

# MÉTHODES D'ANALYSE NON LISSE POUR L'ESTIMATION ROBUSTE

*Mireille L. Bougeard*

**Résumé.** — La validation d'un modèle d'astronomie via l'analyse de données expérimentales conduit souvent à résoudre un problème mal posé, numériquement et statistiquement. Ceci motive un intérêt croissant pour les méthodes d'estimation robuste, parmi lesquelles la classe des M-estimateurs joue un rôle fondamental et optimal. Introduits par Huber, ils sont solutions d'un problème d'optimisation sans expression analytique des solutions, ni même toujours unicité. Nous montrons comment des méthodes d'analyse non lisse permettent de traiter élégamment le problème d'optimisation associé tant du point de vue primal que dual, et comment cette approche a été implémentée en routine pour l'analyse de la rotation terrestre.

**Abstract.** — In recent years, robustness is one problem that has been given much attention in statistical literature. Under the simultaneous occurrence of outliers and/or collinearity, various alternatives to Least Squares are available. Among them, the Huber-M estimators are currently attracting attention when the errors have a contaminated Gaussian distribution. Since they cannot be expressed analytically, finding efficient algorithms to produce them in the case of large data sets is a field of active research. We present new highly parallel algorithms based on the Spingarn Partial Inverse-proximal approach. Practical implementation is described with application to Earth Rotation series.

## 1. Introduction et problématique statistique

Une modélisation classique en astrométrie met en jeu un modèle linéarisé sous forme  $y = Ax + \varepsilon$  encore noté  $\{y, Ax, \sigma^2 I_n\}$  où  $A$  est une matrice  $(n \times m)$  donnée (comportant en colonne les observations des variables « régresseurs »),  $x$  est le vecteur des  $m$  paramètres inconnus,  $y$  celui des  $n$  valeurs de la variable dépendante (obtenues par soustraction entre chaque observation et son estimation théorique). Le vecteur d'erreurs aléatoires  $\varepsilon$  est supposé de moyenne nulle et de matrice de variance-covariance  $\sigma^2 I_n$ ,  $\sigma^2 > 0$ .

Une des approches les plus courantes pour déterminer  $x$  consiste à appliquer un ajustement par *Moindres Carrés* et à résoudre les équations normales  $A^t Ax = A^t y$

où  $A^t$  désigne la matrice transposée de  $A$ . Pour  $A$  de plein rang, l'estimateur des Moindres Carrés (MC),  $\hat{x}$ , possède des propriétés optimales dont la justification théorique repose sur le théorème de Gauss-Markov. Sans biais, il est, néanmoins, peu *robuste* au sens où la présence d'une seule observation atypique peut influencer fortement l'estimation [Huber, 1981].

Par ailleurs, il vient  $E(\hat{x}^t\hat{x}) = x^tx + \sigma^2\text{tr}[(A^tA)^{-1}]$ , ce qui indique qu'un mauvais conditionnement de la matrice  $A$  perturbera aussi les tests statistiques.

Comment construire de « bons estimateurs » de régression dans le cadre d'un modèle potentiellement sujet à des effets de colinéarité ou de points atypiques ? Cette question fondamentale dans les applications a généré de très nombreuses recherches. Ronchetti (1987) fait remonter à l'astronome Galilée (1632), le recours à un estimateur en norme  $L_1$  (valeur médiane) comme alternative à un estimateur en *moindres carrés* en norme  $L_2$  (moyenne arithmétique) pour l'estimation d'une valeur moyenne (encore dit *paramètre de localisation*).

Parmi le grand nombre d'alternatives statistiques à l'emploi de la résolution par *moindres carrés* figure en bonne place le recours à des techniques d'estimation robuste.

**1.1. Vers des estimations robustes.** — La recherche d'estimateurs *robustes* est devenu un sujet de forte actualité depuis ces dernières années en statistique. Dans le cas de l'estimation d'un paramètre de localisation, le manque de robustesse de l'estimateur moyenne arithmétique au regard de données atypiques a tôt attiré l'attention des praticiens. D'où le recours en alternative à des estimateurs comme la médiane, des moyennes tronquées ou à regroupements frontaliers (« winsorisées »)...

Dans le cadre plus général du modèle de régression linéaire qui nous concerne, que faut-il entendre par : *estimer les paramètres directement par une méthode robuste* ? Suivons les propos de [Van Cutsem, 1982] et disons en bref que dans cette démarche : « on considère... un modèle paramétrique (en souhaitant) tenir compte du fait que les observations peuvent, pour diverses raisons, venir non pas d'une loi précisée pour ce modèle mais plutôt d'une loi "assez proche"... représentée par un ensemble de loi "voisines"... ».

