JOINT PROBABILITY DENSITY FUNCTION ESTIMATION BY SPECTRAL ESTIMATE METHODS

A. Pagès-Zamora, Miguel A. Lagunas

Dpt.Teoria Senyal i Comunicacions, ETSE Telecom., Universitat Politècnica de Catalunya (UPC)
Campus Nord UPC, Ed. D-5. c/ Gran Capità, s/n. 08034 Barcelona, Spain.
e-mail: alba@gps.tsc.upc.es

ABSTRACT[*]

Estimation of probability density functions (PDFs) of a given random variable (r.v.) is involved in topics related to codification, speech or whenever a short record of data is available but a greater amount is needed. Existing methods go from the so-called Minimum Description-Length method, up to others based on the maximisation of the differential entropy imposing constraints on the moments of the r.v. In this paper we propose the estimate a PDF function by means of spectral estimate methods, since the positiveness and real character of any PDF function allow us to deal with it as a power spectrum density function. Particularly, the minimum variance method is focused because it can be generalised to multidimensional problems, being used in this paper to estimate the joint-PDF function of a multidimensional r.v.

1. INTRODUCTION

Out of existing PDF estimators we can find the so-called Minimum Description-Length [1], which looks for an accurate and parsimonious estimate, or the methods based on the maximisation of the differential entropy imposing constraints on the moments of the random variable (r.v.) [2], or even the basic and well-known histogram method. In this paper, we propose to estimate the PDF function of a given r.v. by means of well-known Spectral Estimate (SE) methods [3]. This is possible due to the existing parallelism between a PDF function and a Power Spectrum Density (PSD) function, since both are positive and real. This similarity also embraces the autocorrelation function, whose role is played in this particular purpose by the characteristic function.

This parallelism has become evident for the authors from the formulation of a model of non-linear systems (NLSs), on which some previous work has already been done [4]. Thus, the following section (sec.2) is devoted to present this NLS model in order to obtain the link between PDF and PSD estimation problems. Since in a previous paper [5] we have used the more familiar SE methods to estimate the PDF function of a single r.v., we focused in section 3 on those SE methods suitable for multidimensional problems, paying an special attention to the Minimum Variance method. These methods are used to achieve joint-PDF estimates of multidimensional r.v., being then tested in section 4 in a basic inference Bayesian problem.

---

2. THE NLSs MODEL

Although the similarity between the PDF and PSD functions is already known, we propose to infer it from the design of the following NLS model. The reason is twofold. First of all, the model provides the whole approach with a completeness sense and secondly, from our point of view, it can be also useful to motivate future works. In consequence, the purpose of this section is not the model itself which has already been focused [4], but to arrive at the parallelism previously mentioned. Thus, the model $\hat{g}(x)$ of equation (1) can be seen as a shorted Fourier series development of the input/output relation, $g(x)$, of a given memoryless NLS (k denotes discrete time, x(k) the input and y(k) the output).

$$\hat{y}[x(k)] = \hat{g}[x(k)] = \sum_{n=-N}^{N} a_n \cdot e^{jn\omega_o x(k)} \tag{1}$$

Assuming that the input signal, x(k), belongs to the range $[-X_{max}, X_{max}]$, the so-called principal frequency, ω_0 should be less than π/X_{max} in order to avoid an aliasing effect. Concerning the design of the model, and having previously fitted the order of the model N as well as the principal frequency ω_0, the optimum set of coefficients $\{a_n \quad n=-N,..N\}$ is found minimising the Mean Square Error (MSE), defined by the following cost function.

$$\xi_{\hat{g}} \equiv E[|e(x)|^2] = \int_{-\infty}^{+\infty} |g[x] - \hat{g}[x]|^2 \cdot p_X(x)\,dx \tag{2}$$

Thus, equating to zero the gradient of the cost function (eq.2) with respect to each coefficient, the following expression for m=-N,..N is obtained.

$$E[g(x) \cdot e^{-jm\omega_o x(k)}] - \sum_{n=-N}^{N} a_n \cdot E[e^{j(n-m)\omega_o x(k)}] \equiv 0 \tag{3}$$

The resulting linear system of equations allows a matrix-vector notation (eq.4), which involves the so-called *characteristic matrix* Ψ_X, whose elements $\Psi_X(n,m)$ depend on the characteristic function of x denoted by $\psi_X(.)$ (eq.4a), the set of coefficients gathered in vector **a** and the vector γ_g which corresponds to the cross-correlation between the desired output g(x) and the corresponding complex exponential functions (eq.4b). (The bold characters can represent either matrix or vector variables).

