# NONLINEAR AUTOREGRESSIVE MODELING OF NON-GAUSSIAN SIGNALS USING $L_P$-NORM TECHNIQUES

*Ercan E. Kuruoğlu* *William J. Fitzgerald* *Peter J.W. Rayner*

Signal Processing and Communications Laboratory,
Department of Engineering, University of Cambridge, Cambridge, UK
e-mail : eek@eng.cam.ac.uk

## ABSTRACT

In this paper, for the estimation of the model coefficients of a polynomial autoregressive process with non-Gaussian innovations *least $l_p$-norm estimation* (LLPN) is suggested. Simulations showed that LLPN estimation leads to better estimates than the least squares estimation in terms of the mean and the standard deviations of the estimates. The algorithm is also employed in modeling audio data in non-Gaussian noise with the objective of separating signal from noise and superior results have been obtained when compared to the linear autoregressive modeling. Directions of future research is also addressed.

## 1. INTRODUCTION

One of the most popular mathematical models for a wide range of signal processing applications such as speech, audio and image has been linear autoregressive (AR) processes. The popularity of the linear AR models is motivated by the simple structure of the AR processes and the existence of efficient techniques for estimating the model parameters. However, in many applications, the signal under investigation is generated by some nonlinear dynamics, in which case linear models fall short of efficiently describing the data and the identification schemes fail. It has been shown that most biomedical signals such as EEG, [1] ECG and HR time series [2] and some geophysical signals [3] are more successfully represented by *nonlinear autoregressive* (NAR) processes.

Nonlinear autoregression provides one with a wide range of different models to choose from, just to name a few, exponential autoregressive processes (EXPAR), threshold autoregressive processes (TAR), random coefficient autoregressive processes (RCA), etc [4]. However, most of these models do not suggest any procedures for easily estimating the model parameters and their potential use is limited only to specific applications. One family of NAR models, namely *polynomial autoregressive* (PAR) models *polynomial autoregressive models* (PAR) which can be represented as

$$y(n) = \sum_i a_i^{(1)} x(n-i) + \sum_i \sum_j a_{i,j}^{(2)} x(n-i)x(n-j) \ldots + \epsilon(n) \tag{1}$$

do not share these drawbacks. They possess the property of being *linear in the parameters* and therefore many mathematical tools developed for linear models can be extended to accommodate for polynomial models. As can be seen from Eq. (1), polynomial autoregressive models are based on the Volterra series expansion which has been employed with great success in nonlinear system modelling, (e.g. in [5]) and which are theoretically very attractive since it can be shown that a very large class of nonlinear systems which do not have saturating elements can be represented by Volterra series [6]. A main motivation of linear (AR) modelling was the proof by Kolmogorov that every linear system could be represented by a linear AR model of infinite order. Similarly, it can be shown that every nonlinear system with a Volterra series expansion can be represented as a polynomial AR model of infinite order [7]. Therefore, PAR models are general enough to model various nonlinear generation mechanisms for signals. The main drawback of the PAR models is the *curse of dimensionality*: as the order of nonlinearity and the memory size increased, the number of model parameters increases geometrically. However, one can delete some of the terms in the expansion that are of less importance and use a partial model (see e.g. [9]).

## 2. NONLINEAR AUTOREGRESSIVE PROCESSES WITH NON-GAUSSIAN INNOVATIONS

In almost all of the work published so far, it has been assumed that the excitation (or the innovation) sequence $\epsilon(n), n = 1, 2, \ldots$ in Eq. (1) is a time serie of i.i.d. Gaussian random numbers. However, in some applications the time series exhibit characteristics that cannot be accommodated by the Gaussian innovations (e.g. in the case of skewed data). Examples include the wind speed data, the service time in a queue or the daily flows of a river. Lawrance and Lewis suggest employing exponential variables because of its simplity and analytical tractability and they provide simulation results for wind velocity data [8]. However, we believe that a more flexible probability distribution function should be used to model the excitations.

Employing non-Gaussian excitations is further motivated by the fact that many noise processes such as atmospheric noise, underwater acoustic noise and low frequency electromagnetic disturbances show an impulsive character which points out an underlying distribution with heavier tails than the Gaussian p.d.f.. Although various models were suggested for impulsive noise, $\alpha$-stable processes which only recently attracted the attention of the signal processing processing community with the works of Nikias *et al.* (e.g. [10]) seem to be the most general and accurate class since they can be derived rigorously as the limiting distribution of noise processes based on realistic assumptions on the generation mechanisms of noise. One further justification of $\alpha$-stable noise model is the *generalized limit theorem* which states that the limit sum of a large number of random variables with possibly infinite variances is distributed with a stable law [11].

