The Gamma Filter—A New Class of Adaptive IIR Filters with Restricted Feedback

Jose C. Principe, *Senior Member, IEEE*, Bert de Vries, *Member, IEEE*, and Pedro G. de Oliveira, *Member, IEEE*

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

I. Introduction

INFINITE impulse response (IIR) filters are more efficient than finite impulse response (FIR) filters, but in adaptive signal processing, FIR systems are used almost exclusively [5], [12]. 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 nonconvex error surfaces of IIR systems [10].

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 low-pass 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 system: trivial stability, easy adaptation, yet the uncoupling of the region of support of the impulse response and the filter order.

A related filter architecture has been used in the past by Makhoul and Cosell [6] and Amin [1] in the context of spectral analysis. Other related work concerns the Laguerre filter, also a filter structure, although different, with a restricted feedback architecture [7]. Our approach differs from previous research because filters are analyzed as short-term memory structures. This viewpoint is inspired by our work in neural networks for temporal processing [3].

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.

II. Generalized Feedforward Filters—Definitions

Consider the IIR filter architecture described by

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

$$X_k(z) = G(z) X_{k-1}(z), \qquad k = 1, \cdots, K \tag{2}$$

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

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 be either recursive or nonrecursive. When $G(z) = z^{-1}$, this filter structure reduces to a transversal (feedforward) filter. The memory structure of a transversal filter is simply a tapped delay line. By iteration of (2) we can write $Y(z)$ as a function of the input $X(z)$ as a function of the input $X(z)$ as

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

[1] Read $\equiv$ as "is defined as."

Manuscript received June 17, 1991; revised January 6, 1992. This work was supported in part by NSF Grants ECS-8915218 and DDM-8914084. The work of P. Guedes de Oliveira at the University of Florida was supported in part by JNICT.

J. C. Principe and B. de Vries are with the Computational Neuro-engineering Laboratory, Department of Electrical Engineering, University of Florida, Gainesville, FL 32611.

P. G. de Oliveira is with the Departmento Eletronica/INESC, Universidade de Aveiro, Aveiro, Portugal.

IEEE Log Number 9205105.

1053-587X/93$03.00 © 1993 IEEE

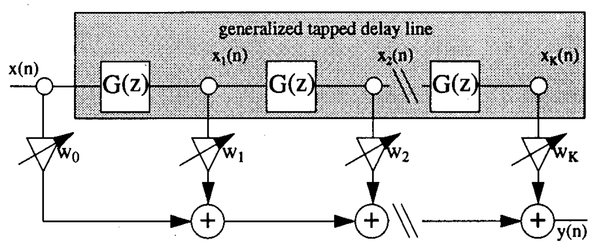


Fig. 1. The generalized feedforward filter.

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

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

It follows from (4) that $H(z)$ is stable whenever $G(z)$ is stable. As the filter structure $H(z)$ is in general more complex than the delay operator $G(z)$, it follows that stability is more easily controlled or guaranteed in the generalized feedforward filter when compared to an unrestricted recursive filter architecture. On the other hand, a judicious choice for the delay operator $G(z)$ may render the filter structure $H(z)$ with some of the desirable properties of recursive filter structures.

The past of $x(n)$ is represented in the tap variables $x_k(n)$ (shaded area in Fig. 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 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, controlled by a performance feedback mechanism such as the LMS algorithm. As a notational convenience, $G(z) = G(z; \mu)$ will be adopted.

This paper analyzes in detail the case $G(z) = \mu/(z - (1 - \mu))$, 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.

III. The Gamma Filter

A. Definitions

The gamma filter[2] is defined in the time domain as

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

$$x_k(n) = (1 - \mu)x_k(n - 1) + \mu x_{k-1}(n - 1), \quad k = 1, \cdots, K \tag{6}$$

where $x_0(n) \equiv x(n)$ is the input signal and $y(n)$ the filter output (Fig. 2). We will assume that the filter parameters $w_0, w_1, \cdots, w_k$ and μ are adaptive.

