Ill-Conditioning and Bandwidth Expansion in Linear Prediction of Speech

Peter Kabal

Department of Electrical & Computer Engineering
McGill University
Montreal, Canada

v1.5 August 2021

Revision History:

- 2003-02 v1.1: First web version
- 2021-08 v1.5: Expanded discussion of correlation lag windowing

Abstract

This report examines schemes that modify linear prediction (LP) analysis for speech signals. First, techniques which improve the conditioning of the LP equations are examined. White noise compensation for the correlations is justified from the point of view of reducing the range of values which the predictor coefficients can take on. Other techniques which modify the correlations are evaluated. These include adding highpass noise and selective power spectrum modification. The efficacy of these procedures is measured over a large speech database. The results show that white noise compensation is the method of choice — it is both effective and simple.

Methods to prematurely terminate the iterative solution of the correlation equations (Durbin recursion) to circumvent ill-conditioning are investigated.

The report also explores the bandwidth expansion of discrete-time filters which have resonances. In speech coding such resonances correspond to the formant frequencies. Bandwidth expansion of the LP filter serves to avoid unnatural sharp resonances that may be artifacts of pitch and formant interaction. Lag windowing of the correlation values has been used with the aim of both bandwidth expansion and improving the conditioning of the LP equations. Experiments show that the benefit for conditioning is minimal. This report also discusses bandwidth expansion of the prediction coefficients after LP analysis using radial scaling of the z-transform. A simple new formula is given which can be used to estimate the bandwidth expansion.

Ill-Conditioning and Bandwidth Expansion in Linear Prediction of Speech

1 Introduction

This report examines techniques which have been employed to modify the linear prediction (LP) analysis of speech signals. One goal of this work is to evaluate methods to improve the conditioning of the LP equations. A motivation for doing so is to limit the dynamic range of the resulting prediction coefficients (for fixed-point implementations). A number of techniques to improve conditioning are described and evaluated.

Secondly, there are a number of approaches to bandwidth expansion of the resonances of LP spectral models. The use of some of these methods has also been proposed to reduce the possibility of ill-conditioning. Indeed this link between bandwidth expansion and improving the conditioning of the LP equations was a starting point for this work. In speech coding resonances correspond to the formant frequencies. Bandwidth expansion is used to dampen the resonances. Modern speech coders employ bandwidth expansion in two places: lag windowing of the correlations before linear predictive (LP) analysis and/or modification of the LP coefficients.

A subset of the material in this report appears in a conference paper [1].

2 Linear Predictive Analysis

Linear predictive analysis fits an all-pole model to the local spectrum of a (speech) signal. The model is derived from the autocorrelation sequence of a segment of the speech. The LP spectral fit is determined by solving a set of linear equations based on the correlation values.

Let the input signal be $x[n]$. This signal is windowed to give a frame of data to be analyzed,

$$x_w[n] = w[n]\, x[n]. \tag{1}$$

The linear prediction formulation minimizes the difference between the windowed signal and a linear combination of past values of the windowed signal,

$$e[n] = x_w[n] - \sum_{k=1}^{N_p} p_k\, x_w[n-k]. \tag{2}$$

The goal is to minimize the total squared error,

$$\varepsilon = \sum_{n=-\infty}^{\infty} |e[n]|^2. \tag{3}$$

For the case that the window is finite in length, the terms in the sum for the squared error will be non-zero only over a finite length interval.

The predictor coefficients (p_k) which minimize ε can be found from the following set of equations

$$\begin{bmatrix} r[0] & r[1] & \cdots & r[N_p-1] \\ r[1] & r[0] & \cdots & r[N_p-2] \\ \vdots & \vdots & \ddots & \vdots \\ r[N_p-1] & r[N_p-2] & \cdots & r[0] \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_{N_p} \end{bmatrix} = \begin{bmatrix} r[1] \\ r[2] \\ \vdots \\ r[N_p] \end{bmatrix}. \tag{4}$$

The autocorrelation values are given by

$$r[k] = \sum_{n=-\infty}^{\infty} x_w[n]\, x_w[n-k]. \tag{5}$$

For finite length windows, the sum needs be evaluated only over a finite interval — the rest of the correlation coefficients will be zero. In vector-matrix notation, the correlation (LP) equations can be written as

$$\mathbf{R}\mathbf{p} = \mathbf{r}. \tag{6}$$

This equation gives the predictor coefficients which minimize the squared error. The mean-square error for a general set predictor coefficients $\mathbf{p}$ is

$$\varepsilon = r[0] - 2\mathbf{p}^T\mathbf{r} + \mathbf{p}^T\mathbf{R}\mathbf{p}. \tag{7}$$

Let the prediction error filter be denoted by $A(z)$,

$$A(z) = 1 - \sum_{k=1}^{N_p} p_k\, z^{-k}. \tag{8}$$

The autocorrelation formulation for the optimal prediction coefficients gives a matrix $\mathbf{R}$ which is Toeplitz. The Levinson-Durbin algorithm can be used to efficiently solve for the predictor coefficients. The prediction error filter ($A(z)$) will be minimum phase and the corresponding synthesis filter ($1/A(z)$) will be stable.

2.1 Conditioning and Predictor Coefficient Values

Of some importance for implementations using fixed-point arithmetic is the dynamic range of the predictor coefficients. If the roots of the prediction error filter are denoted as z_n, we can write the prediction error filter as a product of its root polynomials,

$$A(z) = \prod_{n=1}^{N_p} (1 - z_n z^{-1}). \tag{9}$$

The k'th coefficient of $A(z)$ in the expression of Eq. (8), p_k, is the coefficient of z^{-k}. This coefficient is the sum of the products of the roots (z_n), taken k at a time. For the k'th coefficient, there are $\binom{N_p}{k}$ such products. Since the roots have magnitude less than one (for stability of the synthesis filter), each product of the roots has a magnitude less than unity. The largest possible value for a coefficient occurs for coefficient $N_p/2$ and is $\binom{N_p}{k}$ when all roots have unit magnitude. For the case of $N_p = 10$, common for speech coders, the predictor coefficients have a magnitude which is bounded by 252.

Many modern speech coders (G.729 [2] and SMV [3], for example) store the predictor coefficients in 16-bit fixed-point, with 12 fractional bits (Q12 format). Putting aside one bit for the sign, this leaves 3 bits for the integer part of the representation. The Q12 format can only represent predictor values with values in the range $[-8, 8)$. Methods for conditioning the LP equations will be assessed in terms of the loss in performance required to bring the predictor coefficients into this range.

Problems with large predictor coefficient values will be worst for systems with singularities near the unit circle. These are the systems that are the most predictable. On the other hand, a system with uncorrelated data will have all predictor coefficients equal to zero. The numerical condition of a system of equations can be measured by the condition number. The spectral norm form of the condition number is the ratio of largest eigenvalue to smallest eigenvalue [4],

$$\gamma = \frac{\lambda_{\max}}{\lambda_{\min}}. \tag{10}$$

The *baseline* speech processing system uses a 240 sample Hamming window (200 sample frame advance) and 10'th order LP analysis. The input speech is modified IRS filtered [5] and sampled at 8 kHz. The speech database has 25 speakers, including two children, for a total of 225,779 frames [6].

The condition number of the autocorrelation matrix and the maximum predictor coefficient value were computed for each frame of the database. Figure 1 shows a histogram of the condition number expressed in power dB. The largest condition number encountered was 56.4 dB. The largest predictor coefficient generated was 11.0 for a frame with a condition number of 48.7

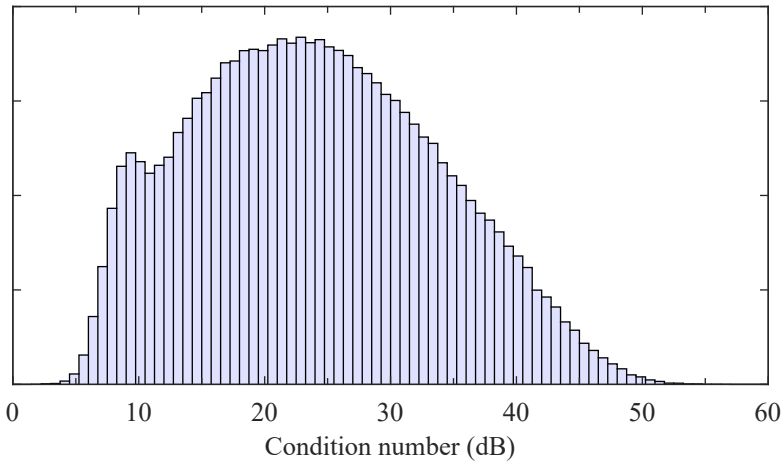

Fig. 1 Histogram of condition numbers (dB) - baseline processing.

dB. These frames occur during normal speech. For instance, the frame with the largest condition number occurs in male speech in the middle of the word "both" in the sentence "The pencil was cut sharp at both ends". The largest predictor coefficient occurs in female speech at the end of the word "floor" in the sentence "Tack the strip of carpet to the worn floor". In both cases, the waveforms in these regions are somewhat sinusoidal.

Figure 2 shows a scatter plot of the values of the largest predictor value[1] against the condition number for each frame in the database. There are only a few frames with large predictor values (exceeding ± 8) and they tend to occur for large condition numbers.

2.2 Power Spectrum Modification

The eigenvalues of a Toeplitz matrix formed from autocorrelation values are bounded by the minimum and maximum values of the power spectrum [4],

$$\min_{\omega} S(\omega) \leq \lambda_{\min} \leq \lambda_{\max} \leq \max_{\omega} S(\omega) \tag{11}$$

$$\gamma = \frac{\lambda_{\max}}{\lambda_{\min}} \leq \frac{\max_{\omega} S(\omega)}{\min_{\omega} S(\omega)}, \tag{12}$$

where the power spectrum is the Fourier transform of the correlation values and is given by

$$S(\omega) = \sum_{k=-\infty}^{\infty} r[k]\, e^{j\omega k}. \tag{13}$$

[1] The largest predictor value can take on either sign.

