PARAMETER ESTIMATION OF AN AUTOREGRESSIVE NON LINEAR PROCESS

Luis F. Rocha

Instituto de Ingeniería Biomédica
Universidad de Buenos Aires
Paseo Colón 850 - Buenos Aires
ARGENTINA

RESUME

Quand la sortie d'un système AR excité par une entrée de spectre ample se met en cascade à travers d'une fonction transfert no-linéal, l'identification du système est afecttée par un erreur dont la magnitude dépende de la nolinéalité . La solution a ce problème que est presenté ici est à travers de l'usage d'un identificateur formé par deux blocs, le premier est un transfert inverse et le seconde est le filtre classique "moving averages" ou "lattice".

Cet identificateur a été réalisé de manière adaptatif et sa vitesse de convergence est relativament lente.

SUMMARY

When the output of an autoregressive (AR) system, excited by a wide spectrum input is cascaded through a nonlinear function, the identification is affected by the nonlinearity, and the classic linear prediction methods of identification cannot be used in the normal way. Some methods presented before requiere the knowledge of the input sequence in order to estimate the nonlinear part of the system, which is a serious drawback in some cases. In this approach the identifier organized as a cascade of two inverse blocks one for the nonlinear part of the system, and the other one for the AR portion. In the adaptive version the speed is reduced.

Introduction

Recently Billing and Voon[1] have presented three algoritms to estimate the parameters of nonlinear systems in terms of input-output observables. Classically the approach has beeing the use of Volterra or Wiener series expansion[2-3], but the method is computer time consuming. In all the cases, the identifier makes use of both the input and the output sequences. As in some problems the input is not obsérvable the identification is only aproximative and for small nonlinearities.

In our approach the system to be identified is an AR model followed in cascade for a nonlinear no-memory transfer, excited by a zero mean, independent white noise $x(n)$.

The AR portion of the system can be represented by a linear difference equation:

$$y(n) = a_i\, y(n-i) + x(n)$$

cascaded by a nonlinear function $z(n)=f(y(n))$ and the observer has access only to the output $z(n)$.

To solve this problem, we can use a two blocks identifier. The first block is an inverse nonlinear no-memory transfer able to linearize the transfer of the system, and the second block is a standard MA or lattice filter.

The first block is expanded in a Taylor truncated polinomial:

$$w(z(n))=b_0+b_1.z(n)+b_2 z^2(n) + \ldots$$

or under the form of a linear picewise function. The second block is expanded as a MA or LPC inverse filter.

In the adaptive version of this identifier all the coefficients are upgraded every sampling time, looking for a flat (white) spectrum of the output.

As the adaptive method of parameter estimation allows the identification of slowly varying coefficients of the system, the speed of the estimation can be important. In this method, for typical cases, the number of samples is around several thousands , but very fast convergence and very small variance in the estimation is obtained if the input $x(n)$ can be measured.

As the model used to represent the system is nonlinear, results very sensitive to the amplitude of the input, but on the other side the truncation of the series limits the range of the identification. So we have to be very careful to select the input amplitude or the number of coefficients of the nonlinear block.

In all the simulations that we have made

no measurement noise is added to the system that is supposed to be free from this problem.

The inverse of the nonlinear function

The block that performs the inverse function $T^{-1} = f^{-1}(y(n))$ of the nonlinear part of the system to be identified $f(y(n))$ can be seen as an operator accomplishing the following operation: (see Fig.1)

$$T^{-1}(T)=1$$

Obviously, T has to be a single valued mapping of y(n) that is to say $z(n)=f(y(n))$. Otherwise the inversion is not longer possible.

In order to simplify the problem we can use non-memory mappings for both, the nonlinear part of the system and its inverse. In that case if $z=f(y)$ the application of the nonlinear inverse f^{-1} gives

$$f^{-1}(f(y)) - y = 0$$

For non perfect inverse the subtraction leaves an error, that in our case will be used to correct the coefficients of f^{-1}.

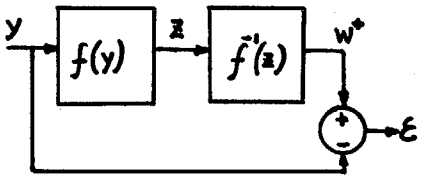

Fig.1

An example of the relation between the two applications is shown in Fig.2. The resulting functions are "mirror symmetric".

