

B. DUBUISSON

J. J. SCHWARZ

**Utilisation de l'algorithme de Kalman pour le
traitement des séries chronologiques**

*Revue française d'automatique, d'informatique et de recherche
opérationnelle. Recherche opérationnelle*, tome 7, n° V1 (1973),
p. 21-29.

http://www.numdam.org/item?id=RO_1973__7_1_21_0

© AFCET, 1973, tous droits réservés.

L'accès aux archives de la revue « Revue française d'automatique, d'informatique et de recherche opérationnelle. Recherche opérationnelle » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/legal.php>). 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/>

UTILISATION DE L'ALGORITHME DE KALMAN POUR LE TRAITEMENT DES SERIES CHRONOLOGIQUES

par B. DUBUISSON ⁽¹⁾ et J. J. SCHWARZ ⁽²⁾

Résumé. — *Cet article reprend le développement de la méthode de filtrage de Kalman Bucy pour l'appliquer aux séries chronologiques.*

Si le modèle de la série est correct, cette méthode adaptative fournit de bons résultats pour la prévision.

L'obtention de méthodes de prévision pour les séries chronologiques est un travail très utile pour ses conséquences (la programmation d'une fabrication, l'établissement d'un budget, sont basées sur l'extrapolation de données passées).

Une méthode de prévision est nécessairement limitée par la qualité du modèle. Le but de cet article n'est pas de développer des procédures d'identification, mais de présenter une méthode de traitement connue et déjà utilisée pour les systèmes dynamiques, le filtrage de Kalman-Bucy [1] dans une application aux séries chronologiques. Dans une première partie, nous indiquerons la forme du modèle nécessaire, dans une seconde partie, l'algorithme, puis nous présenterons deux exemples sur des séries simulées.

1. MODELE DE LA SERIE CHRONOLOGIQUE

Les modèles de séries chronologiques généralement utilisés sont des modèles de type additif ou de type multiplicatif [2], faisant apparaître un terme de tendance et un terme de saisonnalité. Si on s'intéresse à un aspect plus dynamique, on représente une série chronologique par une équation aux différences gouvernées par un aléa $u(t)$ [3]

(1) Maître Assistant à l'Institut National des Sciences Appliquées de Lyon, Département d'Informatique.

(2) Assistant à l'Institut National des Sciences Appliquées de Lyon, Département d'Informatique.

$$x(k) + a_1 x(k-1) + \dots + a_n x(k-n) = u(k) \quad (1)$$

$x(k) = x(kT)$ T étant la période d'observation.

$x(k)$ se présente donc, comme un processus stochastique que nous supposons toujours du second ordre. L'application de la théorie des systèmes dynamiques a fait de grands progrès lorsque est intervenue la notion de variables d'état; c'est cette notion que nous allons utiliser ici. En effet, un processus stochastique donné par une équation aux différences analogue à (1) peut être représenté par un vecteur d'état, donc par un système d'équations aux différences du premier ordre [4]. (Le nombre d'équations est l'ordre de l'équation (1).)

Ce type de représentation se généralise aisément; elle est nécessaire pour l'emploi de la méthode de Kalman.

Nous supposons donc, dans la suite, que le modèle de la série est du type ci-dessous :

$$\begin{aligned} \overrightarrow{x(k+1)} &= A(k+1, k) \overrightarrow{x(k)} + B(k) \overrightarrow{e(k)} + C(k) u(k) \\ y(k) &= D(k) \overrightarrow{x(k)} + v(k) \end{aligned} \quad (2)$$

où : $\overrightarrow{x(k)}$ est la valeur prise par le vecteur d'état au temps $t = kT$,

$\overrightarrow{e(k)}$ est une fonction déterminée intervenant dans le modèle (entrée déterminée du système),

$u(k)$ est l'aléa de génération de la série chronologique (entrée aléatoire du système). Il est de moyenne nulle et non corréllé avec le passé.

$$E[u(k) u(j)] = \sigma_u^2(k) \delta(k - j)$$

$y(k)$ est l'observation au temps kT (valeur prise par la série)

$v(k)$ est l'aléa de mesure, de moyenne nulle, non corréllé avec le passé

$$E[v(k) v(j)] = \sigma_v^2(k) \delta(k - j)$$

et non corréllé avec $u(k)$

$$E[v(k) u(j)] = 0 \quad \forall k, j.$$

Remarquons que $u(k)$ et $v(k)$ peuvent être non stationnaires.

$$A(k+1, k), \quad B(k), \quad C(k), \quad D(k)$$

sont des matrices coefficients qui peuvent dans certains cas être constantes.

