RECURSIVE ESTIMATION OF NONLINEAR STATE - NONLINEAR OBSERVATION SYSTEMS (NSNOS)

PART I: ON-LINE DATA

George Christakos

Advanced Projects Section

K.G.S.

November 1986

INTRODUCTION

The subject of nonlinear estimation, which has been of interest for the last three decades in physics and electrical engineering, is gaining renewed interest in the field of earth sciences.

This interest is perhaps owing to the fact that earth scientists are becoming fully aware that many physical processes cannot be represented adequately by traditional linear models (e.g., Mein et al., 1974; Jackson and Marechal, 1979; Bras and Georgakakos, 1980; Christakos, 1987).

Given the success in linear real-line problems, certain nonlinear recursive optimization algorithms use the related approaches in nonlinear systems with behavior close to that of linear ones. Assuming that the nonlinearities can be expanded in Taylor series, the resulting extended Kalman equations for the estimate and the error variance may be obtained (Gelb, 1974; Anderson and Moore, 1979). Work by Athans et al., (1968) suggests the expansion of the state nonlinearity about the optimal estimate, but the residual of the expansion is unspecified. Then, by establishing a cost criterion one may derive this residual recursively. Similar analyses have been done by Jaswinski (1966) and Culver (1969). Due to the Taylor-series expansions, these methods require that the nonlinear functions of interest are differentiable. Thus, in order to take into account the effect of saturation, threshold and other nonlinearities, approximating any discontinuities or corners in the nonlinear functions may be necessary, and this reduces the accuracy of the resulting expansions.

Some authors (e.g., Sorenson and Stubberud, 1968; Willman, 1981) have viewed the problem of nonlinear estimation as one of approximating the probability density for the state conditioned on all available measurement data by Edgeworth or Gram-Charlier series. This procedure leads to very complicated equations describing the moments of the density; also, obtaining analytic criteria for judging the validity of the approximations is difficult.

Nonlinear recursive optimization based on statistical linearization (Mahalanabis and Farooq, 1971; Gelb, 1974) does not make any assumption regarding function differentiability, but in order to calculate the expansion coefficients, knowledge of the full distribution of the state at every step is required. Furthermore, a great number of calculations are needed depending upon the specific type of the nonlinearities. Therefore, one usually keeps only the first few terms in the expansions and this further questions the accuracy gained.

In this presentation, we develop an estimator which considers expansions of the state and measurement nonlinearities in terms of orthogonal polynomials. These expansions experience properties of significant importance when the states are modeled as factorable random processes. The latter institutes a class of random processes which we define so that their two-dimensional distributions are expressed as the linear combination of the one-dimensional distributions and the orthogonal polynomials of the states. Factorability is a key hypothesis for the present study and it proves to be directly related to the assumption required in order to establish the expansion terms of the nonlinearities. The results obtained are extended to processes which are not necessarily factorable but can be derived from the latter by some kind of transformation. Then the problem is seen to become that of testing to determine if such a transformation can be established. One distinct advantage of the proposed method is that we do not have to calculate a finite number of expansion terms, since the nonlinear estimates are expressed exactly via the state and measurement functions. The gains require no more than polynomial coefficient of degree one to be calculated exactly by means of a dynamic process involving estimated statistics. These features of the proposed method may avoid questionable approximations at this stage of the optimization. On the other hand, the forms of the filtering and prediction expressions assumed cover a wide space of possible nonlinear estimators. Furthermore, because it is a recursive estimator, the method experiences the powerful properties of this type of estimators, connected with the importance of dynamic structure in data-processing methods (Christakos; 1985; 1987). Within this framework, the proposed method may be viewed as a simple, practical substitute to more sophisticated approaches to nonlinear estimation, e.g., martingale theory (Kallianpur, 1980), Lie algebraic and differential geometric methods (Brockett, 1980), stochastic partial differential equations (Pardoux, 1979) or Volterra series expansions (Marcus, 1979).

ORTHOGONAL POLYNOMIALS AND FACTORABLE RANDOM PROCESSES

A basic goal in applied estimation is to develop new types of random processes which conform to the particular estimation models and procedures. This is usually a matter of devising suitable relationships between the associated random variables and the probabilistic hypotheses. Within this context, the notion of using orthogonal expansions related to the probability density has always been an intriguing one (e.g., Stratonovich, 1963; Sorenson and Stubberud, 1968).

Let $p_k(x)$, k=1,2,... be a sequence of polynomials of degree k which are orthogonal with respect to a probability density function g(x), i.e. Szego (1959)

$$\int_R p_k(x)\, p_{k'}(x)\, g(x)\, dx = \delta_{kk'} \qquad (1)$$

where

$\delta_{kk'}=1$ ($k=k'$), $=0$ ($k\neq k'$). Here $p_o(x) = 1$, in consequence of the property of a density ($\int g(x)dx = 1$). Assuming that f(x) is a square integrable function with respect to g(x) ($\int f^2(x)\, g(x)\, dx < \infty$), it is possible to expand f(x) in terms of orthogonal polynomials $p_k(x)$ as follows

$$f(x) = \sum_{k=o}^{\infty} f_k\, p_k\,(x) \qquad (2)$$

where the coefficients f_k are given by

$$f_k = E[f(x)\, p_k(x)] = \int_R f(x)\, p_k(x)\, g(x)\, dx \qquad (3)$$

for all k. Eqs.(2) and (3) offer the best approximation up to order k of the function f(x), by a polynomial of degree k in the mean square sense. This is easily shown, since the minimization of

$$M = E[f(x) - \sum_{k=o}^{\infty} f_k\, p_k(x)]^2 = \int_R [f(x) - \sum_{k=o}^{\infty} f_k\, p_k(x)]^2\, g(x)dx \qquad (4)$$

with respect to the coefficients f_k, i.e. $\partial M/\partial f_k=0$, immediately implies equ.(3). The above procedure may be thought of as a statistical approximation technique and it has the important advantage over the Taylor-series expansion employed by some nonlinear estimation methods: it does not require the existence of derivatives for f(x) and, therefore, numerous nonlinearities can be treated without having to approximate corners and discontinuities in f(x). Moreover, the expansion coefficients are calculated more easily than by the usual series expansions of other statistical linearization methods (Gelb, 1974). Of course, the density function g(x) must be known or approximated and this is a problem to be considered within the context of random processes, below. In the following, working with "normalized" random quantities is convenient. Thus a random variable x may be replaced by the variable $\{x-E[x]\}/\sqrt{Var[x]}$; the latter has zero mean and a unit variance. We consider random (or stochastic) processes in the sense of Doob (1953), and we define the particular class of factorable random processes to be employed in this presentation. (A detailed discussion of this class of random functions is given in the Supplement at the end of this report.)

DEFINITION 1: Let $x_t, t \in T$ (T is a reference set), be a random process, which is assumed to have the same one-dimensional (univariate) density function $g(\chi_i)=g$ for all variables x_i, while the two-dimensional (joint) density $g(\chi_i,\chi_j)$ can be expressed, for any two variables x_i, x_j, as follows

$$g(\chi_i,\chi_j) = \sum_{k=o}^{\infty} u_k(i,j)\ p_k(\chi_i)\ p_k(\chi_j)\ g(\chi_i)\ g(\chi_j) \tag{5}$$

where $u_k(i,j)$ are coefficients such as

$$u_k(i,j) = E[p_k(x_i)\ p_k(x_j)] \tag{6}$$

Random processes satisfying the above assumptions are of significant importance for the purposes of our study and will be called factorable random processes. The corresponding variables x_i, x_j will be called jointly factorable random variables.

DEFINITION 2: Two random processes x_t, y_t will be called jointly factorable if for any t T the corresponding variables are jointly factorable.

Probability models similar to (5) are widely employed in stochastic mathematics. For example, Pearson (1930) used the so-called tetrachoric series model; see also Harris and Soms (1980). Other applications are found in Cramer (1945), Stratonovich (1963), and Matheron (1976). The class of factorable random processes is quite wide and can take account of a broad range of applications. We give some examples below.

REMARK 1: In fact, eq.(6) can be proven by setting

$$E[p_k(x_i)\ p_k(x_j)] = \iint p_k(\chi_i)\ p_k(\chi_j)\ g(\chi_i,\chi_j)\ d\chi_i d\chi_j$$

and then substituting $g(\chi_i,\chi_j)$ form eq.(5)

$$E[p_k(x_i)\ p_k(x_j)] = \sum_{\lambda=o}^{\infty} u_\lambda(i,j) \int p_\lambda(\chi_i)\ p_k(\chi_i)\ g(\chi_i)\ d\chi_i \int p_\lambda(\chi_j)\ p_k(\chi_j)\ g(\chi_j)\ d\chi_j$$

Due to eq.(1) the terms with $\lambda \neq k$ vanish and the integrals are equal to one; thus, eq.(6) is obtained.

EXAMPLE 1: An interesting example of factorable processes are the two-dimensional (but not necessarily completely) Gaussian processes combined with Hermite polynomials; the link between them is that the latter may be viewed as the derivatives of the unidimensional Gaussian density. In this case, $p_k(\chi) = \frac{1}{\sqrt{k!}\ g(\chi)} \frac{d^k}{d\chi^k} [g(\chi)]$, where $g(\chi) = \exp[-\chi^2/2]/\sqrt{2\pi}$

and $u_k(i,j) = E[x_i x_j]^k$. (This result is easily obtained by expressing the corresponding characteristic function in series-form, and then taking the inverse Fourier transform term-by-term.)

EXAMPLE 2: Random processes having one-dimensional γ (gamma) density $g(\chi) = \chi^{c-1}\exp(-\chi)/\Gamma(c)$, $\Gamma(c)$ is the gamma function, and the two-dimensional one is expressed in terms of

$p_k(\chi) = \chi^k L_k^{(c)}(\chi)$, where

$L_k^{(c)}(\chi) = \frac{(-1)^k}{k!\ g(\chi)} \frac{d^k}{d\chi^k}[\chi^k g(\chi)]$ are the Laguerre polynomials and
$u_k(i,j) = \{E[x_i x_j]\}^k$,

are factorable also. In fact they represent a diffusion process (such processes is known to play a fundamental role in nonlinear estimation theory; see, e.g., Zakai, 1969). Depending on the value of c, the γ density can take a variety of shapes: If $c<1$, the g(.) is monotonically decreasing, if $c=1$, the g(.) is exponential, and if $c>1$, it is bell-shaped approaching the Gaussian density as c increases. Various densities of the factorable form (5) may be obtained starting from orthogonal polynomials such as in eq.(1), under the condition that the weighting function g(.) can be used as a density function.

REMARK 2: From eq.(5) and the multiplicaton theorem of densities

$$g(\chi_i,\chi_j) = g(\chi_j)\ g(\chi_i/\chi_j)$$

we find

$$g(\chi_i/\chi_j) = \theta(\chi_i,\chi_j)\ g(\chi_i)$$

where, $\theta(\chi_i, \chi_j) = \sum_{k=o}^{\infty} u_k(i,j)\ p_k(\chi_i)\ p_k(\chi_j)$. Clearly, this equation may be used in Definition 1 in the place of eq.(5).

PROPOSITION 1: Let x_i, x_j be two jointly factorable random variables. Then we may write

$$E[p_k(x_i)\ p_{k'}(x_j)] = \delta_{kk'}\ u_k(i,j) \tag{7}$$

(i.e., $p_k(x_i)$ and $p_{k'}(x_j)$ are uncorrelated for $k \neq k'$) and,

$$E[p_k(x_i)/x_j] = u_k(i,j)\; p_k(\chi_j) \tag{8}$$

Proof: Combining orthogonal polynomials with the factorability assumption implies that the function g(x) of eq.(1) is the density of the factorable process defined by eq.(5). Then we find

$$E[p_k(x_i)p_{k'}(x_j)] = \iint p_k(\chi_i)\; p_{k'}(\chi_j)\; g(\chi_i,\chi_j)\; d\chi_i d\chi_j$$

Substituting $g(\chi_i,\chi_j)$ from eq.(5) we obtain

$$E[p_k(x_i)p_{k'}(x_j)] = \sum_{\lambda=o}^{\infty} u_\lambda(i,j) \int p_k(\chi_i)\; p_\lambda(\chi_i)$$
$$g(\chi_i)d\chi_i \int p_{k'}(\chi_j)p_\lambda(\chi_j)g(\chi_j)d\chi_j$$

It follows from eq.(1) that the terms in the right-hand side vanish except for $\lambda=k=k'$, and so eq.(7) is obtained. To prove eq.(8) one may work along similar lines using the expression for $g(\chi_i/\chi_j)$ given in Remark 2 above.

REMARK 3: More generally it can be shown (Christakos, 1986; manuscript submitted to IEEE, Trans. on Aut. Control) that the bivariate density of any theoretical or observed process subject to the summability condition $\iint \frac{g^2(\chi_i,\chi_i)}{g(\chi_i)\ g(\chi_j)}\, d\chi_i\ d\chi_j < \infty$, can be described by

eq. (5). This fact largely extends the applicability of the estimator presented in this paper.

REMARK 4: Since classification of random processes based upon statistical regularity or memory do not make specific references to the detailed form of the probability densities, a factorable process may be coupled with properties from either one or both of these classifications (e.g., stationary and/or Markovian properties). To illustrate this assume that the process x_t is also Markovian; then

$$E[x_t/X_{t-1}] = E[x_t/x_{t-1}]$$

where, $X_{t-1} = \{x_s : o < s < t-1\}$. By applying the definition of factorability we can express the conditional mean as

$$E[x_t/x_{t-1}] = \int \frac{x_t \sum_k u_k(t,t-1)\ p_k(x_t)\ p_k(x_{t-1})\ g(x_t)\ g(x_{t-1})\ dx_t}{g(x_{t-1})}$$

$$= \{\sum_k u_k(t,t-1) \int x_t p_k(x_t) g(x_t) dx_t\}\ p_k(x_{t-1})$$

$$= b\ x_{t-1}$$

where b $(=u_1(t,t-1))$ is a numerical coefficient (see eq.(8)).

Thus

$$E\{x_t/X_{t-1}\} = b\ x_{t-1}$$

which means that the result of combining factorability and Markov properties may be a martingale-type random process (except for the coefficient b). This is important within the context of estimation.

The numerical coefficients $u_k(i,j)$ in eqs.(6) and (7) express the correlation structure of the process. In fact,

PROPOSITION 2: The coefficients $u_k(i,j)$ are such that

$$|u_k(i,j)| \leq 1 \tag{9}$$

Proof: By the Schwartz's inequality (we always assume that the variables have been "normalized")

$$|u_k(i,j)| = |E[p_k(x_i)\ p_k(x_j)]| \leq \{E[p_k^2(x_i)]\}^{1/2}\{E[p_k^2(x_j)]\}^{1/2}$$

But from eq.(1), both expectations in the right hand side are equal to one and, thus, eq.(9) follows.

For k=0, a straightforward result of the definition of the orthogonal polynomials ($p_o(\cdot) = 1$) is, $u_o(i,j) = 1$.

ESTIMATION ALGORITHMS

First, both the structural and the measurement models are assumed to be scalar. The random process x_t is given by the nonlinear dynamic structural model

$$x_t = f(x_{t-1},t-1) + w_{t-1} \tag{10}$$

where w_{t-1} is a Gaussian white noise, independent of the initial condition x_o. Then the solution of eq.(10) is a Markov process. The observation process y_t is given by the nonlinear expression

$$y_t = h(x_t,t) + v_t \tag{11}$$

where v_t is a Gaussian white noise, uncorrelated with w_t in eq.(10). Eqs.(10) and (11) describe a nonlinear state-nonlinear observation system (NSNOS). Estimating x_t by processing the observed data y_s, $0<s<t$ is desirable, satisfying the minimum variance criterion. More precisely, let τ be a fixed real number and let Y_t denote the σ-field of sample path $\{y_s, 0<s<t\}$. Then, by $\hat{x}_{t+\tau/t}$ we will denote the estimation at time $t+\tau$ of the signal $x_{t+\tau}$ while taking into account the set of values Y_t observed up to time t. If $\tau=0$, the estimation problem

is called filtering, while if $\tau>0$, it is called prediction.

In order to obtain specific results, additional structure will be assumed for the processes $x_{t+\tau}$ and $\hat{x}_{t+\tau/t}$; the state $x_{t+\tau}$ and the estimate $\hat{x}_{t+\tau/t}$ are assumed jointly factorable, so that

$$E[p_k(x_{t+\tau})\ p_{k'}(\hat{x}_{t+\tau/t})] = \delta_{kk'}\ u_k(x_{t+\tau},\ \hat{x}_{t+\tau/t}) \tag{12}$$

where $u_k(x_{t+\tau},\hat{x}_{t+\tau/t}) = E[p_k(x_{t+\tau})\ p_k(\hat{x}_{t+\tau/t})]$. (Later we will relax this hypothesis by assuming processes which may be not factorable themselves but, instead, may result from such processing by some kind of well defined transformation.)

REMARK 5: Eq.(10) may be used to evaluate the probability density of x_t in terms of $g(x_{t-1})$. Indeed, by definition of marginal density and by applying Bayes' rule, we obtain

$$g(x_t) = \int g(x_t/x_{t-1})g(x_{t-1})dx_{t-1} \tag{13}$$

The $g(x_t/x_{t-1})$ can be computed from eq.(10), namely

$$g(x_t/x_{t-1}) = g(w_{t-1}), \quad \text{at } w_{t-1} = x_t-f(x_{t-1},\ t-1) \tag{14}$$

where the density $g(w_{t-1})$ is Gaussian. Thus eq.(14) gives

$$\int x_t g(x_t/x_{t-1})dx_t = f(x_{t-1},t-1) \tag{15}$$

Since x_t is factorable it can be easily shown that (see also Remark 2 above)

$$g(x_t/\ x_{t-1}) = \sum_{k=o}^{\infty} u_k(t,t-1)p_k(x_t)p_k(x_{t-1})g(x_t) \tag{16}$$

where, $u_k(t,t-1) = E[p_k(x_t)\ p_k(x_{t-1})]$. If one substitutes eq.(16) into the left-hand side of eq.(15), one finds

$$\begin{aligned}\int x_t g(x_t/x_{t-1})\ dx_t &= \sum_{k=o}^{\infty} u_k(t,t-1)\ p_k(x_{t-1}) \int x_t p_k(x_t)g(x_t)dx_t \\ &= u_1(t,t-1)\ x_{t-1}\end{aligned} \tag{17}$$

where the factorability property (7) has been employed. Similarly, working with the right-hand side of eq.(15) (w_{t-1}, $p_k(x_{t-1})$ are uncorrelated),

$$\begin{aligned}f(x_{t-1},t-1) &= \sum_{k=o}^{\infty} f_k p_k(x_{t-1}) \\ &= \sum_{k=o}^{\infty} E[x_t p_k(x_{t-1})]\ p_k(x_{t-1}) \\ &= u_1(t,t-1)\ x_{t-1}\end{aligned} \tag{18}$$

i.e. the same result as in eq.(17). Therefore, eq.(15) is valid for a factorable x_t and consequently the structural assumption (10) is compatible with the assumption of factorability.

