Laguerre Filters – An Introduction

Tomás Oliveira e Silva

Abstract – In this tutorial paper we present a generalization of the transversal filter, called Laguerre filter, and study some of its more remarkable properties. This filter is obtained by replacing each delay of the transversal filter by a first order all-pass section, and by applying a first order low-pass filter (with the same pole used in the all-pass sections) to the filter's input signal. Both the transversal and the lattice forms of the Laguerre filter are discussed. We also deduce the stationarity conditions of the mean-square error of a Laguerre filter (transversal or lattice) with respect to its pole position.

Resumo – Neste trabalho apresentamos uma generalização dos filtros transversais, os chamados filtros de Laguerre, e estudamos algumas das suas propriedades mais notáveis. Estes filtros são obtidos substituindo cada atraso dos filtros transversais por uma secção passa tudo de primeira ordem e pré-processando o sinal de entrada do filtro com um filtro passa baixo de primeira ordem (com o mesmo pólo das secções passa tudo). São estudadas as formas transversal e "lattice" do filtro de Laguerre. Deduzimos também as condições de estacionaridade do erro quadrático médio de um filtro de Laguerre (forma transversal ou "lattice") em relação à posição do seu pólo.

I. Introduction

The transversal filter and some other filter structures related with it, such as the lattice filter, are very popular among the models of linear systems, specially if adaptation of its parameters is desired [1]-[3]. Some applications where these filters have attained considerable success include, among others, system identification, linear prediction, channel equalization, and echo cancellation. The reason for this success is, besides the simplicity of the transversal filter structure, the unimodality of its error surface, and the existence of fast and efficient adaptive algorithms to adjust its parameters [1]-[4].

The principal problem of the transversal filter, which is also related to its advantages, is that its impulse response has a finite duration (it is a FIR filter). For this reason, when this filter is used to approximate a system with a long (possibly infinite) impulse response the minimum number of delays of the filter required to provide an acceptable approximation can be quite high. This problem can be partially solved using filters with an infinite impulse response (IIR filters). However, these filters have their own problems, specially if output error models are used [5], [6]. Among these are possible multimodal error surfaces [7], and possible instability problems related to the adaptation of the poles of these filters [6].

Another disadvantage of the transversal filter is that its continuous-time (analog) version requires delay lines,

which are difficult to implement. In an attempt to solve this problem, in [8] each delay of the transversal filter was replaced by an all-pass filter. This preserves many of the properties of transversal filters and gives rise to continuous-time (and discrete-time) generalizations of the transversal filters that have infinite impulse responses. If the all-pass filter is chosen properly, these filters are usually able to produce acceptable approximations of systems with long impulse responses with a much smaller number of parameters that a transversal filter.

The Laguerre filter, which is another generalization of the transversal filter, has its roots in the pioneering work of Wiener and Lee concerning the synthesis of electric networks using Laguerre functions [9], [10]. The early papers about this subject used truncated Laguerre series to approximate the impulse response of a given continuous-time system [11]-[18]. The discrete-time counterparts of these papers, based on the Laguerre sequences [19]-[21], appeared some years later, and gave rise to the so-called Laguerre filters [22]-[30]. In the last few years Laguerre models (or filters) were applied successfully to several problems in the automatic control field [31]-[37], [24], [38]-[40]. Other recent applications of the Laguerre functions and sequences in signal processing can be found in [41]-[43].

The main advantage of the Laguerre filter in relation to the transversal filter is that the former is an IIR filter with one adjustable multiple pole and the latter is a FIR filter with a fixed multiple pole at the origin. As we will see later on, if the pole of the Laguerre filter is placed at the origin the Laguerre filter degenerates into the transversal filter, i.e., we may consider the Laguerre filter to be a generalization of the transversal filter. By adjusting the pole position of the Laguerre filter it is possible to control the rate of decay of its impulse response, which is quite useful to provide good approximations of systems with long impulse responses.

Due to space limitations we will only discuss in this paper the discrete-time Laguerre filters. Similar results can be easily obtained for the continuous-time Laguerre filters, which are left as an exercise to the interested reader (see also [44]). For the same reason we will also not discuss here the adaptation of the weights and of the pole of Laguerre filters.

The structure of this paper is the following. In section II we review some mathematical material necessary for the understanding of this paper. In section III we describe briefly the main properties of transversal filters. These filters are then generalized in section IV, giving rise to the so-called Laguerre filters. As these filters have one additional parameter, the Laguerre pole position, which affects considerably their performance, we present in subsection IV-A a simple condition that the optimal value of this parameter

must satisfy. In section V we introduce the lattice form of the Laguerre filter, which is the base of another proof of the "optimality" condition for the Laguerre pole position. In section VI we present a simple example that illustrates some of the results of this paper. Finally, in section VII we summarize the contents of this paper and describe briefly other generalizations of transversal filters.

II. Notation, definitions, and some useful facts about Hilbert space theory and least mean-square approximations

The majority of the definitions and results presented in this section can be found in [45], [3], [46]-[48].

We will denote matrices and vectors respectively by upper and lower case bold letters. The indexes of the elements of matrices and vectors will start from zero and not from one. The letter k will be the discrete time variable. The Kronecker's delta will be denoted by δ_{ij} (it is equal to one if $i = j$ and equal to zero otherwise).

A. The Hilbert space ℓ^2

A Hilbert space is an inner product space which is a complete metric space with respect to the metric induced by its inner product [46]. This means that a Hilbert space is a linear vector space, possibly of infinite dimension, with an inner product operation defined between any two of its elements (an inner product space). This inner product is used to define the norm of an element of that space (norm induced by the inner product), which is simply the square root of the inner product of that element with itself. This norm is used in turn to define a distance (metric) between any two elements of that space (a metric space), that is the norm of the difference between these two elements. The remaining characteristic of a Hilbert space is that the metric space is complete (or closed). This means that any convergent (Cauchy) sequence of elements of that space converges to an element of that space. Two elements of a Hilbert space are said to be orthogonal if their inner product is zero. An element of a Hilbert space is said to be normal if its norm is equal to one. For an introductory exposition of Hilbert spaces we refer the reader to [46].

The problem of finding the best approximation of an arbitrary element of a Hilbert space by an element of a (linear) subspace of that Hilbert space is solved by the principle of orthogonality (a consequence of the projection theorem [46], [47]), which states that the error of the (unique) best approximation is orthogonal to the subspace in question.

A real sequence $f(k)$ belongs to the Hilbert space ℓ^2, the space of all square-summable causal sequences, if and only if [46]

$$\sum_{k=0}^{+\infty} f^2(k) < \infty.$$

Note that all absolutely summable causal sequences belong to ℓ^2 [45], i.e., the impulse responses of all causal stable linear systems belong to this Hilbert space.

The Fourier transform of a sequence $f(k)$ belonging to ℓ^2 is defined by[1]

$$F(e^{j\omega}) = \sum_{k=0}^{+\infty} f(k) e^{-j\omega k}.$$

The function $F(e^{j\omega})$ is a square-integrable function, in the sense of Lebesgue, on the unit circle [48].

The inner product between any two sequences, $f(k)$ and $g(k)$, of ℓ^2 is defined by

$$\langle f(k), g(k) \rangle = \sum_{k=0}^{+\infty} f(k)g(k).$$

Because both $F(e^{j\omega})$ and $G(e^{j\omega})$ are square-integrable functions on the unit circle we may also evaluate this inner product by the formula (Parseval's theorem)

$$\langle f(k), g(k) \rangle = \frac{1}{2\pi} \int_{-\pi}^{+\pi} F(e^{j\omega})G^*(e^{j\omega}) \, d\omega.$$

In particular,

$$\langle f(k), f(k) \rangle = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \left| F(e^{j\omega}) \right|^2 \, d\omega. \tag{1}$$

A set of sequences of ℓ^2 is said to be complete if *any* sequence of that Hilbert space can be approximated arbitrarily well (in the norm induced by the inner product) by a linear combination of the sequences of that set. If these sequences are orthonormal (both orthogonal and normal) then the set is called an orthonormal basis of ℓ^2.

Let $\{f_i(k)\}_{i=0}^{+\infty}$ be an orthonormal basis of ℓ^2. Then, any sequence $g(k)$ belonging to ℓ^2 can be expanded in the form (orthonormal expansion)

$$g(k) = \sum_{i=0}^{+\infty} c_i f_i(k)$$

where $c_i = \langle g(k), f_i(k) \rangle$ are the Fourier coefficients of $g(k)$ with respect to the orthonormal set $\{f_i(k)\}_{i=0}^{+\infty}$.

The canonical basis of ℓ^2 is the orthonormal set of sequences $\{\delta(k-i)\}_{i=0}^{+\infty}$, where $\delta(k-i) = \delta_{ki}$ are the "pulse sequences" (these sequences are nonzero only for $k = i$). The canonical basis is the simplest example of a complete orthonormal set of ℓ^2.

B. Stochastic processes

Let X and Y be two real random variables with zero mean and finite variance. The inner product between these two random variables is defined by their covariance (or correlation), i.e., by

$$\langle X, Y \rangle = \mathrm{E}[XY]$$

where $\mathrm{E}[\cdot]$ denotes mathematical expectation. In particular, $\langle X, X \rangle = \mathrm{Var}[X]$ is the variance (mean-square value) of X.