Le système (2) doit être complété par la donnée de $\overrightarrow{x(0)}$, pour sa résolution. Ce modèle permet donc de travailler sur des séries non stationnaires; il ne fait, de plus, pas appel à des conditions sur les lois de probabilité des aléas; il peut

toutefois être assez délicat à mettre en œuvre à cause du nombre de paramètres intervenant.

Cependant, ce modèle n'est pas complet car :

a) il ne fait pas intervenir des aléas de moyenne non nulles. On peut, dans ce cas, revenir au modèle (2) en modifiant le vecteur d'état $\overrightarrow{x(k)}$ (on prend comme vecteur d'état $\overrightarrow{x(k)} - E[\overrightarrow{x(k)}]$) et l'équation de l'observation $y(k)$ (celle-ci fait alors intervenir une fonction déterminée qui est la moyenne $E[v(k)]$ de $v(k)$).

b) On suppose que l'aléa $u(k)$ est non corréllé avec le passé. Dans le cas contraire, $u(k)$ peut être représenté par une équation aux différences analogues à (1), donc par un vecteur d'état. Pour se ramener au cas (2), il suffit alors de construire un vecteur d'état $\overrightarrow{x, (k)}$ composé de $\overrightarrow{x(k)}$ et du vecteur correspondant à $u(k)$.

c) On suppose que $v(k)$ est un aléa non corréllé avec le passé. Le cas contraire est plus complexe, mais la méthode peut toujours s'appliquer [5].

Pour voir l'emploi d'un tel modèle, prenons un exemple. Supposons que, par une méthode de régression, on ait pu montrer que la série s'écrivait :

$$x(k) = ax(k - 1) + bx(k - 2) + c + u(k).$$

Créons le vecteur d'état :

$$\overrightarrow{x(k)} = \begin{pmatrix} x_1(k) \\ x_2(k) \end{pmatrix} = \begin{pmatrix} x(k) \\ x(k - 1) \end{pmatrix}.$$

D'où le système d'équation :

$$\overrightarrow{x(k)} = \begin{pmatrix} a & b \\ 1 & 0 \end{pmatrix} \overrightarrow{x(k - 1)} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} c + \begin{pmatrix} 1 \\ 0 \end{pmatrix} u(k)$$

$$x(k) = (1, 0) \overrightarrow{x(k)}.$$

2. PREDICTEUR DE KALMAN

2.1. Prédicteur d'ordre 1

Nous nous intéressons maintenant à l'estimation du vecteur $\overrightarrow{X(k + 1)}$ désignée par $\hat{\overrightarrow{X}}_k(k + 1)$, à partir des observations passées $y(0) \dots y(k)$ (la grandeur intéressante est en fait $\hat{y}_k(k + 1)$, mais il est très facile de la déduire de $\hat{\overrightarrow{X}}_k(k + 1)$). Formuler une estimation suppose la définition d'un critère, nous prenons celui de la distance minimum.

Rappelons que l'espace engendré par les processus stochastiques réels du second ordre représentés par le système (2), peut être muni d'un produit scalaire.

$$\langle \overrightarrow{X(k)}, \overrightarrow{X(j)} \rangle = E[\overrightarrow{X(k)}[\overrightarrow{X(j)}]^T] \quad (3)$$

d'une norme,

$$\|\overrightarrow{X(k)}\|^2 = E[\overrightarrow{X(k)}[\overrightarrow{X(k)}]^T].$$

Cette dernière quantité permet de définir la distance de deux éléments

$$d(\overrightarrow{X(k)}, \overrightarrow{X(j)}) = \|\overrightarrow{X(k)} - \overrightarrow{X(j)}\|.$$

Le problème est donc de trouver l'élément $\hat{\overrightarrow{X}}_k(k+1)$ du sous espace engendré par $y(0) \dots y(k)$ (sous espace désigné par Y_k), situé à une distance minimum de $\overrightarrow{X}(k+1)$: cet élément est évidemment la projection de $\overrightarrow{X}(k+1)$ sur Y_k [6].

