ROBUST ESTIMATION WITH APPLICATION TO HIPPARCOS MINOR PLANET DATA

M.L. Bougeard[1,2], J.-F. Bange[3], C. Caquineau[1], A. Bec-Borsenberger[3]

[1]University UCBL-Lyon1, UFR Math, F-69622 Villeurbanne Cedex, France
[2]URA1125 CNRS, Observatoire de Paris, 61 Av. Observatoire, 75014 Paris, France
[3]Bureau des Longitudes, 77 av. Denfert-Rochereau, F-75014 Paris, France

ABSTRACT

Forty eight minor planets have been observed by the Hipparcos satellite. Pooling their precise position may allow to investigate a possible rotation between the dynamical reference frame and the ICRS-Hipparcos Reference system. Due to the repartition of the observations, ill-conditioning of the global system and outliers simultaneously occur. So, a direct least-squares fit is potentially misleading and we resort to the use of robust statistical alternatives. While it is now clear that no single robust regression is best, the L1 and Huber-M estimators are currently attracting attention under the occurence of contaminated Gaussian errors. Here, we present new algorithms based on the *Spingarn Partial Inverse* proximal decomposition method for L1 and Huber-M estimation that take into account both primal and dual aspects of the optimization problem. The result is a family of highly parallel algorithms attractive for large scale problems. Implemented on the Connection Machine CM5, their computational performances on the data under study are reported and discussed.

Key words: space astrometry; minor planets; robust estimation; L1 fit; Huber-M regression; parallel algorithms.

1. INTRODUCTION

Between 1989 and 1993, 48 minor planets were observed by the ESA astrometric satellite Hipparcos. The observational data were analysed by both consortia FAST and NDAC whose reductions differ in the treatment of the modulated light signal of an observed body. So, the Hipparcos final catalogue includes two sets of results: the FAST one concerns 2657 observations of the 48 minor planets, whereas the NDAC one contains 2837 observations of the same set of minor planets. These data are related to abscissa of observed bodies on different reference great circles.

The comparison between observation and theory leads to an O-C formulation expressed in terms of astrometric parameters of interest, as described by (Bec-Borsenberger et al. 1995). By neglecting the mass perturbation effect during minor planet close encounters, the condition equations can be conveniently written as:

$$\mathbf{O} - \mathbf{C} = \mathbf{M}\Delta\vec{u^0} + \mathbf{B}\vec{\theta}, \quad \text{with } \vec{\theta} = \overrightarrow{\theta_0} + \vec{w}(t - t_0) \quad (1)$$

where O represents the observed position of the minor planet in the Hipparcos reference frame (best optical counterpart of ICRS reference frame), whereas the C value is computed by numerical integration of the equations of motion; JPL-DE200 ephemerides are used for the of the major planets (Standish 1990). The astrometric parameters to be estimated are $\Delta\vec{u^0}$ and $\vec{\theta}$. They stand respectively for the corrections to the minor planet orbital elements and to the initial rotation and spin parameters between the two reference frames; **M**, **B** are the related design matrices. Here, we focus our attention on the quality of any possible estimation of the parameter $\vec{\theta}$. For an extended analysis including some additional minor planet mass corrections, see (Bange & Bec-Borsenberger 1997).

To begin, system (1) was solved by using the Least Square (LS) criterion. It provides highly unstable solutions, in the sense that small perturbations of the data induces large changes in the LS estimate. Futhermore, the observations of 20-Massalia and 27-Eutherpe Minor planets were noticed as highly influential on the LS solution. In addition, Variance Inflation Factors (VIF) were computed (Bougeard et al. 1996, Table 2): their very large values assert that many factors are involved in strong multicollinearity. To quantify the level of ill-conditioning, Singular Value Decomposition (SVD) were computed for each following design matrix (subscript 46 means that both Minor planets, 20-Massalia and 27-Eutherpe, are dropped):

- FAST48: 2657 observation-rows, 294 unknowns;
- FAST46: 2560 observation-rows, 282 unknowns;
- NDAC48: 2837 observation-rows, 294 unknowns;
- NDAC46: 2737 observation-rows, 282 unknowns.

