

Moindres carrés robustes

Hervé Lebret⁽¹⁾ et Laurent El Ghaoui⁽²⁾

⁽¹⁾ENSTA 32 bd Victor
75739 PARIS CEDEX 15
email: lebret@ensta.fr

⁽²⁾email: elghaoui@ensta.fr

RÉSUMÉ

Nous considérons des problèmes de moindres carrés dans lesquels les données dépendent de manière affine de perturbations inconnues mais bornées. Nous calculons l'erreur résiduelle pire-cas minimale par optimisation convexe. Le problème est ainsi résolu en temps polynomial, en utilisant par exemple des méthodes de points intérieurs. Nous montrons que cette approche conduit à la régularisation de problèmes mal-conditionnés. Dans le cas où les perturbations entrent dans les données de manière rationnelle, nous montrons comment calculer et minimiser des bornes supérieures sur les résidus pire-cas.

ABSTRACT

We consider least squares problems in which data matrices depend affinely on an unknown-but-bounded perturbation vector. We compute the smallest worst-case residual error using convex optimization. The problem can thus be solved in polynomial-time, using for instance interior-point methods. We show that the approach can be used to regularize ill-conditioned problems. For the case when the perturbation vector enters rationally in the data matrices, we show how to compute, and minimize, upper bounds on the optimal worst-case residual.

1 Exposé du problème

De nombreux problèmes en traitement de signal (et ailleurs) se formulent comme un problème de moindres carrés :

$$\text{minimiser } \|Ax - b\|. \quad (1)$$

En pratique, les données (dans ce cas, la matrice A et le vecteur b) sont issues de mesures physiques, donc sujettes à des erreurs. Il se trouve que la solution du problème (1) peut être très sensible à de telles erreurs, surtout si A est "presque" singulière. Comment tenir compte de ces erreurs, et si possible "régulariser" le problème ?

On peut par exemple supposer que A, b sont remplacées par $A + \Delta A, b + \Delta b$, où $\Delta = [\Delta A \ \Delta b]$ est une matrice quelconque, inconnue mais bornée ; en d'autres termes, on suppose une incertitude "non structurée" sur les données du problème. En face de telles incertitudes, on peut évaluer la qualité d'une variable x par le résidu pire-cas qu'il engendre,

$$r_{\text{pire}}(x) = \max_{\|\Delta\| \leq \rho} \|(A + \Delta A)x - (b + \Delta b)\|, \quad (2)$$

où $\rho > 0$ est un paramètre qui reflète notre connaissance sur la taille des perturbations affectant les données du problème. On cherche alors la variable x qui minimise le résidu pire-cas r_{pire} . L'interprétation d'une telle solution "robuste" est la suivante : parmi tous les choix de x possibles, on prend celui qui tolère au mieux les incertitudes.

Par ailleurs, on peut généraliser ce problème au cas où les incertitudes $\Delta A, \Delta b$ sont structurées (par exemple, si certains éléments de la matrice A sont parfaitement connus, et d'autres incertains). Ce type de problème intervient notamment en

identification robuste. Si $\Delta = [\Delta A, \Delta b]$ a une structure connue, par exemple Toeplitz, le résidu pire-cas précédent peut être très conservatif. On peut alors considérer la structure suivante : soit $A_0, \dots, A_p \in \mathbf{R}^{n \times m}$, $b_0, \dots, b_p \in \mathbf{R}^n$; définissons pour $\delta \in \mathbf{R}^p$,

$$\mathbf{A}(\delta) \triangleq A_0 + \sum_{i=1}^p \delta_i A_i, \quad \mathbf{b}(\delta) \triangleq b_0 + \sum_{i=1}^p \delta_i b_i. \quad (3)$$

On définit alors pour $\rho \geq 0$, et $x \in \mathbf{R}^m$, un nouveau résidu pire-cas structuré :

$$r_S(\mathbf{A}, \mathbf{b}, \rho, x) \triangleq \max_{\|\delta\| \leq \rho} \|\mathbf{A}(\delta)x - \mathbf{b}(\delta)\|, \quad (4)$$

On dit alors que x est solution des moindres carrés robustes structurés, si x minimise le résidu pire cas r_S .

2 Solution des problèmes

2.1 Problèmes non-structurés

Dans le cas non structuré, le problème de minimisation de (2) est un problème d'optimisation convexe dont les propriétés sont très riches (voir [1, 2]). On peut en particulier montrer pour $\rho = 1$ qu'il est équivalent à un problème de régularisation optimale :

$$\text{minimiser } \lambda \text{ sujet à } \|Ax - b\| \leq \lambda - \tau, \quad \left\| \begin{bmatrix} x \\ 1 \end{bmatrix} \right\| \leq \tau. \quad (5)$$

Un tel problème est appelé problème de programmation au second ordre. Pour plus de détails voir l'article [5].

