ROBUST MODELLING OF NOISY ARMA SIGNALS

Simon J. Godsill

Signal Processing and Communications Group
Cambridge University Engineering Department
Cambridge CB2 1PZ, England

ABSTRACT

In this paper methods are developed for enhancement and analysis of autoregressive moving average (ARMA) signals observed in additive noise which can be represented as mixtures of heavy-tailed non-Gaussian sources and a Gaussian background component. Such models find application in systems such as atmospheric communications channels or early sound recordings which are prone to intermittent impulse noise. Markov Chain Monte Carlo (MCMC) simulation techniques are applied to the joint problem of signal extraction, model parameter estimation and detection of impulses within a fully Bayesian framework. The algorithms require only simple linear iterations for all of the unknowns, including the MA parameters, which is in contrast with existing MCMC methods for analysis of noise-free ARMA models. The methods are illustrated using synthetic data and noise-degraded sound recordings.

1. INTRODUCTION

We present here a fully Bayesian approach to the analysis and extraction of ARMA signals observed in both impulsive noise and continuous background noise. Bayesian computations are performed using a Markov chain Monte Carlo (MCMC) simulation technique based around the Gibbs sampler [2, and references therein].

The framework presented is quite general, allowing for a full ARMA model with impulsive noise in both excitation sequence and observation noise. Special cases of the methods will be of use in less testing conditions; in particular noise reduction for pure AR or MA processes is easily achieved, as is noise reduction for simple Gaussian noise with no impulsive elements. The methods are likely to be of use in areas such as the enhancement of degraded sound recordings and baseband processing of analogue communications channels.

2. SIGNAL AND NOISE MODELS

It is assumed that some underlying signal x_t is observed in additive independent noise v_t:

$$y_t = x_t + v_t \tag{1}$$

2.1. Signal Model

We consider the Autoregressive Moving Average (ARMA) model for the underlying signal x_t:

$$x_t = \frac{\theta(L)}{\phi(L)} e_t \tag{2}$$

where L is the unit delay, $\theta(L) = 1 - \theta_1 L - \ldots - \theta_q L^q$ is the MA polynomial and $\phi(L) = 1 - \phi_1 L - \ldots - \phi_p L^p$ is the AR polynomial, with $q < p$. This ARMA(p, q) process is assumed to be minimum phase and stationary so that all poles and zeros lie strictly within the unit circle. We will also assume that no pole is equal to any zero so that pole-zero cancellation will not occur in $\frac{\theta(L)}{\phi(L)}$. These conditions ensure a unique ARMA representation [4, section 6.4].

In order to facilitate MCMC sampling we use a factorization of the ARMA model into a pure AR process u_t cascaded with an MA filter:

$$x_t = \theta(L) u_t, \; u_t = \frac{1}{\phi(L)} e_t \tag{3}$$

This factorization ensures Gaussian conditional likelihoods for both the AR and the MA coefficients when the ARMA process is observed in additive noise, as noted by [5].

The observation equation (1) and factorized ARMA representation (3) are readily combined into state-space form (not detailed here) which will allow the use of efficient Kalman Filter-based simulation methods in the estimation scheme.

2.2. Noise Models

e_t and v_t are i.i.d. processes. Non-Gaussian impulsive elements are included in the same way as in [1]:

$$v_t \sim \mathcal{N}(0, (1 - i_t)\sigma_v^2 + i_t g_t^2 \sigma_v^2) \tag{4}$$
$$e_t \sim \mathcal{N}(0, (1 - j_t)\sigma_e^2 + j_t h_t^2 \sigma_e^2) \tag{5}$$

where $\mathcal{N}(\mu, \sigma^2)$ denotes the univariate normal distribution. Here the i_t (j_t) are (0/1) indicator variables which switch a noise component between a continuous background component with variance σ_v^2 (σ_e^2) and a volatile impulsive component with variance $g_t^2 \sigma_v^2$ ($h_t^2 \sigma_e^2$). Note that g_t and h_t are time-varying unknown parameters to be determined from the data. Such a model has been found to be flexible and robust in the treatment of impulsive noise from various sources [6, 1]. This noise modelling framework is quite general, allowing for many types of non-Gaussian interference. Useful special cases which can be obtained from the general framework include:

