SPEECH PROCESSING
Linear Predictive Coding (LPC)

Patrick A. Naylor
Spring Term 2008/9

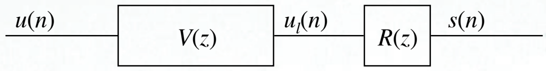
Imperial College London

PART 1

- This lecture studies one of the most important concepts underpinning many applications of speech processing, namely LPC
  - Concept of Linear Prediction
  - Derivation of Linear Prediction Equations
  - Autocorrelation method of LPC
  - Interpretation of LPC filter as a spectral whitener

Imperial College London

Concept of Linear Prediction



- $u(n)$ volume flow at the glottis
- $u_l(n)$ volume flow at the lips
- $s(n)$ pressure at the microphone
- $V(z) = \dfrac{Gz^{-p/2}}{1 - \sum_{j=1}^{p} a_j z^{-j}} = \dfrac{Gz^{-p/2}}{A(z)}$ vocal tract transfer function
- $R(z) = 1 - z^{-1}$ lip radiation model

- The aim of Linear Prediction Analysis (LPC) is to estimate $V(z)$ from the speech signal $s(n)$.

Notes

- We will neglect the pure delay term $z^{-\frac{1}{2}p}$ in the numerator of $V(z)$.

- 50% of the world puts a + sign in the denominator of $V(z)$ (this is almost essential when using MATLAB).

$$V(z) = \frac{Gz^{-p/2}}{1 - \sum_{j=1}^{p} a_j z^{-j}}$$

Preview … in straightforward terms

- Predict sample $s(n)$ from samples $s(n\text{-}1)$, $s(n\text{-}2)$, …, $s(n\text{-}p)$
- Consider prediction of 4 samples from their previous 2

$$
\begin{aligned}
s(2) &= a_1 s(1) + a_2 s(0) \\
s(3) &= a_1 s(2) + a_2 s(1) \\
s(4) &= a_1 s(3) + a_2 s(2) \\
s(5) &= a_1 s(4) + a_2 s(3)
\end{aligned}
$$

- This is an overdetermined system of simultaneous equations
  - If we try to predict only 2 samples then exact solution for the coefficients can be found
  - Otherwise we consider a least squares solution
- Call the prediction $\hat{s}(n)$

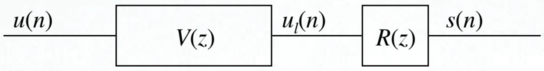
Imperial College London

- Important points to consider in determining the least squares solution
  - The frame $\{F\}$ of samples over which to solve
- Method of solution
  - Formulate the linear algebra problem in the form Xa=b
  - Solve by matrix inversion
- These issue are the main points to discuss in this talk

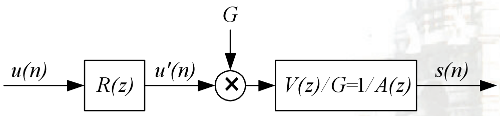
- What should p be to predict successfully:
  - A sinusoid?
  - Voiced speech?
  - Unvoiced speech?
  - The stock market?
- Think of LPC as capturing the harmonic content of a signal.
  - Anything not harmonic is unpredictable and gives a prediction error.

Imperial College London

Linearity



- We can reverse the order of $V(z)$ and $R(z)$ since both are linear and $V(z)$ doesn't change substantially during the impulse response of $R(z)$ or vice-versa:



Prediction Error

$$s(n) = Gu'(n) + \sum_{j=1}^{p} a_j s(n-j)$$

- If the vocal tract resonances have high gain, the second term will dominate:

$$s(n) \approx \sum_{j=1}^{p} a_j s(n-j)$$

- The right hand side of this expression is a *prediction* of $s(n)$ as a *linear sum* of past speech samples. Define the *prediction error* at sample n as

$$e(n) = s(n) - \sum_{j=1}^{p} a_j s(n-j) = s(n) - a_1 s(n-1) - a_2 s(n-2) - \ldots - a_p s(n-p)$$

- In terms of z-transforms

$$E(z) = S(z)A(z)$$

Error Minimization

- Given a frame of speech $\{F\}$, we would like to find the values a_i that minimize

$$Q_E = \sum_{n\in\{F\}} e^2(n) \qquad [1]$$

- To do so, we differentiate w.r.t each a_i

$$\frac{\partial Q_E}{\partial a_i} = \sum_{n\in\{F\}} \frac{\partial\left(e^2(n)\right)}{\partial a_i} = \sum_{n\in\{F\}} 2e(n)\frac{\partial e(n)}{\partial a_i} = -\sum_{n\in\{F\}} 2e(n)s(n-i)$$

Imperial College London

- The optimum values of a_i must satisfy p equations:

$$\sum_{n\in\{F\}} e(n)s(n-i) = 0 \quad \text{for} \quad i = 1,\ldots,p$$

$$\Rightarrow \sum_{n\in\{F\}}\left(s(n)s(n-i) - \sum_{j=1}^{p} a_j s(n-j)s(n-i)\right) = 0 \quad \text{for} \quad i = 1,\ldots,p$$

$$\Rightarrow \sum_{j=1}^{p} a_j \sum_{n\in\{F\}} s(n-j)s(n-i) = \sum_{n\in\{F\}} s(n)s(n-i)$$

$$\Rightarrow \sum_{j=1}^{p} \phi_{ij} a_j = \phi_{i0} \quad \text{where} \quad \phi_{ij} = \sum_{n\in\{F\}} s(n-i)s(n-j)$$

