Helsinki University of Technology Laboratory of Acoustics and Audio Signal Processing
Espoo 2004 Report 71

LINEAR PREDICTIVE MODELLING OF SPEECH - CONSTRAINTS AND LINE SPECTRUM PAIR DECOMPOSITION

Tom Bäckström

Dissertation for the degree of Doctor of Science in Technology to be presented with due permission for public examination and debate in Auditorium S4, Department of Electrical and Communications Engineering, Helsinki University of Technology, Espoo, Finland, on the 5th of March, 2004, at 12 o'clock noon.

Helsinki University of Technology
Department of Electrical and Communications Engineering
Laboratory of Acoustics and Audio Signal Processing

Teknillinen korkeakoulu
Sähkö- ja tietoliikennetekniikan osasto
Akustiikan ja äänenkäsittelytekniikan laboratorio

Helsinki University of Technology
Laboratory of Acoustics and Audio Signal Processing
P.O. Box 3000
FIN-02015 HUT
Tel. +358 9 4511
Fax +358 9 460 224
E-mail lea.soderman@hut.fi

ISBN 951-22-6946-5
ISSN 1456-6303

Otamedia Oy
Espoo, Finland 2004

Let the floor be the limit!

HELSINKI UNIVERSITY OF TECHNOLOGY
P.O. BOX 1000, FIN-02015 HUT
http://www.hut.fi

ABSTRACT OF DOCTORAL DISSERTATION

Author Tom Bäckström

Name of the dissertation

Linear predictive modelling for speech - Constraints and Line Spectrum Pair Decomposition

Date of manuscript 6.2.2004 | Date of the dissertation 5.3.2004

☐ Monograph ☑ Article dissertation (summary + original articles)

Department Electrical and Communications Engineering
Laboratory Laboratory Acoustics and Audio Signal Processing
Field of research Speech processing
Opponent(s) Associate Professor Peter Händel
Supervisor Professor Paavo Alku
(Instructor)

Abstract

In an exploration of the spectral modelling of speech, this thesis presents theory and applications of constrained linear predictive (LP) models. Spectral models are essential in many applications of speech technology, such as speech coding, synthesis and recognition. At present, the prevailing approach in speech spectral modelling is linear prediction. In speech coding, spectral models obtained by LP are typically quantised using a polynomial transform called the Line Spectrum Pair (LSP) decomposition. An inherent drawback of conventional LP is its inability to include speech specific a priori information in the modelling process.

This thesis, in contrast, presents different constraints applied to LP models, which are then shown to have relevant properties with respect to root loci of the model in its all-pole form. Namely, we show that LSP polynomials correspond to time domain constraints that force the roots of the model to the unit circle. Furthermore, this result is used in the development of advanced spectral models of speech that are represented by stable all-pole filters.

Moreover, the theoretical results also include a generic framework for constrained linear predictive models in matrix notation. For these models, we derive sufficient criteria for stability of their all-pole form. Such models can be used to include a priori information in the generation of any application specific, linear predictive model. As a side result, we present a matrix decomposition rule for Toeplitz and Hankel matrices.

Keywords linear prediction, line spectrum pair, minimum-phase property, speech modelling

UDC 534.78:004.934 | Number of pages 132

ISBN (printed) 951-22-6946-5 | ISBN (pdf) 951-22-6947-3

ISBN (others) | ISSN 1456-6303

Publisher Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing

Print distribution Report 71 / HUT, Laboratory of Acoustics and Audio Signal Processing, Espoo, Finland

☑ The dissertation can be read at http://lib.hut.fi/Diss/

Acknowledgements

In the year 2000, while I was occupied with a different subject, at the Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology (HUT), my mentor and supervisor, Professor Paavo Alku presented to me a mathematical problem that he had discovered. Instantly, I became intrigued and excited. A year later, when we had managed to solve that problem, it was obvious to us that this theme, the Line Spectrum Pair (LSP) polynomials, would form the core of my doctoral thesis. How true that was! We found that from the LSP-theory, new results just kept falling into our laps! It became a golden egg. For that golden egg, the unfailing support, vast expertise and friendship, I thank Professor Paavo Alku.

The next important step was the presentation of our results to Professor Bastiaan W. Kleijn. His contribution opened up a completely new angle, a much better toolkit for handling the mathematics involved. I will be forever grateful.

With regard to the mathematical content of this work, I have received support and advice principally from two colleagues, Tuomas Paatero and Carlo Magi. The most helpful contribution of Tuomas was his support of my ideas and his checking of my calculations. He always had some time for my problems. Later on, Carlo became the closest thing to a professional partner by being the only one with whom I could discuss mathematical details on an equal basis.

In the maintenance of proper language, Renée Brinker has been of invaluable help. She has patiently and meticulously corrected the language of all the articles in this work as well as the body of this thesis. Even though she has not enjoyed a mathematical education, she often seemed to understand my thoughts better than I did myself.

While the above mentioned persons have been the most important colleagues in a strictly professional sense, there are many colleagues that have been of

vital importance in daily and administrative matters. Many of these have become friends. It is impossible to list them all, but they include my roommate Henkka, research partners Laura and Mairas, colleagues Hanna, Ville and Mara, our secretary Lea Söderman, Professors Matti Karjalainen, Vesa Välimäki and Unto Laine, the merry people in the plains of Siberia and all others at our laboratory. I do not think any other place could have provided me with the same level of flexibility and richness of sources for inspiration.

For me, it is very important to maintain a balance among the different areas of my life. My involvement in the chamber choir Kampin Laulu has, in this sense, been fulfilling. There, I have enjoyed a colourful and varied social and artistic interaction with, among others, Mari, Minna, Eeppi, Saija, Juha, Harri, Ville, Urkki, William and the "brotherhood of monks". My sincere thanks to you all.

Another strong source of balance and a topic that I feel very much defines who I am today, is systems thinking. Under the supervision of Professors Esa Saarinen and Raimo Hämäläinen, I have participated in courses, seminars and other forums of discussion, where we have delved deeply into systems thinking. The climax of this interaction was the writing of the book "Systeemiäly" (in English, "Systems intelligence") in collaboration with an inspiring group of people. While these studies have not directly contributed to this thesis, I believe that the strengthening of my general awareness and the activation of thinking processes provided by studies in systems thinking, has benefited all aspects of my life, including this thesis. The thank you presented here, is minute compared to the gratitude I feel toward the "Systems intelligence"-group.

In the economic sphere, I am obliged to the Language Technology Graduate School, where I have been a student and which has provided me with the main part of my livelihood. I have received additional support in the form of grants from Jenny ja Antti Wihurin Rahasto, Stiftelsen för Teknikens Främjande and Oskar Öflunds Stiftelse.

Last but certainly not least, I would like to thank my parents and brothers with their families, of their lasting love and support.

Finally, I fear that prefaces like this one all too often seems filled with glib phrases of gratitude and are regularly riddled with clichés better left unwritten. In spite of my determination to avoid these pitfalls, I have not, alas, managed to wriggle myself out of that trap. There are just too many people that I am indebted to and too many favours that I can never return. I am left with no other option than to bow my head in gratitude toward all those mentioned above and others whom I have forgotten. Thank you.

Table of Contents

List of Abbreviations

AR	Auto-Regressive (model)
ARMA	Auto-Regressive Moving-Average (model)
CELP	Code-Excited Linear Prediction
CLP	Constrained Linear Prediction
DAP	Discrete All-Pole (modelling)
DFT	Discrete Fourier Transform
DID	Decimation In Degree
DSP	Digital Signal Processing
ETSI	European Telecommunications Standards Institute
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
FWLP	Frequency Weighted Linear Prediction
GSM	Global System of Mobile communications
HUT	Helsinki University of Technology
IEEE	Institute of Electrical and Electronics Engineers
ISP	Immittance Spectrum Pair
ITU	International Telecommunication Union
ITU-T	ITU Telecommunication Standardization Section
LAR	Log-Area Ratio
LMS	Least Mean Square
LP	Linear Prediction
LPC	Linear Predictive Coding
LPLE	Linear Prediction with Low-frequency Emphasis (model)
LSD	Log Spectral Distortion
LSF	Line Spectrum Frequency
LSP	Line Spectrum Pair
MA	Moving-Average (model)
NN	Neural Network

PDF	Probability Distribution Function
PLP	Perceptual Linear Prediction
SD	Spectral Distortion
SNR	Signal to Noise Ratio
VQ	Vector Quantisation
WLP	Warped Linear Prediction
WLSP	Weighted-sum Line Spectrum Pair (model)

List of Symbols

$\mathbf{a}$, $A(z)$, a_i	Predictor vector, its Z-transform and elements
$\mathbb{C}$	Space of complex values
$D(z, \lambda)$	Z-transform of the WLSP model
δ_i	Kronecker delta function
e_n	Estimation error or residual
$E[\cdot]$	Expected value operator
Γ	Reflection coefficient
m	Model order
$O(\cdot)$	Measure of complexity in relation to size of problem
$P(z)$	Symmetric LSP polynomial
$Q(z)$	Antisymmetric LSP polynomial
$\mathbb{R}$	Space of real values
$\mathbf{R}$, $R(i)$	Autocorrelation matrix (symmetric Toeplitz) and its elements
σ^2	Squared prediction error or residual energy
$u(t)$	Volume velocity
$W(\cdot)$	Spectral weighting function
$\mathbf{x}$, x_n	Input signal vector and its elements
$\#$	Reversal of (vector or matrix) rows operator (superscript)

Chapter 1

Introduction

One ought, everyday, to hear a song,
read a fine poem, and, if possible,
to speak a few reasonable words.

Johann Wolfgang von Goethe

Words and language are at the very heart of human life and intelligence. They are not all there is to humanity, but a great part of it. Specifically, speech is the primary mode of communication that uses words and the accoutrements of language. It is a way of sharing facts, thoughts and emotions, of transferring human intelligence, of information, from one person to another via sound [103, 41, 146].

This thesis goes to the core of speech, to the properties of the sound signal that we perceive as speech. Specifically, we will study spectral models of the speech sound signal. Before going to the main topic, we will, however, present some necessary preliminaries of speech production (Section 1.1) and matrix algebra for speech modelling (Chapter 2). Together, these two topics form the theoretical basis for the spectral models to be discussed in Chapter 3. Finally, in Chapter 4, we will study the Line Spectrum Pair Decomposition frequently used in the representation of linear predictive models.

1.1 Speech Production

Figure 1.1 illustrates the speech organs. Speech production is initiated by the lungs, which generate air pressure that flows through the trachea, vocal folds, pharynx, oral and nasal cavities. Cavities above the vocal folds are collectively called the vocal tract. The actual speech sound can be, roughly speaking, created by two different strategies. Firstly, for voiced speech (e.g. vowels /a/, /o/ and /i/, and nasals /m/ and /n/), the flow of air sets the vocal folds in an oscillating motion, periodically inhibiting the airflow for a short interval. The excitation of voiced speech, the glottal volume velocity waveform, is named after the orifice between the vocal folds, the glottis. Consequently, voiced speech sounds consist of a strong periodic component rich in harmonics. Secondly, for unvoiced speech, airflow is constricted (e.g. fricatives /f/, /s/ and /h/) or completely stopped for a short interval (e.g. stops /t/, /p/ and /k/). Therefore, unvoiced speech is of either noise-like or impulsive-like characteristics, without harmonic structure [103, 41].

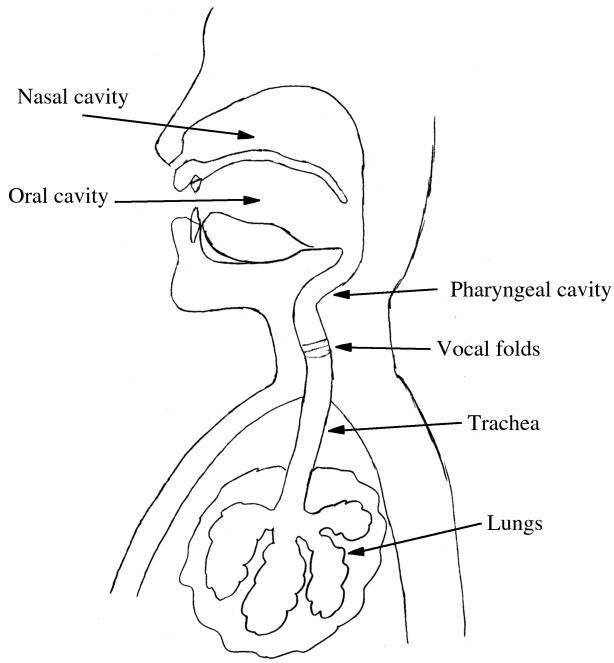

Figure 1.1: Illustration of speech organs.

The smallest unit of distinctive speech sounds is called a phoneme. Of the phonemes in standard English prose, vowels and diphthongs form approximately 38%, voiced consonants 40% and unvoiced consonants 22% [41]. It

is therefore obvious that voiced phonemes form the main part of average English (and most Western languages) and modelling voiced phonemes is, as a result, of primary importance. Consequently, it is understandable that the focus in the development of spectral models of speech, is in material represented by voiced phonemes. For this purpose, this thesis will concentrate on voiced phonemes and in particular, vowels only.

A typical waveform generated in the vocal folds (voiced phoneme) is depicted in Figure 1.2a. It resembles a sinusoidal wave cut off at the bottom and the spectrum will contain a great deal of energy in the harmonics of the fundamental frequency (see Figure 1.2b).

Like any acoustic cavity, the vocal tract has resonances that attenuate and amplify different frequency regions. These resonances are called the formants and can be modified by movements of the vocal organs such as tongue, lips and pharynx [121]. The frequency domain locations of formants are important with respect to recognition of vowels and, consequently, accurate modelling of formants is of primary interest in the spectral modelling of speech. Especially important are the two first formants, which lie mainly in the 200 to 1400 Hz range (F1), and 500 to 2500 Hz range (F2), respectively [41]. A typical formant structure of a vowel is illustrated in Figure 1.2e.

The emitted speech sound is then the combination of the excitation process (the glottal volume velocity waveform) and the filtering process (vocal tract effect) as described above. A typical vowel is depicted in Figure 1.2c-d, such that the harmonic structure generated by the vocal folds and the formant structure, imposed by the vocal tract, are clearly visible in Figure 1.2d.

1.2 Source-Filter Modelling

The theoretical basis widely used for speech modelling is the source-filter model [40]. This model is based on the assumption that speech can be modelled in independent parts, namely, the source and the filter. This approach is schematised in Figure 1.3. The above assumption has two main flaws; namely, it assumes that there are no sub-glottal resonances and worse, it assumes that vocal tract resonances and vocal fold oscillations have no interaction. In practise, however, because the error introduced by these assumptions is small, source-filter modelling yields good results. Furthermore, in this work, source-filter modelling will serve as a general speech acoustical motivation for models presented.

In the digital implementations of source-filter models, the prevalent tech-

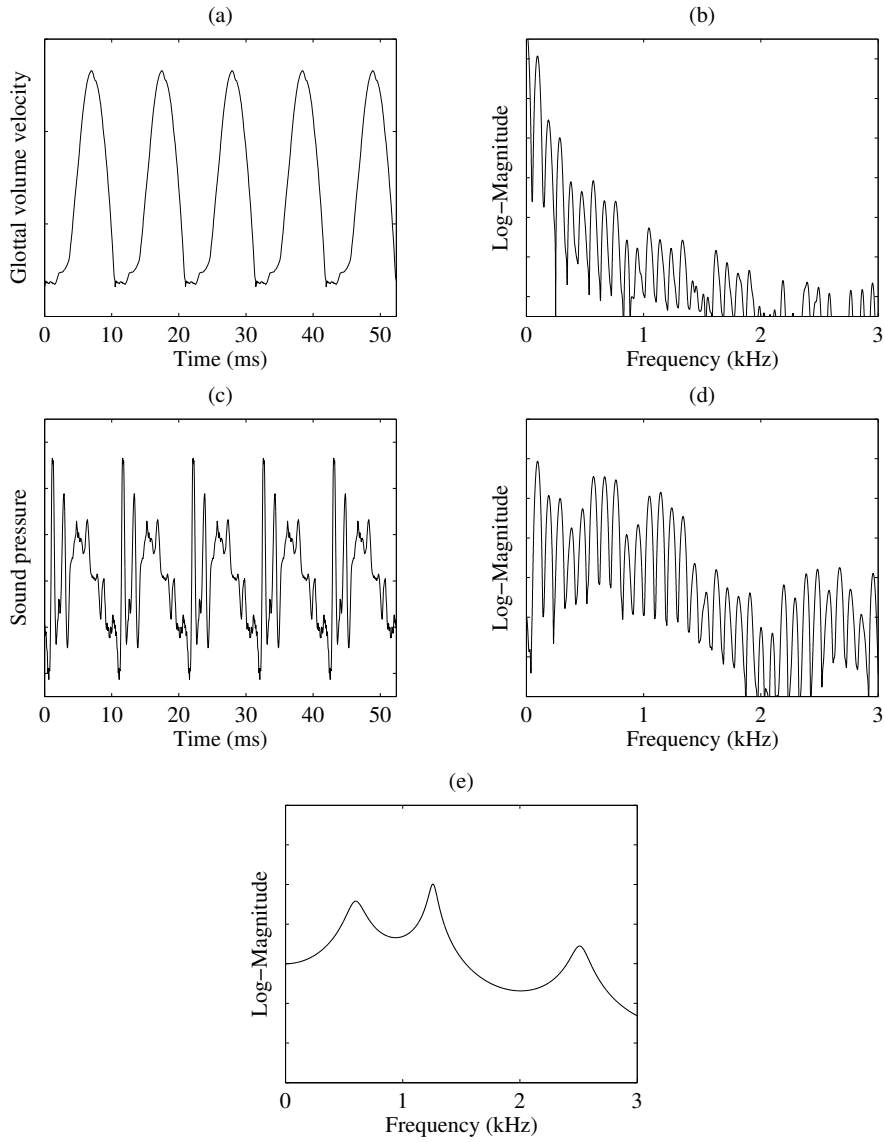

Figure 1.2: Signals in the source-filter model: (a) Glottal flow, (b) magnitude spectrum of the glottal flow, (c) speech signal, (d) spectrum of the speech signal and (e) filtering function (transfer function) of the vocal tract. Signals were generated by concatenating a single period of a real speech signal and the corresponding glottal flow period obtained by inverse filtering.