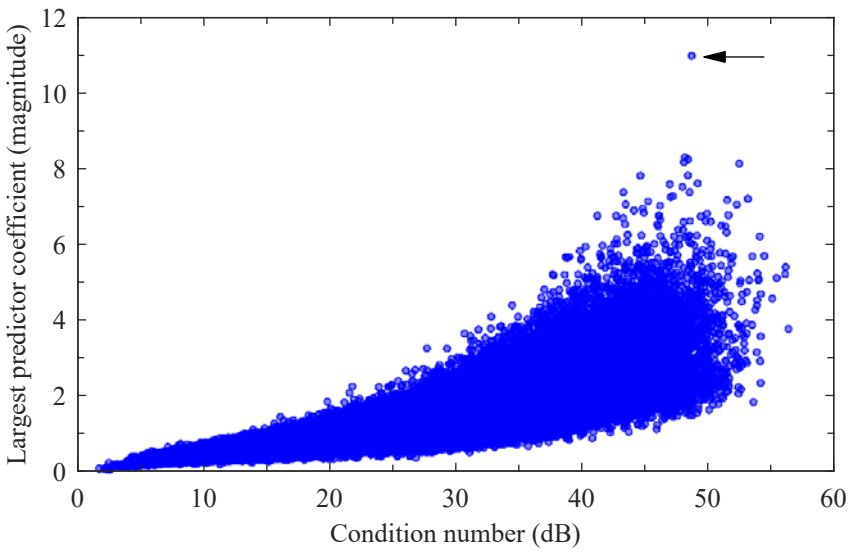

Fig. 2 Scatter plot of the magnitude of the largest predictor coefficient and the condition number for each frame in the database – baseline processing.

Note that this sum involves all of the autocorrelation coefficients, not just those that appear in the correlation matrix. The condition number is related to the dynamic range of the power spectrum — a flat spectrum gives the best conditioned equations, while spectra with large dynamic ranges can give badly conditioned equations.

Consider spectra with large peaks. These can be due to sinusoidal components, for instance dual tones (DTMF) used for signalling. They can, however, also occur in speech. A high pitched voice (female or child) uttering a nasalized sound can generate a surprisingly sinusoidal waveform. Problems occur because these sinusoidal signals are very predictable. In fact for pure sinusoids, only two predictor coefficients per sinusoid are needed to achieve perfect prediction. After applying a tapered window to the input data, we no longer have a pure sinusoids, but the effect of ill-conditioning is still present. As discussed earlier, large condition numbers occur for segments of ordinary speech.

2.2.1 Diagonal loading

A simple modification to reduce the eigenvalue spread is diagonal loading of the correlation matrix (adding a positive term to the zeroth autocorrelation coefficient).[2] In the power spectral domain, this is equivalent to adding a constant white noise term (referred to as white noise compen-

[2]In the matrix computations literature a similar approach is termed *ridge regression* [7].

sation). White noise compensation reduces the bound on the eigenvalue spread.

This same approach is used in fractional spacing equalizers for data transmission [8]. There the problem is ill-conditioning due to over-sampling. Adaptive adjustment algorithms such as LMS are subject to having the taps wander into overload regions. The mean-square error criterion can be modified to constrain the size of the predictor coefficients,

$$\varepsilon' = \varepsilon + \mu \mathbf{p}^T \mathbf{p}. \tag{14}$$

Taking the derivative with respect to the tap weights gives equations of the same form as earlier, but with the correlation matrix replaced by

$$\mathbf{R}' = \mathbf{R} + \mu \mathbf{I}. \tag{15}$$

For the case of a Toeplitz correlation matrix, we can write the modified correlation values as

$$r'[k] = r[k] + \mu \delta[k]. \tag{16}$$

Let μ be proportional to $r[0]$,

$$\mu = \epsilon r[0], \tag{17}$$

giving a multiplicative modification of the correlations

$$r'[k] = r[k](1 + \epsilon \delta[k]). \tag{18}$$

With the modified correlation, the bounds on the condition number ($\epsilon \geq 0$) are as follows,

$$\gamma \leq \frac{\max_\omega (S(\omega) + \epsilon r[0])}{\min_\omega (S(\omega) + \epsilon r[0])} \leq \frac{\max_\omega S(\omega)}{\min_\omega S(\omega)}. \tag{19}$$

Applying diagonal loading by multiplying the zeroth correlation value by the factor 1.0001 ($\epsilon = 0.0001$) improves the condition numbers. The largest condition number is now 48 dB and the largest predictor value is now 4.6 (within the range of ± 8 for a Q12 representation). A histogram of the resulting condition numbers is shown in Fig. 3.

The use of white noise compensation comes at the loss of optimality of the predictor for all frames. The predictor performance can be measured in terms of the ratio of the energy of the input signal to the energy of the prediction residual ε (see Eq. (7)). This ratio is the prediction gain,

$$G_P = \frac{r[0]}{\varepsilon}. \tag{20}$$

A histogram of the prediction gains (in dB) for the baseline configuration (i.e., without white

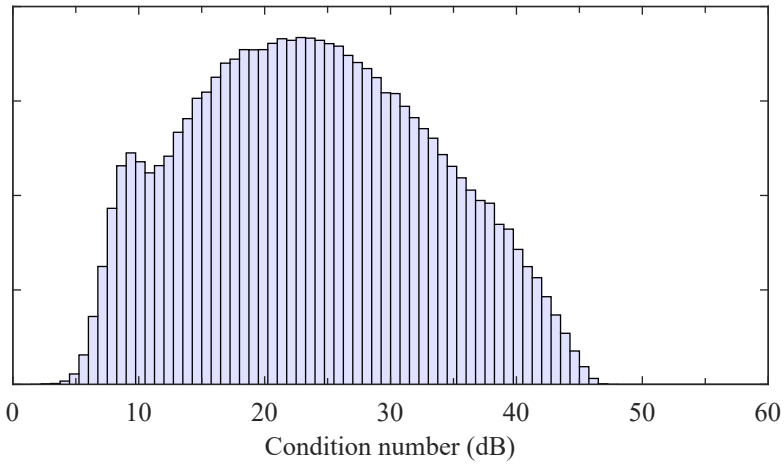

Fig. 3 Histogram of condition numbers (dB) – baseline processing + white noise compensation with $\epsilon = 0.0001$.

noise compensation) is shown in Fig. 4.

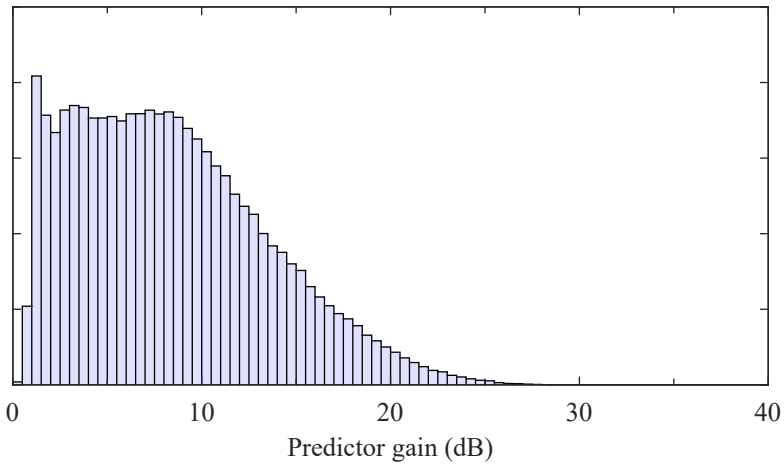

Fig. 4 Histogram of the prediction gains – baseline processing

A plot of the loss in prediction gain due to white noise compensation ($\epsilon = 0.0001$) is shown in Fig. 5. This is measured with the predictor coefficients calculated with white noise compensation acting on the original windowed input. The plot has points corresponding to the bins of the histogram of prediction gains. For each point, the circle indicates the average of the losses in prediction gains (in dB) within the histogram bin. The error bars indicate the smallest and largest prediction gain losses within the bin. One can notice that for frames with prediction gains below about 27 dB, the average prediction gain loss due to the use of white noise compensation (circles on the plot) is near the minimum loss. This indicates that the larger losses in prediction gain are

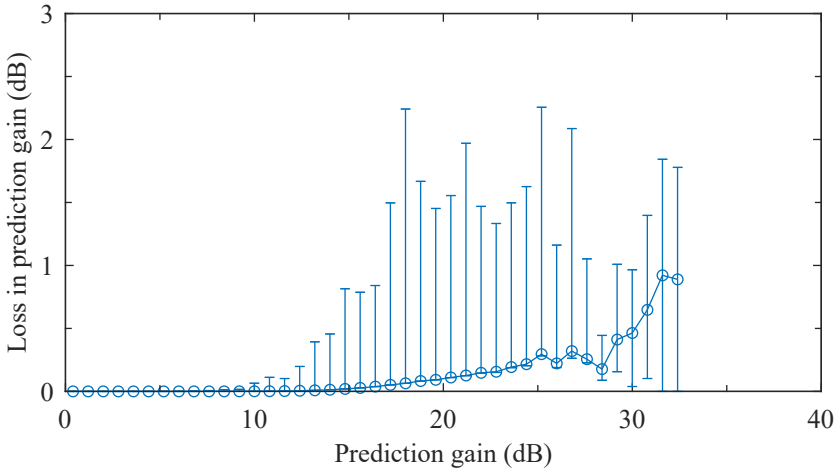

Fig. 5 Loss in prediction gain due to white noise compensation ($\epsilon = 0.0001$) versus the original prediction gain. The circles indicate the average of the prediction gain losses (in dB), while the error bars indicate the minimum and maximum prediction gain losses.

rare. For the frames with higher prediction gains, the mean loss in prediction gain lies close to midway between the minimum and maximum values.

The effect of changing ϵ was tested across the database. Figure 6 shows the maximum condition number as a function of the amount of white noise compensation. The curve is relatively smooth in spite of the fact that the speech frame giving the largest condition number is not the same for the different points on the plot.

Figure 7 shows the largest predictor coefficient value as a function of the amount of white noise compensation. It can be seen that the choice of $\epsilon = 0.0001$ is a reasonable value to keep the largest predictor coefficient below 8.

2.2.2 High frequency compensation

A slightly different tack was taken by Atal and Schroeder [10]. Their concern was that a large power gain (sum of the squares of the predictor filter coefficients) causes excessive feedback of quantization noise in a closed loop quantizer. They identified that the ill-conditioning that leads to large power gains was caused by the use of sharp lowpass filters in the signal path. These filters tend to leave a null in the spectrum near the half-sampling frequency. To compensate for this spectrum null, they applied a high-frequency compensation by adding the correlation for highpass noise to the speech correlation values.