$$\Psi_x \cdot a = \gamma_g \tag{4}$$

$$\Psi_X(n,m) \equiv \psi_x[j(n-m)\omega_0] = \int_{-\infty}^{+\infty} p_x(x)e^{j(n-m)\omega_0 x}\,dx \tag{4a}$$

$$\gamma_g(m) = E[g(x)\cdot e^{-jm\omega_0 x}] \tag{4b}$$

It is worth noting at this point the similarity between the model design solution (eq.4) and the Wiener-Hopf equation [2]. In fact, the characteristic matrix plays the same role as the autocorrelation matrix does in the Wiener-Hopf equation. Moreover, as Ψ_X is also hermitic Toeplitz, the optimum set of coefficients that obey equation (4) can be computed by means of the well-known Levinson-Durbin algorithm [2] under a computational cost of $O(N^2)$. In a similar fashion, adaptive algorithms as the Least-Mean Squares and the Recursive Least Squares can be also considered to design the coefficients in non-stationary environments due, basically, to the linear dependence of the model on the parameters.

A part from the properties of the Fourier model that are out of the scope of this paper, we would like to remark that any linear problem that involves the estimate of the autocorrelation matrix is a priori suitable to be "translated" in terms of the variables of the NLS model. A particular example is the SE methods, where a PSD estimation is provided from a given estimation of the autocorrelation function. Thus, in a similar fashion, from an estimation of the characteristic function we will be able to compute a PDF estimation likewise.

Nevertheless, as this paper is focused on the joint-PDF estimation, it is necessary to introduce the Fourier series based model for NLSs with memory (eq.5), where vector x consists of the temporal diversity components of the input.

$$\hat{g}[x] = \sum_{n_0=-N}^{N}\sum_{n_1=-N}^{N}\cdots\sum_{n_Q=-N}^{N}\left\{a_{n_0,n_1,\ldots,n_Q}\cdot e^{j\left(n_0\,\omega_0\,x(k)+n_1\,\omega_1\,x(k-1)+\ldots+n_Q\,\omega_Q\,x(k-Q)\right)}\right\} \tag{5}$$

The design of this model with memory under a MMSE criterion would lead us to a similar solution to the one obtained in multidimensional MA models, where instead of the multidimensional autocorrelation function, we deal with the multidimensional characteristic function. As it will be seen in the following section, this model of NLSs with memory emerges whenever any multidimensional SE method is applied to the joint-PDF estimation of a multidimensional r.v.

3 THE JOINT-PDF ESTIMATION

In [5], the different SE methods were applied to the PDF estimate of a single r.v. In this case, we deal with the joint-PDF estimate of a 2-dimensional r.v. gathered in vector $x=[x_1\ x_2]^T$, although the approach presented below can be easily generalised to more dimensions. Let's assume then that from each value of x, we build a new data space consisting of the 2-dim. process $z_X(n_1, n_2)$.

$$x = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^T \rightarrow z_x(n_1,n_2) = e^{j\left(n_1\,\omega_1\,x_1 + n_2\,\omega_2\,x_2\right)} \tag{6}$$

It is straightforward to prove the second order stationarity of the $z_X(n1, n2)$ process, being furthermore its autocorrelation function a 2-dim discrete function that equals the joint characteristic function of the r.v. x sampled at the rate of $(\omega 1, \omega 2)$.

$$R_{zz}(m_1,m_2) \equiv E\left[z_x(n_1+m_1,n_2+m_2)\cdot z_x^*(n_1,n_2)\right]$$
$$= \psi_x(m_1\,\omega_1,m_2\,\omega_2) \tag{7}$$

Thus, the existing multidimensional SE methods applied to the 2-dim process $z_X(n1, n2)$ will provide an estimation of the Fourier transform of $\psi_{XX}(m1, m2)$, i.e. the joint-PDF function of x, $p_X(x1, x2)$. In fact, as the characteristic function has been sampled at the rate of $(\omega 1, \omega 2)$, the PDF function is in consequence repeated with a period of $(2\pi/\omega 1, 2\pi/\omega 2)$. As it occurred in the NLS model, the frequencies $\{\omega i$ $i=1,2\}$ depend on the range of the corresponding r.v. in such a way that ω_i should be less than $\pi/X_{max(i)}$, where $[-X_{max(i)}, X_{max(i)}]$ represents the x_i range.

