TIME-FREQUENCY ANALYSIS AND DESIGN OF SIGNALS: THE WIGNER DISTRIBUTION AND ITS APPLICATIONS

By

QU JIN, B.Sc.(Eng.), and M.Eng. (Dalian Maritime University)

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Doctor of Philosophy

McMaster University

October 1991

Dedicated To All Members In My Lovely Family

TIME-FREQUENCY ANALYSIS AND DESIGN OF SIGNALS: THE WIGNER DISTRIBUTION AND ITS APPLICATIONS

DOCTOR OF PHILOSOPHY (1991) MCMASTER UNIVERSITY

(Electrical and Computer Engineering) Hamilton, Ontario

TITLE: **Time-Frequency Analysis and Design of Signals: The Wigner Distribution and Its Applications**

AUTHOR: Qu Jin

B.Sc.(Eng.), and M.Eng. (Dalian Maritime University)

SUPERVISOR(S): Dr. Kon Max Wong

Professor, Chairman of Department of Electrical and Computer Engineering

B.Sc.(Eng.), Ph.D., (University of London)

Dr. Tom Zhi-quan Luo

Assistant Professor of Department of Electrical and Computer Engineering

B.A.Sc. (Beijing University)

Ph.D. (Massachusetts Institute of Technology)

NUMBER OF PAGES: xvii, 234

ABSTRACT

Time-frequency representations provide us with an efficient way to process non-stationary signals and systems with time-dependent spectrum. In this thesis, a tutorial review of several time-frequency representations is presented. The historical development and the properties are discussed with the emphasis being placed on the Wigner distribution (WD).

Three important applications of time-frequency representations are studied. They are the tracking of instantaneous frequency, high resolution estimation of frequencies in the non-stationary environment, and the design of signals for optimum simultaneous estimation of time delay and Doppler shift.

An efficient way to calculate the Cramér-Rao lower bound for the time-varying frequency estimation is obtained. The theoretical performance of the WD when applied to the estimation of this time-varying frequency is derived and the simulation examples are provided.

Using the properties of the instantaneous spectrum, new algorithms are developed which can accurately estimate the frequencies of the non-stationary signals with fast varying amplitude. A particular example of the applications is the frequency estimation in Search and Rescue Satellite (SARSAT). For tests on real SARSAT signals, our new algorithms exhibit superior performance to other spectrum estimation methods.

The Cramér-Rao bound (CRB) for the joint time delay and Doppler shift estimation

is derived and it is proved that the joint estimates of these two parameters using time-frequency representations are optimum under high SNR in the sense that it is unbiased and the variance reaches the CRB. Several efficient methods are developed to obtain the optimum signal for the joint time delay and Doppler shift estimation under different practical constraints.

ACKNOWLEDGEMENT

The author wishes to express sincere gratitude to Dr. K.M.Wong for his constant encouragement, continued assistant and expert guidance and supervision throughout the course of this work. I would like also to thank Dr. T.Z.Luo for his helpful insights and guidance in the work of Chapter 6 and 7. Thanks are also due to Drs. C.R.Carter, G.Slade and P.Yip, members of the Supervisory Committee, for their continued interest and useful suggestions.

Acknowledgement is made to Dr. T.Chung of Communication Research Laboratory, McMaster University, who supplyed the real SARSAT data so that my work in Chapter 5 can be employed in practical situations. In addition, the author appreciates the constructive criticisms of the reviewers of IEEE transactions.

It is the author's pleasure to acknowledge the inspiring discussions with his colleagues Dr. H.Dai, W.Chen, X.Huang and X.Gu.

Many thanks go to other friends for their valuable help.

I would also like to acknowledge the financial support provided by National Science and Engineering Research Council, Department of Electrical and Computer Engineering (McMaster University) and Telecommunication Research Institute of Ontario.

Finally, I am indebted to my wife, Jingyun Li, for her constant understanding, encouragement and support.

Contents

ABSTRACT iii

ACKNOWLEDGEMENT v

1 Introduction 1

1.1 A Stationary Process and Its Spectrum 1

1.2 Non-Stationary Signals and Time-Dependent Spectrum 3

1.3 Time-Frequency Representations 5

1.4 Outline of Thesis 5

2 A Brief Review On Time-Frequency Representations 9

2.1 Instantaneous Frequency and Group Delay 10

2.2 The Spectrogram 12

2.3 Ambiguity Function (AF) 13

2.4 Wigner Distribution (WD) 13

3 Lower Bound on the Estimate of Time-Varying Parameter of a Signal 23

3.1 A Lower Bound On the Mean-Square Error in Estimating a Continuous Random Parameter 24

3.2 The Cramér-Rao Bound for the Frequency Estimation of the Signal . . . 35

3.2.1 Estimation of a continuous Non-Random Parameter 35

3.2.2 The Cramér-Rao Bound for the Frequency Estimation of the Signal 37

4 Estimation of Time-Varying Frequency of a Signal 41

4.1 The Use of WD and PWD to Estimate the instantaneous Frequency of a Signal 41

4.2 Theoretical Performance Analysis 44

4.2.1 Bias and Variance Analysis in the Application of WD to Estimate the Instantaneous Frequency of a Signal 44

4.2.2 Bias and Variance Analysis in the Application of PWD to Estimate the Instantaneous Frequency of a Signal 47

4.3 Comparison of the methods of frequency estimation 50

4.3.1 Discussions on the Theoretical Evaluation of Performance of the PWD 50

4.3.2 Evaluation of Performances by Computer Simulation 57

4.3.3 Comparison of Theoretical and Simulated Performance 58

4.3.4 Comparison with Spectrogram and Discussions 59

5 Accurate Frequency Estimation of SARSAT Signal 69

5.1 Signal Model 69

5.2 Application of PWD in the SARSAT Signal Processing 75

5.2.1 Problem of Direct Use of PWD for the Frequency Estimation . . . 75

5.2.2 The Properties of Averaged PWD 76

5.2.3 Frequency Estimation Using the instantaneous frequency property of PWD 80

5.2.4 Real ELT Signal Test 83

5.3 High Resolution Method for More than One ELT Signals 89

5.3.1 Frequency Estimation combing Burg algorithm and Average PWD 89

5.3.2 Estimation Results with Real and Simulated ELT Signals 91

5.4 Discussions and Comparisons 96

6 Simultaneous Time Delay and Doppler Shift Estimation 99

6.1 Time-Frequency Correlating Receiver 99

6.2 Joint Estimation of Time Delay and Doppler Shift Using WD 103

6.3 Performance Analysis for the WD 106

6.4 The Cramér-Rao Bound of the Joint Estimation 114

6.5 Performance Analysis and Optimum Signal 117

6.5.1 Simulation Results 117

6.5.2 Performance Analysis and the Optimum Signal 119

7 Optimum Signal Design and Uncertainty Principle 129

7.1 Selection of the Optimum Basis 129

7.1.1 Prolate Spheroidal Wave Function and Uncertainty in Fourier Analysis 129

7.1.2 A New Basis for General Time- and Band-Limited Function 134

7.1.3 Conclusion 149

7.2 Lower Bound of $Var(t)Var(f)$ and Optimum Signal 149

7.2.1 Condition for a Minimum $Var(\hat{t}_0)Var(\hat{f}_\Delta)$ 150

7.2.2 The Signal Design Region 152

7.2.3 Optimum Signal in the Unconstrained Design Region 156

7.3 Optimum Signal Design under Practical Constrains 168

7.3.1 Quasi-Simulated Annealing (QSA) Method of Optimization 169

7.3.2 Iterative Signal Synthesis (ISS) Using WD 172

7.3.3 Design Examples and performance of the Optimum Signals 174

8 Conclusion 187

A Gabor Function and the Suitable Window for the Spectrogram 189

B Proof of Eq.(3.45) 193

C Simplification of Eq.(3.51) 195

D Evaluation of the Variance of the Estimated Instantaneous Frequency Using WD 197

E Evaluation of the Bias and Variance in the Application of Spectrogram 201

F Auto-Regression (AR) Model and Burg Algorithm 207

G Evaluation of the Probability of Erroneous Estimations 211

H Evaluation of $\frac{\partial F}{\partial c}$ in Eq.(7.66) 213

I Proof of Theorem 7.5 215

J Proof of Theorem 7.7 219

List of Tables

1.1 Table 1.1 Generating Function $g(\zeta,\tau)$ and Their Time-Frequency Representations $P_r(t,f)$ 6

5.1 The Common Specifications in the 121.5MHz SARSAT System . 72

5.2 The Estimated Frequency For Real ELT Signal: Block Estimation 87

5.3 The Estimated Frequency For Real ELT Signal: Average of Blocks 88

5.4 The Maximum Frequency Variation Using the Block Estimation 89

7.1 Comparison of the Parameters of the Two Optimum Signals $\hat{a}_{o.2}(t)$ and $\hat{a}_{o.3}(t)$ Obtained by the QSA 178

7.2 Comparison of the Parameters of the Two Optimum Signals $\tilde{a}_{o.2}(t)$ and $\tilde{a}_{o.3}(t)$ Obtained by the ISS Method 179

List of Figures

4.1 The Effects of a_2 on the Estimation Bias of the PWD Estimate for the Quadratic FM Signal. 52

4.2 The Effects of a_2 on the Estimation MSE of the PWD Estimate for the Quadratic FM Signal. 53

4.3 The Relation between the MSE of Frequency Estimation using PWD and the Window Size for Linear FM. 55

4.4 The Relation between the MSE of Frequency Estimation Using PWD and the Window Size for Quadratic FM. 56

4.5 The Theoretical and Simulated Bias for the Linear FM Signal. 60

4.6 The Theoretical and Simulated Bias for the Quadratic FM Signal. 60

4.7 The Theoretical and Simulated MSE for the Linear FM Signal. 61

4.8 The Theoretical and Simulated MSE for the Quadratic FM Signal. 61

4.9 The Topographic Plots of the PWD 64

4.10 The Topographic Plots of the PWD under SNR= -5dB and 1 snapshot . . 66

4.11 The Topographic Plots of the PWD under SNR= -5dB and averaging of 15 snapshots 67

5.1 Flat-Earth Model of the SARSAT Geometry. 70

5.2 Illustration of Typical ELT Signal Model in the N Pulse Series. 71

5.3 Non-Coherent ELT Signal with "ON" Time as a Solid Line and "OFF" Time a Dotted Line (Notice the phase discontinuity). 73

5.4 The envelop of ELT signal. 78

5.5 The envelop of $\overline{E}_s(\tau)$. 78

5.6 Real ELT Signals in Time Domain 84

5.7 High Resolution ELT Frequency Estimation for NARELT19. 91

5.8 High Resolution ELT Frequency Estimation for NARELT17. 92

5.9 High Resolution ELT Frequency Estimation for GARELT12. 92

5.10 High Resolution ELT Frequency Estimation for PNTELT01. 93

5.11 Simulated two ELT signals with $f_{c1} = 14KHz$ and $f_{c2} = 16KHz$, d=.5 and SNR=20dB. 94

5.12 High Resolution ELT Frequency Estimation for Simulated Two ELT Signals: $f_{c1} = 16KHz$ and $f_{c2} = 14KHz$, with variable SNR. 95

5.13 High Resolution ELT Frequency Estimation for Simulated ELT Signals: $f_c = 16KHz$. 95

5.14 The Spectrum Obtained by Directly Applying MEM on the Real ELT Signal NARELT17 97

6.1 Simultaneous Estimation of Time Delay and Doppler Shift Using Joint Time-Frequency Representation. 101

6.2 Mean Square Error of the Time Delay Estimation verse ρ_{SN} for Two Commonly Used Signals. 119

6.3 Mean Square Error of the Doppler Shift Estimation verse ρ_{SN} for Two Commonly Used Signals. 120

6.4 The Magnitude of the WD for the Gaussian pulse. 121

6.5 The magnitude of the WD for the Linear FM pulse. 121

6.6 The Relationship Between P_e and γ at various SNR. 123

6.7 The ROC with Different Side Lobe to Main Lobe Ratio, SNR=20dB. . . . 127

6.8 The ROC with Different Side Lobe to Main Lobe Ratio, SNR=10dB. . . . 127

6.9 The ROC with Different Side Lobe to Main Lobe Ratio, SNR=0dB. . . . 128

7.1 The relation between κ and E_t. 148

7.2 Geometric Interpretation of Theorem 7.4. 154

7.3 The Mapping of all Real Signals $a(t)$ onto $\Delta T^2 \sim \Delta F^2$ plane. 156

7.4 Initial Approximation to the Convex Region $\Re_u$ with an Octagon. 157

7.5 Three Different Cases for the Selection of Optimum Solutions on the Octagon Approach. 161

7.6 Selection of the Second Approximation to the Convex Region $\Re_u$ Based on the Octagon. 162

7.7 Optimum Signal Obtained by Polygon Approach and its Corresponding WD. 165

7.8 Optimum Signal Obtained by Multi-dimensional Global Search and its Corresponding WD. 166

7.9 The Relationship Between maximum $\Delta T^2 \dot{\Delta} F^2$ and the Essential Duration-Bandwidth Product FT. 167

7.10 The Optimum Signal $\hat{a}_{0.3}(t)$ Obtained by Simulated Annealing Method in Which We Choose $\gamma = 0.3$ and its Corresponding WD. 176

7.11 The Optimum Signal $\hat{a}_{0.2}(t)$ Obtained by Simulated Annealing Method in Which We Choose $\gamma = 0.2$ and its Corresponding WD. 177

7.12 The Relationship Between the Value $\Delta T^2 \Delta F^2$ of the Optimum Signal and Allowable Sidelobe to Mainlobe Ratio γ. 179

7.13 The Optimum Signal $\tilde{a}_{0.3}(t)$ Obtained by ISS Method in Which We Choose $\gamma = 0.3$ and its Corresponding WD. 180

7.14 The Optimum Signal $\tilde{a}_{0.2}(t)$ Obtained by ISS Method in Which We Choose $\gamma = 0.2$ and its Corresponding WD. 181

7.15 The Theoretical performance of Various Signals Under Different SNR. . . 185

7.16 The Simulated performance of Various Signals Under Different SNR. . . . 185

J.1 Diagram of Expressing the Convergence of Polygon Approach. 220

Chapter 1

Introduction

1.1 A Stationary Process and Its Spectrum

Among all the aspects of Signal Processing, Spectrum Analysis is of special importance. This is due to the fact that most signal characteristics can be described when presented as a function of frequency, i.e. its spectrum. The spectrum of a signal $s(t)$ is defined as its Fourier transform [73]:

$$S(f) \triangleq \int_{-\infty}^{\infty} s(t)e^{-j2\pi ft}dt. \tag{1.1}$$

In many applications, what we are interested in is just the amplitude of $S(f)$. Hence, the power spectrum was defined as

$$P_s(f) \triangleq E\left[\mid S(f) \mid^2\right], \tag{1.2}$$

and much research effort has been attracted to the area of power spectrum estimation [52].

Consider a weakly stationary process $s(t)$, satisfying [35]:

$$E[s(t)] = constant, \tag{1.3}$$

$$C_s(t_1, t_2) \triangleq E\{[s(t_1) - E(s(t_1))][s^*(t_2) - E(s^*(t_2))] = C_s(t_1 - t_2), \tag{1.4}$$

where $E[\cdot]$ denotes the expectation and * denotes the complex conjugate. Without the loss of generality, we can assume that $E[s(t)] = 0$. Furthermore, if the signal is ergotic, i.e., its expected value and its time averaged value are equal with probability one, the power spectrum of $s(t)$ will be the Fourier Transform of its covariance function $C_s(\tau)$ [64], i.e.

$$P_s(f) = \int_{-\infty}^{\infty} C_s(\tau) e^{-j2\pi f\tau} d\tau. \tag{1.5}$$

So far, most spectrum estimation methods concentrate on weakly stationary processes.

However in many practical applications, the signal is observed only for a finite duration. Thus, instead of using Eq.(1.1) directly, the spectrum of a signal $s(t)$ is often approximated by its window Fourier Transform:

$$S(f) \approx \int_{-\infty}^{\infty} h(t)s(t)e^{-j2\pi ft} dt, \tag{1.6}$$

where $h(t)$ is a suitably selected window function.

In order to improve the accuracy of the spectrum estimation with the finite number of data, there are many methods being used, such as Maximum Likelihood Estimation (MLE) [53], Maximum Entropy Method (MEM) [9], MUltiply SIgnal Classification (MUSIC) [75], Forward and Backward Linear Prediction (FBLP) [32]. All these methods can give higher resolution spectrum for the given small duration segment of signal than simply applying the Fourier transform. Unfortunately, all these methods require the signal be a stationary process and in real applications:

- Most of the signals are non-stationary;
- Even if the signals are stationary, they are always transmitted in a non-stationary environment.

To deal with non-stationary signals and environment, some other methods must be introduced.

1.2 Non-Stationary Signals and Time-Dependent Spectrum

For non-stationary signals, Eq.(1.4) will no longer be valid. Generally speaking, it is difficult to define and analyze a non-stationary signal without further specifications. Hence, we begin this section by first introducing some commonly encountered non-stationary signals and defining their characteristics:

- *Piecewise Quasi-Stationary Model [58]:* The statistical properties of a random process change abruptly at some moments, and each segment between jumps can be described as a stationary process. Examples include speech signal, seismic signals, image signals and biomedical signals such as EEG and EKG, etc. [56].

- *Time-Varying Instantaneous Frequency $f_i(t)$[33], [2] and [85]:* The instantaneous frequency $f_i(t)$ of a signal is defined as:

$$f_i(t) = \frac{1}{2\pi}\frac{d\theta(t)}{dt}. \tag{1.7}$$

  The signal having time-varying instantaneous frequency is given by $s(t) = a(t)e^{j\theta(t)}$ which is an analytic signal with $a(t)$ being its envelope and $\theta(t)$ the phase. When the signal is contaminated by additive random noise $n(t)$ with zero mean, then the mean of $x(t) = s(t) + n(t)$ is $E[x(t)] = s(t)$ which varies with time. Hence, $x(t)$ is a non-stationary signal with time varying mean.

- *Locally Stationary Process [77]:* This is defined by their covariance function such that:

$$C_x(t_1, t_2) = m(\frac{t_1 + t_2}{2})q(t_1 - t_2), \tag{1.8}$$

  where $m(\cdot)$ and $q(\cdot)$ could be any kind of functions.

One example is linear FM signal: $s(t) = exp[j(2\pi\beta t^2/2 + \theta_0)]$, where β is a constant number and the initial phase θ_0 is a random variable uniformly distributed between 0 to 2π.

- *Uniformly Modulated Process [71]:*

$$s(t) = b(t)s_0(t), \tag{1.9}$$

where $b(t)$ is a random modulation function and $s_0(t)$ is a stationary process. This kind of signals is typically a non-stationary modulated signal and usually arises in the transmission of communication signals in a non-stationary channel, such as SARSAT (Search And Rescue Satellite) signals (see chapter 5).

For the non-stationary signals, the power spectrum must be time-varying, and we can denote it by $P_s(t, f)$, which is a function related of $s(t)$ and varies with the two variables t and f. We give $P_s(t, f)$ the name of *time-dependent power spectrum* . Naturally, the time-dependent power spectrum must preserve some properties of power spectrum of stationary signals [24]:

1. $P_s(t, f)$ must be a real and non-negative function of time and frequency.
2. $P_s(t, f)$ must be reduced to the ordinary power spectrum if the underlying stochastic process is stationary, i.e. if $C_s(t_1, t_2) = C_s(t_1 - t_2)$, then $P_s(t, f) = P_s(f)$, independent of time.
3. $P_s(t, f)$ should be translational invariant, i.e. if $x(t) = s(t-t_0)e^{j2\pi f_0 t}$,then $P_x(t, f) = P_s(t - t_0, f - f_0)$.

In addition to the above three major properties, it will be more effective if $P_s(t, f)$ can preserve some properties which power spectrum has:

1. *Scale invariance*: if $x(t) = s(ct)$, where c is a constant, then $P_x(t, f) = \frac{1}{|c|}P_s(ct, \frac{f}{c})$.

2. *Scalar product conservation*: if $x(t) = cs(t)$, then $P_x(t,f) = c^2 P_s(t,f)$.

3. *Time reversal compatibility*: if $x(t) = s(-t)$, then $P_x(t,f) = P_s(-t,-f)$.

Unfortunately, it was proved by Wigner, in 1971 [88] that some of above properties are incompatible. He stated that **there is no time-dependent spectrum linearly connected to the covariance function which satisfies all the required properties** [24]. We can only retain some properties of the time-dependent spectrum.

1.3 Time-Frequency Representations

One way to express a time-dependent spectrum is using the *time-frequency representations* which can be generalized as [17]

$$P_s(t,f) = \int_{R^3} g(\zeta,\tau) e^{j2\pi\zeta(u-t)} s(u+\tau/2) s^*(u-\tau/2) e^{-j2\pi f\tau} d\zeta du d\tau, \qquad (1.10)$$

where $s(t)$ is an analytic signal, i.e. its imaginary part is the Hilbert Transform of its real part, $P_s(t,f)$ is the spectrum which depends on two variables: time t and frequency f, R^3 denotes the 3-dimensional Euclidean space. $g(\zeta,\tau)$ is kernel function that generates the particular representation and satisfies:

$$| g(\zeta,\tau) | \leq g(0,0), \qquad | g(\zeta,\tau) | = | g^*(-\zeta,-\tau) | .$$

For different choices of the generating function $g(\zeta,\tau)$, we have the corresponding time-frequency representations as shown in Table 1.1. Detail discussions of the properties of some commonly used representations will be given later on.

1.4 Outline of Thesis

As mentioned above, a time-frequency representation provides us with a time- dependent spectrum, which covers many practical aspects of non-stationary process. Different

Table 1.1: **Table 1.1 Generating Function $g(\zeta,\tau)$ and Their Time-Frequency Representations $P_s(t,f)$**

$g(\zeta,\tau)$	P(t,f)
1	Wigner Distribution (WD)[87], [14] $W_s(t,f) = \int_{-\infty}^{\infty} s(t+\frac{\tau}{2})s^*(t-\frac{\tau}{2})e^{-j2\pi f\tau}d\tau$
$\delta(\tau+t)\delta(\zeta-f)$	Ambiguity Function (AF) [87], [93] $A_s(t,f) = \int_{-\infty}^{\infty} s(\tau+\frac{t}{2})s^*(\tau-\frac{t}{2})e^{j2\pi f\tau}d\tau$
$e^{j\pi\zeta\tau}$ $cos(\pi\zeta\tau)$	Rihaczek's Complex Energy Density Function [73]: $\varepsilon_s(t,f) = s(t)S^*(f)e^{-j2\pi ft}$ where S(f) is the Fourier Transform of s(t). $Re\{\varepsilon_s(t,f)\}$
$\frac{sin(\pi\zeta\tau)}{\pi\zeta}$	Jordan's Representation [43] : $J_s(t,f) = \int_{-\infty}^{\infty}\frac{1}{\tau}\int_{t-\frac{\tau}{2}}^{t+\frac{\tau}{2}} s(u+\frac{\tau}{2})s^*(u-\frac{\tau}{2})e^{-j2\pi f\tau}dud\tau$
$f(\tau,\Delta)\frac{sin(\Delta-\lvert\tau\rvert)\zeta}{\pi\zeta}$ where $f(\tau,\Delta) = \begin{cases} 1 & \text{for } \lvert\tau\rvert < \Delta \\ 0 & \text{for } \lvert\tau\rvert > \Delta \end{cases}$	Flanagan's Short-Time Spectrum [23]: $F_s(t,f) = \lvert \int_{t-\frac{\Delta}{2}}^{t+\frac{\Delta}{2}} s(\tau)e^{-j2\pi f\tau}d\tau \rvert^2$
$e^{j\pi\zeta\lvert\tau\rvert}$	Page's Instantaneous Power Spectrum [63]: $I_s(t,f) = \frac{\partial}{\partial t}\lvert \int_{-\infty}^{t} s(\tau)e^{-j2\pi f\tau}d\tau \rvert^2$
$A_h(\zeta,\tau)$ i.e. Ambiguity Function of a window function $h(t)$ is chosen as a kernel	Spectrogram [68] $G_s(t,f) = \lvert \int_{-\infty}^{\infty} s(\tau)h(t-\tau)e^{-j2\pi f\tau}d\tau \rvert^2$

presentations can be used to detect the change in the dynamical properties of signals and systems [5]. They can also be applied to estimate some signal parameters. In this thesis, we concentrate on the problem related with signal transmission and reception. In such applications, stationarity of the signals and systems is only an ideal assumption. In practice, they are always non-stationary for both signals and channels. Time-frequency representations provide us with an efficient way to deal with these kinds of signals and systems.

We focus on three important applications of time-frequency representations: the tracking of instantaneous frequency, high resolution estimation of frequencies in a non-stationary environment and optimum estimation of time delay and Doppler shift. Although there are several time-frequency representations, some being listed in Table 1.1, in this thesis, we only discuss the three commonly used ones: the Wigner distribution (WD), the Ambiguity function (AF) and the spectrogram, with emphasis on the WD.

In Chapter 2, a tutorial review highlights the historical development and properties of the time-frequency representations mentioned above. Their related application areas will be described.

The third and fourth chapters will mainly deal with the time-varying frequency tracking problem. In order to facilitate comparison, the Cramér-Rao Bound (CRB) for time-varying frequency estimate is derived in chapter 3. The theoretical performance of the WD when applied to the estimation of time-varying frequency will be analyzed in chapter 4. Simulation results will also be provided to compare the results.

In the fifth chapter, a particular application of frequency estimation in SARSAT (Search and Rescue Satellite) signal processing is presented. Detail algorithm will be described and analyzed. Both simulated and real data are used. It is shown that the estimated frequency using the WD is quite stable and the variation is small. This will confirm that the WD can achieve accurate frequency estimation under a non-stationary

environment. A high resolution algorithm which combines the Burg algorithm and the Wigner distribution will be introduced in this chapter. It will be shown that with very short data record, this new algorithm can achieve much higher resolution than other commonly used methods.

Chapter 6 deals with the optimum simultaneous estimation of time delay and Doppler shift. The theoretical estimation bias and variance are derived. The CRB is obtained for comparison. It shows that the joint time- delay and Doppler shift estimates using the peak of a time-frequency representation is optimum in the sense that it is unbiased and its variance reaches the CRB at high Signal-to-Noise Ratio (SNR). The threshold phenomenon will be discussed. The Receiver Operation Curve (ROC) is used as a criterion for the detection of signal.

Chapter 7 continues on from chapter 6. The uncertainty principle in the simultaneous time and frequency estimation for a time- and frequency- constrained signal will be studied in details. According to the uncertainty principle, an accurate frequency estimate cannot be obtained without the loss of accuracy of the time estimate, and vice versa. The problems to be solved in this chapter are: What is the lower bound for the joint variance of the time and frequency estimates? How can a signal be designed to approach this bound? And what are the other practical constraints which will influence the precision of time and frequency estimates. Several efficient methods are developed, which can be used to obtain the optimum signal for the joint time delay and Doppler shift estimation under different practical constraints. The merits of the different methods are discussed and signal design examples are presented.

Results and further research topics will be summarized in the final chapter: chapter 8.

Chapter 2

A Brief Review On Time-Frequency Representations

The motivation to analyze speech signals which change rapidly with time led to the invention of spectrogram in 1940s [68]. Later on the method became a standard and powerful tool to analyze non- stationary signals with time-varying spectra.

The pioneering works were started by Gabor [30], Ville [87] and Page [63]. Afterwards, many different time-frequency representations were developed. In 1966, L. Cohen generalized almost all of these representations in a unified formula, each with different kernels. This formula is given in Eq.(1.10). This chapter is intended to review briefly those of the most commonly used representations: the WD, the AF and the spectrogram with emphasis on the WD. The readers who are interested in others can read the papers [17], [7] and [59].

In order to analyze the time-varying spectrum, the first question to be asked is how time and frequency are defined. Hence, before the review of the time-frequency representations, we will first give the definition of instantaneous frequency and group delay. The importance of these two notions lies in the fact that most signals with time-varying frequencies have their supports in the time-frequency plane located along its instantaneous frequency and group delay[48].

2.1 Instantaneous Frequency and Group Delay

For a real signal, which is real part of a complex signal $a(t)e^{j\theta(t)}$,i.e.

$$v(t) = a(t)\cos\theta(t) = Re\{a(t)e^{j\theta(t)}\}, \tag{2.1}$$

where $Re[\cdot]$ represents the real part, $a(t)$ and $\theta(t)$ are both real functions of t with $a(t)$ being the envelope and $\theta(t)$ being the phase angle, when $a(t)$ is sufficiently low pass compared to $\cos\theta(t)$, the instantaneous frequency is defined by:

$$f_i(t) \triangleq \frac{1}{2\pi}\frac{d\theta(t)}{dt} = \frac{1}{2\pi}\frac{d}{dt}\{Im[\ln(a(t)e^{j\theta(t)})]\}, \tag{2.2}$$

where $Im[\cdot]$ represents the imaginary part. However, if only a real signal is given, in order to form a complex signal, there will be infinite number of choices with arbitrary selection of its imaginary part. Hence the definition of instantaneous frequency is not unique. If we assume that the signal is analytic, the imaginary part of a complex analytic signal $a(t)e^{j\theta(t)}$ is defined as the Hilbert transform of its real part $v(t)$, i.e.

$$a(t)sin\{\theta(t)\} = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{v(t)}{t-\tau}d\tau. \tag{2.3}$$

In this way, the uniqueness of $f_i(t)$ is guaranteed. However, it could be found from Eq.(2.3) that this unique determination of $\theta(t)$ requires a knowledge of the real function $v(t)$ over $(-\infty,\infty)$, which is impossible in practice.

Fortunately, most practical signals are *essentially time-limited*, which means their energy is concentrated in a certain time period $|t| < T$, i.e.

$$v(t) \approx 0; \quad |\,t\,| > T, \tag{2.4}$$

so that Eq.(2.3) can be approximately calculated with an observation of $v(t)$ over finite time.

For the finite observation of $v(t)$, the instantaneous frequency $f_i(t)$ is, in general, bounded as $F_1 > f_i(t) > F_2$ for $|t| \leq T$, where F_1 and F_2 are two real numbers. Hence,

the corresponding analytic signal $s(t) = a(t)e^{j\theta(t)}$ is *essentially band-limited*, which is defined as

$$S(f) \approx 0; \quad f \notin [F_2 - B,\ F_1 + B], \tag{2.5}$$

where $2B$ is the approximate bandwidth of $a(t)$. Since most practical signals satisfy Eqs.(2.4) and (2.5) for suitable values of T, F_1 , F_2 and B, they can be recognized as *essential time- and band-limited* signals. Throughout these thesis, we concentrate on these kinds of signals.

It was also proved ([85] and [33]) that almost all physically measurable frequencies are close to what is obtained by using the unrealizable Hilbert transform. This means that the unique definition of instantaneous frequency of a complex analytic signal having its imaginary part being the Hilbert transform of its real part is not only calculable but also practically meaningful.

Now we define the group delay in a similar way. Given a filter whose Fourier transform is:

$$X(f) = |X(f)|e^{j\Theta(f)},$$

where $|X(f)|$ and $\Theta(f)$, both real functions of f, are the amplitude and the phase of $X(f)$ respectively, then the group delay is defined as

$$\tau_x(f) \triangleq \frac{1}{2\pi}\frac{d}{df}\Theta(f) = \frac{1}{2\pi}\frac{d}{df}Im\{\ln X(f)\}. \tag{2.6}$$

If a signal is amplitude modulated, then the group delay is the measure of delay between the envelope of the input signal and that of the output signal of the filter. It is, in general, different from the phase delay of the carrier, which is the measurement of phasor lagging between the carrier of the output signal and that of the input signal. The group delay is an important concept in the analysis of filter and channel characteristics [21]. It represents the true signal delay.

2.2 The Spectrogram

The spectrogram is based on the Short-Time Fourier Transform (STFT), which is defined as [17]:

$$\begin{aligned} S_s(t,f) &\triangleq e^{j2\pi ft}\int_{-\infty}^{\infty} h(t-\tau)s(\tau)e^{-j2\pi f\tau}d\tau \\ &= \int_{-\infty}^{\infty} H(\nu-f)S^*(\nu)e^{j2\pi\nu t}d\nu, \end{aligned} \tag{2.7}$$

where $h(t)$ is a window function centred at time $t=0$ and $S(f)$ and $H(f)$ are the Fourier transforms of $s(t)$ and $h(t)$ respectively. It should be mentioned here that Eq.(2.7) is one way of defining the STFT. In the more commonly encountered definition ([7] and [14]), the factor $e^{j2\pi ft}$ is absorbed into the window function. The definition used here emphasizes that the commonly used window $h(t)$ is a real function.

To observe the energy distribution of the signal, we use the spectrogram which is defined as:

$$G_s(t,f) \triangleq \mid S_s(t,f) \mid^2, \tag{2.8}$$

which represents the instantaneous power spectrum at time t. It can be proved (see Appendix A) that if the Gaussian function

$$h(t) = (2\pi\sigma_g^2)^{-\frac{1}{4}} e^{-\frac{t^2}{4\sigma_g^2}}, \tag{2.9}$$

is chosen as the window function $h(t)$, $G_s(t,f)$ can best reflect the instantaneous spectrum of $s(t)$ at time instant t for most applications. As a result, if a Gaussian window of Eq.(2.9) is chosen, the spectrogram is just the generalized time-frequency representation (Eq.(1.10)) in which the kernel $g(\zeta,\tau)$ is selected (see in table 1.1) such that it is the AF of the Gaussian window, i.e.,

$$g(\zeta,\tau) = A_h(\zeta,\tau) \triangleq \int_{-\infty}^{\infty} h(u+\frac{\tau}{2})h(u-\frac{\tau}{2})e^{j2\pi\zeta u}du, \tag{2.10}$$

and the properties of the ambiguity function $A_h(\zeta,\tau)$ will be discussed in the next section.

2.3 Ambiguity Function (AF)

Selecting the kernel $g(\zeta,\tau) = \delta(\tau+t)\delta(\zeta-f)$ as shown in table 1.1 of the generalized time-frequency representation, we have the definition of AF of the $s(t)$:

$$A_s(t,f) \triangleq \int_{-\infty}^{\infty} s(\tau+\frac{t}{2})s^*(\tau-\frac{t}{2})e^{j2\pi f\tau}d\tau. \tag{2.11}$$

It could be easily seen from the definition that AF is a quadratic function of $s(t)$. If $f=0$, $A_s(t,0) = \int s(\tau+\frac{t}{2})s^*(\tau-\frac{t}{2})d\tau$ is the autocorrelation function in time, i.e. $A_s(t,0) = R_s(t)$. If $t=0$, $A_s(0,f)$ is the power spectral density of the signal. Many other properties of AF could be found in [17] and [7].

In the early 60's [47], AF was used to provide a pictorial representation in determining the range (time) and velocity (Doppler frequency), two important parameters of a target in radar techniques. The name AF came from the inherent ambiguity in a simultaneous determination of both the range and velocity of a moving target which will be discussed in detail in chapter 6 and 7. The AF has also been used in sonar, radio astronomy [4], geophysics [69], communications [76] and optics [65] [74] as well.