1. i_t(or j_t) = 0, $\forall t$ - pure Gaussian noise
2. i_t(or j_t) = 1, $\forall t$ - heavy-tailed scale mixture of Gaussians

In the general case of (eqs. 4,5) we obtain a mixture of Gaussian background noise and a heavy-tailed scale mixture of Gaussians. The precise form of the scale mixture of Gaussians is determined by the prior distribution of the mixing constants g_t and h_t and can include, for example, the Student-t, α-stable or Generalized Gaussian distributions [7, and references therein].

For computational convenience we describe here only the case where the prior distribution for each of the unknown 'scale' parameters from $V = \{\sigma_e^2, \sigma_v^2, g_t^2, h_t^2, 1 \le t \le \mathrm{N}\}$ is of inverted-Gamma form, as in [1], corresponding to the Student-t form of scale mixtures. We will report on the use of other forms of mixing distribution, in particular those leading to the α-stable distribution, in future work. Also, as in [1], we assume a 1st order Markov dependence for the indicator variables i_t and j_t, which models the observed temporal clustering of impulses in many physical systems.

3. MCMC COMPUTATION

Computations are performed using a Markov chain Monte Carlo (MCMC) scheme which draws samples according to the posterior distribution for the unknown quantities. The primary aim of this work is to obtain an estimate of the underlying data x_t from its marginal posterior density $p(\mathbf{x}|\mathbf{y})$, where $\mathbf{x}$ and $\mathbf{y}$ are vectors made up from N consecutive elements of x_t and y_t, respectively. Of course, a useful by-product of such a general method as MCMC is the possibility of studying the posterior distributions of other system parameters at the same time, and these will be useful both as empirical convergence diagnostics and for applications where a full model analysis is the goal.

3.1. Gibbs Sampler

The scheme is implemented using a form of MCMC called the Gibbs sampler [2]. We first define some useful notation. The ARMA excitation sequence e_t and the signal x_t may be expressed in matrix-vector notation for a contiguous block of N signal values $\mathbf{x}$ using (3) as:

$$\mathbf{e} = A\mathbf{u} = \mathbf{u}_1 - U_\phi \boldsymbol{\phi} \quad (6)$$

$$\mathbf{x} = B\mathbf{u} = \mathbf{u}_1 - U_\theta \boldsymbol{\theta} \quad (7)$$

Here $\mathbf{e}$ denotes an N-vector of excitation samples, $\mathbf{u} = [\mathbf{u}_0^T, \mathbf{u}_1^T]^T$ contains the initial 'state' vector $\mathbf{u}_0 = [u_{-p+1}, \dots, u_0]^T$, and the N succeeding values $\mathbf{u}_1 = [u_1, \dots, u_\mathrm{N}]^T$. Also, define $\boldsymbol{\phi} = [\phi_1, \phi_2, \dots, \phi_p]^T$ and $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_q]^T$. Finally A, B, U_ϕ and U_θ are matrices constructed in such a way as to generate the e_t's and x_t's in the correct row order.

The Gibbs Sampler requires the full conditional posterior distribution for each unknown quantity. In our case the unknowns are $\mathbf{u}$, $\boldsymbol{\theta}$, $\boldsymbol{\phi}$, $\mathbf{i} = [i_1, \dots, i_\mathrm{N}]$, $\mathbf{j} = [j_1, \dots, j_\mathrm{N}]$ and $V = \{\sigma_e^2, \sigma_v^2, g_t^2, h_t^2, 1 \le t \le \mathrm{N}\}$. The method is iterative and involves sampling with replacement from each conditional distribution in turn. It is often advantageous for speed and reliability of convergence to group several unknowns together into a single sampling step, when this is convenient computationally, and we do this for $\mathbf{u}$, $\mathbf{i}$ and $\mathbf{j}$.

