Signal Processing and Communications Laboratory
Department of Engineering, University of Cambridge

Technical Report
CUED/F-INFENG/TR. 304

A Reversible Jump Sampler for Autoregressive Time Series, Employing Full Conditionals to Achieve Efficient Model Space Moves

Paul T. Troughton* and Simon J. Godsill

November 1997

We use reversible jump Markov chain Monte Carlo (MCMC) methods (Green 1995) to address the problem of model order uncertainty in autoregressive (AR) time series within a Bayesian framework. Efficient model jumping is achieved by proposing model space moves from the full conditional density for the AR parameters, which is obtained analytically. This is compared with an alternative method, for which the moves are cheaper to compute, in which proposals are made only for the new parameters in each move. Results are presented for both synthetic and audio time series.

*Work supported by the Engineering and Physical Sciences Research Council
Keywords: Markov chain Monte Carlo, reversible jump, autoregressive time series, model selection, Bayesian model averaging

Correspondence to:
Paul Troughton
Tel: [+44] 1223 3 32767
email: ptt10@cam.ac.uk

Signal Processing and Communications Laboratory, Engineering Department,
Trumpington Street, Cambridge, CB2 1PZ, U.K.
Fax: [+44] 1223 3 32662
http://www-sigproc.eng.cam.ac.uk/~ptt10

1 Introduction

When fitting an autoregressive model to Gaussian time series data, often the correct order of the model is unknown. The model order cannot be estimated analytically by conventional Bayesian techniques when the excitation variance is unknown. We present MCMC methods for drawing samples from the joint posterior of all the unknowns, from which Monte Carlo estimates of the quantities of interest can be made, with the possibility of model mixing, if required, for tasks such as prediction, interpolation, smoothing or noise reduction.

Previous work on MCMC autoregressive model selection has parameterised the model using partial correlation coefficients (Barnett, Kohn & Sheather 1996, Barbieri & O'Hagan 1996) or pole positions[1] (Huerta & West 1997). These have a simple physical interpretation for certain types of signal, and allow stationarity to be enforced in a straightforward manner. We use the AR parameters, $\mathbf{a}$, directly. This allows us to use the full conditional density for $\mathbf{a}$, which is available analytically, to propose efficient reversible jump MCMC moves. Note that our method does not enforce stationarity of the model.

While Barbieri & O'Hagan (1996) also use reversible jump MCMC for AR model sampling, Barnett *et al.* (1996) and Huerta & West (1997) use stochastic search variable selection approaches, which avoid changing the dimension of the parameter vector by including all possible parameters at every iteration (George & McCulloch 1993).

2 Modelling framework

2.1 Autoregressive time series model

We model the signal $\{y_t\}$ as:

$$y_t = e_t + \sum_{i=1}^{k} a_i^{(k)} y_{t-i} \tag{1}$$

where

$$e_t \underset{\text{iid}}{\sim} \mathrm{N}\left(e_t \mid 0, \sigma_e^2\right) \tag{2}$$

is the excitation sequence and $\mathbf{a}^{(k)}$ is the AR parameter vector for a kth order model. This can be rewritten in matrix-vector form as:

$$\mathbf{e} = \mathbf{A}\mathbf{y} = \mathbf{y}_1 - \mathbf{Y}^{(k)}\mathbf{a}^{(k)} \tag{3}$$

where $\mathbf{y}_0$ and $\mathbf{y}_1$ are formed by partitioning $\mathbf{y}$ into, respectively, the first k values and the remainder, and $\mathbf{A}$ and $\mathbf{Y}^{(k)}$ take appropriate forms.

[1] Poles are the inverse of the roots of the characteristic equation.

Since the excitation sequence is Gaussian, the (approximate) likelihood takes the form (Box, Jenkins & Reinsel 1994, §A7.4):

$$p(\mathbf{y} \mid k, \mathbf{a}^{(k)}, \sigma_e^2) \approx p(\mathbf{y}_1 \mid \mathbf{y}_0, k, \mathbf{a}^{(k)}, \sigma_e^2) \tag{4}$$

$$= \mathbf{N}\big(\mathbf{e} \mid \mathbf{0}, \sigma_e^2 \mathbf{I}_{n_e}\big) \tag{5}$$

$$= (2\pi\sigma_e^2)^{-\frac{n_e}{2}} \exp\big(-\tfrac{1}{2\sigma_e^2}\mathbf{e}^T\mathbf{e}\big) \tag{6}$$

where n_e is the length of $\mathbf{e}$ and $\mathbf{y}_1$.

2.2 Prior distributions

We choose simple conjugate prior distributions for the continuous parameters and a uniform prior for the model order:

$$p(k) = \begin{cases} \frac{1}{k_{\max}} & k \in \{0, 1, \dots k_{\max}\} \\ 0 & \text{elsewhere} \end{cases} \tag{7}$$

$$p(\mathbf{a}^{(k)} \mid k) = \mathbf{N}\big(\mathbf{a}^{(k)} \mid \mathbf{0}, \sigma_a^2 \mathbf{I}_k\big) \tag{8}$$

$$p(\sigma_a^2) = \mathrm{IG}\big(\sigma_a^2 \mid \alpha_a, \beta_a\big) \tag{9}$$

$$p(\sigma_e^2) = \mathrm{IG}\big(\sigma_e^2 \mid \alpha_e, \beta_e\big) \tag{10}$$

where the inverted-gamma distribution (see *e.g.* Johnson & Kotz 1970) is defined for positive parameters α and β, and positive x, as:

$$\mathrm{IG}(x \mid \alpha, \beta) \propto x^{-(\alpha+1)} \exp(-\beta/x) \tag{11}$$

which tends to the Jeffreys' prior as $\alpha, \beta \to 0$. k, $\mathbf{a}^{(k)}$, σ_a^2 and σ_e^2 are assumed to be *a priori* independent.

