ESTIMATION OF CAR PROCESSES OBSERVED IN NOISE USING BAYESIAN INFERENCE

Panagiotis Giannopoulos and Simon J. Godsill*

Signal Processing and Communications Group, Department of Engineering
University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, England
pg202@eng.cam.ac.uk, sjg@eng.cam.ac.uk

ABSTRACT

We consider the problem of estimating continuous-time autoregressive (CAR) processes from discrete-time noisy observations. This can be done within a Bayesian framework using Markov chain Monte Carlo (MCMC) methods. Existing methods include the standard random walk Metropolis algorithm. On the other hand, least-squares (LS) algorithms exist where derivatives are approximated by differences and parameter estimation is done in a least-squares manner. In this paper, we incorporate the LS estimation into the MCMC framework to develop a new MCMC algorithm. This new algorithm is combined with the standard Metropolis algorithm and is found to improve performance compared to the standard MCMC algorithm. Simulation results are presented to support our findings.

1. INTRODUCTION

The discrete-time AR model is widely used in a broad range of signal processing applications such as audio modelling [4]. However, most physical systems or phenomena are continuous in time by nature. One example is speech and audio signals. Another example is in astrophysics and modelling sunspot data [9]. Such data can be modelled by continuous-time stochastic processes. CAR processes are a subclass of continuous-time processes and can be used in place of the discrete AR processes.

On the other hand, the advent of digital computers means that analysis of systems is done with discretized data obtained by sampling the underlying continuous-time processes. For example, audio is stored in a computer in digital form obtained through an analog to digital (ADC) conversion of the source signal. This has lead to the greater popularity of discrete models compared to continuous models. Here we model data using continuous-time autoregressions. A continuous-time model becomes necessary for the case of irregularly sampled data.

In this paper, we further develop methods for estimating continuous-time AR processes from discrete samples observed in noise using a Bayesian approach. In section 2, we give a brief review of existing methods and then present our new method in section 3. Simulation results are presented in section 4. Conclusions appear in section 5.

*Work supported by the Engineering and Physical Sciences Research Council and Jesus College Embiricos Foundation

2. THE CAR PROCESS AND PREVIOUS RESULTS

A p^{th} order CAR model can be written as [6]

$$x^{(p)}(t) + \alpha_{p-1}x^{(p-1)}(t) + \cdots + \alpha_0 x(t) = \epsilon(t) \qquad (1)$$

where $x^{(j)}$ denotes the j^{th} derivative of $x(t)$, and $\epsilon(t)$ is a continuous time "white noise" process with spectral density σ_ϵ^2.

For stationarity it is assumed that the roots of the characteristic equation

$$\sum_{j=0}^{p} \alpha_j s^j = 0 \qquad (2)$$

have negative real parts.

This model has a state space representation

$$\frac{d}{dt}\mathbf{x}(t) = A\mathbf{x}(t) + B\epsilon(t) \qquad (3)$$

where the state of the system is given by

$$\mathbf{x}(t) = \begin{bmatrix} x(t) \\ x^{(1)}(t) \\ x^{(2)}(t) \\ \vdots \\ x^{(p-1)}(t) \end{bmatrix} \qquad (4)$$

The matrices A and B are given in [6]. The solution of the system is given by

$$\mathbf{x}(t) = e^{At}\mathbf{x}(0) + \int_0^t e^{A(t-\tau)}B\epsilon(\tau)d\tau \qquad (5)$$

The process can be put in discrete-time state space form as follows

$$\mathbf{x}(t_n) = F_n\mathbf{x}(t_{n-1}) + \mathbf{e}(t_n) \qquad (6)$$

The state transition matrix is given by

$$F_n = e^{A(t_n - t_{n-1})} \qquad (7)$$

This is a matrix exponential and can be evaluated using the eigenvalues and eigenvectors of the matrix A [6]. The covariance matrix $\Sigma_{\mathbf{e}_n}$ of the prediction error $\mathbf{e}(t_n)$ is also given in [6].