The $(i+1)$th iteration may be then be summarized as:

1. $V^{i+1} \sim p(V|\boldsymbol{\theta}^i, \boldsymbol{\phi}^i, \mathbf{u}^i, \mathbf{i}^i, \mathbf{j}^i, \mathbf{y})$
2. $\boldsymbol{\theta}^{i+1} \sim p(\boldsymbol{\theta}|\boldsymbol{\phi}^i, \mathbf{u}^i, \mathbf{i}^i, \mathbf{j}^i, V^{i+1}, \mathbf{y})$
3. $\boldsymbol{\phi}^{i+1} \sim p(\boldsymbol{\phi}|\boldsymbol{\theta}^{i+1}, \mathbf{u}^i, \mathbf{i}^i, \mathbf{j}^i, V^{i+1}, \mathbf{y})$
4. $\{\mathbf{u}, \mathbf{i}, \mathbf{j}\}^{i+1} \sim p(\mathbf{u}, \mathbf{i}, \mathbf{j}|\boldsymbol{\theta}^{i+1}, \boldsymbol{\phi}^{i+1}, V^{i+1}, \mathbf{y})$

After a 'burn-in' period the resulting Markov chain converges and subsequent samples can be thought of as (dependent) draws from the *joint* posterior distribution for all unknowns. Monte Carlo estimates can then be made for any desired *marginal* quantity simply by forming ergodic averages from the converged samples.

The required conditional distributions can be derived from the following likelihood expressions which are obtained directly from the modelling assumptions, coupled with the prior distributions on the unknown parameters:

$$p(\mathbf{y}|\mathbf{u}, \boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{i}, \mathbf{j}, V) = p(\mathbf{y}|\mathbf{u}, \boldsymbol{\theta}, \Sigma_v) = \mathcal{N}_\mathrm{N}(\mathbf{y} - \mathbf{x}|0, \Sigma_v)$$
$$p(\mathbf{u}_1|\mathbf{u}_0, \boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{i}, \mathbf{j}, V) = p(\mathbf{u}_1|\mathbf{u}_0, \boldsymbol{\phi}, \Sigma_e) = \mathcal{N}_\mathrm{N}(\mathbf{e}|0, \Sigma_e)$$
$$p(\mathbf{u}_0|\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{i}, \mathbf{j}, V) = p(\mathbf{u}_0|\boldsymbol{\phi}, \Sigma_e) = \mathcal{N}_p(m_0, \lambda_0^{-1})$$

where $\mathbf{e}$ and $\mathbf{x}$ are as given in (6) and (7) and $\Sigma_e = \mathrm{diag}((1-j_t)\sigma_e^2 + j_t\, h_t^2 \sigma_e^2,\ t = 1, \dots, \mathrm{N})$ and $\Sigma_v = \mathrm{diag}((1-i_t)\sigma_v^2 + i_t\, g_t^2 \sigma_v^2,\ t = 1, \dots, \mathrm{N})$ are the noise covariance matrices. m_0 and λ_0^{-1} are the mean and covariance matrix for the initial state vector. With no prior information about $\mathbf{u}_0$, $m_0 = 0$ and λ_0^{-1} is simply the covariance matrix for p elements of a stationary Gaussian AR process $(\boldsymbol{\phi}, \sigma_e^2)$ [4, section A.7]. Alternatively, $m_0 = 0$ and λ_0^{-1} can be used to encourage continuity with an earlier section of processed data when the methods are applied in segmental fashion to large amounts of data.

The full conditionals, assuming Gaussian priors $\mathcal{N}_q(m_\theta, \lambda_\theta^{-1})$ and $\mathcal{N}_p(m_\phi, \lambda_\phi^{-1})$ for $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$, respectively, are then obtained from straightforward manipulations of the above probability expressions. We do not detail their precise form owing to space limitations. However, the resulting expressions are very closely related to those obtained for the AR case in [1]. Full details for the ARMA case are given in [8].

