CARLO GRILLENZONI (*)

Orthogonal operators in dynamical stochastic systems


1. INTRODUCTION

In the analysis of stochastic systems, the transfer functions which connect input and output are usually specified by “rational polynomials”; that is by a class of fairly general operators, whose properties have been well investigated in mathematics. In the applied context, however, the relationships between these functions are often studied from the point of view of the particular problem to be tackled.

In this work, similarly to Piccolo (1984), we consider the set of polynomial functions in terms of metric space. This enables us to analyze the properties of product and sum of rational operators with the general framework of the length, distance, angle, convex set, etc. Moreover, by means of the transformation which maps each polynomial into the vector of its coefficients, the metrics may be redefined in Euclidean sense, obtaining a suitable operative meaning. The final result is that the conditions of orthogonality in-the-product and stability of-the-sum, of system operators, can be directed to the “observable structure” of their polynomials (orders, delays, roots).

Properties of orthogonality, in particular, have important consequences in simplifying the representations of rational distributed lags.

(*) Dip. di Economia Politica, Università di Modena e Dip. di Scienze Statistiche, Università di Padova.
models, and in turn on the algorithms of identification, estimation and forecasting. The performance of the proposed metrics and/or their ability to characterize the desired properties can be empirically checked.

2. THE SPACE OF POLYNOMIALS

Let \( P(x) \) be the set of linear polynomials \( p(x) \) of degree \( n \) with real coefficients

\[
P(x) = \{ p(x) : (p_0 + p_1 x + \ldots + p_n x^n) ; p_j, x \in \mathbb{R} \}
\]

\( P(x) \) is clearly a linear vector space over the scalar field \( \mathbb{R} \), since it is closed under the usual operations of direct sum and scalar product:

1) \( p(x), q(x) \in P(x) \rightarrow p(x) \oplus q(x) \in P(x) \)

2) \( p(x) \in P(x), \alpha \in \mathbb{R} \rightarrow \alpha \cdot p(x) \in P(x) \)

A basis in this space in given by \( x' = (1, x, \ldots, x^n) \), if \( n = \infty \) the dimension is infinite and the class \( P(x) \) coincides with that of rational polynomials \( R(x) = \{ p(x) / q(x) \} \).

The analysis of the relationships between the elements of \( P(x) \) is developed with the definition of measures of distance, length, angle, etc. Given the continuity of the functions \( p(x) \) in \( x \), we have:

\[
distance \quad d_k(p, q) = \left[ \int_a^b |p(x) - q(x)|^k \, dx \right]^{1/k}, \quad k \geq 0
\]

\[
angle \quad (p, q) = \int_a^b p(x) q(x) \, dx
\]

\[
length \quad \|p\|_k = \left[ \int_a^b |p(x)|^k \, dx \right]^{1/k}
\]

these quantities, satisfying the usual Triangular, Schwarz, Minkowski, Holder inequalities, make \( P(x) \) a metric and/or normed vector space
(see Courtain and Pritchard, 1977). However, the above metrics characterize local properties (relative to the choice of \([a, b]\)), and in passing to the limit they become indefinite \((+ \infty - \infty)\) and useless. Moreover, translation to the complex field with, e.g.,

\[
(p, q) = \int_{-\pi}^{+\pi} p(e^{-i\omega}) q(e^{+i\omega}) \, d\omega
\]

does not improve the situation because once again, we reason in local terms (i.e. on the unit circle \(|e^{\pm i\omega}| = 1\)).

Thus, it is necessary to define a transformation which maps \(P(x)\) into a vector space equivalent with respect to the properties (\(i, ii\)), but over which the metrics can assume a suitable operative meaning. A transformation of this type is that of \(P(x)\) into \(\mathbb{R}^{n+1}\) given by

\[
T[p(x)] = T[p'] = p' = (p_0, p_1, \ldots, p_n)
\]

where \(p\) provides the coordinates of \(p(x)\) in the basis \(x\). \(T(\cdot)\) is clearly isometric because

\[
T[\alpha p(x) + \beta q(x)] = T[\alpha p' + \beta q'] = (\alpha p + \beta q')
\]

but is not isomorphic (it does not preserve the properties of distance, length, etc.), if the metric are redefined in euclidean sense as

\[
d_2(p, q) = [\sum_{j=0}^{n} |p_j - q_j|^2]^{1/2}
\]

\[
(p, q) = \sum_{j=0}^{n} p_j q_j = p' q
\]

\[
\|p\|_2 = (p, p)^{1/2}
\]

Proceeding in this way however, the relationships between the polynomials are directly referable to their "observable structure".

