USC-SIPI REPORT #106

On ESD's In System Identification

by

Ananthram Swami and Jerry M. Mendel

May 1987

Signal and Image Processing Institute
UNIVERSITY OF SOUTHERN CALIFORNIA
Viterbi School of Engineering
Department of Electrical Engineering-Systems
3740 McClintock Avenue, Suite 400
Los Angeles, CA 90089-2564 U.S.A.

On ESD's in System Identification

Ananthram Swami and Jerry M. Mendel

Signal and Image Processing Institute

Department of Electrical Engineering-Systems

University of Southern California

Los Angeles, CA 90089-0781

USCSIPI 106. May 1, 1987

Abstract

In this note, we review several important properties of *Elliptically Symmetric Distributions (ESD's)*. ESD's are generalizations of the Gaussian distribution; they can be represented as scale mixtures of the Gaussian, are closed under linear transformations and have conditional expectations that are linear on the conditioning variable. Several results in estimation theory, such as parameter estimation, Kalman filtering theory etc. hold for ESD's. In several cases, the null distributions of certain invariant test statistics remain unchanged from the Gaussian case. ESD's are also Bussgang processes. Finally, we show that the zeroes of an ARMA system cannot be resolved in two cases: a) when the system is excited by an unobservable ESD process and b) when the output process is Bussgang.

Introduction

Gaussian distributions hold a central place in linear estimation and detection theory. This pre-eminence is largely attributed to three important properties of the Gaussian: closure under linear transformations, linearity of conditional expectations and the Central Limit Theorem. Vershik [1], showed that these two properties characterize a large class of distributions --- the ESD's, of which the Gaussian is a special case. Vershik showed that distributions in this class have characteristic functions of the form $\phi(u)=g(u^t Cu)$; and that linearity of conditional expectations is a necessary and sufficient condition for membership in this class. Blake and Thomas [2] showed that ESD's must have density functions of the form $p(x) = f(x^t C^{-1} x)$, where C is positive definite. The uncorrelated-implies-independence property of the Gaussian translates into uncorrelated implies semi-independence [i.e., $E(xg(y))=E(g(y))E(x \mid y)$] for the ESD. Furthermore, wide-sense stationarity implies strict-sense stationarity for ESD's (a result to be expected since ESD's are second-order processes). ESD's thus represent a generalization of the Gaussian distribution. This generalization preserves the structure of the Gaussian (i.e., the quadratic nature), while relaxing the specific form (i.e., the exponential; although this specific form usually leads to tractable results).

Bussgang [3] proved an important property for Gaussian processes: let x(t) be stationary, and Gaussian, and, let Z denote an arbitrary zero-memory nonlinearity (ZMNL); then,

$$E\{ x(k) Z(x(k-j)) \} = a\ E\{ x(k) x(k-j) \} \tag{1}$$

where a is a constant independent of k and j. The term *Bussgang process* is applied to any process $x(t)$ that satisfies (1). Furthermore, we can represent $Z(x(t))$ as $a\ x(t) + d(t)$, where $d(t)$ is uncorrelated with $x(t)$. Gray [4] and Godfrey and Rocca [5] proved that Minimum Entropy Deconvolution (MED) schemes converge to Bussgang processes. (They assume that the system input is modeled as an i.i.d. non-Gaussian process such as the sub-Gaussian or a two-component Gaussian mixture; see also [7-10]).

McGraw and Wagner [6] showed that ESD's satisfy a condition known as Nutall separability, i.e.,

$$\frac{\partial \phi_2(u, v; \tau)}{\partial u} \Big|_{u=0} = \rho(\tau) \frac{d\, \phi_1(v)}{dv} \tag{2}$$

where ϕ_2 is the joint characteristic function of $x(t)$ and $x(t+\tau)$, and, ϕ_1 is the characteristic function of $x(t)$. Nutall separability is a necessary and sufficient condition for a process to be Bussgang. They also proved an analog of Price's theorem, useful in computing the moments of the output of a ZMNL.

Bussgang processes have also been studied by Barrett and Lampard [11]. They sought expansions of the joint p.d.f. $p(x_1, x_2)$ of the form

$$p(x_1, x_2) = p_1(x_1)\, p_2(x_2) \sum_{m=0, n=0}^{\infty} a_{mn}\ \theta_m^1(x_1)\, \theta_n^2(x_2) \tag{3}$$

where,

$$\int p_i(x_i)\, \theta_n^i(x_i) dx_i = \delta_{mn} \quad i=1,2 \tag{4}$$

in which θ_n^i are orthonormal polynomial functions (with respect to the univariate p.d.f.'s); in particular they showed that the class of distributions with diagonal representations (i.e., $a_{mn}=0,\ m \neq n$ in (4)) are Bussgang and that conditional expectations are linear, i.e., $E\ \{x(t+\tau) \mid x(t)\} = r(\tau) x(t)$. If $x(t)$ is a Markov process belonging to this class, then its autocorrelation must be exponential. If $p(x, y)$ is symmetric, what we have is, in effect, the Mercer expansion. Brown [12] showed that processes which satisfy

$$a_{m1}(\tau) = d_m\, a_{11}(\tau) \quad m=1,2,... \tag{5}$$

are Bussgang, if the d's are real constants, independent of τ.

