Identification de paramètres et optimisation
Cours de Master 2 STIM
2014-2015

Sébastien Adam

16 décembre 2015

Plan du Cours

Plan du Cours

Plan du Cours

Moindres carrés

Inversion

L'estimée θ_{MC} est donc la valeur de θ solution de l'équation :

$$X^T Y_S = X^T X \theta_{MC}$$

Deux cas de figure :

- $X^T X$ est inversible. On a alors $\theta_{MC} = \left(X^T X\right)^{-1} X^T Y_S$
- $X^T X$ n'est pas inversible. θ_{MC} n'est pas unique.

Cas pratiques de non inversibilité :

- Moins d'observations N que de paramètres K : rang$(X^T X) \leq N < K$ (rang(AB)$\leq$ rang(A) $\leq$ min(dim(A))).
- Dépendance linéaire entre les colonnes de X : on ne peut pas résoudre le système

Dans le cas pondéré, on aura : $\theta_{MC} = \left(X^T W X\right)^{-1} X^T W Y_S$

Moindres carrés

Propriétés de l'estimateur

- Modèle : $Y_M = X\theta$
- Minimum du critère $J_{MC}(\theta)$ atteint pour $\theta_{MC} = \left(X^T X\right)^{-1} X^T Y_S$
- On a alors $(Y_M)_{MC} = X\left(X^T X\right)^{-1} X^T Y_S = QY_S$
- L'erreur d'estimation vaut : $Y_S - Y_{M\,MC} = (1 - Q)Y_S$
- Démontrons quelques propriétés de Q :
  - Q est symétrique
  - $Q^2 = Q$
  - $(Y_S - QY_S)^T QY_S = 0$: $Y_S - (Y_M)_{MC}$ est orthogonal à $(Y_M)_{MC}$
- Retour sur l'interprétation géométrique

Moindres carrés

Interprétation stochastique

- On suppose :
  - $Y_S = Y_M(\theta_S) + b = X\theta_S + b$ où θ_S représente la vraie valeur (inconnue) des paramètres du système.
  - b est un bruit blanc (Gaussien, de moyenne nulle et de variance σ^2) indépendant des x_i
- On appelle erreur d'estimation le vecteur $e_\theta = \theta_{MC} - \theta_S$
- On a $e_\theta = \left(X^T X\right)^{-1} X^T \left(Y_S - X\theta_S\right) = \left(X^T X\right)^{-1} X^T b$
- Donc : $E[e_\theta] = \left(X^T X\right)^{-1} X^T E[b] = 0$
- Et $\mathrm{var}(e_\theta) = E[e_\theta e_\theta^T] = \left(X^T X\right)^{-1} X^T E[bb^T] X \left(X^T X\right)^{-1} = \sigma^2 \left(X^T X\right)^{-1}$

Moindres carrés

Bilan moindre carrés

- Modèle linéaire par rapport aux paramètres : $Y_M(\theta) = X\theta$
- Critère erreur quadratique : $J_{MC}(\theta) = \sum_{i=1}^{N} w_i \left(y_{S\,i} - y_{m\,i}\right)^2$
- Si $X^T X$ est inversible : $\theta_{MC} = \left(X^T X\right)^{-1} X^T Y_S$
- Remarque : en pratique $\left(X^T X\right)$ sera calculé par des méthodes QR
- Estimateur non biaisé le moins dispersé (Gauss-Markov) si la contrainte de linéarité est respectée.
- Modèle généralisable :
  - Si les θ apparaissent dans des expressions telles que $\cos(\theta_1)$, θ_1^2 ... : on change de paramètres.
  - Pour certains cas où Y_M n'est pas linéaire par rapport aux paramètres. Exemple : $Y_M = \theta_1^{\alpha_1}\theta_2^{\alpha_2} \rightarrow ln(Y_M) = \alpha_1 ln(\theta_1) + \alpha_2 ln(\theta_2)$: on change de paramètres et de critère.
- Comment mettre à jour une estimation ?