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

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

Inverse z-transformation yields the impulse response for tap k

$$g_k(n) \equiv Z^{-1}\{G_k(z)\} = \binom{n-1}{k-1} \mu^k (1 - \mu)^{n-k} U(n - k) \tag{8}$$

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

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

When $\mu = 1$, the gamma filter reduces to Widrow's *adaptive transversal filter* [12]. For $\mu \neq 1$, the gamma filter transfer function is of IIR type due to the recursion in (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 $1 - \mu$ is global (all feedback loops have the same gain).

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

B. Stability

Stability of the gamma filter is guaranteed if the poles are located within the unit circle. The gamma filter has a Kth order pole at $z_p = 1 - \mu$. As a result, the gamma filter is stable when $0 < \mu < 2$.

C. 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

---

[2] The gamma filter was originally developed in continuous time as part of a neural net model for temporal processing [3]. We showed by transformation $s = (z - 1/T_s)$ that 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 $g_k(t) = (\mu^k/(k - 1)!)t^{k-1}e^{-\mu t}$, $k = 1, \cdots, K$, 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 [13] as a very convenient basis for decomposition of linear systems in a signal processing context. In fact, the functions $g_k(t)$, $k = 1, \cdots, K$, 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 $L_2[0, \infty]$.

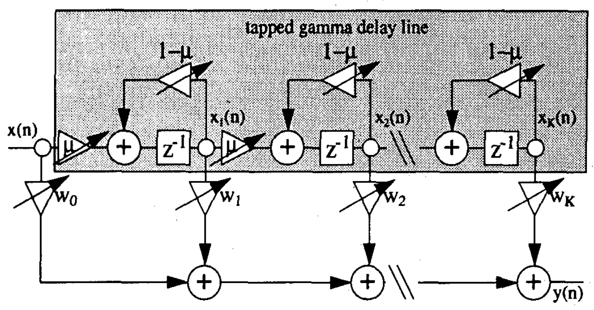


Fig. 2. The gamma filter structure.

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 quantify the notion of memory depth. As a convenient measure of memory depth for a Kth order gamma filter we take the first moment (mean value) of the last (Kth) delay kernel in the filter. Such a measure can be interpreted as the mean sampling time for the last tap. The mean memory depth D for the Kth order filter is thus defined as

$$D \equiv \sum_{n=0}^{\infty} n g_K(n) = Z\{n g_K(n)\}|_{z=l}$$
$$= -z \frac{dG_K(z)}{dz}\bigg|_{z=1} = \frac{K}{\mu}. \quad (10)$$

Next we define the (temporal) resolution R of the filter as the number of free parameters (i.e., the number of tap variables) per unit of time in the filter memory. This is equivalent to the number of taps (K) divided by the mean memory depth D. Thus

$$R \equiv \frac{K}{D} = \mu. \quad (11)$$

The following formula arises which is of fundamental importance for the characterization of the gamma memory structure:

$$K = D \times R. \quad (12)$$

Equation (12) reflects the possible tradeoff of resolution versus memory depth in a memory structure for fixed order K. Such a tradeoff is not possible in a nondispersive 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, the number of weights and depth D of the memory are equal. 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 uncouple the memory order and depth.

The depth of memory as a function of frequency is measured by the group delay. In Fig. 3 we have plotted the group delay of $G_2(z) = (\mu/z - (1 - \mu))^2$ (that is, the input-to-second-tap 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, favoring low frequencies in the memory can be efficient.

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 $i = 1, \cdots, 5$, and $\max_i (n_i) = 10$. If we try to model this signal with a regular FIR structure, the choice $K = 10$ leads to overfitting while $K < 10$ leaves the network unable to incorporate the influence of $x(t - 10)$. In a gamma filter, the choice $K = 5$ and $\mu = 0.5$ leads to 5 free network parameters and mean memory depth of 10, obviously a more efficient memory structure.

