

REVUE DE STATISTIQUE APPLIQUÉE

P. CAZES

Protection de la régression par utilisation de contraintes linéaires et non linéaires

Revue de statistique appliquée, tome 23, n° 3 (1975), p. 37-57

<http://www.numdam.org/item?id=RSA_1975__23_3_37_0>

© Société française de statistique, 1975, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) 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/>

PROTECTION DE LA RÉGRESSION PAR UTILISATION DE CONTRAINTES LINÉAIRES ET NON LINÉAIRES ⁽¹⁾

P: CAZES

Laboratoire de statistique mathématique, Université Paris VI

I – INTRODUCTION

En régression on cherche à expliquer une variable (le rendement d'une culture par exemple) en fonction d'un certain nombre de variables explicatives (telles que certaines caractéristiques du sol dans le cas du rendement d'une culture).

Si le nombre des variables explicatives est grand par rapport au nombre d'observations, ou si ces variables explicatives sont très corrélées, la régression est de mauvaise qualité : coefficients de régression élevés, dépendant fortement de l'échantillonnage : on est dans le cas où pour expliquer une valeur y_i observée, on a une somme d'effets (grands en valeur absolue par rapport à y_i) de signes opposés et donc se retranchant ; la reconstitution de y_i est alors illusoire.

Pour protéger la régression, on peut :

- soit faire une régression pas à pas (i.e. introduire les variables une par une en fonction de leur importance),
- soit imposer des contraintes (i.e. utiliser le critère des moindres carrés sous certaines conditions).

Dans le cas de la régression sous contraintes qui est développée ici, trois types de contraintes nous sont apparues intéressantes :

1/ contraintes sur les coefficients de régression auxquels on impose, pour avoir un sens, d'être positifs (cas que l'on rencontre fréquemment en pratique), ou d'être compris entre deux valeurs données.

2/ contraintes sur la longueur du vecteur des coefficients de régression auquel on impose d'être plus petit qu'une valeur donnée (ridge régression).

3/ contraintes sur la variance résiduelle de la régression à laquelle on impose d'être supérieure à la variance d'erreur (cas de variables entachées d'erreurs)

Après avoir exposé ces trois méthodes, on compare les résultats qu'elles fournissent sur un exemple d'application où la solution usuelle des moindres carrés donne des résultats aberrants ; les résultats obtenus sur cet exemple par ces trois techniques sont comparables et excellents.

(1) Article remis en juillet 1974, révisé en Décembre 1974.

II-1. Formalisation mathématique et interprétation géométrique dans R^n .

Le problème de la régression sous contraintes peut se formaliser ainsi : étant donné une variable à expliquer y et p variables explicatives $x_1, \dots, x_p$ dont on possède un n échantillon $\{y^j, x_1^j, \dots, x_p^j \mid j = 1, n\}$, trouver le vecteur $\vec{b}$ de $E = (R^p)^*$ de composantes $b^i (1 \leq i \leq p)$ telles que :

$$\sum \{ (y^j - \sum \{b^i x_i^j \mid i = 1, p\})^2 \mid j = 1, n \}$$

soit minimum sous la condition : $\vec{b} \in C$ où C désigne la partie de l'espace $E = (R^p)^*$ où sont vérifiées les contraintes que l'on a imposées.

Dans les cas que nous examinerons, C sera :

– soit une partie de E délimitée par une série de sous espaces vectoriels (ou affines) (cas de contraintes linéaires), dans ce cas C est un convexe fermé.

– soit une partie de E délimitée par une surface du second degré (cas de contraintes quadratiques) ; dans ce dernier cas, C peut ne pas être un convexe fermé ; C peut être en particulier le complémentaire d'un convexe ouvert (extérieur d'un ellipsoïde), ou la réunion de deux convexes fermés (cas d'une surface ayant deux branches infinies, les contraintes imposées étant vérifiées à l'intérieur de chaque branche).

Si l'on désigne par $\vec{y}$ le vecteur de $F = R^n$ de composantes

$$\{y^j \mid j = 1, n\} \quad \text{et par} \quad \vec{x}_i (1 \leq i \leq p)$$

le vecteur de R^n de composantes $\{x_i^j \mid j = 1, n\}$, et si l'on pose :

$$X'_{(n,p)} = (\vec{x}_1, \vec{x}_2, \dots, \vec{x}_p) \quad (1)$$

(X' est l'application de $E = (R^p)^*$ dans $F = R^n$, qui à $\vec{b} \in E$ associe $\vec{z} = X' \vec{b}$ $\vec{b} = \sum b^i \vec{x}_i \in F$), on recherchera de façon plus générale, R^n étant muni de la métrique N la combinaison linéaire

$$\vec{z} = \sum \{b^i \vec{x}_i \mid i = 1, p\} = X' \vec{b} \quad (2)$$

telle que $\vec{y} - \vec{z}$ soit de norme minimale, i.e. telle que :

$$\|\vec{y} - \vec{z}\|_N^2 = N(\vec{y} - \vec{z}, \vec{y} - \vec{z}) = N(\vec{y} - X' \vec{b}, \vec{y} - X' \vec{b}) \quad (3)$$

soit minimum, sachant que $\vec{b} \in C$.

Ceci revient encore si l'on pose :

$$D = X'(C) = \{\vec{u} \in R^n : \vec{u} = X' \vec{b} ; \vec{b} \in C\} \subset F$$

à rechercher $\vec{z}$ tel que : $\|\vec{y} - \vec{z}\|_N^2$ soit minimum

avec

$$\vec{z} \in D.$$

$\vec{z}$ est donc la projection de $\vec{y}$ sur D, projection qui est unique si D est un convexe fermé, ce qui est réalisé si C est convexe fermé.

Si D n'est pas convexe, et si D est délimité par une surface du second degré, la projection de $\vec{y}$ sur D peut ne pas être unique (cf. fig. 1) ; ce cas qui se présente si $\vec{y}$ appartient à un axe de symétrie de S à une probabilité nulle d'arriver ; de plus dans les deux types de contraintes quadratiques que nous imposerons, la projection de $\vec{y}$ sur D est unique.

Nous supposons donc dorénavant que la projection de $\vec{y}$ sur D est unique, et nous désignerons par $\vec{t}^+ = X' \vec{b}^+$ cette projection.

Nous supposons de plus que les vecteurs $\vec{x}_1, \dots, \vec{x}_p$ sont linéairement indépendants, i.e. qu'ils forment une base du sous espace H de F qu'ils engendrent. A tout point

$$\vec{z} = X' \vec{b} = \sum \{b^i \vec{x}_i \mid i = 1, p\}$$

de H correspond donc un seul système de coordonnées $\vec{b} = \{b^i \mid i = 1, p\}$.

Nous désignerons par M la métrique induite par N sur H quand on choisit les $\{\vec{x}_i \mid i = 1, p\}$ comme base de H :

$$M = X N X' \quad (4)$$

II-2. Solution. Lien avec la solution des moindres carrés classique (cf. Fig. 2)

Soit $\vec{b}$ la solution classique des moindres carrés ; $\vec{t} = X' \vec{b}$ est la projection orthogonale (au sens de N) de $\vec{y}$ sur le sous espace H :

$$\vec{b} = (X N X')^{-1} X N \vec{y} = M^{-1} X N \vec{y} \quad (5)$$

On a alors (puisque D est contenu dans H) :

$$\forall \vec{u} \in D : \|\vec{y} - \vec{u}\|^2 = \|\vec{y} - \vec{t}\|^2 + \|\vec{t} - \vec{u}\|^2 \quad (6)$$

et minimiser $\|\vec{y} - \vec{u}\|^2$ quand $\vec{u}$ décrit D est équivalent à minimiser $\|\vec{t} - \vec{u}\|^2$, $\vec{u}$ appartenant à D.

