RECURSIVE BAYESIAN LOCATION OF A DISCONTINUITY IN TIME SERIES

J.J. Ó Ruanaidh *W.J. Fitzgerald* *K.J. Pope*
Department of Engineering, University of Cambridge, U. K.

ABSTRACT

This paper presents a Bayesian approach to the location of a discontinuity in linearly modelled data. A matrix formulation is introduced which allows the modelling of changepoints in general linear models. Linear models investigated include abrupt changes in the mean of a Gaussian random variable, and piecewise polynomials such as splines, as well as autoregressive models. The approach facilitates the removal of nuisance parameters by integration. A general recursive technique for updating Bayesian posterior densities, which can result in large savings in computation, is also described.

1. The Detection of Changepoints using the General Piecewise Linear Model

A matrix formulation of the signal model will be employed that enables us to treat all linear-in-the-parameters models (such as polynomial fits to data as well as autoregressive models) in exactly the same way. The data is modelled as

$$y(i) = \begin{cases} \sum_{k=1}^{p} a_k g_k(i) + e(i) & \text{if } i < m \\ \sum_{k=1}^{p} b_k g_k(i) + e(i) & \text{otherwise} \end{cases} \tag{1}$$

where $g_k(i)$ is the value of a time dependent model function $g_k(t)$ evaluated at time t_i and where the noise samples e_i are drawn from a normal distribution. This equation may be expressed in the form $\mathbf{d} = \mathbf{Gb} + \mathbf{e}$ where $\mathbf{d}$ is an $N \times 1$ vector of data points and $\mathbf{G}$ is an $M \times N$ matrix, where each column is a basis function vector used to fit the data $\mathbf{d}$. Each element of the vector $\mathbf{b}$ is the linear coefficient for each basis vector in $\mathbf{G}$.

For example, a two pole AR model may be expressed in the form of equation 2. Note that the number of zero right half-rows is equal to the position of the changepoint, and thus the form of the matrix $\mathbf{G}$ depends explicitly on the position of the changepoint m.

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \\ y_{m+1} \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} y_0 & y_{-1} & 0 & 0 \\ y_1 & y_0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ y_{m-1} & y_{m-2} & 0 & 0 \\ 0 & 0 & y_m & y_{m-1} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & y_{N-1} & y_{N-2} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_{m-1} \\ e_m \\ \vdots \\ e_N \end{bmatrix} \tag{2}$$

The likelihood of the data is given by the joint probability of the noise samples.

$$p(\mathbf{d} \mid \{m\}\, \sigma\, \mathbf{b}\, \mathbf{M}) = \prod_{i=1}^{N} p(e_i) \tag{3}$$

where σ is the standard deviation of the Gaussian noise, $\mathbf{b} \in R^p$ denotes the linear parameters and $\mathbf{M}$ denotes prior information - or any assumptions which led to this particular choice of signal model.

$$p(\mathbf{d} \mid \{m\}\, \sigma\, \mathbf{b}\, \mathbf{M}) = \left(2\pi\sigma^2\right)^{-\frac{N}{2}} \exp\left(-\frac{\mathbf{e}^T\mathbf{e}}{2\sigma^2}\right) \tag{4}$$

where $\{m\}$ denotes the changepoints (non-linear parameters) in the matrix of basis functions $\mathbf{G}$. Substituting $\mathbf{e} = \mathbf{d} - \mathbf{Gb}$ into equation 4 gives,

$$p(\mathbf{d} \mid \{m\}\, \mathbf{b}\, \sigma\, \mathbf{M}) = \frac{\exp\left[-\frac{(\mathbf{d}-\mathbf{Gb})^T(\mathbf{d}-\mathbf{Gb})}{2\sigma^2}\right]}{\left(2\pi\sigma^2\right)^{\frac{N}{2}}} \tag{5}$$

Using a flat prior probability for the parameters a_k, δ_k and a Jeffreys' prior for the standard deviation of the Gaussian noise, Ó Ruanaidh and Fitzgerald [1] and Kheradmandnia [2] show how these parameters can be integrated out to give the following student-t distribution for the position of the changepoint m.