Ainsi, si  $F_0$  est la loi correspondant au modèle paramétrique et  $G$  l'ensemble de toutes les lois sur  $\mathfrak{R}$ , une loi « voisine » sera du type

$$F(\varepsilon) = (1 - \varepsilon) \cdot F_0 + \varepsilon \cdot H, \quad H \in G, \quad \varepsilon \in [0, 1[$$

Comme le note Van Cutsem que nous citons à nouveau : « l'approche robuste consiste intuitivement à chercher des solutions qui

- sont valides et ont une bonne efficacité pour [...] la loi choisie,
- restent relativement valides et efficaces si on s'éloigne un peu ( $\varepsilon$  petit)
- n'ont pas une validité et efficacité nulles si on est très éloigné ( $\varepsilon$  voisin de 1) de  $F_0$ .

Cette approche recherche donc des solutions qui peuvent nous protéger dans l'analyse [...] ».

Différents concepts ont été introduits pour quantifier la robustesse. La notion de *courbe d'influence* proposée par Hampel (1974) consiste à considérer les lois obtenues par translation avec contamination en un point  $t$  en :

$$F(\varepsilon) = (1 - \varepsilon) \cdot F_0 + \varepsilon \cdot \delta_t, \quad \varepsilon \in [0, 1[$$

où  $\delta_t$  est la loi de Dirac de support  $\{t\}$ . Ceci permet via divers outils mathématiques d'analyser l'influence de la contamination — par une valeur aberrante (« outlier ») au point  $t$  — sur le comportement du biais ou du biais asymptotique d'un estimateur particulier lorsque  $t$  prend des valeurs pouvant être très grandes. Nous renvoyons au livre [Hampel *et al*, 1986] pour une discussion approfondie de ces concepts.

**1.2. La classe des M-estimateurs.** — En estimation robuste, la classe des M-estimateurs joue un rôle fondamental et optimal. Ils doivent leur nom au fait qu'ils sont du type « maximum de vraisemblance » en étant obtenus par minimisation du maximum de la variance asymptotique (sous différents types de structure des erreurs observationnelles).

Introduits par Huber dans son article célèbre dans la revue *Annals of Statistics*, 1964, les M-estimateurs sont solutions du problème d'optimisation (\*) suivant<sup>1</sup>

$$(*) \quad \text{Trouver } \hat{x} \text{ réalisant le minimum de la fonction « objectif » } \sum_{i=1}^n \rho\left[\frac{r(x)_i}{\theta}\right]$$

où  $\theta$  est un facteur d'échelle : nous le supposons ici connu ou estimé par ailleurs, prenant  $\theta = 1$  sans perte de généralité. Le terme  $r_i$  est la  $i$ -ème composante du vecteur  $r$  tel que  $r = Ax - y$  (le vecteur d'erreurs aléatoires) et  $\rho$  une fonction de coût convexe choisie en fonction de la structure du terme d'erreur. Comme  $A$  n'est pas nécessairement de plein rang, il sera commode d'introduire  $A^+$ , son inverse généralisé de Moore-Penrose obtenu via une décomposition en valeurs singulières (SVD).

Portons ici notre attention sur le cas correspondant à *une contamination d'une structure gaussienne des erreurs*, autrement dit, où  $\rho$  est la fameuse fonction introduite par Huber en 1964 et dite depuis *M-fonction d'Huber*, définie par :

$$\rho_{\text{Hub},c}(w) = \begin{cases} \frac{w^2}{2} & \text{si } |w| \leq c \\ c|w| - \frac{c^2}{2} & \text{sinon} \end{cases}$$

Le coefficient  $c$  est lié au taux de contamination par des valeurs aberrantes de la distribution gaussienne présumée des erreurs.

On notera que pour ce choix, la fonction objectif dans (\*) est certes différentiable, mais que ce problème même non contraint ne possède pas de solution analytique.

<sup>1</sup>On trouvera des caractérisations du domaine des solutions optimales dans [Bougeard Caquineau 1999]. A ma connaissance, les cas de non unicité posent encore des questions ouvertes.

Plusieurs approches itératives ont donc été proposées, mettant en œuvre des algorithmes soit par gradient ou méthode du second ordre, soit encore par partitionnement des résidus (voir [Dutter, 1977] pour une discussion comparative). De plus, un exemple de [Clark, 1985] montre que même pour  $A$  de plein rang, il n'y a pas toujours unicité des solutions contrairement au cas limite ( $c = \infty$ ) associé au problème en « Moindres Carrés ».

Cette difficulté était bien connue pour l'autre cas limite ( $c = 0$ ) depuis que Charnes a montré en 1955 comment utiliser les techniques de *programmation linéaire* pour résoudre le problème (\*) en norme  $L^1$ , encore dit par *Moindres Valeurs Absolues* avec  $\rho = |\cdot|$ . Cet apport théorique fut décisif pour les praticiens en conduisant à des algorithmes basés sur des méthodes de type simplexe comme celui de Barrodale et Roberts (1977). Bien que très souples pour la prise en compte de contraintes *linéaires* sur les paramètres à estimer, de telles procédures demeurent inopérantes pour les contraintes *non linéaires*. Etant par ailleurs fort coûteuses en place mémoire pour le calcul numérique sur grands tableaux de données, d'autres méthodes sont de nos jours recherchées, d'autant que les méthodes de régression en norme  $L^1$  fournissent des estimateurs permettant de minimiser le plus grand biais asymptotique pouvant être provoqué par une contamination de la loi gaussienne.