At this stage, it is also interesting to review the role of the time k, that we assume it supplies the r.v. $x=x(k)$. As it is well-known, the SE methods need an estimate of the autocorrelation function, before computing the PSD estimation. In a similar fashion, an estimate of the characteristic function should be obtained to be able to compute the PDF estimation. In the simulations, the so-called empirical characteristic estimator has been chosen,

$$\hat{\psi}(m_1\omega_1,m_2\omega_2) = \frac{1}{K}\sum_{k=1}^{K}e^{j\left[m_1\,\omega_1\,x_1(k)+m_2\,\omega_2\,x_2(k)\right]} \tag{8}$$

in which an uniform average of the different exponential functions obtained from each value of x is performed.

3.1 The Minimum Variance Method

The Minimum Variance Method (MVM), due to Capon and also known with the misnomer of Maximum Likelihood Method, looks for a a 2-dim. filter $h_{xo}(n1,n2)$, which supplied by $z_X(n1, n2)$ (eq.9) provides an estimate of the joint-PDF function of x at the value $x_o=[x1,o\ x2,o]^T$.

$$y(n_1,n_2) = z_x(n_1,n_2)*h_{x_o}(n_1,n_2) \tag{9}$$

The x_o subscript in $h_{xo}(n1,n2)$ reveals, in advance, a dependence of the filter on the value at which the joint-PDF function is going to be estimated. Assuming that the 2-dim filter is based on a quarter plane (QP) region of support with $(N+1)^2$ coefficients [3], the output of the filter can be expressed in the following matrix-vector notation,

$$y(n_1,n_2) = \sum_{i_1=0}^{N}\sum_{i_2=0}^{N}a_{i_1,i_2}^* e^{j\left((n_1-i_1)\omega_1\,x_1+(n_2-i_2)\omega_2\,x_2\right)}$$
$$\equiv a^H\cdot z_x(n_1,n_2) \tag{10}$$

where vectors a and z_x gather in a convenient order the coefficients of the filter and the complex exponential functions involved in (eq.10), respectively. The MVM method designs the filter coefficients applying an output power minimisation criterion (eq.11) that leads us to a

2937

quadratic form involving the autocorrelation matrix of z_x, denoted by R_{zz}.

$$a_{MVM} = \arg\min_a \left\{ E\left[\left|a^H z_x\right|^2\right] = E\left[a^H z_x z_x^H a\right] = \\ = a^H R_{zz} a \equiv a^H \Psi_{xx} a \right\} \tag{11}$$

As it is shown, this autocorrelation matrix is equivalent to the named characteristic matrix Ψ_{xx} but now, unlike in the 1-dim case (eq.4), this matrix is no more Toeplitz although it keeps the hermitic property.

Expressing the output power in terms of the transformed domain (eq.12), it can be seen how the filtering process denoted by $H_{xo}(x1,x2)$ (the Fourier transform of the filter impulse response) is applied to the input joint-PDF function $p_X(x)$ to provide the estimation $\hat{p}_X(x_o)$.

$$E\left[\left|y(n_1,n_2)\right|^2\right] = \int_X \left|H_{x_o}(x)\right|^2 p_X(x)dx \equiv \hat{p}_X(x_o) \tag{12}$$

From this point of view, the goal of minimising the output power is to minimise a leakage effect. On the other hand, and in order to avoid the trivial solution, the MVM method forces the transformed filter response to have a unity absolute value where the joint-PDF is estimated.

$$H_{x_o}(x_o) = \sum_{i_1=0}^{N} \sum_{i_2=0}^{N} a_{i_1,i_2}^* e^{-j(x_1 i_1 + x_2 i_2)} \Bigg|_{x_{1,o},x_{2,o}} = \\ \equiv a^H \cdot s_{x_o} \equiv 1 \tag{13}$$

Applying the output power minimisation (eq.11) under the constraint of equation (13), we arrive at the "optimum" set of coefficients given by the MVM method,

$$a_{MVM}(x_o) = \frac{\Psi_{xx}^{-1} \cdot s_{x_o}}{s_{x_o}^H \cdot \Psi_{xx}^{-1} \cdot s_{x_o}} \tag{14}$$

which provides a filter output whose power is the desired joint-PDF estimation.

$$\hat{p}_X^{MVM}(x_o) = \frac{1}{s_{x_o}^H \cdot \Psi_{xx}^{-1} \cdot s_{x_o}} \tag{15}$$

As it has been previously mentioned, it can be verified how the coefficients of the MV filter, and in consequence the filter itself, depend on the value x_o where the joint-PDF function is estimated.