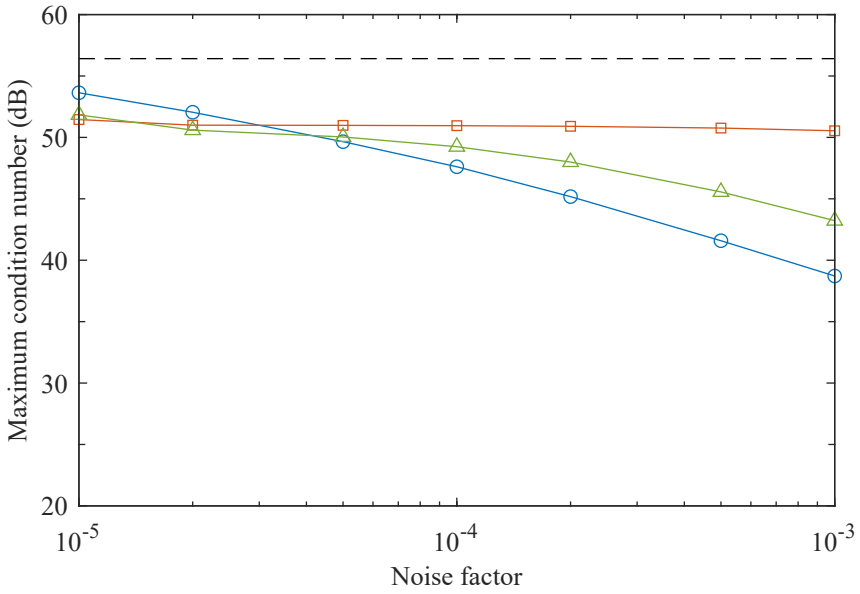

Fig. 6 Largest condition number versus ϵ (circles: white noise compensation, squares: highpass noise compensation, triangles: mixed noise compensation). The dashed line is the largest condition number for $\epsilon = 0$.

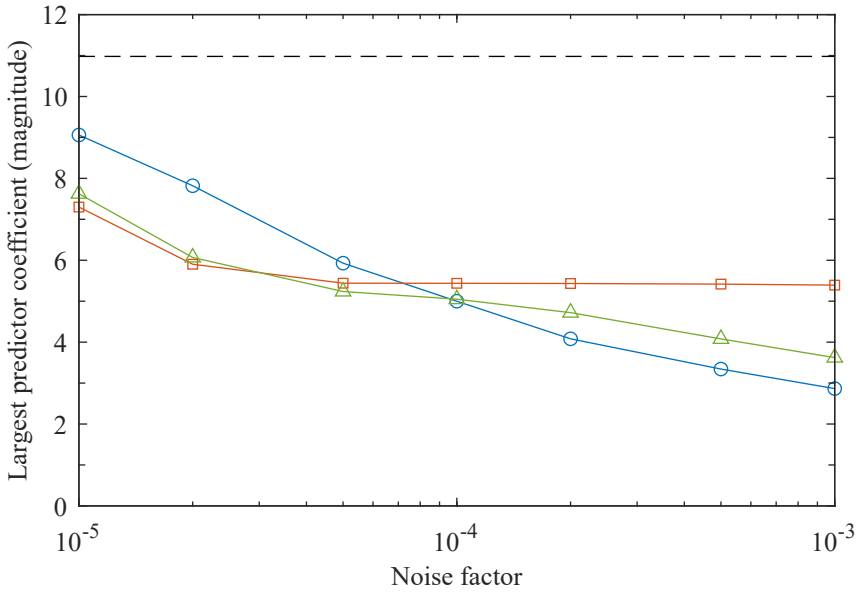

Fig. 7 Largest predictor coefficient value versus ϵ (circles: white noise compensation, squares: highpass noise compensation, triangles: mixed noise compensation). The dashed line is the largest predictor coefficient value for $\epsilon = 0$.

The noise spectrum used in [10] was that of white noise filtered by a highpass filter,

$$H_{\mathrm{HP}}(z) = (1 - z^{-1})^2. \tag{21}$$

The correlation function for the highpass noise, normalized to unit variance, is

$$r_n[k] = \begin{cases} 1 & \text{for } k = 0, \\ -\frac{2}{3} & \text{for } |k| = 1, \\ \frac{1}{6} & \text{for } |k| = 2, \\ 0 & \text{elsewhere.} \end{cases} \tag{22}$$

The power spectrum of the highpass noise has a peak value $8/3$ times that for white noise of the same power. The peak occurs at $\omega = \pi$, while a null occurs at $\omega = 0$.

This correlation for the highpass noise added to the signal correlation gives,

$$r'[k] = r[k] + \epsilon r[0] r_n[k]. \tag{23}$$

A histogram of the condition numbers for highpass noise compensation ($\epsilon = 0.0001$) is shown in Fig. 8. Note that the histogram falls off dramatically above 40 dB, but a small number of frames still have condition numbers which reach up to 51 dB (see Fig. 6). The present histogram for highpass noise compensation can be compared to the previous histogram for white noise compensation of the same energy (see Fig. 3).

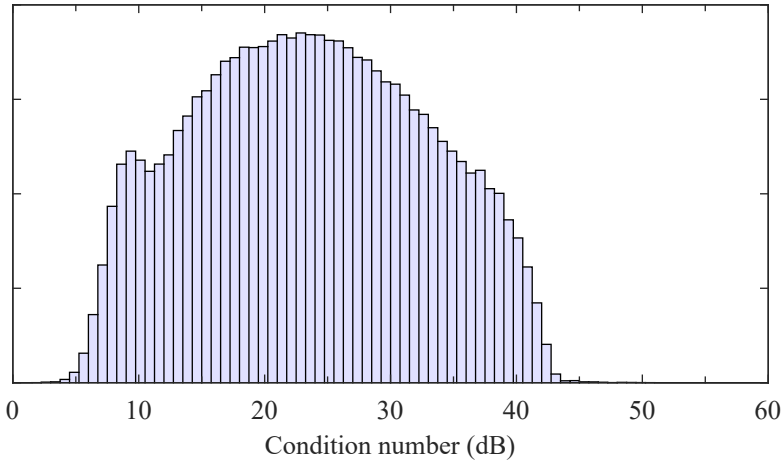

Fig. 8 Histogram of condition numbers (dB) – baseline processing + highpass noise compensation with $\epsilon = 0.0001$.

The effect of varying ϵ on the largest predictor coefficient value is shown as one of the curves

in Fig. 7. One can see that for small values of ϵ, highpass noise is more effective at reducing the maximum coefficient value than white noise of the same energy. When ϵ is large, the predictor coefficients will be those appropriate for predicting the highpass noise. The largest predictor coefficient for this case is 2.18. This is in contrast to the white noise compensation situation, in which the predictor coefficients move towards zero as ϵ increases.

A mixed form of noise compensation was also tested. The mixed noise combines white noise and highpass noise, with the portions adjusted to control the ratio of the power spectrum at zero frequency to the power spectrum at π. That ratio was chosen to be $1/10$. The noise has a highpass nature, but a significant amount of noise remains at low frequencies. The resulting noise correlation values (unit variance) are

$$r_n[k] = \begin{cases} 1 & \text{for } k = 0, \\ -\frac{18}{35} & \text{for } |k| = 1, \\ \frac{9}{70} & \text{for } |k| = 2, \\ 0 & \text{elsewhere.} \end{cases} \tag{24}$$

The modified correlation is obtained by adding the scaled highpass correlation to the signal correlation,

$$r'[k] = r[k] + \epsilon r[0] r_n[k]. \tag{25}$$

The effect of varying ϵ on the condition number and on the maximum predictor coefficient value is shown in Fig. 6 and Fig. 7. The mixed noise is marginally preferable (over white noise) for small values of ϵ. It was initially hoped that for a given value of ϵ, the prediction gain loss for mixed noise would be less than the loss for white noise. This did not turn out to be so. The average prediction loss, for instance at $\epsilon = 0.0001$, was significantly larger than that for white noise compensation.

2.2.3 Selective power spectrum compensation

The previous approaches bias the solutions even when the equations are already well conditioned. Instead of adding a constant power spectrum (white noise compensation), an approach which fills in only spectral valleys was tested,

$$S'(\omega) = \max(S(\omega), \epsilon r[0]), \tag{26}$$

where $\epsilon r[0]$ is the power spectrum of the white noise. This approach modifies only those regions of the power spectrum of the signal that fall below that of the white noise. This process can be viewed as adding a noise correlation tailored to fill in the spectral valleys.

In practice, we have to use a DFT (via an FFT) which only gives us samples of the spectrum. Nonetheless, this process can be approximated as follows.

1. Compute the energy of the frame $r[0]$.
2. Pad the data frame to at least twice its length with zeros to avoid time aliasing. If the length of the frame is N_f, a vector of length at least $2N_f - 1$ is needed.
3. Take the DFT of the vector of samples.
4. Calculate the square magnitude of the frequency response to get the power spectrum.
5. Replace all values of the power spectrum falling below $\epsilon r[0]$ with that value.
6. Take the inverse DFT of the modified power spectrum. The first $N_p + 1$ values of the resulting correlation values will be used in the correlation equations.

This approach was implemented and the value of ϵ was varied. For a given value of ϵ, with selective power spectrum modification, the power spectrum is modified less than for the white noise case. The bounds on the condition number ($\epsilon \geq 0$) are as follows,

$$\gamma \leq \frac{\max_\omega S(\omega)}{\min_\omega \max(S(\omega), \epsilon r[0])} \leq \frac{\max_\omega S(\omega)}{\min_\omega S(\omega)}. \tag{27}$$

Depending on the value of ϵ, the bound for selective power spectrum modification can be smaller or larger than the bound for white noise compensation (see Eq. (19)). Experiments show that the maximum condition number for the white noise case is smaller than the maximum condition number for the selective power spectrum modification case. If the maximum condition number versus ϵ curve is compared to the curve of white noise compensation (as in Fig. 6) it appears slightly above (about 1 dB at $\epsilon = 0.0001$) and nearly parallel to the white noise curve. With $\epsilon = 0.0001$, the selective power spectrum modification approach changes the spectrum in 91% of the frames. The condition number for selective power spectrum compensation for $\epsilon = 0.0001$ is very close to that for white noise compensation with ϵ about 20% smaller. For those respective values of ϵ, the plots of the loss of prediction gain due to the modification of the correlation are very similar (see Fig. 5 for the white noise case). Given the similar results, the extra computation that is entailed by the selective power spectrum modification is hard to justify with respect to simple white noise compensation.

White noise compensation directly controls the sum of the squares of the prediction coefficient values. As such, it also tends to control the maximum prediction coefficient value. The other methods considered also affect the maximum prediction coefficient values, but in an indirect way.

In a later section, lag windowing is applied to the correlation values. This strategy is usually motivated by the need for bandwidth expansion. It does however also have an effect on the conditioning of the LP equations.

