

ESTIMATION DE LA STRUCTURE
D'UN PROCESSUS STOCHASTIQUE MULTIVARIABLE

FUCHS JEAN-JACQUES

IRISA/UNIVERSITE DE RENNES I / GRECO TSI, CAMPUS DE BEAULIEU
35042 RENNES CEDEX

RESUME

Pour modéliser un processus stochastique vectoriel de manière efficace il faut non seulement préciser l'ordre du modèle utilisé (le nombre de coefficients matriciels dans le cas d'un modèle ARMA, la dimension de l'état dans le cas d'une représentation d'état) mais également sa structure (un ensemble de paramètres entiers qui caractérise une forme canonique permettant de représenter le processus). Si on omet de spécifier la structure, la paramétrisation est redondante, le modèle non uniquement identifiable et on peut s'attendre à une procédure d'identification des coefficients du modèle numériquement mal conditionnée. L'estimation de la structure est donc un préalable à toute identification non ad-hoc d'un processus vectoriel.

On propose une méthode qui permet d'estimer la structure d'une représentation d'état d'un processus à partir de l'observation d'une de ses réalisations. On suppose que le processus observé admet une telle représentation, qu'il s'agit donc d'un processus à spectre rationnel, ou encore, qu'il peut être vu comme la sortie d'un système linéaire invariant attaqué par un bruit blanc vectoriel. L'algorithme proposé construit séquentiellement un ensemble de vecteurs lignes linéairement indépendants de la matrice de Hankel construite sur la suite des matrices de covariances estimées. Il fournit les invariants de Kronecker liés à l'observabilité.

In order to estimate in an efficient way a parametric model for a vector stochastic process, one has to specify not only the order of the model but also its structure. We present a procedure allowing to estimate the structure of a state-space representation of a multivariate stochastic process from measured output data. It is assumed that the observed time series is a realization of a process with rational spectrum or the output of a stable, invariant linear system driven by white noise.

We propose an algorithm which selects a maximal set of linearly independent rows of the Hankel matrix built upon the estimated covariance sequence. This set is obtained by sequentially testing the smallest singular value of submatrices of the Hankel matrix and yields estimates of the observability Kronecker invariants. The scheme when compared to the existing ones on simulated examples, exhibits better or at least comparable results for a much lower computational complexity.

FORMULATION DU PROBLEME

On considère un processus y_t vectoriel de dimension p , stationnaire, de rang plein qui admet une représentation d'état d'ordre minimal n :

$$\begin{aligned} X_{t+1} &= F X_t + K e_t \\ y_t &= H X_t + e_t \end{aligned} \quad (1)$$

où e_t est un bruit blanc vectoriel de dimension p (le processus d'innovation de y_t) et de matrice de covariance définie positive Σ . A partir d'une suite d'observations $\{y_t\}$ de longueur T , nous indiquons comment déterminer l'ordre n du processus et un ensemble de p entiers appelés des invariants de structure. Ils permettent de définir parmi l'ensemble des triplets $(HT, T^{-1}FT, T^{-1}K$ avec T une matrice inversible) qui tous représentent le processus, un triplet spécifique identifiable de manière unique dans lequel ne persiste aucune redondance paramétrique. On va choisir un ensemble d'invariants -les invariants d'observabilité de Kronecker- qui vont considérablement diminuer le nombre de paramètres à identifier : en effet alors que les matrices (H,F,K) contiennent n^2+2np éléments, on peut montrer [1] que les p entiers caractérisant la structure et au plus $2np$ paramètres réels suffisent pour représenter la même information. Le triplet (H,F,K) correspondant sera sous forme canonique, rempli de 0 et de 1 et d'au plus $2np$ éléments réels à identifier. On peut espérer que l'élimination de la redondance va conduire à une efficacité accrue des algorithmes d'identification.