$$p(m \mid \mathbf{d}\, \mathbf{M}) \propto \frac{\left[\mathbf{d}^T\mathbf{d} - \mathbf{d}^T\mathbf{G}\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{d}\right]^{\frac{-N+p}{2}}}{\sqrt{\det\left(\mathbf{G}^T\mathbf{G}\right)}} \tag{6}$$

0-7803-1775-0/94 $3.00 © 1994 IEEE

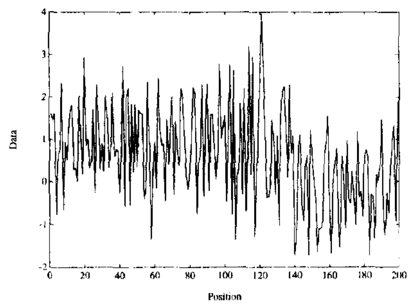

Figure 1. Discrete step in Gaussian noise

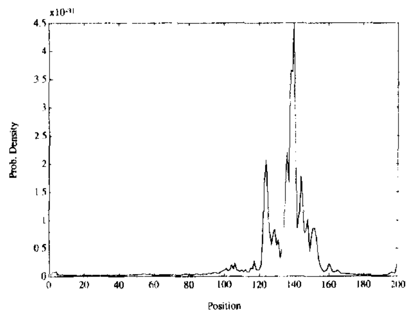

Figure 2. Bayesian step detection

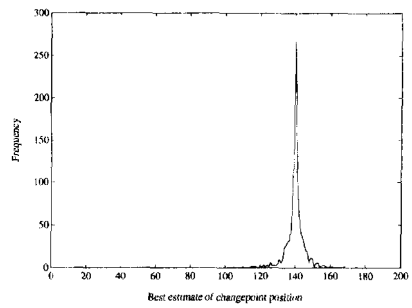

Figure 3. Performance of step detector

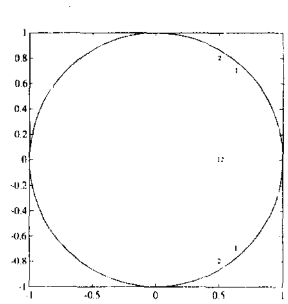

Figure 4. Pole plot

where p is the number of linear coefficients. This expression must be evaluated for candidate values of m, and the most probable selected.

The matrix based approach is very powerful. It is possible to integrate out one parameter at a time as shown by Smith [3] and Ó Ruanaidh and Fitzgerald [1] but this approach suffers the twin disadvantages that one must start from first principles for each new estimation problem and that the resultant mathematical expressions appear far more complicated and difficult to work with than expression 6 above. In contrast, equation 6 is directly applicable to a wide variety of different problems as it stands.

2. Recursive Bayesian Estimation

The approach described above requires N evaluations of expression 6 for the posterior probability. This is generally expensive to compute, especially when the time series is long or the number of linear parameters is large. A recursive implementation resulting in significant savings in computation is therefore highly desirable. A recursive procedure based on the Woodbury matrix inverse update formula [4] for obtaining the posterior probability of the changepoint being at position $m+1$, that is $p\ (m+1 \mid \mathbf{d}\ \mathbf{M})$, from the posterior probability $p\ (m \mid \mathbf{d}\ \mathbf{M})$ will now be defined.

In all, there are three cogent terms in equation 6 which may be updated recursively, namely $\mathbf{G}^T\mathbf{d}$, $\left(\mathbf{G}^T\mathbf{G}\right)^{-1}$ and det $\left(\mathbf{G}^T\mathbf{G}\right)$. The change in posterior probability can be traced back to a change of a *single* row of the matrix $\mathbf{G}$. Let $\mathbf{R}$ be the old m^{th} row and let $\mathbf{Q}$ be the new row that replaces it. For example, in the case of equation 2 above we find that $\mathbf{R} = [y_{m-1}\ y_{m-2}\ 0\ 0]$ and $\mathbf{Q} = [0\ 0\ y_{m-1}\ y_{m-2}]$. We wish to update the key matrix terms as follows,