- which can be written in matrix form

$$\mathbf{\Phi a} = \mathbf{c} \quad \Rightarrow \quad \mathbf{a} = \mathbf{\Phi}^{-1}\mathbf{c} \quad \text{providing } \mathbf{\Phi}^{-1} \text{ exists}$$

  - the matrix $\mathbf{\Phi}$ is symmetric and positive semi-definite.

Imperial College London

Matrices with Special Properties

- Symmetric:

$$\phi_{ji} = \phi_{ij} \Leftrightarrow \Phi^T = \Phi$$

- Positive Definite: for a real symmetric matrix Φ

$$\sum_{i,j} x_i \phi_{ij} x_j > 0 \Leftrightarrow \mathbf{x}^T \Phi \mathbf{x} > 0 \quad \text{for any real vector } \mathbf{x} \neq 0$$

  - There exists a unique lower triangular matrix $\mathbf{L}$ such that $\Phi = \mathbf{L}\mathbf{L}^T$
    - Cholesky factorization

- Positive Semi-Definite: *as above but with ≥*

$$\sum_{i,j} x_i \phi_{ij} x_j \geq 0 \Leftrightarrow \mathbf{x}^T \Phi \mathbf{x} \geq 0 \quad \text{for any real vector } \mathbf{x} \neq 0$$

- Toeplitz: *constant diagonals*

$$\phi_{i+1,j+1} = \phi_{ij} = f(i-j)$$

Imperial College London

Inverting Matrices

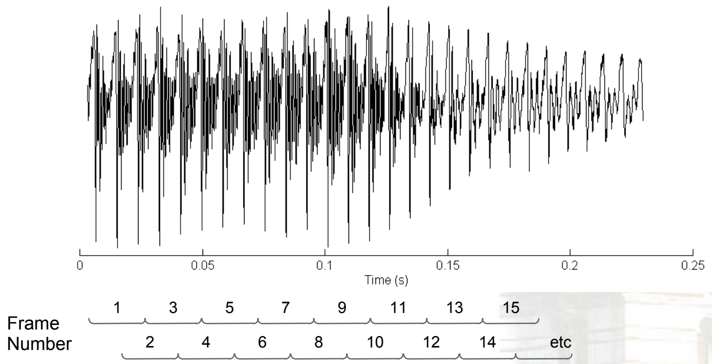
- Any special properties possessed by a matrix can be used when inverting it in order to:
  - reduce the computation time
  - improve the accuracy

Matrix ($p \times p$)	**Computation**
General	$\propto p^3$
Symmetric, +ve definite	$\propto \frac{1}{2} p^3$
Toeplitz, Symmetric, +ve definite	$\propto p^2$

Imperial College London

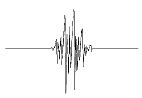
Frame-based Processing

- Consider frame-based processing of a speech signal
  - Extract a set of frames of the speech signal employing a tapered window of duration 20 – 30 ms typically overlapping by 50%



Autocorrelation LPC

- Take $\{F\}$ in equation [1] to be of infinite extent



$$\phi_{ij} = \sum_{n=-\infty}^{+\infty} s(n-i)s(n-j)$$

- Because of the symmetry and the infinite sum, we have

$$\phi_{ij} = \phi_{|i-j|,0} = R_{|i-j|}$$

  - where the sequence R_k is the autocorrelation of the windowed speech

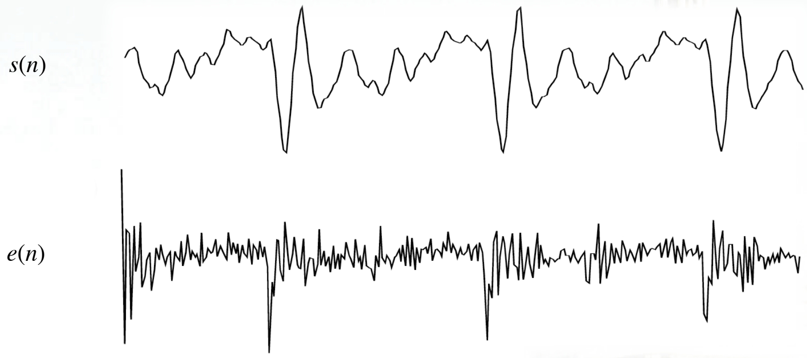
- The matrix $\mathbf{\Phi}$ is now Toeplitz (constant diagonals) and the equations