Unfortunately, we cannot directly observe the state of the system. Instead, we observe a 'noisy' version of it given by

$$y(t_n) = C\mathbf{x}(t_n) + v(t_n) \tag{8}$$

where $y(t_n)$ are the discrete observations of the CAR process corrupted by white Gaussian noise with variance σ_v^2. Again, matrix C is given in [6]. Given observations at N arbitrary points

$$\mathbf{y} \triangleq \Big(y(t_1),\ y(t_2), \ldots,\ y(t_N)\Big) \tag{9}$$

which may be equally or unequally spaced, we are faced with the problem of estimating the CAR process parameters $\theta \triangleq (\mathbf{a}, \sigma_e, \sigma_v)$ and the true signal at these instants. Below we review existing methods that will form the basis of our new method.

2.1. ML estimation using the Kalman filter

Jones [6] evaluates the exact likelihood function by applying the Kalman filter [5] to equations (6) and (8). Let $\hat{\mathbf{x}}(t_{n-1})$ denote the optimal estimator of $\mathbf{x}(t_{n-1})$, based on the observations up to and including $y(t_{n-1})$. Let $P_{t_{n-1}}$ denote the covariance matrix of the estimation error.

Given $\hat{\mathbf{x}}(t_{n-1})$ and $P_{t_{n-1}}$, the optimal estimator of $\mathbf{x}(t_n)$ and the covariance matrix of the estimation error are given by the *prediction equations*:

$$\hat{\mathbf{x}}(t_{n|n-1}) = F_n \hat{\mathbf{x}}(t_{n-1}) \tag{10}$$

$$P_{t_{n|n-1}} = F_n P_{t_{n-1}} F_n' + \Sigma \mathbf{e}_n \tag{11}$$

Given the new observation $y(t_n)$, the *updating equations* are

$$\hat{\mathbf{x}}(t_n) = \hat{\mathbf{x}}(t_{n|n-1}) + K_{t_n}(y(t_n) - C\hat{\mathbf{x}}(t_{n|n-1})) \tag{12}$$

$$P_{t_n} = (I - K_{t_n} C) P_{t_{n|n-1}} \tag{13}$$

where

$$K_{t_n} = P_{t_{n|n-1}} C' (C P_{t_{n|n-1}} C' + \sigma_v^2)^{-1} \tag{14}$$

The Kalman filter output is an optimal state estimator given the data up to that time instant. The Kalman smoother [5] is an optimal state estimator given the whole data set. The likelihood is then evaluated via the prediction error decomposition [5]. Jones [6] then performs maximum-likelihood (ML) estimation of the parameters by searching for the maximum of the likelihood function. However, the likelihood is multimodal and this method will not always lead to the true maximum.

2.2. Least-Squares Parameter Estimation

Söderström et al [1] propose a LS approach for estimating the model parameters from discrete data. They approximate the derivatives with finite differences and the problem is transformed to a linear regression. For example, the k^{th} derivative is approximated by

$$x^{(k)}(t) \approx D^k x(t) \tag{15}$$

$$D^k x(t) \triangleq \frac{1}{h^k} \sum_j \beta_{k,j} x(t + jh) \tag{16}$$

where h is the sampling interval and $\beta_{k,j}$ are some weights.

The natural conditions on the weights $\beta_{k,j}$ to ensure that $D^k x(t) = x^{(k)}(t) + O(h)$ are given by [1]

$$\sum_j \beta_{k,j} j^n = \begin{cases} 0 & n = 0, \ldots, k-1 \\ k! & n = k \end{cases} \tag{17}$$

Standard derivative approximations such as the forward or backward delta operator (δ *or* δ_b) will give a biased estimate [1]. In [1] additional constraints are imposed on the weights to give an asymptotically unbiased solution. In [2] the more realistic case of noisy observations is addressed.