Results are to be found in Table 1. At this stage, the LS regression through any numerical approach (singular values or QR decomposition) is statistically misleading especially in terms of tests and error variance estimates. A great deal of statistical work has been devoted to the construction of good regression estimators when influential and/or collinearity is present in the model. In the sequel, we consider alternatives based on robust techniques.

2. TOWARDS ROBUST ESTIMATION

2.1. Statistical Background and Modelisation

In the past few years, robustness is one problem that has been given much attention in statistical literature. There has been a significant increase in the interest concerning robust estimation methods as an alternative to the LS fit. Although it is now clear (see for instance Hampel et al. 1986) that no single robust procedure is the best (depending on the mean square error or other adequate criteria), the L1 (least absolute value) and the Huber-M estimators are currently attracting considerable attention when the errors have a contaminated Gaussian or thick-tail distribution. They are often recommended in practice as a starting point for iteratively weighted least-squares procedures.

In this framework, finding M-estimators (whose name derive from their similarity to maximum-likelihood estimate) consists in solving the optimization problem:

$$(R_c) \text{ Find } \widehat{x} \in \text{ argmin } (\sum_{i=1}^{n} \rho_c((Ax-b)_i)$$

where r_i is the ith component of the vector $r = Ax-b$ and ρ is some convex cost function. The vectors r, x and b are of dimension n, m and n ($n \geq m$) respectively and the matrix A is of size $(n \times m)$. Since A is not necessarily of full rank, the symbol A^+ will stand for its Moore-Penrose inverse. Here, the convex cost function ρ_c stands for the c-Moreau-Yosida regularisation of the absolute value function as defined by:

$$\forall w \in \Re, \quad \rho_c(w) = inf_{y \in \Re}[|y| + \frac{1}{2c}|y-w|^2])$$

still equal, up to a multiplicative factor $1/c$, to the Huber-M cost function introduced in his famous Paper published in the 1964 Annals of Statistics:

$$\rho_{\text{Hub,c}}(w) = \left\| \begin{array}{ll} \frac{w^2}{2} & \text{if } |w| \leq c \\ c|w| - \frac{c^2}{2} & \text{elsewhere} \end{array} \right.$$

This estimator was proved by Huber to be convenient when the errors are 'contaminated Gaussian'.The tuning constant c depends on the level of contamination (by outliers). The limiting case ($c = 0$) corresponds to an L1 fit, whereas ($c = \infty$) is simply a least-square fit.

Nevertheless, problem R_c cannot be solved directly, since there is no simple analytical representation of the solutions. Finding efficient algorithms to produce such estimates in the case of large data sets is still a field of active research. In (Bougeard & Caquineau 1996), new algorithms were derived, based on the Spingarn Partial Inverse proximal approach that takes into account both primal and dual aspects of the M-estimation problem. We briefly recall the main lines of the approach.

2.2. Re–Parametrization of the Problem

To bring into evidence the linear subspace Range(A), we introduce a reparametrization of problem R_c as:

$$(P_c) : \left\| \begin{array}{ll} \text{Minimize}_{\xi \in \Re^n} & \Phi_c(\xi) = \sum_{i=1}^n \rho_c[\xi - b]_i \\ \text{subject to} & \xi \in \text{Range}(A) \end{array} \right.$$

Setting $c = 0$ provides the $L1$-fit formulation. Since ρ_c is the inf-convolution between the absolute value function and the Moreau-Yosida regularization kernel $\frac{1}{2c}|.|^2$, the Fenchel dual of (P_c) takes the form:

$$(D_c) : \left\| \begin{array}{ll} \text{Minimize}_{p \in \Re^n} & \frac{c}{2}||p||^2 + < p, b > \\ \text{subject to} & A^t p = 0 \text{ and } \gamma_\infty(p) \leq 1 \end{array} \right.$$

where γ_∞ stands for the Chebychev norm. Whenever $c > 0$, due to the strict convexity of the objective function to be minimized, D_c admits an unique optimal solution. As c goes to zero, convergence is expected towards a solution of smallest norm of D_0. For $c = 0$, D_0 is a Linear Programming problem (LP) that has become the basis for some of the best L1-algorithms. The advantage of our formulation is to take into account at the same time the L1 and Huber-M estimations.