$$\mathbf{\Phi a} = \mathbf{c}$$

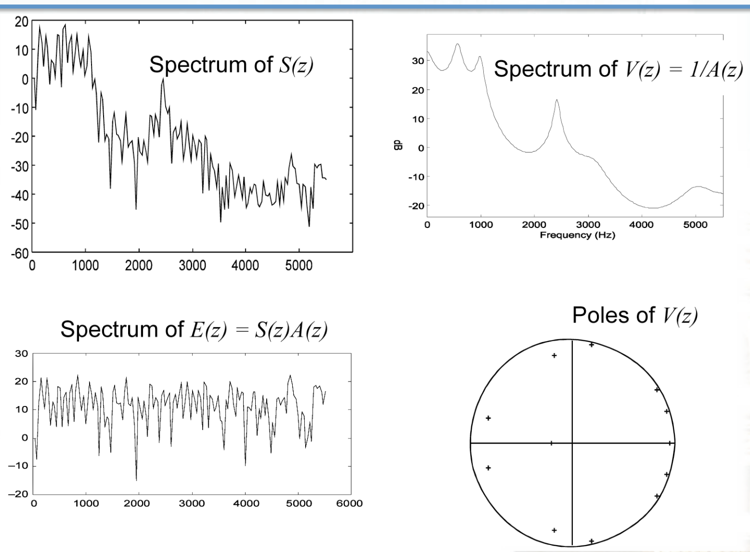
  are called the *Yule-Walker* equations.

- Inverting a symmetric, positive definite, Toeplitz $p \times p$ matrix takes $O(p^2)$ operations instead of the normal $O(p^3)$. Inversion procedure is known as the Levinson or Levinson-Durbin algorithm.

Autocorrelation LPC example: /ɑ/ from "father"



Resulting Spectra and Poles



Spectral Flatness

- Autocorrelation LPC finds the filter of the form

$$A(z) = 1 - a_1 z^{-1} - \ldots - a_p z^{-p}$$

that minimizes the energy of the prediction error.
- We will show that we can also interpret this in terms of flattening the spectrum of the error signal
- Define the normalized power spectrum of the prediction error signal $e(n)$

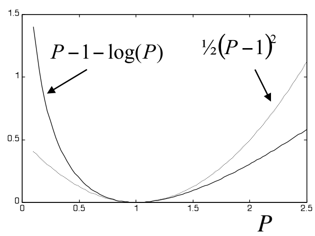
$$P_E(\omega) = \frac{\left|E\left(e^{j\omega}\right)\right|^2}{Q_E} \qquad Q_E = \sum e^2(n) = \frac{1}{2\pi}\int_{\omega=0}^{2\pi}\left|E\left(e^{j\omega}\right)\right|^2 d\omega$$

- where $E(z)$ is the z-transform of the signal and Q_E is the signal energy. The average value of P_E is equal to 1.

- We define the *spectral roughness* of the signal as:

$$R_E = \frac{1}{2\pi}\int_{\omega=0}^{2\pi} P_E(\omega) - 1 - \log\left(P_E(\omega)\right) d\omega$$

- R_E is similar to the variance of P_E since
  - the integrand is similar to $\frac{1}{2}(P_E - 1)^2$ where $\text{mean}(P_E)=1$.



- We can find an alternative expression for R_E

$$R_E = \frac{1}{2\pi}\int_{\omega=0}^{2\pi} P_E(\omega) - 1 - \log\left(P_E(\omega)\right) d\omega$$
$$= \frac{1}{2\pi}\int_{\omega=0}^{2\pi} -\log\left(P_E(\omega)\right) d\omega \quad \text{since} \quad \int P_E(\omega) d\omega = 1$$
$$= \log(Q_E) - \frac{1}{2\pi}\int_{\omega=0}^{2\pi} \log\left(\left|E\left(e^{j\omega}\right)\right|^2\right) d\omega$$

- Thus the spectral roughness of a signal equals the difference between its log energy and the average of its log energy spectrum.

- We know that $E(z) = S(z) \times A(z)$, hence

$$\log\left(\left|E\left(e^{j\omega}\right)\right|^2\right) = \log\left(\left|S\left(e^{j\omega}\right)\right|^2\right) + \log\left(\left|A\left(e^{j\omega}\right)\right|^2\right)$$

- Substituting this in the expression for R_E gives

$$R_E = \log(Q_E) - \frac{1}{2\pi}\int_{\omega=0}^{2\pi} \log\left(\left|E\left(e^{j\omega}\right)\right|^2\right) d\omega$$
$$= \log(Q_E) - \frac{1}{2\pi}\int_{\omega=0}^{2\pi} \log\left(\left|S\left(e^{j\omega}\right)\right|^2\right) d\omega - \frac{1}{2\pi}\int_{\omega=0}^{2\pi} \log\left(\left|A\left(e^{j\omega}\right)\right|^2\right) d\omega$$

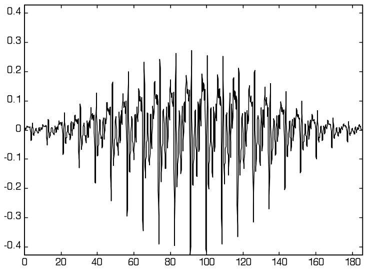
- We saw in the section on filter properties that the term involving A is zero since $a_0=1$ and all roots of A lie in the unit circle. Hence