2.2.4 Notes on the addition of noise correlation

The optimization problem with the addition of noise can be written as

$$\varepsilon' = r[0] - 2\mathbf{p}^T\mathbf{r} + \mathbf{p}^T\mathbf{R}\mathbf{p} + \mu(r_n[0] - 2\mathbf{p}^T\mathbf{r}_n + \mathbf{p}^T\mathbf{R}_n\mathbf{p}). \tag{28}$$

The predictor coefficients which minimize ε' are found from

$$(\mathbf{R} + \mu\mathbf{R}_n)\mathbf{p} = \mathbf{r} + \mu\mathbf{r}_n. \tag{29}$$

For $\mu = 0$, we get the unmodified solution. For large μ, we get the predictor coefficients for prediction of the noise. Note that $\mathbf{r}_n$ is zero for white noise, but is non-zero for coloured noise.

An alternate minimization problem is

$$\varepsilon' = r[0] - 2\mathbf{p}^T\mathbf{r} + \mathbf{p}^T\mathbf{R}\mathbf{p} + \mu\mathbf{p}^T\mathbf{R}_n\mathbf{p}. \tag{30}$$

With this formulation, the predictor coefficients to minimize ε' are found from

$$(\mathbf{R} + \mu\mathbf{R}_n)\mathbf{p} = \mathbf{r}. \tag{31}$$

As μ gets large the predictor coefficients go to zero.

In both formulations the effective correlation matrix is Toeplitz. However, for the second case, the right-hand vector does not use the same correlation elements as the matrix. If one distinguishes the solution methods of Levinson (arbitrary right-hand vector) and Durbin (right-hand vector drawn from the matrix elements), the first formulation can use the Durbin recursion, while in the second the Levinson method must be used. In addition to the (modest) increase in computational complexity for the second formulation, one can no longer guarantee that the prediction error filter coefficients result in a minimum phase filter. It is only for the white noise case that the two formulations coincide.

2.3 Correction in the Levinson-Durbin Recursion

Correction can be applied while solving the correlation equations. Consider the use of trigger parameters that are the by-product of a standard Levinson-Durbin recursion, and hence do not impose a significant additional computational overhead.

2.3.1 Limiting the prediction gain

In the Durbin recursion, the residual energy is available at each iteration. It is easy to modify the procedure to test the prediction gain (ratio of residual energy to initial energy) at each iteration. If the maximum prediction gain is exceeded, the iterations are terminated, keeping the previously calculated predictor values and setting remaining prediction coefficients to zero. All frames with (final) prediction gains below the threshold are unchanged.

This procedure described in the previous paragraph was implemented. A limit of about 17 dB on the prediction gain brings the largest coefficient value below 8. With this threshold, the iterations were prematurely terminated in about 6.5% of the frames. Further insight can be gained from a scatter plot of the largest predictor coefficient versus the prediction gain (see Fig. 9). The results show that large coefficient values can occur for frames with only moderately high prediction gains. In fact it shows that the largest prediction value across all frames occurs for a frame with a prediction gain of about 18 dB.

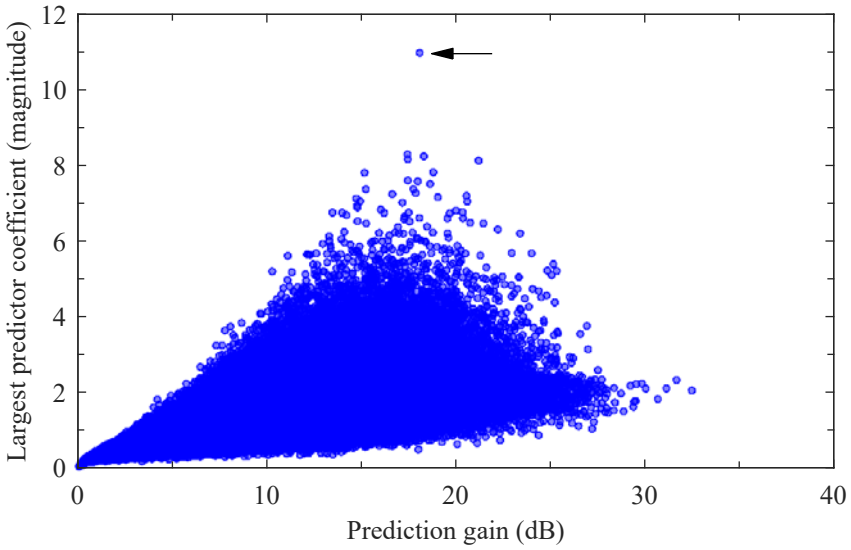

Fig. 9 Scatter plot of the largest predictor coefficient magnitude and the prediction gain for each frame in the database – baseline processing.

2.3.2 Limiting the largest reflection coefficient

Reflection coefficients with magnitude near unity signal ill-conditioning. At stage k in the Durbin recursion, the last coefficient of the predictor coefficients is equal to the kth reflection coefficient. Note that the last predictor coefficient is the product of k roots. If all of the roots are near unity, the reflection coefficient will also have a magnitude near unity.

The Durbin recursion can be terminated when the absolute value of a reflection coefficient exceeds a threshold value. The predictor coefficients from the previous iteration will be retained. This procedure was implemented, but the threshold on the reflection coefficient necessary to bring the largest prediction coefficient value below 8 was much too aggressive. With a threshold of 0.75 on the reflection coefficient, 23% of the frames are affected. At such a setting, the large loss in prediction gain is unacceptable. A scatter plot of the largest predictor value against the largest reflection coefficient shows that large predictor coefficients can occur for relatively small reflection coefficient values (Fig. 10).

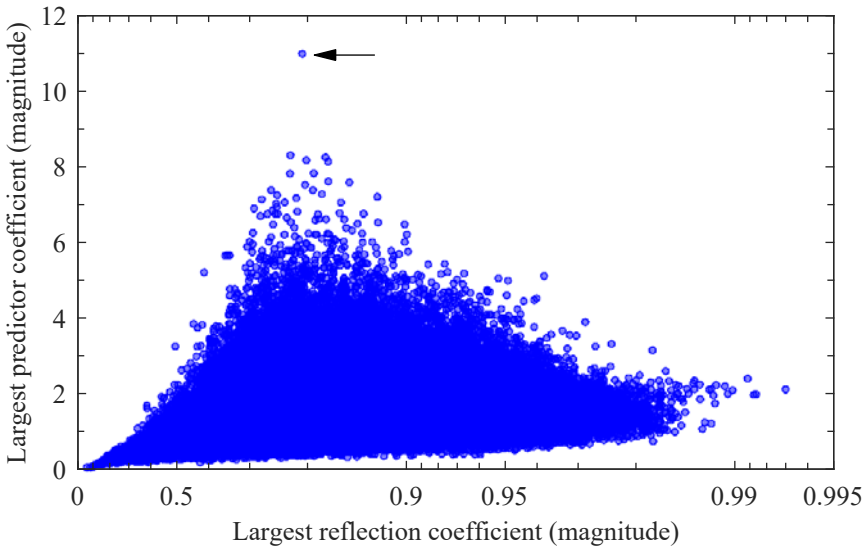

Fig. 10 Scatter plot of the largest predictor value and the largest reflection coefficient for each frame in the database - baseline processing.

3 Bandwidth Expansion

Bandwidth expansion is the process of taking a frequency response, usually with resonances or peaks, and broadening the bandwidths of those peaks. Such an expansion is useful in speech processing to prevent unnatural spectral peaks due to formant/pitch interactions. Such bandwidth expansion can be implemented before LP analysis (time windows or lag windowing) or on the LP coefficients after LP analysis (spectral damping).

3.1 Bandwidth Expansion Before LP Analysis

Bandwidth expansion before LP analysis can serve its purpose of broadening spectral peaks, but since it is done before LP analysis, it will also affect the conditioning of LP equations.

3.1.1 Time windows

The input signal is usually windowed with a tapered window (Hamming or other) prior to calculating the correlation values. The effect of the window in the frequency domain is the convolution of the frequency response of the window with the frequency response of the signal. This in itself constitutes a form of bandwidth expansion since the frequency response of the window has a non-zero main lobe width. For instance, the main lobe width (measured between zero crossings) for a Hamming window is $8\pi/N$ (asymptotically in N), where N is the window length [11].[3] For a 240 sample Hamming window (8 kHz sampling), the main lobe width of the frequency response of the Hamming window is 135 Hz between zero crossings and 44 Hz at the 3 dB points. Some bandwidth expansion due to time windowing is unavoidable.

3.1.2 Correlation lag windowing

Explicit bandwidth expansion prior to LP analysis is done by lag windowing the autocorrelation sequence [12, 13], often with a Gaussian or binomial shaped window. Since the autocorrelation sequence has as its Fourier transform the power spectrum, this correlation windowing corresponds to a periodic convolution of the frequency response of the window with the power spectrum. With lag windowing, a spectral line in the power spectrum will take on the shape of the main lobe of the spectrum of the lag window.

Conceptually, the lag windowing can be viewed as follows. Extended correlation values are computed from the windowed input speech samples. A modified set of correlation values is calculated by applying a lag window to the correlation values. The lag window itself must have a real non-negative Fourier transform to ensure that the power spectrum corresponding to the modified correlation values remains non-negative. This means that the lag window is itself a valid correlation sequence. Finally, the $N_p + 1$ low-lag values are used in the derivation of the prediction coefficients. In practice, only the low-lag correlations of the windowed input speech need be calculated and modified by the low-lag values of the lag window.

Gaussian Window

Consider a Gaussian window which is a function of continuous time,

$$w(t) = \exp(-\frac{1}{2}(at)^2). \tag{32}$$

[3] Reference [11] gives the bandwidth measured between zero crossings and the double-sided 6 dB bandwidth, measured between the half amplitude points. To be consistent with the bandwidths in this report, the double-sided 3 dB bandwidth is more meaningful – this is calculated from the points at which the amplitude-squared falls to one half.

This window is normalized such that $w(0) = 1$. A Gaussian probability density function has the same shape but is normalized to have unit area. The parameter $1/a$ plays the role of the standard deviation for a pdf.

The frequency response of the Gaussian window also has a Gaussian shape,

$$W(\Omega) = \frac{\sqrt{2\pi}}{a} \exp(-\frac{1}{2}(\frac{\Omega}{a})^2). \tag{33}$$

The parameter a is the standard deviation in this formula.

