APPLICATION OF F-TEST METHOD ON MODEL ORDER SELECTION
AND RELATED PROBLEMS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ALPER YAZAR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
ELECTRICAL AND ELECTRONICS ENG.

AUGUST 2015

Approval of the thesis:

APPLICATION OF F-TEST METHOD ON MODEL ORDER SELECTION AND RELATED PROBLEMS

submitted by **ALPER YAZAR** in partial fulfillment of the requirements for the degree of **Master of Science in Electrical and Electronics Eng. Department, Middle East Technical University** by,

Prof. Dr. Gülbin Dural Ünver
Dean, Graduate School of **Natural and Applied Sciences**   ————

Prof. Dr. Gönül Turhan Sayan
Head of Department, **Electrical and Electronics Eng.**   ————

Assoc. Prof. Dr. Çağatay Candan
Supervisor, **Electrical and Electronics Eng. Dept., METU**   ————

Examining Committee Members:

Prof. Dr. Tolga Çiloğlu
Electrical and Electronics Engineering Dept., METU   ————

Assoc. Prof. Dr. Çağatay Candan
Electrical and Electronics Engineering Dept., METU   ————

Assoc. Prof. Dr. Sinan Gezici
Electrical and Electronics Engineering Dept., Bilkent Univ.   ————

Assoc. Prof. Dr. Umut Orguner
Electrical and Electronics Engineering Dept., METU   ————

Assist. Prof. Dr. Sevinç Figen Öktem
Electrical and Electronics Engineering Dept., METU   ————

Date: ————

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: ALPER YAZAR

Signature   :

ABSTRACT

APPLICATION OF F-TEST METHOD ON MODEL ORDER SELECTION AND RELATED PROBLEMS

Yazar, Alper

M.S., Department of Electrical and Electronics Eng.

Supervisor   : Assoc. Prof. Dr. Çağatay Candan

August 2015, 140 pages

Signal modeling is one of the important topics of signal processing area. The input signal should be modeled with a suitable mathematical model first. In statistics related disciplines, there are information theory based criteria for model order selection topic. In this thesis work, F-test based methods are proposed on model order selection and related problems. F-test is used in statistics related disciplines. However, it is not so widely used in signal processing related problems. Solution approaches for signal processing related problems based on known F-test are contributions of this thesis work. This work is focused on signals in linear spaces.

Fundamentally, F-test is a test of significance. It is used to test whether a signal model is sufficient to model the signal of interest or higher order models are needed. This test is made by using two nested models with different orders. RSS (Residual Sum of Squares) values are calculated for each model and they

are compared using F-test. According to the test result, it is determined that whether the lower order model is almost good as the higher order model or the higher order model improves the accuracy significantly. The proposed method is basically an iterative application of F-test. It selects the suitable model order by applying F-test many times.

In this work, some problems related with model order selection topic are solved using F-test based approaches. An analysis window length selection method for zero-crossing point estimation problem using line fit is proposed as the first example. Secondly, a method is proposed for the segmentation of multi tone signals. Similar approach is given as the third example for segmentation of FM signals. As the fourth example, a number of pole selection algorithm is proposed for all-pole signal modeling using Prony's method. Lastly, a segmentation method for damped sinusoidal signals with Prony's method is proposed. Simulation results are provided for each five problems.

Keywords: Signal Modeling, Linear Models, Parameter Estimation, Model Order Selection, Model Validity, Analysis Window Length Selection, Test of Significance, F-test, Nested Models, Zero-Crossing Estimation, Segmentation.

ÖZ

MODEL DERECESİ SEÇİMİ VE İLGİLİ PROBLEMLER İÇİN F-TESTİ
YÖNTEMİNİN UYGULANMASI

Yazar, Alper

Yüksek Lisans, Elektrik ve Elektronik Mühendisliği Bölümü

Tez Yöneticisi   : Doç. Dr. Çağatay Candan

Ağustos 2015 , 140 sayfa

İşaret modellemesi, işaret işleme alanının en önemli konularından biridir. Giriş işareti ilk olarak uygun bir matematiksel model ile gösterilmelidir. İstatistik ile ilgili alanlarda, model derecesi seçimi ile ilgili bilgi kuramı tabanlı çeşitli kriterler bulunmaktadır. Bu tez çalışmasında, model derecesi seçimi ve ilgili problemler için F-testi tabanlı bir yöntem önerilmiştir. F-testi, istastik ile ilgili alanlarda kullanılan bir yöntemdir. Bu tez çalışmasının temel katkısı, çeşitli işaret işleme problemleri için F-testi tabanlı çözümler sunmasıdır. Bu çalışma doğrusal uzaylarda bulunan işaretler üzerinedir.

Temel olarak F-testi bir önemlilik testidir. Bu test, bir modelin ilgilenen işareti gösterebilmek için yeterli olup olmadığını, daha yüksek dereceli modellere ihtiyaç duyulup duyulmadığını anlamak için yapılır. Test için farklı derecelerde iki adet içiçe modele ihtiyaç duyulmaktadır. Her iki model için de RSS (Residual Sum of Squares) veya AKT (Artık Kareler Toplamı) değerleri hesaplanır

ve bu değerler F-testi kullanılarak karşılaştırılır. Test sonucuna göre, düşük dereceli modelin neredeyse yüksek dereceli model kadar iyi olduğu veya yüksek dereceli modelin modelleme doğruluğunu önemli bir biçimde iyileştirdiği kararı verilir. Önerilen yöntem temel olarak F-testi yönteminin tekrarlamalı olarak kullanmaktadır. Yöntem, uygun olan model derecesini birden fazla F-testi yaparak seçmektedir.

Bu çalışmada, model derecesi seçimi ile ilgili bazı probemler F-testi tabanlı yaklaşımlarla çözülmüştür. İlk olarak, doğru oturtularak yapılan sıfır kesim noktası kestirimi problemi için analiz penceresi uzunluğu seçimi problemine bir yöntem önerilmiştir. İkinci örnek problem olarak, çok tonlu işaretlerin bölümlenmesi problemi incelenmiştir. Benzer bir yaklaşımın FM işaretler için uygulanması üçüncü örnek problem olarak verilmiştir. Dördüncü örnek olarak da, bir işaretin Prony yöntemi ile sadece kutuplu süzgeç çıktısı olarak modellenmesi probleminde kutup sayısı seçimi için bir yöntem önerilmiştir. Son problem ise, sönümlü sinüs işaretlerinden oluşan bir işaretin Prony yöntemi kullanılarak bölümlenmesi üzerinedir. Örnek problemler için benzetim sonuçları sunulmuştur.

Anahtar Kelimeler: İşaret Modellemesi, Doğrusal Modeller, Parametre Kestirimi, Model Derecesi Seçimi, Model Geçerliliği, Analiz Penceresi Uzunluğu Seçimi, Önemlilik Testi, F-testi, İçiçe Modeller, Sıfır Kesim Noktası Kestirimi, Bölümleme.

To my loving mother and father

and

Gonca

ACKNOWLEDGMENTS

software, especially LaTeX, I would have to deal with proprietary and hard to use WYSIWYG word editors to write my thesis.

Last but not least, deepest thanks to my mother and father for their love, trust, and every kind of support not only throughout my thesis but also throughout my life.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

ANOVA	ANalysis Of VAriance
CRLB	Cramér–Rao Lower Bound
EFM	Exponential FM
FM	Frequency Modulation
HFM	Hyperbolic FM
LFM	Linear FM
LS	Least Squares
ML	Maximum Likelihood
MLE	Maximum Likelihood Estimator
radar	RAdio Detection And Ranging
QFM	Quadrature FM
RMSE	Root Mean Square Error
RSS	Residual Sum of Squares
sonar	SOund Navigation And Ranging

CHAPTER 1

INTRODUCTION

In many applications of signal processing area, the input signal is modeled with a suitable and mathematically manageable model in the first steps of processing. The selection of a suitable model and its parameters is a fundamentally important signal processing problem in several applications such as power spectrum estimation with all pole modeling, impulse response modeling with Kth order filters etc. Dictionary meaning of the word "model" is given as mathematical description used for guidance or imitation [19]. According to this definition, there is no such thing as correct model. Indeed, a model is suitable if it satisfies requirements of the problem. Once a suitable model is determined, the model parameters are then estimated from the input.

Generally the model accuracy depends on its complexity. Using more parameters provides more detailed and potentially more accurate model. However, generally these parameters are estimated from observed signal and observation contains noise in addition to the actual signal that should be modeled. Estimations are prone to statistical errors caused by noise. Therefore, the model accuracy may get worsen after some point as more and more parameters are estimated from noisy observations. The model complexity and the model accuracy should be balanced. Then, fundamental questions of signal modeling arise: Which signal model should be used for a specific problem? Do we have simple models that satisfy problem requirements or do we need more complex ones?

Ideally, the signal model should be as simple as possible and at the same time represent the signal of interest with high fidelity. In addition the fact that

1

simple models are less prone to the effects of noise on parameter estimation, they also simplify the subsequent signal processing operations. It can be said that the main approach in model selection follows the principle of Occam's which is the utilization of the simplest model, the model with fewest constraints and assumptions, among the useful models.

The problem of model order selection has been examined from different viewpoints. One of the earliest works for model order selection problems is cross-validation. Cross-validation is primarily a way of measuring the predictive performance of a statistical model. Basically, a training set is chosen from observation to apply cross-validation. Then, training is done with the chosen set and the remaining observations are used for parameter estimations. By comparing errors for different set selections, cross-validation tries to find a suitable model for observation. K-fold cross-validation and leave-one-out cross-validation (LOOCV) are some example methods based on this approach. Also, there are information theory based approaches for model order selection. These approaches can be related with cross-validation based ones [31]. One of them is the Akaike Information Criterion, (AIC) which evaluates the generalized likelihood of the model, after estimating its parameters, and penalizes the likelihood with a rate proportional to the number of parameters [1]. Using a higher order model reduces the representation error, i.e. increases the generalized likelihood, at the expense of penalty associated with the higher order model. AIC seeks a balance between representation error and penalty. Several other criteria, similar to AIC, have been proposed in the literature [6,20,30,33]. Among these, Bayesian Information Criterion (BIC) and Generalized Information Criterion (GIC) have also found several applications [9,20,26,36].

In this thesis work, a model order selection rule is proposed for signals in linear spaces that are observed under additive white Gaussian noise. The proposed method is based on a statistical test used for ANOVA in statistics related disciplines called F-test [4,23]. There are some books and papers that utilize F-test for radar, communication, biomedical, array processing and some signal processing problems [3,8,10,13,15,16,34]. Although the origins of F-test date back to 1920's, it is not widely used in signal processing area [22].

In this thesis work, F-test based solutions for various signal processing problems are given. These problems are parameter estimation, model order selection, model validity and analysis window length selection problems. Different from the previously mentioned information theory based criteria, there is not any explicit penalty term related with the number of used parameters in F-test. However as it will be more clear in the following chapters, F-test based approaches for given problems use simple models with a predetermined probability of false model selection. Although fundamentals of all approaches are the same, they may not be used interchangeably for all cases.

1.1 Outline of The Thesis

This thesis work is divided into 5 chapters and the following chapters are organized as follows:

In Chapter 2, the properties of linear signal models are given to explain the basics of F-test based approach. Problems of interest and possible problem types are defined.

In Chapter 3, the basics of F-test are explained. F-test based solution approaches are suggested for the problems of interest.

In Chapter 4, the approaches proposed in Chapter 3 are applied with or without minor modifications on different signal processing problems.

In Chapter 5, a summary of the thesis work and possible future works are given.

1.2 Special Variables

Throughout this thesis work, some variables are given special meanings. List of reserved variables is given in Table 1.1. Although all variables will be defined in the following chapters properly, list is given as reference to reader. However, it should be noted that some variables may be used to represent other quantities inadvertently. Unless explicitly noted, they will be used with these meanings

after defining them in the following chapters properly.

Table 1.1: Special Variables

Variable	Represents
$\boldsymbol{A}$	Design matrix of a linear signal model
A	Amplitude of a sinusoidal signal
$\boldsymbol{e}$	Residuals vector
F	Most of the time F ratio value and sometimes frequency of a discrete time sinusoidal signal (cycles/sample)
f	Frequency of a continuous time sinusoidal signal (Hz)
h	Impulse response of an LTI system
$\boldsymbol{I}$	Identity matrix
K	Order of the actual signal model
L	Order difference between higher and lower order nested models.
l	Linear component of a signal
M	Model order of the model signal. For nested case, order of the lower order model
M_H	Model order of the model signal. For nested case, order of the higher order model
N	Number of observations
$\boldsymbol{p}$	Parameter vector of a linear signal model
p	Number of poles
q	Number of zeros
$\boldsymbol{s}$	Signal vector
$\boldsymbol{w}$	Noise vector
$\boldsymbol{x}$	Signal (in a linear space) vector
$\boldsymbol{y}$	Observation vector
z	Number of zeros
ϵ	Approximation error
Ω	Frequency of a discrete time sinusoidal signal (rad/sample)
ω	Frequency of a continuous time sinusoidal signal (rad/second)

1.3 Publications

The conference article [35] was presented in 23$^{\text{rd}}$ Signal Processing and Communications Applications Conference (SIU'15). Also poster entitled "Model Order Selection Using F-Test" was presented in METU EEE Graduate Research Workshop'15.

CHAPTER 2

PROBLEM DEFINITION

2.1 Linear Signal Model

In this work, real valued discrete time signals in linear spaces are considered. Although real signals are considered, the comments below can be extended to complex signals with proper declaration of operators. Signal model is given as

$$x = Ap. \tag{2.1}$$

Here, x is the **signal vector**.

Consider a signal in a linear space of dimension K observed at N different points. Each x_i in

$$x \triangleq \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}_{Nx1}$$

stands for one signal sample.

In the following equation, p is a column vector and is called the **parameter vector**. Each p_i in

$$p \triangleq \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_K \end{bmatrix}_{Kx1}$$

represents a single parameter of the signal given in (2.1).

In the following equation, $\boldsymbol{A}$ is called the **design matrix**,

$$\boldsymbol{A} \triangleq \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \dots & \boldsymbol{a}_K \end{bmatrix}_{NxK} \tag{2.2}$$

where each $\boldsymbol{a}_i$ is defined as

$$\boldsymbol{a}_i \triangleq \begin{bmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{iN} \end{bmatrix}_{Nx1}. \tag{2.3}$$

The $\boldsymbol{A}$ matrix can also be written as

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1K} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2K} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ a_{N1} & a_{N2} & a_{N3} & \dots & a_{NK} \end{bmatrix}_{NxK}$$

by combining (2.2) and (2.3).

The **signal space** or the column space of $\boldsymbol{A}$ is represented by $C(\boldsymbol{A})$ and is defined as the space spanned by the $\boldsymbol{a}_i$ vectors (columns of $\boldsymbol{A}$). The $\boldsymbol{x}$ vector is an element of this space (i.e. $\boldsymbol{x} \in C(\boldsymbol{A})$).

If $\boldsymbol{x}$ is a uniquely identifiable vector with K parameters, then

$$\text{rank}(A) = K. \tag{2.4}$$

In other words, $\boldsymbol{a}_i$ $(i = 1 : N)$ should form a linearly independent set.

It is assumed that the signal given in (2.1) is observed under zero-mean additive white Gaussian noise (AWGN) with variance σ^2 as

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{w} = \boldsymbol{Ap} + \boldsymbol{w}. \tag{2.5}$$

Here, $\boldsymbol{y}$ is the **observation vector**. Each y_i in

$$\boldsymbol{y} \triangleq \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}_{Nx1}$$

represents an observation point.

In equation (2.5), $\boldsymbol{w}$ is the **noise vector** representing the additive noise. Each w_i in

$$\boldsymbol{w} \triangleq \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix}_{Nx1,}$$

is a random variable with the following distribution

$$w_i \sim N(0, \sigma^2).$$

Mainly two different problems can be defined for signals in linear spaces observed under AWGN namely "Parameter Estimation" and "Model Order Selection".

2.2   Parameter Estimation Problem

One of the most important research topics in signal processing problems is the estimation of the signal parameters from noisy observations. For the case presented in the preceding section, the problem is the calculation of $\hat{\boldsymbol{p}}$ which is the estimate of $\boldsymbol{p}$ from $\boldsymbol{y}$. An effective parameter estimation method which can be applied here is the Maximum Likelihood Estimator (MLE). For this specific case, Least Squares (LS) and MLE solution gives the same result due to the signal model and the noise characteristics [18]. LS solution is found as,

$$\hat{\boldsymbol{p}} = \boldsymbol{A}^+ \boldsymbol{y} \tag{2.6}$$

where $\boldsymbol{A}^+$ is Moore–Penrose pseudoinverse of $\boldsymbol{A}$ for $N \geq K$ case defined as [14]

$$\boldsymbol{A}^+ \triangleq (\boldsymbol{A}^\mathrm{T} \boldsymbol{A})^{-1} \boldsymbol{A}^\mathrm{T}. \tag{2.7}$$

Consequently the LS signal estimate can be written as

$$\hat{\boldsymbol{x}} = \boldsymbol{A}\hat{\boldsymbol{p}}$$

by using parameter estimates.

The LS solution tries to minimize L2-norm of the error defined as

$$e \triangleq y - \hat{x}. \tag{2.8}$$

In other words, the vector $\hat{p}$ satisfies the following,

$$\hat{p} = \arg \min_{p} ||e||^2 \tag{2.9}$$

equality.

In statistics and in some other research fields, the resultant error on the observation vector after minimization given in (2.9) is called as the **Residual Sum of Squares (RSS)** and it is expressed as

$$\text{RSS} = ||e||^2. \tag{2.10}$$

In the following sections, the effect of N on the parameter estimation problems is analyzed for $N < K$ and $N >= K$. It will be assumed that the signal model is known completely. i.e. the matrix $\boldsymbol{A}$ and the parameter K are known.

2.2.1 $N < K$ (Insufficient Number of Observations)

The condition, $\text{rank}(\boldsymbol{A}) = K$, should be satisfied in order to uniquely identify the K different parameters of the signal. However, in that case the following situation will occur: $\text{rank}(\boldsymbol{A}) = K' < K$ which violates the condition given in (2.4). It may be thought that $\boldsymbol{x}$ is a linear combination of K' different parameters, not K. At least K observations should be made to observe and estimate the effects of K different parameters. It also makes sense that estimation of K different parameters from less number of observations causes some problems. Mathematically, $(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})$ product becomes singular. Therefore, $(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}$ does not exist. Consequently, it is general not feasible, to use less than K observations to estimate K different parameters.

2.2.2 $N \geq K$ (Sufficient or More Than Sufficient Number of Observations)

In this case, if $\mathrm{rank}(\boldsymbol{A}) = K$ then the parameters can be estimated without having any trouble in the calculation of $\boldsymbol{A}^+$. The relationship between the parameter estimation accuracy and the number of observations will be examined next. The estimation accuracy can be expressed by using the covariance matrix of the estimates, defined as follows

$$\Sigma_{\hat{p}} \triangleq \mathrm{E}\{\left[\hat{\boldsymbol{p}} - \boldsymbol{\mu}_{\hat{p}}\right]\left[\hat{\boldsymbol{p}} - \boldsymbol{\mu}_{\hat{p}}\right]^{\mathrm{T}}\}, \tag{2.11}$$

$$\Sigma_{\hat{p}} = \begin{bmatrix} \sigma_{p_{11}}^2 & \cdots & \cdots & \cdots & \cdots \\ \vdots & \sigma_{p_{22}}^2 & \cdots & \cdots & \cdots \\ \vdots & \vdots & \sigma_{p_{33}}^2 & \cdots & \cdots \\ \vdots & \vdots & \vdots & \ddots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \sigma_{p_{MM}}^2 \end{bmatrix}_{M \times M.}$$

Here, $\boldsymbol{\mu}_{\hat{p}}$ and $\sigma_{p_{ii}}^2$ are defined as follows

$$\boldsymbol{\mu}_{\hat{p}} \triangleq \mathrm{E}\{\hat{\boldsymbol{p}}\}, \tag{2.12}$$

$$\sigma_{p_{ii}}^2 \triangleq \mathrm{var}(\hat{p}_i).$$

Trace of the covariance matrix can be written as

$$\mathrm{tr}(\Sigma_{\hat{p}}) = \sum_{i=1}^{M} \sigma_{p_{ii}}^2 = \sum_{i=1}^{M} \mathrm{var}(\hat{p}_i).$$

Using (2.6) and (2.7), (2.12) can be written as follows

$$\begin{aligned} \boldsymbol{\mu}_{\hat{p}} &= \mathrm{E}\{\boldsymbol{A}^+ \boldsymbol{y}\} \\ &= \mathrm{E}\{\boldsymbol{A}^+ (\boldsymbol{A}\boldsymbol{p} + \boldsymbol{w})\} \\ &= \mathrm{E}\{\boldsymbol{A}^+ \boldsymbol{A}\boldsymbol{p}\} + \mathrm{E}\{\boldsymbol{A}^+ \boldsymbol{w}\} \\ &= \boldsymbol{p} \end{aligned}$$

which implies that the LS solution is an unbiased estimate of $\boldsymbol{p}$.

Equation (2.11) can expanded further as follows

$$
\begin{aligned}
\Sigma_{\hat{p}} &= E\{\left[\boldsymbol{A}^+\boldsymbol{y} - \boldsymbol{\mu}_{\hat{p}}\right]\left[\boldsymbol{A}^+\boldsymbol{y} - \boldsymbol{\mu}_{\hat{p}}\right]^{\mathrm{T}}\} \\
&= E\{\left[\boldsymbol{A}^+\boldsymbol{w}\right]\left[\boldsymbol{A}^+\boldsymbol{w}\right]^{\mathrm{T}}\} \\
&= \boldsymbol{A}^+\Sigma_w\boldsymbol{A}^{+\mathrm{T}}.
\end{aligned}
\tag{2.13}
$$

Trace of the matrices at both sides of the equation in (2.13) can be written as follows

$$
\begin{aligned}
\mathrm{tr}(\Sigma_{\hat{p}}) &= \mathrm{tr}(\boldsymbol{A}^+\Sigma_w\boldsymbol{A}^{+\mathrm{T}}) \\
&= \mathrm{tr}((\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\Sigma_w((\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}})^{\mathrm{T}}) \\
&= \mathrm{tr}((\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\Sigma_w\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}).
\end{aligned}
\tag{2.14}
$$

Due to the assumed noise characteristics, the covariance matrix of noise can be written as

$$
\Sigma_w = \sigma^2\boldsymbol{I}.
\tag{2.15}
$$

Using (2.15), (2.14) can be written as follows

$$
\begin{aligned}
\mathrm{tr}(\Sigma_{\hat{p}}) &= \sigma^2\,\mathrm{tr}((\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{I}\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}) \\
&= \sigma^2\,\mathrm{tr}((\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}).
\end{aligned}
\tag{2.16}
$$

Define $\boldsymbol{A}_N$ and $\boldsymbol{A}_{N+1}$ matrices as follows

$$
\boldsymbol{A}_N \triangleq
\begin{bmatrix}
a_{11} & a_{12} & \ldots & a_{1K} \\
a_{21} & a_{22} & \ldots & a_{2K} \\
\vdots & \vdots & & \vdots \\
a_{N1} & a_{N2} & \ldots & a_{NK}
\end{bmatrix}_{N\times K,}
$$

10

$$\boldsymbol{A}_{N+1} \triangleq \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1K} \\ a_{21} & a_{22} & \ldots & a_{2K} \\ \vdots & \vdots & & \vdots \\ a_{N1} & a_{N2} & \ldots & a_{NK} \\ a_{(N+1)1} & a_{(N+1)2} & \ldots & a_{(N+1)K} \end{bmatrix}_{N+1 \times K,}$$

$$\boldsymbol{A}_{N+1} = \begin{bmatrix} \boldsymbol{A}_N \\ \boldsymbol{a}^{\mathrm{T}} \end{bmatrix}_{N+1 \times K,}$$

$$\boldsymbol{a} \triangleq \begin{bmatrix} a_{(N+1)1} \\ a_{(N+1)2} \\ \vdots \\ a_{(N+1)K} \end{bmatrix}_{K \times 1.}$$

Here, $\boldsymbol{A}_N$ and $\boldsymbol{A}_{N+1}$ matrices represent the $\boldsymbol{A}$ matrix when N and $N+1$ observations are made, respectively. Consider the $(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}$ term shown in (2.16). It can be written for $N+1$ observations case as follows

$$\boldsymbol{A}_{N+1}^{\mathrm{T}}\boldsymbol{A}_{N+1} = \boldsymbol{A}_N^{\mathrm{T}}\boldsymbol{A}_N + \boldsymbol{a}\boldsymbol{a}^{\mathrm{T}},$$
$$(\boldsymbol{A}_{N+1}^{\mathrm{T}}\boldsymbol{A}_{N+1})^{-1} = (\boldsymbol{A}_N^{\mathrm{T}}\boldsymbol{A}_N + \boldsymbol{a}\boldsymbol{a}^{\mathrm{T}})^{-1}$$

and using Matrix Inversion Lemma,

$$(\boldsymbol{A}_{N+1}^{\mathrm{T}}\boldsymbol{A}_{N+1})^{-1} = (\boldsymbol{A}_N^{\mathrm{T}}\boldsymbol{A}_N)^{-1} - \frac{(\boldsymbol{A}_N^{\mathrm{T}}\boldsymbol{A}_N)^{-1}\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}(\boldsymbol{A}_N^{\mathrm{T}}\boldsymbol{A}_N)^{-1}}{1 + \boldsymbol{a}^{\mathrm{T}}(\boldsymbol{A}_N^{\mathrm{T}}\boldsymbol{A}_N)^{-1}\boldsymbol{a}},$$

$$\mathrm{tr}((\boldsymbol{A}_{N+1}^{\mathrm{T}}\boldsymbol{A}_{N+1})^{-1}) = \mathrm{tr}((\boldsymbol{A}_N^{\mathrm{T}}\boldsymbol{A}_N)^{-1})$$
$$- \mathrm{tr}\left(\frac{(\boldsymbol{A}_N^{\mathrm{T}}\boldsymbol{A}_N)^{-1}\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}(\boldsymbol{A}_N^{\mathrm{T}}\boldsymbol{A}_N)^{-1}}{1 + \boldsymbol{a}^{\mathrm{T}}(\boldsymbol{A}_N^{\mathrm{T}}\boldsymbol{A}_N)^{-1}\boldsymbol{a}}\right). \tag{2.17}$$

Consider the second term in (2.17). It can be written as

$$\mathrm{tr}\left(\frac{(\boldsymbol{A}_N^{\mathrm{T}}\boldsymbol{A}_N)^{-1}\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}(\boldsymbol{A}_N^{\mathrm{T}}\boldsymbol{A}_N)^{-1}}{1 + \boldsymbol{a}^{\mathrm{T}}(\boldsymbol{A}_N^{\mathrm{T}}\boldsymbol{A}_N)^{-1}\boldsymbol{a}}\right) = \mathrm{tr}\left(\frac{\boldsymbol{a}(\boldsymbol{A}_N^{\mathrm{T}}\boldsymbol{A}_N)^{-2}\boldsymbol{a}^{\mathrm{T}}}{1 + \boldsymbol{a}^{\mathrm{T}}(\boldsymbol{A}_N^{\mathrm{T}}\boldsymbol{A}_N)^{-1}\boldsymbol{a}}\right). \tag{2.18}$$

Since $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is a positive semi-definite matrix $(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} \geq 0)$, (2.18) is always positive; hence $\mathrm{tr}\left((\boldsymbol{A}_{N+1}^{\mathrm{T}}\boldsymbol{A}_{N+1})^{-1}\right) \leq \mathrm{tr}\left((\boldsymbol{A}_N^{\mathrm{T}}\boldsymbol{A}_N)^{-1}\right)$.

Consequently, an increase in the number of observations leads to an increase in the parameter estimation accuracy.

2.3 Model Order Selection Problem

In another possible scenario, model order (K) may be unknown. This problem is similar to the previous estimation problem. However, in this case in addition to the parameters to be estimated, model order is also unknown and it should be estimated too.

Let us assume that unknown and actual dimension of the signal space is K. Also assume that M represents the order of the tested model. In other words, order of the signal model is assumed to be M during parameter estimation. According to the relation between K and M, three different cases can be analyzed.

In the following sections, it is assumed that N is fixed. N is also assumed to be high enough to satisfy the condition in (2.4). In other words, N is greater than or equal to the order of the model with the highest order. Dependency of the expressions on N is not shown explicitly. However, the model order is shown explicitly in the expressions as subscript like $X_{\text{model order}}$ if necessary.

2.3.1 $M = K$ (Tested Model Order Matches The True Order)

Let us rewrite the error given in (2.8) as follows

$$
\begin{aligned}
\boldsymbol{e} =& \boldsymbol{y} - \boldsymbol{A}_M \hat{\boldsymbol{p}} \\
=& (\boldsymbol{I} - \boldsymbol{A}_M \boldsymbol{A}_M^+) \boldsymbol{y} \\
=& (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A}_M}) \boldsymbol{y} \\
=& \boldsymbol{P}_{\boldsymbol{A}_M}^{\perp} \boldsymbol{y}.
\end{aligned} \tag{2.19}
$$

Here, $\boldsymbol{P}_{\boldsymbol{A}_M}$ is a projection matrix defined as

$$
\boldsymbol{P}_{\boldsymbol{A}_M} \triangleq \boldsymbol{A}_M \boldsymbol{A}_M^+
$$

and it projects (observation) vectors to the **model signal space** denoted as $\mathrm{C}(\boldsymbol{A}_M)$.

Similarly, $\boldsymbol{P}_{\boldsymbol{A}_M}^{\perp}$ is a projection matrix defined as

$$
\boldsymbol{P}_{\boldsymbol{A}_M}^{\perp} \triangleq \boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A}_M}
$$

and it projects (observation) vectors to the **noise space**. Noise space and model signal space are orthogonal subspaces of the **observation space**. Dimensions of the model signal space, noise space, observation space and signal space are M, $N-M$, N, K, respectively.

The error expression given in (2.19) can further be simplified as

$$
\begin{aligned}
\boldsymbol{e} &= \boldsymbol{P}_{\boldsymbol{A}_M}^{\perp}(\boldsymbol{A}_K \boldsymbol{p}_K + \boldsymbol{w}) \\
&= \boldsymbol{P}_{\boldsymbol{A}_M}^{\perp} \boldsymbol{w}.
\end{aligned}
$$

Rewrite the RSS defined in (2.10) as follows

$$
\begin{aligned}
\mathrm{RSS} &= ||\boldsymbol{e}||^2 = (\boldsymbol{P}_{\boldsymbol{A}_M}^{\perp} \boldsymbol{w})^{\mathrm{T}} \boldsymbol{P}_{\boldsymbol{A}_M}^{\perp} \boldsymbol{w} \\
&= \boldsymbol{w}^{\mathrm{T}} (\boldsymbol{P}_{\boldsymbol{A}_M}^{\perp})^{\mathrm{T}} \boldsymbol{P}_{\boldsymbol{A}_M}^{\perp} \boldsymbol{w} \\
&= \boldsymbol{w}^{\mathrm{T}} \boldsymbol{P}_{\boldsymbol{A}_M}^{\perp} \boldsymbol{w}
\end{aligned}
\tag{2.20}
$$

and this final relation shows that only source of the error is noise. If $\boldsymbol{y}$ were a noiseless signal, i.e., $\boldsymbol{x}$ were used directly for estimation, parameters could be perfectly estimated.

$\boldsymbol{P}_{\boldsymbol{A}_M}^{\perp}$ can be decomposed into its eigenvalues and eigenvectors as

$$
\boldsymbol{P}_{\boldsymbol{A}_M}^{\perp} = \sum_{k=1}^{N} \lambda_k \boldsymbol{e_k} \boldsymbol{e_k}^{\mathrm{T}}.
$$

This is an $N \times N$ square matrix. Since it is a projection matrix, its eigenvalues (λ_K) are either 0 or 1 as

$$
\lambda_k = \begin{cases} 0 & 1 \leq k \leq M \\ 1 & M+1 \leq k \leq N. \end{cases}
$$

Since the dimension of the noise space is $N-M$, $N-M$ and M of eigenvalues are 1 and 0, respectively [14]. Finally, projection matrix expression can be written as

$$
\boldsymbol{P}_{\boldsymbol{A}_M}^{\perp} = \sum_{k=M+1}^{N} \boldsymbol{e_k} \boldsymbol{e_k}^{\mathrm{T}}
$$

where $\boldsymbol{e_k}$'s are $N \times 1$ orthonormal column vectors.

The RSS expression defined in (2.20) is continued as follows

$$\begin{aligned}
\text{RSS} &= \boldsymbol{w}^{\mathrm{T}} \sum_{k=M+1}^{N} \boldsymbol{e_k}\boldsymbol{e_k}^{\mathrm{T}}\boldsymbol{w} \\
&= \sum_{k=M+1}^{N} \boldsymbol{w}^{\mathrm{T}}\boldsymbol{e_k}\boldsymbol{e_k}^{\mathrm{T}}\boldsymbol{w} \\
&= \sum_{k=M+1}^{N} (\boldsymbol{e_k}^{\mathrm{T}}\boldsymbol{w})^2 \\
&= \sum_{k=M+1}^{N} z_k^2
\end{aligned} \qquad (2.21)$$

where

$$z_k \triangleq \begin{cases} \boldsymbol{e_k}^{\mathrm{T}}\boldsymbol{w} & M+1 \le k \le N \\ 0 & \text{otherwise.} \end{cases}$$

The elements of the noise vector $(\boldsymbol{w})$ consist of N independent random variables (w_k) with $N(0,\sigma^2)$ distribution. Consequently, $\Sigma_{\boldsymbol{w}} = \sigma^2\boldsymbol{I}$. Each z_k is also a random variable with distribution $N(0, \boldsymbol{e_k}^T\sigma^2\boldsymbol{I}\boldsymbol{e_k})$. Since, $\boldsymbol{e_k}$'s are mutually orthonormal vectors, $z_k \sim N(0,\sigma^2)$. Notice that similar to w_k, each z_k is an independent identically distributed (I.I.D.) random variable and that makes

$$\text{RSS} \sim \sigma^2\chi^2_{N-M} \qquad (2.22)$$

relation possible.

χ^2_{N-M} term in (2.22) represents a chi-squared distribution with $N-M$ degrees of freedom. Expected value of RSS is given as

$$\text{E[RSS]} = \sigma^2(N-M). \qquad (2.23)$$

2.3.2 $M > K$ (Tested Model Has Higher Order)

In that case, order of the tested model is greater than the order of the actual signal model. Let us expand the matrix $\boldsymbol{A}$, considering different model orders as follows

$$\boldsymbol{A_K} = \begin{bmatrix} \boldsymbol{a_1} & \boldsymbol{a_2} & \dots & \boldsymbol{a_K} \end{bmatrix}_{N \times K,}$$

$$\boldsymbol{A_M} = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \dots & \boldsymbol{a}_K & \boldsymbol{a}_{K+1} & \dots & \boldsymbol{a}_M \end{bmatrix}_{N \times M}$$
$$= \begin{bmatrix} \boldsymbol{A_K} & \boldsymbol{a}_{K+1} & \dots & \boldsymbol{a}_M \end{bmatrix}_{N \times M} \tag{2.24}$$

where

$$L \triangleq M - K.$$

As shown in (2.24), $\boldsymbol{A}_K$ is included in $\boldsymbol{A}_M$ completely. These two different models with model degree of K and M are said to be **nested models**. Model with model degree K is nested in model with model degree M. Model with model degree M has L additional parameters in comparison with the model with degree K. Two models generate the same signal. For the given nested model definition,

$$\mathrm{C}(\boldsymbol{A_K}) \subset \mathrm{C}(\boldsymbol{A_M}),$$
$$\mathrm{C}(\boldsymbol{P}_{\boldsymbol{A_M}}^\perp) \subset \mathrm{C}(\boldsymbol{P}_{\boldsymbol{A_K}}^\perp)$$

relations are valid. Error expressions can be written as follows

$$\begin{aligned} \boldsymbol{e} =& \boldsymbol{y} - \boldsymbol{A_M}\hat{\boldsymbol{p}}_M \\ =& (\boldsymbol{I} - \boldsymbol{A_M}\boldsymbol{A}_M^+)\boldsymbol{y} \\ =& (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A_M}})\boldsymbol{y} \\ =& \boldsymbol{P}_{\boldsymbol{A_M}}^\perp \boldsymbol{y} \\ =& \boldsymbol{P}_{\boldsymbol{A_M}}^\perp(\boldsymbol{A_K}\boldsymbol{p_K} + \boldsymbol{w}) \tag{2.25} \\ =& \boldsymbol{P}_{\boldsymbol{A_M}}^\perp \boldsymbol{w}. \tag{2.26} \end{aligned}$$

Notice that progression from (2.25) to (2.26) is correct since $\boldsymbol{P}_{\boldsymbol{A_M}}^\perp(\boldsymbol{A_K}\boldsymbol{p_K}) = \boldsymbol{0}$. Result of $\boldsymbol{A_K}\boldsymbol{p_K}$ is a column vector which is an element of $\mathrm{C}(\boldsymbol{A_K})$. Due to the fact that the model with model degree K is nested in the model with model degree M, it is also an element of $\mathrm{C}(\boldsymbol{A_M})$. $\boldsymbol{P}_{\boldsymbol{A_M}}^\perp$ projects vectors to the noise space which is orthogonal to $\mathrm{C}(\boldsymbol{A_M})$. Therefore, $\boldsymbol{P}_{\boldsymbol{A_M}}^\perp(\boldsymbol{A_K}\boldsymbol{p_K})$ yields 0.

Summary of the last two sections is that while $M \geq K$, the error expressions given in (2.22) and (2.23) are valid. As long as this condition is met, only source of the error are the noise components projected onto the noise space.