---

[1]The symbol j (not to be confused with j) denotes the square root of -1.

Let $x(k)$ and $y(k)$ be two real wide-sense stationary stochastic processes with zero mean and finite variance.[2] The cross-correlation function between these two stochastic processes is defined by

$$R_{xy}(\tau) = \mathrm{E}[x(k+\tau)y(k)].$$

Because these stochastic processes are wide-sense stationary their cross-correlation does not depend on the time variable k. The cross-power spectral density between $x(k)$ and $y(k)$ is the Fourier transform of $R_{xy}(\tau)$, given by

$$\Phi_{xy}(e^{\mathrm{j}\omega}) = \sum_{\tau=-\infty}^{+\infty} R_{xy}(\tau)\, e^{-\mathrm{j}\omega\tau}.$$

It should be stressed that this expression must be used with care, because it may not converge for some values of ω (on a set of measure zero). If $R_{xy}(\tau)$ is an absolutely summable sequence $\Phi_{xy}(e^{\mathrm{j}\omega})$ can be considered to be the z-transform of $R_{xy}(\tau)$ evaluated on the unit circle. (Note that in this case $\Phi_{xy}(z)$ need not be analytic on the unit circle. Without stronger conditions the best that can be said is that it converges uniformly there.)

It is possible to recover the cross-correlation function from the cross-spectral density by the formula

$$R_{xy}(\tau) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} e^{\mathrm{j}\omega\tau} \Phi_{xy}(e^{\mathrm{j}\omega})\, d\omega.$$

Due to the aforementioned possible convergence problems of $\Phi_{xy}(e^{\mathrm{j}\omega})$ this integral must be evaluated with care.[3]

The functions $R_{xx}(\tau)$ and $\Phi_{xx}(e^{\mathrm{j}\omega})$ are called respectively autocorrelation and power spectral density of the stochastic process $x(k)$. Note that the variance of $x(k)$ is given by

$$\mathrm{Var}[x(k)] = R_{xx}(0) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \Phi_{xx}(e^{\mathrm{j}\omega})\, d\omega.$$

Let $H(z)$ be the transfer function of a stable linear system, $x(k)$ its input, and $y(k)$ its output. The input-output relation of this system has to be expressed in the time domain when $x(k)$ (and hence $y(k)$) is a stochastic process, because is is not clear how to apply z-transforms to this kind of signals (see figure 1). It is easy to verify that the power spectral density of $y(k)$ is given by [45]

$$\Phi_{yy}(e^{\mathrm{j}\omega}) = \left|H(e^{\mathrm{j}\omega})\right|^2 \Phi_{xx}(e^{\mathrm{j}\omega}).$$

[2] Note that in this case $x(k)$ and $y(k)$ are sequences of random variables. We will reserve the letters f, g, and h, to represent nonrandom sequences.

[3] Strictly speaking, the above integral should be replaced by the following Stieltjes integral [47]

$$R_{xy}(\tau) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} e^{\mathrm{j}\omega\tau}\, d\Psi_{xy}(\omega)$$

where

$$\Psi_{xy}(\omega) = \int_{-\pi}^{\omega} \Phi_{xy}(e^{\mathrm{j}\nu})\, d\nu$$

is the cross-spectral distribution function. This formalism accounts for possible impulses (Dirac delta distributions) in $\Phi_{xy}(e^{\mathrm{j}\omega})$ without resorting to the theory of distributions. We will avoid such technicalities here.
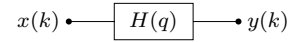

Fig. 1 - Block diagram of a causal stable system with transfer function $H(z) = \sum_{i=0}^{+\infty} h(i)z^{-i}$. The symbol q is the advance operator, i.e., $q[x(k)] = x(k+1)$, and q^{-1} is its inverse (the delay operator). The output of this system is given by the convolution of $x(k)$ with $h(k)$, i.e., by $y(k) = \sum_{i=0}^{+\infty} h(i)x(k-i) = H(q)x(k)$.

Because $H(z)$ converges uniformly and is bounded on the unit circle if the system is stable (remember that $h(k)$ is in this case an absolutely summable sequence), it is easy to verify that if $x(k)$ has zero mean and finite variance then so will $y(k)$. Therefore, the variance of $y(k)$ is given by

$$\langle y(k),y(k)\rangle = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \left|H(e^{\mathrm{j}\omega})\right|^2 \Phi_{xx}(e^{\mathrm{j}\omega})\, d\omega. \quad (2)$$

Let $F(z)$ and $G(z)$ be two stable linear systems excited respectively by the stochastic processes $x(k)$ and $y(k)$, assumed to be correlated, with outputs $u(k) = F(q)x(k)$ and $v(k) = G(q)y(k)$, respectively. The inner product between $u(k)$ and $v(k)$ is given by

$$
\begin{aligned}
\langle u(k),v(k)\rangle &= \sum_{i,j=0}^{+\infty} f(i)\mathrm{E}[x(k-i)y(k-j)]g(j) \\
&= \sum_{i,j=0}^{+\infty} f(i) \left[\frac{1}{2\pi} \int_{-\pi}^{+\pi} e^{\mathrm{j}\omega(j-i)}\Phi_{xy}(e^{\mathrm{j}\omega})\, d\omega\right] g(j) \\
&= \frac{1}{2\pi} \int_{-\pi}^{+\pi} F(e^{\mathrm{j}\omega})G^*(e^{\mathrm{j}\omega})\, \Phi_{xy}(e^{\mathrm{j}\omega})\, d\omega. \quad (3)
\end{aligned}
$$

This formula will be useful later on. Note that (2) is a special case of (3).

C. Least mean-square approximations

Consider the problem of the approximation of a random variable Y by a linear combination of $n+1$ other random variables $X_0, \ldots, X_n$, correlated with Y, such that the variance (mean-square) of the approximation error is as small as possible. Let Y_n be the approximation of Y, given by

$$Y_n = \sum_{i=0}^{n} w_{n,i}X_i,$$

and let $E_n = Y - Y_n$ be the approximation error, whose variance is $\xi_n = \langle E_n, E_n\rangle$. This approximation problem is naturally formulated and solved in the context of Hilbert spaces, in this case the Hilbert space of random variables with zero mean and finite variance [47].

In our specific problem the subspace where the approximation lies is composed of all linear combinations of the random variables $X_0, \ldots, X_n$, and the principle of orthogonality states that

$$\langle E_n, X_i\rangle = 0, \qquad i = 0, \ldots, n. \quad (4)$$

These equations are usually called normal equations and can be deduced without resorting to Hilbert space theory, equating the partial derivatives of ξ_n with respect to each of the $w_{n,i}$'s to zero. Assuming that the normal equations are

satisfied the (least) mean-square error of the approximation is given by

$$\xi_n = \langle E_n, Y \rangle = \langle Y, Y \rangle - \sum_{i=0}^{n} w_{n,i} \langle Y, X_i \rangle. \qquad (5)$$

The normal equations (4) can be put together in only one equation, of the form

$$\begin{bmatrix} \langle X_0, X_0 \rangle & \cdots & \langle X_0, X_n \rangle \\ \vdots & \ddots & \vdots \\ \langle X_n, X_0 \rangle & \cdots & \langle X_n, X_n \rangle \end{bmatrix} \begin{bmatrix} w_{n,0} \\ \vdots \\ w_{n,n} \end{bmatrix} = \begin{bmatrix} \langle Y, X_0 \rangle \\ \vdots \\ \langle Y, X_n \rangle \end{bmatrix}.$$

This equation can be written in the condensed form $\mathbf{R}_n \mathbf{w}_n = \mathbf{p}_n$, with obvious definitions for the square matrix $\mathbf{R}_n$, and for the vectors $\mathbf{w}_n$ and $\mathbf{p}_n$. The matrix $\mathbf{R}_n$ is symmetric and nonnegative definite. This second fact is a trivial consequence of

$$\mathbf{w}_n^{\mathrm{T}} \mathbf{R}_n \mathbf{w}_n = \langle Y_n, Y_n \rangle \geq 0. \qquad (6)$$

A very important case of the approximation problem studied here occurs when $\langle X_i, X_j \rangle = \delta_{ij}$. In this case $\{X_i\}_{i=0}^{n}$ is an orthonormal set and $\mathbf{R}_n$ is the $n+1 \times n+1$ identity matrix, which implies that $w_{n,i} = \langle Y, X_i \rangle \overset{\triangle}{=} c_i$ does not depend on n. The best approximation to Y is then given by the simple formula

$$Y_n = \sum_{i=0}^{n} c_i X_i,$$

and the least mean-square error of the approximation is given by

$$\xi_n = \langle Y, Y \rangle - \sum_{i=0}^{n} c_i^2.$$

Note that $Y_n = Y_{n-1} + c_n X_n$, and that $\xi_n = \xi_{n-1} - c_n^2$, with the initial values $Y_{-1} = 0$ and $\xi_{-1} = \langle Y, Y \rangle$.

