EXTENSION OF THE GENERAL LINEAR MODEL TO INCLUDE PRIOR PARAMETER INFORMATION

Mark C.M. Hsieh and Peter J.W. Rayner

Signal Processing and Communications Laboratory
Department of Engineering, University of Cambridge
Trumpington Street, Cambridge, CB2 1PZ
England
mcmh1@eng.cam.ac.uk

ABSTRACT

A set of approximations have been applied to allow the inclusion of Gaussian distributed priors for the linear parameters of the General Linear Model in order that the parameters may be integrated out alongside the Gaussian error noise variance, to give the model evidence and posterior distributions in analytic form. The extended model achieves greater accuracy in parameter estimation and evidence approximation when applied in a Bayesian inference framework, with no increase in computational load.

1. REVIEW

The General Linear Model [1] models observable data $\mathbf{d}$ as the linear $\mathbf{b}$ weighted combination of a set of basis functions $\mathbf{G}$ with additive noise $\mathbf{e}$ in the form

$$\mathbf{d} = \mathbf{G}\mathbf{b} + \mathbf{e} \tag{1}$$

If a Gaussian white noise model is assumed, the likelihood function is given by

$$p(\mathbf{d}|\mathbf{G}, \mathbf{b}, \sigma) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left[-\frac{\mathbf{e}^T\mathbf{e}}{2\sigma^2}\right] \tag{2}$$

where N is the number of data points, and σ^2 is the noise variance.

If we suppose a set of hypotheses $H_1..H_S$ based on a set of matrices $\mathbf{G}_1..\mathbf{G}_S$ to model the observed data, with the model structure supporting only the linear parameters $\mathbf{b}$, we may apply Bayes rule for model inference.

$$P(H_i|\mathbf{d}) = \frac{P(H_i)p(\mathbf{d}|H_i)}{p(\mathbf{d})} \tag{3}$$

where

$$p(\mathbf{d}|H_i) = \int_0^\infty \int_{\Re^M} p(\mathbf{d}|\mathbf{G}_i, \mathbf{b}, \sigma)p(\mathbf{b}, \sigma)\, d\mathbf{b}\, d\sigma \tag{4}$$

The joint prior $p(\mathbf{b}, \sigma)$ can take various forms which would allow analytical integration of Equation 4 or approximations to be made for analytical results.

1.1. Method 1

Using a uniform prior for the linear parameters and an inverse chi distribution for the noise standard deviation σ.

$$p(\mathbf{b}) = k \tag{5}$$

$$p(\sigma) \propto \sigma^{-P_0} \exp\left[-\frac{P_1}{2\sigma^2}\right] \tag{6}$$

Setting $P_0 = 1$, $P_1 = 0$ gives Jeffrey's non-informative scale prior [2] for the noise standard deviation.

The linear parameters $b_0..b_M$ are integrated out over all real M dimensional space $\Re^M$ by orthogonalisation, and σ is integrated out as a gamma integral.

$$p(\mathbf{d}|\mathbf{G}) \propto \frac{\left[P_1 + \mathbf{d}^T\mathbf{d} - \mathbf{d}^T\mathbf{G}(\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{d}\right]^{-(\frac{N-M+P_0-1}{2})}}{\sqrt{|\mathbf{G}^T\mathbf{G}|}} \tag{7}$$

2. EXTENSION

2.1. Method 2

Using a Gaussian distribution for $\mathbf{b}$ and inverse chi prior for σ

$$p(\mathbf{b}) = \frac{1}{(2\pi)^{\frac{M}{2}} |\mathbf{\Sigma_b}|^{\frac{1}{2}}} \exp\left[-\frac{(\mathbf{b}-\bar{\mathbf{b}})^T\mathbf{\Sigma_b}^{-1}(\mathbf{b}-\bar{\mathbf{b}})}{2}\right]$$

$$p(\sigma) \propto \sigma^{-P_0} \exp\left[-\frac{P_1}{2\sigma^2}\right] \tag{8}$$