A partir de (1) on vérifie, qu'en définissant la suite de covariance :

$$R_j = E(y_{t+j} y_t^T) \quad (2)$$

on a :

$$\begin{aligned} R_k &= H F^{k-1} G \quad k > 0 \\ R_k &= R_k^T \\ R_0 &= H P H^T + \Sigma \end{aligned}$$

où $P = E(X_t X_t^T)$ et $G = E(X_{t+1} y_t^T)$ satisfont :

$$\begin{aligned} P &= F P F^T + K \Sigma K^T \\ G &= F P H^T + K \Sigma \end{aligned} \quad (3)$$

Si les vecteurs infinis du futur Y_t^* et du passé Y_t^- de y_t sont :

$$\begin{aligned} Y_t^* &= [y_t^T \ y_{t+1}^T \ y_{t+2}^T \ \dots]^T \\ Y_t^- &= [y_{t-1}^T \ y_{t-2}^T \ y_{t-3}^T \ \dots]^T \end{aligned} \quad (4)$$



il est facile de vérifier que

$$H = E(Y^*, Y^{*T}) = \begin{bmatrix} R_1 & R_2 & R_3 & \dots \\ R_2 & R_3 & \dots & \dots \\ R_3 & \dots & \dots & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad (5)$$

est une matrice bloc-Hankel de rang n que l'on peut écrire :

$$H = O C \quad (6)$$

avec

$$O = [H^T (HF)^T (HF^2)^T \dots]^T$$

$$C = [G FG F^2G \dots]$$

les matrices d'observabilité et de controllabilité associées au triplet (H, F, G) .

Dans ce contexte, l'ordre n étant supposé connu, estimer la structure revient à indiquer comment spécifier parmi l'infinité de triplets équivalents (H, F, G) qui permettent la factorisation (6), un triplet représentatif de la classe d'équivalence à laquelle appartient le processus d'ordre n considéré, l'ensemble des processus d'ordre n pouvant être décrit par un nombre fini de classes. On choisira en général comme représentant un triplet de matrices contenant un nombre minimal de composantes différentes de 0 ou 1. Ce problème de structure de systèmes linéaires a été longuement étudié [1,2] et nous indiquons simplement que sa solution est équivalente à définir les diverses manières de sélectionner des ensembles de n lignes (colonnes) indépendantes de la matrice infinie $H(5)$. Dans la suite nous supposons que les p premières lignes (colonnes) de H sont indépendantes et ne considérerons que la sélection de lignes indépendantes. On peut noter qu'à cause de la structure bloc-Hankel si la $j^{\text{ème}}$ ligne est linéairement dépendante de celles qui précèdent, il en sera de même des lignes $\{j+kp, k=1, \dots\}$. Si on associe alors à une sélection un n -uplet $S(n) = (i_1, i_2, \dots, i_n)$ avec $i_j < i_{j+1}$ les indices des lignes retenues, on ne considérera que les n -uplets satisfaisant les 2 conditions :

$$\begin{aligned} \text{condition 1 : } & j \in S(\cdot) \text{ pour } j = 1 \text{ à } p \\ \text{condition 2 : } & \text{si } j > p \in S(\cdot) \text{ alors } j-p \in S(\cdot) \end{aligned}$$

Et en fait nous nous limiterons à l'unique sélection qui consiste à choisir les n premières lignes indépendantes de H . Les p entiers n_j associés de la manière suivante à cette sélection $S(n)$:

$$n_j = \arg \min_m (j + mp \notin S(n)) \quad j = 1 \text{ à } p \quad (7)$$

sont alors les invariants de Kronecker. Nous n'indiquons pas quels éléments des matrices H et F peuvent être fixés à 0 ou 1 pour une telle sélection [1,2,3].

Notre but est donc de proposer une manière d'estimer les invariants de Kronecker à partir de T observations d'un processus y_t admettant un modèle de la forme (1). On commence alors par estimer la suite de matrices de covariance du processus :

$$\hat{R}_j = \frac{1}{T-j} \sum_t y_{t+j} y_t^T \quad (8)$$

et on poursuit en testant de manière séquentielle la plus petite valeur singulière d'une suite de sous-matrices de $\hat{H}(\cdot)$ de dimensions non-décroissantes, où $\hat{H}(\cdot)$ est une matrice de Hankel de dimension finie construite à l'aide des covariances estimées (8).

Soit alors $H(S(k), M)$ la sous-matrice de dimension (k, Mp) construite à l'aide des M premières bloc-colonnes de $H(5)$ et des k lignes dont les indices sont

spécifiés par le k -uplet $S(k)$. Supposons pour le moment que M est infini alors sur la matrice $H(5)$ exacte on peut procéder de la manière suivante pour estimer les invariants de Kronecker, pour identifier les premières lignes indépendantes.