The two-sided bandwidth (radians) measured between the points at which the frequency response falls to $\exp(1/2)$ relative to the peak at zero frequency will be designated as $\Omega_{2\sigma}$.

$$\Omega_{2\sigma} = 2a. \tag{34}$$

The two-sided 3 dB bandwidth is

$$\Omega_{\text{bw}} = \sqrt{2\log(2)}\,\Omega_{2\sigma}. \tag{35}$$

The 3 dB bandwidth is about 18% larger than the 2σ bandwidth.

A sampled version of the Gaussian window leads to frequency aliasing of the frequency response. However with reasonably chosen bandwidth expansion factors and given the fact that the Gaussian frequency response dies off quickly with frequency, the effect of aliasing can be largely ignored in the calculation of the effective bandwidth expansion.

The discrete-time Gaussian window is

$$w[k] = \exp(-\frac{1}{2}(\frac{ak}{F_s})^2), \tag{36}$$

where F_s is the sampling rate. The parameter a can be expressed in terms of F_{bw}, the two-sided 3 dB bandwidth in Hz as

$$a = \frac{\pi}{\sqrt{2\log(2)}} F_{\text{bw}}. \tag{37}$$

Using this value, the sampled Gaussian window is

$$w[k] = \exp\big(-\frac{1}{4\log(2)}(\pi k \frac{F_{\text{bw}}}{F_s})^2\big). \tag{38}$$

For fractional bandwidths $F_{\text{bw}}/F_s < 10^{-1}$, the bandwidths determined from the continuous-time Gaussian window are very close to the actual fractional bandwidths, i.e., the effect of aliasing is negligible.

Binomial Window

Appendix A describes a normalized binomial window,

$$w[k] = \begin{cases} \dfrac{\binom{2M}{M+k}}{\binom{2M}{M}}, & -M \le k \le M, \\ 0, & \text{elsewhere.} \end{cases} \tag{39}$$

Direct calculation of the combinatorial terms for typical values of M is not possible. One has to use the log-gamma function to evaluate the factorial terms to avoid numerical overflow. As an alternative, the binomial window coefficients can be computed recursively. Consider $k \ge 0$,

$$w[k] = \begin{cases} 1, & k = 0, \\ \dfrac{M-k+1}{M+k}\, w[k-1], & 1 \le k \le M, \\ 0, & k > M. \end{cases} \tag{40}$$

The Fourier transform of the binomial window is

$$W(\omega) = C_M \big(\cos(\omega/2)\big)^{2M}. \tag{41}$$

This response is a function of the single parameter M. The 3 dB bandwidth can be expressed as a function of M by solving the following equation,

$$\Big(\cos\big(\frac{\pi}{2}\frac{F_{\text{bw}}}{F_s}\big)\Big)^{2M} = \frac{1}{2}, \tag{42}$$

giving

$$\frac{F_{\text{bw}}}{F_s} = \frac{2}{\pi}\cos^{-1}\big(\exp(-\frac{1}{2}\frac{\log(2)}{M})\big). \tag{43}$$

Noting that the binomial window can be approximated by samples of a Gaussian, an approximation to the bandwidth can be found. The standard deviation of the binomial window is $\sqrt{M/2}$ (see Appendix A, with $N = 2M$ and $p = q = 1/2$). Setting $a = \sqrt{2/M}$ in Eq. (36), the approximate bandwidth of the binomial window is

$$\frac{F_{\text{bw}}}{F_s} \approx \frac{2}{\pi}\sqrt{\frac{\log(2)}{M}}. \tag{44}$$

Figure 11 shows a plot of F_{bw}/F_s against M for a binomial window calculated using Eq. (43). For a 8000 Hz sampling rate, the relative bandwidths of 10^{-2} and 10^{-1} correspond, respectively, to

3 dB bandwidths of 80 and 800 Hz. A plot of the approximate formula Eq. (44) is indistinguishable from that of the exact formula on this plot – the maximum relative difference of 0.12% occurs for $M = 50$. The best match to a 3 dB bandwidth of 141.3 Hz (corresponding to $F_{2\sigma} = 120$ Hz for the Gaussian window) relative to a sampling rate of 8000 Hz, occurs for $M = 901$.

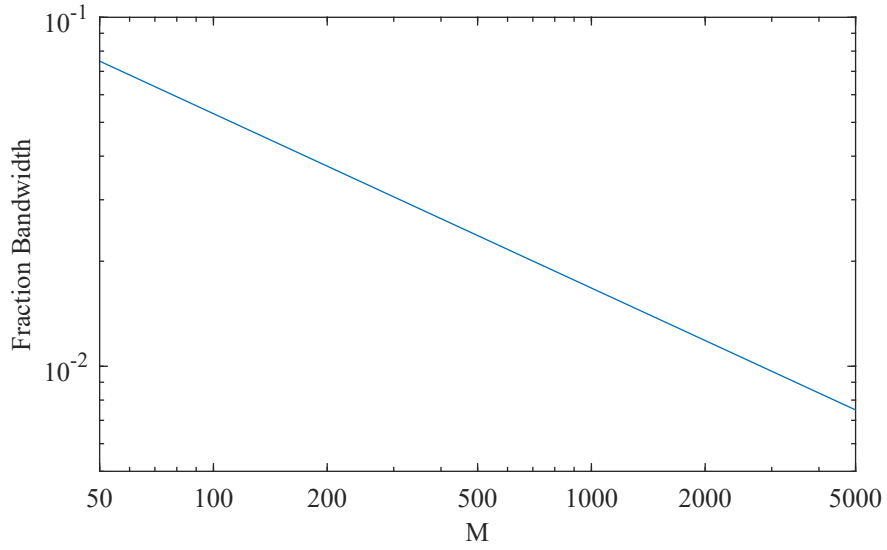

Fig. 11 Fractional 3 dB bandwidth of the binomial window with parameter M.

Lag Windowing

Lag windowing also affects the conditioning of the LP equations. Spreading the spectrum has the effect of helping to fill in spectral valleys and hence potentially reducing the condition number of the correlation matrix. In fact, this effect on conditioning is given as a rationale for using the lag window [9]. The ITU-T G.729 8 kb/s CS-ACELP [2] and SMV [3] speech coders use a Gaussian lag window with bandwidth $F_{2\sigma} = 120$ Hz (referred to as a bandwidth of 60 Hz in the coder standards) relative to a sampling rate $F_s = 8000$ Hz. Tests were run with the G.729/SMV lag window. Surprisingly, the effect on the worst condition number is small (now 56.0 dB, only a little better than the original 56.4 dB). Also the largest predictor coefficient value is now 10.7, only a little smaller than the original value of 11.0.

The above indicates that lag windowing with a Gaussian window by itself does not improve the conditioning by much. Note that white noise compensation can be represented as a lag window (see Eq. (18) which is for the case for white noise alone). In fact, lag windowing and white noise compensation can be combined into a single lag window. The combined effect of lag windowing ($F_{2\sigma} = 120$ Hz) and white noise compensation ($\epsilon = 0.0001$) on the prediction gain is

shown in Fig. 12. The figure can be compared to Fig. 5 which gives the situation for no lag windowing. The use of lag windowing causes additional loss of prediction gain, especially for frames with prediction gains below about 15 dB.

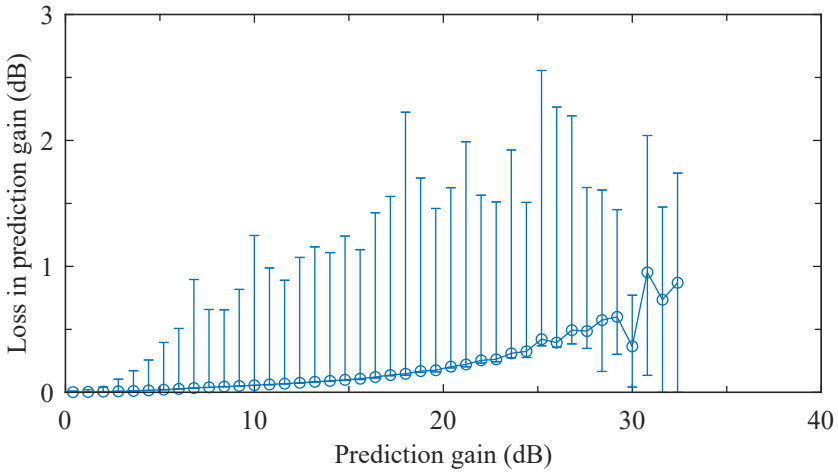

Fig. 12 Loss in prediction gain due to lag windowing (Gaussian window with $F_{2\sigma}$ set to 120 Hz) and white noise compensation ($\epsilon = 0.0001$) versus the original prediction gain. The circles give the average of the prediction gain losses (in dB), while the error bars indicate the minimum and maximum prediction gain losses.

3.2 Bandwidth Expansion After LP Analysis

In this section, bandwidth expansion after LP analysis is investigated. Consider a discrete-time filter $H(z)$ and a bandwidth expanded version of the same filter. In many cases, it is important that the bandwidth expanded version of the filter have the same form as the synthesis filter. For instance if $H(z)$ is an all-pole filter (as would arise from a standard LP analysis), the bandwidth-expanded version should also be all-pole. Replacing z by z/α satisfies this requirement. Consider the all-pole filter,

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 - \sum_{k=1}^{N} p_k z^{-k}}. \tag{45}$$

With bandwidth expansion [14], the filter has the same form, but with a new set of coefficients,

$$p'_k = \alpha^k p_k. \tag{46}$$

Replacing z by z/α moves the singularities of $H(z)$ inward ($\alpha < 1$) or outward ($\alpha > 1$). For a filter with resonances, choosing $\alpha < 1$ has the effect of expanding the bandwidth of the resonances.

3.2.1 Windowing with an exponential sequence

For a causal filter, the effect of replacing $H(z)$ with $H(z/\alpha)$ is such that the impulse response of the filter is modified to become

$$h'[n] = \alpha^n h[n], \tag{47}$$

i.e., the impulse response coefficients are multiplied by an exponential (infinite length) time window. In the frequency domain, the frequency response of the filter is convolved with the frequency response of the window,

$$W(\omega) = \frac{1}{1 - \alpha e^{-j\omega}}. \tag{48}$$

The 3 dB bandwidth of the window frequency response is

$$\omega_{\text{bw}} = 2\cos^{-1}(1 - \frac{(1-\alpha)^2}{2\alpha}), \qquad \text{for } 3 - 2\sqrt{2} \leq \alpha \leq 1. \tag{49}$$

Below the lower limit for α (about 0.172), the response does not decrease sufficiently to fall 3 dB below the peak.