According to (Rockafellar 1970), a vector ξ is an optimal solution for (P_c) and a vector p is optimal for (D_c), if and only if the following optimality conditions (O_c) hold:

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

Where N_{B_∞} stands for the Normal cone to the Chebychev unit ball B_∞. Given ξ, the third condition of (O_c) belongs to the variational inequality class where both ξ and p are unknown. To solve the above duality scheme, let us consider the Spingarn proximal approach that takes into account both primal and dual aspects.

2.3. Spingarn Proximal Method

By letting M=Range(A), the above optimality conditions can be expressed as:

$$(**) \text{ find } \xi \in M \text{ and } p \in M^\perp \text{ such that } p \in T(\xi) \tag{2}$$

where M is a linear subspace of a certain Hilbert space H, $M^\perp$ denotes the orthogonal complementary of M and $T : H \longrightarrow H$ is a maximal monotone multifunction in the sense of Brezis (1973). A convenient resolution method to deal with this general duality scheme is the partial inverse method developed by Spingarn (1983). It acts as follows:

Table 1. Singular Value Decomposition of global design matrix as given by their singular eigenvalues.

eigenvalues	FAST 48	FAST 46	NDAC 48	NDAC 46
$\lambda_1 = \lambda_{\max}$	10405.5	10141.4	10519.1	10290.6
λ_2	9315.4	9249.8	9348.8	9433.6
λ_3	7965.4	7827.3	8207.7	8041.9
λ_4	39.3	39.2	42.7	42.7
λ_5	39.0	38.8	40.4	40.0
λ_6	34.7	34.7	37.0	35.9
λ_7	32.4	32.3	36.1	34.7
$\lambda_{\min}$	4×10^{-5}	4×10^{-5}	14×10^{-5}	14×10^{-5}
	23 eigenv. < 0.1	23 eig. < 0.1	20 eig. < 0.1	20 eig. < 0.1
Condition number	26×10^7	25×10^7	7×10^7	7×10^7

- *initialization stage* : Starting from an arbitrary point $(\xi^0, p^0) \in (M \times M^\perp)$, Spingarn's method generates a sequence according to the updating rules at step k: $(\xi^k, p^k) \in (M \times M^\perp)$
- *stage 1: proximal steps*: find (ξ'^k, p'^k) such that:

$$\left\| \begin{array}{ll} \xi'^k + p'^k = \xi^k + p^k & \text{and} \\ p'^k \in T(\xi'^k) & \text{or equiv. } \xi'^k \in T^{-1}(p'^k) \end{array} \right.$$

- *stage 2: projections steps onto the subspaces M,$M^\perp$*

$$\left\| \begin{array}{lll} \xi^{k+1} & = & Proj_M(\xi'^k) \\ p^{k+1} & = & Proj_{M^\perp}(p'^k) \end{array} \right.$$

The method can be viewed as producing:

$$\xi'^k = Prox_T(\xi^k + p^k), \quad p'^k = Prox_{T^{-1}}(\xi^k + p^k)$$

that proves the existence and uniqueness of the intermediate variables p'^k, ξ'^k. Consequently, Spingarn's method can be computer-implemented provided either the proximal mapping $Prox_T = (Id + T)^{-1}$ associated with T or with its inverse can be evaluated. Because the algorithm is a special instance of the proximal process and provided that the optimal solution set is not empty, the algorithm was proved to be always convergent.