The linear parameters are integrated out over all real M dimensional space $\Re^M$ by orthogonalisation

$$\begin{aligned} p(\mathbf{d}|\mathbf{G}, \sigma) &= \int_{\Re^M} p(\mathbf{d}|\mathbf{G}, \mathbf{b}, \sigma)p(\mathbf{b})\, d\mathbf{b} \\ &= \frac{(2\pi\sigma^2)^{-(\frac{N-M}{2})}}{\sqrt{|\mathbf{G}^T\mathbf{G} + \sigma^2\mathbf{\Sigma_b}^{-1}|}} (2\pi)^{-\frac{M}{2}} |\mathbf{\Sigma_b}|^{-\frac{1}{2}} \\ &\times \exp\left[-\frac{\mathbf{d}^T\mathbf{d} + \sigma^2\bar{\mathbf{b}}^T\mathbf{\Sigma_b}^{-1}\bar{\mathbf{b}} - \mathbf{b}_O^T\mathbf{\Phi}_O^{-1}\mathbf{b}_O}{2\sigma^2}\right] \end{aligned} \tag{9}$$

$$\mathbf{b}_O = \mathbf{G}^T\mathbf{d} + \sigma^2\mathbf{\Sigma_b}^{-1}\bar{\mathbf{b}}$$

$$\mathbf{\Phi}_O = \mathbf{G}^T\mathbf{G} + \sigma^2\mathbf{\Sigma_b}^{-1}$$

The form of $p(\mathbf{d}|\mathbf{G},\sigma)$ does not allow for the noise standard deviation σ to be integrated out analytically.

$$p(\mathbf{d}|\mathbf{G}) = \int_0^\infty p(\mathbf{d}|\mathbf{G},\sigma)p(\sigma)\,d\sigma \qquad (10)$$

MacKay [3] employs Gaussian approximations to the integrand of the evidence integral. Other approaches involve numerical methods or Monte Carlo models, none of which give analytic form for the model evidence $p(\mathbf{d}|\mathbf{G})$.

We derive an integrable form of the integrand in Equation 10 using the following approximations.

Consider σ small such that $\sigma^2 \left|\mathbf{\Sigma_b}^{-1}\right|^{\frac{1}{M}} \ll \left|\mathbf{G}^T\mathbf{G}\right|^{\frac{1}{M}}$

$$\begin{aligned} \mathbf{\Phi}_O^{-1} &= (\mathbf{G}^T\mathbf{G})^{-1}(\mathbf{I}+\sigma^2\mathbf{\Sigma_b}^{-1}(\mathbf{G}^T\mathbf{G})^{-1})^{-1} \\ &\approx (\mathbf{G}^T\mathbf{G})^{-1}(\mathbf{I}-\sigma^2\mathbf{\Sigma_b}^{-1}(\mathbf{G}^T\mathbf{G})^{-1}) \qquad (11) \end{aligned}$$

Thus inserting Equation 11 into Equation 9, expanding, and considering only significant terms in σ

$$\begin{aligned} p(\mathbf{d}|\mathbf{G},\sigma) &\approx \frac{(2\pi)^{-\frac{N}{2}}\sigma^{M-N}}{\sqrt{|\mathbf{G}^T\mathbf{G}|\,|\mathbf{\Sigma_b}|}}\exp\left[-\frac{Q_0}{\sigma^2}\right]\exp\left[Q_1\right] \qquad (12) \\ Q_0 &= \frac{1}{2}\mathbf{d}^T(\mathbf{d}-\mathbf{G}\hat{\mathbf{b}}_L) \\ Q_1 &= -\frac{1}{2}(\bar{\mathbf{b}}-\hat{\mathbf{b}}_L)^T\mathbf{\Sigma_b}^{-1}(\bar{\mathbf{b}}-\hat{\mathbf{b}}_L) \\ \hat{\mathbf{b}}_L &= (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{d} \end{aligned}$$

$\hat{\mathbf{b}}_L$ is in fact the maximum likelihood or least squares estimate for $\mathbf{b}$.

