



Signal Processing 44 (1995) 249–267

**SIGNAL
PROCESSING**

The K-filter: A new architecture to model and design non-linear systems from Kolmogorov's theorem[★]

A. Pagès-Zamora, M.A. Lagunas*, M. Nájar, A.I. Pérez-Neira

Dpt. T.S.C., E.T.S.E. Telecomunicació, c/ Gran Capità s/n, 08034 Barcelona, Spain

Received 4 February 1994; revised 26 January 1995

Abstract

A new architecture to model and design non-linear transfer functions is presented using a new formulation for non-linear systems. This approach follows the guidelines of the mapping theorem due to A. Kolmogorov and it is based on the direct Fourier transform of the transfer function. The resulting scheme is formed by two stages; the first stage contains phase modulators, which, based on random sampling concepts reported by I. Bilinskis, are duplicated with a small perturbation in the modulation factor. This stage depends on the number of diversity data and it is independent of the function. The second step reduces to Volterra systems and a direct combiner of the new diversity kernels. The reported architecture and design seem to be able to cope with both linear and non-linear filtering problems, which can be considered as a formal framework for generalised signal processing.

Zusammenfassung

Eine neue Architektur wird mit neuer Formulierung für nichtlineare Systeme vorgestellt, um nichtlineare Funktionen zu modellieren und zu entwerfen. Dieser Ansatz folgt den Richtlinien des Theorem von A. Kolmogorov und basiert auf direkter Transformation Fouriers von der Funktion des Transfers. Das resultierende Schema besteht aus zwei Etappen. Die Erste hat Phasenmodulationen, die mit einer kleinen Verzerrung im Modulationsfaktor dupliziert werden. Das basiert auf der Auffassung der aleatorischen Stichprobenerhebung von I. Bilinskis. Diese Etappe ist auf die Nummer der Datenverschiedenheit angewiesen, aber ist unabhängig von der Funktion. Die zweite Etappe beschränkt sich auf Volterras Systeme und auf eine direkte Kombination der neuen Kernelverschiedenheit. Das Ergebnis dieser zweiten Etappe ist dann ein Signal, so ähnlich das Signal des gemessenen Ausgang wie möglich. Außerdem scheint es zu sein, daß die verlegte Architektur und der Entwurf mit beiden linearen und nichtlinearen Filterungsproblemen benutzt werden können.

Résumé

Cette communication présente une nouvelle architecture servant à représenter et modéliser des fonctions de transfert pour des systèmes non-linéaires. Cette approche utilise les principales idées développées dans le théorème de A. Kolmogorov ainsi que les transformées de Fourier des fonctions de transfert. Le schéma résultant nécessite deux étapes.

[★] This work has been supported by the Spanish National Research Plan, CICYT, Grant Number TIC 92-0800-C05-05.

* Corresponding author.

La première utilise des modulations de phase où les concepts d'échantillonnage aléatoire décrit par I. Bilinski sont appliqués. Ces dernières sont reproduites avec des petites perturbations dans le facteur de modulation. Cette étape dépend des données de la diversité mais est indépendante de la fonction de transfert. La seconde étape consiste en un système de Volterra et un mélange direct des noyaux de la nouvelle variété. L'architecture et le modèle adoptés semblent s'appliquer tant aux systèmes linéaires qu'aux systèmes non-linéaires et peuvent donc être vus comme une possibilité de résolution générale des problèmes de traitement du signal.

Keywords: Wiener–Kolmogorov filtering; Non-linear systems; Volterra filters; Kolmogorov mapping theorem; Fourier model; K-filter; System identification; Equalisation

1. Introduction

Non-linear signal processing has emerged in the last years as a specific target for signal processing tools, being a direct consequence of the degree of saturation produced in the linear signal processing field during the 1980's. At the beginning, non-linear filtering and modelling were under the scope of the filtering theory and it can be said that the first non-linear kernels, for a generalised convolution integral, were reported at the same time that the filtering of Gaussian processes with linear filters and minimum square error (MSE) criteria were reported [28]. At the same time, three fields evolved, almost separately, finding an alternative to linear schemes in the three aspects which are relevant to signal processing, i.e. the architecture, the associated functions and the algorithms. At the architectural level, new schemes appeared under activities encompassed by neural networks [14]. Concerning associated functions, higher-order processing was developed and proven useful to opening up new signal processing applications [21, 25]. Finally, at the algorithmical level, the 'Estimate and Maximise' algorithm [11, 18] is probably the most valuable tool for the design and learning stage of many of these new alternatives.

In [22] the authors related, using the mapping theorem of Kolmogorov [17], the aforementioned efforts when the source separation problem [2, 3, 9, 15, 16] is faced along the main guidelines contained in the mapping theorem. Precisely, it is this theorem which motivates this work. Starting from the usual architecture and framework of non-linear filtering, a new architecture formed by two processing stages in agreement with the theorem, is intro-

duced. A closed formulation for the design of such an architecture is also presented. The objective is to find a procedure for non-linear filtering design from only short input and output records of measured data, without significant dependence on the input/output statistics and the specific transfer function, either linear or non-linear, which relates both records.

The so-called Volterra filter is briefly outlined in Section 2. Section 3 introduces a Fourier representation of non-linear transfer functions, which proves the efficiency associated with a previous transform of the Volterra processing. In Section 4 the mapping theorem is briefly described in order to support the new architecture. Later on, in Section 5, the new K-filter is presented, where K stands for the Kolmogorov's existence theorem that motivates this work. The architecture is described in detail, starting from the case of memoryless to generalised non-linear systems, which include temporal diversity of the input signal. Finally, in Section 6, some simulations are included in order to support the main points arising from the new proposed techniques. Although these simulations are not enough to assess the overall performance of the new filter, they justify the interest in the procedure and support the potential interest of future work in the reported architecture.

2. Volterra filters

The well-known Volterra modelling of non-linear systems [28] consists of a polynomial approximation of the system response. This polynomial is a function of many variables depending on both

the order and the diversity handled by the model. Thus, $(x(t), x(t - \tau_1), x(t - \tau_2), x(t - \tau_3), \dots, x(t - \tau_Q))$ is the data used by the model to form the so-called Volterra kernels, $(\tau_1, \tau_2, \dots, \tau_Q)$ are the diversity parameters, and the order is the maximum number of diversity components forming a kernel. Without loss of generality, this brief review will focus on the case of memoryless non-linear systems (i.e. no diversity and order M). In this case, the Volterra model for a given function $y = g(x)$ will be

$$y_a(t) = \text{Vol}[x(t), M] = \sum_{m=1}^M a_m x^m(t). \quad (1)$$

The computation of the coefficients a_m can be carried out with minimum squares fitting or with a minimax criteria which are very well known techniques from the classical approximation theory for filter design in the frequency domain.

The one aspect that is important to recall here is that most of the success of the resulting approximation comes from the dynamic ranges of the input/output signals selected to fit the model to the observed data. In any case, the presence of powers of the input signal in the model formulation (1) implies severe errors whenever the input signal evolves over the a priori selected dynamic range, assumed to be $[-X_{\max}, X_{\max}]$ (see Fig. 1). This is an important drawback since most of the systems of interest usually present saturation in their response, and the kernels used in the Volterra model are not adequate to represent this phenomena. It is

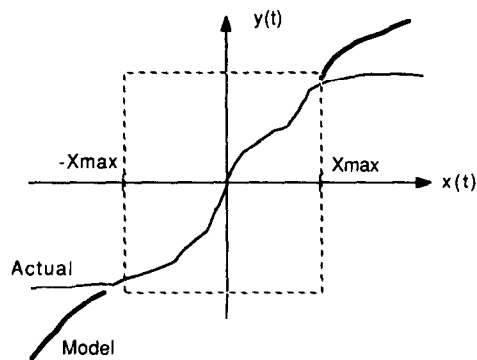


Fig. 1. Volterra modelling in saturated systems.

possible to extend the dynamic range of the modelling design above $[-X_{\max}, X_{\max}]$, but this will have a negative impact on the quality of the approximation around the maxima of the input probability density function (p.d.f.) (located usually nearby zero). Other models may be used to better fit the given function, but they will be at the expense of one of the most important features of the Volterra models, which is the design and learning; in other words, whatever the approximation is, it should be a linear combination of predefined kernels in order to preserve the MSE design and the linear learning.

One way out to overcome, in practice, the above-mentioned drawback is using some data transformation $T[x(t)]$, which will be also non-linear, followed by a redesigned Volterra filter. This architecture is depicted in Fig. 2 and it requires, in general, some a priori knowledge of the transfer function to be modelled. With this approach the linear learning remains in terms of the new input data $z(t)$. The rest of the paper is devoted somehow to the transformation that provides $z(t)$ and how it could be selected. It should be mentioned that many practical models reduce $T(\cdot)$ to a linear-saturated system which avoids the previously mentioned problems (Eq. (2)).

$$T[x(t)] = \begin{cases} x(t) & -X_{\max} < x(t) < X_{\max}, \\ X_{\max} \text{sign}[x(t)] & \text{else.} \end{cases} \quad (2)$$

The next section will reproduce a theorem which is highly related to the role and potential of the mentioned transforms and it is the source of all the new framework for system modelling reported hereafter.