Now, consider in \(P(x)\) the set of monic polynomials \((p_0 = q_0 = 1)\). This set is not a subspace since it is not closed (a sum of monic polynomials do not set up a monic polynomial) and \(T(\cdot)\) associates a set in \(\mathbb{R}^n\). In this context, we wish to investigate the structure of the subset of \(p(x)\) with roots lying outside the unit circle (stability region), and answer to the question whether this subset
forms a group. That is: Does a sum of monic polynomials with “stable roots” still have stable roots? In the sequel we heuristically show that this holds only if the degree \( n < 3 \).

Indeed, let \( \{ p_i (x) \}_{i=1}^{m} \) be \( m \) monic polynomials of degree \( n \) and roots outside the unit circle. This means that the vectors \( p_i' = (p_{i_1}, ... , p_{i_n}) \) are contained in the region of stability \( S_n \subset \mathbb{R}^n \). Their sum generates a non monic polynomial \( p (x) = \sum_{i=1}^{m} p_i (x) = m + \sum_{i=1}^{m} (\sum_{j=1}^{n} p_{ij}) x^j \) whose zeros, however, coincide with those of the monic \( q (x) = p (x) / m \). Now, since the corresponding vector of coordinates \( q = \sum_{i=1}^{m} p_i / m \) is an arithmetic mean (i.e. a convex combination) of the coefficients \( p_i \), we conclude that \( q \in S_n \) (and so \( p (x) \) is stable), only if \( S_n \) is convex. But from Anderson (1975) we have

\[
S_1: \quad |p_1| < 1 \quad S_2: \quad p_2 + p_1 < 1 \quad S_3: \quad p_2 + p_1 + p_3 < 1
\]

\[
p_2 - p_1 < 1 \quad p_2 - p_1 - p_3 < 1
\]

\[
|p_2| < 1 \quad |p_3| < 1
\]

since \( S_3 \) is bounded by nonlinear constraints it may be non-convex, and graphically we have Figure 1. For \( n > 3 \) it is difficult to characterize the structure of \( S \), although Piccolo (1982) have derived the exact measure of its size.

3. Orthogonality and estimation

The question of orthogonality between the operators of dynamical systems has been discussed by Priestley (1983), in the context of the transfer function (TF) models of Box and Jenkins (1976, Part III). This class of systems is fairly general and provides a parsimonious representation of stationary processes \( \{ y_t, x_t \} \) with rational spectral densities

\[
y_t = \frac{\omega_t (B)}{\delta_t (B)} x_{t-b} + \frac{\theta_{y} (B)}{\phi_p (B)} a_t, \quad a_t \sim IN (0, \sigma^2) \tag{3.1a}
\]

\[
\bar{\phi}_p (B) x_t = \bar{\theta}_{\bar{y}} (B)e_t, \quad e_t \sim IN (0, \tilde{\sigma}^2) \tag{3.1b}
\]
In (3.1a), \((y, x, a)\) are the output the input and the disturbance, \((B, b)\) denote the lag operator and the delay factor \((B^b x_i = x_{i-b})\); moreover \((\delta, \omega, \phi, \theta)\) are linear polynomials of degrees \((r, s, p, q)\) \(< \infty\). Some restrictions are needed in order to ensure the structural identifiability (and so the estimability) of the system; in particular

\[
\begin{align*}
\text{i) } & \ [\delta(0), \phi(0), \theta(0), \bar{\phi}(0), \bar{\theta}(0)] = 1, \quad |\omega(z)| < \infty \\
\text{ii) } & \ [\delta(z), \phi(z), \theta(z), \bar{\phi}(z), \bar{\theta}(z)] \neq 0, \quad |z| \leq 1
\end{align*}
\]

that is the monic polynomials are stable and the non monic one has bounded coefficients. Under these conditions we may conceive a convergent linear representation of (3.1a) by means of the Taylor expansions below, in which \(|v_0| < \infty\), \(\psi_0 = \bar{\psi}_0 = 1\)
$$v(B) = \frac{\omega(B)}{\delta(B)} = \sum_{i=0}^{\infty} v_i B^i, \quad \tilde{\psi}(B) = \tilde{\delta}(B) = \sum_{j=0}^{\infty} \tilde{\psi}_j B^j$$

Let us finally recall the parametric expressions of the covariance functions

**Cross**

$$\gamma_{xy}(B) = v(B) B^b \gamma_{xx}(B)$$  \hspace{1cm} (3.2a)

**Auto**

$$\gamma_{xx}(B) = \tilde{\psi}(B) \tilde{\psi}(B^{-1}) \delta^2$$  \hspace{1cm} (3.2b)

these will be utilized in the sequel as spectral densities with $B = z = e^{-i\omega}, \omega \in [-\pi, +\pi]$.

