THE KOLMOGOROV MAPPING THEOREM IN SIGNAL PROCESSING

Miguel A. Lagunas

ETSI de Telecomunicacion, Departamento de TSC
Apdo. 30002, 08080 BARCELONA, SPAIN

ABSTRACT[1]

Since the publication in 1957 of the work of Andrei Kolmogorov /8/ in mapping a function of multiple variables by means functions of a single variable, many mathematicians and engineers try , with different degree of success and not without controversy /19/, to find the direct application of it to multiple extremes problems, rooting of multivariate polynomials, neural networks and pattern recognition. This paper revisits the theorem from the optic of a generalised architecture for signal processing /28/. It is envisaged the high potential of the theorem to handle either linear or non-linear processing problems. A specific implementation following the main guide-lines of the theorem is reported, as well as some preliminary results concerning the design, implementation and performance of non-linear systems. The applications cover non linear transmission channels for communications, instantaneous companders and prediction of chaotic series.

I. INTRODUCTION

Regardless linear systems have been the main component of communications and signal processing engineering, with a clean and formal mathematical background, it is interesting to note the massive presence of non-linear systems and devices in modern communications equipment. Analog mixers and modulators, companders, synchronism units, decoders are examples of a vast list from which it can be said that the role of non-linear systems components and devices is, most of the cases, absolutely crucial in the global systems when compared with the corresponding linear part of them. Also, the inherent non-linear behaviour of many amplifiers and filters, when handling high dynamic ranges, put in scene the issue of linearisation in the analog domain. Surprisingly, there is not a formal manner to model a non-linear system neither an architecture with support the reported schemes with a broad ambit of application, being most of the works reported dealing with the problem of gaussian input /14//15//16//25/26//27/. In fact, it can be said that polynomial modelling constitutes the only way to work with non-linear filters. At the same time, cascading non-

linear memory less with convolutive systems, using the two possible alternatives, was considered as a serious attempt to implement non-linear equalisers for dispersive communication channels in the seventies.

More recently, two topics emerged in an almost separate manner. Neural networks /2/ contributed, at the architecture level, to increase the interest and viability of non-linear processing schemes. On the other hand, high order processing /3//4/, at the algorithmic level, were in the scene as a basic tool to increase the scope of, let us say, classical processing systems developed from second order functions. In any case, the architecture and the processing tools efforts seem to go ahead one apart from the other, without an uniform and global approach to the non-linear processing framework.

The purpose of this work was, at the beginning three years ago, to find out refreshing mathematics, in the line of what has been mentioned in the previous paragraph, focussing the linearisation problem of communication components. The theorem reported by Kolmogorov concerning the representation of multivariate functions with single variable functions /20/ shown the link, in some sense especulative, between neural networks, high order processing and the EM algorithm /6//5/. In fact, a lateral inhibition network with high order learning is a clarifyed example concerning the implementation of the EM algorihtm's principle. The application reported was a DOA tracker for moving sources which, in a blind context, generates a dedicated beam for every source impinging the aperture, for radial velocities below one degree per snapshot. No matter the mapping theorem was the basic motivation to encompass the three techniques in a source separation problem /7/,9//10//11//12//13/, no specific contribution to the architectural design of non-linear systems was concluded from this work. More interesting than the blind DOA tracker was a system, with the same background, designed to solve collision problems in a time reference beamformer for mobile communications. The architecture was again a lateral inhibition network, with two extended Kalman filters to acquire and track two simultaneous carriers, which act as a call signal demanding a dedicated beam. The learning was high order for the inhibition weights, without substantial penalty in acquisition, convergence rate, beamforming and complexity with respect the single call system /30//31/. This last system was an example of up to what degree the EM algorithm involves an architecture, being the scope of

applications enlarged when the estimation steep is considered as an inhibition one, which reduces the multiple extrema problem to a single maximum objective.