2.3 Bayesian hierarchy

The posterior density for the parameters is then:

$$p(k, \mathbf{a}^{(k)}, \sigma_a^2, \sigma_e^2 \mid \mathbf{y}) \propto \underbrace{p(\mathbf{y} \mid k, \mathbf{a}^{(k)}, \sigma_e^2)}_{\text{Likelihood}} \underbrace{p(k)\, p(\mathbf{a}^{(k)} \mid \sigma_a^2)\, p(\sigma_a^2)\, p(\sigma_e^2)}_{\text{Priors}} \tag{12}$$

3 Reversible jump MCMC

Metropolis-Hastings algorithms (Hastings 1970) developed from molecular simulations as means to produce a Markov chain which converges to a required equilibrium distribution $p(\boldsymbol{\theta})$, without the need to sample directly from any inconvenient density.

Each step consists of:

- Choosing which subset of the parameters $\boldsymbol{\theta}$ to update
- Proposing new values for the subset of the parameters associated with that move by drawing from an arbitrary, convenient density:

$$\boldsymbol{\theta}'_u \sim q_u(\boldsymbol{\theta}'_u \mid \boldsymbol{\theta}_{-u}) \tag{13}$$

- Calculating the acceptance probability for this move, $A(\boldsymbol{\theta} \rightarrow \boldsymbol{\theta}')$, which is a function of the density from which the proposal values were drawn and the joint density to which the chain is required to converge:

$$A(\boldsymbol{\theta} \rightarrow \boldsymbol{\theta}') = \min\left(1, \frac{p(\boldsymbol{\theta}'_u \mid \boldsymbol{\theta}_{-u}) q_u(\boldsymbol{\theta}_u \mid \boldsymbol{\theta}_{-u})}{p(\boldsymbol{\theta}_u \mid \boldsymbol{\theta}_{-u}) q_u(\boldsymbol{\theta}'_u \mid \boldsymbol{\theta}_{-u})}\right) \tag{14}$$

- Either:
  - **Accepting** the move, setting the parameters to the proposed values, or
  - **Rejecting** the move, not changing any parameter values

Clearly, the more similar the densities $q_u(\boldsymbol{\theta}_u \mid \boldsymbol{\theta}_{-u})$ and $p(\boldsymbol{\theta}_u \mid \boldsymbol{\theta}_{-u})$, the higher the proportion of moves that will be accepted.

Reversible jump MCMC (Green 1995) is a generalisation which introduces moves between parameter spaces of different dimensionality, whilst retaining detailed balance (Hastings 1970), which is required for convergence, within each type of move.

If $J(k \rightarrow k')$ is the probability of proposing a move from a parameter space of dimension k to one of dimension k', and $\boldsymbol{\phi}$ contains those parameters which are present, and have the same meaning, in models of both dimensionalities, then the required acceptance probability is:

$$A\left((k, \boldsymbol{\theta}^{(k)}) \rightarrow (k', \boldsymbol{\theta}^{(k')})\right) = \min\left(1, \underbrace{\frac{p(k', \boldsymbol{\theta}^{(k')} \mid \boldsymbol{\phi})}{p(k, \boldsymbol{\theta}^{(k)} \mid \boldsymbol{\phi}))}}_{\text{Target density ratio}} \underbrace{\frac{J(k' \rightarrow k)}{J(k \rightarrow k')} \frac{q(\boldsymbol{\theta}^{(k)} \mid k', \boldsymbol{\theta}^{(k')}, \boldsymbol{\phi})}{q(\boldsymbol{\theta}^{(k')} \mid k, \boldsymbol{\theta}^{(k)}, \boldsymbol{\phi})}}_{\text{Transition probability ratio}}\right) \tag{15}$$

Note that the Jacobian term which appears in Green's (1995) formulation does not appear here since we are proposing directly in the new parameter space.

4 Sampling strategy

The parameters to be sampled comprise $k, \mathbf{a}^{(k)}, \sigma_e^2$ and σ_a^2.

4.1 Model moves

Sampling k involves a change in dimensionality, so we use a reversible jump move. First, we choose to propose a move from order k to order k' by sampling k' from the distribution $J(k \rightarrow k')$, for which we use a discretised Laplacian density (see §4.3). We then sample a new parameter vector $\mathbf{a}^{(k')}$ from a proposal density:

$$\mathbf{a}^{(k')} \sim q(\mathbf{a}^{(k')} \mid k', \mathbf{a}^{(k)}, \mathbf{y}, \sigma_a^2, \sigma_e^2) \tag{16}$$

Note that σ_a^2 and σ_e^2 remain unchanged during model moves. A more sophisticated scheme might propose these too.