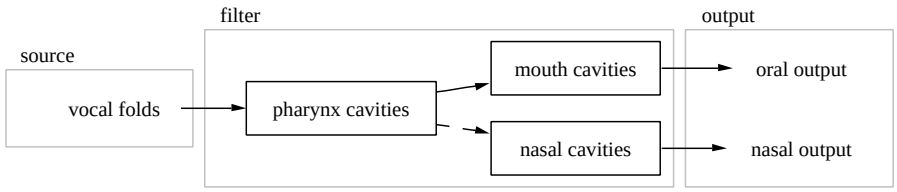

Figure 1.3: Source-filter model of speech production (adapted from Fant [40]).

nique is all-pole modelling or linear predictive modelling [91, 119]. With this method, we aim to model the filtering effects of the speech production mechanisms, with a parametric model, obtained by linear prediction that takes the source signal as input. The all-pole model is theoretically valid, since it can be derived from a tube model of the vocal tract [40, 146, 103]. The tube model is discussed in more detail in Section 1.2.1. While we aim to model only the filtering function, in practise, we end up modelling properties of the glottal source and the radiation effect as well [41, 119]. Thus, when we model a speech signal, the residual will become an impulse train with additive noise, instead of a train of glottal pulses. Yet, since the model performs well in many applications, especially speech coding applications, this difference is not an impediment [118]. The linear predictive model is considered in more detail in Chapter 3.

1.2.1 Acoustic Tube Model

The vocal tract is basically a bent tube of varying diameter, making it a rather complex acoustical system [121]. However, if we ignore the bend, we can make a simple model that explains the formants and resonances quite well. The tube model attempts to model the filter part of the source-filter model described in Section 1.2.

First attempts to directly compute an acoustic tube model of the vocal tract from the speech waveform were presented by Atal [8]. He demonstrated that the formant frequencies and bandwidths are sufficient to uniquely determine the tube model parameters and that this model is always realisable as a transfer function with M poles when the number of cylindrical tubes is M [10].

When using the tube model, we make the following assumptions [95]:

1. The vocal tract consists of M tube sections of equal length but varying

diameter.

2. The length of each section is small enough that sound propagation can be treated as a plane wave.

3. The sections are rigid and internal losses due to wall vibrations etc. can be ignored.

4. The model is linear and uncoupled from the glottis.

5. Interaction with the nasal tract is ignored.

In addition, we make the normal assumptions of wave propagation [95]. In other words, we model the vocal tract by a set of short cylindrical tube segments (see Fig. 1.4).

As the wave travels through the tubes, a portion of it will reflect in the opposite direction at each junction. The proportion of the volume velocity $u_i(t)$ reflected at the junction between segments i and $i+1$ is defined as the reflection coefficient Γ_i. Consequently, from the wave travelling to the right $u_i^+(t)$, the portion reflected to the left at then junction between sections i and $i+1$, is $\Gamma_i u_i^+(t)$. The portion reflected to the right from the wave travelling to the left $u_i^-(t)$ is correspondingly $\Gamma_i u_{i+1}^-(t+1)$ [103, 95]. The signal flow is illustrated in Figure 1.5.

The reflection coefficients depend on the cross-sectional area of the adjacent tubes. If we define the cross-sectional area of the i'th tube as A_i, we obtain the reflection coefficient as [95]

$$\Gamma_i = \frac{A_{i-1} - A_i}{A_{i-1} + A_i}. \tag{1.1}$$

By choosing a suitable tube section length and cross section areas, we obtain surprisingly accurate estimates for formant locations of different vowels. Consequently, the tube model is a well-warranted model of the vocal tract and we can expect that further models based on the tube model will perform as dependably.

1.3 Vocal Tract Estimation and Modelling

The Source-Filter model and the acoustic tube model for speech are used overwhelmingly frequently, due to their ease of use and their reliable results. However, there exist other models for the speech production system as well.

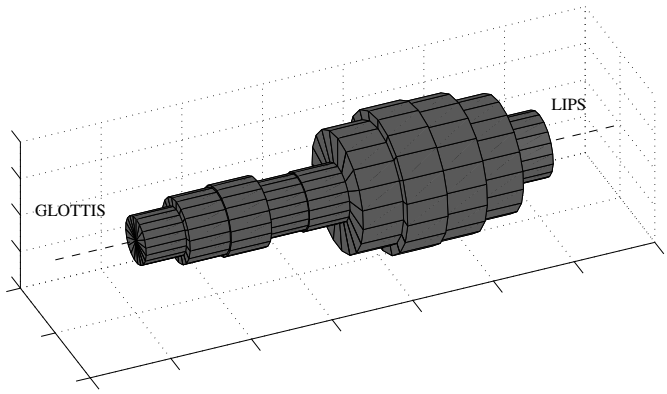

Figure 1.4: Illustration of the tube model (adapted from Markel and Gray, 1980).

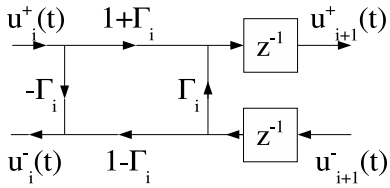

Figure 1.5: Signal flow of the tube model for the vocal tract at the junction of tube sections i and $i+1$. The volume velocities in the right and the left direction are denoted by $u_i^+(t)$ and $u_i^-(t)$, respectively, the reflection coefficient is denoted by Γ_i and unit delay by z^{-1}. (Adapted from O'Shaughnessy, 1987)

Unfortunately, these suffer quite consistently from high computational complexity or requirements of specific *a priori* knowledge of the voice, and are therefore not useful in engineering applications. More suitable application would arise in, for example, medical applications, where the speech production system is studied from a clinical viewpoint.

An analysis of the effect of losses in the tube model is presented in [126]. This model takes into account the internal losses due to, for example, wall vibrations in the walls of the vocal tract, thus providing a slightly more accurate model. It is, however, intended for an analysis of formants and speech transitions, in contrast to telecommunications applications.

In [22], a vocal tract modelling approach employing neural networks was presented. It uses the true glottal pulse as input and estimates the vocal tract transfer function by learning from the output sound. Obviously, in telecommunications and similar applications, the glottal pulse is not available and this model can be used for analysis applications only.

For medical applications, several estimation and modelling approaches are proposed, such as, [140, 147].

Chapter 2

Matrix Algebra for Linear Prediction

Before proceeding to the main topic of this thesis, it is necessary to present some mathematical formulae that may not be familiar to most readers. Namely, we will present matrix algebra of some classes of matrices; i.e. the Toeplitz, Vandermonde and convolution matrices, as well as the solution of related systems by the Levinson recursion and matrix decompositional techniques.

2.1 Toeplitz Matrices

On class of matrices which appears frequently in digital signal processing is the Toeplitz matrices. The characteristic property of these matrices is that each of their diagonals have common entries. In other words, an $n \times k$ Toeplitz matrix $\mathbf{T}$ is defined by $\mathbf{T}_{ij} = t_{j-i}$ or in matrix form

$$\mathbf{T} = \begin{bmatrix} t_0 & t_1 & t_2 & \dots & t_{k-1} \\ t_{-1} & t_0 & t_1 & \ddots & t_{k-2} \\ t_{-2} & t_{-1} & t_0 & \ddots & t_{k-3} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ t_{-n+1} & t_{-n+2} & t_{-n+3} & \dots & t_{k-n-2} \end{bmatrix}. \tag{2.1}$$

The Toeplitz matrices also belong to the more general class of persymmetric matrices, which are symmetric about their northeast-southwest diagonals [46, 60].

While the definition of Toeplitz matrices is straightforward, they have, as mathematical operators, proved to be rich in fruitful properties and to have a wide range of applications. Within the Toeplitz theory, much of the current theoretical interest lies in the special case of circulant (Toeplitz) matrices, since general Toeplitz matrices are in many ways equivalent to circulant matrices [50, 60]. Each row of a circulant matrix is obtained by cyclically shifting the previous row to the right.

In this work however, we shall, concentrate on symmetric Toeplitz matrices which are prevalent in digital signal processing. A symmetric matrix $\mathbf{A}$ has $\mathbf{A}_{ij} = \mathbf{A}_{ji}$ and therefore, a symmetric Toeplitz matrix will have $\mathbf{T}_{ij} = t_{|j-i|}$. Furthermore, we shall limit ourselves to real and positive definite matrices, since these are, in practise, perhaps the most common kind of matrices in speech processing. Positive definite matrices $\mathbf{A}$ are, by definition, such that $\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{x}^T\mathbf{A}\mathbf{x} = 0$ implies $||\mathbf{x}|| = 0$. The positive definite property is, in many signal processing applications, highly important, and therefore, tests for positive definiteness have been devised, such as [93].

2.1.1 Eigenvalues and Eigenvectors

The eigenvalues and eigenvectors of a matrix $\mathbf{A}$ are defined as the pair of scalar λ_i and vector $\mathbf{v}_i$, for which $\mathbf{A}\mathbf{v}_i = \lambda_i\mathbf{v}_i$ [46, 60]. Let superscript $^\#$ denote the reversal of rows operator. If $\mathbf{v}_i$ is an eigenvector of an $m \times m$ symmetric Toeplitz matrix $\mathbf{T}$ corresponding to eigenvalue λ_i; i.e. $\mathbf{T}\mathbf{v}_i = \lambda_i\mathbf{v}_i$, then it also holds that $\mathbf{T}\mathbf{v}_i^\# = \lambda_i\mathbf{v}_i^\#$. If the eigenvalues λ_i are distinct, then $\mathbf{v}_i$ and $\mathbf{v}_i^\#$ must be equal in direction, that is, $\mathbf{v}_i = \pm\mathbf{v}_i^\#$. Consequently, all eigenvalues are either symmetric or antisymmetric, that is, their elements have $v_j^{(i)} = v_{m-j}^{(i)}$. Furthermore, it can be shown that a symmetric Toeplitz matrix of size $m \times m$ has $m/2$ symmetric and antisymmetric eigenvectors for m even, and $(m+1)/2$ and $(m-1)/2$ symmetric and antisymmetric eigenvectors, respectively, for m odd [23, 92].

The eigenvalues of a Hermitian (and thus also the eigenvalues of a real symmetric) matrix are real [78]. Furthermore, the eigenvalues of a positive definite matrix are positive [46]. Since we will be discussing mostly positive definite, real and symmetric Toeplitz matrices, their eigenvalues will be positive and real.

Finding the eigenvalues of a given system is usually expensive and many algorithms have been presented for this task [65, 56, 66]. In our context, however, we will not need the explicit formulae or algorithms for eigenvectors

or eigenvalues.

2.2 Levinson-Durbin Recursion

In most applications where Toeplitz matrices appear, we are dealing with a problem such as

$$\mathbf{Ta} = \mathbf{b} \tag{2.2}$$

where $\mathbf{T}$ is an $n \times n$ Toeplitz matrix and $\mathbf{a}$ and $\mathbf{b}$ are vectors. The problem is to solve $\mathbf{a}$ when $\mathbf{T}$ and $\mathbf{b}$ are known. For the solution, straightforward application of Gaussian elimination is rather inefficient, with complexity $O(n^3)$, since it does not employ the strong structures present in the Toeplitz system.

A first improvement to Gaussian elimination is the Levinson recursion which can be applied to symmetric Toeplitz systems [89, 55, 46]. To illustrate the basics of the Levinson algorithm, first define the $p \times p$ principal sub-matrix $\mathbf{T}_p$ as the upper left block of $\mathbf{T}$. Further, assume that we have the order p solution $\mathbf{a}_p$ to equation

$$\begin{bmatrix} t_0 & t_1 & t_2 & \dots & t_p \\ t_1 & t_0 & t_1 & \ddots & t_{p-1} \\ t_2 & t_1 & t_0 & \ddots & t_{p-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ t_p & t_{p-1} & t_{p-2} & \dots & t_0 \end{bmatrix} \begin{bmatrix} 1 \\ a_1^{(p)} \\ \vdots \\ a_p^{(p)} \end{bmatrix} = \begin{bmatrix} \epsilon_p \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{2.3}$$

Extension of $\mathbf{a}_p$ with a zero yields

$$\begin{bmatrix} t_0 & t_1 & t_2 & \dots & t_{p+1} \\ t_1 & t_0 & t_1 & \ddots & t_p \\ t_2 & t_1 & t_0 & \ddots & t_{p-1} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ t_{p+1} & t_p & t_{p-1} & \dots & t_0 \end{bmatrix} \begin{bmatrix} 1 \\ a_1^{(p)} \\ \vdots \\ a_p^{(p)} \\ 0 \end{bmatrix} = \begin{bmatrix} \epsilon_p \\ 0 \\ \vdots \\ 0 \\ \eta_p \end{bmatrix} \tag{2.4}$$

where $\eta_p = \sum_{i=0}^{p} a_i^{(p)} t_{p-i+1}$ and $a_0^{(p)} = 1$. The salient step comes through the realisation that since $\mathbf{T}_p\mathbf{a}_p = \mathbf{u}_p$ where $\mathbf{u}_p = [\epsilon_p\, 0 \dots 0]^T$ and $\mathbf{T}_p$ is symmetric, we have $\mathbf{T}_p\mathbf{a}_p^{\#} = \mathbf{u}_p^{\#}$, where superscript # denotes reversal of

rows. By defining a reflection coefficient Γ_p, we obtain

$$\mathbf{T}_{p+1}\left(\begin{bmatrix} 1 \\ a_1^{(p)} \\ \vdots \\ a_p^{(p)} \\ 0 \end{bmatrix} + \Gamma_p \begin{bmatrix} 0 \\ a_p^{(p)} \\ a_{p-1}^{(p)} \\ \vdots \\ 1 \end{bmatrix}\right) = \left(\begin{bmatrix} \epsilon_p \\ 0 \\ \vdots \\ 0 \\ \eta_p \end{bmatrix} + \Gamma_p \begin{bmatrix} \eta_p \\ 0 \\ \vdots \\ 0 \\ \epsilon_p \end{bmatrix}\right). \tag{2.5}$$

Obviously, choosing Γ_p so that $\eta_p + \Gamma_p \epsilon_p = 0$ yields the order $p+1$ solution to Eq. 2.3 as

$$\mathbf{a}_{p+1} = \begin{bmatrix} \mathbf{a}_p \\ 0 \end{bmatrix} + \Gamma_p \begin{bmatrix} 0 \\ \mathbf{a}_p^{\#} \end{bmatrix}. \tag{2.6}$$

Consequently, with a suitable choice of initial values ($\mathbf{a}_0 = 1$), this procedure can be used to recursively solve equations of type Eq. 2.3. Furthermore, using the intermediate values $\mathbf{a}_p$, it is straightforward to solve arbitrary problems of type Eq. 2.2. The former algorithm, the solution of Eq. 2.3, is often called the Levinson-Durbin recursion and the latter, the solution of arbitrary equations of type Eq. 2.2, the Levinson recursion [89, 34, 55].

While the Levinson-Durbin recursion has the complexity of $O(n^2)$, it is possible to further improve the algorithm to reduce complexity by half. The algorithm, called the split Levinson-Durbin algorithm, uses a three term recursion instead of the two term recursion in Eq. 2.5. Then either the symmetric or antisymmetric part of two consecutive order solutions, $\mathbf{a}_{p-1}$ and $\mathbf{a}_p$ are used to obtain the next order solution $\mathbf{a}_{p+1}$ [33, 55].

Several other possibilities exist for the solution of Eq. 2.2, such as the Schur or Cholesky decomposition (see Section 2.4), but the split Levinson-Durbin is the most efficient [46, 33]. The others are sometimes preferred when decimal truncations cause numerical instability. Further improvements have been proposed (e.g. [79]), but since they suffer from stability problems [160], they are rarely used.

2.3 Vandermonde and Convolution Matrices

A class of matrices appearing in problems related to polynomial interpolation, control theory and digital signal processing, is the Vandermonde matrix

[46, 60, 73, 61]. The Vandermonde matrix is defined as

$$\mathbf{V} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ \alpha_1 & \alpha_2 & \alpha_3 & \dots & \alpha_m \\ \alpha_1^2 & \alpha_2^2 & \alpha_3^2 & \dots & \alpha_m^2 \\ \vdots & \vdots & \vdots & & \vdots \\ \alpha_1^n & \alpha_2^n & \alpha_3^n & \dots & \alpha_m^n \end{bmatrix}. \tag{2.7}$$

In other words, each column of the Vandermonde matrix $\mathbf{V}$ is a geometric progression of α_i. Sometimes the Vandermonde matrix is also defined in the transposed form of Eq. 2.7, that is, Vandermonde matrices are sometimes defined so that each *row* contains a geometric progression of α_i [74, 60].