15

2.3.3 $M < K$ (Tested Model Has Smaller Order)

In this case, order of the model signal is less than the order of the actual signal. Expressions for $M > K$ case can be defined similarly for this situation as follows

$$\boldsymbol{A_K} = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \ldots & \boldsymbol{a}_M & \boldsymbol{a}_{M+1} & \ldots & \boldsymbol{a}_K \end{bmatrix}_{N \times K}$$

$$= \begin{bmatrix} \boldsymbol{A_M} & \boldsymbol{a}_{M+1} & \ldots & \boldsymbol{a}_K \end{bmatrix}_{N \times K,}$$

$$L \triangleq K - M,$$

$$\mathrm{C}(\boldsymbol{A_M}) \subset \mathrm{C}(\boldsymbol{A_K}),$$

$$\mathrm{C}(\boldsymbol{P}_{\boldsymbol{A_K}}^{\perp}) \subset \mathrm{C}(\boldsymbol{P}_{\boldsymbol{A_M}}^{\perp}),$$

$$\boldsymbol{x} = \boldsymbol{A_K}\boldsymbol{p_K},$$

$$\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{A_M}\hat{\boldsymbol{p}}_{\boldsymbol{M}}$$

$$= (\boldsymbol{I} - \boldsymbol{A_M}\boldsymbol{A}_{\boldsymbol{M}}^{+})\boldsymbol{y}$$

$$= (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A_M}})\boldsymbol{y}$$

$$= \boldsymbol{P}_{\boldsymbol{A_M}}{}^{\perp}\boldsymbol{y}$$

$$= \boldsymbol{P}_{\boldsymbol{A_M}}{}^{\perp}(\boldsymbol{x} + \boldsymbol{w}).$$

In contrast to the previously analyzed $M \geq K$ case, the model signal space is a subspace of the actual signal space. $\boldsymbol{P}_{\boldsymbol{A_M}}{}^{\perp}\boldsymbol{x} \neq \boldsymbol{0}$ for this case. There are remaining components of the actual signal in the noise space after projection. There are L extra non-zero parameters in the actual signal which can not be modeled completely by the model signal. Define c_k term similar to definition given in (2.22) as

$$c_k \triangleq \begin{cases} \boldsymbol{e_k}^{\mathrm{T}}\boldsymbol{x} & 1 \leq k \leq K \\ 0 & \text{otherwise.} \end{cases}$$

Similar to (2.21), RSS can be written as follows

$$\mathrm{RSS} = \sum_{k=M+1}^{N} (\boldsymbol{e_k}^{\mathrm{T}}(\boldsymbol{x} + \boldsymbol{w}))^2$$

$$= \sum_{k=M+1}^{K} c_k^2 + 2 \sum_{k=M+1}^{K} c_k z_k + \sum_{k=M+1}^{N} z_k^2. \tag{2.27}$$

16

Expected value of RSS is given as

$$\mathrm{E}\,[\mathrm{RSS}] = \sigma^2(N - M) + \sum_{k=M+1}^{K} c_k^2.$$

In summary, different than $M \geq K$ case there is a constant error term in the mean value of RSS which is independent from the noise signal as shown in (2.27). This term can be thought as the bias part of the error. This bias part is caused by the insufficiency of the model signal. Different than the other cases, even for noiseless situation RSS $\neq 0$.

2.3.4   Summary of Results

In the previous sections, parameter estimation problem of a signal from its noisy observation was examined. Depending on the relation between M and K, RSS can be expressed as follows

$$\mathrm{RSS} = \begin{cases} \sum_{k=M+1}^{N} z_k^2 & K < M \\ \sum_{k=M+1}^{N} z_k^2 & K = M \\ \sum_{k=M+1}^{K} c_k^2 + 2\sum_{k=M+1}^{K} c_k z_k + \sum_{k=M+1}^{N} z_k^2 & K > M. \end{cases} \qquad (2.28)$$

For all conditions, expected value of RSS is given as follows

$$\mathrm{E[RSS]} = \begin{cases} \sigma^2(N - M) & K < M \\ \sigma^2(N - M) & K = M \\ \sigma^2(N - M) + \sum_{k=M+1}^{K} c_k^2 & K > M. \end{cases}$$

RSS value for noiseless case is given as follows

$$\mathrm{RSS_{noiseless}} = \begin{cases} 0 & K < M \\ 0 & K = M \\ \sum_{k=M+1}^{K} c_k^2 & K > M. \end{cases}$$

17

In the last section, the change in RSS with respect to problem parameters is analyzed. K (true model order) will be varied while M (tested model order) is kept fixed. Since RSS value is the same for $K < M$ and $K = M$ as shown in (2.28), these two cases can be combined as a single $K \leq M$ condition.

2.3.4.1 $K \leq M$

Let us consider two different test models with order M and M_H, both of which are greater than the true model order K. Relationship between orders are given as follows

$$M_H = M + L,$$
$$L > 0,$$
$$M + L < N$$

and two models are considered to be nested.

For nested models, the relations given below

$$C(\boldsymbol{A_M}) \subset C(\boldsymbol{A_{M_H}}),$$
$$C(\boldsymbol{P}^{\perp}_{\boldsymbol{A_{M_H}}}) \subset C(\boldsymbol{P}^{\perp}_{\boldsymbol{A_M}})$$

remain valid as discussed previously.

RSS can be written as

$$\mathrm{RSS}_M = \sum_{k=M+1}^{N} z_k^2,$$
$$\mathrm{RSS}_{M_H} = \sum_{k=M_H+1}^{N} z_k^2 = \sum_{k=M+L+1}^{N} z_k^2$$

where the equation (2.28) is used.

The difference between the two RSS values then becomes

$$\mathrm{RSS}_M - \mathrm{RSS}_{M_H} = \sum_{k=M+1}^{N} z_k^2 - \sum_{k=M+L+1}^{N} z_k^2$$
$$= \sum_{k=M+1}^{M+L} z_k^2. \tag{2.29}$$

Therefore, RSS of the higher order model is smaller than the lower order model where the exact RSS difference is given in (2.29).

2.3.4.2 $K > M$

Similar to the previous case, let us take two nested models with degree M and M_H, both of which have smaller model order than the actual order. Relationship between orders are given as follows

$$
\begin{aligned}
M_H =& M + L, \\
L >& 0, \\
M + L <& N, \\
M_H <& K
\end{aligned}
$$

and two models are considered to be nested.

For nested models, the relations given below

$$
\begin{aligned}
\mathrm{C}(\boldsymbol{A_M}) \subset& \mathrm{C}(\boldsymbol{A_{M_H}}), \\
\mathrm{C}(\boldsymbol{P}^{\perp}_{\boldsymbol{A_{M_H}}}) \subset& \mathrm{C}(\boldsymbol{P}^{\perp}_{\boldsymbol{A_M}})
\end{aligned}
$$

remain valid as discussed previously.

Using equation (2.28), RSS can be written as follows

$$
\begin{aligned}
\mathrm{RSS}_M &= \sum_{k=M+1}^{K} c_k^2 + 2 \sum_{k=M+1}^{K} c_k z_k + \sum_{k=M+1}^{N} z_k^2, \\
\mathrm{RSS}_{M_H} &= \sum_{k=M_H+1}^{K} c_k^2 + 2 \sum_{k=M_H+1}^{K} c_k z_k + \sum_{k=M_H+1}^{N} z_k^2.
\end{aligned}
\tag{2.30}
$$

The difference between two RSS values then becomes

$$
\begin{aligned}
\mathrm{RSS}_M - \mathrm{RSS}_{M_H} &= \sum_{k=M+1}^{M+L} c_k^2 + 2 \sum_{k=M+1}^{M+L} c_k z_k + \sum_{k=M+1}^{M+L} z_k^2 \\
&= \sum_{k=M+1}^{M+L} \left(c_k + z_k \right)^2 \\
&\geq 0.
\end{aligned}
\tag{2.31}
$$

So, RSS of the higher order model is smaller than the lower order model where the exact RSS difference is given in (2.31).

As a special case, $M_H = K$ (true model order) condition can be analyzed separately. Equation (2.30) can be arranged for that special case as follows

$$\text{RSS}_M = \sum_{k=M+1}^{K} c_k^2 + 2 \sum_{k=M+1}^{K} c_k z_k + \sum_{k=M+1}^{N} z_k^2,$$

$$\text{RSS}_{M_H} = \sum_{k=M_H+1}^{N} z_k^2.$$

Then, the difference between two RSS values becomes

$$\begin{aligned}
\text{RSS}_M - \text{RSS}_{M_H} &= \sum_{k=M+1}^{K} c_k^2 + 2 \sum_{k=M+1}^{K} c_k z_k + \sum_{k=M+1}^{K} z_k^2 \\
&= \sum_{k=M+1}^{M+L} \left(c_k + z_k \right)^2 \\
&\geq 0.
\end{aligned} \tag{2.32}$$

So, RSS of the higher order model is smaller than the lower order model where the the exact RSS difference is given in (2.32).

All results obtained about the model order selection problems up to this point can be summarized as follows: Independent from the relation between K and M, as M increases RSS does not increase but decreases on the average. However, components of RSS varies according to the this relation. When $M \geq K$, RSS consists of noise components in the noise space only. In that case, RSS=0 for noiseless observation. When $M < K$, signal components that couldn't be modeled by the signal model remain in the noise space. This additional components contribute to RSS in addition to noise components. In that case, RSS decreases as M increases because of the reduction in both signal and noise components. There is not any signal component projected onto the noise space for $M \geq K$ case. As stated previously, K value is unknown for model order selection problems. Although as M increases, RSS always decreases on the average. This change is governed by the relation between M and K. The goal of the model order selection problem is to estimate the model order of the actual signal observed under noise. For that reason, choosing M which minimizes the RSS value

as the model order estimate is not a suitable approach. This M value makes the estimated signal close to the noisy observations, not to the actual signal. In the limiting case, taking $M = N$ makes $RSS = 0$ and independent from K,L and M but obviously, this is not a valid estimate of K. M should be chosen close to K even though RSS value is lower for higher M values. When $M = K$ all components of the signal lie in the model signal space. This is the best condition where M is minimum and all signal components are in the model signal space. This situation is called as **perfect fit**.[1]

All signal components still continue to stay in the model signal space but new noise components will be an element of the model signal space as M increases beyond K. In that case, the estimated signal is the sum of $\boldsymbol{x}$ and noise components projected onto the model signal space. This is not a desired case for the estimation since the signal to be modeled includes some additional noise components. This case is called as **over fit**.

When $M < K$, the model signal can't model all components of the actual signal. There are some signal components left which are projected onto the noise space and treated like noise. This is the **under fit** case.

As shown in the parameter estimation section, using more observations (N) for parameter estimation improves the estimation accuracy. If results of two types of problems are thought together, when maximum number of observations (N) is used (observing the actual signal), using $M = K$ yields the "the best result" for parameter estimation problems.

2.4 Related Problems

Problems related with parameter estimation of a signal in a linear space from its noisy observations can be classified into four categories. Each type shares some common concepts and approaches. Solutions are proposed for each type in the following chapters.

---

[1] "Perfect fit" should be considered as "perfect order fit" in this thesis work. Perfect fit case does not imply that the error in the parameter estimation is zero. "Perfect fit" case means that the order of the model signal is equal to the order of the actual signal ($K = M$).

2.4.1 Type I

In this type of problem, it is assumed that noisy observations of two signals with different models are concatenated to form a single observation. This scenario is expressed as follows

$$\boldsymbol{y}_1 = \boldsymbol{x}_1 + \boldsymbol{w}_1 = \boldsymbol{A}_1\boldsymbol{p}_1 + \boldsymbol{w}_1,$$

$$\boldsymbol{y}_2 = \boldsymbol{x}_2 + \boldsymbol{w}_2 = \boldsymbol{A}_2\boldsymbol{p}_2 + \boldsymbol{w}_2,$$

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \boldsymbol{0} \end{bmatrix}$$

and shown graphically in Figure 2.1. Models of two signals are known. $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ could be written completely if the length of the each observation was known.

Let us assume that first N' observations from total N observations belong to the signal with Model #1 with parameter vector $\boldsymbol{p}_1$ and remaining observations belong to the signal with Model #2 with parameter vector $\boldsymbol{p}_2$. These two models may be nested models as well. Correct observations should be used to calculate $\hat{\boldsymbol{p}}_1$ and $\hat{\boldsymbol{p}}_2$ using LS approach. $\hat{\boldsymbol{p}}_1$ and $\hat{\boldsymbol{p}}_2$ should be estimated using $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$, respectively. Total observation vector with length N should be split into two observations with length N' and $N - N'$ and each observation should be used for parameter estimation separately. This splitting process and each split section will be called as **segmentation** and **segment**, respectively. N' shown in Figure 2.1 is assumed to be unknown and it should be estimated. If there is a priori information about N', it may be used for segmentation.
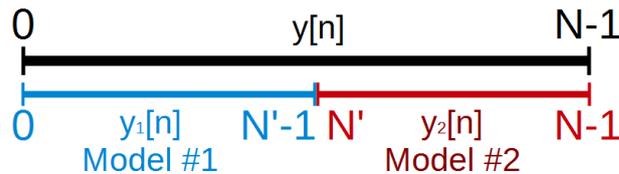
Figure 2.1: Illustration of Problem Type I

2.4.2 Type II

For Type II problem, it is assumed that observed signal is a noisy observation of a signal with unknown model order. This is a model order estimation problem. As stated previously, desired result for this parameter estimation problem is the perfect fit case. One should avoid under fit or over fit cases for estimation problems. For example, it may be known that the observation with length N is an observation of a polynomial function under noise but the polynomial order is unknown.

Estimation of K while N is fixed will be called as Type II problem and it is illustrated in Figure (2.2).
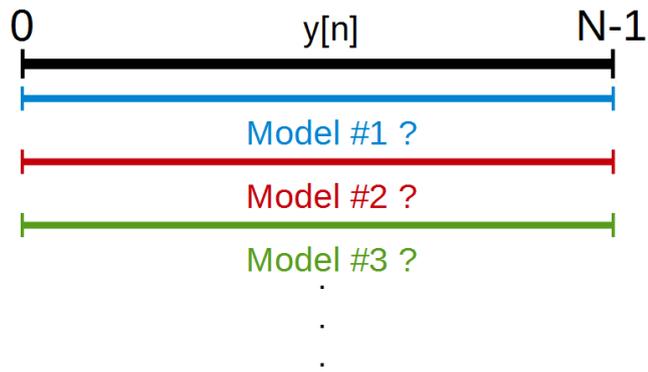
Figure 2.2: Illustration of Problem Type II

2.4.3 Type III

Type III problem is a combination of Type I and Type II. In addition to Type I problem, orders of models that should be used for each model signal, i.e. M values, are also unknown in addition to N' value. In this problem, observation should be segmented as in Type I and suitable signal models should be found for each segment as in Type II. If model orders shown in Figure 2.1 are also unknown, this problem is called as Type III problem.

2.4.4 Type IV

The last problem type is Type IV. In this scenario, parameters of a signal which is approximately in the assumed signal space is estimated. It is assumed that observed signal $\boldsymbol{y}$ with length N is a noisy observation of signal $\boldsymbol{s}$ which is given as

$$\boldsymbol{y} = \boldsymbol{s} + \boldsymbol{w}.$$

In contrast to the earlier problem, $\boldsymbol{s}$ is not an element of a linear space, necessarily, that is, it may not satisfy $\boldsymbol{s} = \boldsymbol{A}\boldsymbol{p}$ condition.

Signal in a linear space is defined as

$$\boldsymbol{x}_1 = \boldsymbol{A}_1 \boldsymbol{p}_1. \tag{2.33}$$

New error term (which results from approximating $\boldsymbol{x}$ using $\boldsymbol{x}_1$) is defined as

$$\epsilon(n) = s(n) - x_1(n) \tag{2.34}$$

and it is called the **approximation error**. This error is assumed to satisfy the following inequality,

$$|\epsilon(n')| > |\epsilon(n)| \qquad \text{where } n' > n.$$

That is, approximation error increases as the signal drifts away from the assumed linear signal space. In certain applications, one may wish to approximate $\boldsymbol{s}$ using $\boldsymbol{x}_1$ and estimate the parameters of $\boldsymbol{x}_1$ from these observations. Question is how many samples should be used to approximate $\boldsymbol{s}$ as a signal in a linear space? In other words, what is a good choice for N' value shown in Figure 2.3. While using more observations for parameter estimation of $\boldsymbol{x}_1$ increases the estimation accuracy, approximation error increases with increasing N. This situation is called **model mismatch**. Window between $[0 \ N' - 1]$ is called **analysis window**. Observations for parameter estimation are taken within this window. Selection of N' is also called as analysis window length selection problem.
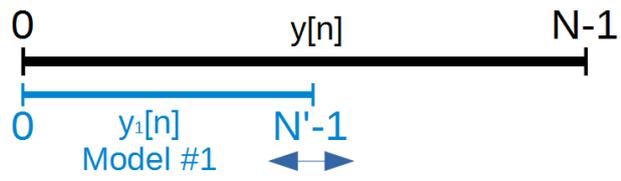
Figure 2.3: Illustration of Problem Type IV

CHAPTER 3

THE PROPOSED METHOD

In this section, the proposed method for the problem types mentioned in the preceding sections is explained. Main idea is using the maximum number of observations for parameter estimation while keeping the model mismatches minimal, i.e., providing the best parameter estimation results by targeting perfect fit to the model.

In the previous chapter, change in RSS is analyzed for different M (tested model order) and K (true model order) values. Basic principle behind the proposed method is to track the change in RSS. If the order of the actual signal model (K) is fixed and the order (M) of the model signal is increased from $M < K$ condition, RSS value drops. This drop is due to the decrease in both signal and noise components projected onto the noise space. After $M = K$ point, reduction in RSS is solely due to the reduction in noise components projected onto the noise space. If it is possible to detect this change, it is also possible to detect when M reaches K, i.e, the perfect fit case. The proposed method tries to make use of the drop in RSS for model order selection.

A test method known as F-test in literature is studied as a solution for four types of problems mentioned earlier. Although F-test is well known in the statistics literature, it is not widely used in signal processing problems, as stated previously. $\boldsymbol{F}$ **ratio (value)** is calculated as

$$F = \frac{\dfrac{\mathrm{RSS}_M - \mathrm{RSS}_{M_H}}{M_H - M}}{\dfrac{\mathrm{RSS}_{M_H}}{N - M_H}} \qquad (3.1)$$

27

and it will be shown that this ratio is suitable for the problems considered.

To define F value properly,

$$M_H = M + L,$$
$$L > 0,$$
$$M_H < N$$

conditions should be satisfied.

In equation (3.1), RSS_M and RSS_{M_H} represent RSS values when the signal is modeled with models with order M and M_H, respectively. It is assumed that the model with order M is nested in the model with order M_H.

In the next section, the characteristics of F value is studied in order to better explain the reasons behind the suggested utilization of RSS in model order selection better.

3.1 Characteristics of F Ratio

Properties of F ratio is given in [29] in detail. In this section, properties that are necessary to understand proposed methods are given. F ratio is analyzed for three different conditions.

3.1.1 $M > K$

Using equation (2.28), RSS can be written as follows

$$\text{RSS}_M = \sum_{k=M+1}^{N} z_k^2$$
$$\sim \sigma^2 \chi_{N-M,}^2$$
$$\text{RSS}_{M_H} = \sum_{k=M_H+1}^{N} z_k^2 = \sum_{k=M+L+1}^{N} z_k^2$$
$$\sim \sigma^2 \chi_{N-M-L}^2. \tag{3.2}$$

Then, the difference between RSS values becomes

$$\text{RSS}_M - \text{RSS}_{M_H} = \sum_{k=M+1}^{M+L} z_k^2$$

$$\sim \sigma^2 \chi_L^2. \tag{3.3}$$

Using the equations (3.1), (3.2) and (3.3), F ratio can be written as follows

$$F = \frac{\dfrac{\sum\limits_{k=M+1}^{M+L} z_k^2}{L}}{\dfrac{\sum\limits_{k=M+L+1}^{N} z_k^2}{N-M-L}} \tag{3.4}$$

$$\sim \frac{\dfrac{\sigma^2 \chi_L^2}{L}}{\dfrac{\sigma^2 \chi_{N-M-L}^2}{N-M-L}}$$

$$\sim \frac{\dfrac{\chi_L^2}{L}}{\dfrac{\chi_{N-M-L}^2}{N-M-L}} \tag{3.5}$$

and F ratio is a random variable with F distribution [17].

Consider a random variable, X, defined as

$$X = \frac{x_1/d_1}{x_2/d_2}. \tag{3.6}$$

Assume that in (3.6), x_1 and x_2 terms represent random variables which have chi-squared distribution with d_1 and d_2 degrees of freedom, respectively. If x_1 and x_2 are independent random variables, X is a random variable with F distribution [17].

If χ_L^2 and χ_{N-M-L}^2 terms are independent, the expression given in (3.5) is a random variable with F distribution. Summation terms shown in (3.4) can be considered to show independence. Due to minimum and maximum limits of summation terms in the nominator and the denominator, a particular z_i is

summed up in either the numerator or the denominator. There is no common z_i term that appears in both the nominator and the denominator. Since each z_i is independent from each other as explained previously, the numerator and the denominator are independent from each other. Therefore, the expression given in (3.5) has an F distribution. F distribution can be characterized by the degrees of freedom of chi-squared random variables in the numerator and the denominator. Probability Density Function (PDF) expression of random variable X defined in (3.6) is given in as

$$
\begin{aligned}
f(x; d_1, d_2) &= \frac{\sqrt{\frac{(d_1\,x)^{d_1}\,d_2^{d_2}}{(d_1\,x+d_2)^{d_1+d_2}}}}{x\,\mathrm{B}\!\left(\frac{d_1}{2}, \frac{d_2}{2}\right)} \\
&= \frac{1}{\mathrm{B}\!\left(\frac{d_1}{2}, \frac{d_2}{2}\right)} \left(\frac{d_1}{d_2}\right)^{\frac{d_1}{2}} x^{\frac{d_1}{2}-1} \left(1 + \frac{d_1}{d_2}\,x\right)^{-\frac{d_1+d_2}{2}}
\end{aligned} \tag{3.7}
$$

for real $x \geq 0$. $f(x; d_1, d_2) = 0$ for $x < 0$. B term in (3.7) is the Beta function and given as

$$
\mathrm{B}(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,\mathrm{d}t \qquad \text{for } \mathrm{Re}(x), \mathrm{Re}(y) > 0.
$$

In this work, PDF expression of F distribution is not used directly. Rather characteristics of F distribution are used. PDFs and CDFs of F distribution for various parameters are shown in Figure 3.1 and Figure 3.2, respectively.
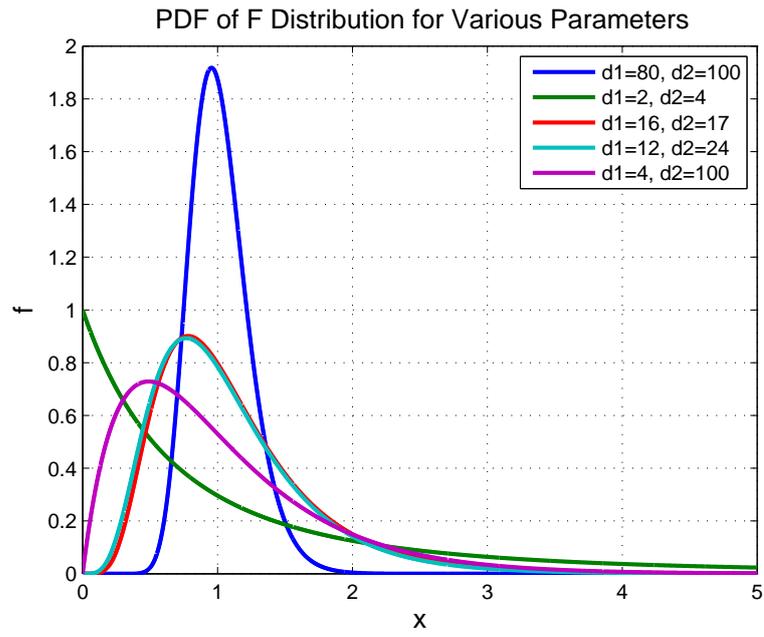
Figure 3.1: PDF of F Distribution
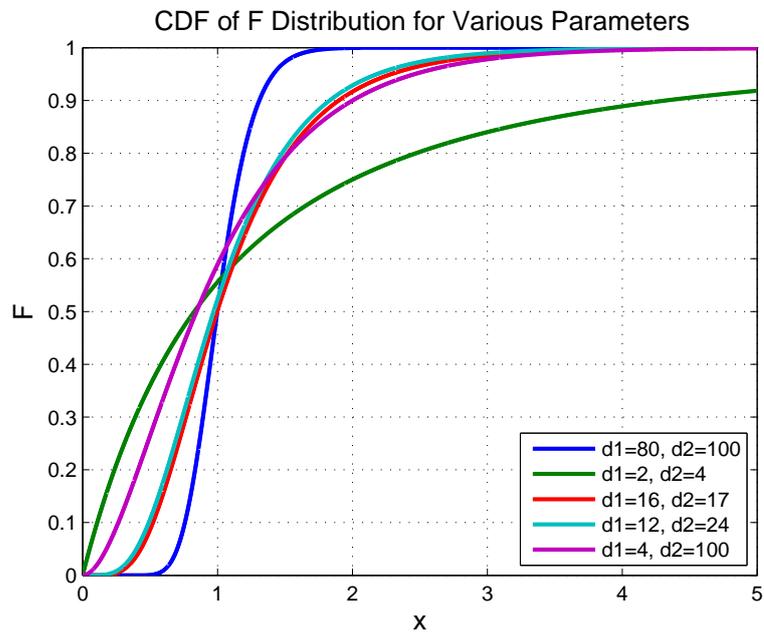
Figure 3.2: CDF of F Distribution

3.1.2 $M < K$ and $M_H \geq K$

RSS expressions can be written as follows

$$\text{RSS}_{\text{M}_\text{H}} = \sum_{k=M_H+1}^{N} z_k^2 = \sum_{k=M+L+1}^{N} z_k^2$$

$$\sim \sigma^2 \chi_{N-M-L}^2,$$

$$\text{RSS}_M - \text{RSS}_{M_H} = \sum_{k=M+1}^{M+L} (c_k + z_k)^2$$

$$= \sum_{k=M+1}^{K} (c_k + z_k)^2 + \sum_{k=K+1}^{M+L} z_k^2$$

$$= \sum_{k=M+1}^{M+L} z_k^2 + \sum_{k=M+1}^{K} c_k^2 + 2 \sum_{k=M+1}^{K} c_k z_k.$$

F ratio defined in (3.1) can be written as

$$F'' = \cfrac{\cfrac{\displaystyle\sum_{k=M+1}^{M+L} z_k^2 + \sum_{k=M+1}^{K} c_k^2 + 2 \sum_{k=M+1}^{K} c_k z_k}{L}}{\cfrac{\displaystyle\sum_{k=M+L+1}^{N} z_k^2}{N-M-L}}$$

$$= \cfrac{\cfrac{\displaystyle\sum_{k=M+1}^{M+L} z_k^2}{L}}{\cfrac{\displaystyle\sum_{k=M+L+1}^{N} z_k^2}{N-M-L}} + \cfrac{\cfrac{\displaystyle\sum_{k=M+1}^{K} c_k^2 + 2 \sum_{k=M+1}^{K} c_k z_k}{L}}{\cfrac{\displaystyle\sum_{k=M+L+1}^{N} z_k^2}{N-M-L}}. \qquad (3.8)$$

After that point let F' denotes the F ratio defined in (3.4) for $M > K$ case. Then, equation (3.8) can be rewritten as

$$F'' = F' + \cfrac{\cfrac{\displaystyle\sum_{k=M+1}^{K} c_k^2 + 2 \sum_{k=M+1}^{K} c_k z_k}{L}}{\cfrac{\displaystyle\sum_{k=M+L+1}^{N} z_k^2}{N-M-L}}. \qquad (3.9)$$

r is defined as

$$r \triangleq \sum_{k=M+1}^{K} c_k^2 + 2 \sum_{k=M+1}^{K} c_k z_k$$

to simplify expressions. It is a random variable defined as

$$r \sim N(\mu_r, \sigma_r^2),$$

$$\mu_r = \sum_{k=M+1}^{K} c_k^2,$$

$$\sigma_r^2 = 4\sigma^2 \sum_{k=M+1}^{K} c_k^2 = 4\sigma^2 \mu_r.$$

When $r > 0$ condition is met, F'' defined in (3.9) becomes greater than F'. The reason behind the analysis of this condition will be clear in the following sections. Probability of this condition can be found as follows

$$\Pr\{r > 0\} = \Pr\{F'' > F'\}$$
$$= Q\left(-\frac{\sqrt{\mu_r}}{2\sigma}\right). \tag{3.10}$$

Plot of (3.10) is given in Figure 3.3 and Figure 3.4 when σ^2 and μ_r are kept constant, respectively. When L, M and K are fixed, Figure 3.4 also can be thought as probability values for different SNR values of observations. As SNR of observed signal increases, probability of $F'' > F'$ increases.

The mean value of F'' can be found as

$$\mathrm{E}\left[F''\right] = \mathrm{E}\left[F'\right] + \mathrm{E}\left[\frac{\dfrac{\sum_{k=M+1}^{K} c_k^2 + 2\sum_{k=M+1}^{K} c_k z_k}{L}}{\dfrac{\sum_{k=M+L+1}^{N} z_k^2}{N-M-L}}\right]$$

$$= \mathrm{E}\left[F'\right] + \frac{N-M-L}{L}\mathrm{E}\left[\frac{\sum_{k=M+1}^{K} c_k^2 + 2\sum_{k=M+1}^{K} c_k z_k}{\sum_{k=M+L+1}^{N} z_k^2}\right]. \tag{3.11}$$

33
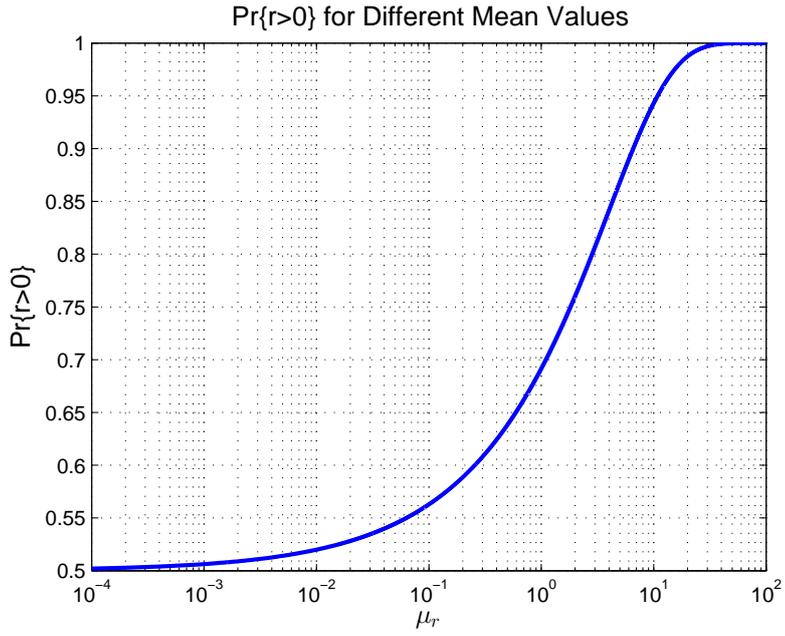
Figure 3.3: $\Pr\{F'' > F'\}$ for Different μ_r Values with Fixed σ^2
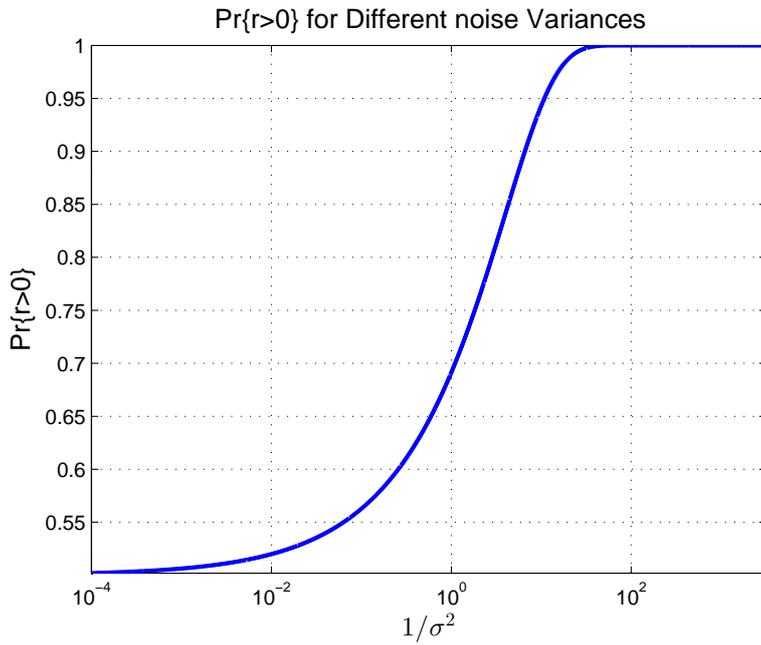
Figure 3.4: $\Pr\{F'' > F'\}$ for Different σ^2 Values with Fixed μ_r

Consider the second term in (3.11). The second expectation term can be written

as follows

$$E\left[\frac{\sum_{k=M+1}^{K} c_k^2 + 2\sum_{k=M+1}^{K} c_k z_k}{\sum_{k=M+L+1}^{N} z_k^2}\right] = E\left[\frac{\sum_{k=M+1}^{K} c_k^2}{\sum_{k=M+L+1}^{N} z_k^2}\right] + 2E\left[\frac{\sum_{k=M+1}^{K} c_k z_k}{\sum_{k=M+L+1}^{N} z_k^2}\right]. \quad (3.12)$$

In this expression numerator and denominator of the first term are both positive. So, expectation of the first term in (3.12) is a positive value. Define

$$r \triangleq \frac{n}{d} = \frac{\sum_{k=M+1}^{K} c_k z_k}{\sum_{k=M+L+1}^{N} z_k^2}$$

to simplify expressions. Notice that due to summation indices, n and d are independent. From law of total expectation, expectation of r can be written as

$$E[r] = E_d\left[E_{r|d}[r|d]\right].$$

Since $E_{r|d}[r|d] = 0$ for all r and d values, $E[r] = 0$. So, expectation of the second term in (3.12) is 0 and it makes $E[F''] > E[F']$. F'' which is the value of F when $M_H \geq K > M$ is greater than F' which is the value of F when $M_H > M \geq K$ in average. The exact expression is given in (3.9).

3.1.3 $M_H < K$

For this case, RSS values can be written as follows

$$RSS_M = \sum_{k=M+1}^{K} c_k^2 + 2\sum_{k=M+1}^{K} c_k z_k + \sum_{k=M+1}^{N} z_k^2,$$

$$RSS_{M_H} = \sum_{k=M_H+1}^{K} c_k^2 + 2\sum_{k=M_H+1}^{K} c_k z_k + \sum_{k=M_H+1}^{N} z_k^2,$$

$$RSS_M - RSS_{M_H} = \sum_{k=M+1}^{M+L} c_k^2 + 2\sum_{k=M+1}^{M+L} c_k z_k + \sum_{k=M+1}^{M+L} z_k^2.$$

F ratio can be written as

$$F''' = \frac{N-M-L}{L}\frac{RSS_M - RSS_{M_H}}{RSS_{M_H}}.$$

In that case, it is not so easy to show the relation between F''' and F' due to c_k terms in both numerator and denominator. Furthermore, final expressions would be dependent on almost all problem parameters and they should be analyzed for various cases. However, one does not have to find all equations for all cases in order to understand the operation of the F-test for $M_H < K$ condition. Therefore, instead of a given set of equations for all cases, important results are emphasized for particular cases.

F-test as defined in the next section is actually a test of significance as stated previously. It gives decisions according to F values. Larger F values for two nested model with order M and M_H mean that the model with order M_H improves the model accuracy or model the actual signal "significantly better" than the model with order M. If the value of F ratio is relatively low, it means that the model with order M_H does not improve the modeling accuracy "significantly" than the model with order M.

F ratio can also be seen as only the ratio of RSS values for different orders of model signal if the ratio formed by skipping N, M and L terms shown in (3.1). F' value is calculated for the case when the model orders of both model signals are greater than the order of the actual signal. In that case, both RSS values are error signals caused only by observation noise.

It was shown that F'' is greater than F' with an increasing probability with an increase in SNR. Also the inequality $F'' > F'$ is true in average. The reason is that RSS_M consists of both noise and signal components that can not be modeled whereas RSS_{M_H} does not have any signal component. So, it makes sense that $\text{RSS}_M - \text{RSS}_{M_H}$ can be greater than the previous case. This situation can also be considered from significance perspective: F ratio increases because the model with order M_H models the actual signal "significantly better" than the model with order M. The model with order M has some missing signal components. The difference between two F values (F' and F'') is highly dependent on the difference between magnitudes of the signal components that can not be modeled by a model with order of M and noise components added to the signal space when the model order is increased from M to M_H. If magnitude of the

mentioned signal components becomes greater than the magnitude of mentioned noise components, change in F ratio becomes more distinguishable.

However, $F'' < F'$ may be true with a non-zero probability depending on the SNR value as shown in Figure 3.4. Low F ratio means that model with order M_H does not model the signal "significantly better" than model with order M. This may be true in low SNR case particularly. In this case, the signal components that can't be modeled by the lower order model, M, may not be distinguishable than the noise components that are included in the model signal space when the model order is changed from M to M_H. To avoid this situation, the missing signal components should be "sufficiently" dominant than the new noise components added in the model signal space when the model order is increased. If the newly added signal components aren't dominant than the noise components, F-test will fail to distinguish the signal components from the noise components.

In the last case, both models with model order M and M_H can't estimate all components of the actual signal. Both RSS values have signal components that can't be modeled by the model signal as an error source. Firstly, let us consider high SNR and $M < M_H < K$ case. In that case, most of the reduction in RSS when the model order is changed from M to M_H is caused by the reduction in signal components that can't be modeled by insufficient models. If this reduction is "signficant", then $F''' > F'$. For low SNR case, reduction in RSS when the model order is increased is dominated by the new noise components included in the new model signal space rather than the actual signal components. This makes $F''' \simeq F'$, meaning that reduction in RSS is probably due to noise components and there is not any "signficant" signal component that is modeled by the higher order model. The lower order model can model the actual signal "as good as" the higher order one.

3.2 The F-test

F-test was initially developed in the statistics literature as the variance ratio by Fisher in 1920s [22].