Les outils de l'analyse convexe vont nous permettre de lever ces difficultés en pouvant intégrer des contraintes non linéaires tout en conduisant à des algorithmes de type primal-dual facilement parallélisables.

## 2. Modélisation du problème et méthodes de résolution

**2.1. Modélisation.** — Introduisons un cadre formel plus apte à la mise en œuvre d'outils de l'analyse non lisse, en considérant les  $M$ -estimateurs comme solutions du problème d'optimisation

$$(R_c) : \left\| \begin{array}{l} \text{Minimiser pour } x \in \mathfrak{X}^m \\ \text{sous la contrainte} \end{array} \right. \sum_{i=1}^n \rho_c((Ax - y)_i) \\ x \in F$$

où  $F$  désigne un convexe fermé d'intérieur relatif non vide (identique à l'espace entier pour un problème non contraint). Ici, la fonction  $\rho_c$  est la *régularisée de Moreau-Yosida* d'indice  $c$  de la fonction *valeur absolue*,

$$\forall w \in \mathfrak{X}, \quad \rho_c(w) = \inf_{z \in \mathfrak{X}} [|z| + \frac{1}{2c}|z - w|^2] = (1/c)\rho_{\text{hub},c}(w)$$

obtenue par *inf-convolution* entre la fonction  $|\cdot|$  et le noyau quadratique  $\frac{1}{2c}|\cdot|^2$ . Ce procédé de régularisation, introduit par Moreau (1964) pour les fonctions et par Yosida pour les opérateurs, possède des propriétés bien connues des spécialistes de l'analyse non lisse [Attouch, 1984].

Un calcul bien mené montre que la fonction  $\rho_c$  est encore égale à un facteur multiplicatif ( $1/c$ ) près à la fameuse M-fonction de coût,  $\rho_{\text{Hub},c}$ , introduite par Huber en 1964. Rappelons qu'ici, le coefficient  $c$  est lié au taux de contamination par des valeurs aberrantes de la distribution gaussienne présumée des erreurs. La relation inf-convolution éclaire en particulier [Bougeard, 1989, 1992] le fait que les deux cas limites ( $c = 0$ ) et ( $c = \infty$ ) sont associés respectivement au problème en norme  $L^1$  et à celui par « Moindres Carrés ».

**2.2. Approche par dualité.** — La résolution numérique d'un problème d'optimisation comme  $R_c$  à  $m$  variables conduit logiquement à mettre en place des algorithmes « directs » en variables *primales*. Le cas *constraint* fait lui intervenir naturellement des variables supplémentaires dites *duales* que l'on peut souhaiter obtenir au lieu des variables primales. Une autre approche algorithmique consiste à chercher directement un couple de solutions primale-duale. Tout dépend du contexte.

Ici, avec notre formalisme, les problèmes  $(R_c)$  relèvent, de par leur structure, du champ de l'optimisation convexe puisque la fonction de coût  $\rho_c$  est convexe. Pour ceux-ci, Rockafellar (1989) a établi une forme d'équivalence entre la théorie minimax lagrangienne et celle de la modélisation par schéma de dualité, où le problème dual est pris au sens de la dualité de Fenchel. Or, en raison de la structure inf-convolution de  $\rho_c$ , on calcule aisément sa fonction duale de Fenchel [Rockafellar, 1970] comme valant

$$\forall t \in \mathfrak{R}, \quad \rho_c^*(t) = \rho_0^*(t) + \frac{c}{2}|t|^2$$

où  $\rho_0^*(t) = \Psi_{B_\infty}(t)$  en notant par  $\Psi_{B_\infty}$  l'indicatrice de la boule unité  $B_\infty$  en norme Chebychev ( $\Psi_{B_\infty} = 0$  sur  $B_\infty$ , et  $\Psi_{B_\infty} = +\infty$  ailleurs).

L'approche par dualité va non seulement nous ouvrir d'autres possibilités de choix algorithmiques mais aussi nous permettre d'en dériver des conditions d'optimalité primal-dual formulées en

$$(**) \quad \text{trouver } \xi \in M \text{ et } p \in M^\perp \text{ tels que } p \in T(\xi)$$

où  $M$  est un sous-espace vectoriel d'un certain Hilbert  $H$ ,  $M^\perp$  son orthogonal et  $T : H \rightrightarrows H$  est un *opérateur maximal monotone* au sens de Brézis (1973). De plus,  $(**)$  peut être résolu par la méthode de l'*Inverse partiel* développée par Spingarn en 1983 en cherchant directement un couple de solutions primale-duale. Elle s'applique sous réserve que l'application proximale  $\text{Prox}_T = (\text{Id} + T)^{-1}$  associée à  $T$ , encore dite *résolvante* de  $T$  puisse être évaluée. Nous l'explicitons brièvement ci-dessous.