2.4 Wigner Distribution (WD)

The cross-Wigner distribution of two signals $x(t)$ and $y(t)$ is defined as

$$W_{xy}(t,f) \triangleq \int_{-\infty}^{\infty} x(t+\frac{\tau}{2})y^*(t-\frac{\tau}{2})e^{-j2\pi f\tau}d\tau. \tag{2.12}$$

Employing the Parseval's theorem, we can equivalently express Eq.(2.12) in the frequency domain such that

$$W_{xy}(t,f) = \int_{-\infty}^{\infty} X(f+\frac{\xi}{2})Y^*(f-\frac{\xi}{2})e^{-j2\pi\xi t}d\xi = W_{XY}(f,t), \tag{2.13}$$

where $X(f)$ and $Y(f)$ are Fourier transform of $x(t)$ and $y(t)$ respectively. Specifically, the auto-Wigner distribution, or simply, the Wigner distribution of a signal $s(t)$ is defined

as in Table 1.1:

$$\begin{aligned} W_s(t,f) &\triangleq \int_{-\infty}^{\infty} s(t+\frac{\tau}{2})s^*(t-\frac{\tau}{2})e^{-j2\pi f\tau}d\tau \\ &= \int_{-\infty}^{\infty} S(f+\frac{\xi}{2})S^*(f-\frac{\xi}{2})e^{-j2\pi\xi t}d\xi \end{aligned} \tag{2.14}$$

with the kernel in Eq.(1.10) being selected as:

$$g(\zeta,\tau)=1.$$

The Wigner distribution has many properties which make it most suitable for the approximation of time-dependent spectrum. All these properties could be found in [7] [14]. Here, we just list some of those which are interested to us

1. *The WD preserves the property of time shift*, i.e.,

$$\text{if} \quad x(t)=s(t-t_0), \qquad \text{then} \quad W_x(t,f)=W_s(t-t_0,f). \tag{2.15}$$

Proof:

$$\begin{aligned} W_x(t,f) &= \int_{-\infty}^{\infty} x(t+\frac{\tau}{2})x^*(t-\frac{\tau}{2})e^{-j2\pi f\tau}d\tau \\ &= \int_{-\infty}^{\infty} s(t-t_0+\frac{\tau}{2})s^*(t-t_0-\frac{\tau}{2})e^{-j2\pi f\tau}d\tau \\ &= W_s(t-t_0,f). \end{aligned}$$

□

2. *The WD preserves the property of frequency shift*, i.e.,

$$\text{if} \quad x(t)=s(t)e^{j2\pi f_0 t}, \qquad \text{then} \quad W_x(t,f)=W_s(t,f-f_0). \tag{2.16}$$

Proof:

$$\begin{aligned} W_x(t,f) &= \int_{-\infty}^{\infty} x(t+\frac{\tau}{2})x^*(t-\frac{\tau}{2})e^{-j2\pi f\tau}d\tau \\ &= \int_{-\infty}^{\infty} s(t+\frac{\tau}{2})e^{j2\pi f_0(t+\frac{\tau}{2})}s^*(t-\frac{\tau}{2})e^{-j2\pi f_0(t-\frac{\tau}{2})}e^{-j2\pi f\tau}d\tau \\ &= \int_{-\infty}^{\infty} s(t+\frac{\tau}{2})s^*(t-\frac{\tau}{2})e^{-j2\pi(f-f_0)\tau}d\tau \\ &= W_s(t,f-f_0). \end{aligned}$$

□

3. *The WD is scale invariance*, i.e.,

$$\text{if} \quad x(t) = s(ct), \qquad \text{then} \quad W_x(t,f) = \frac{1}{|c|} W_s(ct, \frac{f}{c}) \tag{2.17}$$

where c is a constant.

Proof:

$$\begin{aligned} W_x(t,f) &= \int_{-\infty}^{\infty} x(t+\frac{\tau}{2}) x^*(t-\frac{\tau}{2}) e^{-j2\pi f\tau} d\tau \\ &= \int_{-\infty}^{\infty} s(ct+c\frac{\tau}{2}) s^*(ct-c\frac{\tau}{2}) e^{-j2\pi f\tau} d\tau. \end{aligned}$$

Let $\tau_1 = c\tau$, we have

$$\begin{aligned} W_x(t,f) &= \int_{-\infty}^{\infty} s(ct+\frac{\tau_1}{2}) s^*(ct-\frac{\tau_1}{2}) e^{-j2\pi \frac{f}{c}\tau_1} d\frac{\tau_1}{|c|} \\ &= \frac{1}{|c|} W_s(ct, \frac{f}{c}). \end{aligned}$$

□

4. *The WD satisfies scalar product conservation*, i.e.,

$$\text{if} \quad x(t) = cs(t), \qquad \text{then} \quad W_x(t,f) = c^2 W_s(t,f). \tag{2.18}$$

Proof: This property can be directly proved by substituting $x(t) = cs(t)$ into the definition of the WD. □

5. *The WD satisfies time reversal compatibility*, i.e.,

$$\text{if} \quad x(t) = s(-t), \qquad \text{then} \quad W_x(t,f) = W_s(-t,-f). \tag{2.19}$$

Proof: This property is just a special form of property 3 with $c = -1$. □

6. *The WD preserves the instantaneous signal energy*, i.e.

$$| s(t) |^2 = \int_{-\infty}^{\infty} W_s(t,f)df \tag{2.20}$$

Proof:

$$\begin{aligned}\int_{-\infty}^{\infty} W_s(t,f)df &= \int\int_{-\infty}^{\infty} s(t+\frac{\tau}{2})s^*(t-\frac{\tau}{2})e^{-j2\pi f\tau}d\tau df \\ &= \int_{-\infty}^{\infty} s(t+\frac{\tau}{2})s^*(t-\frac{\tau}{2})\left[\int_{-\infty}^{\infty} e^{-j2\pi f\tau}df\right]d\tau \\ &= \int_{-\infty}^{\infty} s(t+\frac{\tau}{2})s^*(t-\frac{\tau}{2})\delta(\tau)d\tau \\ &= |s(t)|^2\end{aligned}$$

where $\delta(\cdot)$ is the Dirac delta function. □)

7. *The WD preserves power density spectrum*, i.e.

$$| S(f) |^2 = \int_{-\infty}^{\infty} W_s(t,f)dt \tag{2.21}$$

Proof: Using the definition of WD in the second equality of Eq.(2.14), Eq.(2.21) can be proved in a similar way as Eq.(2.20). □

From properties 6 and 7, it can be easily obtained that the total signal energy is equal to the global zero order moment of its WD:

$$\|s(t)\|^2 = \int\int_{-\infty}^{\infty} W_s(t,f)dtdf = \int_{-\infty}^{\infty} | s(t) |^2 \, dt = \int_{-\infty}^{\infty} | S(f) |^2 \, df. \tag{2.22}$$

8. *The WD preserves the instantaneous frequency*, i.e.,

$$f_i(t) = \frac{\int_{-\infty}^{\infty} fW_s(t,f)df}{\int_{-\infty}^{\infty} W_s(t,f)df}. \tag{2.23}$$

Proof: Firstly, we present two properties of the Fourier transform which will be used in the proof [64]:

(a)

$$\int_{-\infty}^{\infty} S(f)df = s(0). \tag{2.24}$$

(b)

$$\text{if} \quad s(t) \longleftrightarrow S(f), \qquad \text{then} \quad \frac{ds(t)}{dt} \longleftrightarrow 2\pi j f S(f). \tag{2.25}$$

where the two sides of $\longleftrightarrow$ represents the Fourier transform pairs.

Observing the definition of the WD in Eq.(2.14), we have the following relation

$$s(t+\frac{\tau}{2})s^*(t-\frac{\tau}{2}) \longleftrightarrow W_s(t,f) \tag{2.26}$$

where τ is the time variable of Fourier transform and t is a parameter. Using the Fourier property in Eq.(2.25), we obtain

$$\frac{1}{2\pi j}\frac{d}{d\tau}\left[s(t+\frac{\tau}{2})s^*(t-\frac{\tau}{2})\right] \longleftrightarrow fW_s(t,f). \tag{2.27}$$

Also

$$\frac{d}{d\tau}\left[s(t+\frac{\tau}{2})s^*(t-\frac{\tau}{2})\right] = \frac{1}{2}\dot{s}(t+\frac{\tau}{2})s^*(t-\frac{\tau}{2}) - \frac{1}{2}s(t+\frac{\tau}{2})\dot{s}^*(t-\frac{\tau}{2}), \tag{2.28}$$

where $\dot{s}(t_1) \triangleq \left.\frac{ds(t)}{dt}\right|_{t=t_1}$.

Using Eqs.(2.25) to (2.28) and the property of the Fourier transform in Eq.(2.24), we have

$$\begin{aligned}\int_{-\infty}^{\infty} fW_s(t,f)df &= \frac{1}{4\pi j}\left[\dot{s}(t+\frac{\tau}{2})s^*(t-\frac{\tau}{2}) - s(t+\frac{\tau}{2})\dot{s}^*(t-\frac{\tau}{2})\right]_{\tau=0} \\ &= \frac{1}{4\pi j}\left[\dot{s}(t)s^*(t) - s(t)\dot{s}^*(t)\right]. \end{aligned} \tag{2.29}$$

We also have, from Eq.(2.20),

$$\int_{-\infty}^{\infty} W_s(t,f)df = s(t)s^*(t). \tag{2.30}$$

Combining Eqs.(2.29) and (2.30), we have

$$\begin{aligned}\frac{\int_{-\infty}^{\infty} fW_s(t,f)df}{\int_{-\infty}^{\infty} W_s(t,f)df} &= \frac{1}{4\pi j}\left[\frac{\dot{s}(t)}{s(t)} - \frac{\dot{s}^*(t)}{s^*(t)}\right] \\ &= \frac{1}{2\pi} Im\left[\frac{\dot{s}(t)}{s(t)}\right] \\ &= \frac{1}{2\pi} Im\left\{\frac{d}{dt}\left[ln(s(t))\right]\right\}, \end{aligned}$$

which is the same as the definition of instantaneous frequency given in Eq.(2.2). Thus, Eq.(2.23) is proved. □

9. *The WD preserves the group delay,* i.e.,

$$\tau_s(f) = \frac{\int_{-\infty}^{\infty} tW_s(t,f)dt}{\int_{-\infty}^{\infty} W_s(t,f)dt}. \tag{2.31}$$

Proof: Using the dual property between the time domain and frequency domain in the Fourier transform, this property can be proved in the similar way as in the proof of property 8. □

10. For the linear FM signal, which is defined as

$$s(t) = e^{j2\pi\beta\frac{t^2}{2}}, \tag{2.32}$$

with β being a constant, its instantaneous frequency is

$$f_i(t) = \frac{1}{2\pi}\frac{d}{dt}\left[2\pi\frac{\beta t^2}{2}\right] = \beta t.$$

The WD of linear FM given by Eq.(2.32) will be a Dirac delta function whose non-zero support is located only at its instantaneous frequency $f_i(t)$, i.e.,

$$W_s(t,f) = \delta(f - f_i(t))$$

Proof:

$$\begin{aligned} W_s(t,f) &= \int_{-\infty}^{\infty} e^{j2\pi\beta\frac{(t+\frac{\tau}{2})^2}{2}} e^{-j2\pi\beta\frac{(t-\frac{\tau}{2})^2}{2}} e^{-j2\pi f\tau} d\tau \\ &= \int_{-\infty}^{\infty} e^{-j2\pi(f-\beta t)\tau} d\tau \\ &= \delta(f - \beta t) \end{aligned}$$

where βt is the instantaneous frequency of $s(t)$. □

11. If the signal $s(t)$ is stationary and has a zero mean (non-zero constant mean can be subtracted from $s(t)$ to make its mean zero), then $E[s(t_1)s^*(t_2)] = C_s(t_1 - t_2)$ and

$E[W_s(t,f)]$ will be reduced to the power spectrum of s(t), which is independent on time t, i.e.,

$$E[W_s(t,f)] = P_s(f).$$

Proof:

$$\begin{aligned} E[W_s(t,f)] &= \int_{-\infty}^{\infty} E\left[s(t+\frac{\tau}{2})s^*(t-\frac{\tau}{2})\right] e^{-j2\pi f\tau} d\tau \\ &= \int_{-\infty}^{\infty} C_s(\tau) e^{-j2\pi f\tau} d\tau = P_s(f). \end{aligned}$$

□

For a general s(t), i.e. $E[s(t_1)s^*(t_2)] = C_s(t_1,t_2)$, $E[W_s(t,f)]$ represents the time-dependent spectrum:

$$E\left[W_s(t,f)\right] = \int_{-\infty}^{\infty} C_s(t+\frac{\tau}{2}, t-\frac{\tau}{2}) e^{-j2\pi f\tau} d\tau.$$

12. Since the WD is a bilinear transform, when there are two or more signals present, the WD generates "interference terms" such that for two signals $x(t)$ and $y(t)$, the WD is given by

$$W_{x+y}(t,f) = W_x(t,f) + W_y(t,f) + 2Re[W_{xy}(t,f)]. \tag{2.33}$$

If both $x(t)$ and $y(t)$ are narrow band signals and can be expressed as $x(t) = a_x(t)e^{-j2\pi f_x t}$ and $y(t) = a_y(t)e^{-j2\pi f_y t}$, where $a_x(t)$ and $a_y(t)$ are the corresponding amplitudes, and f_x and f_y are the frequencies. then the interference term $2Re[W_{xy}(t,f)]$ will lie between the two signal components $W_x(t,f)$ and $W_y(t,f)$, and is oscillatory such that its frequency decreases with the decreasing distance between the two signal components in the time-frequency plane [43].

Proof: Substituting $x(t) + y(t)$ into the definition of WD in Eq.(2.14), we have

$$W_{x+y}(t,f) = W_x(t,f) + W_y(t,f) + 2Re[W_{xy}(t,f)]. \tag{2.34}$$

From the assumption that both $a_x(t)$ and $a_y(t)$ are narrow band, we have, for the interference term,

$$\begin{aligned} W_{xy}(t,f) &= \int_{-\infty}^{\infty} x(t+\frac{\tau}{2})y^*(t-\frac{\tau}{2})e^{-j2\pi f\tau}d\tau \\ &= e^{j2\pi(f_x-f_y)t}\int_{-\infty}^{\infty} a_x(t+\frac{\tau}{2})a_y(t-\frac{\tau}{2})e^{2\pi(f-\frac{f_x+f_y}{2})\tau}d\tau. \quad (2.35) \end{aligned}$$

The integrand in Eq. (2.35) shows that the interference term $W_{xy}(t,f)$ lies around the frequency $\frac{f_x+f_y}{2}$ which is half way between the frequency components of $W_x(t,f)$ and $W_y(t,f)$. Also the factor $e^{j2\pi(f_x-f_y)t}$ outside of the integral means that the interference term is oscillatory in the time axis with a frequency which is equal to the distance between the two signal frequency components, i.e., $f_x - f_y$. □

The interference term will yield "false" spectral components in the time-frequency plane when two or more signals present and thus could cause problems in the determination of the signal components. The scale factor 2 in the Eq.(2.33) also shows that the magnitude of the interference term could be as high as twice that of the auto-component.

13. $W_s(t,f)$ *is always real*, i.e.,

$$W_s(t,f) = W_s^*(t,f). \quad (2.36)$$

Proof:

$$W_s^*(t,f) = \int_{-\infty}^{\infty} s^*(t+\frac{\tau}{2})s(t-\frac{\tau}{2})e^{j2\pi f\tau}d\tau$$

Putting $\tau_1 = -\tau$, we have

$$W_s^*(t,f) = \int_{-\infty}^{\infty} s^*(t-\frac{\tau_1}{2})s(t+\frac{\tau_1}{2})e^{-j2\pi f\tau}d\tau_1 = W_s(t,f)$$

□

Nevertheless, although the auto-WD is always real, it is not always positive. This is the property which hinders the WD from being an approximation to the time-dependent power spectrum. However, as we will see later in its application, its positivity is indeed met in a considerably greater number of cases [27] and [39] .

Due to the fact that the WD possesses almost all the properties which time-dependent power spectrum should have (as discussed in section 1.2), it is the most widely used among the group of time-frequency representations. It is especially suitable for the application to non-stationary signal processing and the most appropriate for power spectrum presentation of the non-stationary signal [7].

For computational purposes it is necessary in general to weight the signals by a window h. The window slides along the time axis, and at each instant t, the Wigner distribution of the windowed signal is evaluated. This results in a family of signals $\tilde{s}$ such that

$$\tilde{s}(\tau) = s(\tau)h(\tau - t). \tag{2.37}$$

It can be easily shown [7] that the Wigner distribution of the windowed signal $\tilde{s}$ is also given by

$$W_s(\tau, f) = \int_{-\infty}^{\infty} W_s(\tau, \eta)W_h(\tau - t, f - \eta)d\eta. \tag{2.38}$$

Here t appears as a parameter indicating the position of the window as it slides along the time axis. It is therefore more convenient to consider only the values on the line $\tau = t$, i.e.,

$$\tilde{W}_s(t, f) = W_{\tilde{s}}(\tau, f)\Big|_{\tau=t} = \int_{-\infty}^{\infty} W_s(\tau, \eta)W_h(0, f - \eta)d\eta. \tag{2.39}$$

The function $\tilde{W}_s(t, f)$ which resembles, but is not exactly, a Wigner Distribution, is called a Pseudo Wigner distribution (PWD) and is equal to the original Wigner distribution convolved with the Wigner distribution of a window function with respect to its frequency variable along $t = 0$ line. In general, the window function is chosen to be real and even

symmetric, i.e. $h(t) = h(-t)$. By directly applying the definition of the WD on $\tilde{s}(\tau)$ and letting $\tau = t$, we obtain another useful formula for $\tilde{W}_s(t, f)$, which is

$$\tilde{W}_s(t,f) = \int_{-\infty}^{\infty} h^2(\frac{\tau}{2}) s(t+\frac{\tau}{2}) s^*(t-\frac{\tau}{2}) e^{-j2\pi f\tau} d\tau. \tag{2.40}$$

The above discussions on the properties of the time-frequency representations, especially on the WD, is considered basic to the ensuing work, and accordingly, much of the material will be used in subsequent chapters.

Chapter 3

Lower Bound on the Estimate of Time-Varying Parameter of a Signal

In various environments involving signal transmission and reception, we often encounter signals which possess time-varying frequencies. On many occasions, this parameter of time-varying frequency bears important information, and its accurate measurement is thus necessary. For example, in a radar or sonar system, the frequency of the signal can reveal the velocity of the target. In this chapter and the next one, we focus our attention on signals of the form:

$$s(t, F(t)) = A \exp\{2\pi j \int_0^t F(t_1) dt_1\}, \tag{3.1}$$

with $F(t)$ being time-varying deterministic frequency. Firstly, in this chapter, we will derive the Cramér-Rao Bound (CRB) of estimating a continuous time-varying parameter in Gaussian white noise, so that we may have a common ground for comparing the different time-frequency representations when applied to the time-varying frequency estimate.

The problem of estimating the time-varying frequency of a signal in Gaussian noise falls under the topic of estimating a continuous time-varying parameter. Our attention

is focused on the property that the mapping from the parameter $g(t)$ (here $g(t)$ is a function of time t, which represents a general form of $F(t)$) to the signal $s(t, g(t))$ is a no-memory transformation; i.e., the received signal is

$$x(t) = s(t, g(t)) + n(t), \quad T_S \leq t \leq T_F, \tag{3.2}$$

where T_S and T_F are,respectively, the starting and finishing instant of the signal. By a no-memory transformation we mean that the transmitted signal at some time t dependent only on $g(t)$ and not on the past of $g(t)$. The other assumptions made here are:

1. $n(t)$ is a complex Gaussian white process whose mean and auto-correlation function are given as:

$$E[n(t)] = 0, \tag{3.3}$$

$$R_n(t_1, t_2) = E[n(t_1)n^*(t_2)] = N_0\delta(t_1 - t_2). \tag{3.4}$$

   Due to the zero mean of $n(t)$, its auto-correlation function $R_n(t_1, t_2)$ will be the same as the auto-covariance function $C_n(t_1, t_2)$ defined in Eq.(1.4).

2. $g(t)$ is deterministic and real.

3. The signal $s(t, g(t))$ has a derivative with respect to $g(t)$.

3.1 A Lower Bound On the Mean-Square Error in Estimating a Continuous Random Parameter

To develop the Cramér-Rao bound for the estimate of the continuous deterministic parameter, we will first derive the lower bound when $g(t)$ is a random parameter, then extend the result to case when $g(t)$ is deterministic.

Firstly, with $\hat{g}(t)$ being the estimate of $g(t)$, the error waveform is defined as

$$e_g(t) = g(t) - \hat{g}(t), \tag{3.5}$$

and the error within the observation interval as

$$e_I = \frac{1}{T_I}\int_{T_S}^{T_F} e_g^2(t)dt, \tag{3.6}$$

where $T_I = T_F - T_S$. We note that e_I is a random variable. In the derivation of the lower bound of $E[e_I]$, we follow closely the procedure in [86] and then extend the results to complex signals [91].

Using the Karhunen-Loéve expansion, we expanded $g(t)$ in a series of orthonormal function $\{\psi_{gk}(t)\}$ such that

$$g(t) = \underset{K\to\infty}{\text{l.i.m.}} \sum_{k=1}^{K} g_k\psi_{gk}(t), \tag{3.7}$$

where l.i.m. denotes limit in the mean, and

$$g_k = \int_{T_S}^{T_F} g(t)\psi_{gk}(t)dt, \tag{3.8}$$

and $\psi_{gk}(t)$ satisfies the integral equation

$$\lambda_{gk}\psi_{gk}(t_1) = \int_{T_S}^{T_F} R_g(t_1,t_2)\psi_{gk}(t_2)dt_2. \tag{3.9}$$

The function $R_g(t_1,t_2)$ is defined as the autocorrelation function of $g(t)$:

$$R_g(t_1,t_2) \triangleq E[g(t_1)g^*(t_2)]. \tag{3.10}$$

If g_k are real Gaussian variables with zero mean and variance σ_{gk}^2, then $R_g(t_1,t_2)$ will be the autocovariance function of real Gaussian random process $g(t)$. Since $\psi_{gk}(t)$ are chosen according to Eq.(3.9), g_k are independent Gaussian variables such that[86]

$$E[g_k] = 0, \tag{3.11}$$

$$E[g_k g_l] = \lambda_{gk}\delta_{kl}. \tag{3.12}$$

Note that λ_{gk}, g_k, $\psi_{gk}(t)$ and $R_g(t_1,t_2)$ are all real. Now consider the K term expansion of $g(t)$. i.e.

$$g_K(t) = \sum_{k=1}^{K} g_k\psi_{gk}(t). \tag{3.13}$$

Then

$$\begin{aligned} s(t,g(t)) &= \underset{K\to\infty}{\text{l.i.m.}}\ s(t,g_K(t)) \\ &= \underset{K\to\infty}{\text{l.i.m.}}\ s(t,\sum_{k=1}^{K} g_k\psi_{gk}(t)) \\ &= \underset{K\to\infty}{\text{l.i.m.}}\ s(t,g_K), \end{aligned} \tag{3.14}$$

where

$$g_K = [g_1,\ g_2,\ \cdots\ g_K]^T. \tag{3.15}$$

The maximum a posteriori (MAP) [91] estimate of $g_K(t)$ is defined as

$$\hat{g}_K(t) = \sum_{k=1}^{K} \hat{g}_k\psi_{gk}(t) \qquad T_S \le t \le T_F, \tag{3.16}$$

i.e., the estimate of the waveform $g_K(t)$ over the interval $[T_S, T_F]$ is equivalent to estimating the K coefficients g_k. Extending to a continuous parameter, we write,

$$s(t,\hat{g}(t)) = \underset{K\to\infty}{\text{l.i.m.}}\ s(t,\hat{g}_K(t)) = \underset{K\to\infty}{\text{l.i.m.}}\ s(t,\hat{g}_K). \tag{3.17}$$

Still using the Karhunen-Loéve expansion, we choose a set of orthonormal functions $\{\psi_{ni}(t)\}$ such that they are eigenfunctions of the integral equation

$$\int_{T_S}^{T_F} R_n(t_1,t_2)\psi_{ni}(t_2)dt_2 = \lambda_{ni}\psi_{ni}(t_1), \qquad T_S \le t_1 \le T_F, \tag{3.18}$$

with the property that

$$\int_{T_S}^{T_F} \psi_{ni}^*(t)\psi_{nj}(t)dt = \delta_{ij}. \tag{3.19}$$

We expand $x(t)$ in this coordinate system, i.e.,

$$\begin{aligned} x(t) &= \underset{M\to\infty}{\text{l.i.m.}} \sum_{i=1}^{M} x_i\psi_{ni}(t) \\ &= \underset{M\to\infty}{\text{l.i.m.}} \sum_{i=1}^{M} s_i\psi_{ni}(t) + \underset{M\to\infty}{\text{l.i.m.}} \sum_{i=1}^{M} n_i\psi_{ni}(t). \end{aligned} \tag{3.20}$$

where

$$x_i = \int_{T_S}^{T_F} x(t)\psi_{ni}^*(t)dt, \tag{3.21}$$

$$s_i = \int_{T_S}^{T_F} s(t)\psi_{ni}^*(t)dt, \tag{3.22}$$

and

$$n_i = \int_{T_S}^{T_F} n(t)\psi_{ni}^*(t)dt. \tag{3.23}$$

Since the noise is zero-mean, and since the coefficients of its Karhunen- Loéve expansion are uncorrelated, we can write

$$E[n_i] = 0, \qquad E[n_i n_j^*] = \lambda_{ni}\delta_{ij}. \tag{3.24}$$

Let $\boldsymbol{x} = [x_1,\ x_2,\ \cdots, x_M]^T$. For an unbiased estimate of multiple random parameters, the variance of the estimate is bounded by the Cramér-Rao bound[86]

$$\sigma_{g_k}^2 \triangleq var[\hat{g}_k(\boldsymbol{x}) - g_k] \geq J^{kk}, \qquad k = 1, \cdots K, \tag{3.25}$$

where $\hat{g}_k(\boldsymbol{x})$ is any unbiased estimate of the kth parameter based on the received signal vector $\boldsymbol{x}$ and J^{kk} is the kkth element of $K \times K$ square matrix J^{-1}. The matrix J has two components, i.e.

$$J = J_D + J_P, \tag{3.26}$$

where J_P is the a priori information matrix, the elements of which are:

$$\begin{aligned} J_{Pij} &\triangleq E\left[\frac{\partial \ln P_{\boldsymbol{g}}(\boldsymbol{g})}{\partial g_i} \cdot \frac{\partial \ln P_{\boldsymbol{g}}(\boldsymbol{g})}{\partial g_j}\right] \\ &= -E\left[\frac{\partial^2 \ln P_{\boldsymbol{g}}(\boldsymbol{g})}{\partial g_i \partial g_j}\right], \end{aligned} \tag{3.27}$$

and J_D is the information matrix obtained from the data. The elements of J_D are

$$\begin{aligned} J_{Dij} &\triangleq E\left[\frac{\partial \ln P_{\boldsymbol{x}|\boldsymbol{g}}(\boldsymbol{x} \mid \boldsymbol{g})}{\partial g_i} \cdot \frac{\partial \ln P_{\boldsymbol{x}|\boldsymbol{g}}(\boldsymbol{x} \mid \boldsymbol{g})}{\partial g_j}\right] \\ &= -E\left[\frac{\partial^2 \ln P_{\boldsymbol{x}|\boldsymbol{g}}(\boldsymbol{x} \mid \boldsymbol{g})}{\partial g_i \partial g_j}\right]. \end{aligned} \tag{3.28}$$

Now, since n_k, the coefficients of the Karhunen-Loéve expansion of $n(t)$, are uncorrelated, then the joint conditional PDF (Probability Density Function) of $\boldsymbol{x}$ is Gaussian

given by [60]

$$p_{\boldsymbol{x}|g_K}(\boldsymbol{x}|g_K) = \prod_{k=1}^{M} \frac{1}{\sqrt{\pi\lambda_{nk}}} \exp\left[-\frac{|x_k - s_k(g_K)|^2}{2\lambda_{nk}}\right], \tag{3.29}$$

so that

$$\frac{\partial}{\partial g_i}[\ln p_{\boldsymbol{x}|g_K}(\boldsymbol{x}|g_K)] = \sum_{k=1}^{M}\left[\frac{(x_k - s_k(g_K))}{\lambda_{nk}} \cdot \frac{\partial s_k^*(g_K)}{\partial g_i} + \frac{\partial s_k(g_K)}{\partial g_i} \cdot \frac{(x_k^* - s_k^*(g_K))}{\lambda_{nk}}\right]. \tag{3.30}$$

Using Eqs.(3.21) and (3.22), and letting $M \to \infty$, we have

$$\begin{aligned}
&\frac{\partial}{\partial g_i}[\ln p_{\boldsymbol{x}|g_K}(\boldsymbol{x}|g_K)] \\
&= \sum_{k=1}^{\infty}\left[\int\int_{T_S}^{T_F}\left\{\frac{x(t_1)\psi_{nk}^*(t_1)}{\lambda_{nk}} \cdot \frac{\partial s^*(t_2,g_K)}{\partial g_i} \cdot \psi_{nk}(t_2)\right.\right. \\
&\quad \left. + \frac{\partial s(t_2,g_K)}{\partial g_i} \cdot \psi_{nk}^*(t_2) \cdot \frac{x^*(t_1)\psi_{nk}(t_1)}{\lambda_{nk}}\right\} dt_1 dt_2 \\
&\quad - \int\int_{T_S}^{T_F}\left\{\frac{s(t_1,g_K)\psi_{nk}^*(t_1)}{\lambda_{nk}} \cdot \frac{\partial s^*(t_2,g_K)}{\partial g_i} \cdot \psi_{nk}(t_2)\right. \\
&\quad \left.\left. + \frac{\partial s(t_2,g_K)}{\partial g_i} \cdot \psi_{nk}^*(t_2) \cdot \frac{s^*(t_1,g_K)\psi_{nk}(t_1)}{\lambda_{nk}}\right\} dt_1 dt_2\right] \\
&= \int\int_{T_S}^{T_F} 2Re\left[\{x^*(t_1) - s^*(t_1,g_K)\} \cdot \frac{\partial s(t_2,g_K)}{\partial g_i} \cdot \left\{\sum_{k=1}^{\infty}\frac{\psi_{nk}(t_1)\cdot\psi_{nk}^*(t_2)}{\lambda_{nk}}\right\}\right] dt_1 dt_2 \\
&= \int\int_{T_S}^{T_F} 2Re\left[\{x^*(t_1) - s^*(t_1,g_K)\} \cdot \frac{\partial s(t_2,g_K)}{\partial g_i} Q_n(t_1,t_2)\right] dt_1 dt_2,
\end{aligned} \tag{3.31}$$

where

$$Q_n(t_1,t_2) \triangleq \sum_{k=1}^{\infty}\frac{\psi_{nk}(t_1)\cdot\psi_{nk}^*(t_2)}{\lambda_{nk}}. \tag{3.32}$$

From Eqs.(3.13) and (3.14), we have

$$\begin{aligned}
\frac{\partial s(t_1,g_K)}{\partial g_i} &= \frac{\partial s(t_1,g_K(t_1))}{\partial g_i} \\
&= \left[\frac{\partial s(t_1,g_K(t_1))}{\partial g_K(t_1)}\right] \cdot \frac{\partial g_K(t_1)}{\partial g_i} \\
&:= \left[\frac{\partial s(t_1,g_K(t_1))}{\partial g_K(t_1)}\right] \cdot \psi_{gi}(t_1) \quad i = 1,\cdots K.
\end{aligned} \tag{3.33}$$

To extend the Cramér-Rao bound to a continuous parameter, we use Eq.(3.33) in

(3.31) and obtain

$$\frac{\partial}{\partial g_i}[\ln p_{\underline{x}|\underline{g}_K}(\underline{x}|\underline{g}_K)] = 2\int\int_{T_S}^{T_F} Re\left[\{x^*(t) - s^*(t, g_K(t))\}\frac{\partial s(t_1, g_K(t_1))}{\partial g_K(t_1)}\psi_{gi}(t_1)Q_n(t, t_1)\right] dtdt_1. \tag{3.34}$$

Also, since the elements of $\underline{g}_K$ are independent and Gaussian, then

$$\frac{\partial}{\partial g_i}[p_g(g)] = -\frac{g_i}{\lambda_{gi}}. \tag{3.35}$$

Differentiating the sum of Eqs.(3.34) and (3.35) with respect to g_j, taking the expected value and including the minus sign, we can express the ijth element of the matrix $\boldsymbol{J}$ as

$$\begin{aligned} J_{ij} &= \frac{\delta_{ij}}{\lambda_{gi}} - E\left[2\int\int_{T_S}^{T_F} Re\left(\{x^*(t) - s^*(t, g_K(t))\}\right.\right. \\ &\quad \left.\left.\frac{\partial}{\partial g_j}\left\{Q_n(t, t_1)\psi_{gi}(t_1)\frac{\partial s(t_1, g_K(t_1))}{\partial g_K(t_1)}\right\}\right) dtdt_1\right] \\ &\quad + E\left[2\int\int_{T_S}^{T_F} Re\left(\frac{\partial s^*(t, g_K(t))}{\partial g_K(t)}\psi_{gj}(t)\left\{Q_n(t, t_1)\psi_{gi}(t_1)\frac{\partial s(t_1, g_K(t_1))}{\partial g_K(t_1)}\right\}\right) dtdt_1\right]. \end{aligned} \tag{3.36}$$

But the second term in Eq.(3.36) contains a noise term $n(t) = x(t) - s(t, g_k(t))$ and the expected value of it is zero. Hence, Eq.(3.36) can be simplified to

$$J_{ij} = \frac{\delta_{ij}}{\lambda_{gi}} + 2E\left[\int\int_{T_S}^{T_F} Re\left\{\psi_{gj}(t)\psi_{gi}(t_1)\frac{\partial s^*(t, g_K(t))}{\partial g_K(t)}Q_n(t, t_1)\frac{\partial s(t_1, g_K(t_1))}{\partial g_K(t_1)}\right\} dtdt_1\right], \tag{3.37}$$

where δ_{ij} is the Kronecker delta. Eq.(3.37) expresses the individual element of the matrix $\boldsymbol{J}$. It is not convenient to use in the case of continuous random parameter estimation. To gain deeper insight, we recall that the auto-correlation function of $g(t)$ can be expressed as:

$$R_g(t, t_1) = \sum_{k=1}^{\infty} \lambda_{gk}\psi_{gk}(t)\psi_{gk}(t_1), \quad T_S \le t, t_1 \le T_F. \tag{3.38}$$

Because we use only K terms, we define

$$R_{gK}(t, t_1) \triangleq \sum_{k=1}^{K} \lambda_{gk}\psi_{gk}(t)\psi_{gk}(t_1), \quad T_S \le t, t_1 \le T_F. \tag{3.39}$$

We also define

$$Q_{gK}(t,t_1) \triangleq \sum_{k=1}^{K} \frac{1}{\lambda_{gk}} \psi_{gk}(t)\psi_{gk}(t_1), \quad T_S \le t,t_1 \le T_F. \tag{3.40}$$

We observe that

$$\int_{T_S}^{T_F} Q_{gK}(t,t_1) R_{gk}(t_1,\tau)dt_1 = \sum_{k=1}^{K} \psi_{gk}(t)\psi_{gk}(\tau). \tag{3.41}$$

Now, using Eq.(3.40) and the orthonormality of $\psi_{gk}(t)$, the first term in Eq.(3.36) can be written as

$$\frac{\delta_{ij}}{\lambda_{gi}} = \int\int_{T_S}^{T_F} Q_{gK}(t_1,t_2)\psi_{gi}(t_1)\psi_{gj}(t_2)dt_1dt_2, \tag{3.42}$$

and if we define a kernel:

$$J_K(t_1,t_2) \triangleq Q_{gK}(t_1,t_2) + 2E\left[Re\left\{\frac{\partial s^*(t_1,g_K(t_1))}{\partial g_K(t_1)} Q_n(t_1,t_2) \frac{\partial s(t_2,g_K(t_2))}{\partial g_K(t_2)}\right\}\right], \tag{3.43}$$

then the elements of matrix J can be written as

$$J_{ij} = \int\int_{T_S}^{T_F} [J_K(t_1,t_2)\psi_{gi}(t_1)\psi_{gj}(t_2)]dt_1dt_2. \tag{3.44}$$

To find the elements of the matrix J^{-1}, we can show (Appendix B) that the ijth element of J^{-1} is given by

$$J^{ij} = \int\int_{T_S}^{T_F} [J_K^{-1}(t_1,t_2)\psi_{gi}(t_1)\psi_{gj}(t_2)]dt_1dt_2, \tag{3.45}$$

where the function $J_K^{-1}(t_1,t_2)$ satisfies the equation

$$\int_{T_S}^{T_F} J_K^{-1}(t_1,u)J_K(u,t_2)du = \sum_{k=1}^{K} \psi_{gi}(t_1)\psi_{gj}(t_2). \tag{3.46}$$

Eq.(3.45) gives the formula for finding J^{-1}. However, we have to find $J_K^{-1}(t_1,t_2)$, and for that, we have to put Eq.(3.46) in a more useful form.