D. 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 to adapt a Kth order IIR filter with the LMS algorithm, $O(K^2)$ operations are necessary when the exact error gradients are utilized [12].

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

$$E \equiv \sum_{n=0}^{T} E_n = \sum_{n=0}^{T} \tfrac{1}{2} e^2(n)$$
$$= \sum_{n=0}^{T} \tfrac{1}{2} [d(n) - y(n)]^2 \quad (13)$$

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

$$\Delta w_k = -\eta \frac{\partial E}{\partial w_k} \quad (14)$$

$$\Delta \mu = -\eta \frac{\partial E}{\partial \mu} \quad (15)$$

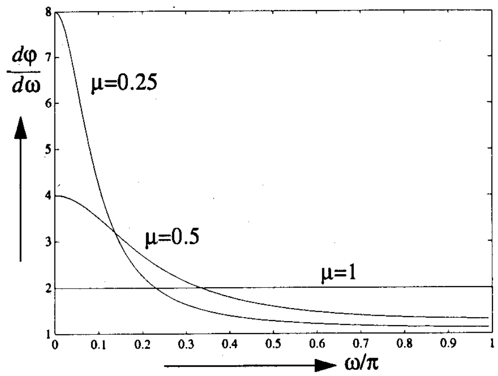


Fig. 3. Group delay of gamma memory at tap $k = 2$.

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

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

Similarly, the update equation for μ evaluates to

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

where $\alpha_k(n) \equiv (\partial x_k(n)/\partial \mu)$. The gradient signal $\alpha_k(n)$ can be computed on-line by differentiating (6) [10], [14], leading to

$$\begin{aligned}\alpha_0(n) &= 0\\ \alpha_k(n) &= (1-\mu)\alpha_k(n-1) + \mu\alpha_{k-1}(n-1)\\ &\quad + [x_{k-1}(n-1) - x_k(n-1)],\\ &\quad k = 1, \cdots, K.\end{aligned} \tag{18}$$

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

$$\Delta w_k(n) = \eta e(n) x_k(n), \qquad k = 0, \cdots, K \tag{19}$$

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

works well if η is sufficiently small. (Equation 19) can be recognized as the update term in the LMS algorithm. Notice that the number of operations per time step for (19) and (20) scale both as $O(K)$. Thus, the entire LMS algorithm scales as $O(K)$, which coincides with the complexity for Widrow's adaptive transversal filter. Equation (20) displays the same complexity as the IIR LMS routine per adaptive feedback parameter, but since the gamma filter has only a single feedback coefficient (the global μ) it has a complexity smaller than a K coefficient IIR filter (which scales as $O(K^2)$). IIR structures with trivial stability tests, such as the cascade of biquads or the lattice, have worse complexity of adaptation.

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

E. Wiener–Hopf Equations for the Gamma Filter

The optimal weights for an adaptive linear combiner in a given stationary environment can be analytically expressed by the Wiener–Hopf or normal equations [5]. Here these equations are extended to the gamma filter. We will show that the gamma normal equations generalize Wiener's formulation for strictly feedforward filters.

Consider the gamma filter structure as described by (5) and (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 $E[\cdot]$ the expectation operator. In order to maintain a consistent notation with respect to the adaptive signal processing literature, we introduce the vectors $X_n \equiv [x_0(n), x_1(n), \cdots, x_K(n)]^T$ and $W \equiv [w_0, w_1, \cdots, w_K]^T$. Note that X_n holds the tap variables and not the input signal samples. Evaluating ξ leads to

$$\xi = E[d^2(n)] + W^T R W - 2P^T W \tag{21}$$

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 $(\partial \xi / \partial w_k) = 0$ and $(\partial \xi / \partial \mu) = 0$ necessarily hold. Partial differentiation of (21) with respect to the system parameters yields the following results:

$$RW = P \tag{22}$$

and

$$W^T[R_\mu W - P_\mu] = 0 \tag{23}$$

where

$$R_\mu \equiv \frac{\partial R}{\partial \mu} = E\left[X_n \frac{\partial X_n^T}{\partial \mu}\right]$$

and

$$P_\mu \equiv \frac{\partial P}{\partial \mu} = E\left[d(n) \frac{\partial X_n}{\partial \mu}\right].$$

