THE GAMMA FILTER - A New Class of Adaptive IIR Filters with Restricted Feedback

Jose C. Principe, Bert de Vries and Pedro Guedes de Oliveira*

Computational Neuro-engineering Laboratory
Electrical Engineering Department, University of Florida
Gainesville, FL 32611
principe@synapse.ee.ufl.edu

*Departamento Eletronica/INESC
Universidade de Aveiro
Aveiro, Portugal

Abstract

In this paper we introduce the generalized feedforward filter, a new class of adaptive filters that combine attractive properties of Finite Impulse Response (FIR) filters with some of the power of Infinite Impulse Response (IIR) filters. A particular case, the adaptive gamma filter, generalizes Widrow's adaptive linear combiner (adaline) to an infinite impulse response filter. Yet, the stability condition for the gamma filter is trivial, and LMS adaptation is of the same computational complexity as the conventional adaline structure. Preliminary results indicate that the adaptive gamma filter is more efficient than adaline. We extend the Wiener-Hopf equation to the gamma filter and develop some analysis tools.

1 INTRODUCTION

Infinite Impulse Response (IIR) filters are more efficient than Finite Impulse Response (FIR) filters, but in adaptive signal processing FIR systems are almost exclusively used (Haykin, 1991; Widrow and Stearns, 1985). This is largely due to the difficulty of ensuring stability during adaptation of IIR systems. Moreover, gradient descent adaptive procedures are not guaranteed to find global optima in the non-convex error surfaces of IIR systems (Shynk, 1989).

Yet IIR systems have an important advantage over FIR systems. For a Kth order FIR system, both the region of support of the impulse response and the number of adaptive parameters equal K. For an IIR system, the length of the impulse response is *uncoupled* from the order (and number of parameters) of the system. Since the length of the impulse response of a filter is closely related to the depth of memory of the system, IIR systems are preferred over FIR systems for modeling of systems and signals characterized by a deep memory and a small number of free parameters. These features are typical for lowpass frequency signals, as is the case for most biological and other real-world signals.

In this paper we introduce the generalized feedforward filter, an IIR filter with restricted feedback architecture. The gamma filter, a particular instance of the generalized feedforward filter, is analyzed in detail. The gamma filter borrows desirable features from both

IIR and FIR systems - trivial stability, easy adaptation and yet the uncoupling of the region of support of the impulse response and the filter order.

This paper is organized as follows. In the next section the generalized feedforward filter is presented. This section is followed by the presentation of the gamma filter, an analysis of its properties, and a comparison with respect to FIR and IIR filter structures. In particular, we analyze stability properties, memory depth, adaptation equations and generalize the Wiener-Hopf equations to the gamma filter. Next a simulation experiment concerning the gamma filter performance in a system identification configuration is presented. Finally, we introduce the γ-transformation which provides a mathematical framework to describe gamma filters as conventional FIR filters in the γ-domain, despite their IIR nature. As a result most FIR tools are applicable to gamma filters.

2 GENERALIZED FEEDFORWARD FILTERS - DEFINITIONS

Consider the IIR filter architecture described by -

$$Y(z) = \sum_{k=0}^{K} w_k X_k(z) \qquad (Eq.1)$$

$$X_k(z) = G(z) X_{k-1}(z) \text{, k = 1,...,K,} \qquad (Eq.2)$$

where $X_0(z) \equiv X(z)$ [1] is the input signal and $Y(z)$ the filter output (Figure 1).

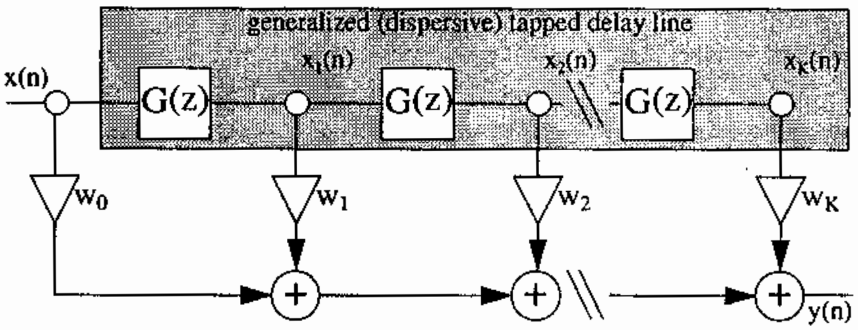

Figure 1 the generalized feedforward filter

We refer to this structure as the *generalized feedforward filter*. The tap-to-tap transfer function $G(z)$ is called the *(generalized) delay operator* and it can either be recursive or non-recursive. When $G(z) = z^{-1}$, this filter structure reduces to an FIR filter. The memory structure of an FIR filter is simply a tapped delay line. By iteration of (Eq.2) we can write $Y(z)$ as a function of the input $X(z)$ as follows -

1. Read $\equiv$ as 'is defined as'.

$$Y(z) = \sum_{k=0}^{K} w_k [G(z)]^k X(z) \qquad (Eq.3)$$

We will also write $G_k(z) \equiv [G(z)]^k$ for the input-to-tap-k transfer function. Thus, the transfer function of the generalized feedforward filter is -

$$H(z) \equiv \frac{Y(z)}{X(z)} = \sum_{k=0}^{K} w_k [G(z)]^k. \qquad (Eq.4)$$

It follows from (Eq.4) that $H(z)$ *is stable whenever* $G(z)$ *is stable.*

The past of $x(n)$ is represented in the tap variables $x_k(n)$ (shaded area in Figure 1). Although conventional digital signal processing structures are built around the tapped delay line ($G(z) = z^{-1}$), we have observed that alternative delay operators may lead to better filter performance. In general, the optimal memory structure $G(z)$ is a function of the input signal characteristics as well as the goal of the filter operation. This observation has led us to consider *adaptive* delay operators $G(z;\mu)$, where μ is an adaptive memory parameter. As a notational convenience, $G(z) = G(z;\mu)$ will be adopted.