More detailed results will be derived by assuming that the filtered state has the form

$$\hat{x}_{t/t} = \phi(\hat{x}_{t/t-1},t) + \beta_t y_t \tag{19}$$

where the first term represents the nonlinear weighting of the previous estimate and the second term is weighted current data sample. The latter is considered reasonable since knowledge of y_t contains information about x_t. The predicted state $\hat{x}_{t/t-1}$ will be reasonably described by

$$\hat{x}_{t/t-1} = \phi(\hat{x}_{t-1/t-1},t-1) \tag{20}$$

Expressions (19) and (20) are approximations of the nonlinear estimation around $\hat{x}_{t/t}$ since the optimal estimates $\hat{x}_{t/t}$ and $\hat{x}_{t/t-1}$ are conditional expectations of x_t with respect to Y_t (the σ -field of sample path $\{y_s,\ 0<s<t\}$) and Y_{t-1} ($\{y_s,\ 0<s<t-1\}$) respectively. (Equations of this form are exact for multi-Gaussian processes.) No approximation is assumed regarding eqs.(10) and(11). Instead, some workers in the area (see, e.g., Sorenson and Stubberud, 1968; Willman, 1981) suppose that perturbations of the system (10), (11) can be approximated by the first

and second order terms of the Taylor series. This leads to a significantly more restricted space of possible nonlinear estimators than those suggested by (19), (20) above. Moreover, the equations expressing the moments of the probability densities are very complicated, while questions arise regarding the validity of the approximations.

We will assume that the functions $f(\cdot)$, $h(\cdot)$, $\phi(\cdot)$, and $\psi(\cdot)$ are square integrable with respect to the density functions of the state and its estimate and, therefore, they may be expanded in terms of orthogonal polynomials. Thus,

$$x_t = \sum_{k=o}^{\infty} f_{k,t-1}\, p_k(x_{t-1}) + w_{t-1} \tag{21}$$

$$y_t = \sum_{k=o}^{\infty} h_{k,t}\, p_k(x_t) + v_t \tag{22}$$

$$\hat{x}_{t/t} = \sum_{k=o}^{\infty} \phi_{k,t}\, p_k(\hat{x}_{t/t-1}) + \beta_t y_t \tag{23}$$

$$\hat{x}_{t/t-1} = \sum_{k=o}^{\infty} \psi_{k,t-1}\, p_k\,(\hat{x}_{t-1/t-1}) \tag{24}$$

Coefficients $\phi_{k,t}$ and $\psi_{k,t}$, and accordingly the functions

$\phi(\cdot)$ and $\psi(\cdot)$, will be determined from the minimization of the mean square errors

$$\sigma^2_{t/t} = E[\hat{x}_{t/t} - x_t]^2 \tag{25}$$

and

$$\sigma^2_{t/t-1} = E[\hat{x}_{t/t-1} - x_t]^2 \tag{26}$$

respectively. The choice of $\phi(\cdot)$ and $\psi(\cdot)$ is the main problem in optimization. Conventional statistical linearization keeps only the first few terms in the corresponding polynomial expansions and, thus, it leads to questionable approximations. As we saw above, approximations are introduced in the proposed estimator only via the filtering and prediction models (19) and (20). We will show below that due to the factorability properties the functions $\phi(\cdot)$ and $\psi(\cdot)$ can be expressed exactly via the functions $f(\cdot)$ and $h(\cdot)$, and there is, finally, no need for approximation by orthogonal expansions.

The estimation algorithm is derived for scalar processes. First we consider factorable random processes and then the results obtained are extended to nonfactorable ones.

A. Factorable random processes

We will attempt to determine the equations of evolution of $\hat{x}_{t+\tau/t}$ and $\sigma^2_{t+\tau/t}$ directly from the NSNOS defined above. Of course these two parameters do not determine the probability function of the state $x_{t+\tau}$, since $x_{t+\tau}$ is not, in general, Gaussian. However, they may offer valuable information about the mean path of the structural model (10) and the dispersion about the path.

PROPOSITION 3: The recursive estimation equations for the NSNOS formulated above, assuming factorable random processes, are as follows:

Filtered state:

$$\hat{x}_{t/t} = \hat{x}_{t/t-1} + \beta_t\ \varepsilon_t \tag{27}$$

where we introduce the nonlinear innovation process

$$\varepsilon_t = y_t - h(\hat{x}_{t/t-1}, t) \tag{28}$$

Filtered error variance:

$$\sigma^2_{t/t} = (1- h_{1,t}\ \beta_t)^2\ \sigma^2_{t/t-1} + \beta^2_t\ E[v^2_t\ ] \tag{29}$$

where

$$\beta_t = h_{1,t}\ \sigma^2_{t/t-1}\ /\ r_t \tag{30}$$

is the gain, $h_{1,t}$ is an expansion coefficient of degree one, see eq.(22), and

$$r_t = h^2_{1,t}\sigma^2_{t/t-1} + E[v^2_t\ ] \tag{31}$$

is the innovation variance.

Predicted state:

$$\hat{x}_{t/t-1} = f(\hat{x}_{t-1/t-1}, t-1) \tag{32}$$

Predicted error variance:

$$\sigma^2_{t/t-1} = f^2_{1,t-1}\ \sigma^2_{t-1/t-1} + E\ [w^2_{t-1}] \tag{33}$$

where $f_{1,t-1}$ is an expansion coefficient of degree one, see eq.(21).

Proof: Substituting eq.(23) into (25) and differentiating with respect to $\phi_{k,t}$ and β_t we obtain, respectively

$$E\,[\,\hat{x}_{t/t}-x_t]\;p_k\,(\hat{x}_{t/t-1}) = E[\,\sum_{k'}\phi_{k',t}\;p_{k'}(\hat{x}_{t/t-1}) + \beta_t y_t - x_t]\;p_k(\hat{x}_{t/t-1})$$

$$= 0 \qquad\qquad \text{for all } k \qquad\qquad (34)$$

$$E[\,\hat{x}_{t/t}-x_t]\;y_t = E[\,\sum_{k'}\phi_{k',t}\;p_{k'}(\hat{x}_{t/t-1}) + \beta_t y_t - x_t]\;y_t = 0 \qquad (35)$$

which are the orthogonality equations of the filtered state. Similarly, for the predicted state it holds

$$E[\,\hat{x}_{t/t-1}-x_t]\;p_k\,(\hat{x}_{t-1/t-1}) = E[\,\sum_{k'}\phi_{k',t-1}\;p_{k'}(\hat{x}_{t-1/t-1}) - x_t]$$

$$p_k(\hat{x}_{t-1/t-1}) = 0 \qquad\qquad \text{for all } k \qquad\qquad (36)$$

Clearly the measures with respect to which orthogonality is considered are assumed to satisfy the factorability property (Definition 1). Before we continue with the proof of Proposition 3, we will prove the Lemma 1, below.

Lemma 1: If the state x_t is jointly factorable with the estimates $\hat{x}_{t/t}$ and $\hat{x}_{t/t-1}$ of the NSNOS defined above, then

$$\phi_{k,t}\, E[p_k(\hat{x}_{t/t-1}) - p_k(x_t)]\, p_k(\hat{x}_{t/t-1}) = 0 \tag{37}$$

and

$$\psi_{k,t-1} E[p_k(\hat{x}_{t-1/t-1}) - p_k(x_{t-1})]\, p_k(\hat{x}_{t-1/t-1}) = 0 \tag{38}$$

for all k ($\phi_{k,t}$ and $\psi_{k,t-1}$ are the polynomial expansion coefficients, see eqs.(23) and (24)).

Proof: From eqs.(24) and (36), we find

$$E[\hat{x}_{t/t-1} - x_t]\, \hat{x}_{t-1/t-1} = 0 \quad \text{or}$$

$$E[\hat{x}_{t/t-1} - x_t][\, \hat{x}_{t/t-1} - \sum_{k\neq 1} \psi_{k,t-1} p_k(\hat{x}_{t-1/t-1})]\, \frac{1}{\psi_{1,t}} = 0$$

Due to the factorability property, the terms with $k\neq 1$ vanish, thus

$$E[\hat{x}_{t/t-1} - x_t]\, \hat{x}_{t/t-1} = 0 \qquad (\psi_{1,t} \neq 0) \tag{39}$$

Similarly, the last equation together with eq.(23) yield

$$E[\hat{x}_{t/t-1}-x_t][\hat{x}_{t/t} \ -\sum_{k\neq 1} \phi_{k,t}\ p_k(\hat{x}_{t/t-1}) - \beta_t y_t]\ \frac{1}{\phi_{1,t}} = 0$$

or (factorability, again)

$$E[\hat{x}_{t/t-1} - x_t][\hat{x}_{t/t} - \beta_t y_t] = 0 \qquad (\phi_{1,t}\neq 0) \tag{40}$$

Eqs.(7), (12), and (39) give

$$E[p_k(\hat{x}_{t/t-1}) - p_k(x_t)]\ \hat{x}_{t/t-1} = 0 \tag{41}$$

for all k. The last equation in turn yields

******

$$p_\lambda(\hat{x}_{t/t-1})]\ \frac{1}{\phi_{1,t}} = 0$$

where eq.(23) has been used. Finally from this, eq.(40), and the factorability properties, ($\phi_{1,t}\neq 0$), we find eq.(37). From eqs.(23) and (35)

$$E[\hat{x}_{t/t} - x_t][\hat{x}_{t/t} - \sum_k \phi_{k,t} p_k(\hat{x}_{t/t-1})] \frac{1}{\beta_t} = 0$$

or due to eq.(34)

$$E[\hat{x}_{t/t} - x_t]\, \hat{x}_{t/t} = 0, \qquad \text{or}$$

$$E[\hat{x}_{t-1/t-1} - x_{t-1}]\, \hat{x}_{t-1/t-1} = 0 \tag{42}$$

After some manipulations, eq.(24) becomes

$$\hat{x}_{t-1/t-1} = \frac{1}{\phi_{1,t-1}} [\hat{x}_{t/t-1} - \sum_{k\neq 1} \phi_{k,t-1}\, p_k(\hat{x}_{t-1/t-1})]$$

which together with eq.(42) give

$$E[\hat{x}_{t-1/t-1} - x_{t-1}]\, \hat{x}_{t/t-1} = 0 \tag{43}$$

(factorability has canceled the terms with $k\neq 1$). Clearly,

$$E[p_k(\hat{x}_{t-1/t-1}) - p_k(x_{t-1})]\, \hat{x}_{t/t-1} = 0 \qquad \text{or}$$

$$E\{[p_k(\hat{x}_{t-1/t-1}) - p_k(x_{t-1})] \sum_\lambda \phi_{\lambda,t-1} p_\lambda(\hat{x}_{t-1/t-1})\} = 0 \qquad \text{or}$$

$$\phi_{k,t-1}\, E[p_k(\hat{x}_{t-1/t-1}) - p_k(x_{t-1})]\, p_k(\hat{x}_{t-1/t-1}) = 0$$

for all k; this completes the proof of Lemma 1.

Proof of eq.(27): Eq.(34), after some manipulations involving (22), writes

$$\phi_{k,t}E[p_k^2(\hat{x}_{t/t-1})] = E[x_t p_k(\hat{x}_{t/t-1})] - \beta_t E[\sum_{k'} h_{k',t} p_{k'}(x_t) p_k(\hat{x}_{t/t-1})]$$

or $$\phi_{k,t}\; E[\; p_k^2\; (\hat{x}_{t/t-1})] = \delta_{1k}\; E[\; x_t \hat{x}_{t/t-1}] - \beta_t\; h_{k,t}\; E[p_k(x_t)$$

$$p_k\; (\hat{x}_{t/t-1})] \qquad \text{for all k} \tag{44}$$

since v_t, $p_k(\hat{x}_{t/t-1})$ are uncorrelated. Taking into account eq.(37), eq.(44) gives

$$\phi_{k,t} = \delta_{1k} - \beta_t h_{k,t} \qquad \text{for all k} \tag{45}$$

Applying this result into eq.(23), we immediately derive (27).

To proceed, we need the Lemma that follows.

Lemma 2: The nonlinear innovation process defined by eq.(28) satisfies

$$E[\varepsilon_t] = 0 \tag{46}$$

$$E[\varepsilon_t^2] = h_{1,t}^2 \ \sigma^2_{t/t-1} + E[v_t^2 \ ] \tag{47}$$

$$E[x_t - \hat{x}_{t/t-1}] \ \varepsilon_t = h_{1,t} \ \sigma^2_{t/t-1} \tag{48}$$

Proof: The filtering condition of unbiasedness

$$E[\hat{x}_{t/t}] = E[x_t] \tag{49}$$

combined with eqs.(22) and (23), gives

$$\sum_k \phi_{k,t} E[p_k(\hat{x}_{t/t-1})] + \beta_t \sum_k h_{k,t} \ E[p_k(x_t)] = E[x_t]$$

Using eq.(45) together with the prediction condition of unbiasedness

$$E[\hat{x}_{t/t-1}] = E[x_t] \tag{50}$$

we obtain

$$E[\hat{x}_{t/t-1}] - \beta_t \sum_k h_{k,t} \ E[p_k(\hat{x}_{t/t-1} \ )] + \beta_t \sum_k h_{k,t} E[p_k(x_t)] = E[x_t] \text{ or}$$

$$\sum_k h_{k,t} \{E[p_k(\hat{x}_{t/t-1})] - E[p_k(x_t)]\} = 0 \tag{51}$$

($\beta_t \neq 0$). Then from eq.(28)

$$E[\varepsilon_t] = E[y_t - h(\hat{x}_{t/t-1},t)]$$

$$= E[h(x_t,t) - h(\hat{x}_{t/t-1},t)]$$

$$= \sum_k h_{k,t}\{E[p_k(x_t) - p_k(\hat{x}_{t/t-1})]\} = 0 \tag{52}$$

due to eq.(51). The innovation variance is

$$E[\varepsilon_t^2] = E[h(x_t,t) + v_t - h(\hat{x}_{t/t-1}, t)]^2$$

$$= E[v_t^2] + E[h(x_t,t) - h(\hat{x}_{t/t-1},t)]^2$$

$$= E[v_t^2] + \sum_k h_{k,t}^2 E[p_k(x_t) - p_k(\hat{x}_{t/t-1})]^2$$

since v_t, $\hat{x}_{t/t-1} - x_t$ are uncorrelated. Now, using eq.(37) (assuming $\phi_{k,t} \neq 0$),

$$E[\varepsilon_t^2] = E[v_t^2] + \sum_k h_{k,t}^2\ E[p_k(x_t) - p_k(\hat{x}_{t/t-1})]p_k(x_t)$$

$$= E[v_t^2] + \sum_k h_{k,t} E[p_k(x_t) - p_k(\hat{x}_{t/t-1})]\ [y_t - v_t - \sum_{\lambda \neq k} h_{\lambda,t}\ p_\lambda(x_t)]$$

$$= E[v_t^2] + h_{1,t}\ E[x_t - \hat{x}_{t/t-1}]\ y_t$$

due to the factorability property. Next by inserting eq.(22) and deleting the terms with $k \neq 1$, the last equation becomes

$$E[\varepsilon_t^2\ ] = E[v_t^2\ ] + h_{1,t}^2\ E[x_t - \hat{x}_{t/t-1}]x_t$$

$$= E[v_t^2] + h_{1,t}^2 E[x_t - \hat{x}_{t/t-1}]^2$$

$$= E[v_t^2\ ] + h_{1,t}^2 \sigma_{t/t-1}^2 \tag{53}$$

where eq.(39) has been used. Furthermore,

$$E[x_t - \hat{x}_{t/t-1}]\ \varepsilon_t = E[\hat{x}_{t/t-1} - x_t][h(x_t,t) - h(\hat{x}_{t/t-1},t)]$$

$$= \sum_k h_{k,t} E[\hat{x}_{t/t-1} - x_t][p_k(x_t) - p_k(\hat{x}_{t/t-1})]$$

$$= h_{1,t}\ E[\hat{x}_{t/t-1} - x_t]^2 \tag{54}$$

(the terms for $k \neq 1$ vanish due to the factorability property), which is eq.(48).

Proof of eqs.(29), (30), and (31): If we substitute eq.(27) into (25) we we find

$$\sigma^2_{t/t} = E[\hat{x}_{t/t-1} - x_t]^2 + \beta_t^2 \; E[\varepsilon_t^2] + 2\beta_t \; E[\; \hat{x}_{t/t-1} - x_t]\varepsilon_t \tag{55}$$

and by inserting eqs.(47) and (48), we obtain eq.(29).

Eq.(31) is, by definition, idential to eq.(47).

Eqs.(11) and (35) give,

$$E[\hat{x}_{t/t} - x_t][h(x_t,t) - \; v_t] = 0 \qquad\qquad \text{or}$$

$$h_{1,t}E[\hat{x}_{t/t} - x_t]x_t = - E[\hat{x}_{t/t} - x_t]v_t \tag{56}$$

By definition,

$$\sigma^2_{t/t} = E[\hat{x}_{t/t} - x_t]^2 = \frac{1}{h_{1,t}} E[\hat{x}_{t/t} - x_t] \; v_t$$

where both eqs.(42) and (56) have been used.