Partant de ces principes, nous allons rappeler les grandes lignes de la démonstration de l'algorithme du prédicteur $\hat{\overrightarrow{X}}_k(k+1)$ [5].

$\hat{\overrightarrow{X}}_k(k+1)$ est nécessairement la somme de deux termes :

- la projection de $\overrightarrow{X}(k+1)$ sur Y_{k-1} , $\hat{\overrightarrow{X}}_{k-1}(k+1)$,
- la projection de $\overrightarrow{X}(k+1)$ sur l'information,

apportée par $y(k)$, $(y(k) - \hat{y}_{k-1}(k))$ ($\hat{y}_{k-1}(k)$ désigne la projection de y_k sur Y_{k-1}).

$$\hat{\overrightarrow{X}}_k(k+1) = \hat{\overrightarrow{X}}_{k-1}(k+1) + \text{Projection de } \overrightarrow{X}(k+1) \text{ sur } (\hat{y}(k) - y_{k-1}(k)). \quad (6)$$

* En utilisant la propriété d'orthogonalité de $v(k)$ et du passé

$$\hat{y}_{k-1}(k) = D(k)\hat{\overrightarrow{X}}_{k-1}(k) \quad (7)$$

et la propriété identique de $u(k)$

$$\hat{\overrightarrow{X}}_{k-1}(k+1) = A(k+1, k)\hat{\overrightarrow{X}}_{k-1}(k) + B(k)e(k). \quad (8)$$

La relation (8) met en évidence le caractère itératif du procédé.

* Le calcul de la projection de $\overrightarrow{X}(k+1)$ sur $y(k) - \hat{y}_{k-1}(k)$ est résolu en utilisant le résultat connu

$$\begin{aligned} \text{Projection de } \overrightarrow{v_1} \text{ sur } \overrightarrow{v_2} &= \langle \overrightarrow{v_1} \overrightarrow{v_2} \rangle \|\overrightarrow{v_2}\|^{-2} \overrightarrow{v_2} \\ \text{Projection de } \overrightarrow{X}(k+1) \text{ sur } (y(k) - \hat{y}_{k-1}(k)) & \\ &= \langle y(k) - \hat{y}_{k-1}(k), \overrightarrow{X}(k+1) \rangle \|\overrightarrow{y(k) - \hat{y}_{k-1}(k)}\|^{-2} (y(k) - \hat{y}_{k-1}(k)). \quad (9) \end{aligned}$$

En utilisant toujours la non corrélation du passé, il vient

$$\hat{X}_k(k+1) = A(k+1, k)\hat{X}_{k-1}(k) + B(k)e(k) + H(k+1, k)(y(k) - D(k)\hat{X}_{k-1}(k)) \quad (10)$$

et

$$\hat{y}_k(k+1) = D(k)\hat{X}_k(k+1)$$

avec

$$H(k+1, k) = \frac{A(k+1, k) d^2(k, k-1) D^T(k)}{D(k) d^2(k, k-1) D^T(k) + \sigma_v^2(k)} \quad (11)$$

$H(k+1, k)$ est le gain du filtre de Kalman.

$$d(k, k-1) = d(\overrightarrow{X}(k), \hat{X}_{k-1}(k)). \quad (12)$$

Ainsi, la prévision du temps $(k+1)$ à partir du passé obtenu jusqu'au temps k est une fonction :

- de l'information nouvelle apportée par l'observation $y(k)$,
- de l'erreur de la prévision précédente, c'est-à-dire de $d^2(\overrightarrow{X}(k), \hat{X}_{k-1}(k))$.

* L'obtention de l'algorithme nécessite le calcul de $d^2(k, k-1)$. C'est un calcul assez long dont nous donnons simplement le résultat (on utilise de la même manière que précédemment la non corrélation des aléas avec le passé) [5].

$$d^2(k+1, k) = A(k+1, k) d^2(k, k-1) A^T(k+1, k) + C(k) \sigma_u^2(k) C^T(k) - \frac{A(k+1, k) d^2(k, k-1) D^T(k)}{D(k) d^2(k, k-1) + \sigma_v^2(k)} D(k) d^2(k, k-1) A^T(k+1, k) \quad (13)$$

* En résumé, le calcul du prédicteur nécessite les quatre phases suivantes :

- 1) calcul de $H(k+1, k)$,
- 2) calcul de $\hat{X}_k(k+1)$,
- 3) calcul de $d^2(k+1, k)$,
- 4) retour en 1) après avoir incrémenté k de 1.