Plan du Cours

Moindres carrés récursifs (1)

Principes

- Objectif : calculer l'estimée $\theta_{MC,n}$ obtenue avec n observations à partir $\theta_{MC,n-1}$ obtenue avec $n-1$ observations
- On connait $\theta_{MC,n-1}$, x_n et ${y_s}_n$. Que vaut $\theta_{MC,n}$?

Méthodologie

- On a par définition $\theta_{MC,n} = \left(X_n^T X_n\right)^{-1} X_n^T {Y_S}_n$
- On pose :
  - $P_n = \left(X_n^T X_n\right)^{-1}$
  - $Q_n = X_n^T {Y_S}_n$
  - D'où $\theta_{MC,n} = P_n Q_n$
- Objectif pour mettre à jour une estimation : exprimer P_n et Q_n en fonction de P_{n-1}, Q_{n-1}, x_n et ${y_s}_n$

Moindres carrés récursifs (2)

Vision par bloc des matrices

$$X_n = \begin{pmatrix} \vdots \\ X_{n-1} \\ \vdots \\ x_n^T \end{pmatrix} \qquad X_n^T = (\ldots X_{n-1} \ldots, x_n) \qquad Y_{S\,n} = \begin{pmatrix} \vdots \\ Y_{n-1} \\ \vdots \\ y_{S\,n} \end{pmatrix}$$

Conséquences

- $Q_n = X_n^T Y_{S\,n} =$

Moindres carrés récursifs (2)

Vision par bloc des matrices

$$X_n = \begin{pmatrix} \vdots \\ X_{n-1} \\ \vdots \\ x_n^T \end{pmatrix} \qquad X_n^T = (\ldots X_{n-1} \ldots, x_n) \qquad Y_{S\,n} = \begin{pmatrix} \vdots \\ Y_{n-1} \\ \vdots \\ y_{S\,n} \end{pmatrix}$$

Conséquences

- $Q_n = X_n^T Y_{S\,n} = X_{n-1}^T Y_{S\,n-1} + x_n y_{S\,n} = Q_{n-1} + x_n y_{S\,n}$

Moindres carrés récursifs (2)

Vision par bloc des matrices

$$X_n = \begin{pmatrix} \vdots \\ X_{n-1} \\ \vdots \\ x_n^T \end{pmatrix} \qquad X_n^T = (\ldots X_{n-1} \ldots, x_n) \qquad Y_{S\,n} = \begin{pmatrix} \vdots \\ Y_{n-1} \\ \vdots \\ y_{S\,n} \end{pmatrix}$$

Conséquences

- $Q_n = X_n^T Y_{S\,n} = X_{n-1}^T Y_{S\,n-1} + x_n y_{S\,n} = Q_{n-1} + x_n y_{S\,n}$
- ${P_n}^{-1} =$

Moindres carrés récursifs (2)

Vision par bloc des matrices

$$X_n = \begin{pmatrix} \vdots \\ X_{n-1} \\ \vdots \\ x_n^T \end{pmatrix} \qquad X_n^T = (\ldots X_{n-1} \ldots, x_n) \qquad Y_{S\,n} = \begin{pmatrix} \vdots \\ Y_{n-1} \\ \vdots \\ y_{S\,n} \end{pmatrix}$$

Conséquences

- $Q_n = X_n^T Y_{S\,n} = X_{n-1}^T Y_{S\,n-1} + x_n y_{S\,n} = Q_{n-1} + x_n y_{S\,n}$
- ${P_n}^{-1} = X_n^T X_n = X_{n-1}^T X_{n-1} + x_n {x_n}^T = {P_{n-1}}^{-1} + x_n {x_n}^T \rightarrow P_n ?$

Problème

- Comment calculer P_n avec : ${P_n}^{-1} = {P_{n-1}}^{-1} + x_n {x_n}^T$

Moindres carrés récursifs (3)

