On approximate system dynamic

Zaka Ratsimalahelo

LATEC

LABORATOIRE D'ANALYSE ET DE TECHNIQUES ÉCONOMIQUES

U.R.A. 342 C.N.R.S.

DOCUMENT de TRAVAIL

UNIVERSITE DE BOURGOGNE

FACULTE DE SCIENCE ECONOMIQUE ET DE GESTION

4, boulevard Gabriel - 21000 DIJON - Tél. 80 39 54 30 - Fax 80 39 56 48

ISSN : 0292-2002

n° 9616

On Approximate System Dynamic

Zaka RATSIMALAHELO*

décembre 1996

*LATEC - UMR 5601 (CNRS) et Université de Franche-Comté

ON APPROXIMATE SYSTEM DYNAMIC[1]

Zaka RATSIMALAHELO

Keywords: time series analysis; state space models, Hankel norm approximations, balanced realization, positive real lemma, unweighted principal components,

ABSTRACT.
In this paper concepts and techniques from system theory are used to obtain state-space (Markovian) models of dynamic economic processes instead of the usual VARMA models. In this respect the concept of state is reviewed as are Hankel norm approximations,and balanced realizations for stochastic models. We clarify some aspects of the balancing method for state space modelling of observed time series. This method may fail to satisfy the so-called positive real condition for stochastic processes. We us a state variance factorization algorithm which does not require us to solve the algebraic Riccati equation. We relate the Aoki-Havenner method to the Arun - Kung method.

1. Introduction.

Linear system approximation has found important applications in model reduction, system identification, spectrum estimation, among others. We will consider the problem of constructing a state-space model (Markovian representation) for a stochastic process from a finite number of estimated covariance lags. The approach is to first obtain a high-order model which exactly matches the estimated covariance sequence, and then use balanced model reduction techniques to obtain a lower order model which approximates the given sequence. The balanced models can then be obtained from a realization algorithm.

The stochastic realization problem deals with the quest for a finite dimensional Markovian representation for a stochastic process from the known covariance information. If the covariance of the intervening random variables are axactly known, then we deal with the exact stochastic realization problem, which has received great attention Akaike (1976), Faurre (1976). For many applications the Markovian representation or state space model may be too complex owing to its high dimensionality, thus barring efficient computational management.

This partially motivates the search for smaller dimensional Markovian realization which approximates the original one in some sense. The high dimensionality of the original state space model can, for instance, be caused by the incorporation of weakly coupled superfluous state components. These components may mask any tendencies hidden in the dynamics.

Another difficulty with the stochastic realization problem is the necessity of the exact covariance information. In most pratical situations, all one has available is an estimate of the covariances based on the real data (i.e., sample covariances). Not only would the noise

[1] This paper was presented at the Second International Conference on Computing in Economics and Finance, Geneva, Switzerland, 26-28 June 1996

fluctuations in the covariance structure lead to models of high dimensions, but what is more essential, the sample covariance sequence may not be positive-real. In such a case, the exact realization algorithm applied to inexact data may have no solution at all.
In the second section we present the different representations of a times series model, a link with the power density spectral is used. Next we recall, in the third section, the stochastic realization problem as background. Finally in the fourth section we clarify some problematic aspects of the method proposed by Aoki-Havenner (1991), we propose a factorization approach which is not used to solve the Riccati equation, and we relate the Aoki-Havenner algorithm to the unweighted principal components algorithm.

2. Different representations of a model

2.1. State-space representation.

Let y_t be a weakly stationary multivariate stochastic process which has a zero mean. A state space time series model can be written in two matrix equations, the state equation and the observation equation:

$$x_{t+1} = Ax_t + u_t \qquad (1.1)$$

$$y_t = Cx_t + v_t \qquad (1.2)$$

where $x \in R^n$, $y \in R^m$, A and C are matrices of appropriate dimensions, v_t and u_t are zero mean white noise with variance-covariance matrix

$$E\left[\begin{pmatrix} u_t \\ v_t \end{pmatrix}\begin{pmatrix} u_s \\ v_s \end{pmatrix}'\right] = \begin{bmatrix} Q & S \\ S' & R \end{bmatrix}\delta_{t,s};\ \forall t,s$$

(the superscript (') indicates the transpose of a vector and matrix and δ_{ts} is the kronecker delta function).
This model is in fact equivalent to the following innovation model

$$\begin{aligned} \hat{x}_{t+1} &= A\hat{x}_t + B_t e_t \\ y_t &= C\hat{x}_t + e_t \end{aligned} \qquad (2)$$

where $\hat{x}_t = E(\hat{x}_t / y_{t-1}, y_{t-2}, \ldots)$, is the optimal prediction estimate of x_t,
$e_t = y_t - E(y_t/y_{t-1}, y_{t-2}, \ldots)$ is the innovation process of y_t,
$B_t = E(x_t e'_t)E(e_t e'_t)^{-1}$ is Kalman gain,
$\Omega_t = E(e_t e'_t)$.

For large times the Kalman gain B_t and Ω_t reaches an asymptotic steady-state $B = \lim_{t\to\infty} B_t$ and $\Omega = \lim_{t\to\infty} \Omega_t$
In what follows, we consider the stationary case, thus the variance-covariance matrix associated with the system is:

$$E\left[\begin{pmatrix} Be_t \\ e_t \end{pmatrix}\begin{pmatrix} Be_t \\ e_t \end{pmatrix}'\right] = \begin{pmatrix} B \\ I \end{pmatrix}\Omega\begin{pmatrix} B \\ I \end{pmatrix}' \qquad (3)$$

The innovation model is of considerable practical value because of its statistical properties. It can be seen that the innovation model is determined by the same dynamic parameters (A, C) as the initial model. However, the triple (Q, R, S) of the initial model is statistically equivalent to the double (B, Ω) of the innovation model (see Ratsimalahelo (1994)).

2.2. ARMA representation.

It is known that there is a one-to-one correspondence between stationary ARMA and state space models, any system (1) and (2) can be expressed in ARMA and conversely, cf Hannan and Deistler (1988), Aoki and Havenner (1991).

