CONTINUOUS ADAPTATION OF LINEAR MODELS WITH IMPULSIVE EXCITATION

S. W. Beet and L. Baghai-Ravary

Department of Electronic and Electrical Engineering,
University of Sheffield, Mappin Street, Sheffield, S1 3JD, UK.

ABSTRACT

This paper presents a new approach to continuously-adaptive system modelling, designed for the analysis of autoregressive (AR) systems excited by signals including an impulsive component. Voiced speech is well represented by such a model, and is used to demonstrate the advantages of the new approach. These include:

1. AR model parameter estimates are more stable in the region of pitch events.

2. A faster adaptation rate can be used, reducing the recovery time after plosives or other sudden changes in signal statistics.

The new method is based on multiple simultaneous estimates of each sample, using separate but related estimators. The general concept is illustrated here using a linear prediction (LP) approach to continuously-adaptive autoregressive (AR) modelling, based on the least mean square (LMS) algorithm.

1. INTRODUCTION

In many applications it is unsatisfactory to treat speech as a piecewise-stationary signal. This precludes the use of frame-based analyses to characterise the signal, and so some form of continuously-varying parameterisation is desired.

A common approach is to use a gradient descent algorithm to improve the accuracy of the parameterisation incrementally as each new sample becomes available [1, 2]. These methods generally assume that the the vocal apparatus is driven by a stationary, stochastic signal. This is quite accurate during those periods when the vocal folds are closed [3], but is quite unrealistic at the moments of opening and closure.

Consequently, the parameters yielded by traditional versions of this method exhibit undesirable perturbations, especially at the onset of each pitch pulse. At this point, the high value of the prediction error causes the estimated model parameters to change rapidly (even though the vocal tract transfer function is only changing slowly). To minimise these problems, the convergence rate of the adaptation algorithm is generally set much lower than could otherwise be possible [4].

The linear prediction approach to continuously adaptive system modelling is exemplified by the LMS algorithm. Firstly, a weighted sum of previous signal values is formed:

$$\hat{x}_n = \sum_{m=1}^{M} a_m x_{n-m} \tag{1}$$

where x_n is the n^{th} input value, a_m is the m^{th} prediction coefficient, and $\hat{x}_n$ is the predicted value of x_n. The error is then calculated:

$$e_n = x_n - \hat{x}_n = x_n - \sum_{m=1}^{M} a_m x_{n-m} \tag{2}$$

The sequence $e[n]$ then forms the residual signal, whose power is minimised by successive application of the LMS steepest descent update rule:

$$\begin{aligned} a_m &\leftarrow a_m - \mu \frac{\partial e_n^2}{\partial a_m} \\ &= a_m - 2\mu\, e_n \frac{\partial e_n}{\partial a_m} \\ &= a_m + 2\mu\, e_n x_{n-m} \end{aligned} \tag{3}$$

In these equations, μ deterimines the speed of convergence, and is chosen according to the properties of the data. Too small a value, and the system will only adapt very slowly; too large a value and the adaptation will not converge.

2. THE FORWARD-BACKWARD MINIMUM ERROR METHOD

The conventional LMS approach described in equations 2 and 3 can be badly affected by outliers in the sample distribution of the stochastic process assumed to be driving the AR system. If an impulse is present in the data, it will cause a large error, which will then be prolonged by the infinite-duration impulse response of the AR system under consideration. That error will give rise to a sudden change in the predictor coefficients, in an attempt to model it as part of the AR transfer function.

On the other hand, during such periods a backward estimator (i.e. a predictor operating backwards in time) can still produce an accurate estimate of the signal, provided the respective coefficients are known. For real, stationary signals, the coefficients for forward and backward estimation are identical, and a backward estimator could just as easily be used to estimate the predictor coefficients:

$$\begin{aligned} \hat{x}'_n &= \sum_{m=1}^{M} a_m x_{n+m} \\ e'_n &= x_n - \hat{x}'_n \\ a_m &\leftarrow a_m + 2\mu\, e'_n x_{n+m} \end{aligned} \tag{4}$$

where $\hat{x}'_n$ is the backward estimate of x_n. Obviously, this simple solution would have little advantage in a practical system, since the backward estimator would be disrupted by any impulses just *prior* to their occurrence, in the same way as a predictor is just *afterwards*.