Problème : inversion difficile

- $P_n = \left({P_n}^{-1}\right)^{-1}$ avec : ${P_n}^{-1} = {P_{n-1}}^{-1} + x_n {x_n}^T$

Moindres carrés récursifs (3)

Problème : inversion difficile

- $P_n = \left({P_n}^{-1}\right)^{-1}$ avec : ${P_n}^{-1} = {P_{n-1}}^{-1} + x_n {x_n}^T$

Solution : Lemme de Sherman-Morrison-Woodburg

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B\left[C^{-1} + DA^{-1}B\right]^{-1} DA^{-1}$$

Conséquences du lemme

$$P_n = P_{n-1} - k_n * {x_n}^T P_{n-1}$$

Avec

$$k_n = P_{n-1}x_n \left(1 + {x_n}^T P_{n-1} x_n\right)^{-1}$$

Moindres carrés récursifs (4)

Conséquences du lemme (2)

- On a obtenu :
  - $P_n = P_{n-1} - k_n * {x_n}^T P_{n-1}$
  - $k_n = P_{n-1} x_n \left(1 + {x_n}^T P_{n-1} x_n\right)^{-1} (**)$
  - $Q_n = Q_{n-1} + x_n {y_S}_n$
- On cherche $\theta_n = P_n Q_n$
  $\theta_n = P_{n-1} Q_n - k_n * {x_n}^T P_{n-1} Q_n$
- Après quelques simplifications : on obtient :
  $\theta_n = \theta_{n-1} + k_n({y_S}_n - {x_n}^T \theta_{n-1})$

Moindres carrés récursifs (5)

Bilan : 3 équations

$$\theta_n = \theta_{n-1} + k_n ({Y_S}_n - {x_n}^T \theta_{n-1})$$

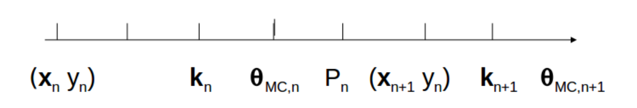
$$k_n = P_{n-1} x_n \left(1 + {x_n}^T P_{n-1} x_n\right)^{-1}$$

$$P_n = P_{n-1} - k_n * {x_n}^T P_{n-1}$$

Remarques

- L'estimée θ_n est une correction de θ_{n-1} de k_n fois la différence entre l'observation ${Y_S}_n$ et la prédiction ${Y_M}_n$ faite avec θ_{n-1}.

Ordre des calculs



Moindres carrés récursifs (6)

Illustration sur l'exemple d'approximation polynomiale

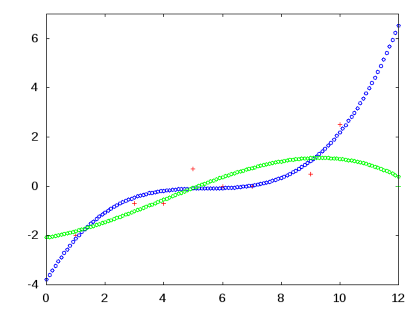
- On suppose l'arrivée d'un nouveau point et on met à jour l'estimation faite précédemment (code polyrec.m)

Moindres carrés récursifs (6)

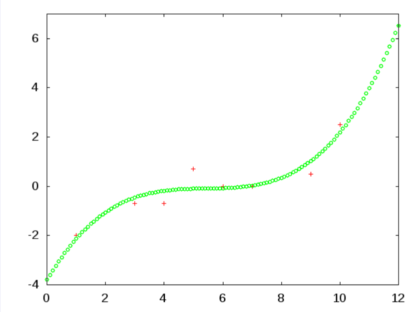
Illustration sur l'exemple d'approximation polynomiale

- On suppose l'arrivée d'un nouveau point et on met à jour l'estimation faite précédemment (code polyrec.m)

(12,0)



(12,6.527)



Moindres carrés récursifs (7)

Initialisation

- Dans les formules, on suppose que $P_n = \left(X_n^T X_n\right)^{-1}$ existe
- Cela implique $N \geq K$: on doit attendre k observations et vérifier inversibilité de P_k