From a given linearly independent set $\{X_i\}_{i=0}^{n}$ it is possible to construct an orthogonal set $\{X_i^b\}_{i=0}^{n}$ where $X_i^b = \sum_{j=0}^{i} b_{ij} X_j$ with the restriction $b_{ii} = 1$, using the Gram-Schmidt orthogonalization procedure [49]. Note that the constants b_{ij} are such that $\langle X_i^b, X_j^b \rangle = 0$ for $0 \leq j < i$, which in turn implies that $\langle X_i^b, X_j \rangle = 0$ also for $0 \leq j < i$. Remembering the orthogonality principle of best approximations in Hilbert spaces it is easy to verify that X_i^b is precisely the error of the best approximation of X_i by a linear combination of the random variables $X_0, \ldots, X_{i-1}$.

D. Approximations of linear systems

Consider the problem of the approximation of a given stable and causal system $H(z)$ by another stable and causal system $H_n(z)$. (The exact form of $H_n(z)$ is irrelevant to the present discussion.) Both systems are excited by the same stochastic process $x(k)$ and the objective of the approximation is to minimize the variance of the error signal $e_n(k)$, which is the difference between the outputs of both
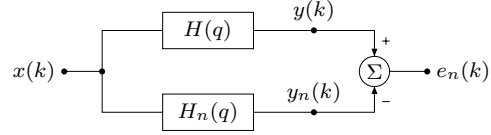

Fig. 2 - Model of the approximation of a general system, $H(z)$, by another system, $H_n(z)$. The objective of the approximation is the minimization of the variance of $e_n(k)$ by adjusting some parameters of $H_n(z)$.
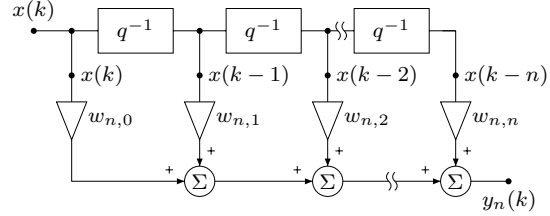

Fig. 3 - Transversal filter of order n (with n delays). The output of this filter is given by $y_n(k) = \sum_{i=0}^{n} w_{n,i} \, x(k-i)$, and is used to approximate a given desired signal, $y(k)$, correlated with $x(k)$.

systems. This situation is depicted in figure 2. Applying (2) to this case gives for the variance of $e_n(k)$ the formula

$$\xi_n = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \left| H(e^{j\omega}) - H_n(e^{j\omega}) \right|^2 \Phi_{xx}(e^{j\omega}) \, d\omega. \quad (7)$$

We are interested in the comparison of this formula with the following ℓ^2 inner product

$$\frac{1}{2\pi} \int_{-\pi}^{+\pi} \left| H(e^{j\omega}) - H_n(e^{j\omega}) \right|^2 \, d\omega. \qquad (8)$$

If $\Phi_{xx}(e^{j\omega})$ is an essentially bounded Lebesgue measurable function on the interval $[-\pi, +\pi]$ and if (8) converges to zero when n goes to infinity then (7) will also converge to zero. For example, this will happen if $R_{xx}(\tau)$ is absolutely summable and if $h_n(k)$, the impulse response of $H_n(z)$, is a linear combination of the first $n + 1$ sequences of a complete set of ℓ^2. (Obviously, each one of the sequences of that set must be absolutely summable, otherwise $H_n(z)$ could be unstable.) For a general power spectral density the same result holds if $H_n(e^{j\omega})$ converges to $H(e^{j\omega})$ for all frequencies where $\Phi_{xx}(e^{j\omega})$ has a Dirac delta distribution. For example, this will happen if the complete set used to form $h_n(k)$ is the canonical basis of ℓ^2.

Note that (7) may be null even when (8) is non-null if $\Phi_{xx}(e^{j\omega})$ vanishes on a set of $[-\pi, +\pi]$ with nonzero measure, i.e., if $x(k)$ is a band limited process. This cannot happen if $\Phi_{xx}(e^{j\omega})$ is strictly positive for (almost) all $\omega \in [-\pi, +\pi]$, a condition usually called persistence of excitation (of infinite order) [50].

III. THE TRANSVERSAL FILTER

Consider the transversal filter of figure 3. The weights of this filter that minimize the variance of the error of the approximation of a given desired signal $y(k)$, correlated with $x(k)$, by $y_n(k)$ satisfy the Wiener-Hopf equations [3] (cf. the normal equations (4))

$$\sum_{i=0}^{n} w_{n,i} \, \langle x(k-i), x(k-j) \rangle = \langle y(k), x(k-j) \rangle$$

for $0 \leq j \leq n$. These equations can be put in the form

$$\mathbf{R}_n \mathbf{w}_n = \mathbf{p}_n \qquad (9)$$

where the elements of $\mathbf{R}_n$, which can be computed easily using (3), are given by

$$
\begin{aligned}
r_{ij} &= \langle x(k-i), x(k-j) \rangle \\
&= \frac{1}{2\pi} \int_{-\pi}^{+\pi} e^{\mathrm{j}\omega(j-i)} \Phi_{xx}(e^{\mathrm{j}\omega}) \, d\omega \qquad (10)
\end{aligned}
$$

$(0 \leq i, j \leq n)$, and where those of $\mathbf{p}_n$ are given by

$$
\begin{aligned}
p_i &= \langle y(k), x(k-i) \rangle \\
&= \frac{1}{2\pi} \int_{-\pi}^{+\pi} e^{\mathrm{j}\omega i} \Phi_{yx}(e^{\mathrm{j}\omega}) \, d\omega
\end{aligned}
$$

$(0 \leq i \leq n)$. Note that r_{ij} only depends on $|i - j|$. This means that the elements of each diagonal of $\mathbf{R}_n$ are equal, i.e., $\mathbf{R}_n$ is a symmetric Toeplitz matrix. It is possible to explore the Toeplitz structure of this matrix to solve (9), as done for example in the Levinson algorithm [3], [49]. This gives rise to the so-called lattice filters, which we will discuss later on in the context of Laguerre filters.

The smallest and largest eigenvalues of $\mathbf{R}_n$ play an important role not only in the resolution of the Wiener-Hopf equations, where they define the numerical stability of the system of normal equations [49], but also in the convergence speed of certain adaptive algorithms of the weights of transversal filters [3]. Fortunately, it is easy to obtain simple bounds for these eigenvalues if we restrict the input signal of the transversal filter to have an absolutely summable autocorrelation function (less restrictive results can be found in [51]). In this case the power spectral density converges uniformly (and is bounded) on the interval $\omega \in [-\pi, +\pi]$.

A well know method to compute the smallest and largest eigenvalues of a symmetric matrix, in this case $\mathbf{R}_n$, is to evaluate the minimum and maximum values of the Rayleigh quotient [49]:

$$\lambda_{\min}(\mathbf{R}_n) = \min_{\mathbf{w}_n \neq \mathbf{0}} \frac{\mathbf{w}_n^{\mathrm{T}} \mathbf{R}_n \mathbf{w}_n}{\mathbf{w}_n^{\mathrm{T}} \mathbf{w}_n},$$

$$\lambda_{\max}(\mathbf{R}_n) = \max_{\mathbf{w}_n \neq \mathbf{0}} \frac{\mathbf{w}_n^{\mathrm{T}} \mathbf{R}_n \mathbf{w}_n}{\mathbf{w}_n^{\mathrm{T}} \mathbf{w}_n},$$

where $\mathbf{w}_n$ is an arbitrary nonnull vector with $n+1$ elements. In our concrete case it is easy to verify that (cf. (6) and (2))

$$\mathbf{w}_n^{\mathrm{T}} \mathbf{R}_n \mathbf{w}_n = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \left| \sum_{i=0}^{n} w_{n,i} \, e^{-\mathrm{j}\omega i} \right|^2 \Phi_{xx}(e^{\mathrm{j}\omega}) \, d\omega,$$

and that (because the functions $e^{-\mathrm{j}\omega i}$ are orthonormal in the interval $[-\pi, +\pi]$)

$$\mathbf{w}_n^{\mathrm{T}} \mathbf{w}_n = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \left| \sum_{i=0}^{n} w_{n,i} \, e^{-\mathrm{j}\omega i} \right|^2 d\omega.$$

Using trivial lower and upper bounds for the power spectral density in the first of these two formulas it is easy to conclude that

$$\inf_{\omega} \Phi_{xx}(e^{\mathrm{j}\omega}) \leq \lambda_{\min}(\mathbf{R}_n) \leq \lambda_{\max}(\mathbf{R}_n) \leq \sup_{\omega} \Phi_{xx}(e^{\mathrm{j}\omega})$$
$$(11)$$

where the infimum and supremum are over all values of ω in the interval $[-\pi, +\pi]$. In particular, if the power spectral density $\Phi_{xx}(e^{\mathrm{j}\omega})$ is bounded away from zero (persistence of excitation), then $\lambda_{\min}(\mathbf{R}_n) > 0$, i.e., $\mathbf{R}_n$ is non-singular. In that case the Wiener-Hopf equations have only one solution, irrespective of the value of n.