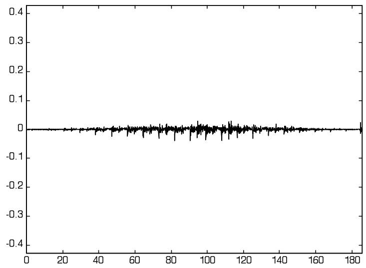
$$R_E = \log(Q_E) - \frac{1}{2\pi}\int_{\omega=0}^{2\pi} \log\left(\left|S\left(e^{j\omega}\right)\right|^2\right) d\omega$$

- The term involving S is independent of A. It follows that if A is chosen to minimize Q_E, it will also minimize R_E, the spectral roughness of $e(n)$. The filter $A(z)$ is a whitening filter because it makes the spectrum flatter.

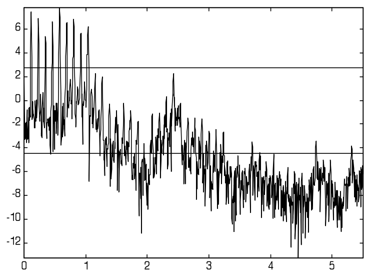
Example

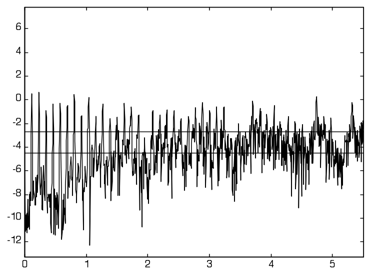
- These two graphs show a windowed speech signal, /ɑ/, and the error signal after filtering by $A(z)$





- These graphs show the log energy spectrum of each signal
  - The two horizontal lines on each graph are the mean value (same for both graphs) and the log of the total energy.
  - The spectral roughness is the difference between the two





PART 2

- In this lecture, we look at further elements under the general heading of Linear Prediction
  - Covariance method of LPC
  - Preemphasis
  - Closed Phase Covariance LPC
  - Alternative LPC parameter sets:
  - Pole positions
  - Reflection Coefficients
  - Log Area Ratios

Variants of LPC

We consider two variants of LPC analysis which differ only in their choice of speech frame, $\{F\}$

- Autocorrelation LPC Analysis
  - Requires a windowed signal
    - tradeoff between spectral resolution and time resolution
  - Requires >20 ms of data
  - Has a fast algorithm because Φ is toeplitz
  - Guarantees a stable filter $V(z)$
- Covariance LPC Analysis (Prony's method)
  - No windowing required
  - Gives infinite spectral resolution
  - Requires >2 ms of data
  - Slower algorithm because Φ is not Toeplitz
  - Sometimes gives an unstable filter $V(z)$

Covariance LPC

- Already seen that $\sum_{j=1}^{p} \phi_{ij} a_j = \phi_{i0}$ where $\phi_{ij} = \sum_{n \in \{F\}} s(n-i)s(n-j)$
- Now we chose $\{F\}$ to be a finite segment of speech:

  $\{F\} = s(n)$ for $0 \le n \le (N\text{-}1)$

  then we have:

$$\phi_{ij} = \sum_{n=0}^{N-1} s(n-i)s(n-j)$$

- The matrix Φ is still symmetric but is no longer Toeplitz
  - Since the matrix is not Toeplitz, the computation involved in inverting Φ is $\propto p^3$ rather than p^2 and so takes longer
- Covariance LPC generally gives better results than Autocorrelation LPC but is more sensitive to the precise position of the frame in relation to the vocal fold closures.

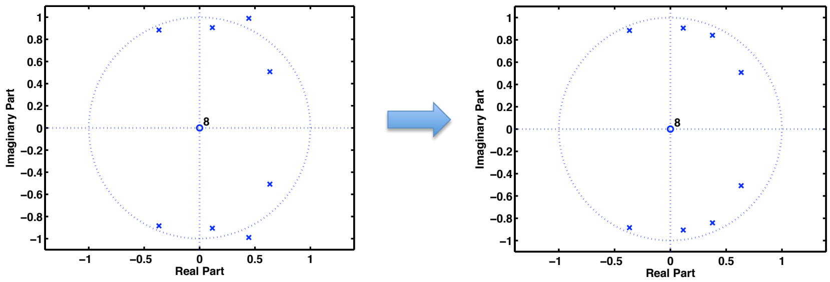
Recursive Computation

- The entire matrix Φ can be calculated recursively from its first row or column.

$$\begin{aligned}
\phi_{ij} &= \sum_{n=-1}^{N-2} s(n-i+1)s(n-j+1) \\
&= s(-i)s(-j) - s(N-i)s(N-j) + \sum_{n=0}^{N-1} s(n-i+1)s(n-j+1) \\
&= s(-i)s(-j) - s(N-i)s(N-j) + \phi_{i-1,j-1}
\end{aligned}$$

Unstable Poles

- Covariance LPC does not necessarily give a stable filter $V(z)$
  - (though it usually does).
- We can force stability by replacing an unstable pole at $z = p$ by a stable one at $z = 1/p^*$



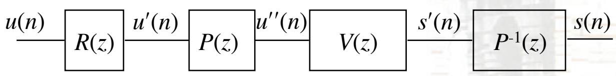
- As we have seen in the section on filter properties, reflecting a pole in the unit circle leaves the magnitude response unchanged except for multiplying by a constant (equal to the magnitude of the pole).
  - spectral flattening property of LPC is unaltered by this pole reflection.
- Discovering which poles lie outside the unit circle is quite expensive
  - this is a further computational disadvantage of covariance LPC.