In the case σ large such that $\sigma^2 \left|\mathbf{\Sigma_b}^{-1}\right|^{\frac{1}{M}} \gg \left|\mathbf{G}^T\mathbf{G}\right|^{\frac{1}{M}}$

$$\begin{aligned} \mathbf{\Phi}_O^{-1} &= (\sigma^2\mathbf{\Sigma_b}^{-1})^{-1}(\mathbf{I}+\mathbf{G}^T\mathbf{G}(\sigma^2\mathbf{\Sigma_b}^{-1})^{-1})^{-1} \\ &\approx \sigma^{-2}\mathbf{\Sigma_b}(\mathbf{I}-\mathbf{G}^T\mathbf{G}\sigma^{-2}\mathbf{\Sigma_b}) \qquad (13) \end{aligned}$$

Thus inserting Equation 13 into Equation 9, expanding, and considering only significant terms in σ

$$\begin{aligned} p(\mathbf{d}|\mathbf{G},\sigma) &\approx (2\pi)^{-\frac{N}{2}}\sigma^{M-N-1}\exp\left[-\frac{Q_2}{\sigma^2}\right] \qquad (14) \\ Q_2 &= \frac{1}{2}(\mathbf{d}-\mathbf{G}\bar{\mathbf{b}})^T(\mathbf{d}-\mathbf{G}\bar{\mathbf{b}}) \end{aligned}$$

A pair of windowing functions is used to portion the noise prior $p(\sigma)$ as given in Equation 8

$$p_w(\sigma) \propto \begin{cases} p(\sigma)\left(1-\exp\left[-\frac{P_2}{\sigma^2}\right]\right) & (\sigma \le \sigma_x) \\ p(\sigma)\exp\left[-\frac{P_2}{\sigma^2}\right] & (\sigma > \sigma_x) \end{cases} \qquad (15)$$

$$P_2 = \log(2)\sigma_x^2\,, \quad \sigma_x^2 = \left[\frac{|\mathbf{G}^T\mathbf{G}|}{|\mathbf{\Sigma_b}^{-1}|}\right]^{\frac{1}{M}}$$

Combining the two portions of the noise prior in Equation 15 with their respective complements in $p(\mathbf{d}|\mathbf{G},\sigma)$ *i.e.* Equation 12 ($\sigma \le \sigma_x$), and Equation 14 ($\sigma > \sigma_x$)

$$\begin{aligned} p(\mathbf{d},\sigma|\mathbf{G}) &= p(\mathbf{d}|\sigma,\mathbf{G})p(\sigma) \\ &\approx \left\{\frac{(2\pi)^{-\frac{N}{2}}\sigma^{M-N}}{\sqrt{|\mathbf{G}^T\mathbf{G}|\,|\mathbf{\Sigma_b}|}}\exp\left[-\frac{Q_0}{\sigma^2}\right]\exp\left[Q_1\right]\right. \\ &\qquad \left.\times\,\sigma^{-P_0}\exp\left[-\frac{P_1}{2\sigma^2}\right]\left(1-\exp\left[-\frac{P_2}{\sigma^2}\right]\right)\right\} \\ &\quad + \left\{(2\pi)^{-\frac{N}{2}}\sigma^{M-N-1}\exp\left[-\frac{Q_2}{\sigma^2}\right]\right. \\ &\qquad \left.\times\,\sigma^{-P_0}\exp\left[-\frac{P_1}{2\sigma^2}\right]\exp\left[-\frac{P_2}{\sigma^2}\right]\right\} \qquad (16) \end{aligned}$$

Integrating over σ gives

$$\begin{aligned} &p(\mathbf{d}|\mathbf{G}) \approx \\ &\quad \frac{(2\pi)^{-\frac{N}{2}}}{2\sqrt{|\mathbf{G}^T\mathbf{G}|\,|\mathbf{\Sigma_b}|}}\Gamma\left(\frac{N_P-1}{2}\right)\exp\left[Q_1\right] \\ &\quad \times\left((Q_0+\frac{P_1}{2})^{-(\frac{N_P-1}{2})}-(Q_0+\frac{P_1}{2}+P_2)^{-(\frac{N_P-1}{2})}\right) \\ &\quad + \frac{1}{2}(2\pi)^{-\frac{N}{2}}\Gamma\left(\frac{N_P}{2}\right)(Q_2+\frac{P_1}{2}+P_2)^{-\frac{N_P}{2}} \qquad (17) \\ &N_P = N-M+P_0 \end{aligned}$$