We then calculate the acceptance probability (eq. 15):

$$A\big((k,\mathbf{a}^{(k)}) \rightarrow (k',\mathbf{a}^{(k')})\big) = \min\left(1, \underbrace{\frac{p(k',\mathbf{a}^{(k')} \mid \mathbf{y},\sigma_a^2,\sigma_e^2)}{p(k,\mathbf{a}^{(k)} \mid \mathbf{y},\sigma_a^2,\sigma_e^2)}}_{\text{Ratio of posteriors}} \underbrace{\frac{J(k' \rightarrow k)}{J(k \rightarrow k')} \frac{q(\mathbf{a}^{(k)} \mid k,\mathbf{a}^{(k')},\mathbf{y},\sigma_a^2,\sigma_e^2)}{q(\mathbf{a}^{(k')} \mid k',\mathbf{a}^{(k)},\mathbf{y},\sigma_a^2,\sigma_e^2)}}_{\text{Ratio of transition probabilities}}\right) \quad (17)$$

4.1.1 Full parameter vector proposals

As the proposal density, we can use the full conditional for the complete parameter vector $\mathbf{a}^{(k')}$, which is available analytically (see appendix A):

$$\begin{aligned} \mathbf{a}^{(k')} \sim q(\mathbf{a}^{(k')} \mid k',\mathbf{a}^{(k)},\mathbf{y},\sigma_a^2,\sigma_e^2) = p(\mathbf{a}^{(k')} \mid k',\mathbf{y},\sigma_a^2,\sigma_e^2) \\ \propto \mathbf{N}\big(\mathbf{a}^{(k')} \mid \boldsymbol{\mu}_{\mathbf{a}^{(k')}}, \mathbf{C}_{\mathbf{a}^{(k')}}\big) \end{aligned} \quad (18)$$

where

$$\mathbf{C}^{-1}_{\mathbf{a}^{(k')}} = \sigma_e^{-2} {\mathbf{Y}^{(k')}}^T \mathbf{Y}^{(k')} + \sigma_a^{-2}\mathbf{I}_{k'} \quad (19)$$

$$\boldsymbol{\mu}_{\mathbf{a}^{(k')}} = \sigma_e^{-2}\mathbf{C}^T_{\mathbf{a}^{(k')}} {\mathbf{Y}^{(k')}}^T \mathbf{y}_{-[1\ldots k']} \quad (20)$$

Rather than drawing a value of $\mathbf{a}^{(k')}$, then simply substituting equation (18) and the likelihood and priors into equation (17), which could lead to numerical problems, we can use the 'Candidate's Identity' (Besag 1989):

$$\frac{p(k,\mathbf{a}^{(k)} \mid \mathbf{y},\sigma_a^2,\sigma_e^2)}{p(\mathbf{a}^{(k)} \mid k,\mathbf{y},\sigma_a^2,\sigma_e^2)} = p(k \mid \mathbf{y},\sigma_a^2,\sigma_e^2) \quad (21)$$

to simplify equation (17) in this case to:

$$A\big((k,\mathbf{a}^{(k)}) \rightarrow (k',\mathbf{a}^{(k')})\big) = \min\left(1, \frac{p(k' \mid \mathbf{y},\sigma_a^2,\sigma_e^2)}{p(k \mid \mathbf{y},\sigma_a^2,\sigma_e^2)} \frac{J(k' \rightarrow k)}{J(k \rightarrow k')}\right) \quad (22)$$

where $p(k \mid \mathbf{y},\sigma_a^2,\sigma_e^2)$ is the posterior model order probability with σ_a^2 and σ_e^2 fixed and known. Godsill (1997) gives some further discussion of this within a general model uncertainty setting.

As shown in appendix A, this can be derived, by marginalising $\mathbf{a}^{(k)}$ and $\mathbf{a}^{(k')}$, as:

$$A\big((k,\mathbf{a}^{(k)}) \rightarrow (k',\mathbf{a}^{(k')})\big) = \min\left(1, \frac{p(k')}{p(k)} \frac{\sigma_a^{-k'}}{\sigma_a^{-k}} \frac{\big|\mathbf{C}_{\mathbf{a}^{(k')}}\big|^{\frac{1}{2}}}{\big|\mathbf{C}_{\mathbf{a}^{(k)}}\big|^{\frac{1}{2}}} \frac{\exp\big(\frac{1}{2}\boldsymbol{\mu}^T_{\mathbf{a}^{(k')}}\mathbf{C}^{-1}_{\mathbf{a}^{(k')}}\boldsymbol{\mu}_{\mathbf{a}^{(k')}}\big)}{\exp\big(\frac{1}{2}\boldsymbol{\mu}^T_{\mathbf{a}^{(k)}}\mathbf{C}^{-1}_{\mathbf{a}^{(k)}}\boldsymbol{\mu}_{\mathbf{a}^{(k)}}\big)} \frac{J(k' \rightarrow k)}{J(k \rightarrow k')}\right) \quad (23)$$

where $\frac{p(k')}{p(k)}$ will cancel because of the uniform prior (eq. 7). This expression is independent of $\mathbf{a}^{(k')}$, so the sampling operation of equation (18) need only be performed if the move is accepted. This point has also been noted in the context of a change-point model by Stark, Fitzgerald & Hladky (1997).