3.2.2 Radial scaling of a bandpass filter

To see the effect of replacing z by z/α from another point of view, consider a bandpass filter with a single resonance. The bandwidth of a resonance has been well studied for continuous-time systems. The discrete-time bandpass filter is formed using a bilinear transformation.

Continuous-time bandpass filter

The second-order bandpass filter is

$$H(s) = \frac{s}{s^2 + (\Omega_o/Q)s + \Omega_o^2}. \tag{50}$$

Assume that $\Omega_o > 0$ and $Q > 0$. The magnitude of the frequency response for this filter has a peak at Ω_o. The 3 dB points of the response have a simple relation that is developed in Appendix B,

$$\Omega_u - \Omega_l = \Omega_o/Q. \tag{51}$$

Note that the frequency response exhibits a resonance even when the poles are real ($Q \leq 1/2$).

Discrete-time bandpass filter

The corresponding discrete-time filter can be found using a bilinear transformation,

$$z = -\frac{s+a}{s-a}. \tag{52}$$

The parameter a is real. This transformation maps the zeros at $s = 0$ and $s = \infty$ to $z = +1$ and $z = -1$, respectively. For complex poles, the poles get mapped to new locations $z_{1,2} = r_p e^{\pm j\omega_o}$,

$$H(z) = \frac{z^2 - 1}{z^2 - 2r_p \cos(\omega_o) z + r_p^2}. \tag{53}$$

Using the bilinear relationship, the 3 dB points in the response can be found for the discrete-time filter. The 3 dB bandwidth has a particularly simple form (see Appendix C),

$$\omega_{\text{bw}} = \pi/2 - 2\tan^{-1}(r_1 r_2). \tag{54}$$

where r_1 and r_2 are the radii of the poles of the discrete-time filter. Note that the poles do not have to be complex for this equation to apply, although the main concern for speech coding is that of resonances for which the poles appear as complex conjugate pairs with $r_1 = r_2$.

Bandwidth expansion

Consider scaling the z-transform

$$H'(z) = H(z/\alpha), \tag{55}$$

where $0 < \alpha \leq 1$. Then the new poles have radii $r_1' = \alpha r_1$ and $r_2' = \alpha r_2$. The 3 dB bandwidth of the resonance is now

$$\omega_u' - \omega_l' = \pi/2 - 2\tan^{-1}(\alpha^2 r_1 r_2). \tag{56}$$

a The difference in bandwidth due to radial scaling by α is

$$\Delta\omega_{\text{bw}} = 2\tan^{-1}\Big(\frac{r_1 r_2(1-\alpha^2)}{1+\alpha^2 r_1^2 r_2^2}\Big). \tag{57}$$

This bandwidth expansion given by Eq. (57) is plotted in Fig. 13 for three different values of $r = r_1 = r_2$. The ordinate is the bandwidth normalized by 2π. The scale runs from 10^{-3} through 10^{-2} to 10^{-1}. For a sampling rate of 8000, these values correspond to bandwidths of 8, 80, and 800 Hz. For a fixed value of α, the bandwidth depends on r.

In speech coding, the goal is to broaden sharp peaks in the spectral envelope. For the sequel, the bandwidth expansion will be the bandwidth of a spectral line (r near unity) after bandwidth

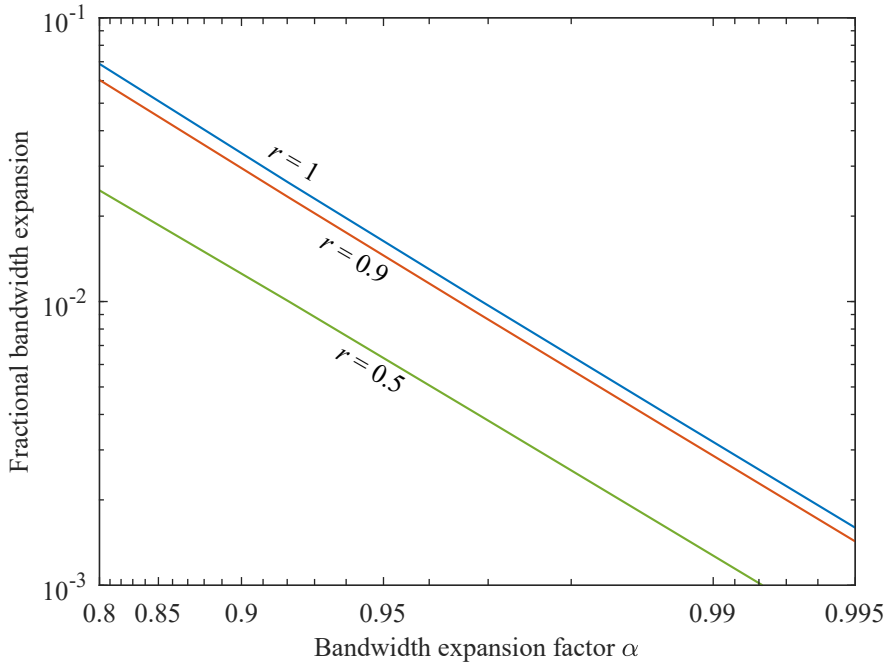

Fig. 13 Fractional bandwidth expansion ($\Delta\omega_{\text{bw}}/(2\pi)$ as a function of the bandwidth expansion parameter α.

expansion. In the limiting case of poles near the unit circle ($r_1 = r_2 \lessapprox 1$), the bandwidth after applying a bandwidth expansion factor of α is

$$\omega_{\text{bw}} = \pi/2 - 2\tan^{-1}(\alpha^2). \tag{58}$$

3.2.3 Comparison of bandwidth formulas

The two formulas, Eqs. (49) and (58) account for the effect of radial scaling of the z-transform. These give the bandwidth of a narrow spectral line after bandwidth expansion. The bandwidth expressions derived here can be compared to approximations that have appeared in the literature. Paliwal and Kleijn [15] give the bandwidth formula (converted to our notation)

$$\omega_{\text{bw}} = -2\log(\alpha). \tag{59}$$

This formulation can be derived from an impulse invariant transformation of a continuous-time exponential sequence. As such it ignores the aliasing effects that are taken care of in the derivation leading to Eq. (49).

For α near unity, all three expressions give similar results. An examination of the Taylor series

for the three expressions shows that they agree in value and the first derivative at $\alpha = 1$,

$$\omega_{\text{bw}} = 2(1-\alpha) + (1-\alpha)^2 + O((1-\alpha)^3). \tag{60}$$

Using just the first two terms gives a simple estimate of the bandwidth,

$$\omega_{\text{bw}} = 2(1-\alpha) + (1-\alpha)^2. \tag{61}$$

Table 1 summarizes the bandwidth formulas. The limits on α are those which give a bandwidth below 2π.

Table 1 Bandwidth formulas

ID	ω_{bw}	limits
$\cos^{-1}$	$2\cos^{-1}\left(1-(1-\alpha)^2/(2\alpha)\right)$	$3-2\sqrt{2} \le \alpha \le 1$
log	$-2\log(\alpha)$	$\exp(-\pi/2) \le \alpha \le 1$
$1-\alpha$	$2(1-\alpha)+(1-\alpha)^2$	$0 \le \alpha \le 1$
$\tan^{-1}$	$\pi/2 - 2\tan^{-1}(\alpha^2)$	$0 \le \alpha \le 1$

The bandwidths estimated by the different expressions are plotted in Fig. 14. The ordinate is the bandwidth normalized to the sampling frequency, $(\omega_{\text{bw}}/(2\pi))$. The figure shows that the simple formula (designated as $1-\alpha$) gives a good estimate of the bandwidth for useful values of α. All of the formulas give nearly identical values for $\alpha \ge 0.7$, corresponding to a normalized bandwidth of about 0.1.

Using the simple formula, a value for α can be determined for a given desired bandwidth expansion,

$$\alpha = 2 - \sqrt{1 + 2\pi F_{\text{bw}}/F_s}, \tag{62}$$

where F_{bw} is the 3 dB bandwidth in Hz and F_s is the sampling frequency.

In speech coding, bandwidth expansion with bandwidth values of 10 Hz to 30 Hz is used. With a sampling rate of 8000 Hz, these bandwidths correspond to values of α between 0.996 and 0.988. Figure 15 shows the prediction gain loss due to bandwidth expansion using $\alpha = 0.99$ (25 Hz bandwidth for $F_s = 8000$). The loss in prediction gain is much less than for the lag windowing considered earlier, but so is the amount of bandwidth expansion.

3.3 Line Spectral Frequencies

Line spectral frequencies (LSF's) are a transformation of the LP coefficients. The LSF's are an ordered set of values in the range $(0, \pi)$. Closely spaced LSF's tend to indicate a spectral resonance

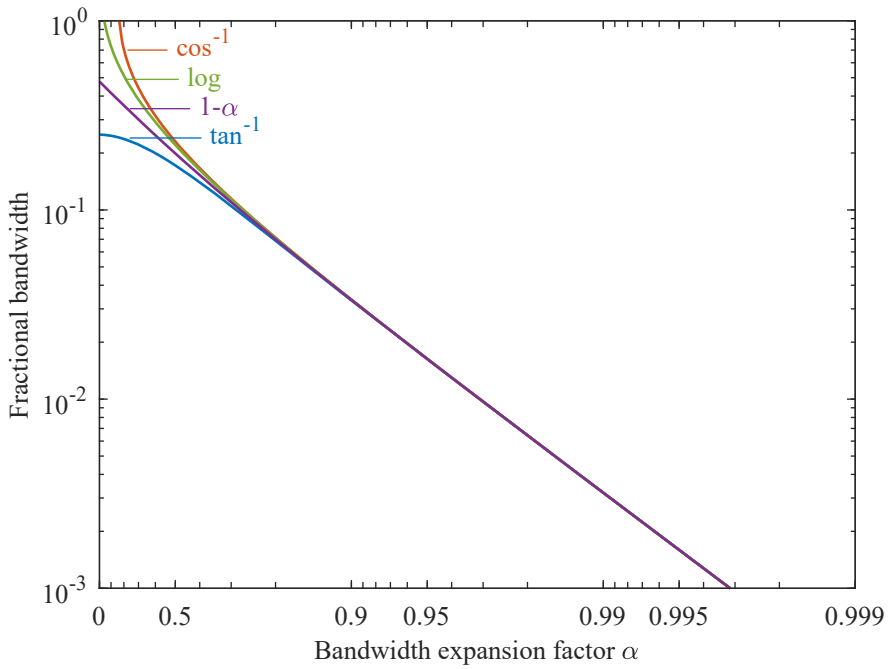

Fig. 14 Relative bandwidth $\omega_{\text{bw}}/(2\pi)$ as a function of the bandwidth expansion parameter α for the different bandwidth formulas. The labels for the curves use the ID designations in Table 1.

at the corresponding frequency. Several standard coders impose minimum separation constraints on the LSF's. This constraint implies a bandwidth expansion that is only applied in exceptional cases.