Let a general vector ARMA(p,q)

$$a(z)\, y_t = b(z)\, \varepsilon_t \qquad (4)$$

where $a(z) = \Sigma A_i\, z^{-i}$ and $b(z) = \Sigma\, B_i\, z^{-i}$ are matrix polynomials in the lag operator z^{-1} of degrees p and q respectively, y_t is vector of data, t = 1,...,T and ε_t is a random error vector such that

$$E(\varepsilon_t) = 0\ ,\ E(\varepsilon_t \varepsilon_s') = \Sigma\delta_{ts},\quad \Sigma > 0.$$

In the stationary case, the ε_t are the innovations if and only if

$$\det a(z) \neq 0 \quad |z| \leq 1 \text{ (stability condition)}$$

and $\det b(z) \neq 0 \quad |z| < 1$ (miniphase condition).

Then the ARMA model may be rewritten

$$y_t = g(z)\, \varepsilon_t$$

with $g(z) = a(z)^{-1}\, b(z)$ is a matrix of rational function, g(z) is said to be the transfer function from ε_t to y_t.

If ε_t is the innovation sequence for y_t again g(z) is analytic for $|z| < 1$, (g(z) is analytic within the unit disc) $\det g(z) \neq 0 \ |z| < 1$.

If we denote the covariance of process y_t by $N_s = E(y_t y_{t+s}')$ then the power spectrum is given by the doubly infinite summation

$$S(z) = \sum_{t=-\infty}^{\infty} N_t z^{-t}$$

evaluated at z = exp(iw). Using the definition of the impulse response,

$$y_t = \sum_{j=0}^{\infty} g_j \varepsilon_{t-j}$$

$$g(z) = \Sigma\, g_j z^{-j},\ g_o = I.$$

It can be easily seen that the covariance N_s is obtained by

$$N_s = \sum_{j=0}^{\infty} g_j \Sigma g_{j+s} \quad s \geq 0$$

$$N_{-s} = N_s'$$

In z-transform language, this translates to

$$S(z) = g(z)\Sigma g(z^{-1})$$

it is known spectral factorization, cf Hannan and Deistler (1988).

2.3. The covariance function , the transfer function and state -space parameters

Definite the covariances by

$$P = E(x_t x'_t) = APA' + B\Omega B' \qquad (5.1)$$

$$M = E(x_{t+1} y'_t) = APC' + B\Omega \qquad (5.2)$$

$$N_s = E(y_t y_{t+s}') = CPC' + \Omega \quad s = 0 \qquad (5.3)$$

$$= CB^{s-1}M \quad s > 0 \qquad (5.4)$$

Using the z transform of equation (2), we obtain:

$$y(z) = [C(zI - A)^{-1}B + I\,]e(z)$$

where y(z) et e(z) are respectively the z transform process y_t and of the innovation process e_t.

Matrix $T(z) = C(zI - A)^{-1}B + I$ is termed the system transfer matrix. Its poles are the eigenvalues of A as $\det(zI - A) = 0$ is a pole . It is a rational fraction matrix wherein the degree of denominator of each item of the matrix exceeds the degree of the corresponding numerator.

2.4. Spectral factorization.
The stochastic realization problem is closely connected with the spectral factorization .

Lemma 1. Given a system of the form (2), the spectral density function S(z) of y_t is defined by

$$S(z) = \sum_{t=-\infty}^{\infty} N_t z^{-t}$$

or

$$S(z) = N_0 + C(zI - A)^{-1}M + M'(z^{-1}I - A')^{-1}C' \qquad (6)$$

where P, M and N_0 are uniquely determined by eqs (5.1) - (5.3).

The spectral density function S(z) may be expressed as a matrix product

$$S(z) = T(z)\Omega\, T(z^{-1}) \qquad (7)$$

with $T(z) = [C(zI - A)^{-1}B + I]$, T(z) is the transfer function matrix.
The spectrum is thus factorized as

$$S(z) = G(z)\, G'(z^{-1}) \qquad (8)$$

with $G(z) = T(z)\, \Omega^{1/2}$. The poles G(z) which are equal to the eigenvalues of A (asymptotically stable by the stationary condition of y_t) are within the unit disc, and the zeros are the eigenvalues of (A - BC). G(z) is analytical outside the unit circle
Replacing z by e^{iw} gives

$$S(e^{iw}) = G(e^{iw})\, G'(e^{-iw}) = T(e^{iw})\, \Omega\, T(e^{-iw}) \quad w \in [0, 2\pi]. \qquad (9)$$

It may be asked what type of information is contained in the spectral density function S(z).The spectrum S(z) provides an external representation of the process y_t, according to the matrix product expression, the spectrum S(z) is determined by the same set of parameters (A, B, C, Ω) as the innovation model.
S(z) is interpreted as an expression of the form

$$S(z) = J(z) + J'(z^{-1}) \qquad (10)$$

where $J(z) = N_0/2 + C(zI - A)^{-1}M$ is the transfer matrix of a linear dynamic system, J(z) is positive real. A link is then established with the problem of realisation of a transfer function.

Notice that, given a spectral density function of the form (6), if the triple (A, C, M) is a minimum system then the parameters A and C can be obtained from equation (6).

To have a minimum system we need the necessary and sufficient conditions provided by the following lemma.

Lemma 2. The stochastic dynamic system (2) admits a spectrum- minimal if and only if (A C) is observable and there are no solutions $x \neq 0$ to the following set of equations:

$$A'x = zx$$
$$(C(zI - A)^{-1}B + I)\, \Omega B'x = 0$$

The proof of this lemma is based on the relations (5). The second condition is closely related to the zeros of the transfer function $T(z) = I + C(zI - A)^{-1}B$.
The following corollary can thus be established.

Corollary 3. The stochastic linear dynamic system of the form (2) admits a minimum spectrum when (A, C, BΩ) is minimum and if

$$\sigma(A) \cap \{z^{-1} / z \neq 0, z \text{ is a zero of } T(z)\} = \varnothing.$$

($\sigma(A)$ denote the set of eigenvalues of the matrix A)

Lemma 4. Given a system of the form (1), the spectral density function S(z) of y_t is the form

$$S(z) = N_0 + C(zI - A)^{-1}M + M'(z^{-1}I - A')^{-1}C'$$

where P, M and N_0 are uniquely determined by

$P = APA' + Q$

$M = APC' + S$

$N_0 = CPC' + R.$

The spectral density function S(z) may be expressed as

$$S(z) = \begin{bmatrix} C(zI - A)^{-1} & I \end{bmatrix} \begin{bmatrix} Q & S \\ S' & R \end{bmatrix} \begin{bmatrix} (z^{-1}I - A')^{-1}C' \\ I \end{bmatrix}$$

The links between the different forms of representation for stochastic linear systems can be summarized in the following figure.

