2004 IEEE Workshop on Machine Learning for Signal Processing

POLYNOMIAL EXPANSION OF THE PROBABILITY DENSITY FUNCTION ABOUT GAUSSIAN MIXTURES

Charles C. Cavalcante[1], J. C. M. Mota[1] and J. M. T. Romano[2]

[1] GTEL/DETI/CT/UFC
C.P. 6005, Campus do Pici, CEP: 60.455-760
Fortaleza-CE, Brazil
E-mails: {charles,mota}@gtel.ufc.br

[2] DSPCom/DECOM/FEEC/UNICAMP
R. Albert Einstein, C.P. 6101, CEP: 13.083-852
Campinas-SP, Brazil
E-mail: romano@decom.fee.unicamp.br

Abstract. A polynomial expansion to probability density function (pdf) approximation about Gaussian mixture densities is proposed in this paper. Using known polynomial series expansions we apply the Parzen estimator to derive an orthonormal basis that is able to represent the characteristics of probability distributions that are not concentrated in the vicinity of the mean point such as the Gaussian pdf. The blind source separation problem is used to illustrate the applicability of the proposal in practical analysis of the dynamics of the recovered data pdf estimation. Simulations are carried out to illustrate the analysis.

INTRODUCTION

Probability density functions (pdf) play a key role in signal processing since the estimation of data is, usually, done by exploiting statistical characteristics of the involved signals.

Blind source separation (BSS) is a general framework on the interference removal. In that case, no knowledge about the transmitted data neither about their pdfs is available. Information-theoretical tools are used to solve the problem but they require the knowledge of the signals pdfs and when they are not known some estimation has to be performed to made a solution possible.

Since only the mixture of data is available, the methods rely on the use of such measures to estimate the pdfs and the source signals. The estimative of the higher-order statistics (HOS) is one of the most used strategies to perform pdf estimation. With those measures, one can use polynomial expansions for the pdf about reference densities. However, the existing expansions are generally developed about a Gaussian density and many signals, such as the originated in digital communication systems, cannot be represented by those expansions.

In this paper we propose a new polynomial expansion developed about Gaussian mixtures that are suitable to represent the data of digital communication systems. We observe that the obtained expansion is a general model that can be used in several fields where some expansion of pdf using HOS is required. Our particular interest is to evaluate the impact of the number of HOS used to approximate the pdf in BSS adaptive algorithms.

The rest of the paper is organized as follows. Next section shows the existing polynomial expansions, namely the Gram-Charlier ann Edgeworth ones. Later, the proposed expansion is presented and an application in a BSS problem is shown in the sequence. Finally, we state our conclusions at the end of the paper.

POLYNOMIAL EXPANSIONS: GRAM-CHARLIER AND EDGEWORTH ONES

The polynomial representation of the probability density function is an expansion in an orthonormal series. The expansion is characterized by the use of the statistics of the signal which pdf we want to represent and a *reference density* [11, 3].

The reference density plays a key role since it has to be as similar as possible to the desired density. Further, the orthonormal series (or basis) is also reference-density-dependent. Let us describe a little the development of a general expansion.

Let the characteristic function (moment generator function) of a real[1] random variable (r.v.) y which fdp is $p_Y(y)$, defined as [9]:

$$\Omega_Y(\omega) \triangleq \int_{-\infty}^{\infty} p_Y(y)\exp(\jmath\omega y)dy, \tag{1}$$

where $\jmath = \sqrt{-1}$ and $\omega \in \mathbb{R}$.

Then, if the k-th moment of the r.v. y exists, we can expand $\Omega_Y(\omega)$ in a power series around $\omega = 0$ as [9]:

$$\Omega_Y(\omega) = 1 + \sum_{k=1}^{\infty} \frac{(\jmath\omega)^k}{k!} \cdot \kappa_k(y), \tag{2}$$

where $\kappa_k(y)$ is the k-th central moment of r.v. y.

[1]For complex-valued variables the development is similar [5], we use real ones for simplicity reason only.