IV. THE LAGUERRE FILTER

Consider again the transversal filter of figure 3. Denote by $H(z)$ the transfer function of the (stable and causal) system that produces $y(k)$ when excited by $x(k)$. Applying the results of subsection II-D to the transversal filter it becomes clear that this filter is indirectly trying to approximate the impulse response of $H(z)$ by the first $n + 1$ sequences of the canonical basis of ℓ^2. Unfortunately, these sequences are extremely localized in time. Hence, the quality of the approximation will be very poor (for small n) when the impulse response of $H(z)$ is very long (e.g., when it decays slowly to zero). Note, however, that because the canonical basis is a complete set of ℓ^2 the approximation error can be made arbitrarily small by using a sufficiently large n (cf. section II-D).

It is possible to use other complete sets of ℓ^2 to build a "transversal-like" filter. In order for the filter to be practical to use each sequence of that set should be easy to generate digitally, i.e., it should have a rational z-transform. Probably the simplest set of such sequences (besides the canonical basis) is the set of the Laguerre sequences [19], which is a complete orthonormal set of ℓ^2 [20]. The z-transforms of these sequences are given by [21]

$$L_i(z, u) = \sqrt{1 - u^2} \, \frac{(z^{-1} - u)^i}{(1 - uz^{-1})^{i+1}}, \qquad i \geq 0 \quad (12)$$

where u is a free (real) parameter, the Laguerre pole position, with modulus smaller that one. Note that $L_i(z, 0) = z^{-i}$, i.e., the sequences of the canonical basis of ℓ^2 are a special case of the Laguerre sequences. Also interesting is the fact that for $i \geq 0$

$$L_{i+1}(z, u) = A(z, u) L_i(z, u) \qquad (13)$$

with

$$A(z, u) = \frac{z^{-1} - u}{1 - uz^{-1}},$$

i.e., these sequences can be generated in cascade, starting with a first order low-pass section ($L_0(z, u)$), followed by first order all-pass sections ($A(z, u)$).

Replacing the "backbone" of the transversal filter, that generates the first $n + 1$ canonical sequences of ℓ^2 when excited by $\delta(k)$, by the equivalent structure that generates the first $n + 1$ Laguerre sequences (when excited by $\delta(k)$) we obtain the Laguerre filter shown in figure 4. This filter structure was introduced in [22] and was studied with some detail in [24]. We will assume, unless stated otherwise, that
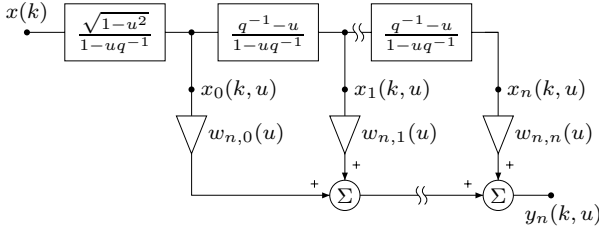
Fig. 4 - Laguerre filter of order n (with n sections). This filter is stable if and only if $|u| < 1$. For $u = 0$ the Laguerre filter degenerates into the familiar transversal filter.

for each value of u the weights of this filter are computed such that $y_n(k,u)$ is the best approximation, in the mean-square sense, to a given desired signal $y(k)$. This explains why these weights are functions of u in figure 4.

The output of an order n Laguerre filter excited by a real wide-sense stationary stochastic process $x(k)$ with zero-mean is given by

$$y_n(k,u) = \sum_{i=0}^{n} w_{n,i}(u)\, x_i(k,u)$$

where

$$x_i(k,u) = L_i(q,u)\, x(k).$$

For each value of u the optimal set of weights of this filter can be computed from the normal equations

$$\sum_{i=0}^{n} w_{n,i}(u)\, \langle x_i(k,u), x_j(k,u)\rangle = \langle y(k), x_j(k,u)\rangle \quad (14)$$

($0 \le j \le n$). These equations can be put in the form

$$\mathbf{R}_n(u)\, \mathbf{w}_n(u) = \mathbf{p}_n(u) \qquad (15)$$

where the elements of $\mathbf{R}_n(u)$ are given by

$$r_{ij}(u) = \int_{-\pi}^{+\pi} \left(\frac{e^{j\omega} - u}{1 - ue^{j\omega}} \right)^{j-i} \frac{\Phi_{xx}(e^{j\omega})(1-u^2)\, d\omega}{2\pi\, |1 - ue^{j\omega}|^2}$$

($0 \le i, j \le n$), and where those of $\mathbf{p}_n(u)$ are given by

$$p_i(u) = \int_{-\pi}^{+\pi} \left(\frac{e^{j\omega} - u}{1 - ue^{j\omega}} \right)^{i} \frac{\Phi_{yx}(e^{j\omega})\sqrt{1-u^2}\, d\omega}{2\pi(1 - ue^{j\omega})}$$

($0 \le i \le n$). These expressions can be obtained easily from (3). Similarly to the transversal filter case, $r_{ij}(u)$ depends only on $|i - j|$, i.e., $\mathbf{R}_n(u)$ is a Toeplitz matrix. We will explore this fact in the next section.

It is possible to simplify considerably the expression for $r_{ij}(u)$ using the change of frequency variable $\omega \mapsto \theta$ defined by the bilinear transformation [24]

$$e^{j\theta} = \frac{e^{j\omega} - u}{1 - ue^{j\omega}}. \qquad (16)$$

It is easy to verify that when ω goes from $-\pi$ to $+\pi$, θ also goes from $-\pi$ to $+\pi$ (see figure 5), that

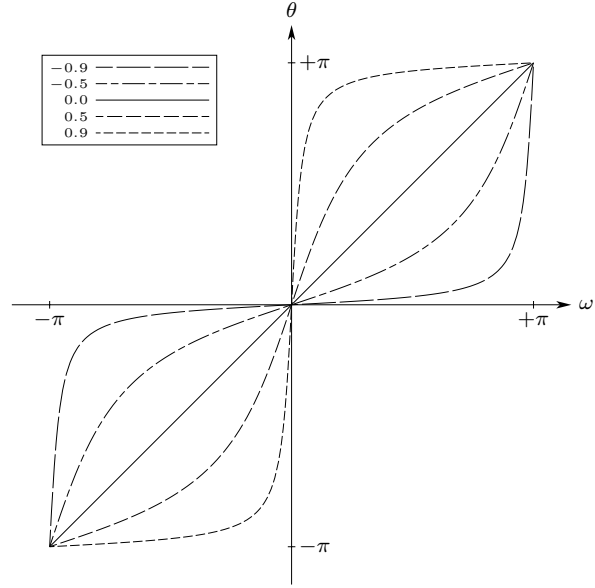$$e^{j\omega} = \frac{e^{j\theta} + u}{1 + ue^{j\theta}}, \qquad (17)$$

Fig. 5 - Graph of the frequency transformation $\omega \mapsto \theta$ for several values of u. Note that there exists a one to one correspondence between ω and θ. The inverse transformation $\theta \mapsto \omega$, which appears in (18), can be visualized easily by replacing u by $-u$. For $u > 0$ this inverse transformation compresses the low frequencies and for $u < 0$ it compresses the high frequencies.

and that

$$d\theta = \frac{1 - u^2}{|1 - ue^{j\omega}|^2}\, d\omega.$$

It is then trivial to verify that $r_{ij}(u)$ can also be given by [24]

$$r_{ij}(u) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} e^{j\theta(j-i)}\, \Phi_{xx}\!\left(\frac{e^{j\theta}+u}{1+ue^{j\theta}} \right) d\theta, \qquad (18)$$

which only differs from (10) in the argument of the power spectral density of $x(k)$. This very important result can be used to determine immediately bounds for the eigenvalues of $\mathbf{R}_n(u)$ based on those for the matrix $\mathbf{R}_n$ appearing in the transversal filter case. Because (17) represents only a distortion of the frequency scale (see figure 5) it turns out that (11) is also valid for this case, i.e., the lower and upper bounds for the eigenvalues of $\mathbf{R}_n(u)$ are exactly the same as those for $\mathbf{R}_n$ and do not depend on u. Also, some results concerning the asymptotic eigenvalue distribution of $\mathbf{R}_n$ when $n \to \infty$ (see [51] or [52] for details) can be adapted with very little effort to the matrices $\mathbf{R}_n(u)$.

A. Stationarity condition of the MSE of a Laguerre filter with respect to u [30]

The variance of the error signal of a Laguerre filter, i.e., its mean-square error (MSE), is a function of u. In order to minimize this function we need to deduce its stationarity condition and then to solve it. One of the solutions of this condition will be the optimal value of u, for which the MSE attains its global minimum.