Concerning the case of systems with memory, it is important to recall that the design procedures reported nowadays [4], are upper bounded by tri-linear systems as long as the input data are Gaussian distributed. Under these premises of tri-linear systems and Gaussian distributed input, the use of the cross-correlation of the input and the output of the non-linearity under modelling [4, 6, 26], allows a closed design of the corresponding model. The rest of the paper is devoted, based on the guidelines provided by Kolmogorov [17], to reporting a new

architecture able to cope with any kind of non-linearity, together with a design framework based directly on the input/output data.

3. The Kolmogorov mapping theorem

In 1957 Andrei Kolmogorov [17] solved the 13th problem of the collection that Hilbert provided as the mathematical problems for the 20th century. He proved the existence of a solution, based on functions of a single variable, of the problem of finding a continuous function mapping Q inputs into S outputs. If x is the input vector of Q components $x(q)$ ($q = 1, \dots, Q$), and y the corresponding output vector with S components $y(s)$ ($s = 1, \dots, S$), the existence theorem specifies that the processor which maps x in y is formed by two stages. The first stage has a mathematical representation given by (3), where λ and the functions $\Phi(\cdot)$ depend on the number of components Q of the input vector and are independent of the specific mapping function, and ε is an arbitrarily chosen value.

$$z_k(x) = \sum_{q=1}^Q \lambda^q \Phi[x(q) + k\varepsilon] + k, \quad k = 1, \dots, 2Q + 1. \quad (3)$$

The second stage contains S functions $\varphi_s(\cdot)$ that produce the s th-component of the output.

$$y(s) = \sum_{k=1}^{2Q+1} \varphi_s(z_k), \quad s = 1, \dots, S. \quad (4)$$

This brief description of the theorem has been obtained from Hecht and Nielsen [14]; further reading of the theorem can be obtained from Lorentz and Kurova [10, 19, 20, 23]. These references shed some light on the most astonishing part of the theorem, which is the independence of the first stage functions $\Phi(\cdot)$ of the specific mapping being implemented.

Although no practical implementation of the theorem has been reported yet, it suggests many important advances in signal processing just following the guidelines contained in it as it will be seen hereafter. It may be argued that the final model and filter, to be reported herein, keep some distance

from the formality of the theorem. In any case, it was really the motivation of the authors to start the searching of alternatives to the inherent limitations of Volterra modelling and Wiener–Kolmogorov (W–K) linear filter theory.

At this moment, it is important to note the similarity between the structure depicted in Fig. 2 and the two-stage scheme of the Kolmogorov’s theorem shown in Fig. 3, for the specific case of a single-input/single-output mapping function.

With independence of the second stage, our attention is focused now on the first stage, where a signal transformation takes place with only two parameters, and it seems to contain the most interesting part of the theorem. Related to this transformation, it is important to be aware of the robustness and the average performance of logarithmic companding for both analog and digital processing in audio noise reduction and coding systems. Furthermore, the practicality and robustness of these techniques suggest to the authors the potential of a logarithmic transformation in order to cover the role assigned by the theorem to the first processing stage. In consequence, the paper will concentrate on the use of logarithmic transformations or, to be more specific, in phase modulated (PM) versions of the input signal. This kind of transformations will allow the connection between

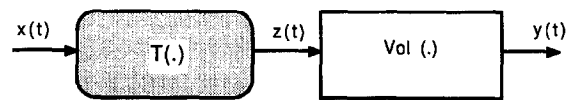


Fig. 2. Two-stage model with a pre-defined non-linear transformation followed by a memoryless Volterra system.

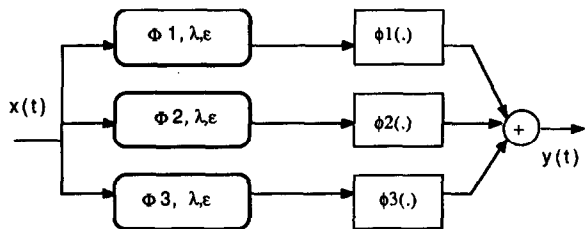


Fig. 3. The Kolmogorov architecture derived directly from the theorem for an instantaneous (memoryless) mapping of $x(t)$ in $y(t)$.

the K-filter and both, the two-stage scheme of Kolmogorov’s theorem and the Fourier model presented in short. A modelling and generalised filtering scheme following the guidelines of the existence theorem will be developed together with an explanation for the requirement of the $2Q + 1$ components. It will be shown that, as predicted, Kolmogorov’s mapping theorem represents a step ahead in the performance of classical Volterra models and Wiener–Kolmogorov linear filters. Starting from the so-called Fourier model for a memoryless non-linear transfer function, the Kolmogorov filter (K-filter) will be introduced and extended to the case of non-linear functions with memory.

4. The Fourier model

The signal $y(t)$, equal to the function $g[x(t)]$, can be represented in terms of the Fourier transform of $g(\cdot)$, denoted by $G(\omega)$ (Eq. (5)).

$$y(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} G(\omega) e^{j\omega x(t)} d\omega. \tag{5}$$

As long as $g(\cdot)$ is assumed periodic with a period $2X_0$ larger than the dynamic range $2X_{\max}$, the above integral can be represented by the Fourier series (Eq. (6a)), which by direct truncation leads to the approximation of Eq. (6b).

$$y(t) = \frac{1}{2X_0} \sum_{m=-\infty}^{+\infty} G(m\omega_0) e^{jm\omega_0 x(t)}, \quad \omega_0 \leq \pi/X_0, \tag{6a}$$

$$y(t) \approx y_p(t) = \frac{1}{2X_0} \sum_{m=-M}^{+M} G(m\omega_0) e^{jm\omega_0 x(t)}, \quad \omega_0 \leq \pi/X_0. \tag{6b}$$

The coefficients of the Fourier series development, $G(m\omega_0)$, correspond to the Fourier transform of $g(\cdot)$ evaluated at multiples of the so-called principal frequency, ω_0 . The problem, motivated by the sampling of the continuous Fourier model, will be commented later.

Taking into account the usual odd character of the mappings of interest in practical applications, the linear combination of exponential functions (Eq. (6b)) evolves to a linear combination of sine

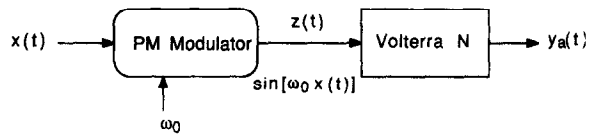


Fig. 4. A phase modulator (PM) followed by a Volterra filter to model any memoryless mapping of two signals.

functions (Eq. (7a)). Thus, the PM modulator proposed to implement the transformation of the first stage of Fig. 1 leads to a real univariate function, as it is specified in the theorem. Furthermore, it can be easily shown that such a sine development allows a summation of infinite odd powers of the modulated signal. Assuming a truncated version of it, expression (7a) can be approximated by Eq. (7b).

$$g(x) = -g(-x) \Rightarrow$$

$$y_p(t) = \frac{1}{X_0} \sum_{m=1}^{+M} |G(m\omega_0)| \sin[m\omega_0 x(t)], \tag{7a}$$

$$y_p(t) \approx y_a(t) = \sum_{n=0}^{+N} a_{2n+1} \sin^{2n+1}[\omega_0 x(t)]. \tag{7b}$$

To state clearly the similarity of the Fourier model with the Kolmogorov structure, Fig. 4 shows the model described in this section (Eq. (7b)). This model avoids the problem of the input signal dynamics and it can be encompassed as one branch of the Kolmogorov architecture, which somehow supports the fitness of the pre-processing transform to any Volterra model. Some questions remain on the additional branches claimed by the existence theorem. This point will be explained in the next section, as it is related to the frequency sampling approximation when passing from the continuous formulation (general mapping definition) to the Fourier series.

5. The K-filter

5.1. The memoryless K-filter

The evolution of the Fourier model to the Kolmogorov model (K-filter) stems from the requirement for additional branches, in order to better fit any mapping function. The reason for a second branch or path in the processor comes

from the potential existence of a non-zero to zero correspondence in the function to be modelled. The term a_0 generalises the mapping as it can be seen in Fig. 5.

$$y_a(t) = a_0 + \sum_{n=0}^{+N} a_{2n+1} \sin^{2n+1} [\omega_0 x(t)]. \quad (8)$$

The problem with Eq. (8) is that it is still an approximation as the periodic structure of the model avoids the convergence to the actual mapping, retaining a few samples of the actual Fourier transform $G(\omega)$.