This paper analyzes in detail the case $G(z) = \frac{\mu}{z-(1-\mu)}$, the *gamma delay operator.* The gamma delay operator can be interpreted as a leaky integrator, where $1 - \mu$ is the gain in the integration (feedback) loop.

3 THE GAMMA FILTER

3.1 Definitions

The *gamma filter*[1] is defined in the time domain as -

1. The gamma filter was originally developed in continuous time as part of a neural net model for temporal processing (de Vries and Principe, 1991). We showed that - by transformation $s = \frac{z-1}{T_s}$ - the impulse response of the continuous time gamma filter can be written as $h(t) = \sum_{k=0}^{K} w_k g_k(t)$, where $g_k(t) = \frac{\mu^k}{(k-1)!} t^{k-1} e^{-\mu t}$, k = 1,...,K, and $g_0(t) = \delta(t)$. The functions $g_k(t)$ are the integrands of the (normalized) gamma function. Hence the name gamma model for structures that utilize tap variables of type $x_k(t) = (g_k \bullet x)(t)$ to store the past of x(t) (here $\bullet$ denotes the convolution operator). Closely related are Laguerre functions, that were proposed by Norbert Wiener (1949) as a very convenient basis for decomposition of linear systems in a signal processing context. In fact, the functions $g_k(t)$, k = 1,...,K, can be easily written in terms of Laguerre functions. Since the Laguerre functions are complete, it follows that the functions $g_k(t)$ are complete in $L_2[0,\infty]$.

$$y(n) = \sum_{k=0}^{K} w_k x_k(n) \qquad (Eq.5)$$

$$x_k(n) = (1-\mu)x_k(n-1) + \mu x_{k-1}(n-1), k = 1,\ldots,K, \qquad (Eq.6)$$

where $x_0(n) \equiv x(n)$ is the input signal and $y(n)$ the filter output (Figure 2). When $w_0, w_1, \ldots, w_k$ and μ are adaptive, this structure is called the *adaptive gamma filter* or *adaline*(μ).

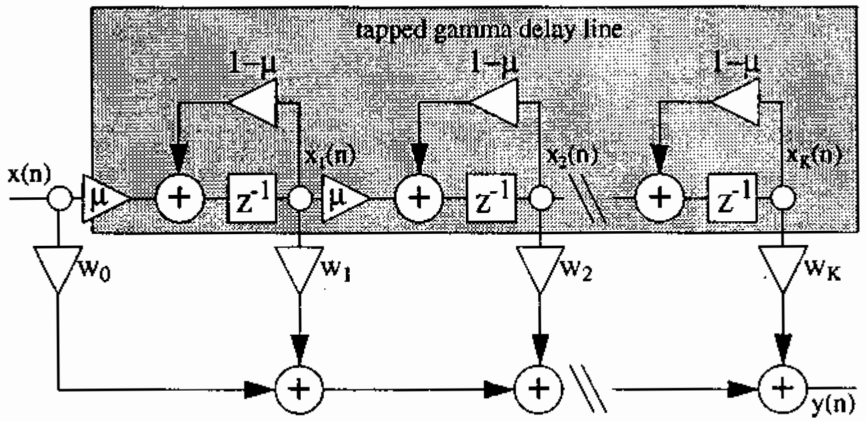

Figure 2 the gamma filter structure

Following the definitions in section 2, the gamma input-to-tap-k transfer function $G_k(z)$ is given by -

$$G_k(z) = \left(\frac{\mu}{z-(1-\mu)}\right)^k. \qquad (Eq.7)$$

Inverse z-transformation yields the impulse response for tap k -

$$g_k(n) \equiv \mathcal{Z}^{-1}\{G_k(z)\} = \binom{n-1}{k-1}\mu^k(1-\mu)^{n-k}U(n-k), \qquad (Eq.8)$$

where $U(n)$ is the unit step function. Note that the gamma delay operator is normalized, that is, -

$$\sum_{n=0}^{\infty} g_k(n) = G_k(z)\big|_{z=1} = 1. \qquad (Eq.9)$$

When $\mu = 1$, the adaptive gamma filter reduces to Widrow's adaline structure (Widrow and Stearns, 1985). For $\mu \neq 1$, the gamma filter transfer function is of IIR type due to the recursion in (Eq.6), and G(z) implements a *dispersive* delay unit. In comparison to a general IIR filter, the feedback structure in the gamma filter is restricted by two conditions -

- (C1): the recurrent loops are *local* with respect to the taps.
- (C2): the loop gain 1 - μ is *global* (all feedback loops have the same gain).