First, we define

$$R_{s'K}(t_1,t_2) \triangleq E\left[\frac{\partial s^*(t_1,g_K(t_1))}{\partial g_K(t_1)} \cdot \frac{\partial s(t_2,g_K(t_2))}{\partial g_K(t_2)}\right], \tag{3.47}$$

and similarly,

$$R_{s'}(t_1,t_2) \triangleq E\left[\frac{\partial s^*(t_1,g(t_1))}{\partial g(t_1)} \cdot \frac{\partial s(t_2,g(t_2))}{\partial g(t_2)}\right]. \tag{3.48}$$

Thus, Eq.(3.43) becomes

$$J_K(t_1,t_2) = Q_{gK}(t_1,t_2) + 2Re[R_{s'K}(t_1,t_2)Q_n(t,t_1)], \tag{3.49}$$

Substituting Eq.(3.49) into Eq.(3.46), multiplying by $R_{gK}(t_1,t_2)$, we have

$$\begin{aligned}&\int_{T_S}^{T_F} J_K^{-1}(t,u)[Q_{gK}(u,t_1) + 2Re\{R_{s'K}(u,t_1)Q_n(u,t_1)\}]R_{gK}(t_1,\tau)du \\ &= \sum_{k=1}^{K} \psi_{gi}(t_1)\psi_{gj}(t_2)R_{gK}(t_1,\tau).\end{aligned} \tag{3.50}$$

Now, integrating Eq.(3.50) with respect to t_1, we obtain

$$\begin{aligned}&\int\int_{T_S}^{T_F} J_K^{-1}(t,u)Q_{gK}(u,t_1)R_{gK}(t_1,\tau)dudt_1 \\ &+2\int\int_{T_S}^{T_F} Re[J_K^{-1}(t,u)R_{s'K}(u,t_1)Q_n(u,t_1)R_{gK}(t_1,\tau)]dudt_1 \\ &= \int_{T_S}^{T_F}\sum_{k=1}^{K}\psi_{gi}(t_1)\psi_{gj}(t_2)R_{gK}(t_1,\tau)dt_1.\end{aligned} \tag{3.51}$$

It can be shown (Appendix C) that the first integral on the left hand side of Eq.(3.51) is equal to $J_K^{-1}(t,\tau)$, and that the right hand side of Eq.(3.51) is $R_{gK}(t,\tau)$. Hence, with $K \to \infty$, we can write

$$J^{-1}(t,\tau) + 2\int\int_{T_S}^{T_F} Re[J^{-1}(t,u)R_{s'}(u,t_1)Q_n(u,t_1)R_g(t_1,\tau)]dudt_1 = R_g(t,\tau). \tag{3.52}$$

Returning to our signal model given in Eq.(3.7), let us define the error waveform as

$$e_g(t) = g(t) - \hat{g}(t), \tag{3.53}$$

and

$$e_I = \frac{1}{T_F - T_S}\int_{T_S}^{T_F}\{g(t) - \hat{g}(t)\}^2 dt = \frac{1}{T_I}\int_{T_S}^{T_F} e_g^2(t)dt, \tag{3.54}$$

where $T_I = T_F - T_S$ and e_I is the interval error. But e_I is a random variable, hence

$$\begin{aligned} T_I \cdot E[e_I] &= E\left[\int_{T_S}^{T_F}\left\{\sum_{k=1}^{\infty}(g_k - \hat{g}_k)\psi_{gk}(t)\right\}\left\{\sum_{l=1}^{\infty}(g_l - \hat{g}_l)\psi_{gl}(t)\right\}dt\right] \\ &= \sum_{k=1}^{\infty} E\left[(g_k - \hat{g}_k)^2\right], \end{aligned} \tag{3.55}$$

where orthonormality of the eigenfunctions has been used in the last step. Now the diagonal elements of J^{-1} are the lower bounds on the mean-square errors of the estimates. Thus,

$$E\{[g_k - \hat{g}_k]^2\} \geq J^{kk}. \tag{3.56}$$

Using Eq.(3.45) in Eq.(3.56) and applying the result to Eq.(3.55), we have

$$T_I \cdot E[e_I] \geq \underset{M\to\infty}{\text{l.i.m.}} \sum_{k=1}^{K} \int\int_{T_S}^{T_F} J_K^{-1}(t_1, t_2)\psi_{gk}(t_1)\psi_{gk}(t_2)dt_1 dt_2. \tag{3.57}$$

Using the property of orthonormal functions $\psi_{gk}(t_1)$ and $\psi_{gk}(t_2)$ such that

$$\sum_{k=1}^{\infty}\psi_{gk}(t_1)\psi_{gk}(t_2) = \delta(t_1 - t_2), \tag{3.58}$$

we have

$$E[e_I] \geq \frac{1}{T_I}\int_{T_S}^{T_F} J^{-1}(t,t)dt. \tag{3.59}$$

To evaluate the lower bound, therefore, we must solve Eq.(3.52) for $J^{-1}(t,\tau)$ and evaluate its trace as in Eq.(3.59). $J^{-1}(t,\tau)$ is referred to as the information kernel. In particular, if the noise is white with power spectral density N_0, then

$$Q_n(u,t_1) = \frac{1}{N_0}\delta(u - t_1), \tag{3.60}$$

and using Eq.(3.60) in Eq.(3.52), we have

$$J^{-1}(t,\tau) + \frac{2}{N_0}\int\int_{T_S}^{T_F} Re[J^{-1}(t,u)R_{s'}(u,t_1)\delta(u,t_1)R_g(t_1,\tau)]dudt_1 = R_g(t,\tau). \tag{3.61}$$

The bracket terms of the integrand in Eq.(3.61) are all real except the term $R_{s'}(u,t_1)$. Therefore, we can interchange the order of integration and taking the real part without

changing the result. This yields

$$J^{-1}(t,\tau) + \frac{2}{N_0} Re\left[\int_{T_S}^{T_F} J^{-1}(t,u) R_{s'}(u,u) R_g(u,\tau) du\right] = R_g(t,\tau). \tag{3.62}$$

But from Eq.(3.48)

$$R_{s'}(u,u) = E\left[\left|\frac{\partial s^*(u,g(u))}{\partial g(u)}\right|^2\right], \tag{3.63}$$

which is real, and since all other terms are also real, then Eq.(3.62) becomes

$$J^{-1}(t,\tau) + \frac{2}{N_0}\int_{T_S}^{T_F} J^{-1}(t,u) R_{s'}(u,u) R_g(u,\tau) du = R_g(t,\tau). \tag{3.64}$$

To solve for $J^{-1}(t)$, we define the following function:

$$P(t) = \sqrt{R_{s'}(t,t)} \tag{3.65}$$

$$Y^{-1}(t,\tau) = P(t) J^{-1}(t,\tau) P(\tau) \tag{3.66}$$

and

$$\Gamma_g(t,\tau) = P(t) R_g(t,\tau) P(\tau). \tag{3.67}$$

Multiplying both side of Eq.(3.64) by $P(t)$ and $P(\tau)$ respectively, we obtain

$$Y^{-1}(t,\tau) + \frac{2}{N_0}\int_{T_S}^{T_F} Y^{-1}(t,u)\Gamma_g(u,\tau) du = \Gamma_g(t,\tau). \tag{3.68}$$

Since the kernel $R_g(t,\tau)$ is symmetric and non-negative definite and $P(t)$ is a non-negative function, then $\Gamma_g(t,\tau)$ is also non-negative definite. Therefore, it can be expanded in a series such that

$$\Gamma_g(t,\tau) = \sum_{k=1}^{\infty} \mu_{gk}\phi_{gk}(t)\phi_{gk}(\tau), \tag{3.69}$$

where μ_{gk} and $\phi_{gk}(t)$ are, respectively, the kth eigenvalue and eigenfunction satisfying the integral equation

$$\int_{T_S}^{T_F} \Gamma_g(t,\tau)\phi_{gk}(\tau) d\tau = \mu_{gk}\phi_{gk}(t), \tag{3.70}$$

and $\{\phi_{gk}\}$ are orthonormal. Also in Eq.(3.45), the quantity J^{ij} is the correlation between the errors of estimating the ith and jth parameters. Since these parameters are independent, we can rewrite Eq.(3.45) such that

$$\int\int_{T_S}^{T_F} J_K^{-1}(t,\tau)\psi_{gi}(t)\psi_{gj}(\tau)dtd\tau = \sigma_{ei}^2\delta_{ij} \geq 0, \tag{3.71}$$

where σ_{ei}^2 is the variance of the error in the estimation of the ith parameter. Eq.(3.71) states that $J^{-1}(t,\tau)$ is symmetric and non-negative definite and further multiplication by the non-negative functions $P(t)$ and $P(\tau)$ yields $Y^{-1}(t,\tau)$ which is also non-negative definite. In order to find the solution $Y^{-1}(t,\tau)$ of Eq.(3.68), we can assume that it has the form of

$$Y^{-1}(t,\tau) = \sum_{k=1}^{\infty} \mu'_{gk}\phi_{gk}(t)\phi_{gk}(\tau). \tag{3.72}$$

Now, substituting (3.69) and (3.72) into (3.68), we obtain

$$\sum_{k=1}^{\infty} \mu'_{gk}\phi_{gk}(t)\phi_{gk}(\tau) + \frac{2}{N_0}\int_{T_s}^{T_F} \left\{\sum_{k=1}^{\infty} \mu'_{gk}\phi_{gk}(t)\phi_{gk}(u)\right\} \cdot \left\{\sum_{l=1}^{\infty} \mu_{gl}\phi_{gl}(u)\phi_{gl}(\tau)\right\} du$$
$$= \sum_{k=1}^{\infty} \mu_{gk}\phi_{gk}(t)\phi_{gk}(\tau). \tag{3.73}$$

Using the orthonormality of $\phi_{gk}(u)$ in the left-hand side of (3.73), the equation is simplified to

$$\sum_{k=1}^{\infty} \left(\mu'_{gk} + \frac{2}{N_0}\mu'_{gk}\mu_{gk}\right)\phi_{gk}(t)\phi_{gk}(\tau) = \sum_{k=1}^{\infty} \mu_{gk}\phi_{gk}(t)\phi_{gk}(\tau). \tag{3.74}$$

Equating coefficients in (3.74), we obtain

$$\mu'_{gk} = \frac{\mu_{gk}}{1 + \frac{2\mu_{gk}}{N_0}}. \tag{3.75}$$

Substituting (3.75) into (3.72) and then using the result in (3.66) with $\tau = t$, we have

$$J^{-1}(t,t) = P^{-2}(t)\sum_{k=1}^{\infty} \frac{\mu_{gk}}{1 + \frac{2\mu_{gk}}{N_0}}\phi_{gk}(t)\phi_{gk}(t). \tag{3.76}$$

Thus, using (3.76) in (3.59), the lower bound for the interval error in white noise can be expressed explicitly as

$$E[e_I] \geq \frac{1}{T_I}\sum_{k=1}^{\infty} \frac{\mu_{gk}}{1 + \frac{2\mu_{gk}}{N_0}} \cdot \int_{T_S}^{T_F} R_{s'}^{-1}(t,t)\phi_{gk}(t)\phi_{gk}(t)dt. \tag{3.77}$$

3.2 The Cramér-Rao Bound for the Frequency Estimation of the Signal $Ae^{2\pi j \int_0^t f(t_1)dt_1}$

3.2.1 Estimation of a continuous Non-Random Parameter

If the parameter to be estimated is non-random, then the lower bound for the variance of the estimation error can be developed in a similar way as in Section 3.1. Here the matrix J has only one component, such that

$$J = J_D, \tag{3.78}$$

where J_{Dij} is given by Eq.(3.28). To extend the theory from estimation of multiple parameters to that of a continuous function, we follow similar development of Eq.(3.29) through Eq.(3.37) with obvious consideration that $g(t)$ is a deterministic function. Hence, we can write

$$g(t) = \lim_{K\to\infty} \sum_{k=1}^{K} g_k \psi_{gk}(t), \tag{3.79}$$

where

$$g_k = \int_{T_S}^{T_F} g(t)\psi_{gk}(t)dt, \tag{3.80}$$

and $\psi_{gk}(t)$ satisfies the integral equation

$$\lambda_{gk}\psi_{gk}(t_1) = \int_{T_S}^{T_F} R_g(t_1,t_2)\psi_{gk}(t_2)dt_2. \tag{3.81}$$

The function $R_g(t_1,t_2)$ is the function defined as

$$R_g(t_1,t_2) = g(t_1)g(t_2). \tag{3.82}$$

The remaining steps are very much the same as those in Section 3.1 so that

$$\begin{aligned} J_{ij} &= 2E\left[\int\int_{T_S}^{T_F} Re\left\{\psi_{gj}(t_1)\psi_{gi}(t_2)\frac{\partial s^*(t_2,g_K(t_1))}{\partial g_K(t_1)}Q_n(t_1,t_2)\frac{\partial s(t_2,g_K(t_2))}{\partial g_K(t_2)}\right\}dt_1dt_2\right] \\ &= \int\int_{T_S}^{T_F} J_K(t_1,t_2)\psi_{gi}(t_1)\psi_{gj}(t_2)dt_1dt_2, \end{aligned} \tag{3.83}$$

where

$$\begin{aligned} J_K(t_1,t_2) &= 2Re\left\{\frac{\partial s^*(t_1,g_K(t_1))}{\partial g_K(t_1)}Q_n(t_1,t_2)\frac{\partial s(t_2,g_K(t_2))}{\partial g_K(t_2)}\right\} \\ &= 2Re\left\{R_{s'}(t_1,t_2)Q_n(t_1,t_2)\right\}, \end{aligned} \tag{3.84}$$

with $R_{s'}(t_1,t_2)$ being a deterministic value in this case and is given by

$$R_{s'}(t_1,t_2) = \frac{\partial s^*(t_1,g(t_1))}{\partial g(t_1)} \cdot \frac{\partial s(t_2,g(t_2))}{\partial g(t_2)}. \tag{3.85}$$

Finally, by Eq.(3.59), the lower bound on the interval error can be expressed as

$$E[e_I] \geq \frac{1}{T_I}\int_{T_S}^{T_F} J^{-1}(t,t)dt, \tag{3.86}$$

where $J^{-1}(t,\tau)$ satisfies the equation (with $Q_n(t_1,t_2)$ being given as Eq.(3.60)):

$$\frac{2}{N_0}\int_{T_S}^{T_F} J^{-1}(t,u)R_{s'}(u,u)R_g(u,\tau)du = R_g(t,\tau). \tag{3.87}$$

The evaluation of the lower bound in the case when the noise is white follows similar procedures as outlined at the end of Section 3.1. With the obvious omission of the first term on the left hand side of Eq.(3.62) in the case of non-random parameters, the expression for the lower bound on the interval error in white noise becomes

$$E[e_I] \geq \frac{N_0}{2T_I}\sum_{k=1}^{\infty}\int_{T_S}^{T_F} R_{s'}^{-1}(t,t)\phi_{gk}(t)\phi_{gk}(t)dt, \tag{3.88}$$

where $\phi_{gk}(t)$ is the eigenvector of the integral equation

$$\int_{T_S}^{T_F} P(t)R_g(t,\tau)P(\tau)\phi_{gk}(\tau)d\tau = \mu_{gk}\phi_{gk}(t), \tag{3.89}$$

and

$$P(t) = \sqrt{R_{s'}(t,t)}. \tag{3.90}$$

3.2.2 The Cramér-Rao Bound for the Frequency Estimation of the Signal

In this section we apply the results in Section 3.2.1 to obtain the lower bound for estimating the instantaneous frequency of the signal of the form

$$s(t) = Ae^{2\pi j \int_0^t F(t_1)dt_1}. \tag{3.91}$$

The estimation is assumed to be performed in white noise. The parameter to be estimated in this case is $F(t)$, and we only consider the case in which $F(t)$ is a deterministic function. Changing the subscript g to F accordingly, we have

$$R_g(t,\tau) \equiv R_F(t,\tau) = F^*(t)F(t). \tag{3.92}$$

Also, substituting (3.91) into (3.85), we have

$$\begin{aligned} R_{s'}(u,u) &= \left|\frac{\partial s(u, F(u))}{\partial F(u)}\right|^2 \\ &= \left|j\left\{2\pi\frac{\partial}{\partial F}\int_0^u F(t)dt\right\} Ae^{j2\pi\int_0^u F(t)dt}\right|^2 \\ &= 4\pi^2u^2|s(u)|^2 = 4\pi^2A^2u^2, \end{aligned} \tag{3.93}$$

since $|s(t)|^2 = A^2$.

Therefore, Eq.(3.88) becomes

$$E[e_I] \geq \frac{N_0}{8\pi^2T_I}\sum_{k=1}^{\infty}\int_{T_S}^{T_F}\frac{1}{A^2t^2}\phi_{Fk}(t)\phi_{Fk}(t)dt, \tag{3.94}$$

where $\phi_{Fk}(t)$ is the eigenfunction satisfying the equation

$$\int_{T_S}^{T_F} 4\pi^2A^2t\tau R_F(t,\tau)\phi_{Fk}(\tau)d\tau = \mu_{Fk}\phi_{Fk}(t). \tag{3.95}$$

The procedure to evaluate the error bound for deterministic instantaneous frequency is as follows.

1. Obtain the time-correlation function of the frequency, i.e., $R_F(t_1,t_2)$.

2. Solve for the eigenvalues and eigenfunctions in Eq.(3.95).

3. Calculate the error bound by Eq.(3.94)

It appears, at first sight, that there are infinitely many terms of finite values to be summed up in Eq.(3.94), and that the error bound is not convergent. However,careful consideration of the problem as outlined below shows that this is not so.

Substituting Eq.(3.92) into Eq.(3.95), we have

$$4\pi^2 A^2 tF(t) \int_{T_S}^{T_F} \tau F(\tau)\phi_{Fk}(\tau)d\tau = \mu_{Fk}\phi_{Fk}(t). \tag{3.96}$$

If we choose

$$\phi_{F1}(t) = ctF(t), \tag{3.97}$$

where c is a constant. It is clear that this choice satisfies the integral equation of Eq.(3.96) since substituting Eq.(3.97) into Eq.(3.96), we have

$$BctF(t) = \mu_{F1}\phi_{F1}(t), \tag{3.98}$$

where

$$B = 4\pi^2 A^2 \int_{T_S}^{T_F} \tau^2 F^2(\tau)d\tau. \tag{3.99}$$

Thus, the eigenvalue μ_{F1} corresponding to this eigenfunction ϕ_{F1} is given by

$$\mu_{F1} = B = 4\pi^2 A^2 \int_{T_S}^{T_F} \tau^2 F^2(\tau)d\tau. \tag{3.100}$$

The other eigenfunctions are orthonormal to this first eigenfunction such that

$$c\int_{T_S}^{T_F} tF(t)\phi_{Fk}(t)dt = \begin{cases} 1 & \text{for } k = 1 \\ 0 & \text{for } k \neq 1 \end{cases}. \tag{3.101}$$

Using this orthonormal condition in Eq.(3.96), we find

$$\mu_{Fk} = 0 \qquad \text{for } k \neq 1. \tag{3.102}$$

This implies that the dimensionality of $P(t)R_F(t,\tau)P(\tau)$ for the signal of Eq.(3.91) is unity if $F(t)$ is deterministic. Thus, there will be only one eigenfunction that is of interest to us and this is given by Eq.(3.97) where the constant c is selected so that

$$\int_{T_S}^{T_F} |\phi_{F1}(t)|^2 dt = 1, \tag{3.103}$$

i.e.

$$c = \left\{ \int_{T_S}^{T_F} t^2 F^2(t) dt \right\}^{-1/2}. \tag{3.104}$$

Thus, substituting Eq.(3.97) and Eq.(3.104) into Eq.(3.94), and ignoring other values of k except for $k = 1$, the error bound becomes

$$E[e_I] \geq \frac{N_0}{8\pi^2 A^2 T_I} \cdot \frac{\int_{T_S}^{T_F} F^2(t) dt}{\int_{T_S}^{T_F} t^2 F^2(t) dt}. \tag{3.105}$$

Now, the two integrals inside the braces in Eq.(3.105) are time varying, and therefore, for a fixed time interval T_I, their values depend on the initial time T_S. Here we assume that the instantaneous frequency of the signal is a polynomial in t such that

$$F(t) = \sum_{m=0}^{M} a_m t^m. \tag{3.106}$$

Note that for $M = 0$, we have a complex sinusoidal; for $M = 1$, we have a linear FM signal; and for $M = 2$, we have a quadratic FM signal. For $F(t)$ given by Eq.(3.106), we notice that the integrand $t^2F^2(t)$ in the denominator of Eq.(3.105) increases more rapidly than the integrand $F^2(t)$ in the numerator. Therefore, the higher is the value of T_S, the smaller is the bound on the right-hand side of Eq.(3.105). The maximum value of this bound is when

$$T_S = -T_I/2 \tag{3.107}$$

$$T_F = T_I/2. \tag{3.108}$$

Using these values as limits of integration, Eq.(3.105) becomes

$$E[e_I] \geq \frac{N_0}{8\pi^2 A^2 T_I} \cdot \frac{\int_{-T_I/2}^{T_I/2} F^2(t) dt}{\int_{-T_I/2}^{T_I/2} t^2 F^2(t) dt}, \tag{3.109}$$

which is the expected interval error of the estimate.

Chapter 4

Estimation of Time-Varying Frequency of a Signal

In the previous chapter, we derived the Cramér-Rao Bound (CRB) of estimating a continuous time-varying parameter in Gaussian white noise. Here in this chapter, the Wigner distribution will be applied to the estimation of time-varying frequency. The estimate bias and variance will be derived for comparisons and the relative comments will be given.

4.1 The Use of WD and PWD to Estimate the instantaneous Frequency of a Signal

As defined in section 2.1, there are two characteristic curves that are often used to analyze signals with time-varying frequency. The first curve gives the instantaneous frequency as a function of time, and the second curve presents the time delay as a function of frequency. For our purpose, we concentrate on the instantaneous frequency plotted as a time function.

As mentioned before, the signal we are interested is of the form

$$s(t) = Ae^{j\theta(t)}, \tag{4.1}$$

which, in general, does not satisfy (2.3). However, as explained in Section 2.1, in practice we only observe the signal for a finite duration. The signal $s(t)$ is *essentially band limited* and can be regarded as an analytic signal. Thus, the definition of instantaneous frequency given by (2.2) can be applied to $s(t)$.

The Wigner distribution offers an alternative method for calculating the instantaneous frequency of the signal $s(t)$ such that the instantaneous frequency is given by (see the properties of WD in the Section 2.4)

$$f(t) = \frac{\int_{-\infty}^{\infty} fW_s(t,f)df}{\int_{-\infty}^{\infty} W_s(t,f)df}. \tag{4.2}$$

Thus, for a fixed time t, the instantaneous frequency of a complex signal is given by the centre of gravity of the Wigner distribution. It has be shown in the section 2.4 that the pseudo-Wigner distribution is the concatenation of slices of Wigner distribution of the windowed signal $s(\tau)$. It can be further shown that [21] the pseudo-Wigner distribution would have the property of Eq.(4.2) provided that the window function has a zero instantaneous frequency which, in particular, is the case for a real valued window. Thus, using a real valued window, the pseudo-Wigner distribution can be employed to find the instantaneous frequency of a signal at any fixed time t by finding the centre of gravity of the distribution.

Although Eq.(4.2) can be used to estimate the instantaneous frequency of a single signal of the form given by Eq.(4.1) when either the WD or PWD is employed, difficulties arise when there present multiple signals. The main difficulty arises from the definition of instantaneous frequency of a sum of signals. Consider, for example, that the received signal consists of two exponential functions such that

$$\begin{aligned} s(t) &= A_1 e^{j\theta_1(t)} + A_2 e^{j\theta_2(t)} \\ &= a(t)e^{j\theta(t)}, \end{aligned} \tag{4.3}$$

where

$$a(t) = \sqrt{A_1^2 + A_2^2 + 2A_1A_2\cos(\theta_1(t) - \theta_2(t))} \tag{4.4}$$

$$\theta(t) = \tan^{-1}\frac{A_1\sin\theta_1(t) + A_2\sin\theta_2(t)}{A_1\cos\theta_1(t) + A_2\cos\theta_2(t)}. \tag{4.5}$$

The instantaneous frequency of this signal, according to the definition in Eq.(2.2), is given by

$$\begin{aligned} f(t) &= \frac{1}{2\pi}\frac{d}{dt}\theta(t) \\ &= a^{-2}(t)\left\{A_1^2\theta_1'(t) + A_2^2\theta_2'(t) + A_1A_2\cos(\theta_1(t) - \theta_2(t))[\theta_1'(t) + \theta_2'(t)]\right\}, \end{aligned} \tag{4.6}$$

where $\theta_i'(t)$ $(i = 1, 2)$ represents the derivative of $\theta_i(t)$ with respect to t.

The result in Eq.(4.6) is intuitively not very satisfactory since, if the signal is made up of two separate components, we would like to obtain the instantaneous frequencies of the two components. This is exactly what Eq.(4.6) fails to tell us. If we rewrite the signal in Eq.(4.3) as

$$s(t) = s_1(t) + s_2(t),$$

then using Eq.(4.2) to obtain the instantaneous frequency of $s(t)$ will lead to the result given by Eq.(4.6), but will not present to us the separate instantaneous frequencies of $s_1(t)$ and $s_2(t)$. Obviously, using the PWD, we will encounter similar difficulties.

To overcome this disadvantage, we assume here that the instantaneous frequency at a fixed time t is located at the peaks of the WD (or the PWD). Thus, by locating the peak of the WD of the signal, we are able to obtain the estimated instantaneous frequencies of a composite signal of the form of Eq.(4.3) at the instant t.

It must be emphasized here that when there are two or more signals present, the WD generates "interference terms" such that for two signals, for example, the WD is given by

$$W_{x+y}(t, f) = W_x(t, f) + W_y(t, f) + 2Re[W_{xy}(t, f)]. \tag{4.7}$$

As discussed in the Section 2.4, the interference term $2Re[W_{xy}(t,f)]$ lies between the two signal components $W_x(t,f)$ and $W_y(t,f)$, and is oscillatory such that its frequency decreases with the decreasing distance between the two signal components in the time-frequency plane. The magnitudes of the interference term could be as high as twice that of the auto-components, and therefore could cause problems in the determination of the peaks of the signal components. However, the interference terms can often be recognized by their oscillations. In the following analysis, we assume that these interference terms have been distinguished from the signal components and have been ignored, thus leaving us with the question of how accurate is the peak location method for the estimation of the instantaneous frequency of a signal. The method of distinguishing the interference terms will be discussed latter.

Another point that must be emphasized is that the assumption that the instantaneous frequency of the signal is located at the peak of the WD is valid only for linear FM signals (or short-time approximate linear FM signals in the case when the PWD is used). This is because the WD of a linear FM signal is an impulse function at the instantaneous frequency at any given instant, as proved in Section 2.4. Thus, when applied to estimate the instantaneous frequency of other signals, the peak of the WD would present bias. The ensuing section presents an analysis of the bias and variance of such an estimate.

4.2 Theoretical Performance Analysis

4.2.1 Bias and Variance Analysis in the Application of WD to Estimate the Instantaneous Frequency of a Signal

Let us consider the case that the received signal is accompanied by noise, i.e., the received signal is given by

$$x(t) = s(t) + n(t), \tag{4.8}$$

where, as was defined in Eq.(3.1), the signal is assumed to be of the form

$$s(t) = Ae^{j2\pi \int_0^t F(t_1)dt_1}, \tag{4.9}$$

and $n(t)$ is a complex Gaussian white process having properties given by Eqs.(3.3) and (3.4). The peak of the WD of the signal at a particular instant satisfies the equation

$$\frac{\partial}{\partial f} W_x(t,f) = 0. \tag{4.10}$$

Now, using the definition of Wigner distribution together with Eq.(4.8), we have

$$\frac{\partial W_x(t,f)}{\partial f} = \frac{\partial}{\partial f}\left[W_s(t,f) + W_n(t,f) + W_{sn}(t,f) + W_{ns}(t,f)\right] \tag{4.11}$$

$$= \int_{-\infty}^{\infty} -j2\pi\tau x(t+\frac{\tau}{2})x^*(t-\frac{\tau}{2})e^{-j2\pi f\tau}d\tau. \tag{4.12}$$

Assuming that at some particular instant t_0, the instantaneous frequency is $F(t_0)$, then in the neighbourhood of $F(t_0)$ we can expand $W_s(t_0,f)$ in a Taylor series such that

$$W_s(t_0,f) = W_s(t_0,F) + (f-F)\left.\frac{\partial W_s(t_0,f)}{\partial f}\right|_{f=F} + \frac{1}{2}(f-F)^2\left.\frac{\partial^2 W_s(t_0,f)}{\partial f^2}\right|_{f=F} + \cdots, \tag{4.13}$$

where $F(t_0)$ has been abbreviated to F.

Substituting Eq.(4.13) into (4.11), and also using Eq.(4.10), we obtain

$$-\left.\frac{\partial W_s(t_0,f)}{\partial f}\right|_{f=F} - (f-F)\left.\frac{\partial W_s^2(t_0,f)}{\partial f^2}\right|_{f=F} = \frac{\partial}{\partial f}\left[W_n(t_0,f) + W_{sn}(t_0,f) + W_{ns}(t_0,f)\right], \tag{4.14}$$

where the terms involving second and higher degrees of $(f-F)$ have been ignored. From Eq.(4.14), we obtain the estimated frequency to be

$$\hat{F}_W(t_0) = F - \left\{\left.\frac{\partial^2 W_s(t_0,f)}{\partial f^2}\right|_{f=F}\right\}^{-1}\left\{\left.\frac{\partial W_s(t_0,f)}{\partial f}\right|_{f=F} + \frac{\partial}{\partial f}\left[W_n(t_0,f) + W_{sn}(t_0,f) + W_{ns}(t_0,f)\right]\right\}, \tag{4.15}$$

which is equivalent to the estimated frequency obtained by locating the peak of $W_x(t,f)$. Note that the last two terms in Eq.(4.15) are complex conjugates, and thus $\hat{F}_W(t_0)$ remains real. From Eq.(4.15), we have

$$\begin{aligned} E[\hat{F}_W(t_0)] &= F - \left\{ \left. \frac{\partial^2 W_s(t_0,f)}{\partial f^2} \right|_{f=F} \right\}^{-1} \left\{ \left. \frac{\partial W_s(t_0,f)}{\partial f} \right|_{f=F} \right. \\ &\quad \left. + E\left[\frac{\partial}{\partial f}[W_n(t_0,f) + W_{sn}(t_0,f) + W_{ns}(t_0,f)] \right] \right\}. \end{aligned} \tag{4.16}$$

Now, by the assumption that the noise is a Gaussian white process with zero mean, we have

$$E\left[\frac{\partial}{\partial f} W_n(t_0,f)\right] = E\left[\int_{-\infty}^{\infty} -j2\pi\tau n(t_0+\frac{\tau}{2}) \cdot n^*(t_0-\frac{\tau}{2}) e^{j2\pi f\tau} d\tau\right] = 0, \tag{4.17}$$

and

$$E[W_{sn}(t_0,f)] = E[W_{ns}(t_0,f)] = 0. \tag{4.18}$$

Hence, Eq.(4.16) becomes

$$E[\hat{F}_W(t_0)] = F - \left\{ \left. \frac{\partial^2 W_s(t_0,f)}{\partial f^2} \right|_{f=F} \right\}^{-1} \left\{ \left. \frac{\partial W_s(t_0,f)}{\partial f} \right|_{f=F} \right\}. \tag{4.19}$$

Thus, the estimation of instantaneous frequency using the peak of the Wigner distribution contains a bias term given by

$$\beta_{\hat{F}_W}(t_0) = -\left\{ \left. \frac{\partial^2 W_s(t_0,f)}{\partial f^2} \right|_{f=F} \right\}^{-1} \left\{ \left. \frac{\partial W_s(t_0,f)}{\partial f} \right|_{f=F} \right\}, \tag{4.20}$$

for the instant of $t = t_0$. This bias term will vanish if the true instantaneous frequency of the signal is sufficiently close to the peak of the WD of the signal. To evaluate the variance of the estimation, we return to Eqs.(4.15) and (4.19) from which we obtain

$$\begin{aligned} \sigma^2_{\hat{F}_W}(t_0) &= E\left[|\hat{F}_W(t_0) - E\{\hat{F}_W(t_0)\}|^2 \right] \\ &= E\left[\left| \frac{\partial}{\partial f} \{W_n(t_0,f) + W_{sn}(t_0,f) + W_{ns}(t_0,f)\} \right|^2 \right] \left\{ \left. \frac{\partial^2 W_s(t_0,f)}{\partial f^2} \right|_{f=F} \right\}^{-2}. \end{aligned} \tag{4.21}$$

Dropping the obvious dependence of the Wigner distribution on t_0 and f for simplicity of notation, and using the fact that $W_{sn} = W_{ns}^*$, we have

$$\sigma^2_{\hat{F}_W}(t_0) = E\left[\left|\frac{\partial W_n}{\partial f}\right|^2 + 2\left|\frac{\partial W_{sn}}{\partial f}\right|^2 + 2\left(\frac{\partial W_n}{\partial f}\right)\left(\frac{\partial W_{sn}}{\partial f} + \frac{\partial W_{sn}^*}{\partial f}\right) + \left(\frac{\partial W_{sn}}{\partial f}\right)^2 + \left(\frac{\partial W_{sn}^*}{\partial f}\right)^2\right]\left(\frac{\partial^2 W_s(t_0,f)}{\partial f^2}\right)^{-2}_{f=F}. \quad (4.22)$$

The evaluation of the numerator of Eq.(4.22) involves the consideration of the second -, third-, fourth-order statistics of the noise process. For band-limited white Gaussian noise, it can be shown (Appendix D) that the variance of the estimated instantaneous frequency at t_0 can be simplified to

$$\sigma^2_{\hat{F}_W}(t_0) = \frac{2\pi^2 N_0 T_I^3}{3}(2A^2 + \sigma_n^2)\left(\frac{\partial^2 W_s(t_0,f)}{\partial f^2}\right)^{-2}_{f=F}, \quad (4.23)$$

where σ_n^2 is the total noise power.

Note that the variance of the frequency is still a function of t_0 since W_s is a function of t_0. We can combine the bias and the variance of the estimate to form the mean square error as the measure of performance, i.e.,

$$e^2_{\hat{F}_W}(t_0) = \beta^2_{\hat{F}_W}(t_0) + \sigma^2_{\hat{F}_W}(t_0). \quad (4.24)$$

4.2.2 Bias and Variance Analysis in the Application of PWD to Estimate the Instantaneous Frequency of a Signal

As mentioned before, for computational purposes, the PWD is commonly used instead of the WD in practice. The received signal $x(t)$ is the same as given by Eq.(4.8). Now, the PWD weights the signal by a window $h(t)$ so that the resultant family of signals is given by

$$\tilde{x}(t) = x(t) \cdot h(t - t_0) \quad (4.25)$$

$$= \tilde{s}(t) + \tilde{n}(t), \quad (4.26)$$

where

$$\tilde{s}(t) = s(t) \cdot h(t - t_0), \tag{4.27}$$

$$\tilde{n}(t) = n(t) \cdot h(t - t_0). \tag{4.28}$$

As defined in Eq. (2.39), the PWD $\tilde{W}_x(t, f)$ is the WD of $\tilde{x}(t)$ at $t = t_0$. Using Eq.(4.25), we have

$$\tilde{W}_x(t, f) = W_{\tilde{x}}(t, f) = \int_{-\infty}^{\infty} x(t + \frac{\tau}{2}) x^*(t - \frac{\tau}{2}) h(\frac{\tau}{2}) h^*(-\frac{\tau}{2}) e^{-j2\pi f\tau} d\tau. \tag{4.29}$$

In general, the window used is a real symmetric function of time such that $h(\tau) = h^*(-\tau)$, and thus Eq.(4.29) can be written as

$$\tilde{W}_x(t, f) = W_{\tilde{x}}(t, f) = \int_{-\infty}^{\infty} h^2(\frac{\tau}{2}) x(t + \frac{\tau}{2}) x^*(t - \frac{\tau}{2}) e^{-j2\pi f\tau} d\tau. \tag{4.30}$$

The analysis of the bias and variance in using the PWD to estimate the instantaneous frequency of a signal follows exactly the same procedure as in the analysis of the WD with the obvious substitutions of $W_{\tilde{s}}(t_0, f)$, $W_{\tilde{s}\tilde{n}}(t_0, f)$, and $W_{\tilde{n}\tilde{s}}(t_0, f)$ for $W_s(t_0, f)$, $W_{sn}(t_0, f)$, and $W_{ns}(t_0, f)$, respectively. In so doing, we employ Eqs.(4.27), (4.28) and (4.30) such that the PWD of the signal or of the noise is no more than the WD of the signal or noise weighted by the factor $h^2(\frac{\tau}{2})$. Thus, the properties of the weighted noise still maintain, i.e.

$$E[\tilde{n}(t)] = E[n(t)h(t - t_0)] = 0, \tag{4.31}$$

$$\begin{aligned} E[\tilde{n}(t_1)\tilde{n}^*(t_2)] &= h(t_1 - t_0)h(t_2 - t_0)E[n(t_1)n^*(t_2)] \\ &= N_0 h^2(t_1 - t_0)\delta(t_1 - t_2). \end{aligned} \tag{4.32}$$

The bias of the frequency estimate by PWD, in parallel to Eq.(4.20), is given by

$$\beta_{\hat{F}_{\tilde{W}}}(t_0) = -\left\{ \frac{\partial^2 W_{\tilde{s}}(t_0, f)}{\partial f^2}\bigg|_{f=F} \right\}^{-1} \left\{ \frac{\partial W_{\tilde{s}}(t_0, f)}{\partial f}\bigg|_{f=F} \right\}. \tag{4.33}$$

Similarly, with s and n substituted by $\tilde{s}$ and $\tilde{n}$, respectively, and bearing in mind that the PWD is defined as the WD of the weighted signal at $t = t_0$, the variance of the frequency estimate by the PWD, is given by

$$\sigma^2_{\hat{F}_W}(t_0) = 8\pi^2 N_0(A^2 + \sigma_n^2)\left[\int_{-\infty}^{\infty} \tau^2 h^4(\frac{\tau}{2})d\tau\right]\left(\frac{\partial^2 W_{\tilde{s}}(t_0, f)}{\partial f^2}\right)^{-2}_{f=F}. \tag{4.34}$$

Note that both the bias and the variance of the PWD depends on the window $h(t)$. In this chapter, we use a Hanning window as $h^2(\frac{\tau}{2})$, which is defined as [68]

$$h^2(\frac{\tau}{2}) = \begin{cases} 0.5 + 0.5\cos(\frac{\pi}{T_w}\tau), & \tau \in (-T_w, T_w) \\ 0, & \text{otherwise} \end{cases}, \tag{4.35}$$

where T_w is the total duration of the window. The reasons for the use of the Hanning window are as follows.