The ultimate goal of the F-test based approaches in this thesis work is to estimate the number of parameters, $\hat{K}$, in other words model order of a signal in a linear space using N noisy observations. If $\hat{K}$ is estimated close to K then $\hat{p}$ can be found accurately as in perfect fit case. Suggested algorithm is given as follows:

1. Initially a false decision probability should be determined. The effect of this parameter will be explained in detail. This value will be denoted by p_{fd}. Since it denotes a probability, p_{fd} can be $0 \leq p_{fd} \leq 1$. However, the inequality given below should be considered to make F-test useful,

$$0 < p_{fd} < 1.$$

2. Two suitable nested models with model degree M and M_H ($M_H > M$) should be determined for the problem. If there is not any a priori information about K, model degrees should be chosen as low as possible. Also, M_H should be close to M as much as possible to increase $\hat{K}$ resolution. If it is possible, taking $M_H = M + 1$ gives the best $\hat{K}$ resolution. However, this may not be logical for all problems which is shown in one of the example problems later.

3. A threshold should be chosen for F-test. This threshold value is calculated by making $K \leq M < M_H$ assumption. Previously, it was shown that F ratio for that case (F') is a random variable with F-distribution: $F' \sim \mathrm{F}(L, N - M_H)$. The calculated F ratio is checked against assumed F-distribution. This is done by calculating a *threshold* value using the CDF of F-distribution and p_{fd} value as given in (3.13).

$$threshold = \mathrm{F_{CDF}}^{-1}(1 - p_{fd}, L, N - M_H) \qquad (3.13)$$

If the assumption is correct, $F < threshold$ condition is satisfied with $1 - p_{fd}$ probability.

4. F ratio given in (3.1) is calculated by using RSS_M and RSS_{M_H} values. Then, F ratio is compared against *threshold* value. If $F \geq threshold$, it is assumed that model with order M_H can model the signal "significantly better" than model with order M. In other words, $M < K$. On the other hand, test may give wrong decision with p_{fd} probability even if the initial assumption ($K \leq M < M_H$) is valid. This is why p_{fd} is called as false decision probability.

5. Both M and M_H are increased by the same amount. Most of the time, increasing them by one is suitable in order to keep $\hat{K}$ resolution high. This increase may be taken higher than one in order to speed up the test at the expense of $\hat{K}$ resolution. New *threshold* and F ratio is calculated and compared for new M and M_H values. This increase, calculate and compare cycle is continued until $F < threshold$ condition is met. When $F < threshold$ condition is met, it is decided that model with order M can model the actual signal "sufficiently good" as model with order M_H. Therefore, model order of the actual signal can be taken as $\hat{K} = M$.

Here, it is assumed that one of the tested models is appropriate for the actual signal. Then, F-test is used to select the appropriate one from the set of tested models.

"Sufficiency" or "goodness" is directly related with p_{fd} parameter. Under any circumstances, $0 < F < \infty$. As p_{fd} decreases, *threshold* increases as it can be seen from Figure 3.2. As *threshold* increases, $F < threshold$ condition is satisfied at lower M and M_H values. As false decision probability decreases, "sufficiency" condition becomes "tighter". One can consider taking p_{fd} values low in order to increase *threshold* and put more strict "sufficiency" conditions. However, in that case there is a risk such that F ratio can be smaller than *threshold* at the beginning of the test, e.g., when $M = 1$ and $M_H = 2$ so K can't be estimated properly. This situation causes under fit. On the other hand, taking p_{fd} so high increases the probability of over fit.

Choice of p_{fd} value affects the performance of F-test directly. Optimum p_{fd} value for the best performance is not the subject of this thesis study and it deserves

special consideration. In this study, p_{fd} is chosen as 0.1 in general empirically. However, the effect of p_{fd} on performance will be shown with simulation results.

One of the main advantages of F-test is that it does not use the noise variance (σ^2) information. Noise variance may be unknown and does not need to be estimated. F ratio is found using RSS values and *threshold* is calculated using the CDF of F-distribution. On the other hand, SNR should be relatively high enough to find $\hat{K}$ close enough to K. This is because, F-test should be able to distinguish reasons behind the decrease in RSS as model order of the model signal increases as explained previously. If SNR is not high enough, it couldn't be possible to understand whether the decrease is caused only by the reduction of noise components or noise components plus signal components. In that case, F-test may decide to stop at a model with order much lower than K and decide that it is "good enough" to model observed signal since further increase in model order does not provide additional "significant" benefits. Obviously, this is not the desired case for model order estimation problems. Effects of SNR on relations between the F ratios for different cases were mentioned briefly in the previous sections with the help of Figure 3.4. The relationship between SNR and parameter estimation error may be analyzed in a separate study.

3.3 Application of F-test to The Related Problems

In this section, solution approaches based on F-test are suggested for previously mentioned four different problem types. In the next chapter, these approaches will be supported by examples.

3.3.1 Type I

Utilization of **search window** is suggested for this type of problems. Length of search window will be denoted by N_{sw}. This window is shifted step by step (for example, one by one) through observation vector from the beginning to the end. This shift corresponds to taking N_{sw} samples from different locations of the observation vector. The lower order model with order M is chosen as

Model #1 which is shown in Figure 3.5. The higher order model with order M_H covers both Model #1 and Model #2. Notice that, the two models become nested models with this selection rule. F-test is done for each window. If $F < threshold$ condition is satisfied for a window, it can be said that this particular window consists of observations from only Model #1, otherwise it is decided that observations from both Model #1 and Model #2 exist in that window.

It is assumed that $N_{sw} = M_H + 1$ which is the minimum allowable N_{sw} not to have problems related with rank and make $\text{RSS}_{M_H} = 0$. Suppose that the first window does not have any observation from Model #2. As the search window is shifted step by step, after some point it includes the point N' shown in Figure 3.5 and the windowed observation consists of observations from both models. In this case $F \geq threshold$ condition is satisfied. For example, centre of the search window may be taken as $\hat{N}'$ when F first becomes greater than $threshold$. Depending on the specific problem, the $\hat{N}'$ estimation can be found in a different way like by taking the starting point, not the centre, of the search window as $\hat{N}'$.

Depending on the desired resolution of N', N_{sw} and shifting step size may be modified. This solution approach is illustrated in Figure 3.5.
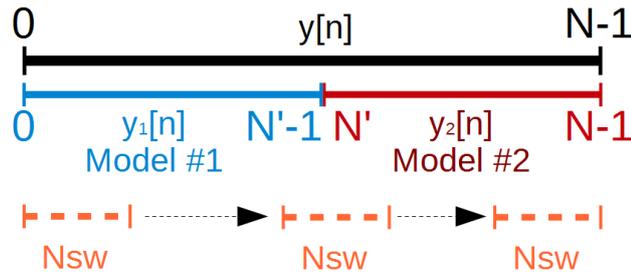
Figure 3.5: Illustration of Proposed F-test Based Approach for Problem Type I

3.3.2 Type II

This type is a straightforward problem for F-test approach. Two models with order M and M_H with $M_H > M$ condition are determined initially. They should be nested models. Furthermore keeping $M_H - M$ difference small, like

one, improves the $\hat{K}$ resolution. RSS values are calculated for each model and F ratio is found. If the condition $F \geq threshold$ is satisfied, it is assumed that $M < K$. M and M_H are increased while keeping the $M_H - M$ difference the same. Then F and $threshold$ are compared. This cycle continues until the $F < threshold$ condition is met. When F becomes smaller than the $threshold$, test is ended up by taking $\hat{K} = M$. You can examine Sec.4.4 for an application of F-test for this type of problem.

3.3.3 Type III

Different than Type I case, model orders of the actual signals are also unknown. This problem includes both segmentation (Type I) and model order selection (Type II) problem. Model orders are estimated first. To estimate parameters of signal #1, the longest possible window is taken from beginning of the observation. If there is a priori information about N', it should be used to decide the proper window length. It should be noted that this window has to exclude observations from the second signal. Parameters of the first observation is estimated using the approach defined for Type II problem. Then the solution approach for Type I problem is applied.Initially, the lower order model with order M stated in Type I approach is formed by the estimated model of signal #1. On the other hand, the higher order with order M_H has to include an information about model #2 in addition to model #1. At that moment, model order of the second signal is also unknown. The lowest possible order can be assumed for model #2 and the higher model (M_H) is formed as if this assumption is valid. At that moment, N' point shown in Figure 2.1 is estimated. The approach explained for Type I problems is used to estimate N'. Now, parameters of model #2 can be estimated properly using the approach explained in Type II. After finishing parameter estimation of model #2, N' may be estimated using the Type I approach again with the improved high order model (M_H). Now, the higher order model can include more parameters from model #2 after estimation of model #2.

As an alternative way, parameters of model #2 can also be estimated prior to the application of Type I approach by taking observations from end of the

observation vector. Two models can be pre-estimated before the application of Type I approach.

Whole solution strategy can be expanded easily for cases where there are more than 2 different observations added consecutively in time. You can examine Sec.4.2 and Sec.4.3 for an application of F-test for this type of problem.

3.3.4   Type IV

The suggested solution for this problem is very similar to the solution for Type II problem. However different than Type II problem, the linear signal model is known prior to application of F-test . Instead of model order, analysis window for the preselected model should be found.

The linear model given in (2.33) is used as the lower order model (M) for F-test. A linear model which covers the model given in (2.33) and approximates s better than x_1, i.e., lowers the error defined in (2.34) is used as the higher order model (M_H). An initial value is selected for N' shown in Figure 2.3. F-test is applied for selected values. If $F < threshold$ condition is satisfied it implies that s can be approximated "good enough" as x_1 for the first N' points. N' is increased and F-test is applied for new analysis window. If $F \geq threshold$ condition is satisfied it implies that x_1 does not approximate s for the first N' points "very well". Since the goal is finding maximum N' value where s can be approximated with x_1 "well enough", it is decided that previously tested window is suitable for that purpose. Previously tested N' value is the largest possible value that makes $F < threshold$ condition possible. Finally, parameters given in (2.33) can be estimated by using the selected observations.

This problem can be thought also as observation of a signal in a linear space under not only AWGN but also AWGN plus a bias given in (2.34). Depending on the signal, noise and bias levels, assuming that the signal is observed under AWGN and ignoring bias may be reasonable at the expense of parameter estimation accuracy. You can examine Sec.4.1 for an application of F-test for this type of problem.

In the next chapter, these suggested methods will be supported by different example problems.

CHAPTER 4

APPLICATION EXAMPLES

In this chapter, the F-test based solution methods for five different example problems are demonstrated. Since the Type I problem is actually part of the Type III, an example for the Type I problem is not given separately.

4.1 Zero-Crossing Point Estimation (Problem Type: IV)

This problem is about estimation of zero-crossing point of a continuous function from its noisy observations in discrete time. This is the first example problem examined during thesis work and it represents the behaviour of F-test clearly. This example was also briefly mentioned in a previous work [35].

Here, t_0 values which satisfy the $0 = f(t_0)$ condition are called **zero-crossing point**s of a real function $f(t)$. For example, a line function has a single zero-crossing point but higher order polynomials may have more than one zero-crossing points or none. In this problem, it is assumed that function of interest has at least one zero-crossing point within a known interval.

Firstly, let us consider the zero-crossing point estimation problem particularly for a line function. Continuous time expression of a line is given as

$$x(t) = at + b = a(t - t_0),$$

where

$$t_0 = -\frac{b}{a}.$$

45

Continuous time signal can be expressed in discrete time as

$$x[n] = an + b = a(n - n_0).$$

This signal is sampled at t values which satisfies the

$$n = t \qquad \text{where } n + 0.5 \text{ is an integer}$$

rule. This rule is selected to simplify the following equations. Therefore, the

$$n_0 \triangleq t_0 \tag{4.1}$$

equality becomes valid.

Similarly, the

$$-0.5 \leq n_0 \leq 0.5 \tag{4.2}$$

inequality is assumed to provide relatively easy calculations.

In continuous time it is assumed that the signal is observed under AWGN as

$$y(t) = x(t) + w(t) = a(t - t_0) + w(t).$$

Here, $y(t)$ and $w(t)$ represents the observed signal and noise, respectively. It also can be written as

$$y[n] = x[n] + w[n] = a(n - n_0) + w[n] \tag{4.3}$$

in discrete time.

n_0 should be estimated as $\hat{n}_0$ by using $y[n]$. Considering the equality given in (4.1), the $\hat{t}_0 = \hat{n}_0$ equality holds. The number of available observations is denoted by N.

Two assumptions given as

$$n_N = -n_1, \tag{4.4}$$

$$n_{i+1} = n_i + 1 \qquad \text{where} \quad 0 < i \leq N \tag{4.5}$$

are made to simplify calculations.

The $\boldsymbol{y}$ vector defined as

$$\boldsymbol{y}_{N\times 1} \triangleq \begin{bmatrix} y[n_1] \\ y[n_2] \\ \vdots \\ y[n_N] \end{bmatrix}$$

is the observation vector. The vectors $\boldsymbol{x}$ and $\boldsymbol{w}$ can be defined similarly. Each w_i term in the $\boldsymbol{w}$ vector is an independent random variable with $\sim N(0, \sigma^2)$ distribution.

Equation (4.3) can be written as

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{\omega} = \boldsymbol{A}\boldsymbol{p} + \boldsymbol{\omega}$$

using vector notation. Also $\boldsymbol{A}$ matrix and $\boldsymbol{p}$ vectors can be represented for this problem as follows

$$\boldsymbol{A}_{N\times 2} \triangleq \begin{bmatrix} n_1 & 1 \\ n_2 & 1 \\ \vdots & \vdots \\ n_N & 1 \end{bmatrix},$$

$$\boldsymbol{p}_{2\times 1} \triangleq \begin{bmatrix} a \\ b \end{bmatrix}.$$

After LS solution is applied, $\hat{\boldsymbol{p}}$ is found. $\hat{n}_0$ can be found as

$$\hat{n}_0 = -\frac{\hat{b}}{\hat{a}}.$$

The Cramér–Rao Lower Bound for $\hat{n}_0$ is given in as

$$\text{CRLB} = \frac{\sigma^2}{Na^2}\left(1 + \frac{12n_0^2}{N^2 - 1}\right).$$

Derivation of the CRLB expression is given in Appendix A. From this expression it can be seen that the CRLB of $\hat{n}_0$ which shows the minimum possible variance of $\hat{n}_0$ depends on many problem variables. One of them is N, the number of observations. As N increases, the CRLB decreases. This is also an intuitive result. An increase in the number of observations decreases the effect of noise on estimation. It should be noted that this is true since there is not any model mismatch.

47

Up to that point, estimation of the zero-crossing point of a line function observed under AWGN was studied. Same concepts will be given for sinusoidal signals with the help of Figure 4.1. General expression of a continuous time sinusoidal signal is given in as

$$s(t) = A \sin 2\pi f(t - t_0). \tag{4.6}$$

A and t_0 terms shown in the equation (4.6) represent amplitude and zero-crossing point of sinusoidal signal, respectively. An example signal is shown in Figure 4.1 when $A = \sqrt{2}$ and $t_0 = 0.3$.
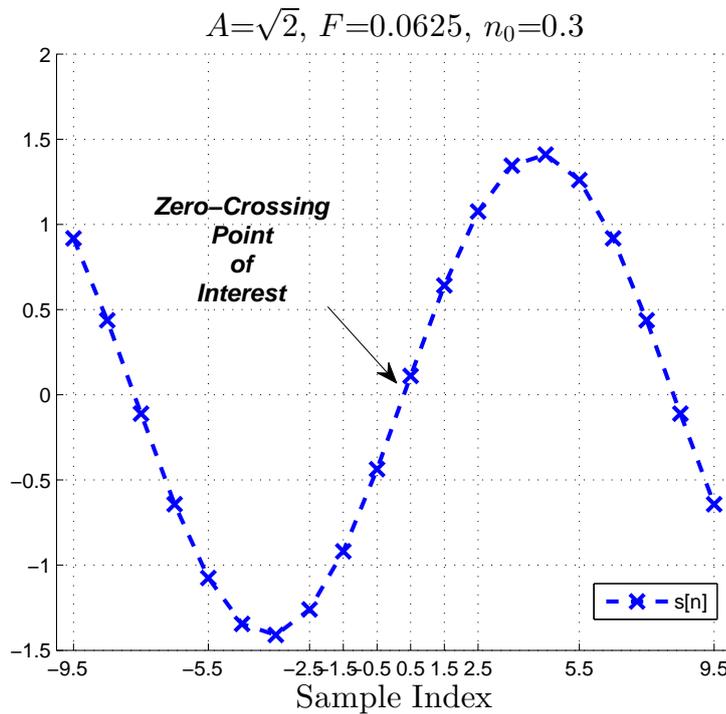
Figure 4.1: An Example Sinusoidal Signal

Similar to line function case, sinusoidal signal is also observed under AWGN and sampled at f_s frequency. For discrete time indices, in addition to the relations given in (4.4), (4.5),

$$n = t \times f_s \qquad \text{where } n + 0.5 \text{ is an integer}$$

is valid. Discrete time signal can be written as

$$s[n] = A \sin \left(2\pi F(n - n_0)\right)$$

while the inequality given in equation (4.2) is still assumed.

48

Using the previously given definitions, the followings can be written

$$n_0 = t_0 \times f_s,$$

$$F \triangleq \frac{f}{f_s},$$

$$\Omega \triangleq 2\pi F = 2\pi \frac{f}{f_s},$$

$$y[n] = s[n] + w[n].$$

As an example, observation signal $(y[n])$ for SNR = 20 dB case and the actual signal $(s[n])$ is shown in Figure 4.2. Similar to line function case, zero-crossing point of the sinusoidal signal shown in Figure 4.2 is to be estimated.
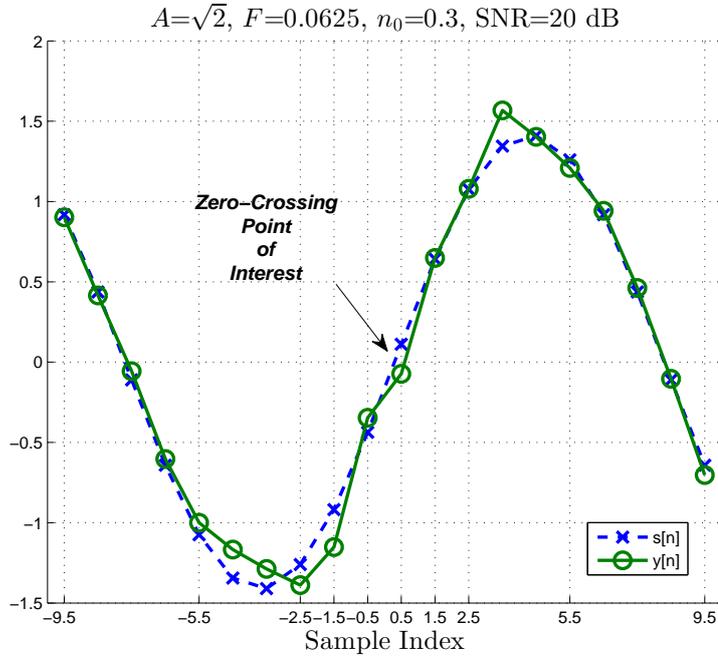
Figure 4.2: Noisy and Noiseless Sinusoidal Signals

One possible approach for finding zero-crossing point of a sinusoidal signal is to fit a straight line to the two samples with different signs around the zero-crossing point [24].Then, the zero-crossing point of the fit line is used as estimate of the zero-crossing point of the sinusoidal signal. This approach assumes that the signal is sampled "fast" enough such that a simple line fit is sufficient to accurately determine the zero-crossing point. Validity of this approach can be

seen from the Taylor series expansion given as [32]

$$\sin x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1}. \tag{4.7}$$

Equation (4.7) can be rearranged as

$$\sin \Omega n = \Omega n + \sum_{k=1}^{\infty} \frac{(-1)^k}{(2k+1)!} (\Omega n)^{2k+1} \tag{4.8}$$

for discrete time signals. Ωn term in (4.8) shows the line component of the Taylor series and the remaining terms represents the higher order polynomials. The line component is represented as

$$l(\Omega n) \triangleq \Omega n$$

and the approximation error is given as

$$\epsilon(\Omega n) \triangleq \sin(\Omega n) - l(\Omega n).$$

$\mathrm{RMSE}(\Omega, N)$ term defined as

$$\mathrm{RMSE}(\Omega, N) \triangleq \sqrt{\frac{1}{N} \sum_{n=-N/2}^{N/2} \left[\epsilon' \left(\Omega \left[n - \frac{1}{2} \right] \right) \right]^2} \tag{4.9}$$

is the root-mean-square of approximation error. Plot of $l(\Omega n)$ ve $\epsilon(\Omega n)$ for the sinusoidal shown in Figure 4.1 are shown in Figure 4.3. $\mathrm{RMSE}(\Omega, N)$ is shown in Figure 4.4 for the same signal.

Suppose that one decides to use line fitting approach to estimate the zero-crossing point of a sinusoidal signal. Figure 4.3 and 4.4 shows that $|\epsilon(\Omega n)|$, which is the absolute value of approximation error, increases as n increases. It is clear that line approximation introduces an estimation error even for noise-less case. This error is caused by model mismatch between the line and the sinusoidal signal and it will be shown up in the zero-crossing estimates as bias.

This problem fits to the scenario given for Type IV problems. N' shown in Figure 2.3 should be estimated first. Then the first N' points of observation vector are considered as observations from a line function under AWGN. A

50
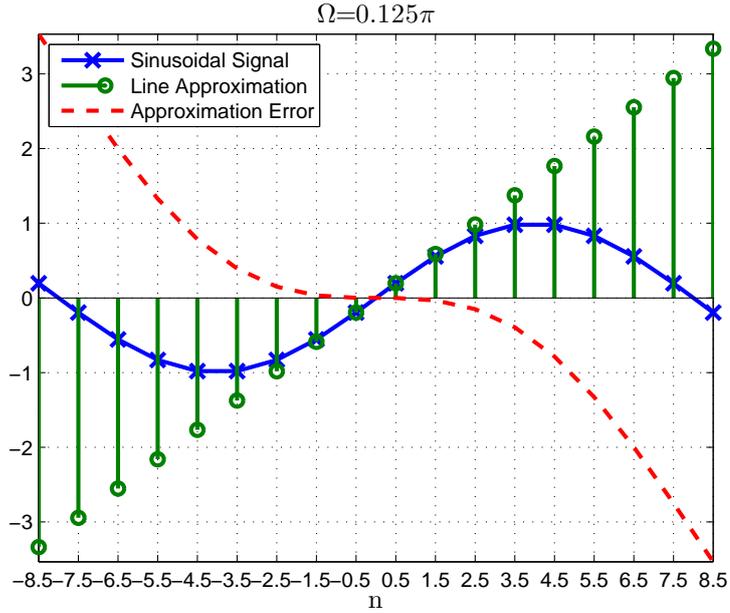
Figure 4.3: Sinusoidal Signal, Line Component and Approximation Error
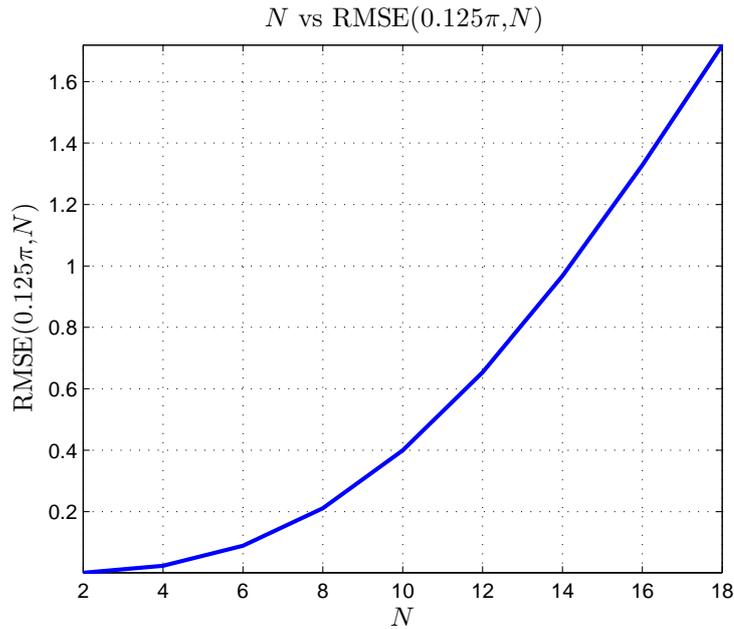
Figure 4.4: RMSE due to Line Approximation for Different Window Lengths

straight line is fitted to these observations and the zero-crossing point of the fitted line is considered as $\hat{n}_0$.

This N' value may be chosen without the proposed method by using only frequency information. In that problem, the frequency is assumed to be unknown.

If discrete time frequency (F) is low, signal can be considered to be sampled "fast". In that case more than 2 points around the zero-crossing point should be used for estimation to increase the accuracy since as frequency decreases more observations tend to lie on a line, approximately. When the frequency is high, less number of observations should be used for line fitting in order to keep the model mismatch at a reasonable level. The question is that how many samples should be used around the zero-crossing point to fit a line? This problem is illustrated in Figure 4.5.
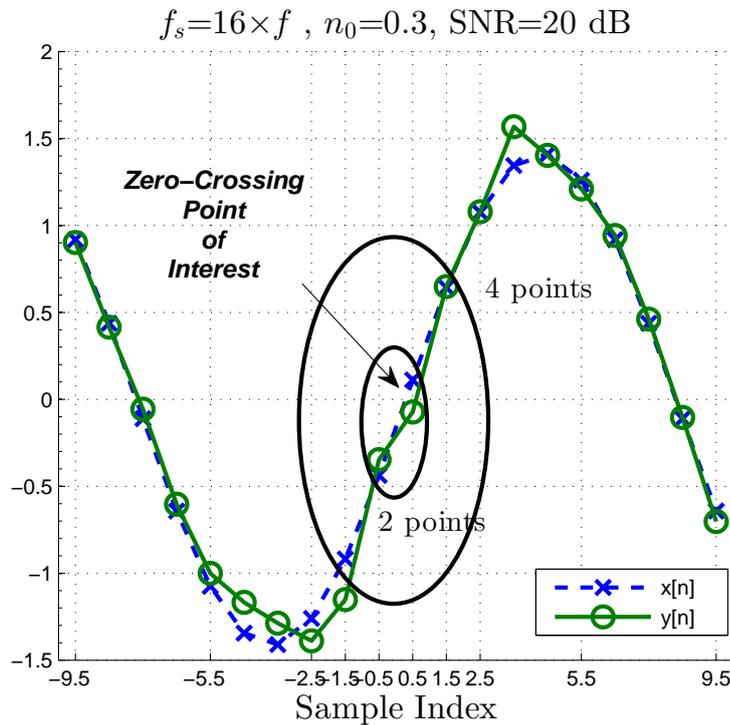
Figure 4.5: Illustration of Analysis Window Length Selection

As explained before, the F-test based approach for Type IV problem needs two nested models with order M and M_H. For that problem, $M = 2$ and the lower order model is a line model. The higher order model is chosen as a polynomial function with degree 3 which makes $M_H = 4$. The reason behind this choice is shown below. Taylor series expansion given as

$$\sin \Omega n = \Omega n - \frac{1}{6}(\Omega n)^3 + \sum_{k=2}^{\infty} \frac{(-1)^k}{(2k+1)!}(\Omega n)^{2k+1}$$

shows that the lowest order term after Ωn term is $(\Omega n)^3$ term.

The solution approach defined for Type IV problems is used with minor modifications. In order to get symmetrical points around the zero-crossing point, N' is increased by 2 in each step. Since $M_H = 4$, the initial value of N' should be greater than 4 and to keep symmetry around the zero-crossing point it is taken as 6. Suppose that $F < threshold$ condition is true for the first run. It means that a polynomial with order 1 (line) fits first 6 points sufficiently "good" as a polynomial with order 3. Then the same test is done by using 8 points around the zero-crossing point. If $F < threshold$ condition is still satisfied, test is repeated for 10 points. Suppose that $F \geq threshold$ condition becomes true when test is done using 10 points. In that case, the higher model fits 10 observations "significantly" better than the lower model. Approximating 10 points around the zero-crossing point as a single line is not a good idea. Therefore, 8 points should be used for line fitting and the zero crossing point of this line can be taken as $\hat{n}_0$.

Suppose that $F \geq threshold$ condition is satisfied when $N' = 6$. In that case a line is fit to 4 points around the zero-crossing point. Especially for higher frequencies, the best estimation results are obtained when the only 2 points are used for line fitting. Using 4 points instead of 2 worsens the estimation performance. As a minor modification, line fit is done by taking 2 not 4 points when $F \geq threshold$ and $N' = 6$ conditions are satisfied. By doing so, the proposed method may give estimates as accurate as the classical 2 points approach especially for high frequencies. Disadvantages of this modification is that for some medium frequencies usage of 4 points may be a better choice than usage of 2 points. However with this modification, the proposed method never selects 4 points.

Depending on SNR and frequency, F may always be smaller than $threshold$ for all N' values for a specific observation. One may want to limit the number of points used by line estimation independent from the F-test result. A variable called $Nusedmax$ is defined for that purpose. If the proposed method can't decide a proper N' value until it reaches $Nusedmax$, test is terminated and $Nusedmax$ points are used for estimation. This may become useful if there is a priori information about the maximum value of frequency of signal of interests.

Independent from the F-test result, it may be used to avoid the cases where the points around negative and positive peaks of the sinusoidal are treated as lying on a straight line. Flow chart of the modified approach is shown in Figure 4.6.
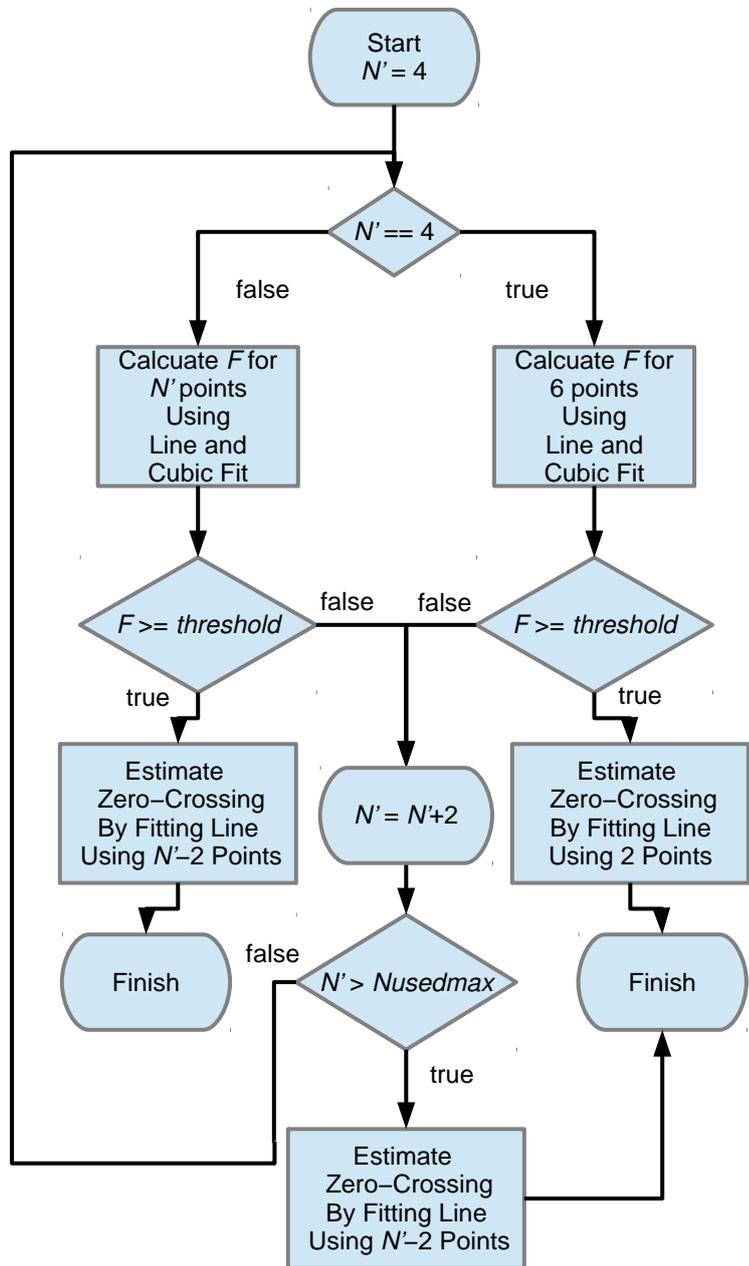
Figure 4.6: F-test Based Proposed Algorithm for Analysis Window Length Selection

4.1.1 Simulation Results

Comparison between the classical 2-points approach and the F-test based approach is given for different scenarios with help of Monte Carlo simulation results. Monte Carlo run number is denoted by MC_{num}. MC_{num} is taken as 10^5 for each scenario. RMSE shown in the following figures is calculated using the

$$\text{RMSE} = \sqrt{\frac{1}{MC_{num}} \sum_{i=1}^{MC_{num}} (n_0 - \hat{n}_{0_i})^2}$$

relation where $\hat{n}_{0_i}$ denotes the estimation for i^{th} run. Notice that this RMSE definition is different from the definition given in (4.9). $Nusedmax$ is taken as 16 for all scenarios.

4.1.1.1 Scenario I

Figure 4.7 shows RMSE values of both approaches for different frequencies. For each frequency value, new Monte Carlo simulation is run. $p_{fd} = 0.1$ is taken for this scenario.
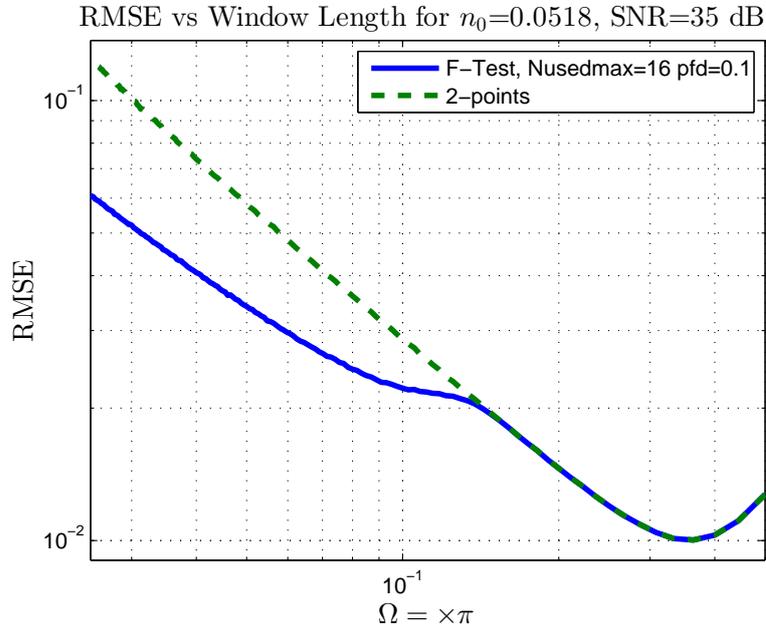
Figure 4.7: Error Comparison Between The Classical and The Proposed Method

Two methods perform the same for high frequencies. This is because the pro-

posed method always choose to use 2 points for line fitting at these frequencies. Figure 4.9 shows the histogram of lengths (N') of selected windows when $\Omega = 0.5\pi$ rad. This histogram shows that the proposed method always takes $N' = 2$.
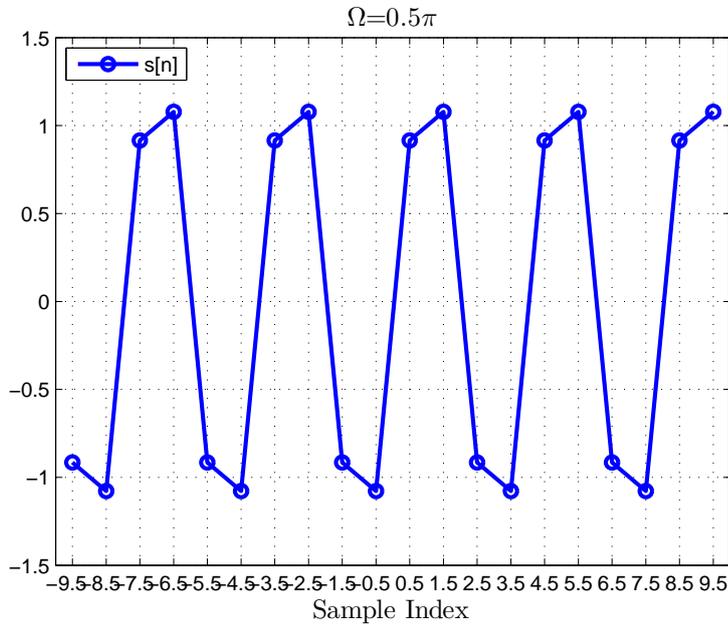
Figure 4.8: $s[n]$ when $\Omega = 0.5\pi$
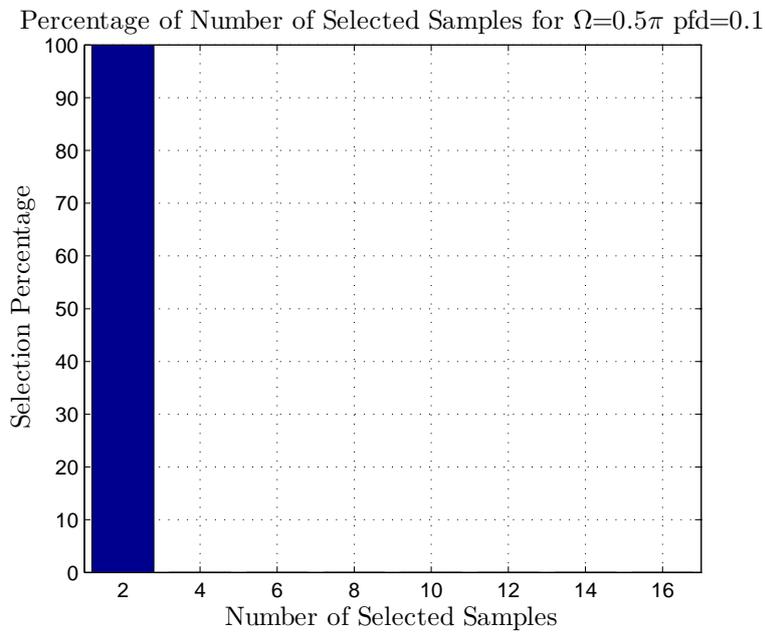
Figure 4.9: Percentage of The Number of Selected Samples by The Proposed Method

This and the following simulation results are obtained when SNR = 35 dB. However, one of the simulations is repeated for $\Omega = 0.5\pi$ rad case when SNR = 0 dB to show the effect of SNR on the performance. As it can be seen from Figure 4.10, F-test fails to select correct number of points at low SNR. For this frequency, 2-points should be selected for line approximation as it can be seen from Figure 4.8. As it is stated in Chapter 3, SNR value should be relatively high for proper operation of the F-test.
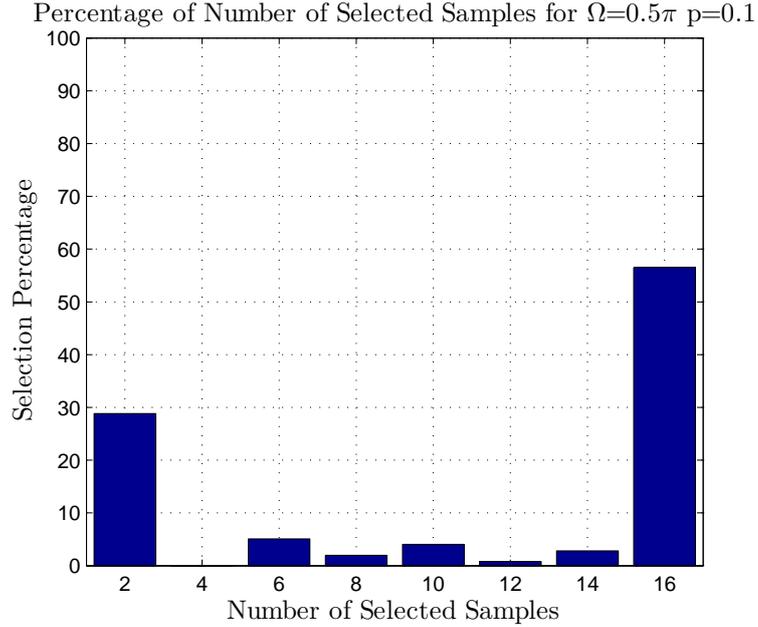
Figure 4.10: Percentage of The Number of Selected Samples by The Proposed Method when SNR = 0 dB

When $\Omega = 0.125\pi$ rad, the proposed method selects $N' = 6$ for more than 10% of experiments as shown in Figure 4.12.