2.3. Existing Bayesian Methods

The estimation of the CAR parameters can be done by using a Bayesian framework [7]. The posterior probability distribution of the joint parameter set is given by Bayes rule:

$$p(\theta|\mathbf{y}) = \frac{p(\mathbf{y}|\theta)p(\theta)}{p(\mathbf{y})} \tag{18}$$

$$\propto p(\mathbf{y}|\theta)p(\theta) \tag{19}$$

Hence, the posterior is proportional to the likelihood times the prior. The likelihood is evaluated by the Kalman filter [5]. Different types of priors can be used for the prior density [8]. They will generally be much broader than the likelihood reflecting our prior ignorance of the parameters. The minimum mean-squared error (MMSE) estimate of the parameters is the mean of the posterior probability density function. The posterior density is not mathematically tractable, hence analytical methods cannot be employed. MCMC methods have to be used for simulating samples from the posterior density, for example the Metropolis-Hastings algorithm [3].

Suppose θ^i is the i^{th} sample of the MCMC chain. It is proposed that the next variate in the random sequence be θ^* which is given by

$$\theta^* \sim T(\theta^*|\theta^i) \tag{20}$$

where $T(\theta^*|\theta^i)$ is the transition probability of θ^* given θ^i, also known as the proposal density.

The probability of accepting θ^* instead of the current sample is given by the acceptance function:

$$A(\theta^i, \theta^*) = min(1, Q(\theta^i, \theta^*)) \tag{21}$$

where

$$Q(\theta^i, \theta^*) = \frac{p(\theta^*|\mathbf{y})T(\theta^i|\theta^*)}{p(\theta^i|\mathbf{y})T(\theta^*|\theta^i)} \tag{22}$$

Li [7] uses a random walk proposal density. The proposal density is a normal distribution centered at the current sample. Hence, it is symmetrical about the origin so the acceptance probability (22) is given by

$$Q(\theta^i, \theta^*) = \frac{p(\theta^*|\mathbf{y})}{p(\theta^i|\mathbf{y})} \tag{23}$$

However, it turns out that the random walk algorithm is a fairly basic one with slow convergence and dependent on the starting position. It works well for low model orders, but performance is compromised for a model order higher than 2.

3. EFFICIENT MCMC ALGORITHM

Our first new MCMC algorithm uses a proposal density for the model parameters which is a Gaussian centered at the LS estimate[1] $\mathbf{a}_{LS}^i$. Its covariance Σ_{LS}^i is also given by the LS estimate. The proposal density for the process and noise variances is still a random walk. Suppose θ^i is the i^{th} element of the chain. Hence, it is proposed that the next variate in the random sequence be θ^* which is given by

$$\theta^* \sim N(\mathbf{a}_{LS}^i, \Sigma_{LS}^i) N(\sigma_e^i, \beta_1^2) N(\sigma_v^i, \beta_2^2) \qquad (24)$$

where $N(\mu, \Sigma^2)$ is a Gaussian distribution with mean μ and covariance matrix Σ^2.

This proposal will start much nearer the true value of the parameters, hence it will have a much smaller burn in. However it will tend to have a lower acceptance ratio, with fewer samples being accepted. When samples are accepted, moves tend to be greater in magnitude. However, as the proposed samples are drawn around the LS estimate, any bias or inaccuracy in the LS estimate will be detrimental to the MCMC chain. Hence, this proposal will not normally be adequate on its own.

Our second new algorithm involves a proposal that combines the two proposal densities so as to have a few successive samples drawn from the random walk proposal and then use the LS proposal and so on. The random walk proposal will explore the vicinity of the current sample and the LS proposal will possibly take us to regions of the distribution that the random walk proposal might not have explored. Simulation results are presented in the next section. MCMC chains from the same dataset are produced by using the random walk and LS proposals on their own plus the combined proposal.

4. SIMULATION RESULTS

Data from a 4^{th} order CAR model were generated and observed in noise. The process variance was $\sigma_e^2 = 1$ and the noise variance was $\sigma_v^2 = 0.2^2$. The poles of the CAR process were located at $2\pi(-0.01 \pm 0.2j)$ and $2\pi(-0.01 \pm 0.4j)$. 500 samples of the process were observed with a sampling interval $h = 0.1$. For the case of the combined proposal, 10 samples were generated using the random walk proposal, then 1 sample using the LS proposal and so on. We used flat priors for the parameters.