Now, the problem of Priestley (1983) was that of establishing the conditions under which the estimates of $v(B)$, yielded by the minimization of the functionals

$$J_1(v) = \sum_{t=1}^{N} n_t^2, \quad n_t = y_t - v(B) x_{t-b}$$  \hspace{1cm} (3.3a)

$$J_2(v, \psi) = \sum_{t=1}^{N} a_t^2, \quad a_t = n_t / \psi(B)$$  \hspace{1cm} (3.3b)

are parametrically equivalent. In particular, he recognized that “in order to have $\tilde{\nu}_1(B) \equiv \tilde{\nu}_2(B)$, one must appeal to some form of orthogonality property between the functions $v(B), \psi(B)$” (Priestley 1983, p. 277).

The separability of the estimates of $v(B), \psi(B)$, induced by their orthogonality, have useful consequences on many grounds. We point out the simplification of the algorithms of calculation, the improvement of the efficiency of estimates, the reliability of the control rules $x_i \rightarrow y_i$, etc. An important application of these properties occurs in the identification of the orders of the system $(r, s, b, p, q)$. Recently, Poskitt (1989) has outlined a selection strategy based on the minimization of the criterion

$$BIC(r, s, p, q | b) = \log \frac{J_2}{N} + (r + s + 1 + p + q) \log \frac{N}{N}$$

with maximum likelihood. The computational problems are clearly hard; however, if $\tilde{\nu}_1 \equiv \tilde{\nu}_2$, then one can identify $v(B)$ with $BIC(r, s | b) = \log (\tilde{J}_1 / N) + (r + s) \log N / N$ and least squares methods.
The general condition established by Priestley (1983) so that 
\[ \hat{v}_1 \equiv \hat{v}_2 \], was

\[
\begin{bmatrix}
\gamma_{xx}(z) \\
\tilde{\Psi}(z^{-1})
\end{bmatrix} = \sum_{k=-\infty}^{-1} \alpha_k z^k \equiv 0
\]

where \([ \cdot ]_-\) denotes the “forward” expansion. This condition is certainly satisfied if \( \{x_t, n_t\} \) are uncorrelated; but this is a typical feature of open-loop systems, where \( \gamma_{xy}(k) = 0 \), \( k < b \geq 0 \), \([\gamma_{xy}(z)]_- \equiv 0 \). To show this situation for the model (3.1), and its pendant in terms of orthogonal operators, note that for \( \Psi(\cdot) \) fixed, (3.2) and (3.3) imply

\[
\hat{v}_1(z) = \frac{\gamma_{xy}(z)}{\gamma_{xx}(x) z^b}, \quad \hat{v}_2(z) = \frac{\gamma_{xy}(z)}{\gamma_{xx}(x) z^b}
\]

where \( \bar{y}_i = y_t / \Psi(B), \bar{x}_i = x_t / \Psi(B) \) in (3.3b). Now, since

\[
\gamma_{xy}(z) = \frac{\gamma_{xy}(z)}{|\psi(z)|^2}, \quad \gamma_{xx}(z) = \frac{\gamma_{xx}(z)}{|\psi(z)|^2}
\]

we have \( \hat{v}_1 \equiv \hat{v}_2 \) only if \([\gamma_{xy} / \gamma_{xx}] \) is a backward transform, i.e. the feedback \( y_t \rightarrow x_t \) does not occur. Note finally that \( \tilde{\Psi}(z) = \sqrt{\gamma_{nn}(z)} \sigma \), in terms of spectral factorization theorem.