If we take the logarithm of the characteristic function we have the cumulant generator function (second characteristic function) which is given by [9, 12]:

$$\ln\left[\Omega_Y(\omega)\right] = \sum_{k=1}^{\infty} \frac{c_k}{k!} \cdot (\jmath\omega)^k, \tag{3}$$

where c_k is the k-th order cumulant (or semi-invariant) of r.v. y. The cumulants are related to the moments by the following recursion [7]:

$$c_k = \kappa_k - \sum_{i=1}^{k-1} \begin{pmatrix} k-1 \\ i-1 \end{pmatrix} c_i \cdot \kappa_{k-i}. \tag{4}$$

A reference variable is defined by convenience, and its characteristic function is given by $\Omega_0(\omega)$, such as

$$\ln\left[\Omega_0(\omega)\right] = \frac{1}{2}(\jmath\omega)^2 + \sum_{k=3}^{\infty} \frac{c_{k,0}}{k!}(\jmath\omega)^k, \tag{5}$$

where $c_{k,0}$ is the k-th order cumulant of $\Omega_0(\omega)$.

If we subtract the cumulant generator function of r.v. y from the one of the reference density we have, after some manipulation, the following expression [3, 1]:

$$\frac{\Omega_Y(\omega)}{\Omega_0(\omega)} = 1 + \sum_{l=1}^{\infty} \frac{1}{l!} \cdot \left(\sum_{k=3}^{\infty} \frac{c_k - c_{k,0}}{k!} (\jmath\omega)^k \right). \tag{6}$$

Then, performing an inverse Fourier transform on $\Omega_Y(\omega)$ it is possible to write the following approximation for the probability density function of r.v. y about a reference density $p_0(y)$ [5, 6, 3]:

$$p_Y(y) = p_0(y) \sum_{k=1}^{\infty} C_k \cdot \mathfrak{b}_k(y), \tag{7}$$

where C_k are the coefficients of an orthonormal series expansion. The $\mathfrak{b}_k(y)$ are appropriate mathematical functions that compose the orthonormal basis given by the following expression [11]:

$$\mathfrak{b}_k(y) = (-1)^k \cdot \frac{1}{p_0(y)} \cdot \frac{d^k p_0(y)}{dy^k}, \tag{8}$$

Thus, the coefficients C_k of the orthonormal series are defined in terms of the k-th order cumulants of y. The first 8 ones are given by [3]:

$$
\begin{aligned}
C_1 &= 0 \\
C_2 &= 0 \\
C_3 &= \frac{c_3}{6} \\
C_4 &= \frac{c_4}{24} \\
C_5 &= \frac{c_5}{120} \\
C_6 &= \frac{1}{720}\left(c_6 + 10c_3^2\right) \\
C_7 &= \frac{1}{5040}\left(c_7 + 35c_4c_3\right) \\
C_8 &= \frac{1}{40320}\left(c_8 + 56c_5c_3 + 35c_4^2\right).
\end{aligned}
\tag{9}
$$

It is known that the C_k can be also written as polynomials of the k-th central moments [6]

Gram-Charlier Expansion

When the Gaussian distribution is considered as the reference one we obtain the *Gram-Charlier series expansion* which is given by [11]:

$$p_Y(y) = p_G(y)\left(1 + \sum_{k=3}^{\infty} C_k \cdot \mathfrak{h}_k(y)\right), \tag{10}$$

where $p_G(y) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{y^2}{2}\right)$.

The elements of the orthonormal basis for the Gram-Charlier expansion are then derived following Equation (8) when $p_0(y) = p_G(y)$ and in this case we denote $\mathfrak{b}_k = \mathfrak{h}_k$. The obtained polynomials are the *Hermite polynomials* which recursion generation rule is given by [3]:

$$\mathfrak{h}_{k+1}(y) = y \cdot \mathfrak{h}_k - k \cdot \mathfrak{h}_{k-1}(y). \tag{11}$$

The first 8 Hermite polynomials are [6]

$$
\begin{aligned}
\mathfrak{h}_0(y) &= 1 \\
\mathfrak{h}_1(y) &= y \\
\mathfrak{h}_2(y) &= y^2 - 1 \\
\mathfrak{h}_3(y) &= y^3 - 3y \\
\mathfrak{h}_4(y) &= y^4 - 6y^2 + 3 \\
\mathfrak{h}_5(y) &= y^5 - 10y^3 + 15y \\
\mathfrak{h}_6(y) &= y^6 - 15y^4 + 45y^2 - 15 \\
\mathfrak{h}_7(y) &= y^7 - 21y^5 + 105y^3 - 105y \\
\mathfrak{h}_8(y) &= y^8 - 28y^6 + 210y^4 - 420y^2 + 105.
\end{aligned}
\tag{12}
$$