2.4. Proximal-Projection Algorithms for Robust Estimation

Transposing the Spingarn approach by evaluating the specific proximal mappings related to the duality scheme O_c (for details see: Bougeard & Caquineau 1996) leads to the following computational algorithms parametrized by the tuning constant $c \geq 0$. The Proximal-Projection c-algorithm consists of:

- *Initialization*: Start from a vector x^0 randomly chosen, a vector p^0 such that $A^t p^0 = 0$ (for instance $x^0 = 0$ and $p^0 = 0$) and assign $\xi^0 = Ax^0$. Then, given the kth iterate (ξ^k, p^k) satisfying respectively to $\xi^k \in \text{range}(A)$ and $A^T p^k = 0$, evaluate the next iterate using the following lines.
- *Proximal Phase (in dual form)*: Set $m = 1 + c$ and $z_k = \xi^k + p^k - b$, then:

calculate $p'^k = Proj_{B_\infty}[\frac{(\xi^k + p^k - b)}{(1+c)}]$that is:
$\forall j = 1, \cdots, n$

$$\left\{ \begin{array}{llllll} if & |(z_k)|_j & \leq & m & set & (p'^k)_j = (z_k)_j/m \\ if & (z_k)_j & > & m & set & (p'^k)_j = 1 \\ if & (z_k)_j & < & -m & set & (p'^k)_j = -1 \end{array} \right.$$

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

- *Projection Phase (onto M = range(A) and its complement $M^\perp$)*

$$\left\| \begin{array}{lll} \xi^{k+1} & = & Proj_M(\xi'^k) \\ p^{k+1} & = & p^k - \xi'^k + \xi^{k+1} \end{array} \right.$$

It can be noticed that the process leads to quite simple computational steps and allows parallel computations that can be of importance for large-scale problems. Moreover, it simultaneously generates two sequences, (ξ^k) and (p^k), globally converging respectively to a primal and a dual solution, whenever the set of solutions is nonempty. In addition, since the algorithm provides a dual solution, this allows us to deduce (from relation O_c) all the optimal-primal solutions and to conclude about a possible *non-uniqueness* of the estimate so far performed (Michelot & Bougeard 1994). Once regression coefficients obtained for the reparametrized model, a transformation is made back to the original model. It is not true, however that we totally overcome the inherent ill-condition of the initial astrometric model, due to this last transformation.

2.5. Implementation for the CM5 Family

The above algorithm was implemented on the CNCPST CM5 that is a parallel computer. The algorithm is clearly data regular so we chose to use the global data-parallel CM Fortran (CMF) language. For details, we refer to the 1996 Paper by the two first authors. The CM5 is a collection of Vector Units (512 VU at the CNCPST) hooked up to a very fast network. A sparc processor drives 4 VUs leading to 128 processing nodes. The CM-runtime system splits the giant matrices (or vectors) by partitioning the matrix

Table 2. Solutions FAST and NDAC (θ in mas, ω in mas/yr).

	FAST 48		FAST 46			
c:	∞	0	∞	0	0.01	0.1
	(svd − ls)	($L1$)	(svd − ls)	($L1$)		
θ_1	19.8	1.9	3.7	1.3	1.2	1.5
θ_2	8.2	-12.1	-11.9	-12.3	-12.3	-12.3
θ_3	-41.3	-15.0	-12.6	-13.3	-13.3	-13.5
ω_4	2.7	3.5	4.3	3.7	3.8	3.8
ω_5	-18.4	-8.5	-9.5	-8.0	-8.0	-8.2
ω_6	4.0	14.3	14.9	12.6	12.7	13.1

	NDAC 48		NDAC 46			
c:	∞	0	∞	0	0.01	0.1
	(svd − ls)	($L1$)	(svd − ls)	($L1$)		
θ_1	0.6	3.7	6.5	3.7	3.7	3.9
θ_2	15.7	-12.2	-25.2	-17.1	-17.2	-16.6
θ_3	-48.1	-16.9	7.1	-5.1	-5.0	-5.1
ω_4	17.3	1.4	4.2	2.7	2.7	2.4
ω_5	-20.9	-8.4	-12.2	-9.9	-9.9	-10.2
ω_6	19.1	14.3	17.2	16.4	16.4	16.7

data across the VU pool making none of them idle. All the computational resources are used if we use large data sets. The main difficulty is numerically solving the 'projection step' by computing the projector onto the range of matrix A. To be protected against all possible ill-conditionning of the A design matrix, our approach is first to perform a Singular Value Decomposition (SVD).