The MSE of a Laguerre filter of order n is given by

$$\xi_n(u) = \langle e_n(k,u), e_n(k,u)\rangle.$$

Assume for the moment that the weights of the Laguerre filter are arbitrary, i.e., that they do not depend on u and that

they do not need to satisfy the normal equations. Then, the partial derivative of the MSE with respect to u is given by (a prime denotes differentiation with respect to this variable)

$$\xi_n'(u, \mathbf{w}_n) = -2 \sum_{i=0}^{n} w_{n,i} \langle e_n(k, u), x_i'(k, u) \rangle \qquad (19)$$

where $\mathbf{w}_n$ is a vector whose elements are the weights of the Laguerre filter. Notice that the stationarity conditions of $\xi_n(u, \mathbf{w}_n)$ require (19) to be zero, and they also force the normal equations to be satisfied. This latter condition would have been unnecessary if we had assumed that the weights where computed from the normal equations for each value of u. (This assumption would have made the analysis of the problem much more difficult.)

Because of the remarkable formula [30]

$$L_i'(z, u) = \frac{(i+1)L_{i+1}(z, u) - i\, L_{i-1}(z, u)}{1 - u^2}, \qquad (20)$$

the derivative of $x_i(k, u) = L_i(q, u)x(k)$ with respect to u is given by

$$x_i'(k, u) = \frac{(i+1)x_{i+1}(k, u) - i\, x_{i-1}(k, u)}{1 - u^2}. \qquad (21)$$

Applying this formula in (19) and simplifying the result with the normal equations (14) we obtain

$$\xi_n'(u) = -2(n+1)\frac{w_{n,n}(u)\,\langle e_n(k, u), x_{n+1}(k, u) \rangle}{1 - u^2}. \qquad (22)$$

We emphasize that this formula is only valid if the normal equations are satisfied. Hence, the weights are again functions of u, i.e., they are again computed from the normal equations. (In fact, this formula is the total derivative of $\xi_n(u)$.) Equating this formula to zero gives the stationarity condition of the MSE with respect to u.

The next step is to find the value of the only complicated term of (22): $\langle e_n(k, u), x_{n+1}(k, u) \rangle$. In order to do so we need to orthonormalize the signals $x_i(k, u)$.

Let $x_0^o(k, u), \ldots, x_i^o(k, u)$ be the orthonormalized signals obtained by applying the Gram-Schmidt orthogonalization algorithm (with normalization) to the signals $x_0(k, u), \ldots, x_i(k, u)$. The linear transformation performed by this algorithm can be expressed by

$$\mathbf{x}_i^o(k, u) = \mathbf{T}_i(u)\, \mathbf{x}_i(k, u) \qquad (23)$$

where

$$\mathbf{x}_i(k, u) = [\, x_0(k, u) \quad \cdots \quad x_i(k, u)\,]^{\mathrm{T}}$$

is a vector holding the original signals, and where

$$\mathbf{x}_i^o(k, u) = [\, x_0^o(k, u) \quad \cdots \quad x_i^o(k, u)\,]^{\mathrm{T}}$$

is the corresponding vector holding the orthonormalized signals. The matrix $\mathbf{T}_i(u)$ is a lower-triangular matrix. It will be nonsingular if and only if the signals $x_0(k, u), \ldots, x_i(k, u)$ are linearly independent (we will assume this condition to hold in the sequel). For example,

this will happen if $x(k)$ satisfy the persistence of excitation condition.

It is possible to prove that [53]

$$\mathbf{T}_i(u)\, \mathbf{R}_i(u)\, \mathbf{T}_i^{\mathrm{T}}(u) = \mathbf{I}_i,$$

i.e., $\mathbf{T}_i(u)$ is the inverse of the first Cholesky factor of $\mathbf{R}_i(u)$ [49]. For notational convenience we will denote the element of the last line and column of $\mathbf{T}_i(u)$ by $t_i(u)$. Note that $t_i(u) = 1$ for all i if and only if $\Phi_{xx}(e^{j\omega}) = 1$ for all ω, i.e., if and only if $x(k)$ is white noise with unitary variance. Note also that $t_i(u)$ is strictly positive, and that the element in the same position on the lower-triangular matrix $\mathbf{T}_i^{-1}(u)$ is $1/t_i(u)$.

Because the signals $x_i^o(k, u)$ are obtained by a linear combination of the signals $x_j(k, u)$, $j = 0, \ldots, i$, it is clear that the output of a Laguerre filter of order n can also be given by the orthonormal expansion

$$y_n(k, u) = \sum_{i=0}^{n} c_i(u)\, x_i^o(k, u) \qquad (24)$$

with $c_i(u) = \langle y(k), x_i^o(k, u) \rangle$. Note that $c_i(u)$ does not depend on n. It is also clear that the error signal of a Laguerre filter of order $n + 1$ is given by

$$e_{n+1}(k, u) = e_n(k, u) - c_{n+1}(u)\, x_{n+1}^o(k, u).$$

This formula implies that

$$\langle e_n(k, u), x_{n+1}(k, u) \rangle = \langle e_{n+1}(k, u), x_{n+1}(k, u) \rangle$$
$$+ \quad c_{n+1}(u)\langle x_{n+1}^o(k, u), x_{n+1}(k, u) \rangle.$$

The last normal equation for the Laguerre filter of order $n + 1$ gives $\langle e_{n+1}(k, u), x_{n+1}(k, u) \rangle = 0$. It is also clear that

$$\begin{aligned} y_{n+1}(k, u) &= \mathbf{w}_{n+1}^{\mathrm{T}}(u)\, \mathbf{x}_{n+1}(k, u) \\ &= \mathbf{c}_{n+1}^{\mathrm{T}}(u)\, \mathbf{x}_{n+1}^o(k, u) \end{aligned}$$

with obvious definitions for the vectors $\mathbf{w}_{n+1}(u)$ and $\mathbf{c}_{n+1}(u)$. Due to (23) and to the special form of $\mathbf{T}_{n+1}(u)$ it is then easy to verify that

$$c_{n+1}(u) = w_{n+1,n+1}(u)/t_{n+1}(u). \qquad (25)$$

Due to the special form of $\mathbf{T}_{n+1}^{-1}(u)$ and to the orthonormality of the signals $x_i^o(k, u)$ it is also easy to verify that

$$\langle x_{n+1}^o(k, u), x_{n+1}(k, u) \rangle = 1/t_{n+1}(u).$$

Putting all these facts together gives

$$\langle e_n(k, u), x_{n+1}(k, u) \rangle = \frac{w_{n+1,n+1}(u)}{t_{n+1}^2(u)}.$$

Applying this formula in (22) we obtain

$$\xi_n'(u) = -\frac{2(n+1)\, w_{n,n}(u)\, w_{n+1,n+1}(u)}{(1 - u^2)\, t_{n+1}^2(u)}. \qquad (26)$$

Therefore, the stationarity points of the MSE with respect to u satisfy the simple condition [30]

$$w_{n,n}(u)\, w_{n+1,n+1}(u) = 0. \qquad (27)$$

This condition is a generalization of the condition presented in [26] for the particular case where $x(k) = \delta(k)$ (which is the deterministic signal equivalent to white noise). It is interesting to verify that if $w_{n,n}(u) = 0$ then $\xi_n(u) = \xi_{n-1}(u)$, and that if $w_{n+1,n+1}(u) = 0$ then $\xi_n(u) = \xi_{n+1}(u)$. Hence, in each stationary point of $\xi_n(u)$ the graph of this function touches the graph of $\xi_{n-1}(u)$ and/or of $\xi_{n+1}(u)$. We will illustrate this phenomenon in section VI. It is important to stress that usually, *but not always*, the local minima of $\xi_n(u)$ satisfy the conditions $w_{n,n}(u) \neq 0$ and $w_{n+1,n+1}(u) = 0$, in which case $\xi_{n-1}(u) > \xi_n(u) = \xi_{n+1}(u)$.

For simple and efficient ways of solving approximately (27) we refer the reader to [54]. Basically, we approximate $w_{i,i}(u)$ by a truncated Taylor series or by a Padé approximant, and then find the zeros of that approximation. The derivatives of $w_{i,i}(u)$ required to form these approximations can be computed differentiating (15) and using (20) to simplify the result.

V. THE LAGUERRE LATTICE FILTER [55]

As promised earlier we are going to explore the Toeplitz structure of the matrix $\mathbf{R}_n(u)$. This will give rise to the lattice form of the Laguerre filter. The following line of reasoning is a simple generalization of the ideas that led to the standard lattice filter. These ideas can be found in any good book about adaptive filter theory, such as [1]-[3]. Another important work related to the material presented here is [8]. As before, we assume that the signals $x_i(k, u)$ of the Laguerre filter are linearly independent. To simplify the notation we will use the definition $r_{|i-j|}(u) \triangleq r_{ij}(u)$ when referring to the elements of the Toeplitz matrix $\mathbf{R}_n(u)$.

In order to orthogonalize the signals $x_i(k, u)$ it is useful to consider the problem of the minimization of the variance of the following signals, with the restrictions $a_{i0}(u) = b_{i0}(u) = 1$:

$$x_i^f(k, u) = \sum_{j=0}^{i} a_{ij}(u)\, x_j(k, u); \qquad (28)$$

$$x_i^b(k, u) = \sum_{j=0}^{i} b_{ij}(u)\, x_{i-j}(k, u). \qquad (29)$$

As explained in subsection II-C, $x_i^f(k, u)$ will be orthogonal to $x_j(k, u)$ for $0 < j \leq i$, and $x_i^b(k, u)$ will be orthogonal to $x_j(k, u)$ for $0 \leq j < i$. This implies that the signals $x_i^b(k, u)$ are the result of the Gram-Schmidt orthogonalization procedure applied to the signals $x_i(k, u)$. We will denote the standard deviation (the square root of the variance) of $x_i^f(k, u)$ by $\sigma_i^f(u)$, and that of $x_i^b(k, u)$ by $\sigma_i^b(u)$. Both of these standard deviations are strictly positive because we have assumed that the signals $x_i(k, u)$ are linearly independent.