Le démarrage de cet algorithme demande la connaissance de $\hat{X}_0(1)$ et de $d(1, 0)$. $\overrightarrow{X}(0)$ est un vecteur aléatoire dont la connaissance est nécessaire pour la résolution de (2) : on en connaît donc sa moyenne $\overline{m}(0)$ et sa variance $\sigma^2(0)$.

On convient de prendre pour estimateur de $\overrightarrow{X}(1)$, $\hat{X}_0(1)$, la quantité

$$\hat{X}_0(1) = A(1, 0)\overline{m}(0) + B(0)e(0) \quad (14)$$

$$d^2(1, 0) = A(1, 0)\sigma^2(0)A^T(1, 0) + C(0)\sigma_u^2(0)C^T(0) \quad (15)$$

Ces conditions initiales permettent le calcul de $H(1, 0)$, donc de débiter l'algorithme.

2.2. Prédicteur d'ordre supérieur à 1

Le prédicteur d'ordre N ($N > 1$) se déduit directement du prédicteur d'ordre 1. En effet :

$$\begin{aligned} \overrightarrow{X}(k+N) &= A(k+N, k+N-1)[A(k+N-1, k+N-2)[\dots \\ &\quad [A(k+2, k+1)\overrightarrow{X}(k+1) + B(k+1)\overrightarrow{e}(k+1) + C(k+1)u(k+1)] \\ &+ B(k+2)\overrightarrow{e}(k+2) + C(k+2)u(k+2)] + \dots \\ &\quad + B(k+N-1)\overrightarrow{e}(k+N-1) + C(k+N-1)u(k+N-1). \end{aligned} \quad (16)$$

Donc, en utilisant le théorème de la projection

$$\begin{aligned} \hat{\overrightarrow{X}}_k(k+N) &= A(k+N, k+N-1)[A(k+N-1, k+N-2) \\ &\quad [\dots [A(k+2, k+1)\hat{\overrightarrow{X}}_k(k+1) + B(k+1)\overrightarrow{e}(k+1)] \\ &\quad + B(k+2)\overrightarrow{e}(k+2)] + \dots] + B(k+N-1)\overrightarrow{e}(k+N-1). \end{aligned} \quad (17)$$

Il est bien évident que la distance entre la réalisation et la valeur prédite est d'autant plus importante que le décalage vers l'avenir est grand.

3. INTERET DE LA METHODE DE KALMAN

L'algorithme de Kalman a comme premier intérêt d'être adaptatif : chaque itération tient compte de l'erreur du pas précédent. Cette erreur est l'erreur estimée puisque on peut ne pas connaître la réalisation exacte (présence possible d'un aléa de mesure). L'algorithme est convergent mais l'erreur ne peut descendre au dessous d'un certain seuil défini par les variances des deux aléas.

Nous pensons que son grand intérêt pour les séries chronologiques est le traitement sur un modèle non figé : en effet, toutes les matrices dépendent de l'instant d'observation.

Cependant, un tel modèle est difficile à appliquer dans toute sa variété. Pour une application pratique, on aura donc en général :

- 1) les coefficients σ_u^2 et σ_v^2 constants,
- 2) les matrices $D(k)$, $B(k)$, $C(k)$, constantes,
- 3) la matrice $A(k+1, k)$ variable.

En effet, pour les premières quantités, il est difficile d'avoir de l'information sur les aléas ; de même pour les secondes, on fixe le modèle par manque de précision. Nous pensons que la matrice paramètre $A(k+1, k)$ suffit pour les séries chronologiques (remarquons que sa variabilité entraîne aussi la variabilité du modèle de l'aléa $u(k)$ dans le cas d'un aléa corréllé).