We need to choose a real and symmetric window such that the distortion is small, i.e., the PWD should be close to the WD. But from Eq.(2.39),

$$\bar{W}_s(t, f) = \int_{\infty}^{\infty} W_s(t, f) W_h(0, f - \eta)d\eta, \tag{4.36}$$

where

$$\begin{aligned} W_h(0, f - \eta) &= \int_{-\infty}^{\infty} h(\frac{\tau}{2})h(-\frac{\tau}{2})e^{-j2\pi(f-\eta)\tau}d\tau \\ &= \int_{-\infty}^{\infty} h^2(\frac{\tau}{2})e^{-j2\pi(f-\eta)\tau}d\tau. \end{aligned} \tag{4.37}$$

Thus, from Eq.(4.36), for $\bar{W}_s(t, f)$ to be identical to $W_s(t, f)$, $W_h(0, f - \eta)$ must behave like $\delta(f - \eta)$ on the t=0 line. Parallel to the theory of Fourier transform [68] , since the window $h^2(\frac{\tau}{2})$ has to be finite in duration for practical purposes, it follows that the spectrum of $h^2(\frac{\tau}{2})$ should have low sidelobes and a narrow mainlobe. The Hanning window satisfies these requirements and, furthermore, its functional form is simple.

Again we may use the mean square error to measure the performance of the PWD in estimating the frequency such that

$$e^2_{\hat{F}_W}(t_0) = \beta^2_{\hat{F}_W}(t_0) + \sigma^2_{\hat{F}_W}(t_0). \tag{4.38}$$

4.3 Comparison of the methods of frequency estimation

4.3.1 Discussions on the Theoretical Evaluation of Performance of the PWD

The bias and variance of the Wigner distribution have been developed in section 4.2. However, as mentioned before, due to the necessity in computation, a window of finite duration is used in practice, and this results in the use of the PWD, the bias and variance of which have also been evaluated for the case of band-limited white Gaussian noise having total noise power σ_n^2. Here we compare the performances of the PWD with reference to the Cramér-Rao lower bound. We confine our consideration to the cases of signals of the form in Eq.(4.9) which have deterministically time-varying frequencies. More specifically, we focus our attention to the signals having a single frequency of the linear FM or the quadratic FM type, i.e., the two types of time-varying frequencies that are considered are

$$F_1(t) = a_0 + a_1 t, \tag{4.39}$$

and

$$F_2(t) = a_0 + a_1 t + a_2 t^2. \tag{4.40}$$

In the small observation period or a suitable narrow window, most practical signal can be approximately represented as either linear FM or quadratic FM. For these signals, after deciding the period of observation T_I, the lower bound of the interval error can be evaluated for different signal to noise ratios using Eq.(3.109). Furthermore, for various signal to noise ratios defined as $10 \log \frac{A^2}{\sigma_n^2}$ (db) and for the two types of FM signals shown in Eqs.(4.39) and (4.40), we can evaluate $\beta_{\hat{F}_W}(t_0)$ and $\sigma^2_{\hat{F}_W}(t_0)$ and hence the mean square error $e^2_{\hat{F}_W}(t_0)$. Over the period of observation T_I, we can find the average mean square error $\bar{e}^2_{\hat{F}_W}(t_0)$.

Although it is true, in general, that the bias and variance of the frequency estimate using the PWD is dependent on t_0, for the linear and quadratic FM signals, it can be shown in the following that the bias and variance are independent of t_0.

Since linear FM signal is a particular situation of quadratic FM signal with $a_2 = 0$, we consider only quadratic FM signal. Using Eq.(4.40) in Eq.(4.9), we have

$$\begin{aligned} s(t_0 + \frac{\tau}{2})s^*(t_0 - \frac{\tau}{2}) &= A^2 e^{j2\pi[(a_0 + a_1 t_0 + a_2 t_0^2)\tau + (a_2\tau^3/12)]} \\ &= A^2 e^{j2\pi(F(t_0)\tau + a_2\tau^3/12)}. \end{aligned} \tag{4.41}$$

Substituting Eq.(4.41) into the definition of PWD of Eq.(4.29) where a real and symmetric window with finite duration $[-T_w, T_w]$ is used, we obtain, after differentiating

$$\left.\frac{\partial^n W_s(t_0, f)}{\partial f^n}\right|_{f=F} = A^2 \int_{-\infty}^{\infty} (-j2\pi\tau)^n h^2(\frac{\tau}{2}) e^{ja_2\tau^3/12} d\tau, \tag{4.42}$$

which is independent of t_0. Note that for a given linear or quadratic FM of a constant amplitude and for a particular window size, the bias of the frequency estimate remains constant. Notice also that $\beta_{\hat{F}_W}(t_0)$ is zero for linear FM signal since for $n = 1$, Eq.(4.42) is the integral of an odd function. This, of course, is due to the fact that the WD of a linear FM is an impulse function at the instantaneous frequency for any given time instant. From Eqs.(4.33), (4.34), and (4.42), we can conclude that for the linear and quadratic FM signal, the bias and variance of the estimate by the PWD is independent of t_0. Thus, for these two types of signals, the mean square error at any instant t_0 is in fact the mean square error averaged over the entire observation interval T_I, i.e.,

$$\begin{aligned} \bar{e}^2_{\hat{F}_W} &= \frac{1}{T_I}\int_{T_S}^{T_F} e^2_{\hat{F}_W}(t_0)dt_0 \\ &= \beta^2_{\hat{F}_W}(t_0) + \sigma^2_{\hat{F}_W}(t_0) \\ &= \varepsilon^2_{\hat{F}_W}(t_0) \quad \forall t_0 \in T_I. \end{aligned} \tag{4.43}$$

It can also be observed from Eq.(4.42) that for the linear FM signal, the derivatives of the WD with respect to f are independent of the parameters a_0 and a_1, whereas, for

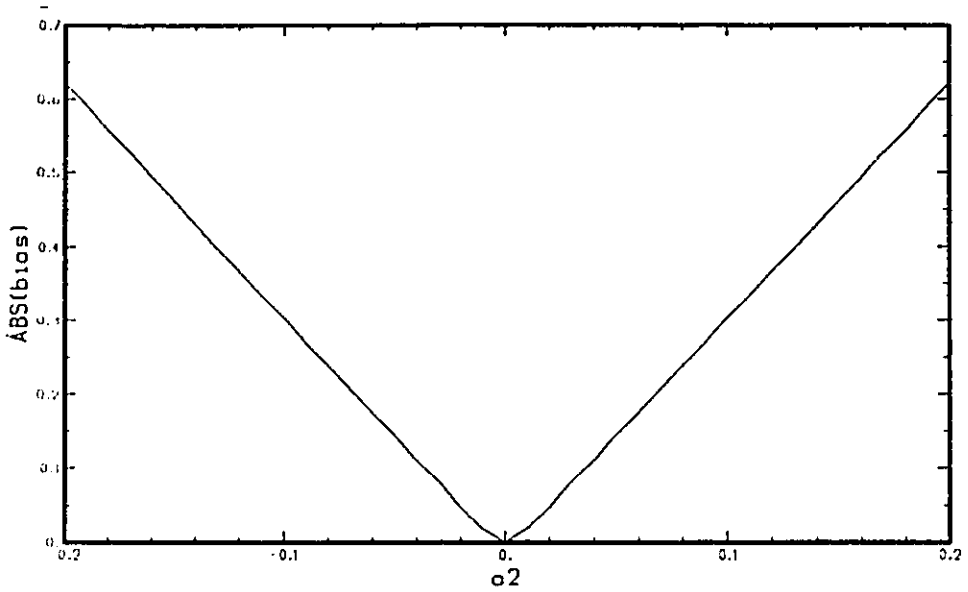


Figure 4.1: The Effects of a_2 on the Estimation Bias of the PWD Estimate for the Quadratic FM Signal.

the quadratic FM signal, the derivatives are dependent only on a_2 but not on a_0 or a_1. Therefore, a_0 and a_1 have no bearing on the bias and variance of the estimate. Figs 4.1 and 4.2 depict, respectively, the effects of the parameter a_2 on the bias and the MSE of the PWD estimate for the quadratic FM signal characterized by Eq.(4.40) with the window width $T_w = 8$ *sec.*. Judging from these two figures, it can be concluded that the larger a_2 is, the worse is the bias and the error.

Further insight to this observation can be gained by considering the WD of the quadratic FM signal which, after straightforward substitution, can be written as

$$\begin{aligned} W_s(t_0, f) &= 2\int_0^\infty A^2 \cos\left[\frac{\pi a_2}{6}\tau^3 - 2\pi\{f - F(t_0)\}\right] d\tau \\ &= 2\pi A^2 \left(\frac{\pi a_2}{2}\right)^{-\frac{1}{3}} Ai\left[-2\pi\left(\frac{\pi a_2}{2}\right)^{-\frac{1}{3}}\{f - F(t_0)\}\right], \end{aligned} \tag{4.44}$$

where $Ai(\cdot)$ is the Airy function [1] the peak of which occurs at

$$-2\pi\left(\frac{\pi a_2}{2}\right)^{-\frac{1}{3}}[f - F(t_0)] = 1. \tag{4.45}$$

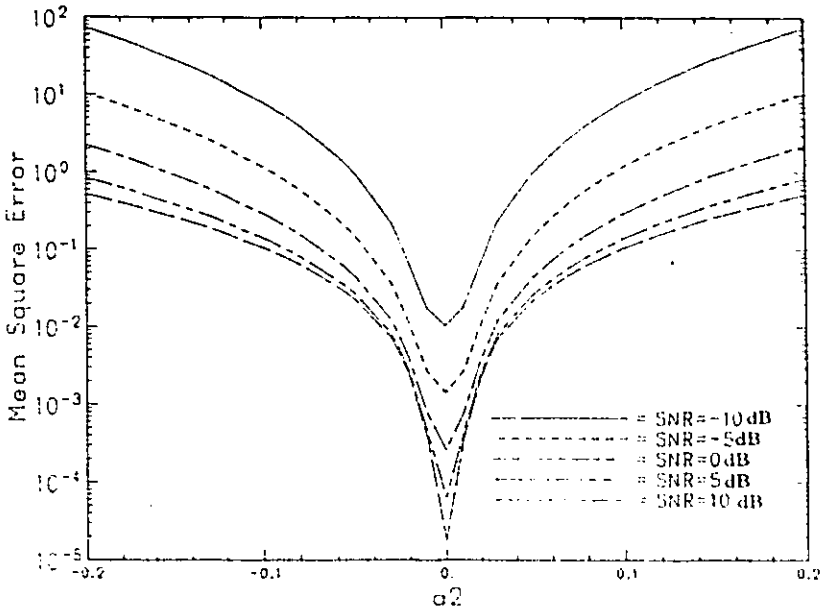


Figure 4.2: The Effects of a_2 on the Estimation MSE of the PWD Estimate for the Quadratic FM Signal.

Thus, if the peak frequency is used as an estimate for the instantaneous frequency, the bias is given by

$$|\beta_{\hat{F}_W}(t_0)| = \frac{1}{2\pi}\left(\frac{\pi a_2}{2}\right)^{-\frac{1}{3}}. \tag{4.46}$$

Eq.(4.46) shows that the bias of the estimate of the instantaneous frequency of the quadratic FM signal using the peak of its WD is independent of t_0 as well as independent of a_0 and a_1 as was pointed out earlier. Eq.(4.46) also states that the bias is proportional to the cube root of the coefficient a_2, whereas Fig 4.1 depicts that the relationship is more or less linear. The difference is due to the fact the Eq.(4.46) is derived from the WD of the signal, whereas Fig.4.1 is evaluated for the PWD of the signal in which, as mentioned in Section 4.2.2, a Hanning window is applied together with the assumption that the peak frequency is in the close neighbourhood of the instantaneous frequency, in which case the higher order terms of $(f - F)$ can be ignored. Under such an assumption, if a relatively narrow window is employed or, equivalently, if a_2 is small, the Fig 4.1 holds true.

An important factor that affects the performance of the various methods of frequency estimation is the window size used. As mentioned in Section 4.2.2, we use the Hanning window for the PWD. However, for different window sizes, the mean square errors are different as is obvious from Eqs.(4.33) and (4.34). Thus, the problem is to choose a total window duration T_w so that the mean square error is a minimum. This depends on the signal that is being considered.

As examples, we choose linear and quadratic FM signals whose frequencies are given, respectively, by

$$F_1(t) = -2 + 0.5t, \tag{4.47}$$

and

$$F_1(t) = -1 + 0.025t^2. \tag{4.48}$$

For the linear FM signal, we have already seen that the PWD estimate is unbiased, therefore, the window size is chosen to minimize the variance. Substituting the expression of Hanning window into Eq.(4.34) and using Eq.(4.42), we have

$$\sigma^2_{\hat{F}_W} = \frac{N_0(2A^2 + \sigma_n^2)}{2\pi^2 A^4} \cdot \frac{\int_{-T_w}^{T_w} \frac{\tau^2}{4}\left(1 + \cos\frac{2\pi\tau}{T}\right)^2 d\tau}{\left[\int_{-T_w}^{T_w} \frac{\tau^2}{2}\left(1 + \cos\frac{2\pi\tau}{T}\right) d\tau\right]^2}. \tag{4.49}$$

We note that the expression of the variance for the linear FM signal in Eq.(4.49) is independent of the signal parameters a_0 and a_1. As T_w increases, the denominator of Eq.(4.49) increases faster than the numerator. Therefore, the larger the window is, the lower the variance and the higher the accuracy of the estimate. The mean square errors of the frequency estimate at various SNR are plotted in Fig 4.3. Thus, for the linear FM signal, we should choose T_w as large as possible. However, a large T_w increases the computational burden. Thus, the choice of T_w for a linear FM signal is a compromise between the acceptable accuracy and the computational complexity.

For the quadratic FM signal, the PWD estimate is no longer unbiased, and hence we choose T_w to minimize the mean square error. Again using the expression of Hanning

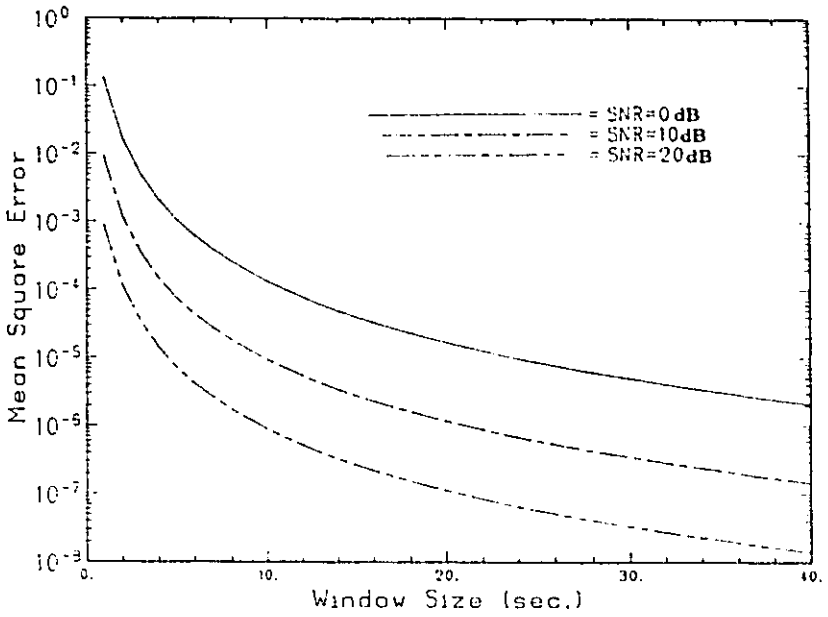


Figure 4.3: The Relation between the MSE of Frequency Estimation using PWD and the Window Size for Linear FM.

window and Eq.(4.42) in Eqs.(4.33) and (4.34), we can obtain an expression of the mean square error in terms of T_w. This expression is also dependent on the SNR as well as on the parameter a_2 of the quadratic FM signal. For various SNR and $a_2 = 0.025$, the dependence of the mean square values on the window size is shown in Fig 4.4. The minima of these curves can, in principle, be obtained by differentiating the expression with respect to T_w. However, this turns out to be quite complicated, and a general closed-form expression for the optimum value of T_w may not be easily obtained. Also, a word of caution has to be mentioned here. In computing the PWD, the DFT (Discrete Fourier Transform) is employed. For the DFT, the width of the frequency bin is inversely proportional to the length of the data (window size) if no padding of zeros is employed, i.e., the bin-width is given by

$$f_b = \frac{1}{2T_w}.$$

If we assume that the probability density function of locating the true frequency is

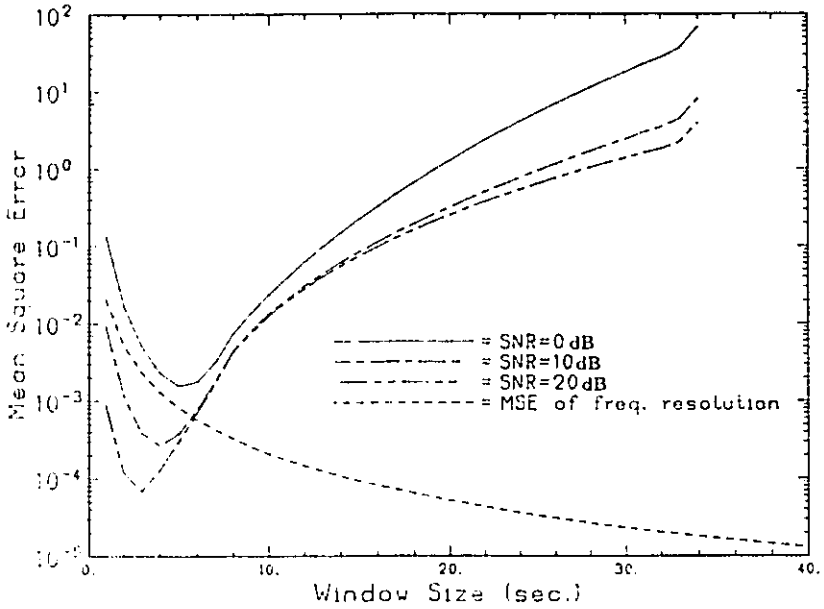


Figure 4.4: The Relation between the MSE of Frequency Estimation Using PWD and the Window Size for Quadratic FM.

uniformly distributed within the frequency bin, then the mean square error of frequency resolution is given by

$$e_f^2 = \frac{1}{12(2T_w)^2}. \tag{4.50}$$

Therefore, to choose an optimum window size T_w for the quadratic FM signal, we have to consider two factors, viz., the mean square error of estimation $\bar{e}^2_{\hat{F}_W}$ and the mean square error of resolution e_f^2. If we know the signal parameters and SNR, the values of $\bar{e}^2_{\hat{F}_W}$ can be plotted for various window sizes T_w as shown in Fig 4.4. Furthermore, e_f^2 can be plotted, as is also shown in Fig 4.4, for various values of T_w according to Eq.(4.50). We can then choose an optimum value of T_w which minimizes the higher of these two errors, i.e., our choice is

$$\min_{T_w}[\max(\bar{e}^2_{\hat{F}_W}, e_f^2)],$$

which results in a choice of T_w such that

$$\bar{e}^2_{\hat{F}_W}(T_w) = e_f^2(T_w). \tag{4.51}$$

In practice, we may not know either the parameters of the signal or the SNR. Under such circumstances, a guess of the window size taking into account the computational burden will be the only plausible choice.

4.3.2 Evaluation of Performances by Computer Simulation

The theoretical analysis of the errors involved in the frequency estimate using the WD (or the PWD) is based on the location of the peaks at various instants. Furthermore, it is assumed that the true instantaneous frequency indeed lies at the peaks of the time-frequency plane of PWD. This assumption may not be valid under very low SNR.

In order to test the validity of the theoretical evaluation of the errors in the various estimates, computer simulations are carried out in which a signal (of either the linear FM or quadratic FM type) is mixed with additive Gaussian white noise. The signals we use here are the same as given by Eqs.(4.47) and (4.48). The total duration of the received data is 16 seconds and the sampling rate of the data is 0.125 second. For the PWD estimation, we use a Hanning window of duration 8 seconds which is close to the optimum window size for the quadratic FM signal with $a_2 = 0.025$ in the theoretical evaluation of the PWD. For the linear FM signal, we choose the window size which is the same as that for the quadratic FM signal, so that the computational complexity in the two cases are the same. The following is a general description of the simulation procedure.

1. For a particular SNR, at each shift of the window, the data samples within the window are used to evaluate the frequency on the time-frequency plane.

2. At this point, the PWD is a presentation of the received data in the time-frequency plane. There are 64 steps in the time direction, each of 0.125 sec. in length.

3. At each time instant, the peak of PWD is located, and the frequency at which

the peak situated is assigned as the estimate of the instantaneous frequency of the signal.

4. Now, at this stage, the PWD will now have a time-frequency plane on which the estimated instantaneous frequency is shown at various instants. At each time instant the difference between the true instantaneous frequency and the estimated one is measured. The bias and the error square averaged over the observation time are then computed.

5. The above procedure is then repeated 300 times with Gaussian white noise of the same power so that a mean bias and a mean square error (averaged over the observation period) are computed.

6. The above procedure is repeated for various SNR.

4.3.3 Comparison of Theoretical and Simulated Performance

Fig.4.5 shows the theoretical and simulated biases for the linear FM signal whose frequency is given by Eq.(4.47). As was derived earlier, the PWD estimate is theoretically unbiased. The simulation results confirm this under high SNR, which means that under high SNR, our assumptions made in the theoretical derivations are reasonable and the simulation results are quite close to what we expect in the theoretical performance. However, when the SNR decreases below 0dB, the computer simulation depart from the theoretical calculations. The reason is that our derivations are based on the high SNR and the peak is close to the true instantaneous frequency and this assumption will not be true under low SNR.

Fig.4.6 shows the theoretical and simulated biases of the estimates for the quadratic FM signal whose frequency is given by Eq.(4.48). Here, simulation results show that the estimate has a constant bias at high SNR as predicted in the theoretical analyses

for PWD. However, as SNR decrease, the biases depart from the theoretical derivation. The explanation for this is similar to that in the case of the linear FM. One thing should be noted: since the assumption that the peak position is the position of instantaneous frequency may not be exactly true for quadratic FM signal (it is only an approximation), the theoretical and simulated biases, even at high SNR, will not coincide as well as we observed in linear FM situations, and a small difference can be noticed.

Figs.4.7 and 4.8 show, respectively, the averaged mean square error of the PWD estimate of the linear FM and quadratic FM signals. The Cramér-Rao lower bounds were also plotted as references. It can be observed that at high SNR (above 5dB), the performances are almost identical to those predicted by theoretical analysis, which confirms that the higher order terms of $(f - F)$ are truly negligible in the derivation. However, at low SNR the assumption will not be true any more and the higher order terms of $(f - F)$, as stated before, cannot be ignored. It can be observed that as SNR continues to increase, the MSE for the linear FM signal keeps on increasing. On the other hand, for the quadratic FM signal, there exists an irreducible error no matter how much the SNR increases. This is due to the fact that for the linear FM signal, the PWD estimate is theoretically unbiased (and is unbiased in practice if the SNR is high). However, for the quadratic FM signal, there is a constant bias as illustrated in Fig.4.6. This constant bias contributes to the irreducible error in the estimate.

4.3.4 Comparison with Spectrogram and Discussions

Estimate performance

As a matter of interest, we compare the performance of the spectrogram in the estimation of the instantaneous frequency of the two FM signals to that of the PWD. A theoretical analysis of the performance can be found in Appendix E. The results from simulations

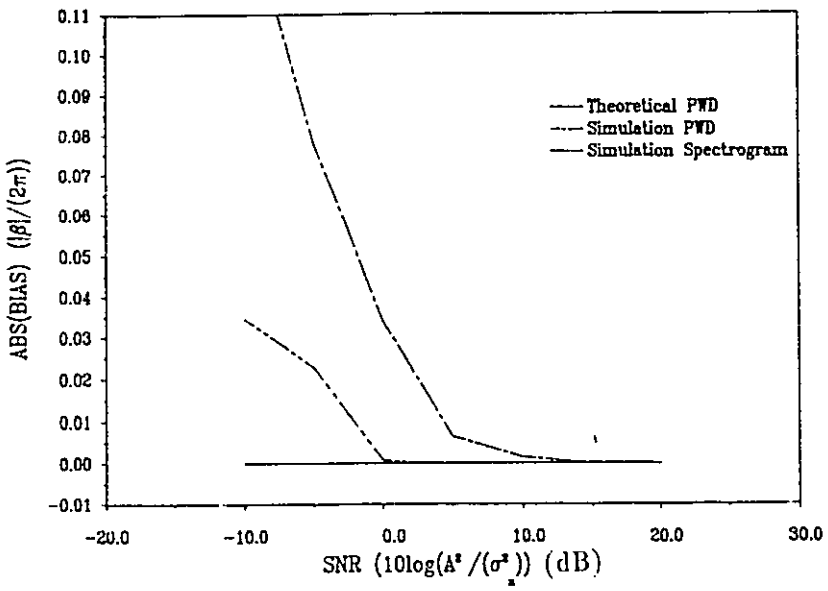


Figure 4.5: The Theoretical and Simulated Bias for the Linear FM Signal.

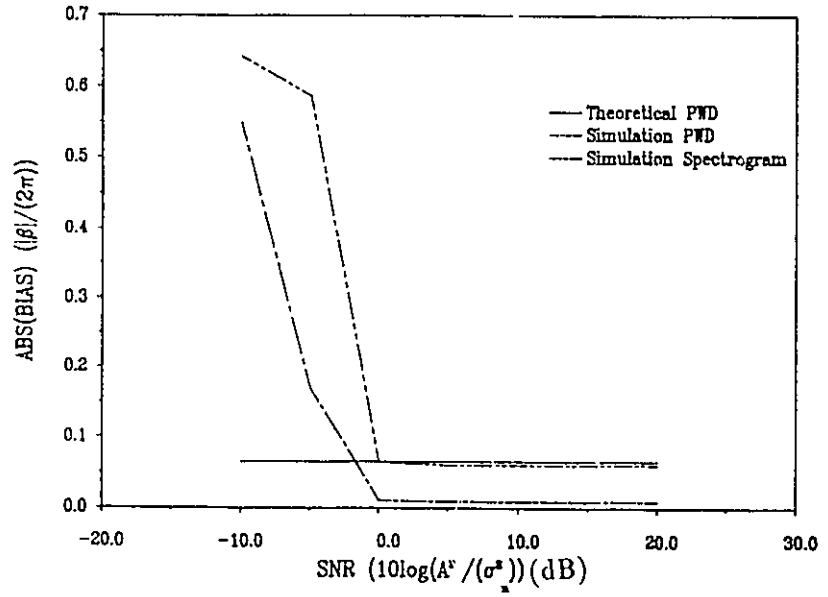


Figure 4.6: The Theoretical and Simulated Bias for the Quadratic FM Signal.

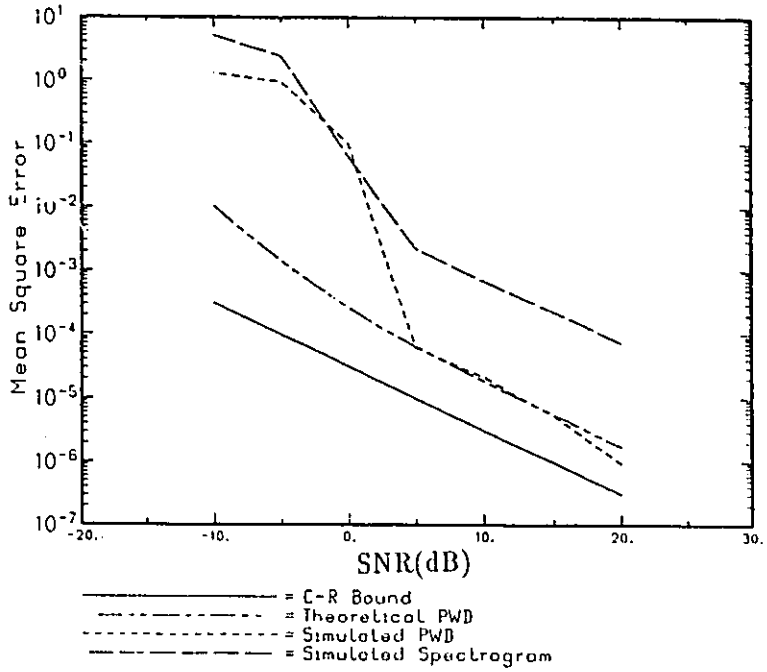


Figure 4.7: The Theoretical and Simulated MSE for the Linear FM Signal.

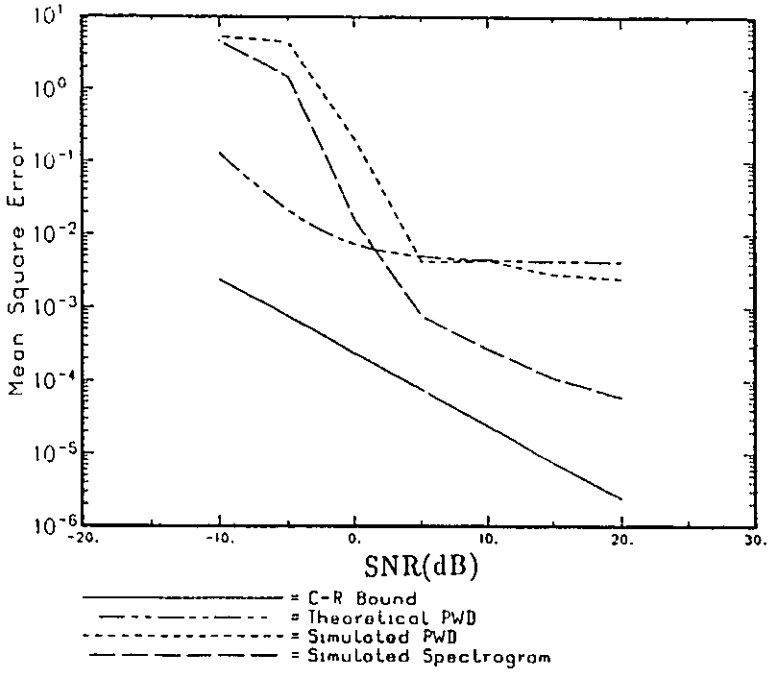


Figure 4.8: The Theoretical and Simulated MSE for the Quadratic FM Signal.

are plotted correspondingly in Figs.4.5 to 4.8, where the Gaussian function discussed in Eq.(2.9) with $\sigma_y = .05msec.$ is selected as the window function.

It can be observed that for the linear FM signal, the spectrogram has a much larger bias than the PWD. On the other hand, for the quadratic FM signal, the bias obtained by using the spectrogram is consistently lower than that of the PWD. The MSE (Mean Square Error) obtained in estimating both types of signals by the spectrogram do not differ significantly, and in the case of the quadratic FM signal actually outperforms the PWD under very high SNR (above 10 dB) due to the irreducible bias exhibited by the use of the PWD. However, in general, the accuracy in the estimation is considerably high when the PWD is employed.

Effect of the interference terms

It has been mentioned that when there are two or more signals present, the PWD generates the interference terms which may be detrimental to the location of the peaks. It has also been suggested that the interference terms can be distinguished from the signal terms by their oscillatory nature. Here we illustrate by examples how the interference terms may affect the location of the peaks. Fig.4.9 (a), (b) and (c) show, respectively, the topographic plots of the PWD with two linear FM signals, one linear FM and one quadratic FM signals, and two quadratic FM signals. In all these cases, no noise is added. And in Fig.4.9, a point whose value is positive and larger than 70% of the peak value at a particular instant is represented by a $\times$, whereas a point whose value is negative and less than 70% of the minimum value at a particular instant is represented by a $\odot$. It is clear from the examples that the interference terms exist, but they are characterized by the alternating maxima and minima. Quite apart from the interference terms, it can also be observed that when the instantaneous frequencies are close to each other, the peaks of the signal terms interfere with each other thus causing bias in the location of

the peaks. This is especially serious in the cases involving the quadratic FM signal. The reason is that the PWD of a linear FM signal has much narrower main lobe than that of a quadratic FM signal. Thus the error analysis described in the previous section is applicable only when the two instantaneous frequencies are sufficiently separated from each other such that their individual PWDs do not interfere with each other.

Low SNR test

Fig.4.10 shows the PWD of the same signals embedded in White Gaussian noise. The SNR in all cases is -5dB. In each case, we only take one snap shot of the signals and noise. Now the situation of having spurious peaks and minima worsens. The oscillatory characteristics of the interference terms of PWD cannot be recognized easily leading to possible errors in the location of peaks of the signal terms. Again, the deterioration is more serious in the cases involving quadratic FM signals.

In many applications such as radar, active sonar, and SARSAT systems, the FM signal bursts are usually repetitive. Hence, we may take several snapshot. This process helps to eliminate some of the spurious peaks and minima, and increases the accuracy of locating the peaks of the signal terms. Figs.4.11 shows, respectively, the averaged PWD of the same signals embedded in white Gaussian noise of -5dB. The plots are obtained by averaging over 15 snapshots. It can be observed that most of the spurious peaks and minima have been eliminated. The identification of the interference terms of PWD using the oscillatory characteristics is possible even at SNR=-5dB using 15 snapshots.