Another bandwidth expansion scheme for LSF's, albeit for use as a postfilter for speech processing, is described in [16]. In that paper, the LSF's are pushed apart by interpolating the given LSF's with a set corresponding to a flat spectrum (equally spaced LSF's),

$$\omega'[k] = (1-\mu)\omega[k] + \mu\omega_u[k], \qquad k = 1, \ldots, N_p, \tag{63}$$

where the $\omega_u[k] = k\pi/(N_p+1)$ are the LSF's corresponding to a flat spectrum. Reference [16] also suggests that the target LSF vector ($\omega_u[k]$) can be generalized, as can the value μ, by making it a function of frequency.

The relationship between LSF separation and the bandwidth of a resonance can be investigated as follows. Consider a two pole resonance at frequency ω_o and radius r near unity.

$$A(z) = 1 + 2r\cos(\omega_o)z^{-1} + r^2 z^{-2}. \tag{64}$$

The LSF's of this error prediction filter (extended with zeros to N_p+1 terms) are calculated. For

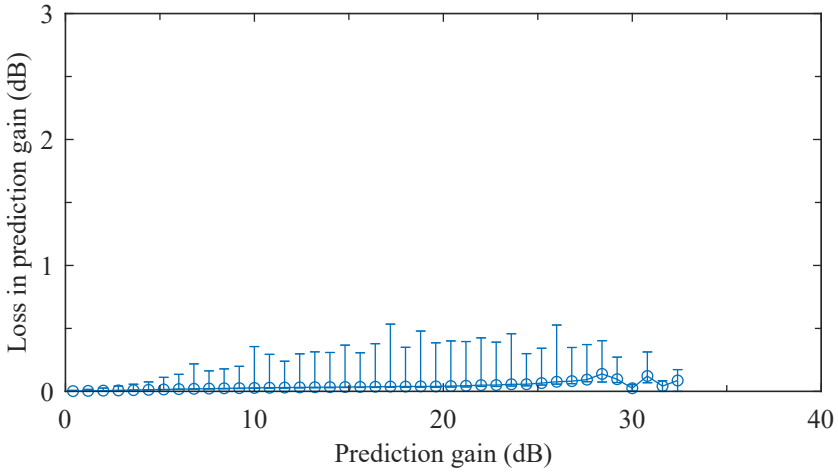

Fig. 15 Loss in prediction gain due to bandwidth expansion of the predictor coefficients ($\alpha = 0.99$) versus the original prediction gain. The circles indicate the average of the prediction gain losses (in dB), while the error bars indicate the minimum and maximum prediction gain losses.

some values of ω_o, two of the LSF's are closely spaced near ω_o. The remaining LSF's are equally spaced from 0 to π. However, the situation is more complicated for other values of ω_o when 3 LSF's play a role in defining the resonance. This situation is illustrated in Fig. 16 for $N_p = 10$, $r = 0.97$ and sampling frequency 8000 Hz.

The effective bandwidth expansion was investigated for the same resonant system. The resonance is now given a much narrower bandwidth by setting $r = 0.999$. The bandwidth expansion formula Eq. (63) is applied to the LSF's using $\mu = 0.15$. The bandwidth-expanded LSF's are converted back to give the coefficients of a prediction error filter $A'(z)$. The bandwidth of the resonance in this filter was measured. Figure 17 shows the 3 dB bandwidth of the bandwidth-expanded prediction error filter as a function of the resonance frequency. It can be seen that the bandwidth depends on the resonance frequency. This indicates that bandwidth expansion using LSF's gives inconsistent results.

4 Summary

This report has demonstrated the problems of ill-conditioning of the LP equations for speech signals. The standard technique of white noise compensation brings the predictor coefficients to a reasonable range. This approach can be viewed as constraining the sum of the squared coefficients through a Lagrange multiplier. The other approaches to power spectrum modification (highpass noise, mixed highpass and white noise, and selective power spectrum modification) are all effective at controlling both the condition number of the correlation matrix and the range of predictor

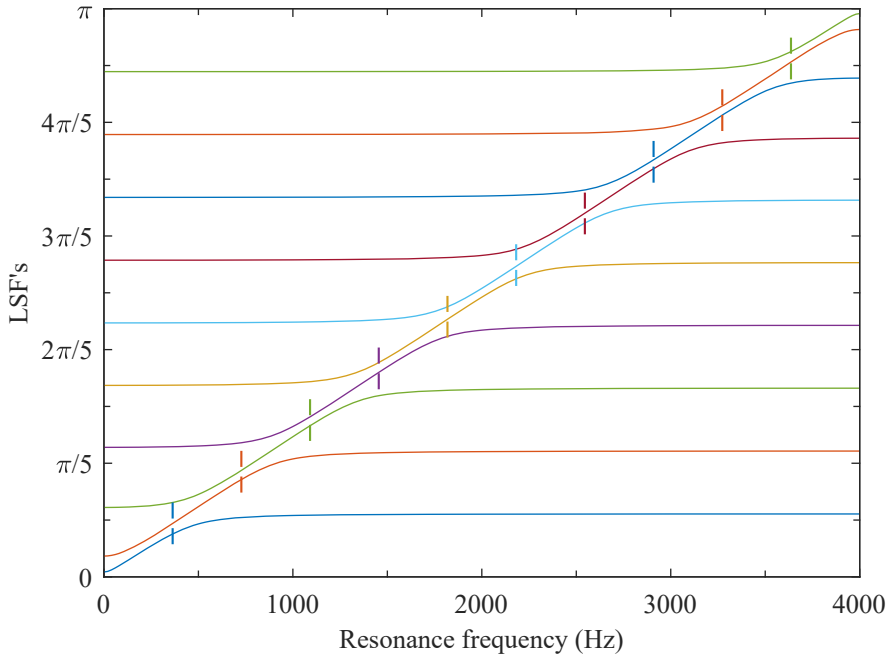

Fig. 16 LSF values for a resonant system ($r = 0.97$). The vertical dashed lines lines indicate the positions of the equally spaced LSF values corresponding to a flat spectrum.

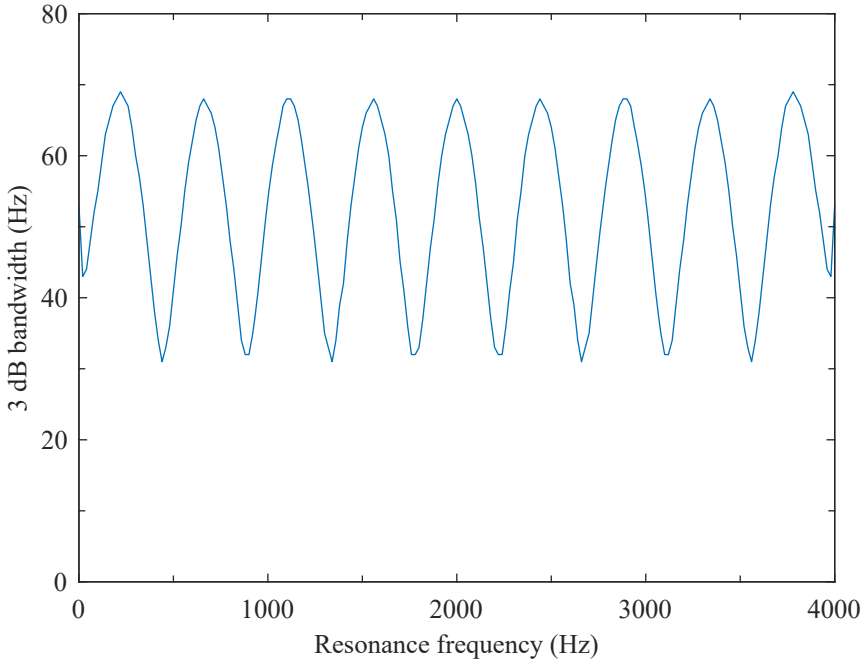

Fig. 17 3 dB bandwidth of the bandwidth-expanded prediction error filter as a function of the resonance frequency ($r = 0.999$ and LSF modification factor $\mu = 0.15$).

coefficients. However, white noise compensation still stands out as being both effective and simple to implement, while not imposing an excessive loss in prediction gain.

Methods which prematurely terminate the iterations for the solution of the correlation equations were examined. In one approach, the prediction gain was limited to be below a threshold. In another approach, the absolute value of the reflection coefficients was limited to be below a threshold. The Durbin recursion was halted at the point just before the threshold was exceeded and the prediction coefficients obtained at that point were used. For both strategies, the thresholds required to ensure that the resulting prediction coefficients were in a reasonable range (± 8) resulted in significant performance loss for a substantial portion of the frames processed.

Bandwidth expansion to prevent abnormally narrow formant peaks can be provided by lag windowing of the correlation values. This approach has surprisingly little effect on the conditioning of the LP equations. Other approaches to bandwidth expansion are applied after LP analysis. We have compared the formulas to calculate the amount of bandwidth expansion due to a radial scaling of the singularities of the LP filter. A new simple formula for calculating the bandwidth expansion by spectral damping has been developed.

We have also examined a method to modify the spacing between line spectral frequencies for the purpose of bandwidth expansion. It is shown that this method gives is not consistent in the amount of bandwidth expansion across frequencies.

Appendix A Binomial Window

This appendix derives the properties of a binomial window using a signal processing formalism. Start with a two-coefficient generating function defined by its z-transform,

$$g(z) = q + pz^{-1}. \tag{65}$$

The N-fold product of $g(z)$ is

$$\begin{aligned} g_N(z) &= \underbrace{g(z) \dots g(z)}_{N} \\ &= \sum_{k=0}^{N} g_k z^{-k}, \end{aligned} \tag{66}$$

where the coefficient g_k is given by

$$g_k = \binom{N}{k} p^k q^{N-k}, \tag{67}$$

and where the combinatorial term is defined as

$$\binom{N}{k} = \frac{N!}{(N-k)!\,k!}. \tag{68}$$

If p is interpreted as a probability and q is set to $1 - p$, the coefficients g_k are the probability masses of a binomial distribution [17] with mean Np and standard deviation $\sqrt{Npq}$.