An approach, reported by Bilinskis [5] some years ago concerning random sampling, inspired the solution of the sampling problem and the inclusion of the third branch claimed in the theorem. Let us assume that a third branch with a slightly different modulation frequency ($\omega_0 + \varepsilon$) is set. The new frequency is set in such a manner that the coincidence between harmonics of both frequencies is reduced as much as possible. With these two frequencies we may say that the frequency sampling is produced almost at random in $G(\omega)$. In fact this is the important objective of the two branches and the perturbation parameter ε , i.e. to produce a random sampling of the continuous function to be modelled. It can be said that the Volterra model in every branch, which generates the corresponding frequency scanning, can be modified in order to produce the random effect of the frequency sampling better. The answer is positive but the resulting complexity increases. Thus, assuming that in the band of interest, the random effect takes place, the

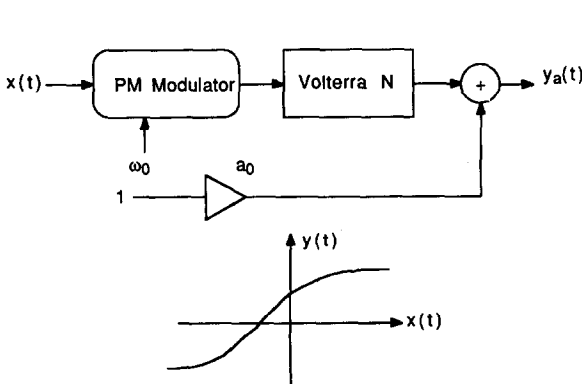


Fig. 5. Two path for generalised instantaneous x - y mapping.

corresponding model can be formulated as

$$y_k(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} G(\omega) \left[\sum_{i=-\infty}^{+\infty} \delta(\omega - \omega_i) \right] e^{j\omega x(t)} d\omega. \quad (9)$$

If $P_{\Omega_i}(\omega_i)$ is the p.d.f. of sample ω_i , then the expected value of the model will be

$$\begin{aligned} E[y_k(t)] &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} G(\omega) \left\{ \sum_{i=-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} \delta(\omega - \omega_i) \right. \right. \\ &\quad \left. \left. \times P_{\Omega_i}(\omega_i) d\omega_i \right] \right\} e^{j\omega x(t)} d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} G(\omega) \left\{ \sum_{i=-\infty}^{+\infty} P_{\Omega_i}(\omega) \right\} e^{j\omega x(t)} d\omega, \end{aligned} \quad (10)$$

revealing that, as long as the superposition of the p.d.f. of the sampling frequency is constant, the expected value of the model will coincide with the actual transfer function of $g[x(t)]$. That is, the proposed model implements an unbiased estimator (Eq. (11)).

If

$$\sum_{i=-\infty}^{+\infty} P_{\Omega_i}(\omega) = ct \Rightarrow E[y_k(t)] = y(t). \quad (11)$$

The resulting architecture for a finite number of samples (i.e. ω_0 not too small and Volterra systems of finite order) is shown in Fig. 6. Note that a quality improvement on the resulting approximation

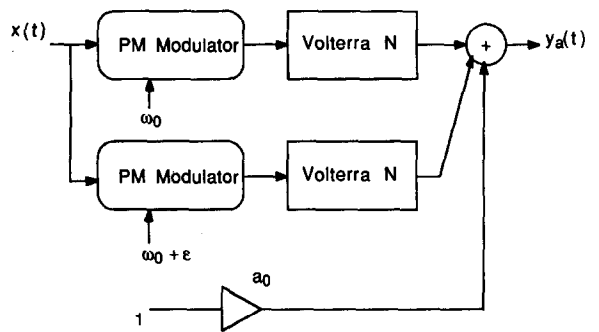


Fig. 6. The K-filter for a memoryless non-linear function $y(t) = g(x(t))$ (top), and the extension of the basic function to the coefficients computation.

can be obtained by decreasing ω_0 , at the expense of increasing the order of the Volterra systems. It is also important to remark that the learning stage of the K-filter can be performed using gradient techniques (least-mean squares (LMS), normalised-LMS) in the same manner that they are applied to the classic Volterra models. With respect to the PM modulators, note that they contribute, better than the powers of the input signal, to enlarge the bandwidth of the new frequencies not contained in the input signal power spectral density. The richness of the outband frequency content of the output $y(t)$ is associated with most of non-linear devices and systems used in practice. In terms of convergence it is expected to improve with respect to single Volterra models, since the independence between PM modulators of different modulating frequencies of the input signal is, in general, better than using powers of it. An important issue that justifies the superior performance in terms of adaptive convergence of the K-filter versus Volterra models is that the fixed dynamic range of the PM diversity (i.e. $\sin(n\omega_0 x(t))$), generated in the first stage, allows an accurate setting of the μ parameter in an LMS algorithm. In this sense note that the power diversity in the Volterra case (i.e. $x^n(t)$) poses a problem for the step size setting, since the dynamic range changes dramatically from low n to the approximation order (high n).

Summarising the previous paragraphs, note that the phase modulators produce a new diversity set from the original one. The perturbation parameter, ε in the theorem, together with the modulation frequency are the basic parameters for the frequency scanning of the corresponding Fourier Transform $G(\omega)$. Note that every Volterra system contains a finite series of smooth functions somehow reminiscent of the functions described in the previous section. The weights of the series involved in the Volterra systems are function dependent, as predicted by the theorem. In any case, it is worthwhile to remark that the Volterra systems represent a specific choice for the task assigned which is to scan $G(\omega)$ properly, in order to get accuracy in the resulting implementation of function $g(\cdot)$. This is the synthesis of the existence theorem in our approach: first step: generate transform kernels; second step: scan the transform domain.

5.2. The K-filter with memory

To generate the K-filter to the case of non-linear functions with memory and without loss of generality, we will set the corresponding K-filter for a function $g(\cdot, \cdot)$ of two values of the signal input, $x(t)$ and $x(t - \tau)$. Note that the system in (12) manages input data diversity, which is associated with any filtering scheme.

$$y(t) = g[x(t), x(t - \tau)]. \tag{12}$$

The two-dimensional inverse Fourier transform which leads to $y(t)$ is as follows.

$$y(t) = \frac{1}{(2\pi)^2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} G(\omega_1, \omega_2) \times e^{j[\omega_1 x(t) + \omega_2 x(t - \tau)]} d\omega_1 d\omega_2. \tag{13}$$

At this point, it is worthwhile to remark that the crucial issue when designing the K-filter is the way in which the frequency plane is scanned, which is associated with the random assumption. As the frequency plane is now two-dimensional, the exploration performed by the corresponding Volterra-like system has to cover the domain of both frequencies simultaneously. Nevertheless, before discussing how the randomised sampling can be generalised to a two-dimensional plane, it is necessary to obtain a formulation similar to the linear combination of powers of sine functions (Eq. (7b)) but in the two-dimensional case. In order to achieve this, a truncated version of a two-dimensional Fourier series development is proposed in Eq. (14), as it was done in the case of memoryless systems.

$$y_p(t) = \frac{1}{(2X_0)^2} \sum_{m_1=-M}^M \sum_{m_2=-M}^M G(m_1\omega_0, m_2\omega_0) \times e^{j[m_1\omega_0 x(t) + m_2\omega_0 x(t - \tau)]}. \tag{14}$$

The main frequency is chosen as $\omega_0 = \pi/X_0$ where $2X_0$ should be larger than the input range $2X_{\max}$. Then, assuming that the mapping of interest will have odd symmetry with respect to the origin, the Fourier series can be arranged in

a summation of sine functions.

$$y_p(t) = \frac{2}{(2X_0)^2} \sum_{m_1=0}^M \sum_{m_2=0}^M |G(m_1\omega_0, m_2\omega_0)| \times \sin [m_1\omega_0x(t) + m_2\omega_0x(t - \tau)]. \quad (15)$$

Note that it is not possible to arrange the terms of expression (15) in powers of sine functions, although it is required to follow the same steps as in the memoryless case. In consequence, we propose the following development in order to achieve an approximation of Eq. (15) in powers of sine functions. Although such a development only leads us to an approximation and it also seems to be quite heuristic, it is important since it will allow the connection between the K-filter and the Kolmogorov theorem. Thus, Eq. (15) results in expression (16) when it is written in terms of two new functions denoted by $f_1(m)$ and $f_2(m)$. These functions play the same role as the instantaneous frequency in a phase modulation scheme, with index equal to the input diversity components (Fig. 7). On the other hand, the new coefficients a_m can be directly related to the samples of the two-dimensional Fourier transform, $G(m_1\omega_0, m_2\omega_0)$.

$$y_p(t) = \sum_{m=0}^{(M+1)^2-1} a_m \sin [f_1(m)x(t) + f_2(m)x(t - \tau)]. \quad (16)$$

The approximation we propose consists of substituting the functions $f_1(m)$ and $f_2(m)$ by a first-order approximation which corresponds, let us say,

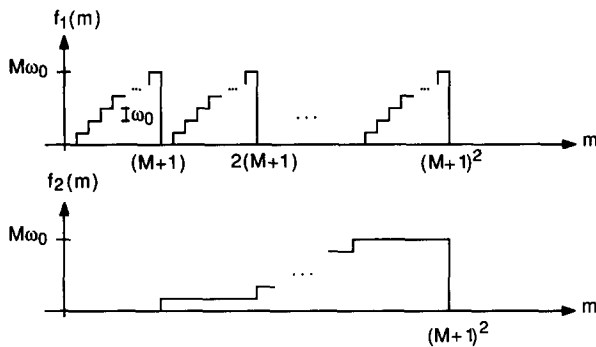


Fig. 7. Representation of functions $f_1(m)$ and $f_2(m)$ versus m .

with their unwrapped versions.