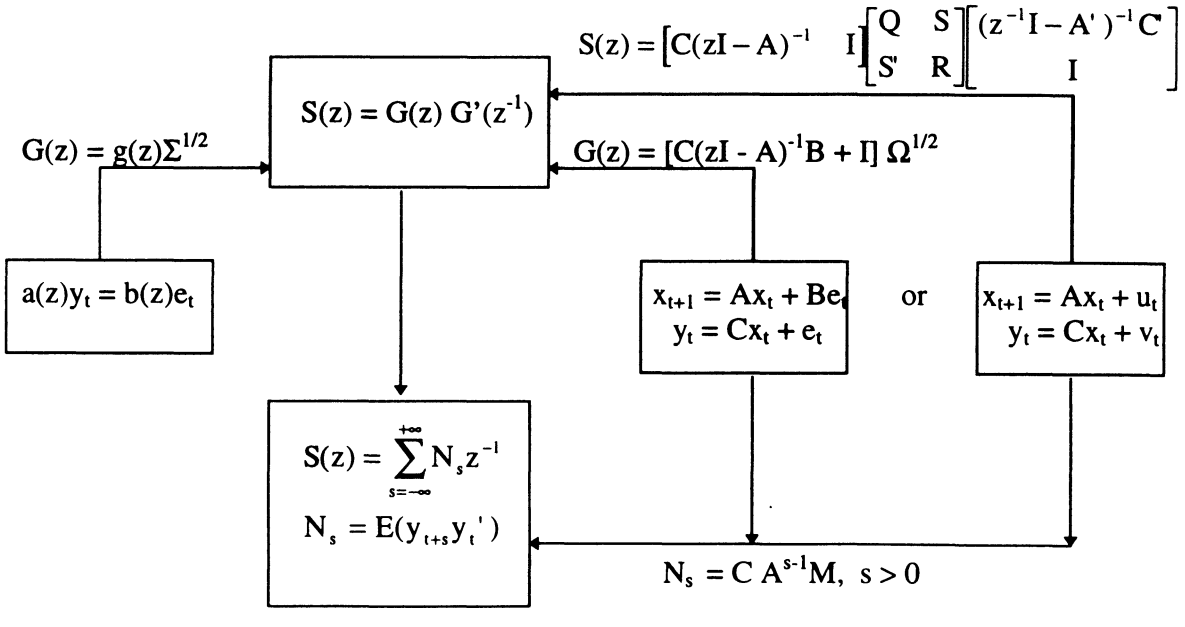

Fig 1. The different representations of stochastic processes.

In a stochastic process y_t representations in ARMA form, in the form of (spectrum) transfer functions, or using the autocorrelation function N_s, are termed external descriptions while a state representation is termed an internal description.

3. Stochastic realization problem.

The realization problem consists in determining an internal representation from an external representation of a system such that the state model obtained has the same input-output behaviour as that specified by the external representation.

Thus, for a deterministic model defined by

$$x_{t+1} = Ax_t + Bu_t$$
$$y_t = Cx_t + Du_t$$

where u is a deterministic command. The realization problem consists in determining a quadruple (A, B, C, D) from:

- the transfer function $G(z) = C(zI - A)^{-1}B + D$
- or Markov parameters $L_s = CA^{s-1}B \quad s > 0, \quad L_0 = D$

Similarly for a stochastic system, realization consists in determining the quadruple (A, C, B, Ω) or the quintuple (A, C, Q, R, S) from:

- the spectrum $S(z) = N_0 + C(zI - A)^{-1}M + M'(z^{-1}I - A')^{-1}C'$
- the autocorrelation function $N_s = CA^{s-1}M \quad s > 0.$

Remarks.

i) The realization problem is also known as the spectrum factorization problem, Lindquist and Picci (1979), Caines (1987) i.e. it seeks a transfer function G(z) such that $S(z) = G(z)G'(z^{-1})$.

ii) Notice the similarity between the following problems:

- deterministic case, determining a triple (A, B, C) from Markov parameters
  $L_s = CA^{s-1}B \quad s > 0,$
- stochastic case, determining a triple (A, M, C) from the autocorrelation function
  $N_s = CA^{s-1}M \quad s > 0.$

An essential difference is that every triple (A, B, C) describes a corresponding input-output system, but that not every (A, M, C) describes a corresponding stochastic process. This is expressed by the well-known positive real lemma.

Given the covariance sequence, one forms the (infinite) Hankel matrix which is factorizable into an observability matrix O and a controllability matrix K

$$\begin{bmatrix} N_1 & N_2 & N_3 & . & . \\ N_2 & N_3 & N_4 & . & . \\ N_3 & N_4 & N_5 & . & . \\ . & . & . & . & . \\ . & . & . & . & . \end{bmatrix} = \begin{bmatrix} C \\ CA \\ CA^2 \\ . \\ . \end{bmatrix} \begin{bmatrix} M & AM & A^2M & . & . \end{bmatrix} \quad (12)$$

$$H = O\,K$$

The time sequence is rational if and only if this Hankel matrix has finite rank (say n). It follows then from the deterministic realization theory Faurre (1976) that the order of any minimal Markovian representation of y_t is precisely n, and a triple (A, M, C) can be constructed such that

$$N_s = CA^{s-1}M \qquad s > 0,$$

where in order to have a Markovian representation, (internal representation) the following needs to be satisfied

$$\begin{bmatrix} P - APA' & M - APC' \\ M' - CPA' & N_0 - CPC' \end{bmatrix} \geq 0, \qquad P > 0$$

this is known as the positive real lemma.
Using the innovation model (2) and from eqs (5.1)-(5.3) we have

$$(B'\ I)'\ \Omega\ (B'\ I) \geq 0, \quad P > 0$$

This positivity condition is equivalent to the spectral factorization theorem, (see section 2.4) cf Faurre and al. (1979) Furthore this positive real condition is neglected in the balancing method as it does not arise for deterministic systems.
The triple (A, M, C) together with N_0 do not uniquely specify the covariances P, Ω and matrix B. However, P completely specifies Ω and B, and therefore characterizes the Markovian representation. Furthermore, note that any minimal realization of the covariance sequence is unique, modulo a similarity transformation.
This classical algorithm gives no indication of how a good reduced-order model can be obtained if an approximate stochastic realization is desired.

