On the Adaptation of the Pole of Laguerre-Lattice Filters

T. Oliveira e Silva*
Universidade de Aveiro / INESC Aveiro
Portugal
tos@inesca.pt

ABSTRACT

The main purpose of this paper is to present some experimental results concerning the adaptation of the pole position of the lattice version of the Laguerre filter. Basically, we propose the adaptation of the Laguerre-lattice parameters with the GAL-L algorithm, and the adaptation of the pole with a suitable sign algorithm. An example and suggestions about how to attempt to avoid local minima (with respect to the pole position) are also given.

1 Introduction

The main problem of (feedforward) lattice filters, or, in general, FIR filters, is that they are not well suited to approximate systems with very long impulse responses [1]. One possible way to attenuate this problem is to replace the FIR filter by the so-called Laguerre filter [2], [3], which, instead of having a multiple pole at the origin, has a multiple pole somewhere in the interval $(-1, 1)$. These Laguerre filters also have a corresponding lattice version, as shown recently [4], [5]. Unfortunately, this lattice structure is only valid for the stationary case, with the result that its parameters have to be adapted using, e.g., the gradient adaptive lattice (GAL) algorithm [1]. In this paper we are interested in the much more difficult problem of adapting the (multiple) pole of the filter, since this parameter is crucial for its optimal performance. Unfortunately, this minimization problem in non-convex and so our algorithm, which uses only local information, may be trapped in local minima.

The structure of this paper is the following. In section 2 we describe succinctly the Laguerre-lattice filter. Section 3 presents a trivial generalization of the GAL formulas to our more general lattice filter. In section 4 we describe our pole adaptation algorithm. In section 5 we present the results of a simple system identification simulation using the proposed algorithm. Lastly, in section 6 we describe some possible ways to attempt to escape from local minima of the performance surface.

* The author gratefully acknowledges the financial support of JNICT, Junta Nacional de Investigação Científica e Tecnológica, Portugal.

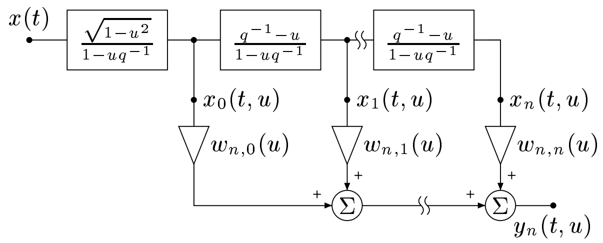

Figure 1: The transversal form of the Laguerre filter of order n. Note that for $u = 0$ this filter degenerates into the familiar transversal filter.

In this paper q will denote the advance operator, q^{-1} its inverse (the delay operator), and t the (discrete) time variable. $E[\cdot]$ will denote the expectation operator.

2 The Laguerre-lattice filter

The transfer function of the transversal form of the Laguerre filter of order n, which is depicted in Fig. 1, is given by [2]

$$H_n(z,u) = L(z,u) \sum_{i=0}^{n} w_{n,i} A^i(z,u) \qquad (1)$$

where $L(z,u) = \sqrt{1-u^2}/(1-uz^{-1})$ is a first order low-pass filter and $A(z,u) = (z^{-1}-u)/(1-uz^{-1})$ is a first order all-pass filter (both with the same pole). Note that for $u = 0$ (1) becomes exactly the transfer function of a FIR filter. It is very interesting to verify that in the stationary case the correlation matrix of the internal signals of the Laguerre filter, which are given by

$$x_i(t,u) = L(q,u)A^i(q,u)x(t),$$

is a Toeplitz matrix [2]. This observation suggests immediately that these filters have also a lattice version, which is indeed the case [3], [5]. It is interesting to note that this so-called Laguerre-lattice filter is very similar to the generalized lattice filter proposed by Messerschmitt in 1980 [6], in which the delays of the standard lattice filter are replaced by copies of an arbitrary all-pass filter. (In the Laguerre-lattice filter this all-pass