$$f_1(m) \cong m\omega_0, \quad (17a)$$

$$f_2(m) \cong \frac{m}{M}\omega_0. \quad (17b)$$

Thus, if only these frequencies and their harmonics are kept, Eq. (16) results in Eq. (18a) that can be already arranged in a truncated linear combination of powers of sine functions (Eq. (18b)). It is assumed that the number of terms in Eq. (18b) is much less than in Eq. (18a), with the new coefficients c_{2n+1} different from the former ones a_m , although they are related.

$$y_p(t) = \sum_{m=0}^{(M+1)^2-1} a_m \sin \left[m\omega_0x(t) + \frac{m}{M}\omega_0x(t - \tau) \right], \quad (18a)$$

$$y_p(t) \approx y_a(t) = \sum_{n=0}^N c_{2n+1} \sin^{2n+1} \left[\omega_0x(t) + \frac{\omega_0}{M}x(t - \tau) \right], \quad (18b)$$

Although it will be justified in the next section, note that if in expression (16) the function $f_1(m)$ had been applied to $x(t - \tau)$ and $f_2(m)$ to $x(t)$, expression (18b) would have resulted in the following one:

$$y_p(t) \approx y_a(t) = \sum_{n=0}^N d_{2n+1} \sin^{2n+1} \left[\frac{\omega_0}{M}x(t) + \omega_0x(t - \tau) \right]. \quad (19)$$

It is important to remark that, forming the instantaneous phase, the phase modulation and the corresponding Volterra system involve real-valued univariate functions as predicted by the theorem. On the other hand, taking into account the approximation involved in (18), which gives priority to the scanning associated with $x(t - \tau)$ with a high modulation index ω_0/M , it is clear that in order to balance the role of every diversity input in the scanning procedure, expression (19) should be included in addition to (18) in the filter architecture. This argument, which justifies a second branch in the architecture, can be considered heuristic, but it reveals that even under strong approximations to

the first stage of Kolmogorov’s architecture, the number of outputs supplying the second stage remains unchanged. In summary, the first stage produces the following formulation for the system under consideration:

$$y_p(t) = \sum_{n=0}^N \left\{ c_{2n+1} \sin^{2n+1} \left[\omega_0 x(t) + \frac{\omega_0}{M} x(t - \tau) \right] + d_{2n+1} \sin^{2n+1} \left[\frac{\omega_0}{M} x(t) + \omega_0 x(t - \tau) \right] \right\}. \quad (20)$$

Once the two-dimensional Fourier series development has been arranged as a linear combination of powers of sine functions, we should face the problem of how the randomised sampling, mentioned in the previous Section 5.1, can be generalised to the two-dimensional case. The scanning of the 2D frequency plane that we propose can be done by setting two different PM modulators for every diversity component with frequencies equal to ω_0 and $(\omega_0 + \varepsilon)$. The diversity components of the input signal are combined before being modulated and applied to the Volterra systems. Thus, the approximation of function (12) that the K-filter will perform is summarised as follows:

$$y_{\text{out}}(t) = z_0 + \sum_{k=1}^4 z_k(t), \quad (21a)$$

$$z_1(t) = \sum_{n=0}^N c_{0,2n+1} \sin^{2n+1} \left[\omega_0 x(t) + \frac{\omega_0}{M} x(t - \tau) \right], \quad (21b)$$

$$z_2(t) = \sum_{n=0}^N c_{\varepsilon,2n+1} \times \sin^{2n+1} \left[(\omega_0 + \varepsilon)x(t) + \frac{(\omega_0 + \varepsilon)}{M} x(t - \tau) \right], \quad (21c)$$

$$z_3(t) = \sum_{n=0}^N d_{0,2n+1} \sin^{2n+1} \left[\frac{\omega_0}{M} x(t) + \omega_0 x(t - \tau) \right], \quad (21d)$$

$$z_4(t) = \sum_{n=0}^N d_{\varepsilon,2n+1} \times \sin^{2n+1} \left[\frac{(\omega_0 + \varepsilon)}{M} x(t) + (\omega_0 + \varepsilon)x(t - \tau) \right]. \quad (21e)$$

This processing produces, together with the perturbation, the effect of an almost random sampling in the 2D frequency plane. The formulation in Eqs. (21) provides the guidelines for higher-order systems; these guidelines can be summarised in three steps. The first step contains the combination of the diversity components including the perturbed version ($2Q$ components plus a dc level). The second step applies the modulators preserving the number of variables, $2Q + 1$. Note that these two steps do not depend on the specific mapping to be performed. The third step is formed by Volterra systems acting on the diversity combinations and they depend on the specific function being modelled.

5.3. The formulation of the K-filter for the general case

In general, a filter maps a given data signal $x(t)$ into a signal $y(t)$, which has to be as close as possible to a reference $d(t)$ in accordance with some distance criterion. The diversity used in the filtering depends on the application under analysis. In general, equally spaced time diversity is the most general (i.e. the diversity vector x_t formed by $x(t - q\tau)$ for $q = 0, \dots, Q - 1$). Also spatial, or mixed spatial-temporal, diversity is used in filtering (beam-forming) for narrowband or wideband array processing. Without loss of generality we will analyse the case of uniform time diversity since it is the most common use of W-K linear filtering theory.

The signal $y(t)$ can be viewed as a multivariate function of the diversity components (Eq. (22a)) and can be also expressed in terms of its corresponding Fourier modelling (Eq. (22b)).

$$y(t) = g[x(t), x(t - \tau), \dots, x(t - (Q - 1)\tau)], \quad (22a)$$

$$y(t) = \frac{1}{(2\pi)^Q} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} G(\omega_1, \omega_2, \dots, \omega_Q) \times e^{j[\omega_1 x(t) + \omega_2 x(t - \tau) + \dots + \omega_Q x(t - (Q - 1)\tau)]} \times d\omega_1 d\omega_2 \dots d\omega_Q. \quad (22b)$$

In line with what we proposed previously, the first stage combines the Q diversity components in

such a manner that it generates Q outputs. Following the steps already taken in the K-filter in Section 5.3, the following combinations are proposed:

$$z_1(t) = \sin \left[\omega_0 x(t) + \frac{\omega_0}{M} x(t - \tau) + \dots + \frac{\omega_0}{M^{Q-1}} x(t - (Q - 1)\tau) \right], \quad (23a)$$

$$z_2(t) = \sin \left[\frac{\omega_0}{M} x(t) + \frac{\omega_0}{M^2} x(t - \tau) + \dots + \omega_0 x(t - (Q - 1)\tau) \right], \quad (23b)$$

...

$$z_Q(t) = \sin \left[\frac{\omega_0}{M^{Q-1}} x(t) + \omega_0 x(t - \tau) + \dots + \frac{\omega_0}{M^{Q-2}} x(t - (Q - 1)\tau) \right]. \quad (23c)$$

The argument of the sine functions correspond to the first-order approximation of the instantaneous frequencies, obtained in the same way as it was done for the K-filter in the two-dimensional case. On the other hand, since the Kolmogorov theorem claims the existence of $(2Q + 1)$ variables, we propose to add a perturbed version of each $z_k(t)$, $k = 1, \dots, Q$, with a different principal frequency (i.e. substituting ω_0 by $\omega_0 + \varepsilon$). This processing produces the effect of an almost random sampling in the Q -dimensional frequency hyperplane in the same way as it was proposed in the two-dimensional K-filter. Thus, the perturbed Q functions are chosen as follows:

$$z_{Q+k}(t) = z_k(t)|_{\omega_0 = \omega_0 + \varepsilon}, \quad k = 1, \dots, Q. \quad (24)$$

Then, these $2Q$ functions $z_k(t)$ ($k = 1, \dots, 2Q$), are applied to a Volterra memoryless system of order N . Thus, the output of the Volterra system whose input is $z_k(t)$ is written in powers of the input:

$$\gamma_k(t) = \sum_{n=1}^N a_{k,2n+1} z_k^{2n+1}(t), \quad k = 1, \dots, 2Q. \quad (25)$$

Finally, the output consists of a linear combination of these $2Q$ functions together with the

dc component denoted by γ_0 (Eq. (26)).

$$y_{out}(t) = \gamma_0 + \sum_{k=1}^{2Q} \gamma_k(t). \quad (26)$$

5.4. The architecture

The formulation provided from Eqs. (22)–(26) can be summarised into three steps as it is shown in Fig. 8. The first step combines the Q diversity components generating $(2Q + 1)$ variables: the first Q ones correspond to different linear combinations of the instantaneous phases, the second group of also Q variables is obtained from perturbing the previous ones and finally the dc level. Then, the second step includes the PM modulators implemented by sine functions that are applied to the first $2Q$ variables (Eqs. (23) and (24)). Finally, the third step is formed by the Volterra systems which act on the modulated signals (Eq. (25)) and the

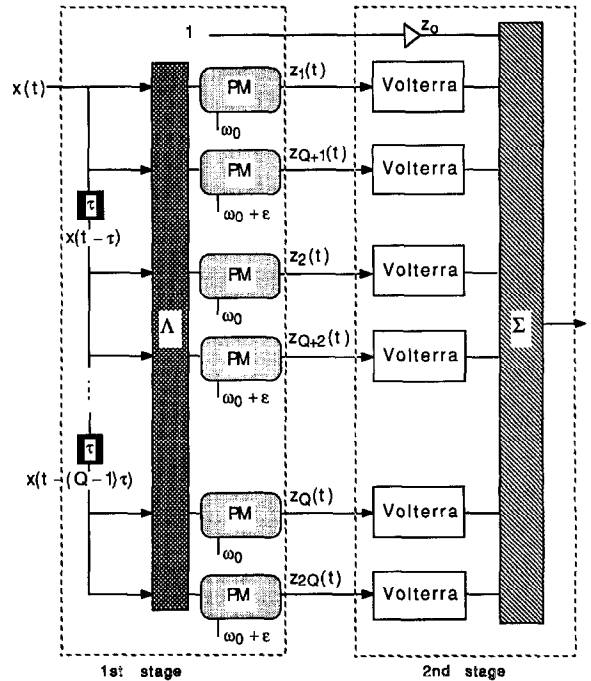


Fig. 8. The two-stage architecture of the K-filter showing from left to right the diversity, the diversity combiner, the modulation and the Volterra modules.

subsequent linear combination to obtain the output (Eq. (26)).

