

ESTIMATION OF THE NONLINEARITY DEGREE FOR POLYNOMIAL AUTOREGRESSIVE PROCESSES WITH RJMCMC

Oktay Karakuş*, Ercan E. Kuruoğlu†, Mustafa A. Altinkaya*

* İzmir Institute of Technology
Electrical-Electronics Engineering
Urla, İzmir, Turkey

† ISTI-CNR,
via G. Moruzzi 1, 56124,
Pisa, Italy

ABSTRACT

Despite the popularity of linear process models in signal and image processing, various real life phenomena exhibit nonlinear characteristics. Compromising between the realistic and computationally heavy nonlinear models and the simplicity of linear estimation methods, linear in the parameters nonlinear models such as *polynomial autoregressive* (PAR) models have been accessible analytical tools for modelling such phenomena. In this work, we aim to demonstrate the potentials of Reversible Jump Markov Chain Monte Carlo (RJMCMC) which is a successful statistical tool in model dimension estimation in nonlinear process identification. We explore the capability of RJMCMC in jumping not only between spaces with different dimensions, but also between different classes of models. In particular, we demonstrate the success of RJMCMC in sampling in linear and nonlinear spaces of varying dimensions for the estimation of PAR processes.

Index Terms— Polynomial AR, Reversible Jump MCMC, Nonlinearity degree estimation.

1. INTRODUCTION

Nonlinear models may be more favorable than linear models when data is generated from a nonlinear source or when linear modelling requires too many model parameters. In spite of the popularity of nonlinear models, e.g. Nonlinear autoregressive (NAR) models, most of these models do not provide a solution to estimate the model parameters easily and their potential use is very limited. Polynomial autoregressive (PAR) differs from the other NAR processes since they are *linear-in-the-parameters* and thus many mathematical applications developed for linear models can be employed without much difficulty [1].

PAR models can be represented as in (1) where $\epsilon(n)$ is excitation sequence with distribution $\mathcal{N}(0, \sigma_\epsilon^2)$, $a_i^{(1)}$, $a_{i,j}^{(2)}$ and $a_{i,\dots}^{(p)}$ are PAR coefficients for first order, second order and p^{th} order polynomials, respectively, p is the nonlinearity degree and k is the AR memory of the PAR model. A PAR model

can be represented in the notation: $P^{(p)}AR(k)$.

$$x(n) = \sum_i^k a_i^{(1)} x(n-i) + \sum_i^k \sum_j^k a_{i,j}^{(2)} x(n-i)x(n-j) + \dots + \sum_{i,\dots}^{k,\dots} a_{i,\dots}^{(p)} x(n-i)\dots + \epsilon(n). \quad (1)$$

PAR models are based on the Volterra series expansion as suggested by (1). This expansion has great success in modelling many real life phenomena such as a pilot plant control system [2], short term wind speed prediction [3], brain signals [4], seismology [5] and communications [6].

Reversible Jump Markov Chain Monte Carlo (RJMCMC) was first introduced in [7] as a Bayesian model identification tool. It is an extended and generalized version of the MCMC algorithm which is used when the parameter subspace dimension is fixed. RJMCMC provides an algorithm for the construction of reversible Markov chain samplers which jump between parameter subspaces of different dimensions. In the literature, RJMCMC has been used in linear model identification problems, e.g. in [8, 9] for autoregressive (AR) time series model selection. However, the sampling strategy of RJMCMC is not limited to linear models [7] and can also be used for nonlinear model identification problems in a wide range of applications.

In this study, we are going to use RJMCMC algorithm to estimate the nonlinearity degree and the AR length of various synthetically generated PAR processes. Generally [2–6], the nonlinearity degree of selected PAR model, is assumed to be known. This is not a realistic assumption and in practical problems the nonlinearity degree needs to be estimated as well.

In this paper, we contribute to the literature on RJMCMC by demonstrating the potentials of this method beyond linear models and in particular we provide results on PAR process estimation where we estimate the nonlinearity order as well in contrast to previous works.

The rest of the paper is organized as follows: PAR models and RJMCMC methodology for PAR model selection problem are examined in Section 2. The results of the simulations

are provided in Section 3. Section 4 concludes the paper with a brief summary.