Note that (22) is the same expression as the Wiener–Hopf equation for the adaptive transversal filter. 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 (23) is a result of requiring $(\partial \xi / \partial \mu) =$

TABLE I
COMPARISON OF FIR, IIR, AND GAMMA FILTER PROPERTIES

Kth Order Filter	FIR	Gamma	IIR
Stability	Always stable	Trivial stability $0 < \mu < 2$	Nontrivial stability
Memory depth vs. order	Coupled K	Uncoupled K/μ	Free
Complexity of adaptation	$O(K)$	$O(K)$	$O(K^2)$

0. Thus, (23) provides an analytical expression for the optimal memory structure. This expression also reveals that the signal $\alpha_k(n) \equiv (\partial x_k(n)/\partial \mu)$ 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 gamma normal equations (22), (23) in terms of the input signal $x(n)$. Let us define the delay kernel vector $G(n) \equiv [g(n), g^2(n), \cdots, g^K(n)]^T$. Then (22) and (23) evaluate to

$$E\{[G(n) \bullet x(n)][G(n) \bullet x(n)]^T\}W = E\{d(n)[G(n) \bullet x(n)]\} \tag{24}$$

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

where $\bullet$ denotes the convolution operator.

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 as long as $G(z)$ is a rational function of z.

The optimal coefficients for the class of generalized feedforward filters can be computed from (22) (or (24)), augmented with a number of adjoint scalar equations (23) or (25), one for each adaptive parameter in $G(z)$.

IV. Experimental Results

We have presented two frameworks to obtain an optimal gamma filter architecture. In Section III-D the LMS adaptation algorithm was derived and Section III-E was devoted to the Wiener–Hopf equations for the gamma filter. In this section we present numerical simulation results for both optimization models when the gamma filter 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 adaptive linear combiner. Also, it will be shown that the filter coefficients w_k converge to the optimal values if the LMS update rules of Section III-D are used.

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

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

The performance index ξ as a function of μ was computed by evaluating (21) in the z domain (residue theorem). The optimal weight vector W^* is computed by solving the Wiener–Hopf equation (22), and substituting back into (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]. In this particular case where we model a low-pass filter, the range $0 < \mu < 1$ is most interesting despite the fact that the gamma filter is stable for a larger domain ($0 < \mu < 2$). All computations were performed with Mathematica [15] on a NeXT computer. The results are displayed in Fig. 4(a). Restrictions on the computational resources limited evaluations to $K \leq 3$. Note that these results present theoretical rather than experimental results since the Wiener–Hopf equations were solved analytically in the frequency domain. Observe that for all memory orders K the optimal performance is obtained for $\mu < 1$. Hence, the optimal gamma filter outperforms the conventional adaptive transversal filter by a large margin. Note that the optimal memory depth $D_{\text{opt}} \equiv K/\mu_{\text{opt}} \approx 5$ is constant for different memory orders.

In Fig. 4(b) we show the relative total error $J = \sigma_e^2/\sigma_d^2$ after convergence using the LMS update rule (19). We parametrized μ over the domain [0, 1] using a step size $\Delta\mu = 0.1$. The experimental results of Fig. 4(b) match the theoretical optimal performance (Fig. 4(a)) very well. This experiment shows that the filter weights $\{w_k\}$ can indeed be learned by on-line LMS learning. When $K = 5$, the adaptive linear combiner performs as well as a third-order ($K = 3$) gamma filter with $\mu = 0.6$. In practice we prefer the third-order gamma filter (5 free parameters) over the fifth-order transversal filter (6 parameters). Parsimony in the number of free parameters provides the gamma filter with better modeling (generalization) characteristics.