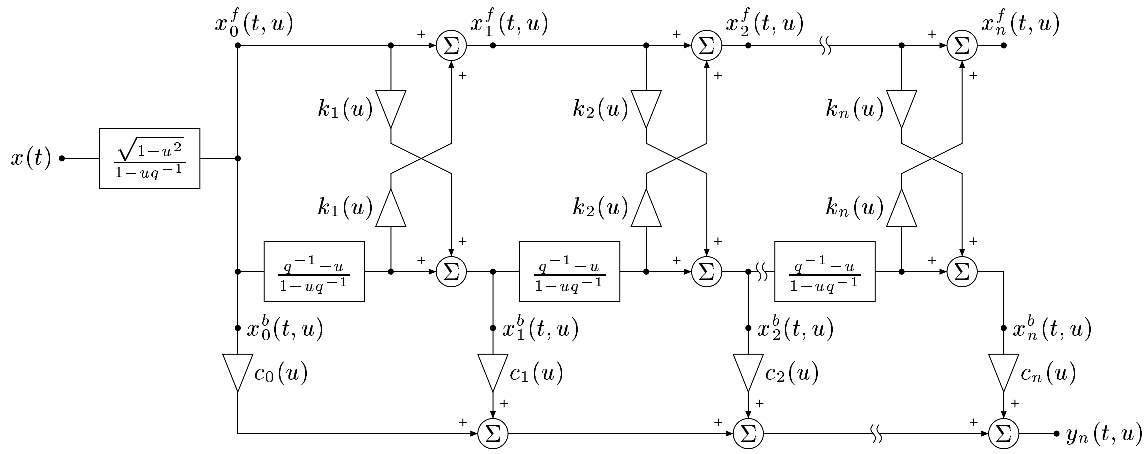

Figure 2: Laguerre-lattice filter of order n. Note that for $u = 0$ this filter degenerates into the familiar lattice filter.

filter is of first order, and there is a first-order low-pass preprocessing stage. These are also the differences between the transversal filters and Laguerre filters.)

For the lattice version of the Laguerre filter the following formulas hold [3], [5] (see also Fig. 2):

$$x^f_{i+1}(t,u) = x^f_i(t,u) + k_{i+1}(u)\,A(q,u)\,x^b_i(t,u)$$

$$x^b_{i+1}(t,u) = A(q,u)\,x^b_i(t,u) + k_{i+1}(u)\,x^f_i(t,u)$$

with $x^f_0(t,u) = x^b_0(t,u) = L(q,u)x(t)$;

$$k_{i+1}(u) = -\frac{E[A(q,u)x^b_i(t,u)\,x^f_i(t,u)]}{\sigma^2_i(u)}$$

with

$$\sigma^2_i(u) = E[x^f_i(t,u)\,x^f_i(t,u)] = E[x^b_i(t,u)\,x^b_i(t,u)];$$

$$\sigma^2_{i+1}(u) = [1 - k^2_{i+1}(u)]\sigma^2_i(u);$$

and

$$c_i(u) = \frac{E[y(t)\,x^b_i(t,u)]}{\sigma^2_i(u)}$$

where $y(t)$ is the desired output signal of the filter. The mean squared error (MSE) of the filter is given by

$$\xi_n(u) = E[y^2(t)] - \sum_{i=0}^{n} \sigma^2_i(u)c^2_i(u),$$

and the normalized MSE is given by

$$J_n(u) = \frac{\xi_n(u)}{E[y^2(t)]}.$$

Lastly, the derivative of $\xi_n(u)$ with respect to u is given by [3], [5]

$$\xi'_n(u) = -2(n+1)\frac{\sigma^2_{i+1}(u)}{1-u^2}\,c_n(u)c_{n+1}(u) \qquad (2)$$

($c_{n+1}(u)$ can be computed easily adding one extra section to the filter.)

3 The GAL-L adaptation algorithm

The expectation operator that appears in some formulas of the previous section is not suited for use in an adaptive filter. In the GAL algorithm it is replaced by the operator

$$E_\lambda[x(t)] = \begin{cases} 0, & t < 0 \\ \lambda\,E_\lambda[x(t-1)] + x(t), & t \ge 0 \end{cases} \qquad (3)$$

which, in essence, is an expectation-like operator (for ergodic signals) with an exponential forgetting window. Of course $0 < \lambda \le 1$. Actually, $E_\lambda[x(t)]$, as defined by (3), should be divided by $E_\lambda[1]$ in order to get a running estimate of $E[x(t)]$. [A popular alternative, for $\lambda < 1$, is to multiply $x(t)$ in the right hand side of (3) by $(1-\lambda)$.] However, it is not necessary to do so (in a floating point implementation of these equations) because in the Laguerre-lattice equations, with the exception of the formulas involving $\xi_n(u)$ and its first derivative, these expectations appear only as ratios. The same will be true for these exceptions if instead of working with $\xi_n(u)$ we work with its normalized version, $J_n(u)$.