2.2. Method 3

Using a Gaussian distribution for $\mathbf{b}$ and inverse chi prior for σ as Equation 8, the noise standard deviation σ is integrated out first

$$\begin{aligned} p(\mathbf{d}|\mathbf{G},\mathbf{b}) &= \int_0^\infty p(\mathbf{d}|\mathbf{G},\mathbf{b},\sigma)p(\sigma)\,d\sigma \\ &= \frac{\pi^{-\frac{N}{2}}(P_1+\mathbf{e}^T\mathbf{e})^{\frac{1-N-P_0}{2}}}{2^{\frac{3-P_0}{2}}}\Gamma\left(\frac{N+P_0-1}{2}\right) \qquad (18) \\ \mathbf{e}^T\mathbf{e} &= (\mathbf{d}-\mathbf{G}\mathbf{b})^T(\mathbf{d}-\mathbf{G}\mathbf{b}) \end{aligned}$$

This is in the form of a Student-t distribution for $\mathbf{b}$, which may be approximated by a Gaussian.

Consider the general form $[a+b(x-c)^2]^{-\frac{p}{2}}$, which has variance

$$Var\left([a+b(x-c)^2]^{-\frac{p}{2}}\right) = \frac{a}{b(p-3)} \qquad (p \ge 4) \qquad (19)$$

Equalising the variance and the height of the mode of the approximation to that of the Student-t distribution, the Gaussian approximation is

$$[a+b(x-c)^2]^{-\frac{p}{2}} \approx a^{-\frac{p}{2}}\exp\left[-\frac{p-3}{2a}b(x-c)^2\right] \qquad (20)$$

Extending this result to the multidimensional case

$$\begin{aligned} p_{approx}(\mathbf{d}|\mathbf{G},\mathbf{b}) &= \frac{\pi^{-\frac{N}{2}}a^{\frac{1-N-P_0}{2}}}{2^{\frac{3-P_0}{2}}}\Gamma\left(\frac{N+P_0-1}{2}\right) \\ &\quad \times\exp\left[\frac{N+P_0-4}{2a}\left(a-P_1-\mathbf{e}^T\mathbf{e}\right)\right] \qquad (21) \\ a &= P_1+(\mathbf{d}-\mathbf{G}\hat{\mathbf{b}})^T(\mathbf{d}-\mathbf{G}\hat{\mathbf{b}}) \\ \hat{\mathbf{b}} &= (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{d} \end{aligned}$$

The model evidence is now readily calculated as the integral of the product of two Gaussians.

$$p(\mathbf{d}|\mathbf{G}) = \int_{\Re^M} p_{approx}(\mathbf{d}|\mathbf{G},\mathbf{b})p(\mathbf{b})\,d\mathbf{b} \qquad (22)$$

The posterior distribution of $\mathbf{b}$ is given by

$$p(\mathbf{b}|\mathbf{G},\mathbf{d}) \propto p_{approx}(\mathbf{d}|\mathbf{G},\mathbf{b})p(\mathbf{b}) \quad (23)$$

which has the following characteristics

$$\begin{aligned} E[\mathbf{b}|\mathbf{d},\mathbf{G}] &= \left(\frac{N+P_0-4}{a}\mathbf{G}^T\mathbf{G}+\mathbf{\Sigma_b}^{-1}\right)^{-1} \\ &\quad \times \left(\frac{N+P_0-4}{a}\mathbf{G}^T\mathbf{d}+\mathbf{\Sigma_b}^{-1}\bar{\mathbf{b}}\right) \\ Cov[\mathbf{b}|\mathbf{d},\mathbf{G}] &= \frac{N+P_0-4}{a}\mathbf{G}^T\mathbf{G}+\mathbf{\Sigma_b}^{-1} \quad (24) \end{aligned}$$

3. APPLICATION

3.1. Channel Estimation in non-stationary noise

The General Linear Model can represent an FIR filter with the linear parameters as the filter coefficients

$$\begin{aligned} d_k &= b_0 g_k + b_1 g_{k-1} \ldots + b_{M-1} g_{k-M+1} + e_k \quad (25) \\ \mathbf{d} &= \mathbf{G_g b} + \mathbf{e} \\ \mathbf{G_g} &= \begin{bmatrix} g_0 & g_{-1} & g_{-2} & \cdots & g_{1-M} \\ g_1 & g_0 & g_{-1} & \cdots & g_{2-M} \\ g_2 & g_1 & g_0 & \cdots & g_{3-M} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ g_{N-1} & g_{N-2} & g_{N-3} & \cdots & g_{N-M} \end{bmatrix} \end{aligned}$$

If we have prior knowledge of g_k, and if the additive noise $\mathbf{e}$ is modelled well by an independent Gaussian distribution of standard deviation σ, we can use the General Linear Model to estimate the linear parameters $\mathbf{b}$ providing they are stationary.