Edgeworth Expansion

The Edgeworth expansion consists on the Gram-Charlier polynomial series ordered in decreasing order of its coefficients. Although, they have the same structure of the Gram-Charlier expansion they were developed independently [13]. Therefore, we can write Edgeworth expansion as [5, 3]:

$$\begin{aligned} p_Y(y) = p_G(y)\cdot \Bigg(1 &+ \frac{c_3}{3!}\mathfrak{h}_3(y) + \frac{c_4}{4!}\mathfrak{h}_4(y) + \frac{10c_3^2}{6!}\mathfrak{h}_6(y) + \frac{\kappa_5}{5!}\mathfrak{h}_5(y) \\ &+ \frac{35c_3c_4}{7!}\mathfrak{h}_7(y) + \frac{280c_3^3}{9!}\mathfrak{h}_9(y) + \frac{c_6}{6!}\mathfrak{h}_6(y) + \frac{56c_3c_5}{8!}\mathfrak{h}_8(y) \\ &+ \frac{35c_4^2}{8!}\mathfrak{h}_8(y) + \frac{2100c_3^2c_4}{10!}\mathfrak{h}_{10}(y) + \frac{15400c_3^4}{12!}\mathfrak{h}_{12}(y) + \cdots \Bigg). \end{aligned} \tag{13}$$

The use of the Edgeworth expansion is useful when a truncated version of the polynomial expansion is necessary and the most important terms should be retained. Both polynomial expansions are able to approximate a large number of probability density functions. However, typical densities encountered in digital communication systems cannot be approximated by none of the presented expansions. To cope with that a new polynomial expansion is proposed in next section.

POLYNOMIAL EXPANSION ABOUT GAUSSIAN MIXTURES DENSITIES

Digital communication systems have the characteristic that the transmitted signals are discrete, belonging to an alphabet $\mathcal{A}$ with cardinality S, and usually an additive noise with a Gaussian pdf. This implies that the pdf of the signals at the receiver output have the signals y with the following density distribution [1]:

$$p_{GM}(y) = \sum_{i=1}^{S} \frac{1}{\sqrt{2\pi\sigma_\vartheta^2}} \exp\left[-\frac{(y-a_i)^2}{2\sigma_\vartheta^2}\right] \cdot \Pr(a_i), \tag{14}$$

where a_i are the symbols of the transmitted alphabet, σ_ϑ^2 is the variance of each model and GM stands for Gaussian mixture.

If we want to exploit some characteristics of the signal at the receiver in a digital communication system we have to evaluate the cumulants of a distribution like the one described in Equation (14).

For this sake we have developed an orthonormal basis using Equation (8) for $p_0(y) = p_{GM}(y)$ using the fact that a sum of Gaussians can model a large number of densities as stated by Parzen [10]. Further, assuming that the alphabet $\mathcal{A}$ has zero mean, we achieve the following recursion rule for the orthonormal basis when the reference density equals to a sum of Gaussian mixtures with different means [1]:

$$\mathfrak{c}_{k+1}(y) = S \cdot y \cdot \mathfrak{c}_k(y) - k\mathfrak{c}_{k-1}(y), \tag{15}$$

where $\mathfrak{c}_k$ is the obtained orthonormal basis for the polynomial expansion about a Gaussian mixture density function. The first 8 $\mathfrak{c}_k$ are written as [1]:

$$\begin{aligned}
\mathfrak{c}_0(y) &= 1 \\
\mathfrak{c}_1(y) &= S \cdot y \\
\mathfrak{c}_2(y) &= (S \cdot y)^2 - 1 \\
\mathfrak{c}_3(y) &= (S \cdot y)^3 - 3(S \cdot y) \\
\mathfrak{c}_4(y) &= (S \cdot y)^4 - 6((S \cdot y)^2 + 3 \\
\mathfrak{c}_5(y) &= (S \cdot y)^5 - 10(S \cdot y)^3 + 15(S \cdot y) \\
\mathfrak{c}_6(y) &= (S \cdot y)^6 - 15(S \cdot y)^4 + 45(S \cdot y)^2 - 15 \\
\mathfrak{c}_7(y) &= (S \cdot y)^7 - 21(S \cdot y)^5 + 105(S \cdot y)^3 - 105(S \cdot y) \\
\mathfrak{c}_8(y) &= (S \cdot y)^8 - 28(S \cdot y)^6 + 210(S \cdot y)^4 - 420(S \cdot y)^2 + 105,
\end{aligned} \tag{16}$$

that are also a class of Hermite polynomial since the Hermite polynomials properties are still valid [3].

