Thesis for the degree of Doctor of Philosophy

Radio Channel Prediction Based on Parametric Modeling

by

Ming Chen

Department of Signals and Systems
Signal Processing Group
CHALMERS UNIVERSITY OF TECHNOLOGY

Göteborg, Sweden 2007

Radio Channel Prediction Based on Parametric Modeling
Ming Chen
ISBN 978-91-7385-009-4

This thesis has been prepared using LaTeX.

Doktorsavhandlingar vid Chalmers Tekniska Högskola
Ny serie nr 2690.
ISSN 0346-718X

Department of Signals and Systems
Signal Processing Group
Chalmers University of Technology
SE-412 96 Göteborg, Sweden

tel: +46 31 772 1000
fax: +46 31 772 1782
email: ming.chen@chalmers.se

Cover:
Frequency response vs. time for a wideband multipath channel.

Printed by Chalmers Reproservice
Göteborg, Sweden 2007

To Hong

Abstract

Long range channel prediction is a crucial technology for future wireless communications. The prediction of Rayleigh fading channels is studied in the frame of parametric modeling in this thesis.

Suggested by the Jakes model for Rayleigh fading channels, deterministic sinusoidal models were adopted for long range channel prediction in early works. In this thesis, a number of new channel predictors based on stochastic sinusoidal modeling are proposed. They are termed conditional and unconditional LMMSE predictors respectively. Given frequency estimates, the amplitudes of the sinusoids are modeled as Gaussian random variables in the conditional LMMSE predictors, and both the amplitudes and frequency estimates are modeled as Gaussian random variables in the unconditional LMMSE predictors. It was observed that a part of the channels cannot be described by the periodic sinusoidal bases, both in simulations and measured channels. To pick up this un-modeled residual signal, an adjusted conditional LMMSE predictor and a Joint LS predictor are proposed.

Motivated by the analysis of measured channels and recently published physics based scattering SISO and MIMO channel models, a new approach for channel prediction based on non-stationary Multi-Component Polynomial Phase Signal (MC-PPS) is further proposed. The so-called LS MC-PPS predictor models the amplitudes of the PPS components as constants. In the case of MC-PPS with time-varying amplitudes, an adaptive channel predictor using the Kalman filter is suggested, where the time-varying amplitudes are modeled as auto-regressive processes. An iterative detection and estimation method of the number of PPS components and the orders of polynomial phases is also proposed. The parameter estimation is based on the Nonlinear LS (NLLS) and the Nonlinear Instantaneous LS (NILS) criteria, corresponding to the cases of constant and time-varying amplitudes, respectively.

The performance of the proposed channel predictors is evaluated using both synthetic signals and measured channels. High order polynomial phase parameters are observed in both urban and suburban environments. It is observed that the channel predictors based on the non-stationary MC-PPS models outperform the other predictors in Monte Carlo simulations and examples of measured urban and suburban channels.

Keywords: Wireless communications, channel modeling, channel prediction, polynomial phase signal, sinusoidal modeling, MIMO, GLRT, CRLB, Kalman filter, NLLS, NILS.

Contents

Acknowledgements

To make this thesis possible, many people have helped and for this I would like to express my appreciation.

First, I would like to thank my supervisor, Prof. Mats Viberg, for letting me join the signal processing group at Chalmers, and guiding me through the interesting world of signals. Mats is such a remarkably good teacher, that he is always able to make the most abstract and complicated concepts and theories easy to understand. The discussions with him have always been interesting, heuristic, and encouraging. In fact, besides from the technical knowledge that one can learn from Mats, one can also learn from his unique positive attitude to life and work. Personally I also owe Mats for developing my taste for wines at the great social activities he organized.

I am grateful to Mr. Stefan Felter who previously worked at Ericsson Research and Dr. Torbjörn Ekman at NTNU in Trondheim, Norway, for their valuable input to our coauthored papers. To Stefan, I am also grateful for all the other interesting conversations about almost everything as well.

I would like to sincerely thank Prof. Mikael Sternad at Uppsala University and all other professors and Ph.D students in the Wireless IP project for the intriguing discussions during the project meetings.

I owe many thanks to my previous and current colleagues in the signal processing group for their help at different time and different aspects during my study. Special thanks goes to Dr. Lennart Svensson and Andy Backhouse for their proofreading of a early version of the thesis. I would also like to thank Agneta Kinnander and Lars Börjesson for their help with many practical issues.

Many thanks also goes to my friends in Göteborg for their friendship and help in various aspects of life. The list of the names will be too long, but some of them are Rui Li, Quan Zheng, and Yaoxin Bai.

Special thanks goes to my uncle Suiji Liu in China for his abundant help and support on many family issues.

Finally, I would like to express my deepest appreciation to my family for always being there for me. Especially, my thanks and love goes to Hong for her support, patience and love.

Ming Chen
Göteborg, September 2007

Abbreviations and Acronyms

ACF	Auto-Correlation Function
ANMSE	Adjusted Normalized Mean Square Error
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BF	Beam Forming
BS	Base Station
CDF	Cumulative Density Function
CRLB	Cramér-Rao Lower Bound
CSI	Channel Status Information
DDSR	Double Direction Single Reflection
DFT	Discrete Fourier Transform
DOA	Direction-Of-Arrival
DOD	Direction-Of-Departure
DPS	Discrete Prolate Spheroidal
DSR	Down Sampling Ratio
ESPRIT	Estimation of Signal Parameters via Rotation Invariance Techniques
EV	EigenVector
EVD	EigenValue Decomposition
FIR	Finite Impulse Response
GLRT	Generalized Likelihood Ratio Test
HAF	High-order Ambiguity Function
HIM	High order Instantaneous Moment
IID	Independent Identical Distributed
IVM	Instrumental Variable Method
JMAS	Joint Moving Average and Sinusoidal
LMMSE	Linear Minimum Mean Square Error
LMS	Least Mean Square
LOS	Line-Of-Sight

LP	Linear Prediction
LS	Least Square
MA	Moving Average
MARS	Multivariate Adaptive Regression Splines
MC-PPS	Multi-Component Polynomial Phase Signal
MIMO	Multiple-In-Multiple-Out
MLE	Maximum Likelihood Estimation
MMSE	Minimum Mean Square Error
MPS	MUSIC Pseudo Spectrum
MSE	Mean Square Error
MUSIC	MUltiple SIgnal Classification
NILS	Nonlinear Instantaneous Least Square
NLLS	NonLinear Least Square
NLOS	Non-Line-Of-Sight
NMSE	Normalized Mean Square Error
NSE	Normalized Square Error
OFDM	Orthogonal Frequency Division Multiplexing
PDF	Probability Density Function
PPS	Polynomial Phase Signal
PPT	Polynomial Phase Transform
PSK	Phase-Shift Keying
QAM	Quadrature Amplitude Modulation
RELAX	RELAXation algorithm
RI	Rotation Invariant
RMSE	Root Mean Square Error
SAGE	Space-Alternating Generalized Expectation-maximization
SAR	Synthetic Aperture Radar
SIMO	Single-In-Multiple-Out
SISO	Single-In-Single-Out
SNR	Signal Noise Ratio
SVD	Singular Value Decomposition
TOA	Time-Of-Arrival
ULA	Uniform Linear Array

Notations

In this thesis, matrices and vectors are denoted on boldface. The upper-case letters are assigned to matrices, and lower-case is assigned to vectors. If there is no explicitly statement, the meaning of these notations are

$\mathbf{A}^T$	The transpose of $\mathbf{A}$.
$\mathbf{A}^H$	The Hermitian transpose of $\mathbf{A}$.
$\bar{\mathbf{A}}$	The complex conjugation without transposition.
$\mathbf{A}^{1/2}$	The Hermitian (*i.e.* $(\mathbf{A}^{1/2})^H = \mathbf{A}^{1/2}$) square root factor of a square matrix, *i.e.* $\mathbf{A} = \mathbf{A}^{1/2}\mathbf{A}^{1/2}$.
$\mathbf{A}^{-1}$	The inverse matrix of $\mathbf{A}$.
$\lvert\mathbf{A}\rvert$	The matrix determinant.
$[\mathbf{A}]_{ij}$ or a_{ij}	The $(i,j)^{\text{th}}$ element of the matrix $\mathbf{A}$.
$\lVert\mathbf{A}\rVert_F$	The Frobenius norm of $\mathbf{A}$.
$\lVert\mathbf{a}\rVert$	The Euclidian norm of vector $\mathbf{a}$.
$\mathbf{I}_N$	The $N \times N$ identity matrix.
$\mathbf{X}(m:n, p:q)$	The submatrix of $\mathbf{X}$ contains the elements from the m^{th} to the n^{th} row and from the p^{th} to the q^{th} column.
$\mathbf{X}^\dagger$	The pseudo inverse of $\mathbf{X}$, *i.e.* $(\mathbf{X}^H\mathbf{X})^{-1}\mathbf{X}^H$, if $\mathbf{X}$ has full column rank.
$\lvert x\rvert$	The modulus of the (possibly complex) scalar x.
x^*	The complex conjugate of x.
$\hat{x}$	An estimate of x.
$\mathbf{E}[\mathbf{x}]$	The expectation of $\mathbf{x}$.
$\mathbf{E}[\mathbf{x}\lvert\mathbf{y}]$	The conditioned expectation of $\mathbf{x}$ given $\mathbf{y}$.
j	The imaginary unit; $j^2 = -1$.
$\delta_{k,l}$	The Kronecker delta, *i.e.* $\delta_{k,l} = 1$, if $k = l$ and $\delta_{k,l} = 0$ otherwise.
$\delta(t)$	The Dirac delta function.

$\arg\min_{\boldsymbol{\theta}} V(\boldsymbol{\theta})$	The minimizing argument of the function $V(\boldsymbol{\theta})$.
$\mathrm{Tr}[\mathbf{A}]$	The Trace operation of a square matrix $\mathbf{A}$.
$\mathbf{A} \odot \mathbf{B}$	The Hadamard product of $\mathbf{A}$ and $\mathbf{B}$.
$\mathbf{A} \otimes \mathbf{B}$	The Kronecker product of $\mathbf{A}$ and $\mathbf{B}$.
$diag(\mathbf{A})$	The column vector formed from the diagonal elements of a square matrix $\mathbf{A}$.
$vec(\mathbf{A})$	The column vector obtained by stacking the columns of $\mathbf{A}$.
$\mathrm{Re}[\cdot]$	The real part of a variable, vector or matrix.
$\mathrm{Im}[\cdot]$	The imaginary part of a variable, vector or matrix.
$\mathcal{F}\{\cdot\}$	The DFT operation.
$\lceil \cdot \rceil$	Rounding towards plus infinity.
$x(t) \star y(t)$	The convolution of $x(t)$ and $y(t)$.

Chapter 1

Introduction

The history of radio communications can be traced back to the beginning of the 20th Century, when Guglielmo Marconi invented the wireless telegraphy and sent his famous "S" (dit dit dit) in Morse code from England to Canada in 1901. By the 1940's, the applications of radio communications had expanded into radio broadcasting, television, and aircraft navigation etc. The so-called wireless communications/wireless telephone for civil service was first introduced in 1946 in the U.S. by Bell Systems. However, the capacity of the early systems was very limited, and no more than a hundred users could be in service simultaneously within an area of thousands of square kilometers. The early mobile stations were cumbersome and had to be installed in the trunks of cars. The real advent of the massive wireless communications was in the 1980's, when the frequency reuse techniques[1] were invented together with the development of semiconductor industry. These systems are usually called the *first* generation of mobile communications. It took another 10 years for the wireless industry to develop the *second* generation of mobile communications. The dominant second generation mobile communication systems are GSM (Global System for Mobile communications) from Europe and IS-95 from the U.S.. With the development of digital technologies, the mobile terminal has become even cheaper, smaller and more power efficient than analogue ones. It has been a big success for the 2G systems that now more than two billion of users are using different mobile services every day and everywhere in the world.

With the aim of providing multimedia services, such as mobile TV, mobile video telephone, and mobile gaming etc., the *third* generation (3G) mobile communications were launched early in the new millennium, around 100

[1]The whole coverage area is split into small cells and the same radio frequencies are reused in non-neighboring cells.

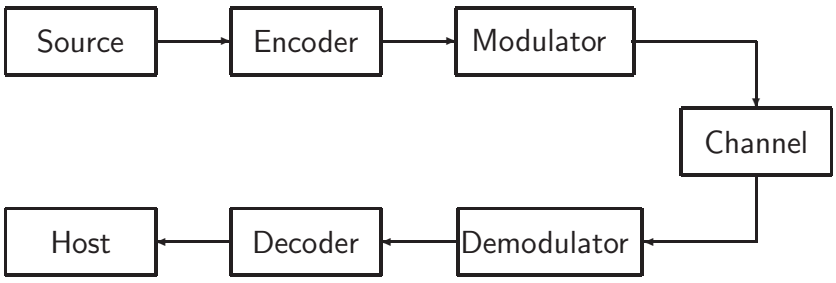

Figure 1.1: A block diagram of a digital communication system.

years after Marconi's famous *transatlantic transmission*. Meanwhile, the research and development of the *forth* generation (4G) mobile communication systems began being carried out across the world. The major features of the new systems are more frequency efficient modulation techniques, such as Orthogonal Frequency-Division Multiplexing (OFDM), and the Multiple-In-Multiple-Out (MIMO) technology.

As a part of the 4G study, the radio channel prediction, which helps to collect channel status information to improve system capacity and frequency efficiency, is investigated based on parametric modeling in this thesis.

1.1 Digital Communications

In digital communications, an analogue source signal is first quantized and then mapped into a binary sequence, i.e. a stream containing ones and zeros. Each of them is called a *bit*. These binary bits are then sent through a system and converted back to their original form at the receiver.

A block diagram of a digital communication system is given in Figure 1.1. The whole system includes a transmitter, a receiver and a channel. The transmitter consists of three basic blocks named *source*, *encoder*, and *modulator*. The three counterparts *host*, *decoder*, and *demodulator* form the receiver. The *channel* provides a physical link between a transmitter and a receiver.

In such a system, the *source* collects information-bearing signals and converts them into a stream of bits, where the raw signal might have been *compressed* by *source coding* in order to reduce the consumption of transmission resources. It is also possible to have digital sources of data, such as

file transfer or streaming digital video. The *encoder* is an entity that helps to improve the reliability of data transmission by adding extra control bits. This processing is known as *channel coding*. The resulting bit stream is then fed into a *modulator*, which maps the block-wised bit stream into a stream of *symbols*, which are selected from a symbol constellation. In a receiver, the *demodulator* first performs symbol detection and converts the noise corrupted symbols back to a binary sequence. The errors made in symbol detection give rise to bit errors. A part of the bit errors might be corrected by the *decoder* using the redundancy bits added by the *encoder*, while the others remain in the decoded bit stream. Finally, the received bit stream is forwarded to a *host*, and its original form, such as voice, text, or image, is recovered.

In practice, the physical channel can be set up over various media, such as a wired line, coaxial cable, electromagnetic field, or optical fibre, and its time-frequency property is described by the *channel impulse response* $h(t)$, and

$$x_R(t) \quad = \quad h(t) \star x_T(t) + n(t), \tag{1.1}$$

where t is discrete time, $x_T(t)$ and $x_R(t)$ are the transmitted and received signal, $n(t)$ is additive noise, and $\star$ represents convolution. A mathematically *ideal* channel allows the signal to pass through without distortion besides inevitable delay and power attenuation. Neglecting the delay, its impulse response can be modeled as

$$h(t) = g \cdot \delta(t) = \begin{cases} g & \text{t=0} \\ 0 & \text{otherwise} \end{cases}$$

where g is a constant channel gain and $\delta(t)$ is a Dirac Delta function. Unfortunately a practical channel can be much different from the ideal case due to delay spread, and random channel gain, which makes data transmission over radio difficult.

1.2 Mobile Communications

In mobile communications, the information carrying signal is transmitted over the air interface, which is formed by a multipath propagation environment as illustrated in Figure 1.2. In this figure, p dominant paths are present. The received radio signals which come via different paths suffer independent or correlated attenuations, phase shifts, frequency drifts and delays. Due to the relative movement between the Mobile Station (MS), Base Station (BS) and reflection clusters during the communication, the superimposed radio

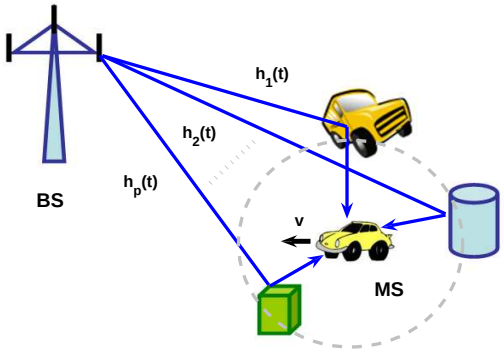

Figure 1.2: Multi-path propagation environment.

waves are added constructively or destructively at the locations travelled by the mobile. This gives rise to the *fast* fading of a radio channel as seen in Figure 1.3, where a measured radio channel is presented and the channel experiences a deep minimum every few portions of the wavelength. In this measurement, the wavelength is $\lambda = 0.15$ m, and the mobile velocity is around 13 m/s. These deep fades make the data transmission over the radio channel challenging. When the number of paths, p, is large, the channel can be modeled as a complex Gaussian random process with a Rayleigh distributed magnitude. Such a channel is called a *Rayleigh* fading channel.

1.3 Adaptive Transmissions

In the history of digital wireless communications, a number of techniques were introduced to combat the deep fades, such as power control and receiver diversity. The design at the link level was aimed at achieving a certain desired Bit Error Rate (BER) for a given data transition rate. This implied that modulation and coding selections were made based on the worst channel status in terms of a low SNR. These selections did not utilize the advantage of instantaneous high SNR values during data transmissions. Later, to achieve the Shannon capacity given by the instantaneous SNR of a channel [Sha48], adaptive transmissions were proposed [GC98, QC99, KH00, CG01, CEGJ02], where the transmitter adapts the power, data rate, and coding scheme based on the knowledge of the fade level. The adaption scheme in space, time, and

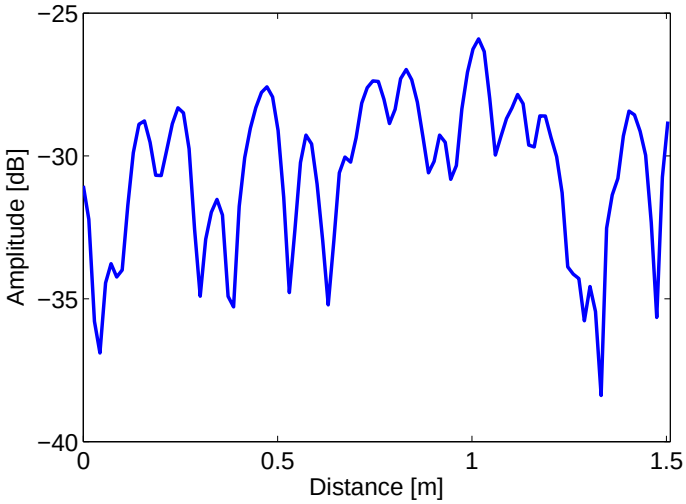

Figure 1.3: A measured Rayleigh fading channel, where the speed of the mobile is 13 m/s and the wavelength is 0.15 m.

frequency on a link level, is called *link adaptation*.

In the *time* domain, adaptive transmissions select different *modes*, which contain fixed combinations of modulations and coding rates, to send data depending on the Channel Status Information (CSI). This results in a time-vary data transmission rate. For example, when the SNR is higher than a given threshold, a high density symbol constellation is selected, where a symbol represents more binary bits. Otherwise a symbol constellation with lower density is used. See Figure 1.4 for an illustration. Together with the modulation selection, different channel coding schemes can be selected as well.

In the *frequency* domain, the fades at different frequency bins in a wide band channel are selective, which can be seen in the frequency response of a measured wide band channel shown in Figure 1.5. The adaptive transmissions explore the frequency selectivity of the channel, such as Orthogonal Frequency Division Multiplexing (OFDM) techniques, where the data is transmitted over a set of subcarriers with high instantaneous SNR instead of the whole bandwidth.

Adaptive transmission in the *spatial* domain is obtained by employing multiple antennas at both the transmitter and receiver. This is called a Multiple-In-Multiple-Out (MIMO) system. Depending on the spatial statistical property of the channel, one can choose to improve the transmission reliability by utilizing the channel redundancy (existing in highly correlated subchannels) [Ala98], or by maximizing the data throughput when the sub-

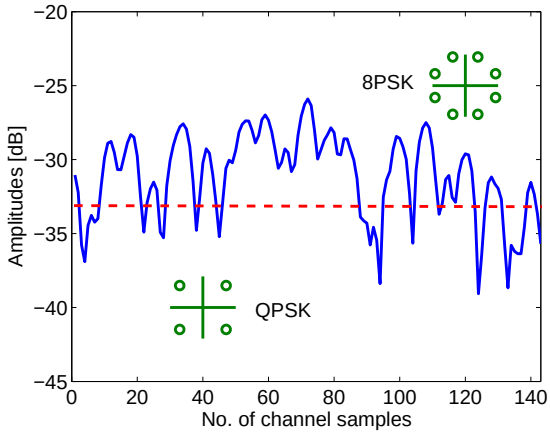

Figure 1.4: Demonstration example of adaptive modulation. When the channel is higher than the threshold (dash line), 8-PSK is selected. Otherwise, QPSK is used.

channels are independent [TSC98].

Besides *link adaptation*, adaption technique on system level have also been developed, i.e. multiuser diversity or smart scheduling in [Tse01]. This method takes the advantage of the independence of the fading statistics between different users. As seen in Figure 1.6, the channels (the dash curve and the dash-dot curve) are associated with two different users. They have independent channel fades. The BS might allocate the radio resource to the one with the highest instantaneous SNR (solid curve) to improve the throughput on the system level as shown in the figure.

1.4 Channel Estimation and Channel Prediction

In the aforementioned adaptive transmissions, good knowledge of the channel in the future is assumed to be known at the transmitter. In practice, the current channel status information is collected at the receiver, which is termed channel estimation. The CSI is then fed back to the transmitter via an *idle* logical uplink channel. If all logical channels are in busy mode, the mobile station might need to wait until one was allocated. The possible feedback delay gives rise to an outdated CSI. The remedy is to use channel prediction.

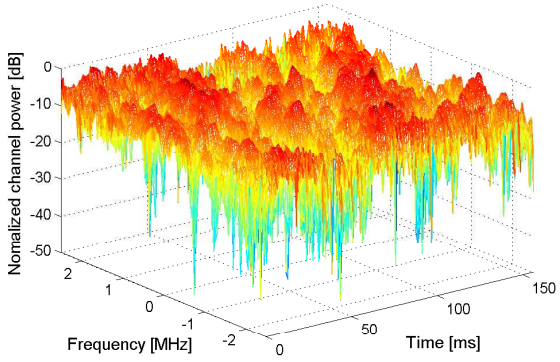

Figure 1.5: Frequency response vs. time for a wide band channel.