It must be noted that since the recursive operator $E_\lambda[\cdot]$ depends on t (time) the variables $k_i(u)$, $\sigma_i(u)$, and $c_i(u)$ of the Laguerre-lattice filter become time variant (as they should in an adaptive algorithm). Furthermore, for safety reasons, it is convenient to add a small positive number to the first estimate of $\sigma^2_i(u)$. This can be easily accomplished with a suitable initial condition in (3).

4 The adaptation of the pole

The adaptation of u is a difficult problem. Our approach is to use the time variant version of (2), in which $c_n(u)$, $c_{n+1}(u)$, and $\sigma^2_{n+1}(u)$ are estimated on a sample by sample basis, to estimate the "correct" direction of change of u, also on a sample by sample basis. In practice, this amounts to verify if $c_n(t,u)$ and $c_{n+1}(t,u)$ have the same sign or not. This leads to an adaptation rule of the form

$$u(t+1) = u(t) + \mu(t)\,\mathrm{sign}[c_n(t,u)c_{n+1}(t,u)] \qquad (4)$$

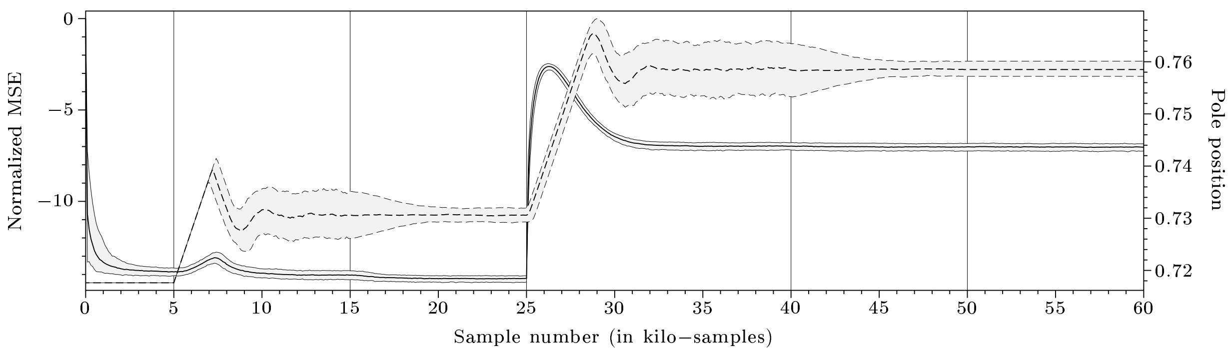

Figure 3: Normalized MSE (solid line) and pole position (broken line) as a function of the sample number. These curves represent the 10%, median (50%) and 90% percentiles of 501 different runs of the simulation.

where $\mu(t)$ should be *very small* (say 10^{-6}). Note that any change in u will induce transients in the GAL-L algorithm, and this will increase the excess MSE of the filter. Also, if these transients are too large, i.e., if u is adapted too quickly, (2) will be meaningless, since it is only valid for the stationary case. If that happens, then there is no guarantee that the adaptation process remains stable. Finally, note that $\mu(t)$ should decrease when $|u|$ approaches one, since in that case the transients take more time to disappear. This suggests the introduction of a factor of the form $(1-u^2)^a$, for some $a>0$, in the last term of (4). [We have not done so in the simulation experiment.] In [7] it is suggested that a should be $3/2$ for a transversal Laguerre filter with weights adapted by the RLS algorithm.

During the preparation of this paper we have also considered to use an adaptation rule for u that was of the form

$$u(t+1) = u(t) - \nu(t)\xi_n'(u),$$

i.e., an LMS-like adaptation rule. There are two problems with this approach. The main one is the selection of an appropriate step size, $\nu(t)$, which ensures that the adaptation remains stable. This is not an easy problem to solve. The other is its very slow speed of adaptation when u happens to be near a saddle point. The sign-like adaptation rule is easier to implement, is safer, and works well near saddle points, at the expense of a worse behaviour near local minima.