Figures (1-3) show plots of the MCMC chains using the random walk, LS and combined proposal respectively. The estimated values of the poles and the estimated process and noise variances for each proposal are given in table 1. The combined proposal gives slightly better estimates compared to the other two. Figure 4 shows the noisy observations, true data and estimated data. The estimate of the signal is obtained by averaging out the Kalman smoother output over all iterations. Hence it is the MMSE estimate of the signal $E(\mathbf{x}|\mathbf{y})$. Figure 5 shows the autocorrelation function of the three MCMC chains. The combined proposal has also got better properties in terms of the correlation of successive samples. The autocorrelation function for the combined proposal drops faster than the other two.

[1] We used a modified LS algorithm were estimates of the derivatives up to order $p-1$ are taken from the Kalman smoother.

Proposal	Pole 1 ($/2\pi$)	Pole 2 ($/2\pi$)	σ_e	σ_v
RW	$-0.004 \pm 0.196j$	$-0.007 \pm 0.401j$	0.827	0.211
LS	$-0.003 \pm 0.196j$	$-0.03 \pm 0.3924j$	1.178	0.210
Comb.	$-0.005 \pm 0.197j$	$-0.009 \pm 0.401j$	0.895	0.209

Table 1: Comparative results using the 3 proposals

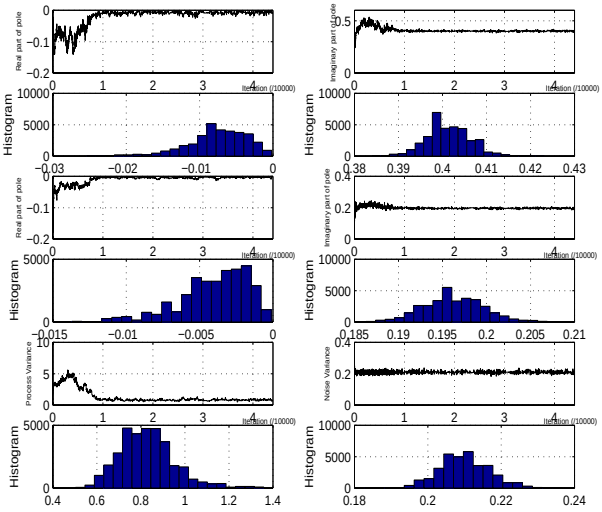

Figure 1: MCMC chain using the random walk proposal and estimates of the marginal posterior probabilities of the parameters.

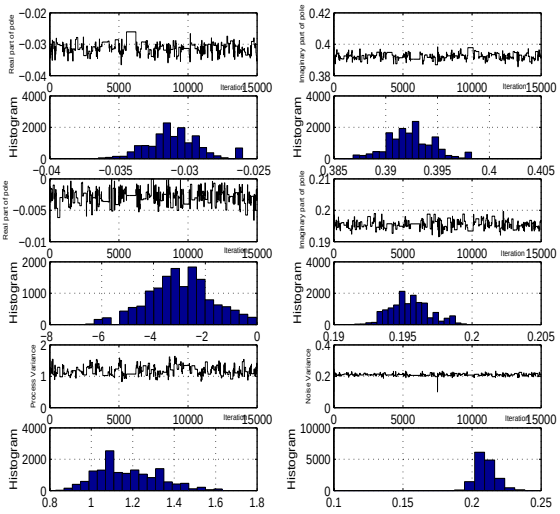

Figure 2: MCMC chain using the LS proposal and estimates of the marginal posterior probabilities of the parameters.

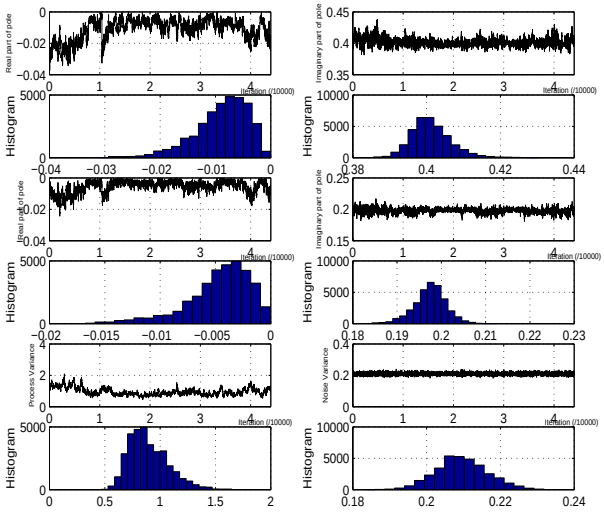

Figure 3: MCMC chain using the combined proposal and estimates of the marginal posterior probabilities of the parameters.