**2.3. Méthode proximale de l'inverse partiel.** — Partant d'un point origine arbitraire,  $(\xi^0, p^0) \in (M \times M^\perp)$ , la méthode de Spingarn génère une double suite selon les règles de mise à jour à l'étape  $k : (\xi^k, p^k) \in (M \times M^\perp)$ .

**Étape 1 : pas proximal.** — Trouver  $(\xi'^k, p'^k)$  tels que :

$$\left\| \begin{array}{l} \xi'^k + p'^k = \xi^k + p^k \text{ avec} \\ p'^k \in T(\xi'^k) \end{array} \right. \quad \text{ou encore } \xi'^k \in T^{-1}(p'^k)$$

**Étape 2 : pas de projection sur les sous-espaces  $M$  et  $M^\perp$**

$$\left\| \begin{array}{l} \xi^{k+1} = \text{Proj}_M(\xi'^k) \\ p^{k+1} = \text{Proj}_{M^\perp}(p'^k) \\ \quad = p^k - \xi'^k + \xi^{k+1} \end{array} \right.$$

On vérifie par ailleurs aisément que :

$$\begin{aligned} \xi'^k &= \text{Prox}_T(\xi^k + p^k) \\ p'^k &= \text{Prox}_{T^{-1}}(\xi^k + p^k) \end{aligned}$$

ce qui prouve l'existence et l'unicité des variables intermédiaires  $p'^k, \xi'^k$ . D'où la possibilité d'implémenter la méthode en fait dès que l'une des applications proximales  $\text{Prox}_T$  ou  $\text{Prox}_{T^{-1}}$  est calculable explicitement.

Il se trouve que ce calcul peut être fait dans le contexte étudié. Le lecteur en trouvera le détail dans [Bougeard et Caquineau, 1996]. Nous présentons seulement les grandes lignes des algorithmes dans la section suivante.

### 3. Algorithmes de résolution de type primal-dual

**3.1. Une famille paramétrée pour le problème du M-estimateur d'Huber non contraint.** — Considérons tout d'abord le problème non contraint que nous reformulons en

$$(P_c) : \left\| \begin{array}{l} \text{Minimiser}_{\xi \in \mathbb{R}^n} \Phi_c(\xi) = \sum_{i=1}^n \rho_c[\xi - y]_i \\ \text{avec} \quad \xi \in \text{Image}(A) \end{array} \right.$$

dont le problème dual de Fenchel se formule en

$$(D_c) : \left\| \begin{array}{l} \text{Minimiser}_{p \in \mathbb{R}^n} \Phi_c^*(p) = \frac{c}{2} \|p\|^2 + \langle p, y \rangle + \Psi_{B_\infty}(p) \\ \text{avec} \quad A^t p = 0 \end{array} \right.$$

Ce problème dual est de *type quadratique* : il possède une solution unique par stricte convexité de la fonction objectif (la fonction dont on cherche le minimum). Selon Rockafellar (1989), un vecteur  $\xi$  est optimal pour  $(P_c)$  et un vecteur dual  $p$  est optimal pour  $(D_c)$ , si et seulement si les conditions suivantes d'optimalité  $(O_c)$  sont satisfaites :

$$(O_c) : \left\| \begin{array}{l} \xi \in \text{Image}(A) \\ A^t p = 0 \\ \xi \in \{c p + y\} + N_{B_\infty}(p) \end{array} \right.$$

où on note par  $N_J$  le cône normal au convexe  $J$ .

La résolution de ce schéma de dualité par la méthode de l'inverse partiel [Bougeard et Caquineau, 1996] conduit à une famille d'algorithmes paramétrés par la constante  $c$ , à savoir :

### Algorithme Proximal-Projection A

– *Initialisation de l'algorithme :*

Partant de  $x^0$  choisi arbitrairement et d'un vecteur  $p^0$  tel que  $A^T p^0 = 0$ , on pose  $\xi^0 = Ax^0$ . Alors, considérant le  $k$ -ième itéré  $(\xi^k, p^k)$  vérifiant  $\xi^k \in \text{Image}(A)$  et  $A^T p^k = 0$ , on calcule en deux phases :

– *Phase Proximale (de type dual) :*

Calculer  $p'^k = \text{Proj}_{B_\infty} \left[ \frac{(\xi^k + p^k - y)}{(1+c)} \right]$  c'est à dire :