The effect of the memory parameter μ on the filter performance increases when we model a system with longer impulse response (smaller cutoff frequency) but the same number of parameters. In Fig. 5 the performance index ξ versus μ is plotted for a third-order elliptic low pass filter $H_2(z)$ with smaller cutoff frequency ($w_{co} = 0.06\pi$ rad). As in the previous experiments, we employed a system identification protocol and computed the optimal performance error as a function of μ by solving the Wiener–Hopf equations. The unknown system was given by

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

[3]This filter has been described in Oppenheim and Schafer [8].

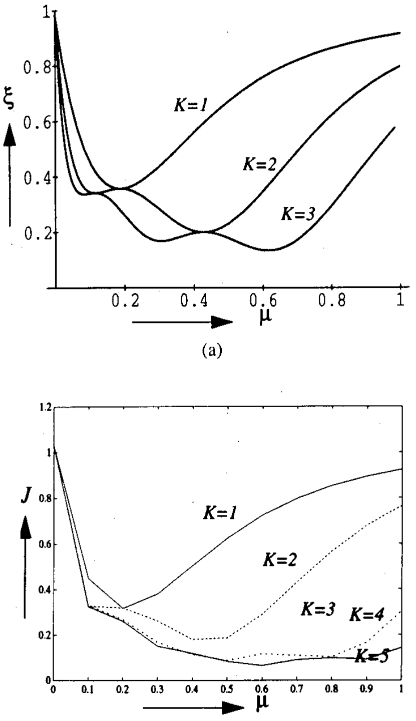


Fig. 4. Optimal performance index as a function of μ for identification of elliptic filter $H(z)$. (a) ξ is computed using from the Wiener–Hopf equations for the gamma filter. (b) $J_{\min} = \text{var}[e(n)]/\text{var}[d(n)]$ is computed after adaptation using the LMS update rule

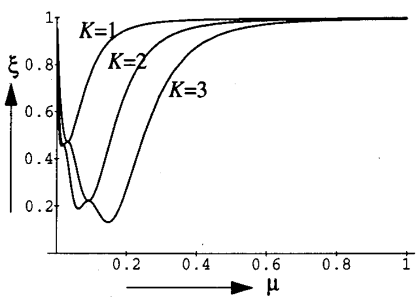


Fig. 5. ξ versus μ for $H_2(z)$.

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

We have experimented with several signals (sinusoids in noise, Feigenbaum map, electroencephalogram (EEG)) for various processing protocols (prediction, system identification, classification) [3]. Invariably the optimal memory structure[4] was obtained for $\mu < 1$. These data will be reported in a forthcoming publication.

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

V. The Gamma Transform

Thus far we have analyzed the 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

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

In the γ-domain, generalized feedforward filters are ordinary FIR filters, defined around delay operators γ^{-1}. The Gamma transformation is interesting as it provides a very appealing interpretation of the gamma filter in the transform domain. More importantly, when gamma filters are expressed in the gamma domain, feedforward filter design and analysis techniques can be applied directly.

For gamma filters, (28) evaluates to

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

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

$$\begin{aligned} X(\gamma) &\equiv X(z)|_{z = \mu\gamma + (1-\mu)} \\ &= \sum_{n=0}^{\infty} x(n) z^{-n} \Big|_{z = \mu\gamma + (1-\mu)} \\ &= \sum_{n=0}^{\infty} \mu^{-n} x(n) \left\{ \gamma + \frac{1-\mu}{\mu} \right\}^{-n}. \end{aligned} \tag{30}$$

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

The corresponding time series obtained by the inverse gamma transform can be computed as

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

or

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

where C is a closed contour that encircles the point γ_0. Equations (30) and (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 γ-domain.