Coming back to the mapping theorem, it was considered the importance of the architectural aspect involved in the Kolmogorov's theorem. In fact, what is going to be reported hereafter is more an architecture than an algorithm for signal processing. Ir seems clear that most of the possibilities of success in non-linear processors, as well as in neural network approach for a problem solver, are based in the architecture choice, locating in a secondary plane the corresponding algorithm for design and/or learning. The architecture will be refereed from now on as the kol-filter. Regardless the paper focuss its content in non-linear filtering, the impact of the theorem in many different problems, with different applications, rises the doubt of up to what degree the reported scheme is just a very small part of the potential applications and consequences of it. The author is highly convinced that the theorem will be still alive and that the reported material are really very small and specific results, with a high component of speculation and lack of formality, that in any case do not preclude further contributions from the work of Andrei Kolmogorov.

2. THE MAPPING THEOREM.

The mapping theorem due to A. Kolmogorov proves the existence of a two stage architecture to map and input vector $\underline{x}$ in an output vector $\underline{y}$. Being an existence theorem, there is not reported yet any practical implementation, for non-linear processing, following in strict sense its guide-lines. In fact, it seems that the issue will remain in the current status and that, in future work, the theorem will act as fundamental motivation to formalise the solutions of application driven problems. It may be that this way is the adequate manner to arise at a full profit, at the engineering level, of a mathematical work of outstanding interest in signal processing.

No matter in the literature we can find, together with the original version, many descriptions as well as fundamentals and controversy about the theorem, hereafter we will reduce its reading to the aspects of interest for signal processing; further reading in the theorem can be found in the following references /8//17//19//18//20/. Given a continuous mapping g(.) function of n input variables, the components of the input vector $\underline{x}$, it exists an architecture, based in single variable functions that implements the mapping $\underline{x} \rightarrow g(\underline{x})$. The architecture contains two stages of processing. Selecting the case of single output (i.e. the output vector reduces to a scalar y), the theorem in its original formulation had the following representation:

$$y = g(x_1, ..., x_n) = \sum_{p=0}^{2n} \gamma_p (\sum_{q=1}^{n} \phi_{pq}(x_q)) \qquad (1)$$

Later on, it was reported the possibility of changing functions γ_p by a single function γ (without any change in the definition interval); at the same time, functions $\phi(.)$ of

the first stage can be replaced by 2n+1 functions ϕq (q=1,2n+1) plus a chosen parameters λ_p which depend on the number of input components n. The formulation that is going to be used here is another possibility, which reduces to a single function the first stage and preserves the 2n+1 of the second stage. Thus, the first stage of processing is formed by (2),

$$z_p = \sum_{q=1}^{n} \lambda_q \phi(x(q), \varepsilon_q) \qquad (2)$$

where the function or transform represented by $\phi(.)$ does not depend on the specific mapping, and the parameters λ_q and ε_q also do not depend on the mapping, but they do on the number of components of the input vector n.

The second stage of processing is formed by 2n+1 functions γ_p (.), which, by direct superposition, form the output function y.

$$y = \sum_{p=1}^{2n+1} \gamma_p (z_p) \qquad (3)$$

Note that the most interesting part of the theorem is the, let us say, absolute value or the wide ambit of application of the functions forming the first stage. It seems to be like a general transform which is mandatory to any processor architecture for the mapping problem. An approach that keeps some similarity with this transform is the use of saturated devices (limiters) prior a Volterra approximation in modelling a compander. This scheme arises when a given non-linearity is modelled by a polynomial approximation. The approximation by itself does not include saturation , in consequence, and to avoid that the polynomial approximation of the compander may produce out of range values at the output, it is necessary to saturate the dynamic of the input values applied to the Volterra system.
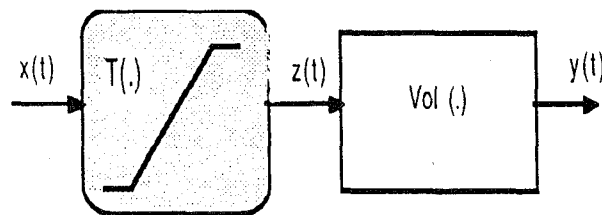

Fig. 1. A transform before the Volterra approximation of a given compander helps to avoid high dynamic output ranges and protects later stages in communications equipment.