A random variable is called $\alpha$-stable if its characteristic function (which is the Fourier transform of the pdf) can be expressed in the following form:

$$\varphi(z) = \exp\{jaz - \gamma|z|^{\alpha}[1 + j\beta \operatorname{sign}(z)\, w(z,\alpha)]\} \quad (2)$$

where $w(z,\alpha) = \tan\frac{\alpha\pi}{2}$, if $\alpha \neq 1$ and $w(z,\alpha) = \frac{2}{\pi}\log|z|$ if $\alpha = 1$ and $-\infty < a < \infty$, $\gamma > 0$, $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$. The *characteristic exponent*, $\alpha$, is a measure of the thickness of the tails of the distribution. The special cases $\alpha = 2$ and $\alpha = 1$ with $\beta = 0$ correspond to the Gaussian distribution and the Cauchy distribution respectively. The *symmetry parameter*, $\beta = 0$, corresponds to a distribution that is symmetric around $a$, in which case the distribution is called *Symmetric $\alpha$-Stable* ($S\alpha S$). The *location parameter*, $a$, is the symmetry axis. Finally, the *dispersion*, $\gamma$, in analogy to the variance of the Gaussian distribution, is a measure of the deviation around the mean. A thorough discussion of $\alpha$-stable distributions can be found in [10].

We believe that based on the above facts $\alpha$-stable distribution is a very appropriate model for the excitation of PAR. It is flexible, has the special case of Gaussian distribution and general enough to investigate the potentials of nonlinear autoregressive modelling.

## 3. POLYNOMIAL AUTOREGRESSIVE PROCESS COEFFICIENT ESTIMATION

### 3.1. Least Squares Estimation

Since the polynomial autoregressive model is linear in the parameters, we can incorporate the linear and polynomial data terms and the coefficients for the linear and polynomial terms into an extended data vector and an extended parameter vector respectively and express PAR with the following linear equation:

$$x(n) = \mathbf{x}^T_{N,ext}(n)\mathbf{a}_{N,ext} \quad (3)$$

where

$$\mathbf{x}_{N,ext}(n) = [\mathbf{x}^{(1)}_N(n)\mathbf{x}^{(2)}_N(n) \ \dots \ \mathbf{x}^{(P)}_N(n)]^T \quad (4)$$

and

$$\mathbf{a}_{ext} = [\mathbf{a}^{(1)}_N\mathbf{a}^{(2)}_N \ \dots \ \mathbf{a}^{(P)}_N]. \quad (5)$$

Above, $P$ is the order of the truncated polynomial autoregressive model, $N$ is the memory size of the filter and $\mathbf{x}^{(1)}_N = [x(n-1)\ x(n-2)\dots x(n-N)]^T$, $\mathbf{x}^{(2)}_N = [x^2(n-1)\ x(n-1)x(n-2)\dots x^2(n-N)]^T$, etc, and $\mathbf{a}^{(k)}_N$'s are the corresponding PAR coefficients. Then, the usual least squares problem can be solved to obtain the coefficients as

$$\mathbf{a}_{N,ext} = E[\mathbf{x}_{N,ext}\mathbf{x}^T_{N,ext}]^{-1}E[\mathbf{x}_{N,ext}x(n)]. \quad (6)$$

Almost all of the previous work using PAR for modelling have suggested using least squares for estimating the PAR coefficients. However, least squares estimate is optimal only for Gaussian data. Specifically, in the case of $\alpha$-stable data for which the variance diverges to infinity, the estimates of the second order moments employed in Eq. (6) are not reliable and therefore alternative techniques should be employed.