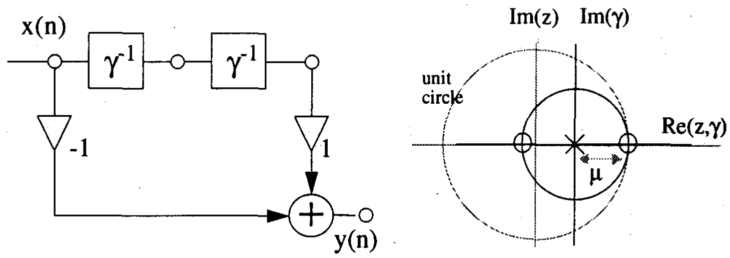


Fig. 6. A second-order gamma filter, and the relation between z and γ domains.

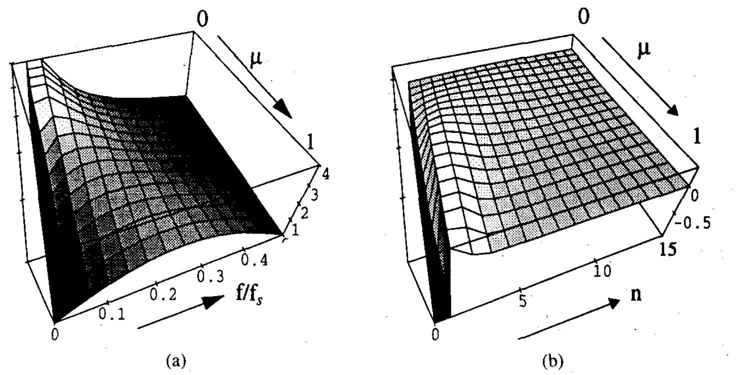


Fig. 7. (a) Frequency magnitude response and (b) impulse response of gamma filter (35) as a function of μ.

As an example, let us analyze a simple second-order gamma filter with $w_0 = -1$, $w_1 = 0$ and $w_2 = 1$ (Fig. 6). Note that the pole(s) of a gamma filter are located at the origin in the γ plane. The zeros are located at $\gamma = -1$ and $\gamma = 1$. Thus, the system is feedforward in the γ-domain. The transfer function in the γ-domain is easily obtained by inspection:

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

Substitution of (29) gives for the transfer function in the z domain

$$H(z) = -1 + \left\{\frac{z - (1 - \mu)}{\mu}\right\}^{-2}. \qquad (34)$$

The impulse response of a gamma filter can be expressed as

$$\begin{aligned} h(n) &= \sum_{k=0}^{K} w_k g_k(n) \\ &= -\delta(n) + (n - 1)\mu^2(1 - \mu)^{n-2}U(n - 2). \qquad (35) \end{aligned}$$

In Fig. 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.

VI. 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 $\mu_0 = 1$. The structure of the error surface as a function of μ (Fig. 4) is in general not convex which potentially leads to adaptation problems for a simple gradient descent algorithm. We are currently studying this problem.

We have shown with a system identification problem that the gamma filter outperforms the conventional adaptive transversal filter 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 where 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 has been conducted in the area of Laguerre filters where the delay operators are first-order all-pass filters [7], [11]. The work in this paper, however, takes a different viewpoint. We have analyzed filter structures as networks for storing history traces of a signal, that is, in terms of its computational properties as a short-term memory structure. Second, unlike previous work, we present the theoretical framework for adaptation of the memory structure by gradient descent in order to match the filter memory to the input signal.

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 [9] showed that alternative basis functions (different from $G_k(z) = z^{-k}$) may indeed outperform conventional transversal filters, but the authors did not use an adaptive $G_k(z)$ and did not provide an extended framework.

In the neural network community it is common to speak of the adaline (adaptive linear neuron) structure instead of adaptive transversal filter. In related work, we have also referred to the adaptive gamma filter as adaline (μ) [4]. In this terminology, adaline(1) is equivalent to Widrow's adaline.

References

[1] M. Amin, "Sliding spectra: A new perspective" in *Proc. 4th Annu. ASSP Workshop Spectrum Estimation*, 1988, pp. 55–59.