There are several important points to note about the conditional sampling operations:

- **MA parameters $\boldsymbol{\theta}$.** The conditional distribution for the MA parameters, required in step 2. of the iteration is multivariate normal. This results from assuming non-zero observation noise and means that the MA estimation step involves only simple linear operations.
- **Signal u and indicators i and j.** As in [1], sampling step 4. is reduced to a simpler scheme which draws each triple $\{u_t, i_t, j_t\}$ conditional upon the rest. Also as in [1] we occasionally substitute a sampling step which draws each of **u**, **i** and **j** from their conditionals using fast Kalman Filter-based sampling (see e.g. [9] and others).
- **Stability and invertibility.** These are enforced upon the AR and MA parameters by rejection sampling, as in [3].
- **Near-cancellation of poles/zeros.** This will generally mean that the model order has been selected too high. We can avoid the numerical problems associated with this by incorporating a prior constraint which disallows models with poles very close to zeros. This is implemented by rejection sampling in a similar way to the previous point.

4. OTHER ISSUES

We note without stating the detail that it is straightforward to incorporate non-white noise sources and indirectly observed outputs with known transfer functions into the same framework, by inclusion of additional linear filtering operators in the model. Further modifications will include the development of informative priors which are suitable for use with real acoustical signals, time varying models and model order selection (including detection of source presence). All of these can be achieved naturally within a Bayesian numerical framework.

5. EXAMPLES

The methods are first applied to a synthetic ARMA(4,1) process with coefficients $\boldsymbol{\phi} = [2.8826, -3.8438, 2.8351, -0.9703]$ and $\boldsymbol{\theta} = [0.9]$. There are additive observational outliers ($i_t = 1$) of amplitude 1000 and 300 at sample numbers 200 and 600, respectively, and an innovational outlier ($j_t = 1$) of size 300 at sample number 400.

Figure 1 shows the noisy ARMA data and figure 2 gives the corresponding estimated MMSE reconstruction of **x**. (i.e. the arithmetic mean of the data samples following a 'burn-in' of 200 iterations) . Comparison with the true original data shows a noise reduction of about 8dB. Figure 3 shows normalized histograms of the outlier indicators, showing high probabilities at the correct outlier positions plus one 'false alarm' around sample number 710. Figure 4 gives some posterior analysis of parameter values following the burn-in period. Histograms are centered around the true model parameter values.

Figure 5 shows a 1100-sample extract from an early musical recording. Here we employ an ARMA(15,5) model for the data. The resulting estimated MMSE reconstruction is given in figure 6, in which both impulse and background noise have visibly been removed. The indicator histograms are given in figure 7. The impulsive noise has clearly been identified and there is little evidence for any innovational outliers in this section of data. The estimated ARMA power spectrum is given in figure 8, which indicates the sort of 'low-pass' spectrum one might expect from an early recording.

REFERENCES

[1] S. J. Godsill and P. J. W. Rayner. Robust noise reduction for speech and audio signals. In *Proc. ICASSP*, May 1996.

[2] A. E. Gelfand and A. F. M. Smith. Sampling-based approaches to calculating marginal densities. *J. Am. Statist. Assoc.*, 85:398–409, 1990.

[3] J. Marriott, N. Ravishanker, A. Gelfand, and J. Pai. Bayesian analysis of ARMA processes: complete sampling-based inference under exact likelihoods. In D.A. Berry, K.M. Chaloner, and J. Geweke, editors, *Bayesian analysis in statistics and econometrics*, pages 243–256. John Wiley and Sons, 1996.

[4] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel. *Time Series Analysis Forecasting and Control.* Prentice Hall, 3rd edition, 1994.

[5] B. R. Musicus and J. S. Lim. Maximum likelihood parameter estimation of noisy data. *Proc. ICASSP*, pages 224–227, April 1979.

[6] S. J. Godsill and P. J. W. Rayner. A Bayesian approach to the restoration of degraded audio signals. *IEEE Trans. on Speech and Audio Processing*, 3(4):267–278, July 1995.

[7] M. West. On scale mixtures of normal distributions. *Biometrika*, 74(3):99–102, 1987.

[8] S.J. Godsill. Bayesian enhancement of speech and audio signals in the presence of both impulsive and background noise. (Unpublished) [1], July 1996.

[9] S. Fruhwirth-Schnatter. Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, 15:183–202, 1994.