### 3.2. Least $l_p$-Norm Estimation

As mentioned above, the analogue of variance in an $\alpha$-stable variable is the dispersion $\gamma$ which is related to its moments with the following equation [10]:

$$\mathbf{E}(|X|^p) = C(p,\alpha)\gamma^{p/\alpha} \qquad -1 < p < \alpha. \quad (7)$$

Then similar to the MMSE criterion for Gaussian signals which lead to least squares estimation, *minimum dispersion* (MD) criterion [10] can be defined which leads to *least $l_p$-norm estimation*:

$$\hat{\mathbf{a}}_{N,ext} = \operatorname{argmin}\sum_n |x(n) - \mathbf{x}^T_{N,ext}(n)\hat{\mathbf{a}}_{N,ext}|^p, \quad (8)$$

where $\hat{\mathbf{a}}$ is the estimate of the actual $\mathbf{a}$. It should be emphasized that the least $l_p$-norm estimates are not only optimal in MD sense for $\alpha$-stable data but also are optimal in maximum likelihood sense for another family of distributions namely the *generalized Gaussian distribution* which recently found applications in the field of image processing [12].

For the solution of this $l_p$-norm minimization problem, we propose an extension of the *iteratively reweighted least squares algorithm* (IRLS) algorithm which can be coded compactly as:

1. $\hat{\mathbf{a}}(0) = (\mathcal{X}^T\mathcal{X})^{-1}\mathcal{X}^T\mathbf{x_N}$
2. $r_i(k) = (\mathbf{x}_N - \mathcal{X}\hat{\mathbf{a}}(k))_i$
3. $W_{ii}(k) = p|r_i(k)|^{p-2}$
4. $\hat{\mathbf{a}}(k+1) = (\mathcal{X}^TW(k)\mathcal{X})^{-1}\mathcal{X}^TW(k)\mathbf{x}_N$
5. if $\frac{||\mathbf{r}(k+1)||_{(p)} - ||\mathbf{r}(k)||_{(p)}}{||\mathbf{r}(k)||_{(p)}} < \epsilon$ then stop, else go to step 2.

where

$$\mathcal{X} = \begin{bmatrix} x[1] & 0 & \dots & 0 \\ x[2] & x[1] & \dots & 0 \\ \vdots & \vdots & & \vdots \\ x[L] & x[L-1] & \dots & x[L-N+1] \end{bmatrix}. \quad (9)$$

and the weigthing matrix $W$ is a diagonal matrix with diagonal entries

$$W_{ii} = \begin{cases} |r_i|^{p-2}, & |r_i| > \varepsilon \\ \varepsilon^{p-2}, & |r_i| < \varepsilon \end{cases}. \quad (10)$$

For the solution of Eq.(8), we replace the data matrix $\mathcal{X}$ with the Volterra data matrix $\mathbf{X}$ defined as

$$\mathbf{X}^{(K)} = [\ \mathcal{X}^{(1)}\ \ \mathcal{X}^{(2)}\ \ \dots\ \ \mathcal{X}^{(K)}\ ] \quad (11)$$

| α | L | method | std | estimates |
|---|---|---|---|---|
| | | actual | | 0.5000 -0.0100 -0.0020 |
| 1.0 | 100 | LLMN | 0.2338 | 0.5046 -0.0147 -0.0022 |
| | | LS | 0.5541 | 0.5308 -0.0502 -0.0028 |
| | 1000 | LLMN | 0.5719 | 0.4987 -0.0163 0.0010 |
| | | LS | 1.1270 | 0.5008 -0.0208 0.0010 |
| 1.5 | 100 | LLMN | 0.5047 | 0.5151 -0.0260 -0.0055 |
| | | LS | 0.5327 | 0.5050 -0.0325 -0.0045 |
| | 1000 | LLMN | 0.2520 | 0.4982 -0.0130 -0.0011 |
| | | LS | 0.5243 | 0.4950 -0.0194 -0.0009 |

**Table 1. Estimates of the model (I); L: block length, std: total standard deviation**

| α | L | method | std | estimates |
|---|---|---|---|---|
| | | actual | | 0.5000 -0.0500 0.0100 0.0020 -0.0010 0.0005 0.0001 |
| 1.0 | 100 | LLMN | 0.1483 | 0.5026 -0.0488 0.0044 0.00089 0.0015 -0.00088 0.00021 |
| | | LS | 0.8310 | 0.4728 -0.0557 -0.0183 0.000017 0.0061 0.0021 0.00011 |
| | 1000 | LLMN | 0.0122 | 0.5002 -0.0507 0.0097 0.00098 -0.00056 0.00007 -0.00001 |
| | | LS | 0.1131 | 0.5097 -0.0646 0.0207 0.00101 -0.00010 -0.00013 -0.00001 |
| 1.5 | 100 | LLMN | 0.3677 | 0.5000 -0.0500 0.0100 0.0010 -0.0005 0.0001 -0.00001 |
| | | LS | 0.6209 | 0.5056 -0.0594 0.00032 0.0051 -0.0039 0.00027 -0.00199 |
| | 1000 | LLMN | 0.0654 | 0.4983 -0.0532 0.0090 0.0018 -0.00034 0.00041 0.00005 |
| | | LS | 0.1351 | 0.4973 -0.0497 0.0035 0.00115 0.00061 -0.00073 0.00001 |