If the α_i's are distinct ($\alpha_i \neq \alpha_j$ for $i \neq j$), then matrix $\mathbf{V}$ in Eq. 2.7 is of full rank [46]. However, if $\alpha_i \to \alpha_j$ then the matrix becomes rank deficient. The standard solution is the usage of confluent Vandermonde matrices, whereby we replace columns of overlapping α_i with the confluent Vandermonde matrix

$$[\mathbf{V}_c(\alpha_i)]_{jk} = \begin{cases} \binom{j-1}{k-1} \alpha_i^{j-k}, & \text{for } j \geq k \\ 0, & \text{otherwise.} \end{cases} \tag{2.8}$$

The formula of confluent Vandermonde matrix can readily be acquired by taking the limit of series $\alpha_i^k - \alpha_j^k$ scaled to unity when $\alpha_i \to \alpha_j$. In other words, the confluent Vandermonde matrix has derivatives $d \cdot /d\alpha_i$ of the overlapping geometric progression α_i^k [20]. For example, if we construct a confluent Vandermonde matrix with distinct geometric progression for α_2 and α_4, and overlapping progressions for α_1 and α_3 with multiplicity 2 and 3, respectively, we have

$$\mathbf{V} = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ \alpha_1 & 1 & \alpha_2 & \alpha_3 & 1 & 0 & \alpha_4 \\ \alpha_1^2 & 2\alpha_1 & \alpha_2^2 & \alpha_3^2 & 2\alpha_3 & 2 & \alpha_4^2 \\ \alpha_1^3 & 3\alpha_1^2 & \alpha_2^3 & \alpha_3^3 & 3\alpha_3^2 & 6\alpha_3 & \alpha_4^3 \\ \alpha_1^4 & 4\alpha_1^3 & \alpha_2^4 & \alpha_3^4 & 4\alpha_3^3 & 12\alpha_3^2 & \alpha_4^4 \\ \alpha_1^5 & 5\alpha_1^4 & \alpha_2^5 & \alpha_3^5 & 5\alpha_3^4 & 20\alpha_3^3 & \alpha_4^5 \\ \alpha_1^6 & 6\alpha_1^5 & \alpha_2^6 & \alpha_3^6 & 6\alpha_3^5 & 30\alpha_3^4 & \alpha_4^6 \end{bmatrix}. \tag{2.9}$$

The importance of Vandermonde matrices for digital signal processing becomes evident when we study their relation to convolution matrices. A

convolution matrix $\mathbf{A}$ for polynomial $A(z) = \sum_{i=0}^{m-1} a_i z^{-i}$ is defined as [98]

$$\mathbf{A}^T = \begin{bmatrix} a_0 & a_1 & \dots & a_{m-1} & 0 & 0 & \dots & 0 \\ 0 & a_0 & a_1 & \dots & a_{m-1} & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & a_0 & a_1 & \dots & a_{m-1} & 0 \\ 0 & 0 & \dots & 0 & a_0 & a_1 & \dots & a_{m-1} \end{bmatrix}. \tag{2.10}$$

The name, convolution matrix, springs from the fact that multiplication of a vector $\mathbf{b} = [b_0\, b_1\, \dots\, b_n]^T$ with $\mathbf{A}$, i.e. $\mathbf{A}^T\mathbf{b}$ corresponds to convolution of sequences b_n and a_n. This is, stated by means of polynomials, equivalent to multiplication of polynomial $B(z) = \sum_{i=0}^{m} b_i z^{-i}$ with $A(z)$, i.e. $A(z)B(z)$.

The intriguing connection to Vandermonde matrices can now be readily shown. If polynomial $A(z)$ has zeros α_i^{-1}, then from Eqs. 2.7 and 2.10 we obtain [98]

$$\mathbf{A}^T\mathbf{V} = 0. \tag{2.11}$$

Consequently, the (possibly confluent) Vandermonde matrix $\mathbf{V}$ is in the null-space of the convolution matrix $\mathbf{A}$. If matrix $\mathbf{A}$ is of size $m \times (m-n)$ and $\mathbf{V}$ of size $m \times n$ then matrix $\mathbf{V}$ is the complete null-space of $\mathbf{A}$.

2.4 Matrix Decompositions

Operations called matrix decompositions are factorisations of matrices that express them in some canonical form. For example, the singular value decomposition of a matrix $\mathbf{A}$ is

$$\mathbf{U}^T\mathbf{A}\mathbf{V} = \mathbf{D} \tag{2.12}$$

where matrix $\mathbf{D}$ is diagonal with eigenvalues λ_i of $\mathbf{A}$ as elements and matrices $\mathbf{U}$ and $\mathbf{V}$ are orthogonal [46, 60]. This kind of decomposition could be intuitively described as a rotation of $\mathbf{A}$ by $\mathbf{U}$ and $\mathbf{V}$, such that only the essential information, the eigenvalues, are extracted.

Two matrix decompositions often used in digital signal processing are the Schur and Cholesky decompositions [95, 46, 55]. For example, the Schur decomposition is used in the first GSM standards [1].

The Schur decomposition for matrix $\mathbf{A}$ is such that

$$\mathbf{Q}^H\mathbf{A}\mathbf{Q} = \mathbf{T} = \mathbf{D} + \mathbf{N} \tag{2.13}$$

where $\mathbf{Q}$ is unitary, $\mathbf{D}$ diagonal with eigenvalues of $\mathbf{A}$ on the diagonal, and $\mathbf{N}$ is strictly upper triangular [46].

Similarly, the Cholesky decomposition for a real, symmetric and positive definite matrix $\mathbf{A}$ is

$$\mathbf{A} = \mathbf{G}\mathbf{G}^T \tag{2.14}$$

where $\mathbf{G}$ is a unique lower triangular matrix [46].

Since inversion of a triangular matrix is easy, and efficient algorithms for finding the Schur and Cholesky decompositions exist, these decompositions can be used for the solution of problems of type Eq. 2.2 [55].

Chapter 3

Predictive Models of Speech

In this chapter, we will discuss the linear predictive (LP) method for spectral modelling of speech. It must be emphasised that while our motivation for using linear prediction lies within spectral modelling of speech, the possible fields of application are vast. For example, classic applications of linear predictive models outside speech processing include fields such as economy and geology [91, 116].

3.1 Linear Prediction

Let us assume that we have m past samples of a signal x_n in an interval $[n-m, n-1]$ and our task is to estimate a sample x_n. Moreover, let us restrict ourselves to the case where we construct our estimate as a linear combination of m past samples. The estimate $\hat{x}_n$ can then always be stated as [55, 95, 91]

$$\hat{x}_n = -\sum_{i=1}^{m} a_i x_{n-i} \tag{3.1}$$

where the minus sign has been added for convenience and a_i $(1 \leq i \leq m)$ are the model parameters. The error e_n between the estimated sample $\hat{x}_n$ and the true sample x_n is then

$$e_n = x_n - \hat{x}_n = x_n + \sum_{i=1}^{m} a_i x_{n-i} = \sum_{i=0}^{m} a_i x_{n-i} = \mathbf{a}^T \mathbf{x} \tag{3.2}$$

with constraint $a_0 = 1$ and where $\mathbf{a}^T = [a_0 \ \ldots \ a_m]$ and $\mathbf{x}^T = [x_n \ \ldots \ x_{n-m}]$.

Our objective is to find the best possible estimate, that is, the set of parameters a_i such that the error e_n is minimised. Historically, the optimisation criterion has been minimisation of the sum of squares of some specified number of error samples with respect to the coefficients. The main reason for this choice of a minimisation criterion is simply that the solution is linear, tractable and produces results, which, in most applications, are satisfactory for the analysis of speech [95].

The *total squared error* is

$$\begin{aligned} \alpha &= \sum_{n=n_0}^{n_1} e_n^2 = \sum_{n=n_0}^{n_1} \sum_{i=0}^{m} \sum_{j=0}^{m} a_i x_{n-i} x_{n-j} a_j \\ &= \mathbf{a}^T \mathbf{X}^T \mathbf{X} \mathbf{a} = \mathbf{a}^T \mathbf{C}_x \mathbf{a}, \end{aligned} \tag{3.3}$$

where $\mathbf{X}$ is the $(n_1 - n_0 + 1) \times m$ convolution matrix of x_n (see Section 2.3) and the covariance matrix $\mathbf{C}_x$ is defined as

$$\mathbf{C}_x = \mathbf{X}^T \mathbf{X}. \tag{3.4}$$

To be able to minimise α, we need to set a constraint $a_0 = 1$, which can be obtained using Lagrange multiplier λ with objective function

$$\eta(\mathbf{a}, \lambda) = \mathbf{a}^T \mathbf{C}_x \mathbf{a} - 2\lambda \mathbf{a}^T \mathbf{u} \tag{3.5}$$

where $\mathbf{u} = [1\, 0\, 0\, \ldots\, 0]^T$.

The minimum is found by setting the partial derivatives $\partial / \partial a_k$ to zero, and we obtain

$$0 = \frac{\partial \eta(\mathbf{a}, \lambda)}{\partial \mathbf{a}} = 2\mathbf{C}_x \mathbf{a} - 2\lambda \mathbf{u} \tag{3.6}$$

which yields the solution as

$$\mathbf{C}_x \mathbf{a} = \lambda \mathbf{u}. \tag{3.7}$$

The prediction error $\mathbf{e} = [e_{n_0}\, e_{n_0+1}\, \ldots\, e_{n_1}]^T$ is found by

$$\mathbf{e} = \mathbf{X}\mathbf{a}. \tag{3.8}$$

This method is generally known as the *covariance method*.

In mathematical terms, a more securely warranted minimisation criterion is to minimise the expected value $E[\cdot]$ of the squared error e_n^2. Thus, from Eq. 3.2 we obtain

$$E\left[e_n^2\right] = E\left[\mathbf{a}^T \mathbf{x}^T \mathbf{x} \mathbf{a}\right] = \mathbf{a}^T E\left[\mathbf{x}^T \mathbf{x}\right] \mathbf{a} = \mathbf{a}^T \mathbf{R} \mathbf{a} \tag{3.9}$$

where $\mathbf{R}$ is the autocorrelation matrix, which is real, symmetric and Toeplitz. As in the covariance method, we have the objective function

$$\eta(\mathbf{a}, \lambda) = \mathbf{a}^T \mathbf{R} \mathbf{a} - 2\lambda \mathbf{a}^T \mathbf{u} \tag{3.10}$$

solution which yields the normal equations

$$\mathbf{R}\mathbf{a} = \sigma^2 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{3.11}$$

where $\sigma^2 = \lambda$ is the error energy, since

$$E[e_n^2] = \mathbf{a}^T \mathbf{R} \mathbf{a} = \sigma^2 \mathbf{a}^T \mathbf{u} = \sigma^2 a_0 = \sigma^2. \tag{3.12}$$

This method is generally known as the *autocorrelation method*.

A crucial property of this method is that, if we assume that the autocorrelation matrix $\mathbf{R}$ is a positive definite Toeplitz matrix, then the model yields a predictor $\mathbf{a}$ whose Z-transform $A(z) = \sum_{i=0}^{m} a_i z^{-i}$ has all its roots inside the unit circle. This property, also known as the minimum-phase property or the stability of the inverse model $A^{-1}(z)$, ensures that when it is used as a autoregressive filter, the impulse response converges [91, 55]. Root loci of linear predictors is demonstrated in Figure 3.1.

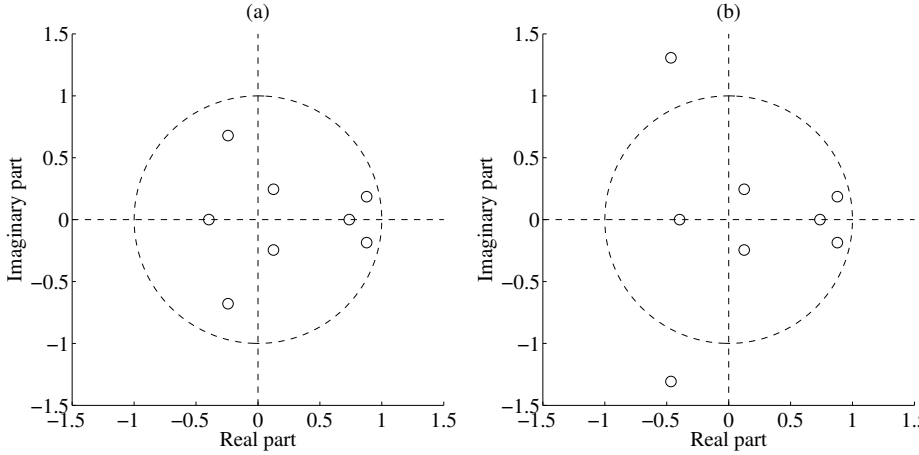

Figure 3.1: Illustration of root loci of linear predictors: (a) minimum-phase predictor and (b) non-minimum-phase predictor.

The root loci of the LP model is so important, that it has been extensively studied and a large number of equivalent proofs of the minimum-phase property have been presented [150, 82, 45, 104, 138, 55]. While the linear predictor is undoubtedly minimum-phase if the above assumption holds, that

is, the model is minimum-phase if the autocorrelation matrix $\mathbf{R}$ is positive definite, in practical applications this assumption is not always true. Therefore, and due to its importance to other applications, zero location of general polynomials has been thoroughly studied [14, 15, 16, 17].

3.1.1 Relation to the Acoustic Tube Model

In Section 1.2.1, we discussed the acoustic tube model of the vocal tract. We concluded, that the vocal tract can be accurately modelled with the tube model, which, in turn, can be described by a set of reflection coefficients Γ_i. Furthermore, there is a unique mapping from the reflection coefficients to the transfer function of the vocal tract model. Specifically, if we let $A_i(z)$ denote the transfer function on iteration i, we can recursively construct the transfer function with [95, 103, 119]

$$A_{i+1} = A_i(z) + \Gamma_i z^{-i-1} A_i(z). \tag{3.13}$$

When comparing with Eq. 2.6, we note that the the reflection coefficients Γ_i of the Levinson recursion and the acoustics tube model are, in fact, equal. Each iteration of the Levinson recursion can, therefore, in the tube model, be interpreted as the attachment of an additional tube segment.

In conclusion, since LP models are presented by means of a transfer function $A(z)$, which is generally a solution to a Toeplitz system, the model can be interpreted as the acoustic tube model of the vocal tract. Hence, the LP model is a justified model for modelling of the spectral of speech.

3.2 Symmetric Predictors

Symmetric and antisymmetric predictors $A^{\pm}(z) = \sum_{i=0}^{m} a_i^{\pm} z^{-i}$ are such that their coefficients have

$$\begin{aligned} a_i^+ &= +a_{m-i}^+ \\ a_i^- &= -a_{m-i}^- \end{aligned} \tag{3.14}$$

In other words, in vector form we have $\mathbf{a}_{\pm} = \pm \mathbf{a}_{\pm}^{\#}$, where superscript $\#$ denotes reversal of rows. A trivial property of such real polynomials is that their roots may appear in any of the following constellations:

- Root pairs either on the unit circle as conjugate root pairs $z = a$ and $z = a^*$ with $|a| = 1$ (see Fig. 3.2(a)), $a \in \mathbb{C}$ or on the real axis $z = a$ and $z = a^{-1}$ with $a \in \mathbb{R}$ (see Fig. 3.2(b)).
- Root quadruples at $z = a$, $z = a^*$, $z = a^{-1}$ and $z = a^{*-1}$ with $a \in \mathbb{C}$ (see Fig. 3.2(c)).
- Single root at $z = -1$ for both symmetric and antisymmetric polynomials. In addition, antisymmetric polynomials always have a root at $z = +1$ (see Fig. 3.2(d)).

Observe that, consequently, symmetric and antisymmetric predictors are never minimum-phase.

Furthermore, since predictors are usually calculated as solutions to Toeplitz systems of type $\mathbf{Ra}_\pm = \mathbf{c}_\pm$, we note that also vector $\mathbf{c}_\pm$ has to be symmetric or antisymmetric because $\mathbf{R}$ is symmetric Toeplitz. Another formulation of the same property is that if $\mathbf{Ra} = \mathbf{c}$ then $\mathbf{Ra}^\# = \mathbf{c}^\#$.

It has been claimed [137], that the roots of symmetric linear prediction models lie on the unit circle. While this claim is valid in certain cases, it is certainly not universally true. For example, in Section 3.4.2 we find a linear predictive model that can have roots in all the symmetric constellations presented above.

3.3 Spectral Distortion Measures

In many applications it is important to measure how much distortion an operation exerts on the spectrum, that is, to measure how much the spectrum changes, for example, in quantisation of parameters.

Define the spectral error or difference $V(\omega)$ as

$$V(\omega) = 10 \log_{10}[A(\omega)] - 10 \log_{10}[\hat{A}(\omega)]. \tag{3.15}$$

The simplest and most used of spectral distortion measures, log spectral distortion (SD), is defined as

$$d^2 = \frac{1}{\pi} \int_0^\pi |V(\omega)|^2 \, d\omega, \tag{3.16}$$

where $A(\omega)$ is the original spectrum and $\hat{A}(\omega)$ the distorted spectrum [49].