Clearly,

$\sigma^2_{t/t} = \dfrac{1}{h_{1,t}} E[\hat{x}_{t/t}v_t]$, since the x_t, v_t are uncorrelated. Thus

$$\sigma^2_{t/t} = \frac{1}{h_{1,t}} \; E[\; \sum_k \phi_{k,t}p_k(\hat{x}_{t/t-1}) + \beta_t y_t]v_t$$

$$= \frac{\beta_t}{h_{1,t}} E[y_t v_t]$$

since $v_t, \hat{x}_{t/t-1}$ are uncorrelated. Finally,

$$\sigma^2_{t/t} = \frac{\beta_t}{h_{1,t}} \{h_{1,t} E[x_t v_t] + E[v_t^2]\} = \frac{\beta_t}{h_{1,t}} E[v_t^2] \qquad (57)$$

Combining eqs.(29) and (57) one may solve for β_t to find equ(30).

Proof of eqs.(32) and (33): To proceed with the proof we write eq.(36) as follows

$$\phi_{k,t-1} E[p_k^2(\hat{x}_{t-1/t-1})] = E[x_t p_k(\hat{x}_{t-1/t-1})] \qquad \text{or}$$

$$\phi_{k,t-1} E[p_k^2(\hat{x}_{t-1/t-1})] = f_{k,t-1} E[p_k(x_{t-1}) p_k(\hat{x}_{t-1/t-1})] \qquad (58)$$

since w_{t-1}, $p_k(\hat{x}_{t-1/t-1})$ are uncorrelated. From eqs.(38) and (58)

$$\phi_{k,t-1} = f_{k,t-1} \qquad \text{for all } k \qquad (59)$$

which immediately yields eq.(32). Finally, to obtain eq.(33), one may substitute eq.(32) into (26)

$$\sigma^2_{t/t-1} = E[f(x_{t-1}, t-1) - f(\hat{x}_{t-1/t-1},t-1)]^2 + E[w^2_{t-1}]$$

noting that w_{t-1}, $f(x_{t-1}, t-1) - f(\hat{x}_{t-1/t-1},t-1)$ are uncorrelated.

The last equation becomes

$$\sigma^2_{t/t-1} = \sum_k f^2_{k,t-1}E[p_k(x_{t-1}) - p_k(\hat{x}_{t-1/t-1})]^2 + E[w^2_{t-1}]$$

where the terms with $k' \neq k$ have disappeared due to factorability properties; then

$$\sigma^2_{t/t-1} = \sum_k f^2_{k,t-1}E[p_k(x_{t-1}) - p_k(\hat{x}_{t-1/t-1})]\ p_k(x_{t-1}) + E[w^2_{t-1}]$$

the terms $E[p_k(x_{t-1}) - p_k(\hat{x}_{t-1/t-1})]p_k(\hat{x}_{t-1/t-1})$

being all zero (see eq.(38)). Clearly

$$\sigma^2_{t/t-1} = f_{1,t-1}E[x_{t-1} - \hat{x}_{t-1/t-1}][x_t - w_{t-1}] + E[w^2_{t-1}]$$

where eq.(21) and the factorability properties have been used. But $x_{t-1} - \hat{x}_{t-1/t-1}$, w_{t-1} are uncorrelated and x_t may be expressed by eq.(21) again, so that

$$\sigma^2_{t/t-1} = f^2_{1,t-1}\ E[x_{t-1} - \hat{x}_{t-1/t-1}]\ x_{t-1} + E[w^2_{t-1}]$$

This equation combined with eq.(42) yields eq.(33).

REMARK 6; Eqs.(27) and (29) may be rephrased in terms of the previous estimated state to give

$$\hat{x}_{t/t} = f(\hat{x}_{t-1/t-1},t-1) + \beta_t \{ y_t - h[f(\hat{x}_{t-1/t-1},t-1)]\} \tag{60}$$

$$\sigma^2_{t/t} = (1- h_{1,t}\ \beta_t)^2 \{ f^2_{1,t-1}\ \ \sigma^2_{t-1/t-1} + E[w^2_t]\} + \beta^2_t\ E[v^2_t] \tag{61}$$

If in eqs.(29) and (61), we substitute $E[v^2_t]$ from eqs.(30) and (31), we find the simpler expressions

$$\sigma^2_{t/t} = (1- h_{1,t}\beta_t)\ \sigma^2_{t/t-1} \tag{62}$$

$$\sigma^2_{t/t} = (1- h_{1,t}\beta_t)\ \{ f^2_{1,t-1}\ \sigma^2_{t-1/t-1} + E[w^2_t\ ]\ \} \tag{63}$$

respectively. The recursive formulation of the approach propagates a priori information $\hat{x}_{o/o}$, $\sigma^2_{o/o}$ through the system to give the estimate $\hat{x}_{t/t}$ and the associated accuracy $\sigma^2_{t/t}$ at each point t.

It is remarkable that the nonlinear functions $\phi(.)$ and $\psi(.)$ of eqs.(19),and (20) are determined exactly from the known functions f(.), h(.), and are not approximated by the expansions (21)-(24). Therefore, no approximation is introduced at this stage of estimation. Also, for the gain and the error variances we only need to calculate the expansion coefficients which correspond to k=1 (the other terms vanish due to factorability properties). These coefficients can be obtained by means of the statistics of the estimated states, i.e.

$$h_{1,t} = E[\, h(\hat{x}_{t/t-1},t)\ \hat{x}_{t/t-1}] \tag{64}$$

$$f_{1,t-1} = E[f(\hat{x}_{t-1/t-1},t-1\ )\ \hat{x}_{t-1/t-1}] \tag{65}$$

and inserted into the estimation algorithm (see eq.(3), also following sections).

Under the hypothesis of factorability, the estimations obtained are optimal for the nonlinear approximation (19), (20) (we will see below that the same holds true for certain nonfactorable processes as well). Surely, they are not the best of all possible estimators, which is the case of conditional expectation, but the prerequisites are less and the theory simpler. An assumption is required for the density function of the state x_t; however, such an assumption is already inherent in the hypothesis of factorability.

We saw that the innovation process ε_t is a zero mean white noise, and it is, by definition, orthogonal to $y_0,y_1,\ldots,y_{t-1}$. In fact, this is a quite general result of the stochastic estimation theory (Kailath, 1971; Allinger and Mitter, 1981). (Moreover, under certain circumstances the ε_t is not only white, but also Gaussian.)

In conclusion, another appealing property of the proposed estimator is that in the case of linear state-linear observation systems, like

$$x_t = a\, x_{t-1} + w_{t-1} \tag{66}$$

$$y_t = b\, x_t + v_t \tag{67}$$

it is equivalent to the Kalman filter (Nahi, 1969). Indeed, if in eqs. (27)-(33), we replace $f_{1,t-1}$, $h_{1,t}$, $f(\hat{x}_{t-1/t-1},t-1)$ and $h(\hat{x}_{t/t-1},t)$ by a, b, a $\hat{x}_{t-1/t-1}$ and b $\hat{x}_{t/t-1}$ respectively, these equations become identical to the standard Kalman ones.

B. Nonfactorable random processes

The presented method may be extended to random processes which do not satisfy the assumptions of factorability. The key point is to establish a suitable transformation of the given nonfactorable to a factorable process. If such a transformation exists, then estimation may be made in an analogous way. The relevant results are summarized below:

PROPOSITION 4: Let z_t be a nonfactorable random process generated by a NSNOS as above, for which we assume that it can be established a transformation $\gamma(.)$ such as

$$z_t = \gamma(x_t) \tag{68}$$

and

$$x_t = \gamma^{-1}(z_t) \tag{69}$$

where the process x_t is factorable. Then, the recursive estimation equations for the process z_t are as follows:

Filtered state:

$$\hat{z}_{t/t} = \hat{z}_{t/t-1} + \theta_t \, \varepsilon_t \tag{70}$$

where

$$\varepsilon_t = y_t - h(\hat{z}_{t/t-1}, t) \tag{71}$$

is the innovation process.

Filtered error variance:

$$\sigma^2_{t/t} = (1 - h_{1,t}\theta_t)^2 \, \sigma^2_{t/t} + \theta^2_t \, E\,[v^2_t\,] \tag{72}$$

where

$$\theta_t = h_{1,t} \, \sigma^2_{t/t-1} / r_t \tag{73}$$

$h_{1,t}$ is the expansion coefficient as before and

$$r_t = h^2_{1,t}\ \sigma^2_{t/t-1} + E\ [v^2_t\ ] \tag{74}$$

Predicted state:

$$\hat{z}_{t/t-1} = f(\hat{z}_{t-1/t-1}, t-1) \tag{75}$$

Predicted error variance:

$$\sigma^2_{t/t-1} = f^2_{1,t-1}\ \sigma^2_{t-1/t-1} + E[w^2_{t-1}] \tag{76}$$

where $f_{1,t-1}$ is the expansion coefficient of degree one for the function $f(\cdot)$.

Proof: The proof can be made following an analogous procedure with that of propositon 3; therefore, we will restrict ourselves to a brief description of its main points. Let $\hat{z}_{t+\tau/t} = \lambda(\hat{x}_{t+\tau/t})$, $\tau \geq 0$. Since eqs.(68) and (69) are valid, the orthogonality equation $E[\ \hat{z}_{t+\tau/t} - z_{t+\tau}]\ p_k(\hat{x}_{t+\tau/t}) = 0$, immediately gives $\lambda_{k,t} = \gamma_{k,t}$, where $\lambda_{k,t}$ and $\gamma_{k,t}$ are the expansion coefficients of the measurable functions $\lambda(\cdot)$ and $\gamma(\cdot)$ respectively. Thus,

$$\hat{z}_{t+\tau/t} = \gamma(\hat{x}_{t+\tau/t}) \tag{77}$$

This is an interesting result: if a random process z_t can be expressed as a function of a factorable process x_t, see eq.(68), then its estimator is expressed by the same function, see eq.(77).

Let the filtered state be

$$\hat{z}_{t/t} = \eta(\hat{x}_{t/t-1}) + \theta_t y_t = \sum_k \eta_{k,t}\, p_k(\hat{x}_{t/t-1}) + \theta_t y_t \tag{78}$$

where $\hat{x}_{t/t-1}$ is the factorable estimator (eq.(77)). The corresponding orthogonality equation

$E[\hat{z}_{t/t} - z_t]\, p_k(\hat{x}_{t/t-1}) = 0$, leads to $\eta_{k,t} = \gamma_{k,t} - \theta_t \mu_{k,t}$,

where $\mu_{k,t}$ are the expansion coefficients of the observation term $\mu(x_t) = h[\gamma(x_t)] = h(z_t)$. Substituting $\eta_{k,t}$ into eq.(78) and after taking into account eq.(77) we find eq.(70). If we insert (70) into the error variance $\sigma^2_{t/t} = E[\hat{z}_{t/t} - z_t]^2$, we derive eq.(72). Then eq.(73) is obtained if we combine eq.(72) and $\sigma^2_{t/t} = \theta_t E[v_t^2]/h_{1,t}$, and solve for θ_t.

Next, let the predicted state be

$$\hat{z}_{t/t-1} = \upsilon(\hat{x}_{t-1/t-1}) = \sum_k \upsilon_{k,t-1}\, p_k(\hat{x}_{t-1/t-1}) \tag{79}$$

Using the orthogonality equation

$E[\hat{z}_{t/t-1} - z_t]\, p_k(\hat{x}_{t-1/t-1}) = 0$, and following similar lines to that of the proof of Proposition 3, we yield eqs.(75) and (76).

At this point a fact to be stressed is that the transformation $\gamma(.)$ in eqs.(68) and (69) is not involved in the estimation algorithm. Though it is used throughout the proof to show that, if the proof of Proposition 3, we yield eqs.(75) and (76).

At this point a fact to be stressed is that the transformation $\gamma(.)$ in eqs.(68) and (69) is not involved in the estimation algorithm. Though it is used throughout the proof to show that, if such a transformation can be established between a factorable process x_t and a nonfactorable one z_t, the latter can be estimated by the algorithm discussed above. However, we need a priori values for the initial statistics $\hat{x}_{o/o}$, $\sigma^2_{o/o}$ and the error variances $E[w_t^2]$, $E[v_t^2]$. These are model parameters and may be selected through analysis of the empirical data or computer simulations (see following section for some detail).

REMARK 7: Particularly interesting may be these transformations $\gamma(.)$, which lead to two-dimensional Gaussian processes x_t. In fact, it is possible to construct a transformation such as in eq.(68) and (69) where x_t will be a random process with Gaussian one-dimensional distribution (see e.g. Cramer and Leadbetter (1967)). Furthermore, for many practical applications it is reasonable to assume that the two-dimensional distribution of x_t is Gaussian too. By way of example, consider the lognormal model commonly used in geostatistical applications: if z_t is a lognormal process, the transformation $\gamma(.)$ which maps a Gaussian process x_t into z_t is

$$z_t = \gamma(x_t) = \exp(a+cx_t) \tag{80}$$

where a and c are the expectation and the variance of the Gaussian process $\log(z_t)$. Many nonfactorable distributions which cannot be evaluated in closed form can be represented by approximations based on the Gaussian law (the best are those which are asymptotic, converging in some sense to normality; see, e.g., Patel and Read, 1982). Using the above approach consequently, the estimation problem is seen to become a problem of testing if a transformation can be established between the given process z_t and a factorable process x_t.

SOME PRACTICAL ASPECTS OF THE NSNOS ESTIMATION ALGORITHM

We distinguish two groups of parameters available in the proposed estimator: (1) those which control its operation and must be specified by the scientist in proper time, and (2) those which reflect the quality of the algorithm performance and, therefore, may be used to check if it is properly functioning. The group of control-parameters includes the variances $E[w_t^2]$, $E[v_t^2]$, the coefficients $f_{1,t-1}$, $h_{1,t}$ and the initial conditions $\hat{x}_{o/o}$, $\sigma^2_{o/o}$. To the group of test-parameters belong the gain β_t and the innovation process ε_t .

The parameters $E[w_t^2]$, $E[v_t^2]$, $\hat{x}_{o/o}$ and $\sigma^2_{o/o}$ have to be determined a priori. Their influence on the estimation accuracy has as follows: The accuracy of the estimator improves as $E[w_t^2]$ and/or $E[v_t^2]$ decrease. The initial estimation variance $\sigma^2_{o/o}$ affects the initial state estimates. However, it decreases drastically as more observations become available, meaning that errors in choosing a $\sigma^2_{o/o}$ value as well as errors in calculating further $\sigma^2_{t/t}$ values vanish asymptotically.

As follows from the algorithm, the gain β_t is directly related to variations in $E[w_t^2]$, $E[v_t^2]$ as well as in $\hat{x}_{t/t}$, $\sigma^2_{t/t}$. More specifically, if $E[w_t^2]$ is large (or $E[v_t^2]$ is small), the β_t is large and thus the estimator is governed by the observation model, the structural model being inadequate (in this case the estimates obtained are rather scaled observations.) If, instead the $E[w_t^2]$ is small (or $E[v_t^2]$ is large), the β_t will be small too and then the algorithm followss naturally the structural model. Innovation ε_t is more sensitive to changes in $E[v_t^2]$ than in $E[w_t^2]$ and this follows easily from eq.(31).

The structure of the algorithm requires the expansion coefficients $f_{1,t-1}$ and $h_{1,t}$ to be calculated on-line in terms of the statistics of the estimates; this can be done using eqs.(64), (65). However, the form of these equations must be determined a priori. For instance, eq.(64) writes

$$h_{1,t} = \int_R h(\hat{x}_{t/t-1},t)\ \hat{x}_{t/t-1} g(\hat{x}_{t/t-1})\ d\hat{x}_{t/t-1} \tag{81}$$

If the g(.) is known or is assumed, this integral may be calculated analytically (e.g, if g(.) is Gaussian and the h(.) is a polynomial, the form of the equation expressing $h_{1,t}$ is easily determined in terms of the moments of $\hat{x}_{t/t-1}$; see example below). In more complex situations, one may apply numerical integration methods. We will briefly describe such a situation here: assume that for the given observation process y_t, there exists a transformation $\gamma(.)$ such as $y_t = \gamma(u_t)$, where u_t is a standard Gaussian process. Numerical values $\gamma(u_{t,i})$ of this transformation may be determined graphically by studying the standard cumulative distribution curve of u_t and the experimental cumulative distribution curve of y_t. Using these values together with Hermite polynomials, the $\gamma(.)$ may be completely defined by

$$y_t = \gamma(u_t) = \sum_{k=0}^{K} \gamma_k\ p_k\ (u_t) \tag{82}$$

for proper K. The coefficients γ_k are calculated by the Hermite integration formula

$$\gamma_k = \frac{1}{k!\sqrt{2\pi}} \sum_{i=0}^{M} w_i\ \gamma(u_{t,i})\ p_k(u_{t,i})\ \exp[\ -\frac{u_{t,i}^2}{2}]$$

where the abscissas $u_{t,i}$ and the weight factors w_i are given (see, e.g., Abramowitz and Stegun, 1965). Next, since $\gamma(.)$ is determined, the probability density of y_t may be defined from eq.(82) in terms of the Gaussian density of u_t. Furthermore, using eq.(11) one may calculate the density of x_t. Supposing that x_t and $\hat{x}_{t/t-1}$ have the same density the coefficient $h_{1,t}$ can be readily calculated from eqs.(64). (Similar comments hold for the computation of the coefficient $f_{1,t-1}$.)

We will now discuss some statistical tests for the evaluation of the algorithm, from simulation results. If the true process x_t is known, one may test statistically the reasonableness of the obtained error variances by plotting a confidence interval, say $\pm 2\sqrt{\sigma^2_{t/t}}$, around the estimation error $\tilde{x}_{t/t} = x_t - \hat{x}_{t/t}$. The $\sigma^2_{t/t}$ will be considered reasonable if about 95% of the $\tilde{x}_{t/t}$ values lie within these limits. Since ε_t is a zero mean white noise, one may check the values provided by the algorithm; if the plot of ε_t vs t does not oscillate around zero, the initial conditions may need to be changed. More specifically, a number of statistical hypothesis tests about the postulated zero mean white noise ε_t may complete the evaluation of the algorithm. For example, at a 5% significance level the sample mean μ_ε of a Gaussian ε_t should be less than $1.96\sqrt{\sigma^2_\varepsilon/n}$ (n is the number of data and σ^2_ε is the sample variance), and 95% of the sample correlation coefficient $\rho_\varepsilon(s)$ obtained at several intervals $s = t_j - t_i$ should lie within $\pm 1.96/\sqrt{n}$,n>30. Furthermore, proper functioning of the algorithm implies that the interval $\pm 2\sqrt{r_t}$, where r_t is the innovation variance provided by the algorithm, should include about 95% of the sampled ε_t.(See also the example that follows.)