In fact, the conditions C1 and C2 are typical for all generalized feedforward structures. Now let us analyze some of the properties of the adaptive gamma filter.

3.2 Stability

Due to the restricted nature of the feedback loops it is easily verified that stability is guaranteed when $0 < \mu < 2$.

3.3 Memory Depth versus Filter Order

We have discussed the strict coupling of the memory depth to the number of free parameters in the adaptive FIR filter structure and argued that this property leads to poor modeling of low pass frequency bounded signals. IIR filters on the other hand have feedback connections, and consequently the memory depth is not coupled to the number of filter parameters. In this section an effort is made to quantify the relation memory depth versus filter order for the gamma filter. It will be shown that the memory parameter μ provides a mechanism to uncouple depth from the filter order.

First, let us first make the notion of memory depth more quantitative. The *mean sampling time* n_k for the kth tap is defined as -

$$n_k \equiv \sum_{n=0}^{\infty} n g_k(n) = Z\{n g_k(n)\}\big|_{z=1} = -z \frac{dG_k(z)}{dz}\bigg|_{z=1} = \frac{k}{\mu}. \qquad (\textit{Eq.10})$$

We also define the *mean sampling period* Δn_k (at tap k) as $\Delta n_k \equiv n_k - n_{k-1} = \frac{1}{\mu}$. The *mean memory depth* D_k for a gamma memory of order k then becomes -

$$D_k \equiv \sum_{i=1}^{k} \Delta n_i = n_k - n_0 = \frac{k}{\mu}. \qquad (\textit{Eq.11})$$

If the *resolution* R_k is defined as $R_k \equiv \frac{1}{\Delta n_k} = \mu$, the following formula arises which is of fundamental importance for the characterization of the gamma memory structure (we dropped the subscript when k = K)-

$$K = D \times R. \qquad (\textit{Eq.12})$$

(Eq.12) reflects the possible trade-off of resolution versus memory depth in a memory structure for fixed order K. Such a trade-off is not possible in a non-dispersive tapped delay line, since the fixed choice of $\mu = 1$ sets the depth and resolution to $D = K$ and $R = 1$ respectively. However, in the gamma memory, depth and resolution can be adapted by variation of μ. The choice $\mu = 1$ represents a memory structure with maximal resolution and minimal depth. In this case, the order K and depth D of the memory are equal. In FIR and gamma filter structures, the number of adaptive parameters and the filter order are coupled (both K). Thus, when $\mu = 1$, the number of weights equals the memory depth. Very often this coupling leads to overfitting of the data set (using parameters to model the noise). Hence, the parameter μ provides a means *to decouple the memory order and depth.*

The depth of memory as a function of frequency is measured by the group delay. In

Figure 3 we have plotted the group delay of $G_2(z) = \left(\frac{\mu}{z-(1-\mu)}\right)^2$ (that is, the input-to-tap-2 transfer function) for three values of μ. Note that for $\mu < 1$ the group delay at low frequencies is greater than the tap index k = 2. Thus, for $\mu < 1$ additional memory depth is obtained for low frequencies at the cost of group delay for the high frequencies. When most of the information of a signal is in the low pass region, the favoring of low frequencies in the memory can be efficient.

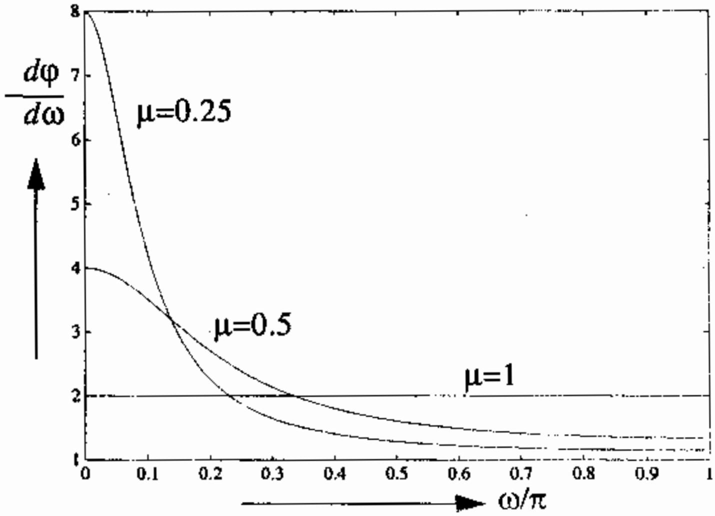

Figure 3 group delay of gamma memory at tap k = 2.

As an example, assume a signal whose dynamics are described by a system with 5 parameters and maximal delay 10, that is, $y(t) = f(x(t-n_i), w_i)$ where $i = 1,...,5$, and $max_i(n_i) = 10$. If we try to model this signal with an adaline structure, the choice $K = 10$ leads to overfitting while $K < 10$ leaves the network unable to incorporate the influence of $x(t-10)$. In an adaline(μ) network, the choice $K = 5$ and $\mu = 0.5$ leads to 5 free network parameters and mean memory depth of 10, obviously a better compromise.

3.4 LMS Adaptation

In this section the least mean square (LMS) adaptation update rules for the gamma filter parameters w_k and μ are derived. In particular, our interest is to show that the update equations can be computed by an algorithm where the number of operations per time step scales by $O(K)$, K being the filter order. This is interesting, since LMS type algorithms scale by $O(K^2)$ for general IIR filters.