The choice presented in (26) has been used successfully in the simulations, but the overall rule is the independence of the Volterra kernels. The problem encountered when selecting the modulation frequencies is in accordance with the mapping theorem, since the transforms do not depend on the mapping, but their design depends on the number of diversity data managed by the filter. If the order of M and Q is increased, the adequate setting for the modulation frequencies and perturbations in order to guarantee, as much as possible, the independence of the kernels, will be more difficult.

The first stage of processing, which will not depend on the specific mapping, is constrained to be a linear combination of the diversity components (Eq. (3)). This stage combines the diversity set in order to allow a multifrequency scan with functions of a single variable. This is achieved by a matrix A (with inputs λ_{kp}) that combines the diversity components.

$$z_k(t) = \sin \left[\omega_0 \sum_{p=0}^{Q-1} \lambda_{kp} x(t - p\tau) \right], \quad k = 1, \dots, Q, \tag{27.1}$$

$$z_k(t) = \sin \left[(\omega_0 + \varepsilon) \sum_{p=0}^{Q-1} \lambda_{kp} x(t - p\tau) \right],$$

$$k = Q + 1, \dots, 2Q, \tag{27.2}$$

Parameters λ_{kp} are given by

$$\lambda_{k,0} = 1, \quad k = 1, \dots, 2Q, \tag{28.1}$$

$$\lambda_{k,p} = \begin{cases} \frac{\lambda_{k,p-1}}{M} & \text{if } \lambda_{k,p-1} < \frac{1}{M^{Q-1}}, \\ 1 & \text{if } \lambda_{k,p-1} = \frac{1}{M^{Q-1}}, \end{cases}$$

$$k = 1, \dots, 2Q, \quad p = 1, \dots, Q - 1. \tag{28.2}$$

The outputs γ_k of the Volterra systems with input z_k for $k = 1, \dots, 2Q$ are added together to the dc component γ_0 to produce the filter output as it is shown in (25).

When, for the sake of generality, we include a weighting c_k in the addition of the Volterra out-

puts, we have

$$y_{\text{out}}(t) = \sum_{k=0}^{2Q} c_k \gamma_k(t). \tag{29}$$

This structure can be seen as the cascade of a non-linear system for each combined diversity component with an optimum linear combiner (generalised convolution). It should be recognised as a step ahead in the generalised convolution theorem having a strong similarity with the so-called Wiener and Hammerstein processors [7] that were applied in non-linear echo cancellation problems in the 1980's [1, 30]. The architecture represented by the K-filter is more sophisticated since it introduces the Hammerstein architecture at the diversity components level, instead of cascading a single linear filter with Volterra systems; this is, to the author's knowledge, the first time that this architecture is proposed [4, 27, 31] in a formal framework design. It should be recognised that the so-called functional link equalisers share the same basic principle [8, 12]; the performance of these architectures versus conventional Hammerstein structures in digital communications equalisers for highly dispersing channels is remarkable. Also the interest of orthogonal functions, connected with the Wiener functionals [24] in non-linear filtering, has been shown in [29] in a very elegant manner.

Before formalising the filter design, it is important to remark that the architecture proposed in Fig. 8 perfectly agrees with the mapping existence theorem of Kolmogorov. First of all, it has been justified that $2Q + 1$ transformed diversity inputs are needed in order to generate an almost randomised sampling in the frequency domain providing an unbiased estimator. Furthermore, as it was claimed by Kolmogorov, the first two steps are independent of the specific mapping that the filter tries to emulate, whereas the second step already depends on it since it includes the coefficients $a_{n,l}$ related to the coefficients of the Q -dimensional Fourier series of $g(\mathbf{x})$. In fact, it is worthwhile to remark that the mentioned architecture emulates perfectly the linear W-K filtering for linear system identification problems, and it surpasses the W-K performance when non-linear systems are embedded in the model of the measured signal.

5.5. The design of the K-filter

In order to formalise the filter design, a diversity vector with $(1 + 2QM)$ components that contains the phase modulated diversity inputs is introduced and denoted by \mathbf{u} (Eq. (30)).

$$\mathbf{u}_t = [1 \ u_{(1,1)} \ u_{(1,2)} \ \dots \ u_{(1,M)} \ u_{(2,1)} \ u_{(2,2)} \ \dots \ u_{(2,M)} \ \dots \ u_{(2Q,1)} \ u_{(2Q,2)} \ \dots \ u_{(2Q,M)}]^\top. \quad (30)$$

Where the components of vector \mathbf{u} are equal to expressions (31) in accordance with (27).

$$u_{(k,l)}(t) = \sin \left[1 \cdot \omega_0 \sum_{p=0}^{Q-1} \lambda_{kp} x(t - p\tau) \right],$$

$$k = 1, \dots, Q, \quad (31.1)$$

$$u_{(k,l)}(t) = \sin \left[1 \cdot (\omega_0 + \varepsilon) \sum_{p=0}^{Q-1} \lambda_{kp} x(t - p\tau) \right],$$

$$k = Q + 1, \dots, 2Q. \quad (31.2)$$

Thus, the output of the K-filter can be expressed in a vectorial way as in Eq. (32) where \mathbf{w} denotes the vector of coefficients.

$$y_{\text{out}}(t) = \mathbf{w}^\top \mathbf{u}_t. \quad (32)$$

A global MSE criterion for the design of the coefficients of the Volterra system is performed from the snapshots \mathbf{u}_t . If \mathbf{C} is the corresponding covariance matrix and \mathbf{p} the cross-correlation of the mentioned snapshots with the reference signal $d(t)$, the optimum weights for each Volterra kernel are given by

$$\mathbf{w} = \mathbf{C}^{-1} \mathbf{p}. \quad (33)$$

The matrix and vector involved are computed as follows:

$$\mathbf{C} = E[\mathbf{u}_t \mathbf{u}_t^\top], \quad (34)$$

$$\mathbf{p} = E[\mathbf{u}_t d(t)], \quad (35)$$

where $E(\cdot)$ denotes the expected value, or average values in the implementation of the filter.

It should be reminded that adequate selection of the modulation and perturbed frequencies (ω_0 and $\omega_0 + \varepsilon$) will have a considerable impact on the condition number of the new diversity covariance data matrix \mathbf{C} used in the design of the K-filter. In other words, there is no general rule to select the

modulation frequencies, as well as the perturbations, but an MSE criteria to find the coefficients of the Volterra models gives priority to a good condition number of the covariance matrix.

The design procedure mentioned here is based on a pure MSE fit for the combined and PM modulated diversity components. Although other approaches may be suggested for finding the optimum weights \mathbf{w} , the main goal of this work is to prove the effectiveness of the new diversity data, formed by the combiner plus the modulators, to compete with existing alternatives to model both linear or non-linear transfer functions. The potential of the new architecture designed with different objectives other than the MSE is out of the scope of this paper.

6. Simulations

It is difficult to provide a global proof of performance and evaluation of the presented filtering architecture. Nevertheless, three examples have been selected to support the potential and interest of the signal-processing architectures reported in this work. The first one is the identification of a memoryless non-linear system by means of the memoryless K-filter. The second one corresponds to a filtering problem of a non-linear system with memory (Fig. 9). Finally, the equalisation of a digital non-minimum phase channel is included (Fig. 10).

For the memoryless K-filter, the approximation of a non-linear memoryless transformation with

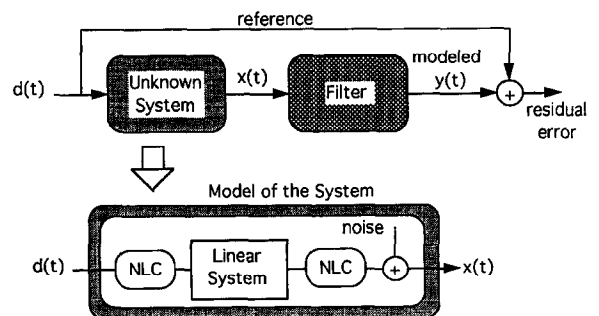


Fig. 9. The filtering problem and model used in the simulations in Fig. 12 and Tables 1 and 2.