$\forall j = 1, \dots, n$

$$\begin{cases} \text{if } |(\xi^k + p^k - y)_j| \leq 1 + c & \text{faire } (p'^k)_j = (\xi^k + p^k - y)_j / (1 + c) \\ \text{if } (\xi^k + p^k - y)_j > 1 + c & \text{faire } (p'^k)_j = 1 \\ \text{if } (\xi^k + p^k - y)_j < -(1 + c) & \text{faire } (p'^k)_j = -1 \end{cases}$$

poser  $\xi'^k = (\xi^k + p^k) - p'^k$

– *Phase de Projection (sur  $M = \text{image}(A)$  et son complémentaire  $M^\perp$ )*

$$\begin{cases} \xi^{k+1} = \text{Proj}_M(\xi'^k) \\ p^{k+1} = p^k - \xi'^k + \xi^{k+1} \end{cases}$$

Notons la souplesse d'évaluation en fonction du paramètre  $c$  pour l'utilisateur. L'algorithme basé sur la méthode proximale donne deux suites convergeant globalement respectivement vers une solution du primal et une du dual de sorte qu'il devient possible de déterminer l'ensemble de toutes les solutions primales à partir de la relation  $O_c$  et de conclure quant à une possible *non-unicité*. Il reste alors, dans une dernière étape, à revenir par transformation inverse aux solutions du problème initial. Ici resurgit le problème d'un mauvais conditionnement possible de la matrice  $A$  dont il conviendra de se protéger numériquement.

**3.2. Algorithmes de base pour l'estimation robuste avec contraintes.** — Considérons maintenant le problème avec contraintes

$$(R_c) : \begin{cases} \text{Minimiser}_{x \in \mathcal{D}^m} & \sum_{i=1}^n \rho_c((Ax - y)_i) \\ \text{sous la contrainte} & x \in F \end{cases}$$

où  $F$  désigne un convexe fermé d'intérieur relatif non vide. Dans les applications d'astrométrie, les contraintes additionnelles sur l'espace des paramètres sont de l'un des types suivants :

– des contraintes de signe positif (évaluation d'une masse, d'une parallaxe... en astrométrie),

- des contraintes linéaires de type égalité lorsque les données sont insuffisantes pour « séparer » correctement les inconnues ; ces équations peuvent provenir de théories physiques ou être interprétées mathématiquement (cas de contrainte « moyenne instrumentale » par exemple)
- des relations non linéaires par exemple de type ellipsoïdal en régression « ridge »... etc.

Il suffit alors de penser à introduire deux types de transformation sur le problème original :

- Tout d'abord, poser comme dans le cas non contraint ( $\xi = Ax$ ),
- Puis, introduire une copie de la variable d'origine : ( $\mu = x$ ).

Ceci permet une décomposition de la fonction objectif à minimiser. En effet, en posant  $z^t = (\xi^t | \mu^t)$ ,  $\xi \in \mathfrak{R}^n$ ,  $\mu \in \mathfrak{R}^m$  et  $\mathcal{A}^t = [A^t | I_m]$ , le problème ( $R_c$ ) se reformule dans  $\mathfrak{R}^n \times \mathfrak{R}^m$  comme :

$$(RP_c) : \left\| \begin{array}{l} \text{Minimiser } \Omega(\xi, \mu) = \Phi_c(\xi) + \Psi_F(\mu) \\ \text{avec} \qquad \qquad \qquad z \in \text{image}(\mathcal{A}) \end{array} \right.$$

dont le dual Fenchel devient :

$$(RD_c) : \left\| \begin{array}{l} \text{Minimiser } \Omega^*(p, q) = \Phi_c^*(p) + \Psi_F^*(q) \\ \text{sous contrainte} \qquad \qquad A^t p + q = 0 \end{array} \right.$$

Pour paraphraser les conditions d'optimalité sous la forme voulue (\*\*), on pose  $H = \mathfrak{R}^{n+m}$ , et on considère les sous-espaces complémentaires

$$\left\| \begin{array}{l} M = \{z \in \text{image}(\mathcal{A})\} \\ M^\perp = \text{kernel}(\mathcal{A}^t) = \{Q = (p, q) \in \mathfrak{R}^n \times \mathfrak{R}^m | A^t p + q = 0\} \end{array} \right.$$

ainsi que l'opérateur maximal monotone

$$T = T_c \times T_F \text{ avec } T_c(z) = \partial\Phi_c(\xi) \text{ et } T_F(z) = N_F(\mu) (= \partial\Psi_F)$$

où  $\partial\Phi_c$  désigne le sous-différentiel au sens de l'analyse convexe de la fonction  $\Phi_c$  et  $N_F$  le cône normal au convexe  $F$ . Les applications proximales se calculent comme précédemment sachant de plus que (en notant par  $\text{Proj}$  l'opérateur de projection) :  $\text{Prox}_{\partial\Psi_F} = \text{Proj}_F$ .