However, one possible solution can be found by considering both a forward and a backward estimate *of the same sample*: the more accurate of the two estimates can be found, and the prediction coefficients updated using only the equations for the respective estimator. This can be termed a forward-backward minimum error (FBME) update scheme, but is different to the forward-backward adaptation employed in lattice filters, since here it is the same sample which is being estimated in both directions.

3. OPTIMISING μ

In many applications, the only practical method for optimising the speed of adaptation is by experiment. Figure 1 shows some typical data obtained when natural speech is processed using the conventional LMS coefficient update (equation 3), when using LMS updating based on *backward* prediction (equation 4), and when using the new forward-backward minimum error (FBME) approach. The average power of the forward residual, $e[n]$, relative to the input, $x[n]$, is plotted (in dB) as a function of the adaptation parameter, μ.

There are clear optima at μ = 7.5 for the conventional (forward) LMS case, μ = 8.5 for the backward LMS, and μ = 11 for FBME. Thus, despite the degradation of about 1.8dB which results from FBME, it is able to adapt nearly 50% faster than conventional LMS and 30% faster than the backward version.

Of interest is the observation that backward prediction appears to perform noticeably better than forward prediction when updating the predictor coefficients. It is also of note, that the performance of the FBME method is much less severely affected by over-specifying μ.

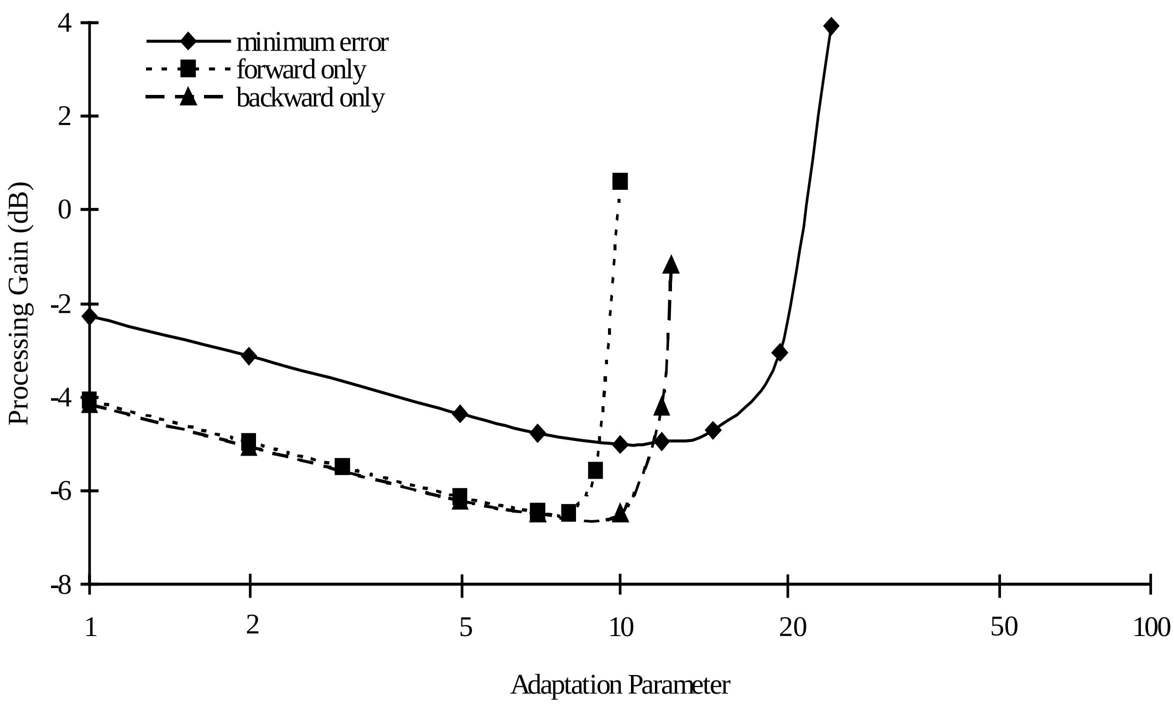

Figure 1: Variation in mean processing gain for natural speech as a function of adaptation parameter, μ, for three different update algorithms.