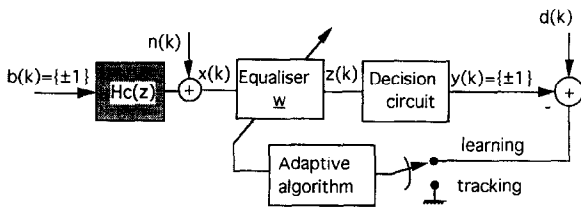


Fig. 10. The equalisation problem of a digital non-minimum phase channel.

a transfer function between input and output data sequences given by (36) has been selected. Figs. 11(a) and 11(b) show the actual transfer function in solid line assuming that the input signal is limited to the range $[-1, 1]$.

$$g(x) = [1 - e^{-x^2/0.05}] \text{sign}(x),$$

$$x \in [-X_{\max}, X_{\max}]. \quad (36)$$

Two different approximations have been considered in order to compare the results: first, the approximation achieved by a memoryless Volterra filter (polynomial model) of order 20 and second, the approximation obtained by a memoryless K-filter with two PM modulators and ten coefficients each one. These transfer functions, respectively shown in Figs. 11(a) and 11(b), have been achieved after a training period of adaptive computation in which the coefficients of the respective model are updated by means of the standard LMS algorithm with a step-size parameter equal to $\mu = 0.1$. The input sequence is a realisation of 512 samples of a uniform distributed random process with the same dynamic range of the non-linearity being modelled, that is, $[-1, 1]$. For the K-filter, a modulation frequency of $\omega_0 = \pi/2X_{\max}$ and a perturbation parameter equal to $\varepsilon = \omega_0/10$ have been selected.

Note that with the same computational load (both filters have the same number of coefficients) the approximation obtained by the K-filter is better than the one achieved by the Volterra one. This can be realised by observing the behaviour nearby the (0, 0) point and the flatness of the K-filter response for input dynamics above 0.5. The differences between the approximations achieved by the K-filter (Fig. 11(b)) and by the Volterra filter (Fig. 11(a)) are basically due to the cross-correlation among the Volterra kernels in the Volterra model, and also to

the difficulties associated with the optimum setting of the step-size parameter of the LMS algorithm for the Volterra kernels. Thus, since all the kernels of the K-filter have the same dynamic range and they are less correlated than in the Volterra model, the K-filter represents a better approach than the Volterra model. On the other hand, it is important to remark that this superiority of the K-filter has been tested for different orders and non-linearities.

In order to test the K-filter with memory, the authors have selected the inversion problem (Fig. 9) of a non-linear system with memory that emulates a communication channel with non-linear amplifiers at both sides, at the transmitter and at the receiver.

The reference signal is obtained by passing 256 samples of a normal distributed noise through a filter with the following transfer function: $H_r(z) = (1 + z^{-1})/(1 - 0.98z^{-1})$. Thus, the resulting signal is a low pass normally distributed noise. Then, this signal is driven into a non-linear device with an input/output transfer function denoted by $y = g_1(x) = x + 0.6\text{sign}(x)x^2$, which emulates the non-linear behaviour of the amplifier located at the transmitter. Next, the signal supplies a linear system with a transfer function given by $H_c(z) = (0.5 + 2z^{-1})$ that represents the communication channel. Following this linear system, a second non-linearity is set with a transfer function equal to $y = g_2(x) = x + 0.8\text{sign}(x)x^2 + 0.4x^3$ in order to include the non-linear amplifier at the receiver. Finally, a -35 dB additive Gaussian noise was added at the output. In summary, the K-filter has to reproduce a low pass reference signal that passes the cascade of the first non-linearity (transmitter amplifier), the band-pass linear system (communication channel), the second non-linearity (receiver amplifier) and the additive Gaussian noise, only by means of the output signal of the whole system.

The filters under test are a linear Wiener filter of order 10 (FIR filter), a K-filter (10, 1) (a temporal diversity vector of dimension 10, i.e. $Q = 10$ (Eq. (30)) and a Volterra filter of order 1, i.e. $M = 1$ (Eq. (30)) and another K-filter (10, 5). It is important to remark that these K-filters have two branches for each diversity component (Fig. 8) and that the principal frequency and the perturbation parameter have been chosen as before, i.e.

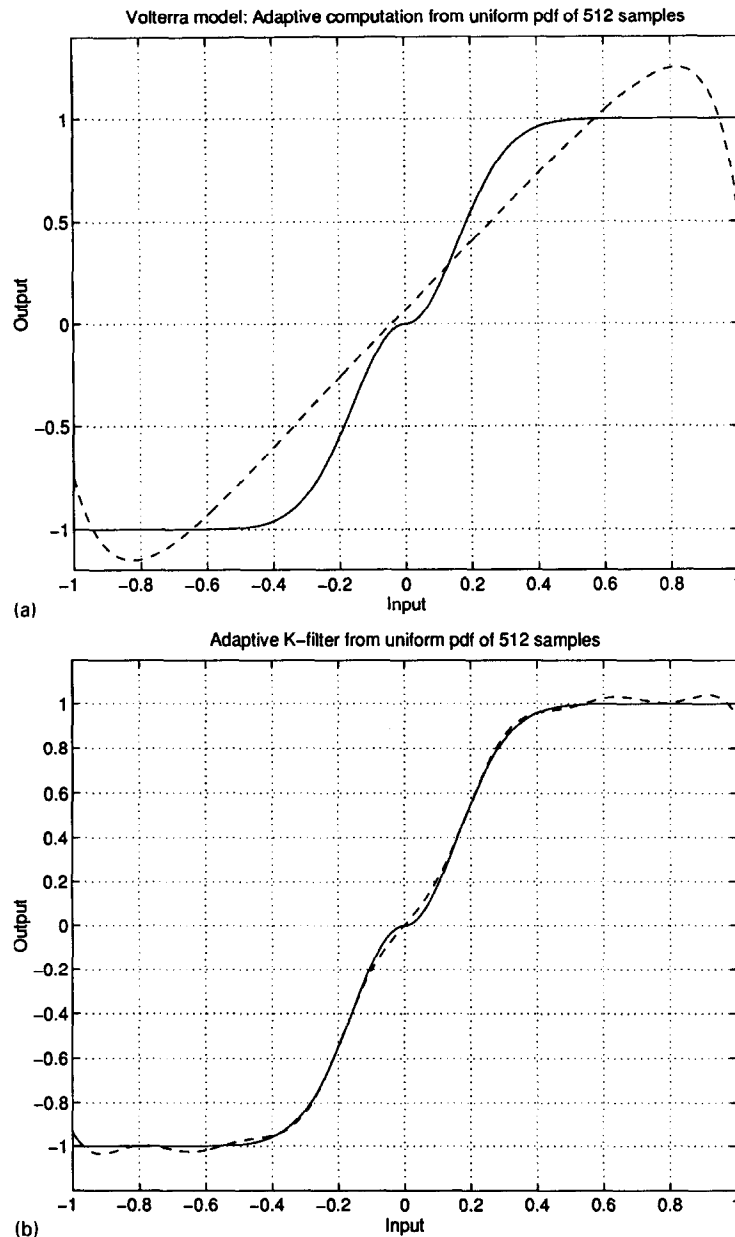


Fig. 11. Adaptive (LMS) computation of the transfer function from 512 samples of the output data with a uniformly distributed random input. (a) Actual transfer function (solid line). Approximated transfer function by a memoryless Volterra filter (polynomial filter) of order 20 with $\mu = 0.1$ (dashed line). (b) Actual transfer function (solid line). Approximated transfer function by a memoryless K-filter ($\omega_0 = \pi/2X_{\max}$, $\varepsilon = \omega_0/10$) of order 10, i.e. $M = 10$ in Eq. (7a) with $\mu = 0.1$ (dashed line).

$\omega_0 = \pi/2X_{\max}$ and $\varepsilon = \omega_0/10$. Figs. 12(a)–(c) show the corresponding plots where the actual reference can be compared with the output of each proposed filter. A visual inspection reveals the superiority of

the K-filter. The mean residual power normalised at the power of the reference of 30 independent realisations for different data record lengths are in Table 1. The values in brackets correspond to the

standard deviation (σ) with respect to the mean value.