Although previous results are “purely algebraic” ones (Priestley 1983, p. 285), the question of the separability was already introduced by Pierce (1972) in statistical terms. Indeed, he showed that the maximum likelihood estimates \( \hat{\nu}_N, \hat{\psi}_N \) are asymptotically independent in absence of feedback. This fact can easily be seen in the Gauss-Newton estimator, by computing its gradient

\[
\frac{\partial \alpha_t}{\partial \beta} = \left[ \frac{\phi(B)}{\theta(B) \delta(B)} \right] (m_{t-1} \ldots m_{-r} x_{t-b} \ldots x_{t-b-s}),
\]

\[
\frac{1}{\theta(B)} (n_{t-1} \ldots n_{t-p} a_{t-1} \ldots a_{t-q})
\]

(3.4)

\[
\beta' = [\delta_1 \ldots \delta_r, \omega_0 \ldots \omega_s, \phi_1 \ldots \phi_p, \theta_1 \ldots \theta_q] = [\beta_1, \beta_2], \quad m_t = y_t - n_t
\]
Thus, since \( (x_t, a_t) \) are independent, the same is true for \( m_t = \nu (B) x_{t-b} \), \( n_t = \psi (B) a_t \), so that the dispersion of the nonlinear estimator
\[ E (\xi_t \xi_t')^{-1} \sigma^2 \]
is block diagonal.

This remark enables us to state that the estimation of \( \nu (B) \) in
(3.3a), cannot be carried out with the least squares method applied to
the linearized model
\[ \delta_r (B) y_t = \omega_r (B) x_{t-b} + \eta_t, \quad \eta_t = \delta_r (B) n_t \]

4. ORTHOGONALITY AND REPRESENTATION

The problem of orthogonality early posed by Priestley (1983)
could be summarized in the following way: Under what conditions,
does the estimation of the models
\[ y_t = \nu (B) x_{t-b} + n_t \]
\[ y_t = \nu (B) x_{t-b} + \psi (B) a_t \]
yield equivalent results for \( \nu (B) \)? Instead, the question we address in
this and next section in given by: Under what conditions, does the
nonlinear estimation of
\[ \pi (B) y_t - \nu (B) x_{t-b} = a_t, \quad \pi (B) = 1 / \psi (B) \]
\[ y_t = \nu (B) e_{t-b} + \psi (B) a_t, \quad e_t = x_t / \sqrt{\psi} (B) \]
lead to equivalent results for \( \nu (B) \), \( \psi (B) \)? Notice that the above are
just AR and MA representations with the same operators as the
transfer function model, so that our answer must lie in the algebraic
relationships between the polynomials of (3.1). However, more gen-
eral and substantial issues are involved.

**Simplified AR** - Consider the “exact” autoregressive representation
of the system (4.1): \( \pi (B) y_t = w (B) x_{t-b} + a_t \). Consistent with
the univariate (ARMA) analysis, we would expect the number of
parameters involved and/or the order of the polynomials not to
change. Only the nature of the coefficients (linear or rational) or their value might be allowed to change. In the strictly arithmetical sense, however, the new impulse respose function \( w (B) = \pi (B) \nu (B) \) contains \( p + q \) new parameters, and this seems in contradiction with the mechanism of sequential filtering implicit in (4.1)

\[
\begin{align*}
  y_t - \nu (B) x_{t-b} &= n_t , \\
  n_t &= \pi_1 (B) n_t + a_t = \pi_1 (B) y_t + a_t
\end{align*}
\]

Where, in the first step cross-covariance \( \gamma_{xy} \) is filtered independently of auto-covariance \( \gamma_{yy} \), and in the second step, since \( \gamma_{nn} \) is a function of \( \gamma_{yy} \)

\[
\gamma_{nn} (k) = \gamma_{yy} (k) - [v_k \gamma_{xy} (b) + v_{k+1} \gamma_{xy} (b+1) + ...] \quad (4.4)
\]

(see Appendix for the proof) \( n_t \) may be represented with the basis \( \{y_t\} \), by means of a suitable polynomial \( \pi^* (B) \).

Now, the only possible way to have (4.1), (4.2) parametrically equivalent, is to require some form of orthogonality between the operators of the system. Setting \( v_b (B) = \nu (B) B^b \), \( w_b (B) = \nu (B) B^b \), this means

\[
w_b (B) = \pi (B) v_b (B) = [1 - \pi_1 (B)] v_b (B) \approx v_b (B)
\]

i.e.

\[
\pi_1 (B) = \left[ \sum_{j=1}^{\infty} \rho_j B^j \right] \perp \left[ \sum_{i=0}^{\infty} v_i B^i \right] = v_b (B)
\]

To this end, by extending the ideas of Section 1 to the space of infinite linear polynomials of \( B \), a measure of orthogonality is provided by (see also Piccolo, 1984)

\[
((\pi_1, \nu_b)) = [\pi' \nu / \sqrt{\pi' \pi \cdot \nu' \nu}] \in (-1, +1) \quad (4.5)
\]

\[
\pi' = [\pi_1 \pi_2 ... \pi_b \pi_{b+1} ...] \\
\nu' = [00 ... \nu_0 \nu_1 \nu_2 ...]
\]

which is close to zero according to:
i) \{\pi_i, \nu_i\} decay rapidly, i.e. \(\theta(B), \delta (B)\) have roots far from the unit circle;

ii) \(b\) is relatively high;

iii) \{\pi_j, \nu_i\} are non-monotonic, i.e. the polynomials have negative or complex roots.