It is simple to verify that the augmented normal equations[4] for these two problems are (in the following two equations

we have omitted, for aesthetical reasons, the dependence on u of all variables)

$$\begin{bmatrix} r_0 & \cdots & r_i \\ \vdots & \ddots & \vdots \\ r_i & \cdots & r_0 \end{bmatrix} \begin{bmatrix} a_{i0} & b_{ii} \\ \vdots & \vdots \\ a_{ii} & b_{i0} \end{bmatrix} = \begin{bmatrix} (\sigma_i^f)^2 & 0 \\ \vdots & \vdots \\ 0 & (\sigma_i^b)^2 \end{bmatrix}. \quad (30)$$

Due to the symmetry of these two problems it is clear that $a_{ij}(u) = b_{ij}(u)$ for all $i \geq 0$ and for $0 \leq j \leq i$, and that $\sigma_i^f(u) = \sigma_i^b(u) \triangleq \sigma_i(u)$ also for all $i \geq 0$.

The reason why the signals $x_i^f(k, u)$ are also useful is related to the special form of (30), that implies that

$$\begin{bmatrix} r_0 & \cdots & r_{i+1} \\ \vdots & \ddots & \vdots \\ r_{i+1} & \cdots & r_0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ a_{i1} & a_{ii} \\ \vdots & \vdots \\ a_{ii} & a_{i1} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \sigma_i^2 & \Delta_i \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ \Delta_i & \sigma_i^2 \end{bmatrix}. \quad (31)$$

Note that $\Delta_i(u)$ and $\sigma_i(u)$ can be computed as soon as the coefficients $a_{ij}(u)$ are know.

From (31) it is very easy to obtain the order update formulae for the weights $a_{ij}(u)$ (and also for the weights $b_{ij}(u)$), which are

$$a_{i+1,j}(u) = a_{i,j}(u) + k_{i+1}(u)\, a_{i,i+1-j}(u)$$

for $0 \leq j \leq i + 1$, with $k_{i+1}(u) = -\Delta_i(u)/\sigma_i^2(u)$, and with $a_{i,i+1}(u) \triangleq 0$. The application of these formulae in (28) and in (29), together with (13), gives

$$x_{i+1}^f(k, u) = x_i^f(k, u) + k_{i+1}(u)\, A(q, u)\, x_i^b(k, u)$$

and

$$x_{i+1}^b(k, u) = A(q, u)\, x_i^b(k, u) + k_{i+1}(u)\, x_i^f(k, u),$$

with $x_0^f(k, u) = x_0^b(k, u) = x_0(k, u)$. These recursion equations define part of the Laguerre lattice filter. It is also easy to show that

$$k_{i+1}(u) = -\frac{\langle A(q, u) x_i^b(k, u), x_i^f(k, u) \rangle}{\sigma_i^2(u)},$$

and that

$$\sigma_{i+1}^2(u) = [1 - k_{i+1}^2(u)]\, \sigma_i^2(u).$$

From this last formula we conclude, if $\sigma_{i+1}(u) > 0$, that $|k_{i+1}(u)| < 1$. This implies that the inverse Laguerre lattice filter is stable [56]. The coefficients $k_i(u)$ are sometimes called reflection coefficients.

It is clear that the output of the Laguerre filter is given by the orthogonal expansion (compare with (24))
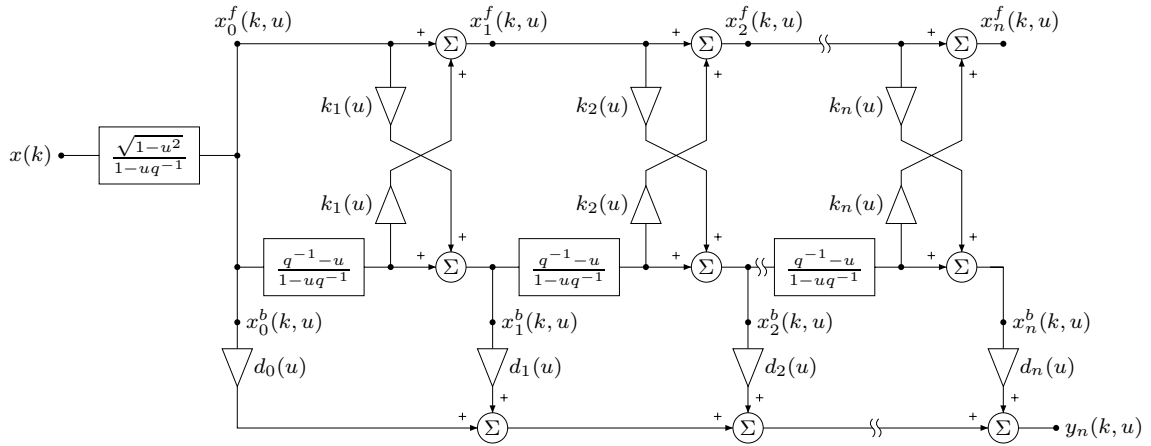
$$y_n(k, u) = \sum_{i=0}^{n} d_i(u)\, x_i^b(k, u)$$

where

$$d_i(u) = \frac{\langle y(k), x_i^b(k, u) \rangle}{\sigma_i^2(u)}$$

does not depend on n. This formula defines the joint-process part of the Laguerre lattice filter, shown in figure 6.

---

[4]The augmented normal equations are the equations (4) and (5) put together in only one equation.

Fig. 6 - Laguerre lattice filter of order n. For $u = 0$ this is the familiar lattice filter.

A. Stationarity condition of the MSE of a Laguerre lattice filter with respect to u [55]

We start by normalizing the signals $x_i^b(k, u)$, obtaining the signals

$$x_i^o(k, u) = \frac{x_i^b(k, u)}{\sigma_i(u)}.$$

Using these signals the output of the Laguerre filter can also be given by (this is a repetition of (24))

$$y_n(k, u) = \sum_{i=0}^{n} c_i(u) \, x_i^o(k, u)$$

with $c_i(u) = \sigma_i(u) \, d_i(u)$, and its MSE is given by

$$\xi_n(u) = \langle y(k), y(k) \rangle - \sum_{i=0}^{n} c_i^2(u).$$

Because

$$x_i^o(k, u) = \sum_{j=0}^{i} \frac{b_{i,i-j}(u)}{\sigma_i(u)} \, x_j(k, u)$$

it is not very difficult to show, using (21), that

$$\frac{dx_i^o(k, u)}{du} = \sum_{j=0}^{i+1} \alpha_{ij}(u) \, x_j^o(k, u) \tag{32}$$

with

$$\alpha_{i,i+1}(u) = \frac{(i+1) \, \sigma_{i+1}(u)}{(1-u^2) \, \sigma_i(u)}.$$

The exact value of the other α_{ij}'s will not be needed. It will prove useful to change the upper limit of the summation in (32) from $i+1$ to ∞. This is accomplished with the definition $\alpha_{ij}(u) \overset{\Delta}{=} 0$ for $j > i+1$.

Differentiating the orthonormality condition

$$\langle x_i^o(k, u), x_j^o(k, u) \rangle = \delta_{ij}$$

with respect to u and using (32) it is easy to verify that for all $i, j \geq 0$

$$\alpha_{ij}(u) + \alpha_{ji}(u) = 0.$$

This formula implies that $\alpha_{ij}(u) = 0$ for $j = i$ and also for $j < i - 1$. Because

$$c_i(u) = \langle y(k), x_i^o(k, u) \rangle$$

it is then clear that

$$c_i'(u) = \alpha_{i,i-1}(u) \, c_{i-1}(u) + \alpha_{i,i+1}(u) \, c_{i+1}(u). \tag{33}$$

The derivative of the MSE with respect to u is given by

$$\xi_n'(u) = -2 \sum_{i=0}^{n} c_i(u) \, c_i'(u).$$

Using (33) and $\alpha_{i,i-1}(u) = -\alpha_{i-1,i}(u)$ this summation becomes a telescopic series (!) whose sum is

$$\xi_n'(u) = -2\alpha_{n,n+1}(u) \, c_n(u) \, c_{n+1}(u).$$

Note that this formula is in accord with (26) because $t_i(u) = 1/\sigma_i(u)$. It is then very easy to verify that

$$\xi_n'(u) = -\frac{2(n+1) \, \sigma_{n+1}^2(u) \, d_n(u) \, d_{n+1}(u)}{(1-u^2)}. \tag{34}$$

Hence, the stationarity condition is

$$d_n(u) \, d_{n+1}(u) = 0. \tag{35}$$

Note how easily this condition can be interpreted: the MSE of a Laguerre lattice filter has a stationary point with respect to u if and only if the last weight used to compute the output signal vanishes and/or the first unused weight vanishes.