At the same time, 't is well known the extraordinary performance of logarithm companders in many applications for coding and noise reduction both in speech and high fidelity audio /23//24/. Somehow the mentioned experience about certain transforms, together with what is stated in the

theorem, for the first stage of processing, arise to the conclusion that it may exist a data transform which, independently of the mapping and connected with the extended use of logarithmic and sigmoidal companders, reduces the complexity and improves the performance of a second stage of processing based only in Volterra systems. As we will see later on this paper, this architecture has a formal approach and somehow provides additional reasoning to the use of the above functions. To be more specific, the architecture of the kol-filter will be a first stage formed by phase modulators (PM) followed by Volterra filters.

Before we pass to the next section, note that the problem to be considered is the filtering one, where a data diversity input, that could be either time (FIR filters) or spatial (array processing), is the input to the processor which will map this data diversity in a given output function $y(t)$.

$$y(t)=g(x(t),x(t-T),....,x(t-Q.T)) \qquad (4)$$

3.VOLTERRA, FOURIER AND KOL-FILTERS.

Regardless the main interest of non-linear processing use to be reduced to the case of data diversity input mapped to a single output (i.e. non-linear system with memory), the presentation will start with the case of non diversity data input. When adequate, an indication will be provided about the extension to the general case.

Thus, assuming a memory less non-linear function $g(.)$,
$$y(t)=g(x(t)) \qquad (5)$$

the polynomial approximation models the above system as (6), which will be refereed as the Volterra filter or model.

$$y_v(t)= \sum_{q=1,3,...}^{Q} d(q).x^q(t) \qquad (6)$$

The corresponding filter for the case of diversity in the input reduces to include kernels with all suitable combinations of the diversity components, both in number and exponents. Continuing with the diversity case, a reduced complexity schemes, at the expense of approximation quality of the actual mapping, are the so-called Wiener and Hammerstein architectures /1//22/, which basically are the cascade of a non-linear system without memory with a linear system or viceversa.

An alternative to the above representation, more in line with the architecture involved in the mapping theorem, is to use a Fourier representation of the mapping, as it is indicated in (7).

$$y_f(t)= \int_{-\infty}^{\infty} G(w).exp(j.w.x(t)).dw \qquad (7)$$

Furthermore, given the input dynamics, the above mapping can be reduced to a Fourier Series which, under an approximation error, can be reduced to a finite Fourier series as in (8).

$$y_f(t)= \sum_{p=-P}^{P} a(p).exp(j.q.w_o.x(t)) \qquad (8)$$

Before going into details on the mentioned approximation, note that the Fourier model keeps a close relationship with the theorem, in the sense that a first stage of PM modulation is always required, independently of the mapping itself, and a second stage, specific for the given mapping function , follows. This second stage can be implemented with a Volterra filter.
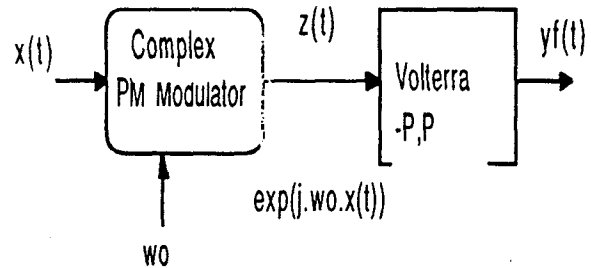

Fig 2. Fourier model of the mapping function showing some similarity with the two stage processor of the mapping theorem.

Some other guide-lines of the theorem are pending like the necessary 2Q+1 functions of the first stage and the parameters λ and ε.

In the case that the mapping has an odd symmetry the above model can be represented as a power series of non-complex PM modulators. At the same time, to do not exclude the case of non-cero output for zero input, the dc term has to be added to the model as it is denoted in (9), where the sine model has been reformulated in terms of the harmonics of the PM carrier frequency.

$$y_f(t)=b(0)+ \sum_{q=1}^{Q} b(q).sin^q(w_o.x(t)) \qquad (9.a)$$

$$y_f(t)=c(0)+ \sum_{q=1}^{Q} c(q).sin(q.w_o.x(t)) \qquad (9.b)$$

With respect the dc term in the above formulas, note that, for the case of (9.a), it will be necessary, even there is a zero input to zero output conversion in the mapping, since internal modulation products may produce a dc term that has to be adjusted or compensated to reproduce the corresponding mapping function. At this moment, it can be said that for the case of Q-components input it is necessary to have n PM modulators plus the dc term. In other words, it remains Q additional functions to match the above processor with the theorem.