3.2 The Normalised MVM method and Periodogram method

In a similar way, other multidimensional SE methods can be developed to estimate the joint-PDF function of a multidimensional r.v. Thus, first of all we propose the so-called Normalised MVM method (NMVM) that, briefly, consists in dividing the output power of the MVM filter by the bandwidth of the filter in order to deal with an spectral estimate instead of a power one [6].

$$\hat{p}_X^{NMVM}(x_o) = \frac{s_{x_o}^H \cdot \Psi_{xx}^{-1} \cdot s_{x_o}}{s_{x_o}^H \cdot \Psi_{xx}^{-2} \cdot s_{x_o}} \tag{16}$$

The NMVM method is characterised by an improvement in front of the MVM method of the estimate resolution, providing a sharper estimation which, in general, is not suitable for smooth PDF shapes.

Another SE method also taken into account is the well-known Periodogram estimator (PM).

$$\hat{p}_X^{PM}(x_o) = \frac{s_{x_o}^H \cdot \Psi_{xx} \cdot s_{x_o}}{(N+1)^2} \tag{17}$$

It is important to remark that all the joint-PDF estimators presented involve the previous computation of the so-called characteristic matrix (eq.11). As it has been mentioned before, this matrix is built from an estimation of the characteristic function obtained from all the different values of the r.v. x that are given (eq.8).

SIMULATION RESULTS

In this section, the performance of the different presented methods, along with the Histogram Method (HM), in estimating 2-dim joint-PDF functions is checked by addressing an equalisation problem. The proposed equaliser builds an approximation of the Bayes estimator from an estimate of the conditioned PDF function. The input signal x(k) is normal distributed with zero mean and unity variance. It supplies a memoryless non-linear system (eq.18), which emulates the response of a non-linear amplifier with a saturation effect, followed by and additive gaussian noise of 0dB, denoted by w(k).

$$y(n) = g[x(n)] = sign(x) \cdot \left(1 - e^{-x^2/0.05}\right) \tag{18}$$

The output of the system is thus v(k)=y(k)+w(k), and it can be easily seen that its PDF function conditioned to x is normal distributed with mean equal to g(x) and variance equal to the noise variance.

The estimates of the joint-PDF function of (x,v) obtained by the different SE methods are shown in figure (1). In these simulations, the principal frequencies ($\omega1$, $\omega2$) corresponding to the r.v. x and v, respectively, have been chosen equal to (π/X_{max}, π/V_{max}) since the expected PDF functions will have even symmetry with respect to the origin. On the other hand, the estimate of the characteristic matrix is built from the characteristic function estimate provided by expression (8), with K=1000 different values of the r.v. x(k) and v(k), and also considering a quarter plane of region of support with N=6, i.e. estimating the characteristic function in a square plane of 7x7 dots.

The resulting joint-PDF estimates are included in figure (1) where it can be seen how the shape of the concentric ellipses follows the non-linear transformation, in dashed line, as it was expected. The methods that achieve a better estimation are clearly the MVM and PM methods. Nevertheless, the first one seems to achieve a best resolution than the PM method. On the other hand, the expected sharp behaviour of the NMVM is evident.

2938

The equaliser is designed in order to perform an approximation of the Bayes estimator (eq.19) using the presented SE methods to obtain the needed PDF functions.

$$\hat{x}_{BAYES} \equiv E[x/v] = \int_{-\infty}^{+\infty} x \cdot p_{X/V}(x/v)dx =$$

$$= \int_{-\infty}^{+\infty} x \cdot \frac{p_{X,V}(x,v)}{p_V(v)} dx \qquad (19)$$

Thus, the expected value of the Bayes estimator is computed by averaging for each value of v, the multiplication of x by the PDF function of x conditioned to v, over the x axis. The conditioned PDF function is approximated by means of the Bayes theorem (eq.19) in terms of the joint-PDF function of (x,v) and the single PDF function of v. In fact, instead of using the actual PDF functions, the estimations obtained by the different methods are used.