Although (35) could also have been obtained much more easily from (27) and (25), its deduction given above is entirely based on the Laguerre lattice filter, and is interesting in its own right.

VI. AN EXAMPLE

To illustrate the approximation capabilities of Laguerre filters we used a Laguerre lattice filter with 10 sections to approximate the output of a third order elliptic low pass filter with the following transfer function

$$H(z) = \frac{0.01624(1 + z^{-1})(1 - 1.7313z^{-1} + z^{-2})}{(1 - 0.8957z^{-1})(1 - 1.8445z^{-1} + 0.9282z^{-2})},$$
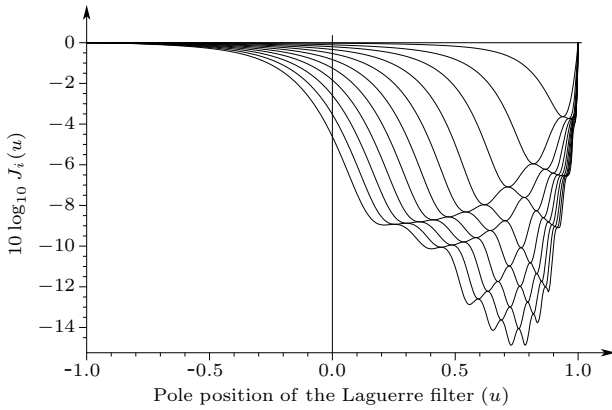
Fig. 7 - Normalized MSE (in dB) of the Laguerre lattice filters of orders from 0 (top) to 10 (bottom) as functions of u. The Laguerre filters were used to approximate a low pass system excited by colored Gaussian noise. Note that these curves touch only at their local extrema. Note also the bad performance of the transversal filter ($u = 0$) when compared to the best Laguerre filter, i.e., with optimal u, of the same order.

excited by colored Gaussian noise generated by feeding (pseudo) white Gaussian noise of unitary variance to a filter with transfer function

$$N(z) = \frac{0.5 + 1.5z^{-1}}{1 + 0.4z^{-1}}.$$

The same signal was used as input of the Laguerre lattice filter. The approximation was performed off-line using 2500 samples of the input signal, previously recorded from one realization of the (pseudo) white Gaussian process. To reduce the effects of the null initial conditions, the first 500 samples were used only to initialize the Laguerre lattice filter. The other 2000 samples were used to compute the reflection coefficients ($k_i(u)$), the joint-process weights ($d_i(u)$), and the MSE ($\xi_i(u)$) of the Laguerre lattice filters of orders from 0 to 10. (Remember that a Laguerre lattice filter of order n effectively contains all Laguerre lattice filters of lower orders.) The algorithm used to compute these coefficients is presented in the appendix.

The normalized MSE error[5] curves for the eleven Laguerre lattice filters are presented in figure 7. Note that consecutive curves touch only where they have local extrema, which is in accord with (35). Although in this example all local minima (maxima) of $J_i(u)$ are associated with the condition $d_{i+1}(u) = 0$ ($d_i(u) = 0$) this is not always the case. Figure 7 also shows that the MSE curve of a Laguerre filter usually has local minima. This is an usual characteristic of insufficient order IIR filters used in an output error configuration.

VII. CONCLUSIONS

We have seen that the transversal filter can be generalized to a filter structure, the Laguerre filter, which has one additional free parameter that controls the filter's (multiple) pole position. Setting this parameter to 0 puts the (multiple) pole of the filter at the origin, turning the Laguerre filter into a transversal filter. By adjusting properly this

---

[5]That is, the MSE error divided by the variance of the signal being approximated: $J_i(u) = \xi_i(u)/\langle y(k), y(k)\rangle$.

parameter, which controls the rate of decay to zero of the filter's impulse response, allows this filter to provide good approximations to systems with slowly decaying impulse responses.

It is easy to devise algorithms to adapt the weights of Laguerre filters similar to the LMS or the RLS developed for the transversal filter [1]-[3]. It is also easy to adapt the reflection coefficients and joint-process weights of the Laguerre lattice filters using a stochastic gradient approach similar to the one used for transversal filters [1]-[3]. Unfortunately, it appears that it is not easy to generalize the FTF and LSL fast adaptation algorithms [1]-[4] to the Laguerre filter. Finally, it is possible to extend the adaptation (using a LMS scheme) to the Laguerre pole position. In this respect, the equations (26) and (34) are useful (specially the latter).

Besides the Laguerre functions there are other complete orthonormal sets of ℓ^2 whose sequences have rational z-transforms. In this respect the Kautz functions [57] and sequences [21] are particularly useful. Replacing the Laguerre sequences by the Kautz sequences we obtain a Kautz filter, which appears to be very promising in the approximation of systems with a dominant complex pole pair [58]-[62].

APPENDIX

The following C code implements an algorithm to compute the coefficients of the orthonormal expansion (24) given the symmetric Toeplitz matrix $\mathbf{R}_n(u)$ (more properly, its first line) and the vector $\mathbf{p}_n(u)$. From these coefficients it is very easy to compute the MSE of the Laguerre lattice filters with up to n sections. Note that this algorithm is slightly different than the usual Levinson algorithm [49] that solves the system $\mathbf{R}_n \mathbf{w}_n = \mathbf{p}_n$. Here we are not interested in $\mathbf{w}_n$ but in the coefficients of the orthonormal expansion (24).

```c
/*
** Modified Levinson algorithm
**
** Inputs:
**   n        --- Number of sections
**   r[0..n] --- Elements of the first line
**                 of the R Toeplitz matrix
**   p[0..n] --- Elements of the p vector
** Outputs:
**   c[0..n] --- Weights of the orthonormal
**                 expansion
** Internal variables:
**   k   --- Reflection coefficient of each
**             section
**   s2 --- Variance of orthogonal output
**             signal of each section
*/

typedef double real;
#define nMax 10

void modLevinson(int n,real *r,real *p,real *c)
{
  real k,s2,a[1 + nMax],b[1 + nMax];
  int i,j;

  a[0] = 1.0;
  s2 = r[0];
  c[0] = p[0] / sqrt(s2);
  for(i = 1;i <= n;i++)
```

```
  {
    a[i] = 0.0;
    k = 0;
    for(j = 0;j < i;j++)
      k += r[i - j] * a[j];
    k /= s2;
    s2 *= 1.0 - k * k;
    for(j = 0;j <= i;j++)
      b[j] = a[j];
    for(j = 0;j <= i;j++)
      a[j] -= k * b[i - j];
    c[i] = 0.0;
    for(j = 0;j <= i;j++)
      c[i] += a[i - j] * p[j];
    c[i] /= sqrt(s2);
  }
}
```

This algorithm can be easily modified to generate the $d_i(u)$ coefficients of the Laguerre lattice filter. It is only necessary to replace the divisions by `sqrt(s2)` with divisions by `s2`.

REFERENCES

[1]   Michael L. Honig and David G. Messerschmitt, *Adaptive Filters: Structures, Algorithms, and Applications*, Kluwer Academic Publishers, Boston, 1984.

[2]   Bernard Widrow and Samuel D. Stearns, *Adaptive Signal Processing*, Prentice–Hall, Englewood Cliffs, 1985.

[3]   Simon Haykin, *Adaptive Filter Theory*, Prentice–Hall, Englewood Cliffs, NJ, second edition, 1991.

[4]   Ali H. Sayed and Thomas Kailath, "A state-space approach to adaptive RLS filtering", *IEEE Signal Processing Magazine*, vol. 11, no. 3, pp. 18–60, July 1994.

[5]   C. Richard Johnson, Jr., "Adaptive IIR filtering: Current results and open issues", *IEEE Transactions on Information Theory*, vol. 30, no. 2, pp. 237–250, Mar. 1984.

[6]   John J. Shynk, "Adaptive IIR filtering", *IEEE ASSP Magazine*, vol. 6, no. 2, pp. 4–21, Apr. 1989.

[7]   T. Söderström and P. Stoica, "Some properties of the output error method", *Automatica*, vol. 18, no. 1, pp. 93–99, Jan. 1982.

[8]   David G. Messerschmitt, "A class of generalized lattice filters", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 2, pp. 198–204, Apr. 1980.

[9]   Y. W. Lee, "Synthesis of electric networks by means of the Fourier transforms of Laguerre's functions", *Journal of Mathematics and Physics*, vol. XI, pp. 83–113, 1933.

[10] Y. W. Lee, *Statistical Theory of Communication*, John Wiley and Sons, New York, 1960.

[11] D. G. Lampard, "A new method of determining correlation functions of stationary time series", *The Proceedings of the Institution of Electrical Engineers*, vol. 101, Part III, pp. 343–346, 1954, Monograph No. 104.

[12] J. W. Head, "Approximation to transients by means of Laguerre series", *Proceedings of the Cambridge Philosophical Society*, vol. 52, pp. 640–651, 1956.

[13] G. J. Clowes, "Choice of the time-scaling factor for linear system approximations using orthonormal Laguerre functions", *IEEE Transactions on Automatic Control*, vol. 10, no. 4, pp. 487–489, Oct. 1965.