4.1.2 Partial parameter vector proposals

Alternatively, we can propose only the additional AR parameters introduced by the move. Again, we propose using the full conditional for the new parameters:

$$q(\mathbf{a}^{(k')} \mid k', \mathbf{a}^{(k)}, \mathbf{y}, \sigma_a^2, \sigma_e^2) = \begin{cases} \delta\big(\mathbf{a}^{(k')} - \mathbf{a}^{(k)}_{[1,\ldots k']}\big) & k' \leq k \quad (\text{'Death'}) \\ p\big(\mathbf{a}^{(k')}_{[k+1,\ldots k']} \mid k', \mathbf{y}, \sigma_a^2, \sigma_e^2, \mathbf{a}^{(k')}_{[1,\ldots k]} = \mathbf{a}^{(k)}\big) & \\ \qquad \cdot\, \delta\big(\mathbf{a}^{(k')}_{[1,\ldots k]} - \mathbf{a}^{(k)}\big) & k' > k \quad (\text{'Birth'}) \end{cases} \tag{24}$$

where $\boldsymbol{\theta}_{[1,\ldots k]}$ denotes the first k components of $\boldsymbol{\theta}$.

This proposal is quicker to compute than equation (18), but is likely to be more susceptible to convergence problems because lower order parameters remain fixed during moves between high order models. It is hence important also to make 'within-model' moves such as $\mathbf{a}^{(k)} \sim p(\mathbf{a}^{(k)} \mid k, \mathbf{y}, \sigma_a^2, \sigma_e^2)$ from time to time (see §B.1).

For partial parameter vector proposals, the acceptance probability takes different forms for 'birth' ($k' > k$) and 'death' ($k' < k$) moves:[2]

'Birth' move In this case, we are proposing

$$\mathbf{a}^{(k')} = \begin{bmatrix} \mathbf{a}^{(k)} \\ \mathbf{a}_u \end{bmatrix} \tag{25}$$

where $\mathbf{a}_u$ are the n new parameters drawn from the full conditional posterior density (Troughton & Godsill 1997):

$$\mathbf{a}_u \sim q(\mathbf{a}_u \mid k', \mathbf{a}^{(k)}, \mathbf{y}, \sigma_a^2, \sigma_e^2) \propto \mathbf{N}\big(\mathbf{a}_u \mid \boldsymbol{\mu}_{\mathbf{a}_u}, \mathbf{C}_{\mathbf{a}_u}\big) \tag{26}$$

where

$$\mathbf{C}_{\mathbf{a}_u}^{-1} = \sigma_e^{-2} {\mathbf{Y}_u^{(k')}}^T \mathbf{Y}_u^{(k')} + \sigma_a^{-2} \mathbf{I}_n \tag{27}$$

$$\boldsymbol{\mu}_{\mathbf{a}_u} = \sigma_e^{-2} \mathbf{C}_{\mathbf{a}_u} {\mathbf{Y}_u^{(k')}}^T \big(\mathbf{y}_1 - \mathbf{Y}_{-u}^{(k')} \mathbf{a}^{(k)}\big) \tag{28}$$

where the matrix $\mathbf{Y}$ is partitioned columnwise as $\mathbf{Y}_u$ and $\mathbf{Y}_{-u}$. Again, we can simplify equation (17) by marginalising $\mathbf{a}_u$:

$$A\big((k, \mathbf{a}^{(k)}) \to (k', [\,{}^{\mathbf{a}^{(k)}}_{\mathbf{a}_u}\,])\big) = \min\left(1, \frac{J(k' \to k)}{J(k \to k')} \frac{p(k' \mid \mathbf{a}^{(k')}_{[1,\ldots k]} = \mathbf{a}^{(k)}, \mathbf{y}, \sigma_a^2, \sigma_e^2)}{p(k \mid \mathbf{a}^{(k)}, \mathbf{y}, \sigma_a^2, \sigma_e^2)}\right) \tag{29}$$

$$= \min\left(1, \frac{J(k' \to k)}{J(k \to k')} \frac{p(k')}{p(k)} \frac{\int p(\mathbf{y} \mid k', [\,{}^{\mathbf{a}^{(k)}}_{\mathbf{a}_u}\,], \sigma_a^2, \sigma_e^2)\, p(\mathbf{a}_u \mid \sigma_a^2)\, d\mathbf{a}_u}{p(\mathbf{y} \mid k, \mathbf{a}^{(k)}, \sigma_a^2, \sigma_e^2)}\right) \tag{30}$$

$$= \min\left(1, \frac{J(k' \to k)}{J(k \to k')} \frac{p(k')}{p(k)} \sigma_a^{-n} |\mathbf{C}_{\mathbf{a}_u}|^{\frac{1}{2}} \exp\big(\tfrac{1}{2} \boldsymbol{\mu}_{\mathbf{a}_u}^T \mathbf{C}_{\mathbf{a}_u}^{-1} \boldsymbol{\mu}_{\mathbf{a}_u}\big)\right) \tag{31}$$

[2] 'Life' moves ($k' = k$) need not necessarily be proposed at all.

'Death' move Here, no new parameters are being proposed; we merely truncate $\mathbf{a}^{(k)}$ at the k'th parameter. By definition, if

$$A\big((k,\mathbf{a}^{(k)}) \to (k',\mathbf{a}^{(k')})\big) = \min\big(1, w(k,\mathbf{a}^{(k)},k',\mathbf{a}^{(k')},\sigma_a^2,\sigma_e^2)\big) \tag{32}$$

then

$$A\big((k',\mathbf{a}^{(k')}) \to (k,\mathbf{a}^{(k)})\big) = \min\Big(1, \frac{1}{w(k,\mathbf{a}^{(k)},k',\mathbf{a}^{(k')},\sigma_a^2,\sigma_e^2)}\Big) \tag{33}$$

so the calculations are similar to those for the acceptance probability for the corresponding 'birth' move.