Examples of the Bussgang include any i.i.d. process, colored Gaussian processes (the expansion in (3) involves Hermite polynomials); exponential (involves Laguerre polynomials; e.g., the square of the envelope of a narrow bandpass filter excited by white Gaussian noise); sine waves of constant amplitude (Chebyshev polynomials); and, more generally, any ESD process.

Properties of ESD's

The material in this section is largely based on Chu [13] and uses Chu's notation; proofs, where omitted, may be found in Chu.

Let x denote a real-valued, n-component ESD random vector.

(0) **Density function** : the p.d.f. p(x) and the characteristic function $\phi(v)$ are given by

$$p(x) = f(x^t C^{-1} x),\ C > 0 \tag{6}$$

and,

$$\phi(v) = h(v^t C v) \tag{7}$$

(1) **Gaussian representation** : ESD's can be represented as scale mixtures of the Gaussian, i.e., as a weighted average of Gaussians, each with the same correlation matrix C, but with different scales. In some cases, the weight function may be non-negative, and hence may be interpreted as the p.d.f. of the scale factor; such a process is called compound Gaussian. Any symmetric marginal distribution may be used to construct an ESD.

$$p(x) = \int_0^\infty w(t)\, N_x(C/t)\, dt \tag{8}$$

where $N_x(C/t)$ represents the zero-mean Gaussian density with covariance matrix C/t and the weighting function, $w(t)$, is given by

$$w(t) = (2\pi)^{n/2} \mid C \mid^{1/2} t^{-n/2} L^{-1}(f(s)) \tag{9}$$

where s is a scalar, and, L^{-1} denotes the inverse Laplace transform. *ESD's are thus characterized by the correlation matrix C and the weighting function w. $E(w,C)$ will denote such a distribution.*

(2) **Averages** : If g(x) is Borel measurable, then,

$$E\ \{g(x)\} = \int_0^\infty w(t)\, E\{g(x) \mid N_x(C/t)\}\, dt \tag{10}$$

where $E\{y \mid N(C/t)\}$ denotes E{y}, computed as if y had the zero-mean Gaussian distribution with covariance C/t. It follows, at once, that C is a scalar multiple of the covariance matrix of x.

(3) **Closure under Linear Transformations** : if x is ESD(w,C), then, $y = Ax$ is ESD(w,ACA^t).

(4) **Marginals** of ESD's are also ESD and have the same form and hence the same weighting function $w(t)$.

(5) **Linearity of Conditional Expectations** : if $x = col(x_1, x_2)$ then,

$$E(x_2 \mid x_1) = C_{21} C_{11}\, x_1 \tag{11}$$

where C_{21} and C_{11} are appropriate sub-matrices of C.

(6) **Conditional covariance** has the same form as in the Gaussian case, except for a scalar multiplier that depends on the conditioning variable.

$$E(x_2 x_2^t \mid x_1) = \frac{\int_s^\infty f_1(s_1) ds_1}{f_1(s)} [C_{22} - C_{21} C_{11}^{-1} C_{12}] \tag{12}$$

where $s = x_1^t C_{11} x_1$.

(7) **Bussgang property** : x(t) is Bussgang (McGraw and Wagner [6]).

(8) **i.i.d. process** : if x is ESD and has independent components, then x is necessarily Gaussian; hence, if x(t) is i.i.d. and non-Gaussian, then, it cannot be ESD.

(9) **Central Limit Theorem (CLT)** : property (3) implies that the CLT does not apply to ESD's; i.e., if x_i are ESD(w,C), then $x = \sum x_i$ is also ESD(w, C_x) (Picinbono [14]).

(10) **Ergodicity** : Vershik [1] seems to prove that a stationary ergodic ESD process is necessarily Gaussian.