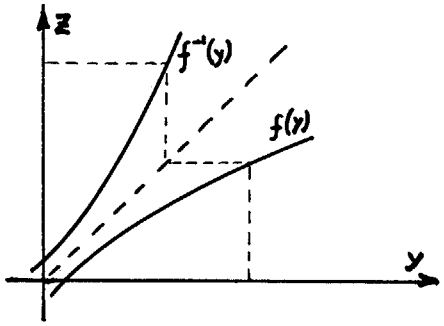

Fig.2

As normally the variable y is not given in closed form but as a set of samples of a signal x(n) used as input to the system and filtered by the AR part of the system, a method to build up the inverse function is to use an adaptive method to change the coefficients of f^{-1}. In that case (see Fig.1) $w^* - y = \varepsilon$ or

$$f^*(f(y)) - y = \varepsilon$$

where f^* is an aproximated inverse function and ε is the error.

The classic Widrow method of convergence to f^{-1} is to decrease (in the mean square sense) the value of ε using the algorithm:

$$f^*(k+1)=f^*(k) - \nabla f^*(k)$$

where the last term is the upgrading gradient calculated in every adaptation cycle.k.

In order to use this algorithm we have to expand f^* in an adecuate class of coefficients. If we use orthogonal coefficients the convergence will be very fast, but in order to simplify the operations we can use non orthogonal, simpler but with slow convergence.

Here we give two of the many possible expansions, one is the truncated Taylor polinomial and other one a linear picewise function.

In the Taylor expansion - useful for small nonlinearities - the upgrading is

$$c_i(k+1)=c_i(k) - \mu.\varepsilon(k).z^i$$

where

$$f(z)= c_0 + c_1 z + c_2 z^2 + \ldots$$

is the expanded nonlinear function and μ is a convergence constant, that in order to reduce the error variance is sometimes decreased very slowly if the system to be identified has constant output statistics.

On the other side, in the linear picewise expansion (see Fig.3) with N sections of the same size the ordinate w^* is every time calculated using: $j=z/L$ $i=\text{integer}(j).\text{sign}(j)$ $\Delta z = z - i.L$ in order to separate the value of the abscissa z in two components the first modulo-L and the rest Δz. After that we can define the function $w^* = A(i) + B(i).\Delta z$ where $A(i) = A(i-1) + B(i).L$

The coefficients A(i) and B(i) are upgraded with the following criterion: A(0) is corrected and also all B(i), but the remaining A(i) are automatically upgraded because they depend on A(0) and B(i) as expresed by the equation:

$$A(i)= A(i-1) + B(i).L \qquad 1 \leq i \leq N$$

So the upgrading is:

$$A^k(0)= A^{k-1}(0) - \mu.\varepsilon(k)$$

$$B^k(i)= B^{k-1}(i) - \mu\varepsilon(k)$$

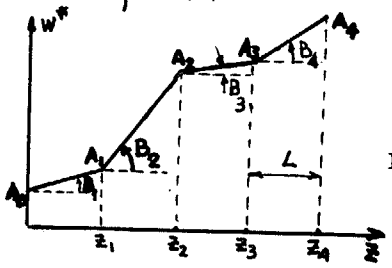

Fig.3

The linear picewise function has the advantage that can cope with very strong nonlinearities and it is very simple to program but as the coefficients with higher subindex i are strongly dependent on the others, this correlation decreases the speed of convergence.

The inverse of the AR block

As the identification is based on the cascade or series method using a zero mean independent white noise input (that is a "hidden" or not measured variable) exciting the linear AR block in the process, we can model its inverse using an adaptive method to update the coefficients of a moving averaged (MA) or a lattice filter. The latter has normally faster convergence. (Fig.4)

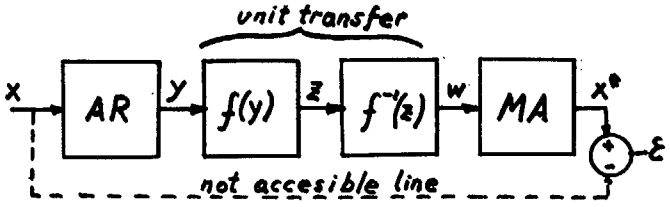

Fig.4

If the MA identification filter is expanded in the following form:

$$x^*(n) = 1 + \sum_{i=1}^{p} b_i . w(n-i) \qquad 1 \le i \le p$$

where b_i are the coefficients to be upgraded and p is the (supposed) maximum number of poles of the AR system filter, the upgrading algorithm is:

$$b_i^{k+1} = b_i^k - \mu \varepsilon^k w(n-i)$$

where k is the number of upgrading steps that frequently is the same that the sampling interval n, but can be changed to a multiple or submultiple of that interval.