The notion of state.
We first need to define some more notation, denote by

$$y_t^+ = \begin{bmatrix} y_t \\ y_{t+1} \\ . \\ . \\ . \end{bmatrix} \qquad y_t^- = \begin{bmatrix} y_t \\ y_{t-1} \\ . \\ . \\ . \end{bmatrix} \qquad (13)$$

the future and the past, respectively, of y_t.
Let S (y_t^+) and S (y_t^-) be the Hilbert spaces obtained by taking the closed linear span of the random variables in y_t^+ and y_t^-, respectively. (In general, S (.) will be the Hilbert space generated by the closed linear span of elements of « . ».) The inner product on this space is the cross covariance. Orthogonal projection in these Hilbert spaces is equivalent to conditional expectation i.e. x \y = E(x \y) denotes the orthogonal projection of x onto S(y). This projection is given explicitly by

$$x \setminus y = E(xy)\ \{E(yy')\}^{-1} x \qquad (14)$$

where x and y are zero-mean random vectors.
Intuitively, the state of a linear system is a summary of the information in the past input history that is both necessary and sufficient to predict the future output . In the stochastic case, for the innovations representation, the state can also be interpreted as a as a summary of the past output history (instead of past input history) with regards to the prediction of the future output.
Construct a system of the following form

$$x_{t+1} = (A - BC)\ x_t + B y_t \qquad (15)$$
$$e_t = - C x_t + y_t$$

this is a formal inverse of the innovations representation and has all zeros within the unit circle i.e the eigenvalue of (A - BC) lie within the unit circle. Thus the state process x_t as well as the input e_t can be obtained causally from the output y_t using the above filter.
Let $A_* = A - BC$ is in linear system terminology the matrix of loop filter. The matrix A_* is the dynamic matrix in the Kalman filter for (15).

The state transition equation

$$x_t = [B \;\; A_*B \;\; A_*^2 B \;\; A_*^3 B \;\; \dots \;] \begin{bmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ . \\ . \end{bmatrix} \tag{16}$$

$$x_t = Sy_t^- \tag{17}$$

Thus the state of the innovations model is completely reproduceable from the infinite past y_t^-.
The state x_t of the innovations model (minimum-phase model) is also the state of the Kalman filter which is an estimate of the state every other non-minimum phase model with the same parameters A and C matrices estimated from the infinite past y^-. The state variance of the innovations model has to be smaller than the same A and C matrices.
Thus for the innovation model, beginning with the solution of the difference equation for the states equation and substituting into the observation equation produces

$$y^+ = Ox_t + Le_t^+ \tag{18}$$

where e_t^+ is defined the same way as y_t^+ and L is a block lower triangular matrix containing the Markov parameters

$$L_i = CA^{i-1}B \quad i > 0$$
$$= I \quad i = 0$$

and L is given by

$$L = \begin{bmatrix} I & 0 & 0 & . & 0 \\ L_1 & I & 0 & . & 0 \\ L_2 & L_1 & I & . & 0 \\ . & . & . & . & . \\ . & . & . & . & . \end{bmatrix}$$

Hence, from eq (17) we can write

$$y^+ = OSy^- + Le^+ \tag{19}$$

Since e_t^+ is independent of y_t^-

$$y^+ \setminus y^- = OS\, y^- = Ox \tag{20}$$

Thus, for the innovations model, we have

$$x_t = Sy_t^- \quad \text{and} \quad y^+ \setminus y^- = Ox \tag{21}$$

which means that the state of innovations model summarizes the past output history for predicting the future output.

The projection $y_t^+ \setminus y_t^-$ defines the state space of a Markovian model of the form (2). Thus this projection is central to any stochastic realization algorithm. From (14), it can be seen that this projection is given in terms of the following matrices formed from the covariance sequence:

$$y_t^+ \setminus y_t^- = H R^{-1} y^- \tag{22}$$

where $H = E[y_t^+(y_t^-)']$ and $R = E[y_t^-(y_t^-)']$ are the Hankel and Toeplitz matrices formed from the covariance lags of the output. H is given from eq(12) and R is given by

$$R = E[y_t^-(y_t^-)'] = \begin{bmatrix} N_0 & N_1 & N_3 & . & . \\ N_1 & N_0 & N_1 & . & . \\ N_2 & N_1 & N_0 & . & . \\ . & . & . & . & . \\ . & . & . & . & . \end{bmatrix} \tag{22}$$

Since $x_t = Sy^-$, in this case the state covariance matrix is

$$P = E(x_t x_t') = SRS'$$

A comparison eq (20) with $y_t^+ \setminus y^- = HR^{-1}y^-$ indicates that HR^{-1} must be factorizable into O and S , and consequently it must have rank equal to the size of the state vector.
Thus we have

$$x_t = Sy_t^- \text{ and } y_t^+ \setminus y_t^- = H R^{-1} y^- = Ox_t \tag{23}$$

The Hankel matrix can be factorised as $H = OK$, then the state vector as $x_t = Sy^-$, with $S = KR^{-1}$. Thus S is said to be a matrix aggregation. The state is considered by Aoki et Havenner (1991) as sufficient statistics.
In this case the state covariance matrix of the innovations model is

$$P = SRS' = K(R^{-1})'K' \tag{24}$$

The state is in fact the information interface between past and future and its dimension is equal to the order of the system. When the system has to be approximated by a lower order model, it is a question of compressing this information-interface into a lower dimensional which contains most of the information in the full-order state.