Applications : Well-known results in Gaussian linear estimation/detection theory are readily applicable to ESD's. For example, Chu shows that if the plant and observation noises and the initial state vector are jointly ESD, then the classical Kalman filter theory holds (except that the state covariance matrix, useful in performance analysis, needs to be computed separately). Applications in the area of estimation/detection are also discussed by Picinbono [14], Yao [15], Gualtieroti [16], Goldman [17] and Masreliez and Martin [18]. Chmielwski [19] gives an excellent review of applications in other areas as well as an extensive bibliography. Applications are also discussed by Devlin et al [22]. A very readable account of ESD theory may be found in Muirhead [20].

ARMA identifiability : In this section we show, via a counter-example, that non-Gaussianity of the system input is insufficient to resolve the zeroes of an ARMA system, using only the observed output.

Theorem 1 : The zeros of an ARMA system excited by an ESD process, cannot be resolved using the output alone.

Proof : If the input x(t) to an ARMA system is ESD, then the output y(t) is also ESD. The p.d.f. of y(.) is completely characterized by its correlation matrix C and the weighting function w(t). Applying an all-pass filter does not change C (since it is a correlation matrix!) or w(t) (via property 3); consequently, the p.d.f.

of y(t) is insensitive to phase. Thus, as in the Gaussian case, the true zeroes of an ARMA system, excited by an ESD input, cannot be resolved using the output alone.□

We note that our result is much stronger than that in Rocca and Kostov [21] who prove this result for a particular gradient-based method.

All joint moments and cumulants of an ESD process can be described in terms of its correlation function and w(t) (recall that x(t) is a second-order process whose p.d.f. is completely specified by C and w(t)). For example, Muirhead [20] shows that

$$E\{x(t)x(t+u)x(t+v)x(t+w)\} = K\,[R(u)R(v-w)+R(v)R(w-u)+R(w)R(u-v)] \tag{13}$$

where $K = E\{x^4\}/3E^2\{x^2\}$, and is given by,

$$K = \frac{\int_0^\infty w(t)/t^2\,dt}{[\int_0^\infty w(t)/t\,\,dt]^2} \tag{14}$$

In particular, for $u=v=w$, we have, $E\{x(t)x^3(t+u)\} = 3KR(0)R(u)$. Since an ESD process is Bussgang, we have the more general result- $E\{x(t)x^n(t+u)\} = K_n R(0)R(u)$, where, K_n may be obtained via Eqn (10).

The k-th order *cumulant*, $cum_k^y\{y(t_1), y(t_2), \cdots y(t_k)\}$, of a process y(t) is defined in terms of its joint moments of orders up to k; definitions and details may be found in Giannakis [23]. For a stationary process, the k-th order cumulant is a function of (k-1) variables. The diagonal slice is obtained by setting $t_2 = t_3 = \cdots = t_k = t_1 + u$. The cumulants of a non-Gaussian process cannot all be zero (for k>2); and, as such, they have been used in system identification (see [23], and references therein). Let $C_k^y(u)$ denote the diagonal slice of the k-th order cumulant; let $M_k^y(u)$ denote $E\{y(t)y^{k-1}(t+u)\}$.

Theorem 2 : The zeros of an ARMA system, whose output $y(t)$ is Bussgang, cannot be resolved using only the diagonal slice of the output cumulants.

Proof $M_m^y(u) = E\{y(t)y^{m-1}(t+u)\}$. Now, $Z(y) = y^{m-1}$, is a ZMNL and $y(t)$ is Bussgang. Hence, from Eqn. (1),

$$M_m^y(u) = g_m \; R(u) \tag{15a}$$

It is easily shown that,

$$C_k^y(\tau) = \sum_{m=1}^{k} \alpha_m M_m^y(\tau) \tag{15b}$$

$$= R(\tau) \sum_{m=1}^{k} \alpha_m g_m \tag{15c}$$

The sum in Eqn. (15c) is easily evaluated, and, yields,

$$C_k^y(\tau) = R(\tau) C_k^y(0)/R(0) \tag{16}$$

Thus, for a Bussgang process, the diagonal slices of its cumulant are proportional to its autocorrelation, and are therefore phase-blind. Consequently, if the output of an ARMA system is Bussgang, the system zeros cannot be resolved using only the diagonal slices of the output cumulant.□

In [23], it is claimed that a "fourth-order whiteness" (i.e., the fourth-order cumulant is a delta function) assumption for the input suffices to permit system identification using only the diagonal slices of the output cumulant. However, it is possible for the input process to be ESD and fourth-order white. In this case, the output being ESD, is also Bussgang; consequently, the true system zeroes cannot be identified from the diagonal slice of the output cumulant alone. We note that the methods presented in [23], are valid when the input is modeled as an i.i.d. non-Gaussian process (and hence not ESD). Further, this theorem does not rule out wavelet recovery using 1-D cumulant slices of the form considered in Giannakis [24].