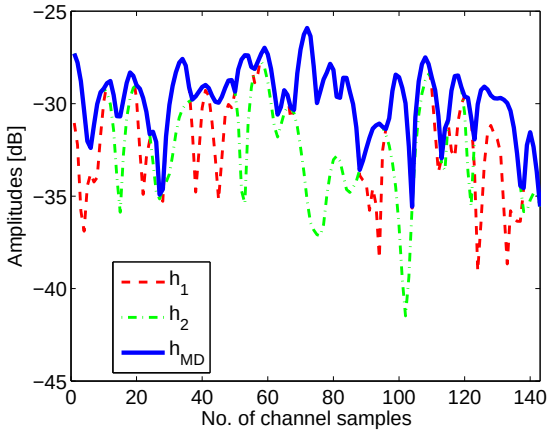

Figure 1.6: Multiuser diversity.

1.4.1 Channel Estimation

Due to the fast fading property of a radio link, knowledge of a channel is necessary for coherent symbol detection in the *demodulator*. Channel estimation can be made by transmitting *pilot* symbols, which are known at the receiver. The channel status is assumed to be constant before the next channel estimation made [TSD04].

Increasing the number of pilot symbols helps to improve the estimation accuracy, but it reduces the throughput of the information (payload) symbols. One of the major tasks of channel estimation is to optimize the quantity ratio, power ratio, and placement between the pilot symbols and the payload symbols in a data frame. As an example, the pilot symbol allocation in an

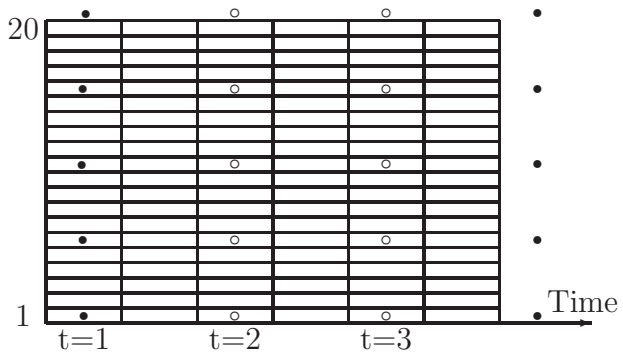

Figure 1.7: The pilot and control symbol allocation in a time-frequency bin of size 0.667 ms and 200 KHz, proposed by Wireless IP project. One time-frequency bin contains 20 subcarriers with 6 symbols each. Known 4-QAM pilot symbols (black) and 4-QAM control symbols (rings) are placed on four pilot subcarriers.

OFDM time-frequency bin proposed by the Wireless IP[1] project is given in Figure 1.7. In such a system, each time-frequency bin of size 0.667 ms and 200 KHz contains 120 symbols. Four pilot symbols and eight control symbols are placed as shown in the plot. The non-pilot subcarriers can be estimated by interpolation.
Usually an observed channel (channel estimation) is written as

$$y(t) = h(t) + e(t), \tag{1.2}$$

where $y(t)$ is the observed channel, $h(t)$ is the true channel, $e(t)$ is the estimation error with zero-mean and variance σ_e^2. The $h(t)$ can correspond to a narrow band Rayleigh fading channel or an FIR tap of a frequency selective channel.

1.4.2 Channel Prediction

In a real system, the mobile stations in service need to listen to all physical channels in the down link, and report the measured CSI's to the BS for radio resource allocation and adaptive transmissions. In a multiple access system, it is non trivial to feed back this information from different users to the BS simultaneously due to the limited bandwidth in the uplink. A well synchronized and scheduled uplink is necessary. In such a scenario, the

[1]Wireless IP is a 4G-oriented research project founded by Swedish Foundation for Strategy Research (SSF).

feed-back delay results in an outdated CSI at the transmitter, which can lead to uncertainty in the adaptive transmission selection. The influence of the imperfect channel information was studied in [DHHH00, FSES04]. However, such a problem might be alleviated by channel prediction using previous channel observations.

In channel prediction, a finite number, N, of channel estimates is collected in $\mathbf{y}$, and

$$\mathbf{y} = \mathbf{h} + \mathbf{e}, \tag{1.3}$$

where

$$\begin{aligned} \mathbf{y} &= [y(t), y(t-1), \cdots, y(t-N+1)]^T, & (1.4)\\ \mathbf{h} &= [h(t), h(t-1), \cdots, h(t-N+1)]^T, & (1.5)\\ \mathbf{e} &= [e(t), e(t-1), \cdots, e(t-N+1)]^T. & (1.6) \end{aligned}$$

The value $h(t+L)$ is to be predicted, where L is the prediction horizon. Such a scenario is shown in Figure 1.8, where $N = 100$ and $t = 0$.

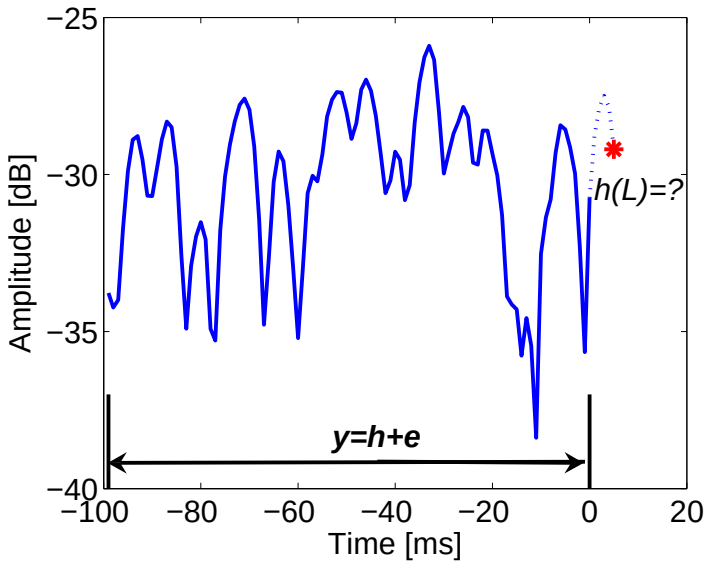

Figure 1.8: Channel prediction.

In previous studies, parametric and nonparametric channel prediction methods were developed in [AJHF99, HW98, DHHH00, Ekm02]. These reported predictors mainly fall into two categories: the classical Linear Prediction (LP) and sinusoidal modeling based prediction. Beside these prediction methods, a nonlinear prediction of radio channel using Multivariate Adaptive

Regression Splines (MARS) was studied in [EK99]. Recently, a new channel estimation and prediction method using Discrete Prolate Spheroidal (DPS) sequences was proposed in [ZM05]. But it was reported that the radio channel is only predictable a few portions of wavelength into the future [TV01]. A long range channel prediction (half a wavelength) is expected to be necessary to achieve full advantage of adaptive communication.

1.5 Contributions and Thesis Outline

The contributions of the author in the research field of channel modeling and long range channel prediction based on parametric modeling is summarized as follows:

Physics based scattering model for narrow band channels

By modeling the radio waves scattered on a rough surface, a physics base scattering channel model is proposed for narrow band SISO channels [CVF07b]. It is extended into MIMO channels following the double directional structure [SWS03]. In these models, there is no predefined model parameters, such as Doppler frequencies and amplitudes. These setups are critical for the performance evaluation of model based channel predictors.

LMMSE channel predictors based on sinusoidal modeling

A number of LMMSE channel predictors are proposed based on statistical sinusoidal modeling of Rayleigh fading channels, where both the amplitudes and frequencies are modeled as Gaussian random variables. These predictors are named conditional LMMSE predictor, adjusted conditional LMMSE predictor, and unconditional LMMSE predictor respectively. These methods outperform the deterministic sinusoidal model based channel predictor and LP using synthetic data, but underperform LP using measured channels [CV04, CEV05].

Joint Moving Average and Sinusoidal model and Joint LS channel predictor

By splitting the channel into the sinusoidal (periodic) part and non-sinusoidal

(non-periodic) part, a Joint Moving Average and Sinusoidal (JMAS) model for channel prediction is proposed, which leads to a Joint LS predictor. The prediction is then based on both autoregressive bases and selected sinusoidal bases. Together with a simple SVD based model selection method, the JLS predictor outperforms all the LMMSE predictors in performance evaluation using measured channels [CEV07], but still slightly underperform the LP.

Adaptive channel prediction based on Polynomial Phase Signals

Motivated by the analysis of real world channels and the physics based scattering models, adaptive channel prediction based on a Polynomial Phase Signal (PPS) model is proposed. To mitigate the influence of the unknown time-varying amplitudes, an iterative parameter estimation of the polynomial phases using the Nonlinear *Instantaneous* LS (NILS) criterion is proposed. The time-varying amplitude is modeled as an AR(d) process. The new predictor outperforms the LP in Monte Carlo simulations with known model orders and number of components [CVF07a, CV07].

Detection of the number of signal components and model order selection of polynomial phase signals

An iterative procedure to detect the number of signal components and the order of the polynomial phases is proposed in [CV07]. The detection of the signal component is based on an examination of the size of the residual signal. The model order selection of the polynomial phase is based on a Generalized Likelihood Ratio Test (GLRT) or Wald test. High order parameters of the polynomial phases are observed in both urban and suburban environments. The adaptive channel predictors, together with the detection and estimation method, outperform LP in simulations and examples of measured urban and suburban channels.

The content in this thesis are partially contained in the following publications:

- M. Chen, M. Viberg. "LMMSE channel prediction based on sinusoidal modeling." In *Proc. of 3rd IEEE Sensor Array and Multichannel Signal Processing Workshop*, pp. 377- 381, Barcelona, Spain, Jul. 2004. [CV04]
- M. Chen, T. Ekman, and M. Viberg, "Two new approaches for channel

prediction based on sinusoidal modeling." In *Proc. of IEEE Workshop on Statistical Signal Processing*, Bordeaux, France, Jul. 2005.[CEV05]

- M. Chen, T. Ekman, and M. Viberg, "New Approaches for Channel Prediction Based on Sinusoidal Modeling." In *EURASIP Journal on Advances in Signal Processing*, vol. 2007, Article ID 49393, 13 pages, 2007. doi:10.1155/2007/49393. [CEV07]

- M. Chen, M. Viberg, and S. Felter, "Models and Predictions of Scattered Radio Waves on Rough Surfaces." In *Proc. IEEE ICASSP*, vol. 3, pp. 785-788, Honolulu, Hawaii, USA, 2007. [CVF07b]

- M. Chen, M. Viberg, and S. Felter, "Adaptive Channel Prediction Based on Polynomial Phase Signals." Submitted to *IEEE ICASSP 2008*, Las Vegas, USA, Sept. 2007. [CVF07a]

- M. Chen, M. Viberg, "Long Range Channel Prediction Based on Non-Stationary Parametric Modeling." Submitted to *IEEE Trans. on Signal Processing*, Sept. 2007. [CV07]

Other publications of the author:

- M. Chen, S. Felter, "Feasibility Study of Channel Prediction Based on Sinusoidal Modeling with Time Variant Model Parameters." *Technical Report*, Ericsson Research, Stockholm, Sweden, Nov. 2005. [CF05]

- S. Felter, M. Chen and M. Viberg, "Method and Arrangement for Channel Prediction." *Patent application*, P22729US1, USA, Sept. 2006. [FCV06].

- M. Chen, "Channel Prediction Based on Sinusoidal Modeling." *Licentiate Thesis*, Chalmers University of Technology, August, 2005. [Che05]

- M. Chen, "Mobile Positioning in Distributed Antenna Systems." *Patent application*, Telefonaktiebolaget LM Ericsson, USA, Dec, 2002. [Che02]

- M. Chen, H. Koorapaty, and A. Kangas, "Enhanced Positioning Method in Cellular Systems." In *Proc. of IEEE International Conference of Telecommunications*, Beijing, China, June, 2002. [CKK02]

- M. Chen, H. Asplund, "Measurements and Models for Direction of Arrival of Radio Waves in LOS in Urban Microcells." In *Proc. of the 12th IEEE International Symposium of Personal, Indoor and Mobile Radio Communications*, vol. 1, pp. B100–104, San Diego, USA, 2001. [CA01]

A brief introduction to the content in each chapter is given below.

Chapter 2 *Channel Modeling*

This chapter is devoted to the modeling of narrow band Rayleigh fading channels. The sinusoidal modeling is introduced first. Then the physics based scattering model is addressed, followed by the non-stationary polynomial phase signal modeling. These models are also extended into MIMO scenarios.

Chapter 3 *Parameter Estimation*

Parameter estimation techniques involved in the predictor design are discussed in this chapter. These techniques include the classical Wiener filters and estimators such as LS estimator, MMSE estimator, ML estimator, NLLS and NILS estimator. A number of frequency estimation algorithms, such as MUSIC and ESPRIT are introduced as well. To reduce the computational complexity of the frequency estimate, the well known iterative method (SAGE/RELAX) is addressed.

Chapter 4 *Model Order Selection and Detection of MC-PPS*

The detection of the model order of a polynomial phase signal using the Wald Test is discussed. It is in principle equivalent to the GLRT based detection method. An iterative detection method based on examination of the size of residual signal is also introduced in this chapter.

Chapter 5 *Model-Based Channel Prediction*

This chapter summarizes the predictors proposed by the authors using the different parametric models.

Chapter 6 *Conclusions and Future Works*

This chapter contains the conclusions of the thesis. The comments and discussions on future works are given.

Chapter 2

Channel Modeling

The study of modeling and characterization of radio channels plays an important role in the research and development of wireless communications. It helps to understand the challenge to design a mobile communication system and it provides a tool for simulation. Serving for different purposes, numerous publications on this topic can be found in [ECS+98, YO02], and the references therein. In principle, all these models can be divided into *non-physical* models and *physical* models. The *non-physical* models characterize only the statistical property of the channels, such as their spatial and temporal correlations. These models provide a fast and convenient method to generate radio channels in simulations. However, in some studies, such as Beamforming (BF) and MIMO etc, not only the statistical property, but also detailed propagation parameters, such as Direction-Of-Arrival (DOA), Direction-Of-Departure (DOD), and Time-Of-Arrival (TOA) etc., need to be characterized. This information can be provided by a *physical* model, but a large number of parameters are needed to fully describe the wave propagation scenarios, such as the locations of BS, MS and reflectors, the speed of the MS, and the configurations and orientations of the antenna arrays. In this thesis, the *physical* models are of interest, and the wave propagation structure is explored for long range channel prediction.

In this chapter, a *physical* ellipsoidal channel model is presented first. Then, motivated by the Jakes model [Jak74], a statistical sinusoidal model is introduced [CEV07]. Later, a physics based scattering model for SISO channels is discussed [CVF07b], which leads to the Multi-Component Polynomial Phase Signal (MC-PPS) [CVF07a], which is a non-stationary signal. A discussion on the physics based scattering MIMO channel models is given at the end of this chapter.

2.1 Physical Ellipsoidal Models

Radio wave propagation over a local area can be described by an ellipsoidal model as in Fig. 2.1. In this model, the transmitter and receiver are placed separately on two foci of a series of ellipsoids. A number of reflection objects are distributed on these ellipsoids. The inter-distance (resolution) Δl between neighboring ellipsoids is limited by the bandwidth of the channel, B,

$$\Delta l \leq \frac{c}{4B}, \tag{2.1}$$

where c is the speed of light. If one assumes that all paths experience no more than one reflection, then all paths, which pass via reflection objects that lie on the same ellipsoid, will share the same path length and delay. The path loss increases exponentially with the increase of the path length. The signals via reflection objects on outer ellipsoids are weak and could be neglected.

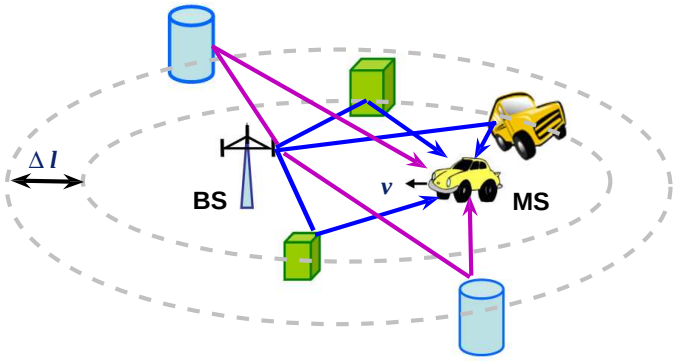

Figure 2.1: Physical ellipsoidal models.

Mathematically, the channel impulse response of the ellipsoidal model can be formed as a Finite Impulse Response (FIR) filter,

$$h(t,\tau) = \sum_{k=0}^{K-1} h(t,\tau_k)\delta(\tau-\tau_k), \tag{2.2}$$

where τ is the excessive delay, K is the number of taps, $h(t,\tau_k)$ contains all paths sharing the same delay of τ_k. When $K=1$, the channel has only one

tap and can be written as

$$h(t,\tau) = h(t,\tau_0) = h(t)\delta(\tau - \tau_0), \tag{2.3}$$

which is different from the *ideal* channel by a time-varying scaling factor. Since its frequency response is constant over the whole bandwidth, it is called a *flat* fading channel, or a *narrow* band channel. For simplicity, the narrow band channel can be denoted as $h(t)$, where the delay parameter τ is dropped.

When $K > 1$, the channel introduces inter-symbol interference and the frequency response becomes selective. Such a channel is termed a frequency *selective* fading channel, or a *wide* band channel. An example of a measured wide band channel is given in Fig. 2.2, where the channel impulse response contains 120 taps along the axis of excessive delay, and 143 impulse responses are plotted along the time axis.

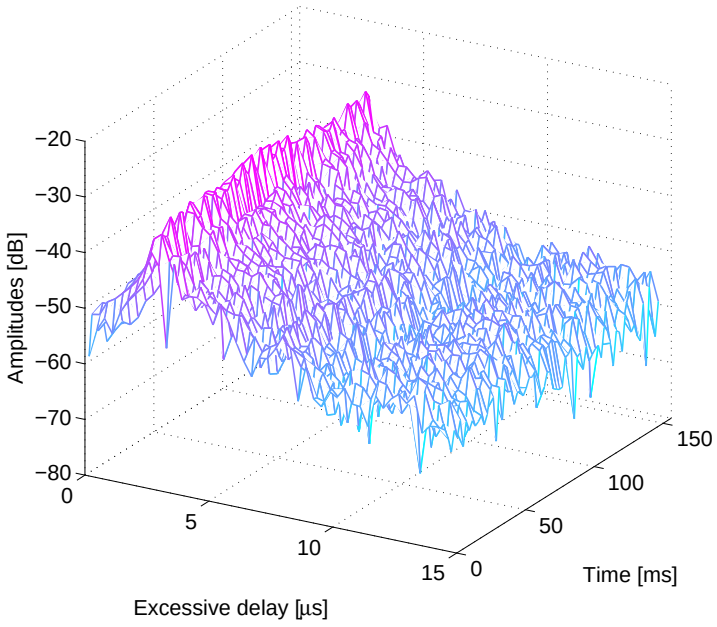

Figure 2.2: Example of a measured wide band channel. Each channel impulse response contains 120 taps along the axis of excessive delay, and 143 impulse responses are plotted along the axis of time.

2.2 The Jakes Model

Jakes model is one of the most widely used models for a flat Rayleigh fading channel [Jak74], in which the second order property of a channel is approxi-

mated by the *tub* shape power spectrum as in (2.4) [Rap96],

$$p(f) = \frac{1}{\pi f_m \sqrt{1 - (\frac{f - f_c}{f_m})^2}}, \tag{2.4}$$

where f_m is the maximum Doppler frequency and f_c is the carrier frequency. Both are in Hz. To complete this definition, define $p(f) = 0$, when $|f - f_c| > f_m$. Such a spectrum are based on two assumptions:

- Rich scattering environment, where a large number of scatters are uniform distributed in the vicinity of the mobile terminal (in the far field);
- The scattered signals from different objects have equal power.

These assumptions are true in statistics, and this model is suitable for generating channels when only properties over time are required. The Auto-Correlation Function (ACF) of (2.4) is a zero order Bessel function of the first kind, i.e.,

$$r_h(\tau) = \mathcal{J}_o(\omega_m \tau), \tag{2.5}$$

where $r_h(\tau) = \mathbf{E}[h(t)h^*(t - \tau)]$ and $\omega_m = 2\pi f_m$ is the maximum Doppler frequency in radian. Since $\frac{\omega_d \tau}{2\pi}$ is a distance measured in wavelength, one can plot (2.5) as a function of distance as in Figure 2.3, where the initial zero-crossing appears at around 0.5. For this reason, a prediction of such a channel over a half wavelength is considered difficult [SEA01].

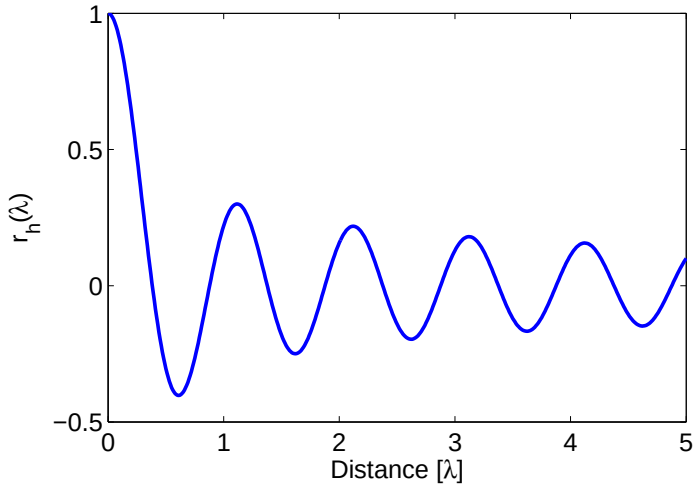

Figure 2.3: Auto-Correlation Function (ACF) of Rayleigh fading channels.

2.3 Sinusoidal Modeling

2.3.1 SISO Channel Based on Sinusoidal Modeling

In a local area, the number of dominant paths are limited [CA01]. This motivates a deterministic sinusoidal model for a Rayleigh fading channel [AJHF99, HW98]. Assume p paths contained in a narrow band channel $h(t)$, and

$$h(t) = \sum_{i=1}^{p} h_i(t), \tag{2.6}$$

where $h_i(t)$ corresponds to the i^{th} path. When the mobile is moving at the speed of v m/s, we have

$$h_i(t) = s_i e^{j\omega_i t}, \tag{2.7}$$

where s_i is the complex amplitude, and ω_i is the doppler frequency, which is

$$\omega_i = \frac{2\pi v}{\lambda} \cos\theta_i, \tag{2.8}$$

where θ_i is the DOA, the angle between v and the impinging path, and λ is the wavelength.