Pre-emphasis

- The matrix Φ is always non-singular, but not necessarily by very much.
- A measure of how close a matrix is to being singular is given by its condition number
  - for a symmetric +ve definite matrix, this is the ratio of its largest to its smallest eigenvalue.
- For large p, the condition number of Φ tends to the ratio $S_{max}(\omega)/S_{min}(\omega)$.
- We can thus improve the numerical properties of the LPC analysis procedure by flattening the speech spectrum before calculating matrix Φ .
- For voiced speech, the input to $V(z)$ is $u_g'(n)$ whose spectrum falls off at high frequencies at around –6 dB/octave
  - This can be compensated with a 1st-order high-pass filter with a zero near $z=1$

$$P(z) = 1 - \alpha z^{-1}$$

- $P(z)$ is approximately a differentiator
- The normalised corner frequency of $P(z)$ is approximately (1–a)/2π
- This is typically placed in the range 0 to 150 Hz.
- From a spectral flatness point of view, the optimum value of a is ϕ_{10}/ϕ_{00} (obtained from autocorrelation LPC with $p = 1$).



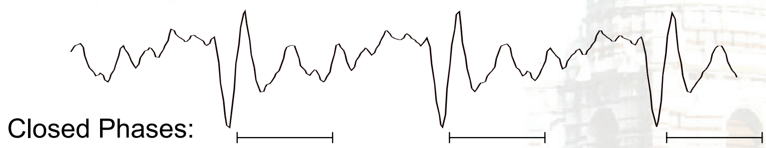
Closed-phase Covariance LPC

- We have already seen that $s(n) = Gu'(n) + \sum_{j=1}^{p} a_j s(n-j)$
- We have neglected the term $Gu'(n)$ because we don't know what it is and it is assumed to be much smaller than the second term
- If we knew when the vocal folds were closed, we could restrict $\{F\}$ to those particular intervals. We can estimate the times of vocal fold closure in two ways
  - Looking for spikes in the $e(n)$ signal
  - Using a Laryngograph (or Electroglottograph or EGG): this instrument measures the radio-frequency conductance across the larynx.
    - Conductance ∝ Vocal fold contact area.
    - Accurate but inconvenient.

Imperial College London

31

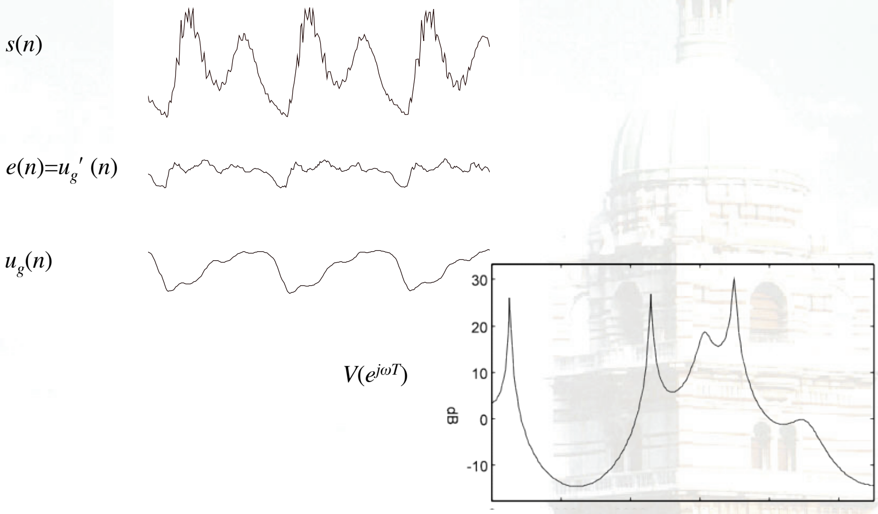
- In Closed-Phase LPC, we choose our analysis interval $\{F\}$ to consist of one or more closed phase intervals
  - (not necessarily contiguous).
- No preemphasis is necessary because the excitation now has a flat spectrum



Imperial College London

32

Closed Phase Analysis of /i/ from 'bee'



Alternative Parameter Sets

- The vocal tract filter is defined by $p+1$ parameters:

$$V(z) = \frac{G}{1 - \sum_{k=1}^{p} a_k z^{-k}}$$

- The LPC (or AR) coefficients a_k have some bad properties:
  – The frequency response is very sensitive to small changes in a_k
    - (such as quantizing errors in coding)
  – There is no easy way to verify that the filter is stable
  – Interpolating between the parameters that correspond to two different filters will not vary the frequency response smoothly from one to the other: stability is not even guaranteed.
- There are several alternative parameter sets that are equivalent to the a_k
  – most require G to be specified as well

Pole Positions

- We can factorize the denominator of $V(z)$ to give its poles:

$$1-\sum_{k=1}^{p} a_k z^{-k} = \prod_{k=1}^{p}\left(1-x_k z^{-1}\right)$$

- The polynomial roots x_k are either real or occur in complex conjugate pairs. $|x_k|$ must be <1 for stability
- Factorizing polynomials is computationally expensive
- The frequency response is sensitive to pole position errors near $|z|=1$.