2. METHODOLOGY

2.1. PAR Models

A $P^{(p)}$ AR(k) model given by (1) can be represented in matrix-vector form by using the *linear-in-the-parameters* property;

$$\mathbf{x} = \mathbf{X}\mathbf{a}^{(p,k)} + \epsilon, \quad \text{where} \quad \epsilon \sim \mathcal{N}(0, \sigma_e^2), \quad (2)$$

where \mathbf{x} is $n \times 1$ data vector, \mathbf{X} is $n \times w$ matrix of past samples and polynomial products of the data, $\mathbf{a}^{(p,k)}$ is $w \times 1$ coefficient vector of $P^{(p)}$ AR(k) model and ϵ is $n \times 1$ vector of excitation sequence. In the last row of Table 1, the number of model coefficients, w , for different p and k pairs are shown.

$$\mathbf{x} = [x[1], \dots, x[n]]^T, \quad \epsilon = [\epsilon[1], \dots, \epsilon[n]]^T, \quad (3)$$

$$\mathbf{X} = \begin{bmatrix} x[0] & \dots & x[1-k] & x^2[0] & x[0]x[-1] & \dots & x^p[1-k] \\ x[1] & \dots & x[2-k] & x^2[1] & x[0]x[0] & \dots & x^p[2-k] \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x[n-1] & \dots & x[n-k] & x^2[n-1] & x[n-1]x[n-2] & \dots & x^p[n-1] \end{bmatrix}, \quad (4)$$

$$\mathbf{a}^{(p,k)} = \begin{bmatrix} a_1, a_2, \dots, a_k, a_{1,1}, a_{1,2}, \dots, a_{k,k}, \dots, \underbrace{a_{k,k,\dots,k}}_{p^{\text{th}} \text{ order}} \end{bmatrix}^T. \quad (5)$$

Since the excitation sequence is Gaussian, the approximate likelihood of a PAR data \mathbf{x} , can be defined by using the derivation about AR model order selection problem in [9];

$$f(\mathbf{x}|p, k, \mathbf{a}^{(p,k)}, \sigma_e^2) \approx \mathcal{N}(\epsilon|\mathbf{0}, \sigma_e^2 \mathbf{I}_n). \quad (6)$$

2.2. Hierarchical Model and Priors

The posterior density, $f(\theta|\mathbf{x})$, can be written from Bayes Theorem easily where parameter vector θ consists of the parameters, $\{p, k, \mathbf{a}^{(p,k)}, \sigma_e^2, \sigma_a^2\}$;

$$f(p, k, \mathbf{a}^{(p,k)}, \sigma_e^2, \sigma_a^2|\mathbf{x}) \propto f(\mathbf{x}|p, k, \mathbf{a}^{(p,k)}, \sigma_e^2) \times f(\mathbf{a}^{(p,k)}|p, k, \sigma_a^2) f(\sigma_e^2) f(k) f(p). \quad (7)$$

The noninformative priors for the AR memory k , and the nonlinearity degree p , are chosen as uniform with maximum values k_{max} and p_{max} respectively;

$$f(k) = \mathcal{U}(1, k_{max}) \quad \text{and} \quad f(p) = \mathcal{U}(1, p_{max}). \quad (8)$$

We also choose conjugate priors for the other parameters [9];

$$f(\mathbf{a}^{(p,k)}|p, k, \sigma_a^2) = \mathcal{N}(\mathbf{a}^{(p,k)}|\mathbf{0}, \sigma_a^2 \mathbf{I}_w), \quad (9)$$

$$f(\sigma_a^2) = \mathcal{IG}(\sigma_a^2|\alpha_a, \beta_a), \quad (10)$$

$$f(\sigma_e^2) = \mathcal{IG}(\sigma_e^2|\alpha_e, \beta_e). \quad (11)$$

The key property of RJMCMC or all MCMC based methods is to provide convergence, even though we have nonstandard priors and likelihoods with systems that cannot be observed directly or with missing data [10].

2.3. RJMCMC Methodology

Following [7], when the current state is κ , we propose a move type m with probability $Pr(\kappa \rightarrow \kappa')$, which changes dimension, and takes the state to κ' . The acceptance probability, which is denoted by $\alpha(\kappa \rightarrow \kappa')$, needs to be calculated.