Modeling θ_i as a random variable with uniform distribution, $\mathcal{U}[-\pi, \pi)$, the PDF of the normalized Doppler frequency ω_i is

$$p(\omega_i) = \frac{1}{\pi\sqrt{1-\omega_i^2}}, \quad -1 < \omega_i < 1, \tag{2.9}$$

which is identical (except for a scaling factor) to the expression of the power spectrum given by the Jakes model [Jak74]. A plot of $p(\omega_i)$ is shown in Figure 2.4. In vector form, the model is, given $\boldsymbol{\omega} = [\omega_1, \cdots, \omega_p]^T$,

$$\mathbf{y} = \mathbf{A}(\boldsymbol{\omega})\mathbf{s} + \mathbf{e}, \tag{2.10}$$

where

$$\begin{aligned} \mathbf{A}(\boldsymbol{\omega}) &= [\mathbf{a}(\omega_1), \cdots, \mathbf{a}(\omega_p)], & (2.11)\\ \mathbf{a}(\omega_i) &= [e^{j\omega_i(t)}, \cdots, e^{j\omega_i(t-N+1)}]^T, & (2.12)\\ \mathbf{s} &= [s_1, \cdots, s_p]^T. & (2.13) \end{aligned}$$

2.3.2 MIMO Channel Based on Sinusoidal Modeling

Such a ray tracing based model is extended into the MIMO scenario in [SWS03, Che05]. In [SWS03], a MIMO channel with n_T and n_R transmit

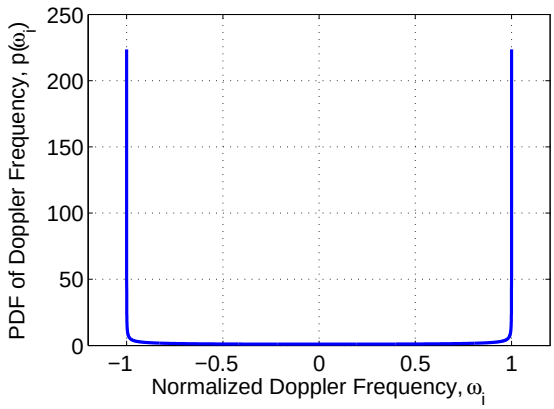

Figure 2.4: PDF of the normalized Doppler frequency.

and receive antennas is modeled as

$$\mathbf{H}(t) = \sum_{i=1}^{p} s_i e^{j\omega_i t} \mathbf{a}(\theta_{R,i}) \mathbf{a}^T(\theta_{T,i}), \tag{2.14}$$

where $\theta_{R,i}$ and $\theta_{T,i}$ are the DOA and DOD associated with the i^{th} path, and $\mathbf{a}(\theta_{R,i})$ and $\mathbf{a}(\theta_{T,i})$ are the array response vector to the i^{th} path at the transmitter and receiver respectively. In the case of a Uniform Linear Array (ULA),

$$\mathbf{a}(\theta_{T,i}) = [1, e^{-j\Omega_{T,i}}, \cdots, e^{-j(n_T-1)\Omega_{T,i}}]^T, \tag{2.15}$$

$$\mathbf{a}(\theta_{R,i}) = [1, e^{-j\Omega_{R,i}}, \cdots, e^{-j(n_R-1)\Omega_{R,i}}]^T, \tag{2.16}$$

where the spatial frequencies at the transmitter and receiver are

$$\Omega_{T,i} = 2\pi\Delta_T \sin(\theta_{T,i})/\lambda, \tag{2.17}$$

$$\Omega_{R,i} = 2\pi\Delta_R \sin(\theta_{R,i})/\lambda, \tag{2.18}$$

where Δ_T and Δ_R are the element separations of the transmit and receive array elements. Note that it is assumed in (2.14) that the observed Doppler frequencies and amplitudes at different array elements associated with the same path are identical.

In [Che05], a MIMO channel based on statistical sinusoidal modeling is proposed as

$$\mathbf{H}(t) = \sum_{i=1}^{p} e^{j\omega_i t} \mathbf{S}_i, \tag{2.19}$$

where the $n_R \times n_T$ matrix $\mathbf{S}_i$ contains the random amplitudes associated with the i^{th} path, and vec($\mathbf{S}_i$) has PDF, $\mathcal{CN}(\mathbf{0}_{n_R n_T}, \sigma_s^2 \mathbf{I}_{n_R n_T})$. The amplitudes associated with the same path are independent for different subchannels.

2.4 Physics Based Scattering Models

A number of physical MIMO channel models were published during the last several years [GBGP02, AK02, SFGK00, Sva01, WJ01, FMB98, Cor01]. The common feature of these models is to approximate the spatial and temporal correlation of a MIMO channel by modeling the relative locations of the BS, MS and a number of distributed reflection clusters. In [AK02, Sva01, Cor01, FMB98], the time evolution of a MIMO channel is modeled by taking the mobile velocity into account. However, the reflection of wave propagation in these models are assumed to be on specular surfaces, i.e. the amplitude to each path is constant. In [WJ01, FMB98], the scattering effect is modeled by the angular spread, which corresponds to distributed sources. In this section, a physics based scattering SISO channel model proposed in [CVF07b] is addressed. A similar model can be found in [3GP03]. It is also extended into MIMO channel models with single and double reflections.

2.4.1 Physics Based Scattering SISO Models

In the sinusoidal modeling of a Rayleigh fading channel (2.7), the reflection is assumed to be on a specular surface, and the DOA is constant. But these assumptions might not be true in practice. The reflection of radio wave can be on a rough surface and the DOA will be time-varying when the mobile is close to the reflection object. In such a scenario, the reflected wave becomes scattered from a large number of scatters on the surface, which is termed a *cluster* as in Figure 2.5, where $(x_{i,c}, y_{i,c})$ is the *center of gravity* of the i^{th} cluster, and

$$(x_{i,c}, y_{i,c}) = \sum_{j=1}^{q_i} (x_{i,j}, y_{i,j}) / q_i, \tag{2.20}$$

where $(x_{i,j}, y_{i,j})$ is the coordinate of the j^{th} scatter in the i^{th} scattering surface, and q_i is the number of scatters in the i^{th} clusters. The radio channel scattered on the i^{th} rough surface can be approximated as

$$h_i(t) = s_i(t) e^{j2\pi l_{i,c}(t)/\lambda}, \tag{2.21}$$

$$s_i(t) = \frac{s_i}{\sqrt{q_i}} \sum_{j=1}^{q_i} e^{j2\pi \Delta l_{i,j}(t)/\lambda}, \tag{2.22}$$

$$\Delta l_{i,j}(t) = l_{i,j}(t) - l_{i,c}(t), \tag{2.23}$$

where $l_{i,c}(t)$ and $l_{i,j}(t)$ are the lengths of the propagation path from the transmitter antenna to the receiver antenna via $(x_{i,c}, y_{i,c})$ and the scatter at

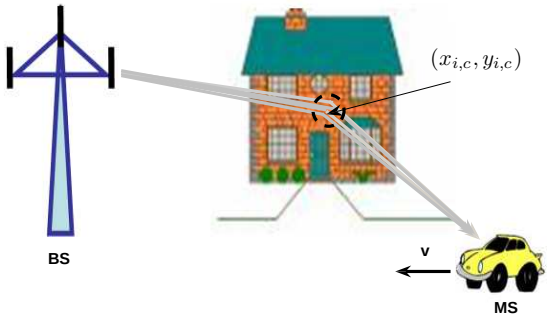

Figure 2.5: Scattered Radio Waves on Rough Surface.

$(x_{i,j}, y_{i,j})$ respectively, s_i is the impinging amplitude, which is assumed to be identical over the scattering surface and time.

Assume a circular reflection area with a radius of $\gamma\lambda$ as in Figure 2.6, the degree of scattering or roughness of the surface is determined by γ. According to the Rayleigh criterion [Sau99], a surface is considered as smooth, if γ is less than 0.25, which results in the maximum path length difference of a half wavelength. When $\gamma = 0$, this scattering model degenerates to the specular reflection model. In general, the larger γ is, the larger/rougher the surface is in this model. Note that it was found that the time-varying amplitude $s_i(t)$ in (2.22) has a nonzero mean in general [CVF07b].

2.4.2 Physics Based Scattering MIMO Models

Let a MIMO channel with n_T transmit antennas at BS and n_R receive antennas at MS, and p clusters be located in the vicinity of the mobile. Following the idea in [SWS03], a Double-Direction-Single-Reflection (DDSR) MIMO channel using the physics based scattering scheme is

$$\mathbf{H}_{DDSR}(t) = \sum_{i=1}^{p} s_i(t) e^{j2\pi l_{i,c}(t)/\lambda} \mathbf{a}(\theta_{R,i}(t)) \mathbf{a}(\theta_{T,i})^H, \tag{2.24}$$

where $\theta_{R,i}(t)$ and $\theta_{T,i}$ are the *time-varying* DOA and *constant* DOD associated with the i^{th} cluster as in Figure 2.7, and $\mathbf{a}(\theta_{R,i}(t))$, $\mathbf{a}(\theta_{T,i})$ are the associated steering vector of the receive and transmit array. For the Uniform Linear Array (ULA), $\mathbf{a}(\theta_{T,i})$ is given in (2.15), and

$$\mathbf{a}(\theta_{R,i}(t)) \quad = \quad [1, e^{-j\Omega_{R,i}(t)}, \cdots, e^{-j(n_R-1)\Omega_{R,i}(t)}]^T, \tag{2.25}$$

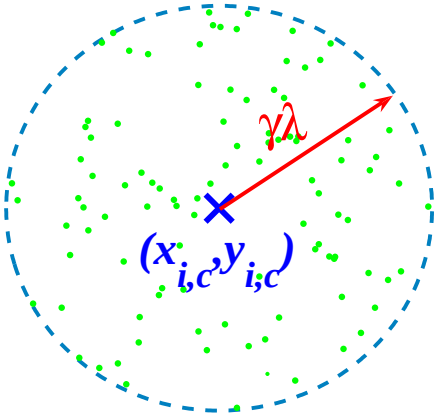

Figure 2.6: A circular scattering cluster with a radius of $\gamma\lambda$, and $(x_{i,c}, y_{i,c})$ is the center of gravity.

where

$$\Omega_{R,i}(t) \quad = \quad 2\pi\Delta_R \sin\left(\theta_{R,i}(t)\right)/\lambda. \tag{2.26}$$

Note that the time-varying DOA, $\theta_{R,i}(t)$, is due to the short distance between the MS and the reflection clusters. When the distance from the BS to the scattering clusters is large, the DOD is constant over a short observation interval.

2.5 Polynomial Phase Signal Modeling

The time-varying phase associated with the i^{th} cluster in (2.21) is approximated by a polynomial of time t with order M_i, i.e.

$$\phi_i(t) = 2\pi l_{i,c}(t) \approx \sum_{m=1}^{M_i} \beta_{i,m} t^m. \tag{2.27}$$

Then, the channel is

$$h_i(t) = s_i(t) e^{j\phi_i(t)}. \tag{2.28}$$

When $M_i = 1$, (2.28) becomes a sinusoidal model with time-varying amplitudes [CVF07b]. When $M_i = 2$, (2.28) is a quadratic phase signal, which has found an application in Synthetic Aperture Radar (SAR) [CM91].

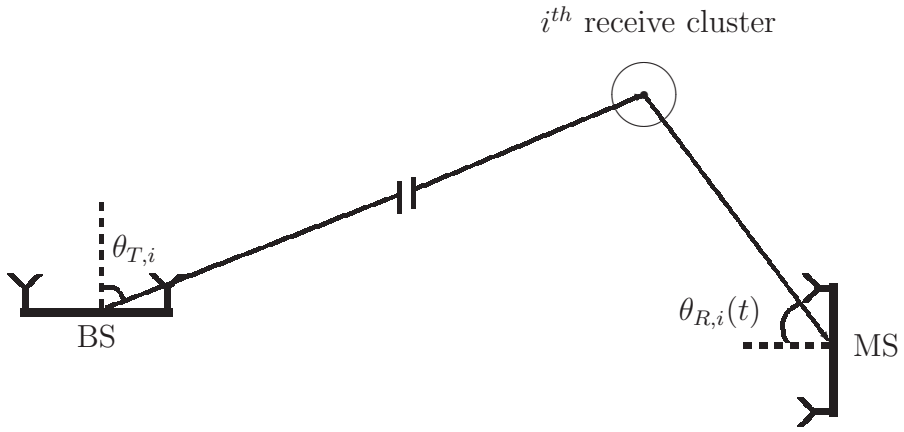

Figure 2.7: Physics Based Scattering DDSR MIMO Model.

In vector form, the MC-PPS signal

$$\mathbf{y} = \sum_{i=1}^{p} \mathbf{a}_i(\boldsymbol{\theta}_i) \odot \mathbf{s}_i + \mathbf{e}, \tag{2.29}$$

where $\odot$ is the Hadamard product (elementwise multiplication), and

$$\begin{aligned} \mathbf{a}_i(\boldsymbol{\theta}_i) &= [e^{j\phi_i(t)}, \cdots, e^{j\phi_i(t-N+1)}]^T, \\ \mathbf{s}_i &= [s_i(t), \cdots, s_i(t-N+1)]^T. \end{aligned}$$

The model parameters are collected in the vectors $\boldsymbol{\theta} = [\boldsymbol{\theta}_1^T, \cdots, \boldsymbol{\theta}_p^T]^T$, where $\boldsymbol{\theta}_i = [\beta_{i,1}, \cdots, \beta_{i,M_i}]^T$. The time varying amplitudes are considered as unknown deterministic/stochasitc signals, but not as model parameters.

Chapter 3

Parameter Estimation

Parameter estimation is one of the typical problems in signal processing, where the model parameters collected in a vector $\boldsymbol{\theta} = [\theta_1, \cdots, \theta_p]^T$ need to be estimated from a finite number of observations,

$$\mathbf{y} = \mathbf{h}(\boldsymbol{\theta}) + \mathbf{e}, \tag{3.1}$$

where $\mathbf{h}$ in (1.3) is written as $\mathbf{h}(\boldsymbol{\theta})$ to emphasize the parametric dependence. The number of unknown parameters is assumed to be less than the signal observations, i.e., $p < N$. The estimate of $\boldsymbol{\theta}$ can usually be formed as an optimization problem as

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} V(\boldsymbol{\theta}), \tag{3.2}$$

where $V(\boldsymbol{\theta})$ is the *cost* function. Various estimators can be obtained by applying different optimal or suboptimal criterion in (5.48). An extensive discussion on this topic can be found in [Kay93b]. In this chapter, estimation techniques related to the channel predictor design are addressed briefly.

After a brief introduction of the Cramér-Rao Lower Bound (CRLB), the Maximum Likelihood estimator (MLE) is introduced, which is an optimal estimator when the statistical property of the signal is known. Then the estimators for linear models, such as LS and MMSE, are presented. After that, the celebrated adaptive parameter estimation method using a Kalman filter is discussed, which is suitable for non-stationary signal processing. Later, the Non-linear LS (NLLS) and Non-linear *Instantaneous* LS method is introduced for parameter estimation of MC-PPS signals with time-varying amplitudes. At the end, the subspace based frequency estimation methods, such as MUSIC and ESPRIT, are presented.

3.1 Cramér-Rao Lower Bound

Given $\mathbf{y}$, the model parameters in $\boldsymbol{\theta}$ need to be estimated. Due to the random noise, the estimation of $\boldsymbol{\theta}$ becomes random variables as well. The mean, $\mathbf{E}[\hat{\boldsymbol{\theta}}]$, and the variance, $\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \mathbf{E}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T]$, of $\hat{\boldsymbol{\theta}}$ are usually used to evaluate the performance of different estimators. The notation $\mathbf{E}[\cdot]$ is expectation. When $\mathbf{E}[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta}$, an estimator is *unbiased*, otherwise it is called a *biased* estimator. If a *regularity* condition is satisfied, for example

$$\mathbf{E}\left[\frac{\partial \ln p(\mathbf{y}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right] = 0, \text{for all } \boldsymbol{\theta}, \tag{3.3}$$

where $p(\mathbf{y}, \boldsymbol{\theta})$ is the PDF of $\mathbf{y}$, the variance of any unbiased estimator is bounded by the Cramér-Rao Lower Bound (CRLB), which is

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} \geq \mathbf{F}^{-1}(\boldsymbol{\theta}), \tag{3.4}$$

where $\mathbf{F}(\boldsymbol{\theta})$ is the Fisher information matrix, defined by

$$[\mathbf{F}(\boldsymbol{\theta})]_{ij} = -\mathbf{E}\left[\frac{\partial^2 \ln p(\mathbf{y}, \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}\right]. \tag{3.5}$$

For the general Gaussian case, where

$$\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{\theta}), \mathbf{C}(\boldsymbol{\theta})), \tag{3.6}$$

the Fisher information matrix is given by, i.e. [Kay93b],

$$\begin{aligned}[\mathbf{F}(\boldsymbol{\theta})]_{ij} &= \left[\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_i}\right] \mathbf{C}^{-1}(\boldsymbol{\theta}) \left[\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_j}\right] \\ &+ \frac{1}{2}\mathrm{Tr}\left[\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_i}\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_j}\right],\end{aligned} \tag{3.7}$$

where

$$\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_i} = \begin{bmatrix} \frac{\partial [\boldsymbol{\mu}(\boldsymbol{\theta})]_1}{\partial \theta_i} \\ \frac{\partial [\boldsymbol{\mu}(\boldsymbol{\theta})]_2}{\partial \theta_i} \\ \vdots \\ \frac{\partial [\boldsymbol{\mu}(\boldsymbol{\theta})]_N}{\partial \theta_i} \end{bmatrix} \tag{3.8}$$

and

$$\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_i} = \begin{bmatrix} \frac{\partial [\mathbf{C}(\boldsymbol{\theta})]_{11}}{\partial \theta_i} & \frac{\partial [\mathbf{C}(\boldsymbol{\theta})]_{12}}{\partial \theta_i} & \cdots & \frac{\partial [\mathbf{C}(\boldsymbol{\theta})]_{1N}}{\partial \theta_i} \\ \frac{\partial [\mathbf{C}(\boldsymbol{\theta})]_{21}}{\partial \theta_i} & \frac{\partial [\mathbf{C}(\boldsymbol{\theta})]_{22}}{\partial \theta_i} & \cdots & \frac{\partial [\mathbf{C}(\boldsymbol{\theta})]_{2N}}{\partial \theta_i} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial [\mathbf{C}(\boldsymbol{\theta})]_{N1}}{\partial \theta_i} & \frac{\partial [\mathbf{C}(\boldsymbol{\theta})]_{N2}}{\partial \theta_i} & \cdots & \frac{\partial [\mathbf{C}(\boldsymbol{\theta})]_{NN}}{\partial \theta_i} \end{bmatrix}. \tag{3.9}$$

3.2 Maximum Likelihood Estimation

When the statistical properties of the observed data is known, the MLE provides the *best* estimate for the parameter estimate in the sense that the variance of $\hat{\boldsymbol{\theta}}$ approaches the CRLB as $N \to \infty$ [Kay93b].

Let $p(\mathbf{y}; \boldsymbol{\theta})$ be the probability density function of $\mathbf{y}$ given parameter $\boldsymbol{\theta}$. The MLE of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}}_{ML} = \arg\max_{\boldsymbol{\theta}} \; p(\mathbf{y}; \boldsymbol{\theta}). \tag{3.10}$$

Due to the monotonically increasing property of the logarithm function, (3.10) is equivalent to

$$\hat{\boldsymbol{\theta}}_{ML} \;\; = \;\; \arg\max_{\boldsymbol{\theta}} \; \ln p(\mathbf{y}; \boldsymbol{\theta}). \tag{3.11}$$

When the signal is a complex Gaussian random variable, $\mathcal{CN}(\boldsymbol{\mu}_y, \mathbf{C}_y)$,

$$p(\mathbf{y}; \boldsymbol{\theta}) = \frac{1}{\pi^N \det(\mathbf{C}_y)} e^{-(\mathbf{y}-\boldsymbol{\mu}_y)^H \mathbf{C}_y^{-1} (\mathbf{y}-\boldsymbol{\mu}_y)}. \tag{3.12}$$

Substitute (3.12) into (3.11), and drop the constant term, which leads to

$$\hat{\boldsymbol{\theta}}_{ML} = \arg\min_{\boldsymbol{\theta}} \; \ln\det(\mathbf{C}_y) + \left(\mathbf{y} - \boldsymbol{\mu}_y\right)^H \mathbf{C}_y^{-1} \left(\mathbf{y} - \boldsymbol{\mu}_y\right). \tag{3.13}$$

This optimization involves a multiple dimension searching procedure, which is computational intensive. For this reason, some fast and suboptimal searching algorithms, e.g. SAGE/RELAX, are attractive in practice [FH94, LS96, CB05]. In these methods, only one or a subset of the parameters are optimized at a time, and the parameter estimation is performed iteratively. These methods are much cheaper than MLE, and the price to pay is the risk to obtain a local minimum instead of a global minimum when the likelihood function is not convex. Empirically a good initial parameter setting is critical in order to converge globally.

3.3 Parameter Estimation for Linear Models

A linear signal model is

$$\mathbf{y} = \mathbf{H}\boldsymbol{\theta} + \mathbf{e}, \tag{3.14}$$

where $\mathbf{H}$ is an $N \times p$ known matrix. The observations is a linear function of the model parameters.

3.3.1 LS Estimate

When $\mathbf{H}$ is a deterministic matrix, the LS estimate of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}}_{LS} = \arg\min_{\boldsymbol{\theta}} \ \|\mathbf{y} - \mathbf{H}\boldsymbol{\theta}\|^2. \tag{3.15}$$

Setting

$$\frac{\partial \|\mathbf{y} - \mathbf{H}\boldsymbol{\theta}\|^2}{\partial \boldsymbol{\theta}} = 0, \tag{3.16}$$

and solving the resulting linear equations, the LS estimate of $\boldsymbol{\theta}$ is

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{LS} &= (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{y}, & (3.17)\\ &= \mathbf{H}^\dagger\mathbf{y}, & (3.18) \end{aligned}$$

where $\mathbf{H}^\dagger$ is the pseudo inverse of $\mathbf{H}$.
For instance, a d^{th} order linear prediction of $h(t+L)$ based on $\mathbf{y}$ is

$$\begin{aligned} \hat{h}(t+L) &= \sum_{k=0}^{d-1} \alpha_k y(t-k), & (3.19)\\ &= \boldsymbol{\alpha}_d^T \mathbf{y}_d, & (3.20) \end{aligned}$$

where L is the prediction horizon, $\boldsymbol{\alpha}_d = [\alpha_0, \alpha_1, \cdots, \alpha_{d-1}]^T$ is the predictor coefficient vector, and $\mathbf{y}_d = [y(t), y(t-1), \cdots, y(t-d+1)]^T$. In vector form, (3.20) can be written as

$$\mathbf{y}_{ls} = \mathbf{Y}_{ls}\boldsymbol{\alpha}_d, \tag{3.21}$$

where

$$\begin{aligned} \mathbf{y}_{ls} &= [y(t), y(t-1), \cdots, y(t-N+L+d)]^T, & (3.22)\\ \mathbf{Y}_{ls} &= \begin{bmatrix} y(t-L) & y(t-L-1) & \cdots & y(t-L-d+1) \\ \vdots & \vdots & \vdots & \vdots \\ y(t-N+d) & y(t-N+d-1) & \cdots & y(t-N+1) \end{bmatrix}. & (3.23) \end{aligned}$$

The LS estimate of $\boldsymbol{\alpha}_d$ is therefore

$$\hat{\boldsymbol{\alpha}}_d = \mathbf{Y}_{ls}^\dagger \mathbf{y}_{ls}, \tag{3.24}$$

where $\mathbf{Y}_{ls}^\dagger$ is the pseudo inverse of $\mathbf{Y}_{ls}$. A *forward-backward* LS estimate of $\boldsymbol{\alpha}_d$, that takes advantage of the stationary property of the model, can be found in Appendix A.