Imperial College London

Reflection Coefficients

- Any all-pole filter is equivalent to a tube with p sections: this is characterised by p reflection coefficients (assuming $r_0=1$)
- We can convert between the reflection coefficients and the polynomial coefficients by using the formulae given earlier in the course
- Properties:
  - An all-pole filter is stable iff the corresponding reflection coefficients all lie between -1 and +1.
  - Interpolating between two of reflection coefficient sets will give a smoothly changing frequency response.
  - High coefficient sensitivity near ±1.
- The negative reflection coefficients are sometimes called the *PARCOR* coefficients (PARCOR = partial correlation)

Imperial College London

Log Area Ratios

- Log area ratios are derived from the lossless tube model

$$g_i = \log\left(\frac{A_{i+1}}{A_i}\right) = \log\left(\frac{1+r_i}{1-r_i}\right) \quad \Leftrightarrow \quad r_i = \frac{e^{g_i}-1}{e^{g_i}+1} = \tanh(g_i)$$

- Stability is guaranteed for any values of g_i.

Imperial College London 37

PART 3

- In this lecture, we at more alternative sets of LPC coefficients and their applications
  - Cepstral Coefficients
    - Relation to pole positions
    - Relation to LPC filter coefficients
  - Line Spectrum Frequencies
    - Relation to pole positions and to formant frequencies
  - Summary of LPC parameter sets

Imperial College London 38

Cepstral Coefficients

- Most *speech recognisers* describe the spectrum of speech sounds using *cepstral coefficients*
  - good at discriminating between different phonemes
  - fairly independent of each other
  - have approximately Gaussian distributions for a particular phoneme.
- Cepstrum is defined as inverse fourier transform of log spectrum
  - (periodic spectrum ⇒ discrete cepstrum)

$$c_n = \frac{1}{2\pi}\int_{\omega=-\pi}^{+\pi} \log\left(V(e^{j\omega})\right)e^{j\omega n}d\omega$$

  - Can be computed either from roots of the prediction filter polynomial
  - Can be computed alternatively from the coefficients of the prediction filter polynomial.

Computation from Roots x_k

- Define the cepstral coefficients c_n in terms of

$$C(z) = \sum_{n=-\infty}^{+\infty} c_n z^{-n} \quad \Rightarrow \quad c_n = \frac{1}{2\pi}\int_{\omega=-\pi}^{+\pi} C(e^{j\omega})e^{j\omega n}d\omega$$

- This is the standard inverse z-transform derived by taking the inverse Fourier transform of both sides of the first equation.
- By equating the Fourier transforms of the two expressions for c_n, we get

$$\begin{aligned} C(z) &= \log\left(V(z)\right) \\ &= \log\left(\frac{G}{A(z)}\right) = \log(G) - \log\left(A(z)\right) \end{aligned}$$

$$\text{where} \quad A(z) = 1 - \sum_{k=1}^{p} a_k z^{-k} = \prod_{k=1}^{p}\left(1 - x_k z^{-1}\right)$$

- Next, using the Taylor series $\log(1-y) = -\sum_{n=1}^{\infty}\frac{y^n}{n}$ for $|y|<1$

$$C(z) = \log(G) - \log(A(z))$$
$$= \log(G) - \sum_{k=1}^{p}\log\left(1 - x_k z^{-1}\right)$$
$$= \log(G) + \sum_{k=1}^{p}\sum_{n=1}^{\infty}\frac{x_k^n}{n}z^{-n}$$

- By collecting all the terms in z^{-n}, we obtain c_n in terms of x_k:

$$c_n = \begin{cases} 0 & \text{for } n<0 \\ \log(G) & \text{for } n=0 \\ \sum_{k=1}^{p}\frac{x_k^n}{n} & \text{for } n>0 \end{cases}$$

- Because $|x_k|<1$ the c_n decrease exponentially with n.