4.2 'Null' moves

Sampling from $\mathbf{a}^{(k)}$, σ_e^2 or σ_a^2 with fixed model order k does not involve any change of dimensionality, so the treatment is straightforward, using a standard Gibbs sampler move. See appendix B for details.

4.3 Move selection

We have described in previous sections the following types of move:

- Model moves
  - 'Birth' moves
  - 'Death' moves
- 'Null' moves
  - $\mathbf{a}^{(k)}$ move
  - σ_e^2 move
  - σ_a^2 move

For simplicity, a partially systematic scan was chosen: each proposed model move is followed by a σ_e^2 move and a σ_a^2 move, but sampling $\mathbf{a}^{(k)}$ is a relatively expensive move, so it is performed randomly, less frequently.

The choice of model move is determined by the function $J(k \to k')$. To ensure good convergence, we want most proposed jumps to be small, but occasional large ones to occur too. We have chosen a discretised Laplacian density:

$$J(k \to k') \propto \exp(-\lambda\big|k'-k\big|) \qquad k \in \{0,\dots k_{\max}\} \tag{34}$$

5 Results

The sampler was implemented as described, comparing both full and partial parameter vector proposals.

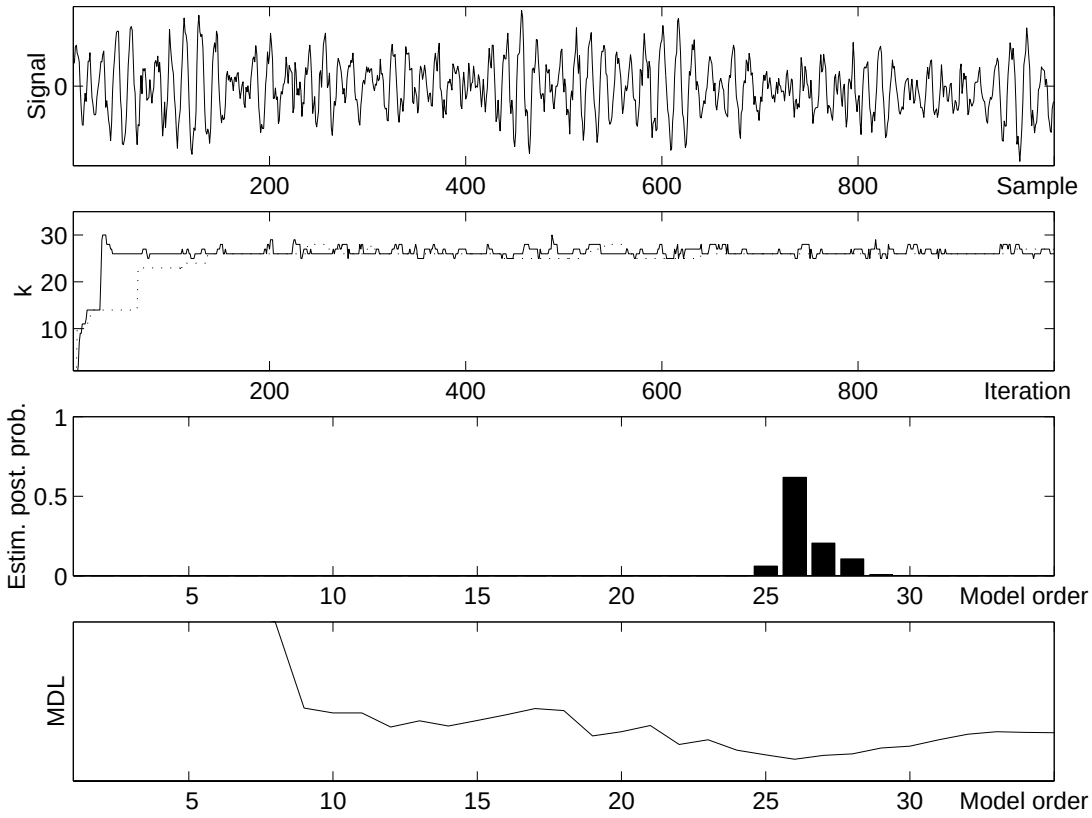

Figure 1: *Orchestral recording* (from top)*: Signal; Sampled model order values for full proposals* (solid) *and partial proposals* (dotted)*; Monte Carlo estimate of* $p(k \mid \mathbf{y})$*; MDL values.*

5.1 Audio data

The samplers were used to fit an AR model to a block of 1000 samples from a 44.1kHz sampled orchestral music recording. Figure 1 shows the signal, along with the results of running the both samplers.

Although both versions converge to roughly the same model order distribution, it can be seen that the sampler using partial proposals appears to generate a less well mixed chain than the full proposal sampler.

The Monte Carlo estimate of the marginal posterior density $p(k \mid \mathbf{y})$ was obtained by calculating the histogram of k having discarded the values from the first 50 iterations as burn-in from the full parameter proposal sampler's output. Such a short burn-in period is deemed satisfactory from observation of many MCMC runs on the same data. Clearly, the maximum *a posteriori* estimate of the model order is 26, which agrees with the global minimum of the Minimum Description Length (MDL) criterion, which is plotted for comparison.