When $\Omega = 0.08\pi$ rad, the proposed method starts to select $N' = 8$ in addition to 6 points as shown in Figure 4.14. This provides lower RMSE values compared to the classical 2-points approach.

When $\Omega = 0.0025\pi$ rad, the proposed method selects 16 points most of the time as it can be seen from Figure 4.16. Notice that *Nusedmax* is also 16 for that problem.
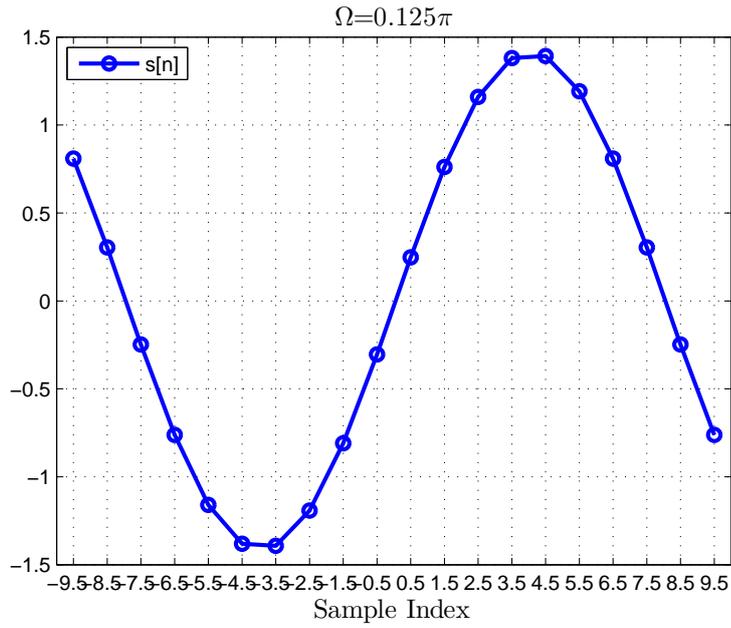
Figure 4.11: $s[n]$ when $\Omega = 0.125\pi$
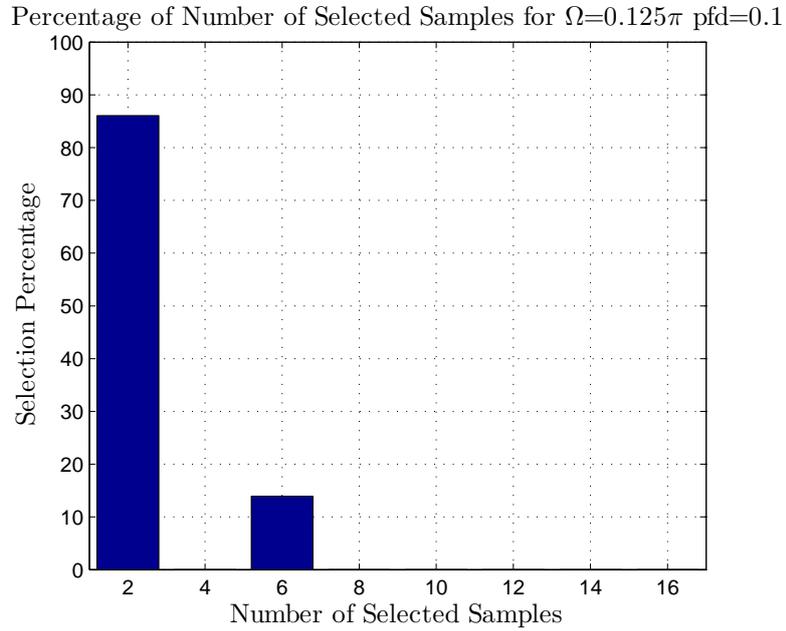
Figure 4.12: Percentage of The Number of Selected Samples by The Proposed Method
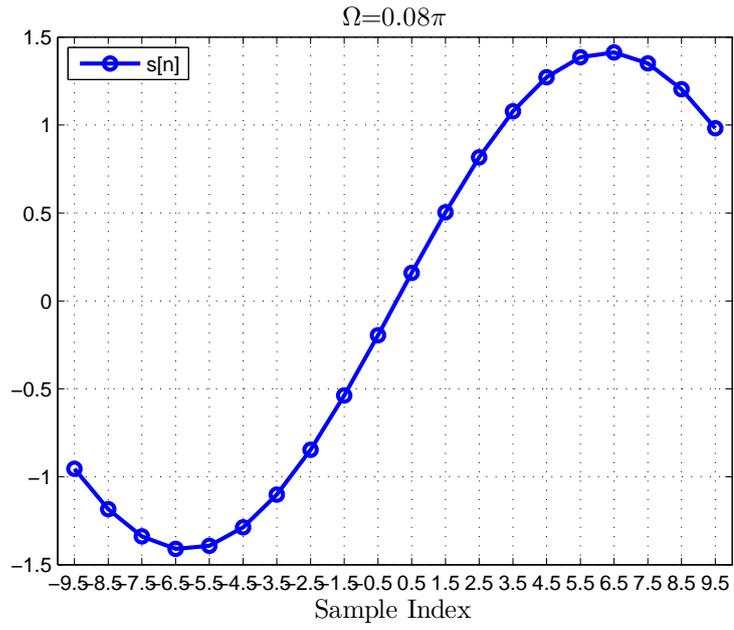
Figure 4.13: $x[n]$ when $\Omega = 0.08\pi$
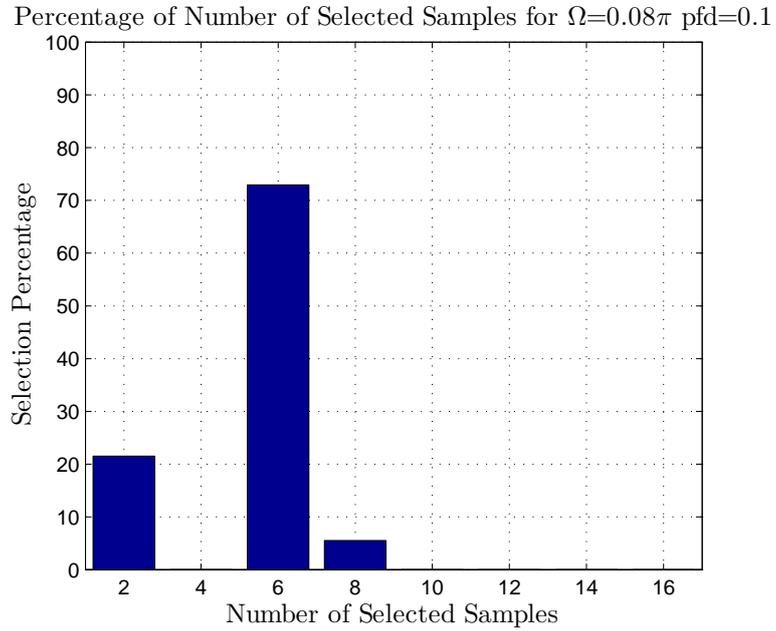
Figure 4.14: Percentage of The Number of Selected Samples by The Proposed Method
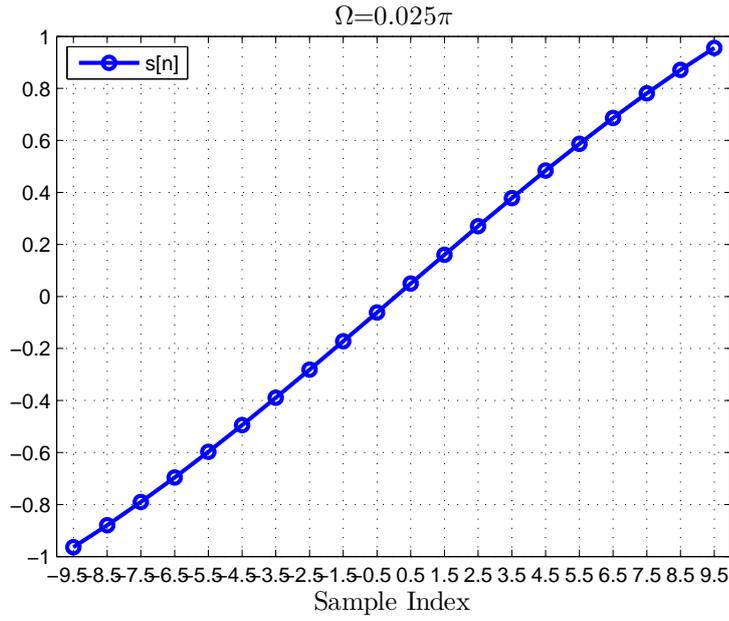
Figure 4.15: $s[n]$ when $\Omega = 0.025\pi$
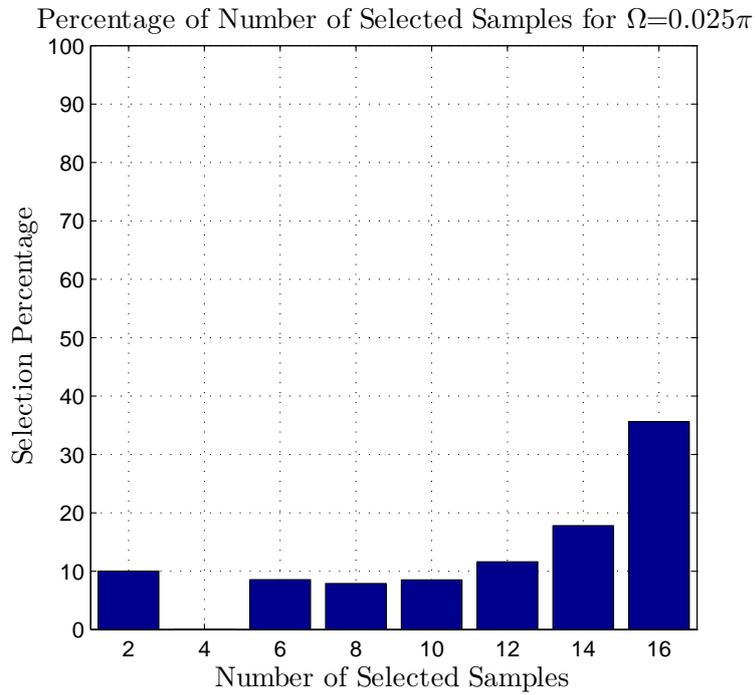
Figure 4.16: Percentage of The Number of Selected Samples by The Proposed Method

4.1.1.2 Scenario II

Similar experiments are repeated for by taking $p_{fd} = 0.9$. In that case, *threshold* has lower values than the previous case. F becomes $F \geq threshold$ at lower N' values. Thus, the proposed method tends to select lower N' values. RMSE value for $p_{fd} = 0.9$ case is shown in Figure 4.17. Since the proposed method selects lower N' values, RMSE value of the proposed method is close to RMSE value of the classical 2-points method.
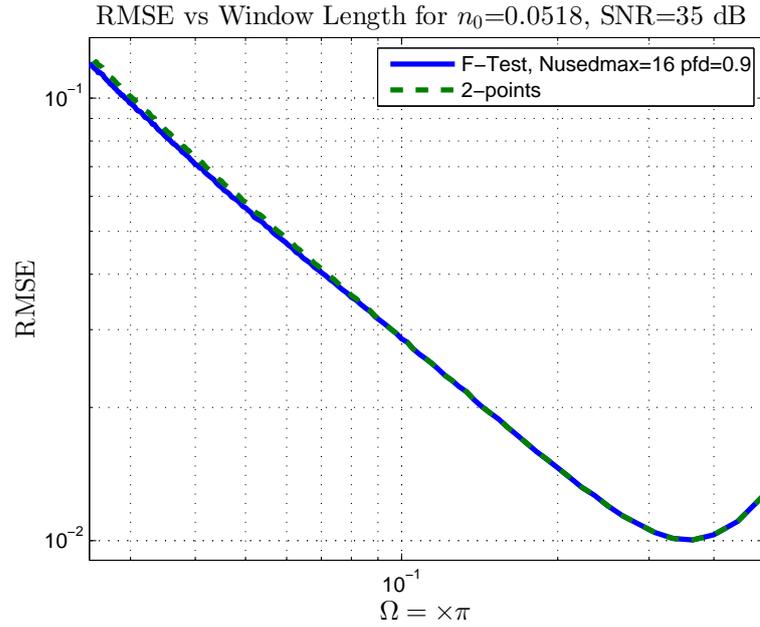
Figure 4.17: Error Comparison Between The Classical and The Proposed Method

Larger N' values are chosen for low frequencies as seen previously. Even for low frequencies, the proposed method chooses 2-points 90% of the time as seen from Figure 4.18 when $p_{fd} = 0.9$. In that case the maximum value of N' is 6 where it was 16 for the $p_{fd} = 0.1$ case.
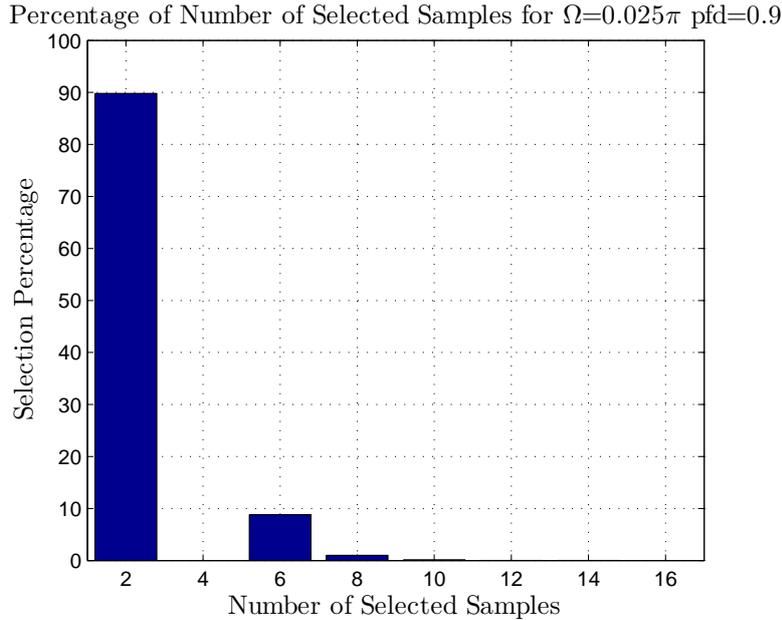
Figure 4.18: Percentage of The Number of Selected Samples by The Proposed Method

4.1.1.3 Scenario III

Lastly, simulations are repeated when $p_{fd} = 0.01$. In that case *threshold* reaches the highest value among three scenarios. As *threshold* increases N' value selected by the proposed method increases. RMSE graph is given in Figure 4.19 for $p_{fd} = 0.01$. N' value tends to increase so much that especially for high frequencies the proposed method does not recognize model mismatches and uses wrong number of points for line approximation.

This example is given to demonstrate the proposed method for the type IV problem. Also, effects of p_{fd} and SNR value on performance are demonstrated by using three different values. As stated in the previous chapter, as p_{fd} increases *threshold* deceases. "False decision" probability of F-test should not be confused with performances of the suggested algorithms. For example, 2 points are selected with $\sim 90\%$ probability when $p_{fd} = 0.9$ as shown in Figure 4.18. However, it does not mean that selection of 2 points is wrong. Value of the p_{fd} changes behaviour of the F-test. Depending on application, change in behaviour affects the overall performance differently.
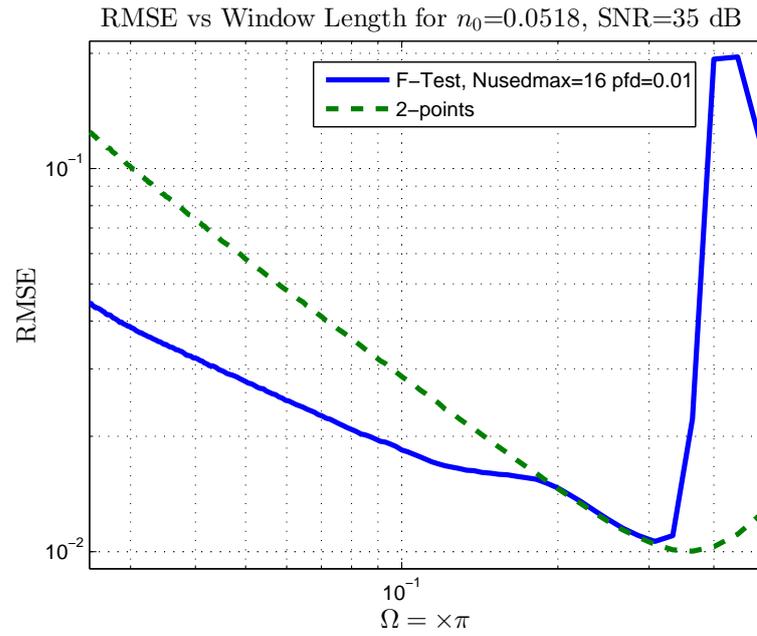
Figure 4.19: Error Comparison Between The Classical and The Proposed Method
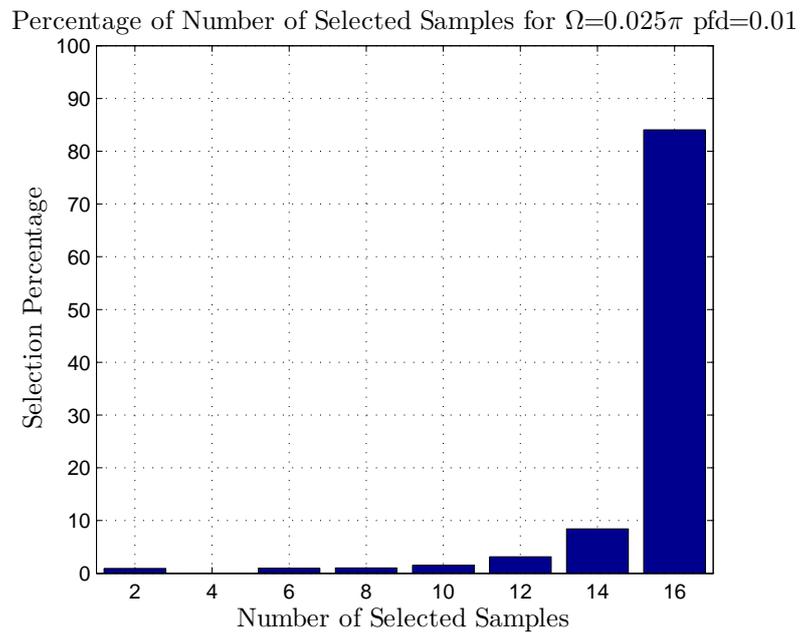
Figure 4.20: Percentage of The Number of Selected Samples by The Proposed Method

4.2 Segmentation of Multi Tone Signals (Problem Type: III)

As a demonstrative example for the type III problem, segmentation of multi tone signals is selected. A multi tone signal with length N is expressed as

$$s[n] = \begin{cases} \sum_{t=1}^{T} T(n, n_{1t}, n_{2t}, A_t, \Omega_t, \Phi_t) & 0 \leq n \leq N - 1 \\ 0 & \text{otherwise} \end{cases} \qquad (4.10)$$

where

$$0 \leq n_{1t}, n_{2t} \leq N - 1 \qquad , \forall t,$$

$$T \geq 1,$$

$$T(n, n_1, n_2, A, \Omega, \Phi) \triangleq \begin{cases} A \sin(\Omega n + \Phi) & n_1 \leq n \leq n_2 \\ 0 & \text{otherwise.} \end{cases} \qquad (4.11)$$

The signal $s[n]$ is generated by adding up tones with different or not amplitude, frequency, phase, start and stop times. One interest of signal processing area is the characterization of this kind of signals. The only constraint on $s[n]$ is that there is at least one non-zero tone component at any time instance. One may want to plot spectrograms of multi tone signals. In that problem amplitude(A), frequency(Ω), phase(Φ), start(n_1) and stop times(n_2) of signals are assumed to be unknown. The goal is to improve the frequency resolution in spectrogram plots.

Let us review some basics of spectrogram before the application of F-test. At this moment, the observation vector is assumed to be noiseless.

Spectrogram is generated by plotting magnitude of DFT of windowed data. A vector (segment) with length $Nwindow$ is taken from the beginning of the observation vector. The length of the second segment is also $Nwindow$. $Noverlap$ samples are taken from end of the first segment for second segment. In other words, segments are overlapped. This segmentation procedure is repeated as shown in Figure 4.21. We call it as **classical segmentation**. It is illustrated in . After segmentation, $NDFT$ point DFT is calculated for each segment. In practice, generally DFT is calculated using FFT. Magnitude plots of DFTs are

placed side by side in time to generate spectrogram plot. Overlapping provides smooth transitions in spectrogram plot.
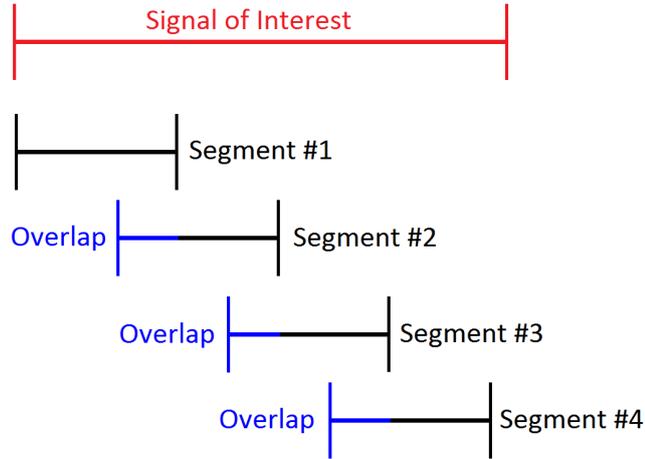
Figure 4.21: Illustration of The Classical Segmentation

As an example, let us consider $s[n]$ generated by using (4.10), (4.11) and the parameters given in Table 4.1. Spectrogram of this signal is shown in Figure 4.22. This is generated by taking $Nwindow = 1024$, $NDFT = 1024$ and $Noverlap = 512$. It can be seen that there are tone changes around 20000 and 25000 at time axis.

Table 4.1: Parameters of $s[n]$, $N = 30000$

Tone # / Parameter	n_1	n_2	A	Ω	Φ
1	0	20000	1	0.3π	0
2	20000	30000	1	0.2π	0
3	25000	30000	1	0.23π	0

Figure 4.23 shows spectrogram of the same signal with different spectrogram parameters: $Nwindow = 64$, $NDFT = 64$ ve $Noverlap = 32$. Unlike the previous spectrogram shown in Figure 4.22, it is not possible to distinguish existence of two tones after 25000th sample visually. Similarly, the plot is spread over frequency axis for all tones. Although an accurate frequency estimation is much more difficult than the previous case, the tone transition points are much more clear than Figure 4.22.

Value of $Nwindow$ parameter has direct impact on spectrogram plots. Basics of DTFT should be analyzed to understand the differences between two plots. DFT

65

Spectrogram in dB Scale NDFT=1024 Nwindow=1024 Noverlap=512
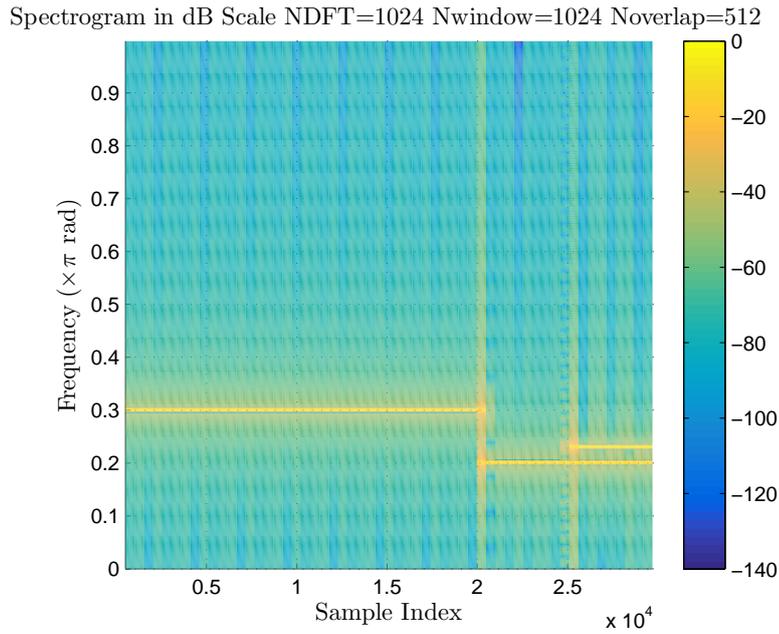
Figure 4.22: Spectrogram of Example Signal with The Classical Approach Using Long Segments

Spectrogram in dB Scale NDFT=64 Nwindow=64 Noverlap=32
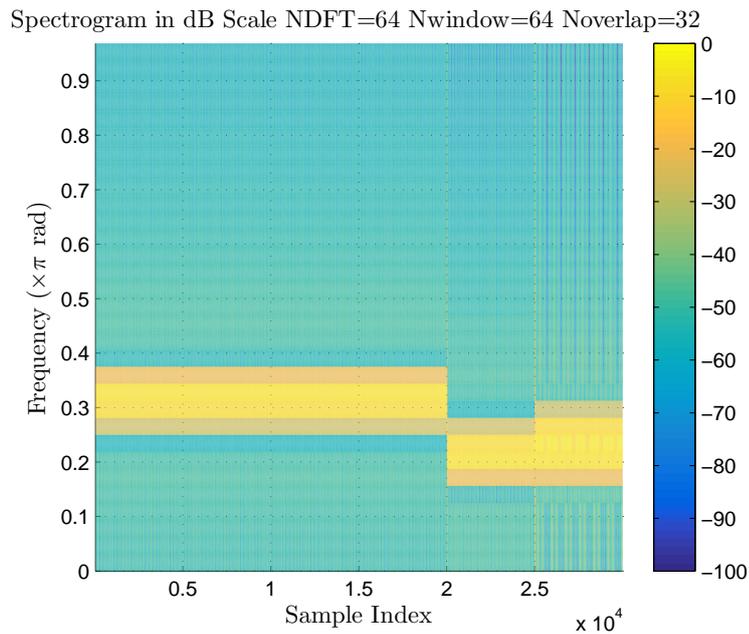
Figure 4.23: Spectrogram of Example Signal with The Classical Approach Using Short Segments

can be thought as sampled version of DTFT between 0 and 2π rad frequencies with $2\pi/NDFT$ rad steps [2]. As it will be explained later, main reason behind the differences between two spectrograms is windowing effect not the relations

66

between DTFT and DFT functions. Therefore, explanations are given by using concepts related with DTFT.

DTFT of an infinite length signal $x[n]$ is found as

$$S(\Omega) = \sum_{n=-\infty}^{\infty} s[n]e^{-j\Omega n}.$$

Let us consider an infinite length sinusoidal signal given as

$$s[n] = \cos(\Omega_s n) \qquad -\pi \leq \Omega_s \leq \pi$$

and its DTFT can be written as

$$S(\Omega) = \pi\delta(\Omega - \Omega_s) + \pi\delta(\Omega + \Omega_s) \qquad -\pi \leq \Omega \leq \pi.$$

The classical segmentation for spectrogram generation can be expressed mathematically as

$$s'[n] = \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} s[n] = s[n]w[n],$$

$$w[n] \triangleq \begin{cases} 1 & |n| \leq \dfrac{N-1}{2} \\ 0 & \text{otherwise} \end{cases} \qquad (4.12)$$

where $s'[n]$ represents one of the segments with length $N+1$.

Using DTFT properties [2] $S'(\Omega)$ can be written as

$$S'(\Omega) = \sum_{n=-\infty}^{\infty} s'[n]e^{-j\Omega n} = S(\Omega) * W(\Omega) = \frac{1}{2\pi} \int_{-\pi}^{\pi} S(\Omega - \lambda)W(\lambda)\mathrm{d}\lambda. \quad (4.13)$$

Equation given in (4.12) is known as rectangular window function and its DTFT given as

$$W(\Omega) = \frac{\sin\left(N\dfrac{\Omega}{2}\right)}{\sin\left(\dfrac{\Omega}{2}\right)} \qquad \text{if } N \text{ is odd}$$

is a well known function [25].

Although this equation is valid when N is an odd number, an extra phase term is added when N is an even number. Since phase of DTFT is discarded to plot

spectrogram and only magnitude of DTFT is used, DTFT for odd number case is considered in order to get rid of extra phase term to simplify expressions [25].

Plot of $W(\Omega)$ for $N = 17$ case is shown in Figure 4.24. $W(\Omega)$ has null points at $\Omega = 2\pi/N$ frequencies. Main lobe width of the rectangular window function is $4\pi/N$.
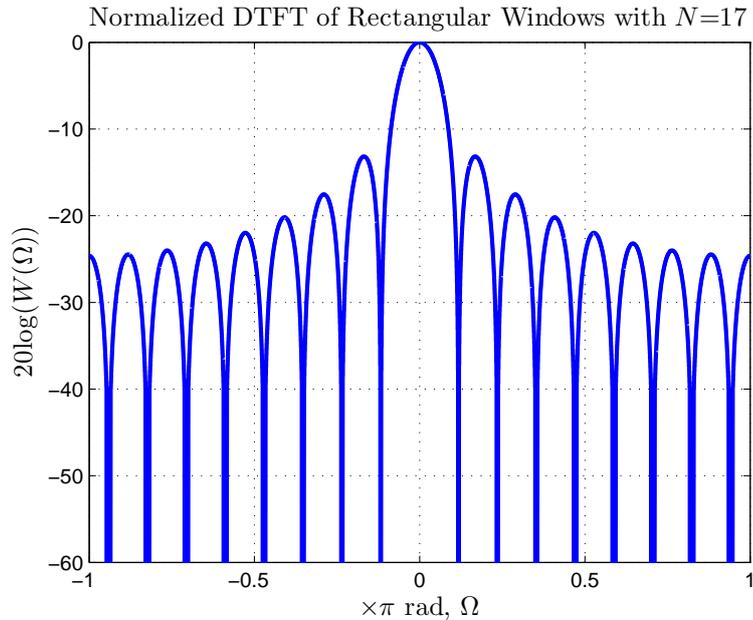
Figure 4.24: Normalized DTFT of Rectangular Window with Length 17

Equation (4.13) can be continued as

$$S(\Omega) = \frac{1}{2}\left[W(\Omega - \Omega_x) + W(\Omega + \Omega_x)\right]$$

and as it can be seen from this equation, DTFT of a finite length sinusoidal is equivalent to shifting DTFT of the window function to the frequency of sinusoidal signal which is known as modulation property. Magnitudes of DTFT of rectangular windows with $N = 64$ and $N = 1024$ are plotted in 4.25. As it can be seen from the figure, as N increases main lobe width decreases. If it assumed that $NDFT$ is fixed, energy of sinusodial signal spreads over less number of DFT points as N increases. This improves both frequency estimation accuracy for single tone case and frequency clearance between tones for multi tone case.

The problem which is defined by parameters given in Table 4.1 contains single tone up to 20000th point. As it can be seen from Figure 4.22 and 4.23, frequency
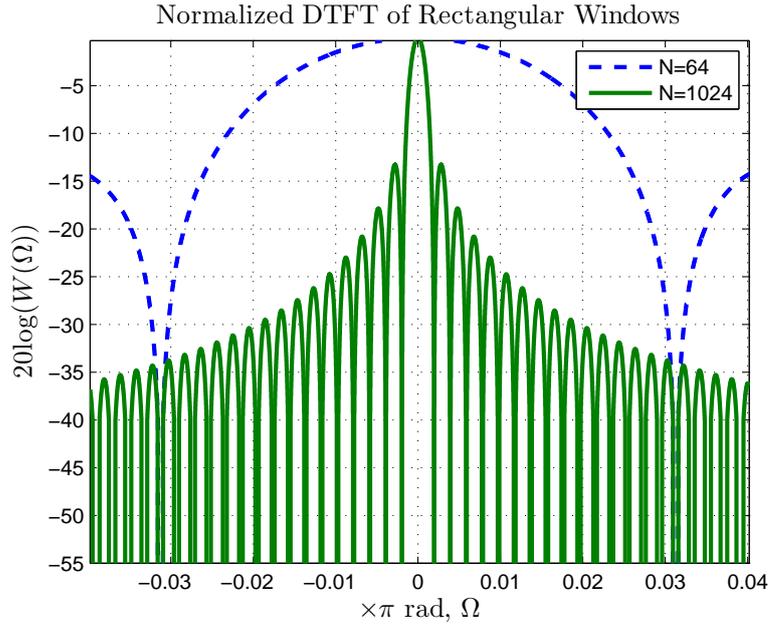
Figure 4.25: Effect of Window Length on DTFT Results

resolution increases as $Nwindow$ increases. Different from Figure 4.25, there is an extra parameter involved in spectrogram generation: $NDFT$. However, this is not the reason of the mentioned difference. This parameter only changes the sampling interval of DTFT. As $NDFT$ increases, sampling interval in the frequency axis decreases. However, the difference shown in Figure 4.25 is purely due to DTFT properties. Sampling these DTFT functions more and more closely does not improve the result. To show this, Figure 4.23 is plotted again in Figure 4.26 by taking $NDFT = 1024$ as in Figure 4.22. The frequency resolution problem still exists even if $NDFT$ is same as the condition where frequency resolution is fine as in Figure 4.22.

Taking long segments for spectrogram generation improves the frequency resolution. However as it can be seen from Figure 4.22 and Figure 4.23 around 20000th and 25000th sample, tone transitions can't be resolved in time axis very well. On the other hand, transitions are resolved in time axis when shorter segments are used as shown in Figure 4.26. The reason behind is that when segments are long, they include higher number of samples and the distance between two adjacent segment is big. For example, for $Nwindow = 64$ and $Noverlap = 32$ case each segment contains 64 samples and the distance between each of them is 32
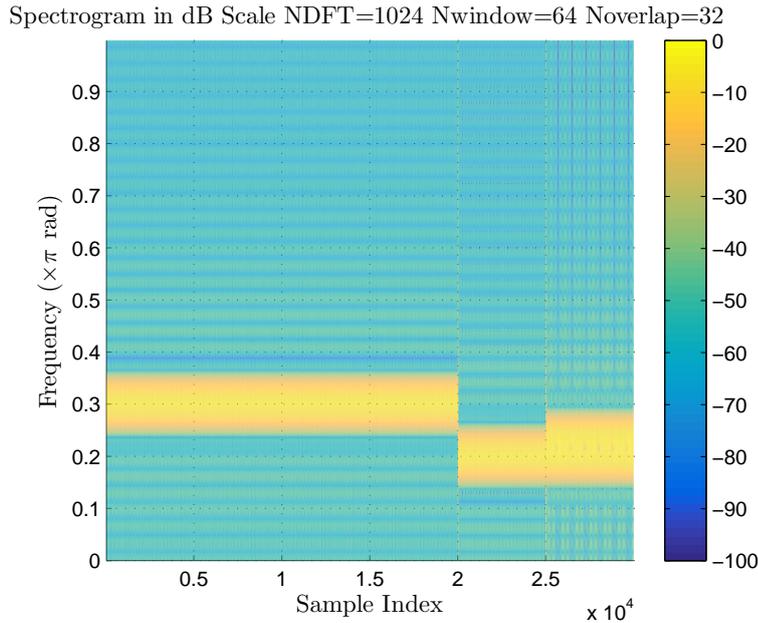
Figure 4.26: Effect of $NDFT$ Parameter on The Frequency Resolution

samples whereas for $Nwindow = 1024$ and $Noverlap = 512$ case each segment contains 1024 samples and the distance between each of them is 512 samples. Since DFT is computed for each individual segment, time resolution of the long segments is lower than the short segments. Lower time resolution causes time ambiguity around tone transition points.

In summary, long segments and short segments increase the frequency and the time resolution, respectively.