Comme les p premières lignes sont (supposées) indépendantes ; posons $k = p$, $S(k) = (1, 2, \dots, p)$; déclarons autorisées toutes les autres lignes et réalisons l'algorithme suivant :

- 1) ajouter la ligne autorisée suivante d'indice ℓ à la sous matrice $H(S(k), M)$ et calculer sa plus petite valeur singulière
- si elle est non-nulle : ajouter l'indice ℓ à $S(k)$ pour obtenir $S(k+1)$ et retourner en 1)
- si elle est nulle : enlever la dernière ligne ℓ , déclarer interdites toutes les lignes d'indice $\{\ell+jp; j=0, 1, \dots\}$; retourner en 1) s'il reste des lignes autorisées sinon la procédure est terminée.

A la sortie $k=n$, $S(k=n)$ contient les indices des n premières lignes indépendantes dont on déduit les invariants de Kronecker (7). Le nombre de bloc-colonnes M doit être suffisamment grand pour ne pas interférer avec la procédure décrite, si on suppose connaître n , un majorant de n , on peut montrer [4] qu'il suffit de prendre :

$$M = \bar{n} - p + 1 \quad (9)$$

Notons que cette hypothèse, la connaissance a priori d'un majorant de n , apparaît toujours et est inhérente à ce type de problème. L'algorithme va en fait être appliqué à $\hat{H}(S(k), M)$ de dimension (k, Mp) une estimée de $H(S(k), M)$ et l'on a :

$$\hat{H}(\cdot) = H(\cdot) + \delta H(\cdot) \quad (10)$$

où le terme dominant dans $\delta H(\cdot)$ est d'ordre $T^{-\frac{3}{2}}$, noté $O(T^{-\frac{3}{2}})$. En effet les estimées (8) des covariances R_j sont convergentes et asymptotiquement normales, il en est donc de même pour $\hat{H}(\cdot)$ et $\delta H(\cdot)$ peut être vue comme une matrice aléatoire gaussienne, centrée, de variance d'ordre T^{-1} . Pour T fini, $\hat{H}(\cdot)$ est alors génériquement de rang plein et puisque dans l'algorithme proposé plus haut les sous-matrices à tester possèdent au plus une valeur singulière nulle, il nous faut définir un test permettant de décider si $\hat{\sigma}_{\min}$, la plus petite valeur singulière courante, doit être déclarée nulle ou non.

ANALYSE

Nous allons étudier les propriétés statistiques de $\hat{\sigma}_{\min}$, la plus petite valeur singulière de $\hat{H}(S(k), M)$ sous l'hypothèse que la sous-matrice exacte correspondante possède une valeur singulière nulle, unique et bien isolée. De ces propriétés nous déduirons un test qui, inséré dans l'algorithme précédent, constituera la procédure d'estimation de structure recherchée.

La décomposition en valeurs singulières de la matrice (rectangulaire) $H(S(k), M)$ de dimension (k, Mp) avec $Mp > k$ est notée

$$H(\cdot) = U \Sigma V^T$$

avec U et V des matrices carrées orthogonales. Nous partitionnons ces matrices :

$$\Sigma = \begin{bmatrix} \Sigma_1 & 0 & 0 \\ 0 & \sigma_k & 0 \end{bmatrix} \quad U = [U_1 \ U_k] \quad V = [V_1 \ V_2] \quad (11)$$



avec $\Sigma_1 = \text{diag}(\sigma_i)$ et $\sigma_1 \geq \sigma_2 \geq \dots \sigma_{k-1} > 0$, $\sigma_k = 0$. u_k est le vecteur propre nul unique de $\mathbb{H}(\cdot) \mathbb{H}(\cdot)^T$ et V_2 une base orthogonale du noyau de $\mathbb{H}(\cdot)^T \mathbb{H}(\cdot)$ de dimension $Mp-(k-1)$. En désignant par \hat{U} , $\hat{\Sigma}$, \hat{V} ... les éléments de la décomposition en valeurs singulières de $\hat{\mathbb{H}}(\cdot)$, on peut établir la propriété suivante [4] :

Proposition : Pour T suffisamment grand et pour une matrice perturbée $\hat{\mathbb{H}}(\cdot)$, il existe une base \hat{V}_2 du noyau de $\mathbb{H}(\cdot)^T \mathbb{H}(\cdot)$ pour laquelle :

$$\hat{V}_2 = \check{V}_2 + o(T^{-1/2}) \quad (12)$$

$$\hat{U} = U + o(T^{-1/2})$$

$$\hat{\sigma}_k = \|u_k^T \hat{\mathbb{H}}(\cdot) V_2\| + o(T^{-1}) \quad (13)$$

où $\|\cdot\|$ désigne la norme euclidienne.