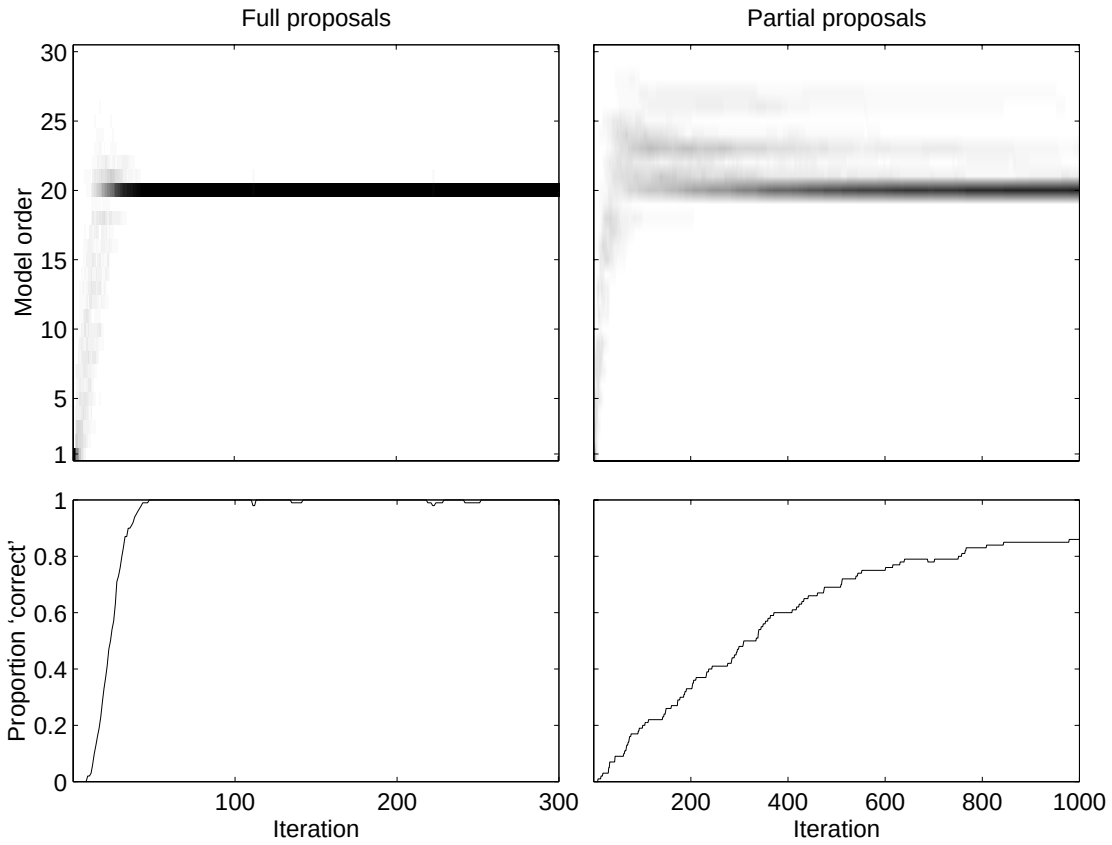

Figure 2: *Convergence behaviour from* 100 *runs with synthetic AR(20) data:* (left) *Full proposals;* (right) *Partial proposals;* (top) *Evolution of the model order histogram – darkness represents frequency;* (bottom) *Frequency of choosing* $k = 20$. *Note differing* x *scales.*

5.2 Synthetic data

To demonstrate the different convergence behaviour, 3500 samples were synthesised from an AR(20) process (see appendix C), and an ensemble of 100 runs were made with each sampler. The results are shown in figure 2. For each iteration, the top plots show the model order histogram, across the ensemble, and the bottom plots show the proportion of the ensemble which have the 'correct' value of model order.

It can be seen that all the runs of the full proposal sampler appear to converge within 50 iterations, whereas, even after 1000 iterations, some 15% of the partial proposal sampler runs have not converged.

6 Discussion

This reversible jump sampler provides a fast, straightforward way to cope with AR model order uncertainty in an MCMC framework. Using the raw AR parameters allows

the methods presented here to take advantage of the partially analytic structure of the AR model to speed convergence. The computation involved could probably be further reduced, for a given class of problems, by using a fully random scan and adjusting the move probabilities.

Proposing the full parameter vector in each move leads to a reliable sampler. Whilst proposing only part of the parameter vector makes acceptance probabilities faster to compute, the resulting Markov chain is more highly correlated.

Furthermore, whilst the two methods behave similarly for many modelling problems, in some cases, such as the AR(20) process of figure 2, the partial parameter vector proposals method is consistently very slow to converge. This is probably due to the transition of equation (25) being quite unnatural when considered in terms of, for example, pole positions; models of different order to the correct one, but with similar values for common parameters, may have low likelihood.

We do not enforce model stationarity, as this is difficult to incorporate into a prior on the AR parameters. Rejection sampling could be used, but the marginalisation in equation (36) would need to be over a parameter space containing only stationary models. This issue will be explored in future work.