If spectrogram of the example signal was generated by using not equal segments like, first segment was taken between 0th and 20000th sample, second one was taken between 20000th and 25000th sample and the last one was taken between 25000th and 30000th, spectrogram would be like in Figure 4.27 and Figure 4.28. Notice that the colormap is different from the previous spectrograms for visual simplicity. Also, there is no overlap between these three segments. Segmentation is illustrated in Figure 4.29. We call it as **dynamic segmentation.**

If tone changes are detected and segments were taken between transition points, spectrograms have better frequency resolutions than classical spectrograms. F-test is used to estimate the tone transition points. After these points are es-
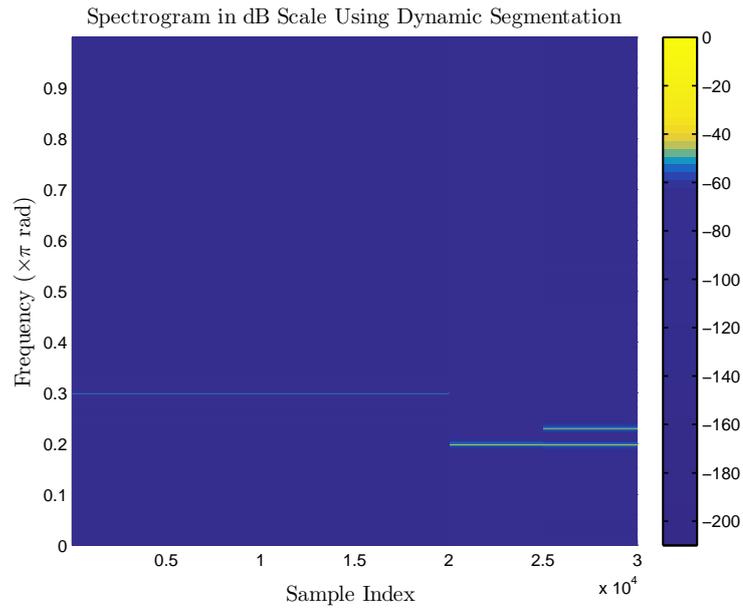
Figure 4.27: Spectrogram of Example Signal Using The Dynamic Segmentation (3 Segments), View #1
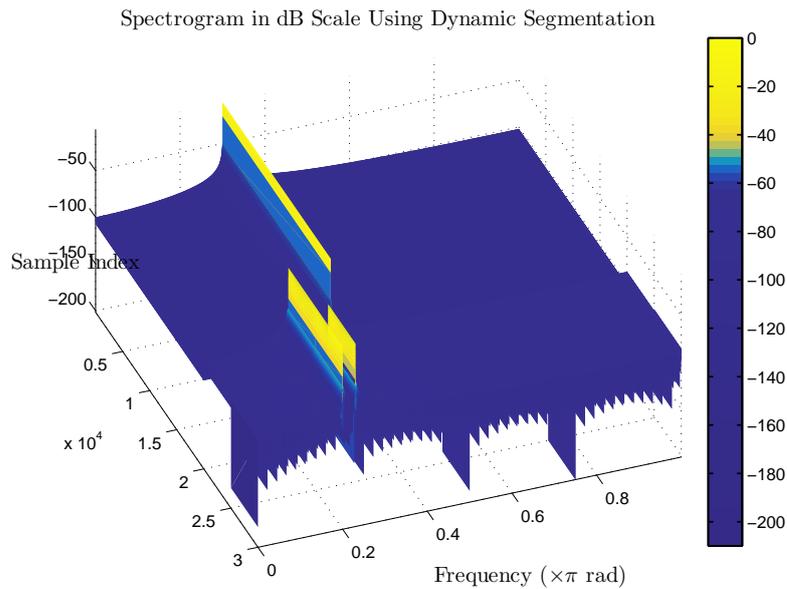
Figure 4.28: Spectrogram of Example Signal Using The Dynamic Segmentation (3 Segments), View #2

timated, (dynamic) segments and spectrogram plots are generated. Similar to the previous cases, $s[n]$ is observed under AWGN as $y[n]$.

This problem is similar to Type III problem but the solution given for Type III is modified a lot. Two nested and linear models should be determined to

71
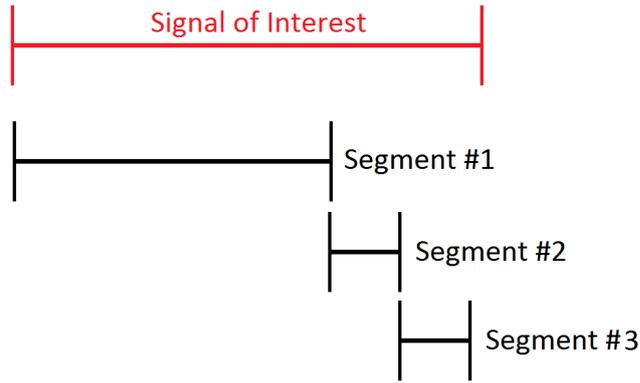
Figure 4.29: Illustration of The Dynamic Segmentation

apply F-test. However, there is no prior information about added tones, so the necessary ones should be estimated. Rather than giving F-test approach as one stage, it is divided into four stages.

First of all, proper linear models should be selected. First stage finds the model candidates. The flow is illustrated in Figure 4.30. This stage is called as model pre-selection algorithm.

Logic behind the model pre-selection algorithm is finding frequencies that have high DFT values in terms of magnitude. In that stage, a search window with length N_{sw} is selected and shifted through the observation vector. For each shift, DFT is taken and possible tones are determined. It is thought that tones with higher DFT magnitudes have higher probability of existence. If there is a priori information about frequency components, the maximum number of allowable tones can be limited in each search window. In the next section, validity of these pre-selected tones is tested.

The pre-selection algorithm uses following parameters:

- N_{sw}: *sw* is abbreviation for *Search Window*. It shows the length of search window. This length affects the frequency resolution for searching and the final segmentation performance. If search window is considered as rectangular window, two tones become separable when their frequencies are separated by more than $2\pi/N_{sw}$ rad.

- N_{sdft}: *sdft* is an abbreviation for *Search DFT*. N_{sdft} point DFT is taken

72
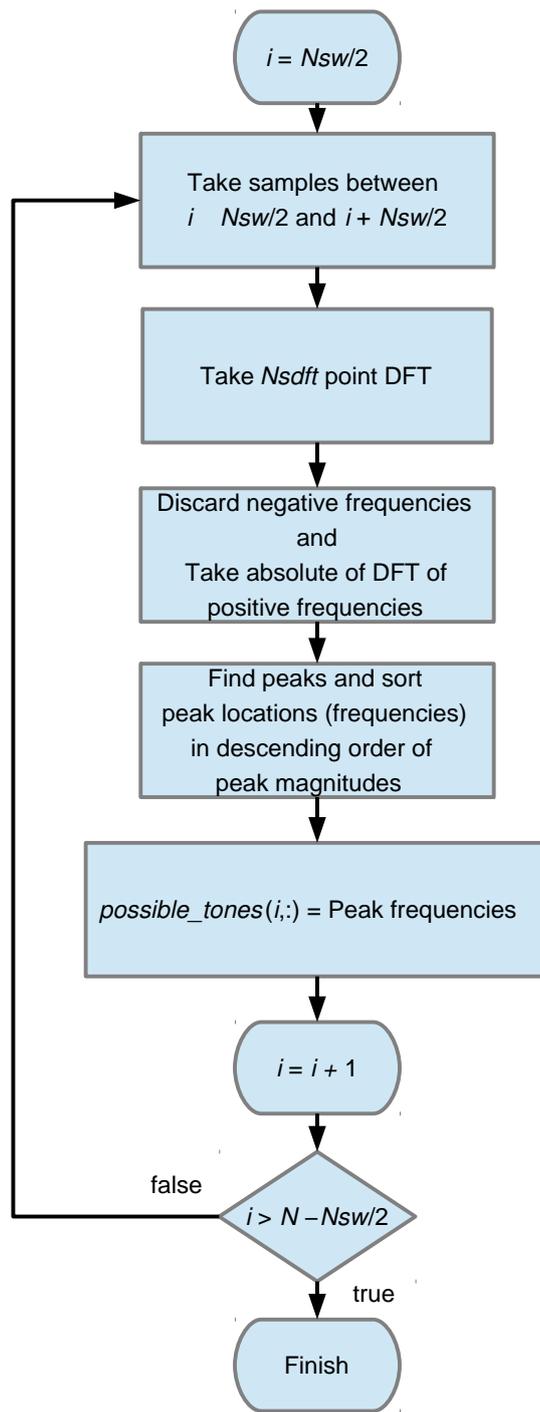
Figure 4.30: The Proposed Model Pre-Selection Algorithm

at each search window. Since DFT peaks are used for the pre-selection algorithm, at least one DFT bin should exists between two separable fre-

quencies. Therefore, this parameter should be at least double of the desired search frequency resolution between 0 and 2π.

At second stage, validity of the pre-selected tones is checked using F-test. This is similar the approach used for Type II problem. The flow is illustrated in Figure 4.31. This stage is called as model validity stage.

For each search window, the pre-selected tones are tested using the Type II approach. Consider the F-test done for the ith search window. Initially, F-test checks whether this window contains a single or dual tones. To do this, the lower order model is constructed using the frequency in $possible\_tones(i, 1)$. This is the frequency with the highest DFT magnitude for the ith search window. The higher order model is constructed using two tones with frequencies $possible\_tones(i, 1)$ and $possible\_tones(i, 2)$. $\boldsymbol{A}$ matrix for the first test is given as

$$\boldsymbol{A}_M^i = \begin{bmatrix} \boldsymbol{a}_1^i & \overline{\boldsymbol{a}_1^i} \end{bmatrix}$$

for the lower order (M) model and as

$$\boldsymbol{A}_{M_H}^i = \begin{bmatrix} \boldsymbol{a}_1^i & \overline{\boldsymbol{a}_1^i} & \boldsymbol{a}_2^i & \overline{\boldsymbol{a}_2^i} \end{bmatrix}$$

for the higher order (M_H) model. In that case $M = 2$ and $M_H = 4$. Notice that different from the previous cases, $\boldsymbol{A}$ matrices (and also $\boldsymbol{p}$ vectors) have complex terms. Although previous expressions given for real cases, replacing transpose matrices with Hermitian matrices is sufficient to apply all previous results and comments about RSS values and F ratios for complex cases.

In general, elements of $\boldsymbol{A}$ matrix can be written as

$$\boldsymbol{a}_l^i = \begin{bmatrix} a_{1,l}^i \\ a_{2,l}^i \\ \vdots \\ a_{N_{sw},l}^i \end{bmatrix}$$

where

$$\boldsymbol{a}_{k,l}^i = e^{j \times possible\_tones(i,l) \times k}.$$

74
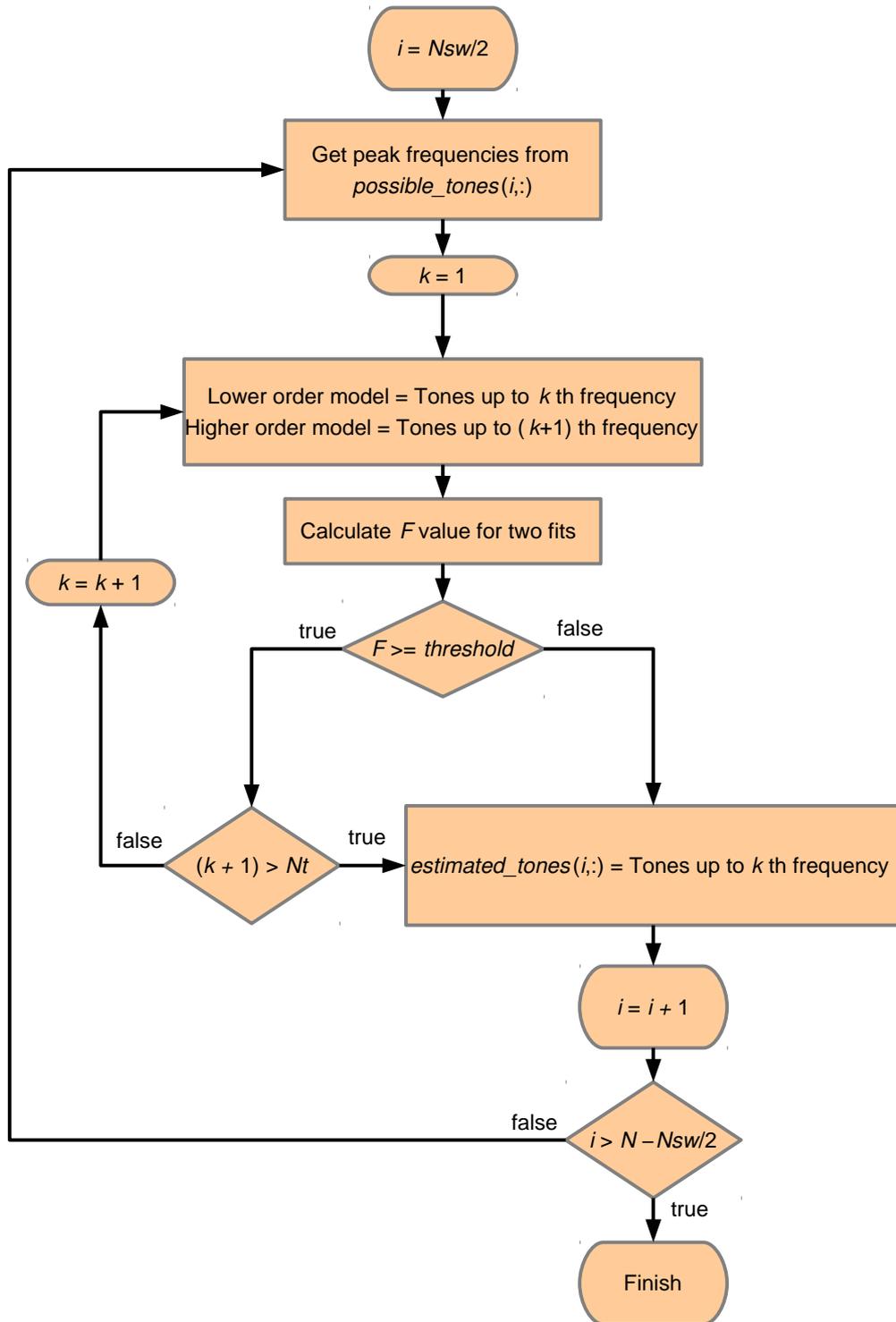
Figure 4.31: The Proposed Model Validation Algorithm

If F-test concludes that the dual tone model is "more suitable" than the single tone model for the ith search window, then the dual tone and the triple tone

models are compared. In that case, $\boldsymbol{A}$ matrix can be written as follows

$$\boldsymbol{A}_M^i = \begin{bmatrix} \boldsymbol{a}_1^i & \overline{\boldsymbol{a}_1^i} & \boldsymbol{a}_2^i & \overline{\boldsymbol{a}_2^i}, \end{bmatrix}$$

$$\boldsymbol{A}_{M_H}^i = \begin{bmatrix} \boldsymbol{a}_1^i & \overline{\boldsymbol{a}_1^i} & \boldsymbol{a}_2^i & \overline{\boldsymbol{a}_2^i} & \boldsymbol{a}_3^i & \overline{\boldsymbol{a}_3^i} \end{bmatrix}$$

for $M = 4$ and $M_H = 6$. If F-test concludes that the triple tone model is "more suitable" than the dual tone model for the ith search window, the quadruple tone model is tested against the triple tone model until the correct model order is found. It is the algorithm given for Type II problem.

The validity algorithm uses the following parameter:

- N_t, *Number of maximum tones*: It is the maximum number of tones that can be found in a search window. If F-test couldn't verify tones even if the number of tested tones reaches that parameter, number of N_t tones are assumed to exist in that search window and test is ended for that particular search window. If there is a priori information about the possible number of tones at a time instant, this information can be used for N_t selection.

After the validation stage, a post processing is done to improve the validation results. At least, F-test may give wrong decisions with p_{fd} probability. These decisions may be reduced using some a priori information about the problem if available. This stage is not related with F-test and is not subject of this thesis work. However, it is given to show that F-test results may be improved further by some post processing depending on the problem specifications.

Due to nature of F-test, it may found some tones with very short durations. For example, tone with frequency 0.1π with 5 samples duration may not be a practical case. These impractical "glitchy" results may be eliminated by utilization of some simple methods. The post-processing algorithm is illustrated in Figure 4.32.

The proposed post-processing algorithm uses the following parameter:

- N_{mp}, *Minimum Period*: A tone should be detected at least N_{mp} times of
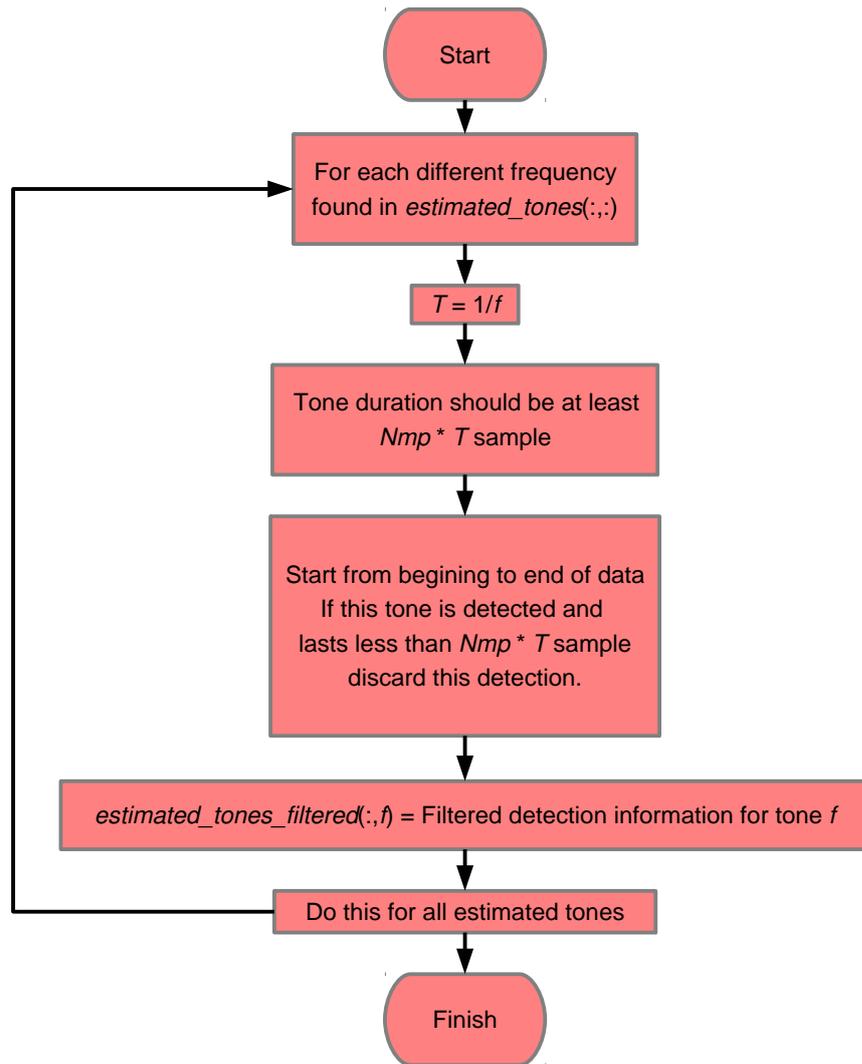
Figure 4.32: The Proposed Post-Processing Algorithm

its period to declare its existence. If there is a priori information about the signal it can be used for N_{mp} selection.

The last stage is generation of dynamic segments and resultant spectrogram using the post-processed tone profile. The flow is shown in Figure 4.33. If a tone disappears or a new tone appears, a new segment should start at that instant. After segmentation, spectrogram is plot similar to the classical case. DFT is taken for each spectrogram and plotted side by side in time. Different from the classical approach, there is no overlapped samples between segments.

Notice that the final frequency resolution of dynamic spectrogram should not be confused with N_{sw} and N_{sdft} parameters used in the model pre-selection section.

The number of DFT points is selected as length of each segment which is generally larger than the length of search window to have resolution improvements.
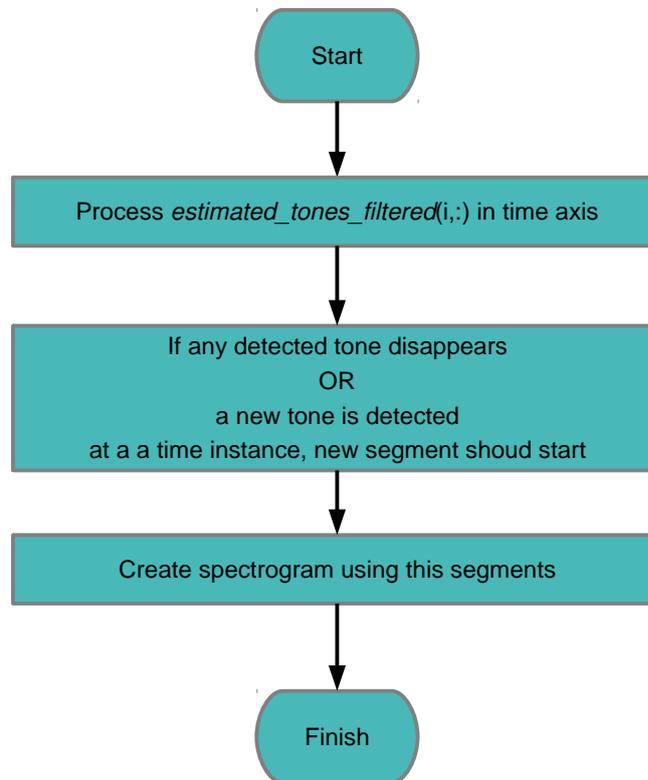
Figure 4.33: The Proposed Dynamic Spectrogram Generation

The overall proposed algorithm is shown in Figure 4.34.

Performance of the proposed method is demonstrated using example scenarios at the next section.
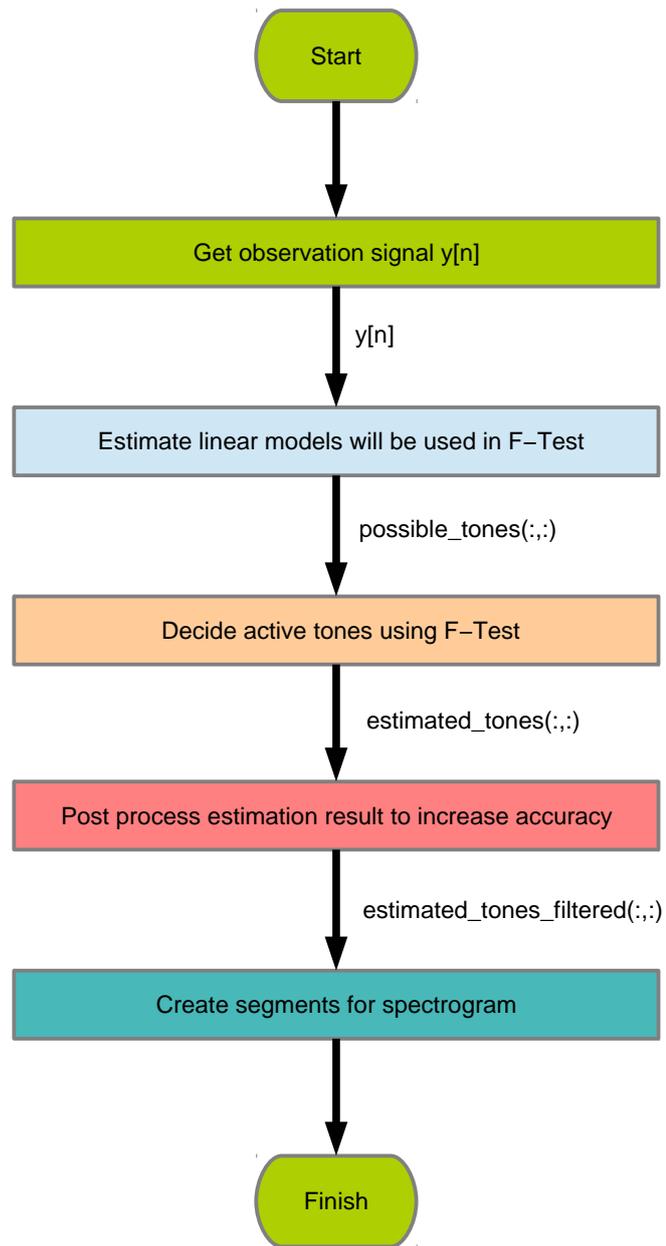
Figure 4.34: The Proposed Dynamic Segmentation Algorithm

4.2.1 Simulation Results

4.2.1.1 Scenario I

Problem Definition

Table 4.2: Problem Parameters for Scenario I

Parameter	Value
N	4000
N_{sw}	96
N_{sdft}	96
N_t	4
N_{mp}	10
SNR	60 dB
p_{fd}	0.1

Table 4.3: Tone Parameters for Scenario I

Tone # / Parameter	n_1	n_2	A	Ω	Φ
1	0	1000	1	0.05π	Random
2	1001	2000	1	0.12π	Random
3	2001	3000	1	0.22π	Random
4	3001	4000	1	0.32π	Random

Results

Ideally, indices of dynamic segments should be at 1001, 2001 and 3001. The proposed method finds starting indices of dynamic segments at 942, 981, 1943 , 1945, 2948, 2985. Although the proposed method couldn't find the ideal indices exactly, it starts new segments around them.

Spectrogram shown in Figure 4.35 is generated using the classical segmentation. Figure 4.36 shows the spectrogram generated using dynamic segments. Even if dynamic segmentation is not the same as the ideal one, the dynamic spectrograph increases the frequency resolution. Figure 4.37 shows the number of actual and detected tones using F-test.
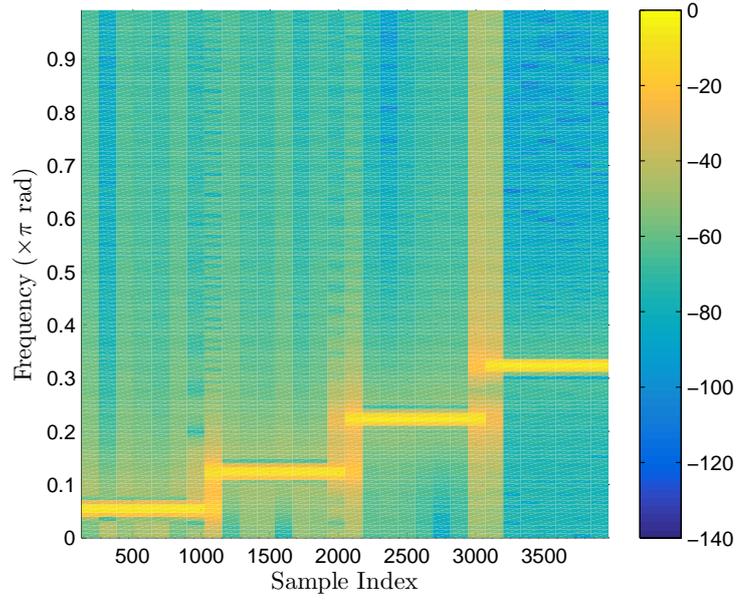
Figure 4.35: Classical Spectrogram
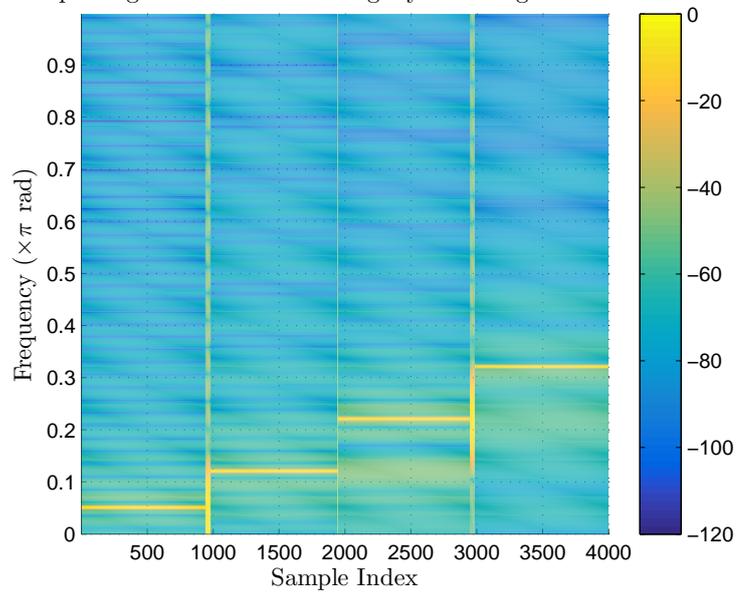
Figure 4.36: Dynamic Spectrogram
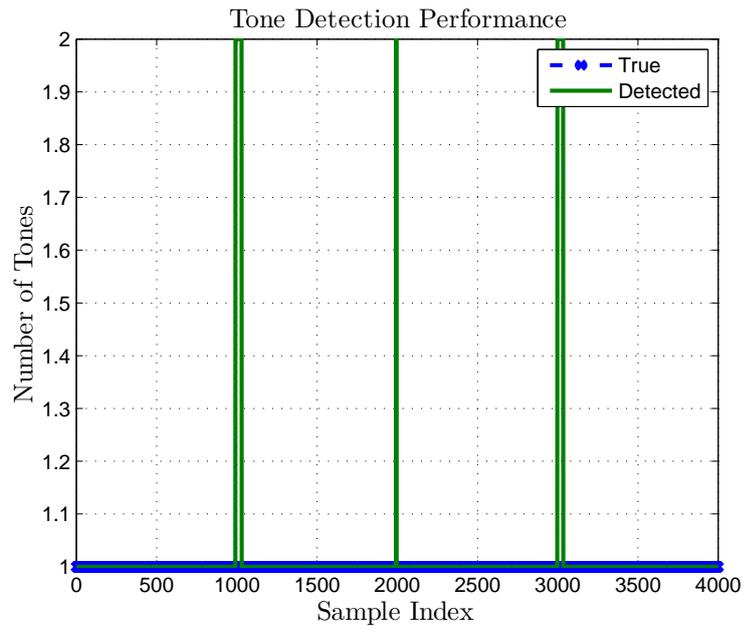
Figure 4.37: True and Estimated Number of Tones

4.2.1.2 Scenario II

Problem Definition

Table 4.4: Problem Parameters for Scenario II

Parameter	Value
N	30000
N_{sw}	24
N_{sdft}	64
N_t	4
N_{mp}	2
SNR	60 dB
p_{fd}	0.1

Table 4.5: Tone Parameters for Scenario II

Tone # / Parameter	n_1	n_2	A	Ω	Φ
1	0	10000	1	0.20π	Random
2	7500	22500	1	0.80π	Random
3	15000	22500	1	0.25π	Random
4	22501	30000	1	0.30π	Random

Results

Ideally, indices of dynamic segments should be at 7500, 10001, 15001, 22501. The proposed method finds starting indices of dynamic segments at 7419, 9989, 14933, 22422, 22444, 22483, 22492. Although the proposed method couldn't find ideal indices exactly, it starts new segments around them.

Spectrogram shown in Figure 4.38 is generated using the classical segmentation. Figure 4.39 show spectrogram generated using dynamic segments. Even if the dynamic segmentation is not same as the ideal one, the dynamic segmentation increases the frequency resolution. Figure 4.42 show the number of actual and detected tones using F-test.

From Figure 4.39, it may not be easy to see the detected frequencies due to the problems related with the MATLAB renderer, Figure 4.40 and Figure 4.41 are given to show the same spectrogram from different angles. As it can be seen

from these figures, each DFT segments have sharp frequency plots.
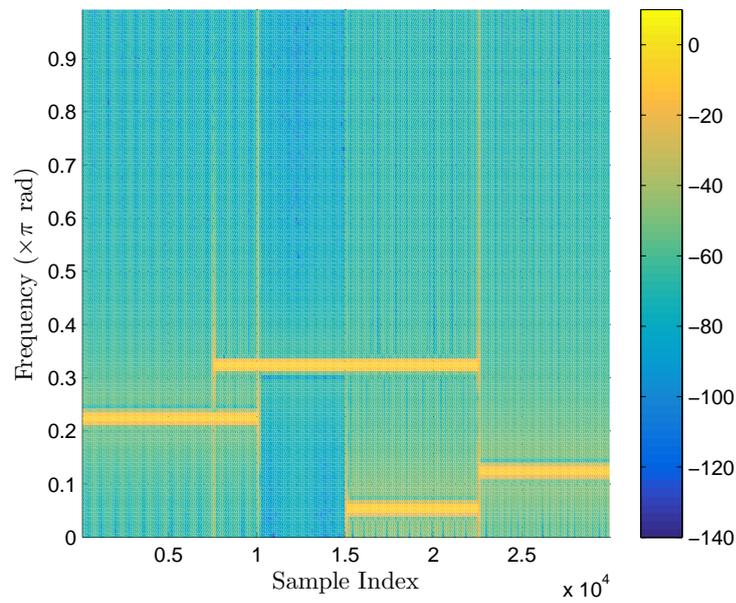
Figure 4.38: Classical Spectrogram
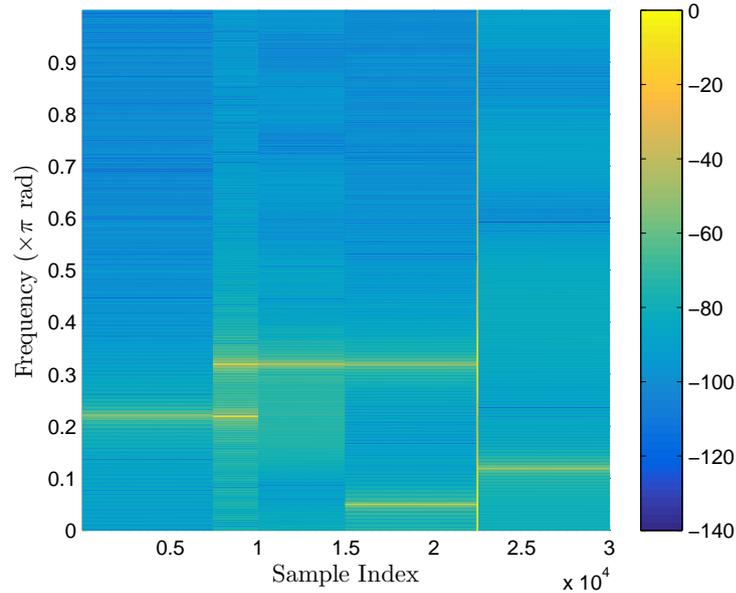
Figure 4.39: Dynamic Spectrogram, View #1

84
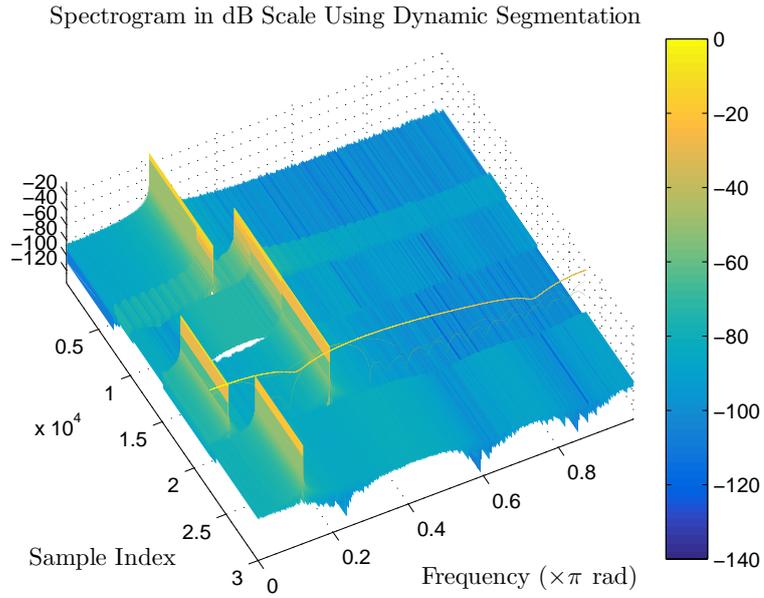
Figure 4.40: Dynamic Spectrogram, View #2
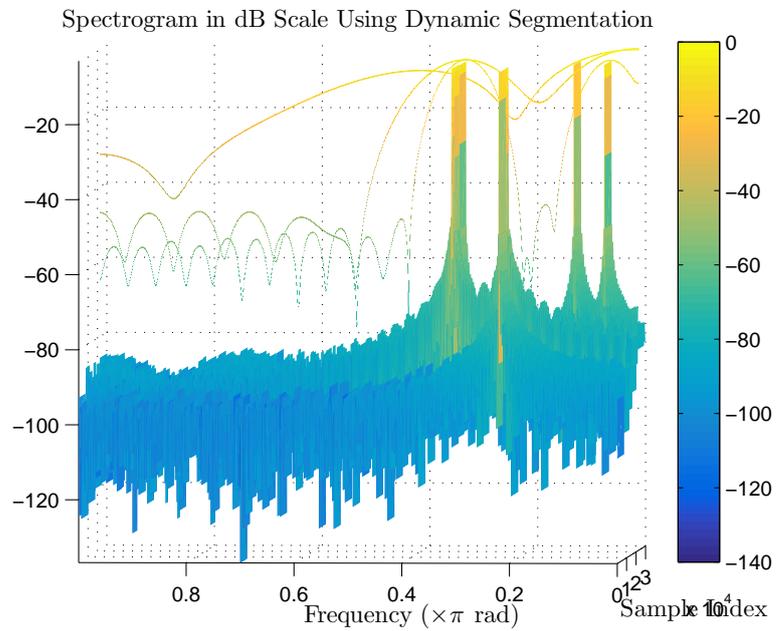
Figure 4.41: Dynamic Spectrogram, View #3
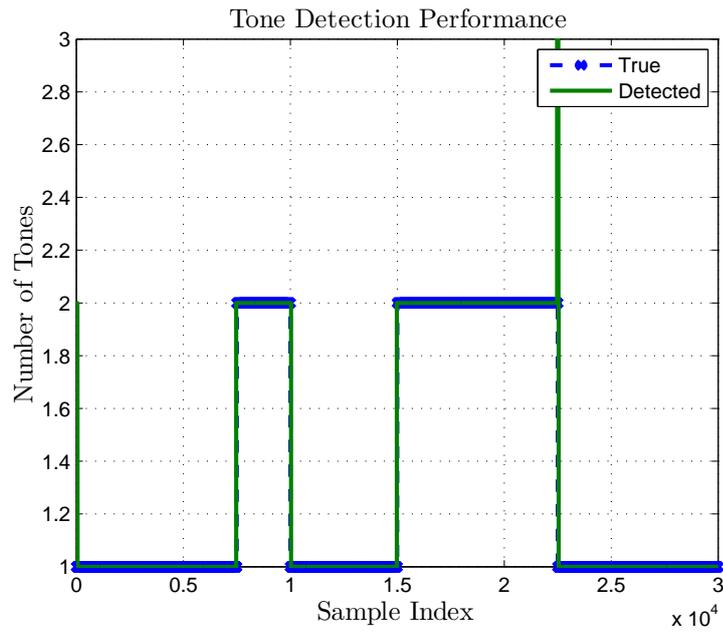
Figure 4.42: True and Estimated Number of Tones

4.3   Segmentation of FM Signals (Problem Type: III)

This example problem is similar to the segmentation of the multi tone signal problem. This problem is based on a single tone signal which has a time varying instantaneous frequency. For example, various FM signals, like linear, hyperbolic FMs, have time-varying instantaneous frequencies. Since various FM pulses are used in radar and sonar applications, these signals are specially considered in this section [28]. The aim is improvements in the frequency resolution of spectrograms of FM signals observed under AWGN. Let us consider the signal given as

$$s(t) = A\sin(\theta(t)).$$

The instantaneous frequency of the signal $s(t)$ is defined as

$$f(t) \triangleq \frac{1}{2\pi}\frac{\mathrm{d}\theta(t)}{\mathrm{d}t}. \tag{4.14}$$

It is assumed that $f(t)$ is a continuous function and has its derivative is well defined at every point. Noisy version of this signal is observed in discrete time as

$$y[n] = s[n] + \omega[n].$$

If $F[n]$ denotes the instantaneous frequency of $x[n]$ is discrete time, the relation $f(t) = F[tf_s]f_s$ is valid. Notice that $n = tf_s$ where f_s is the sampling frequency.