Nous proposons donc une utilisation comme suit :

- 1) A partir de l'historique obtention d'un modèle aussi complet que possible de la série (plus le passé est important, plus le modèle est sûr).
- 2) Utilisation du passé pour adapter la méthode de Kalman à la série (on calcule les « prévisions » sur toutes les valeurs passées).
- 3) Premières prévisions.
- 4) Observations des réalisations et mise à jour du modèle.
- 5) Nouvelles prévisions avec le nouveau modèle et retour en (4).

C'est de cette manière, à notre avis, qu'on utilise les deux caractéristiques du filtre de Kalman : adaptivité, modèle non fixe.

4. EXEMPLES D'UTILISATION

Nous proposons ici deux exemples d'utilisation sur deux séries simulées extraites du livre de M. Kaufmann et Groboillot déjà cité [3]. Travaillant sur des séries simulées il n'est pas besoin de passer par la première phase du traitement d'une série chronologique, l'identification de son modèle. Les deux exemples correspondent à des modèles dits « à deux aléas ».

a) *Premier exemple :*

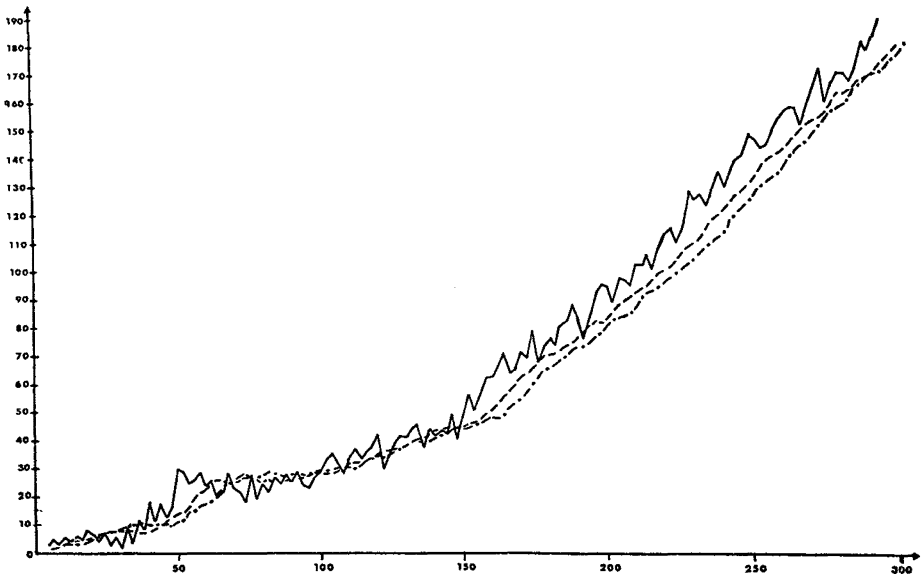


Fig. 1

Premier exemple sur série simulée

Courbe trait plein : réalisation
 Courbe pointillée : prévision à $k + 2$
 Courbe trait discontinu, prévision à $k + 6$

C'est un exemple de modèle à deux aléas avec rupture de tendance :

$$x(kT) = x(k-1)T + e(kT) + u(kT)$$

$$y(kT) = x(kT) + v(kT)$$

$$\sigma_u^2(kT) = 1 \quad \sigma_v^2(kT) = 16.$$

$e(kT) = 0,5$ pour les 75 premières observations.

= 2 pour les suivantes.

Nous avons réalisé (fig. 1), des estimations de la valeur prise par la série au temps $(k+1)$ et $(k+6)$, en partant d'une valeur de départ égale à 1 avec une variance de 4. On peut remarquer que tant que le modèle convient, il y a une remarquable concordance entre les réalisations et les prévisions, ce qui est normal. Lorsque le modèle change, le prédicteur s'adapte avec un retard plus ou moins grand suivant la valeur de la projection vers l'avenir de la prévision. Ainsi, apparaît un écart qui est ici, pour employer le langage des systèmes, un écart de trainage.