Consider the gamma filter as described by the set of equations (Eq.5) and (Eq.6). Let the performance of the system be measured by the *total error E*, defined as -

$$E \equiv \sum_{n=0}^{T} E_n = \sum_{n=0}^{T} \frac{1}{2} e^2(n) = \sum_{n=0}^{T} \frac{1}{2} (d(n) - y(n))^2 \qquad (Eq.13)$$

where $d(n)$ is a *target signal*. The LMS algorithm corrects the filter coefficients proportionally to the negative of the local gradient, i.e. the coefficient update equations are in the direction of the negative gradients -

$$\Delta w_k = -\eta \frac{\partial E}{\partial w_k} \qquad (Eq.14)$$

$$\Delta \mu = -\eta \frac{\partial E}{\partial \mu}, \qquad (Eq.15)$$

where η is a *step size* parameter. We first expand for w_k, yielding-

$$\Delta w_k = -\eta \frac{\partial E}{\partial w_k} = \eta \sum_{n=0}^{T} e(n) \frac{\partial y(n)}{\partial w_k} = \eta \sum_{n=0}^{T} e(n) x_k(n) \qquad (Eq.16)$$

Similarly, the update equation for μ evaluates to -

$$\Delta \mu = -\eta \frac{\partial E}{\partial \mu} = \eta \sum_{n=0}^{T} e(n) \sum_{k=0}^{K} w_k \frac{\partial x_k(n)}{\partial \mu} = \mu \sum_{n=0}^{T} \sum_{k=0}^{K} e(n) w_k \alpha_k(n) \qquad (Eq.17)$$

where $\alpha_k(n) \equiv \frac{\partial x_k(n)}{\partial \mu}$. The gradient signal $\alpha_k(n)$ can be computed on-line by differentiating (Eq.6) (Shynk, 1989; Williams and Zipser, 1989), leading to -

$$\alpha_0(n) = 0$$

$$\alpha_k(n) = (1-\mu)\alpha_k(n-1) + \mu\alpha_{k-1}(n-1) + \mu[x_{k-1}(n-1) - x_k(n-1)] \quad k=1,..,K \qquad (Eq.18)$$

The set of equations (Eq.16), (Eq.17) and (Eq.18) constitute the update algorithm in *block mode adaptation*. In practice, a *local in time* approximation (i.e. sample by sample) of the form -

$$\Delta w_k(n) = \eta e(n) x_k(n),\ k=0,..,K \qquad (Eq.19)$$

$$\Delta \mu = \eta \sum_{k=0}^{K} e(n) w_k \alpha_k(n) \qquad (Eq.20)$$

works well if η is sufficiently small. (Eq. 19) can be recognized as the update term in the LMS algorithm. Notice the number of operations per time step for (Eq.19) and (Eq.20) scale both as $O(K)$ whereas (Eq.18) is $O(1)$. Thus, the entire LMS algorithm scales as $O(K)$, which coincides with the complexity for Widrow's adaline. The gain with respect to a general IIR LMS routine (scales as $O(K^2)$) is due to the restricted nature of the feedback loops in the gamma filter.

The results of the last three sections are summarized in Table 1. Clearly the gamma filter shares desirable features from both FIR and IIR filters.

Kth order filter	FIR	GAMMA	IIR
STABILITY	always stable	trivial stability $0 < \mu < 2$	non-trivial stability
MEMORY DEPTH vs. ORDER	coupled K	decoupled K/m	decoupled
COMPLEXITY of ADAPTATION	O(K)	O(K)	O(K^2)

Table 1 comparison of FIR, IIR and Gamma filter properties.

3.5 Wiener-Hopf Equations for the Gamma Filters

The optimal weights for an adaline structure in a given stationary environment can be analytically expressed by the *Wiener-Hopf* or *normal equations* (Haykin, 1991). Here these equations are extended to the gamma filter. We will show that the gamma normal equations generalize Wiener's formulation for strict feedforward filters.

Consider the adaline(μ) structure as described by (Eq.5) and (Eq.6). We define a *performance index* $\xi \equiv E[e^2(n)]$ where $e(n) \equiv d(n) - y(n)$ is an *error signal* and E[.] the expectation operator. In order to maintain a consistent notation with respect to the adaptive signal processing literature, we introduce the vectors $X_n = [x_0(n), x_1(n), \ldots, x_K(n)]^T$ and $W = [w_0, w_1, \ldots, w_K]^T$. Note that X_n *holds the tap variables and not the input signal samples.* Evaluating ξ leads to -

$$\xi = E[d^2(n)] + W^T R W - 2P^T W, \qquad (Eq.21)$$

where $R \equiv E[X_n X_n^T]$ and $P \equiv E[d(n) X_n]$. The goal of adaptation is to minimize ξ in the space of K+1 weights and μ. When ξ is minimal, the conditions $\frac{\partial \xi}{\partial w_k} = 0$ and $\frac{\partial \xi}{\partial \mu} = 0$ necessarily hold. Partial differentiation of (Eq.21) with respect to the system parameters yields the following results -

$$RW = P \text{ and} \qquad (Eq.22)$$

$$W^T [R_\mu W - 2P_\mu] = 0, \qquad (Eq.23)$$

where $R_\mu \equiv \frac{\partial R}{\partial \mu} = 2E\left[X_n \frac{\partial X_n^T}{\partial \mu}\right]$ and $P_\mu \equiv \frac{\partial P}{\partial \mu} = E\left[d(n) \frac{\partial X_n}{\partial \mu}\right]$.