If $p = q$, the binomial coefficients are symmetrical, i.e., $g_{N-k} = g_k$. Furthermore, if N is even (giving an odd number of coefficients), the binomial coefficients have a unique maximum value at $k = N/2$. Let $p = q = 1/2$ and $M = N/2$,

$$g_k = \frac{1}{2^{2M}} \binom{2M}{k}. \tag{69}$$

The first factor normalizes the sum of the g_k to unity. The generating function $g(z)$ for $p = q = 1/2$ has a Fourier transform

$$G(\omega) = e^{-j\omega/2} \cos(\omega/2), \tag{70}$$

and the Fourier transform corresponding to $g_N(z)$ is

$$G_N(\omega) = e^{-j\omega M} \big(\cos(\omega/2)\big)^{2M}. \tag{71}$$

A normalized binomial window, symmetric about $k = 0$, can be generated as

$$w[k] = \begin{cases} \dfrac{\binom{2M}{M+k}}{\binom{2M}{M}}, & -M \leq k \leq M, \\ 0, & \text{elsewhere.} \end{cases} \tag{72}$$

The Fourier transform of this window is

$$W(\omega) = \frac{1}{\binom{2M}{M}} \big(\cos(\omega/2)\big)^{2M}. \tag{73}$$

As the number of terms in the binomial expansion is increased, the values of the binomial coefficients can be approximated by samples of the Gaussian distribution [17, 18].

Appendix B Bandwidth of a Continuous-Time Bandpass Filter

Consider the second-order bandpass filter

$$H(s) = \frac{s}{s^2 + (\Omega_o/Q)s + \Omega_o^2}. \tag{74}$$

This bandpass filter has zeros at $s = 0$ and $s = \infty$. The poles occur at

$$s = -\frac{\Omega_o}{2Q}(1 \pm j\sqrt{4Q^2 - 1}). \tag{75}$$

The poles are complex for $Q > 1/2$. These complex conjugate poles are at a distance Ω_o from the origin, with a real component equal to $-\Omega_o/(2Q)$. As Q increases beyond $1/2$, the poles move along a circle of radius Ω_o.

The frequency response of the bandpass-filter can be written as

$$H(j\Omega) = \frac{Q/\Omega_o}{1 + jQ(\Omega/\Omega_o - \Omega_o/\Omega)}. \tag{76}$$

The maximum of the magnitude-squared of this response occurs for $\Omega = \pm\Omega_o$. The 3 dB points occur when $|Q(\Omega/\Omega_o - \Omega_o/\Omega)| = 1$. Let $u = \Omega/\Omega_o$. The 3 dB points satisfy

$$Q(u - \frac{1}{u}) = \pm 1 \quad \text{or} \quad Qu^2 \mp u - Q = 0. \tag{77}$$

This gives the solutions,

$$u = \frac{1}{2Q}(\pm 1 \pm \sqrt{1 + 4Q^2}). \tag{78}$$

The equation has four solutions, two positive and two negative. Consider the positive solutions,

$$u = \frac{1}{2Q}(\sqrt{1 + 4Q^2} \pm 1). \tag{79}$$

The two solutions are related by $u_1 u_2 = 1$ and $u_2 - u_1 = 1/Q$. In terms of the 3 dB points Ω_l and Ω_u,

$$\Omega_l \Omega_u = \Omega_o^2 \quad \text{and} \quad \Omega_u - \Omega_l = \Omega_o/Q. \tag{80}$$

The difference $\Omega_u - \Omega_l$ is the 3 dB bandwidth of the resonance at frequency Ω_o. Note that the response has a resonance even when the poles are real — the response is zero at zero frequency, rising to a maximum at Ω_o, and falling to zero again for increasing frequency.

Appendix C Bandwidth of a Discrete-Time Bandpass Filter

A discrete-time bandpass filter will be formed from the continuous-time bandpass filter described in Appendix B.

The bilinear transformation relating a continuous-time filter to a discrete-time filter is

$$z = -\frac{s+a}{s-a} \quad \text{or} \quad s = a\frac{z-1}{z+1}. \tag{81}$$

Using the bilinear relationship, the 3 dB points in the response can be found for the discrete-time filter. The frequency response point at $s = j\Omega$ in the continuous-time system is mapped to the frequency point at $z = e^{j\omega}$, where

$$\omega = 2\tan^{-1}(\Omega/a). \tag{82}$$

Consider the identity (valid for $\theta_1\theta_2 > -1$)

$$\tan^{-1}(\theta_1) - \tan^{-1}(\theta_2) = \tan^{-1}\left(\frac{\theta_1 - \theta_2}{1+\theta_1\theta_2}\right). \tag{83}$$

The bandwidth for the resonance of the discrete-time filter is then

$$\begin{aligned} \omega_u - \omega_l &= 2\tan^{-1}(\Omega_u/a) - 2\tan^{-1}(\Omega_l/a) \\ &= 2\tan^{-1}\left(\frac{a(\Omega_u - \Omega_l)}{a^2 + \Omega_l\Omega_u}\right) \end{aligned} \tag{84}$$

This expression can be further simplified. The mapping of the poles is through the bilinear transformation. The poles at location s_1 and s_2 are mapped to poles at locations z_1 and z_2 with magnitudes r_1 and r_2. The product of the magnitudes satisfy (from the bilinear transform)

$$r_1 r_2 = \frac{s_1 s_2 + a(s_1 + s_2) + a^2}{s_1 s_2 - a(s_1 + s_2) + a^2}. \tag{85}$$

The left side of this equation is $z_1 z_2$ which is equal to $r_1 r_2$ since the roots are either both real or complex conjugates of each other. Similarly the terms $s_1 s_2$ and $s_1 + s_2$ are real, since s_1 and s_2 are either both real or complex conjugates of each other. But from the formulation of the continuous-time filter, $s_1 s_2 = \Omega_o^2$, and $s_1 + s_2 = -\Omega_o/Q$. Furthermore in Appendix B, it was found that the 3 dB points of the continuous-time filter satisfy,

$$\Omega_l\Omega_u = \Omega_o^2 \quad \text{and} \quad \Omega_u - \Omega_l = \Omega_o/Q. \tag{86}$$

Then Eq. (85) becomes

$$r_1 r_2 = \frac{\Omega_l \Omega_u - a(\Omega_u - \Omega_l) + a^2}{\Omega_l \Omega_u + a(\Omega_u - \Omega_l) + a^2}. \tag{87}$$

This can be rearranged to become

$$\frac{1 - r_1 r_2}{1 + r_1 r_2} = \frac{a(\Omega_u - \Omega_l)}{a^2 + \Omega_l \Omega_u}. \tag{88}$$

Finally, using this result in Eq. (84), the 3 dB bandwidth of the discrete-time filter can be written as follows,

$$\begin{aligned} \omega_u - \omega_l &= 2 \tan^{-1}\left(\frac{1 - r_1 r_2}{1 + r_1 r_2}\right) \\ &= \pi/2 - 2 \tan^{-1}(r_1 r_2). \end{aligned} \tag{89}$$

References

[1] P. Kabal, "Ill-Conditioning and Bandwidth Expansion in Linear Prediction of Speech", *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing* (Hong Kong), pp. I-824–I-827, April 2003 (doi:10.1109/ICASSP.2003.1198908).

[2] ITU-T Recommendation G.729 *Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP)*, Jan. 2007.

[3] 3GPP2 Document C.S0030-0, *Selectable Mode Vocoder Service Option of Wideband Spread Spectrum Communication Systems*, Version 3.0, Jan. 2004.

[4] S. Haykin, *Adaptive Filter Theory*, 5th ed., Prentice-Hall, 2014.

[5] ITU-T Recommendation P.830 *Methods for Objective and Subjective Assessment of Quality*, Feb. 1996.

[6] P. Kabal, *TSP Speech Database*, Technical Report, Dept. Electrical & Computer Engineering, McGill University, Version 2, Nov. 2018, (on-line at www-MMSP/Documents).

[7] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed., John Hopkins University Press, 2013.

[8] R. D. Gitlin, J. F. Hayes, and S. B. Weinstein, *Data Communications Principles*, Springer, 1992.

[9] R. Salami, C. Laflamme, J.-P. Adoul, A. Kataoka, S. Hayashi, T. Moriya, C. Lamblin, D. Massaloux, S. Proust, P. Kroon, and Y. Shoham, "Design and description of CS-ACELP: A toll quality 8 kb/s speech coder", *IEEE Trans. Speech and Audio Processing*, vol. 6, pp. 116–130, March 1998 (doi:10.1109/89.661471).

[10] B. S. Atal and M. R. Schroeder, "Predictive coding of speech signals and subjective error criteria", *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-27, pp. 247–254, June 1979 (doi:10.1109/TASSP.1979.1163237).

[11] P. Kabal, *Time Windows for Linear Prediction of Speech*, Technical Report, Dept. Electrical & Computer Engineering, McGill University, Version 2a, Nov. 2009, (on-line at www-MMSP/Documents).

[12] Y. Tohkura, F. Itakura, and S. Hashimoto, "Spectral smoothing technique in PARCOR speech analysis-synthesis", *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-26, pp. 587–596, Dec. 1978 (doi:10.1109/TASSP.1978.1163165).

[13] Y. Tohkura and F. Itakura, "Spectral sensitivity analysis of PARCOR parameters for speech data compression", *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-27, pp. 273–280, June 1979 (doi:10.1109/TASSP.1979.1163241).

[14] R. Viswanathan and J. Makhoul, "Quantization properties of transmission parameters in linear predictive systems", *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-23, pp. 309–321, June 1975 (doi:10.1109/TASSP.1975.1162675).

[15] K. K. Paliwal and W. B. Kleijn, "Quantization of LPC parameters", in *Speech Coding and Synthesis* (W. B. Kleijn and K. K. Paliwal, eds.), ch. 12, pp. 433–466, Elsevier, 1995.

[16] H. Tasaki, K. Shiraki, K. Tomita, and S. Takahashi, "Spectral postfilter design based on LSP transformation", *Proc. IEEE Workshop on Speech Coding* (Pocono Manor, PA), pp. 57–58, Sept. 1997 (doi:10.1109/SCFT.1997.623894).

[17] W. Feller, *An Introduction to Probability Theory and Its Applications, Vol. 1*, 3rd ed., Wiley, 1968.

[18] G. E. P. Box, J. S. Hunter, and W. G. Hunter, *Statistics for Experimenters : Design, Discovery, and Innovation*, 2nd ed. Wiley-Interscience, 2005.