4. Stochastic balanced realization.

Balanced realizations originate in deterministic control theory, Moore (1981), this concept is well known in the literature. In words one may say that a balanced realization of a system has the property that the amount of controllability of a certain component of the state vector is equal to the amount of observability of this component. As shown by for instance Glover (1984) we can more or less consider the Gramians of a system as a tool to measure the controllability and observability of a realization.

4.1. Aoki-Havenner's Method.

For convenience we summarise Aoki-Havenner's (1991) method. We wish to predict future values of data vector y_t from the past values. Let the upper bound on number of lags needed to model y_t be p and the desired forecast horizon be f. Define the (mf x 1) and (mp x 1) future and past data vectors as $y_t^{+\prime} = (y_t', y_{t+1}', \ldots, y_{t+f-1}')$ and $y_t^{-\prime} = (y_t', y_{t-1}', \ldots, y_{t-p+1})$ and let $N_s = E(y_{t+s}y_t')$, s = 0, 1, 2 ...the autocovariance sequence (N_s is a m x m matrix)

Step 1. Calculate the sample autocovariances N_s , s = 0, 1, 2, ..., p+f

$$N_s = T^{-1} \Sigma\, y_{t+s}\, y_t'$$

where T is the sample size and arrange them in the Hankel matrix

$$H = \begin{bmatrix} N_1 & N_2 & . & . & N_p \\ N_2 & N_3 & . & . & N_{p+1} \\ . & . & . & . & . \\ . & . & . & . & . \\ N_f & N_{f+1} & . & . & N_{p+f-1} \end{bmatrix}$$

H has dimensions mf x mp.

The N_0 and N_{p+f} matrices do not appear in H but are used in estimating the A and B matrices.

Step 2. Obtain the singular value decomposition of the estimated Hankel matrix (HSV : Hankel Singular Value)

$$H = UDV' = \sum_{k=1}^{mf} \sigma_k u_k v_k'$$

where $\sigma_1 \geq \sigma_2 \geq \ ... \geq \sigma_{mf}$ are the singular values of the matrix H and u_k, v_k the corresponding eigenvectors. The matrices $U = (u_1,...,u_{mf})$ and $V = (v_1,...,v_{mp})$ are orthogonal, and $D = diag\ (\sigma_1,...,\sigma_{mf})$

Choose the importantly nonzero singular values to be included in the approximation. Since the singular values are scale-dependent, the series should be scaled so that the covariances are all of approximately the same magnitude. By the well known Kronecker theorem, cf Kailath (1980), Aoki (1987, 1990), Kung and Lin (1981) the rank of H is equal to the number of states (n) required to summarize the past history of the system. If the covariances that comprise H were the actual values rather than estimates , the largest n singular values would be nonzero with the remaining mf - n exactly zero, thus directly providing the order information critical to time-series modeling. Given the Hankel matrix with rank mf the problem is to find a Hankel matrix $\tilde{H}$ with $r(\tilde{H}) = n < mf$ such that

$$|| H - \tilde{H} \ || \text{ is minimum}$$

An (non-unique) approximant H_n of H with $r(H_n) = n$ which minimizes the Hankel norm is

$$H_n = U_n D_n V_n' = \sum_{k=1}^{n} \sigma_k u_k v_k'$$

By the singular value decomposition theorem we have

$$|| H - H_n ||_H = \sigma_{n+1}$$

The matrix H_n is in general not a Hankel matrix.
If the spectral norm of the L_2 is used to characterize the approximation we have

$$|| H - H_n ||_2 = [\sigma_{n+1} + \sigma_{n+2} + ... + \sigma_{mp}]^{1/2}.$$

The Hankel norm lies between the conventional L_2 and L_∞ norm. These norms provide a formal measure of the degree to which the time-series model reproduces the sample autocovariances.

Step 3. Estimate the parameters A, C, M by least squares based on the singular value decomposition.

Note that the covariance matrix between y_{t+1}^+ and y_{t-1}^- is

$$\bar{H} = O\bar{K} = OAK \qquad (24)$$

the first submatrix row of H, denoted by H_{1*}, is

$$H_{1*} = [N_1, N_2, ..., N_p] = CK \qquad (25)$$

and the first submatrix column of H, denoted by H_{*1}, is

$$H_{*1} = (N_1, N_2, ..., N_f]' = OM \qquad (26)$$

the system matrices A, C and M must be choosen to satisfy (24) through (26)

These equations can be solved easily by using the singular value decomposition of the Hankel matrix. Compare the singular value decomposition

$$H_n = U_n D_n V_n'$$

with the factorization in terms of O and K.(observability-controllability factorization)

$$(U_n D_n^{1/2})(D_n^{1/2} V_n') = OK$$

Theoretically, the number of positive singular values is equal to the theoretical rank of the Hankel matrix.

Choose a coordinate system in which

$$O = UD^{1/2} \text{ and } K = D^{1/2}V'$$

in this coordinate system we have

$$O'O = KK' = D$$

where O'O and KK' are, respectively, the observability and controllability gramians. Thus these system matrices are said to be balanced. Putting models in balanced representation or balanced form uniquely determines the system matrices, i.e., the model is uniquely identified. The word « balanced » in the above description comes from the fact that the representation balances two sources of errors: one related to the observability , i.e., error in reconstructing past state vector values from future observations, and the other errors related to reachability, i.e, the error due to past prediction errors.

The solutions:

$$C = H_{1*} V D^{-1/2} \qquad (27)$$

$$M = D^{-1/2} U' H_{*1} \qquad (28)$$

$$A = D^{-1/2} U' \bar{H} V D^{-1/2} \qquad (29)$$

Where $VD^{-1/2}$ and $D^{-1/2}U'$ are a singular value generalized inverse of O and K respectively so that

$$O^+ O = (D^{-1/2}U')(UD^{1/2}) = I$$

$$K K^+ = (D^{1/2}V')(VD^{-1/2}) = I$$

The balanced model representation uses the system matrices computed in eqs (27) - (29). It has several useful properties. One is that system matrices A, C and M of any lower-dimensional balanced model than a given one are obtained by merely taking a leading principal submatrix of the original dynamical matrix A and the corresponding submatrices C and M. Aoki and Havenner call this the nested property or orthogonality property . Another property is that these lower-dimensional models are all asymptotically stable if the original matrix A is asymptotically stable.