The polynomial expansion is written as:

$$p_Y(y) = p_{GM}(y)\left(1 + \sum_{k=3}^{\infty} C_k \cdot \mathfrak{c}_k(y)\right). \tag{17}$$

This new expansion made possible the investigation on the dynamics evolution on blind source separation problems. This is, explored in the simulations on next section.

COMPUTATIONAL RESULTS: BLIND SOURCE SEPARATION

In the design of blind source separation criteria, the use of higher order statistics is required in order to recover the sources [4]. Figure 1 shows the scheme of BSS.

When discrete sources are considered, two existing approaches that use a different number of higher order statistics in their structures can be evaluated. Namely, the multi-user kurtosis (MUK) [8] and the multi-user constrained fitting pdf (MU-CFP) [2].

Both of them use an equalization criterion constrained to have an orthogonal global system response. We can write the MUK as

$$\begin{cases} \max\limits_{\mathbf{G}} J_{\text{MUK}}(\mathbf{G}) = \sum\limits_{j=1}^{K} |\kappa[y_k]| \\ \text{subject to: } \mathbf{G}^H\mathbf{G} = \mathbf{I} \end{cases}, \tag{18}$$

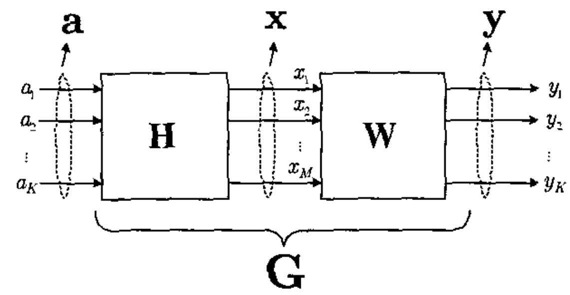


Figure 1: General scheme of blind source separation

where $\kappa[\cdot]$ is the kurtosis operator, $\mathbf{G} = \mathbf{W}^H\mathbf{H}$ is the global response matrix, $\mathbf{H}$ is the mixing system, W is the demixing system and $\mathbf{I}$ is the identity matrix. The adaptation procedure for the MUK is given by

$$\mathbf{W}^e(n+1) = \mathbf{W}(n) + \mu \operatorname{sign}(\kappa_a)\, \mathbf{x}^*(n)\mathcal{Z}(n), \tag{19}$$

where $\mathcal{Z}(n) = \left[\ |y_1(n)|^2 y_1(n) \quad \cdots \quad |y_K(n)|^2 y_K(n)\ \right]$. The orthogonalization procedure is performed using a Gram-Schimdt one [8].

The MU-CFP is given by

$$\begin{cases} \min\limits_{\mathbf{W}} J_{\text{MU-CFP}}(\mathbf{W}) = \sum\limits_{k=1}^{K} D_{p_Y(y)\|\Phi(y_k,\sigma_r^2)} \\ \text{subject to: } \mathbf{G}^H\mathbf{G} = \mathbf{I} \end{cases}, \tag{20}$$

where $D_{\circ\|\bullet}$ is the KLD between the pdfs, p_Y is the pdf of the ideally equalized signal and $\Phi(y, \sigma_r^2)$ is the parametric model given by

$$\Phi(y_k) = \frac{1}{\sqrt{2\pi\sigma_r^2}} \sum_{i=1}^{S} \exp\left(-\frac{\left|\mathbf{w}_k^H(n)\mathbf{x}(n) - a_i\right|^2}{2\sigma_r^2}\right) \cdot \Pr(a_i), \tag{21}$$

where σ_r^2 is the assumed variance of each Gaussian in the model. The stochastic gradient of the MU-CFP, for the k-th column of $\mathbf{W}$, is given by [2]:

$$\begin{aligned} \nabla J_{\text{MU-CFP}}\left(\mathbf{w}_k(n)\right) &= \frac{\sum\limits_{i=1}^{S} \exp\left(-\frac{|y_k(n)-a_i|^2}{2\sigma_r^2}\right)\left(y_k(n) - a_i\right)}{\sigma_r^2 \cdot \sum\limits_{i=1}^{S} \exp\left(-\frac{|y_k(n)-a_i|^2}{2\sigma_r^2}\right)} \mathbf{x}^* \\ \mathbf{w}_k(n+1) &= \mathbf{w}_k(n) - \mu\nabla J_{\text{FP}}\left(\mathbf{w}_k\right), \end{aligned} \tag{22}$$