[14] Kenneth Steiglitz, "Rational transform approximation via the Laguerre spectrum", *Journal of the Franklin Institute*, vol. 280, no. 5, pp. 387–394, Nov. 1965.

[15] J. J. King and T. O'Canainn, "Optimum pole positions for Laguerre-function models", *Electronics Letters*, vol. 5, no. 23, pp. 601–602, Nov. 1969.

[16] Martin Schetzen, "Power-series equivalence of some functional series with applications", *IEEE Transactions on Circuit Theory*, vol. 17, no. 3, pp. 305–313, Aug. 1970.

[17] Martin Schetzen, "Asymptotic optimum Laguerre series", *IEEE Transactions on Circuit Theory*, vol. 18, no. 5, pp. 493–500, Sept. 1971.

[18] T. W. Parks, "Choice of time scale in Laguerre approximations using signal measurements", *IEEE Transactions on Automatic Control*, vol. 16, no. 5, pp. 511–513, Oct. 1971.

[19] Morris J. Gottlieb, "Concerning some polynomials orthogonal on a finite or enumerable set of points", *American Journal of Mathematics*, vol. 60, pp. 453–458, 1938.

[20] Gabor Szegö, *Orthogonal Polynomials*, vol. XXIII of *American Mathematical Society Colloquium Publications*, American Mathematical Society, Providence, Rhode Island, fourth edition, 1975.

[21] Paul W. Broome, "Discrete orthonormal sequences", *Journal of the Association for Computing Machinery*, vol. 12, no. 2, pp. 151–168, Apr. 1965.

[22] R. E. King and P. N. Paraskevopoulos, "Digital Laguerre filters", *International Journal of Circuit Theory and Applications*, vol. 5, pp. 81–91, 1977.

[23] Bruno Maione and Biagio Turchiano, "Laguerre z-transfer function representation of linear discrete-time systems", *International Journal of Control*, vol. 41, no. 1, pp. 245–257, 1985.

[24] Bo Wahlberg, "System identification using Laguerre models", *IEEE Transactions on Automatic Control*, vol. 36, no. 5, pp. 551–562, May 1991.

[25] Svante Gunnarsson and Bo Wahlberg, "Some asymptotic results in recursive identification using Laguerre models", *International Journal of Adaptive Control and Signal Processing*, vol. 5, pp. 313–333, 1991.

[26] Mohammad A. Masnadi-Shirazi and N. Ahmed, "Optimal Laguerre networks for a class of discrete-time systems", *IEEE Transactions on Signal Processing*, vol. 39, no. 9, pp. 2104–2108, Sept. 1991.

[27] Y. Fu and G. Dumont, "An optimum time scale for discrete Laguerre network", *IEEE Transactions on Automatic Control*, vol. 38, no. 6, pp. 934–938, June 1993.

[28] Y. Fu and G. A. Dumont, "On determination of Laguerre filter pole through step or impulse response data", in *Preprints of the 12-th IFAC World Congress*, July 1993, pp. 5:303–307.

[29] Tomás Oliveira e Silva, "On the equivalence between Gamma and Laguerre filters", in *1994 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Apr. 1994, pp. IV:385–388.

[30] Tomás Oliveira e Silva, "Optimality conditions for truncated Laguerre networks", *IEEE Transactions on Signal Processing*, vol. 42, no. 9, pp. 2528–2530, Sept. 1994.

[31] Ü. Nurges, "Laguerre models in problems of approximation and identification of discrete systems", *Automation and Remote Control*, vol. 48, pp. 346–352, 1987.

[32] C. Zervos, P. R. Bélanger, and G. A. Dumont, "On PID controller tuning using orthonormal series identification", *Automatica*, vol. 24, no. 2, pp. 165–175, 1988.

[33] Guoxiang Gu, Pramod P. Khargonekar, and E. Bruce Lee, "Approximation of infinite-dimensional systems", *IEEE Transactions on Automatic Control*, vol. 34, no. 6, pp. 610–618, June 1989.

[34] Guy A. Dumont, Christos C. Zervos, and Gerry L. Pageau, "Laguerre-based adaptive control of pH in an industrial bleach plant extraction stage", *Automatica*, vol. 26, no. 4, pp. 781–787, 1990.

[35] Pertti M. Mäkilä, "Approximation of stable systems by Laguerre filters", *Automatica*, vol. 26, no. 2, pp. 333–345, 1990.

[36] P. M. Mäkilä, "Laguerre series approximation of infinite dimensional systems", *Automatica*, vol. 26, no. 6, pp. 985–995, 1990.

[37] Jonathan R. Partington, "Approximation of delay systems by Fourier-Laguerre series", *Automatica*, vol. 27, no. 3, pp. 569–572, 1991.

[38] Osvaldo Agamennoni, Eduardo Paolini, and Alfredo Dasages, "On robust stability analysis of a control system using Laguerre series", *Automatica*, vol. 28, no. 4, pp. 815–818, 1992.

[39] P. R. Bélanger, O. Arafat, M. Gaber, S. Gendron, and D. Vurgait-Cherson, "Direct performance optimization using Laguerre models", *Automatica*, vol. 30, no. 5, pp. 883–886, May 1994.

[40] L. Wang and W. R. Cluett, "Optimal choice of time-scaling factor for linear system approximations using Laguerre models", *IEEE Transactions on Automatic Control*, vol. 39, no. 7, pp. 1463–1467, July 1994.

[41] A. C. den Brinker, "Adaptive modified Laguerre filters", *Signal Processing*, vol. 31, pp. 69–79, 1993.

[42] A. C. den Brinker, "Calculation of the local cross-correlation function on the basis of the Laguerre transform", *IEEE Transactions on Signal Processing*, vol. 41, no. 5, pp. 1980–1982, May 1993.

[43] Albertus C. den Brinker, "Laguerre-domain adaptive filters", *IEEE Transactions on Signal Processing*, vol. 42, no. 4, pp. 953–956, Apr. 1994.

[44] K. Steiglitz, "The equivalence of digital and analog signal processing", *Information and Control*, vol. 8, pp. 455–467, 1965.

[45] Alan V. Oppenheim and Ronald W. Schafer, *Digital Signal Processing*, Prentice–Hall, Englewood Cliffs, NJ, 1975.

[46] Nicholas Young, *An Introduction to Hilbert Space*, Cambridge University Press, Cambridge, UK, 1988.

[47] Boaz Porat, *Digital Processing of Random Signals : Theory and Methods*, Prentice–Hall, 1994.

[48] Walter Rudin, *Real and Complex Analysis*, McGraw–Hill Book Company, New York, third edition, 1987.

[49] Gene H. Golub and Charles F. van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, second edition, 1989.

[50] Lennart Ljung, *System Identification. Theory for the user*, Prentice–Hall, Englewood Cliffs, NJ 07632, 1987.

[51] Ulf Grenander and Gabor Szegö, *Toeplitz forms and their applications*, Chelsea Publishing Company, New York, second edition, 1984.

[52] Robert Molten Gray, "On the asymptotic eigenvalue distribution of Toeplitz matrices", *IEEE Transactions on Information Theory*, vol. 18, no. 6, pp. 725–730, Nov. 1972.

[53] Philip J. Davis, *Interpolation and Approximation*, Dover Publications, Inc., New York, 1975, Republication.

[54] Tomás Oliveira e Silva, "On the determination of the optimal pole position of Laguerre filters", *IEEE Transactions on Signal Processing*, vol. 43, no. 9, pp. 2079–2087, Sept. 1995.

[55] Tomás Oliveira e Silva, "Optimality conditions for Laguerre lattice filters", *IEEE Signal Processing Letters*, vol. 2, no. 5, pp. 97–98, May 1995.

[56] P. P. Vaidyanathan and Sanjit K. Mitra, "A unified structural interpretation of some well-known stability-test procedures for linear systems", *Proceedings of the IEEE*, vol. 75, no. 4, pp. 478–497, Apr. 1987.

[57] William H. Kautz, "Transient synthesis in the time domain", *IRE Transactions on Circuit Theory*, vol. 1, pp. 29–39, 1954.

[58] P. Lindskog and B. Wahlberg, "Applications of Kautz models in system identification", in *Preprints of the 12-th IFAC World Congress*, July 1993, pp. 5:309–312.

[59] Bo Wahlberg, "System identification using Kautz models", *IEEE Transactions on Automatic Control*, vol. 39, no. 6, pp. 1276–1282, June 1994.

[60] Tomás Oliveira e Silva, "Optimality conditions for truncated second order Kautz networks with two complex conjugate poles", *IEEE Transactions on Automatic Control*, vol. 40, no. 2, pp. 342–346, Feb. 1995.

[61] Tomás Oliveira e Silva, *Sobre os filtros de Kautz e sua utilização na aproximação de sistemas lineares invariantes no tempo*, PhD thesis, Universidade de Aveiro, July 1994.

[62] Tomás Oliveira e Silva, "Kautz filters", English translation of a work written in Portuguese for the "Prémio Científico IBM 94", Sept. 1994.