3.3.2 MMSE Estimate

When $\boldsymbol{\theta}$ is random, and $\mathbf{H}$ in (3.14) may or may not be random, the MMSE estimate of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}}_{MMSE} = \arg\min_{\boldsymbol{\theta}} \; \mathbf{E}\left[(\mathbf{y}-\mathbf{H}\boldsymbol{\theta})^H(\mathbf{y}-\mathbf{H}\boldsymbol{\theta})\right]. \tag{3.25}$$

Similarly setting

$$\frac{\partial \mathbf{E}\left[(\mathbf{y}-\mathbf{H}\boldsymbol{\theta})^H(\mathbf{y}-\mathbf{H}\boldsymbol{\theta})\right]}{\partial \boldsymbol{\theta}} = 0, \tag{3.26}$$

and solving the resulting linear equations, the MMSE of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}}_{MMSE} = \mathbf{R}_{HH}^{-1}\mathbf{r}_{Hy}, \tag{3.27}$$

where $\mathbf{R}_{HH} = \mathbf{E}\left[\mathbf{H}\mathbf{H}^H\right]$ and $\mathbf{r}_{Hy} = \mathbf{E}\left[\mathbf{H}\mathbf{y}\right]$.

For the linear prediction model in (3.20), the MMSE estimate of $\boldsymbol{\alpha}_d$ is

$$\hat{\boldsymbol{\alpha}}_d \;=\; \arg\min_{\boldsymbol{\alpha}_d} \; \mathbf{E}\left[\|h(t+L)-\boldsymbol{\alpha}_d^T\mathbf{y}_d\|^2\right]. \tag{3.28}$$

From (3.27),

$$\hat{\boldsymbol{\alpha}}_d \;=\; \mathbf{R}_d^{-1}\mathbf{r}_d, \tag{3.29}$$

where

$$\mathbf{R}_d \;=\; \begin{bmatrix} r_{yy}(0) & r_{yy}(1) & \cdots & r_{yy}(d-1) \\ r_{yy}(-1) & r_{yy}(0) & \cdots & r_{yy}(d-2) \\ \vdots & \vdots & \vdots & \ddots \\ r_{yy}(-d+1) & r_{yy}(-d+2) & \cdots & r_{yy}(0) \end{bmatrix}, \tag{3.30}$$

$$\mathbf{r}_d \;=\; \begin{bmatrix} r_{yy}(L) \\ r_{yy}(L+1) \\ \vdots \\ r_{yy}(L+d-1) \end{bmatrix}, \tag{3.31}$$

$$r_{yy}(\tau) \;=\; \mathbf{E}\left[y(t)y^*(t-\tau)\right]. \tag{3.32}$$

In practice, $\mathbf{R}_d$ might be unknown. It can be estimated from the observations as

$$\hat{\mathbf{R}}_d = \frac{1}{N-d+1}\mathbf{Y}_d\mathbf{Y}_d^H, \tag{3.33}$$

where $\mathbf{Y}_d$ is a $d \times (N-d+1)$ Hankel matrix as

$$\mathbf{Y}_d = \begin{bmatrix} y(t) & y(t-1) & \cdots & y(t-N+d) \\ y(t-1) & y(t-2) & \cdots & y(t-N+d+1) \\ \vdots & \vdots & \ddots & \vdots \\ y(t-d+1) & y(t-d) & \cdots & y(t-N+1) \end{bmatrix}. \tag{3.34}$$

When N goes to infinite, it can be shown that $\hat{\mathbf{R}}_d \to \mathbf{R}_d$ in probability, *i.e.* $\hat{\mathbf{R}}_d$ is a consistent estimator. The correlation coefficients in $\mathbf{r}_d$ can be estimated as

$$\hat{r}_{yy}(k) = \frac{1}{N-k} \sum_{l=t}^{t-N+k+1} y(l)y^*(l-k). \tag{3.35}$$

3.4 State Space Model and Kalman Filter

Signals with a non-stationary property exist widely in practice. One of the most celebrated methods for non-stationary signal processing is the Kalman filter [Kal60]. Suppose a measured signal $y(t)$, for $\mathbf{t} = [t, t-1, \cdots, t-N+1]^T$, can be written in a state space form as:

$$\mathbf{x}(t+1) = \mathbf{\Gamma}_K \mathbf{x}(t) + \mathbf{u}(t), \tag{3.36}$$

$$y(t) = \mathbf{c}^T(t)\mathbf{x}(t) + e(t), \tag{3.37}$$

where $\mathbf{x}(t)$ is the state vector, $\mathbf{\Gamma}_K$ is the state transition matrix, $\mathbf{c}(t)$ is the observation vector. The $\mathbf{u}(t)$ is a complex Gaussian noise vector with PDF, $\mathcal{CN}(\mathbf{0}, \mathbf{Q})$. The state vector $\mathbf{x}(t-N)$ is a complex Gaussian vector with PDF $\mathcal{CN}(\boldsymbol{\mu}_\mathbf{x}, \mathbf{C}_\mathbf{x})$, and $\mathbf{x}(t-N)$ is independent of the $\mathbf{u}(k)$ for $k \geq t-N+1$. The observation noise $e(t)$ is a zero-mean white Gaussian noise with variance σ_e^2. We seek $\hat{\mathbf{x}}(t+1|t)$ and $\hat{\mathbf{x}}(t|t)$, where $\hat{\mathbf{x}}(t+1|t)$ is the estimate of $\mathbf{x}(t+1)$ until $y(t)$ is observed, and $\hat{\mathbf{x}}(t|t)$ is the estimate of $\mathbf{x}(t)$ until $y(t)$ is observed.

Such a problem is solved recursively by the following steps in the Kalman filter:

- **Prediction:**

$$\hat{\mathbf{x}}(t|t-1) = \mathbf{\Gamma}_K \hat{\mathbf{x}}(t-1|t-1). \tag{3.38}$$

- **Minimum Prediction MSE Matrix**:

$$\mathbf{C}_\mathbf{x}(t|t-1) = \mathbf{\Gamma}_K \mathbf{C}_\mathbf{x}(t-1|t-1)\mathbf{\Gamma}_K^H + \mathbf{Q}. \tag{3.39}$$

- **Kalman Gain Vector**:

$$\mathbf{K}(t) = \frac{\mathbf{C}_{\mathbf{x}}(t|t-1)\mathbf{c}(t)}{\sigma_e^2 + \mathbf{c}^T(t)\mathbf{C}_{\mathbf{x}}(t|t-1)\mathbf{c}(t)}. \tag{3.40}$$

- **Correction**:

$$\hat{\mathbf{x}}(t|t) = \hat{\mathbf{x}}(t|t-1) + \mathbf{K}(t)(y(t) - \mathbf{c}^T(t)\hat{\mathbf{x}}(t|t-1)). \tag{3.41}$$

- **Minimum MSE Matrix**:

$$\mathbf{C}_{\mathbf{x}}(t|t) = (\mathbf{I} - \mathbf{K}(t)\mathbf{c}^T(t))\mathbf{C}_{\mathbf{x}}(t|t-1). \tag{3.42}$$

where the mean square error matrices are defined as

$$\mathbf{C}_{\mathbf{x}}(t|t) = \mathbf{E}\left[(\mathbf{x}(t) - \hat{\mathbf{x}}(t|t))(\mathbf{x}(t) - \hat{\mathbf{x}}(t|t))^H\right], \tag{3.43}$$

$$\mathbf{C}_{\mathbf{x}}(t|t-1) = \mathbf{E}\left[(\mathbf{x}(t) - \hat{\mathbf{x}}(t|t-1))(\mathbf{x}(t) - \hat{\mathbf{x}}(t|t-1))^H\right]. \tag{3.44}$$

A detailed derivation of these equation systems from (3.38) to (3.42) can be found in [Kay93b].

3.5 Parameter Estimation for Polynomial Phase Signal Models

3.5.1 Parameter Estimation of a Single PPS

A PPS signal with constant amplitude is

$$\mathbf{y}_1 = \mathbf{h}_1 + \mathbf{e}_1 = \mathbf{a}_1 s_1 + \mathbf{e}_1, \tag{3.45}$$

where

$$\mathbf{a}_1 = [e^{j\phi_1(t)}, \cdots, e^{j\phi_1(t-N+1)}]^T, \tag{3.46}$$

$$\phi_1(t) = \sum_{m=1}^{M_1} \beta_{1,m} t^m. \tag{3.47}$$

The model parameters are

$$\boldsymbol{\theta}_1 = [\beta_{1,1}, \cdots, \beta_{1,M_1}]^T. \tag{3.48}$$

The estimation of polynomial phase parameters has drawn considerable attractions due to its wide wide applicability, such as in radar, sonar and mobile

communications. Besides the ML estimate in [LA92], a least square method was applied to the unwrapped phases in high SNR scenarios [DK90]. Later estimation methods utilizing the special PPS structure were proposed, such as the High-order Ambiguity Function (HAF), Polynomial Phase Transform (PPT), and High order Instantaneous Moment (HIM) method [PF96, PF95, ZGS96]. All these methods assume the order of the polynomial phase to be known. The CRLB in case of constant and time-varying amplitudes of parameter estimation of PPS signal were derived in [PP91a] and [FF95, GNS99] respectively, where it was shown that the estimation of the parameters of the polynomial phase is not affected significantly by modulated amplitudes at high SNR. So that the maximum likelihood type estimators, which optimize a criterion with respect to only phase parameters, is approximated by the Nonlinear LS estimate (NLLS), i.e.

$$\hat{\boldsymbol{\theta}}_{1,NLLS} = \arg\min_{\boldsymbol{\theta}_1} \left[(\mathbf{y}_1 - \mathbf{h}_1)^H (\mathbf{y}_1 - \mathbf{h}_1) \right]. \tag{3.49}$$

Generally there is no closed form solution to the NLLS optimization problem. A multiple dimension search is inevitable. In case of MC-PPS, an iterative parameter estimation of multiple chirp signals was proposed in [IAMH97].

An alternative to the NLLS criterion, a Nonlinear *Instantaneous* Least Square (NILS) estimate of $\boldsymbol{\theta}$ is, i.e. [GNS99, PZ07],

$$\hat{\boldsymbol{\theta}}_{1,NILS} = \arg\min_{\boldsymbol{\theta}} \sum_{k=t}^{t-N+n} \|\mathbf{y}_1(k) - \mathbf{h}_1(k)\|^2, \tag{3.50}$$

where

$$\mathbf{y}_1(k) = [y_1(k), \cdots, y_1(k-n+1)]^T, \tag{3.51}$$
$$\mathbf{h}_1(k) = [h_1(k), \cdots, h_1(k-n+1)]^T, \tag{3.52}$$

where n is the length of the local interval and is a user's choice. The LS estimate of the instantaneous amplitude $s_1(k)$ is

$$\hat{s}_1(k) = \left(\hat{\mathbf{a}}_{1,k}^H \hat{\mathbf{a}}_{1,k}\right)^{-1} \hat{\mathbf{a}}_{1,k}^H \mathbf{y}_1(k), \tag{3.53}$$
$$= \hat{\mathbf{a}}_{1,k}^{\dagger} \mathbf{y}_1(k), \tag{3.54}$$

where

$$\mathbf{a}_{1,k} = [e^{j\phi_1(k)}, \cdots, e^{j\phi_1(k-n+1)}]^T. \tag{3.55}$$

The selection of n depends on the rate of the time variation of the amplitude. A small n gives a smoother cost function for model parameters compared to

those given by a large n [JAG99]. This property makes the convergence of parameter searching in multiple dimension spaces less sensitive to the initial parameter setting when n is small. Meanwhile the resolution of the model parameters is reduced. When $n = N$, NILS is the same as the standard NLLS [BW88].

3.5.2 Iterative Parameter Estimation of MC-PPS

In MC-PPS scenarios with p PPS components, the model parameters are

$$\boldsymbol{\theta} = [\boldsymbol{\theta}_1^T, \cdots, \boldsymbol{\theta}_p^T]^T, \tag{3.56}$$

$$\boldsymbol{\theta}_i = [\beta_{i,1}, \cdots, \beta_{i,M_i}]^T, \tag{3.57}$$

where $\boldsymbol{\theta}$ contains $\sum_{i=1}^{p} M_i$ parameters. Motivated by the idea in SAGE/RELAX [FH94, LS96], an iterative parameter estimation using NILS and NLLS criterion is proposed in [CVF07a] and [CV07] respectively. A block diagram of the method using NLLS criterion proposed in [CV07] is given in Figure 3.1, where the number of PPS and the order of the polynomial phase are assumed to be known.

A demonstration example of the iterative method is given below. The number of PPS is $p = 3$, and the signal parameters are given in Table 3.1.

Table 3.1: Parameter of MC-PPS

	$i = 1$	$i = 2$	$i = 3$
$\beta_{i,1}$	-0.31	0.1	0.2
$\beta_{i,2}$	-0.05×10^{-3}	-0.3×10^{-3}	-0.16×10^{-3}
$s_i(t)$	1	1	1

The total data length $N = 250$, the SNR is 10 dB. The power spectrum of original signal $\mathbf{y}$, the residual signal $\mathbf{y}_r$, and the reconstructed signal $\hat{\mathbf{h}}_i$ associated with the PPS No. 1, 2 and 3 are plotted in the subplots (a), (b), and (c) in Figure 3.2 respectively. From these plots it can be seen that the overlapped spectrum can be classified by the proposed methods. Note that the number of clusters and the order of the polynomial are assumed to be known.

3.6 Subspace Based Frequency Estimation

Frequency estimation has been studied extensively in array signal processing in [KV96] and the reference therein. Subspace based frequency estimation

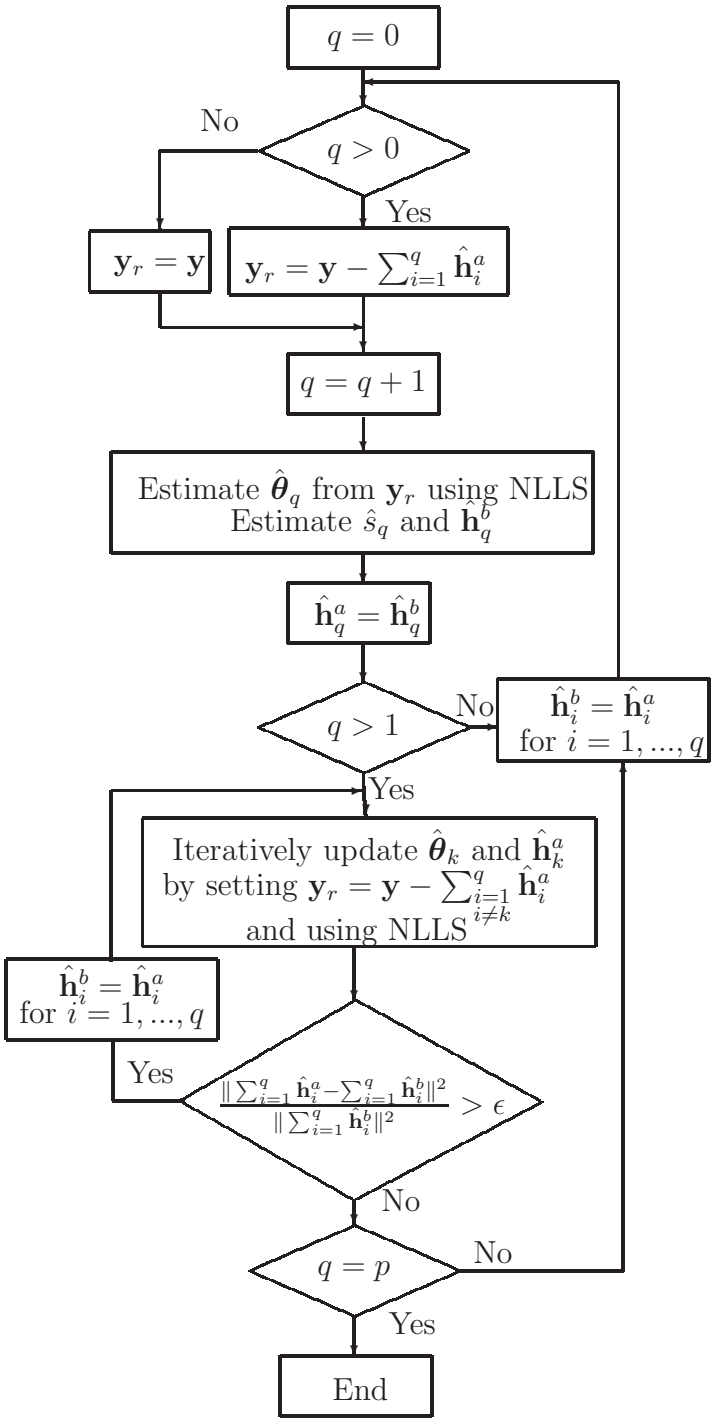

Figure 3.1: A block diagram of the iterative parameter estimate of MC-PPS using NLLS.

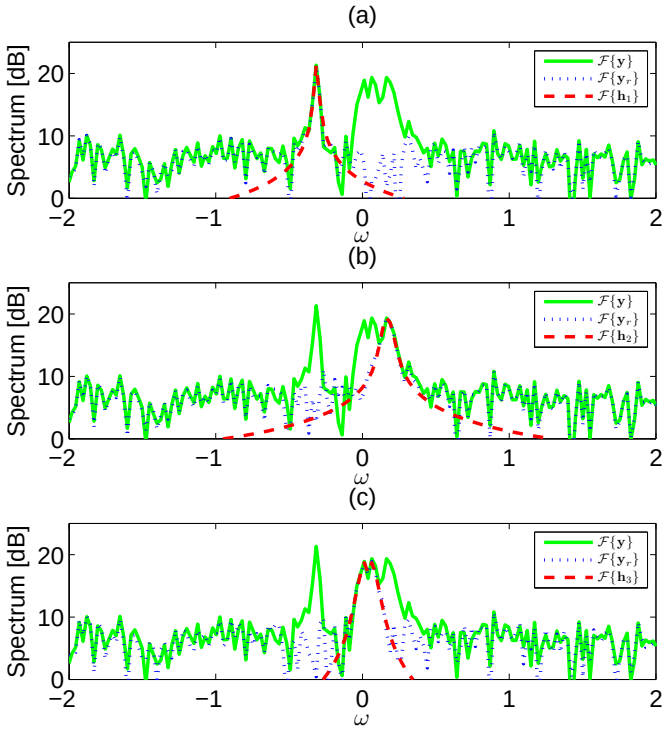

Figure 3.2: Classified power spectrum of a 3-component PPS signal with parameters given in Table 3.1.

algorithms, such as MUSIC (Multiple Signal Classification) [Sch81] and ESPRIT (Estimation of Signal Parameters via Rotation Invariance Techniques) [PRK85, RPK86] are introduced in this section.

3.6.1 MUSIC Pseudo Spectrum

The MUSIC Pseudo Spectrum (MPS) is defined as

$$p_{mu}(\omega) = \frac{\mathbf{a}(\omega)^H \mathbf{a}(\omega)}{\mathbf{a}(\omega)^H \mathbf{\Pi}_{\mathbf{a}}^{\perp} \mathbf{a}(\omega)}, \quad \omega \in [-\omega_m, \omega_m], \tag{3.58}$$

where $\mathbf{\Pi}_{\mathbf{a}}^{\perp} = \mathbf{U}_n \mathbf{U}_n^H$, and the columns of $\mathbf{U}_n$ are the eigenvectors spanning the noise space of the covariance matrix of the data sequence $\mathbf{y}$, i.e.,

$$\mathbf{R}_{yy} = \mathbf{E}[\mathbf{y}\mathbf{y}^H]. \tag{3.59}$$

It can be obtained by taking eigenvalue decomposition of $\mathbf{R}_{yy}$ as

$$\mathbf{R}_{yy} = [\mathbf{U}_s \ \mathbf{U}_n] \begin{bmatrix} \mathbf{\Lambda}_s & \\ & \mathbf{\Lambda}_n \end{bmatrix} \begin{bmatrix} \mathbf{U}_s^H \\ \mathbf{U}_n^H \end{bmatrix}, \tag{3.60}$$

$$= \mathbf{U}_s \mathbf{\Lambda}_s \mathbf{U}_s^H + \mathbf{U}_n \mathbf{\Lambda}_n \mathbf{U}_n^H, \tag{3.61}$$

where a rank estimation algorithm is necessary, and the eigenvalues are ordered in non-increasing order. In practice, an estimate of $\mathbf{R}_{yy}$ similar to (3.33) is used. In case of high SNR, MPS is not sensitive to under-estimated dimension of noise subspace, while an over-estimated dimension of noise space will attenuate certain weak frequency components. The $\mathbf{a}(\omega)$ is the DFT vector associated with frequency ω. The frequency estimate can be obtained from the peaks of $p_{mu}(\omega)$.

3.6.2 ESPRIT

ESPRIT utilizes the rotation invariance property of the sinusoidal signal subspace for frequency estimation. In (2.10), the signal space is spanned by the columns of $\mathbf{A}(\boldsymbol{\omega})$. For convenience, $\mathbf{A}(\boldsymbol{\omega})$ is denoted as $\mathbf{A}$ in the following discussion in this chapter. Let

$$\mathcal{O}_1 = \mathbf{A}_{1:N-1}, \tag{3.62}$$

$$\mathcal{O}_2 = \mathbf{A}_{2:N}, \tag{3.63}$$

which contain the first $N-1$ and the last $N-1$ rows of $\mathbf{A}$ respectively. Note that $\mathcal{O}_1$ and $\mathcal{O}_2$ are two temporally displaced subsets of the basis functions of $\mathbf{A}$. The rotation invariance property is interpreted as

$$\mathcal{O}_2 = \mathcal{O}_1 \mathbf{\Phi}, \tag{3.64}$$

where

$$\mathbf{\Phi} = \begin{bmatrix} e^{j\omega_1} & & & \\ & e^{j\omega_2} & & \\ & & \ddots & \\ & & & e^{j\omega_p} \end{bmatrix}, \tag{3.65}$$

The matrix $\mathbf{\Phi}$ is the rotation operator, which is a diagonal unitary matrix in this case. Assume a matrix $\mathbf{B}_{N\times p} = [\mathbf{b}_1, \cdots, \mathbf{b}_p]$ contains another set of basis function spanning the same signal subspace of the sinusoidal subspace, and

$$\mathbf{b}_i = t_{i,1}\mathbf{v}_1 + t_{i,2}\mathbf{v}_2 + \cdots + t_{i,p}\mathbf{v}_p, \tag{3.66}$$

$$= \mathbf{A}\mathbf{t}_i, \tag{3.67}$$

where $\mathbf{t}_i = [t_{i,1}, \cdots, t_{i,p}]^T$ is the coordinates of $\mathbf{b}_i$ with respect to the basis in $\mathbf{A}$. So there exists a $p \times p$ transform matrix

$$\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \cdots, \mathbf{t}_p], \tag{3.68}$$

which satisfies

$$\mathbf{B} = \mathbf{AT}, \tag{3.69}$$

where $\mathbf{T}$ is an invertible matrix [HJ85]. Define two sub matrices

$$\mathcal{B}_1 = \mathbf{B}(1 : N-1, :), \tag{3.70}$$
$$\mathcal{B}_2 = \mathbf{B}(2 : N, :), \tag{3.71}$$

which contain the first $N-1$ and the last $N-1$ rows of $\mathbf{B}$ respectively. So

$$\begin{aligned} \mathcal{B}_2 &= \mathcal{O}_2\mathbf{T} = \mathcal{O}_1\mathbf{\Phi T} = (\mathcal{O}_1\mathbf{T})(\mathbf{T}^{-1}\mathbf{\Phi T}) \\ &= \mathcal{B}_1\mathbf{\Psi}, \end{aligned} \tag{3.72}$$

where $\mathbf{\Psi} = \mathbf{T}^{-1}\mathbf{\Phi T}$. The matrices $\mathbf{\Phi}$ and $\mathbf{\Psi}$ are similar, and have the same eigenvalues. So the frequency estimate can be obtained from the phases of the eigenvalues of $\mathbf{\Psi}$.