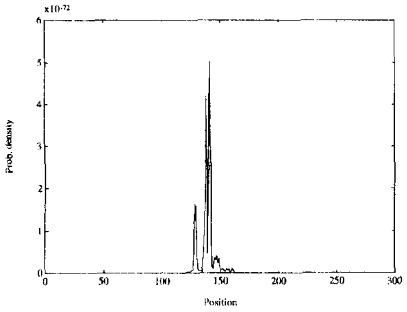

Figure 5. Bayesian changepoint detection in autoregressive data

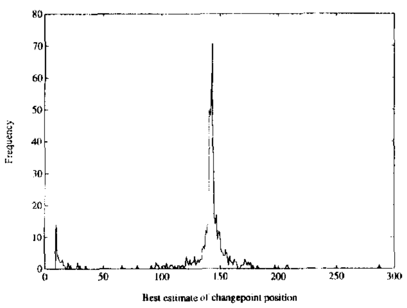

Figure 6. Three pole autoregressive data with a single changepoint

$$\left(\mathbf{G}^T\mathbf{G}\right)^{-1} \leftarrow \left(\mathbf{G}^T\mathbf{G} - \mathbf{R}^T\mathbf{R} + \mathbf{Q}^T\mathbf{Q}\right)^{-1} \quad (7)$$

$$\det\left(\mathbf{G}^T\mathbf{G}\right) \leftarrow \det\left(\mathbf{G}^T\mathbf{G} - \mathbf{R}^T\mathbf{R} + \mathbf{Q}^T\mathbf{Q}\right) \quad (8)$$

$$\mathbf{d}^T\mathbf{G} \leftarrow \mathbf{d}^T\mathbf{G} - d_m\mathbf{R} + d_m\mathbf{Q} \quad (9)$$

This can be carried out in two stages. The first stage is to replace the m^{th} row of $\mathbf{G}$ with a row of zeroes. The second stage is to replace this row of zeros with the row matrix $\mathbf{Q}$. The recursive update scheme is as follows. Define the matrices, $\mathbf{\Phi} = \left(\mathbf{G}^T\mathbf{G}\right)^{-1}$, $\chi = \mathbf{d}^T\mathbf{G}$, $D = \mathbf{d}^T\mathbf{d}$ and $\Delta = \det\left(\mathbf{G}^T\mathbf{G}\right)$.

The posterior density $p\,(m \mid \mathbf{d}\,\mathbf{M})$ is given by,

$$p\,(m \mid \mathbf{d}\,I) \propto \Delta^{-\frac{1}{2}}\left[D - \chi\,\Phi\,\chi^T\right]^{\frac{-N+p}{2}} \quad (10)$$

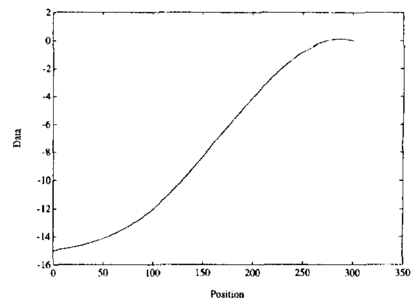

Figure 7. Noisy cubic spline data with a single changepoint

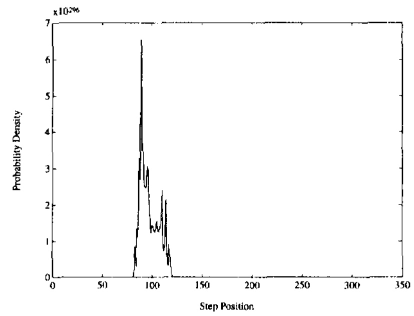

Figure 8. Bayesian changepoint detection in polynomial data

The following steps are performed twice, firstly with $\mathbf{S} = -R$ and secondly with $\mathbf{S} = Q$.