Before going ahead, it is worthwhile to compare the approximations associated to the single Volterra model and the Fourier model. Both approximations are in principle equivalent since, assuring no over modulation (i.e. $Q.w_o.x(t)<1<\pi$), the exponential term in (7) can be

approximated by a Taylor's series producing (10) which, in a very generic form, relates the two approaches.

$$a(p)= \int_{-\infty}^{\infty} j.G(w) . \frac{(jw)^q}{q!} . (-1)^{q+1} . dw \qquad (10)$$

Nevertheless, the above derivation reduces to a formal relationship since the convergence of the two approximations, with increasing order, to the actual mapping could be quite different (an extreme case is for sinusoidal-like mappings which are very common in practice). More important that the previous argument, in an adaptive design of the approximation weights, the bounded dynamic range of the Fourier kernels acts very much in favour of it, with respect the Volterra modelling, improving dramatically the convergence rate for similar adaptive algorithm or learning rule. Also, it should be noticed that the cross-correlation of the kernels involved in each approximation, easy to compute for the case of gaussian distributed input data, is lower in the Fourier model than in the Volterra case; this improves convergence rate and provides an effective order reduction for similar levels of performance.

With respect to the design of the co-efficients, it can be carried out by direct frequency sampling of the Fourier transform of the mapping $G(w)$, when the desired mapping is a priori known, or by a LMS like algorithm for the sinusoidal kernels, using standard MSE objectives.

The extension of the model to the case of data diversity input is done by setting a PM modulator for every diversity component followed by a specific Volterra system for every branch. This generates n+1 functions $z_p(t)$; the addition of these functions provides the output of the filter.

$$z_p(t)= \sin (w_p . x(t-pT)); \ p=1,n \qquad (11)$$

$$f_p(t)= \sum_{q=1}^{Q} c(q) . z_p^q(t); \ p=1,n \qquad (12.a)$$

$$f_o \qquad (12.b)$$

$$y(t)= \sum_{p=0}^{n} f_p(t) \qquad (13)$$

This structure makes evident the combination of PM modulators, with different frequency for every diversity component, and the specific Volterra systems.

Let us to revisit, at this moment, the approximation involved in passing from the continuous Fourier model to the Fourier series in (8). Note that the formulation we arise in (11)-(13) sets up the roll of the PM modulators, followed by memory less Volterra systems, as the way out to implement the sampling of the continuous Fourier Model. In other words, (11) with (12) represent the multidimensional sampling of the corresponding Fourier representation of the model. This sampling is uniform and concentrated in the axes of the multidimensional function $G(w_1,...,w_p,...,w_Q)$, but not in the plane, which should be more adequate in general. In this sampling problem are imbeded two arguments claimed in the mapping theorem as we will see in the next paragraphs.

About the sampling involved in the Fourier series representation of the continuous model, and focussing again the case of non-diversity input, the corresponding model can be re-formulated as(14),

$$z_1(t)= \sin (w_1 . x(t)) \qquad (14.a)$$

$$f_1(t)= \sum_{q=1}^{Q} c(q) . z_1^q(t) \qquad (14.b)$$

$$f_o \qquad (14.c)$$

$$y_{kol}(t)= \sum_{p=0}^{1} f_p(t) \qquad (14.d)$$

where it is evident the uniform sampling performed by the Volterra system at the harmonics of the modulation frequency.

Related with the problem of sampling, it is interesting to remind the important consequences of random sampling for discrete processing in order to emulate continuous systems. It is possible by suitable randomising schemes to design, and to implement, discrete systems with reduced, in practice negligible, aliasing effect (see /21/ for a detailed study on randomised signal processing). The basic idea is to achieve an almost uniform distribution of sampling intervals because the expected value of the discrete processing converges to the actual or continuous value without aliasing effects.In consequence, a better approach would be to randomise the mentioned sampling in order to obtain an approximation $y_{kol}(t)$, such that the expected value of it coincides with the actual function $y(t)$. Just to describe briefly the effect claimed under the randomised assumption, and taking w_s as the random samples of the continuous $G(w)$, the expected value of the model output will be:

$$E< y_r(t) >= \int_{-\infty}^{\infty} G(w). \exp(jwx(t)) . (\sum_{s=-\infty}^{\infty} \delta(w-w_s).$$
$$p_s(w_s).dw_s). \ dw \qquad (15)$$

where, it is clear that, whenever the distributions of the sampling frequencies overlap in a constant or uniform pdf, the expected value will coincide with the actual one.