Summary

In this chapter, we have reviewed the methods of the PWD in the representation of a signal in the time-frequency plane. We focused our attention on the application of the method to estimate the instantaneous frequency of a signal. The performance was

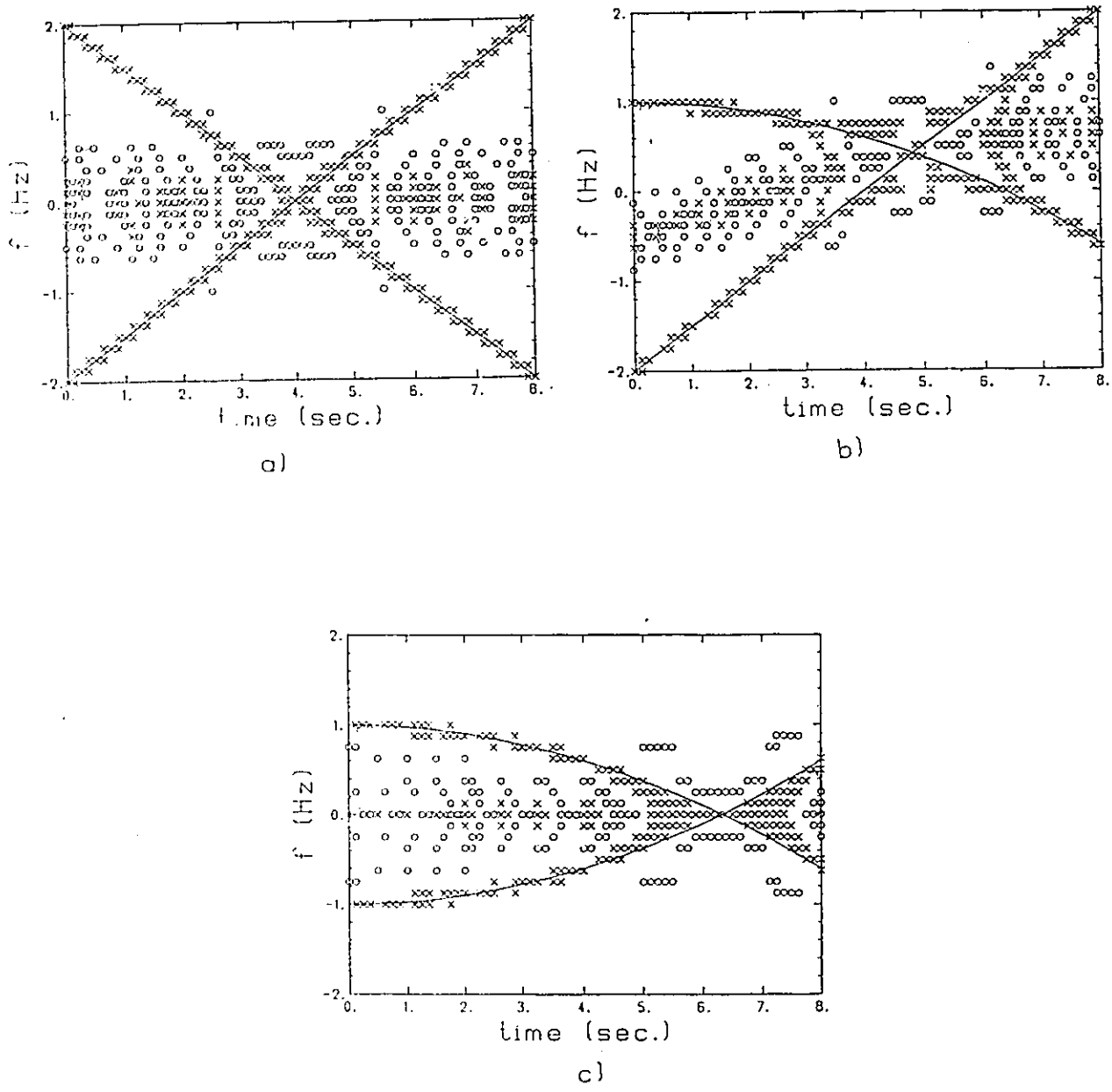


Figure 4.9: The Topographic Plots of the PWD with a) Two Linear FM Signals b) One Linear FM and One Quadratic FM Signals and c) Two Quadratic FM Signals.

analyzed. Applying of the method to specific cases of linear and quadratic FM signals revealed various interesting results:

1. The PWD is very suitable for the tracking of linear FM but not very good for the quadratic FM due to the bias. Also, its resolution is restricted by the frequency bin width which is equal to the reverse of window width.

2. The use of the PWD has the main problem of cross- interference terms, especially when two or more signals are received simultaneously. However, this cross- interference terms can be distinguished from the true instantaneous frequency peak by their oscillatory characteristic.

3. The performance of the PWD estimate is not close to the theoretical limit given by the Cramér-Rao Bound, even for linear FM signals. It also exhibits threshold phenomenon at low SNR.

It is possible to improve the performance of instantaneous estimation by using the smoothed pesudo Wigner distribution (SPWD), introduced in 1983 by P. Flandrin [67]. The SPWD offers sdditional smoothing of the cross interference terms and allows independent smoothing in time and frequency [30]. However, the PWD is simpler in computation and easier for analysis, and the idea employed in this chapter for the performance analysis can be easily adopted to the SPWD.

There also exist some other methods which can be used in the time-varying frequency estimate, such as PIE (Phase Interpolate Estimate) [65], spectrogram (which has been mentioned briefly in this chapter) and transformed spectrogram [56]. The limitations and advantages of all these methods have been discussed in [106]. Cohen and others [22] also proposed a method of tracking unknown chirp signals using unsupervised clustering. The method involves the maximum likelihood estimation of the signal parameters and the decision of the number of nonoverlapping line segments representing the instantaneous

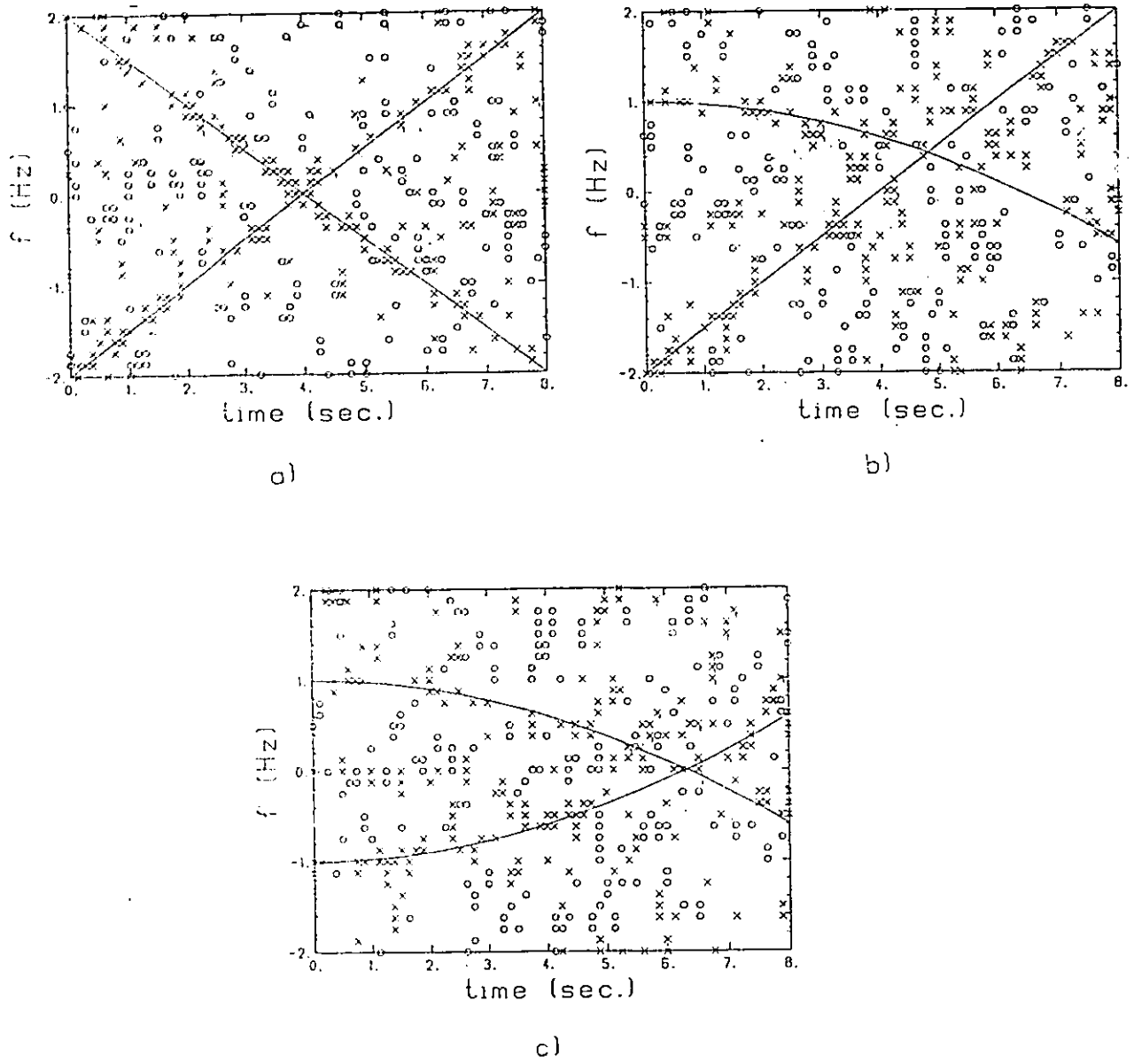


Figure 4.10: The Topographic Plots of the PWD under SNR= -5dB and 1 snapshot, with a) Two Linear FM Signals b) One Linear FM and One Quadratic FM Signals c) Two Quadratic FM Signals.

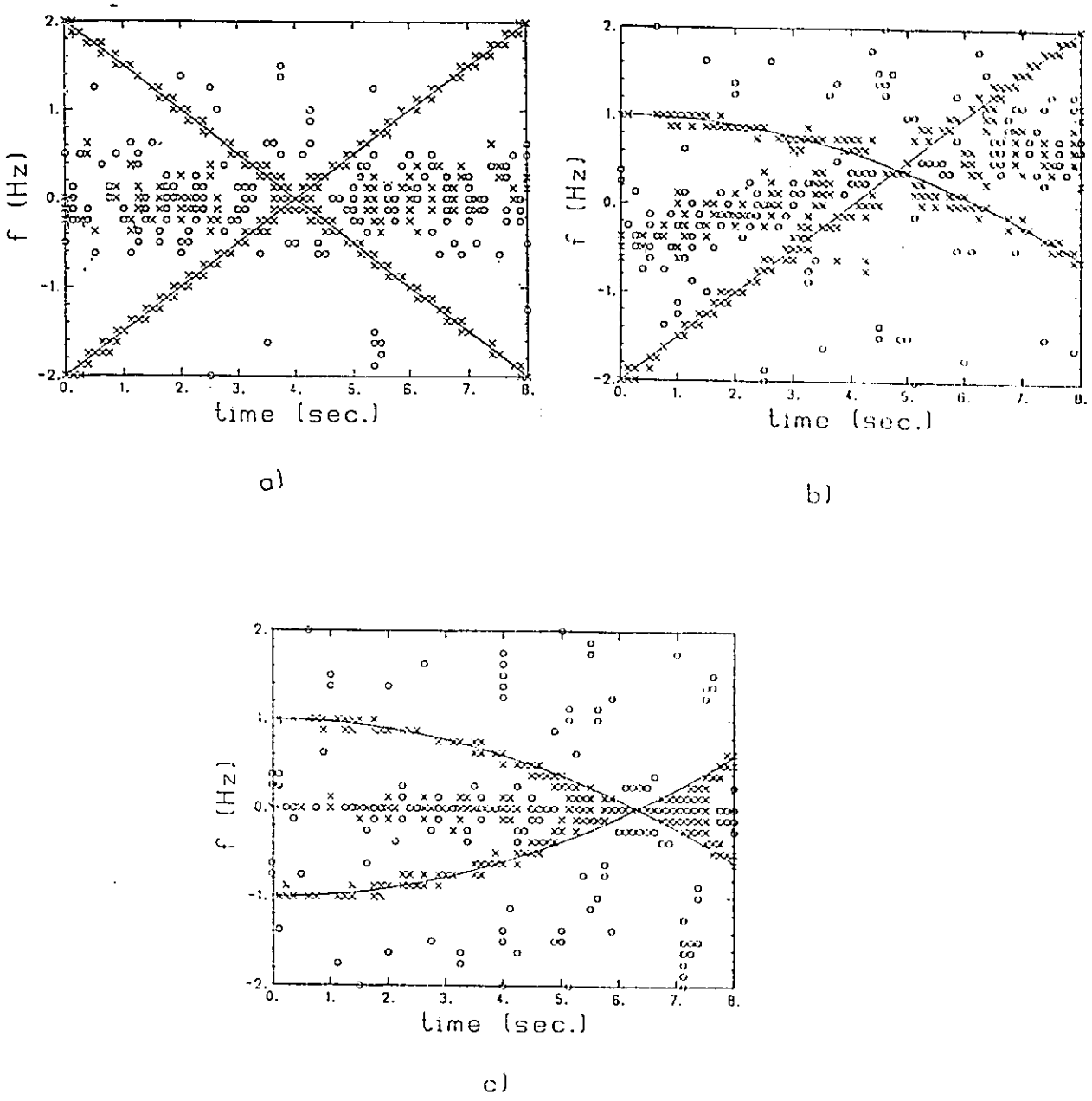


Figure 4.11: The Topographic Plots of the PWD under SNR= -5dB and averaging of 15 snapshots, with a) Two Linear FM Signals b) One Linear FM and One Quadratic FM Signals c) Two Quadratic FM Signals.

frequencies of the signal. This method, which is based on the PWD as the first step, is applicable to signals with multi-frequencies and appears to work well under low SNR. However, the peak location method discussed in this chapter has the main advantage of being simple in computation.

All the work in this chapter and the previous one, which includes the derivation, the analyses and the simulations, was finished in 1987 [106], and a full paper was published on the IEEE trans. on ASSP in March 1989 [105]. Since then, the research on the instantaneous frequency estimation is still very active. Boashash and his research group have published many papers over these years [7], [8], [9] and [11]. They described the methods similar conceptually to ours in that they estimate the instantaneous frequency of a signal using WD techniques. They have also extended the method to other application fields [6], [10], [103] and [104], and discussed its performance under various channel backgrounds [99] to [102].

Chapter 5

Accurate Frequency Estimation of SARSAT Signal

Search and rescue satellite-aided tracking (SARSAT) system employs satellites to locate the accident site in both air and marine search and rescue. The signal from the emergency locator transmitter (ELT) of a down aircraft or the emergency position-indicating radio beacon(EPIRB) unit of a marine vessel in distress is received by satellites and then is sent to an earth station, where the signal is analyzed to extract the position of the accident. In this chapter, we concentrate on the signal from ELT unit. The analysis of the signal from EPIRB unit will be similar to that of the signal from ELT unit.

In this chapter, we will apply the PWD to the ELT signals and show how it is superior to other spectrum estimation methods [18]. A high resolution method will be introduced in combination with PWD, which works quite well for the non-stationary short data record. The real SARSAT data will be used to test the effectiveness of this method.

5.1 Signal Model

Since the satellite moves in a low polar orbit, a Doppler frequency drift occurs due to the relative motion of the spacecraft with respect to the transmitter. If the carrier frequency

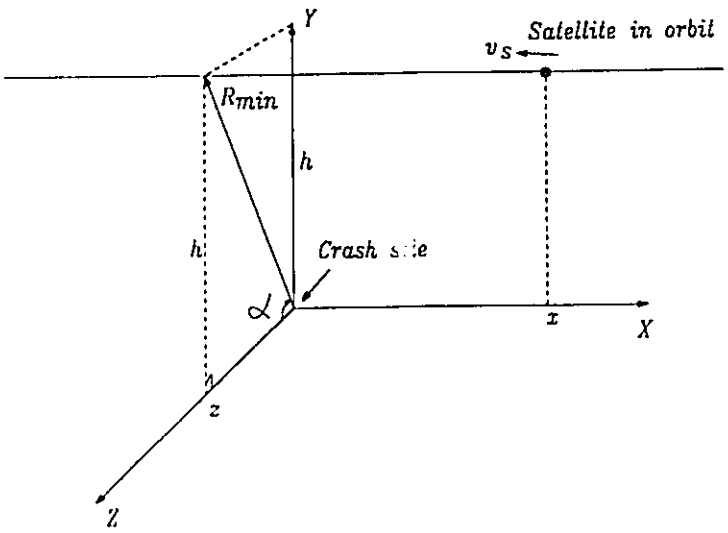


Figure 5.1: Flat-Earth Model of the SARSAT Geometry.

of ELT unit is f_0, the received frequency at the satellite will be (see Fig. 5.1) [17] [18]:

$$f_c = f_0 + f_d \left[\frac{x}{\sqrt{x^2 + h^2 + z_0^2}} \right], \tag{5.1}$$

where f_d = Doppler shift of the signal along the flight path;

x = distance along the flight path;

h = altitude of the satellite;

z_0 = displacement of unknown magnitude.

Hence, the accurate location depends on the accurate estimation of Doppler shift, or in other words, the received carrier frequency f_c.

The typical ELT signal model is illustrated in Fig.5.2. It comprises N pulse series, each with duration T_i, $0 \leq i \leq N-1$. The pulse "ON" duration is dT_i and the pulse "OFF" duration is $(1-d)T_i$, where d is the percentage proportion of pulse "ON" duration relative to the total pulse period T_i and is called the duty cycle.

Since both the satellite and the target move in relative lower speed in comparison with the speed of light, the received carrier frequency is almost unchanged in the observation

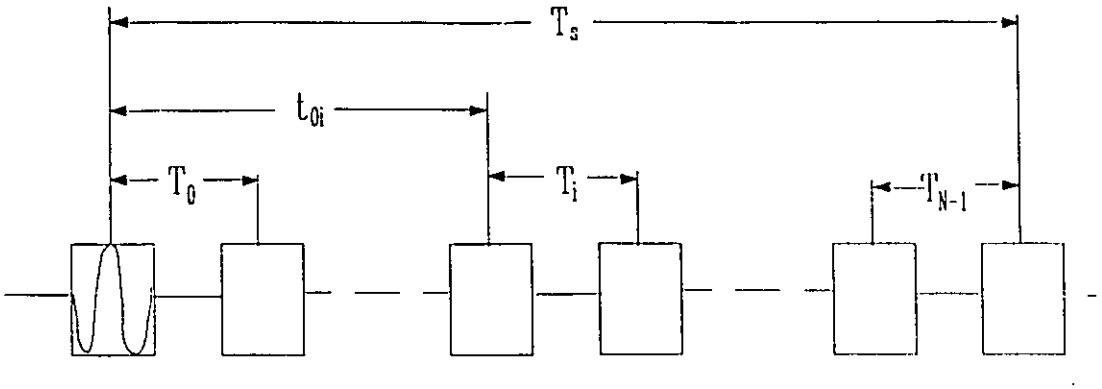


Figure 5.2: Illustration of Typical ELT Signal Model in the N Pulse Series.

period of about 1 second, consisting of N pulse series. Under this assumption, the mathematic model of ELT signal was developed by C.R.Carter and T. Chung[17], in which the received signal component is

$$s_r(t) = \sum_{i=0}^{N-1} g_i(t), \tag{5.2}$$

where $g_i(t)$ is the signal of the ith pulse. Thus

$$g_i(t) = \begin{cases} A_i(t)\cos(2\pi f_c t + \theta_i), & T_{0i} - \frac{dT_i}{2} \le t \le T_{0i} + \frac{dT_i}{2} \\ 0, & \text{otherwise.} \end{cases}, \tag{5.3}$$

where $i = 0, 1, 2, \cdots, N-1$, $A_i(t)$ is the envelope of the ith pulse; d is the duty cycle, T_i and θ_i are respectively the duration and the initial phase of the ith pulse, and T_{0i} is the centre of the ith pulse period.

The common specifications in the 121.5MHz SARSAT system are listed in Table.5.1

Table 5.1: **The Common Specifications in the 121.5MHz SARSAT System**

Carrier frequency	121.5 MHz
power output	approximately 100 mW
Modulation type	pulse
Duty cycle	between 33% to 55%
pulse repeat period	0.7ms to 1.6ms

Since each pulse "ON" period is non-overlapped with others, Eqs. (5.2) and (5.3) can be combined into a single formula as

$$s_r(t) = \begin{cases} A_i(t)\cos(2\pi f_c t + \theta_i), & T_{0i} - \frac{dT_i}{2} \le t \le T_{0i} + \frac{dT_i}{2} \\ 0, & \text{otherwise.} \end{cases} \qquad (5.4)$$

$$i = 0, 1, 2, \cdots N-1$$

If the signal is represented by an analytical signal $s(t)$

$$s(t) = \begin{cases} A_i(t)e^{j(2\pi f_c t + \theta_i)}, & T_{0i} - \frac{dT_i}{2} \le t \le T_{0i} + \frac{dT_i}{2} \\ 0, & \text{otherwise,} \end{cases} \qquad (5.5)$$

then $s_r(t)$ will be the real part of it. Defining the complex envelope as

$$\tilde{A}(t) \triangleq \begin{cases} A_i(t)e^{j\theta_i}, & T_{0i} - \frac{dT_i}{2} \le t \le T_{0i} + \frac{dT_i}{2} \\ 0, & \text{otherwise.} \end{cases} \qquad (5.6)$$

Eq. (5.5) can be rewritten as

$$s(t) = \tilde{A}(t)e^{j2\pi f_c t}. \qquad (5.7)$$

According to the phase characteristic ($\theta(t) = Im[\ln(s(t))]$), C.R.Carter and T.Chung[17] classify the signal into three different models:

1. Ideal Coherent Model: In this model, the phase is continuous from pulse to pulse, i.e., $\theta_i = \theta_0$ in Eq. (5.6) and

$$\theta(t) = 2\pi f_c t + \theta_0.$$

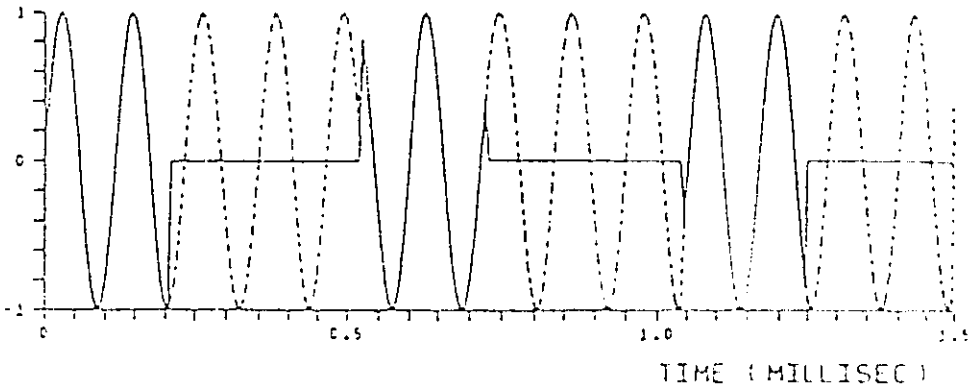


Figure 5.3: Non-Coherent ELT Signal with "ON" Time as a Solid Line and "OFF" Time a Dotted Line (Notice the phase discontinuity).

2. Non Ideal Coherent Model: In this model, in addition to the continuous phase, there also exists a phase shift ($\Delta\theta_i$) in the ith pulse period of the signal, which is due to the frequency variation between pulse "ON" and "OFF" period, i.e.,

$$\theta(t) = 2\pi f_c t + \Delta\theta_i + \theta_0$$

where $\Delta\theta_i = \sum_{k=1}^{i-1} 2\pi f_p(1-d)T_k$ and f_p is the difference between oscillator frequency while switch is "ON" and "OFF".

3. Non-Coherent Model: In this model, $\theta(t) = 2\pi f_c t + \theta_i$, where θ_i is a random variable, uniformly distributed between 0 and 2π, and independent with each other.

An example of non-coherent signal is shown in Fig.5.3.

It can be observed from the above discussions that the typical ELT signal consists of N pulse series, each with very short pulse duration. Hence, accurate frequency estimation becomes a difficult job, even with the ideal coherent model. Due to the narrow pulse

duration, the side lobes in the spectrum will give a false indication of signal. This becomes even more serious when SNR is low, not to mention that the real ELT signals are far from ideal. In general, the signals are quite unstable and there may exist amplitude modulation. As for the Non Ideal Coherent Model and Non-Coherent Model, the phase is not continuous from pulse to pulse. All the above facts will seriously influence the accuracy of frequency estimation using traditional methods, which are based on a long period of data record [19].

For the Non-Coherent Model, the initial phase of each pulse is totally independent of each other. If we take an observation period which is longer than two pulse periods, due to the phase discontinuity between the adjacent pulse, the Fourier spectrum of this observation period will not present an energy concentration at the carrier frequency f_c. It means that no strong sinusoidal component with carrier frequency f_c is presented in the observation period longer than two pulse periods. Therefore, the frequency estimation is inefficient with traditional spectrum estimate method[19] using a long period observation. If we take a short period of observation and estimate the frequency only based on one "ON" pulse period, the extremely short pulse period makes the estimate result very unsatisfactory. The low SNR under practical circumstances adds to the difficulty of accurately estimating the frequency using very short data record.

The mixed time-frequency representations yield a time-dependent spectrum, or in other words, the instantaneous spectrum is obtained. Therefore, an accurate instantaneous frequency estimation is expected to be reached by adopting time-frequency representations, which may get rid of all unstable or time variant factors, such as phase discontinuity and amplitude modulation.

5.2 Application of PWD in the SARSAT Signal Processing

5.2.1 Problem of Direct Use of PWD for the Frequency Estimation

After the appropriate preprocessing of the input ELT signal, we can obtain the analytical signal:

$$x(t) = s(t) + n(t), \tag{5.8}$$

where $s(t)$ is the complex analytical ELT signal as given in Eq. (5.5);

$n(t)$ is complex white Gaussian noise with power density N_0.

The PWD is defined as (see Eq.(2.40)):

$$\tilde{W}_x(t,f) = \int_{-\infty}^{\infty} h^2(\frac{\tau}{2}) x(t+\frac{\tau}{2}) x^*(t-\frac{\tau}{2}) e^{-j2\pi f\tau} d\tau. \tag{5.9}$$

Usually the "ON" pulse duration (dT_i) is short. In order that $\tilde{W}_x(t,f)$ in Eq. (5.9) can reflect the instantaneous frequency, the window $h^2(\frac{\tau}{2})$ should be very narrow in comparison with both dT_i and $(1-d)T_i$, so that in the "ON" pulse duration (t around T_{0i}), $A_i(t)$ is almost unchanged, i.e. $A_i(t) \approx A_i(T_{0i})$ and the signal component in $\tilde{W}_x(t,f)$ is

$$\begin{aligned} \tilde{W}_s(t,f) &= \int_{-\infty}^{\infty} h^2(\frac{\tau}{2}) s(t+\frac{\tau}{2}) s^*(t-\frac{\tau}{2}) e^{-j2\pi f\tau} d\tau \\ &\approx A_i^2(T_{0i}) \int_{-\infty}^{\infty} h^2(\frac{\tau}{2}) e^{-j2\pi(f-f_c)\tau} d\tau. \end{aligned} \tag{5.10}$$

While in the pulse "OFF" duration (t around $T_{0i} + \frac{T_i}{2}$, where $s(t) = 0$):

$$\tilde{W}_s(t,f) \approx 0. \tag{5.11}$$

Therefore, by locating the peak point of $\tilde{W}_x(t,f)$ at given time t in the pulse duration, the corresponding f is just the instantaneous frequency at that time. The detail analysis of this peak locating method will be similar to that described in chapter 4.

However, as just mentioned before, the window width ($2T_W$) is very small and from the discussions in the Chapter 4, the frequency bin-width Δf is inversely proportional to T_W, i.e. $\Delta f = \frac{1}{2T_W}$. For example, if $dT_i = 0.2$msec. and $T_W = .15$msec. $< dT_i$, then $\Delta f = 3.3KHz$. Since the maximum carrier frequency variation due to the Doppler shift is less than 10KHz, the frequency bin-width of 3.3KHz is too large for the accurate carrier frequency estimation and thus the accurate location of the incident site. Therefore, we must develop a more efficient way for the accurate estimation of frequency f_c. Also, due to the low SNR, the frequency estimation is not accurate enough with only one pulse period. Hence, our method should be able to eliminate the random initial phase so that the average of many pulse periods becomes possible.

5.2.2 The Properties of Averaged PWD

Due to the low SNR, it is often difficult to distinguish pulse "ON" and "OFF" duration. We take the average of the PWD at every possible time instant and define the averaged PWD over a period of M discrete points as:

$$\overline{W}_s(f) \triangleq \frac{1}{M}\sum_{i=1}^{M}\int_{-\infty}^{\infty} h^2(\frac{\tau}{2})s(t_i+\frac{\tau}{2})s^*(t_i-\frac{\tau}{2})e^{-j2\pi f\tau}d\tau, \tag{5.12}$$

where $t_i = t_0 + (i-1)\Delta t$, t_0 is the initial time and Δt is the discrete time interval suitably selected for the summation. This will be discussed in further details when we perform the testing with real ELT signals. We also define

$$\begin{aligned}\overline{E}_s(\tau) &\triangleq \frac{1}{M}\sum_{i=1}^{M} h^2(\frac{\tau}{2})s(t_i+\frac{\tau}{2})s^*(t_i-\frac{\tau}{2}) \\ &= \frac{1}{M}\sum_{i=1}^{M} h^2(\frac{\tau}{2})\tilde{A}(t_i+\frac{\tau}{2})\tilde{A}^*(t_i-\frac{\tau}{2})e^{j2\pi f_c\tau},\end{aligned} \tag{5.13}$$

where Eq. (5.5) is used in the second equality, then, $\overline{W}_s(f)$ is the Fourier transform of $\overline{E}_s(\tau)$, i.e.

$$\overline{W}_s(f) = \int_{-\infty}^{\infty}\overline{E}_s(\tau)e^{-j2\pi f\tau}d\tau. \tag{5.14}$$

If $\overline{E}_s(\tau)$ has a strong sinusoidal component on frequency f_c, we may use the spectrum estimation methods to estimate this frequency. Next, we will discussed the properties of $\overline{E}_s(\tau)$ and prove that it really has strong sinusoidal component of frequency f_c. Finally, suitable methods will be applied to extract this frequency component.

With $s(t)$ being the analytical signal given by Eq. (5.5), $\overline{E}_s(\tau)$ will possess the following properties:

1. $\overline{E}_s(\tau)$ eliminate random phase, decrease noise power, and retain the sinusoidal component

It is easy to see that the average by Eq. (5.13) can certainly reduce the uncorrelated noise components. Here, we just explain in detail how the operation to obtain $\overline{E}_s(\tau)$ from $s(t)$ can eliminate the random phase θ_i and retain the sinusoidal component.

Notice the definition of $\tilde{A}(t)$ in Eq. (5.6). If T_W is small enough, i.e.,

$$T_W < \min_i(dT_i, (1-d)T_i), \tag{5.15}$$

then in Eq. (5.13), $\tilde{A}(t_i + \frac{\tau}{2})$ and $\tilde{A}(t_i - \frac{\tau}{2})$ have the same phase θ_i in the given window period of $h^2(\frac{\tau}{2})$ and this phase θ_i will be cancelled by the multiplication of $\tilde{A}(t_i + \frac{\tau}{2})$ and $\tilde{A}^*(t_i - \frac{\tau}{2})$. Hence, Eq. (5.13) becomes

$$\overline{E}_s(\tau) = \frac{1}{M}\sum_{i=1}^{M} h^2(\frac{\tau}{2})|\tilde{A}(t_i + \frac{\tau}{2})||\tilde{A}^*(t_i - \frac{\tau}{2})|e^{j2\pi f_c \tau}. \tag{5.16}$$

Assuming an ideal pulse such that $A_i(t) = 1$ and $T_i = T$ in Eq. (5.5) for all i, then $|\tilde{A}(t)|$ is as shown in Fig.5.4. The amplitude of $\overline{E}_s(\tau)$ will be

$$|\overline{E}_s(\tau)| = \frac{1}{M}\sum_{i=1}^{M} h^2(\frac{\tau}{2})|\tilde{A}(t_i + \frac{\tau}{2})||\tilde{A}^*(t_i - \frac{\tau}{2})|. \tag{5.17}$$

If a rectangular window is adopted in the above equation, i.e.,

$$h(\frac{\tau}{2}) = \begin{cases} 1, & |\tau| \le T_W \\ 0, & \text{otherwise} \end{cases}, \tag{5.18}$$

If $\overline{E}_s(\tau)$ has a strong sinusoidal component on frequency f_c, we may use the spectrum estimation methods to estimate this frequency. Next, we will discussed the properties of $\overline{E}_s(\tau)$ and prove that it really has strong sinusoidal component of frequency f_c. Finally, suitable methods will be applied to extract this frequency component.

With $s(t)$ being the analytical signal given by Eq. (5.5), $\overline{E}_s(\tau)$ will possess the following properties:

1. $\overline{E}_s(\tau)$ eliminate random phase, decrease noise power, and retain the sinusoidal component

It is easy to see that the average by Eq. (5.13) can certainly reduce the uncorrelated noise components. Here, we just explain in detail how the operation to obtain $\overline{E}_s(\tau)$ from $s(t)$ can eliminate the random phase θ_i and retain the sinusoidal component.

Notice the definition of $\tilde{A}(t)$ in Eq. (5.6). If T_W is small enough, i.e.,

$$T_W < \min_i(dT_i, (1-d)T_i), \tag{5.15}$$

then in Eq. (5.13), $\tilde{A}(t_i + \frac{\tau}{2})$ and $\tilde{A}(t_i - \frac{\tau}{2})$ have the same phase θ_i in the given window period of $h^2(\frac{\tau}{2})$ and this phase θ_i will be cancelled by the multiplication of $\tilde{A}(t_i + \frac{\tau}{2})$ and $\tilde{A}^*(t_i - \frac{\tau}{2})$. Hence, Eq. (5.13) becomes

$$\overline{E}_s(\tau) = \frac{1}{M}\sum_{i=1}^{M} h^2(\frac{\tau}{2})|\tilde{A}(t_i + \frac{\tau}{2})||\tilde{A}^*(t_i - \frac{\tau}{2})|e^{j2\pi f_c \tau}. \tag{5.16}$$

Assuming an ideal pulse such that $A_i(t) = 1$ and $T_i = T$ in Eq. (5.5) for all i, then $|\tilde{A}(t)|$ is as shown in Fig.5.4. The amplitude of $\overline{E}_s(\tau)$ will be

$$|\overline{E}_s(\tau)| = \frac{1}{M}\sum_{i=1}^{M} h^2(\frac{\tau}{2})|\tilde{A}(t_i + \frac{\tau}{2})||\tilde{A}^*(t_i - \frac{\tau}{2})|. \tag{5.17}$$

If a rectangular window is adopted in the above equation, i.e.,

$$h(\frac{\tau}{2}) = \begin{cases} 1, & |\tau| \le T_W \\ 0, & \text{otherwise} \end{cases}, \tag{5.18}$$

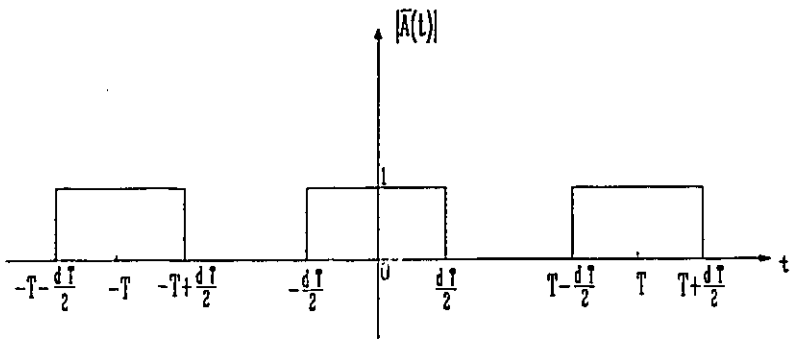


Figure 5.4: The envelop of ELT signal.

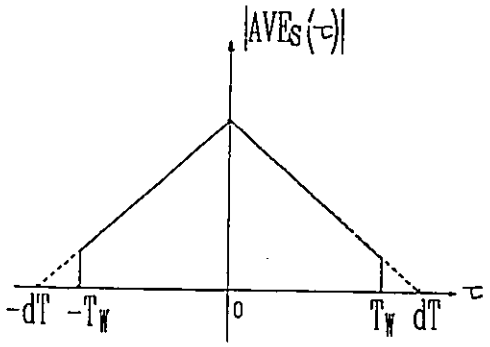


Figure 5.5: The envelop of $\overline{E}_s(\tau)$.

and if T_W is chosen such that Eq. (5.15) is satisfied, then $|\overline{E}_s(\tau)|$ is as shown in Fig.5.5.

From Eq. (5.16), it can be seen that the random phase has been eliminated by a narrow window $h^2(\frac{\tau}{2})$ and $\overline{E}_s(\tau)$ contains a sinusoidal component $e^{j2\pi f_c t}$ modulated by $|\overline{E}_s(\tau)|$, which is given in Eq. (5.17) and shown in Fig.5.5 for a rectangular window.