Solution approchée

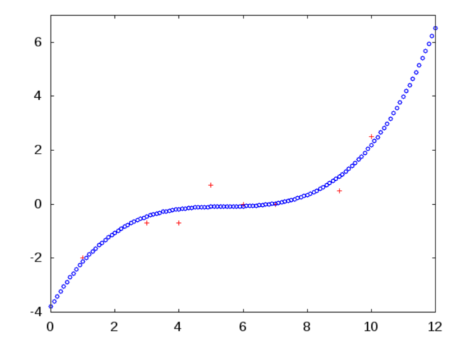
- On pose $P_0 = \alpha I$ et $\theta_0 = 0$, avec α grand (ex. 10^6)
- Or : $P_1^{-1} = P_0^{-1} + x_1 x_1^T$
  $\rightarrow P_1$ inversible car $X^T X + \alpha I$ est toujours inversible
- $P_2^{-1} = P_1^{-1} + x_2 x_2^T = P_0^{-1} + x_1 x_1^T + x_2 x_2^T$
- $\ldots$
- $P_n^{-1} = P_0^{-1} + \sum_{i=1}^{n} x_i x_i^T = P_0^{-1} + X^T X$

Moindres carrés récursifs (8)

Illustration sur l'exemple d'approximation polynomiale

- On pose $P_0 = \alpha I$ et $\theta_0 = 0$
- On simule l'arrivée de chaque nouveau point
- Démo (code polyrec2.m)

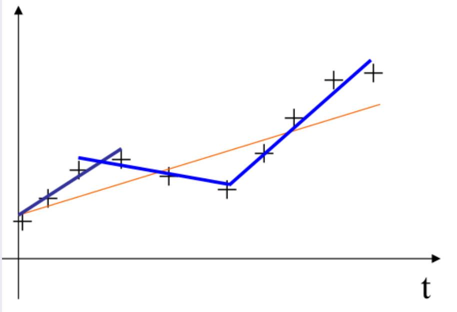
Résultat final



Moindres carrés récursifs (9)

Cas des paramètres lentement variables

- Contexte : les paramètres du modèle évoluent (lentement) dans le temps



Moindres carrés récursifs (9)

Cas des paramètres lentement variables

- Contexte : les paramètres du modèle évoluent (lentement) dans le temps
- Première option : intégration d'un facteur d'oubli dans le critère : $J_{MC}(\theta) = \sum_{i=0}^{N-1} \gamma^i \left(ys_{N-i} - ym_{N-i}\right)^2 \quad \text{avec} \quad \gamma < 1$
- Conséquence : les formules de récurrences sont modifiées

Moindres carrés récursifs (9)

Cas des paramètres lentement variables

- Contexte : les paramètres du modèle évoluent (lentement) dans le temps
- Première option : intégration d'un facteur d'oubli dans le critère : $J_{MC}(\theta) = \sum_{i=0}^{N-1} \gamma^i \left({y_S}_{N-i} - {y_m}_{N-i} \right)^2 \quad \text{avec} \quad \gamma < 1$
- Conséquence : les formules de récurrences sont modifiées

$$\theta_n = \theta_{n-1} + k_n({Y_S}_n - {x_n}^T \theta_{n-1})$$

$$k_n = P_{n-1} x_n \left(\gamma + {x_n}^T P_{n-1} x_n \right)^{-1}$$

$$P_n = (P_{n-1} - k_n * {x_n}^T P_{n-1})/\gamma$$

- Démo sur l'exemple de l'approximation polynomiale

Moindres carrés récursifs (10)

Cas des paramètres lentement variables

- Seconde option : moindres carrés à horizon fini

$$J_{MC}(\theta) = \sum_{i=0}^{L} \left(y_{S\,N-i} - y_{m\,N-i}\right)^2$$

- On en déduit

$$P_n{}^{-1} = P_{n-1}{}^{-1} + x_n x_n^T - x_{n-L-1} x_{n-L-1}^T$$

$$Q_n = Q_{n-1} + x_n Y_{S\,n} - x_{n-L-1} Y_{S\,n-L-1}$$

- Le lemme d'inversion matricielle ne s'applique plus

Moindres carrés récursifs (11)