One way out to produce an almost random sampling, or scanning, of the frequency axis is to modify the Volterra system that follows the PM modulator; nevertheless, this will increase the implementation complexity and somehow will deviate the architecture from the theorem, in the sense that the random sampling effect takes part inside the first stage of processing, i.e. it should be independent of the mapping itself. The manner selected to produce this random sampling effect is to add another PM modulator with modulation frequency slightly deviated from the initial one. This implies the presence of another PM modulator with modulation frequency equal to $w_1+ \varepsilon_1$. Note that this additional modulator may introduce the random effect whenever the harmonics of the second frequency do not

coincide with the original one. In fact, taking an MSE technique to find out the co-efficients, c(q) in (14), the selection of the perturbation influences the condition number of the corresponding kernel matrix. Note also that in the case of diversity data input, the selection of the perturbations ε_p (p=1,Q) depends on the number of input diversity components Q and not in the specific mapping, as it was anticipated in the existence theorem. At the same time, the perturbed modulators, together with the dc term, seems to be the necessary 2Q+1 functions claimed in the theorem.
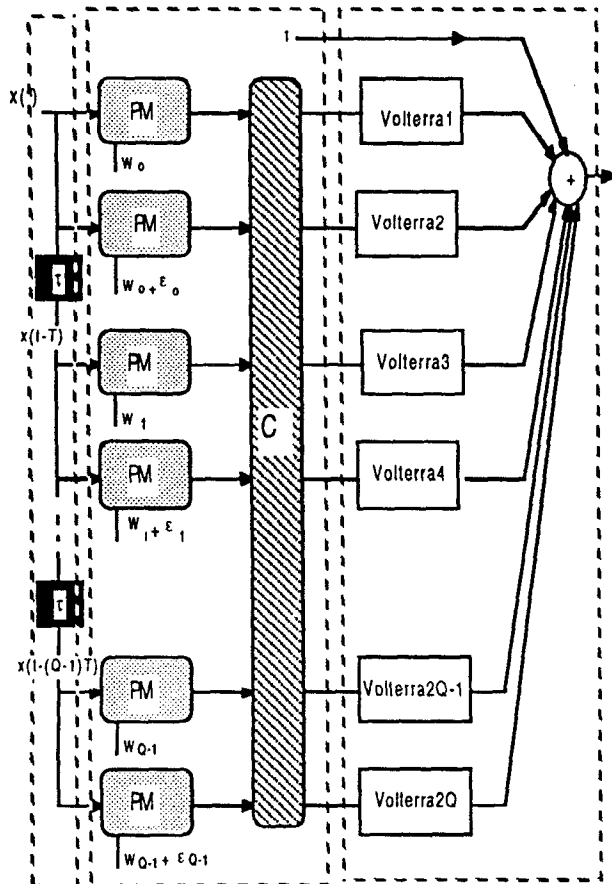

Figure 3. The architecture of the Kol-filter. The three steeps are imbeded in dashed squares, these squares indicate the diversity stage, the transforms or phase modulators with the nominal and perturbed frequency plus the linear combiner, and the specific Volterra systems.

Still there is a problem associated with sampling only along the frequency axis and not in the multidimensional frequency plane. In other words, with the current scheme, every diversity component is scanned, almost at random, in its corresponding frequency axis of the function $G(w_1,....,w_p,....,w_n)$. It is clear that combining the modulated versions prior to the Volterra processing will produce a frequency scan in the multidimensional frequency plane, leaving to the Volterra order the convergence to the actual system to model. The solution was to include a linear combiner of the modulated diversity components, prior to

the Volterra systems of the second stage, which was formulated as (16).

$$C = \begin{matrix} 100..0 \\ 110..0 \\ 111..0 \\ 111...\end{matrix} \qquad (16)$$

This combiner, which is also independent of the specific mapping, seems to play the role assigned to parameters λ included in the theorem (see (2)). It is clear that the choice for the linear combiner reduces to a very specific scanning mode of the multidimensional frequency plane. In the case that some a priori knowledge of the mapping Fourier transform is available, both, the selection of modulation frequencies, perturbations and combiner, could be oriented to ensure adequate sampling of every local extremes of the mentioned Fourier representation of the mapping.