If our choice is the lattice method we can use many convergence methods, each one with a particular merit[4-5]. In our simulations we have used (see Fig.5)

$$EF(n)=EF(n-1) - K(n).ED(n-1)$$
$$EB(n)=EB(n-1) - K(n).EF(n-1)$$
$$K(n)=K(n-1)+\mu_2(EF(n-1).(ED(n-1)-EF(n-1)K(n-1))$$

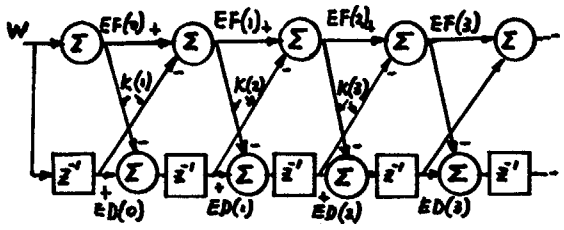

Fig.5

Experimental results

In a typical experiment, we have used as the process the following system:

$$y(n)= x(n)+1.2\ y(n-1) - 0.7\ y(n-2)$$

as the linear, second order AR process and

$$z(n)= y(n) + 0.2\ (y(n))^2$$

as a linear and quadratic nonlinearity.

Our choice as the nonlinear part of the identifier is a third order truncated Taylor polinomial:

$$w(n)= c_0 + c_1\ z(n) + c_2\ z^2(n) + c_3\ z^3(n)$$

and for the MA inverse filter we have used a lattice filter of two coefficients.

In the nonlinear identifier, we have normalized the value of c_0 forcing it to be the unity.

The initial value of μ used in the nonlinear identifier is set to .1 and the starting value of μ_2 of the lattice is set to .15.

Both coefficients are reduced to decrease the variance of the errors in the following form:

$$\mu^k = 0.9995\ \mu^{k-1}$$
$$\mu_2^k = 0.99995\ \mu_2^{k-1}$$

The convergence is slower than in linear identification, and an error remains in the value of the coefficients mainly due to the low number of terms in the Taylor polinomial but the results are fairly good.

In 1700 steps the results are:

c_0=0.0207 (should be zero)
c_1= 1 (forced)
c_2=-0.3705
c_3=-0.0562
K(1)= 0.66259 K(2)= -0.5088

In 10^4 steps the results are:

c_0=0.03218
c_1= 1
c_2= -0.2768
c_3= -0.08823
K(1)=0.6975 K(2)= -0.596

Conclusions

A "blinded" identification of a non-linear AR process can be realized with the method already presented. We mean blinded because we use only the output of the process and the statistics of the input (zero mean, independent white noise).

Two identification methods are used in

the nonlinear part of the identifier: a truncated Taylor series or a picewise expansion of the inverse function. The latter gives a better aproximation when the nonlinearity is strong, but its adaptation has to be done more slowly because the interdependence among the coefficients.

Both methods are only aproximate in the sense that always remains a variance in the estimation of the coefficients of the process due mainly to the imperfect inversion in the nonlinear block of the identifier.

Acknowledgment

This work was supported by the Consejo Nacional de Investigaciones Científicas y Técnicas de Argentina.

References

(1) S. A. BILLINGS and W.S.F. VOON "Least squares parameter estimation algorithms for non-linear systems" Int. J. Systems Sci. (1984) vol. 15 No 6, pp.601-615

(2) M. SCHETZEN "Nonlinear system modelling based on the Wiener theory" Proc. IEEE vol. 69 , pp.1557-1573, Dec. 1981.

(3) T. KOH and E. J. POWERS "Second-Order Volterra filtering and its application to nonlinear system identification" IEEE Trans. on Acoust. Speech and Signal Processing vol.ASSP-33 No. 6, pp.1445-1455, Dec. 1985.

(4) J. D. MARKEL and A. H. GRAY Linear Prediction of Speech Ed. Springer-Verlag N.Y. 1976.

(5) B. FRIEDLANDER "Lattice filters for adaptive processing" Proc. IEEE, vol.70 pp.829-867 Aug. 1982.