Let us consider an HFM signal as an example. General equation for an HFM signal is given in as

$$s(t) = \cos\left(\frac{2\pi}{b}\ln\left(1 + bf_0 t\right) + \theta_0\right)$$

$$b \triangleq \frac{f_0 - f_1}{f_0 f_1 T} \tag{4.15}$$

where f_0, f_1, T, θ_0 in (4.15) represents start frequency, stop frequency, duration and initial phase of FM pulse, respectively. The instantaneous frequency of the signal given can be written as

$$f(t) = \frac{f_0}{1 + \dfrac{f_0 - f_1}{f_1 T}t}$$

by using (4.14).

An example plot of $f(t)$ where $f_0 = 300$Hz, $f_1 = 3800$ Hz, $T = 5$ sec, $\theta_0 = 0$ is given in Figure 4.43. Spectrogram of $s[n]$ which is generated by sampling $s(t)$ with $f_s = 8000$ Hz is shown Figure 4.44. This is a classical spectrogram, i.e., lengths of segments and overlapped samples are fixed. Parameters $Nwindow$, $Noverlap$ and $NDFT$ shown in classical spectrogram are the same as previous definitions given in the multi tone segmentation problem.
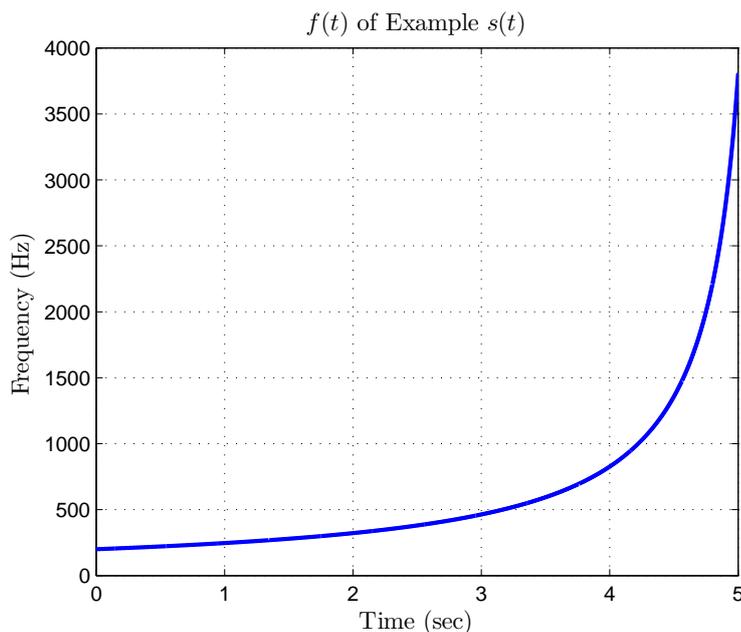
Figure 4.43: The Instantaneous Frequency of Example Signal

As it can be seen from Figure 4.43 or Figure 4.44, the instantaneous frequency seems to be almost constant at the beginning of the pulse. It rapidly increases at the end of the pulse. Frequency resolution of spectrogram can be increased by taking long segments where instantaneous frequency changes slowly and short segments where the instantaneous frequency changes rapidly. These concepts are explained in detail in the previous problem. It is assumed that FM pulse parameters and properties of $f(t)$ function, like monotone increasing/decreasing, are unknown.

The proposed approach uses the fact that instantaneous frequency is a continuous function. A single segment is used while the instantaneous frequency stays almost constant. As the instantaneous frequency starts to deviate, new seg-
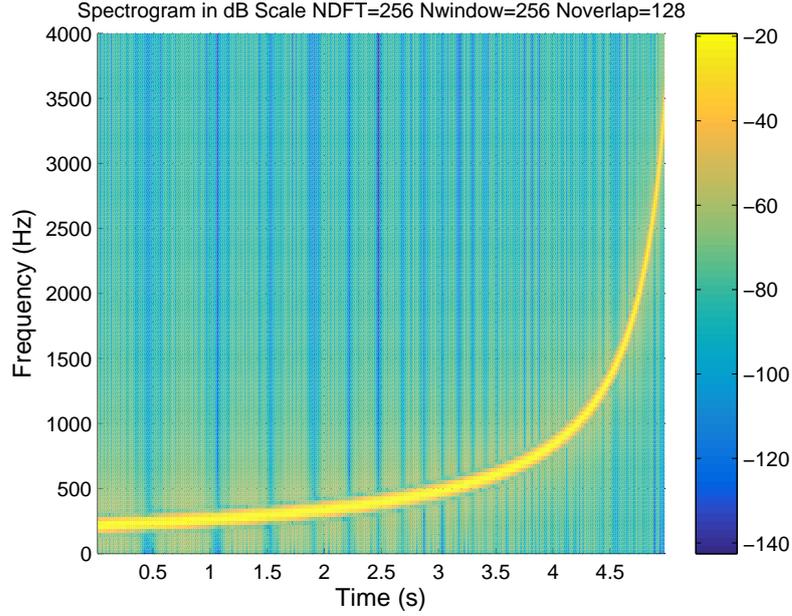
Figure 4.44: Classical Spectrogram of Example Signal

ment starts. Whole observation is dynamically segmented using this approach. After segmentation, spectrogram is plotted as described previously. One may also limit the minimum length of the dynamic segments. It guarantees that frequency resolution of the dynamic spectrogram is as good at least as the classical spectrogram.

The algorithm proposed for the segmentation of FM pulses is very similar to the algorithm proposed for segmentation of multi tone signal. The blocks shown in Figure 4.34 from begging up to the post process block are same for FM case. The F-test based verification algorithm proposed in the previous problem is used also for FM case without any modification. Last two blocks are different for FM case. New post processing algorithm is illustrated in Figure 4.45.

The following parameters are used in the segmentation algorithm:

- W_{min}, *Minimum Window*: This is the minimum length for dynamic segments. Length of each dynamic segment is at least W_{min} samples.

- T_{min}, *Minimum Period*: Segment length should be at least T_{min} times its fundamental period.

89
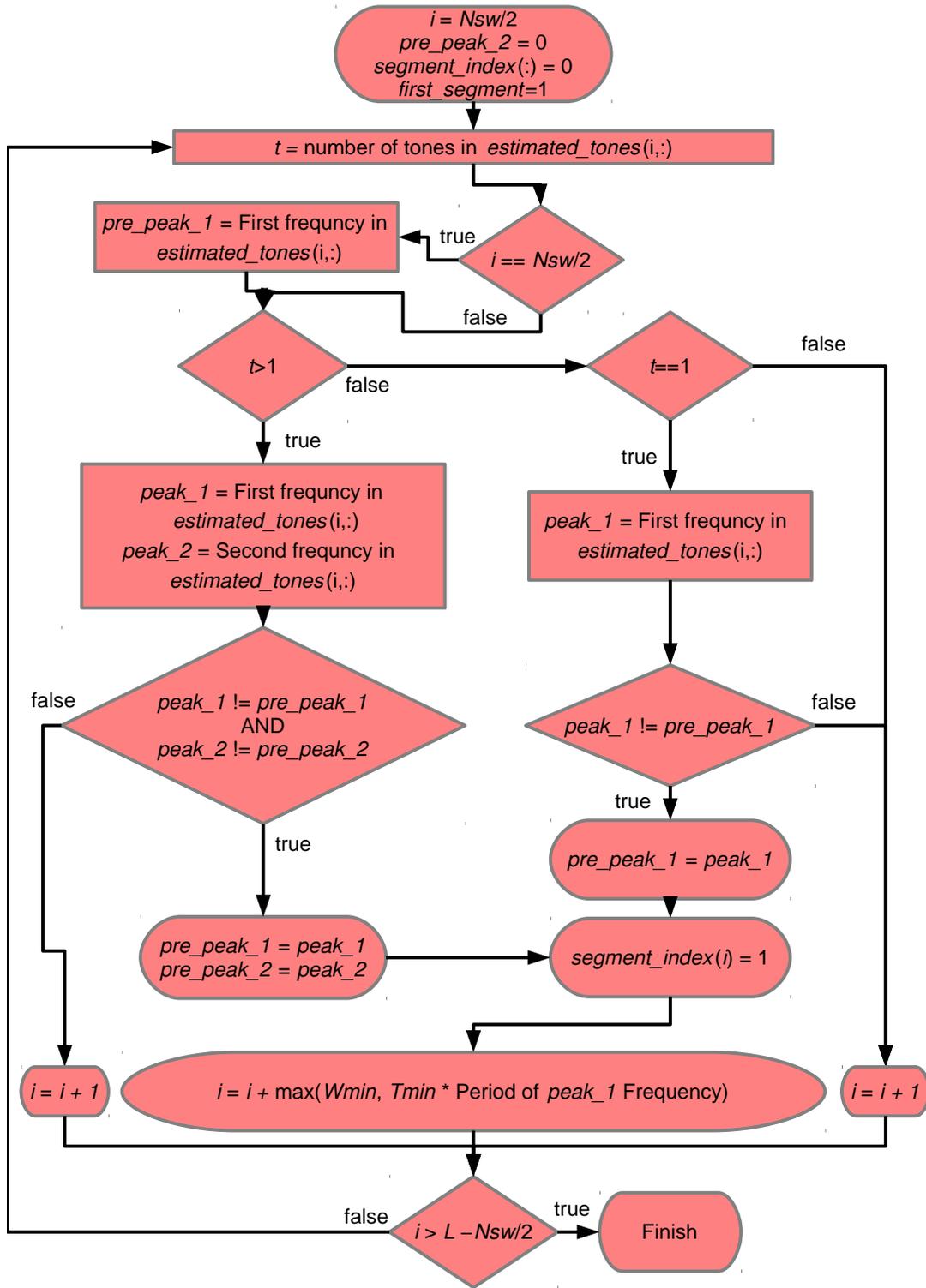
Figure 4.45: The Proposed Segmentation Algorithm for FM Pulses

This approach divides $[0, 2\pi)$ radial frequency interval to $Nsdft$ pieces. The frequency resolution of dynamic spectrogram will be $2\pi/N_{sdft}$ rad. The proposed

algorithm basically divides spectrogram into $2\pi/N_{sdft}$ rad steps and determines which parts of observation vector can be considered as a single tone. If the instantaneous frequency is almost constant at a specific part of observed signal, it can be considered as a single constant tone.

Two rules are used to end the current segment and start a new segment:

1. Suppose, F-test verifies that both i^{th} and $i-1^{\text{th}}$ search window contains dual tone. In other words observation vectors in both search windows can be modeled as summation of two tones. It is decided that current segment should end at location of $i-1^{\text{th}}$ window and new one should start at the location of ith window if at least one of these detected frequencies is not same for both $i-1^{\text{th}}$ and i^{th} search windows.

2. Suppose, F-test verifies that both i^{th} and $i-1^{\text{th}}$ search window contains singe tone. In other words observation vector in both search windows can be modeled as single tone. It is decided that current segment should end at the location of $i-1^{\text{th}}$ window and new one should start at the location of ith window if frequency of the single tone is not same for both $i-1^{\text{th}}$ and i^{th} search windows.

After a new segment starts, it is not tested against a new segment for specific time determined by W_{min} and T_{min} parameter.

$Nsdft$ and Nsw parameters shown in Figure 4.30 effects lengths of segments directly. As values of these parameters are increased, frequency resolution of the dynamic spectrogram also increases. As resolution increases, small changes in detected frequencies causes F-test to decide end the current segment and start a new one more quickly.

When this stage is competed, $segment\_index()$ stores start and end points of dynamic segments.

At the last stage, dynamic spectrogram is plotted using the dynamic segments as shown in Figure 4.46.

91
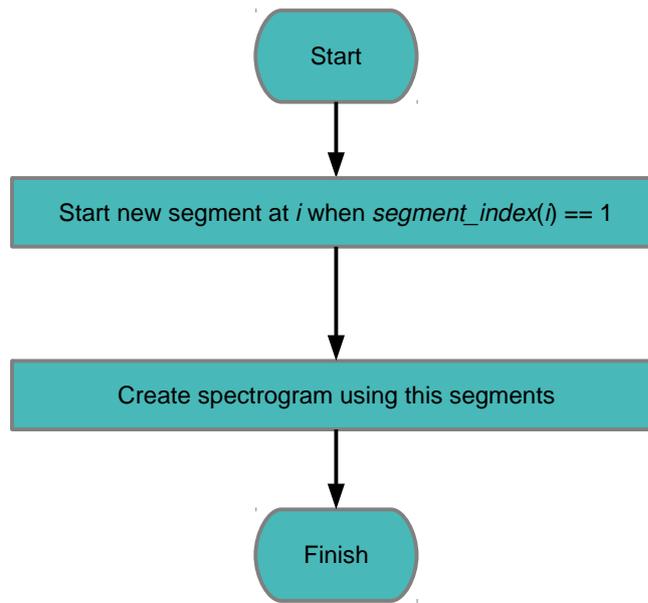
Figure 4.46: Dynamic Spectrogram Generation Using Dynamic Segments

The proposed algorithm is shown in Figure 4.47. As stated before, only the last two blocks are not same for Figure 4.47 and Figure 4.34.
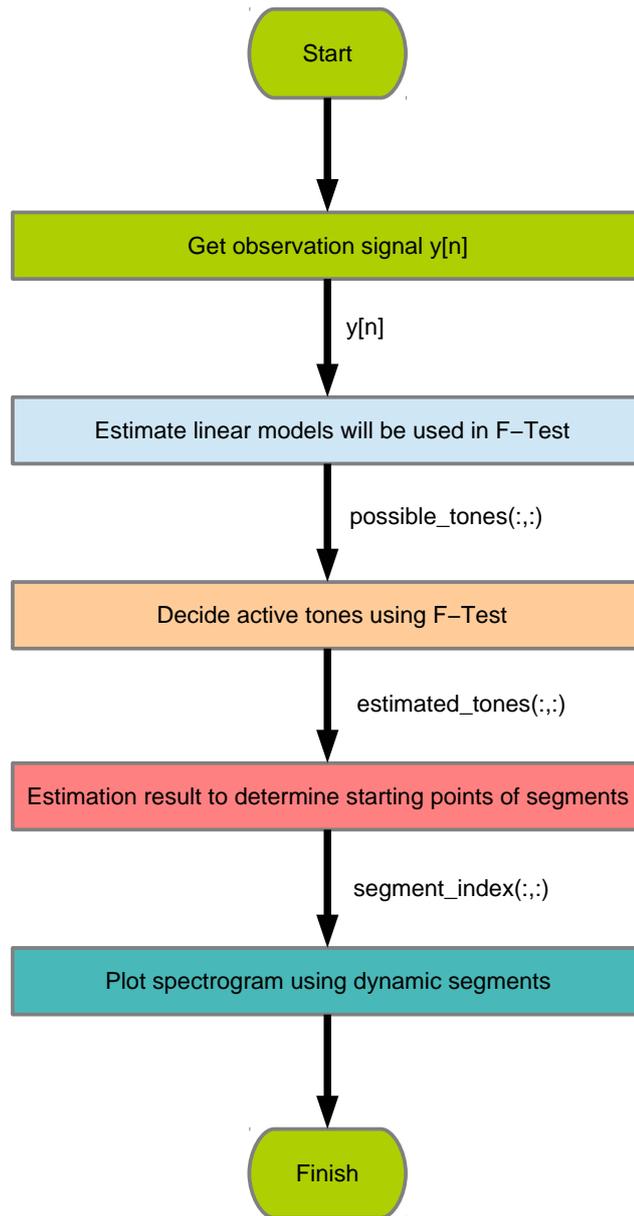
Figure 4.47: General Flow of The Proposed Dynamic Segmentation Approach

4.3.1 Simulation Results

4.3.1.1 Scenario I

Problem Definition

Table 4.6: Problem Parameters for Scenario I

Parameter	Value
N_{sw}	128
N_{sdft}	256
T_{min}	5
W_{min}	256
SNR	60 dB
p_{fd}	0.1
f_s	8000 Hz

Interested signal is a noisy observation of HFM pulse with the instantaneous frequency given in (4.15). Pulse parameters are given in Table 4.7.

Table 4.7: Signal Parameters for Scenario I

Parameter	Value
f_0	300
f_1	3800
T	5
θ_0	Random

Results

Figure 4.48 shows plot of $f(t)$ given in (4.15). Classical and dynamic spectrogram of FM pulse are shown in Figure 4.49 and 4.51, respectively. Since the instantaneous frequency of pulse increases rapidly at the end of pulse, length of segments decreases. Length of segment is limited by W_{min} parameter approximately after 30th segment. After that point, dynamic spectrogram looks similar to the classical spectrogram. Figure 4.50 show the length of each segment.
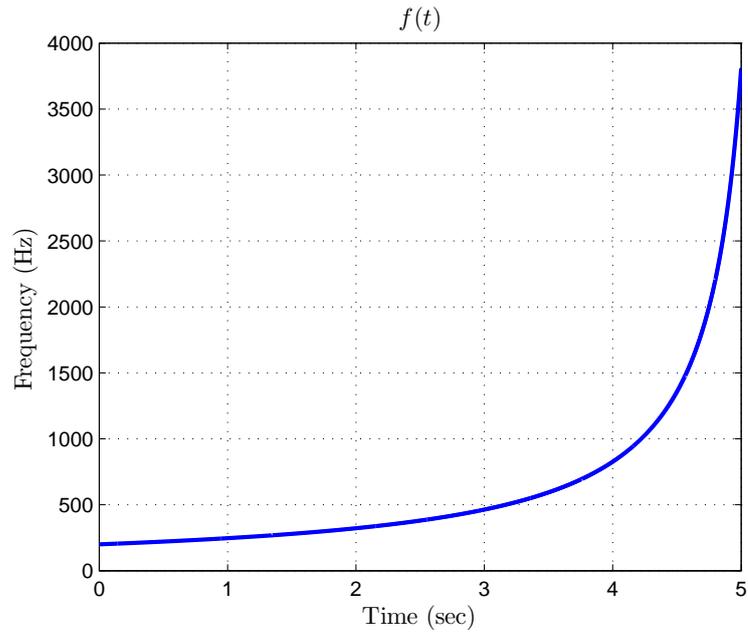
Figure 4.48: The Instantaneous Frequency of Signal, Scenario I
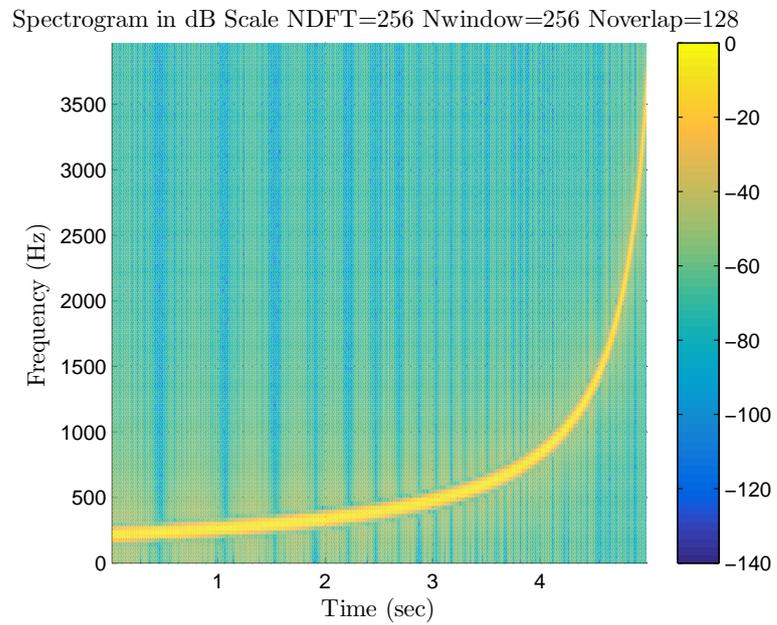
Figure 4.49: Classical Spectrogram, Scenario I

95
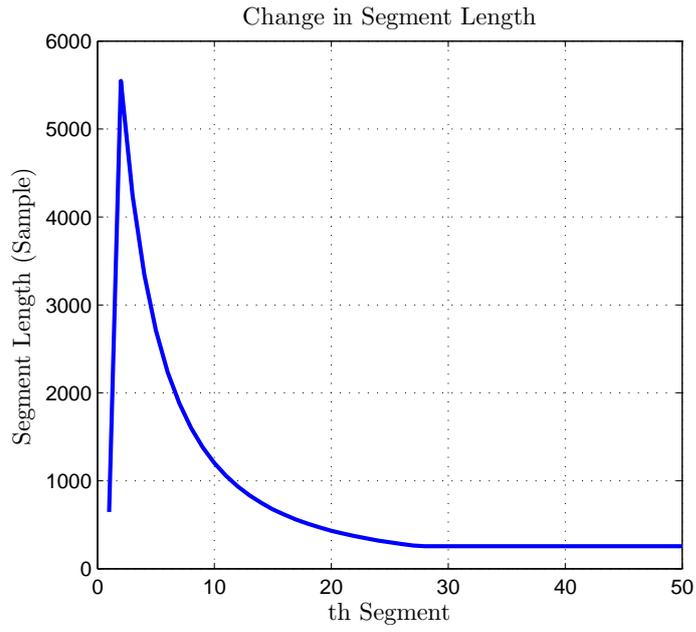
Figure 4.50: Change in Dynamic Segment Length, Scenario I
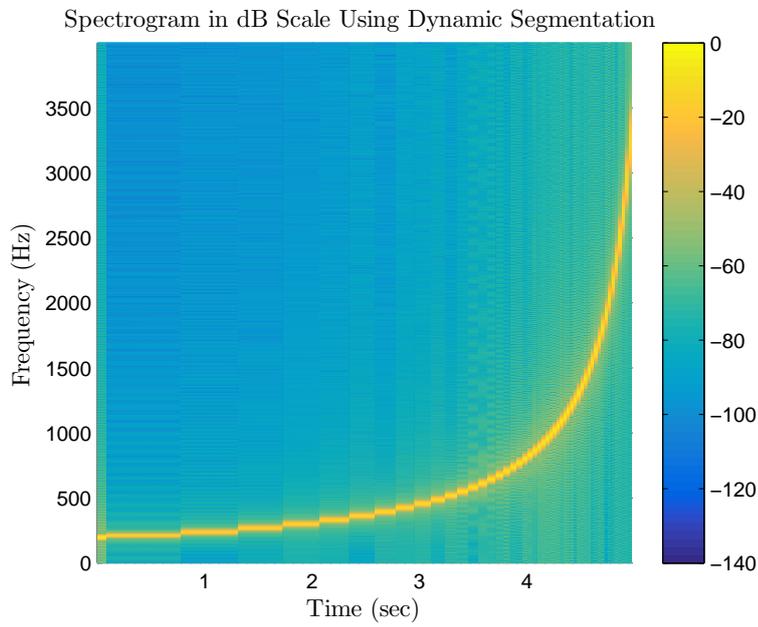
Figure 4.51: Dynamic Spectrogram, Scenario I

4.3.1.2   Scenario II

Problem Definition

<p align="center">Table 4.8: Problem Parameters for Scenario II</p>

Parameter	Value
N_{sw}	128
N_{sdft}	256
T_{min}	5
W_{min}	256
SNR	60 dB
p_{fd}	0.1
f_s	10000 Hz

This example is about LFM pulse. General expression and instantaneous frequency of LFM signal is given as follows

$$x(t) = \cos\left(2\pi\left(f_0 + \frac{kt}{2}\right)t + \theta_0\right),$$

$$f(t) = f_0 + kt. \tag{4.16}$$

<p align="center">Table 4.9: Signal Parameters for Scenario II</p>

Parameter	Value
f_0	2000
k	1000
T	2
θ_0	Random

Results

Figure 4.52 shows plot of the instantaneous frequency given in (4.16). Classical and dynamic spectrogram of FM pulse is shown in Figure 4.53 and 4.55, respectively. Since slope of the instantaneous frequency of LFM signal is constant, F-test selects the length of each segment almost same. Figure 4.54 shows the length of each segment.
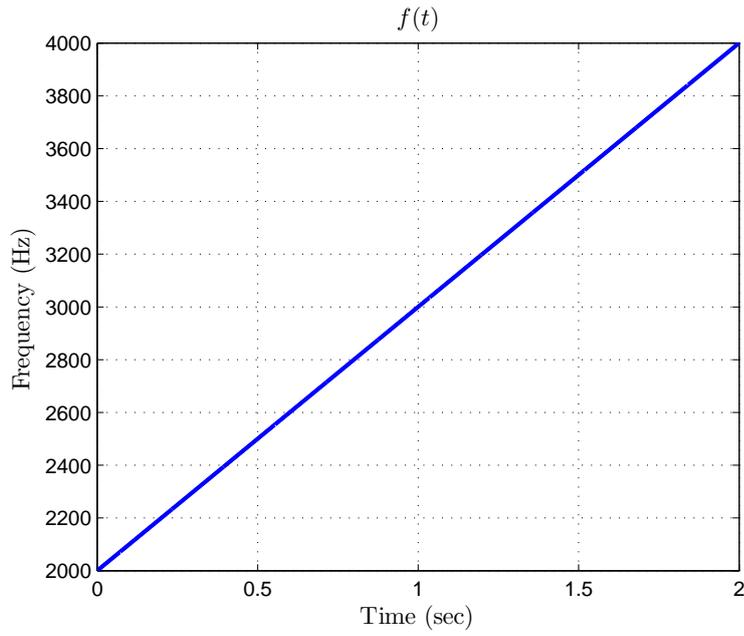
Figure 4.52: The Instantaneous Frequency of Signal, Scenario II
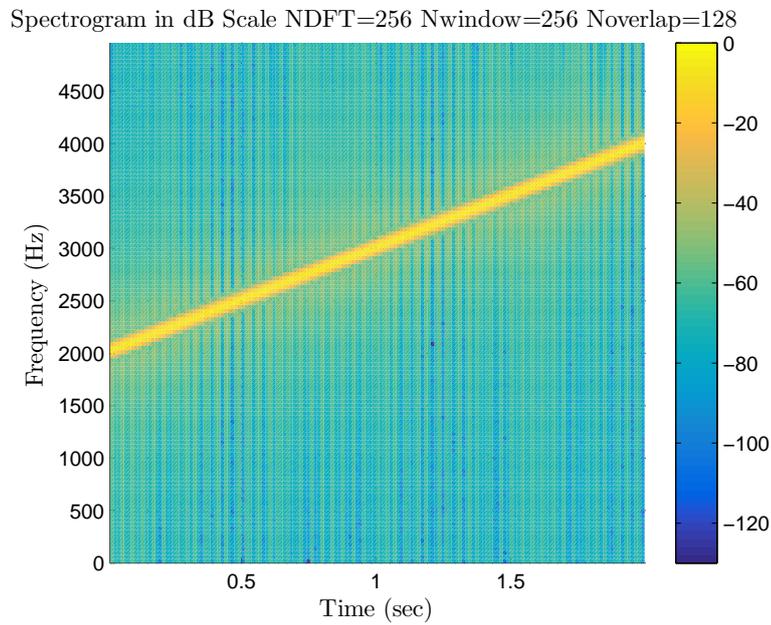
Figure 4.53: Classical Spectrogram, Scenario II

98
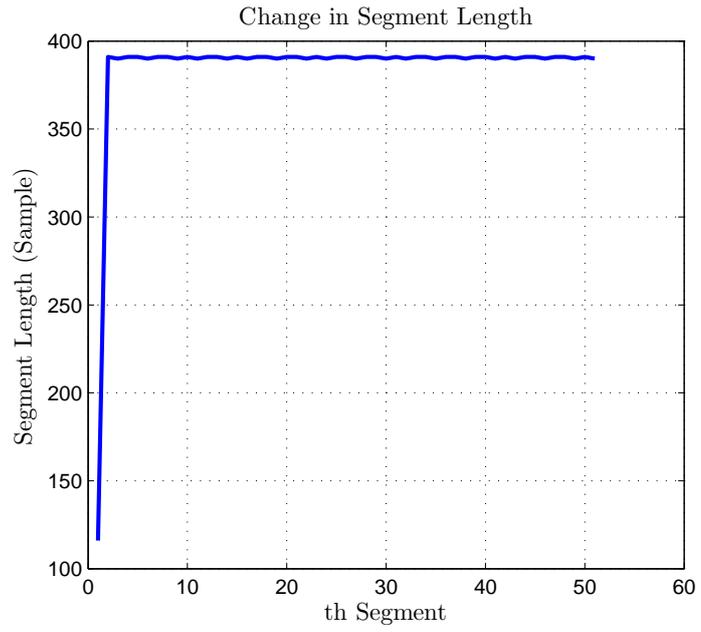
Figure 4.54: Change in Dynamic Segment Length, Scenario II
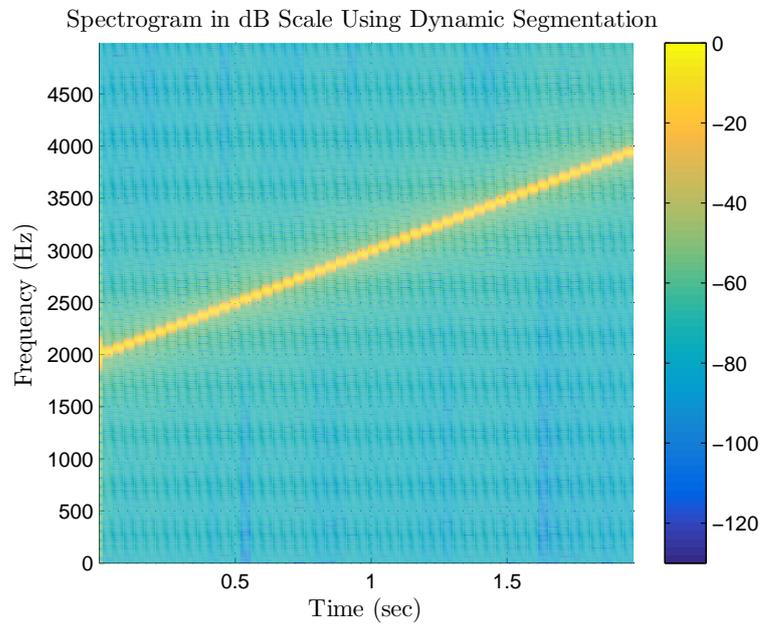
Figure 4.55: Dynamic Spectrogram, Scenario II

4.3.1.3 Scenario III

Problem Definition

Table 4.10: Problem Parameters for Scenario III

Parameter	Value
N_{sw}	128
N_{sdft}	256
T_{min}	5
W_{min}	256
SNR	60 dB
p_{fd}	0.1
f_s	8000 Hz

This example is about EFM pulse. General expression and the instantaneous frequency of EFM signal is given as follows

$$x(t) = \cos\left(2\pi f_0 \left(\frac{k^t}{\ln k}\right) + \theta_0\right),$$

$$f(t) = f_0 k^t. \tag{4.17}$$

Table 4.11: Signal Parameters for Scenario III

Parameter	Value
f_0	2000
k	1.8206
T	5
θ_0	Random

Results

Figure 4.56 shows plot of the instantaneous frequency given in (4.17). Classical and dynamic spectrogram of FM pulse is shown in Figure 4.57 and 4.59, respectively. Figure 4.58 show the length of each segment. As slope of the instantaneous frequency increases, length of segments decreases. The length of segments are limited approximately after 50th segment by W_{min} parameter.

100
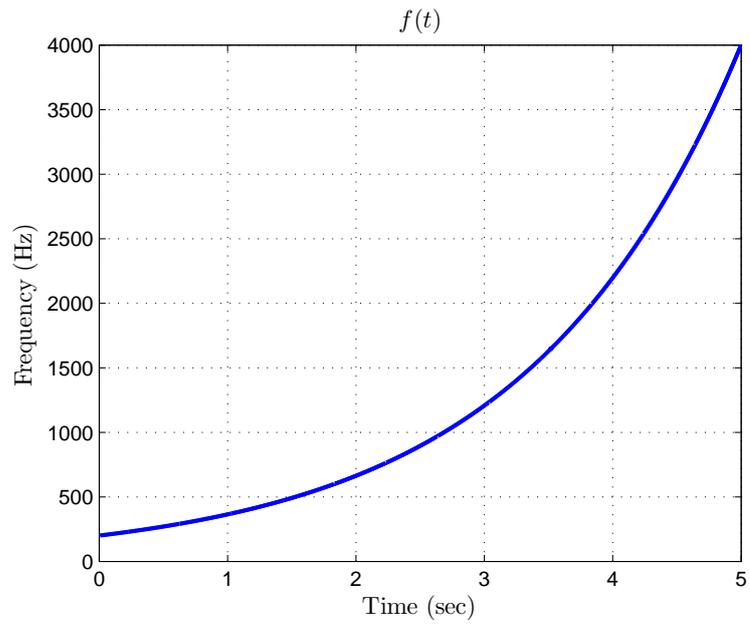
Figure 4.56: The Instantaneous Frequency of Signal, Scenario III
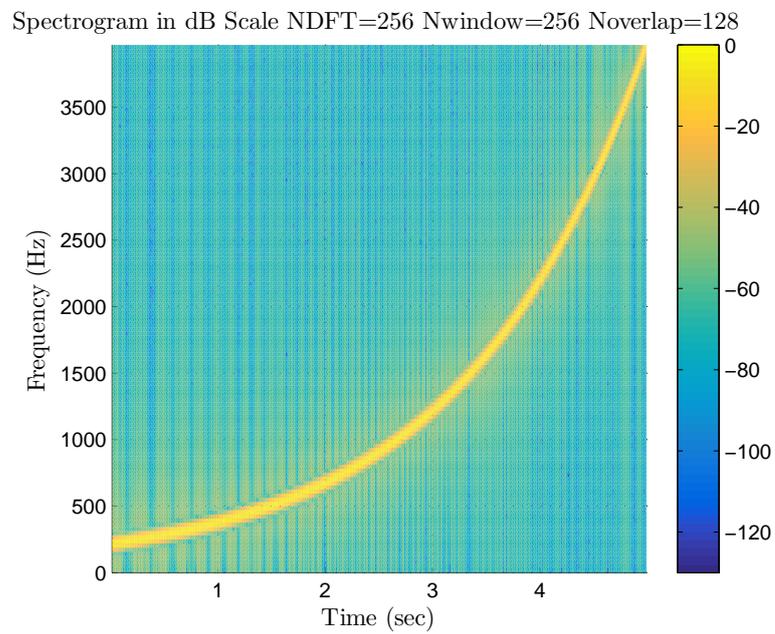
Figure 4.57: Classical Spectrogram, Scenario III

101
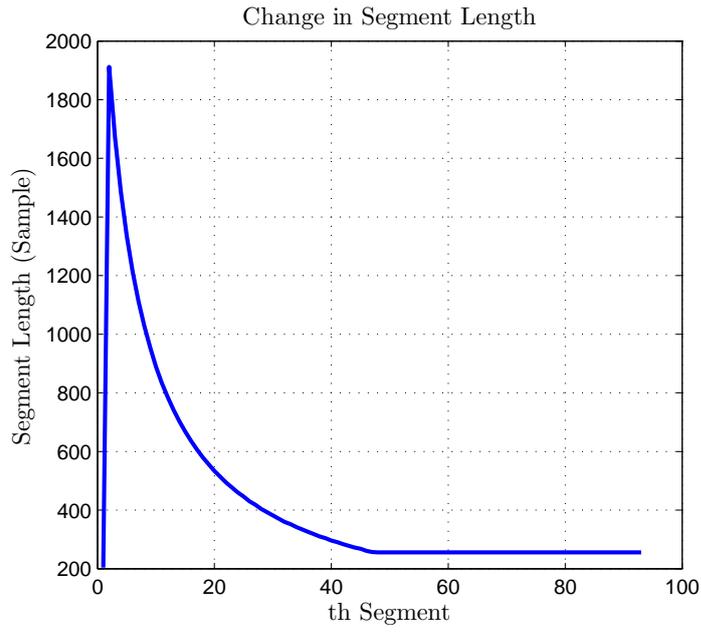
Figure 4.58: Change in Dynamic Segment Length, Scenario III
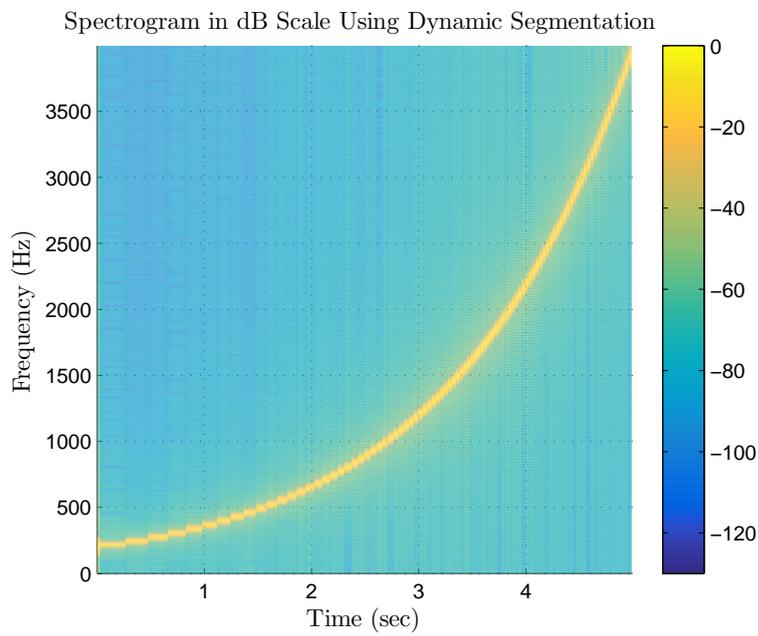
Figure 4.59: Dynamic Spectrogram, Scenario III