Step 4. Determination of the matrices B and Ω requires solving a certain matrix Riccati equation. From eqs(5) note the relations

$$P = APA' + BHB' \text{ and } \Omega = N_0 - CPC' \text{ with } \Omega > 0$$

From the definition of the matrix M we have

$$B\,\Omega = M - APC'.$$

From these we obtain the equation for P as

$$P = APA' + (M - APC')(N_0 - CPC')^{-1}(M - APC')' \qquad (30)$$

The smallest positive definite solution of (30), which exists under certain technical conditions (see Faurre et al. (1979)), is used to determine B and Ω.

Eq (30) can be solved by converting it into a (2nx2n) symplectic matrix Φ

$$\Phi = \begin{bmatrix} \Psi' - Q\Psi^{-1}D & Q\Psi^{-1} \\ -\Psi^{-1}D & \Psi^{-1} \end{bmatrix}$$

where $\Psi = A - MN_0^{-1}C$, $Q = C'N_0^{-1}C$ and $D = MN_0^{-1}M'$. This is a symplectic matrix, i.e.,

$$\Phi^{-1} = J'\Phi' J$$

with $J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \Leftrightarrow J^{-1} = -J = J'$

Given N_0 and A, M and C from eqs (27)-(29) then Ω and B are estimated by the following equations

$$\Omega = N_0 - CPC' \qquad (31)$$

$$B = (M - APC')(N_0 - CPC')^{-1} \qquad (32)$$

The Aoki - Havenner's method (called A-H) is summarized by

A-H's Algorithm.

1) Calculate the sample autocovariance N_s s = 0, 1, 2, ...f+p, form the covariance Hankel H

2) Determine approximate rank of Hankel matrix which specifies the number n of states, and using the singular value decomposition of the Hankel matrix $H_n = U_n D_n V_n'$

3) Factor H_n as $H_n = (U_n D_n^{1/2})(D_n^{1/2} V_n') = OK$ and estimate the parameters A, M, C by least squares based on the singular value decomposition

$$C = H_{1*} V D^{-1/2}$$

$$M = D^{-1/2} U' H_{*1}$$

$$A = D^{-1/2} U' \bar{H} V D^{-1/2}$$

4) Solve the Ricatti equation for P and determine the parameters B and Ω

$$B = (M - APC')(N_0 - CPC')^{-1}$$

$$\Omega = N_0 - CPC'$$

Hankel norm approximation.

The Hankel norm approximation problem was first shown to be solvable (in discrete time domain) by Adamyan, Arov and Krein (1971) and the first reliable state space algorithm was developed by Glover (1984, 1989) (in the continuous time domain). Many other authors have contribued to this field, among which Kung and Lin (1981), Ball and Ran (1987).

Step 2 concerns the approximation of a high dimensional system by one containing fewer states, in such a way that norm of the difference between the corresponding Hankel operators is minimal. This involves the approximation of estimated Hankel matrix by Hankel matrix of

smaller rank n. This « numerical rank » is determined by the number of « dominant » singular values of H. In this case the SVD can be partioned as folllows:

$$H = \begin{bmatrix} U & \overline{U} \end{bmatrix} \begin{bmatrix} D & 0 \\ 0 & \overline{D} \end{bmatrix} \begin{bmatrix} V' \\ \overline{V}' \end{bmatrix}$$

where D contains the n dominant singular values and the singular vectors are partioned accordingly. Then the optimal (in Frobenius norm) rank n approximation to H is

$$H_n = U_n D_n V_n' = OK$$

H_n is not a Hankel matrix in general. Thus the singular value approximation of step 2 is simply not optimal in the sense of the Hankel norm for linear systems.
This clarifies the incorrectness of the theorem on specification efficiency in Aoki and Havener (1991), as well as the corresponding reply of Havenner to comments of Deistler and Mittnik (1991) on this point.
A possible modification of the algorithm would be to determine an optimal Hankel norm approximation error σ_{n+1} while the Hankel structure is preserved. For this modification the theorem on specification efficiency would hold true.

We saw in section 3 that stochastic realization must satisfy the positivity condition, this is a basic problem in spectrum analysis cf Byrnes and Lindquist (1994). This positivity constraint is neglected in Aoki-Havenner's method and it can cause problems for the algorithm. The estimates in step 3 may be such that the positivity condition is not satisfied, in which case no model can be determined. This can be checked in terms of the characteristic roots of a symplectic matrix arising in the algorithm.

Symplectic matrix and unit root.
The eigenvalues of symplectic matrix Φ are not only symmetric with respect to the real axis but also with respect to the unit circle. Thus the algorithm of Vaughan fails only when Φ has a unit root i.e., an eigenvalue on the unit circle.

Symplectic Unit Root theorem. (Heij et al. (1992))
If Φ has an eigenvalue at $z = \exp(iw)$, then either (A, C, M, N_0) does not satisfy the positivity condition or spectrum is singular at $z = \exp(iw)$
As shown by Hannan and Poskitt (1988) on the unit circle the spectrum has a rank deficiency i.e. singularity, this is related to the existence of deterministic components in the process. For stationary ARMA processes singularity on the unit circle is equivalent to unit roots in the MA part, i.e., it is equivalent to non-invertibility.
Heij et al (1992) are shown that the symplectic matrix Φ has a unit root if and only if the power density spectral S(z) has a zero on the unit circle. Then if this root is + 1 or - 1 this corresponds to a deterministic component. If the unit root is not real then the positivity condition is violated because $S(1)S(-1) < 0$. This case causes problems for Aoki-Havenner's algorithm.

4.2. Factorization Approach.
Aoki-Havenner's algorithm is characterized by the fact that it requires the solution to an algebraic Riccati equation. We propose an algorithm which is representative of a class of algorithms which do not require a Riccati solution but only a matrix factorization to obtain the stochastic balanced realization.
It will be recalled that the state vector of any innovation model must always be a linear function of y_t^- then we have $x_t = Sy_{t-1}^-$ and $y_t^+ \setminus y_{t-1}^- = H_n R^{-1} y_{t-1}^- = Ox_t$

The aggregating matrix S is

$$S = KR^{-1}$$

In the balanced realization form, the coordinate system is choosen so that $O = UD^{1/2}$ and $K = D^{1/2}V'$. hence the Hankel norm method summarizes past data in the state vector

$$x_t = D^{1/2}V'R^{-1}y_t^-$$

The aggregating matrix $S = D^{1/2}V'R^{-1}$, given the model specification, as the aggregating matrix has full row rank the information in y_{t-1}^- is preserved, and the states are sufficient statistics for information in the data.