Note that (Eq.22) is the same expression as the Wiener-Hopf equation for the adaline network. The difference lies in the fact that the vector X_n holds the tap variables $x_k(n)$ and not the samples $x(n-k)$. The extra *scalar* condition (Eq.23) is a result of requiring $\frac{\partial \xi}{\partial \mu} = 0$. Thus, (Eq.23) provides *an analytical expression for the optimal memory depth*. This expression also reveals that the signal $\alpha_k(n) \equiv \frac{\partial x_k(n)}{\partial \mu}$ is needed in order to compute the optimal memory structure (that is, the optimal value of μ). This observation is confirmed in the expressions for the LMS algorithm.

It is insightful to rewrite the Wiener-Hopf equations in terms of the input signal $x(n)$. Let us define the delay kernel vector $G(n) \equiv [g(n), g^2(n), \dots, g^K(n)]^T$. Then (Eq.22) and (Eq. 23) evaluate to [1] -

$$E[(G(n) \bullet x(n))(G(n) \bullet x(n))^T] W = E[d(n)(G(n) \bullet x(n))] \qquad (Eq.24)$$

$$W^T E\left[(G(n) \bullet x(n)) \left(\frac{\partial G(n)}{\partial \mu} \bullet x(n)\right)^T\right] W = 2W^T E\left[d(n) \left(\frac{\partial G(n)}{\partial \mu} \bullet x(n)\right)\right]. \qquad (Eq.25)$$

(Eq.24) and (Eq.25) extend the Wiener-Hopf equations to generalized feedforward structures. Note that these equations in the time domain include infinite summations ($g(n)$ may be of infinite length), but in the z-domain they can be computed exactly by contour integration.

4 EXPERIMENTAL RESULTS

We have presented two frameworks to obtain an optimal filter architecture adaline(μ). In section 3.4 the LMS adaptation algorithm was derived and section 3.5 was devoted to the Wiener-Hopf equations for the gamma filter. In this section we present numerical simulation results for both optimization models when adaline(μ) is used in a system identification configuration. The goal of this section is twofold. First we will show that the optimal filter architecture indeed outperforms Widrow's adaline(1). Also, it will be shown that the filter coefficients converge to the optimal values if the LMS update rules of section 3.4 are used.

The system to be identified is the 3rd order elliptic low pass filter described by[2] -

1. $\bullet$ denotes the convolution operator.

2. This filter has been described in Oppenheim and Schafer, 1975, pg.226

$$H(z) = \frac{0.0563 - 0.0009z^{-1} - 0.0009z^{-2} + 0.0563z^{-3}}{1 - 2.1291z^{-1} + 1.7834z^{-2} - 0.5435z^{-3}}. \quad (Eq.26)$$

The performance index ξ as a function of μ was computed by evaluating (Eq.21) in the z-domain (residue theorem). The optimal weight vector W^* is computed by solving the Wiener-Hopf equation (Eq.22), and substituting back into (Eq.21). We assumed a normal (0,1)-distributed white noise input, which translates to a constant spectrum in the z-domain. μ was parametrized over the real domain [0,1]. The simulations, plotted in Figure 4(a), were performed with Mathematica (Wolfram, 1989) on a NeXT computer. The time duration of the simulations restricted the evaluation to $K \le 3$. Observe that for all memory orders K the optimal performance is obtained for $\mu < 1$. Hence, the optimal gamma net outperforms the conventional adaline by a large margin.There is a lot of structure in the ξ curves. Note that the optimal memory depth $D_{opt} \equiv \frac{K}{\mu_{opt}} \approx 5$ is constant for different memory orders.

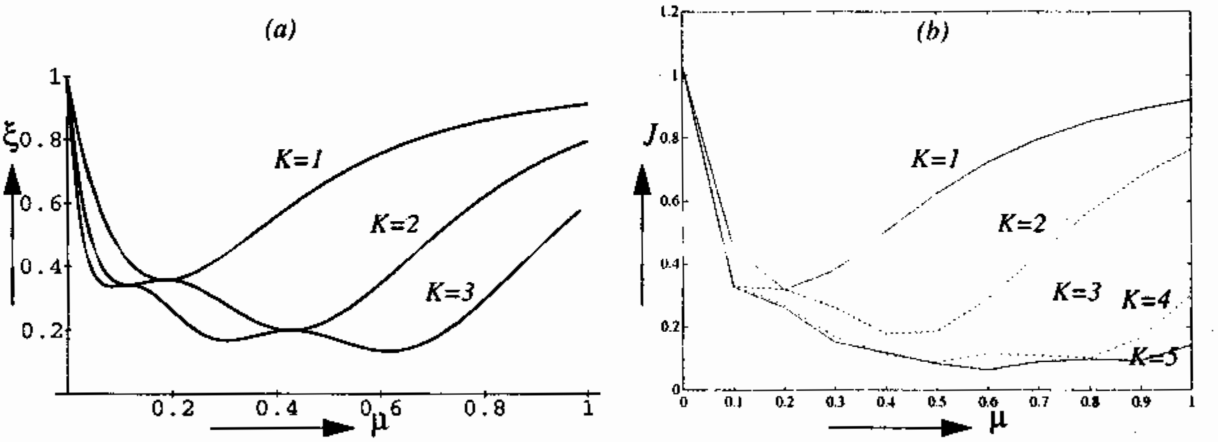

Figure 4 Optimal performance index as a function of μ for identification of elliptic filter H(z). (a) ξ is computed using from the Wiener-Hopf equations for adaline(μ). (b) J_{min}=var[e(n)]/var[d(n)] is