5 Simulation experiment

The following experimental setup was used. An input signal was generated by passing pseudo-random white Gaussian noise through a filter with transfer function $N(z) = 1 - z^{-1} + z^{-2}$. This colored signal was fed to a Laguerre-lattice filter of order $n = 8$ and to a system with the time-variant "transfer function"

$$H(z) = \frac{1}{1-\alpha z^{-1}+0.9z^{-2}}, \qquad \alpha = \begin{cases} 1.8, & t < 25000 \\ 1.6, & t \geq 25000. \end{cases}$$

Our desired signal was the output of this system plus pseudo-random Gaussian noise with variance $(0.01)^2$. The forgetting factor of the GAL-L algorithm was set to $\lambda = 0.999$ during the entire simulation, and the Laguerre pole was set initially to $u \approx 0.7176$ (this value is the theoretical asymptotically best value for u, which in this case is given by the root of $0.9x^2 - (1+0.9^2+0.3^2)x + 0.9 = 0$ with modulus smaller than 1). To illustrate the various regimes of adaptation $\mu(t)$ was varied in the following way:

$$\mu(t) = \begin{cases} 0, & 0 \leq t \leq 5000; \\ 10^{-5}, & 5000 < t \leq 15000; \\ 10^{-6}, & 15000 < t \leq 25000; \\ 10^{-5}, & 25000 < t \leq 40000; \\ 10^{-6}, & 40000 < t \leq 50000; \\ 0, & 50000 < t. \end{cases}$$

The results of 501 different runs of the simulation are presented in Fig. 3. Note the triangular-wave behaviour of the pole movements during the initial stages of the adaptation [in our case, those with a "large" $\mu(t)$]. Note also the just noticeable decrease of the MSE after $t = 15000$, i.e., when $\mu(t)$ was decreased tenfold. A similar phenomenon occurs after $t = 40000$ but is not clearly detectable in the figure. Finally, note the decrease of the pole wanderings after $t = 15000$ and $t = 40000$, which is due to the decrease of the value of $\mu(t)$. This decrease was not, however, by a factor of ten.

It is instructive to analyze the MSE curves of Laguerre filters of different orders in steady state (stationary) conditions. We need two sets of curves, one for $t < 25000$ ($\alpha = 1.8$), and the other for $t \geq 25000$ ($\alpha = 1.6$). These are shown, respectively, in Figs. 4 and 5. From these figures it is clear that the value to which the pole converges after $t > 25000$ corresponds to a local minimum of the MSE, which happens to be the one closer to the global minimum for $t < 25000$.

With a λ not so close to 1 (we experimented with $\lambda = 0.99$ and $\mu(t)$ as in the first experiment), in some runs of the simulation u converged to a value corresponding to a local minimum closer to the global minimum. The conclusion to be extracted from this experiment is that smaller values of λ make the estimated sign of $\xi_n'(u)$ noisier (as well as speeding up the adaptation

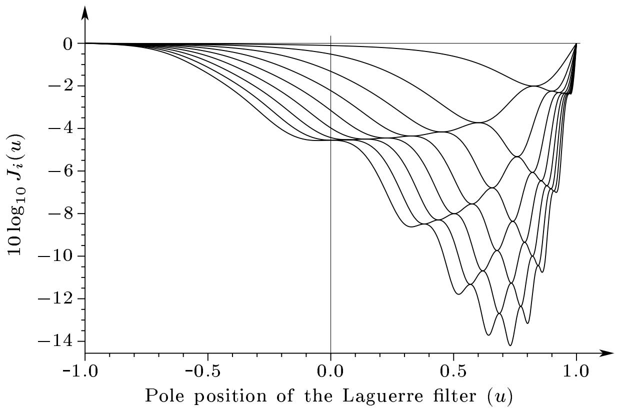

Figure 4: Normalized MSE curves for Laguerre filter of orders up to 8 when $\alpha = 1.8$ ($t < 25000$). The global minimum of the last curve ($n = 8$) occurs for $u \approx 0.73$.