The architecture that summarises all the points included in this section, refereed herein as the kol-filter, is depicted in figure 3.

It is worthwhile to remind that this structure includes a dedicated processing for every diversity branch, being a steep ahead from the coarse approximation which is associated with the existing alternatives, refereed as Wiener and Hammerstein structures. Finally, note that the Volterra kernels can be designed with and MSE criteria, when a training output is available or by direct sampling of the multidimensional Fourier function associated with the system. The adaptive versions can be applied also to the mentioned co-efficients, having the benefit of the bounded dynamic range of the corresponding kernels.

4. SOME APPLICATIONS OF INTEREST

The kol-filter was used to equalise a communication system with non-linear behaviour in the transmitter (above 6% distortion to the second and third harmonic), dispersive communication channel, and non-linear receiver effects (2% distortion to the second harmonic). The second stage weights were estimated by using a MSE objective. The performance of the kol-filter was highly above the corresponding Volterra or Hammerstein alternatives. The residual, with respect the undistorted transmitted signal was 2% of the reference power, 16 and 42% for the kol-filter, Volterra and Hammerstein respectively. A very important result was that even when no distortion was set in the overall communication link (i.e. the system to be equalised was linear) the kol-filter shown the same performance that the optimum linear Wiener filter.

Other aspect of interest is the ability of the herein reported filter to predict, with no appreciable error, chaotic signal series. The so-called Tent-map was used to evaluate the prediction capabilities of the non-linear kol-filter with excellent results.

The design, block or adaptive, of instantaneous companders was done in comparison with classical Volterra models. Again the quality of the mapping, in block processing mode, and also the convergence rate in adaptive

mode were dramatically good when compared with the classical alternative. Similar results were obtained for linearising power amplifiers, reducing the scheme to an array of medium power amplifiers with no relevant distortion in the global response.

Finally, the kol-filter ideas has been used in order to obtain high rate 1-bit A/D and D/A converters. Being this work at a very preliminary stage, the result are very promising to encourage further work in the topic as a valid alternative to the CD converters used currently.

5. CONCLUSIONS

The purpose of this work was to shown the potential associated to an important and, at the same time, more surprising theorem concerning non-linear mapping or filtering. Motivated and written by an engineer, the style of the paper has been speculative and positive, versus the relevance of the theorem and the extraordinary contribution of Andrei Kolmogorov. Probably, almost certain, that this work can be improved even in fundamental aspects of it. The search for the unknown transform of the first stage, even the special character of it, seems to be the main issue to come out with any efficient implementation of the theorem. In any case, and in favour of the importance of further work on the theorem, is important to remark the excellent results obtained, for the applications listed in section four, with an architecture that just, again in a very speculative manner, was trying to follow the guide-lines of the theorem.

5 REFERENCES

/1/ M. Schetzen. "The Volterra and Wiener theories of non-linear systems". Wiley, New York, 1980.

/2/ R. Hecht-Nielsen. "Neurocomputing". Addison Wesley, 1990, pp 122-124.

/3/ C.L.Nikias, J.M. Mendel, "Signal processing with high order spectra", IEEE Signal Processing Magazine, July 1993.

/4/ J.L. Lacoume Ed. ,"HOS", Proc. Signal Processing on High Order Statistics, Chamrouse, Francia, July 1991.

/5/ D.Kraus, D.Maiwald, J.F. Bhöme, "Maximum likelihood source location estimation via EM algorithm", Signal Processing VI, J. Vandewalle Ed., Eusipco 92, Brussels, Belgium,Elsevier 1992.

/6/ M. Feder, E. Weinstein. "Parameter estimation of superimposed signals using the EM algorithm" IEEE ASSP, Vol 36, no.4, pp 477-489, April 1988.

/7/ M.A.Lagunas et al., "The Kolmogorov Signal Processor". Lecture Notes in Computer Science, Vol 686. New Trends in Neural Computation. Springer Verlag. pp 494-512, June 1993.