In this test, the channel is carrying a binary data stream $s_0..s_{N_b-1}$. Each bit is encoded as $\mathbf{g}_0$ or $\mathbf{g}_1$ of length L. Thus the columns of $\mathbf{G_g}$ are composed of the successive vectors $\mathbf{g}_{s_0}..\mathbf{g}_{s_1}..\mathbf{g}_{s_2}..$ with their respective column to column displacements for correct FIR filter implementation.

The channel filter coefficients can be estimated by integrating and summing over the nuisance parameters using Method 1.

$$\begin{aligned} p(\mathbf{b}|\mathbf{d}) &= \sum_{\mathbf{s}} \int_0^\infty \frac{p(\mathbf{d}|\mathbf{b},\mathbf{s},\sigma)}{p(\mathbf{d})} p(\mathbf{b})p(\mathbf{s})p(\sigma)\, d\sigma \quad (26) \\ p(\mathbf{d}|\mathbf{b},\mathbf{s},\sigma) &= (2\pi\sigma^2)^{-\frac{LN_b}{2}} \exp\left[-\frac{\mathbf{e}^T\mathbf{e}}{2\sigma^2}\right] \\ \mathbf{e}^T\mathbf{e} &= (\mathbf{d}-\mathbf{G_{g_s}b})^T(\mathbf{d}-\mathbf{G_{g_s}b}) \end{aligned}$$

Clearly, as N_b increases, the summation over all combinations of $\mathbf{s}$ becomes unmanageable. The summation is therefore approximated by retaining the most probable N_{nodes} combinations of $\mathbf{s}$ as successive observations are received. This allows for efficient recursive updating of the evidence and posterior density as new data arrives, using the Woodbury formula [4] [1, page 105].

If we further propose that the channel is subject to non-stationary noise, we can modify the model to assume the noise is stationary within small segments of observed data. For convenience of analysis, we assume the segment boundaries coincide with the data bit $\mathbf{s}$ boundaries.

For a segment length of $2L$, and $M \le L$

$$\begin{aligned} p(\mathbf{b}|\mathbf{d}) &= \sum_{\mathbf{s}} \int_{\sigma=0}^{\infty} \frac{p(\mathbf{d}|\mathbf{b},\mathbf{s},\sigma)}{p(\mathbf{d})} p(\mathbf{b})p(\mathbf{s})p(\sigma)\, d\sigma \\ &= \frac{1}{p(\mathbf{d})} \sum_{\mathbf{s}} \int_{\sigma=0}^{\infty} p(d_0..d_{2L-1}|\mathbf{b}, s_{-1}, s_0, s_1, \sigma_0) \ldots \\ &\quad \times p(d_{2L}..d_{4L-1}|\mathbf{b}, s_1, s_2, s_3, \sigma_1) \ldots p(\mathbf{b})p(\mathbf{s})p(\sigma)\, d\sigma \\ &= \frac{1}{p(\mathbf{d})} \sum_{\mathbf{s}} p(d_0..d_{2L-1}|\mathbf{b}, s_{-1}, s_0, s_1) \ldots \\ &\quad \times p(d_{2L}..d_{4L-1}|\mathbf{b}, s_1, s_2, s_3) \ldots p(\mathbf{b})p(\mathbf{s}) \quad (27) \end{aligned}$$