As stated above, RJMCMC enables to jump between spaces with different dimensions. There occurs a dimension difference between the two states and to provide dimension matching, auxiliary variables u_1 with length m_1 , are proposed from distributions $q_1(u_1)$. The resulting dimension matching is; $dim_{\kappa'} = dim_{\kappa} + m_1$ [7].

The general expression for acceptance ratio which defines a move with dimension change from state κ to a proposed state κ' is given below. This form of acceptance ratio is defined as Equation (8) of [7];

$$\alpha_{\text{RJMCMC}}(\kappa \rightarrow \kappa') = \min \left\{ 1, \frac{f(\theta^*|\mathbf{x}) Pr(\kappa' \rightarrow \kappa)}{f(\theta|\mathbf{x}) Pr(\kappa \rightarrow \kappa') q_1(u_1)} \left| \frac{\partial \theta^*}{\partial(\theta, u_1)} \right| \right\}, \quad (12)$$

where $f(\cdot|\mathbf{x})$ is the target distribution of interest, $Pr(\kappa \rightarrow \kappa')$ and $Pr(\kappa' \rightarrow \kappa)$ represent the probabilities for move m and its reverse move, $q_1(u_1)$ is the proposal distribution for auxiliary variable vector u_1 for move m , θ and θ^* are parameter vectors for states κ' and κ respectively and $\left| \frac{\partial \theta^*}{\partial(\theta, u_1)} \right|$ is the magnitude of the Jacobian determinant.

Sampling AR length k and nonlinearity degree p require changing the dimension in the RJMCMC. Firstly we define the move pairs for this problem. There are 4 types of moves for this problem which are, *birth of a new parameter*, *death of an existing parameter*, *updating the AR coefficients via MCMC algorithm (life move)*, *updating σ_e^2 via Gibbs Sampling*.

Each move has probabilities P_{birth} , P_{death} and P_{life} satisfying $P_{\text{birth}} + P_{\text{death}} + P_{\text{life}} = 1$.

Constructing an RJMCMC for this problem begins firstly by defining a birth move (birth of a new variable causes an increase in dimension) from AR memory parameter k to k' where nonlinearity degree p is fixed. After computing the posterior distributions as in (7), the acceptance ratio, $\alpha_{\text{birth}} = \min\{1, r_{\text{birth}}\}$;

$$r_{\text{birth}} = \frac{f(\mathbf{x}|p, k', \mathbf{a}^{(p,k')}, \sigma_e^2)}{f(\mathbf{x}|p, k, \mathbf{a}^{(p,k)}, \sigma_e^2)} \times \frac{f(\mathbf{a}^{(p,k')})|p, k', \sigma_a^2}{f(\mathbf{a}^{(p,k)})|p, k, \sigma_a^2} \times \frac{P_{\text{death}}}{P_{\text{birth}} q_1(\mathbf{u}_1)} \times \left| \frac{\partial \mathbf{a}^{(p,k')}}{\partial(\mathbf{a}^{(p,k)}, \mathbf{u}_1)} \right|. \quad (13)$$

For a death move, we propose a move m which changes states from k to k' where $k' < k$. Here death move will be applied and no new parameters are proposed. We remove the coefficients which belong to AR memory k of parameter vector $\mathbf{a}^{(p,k)}$.

The acceptance ratio of death move, $\alpha_{\text{death}}(k \rightarrow k') = \min\{1, r_{\text{death}}\}$, is calculated directly from;

$$\alpha_{\text{death}}(k \rightarrow k') = \min\{1, 1/r_{\text{birth}}\}, \quad (14)$$

where r_{birth} is calculated from $\alpha_{\text{birth}}(k' \rightarrow k)$ as in (13).