In Figure 4(b) we show the relative total error $J = \frac{\sigma_e^2}{\sigma_d^2}$ after convergence using the LMS update rule (Eq.19). We parametrized μ over the domain [0,1] using a step size $\Delta\mu = 0.1$. The results match the theoretical optimal performance (Figure 4(a)) very well. This experiment shows that the filter weights $\{w_k\}$ can indeed be learned by on-line LMS learning. Figure 4(b) shows that it is irrelevant to use memory orders higher than $K = 5$ for this identification problem. In fact, when $K = 5$, adaline(1) performs as well as $K = 3$ for adaline(0.6). However, we still prefer $K = 3$, since this structure has 5 free parameters whereas adaline uses 7 parameters (K+1 weights w_k plus μ). Parsimony in the number of free parameters provides adaline(0.6) with better modeling (generalization) characteristics.

The effect of the memory parameter μ on the filter performance increases when we model a system with smaller cut-off frequency but the same number of parameters. In Figure 5 the performance index ξ versus μ is plotted for a third-order elliptic low pass filter $H_2(z)$ with smaller cut-off frequency ($w_{co} = 0.06\pi$ rad) -

$$H_2(z) = \frac{1 - 0.8731z^{-1} - 0.8731z^{-2} + z^{-3}}{1 - 2.8653z^{-1} + 2.7505z^{-2} - 0.8843z^{-3}} \quad (Eq.27)$$

It is clear that the third-order adaline structure performs very poorly ($\xi \approx 1$) whereas the third-order adaline(0.15) performs at $\xi = 0.1$.

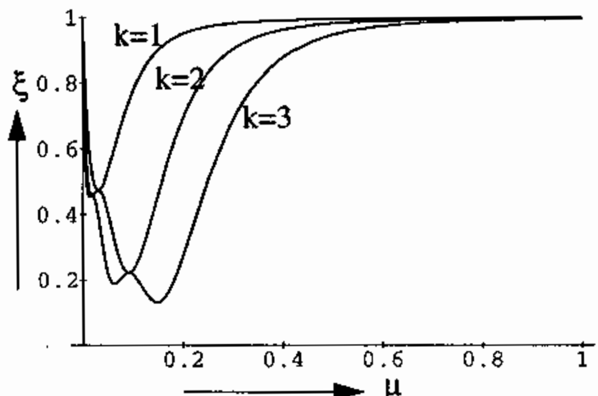

Figure 5 ξ vs. μ for $H_2(z)$

We have experimented with several signals (sinusoids in noise, Feigenbaum map, electroencephalogram (EEG)) for various processing protocols (prediction, system identification, classification). Invariably the optimal memory structure[1] was obtained for $\mu <$ 1. These data will be reported in a forthcoming publication.

5 THE GAMMA TRANSFORM - A DESIGN AND ANALYSIS TOOL FOR GAMMA FILTERS

Sofar we have analyzed the (adaptive) gamma filter properties using the time and z-domain tools and compared the results to FIR and IIR filters. We have shown that the restricted nature of the feedback connections in the gamma filter has rendered this model with desirable properties from both filter classes. In fact, the gamma filter can be viewed as an instance of a hybrid filter class, the generalized feedforward filter. An interesting feature of gamma filters, already explored for the extension of the Wiener-Hopf solution, is that they can be formulated as FIR filters with respect to a delay operator $G(z)$. In this section we explore the implications of describing the system in a new transform domain, the γ-domain, which we define as -

$$\gamma^{-1} \equiv G(z) . \quad (Eq.28)$$

1. The optimal memory structure is defined as the structure of lowest dimensionality that minimizes the performance index J.

In the γ-domain, generalized feedforward filters are ordinary FIR filters, defined around delay operators γ^{l}. For gamma filters, (Eq.28) evaluates to -

$$\gamma = \frac{z-(1-\mu)}{\mu}. \qquad (Eq.29)$$

A signal $x(n)$ can be expressed in the γ-domain by substituting (Eq.29) in the z-transform. This leads to the following expression for the *γ-transform* of a signal $x(n)$ -

$$X(\gamma) \equiv X(z)\big|_{z=\mu\gamma+(1-\mu)}$$

$$= \sum_{n=0}^{\infty} x(n)\,z^{-n}\bigg|_{z=\mu\gamma+(1-\mu)}$$

$$= \sum_{n=0}^{\infty} \mu^{-n}x(n)\,\left\{\gamma+\frac{1-\mu}{\mu}\right\}^{-n} \qquad (Eq.30)$$

Thus, the γ-transform is equivalent to the Laurent series expansion of the signal $\mu^{-n}x(n)$ evaluated at the point $\gamma_0 = \frac{\mu-1}{\mu}$. This idea is displayed in Figure 6.