AN ILLUSTRATIVE EXAMPLE

In order to give some insight into the theoretical results of the preceding sections and to compare the proposed algorithm (abr. PA) and the well established extended Kalman algorithm (abr. EKA, see Jazwinski, 1970; Candy, 1986), a simple example has been simulated. The NSNOS is given by

$$x_t = 99.95 \times 10^{-2} x_{t-1} + 4 \times 10^{-4} x_{t-1}^2 + w_{t-1} \tag{84}$$

and

$$y_t = x_t^2 + x_t^3 + v_t \tag{85}$$

where the initial state x_o is Gaussian with mean 2 and variance 10^{-2}, and w_{t-1}, v_t are zero mean Gaussian white noises with variances 5×10^{-5} and 9×10^{-2}, respectively. The simulated measurements are shown in Figure 1. On the basis of eqs.(84) and (85) and the Gaussian assumption, the PA coefficients are found to be

$f_{1,t-1} = 99.95 \times 10^{-2}\, E[\hat{x}^2_{t-1/t-1}]$, $h_{1,t} = 3\, E[\hat{x}^2_{t/t-1}]^2$;

the EKA jacobians will be

$a_{1,t} = 99.95 \times 10^{-2} + 8 \times 10^{-4}\, \hat{x}_{t/t-1}$, $b_{1,t} = 2\, \hat{x}_{t/t-1} + 3\, \hat{x}^2_{t/t-1}$.

For both the PA and EKA we assume the initial conditions, $\hat{x}_{o/o} = 2$ and $\sigma^2_{o/o} = 10^{-2}$. The PA estimated state $\hat{x}_{t/t}$ is shown in Figure 2. Since we know the true state x_t, we can plot the estimation error $\tilde{x}_{t/t} = x_t - \hat{x}_{t/t}$ and use $\pm 2 \sqrt{\sigma^2_{t/t}}$ as confidence intervals, see Figure 3. The error $\tilde{x}_{t/t}$ is very small (its mean is $\sim 2.4\ 10^{-4}$) and only about 0.7% lies outside the bounds. The last PA variance available is close to the sample variance ($0.25\ 10^{-4} \sim 0.20\ \ 10^{-4}$) . In Figure 4 we show the nonlinear innovation process ε_t together with the confidence interval $\pm\ 2 \sqrt{r_t}$ provided by the PA. Here none of the ε_t values exceeds the bounds, while the mean is very close to zero ($\sim 3.4 \times 10^{-3} < 1.96 \sqrt{\ \sigma_\varepsilon^2/n} = 5.84 \times 10^{-2}$) . Figure 5 shows that only 2.67% of the ρ_ε lies outside the bounds $1.96/\sqrt{\ n} = 0.16$. Therefore, the ε_t is a zero mean white noise, fact that implies proper functioning of the PA. In Figure 6 we compare the estimation error $\tilde{x}_{t/t}$(PA) provided by the PA and that provided by the EKA, $\tilde{x}_{t/t}$ (EKA): the plot of $\Delta x_{t/t} = |\tilde{x}_{t/t}(PA)| - |\tilde{x}_{t/t}\ (EKA)|$ establishes the superiority of the PA (note that negative Δx values indicate more accurate PA, while positive Δx values indicate more accurate EKA).

To obtain the sensitivity of the two algorithms to variations in the statistics of the state noise w_t, we again consider the NSNOS of eqs.(84) and (85) where now the variance $E[w_t^2\ ]$ takes the values 10^{-5}, $5\ 10^{-5}$, and $9\ 10^{-5}$. The plots in Figure 7 show the PA estimation variance $\sigma^2_{t/t}$. Note the rapid reduction of the $\sigma^2_{t/t}$ with the number of observations processed leading to increasingly accurate estimates. This reduction seems to tend to continue and after the last observation available. The same for the EKA are shown in Figure 8. As the last two figures verify, the PA gives better results in this regard also (smaller $\sigma^2_{t/t}$ values and less sensitive to changes in $E[w_t^2]$).

SUMMARY - CONCLUSIONS

In this paper, we presented on-line recursive estimation algorithms for nonlinear state-nonlinear observation physical systems generating factorable random processes, or processes which, while are not factorable themselves, can be transformed to such processes. Such situations may arise in diffusion models of transport phenomena, real-time estimation of physical parameters, identification of environmental models, mine planning, etc. The proposed method combines these processes with orthogonal expansions of the nonlinearities in the state and observation models leading to efficient and mathematically meaningful estimates.

In summary, some notable features of the method proposed are: It takes advantage of exact nonlinear dynamic structural and measurement models to adequately describe spatial variations and measurement errors in the underlying processes. Approximations are introduced only via the nonlinear filtering and prediction models. Estimation is carried out using sequential algorithms that enable the available information to be analyzed and weighted with less computational cost than other methods. If processing a priori information, new data can be handled immediately and new estimates made. The present algorithm is suboptimal but very comprehensive, so it may serve as a simple substitute to substantially more complicated approaches, like martingale theory, Lie algebraic and differential geometric methods, etc. As the example illustrates, the method compares favorably in practice with the well established extended Kalman algorithm.

ACKNOWLEDGMENTS

The author wishes to express his gratitude to Prof. Tyrone Duncan, Department of Mathematics of the University of Kansas, for his very helpful discussions and advice concerning this study. Thanks are also owed to Dr. Michel Gevers, Universite Catholique de Lourain, Belgium, for his valuable recommendations, and to Cora Cowan and Marla Adkins-Heljeson for the careful editing and processing of this material.

REFERENCES

Abramowitz, M. and D. A. Stegun, 1965, "Handbook of Mathematical Functions," Dover, New York.

Anderson, B. D. and J. B. Moore, 1979, "Optimal Filtering," Englewood Cliffs, Prentice-Hall, New Jersey.

Allinger, D. F. and S. K. Mitter, 1981, "New Results on the Innovations Problem for Nonlinear Filtering", Stochastics, Vol. 4, No. 4, p. 339-348.

Athans, M., R. P. Wishner and A. Bertolini, 1968, "Suboptimal State Estimation for Continuous-time Nonlinear System from Discrete Noisy Measurements," IEEE Trans. AC-13, 5, p. 504-514.

Bras, R. L. and K. P. Georgakakos, 1980," Real Time Non-linear Filtering Techniques in Streamflow Forecasting: A Statistical Linearization Approach." In Proc. 3rd Intern. Sympos. on Stochastic Hydraulics, Aug. 5-7, Tokyo, p. 95-105.

Brockett, R.W., 1980, "Nonlinear Systems and Nonlinear Estimation

Theory," Proceedings NATO/ASI, Savoie, France, p. 441-477.

Candy, J. C., 1986,"Signal Processing: The Model-based Approach," McGraw-Hill, New York.

Christakos, G., 1985, "Recursive Parameter Estimation with Applications in Earth Sciences," Jour. of Mathematical Geology, v. 17, no. 5, p. 489-515.

Christakos, G., 1987, "On-line estimation of nonlinear physical systems," Jour. of Mathematical Geology, v. 19, no. 8, in press.

Cramer, H., 1945, "Mathematical Methods of Statistics," Princeton Univ. Press, New Jersey.

Cramer, H. and M. R. Leadbetter, 1967, "Stationary and Related Stochastic Processes," J. Wiley, New York.

Culver, C. O., 1969, "Optimal Estimation for Nonlinear Systems," AIAA Conf. Princeton, Paper no. 69-852, New Jersey.

Doob, J. L., 1953, "Stochastic Processes," J. Wiley, New York.

Gelb, A., 1974, "Applied Optimal Estimation," MIT Press, Cambridge, MA.

Harris, B. and A. P. Soms, 1980, "The Use of the Tetrachoric Series for Evaluating Multivariate Normal Probabilities," Jour. of Multivariate Anal., Vol 10, p. 252-267.

Jackson, M. and Marechal, A., 1979, "Recoverable Reserves Estimated by Disjunctive Kriging: a Case Study," in 16th Appl. of Comput. and Operat. Research in the Mineral Industry (O'Neil, T.J., ed): Soc. of Mining Eng. of American Inst. of Mining and Metallurgy, New York, p. 240-249.

Jaswinski, A. H. 1966, "Filtering of Nonlinear Systems," IEEE Trans. AC-11, p. 765-766.

Jaswinski, A. H., 1970, "Stochastic Processes and Filtering Theory," Academic Press, New York.

Kailath, T., 1971, "Some Extensions of the Innovation Theorem," Bell Syst. Tech. J., v. 50, p. 1487-1494.

Kallianpur, G., 1980, "Stochastic Filtering Theory," Springer-Verlag, New York.

Mahalanabis, A. K. and M. Farooq, 1971, "A Second-Order Method for State Estimation of Nonlinear Dynamic Systems": Int. J. Control, 14, p. 631-639.

Marcus, S. I., 1979, "Optimal Nonlinear Estimation for a Class of Discrete-Time Stochastic Systems," IEEE Trans. AC-24, no. 2, p. 197-302.

Matheron, G., 1976, "A Simple Substitute for Conditional Expectation: the Disjunctive Kriging," in Adv. Geostatistics in Mining Industry (Guarascio, M., David, M., and Huijbregts, eds.): Reidel, Boston, p. 221-236.

Mein, R. G., E. M. Laurenson, and T. A. McMahon, 1974, "Simple Non-linear Model for Flood Estimation," ASCE Jour. Hydraulic, Vol. 100, No. HY11, p. 1507-1518.

Nahi, N. E., 1969, "Estimation Theory and Applications," J. Wiley, New York.

Pardoux, E., 1979, "Stochastic Partial Differential Equations and Filtering of Diffusion Processes," Stochastics, 2, p. 127-168.

Patel, J. K. and C. B. Read, 1982, "Handbook of the Normal Distribution", Marcel Dekker Inc., New York.

Pearson, K., 1930, "Tables for Statisticians and Biometricians", Part I (3rd ed.), Cambridge University, Cambridge.

Sorenson, H. W. and A. R. Stubberud, 1968, "Nonlinear Filtering by Approximation of the A Posteriori Density," Int. Control, Vol. 8, No. 1, p. 33-51.

Stratonovich, R. L., 1963, "Topics in the Theory of Random Noise," Vols. 1 and 2, Gordan and Breach, New York.

Szego, G., 1959, "Orthogonal Polynomials," J. Wiley, New York.

Willman, W., 1981, "Edgeworth Expansions in State Perturbation Estimation," IEEE Trans. Automatic Control, Vol. AC-26, No. 2, p. 493-498.

Zakai, M., 1969," On the Optional Filtering of Diffusion Processes," Z. Wahr. Verw. Geb., Vol. 11, p. 230-243.

FIGURE LEGENDS

Figure 1: Simulated measurements.

Figure 2: Estimated states.

Figure 3: Error estimates and confidence intervals.

Figure 4: Nonlinear innovation process and confidence intervals.

Figure 5: Differences in error estimates, Proposed vs extended Kalman estimators.

Figure 6: Sampled correlation and confidence intervals.

Figure 7: Error variances for Proposed estimator assuming,

$E[w_t^2] = 9\ 10^{-5}$ (labeled --), $= 5\ 10^{-5}$ (labeled ···)

and $= 10^{-5}$ (labeled —).

Figure 8: Error variances for extended Kalman estimator assuming,

$E[w_t^2] = 9\ 10^{-5}$ (labeled --), $= 5\ 10^{-5}$ (labeled ···)

and $= 10^{-5}$ (labeled —).

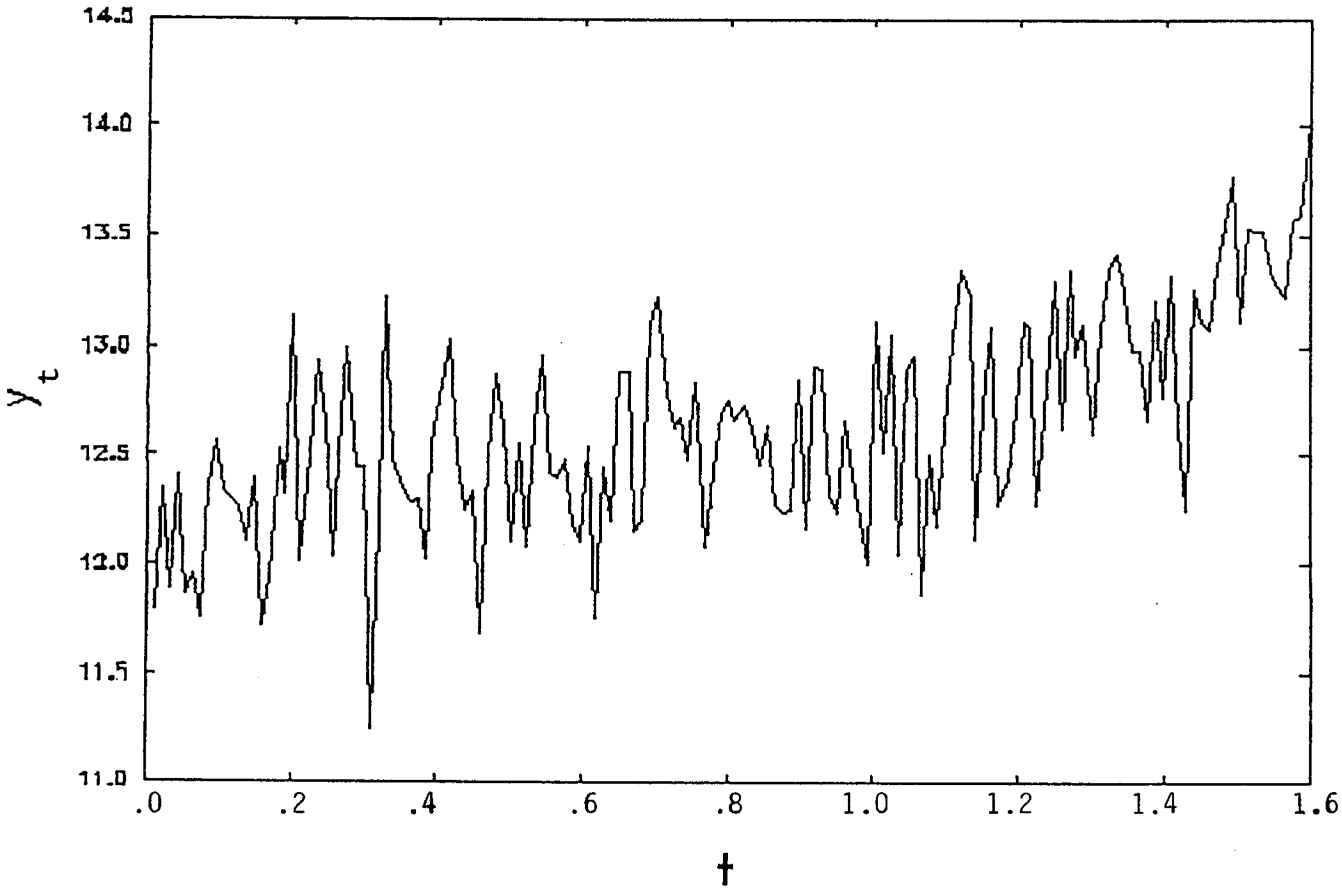
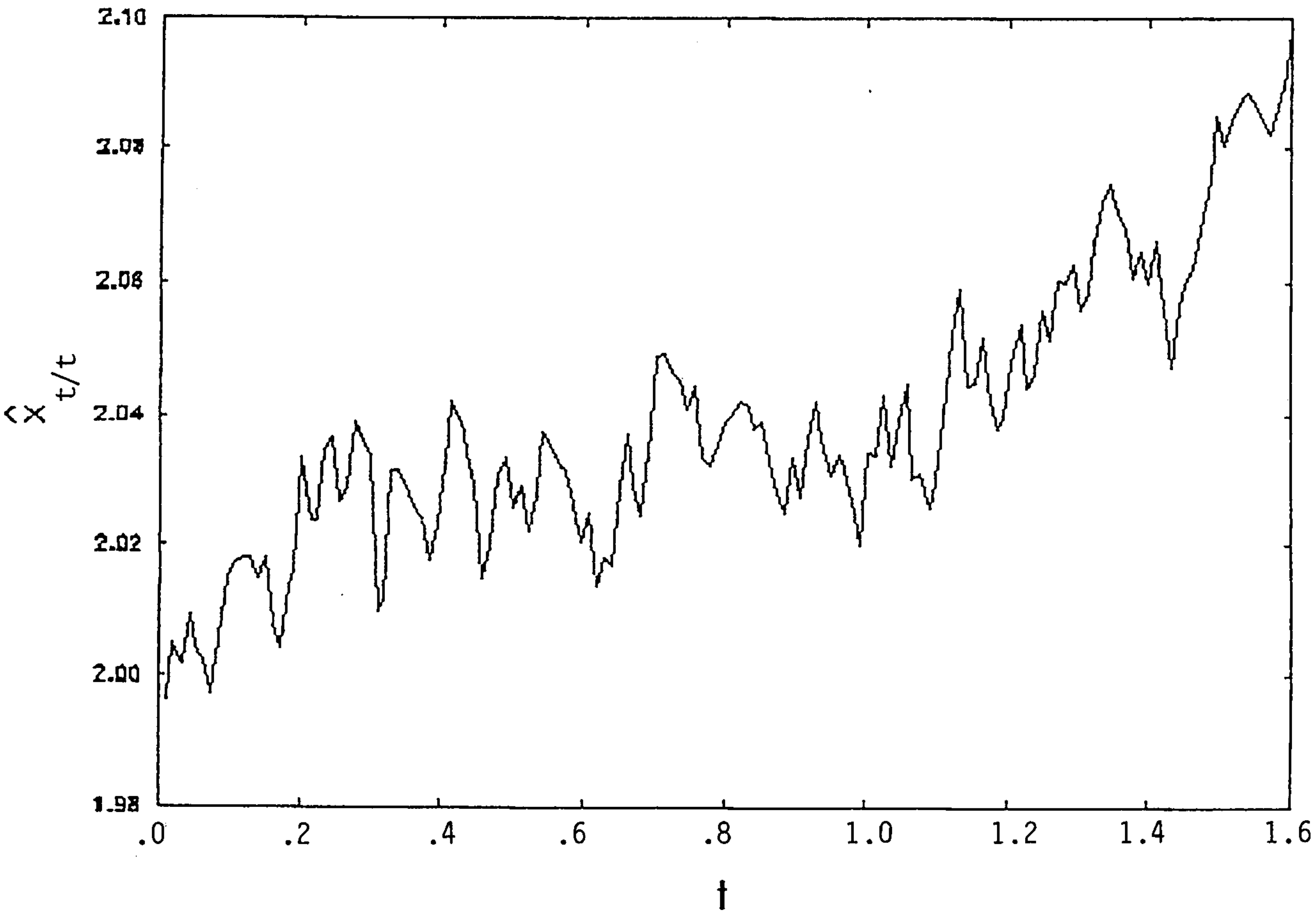
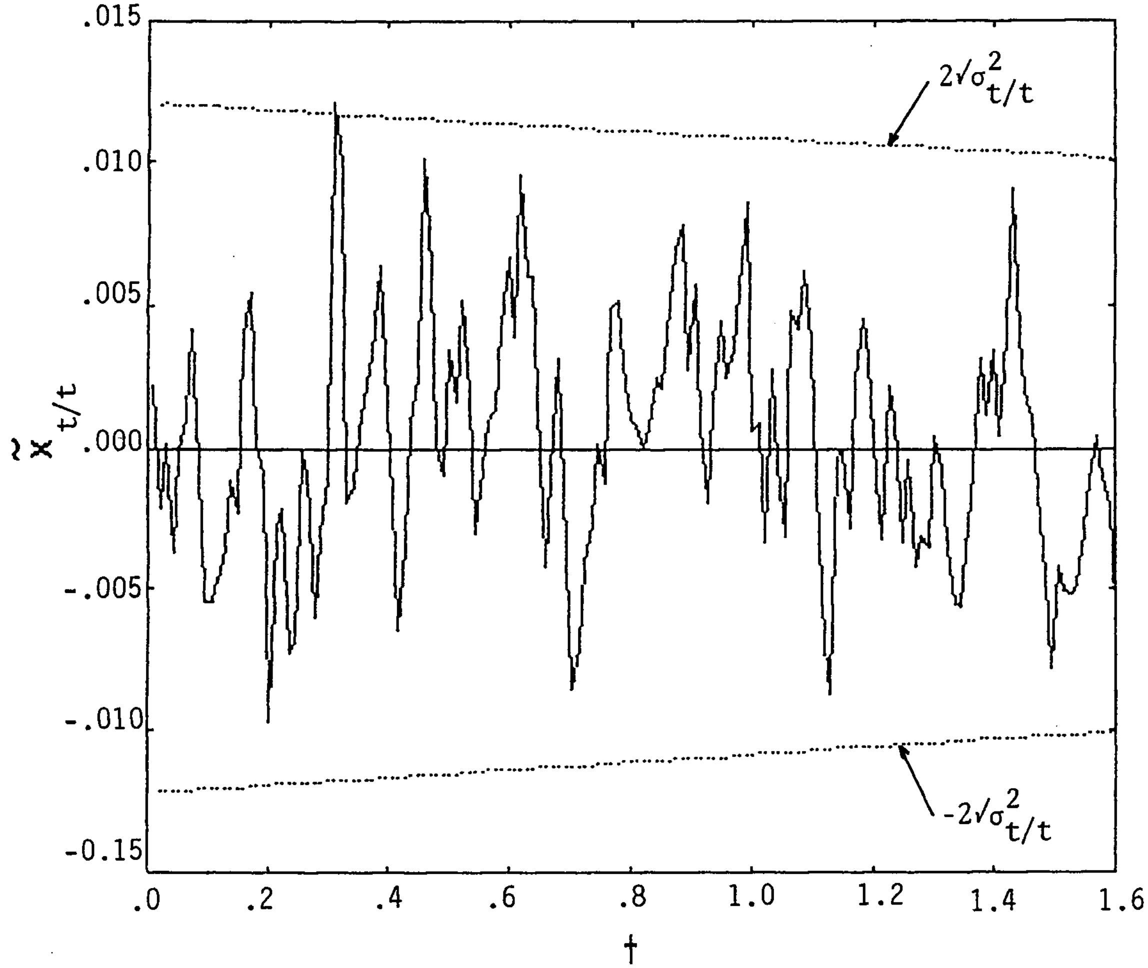