La proposition nous apprend (12) que la base \hat{V}_2 estimée est proche d'une certaine base exacte. Dans (13) par contre la base V_2 est arbitraire et comme $\mathbb{H}(\cdot) V_2 = 0$, (13) se réécrit

$$\hat{\sigma}_k \simeq \|u_k^T \hat{\mathbb{H}}(S(k), M) V_2\| \quad (14)$$

Comme la valeur singulière nulle perturbée $\hat{\sigma}_k$ peut être approchée par la norme euclidienne d'un vecteur, nous allons étudier la loi de ce vecteur que nous notons Z, de dimension r et de composantes z_i :

$$Z^T = u_k^T \hat{\mathbb{H}}(S(k), M) V_2 \quad (15)$$

$$r = \dim Z = Mp-(k-1)$$

$$z_i = u_k^T \hat{\mathbb{H}}(\cdot) v_i \quad i = 1 \text{ à } r \quad (16)$$

On peut voir la dernière relation (16) comme la définition de la variable aléatoire z_i en fonction de la matrice aléatoire $\hat{\mathbb{H}}(\cdot)$ dont une estimée peut être obtenue de la manière suivante :

$$\hat{\mathbb{H}}(S(k), M) \simeq \frac{1}{T} \sum_t Y_t^*(S(k)) Y_t^T(M)^T \quad (17)$$

avec, cf. (4) (5) :

$$Y_t^T(M) = [y_{t-1}^T \ y_{t-2}^T \ \dots \ y_{t-M}^T]^T \quad (18)$$

et $Y_t^*(S(k))$ un vecteur colonne construit à l'aide des k composantes de Y_t^T (4) spécifiées par S(k). On a alors pour z_i (16,17,18) :

$$z_i \simeq \frac{1}{T} \sum_t u_t v_t(i) \quad (19)$$

avec :

$$u_t = u_k^T Y_t^*(S(k)) \quad (20)$$

$$v_t(i) = v_i^T Y_t^T(M)$$

Ces relations montrent que z_i peut être vu comme une estimée de la covariance entre les deux processus u_t et $v_t(i)$ qui sont eux-mêmes des combinaisons linéaires des composantes de y_t à différents instants. Ces processus sont donc centrés, stationnaires et z_i va vérifier un théorème limite centrale. Il en est de même pour le vecteur Z qui converge donc en loi vers un vecteur gaussien centré de matrice de covariance Q que nous allons évaluer. Dans un premier temps, on peut vérifier que :

$$E(u_t v_{t+j}(i)) = u_k^T \{E(Y_t^*(S(k)) Y_{t+j}^T(M)^T)\} v_i \quad (21)$$

est nul pour $j \geq 0$ et $i = 1$ à r. En effet, la matrice intervenant dans (21) est construite à l'aide de bloc-colonnes de $\mathbb{H}(S(k), \infty)$ qui par définition de u_k satisfait $u_k^T \mathbb{H}(S(k), \infty) = 0$. Dans un deuxième temps, il est important de noter que u_t est un processus m-dépendant. En effet si on définit

$$\gamma_u(j) = E(u_t u_{t-j}) = u_k^T T_u(j) u_k \quad (22)$$

$$\text{avec : } T_u(j) = E(Y_t^*(S(k)) Y_{t-j}^T(S(k))^T) \quad (23)$$

et si l'on note N_k le nombre de bloc-lignes dans lesquelles S(k) sélectionne au moins une ligne, il est facile de réaliser que pour $j \geq N_k$ les composantes de $Y_{t-j}^*(S(k))$ sont également des composantes de Y_t^* et que par conséquent les k colonnes de $T_u(j)$ sont aussi des colonnes de $\mathbb{H}(S(k), \infty)$. On en déduit :

$$\gamma_u(j) = 0 \quad \text{pour } j \geq N_k$$

ce qui établit que u_t (20) est un processus m-dépendant d'ordre N_k-1 . Cette propriété est importante car elle implique qu'il n'y a qu'un nombre fini de termes non nuls dans l'expression de $E(z_i z_j)$ et donc dans celle de $Q = E(ZZ^T)$ dont on peut vérifier qu'elle vaut

$$Q = \frac{1}{T} V_2^T \left\{ \gamma_u(0) T_v(0) + \sum_{\ell=1}^{N_k-1} \gamma_u(\ell) (T_v(\ell) + T_v^T(\ell)) \right\} V_2 \quad (24)$$

$$\text{avec : } T_v(\ell) = E(Y_t^T(M) Y_{t+\ell}^T(M)^T) ,$$

$\gamma_u(\cdot)$ définit en (22,23) et N_k le nombre de bloc-lignes affectées par S(k).