Based on (3.64), the frequency estimate can be obtained by the following steps:

1. Find a set of basis functions which span the signal subspace, *i.e.* a similar matrix to $\mathcal{O}$;

2. Forming $\mathcal{O}_1$ and $\mathcal{O}_2$ as in (3.62) and (3.63);

3. The LS estimate of the rotation matrix $\mathbf{\Phi}$ is

$$\begin{aligned} \hat{\mathbf{\Phi}} &= (\mathcal{O}_1^H\mathcal{O}_1)^{-1}\mathcal{O}_1^H\mathcal{O}_2 \\ &= \mathcal{O}_1^\dagger\mathcal{O}_2; \end{aligned} \tag{3.73}$$

4. The frequency estimate $\hat{\boldsymbol{\omega}}$ can be obtained from the phases of the eigenvalues of $\hat{\mathbf{\Phi}}$.

3.6.3 Signal Subspace Estimation Using SVD

The signal subspace estimate method which uses Singular Value Decomposition (SVD) is termed *Kung's algorithm* [Kun78]. A Hankel matrix $\mathbf{H}$ is formed by the data sequence as

$$\mathbf{H} = \begin{bmatrix} y(t) & y(t-1) & \dots & y(t-N+M) \\ y(t-1) & y(t-2) & \dots & y(t-N+M-1] \\ \vdots & \vdots & \ddots & \vdots \\ y(t-M+1) & y(t-M) & \dots & y(t-N+1) \end{bmatrix}, \tag{3.74}$$

where M is chosen as an integer close to $N/2$ as a rule of thumb. The SVD of $\mathbf{H}$ is

$$\mathbf{H} = [\mathbf{U}_s \;\; \mathbf{U}_n] \begin{bmatrix} \mathbf{\Sigma}_s & \\ & \mathbf{\Sigma}_n \end{bmatrix} \begin{bmatrix} \mathbf{V}_s^H \\ \mathbf{V}_n^H \end{bmatrix}, \tag{3.75}$$

where the singular values are ordered in non-increasing order. The *observability* matrix $\mathcal{O}$ is formed as

$$\mathcal{O} = \mathbf{U}_s \cdot \mathbf{\Sigma_s}^{1/2}, \tag{3.76}$$

and $\mathcal{O}_1$ and $\mathcal{O}_2$ can be obtained as in (3.62) and (3.63). Finally, $\hat{\mathbf{\Phi}}$ is obtained as in (3.73).

3.6.4 Unitary ESPRIT

It was observed that the eigenvalues of $\mathbf{\Phi}$ have unit norm. To exploit this property, the Unitary ESPRIT algorithms was proposed in [HN95]. The Unitary ESPRIT algorithms begin with forming the data matrix $\mathbf{H}_u$,

$$\mathbf{H}_u = [\mathbf{H} \;\; \boldsymbol{\Pi}_M \bar{\mathbf{H}}], \tag{3.77}$$

where $\mathbf{H}$ is the same as in (3.74). The over bar denotes complex conjugation without transposition. The $\boldsymbol{\Pi}_M$ is the $M \times M$ exchange matrix with ones on its anti-diagonal and zeros elsewhere as given in (3.78).

$$\boldsymbol{\Pi}_M = \begin{bmatrix} & & & 1 \\ & & 1 & \\ & \cdot & & \\ 1 & & & \end{bmatrix}. \tag{3.78}$$

Then the SVD is performed upon $\mathbf{H}_u$ instead of on $\mathbf{H}$. The remaining steps are identical to Kung's algorithm. It was found that Unitary ESPRIT has better performance than ESPRIT [HN95].

Chapter 4

Model Order Selection and Detection of MC-PPS

Model order selection and detection of MC-PPS is critical for the parameter estimation. Such a problem is formulated as the determination of the number of PPS signals, p, and the orders of polynomial phases of PPS components, M_i for $i = 1, \cdots, p$. For convenience, the signal model is repeated below:

$$y(t) = \sum_{i=1}^{p} h_i(t) + e(t), \tag{4.1}$$

$$h_i(t) = s_i(t) e^{j\phi_i(t)}, \tag{4.2}$$

$$\phi_i(t) = \sum_{m=1}^{M_i} \beta_{i,m} t^m, \tag{4.3}$$

where $e(t)$ is an additive noise with zero-mean and variance σ_e^2. The model parameters are collected in $\boldsymbol{\beta}$ as

$$\boldsymbol{\beta} = \left[\boldsymbol{\beta}_1^T, \cdots, \boldsymbol{\beta}_p^T\right]^T, \tag{4.4}$$

$$\boldsymbol{\beta}_i = [\beta_{i,M_i}, \beta_{i,M_i-1}, \cdots, \beta_{i,1}]^T, \tag{4.5}$$

where $\boldsymbol{\beta}_i$ contains the polynomial phase parameters associated with the i^{th} component, but the parameters are put in a reverse sequence as compared to (3.57). The orders of the PPS components are mixed and the total number of model parameters is $\sum_{i=1}^{p} M_i$. The time-varying amplitude $s_i(t)$ has a low pass property. In vector form,

$$\mathbf{y} = \mathbf{h} + \mathbf{e} = \sum_{i=1}^{p} \mathbf{h}_i + \mathbf{e}, \tag{4.6}$$

where

$$\mathbf{y} = [y(t), \cdots, y(t-N+1)]^T, \tag{4.7}$$
$$\mathbf{e} = [e(t), \cdots, e(t-N+1)]^T, \tag{4.8}$$
$$\mathbf{h}_i = [h_i(t), \cdots, h_i(t-N+1)]^T. \tag{4.9}$$

Due to the non-stationary property of PPS, the classical nonparametric model order selection methods, such as Akaike Information Criterion [Aka74] and Minimum Description Length [Ris83], cannot be used directly. In this chapter, an iterative model order selection and detection of MC-PPS is introduced. This method does not require the highest model order and the number of PPS as a priori knowledge. Most of the following content is contained in [CV07], except that a Wald test based model order selection of the polynomial phases is proposed [Kay93a]. The Wald test based model order selection in single PPS scenarios is addressed first. Then the detection of the number of PPS components based on the size of the residuals is presented. Finally, the iterative detection and estimation method is proposed.

It is worth noting that all the detectors proposed in this chapter assume that the parameter estimates are correct, i.e $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$. The estimation of $\boldsymbol{\beta}$ can be made using the NILS criterion in (3.50) [CVF07a], and the NLLS criterion in (3.49) [PZ07, CV07]. Using the NILS criterion, the cost function is smoothed regarding the model parameters, and the local searching of the optimization is less sensitive to the initial parameter settings. But the resolution of parameter estimation is reduced. For this reason, a higher resolution parameter estimation using the NLLS criterion is adopted in this chapter.

4.1 Model Order Selection of Polynomial Phase Signals Using Wald Test

In this section, a parametric model order selection using Wald test is proposed [Kay93a]. The highest model order is not required as a priori knowledge. A detailed description of this method is presented using a toy example of a single PPS with a real and constant amplitude as follows. Note that the Wald test is in principle equivalent to the Generalized Likelihood Ratio Test (GLRT) based detector [Kay93a].

A toy example of model order selection using Wald test

A single PPS corrupted by noise with model order $M_1 = 2$ is

$$y_1(t) = s_1 e^{j\phi_1(t)} + e_1(t), \tag{4.10}$$

where s_1 is a real constant amplitude, $e_1(t)$ is an additive noise with zero-mean and variance $\sigma_{e_1}^2$, and $\phi_1(t)$ is the polynomial phase,

$$\phi_1(t) = \sum_{m=1}^{M_1} \beta_{1,m} t^m = \beta_{1,1} t + \beta_{1,2} t^2. \tag{4.11}$$

The model parameters are collected in

$$\boldsymbol{\beta}_1 = [\beta_{1,2}, \beta_{1,1}]^T, \tag{4.12}$$

and $\mathbf{y}_1 = [y_1(t), y_1(t-1), \cdots, y_1(t-N+1)]^T$. Note that the amplitude s_1 and the noise variance $\sigma_{e_1}^2$ are not listed as model parameters, although they are estimated in different places.

The detection problem is:

$$\mathcal{H}_0 \;:\; \beta_{1,m+1} = 0, \tag{4.13}$$
$$\mathcal{H}_1 \;:\; \beta_{1,m+1} \neq 0, \tag{4.14}$$

where the *null* hypothesis assumes m is the highest model order, while the *alternative* hypothesis assumes the highest model order is at least $m+1$. The already detected and estimated parameters under $\mathcal{H}_0$ and $\mathcal{H}_1$ are collected in

$$\hat{\boldsymbol{\beta}}_{1,\mathcal{H}_0} = [0, \hat{\beta}_{1,m}, \cdots, \hat{\beta}_{1,1}]^T, \tag{4.15}$$
$$\hat{\boldsymbol{\beta}}_{1,\mathcal{H}_1} = [\hat{\beta}_{1,m+1}, \hat{\beta}_{1,m}, \cdots, \hat{\beta}_{1,1}]^T. \tag{4.16}$$

Given $\hat{\boldsymbol{\beta}}_{1,\mathcal{H}_1}$, the Wald test for model order selection is, (see. e.g. [Kay93a]),

$$\mathcal{T}_W = \hat{\beta}_{1,m+1} \left(\left[\mathbf{F}^{-1} \right]_{1,1} \right)^{-1} \hat{\beta}_{1,m+1} \underset{m+1}{\overset{m}{\lessgtr}} \gamma_m, \tag{4.17}$$

where $\mathbf{F}$ is the Fisher information matrix, given in Appendix D, and $[\mathbf{F}^{-1}]_{1,1}$ is the element at the first row and first column of $\mathbf{F}^{-1}$.

To complete the design of the detector, the threshold γ_m can be selected at a desired level of *false alarm* according to the following test distribution: i.e. [Kay93a],

$$\mathcal{T}_W = \chi_1^2 \text{ under } \mathcal{H}_1,\ m = M_1, \tag{4.18}$$

where χ_1^2 is a chi-square distribution with one degree of freedom. In practice, the Fisher information matrix $\mathbf{F}$ is heavily ill-conditioned, which makes difficulty to implement the test in (4.17). An approximation of the CRLB of the parameter estimation was derived in [PP91b], but it is valid for a small N only. The remedy is assuming the first m parameter estimate to be exact,

i.e. $[\hat{\beta}_{1,m}, \cdots, \hat{\beta}_{1,1}]^T = [\beta_{1,m}, \cdots, \beta_{1,1}]^T$, one can use $F_{1,1}$ as $\left([\mathbf{F}^{-1}]_{1,1}\right)^{-1}$ in (4.17). In other words, only $\beta_{1,m+1}$ is considered as the unknown parameter. Such an approximation is also motivated by the fact that the CRLB for the polynomial phase parameters decreases *super* fast with N, e.g. $1/N^3$ for the frequency. So the approximate Wald test is

$$\mathcal{T}_W \approx \hat{\beta}_{1,m+1}\mathbf{F}_{1,1}\hat{\beta}_{1,m+1} \underset{m+1}{\overset{m}{\lessgtr}} \gamma_m. \tag{4.19}$$

The detection using (4.19) starts with a test on the first order parameter, $m = 0$, $\beta_{m+1} = [\beta_{1,1}]$. Once $\mathcal{H}_1$ is accepted, or $\beta_{1,1}$ is estimated and $\beta_{1,1} \neq 0$ is detected, a new test on $\beta_{m+1} = [\beta_{1,2}]$ is performed. Such a procedure is carried out until $\mathcal{H}_1$ is rejected, and $m = M_1$ is selected.

In this example, s_1 is assumed to be *real*, and the LS estimate of s_1, given $\hat{\boldsymbol{\beta}}_{1,\mathcal{H}_1}$, is

$$\hat{s}_1 = \tilde{\mathbf{a}}^{\dagger}_{\beta_1}\tilde{\mathbf{y}}_1, \tag{4.20}$$

where

$$\begin{aligned}
\tilde{\mathbf{y}}_1 &= [\mathrm{Re}\left[\mathbf{y}_1^T\right], \mathrm{Im}\left[\mathbf{y}_1^T\right]]^T, && (4.21)\\
\tilde{\mathbf{a}}^{\dagger}_{\beta_1} &= (\tilde{\mathbf{a}}^T_{\beta_1}\tilde{\mathbf{a}}_{\beta_1})^{-1}\tilde{\mathbf{a}}^T_{\beta_1}, && (4.22)\\
\tilde{\mathbf{a}}_{\beta_1} &= [\mathrm{Re}\left[\hat{\mathbf{a}}^T_{\beta_1}\right], \mathrm{Im}\left[\hat{\mathbf{a}}^T_{\beta_1}\right]]^T, && (4.23)\\
\hat{\mathbf{a}}_{\beta_1} &= [e^{j\hat{\phi}_{1,m+1}(t)}, \cdots, e^{j\hat{\phi}_{1,m+1}(t-N+1)}]^T, && (4.24)\\
\hat{\phi}_{1,m+1}(t) &= \sum_{k=1}^{m+1}\hat{\beta}_{1,k}t^k, && (4.25)
\end{aligned}$$

where $\mathrm{Re}\left[\cdot\right]$ and $\mathrm{Im}\left[\cdot\right]$ are the real part and the imaginary part of the argument respectively. The ML estimate of $\sigma^2_{e_1}$ is

$$\hat{\sigma}^2_{e_1} = \frac{1}{N}\left\|\mathbf{\Pi}^{\perp}_{\tilde{\mathbf{a}}_{\beta_1}}\tilde{\mathbf{y}}_1\right\|^2, \tag{4.26}$$

where

$$\mathbf{\Pi}^{\perp}_{\tilde{\mathbf{a}}_{\beta_1}} = \mathbf{I}_N - \tilde{\mathbf{a}}_{\beta_1}\tilde{\mathbf{a}}^{\dagger}_{\beta_1}. \tag{4.27}$$

The detection statistics of $\mathcal{T}_W$ in (4.19) is evaluated by simulations:

- The constant amplitude is $s_1 = 1$;
- The order of the polynomial phase is $M_1 = 2$, and the parameters are $\beta_{1,1} = 0.41$, and $\beta_{1,2} = 3 \times 10^{-5}$;
- SNR = 0 dB and 10dB;

- The estimate of $\boldsymbol{\beta}_1$ in (4.12) using the NLLS (3.49) is obtained using the *Simplex* method [NM65];
- The length of data is $N = 100$;
- Number of simulation is 200;
- The initial parameter setting for a local search is

$$\text{Initial } \beta_{1,1} = \arg\max_{\beta_{1,1}} \mathcal{F}\{\mathbf{y}_1\}, \tag{4.28}$$

  where $\mathcal{F}\{\mathbf{y}_1\}$ is the DTFT of $\mathbf{y}_1$. The initial values for $\beta_{1,m}$, $m > 1$, are taken to be 0.

The Cumulative Distribution Functions (CDF) of $\mathcal{T}_W$ for different SNR's are given in Figure 4.1, together with the CDF of χ^2_1. From this figure, it can be seen that the CDF of $\mathcal{T}_W$ is close to that of a random variable with χ^2_1 distribution.

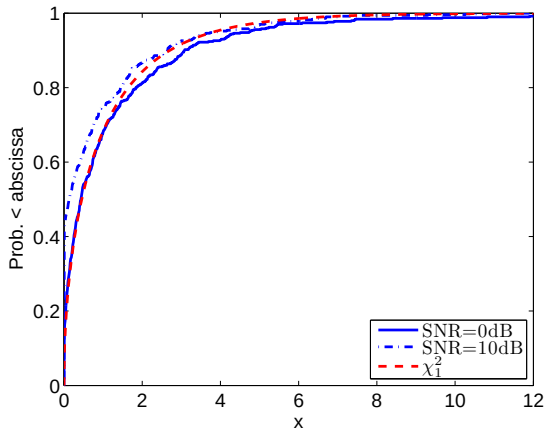

Figure 4.1: The CDF's of the Wald test from Monte Carlo simulations and χ^2_1. In the simulations, $p = 1$, $M_1 = 2$, $\beta_{1,1} = 0.41$, $\beta_{1,2} = 3 \times 10^{-5}$, $s_1 = 1$, $N = 100$, SNR $= 0, 10$dB. In total, 200 simulations are performed for each SNR.

4.2 Detection of the Number of PPS Components

Detection of the number of MC-PPS, p, in (4.1), without a priori knowledge of the mixed model orders of the polynomial phases is a nontrivial task.

However, an iterative solution is to detect a single PPS at a time until all PPS's are discovered. Assume the model parameters associated with the first p PPS's have been detected and estimated, and are put in the vector

$$\hat{\boldsymbol{\theta}} = \left[\hat{\boldsymbol{\theta}}_1^T, \cdots, \hat{\boldsymbol{\theta}}_p^T\right]^T. \tag{4.29}$$

Define the residual signal

$$\begin{aligned}
\mathbf{y}_r &= \mathbf{y} - \sum_{i=1}^{p} \hat{\mathbf{h}}_i, && (4.30)\\
\hat{\mathbf{h}}_i &= [\hat{h}_i(t), \cdots, \hat{h}_i(t-N+1)]^T, && (4.31)\\
\hat{h}_i(t) &= \hat{s}_i(t) e^{j\hat{\phi}_i(t)}, && (4.32)\\
\hat{\phi}_i(t) &= \sum_{m=1}^{\hat{M}_i} \hat{\beta}_{i,m} t^m, && (4.33)
\end{aligned}$$

where $\hat{s}_i(t)$ can be estimated using (3.54). Then the detection problem is formulated as

$$\begin{aligned}
\mathbb{H}_0 &: \mathbf{y}_r = \mathbf{e} && (4.34)\\
\mathbb{H}_1 &: \mathbf{y}_r = \mathbf{e} + \text{“Undetected components”}. && (4.35)
\end{aligned}$$

The detector based on the size of residual is

$$\mathcal{T}_p = \frac{2}{\sigma_e^2} \left\| \mathbf{\Pi}^{\perp}_{\hat{\mathbf{A}}_p} \mathbf{y} \right\|^2 \underset{\mathbb{H}_1}{\overset{\mathbb{H}_0}{\lessgtr}} \gamma_p, \tag{4.36}$$

where

$$\begin{aligned}
\mathbf{\Pi}^{\perp}_{\hat{\mathbf{A}}_p} &= \mathbf{I}_N - \hat{\mathbf{A}}_p \hat{\mathbf{A}}_p^{\dagger}, && (4.37)\\
\hat{\mathbf{A}}_p^{\dagger} &= \left(\hat{\mathbf{A}}_p^H \hat{\mathbf{A}}_p\right)^{-1} \hat{\mathbf{A}}_p^H, && (4.38)\\
\hat{\mathbf{A}}_p &= [\hat{\mathbf{a}}_1, \cdots, \hat{\mathbf{a}}_p], && (4.39)\\
\mathbf{a}_i &= [e^{j\hat{\phi}_i(t)}, \cdots, e^{j\hat{\phi}_i(t-N+1)}]^T. && (4.40)
\end{aligned}$$

Note that σ_e^2 is assumed to be known or can be estimated in advance. For example, one can estimate σ_e^2 from the out-of-band spectra of a band limited signal in $\omega_B = [\omega_l, \omega_h]$, as

$$\hat{\sigma}_e^2 = \frac{1}{2\pi - (\omega_h - \omega_l)} \int_{\omega \notin \omega_B} p(\omega) d\omega, \tag{4.41}$$

where $p(\omega)$ is the power spectrum of $y(t)$. When p is the true number of the MC-PPS, and the parameters in $\hat{\boldsymbol{\theta}}$ are correct, the distribution of $\mathcal{T}_p$ is

$$\mathcal{T}_p = \chi^2_v \quad \text{under } \mathcal{H}_0 \text{ and } p \text{ is true,} \tag{4.42}$$

where v is the degrees of freedom, $v = 2N - p$. The reduction of degrees of freedom by p is due to the projection operation. With the distribution in (4.42), one can select γ_p at a desired level of *false alarm* for the detection of the number of MC-PPS. However, this detector cannot work alone. It has to be implemented together with the model order selection of the polynomial phase in (4.19) or the GLRT detector in [CV07].

4.3 Summary of Iterative Model Order Selection and Detection of MC-PPS

An iterative detection and estimation procedure is suggested as follows:

1. Initialize the number of signal components $p = 0$, and let the residual signal be $\mathbf{y}_r = \mathbf{y}$;

2. Decide if any new signal components exist in $\mathbf{y}_r$ using $\mathcal{T}_p$. If no new component is detected, Stop. Otherwise go to Step 3);

3. Let $p = p+1$ and estimate and select the model order using $\mathcal{T}_W$, where the parameter search of $\beta_{p,1}$ starts with

$$\text{Initial } \beta_{p,1} = \arg\max_{\beta_{p,1}} \mathcal{F}\{\mathbf{y}_r\}, \tag{4.43}$$

   where $\mathcal{F}\{\mathbf{y}_r\}$ is the DTFT of $\mathbf{y}_r$. The initial values for $\beta_{p,m}$, $m > 1$, can be taken to be 0;

4. If $p > 1$ components are detected, iteratively update the model order $\hat{M}_k$ and re-estimate $\hat{\boldsymbol{\theta}}_k$, for $k = 1, \cdots, p$, by setting

$$\mathbf{y}_r = \mathbf{y} - \sum_{\substack{i=1 \\ i \neq k}}^{p} \hat{\mathbf{h}}_i. \tag{4.44}$$

   In this step, the old parameter estimation is set to be the initial values of new NLLS estimate;

5. Let

$$\mathbf{y}_r = \mathbf{y} - \sum_{i=1}^{p} \hat{\mathbf{h}}_i, \tag{4.45}$$

and go to Step 2).

A block diagram of the suggested procedure is given in Figure 4.2. Note that the initial parameter setting in Step 3) give rise to a local search of the NLLS criterion, for example the Simplex method [NM65].

The distribution of $\mathcal{T}_p$ is investigated using Monte Carlo simulations, with known p and M_i. The simulation setting-up is:

- The number of PPS is $p = 3$;
- The signal parameters are given in Table 4.1;

Table 4.1: Parameter of MC-PPS

	$i = 1$	$i = 2$	$i = 3$
$\beta_{i,1}$	0.41	-0.1	-0.17
$\beta_{i,2}$	3×10^{-5}	2×10^{-5}	-3×10^{-5}
$s_i(t)$	1	1	1

- SNR = 10 dB;
- The NLLS estimate of $\boldsymbol{\theta}_k$ in (3.49) is obtained using the *Simplex* method [NM65];
- The length of data is $N = 100$;
- Number of simulations is 200;
- The degree of freedom $v = 2N - p = 197$;
- The orders of the polynomial phases and the noise variance are known.

The other simulation set-up are same as before. The CDF of $\mathcal{T}_p$ is given in Figure 4.3, together with the CDF of χ^2_v and $v = 197$. A good agreement between these two CDF's is observed.