By means of the metric (4.5), the properties of orthogonality of the polynomials are then directly referable to the dinamic structure of the system. In particular, to emphasize the role of the delay \(b\), note that if \(n_i \sim AR(p)\) with \(p < b\), it follows that \(\pi' = [\phi_1 \ldots \phi_p\] hence \((\pi_1, \nu_0) = 0\). Final consequence of the orthogonality \(\pi_1 (B) ~ \nu_b (B)\) is the simplified autoregressive representation (4.2).

**Simplified MA** - In the identification of the impulse response function \(\nu (B) B^b\), use of the partial cross covariance function has never been required. This function may be defined in practice, as proportional to the sequence of marginal regression coefficients

\[
\{\nu_{kk}\} \in y_t = \sum_{i=0}^{k} \nu_{ik} x_{t-i} + n_t
\]

or equivalently with

\[
\begin{pmatrix}
\gamma_{xy}(0) \\
\gamma_{xy}(1) \\
\vdots \\
\gamma_{xy}(k)
\end{pmatrix} =
\begin{pmatrix}
\gamma_{xx}(0) & \gamma_{xx}(1) & \ldots & \gamma_{xx}(k) \\
\gamma_{xx}(1) & \gamma_{xx}(0) & \ldots & \vdots \\
\vdots & \vdots & \ddots & \vdots \\
\gamma_{xx}(k) & \ldots & \gamma_{xx}(0)
\end{pmatrix}
\begin{pmatrix}
\nu_{0k} \\
\nu_{1k} \\
\vdots \\
\nu_{kk}
\end{pmatrix}
\]

\[
\gamma_{xy} = \Gamma_{xx} \nu_{kk}
\]

The unnecessity of \(\{\nu_{kk}\}\) suggests that cross covariance and partial cross covariance should have the same information content, i.e. the same pattern, and indeed, if \(\{x_t\}\) is white noise (\(\Gamma_{xx}\) diagonal) we have \(\gamma_{xy} \propto \nu_{kk}\). This equivalence can be extended to the case of \(\{x_i\}\) correlated, by assuming orthogonality between \(\bar{\psi}_1 (B) = [\bar{\psi} (B) - 1]\) and \(\nu_b (B)\), obtaining

\[
\gamma_{xy} (B) = \nu (B) B^b \sigma^2 \propto \nu_b (B)
\]

(4.7)
Also in this case, by using the metric (4.5), the orthogonality is
directly referable to the dynamic structure of the system. In
particular, assuming $x_i \sim MA(q)$ with $q < b$, we have $\tilde{\psi} = [\theta_1 \ldots \theta_q]$, 
hence $(v_b, \tilde{\psi}_1) = 0$. As a consequence of the orthogonality $v_b(B) \perp 
\tilde{\psi}_1(B)$ we have the simplified moving average representation (4.3).

Let us finally emphasize that the measure $(,)$ not only has a
formal meaning in terms of space of polynomials and orthogonality.
Indeed, it summarizes two substantial facts, responsible of the sim-
plified representations:

   i) When the monic polynomials have stable roots far from
the unit circle, their products generate near singular components that
cancel. For example, if $\phi(B) = (1 - \phi_1 B)$, $\phi_1 \ll 1$, then $\omega(B) \phi_1 \approx 
0$; this fact is clearly strengthened by the presence of a second root
with opposite sign.

   ii) When the polynomials have negative or complex roots and
$b$ is relatively high, the impulses $\{v_i, \psi_j, \tilde{\psi}_b\}$ inside the system are
either positive or negative and compensate each other. A typical
example is the gain $g = \sum_{i=0}^{\infty} v_i$, $v_i = \sum_{j=1}^{\infty} \delta_j v_{i-j} - \omega_i$; while the
contribution of $b$ is well established by (4.4).