2.2 Problèmes structurés

Pour le cas structuré, en supposant à nouveau que $\rho = 1$, on a montré dans [2] que le problème de minimisation du résidu dans l'équation (4) peut être résolu par le problème de programmation semi-définie (voir [7])

$$\text{minimiser } \lambda \text{ sujet à } \begin{bmatrix} \lambda - \tau & 0 & (A_0x - b_0)^T \\ 0 & \tau I & M(x)^T \\ A_0x - b_0 & M(x) & I \end{bmatrix} \geq 0,$$

où $M(x) \triangleq \begin{bmatrix} A_1x - b_1 & \dots & A_px - b_p \end{bmatrix}$. Pour plus de détails, consulter l'article [2].

2.3 Problèmes linéaire-fractionnels

Soit \mathcal{D} un sous-espace vectoriel de $\mathbf{R}^{N \times N}$, $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, $L \in \mathbf{R}^{m \times N}$, $R_A \in \mathbf{R}^{N \times n}$, $R_b \in \mathbf{R}^N$, $D \in \mathbf{R}^{N \times N}$. Pour toute matrice $\Delta \in \mathcal{D}$ telle que $\det(I - D\Delta) \neq 0$, on définit

$$\begin{bmatrix} \mathbf{A}(\Delta) & \mathbf{b}(\Delta) \end{bmatrix} = \begin{bmatrix} A & b \end{bmatrix} + L\Delta(I - D\Delta)^{-1} \begin{bmatrix} R_A & R_b \end{bmatrix}.$$

Pour $x \in \mathbf{R}^n$, on définit également le résidu pire-cas par

$$r_{\mathcal{D}}(\mathbf{A}, \mathbf{b}, \rho, x) \triangleq \max_{\Delta \in \mathcal{D}, \|\Delta\| \leq \rho} \|\mathbf{A}(\Delta)x - \mathbf{b}(\Delta)\| \quad (6)$$

si $\det(I - D\Delta) \neq 0$, (l'infini sinon), ce qui généralise les deux cas précédents. On introduit ensuite les sous-espaces suivants :

$$\begin{aligned} \mathcal{B} &\triangleq \{B \in \mathbf{R}^{N \times N} \mid B\Delta = \Delta B \text{ for every } \Delta \in \mathcal{D}\}, \\ \mathcal{S} &\triangleq \{S \in \mathcal{B} \mid S = S^T\}, \quad \mathcal{G} \triangleq \{G \in \mathcal{B} \mid G = -G^T\}. \end{aligned} \quad (7)$$

Calculer le résidu pire-cas est un problème NP-complet. On peut par contre calculer des *bornes supérieures* sur ce résidu. Le résultat suivant est démontré dans [1].

Théorème 2.1 — Une borne supérieure sur le résidu pire-cas peut-être obtenue par programmation semi-définie :

$$\inf_{S, G, \lambda, x} \lambda \text{ sujet à } S \in \mathcal{S}, \quad G \in \mathcal{G}, \quad (8)$$

et

$$\begin{bmatrix} \lambda I - LSL^T & -L(SD^T - G) & Ax - b \\ -(DS + G)L^T & \Sigma & R_Ax - R_b \\ (Ax - b)^T & (R_Ax - R_b)^T & \lambda \end{bmatrix} > 0. \quad (9)$$

avec $\Sigma = S + DG - GD^T - DSD^T$. La borne supérieure est toujours exacte lorsque $\mathcal{D} = \mathbf{R}^{N \times N}$. Définissons

$$\Theta \triangleq \begin{bmatrix} \lambda I - LSL^T & -LSD^T + LG \\ -DSL^T + G^T L^T & S + DG - GD^T - DSD^T \end{bmatrix}. \quad (10)$$

et

$$\mathcal{A} = \begin{bmatrix} A \\ R_A \end{bmatrix} \quad \mathcal{B} = \begin{bmatrix} b \\ R_b \end{bmatrix}$$

Si $\Theta > 0$ à l'optimum, la borne est aussi exacte. La solution optimale x est alors unique et donnée par

$$x = (\mathcal{A}^T \Theta^{-1} \mathcal{A})^{-1} \mathcal{A}^T \Theta^{-1} \mathcal{B}. \quad (11)$$

où R_A est de rang plein.