L'algorithme du cas contraint devient :

### Algorithme Proximal-Projection (Proxi-Proj) C

- *Initialisation de l'algorithme* :  
Partant de  $(z^0, Q^0) \in M \times M^\perp$ , par exemple  $(0, 0)$ , considérant le k-ième itéré  $(z^k, Q^k)$ , on calcule en deux phases
- *Phase Proximale* :

a) *Partie non contrainte comme dans l'Algorithme A,*

$$\text{Calculer } p'^k = \text{Proj}_{\mathbb{B}_\infty} \left[ \frac{(\xi^k + p^k - y)}{(1 + c)} \right]$$

puis poser  $\xi'^k = (\xi^k + p^k) - p'^k$

b) *Pour la restriction au convexe fermé F,*

$$\text{Calculer } \mu'^k = \text{Proj}_F(\mu^k + q^k)$$

– *Phase de Projection sur les sous-espaces M, M<sup>⊥</sup>*

– *Variables Primales*

$$\begin{aligned} \mu^{k+1} &= (A^t A + I_m)^{-1} \cdot (A^t \xi'^k + \mu'^k) \\ \xi^{k+1} &= A \cdot \mu^{k+1} \end{aligned}$$

– *Variables Duales*

$$\begin{aligned} p^{k+1} &= p^k + \xi^{k+1} - \xi'^k \\ q^{k+1} &= q^k + \mu^{k+1} - \mu'^k \end{aligned}$$

Il suffit ensuite de particulariser le calcul de  $\text{Proj}_F$  au cas particulier que l'on souhaite traiter.

#### 4. Note sur l'implémentation et application

Différentes simulations présentées dans [Bougeard et Caquineau, 1999] montrent un gain très important en vitesse de convergence (nombre d'itérations) lorsque la variable  $c$  croît de 0 à 1. Ceci résulte de la régularisation de Moreau-Yosida qui induit de plus une régularisation dans le problème dual associé.

**4.1. Implémentation sur une architecture parallèle.** — Le lecteur trouvera dans [Bougeard *et al*, 1997] un premier cas d'application de ces algorithmes — dont les calculs sur la Connection Machine CM5 du CNCPST ont été faits en langage CM Fortran (CMF)- à l'étude des données d'observation de petites planètes par le satellite Hipparcos de l'ESA. Rappelons qu'entre 1989 et 1993, 48 petites planètes furent observées par ce satellite. Les paramètres d'intérêt concernaient des corrections aux éléments orbitaux d'asteroïdes observés ainsi que des variables de position entre différents repères de référence. Du à la répartition des observations, ce problème astrométrique à matrice creuse était très mal conditionné (nombre de condition dépassant allégrement  $10^7$ ), pour environ 3000 lignes et 300 inconnues.

Clairement, la difficulté principale qui surgit dans la partie programmation est la façon de résoudre l'étape de projection sur l'image de  $A$  en calculant l'opérateur associé. Pour se protéger d'un mauvais conditionnement de  $A$ , notre approche consiste à

effectuer, en premier lieu, une décomposition en valeurs singulières (SVD) de  $A$ . Le projecteur sur l'image de  $A$  se calcule alors par :

$$\text{Proj}_M = A \cdot A^+ = A \cdot (A^t A)^+ \cdot A^t$$

tandis que la solution (en Moindres Carrés) du système  $b = Ax$  est donné par

$$x^+ = A^+ \times b.$$

Ceci a l'avantage de permettre la comparaison entre l'estimateur des Moindres Carrés  $x^+$  et les solutions obtenues sous différentes valeurs du paramètre  $c$  (en théorie liée au taux de contamination inconnu des données par des valeurs aberrantes). L'utilisateur peut ainsi apprécier la stabilité des estimations obtenues, avec pour cas extrêmes  $c = \infty$  (moindres carrés) et  $c$  voisin de 0 (norme  $L^1$ ) et décider avec succès des données à garder [Bange et al, 1996].

**4.2. Implémentation en environnement UNIX.** — Dans un second temps, les algorithmes ont été implémentés *en routine* sur une station HP9000 travaillant en environnement UNIX au service international IERS (voir annexe) de l'Observatoire de Paris. Un sous-programme y faisant appel a été inclus par R. Ray en alternative aux ajustements par Moindres Carrés dans le *IERS combination software* Fortran. On trouvera dans [Bougeard, 2001], la description de différents outils introduits pour accélérer la convergence.

## 5. Conclusion

Dans cet article, nous avons montré comment les méthodes de l'analyse non lisse conduisent à de nouveaux algorithmes pour résoudre à la fois les problèmes d'ajustement par M-estimateur d'Huber. Ils ont été obtenus via la méthode proximale de l'Inverse Partiel pour les opérateurs maximaux monotones et sont paramétrés par une constante  $c$  en théorie liée au taux de contamination des données par des valeurs aberrantes. Non seulement convergeant globalement vers un couple de solutions primale-duale, leur structure les rend facilement parallélisables. Différentes implémentations ont été réalisées, pour répondre à des applications astrométriques différentes.