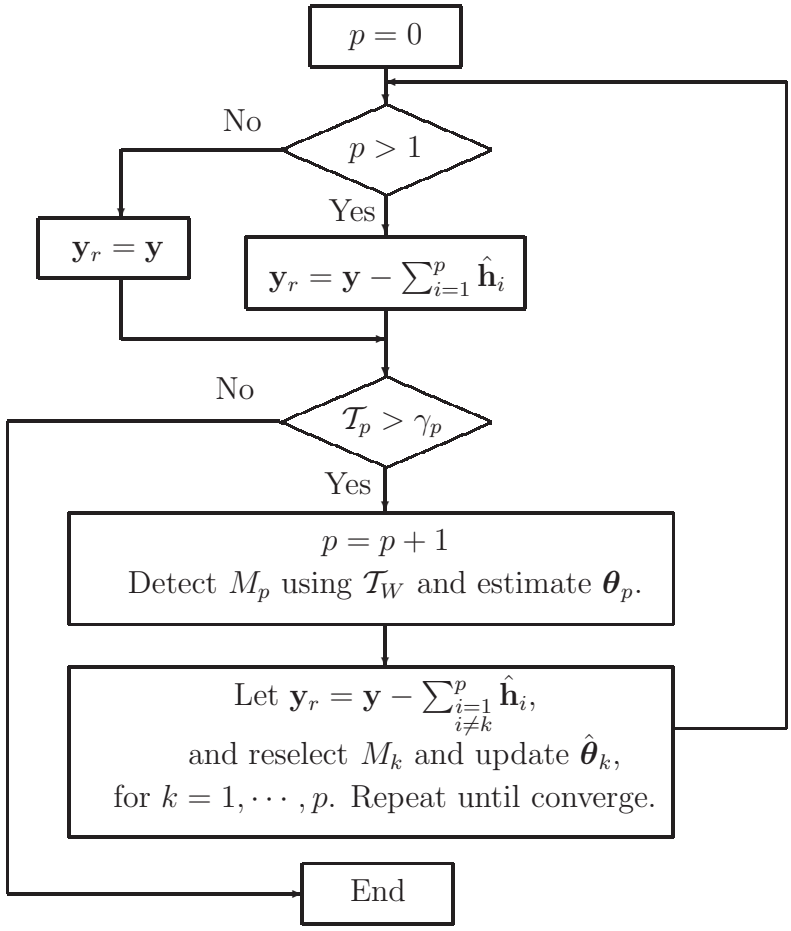

Figure 4.2: A block diagram of model order selection and detection of MC-PPS.

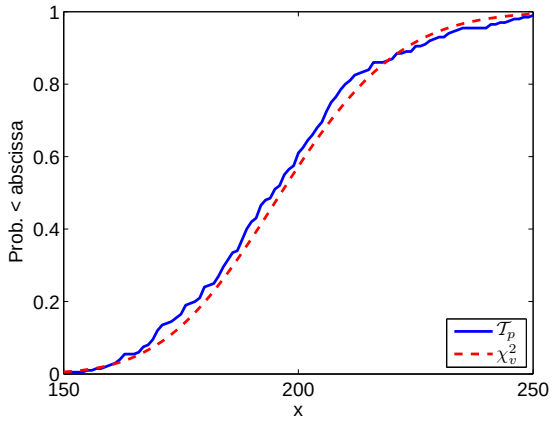

Figure 4.3: The CDF of $\mathcal{T}_p$ from Monte Carlo simulations and χ^2_v. In the simulations, $p = 3$, the signal parameters are given in Table 4.1, $N = 200$, SNR=10 dB and $M_1 = M_2 = M_3 = 2$ and σ_e^2 are known. In total, 200 simulations are performed.

Chapter 5

Model Based Channel Prediction

This chapter contains a summary of the channel predictors proposed by the author based on sinusoidal models and polynomial phase signals. These predictors are published in [CEV07, CVF07b, CVF07a, CV07]. A short overview of previous studies of channel prediction is presented as well.

5.1 Previous Studies of Channel Prediction

Early works on radio channel prediction includes [DHHH00, HW98, AJHF99, EK99]. The Linear Prediction is proposed and investigated in [DHHH00, Ekm02]. Deterministic sinusoidal modeling based channel prediction is suggested in [HW98, AJHF99]. In [EK99], a nonlinear predictor based on Multivariate Adaptive Regression Splines (MARS) is introduced. Recently, a new channel estimation and prediction method using Discrete Prolate Spheroidal Sequences was proposed in [ZM05], where the band limited property of radio channels is utilized.

Linear prediction

A d^{th} order LP of $h(t+L)$, as seen in Figure 5.1, is

$$\hat{h}(t+L) = \sum_{k=0}^{d-1} \alpha_k y(t-k) = \boldsymbol{\alpha}_d^T \mathbf{y}_d, \tag{5.1}$$

where $\boldsymbol{\alpha}_d$ is the coefficient vector of the LP, which is estimated from $\mathbf{y}$ as in (3.24). This method is found to be simple and robust in different performance evaluations [SK03, TV01]. It becomes computationally intensive in MIMO scenarios [SWS03].

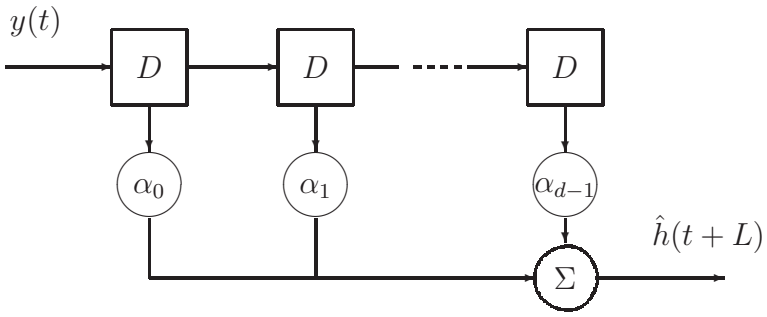

Figure 5.1: A linear channel predictor with order d.

Channel prediction based on deterministic sinusoidal modeling

Motivated by the Jakes model [Jak74], a radio channel is modeled as superimposed complex sinusoids as

$$\mathbf{y} = \mathbf{h} + \mathbf{e} = \mathbf{A}\mathbf{s} + \mathbf{e}, \tag{5.2}$$

where

$$\begin{aligned} \mathbf{A} &= [\mathbf{a}(\omega_1), \cdots, \mathbf{a}(\omega_p)], & (5.3)\\ \mathbf{a}(\omega_i) &= [e^{j\omega_i(t)}, e^{j\omega_i(t-1)}, \cdots, e^{j\omega_i(t-N+1)}]^T, & (5.4)\\ \mathbf{s} &= [s_1, \cdots, s_p]^T. & (5.5) \end{aligned}$$

Given $\boldsymbol{\omega}$, the LS estimate of $\mathbf{s}$ is

$$\hat{\mathbf{s}}_{LS} = \mathbf{A}^{\dagger}\mathbf{y}, \tag{5.6}$$

where $\mathbf{A}^{\dagger} = \left(\mathbf{A}^H\mathbf{A}\right)^{-1}\mathbf{A}^H$. In practice, the frequency vector is estimated and assume $\hat{\boldsymbol{\omega}}$ is unbiased, (5.6) can be written as

$$\hat{\mathbf{s}}_{LS} = \hat{\mathbf{A}}^{\dagger}\mathbf{y}. \tag{5.7}$$

In (5.7), $\hat{\mathbf{A}}$ might be ill-conditional due to closely separated frequency components. Such a problem can be solved by *regularization*.

Since the ill-conditional $\hat{\mathbf{A}}$ matrix leads to large $\|\mathbf{s}\|^2$, a penalty term could be added into the LS estimate

$$\hat{\mathbf{s}}_{Reg} = \arg\min_{\mathbf{s}} \{\|\mathbf{y} - \hat{\mathbf{A}}\mathbf{s}\|^2 + \alpha\|\mathbf{s}\|^2\}, \tag{5.8}$$

where $\alpha > 0$ is called the *regularization* parameter. Let

$$\frac{\partial\{\|\mathbf{y} - \hat{\mathbf{A}}\mathbf{s}\|^2 + \alpha\|\mathbf{s}\|^2\}}{\partial\mathbf{s}} = 0, \tag{5.9}$$

the regularized LS estimate of $\mathbf{s}$ is

$$\begin{aligned} \hat{\mathbf{s}}_{Reg} &= (\hat{\mathbf{A}}^H\hat{\mathbf{A}} + \alpha\mathbf{I}_p)^{-1}\hat{\mathbf{A}}^H\mathbf{y} && (5.10)\\ &= \mathbf{R}_{Reg}^{-1}\hat{\mathbf{A}}^H\mathbf{y}, && (5.11) \end{aligned}$$

where $\mathbf{R}_{Reg}$ is the regularized covariance matrix. Its eigenvalues are displaced from those of $(\hat{\mathbf{A}}^H\hat{\mathbf{A}})$ by the regularization parameter α. So the ill-conditional problem (zero eigenvalues) is alleviated [Che05].

Channel estimation and prediction method using Discrete Prolate Spheroidal Sequences

Recently, a new channel estimation and prediction method using Discrete Prolate Spheroidal Sequences [Sle78] was proposed in [ZM05], where a time-limited snapshot of bandlimited radio channel observations is expanded using Slepian basis as

$$h(t) \approx \sum_{i=0}^{D^{(S)}-1} u_i^{(S)}(t)\gamma_i^{(S)}, \tag{5.12}$$

where $u_i^{(S)}(t)$ is the Slepian sequence whose maximum energy concentration in an interval with length N,

$$D' < D^{(S)-1} < N, \tag{5.13}$$

where

$$D' = \lceil 2\omega_m N \rceil + 1, \tag{5.14}$$

where $\lceil\cdot\rceil$ is round towards plus infinity of a real number. In this method only the maximum Doppler frequency ω_m and the length of channel observation sequence N are required.

A thorough performance evaluation of these sinusoid model based methods and LP were made in [SK03], where both simulation data and real world data are used. It was reported that the LP outperforms the other methods in general. In [TV01], it was claimed that the channel is only predictable a few portions of wavelength into the future in a SISO channel. The performance bound of MIMO channel predictors based on sinusoidal modeling was derived in [SS03]. A longer range prediction revealing structures in a MIMO channel is possible. In [ZG04], a study of the required prediction accuracy was performed at link level.

5.2 Channel Prediction Based on Statistical Sinusoidal Modeling

A stationary sinusoidal model for a Rayleigh fading channel is given in (5.2). The unknown model parameters are $\boldsymbol{\theta} = [\boldsymbol{\omega}^T, \mathbf{s}^T]^T$. In these methods, the Doppler frequency, $\boldsymbol{\omega}$, is estimated using a high resolution algorithm, i.e. Unitary-ESPRIT [HN95], while the complex amplitude, $\mathbf{s}$, can be estimated using different methods [CEV07].

5.2.1 Conditional LMMSE Predictors

Given the unbiased frequency estimate $\hat{\boldsymbol{\omega}} = [\hat{\omega}_1, \cdots, \hat{\omega}_p]^T$, the LS estimate of the complex amplitudes is given in (5.7). Substituting (5.2) into (5.7),

$$\begin{aligned} \hat{\mathbf{s}}_{LS} &= \left(\hat{\mathbf{A}}^H \hat{\mathbf{A}}\right)^{-1} \hat{\mathbf{A}}^H \left(\hat{\mathbf{A}}\mathbf{s} + \mathbf{e}\right), \\ &= \mathbf{s} + \hat{\mathbf{A}}^{\dagger}\mathbf{e}, \end{aligned} \tag{5.15}$$

where $\hat{\mathbf{A}}^{\dagger}\mathbf{e}$ is the estimation error with covariance matrix $\sigma_e^2(\hat{\mathbf{A}}^H\hat{\mathbf{A}})^{-1}$. Either in simulations or measured channels, the doppler frequencies might be closely separated, which can be seen from the PDF of the Doppler frequency in Figure 2.4, where the probability in the region close to ± 1 in the normalized spectrum is much higher than other parts. Therefore, the *true* $\mathbf{A}$ matrix might be ill-conditioned. As a result, the noise will be amplified by the pseudo inverse matrix $\hat{\mathbf{A}}^{\dagger}$. Such a problem is solved by regularization as in (5.11). To determine the optimal regularization parameter α, the amplitude $\mathbf{s}$ is modeled as a Gaussian random vector with PDF, $\mathcal{CN}(\mathbf{0}_p, \delta_s^2\mathbf{I}_p)$, which is a *stochastic* sinusoidal model. The conditional LMMSE estimate of $\mathbf{s}$ is, i.e. [Kay93b],

$$\hat{\mathbf{s}}_{LMMSE} = \mathbf{R}_{sy}\mathbf{R}_{yy}^{-1}\mathbf{y}, \tag{5.16}$$

where

$$\mathbf{R}_{sy} = \mathbf{E}[\mathbf{s}\mathbf{y}^H] = \sigma_s^2\hat{\mathbf{A}}^H, \tag{5.17}$$

$$\mathbf{R}_{yy} = \mathbf{E}[\mathbf{y}\mathbf{y}^H] = \sigma_s^2\hat{\mathbf{A}}\hat{\mathbf{A}}^H + \sigma_e^2\mathbf{I}_N. \tag{5.18}$$

Therefore,

$$\hat{\mathbf{s}}_{LMMSE} = \hat{\mathbf{A}}^H \left(\hat{\mathbf{A}}\hat{\mathbf{A}}^H + \alpha\mathbf{I}_N\right)^{-1}\mathbf{y}, \tag{5.19}$$

$$= (\hat{\mathbf{A}}^H\hat{\mathbf{A}} + \alpha\mathbf{I}_p)^{-1}\hat{\mathbf{A}}^H\mathbf{y}, \tag{5.20}$$

$$= \mathbf{R}_{Reg}^{-1}\hat{\mathbf{A}}^H\mathbf{y}, \tag{5.21}$$

where $\mathbf{R}_{Reg}$ is the regularized covariance matrix as in (5.11), $\alpha = \sigma_e^2/\sigma_s^2$, which is the inverse of SNR, and the matrix inversion lemma[1] is used in the second equation. Then the conditional LMMSE predictor is

$$\hat{h}(t+L) = \hat{\mathbf{a}}^H(L)\hat{\mathbf{s}}_{LMMSE}, \tag{5.22}$$

where $\hat{\mathbf{a}}^H(L) = [e^{j\hat{\omega}_1(t+L)}, \cdots, e^{j\hat{\omega}_p(t+L)}]$. Note that for Gaussian signals, the conditional LMMSE predictor is also an MMSE predictor. The proof is given in Appendix C.

5.2.2 Unconditional LMMSE Predictors

Since the frequency estimates is subject to errors in practice, $\boldsymbol{\omega}$ can be modeled as a Gaussian vector with PDF, $\mathcal{N}(\hat{\boldsymbol{\omega}}, \sigma_{\Delta\omega}^2\mathbf{I}_p)$. The variance $\sigma_{\Delta\omega}^2$ is determined by the SNR and the number of samples [Kay93b]. Such a problem was investigated in [SS91]. The unconditional LMMSE predictor is

$$\hat{h}(t+L) = \mathbf{R}_{hy}\mathbf{R}_{yy}^{-1}\mathbf{y}, \tag{5.23}$$

where

$$\begin{aligned}
\mathbf{R}_{hy} &= \sigma_s^2\hat{\mathbf{a}}^H(L)\hat{\mathbf{A}}^H \odot \boldsymbol{\gamma}, && (5.24)\\
\boldsymbol{\gamma} &= \left[e^{\frac{-\sigma_{\Delta\omega}^2}{2}L^2}, \cdots, e^{\frac{-\sigma_{\Delta\omega}^2}{2}(L+N-1)^2}\right], && (5.25)\\
\mathbf{R}_{yy} &= \mathbf{R}_{hh} + \sigma_e^2\mathbf{I}_N, && (5.26)\\
\mathbf{R}_{hh} &= \sigma_s^2\hat{\mathbf{A}}\hat{\mathbf{A}}^H \odot \boldsymbol{\Gamma}, && (5.27)\\
[\boldsymbol{\Gamma}]_{mn} &= e^{-\frac{\sigma_{\Delta\omega}^2}{2}(n-m)^2}. && (5.28)
\end{aligned}$$

Note that the random frequencies are assumed to be independent on the amplitudes. The unconditional LMMSE predictor is

$$\begin{aligned}
\hat{h}(t+L) &= \left(\sigma_s^2\hat{\mathbf{a}}^H(L)\hat{\mathbf{A}}^H \odot \boldsymbol{\gamma}\right)\left(\sigma_s^2\hat{\mathbf{A}}\mathbf{A}^H \odot \boldsymbol{\Gamma} + \sigma_e^2\mathbf{I}_N\right)^{-1}\mathbf{y},\\
&= \left(\hat{\mathbf{a}}^H(L)\hat{\mathbf{A}}^H \odot \boldsymbol{\gamma}\right)\left(\hat{\mathbf{A}}\hat{\mathbf{A}}^H \odot \boldsymbol{\Gamma} + \alpha\mathbf{I}_N\right)^{-1}\mathbf{y}. && (5.29)
\end{aligned}$$

So the influence of old observations is reduced by the damping matrix $\boldsymbol{\Gamma}$ and the damping vector $\boldsymbol{\gamma}$. The damping vector $\boldsymbol{\gamma}$ is dependent on the prediction range as well. The way the frequency error is taken into account can be interpreted as a convolution of the line spectrum of the signal with the error distribution. The filter design is thus done for distributed sources.

[1] $(\mathbf{A}+\mathbf{BCD})^{-1}=\mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{B}(\mathbf{DA}^{-1}\mathbf{B}+\mathbf{C}^{-1})^{-1}\mathbf{DA}^{-1}$.

5.2.3 Adjusted Conditional LMMSE Predictors

Due to the limit on frequency resolution, the frequency estimate $\hat{\boldsymbol{\omega}}$ might be biased. This gives rise to a colored *residual* signal, which is

$$\boldsymbol{\epsilon} = \mathbf{y} - \hat{\mathbf{h}} = \mathbf{y} - \hat{\mathbf{A}}\hat{\mathbf{s}}, \tag{5.30}$$

where $\boldsymbol{\epsilon} = [\epsilon(t), \cdots, \epsilon(t-N+1)]^T$, and $\boldsymbol{\epsilon}$ contains both the channel estimation error $\mathbf{e}$ and the prediction error $\mathbf{h}-\hat{\mathbf{h}}$. This residual signal could be predicted by a low order LP predictor. The adjusted conditional LMMSE predictor is

$$\hat{h}_{adj}(t+L) = \hat{h}(t+L) + \hat{\epsilon}(t+L), \tag{5.31}$$

where

$$\hat{h}(t+L) = \hat{\mathbf{a}}^H(L)\hat{\mathbf{s}}, \tag{5.32}$$

$$\hat{\epsilon}(t+L) = \sum_{k=0}^{d_\epsilon-1} \alpha_{\epsilon,k}\epsilon(t-k) = \boldsymbol{\alpha}_\epsilon^T \boldsymbol{\epsilon}_d. \tag{5.33}$$

Then (5.31) can be rewritten as

$$\hat{h}_{adj}(t+L) = \begin{bmatrix} \boldsymbol{\epsilon}_d^T & \hat{\mathbf{a}}^H(L) \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_\epsilon \\ \mathbf{s} \end{bmatrix}. \tag{5.34}$$

5.2.4 JMAS Prediction Model and Joint LS Predictor

Motivated by the real data, a channel can be split into two parts, the periodic sinusoidal part and the non-periodic autoregressive part. A Joint Moving Average and Sinusoidal (JMAS) prediction model is

$$\hat{h}(t+L) = \mathbf{y}_d^T \boldsymbol{\alpha}_J + \hat{\mathbf{a}}^H(L)\mathbf{s} + e(t+L), \tag{5.35}$$

where $\hat{\boldsymbol{\omega}}$ contains only the stable frequencies from some model selection methods. Given $\hat{\boldsymbol{\omega}}$, (5.35) can be rewritten as

$$\hat{h}(t+L) = \begin{bmatrix} \mathbf{y}_d^T & \hat{\mathbf{a}}^H(L) \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_J \\ \mathbf{s} \end{bmatrix}, \tag{5.36}$$

where the parameters associated with the two parts can be estimated jointly, so that this predictor is termed Joint LS predictor. Note that the autoregressive basis and the sinusoidal basis in (5.36) are correlated, but not in (5.34).

5.3 Channel Prediction Based on Polynomial Phase Signals

The channel model based on MC-PPS is given in (2.29). To be able to predict $y(t)$, the time-varying amplitude, $s_i(t)$ is modeled as AR(d)-processes with nonzero mean, i.e.

$$s_{z,i}(t) = s_i(t) - \mu_{s,i}(t), \tag{5.37}$$

where $\mu_{s,i}(t)$ is the mean of $s_i(t)$, which should be constant, and $s_{z,i}(t)$ is the amplitude with zero mean. Further, let

$$\begin{aligned} s_{i,z}(t+1) &= \sum_{l=1}^{d} \alpha_{i,l} s_{i,z}(t-l+1) + v_i(t), & (5.38)\\ &= \boldsymbol{\alpha}_{i,d}^T \mathbf{s}_{i,z}(t) + v_i(t), & (5.39) \end{aligned}$$

where $v_i(t)$ is a white noise with non-zero mean and variance $\sigma_{v,i}^2$.

Then the signal model (2.29) can be expressed in a state-space structure as

$$\begin{aligned} \mathbf{x}(t+1) &= \boldsymbol{\Gamma}_\alpha \mathbf{x}(t) + \mathbf{u}(t), & (5.40)\\ y(t) &= \mathbf{c}^T(t)\mathbf{x}(t) + e(t), & (5.41) \end{aligned}$$

where

$$\begin{aligned} \mathbf{x}(t) &= [\mathbf{x}_1^T(t), \cdots, \mathbf{x}_p^T(t)]^T,\\ \mathbf{x}_i(t) &= [s_{z,i}(t), \cdots, s_{z,i}(t-d+1), \mu_{s,i}(t)]^T,\\ \boldsymbol{\Gamma}_\alpha &= \begin{bmatrix} \boldsymbol{\Gamma}_1 & & \\ & \ddots & \\ & & \boldsymbol{\Gamma}_p \end{bmatrix}, & (5.42)\\ \boldsymbol{\Gamma}_i &= \begin{bmatrix} \alpha_{i,1} & \alpha_{i,2} & \cdots & \alpha_{i,d-1} & \alpha_{i,d} & 0\\ 1 & 0 & \cdots & 0 & 0 & 0\\ & & \vdots & & & \\ 0 & 0 & \cdots & 1 & 0 & 0\\ 0 & 0 & \cdots & 0 & 0 & 1 \end{bmatrix},\\ \mathbf{u}(t) &= [\mathbf{u}_1^T(t), \cdots, \mathbf{u}_p^T(t)]^T,\\ \mathbf{u}_i(t) &= [v_i(t), \mathbf{0}_{d-1}^T, w_i(t)]^T,\\ \mathbf{c}(t) &= [\mathbf{c}_1^T(t), \cdots, \mathbf{c}_p^T(t)]^T,\\ \mathbf{c}_i(t) &= [e^{j\phi_i(t)}, \mathbf{0}_{d-1}^T, e^{j\phi_i(t)}]^T. \end{aligned}$$

The variance of $w_i(t)$, $\sigma^2_{w,i}$, should be set much smaller than the variance of $v_i(t)$, since $w_i(t)$ is the innovation noise for the mean amplitude which is constant. Then, the prediction of the $h(t+L)$ is

$$\hat{h}(t+L|t) = \mathbf{c}^T(t+L)\mathbf{\Gamma}^L_\alpha \hat{\mathbf{x}}(t|t), \tag{5.43}$$

where $\hat{\mathbf{x}}(t|t)$ is obtained by the Kalman filter [Kal60].