The different equaliser responses are shown in figure (2) along with the ideal equaliser (dashed line) that would implement the inverse function of the non-linearity (eq.18). The equaliser response is computed with K=1000 samples of x(k) and v(k), useful to estimate the joint characteristic function of (x,v), in a square of 7x7 values (QP region of support with N=6). Each title includes the MSE in % between the ideal equaliser and the implemented one, averaged over 25 independent realisations.

In this particular case, the MVM method shows a superior performance in front of the others, becoming also evident the sharp character of the NMVM estimator and the too smooth estimation provided by the PM.
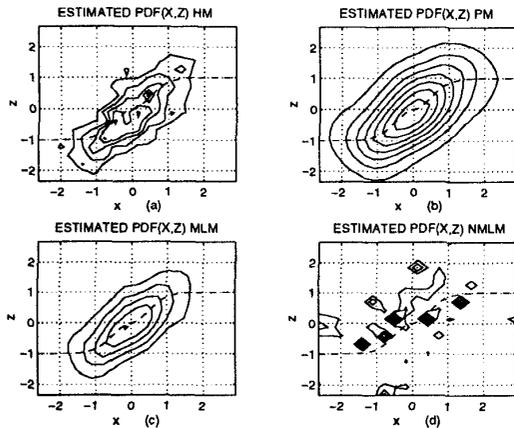
Figure 1. Estimation of the 2-dim joint-PDF function of the pair [x v] by different methods along with the non-linearity (dashed line). (a) Histogram Method (HM). (b) Periodogram Method (PM). (c) Minimum Variance Method (MVM). (d) Normalised MVM (NMVM). A record of 1000 values of (x(k),v(k)) are used to estimate the 7x7 dots of the characteristic function (N=6).
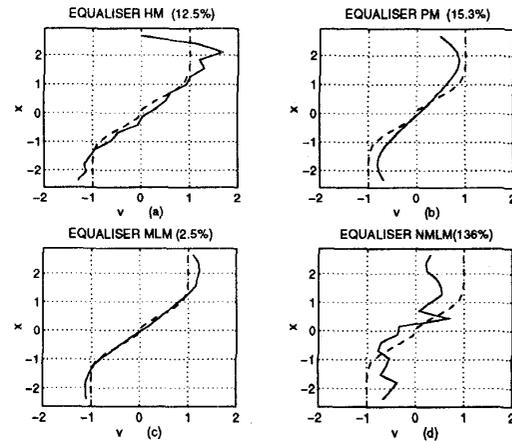
Figure 2. Ideal equaliser (dashed line) and estimated one (solid line) computed applying the Bayes estimation and using the estimated PDF functions by different methods. Each title includes the MSE (in % averaging 25 results obtained from independent realisations) of the estimated equaliser with respect to the ideal response.

CONCLUSIONS

The parallelism between the PSD function and the PDF of a r.v. has been established from the problem of modeling non-linear systems. This relation allows the use of SE methods to estimate any PDF function, paying in this paper an special attention to multidimensional methods that provide an estimation of the joint-PDF function of a multidimensional r.v.

REFERENCES

[1] Barron A.R., Cover T.M., "Minimum Complexity Density Estimation", IEEE Trans. on Information Theory, Vol.37, N°4, pp.1034-1054, July 1991.

[2] Zivojnovic V., "Higher-Order Statistics and Huber's Robustness", IEEE-ATHOS Workshop on Higher-Order Statistics, pp.236-240, Aiguablava, Girona, Spain, June 1995.

[3] Kay S.M.,Marple S.L.,"Spectrum Analysis-A Modern Perspective", Proc.IEEE, Vol.69, No.11, pp.1380-1419, November 1981.

[4] Pagès-Zamora A., Lagunas M.A., Nájar M., Pérez-Neira A., "The K-filter: A new architecture to model and design non-linear systems from Kolmogorov's theorem", EURASIP Signal Processing, Vol. 44, No. 3, pp. 249-267, July 1995.

[5] Pagès-Zamora A., Lagunas M.A.,"New Approaches in Non-linear Signal Processing: Estimation of the PDF function by Spectral Estimation Methods" IEEE-ATHOS Workshop on Higher-Order Statistics,pp.204-208, Aiguablava, Girona, Spain, June 1995.

[6] Lagunas M.A.,Santamaria M.E.,Gasull A,Moreno A., "Maximum Likelihood Filters in Spectral Estimation Problems", Signal Processing, Vol.10, No.1, pp. 19-34, January 1986.