Fig. 1

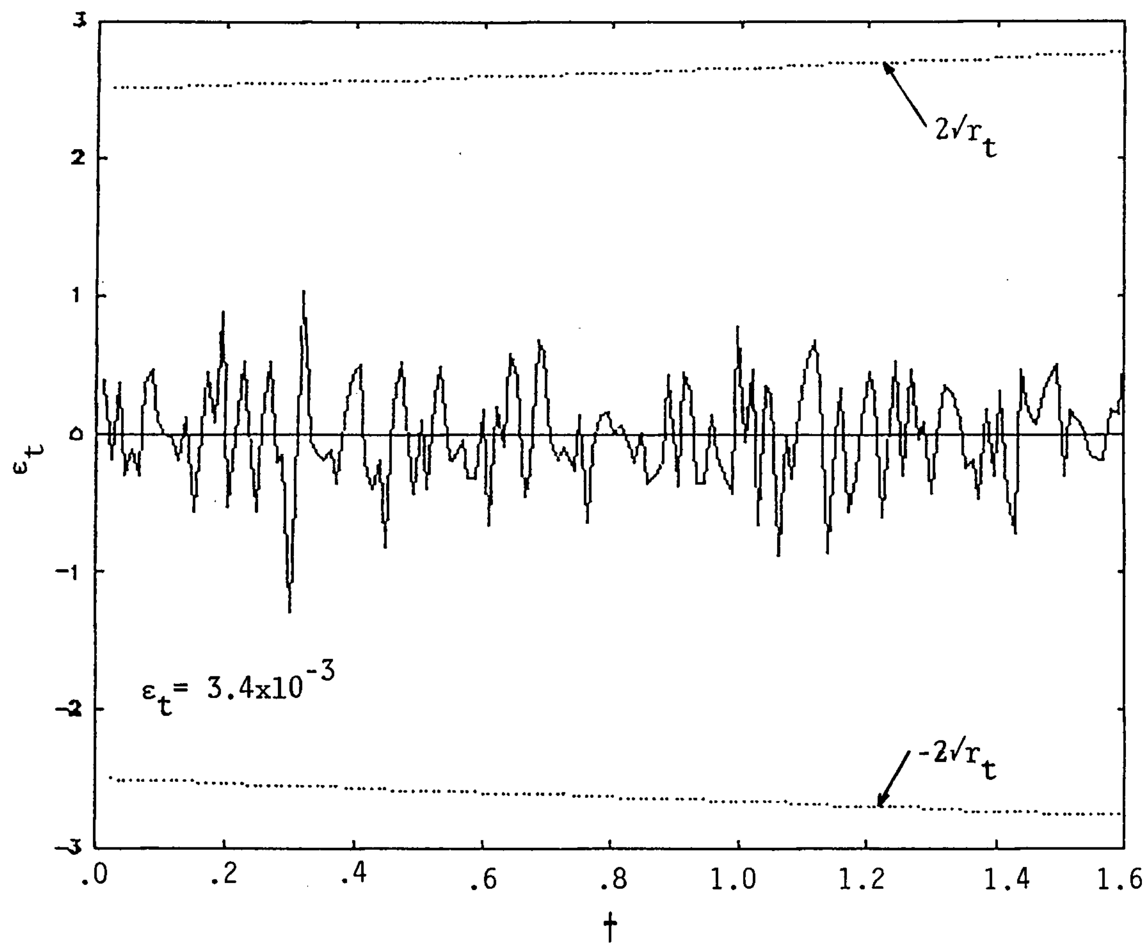


2.10
2.08
2.06
2.04
2.02
2.00
1.98
$\hat{x}_{t/t}$
.0
.2
.4
.6
.8
1.0
1.2
1.4
1.6
t



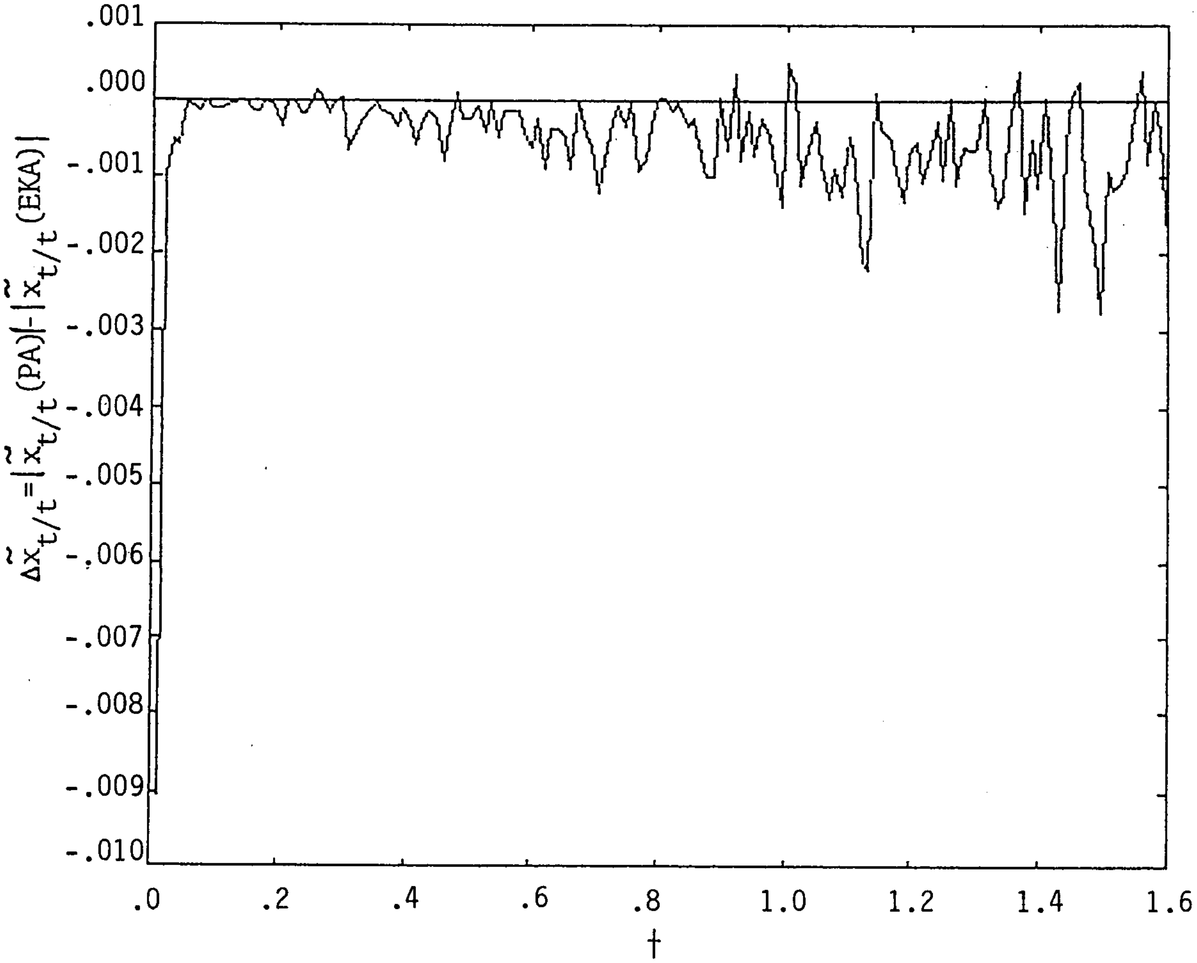
.015
.010
.005
.000
-.005
-.010
-0.15
$\tilde{x}_{t/t}$
$2\sqrt{\sigma^2_{t/t}}$
$-2\sqrt{\sigma^2_{t/t}}$
.0
.2
.4
.6
.8
1.0
1.2
1.4
1.6
t

Fig. 3



3
2
1
0
-1
-2
-3
ε_t
$2\sqrt{r_t}$
$-2\sqrt{r_t}$
$\varepsilon_t = 3.4\times10^{-3}$
.0
.2
.4
.6
.8
1.0
1.2
1.4
1.6
t

Fig. 4



.001
.000
-.001
-.002
-.003
-.004
-.005
-.006
-.007
-.008
-.009
-.010
$\Delta\tilde{x}_{t/t}=|\tilde{x}_{t/t}(PA)|-|\tilde{x}_{t/t}(EKA)|$
.0
.2
.4
.6
.8
1.0
1.2
1.4
1.6
t

Fig 5

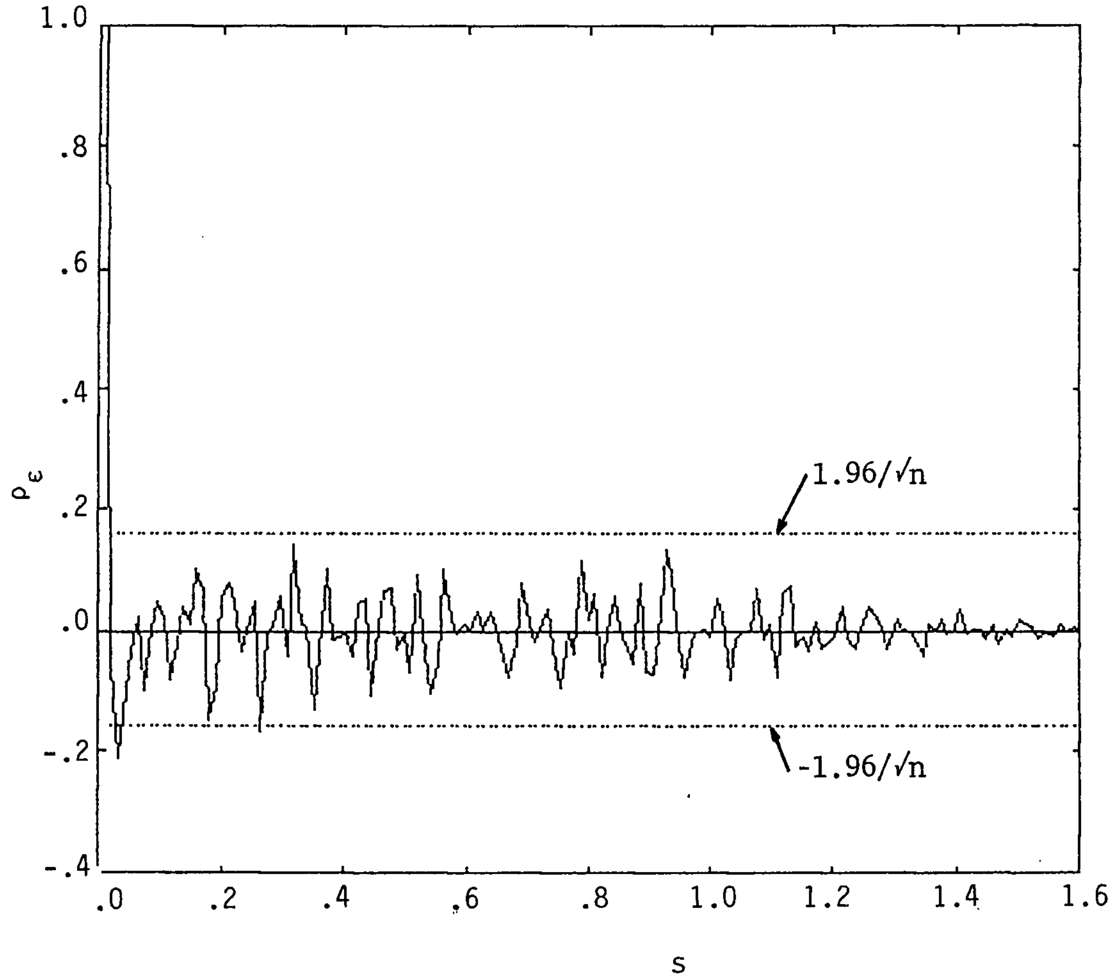
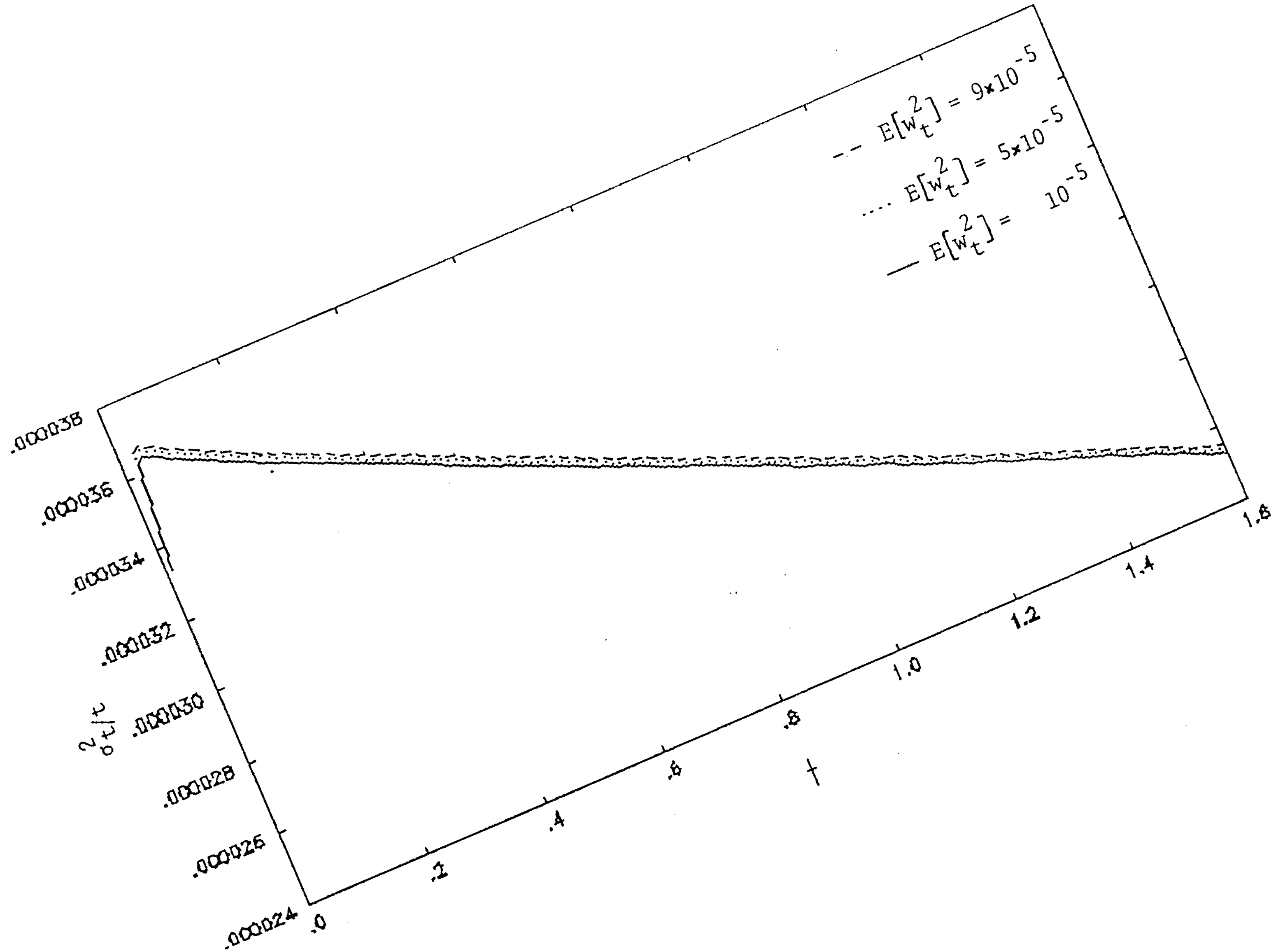
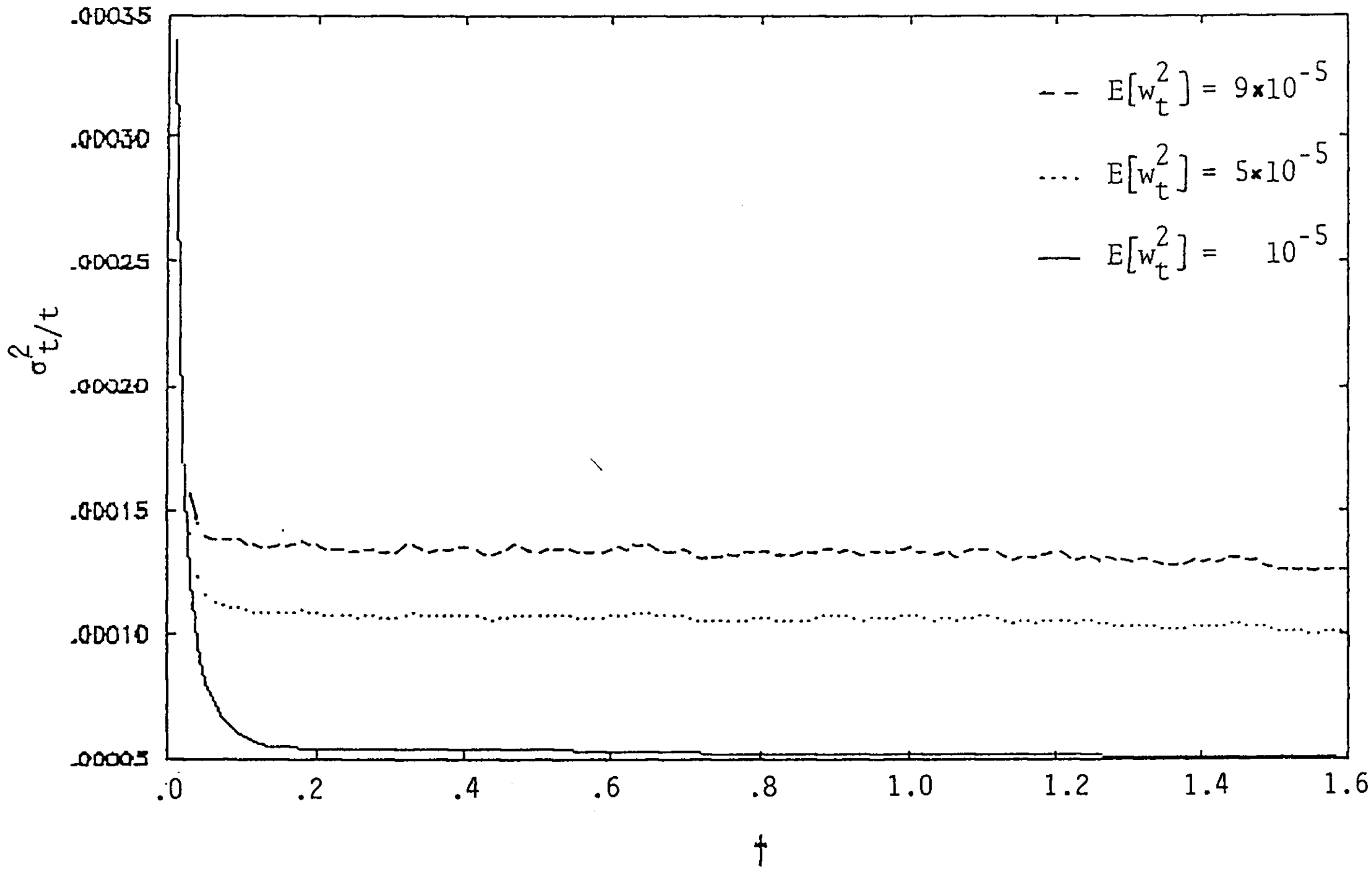