Applications - In what follow, we investigate the practical impli-
cations of orthogonality on the algorithms of identification, estimation
and forecasting:

Identification - Recalling previous results (4.4) and (4.7), we have

\[
\pi_1(B) \perp v_b(B) \rightarrow \gamma_{m}(B) = \gamma_{yy}(B) \\
v_b(B) \perp \tilde{\psi}_1(B) \rightarrow \gamma_{xy}(B) \propto v(B) B^b
\]

In this way, the identification of $v(B), \psi(B)$ may be directly
developed on the sample correlation functions (auto and cross) of the
observable series $\{y_t, x_t\}$. In practice, the procedures of filtering (to
estimate $n_t$) and prewhitening (to cancel $\tilde{\psi}(B)$), see Box and Jenkins
(1976), are no longer necessary. Moreover, all the problems of
choosing among alternative filters can be avoided.

Estimation - The estimation of the parameters $\beta$ utilizing the repre-
sentation (4.2), leads to a substantial saving of calculation. In fact,
with respect to (3.4), the gradient of the Gauss-Newton estimator takes on the analytic expression

\[
\frac{\xi'_t}{\xi_t} (\beta) = \frac{\partial a_t}{\partial \beta} = \left[ \frac{1}{\delta (B)} (m_{t-1} \ldots m_{t-r} x_{t-b} \ldots x_{t-b-s}) \right] \]

\[
\frac{1}{\theta (B)} (y_{t-1} \ldots y_{t-\rho} u_{t-1} \ldots u_{t-\psi})
\]

\[
u_t = m_t + \frac{a_t}{\pi (B) y_t}, \quad u_t = \pi (B) y_t
\]

The resulting nonlinear estimates \(\hat{v}_N, \hat{\psi}_N\) are no more statistically independent. However, their gradients are computed separately from the observable series \(y_t, x_t\), with a procedure of filtering that involves 1/3 the previous calculations only.

**Forecasting** - In the context of the representation (4.3), it is easy to see that the variance of the \(l\)-steps ahead prediction error becomes

\[
E [\hat{\sigma}_t (l)]^2 = \sum_{i=0}^{l-b} v_i^2 \hat{\sigma}^2 + \sum_{j=0}^{l} \psi_j^2 \sigma^2
\]

This simplified expression has been already introduced by Box and Jenkins (1976, p. 405); we remark, however, that without the assumption of \(v_0 (B) \perp \hat{\psi}_1 (B)\) it cannot be maintained. Finally, using the representation (4.2) we may split the predictor \(\hat{y}_t (l)\) into two quantities, both operating on the observable series

\[
\hat{y}_{1t} (l) = \phi_1 \hat{y}_{1t} (l-1) + \ldots + \phi_p y_{t+l-p} + \theta_l u_t + \ldots + \theta_q u_{t+l-q}
\]

\[
\hat{y}_{2t} (l) = \delta_1 \hat{y}_{2t} (l-1) + \ldots + \delta_r m_{t+l-r} + \omega_0 \hat{x}_t (l-b) + \ldots + \omega_s x_{t+l-b-s}
\]

5. **AN EMPIRICAL EXAMPLE**

The criterion of orthogonality ((.,)) has practical utility and is reasonable since it refers to observable and spectral properties of the operators. It is not sure, however, that if \(((\rho, q)) = 0\) the polynomials
$p(x)$, $q(x)$ would actually be orthogonal; indeed other types of metrics could be introduced. Anyway, orthogonality itself is not an absolute concept: We should distinguish a polynomial orthogonality in the arithmetical sense from one relating to the filtering performance of operators. In this section we empirically check the ability of (4.5) to characterize the orthogonality in the second sense.

Let us define the three macroeconomic series

\begin{align*}
B_t &= \text{Balance of foreign trade}, (1 - B) B_t = Y_t; \\
PM_t &= \text{Index of import prices}, (1 - B) PM_t = X_t; \\
PI_t &= \text{Index of wholesale proces}, (1 - B) PI_t = Z_t; \\
t &= \text{Monthly data in the period 1973.01 - 1985.12}.
\end{align*}

For all the series the stationarity in mean was reached with difference of order one $(1 - B)$; Figures 2, 3 report the corresponding sample covariance functions.

In Figure 2 we note that whereas the series $X_t$ is practically a white noise, the others $Y_t, Z_t$ still exhibit a considerable autocovariance, of $MA(1), AR(1)$ type respectively. The implied univariate models were

\[
Y_t = \left(1 + \frac{.763 B}{(8.5)}\right)y_t, \quad X_t \approx x_t, \quad \left(1 - \frac{.612 B}{(6.9)}\right)Z_t = z_t.
\]

In Figure 3 both the pairs $(Y, X), (X, Z)$ seem related by feedback. The identification of transfer function models, however, requires analysis of the "prewhitened" cross covariance functions, given in Figure 4.