**Table 2. Estimates of the model (II); L: block length, std: total standard deviation**

where $K$ is the order of the nonlinearity, and $\mathcal{X}^{(i)}$ are the data matrices containing the polynomial terms of degree $i$ from the Volterra expansion, e.g. :

$$\mathcal{X}^{(2)} = \begin{bmatrix} x^2[1] & 0 & \dots & 0 \\ x^2[2] & x[2]x[1] & \dots & 0 \\ \vdots & \vdots & & \vdots \\ x^2[N] & x[N]x[N-1] & \dots & x^2[1] \\ \vdots & \vdots & & \vdots \\ x^2[L] & x[L]x[L-1] & \dots & x^2[L-N+1] \end{bmatrix}, \quad (12)$$

etc.

## 4. SIMULATION STUDIES

### 4.1. PAR Process Coefficient Estimation

To evaluate the perfomance of least $l_p$-norm minimization in estimation of the parameters of polynomial autoregressive processes with non-Gaussian innovations, and to be able to compare with the performance of least squares technique, the following PAR processes were generated: $(I) : x(n) = 0.5x(n-1) - 0.01x(n-2) - 0.002x(n-1)x(n-2) + \epsilon(n)$ and $(II) : x(n) = 0.5x(n-1) - 0.05x(n-2) + 0.01x(n-3) + 0.002x(n-1)x(n-2) - 0.001x(n-1)x(n-3) + 0.0005x(n-2)x(n-3) + 0.0001x(n-1)x(n-2)x(n-3) + \epsilon(n)$ where $\epsilon(n)$ are $\alpha$-stable distributed (specifically with $\alpha = 1$ or $\alpha = 1.5$). For each model and each $\alpha$ choice 40 sequences of lengths 100 and 1000 were generated. The results are presented in Table 1, Table 2.

From the simulation results, we observe that sequences of length 100 are not enough for efficient estimation since very high standard deviations were observed in this case. Considering results obtained with sequences of length 1000, Least $l_p$-norm parameter estimates have significantly lower standard deviation and are closer to the actual parameter values. Despite LLPN's superiority to LS estimation, for tap weights with relatively smaller magnitudes, percentage estimation errors in some cases were seen to be large. This characteristic is mainly due to the high condition numbers of the matrices employed in the equations which is caused by the fact that the eigenvalues of the polynomial systems are dispersed over a much wider range. Singular Value Decomposition can be applied to improve the solutions.

### 4.2. Modelling Audio Data in Impulsive Noise

Next, least $l_p$-norm parameter estimation method is applied in modeling real audio signal corrupted with impulsive noise with the PAR model in Eq. (1). The data shown in Fig.(1) is modeled with first (linear) and second (quadratic) order PAR processes with $\alpha$-stable innovations, the results of which are given in Fig.(2) and Fig.(3), respectively.

The modelling process with quadratic LLPN filter effectively seperated the audio signal and the impulsive noise which acts as the innovations sequence for the PAR process, while the linear filter has been very unsuccessful in modeling the the audio signal and in seperating it from the impulsive noise. The appeal of the suggested algorithm is that it can be applied to any data with outliers. Epileptic EEG is one of the many possible application areas where nonlinear autoregressive modeling with least $l_p$-norm estimation can lead to a better understanding of the data.

## 5. CONCLUSION

In this work, a new technique is suggested for the estimation of the parameters of PAR processes with non-Gaussian innovations. The innovations are modelled with $\alpha$-stable distribution which provides a very wide class of distributions. The suggested method in general showed better performance than least squares estimation in terms of the mean and the standard deviation of the estimates. One potential application of the new technique is demonstrated with an example which gave encouraging results. Modelling with other nonlinear autoregressive processes and model order estimation are the subjects of the current research. Biomedical and other types of data are also under investigation.