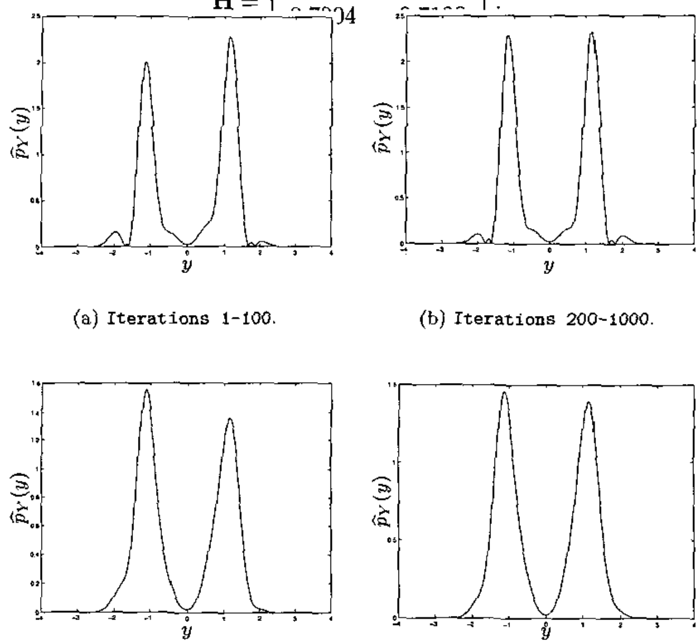
and the same Gram-Schimdt procedure in [8] is used to assure the ortoghonalization of the global response.

The main difference between both criteria is the number of used higher order statistics. The MUK uses the kurtosis only and the estimation of the pdf at the output of the separation device uses this HOS only. On the other hand, the MU-CFP is pdf-estimation-based and *all* HOS are involved in the process.

The polynomial expansion about a Gaussian mixture is used to evaluate the pdf estimation during the adaptation procedure. In the MUK case, we use only the kurtosis to approximate the pdf through $p_{GM}(y)$. The kurtosis is computed for some time intervals using the available data at the moment. A similar procedure is used for the MU-CFP. The difference is the greater number of considered HOS. For simplicity we use only 8 terms of the expansion in Equation (14). Figures 2 and 3 show the evolution of the pdf estimation for the MUK and MU-CFP algorithms at different time instants.

The simulation is done using two BPSK sources and two sensors. At the receiver is considered additive white gaussian noise which power is given by the signal-to-noise ratio (SNR) equals 30 dB. The mixture matrix is given by:

$$\mathbf{H} = \begin{bmatrix} 0.7138 & 0.7004 \\ \text{[illegible]}04 & \text{[illegible]} \end{bmatrix}. \quad (23)$$



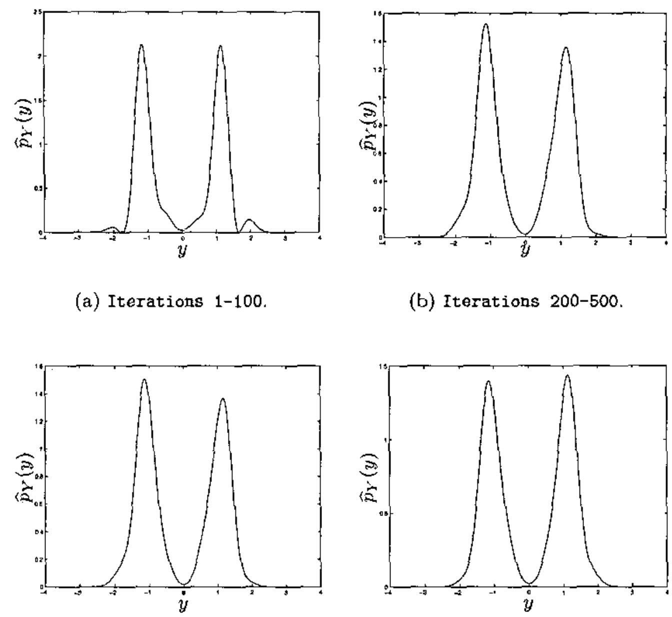
(a) Iterations 1-100. (b) Iterations 200-1000.