Imperial College London

41

Computation from Coefficients a_k

- Differentiating $C(z) = \log(G)$ - $\log(A(z))$ with respect to z:

$$C'(z) = \frac{-A'(z)}{A(z)} \Rightarrow A(z)C'(z) = -A'(z)$$
$$\Rightarrow A(z)zC'(z) = -zA'(z)$$

- Gives

$$\left(1-\sum_{k=1}^{p}a_k z^{-k}\right)\left(z\sum_{m=0}^{\infty}-mc_m z^{-(m+1)}\right) = -z\sum_{n=1}^{p}+na_n z^{-(n+1)}$$
$$\Rightarrow \left(1-\sum_{k=1}^{p}a_k z^{-k}\right)\left(\sum_{m=1}^{\infty}mc_m z^{-m}\right) = \sum_{n=1}^{p}na_n z^{-n}$$
$$\Rightarrow \sum_{n=1}^{\infty}nc_n z^{-n} - \sum_{k=1}^{p}\sum_{m=1}^{\infty}mc_m a_k z^{-(m+k)} = \sum_{n=1}^{p}na_n z^{-n}$$

- Replacing m by n-k (to make the z exponent uniform) gives:

$$\Rightarrow \sum_{n=1}^{\infty}nc_n z^{-n} = \sum_{n=1}^{p}na_n z^{-n} + \sum_{k=1}^{p}\sum_{n=k+1}^{\infty}(n-k)c_{(n-k)}a_k z^{-n}$$

Imperial College London

42

- Now take the coefficient of z^{-n} in the above equation noting that
  $n \geq k+1 \quad \Rightarrow \quad k \leq n-1$

$$nc_n = na_n + \sum_{k=1}^{\min(p,n-1)} (n-k)c_{(n-k)}a_k$$

$$\Rightarrow \quad c_n = a_n + \frac{1}{n}\sum_{k=1}^{\min(p,n-1)} (n-k)c_{(n-k)}a_k$$

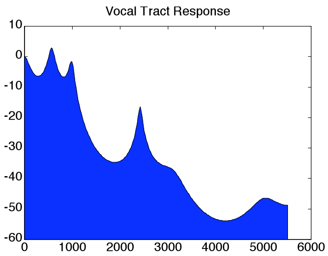
- Thus we have a recurrence relation to calculate the c_n from the a_k coefficients

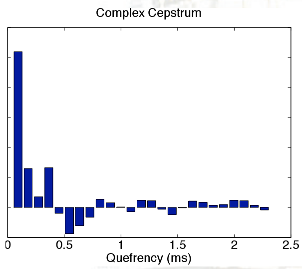
$$c_n = a_n + \frac{1}{n}\sum_{k=1}^{\min(p,n-1)} (n-k)c_{(n-k)}a_k$$

- From which

$$c_1 = a_1$$
$$c_2 = a_2 + \tfrac{1}{2}c_1a_1$$
$$c_3 = a_3 + \tfrac{1}{3}\left(2c_2a_1 + c_1a_2\right)$$
$$c_4 = a_4 + \tfrac{1}{4}\left(3c_3a_1 + 2c_2a_2 + c_1a_3\right)$$
$$c_5 = \cdots$$

- These coefficients are called the *complex cepstrum* coefficients
  - even though they are real
- The *cepstrum* coefficients use $\log|V|$ instead of $\log(V)$
  - half as big, except for c_0
- Note the cute names:
  - spectrum→cepstrum ; frequency→quefrency ; filter→lifter ; etc





Line Spectrum Frequencies (LSF)

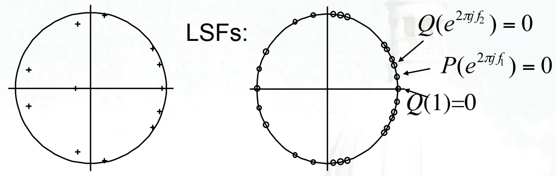
- Consider $A(z) = G \times V^{-1}(z) = 1 - \sum_{j=1}^{p} a_j z^{-j} = 1 - a_1 z^{-1} - a_2 z^{-2} - \ldots - a_p z^{-p}$
- We can form symmetric and antisymmetric polynomials:

$$P(z) = A(z) + z^{-(p+1)} A^*(z^{*-1})$$
$$= 1 - (a_1 + a_p) z^{-1} - (a_2 + a_{p-1}) z^{-2} - \ldots - (a_p + a_1) z^{-p} + z^{-(p+1)}$$
$$Q(z) = A(z) - z^{-(p+1)} A^*(z^{*-1})$$
$$= 1 - (a_1 - a_p) z^{-1} - (a_2 - a_{p-1}) z^{-2} - \ldots - (a_p - a_1) z^{-p} - z^{-(p+1)}$$

- $V(z)$ is **stable** if and only if the roots of $P(z)$ and $Q(z)$ **all lie on the unit circle** and **they are interleaved**.

Imperial College London 45

- Example



- If the roots of $P(z)$ are at $\exp(2\pi j f_i)$ for i=1, 3, … and the roots of $Q(z)$ are at $\exp(2\pi j f_i)$ for i=0, 2, … with $f_{i+1} > f_i \geq 0$
  - then the LSF frequencies are defined as $f_1, f_2, \ldots, f_p$.
- Note that it is always true that f_0=+1 and f_{p+1}=−1

E.g. $A(z) = 1 - 0.7z^{-1} + 0.5z^{-2}$ $\quad P(z) = 1 - 0.2z^{-1} - 0.2z^{-2} + z^{-3}$

$z^{-3} A^*(z^{*-1}) = 0.5z^{-1} - 0.7z^{-2} + z^{-3}$ $\quad Q(z) = 1 - 1.2z^{-1} + 1.2z^{-2} - z^{-3}$