Deux cas se présentent alors

. ou $\vec{b} \in C$, i.e. $\vec{t} \in D$

auquel cas la solution des moindres carrés vérifiant les contraintes convient : $\vec{t}^+ = \vec{t}$; $\vec{b}^+ = \vec{b}$

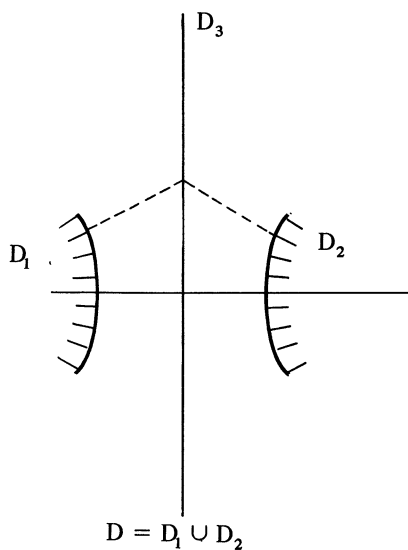
. ou $\vec{b} \notin C$; i.e. $\vec{t} \notin D$.

Dans ce cas la solution $\vec{b}^+$ des moindres carrés sous contraintes est telle que :

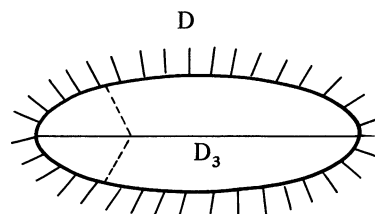
$$\vec{b}^+ \in \delta C$$

ou

$$\vec{t}^+ = X' \vec{b}^+ \in \delta D$$

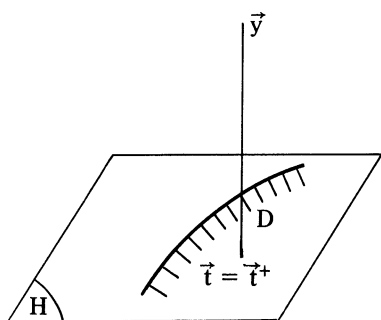


1^{er} cas de figure :
tous les points de D_3 sont à égale distance de D_1 et D_2 , et ont donc deux projections sur D .

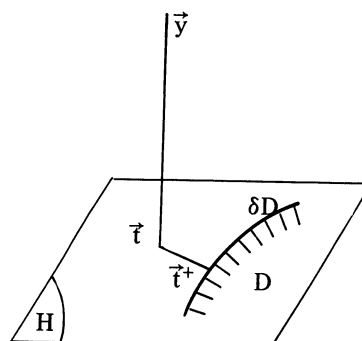


2^{ème} cas de figure :
tous les points du segment D_3 ont deux projections sur D

Figure 1



1^{er} cas de figure



2^{ème} cas de figure

$$R^2 = \|\vec{t}^+\|^2 / \|\vec{y}\|^2$$

$$S^2 = \|\vec{y} - \vec{t}^+\|^2 / \|\vec{y}\|^2$$

Figure 2

δC (resp. δD) désignant la frontière de C (resp. D) ; en effet, si ce n'était pas le cas, $\vec{t}^+$ appartiendrait à l'intérieur $\overset{\circ}{D}$ de D , et la droite joignant $\vec{t}$ à $\vec{t}^+$ couperait δD en un point $\vec{u}$ plus proche de $\vec{t}$ (et donc de $\vec{y}$, d'après (6)) que $\vec{t}^+$.

On projettera donc dans ce cas $\vec{t}$ sur δD .

II-3. Critère de qualité de la régression

La qualité de l'approximation de $\vec{y}$ par $\vec{t}^+$ sera mesurée par le rapport

$$R^2 = \|\vec{t}^+\|_N^2 / \|\vec{y}\|_N^2 \quad (1)$$

rapport qui n'est autre, si les variables $\vec{y}, \vec{x}_1, \dots, \vec{x}_p$ sont de moyenne nulle, et si $\vec{t}^+ = \vec{t} \in D$ (i.e. si la solution classique des moindres carrés vérifie les contraintes imposées), que le carré du coefficient de corrélation multiple de y par rapport à $x_1, \dots, x_p$.

Nous désignerons par résidu, le rapport

$$S^2 = \|\vec{y} - \vec{t}^+\|_N^2 / \|\vec{y}\|_N^2 \quad (1)$$

rapport qui est toujours plus petit ou égal à 1 si D (ou C) contient l'origine. Si $\vec{t} = \vec{t}^+$, de la relation

$$\begin{aligned} \|\vec{y}\|^2 &= \|\vec{y} - \vec{t}\|^2 + \|\vec{t}\|^2 \\ &= \|\vec{y} - \vec{t}^+\| \|\vec{t}^+\| \end{aligned} \quad (9)$$

$$\text{l'on tire :} \quad \underline{R^2 + S^2 = 1} . \quad (10)$$

De façon générale, de la relation :

$$\|\vec{y}\|^2 = \|\vec{y} - \vec{t}^+\|^2 + \|\vec{t}^+\|^2 + 2 \langle \vec{y} - \vec{t}^+, \vec{t}^+ \rangle \quad (11)$$

l'on tire

$$R^2 + S^2 = 1 - \frac{2 \langle \vec{y} - \vec{t}^+, \vec{t}^+ \rangle}{\|\vec{y}\|^2} \quad (12)$$

Dans le cas où D (ou C) est un convexe fermé contenant l'origine, on a

$$\langle \vec{y} - \vec{t}^+, \vec{t}^+ \rangle > 0$$

$$\text{et donc :} \quad \underline{R^2 + S^2 \leq 1} \quad (13)$$

II-4. Etude dans $E = (R^p)^*$.

Le vecteur $\vec{b}$ de E que l'on recherche doit minimiser

$$Q(\vec{b}) = N(\vec{y} - X'\vec{b}, \vec{y} - X'\vec{b}) = \|\vec{y} - X'\vec{b}\|_N^2 \quad (14)$$

$$\text{avec} \quad \vec{b} \in C \quad (15)$$

(1) Nous avons déjà introduit les quantités R^2 et S^2 dans le cas de contraintes linéaires (cf. (2)).

$\vec{b}$ désignant toujours la solution des moindres carrés, on a :

$$\begin{cases} Q(\vec{b}) = \|\vec{y} - X'\vec{b}\|_N^2 = \|\vec{y} - X'\vec{b}\|_N^2 + \|X'(\vec{b} - \vec{b})\|_N^2 \\ = Q_0 + \|\vec{b} - \vec{b}\|_{M=XX'}^2 \end{cases} \quad (16)$$

où $Q_0 = \|\vec{y} - X'\vec{b}\|^2$ est indépendant de $\vec{b}$.

Minimiser (14) sous la contrainte $\vec{b} \in C$ est donc équivalent à minimiser

$$\begin{cases} Q_1(\vec{b}) = \|\vec{b} - \vec{b}\|_{M=XX'}^2 \\ \text{avec } \vec{b} \in C \end{cases} \quad (17)$$

La solution $\vec{b}^+$ est donc, si l'on munit E de la métrique $M = XX'$ la projection de $\vec{b}$ sur C.

Si $\vec{b} \in C$, $\vec{b}^+ = \vec{b}$ et $Q_1(\vec{b}^+) = 0$

Si $\vec{b} \notin C$, $\vec{b}^+ \in \delta C$; dans ce cas $\vec{b}^+$ est la projection de $\vec{b}$ sur δC .

Dans ce dernier cas, pour rechercher $\vec{b}^+$, on pourra utiliser un algorithme itératif qui à partir d'un point initial $\vec{b}_0 \in C$ (en fait on choisira $\vec{b}_0 \in \delta C$, puisque $\vec{b} \notin C$, $\vec{b}^+ \in \delta C$) converge vers $\vec{b}^+$.