2. $\overline{E}_s(\tau)$ eliminate the cross interference in PWD

If the received signal $s(t)$ consists of two ELT signals generated by two different sources such that

$$s(t) = s_1(t) + s_2(t), \tag{5.19}$$

then the PWD will be [13]:

$$W_s(t,f) = W_{s1}(t,f) + W_{s2}(t,f) + 2Re[W_{s_1 s_2}(t,f)], \tag{5.20}$$

where the third component $W_{s_1 s_2}(t,f)$ is the cross component between $s_1(t)$ and $s_2(t)$. The existence of $W_{s_1 s_2}(t,f)$ will yield a "false" spectrum components between the spectrum of $s_1(t)$ and $s_2(t)$, and both the resolution and the frequency estimate accuracy will deteriorate because of it (see the discussions in section 4.2). Now, we will prove that with the averaging in Eq. (5.12), the cross-components will approach zero.

Suppose $s(t)$ has two components as in Eq. (5.19) such that

$$s_1(t) = \tilde{A}_1(t)e^{j2\pi f_{c1}t}, \tag{5.21}$$

$$s_2(t) = \tilde{A}_2(t)e^{j2\pi f_{c2}t}, \tag{5.22}$$

where f_{c1} and f_{c2} are the received carrier frequencies of $s_1(t)$ and $s_2(t)$ respectively, $\tilde{A}_1(t)$ and $\tilde{A}_2(t)$ are the corresponding complex envelopes and have the same form as Eq. (5.6), having initial phase θ_{1i} and θ_{2i}, respectively, in their own pulse "ON" duration. Now, we have

$$\overline{E}_s(\tau) = \overline{E}_{s_1}(\tau) + \overline{E}_{s_2}(\tau) + 2Re[\overline{E}_{s_1 s_2}(\tau)] \tag{5.23}$$

where

$$\begin{aligned}
\overline{E}_{s_1}(\tau) &= \frac{1}{M}\{\sum_{i=1}^{M} h^2(\frac{\tau}{2})s_1(t_i+\frac{\tau}{2})s_1^*(t_i-\frac{\tau}{2})\} \\
&= |\overline{E}_{s_1}(\tau)|e^{j2\pi f_{c1}\tau} \qquad (5.24)\\
\overline{E}_{s_2}(\tau) &= \frac{1}{M}\{\sum_{i=1}^{M} h^2(\frac{\tau}{2})s_2(t_i+\frac{\tau}{2})s_2^*(t_i-\frac{\tau}{2})\} \\
&= |\overline{E}_{s_2}(\tau)|e^{j2\pi f_{c2}\tau} \qquad (5.25)\\
\overline{E}_{s_1 s_2}(\tau) &= \frac{1}{M}\{\sum_{i=1}^{M} h^2(\frac{\tau}{2})s_1(t_i+\frac{\tau}{2})s_2^*(t_i-\frac{\tau}{2})\} \\
&= \frac{1}{M}e^{j\pi(f_{c1}+f_{c2})\tau}\sum_{i=1}^{M} h^2(\frac{\tau}{2})[|\tilde{A}_1(t_i-\frac{\tau}{2})||\tilde{A}_2(t_i+\frac{\tau}{2})|e^{j2\pi(f_{c1}-f_{c2})t_i}e^{j(\theta_{1i}-\theta_{2i})}]. \qquad (5.26)
\end{aligned}$$

For the ideal pulse, $|\tilde{A}_1(t)|$ and $|\tilde{A}_2(t)|$ have similar shapes as that given in Fig.5.4. If both $s_1(t)$ and $s_2(t)$ are of non-coherent model, and θ_{1i} and θ_{2i} are independent, then

the average of $e^{j(\theta_{1i}-\theta_{2i})}$ will approach zero for large M. Hence $\overline{E}_{s_1 s_2}(\tau) = 0$. If both $s_1(t)$ and $s_2(t)$ are of coherent model, then we have $\theta_{1i} - \theta_{2i} = \theta_{00}$(constant). However, since $f_{c1} \neq f_{c2}$, averaging $e^{j2\pi(f_{c1}-f_{c2})t_i}$ over t_i will also approach zero for large M. Therefore, regardless of whether the two signals are of coherent or non-coherent models, the average given in Eq. (5.12) can eliminate the cross component interference in PWD. Hence, Eq. (5.23) becomes:

$$\overline{E}_s(\tau) = \overline{E}_{s1}(\tau) + \overline{E}_{s2}(\tau), \tag{5.27}$$

which contains only the frequency components f_{c1} and f_{c2}.

Although $\overline{E}_s(\tau)$ contains the frequency information, the direct use of Fourier transform onto it as in Eq. (5.14) will not give a good estimation of the frequencies due to the extremely narrow window width of $h^2(\frac{\tau}{2})$ which should satisfy Eq. (5.15). In the following part, we will present more accurate ways for the frequency estimation. In the section 5.2.3, we use the property of the WD to estimate the frequency under the assumption that only one ELT signal is received. In most situations, this assumption is valid since an ELT signal is transmitted only when an emergency happens to the aircraft and it is unlikely that more than one emergency would occur at the same time. The less likely situation when more than one ELT signals occur will be discussed in section 5.3.

5.2.3 Frequency Estimation Using the instantaneous frequency property of PWD

In section 2.4, we have seen that the centre of gravity of the WD (or PWD) at a fix time t is equal to the instantaneous frequency at the same time. This property can be equally applied to the average PWD as well, i.e,

Property 1:

$$f_c = \frac{\int_{-\infty}^{\infty} f\overline{W}_s(f)df}{\int_{-\infty}^{\infty} \overline{W}_s(f)df}. \tag{5.28}$$

Proof: If $h(t)$ is real and even, it is straightforward from Eq. (5.17) that $|\overline{E}_s(\tau)|$ is even and real. Also from Eq. (5.16), $\overline{E}_s(\tau)$ is the amplitude modulation of carrier frequency f_c with envelope $|\overline{E}_s(\tau)|$ and from Eq. (5.14), $\overline{W}_s(f)$ is the Fourier Transform of $\overline{E}_s(\tau)$. Therefore, $\overline{W}_s(f)$ is real and symmetrical with centre at f_c. Hence the centre of gravity of $\overline{W}_s(f)$ is the required frequency f_c. □

Eq. (5.28) gives an accurate evaluation of carrier frequency of $s(t)$. However, our received signal $x(t)$, as given in Eq. (5.8), is contaminated by additive Gaussian white noise. $\overline{W}_s(f)$ should be, therefore, replaced by $\overline{W}_x(f)$ if Eq.(5.28) is used to evaluate f_c. In general, $\overline{W}_s(f)$ is essentially band-limited in $[f_c - B_s, f_c + B_s]$, where $2B_s$ is the bandwidth of $\overline{W}_s(f)$, i.e. $\overline{W}_s(f) \approx 0$ for $|f - f_c| \geq B_s$. The integration of frequency from $-\infty$ to ∞ in Eq. (5.28) becomes unnecessary. Hence, in practice, Eq. (5.28) can be modified as

$$\hat{f}_c \approx \frac{\int_{f_p - B_s}^{f_p + B_s} f\overline{W}_x(f)df}{\int_{f_p - B_s}^{f_p + B_s} \overline{W}_x(f)df}. \tag{5.29}$$

Here, in the right side of Eq. (5.29), we encounter two unknown parameters f_p and B_s. The centre frequency f_p can be evaluated by a pre-estimated value which is the peak position of $\overline{W}_x(f)$, denoted by f_p. The suitable B_s should be selected. If it is too small, a large amount of signal energy will be cut off. This will yield a large bias in the estimation of $\hat{f}_c$, especially when f_p is not equal to f_c. On the other hand, if B_s is too large, more noise power will be added in, so that the estimation variance will increase. Therefore the suitable integration bound B_s in Eq. (5.29) will depend on two facts: the bandwidth of $\overline{W}_s(f)$ and signal to noise ratio(SNR).

Observing Eq. (5.12) and using the assumption that the window width of $h^2(\frac{\tau}{2})$ is so narrow that $s(t_i + \frac{\tau}{2})s^*(t_i - \frac{\tau}{2})$ has almost unchanged amplitude in the short window period, the bandwidth of $\overline{W}_s(f)$ will merely depend on the window width of $h^2(\frac{\tau}{2})$. Given a window function, there are many ways of defining its window width. Here,

we assume that when the level $h^2(\frac{\tau}{2})$ is lower than $0.1 \times h^2(0)$, the contribution of $h^2(\frac{\tau}{2})s(t_i + \frac{\tau}{2})s^*(t_i - \frac{\tau}{2})$ to the $\overline{W}_s(f)$ is negligible. Hence, we define the equivalent window width T_E as the instant when $h^2(\frac{\tau}{2}) = 0.1h^2(0)$, i.e.,

$$T_E \triangleq \arg_\tau \left\{ \frac{h^2(\frac{\tau}{2})}{h^2(0)} = 0.1 \right\}. \tag{5.30}$$

With the window width defined above, we then subjectively choose the B_s as

$$B_s = \frac{1}{T_E}. \tag{5.31}$$

Eq. (5.30) gives the definition of the window width T_E. Next we will discuss how to assign a suitable value for it and finally to obtain B_s using Eq. (5.31).

From Table 5.1. d=33% to 55%, T_i=0.7 to 1.6 ms., hence the minimum value of d and T_i are $d_{min} = .33$ and $T_{imin} = 0.7ms$. Also considering the facts that the pulse is not ideal, i.e. it is distorted by the amplitude variation, pulse rising time and pulse dropping time, we finally select T_E such that

$$T_E = 0.15ms < (dT_i)_{min} = 0.23ms, \tag{5.32}$$

where $(dT_i)_{min}$ represents the minimum value of dT_i. It could be expected that the signal is quite stable in this short duration. Now, the suitable B_s is selected as,

$$B_s = \frac{1}{T_E} = 6.67KHz \tag{5.33}$$

Finally, we come to choose our window function. A lot of existing windows are available [70], among which, the Gaussian window has the property that the product of rms bandwidth and rms duration has the minimum possible value [36]. Hence, the Gaussian window is adopted here:

$$h^2(\frac{\tau}{2}) = e^{-\frac{\tau^2}{4\sigma_g^2}}, \tag{5.34}$$

where σ_g is the parameter which control the window width. With T_E given by Eq. (5.32) and its definition in Eq. (5.30), we get the conservative selection of σ_g as

$$\sigma_g = 0.05ms. \tag{5.35}$$

The application of the method discussed in this section to the frequency estimation of a real ELT signal will be given next.

5.2.4 Real ELT Signal Test

Four groups of real ELT signals are used here for the frequency estimation: they are Pointer ELT (PNTELT01), Garrett ELT (GARELT12), and two Narco ELT (NARELT17 and NARELT19), while PNTELT01 and GARELT12 are of the coherent model, NARELT17 and NARELT19 are of the Non-ideal coherent model (see the discussions in [18]). Figs 5.6 (a) to (d) show, respectively, the four groups of real ELT signals in time domain for a certain period of observation. It can be observed that the noise components are quite high and the pulse shape is not the same as the ideal pulse shown in Fig.5.4. Our estimation procedures are the same for all four groups of signals and can be briefly summarized as follows:

1. Collecting the data: the input signals are first translated to the intermediate frequency band by mixing with a local oscillate signal in the receiver. The output intermediate frequency signals are then sampled with sampling frequency 102KHz, i.e., sampling period $\Delta t_s = \frac{1}{102KHz}$, and stored on the disk. The collected data for each group of ELT signals are divided into 25 blocks, each block containing 4K data. Hence, the total data available is $25 \times 4K = 100K$ data points. The total observation period is $100K\Delta t_s \approx 1\ sec.$ With the assumption made in Section 5.1, the received carrier frequency is essentially unchanged in this period of observation.

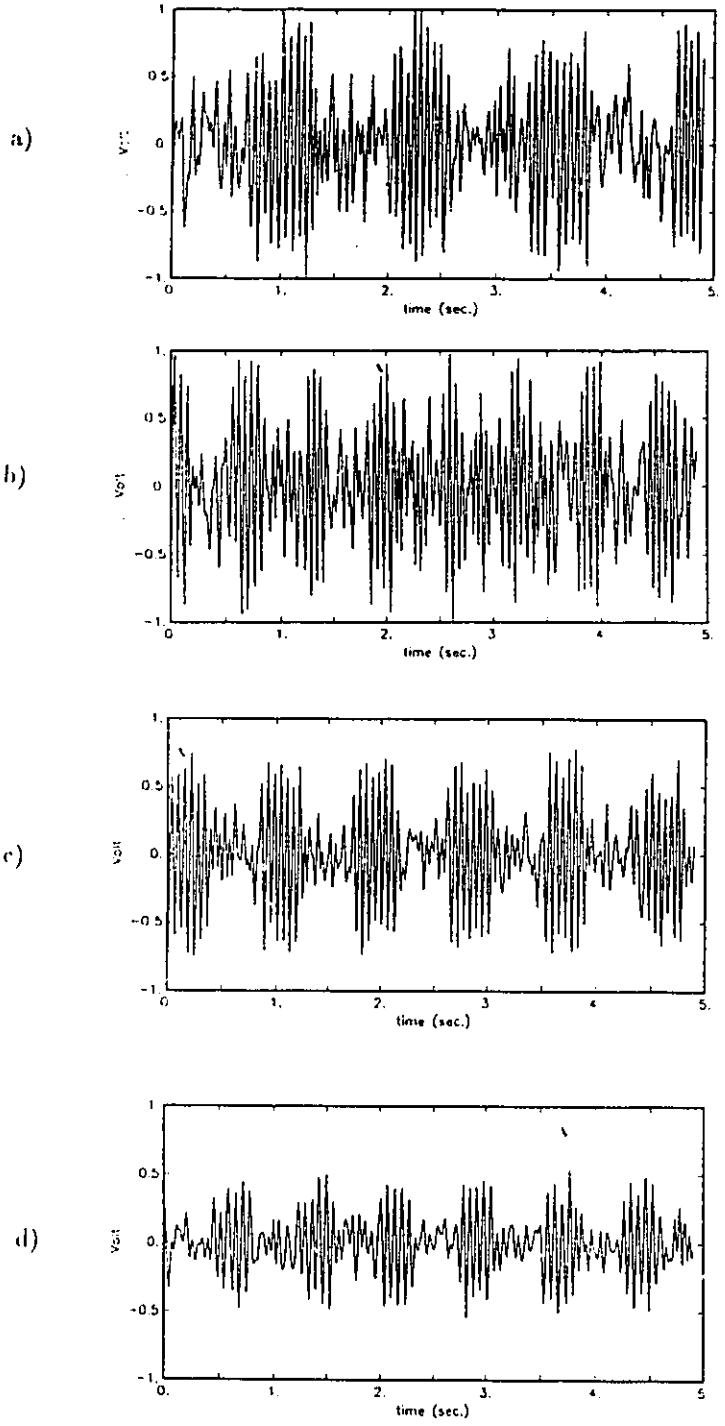


Figure 5.6: Real ELT Signals in Time Domain (a)PNTELT01, (b)GARELT12, (c)NARELT17 and (d)NARELT19.

2. The Gaussian window in Eq. (5.34) with σ_g given by Eq. (5.35) is used for $h^2(\frac{\tau}{2})$ in Eq. (5.13). Since the window width is very narrow and $\overline{E}_x(\tau)$ will vanish to zero for $|\tau| \gg T_E$. Hence, it is unnecessary to calculate $\overline{E}_x(\tau)$ for all τ. Here, we choose 128 sampling points on $\overline{E}_s(\tau)$, i.e. $-64\Delta t_s < \tau \leq 64\Delta t_s$, where the sampling period of $\overline{E}_s(\tau)$ takes the same value as the data sampling period Δt_s.

3. For each block of data, we calculate $h^2(\frac{\tau}{2})s(t_i+\frac{\tau}{2})s^*(t_i-\frac{\tau}{2})$ and obtain the discrete $\overline{E}(\tau)$ using Eq. (5.13) with averaging on t_i, where $t_i = t_0+(i-1)\Delta t$, $i = 1,2,\cdots,2\times 4096$, t_0 is the initial time and $\Delta t = \frac{\Delta t_s}{2}$. The reasons that Δt can be chosen as half of the sampling interval are explained as follows: When i is an odd number, let $k = \frac{i-1}{2}$, then $t_i = t_0 + k\Delta t_s$. Since $s(t)$ is sampled with sampling interval Δt_s, $h^2(\frac{\tau}{2})s(t_i+\frac{\tau}{2})s^*(t_i-\frac{\tau}{2})$ yields values for $\tau = \cdots,-4\Delta t_s,-2\Delta t_s,0,2\Delta t_s,4\Delta t_s,\cdots$, i.e., contributes to the even sampling points of $\overline{E}_s(\tau)$ if $\overline{E}_s(\tau)$ is also sampled with the sampling interval Δt_s. When i is an even number, let $k_1 = \frac{i}{2}$, then $t_i = t_0 + k_1\Delta t_s - \frac{t_s}{2}$. For the same reason, $h^2(\frac{\tau}{2})s(t_i+\frac{\tau}{2})s^*(t_i-\frac{\tau}{2})$ yields values for $\tau = \cdots,-3\Delta t_s,-\Delta t_s,\Delta t_s,3\Delta t_s,\cdots$, i.e., contributes to the odd sampling points of $\overline{E}_s(\tau)$. Combining the odd and even sampling points of $\overline{E}_s(\tau)$ and averaging for $i = 1,2,\cdots,2\times 4096$, we obtain the discrete $\overline{E}_s(\tau)$ whose sampling interval is Δt_s, half of what can be obtained with averaging either even i or odd i. If we want to decrease the noise components in $\overline{E}_x(\tau)$ further more, we must get rid of those $h^2(\frac{\tau}{2})s(t_i+\frac{\tau}{2})s^*(t_i-\frac{\tau}{2})$ in the average, where t_i is in the pulse "OFF" duration. These $h^2(\frac{\tau}{2})s(t_i+\frac{\tau}{2})s^*(t_i-\frac{\tau}{2})$ contribute only noise components to the $\overline{E}_x(\tau)$. However, it was mentioned before that it is difficult to distinguish pulse "ON" and "OFF" duration under low SNR. Here, we intend to eliminate those $h^2(\frac{\tau}{2})s(t_i+\frac{\tau}{2})s^*(t_i-\frac{\tau}{2})$ from the average, which has smaller energy and hence is very possible to be in the pulse "OFF" duration.

Considering that the minimum duty cycle is $d_{min} = 33\%$, we choose only 33% of $h^2(\frac{\tau}{2})s(t_i + \frac{\tau}{2})s^*(t_i - \frac{\tau}{2})$ in the average. We first calculate the energy of $h^2(\frac{\tau}{2})s(t_i + \frac{\tau}{2})s^*(t_i - \frac{\tau}{2})$, i.e, $\int |h^2(\frac{\tau}{2})s(t_i + \frac{\tau}{2})s^*(t_i - \frac{\tau}{2})|^2 d\tau$, for given t_i. We then rank the energy for different t_i in the order of decreasing values and take only those $h^2(\frac{\tau}{2})s(t_i + \frac{\tau}{2})s^*(t_i - \frac{\tau}{2})$ into the average, whose energy is ranked at the top 33%. In this way, we decrease the number of average by 67% .

Now, we obtain $E_x(\tau)$ which is the average of $h^2(\frac{\tau}{2})s(t_i + \frac{\tau}{2})s^*(t_i - \frac{\tau}{2})$ for some t_i. The number we take into the average for each block of data is $2 \times 4096 \times 0.33 \approx 2693$.

4. Apply Fourier transform to $\overline{E}_x(\tau)$ as in Eq. (5.14). Now $\overline{E}_x(\tau)$ has only 128 points, so that a 128 point discrete Fourier transform [73] is applied to obtain $\overline{W}_x(f)$.

5. Eq. (5.29) is applied to the $\overline{W}_x(f)$ to get the estimated value of the carrier frequency $\hat{f}_c$, where in the right side of Eq. (5.29), the frequency f_p, is assigned the value of the maximum point of $|\overline{W}_x(f)|$, and the value of B_s is given by Eq. (5.33).

6. Repeat step 2 and 3 for another block of data until all 25 blocks are used.

Using the above estimate procedures to all four groups of signals presented in Fig.5.6, we obtain the estimation results which are listed in Tables 5.2, where the unit of the estimate frequency is Hz. The "block estimation" means the frequency estimation based on the ith block and the block number i are given at the most left column of the table. Similarly, in Table 5.3, we take more samples in the average. The "average of blocks" means that the frequencies be estimated using $\overline{E}_x(\tau)$, which is calculated based on the connection of previous i blocks of data, i.e., we have $i \times 4096$, instead of 4096, sampling points of data .

Since the true frequency is unknown in our real ELT signals, we can only observe its frequency estimation stability. From table 5.2, we can obtain the the average frequency,

Table 5.2: **The Estimated Frequency For Real ELT Signal: Block Estimation**

NO. OF BL.	NARELT19	NARELT17	GARELT12	PNTELT01
1	15057.10	16823.35	17157.94	15797.02
2	15072.88	16864.77	17108.31	15799.84
3	15002.69	16818.62	17109.30	15823.29
4	14971.54	16748.74	17160.48	15725.73
5	15022.97	16803.36	17226.94	15754.37
6	14990.93	16781.63	17142.43	15759.52
7	14973.11	16784.80	17058.33	15692.42
8	14981.27	16797.31	17097.29	15778.86
9	14972.23	16809.81	16983.87	15823.56
10	14961.96	16767.31	17006.60	15816.16
11	15049.60	16815.75	17107.32	15789.60
12	15022.87	16864.87	17110.48	15702.10
13	15030.12	16812.33	17196.30	15763.80
14	15033.19	16767.68	17146.32	15757.91
15	15004.85	16851.37	17118.17	15733.88
16	15019.45	16936.62	17206.95	15824.22
17	15045.64	16777.62	17075.78	15790.10
18	14946.59	16794.35	17118.74	15823.99
19	14943.19	16862.72	17121.92	15762.83
20	15011.87	16731.94	17128.06	15780.21
21	14985.41	16824.71	17147.09	15750.17
22	15123.70	16815.43	17104.82	15730.02
23	15121.98	16806.95	17072.07	15799.59
24	15012.23	16786.78	17054.03	15693.10
25	15018.36	16777.46	17081.56	15805.45

Table 5.3: The Estimated Frequency For Real ELT Signal: Average of Blocks

Ave.of Bl.	NARELT19	NARELT17	GARELT12	PNTELT01
1	15057.10	16823.35	17157.94	15797.02
2	15092.01	16840.50	17086.45	15798.39
3	15072.02	16829.47	17080.78	15806.72
4	15053.56	16807.41	17093.14	15786.62
5	15054.24	16806.19	17114.19	15780.42
6	15047.38	16801.97	17112.78	15776.89
7	15040.66	16799.13	17098.84	15765.19
8	15036.02	16798.44	17094.41	15766.74
9	15031.18	16798.52	17077.06	15772.76
10	15026.61	16795.28	17065.46	15777.10
11	15031.24	16796.93	17065.68	15778.22
12	15034.02	16801.88	17066.19	15771.61
13	15035.90	16802.36	17073.82	15771.04
14	15037.57	16799.18	17075.72	15770.13
15	15037.38	16802.38	17075.98	15767.87
16	15038.01	16824.67	17081.20	15771.06
17	15039.57	16821.46	17078.11	15772.09
18	15036.09	16819.54	17078.55	15774.71
19	15032.77	16821.17	17078.57	15774.12
20	15032.75	16816.34	17079.27	15774.42
21	15031.68	16816.42	17080.44	15773.32
22	15036.66	16816.10	17079.10	15771.37
23	15040.98	16815.59	17076.94	15772.54
24	15041.06	16814.30	17073.92	15769.28
25	15041.40	16812.54	17072.56	15770.71

Table 5.4: **The Maximum Frequency Variation Using the Block Estimation**

	Maximum Freq. Variation	Average Freq.	Maximum Relative Freq. Variation
NARELT19	±108.67Hz	15015.03Hz	.72%
NARELT17	±127.60Hz	16809.05Hz	.76%
GARELT12	±129.77Hz	17113.64Hz	.76%
PNTELT01	±78.01Hz	15771.11Hz	.49%

maximum frequency variation and the maximum relative frequency variation, defined as the ratio of the maximum frequency variation to the average frequency, for different groups of data. The results are listed in Table 5.4. It can be seen that the relative frequency variation is smaller than .8%. If we take even large number of samples, the SNR will increase and more stable results can be obtained. Examination of the last ten row of table 5.3 shows that the frequency variation is within a few Hz. Therefore, it can be concluded that using our method, a very stable frequency estimation can be reached within about 1 second record of data.

5.3 High Resolution Method for More than One ELT Signals

5.3.1 Frequency Estimation combing Burg algorithm and Average PWD

In the previous discussion, we assumed that only one ELT signal is received. Although this assumption is true for most of the cases, in some practical situations, it is still possible that the earth station may receive more than one ELT signal at the same time. In order to locate these incident sites, it is necessary that the earth station can separate these signals.

From the previous analysis, it is known that in order to make the sinusoidal signal stable in the given window duration, the window width T_E should be very narrow. After considering the factors of: minimum pulse width, and the rising and dropping time of pulse, T_E was selected to be 0.15 ms in Eq. (5.32). This window is too narrow if a high resolution is desired. Considering the averaged PWD, it could be seen that the bandwidth of spectrum is about $2B_s = 13.33KHz$. If two signals are closer together in frequency than this bandwidth, it will be impossible to distinguish them from each other. Hence the formula for the calculation of instantaneous frequency can no longer be used. In this section, we use the property of $\overline{E}_x(\tau)$, that it contains the desired frequency components modulated by their respective envelope, we try to find a suitable high resolution method and apply it on $\overline{E}_x(\tau)$, so that these frequency components can be extracted and the separation of two or more close signals can be realized with even very narrow window width. Here, $\overline{E}_x(\tau)$ is calculated using the same procedures stated in Section 5.2.4. However, since for two and more ELT signals, it is impossible to distinguish the respective pulse "ON" and "OFF" duration, therefore, we estimate $\overline{E}_x(\tau)$ by taking the average of $h^2(\frac{\tau}{2})s(t_i + \frac{\tau}{2})s^*(t_i - \frac{\tau}{2})$ at every discrete instant $t_i = t_0 + i\frac{\Delta t_s}{2}$, where Δt_s is the sampling frequency (see section 5.2.4).

The Burg algorithm [16] is a high resolution, all pole model algorithm, which is robust for the amplitude variations and hence is suitable for extracting the implicit periodic components regardless of the amplitude. The detailed procedures of the algorithm are presented in Appendix F and the basic idea is also explained briefly there. In this section, we will first apply the Burg algorithm on the $\overline{E}_x(\tau)$ of real ELT signals to show that this method really works on ELT signals. Then we will use the simulated ELT signals to show that the application of the Burg algorithm on $\overline{E}_x(\tau)$ can provide much higher resolution than what we could achieve before.

5.3.2 Estimation Results with Real and Simulated ELT Signals

For each of the four ELT signals as given in section 5.2.4, we adopt the rectangular window of Eq. (5.18) instead of Gaussian window for less amplitude distortion. The window width is chosen to be $2T_W = 0.28$ms which is about the same as $2T_E$ in Eq. (5.32) and this window contains 29 sampling points. $\overline{E}_x(\tau)$ is calculated by taking the average over 25 blocks of available data. Now, we have an $N_1 = 29$ point time series $\overline{E}_x(\tau)$ for $-14\Delta t_s \leq \tau \leq 14\Delta t_s$ where Δt_s is the sampling interval. We apply the Burg algorithm described in Appendix F to the discrete time series $\overline{E}_x(\tau)$. For the given four groups of ELT signals, the corresponding power spectrum are shown in Figs.5.7 to 5.10 with the order of AR (Auto Regression) model being $M_1 = 5$ (the suitable order selection will be discussed later) and the peaks are located respectively at following frequencies:

NARELR19	15.12KHz
NARELT17	16.84KHz
GARELT12	17.04KHz
PNTELT01	15.74KHz

These frequencies represent the estimated carrier frequencies for different ELT signals. Comparing with Table 5.3. it could be found that the estimated frequencies using Burg algorithm are quite close to what we obtained using the instantaneous frequency property of the PWD. Therefore, we can confidently say that both methods work very well on real ELT signals.

Now, we come to test the resolution of this new method. Since no real data are available for two ELT signals, we use simulated data instead. Fig.5.11 shows two simulated ELT signals in time domain: one has a carrier frequency $f_{c1} = 14KHz$ and another has $f_{c2} = 16KHz$. Both signals have duty cycle $d = .5$ and the pulse repeat period .7 msec. These are typical ELT signal specifications. The SNR (here SNR is defined as average

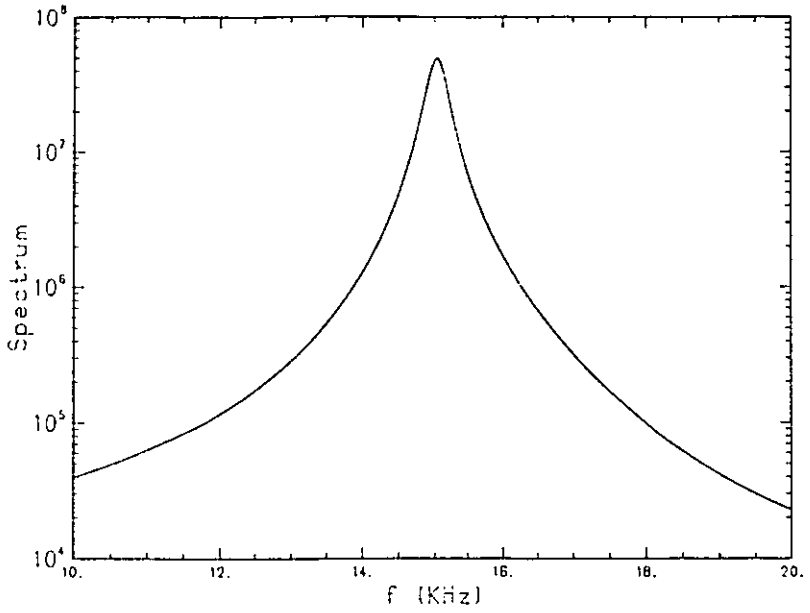


Figure 5.7: High Resolution ELT Frequency Estimation for NARELT19.

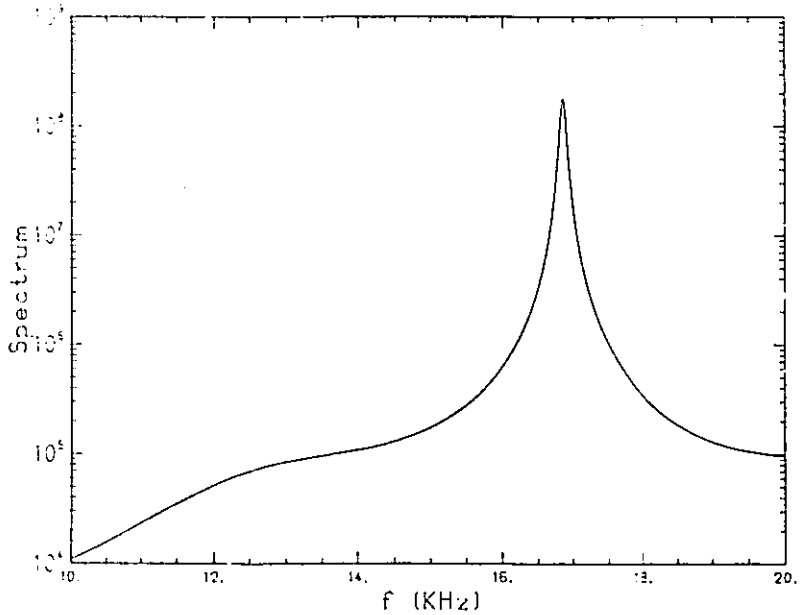


Figure 5.8: High Resolution ELT Frequency Estimation for NARELT17.

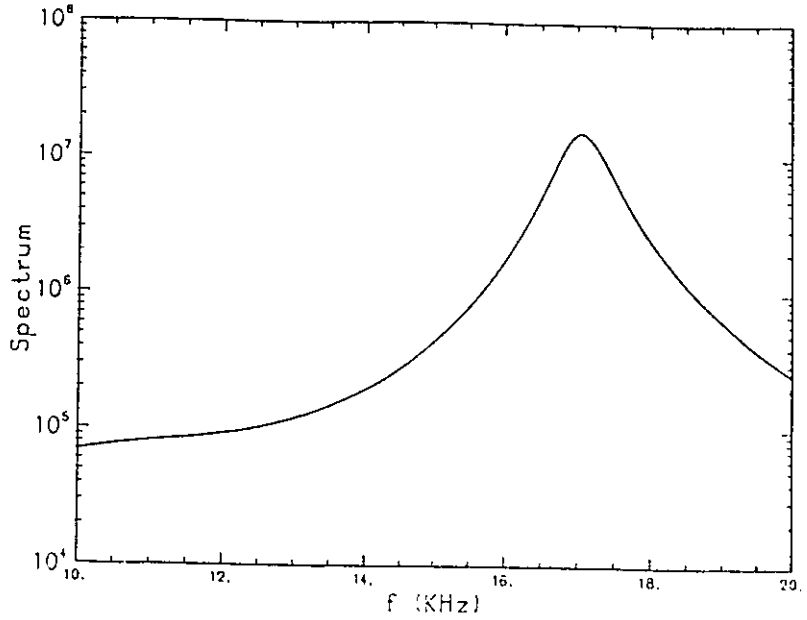


Figure 5.9: High Resolution ELT Frequency Estimation for GARELT12.

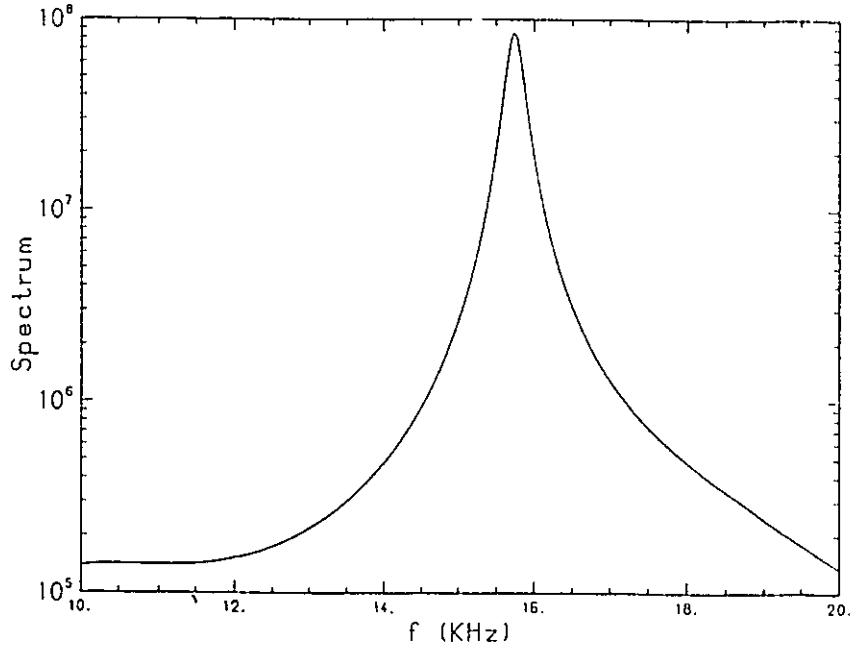


Figure 5.10: High Resolution ELT Frequency Estimation for PNTELT01.