Fig. 6



-- $E[w_t^2] = 9 \times 10^{-5}$
.... $E[w_t^2] = 5 \times 10^{-5}$
— $E[w_t^2] = 10^{-5}$
.000038
.000036
.000034
.000032
.000030
.000028
.000026
.000024
σ_t^2/t
.0
.2
.4
.6
.8
1.0
1.2
1.4
1.6
t

Fig-7



.00035
.00030
.00025
.00020
.00015
.00010
.00005
σ_t^2/t
.0
.2
.4
.6
.8
1.0
1.2
1.4
1.6
t
$E[w_t^2] = 9\times10^{-5}$
$E[w_t^2] = 5\times10^{-5}$
$E[w_t^2] = 10^{-5}$

Fig. 8

S U P P L E M E N T: The factorable random processes

ON THE POLYNOMIAL EXPANSION OF BIVARIATE PROPABILITY DENSITIES

Let x_t, $t \epsilon T$ (T is a reference set) be a random function and let $g(\chi_i, \chi_j)$ be the bivariate probability density of the random variables x_i and x_j at $t = t_i$ and $t = t_j$ respectively. Furthermore, assume that $p_k(\chi_i)$ and $q_\lambda(\chi_j)$, $k, \lambda = 0, 1, \ldots$, are two complete sets of orthonormal polynomials having weighting functions the univariate densities $g_i(\chi_i)$ and $g_j(\chi_j)$ respectively, i.e.,

$$\int p_k(\chi_i)\, p_{k'}(\chi_i)\, g_i(\chi_i)\, d\chi_i = \delta_{kk'}, \qquad \text{and} \tag{1}$$

$$\int q_\lambda(\chi_j)\, q_{\lambda'}(\chi_j)\, g_j(\chi_j)\, d\chi_j = \delta_{\lambda\lambda'} \tag{2}$$

where $\delta_{k\lambda}$ (=1 if $k=\lambda$, =0 otherwise) is the Kronecker-delta. Accordingly, one may define the correlation coefficients $u_{k\lambda}$ as

$$u_{k\lambda} = \iint p_k(\chi_i)\, q_\lambda(\chi_j)\, g(\chi_i, \chi_j)\, d\chi_i\, d\chi_j \tag{3}$$

and the coefficient r by

$$r = \int \int K^2(\chi_i, \chi_j)\, g_i(\chi_i)\, g_j(\chi_j)\, d\chi_i\, d\chi_j \tag{4}$$

where $K(\chi_i, \chi_j)$ is such that

$$K(\chi_i,\chi_j) = \frac{g(\chi_i,\chi_j)}{g_i(\chi_i)\ g_j(\chi_j)} \tag{5}$$

If r is finite, the bilinear form

$$U(p_k,q_\lambda) = \sum_k \sum_\lambda u_{k\lambda}\ p_k(\chi_i)\ q_\lambda(\chi_j), \qquad \text{or} \tag{6}$$

$$U(p_k,q_\lambda) = P^T U\ Q \tag{7}$$

where,

$$P = \begin{bmatrix} p_0(\chi_i) \\ \vdots \\ p_k(\chi_i) \end{bmatrix}, \quad Q = \begin{bmatrix} q_0(\chi_j) \\ \vdots \\ q_\lambda(\chi_j) \end{bmatrix} \quad \text{and} \quad U = \begin{bmatrix} u_{00} & \cdots & u_{0\lambda} \\ \vdots & & \vdots \\ u_{k1} & \cdots & u_{k\lambda} \end{bmatrix}$$ is the matrix of the

form, offers the best approximation of the quantity $K(\chi_i,\chi_j)$ of Eq. (5) in the mean square sense. Indeed, since $p_k(\chi_i)$ and $q_\lambda(\chi_j)$, $k,\lambda = 0,1, \ldots,$ are two complete sets of orthonormal polynomials, the functions $p_k(\chi_i)\ q_\lambda(\chi_j)$ also form a complete orthonormal set of polynomials (Courant and Hilbert, 1953). The $K(\chi_i,\chi_j)$ is square integrable by Eq. (4) and the hypothesis that r is finite and, thus, the minimization of

$$\int \int [K(\chi_i,\chi_j) - U^*(p_k,q_\lambda)]^2\ g_i(\chi_i)\ g_j(\chi_j)\ d\chi_i\ d\chi_j$$

where

$$U^*(p_k,q_\lambda) = P^T U^* Q\ , \qquad \text{and}$$

$$U^* = \begin{bmatrix} \tau_{00} & \ldots & \tau_{1\lambda} \\ \vdots & & \vdots \\ \tau_{k1} & \ldots & \tau_{k\lambda} \end{bmatrix}$$

with respect to the coefficients $\tau_{k\lambda}$, $k,\lambda=0,1,\ldots$, implies

$$U = U^{*} \tag{8}$$

When $k,\lambda \to \infty$, the approximation (6) becomes equal to $K(\chi_i,\chi_j)$, i.e.

$$K(\chi_i,\chi_j) = U(p_k,q_\lambda) \tag{9}$$

Moreover it is valid

$$\sum_{k=0}^{\infty} \sum_{\lambda=0}^{\infty} u_{k\lambda}^2 = r \tag{10}$$

which is the completeness relation of the linear algebra theory (Gel'fand, 1961). For instance, if the density $g(\chi_i,\chi_j)$ is bivariate Gaussian, $r = 1/(1-\rho^2)$, $|\rho|<1$; ρ is the correlation coefficient.

The polynomials $p_k(\chi_i)$ and $q_\lambda(\chi_j)$ depend, in general, on the parameters t_i and t_j, and so do the correlation coefficients $u_{k\lambda}$ and the univariate and the bivariate densities. Taking into account that the $g_i(\chi_i)$, $g_j(\chi_j)$ and $g(\chi_i,\chi_j)$ are probability densities, one finds

$$p_0(\chi_i) = q_0(\chi_j) = 1$$

$$p_1(\chi_i) = \{\chi_i - E[x_i]\} \;/\; \{Var[x_i]\}^{1/2} \tag{11}$$

$$q_1(\chi_j) = \{\chi_j - E[x_j]\} \;/\; \{Var[x_j]\}^{1/2}$$

These results demonstrate that the mean and variance of the two variables may be derived from the orthogonal polynomials and vice versa.

The Importance of the Correlation Coefficients $u_{k\lambda}$

The above procedure shows that the bivariate probability density may be defined in terms of the two univariate densities and the coefficients $u_{k\lambda}$. This fact depicts that it may be useful that the latter be studied in more depth. A straight-forward application of the Schwartz's inequality yields

$$|u_{k\lambda}| = |E[p_k(x_i)\ q_\lambda(x_j)]| < \{E[p_k^2(x_i)]\}^{1/2} \{E[q_\lambda^2(x_j)]\}^{1/2}$$

or, due to Eqs. (1) and (2),

$$|u_{k\lambda}| < 1 \tag{12}$$

for all k,λ. Furthermore, since $g_j(\chi_j)$ is given by

$$g_j(\chi_j) = \int g(\chi_i,\chi_j)\ d\chi_i$$

one may use Eqs. (1) through (4) to find

$$1 = \sum_{j=0}^{\infty} u_{0j}\ q_j(\chi_j)$$

or, due to eq. (11),

$$u_{00}=1, \qquad \text{and} \tag{13}$$

$$u_{0\lambda} = 0 \tag{14}$$

for $\lambda > 0$. Similarly one may prove that

$$u_{k0} = 0 \tag{15}$$

for $k > 0$. Finally an important property of the coefficients $u_{k\lambda}$ can be established as follows: by defining the covariance function

$$\begin{aligned} c_x(t_i,t_j) &= E\,\{x_i - E[x_i]\}\,\{x_j - E[x_j]\} \\ &= \iint \{\chi_i - E[x_i]\}\,\{\chi_j - E[x_j]\}\; g(\chi_i,\chi_j)\, d\chi_i\, d\chi_j, \end{aligned}$$

and by using eqs. (1), (2), (3) and (9), one finds

$$c_x(t_i,t_j) = u_{11}\,\{\mathrm{Var}\,[x_i]\}^{1/2}\,\{\mathrm{Var}\,[x_j]\}^{1/2} \tag{16}$$

(i.e., u_{11} is the correlation coefficient of the variables x_i and x_j). Note that in this section no restriction regarding the statistical regularity (such as stationarity) of the function x_t is being imposed. So, the covariance $c_x(t_i, t_j)$ is in general a function of t_i and t_j.

Since eq. (6) is a bilinear form in the polynomials $p_k(\chi_i)$ and $q_\lambda(\chi_j)$ and $u_{k\lambda}$ is the mean value of the product $p_k(\chi_i)\, q_\lambda(\chi_j)$, one is able to go from the bilinear form (6) to its matrix representation (7) and vice versa by inspection.

Example 1: Assume that the univariate density is uniform over the range [0, 1]. The associated orthonormal polynomials are Legendre and the bilinear form is

$$U(p_k,q_\lambda) = 1 + \sqrt{0.75}\times10^{-1}(2\chi_i-1)(2\chi_j-1) + 3\sqrt{5}\times10^{-2}[3(2\chi_i-1)^2-1][3(2\chi_j-1)^2-1] + 2.5\sqrt{7}\times10^{-2}[5(2\chi_i-1)^3-3(2\chi_i-1)]\,[5(2\chi_j-1)^3-3(2\chi_j-1)]$$

Writing the last equation according to the terminology of eq. (6) one finds

$$U(p_k,p_k) = u_{00}\, p_0(\chi_i)\, p_0(\chi_j) + u_{11}\, p_1(\chi_i)\, p_1(\chi_j) + u_{22}\, p_2(\chi_i)\, p_2(\chi_j) + u_{33}\, p_3(\chi_i)\, p_3(\chi_j),$$

where, $p_0(\chi)= 1$, $p_1(\chi)= \sqrt{12}\chi-\sqrt{3}$, $p_2(\chi)= 6\sqrt{5}\,[3(2\chi-1)^2-1]$ and

$$p_3(\chi) = \frac{5\sqrt{7}}{2}\,[5(2\chi-1)^3-3(2\chi-1)].$$

Therefore, the matrix U in eq. (7) is found, by inspection, to be

$$U = \begin{bmatrix} u_{00} & 0 & 0 & 0 \\ 0 & u_{11} & 0 & 0 \\ 0 & 0 & u_{22} & 0 \\ 0 & 0 & 0 & u_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.05 & 0 & 0 \\ 0 & 0 & 0.08 & 0 \\ 0 & 0 & 0 & 0.04 \end{bmatrix}$$

Nonlinear Transformations of Random Functions

In many applications one may define an output function z_t which is the nonlinear, in general, transformation of the input function x_t, i.e.

$$z_t= f(x_t,\ t) \tag{17}$$

Such transformations are very common in geotechnical engineering: for example, numerical models (finite elements, etc) or elastoplastic methods. In geostatistics nonlinear transformations arise when data must be transformed to normality so that nonlinear kriging estimators apply. The correlation structure of the output function will be defined by the covariance below

$$
\begin{aligned}
c_z(t_i,t_j) &= E\{z_i - E[z_i]\}\ \{z_j - E[z_j]\} \\
&= E\{f(x_i,t_i) - E[f(x_i,t_i)]\}\ \{f(x_j,t_j) - E[f(x_j,t_j)]\} \\
&= \iint \{f(\chi_i,t_i) - E[f(x_i,t_i)]\}\ \{f(\chi_j,t_j) - E[f(x_j,t_j)]\} \\
&\qquad g(\chi_i,\chi_j)\ d\chi_i\ d\chi_j
\end{aligned} \tag{18}
$$

It is easy to see that

$$
\begin{aligned}
&\iint E[f(x_i,t_i)]\ \{f(\chi_j,t_j) - E[f(x_j,t_j)]\}\ g(\chi_i,\chi_j)\ d\chi_i\ d\chi_j \\
&= E[f(x_i,t_i)]\ \{E[f(x_j,t_j)] - E[f(x_j,t_j)]\} = 0
\end{aligned}
$$

and thus

$$
c_z(t_i,t_j) = E\{f(x_j,t_j) - E[f(x_j,t_j)]\}\ f(x_i,t_i),\ \text{and} \tag{19}
$$

$$
c_z(t_i,t_j) = E\{f(x_i,t_i) - E[f(x_i,t_i)]\}\ f(x_j,t_j) \tag{20}
$$

To proceed one may assume

$$
f(\chi_i,t_i) = \sum_{k=0}^{\infty} f_{k,i}\ p_k(\chi_i) \tag{21}
$$

and

$$f(\chi_j,t_j) = \sum_{\lambda=0}^{\infty} f_{\lambda,j}\, q_\lambda(\chi_j) \tag{22}$$

where

$$f_{k,i} = E\left[f(x_i,t_i)\, p_k(x_i)\right] \tag{23}$$

and

$$f_{\lambda,j} = E\left[f(x_j,t_j)\, q_\lambda(x_j)\right] \tag{24}$$

respectively. (Here it is assumed that the functions used are such that the expansions exist.) Then eq. (18) writes

$$c_z(t_i,t_j) = \iint \sum_{k=1}^{\infty} f_{k,i}\, p_k(\chi_i) \sum_{\lambda=1}^{\infty} f_{\lambda,j}\, q_\lambda(\chi_j)$$
$$\sum_{\mu=0}^{\infty} \sum_{\nu=0}^{\infty} u_{\mu\nu}\, p_\mu(\chi_i)\, q_\nu(\chi_j)\, g(\chi_i)\, g(\chi_j)\, d\chi_i\, d\chi_j$$

where eq. (9) has been inserted. Taking advantage of the orthonormality relationships (1), (2) one finds

$$c_z(t_i,t_j) = \sum_{k=1}^{\infty} \sum_{\lambda=1}^{\infty} u_{k\lambda}\, f_{k,i}\, f_{\lambda,j} \tag{25}$$

(Eqs. (16) and (25) stress the importance of the coefficients $u_{k\lambda}$ in the determination of the correlation structure of a random function.)