III – CAS DE CONTRAINTES LINEAIRES

Nous supposons pour simplifier que les coefficients de régression doivent être positifs :

$$C = \{\vec{b} \in E : \forall i \in [1, p] : b_i \geq 0\} \quad (18)$$

C est donc le cône positif de E, délimité par l'ensemble des sous-espaces vectoriels où une ou plusieurs coordonnées sont nulles.

De même D est cône positif de H, quand on prend $\vec{x}_1, \dots, \vec{x}_p$ comme base dans H.

D contenant l'origine, et étant délimité par une série de sous espaces vectoriels, le théorème de Pythagore reste valable :

$$\|\vec{y}\|^2 = \|\vec{t}^+\|^2 + \|\vec{y} - \vec{t}^+\|^2 \quad (9)$$

ou encore

$$R^2 + S^2 = 1 \quad (10)$$

En effet si $\vec{t}^+ = \vec{t}$, (9) est vérifié, tandis que si $\vec{t}^+ \neq \vec{t}$, comme $\vec{t}^+ \in \delta D$, si H^+ désigne la face (le sous espace vectoriel) où se trouve $\vec{t}^+$, $\vec{t}^+$ est la projection de $\vec{y}$ sur H^+

Si $\vec{b} \notin C$, i.e. si un au moins des coefficient de régression classique est négatif, on doit minimiser

$$Q_1(b) = \|\vec{b} - \vec{b}\|_M^2$$

sous les contraintes linéaires : $\vec{b} \in \delta C$

On pourra pour rechercher la solution $\vec{b}^+$, utiliser une procédure itérative comme celle du gradient réduit de Wolfe, ou des méthodes dérivées. Nous n'insisterons pas plus sur ces procédures classiques que nous avons déjà développées longuement par ailleurs (cf. (2), (3)).

IV – CAS DE CONTRAINTES NON LINEAIRES

IV-1. La ridge régression (cf. (4), (5), (6))

Dans la ridge régression, on impose au vecteur de régression d'avoir une norme pas trop élevée de façon à éviter que la reconstitution de la variable y à expliquer se fasse à l'aide de termes de valeurs élevées, qui se retranchant donnent une reconstitution illusoire de y .

De façon plus précise, si E est muni de la métrique M_1 on impose donc que

$$\|\vec{b}\|_{M_1}^2 \leq a^2 \quad (19)$$

Si la solution $\vec{b}$ des moindres carrés vérifie (19), la solution est acceptable, sinon la solution minimisant $Q(\vec{b})$ (ou $Q_1(\vec{b})$) sous la contrainte (19) sera telle que :

$$\|\vec{b}\|_{M_1}^2 = a^2 \quad (20)$$

Employant le multiplicateur de Lagrange k , on devra donc minimiser

$$Q_1(\vec{b}) + k(\|\vec{b}\|_{M_1}^2 - a^2) = \|\vec{b} - \vec{\hat{b}}\|_M^2 + k(\|\vec{b}\|_{M_1}^2 - a^2) \quad (21)$$

d'où l'on déduit par dérivation matricielle

$$M(\vec{b} - \vec{\hat{b}}) + kM_1 \vec{b} = 0 \quad \text{pour} \quad \vec{b} = \vec{b}^+$$

$$\text{soit} \quad (M + kM_1) \vec{b}^+ = M\vec{\hat{b}}$$

$$\text{ou} \quad (XNX' + kM_1) \vec{b}^+ = XNX'\vec{\hat{b}} = XN\vec{y} \quad (22)$$

La solution $\vec{b}^+$ est donc donnée par :

$$\vec{b}^+ = (XNX' + kM_1)^{-1} XN\vec{y} \quad (23)$$

k étant déterminé de telle sorte que

$$\|\vec{b}^+\|_{M_1}^2 = a^2 \quad (24)$$

Posons

$$\vec{b}_k = (XNX' + kM_1)^{-1} XN\vec{y} \quad (25)$$

$$f(k) = \|\vec{b}_k\|_{M_1}^2 \quad (26)$$

On montre (cf. annexe 1 de (3)) qu'au voisinage de $k = 0$ et pour $k \geq 0$, $f(k)$ est une fonction décroissante de k , et que le minimum de (21) sous la contrainte $f(k) = a^2$ (sachant que $f(0) = \|\vec{b}\|_{M_1}^2 > a^2$) est atteint pour une valeur positive de k .

La recherche de k , tel que $f(k) = a^2$ étant assez complexe, car elle exige la diagonalisation de $XNX'M_1^{-1}$, il semble plus pratique pour déterminer $\vec{b}^+$ d'employer un processus itératif en appliquant la formule (25) pour diverses valeurs de k , et en comparant pour chaque valeur de k la norme du vecteur $\vec{b}_k$ ainsi obtenu à a .

Cette façon de procéder est également intéressante si on ne sait pas à priori quelle limitation donner à $\|\vec{b}^+\|$; on appliquera alors la formule (25) pour $k = 0 ; 0,1 ; 0,2 ; \dots$; et l'on s'arrêtera quand tous les coefficients de régression se seront stabilisés.

Du point de vue de la métrique M_1 , il semble logique, soit de prendre la métrique unité (cas usuel que l'on rencontre dans la littérature), soit la métrique diagonale ayant mêmes éléments diagonaux que XNX' .

IV-2. Régression sur variables entachées d'erreurs (cf. (1))

IV-2.1. Introduction

On suppose ici que du fait des conditions d'observation, chaque variable observée (variable explicative $x_1, \dots, x_p$ ou variable à expliquer y) est entachée d'une erreur, ce qui revient à dire que chaque variable observée est la somme des deux termes : la variable que l'on veut effectivement observer plus une erreur.

On a donc

$$\begin{cases} y = v_0 + r_0 \\ x_i = v_i + r_i \end{cases} \quad (i = 1, 2, \dots, p) \quad (27)$$

où $v_0, v_1, \dots, v_p$ sont les variables vraies supposées centrées

$y, x_1, \dots, x_p$ sont les variables observées

$r_0, r_1, \dots, r_p$ sont les erreurs aléatoires centrées, de variances connues respectivement égales à $s_0^2, s_1^2, \dots, s_p^2$.

Nous supposons ces distorsions indépendantes entre elles, et indépendantes des variables $v_0, v_1, \dots, v_p$, et nous désignerons par C_1 la matrice de variance de $r_1, r_2, \dots, r_p$:

$$C_1 = \begin{pmatrix} s_1^2 & & 0 \\ & \ddots & \\ 0 & & s_p^2 \end{pmatrix} \quad (28)$$

Désirant expliquer v_0 par rapport aux v_i ($1 \leq i \leq p$), mais ne connaissant $v_0, v_1, \dots, v_p$ que par les variables observées entachées d'erreurs $y, x_1, \dots, x_p$, on cherche la combinaison linéaire $y - \sum b^i x_i$ de variance minimale ; or ayant :

$$y - \sum_{i=1}^p b^i x_i = v_0 - \sum_{i=1}^p b^i v_i + r_0 - \sum_{i=1}^p b^i r_i$$

et du fait des hypothèses d'indépendance sur les erreurs, on a

$$\left\{ \begin{array}{l} \text{Var}(y - \sum b^i x_i) = \text{Var}(v_0 - \sum b^i v_i) + \text{Var}(r_0 - \sum b^i r_i) \\ \geq \text{Var}(r_0 - \sum b^i r_i) = s_0^2 + \sum b_i^2 s_i^2 \\ = s_0^2 + \|\vec{b}\|_{C_1}^2 \end{array} \right\} \quad (29)$$

La variance résiduelle est supérieure à la variance d'erreur.