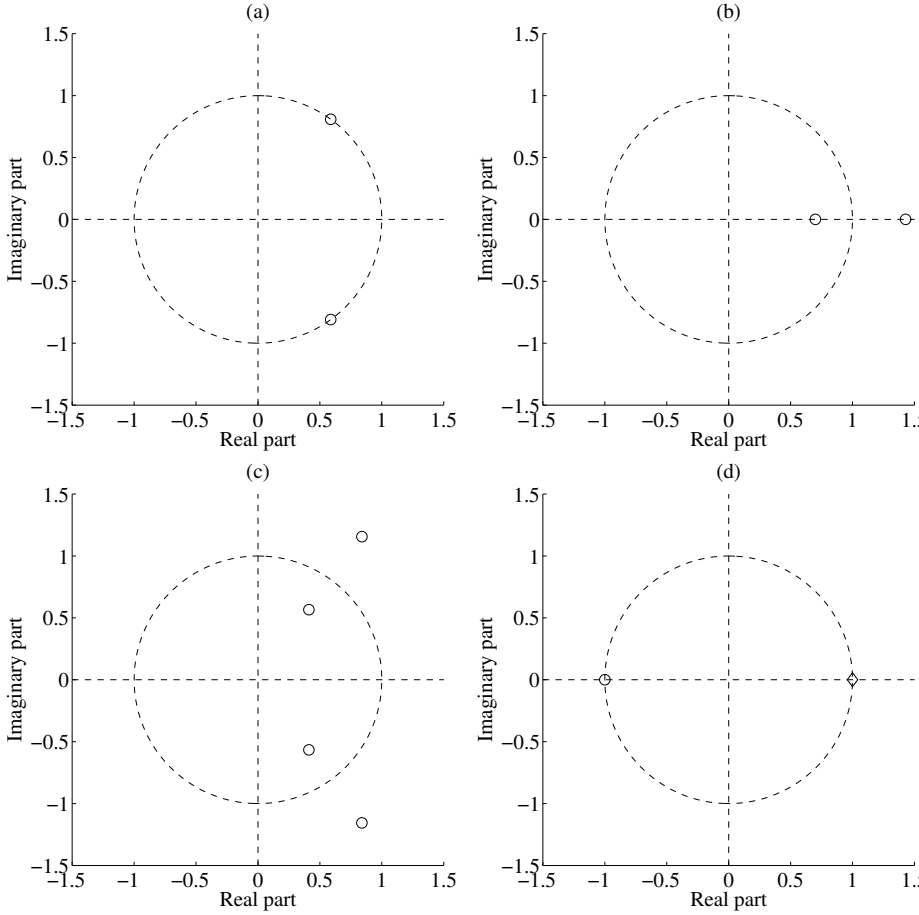

Figure 3.2: Illustration of root constellations of real symmetric and antisymmetric polynomials in the complex plain $\mathbb{C}$: (a) unit circle root pair (b) real root pair (c) root quadruple (d) trivial root at $z = -1$ (circle 'o') and for antisymmetric polynomials only at $z = +1$ (diamond '$\Diamond$').

However, log spectral distortion does not take into account any perceptual features of the sound. Different areas of the spectrum are perceptually of different importance and should therefore be weighted differently in the distortion measure. One formulation of the weighted distance measure is [30, 29]:

$$d_W^2 = \frac{1}{\pi W_0} \int_0^\pi W_B(\omega)^2 \left|V(\omega)\right|^2 d\omega, \tag{3.17}$$

where $W_B(\omega)$ is the weighting function on frequency range $\omega \in [0, \pi]$ and W_0 a scaling coefficient defined as the integral of (or in the discrete case, the sum of) $W_B(\omega)$ over the frequency range. The weighting function can be defined to adjust to the Bark scale by setting

$$W_B(\omega) = \frac{1}{25 + 75\left[1 + 1.4\left(\frac{\omega}{1000}\right)^2\right]^{0.69}}. \tag{3.18}$$

Another approach is to evaluate the maximum likelihood distortion measure, usually denoted as the Itakura-Saito distortion measure [95, 49], first presented in [70]. The distortion measure then becomes [95, 49]

$$I = \frac{1}{\pi} \int_0^\pi \left[e^{V(\omega)} - V(\omega) - 1\right] d\omega. \tag{3.19}$$

Similar to Eq. 3.17, we can introduce a spectral weighting factor $W_B(\omega)$ to facilitate perceptual weighting and we obtain

$$I_W = \frac{1}{\pi W_0} \int_0^\pi W_B(\omega)^2 \left[e^{V(\omega)} - V(\omega) - 1\right] d\omega. \tag{3.20}$$

3.4 Modified Linear Predictive Models

The idea of improving LP models, as presented in this thesis for example, by perceptual criteria, is not at all new. In this section, we present some preceding work on the same problem. The motivation is as follows: Conventional LP models the input signal with constant weighting for the whole frequency range [11]. However, human perception does not have constant frequency perception in the whole frequency range. For example, in general, low frequencies are perceived with higher accuracy than high frequencies [41, 121]. Therefore, since LP treats all frequencies equally, effort is wasted on high frequencies while important information in the low frequencies is discarded.

Researchers in the field seem to have a peculiar but entertaining predilection for acronyms. Hence, we will in this section present, among others, linear predictive models known as WLP, FWLP, DAP, PLP, LPES, LPLE and SLP.

3.4.1 Warped Linear Prediction

Warping is a technique that enhances the frequency resolution of a model by transformation to a "warped" frequency scale, where the model is constructed. An inverse transform maps the model to the original frequency domain. The motivation is, that if the warped frequency domain has a frequency resolution which resembles that of the perception, then a model generated in the warped domain will seize those characteristics of the signal that are perceptually important. An early attempt toward frequency warping can be found in [94], but the current formulation has its roots in [141].

The warped linear predictive model is acquired by first Z-transforming Eq. 3.1 to obtain

$$\hat{X}(z) = -\sum_{i=1}^{m} a_i z^{-i} X(z). \tag{3.21}$$

We then replace the delay z^{-i} by a linear all-pass filter $D(z)$ and we have [63, 62]

$$\hat{X}(z) = -\sum_{i=1}^{m} a_i D^i(z) X(z). \tag{3.22}$$

A common warping method is to use the all-pass model

$$D(z) = \frac{z^{-1} - \lambda}{1 - \lambda z^{-1}}. \tag{3.23}$$

We can then generate an LP model for the warped signal as we did in Section 3.1, to obtain a warped model $A_w(z)$. The model $A_w(z)$ can, in turn, be readily inverse transformed to the original domain.

3.4.2 Two-sided Linear Prediction

A reformulation of the linear predictive problem that has been frequently reinvented, is the two-sided linear predictive model [5, 85, 32, 64, 88]. Due to the different perspectives of the many authors, two-sided LP has been presented in a variety forms. We will use the current notation to minimise confusion. Instead of Eq. 3.1, this model is based on an estimate $\hat{x}_n$ of signal

x_n that is a linear combination of both past and future samples x_n. The model is then, by design, non-causal. The estimate is defined as

$$\hat{x}_n = -\left[\sum_{i=-m/2}^{-1} a_i x_{n-i} + \sum_{i=1}^{m/2} a_i x_{n-i}\right]. \tag{3.24}$$

By a derivation similar to Section 3.1, we can readily find the solution as the minimum of the expected value of the squared residual as

$$\mathbf{Ra} = \sigma^2 \left[0\,0 \ldots 0\,0\,1\,0\,0 \ldots 0\,0\right]^T \tag{3.25}$$

where $\mathbf{a} = [a_{-m/2}\, a_{-m/2+1} \ldots a_0 \ldots a_{m/2-1}\, a_{m/2}]^T$ and $a_0 = 1$.

The advantage of two-sided LP is that its roots often lie in the quadruple constellation of Fig. 3.2(c). If the roots outside the unit circle are reflected symmetrically to the inside, we obtain an optimal "quadratic" model, that is, a model where all the roots are double and the match to the spectrum is optimal. This kind of model has enhanced spectral dynamics, which is beneficial in some cases. Similar algorithms have been presented in [7, 59].

3.4.3 Frequency Weighted Linear Prediction

A straightforward approach to perceptual weighting of the LP problem, called Frequency Weighted Linear Prediction (FWLP), is presented in [28, 27]. It is based on the creation of an optimal LP model with respect to the Itakura-Saito spectral distortion measure presented in Section 3.3, Eq. 3.19.

It is shown that if the spectral weighting function $W(z)$ is a one-pole polynomial, then the resulting model is minimum-phase. However, for more complex functions $W(z)$, the authors were unable to find criteria for the minimum-phase property.

The one-pole FWLP model was shown to be solvable by Cholesky decomposition. The multiple-pole case is more complex, but was shown to be iteratively solvable by the Newton-Raphson method.

3.4.4 Discrete All-Pole Modelling

Another modelling method based on the Itakura-Saito distortion criterion is the Discrete All-Pole (DAP) modelling method, originally presented in [35, 36]. This method is based on a discretisation of the spectral distortion

criterion in Eq. 3.19 as

$$I_{DAP} = \frac{1}{N} \sum_{k=1} N \left[e^{V(\omega_k)} - V(\omega_k) - 1 \right]. \tag{3.26}$$

In the conventional discretisation of Itakura-Saito, the points ω_k would be equally spaced points of the spectrum, but in DAP, the ω_k may be chosen otherwise. Consequently, the system designer can choose the parts of the spectrum to be modelled accurately. The proposed approach is, however, to choose ω_k so that they fit the harmonic peaks of the input signal. It is then assumed that the input signal spectrum has harmonic structure.

Straightforward minimisation of the spectral distortion criterion results in non-linear equations, which cannot be easily solved. Fortunately, we can readily obtain criteria for optimality and, by a recursive algorithm, find estimates for the model [35, 36].

It should be noted that the basic DAP algorithm does not involve any perceptual weighting. There exists, however, some recent work that includes perceptual weighting with a discrete mean square all-pole model [113].

3.4.5 Perceptual Linear Prediction

An approach for linear prediction completely based on perceptual criteria is the Perceptual Linear Prediction (PLP) [58]. This model includes the following perceptually motivated analyses:

1. Critical-band spectral resolution. The spectrum of the original signal is warped (see Section 3.4.1) into the Bark frequency scale, where a critical-band masking curve is convolved to the signal.

2. Equal-loudness pre-emphasis. The signal is pre-emphasised by a simulated equal-loudness curve to match the frequency magnitude response of the ear.

3. Intensity-loudness power law. The signal amplitude is compressed by the cubic-root to match the nonlinear relation between intensity of sound and perceived loudness.

After these operations, all signal components are perceptually equally weighted and we can, from the modified signal, make a regular LP model.

In [58], the author continues to analyse the PLP model in view of speech recognition and vowel perception, and argues, that due to the perceptual

motivation of PLP and due to the experiments presented, PLP behaves well even with low-order LP models (such as $m = 5$).

3.4.6 Linear Prediction by Sample Grouping

The LP problem can also be reformulated in the time domain by sample grouping [154, 6, 152]. Similar work related to warped LP models (see Section 3.4.1) can be found in [62]. In contrast to the earlier studies and with pedagogical intentions, we will derive the reformulations in matrix form. We will start by a modification of Eq. 3.1 to obtain

$$\hat{x}_n = -\sum_{i=1}^{m} a_i F(x_n,\, x_{n-1}, \ldots, x_{n-N}) \tag{3.27}$$

where $F(\cdot)$ is a linear function of its parameters and N is a scalar $N > m$. The motivation for this formulation is, that if $F(\cdot)$ is chosen to emphasise certain components of the input signal and to attenuate others; when optimising, the optimal predictor will then give more attention to the emphasised components and less attention to the attenuated components. Consequently, the designer of the model can choose which components of the input signal are given more attention in the optimisation process.

This approach is well known in systems identification and often denoted as rank-reduction [128, 162]. The name emerges from the fact that $F(\cdot)$ takes as input N parameters when the model order is m with $N > m$. The model therefore uses fewer parameters a_i than it uses input signal values x_n. Consequently, the degrees of freedom or rank of the model is reduced.

Similarly to Section 3.1, the prediction error becomes

$$\begin{aligned} e_n &= x_n - \hat{x}_n = x_n + \sum_{i=1}^{m} a_i F(x_n,\, x_{n-1}, \ldots, x_{n-N}) \\ &= \mathbf{u}^T\mathbf{x} + \mathbf{a}^T\mathbf{F}^T\mathbf{x} \end{aligned} \tag{3.28}$$

where $\mathbf{F}$ is an $N \times m$ matrix corresponding to the linear combination of $F(\cdot)$, $\mathbf{u} = [1\,0\ldots0]^T$ and with constraint $a_0 = 1$. As in the autocorrelation method, we minimise the expected value of the squared residual

$$\begin{aligned} E\left[e_n^2\right] &= E\left[\mathbf{a}^T\mathbf{F}^T\mathbf{x}\mathbf{x}^T\mathbf{F}\mathbf{a} + 2\mathbf{u}^T\mathbf{x}\mathbf{x}^T\mathbf{F}\mathbf{a} + \mathbf{u}^T\mathbf{x}\mathbf{x}^T\mathbf{u}\right] \\ &= \mathbf{a}^T\mathbf{F}^T E\left[\mathbf{x}\mathbf{x}^T\right]\mathbf{F}\mathbf{a} + 2\mathbf{u}^T E\left[\mathbf{x}\mathbf{x}^T\right]\mathbf{F}\mathbf{a} + \mathbf{u}^T E\left[\mathbf{x}\mathbf{x}^T\right]\mathbf{u} \\ &= \mathbf{a}^T\mathbf{F}^T\mathbf{R}\mathbf{F}\mathbf{a} + 2\mathbf{u}^T\mathbf{R}\mathbf{F}\mathbf{a} + \mathbf{u}^T\mathbf{R}\mathbf{u}. \end{aligned} \tag{3.29}$$

The zero of the partial derivative $\partial/\partial\mathbf{a}$ yields the minimum at

$$\mathbf{F}^T\mathbf{R}\mathbf{F}\mathbf{a} = 2\mathbf{F}^T\mathbf{R}\mathbf{u}. \tag{3.30}$$

This solution is derived in scalar form in, among others, [154, 6, 5, 152]. It is instructive to note, however, that Eq. 3.30 does not, *per se*, say anything about the stability of the corresponding all-pole model.

Interestingly, there is another equivalent solution method. By defining $\hat{\mathbf{a}} = \mathbf{F}\mathbf{a} + \mathbf{u}$, we obtain from Eq. 3.28

$$e_n = \hat{\mathbf{a}}^T\mathbf{x}. \tag{3.31}$$

To accommodate this change, the constraints have to be modified accordingly. Since $\mathbf{F}\mathbf{a} = \hat{\mathbf{a}} - \mathbf{u}$, then $\mathbf{a} - \mathbf{u}$ must be in the column space of $\mathbf{F}$ and $\mathbf{F}_0^T(\mathbf{a} - \mathbf{u}) = 0$ where matrix $\mathbf{F}_0$ is the null-space of $\mathbf{F}$; that is, $\mathbf{F}^T\mathbf{F}_0 = 0$. We can then write the objective function as

$$\eta(\hat{\mathbf{a}}, \mathbf{g}) = \hat{\mathbf{a}}^T\mathbf{R}\hat{\mathbf{a}} - 2\mathbf{g}^T\mathbf{F}_0^T(\hat{\mathbf{a}} - \mathbf{u}) \tag{3.32}$$

where $\mathbf{g}$ is a vector of Lagrange coefficients.

Again, the minimum is found by setting the partial derivative to zero, and we have the set of equations [128]

$$\begin{aligned} \mathbf{R}\hat{\mathbf{a}} &= \mathbf{F}_0\mathbf{g} \\ \mathbf{F}_0^T\mathbf{R}^{-1}\mathbf{F}_0\mathbf{g} &= \mathbf{F}_0\mathbf{u} \end{aligned} \tag{3.33}$$

While Eq. 3.30 provides a solution by inversion of an $m \times m$ matrix $\mathbf{F}^T\mathbf{R}\mathbf{F}$, whereas Eq. 3.33 provides a solution through inversion of an $N \times N$ matrix $\mathbf{R}$ and an $(N - m) \times (N - m)$ matrix $\mathbf{F}_0^T\mathbf{R}^{-1}\mathbf{F}_0$. The latter can still be more efficient since we can use efficient algorithms such as Levinson-Durbin recursion for solution of Eq. 3.33 (see Section 2.2). Moreover, the form in Eq. 3.33 can present additional information of the model not readily seen in Eq. 3.30. For educational purposes, let us study three algorithms presented in [153, 154, 6] and collectedly in [152], in more detail, with the above matrix notation.

In the current notation, from [153, 154, 6, Eqs. 2] and Eq. 3.1 we obtain

$$\hat{x}_n^{LPES} = \sum_{i=1}^{m} a_i \left\{(i+1)\left[x_{n-i} - x_{n-i-1}\right] + x_{n-i-1}\right\} \tag{3.34}$$

$$\hat{x}_n^{SLP} = \sum_{i=1}^{m} a_i \left\{\frac{i+1}{i}\left[x_{n-1} - x_{n-i-1}\right] + x_{n-i-1}\right\} \tag{3.35}$$

$$\hat{x}_n^{LPLE} = \sum_{i=1}^{m} a_i \left\{2i\left[x_{n-2i+1} - x_{n-2i}\right] + x_{n-2i}\right\}. \tag{3.36}$$

Immediately we observe, that Eqs. 3.34 and 3.35 use $m+1$ samples of the input signal x_n to estimate a model of order m, while Eq. 3.36 uses $2m+1$ samples for a model of order m. The models LPES and SLP thus provide a rank reduction of order one, while LPLE provides a rank reduction of order $m+1$.

The linear combination matrices $\mathbf{F}$ (see Eq. 3.2) of Eqs. 3.34–3.36 can be written as

$$\mathbf{F}_{LPES} = \begin{bmatrix} 0 & 2 & -1 & 0 & 0 & 0 & \dots \\ 0 & 0 & 3 & -2 & 0 & 0 & \dots \\ 0 & 0 & 0 & 4 & -3 & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots \end{bmatrix} \tag{3.37}$$

$$\mathbf{F}_{SLP} = \begin{bmatrix} 0 & \frac{2}{1} & -\frac{1}{1} & 0 & 0 & 0 & \dots \\ 0 & \frac{3}{2} & 0 & -\frac{1}{2} & 0 & 0 & \dots \\ 0 & \frac{4}{3} & 0 & 0 & -\frac{1}{3} & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots \end{bmatrix} \tag{3.38}$$

$$\mathbf{F}_{LPLE} = \begin{bmatrix} 0 & 2 & -1 & 0 & 0 & 0 & \dots & \\ 0 & 0 & 0 & 4 & -3 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 8 & -7 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots\ddots & \ddots & \ddots & \end{bmatrix}. \tag{3.39}$$

The corresponding null spaces are

$$\mathbf{F}^T_{LPES,0} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 1 & 2 & 3 & 4 & 5 & \dots \end{bmatrix} \tag{3.40}$$

$$\mathbf{F}^T_{SLP,0} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 1 & 2 & 3 & 4 & 5 & \dots \end{bmatrix} \tag{3.41}$$

$$\mathbf{F}^T_{LPLE,0} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 1 & 2 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 3 & 4 & 0 & \dots \\ \vdots & & \vdots & & \vdots & & \ddots \end{bmatrix}. \tag{3.42}$$

Looking at Eqs. 3.40–3.42 we recognise the similarity between the null-spaces. Specifically, the null spaces of LPES and SLP are exactly equal while LPLE is very similar. We can conclude that the LPES and SLP will yield equal result, even though the Eqs. 3.34–3.36 gives no hint of this to the casual reader.