<u>Example 2</u>: Let x_i, x_j be two jointly Gaussian random variables with bivariate density

$$g(\chi_i,\chi_j) = [2\pi\sigma_i^2\ \sigma_j^2 \sqrt{1-\rho_{i,j}^2}]^{-1} \exp\{-[\frac{\chi_i^2}{\sigma_i^2} + \frac{\chi_j^2}{\sigma_j^2}$$

$$- 2\rho_{i,j} \frac{\chi_i}{\sigma_i} \frac{\chi_j}{\sigma_j}]\ [1 - \rho_{ij}^2]^{-1}\} \tag{26}$$

where, $\sigma_i = \{Var[x_i]\}^{1/2}$ and $\sigma_j = \{Var[x_j]\}^{1/2}$. This density may be expressed in a bilinear form, where

$g_i(\chi_i) = (2\pi\sigma_i^2)^{-1/2}\exp[-\chi_i^2/2\sigma_i^2]$, $g_j(\chi_j) = (2\pi\sigma_j^2)^{-1/2}\exp[-\chi_j^2/2\sigma_j{}^2]$ and

$p_k(\chi_i) = (2^k k!)^{-1/2}\ H_k(\chi_i/\sqrt{2}\sigma_i)$, $q_\lambda(\chi_j) = (2^\lambda \lambda!)^{-1/2}\ H_\lambda(\chi_j/\sqrt{2}\sigma_j)$ and

$$u_{k\lambda} = \begin{cases} u_k = \rho_{ij}^k, & \text{if } k=\lambda \\ 0, & \text{otherwise.} \end{cases}$$

Now, let $z_t = f(x_t,t) = x_t^2$. The $c_z(t_i,t_j)$ will be of the form (25) which, taking into account the above results, reduces to

$$c_z(t_i,t_j) = \sum_{k=1}^{\infty} u_k\ f_{k,i}\ f_{k,j} \tag{27}$$

where, in this case

$f_{k,i} = f_{k,j} = 0$ for $k \neq 2$, and $f_{2,i} = \sqrt{2}\ \sigma_i^2$, $f_{2,j} = \sqrt{2}\ \sigma_j^2$. Thus,

$$c_z(t_i,t_j) = 2\ \sigma_i^2\ \sigma_j^2\ \rho_{ij}^2 \tag{28}$$

(Note that if x_i and x_j are uncorrelated, then so are $f(x_i,t_i)$ and $f(x_j,t_j)$).

<u>Example 3</u>: An interesting special case of Eq. (25) is to choose $u_{k\lambda}$ such as

$$u_{k\lambda} = b_k c_\lambda u_{11} \tag{29}$$

where $k,\lambda>1$; b_k, c_λ are real coefficients and $b_1 = c_1 = 1$. (This choice is not without a good reason, for eq. (29) is a linear relation of the correlation structure of the random variables $p_k(x_i)$, $q_\lambda(x_j)$ and that of x_i, x_j.) Eq. (25) simplifies to

$$c_z(t_i,t_j) = u_{11} \sum_{k=1}^{\infty} \sum_{\lambda=1}^{\infty} b_k\, c_\lambda\, f_{k,i}\, f_{\lambda,j} \tag{30}$$

which using Eq. (16) writes

$$c_z(t_i,t_j) = \alpha_f\, c_x(t_i,t_j) \tag{31}$$

where the coefficient α_f depends on the functions $f(\chi_i,t_i)$ and $f(\chi_j,t_j)$ and is such as

$$\alpha_f = \frac{\sum_{k=1}^{\infty} \sum_{\lambda=1}^{\infty} b_k\, c_\lambda\, f_{k,i}\, f_{\lambda,j}}{\{\mathrm{Var}\,[x_i]\}^{1/2}\,\{\mathrm{Var}\,[x_j]\}^{1/2}} \tag{32}$$

(In effect, it is not difficult to show that Eq. (29) is a necessary and sufficient condition for the validity of Eq. (31).)

A Method for Deriving Orthogonal Polynomials

The choice of the polynomials which are orthogonal with respect to the univariate density is of primary importance in the establishment of polynomial expansions for the bivariate density. Here we will discuss a powerful and comprehensive approach of constructing set of polynomials which are orthogonal with respect to a given probability density. The approach is based on some results by Lavrenter and Schabat (1967) and has as follows:

Let $g(\chi)$ be a density function and let χ_1 and χ_2 be the limits of the range of interest of χ. If $g(\chi)$ is such as

$$\frac{\frac{dg(\chi)}{d\chi}}{g(\chi)} = \frac{w(\chi)}{v(\chi)} \tag{33}$$

where $$w(\chi) = w_1\chi + w_o$$
$$v(\chi) = v_2\chi^2 + v_1\chi + v_o \tag{34}$$

and

$$g(\chi_1)\ v(\chi_1) = g(\chi_2)\ v(\chi_2) = 0, \tag{35}$$

then the set of orthogonal polynomials corresponding to $g(\chi)$ is given by

$$p_k(\chi) = \frac{1}{g(\chi)} \frac{d^k}{d\chi^k} [v^k(\chi)\ g(\chi)] \tag{36}$$

For several types of important orthogonal polynomials which include the Hermite, Chebyshev, Laguerre, Legendre and Gegenbauer polynomials, eq. (36) is in agreement with the Rodrigues' formula.

Example 4: Let

$$g(\chi) = (1- \chi^2)^{\alpha- 1/2} \tag{37}$$

where $\alpha > o$ and $-1 < \chi < 1$. Following the procedure above one finds

$$\frac{\frac{dg(\chi)}{d\chi}}{g(\chi)} = \frac{(1-2\alpha)\ \chi}{1-\chi^2}\ ;\quad \text{thus, } v(\chi) = -\chi^2 + 1. \quad \text{Then, } g(\chi_1 = -1)\ v(x_1 = -1)$$

$= g(\chi_2 = 1)\ v(\chi_2 = 1) = 0$, so that

$$p_k(\chi) = \frac{1}{(1-\chi^2)^{\alpha-1/2}} \frac{d^k}{d\chi^k} [(1-\chi^2)^{k+\alpha-1/2}] = \text{const. } C_k^{(\alpha)}(\chi) \tag{38}$$

where $C_k^{(\alpha)}(.)$ are Gegenbauer polynomials.

FACTORABLE RANDOM FUNCTIONS

Definitions

In the previous section some quite general results have been presented about random functions. In this section the attention will be restricted to more specific classes of random function. It is computationally convenient to require that the coefficients in Eq. (7) are such as

$$u_{k\lambda} = E[p_k(x_i)\ q_\lambda(x_j)] = \delta_{k\lambda}\ u_k \tag{39}$$

for all $k \neq \lambda$. Essentially one needs to choose the associated orthogonal polynomials so that eq. (39) is valid, where

$$u_k = E[p_k(x_i)\ q_k(x_j)] \tag{40}$$

corresponds to the maximum correlation between members of the complete sets $p_k(\chi_i)$ and $q_\lambda(\chi_j)$, $k,\lambda = 0,1,\ldots$ Then, by a straightforward application of the theory of diagonalization of bilinear forms (Gel'fand, 1961), the bivariate density can be expressed as

$$g(\chi_i,\chi_j) = g_i(\chi_i)\, g_j(\chi_j) \sum_{k=0}^{\infty} u_k\, p_k(\chi_i)\, q_k(\chi_j) \tag{41}$$

The corresponding random function will be called a factorable random function. In this case it is easily seen that

$$u_o = 1, \qquad \text{and} \tag{42}$$

$$u_1 = c_x(t_i,\ t_j)\ /\ \{\mathrm{Var}[x_i]\}^{1/2}\ \{\mathrm{Var}[x_j]\}^{1/2} \tag{43}$$

(see also, eqs. (13) and (16)).

In engineering applications it is many times convenient to work in the frequency domain. The characteristic function, that forms a Fourier transform pair with the bivariate density of eq. (41), will then take the form

$$\phi(w_i,w_j) = \sum_{k=0}^{\infty} u_k\Big[\sum_{\lambda,\nu=0}^{k} \frac{c_{\lambda,\nu}}{i^{\lambda+\nu}} \frac{d^{\lambda}\phi_i(w_i)}{dw_i^{\lambda}} \frac{d^{\nu}\phi_j(w_j)}{dw_j^{\nu}} \Big] \tag{44}$$

for suitable coefficients $c_{\lambda,\nu}$. (The proof is given in the Appendix A.)

Some Criteria of Factorability

Let us now discuss some criteria of factorability. It can be proven (Appendix B) that a necessary and sufficient condition for a random function x_t, to be factorable in the sense defined above is that

$$E[x_i^m/\chi_j] = w_m(\chi_j) \tag{45}$$

$$E[x_j^m/\chi_i] = w_m(\chi_i) \tag{46}$$

where $w_m(\cdot)$ is a polynomial of degree m. It is important to note that this condition does not require the knowledge of the associated orthogonal polynomials.

Example 5: Consider the bivariate density

$$g(\chi_i,\chi_j) = \frac{\alpha}{\pi(1-\rho)^{2\alpha - 1/2}}[(1-\rho^2) - (\chi_i^2 + \chi_j^2) + 2\rho\chi_i\chi_j]^{\alpha-1} \tag{47}$$

where $\alpha>o$ and $|\rho|<1$, defined in the region $\chi_i^2 + \chi_j^2 - 2\rho\chi_i\chi_j < 1 - \rho^2$. We want to test if equ. (47) is of the factorable form. First, by integration one finds

$$g(\chi) = \frac{\Gamma(\alpha+1)}{\sqrt{\pi}\,\Gamma(\alpha + \frac{1}{2})}(1-\chi^2)^{\alpha - 1/2}, \quad \alpha>0,\ -1< \chi <1 \tag{48}$$

Then one needs to check if Eqs. (45) and (46) are valid. Indeed

$$E[x_i^m/\chi_j] = \int_{S(\chi_i)} \chi_j^m \frac{g(\chi_i,\chi_j)}{g(\chi_j)}\, d\chi_i$$

where $S(\chi_i) = [\rho\chi_i - \sqrt{(1-\rho^2)(1-\chi_j^2)},\ \rho\chi_i + \sqrt{(1-\rho^2)(1-\chi_j^2)}]$, is obviously a polynomial of degree m in χ_j. Similarly, the $E[x_j^m/\chi_i]$ is a polynomial of degree m in χ_i and, thus, the criterion of factorability is satisfied. One may now proceed to the derivation of the polynomial expansion of eq. (47). For the univariate density (48) one uses the result of eq. (38). The coefficient u_k are determined from eq. (40), i.e.

$$u_k = \left[\frac{k!\,\Gamma(2\alpha)}{\Gamma(2\alpha+k)}\right] C_k^{(\alpha)}(\rho) \tag{49}$$

where, k= 0,1,2, ..., and ρ is the correlation coefficient between x_i and x_j. (When k= 1, $u_1 = \rho$.) Consequently the factorable expansion of Eq. (47) has as follows (α>o)

$$g(\chi_i,\chi_j) = \frac{[\Gamma(\alpha+1)]^2}{\pi[\Gamma(\alpha+\frac{1}{2})]^2} [(1-\chi_i)(1-\chi_j)]^{\alpha-\frac{1}{2}} \sum_{k=0}^{\infty} \frac{(k+\alpha)(k!)^2[\Gamma(2\alpha)]^2}{\alpha[\Gamma(2\alpha+k)]^2}$$

$$C_k^{(\alpha)}(\rho)\, C_k^{(\alpha)}(\chi_i)\, C_k^{(\alpha)}(\chi_j). \tag{50}$$

Another criterion of factorability may be established in the frequency domain: A necessary and sufficient condition for a random function x_t to be factorable is that

$$\left.\frac{\partial^m \phi(w_i,\ w_j)}{\partial w_1^m}\right|_{w_i=0} = \sum_{\lambda=0}^{m} \frac{\alpha_\lambda}{i^\lambda} \frac{\partial^\lambda \phi(0,w_j)}{\partial w_j^\lambda} \tag{51}$$

$$\left.\frac{\partial^m \phi(w_i,\ w_j)}{\partial w_j^m}\right|_{w_j=0} = \sum_{\lambda=0}^{m} \frac{\beta_\lambda}{i^\lambda} \frac{\partial^\lambda \phi(w_i,0)}{\partial w_i^\lambda} \tag{52}$$

for all m; a_λ, β_λ are suitable polynomial coefficients. (The proof is given in the Appendix C.)

Further manipulations on eq. (B2) and (B3) of Appendix B lead to the following useful expressions

$$E[p_k(x_i)/\chi_j] = \int p_k(\chi_i) \sum_{\lambda=0}^{\infty} u_\lambda \, p_\lambda(\chi_i) \, q_\lambda(\chi_j) \, g(\chi_i) \, d\chi;$$

$$= u_k \, q_k(\chi_j) \tag{53}$$

and, similarly,

$$E[q_\lambda(x_j)/\chi_i] = u_\lambda p_\lambda(\chi_i). \tag{54}$$

Nonlinear Transformations

For the random function z_t defined by eq. (17) above, if $f(.,.)$ is a strictly monotonic function and x_t is a factorable random function, z_t is factorable too; this is easily seen since in this case it holds

$$g(\zeta_i,\zeta_j) = g(\zeta_i) \; g(\zeta_j) \sum_{k=0}^{\infty} u_k \; p_k(f^{-1}(\zeta_i)) \; q_k(f^{-1}(\zeta_j)) \tag{55}$$

(assuming that the denominator is non-zero). The formula relating the corresponding covariances will be as follows (see, also, eq.(25)

$$c_z(\tau_{ij} = t_i - t_j) = \sum_{k=0}^{\infty} u_k^2 \, f_k \tag{56}$$

Example 6: Let

$$z_t = f(x_t) = \mathrm{sgn}[x_t] = \begin{cases} 1 & \text{if } x_t > 0 \\ -1 & \text{if } x_t < 0 \end{cases} \tag{57}$$

where x_t is a factorable (Gaussian) function with unit variance. According to the results above one finds

$$f(\chi_i) = \sum_{k=0}^{\infty} f_{2k+1,i} \, H_{2k+1}(\chi_i)$$

since the even terms vanish;

$$f_{2k+1,i} = \sqrt{\frac{2}{\pi}} \, \frac{(-1)^k}{2k!(2k+1)}$$

and $H_{2k+1}(\cdot)$ are the Hermite polynomials. In this case, $u_k = \rho^k$(ρ is the correlation coefficient) and

$$c_z(\tau = t_i - t_j) = \frac{2}{\pi} \sum_{k=0}^{\infty} \frac{(2k!)}{2^{2k}(k!)^2 \, (2k+1)} \rho^{2k+1} \tag{58}$$

Using Gradstheyn and Ryzhik (1965), one finds that the right-hand side of the last equation is equal to $\frac{2}{\pi} \sin^{-1}\rho$, and thus,

$$c_z(\tau) = \frac{2}{\pi} \sin^{-1} c_x(\tau) \tag{59}$$

Expressions like the eq. (58) are very tractable. For instance, by taking the Fourier transform of eq. (58) one easily finds that

$$C_z(w) = \frac{2}{\pi} \sum_{k=0}^{\infty} \frac{(2k!)}{2^{2k}(k!)^2 \, (2k+1)} [C_x^*(w)]^{2k+1} \tag{60}$$

where, $C_z(w)$, $C_x(w)$ are the spectral functions of the covariances

$c_z(\tau)$ and $c_x(\tau)$ respectively; the $[C_x^*(w)]^{2k+1}$ means the (2k+1)th self-convolution of $C_x(w)$.

THE SYMMETRIC MODEL

This section examines that subclass of factorable processes for which the bivariate density is symmetric in the variables x_i and x_j, i.e.

$$g(\chi_i,\chi_j) = g(\chi_j,\chi_i) \tag{61}$$

Then Eq. (41) becomes

$$g(\chi_i,\chi_j) = g(\chi_i)\; g(\chi_j) \sum_{k=0}^{\infty} u_k\, p_k(\chi_i)\; p_k(\chi_j) \tag{62}$$

where

$$u_k = E[p_k(x_i)\; p_k(x_j)] \tag{63}$$

The hypothesis of stationarity implies that the univariate densities and the polynomials in eq. (62) are independent of t_i, t_j. Also, the coefficients u_k depend only on the difference $\tau_{ij} = t_i - t_j$. In this case, eq. (62) is identical with the well known isofactorial model of Geostatistics (Matheron, 1976; Armstrong and Matheron, 1986). Clearly, a necessary and sufficient condition for a stationary function to have a bivariate density of the isofactorial form (62) is that its characteristic function has a similar factorable form.