Si l'on estime cette variance résiduelle à partir d'un n échantillon

$$(\vec{y}, \vec{x}_1, \dots, \vec{x}_p) \text{ de } (y, x_1, \dots, x_p)$$

et si l'on suppose que chaque observation j ($1 \leq j \leq n$) est affectée du poids p_j

$$\left(\sum_{j=1}^n p_j = 1 ; \quad \text{en général} \quad p_j = \frac{1}{n} \right),$$

et que l'on a centré les variables observées,

$$\left(\sum_j p_j y^j = 0, \quad \sum_j p_j x_i^j = 0 \quad (1 \leq i \leq p) \right),$$

cette variance résiduelle est estimée par la variance empirique égale à :

$$Q(\vec{b}) = \sum \{ p_j (y^j - \sum \{ b^i x_i^j | i = 1, p \})^2 | j = 1, n \} = \|\vec{y} - X' \vec{b}\|_{N=D_p}^2 \quad (30)$$

en choisissant comme métrique N dans R^n .

$$N = D_p = \begin{pmatrix} p_1 & & 0 \\ & \ddots & \\ 0 & & p_n \end{pmatrix} \quad (31)$$

Pour qu'elle ait un sens, cette variance doit être supérieure à la variance d'erreur :

$$Q(\vec{b}) \geq s_0^2 + \|\vec{b}\|_{C_1}^2 \quad (32)$$

On sera donc ramené à minimiser $Q(\vec{b})$, sous la contrainte (32).

IV-2.2. Recherche de la solution

Si la solution $\vec{b}$ des moindres carrés vérifie (32), alors elle convient ; si par contre

$$Q(\vec{b}) < s_0^2 + \|\vec{b}\|_{C_1}^2$$

le minimum de $Q(\vec{b})$ sous la contrainte (32) sera réalisé pour $\vec{b}^+$ tel que :

$$Q(\vec{b}^+) = s_0^2 + \|\vec{b}^+\|_{C_1}^2 \quad (33)$$

On est donc ramené à minimiser $Q(\vec{b})$ sous la contrainte

$$Q(\vec{b}) = s_0^2 + \|\vec{b}\|_{C_1}^2 \quad (33 \text{ bis})$$

Employant le multiplicateur de Lagrange λ , on doit donc minimiser

$$\left. \begin{aligned} & Q(\vec{b}) + \lambda (s_0^2 + \|\vec{b}\|_{C_1}^2 - Q(\vec{b})) \\ \text{ou encore} & (1 - \lambda) Q(\vec{b}) + \lambda (s_0^2 + \|\vec{b}\|_{C_1}^2) \end{aligned} \right\} \quad (34)$$

d'où l'on déduit par dérivation matricielle, puisque $Q(\vec{b}) = \|\vec{y} - X'\vec{b}\|_N^2$:

$$- (1 - \lambda) XN(\vec{y} - X'\vec{b}) + \lambda C_1 \vec{b} = 0 \quad \text{pour} \quad \vec{b} = \vec{b}^+$$

d'où

$$\left(XNX' + \frac{\lambda}{1 - \lambda} C_1 \right) \vec{b}^+ = XN\vec{y} \quad (35)$$

λ étant déterminé de telle sorte que (33) soit vérifiée.

Posant $k = \frac{\lambda}{1 - \lambda}$, on retrouve l'expression trouvée dans le cas de la ridge régression, quand on munit E de la métrique $M_1 = C_1$, l'expression que l'on minimise s'écrivant en mettant $1 - \lambda$ en facteur, et du fait que $k + 1 = 1/(1 - \lambda)$:

$$(Q(\vec{b}) + k(\|\vec{b}\|_{C_1}^2 + s_0^2))/(k + 1) \quad (36)$$

De même que pour la ridge régression, on montre (cf. annexe 2 de (3)) que la solution est fournie pour une valeur de k positive. On pourra donc, comme dans le cas de la ridge régression calculer $\vec{b}^+$ par une procédure itérative basée sur la formule (25), en comparant pour chaque valeur de k ,

$$Q(\vec{b}_k) \quad \text{et} \quad s_0^2 + \|\vec{b}_k\|_{C_1}^2.$$

Nous proposons ci-après une recherche directe de $\vec{b}^+$, basée sur les considérations géométriques exposées au II-4.

IV-2.3. Recherche directe de la solution

Désirant minimiser $Q(\vec{b}) = Q(\vec{b}) + \|\vec{b} - \vec{b}\|_{M=XNX'}^2$ sous la contrainte

$$Q(\vec{b}) = s_0^2 + \|\vec{b}\|_{C_1}^2 \quad (33 \text{ bis})$$

revient si l'on désigne par S la quadrique définie par (33 bis) soit à minimiser $\|\vec{b} - \vec{b}\|_M^2$ avec $\vec{b} \in S$ soit à minimiser, du fait de (33 bis) $\|\vec{b}\|_{C_1}^2$ avec $\vec{b} \in S$.

La solution $\vec{b}^+$ est donc à la fois

- la projection de $\vec{b}$ sur S , $E = (R^p)^*$ étant muni de la métrique M
- la projection de l'origine O sur S , E étant muni de la métrique $M_1 = C_1$.

Nous nous servirons de ce dernier point de vue pour rechercher la solution, et chercherons donc à minimiser $\|\vec{b}\|_{C_1}^2$ sachant que $\vec{b} \in S$.

Nous supposons donc E muni de la métrique $M_1 = C_1$.

L'algorithme pour trouver la projection de l'origine sur S, E étant muni de la métrique C_1 , (algorithme suggéré par le Professeur J.-P. Benzécri, et que nous avons légèrement modifié, pour tenir compte de certains cas particuliers qui peuvent se présenter) est exposé en annexe.

IV-2.4 Remarque (cas d'un modèle non centré)

Si $\vec{y}, \vec{x}_1, \dots, \vec{x}_p$ ne sont pas centrées (empiriquement), et si la variable certaine ne fait pas partie des $\vec{x}_i$ ($1 \leq i \leq p$), $Q(\vec{b})$ n'est plus un estimateur de la variance de $z = y - \sum \{b^i x_i \mid i = 1, p\}$, mais du moment d'ordre 2 de z.

Posant

$$\begin{aligned}\bar{z} &= \sum \{p_j(y^j - \sum \{b^i x_i^j \mid i = 1, p\}) \mid j = 1, n\} \\ &= \bar{y} - \sum \{b^i \bar{x}_i \mid i = 1, p\}\end{aligned}\quad (37)$$

en désignant par $\bar{y}$ et $\bar{x}_i$ les moyennes empiriques respectives de y et des x_i ($1 \leq i \leq p$) (l'observation j étant toujours supposée munie du poids p_j , et R^n étant muni de la métrique $N = D_p$ définie par (31)) on devra avoir :

$$Q(\vec{b}) = \|\vec{y} - X'\vec{b}\|_N^2 \geq \|\vec{b}\|_{C_1}^2 + s_0^2 + \bar{z}^2 \quad (38)$$

Si donc $\vec{b}$ vérifie (38), la solution usuelle des moindres carrés convient, sinon la solution $\vec{b}^+$ minimisant $Q(\vec{b})$ sous la contrainte (38) sera telle que :

$$Q(\vec{b}^+) = \|\vec{b}^+\|_{C_1}^2 + s_0^2 + \bar{z}^2 \quad (39)$$

Employant comme au IV 2.2 la méthode des multiplicateurs de Lagrange, on obtient :

$$(M + k(C_1 + D))\vec{b}^+ = XN\vec{y} + kE \quad (40)$$

où — D est la matrice (p, p) de terme général $d_{ij} = \bar{x}_i \bar{x}_j$

E le vecteur (p, 1) de composante $e_i = \bar{x}_i \bar{y}$

k étant déterminé de telle sorte que (39) soit vérifiée, ce qui permet de trouver la solution, comme dans le cas où l'on est centré, à l'aide d'un processus itératif, basé sur la formule (40)

V Un exemple d'application de la régression sous contraintes

Nous reprenons ici un exemple que nous avons déjà exposé, (cf. (2)) et où nous avons appliqué la régression sous contraintes de positivité des coefficients de régression, contraintes linéaires qui s'imposaient pour cet exemple. Nous avons appliqué la ridge régression et la régression sur variables entachées d'erreurs pour comparer les résultats obtenus par ces méthodes avec la précédente.