It is also important to remark that we cannot remove the perturbation branches since, for

example, a K-filter with 10-perturbed branches (20 PM modulators) performs much better than another K-filter with 20 diversity components but with a single PM modulator per branch or even

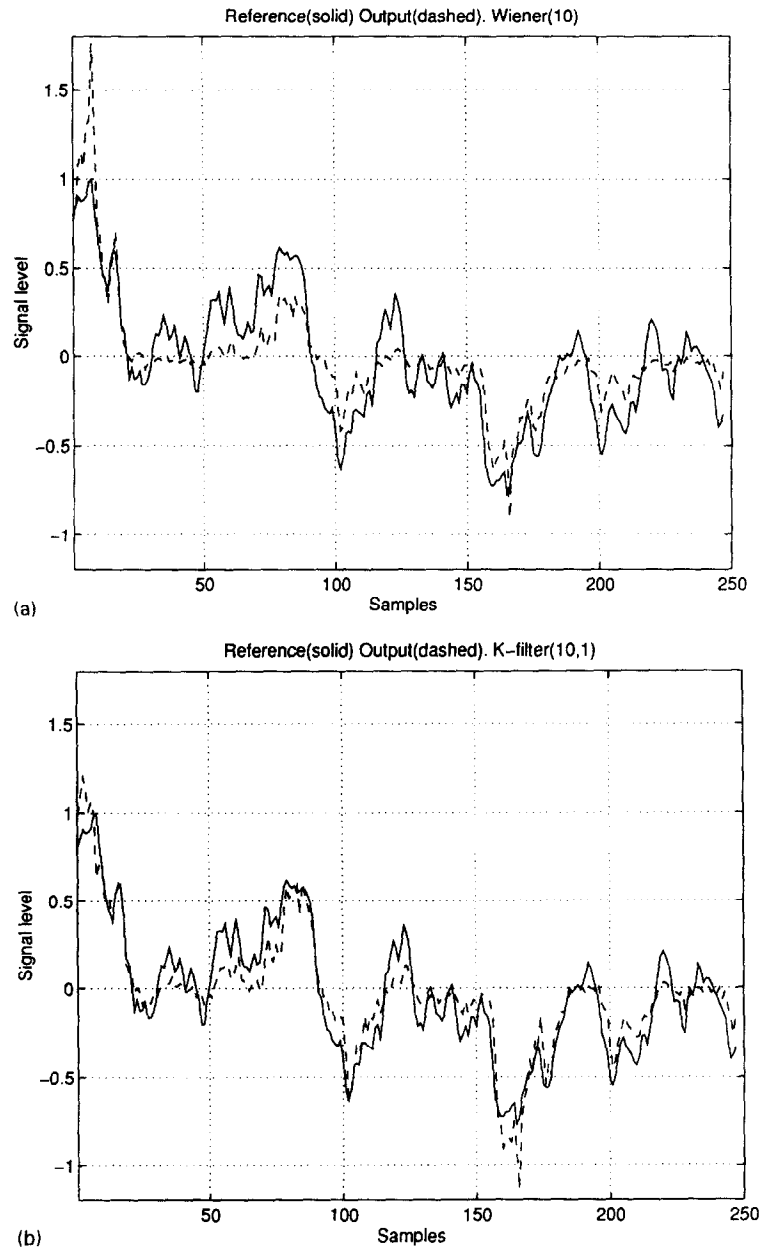


Fig. 12. Representation of a 256 samples of reference data (solid line) and the approximation achieved by the corresponding filter (dashed line) with an SNR of 35 dB. (a) Reference data (solid line). Approximated output by a Wiener (FIR) filter of 10 coefficients (dashed line) (25.3% of residual power/reference power). (b) Reference data (solid line). Approximated output by a K-filter (10, 1) ($\omega_0 = \pi/2X_{\max}$, $\epsilon = \omega_0/10$), i.e. 10 components of diversity and Volterra filters of order 1 (dashed line) (12.9% of residual power/reference power).

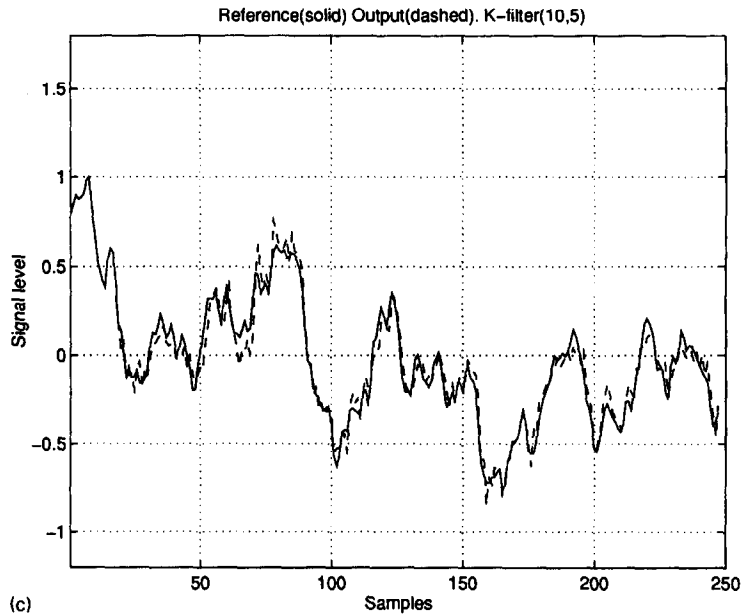


Fig. 12 (c) Reference data (solid line). Approximated output by a K-filter (10, 5) ($\omega_0 = \pi/2X_{\max}$, $\varepsilon = \omega_0/10$), i.e. 10 components of diversity and Volterra filters up to order 5 (dashed line) (3.8% of residual power/reference power).

Table 1

The mean residual power normalised at the power of the reference of 30 independent realisations for different data record lengths and for three different models used in the non-linear filtering problem (Wiener (10), K-filter (10, 1) and K-filter (10, 5)). The model of the system that is filtered is a non-linear system with memory (Fig. 9). The values in brackets correspond to the standard deviation (σ) with respect to the mean value

Data record length	Wiener (10)	K-filter (10, 1)	K-filter (10, 5)
256	22.2% (5.6)	10.9% (3.0)	7.2% (4.4)
512	24.4% (5.2)	12.9% (4.6)	3.1% (1.2)
1024	24.3% (4.5)	12.5% (3.4)	3.1% (0.8)
2048	24.6% (3.6)	12.5% (1.4)	3.7% (1.2)

better than one K-filter with 10 diversity components with a single PM modulator. That is, a K-filter (20, 1) and a K-filter (10, 1) with one branch per diversity component ($\omega_0 = \pi/2X_{\max}$) have a residual power of 16.9% ($\sigma = 4.8$) and 17.6% ($\sigma = 5.4$) respectively, whereas the K-filter (10, 1) with two branches per diversity component ($\omega_0 = \pi/2X_{\max}$ and $\varepsilon = \omega_0/10$) has a residual power

of 12.4% ($\sigma = 3.43$) for the same reference and observed data of 1024 samples length (30 realisations). Note that the procedure works properly without requiring long data records, as it does for other design alternatives which use higher-order moments (order four for tri-linear systems). The K-filter design is a deterministic tool which performs very well for short data records.

A valuable feature of the K-filter is its ability to cope with the linear case. The same experiment for the second example where no non-linearities were present in the model, has been reproduced. In this case the observed data is an AR filtered version of the reference and the residual error powers are in Table 2. These results reveal that the PM modulators produce valid diversity data even when the model to be identified is linear, adding robustness to the K-filter to cover any filtering situation.

Before concluding this section, the performance of the K-filter in the problem of equalising a non-minimum phase channel is considered. This is a widely used problem in non-linear signal processing since, as it will be seen after, a non-linear structure is required to equalise this kind of channels.

In digital communications where the receiver has to decide between two possible values, the equaliser can be seen as a filter that implements a boundary in the input data space in such a way that, when the input data vector is at one side of the boundary, the equaliser decides one value and vice versa. As it is well-known, the most commonly used architecture to equalise a channel is a FIR filter, which is only able to implement a linear boundary in the input data space. It can be demonstrated [13] that when

the channel is non-minimum phase, a non-linear boundary is sometimes required and, thus, a FIR filter is not suitable to separate properly the input data even in the absence of noise. It is in this point where a non-linear structure as the K-filter can be used to equalise the channel in a similar way as the multilayer perceptron [13] or a model based on functional-link [12] have been proposed.

In the concrete problem that the authors propose (Fig. 10), the input consists of a binary polar signal $\{\pm 1\}$ that is driven into a non-minimum phase channel with a transfer function equal to $H_c(z) = (0.5 + z^{-1})$ and followed by an additive Gaussian noise of -20 dB. The non-linear equaliser proposed is a K-filter (2, 2) with two PM modulators per input diversity component ($\omega_0 = \pi/2X_{\max}$ and $\varepsilon = \omega_0/10$), followed by a decision circuit implemented by a sigmoid function of the type $\tanh(\cdot)$ (Fig. 10).

Fig. 13 shows the resulting boundary performed by the K-filter after a training sequence of 200 symbols in which the coefficients of the K-filter have been updated with the LMS algorithm with a step-size parameter equal to $\mu = 0.1$ (the error

Table 2

The same as in Table 1, but in this case the model of the system that is filtered is a linear system (Fig. 9) since the non-linear amplifiers at the transmitter and at the receiver have been avoided

Data record length	Wiener (10)	K-filter (10, 1)	K-filter (10, 5)
256	1.35% (0.57)	1.38% (0.64)	1.17% (0.67)
512	0.98% (0.36)	0.98% (0.32)	0.86% (0.35)
1024	0.81% (0.17)	0.80% (0.17)	0.79% (0.17)
2048	0.76% (0.14)	0.76% (0.02)	0.76% (0.20)