In conclusion, sample grouping is an engineering approach to linear prediction with a perceptual goal in mind. However, no thorough analysis of perceptual properties of presented algorithms exists as yet.

3.5 Non-linear Prediction

Even though the linear predictive model, based on the acoustic tube model (see Section 1.2.1), is a good approximation of the vocal tract, in real life, the speech production process is non-linear. Therefore, performance of linear models will always remain limited. In order to improve accuracy, we can study non-linear models. While this is outside the main scope of this thesis, a few notes may be helpful. Some of the presented models have been developed specifically for speech, others for any prediction problem, but we will consider them all since distinguishing between them is sometimes difficult.

Studies in spoken speech have shown that most of the points in the state space lie very close to a manifold of surprisingly low dimensionality, fewer than three. This result indicates that one should be able to construct non-linear models of speech of low dimensionality that significantly outperform linear models [149].

A simple approach to non-linear prediction is to assign a code-book for the vector quantised past of the input signal. Both the analysis and synthesis stages are then merely table look-ups, which can be performed very efficiently. However, construction of the code-book might prove costly [159, 131].

Neural networks (NN) for speech processing have been studied extensively [57, 76] and it is only natural to attempt to apply those methods to the speech spectral modelling task as well. The methods have developed along the current trends in NN research, ranging from Radial Basis Function methods [100, 101], and Hidden Markov Models [42], through Time Delay NN [145] to Real-Time Recurrent Learning NN [164].

Non-linear prediction methods are usually computationally complex, since solution of model parameters is generally non-linear. However, some non-linear autoregressive models, which are linear to their parameters, have been developed [44]. A final non-linear prediction method is application of Volterra filters, which has been used with some success in, among others, [145, 101, 130].

Chapter 4

Line Spectrum Pair Decomposition

In speech transmission, it is often essential to find a representation of the model parameters which tolerates small quantisation errors. One such technique is the Line Spectrum Pair (LSP) polynomials [69, 132]. This representation includes the decomposition of a polynomial, such as the linear predictor, into another domain, the Line Spectrum Frequency (LSF) domain, where the parameters can be represented as angles of polynomial zeros on the unit circle.

The prevalent method for speech transmission is code excited linear prediction (CELP) with LSP polynomials. This combination appears in all major coding standards [1, 3, 4, 2, 31]. CELP is based on modelling the speech spectrum with a linear predictive model and using a code-book model for the residual [119]. The LP parameters are transferred in terms of line spectrum frequencies (LSFs); that is, as the zero locations of the LSP polynomials. This approach has several advantages. Firstly, LSFs are robust against quantisation errors as long as they are kept interlaced [132]. Secondly, interframe interpolation is well-behaved [108]. Finally, the computational load associated with root-finding of LSFs is reasonable compared to other coding schemes [123].

Prior to LSP, the most suitable representations of LP coefficients was log area ratios (LAR) and inverse sine quantisation. In terms of spectral deviation due to quantisation errors, either of these representations can be optimal, depending on the deviation measure [71]. However, a dissenting article from the same time period claims that the reflections coefficients would

be superior to other coding schemes [156]. Either way, the LSP decomposition was soon found to be insurmountable to all of the above mentioned coding schemes. Quality-wise, coding with a 31-bit representation of the LSP polynomials is equivalent to, or better than, a 41-bit representation with reflection coefficients [75].

In this present chapter, we will present the most important properties and applications of LSP decomposition.

4.1 Basics

The LSP polynomials are defined, for an order m predictor $A(z)$, with [69, 132]

$$\begin{aligned} P(z) &= A(z) + z^{-m-1}A(z^{-1}) \\ Q(z) &= A(z) - z^{-m-1}A(z^{-1}). \end{aligned} \tag{4.1}$$

We can readily see that, using $P(z)$ and $Q(z)$, polynomial $A(z)$ can be reconstructed as

$$A(z) = \frac{1}{2}\left[P(z) + Q(z)\right]. \tag{4.2}$$

The roots α_i and β_i of $P(z)$ and $Q(z)$, respectively, have a number of useful properties, namely, it holds that [127, 132, 137]

1. α_i and β_i are on the unit circle $|\alpha_i| = |\beta_i| = 1$ and can be presented as $\alpha_i = e^{i\pi\lambda_i}$ and $\beta_i = e^{i\pi\gamma_i}$.

2. λ_i and γ_i are simple and distinct $\lambda_i \neq \lambda_j$ and $\gamma_i \neq \gamma_j$ for $i \neq j$, and $\lambda_i \neq \gamma_i$ for all i.

3. λ_i and γ_i are interlaced, that is, $\gamma_i < \lambda_i < \gamma_{i+1}$ for all i.

Polynomials $P(z)$ and $Q(z)$ can be reconstructed from λ_i and γ_i, and since we can reconstruct $A(z)$ from $P(z)$ and $Q(z)$, we can use the angles λ_i and γ_i to uniquely describe $A(z)$. This description is bounded since $\lambda_i, \gamma_i \in [0, 1]$ (the complex conjugate λ_i^* and γ_i^* are redundant and can be ignored).

Conversely, if two polynomials have zeros interlaced on the unit circle, their sum is minimum-phase [132]. Therefore, if we make sure that we retain the interlacing property, then the description is robust in terms of stability of the all-pole model.

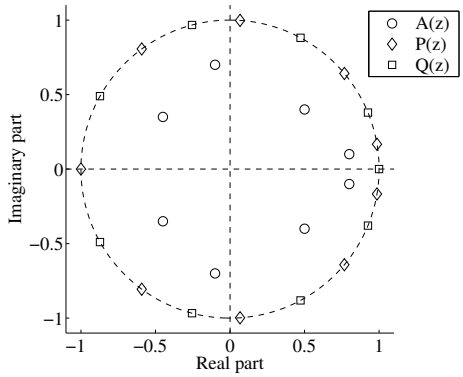

Figure 4.1: Illustration of root loci of LSP polynomials $P(z)$ and $Q(z)$ calculated from polynomial $A(z)$.

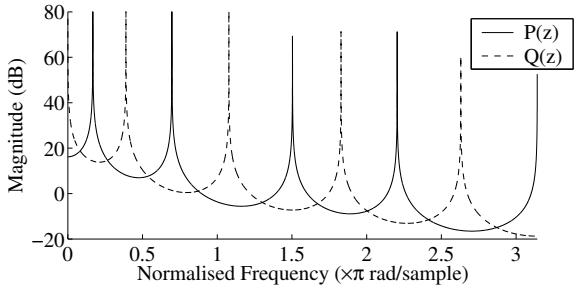

Figure 4.2: Illustration of the Line Spectrum Frequencies in the spectra of polynomials $P(z)$ and $Q(z)$.

Properties 1 and 3 are called the *unit circle property* and the *intra-model interlacing property* of LSP polynomials. These properties are illustrated in Figure 4.1.

Since the roots lie on the unit circle, the all-pole models $P^{-1}(z)$ and $Q^{-1}(z)$ will have infinite values at these locations. In terms of the spectrum, these roots can be seen as vertical lines at frequencies corresponding to the angle of each root. These lines are known as the Line Spectrum Frequencies (LSFs) of the corresponding model.

Depending on the model order m, the LSP polynomials have trivial zeros at $z = \pm 1$ as follows [122]:

m	$P(z)$	$Q(z)$
odd	$z = -1$	$z = +1$
even	none	$\begin{cases} z = +1 \\ z = -1 \end{cases}$

Cancelling the trivial zeros of mth order LSP polynomials $P_m(z)$ and $Q_m(z)$, yields symmetric polynomials $R_{P_m}(z)$ and $R_{Q_m}(z)$ of even order as [122]

$$\begin{aligned} R_{P_m}(z) &= \frac{P_m(z)}{1+z^{-1}}, \quad R_{Q_m}(z) = \frac{Q_m(z)}{1-z^{-1}}, \quad m \text{ odd} \\ R_{P_m}(z) &= P_m(z), \qquad R_{Q_m}(z) = \frac{Q_m(z)}{1-z^{-2}}, \quad m \text{ even}. \end{aligned} \tag{4.3}$$

Note that $R_{P_m}(z)$ and $R_{Q_m}(z)$ are generally no longer of order m, but since they are generated from mth order equations, we retain the subscript.

Consecutive order LSP polynomials have another intriguing interlacing property called the *inter-model interlacing property*. Namely, the roots of an order m LSP polynomial $P_m(z)$ are interlaced with that of $P_{m+1}(z)$ and similarly for the roots of $Q_m(z)$ and $Q_{m+1}(z)$ [97, 77]. In other words, if $\lambda_i^{(m)}$ are the angles of the roots of the order m polynomial $P_m(z)$, and $\lambda_i^{(m+1)}$ are the angles of the roots of the order $(m+1)$ polynomial $P_{m+1}(z)$, then

$$\lambda_i^{(m+1)} < \lambda_i^{(m)} < \lambda_{i+1}^{(m+1)}, \qquad \text{for all } i. \tag{4.4}$$

The same holds for roots of $Q_m(z)$ and $Q_{m+1}(z)$.

Further elaboration on the bounds of LSFs through recursive evaluation at subsequent model orders is presented in [77]. The article also shows another proof for the inter-model interlacing property. Furthermore, the article proposes a method for determining a proper model order in a view of representing formant frequencies.

The root loci of $A_m(z)$ for $\Gamma_m \in \mathbb{R}$ in Eq. 2.6 are studied in [87]. This article also presents an alternative and somewhat more intuitive proof for the unit circle property than [132]. The root loci concept is a classical method in polynomial analysis and can be studied further in some basic signal processing tutorials, e.g. [102].

A physical motivation for the LSF frequencies based on the glottal driving-point impedance of a discrete matched-impedance vocal tract model is given in [54]. This model also provides yet another alternative proof for the interlacing property and stability of the model.

As an extension to LSP noted in [97], it is possible to prove that unit circle and intra-model interlacing properties hold also for the polynomial pair $P_{m+1}^l(z) = A_m(z) + z^{-l}A_m(z^{-1})$ and $Q_{m+1}^l(z) = A_m(z) - z^{-l}A_m(z^{-1})$ with $l \geq 0$ [127, 45]. It is, in fact, sufficient to prove the properties for $l = 0$ since we can always insert zeros at the end of the coefficient vector. The number of added zeros then corresponds to the value of l.

At first sight, it would be tempting to use LSFs as a basis for sinusoidal modelling, similarly to the way eigenvectors are used in Pisarenko's harmonic decomposition [117, 55]. Unfortunately, LSFs are biased estimators which makes them unsuitable for sinusoidal modelling [139].

It can be concluded that it is possible to describe the spectral envelope of a signal through the angles of the zeros of LSP polynomials $P_{m+1}(z)$ and $Q_{m+1}(z)$ calculated from an LP polynomial $A_m(z)$. This yields a convenient representation of the LP coefficients, since the range of the angles is limited to $[0, \pi]$, and the stability of the all-pole model corresponding to $A_m(z)$ is guaranteed if the interlacing property is retained.

4.1.1 Relation to Levinson-Durbin Recursion

The Levinson recursion is an iterative solution method for matrix equations. The problem is formulated as $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A}$ is a known symmetric Toeplitz $n \times n$ matrix, $\mathbf{b}$ a known $n \times 1$ vector and $\mathbf{x}$ the unknown $n \times 1$ vector. The recursion algorithm computes the solution vector $\mathbf{x}$ iteratively for $i \times i$ sub-matrices using the solution of the $(i-1) \times (i-1)$ sub-matrix [89].

The Levinson-Durbin recursion is a special case of the Levinson recursion, where the problem is defined as $\mathbf{Ra} = \mathbf{r}$, where $[\mathbf{R}]_{i,j} = R(i-j)$, $[\mathbf{r}]_j = R(j)$ and $\mathbf{a}$ is the unknown vector [34]. In the Levinson-Durbin recursion, consecutive order solutions are calculated using the following relation

$$A_m(z) = A_{m-1}(z) + \Gamma_m z^{-m-1} A_{m-1}(z^{-1}), \tag{4.5}$$

where Γ_m is called the reflection coefficient.

Setting $\Gamma_m = \pm 1$ in the above equation yields the LSP polynomials $P_m(z)$ and $Q_m(z)$ as can be seen from Eq. 4.1 [33]. Using equation 4.5 it is also possible to form recursive relations between consecutive order LSP polynomials as [97]

$$2P_m(z) = (1 + k_{m-1})\left(1 + z^{-1}\right) P_{m-1} + (1 - k_{m-1})\left(1 - z^{-1}\right) Q_{m-1} \tag{4.6}$$

$$2Q_m(z) = (1 + k_{m-1})\left(1 - z^{-1}\right) P_{m-1} + (1 - k_{m-1})\left(1 + z^{-1}\right) Q_{m-1}. \tag{4.7}$$

The inter-model interlacing property (in Section 4.1) follows from the above equations.

4.2 Finding the Roots of LSP Polynomials

When using LSP in speech coding, it is not enough to calculate the LSP polynomials $P(z)$ and $Q(z)$ through Eqs. 4.1, and transmit their coefficients. The LSP coefficients suffer from the same problems as the LP coefficients; that is, they are sensitive to quantisation errors. The motivation of using the LSP decomposition in speech coding, is to transmit the angles of the zeros of $P(z)$ and $Q(z)$ (which lie on the unit circle), since they are robust and tolerate small errors. In order to transmit the LP information using LSFs, the root locations of the LSP polynomials have to be determined. In this section, we will discuss some of the proposed methods for finding these roots.

It is well known that the zeros of a polynomial up to order four can be solved analytically through its radicals [161]. However, in the coding of telephone-band speech, the order of LP is typically $m = 8$ or $m = 10$, in order to represent the relevant formants in speech. The root finding methods must therefore be numerical.

Before a discussion of the different root finding algorithms, the reader should note that there are several facts known in advance about the roots of the LSP polynomials as described in in Section 4.1. Namely, the zeros are interlaced on the unit circle, and therefore, when the zeros of $P(z)$ are known, the required search area for the zeros of $Q(z)$ is roughly halved, and vice versa if the zeros of $Q(z)$ are located first. In addition, because $P(z)$ and $Q(z)$ are real, their zeros come in complex conjugate pairs, so that only half the number of zeros of the model order has to be located [132]. Furthermore, as mentioned before, [87] presents some properties of root loci.

The very first LSP polynomial root-finding algorithm was proposed in [132]. This algorithm basically utilises most of the properties described above. In short, the algorithm can be described as follows:

1. Calculate LP polynomial $A_m(z)$, and the corresponding LSP polynomials $P_{m+1}(z)$ and $Q_{m+1}(z)$ (Eq. 4.1).
2. Remove trivial zeros using Eq. 4.3 (polynomials are now denoted $R_P(z)$ and $R_Q(z)$).
3. Evaluate polynomials on unit circle using $z = e^{j\omega}$ which yields polynomials $R_P(\omega)$ and $R_Q(\omega)$.
4. Calculate the cosine transform of polynomials $R_P(\omega)$ and $R_Q(\omega)$, which yields polynomials $\hat{R}_P(x)$ and $\hat{R}_P(x)$ where $x = \cos\omega$, and thus

$x \in [-1, 1]$. The polynomials $\hat{R}_P(x)$ and $\hat{R}_P(x)$ are of order $m/2$ (for even m) and are thus much easier to evaluate than $R_P(z)$ and $R_Q(z)$.

5. Divide the range of x into an adequately fine grid and look for changes of sign.

6. Search for the zeros in the intervals with sign changes.

This algorithm is a rather brute-force method in the final stages and is therefore computationally fairly expensive.

A computationally more efficient approach using Chebyshev polynomials was proposed in [72]. Differing from antecedent methods, the proposed method does not include any calculations of trigonometric functions. The algorithm has two stages. Firstly, the polynomials $R_P(x)$ and $R_Q(x)$ are represented by Chebyshev polynomials. Secondly, the roots of this new representation are calculated with an efficient numerical algorithm. The precision of LSFs given by this method should be adequate, but if a higher accuracy is desired methods exist for more exact localisation of zeros, e.g. [21].

Further improvement to the Chebyshev polynomial approach is presented in [163], where a decimation in degree (DID) algorithm is used to simplify the equations. The DID algorithm closely resembles that of Chebyshev polynomials but is slightly simpler. In addition, it obviates prior storage or the large calculation of transcendental functions and complex computation and, of special importance, it provides more selections of rapid numerical methods. However, for model orders larger than $m = 8$, this algorithm still requires searching of zero crossings in an adequately fine grid. Some improvements to the DID algorithm are found in [122].

The zero finding algorithms (step 6 in above algorithm) are further developed in [123]. It introduces an accelerated variation of the classical Newton's method. The method is based on a double-step idea, where the gradient search is accelerated by taking double-length steps in the direction of the gradient. A zero crossing is indicated by changes in the sign of polynomial or gradient, and it is therefore possible to slow down to the classical Newton's method when the immediate vicinity of the zero is reached. The rationale in the double-step idea is similar to many convex optimisation algorithms, and further improvements could be achieved by studies in classical optimisation, e.g. [12].

By using a property similar to the inter-model interlacing property (see Section 4) the zero crossings search can be avoided for model order $m = 10$ [48]. The zeros of the lower order polynomial offer an ordered list of in-

tervals where the zeros of the full order polynomial can be found. (The actual full order polynomials used are of order $m/2 = 5$, and the lower order polynomial is $m/2 - 1 = 4$, and the roots can thus be solved analytically.) This approach presents a substantial improvement to the number of multiplications required for the zero finding problem, since it is possible to use gradient search algorithms directly, without an adequately fine grid. This method was later improved in [96].