4. SOME COMPARISONS

Although the experiments in section 3 demonstrated that it was possible to use a higher rate of adaptation with the new method, it is necessary to look at some examples of the residual signals produced by the respective predictors, in order to visualise the causes and effects of the differences. There follow some examples of both synthetic and natural data to illustrate the behaviour of the FBME system.

4.1. Synthetic Data

To test the response of the FBME method to an 'ideal' data sequence, a set of 12th order predictor coefficients were extracted from a steady vowel taken from a sample of natural speech. The corresponding AR system was then driven by a sequence of randomly-occurring, random-valued impulses. This produced the waveform shown as a dotted line shown in figure 2. The forward prediction error of the FBME analysis, e_n, is shown as a solid line in the same figure.

For comparison, the dotted line in figure 3 shows the corresponding residual for the conventional LMS method. It is clear that, on the occurrence of each impulse, the LMS method is severely disrupted, with rapid over-compensation and consequent prolongation of the large error component. Although slightly perturbed by each impulse, the FBME method does not prolong the disturbance: The only large error occurs at the moment at which the impulse occurs.

In these and the graphs which follow, the sampling rate was 16,000 samples per second, so the total time interval shown in each graph corresponds to 200 samples.

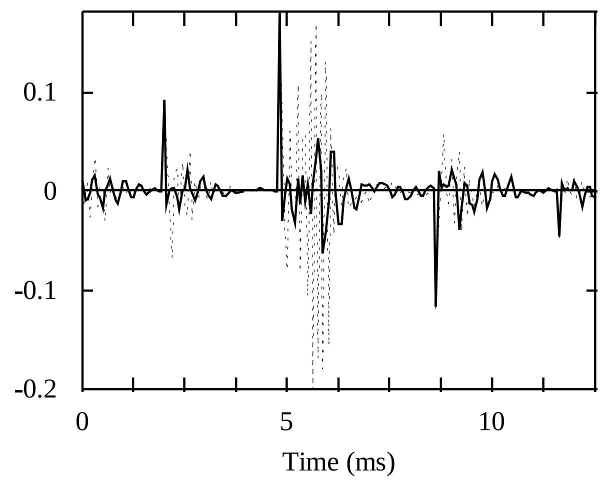

Figure 3: Synthetic source (as in figure 2) forward residual signals from a 12th order 'minimum error' method (solid line) and a conventional 12th order 'forward prediction' LMS algorithm (dotted line).

4.2. Natural Data

Corresponding graphs have been plotted in figures 4 and 5 for a segment of voiced speech. The speech was preemphasised and μ was set to 10. The FBME method behaves much as it did on the synthetic data. However, the residual for the conventional LMS algorithm in figure 5 shows a transient period of instability following each pitch event. This becomes increasingly common if the adaptation rate parameter, μ, is set larger.

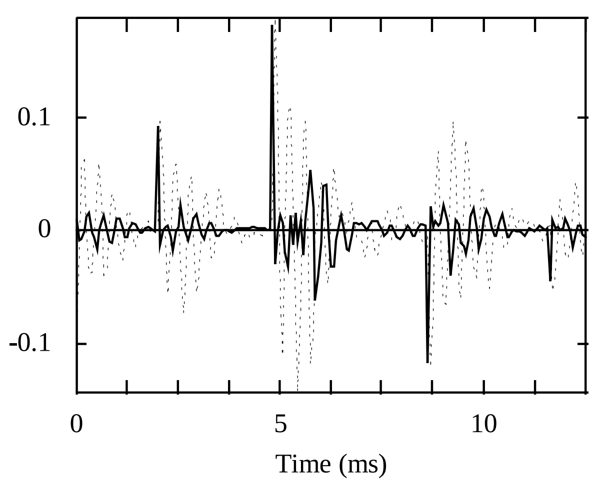

Figure 2: Forward residual signal from a 12th order 'minimum error' method (solid line) applied to the output of an ideal 12th order AR system driven by random impulses (dotted line).