Résumons les résultats obtenus et spécifions le test que nous proposons à chacun des pas de l'algorithme précédent. Etant donné $\hat{\mathbb{H}}(S(k), M)$ de vrai rang k-1, nous savons que sa valeur singulière minimale peut être approchée par la norme euclidienne du vecteur Z(15) qui est asymptotiquement en T, gaussien, centré et de variance Q(24). On sait qu'alors

$$\mu = Z^T Q^{-1} Z \quad (25)$$

est une variable aléatoire du Chi-deux à r degrés de liberté et que dans le cas contraire, si Z n'est plus centré (associé à une valeur singulière non nulle), sa loi est différente. La règle que nous proposons est alors de la forme :

$$\mu \leq t_r \rightarrow \text{rang } \hat{\mathbb{H}}(S(k), M) = k-1$$

$$\mu > t_r \rightarrow \text{rang } \hat{\mathbb{H}}(\cdot) = k \quad (26)$$

où le seuil t_r est fixé de manière à réaliser une certaine probabilité de déclarer de rang plein une matrice de rang déficient.

Il nous reste à indiquer comment, à l'aide des données, construire des estimées de Z, Q et μ ces quantités n'étant pas disponibles. Il est alors crucial de vérifier que la valeur numérique prise par μ ne dépend pas de la base V_2 (11,12) utilisée dans le calcul de Z et Q. On peut donc utiliser \check{V}_2 (12) la base qui dépend des observations. En fait on remplace \check{V}_2 par \hat{V}_2 la seule base disponible en vérifiant que l'erreur ainsi introduite dans l'évaluation de \hat{Z} , \hat{Q} et $\hat{\mu}$ est bien négligeable. On a alors (15) :



$$\hat{Z}^T = \hat{u}_k^T \hat{H}(S(k), M) \hat{V}_2^T \quad (27)$$

$$= [\hat{\sigma}_k \ 0 \ 0 \ \dots \ 0] .$$

L'estimée \hat{Q} de Q s'obtient de même (22,24) en remplaçant les quantités inconnues par leurs estimées. Les composantes nulles de \hat{Z} (27) ne devraient pas trop surprendre, ce sont des estimées de termes de l'ordre de T^{-1} que nous négligeons quand nous remplaçons \check{Z} par \hat{Z} .

RESULTATS DE SIMULATION

Nous présentons des résultats de simulation sur un exemple proposé dans [5] et repris dans [6,7]. Peu de méthodes non heuristiques existent pour estimer la structure. Nous appliquons l'algorithme décrit dans le deuxième paragraphe avec le test (25,26). Il en résulte les différentes étapes suivantes :

- estimer les premiers termes de la suite de covariances (8)
- construire la matrice bloc-Hankel (5)
- extraire des sous-matrices $\hat{H}(S(k), M)$ avec un nombre non décroissant k de lignes et $M(9)$ bloc-colonnes
- pour chacune des ces sous-matrices :
 - faire la décomposition en valeurs singulières pour obtenir $\hat{\sigma}_k$, \hat{u}_k et \hat{V}_2 (11)
 - construire \hat{Z} (27) et évaluer \hat{Q} (22-24) en utilisant \hat{u}_k , \hat{V}_2 , \hat{T}_u et \hat{T}_v .
 - former $\hat{\mu}$ (25) pour le comparer à t_r (26) le seuil fonction de r (16) le nombre de degrés de liberté et la probabilité de "fausse alarme" que l'on a choisie.