4.3.1.4 Scenario IV

Problem Definition

<div align="center">

Table 4.12: Problem Parameters for Scenario IV

Parameter	Value
N_{sw}	128
N_{sdft}	256
T_{min}	5
W_{min}	256
SNR	60 dB
p_{fd}	0.1
f_s	8000 Hz

</div>

This example is about QFM pulse. General expression and the instantaneous frequency of QFM signal is given as

$$x(t) = \cos\left(2\pi f_0 \left(\frac{a}{3}t^3 + \frac{b}{2}t^2 + ct\right) + \theta_0\right),$$

$$f(t) = at^2 + bt + c. \tag{4.18}$$

<div align="center">

Table 4.13: Signal Parameters for Scenario IV

Parameter	Value
a	-862.5
b	3425
c	100
T	4
θ_0	Random

</div>

Results

Figure 4.60 shows plot of the instantaneous frequency given in (4.18). Classical and dynamic spectrogram of FM pulse is shown in Figure 4.61 and 4.63, respectively. Figure 4.62 show the length of each segment. Most of time the length of segments are limited by W_{min} parameter. The length of segments is only above W_{min} limit around 2 seconds where slope of the instantaneous frequency decreases and changes sign.
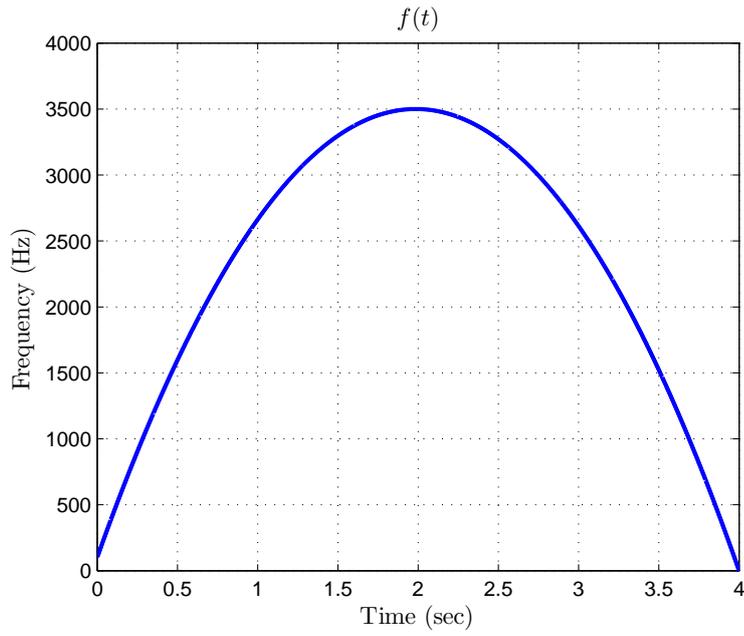
Figure 4.60: The Instantaneous Frequency of Signal, Scenario IV
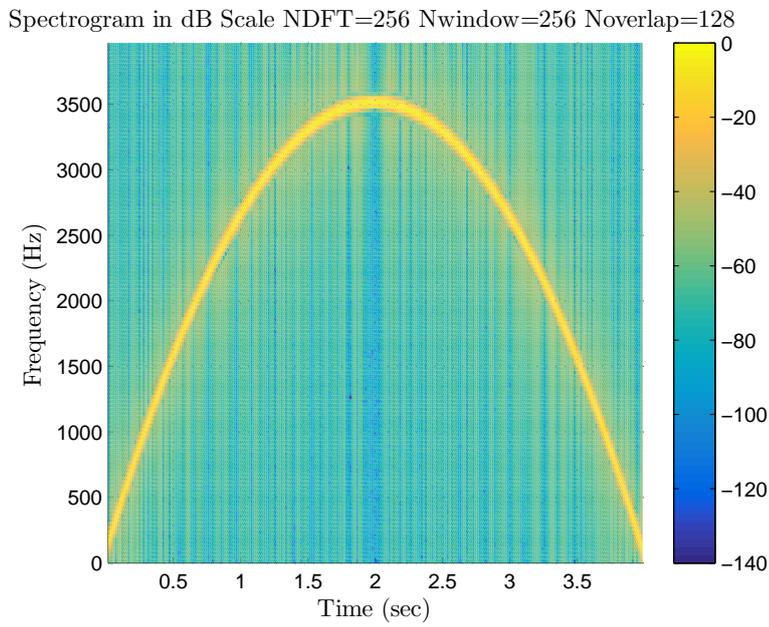
Figure 4.61: Classical Spectrogram, Scenario IV
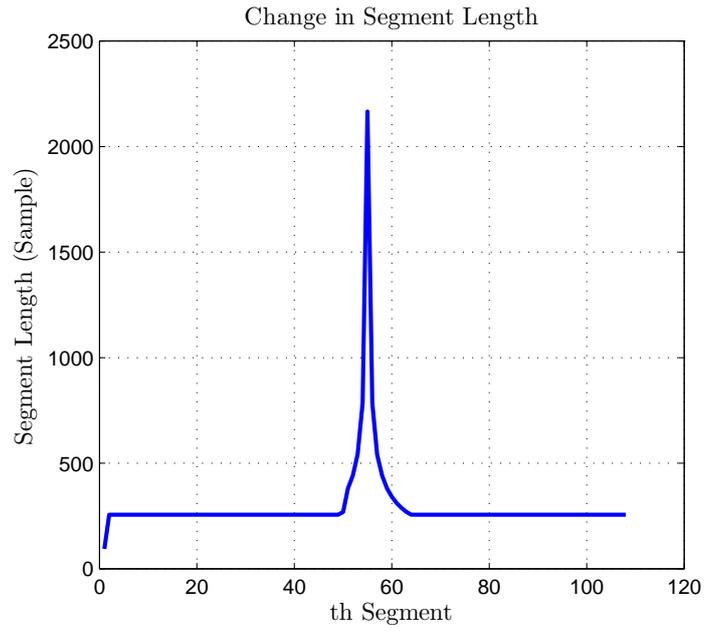
Figure 4.62: Change in Dynamic Segment Length, Scenario IV
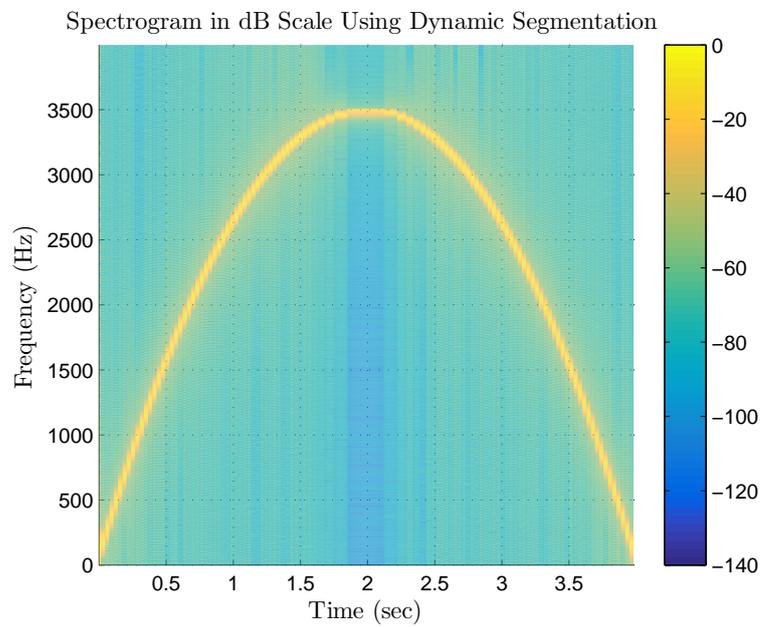
Figure 4.63: Dynamic Spectrogram, Scenario IV

105

4.4   Model Order Selection for Prony's Method (Problem Type: II)

This problem is about signal modeling in discrete time. Impulse response modeling is one of the practical modeling approaches used in signal processing area. Impulse response expression of an LTI filter is given as

$$H(z) = \frac{B_q(z)}{A_p(z)} = \frac{\sum_{k=0}^{q} b_q(k) z^{-k}}{1 + \sum_{k=1}^{p} a_p(k) z^{-k}}. \qquad (4.19)$$

If input of this filter is excited by $v[n]$, the output signal $s[n]$ can be written as

$$S(z) = H(z)V(z) \qquad (4.20)$$

in frequency domain and as

$$s[n] = h[n] * v[n] \qquad (4.21)$$

in time domain.

For example, when $v[n]$ is selected as

$$v[n] = \sum_{k=0}^{p} \delta(n - kn_0)$$

which is an impulse train, $s[n]$ may be used to model a speech signal [27].

When $v[n]$ is selected as

$$v[n] = \sum_{k=0}^{p} a_k \delta(n - n_k),$$

$s[n]$ may be used for multipulse linear predictive coding problems [5, 7, 11, 12].

In this problem, $v[n]$ is assumed to be a unit impulse as given below

$$v[n] = \delta[n].$$

In this case, equation (4.20) can be written as

$$S(z) = H(z)$$

and equation (4.21) can be written as

$$s[n] = h[n].$$

106

Let us assume that q, $b_q(k)$, p and $a_p(k)$ terms in (4.19) are unknown. Length of $s[n]$ is assumed to be N. Once these parameters are estimated, the estimate of $s[n]$, i.e. $\hat{s}[n]$, can be generated from these estimates using the same equality given in (4.19). Least squares solution of this problem minimizes the RSS value defined as [21]

$$\text{RSS}' \triangleq \sum_{n=0}^{\infty} |e'[n]|^2$$

where

$$e'[n] \triangleq s[n] - \hat{s}[n]. \tag{4.22}$$

However, this is not a linear least squares problem and requires solutions for set of $p + q + 1$ non-linear equations [14]. Instead of non-linear equations, this problem can be solved approximately by set of linear equations. Let us consider the case where the transfer function has no zeros and has fixed number of poles. In this case by setting $q = 0$, (4.19) can be written as

$$S(z) = \frac{B_q(z)}{A_p(z)} = \frac{b_q(0)}{1 + \sum_{k=1}^{p} a_p(k) z^{-k}}. \tag{4.23}$$

This kind of signal is called as all-pole signal and widely used in signal processing problems like speech signal modeling [27]. Also, many signals in practical systems can be modeled as all-pole signals, as well [14].

Equation (4.23) can be written as

$$s[n] + a_p(1)s[n-1] + a_p(2)s[n-2] + \ldots + a_p(p)s[n-p] = b_q(0)\delta[n] \tag{4.24}$$

in time domain. Additionally, if signal is considered as an impulse response of a casual filter, $s[n] = 0$ is satisfied when $n < 0$.

The signal $s[n]$ can be written recursively using the characteristics of IIR filters as follows

$$a_p(1)s[0] = -s[1],$$
$$a_p(1)s[1] + a_p(2)s[0] = -s[2],$$
$$a_p(1)s[2] + a_p(2)s[1] + a_p(3)s[0] = -s[3],$$
$$\ldots = \ldots .$$

These equations can be expressed in matrix form as

$$
\begin{bmatrix}
s[0] & 0 & 0 & \ldots & 0 \\
s[1] & s[0] & 0 & \ldots & 0 \\
s[2] & s[1] & s[0] & \ldots & 0 \\
\vdots & \vdots & \vdots & & \vdots
\end{bmatrix}
\begin{bmatrix}
a_p(1) \\
a_p(2) \\
\vdots \\
a_p(p)
\end{bmatrix}
= -
\begin{bmatrix}
s[1] \\
s[2] \\
\vdots \\
s[N-1]
\end{bmatrix} .
$$

Using the standard notation, they also can be written as

$$
\boldsymbol{Ap} = \boldsymbol{x}. \tag{4.25}
$$

Estimation of unknown parameters of an all-pole signal by using the least squares solution of (4.25) is known as Prony's method.

$$
\text{RSS}_{\text{prony}} \triangleq \sum_{n=0}^{\infty} |e_{\text{prony}}[n]|^2 \tag{4.26}
$$

where

$$
e_{\text{prony}}[n] \triangleq s[n] + \sum_{k=1}^{p} \hat{a}_p(k)s[n-k] \tag{4.27}
$$

is the **Prony error**.

Notice that the error defined in (4.27) is different than the error defined in (4.22). Linear LS solution tries to minimize RSS defined in (4.26). Therefore, this solution may not minimize the actual approximation error. This is a disadvantage of Prony's solution.

After finding $\hat{\boldsymbol{p}}$ by using LS approach for the equation (4.25), $b_q(0)$ becomes the only unknown term in (4.23). One may use the

$$
\hat{b}_q(0) = s[0] \tag{4.28}
$$

equation directly which can be written from (4.24). When all parameters are estimated, estimate signal can be written as

$$
\hat{S}(z)_{\text{prony}} = \frac{\hat{B}_q(z)}{\hat{A}_p(z)} = \frac{\hat{b}_q(0)}{1 + \sum_{k=1}^{p} \hat{a}_p(k)z^{-k}} .
$$

In this equation, each $\hat{a}_p(k)$ is taken from $\hat{\boldsymbol{p}}$. $\hat{s}[n]_{\text{prony}}$ can be found by taking the inverse z transform of $\hat{S}(z)_{\text{prony}}$.

Although (4.28) means $\hat{s}[0]_{\text{prony}} = s[0]$, this selection may cause an improper selection of $\hat{b}_q(0)$ if $s[0]$ is a badly scaled value. To avoid the dependency of $\hat{b}_q(0)$ on a single sample, $s[0]$, $\hat{b}_q(0)$ can be estimated as [19]

$$\hat{b}_q(0) = \sqrt{r_s(0) + \sum_{k=1}^{p} \hat{a}_p(k) r_s(k)},$$

$$r_s(k) \triangleq \sum_{n=0}^{\infty} s(n)s(n-k).$$

This selection rule matches not their first samples of both signals but their energy [19]. It will be used to calculate $\hat{b}_q(0)$ in the following example scenarios.

The difference between the actual signal and the approximated signal using Prony's method is called as **Prony approximation error** and it is calculated as

$$\epsilon_{\text{prony}}[n] = s[n] - \hat{s}[n]_{\text{prony}}. \tag{4.29}$$

In this problem, $s[n]$ with length N is observed under AWGN as $y[n]$. Parameters given in (4.23) are estimated from noisy observations using Prony's method. Observation model is expressed as

$$y[n] = s[n] + w[n].$$

To apply Prony's solution, a suitable p value which denotes the number of poles of all-pole model should be selected.

The $\boldsymbol{A}$ and $\boldsymbol{p}$ matrices defined in (4.25), when the number of poles in Prony's solution is p and $p+1$, are given as follows

$$\boldsymbol{A}_p \triangleq \begin{bmatrix} s[0] & 0 & 0 & \ldots & 0 \\ s[1] & s[0] & 0 & \ldots & 0 \\ s[2] & s[1] & s[0] & \ldots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ s[N-2] & s[N-3] & s[N-4] & \ldots & s[N-p-1] \end{bmatrix}_{N-1 \times p,} \tag{4.30}$$

$$\boldsymbol{p}_p \triangleq \begin{bmatrix} a_p(1) \\ a_p(2) \\ \vdots \\ a_p(p) \end{bmatrix}_{p \times 1,}$$

$$\boldsymbol{A}_{p+1} = \begin{bmatrix} s[0] & 0 & 0 & \dots & 0 & 0 \\ s[1] & s[0] & 0 & \dots & 0 & 0 \\ s[2] & s[1] & s[0] & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ s[N-2] & s[N-3] & s[N-4] & \dots & s[N-p-1] & s[N-p] \end{bmatrix}_{N-1 \times p+1,}$$

$$\boldsymbol{A}_{p+1} = \begin{bmatrix} & 0 \\ & 0 \\ \boldsymbol{A}_p & 0 \\ & 0 \\ & \vdots \end{bmatrix}_{N-1 \times p+1,} \tag{4.31}$$

$$\boldsymbol{p}_{p+1} = \begin{bmatrix} a_p(1) \\ a_p(2) \\ \vdots \\ a_p(p) \\ a_p(p+1) \end{bmatrix}_{p \times 1.}$$

As it can be seen from (4.31), $\boldsymbol{A}_p$ is nested in $\boldsymbol{A}_{p+1}$ and each column is independent from each other except some rare and special conditions. Therefore, F-test can be used to find the proper p number. This case is similar to Type II problem. For this problem, K and $\hat{K}$ defined in Type II are denoted as p and $\hat{p}$, respectively. F-test can be applied straightforwardly to this problem as shown in Figure 4.64.

At each iteration two nested models are compared using F-test. If $F \geq threshold$, it is decided that higher order model (M_H) models the actual signal "significantly better" than the model with lower order (M). Then, each model order is increased by same amount and F-test is applied until $F < threshold$. In that case,
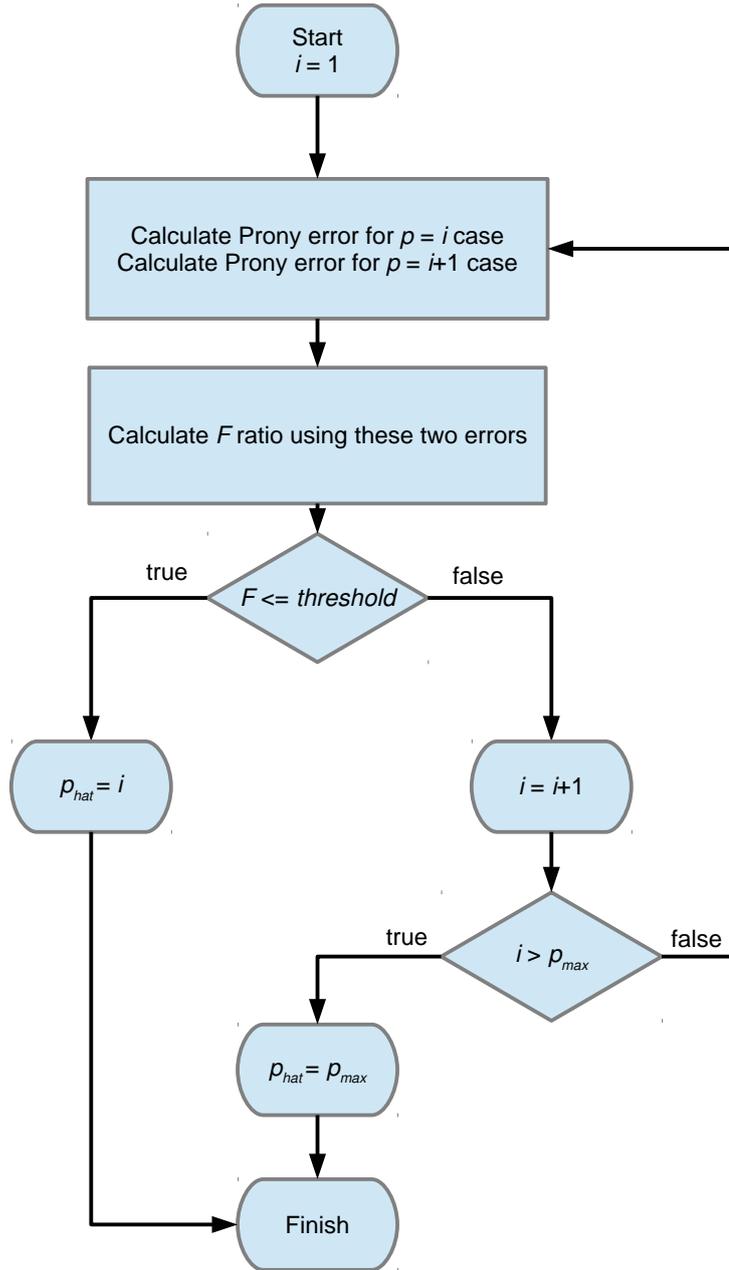
Figure 4.64: The Proposed Model Order Selection Algorithm for Prony's Method

it is decided that the model with order M is suitable to model the actual signal, i.e. $\hat{p} = M$. Then poles and zero are estimated as mentioned before.

In each ith iteration $\boldsymbol{A}_M$ and $\boldsymbol{A}_{M_H}$ are formed as stated previously in (4.30) and (4.31). For each estimation, $\text{RSS}_{\text{prony}}$ is calculated as in (4.26). F ratio and $threshold$ are calculated using the formulas given in (3.1) and (3.13), respectively.

New algorithm parameter p_{max} is introduced as shown in Figure 4.64.

- p_{max}: Limits the maximum number of poles used by Prony's method. One may want to set a upper limit for $\hat{p}$ if there is a priori information about $s[n]$.

Since only noisy observation of the signal $s[n]$ is available, $\boldsymbol{A}$ matrices are formed by using $\boldsymbol{y}$ vector, not $\boldsymbol{s}$.

4.4.1 Simulation Results

In this section, performance of the proposed method for number of pole selection is demonstrated with different scenarios. In each scenario, as many as MCNum parameter Monte Carlo simulation are run.

The following figures show RMSE values of Prony errors and Prony approximation errors. RMSE stands for root-mean-square error and is calculated using all Monte Carlo simulations.

4.4.1.1 Scenario I

Problem Definition

$s[n]$ is generated using the parameters given in Table 4.14. Plot of the signal is given in Figure 4.65.

Table 4.14: Signal Parameters for Scenario I

Parameter	Value
Length	96
Number of Zeros (q)	0
Zero Location #0	1
Number of Poles (p)	5
Pole Location #1	-0.92
Pole Location #2	-0.4
Pole Location #3	-0.3
Pole Location #4	-0.1
Pole Location #5	0.4
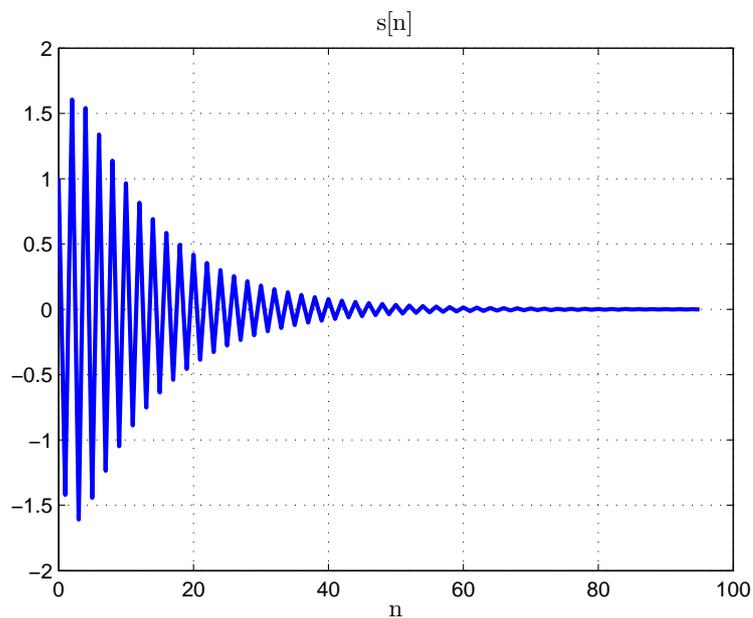
Figure 4.65: Plot of $s[n]$, Scenario I

Problem parameters are given in Table 4.15

Table 4.15: Problem Parameters for Scenario I

Parameter	Value
p_{max}	10
SNR	60 dB
p	0.1
MCNum	10000

Results

113

The poles of all-pole model signals are estimated from noisy observations of the signal. The proposed method is used to select the number of poles that should be used in the model. Figure 4.66 shows the RMS value of Prony error with changing number of poles. The error decreases monotonically as p increases. However, Prony approximation error does not decrease monotonically as shown in Figure 4.67. This is due to the fact that Prony's method is an approximation to the non-linear LS problem. Its error is different than the actual error defined in (4.22). The proposed method tracks the changes in the Prony error, not Prony approximation error. Histogram of the selected number of poles by the proposed method is given in Figure 4.68. The proposed method selects all-pole models with 4 or 5 poles in most of the experiments. From Figure 4.67, it can be seen that using all-pole signal with 5 poles is suitable. However, since the Prony error is almost same for $p = 4$ and $p = 5$ the proposed method chooses both of them. This difference is due to the performance of Prony's method not the proposed method itself. Nevertheless, selecting $p = 4$ is not the worst choice according to Figure 4.67.
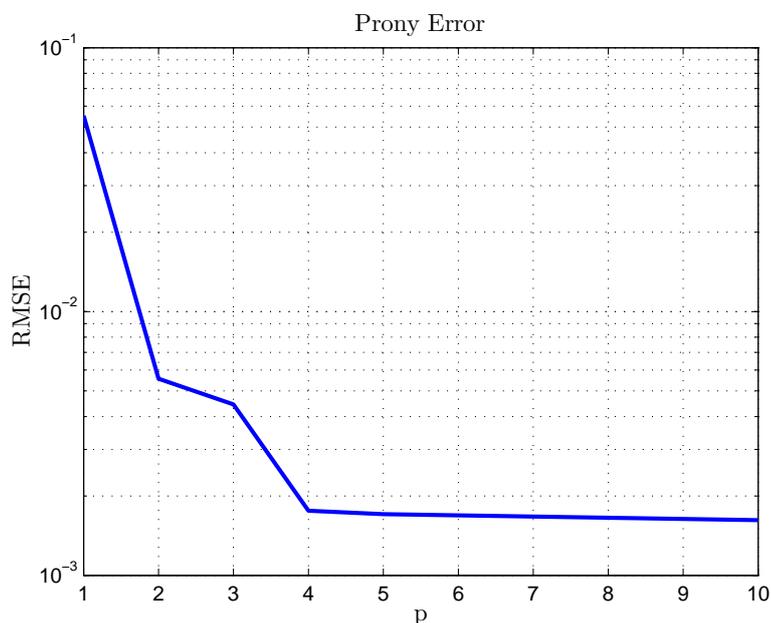
Figure 4.66: $e_{\mathrm{prony}}[n]$ vs p, Scenario I
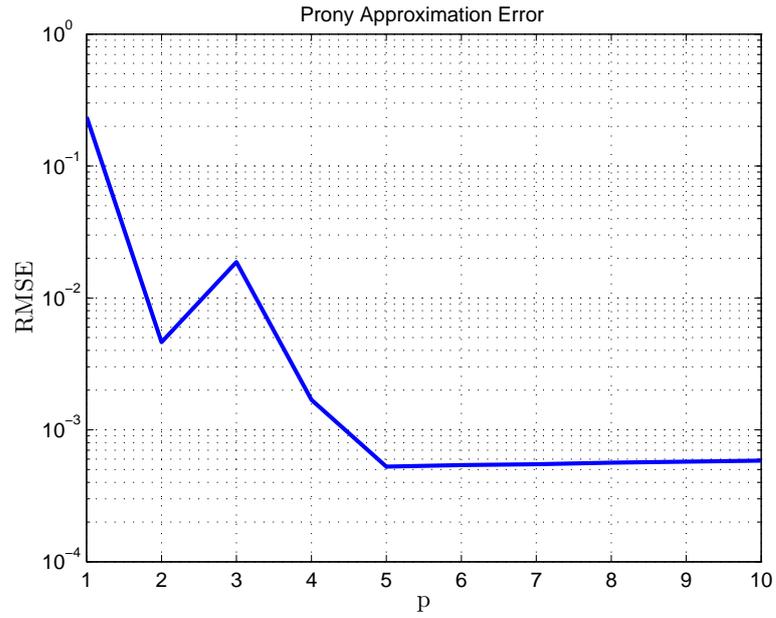
Figure 4.67: $\epsilon_{\mathrm{prony}}$ vs p, Scenario I
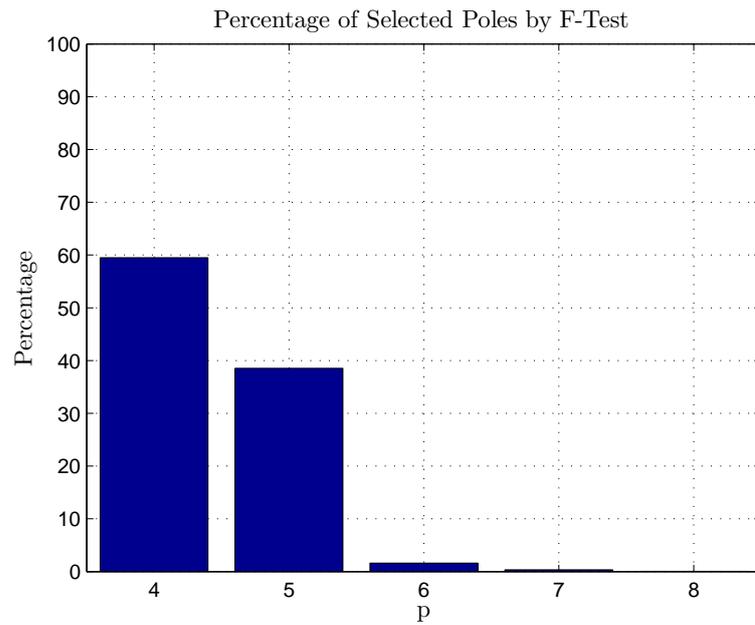
Figure 4.68: Selection Percentage of p by The Proposed Algorithm, Scenario I

4.4.1.2 Scenario II

Problem Definition

Parameters of $s[n]$ signal are given in Table 4.16. Notice that this is not an

all-pole signal, it has 2 zeros in addition to 4 poles. All-pole signal can also be used to model more general signals in practice like this signal [19]. In this scenario, $s[n]$ is generated using more general formula given in (4.19) not using (4.23) which is valid for only all-pole signals. However, during modeling signal is assumed to be an all-pole signal as given in (4.23) and modeled using this assumption. Plot of the signal is given in Figure 4.69.

Table 4.16: Signal Parameters for Scenario II

Parameter	Value
Length	98
Number of Zeros (q)	2
Zero Location #0	1
Zero Location #1	-0.2
Zero Location #2	0.1
Number of Poles (p)	4
Pole Location #1	-0.8 + 0.1j
Pole Location #2	-0.8 - 0.1j
Pole Location #3	-0.7 + 0.6j
Pole Location #4	-0.7 - 0.6j
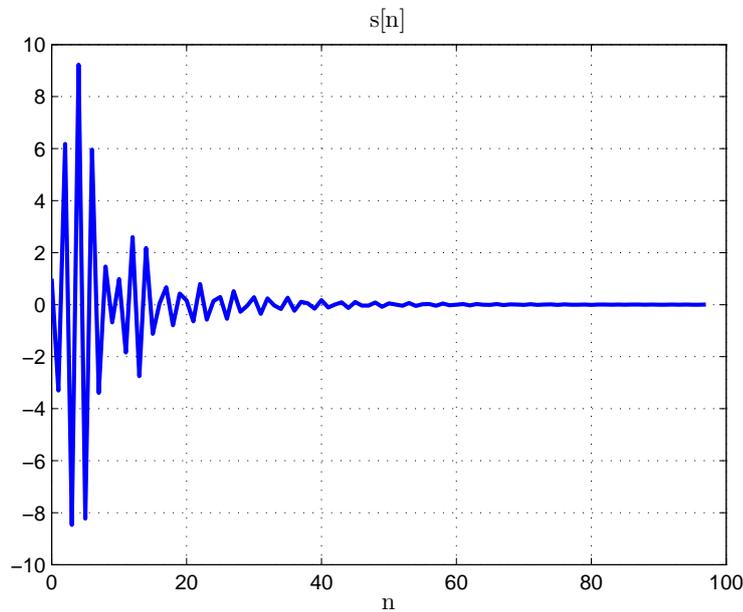
Figure 4.69: Plot of $s[n]$, Scenario II

Problem parameters are given in Table 4.17.

116

Table 4.17: Problem Parameters for Scenario II

Parameter	Value
p_{max}	10
SNR	60 dB
p	0.1
MCNum	10000

Results

From Figure (4.72), the proposed method selects the suitable number of poles at most of time according to Prony error graph shown in Figure 4.70. However, similar to Scenario I, the Prony approximation error does not follow the Prony error exactly. This situation is independent from the proposed method as explained previously.
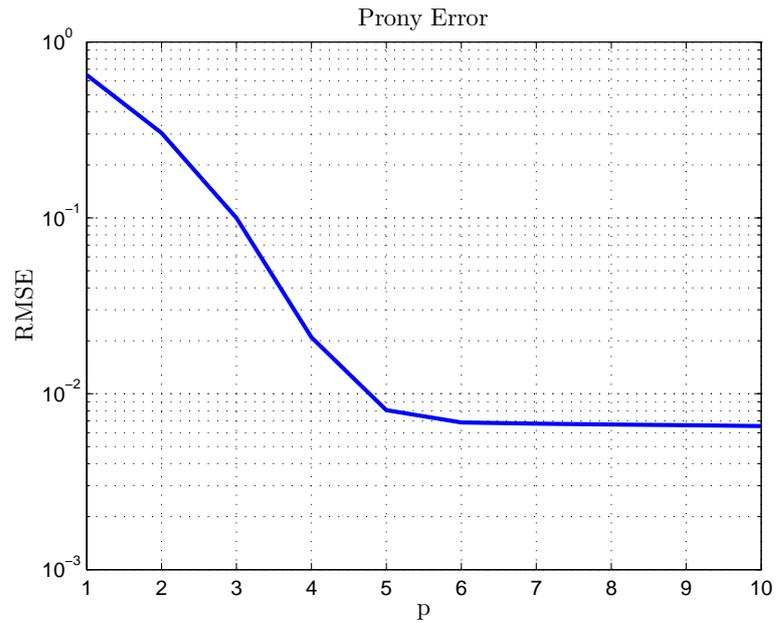
Figure 4.70: $e_{\mathrm{prony}}[n]$ vs p vs p, Scenario II
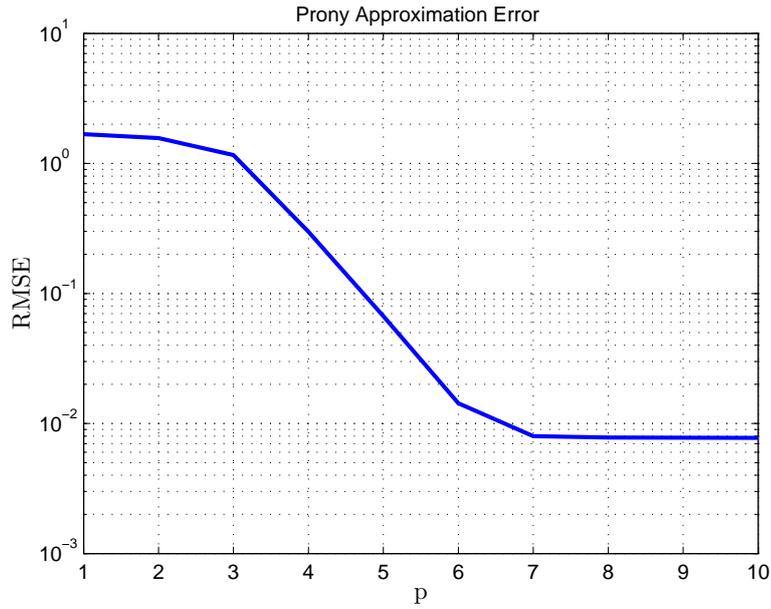
Figure 4.71: $\epsilon_{\mathrm{prony}}[n]$ vs p vs p, Scenario II
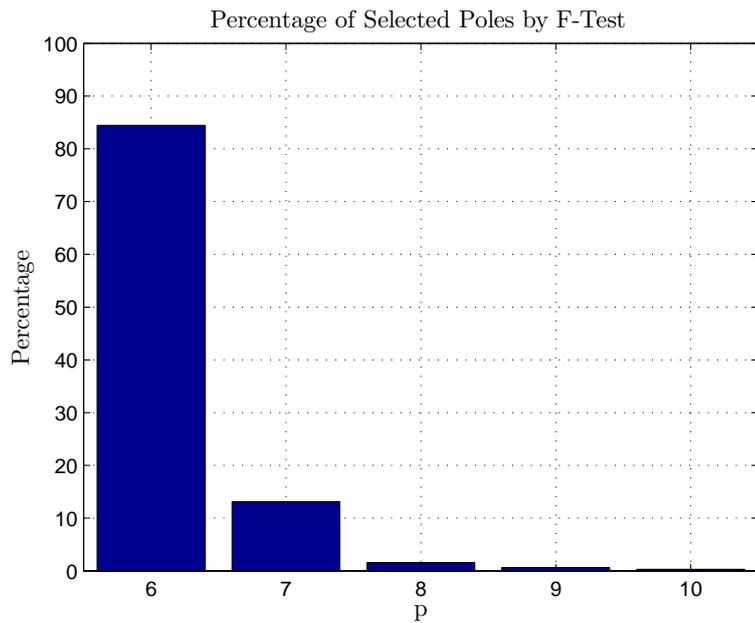
Figure 4.72: Selection Percentage of p by The Proposed Algorithm, Scenario II

4.4.1.3 Scenario III

Problem Definition

Parameters of $s[n]$ are given in Table 4.18. Notice that this is same signal introduced in Scenario I. Therefore, plot of $s[n]$ is not given again.

Table 4.18: Signal Parameters for Scenario III

Parameter	Value
Length	96
Number of Zeros (q)	0
Zero Location #0	1
Number of Poles (p)	5
Pole Location #1	-0.92
Pole Location #2	-0.4
Pole Location #3	-0.3
Pole Location #4	-0.1
Pole Location #5	0.4

Problem parameters are given in Table 4.19. The only difference form Scenario I is value of SNR, which is lower than Scenario I in this case. This scenario is given to show the effect of SNR on the performance.

Table 4.19: Problem Parameters for Scenario III

p_{max}	10
SNR	20 dB
p	0.1
MCNum	10000

Results

The first thing that is different from the previous scenarios is the changes in both Prony error and Prony approximation error with p. In this case errors do not change significantly with p as shown in Figures 4.73 and 4.74. Notice the limits of y-axes. As shown in Figure 4.75, the proposed method generally selects one pole model. This selection is logical if Figure 4.73 is considered. According to this figure, model with order 2 does not provide a "significant" decrease in RMSE than model with order 1. However, selected number of poles should be close to 5 in order to estimate the model order accurately. For this SNR value, the reduction in RSS is dominated by not only signal components but also noise components. Therefore, it is not possible to estimate the correct model order using RSS values.