Imperial College London 46

Proof that roots of $P(z)$ and $Q(z)$ lie on the unit circle

- Given $P(z)=0 \quad \Leftrightarrow \quad A(z)=-z^{-(p+1)}A^*(z^{*-1}) \quad \Leftrightarrow \quad H(z)=-1$

  $Q(z)=0 \quad \Leftrightarrow \quad A(z)=+z^{-(p+1)}A^*(z^{*-1}) \quad \Leftrightarrow \quad H(z)=+1$

  $$\text{where} \quad H(z)=\frac{A(z)}{z^{-(p+1)}A^*(z^{*-1})}=z\prod_{i=1}^{p}\frac{(1-x_i z^{-1})}{z^{-1}(1-x_i^* z)}=z\prod_{i=1}^{p}\frac{(z-x_i)}{(1-x_i^* z)}$$

- here the x_i are the roots of $A(z)=V^{-1}(z)$.
- Providing all the x_i lie inside the unit circle, the absolute values of the terms making up $H(z)$ are either all > 1 or else all < 1.

- Taking | | of a typical term:

$$\left|\frac{(z-x_i)}{(1-x_i^* z)}\right|>1 \quad \Leftrightarrow \quad \left|1-x_i^* z\right|<\left|z-x_i\right|$$

$$\Leftrightarrow \quad \left(1-x_i^* z\right)\left(1-x_i^* z\right)^*<\left(z-x_i\right)\left(z-x_i\right)^*$$

$$\Leftrightarrow \quad \left(1-x_i^* z\right)\left(1-x_i z^*\right)<\left(z-x_i\right)\left(z^*-x_i^*\right)$$

$$\Leftrightarrow \quad 1-x_i^* z-x_i z^*+x_i x_i^* z z^*<z z^*-x_i^* z-x_i z^*+x_i x_i^*$$

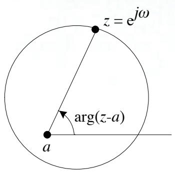
$$\Leftrightarrow \quad 1-x_i x_i^*-z z^*+x_i x_i^* z z^*<0$$

$$\Leftrightarrow \quad \left(1-\left|x_i\right|^2\right)\left(1-\left|z\right|^2\right)<0 \quad \Leftrightarrow \quad |z|>1 \quad \text{since each } \left|x_i\right|<1$$

- Thus each term is greater or less than 1 according to whether $|z|>1$ or $|z|<1$
- Hence $|H(z)|=1$ if and only if $|z|=1$ and so the roots of $P(z)$ and $Q(z)$ must lie on the unit circle.

Proof that the roots of $P(z)$ and $Q(z)$ are interleaved

- We want to find the values of $z = e^{j\omega}$ that make $H(z) = \pm 1$ or equivalently that make $\arg(H(z))$ = a multiple of π.
- If $z = e^{j\omega}$ then



$$\arg\left(H(e^{j\omega})\right) = \arg\left(e^{j(1-p)\omega}\prod_{i=1}^{p}\frac{(e^{j\omega}-x_i)}{(e^{-j\omega}-x_i^*)}\right)$$

$$= (1-p)\omega + \sum_{i=1}^{p}\left(\arg(e^{j\omega}-x_i) - \arg(e^{-j\omega}-x_i^*)\right)$$

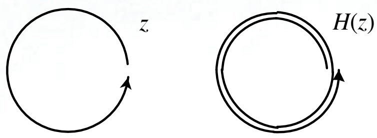
$$= (1-p)\omega + 2\sum_{i=1}^{p}\arg(e^{j\omega}-x_i)$$

- As ω goes from 0 to 2π, $\arg(z-a)$ changes monotonically by $+2\pi$ if $|a|<1$
- Therefore as ω goes from 0 to 2π, $\arg(H(e^{j\omega}))$ increases by

$$(1-p)\times 2\pi + 2p\times 2\pi = (1+p)\times 2\pi$$

Imperial College London

- Since $H(e^{j\omega})$ goes round the unit circle $(1+p)$ times, it must pass through each of the points +1 and –1 alternately $(1+p)$ times



- $\arg(H(z))$ varies most rapidly when z is near one of the x_i so the LSF frequencies will cluster near the formants

Imperial College London

Summary of LPC parameter sets

- Filter Coefficients: a_i
  - – Stability check difficult; Sensitive to errors; Cannot interpolate
- Pole Positions: x_i
  - + Stability check easy; Can interpolate but unordered.
  - – Hard to calculate; Sensitive to errors near $|x_i|=1$
- Reflection Coefficients: r_i
  - + Stability check easy; Can interpolate
  - – Sensitive to errors near ± 1
- Log Area Ratios: g_i
  - + Stability guaranteed; Can interpolate
- Cepstral Coefficients : c_i
  - + Good for speech recognition
  - – Stability check difficult
- Line Spectrum Frequencies: f_i
  - + Stability check easy; Can interpolate; Vary smoothly in time; Strongly correlated $\Rightarrow$ better coding; Related to spectral peaks (formants).
  - – Awkward to calculate

Imperial College London

51