Then, the projector onto the range of A is given by $Proj_M = A.A^+ = A.(A^tA)^+.A^t$ whereas the generalized least-squares solution of the system of linear equations $b = Ax$ is given by $x^+ = A^+.b$. This will allow to compare the behaviour of the solutions for different c-fits to the SVD least-squares estimate x^+. We are now ready to deal with the astrometric problem under study.

3. APPLICATION TO HIPPARCOS MINOR PLANET DATA

The astrometric system (1) was solved: by a SVD least-squares fit ($c = \infty$); and by computation of M-estimations for different c-values. According to different values of the c parameter, it is possible to evaluate the sensitivity of the adjustment with respect to potential outliers. As shown in Table 2, least-squares solutions are 'unstable' (in the sense where they significantly differ for 48 and 46 minor planets), whereas solutions obtained with $c = 0$ and c near 0 are coherent. Thus, the least-squares solution for the rotation parameters is of poor confidence. This result was expected after the study of the system condition numbers. The L_1 solution give some more 'stable' results. However, an extensive study of outliers is needed in order to improve the estimation.

4. CONCLUSION

We have been concerned with the evaluation of a potential rotation between the dynamical reference frame and the ICRS-Hipparcos Reference system from the Hipparcos observations of minor planets. Due to the repartition of the observations, influential points and collinearity were proved to be present in the astrometric model. So, alternatives to LS fit have been considered based on robust estimation.

New algorithms were defined in Section 2 for solving both L1 and Huber-M estimation problems. As special instances of the partial inverse proximal method, they take into account at the same time, the primal and dual structures of the optimization problem; generated sequences are globally convergent. Their implementation on the Connection Machine CM5 has been performed in such a way that allows a direct comparaison to the SVD LS solution in case of ill-conditioning of the design matrix. These algorithms refer explicitly to the tuning constant c that depends on the level of contamination by outliers; its true value is unfortunately unknown in practice. Section 3, introductory experiments were presented. The FAST46, NDAC46 solutions were proved to be more 'stable' and meaningful. Additional work is in progress in this encouraging direction.

REFERENCES

Bange J-F., Bec-Borsenberger A., 1997, Determination of the Masses of Minor Planets, ESA SP–402, this volume

Bec-Borsenberger A., Bange J-F., Bougeard M-L., 1995, A&A, 304, 176

Bougeard, M. L., Bange J.-F., Mahfouz M., Bec-Borsenberger A., 1996, Dynamics and astrometry of the solar system, IAU symposium 172 Proc., eds S. Ferraz-Mello et al., Kluwer Acad. Publ., 447

Bougeard M-L., Caquineau C.D., 1996 Parallel Primal-Dual Algorithms for Robust Estimation, preprint Univ. Lyon1-UCBL, submitted Annals of O.R.

Brezis, 1973, Opérateurs maximaux monotones, Math. Studies North Holland

Hampel P.W., Ronchetti E.M., Rousseeuw, P.J., Stahel W.A., 1986, *Robust Statistics*, Wiley New York

Michelot C., Bougeard M.L., 1994, Duality results and proximal solutions of the Huber-M estimator problem, J. Optimization Theory and Appl., 30, 203

Rockafellar R.T.,1970, *Convex Analysis*, Princeton University Press

Spingarn J.E.,1983, Partial inverse of a monotone operator, J.Appl. Math. Optimization, 10, 247

Standish E. M. Jr, 1990, A&A, 233, 252

Thinking Machine, 1994, CM-Scientific Subroutines Library for CM Fortran, Volumes I and II