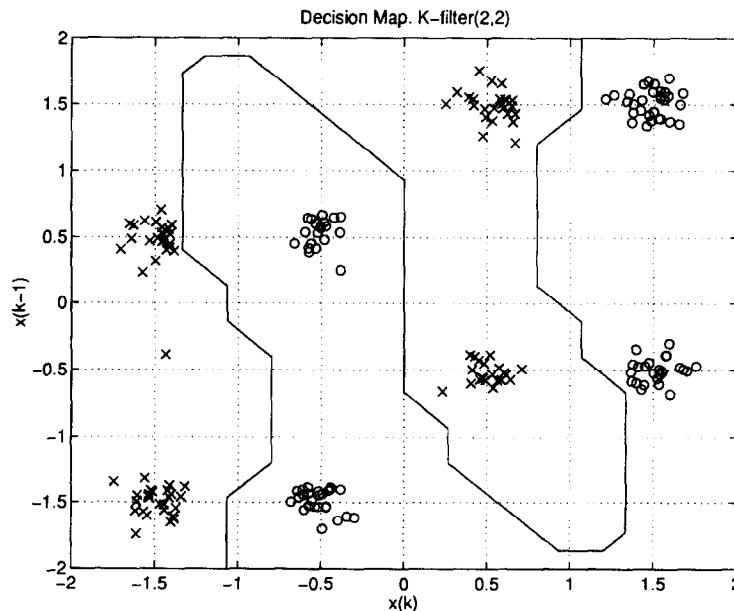


Fig. 13. Representation of the input data space, $x(k)$ versus $x(k - 1)$, with an SNR of 20 dB. Input data of the equaliser when $\{-1\}$ was emitted (crosses) and $\{+1\}$ (circles). The solid line shows the boundary implemented by a K-filter (2, 2) ($\omega_0 = \pi/2X_{\max}$, $\varepsilon = \omega_0/10$) after an adaptive training sequence of 200 symbols, being the weights updated by the LMS algorithm with $\mu = 0.1$

signal is the difference between the real symbol emitted and the output of the decision circuit). During this training period the K-filter has made 9 errors concentrated on the first 25 emitted symbols. This confirms the great ability of the K-filter to learn. Although the K-filter does not achieve the optimum boundary, the result obtained with such a small number of symbols is encouraging to go on to further works.

7. Conclusions

The K-filter architecture reported in this work can be considered as an important contribution to the use of non-linear transforms in both time or spatial diversity FIR filters. Using the Fourier description of a memoryless non-linear system some transformations, which basically are phase modulators, have been introduced. Both in block processing, as well as in adaptive processing, the resulting K-filter performs better than classical Volterra systems. In fact, the K-filter is strongly connected with the Kolmogorov mapping existence theorem, which was the origin of this work. This theorem totally provides the guidelines of the modelling and filtering problem. Starting from the K-filter transformations, a new architecture, which somehow can be viewed as a generalised Hammerstein filter in the sense that the basic idea of this filter is introduced at the diversity level, has been proposed. Regardless that our K-filter can be related to the Hammerstein architecture, our derivation comes from the suggestions of Kolmogorov's theorem. Furthermore, it is proven that with Q diversity inputs the K-filter requires $2Q + 1$ diversity data. Surprisingly enough, the K-filter outperforms the classical Wiener competence and produces the same degree of performance for linear problems, where it was not expected. This feature of the K-filter is not obtained from other alternatives which usually fail when a linear model is assumed. A block processing technique has been suggested for the weighting vector of the K-filter. In summary, further work will be necessary to exploit the potential of the K-filter and in order to justify better the design of the weighting vector with objectives different from the global MSE. Regardless of

the fact that an optimum choice for parameter selection is not provided, the importance of the architecture and the relevance of the Kolmogorov's theorem were considered the main objective of this work. Future work on the application of this architecture to optimum combiners for mobile antennas scenarios will be carried on, since the presence of non-linearities are very common in the spread delay model assumed for the received snapshot.

References

- [1] O. Agazzi, D.G. Messerschmitt and D.A. Hodges, "Non-linear echo cancellation of data signals", *Proc. IEEE Internat. Conf. Comm.*, 1982, pp. 7G6.1–7G6.5.
- [2] Y. Bar-Ness et al., "Cross-coupled phase-locked loop with closed loop amplitude control", *IEEE Trans. Comm.*, Vol. COM-32, No. 2, February 1984, pp. 195–199.
- [3] Y. Bar-Ness and J. Rokach, "Cross coupled bootstrapped interference canceller", *Proc. Internat. Symp. IEEE AP*, 1981, pp. 292–295.
- [4] J.S. Bendat, *Nonlinear Systems Analysis and Identification from Random Data*, Wiley, New York, 1990.
- [5] I. Bilinskis and A. Michlesons, *Randomised Signal Processing*, Prentice Hall, Englewood Cliffs, NJ, 1992.
- [6] J.J. Bussgang, "Cross-correlation functions of amplitude distorted Gaussian inputs", in: A.H. Haddad, ed., *Nonlinear Systems*, Dowden, Hutchinson & Ross, Stroudsburg, PA, 1975.
- [7] J.R. Casar and M. Garcia-Otero, "A class of adaptive non-linear filters", in: D. Abontadjine, ed., *Proc. MWPSA Workshop*, Marrakech, Morocco, 24–26 September 1984, paper A4.2.
- [8] S. Chen, C. Cowans and P. Grant, "Orthogonal least squares learning algorithm for radial basis functions networks", *Proc. Internat. Conf. IEEE Acoust. Speech Signal Process.*, Vol. 2, No. 2, March 1991, pp. 302–309.
- [9] P. Comon, C. Jutten and J. Herault, "Blind separation of sources, Part II: Problem statement", *Signal Processing*, Vol. 24, No. 1, July 1991, pp. 11–20.
- [10] G. Cybenko, "Approximation by superposition of a single function", *Mathematical Control, Signal and Systems*, Vol. 2, 1989, pp. 303–314.
- [11] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm", *Proc. Internat. Conf. IEEE Acoust. Speech Signal Process.*, Vol. 36, No. 4, April 1988, pp. 477–489.
- [12] W.S. Gan, J.J. Sorahgan and T. Durrani, "Functional link models for adaptive equalisation", *Proc. Internat. Conf. IEEE Internat. Conf. Acoust. Speech Signal Process.*, Adelaide, Vol. 3, 1994, pp. 309–312.
- [13] G.J. Gibson, S. Siu and C.F.N. Cowan, "The Application of non-linear structure to the reconstruction of binary signals", *IEEE Trans. Signal Process.*, Vol. 39, No. 8, August 1991, pp. 1877–1884.

- [14] R. Hecht-Nielsen, *Neurocomputing*, Addison-Wesley, Reading, MA, 1990, pp. 122–124.
- [15] C. Jutten, Calcul neuromimetique et traitement du signal: Analyse en composants independents, Ph.D. Thesis (in French), INPG-USMG, Grenoble, 1987.
- [16] C. Jutten and J. Hérault, “Independent component analysis (INCA) versus principal component analysis”, *Proc. 5th Internat. Conf. EURASIP EUSIPCO’92*, Grenoble, France, Elsevier, Amsterdam, 1988.
- [17] 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*, Vol. 114, 1957, pp. 953–956.
- [18] D. Kraus, D. Maiwald and J.F. Böhme, “Maximum likelihood source location estimation via EM algorithm”, *Proc. 7th Internat. Conf. EURASIP EUSIPCO’92*, Brussels, Belgium, Elsevier, Amsterdam, August 1992.
- [19] V. Kurkova, “Kolmogorov’s theorem is relevant”, *Neural Computation*, Vol. 3, No. 4, MIT, 1991, pp. 617–622.
- [20] V. Kurkova, “Kolmogorov theorem and multilayers neural networks”, *Neural Networks*, Vol. 5, 1992, pp. 501–506.
- [21] J.L. Lacoume, ed., “HOS”, *Proc. Signal Processing on High Order Statistics*, Chamrouse, France, July 1991.
- [22] M.A. Lagunas et al., “The Kolmogorov Signal Processor”, *Lecture Notes in Computer Science, New Trends in Neural Computation*, Springer, Berlin, Vol. 686, June 1993, pp. 494–512.
- [23] G. Lorentz, *Approximation Functions*, Chelsea, New York, 1986, pp. 168–179.
- [24] B. Mulgrew, “Orthogonal functions for non-linear signal processing for adaptive filtering”, *Proc. Internat. Conf. IEEE Internat. Conf. Acoust. Speech Signal Process.*, Adelaide, Vol. 3, 1994, pp. 509–512.
- [25] C.L. Nikias and J.M. Mendel, “Signal processing with high order spectra”, *IEEE Signal Process. Mag.*, July 1993.
- [26] R. Price, “A useful theorem for non-linear devices having Gaussian inputs”, in: A.H. Haddad, ed., *Nonlinear Systems*, Dowden, Hutchinson & Ross, Stroudsburg, PA, 1975.
- [27] W.J. Rugh, *Nonlinear System theory*, Johns Hopkins Univ. Press, Baltimore, MD, 1981.
- [28] M. Schetzen, *The Volterra and Wiener Theories of Non-linear Systems*, Wiley, New York, 1980.
- [29] M. Schetzen, “Non-linear system modeling based on the Wiener theory”, *Proc. IEEE*, Vol. 69, No. 12, December 1981, pp. 1557–1573.
- [30] G.L. Sicuranza, A. Bucconi and P. Mitri, “Adaptive echo cancellation with non-linear digital filters”, *Proc. IEEE Internat. Conf. Acoust. Speech Signal Process.*, San Diego, California, 1984, pp. 3.10.1–3.10.4.
- [31] M. Vidyasagar, *Non-linear Systems Analysis*, Networks Series, R.W. Newcomb, ed., Prentice Hall, Englewood Cliffs, NJ, 1978.