Cas des paramètres lentement variables

- Troisième option : supprimer l'influence d'une mesure
- Pour supprimer l'influence de $(x_k, Y_{S\,k})$, on sait faire :

$$P_n{}^{-1} = P_{n-1}{}^{-1} - x_k x_k^T$$

$$Q_n = Q_{n-1} - x_k Y_{S\,k}$$

- On a alors :

$$\theta_n = \theta_{n-1} + k_n (Y_{S\,k} - x_k{}^T \theta_{n-1})$$

$$k_n = P_{n-1} x_k \left(-1 + x_k{}^T P_{n-1} x_k \right)^{-1}$$

$$P_n = P_{n-1} - k_n * x_k{}^T P_{n-1}$$

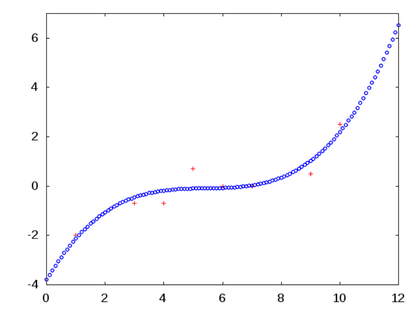
- On peut donc appliquer les MCR normaux, puis supprimer la mesure la plus ancienne.

Moindres carrés récursifs (12)

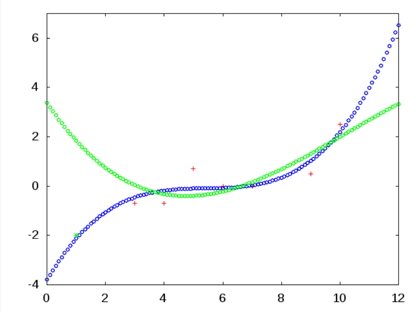
Illustration sur l'exemple d'approximation polynomiale

- On suppose supprimer l'influence du premier point et on met à jour l'estimation faite précédemment

Initial

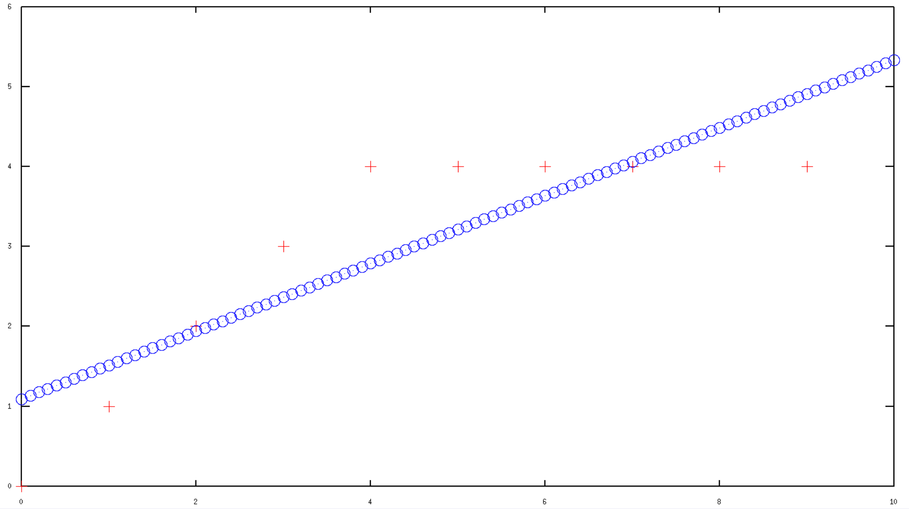


Point 1 supprimé



Moindres carrés récursifs (13)

Exemple complet sur un problème de regression linéaire



Plan du Cours

Plan du Cours

Plan du Cours

Plan du Cours