A rather different approach to the root finding problem based on a quotient-difference algorithm is presented in [112]. This iterative approach calculates an estimate for *all* zeros on each iteration. It thus offers the possibility for a trade-off between accuracy and complexity.

4.2.1 Tracking Roots

Speech coding is usually performed by dividing long speech signals into windows and processing these windows separately. Assuming the windowing interval compared to phoneme length is short, the consecutive windows will be similar, and thus, consecutive LP models (and the corresponding LSP polynomials) will be similar in spectral structure. It should be obvious that usage of such information could improve the efficiency of the coding algorithm.

Several root tracking methods for general high-order polynomials have been proposed [165, 136], but the first tracking algorithm developed specifically for LSP was proposed in [115]. This approach defines continuous paths from known roots of the LSP polynomials of a prior speech frame to the unknown roots of the next frame in the sequence. A gradient search based numerical predictor-corrector procedure is used for tracking these paths in order to compute the unknown roots.

4.3 Immittance Spectrum Pair

Another representation of the LP coefficients is the Immittance Spectrum Pair (ISP), which is also closely related to LSP. The immittance model is defined using polynomials $F_m(z)$ and $G_m(z)$, which are defined as [19]

$$F_m(z) = A_m(z) + z^{-m}A_m(z), \quad G_m(z) = A_m(z) - z^{-m}A_m(z). \tag{4.8}$$

Note the similarity between $F_m(z)$ and $G_m(z)$ to the LSP polynomials $P_m(z)$ and $Q_m(z)$ in Eq. 4.1. The immittance polynomials $F_m(z)$ and $G_m(z)$ may

represent the wave field variables sound pressure and volume velocity, whose ratios are impedance or admittance at the glottis [19].

The immittance model is then defined as

$$I_p(z) = \frac{A_m(z) - z^{-m}A_m(z)}{A_m(z) + z^{-m}A_m(z)}. \tag{4.9}$$

This model has properties similar to the LSP polynomials; namely, if polynomial $A_m(z)$ is minimum-phase, then the zeros and poles of $I_p(z)$ are simple, on the unit circle and interlaced [127]. Additionally, it has trivial zeros and poles similar to the LSP polynomials. When polynomial $A_m(z)$ is stable, the immittance model can therefore be written for m even as as

$$I_m(z) = \frac{K(1-z^{-2})\prod_{i=1}^{(m/2)-1}\left(1-2z^{-1}x_{2i}+z^{-2}\right)}{\prod_{i=1}^{m/2}\left(1-2z^{-1}x_{2i-1}+z^{-2}\right)}. \tag{4.10}$$

A similar equation for m odd exists with different trivial zeros and poles. The coefficients x_i are the ISP coefficients and it holds that

$$-1 < x_{m-1} < \ldots < x_2 < x_1 < 1. \tag{4.11}$$

Apart from speech coding, immittance polynomial forms can also be used in efficient inversion of Toeplitz and quasi-Toeplitz matrices in manner similar to the Levinson-Durbin recursion [18].

4.4 Line and Immittance Spectrum Statistics

The significance of LSP polynomials can be appreciated when it is noted that the LSFs and ISP coefficients are statistically uncorrelated when estimated from stationary autoregressive processes, in contrast to other representations [39]. Simplifying, this means that a small error in one LSF or ISP coefficient distorts the spectrum only in a local region.

In [148], an extensive study of theoretical statistical properties of the LSP polynomial roots is presented. The results presents estimates of the LSF probability distribution functions (PDF) in terms of the corresponding autoregressive (AR) model. Furthermore, it states that both the LSFs and AR-parameters cannot follow the Gaussian distribution, but that this theoretical result does not inhibit almost-Gaussian PDF's for both.

4.5 Quantisation of LSFs

As mentioned before, the primary usage of LSP polynomials is in speech coding, where LSFs are used as a representation of the LP information. However, the LSFs cannot be transmitted with arbitrary accuracy, that is, the LSF locations have to be quantised to a fixed number of bits, before they can be transmitted. This section will present some of the methods developed for quantisation of LSFs.

4.5.1 Uniform Quantisation

The first choice for quantisation of LSFs is simply to quantise the value of the LSF directly. The properties of such a quantisation is presented in [132]. The LSFs are limited in range to the open interval $(0, 1)$ and should thus be suitable for quantisation. However, direct uniform quantisation of the LSFs does not use any *a priori* information. As described in Section 4.1, the LSFs are interlaced on the unit circle, and they are thus correlated (in the sense that knowing zeros of $P(z)$ gives us the intervals of the zero locations of $Q(z)$ and vice versa). To utilise this property, [132] proposes an LSF difference quantisation scheme. In this approach, the difference of subsequent LSFs is quantised instead of the value of each LSF. The histograms of the LSF differences are concentrated close to zero and usually have a smaller range of values, and thus requiring fewer bits for quantisation. The LSF difference methods yields a 30% efficiency improvement over LAR [132].

In LSF difference quantisation, only a partial range of the theoretically complete LSF difference is used, since large LSF differences are rare. However, these large LSF differences can sometimes occur and when present, the quantiser will produce a large spectral error. Additionally, since quantisers are designed using only single parameter spectral sensitivity measures, a moderate weighted Euclidean distance between two given sets of LSP parameters can produce an unexpected, large spectral error. To overcome these problems, [134] presents a coding scheme that finds an optimal combination of quantised values in the sense of spectral distortion, using delayed decision coding.

Another solution to the problem is presented in [142]. This approach utilises the fact that after quantising one LSF, the range of possible values for the next LSF is smaller. The proposed method, Forward Sequential Adaptive Quantisation, applies uniform quantisation for each LSF ω_i and uses (ω_i, π) as the range for the next LSF ω_{i+1}, thus making it possible to reduce the

number of bits required for later LSFs. This algorithm can be further improved by recognising two things. Firstly, we note that the lower frequencies are perceptually more important than the higher frequencies. Secondly, we note that as the quantisation range decreases, the accuracy improves (provided that the number of bits used does not decrease substantially). The Backward Sequential Adaptive Quantisation utilises this property by quantising LSFs starting from the last.

4.5.2 Non-Uniform Quantisation

An alternative to uniform quantisation is non-uniform quantisation, where probable LSF values are given better accuracy than rare values. It is well known that non-uniform quantisation is superior to uniform quantisation in the sense that it results in a smaller average quantisation error, except in the situation where the source has a uniform distribution [142]. Non-uniform quantisation can be applied either directly to the LSFs, the LSF differences, or the Forward/Backward Sequential Adaptive Quantisation algorithms. According to [142], the Backward Sequential Adaptive Quantisation algorithm seems to be the most prolific of these.

Further improvements are presented in [125], which proposes a non-uniform adaptive quantisation forward method. It is based on an improved LSF probability distribution (PDF) estimate using a non-parametric method. The optimum bit allocation is obtained by the Fox-Makhoul procedure.

A simpler algorithm with an optimal quantisation scheme has been presented in [133, 135]. The optimal bit allocation is calculated by iterative subdivisions of the LSF distribution. At 32 bits/frame, this approach reaches a 1 dB average log spectral distortion, a commonly accepted level for reproducing perceptually transparent spectral information.

4.5.3 Vector Quantisation

In vector quantisation, the LSFs denoted by x_i, are presented in vector form as

$$\mathbf{x} = [x_1, x_2, \ldots, x_p]^T, \tag{4.12}$$

where p is the number of LSFs and $\mathbf{x}$ belongs to the set $\mathbf{x} \in G = \{\mathbf{x} \in \mathbb{R}^p : 0 < x_1 < \ldots < x_p < \pi\}$ [114].

The fundamental idea in vector quantisation (VQ) is the fact that the set G is considerably smaller than $\mathbb{R}^p$, and the set of probable values of G is still

smaller. That is, the distributions of individual LSFs are not independent and the use of this information can improve the quantisation. However, the cost of vector quantisation methods is a prohibitively increased complexity in the algorithms [107].

For reducing the complexity of VQ algorithms, several different approaches have been proposed [107, 114]. These algorithms are sub-optimal and therefore they provide a trade-off between complexity and accuracy. For example, split vector quantisation divides the vector space into two smaller subspaces and quantises these spaces separately. The two spaces are slightly correlated and information is thus lost resulting in reduced accuracy. The trade-off is that the complexity is considerably reduced. This approach has been shown to yield better quality than other antecedent LP quantisers at 24 bits/frame [107].

Another approach is the multistage vector quantisation method which uses two (or more) consecutive quantisers. The first stage quantiser yields a rough quantisation of the vector and the second stage quantises the error of the first stage [107, 109]. Basically, it would be possible to use more consecutive quantisers but such constructions have not appeared in any pertinent literature.

Other versions of vector quantisation are, for example, structured VQ [83], segmented VQ-methods [53], multiple scale lattice VQ [155] and safety-net pyramid VQ [25].

Further improvement can be achieved through usage of inter-frame correlation. In other words, sounds often change slowly and consecutive frames are therefore often similar. Different approaches to usage of this information appears in [114, 37]. A comparison of intra- and inter-frame LSF quantisation is presented in [158].

Vector values are usually not transmitted as such, nor in some non-uniform quantisation, but with different code-book methods. The different vector values are given a code which is transmitted and the transmitter translates the code back to the vector form using a code-book. Such methods are described in, among others, [114, 52, 53], a two code-book approach in [120], and a more general approach (applied to a general case VQ instead of LSF specific VQ) is presented in [51].

Vector quantisation methods are usually optimised using some spectral distortion measure (see Section 3.3) to yield the best quantisation. A theoretical analysis of vector quantisation methods and corresponding spectral distortion measures is given in [43]. It also presents an extension to scalar spectral

sensitivity measures through a sensitivity matrix.

4.5.4 Perceptual Sensitivity

From the viewpoint of perception, the upper frequency range is not as important as the middle range. The relative hearing sensitivity at 4 kHz is approximately half that at 0.5 kHz [75]. LSFs in the upper range could therefore be represented with lesser accuracy, as in the Backward Sequential Adaptive Quantisation presented in Section 4.5.1. A cruder approach would be to omit the upper LSFs completely and replace them with either equally spaced LSFs or according to the statistical distribution of LSFs [75].

A quantisation is said to be transparent if the errors introduced by quantisation are inaudible. In terms of spectral distortion (see Section 3.3) the generally accepted limits for transparent quantisation are [107]:

- The average distortion is $SD \leq 1$ dB.
- There are no outliers having $SD > 4$ dB.
- The number of outliers having $2 \leq SD \leq 4$ dB is less than 2%.

Quantiser performance with a weighed spectral distortion measure is analysed in [29].

4.5.5 Source and Channel Coding

Mobile communication applications usually require not only source coding (i.e., quantisation of the LP), but also a separate coding for the communication channel. The channel coder makes the application more tolerant to transmission errors. Some attempts to combine these stages, source and channel coding, using Trellis coding has been presented in [129, 110, 81].

Another solution for increasing error tolerance is index assignment [67, 124], which refers to the indexing of the code-books used in vector quantisation. The purpose is to ensure that a small error in the code-book index produces a small error in the LSF-domain and thus a small spectral distortion.

4.6 Adaptive LSP Filters

Speech processing is most often done in a window-by-window fashion, and the LSP polynomials are thus calculated for each window separately. However, sometimes it is beneficial, or even required, to have a continuous estimate of the LSFs. An adaptive algorithm would provide such estimate. It is well known that the reflection coefficients of Levinson-Durbin recursion can be implemented by a lattice filter. The lattice filter structure can be used to create an adaptive LSP filter that yields continuous estimates of LSP parameters [24]. Better convergence can be achieved with a least mean square (LMS) type algorithm for end-point error [143].

The reader should note that adaptive LSP filters should not be confused with, for example, Forward Sequential Adaptive Quantisation (described in Section 4.5.1), which uses an adaptive method for quantisation accuracy, not for the LSP filter parameters.

4.7 Interpolation Properties of LSFs

In speech coding, LP parameters are usually calculated and transfered frame-wise. The frames are typically 20 ms in length, which is relatively long and can lead to large differences in the LP parameters of consecutive frames. The large differences can result in audible clicks or distortion on the frame borders. It is therefore often useful to interpolate parameters in such a way that the border frames are inaudible, e.g. in 5 ms sub-frames [151, 106, 68, 111]. Such interpolation properties of LSFs were first studied in [9].

To reiterate, LSFs are more suitable for interpolation than reflection coefficients, log area ratios and other representations [106, 108, 38]. It has also been shown that using interpolation weighting of frame energy can improve the interpolation process [38].

4.8 Spectral Distortion Measures Using LSP

Vector quantisation requires a distance measure between the original and quantised spectra. Such measures were presented in Section 3.3, but these formulae did not use any information gained from the LSP polynomial properties. A rigorous derivation using LSP polynomials explicitly is found in

[84] which yields the following distortion measure

$$d^2 = \sum_{i=1}^{m} \frac{\prod_{i+2j\geq 1, i+2j\leq m, j\neq 0} |c_i - c_{i+2j}|}{\prod_{i+2j+1\geq 0, i+2j+1\leq m+1} |c_i - c_{i+2j+1}|} dc_i^2 \tag{4.13}$$

where d^2 is the spectral distortion, m is model order, we have $dc_i = c_i - c_{i,opt}$, where $c_{i,opt}$ is the correct value of the coefficient and the c_i's are given by

$$\begin{aligned} P(z) &= \left(1 + z^{-1}\right) \prod_{i=1}^{m/2} \left(1 - 2c_{2i-1}z^{-1} + z^{-2}\right) \\ Q(z) &= \left(1 - z^{-1}\right) \prod_{i=1}^{m/2} \left(1 - 2c_{2i}z^{-1} + z^{-2}\right) \end{aligned} \tag{4.14}$$

for m even, where $P(z)$ and $Q(z)$ are as defined in Eq. 4.1.

Equation 4.13 is valid only for small quantisation errors, but is still rather complicated to use. A convenient approximation is

$$d^2 \approx \sum_{i=1}^{m} W(i) \frac{dc_i^2}{(c_{i-1} - c_i)(c_i - c_{i+1})} \tag{4.15}$$

where $W(i)$ is a weighting function, similar to that in Section 3.3.

A heuristically designed alternative is the Euclidean distance between an m'th order LSF vector $\omega = [\omega_1, \ldots, \omega_m]$ and the quantised $\hat{\omega} = [\hat{\omega}_1, \ldots, \hat{\omega}_m]$ which is defined as [39]

$$d^2 = \sum_{i=1}^{m} w_i (\omega_i - \hat{\omega}_i)^2 \tag{4.16}$$

where w_i is a weight vector reflecting the perceptual weight of each LSF. This weight function can be improved by applying a formulation of the LSF spectral sensitivity [157] or by using the intra-model interlacing property [86].

4.9 Applications and Interpretations of LSP

The primary application of LSP polynomials is speech coding, discussed in Section 4.5. There are, however, other usages of LSP, which will be discussed in this section together with some interpretations of LSP.

4.9.1 Speech and Speaker Recognition and Enhancement

The LSP polynomials and the LSFs contain all information of the spectral envelope found in the corresponding LP polynomial. The LSFs could therefore also be used in speech and speaker recognition applications. In speech recognition, the LSFs have been found to perform better than other LP representations [105, 90]. Furthermore, if speech is already coded, e.g., for transmission, speaker-independent speech recognition directly from the LSFs is more efficient than recognition by any other known methods from the reconstructed signal [26]. Moreover, LSFs can be used in speech enhancement (e.g. [99]) and speech analysis (e.g. [54]).

4.9.2 Time-Frequency Duality

The duality of time and frequency representations is well-known. In the case of LSP polynomials this can be appreciated by noting the correspondence between the zeros of the LSP generated from the frequency domain signal and zero crossings of the original signal [80]. That is, if we generate a LP model on the spectrum of a signal, the LP model (which we will call the dual LP model) will then describe the time domain envelope of the original signal. Furthermore, the zeros of LSP polynomials computed from the dual LP model will be located at the zero crossings of the original signal.

4.9.3 Filter Design

The stability properties provided by the LSP polynomials yield efficient tools for filter design applications. Article [144] presents one such application involving post-filter design.

Post-filters are used in speech decoders to weigh quantisation noise that arises from low bit-rate speech coding, such that it is perceptually less disturbing. The conventional LP-based ARMA-type spectral post-filter is, with constant scalars η and ν, described by

$$H(z) = A(z/\eta)/A(z/\nu), \tag{4.17}$$

which is a transformation that guarantees the stability of $H(z)$. The post-filter proposed by [144] is defined as

$$\omega'_n(i) = c_n(i)\omega_T(i) + [1 - c_n(i)]\,\omega(i), \qquad (i = 0, \ldots, m-1), \tag{4.18}$$

where ω is an input LSF, ω'_n an LSF after transformation, ω_T is a target LSF, c_n division ratio, m is model order and n the LSP transformation

block number. This transformation yields a continuous mapping that shifts modelled LSFs toward desired LSFs, and thus provides means for enhancing the spectral properties.

A filter design method that takes the effect of noise into account has been presented in [166].

List of Publications

P1 Bäckström, T., Alku, P., Paatero, T., Kleijn, B. W., "*A Time Domain Interpretation for the LSP Decomposition*", accepted to IEEE Transactions on Speech and Audio Processing, 2004.

P2 Kleijn, B. W., Bäckström, T., and Alku, P., "*On Line Spectral Frequencies*", IEEE Signal Processing Letters, vol. 10, no. 3, pp. 75–77, March 2003.

P3 Bäckström, T., and Alku, P., "*All-pole Modeling Technique Based on Weighted Sum of LSP Polynomials*", IEEE Signal Processing Letters, vol. 10, no. 6, pp. 180–183, June 2003.