Les données sont générées à l'aide du système suivant considéré dans [6,7].

$$X_{t+1} = F X_t + B u_t + K e_t \quad (28)$$

$$y_t = H X_t + e_t$$

avec

$$F = \begin{bmatrix} 0 & 1 & 0 \\ .25 & 0 & .25 \\ 1 & 0 & 0 \end{bmatrix} \quad B = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$$K = \begin{bmatrix} .00547 & .063 \\ .119 & .157 \\ .674 & .000666 \end{bmatrix} \quad H = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

avec e_t du bruit blanc centré de variance $\Sigma = 0,2 I$ et u_t une suite scalaire aléatoire centrée prenant les valeurs ± 1 . En considérant une longueur d'observations de y_t de $T = 500$ et un ordre maximal a priori $\bar{n} = 4$ dans (9), on obtient typiquement les résultats suivants où k désigne le nombre de lignes dans le k-uple $S(k)$ et r le nombre de degrés de liberté (15) de μ (25) sous l'hypothèse que $H(S(k), 3)$ possède une valeur singulière nulle. Pour cet exemple, l'ordre exact est 3 et la structure correspondante $S(3) = (1,2,3)$.

k	S(k)	$\hat{\mu}$	r	décision
3	1,2,3	13.37	4	garder la dernière ligne
4	1,2,3,4	3.09	3	rejeter la dernière ligne
4	1,2,3,5	4.31	3	rejeter la dernière ligne

Dans la dernière colonne nous avons indiqué la décision découlant de l'application de la règle (26). Il faut, bien entendu, choisir a priori une probabilité de "fausse alarme" pour en déduire la valeur des seuils t_r . Mais dans le cas de cet exemple qui, d'après les résultats de [6,7], ne paraissait pas trivial, on voit que la structure estimée sera la bonne dans une large plage. Notons en plus que la charge de calcul de notre approche est sans commune mesure avec celle de la méthode proposée dans [6,7].

CONCLUSION

Nous avons proposé une procédure qui permet d'estimer la structure et l'ordre d'un processus stochastique vectoriel. Le test a été développé pour estimer les invariants de Kronecker (relatifs à l'observabilité) mais ils peuvent bien sur être utilisés comme des indices de structure dans des représentations à paramétrisation non minimale. Le test peut également être modifié pour estimer d'autres invariants correspondant notamment à d'autres manières de sélectionner les lignes (ou colonnes) indépendantes de la matrice bloc-Hankel.

Par ailleurs, si on désire uniquement obtenir une estimée de l'ordre n d'un processus vectoriel, l'approche proposée peut sembler longue. Une façon plus directe de procéder consiste alors à réaliser la décomposition en valeurs singulières d'une matrice carrée à M blocs $\hat{H}(M, M)$ unique et à développer un test qui permette de décider combien de valeurs singulières doivent être déclarées nulles. Un tel test peut être obtenu à partir de l'analyse que nous avons proposée mais de manière générale on observe sur les simulations que le saut entre la n -ième valeur singulière (la plus petite non-nulle) et la $(n+1)$ -ième (la plus grande "nulle") décroît quand l'ordre de la matrice considérée augmente. Dans la pratique, il est donc important de ne pas prendre un nombre de bloc-colonnes trop important.

REFERENCES

- [1] AKAIKE H. : "Canonical correlation analysis of time series and the use of an information criterion" in System Identification: Advances and Cases Studies, Mehra R. and Lainotis D. Editors, Academic Press, NY, 1976.
- [2] GEVERS M. and WERTZ V. : "Uniquely identifiable state space and ARMA parametrizations for multivariable linear systems", *Automatica*, vol. 20, pp. 333-347, 1984.
- [3] RISSANEN J. and LJUNG L. : "Estimation of optimum structures and parameters for linear systems", *Lect. Notes in Economic and Mathematics Systems*, vol. 131, Springer, pp. 75-91, 1975.
- [4] FUCHS J.J. : "Structure and order estimation of multivariable stochastic processes", soumis à IEEE T-AC, 1988.
- [5] VAN OVERBEEK A.J.M. and LJUNG L. : "One-line structure selection for multivariable state-space models", *Automatica*, vol. 18, pp. 529-544, 1982.
- [6] RISSANEN J. and WERTZ V. : "Structure estimation by accumulated prediction error criterion", 8th IFAC Symp. on Identification and System Parameter Estimation, York, UK, pp. 757-758, 1985.
- [7] RISSANEN J. : "Predictive and non predictive minimum description length principles", *Lecture Notes in Control and Information Sciences*, vol. 86, Springer, pp. 115-140, 1986.