Life move applies MCMC to update the PAR coefficients due to there is no dimension change for the life move. Acceptance ratio is defined as $\alpha_{\text{life}} = \min\{1, r\}$. The corresponding value of r is given below;

$$r = \frac{f(\mathbf{x}|p, k', \mathbf{a}^{(\mathbf{p}, k')}, \sigma_e^2)}{f(\mathbf{x}|p, k, \mathbf{a}^{(\mathbf{p}, k)}, \sigma_e^2)} \times \frac{f(\mathbf{a}^{(\mathbf{p}, k')}|p, k', \sigma_a^2)}{f(\mathbf{a}^{(\mathbf{p}, k)}|p, k, \sigma_a^2)} \quad (15)$$

$$\times \frac{q(\mathbf{a}^{(\mathbf{p}, k)}|p, k', \mathbf{a}^{(\mathbf{p}, k')})}{q(\mathbf{a}^{(\mathbf{p}, k')}|p, k, \mathbf{a}^{(\mathbf{p}, k)})},$$

where $f(\mathbf{x}|\cdot)$ is likelihood distribution and $f(\mathbf{a}^{(\mathbf{p}, k')}|\cdot)$ is prior distribution for parameter vector. Updated values of parameters are proposed from the distribution $q(\mathbf{a}^{(\mathbf{p}, k')}|p, k, \mathbf{a}^{(\mathbf{p}, k)})$ which is defined below;

$$\mathbf{a}^{(\mathbf{p}, k')} \sim q(\mathbf{a}^{(\mathbf{p}, k')}|p, k, \mathbf{a}^{(\mathbf{p}, k)}) = \mathcal{N}(\mathbf{a}^{(\mathbf{p}, k')}|\mu_n, \sigma_e^2 \Sigma_n^{-1}), \quad (16)$$

where μ_n and Σ_n is,

$$\mu_n = \mathbf{a}^{(\mathbf{p}, k)}, \quad \text{and} \quad \Sigma_n = \mathbf{X}^T \mathbf{X} + \frac{1}{\sigma_a^2} \mathbf{I}_w. \quad (17)$$

Independently from defined reversible moves, excitation variance σ_e^2 is updated at each iteration as in (20) by using all other parameters which are sampled at recent iteration. The updating mechanism follows Gibbs Sampling methodology. The posterior distribution for σ_e^2 is calculated as [9];

$$f(\sigma_e^2|\mathbf{x}, p, k, \mathbf{a}^{(\mathbf{p}, k)}) \propto f(\mathbf{x}|k, \mathbf{a}^{(\mathbf{p}, k)}, \sigma_e^2) f(\sigma_e^2) \quad (18)$$

$$f(\sigma_e^2|\mathbf{x}, p, k, \mathbf{a}^{(\mathbf{p}, k)}) \approx \mathcal{N}(\epsilon|\mathbf{0}, \sigma_e^2 \mathbf{I}_n) \mathcal{IG}(\sigma_e^2|\alpha_n, \beta_n) \quad (19)$$

$$= \mathcal{IG}(\sigma_e^2|\alpha_n, \beta_n), \quad (20)$$

where $\alpha_n = \alpha_e + \frac{1}{2}n$ and $\beta_n = \beta_e + \frac{1}{2}\epsilon^T \epsilon$.

PAR model selection procedure requires 2 cascaded RJMCMC steps for AR memory k , and nonlinearity degree p at each iteration. For each parameter, the algorithm proposes a candidate parameter according to the selected move type from proposal distribution. Then, it calculates acceptance ratio and tests the decision for this candidate. Equations between (13)-(17) which are derived for AR length, k , when p is fixed, will be converted to a state that k is fixed and p will be estimated, when 2^{nd} RJMCMC step is active. The steps of the algorithm are given in Algorithm 1.

Each RJMCMC step updates one of the model parameters that is either p or k . As one might see clearly from Table 1, e.g. if a birth move corresponding to a model change from $P^{(2)}\text{AR}(2)$ to $P^{(3)}\text{AR}(2)$ or in a similar manner, a jump from