P4 Alku, P., and Bäckström, T., "*Linear Predictive Method for Improved Spectral Modeling of Lower Frequencies of Speech with Small Prediction Orders*", accepted to IEEE Transactions on Speech and Audio Processing, March 2004.

P5 Bäckström, T., and Alku, P., "*A Constrained Linear Predictive Model with the Minimum-Phase Property*," Signal Processing, vol. 83, no. 10, pp. 2259–2264, October 2003.

P6 Bäckström, T., "*Root-exchange Property of Constrained Linear Predictive Models*", in Proc. of the 2003 IEEE Workshop on Statistical Signal Processing (SSP03), pp. 81–84, St. Louis, MO, September 28 – October 1, 2003.

Summary of Articles and Author's Contribution

This chapter provides a summary of the contents of articles included in this thesis as well as a description of the present author's contribution to those articles. Firstly, we will list each author's contribution to the articles and secondly, we will present an overview of the results in the two following sections. Finally, the last section attempts to give a forecast of possible further directions.

Authors' Contributions

The starting point for the current work was a connection between Line Spectrum Pair (LSP) polynomials and time domain constraints to the linear predictive (LP) problem that Prof. Paavo Alku had discovered. This connection was found in experiments with different formulations of the linear predictive problem. The present author managed to prove that connection mathematically in Publication P1. Prof. Kleijn and Tuomas Paatero assisted with corrections and clarifications.

The aforementioned connection between LSP and the time domain reformulation of LP was better formalised in Publication P2. The basic matrix form concept was provided by Prof. Kleijn, whereas the present author completed the new proof and extended the formulation to all types of time domain formulations of LP.

On a parallel track with the theoretical results, Prof. Paavo Alku applied the results to practical uses. In Publication P3, the new-found insight into the theory of linear prediction allowed a new type of optimisation criteria for the linear predictive model for speech envelope modelling. Prof. Paavo Alku implemented the model and made some experiments, while both the

idea behind the model and the supporting theory were contributed by the current author.

A similar approach was used in Publication P4, where again, Prof. Paavo Alku implemented the model and made some experiments, while both the idea behind the model and the supporting theory was contributed by the current author. However, the theoretical importance of Publication P3 is greater than that of Publication P4 since the latter is in a sense, an improvement on, or enhancement of, the ideas presented in the former. Consequently, the weight of Prof. Paavo Alku's contribution to Publication P4 was more significant than that of the present author's, whereas for Publication P3 the roles were reversed.

These theoretical insights leaped forward in Publication P5 which provides a complete review of the time domain constraints in matrix form based on the original idea of Prof. Kleijn. However, the most important results of Publication P5 are the stability criteria of the constrained LP models presented. These results were entirely contributions of the current author. Notably, however, a colleague and student in mathematics, Carlo Magi, had an important part in the verification of the results even though his name was regrettably not included in the author list.

Analysis of the framework presented in Publication P5 was continued in Publication P6 which relates the root location of the constraints to the predictor roots with a root-exchange rule. In addition, the root-exchange rule was applied to yield a new matrix decomposition for Toeplitz and Hankel matrices. These results of Publication P6 were entirely contributions of the current author.

Theoretical Results

The theoretical results of this work can be divided into three partly overlapping categories. Firstly, the connection between LSP and reformulations of LP provide insight into the rationale of LSP [P1]. While LSP had formerly been a rather ad hoc method, used in speech coding *since* it works, but without consideration as to *why* it works, the current results give a better rationale for its usage. Briefly, we have found that an LP predictor, when separated into its symmetric and antisymmetric parts, i.e. the LSP polynomials, corresponds to predictors which use, respectively, low-pass and high-pass data only. This could give us, for example, better information on the distribution of line spectral frequencies (cf. [39, 148]) and consequently aid in the design of speech coding applications.

Secondly, a significant part of this work is the formalisation of reformulations of the linear predictive problem. While many authors have worked with reformulations of linear prediction [154, 6, 5], the mathematical formulations have been rather complex. This work provides a simple and generic basis for analysis of such systems in matrix form and thus enables us to use the whole power of matrix analysis.

In addition, together with the third part of the theoretical results, this work provides basic tools for stability, i.e. minimum-phase property, analysis of linear predictive models [P3, P4, P5, P6]. Earlier works have considered only the theoretical stability of unconstrained linear predictive models [150, 91, 82, 104] and root loci of symmetric predictor polynomials [127, 7, 47, 132, 137, 139, 97, 77]. In contrast, we have studied the stability of more general models, using either known interlacing properties [P3, P4] or whole classes of models with the minimum-phase property [P5]. Furthermore, Publication [P6] present completely new theory in an analysis of Toeplitz matrices. These results employ the Toeplitz structure in a completely novel manner in order to make a connection to Vandermonde matrices [20, 46, 60]. This approach owes much to [55], where a similar but simplified idea was used to proof stability of the conventional linear predictive model.

Applications

While the most significant results of this work are, without doubt, in its theoretical parts, some preliminary experiments have been made [P3, P4]. The models have been applied to the speech spectral envelope modelling task, which is crucial for speech coding applications. The conventional linear predictive model gives equal weight to all spectral regions, but it is in our interest to emphasise those spectral features that are important for perception of speech. In this aim, both WLSP and LPLE models in Publications P3 and P4, respectively, perform better than conventional LP models. Specifically, both models separate the two first formants better than LP. This improvement comes, however, with some trade-offs since the formant band-width tends to decrease and sometimes the centre frequency of formants moves slightly. Nevertheless, since the format band-width problem can be solved by windowing and formant movement is in practise negligible, our view is that both models are well warranted.

Future Directions

It is a considerable understatement that there is a lot of work remaining. Perhaps the most obvious task is a systematic evaluation of constrained and other linear predictive models. These should be done both for isolated phonemes with objective measurements, as well as in a setting mimicking the real-life speech coding problem with the new models applied to existing speech codecs. However, the magintude of such a process would require at least another dissertation and cannot possibly be included in the current work.

Furthermore, while our theoretical results have been considered only from the speech modelling perspective, they are applicable to any area of problems where linear predictive models are used. Such areas include, but are not restricted to, economic and geological models, systems identification and control theory. The special requirements and possibilities of any such field are beyond the scope of our current theory. In fact, a bibliographical study of all applications of linear prediction would be a huge task in itself, let alone trying to apply our constrained models to more general problems.

A specific direction of work would be a more detailed analysis of the relation of root loci and constraints in linear predictive models. Some results in this direction, not included in this thesis, have already been published [13].

While it might, at first hand, seem far fetched, one possible practical application of the theory of Toeplitz and Vandermonde functionals in Publication P6, is the generation of new root-finding algorithms. By relating the roots and constraints of linear predictive models by using a mapping through Toeplitz matrices, we could obtain a new domain for the root-finding problem. In the mathematical world, the solution of polynomial roots can be considered, one of the most formidable problems, and any improvements to existing algorithms would surely grant eternal fame and ensure a place in heaven. Therefore, any novel attempts in that aim are surely worthwhile.

Bibliography

[1] *Recommendation GSM 06.10 "GSM full rate speech transcoding".* ETSI, 1992.

[2] *Recommendation G.729-Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP).* ITU-T, March 1996.

[3] *Digital cellular telecommunications system (Phase 2); Enhanced full rate (EFR) speech transcoding, GSM 06.60 version 4.0.0.* ETSI, August 1997.

[4] *3rd Generation Partnership Project: AMR Wideband Speech Codec, 3GPP TS 26.190 (Release 5).* ETSI, March 2001.

[5] P. Alku and U. Laine. Bi-directional linear prediction – a new LPC-based method for signal processing. In *Proc. AMSE International Conference Signals & Systems*, volume 1, pages 101–114, Brighton, UK, 1989.

[6] P. Alku and S. Varho. A new linear predictive method for compression of speech signals. In *Proc. International Conference on Spoken Language Processing (ICSLP)*, volume 6, pages 2563–2566, Sydney, Australia, 1998.

[7] T. V. Ananthapadmanabha and H. S. Srinivasa Chakravarty. Linear prediction based on square root autocorrelation. *Electro-Technology*, pages 31–34, September 1980.

[8] B. S. Atal. Determination of the vocal tract shape directly from the speech wave. *J. Acoust. Soc. Am.*, 47(65(A)), 1970.

[9] B. S. Atal, R. V. Cox, and P. Kroon. Spectral quantization and interpolation for CELP coders. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'89*, volume 1, pages 69–72, 1989.

[10] B. S. Atal and S. L. Hanauer. Speech analysis and synthesis by linear prediction of the speech wave. *J. Acoust. Soc. Am.*, 50:637–655, 1970.

[11] B. S. Atal and M. R. Schroeder. Predictive coding of speech signals and subjective error criteria. *IEEE Trans. Acoust. Speech Sig. Proc.*, ASSP-27(3), June 1979.

[12] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. *Nonlinear programming: Theory and algorithms*. J. Wiley & Sons, Inc., New York, 2nd edition, 1993.

[13] T. Bäckström and P. Alku. On the stability of constrained linear predictive models. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'03*, volume 6, pages 258–288, Hong Kong, April 6–10 2003.

[14] Y. Bistritz. Zero location with respect to the unit circle of discrete-time linear system polynomials. *Proc. IEEE*, 72(9):1131–1142, September 1984.

[15] Y. Bistritz. A modified unit-circle zero location test. *IEEE Trans. Circuits Systems*, 43(6):472–474, June 1996.

[16] Y. Bistritz. Reflections on schur-cohn matrices and jury-marden tables and classification of related unit circle zero location criteria. *Circuits Systems Signal Processing*, 72(9):1131–1142, September 1996.

[17] Y. Bistritz. Zero location of polynomials with respect to the unit-circle unhampered by nonessential singularities. *IEEE Trans. Circuits Systems*, 49(3):305–314, March 2002.

[18] Y. Bistritz, H. Lev-Ari, and T. Kailath. Immittance-domain Levinson algorithms. *IEEE Trans. Information Theory*, 35:675–682, May 1989.

[19] Y. Bistritz and S. Peller. Imittance spectral pairs (ISP) for speech encoding. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'93*, volume 2, pages 9–12, 1993.

[20] L. Brand. The companion matrix and its properties. *The American Mathematical Monthly*, 71(6):629–634, June–July 1964.

[21] R. P. Brent. An algorithm with guaranteed convergence for finding a zero of a function. *The Comput. J.*, 14:422–425, 1971.

[22] T. L. Burrows and M. Niranjan. Vocal tract modelling with recurrent neural networks. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'95*, volume 5, pages 3315–3318, 9–12 May 1995.

[23] A. Cantoni and P. Butler. Eigenvalues and eigenvectors of symmetric centrosymmetric matrices. *Linear Algebra and its Applications*, 13: 275–288, 1976.

[24] B. M. G. Cheetham. Adaptive LSP filter. *Electronics Letters*, 23(2): 89–90, 1987.

[25] S. Chi, S. Kang, and C. Lee. Safety-net pyramid VQ of LSF parameters for wideband speech codecs. *Electronics Letters*, 37(11):724–726, May 2001.

[26] S. H. Choi, H. K. Kim, H. S. Lee, and R. M. Gray. Speech recognition method using quantised LSP parameters in CELP-type coders. *Electronics Letters*, 34(2):156–157, 1998.

[27] P. L. Chu and D. G. Messerschmitt. A frequency weighted Itakura-Saito spectral distance measure. *IEEE Trans. Acoust., Speech, Signal Processing*, 30:545–560, August 1982.

[28] P. L. Chu and D. G. Messerschmitt. Frequency weighted linear prediction. In *Proc. ICASSP'82*, pages 1318–1321, May 1982.

[29] R. P. Cohn and J. S. Collura. Incorporating perception into LSF quantization some experiments. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'97*, volume 2, pages 1347–1350, 1997.

[30] J. S. Collura, A. McCree, and T.E. Tremain. Perceptually based distortion measurements for spectrum quantization. In *Proc. IEEE Workshop Speech Coding for Telecomm.*, pages 49–50, 1995.

[31] R. V. Cox. Speech coding standards. In W. B. Kleijn and K. K. Paliwal, editors, *Speech Coding and Synthesis*, pages 49–78. Elsevier, Amsterdam, The Netherlands, 1995.

[32] S. David and B. Ramamurthi. Two-sided filters for frame-based prediction. *IEEE Trans. Signal Processing*, 39(4):789–794, 1991.

[33] P. Delsarte and Y. V. Genin. The split Levinson algorithm. *IEEE Trans. Acoust. Speech Sig. Proc.*, ASSP-34(3):470–478, 1986.

[34] J. Durbin. The fitting of time series models. *Rev. Internat. Statist. Inst.*, 23:233–244, 1960.

[35] A. El-Jaroudi and J. Makhoul. Discrete all-pole modeling for voiced speech. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'87*, volume 12, pages 320–323, April 1987.

[36] A. El-Jaroudi and J. Makhoul. Discrete all-pole modeling. *IEEE Trans. Signal. Proc.*, 39:411–423, 1991.

[37] T. Eriksson, J. Linden, and J. Skoglund. Interframe LSF quantization for noisy channels. *IEEE Trans. Speech Audio Proc*, 7(5):495–509, Sept. 1999.

[38] J. S. Erkelens and P. M. T. Broersen. Analysis of spectral interpolation with weighting dependent on frame energy. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'94*, volume 1, pages 481–484, 1994.

[39] J. S. Erkelens and P. M. T. Broersen. On the statistical properties of line spectrum pairs. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'95*, volume 1, pages 768–771, 1995.

[40] G. Fant. *Acoustic Theory of Speech Production.* Mouton, The Hague, 1960.

[41] J. L. Flanagan. *Speech Analysis Synthesis and Perception.* Springer Verlag, Berlin, 1972.

[42] A. Fraser. Modeling nonlinear time series. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'92*, volume 5, pages 313–316, 23–26 March 1992.

[43] W. R. Gardner and B. D. Rao. Theoretical analysis of the high-rate vector quantization of LPC parameters. *IEEE Trans. Speech Audio Proc.*, 3(5), Sept. 1995.

[44] E. E. Kuruoğlu, W. J. Fitzgerald, and P. J. W. Rayner. Nonlinear autoregressive modeling of non-gaussian signals using l_p-norm techniques. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'97*, volume 5, pages 3533–3536, 21–24 April 1997.

[45] R. Gnanasekaran. A note on the new 1-D and 2-D stability theorems for discrete systems. *IEEE Trans. Acoust. Speech Signal Proc.*, ASSP-29(6):1211–1212, December 1981.

[46] G. Golub and C. van Loan. *Matrix Computations.* The Johns Hopkins University Press, London, 1996.

[47] D. M. Goodman and E. K. Miller. A note on minimizing the prediction error when the zeros are restricted to the unit circle. *IEEE Trans. Acoust. Speech Signal Proc.*, ASSP-30(3):503–505, June 1982.

[48] S. Grassi, A. Dufaux, M. Ansorge, and F. Pellandini. Efficient algorithm to compute LSP parameters from 10th-order LPC coefficients. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'97*, volume 3, pages 1547–1550, 1997.

[49] A. H. Gray and J. D. Markel. Distance measures for speech processing. *IEEE Trans. Acoust. Speech Signal Proc.*, ASSP-24:380–391, Oct. 1976.

[50] R. M. Gray. *Toeplitz and Circulant Matrices: A Review*. Information Systems Laboratory, Stanford University, `http://www-ee.stanford.edu/~gray/toeplitz.html`, 2002. published on the internet only.

[51] R. M. Gray and Y. Linde. Vector quantizers and predicitve quantizers for gauss-markov sources. *IEEE Trans. Commun.*, COM-30(2):381–389, Feb. 1982.

[52] R. Hagen and P. Hedelin. Low bit-rate spectral coding in CELP, a new LSP-method. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'90*, volume 1, pages 189–192, 1990.

[53] R. Hagen and P. Hedelin. Spectral coding by LSP frequencies – scalar and segmented vq-methods. In *IEEE Proc Comm Speech and Vision*, volume 139, pages 118–122, 1992.

[54] M. Hasegawa-Johnson. Line spectral frequencies are poles and zeros of the glottal driving-point impedance of a discrete matched-impedance vocal tract model. *J. Acoust. Soc. Amer.*, 108:457–460, July 2000.

[55] M. H. Hayes. *Statistical Digital Signal Processing and Modeling*. J. Wiley & Sons, Inc., New York, 1996.

[56] M. H. Hayes and M. A. Clements. An efficient algorithm for computing Pisarenko's harmonic decomposition using Levinson's recursion. *IEEE Trans. Acoustics, Speech and Signal Proc.*, ASSP-34:470–478, 1986.

[57] Simon Haykin. *Neural Networks – A Comprehensive Foundation*. Prentice-Hall, Upper Saddle River, NJ, 2nd edition, 1999.

[58] H. Hermansky. Perceptual linear predictive (PLP) analysis of speech. *J. Acoust. Soc. Amer.*, 87(4):1738–1751, 1990.

[59] H. Hermansky, H. Fujisaki, and Y. Sato. Spectral transform linear prediction method for analysis-synthesis of speech. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'83*, pages 43, 1–4, Paris, France, 1983.

[60] R. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, UK, 1999.

[61] S.-H. Hou and W.-K. Pang. Inversion of confluend vandermonde matrices. *Computers & Mathematics*, 43:1539–1547, 2002.

[62] A. Härmä. Linear predictive coding with modified filter structures. *IEEE Trans. Speech Audio Proc.*, 9(8):769–777, November 2001.

[63] A. Härmä and U. K. Laine. A comparison of warped and conventional linear predictive coding. *IEEE Trans. Speech Audio Proc.*, 9(5):579–588, July 2001.

[64] J.-J. Hsue and A. E. Yagle. Blind deconvolution of symmetric noncausal impulse responses using two-sided linear prediction. *IEEE Trans. Signal Processing*, 42(6):1509–1518, June 1994.

[65] Y. H. Hu. Toeplitz eigensystem solver. *IEEE Trans. Acoustics, Speech and Signal Proc.*, ASSP-33(4):1264–1271, October 1985.