A Simplification of $A\big((k, \mathbf{a}^{(k)}) \to (k', \mathbf{a}^{(k')})\big)$

We can express $p(k \mid \mathbf{y}, \sigma_a^2, \sigma_e^2)$ as:

$$p(k \mid \mathbf{y}, \sigma_a^2, \sigma_e^2) \propto p(k)\, p(\mathbf{y} \mid k, \sigma_a^2, \sigma_e^2) \tag{35}$$

$$= p(k) \int p(\mathbf{y}, \mathbf{a}^{(k)} \mid k, \sigma_a^2, \sigma_e^2)\, d\mathbf{a}^{(k)} \tag{36}$$

$$= p(k) \int p(\mathbf{y} \mid k, \mathbf{a}^{(k)}, \sigma_a^2, \sigma_e^2)\, p(\mathbf{a}^{(k)} \mid k, \sigma_a^2)\, d\mathbf{a}^{(k)} \tag{37}$$

but

$$p(\mathbf{y} \mid k, \mathbf{a}^{(k)}, \sigma_a^2, \sigma_e^2)\, p(\mathbf{a}^{(k)} \mid k, \sigma_a^2) \tag{38}$$

$$= \mathbf{N}\big(\mathbf{y}_1 - \mathbf{Y}^{(k)}\mathbf{a}^{(k)} \mid \mathbf{0}, \sigma_e^2 \mathbf{I}_{n_e}\big)\, \mathbf{N}\big(\mathbf{a}^{(k)} \mid \mathbf{0}, \sigma_a^2 \mathbf{I}_k\big) \tag{39}$$

$$= (2\pi\sigma_e^2)^{-\frac{n_e}{2}} (2\pi\sigma_a^2)^{-\frac{k}{2}} \exp\big(-\tfrac{1}{2}\big(\mathbf{a}^{(k)T}(\sigma_e^{-2}\mathbf{Y}^{(k)T}\mathbf{Y}^{(k)} + \sigma_a^{-2}\mathbf{I}_k)\mathbf{a}^{(k)} - 2\sigma_e^{-2}\mathbf{y}_1^T\mathbf{Y}^{(k)}\mathbf{a}^{(k)} + \sigma_e^{-2}\mathbf{y}_1^T\mathbf{y}_1\big)\big) \tag{40}$$

$$= (2\pi\sigma_e^2)^{-\frac{n_e}{2}} \sigma_a^{-k} \big|\mathbf{C}_{\mathbf{a}^{(k)}}\big|^{\frac{1}{2}} \exp\big(-\tfrac{1}{2}\big(\sigma_e^{-2}\mathbf{y}_1^T\mathbf{y}_1 - \boldsymbol{\mu}_{\mathbf{a}^{(k)}}^T \mathbf{C}_{\mathbf{a}^{(k)}}^{-1} \boldsymbol{\mu}_{\mathbf{a}^{(k)}}\big)\big) \mathbf{N}\big(\mathbf{a}^{(k)} \mid \boldsymbol{\mu}_{\mathbf{a}^{(k)}}, \mathbf{C}_{\mathbf{a}^{(k)}}\big) \tag{41}$$

so

$$p(k \mid \mathbf{y}, \sigma_a^2, \sigma_e^2) \propto p(k)(2\pi\sigma_e^2)^{-\frac{n_e}{2}} \sigma_a^{-k} \left|\mathbf{C}_{\mathbf{a}^{(k)}}\right|^{\frac{1}{2}} \exp\left(-\tfrac{1}{2}\left(\sigma_e^{-2}\mathbf{y}_1^T\mathbf{y}_1 - \boldsymbol{\mu}_{\mathbf{a}^{(k)}}^T \mathbf{C}_{\mathbf{a}^{(k)}}^{-1} \boldsymbol{\mu}_{\mathbf{a}^{(k)}}\right)\right) \cdot \int \mathbf{N}\left(\mathbf{a}^{(k)} \mid \boldsymbol{\mu}_{\mathbf{a}^{(k)}}, \mathbf{C}_{\mathbf{a}^{(k)}}\right) d\mathbf{a}^{(k)} \tag{42}$$

$$= p(k)(2\pi\sigma_e^2)^{-\frac{n_e}{2}} \sigma_a^{-k} \left|\mathbf{C}_{\mathbf{a}^{(k)}}\right|^{\frac{1}{2}} \exp\left(-\tfrac{1}{2}\left(\sigma_e^{-2}\mathbf{y}_1^T\mathbf{y}_1 - \boldsymbol{\mu}_{\mathbf{a}^{(k)}}^T \mathbf{C}_{\mathbf{a}^{(k)}}^{-1} \boldsymbol{\mu}_{\mathbf{a}^{(k)}}\right)\right) \tag{43}$$

because $\int \mathbf{N}(\boldsymbol{\theta} \mid \dots)\, d\boldsymbol{\theta} = 1$.

To eliminate dependence on the scale of the signal, the same vector $\mathbf{y}_1$, and hence length n_e, is used for both model orders being considered, *i.e.* all probabilities are conditional on the first $\max(k, k')$ values of $\mathbf{y}$.

B 'Null' moves

B.1 Sampling the AR parameter vector

We can sample $\mathbf{a}^{(k)}$ directly from its full conditional (eq. 18) in a Gibbs move, for which the acceptance probability is always 1.