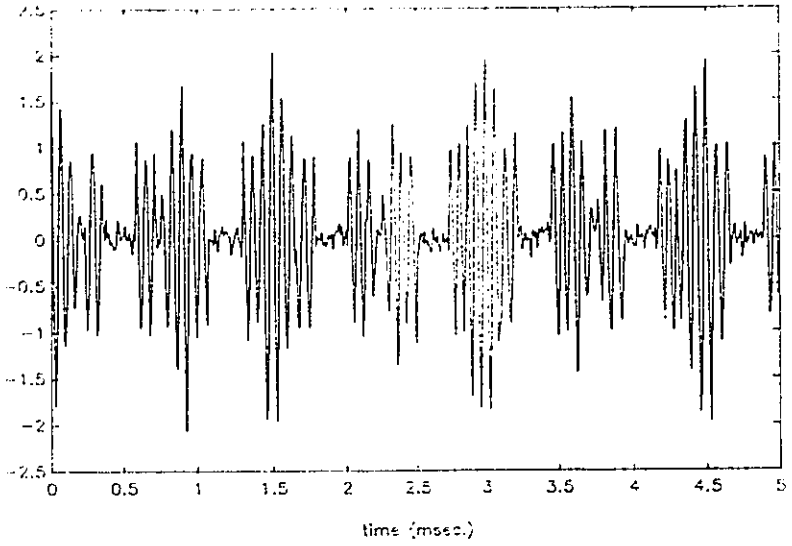


Figure 5.11: Simulated two ELT signals with $f_{c1} = 14KHz$ and $f_{c2} = 16KHz$, d=.5 and SNR=20dB.

signal power in the pulse "ON" duration to average noise power σ_n^2 ratio) in Fig. 5.11 is 20dB. We use the sampling frequency of 100KHz and estimate with the same rectangular window as we used in Figs. 5.7 to 5.10. Total data available is 4096 sampling points, which is just one block of data collected for the real ELT signal. Fig.5.12 shows the estimated spectra of the simulated ELT signals given above. The SNR is chosen to be SNR=0 and 20dB respectively.

From Fig.5.12, it can be observed that at SNR=20dB, two signals can be resolved. When SNR=0dB, only one peak at $f = 15KHz$ can be observed. Therefore, the Burg algorithm can only resolve two close frequency signals at high SNR.

Although the Burg algorithm is a high resolution algorithm, its resolution is still restricted by the window width. For the window width $2T_W = 0.28$ms. which is selected by considering the minimum pulse duration of ELT signals, the Burg algorithm cannot resolve two signals whose frequency distance is smaller than $1KHz$, even at high SNR.

It can also be noticed in Fig. 5.12 that for two or more ELT signals, a noticeable bias can be observed in the estimated frequencies. This bias does not appear to occur

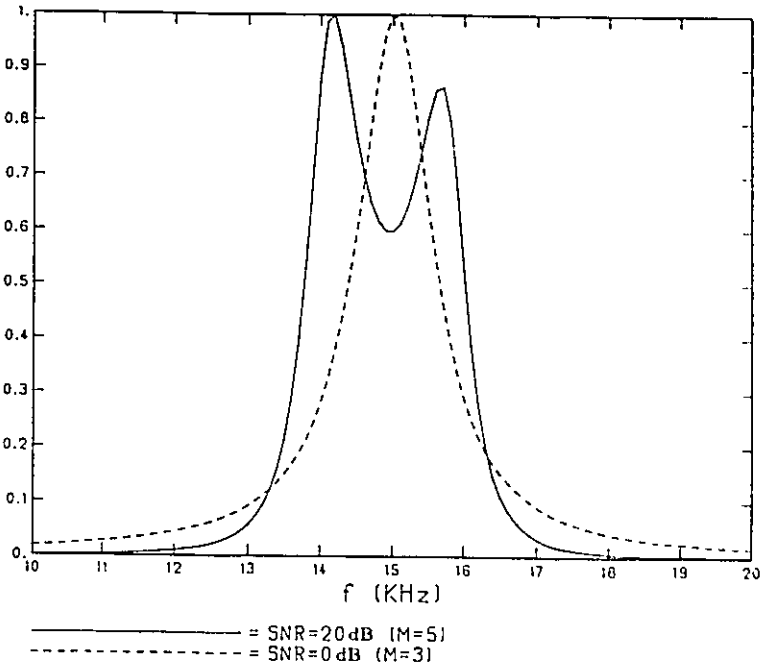


Figure 5.12: High Resolution ELT Frequency Estimation for Simulated Two ELT Signals: $f_{c1} = 16KHz$ and $f_{c2} = 14KHz$, with variable SNR.

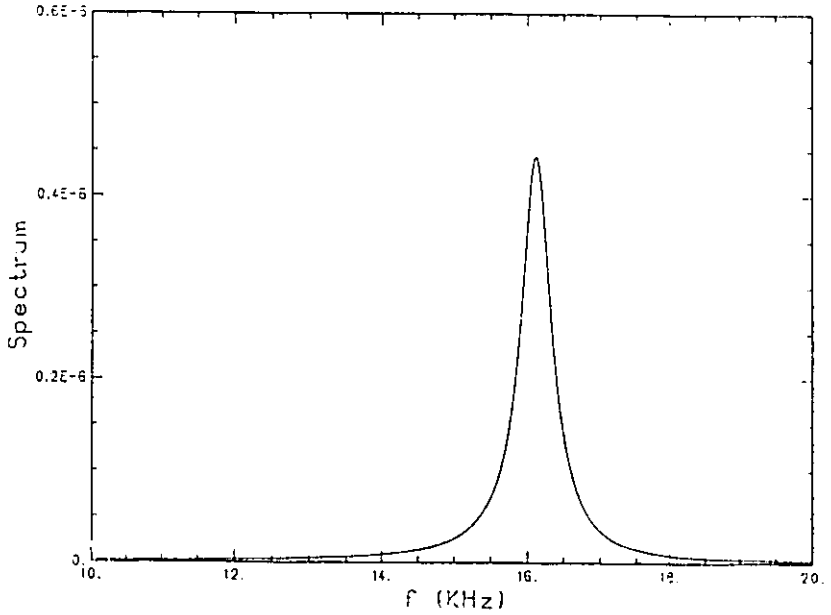


Figure 5.13: High Resolution ELT Frequency Estimation for Simulated ELT Signals: $f_c = 16KHz$.

in the case of only one ELT signal (see Fig.5.13, the spectrum of one ELT signal with $f_c = 16KHz$, SNR=20dB and other parameters are the same as before. Peak position presents no bias).

5.4 Discussions and Comparisons

Two methods are presented for the ELT frequency estimate in this chapter and both of them are based on the averaged quality $\overline{E}_x(\tau)$. It could be observed from estimated results using the real ELT signals that both methods yield almost the same estimated frequencies for one single signal. Now, we compare our results with that using the classical spectrum estimation methods [18]. Fig.5.14 shows the frequency estimation using the Maximum Entropy (Burg's algorithm) Methods (MEM) on the real ELT signals NARELT17, which is copied from the thesis of T. Chung [18]. According to the results in his thesis, the MEM gives very good frequency estimation in comparison with other classic spectrum estimation methods. However, there still exist spurious peaks in the spectrum of Fig.5.14 and it is very difficult to tell which peak corresponds to the signal frequency if no previous knowledge is given about the received ELT signal. Comparing the result in Fig.5.14 to that in Fig.5.8, we can say that the method based on the instantaneous spectrum can provide more reliable frequency estimation.

As for the two methods presented in this chapter, using the instantaneous frequency property of the PWD (Eq. (5.29)), the frequency estimation by finding the centre of gravity of $\overline{W}_x(f)$ is accurate and stable. Also the computation is relatively simple. The major drawback is that it depends on the assumption that only one ELT signal is received and will be invalid if two or more ELT signals exist. The application of the Burg algorithm to $\overline{E}_x(\tau)$ has the advantage that it can estimate two or more ELT frequencies at the same time. However, the Burg algorithm is not very stable [42] and

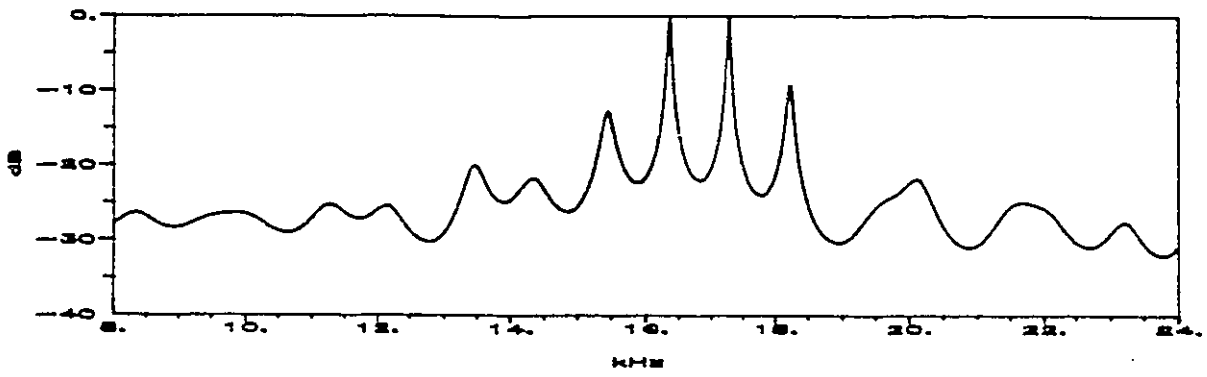


Figure 5.14: The Spectrum Obtained by Directly Applying MEM on the Real ELT Signal NARELT17 (Original copy from Fig4.14(c) of [18]).

its performance depends on the suitable selection of the order of AR model [41]. If the order is too small, proper resolution may not be obtained, and if the order is too large, spurious peak may yield false peaks resulting in erroneous signal frequency estimates and sometimes distorting the spectrum [42]. Therefore, we need a criterion to determine the suitable order M_1 for the AR model. The commonly used criteria for the proper order selection are the AIC (information theory criterion) and the FPE (Final Prediction Error), which have been briefly described in Appendix F. For both criteria, the suitable order M_1 will be the value which minimizes the corresponding criterion AIC(M_1) or FPE(M_1). Detailed analysis can be found in [41].

From the testing using real ELT data and the analysis, we have the following conclusions:

1. both methods presented in this chapter are robust to the different ELT signal models given in Section 5.1.

2. both methods yield almost the same frequency estimation for four real ELT signals.

Chapter 6

Simultaneous Time Delay and Doppler Shift Estimation

6.1 Time-Frequency Correlating Receiver

In a radar or sonar system, the received signals often contain both time delay (TD) and Doppler shift (DS). In order to estimate certain parameters associated with the targets, such as the range and the relative velocity, the accurate estimation of time delay and Doppler shift is crucial [87] [110].

Consider a signal processing unit with impulse response $h(t)$ in a radar or a sonar system such that it is designed to process the bandpass transmitted signal

$$s_0(t) = a(t)e^{j2\pi f_c t}, \tag{6.1}$$

where the real and imaginary parts of $s_0(t)$ form a pair of Hilbert transforms, $a(t)$ is the envelope of $s_0(t)$ and f_c is the carrier frequency. It is well known that the maximum output signal-to-noise ratio (SNR) can be obtained if $h(t)$ is matched to $s_0(t)$ such that

$$h(t) = s_0^*(-t). \tag{6.2}$$

The system $h(t)$ is called the matched filter and its ideal output is

$$s_0(t) \otimes h(t) = \left[\int a(\tau)a^*(\tau - t)d\tau\right] e^{j2\pi f_c t}, \tag{6.3}$$

where $\otimes$ denotes convolution. The output amplitude is maximum at $t = 0$.

However, due to the range and motion of the target, the received signal $s_R(t)$ often contains an unknown TD t_0 and an unknown DS f_Δ, i.e ,

$$s_R(t) = a(t - t_0)e^{j2\pi(f_c+f_\Delta)(t-t_0)}. \tag{6.4}$$

The output of the filter $h(t)$ will then be a function of $t-t_0$ and f_Δ and can be expressed, after a little simplification, as

$$\begin{aligned} & y(t - t_0, f_\Delta) = s_R(t) \otimes \{a^*(-t)e^{j2\pi f_c t}\} \\ = & \left[\int a(\tau + \frac{t - t_0}{2})a^*(\tau - \frac{t - t_0}{2})e^{j2\pi f_\Delta \tau}d\tau\right] e^{j2\pi(f_c+\frac{f_\Delta}{2})(t-t_0)}. \end{aligned} \tag{6.5}$$

The amplitude of the output $\mid y(t - t_0, f_\Delta) \mid$ reaches a maximum value at $t = t_0$ and $f_\Delta = 0$. The filter and the input is no longer matched in this case due to the existence of the Doppler shift f_Δ. At $t = t_0$, $\mid y(0, f_\Delta) \mid < \mid y(0,0) \mid$, and the output signal-to-noise ratio (SNR) at these parameter points is reduced rendering the estimation of these parameters less accurate.

This problem can be overcome by the use of a joint time-frequency representation (JTFR). The block diagram of a signal processing system employing a JTFR is shown in Fig.6.1. The input $x_i(t)$ is first demodulated and low-pass filtered. The resultant signal $x(t)$ is passed into the JTFR. The output $G_{xa}(t, f)$ is then used for the simultaneous estimation of t_0 and f_Δ.

The ambiguity function (AF) and the Wigner distribution (WD) are two commonly used examples of bilinear JTFR and their respective definitions were given in sections 2.3 and 2.4. Here we rewrite the expression of cross-representations as following:

$$A_{xa}(t, f) = \int x(\tau + \frac{t}{2})a^*(\tau - \frac{t}{2})e^{j2\pi f\tau}d\tau, \tag{6.6}$$

$$\begin{aligned} W_{xa}(t, f) &= \int x(t + \frac{\tau}{2})a^*(t - \frac{\tau}{2})e^{-j2\pi f\tau}d\tau \\ &= \int X(f + \frac{\nu}{2})A^*(f - \frac{\nu}{2})e^{j2\pi\nu t}d\nu. \end{aligned} \tag{6.7}$$

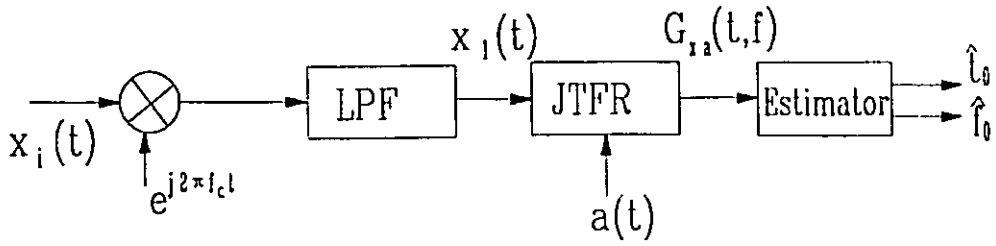


Figure 6.1: Simultaneous Estimation of Time Delay and Doppler Shift Using Joint Time-Frequency Representation.

where $X(f)$ and $A(f)$ are the Fourier transforms of $x(t)$ and $a(t)$, respectively. The WD has many desirable properties in signal analysis as described in Section 2.4 and it will be considered exclusively in this chapter although the development can be easily extended to other JTFR.

Let the WD be employed as the JTFR in Fig.6.1. Suppose that the input $x_i(t)$ to the processor is the bandpass analytic signal given by Eq.(6.4). For simplicity, we assume the envelope $a(t)$ to be real and symmetric, i.e.,

$$a(t) = a^*(t) = a(-t). \tag{6.8}$$

This assumption is often true in practice with the signal being time translated and represents no harsh constraint in practical applications. Under this assumption, the relationship between AF and WD can be derived as follows:

$$W_{xa}(t,f) = \int x(t+\frac{\tau}{2})a^*(t-\frac{\tau}{2})e^{-j2\pi f\tau}d\tau.$$

Putting $\tau_1 = \frac{\tau}{2}$ and using Eq.(6.8), we have

$$W_{xa}(t,f) = \int x(\tau_1+t)a^*(\tau_1-t)e^{-j4\pi f\tau_1}2d\tau_1 = 2A_{xa}(2t,-2f). \tag{6.9}$$

After the input $x_i(t)$ has been demodulated and low-pass filtered, the input to the JTFR is, to within a constant, given by

$$x_1(t) = s(t, t_0, f_\Delta) \equiv a(t - t_0)e^{j2\pi f_\Delta t}, \tag{6.10}$$

and the output of the JTFR is then

$$\begin{aligned} G_{sa}(t,f) &\equiv W_{sa}(t,f) \\ &= \int s(t+\frac{\tau}{2})a^*(t-\frac{\tau}{2})e^{-j2\pi f\tau}d\tau \\ &= e^{j2\pi f_\Delta t}\int a(t-t_0+\frac{\tau}{2})a^*(\frac{\tau}{2}-t)e^{-j2\pi(f-\frac{f_\Delta}{2})\tau}d\tau, \end{aligned} \tag{6.11}$$

where, in the last step of Eq.(6.11), Eqs.(6.8) and (6.10) have been used. The magnitude $|W_{sa}|$ in Eq.(6.11) attains its maximum at $t = \frac{t_0}{2}$ and $f = \frac{f_\Delta}{2}$ so that the estimation of TD and DS becomes a problem of locating the maximum point of $|W_{sa}|$. Comparing Eq.(6.11) with the matched filter given by Eq.(6.3), the simultaneous estimation using JTFR can be recognized as the **time-frequency correlating receiver**, which matches with signal in both time and frequency.

In practice, the signal is always accompanied by noise and hence the input to JTFR is given by

$$x(t) = a(t - t_0)e^{j2\pi f_\Delta t} + n(t), \tag{6.12}$$

where the noise $n(t)$ is assumed to be white Gaussian with zero mean such that

$$E[n(t_1)n^*(t_2)] = N_0\delta(t_1 - t_2), \tag{6.13}$$

and

$$E[n(t_1)n(t_2)] = 0 \qquad \forall t_1,\ t_2, \tag{6.14}$$

with N_0 being the power spectral density.

Therefore, the JTFR will have to process both signal and noise. The same strategy of locating the maximum point of $|W_{xa}|$ can still be employed to estimate TD and DS.

In this chapter, we examine the use of the WD as a JTFR in the estimation of time-delay and Doppler shift. We first analyze the statistical performance of the WD as an estimator, and then derive the Cramér-Rao bound for the joint estimation of TD and DS. We then show that the employment of the WD reaches the Cramér-Rao bound under reasonable high SNR. Finally, we present results of computer simulations showing the performance of the WD as an estimator, and describe what kind of signals are optimum for the best TD and DS estimation.

6.2 Joint Estimation of Time Delay and Doppler Shift Using WD

In the previous section, it has been shown that the use of WD as a JTFR for the joint estimation of TD and DS amounts to the location of the maximum point in $|W_{xa}(t,f)|$, i.e., we solve the pair of equation

$$\frac{\partial}{\partial t}|W_{xa}(t,f)|^2 = 0, \tag{6.15}$$

$$\frac{\partial}{\partial f}|W_{xa}(t,f)|^2 = 0. \tag{6.16}$$

If the solution of Eqs.(6.15) and (6.16) are t_p and f_p respectively, then from Eq.(6.11) and the discussion following that, the estimated values of TD and DS are then

$$\hat{t}_0 = 2t_p, \tag{6.17}$$

$$\hat{f}_\Delta = 2f_p. \tag{6.18}$$

Here in this section, we develop an approximate expression for the solution of Eqs.(6.15) and (6.16) so that an analysis of the performance of the estimation can be carried out.

Eqs.(6.15) and (6.16), when evaluated at the solution of $t = t_p$ and $f = f_p$, can be equivalently written as

$$Re\left\{W_{xa}(t_p,f_p)\cdot \dot{W}^*_{xa}(t_p,f_p)\right\} = 0, \tag{6.19}$$

$$Re\left\{W_{xa}(t_p, f_p) \cdot \dot{W}^{*}_{xa}(t_p, f_p)\right\} = 0, \tag{6.20}$$

where, for convenience, we introduce the notation that

$$\begin{aligned} \dot{W}_{xa}(t_p, f_p) &\equiv \frac{\partial}{\partial t} W_{xa}(t, f)\Big|_{\substack{t=t_p \\ f=f_p}}, \\ W'_{xa}(t_p, f_p) &\equiv \frac{\partial}{\partial f} W_{xa}(t, f)\Big|_{\substack{t=t_p \\ f=f_p}}. \end{aligned}$$

When the input to the JTFR is the data $x(t)$ in Eq.(6.12) containing both signal and noise, then using the definition of the WD in Eq.(6.7), the output of the JTFR is

$$W_{xa}(\tau, f) = W_{sa}(t, f) + W_{na}(t, f), \tag{6.21}$$

where $W_{sa}(t, f)$ is given by Eq.(6.11). Now, with no loss of generality, we let $t_0 = 0$ and $f_\Delta = 0$ in Eq.(6.11) since with non-zero delay and Doppler shift, Eq.(6.11) is only a translated version of the case when both quantities are zero. Hence, we can write,

$$W_{xa}(t, f) = W_a(t, f) + W_{na}(t, f). \tag{6.22}$$

Eq.(6.19) then becomes

$$Re\left\{[W_a(t_p, f_p) + W_{na}(t_p, f_p)] \cdot \dot{W}^{*}_{a}(t_p, f_p)\right\} = -Re\left\{[W_a(t_p, f_p) + W_{na}(t_p, f_p)]\, \dot{W}^{*}_{na}(t_p, f_p)\right\}. \tag{6.23}$$

But $W_a(t, f)$ is always real (see Section 2.4) and so is $\dot{W}_a(t, f)$. Thus if $W_a(t_p, f_p) \neq 0$, we can write Eq.(6.23) as

$$Re\left\{1 + \frac{W_{na}(t_p, f_p)}{W_a(t_p, f_p)}\right\} \cdot \dot{W}^{*}_{a}(t_p, f_p) = -Re\left\{\left[1 + \frac{W_{na}(t_p, f_p)}{W_a(t_p, f_p)}\right] \cdot \dot{W}^{*}_{na}(t_p, f_p)\right\}. \tag{6.24}$$

Under the high signal-to-noise ratio (SNR),

$$|W_{na}(t_p, f_p)| \ll |W_a(t_p, f_p)|. \tag{6.25}$$

Therefore we obtain

$$\dot{W}_a(t_p, f_p) \approx -Re\left\{\dot{W}^{*}_{na}(t_p, f_p)\right\}. \tag{6.26}$$

Similarly, under high SNR,

$$W'_a(t_p, f_p) \approx -Re\left\{W'^{*}_{na}(t_p, f_p)\right\}. \tag{6.27}$$

Now, expanding $W_a(t, f)$ in a Taylor series around $t = 0$ and $f = 0$, and ignoring terms higher than second order, we have

$$W_a(t, f) \approx W_a(0,0) + \frac{1}{2}t^2\ddot{W}_a(0,0) + \frac{1}{2}f^2W''_a(0,0) + ft\dot{W}'_a(0,0). \tag{6.28}$$

Notice that since $\dot{W}_a(0,0) = W'_a(0,0) = 0$, i.e., maximum at $t = 0$, $f = 0$, the first order terms in the expansion have vanished. Then, differentiating Eq.(6.28) with respect to t and f, we have

$$\dot{W}_a(t, f) \approx W_{11}t + W_{12}f, \tag{6.29}$$

$$W'_a(t, f) \approx W_{12}t + W_{22}f, \tag{6.30}$$

where we have simplified the notation by putting

$$W_{11} = \ddot{W}_a(0,0), \quad W_{12} = \dot{W}'_a(0,0), \quad W_{22} = W''_a(0,0). \tag{6.31}$$

Assuming that the solutions of Eqs.(6.15) and (6.16), $t = t_p$ and $f = f_p$, are sufficiently close to the true solution $t = 0$ and $f = 0$ so that the terms higher than the second order in the Taylor expansion above can be ignored, then Eqs.(6.29) and (6.30) are true for $t = t_p$ and $f = f_p$. Thus, combining Eqs.(6.26), (6.27) and (2.29), (6.30), we obtain

$$W_{11}t_p + W_{12}f_p \approx -Re\left\{\dot{W}^{*}_{na}(t_p, f_p)\right\}, \tag{6.32}$$

$$W_{12}t_p + W_{22}f_p \approx -Re\left\{W'^{*}_{na}(t_p, f_p)\right\}. \tag{6.33}$$

Eqs.(6.32) and (6.33) are a pair of linear equations the solutions of which are

$$t_p \approx \frac{1}{\Delta}\left\{-W_{22}Re\left[\dot{W}^{*}_{na}(t_p, f_p)\right] + W_{12}Re\left[W'^{*}_{na}(t_p, f_p)\right]\right\}, \tag{6.34}$$

$$f_p \approx \frac{1}{\Delta}\left\{W_{12}Re\left[\dot{W}^{*}_{na}(t_p, f_p)\right] - W_{11}Re\left[W'^{*}_{na}(t_p, f_p)\right]\right\}, \tag{6.35}$$

where

$$\Delta = \begin{vmatrix} W_{11} & W_{12} \\ W_{12} & W22 \end{vmatrix}. \tag{6.36}$$

The values of W_{11}, W_{22} and W_{12} can be obtained directly:

$$\begin{aligned} W_{12} &= \frac{\partial^2}{\partial t \partial f} W_a(t,f)\Big|_{\substack{t=0 \\ f=0}} \\ &= \frac{\partial^2}{\partial t \partial f}\left[\int a(t+\frac{\tau}{2})a^*(\frac{\tau}{2}-t)e^{-j2\pi f\tau}\,d\tau\right]\Big|_{\substack{t=0 \\ f=0}} \\ &= \int(-j2\pi\tau)[\dot{a}(\frac{\tau}{2})a^*(\frac{\tau}{2}) - a(\frac{\tau}{2})\dot{a}^*(\frac{\tau}{2})]d\tau \\ &= 4\pi\int\tau Im[\dot{a}(\frac{\tau}{2})a^*(\frac{\tau}{2})]d\tau \\ &= 16\pi\int\tau Im[\dot{a}(\tau)a^*(\tau)]d\tau, \end{aligned} \tag{6.37}$$

where $\dot{a}$ denotes $\frac{\partial}{\partial t}a$. Notice that when $a(t)$ is real, $W_{12} = 0$.

Similarly,

$$W_{22} = -32\pi^2\int\tau^2|a(\tau)|^2d\tau. \tag{6.38}$$

For the evaluation of W_{11}, we can employ the definition of WD in its Fourier transform, i.e. the second equality of Eq.(6.7) and carry out the differentiation similarly, resulting in

$$W_{11} = -32\pi^2\int\nu^2|A(\nu)|^2d\nu. \tag{6.39}$$

Eqs.(6.34) and (6.35) yield an expressions for the peak of the WD $|W_{xa}(t,f)|$ from which we can estimate TD and DS. These expressions would help to analyze the performance of such an estimator. In practice, of course, the peak of $|W_{xa}(t,f)|$ is located by suitably thresholding the output of the JTFR. Detail will be discussed in Section 6.5.2.

6.3 Performance Analysis for the WD

Let us now examine how the WD performance when used to jointly estimate the TD and the DS of a signal. Eqs (6.17) and (6.18) express the relationship between the time

delay and Doppler shift estimates to the location of the peak of the WD. Eqs.(6.34) and (6.35) further tell us where the peak is in relation to the cross WD between the noise and the signal. Thus, we are in a position to analyze the bias and the variance of the estimates.

From Eqs.(6.34) and (6.35), we see that $W_{na}(t_p, f_p)$ is a random variable. Hence the biases and the variances of the estimates depend on the means and correlations of the derivatives of $W_{na}(t_p, f_p)$. Before evaluating these quantities, we first need the following lemmas which show some important properties of the noise.

Lemma 6.1 *Given that n(t) is zero-mean Gaussian with properties stated in Eqs.(6.13) and (6.14), then*

$$E[N(f)] = 0, \tag{6.40}$$

$$E[N(f_1)N(f_2)] = E[N^*(f_1)N^*(f_2)] = 0, \tag{6.41}$$

$$E[N(f_1)N^*(f_2)] = N_0\delta(f_1 - f_2), \tag{6.42}$$

where $N(f)$ is the Fourier transform of $n(t)$.

Proof:

$$E[N(f)] = \int E[n(t)]e^{-j2\pi ft}dt = 0,$$

which is Eq.(6.40). Also using Eq.(6.13),

$$E[N(f_1)N(f_2)] = \int\int E[n(t_1)n(t_2)]e^{-j2\pi f_1 t_1 - j2\pi f_2 t_2}dt_1 dt_2 = 0$$

which proves Eq.(6.41). Finally using Eq.(6.14),

$$\begin{aligned} E[N(f_1)N^*(f_2)] &= \int\int E[n(t_1)n^*(t_2)]e^{-j2\pi(f_1 t_1 + f_2 t_2)}dt_1 dt_2 \\ &= N_0 \int\int \delta(t_1 - t_2)e^{-j2\pi(f_1 t_1 + f_2 t_2)}dt_1 dt_2 \\ &= N_0 \int e^{-j2\pi(f_1 - f_2)t_1}dt_1 \\ &= N_0\delta(f_1 - f_2), \end{aligned}$$

which is Eq.(6.42). □

Lemma 6.2 *Given that n(t) is zero-mean Gaussian with properties stated in Eqs.(6.13) and (6.14), then*

$$E[n(t)N(f)] = E[n^*(t)N^*(f)] = 0 \qquad \forall t, f, \tag{6.43}$$

and

$$E[n^*(t)N(f)] = N_0 e^{j2\pi ft}. \tag{6.44}$$

Proof: Since

$$N(f) = \int n(\tau) e^{-j2\pi f\tau} d\tau,$$

we have, from Eq.(6.14),

$$E[n(t)N(f)] = \int E[n(t)n(\tau)] e^{-j2\pi f\tau} d\tau = 0.$$

Similarly

$$E[n^*(t)N^*(f)] = 0.$$

Furthermore,

$$\begin{aligned} E[n(t)N^*(f)] &= \int E[n(t)n^*(\tau)] e^{-j2\pi f\tau} d\tau \\ &= N_0 \int \delta(t-\tau) e^{-j2\pi f\tau} d\tau = N_0 e^{-j2\pi ft}, \end{aligned}$$

which proves Eq.(6.44) □

Now, we are in the position to evaluate the mean and correlations of the derivatives of $W_{na}(t_p, f_p)$. These are stated in the following theorems.

Theorem 6.1 *Given that n(t) is zero-mean Gaussian, then*

$$E[\dot{W}_{na}(t,f)] = 0 \quad \forall t, f, \tag{6.45}$$

$$E[W'_{na}(t,f)] = 0 \quad \forall t, f. \tag{6.46}$$

Proof: Eqs.(6.45) and (6.46) can be easily proved by direct substitution of the WD in Eq.(6.7) and employing the fact that $E[n(t)] = 0$. □

Theorem 6.2 *Given that n(t) is zero-mean Gaussian with properties stated in Eq.(6.13) and (6.14), then*

$$E[\{\dot{W}_{na}(t,f)\}^2] = 0 \quad \forall t, f, \tag{6.47}$$

$$E[\{W'_{na}(t,f)\}^2] = 0 \quad \forall t, f, \tag{6.48}$$

$$E[\dot{W}_{na}(t,f)W'_{na}(t,f)] = 0 \quad \forall t, f. \tag{6.49}$$

Proof: Using the definition of the WD in Eq.(6.7) and Eq.(6.41), we have

$$\begin{aligned} & E[\{\dot{W}_{na}(t,f)\}^2] \\ = \; & -4\pi^2 \int\int \nu_1\nu_2 E[N(f+\tfrac{\nu_1}{2})N(f+\tfrac{\nu_2}{2})]A^*(f-\tfrac{\nu_1}{2})A^*(f-\tfrac{\nu_2}{2})e^{j2\pi(\nu_1+\nu_2)t}d\nu_1 d\nu_2 \\ = \; & 0, \end{aligned}$$

which proves Eq.(6.47). Also using Eq.(6.14) and the definition of the WD in the first equality of Eq.(6.7), we have

$$\begin{aligned} & E[\{W'_{na}(t,f)\}^2] \\ = \; & -4\pi^2 \int\int \tau_1\tau_2 E[n(t+\tfrac{\tau_1}{2})n(t+\tfrac{\tau_2}{2})]a^*(t-\tfrac{\tau_1}{2})a^*(t-\tfrac{\tau_2}{2})e^{-j2\pi f(\tau_1+\tau_2)}d\tau_1 d\tau_2 \\ = \; & 0, \end{aligned}$$

which is Eq.(6.48). Finally, using Eq.(6.43),

$$\begin{aligned} & E[\dot{W}_{na}(t,f)W'_{na}(t,f)] \\ = \; & -4\pi^2 \int\int \nu\tau E[N(f+\tfrac{\nu}{2})n(t+\tfrac{\tau}{2})]A^*(f-\tfrac{\nu}{2})a^*(t-\tfrac{\tau}{2})e^{j2\pi(\nu t-f\tau)}d\tau d\nu \\ = \; & 0, \end{aligned}$$

which proves Eq.(6.49) □

Theorem 6.3 *Given that n(t) is zero-mean Gaussian with properties stated in Eq.(6.13) and (6.14), then*

$$E\left[\dot{W}_{na}(t_p,f_p)\dot{W}^*_{na}(t_p,f_p)\right] \approx -2N_0W_{11}, \tag{6.50}$$

$$E\left[W'_{na}(t_p,f_p)W'^*_{na}(t_p,f_p)\right] \approx -2N_0W_{22}, \tag{6.51}$$

and

$$E\left[W'^*_{na}(t_p,f_p)\dot{W}_{na}(t_p,f_p) + W'_{na}(t_p,f_p)\dot{W}^*_{na}(t_p,f_p)\right] \approx -4N_0W_{12}, \tag{6.52}$$

where $\dot{W}_{11}$, W_{22}, and W_{12} are defined in Eq.(6.31) and are evaluated in Eqs.(6.38), (6.39) and (6.37) respectively.