3 Exemples numériques

3.1 LS, TLS and RLS

Par un exemple simple, nous montrons ici les propriétés respectives des moindres carrés (LS), des moindres carrés totaux (TLS, voir [4]) et robustes (RLS). Soit

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix}^T, \quad b = \begin{bmatrix} 3 & 7 & 1 & 3 \end{bmatrix}^T.$$

Les parties gauches et droites de la figure 1 montrent les quatre points (A_i, b_i) par les signes '+'. La régression linéaire est indiquée par les droites en trait plein alors que les moindres carrés totaux sont indiqués par les pointillés et les moindres carrés robustes par les lignes discontinues pour $\rho = 1, 2$. La figure de gauche correspond à $[A + \Delta A, b + \Delta b]$ alors que celle de droite correspond à des incertitudes sur A seulement $[A + \Delta A, b]$. Sur chaque figure, les points pire-cas sont indiqués par 'o' pour $\rho = 1$ et '*' pour $\rho = 2$. La figure montre le caractère "précis" du TLS et robuste de RLS relativement aux moindres carrés standards.

3.2 RLS et régularisation

Considérons par exemple le problème de moindres carrés tel que (1), où la matrice A dépend linéairement d'un paramètre α . Pour $\alpha = 5$, A est singulière. La figure 2 montre les résidus optimaux en fonction de α pour différents degrés de perturbation mesurés par ρ . La solution du problème (1), qui correspond aussi à la minimisation du résidu (2) pour $\rho = 0$, est discontinue par rapport à α . D'un point de vue pratique, cette solution n'est pas donc satisfaisante. On constate que lorsque ρ augmente, le pic de discontinuité de la solution classique est lissé.

3.3 Interpolation robuste

Pour des entiers donnés $n \geq 1, k$, on cherche un polynôme de degré $q-1$ $p(t) = x_1 + \dots + x_n t^{n-1}$ qui interpole des points $(a_i, b_i), i = 1, \dots, k$, c'est à dire $p(a_i) = b_i, i = 1, \dots, k$. Nous supposons que les b_i sont connus alors que les a_i sont incertains mais bornés,

$$a_i(\delta) = a_i + \delta_i, \quad i = 1, \dots, k,$$

avec $\|\delta\|_{\infty} \leq \rho$. Nous cherchons alors un interpolant robuste, c'est à dire un minimiseur de

$$\max_{\|\delta\|_{\infty} \leq \rho} \|\mathbf{A}(\delta)x - b\|,$$

où $\mathbf{A}(\delta)$ a une structure de matrice de Vandermonde. Comme il est démontré dans [1], ce problème peut être formulé comme un problème linéaire-fractionnel introduit en section 2.3.

En figure 3, nous montrons les résultats pour $n = 3, k = 1, \rho = 0.2$, et

$$\{(a_i, b_i), i = 1, \dots, 3\} = \{(1, 1), (2, -0.5), (4, 2)\}.$$

La solution par moindres carrés (LS) est très précise (le résidu est nul, l'interpolation est exacte) mais a un résidu pire-cas (prédit) de 1.7977. La solution robuste (RLS) est un compromis entre la précision (un seul point est interpolé

exactement avec un résidu nominal de 0.8233) et la robustesse (avec un résidu pire-cas inférieur à 1.1573). Quand $\rho \rightarrow \infty$, le polynôme obtenu par RLS est de plus en plus horizontal, ce qui est en accord avec le fait que nous permettons des variations de a uniquement. A la limite, le polynôme d'interpolation est la droite $p(t) = 0$.

4 Conclusion

Ce nouveau concept d'approche robuste des moindres carrés se généralise aisément à de nombreux problèmes d'optimisation (voir [6]). Il se généralise également à de nombreux problèmes sur les systèmes linéaires (voir [3]). Il nous semble donc que la robustesse pour des problèmes incertains aura des applications très intéressantes, surtout lorsque que l'on cherche à évaluer les dégradations des incertitudes sur un problème donné.