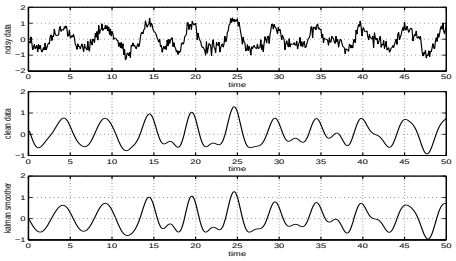

Figure 4: Noisy, clean and estimate of data using MMSE estimator $E(\mathbf{x}|\mathbf{y})$ for the combined proposal.

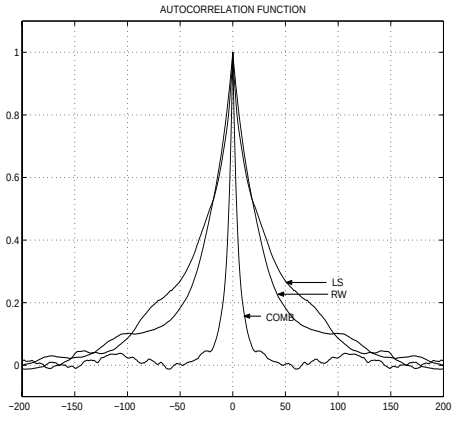

Figure 5: Autocorrelation function of the MCMC chain for the 3 proposals.

5. CONCLUSIONS

We have performed Bayesian inference of the CAR model parameters from discrete noisy observations using an MCMC method. Our new algorithm combined the basic random walk Metropolis algorithm with LS estimation to give a new efficient MCMC algorithm. Results are shown to be superior than those of the standard Metropolis algorithm or those of the Metropolis algorithm using a LS proposal. Future work can include a model order selection for the case of an unknown model order by using a reversible jump sampler [10, 11].

6. REFERENCES

[1] S. Fan, T. Söderström, M. Mossberg, B. Carlsson, and Y. Zou. Estimation of continuous-time AR process parameters from discrete-time data. *IEEE Transactions on Signal Processing*, 47(5):1232–1244, May 1999.

[2] S. Fan, T. Söderström, and Y. Zou. Continuous-time AR process parameter estimation in presence of additive white noise. *IEEE Transactions on Signal Processing*, 47(12):3392–3398, December 1999.

[3] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman & Hall, 1996.

[4] S. J. Godsill and P. J. W. Rayner. *Digital Audio Restoration*. Springer-Verlag, 1998.

[5] A. C. Harvey. *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press, 1989.

[6] R. H. Jones. Fitting a continuous time autoregression to discrete data. In *Applied Time Series Analysis*, volume 2, pages 117–152, 1980.

[7] F. S. Li. *Time deformation modeling: theory and application*. PhD thesis, Institute of Statistics and Desicion Sciences, Duke University, 1997.

[8] J. J. K. O' Ruanaidh and W. J. Fitzgerald. *Numerical Bayesian Methods Applied to Signal Processing*. Springer-Verlag, 1996.

[9] M. S. Phadke and S. M. Wu. Modeling of continuous stochastic processes from discrete observations with application to sunspots data. *Journal of American Statistical Association*, 69(346), 1974.

[10] P. T. Troughton and S. J. Godsill. A reversible jump sampler for autoregressive time series, employing full conditionals to achieve efficient model space moves. Technical Report CUED/F-INFENG/TR.304, Cambridge University Engineering Department, 1997.

[11] J. Vermaak, C. Andrieu, A. Doucet, and S. J. Godsill. Bayesian model selection of autoregressive processes. Technical Report CUED/F-INFENG/TR.360, Cambridge University Engineering Department, January 2000.