Algorithm 1 RJMCMC for PAR model selection

```

1: Given data  $\mathbf{x}$ 
2:  $k_0, p_0, N_{iter}, a^{(p_0, k_0)}, \sigma_e^2(0)$  values are initialize at  $t = 0$ 
3: for do  $t = 1 : N_{iter}$ 
4:   RJMCMC for  $k$  Use  $p_{t-1}$  for  $p$ 
5:   Propose a move  $m$  with  $P_{\text{birth}}, P_{\text{death}}, P_{\text{life}}$ 
6:   Assign  $k^*$  for selected move  $m$ 
7:   Sample  $u_1 \sim q_1(u_1)$ 
8:   Construct  $a^{(p_{t-1}, k^*)}$  from  $u_1$  and  $a^{(p_{t-1}, k_{t-1})}$ .
9:   Use  $\sigma_e^2(t-1)$  for excitation sequence variance
10:  Calculate Acceptance ratio,  $\alpha_k$ 
11:  if  $\eta \leq \alpha_k$  where  $\eta = \mathcal{U}(0, 1)$  then
12:     $k_t = k^*$ 
13:     $a^{(p_{t-1}, k_t)} = a^{(p_{t-1}, k^*)}$ 
14:  else
15:     $k_t = k_{t-1}$ 
16:     $a^{(p_{t-1}, k_t)} = a^{(p_{t-1}, k_{t-1})}$ 
17:  RJMCMC for  $p$  Use  $k_t$  for  $k$ 
18:  Repeat same procedure in lines between 5 and 16
    for nonlinearity degree  $p$ .
19:
20:  Sample  $\sigma_e^2(t)$  from eq. (20) via Gibbs Sampling
21: end

```

$p = 2$ to $p = 3$, this move requires $\lambda = 9 - 5 = 4$ candidate coefficients to be proposed from a proposal distribution to satisfy the dimension matching criteria. Each element, u_i , of the proposed vector \mathbf{u}_1 , where $i = 1, \dots, \lambda$ is proposed from uniform distribution $\mathcal{U}(-\delta, \delta)$ and the joint distribution $q_1(\mathbf{u}_1)$ is defined as;

$$q_1(\mathbf{u}_1) = \prod_{i=1}^{\lambda} \mathcal{U}(-\delta, \delta), \quad \text{and} \quad \delta = \frac{0.1}{E[|\mathbf{x}|]}, \quad (21)$$

where $E[|\mathbf{x}|]$ is the expected value of the absolute value of the data vector \mathbf{x} .

Proposal distribution for candidate coefficients is chosen to make the candidates independent from recent coefficients for PAR model selection problem. This selection makes the Jacobian determinant equal to unity.

3. SIMULATION & RESULTS

In all simulations, initial values for excitation variance prior distribution are selected as $\alpha_e = 1$ and $\beta_e = 2$. The initial values for AR length k_0 and nonlinearity degree p_0 are set to 1. Maximum values for these parameters are set to 8 and 7, respectively. $\mathbf{a}^{(p_0, k_0)}$ is sampled from the prior distribution which is defined in (9).

Move probabilities are selected as 0.15, 0.15 and 0.7 respectively for $P_{\text{birth}}, P_{\text{death}}$ and P_{life} . 100,000 iterations are simulated to let sampled parameters converge.

Table 1. Detected Percentage of True Models

	P ⁽¹⁾ AR(1)	P ⁽¹⁾ AR(2)	P ⁽¹⁾ AR(3)	P ⁽²⁾ AR(1)	P ⁽²⁾ AR(2)	P ⁽²⁾ AR(3)	P ⁽³⁾ AR(1)	P ⁽³⁾ AR(2)	P ⁽³⁾ AR(3)	P ⁽⁴⁾ AR(1)	P ⁽⁵⁾ AR(1)
RJMCMC	96%	100%	98%	82%	90%	88%	76%	76%	75%	60%	84%
AIC	36%	45%	57%	51%	77%	73%	66%	51%	43%	17%	17%
BIC	100%	100%	98%	100%	96%	66%	92%	4%	1%	2%	2%
# of Coef.	1	2	3	2	5	9	3	9	19	4	5

model are plotted, respectively. Results of first 5,000 iterations are discarded as burn-in period. P⁽⁵⁾AR(1) model coefficients, $\mathbf{a}^{(5,1)} = [0.5, 0.2, -0.08, 0.11, -0.07]$, are estimated via RJMCMC with Mean Square Error (MSE) of 2.3×10^{-3} when MSE of the Least Squares (LS) estimate is 3.1×10^{-3} .