The state covariance is

$$P = F'F \text{ with } F = R^{-1/2}VD^{1/2} \text{ (where } R^{-1} = R^{-1/2}(R^{-1/2})')$$

We propose a method which is not used to solve the Riccati equation to find B and Ω estimate.

Proposition.

The estimates B and Ω are obtained by choosing the state vector as $x_t = D^{1/2}V'R^{-1}y_{t-1}^-$, the state covariance matrix is $P = F'F$ with $F = R^{-1/2}VD^{1/2}$.

In this proposition the state covariance matrix is specified in terms of the factorization, thus no Riccati equation has to be solved, then P can be used directly to find B and Ω. This method modified step 4 of Aoki- Havenner's method which can be summarized by

Algorithm factorization or A-H's modified Algorithm.

1) Calculate the sample autocovariance N_s s = 0, 1, 2, ...f+p, form the covariance Hankel H
2) Determine approximate rank of Hankel matrix which specifies the number n of states, and using the singular value decomposition of the Hankel matrix $H_n = U_nD_nV_n'$
3) Factor H_n as $H_n = (U_nD_n^{1/2})(D_n^{1/2}V_n') = OK$ and estimate the parameters A, M, C by least squares based on the singular value decomposition
4) Calculate the state covariance matrix $P = F'F$ and estimate Ω and B.

$$\Omega = N_0 - CPC'$$
$$B = (M - APC')\,\Omega^{-1}.$$

In this approach the state covariance matrix is also specified in terms of factorization. Thus no Riccati equation has to be solved, but P can be used directly to determine B and Ω.

4.3. Arun-Kung's Method.

The state vector can be introduced in yet another way. Recall from eq (22) that the state space of stochastic realization is spanned by $HR^{-1}y_{t-1}^-$. If H is of rank mp, then the dimension of the state space is mp. In order to obtain an n-th order model, $n < mp$, it is necessary to find a matrix S such that Sy_t^- nearly spans the state space. Arun and Kung (1986) have suggested choosing S to minimize the following criterion:

$$\text{Var}\,[y_t^+ - y_t^+ \setminus x_t].$$

Such a criterion is called the predictive efficiency criterion which was first used by Rao (1964) in multivariate statistics for the 2-vector problem.

Noting that

$$y^+ \setminus x = HS'(SRS')^{-1}\,Sy_t^-$$

and

$$\text{Var}\,[y_t^+ - y_t^+ \backslash x] = R - HS'(SRS')^{-1}SH' \tag{33}$$

thus we have

$$\min \text{tr}\,[R - HS'(SRS')^{-1}SH'] = \max \text{tr}\,[(SH'HS')\,(SRS')^{-1}]. \tag{34}$$

(tr (A) denotes trace A)

The solution to this optimization problem is: the n rows of S must be a basis for the space spanned by the n generalized eigenvectors corresponding to the n largest generalized eigenvalues of the matrix pencil ($\lambda R - H'H$).

The approximation is based on the eigenvalues values of $HR^{-1}H'$

$$HR^{-1}H' = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1^2 & 0 \\ 0 & \Sigma_2^2 \end{bmatrix} \begin{bmatrix} U_1' \\ U_2' \end{bmatrix} \tag{35}$$

where Σ_1^2 contains the n largest generalized eigenvalues i.e. the principal components, then the predictive efficience criterion is optimized by choosing S to be

$$S = EU_1'HR^{-1} \tag{36}$$

where E is any nxn invertible matrix. Different choices of E will correspond to different coordinate transformations of the state vector. This solution is called the unweighted principal components (UPC).

The state vector will be in balanced coordinates by choosing E by $\Sigma_1^{-1/2}$

$$S = \Sigma_1^{-1/2}U_1'HR^{-1} \tag{37}$$

which implies that the observability matrix is $O = U_1\Sigma_1^{1/2}$ and the observability grammians, which is O'O is equal to Σ_1

The state covariance is

$$P = SRS' = \Sigma_1 \tag{38}$$

The UPC Algorithm.

Step 1. Perform an eigendecomposition of

$$HR^{-1}H' = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1^2 & 0 \\ 0 & \Sigma_2^2 \end{bmatrix} \begin{bmatrix} U_1' \\ U_2' \end{bmatrix}$$

and retain only the principal components. The aggregating matrix S is

$S = EU_1'HR^{-1}$ where E is any nxn invertible matrix.

In balanced coordinates

$$S = \Sigma_1^{-1/2}U_1'HR^{-1}$$

The observability matrix is

$$O = U_1\Sigma_1^{-1/2}$$

and the observability grammian $O'O = \Sigma_1$

Step 2. The parameter estimation step is taken from the deterministic algorithm

$$A = O_1^+ O_2$$

where O_1 denotes the first m rows of O and the superscript $^+$ stands for the pseudoinverse and O_2 the last m rows of O.

C = first row of O

B = first column of S

$P = \Sigma_1$

$M = N_0 - C\Sigma_1C'$

An advantage of UPC algorithm is that it can be applied to sample covariance sequence that are not positive definite. This condition often arises when the covariance sequence is estimated from a short data record.

To relate the UPC algorithm to the Aoki-Havenner's algorithm , it is necessary to work with the « square root »of (35)

$$HR^{-1/2} = U\Sigma V' \text{ (where } R^{-1} = R^{-1/2} R^{-1/2})$$

which yields the following factorization of the Hankel matrix:

$$H = U\Sigma V' R^{-1/2} = OK$$

The resulting stochastic realization can be obtained from step 3 of the Aoki-Havenner's algorithm.

Note that Step 4 of A-H's algorithm provides an explicit formula for Ω, while the UPC algorithm does not.

Comparison to the Canonical Correlations Criterion and Akaike's Method.