Une utilisation plus normale aurait été de recomposer à chaque pas le modèle, ce qui aurait eu pour conséquence de faire disparaître cet écart.

b) *Deuxième exemple :*

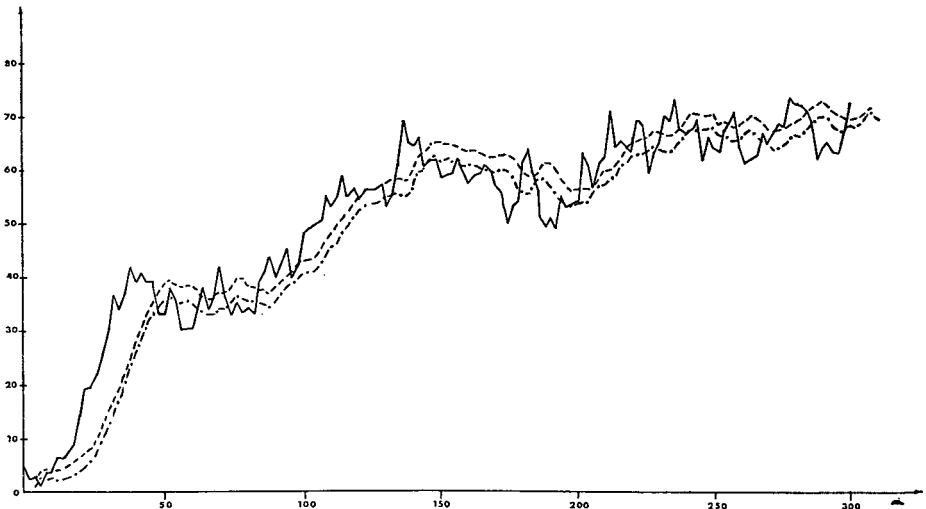


Figure 2

Deuxième exemple sur série simulée

Courbe trait plein : réalisation

Courbe trait pointillé : prévision à $i+2$ (tendance égale à 1, $\sigma_u^2 = 16$, $\sigma_v^2 = 1$)

Courbe trait discontinu : prévision à $i+2$ (tendance égale à 0,5, $\sigma_u^2 = 16$, $\sigma_v^2 = 1$).

Le modèle exact correspond à une tendance égale à 0,5, à des variances σ_u^2 et σ_v^2 respectivement égales à 9 et 1.

C'est encore un modèle à deux aléas, mais ici à tendance stationnaire, égale à 0,5. Une application du modèle exact donne évidemment des résultats très bons à une variance des aléas près (variances exactes des aléas respectifs 9 et 1). La prévision à $(k + 2)$ que nous présentons (fig. 2) correspond à un modèle inexact de la série ($e(kT) = 1$, $\sigma_u^2 = 16$, $\sigma_v^2 = 1$, valeur de départ égale à 1 avec une variance de 9). On peut voir ici que le prédicteur s'adapte remarquablement bien, malgré un modèle faux, à la série. Simplement, si on compare les deux exemples, il lui faut dans ce dernier cas, un historique beaucoup plus long (historique double de celui nécessaire à l'exemple 1).

5. CONCLUSION

La méthode de Kalman est certainement très puissante; cependant, elle nécessite, comme d'autres, la connaissance d'un modèle de la série. Si le modèle est éloigné de la réalité, il faudra un long historique pour que le prédicteur s'adapte. Si le modèle évolue, la méthode de Kalman retrouve tout son intérêt, car elle permet de travailler sur des modèles non figés.

En conclusion, la méthode offre un éventail de possibilités assez grand pour permettre la réalisation de prévision assez bonnes dans la majorité des cas.

BIBLIOGRAPHIE

- [1] KALMAN R. E. et BUCY R., *New results in linear filtering and prediction theory*, Trans. A.S.M.E., J. Basic Eng., vol. 83 D., mars 1961, p. 95-108.
- [2] D'HOERAENE J. et LEDOUX Y., *La prévision dans l'entreprise*, Dunod, 1971.
- [3] KAUFMAN H. et GROBOILLOT J. L., *La prévision économique à court terme. Méthodes générales. Lissage exponentiel*, Dunod, 1968.
- [4] VAN TREES H. L., *Detection, estimation and modulation theory* (Part 1), John Wiley and Sons, New York, 1968.
- [5] SAGE A. P. et SELSA J. L., *Estimation theory with applications to communications and control*, Mac Graw Hill, 1971.
- [6] PAPOULIS A., *Probability random variables and stochastic process*, Mac Graw Hill, New York, 1965.