[66] D. Huang. Symmetric solutions and eigenvalue problems of Toeplitz systems. *IEEE Trans. Signal Processing*, 40(12):3069–3074, December 1992.

[67] R. Iordache and I. Tabus. Index assignment for transmitting vector quantized LSF parameters over binary markov channels. In *Proc. IEEE Circuits and Systems ISCAS '99*, volume 4, pages 544–547, 1999.

[68] T. Islam and P. Kabal. Partial-energy weighted interpolation of linear prediction coefficients. In *Proc. IEEE Workshop on Speech Coding*, pages 105–107, Delavan, Wisconsin, Sept. 2000.

[69] F. Itakura. Line spectrum representation of linear predictive coefficients of speech signals. *J. Acoust. Soc. Amer.*, 57, suppl. 1:35, 1975.

[70] F. Itakura and S. Saito. Analysis synthesis telephony based on the maximum likelihood method. In *Proc. 6th Int'l Congress of Acoustics*, pages C–17–C–20, Tokyo, Japan, August 1968.

[71] A. H. Gray Jr. and J. D. Markel. Quantization and bit allocation in speech processing. *IEEE Trans. Acoust Speech Signal Processing*, ASSP-24(6):459–473, Dec. 1976.

[72] P. Kabal and R. P. Ramachandran. The computation of line spectral frequencies using Chebyshev polynomials. *IEEE Trans. Acoust. Speech Sig. Proc.*, ASSP-34(6):1419–1426, 1986.

[73] T. Kailath. *Linear systems*. Prentice-Hall, Englewood Cliffs NJ, 1980.

[74] T. Kailath and A. H. Sayed, editors. *Fast Reliable Algorithms for Matrices with Structure*. SIAM Society for Industrial and Applied Mathematics, Philadelphia, USA, 1999.

[75] G. S. Kang and L. J. Fransen. Application of line-spectrum pairs to low-bit-rate speech encoders. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'85*, volume 1, pages 244–247, 1985.

[76] S. Katagiri, editor. *Handbook of Neural Networks for Speech Processing*. Artech House, London, 2000.

[77] H. K. Kim and H. S. Lee. Interlacing properties of line spectrum pair frequencies. *IEEE Trans. Speech Audio Proc.*, 7(1):87–91, 1999.

[78] E. Kreyszig. *Advanced engineering mathematics*. John Wiley & Sons, Inc., New York, 1993.

[79] H. Krishna. New split Levinson, Schur and lattice algorithms for digital signal processing. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'88*, volume 3, pages 1640–1642, 1988.

[80] R. Kumaresan and Y. Wang. On the duality between line-spectral frequencies and zero-crossings of signals. *IEEE Trans. Speech Audio Proc.*, 9(4):458–461, 2001.

[81] F. Lahouti and A. K. Khandani. Quantisation of LSF parameters using a Trellis modeling. *IEEE Trans. Speech Audio Proc.*, 11(5):400–412, September 2003.

[82] S. W. Lang and J. H. McClellan. A simple proof of stability for all-pole linear prediction models. *Proceedings of the IEEE*, 67(5):860–861, May 1979.

[83] R. Laroia, N. Phamdo, and N. Farvardin. Robust and efficient quantization of speech LSP parameters using structured vector quantizers. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'91*, volume 1, pages 641–644, 1991.

[84] P. A. Laurent. Expression of spectral distortion using line spectral frequencies. *IEEE Trans. Speech Audio Proc.*, 5(5):481–484, 1997.

[85] A.-C. Lee. A new autoregressive method for high-perfomance spectrum analysis. *J. Acoust. Soc. Am.*, 86(1):150–157, 1989.

[86] M. S. Lee, H. K. Kim, and H. S. Lee. A new distortion measure for spectral quantization based on the LSF intermodel interlacing property. *Speech Commun.*, 35:191–202, 2001.

[87] A. Lepschy, G. A. Mian, and U. Viaro. A note on line spectral frequencies. *IEEE Trans. Acoust. Speech Audio Proc.*, 36(8):1355–1357, Aug. 1988.

[88] S. H. Leung, H. C. Ng, and K. F. Wong. Speech synthesis using two-sided linear prediction parameters. In *Proc. Speech, Image Processing and Neural Networks, ISSIPNN'94*, Hong Kong, April 1994.

[89] N. Levinson. The Wiener RMS error criterion in filter design and prediction. *J. Math. Phys.*, 25:261–278, 1947.

[90] C. S. Liu, M. T. Lin, W. J. Wang, and H. C. Wang. Study of line spectrum pair frequencies for speaker recognition. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'90*, pages 277–280, 1990.

[91] J. Makhoul. Linear prediction: A tutorial review. *Proc. IEEE*, 63(5): 561–580, April 1975.

[92] J. Makhoul. On the eigenvectors of symmetric Toeplitz matrices. *IEEE Acoust Speech Signal Processing*, ASSP-29(4):868–872, Aug 1981.

[93] J. Makhoul. Toeplitz determinants and positive semidefiniteness. *IEEE Trans. Signal Processing*, 39(3):743–746, March 1991.

[94] J. Makhoul and L. Cosell. LPCW: An LPC vocoder with linear predictive spectral warping. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'84*, volume 1, pages 466–469, April 1976.

[95] J. D. Markel and A. H. Gray. *Linear prediction of speech.* Springer, Berlin, 2nd edition, 1980.

[96] X.-D. Mei and S.-H. Sun. An efficient method to compute LSFs from LPC coefficients. In *Int Conf Sig Processing Proc. WCCC-ICSP 2000*, volume 2, pages 655–658, 2000.

[97] G. A. Mian and G. Riccardi. A localization property of line spectrum frequencies. *IEEE Trans. Speech Audio Proc.*, 2(4):536–539, 1994.

[98] T. K. Moon and W. C. Stirling. *Mathematical Methods and Algorithms for Signal Processing.* Prentice Hall, Upper Saddle River, NJ, 2000.

[99] R. W. Morris and M. A. Clements. Modification of formants in the line spectrum domain. *IEEE Signal Processing Letters*, 9(1):19–21, January 2002.

[100] M. Niranjan and V. Kadirkamanathan. A nonlinear model for time series prediction and signal interpolation. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'91*, volume 3, pages 1713–1716, 14–17 April 1991.

[101] K. C. Nisbet, B. Mulgrew, and S. McLaughlin. Reduced state methods in nonlinear prediction. *Signal Processing*, 48:37–49, 1996.

[102] A. V. Oppenheim, A. S. Wilsky, and S. H. Nawab (2nd ed.). *Signals and systems*. Prentice-Hall, London, UK, 2nd edition, 1983.

[103] D. O'Shaughnessy. *Speech Communication – Human and Machine*. Addison-Wesley, New York, 1987.

[104] L. Pakula and S. Kay. Simple proofs of the minimum phase property of the prediction error filter. *IEEE Trans. Acoust. Speech Signal Proc.*, ASSP-31(2):501, April 1983.

[105] K. K. Paliwal. A study of line spectrum pair frequencies for vowel recognition. *Speech Commun.*, 8:27–33, 1989.

[106] K. K. Paliwal. Interpolation properties of linear prediction parametric representations. In *Proc. IEEE Speech Coding Workshop*, Annapolis, MD, 1995.

[107] K. K. Paliwal and B. S. Atal. Efficient vector quantization of LPC parameters at 24 bits/frame. *IEEE Trans. Speech Audio Proc.*, 1(1): 3–14, 1993.

[108] K. K. Paliwal and W. B. Kleijn. Quantization of LPC parameters. In W. B. Kleijn and K. K. Paliwal, editors, *Speech Coding and Synthesis*, pages 433–466. Elsevier, Amsterdam, 1995.

[109] J. Pan and T. Fischer. Vector quantization of speech line spectrum pair parameters and reflection coefficients. *IEEE Trans. Speech Audio Proc.*, 6(4):106–115, 1998.

[110] J. Pan and T. R. Fischer. Vector quantization of speech LSP parameters using Trellis codes and l_1-norm constraints. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'93*, volume 2, pages 17–20, 1993.

[111] W. Pereira and P. Kabal. Improved spectral tracking using interpolated linear prediction parameters. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'02*, volume 1, pages 261–264, Orlando, Florida, May. 2002.

[112] D. Petrinovic. Calculation of the line spectrum frequencies using the quotient-difference scheme. In *Proc 10th Mediterranean Electrotech Conf. MELECON 2000*, volume 2, pages 831–834, 2000.

[113] D. Petrinovic. Discrete weighted mean square all-pole modeling. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'03*, volume 1, pages 828–831, Hong Kong, April 2003.

[114] N. Phamdo and N. Farvardin. Coding of speech LSP parameters using TSVQ with interblock noiseless coding. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'90*, volume 1, pages 193–196, 1990.

[115] U. Pillai and V. Stonick. A scalar homotopy method for parallel and robust tracking of line spectral pairs. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'96*, volume 2, pages 805–808, 1996.

[116] S. P. Pindryck and D. L. Rubinfeld. *Econometric models and econometric forecasts*. McGraw-Hill, New York, 1998.

[117] V. F. Pisarenko. The retrieval of harmonics from a covariance function. *Geophys. J. R. Astr. Soc.*, 33:347–366, 1973.

[118] L. R. Rabiner and B. H. Juang. *Fundamentals of speech recognition*. Prentice-Hall, Englewood Cliffs, New Jersey, 1993.

[119] L. R. Rabiner and R. W. Schafer. *Digital processing of speech signals*. Prentice-Hall, Englewood Cliffs, New Jersey, 1978.

[120] R. P. Ramachandran, M. M. Sondhi, N. Seshadri, and B.S. Atal. A two codebook format for robust quantization of line spectral frequencies. *IEEE Trans. Speech Audio Proc.*, 3(3), May 1995.

[121] T. D. Rossing. *The Science of Sound*. Addison-Wesley, Reading, MA, 2nd edition, 1990.

[122] J. Rothweiler. On polynomial reduction in the computation of LSP frequencies. *IEEE Trans. Speech Audio Proc.*, 7(5):592–594, 1999.

[123] J. Rothweiler. A rootfinding algorithm for line spectral frequencies. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'99*, volume 2, pages 661–664, 1999.

[124] V. T. Ruoppila and S. Ragot. Index assignment for predictive wideband LSF quantization. In *Proc IEEE Workshop Speech Coding*, pages 108–110, 2000.

[125] S. Saoudi, J. M. Boucher, and A. Le Guyader. Optimal scalar quantization of the LSP and the LAR for speech coding. In *Proc. ICSLP*, volume 1, pages 101–103, Kobe, Japan, Nov. 1990.

[126] K. Schnell and A. Lacroix. Analysis o f lossy vocal tract models for speech production. In *Proc. 8th European Conference on Speech Communication and Technology (EUROSPEECH 2003)*, pages 2369–2372, Sept. 1-4 2003.

[127] H. W. Schüssler. A stability theorem for discrete systems. *IEEE Trans. Acoust. Speech Signal Proc.*, ASSP-24(1):87–89, February 1976.

[128] T. Söderstrom and P. Stoica. *System Identification*. Prentice Hall, New York, 1989.

[129] P. Secker and A. Perkis. Joint source and channel trellis coding of line spectrum pair parameters. *Speech Commun.*, 11:149–158, 1992.

[130] T. Shimamura and H. Hayakawa. Adaptive nonlinear prediction based on order statistics for speech signals. In *Proc. Eurospeech'99*, pages 347–350, Budapest, Hungary, 1999.

[131] A. C. Singer, G. W. Wornell, and A. V. Oppenheim. Codebook prediction: a nonlinear signal modeling paradigm. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'92*, volume 5, pages 325–328, 23-26 March 1992.

[132] F. K. Soong and B.-H. Juang. Line spectrum pair (LSP) and speech data compression. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'84*, volume 1, pages 1.10.1–1.10.4, San Diego, CA, March 1984.

[133] F. K. Soong and B.-H. Juang. Optimal quantization of LSP parameters (speech coding). In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'88*, volume 1, pages 394–397, 1988.

[134] F. K. Soong and B. H. Juang. Optimal quantization of LSP parameters using delayed decisions. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'90*, volume 1, pages 185–188, 1990.

[135] F. K. Soong and B. H. Juang. Optimal quantization of LSP parameters. *IEEE Trans. Speech Audio Proc.*, 1(1):15–24, 1993.

[136] D. Starer and A. Nehorai. High-order polynomial root tracking algorithm. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'92*, volume 4, pages 465–268, 1992.

[137] P. Stoica and A. Nehorai. The poles of symmetric linear prediction models lie on the unit circle. *IEEE Trans. Acoust. Speech Signal Proc.*, ASSP-34:1419–1426, Dec. 1986.

[138] P. Stoica and A. Nehorai. On stability and root location of linear prediction models. *IEEE Trans. Acoust. Speech Signal Proc.*, ASSP-35(4):582–584, April 1987.

[139] P. Stoica and A. Nehorai. On linear prediction models constrained to have unit-modulus poles and their use for sinusoidal frequency estimation. *IEEE Trans. Acoust. Speech Signal Proc.*, 36(6):940–942, June 1988.

[140] B. Story, I. R. Titze, and E. A. Hoffman. Vocal tract area functions from magnetic resonance imaging. *J. Acoust. Soc. Amer.*, 100(1):537–554, 1996.

[141] H. W. Strube. Linear prediction on a warped frequency scale. *J. Acoust. Soc. Am.*, 68(4):1071–1076, 1980.

[142] N. Sugamura and N. Farvardin. Quantizer design in LSP speech analysis-synthesis. *IEEE J. Selected Areas in Comm.*, 6(2):432–440, 1992.

[143] J. Taihui, Z. Youwei, and G. Junying. A new algorithm for adaptive line spectral pair filter and its convergence behaviour. In *Int. Conf. Communication Technology Proc., ICCT'96*, volume 1, pages 436–440, 1996.

[144] H. Tasaki, K. Shiraki, K. Tomita, and S. Takahashi. Spectral postfilter design based on LSP Transformation. In *Proc. IEEE Workshop Speech Coding For Telecomm.*, pages 57–58, 1997.

[145] J. Thyssen, H. Nielsen, and S. D. Hansen. Non-linear short-term prediction in speech coding. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'94*, volume 1, pages I/185–I/188, 19–22 April 1994.

[146] I. R. Titze. *Principles of Voice Production.* Prentice-Hall, Englewood Cliffs, NJ, 1994.

[147] K. Tom, I. R. Titze, E. A. Hoffman, and B. H. Story. Three-dimensional vocal tract imagining and formant structure: Varying vocal regster, pitch, and loudness. *J. Acoust. Soc. Amer.*, 109(2): 742–747, 2001.

[148] J.-W. Tourneret. Statistical properties of line spectrum pairs. *Signal Processing*, 65:239–255, 1998.

[149] B. Townshend. Nonlinear prediction of speech. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'91*, volume 1, pages 425–428, 14–17 April 1991.

[150] S. A. Tretter. The all-pole model is theoretically stable. *IEEE Trans. Audio and Electroacoustics*, AU-20(4):316, Oct. 1972.

[151] T. Umezaki and F. Itakura. Analysis of time fluctuating characteristics of linear predictive coefficients. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'86*, volume 2, pages 1257–1260, Tokyo, Japan, 1986.

[152] S. Varho. *New Linear Predictive Methods for Digital Speech Processing.* PhD thesis, Helsinki University of Technology, Helsinki, Finland, 2001.

[153] S. Varho and P. Alku. A linear predictive method using extrapolated samples for modelling of voiced speech. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 13–16, New Paltz, NY, 1997.

[154] S. Varho and P. Alku. Separated linear prediction – a new all-pole modeling technique for speech analysis. *Speech Commun.*, 24:111–121, 1998.

[155] A. Vasilache, M. Vasilache, and I. Tabus. Predictive multiple-scale lattice VQ for LSF quantization. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'99*, volume 2, pages 657–660, 1999.

[156] R. Viswanathan and J. Makhoul. Quantization properties of transmission parameters in linear predictive systems. *IEEE Trans. Acoust Speech Signal Processing*, ASSP-23(3):309–321, June 1975.

[157] H. L. Vu and L. Lois. A new general distance measure for quantization of LSF and their transformed coefficients. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'98*, volume 1, pages 45–48, 1998.

[158] J. Wang and J.D. Gibson. Performance comparison of intraframe and interframe LSF quantization in packet networks. In *Proc IEEE Workshop Speech Coding*, pages 126–128, 2000.

[159] S. Wang, E. Paksoy, and A. Gersho. Performance of nonlinear prediction of speech. In *Proc. 3rd International Conference on Spoken Language Processing (ICSLP'90)*, volume 1, pages 29–32, Kobe, Japan, 1990.

[160] Y. Wang, H. Krishna, and B. Krishna. Split Levinson algorithm is weakly stable. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'89*, pages 1215–1218, Glasgow, UK, 1989.

[161] L. Weisner. *Introduction to the Theory of Equations.* Macmillan, NY, 1938.

[162] H. E. Witzgall and J. S. Goldstein. Reduced rank predictive source coding. In *Proc. IEEE Workshop on Statistical Signal Processing*, pages 225–228, St. Louis, Missouri, 2003.

[163] C.-H. Wu and J.-H. Chen. A novel two-level method for the computation of the LSP frequencies using a decimation-in-degree algorithm. *IEEE Trans. Speech Audio Proc.*, 5:106–115, Mar. 1997.

[164] L. Wu and M. Niranjan. On the design of nonlinear speech predictors with recurrent nets. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'94*, volume 2, pages II/529–II/532, 19–22 April 1994.

[165] J. F. Yang and M. Kaveh. Adaptive algorithms for tracking roots of spectral polynomials. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'89*, volume 2, pages 1162–1165, 1989.

[166] A.-T. Yu and H.C. Wang. Model adaptation in line spectrum domain. In *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'00*, volume 3, pages 1547–1550, 2000.