$$\chi_{i+1} = \chi_i + d_m\mathbf{S} \quad (11)$$

$$\mathbf{W} = \mathbf{\Phi}_i\mathbf{S}^T \quad (12)$$

$$\lambda = (\mathbf{1} + \mathbf{SW}) \quad (13)$$

$$\Delta_{i+1} = \lambda\,\Delta_i \quad (14)$$

$$\mathbf{\Phi}_{i+1} = \mathbf{\Phi}_i - \mathbf{W}\mathbf{W}^T/\lambda \quad (15)$$

Finally, the posterior density $p\,(m+1 \mid \mathbf{d}\,\mathbf{M})$ may be computed as,

$$p\,(m+1 \mid \mathbf{d}\,\mathbf{M}) \propto \Delta^{-\frac{1}{2}}\left[D - \chi\,\Phi\,\chi^T\right]^{\frac{-N+p}{2}} \quad (16)$$

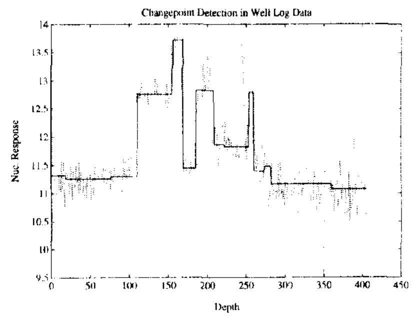

Figure 9. Well Log Data

3. Simulations

200 data points consisting of Gaussian noise of standard deviation $\sigma = 1$ added to a downward step from unity to zero at position $m = 140$ are plotted in figure 1. Figure 2 shows the posterior probability of the changepoint position. Examining the data by eye indicates a changepoint at $m \approx 120$. The Bayesian approach clearly determines the correct answer.

200 data points were generated by an $AR(3)$ system whose pole positions shifted instantaneously at $m = 140$ in figure 3. The pole positions are plotted in figure 4, where '1' and '2' signify the pole positions both before and after the changepoint respectively. Figure 5 shows the posterior probability of the changepoint position for a particular noise realisation for which the correct changepoint position is inferred. The question arises as to how well the method works in general. Monte Carlo simulations involving 1000 trials of this detector for different Gaussian noise realisations were carried out. The histogram of the changepoint positions chosen in each trial is shown in figure 6 which indicates that the detector performs well in most cases.

Two sections of a cubic spline fit, with added Gaussian noise and standard deviation $\sigma = 0.01$, comprise the 300 data points in figure 7. The changepoint which cannot be determined by eye is at position $m = 100$. The posterior probability of the changepoint position is shown in figure 8. Simply applying a smoothing filter to the data followed by numerical differentiation does not give good results. Even for this extremely taxing problem, the Bayesian detector performs well. The peak is biased, but the probability mass is centred about the "true" changepoint position.

Consider the data plotted in figure 9. This geophysical data contains measurements of nuclear magnetic response and conveys information about rock structure and in particular, the boundaries between different rock strata. Ó Ruanaidh and Fitzgerald [1] modelled this data as a Laplacian random process with thirteen abrupt changes in the mean. The results of a Bayesian analysis using the Gibbs' sampler and the Metropolis algorithm are shown with the data in figure 9.

4. Conclusion

Modelling changepoints by means of linear-in-the-parameters model based Bayesian analysis has been discussed. A recursive procedure, based on the Woodbury formula, has been derived whose application generally results in significant reductions in computation time. Simulations using polynomials and autoregressive models have given excellent results. Results for non Gaussian data with multiple changepoints has also been presented.

Further work to be carried out by the authors at present includes extending the above to click detection in audio signals, as part of the work being carried out in the Signal Processing Laboratory on techniques for the restoration of old gramophone records.

References

[1] J.J. Ó Ruanaidh and W.J. Fitzgerald. The identification of discontinuities in time series using the bayesian piecewise linear model. Technical report, Department of Engineering, University of Cambridge, 1993.

[2] M. Kheradmandnia. *Aspects of Bayesian Threshold Autoregressive Modelling.* PhD thesis, University of Kent, England, 1991.

[3] A.F.M. Smith. A bayesian approach to inference about a changepoint in a sequence of random variables. In *Biometrika*, pages 407–416, 1975.

[4] Press W.H. et al. *Numerical Recipes in C.* Cambridge University Press, 1992.