The initial covariance matrix of the state vector for the Kalman filter is chosen as

$$\mathbf{C}_\mathbf{x} = \begin{bmatrix} \mathbf{C}_{\mathbf{x},1} & & \\ & \ddots & \\ & & \mathbf{C}_{\mathbf{x},p} \end{bmatrix}, \tag{5.44}$$

$$\mathbf{C}_{\mathbf{x},i} = \begin{bmatrix} \sigma^2_{s,i}\mathbf{I}_d & \\ & 0 \end{bmatrix}. \tag{5.45}$$

The initial state vector is taken to be $\mathbf{x}_i(t-N+1|t-N+1) = [\mathbf{0}^T_d, \mu_i(t)]^T$. The covariance matrix of $\mathbf{u}(t)$ is selected as

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_1 & & \\ & \ddots & \\ & & \mathbf{Q}_p \end{bmatrix}, \tag{5.46}$$

$$\mathbf{Q}_i = \begin{bmatrix} \sigma^2_{v,i} & & & & \\ & 0 & & & \\ & & \ddots & & \\ & & & 0 & \\ & & & & \sigma^2_{w,i} \end{bmatrix}. \tag{5.47}$$

The instantaneous amplitude $s_i(k)$ can be estimated using (3.54), where $\mathbf{y}_1(k)$ is replaced by $\mathbf{y}(k)$. Then the AR parameters $\boldsymbol{\alpha}_{i,d}$ in (5.39) can be estimated using LS as:

$$\hat{\boldsymbol{\alpha}}_{i,d} = \arg\min_{\boldsymbol{\alpha}_{i,d}} \sum_{k=t}^{t-N+n+1} \|\hat{s}_{i,z}(k+1) - \boldsymbol{\alpha}^T_{i,d}\hat{\mathbf{s}}_{i,z}(k)\|^2, \tag{5.48}$$

where

$$\hat{s}_{i,z}(k) = \hat{s}_i(k) - \hat{\mu}_i(k), \tag{5.49}$$

$$\hat{\mu}_i(k) = \frac{1}{N-n+1} \sum_{k=t}^{t-N+n} \hat{s}_i(k). \tag{5.50}$$

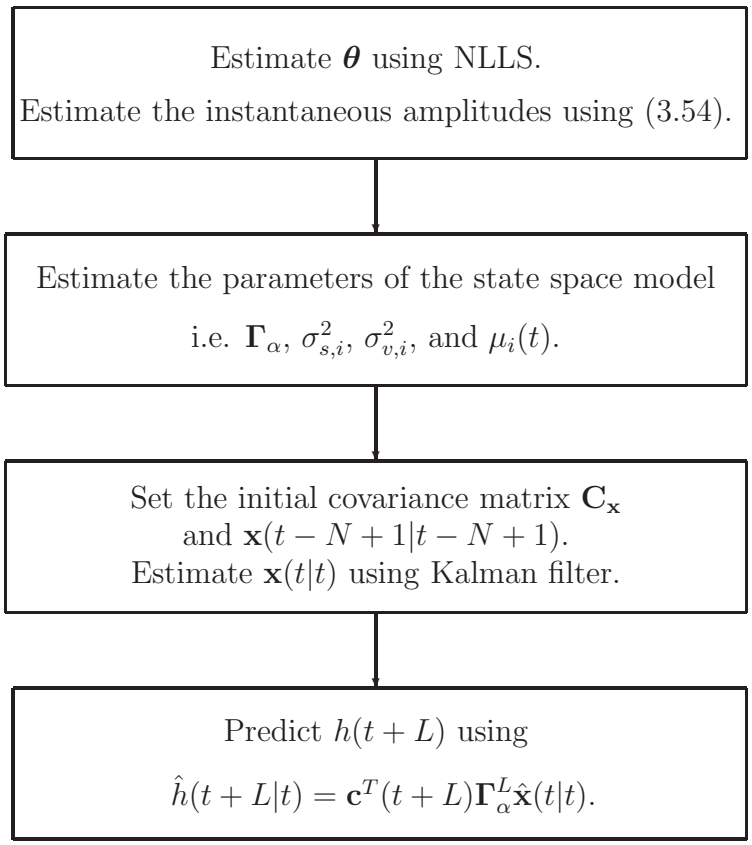

Figure 5.2: A block diagram of the adaptive channel predictor using the Kalman filter.

The variance of $s_{i,z}(t)$ is

$$\hat{\sigma}^2_{s,i} = \frac{1}{N-n+1} \sum_{k=t}^{t-N+n} \|\hat{s}_i(k) - \hat{\mu}_i(k)\|^2. \tag{5.51}$$

The variance of $v_i(t)$ is

$$\hat{\sigma}^2_{v,i} = \frac{1}{N-n} \sum_{k=t}^{t-N+n+1} \|\hat{s}_{i,z}(k+1) - \hat{\boldsymbol{\alpha}}^T_{i,d} \hat{\mathbf{s}}_{i,z}(k)\|^2. \tag{5.52}$$

A block diagram of the adaptive channel predictor is given in Figure 5.2.

Particularly, when the time-varying amplitude $s_i(t)$ is approximated by a constant, $\hat{s}_i$, which is reasonable when N is small, the channel prediction

based on the MC-PPS model is

$$\hat{h}(t+L) = \mathbf{a}_{\phi}^{T}(L)\hat{\mathbf{s}}, \tag{5.53}$$

where

$$\begin{aligned} \mathbf{a}_{\phi}(L) &= [e^{j\phi_1(t+L)}, \cdots, e^{j\phi_p(t+L)}]^T, & (5.54)\\ \hat{\mathbf{s}} &= [\hat{s}_1, \cdots, \hat{s}_p]^T. & (5.55) \end{aligned}$$

This predictor is termed LS MC-PPS prediction.

Chapter 6

Conclusions and Future Works

An extensive study on radio channel prediction based on parametric modeling is performed in this thesis.

6.1 Conclusions

Motivated by the Jakes model, a number of channel prediction methods based on stochastic sinusoidal modeling are proposed. Two LMMSE predictors are proposed, which are termed conditional LMMSE prediction and unconditional LMMSE prediction. In the conditional LMMSE predictors, the Doppler frequencies or frequency estimates are given and the amplitudes are modeled as Gaussian random variables. In the unconditional LMMSE prediction, both the amplitudes and frequency estimates are modeled as Gaussian random variables. In the conditional LMMSE predictor, the ill-conditioned LS estimate of the complex amplitudes is alleviated by *regularization*. While the uncertainty of the frequency estimate in the unconditional LMMSE prediction is interpreted as *distributed* sources in the frequency domain, or an exponential weighting of the channel observations in the time domain. The unconditional LMMSE predictor outperforms the conditional LMMSE predictor both in simulations and test using measured channels.

However, it was found that a part of the channels, both in simulations and measurements, cannot be described using pure periodic sinusoidal models. To pick-up this part of the channels, an adjusted conditional LMMSE prediction and JLS prediction are proposed using mixed sinusoidal bases and autoregressive bases, where the prediction coefficients are estimated separately and jointly respectively. These methods approach the compressed CRLB of channel prediction in simulations, and outperform the other two LMMSE

predictors in measured channels.

Based on a detailed modeling of the radio wave scattering on rough surfaces, a physics based scattering model is proposed to narrow band SISO channels. It is extended into MIMO scenarios with single reflection. Similar models are found in 3GPP standards.

Motivated by these physical channel models, a new category of channel prediction based on the non-stationary Multi-Component Polynomial Phase Signal with time-varying amplitudes is proposed. An adaptive channel predictor using the Kalman filter is proposed, where the time-varying amplitudes are modeled as AR(d) processes. Following the idea in SAGE/RELAX, an iterative parameter estimation method of MC-PPS is proposed with known model order and number of PPS components, where a Nonlinear *instantaneous* LS criterion can be used to alleviate the influence of the time-varying amplitudes on parameter estimation. Alternative to the NILS criterion, the parameter estimation of the polynomial phases using the standard NLLS criterion is studied as well, where a high resolution on parameter estimation is obtained.

An iterative detection and estimation method of MC-PPS is proposed. The detection of the number of signal components is based on examination of the size of the residual signals with known noise variance. The detection of the order of polynomial phase is based on GLRT or Wald test.

The channel predictors based on MC-PPS are evaluated using both synthetic data and measured channels. These new methods outperform the sinusoidal model based predictors and the classical LP in simulations. High order polynomial phase parameters are observed in both urban and suburban environments. The PPS based channel predictors outperform the other methods in examples of measured urban and suburban channels.

6.2 Discussions and Future works

In this thesis, the study of channel prediction methods based on parametric modeling focus on narrow band SISO channels or the channels in a single frequency bin in OFDM systems.

In a MIMO system, due to the close separation between antenna elements, the subchannels might share the same or highly correlated model parameters, which provide two potential benefits: the first one is the improvement of the accuracy of parameter estimation and channel prediction, and the second is the reduction on computational complexity of MIMO channel prediction. These potential advantages are definitely worth to investigate. It is also of interested to evaluate these models and methods with MIMO systems with

different polarized antenna elements.

In wide band systems or OFDM system, the channels most likely experience frequency selective fading. Typical channel estimation techniques in OFDM systems are allocating pilot symbols in a number of equally separated frequency bins. Channel prediction can be performed on these pilot bins. Then the frequency domain interpolation can be used to predict the non-pilot bins. Due to the similar wavelengths of the radio signals in different frequency bins, the channels might have highly correlated model parameters as well. The same advantages can be taken as in a MIMO system.

Regarding the measured channels, the study has been performed on the strongest tap in the channel impulse response. It is also interesting to have similar studies on the following taps. It would of course be beneficial to take into account the correlation (if any) that exists between the different taps. This gives a 2D structure, even in the SISO case. It might lead to very computationally costly algorithms in the SIMO and MIMO case, but it might improve performance.

An extensive performance evaluation of these parametric model based channel predictors using measured SISO and MIMO channels is of great interest.

In the implementation of these parametric modeling based prediction methods, a good parameter tracking scheme is important, which helps to reduce the calculation complexity and improve the prediction accuracy. This aspect should be studied in the future work.

Appendix A

Forward-Backward LS Estimate of LP Coefficients

The LS estimate of $\boldsymbol{\alpha}_d$ of LP in (3.24) is usually named a *forward* LS estimation. A *backward* LP is

$$\hat{h}(t-L) = \boldsymbol{\alpha}_{d,b}^T \mathbf{y}_{d,b},$$

where $\mathbf{y}_{d,b} = [y(t), \cdots, y(t+d-1)]^T$. In vector form,

$$\mathbf{y}_b = \mathbf{Y}_b \boldsymbol{\alpha}_{d,b}, \tag{A.1}$$

where

$$\mathbf{y}_b = [y(t-N+1), \cdots, y(t-L-d+1)]^T, \tag{A.2}$$

$$\mathbf{Y}_b = \begin{bmatrix} y(t-N+L+1) & y(t-N+L+2) & \cdots & y(L+d) \\ \vdots & \vdots & \vdots & \vdots \\ y(t-d+1) & y(t-d+2) & \cdots & y(t) \end{bmatrix}, \tag{A.3}$$

$$\boldsymbol{\alpha}_{d,b} = [\alpha_{0,b}, \cdots, \alpha_{d-1,b}]^T. \tag{A.4}$$

The backward LS estimate of $\boldsymbol{\alpha}_{d,b}$ is

$$\boldsymbol{\alpha}_{d,b} = (\mathbf{Y}_b^H \mathbf{Y}_b)^{-1} \mathbf{Y}_b^H \mathbf{y}_b.$$

It is easy to show that $\bar{\boldsymbol{\alpha}}_{d,b} \approx \boldsymbol{\alpha}_d$, where $\bar{\boldsymbol{\alpha}}_{d,b}$ is the conjugate of $\boldsymbol{\alpha}_{d,b}$ without transpose, so that the *forward-backward* LS estimate of $\boldsymbol{\alpha}_d$ can be obtained by solving

$$\begin{bmatrix} \mathbf{y}_{ls} \\ \bar{\mathbf{y}}_b \end{bmatrix} = \begin{bmatrix} \mathbf{Y}_{ls} \\ \bar{\mathbf{Y}}_b \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \vdots \\ \alpha_{d-1} \end{bmatrix}$$

$$\mathbf{y}_{fb} = \mathbf{Y}_{fb} \boldsymbol{\alpha}_d. \tag{A.5}$$

Then the *forward-backward* LS estimate of $\boldsymbol{\alpha}_d$ is

$$\hat{\boldsymbol{\alpha}}_d = (\mathbf{Y}_{fb}^H \mathbf{Y}_{fb})^{-1} \mathbf{Y}_{fb}^H \mathbf{y}_{fb}. \tag{A.6}$$

Appendix B

Instrumental Variable Method for Frequency Estimation in Colored Noise

The classical so-called eigenvector (EV) techniques, such as ESPRIT, MUSIC etc., can produce highly biased frequency estimates in colored noise or interference. An Instrumental Variable Method (IVM) for frequency estimation in colored noise is proposed in [SVO94, Gus98].

Assume the noise $\mathbf{e}$ in signal model (2.10) is colored with zero mean and covariance matrix $E[\mathbf{e}\mathbf{e}^H] = \mathbf{C}_e$, which is a positive definite matrix. Then the covariance matrix $\mathbf{R}_{yy}$ in (3.59) becomes

$$\mathbf{R}_{yy} = \mathbf{A}\mathbf{S}\mathbf{A}^H + \mathbf{Q}. \tag{B.1}$$

The signal subspace and the noise subspace cannot be separated by the eigenvalue decomposition as in (3.61). However if the colored noise $e(t)$ could be modelled as a q^{th} order Moving Average (MA) process, which is

$$e(t) = u(t) + b_1 u(t-1) + \cdots + b_q u(t-q), \tag{B.2}$$

where $u(t)$ is complex white Gaussian noise with zero mean and variance σ_u^2. The noise covariance matrix $\mathbf{Q} = E[\mathbf{e}\mathbf{e}^H]$, which is a banded Hermitian Toeplitz matrix with first row of $[q_0, q_1, \cdots, q_q, 0, \cdots, 0]^T$, where $q_k = E[e(t)e(t-k)^*]$, and $q_k = 0$, when $k > q$. Motivated by this property, the signal subspace can be separated from the noise subspace by using the IVM.

Define the data vector $\boldsymbol{\zeta}_y$ and instrumental vector $\boldsymbol{\psi}_y$ as

$$\boldsymbol{\zeta}_y = [y(\tau), y(\tau-1), \cdots, y(\tau-K+1)]^T, \tag{B.3}$$

$$\boldsymbol{\psi}_y = [y(\tau-M), y(\tau-M-1), \cdots, y(\tau-M-L+1)]^T, \tag{B.4}$$

where $L \geq K \geq q$, and M is the temporal displacement. The cross-correlation $\mathbf{R}_{\psi\zeta} = E[\boldsymbol{\psi}_y \boldsymbol{\zeta}_y^H]$ is

$$\mathbf{R}_{\psi\zeta} = \begin{bmatrix} r_{yy}(M) & r_{yy}(M-1) & \cdots & r_{yy}(M-K+1) \\ r_{yy}(M+1) & r_{yy}(M) & \cdots & r_{yy}(M-K+2) \\ \vdots & \vdots & \vdots & \vdots \\ r_{yy}(M+L-1) & r_{yy}(M+L-2) & \cdots & r_{yy}(M+L-K) \end{bmatrix}, \quad \text{(B.5)}$$

$$= \begin{bmatrix} r_{xx}(M) & r_{xx}(M-1) & \cdots & r_{xx}(M-K+1) \\ r_{xx}(M+1) & r_{xx}(M) & \cdots & r_{xx}(M-K+2) \\ \vdots & \vdots & \vdots & \vdots \\ r_{xx}(M+L-1) & r_{xx}(M+L-2) & \cdots & r_{xx}(M+L-K) \end{bmatrix}$$

$$+ \begin{bmatrix} r_{ee}(M) & r_{ee}(M-1) & \cdots & r_{ee}(M-K+1) \\ r_{ee}(M+1) & r_{ee}(M) & \cdots & r_{ee}(M-K+2) \\ \vdots & \vdots & \vdots & \vdots \\ r_{ee}(M+L-1) & r_{ee}(M+L-2) & \cdots & r_{ee}(M+L-K) \end{bmatrix}. \quad \text{(B.6)}$$

Since $r_{ee}(\tau) = 0$ for $\tau > q$. It can be seen that when $(M-K+1) > q$, the second term in the last equation in (B.6) becomes zeros. This means that the sinusoidal signal subspace can be obtained from SVD of $\mathbf{R}_{\psi\zeta}$ without interference of the colored noise.

In practice, assume $\mathbf{y} = [y(t), y(t-1), \cdots, y(t-N+1))]^T$ is observed, the estimate of $\mathbf{R}_{\zeta\psi}$ can be obtained as

$$\hat{\mathbf{R}}_{\zeta\psi} = \frac{1}{N}\mathbf{Z}^T\bar{\mathbf{Y}}, \quad \text{(B.7)}$$

$$\text{(B.8)}$$

where

$$\mathbf{Z} = \begin{bmatrix} y(t-N+L) & \cdots & y(t-N+1) \\ \vdots & \vdots & \vdots \\ y(t-M-1) & \cdots & y(t-M-L) \end{bmatrix}, \quad \text{(B.9)}$$

$$\mathbf{Y} = \begin{bmatrix} y(t-N+M+L) & \cdots & y(t-N+M+L-K+1) \\ \vdots & \vdots & \vdots \\ y(t-1) & \cdots & y(t-K) \end{bmatrix}. \quad \text{(B.10)}$$

Both $\mathbf{Z}$ and $\mathbf{Y}$ are Toeplitz matrices. It is worth to note that, with limited number of observations, the IVM frequency estimate produces unbiased

frequency estimates, but increased variance due to reduced number of effective number of observations. The minimum variance is obtained when $M = K + q - 1$.

Appendix C

Equivalence of LMMSE and MMSE Predictors

For the given signal model in (5.2), the conditional LMMSE predictor of $h(t+L)$ is given in (5.22). Both the signal model and the predictor are represented here for convenience.

$$\mathbf{y} = \mathbf{A}\mathbf{s} + \mathbf{e}, \tag{C.1}$$

and

$$\hat{h}(t+L) = \hat{\mathbf{a}}(L)^H \mathbf{R}_{Reg}^{-1} \hat{\mathbf{A}}^H \mathbf{y}. \tag{C.2}$$

In this appendix, the Conditioned LMMSE predictor in (C.2) is shown to be equivalent to the MMSE predictor given observations $\mathbf{y}$ and frequency estimates $\hat{\boldsymbol{\omega}}$.

Assume that the amplitude vector $\mathbf{s}$ is a Gaussian random vector with PDF, $\mathcal{CN}(\mathbf{0}, \sigma_s^2 \mathbf{I}_p)$. The MMSE prediction of $h(t+L)$ given $\mathbf{y}$ and $\hat{\boldsymbol{\omega}}$ is

$$\begin{aligned} \hat{h}(t+L) &= E[h(t+L)|\mathbf{y}, \hat{\boldsymbol{\omega}}], \\ &= E\left[\hat{\mathbf{a}}(L)^H \mathbf{s}|\mathbf{y}, \hat{\boldsymbol{\omega}}\right], \\ &= \hat{\mathbf{a}}(L)^H E\left[\mathbf{s}|\mathbf{y}, \hat{\boldsymbol{\omega}}\right], \end{aligned} \tag{C.3}$$

where $E[\mathbf{s}|\mathbf{y}, \hat{\boldsymbol{\omega}}]$ is the MMSE estimate of $\mathbf{s}$ given $\mathbf{y}$ and $\hat{\boldsymbol{\omega}}$. The Bayes's rule is used to find the PDF of $\mathbf{s}$ conditioned on $\mathbf{y}$ and $\hat{\boldsymbol{\omega}}$, $f(\mathbf{s}|\mathbf{y}, \hat{\boldsymbol{\omega}})$. We write

$$f(\mathbf{s}|\mathbf{y},\hat{\boldsymbol{\omega}}) = \frac{f(\mathbf{s},\mathbf{y}|\hat{\boldsymbol{\omega}})}{f(\mathbf{y}|\hat{\boldsymbol{\omega}})} = \frac{f(\mathbf{y}|\mathbf{s},\hat{\boldsymbol{\omega}})f(\mathbf{s})}{f(\mathbf{y}|\hat{\boldsymbol{\omega}})}. \tag{C.4}$$

The numerator in (C.4) can be expressed as

$$f(\mathbf{y}|\mathbf{s},\hat{\boldsymbol{\omega}})f(\mathbf{s}) = \frac{1}{\pi^N\sigma_e^{2N}} e^{\frac{-(\mathbf{y}-\hat{\mathbf{A}}\mathbf{s})^H(\mathbf{y}-\hat{\mathbf{A}}\mathbf{s})}{\sigma_e^2}} \cdot \frac{1}{\pi^p\sigma_s^{2p}} e^{-\frac{\mathbf{s}^H\mathbf{s}}{\sigma_s^2}}. \tag{C.5}$$

Define $\mathbf{R}_{Reg} = \left(\hat{\mathbf{A}}^H\hat{\mathbf{A}} + \frac{\sigma_e^2}{\sigma_s^2}\mathbf{I}_p\right)$ as in (5.21) and complete the square, (C.5) can be written as

$$\begin{aligned} f(\mathbf{y}|\mathbf{s},\hat{\boldsymbol{\omega}})f(\mathbf{s}) &= \frac{1}{\pi^p|\sigma_e^2\mathbf{R}_{Reg}^{-1}|} e^{-(\mathbf{s}-\mathbf{R}_{Reg}^{-1}\hat{\mathbf{A}}^H\mathbf{y})^H(\sigma_e^2\mathbf{R}_{Reg}^{-1})^{-1}(\mathbf{s}-\mathbf{R}_{Reg}^{-1}\hat{\mathbf{A}}^H\mathbf{y})} \\ &\quad \cdot|\alpha\mathbf{R}_{Reg}^{-1}| \frac{1}{\pi^N\sigma_e^{2N}} e^{\frac{-(\mathbf{y}^H\mathbf{y}-\mathbf{y}^H\hat{\mathbf{A}}^H\mathbf{R}_{Reg}^{-1}\hat{\mathbf{A}}\mathbf{y})}{\sigma_e^2}}, \end{aligned} \tag{C.6}$$

where $|\mathbf{R}_{Reg}|$ is the determinant of $\mathbf{R}_{Reg}$. In (C.6), the first term is the PDF of a complex Gaussian multivariate $\mathbf{s}$ with PDF of $\mathcal{CN}(\mathbf{R}_{Reg}^{-1}\hat{\mathbf{A}}^H\mathbf{y},\sigma_e^2\mathbf{R}_{Reg}^{-1})$, while the second term is independent on $\mathbf{s}$.