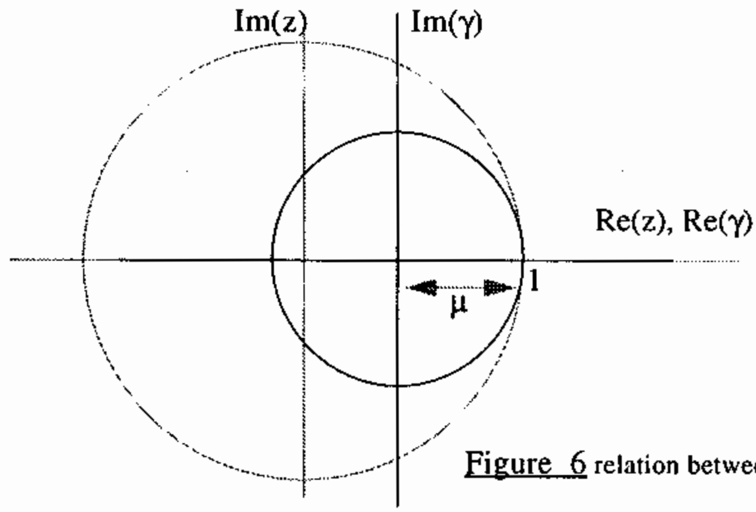

Figure 6 relation between z- and γ-domain

The corresponding time series obtained by the *inverse γ-transform* can be computed as -

$$x(n) = \frac{1}{2\pi j}\oint_C X(z)\,z^{n-1}dz$$

$$= \frac{1}{2\pi j}\oint_C X(\gamma)\,\{\mu\gamma+(1-\mu)\}^{n-1}\mu d\gamma \qquad (Eq.31)$$

or -

$$\mu^{-n} x(n) = \frac{1}{2\pi j} \oint_C X(\gamma) \left\{ \gamma + \frac{1-\mu}{\mu} \right\}^{n-1} d\gamma. \quad (Eq.32)$$

where C is a closed contour that encircles the point γ_0. The equations (Eq.30) and (Eq.32) relate the time domain and the γ-domain. Since the gamma filter is a FIR filter in the γ-domain, all design and analysis tools available for this class of filters are without restriction applicable to gamma filters in the gamma domain.

As an example, let us analyze a simple second order gamma filter with w_0=-1, w_1=0 and w_2=1 (Figure 7). Note that the pole(s) of a gamma filter are located at the origin in the γ-plane. The

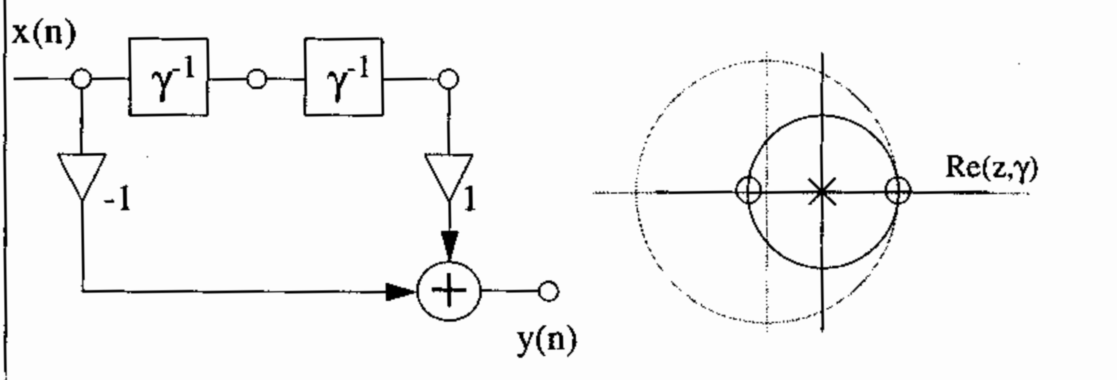

Figure 7 a second order gamma filter.

zeros are located at γ = -1 and γ = 1. Thus, the system is feedforward in the γ-domain. The transfer function in the γ-domain is easily obtained by inspection -

$$H(\gamma) \equiv \frac{Y(\gamma)}{X(\gamma)} = -1 + \gamma^{-2}. \quad (Eq.33)$$

Substitution of (Eq.29) gives the transfer function in the z-domain -

$$H(z) = -1 + \{\mu z + (1-\mu)\}^{-2}. \quad (Eq.34)$$

The impulse response of a gamma filter can be expressed as follows -

$$h(n) = \sum_{k=0}^{K} w_k g_k(n)$$

$$= -g_0(n) + g_2(n)$$

$$= -\delta(n) + (n-1)\mu^2 (1-\mu)^{n-2} U(n-1). \quad (Eq.35)$$

In Figure 7 the system's magnitude frequency and impulse responses are displayed as a function of μ. Note that if μ is close to 1, the gamma filter behaves as the FIR system $H(z) = z^{-2}-1$. When μ gets smaller, the "peak" of the frequency response becomes sharper, which is typical for IIR filters as compared to FIR filters of the same order. Thus, the global filter parameter μ determines whether FIR or IIR filter characteristics are obtained.

