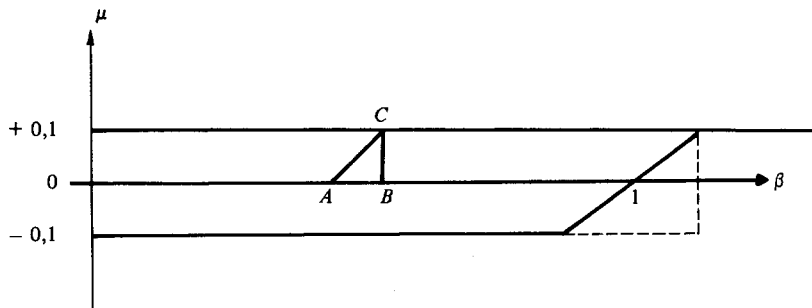


Figure 10

Ce quadrilatère est "le" cas particulier de l'analyse du paragraphe 4.3. Par une méthode identique, on montre que les origines optimales sont à l'intérieur du triangle  $A = (0,5 ; 0)$ ,  $B = (0,55 ; 0)$ ,  $C = (0,55 ; 0,1)$ . Plaçant l'origine en  $B$ , le domaine  $D^{\circ\circ}$  est un rectangle dont les côtés valent 1,1 et 0,2. Les formes quadratiques optimales sont alors de la forme :

$$Q = \begin{bmatrix} \varphi \cdot (0,55)^{-2} & 0 & 0 \\ 0 & (1 - \varphi) (0,1)^{-2} & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 3,306 \varphi & 0 & 0 \\ 0 & 100(1 - \varphi) & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad 0 \leq \varphi \leq 1$$



On obtient alors les estimations suivantes (la figure 11 montre les variations entre les diverses estimations des paragraphes 4.2 et 4.4) :

$\varphi$	$k^2 = 0,1$		$k^2 = 0,2$	
	$\hat{\beta}$	$\hat{\mu}$	$\hat{\beta}$	$\hat{\mu}$
0	0,192	-0,001	0,192	-0,001
0,25	0,224	-0,002	0,249	-0,002
0,5	0,250	-0,004	0,291	-0,003
0,75	0,273	-0,009	0,323	-0,007
0,85	0,282	-0,013	0,334	-0,011
1	0,297	-0,034	0,353	-0,046

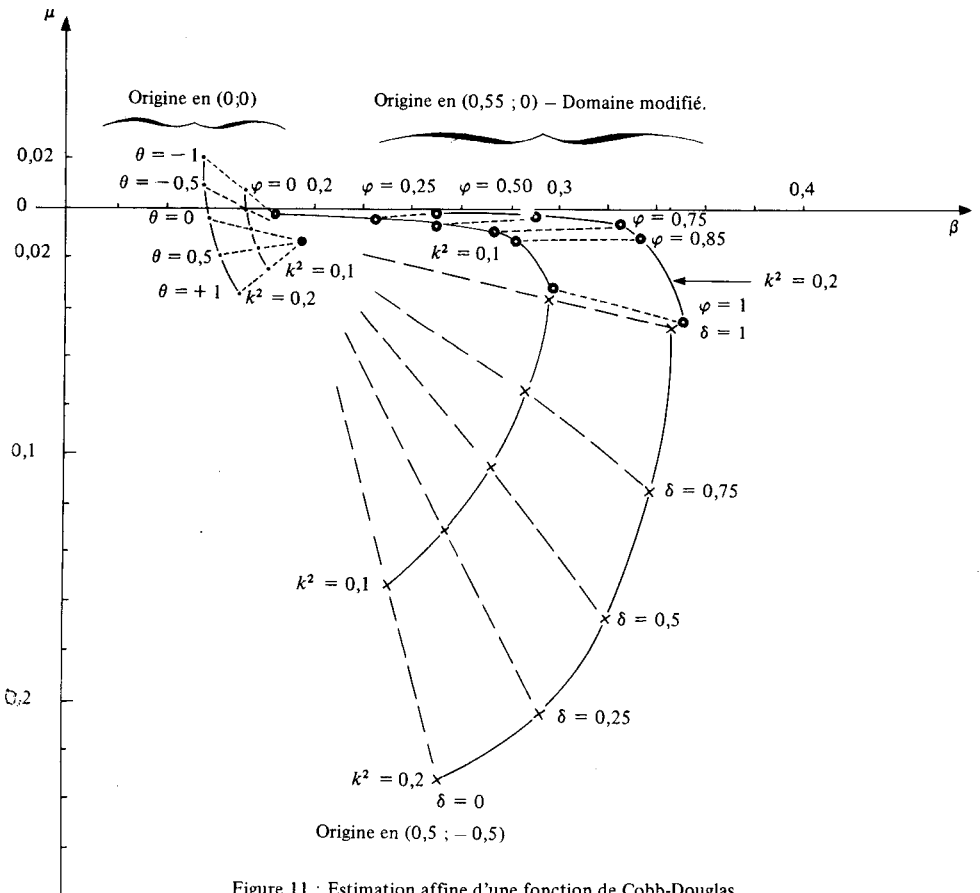


Figure 11 : Estimation affine d'une fonction de Cobb-Douglas.