[1] http://www-com-serv.eng.cam.ac.uk/~sjg/papers/96/arma.ps

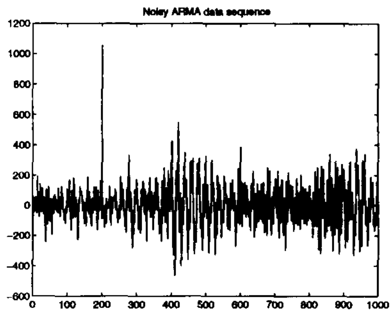

Figure 1. Noisy ARMA(4,1) data

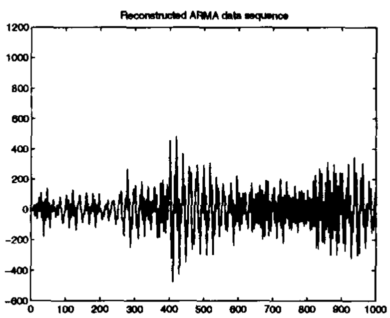

Figure 2. Reconstructed ARMA(4,1) data

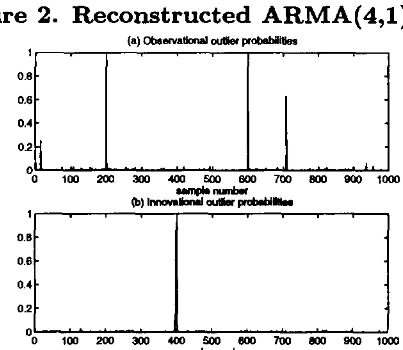

Figure 3. Outlier probabilities for ARMA(4,1) data

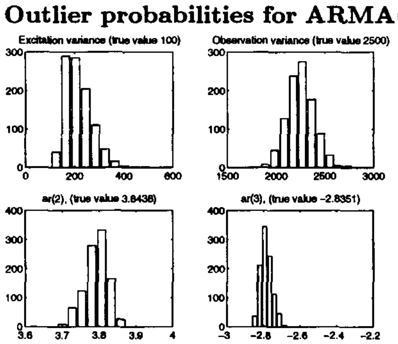

Figure 4. Parameter analysis for ARMA(4,1) data

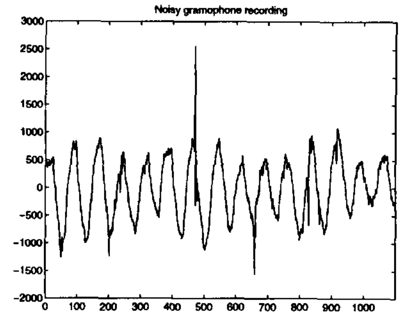

Figure 5. Noisy gramophone recording

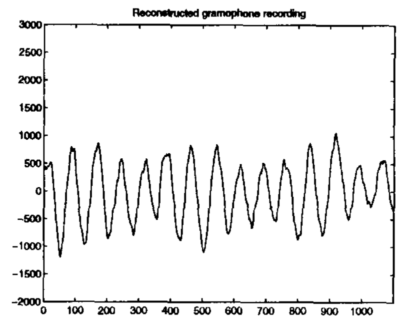

Figure 6. Reconstructed gramophone recording

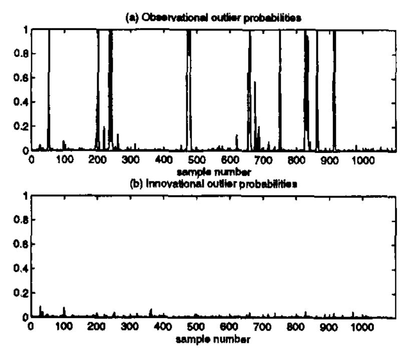

Figure 7. Outlier probabilities for gramophone recording

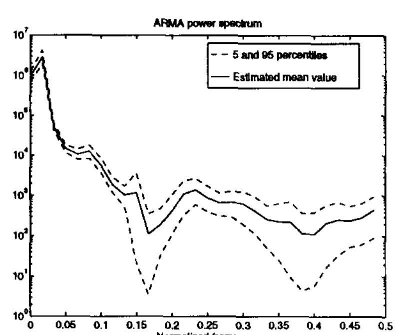

Figure 8. ARMA(15,5) power spectrum for gramophone data