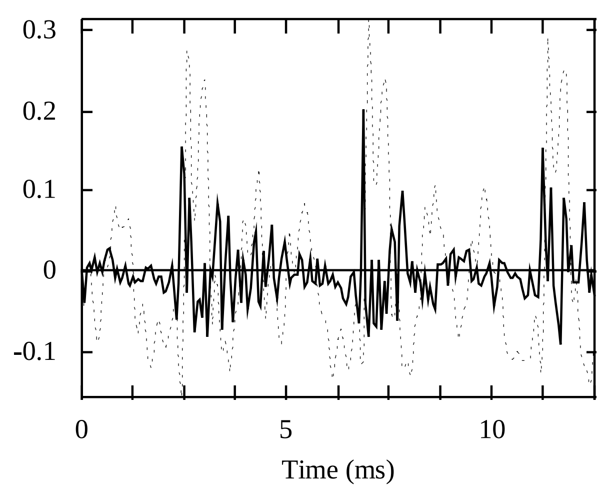

Figure 4: Forward residual signal from a 12th order 'minimum error' method (solid line) applied to a segment of voiced speech (dotted line).

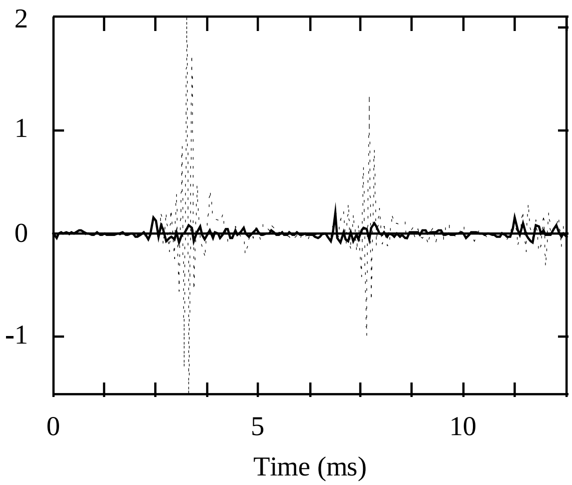

Figure 5: Natural speech (as in figure 4) forward residual signals from a 12th order ‘minimum error’ method (solid line) and a conventional 12th order ‘forward prediction’ LMS algorithm (dotted line).

5. CONCLUSIONS

It has been shown in passing in this paper that the LMS algorithm is better suited to modelling of speech when applied backwards in time. This appears to be due to the temporal asymmetry in the ringing produced by the vocal tract resonances.

However, the main conclusion to be drawn from this work is that the automatic selection of forward or backward (in time) versions of the LMS algorithm can be beneficial if the fastest adaptation rate is required.

It has been demonstrated that even a simple decision based on the local magnitude of the forward and backward residuals is sufficient to allow nearly 50% faster adaptation. It seems likely that more sophisticated algorithms for selecting the optimum update from the two candidates could yield even better performance.

6. REFERENCES

1. B. Widrow *et al.*, "Adaptive antenna systems", *Proc. IEEE*, vol. 55, no. 12, pp. 2143-2159, 1967.

2. S. S. Reddi, "A time-domain adaptive algorithm for rapid convergence", *Proc. IEEE*, vol. 72, no. 4, pp. 533-535, 1984.

3. B. S. Atal and J. R. Remde, "A new model of LPC excitation for producing natural-sounding speech at low bit rates", *Proc. IEEE ICASSP-82*, May 1982, pp. 614-617.

4. W. A. Gardner, "Learning characteristics of stochastic-gradient-descent algorithms: a general study, analysis and critique", *Signal Processing*, vol. 6, pp. 113-133, 1984.