#### 4.5 Remarques et conclusions

Il peut paraître curieux de vouloir estimer linéairement un paramètre borné. L'estimation risque, en effet, de tomber en dehors du domaine, ce qui la rend pratiquement à peu près inutilisable.

Néanmoins, on peut remarquer qu'il s'agit là pourtant d'une pratique courante. Nous avons simplement essayé d'indiquer un procédé qui permet d'éviter cette difficulté plus souvent qu'avec les estimateurs linéaires sans biais. Un contexte linéaire permet par ailleurs un traitement analytique satisfaisant : on connaît à peu près bien les propriétés des estimateurs qui sont ici proposés.

Les estimateurs concurrents ont aussi bien des défauts : des estimateurs à distance minimale, par exemple, envoient systématiquement à la frontière du domaine dans le cas où l'estimation sans biais tombe en dehors. On ne voit pas très bien le surcroît de confort que procure au statisticien une telle situation ! Remarquons enfin qu'aucun des estimateurs raccourcis non linéaires, du type James-Stein par exemple, n'évitent ce défaut. Enfin, qu'ils soient de James-Stein ou à distance minimum, ces estimateurs ne peuvent guère trouver de justification que dans le cadre d'hypothèses de normalité des lois de probabilité intervenant dans le modèle. De telles hypothèses sont, très souvent, parfaitement injustifiables, surtout dans la pratique économétrique.

#### BIBLIOGRAPHIE

- [1] F. CAILLIEZ et J.P. PAGES. — Introduction à l'analyse des données — *SMASH* (1976).
- [2] P. CAZES. — Estimation linéaire et estimation sous contraintes dans le modèle linéaire. *Journées internationales sur l'Analyse des données et l'Informatique* — IRIA (1977).
- [3] A. DEMPSTER. — Alternatives to least squares in multiple regression dans *Multivariate Statistical Inference*. KABE et GUPTA editors — North-Holland — Elsevier (1973).
- [4] A. DEMPSTER, M. SCHATZOFF, N. WERMUTH. — A simulation study of alternatives to least squares. *JASA* vol. 72 pp. 77-90 (1977).
- [5] I. GLATZMAN et Y. LIOUBITCH. — Analyse linéaire dans les espaces de dimension finie. Editions de Moscou (1976).

- [6] R.F. GUNST, R.L. MASON. — Biassed estimation in regression : an evaluation using mean square error. *JASA* vol. 72 pp. 616-628 (1977).
- [7] P.R. HALMOS. — Finite dimensional vector spaces. Van-Nostrand (1954).
- [8] W. JAMES et C. STEIN. — Estimation with quadratic loss. *Pr. of the fourth Berkeley Symposium*. Vol. 1 pp. 361-379 (1961).
- [9] W. KRUSKAL. — The coordinate free approach to Gauss-Markov estimation and its applications to missing and extra-observations. *Pr. of the fourth Berkeley Symposium* — Vol. 1, pp. 435-451 (1961).
- [10] J. MAIRESSE. — Comparison of production functions estimates on the french and norvegian censuses of manufacturing industries, dans "On the measurement of factor productivities". Van der Hoech et Ruprecht — Göttingen (1976).
- [11] E. MALINVAUD. — Méthodes statistiques de l'économétrie. Dunod (1969).
- [12] P. MAZODIER. — Exposé sur la "ridge-regression". *Séminaire de l'ENSAE* — janvier 1977.
- [13] J.L. PHILOCHE. — A propos du théorème de Gauss-Markov. *Annales de l'I.H.P.*, pp. 271-281., vol. 4. (1971).
- [14] J.L. PHILOCHE, PH. COURREGÉ et P. PRIOURET. — Régression linéaire et estimation par la méthode des moindres carrés. *Annales de l'I.H.P.*, pp. 253-270 vol. 4. (1971).
- [15] C.R. RAO. — Estimation of parameters in a linear model. *Annals of statistics* pp. 1023-1037 (1976).
- [16] A. ROBINSON. — Non standart Analysis. North-Holland (1970).
- [17] L. SCHWARTZ. — Cours d'analyse de l'Ecole Polytechnique : chapitre VI : calcul différentiel extérieur, § 6 (1964).