Each of the terms $p(d_0..d_{2L-1}|\mathbf{b}, s_{-1}, s_0, s_1)$, $p(d_{2L}..d_{4L-1}|\mathbf{b}, s_1, s_2, s_3)\ldots$ is a Student-t distribution in $\mathbf{b}$ which is consistent with Method 3 for Gaussian approximation. The summation is again approximated by retaining the most probable N_{nodes} combinations of $\mathbf{s}$ as successive observations are received. The evidence for the combinations in $\mathbf{s}$ are calculated using Method 2. As each new observation arrives, the current set of N_{nodes} provide prior information for $\mathbf{b}$ given each combination of $\mathbf{s}$ they represent in the form of the product of Gaussian approximations, as calculated within the summation in Equation 27. New candidate nodes are compared and selected by integrating over the likelihoods and priors in $\mathbf{b}$ and σ_k using Method 2

$$\begin{aligned} & p(d_0..d_{2Lk-1}|s_{-1}..s_{2k-1}) = \quad (28) \\ & \quad \int_{\Re^M} \int_0^\infty p(d_{2L(k-1)}..d_{2Lk-1}|\mathbf{b}, \sigma_k, s_{2k-3}..s_{2k-1}) \ldots \\ & \quad \times p(d_0..d_{2L(k-1)-1}|\mathbf{b}, s_{-1}..s_{2k-3}) p(\sigma_k) p(\mathbf{b})\, d\sigma_k\, d\mathbf{b} \\ & p(d_0..d_{2L(k-1)-1}|\mathbf{b}, s_{-1}..s_{2k-3}) = \\ & \quad p(d_0..d_{2L-1}|\mathbf{b}, s_{-1}, s_0, s_1) p(d_{2L}..d_{4L-1}|\mathbf{b}, s_1, s_2, s_3) \ldots \end{aligned}$$

The observed data in this test is generated by filtering the sequence of vectors $\mathbf{g}_{s_0}$, $\mathbf{g}_{s_1}$, $\mathbf{g}_{s_2} \cdots$ and adding Gaussian white noise whose standard deviation is determined by a 2 state Markov process to model burst noise. The transition matrix is given by

$$\begin{aligned} \mathbf{T}_\sigma &= \begin{bmatrix} p_{11} & p_{21} \\ p_{12} & p_{22} \end{bmatrix} \quad (29) \\ p_{12} &= 1 - p_{11} \\ &= \frac{\textit{mean burst rate}}{(1 - \textit{mean burst rate}) \times \textit{mean burst length}} \\ p_{22} &= 1 - p_{21} \\ &= 1 - \frac{1}{\textit{mean burst length}} \end{aligned}$$

States are recomputed for every data sample $d_0..d_{LN_b-1}$.

3.2. Results

A random bitstream of length $N_b = 40$ was used to produce the original data composed of the stream of successive vectors $\mathbf{g}_{s_0}$, $\mathbf{g}_{s_1}$, $\mathbf{g}_{s_2} \cdots$ where both $\mathbf{g}_0$ and $\mathbf{g}_1$ are Gaussian random vectors of standard deviation 1 and length $L = 8$.

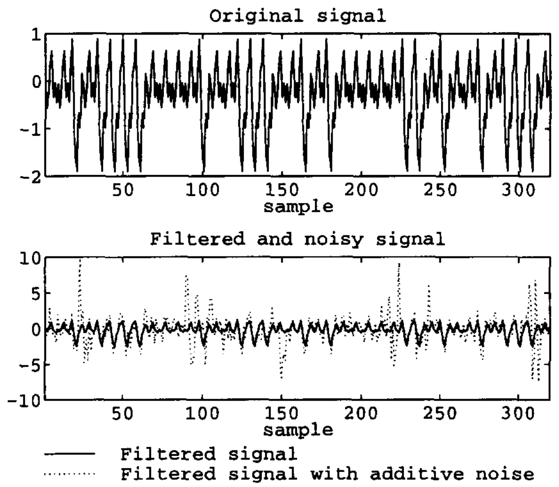

Figure 1: Examples of original signal, filtered signal and noisy filtered signal

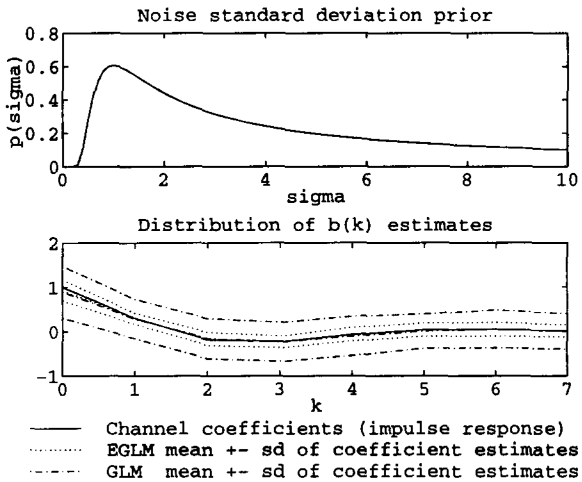

Figure 2: Inverse chi noise prior and Channel coefficient estimates based on full data