Rappelons de quel type de problème il s'agissait : on possède l'histogramme d'une variable dont la loi de probabilité G s'écrit :

$$G = \sum \{b^i P_i \mid i = 1, p\}$$

où P_i est la loi de Poisson de paramètre α_i connu ; les b^i sont les proportions de particules d'un nuage de particules en suspension dans un gaz, de diamètre donné (ou plus exactement appartenant à un intervalle de diamètre donné d_{i-1}, d_i), et le problème est d'estimer les b^i à partir de l'histogramme précédent (obtenu de façon expérimentale).

Si on divise l'histogramme expérimental en n classes, et si N_j désigne le nombre d'observations tombé dans la classe j , et N le nombre total d'observations, $y^j = N_j/N$ est l'estimation du maximum de vraisemblance de la probabilité G^j de se trouver dans la $j^{\text{ème}}$ classe :

$$G^j = \sum \{b^i x_i^j | i = 1, p\}$$

où x_i^j désigne la probabilité pour la variable de Poisson de loi P_i d'appartenir à la $j^{\text{ème}}$ classe.

On cherchera donc si l'on pose

$$e^j = y^j - \sum \{b^i x_i^j | i = 1, p\}$$

à minimiser $\sum \{(e^j)^2 | j = 1, n\}$, ce qui revient à munir l'espace $F = R^n$ de la métrique unité.

Remarquons que pour que les résultats aient un sens, il faut que les b^i soient positifs et de somme 1.

Nous donnons ici les résultats relatifs à deux histogrammes, l'un H_1 à 16 classes, l'autre H_2 à 12 classes (n vaut soit 16, soit 12), le nombre p de loi de Poisson étant dans les deux cas égal à 8.

Pour chacun des deux histogrammes, nous avons fait :

- La régression classique.
- La régression classique en imposant que la somme des estimations $\hat{b}^i$ des coefficients b^i vale 1.
- La ridge régression classique.
- La ridge régression en imposant à la somme des estimations $(b^i)^+$ des b^i de valoir 1.
- La régression sous contraintes (estimations $(b^i)^+$ des b^i positives).
- La régression sous contraintes (estimations $(b^i)^+$ des b^i , positives et de somme 1).
- La régression sur variable entachées d'erreurs⁽¹⁾ (sans imposer à la somme des $(b^i)^+$ de valoir 1⁽²⁾).

La somme des estimations $\hat{b}^i$ des b^i obtenues par la régression classique valant 1,0001 (histogramme H_1 à 16 classes), et 0,998 (histogramme H_2 à 12 classes), les résultats de cette régression sont identiques à ceux de la régression où l'on impose à cette somme de valoir 1 ; nous ne donnerons donc que les résultats relatifs à cette dernière régression.

(1) Nous avons utilisé cette méthode, quoique n'étant pas exactement dans le cadre où elle s'applique, pour comparer numériquement la protection qu'elle donne à la régression par rapport aux autres méthodes également utilisées. (On est dans le cas où l'on a un modèle non centré empiriquement (cf. IV-2.4)).

(2) Si on impose à la somme des $(b^i)^+$ d'être égale à 1, on a un modèle centré, et on retombe donc sur les résultats fournis par la ridge régression où $\sum (b^i)^+ = 1$.

Les résultats sont résumés sur les tableaux 1 à 4 ; sur les tableaux 1 et 3, on a reporté les résultats fournis par la régression usuelle, la ridge régression ($\Sigma (b^i)^+ = 1$), la régression sous contrainte de positivité des $b^i (b^i > 0 ; b^i > 0 \text{ et de somme } 1)$.

On voit que la régression classique donne des résultats aberrants, certains coefficients de régression étant négatifs ou de module supérieur à 1 (dans le cas du tableau 3) ; ceci est essentiellement dû au fait que les variables explicatives sont très corrélées. De plus le résidu (qui est proportionnel à la variance résiduelle, ou plus exactement au moment d'ordre 2) est très petit (inférieur à $5 \cdot 10^{-3}$ dans les deux cas exposés ici).

C'est la raison pour laquelle, outre la régression, où l'on imposait aux valeurs $(b^i)^+$ de b^i d'être positives on a appliqué la ridge régression de façon à limiter la valeur de la norme de $\vec{b}$, et la régression sur variables entachées d'erreur, de façon à ce que le moment résiduel d'ordre 2 ne soit pas trop faible.

Pour la ridge régression, nous avons choisi comme métrique M_1 dans $(R^p)^*$, la métrique diagonale ayant mêmes éléments diagonaux que la matrice $XX' = X'X$ et nous avons fait varier k (cf. IV 1) par pas de 0,05 de 0 à 0,95.

A partir d'une valeur k_0 de k (du moins quand on impose à la somme des coefficients de régression de valoir 1 : $k_0 = 0,3$ pour la première régression ; $k_0 = 0,35$ pour la seconde), toutes les estimations des coefficients de régression deviennent positives. Pour cette valeur k_0 de k , les résultats de la ridge régression ($\Sigma (b^i)^+ = 1$) sont voisins de ceux fournis par la régression sous contraintes linéaires (coefficients de régression positifs ; coefficients de régression positifs et de somme 1).

Les tableaux 2 et 4 résument les résultats fournis par la ridge régression ($\Sigma (b^i)^+ = 1$) pour différentes valeurs de k ; on a également reporté sur ces tableaux les résultats de la ridge régression où l'on n'impose pas à la somme des $(b^i)^+$ de valoir 1. On voit que dans ce cas, dès que $k \geq 0,05$, les valeurs négatives des $(b^i)^+$ deviennent faibles en module et pour $\|\vec{b}^+\|_{M_1}^2 = 0,05^{(1)}$, les résultats obtenus sont voisins de ceux fournis par la ridge régression où $\Sigma (b^i)^+ = 1$, pour la même valeur 0,05 de $\|\vec{b}^+\|_{M_1}^2$.

En ce qui concerne la seconde régression, (cf. tableau 4), dès que k est supérieur ou égal à 0,05, toutes les estimations $(b^i)^+$ des b^i (que l'on impose ou non $\Sigma (b^i)^+ = 1$) deviennent inférieures en module à 1, le carré de norme de $\vec{b}^+$ passant de 58,95 ($k = 0$), à une valeur plus petite ou égale à 0,06 ($k \geq 0,05$).

Quand k augmente, dans les deux cas étudiés (histogramme H_1 et histogramme H_2), les valeurs $(b^i)^+$ peu stables dans la ridge régression ($\Sigma (b^i)^+ = 1$), correspondent aux $(b^i)^+$ annulés par les contraintes de positivité.

En fait, la somme pour $i \geq i_0$ des $(b^i)^+$

$$(\Sigma \{(b^i)^+ \mid i = i_0, p\}) \quad (i_0 = 5 \text{ pour } H_1 ; i_0 = 6 \text{ pour } H_2)$$

(1) Ce qui correspond à $k = 0,15$ pour la première régression, et $k = 0,10$ pour la seconde.

obtenus par la ridge régression ($\sum \{(b^i)^+ \mid i = 1, p\} = 1$) varie peu quand k varie, cette somme étant voisine de la même somme calculée sur les $(b^i)^+$ fournis par la régression sous contraintes de positivité ($(b^i)^+ > 0$; et $(b^i)^+ > 0$ et de somme 1) ; dans ce dernier cas, les $(b^i)^+$ étant nuls pour $i > i_0$, cette somme se réduit à $(b^{i_0})^+$.