In Figures 3 and 4, the instantaneous estimates and histograms of parameters p and k for model P⁽²⁾AR(3) are shown, respectively. Algorithm decides true order pair, $(p, k) = (2, 3)$, nearly 80% of the iterations. By using these facts, it can be clearly said that, by using RJMCMC, we can decide the nonlinearity degree of PAR model and the resulting parameters can be used in an application to find the best model.

Model detection performances of RJMCMC, AIC and BIC are shown in Table 1. The performance of RJMCMC is superior to AIC in all these cases. BIC performs better than RJMCMC for linear models. However, this good performance of BIC is caused by the bias of penalizing the model complexity. The cost of this bias becomes evident in the detection probability of BIC for nonlinear models shown in the last four columns of Table 1 which cannot reach even 5%. This bias of BIC is clearly visible in the percentages of model decisions of methods RJMCMC, AIC and BIC plotted as a function of nonlinearity degree, p , in Figure 5 for models P⁽²⁾AR(2) and P⁽³⁾AR(2). Figure 5 and Table 1 show that RJMCMC performs nearly as good as BIC for models with small number of model coefficients and outperforms both AIC and BIC when the nonlinearity degree is increased accompanied by an increasing number of model coefficients.

4. CONCLUSIONS

RJMCMC algorithm provides an important success to the solution of model uncertainty of linear processes by using Bayesian methods. In addition to the studies which show this success in linear case, this study demonstrates the potential of the RJMCMC algorithm when the model-to-be identified is nonlinear.

The RJMCMC is shown to be superior to the common model selection methods, AIC and BIC, in estimating the nonlinearity degree and AR memory of a PAR model which is a nonlinear model based on Volterra series expansion.

Furthermore, we also show that RJMCMC algorithm can jump not only between spaces with different dimensions, but also between different classes of models. This potential of RJMCMC is demonstrated by sampling in linear and nonlin-

ear spaces of varying dimensions for the estimation of the parameters p and k .

Usage of RJMCMC in nonlinear case reduces the complexity of many nonlinear real life problems in areas such as seismology, control systems, wind energy and communications. As a future work, the proposed method will be applied to hourly average wind speed data to reveal the nonlinear character of wind speed and to make predictions in cases when the prediction time is long, e.g. over 12 hours.

REFERENCES

- [1] EE Kuruoğlu, "Nonlinear least l_p -norm filters for nonlinear autoregressive α -stable processes," *Digital Signal Processing*, vol. 12, no. 1, pp. 119–142, 2002.
- [2] JK Gruber, C Bordons, R Bars, and R Haber, "Nonlinear predictive control of smooth nonlinear systems based on Volterra models. Application to a pilot plant," *International Journal of Robust and Nonlinear Control*, vol. 20, no. 16, pp. 1817–1835, 2010.
- [3] D Lee, "Short-term prediction of wind farm output using the recurrent quadratic Volterra model," in *Power and Energy Society General Meeting, 2011 IEEE*. IEEE, 2011, pp. 1–8.
- [4] PJ Lahaye, JB Poline, G Flandin, S Dodel, and L Garnero, "Functional connectivity: studying nonlinear, delayed interactions between BOLD signals," *NeuroImage*, vol. 20, no. 2, pp. 962–974, 2003.
- [5] S Bekleric, "Nonlinear prediction via Volterra series and applications to geophysical data," in *Masters Abstracts International*, 2008, vol. 47.
- [6] CAR Fernandes, JCM Mota, G Favier, et al., "MIMO Volterra modeling for nonlinear communication channels," *Learning and Nonlinear Models*, vol. 2, no. 8, pp. 71–92, 2010.
- [7] PJ Green, "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, vol. 82, no. 4, pp. 711–732, 1995.
- [8] J Vermaak, C Andrieu, A Doucet, and SJ Godsill, "Reversible jump Markov chain Monte Carlo strategies for Bayesian model selection in autoregressive processes," *Journal of Time Series Analysis*, vol. 25, no. 6, pp. 785–809, 2004.
- [9] PT Troughton and SJ Godsill, "A reversible jump sampler for autoregressive time series," in *ICASSP 1998 IEEE*. IEEE, 1998, vol. 4, pp. 2257–2260.
- [10] J Besag, P Green, D Higdon, and K Mengersen, "Bayesian computation and stochastic systems," *Statistical Science*, pp. 3–41, 1995.