## Bibliographie

- ATTOUCH H., 1984 *Variational convergence for functions and operators*, (Pitman Publ.).  
 BANGE J., BEC-BORSENBERGER A., BOUGEARD M.L., CAQUINEAU C., 1996, Lien entre référentiel dynamique et ICRS-Hipparcos, in : *Journées Systemes de Reference spatio-temporels 1996*.  
 BARRODALE, ROBERTS, 1977, An improved algorithm for discrete l1 linear approximation, *SIAM J. Num. Anal.*, 10 (1977), p. 839-848.

- BOUGEARD M.L., 1989, Connection between some statistical estimation criteria, lower  $c_2$  functions and Moreau-Yosida approximates, in : *Bulletin International Statistical Institut 47th Session*, (INSEE Paris press), 1 (1989), p. 159-160.
- BOUGEARD M.L., 1992, Morse theory for favorable classes of nondifferentiable functions in statistical optimization, in : *Transactions 11th Prague Conference (1990) on Information theory*, (Academia Publ House of the Czechoslovak Academy of Sciences, 1992) pp. 291-301.
- BOUGEARD M.L., 2001, Computational Aspects of primal-dual algorithms for estimation with constraints, in *Approximation Optimization and math Economics*, Lassonde ed. Springer-Verlag, p. 59-69.
- BOUGEARD M-L., BANGE, CAQUINEAU C.D., BEC-BORSENBARGER A., 1997, Robust Estimation with application to Hipparcos Minor Planets Data, in *Hipparcos-Venice'97*, vol. ESA-SP402 (july 1997), p. 165-168.
- BOUGEARD M-L., CAQUINEAU C.D., 1999, *Parallel Proximal Decomposition Algorithms for Robust Estimation*, Annals of Op. Research, 90, p. 247-270.
- BREZIS, 1973, *Opérateurs maximaux monotones*, North Holland.
- CANDAHL E., 1995, Applications algorithmiques de l'analyse proximale, memoire D.E.A., University Lyon 1, France (mars 1995)
- CHARNES A., COOPER W., FERGUSON R.O., 1955, *Management Science*, 1, p. 138-151.
- CLARK D.I., 1985, The mathematical structure of Huber's M estimator, *SIAM J. of Scientific and Statistical Computing*, 6, p. 209-219.
- DUTTER R., 1977, *J. of Statistical Computation and simulation*, 15, p. 207-238.
- HAMPEL, RONCHETTI, ROUSSEEW, STAHEL, 1986 *Robust Statistics*(Wiley, New York, 1986)
- LAPLACE P.S., 1793 *Sur quelques points du Systeme du monde, Mémoire Académie des Sciences Paris*.
- MICHELOT C., BOUGEARD M.L., 1994, Duality results and proximal solutions of the Huber-M estimator problem, *J. Optimization Theory and Applications*, 30, p. 203-221.
- MOREAU J.J., 1962, Fonctions convexes duales, points proximaux, C. R. Acad. Sci. Paris, A255, p. 2897-2899.
- ROCKAFELLAR R.T., 1970 *Convex Analysis*, (Princeton University Press).
- ROCKAFELLAR R.T., 1989, *Conjugate duality and optimization*, ed. SIAM.
- RONCHETTI E., 1987, Bounded influence inference, in *Statistical Data analysis based on the  $L^1$  norm*, ed Dodge Y., Elsevier Sci.
- SPINGARN J.E., 1983, Partial inverse of a monotone operator, *Appl. Math. Optimization*, 10 (1983), p. 247-265.
- TICHONOV A., ARSENINE V., 1972, *Méthodes de résolution de problèmes mal posés*, (Mir Moscou, 1972).
- VAN CUTSEM, CAPERA, 1982, *Méthodes non paramétriques*, (Dunod éd.).

### **Annexe. Rotation terrestre et IERS**

L'Observatoire de Paris abrite en ses locaux un centre de recherche lié au Service International de la Rotation Terrestre (IERS) qui fournit au monde entier la position et l'orientation de la Terre dans l'Univers. Ceci concerne la navigation terrestre et spatiale (les sondes planétaires), la détermination d'orbites de satellites artificiels comme ceux de télécommunication.

Observer, interpréter et prévoir positions et mouvements d'objets célestes exige des repères d'espace et de temps bien définis. Le repère céleste international, ICRS, adopté en 1998 par l'Union Astronomique Internationale, IAU, est matérialisé par des directions de quasars extra-galactiques, fig 1, observées en VLBI (interférométrie radio à très longue base). Le repère terrestre international ITRF, quant à lui, est lié aux stations d'observation au sol, fig.2. Par rapport à ces deux repères, le service IERS évalue en permanence les paramètres de la rotation terrestre, EOP, donnant le mouvement de l'axe de rotation et le Temps Universel UT1 résultant de l'angle de rotation de la Terre dans l'espace. Différentes techniques d'observation sont en permanence utilisées : l'interférométrie radio à très longue base VLBI, la télémétrie laser sur satellites artificiels SLR ou récepteurs lunaires LLR, les systèmes satellitaires GPS, DORIS. La Table 1 montre la part de ces techniques et la Table 2 leur évolution en précision. Au stade final, le Service EOP-IERS de l'Observatoire de Paris est en charge de l'estimation moyenne des EOP par *combinaison* des séries de mesures propres à chaque technique et de leur prédiction, via des théories et modélisations astrométriques, géodésiques et mathématiques.