D'un point de vue physique, cela revient à dire que ce que l'on estime avec précision, c'est le pourcentage de particules de diamètre supérieur ou égal à d_{i_0-1} .

En ce qui concerne la régression sur variables entachées d'erreurs que nous avons appliquée, comme on l'a déjà signalé, pour la comparer à la ridge régression, quoiqu'on ne soit pas exactement dans le cadre d'application de cette technique, nous avons adopté pour matrice variance d'erreurs (cf. IV 2.1) la matrice $C_1 = \alpha M_1$, où M_1 est la métrique que nous avons adoptée dans la ridge régression, et α un coefficient positif plus petit que 1. De même on a pris

$$s_0^2 = (\alpha/n) \sum_{j=1}^n (y^j)^2.$$

Nous avons fait varier α à partir de 0 (régression classique)

par pas de 0,01. Dans le cas du premier histogramme tous les coefficients de régression deviennent positifs pour $\alpha = 0,02$, ce qui n'a lieu que pour $\alpha = 0,03$ pour le second histogramme. Dans les deux cas la somme des $(b^i)^+$ est élevée et voisine de 1 ($\sim 0,99$). Les résultats obtenus sont voisins, ce qui est normal d'après le choix de C_1 , des résultats fournis par la ridge régression (cf. tableaux 2 et 4). Il semble néanmoins que la protection de la régression soit meilleure, ceci étant dû au fait que l'on a introduit pour la variable à expliquer une variance d'erreur, et qu'ayant un modèle non centré, on est dans le cadre du IV 2.4 qui assure une protection plus efficace (cf. équations (37) à (40)).

L'approximation réalisée par ces différents types de régression est excellente, puisque le résidu S^2 (égal à $\|\vec{y} - \vec{y}^*\|^2 / \|\vec{y}\|^2$, si $\vec{y}^*$ désigne l'approximation de $\vec{y}$) est toujours inférieur à 0,1, sauf dans le cas de la ridge régression où l'on n'impose pas à la somme des $(b^i)^+$ de valoir 1, S^2 devenant supérieur à 0,1 pour $k = 0,75$, et atteignant pour $k = 0,95$ la valeur 0,136 dans le cas de la première régression, et 0,128 dans le cas de la seconde.

En conclusion, on voit que sur les deux cas étudiés, et en particulier dans le second, la ridge régression, la régression sous contrainte de positivité des coefficients, et la régression sur variables entachées d'erreurs protègent bien la régression, et donnent des résultats équivalents, la régression sous contrainte de positivité semblant s'imposer dans ces deux exemples du fait de la nature même du modèle.

TABLEAU I

i	$\hat{b}^i$	$\hat{\sigma}_i$	$(b^i)^+$	$(b^i)^+$	$(b^i)^+$
1	0.087	0.055	0.076	0.080	0.064
2	0.104	0.083	0.104	0.138	0.141
3	0.131	0.040	0.125	0.105	0.110
4	0.190	0.021	0.212	0.200	0.212
5	0.572	0.033	0.404	0.477	0.490
6	-0.201	0.088	0.058	0	0
7	0.161	0.128	0.001	0	0
8	-0.044	0.07	0.020	0	0
Σb^i	1		1	1	1.017
S^2	0.0041		0.026	0.0098	0.0095
Modèle ou technique	Moindres carrés classique ($\Sigma b^i = 1$)		ridge régression ($\Sigma (b^i)^+ = 1$) $\ \hat{b} \ ^2 \leq 0.05$ (k = 0,3)	contraintes $(b^i)^+ > 0$ $\Sigma (b^i)^+ = 1$	contraintes $(b^i)^+ > 0$

Comparaison des résultats obtenus entre la régression classique la ridge régression, et la régression sous contraintes de positivité $((b^i)^+ > 0)$ dans le cas de l'histogramme à 16 classes.

$$S^2 = \frac{\| \vec{y} - \vec{y}^+ \|^2}{\| \vec{y}^+ \|^2}$$

avec $\vec{y}^+$ approximation de $\vec{y}$ pour le modèle considéré.

$\hat{\sigma}_i$ estimation de l'écart type de $\hat{b}^i$.

Tableau II

i	$(b^i)_0^+ = \hat{b}_0$	$(b^i)_{0,05}^+$	$(b^i)_{0,30}^+$	$(b^i)_{0,95}^+$	$(b^i)_{0,05}^+$	$(b^i)_{0,15}^+$	$(b^i)_{0,95}^+$	$(b^i)^+$	$(b^i)^+$
1	0,087	0,080	0,076	0,066	0,080	0,077	0,059	0,078	0,075
2	0,104	0,112	0,104	0,092	0,111	0,106	0,080	0,107	0,103
3	0,131	0,125	0,125	0,123	0,122	0,118	0,096	0,124	0,124
4	0,190	0,204	0,212	0,203	0,198	0,195	0,143	0,210	0,210
5	0,572	0,503	0,404	0,319	0,496	0,432	0,246	0,439	0,393
6	-0,201	-0,046	0,058	0,120	-0,049	0,004	0,059	0,025	0,064
7	0,161	-0,009	0,001	0,046	-0,012	-0,023	-0,002	-0,012	0,003
8	-0,044	0,031	0,020	0,031	0,027	0,017	-0,004	0,023	0,019
$\Sigma (b^i)^+$	1	1	1	1	0,973	0,926	0,677	0,994	0,991
k	0	0,05	0,30	0,95	0,05	0,15	0,95		
$\ \vec{b}^+ \ ^2$	0,081	0,062	0,049	0,038	0,060	0,050	0,023	0,053	0,047
S^2	0,0041	0,0063	0,026	0,082	0,0067	0,016	0,016	0,0154	0,0289
	Ridge régression $\Sigma (b^i)_k^+ = 1$				Ridge régression classique			Régression sur variables entachées d'erreurs $\alpha = 0,01 \quad \alpha = 0,02$	

Etude en fonction de k des résultats fournis par la ridge régression, et comparaison avec les résultats de la régression sur variables entachées d'erreurs, dans le cas de l'histogramme à 16 classes.

Tableau III

i	$\hat{b}^i$	$\hat{\sigma}_i$	$(b^i)^+$	$(b^i)^+$	$(b^i)^+$
1	5.095	5.28	0.052	0.012	0.077
2	-6.041	6.38	0.059	0.079	0.0003
3	1.246	0.18	0.102	0.129	0.161
4	0.050	0.10	0.141	0.110	0.106
5	0.240	0.036	0.280	0.291	0,288
6	0.671	0.084	0.308	0.379	0.374
7	-0.560	0.178	0.056	0	0
8	0.299	0.130	0.002	0	0
Σb^i	1		1	1	1.0063
S^2	0.0034		0.032	0.0194	0.0191
Modèle ou technique	Moindres carrés classiques ($\Sigma \hat{b}^i = 1$)		ridge régression ($\Sigma (b^i)^+ = 1$) $\ \hat{b}^+\ ^2 \leq 0,046$ ($k = 0.35$)	contraintes ($(b^i)^+ > 0$) $\Sigma (b^i)^+ = 1$	contraintes ($(b^i)^+ > 0$)

Comparaison des résultats obtenus entre la régression classique, la ridge régression, et la régression sous contraintes linéaires ($(b^i)^+ > 0$) (histogramme à 12 classes)

$$S^2 = \frac{\|\vec{y} - \vec{y}^+\|^2}{\|\vec{y}^+\|^2}$$

où y^+ est l'approximation de y pour le modèle considéré. $\hat{\sigma}_i$ estimation de l'écart type de $\hat{b}^i$.