The denominator in (C.4) is

$$f(\mathbf{y}|\hat{\boldsymbol{\omega}}) = \int f(\mathbf{y}|\mathbf{s},\hat{\boldsymbol{\omega}})f(\mathbf{s})d\mathbf{s}, \tag{C.7}$$

where the integrand is the same as the numerator in (C.4). So we have

$$\begin{aligned} f(\mathbf{y}|\hat{\boldsymbol{\omega}}) &= \int \frac{1}{\pi^p|\sigma_e^2\mathbf{R}_{Reg}^{-1}|} e^{-(\mathbf{s}-\mathbf{R}_{Reg}^{-1}\hat{\mathbf{A}}^H\mathbf{y})^H(\sigma_e^2\mathbf{R}_{Reg}^{-1})^{-1}(\mathbf{s}-\mathbf{R}_{Reg}^{-1}\hat{\mathbf{A}}^H\mathbf{y})} d\mathbf{s} \\ &\quad \cdot|\alpha\mathbf{R}_{Reg}^{-1}| \frac{1}{\pi^N\sigma_e^{2N}} e^{\frac{-(\mathbf{y}^H\mathbf{y}-\mathbf{y}^H\hat{\mathbf{A}}^H\mathbf{R}_{Reg}^{-1}\hat{\mathbf{A}}\mathbf{y})}{\sigma_e^2}}, \end{aligned} \tag{C.8}$$

where the first term integrates to one. Finally,

$$f(\mathbf{y}|\hat{\boldsymbol{\omega}}) = |\alpha\mathbf{R}_{Reg}^{-1}| \frac{1}{\pi^N\sigma_e^{2N}} e^{\frac{-(\mathbf{y}^H\mathbf{y}-\mathbf{y}^H\hat{\mathbf{A}}^H\mathbf{R}_{Reg}^{-1}\hat{\mathbf{A}}\mathbf{y})}{\sigma_e^2}}. \tag{C.9}$$

Not surprisingly, this is exactly the second term in (C.6). Substituting (C.6) and (C.9) into (C.4), we get the complex Gaussian PDF as $\mathcal{CN}(\mathbf{R}_{Reg}^{-1}\hat{\mathbf{A}}^H\mathbf{y},\sigma_e^2\mathbf{R}_{Reg}^{-1})$, which is given below.

$$f(\mathbf{s}|\mathbf{y},\hat{\boldsymbol{\omega}}) = \frac{1}{\pi^p|\sigma_e^2\mathbf{R}_{Reg}^{-1}|}e^{-(\mathbf{s}-\mathbf{R}_{Reg}^{-1}\hat{\mathbf{A}}^H\mathbf{y})^H(\sigma_e^2\mathbf{R}_{Reg}^{-1})^{-1}(\mathbf{s}-\mathbf{R}_{Reg}^{-1}\hat{\mathbf{A}}^H\mathbf{y})}. \tag{C.10}$$

Then, the conditional expectation of $\mathbf{s}$ given $\mathbf{y}$ and $\hat{\boldsymbol{\omega}}$ is

$$E\left[\mathbf{s}|\mathbf{y},\hat{\boldsymbol{\omega}}\right] = \mathbf{R}_{Reg}^{-1}\hat{\mathbf{A}}^H\mathbf{y}. \tag{C.11}$$

This is the same as the LMMSE estimate of $\mathbf{s}$ as in (5.21). Furthermore, the MMSE prediction of $h(t+L)$ in (C.3) is

$$E[h(t+L)|\mathbf{y},\hat{\boldsymbol{\omega}}] \;=\; \hat{\mathbf{a}}(L)^H\mathbf{R}_{Reg}^{-1}\hat{\mathbf{A}}^H\mathbf{y}, \tag{C.12}$$

which is identical to the LMMSE prediction (C.2). The MSE of (C.12) is obtained as

$$Var(h(t+L)|\mathbf{y},\hat{\boldsymbol{\omega}}) = \sigma_e^2\hat{\mathbf{a}}(L)^H\mathbf{R}_{Reg}^{-1}\hat{\mathbf{a}}(L). \tag{C.13}$$

Appendix D

CRLB for PPS Parameter Estimation

The CRLB of the model parameter estimation for a PPS signal with a constant amplitude is derived in this appendix. Note that a similar derivation of the bound can be found in [PP91b].

A single PPS buried in noise is

$$y(t) = se^{j\phi(t)} + e(t), \tag{D.1}$$

where s is a *real* amplitude, $e(t)$ is additive Gaussian noise with PDF, $\mathcal{CN}(0, \sigma_e^2)$, $\phi(t)$ is the polynomial phase

$$\phi(t) = \sum_{m=1}^{M} \beta_m t^m. \tag{D.2}$$

The model parameters are collected in

$$\boldsymbol{\theta} = [\beta_M, \beta_{M-1}, \cdots, \beta_1, s, \sigma_e^2]. \tag{D.3}$$

The N observed signals are collected in $\mathbf{y} = [y(0), y(1), \cdots, y(N-1)]^T$, and the noise vector $\mathbf{e} = [e(0), e(1), \cdots, e(N-1)]^T$.

Let $\mathrm{Re}\,[\mathbf{y}]$ and $\mathrm{Im}\,[\mathbf{y}]$ be the real part and the imaginary part of $\mathbf{y}$, and

$$\tilde{\mathbf{y}} = \begin{bmatrix} \mathrm{Re}\,[\mathbf{y}] \\ \mathrm{Im}\,[\mathbf{y}] \end{bmatrix} = \begin{bmatrix} s\cos(\boldsymbol{\phi}) \\ s\sin(\boldsymbol{\phi}) \end{bmatrix} + \begin{bmatrix} \mathrm{Re}\,[\mathbf{e}] \\ \mathrm{Im}\,[\mathbf{e}] \end{bmatrix}, \tag{D.4}$$

$$= \tilde{\mathbf{h}} + \tilde{\mathbf{e}}, \tag{D.5}$$

where $\boldsymbol{\phi} = [\phi(0), \cdots, \phi(N-1)]^T$, and the PDF of $\tilde{\mathbf{e}}$ is $\mathcal{N}(\mathbf{0}_{2N}, \frac{\sigma_e^2}{2}\mathbf{I}_{2N})$. So that the mean and covariance matrix of $\tilde{\mathbf{y}}$ are

$$\boldsymbol{\mu}_y = \begin{bmatrix} s\cos(\boldsymbol{\phi}) \\ s\sin(\boldsymbol{\phi}) \end{bmatrix}, \tag{D.6}$$

$$\mathbf{C}_y = \frac{\sigma_e^2}{2}\mathbf{I}_{2N}. \tag{D.7}$$

The CRLB is, i.e. [Kay93b],

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} \geq \mathbf{F}^{-1}(\boldsymbol{\theta}), \tag{D.8}$$

where $\mathbf{F}$ is the fisher information matrix, which is

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}_{\beta\beta} & \mathbf{F}_{\beta s} & \mathbf{F}_{\sigma\sigma} \\ \mathbf{F}_{s\beta} & \mathbf{F}_{ss} & \mathbf{F}_{s\sigma} \\ \mathbf{F}_{\sigma\beta} & \mathbf{F}_{\sigma s} & \mathbf{F}_{\sigma\sigma} \end{bmatrix}, \tag{D.9}$$

where

$$[\mathbf{F}_{\beta,\beta}]_{kl} = \frac{2s^2}{\sigma_e^2} \sum_{t=0}^{N-1} t^{(M-k+1)+(M-l+1)}, \tag{D.10}$$

$$[\mathbf{F}_{ss}] = \frac{2N}{\sigma_e^2}, \tag{D.11}$$

$$[\mathbf{F}_{\sigma\sigma}] = \frac{N}{2\sigma_e^4}, \tag{D.12}$$

and $[\mathbf{F}_{\beta s}] = [\mathbf{F}_{\beta\sigma}] = \mathbf{0}_M$, $[\mathbf{F}_{s\sigma}] = 0$. Note that $[\mathbf{F}_{\beta s}] = [\mathbf{F}_{s\beta}]^T$, and $[\mathbf{F}_{\beta\sigma}] = [\mathbf{F}_{\sigma\beta}]^T$.

Bibliography

[3GP03] 3GPP. *Spatial channel model for multiple input multiple output (MIMO) simulations*. Number 25.996 V6.1.0. Technical Report, Chichester, UK, Sept. 2003.

[AJHF99] J. B. Andersen, J. Jensen, S. Holdt, and F. Frederiksen. Prediction of future fading based on past measurements. In *Proc. of IEEE VTC Fall*, volume 1, Amsterdam, Netherlands, 1999.

[AK02] A. Abdi and M. Kaveh. A space-time correlation model for multielement antenna systems in mobile fading channels. *IEEE Transactions on Selected Areas in Communications*, 20:550–560, Apr. 2002.

[Aka74] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716–723, Dec. 1974.

[Ala98] S.M. Alamouti. A simple transmit diversity technique for wireless communications. *IEEE Journal on Selected Areas in Communications*, 16:14511458, Oct. 1998.

[BW88] D.M. Bates and D.G. Watts. *Nonlinear Regression and Its Applications*. John Wiley and Sons, New York, 1988.

[CA01] M. Chen and H. Asplund. Measurements and models for direction of arrival of radio waves in los in urban microcells. In *Proc. of the 12th IEEE International Symposium of Personal, Indoor and Mobile Radio Communications*, San Diego, USA, Sept 2001.

[CB05] P. J. Chung and J. F. Böhme. Recursive em and sage-inspired algorithms with application to doa estimation. *IEEE Trans. on Signal Processing*, 53:2664–2676, Aug. 2005.

[CEGJ02] S. Catreux, V. Erceg, D. Gesbert, and R.W. Heath Jr. Adaptive modulation and mimo coding for broadband wireless data networks. *IEEE Communications Magazine*, 40:108–115, Jun. 2002.

[CEV05] M. Chen, T. Ekman, and M. Viberg. Two new approaches for channel prediction based on sinusoidal modeling. In *Proc. of IEEE Workshop on Statistical Signal Processing*, Bordeaux, France, Jul. 2005.

[CEV07] M. Chen, T. Ekman, and M. Viberg. New approaches for channel prediction based on sinusoidal modeling. *EURASIP Journal on Advances in Signal Processing*, 2007:Article ID 49393, 13 pages, 2007. doi:10.1155/2007/49393.

[CF05] M. Chen and S. Felter. Feasibility study of channel prediction based on sinusoidal modeling with time variant model parameters. *Technical Report*, 2005.

[CG01] S.T. Chung and A.J. Goldsmith. Degrees of freedom in adaptive modulation: a unified view. *IEEE Transactions on Communications*, 49:1561–1571, Sept. 2001.

[Che02] M. Chen. *Mobile Positioning in Distributed Antenna Systems.* Patent application, USA, Dec 2002.

[Che05] M. Chen. *Channel Prediction Based on Sinusoidal Modeling.* Göteborg, Sweden, 2005.

[CKK02] M. Chen, H. Koorapaty, and A. Kangas. Enhanced positioning method in cellular systems. In *Proc. of the International Conference of Telecommunications*, Beijing, China, Jun. 2002.

[CM91] C. Curlander and R. N. McDonough. *Synthetic Aperture Radar-Systems and Signal Processing.* John Wiley and Sons, New York, 1991.

[Cor01] L.M. Correia. *Wireless Flexible Personalised Communications.* John Wiley and Sons Inc., 2001.

[CV04] M. Chen and M. Viberg. Lmmse channel prediction based on sinusoidal modeling. In *Proc. of 3rd IEEE Sensor Array and Multichannel signal Processing Workshop*, Barcelona, Spain, Jul. 2004.

[CV07] M. Chen and M. Viberg. Long range channel prediction based on non-stationary parametric modeling. *submitted to IEEE Transactions on Signal Processing*, Sept. 2007.

[CVF07a] M. Chen, M. Viberg, and S. Felter. Adaptive channel prediction based on polynomial phase signals. *submitted to IEEE ICASSP 2008*, Sept. 2007.

[CVF07b] M. Chen, M. Viberg, and S. Felter. Models and predictions of scattered radio waves on rough surfaces. In *Proc. of 32th International Conference of Acoustic, Speech, and Signal Processing*, Honolulu, USA, April 2007.

[DHHH00] A. Duel-Hallen, S. Hu, and H. Hallen. Long range prediction of fading signals: Enabling adaptive transmission for mobile radio channels. *IEEE Signal Processing Magazine*, 17(3):62–75, May 2000.

[DK90] P.M. Djuric and S.M. Kay. Parameter estimation of chirp signals. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 38:2118–2126, Dec. 1990.

[ECS+98] R.B. Ertel, P. Cardieri, K.W. Sowerby, T.S. Rappaport, and J.H. Reed. Overview of spatial channel models for antenna array communication systems. *IEEE Personal Communications*, 5(1):10–22, Feb. 1998.

[EK99] T. Ekman and G. Kubin. Nonlinear prediction of mobile radio channels: Measurements and mars model designs. In *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Phoenix, AZ, Mar. 1999.

[Ekm02] T. Ekman. *Prediction of Mobile Radio Channels, Modeling and Design*. Uppsala, Sweden, 2002.

[FCV06] S. Felter, M. Chen, and M. Viberg. *Method and Arrangement for Channel Prediction*. Number P22729US1. Patent application, USA, Sept 2006.

[FF95] B. Friedlander and J.M. Francos. Estimation of amplitude and phase parameters of multicomponent signals. *IEEE Transactions on Signal Processing*, 43:917–926, Apr. 1995.

[FH94] J.A. Fessler and A.O. Hero. Space-alternating generalized expectation-maximization algorithm. *IEEE Trans. on Signal Processing*, 42:2664 – 2677, Oct. 1994.

[FMB98] J. Fuhl, A.F. Molisch, and E. Bonek. Unifiled channel model for mobile radio systems with smart antennas. *IEE Proc.-Radar, Sonar Navig.*, 145(1):32–41, Feb. 1998.

[FSES04] S. Falahati, A. Svensson, T. Ekman, and M. Sternad. Adaptive modulation systems for predicted wireless channels. *IEEE Transactions on Communications*, 52:307– 316, Feb. 2004.

[GBGP02] D. Gesbert, H. Bölcskei, D. Gore, and A. Paulraj. Outdoor mimo wireless channels: Models and performance prediction. *IEEE Transactions on Communications*, 50(12):1926–1934, Dec. 2002.

[GC98] A.J. Goldsmith and S.G. Chua. Adaptive coded modulation for fading channels. *IEEE Transactions on Communications*, 46:595–602, May. 1998.

[GNS99] M. Ghogho, A.K. Nandi, and A. Swami. Cramer-rao bounds and maximum likelihood estimation for randomamplitude phase-modulated signals. *IEEE Transactions on Signal Processing*, 47:2905–2916, Nov. 1999.

[Gus98] T. Gustafsson. Instrumental variable subspace tracking using projection approximation. *IEEE Transactions on Signal Processing*, 46:669–681, Mar. 1998.

[HJ85] R.A. Horn and C.R. Johnson. *Matrix Analysis*. Cambridge University Press, UK, 1985.

[HN95] M. Haardt and J.A. Nossek. Unitary esprit: How to obtain increased estimation accuracy with a reduced computational burden. *Proc. IEEE Trans. on Signal Processing*, 43(5):1232–1242, May 1995.

[HW98] J.K. Hwang and J.H. Winters. Sinusoidal modeling and prediction of fast fading processes. In *Proc. of IEEE GLOBECOM*, Sydney, Australia, Nov. 1998.

[IAMH97] M. Z. Ikram, K. Abed-Meraim, and Y. Hua. Iterative parameter estimation of multiple chirp signals. *IEEE Electronic Letters*, 33(8), April 1997.

[JAG99] M. Viberg J. Ängeby and T. Gustafsson. Non-linear instantaneous least squares and its high snr analysis. In *Proc. of IEEE ICASSP 99*, volume 3, 1999.

[Jak74] W.C. Jakes. *Microwave Mobile Communications*. IEEE Press, Piscataway, NY, USA, 1974.

[Kal60] R.E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(D):35–45, 1960.

[Kay93a] S.M. Kay. *Fundamentals of Statistical Signal Processing - Detection Theory*. Prentice-Hall, Englewood Cliffs, NJ, 1993.

[Kay93b] S.M. Kay. *Fundamentals of Statistical Signal Processing - Estimation Theory*. Prentice-Hall, Englewood Cliffs, NJ, 1993.

[KH00] T. Keller and L. Hanzo. Adaptive modulation techniques for duplex ofdm transmission. *IEEE Transactions on Vehicular Technology*, 49:1893–1906, Sept. 2000.

[Kun78] S. Kung. A new identification and model reduction algorithm via singular value decomposition. In *Proc. of 12th Asilomar Conference on Circuits, Systems and Computers*, Pacific Grove, CA, 1978.

[KV96] H. Krim and M. Viberg. Two decades of array signal processing research-the parametric approach. *IEEE Signal Processing Magazine*, pages 67–94, Jul. 1996.

[LA92] R.M. Liang and K.S. Arun. Parameter estimation for superimposed chirp signals. In *Proc. IEEE International Conference Acoustics, Speech, and Signal Processing*, San Francisco, CA, USA, Mar. 1992.

[LS96] J. Li and P. Stoica. Efficient mixed-spectrum estimation with applications to targetfeature extraction. *IEEE Trans. on Signal Processing*, 44:281–295, Feb. 1996.

[NM65] J.A. Nelder and R. Mead. A simplex method for function minimization. *Comput. J.*, pages 308–313, Jul. 1965.

[PF95] S. Peleg and B. Friedlander. The discrete polynomial-phase transform. *IEEE Transactions on Signal Processing*, 43:1901–1914, Aug. 1995.

[PF96] S. Peleg and B. Friedlander. Multicomponent signal analysis using the polynomial-phase transform. *IEEE Transactions on Aerospace and Electronic Systems*, 32:378–387, Jan. 1996.

[PP91a] S. Peleg and B. Porat. The cramer-rao lower bound for signals with constant amplitude and polynomial phase. *IEEE Transactions on Signal Processing*, 39:749–752, Mar. 1991.

[PP91b] S. Peleg and B. Porat. The cramer-rao lower bound for signals with constant amplitude andpolynomial phase. *IEEE Transactions on Signal Processing*, 39:749–752, Mar. 1991.

[PRK85] A. Paulraj, R. Roy, and T. Kailath. Estimation of signal parameters via rotational invariance techniques - ESPRIT. In *Proc. of 19th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Nov. 1985.

[PZ07] D.S. Pham and A.M. Zoubir. Analysis of multicomponent polynomial phase signals. *IEEE Transactions on Signal Processing*, 55(1):56–65, Jan. 2007.

[QC99] X. Qiu and K. Chawla. On the performance of adaptive modulation in cellular systems. *IEEE Transactions on Communications*, 47:884–895, Jun. 1999.

[Rap96] T.S. Rappaport. *Wireless Communications Principles and Practice*. Prentice-Hall Inc., 1996.

[Ris83] J. Rissanen. A universal prior for integers and estimation by minimum description length. *The Annals of Statistics*, 11(2):416–431, Jun. 1983.

[RPK86] R. Roy, A. Paulraj, and T. Kailath. ESPRIT - a subspace rotation approach to estimation of parameters of cisoids in noise. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 34(4):1340–1342, Oct. 1986.

[Sau99] S. R. Saunders. *Antennas and Propagation for Wireless Communication Systems*. Wiley John and Sons, Inc., Philadelphia, PA, June 1999.

[Sch81] R.O. Schmidt. *A Signal Subspace Approach to Multiple Emitter Location and Spectral Estimation*. Nov. 1981.

[SEA01] M. Sternad, T. Ekman, and A. Ahlén. Power prediction on broadband channels. In *Proc. of VTC Fall*, Rhodes, Greece, May 2001.

[SFGK00] D.S. Shiu, G.J. Foschini, M.J. Gans, and J.M. Kahn. Fading correlation and its effect on the capacity of multielementantenna systems. *IEEE Transactions on Communications*, 48(3):502–513, Mar. 2000.

[Sha48] C.E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27:379–423, Jul. 1948.

[SK03] S. Semmelrodt and R. Kattenbach. Performance analysis and comparison of different fading forecast schemes for flat fading radio channels. In *Proc. of COST 273 TD(03)045*, 2003.

[Sle78] D. Slepian. Prolate spheroidal wave functions, fourier analysis, and uncertaintyv: The discrete case. *Bell Syst. Tech. J.*, 57(5):13711430, May-Jun 1978.

[SS91] P. Stoica and T. Söderström. Statistical analysis of MUSIC and subspace rotation estimates of sinusoidal frequencies. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 39(8):1836–1847, Aug. 1991.

[SS03] T. Svantesson and A. Lee Swindlehurst. A performance bound for prediction of mimo channels. *IEEE transactions on signal processing*, 1:233– 237, Nov. 2003.

[Sva01] T. Svantesson. A physical mimo radio channel model for multielement multi-polarized antenna systems. In *Proc. IEEE Vehicular Technology Conference, Fall*, Atlantic City, NJ, USA, Oct. 2001.

[SVO94] P. Stoica, M. Viberg, and B. Ottersten. Instrumental variable approach to array processing in spatially correlated noise fields. *IEEE Transactions on Signal Processing*, 42(3):121–133, Jan. 1994.

[SWS03] T. Svantesson, J. W. Wallace, and S. Semmelrodt. Performance evaluation of MIMO channel prediction algorithms using measurements. In *Proc. of the 13th IFAC Symposium on System Identification*, Rotterdam, The Netherlands, Aug. 2003.

[TSC98] V. Tarokh, N. Seshadri, and A. R. Calderbank. Spacetime codes for high data rate wireless communication: Performance analysis and code construction. *IEEE Transactions on Information Theory*, 44:744765, Mar. 1998.

[TSD04] L. Tong, B.M. Sadler, and M. Dong. Pilot-assisted wireless transmissions: general model, design criteria, and signal processing. *IEEE Signal Processing Magazine*, 21:12– 25, Nov. 2004.

[Tse01] D.N.C. Tse. Multiuser diversity in wireless networks. In *Wireless communications semenar*, Stanford University, CA, 2001.

[TV01] P.D. Teal and R.G. Vaughan. Simulation and performance bounds for real-time prediction of the mobile multipath channel. In *Proc. of the 11th IEEE Signal Processing Workshop on Statistical Signal Processing*, Singapore, Aug. 2001.

[WJ01] J.W. Wallace and M.A. Jensen. Statistical characteristics of measured mimo wireless channel data and comparison to conventional models. In *Proc. of IEEE Vehicular Technology Conference, 2001. Fall.*, Atlantic City, NJ, USA, Oct. 2001.

[YO02] K. Yu and B. Ottersten. Models for MIMO propagation channels, a review. *Special Issue on Adaptive Antennas and MIMO Systems, Wiley Journal on Wireless Communications and Mobile Computing*, 2002.

[ZG04] S. Zhou and G. B. Giannakis. How accurate channel prediction needs to be for transmit-beamforming with adaptive modulation over rayleigh mimo channels? *IEEE Transactions on Wireless Communications*, 3:1285– 1294, Jul. 2004.

[ZGS96] G.T. Zhou, G.B. Giannakis, and A. Swami. On polynomial phase signals with time-varying amplitudes. *IEEE Transactions on Signal Processing*, 44(7):848–861, Apr. 1996.

[ZM05] T. Zemen and C.F. Mecklenbrauker. Time-variant channel estimation using discrete prolate spheroidal sequences. *IEEE Transactions on Signal Processing*, 53:3597– 3607, Sept. 2005.