TABLE 1. *Contribution des techniques d'observation pour l'orientation terrestre et les repères fondamentaux. Le nombre de \* sert d'indicateur pour le niveau de contribution en précision et volume des mesures (source : IERS-Paris Observatory)*

PRODUITS	LLR	VLBI	SLR	GPS	DORIS
<hr/>					
ITRF repère terrestre					
- position du géocentre	*		***	**	*
- Tectonic plate motion		***	**	***	**
<hr/>					
ICRF repère Céleste Extragalactique		***			
Lien au système solaire	***				
<hr/>					
Terre					
- Precession-nutation	**	***	*	*	
- Temps Universel	*	***			
<hr/>					
Rotation Terrestre					
- High-frequency UT		***	**	***	
- Polar motion (x, y)		***	***	***	*

TABLE 2. *Observation des EOP et Evolution temporelle des techniques en précision; PM=polar motion (x, y); CP=celestial pole coordinates ( $d\psi \sin \varepsilon$ ,  $d\varepsilon$ ), UT=temps universel UT1 (source ; IERS -Paris Observatory)*

Période	SLR	VLBI	GPS	resol.	Précision
1988-92	UT	*	***		0.00005s
	PM	**	**		0.0005"
	CP		****	jours	0.0005"
1992-95	UT	*	***	*	0.00003s
	PM	***	***	***	0.0003"
	CP		****	jour	0.0003"
1996-99	UT	*	***	*	0.00002s
	PM	***	***	***	0.0001"
	CP		****	jour	0.0003"

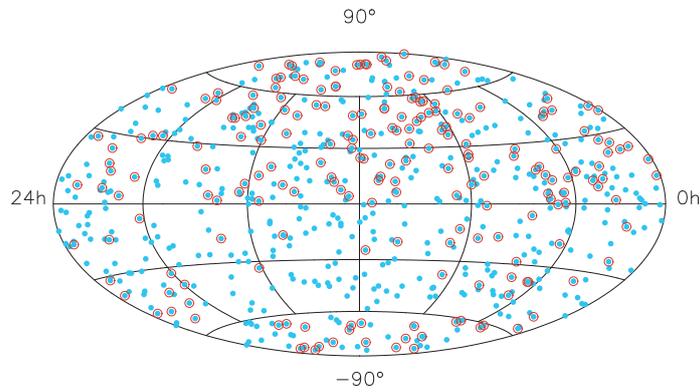


FIGURE 1. Repère céleste international ICRF-1998 matérialisé par les directions de 610 sources extra-galactiques (cf. *the Astronomical Journal*, 1998, 116, 516)

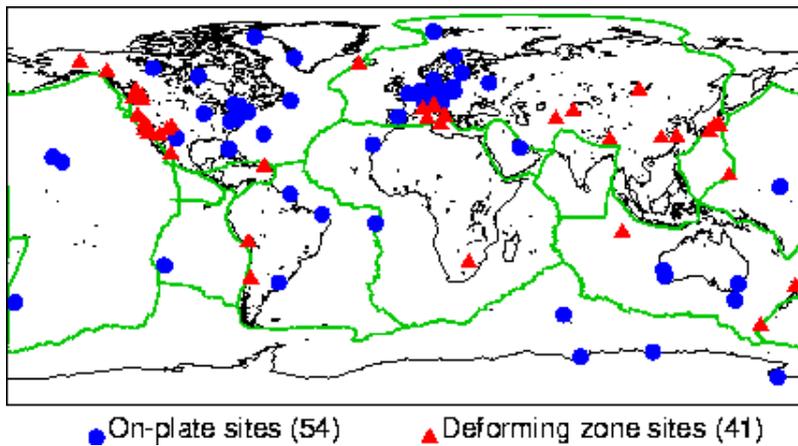


FIGURE 2. Repère terrestre international ITRF-2000 matérialisé par une sélection de 54 sites primaires d'observation complétée par 41 sites localisés en zone de déformation géophysique (plaques tectoniques) (source : Z. Altamimi <http://lareg.ensg.ign.fr/ITRF/>\*)

*Mireille L. Bougeard*

[1] Université LYON.1-IGD, UFR Mathématique, 69622 Villeurbanne Cedex, France.

[2] Observatoire de Paris, IERS, UMR8630, 61 av. de l'Observatoire, 75014 Paris, France.

*E-mail* : bougeard@hpopa.obspm.fr