Tableau IV

i	$(b^i)_0^+ = \hat{b}_i$	$(b^i)_{0,05}^+$	$(b^i)_{0,35}^+$	$(b^i)_{0,95}^+$	$(b^i)_{0,05}^+$	$(b^i)_{0,10}^+$	$(b^i)_{0,60}^+$	$(b^i)_{0,95}^+$	$(b^i)^+$	$(b^i)^+$
1	5,095	0,048	0,052	0,051	0,047	0,049	0,047	0,044	0,049	0,052
2	-6,041	0,056	0,059	0,057	0,056	0,056	0,052	0,049	0,057	0,058
3	1,246	0,119	0,102	0,095	0,118	0,110	0,086	0,077	0,114	0,098
4	0,050	0,132	0,141	0,144	0,126	0,126	0,109	0,098	0,133	0,141
5	0,240	0,271	0,280	0,267	0,260	0,258	0,208	0,179	0,274	0,272
6	0,671	0,427	0,308	0,256	0,417	0,372	0,222	0,182	0,394	0,279
7	-0,560	-0,033	0,056	0,088	-0,035	0,003	0,045	0,047	-0,000	0,069
8	0,299	-0,02	0,002	0,042	-0,025	-0,033	0,001	0,010	-0,028	0,017
$\Sigma (b^i)^+$	1	1	1	1	0,964	0,941	0,770	0,686	0,993	0,986
k	0	0,05	0,35	0,95	0,05	0,10	0,60	0,95		
$\ \hat{b}^+\ ^2$	58,95	0,060	0,046	0,041	0,057	0,050	0,028	0,021	0,055	0,043
S^2	0,0034	0,013	0,032	0,061	0,0133	0,018	0,081	0,128	0,015	0,042
	Ridge régression $\Sigma (b^i)_k^+ = 1$				Ridge régression classique				Régression sur variables entachées d'erreurs $\alpha = 0,01 \quad \alpha = 0,03$	

Etude en fonction de k des résultats fournis par la ridge régression et comparaison avec les résultats de la régression sur variables entachées d'erreurs, dans le cas de l'histogramme à 12 classes

ANNEXE

Recherche dans E de la projection $\vec{b}^+$ de l'origine sur la quadrique S d'équation $Q(\vec{b}) = s_0^2 + \|\vec{b}\|_{C_1}^2$, E étant muni de la métrique

$$C_1 = \begin{pmatrix} s_1^2 & & & \\ & \ddots & & \\ & & \circ & \\ & & & \ddots \\ & & \circ & & \\ & & & & s_p^2 \end{pmatrix}$$

on est dans le cas où

$$Q(\vec{b}) < s_0^2 + \|\vec{b}\|_{C_1}^2 \quad (1)$$

avec, d'après la relation (30) du IV 2.1 dont on conserve les notations :

$$Q(0) = \|\vec{y}\|_{N=D_p}^2 = \text{Var } y > s_0^2 \quad (2)$$

la variance d'erreur sur y ne pouvant être supérieure à la variance empirique de y.

Les relations (1) et (2) montrent que la solution est en quelque sorte comprise entre 0 et $\vec{b}$, d'où l'algorithme exposé ci-dessous :

1/ Point de départ :

On recherche l'intersection $\vec{b}^0$ du segment joignant 0 à l'extrémité de $\vec{b}$ avec S.

2/ Connaissant à l'itération it, $\vec{b}^{it}$ (it ≥ 0)

a) On calcule le vecteur gradient $\vec{N}^{it}$ en $\vec{b}^{it}$ (i.e. le vecteur normal à S en $\vec{b}^{it}$), vecteur dont la j^{ème} composante vaut $(1/s_j) (\partial S / \partial b^j)_{\vec{b}=\vec{b}^{it}}$, en posant :

$$S = Q(\vec{b}) - s_0^2 - \|\vec{b}\|_{C_1}^2$$

On norme $\vec{N}^{it}$ ($\|\vec{N}^{it}\|_{C_1} = 1$)

b) On calcule $c = \langle \vec{b}^{it}, \vec{N}^{it} \rangle_{C_1}$

c) On calcule $\cos A = \langle \vec{b}^{it}, \vec{N}^{it} \rangle_{C_1} / \|\vec{b}^{it}\|_{C_1}$

Si $\cos A < \text{PRECI}$, on va en g), sinon on va en d)

d) Si $\sin A < \epsilon$, on arrête le processus, car on a atteint la convergence ; $\vec{b}^{it}$ est la solution

Si $\sin A > \epsilon$, on va en e)

e) On calcule le vecteur tangent $\vec{T}^{it}$ à S en $\vec{b}^{it}$, dans le plan $(\vec{b}^{it}, \vec{N}^{it})$:

$$\vec{T}^{it} = (-\vec{b}^{it} + c \vec{N}^{it}) / \|\vec{b}^{it} + c \vec{N}^{it}\|_{C_1}$$

f) On calcule

$$t = -\langle \vec{b}^{it}, \vec{T}^{it} \rangle_{C_1}$$

(i.e. $\vec{b}^{it} + t \vec{T}^{it}$ est la projection de 0 sur la tangente en $\vec{b}^{it}$ à S, dans le plan $(\vec{b}^{it}, \vec{N}^{it})$).

On va en 3)

g) On calcule $t (t \neq 0) : \vec{b}^{it} + 2 t \vec{N}^{it} \in S$

– On pose $\vec{T}^{it} = \vec{N}^{it}$

(dans ce cas $\vec{T}^{it}$ désigne le vecteur normal à S) :

– On va en 3)

3/

a) On recherche u tel que $\vec{b} = (1 - u) \vec{b}^{it} + t \vec{T}^{it}$ appartienne à S (i.e. on recherche l'intersection la plus proche de 0 de la droite parallèle à $\vec{b}^{it}$ et passant par le point $\vec{b}^{it} + t \vec{T}^{it}$, avec S)

b) Si cette intersection existe, et si $\|\vec{b}\|^2 < \|\vec{b}^{it}\|^2$, alors on pose $it = it + 1$, $\vec{b}^{it} = \vec{b}$, et l'on revient en 2), sinon, on fait $t = t/2$

si $t < \epsilon$, on arrête le processus

$t > \epsilon$, on revient en 3) a).

Commentaire.

Pour trouver à partir de $\vec{b}^{it}$ appartenant à S, un point de S plus proche de 0 que $\vec{b}^{it}$, on va se placer dans le plan P formé par $\vec{b}^{it}$ et par la normale $\vec{N}^{it}$ à S en $\vec{b}^{it}$.

Si $\vec{b}^{it}$ est parallèle à $\vec{N}^{it}$ seul cas où le plan P précédent se réduit à une droite, $\vec{b}^{it}$ est la projection de 0 sur S ; on a donc obtenu la solution, et on arrête le processus.

De façon pratique, si A désigne l'angle entre $\vec{N}^{it}$ et $\vec{b}^{it}$, on arrête les itérations, si $|\sin A| < \epsilon$ ($\epsilon = 10^{-7}$ par exemple).

Si $|\sin A| > \epsilon$, on pose :

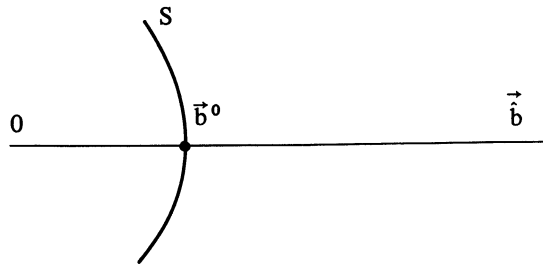
$$\vec{b} = \vec{b}^{it} + t \vec{T}^{it} - u \vec{b}^{it} \quad (3)$$

et l'on détermine t et u de telle façon que $\vec{b}$ soit un point de S plus proche de 0 que $\vec{b}^{it}$.