(c) Iterations 1500-2000. (d) Iterations 2500-4000.

Figure 2: Dynamics of the pdf estimation for the MUK algorithm.

As one can see, the pdf of the data is faster estimated by the MU-CFP than using the MUK. This behavior has been also observed in a wide range of channels and situations.

The presented approach works at the extrems: *one* HOS and *all* HOS.



(a) Iterations 1-100.

(b) Iterations 200-500.

(c) Iterations 500-1000.

(d) Iterations 2500-4000.

Figure 3: Dynamics of the pdf estimation for the MU-CFPA algorithm.

The gain in terms of convergence rate is very interesting for adaptive algorithms and such analysis can help the design of new methods for blind source separation.

CONCLUSIONS AND PERSPECTIVES

We have proposed a new polynomial expansion for probability density function approximation about Gaussian mixtures densities using higher-order statistics .

The proposed expansion is based on the Parzen estimation and uses the derivatives of known expansions (Gram-Charlier and Edgeworth) to a sum of Gaussians with different means. The obtained result is interesting in the sense that investigation about systems that have densities given by Gaussian mixtures, such as digital communication systems, can be done at each time instant regarding the pdf estimation of the data.

The blind source separation problem benefits from the proposed approach since the criterion may be used to guide the project of criteria to maximize the trade-off complexity $\times$ performance.

A natural extension to this work is investigate the most important higher-order statistics to provide a more efficient blind source separation criterion in terms of performance (faster convergence) and complexity.

REFERENCES

[1] C. C. Cavalcante, **On Blind Source Separation: Proposals and Analysis of Multi-User Processing Strategies**, Ph.D. thesis, State University of Campinas (UNICAMP) - DECOM, Campinas, SP - Brazil, April 2004.

[2] C. C. Cavalcante, F. R. P. Cavalcanti, J. C. M. Mota and J. M. T. Romano, "A Constrained Version of Fitting PDF Algorithm for Blind Source Separation," in **Proceeding of IEEE Signal Processing Advances for Wireless Communications (SPAWC 2003)**, Rome, Italy, June, 15-18 2003.

[3] S. Haykin, **Neural Networks: A Comprehensive Foundation**, Prentice Hall, 2nd edn., 1998.

[4] A. Hyvärinen, E. Oja and J. Karhunen, **Independent Component Analysis**, John Wiley & Sons, 2001.

[5] J.-L. Lacoume, P.-O. Amblard and P. Comon, **Statistiques d'Ordre Supérieur pour le Traitement du Signal**, (Traitement du Signal), Paris: Masson, 1997.

[6] J. D. Laster, **Robust GMSK Demodulation Using Demodulator Diversity and BER Estimation**, Ph.D. thesis, Faculty of the Virginia Polytechnic Institute and State University, Blacksburg, Virginia, March 1997.

[7] C. L. Nikias and A. P. Petropulu, **Higher-Order Spectra Analysis**, Prentice Hall, 1993.

[8] C. B. Papadias, **Blind Separation of Independent Sources Based on Multiuser Kurtosis Optimization Criteria**, John-Wiley & Sons, vol. 2, chap. 4, pp. 147–179, 2000.

[9] A. Papoulis, **Probability, Random Variables and Stochastic Processes**, (Electrical & Electronic Engineering Series), McGraw-Hill International, 3rd edn., 1991.

[10] E. Parzen, "On Estimation of a Probability Density Function and Mode," **The Annals of Mathematical Statistics**, vol. 33, no. 3, pp. 1066–1076, September 1962.

[11] B. W. Silverman, **Density Estimation for Statistics and Data Analysis**, (Monographs on Statistics and Applied Probability), Bristol, Great Britain: Chapman and Hall, 1986.

[12] C. W. Therrien, **Discrete Random Signals and Statistical Signal Processing**, (Prentice-Hall Signal Processing Series), Prentice-Hall International, 1992.

[13] E. J. Wegman, "Nonparametric Probability Density Estimation: I. A Summary of Available Methods," **Technometrics**, vol. 14, no. 3, pp. 533–546, August 1972.