Arun and Kung (1986) point out that canonical realization algorithm is not well suited for model reduction due to the smallness of the canonical correlations and the fact that it works with unapproximated data.

Akaike (1974) uses Hotelling's concept of canonical correlations coefficients. This method optimizes mutual information in the past and future observations. The method introduces canonical variables by normalizing the data vector by the inverse of square root of the covariance matrix of the data data vector. The selection of the state vector is based on the singular value decomposition of matrix $R^{-1/2} H (R^{-1/2})'$, where H is the same Hankel matrix of the covariances of the future and past observations and the covariance matrix R is expressed as $R = R^{1/2}(R^{1/2})'$ where $R^{1/2}$ is any full-rank factorization R. Thus, this Akaike's solution is different from Arun- Kung's solution because under perturbations, H will be full rank and the principal components of $HR^{-1}H'$, $R^{-1/2} H (R^{-1/2})'$, and H will not be same. In addition the Akaike's method use the matrix $R^{1/2}$ for normalize the data vector by $Z^+ = R^{-1/2}Y^+$. This indicate that the state vector component by Akaike's method do not do the best of predicting the future. It is inherent normalization in the canonical correlation criterion that causes it to ignore the strength of modes in the process. Following C.R. Rao the generalized principal component s analysis for studing the association between two randon vectors is different from Hotelling's canonical correlation analysis.

References.

Adamyan, V. M., D. Z. Arov and M. G . Krein (1971). « Analytic properties of Schidt pairs for Hankel operator and the generalized Schur-Takagi problem. Math. » USSR Sbornik, vol 15, no.1, pp 31-73.

Akaike, H., (1974a) « A New Look at Statistical Model Identification, » IEEE Trans. Automatic Control, AC-19, 6, pp 716-723.

Akaike, H., (1974b) « Markovian Representation of Stochastic Processes and its Application to the Analysis of Autoregressive Moving Average Processes, » Annals of Institute of Statistical Mathematics, 26, pp363-387.

Akaike, H., (1976) « Canonical Correlation Analysis of Time Series and Use of an Information Criterion, » in System Identification : Advances and Case Studies, Chap 2, R. K. Mehra and D. G. Lainiotis (Eds.), Academic, New York.

Anderson , B.D.O., and Moore, J. B., (1979) « Optimal Filtering Information and Systems Sciences Series, Prentice-Hall, Englewood Cliffs, N. J.

Aoki, M., (1987) « State space modelling of time series », Springer -Verlag, Berlin.

Aoki, M. and Havenner A. (1989) « A method for approximate representation of vector-valued time-series and its relation to two alternatives » Journal of Econometrics, 42, pp 181-199.

Aoki, M., (1990) « State space modelling of time series », 2nd edition, Springer Verlag, Berlin.

Aoki, M., and Havenner, A., (1991), « State space modeling of multiple time series », Econometric Reviews, 10, 1-59, 93-99.

Arun, K. S. and Kung, S. Y., (1986) « Generalized Principal Component Analysis and its Application of stochastic Processes, » in Modelling and Application of Stochastic Processes, Chap. 4,U. B. Desai (Ed.), Kluwer Academic, Boston.

Ball, J. A. and A. C. M. Ran (1987). Optimal Hankel norm model reductions and Wiener-Hopf factorization I: The canonical case. Siam J. contr. and optim. vol 25-2, 362-382.

Byrnes, C. I., and Lindquist A. (1994), « Toward a solution of the minimal partial stochastic realization problem », Comptes Rendus des Academies de Sciences t. 319, Série I, pp 1231-1236

Caines, P. E. (1988): Linear Stochastic Systems, New York: John Wiley and Sons.

Desai, U. B. and Pal, D., (1984) « A Transformation Approach to Stochastic Model Reduction, » IEEE Trans. Automatic Control A-C 29, 12, pp. 1097-1099.

Deistler, M. (1989). « Linear system identification- a survey ». In: From Data to Model, J. Willems (Ed.). Springer-Verlag, Berlin, 1-25.

Deistler, M., Dorfman, J. H., Mittnik, S. and Zellner, A., (1991) « Comment on state space modeling of multiple time series », Econometric Reviews, 10, 61-65, 67-73, 75-90, 91-92.

Faurre, P. (1976). « Stochastic realization algorithm » in System Identification : Advances and Case Studies, R. K. Mehra and D. G. Lainiotis (Eds.), Academic, New York.

Faurre, P., M. Clerget, et F. Germain (1979) « Opérateurs rationnels positifs » Dunod, Paris.

Kailath, T., (1980), « Linear systems » Prentice Hall, Englewood Cliffs, NJ.

Glover, K. (1984). All optimal Hankel norm approximations of linear multivariable systems and their L_∞ - error bound. Int. J. Contr. 39-6, 1115-1193.

Glover, K. (1989). « A tutorial on Hankel-norm Approximation. » In: From Data to Model, J. Willems (Ed). Springer-Verlag, Berlin, 26-48.

Hannan, E. J., and Deistler M. (1988). « The statistical theory of linear systems » John Wiley New-York.

Kung, S. Y., and Lin, D. (1981) « Optimal Hankel-Norm Model Reduction: Multivariable Systems », IEEE Trans. Automatic Control, A-C 26,

Linquist, A. and Picci, G. (1979) « On the stochastic realization problem, » SIAM J. Contr. Optimiz, vol 17, pp 365-389.

Mittnik, S. (1989). « Multivariate time series analysis and state space models, » Comp. Math. Applications, 17, pp 1189-1201.

Moore, B.C., (1981). « Principal component analysis in linear systems: controllability, observability, and model reduction ». IEEE Transaction on Automatic Control 26, pp 17-32.

Rao, C. R. (1964): « The Use and Interpretation of Principal Component Analysis in Applied Research » Sankhya Series A, 26, pp 329-358.

Ratsimalahelo Z. (1994) « Théorie de système et Séries Temporelles » Document de travail L.AT.E.C. N° 9401.

Vaccaro, R. J., (1985) « Deterministic Balancing and Stochastic Model Reduction, » IEEE Transaction on Automatic and Control, Vol AC- 30, n° 9, pp. 921-923.