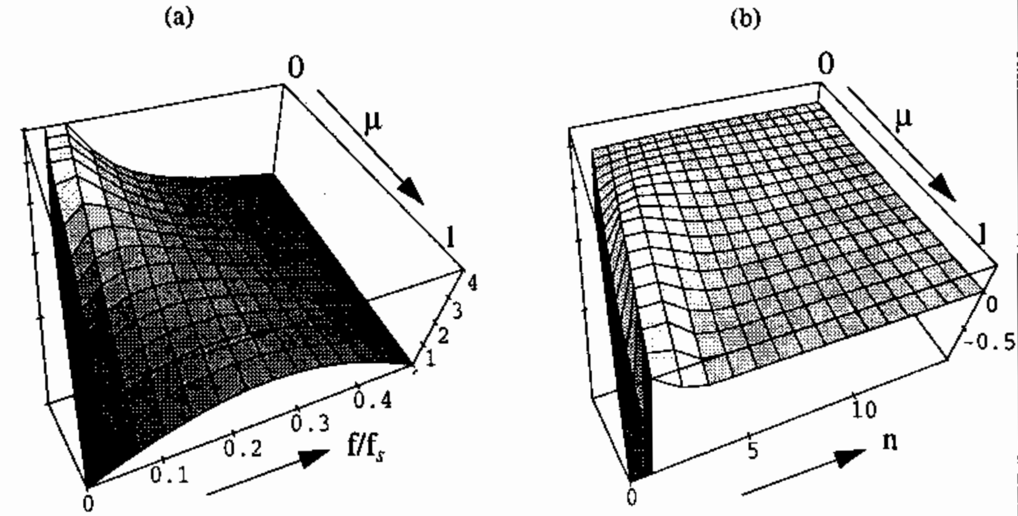

Figure 7 Frequency magnitude response (a) and impulse response (b) of gamma filter (Eq.35) as a function of μ.

6 DISCUSSION

In this paper the analytical development of a new class of adaptive filters -the gamma filters- has been presented. In FIR filter structures, filter memory depth and filter order are coupled. As a result, when long impulse responses are required in an FIR filter, the filter order must be high. Thus the FIR filter order usually exceeds the number of degrees of freedom of the system to be modeled, leading to poor modeling performance. In IIR filters these two aspects appear uncoupled. However, the simplicity of the adaptation of the FIR and its inherent stability are normally practical factors for the choice of the FIR over IIR designs.

The gamma filters implement a remarkable compromise between these two extremes. While the memory depth is adjustable independently from the filter order, the stability and adaptation characteristics of the gamma filter are similar to FIR structures. The error surface is still quadratic with respect to the filter weights $\{w_k\}$, but it is not convex in μ. As an experimental rule of thumb, we have observed that gradient descent adaptation of μ leads to the global minimum if we choose the initial value $\mu_0 = 1$.

We have shown by a system identification problem that the gamma filter outperforms the conventional FIR (adaline) of the same order. In general, the gamma filter is preferable if the processing problem involves signals with energy concentrated at low frequencies and relatively few degrees of freedom. Applications involving long delays as in channel equalization, room acoustics or identification of systems with long impulse responses seem to be particularly appropriate for the gamma filter. Yet the identification of application areas were the gamma filter outperforms adaline is still an open question. We are also currently investigating the practical importance of alternative delay operators $G(z)$.

Related work was recently presented by Amin who used gamma filter-like structures in spectral analysis (Amin, 1988). Although not explored here, it is possible to treat the generalized feedforward filter as an approximation problem using the basis functions $G_k(z)$. In this context, recently Perez and Tsujii (1991) showed that alternative basis functions (different from $G_k(z) = z^{-k}$) may indeed outperform conventional linear combiners, but the authors did not use an adaptive $G_k(z)$ and did not provide an extended framework.

Acknowledgments

This work was partially supported by NSF grant ECS-8915218 and DDM-8914084. The stay of the third author (P.G.O.) at the University of Florida was partially sponsored by JNICT.

References

Amin M., Sliding Spectra: a New Perspective, *Proc. 4th Ann. ASSP Workshop on Spectrum Estimation*, pp 55-59, 1988.

de Vries B. and Principe J.C., A Theory for Neural Nets with Time Delays, *NIPS-90 Proceedings, Lippmann R., Moody J., and Touretzky D. (eds.)*, San Mateo, CA, Morgan Kaufmann, 1991.

Haykin S., *Adaptive Filter Theory 2nd ed.*, Prentice-Hall, Englewood Cliffs, NJ, 1991

Oppenheim A. and Schafer R., *Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ,1975.

Perez H., Tsujii S., A System Identification Algorithm using Orthogonal Functions, *IEEE Trans. Signal Proc.*, vol 39, #3, 752-755, 1991.

Shynk J.J., Adaptive IIR Filtering, *IEEE ASSP Magazine*, pp. 4-21, April 1989.

Widrow B. and Stearns S., *Adaptive Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ,1985.

Wiener N., *Extrapolation, Interpolation and Smoothing of Stationary Time Series, with Engineering Applications*, New York, Wiley, 1949.

Williams R.J. and Zipser D., A Learning Algorithm for Continually Running Fully Recurrent Neural Networks, *Neural Computation, vol.1 no.2*, pp. 270-281, the MIT Press, Cambridge, MA, 1989.

Wolfram S., *Mathematica, A System for Doing Mathematics by Computer*, Addison-Wesley, Reading, MA, 1989.