[2] B. De Vries, "Temporal processing with neural networks—the development of the gamma model," Ph.D. dissertation, Univ. Florida, 1991.

[3] B. De Vries and J. C. Principe, "A theory for neural nets with time delays," in *NIPS-90 Proceedings*, R. Lippmann, J. Moody, and D. Touretzky, eds. San Mateo, CA: Morgan Kaufmann, 1991.

[4] B. De Vries, J. C. Principe, and P. G. De Oliveira, "Adaline with adaptive recursive memory," in *Proc. IEEE Workshop Neural Networks Signal Processing*, Princeton, NJ, Sept. 29–Oct. 2, 1991.

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

[6] J. I. Makhoul and L. K. Cosell, "Adaptive lattice analysis of speech," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 29, no. 3, June 1981.

[7] M. A. Masnadi-Shirazi and N. Ahmed, "Optimum Laguerre networks for a class of discrete-time systems," *IEEE Trans. Signal Processing*, vol. 39, no. 9, pp. 2104–2108, 1991.

[8] A. Oppenheim and R. Schafer, *Digital Signal Processing*. Englewood Cliffs, NJ: Prentice Hall, 1975.

[9] H. Perez and S. Tsujii, "A system identification algorithm using orthogonal functions," *IEEE Trans. Signal Processing*, vol. 39, no. 3, pp. 752–755, 1991.

[10] J. J. Shynk, "Adaptive IIR filtering," *IEEE ASSP Mag.*, pp. 4–21, Apr. 1989.

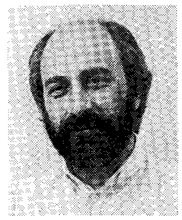
[11] H. W. Strube, "Linear prediction on a warped frequency scale," *J. Acoust. Soc. Amer.*, vol. 68, 1071–1076, 1980.

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

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

[14] R. J. Williams and D. Zipser, "A learning algorithm for continually running fully recurrent neural networks," *Neural Computation*, vol. 1, no. 2, pp. 270–281, MA, 1989.

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



Jose C. Principe (M'83–SM'90) was born in Porto, Portugal, in 1950. He received the Engenheiro degree from the University of Porto, Portugal, in 1972 and the M. Sc. and Ph.D. degrees from the University of Florida, Gainesville, in 1975 and 1979, respectively.

He has been a faculty member of the University of Aveiro, Portugal, since 1980, where he is a Professor of Electrical Engineering. He is a senior member of INESC, Portugal. He joined the Faculty of the Electrical Engineering Department of the University of Florida in 1987, and is the Director of the Computational NeuroEngineering Laboratory. His interests are centered in adaptive signal processing and nonlinear dynamical modeling (with nonlinear dynamical systems) of biological signals, primarily the electroencephalogram, as well as neural networks and machine intelligence.

Bert de Vries (S'86–M'92) was born in Utrecht, The Netherlands. He received the Ingenieur degree in 1986 from the Technical University Eindhoven, The Netherlands, and the Ph.D. degree in electrical engineering in December 1991 from the University of Florida, Gainesville.

Currently he is working for David Sarnoff Research Center in Princeton, NJ. His interests are digital signal processing and neural network modeling of time varying signals.

Pedro G. de Oliveira (M'87) was born in Oporto, Portugal, in 1945. He received the B.S.E.E. degree from the University of Oporto and the Ph.D. degree from the University of Aveiro, Portugal.

He was a Guest Researcher at the Institute of Medical Physics, TNO, in Utrecht, The Netherlands, in 1978–1979 and spent part of his sabbatical leave in 1991 at the Computational Neuro-Engineering Lab of the Department of Electrical Engineering, University of Florida, Gainesville.

He is a Professor at the University of Aveiro and is also a member of the Executive Board of INESC (a nonprofit Research Institute, belonging to three Portuguese Telecom operators and the University of Porto, University of Aveiro, and the Technical University of Lisbon). His research interests are biological signal processing, neural networks, and analog circuits.