Comparison of Figures 3, 4 provides a first check on the performance of $((,))$.

Indeed, since

\[
x_t \sim MA(q < b) \rightarrow ((v_b, \bar{\psi}_1)) = 0 \\
v_b(B) \perp \bar{\psi}_1(B) \rightarrow \gamma_{xy}(B) \propto v_b(B)
\]

then to test if

\[
((v_b, \bar{\psi}_1)) = 0 \rightarrow v_b(B) \perp \bar{\psi}_1(B)
\]
Fig. 2 - Sample Autocovariance Functions

Fig. 3 - Sample Cross Covariance Functions

Fig. 4 - Prewhitened Cross Covariances
one can investigate the effect of the prewhitening on the behaviour of the cross covariances. Now, in our data we observe that \( \gamma_{yx} \approx \gamma_{yx} \), since \( Y_t \sim MA (1), \ b > 0 \); while \( \gamma_{xz} \neq \gamma_{xz} \), because \( Z_t \sim AR (1), \ b = 0 \).

An efficient test on the reliability of ((c)) follows from the estimation of the various TF representations. From Figures 3 and 4 we identify the system

\[
Y_t = \left( \frac{\omega}{1 - \delta B^3} \right) X_{t-1} + (1 + \theta B) a_t
\]

and empirically

\[
TF 1: Y_t = \left[ -0.00985 / \left( 1 - 0.514 B^3 \right) \right] X_{t-1} + \left( 1 + 0.632 B \right) a_t, \ \hat{\sigma}_1 = 0.674
\]

\[
TF 2: \left[ 1 / \left( 1 + 0.588 B \right) \right] Y_t = \left[ -0.00913 / \left( 1 - 0.449 B^3 \right) \right] X_{t-1} + a_t, \ \hat{\sigma}_2 = 0.687
\]

Apart from the relative loss of significance of the parameter \( \delta \) in the second representation, the two estimates are not statistically different on the basis of the tests \( (\hat{\beta}_1 - \hat{\beta}_2) / \sqrt{s_1^2 + s_2^2} \to N (0, 1), (\hat{\sigma}_2 / \hat{\sigma}_1)^2 \to F (n, n) \). Hence it should be \( \pi_1 (B) \perp v_b (B) \); indeed, since

\[
\pi' = [ - \theta - \theta^2 - \theta^3 - \theta^4 ... ] \\
\psi' = [ - \omega 0 0 0 + \omega \delta 0 0 - \omega \delta^2 ... ]
\]

we obtain the estimate \( (\hat{\pi}_1, \hat{\psi}_b) = 0.23 \); but \( \cos (\alpha) = 0.23 \implies \alpha = 85 / 100 \).

A further check on the orthogonality of \( v_b (B), \ \psi_1 (B) \), comes from the estimation of the second system

\[
X_t = \left( \frac{\omega}{1 - \delta B^3} \right) Y_{t-1} + e_t, \quad Y_t = (1 + \theta B) y_t
\]
It gave

$$TF 1 \quad X_t = \left[ \frac{6.32}{(3.5)} \right] \left( 1 - \frac{.554}{(-2.1)} B^3 \right) Y_{t-1} + e_t, \quad \hat{\sigma}_1 = 18.07$$

$$TF 3 \quad X_t = \left[ \frac{6.52}{(3.1)} \right] \left( 1 - \frac{.637}{(-3.5)} B^3 \right) y_{t-1} + e_t, \quad \hat{\sigma}_2 = 18.17$$

where $y_t = (1 + .763 \ B)^{-1} Y_t$; also in this case, we check the statistical equivalence of the estimates. In fact the above coefficients yield a value $\langle \langle \hat{\psi}_b, \hat{\psi}_1 \rangle \rangle = .12$, such that $\cos (\beta) = .12 \Rightarrow \beta = 92/100$.