B.2 Sampling the noise variance

We can also sample σ_e^2 using a Gibbs move. To do this, we require the full conditional posterior distribution:

$$p(\sigma_e^2 \mid \mathbf{y}, k, \mathbf{a}^{(k)}, \sigma_a^2) \propto \overbrace{p(\mathbf{y} \mid k, \mathbf{a}^{(k)}, \sigma_a^2, \sigma_e^2)}^{\text{Likelihood}} \cdot \overbrace{p(\sigma_e^2)}^{\text{Prior}} \tag{44}$$

$$\approx \mathbf{N}\left(\mathbf{e}^T\mathbf{e} \mid \mathbf{0}, \sigma_e^2\mathbf{I}_{n_e}\right) \cdot \mathrm{IG}\left(\sigma_e^2 \mid \alpha_e, \beta_e\right) \tag{45}$$

$$\propto (2\pi\sigma_e^2)^{-\frac{n_e}{2}} \exp\left(-\tfrac{1}{2\sigma_e^2}\mathbf{e}^T\mathbf{e}\right) \cdot \sigma_e^{-2(\alpha_e+1)} \exp\left(-\tfrac{\beta_e}{\sigma_e^2}\right) \tag{46}$$

$$= \sigma_e^{-(n_e+2\alpha_e+2)} \exp\left(-\tfrac{\beta_e + \frac{1}{2}\mathbf{e}^T\mathbf{e}}{\sigma_e^2}\right) \tag{47}$$

$$= \mathrm{IG}\left(\sigma_e^2 \mid \alpha_{se}, \beta_{se}\right) \tag{48}$$

where

$$\alpha_{se} = \alpha_e + \tfrac{1}{2}n_e \qquad \text{and} \qquad \beta_{se} = \beta_e + \tfrac{1}{2}\mathbf{e}^T\mathbf{e} \tag{49}$$

We can sample from this inverted-gamma density directly.

B.3 Sampling the parameter variance

Similarly, we can use a Gibbs move to sample the hyperparameter σ_a^2:

$$\begin{aligned} p(\sigma_a^2 \mid \mathbf{y}, k, \mathbf{a}^{(k)}, \sigma_e^2) &= p(\sigma_a^2 \mid \mathbf{a}^{(k)}) && (50)\\ &\propto p(\mathbf{a}^{(k)} \mid \sigma_a^2) \cdot p(\sigma_a^2) && (51)\\ &= \mathbf{N}\big(\mathbf{a}^{(k)} \mid \mathbf{0}, \sigma_a^2 \mathbf{I}_k\big) \cdot \mathrm{IG}\big(\sigma_a^2 \mid \alpha_a, \beta_a\big) && (52)\\ &= \mathrm{IG}\big(\sigma_a^2 \mid \alpha_{sa}, \beta_{sa}\big) && (53) \end{aligned}$$

where

$$\alpha_{sa} = \alpha_a + \tfrac{1}{2}k \qquad \text{and} \qquad \beta_{sa} = \beta_a + \tfrac{1}{2}\mathbf{a}^{(k)^T}\mathbf{a}^{(k)} \tag{54}$$

C Test AR process

The data used in §5.2 was generated from the AR process with parameters:

$$\mathbf{a} = \begin{bmatrix} -0.5078 \\ 4.5564 \\ 1.9504 \\ -11.2203 \\ -3.5378 \\ 19.1868 \\ 3.8193 \\ -24.8657 \\ -2.4029 \\ 25.0465 \\ 0.2678 \\ -19.7237 \\ 1.1703 \\ 12.0275 \\ -1.3091 \\ -5.5202 \\ 0.6804 \\ 1.7487 \\ -0.1543 \\ -0.2984 \end{bmatrix} \tag{55}$$

References

Barbieri, M. M. & O'Hagan, A. (1996). A reversible jump MCMC sampler for Bayesian analysis of ARMA time series. Tech. rep. Dipartimento di Statistica, Probabilità e Statistiche Applicate, Università "La Sapienza", Roma and Department of Mathematics, University of Nottingham.

BARNETT, G., KOHN, R. & SHEATHER, S. (1996). Bayesian estimation of an autoregressive model using Markov chain Monte Carlo. *Journal of Econometrics* **74**(2) 237–254.

BESAG, J. (1989). A candidate's formula — a curious result in Bayesian prediction. *Biometrika* **76**(1) 183.

BOX, G. E. P., JENKINS, G. M. & REINSEL, G. C. (1994). *Time Series Analysis: Forecasting and Control*. Holden-Day 3rd edn.

GEORGE, E. I. & MCCULLOCH, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association* **88**(423) 881–889.

GODSILL, S. J. (1997). Some new relationships between MCMC model uncertainty methods and implications for the development of new methods. Tech. Rep. CUED/F-INFENG/TR.305 Department of Engineering, University of Cambridge.
URL: http://www.sigproc.eng.cam.ac.uk/~sjg/papers/97/model.ps.gz

GREEN, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**(4) 711–732.

HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**(1) 97–109.

HUERTA, G. & WEST, M. (1997). Priors and component structures in autoregressive time series models. Tech. rep. Institute of Statistics & Decision Sciences, Duke University.

JOHNSON, N. L. & KOTZ, S. (1970). *Distributions in Statistics: Continuous Univariate Distributions*. Wiley.

STARK, J. A., FITZGERALD, W. J. & HLADKY, S. B. (1997). Multiple-order Markov chain Monte Carlo sampling methods with application to a changepoint model. Tech. Rep. CUED/F-INFENG/TR.302 Department of Engineering, University of Cambridge.

TROUGHTON, P. T. & GODSILL, S. J. (1997). Bayesian model selection for time series using Markov chain Monte Carlo. *Proceedings of IEEE ICASSP-97* **V** 3733–3736.