## REFERENCES


[1] N. D. Schiff, J. D. Victor, A. Canel, and D. R. Labar, "Characteristic nonlinearities of the 3/s ictal electroencephalogram identified by nonlinear autoregressive analysis," *Biological Cybernetics,* Vol. 72, pp. 519-526, 1995.

[2] D. J. Christini, F. M. Bennett, K. R. Lutchen, H. M. Ahmed, J. M. Hausdorff, and N. Oriol, "Application of linear and nonlinear time series modeling to heart rate dynamics analysis," *IEEE Trans. Biomedical Eng.,* Vol. 42, No. 4, pp. 411-415, April 1995.

[3] M. Bracalari and E. Salusti, "On a nonlinear autoregressive method applied to geophysical signals," *Geophysics,* Vol.59, No.8, pp.1270-1277, 1994.

[4] H. Tong, *Non-linear time series,* Oxford University Press, Oxford, UK, 1995.

[5] T. Koh and E.J. Powers, "Second-Order Volterra Filtering and Its Application to Nonlinear System Identification," *IEEE Trans. Acoust. Speech, Signal Processing,* Vol. 33, No. 6, December 1985.

[6] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems,* New York: John Wiley & Sons, 1980.

[7] L. R. Hunt, R. D. DeGroat, and D. A. Linebarger, "Nonlinear AR modeling," *Circuits, Systems, Signal Proc.,* Vol. 14, No. 5, pp. 689-705, 1997.

[8] A. J. Lawrance and P. A. W. Lewis, "Modelling and residual analysis of nonlinear autoregressive time series in exponential variables," *J. Royal Statist. Soc. B,* Vol. 47, No. 2, pp. 165-202, 1985.

[9] K. C. Nisbet, B. Mulgrew, and S. Mclauglin, "Reduced state methods in nonlinear prediction," *Signal Processing,* Vol. 48, pp. 37-49, 1996.

[10] M. Shao, C.L. Nikias, "Signal Processing with Fractional Lower Order Moments: Stable Processes and Their Applications," *Proc. IEEE,* vol.81, no.7, July 1993.

[11] W. Feller, *An Introduction to Probability Theory and Its Applications,* Vol. II, John Wiley & Sons, 1966.

[12] F. Muller, "Distribution shape of two-dimensional DCT coefficients of natural images," *Electronics Letters,* Vol. 29, No. 22, pp. 1935-1936, 1993.


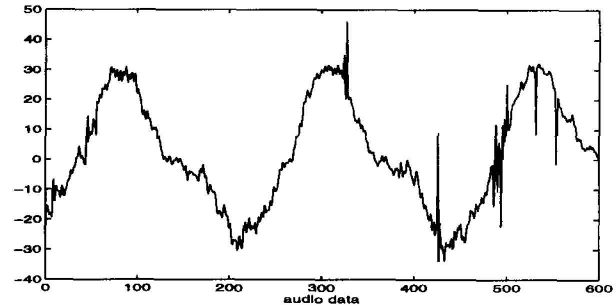


**Figure 1.**

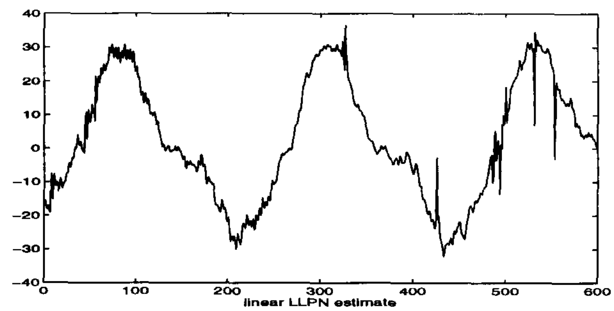


**Figure 2.**

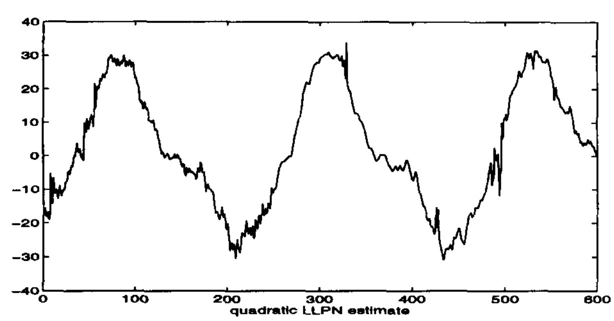


**Figure 3.**