Conclusions

We have reviewed several important properties of Elliptically distributed processes; we have seen that several results in Gaussian theory, e.g., Kalman filter theory, are applicable to these processes. Two new results have been presented:

1. The zeroes of an ARMA system excited by an ESD process cannot be resolved using the output alone.
2. The zeroes of an ARMA system cannot be resolved using 1-D cumulants of the form C(u,...u), if the observed process is Bussgang.

Acknowledgements : The work reported in this paper was performed at the University of Southern California, Los Angeles, under National Science Foundation Grant ECS-8602531 and Naval Ocean Systems Center Contract N66001-85-D-0203. The authors wish to thank G. Giannakis for stimulating discussions on this and related topics.

References

[1] A.M. Vershik, "Some Characteristic Properties of Gaussian Processes," *Theory of Prob and Appln,* v. 9, 353-6, 1964.

[2] I.F. Blake and Thomas, "On a class of processes arising in linear estimation theory," *IEEE Trans. on Info Theory,* IT-14, 12-16, 1968.

[3] J.J. Bussgang, "Cross-correlation functions of amplitude distorted Gaussian signals," *Tech. Rep. No 216, Res. Lab. Elec., MIT,* March 1952.

[4] W.C. Gray, "Variable Norm Deconvolution,", Ph.D. Thesis, Stanford, 1979.

[5] R. Godfrey and F. Rocca, "Zero-memory Non-linear Deconvolution," *Geophysical Prospecting,* 189-228, 1981.

[6] D.K. McGraw and J.F. Wagner, "Elliptically Symmetric Distributions," *IEEE Trans. on Info Theory,* IT-14, 110-120, 1968.

[7] A.T. Walden, "Non-Gaussian Reflectivity, Entropy and Deconvolution," *Geophysics,* v. 50, 2862-88, 1985.

[8] A.T. Walden and J.W.J. Hosken, "The nature of the non-Gaussianity of primary reflection coefficients and its significance for deconvolution," *Geophysical Prospecting,* 1038-66, 1981.

[9] S. Bellini and F. Rocca, "Blind Deconvolution: Polyspectra or Bussgang processes," in 'Digital Communications', Ed. E.Biglieri and G. Prati, *Elsevier Science,* 1986.

[10] S. Bellini, "Bussgang techniques for blind deconvolution," *Globcom 86, Houston,* 1986.

[11] J.F. Barrett and D.G. Lampard, "An expansion for some second-order probability distributions and its applications to noise problems," *IRE Trans. on Info Theory,* 10-15, March 1955.

[12] J.L. Brown Jr., "On a cross-correlation property for stationary random processes," *IRE Trans on Info Theory,* 28-31, 1957.

[13] K'ai Chang Chu, "Estimation and Decision for Linear Systems with Elliptical Random Processes," *IEEE Trans. on Auto. Control,* AC-18, 499-505, 1973.

[14] B. Picinbono,"Spherically Invariant and compound Gaussian stochastic processes," *IEEE Trans. on Info Theory,* IT-16 , 77-79, 1970.

[15] K. Yao, "A representation theorem and its applications to spherically invariant processes," *IEEE Trans. on Info Theory,* IT-19, 600-8, 1973.

[16] A.F. Gualtieroti, "A likelihood ratio formula for spherically invariant processes," *IEEE Trans. on Info Theory,* IT-22, p 610, 1976.

[17] J. Goldman, "Detection in the presence of spherically symmetric random vectors," *IEEE Trans. on Info Theory,* IT-22, 52-59, 1976.

[18] C.J. Masreliez and R.D. Martin, "Robust Bayesian Estimation for the linear model and robustifying the Kalman filter," *IEEE Trans. on Auto. Control,* AC-22,361-71, 1977.

[19] M.A. Chmielwski, "Elliptically Symmetric Distributions: a review and bibliography," *Intl. Stat. Review,* V 49, 67-74, 1981.

[20] R.J. Muirhead, "Aspects of multivariable statistical theory," *John Wiley & Sons,* Ch 1, 1-49, 1982.

[21] F. Rocca and C. Kostov, "Estimation of Residual Wavelets," 1987.

[22] S.J. Devlin et al, "Some multivariate applications of elliptical distributions," in "Essays on Probability and Statistics," Ed. S. Ikeda, *Shinko Tsusho Co., Tokyo,* 1976.

[23] G. Giannakis, "Signal Processing via higher-order statistics," *Ph.D. thesis, U.S.C.* 1986.

[24] G. Giannakis, "Cumulants: a powerful tool in signal processing," *IEEE Proceedings,* 1987 (to appear).