Références

- [1] L. El Ghaoui and H. Lebret. Robust solutions to least-squares problems with uncertain data matrices. *SIAM J. Matrix Anal. Appl.*, 1996. To appear.
- [2] L. El Ghaoui and H. Lebret. Robust solutions to least squares problems with uncertain data. In SIAM, editor, *Recent Advances in Total Least Squares techniques and in Errors-in-Variables Modeling*, pages 161–170. SIAM, 1997.
- [3] L. El Ghaoui and H. Lebret. Inversion error, condition number and accurate inverses of structured matrices. In Preparation, 1997.
- [4] G. H. Golub and C. F. Van Loan. An analysis of the total least squares problem. *SIAM J. on Numerical Analysis*, 17 :883–893, 1980.
- [5] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret. Second-order cone programming. 1997. Submitted to *Linear Algebra and Applications*, special issue on linear algebra in control, signals and image processing.
- [6] F. Oustry, L. El Ghaoui, and H. Lebret. Robust solutions to uncertain semidefinite programs. Submitted to *SIAM J. Opt.*, 1996.
- [7] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Review*, 38(1) :49–95, March 1996.

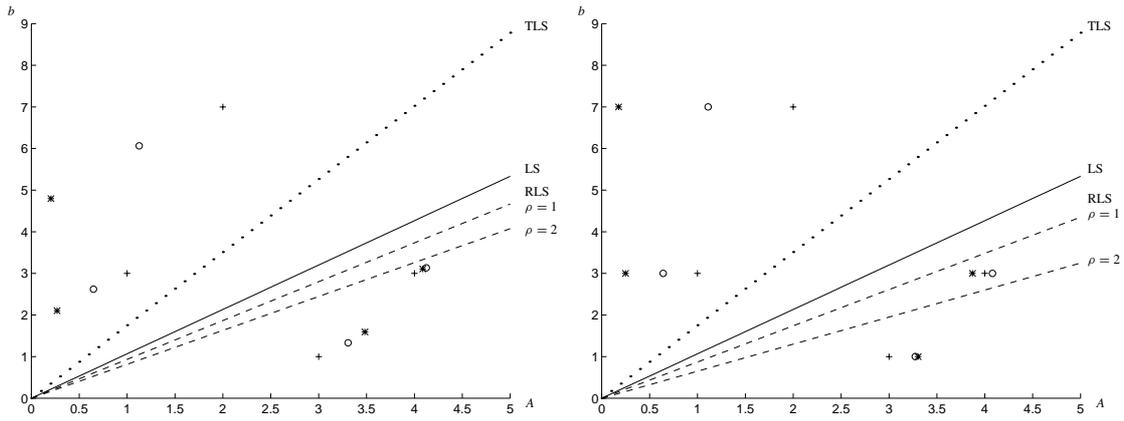


FIG. 1 — Comparaison des moindres carrés (LS), totaux (TLS) et robustes (RLS)

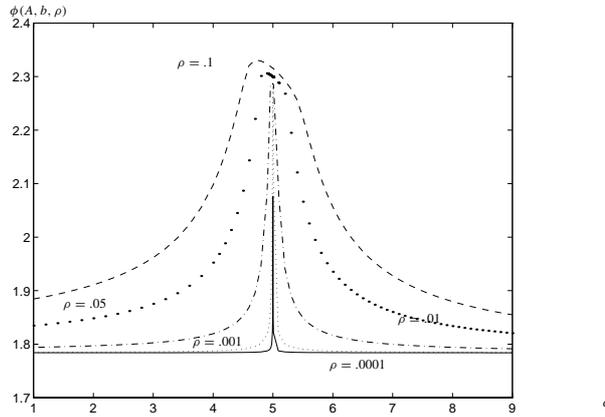


FIG. 2 — Résidu optimal en fonction de α pour différentes valeurs de la taille de l'incertitude ρ . La figure montre l'effet régularisateur de la solution robuste.

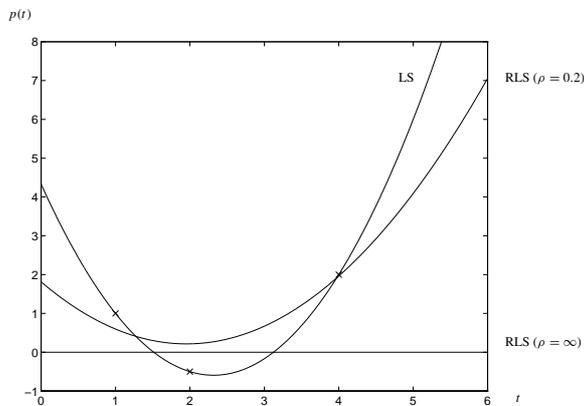


FIG. 3 — Interpolation polynômiale : solutions LS et RLS solutions pour $\rho = 0.2$. La solution LS donne une interpolation parfaite alors que la solution RLS garantit un résidu pire-cas de 1.1573. Pour $\rho = \infty$, la solution RLS donne le polynôme nul.