6. Conclusions

Empirical analysis points out the role of negative roots far from the unit circle, in characterizing the polynomial orthogonality in terms of the measure $\langle \langle, \rangle \rangle$. The general aim of the study is that of finding simplified AR and MA representations which utilize the same system operators. This question is interesting in its own right, but also have useful consequences in building dynamic models. For multivariate closed-loop systems (simultaneous transfer functions)

$$z_t = V (B) z_t + \Psi (B) a_t, \quad z_t = [z_1, ... z_m]$$

$$V (B) = \{v_y (B) B^{by}\}, \quad \Psi (B) = \text{Diag} [\psi_t (B)]$$

this problem is more urgent, but more difficult to solve. A parametrically equivalent AR representation $[\Pi (B) - V (B)] z_t = a_t, \Pi (B) = \Psi (B)^{-1}$ is conceivable by assuming polynomial orthogonality by rows: $\Pi (B) \perp V_b (B)$ where $b = \min b_y$. However, for the MA representation, whose exact algebraic expression is unacceptably complicated $z_t = [I - V (B)]^{-1} \Psi (B) a_t$ (and thus non estimable), a solution of stochastic type must be sought.
Appendix

Proof of Formula (4.4)

From the assumptions of section 3 we have

\[ \gamma_{xy}(k) = E \{ y_i \, x_i \, x_{i-k} \} \begin{cases} 0 & k < b \\ \neq 0 & k \geq b \end{cases} \quad b \geq 0 \]

Now since \( n_i = \nu(B) x_{i-b} \), \( \nu(B) = \sum_0^\infty \nu_i B^i \) it follows

\[ \gamma_{n} (k) = E [n_i, n_{i-k}] = E \left[ (y_i - \nu_0 x_{i-b} - \nu_1 x_{i-b-1} - \nu_2 x_{i-b-2} - \ldots - \nu_k x_{i-b-k} - \ldots) \right. \]

\[ \left. (y_{i-k} - \nu_0 x_{i-b-k} - \nu_1 x_{i-b-k-1} - \ldots - \nu_k x_{i-b-2k} - \ldots) \right] \]

Hence

\[ \gamma_{n} (k) = \gamma_{yy}(k) - \nu_k \gamma_{xy}(b) - \nu_{k+1} \gamma_{xy}(b+1) - \ldots \hspace{1cm} (0) \]

\[ - \nu_0 \gamma_{xy}(b+k) - \nu_1 \gamma_{xy}(b+k+1) - \nu_2 \gamma_{xy}(b+k+2) - \ldots \hspace{1cm} (1) \]

\[ + \nu_0 \nu_0 \gamma_{xx}(k) + \nu_0 \nu_1 \gamma_{xx}(k-1) + \nu_0 \nu_2 \gamma_{xx}(k-2) + \ldots \hspace{1cm} (2) \]

\[ + \nu_1 \nu_0 \gamma_{xx}(k+1) + \nu_1 \nu_1 \gamma_{xx}(k) + \nu_1 \nu_2 \gamma_{xx}(k-1) + \ldots \hspace{1cm} (3) \]

\[ + \nu_2 \nu_0 \gamma_{xx}(k+2) + \ldots \]

But \( y_i = \nu(B) x_{i-b} + \psi(B) a_i \) so that

\[ \nu_0 \gamma_{xy}(b+k) = \nu_0 E \{ [(\nu_0 x_{i-b} + \nu_1 x_{i-b-1} + \nu_2 x_{i-b-2} + \ldots + \psi(B) a_i) x_{i-b-k} \] \]

\[ = \nu_0 \left[ \nu_0 \gamma_{xx}(k-1) + \nu_2 \gamma_{xx}(k-2) + \ldots \right] \]

\[ \nu_1 \gamma_{xy}(b+k+1) = \nu_1 \left[ \nu_0 \gamma_{xx}(k+1) + \nu_1 \gamma_{xx}(k) + \ldots \right] \]

Hence, the first term of (1) cancels with row (2); the second term of (1) cancels with row (3), and so on. Only the row (0) does not vanish, being indeed formula (4.4).
REFERENCES


SUMMARY

We propose a measure of orthogonality for real polynomials which can be suitably applied to the rational operators of dynamical stochastic systems. The measure emphasizes the role of the dynamic structure of the models, in particular roots and delay factors. The existence of orthogonality has important consequences in simplifying autoregressive and moving average representations and thus on the algorithms of identification, estimation and forecasting. The performance of the measure in characterizing the required propered property will be empirically checked.

RIASSUNTO

Viene proposta una misura di ortogonalità per polinomi a coefficienti reali che può essere utilmente applicata agli operatori razionali di sistemi dinamici stocastici. La misura sottolinea il ruolo che ha la struttura dinamica dei modelli, in particolare radici e ritardi. L’esistenza di ortogonalità ha importanti conseguenze nel semplificare le rappresentazioni autoregressive ed a media mobile e quindi sugli algoritmi di identificazione, stima e previsione. Il comportamento della misura nel caratterizzare le proprietà richieste è infine sottoposto a verifica empirica.