As an example, one of the observed signals used during the simulation is shown in (4.75).
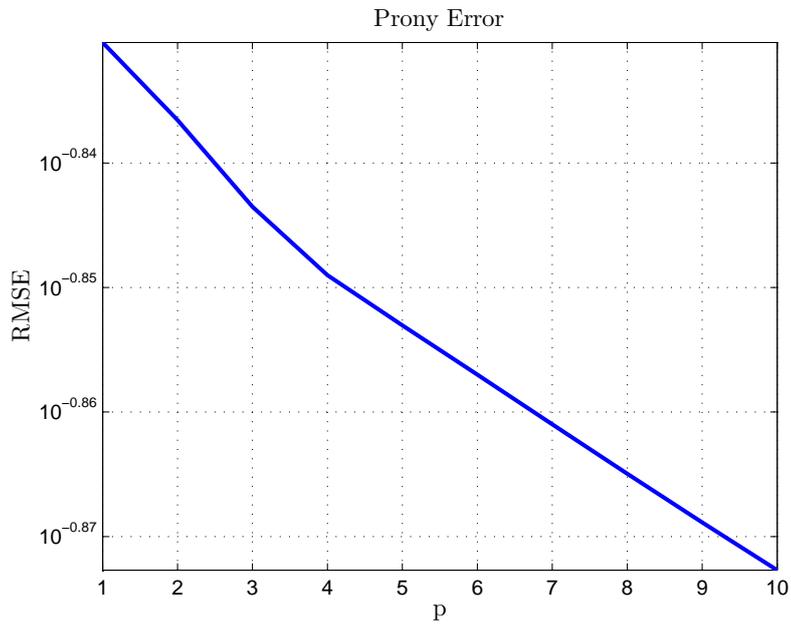
Figure 4.73: $e_{\text{prony}}[n]$ vs p, Scenario III
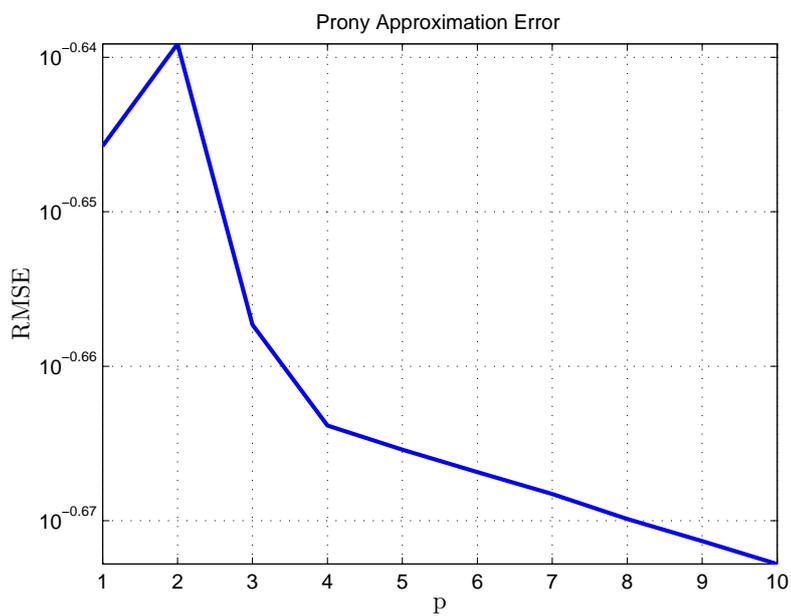
Figure 4.74: $\epsilon_{\text{prony}}[n]$ vs p, Scenario III
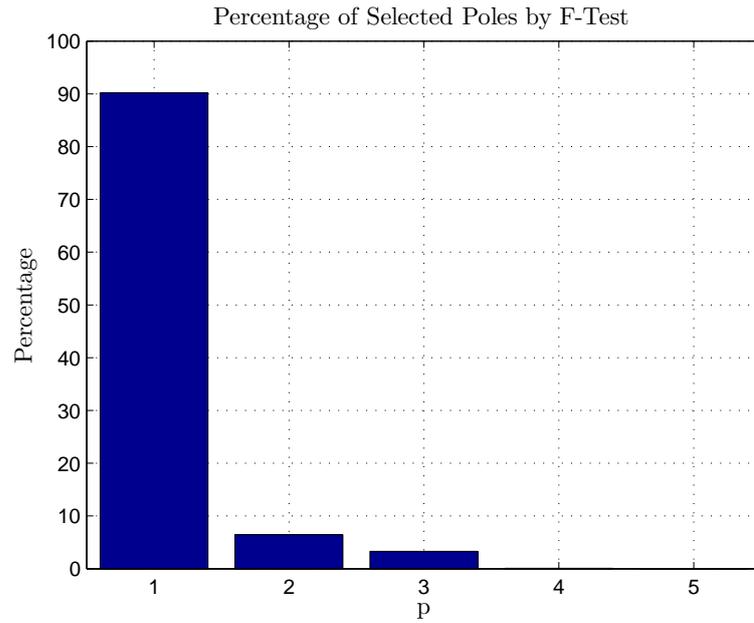
Figure 4.75: Selection Percentage of p by The Proposed Algorithm, Scenario III
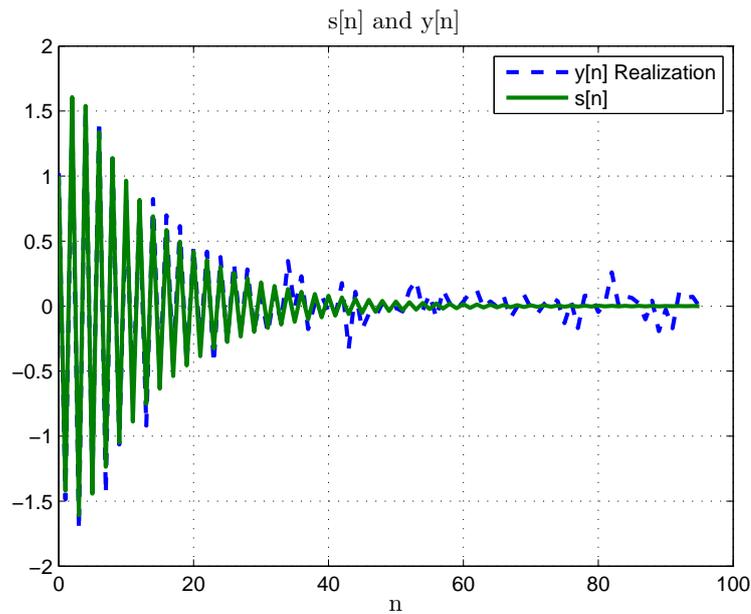
Figure 4.76: $s[n]$ and $y[n]$ Realization, Scenario III

4.5 Segmentation of Damped Sinusoidals

As a last example problem, segmentation of the impulse responses of casual LTI filters will be considered. Similar to segmentation of multi tone signals, impulse responses are concatenated over time. Noisy observation is segmented

into multiple segments such that each contains observation from a single impulse response. Segmented observations may be modeled later as all-pole signals explained in the previous example problem. However, this example problem is about only segmentation of observations. This is also an example problem for usage of the F-test based proposed method with other algorithms as it will be explained soon.

$s[n]$ is generated by summing impulse responses of H different casual LTI filters as given below

$$
s[n] = \begin{cases} \displaystyle\sum_{h=1}^{H} s_h(n, n_{1_h}, n_{2_h}, \boldsymbol{a}_{p_h}, \boldsymbol{b}_{q_h}, p_h, q_h) & 0 \le n \le N-1 \\ 0 & \text{otherwise,} \end{cases}
$$

$$
s_h(n, n_{1_h}, n_{2_h}, \boldsymbol{a}_{p_h}, \boldsymbol{b}_{q_h}, p_h, q_h) \triangleq \begin{cases} \mathcal{Z}^{-1}\{S_h(z)\} & n_{1_h} \le n \le n_{2_h} \\ 0 & \text{otherwise,} \end{cases}
$$

$$
S_h(z, \boldsymbol{a}_{p_h}, \boldsymbol{b}_{q_h}, p_h, q_h) = \frac{\displaystyle\sum_{k=0}^{q_h} b_q(k, h) z^{-k}}{1 + \displaystyle\sum_{k=1}^{p_h} a_p(k, h) z^{-k}},
$$

$$
\boldsymbol{a}_{p_h} \triangleq \begin{bmatrix} a_p(1, h) \\ a_p(2, h) \\ a_p(3, h) \\ \vdots \\ a_p(p_h, h) \end{bmatrix}_{p_h \times 1,}
$$

$$
\boldsymbol{b}_{q_h} \triangleq \begin{bmatrix} b_q(0, h) \\ b_q(1, h) \\ b_q(2, h) \\ b_q(3, h) \\ \vdots \\ b_q(g_h, h) \end{bmatrix}_{(q_h+1) \times 1,}
$$

$$
N_h \triangleq n_{2_h} - n_{1_h} + 1.
$$

$a_p(j, i)$ and $b_q(j, i)$ denote j^{th} pole and j^{th} zero of ith response, respectively. To demonstrate the structure of $s[n]$, an example is shown in Figure 4.77 where

$H = 4$. It was mentioned that all-pole modeling can also be used to model signals with number of zeros greater than one. Therefore, there is no constraint on number of zeros for any response.
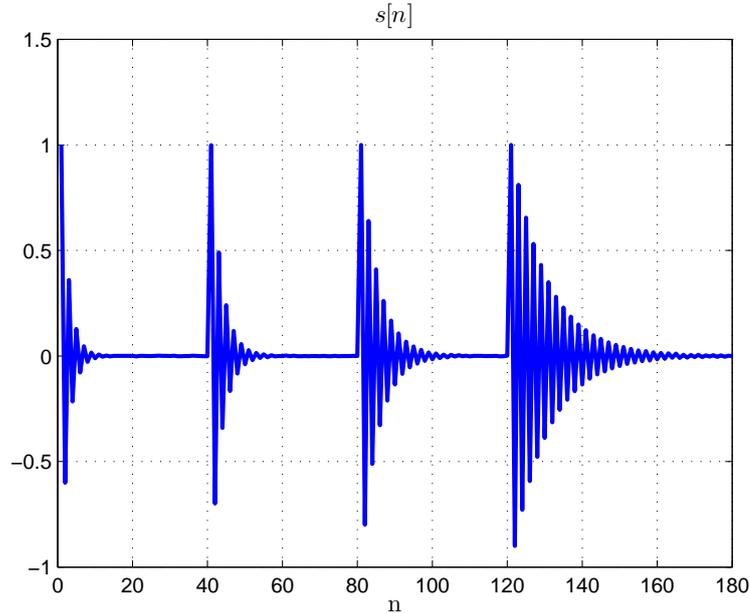
Figure 4.77: An Example $s[n]$ Signal

Segmentation is based on the difference between approximated signal using Prony's method and the observation signal. Since Prony's method approximates to the signal as a sum of damped sinusoidals, at starting point of the next segment the error between the approximation and the observation starts to increase. Segmentation is done by tracking this error which is not related with F-test. F-test is used as in the previous model order selection problem for Prony's method. Once the error reaches to *error_threshold*, which is different than *threshold* used in F-test, new segment starts. Observation in the new segment is modeled as all-pole signal using Prony's method with the help of the proposed F-test based number of poles selection approach. The segmentation algorithm is shown in Figure 4.78.

Blue block shown in Figure 4.78 represents the algorithm previously given in the number of pole selection problem. It represents the algorithm shown in Figure 4.64. It can be seen that the segmentation algorithm given for this problem is quite different than the algorithm given for multi tone signals. In

123
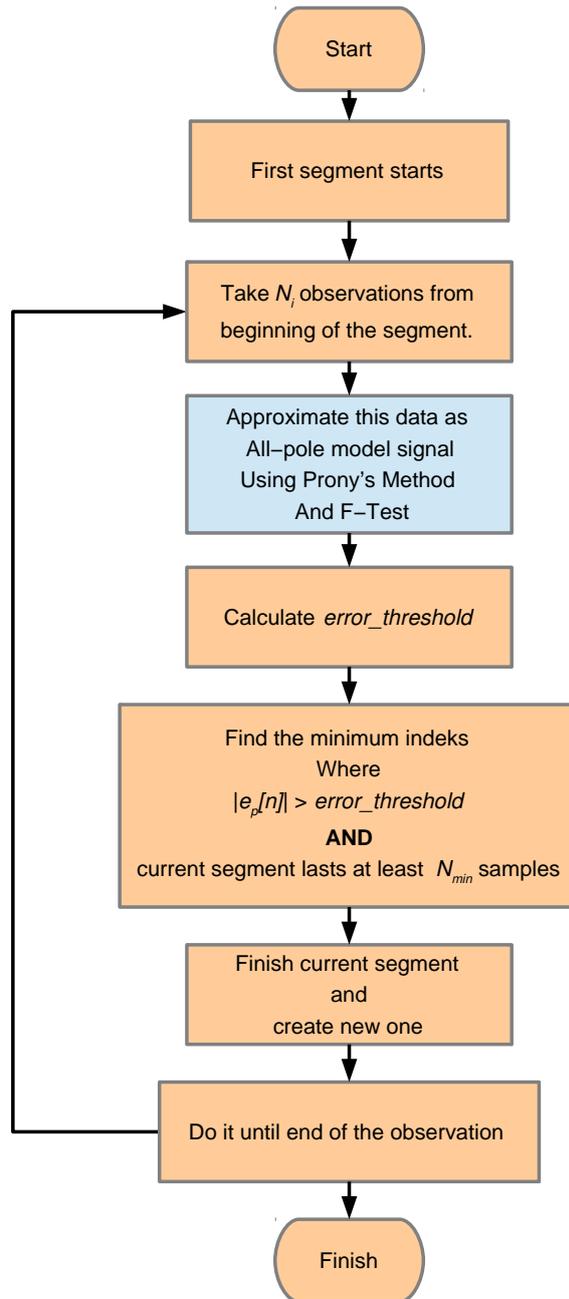
Figure 4.78: The Proposed Segmentation Algorithm

multi tone segmentation problem, power of tone does not decrease as search window moves. However in this case, power of damped sinusoidal decreases if a search window is used and shifted. Therefore, segmentation algorithm is not F-test based completely.

The algorithm works as follows: It is assumed that observation starts with a

noisy impulse response. N_i samples are selected from the beginning of the first segment. Using the proposed algorithm given in the previous problem, selected samples are modeled as all-pole signal with pole number selected by the proposed method. Prony approximation error which is defined in (4.29) is calculated for the rest of the observation. A threshold called *error_threshold* is calculated for segmentation.

To calculate *error_threshold*, absolute value of Prony approximation error is smoothed using a moving average filter with window length N_{MA}. Then, *error_threshold* is set as mean value of the first N_{min} samples. Minimum length of each segment is also taken as N_{min}. The lowest possible index where the absolute value of Prony approximation error is greater than *error_threshold* is found after segment lasts at least N_{min} samples. This index is taken as starting point of the new segment. The flow is repeated for the new segment and the segmentation continues until the end of the observation.

error_threshold is calculated as [1]

$$error\_threshold = \frac{1}{N_{min}} \sum_{k=0}^{N_{min}-1} \text{smooth}(|e_{\text{prony}}[n]|, N_{MA}). \qquad (4.32)$$

The following parameters are used in the algorithm shown in Figure 4.78.

- N_i ($N_{initial}$): Number of samples taken from the beginning of the each new segment to approximate to the observation using Prony's method.

- N_{min} ($N_{minimum}$): Minimum length of each segment. If there is a priori information about minimum separation between two consecutive filter responses, it may be used to set this parameter.

- N_{MA} ($N_{Moving\ Average}$): Length of the moving window used to smooth out Prony error as given in (4.32).

- p_{max}: Limits the maximum number of poles used in Prony's method. One may want to set an upper limit for $\hat{p}$ if there is a priori information about

---

[1] smooth() function is used to demonstrate moving average filter. Indeed, this syntax is used same as smooth() function of Curve Fitting Toolbox of MATLAB®. For detailed information please see the MATLAB® R2015a documentation or visit: `http://www.mathworks.com/help/curvefit/smooth.html`

$s[n]$. It is used in blue box shown in Figure 4.78. It is the same parameter introduced in the previous problem.

4.5.1 Simulation Results

4.5.1.1 Scenario I

Problem Definition

In this scenario H is selected as 4. Parameters for each impulse response are given in Table 4.20.

Table 4.20: Signal Parameters for Scenario I

Parameter	$s_1[n]$	$s_2[n]$	$s_3[n]$	$s_4[n]$
Starting Index	0	40	80	120
End Index	39	79	119	179
Length	40	40	40	60
Number of Zeros (q)	0	0	0	0
Zero Location #0	1	1	1	1
Number of Poles (p)	1	1	1	1
Pole Location #1	-0.6	-0.7	-0.8	-0.9

Problem parameters are given in Table 4.23.

Table 4.21: Problem Parameters for Scenario I

Parameter	Value
p_{max}	10
SNR	60 dB
p_{fd}	0.1
N_i	10
N_{MA}	20
N_{min}	30

Results

Indices of ideal segment borders should be 40, 80, 120. The proposed method estimates these values as: 39, 179, 119. It finds the segment borders almost

exactly. The segment borders are shown in Figure 4.79.
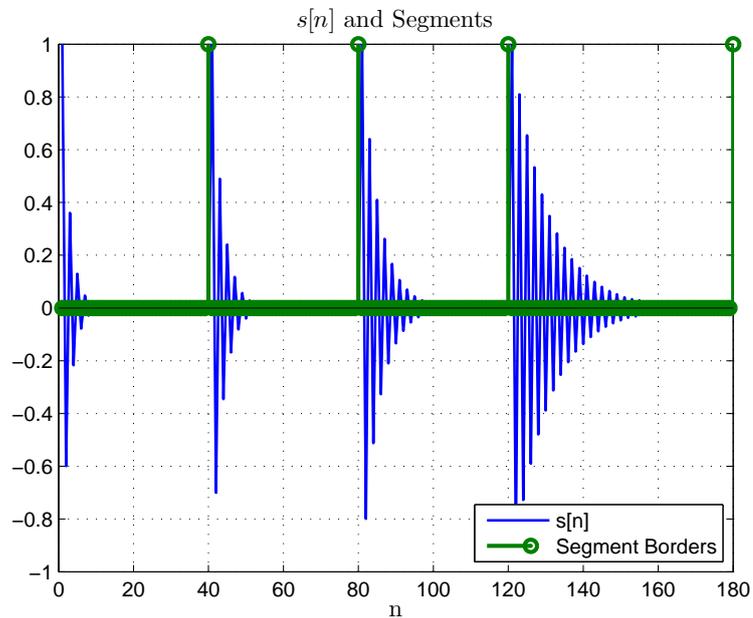
Figure 4.79: $s[n]$ and Borders of Dynamic Segments Selected by The Proposed Algorithm, Scenario I

4.5.1.2 Scenario II

Problem Definition

In this scenario H is selected as 4. Parameters for each impulse response are given in Table 4.22.

Table 4.22: Signal Parameters for Scenario II

Parameter	$s_1[n]$	$s_2[n]$	$s_3[n]$	$s_4[n]$
Starting Index	0	75	172	269
End Index	74	171	268	367
Length	75	97	97	99
Number of Zeros (q)	0	0	0	2
Zero Location #0	1	1	1	1
Zero Location #1	-	-	-	-0.9 + 0.7j
Zero Location #2	-	-	-	-0.9 - 0.7j
Number of Poles (p)	6	4	5	2
Pole Location #1	-0.2 + 0.5j	-0.2 + 0.1j	-0.92	-0.9 + 0.01j
Pole Location #2	-0.2 - 0.5j	-0.2 - 0.1j	-0.4	-0.9 - 0.01j
Pole Location #3	0.1	-0.7 - 0.6j	-0.3	-
Pole Location #4	-0.1	-0.7 + 0.6j	-0.1	-
Pole Location #5	-0.6 + 0.7j	-	-0.4	-
Pole Location #6	-0.6 - 0.7j	-	-	-

Problem parameters are given in Table 4.23.

Table 4.23: Problem Parameters for Scenario II

Parameter	Value
p_{max}	10
SNR	60 dB
p_{fd}	0.1
N_i	10
N_{MA}	20
N_{min}	50

Results

Indices of ideal segment borders should be 75, 172, 269. The proposed method estimates these values as: 74, 171, 271. It finds the segment borders with relatively small errors. The segment borders are shown in Figure 4.80.

128
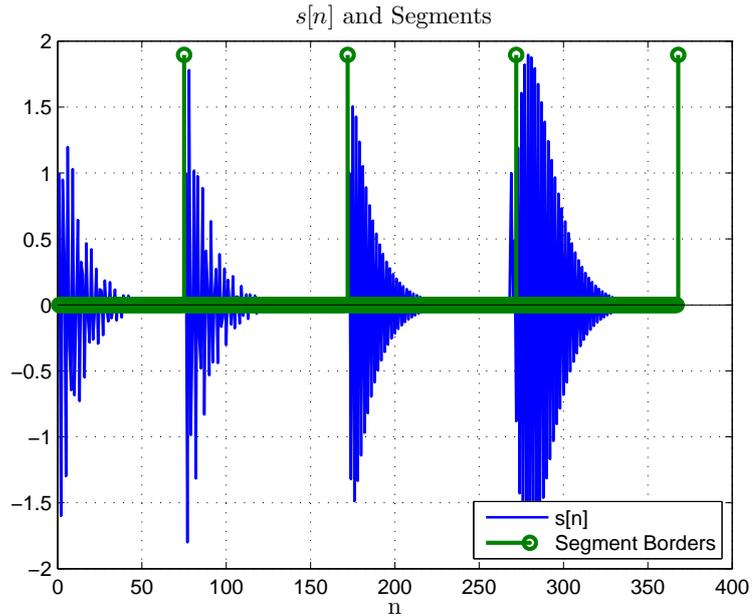
Figure 4.80: $s[n]$ and Borders of Dynamic Segments Selected by The Proposed Algorithm, Scenario II

4.5.1.3 Scenario III

Problem Definition

As a last example, a real world data is taken from p. 133 of [19]. Raw data can be found in the CD included in [19] with the name *FSSP3exer5_2.mat*. This data is taken from output of an accelerometer of a faulty bearing machine. Since data is not synthetically generated, it is not possible to give a mathematical representation. Instead, a plot is given in Figure 4.81.

Problem parameters are given in Table 4.24.

Table 4.24: Problem Parameters for Scenario III

Parameter	Value
p_{max}	10
SNR	?
p_{fd}	0.4
N_i	10
N_{MA}	20
N_{min}	100

129
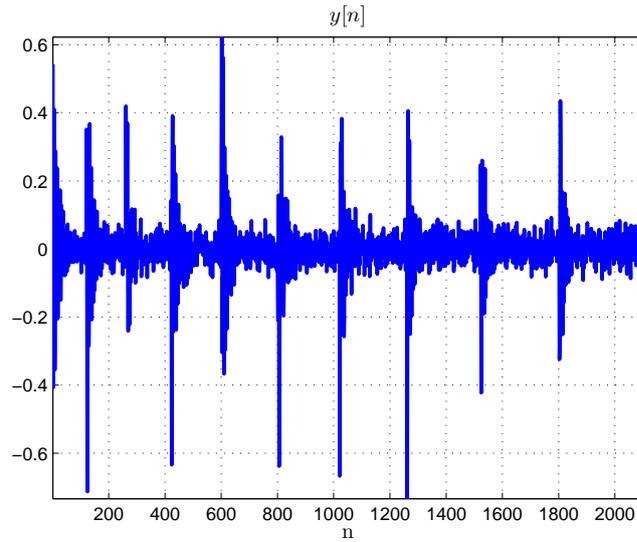
Figure 4.81: Plot of $y[n]$

Although expression and characteristics of noise aren't known, proposed method is applied anyway.

Results

Since the mathematical expression of the actual signal is unknown, results can be only seen visually in Figure 4.82. Except few false decisions, segments seem to be at correct positions.
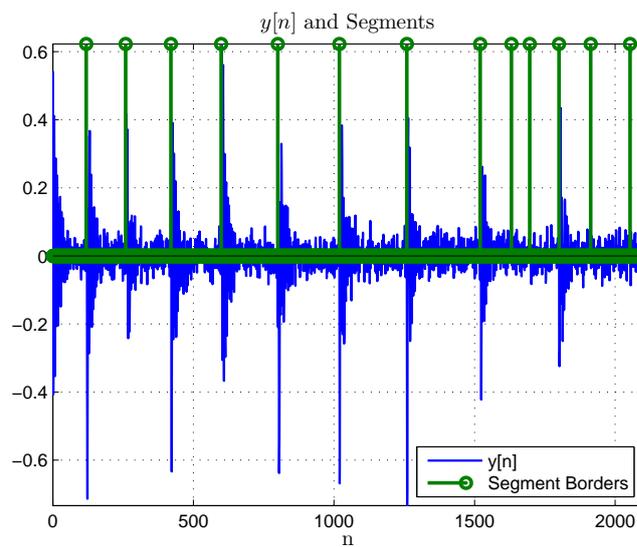
Figure 4.82: $y[n]$ and Borders of Dynamic Segments Selected by The Proposed Algorithm, Scenario III

CHAPTER 5

CONCLUSIONS

In this thesis, different F-test based methods are proposed for model order selection related problems. F-test is studied in the statistical literature whereas it is not equally used in the signal processing area. The fundamentals of F-test are adapted to signal processing problems in this work.

Initially, the necessary background information is given. After giving necessary information, the proposed methods are supported by example problems. In each problem, the proposed algorithms are shown using flow charts and supported by simulations. The effects of SNR and F-test parameters as p_{fd} on the performance are studied in some simulations.

In general, F-test based methods looks suitable for some signal processing problems, provided that the signal can be indeed modeled as an element of a linear space.

One of the main advantages of F-test is that it does not require noise variance information. On the other hand, SNR should be relatively high to estimate the model orders accurately.

The F-test based approaches also look suitable for embedded and real time applications. The most expensive operation is the calculation of $\boldsymbol{A}^+$ matrix which involves a matrix inversion operation. Apart form this calculation, application of F-test is quite easy to implement.

Future Work

Although the effects of p_{fd} and SNR are shown in some simulations, their effects aren't analyzed in detail. It is shown and explained that both of them affect the performance of the suggested methods directly. For a given problem, an optimum selection of p_{fd} parameter may be analyzed. Also the SNR threshold beyond which F-test works well can be studied.

REFERENCES

[1] H. Akaike. A new look at the statistical model identification. *IEEE Trans. Automatic Control*, 19(6):716–723, Dec 1974.

[2] O. Alan, O. Schafer, and S. W. *Discrete Time Signal Processing*. Prentice-Hall Of India Pvt. Limited, 1996.

[3] E. Alpaydım. Combined 5 x 2 cv f test for comparing supervised classification learning algorithms. *Neural Computation*, 11(8):1885–1892, Nov 1999.

[4] T. Amemiya. *Advanced Econometrics*. Harvard University Press, 1985.

[5] B. Atal and J. Remde. A new model of lpc excitation for producing natural-sounding speech at low bit rates. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '82.*, volume 7, pages 614–617, May 1982.

[6] H. Bozdogan. Akaike's Information Criterion and Recent Developments in Information Complexity. *Journal of mathematical psychology*, 44(1):62–91, 2000.

[7] F. Brophy and A. Salazar. Considerations of the pade approximant technique in the synthesis of recursive digital filters. *Audio and Electroacoustics, IEEE Transactions on*, 21(6):500–505, Dec 1973.

[8] D.-H. Chen, Y.-C. Chang, P.-J. Huang, and C.-H. Wei. The Correlation Analysis between Breast Density and Cancer Risk Factor in Breast MRI Images. In *2013 International Symposium on Biometrics and Security Technologies*, pages 72–76, 2013.

[9] S. S. Chen and P. Gopalakrishnan. Clustering via the bayesian information criterion with applications in speech recognition. In *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, volume 2, pages 645–648 vol.2, May 1998.

[10] P.-J. Chung, J. F. Böhme, A. O. Hero, and C. F. Mecklenbräuker. Detection of the number of signals using a multiple hypothesis test. In *Sensor Array and Multichannel Signal Processing Workshop Proceedings, 2004*, pages 221–224. IEEE, 2004.

[11] R. Churchill and J. Brown. *Complex Variables and Applications*. Tan Chiang Book Company, 1984.

[12] J. Deller, J. Hansen, and J. Proakis. *Discrete-Time Processing of Speech Signals*. An IEEE Press classic reissue. Wiley, 2000.

[13] M. Dorozhovets. Using f - test for the indirect detecting of the autocorrelation of random observations. In *Intelligent Data Acquisition and Advanced Computing Systems (IDAACS), 2013 IEEE 7th International Conference on*, volume 01, pages 143–148, Sept 2013.

[14] M. Hayes. *Statistical digital signal processing and modeling*. John Wiley & Sons, 1996.

[15] S. Haykin. *Adaptive Radar Signal Processing*. Wiley, 2007.

[16] Q. Huang and P.-J. Chung. An F-Test Based Approach for Spectrum Sensing in Cognitive Radio. *Wireless Communications, IEEE Transactions on*, 12(8):4072–4079, 2013.

[17] N. Johnson, S. Kotz, and N. Balakrishnan. *Continuous univariate distributions*. Number 2. c. in Wiley series in probability and mathematical statistics: Applied probability and statistics. Wiley & Sons, 1995.

[18] S. Kay. *Fundamentals of Statistical Signal Processing: Estimation theory*. Fundamentals of Statistical Signal Processing. Prentice-Hall PTR, 1993.

[19] S. Kay. *Fundamentals of Statistical Signal Processing: Practical algorithm development*. Fundamentals of Statistical Signal Processing. Prentice-Hall PTR, 2013.

[20] S. Konishi and G. Kitagawa. Generalised information criteria in model selection. *Biometrika*, 83(4):pp. 875–890, 1996.

[21] J. Lim and A. Oppenheim. *Advanced topics in signal processing*. Prentice-Hall signal processing series. Prentice-Hall, 1988.

[22] R. Lomax and D. Hahs-Vaughn. *Statistical Concepts: A Second Course*. Taylor & Francis, 2013.

[23] R. Miller. *Beyond ANOVA: Basics of Applied Statistics*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 1997.

[24] G. E. Mog and E. P. Ribeiro. Zero crossing determination by linear interpolation of sampled sinusoidal signals. In *Transmission and Distribution Conference and Exposition: Latin America, 2004 IEEE/PES*, pages 799–802. IEEE, 2004.

[25] B. Porat. *A course in digital signal processing*. John Wiley, 1997.

[26] W. Pu and X.-F. Niu. Selecting mixed-effects models based on a generalized information criterion. *Journal of Multivariate Analysis*, 97(3):733–758, 2006.

[27] L. Rabiner and R. Schafer. *Digital Processing of Speech Signals*. Prentice-Hall signal processing series. Prentice-Hall, 1978.

[28] M. Richards. *Fundamentals of Radar Signal Processing*. Professional Engineering. McGraw-Hill Education, 2005.

[29] H. Scheffé. *The Analysis of Variance*. A Wiley publication in mathematical statistics. Wiley, 1999.

[30] G. Schwarz et al. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.

[31] J. Shao. An asymptotic theory for linear model selection. *Statistica Sinica*, 7(2):221–242, 1997.

[32] J. Stewart. *Calculus*. Cengage Learning, 2011.

[33] P. Stoica and Y. Selen. Model-order selection: a review of information criterion rules. *IEEE Signal Process. Mag.*, 21(4):36–47, July 2004.

[34] O. Vanjukevich and A. Popov. The f-test by testing of hypotheses about structure variations in the fuzzy regression models. In *Science and Technology, 2005. KORUS 2005. Proceedings. The 9th Russian-Korean International Symposium on*, pages 104–106, June 2005.

[35] A. Yazar and C. Candan. Analysis window length selection for linear signal models. In *Signal Processing and Communications Applications Conference (SIU), 2015 23th*, pages 1301–1304, May 2015.

[36] B. Zhou and J. H. L. Hansen. Efficient audio stream segmentation via the combined T2 statistic and bayesian information criterion. *IEEE Transactions on Speech and Audio Processing*, 13(4):467–474, 2005.

APPENDIX A

CRAMÉR–RAO LOWER BOUND FOR
ZERO-CROSSING POINT ESTIMATION

Consider the signal given as

$$x[n] = a(n - n_0) = an + b, \quad n = -\frac{N-1}{2}, \dots, -1.5, -0.5, 0.5, 1.5, \dots, \frac{N-1}{2}.$$
$$\text{(A.1)}$$

n_0 is the zero-crossing point of $x[n]$. a and b, in other words n_0, are unknown parameters of the signal. Similar to example problem about zero-crossing point estimation problem, it is assumed that n_0 can take any real value between -0.5 and 0.5. At this section, CRLB will be derived for $\hat{n}_0$.

This problem is similar to CRLB calculation for vector parameter case [18].

N is an integer. $x[n]$ is observed under AWGN ($w[n]$) as

$$y[n] = x[n] + w[n] = An + B + \omega[n].$$

This can be written as

$$\boldsymbol{y}_{N\times1} = \boldsymbol{A}_{N\times2}\boldsymbol{p}_{2\times1} + \boldsymbol{w}_{N\times1},$$

in vector notation where

$$\boldsymbol{p} = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix},$$

$$\boldsymbol{A} = \begin{bmatrix} -\frac{N-1}{2} & 1 \\ \vdots & \vdots \\ -0.5 & 1 \\ 0.5 & 1 \\ \vdots & \vdots \\ \frac{N-1}{2} & 1 \end{bmatrix}.$$

Fisher information matrix $(\boldsymbol{I}(\boldsymbol{p}))$ can be written as

$$\boldsymbol{I}(\boldsymbol{p}) = \begin{bmatrix} -E\left[\dfrac{\partial^2 \ln p(\boldsymbol{y};\boldsymbol{p})}{\partial a^2}\right] & -E\left[\dfrac{\partial^2 \ln p(\boldsymbol{y};\boldsymbol{p})}{\partial a \partial b}\right] \\ -E\left[\dfrac{\partial^2 \ln p(\boldsymbol{y};\boldsymbol{p})}{\partial b \partial a}\right] & -E\left[\dfrac{\partial^2 \ln p(\boldsymbol{y};\boldsymbol{p})}{\partial b^2}\right] \end{bmatrix}.$$

The likelihood function is given as

$$p(\boldsymbol{y};\boldsymbol{p}) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} (y[n] - an - b)^2 \right\}$$

and

$$\frac{\partial \ln p(\boldsymbol{y};\boldsymbol{p})}{\partial a} = \frac{1}{\sigma^2} \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} (y[n] - an - b)n,$$

$$\frac{\partial^2 \ln p(\boldsymbol{y};\boldsymbol{p})}{\partial a^2} = -\frac{1}{\sigma^2} \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} n^2, \qquad (A.2)$$

$$\frac{\partial \ln p(\boldsymbol{y};\boldsymbol{p})}{\partial b} = \frac{1}{\sigma^2} \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} (y[n] - an - b),$$

$$\frac{\partial^2 \ln p(\boldsymbol{y};\boldsymbol{p})}{\partial b^2} = -\frac{N}{\sigma^2},$$

$$\frac{\partial^2 \ln p(\boldsymbol{y};\boldsymbol{p})}{\partial a \partial b} = -\frac{1}{\sigma^2} \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} n = 0,$$

$$\frac{\partial^2 \ln p(\boldsymbol{y};\boldsymbol{p})}{\partial b \partial a} = -\frac{1}{\sigma^2} \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} n = 0.$$

Consider summation term given (A.2) as follows;

$$
\sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} n^2 = 2 \sum_{n=\frac{1}{2}}^{\frac{N-1}{2}} n^2
$$

$$
= 2 \sum_{k=0}^{\frac{N}{2}-1} \left(k + \frac{1}{2} \right)^2, \qquad k = n - \frac{1}{2}
$$

$$
= 2 \sum_{l=1}^{\frac{N}{2}} \left(l - \frac{1}{2} \right)^2, \qquad l = 1 + k
$$

$$
= \frac{1}{2} \sum_{l=1}^{\frac{N}{2}} (2l-1)^2, \qquad l = 1 + k
$$

$$
= \frac{1}{2} \sum_{l=1}^{M} (2l-1)^2, \qquad M = \frac{N}{2}
$$

$$
= 2 \sum_{l=1}^{M} l^2 - 2 \sum_{l=1}^{M} l + \frac{M}{2}
$$

$$
= \frac{M(M+1)(2M+1)}{3} - M(M+1) + \frac{M}{2}
$$

$$
= \frac{M}{6} \left[4M^2 - 1 \right]
$$

$$
= \frac{N(N^2 - 1)}{12}
$$

and then

$$
\frac{\partial^2 \ln p(\boldsymbol{y}; \boldsymbol{p})}{\partial a^2} = \frac{1}{\sigma^2} \frac{N(N^2 - 1)}{12}
$$

can be written.

Now the Fisher information matrix can be written as

$$
\boldsymbol{I}(\boldsymbol{p}) = \frac{1}{\sigma^2} \begin{bmatrix} \dfrac{N(N^2-1)}{12} & 0 \\[2em] 0 & N \end{bmatrix}.
$$

Finally using (A.1), the followings can be written,

$$
g(\boldsymbol{p}) \triangleq n_0 = -\frac{b}{a} = -\frac{p_2}{p_1},
$$

$$
\frac{\partial g(\boldsymbol{p})}{\partial \boldsymbol{p}} = \left[\frac{\partial g(\boldsymbol{p})}{\partial p_1} \quad \frac{\partial g(\boldsymbol{p})}{\partial p_2} \right] = \left[\frac{b}{a^2} \quad -\frac{1}{a} \right],
$$

139

$$\frac{\partial g(\boldsymbol{p})}{\partial \boldsymbol{p}} \boldsymbol{I}^{-1}(\boldsymbol{p}) \frac{\partial g(\boldsymbol{p})^{\mathrm{T}}}{\partial \boldsymbol{p}} = \frac{\sigma^2}{Na^2} \left(1 + \frac{12n_0^2}{N^2 - 1} \right),$$

$$\mathrm{var}\{\hat{n}_0\} \geq \frac{\sigma^2}{Na^2} \left(1 + \frac{12n_0^2}{N^2 - 1} \right).$$