Si $\vec{b}^{it}$ n'est pas parallèle au vecteur tangent en $\vec{b}^{it}$ à S dans P, on prend pour $\vec{T}^{it}$ ce vecteur tangent.

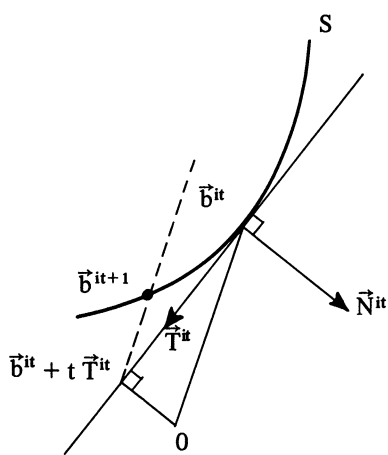
Si $\vec{b}^{it}$ est parallèle à ce vecteur tangent, on ne peut plus prendre pour $\vec{T}^{it}$ ce vecteur tangent, puisque le seul $\vec{b}$ défini par (3) et appartenant à S correspondrait dans ce cas à $t = u = 0$, (i.e. $\vec{b} = \vec{b}^{it}$).

Si donc $\vec{b}^{it}$ est parallèle à ce vecteur tangent ($\cos A = 0$; en pratique $|\cos A| < \text{PRECI} \approx 0,1$), on prendra pour $\vec{T}^{it}$ le vecteur normal $\vec{N}^{it}$ à S en $\vec{b}^{it}$,

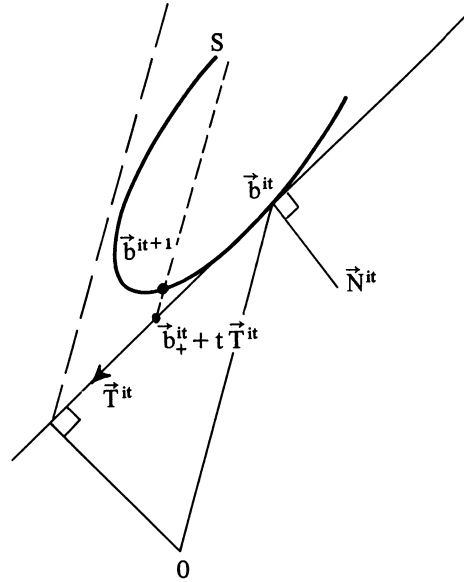


point de départ de l'algorithme

Figure 1 a

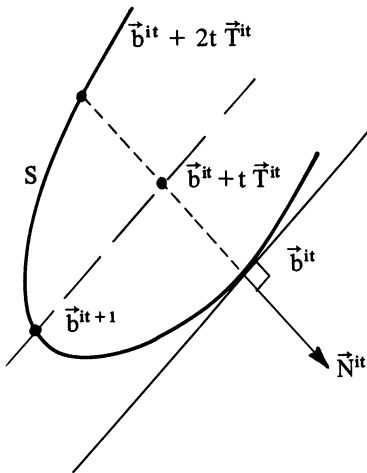


1^{er} cas de figure



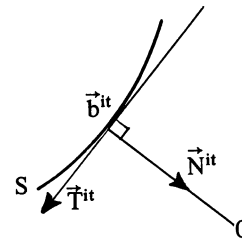
2^{ème} cas de figure :

on a dû diviser t par 2 pour obtenir u et $\vec{b}^{it+1}$



3^{ème} cas de figure :

$\vec{b}^{it}$ est parallèle à la tangente en $\vec{b}^{it+1}$ à S , on prend $\vec{T}^{it} = \vec{N}^{it}$



4^{ème} cas de figure :

$\vec{b}^{it}$ est parallèle à $\vec{N}^{it}$: on a atteint la convergence

Passage de $\vec{b}^{it}$ à $\vec{b}^{it+1}$

Figure 1b

Recherche de la projection de l'origine 0 sur la quadrique S par l'algorithme exposé en annexe

Figure 1 de l'annexe

t est alors déterminé comme on l'a déjà indiqué de telle façon que
 - $\vec{b}^{it} + t \vec{T}^{it}$ soit la projection de 0 sur la tangente en $\vec{b}^{it}$ à S, dans P, dans le cas où $\vec{T}^{it}$ est le vecteur tangent à S dans P en $\vec{b}^{it}$.
 - $\vec{b}^{it} + 2 t \vec{T}^{it}$ soit la seconde intersection de la normale $\vec{N}^{it}$ en $\vec{b}^{it}$ à S, avec S (dans le cas où $\vec{T}^{it} = \vec{N}^{it}$).

Ayant ainsi déterminé t , s'il existe une valeur de u telle que $\vec{b}$ défini par (3) appartienne à S et soit plus proche de 0 que $\vec{b}^{it}$, on a obtenu le nouvel itéré $\vec{b}^{it+1} = \vec{b}$, sinon on fera $t = \frac{t}{2}$ (i.e. on partira pour rechercher u d'un point plus proche de $\vec{b}^{it}$), et l'on recherchera à nouveau u tel que $\vec{b}$ défini par (3) appartienne à S, et soit plus proche de 0 que $\vec{b}^{it}$. Si un tel u existe, on posera $\vec{b}^{it+1} = \vec{b}$, sinon on fera $t = \frac{t}{2}$, et l'on recherchera à nouveau u ; l'on s'arrêtera si $t < \epsilon \approx 10^{-7}$.

La figure 1b) illustre un certain nombre de cas se présentant pour passer de $\vec{b}^{it}$ à $\vec{b}^{it+1}$.

Remarque : Dans l'algorithme décrit précédemment, on peut au lieu de partir de $\vec{b}$ pour trouver $\vec{b}^+$, partir de $\vec{b}_k$ tel que :

$$\vec{b}_k = (X N X' + k C_1)^{-1} X N \vec{y} \quad (4)$$

$$Q(\vec{b}_k) < s_0^2 + \|\vec{b}_k\|_{C_1}^2 \quad (5)$$

En particulier, on peut pour rechercher la solution $\vec{b}^+$ utiliser d'abord le processus itératif défini à partir de l'équation (4), k variant à partir de 0 avec un pas égal à h , et si

$$Q(\vec{b}_k) < s_0^2 + \|\vec{b}_k\|_{C_1}^2$$

$$Q(\vec{b}_{k+h}) > s_0^2 + \|\vec{b}_{k+h}\|_{C_1}^2$$

Obtenir la solution exacte (qui est en quelque sorte comprise entre $\vec{b}_k$ et $\vec{b}_{k+h}$) en utilisant l'algorithme que nous venons de décrire (où l'on part de $\vec{b}_k$, au lieu de $\vec{b}$, pour déterminer $\vec{b}^+$).

BIBLIOGRAPHIE

- [1] BENZECRI — La régression (publication du laboratoire de statistique mathématique) 1970.
- [2] CAZES-TURPIN — Régression sous contraintes. Application à l'estimation de la courbe granulométrique d'un aérosol. *Revue de Statistique appliquée*, 1971, 19, 4.
- [3] CAZES Protection de la régression
 Régression sous contraintes Doc. 74/17 (publication du laboratoire de biométrie du CNRZ, Jouy en Josas, et du laboratoire de statistique mathématique), Sept. 74.

- [4] HOERL-KENNARD — Ridge régression : Biased estimation for non orthogonal problems, *Technometrics*, Vol. 12, 1, 1970.
- [5] HOERL-KENNARD — Ridge régression : Applications to non orthogonal problems, *Technometrics*, vol. 12, 1, 1970.
- [6] MARQUARDT — Generalized inverses, Ridge régression, biased linear estimation and non linear estimator. *Technometrics*, Vol. 12, n° 3, 1970.