The $M = 8$ filter coefficients are based on a damped sinusoid as shown in Figure 2. A mean burst rate of 0.2 and mean burst length of 10 were used to model the non-stationary channel noise. State 1 has noise standard deviation $\sigma = 1$ and state 2 has $\sigma = 5$. Examples of the original signal, filtered signal and noisy filtered signal are shown in Figure 1.

The corresponding noise prior (Figure 2) for both the General Linear Model (Method 1) and Extended General Linear Model (Methods 2 and 3) has inverse chi distribution parameters $P_0 = 1$ and $P_1 = 1$. The linear parameter **b** priors are uniform.

A noise segment length of $2L$ was used to model the non-stationary noise, giving a total of $\frac{N_b}{2} = 20$ segment observations. Using $N_{nodes} = 16$, the mean and standard deviation of the estimates for the channel coefficients over 100 randomly generated input data streams are plotted in Figure 2. Figure 3 shows the progress of the mean and standard deviation of the estimates as the number of obser-

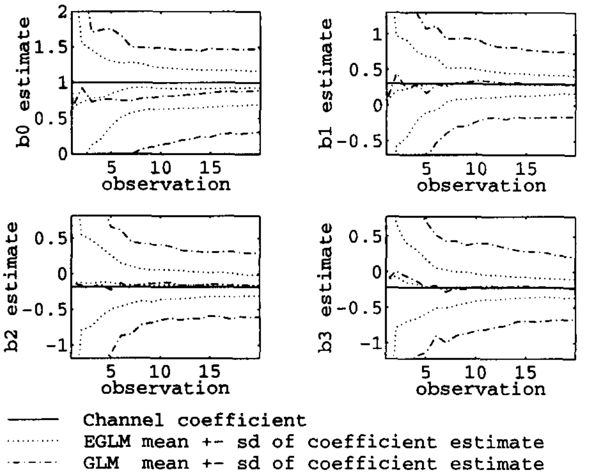

Figure 3: Channel coefficient estimates for $b_0..b_3$ based on increasing number of observations

vations increases.

The MAP estimates for **s** were compared against the original bitstreams as a measure of classification performance. Over the 100 random samples, the error rate was 16.4% for the combined EGLM (Methods 2 and 3), and 20.2% for the GLM (Method 1). Performing the same test using the EGLM Method 3 for both the posterior and evidence calculations resulted in very similar coefficient estimate characteristics, but an error rate of 20.1%. This is due to the thin tails of the Gaussian approximation to the Student-t underestimating the evidence.

Both the GLM and EGLM require $\mathcal{O}(M^2 L N_b N_{nodes})$ computational cycles to process a random sample of data.

4. CONCLUSION

Due to its ability approximately to model non-stationary noise, the Extended General Linear Model produces estimates for the channel coefficients with significantly narrower standard deviation and therefore greater accuracy. Classification is correspondingly improved when extension Methods 2 and 3 are combined. The improvement in performance is achieved with no increase in computational load.

5. REFERENCES

[1] J.Ó RUANAIDH AND W.FITZGERALD, *Numerical Bayesian Methods applied to Signal Processing*, Cambridge University Engineering Department, 1995.

[2] H.JEFFREYS, *Theory of Probability*, Oxford University Press, 1939.

[3] D.MACKAY, *Bayesian Methods for Adaptive Models*, PhD Thesis submitted to California Institute of Technology, 1991.

[4] W.PRESS, S.TEUKOLSKY, W.VETTERLING, AND B.FLANNERY, *Numerical Recipes in C*, Cambridge University Press, 2nd edition, 1992.