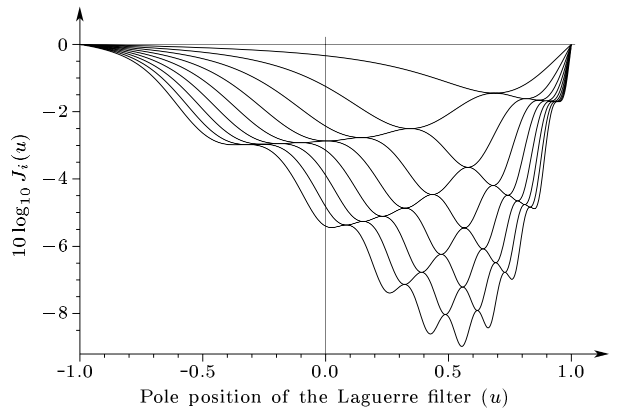

Figure 5: Normalized MSE curves for Laguerre filter of orders up to 8 when $\alpha = 1.6$ ($t \geq 25000$). The global minimum of the last curve ($n = 8$) occurs for $u \approx 0.56$.

in non-stationary environments), and this may help the algorithm to escape from local minima. This may be useful in some cases (see next section).

In another experiment we made, α changed linearly from 1.8 to 1.6 in the interval $25000 < t < 30000$, keeping λ and $\mu(t)$ as in the first experiment. In this experiment we observed that u was not able to follow the global minimum of the slowly changing MSE curve, although it was able to reach a better local minimum of the final MSE curve. That was to be expected, because u could not change quickly enough: in 5000 samples u could change by, at most, $5000 \times 10^{-5} = 0.05$, far less than the $0.73 - 0.56 = 0.17$ required (cf. Figs. 4 and 5). We have also observed that the global minimum was tracked successfully when $\mu(t)$ was changed to 10^{-4} in the interval $25000 < t \leq 40000$. With this "large" value of $\mu(t)$ the pole of the Laguerre filter was able to change quickly enough to follow the changing global minimum of the MSE. However, this large $\mu(t)$ gave rise to a clearly observable excess of MSE.

6 Suggestions about how to escape from local minima

Some standard ways of attempting to avoid local minima can be used in this problem. One of them is simulated annealing, which in our case can be implemented by adding a random term with an appropriately decreasing variance to the update formula for $u(t)$. One of our experiments suggests that decreasing λ produces an effect somewhat similar, since all estimated quantities will have higher variances. However, using this solution may increase substantially the excess of MSE.

In a stationary environment it would be possible (but not very easy) to use the results of [8] to estimate the position of local minima, with respect to u, of the error surface nearer to the current value of u. It would then be possible to jump to the best nearby local minimum, and, after a few such jumps, to reach, perhaps, the global minimum. The reason why this strategy might work is the experimental observation, with no exceptions found to date, that drawing line segments between consecutive local minima of the MSE of a Laguerre filter produces a unimodal piecewise linear curve.

References

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

[2] B. Wahlberg. System identification using Laguerre models. *IEEE Transactions on Automatic Control*, 36(5):551–562, May 1991.

[3] T. Oliveira e Silva. Laguerre filters – An introduction. *Revista do DETUA*, 1(3):237–248, January 1995. **URL:** `ftp://inesca.inesca.pt/pub/tos/English/revdetua.ps.gz`.

[4] Z. Fejzo and H. Lev-Ari. Adaptive Laguerre filters with lattice orthogonalization. In *Proceedings of Seventh SP Workshop on Statistical Signal & Array Processing*, pages 285–288, June 1994.

[5] T. Oliveira e Silva. Optimality conditions for Laguerre lattice filters. *IEEE Signal Processing Letters*, 2(5):97–98, May 1995.

[6] David G. Messerschmitt. A class of generalized lattice filters. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28(2):198–204, April 1980.

[7] H. J. W. Belt and A. C. den Brinker. Laguerre filters with adaptive pole optimization. In *Proceedings of the 1996 IEEE International Symposium on Circuits and Systems (ISCAS'96)*, Atlanta, GA, May 1996.

[8] T. Oliveira e Silva. On the determination of the optimal pole position of Laguerre filters. *IEEE Transactions on Signal Processing*, 43(9):2079–2087, September 1995.