/8/ A.N.Kolmogorov. "On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition". (in Russian), Dokl. Akad. Nauk USSR, 114, pp953-956, 1957.

/9/ C. Jutten, "Calcul neuromimetique et traitement du signal: Analyse en composants independents". Ph.D. Thesis, in French, INPG-USMG, Grenoble 1987.

/10/ C. Jutten, J. Herault. "Independent component analysis (INCA) versus principal component analysis",

Signal Processing IV, Eusipco 88, J.L. Lacoume Ed.,Elsevier, Grenoble, Francia 1988.

/11/ P. Common, C. Jutten, J. Herault. " Blind separation of sources Part II: Problem statement" Signal Processing, Vol. 24, no 1, pp. 11-21, July 1991.

/12/ Y. Bar-Ness et al.,"Cross-coupled phase-locked loop with closed loop amplitude control", IEEE Trans. Communications, Vol. COM-32, no. 2, pp 195-199, February 1984.

/13/ Y. Bar-Ness, J. Rokach, "Cross coupled bootstrapped interference canceller". IEEE AP Int. Symp., pp 292-295, 1981.

/14/J.S. Bendat. "Nonlinear systems analysis and identification from random data". Wiley. 1990.

/15/ R. Price. "A useful theorem for non-linear devices having Gaussian inputs". Nonlinear Systems, A.H. Haddad EJ. Dowden, Hutchinson & Ross, Stroudsburg, Pennsylvania. 1975.

/16/ J.J. Bussgang. "Cross-correlation functions of amplitude distorted Gaussian inputs". Nonlinear Systems, A.H. Haddad Ed. Dowden, Hutchinson & Ross, Stroudsburg, Pennsylvania. 1975.

/17/ G. Lorentz. "Approximation functions". pp 168-179, Chelsea Publishing Co., New York, 1986.

/18/ V. Kurkova. " Kolmogorov theorem and multilayers neural networks". Neural Networks, Vol. 5, pp 501-506, 1992.

/19/ V. Kurkova. "Kolmogorov's theorem is relevant". Neural Computation 3, pp 617-622, 1991, MIT.

/20/ G. Cybenko. "Approximation by superposition of a single function". Mathimatical Control, Signal and Systems 2, pp 303-314, 1989.

/21/ I. Bilinskis, A. Michlesons, "Randomised signal processing". Prentice Hall, 1992.

/22/ J.R.Casar,M.Garcia-Otero,"A class of adaptive non-linear filters", Proceedings MWPSA Workshop, paper A4.2, Ed. D. Aboutadjine, Marrakech, Morocco, September 24-26, 1984.

/23/ O.Agazzi, D.G. Messerschmitt, D.A. Hodges, "Nonlinear echo cancellation of data signals", Proc. IEEE Int. Conf. Comm., 1982, pp.7G6.1-7G6.5,1982.

/24/ G.L. Sicuranza, A. Bacconi, P. Mitri. "Adaptive echo cancellation with non-linear digital filters". Proc. IEEE-ICASSP, pp 3.10.1-3.10.4, San Diego, California, USA, 1984.

/25/ W.J.Rugh. "Nonlinear system theory", Johns Hopkins University Press, 1981.

/26/ M. Vidyasagar. "Non-linear systems analysis". Networks Series, R.W. Newcomb Ed., Prentice Hall, 1978.

/27/ A. Papoulis, "Probability, Random Variables, and Stochastic Processes", McGraw-Hill,1965, pp. 476-477.

/28/ M.A. Lagunas et al.,"The K-filter" Submitted for publication to the Eurasip Journal, Signal Processing.

/29/ A. Pages, M.A. Lagunas."Identification of non-linear systems". Procc. 7 SP Workshop on Statistical Signal and Array Processing, Quebec City, Canada June 26-29, 1994.

/30/ A. Pages, M.A.Lagunas," High order learning in TRB beamforming", Siganl Processing V, Proc. Eusipco 92, Brussells, pp , Belgium 1992.

/31/ A. I. Perez-Neira."Antenna array multiparameter / signal estimation based on a neural network mapping theorem ", Ph D Tesis, Univ. Pol. Catalunya, in spanish, Nov 1994.