Below, some interesting properties of the model (62) will be discussed:

An Integral Representation of the Isofactorial Model

The theory of integral equations enables one to consider a variety of different problems from a unified point of view (Courant and Hilbert, 1953). In effect, Eq. (62) may be written as

$$\int g(\chi_i,\chi_j)\ p_k(\chi_j)\ d\chi_j = g(\chi_i) \sum_{k=0}^{\infty} u_k\ p_k(\chi_i) \int p_k(\chi_j)\ g(\chi_j)\ d\chi_j$$
$$= u_k\ g(\chi_i)\ p_k(\chi_i) \tag{64}$$

or by setting

$$F(\chi_i,\chi_j) = g(\chi_i,\chi_j)\ [g(\chi_i)\ g(\chi_j)]^{-1/2} \tag{65}$$

$$\theta_k(\chi_i) = g(\chi_i)^{1/2}\, p_k(\chi_i) \tag{66}$$

$$\theta_k(\chi_j) = g(\chi_j)^{1/2}\, p_k(\chi_j) \tag{67}$$

and

$$\gamma_k = u_k \tag{68}$$

one obtains the well known linear homogeneous integral equation

$$\int F(\chi_i,\chi_j)\ \theta_k(\chi_j)\ d\chi_j = \gamma_k\ \theta_k(\chi_i) \tag{69}$$

Here, $F(\chi_i,\chi_j)$ is the kernel of the integral equation, the eigenfunctions $\theta_k(\cdot)$ are orthogonal and normal, and γ_k are the corresponding eigenvalues. Clearly, the problem of deriving a symmetric, factorable bivariate density expansion is now a problem of the eigenvalue theory, where $F(\chi_i,\chi_j)$ is given by

$$F(\chi_i,\chi_j) = \Theta_i^T \Gamma \Theta_j \tag{70}$$

where

$$\Theta_i = \begin{bmatrix} \theta_1(\chi_i) \\ \vdots \\ \theta_k(\chi_i) \end{bmatrix}, \quad \Theta_j = \begin{bmatrix} \theta_1(\chi_j) \\ \vdots \\ \theta_k(\chi_j) \end{bmatrix}, \quad \text{and} \quad \Gamma = \begin{bmatrix} \gamma_1 & 0 & \cdots & 0 \\ 0 & & & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & \gamma_k \end{bmatrix}$$

Eq. (70), combined with eq. (65), yields the bilinear form of the bivariate density. However, expansion (70) holds under certain conditions only: the eigenfunctions $\theta_k(\cdot)$ must form a complete set and they should not depend on $\tau_{ij}= t_i-t_j$. Furthermore, by comparing Eqs. (4), (5) and (65) one finds

$$\begin{aligned} r &= \int \int F^2(\chi_i,\ \chi_j)\ d\chi_i\ d\chi_j \\ &= \sum_{k=1}^{\infty} \gamma_k^2 \end{aligned} \tag{71}$$

which must be finite.

Example 7: Let $g(\chi) = \chi^c \exp[-\chi]$ if $\chi>0$, $= 0$ otherwise. For $c= n - 1/2$, $n=0,1,2,\ldots$, this is the density of the chi-square distribution. The corresponding polynomials which are orthogonal and normal for the interval $(0,\infty)$ are (see, eg, Watson, 1933)

$$p_k(\chi) = \frac{\sqrt{k!}}{\sqrt{\Gamma(k+c+1)}}\ L_k^{(c)}(\chi) \tag{72}$$

where

$$L_k^{(c)}(\chi) = \frac{1}{\chi^c \exp[-\chi]} \frac{d^k}{d\chi^k} \{\chi^c \exp[-\chi]\} \tag{73}$$

are the generalized Laguerre polynomials, and

$$\gamma_k = \rho^k$$

are the eigenvalues (ρ is the correlation coefficient). Eqs. (66) and (67) write

$$\theta_k(\chi_i) = \{\frac{k!g(\chi_i)}{\Gamma(k+c+1)}\}^{1/2} L_k^{(c)}(\chi_i) \qquad \text{and}$$

$$\theta_k(\chi_j) = \{\frac{k!\; g(\chi_j)}{\Gamma(k+c+1)}\}^{1/2} \; L_k^{(c)}(\chi_j),$$ respectively. By making use of eqs. (65) and (70) one finds

$$g(\chi_i,\chi_j)=(\chi_i\chi_j)^c \exp[-(\chi_i+\chi_j)] \sum_{k=0}^{\infty} \rho^k \frac{k!}{\Gamma(k+c+1)} \; L_k^{(c)}(\chi_i) \; L_k^{(c)}(\chi_j) \tag{74}$$

which, for $c= n-1/2$, $n=0,1,2,...$, is the factorable form of the bivariate chi-square density.

On-Line Stationary Markov Functions

Let x_t be an on-line stationary Markov function; then (Melsa and Sage, 1972)

$$g(\chi_i, \chi_m, \chi_j) = \frac{g(\chi_i, \chi_m)\ g(\chi_m,\chi_j)}{g(\chi_m)}$$

Since x_t is a factorable function one may apply eq. (39) in the right-hand side of the above equation to find

$$g(\chi_i, \chi_m,\chi_j) = g_i(\chi_i)\ g_j(\chi_j)\ g_m(\chi_m) \sum_{\lambda=0}^{\infty} \sum_{k=0}^{\infty} u_\lambda\ u_k\ q_\lambda(\chi_i)$$

$$q_\lambda(\chi_m)\ p_k(\chi_m)\ p_k(\chi_j)$$

where $u_\lambda = u_\lambda(\tau_{mi} = t_m - t_i)$ and $u_k = u_k(\tau_{jm} = t_j - t_m)$ due to stationarity. Furthermore, by integrating the above equation with respect to χ_m one finds

$$g(\chi_i,\chi_j) = g_i(\chi_i)\ g_i(\chi_j) \sum_{k=0}^{\infty} u_k(\tau_{jm})\ u_k(\tau_{mi})\ q_k(\chi_i)\ p_k(\chi_j)$$

Comparing the last equation with Eq. (39) where $u_k = u_k(\tau_{ji} = t_j - t_i)$ one obtains the equation

$$u_k(\tau_{ji}) = u_k(\tau_{jm})\ u_k(\tau_{mi})$$

The solution of this equation is

$$u_k(\tau) = \exp[-\alpha_k \tau] \tag{75}$$

where $0<\alpha_1<\alpha_2<\dots$. Therefore, an on-line stationary Markov function with an isofactorial bivariate density has exponential correlation functions of the form (75).

Furthermore, the Markov property implies

$$E[x_t/X_{t-1}] = E[x_t/x_{t-1}] \tag{76}$$

where $X_{t-1} = \{x_s : 0 < s < t-1\}$. A straightforward application of eq. (39) yields

$E[x_t/x_{t-1}] = bx_{t-1}$, or due to Eq. (76),

$$E[x_t/X_{t-1}] = bx_{t-1} \tag{77}$$

where $b = u_1(t, t-1)$ is a numerical coefficient. Eq. (77) demonstrates that the function x_t is a martingale-type function. Properties (75) and (77) are of significant importance in nonlinear estimation.

Frequency Domain

The expression (44) for the characteristic function holds for the symmetric model, but the $\phi_i(\cdot)$ and $\phi_j(\cdot)$ have the same functional form $\phi(\cdot)$ in this case. Moreover, Armstrong and Matheron (1986) have shown that for a stationary random function one finds

$$\phi(w_i)^{1-\rho} \frac{\partial^n}{\partial w_i^n} [\phi(w_i)]^\rho = \sum_{k=0}^{n} A_{n,k} \frac{\partial^k}{\partial w_i^k} \phi(w_i) \tag{78}$$

for suitable correlation factor and coefficients $A_{n,k}$.

APPENDIX A

The characteristic function of the bivariate density $g(\chi_i, \chi_j)$ is

$$\phi(w_i, w_j) = \int \int g(\chi_i, \chi_j) \exp [i(w_i \chi_i + w_j \chi_j)] \, d\chi_i \, d\chi_j \tag{A1}$$

or, by inserting eq. (41),

$$\phi(w_i,w_j) = \sum_{k=0}^{\infty} u_k E[p_k(x_i) \, e^{iw_i x_i}] E[q_k(x_j) \, e^{iw_j x_j}] \tag{A2}$$

Let $$p_k(\chi_i) = \sum_{\lambda=0}^{k} \alpha_\lambda \chi_i^\lambda \,, \tag{A3}$$

$$q_k(\chi_j) = \sum_{\nu=0}^{k} \beta_\nu \chi_j^\nu \tag{A4}$$

for proper coefficients $\alpha_\lambda, \beta_\lambda$. Taking into account the Eqs. (A3) and (A4), Eq. (A2) writes

$$\phi(w_i,w_j) = \sum_{k=0}^{\infty} u_k \int \sum_{\lambda=0}^{k} \alpha_\lambda \, \chi_i^\lambda \, g_i(\chi_i) \exp[iw_i\chi_i] \, d\chi_i \int \sum_{\nu=0}^{k} \beta_\nu$$

$$\chi_j^\nu \, g_j(\chi_j) \exp[iw_j x_j] \, d\chi_j \tag{A5}$$

The below Fourier transform pairs hold

$$\{(i\chi_i)^\lambda g_i(\chi_i), \; \frac{d^\lambda \phi_i(w_i)}{dw_i^\lambda} = i^\lambda \; E[x_i^\lambda \; e^{iw_i x_i}]\}, \text{ and} \tag{A6}$$

$$\{(i\chi_j)^\nu g_i(\chi_j), \frac{d^\nu \phi_j(w_j)}{dw_j^\nu} = i^\nu \; E[x_j^\nu \; e^{iw_j x_j}]\} \tag{A7}$$

which, combined with Eq. (A5), yield Eq. (44), where $c_{\lambda,\nu} = \alpha_\lambda \beta_\nu$.

APPENDIX B

First, by assuming that x_t is factorable one finds that

$$E[x_i^m/\chi_j] = \int \chi_i^m \frac{g(\chi_i,\chi_j)}{g_j(\chi_j)} \, d\chi_i$$

$$= \sum_{k=0}^{\infty} u_k \, p_k(\chi_j) \int \chi_i^m \, q_k(\chi_i) \, g_i(\chi_i) \, d\chi_i \tag{B1}$$

and since, due to orthonormality properties the term with $k \neq m$ vanish, the right-hand side of the above equation is a polynomial of degree m. Similarly, the conditional movement $E[x_j^m/\chi_i]$ is a polynomial of degree m. Thus, Eq. (45) results; Eq. (46) is obtained in exactly the same way.

Conversely, assume that eqs. (45) and (46) are valid. They imply respectively,

$$E[p_k(x_i)/\chi_j] = \int p_k(\chi_i) \frac{g(\chi_i,\chi_j)}{g_j(\chi_j)} \, d\chi_i = w_k(\chi_j) \tag{B2}$$

$$E[q_\lambda(x_j)/\chi_i] = \int q_\lambda(\chi_j) \frac{g(\chi_i,\chi_j)}{g_i(\chi_i)} \, d\chi_j = v_\lambda(\chi_i) \tag{B3}$$

where $w_k(\cdot)$ and $v_\lambda(\cdot)$ are polynomials of degree k and λ respectively. The coefficients $u_{k\lambda}$ of the general expansion (6) become

$$u_{k\lambda} = \iint p_k(\chi_i) \, q_\lambda(\chi_j) \, g(\chi_i,\chi_j) \, d\chi_i \, d\chi_j$$

$$= \int q_\lambda(\chi_j)[\int p_k(\chi_i) \, g(\chi_i,\chi_j) \, d\chi_i] \, d\chi_j$$

Now, by using eq. (B2) one finds

$$u_{k\lambda} = \int q_\lambda(\chi_j)\, w_k(\chi_j)\, g_j(\chi_j)\, d\chi_j$$
$$= 0$$

for all $\lambda > k$, since the orthogonal polynomial $q_\lambda(\cdot)$ is orthogonal to any polynomial $w_k(\cdot)$ of degree $k < \lambda$. Similarly, using eq. (B3) one gets

$$u_{k\lambda} = \int p_k(\chi_i)\, v_\lambda(\chi_i)\, g_i(\chi_i)\, d\chi_i$$
$$= 0$$

for all $k > \lambda$. Therefore, since $u_{k\lambda} = 0$ for all $k \neq \lambda$ by the definition (39) the bivariate density will have the factorable form of Eq. (41). Concluding, eqs. (45) and (46) constitute a necessary and sufficient condition for a random function x_t to be factorable. (Another proof for the special case of isofactorial bivariate densities is given in Armstrong and Matheron, 1986.)

APPENDIX C

It will be proven that Eqs. (51) and (52) are necessary and sufficient conditions for the validity of Eq. (45) and (46) respectively (and, consequently, they constitute themselves a necessary and sufficient criterion of factorability). To prove the necessity of Eq. (51), assume that Eq. (45) holds, multiply both sides of Eq. (45) by $\exp[iw_j\chi_j]$ and take the mean value, i.e.

$$E\{e^{iw_jx_j}\; E[x_i^m/\chi_j]\} = E[e^{iw_jx_j}\; w_m(x_j)], \qquad \text{or}$$

$$E[e^{iw_jx_j}\; x_i^m] = \sum_{\lambda=0}^{m} \alpha_\lambda E[e^{iw_jx_j}\; x_j^\lambda] \tag{C1}$$

$$E[e^{iw_jx_j}\, x_i^m] = \sum_{\lambda=0}^{m} \alpha_\lambda E[e^{iw_jx_j}\, x_j^\lambda] \tag{C1}$$

where α_λ are the coefficients of the polynomial $w_m(\chi_j)$. But it holds

$$E[e^{iw_jx_j}\, x_i^m\,] = i^{-m} \left.\frac{\partial^m \phi(w_i, w_j)}{\partial w_i^{\,m}}\right|_{w_i=0} \quad \text{and} \tag{C2}$$

$$E[e^{iw_jx_j}\, x_j^\lambda] = i^{-\lambda} \frac{\partial^\lambda \phi(0,\, w_j)}{\partial w_j^{\,\lambda}} \tag{C3}$$

and Eq. (51) follows. To prove the sufficiency of the conditon, assume that Eq. (51) is true for all m, fact that implies (Eqs. (C2), (C3))

$$E\{e^{iw_jx_j}\, [x_i^m - \sum_{\lambda=0}^{m} \alpha_\lambda x_j^\lambda\,]\} = 0 \text{ , or}$$

$$E\{e^{iw_jx_j}\, E[x_i^m - \sum_{\lambda=0}^{m} \alpha_\lambda\, x_j^\lambda/\chi_j]\} = 0. \tag{C4}$$

But Eq. (C4) is true only if

$$E[x_i^m - \sum_{\lambda=0}^{m} \alpha_\lambda x_j^\lambda/\chi_j] = 0 \tag{C5}$$

Then, Eq. (C5) yields

$$E[x_i^m/\chi_j] = \sum_{\lambda=0}^{m} \alpha_\lambda\, \chi_j^\lambda = w_m(\chi_j) \text{ , i.e. Eq. (45).}$$

In exactly the same way one may prove that Eq. (52) is a necessary and sufficient condition for the validity of Eq. (46) and this completes the proof.

APPENDIX D : Some more examples of factorable processes

Example: The bivariate density

$$g(\chi_i, \chi_j) = g(\chi_i)\, g(\chi_j) \frac{\exp[-\rho^2(\chi_i^2+\chi_j^2)/2(1-\rho^2)]}{\sqrt{1-\rho^2}} \cosh \frac{\rho\chi_i\chi_i}{1-\rho^2} \qquad \text{(D1)}$$

where $g(\chi_i)=\frac{1}{\sqrt{2\pi}} \exp[-\chi_i^2/2]$, $g(\chi_j) = \frac{1}{\sqrt{2\pi}} \exp[-\chi_j^2/2], -\infty<\chi_i, \chi_j<\infty$, and ρ is the correlation coefficient, $|\rho|<1$, may be used as a symmetric (isofactorial) model, for Eq(D1) writes

$$g(\chi_i, \chi_j) = g(\chi_i)\, g(\chi_j) \sum_{k=0}^{\infty} \rho^{2k} H_{2k}(\chi_i)\, H_{2k}(\chi_j) \qquad \text{(D2)}$$

where $H_{2k}(\cdot)$ are Hermite polynomials of even order.

By comparing the model (D2) with the well known bivariate Gaussian model of nonlinear geostatistics

$$g(\chi_i, \chi_j) = g(\chi_i)\, g(\chi_j) \sum_{k=0}^{\infty} \rho^{k} H_{k}(\chi_i)\, H_{k}(\chi_j) \qquad \text{(D3)}$$

one finds that

Eq. (D1) $= \frac{1}{2}$ [Eq. (D3) with correlation coefficient ρ] +
$\frac{1}{2}$ [Eq. (D3) with correlation coefficient $-\rho$].

Example: Consider the bivariate density

$$g(\chi_i, \chi_j) = \frac{\exp[-(\chi_i+\chi_j)/(1-\mu^2)]}{1-\mu^2}\, I_0\left(2\, \frac{\mu\sqrt{\chi_i\chi_j}}{1-\mu^2}\right) \qquad \text{(D4)}$$

where $0 <\chi_i, \chi_j < \infty$. Eq. (D4) may be expanded as follows (see

Gradshteyn and Ryzhik, 1965)

$$g(\chi_i, \chi_j) = g(\chi_i)\, g(\chi_j) \sum_{k=0}^{\infty} \mu^{2k} L_k(\chi_i)\, L_k(\chi_j) \qquad \text{(D5)}$$

where, $g(-\chi_i) = \exp[-\chi_i]$, $g(\chi_j) = \exp[-\chi_j]$ and $L_k(.)$ are the Laguerre polynomials which are orthonormal with respect to $g(.)$. The correlation coefficient is, $\rho = \mu^2 < 1$.

REFERENCES

Armstrong, M. and G. Matheron, 1986, "Disjunctive Kriging Revisited: Part I", Mathematical Geology, v.18, n.8, p.711-728.

Courant, R. and Hilbert, D., 1953, "Methods of Mathematical Physics: Vol.1", J.Wiley and Sons Publ., New York.

Gel'fand, I.M., 1961, "Lectures on Linear Algebra", Interscience Publ. Inc., New York.

Lavrentev, M.A. and Schabat, B.W., 1967, "Methoden der Komplexen Functionen Theorie", Deutsches Verlag der Wissenschaften, Berlin.

Matheron, G., 1976, "A Simple Substitute for Conditional Expectation: the Disjunctive Kriging", in Adv. Geostatistics in Mining Industry (Guarascio, M., David, M., and Huijbregts, Ch.J., eds.), Reidel, Boston.

Watson, G.N., 1933, "Notes on Generating Functions of Polynomials: (1) Laguerre Polynomials, (2) Hermite Polynomials, J. London Math. Soc., 8, p.189-199.