Proof: Using the definition of the WD in Eq.(6.7), we have

$$\begin{aligned}
& E\left[\dot{W}_{na}(t_p,f_p)\dot{W}^*_{na}(t_p,f_p)\right] \\
= \; & 4\pi^2\int\int \nu_1\nu_2 E\left[N(f+\frac{\nu_1}{2})N^*(f+\frac{\nu_2}{2})\right]A^*(f-\frac{\nu_1}{2})A(f-\frac{\nu_2}{2})e^{j2\pi(\nu_1 t-\nu_2 t)}d\nu_1 d\nu_2\Big|_{\substack{t=t_p\\f=f_p}} \\
= \; & 4\pi^2 N_0\int\int \nu_1\nu_2\delta(\frac{\nu_1-\nu_2}{2})A^*(f-\frac{\nu_1}{2})A(f-\frac{\nu_2}{2})e^{j2\pi(\nu_1 t-\nu_2 t)}d\nu_1 d\nu_2\Big|_{\substack{t=t_p\\f=f_p}} \\
= \; & 64\pi^2 N_0\int \nu_1^2 A^*(f-\nu_1)A(f-\nu_1)d\nu_1\Big|_{\substack{t=t_p\\f=f_p}}.
\end{aligned}$$

Let $f-\nu_1=\mu$, then

$$\begin{aligned}
& E\left[\dot{W}_{na}(t_p,f_p)\dot{W}^*_{na}(t_p,f_p)\right] \\
= \; & 64\pi^2 N_0\int (f-\mu)^2|A(\mu)|^2 d\mu\Big|_{\substack{t=t_p\\f=f_p}} \\
= \; & 64\pi^2 N_0\left\{\int f^2|A(\mu)|^2 d\mu - 2f\int \mu|A(\mu)|^2 d\mu + \int \mu^2|A(\mu)|^2 d\mu\right\}_{f=f_p}.
\end{aligned}$$

Now, $|A(\mu)|^2$ is an even function for real $a(t)$, so that the second term in the braces is zero. Also the last term is recognized as $-W_{11}$ from Eq.(6.38), then

$$E\left[\dot{W}_{na}(t_p,f_p)\dot{W}^*_{na}(t_p,f_p)\right] = 2N_0\left\{-W_{11}+32\pi^2 f^2\int |A(\mu)|^2 d\mu\right\}_{f=f_p}.$$

Eq.(6.50) follows from the fact that $f_p \approx 0$ at high SNR. Using the definition of the WD in the first equality of Eq.(6.7), we have

$$\begin{aligned}
& E\left[W'_{na}(t_p,f_p)W'^{*}_{na}(t_p,f_p)\right] \\
= \;& 4\pi^2 \int\int \tau_1\tau_2 E\left[n(t+\frac{\tau_1}{2})n^*(t+\frac{\tau_2}{2})\right] a^*(t-\frac{\tau_1}{2})a(t-\frac{\tau_2}{2})e^{-j2\pi f(\tau_1-\tau_2)}d\tau_1 d\tau_2\Big|_{\substack{t=t_p\\f=f_p}} \\
= \;& 64\pi^2 N_0 \int \tau_1^2 a^*(t-\tau_1)a(t-\tau_1)d\tau_1\Big|_{\substack{t=t_p\\f=f_p}} \\
\overset{\tau=t-\tau_1}{=} \;& 64\pi^2 N_0 \left\{\int \tau^2|a(\tau)|^2 d\tau + t^2\int |a(\tau)|^2 d\tau\right\}_{t=t_p} \\
= \;& -2N_0W_{22} + 64\pi^2 N_0 t_p^2 \int |a(\tau)|^2 d\tau \\
\approx \;& -2N_0W_{22},
\end{aligned}$$

where, in the last step, Eq.(6.39) has been used and high SNR for which $t_p \approx 0$ has been assumed. This proves Eq.(6.51). Using the definitions of the first and second equality of Eq.(6.7) respectively in W'_{na} and $\dot{W}_{na}$, we have

$$\begin{aligned}
& E\left[W'^{*}_{na}(t_p,f_p)\dot{W}_{na}(t_p,f_p)\right] \\
= \;& -4\pi^2 \int\int \nu\tau E[n^*(t+\frac{\tau}{2})N(f+\frac{\nu}{2})]a(t-\frac{\tau}{2})A^*(f-\frac{\nu}{2})e^{j2\pi(f\tau+\nu t)}d\tau d\nu\Big|_{\substack{t=t_p\\f=f_p}} \\
= \;& -4\pi^2 N_0 \int\int \tau\nu e^{-j2\pi(f+\frac{\nu}{2})(t+\frac{\tau}{2})}a(t-\frac{\tau}{2})A^*(f-\frac{\nu}{2})e^{j2\pi(f\tau+\nu t)}d\tau d\nu\Big|_{\substack{t=t_p\\f=f_p}} \\
= \;& -4\pi^2 N_0 \int \left[\tau a(t-\frac{\tau}{2})d\tau\right]\cdot\left[\int \nu A^*(\frac{\nu}{2}-f)e^{-j2\pi(\frac{\nu}{2}-f)(\frac{\tau}{2}-t)}d\nu\right]_{\substack{t=t_p\\f=f_p}}, \qquad (6.53)
\end{aligned}$$

where in the second equality, Eq.(6.44) is applied and in the final equality, the fact that if $a(t) = a(-t)$ is used, then $A(f) = A(-f)$. Furthermore by letting $\mu = \frac{\nu}{2} - f$, the second integral in Eq.(6.53) becomes

$$\begin{aligned}
& \int \nu A^*(\frac{\nu}{2}-f)e^{-j2\pi(\frac{\nu}{2}-f)(\frac{\tau}{2}-t)}d\nu \\
= \;& 4\left[\int \{\mu A^*(\mu) + fA^*(f)\}\, e^{-j2\pi\mu(\frac{\tau}{2}-t)}d\mu\right] \\
= \;& 4\left[-\frac{j}{2\pi}\dot{a}^*(t-\frac{\tau}{2}) + fa^*(t-\frac{\tau}{2})\right]. \qquad (6.54)
\end{aligned}$$

Substituting Eq.(6.54) into (6.53), we have

$$\begin{aligned}
& E\left[W'^{*}_{na}(t_p,f_p)\dot{W}_{na}(t_p,f_p)\right] \\
= \;& 16\pi^2 N_0 \int \tau \left[\frac{j}{2\pi}a(t-\frac{\tau}{2})\dot{a}^{*}(t-\frac{\tau}{2}) - f a(t-\frac{\tau}{2})a^{*}(t-\frac{\tau}{2})\right] d\tau \Big|_{\substack{t=t_p \\ f=f_p}} \\
\overset{t_1=t-\frac{\tau}{2}}{=} \;& 64\pi^2 N_0 \int (t-t_1)\left[\frac{j}{2\pi}a(t_1)\dot{a}^{*}(t_1) - f a(t_1)a^{*}(t_1)\right] dt_1 \Big|_{\substack{t=t_p \\ f=f_p}} \\
= \;& 64\pi^2 N_0 \left[-\frac{j}{2\pi}\int t_1 a(t_1)\dot{a}^{*}(t_1)dt_1 - ft\int |a(t_1)|^2 dt_1\right]_{\substack{t=t_p \\ f=f_p}} \\
\approx \;& -32\pi N_0 \int j t_1 a(t_1)\dot{a}^{*}(t_1)dt_1, \qquad (6.55)
\end{aligned}$$

where, to arrive at the third equality we have used the facts that for real and even $a(t_1)$,

$$\int a(t_1)\dot{a}^{*}(t_1)dt_1 = 0, \quad \int t_1|a(t_1)|^2 dt_1 = 0,$$

and the last equality of Eq.(6.55) is obtained using the fact that at high SNR, $t_p \approx 0$ and $f_p \approx 0$. Thus, by comparing to Eq.(6.37)

$$\begin{aligned}
& E\left[W'^{*}_{na}(t_p,f_p)\dot{W}_{na}(t_p,f_p) + W'_{na}(t_p,f_p)\dot{W}^{*}_{na}(t_p,f_p)\right] \\
\approx \;& -32\pi N_0 \int j t_1 \left\{a(t_1)\dot{a}^{*}(t_1) - a^{*}(t_1)\dot{a}(t_1)\right\} dt_1 \\
= \;& -4N_0 W_{12},
\end{aligned}$$

which proves Eq.(6.52). □

With the aid of the above theorems, we can evaluate the performance of the WD as a joint estimator of TD and DS. To analyze the performance, we examine Eqs.(6.17) and (6.18) which express TD and DS in terms of the peak on the time and frequency axes respectively. The location of the peak is governed by Eqs.(6.34) and (6.35). From these two equations and Theorem 6.1, we obtain

$$\begin{aligned}
E[t_p] &= \frac{1}{\Delta}\left\{-W_{22}E\left[Re\left(\dot{W}^{*}_{na}(t_p,f_p)\right)\right] + W_{12}E\left[Re\left(W'^{*}_{na}(t_p,f_p)\right)\right]\right\} \\
&= 0, \qquad (6.56)
\end{aligned}$$

$$E[f_p] = \frac{1}{\Delta}\left\{W_{12}E\left[Re\left(\dot{W}^*_{na}(t_p,f_p)\right)\right] - W_{11}E\left[Re\left(W'^*_{na}(t_p,f_p)\right)\right]\right\}$$
$$= 0. \tag{6.57}$$

Since the true solution is $t_0 = 0$ and $f_\Delta = 0$, we can conclude from Eqs.(6.17) and (6.18) that the joint estimation is unbiased. To evaluate the variance of the estimate, we evaluate the covariance matrix of the estimates:

$$Cov(\hat{t}_0, \hat{f}_\Delta) = E\left\{\begin{bmatrix}(\hat{t}_0 - t_0)\\ (\hat{f}_\Delta - f_\Delta)\end{bmatrix}\begin{bmatrix}(\hat{t}_0 - t_0) & (\hat{f}_\Delta - f_\Delta)\end{bmatrix}\right\}$$
$$= 4\begin{bmatrix}E(t_p^2) & E(f_p t_p)\\ E(f_p t_p) & E(f_p^2)\end{bmatrix}. \tag{6.58}$$

We now proceed on to evaluate the elements of the matrix in Eq.(6.58). From Eq.(6.34), we have

$$E(t_p^2) = \frac{1}{\Delta^2}\left\{W_{22}^2E\left[Re\left(\dot{W}^*_{na}(t_p,f_p)\right)\right]^2 + W_{12}^2E\left[Re\left(W'^*_{na}(t_p,f_p)\right)\right]^2 - 2W_{12}W_{22}E\left[Re\left(\dot{W}^*_{na}(t_p,f_p)\right)\cdot Re\left(W'^*_{na}(t_p,f_p)\right)\right]\right\}, \tag{6.59}$$

where Δ is given by Eq.(6.36).

But for any two complex quantities F and G,

$$E[Re(F^*)Re(G^*)] = \frac{1}{4}E[(F+F^*)(G+G^*)]$$
$$= \frac{1}{4}E[FG + F^*G^* + F^*G + FG^*]. \tag{6.60}$$

Thus using Eq.(6.59) and Theorems 6.2 and 6.3, we have

$$E\left\{\left(Re\left[\dot{W}^*_{na}(t_p,f_p)\right]\right)^2\right\} = \frac{1}{4}\left\{2E\left[\dot{W}_{na}(t_p,f_p)\dot{W}^*_{na}(t_p,f_p)\right]\right\}$$
$$= -N_0W_{11}. \tag{6.61}$$

Similarly

$$E\left\{\left(Re\left[W'^*_{na}(t_p,f_p)\right]\right)^2\right\} = -N_0W_{22}, \tag{6.62}$$

and

$$E\left[\left(Re\dot{W}_{na}^{*}(t_p,f_p)\right)\left(ReW_{na}^{'*}(t_p,f_p)\right)\right] = -N_0W_{12}. \tag{6.63}$$

Thus we have

$$E[t_p^2] = -\frac{N_0}{\Delta^2}W_{22}\{W_{11}W_{22} - W_{12}^2\} = \frac{-N_0}{\Delta}W_{22}. \tag{6.64}$$

Similarly,

$$E[f_p^2] = \frac{-N_0}{\Delta}W_{11}, \tag{6.65}$$

and

$$E[f_pt_p] = \frac{-N_0}{\Delta}W_{12}, \tag{6.66}$$

and the covariance matrix of t_0 and f_Δ is then given by

$$Cov(\hat{t}_0,\hat{f}_\Delta) = -\frac{4N_0}{\Delta}\begin{bmatrix} W_{22} & -W_{12} \\ -W_{12} & W_{11} \end{bmatrix}. \tag{6.67}$$

Note that for real $a(t)$, $W_{12} = 0$ and we have

$$Cov(\hat{t}_0,\hat{f}_\Delta) = -4N_0\begin{bmatrix} \frac{1}{W_{11}} & 0 \\ 0 & \frac{1}{W_{22}} \end{bmatrix}. \tag{6.68}$$

6.4 The Cramér-Rao Bound of the Joint Estimation

The Cramér-Rao bound (CRB) expresses a lower bound for the variance of an unbiased estimate and is, in general, not too complicated to compute. By comparing the performance of an estimator to the CRB, we can often obtain an indication of how well the estimator performs. The CRB for the unbiased estimation of M deterministic parameters g can be stated as follows [95]:

$$Cov\{\hat{g}\} \equiv E[(\hat{g}-g)\cdot(\hat{g}-g)^{\dagger}] \geq \mathbf{J}^{-1}, \tag{6.69}$$

where $Cov(\cdot)$ denotes the covariance matrix, $(\hat{g}-g)$ is the M-dimensional vector of estimation errors, the superscript $\dagger$ denotes the Hermitian transpose of a matrix or

vector, and $\mathbf{J}$ is the $M \times M$ Fisher information matrix defined as

$$\mathbf{J} \triangleq E\left[(\nabla_g\{\ln p(x|g)\}) \cdot (\nabla_g\{\ln p(x|g)\})^\dagger\right], \tag{6.70}$$

with

$$\nabla_g \triangleq [\frac{\partial}{\partial g_1} \cdots \frac{\partial}{\partial g_M}]^T. \tag{6.71}$$

The inequality for matrices $\mathbf{A} \geq \mathbf{B}$ in Eq.(6.69) means that $(\mathbf{A} - \mathbf{B})$ is non-negative definite. In particular, the variance of the mth estimated parameter is given by

$$Var(\hat{g}_i - g_i) \geq J^{ii}, \tag{6.72}$$

where J^{ii} is the iith element of $\mathbf{J}^{-1}$.

To evaluate the CRB for the joint estimation of time-delay and Doppler shift, we have to evaluate $\mathbf{J}$ which, in turn, needs the conditional probability density function (PDF) $p(x|g)$ where g contains the two parameters t_0 and f_Δ, i.e.,

$$g = [t_0 \quad f_\Delta]^T. \tag{6.73}$$

From our assumptions that the noise is white Gaussian with power N_0, following standard procedures [95] of taking samples of the received data and letting the sampling period tend to zero, we can write

$$\ln p(x|g) = -\frac{1}{N_0}\int |x(t) - s(t; t_0, f_\Delta)|^2 dt + k_0, \tag{6.74}$$

where k_0 is a constant and $s(t; t_0, f_\Delta)$ is given by Eq.(6.10).

Differentiating Eq.(6.74) with respect to g_i, $i = 1, 2$, and using Eq.(6.12), we have

$$\frac{\partial \ln p(x|t_0, f_\Delta)}{\partial g_i} = \frac{2}{N_0} Re\left[\int n(t) \frac{\partial}{\partial g_i} s^*(t; t_0, f_\Delta) dt\right]. \tag{6.75}$$

Substituting Eq.(6.75) into Eq.(6.70), and using Eqs.(6.13) and (6.14), the ijth element of $\mathbf{J}$ is given by

$$J_{ij} = \frac{1}{N_0^2} E\left[\int\int \left\{n(t_1)\frac{\partial}{\partial g_i} s^*(t_1; t_0, f_\Delta) + n^*(t_1)\frac{\partial}{\partial g_i} s(t_1; t_0, f_\Delta)\right\}\right.$$

$$\left\{ n(t_2)\frac{\partial}{\partial g_j}s^*(t_2;t_0,f_\Delta) + n^*(t_2)\frac{\partial}{\partial g_j}s(t_2;t_0,f_\Delta) \right\} dt_1 dt_2 \Bigg]$$

$$= \frac{2}{N_0} Re\left[\int\int \delta(t_1-t_2)\frac{\partial}{\partial g_i}s^*(t_1;t_0,f_\Delta)\frac{\partial}{\partial g_j}s(t_2;t_0,f_\Delta)dt_1 dt_2 \right]$$

$$= \frac{2}{N_0} Re\left[\int \frac{\partial}{\partial g_i}s^*(t_1;t_0,f_\Delta)\frac{\partial}{\partial g_j}s(t_1;t_0,f_\Delta)dt_1 \right]. \tag{6.76}$$

From Eq.(6.10), we have

$$\begin{aligned}\frac{\partial}{\partial g_1}s(t;t_0,f_\Delta) &= \frac{\partial}{\partial t_0}a(t-t_0)e^{j2\pi f_\Delta t} \\ &= -\dot{a}(t-t_0)e^{j2\pi f_\Delta t},\end{aligned} \tag{6.77}$$

and

$$\begin{aligned}\frac{\partial}{\partial g_2}s(t;t_0,f_\Delta) &= \frac{\partial}{\partial f_\Delta}a(t-t_0)e^{j2\pi f_\Delta t} \\ &= 2\pi j t a(t-t_0)e^{j2\pi f_\Delta t}.\end{aligned} \tag{6.78}$$

Thus, substituting Eqs.(6.77) and (6.78) into (6.76), we have

$$\begin{aligned}J_{11} &= \frac{2}{N_0}\int |\dot{a}(t-t_0)|^2 dt \\ &= \frac{2}{N_0}\int \nu^2 |A(\nu)|^2 d\nu \\ &= -\frac{1}{4N_0}W_{11},\end{aligned} \tag{6.79}$$

where the Parseval's theorem has been employed. Similarly

$$J_{22} = -\frac{1}{4N_0}W_{22}, \tag{6.80}$$

$$J_{12} = -\frac{1}{4N_0}W_{12}. \tag{6.81}$$

Hence the Cramér-Rao bound in the joint estimation of time delay and Doppler shift is given by

$$Cov(\hat{t}_0,\hat{f}_\Delta) \geq -\frac{4N_0}{\Delta}\begin{bmatrix} W_{22} & -W_{12} \\ -W_{12} & W_{11} \end{bmatrix}, \tag{6.82}$$

where Δ is defined in Eq.(6.36).

Comparing Eqs.(6.67) and (6.82), it can be seen that under high SNR, the use of the WD to jointly estimate TD and DS is efficient in the sense that it is unbiased and its variance reaches the Cramér-Rao bound.

6.5 Performance Analysis and Optimum Signal

6.5.1 Simulation Results

In this section, we present some computer simulation results which are obtained by employing two commonly used real signals and using the WD for jointly estimating TD and DS. The two commonly used signals are (i) the Gaussian pulse and (ii) the linear FM pulse. The input to the JTFR in Fig.6.1 is modelled according to Eq.(6.12) in which Gaussian noise with spectral power density N_0 watts/Hz has been added to the signal. The signal energy is defined as

$$E_a = \int_{-\infty}^{\infty} a^2(t)dt. \tag{6.83}$$

The signal-to-noise ratio in our simulation is defined as

$$\rho_{SN} = \frac{E_a}{N_0}. \tag{6.84}$$

The following is a general description of the simulation which is similar to that given in Chapter 4:

1. For a particular ρ_{SN}, the received data $x(t)$ and the locally generated signal $a(t)$ are used to evaluate the WD $W_{xa}(t, f)$.

2. the peak of the WD $W_{xa}(t, f)$ is located and the corresponding co-ordinates t_p and f_p are recorded.

3. The difference between the true time delay t_0 and $2t_p$, as well as the difference between the true Doppler shift f_Δ and $2f_p$ are measured. These are the errors of the estimates.

4. The above procedure is repeated 300 times with different realizations of white Gaussian noise of the same spectral power density N_0 and the mean square error is then calculated.

5. The above procedure is repeated for various values of ρ_{SN}.

In the examples shown below, we choose following two time-limited signals:

$$\text{Gaussian Pulse:} \qquad a(t) = e^{-\pi t^2/16}[u(t+4) - u(t-4)], \tag{6.85}$$

$$\text{Linear FM Pulse:} \qquad a(t) = cos(0.32\pi t^2)[u(t+4) - u(t-4)], \tag{6.86}$$

where $u(t)$ is the unit step function.

Fig.6.2 shows the mean square error of the time delay estimate versus ρ_{SN}, the ratio of signal energy to noise spectral power density. The theoretical performance and the simulation performance for both signals are plotted for comparison. The theoretical performance assumes high SNR and, as shown in last section, is identical to the Cramér-Rao bound. At high SNR, the simulated performance of both signals agree closely with the theoretical evaluation. It can be seen that for the same accuracy of the estimate, the Gaussian pulse necessitates a much higher SNR (nearly 20dB) than the linear FM pulse when ρ_{SN} is large. When $\rho_{SN} \approx$ 15dB, both signals experience threshold effects and their performance deteriorate quickly. This is because in deriving the theoretical performance, we have assumed that t_p and f_p are close to the origin and ignored the second and higher order terms in the Taylor expansion in Eq.(6.28). Furthermore, when SNR is low, $|W_{na}(t_p, f_p)|$ is no longer negligible compared to $|W_a(t_p, f_p)|$ and the approximations in Eqs.(6.26) and (6.27) would not be accurate any more.

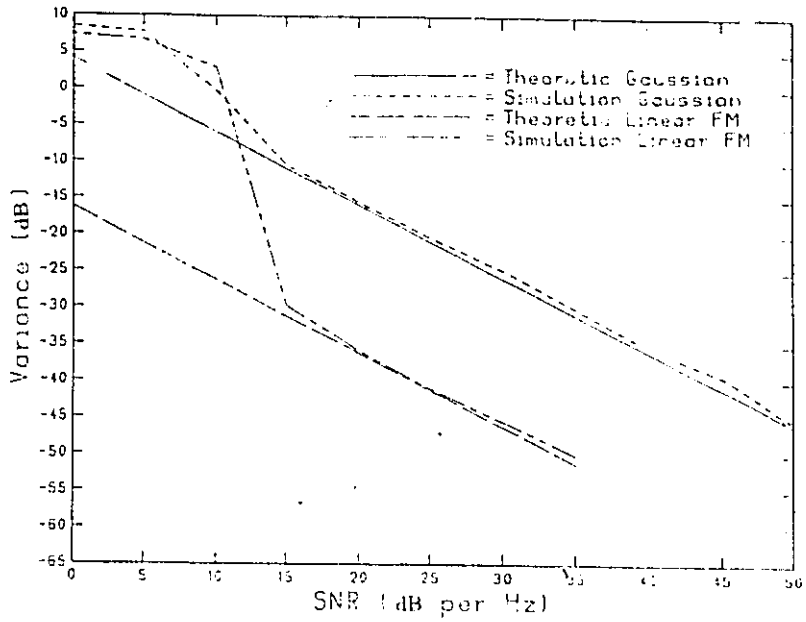


Figure 6.2: Mean Square Error of the Time Delay Estimation verse ρ_{SN} for Two Commonly Used Signals.

Fig.6.3 shows the mean square error of the Doppler shift estimate versus ρ_{SN}. Again, the theoretical performance and simulation performance for both signals are plotted. This time, the linear FM pulse outperforms the Gaussian pulse by a margin of 6dB for the same accuracy when ρ_{SN} is large. Again at $\rho_{SN} \approx 15$dB, thresholds occur for both signals.

6.5.2 Performance Analysis and the Optimum Signal

The reason for the linear FM pulse to yield better performance in the estimation of both the time delay and the Doppler shift can be easily seen from Figs.6.4 and 6.5 which depict the WD for the Gaussian pulse and the linear FM pulse respectively. The WD of the Gaussian pulse in this example exhibits relatively sharply rising peak along the frequency axis whereas the rise and fall of the peak along the time axis is much gentler. As a result the resolution in the frequency axis is higher than that in the time axis which leads to a comparatively lower estimation error in the Doppler shift. On

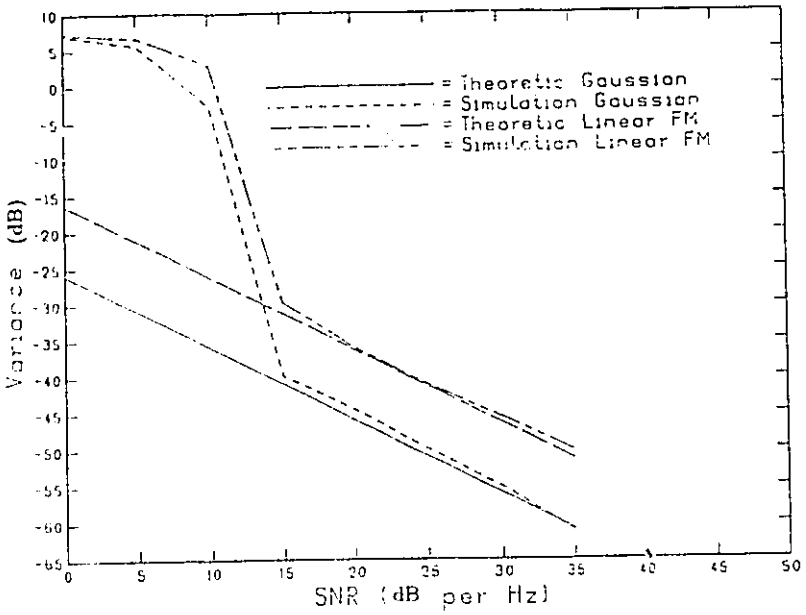


Figure 6.3: Mean Square Error of the Doppler Shift Estimation verse ρ_{SN} for Two Commonly Used Signals.

the other hand, the gradient of the peak of the WD for the linear FM pulse in this example along the frequency axis is not vastly different from that along the time axis. Furthermore, the gradients are sharper than those of the Gaussian pulse. As a result, the estimation errors in both TD and DS are of similar magnitudes, and the combined error $Var(\hat{t}_0 - t_0)Var(\hat{f}_\Delta - f_\Delta)$ is smaller than that for the Gaussian pulse.

The estimation errors for both TD and DS can be adjusted relative to each other by changing the signal parameters. For example, in the case of the Gaussian signal, the denominator in the exponent can be decreased resulting in the sharpness of the gradient along the time axis and the flattening of the gradient along the frequency-axis. Consequently, the estimation error in time delay will decrease at the expense of increasing the estimation error in Doppler shift. Similarly, the instantaneous frequency of the linear FM pulse can also be adjusted so that the estimation error in one parameter can be decreased at the expense of increasing the error in the other parameter. This is a fundamental result of the principle of uncertainty [36], [88] [62]. It can be shown that for

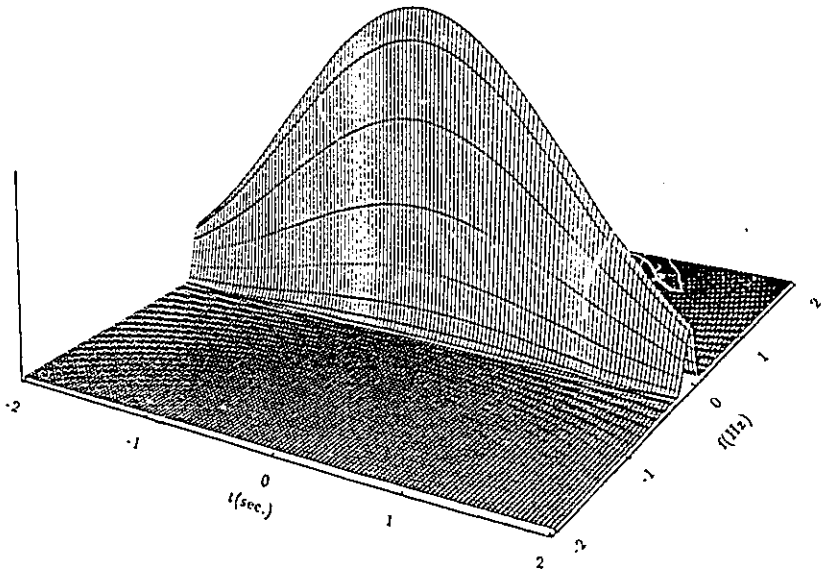


Figure 6.4: The Magnitude of the WD for the Gaussian pulse.

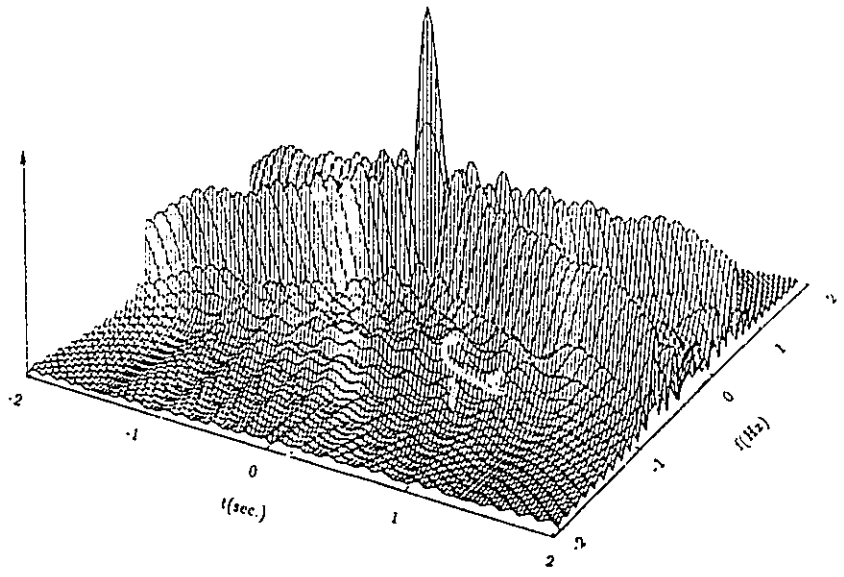


Figure 6.5: The magnitude of the WD for the Linear FM pulse.

the Gaussian pulse, $E[(\hat{t}_0 - t_0)^2] \cdot E[(\hat{f}_\Delta - f_\Delta)^2]$ is larger than any other signal since it has the smallest duration-bandwidth product as discussed in Appendix A. Thus, in general, the linear FM pulse will have a better performance than the Gaussian pulse in the joint estimation of TD and DS as is depicted by this example. However, a word of caution has to be applied here: The magnitude of the WD of the linear FM pulse has one main lobe and many side lobes as is evident from Fig.6.5. Under very noisy circumstances, some sidelobes may shoot over the main lobe causing the false location of a peak and we may have an entirely erroneous estimation of t_0 and f_Δ, and this is the main reason which yields the threshold phenomena. Thus, a further condition in practical applications is

$$\frac{\max[W_a(t,f)]}{W_a(0,0)} \leq \gamma \qquad \begin{matrix} |t| > \Delta t_0, \\ |f| > \Delta f_0 \end{matrix}, \tag{6.87}$$

where Δt_0 and Δf_0 are the mainlobe width of $W_a(t,f)$ in the frequency and time axes respectively, and γ, the ratio of the maximum sidelobe to mainlobe is a constant specified by the allowable probability of erroneous estimation which can be shown (Appendix G) to be given by

$$P_e \approx \frac{1}{2}\mathrm{erfc}(\sqrt{\frac{1-\gamma^2}{2}\rho_{SN}}), \tag{6.88}$$

with

$$\mathrm{erfc}(z) \triangleq \frac{2}{\sqrt{2\pi}} \int_z^\infty e^{-\frac{\zeta^2}{2}} d\zeta. \tag{6.89}$$

Eq.(6.88) assumes that erroneous estimation caused by sidelobes other than the largest one has negligible probability. The probability of erroneous estimation given by Eq.(6.88) can be numerically integrated for different γ at different signal-to-noise ratio (SNR) (ρ_{SN}) which is defined the same as in Eq.(6.84)

Fig.6.6 depicts the relationship between P_e and γ at various SNR. For a prescribed P_e at a certain SNR, γ can be determined.

Side lobe level is also a main factor which will influence the detection performance.

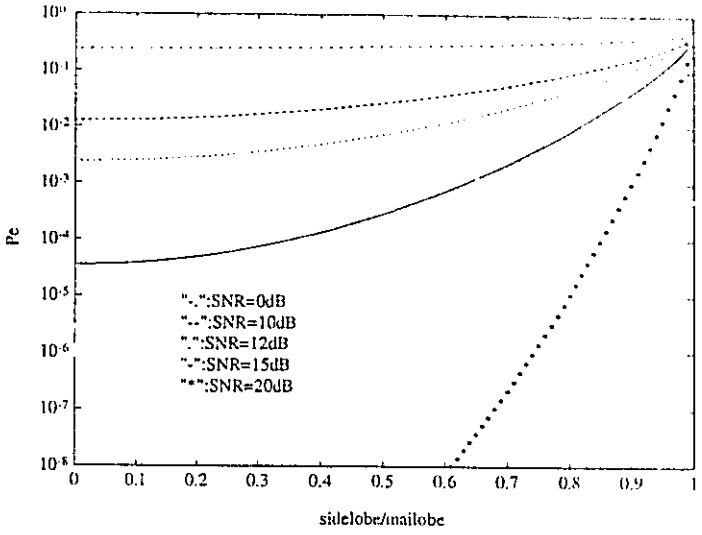


Figure 6.6: The Relationship Between P_e and γ at various SNR.

As mentioned in the last part of Section 6.2, the peak of $W_{xa}(t,f)$ is located firstly by suitably thresholding the output of JTFR. It means that before estimating TD and DS, we must determine whether the peak is due to the signal or due to the noise and side lobe. This forms the detection problem, which is especially important for multi-signal classifications and estimations. In order to have a better understanding of the influence of side lobe on the detection performance, the ROC (Receiver Operation Curve) will be introduced as follows:

Rewriting Eq.(6.22), we have

$$W_{xa}(t,f) = W_a(t,f) + W_{na}(t,f).$$

Under noisy situations, if a threshold is set, false alarm happens when $|W_{xa}(t,f)|$ is larger than the threshold, (where $|f| > \Delta f_0, |t| > \Delta t_0$). A miss happens when $|W_{xa}(0,0)|$ is less than the threshold. With given side lobe level γ and threshold z, the probability of false alarm P_f and the probability of detection P_d can be calculated using the following method:

The ratio $\left|\frac{W_{xa}(t,f)}{W_a(0,0)}\right|^2$ can be written as

$$\left|\frac{W_{xa}(t,f)}{W_a(0,0)}\right|^2 = \left[\frac{W_a(t,f)}{W_a(0,0)} + Re\left\{\frac{W_{na}(t,f)}{W_a(0,0)}\right\}\right]^2 + \left[Im\left\{\frac{W_{na}(t,f)}{W_a(0,0)}\right\}\right]^2.$$

Assuming that the signal has unity energy ($E_a = 1$) and $n(t)$ being defined as in Eq.(6.13) and (6.14), and using the same derivation as in Theorems 6.1 to 6.3, we can easily prove the following relations:

$$E\left[W_{na}(t,f)\right] = E\left[W_{na}^*(t,f)\right] = 0, \tag{6.90}$$

$$E\left|W_{na}(t,f)\right|^2 = 2N_0 W_a(0,0), \tag{6.91}$$

$$E\left[(W_{na}(t,f))^2\right] = 0. \tag{6.92}$$

Using the above relations and Eq.(6.60), we can easily see that

$$Re\left\{\frac{W_{xa}(t,f)}{W_a(0,0)}\right\} = Re\left\{\frac{W_a(t,f)}{W_a(0,0)} + \frac{W_{na}(t,f)}{W_a(0,0)}\right\}$$

is linear related with Gaussian variables $n(t)$ and hence is also Gaussian distribution [35] with mean

$$E\left[Re\left\{\frac{W_a(t,f)}{W_a(0,0)} + \frac{W_{na}(t,f)}{W_a(0,0)}\right\}\right] = Re\left\{\frac{W_a(t,f)}{W_a(0,0)}\right\} = \frac{W_a(t,f)}{W_a(0,0)},$$

and variance

$$\begin{aligned}
& E\left[Re\left\{\frac{W_a(t,f)}{W_a(0,0)} + \frac{W_{na}(t,f)}{W_a(0,0)}\right\} - \left\{\frac{W_a(t,f)}{W_a(0,0)}\right\}\right]^2 \\
= \; & E\left\{\left(Re\left[W_{na}(t,f)\right]\right)^2\right\}\left[W_a(0,0)\right]^{-2} \\
= \; & \frac{2}{4}E\left|W_{na}(t,f)\right|^2\left[W_a(0,0)\right]^{-2} \\
= \; & N_0 W_a^{-1}(0,0) = \frac{N_0}{2},
\end{aligned}$$

where we have assumed that the signal energy is unity, so that

$$W_a(0,0) = \int a^2(\frac{t}{2})dt = 2\int a^2(t)dt = 2.$$

We denote the Gaussian distribution as $Re\left\{\frac{W_{xa}(t,f)}{W_a(0,0)}\right\} \sim \mathcal{N}(\frac{W_a(t,f)}{W_a(0,0)}, \frac{N_0}{2})$. In the same way $Im\left\{\frac{W_{xa}(t,f)}{W_a(0,0)}\right\}$ is also Gaussian distribution and can be noted as $\sim \mathcal{N}(0, \frac{N_0}{2})$.

The covariance function of the real and imaginary parts of $\frac{W_{xa}(t,f)}{W_a(t,f)}$ can be written as

$$\begin{aligned} & E\left[\left(Re\left\{\frac{W_a(t,f)}{W_a(0,0)}+\frac{W_{na}(t,f)}{W_a(0,0)}\right\}-\left\{\frac{W_a(t,f)}{W_a(0,0)}\right\}\right)\left(Im\left\{\frac{W_{na}(t,f)}{W_a(0,0)}\right\}\right)\right] \\ = \; & E\left[\left(Re\left\{\frac{W_{na}(t,f)}{W_a(0,0)}\right\}\right)\left(Im\left\{\frac{W_{na}(t,f)}{W_a(0,0)}\right\}\right)\right]. \end{aligned}$$

Let $F = W_{na}(t,f)$, then using Eqs.(6.92), we have $E[FF] = 0$. We also have $E[FF^*] = E[F^*F]$. Therefore

$$\begin{aligned} & E\left[\left(Re\left\{\frac{W_{na}(t,f)}{W_a(0,0)}\right\}\right)\left(Im\left\{\frac{W_{na}(t,f)}{W_a(0,0)}\right\}\right)\right] \\ = \; & \frac{1}{4j}E[(F+F^*)(F-F^*)] \\ = \; & \frac{1}{4j}E[F\cdot F - F^*\cdot F^* - F\cdot F^* + F^*\cdot F] = 0. \end{aligned}$$

It means that the real part and imaginary part of $\frac{W_{xa}(t,f)}{W_a(0,0)}$ are uncorrelated.

In order to derive the probability of detection P_d, we further define following two variables, at $t = 0$ and $f = 0$,

$$L_c \triangleq Re\left[\frac{W_{xa}(0,0)}{W_a(0,0)}\right] \quad \text{and} \quad L_s \triangleq Im\left[\frac{W_{xa}(0,0)}{W_a(0,0)}\right].$$

With the above derivation, we know that L_c and L_s are of joint Gaussian distribution and their joint probability density function can be expressed as

$$p(L_c, L_s) = \frac{1}{\pi N_0}\exp\left\{-\frac{(L_c-1)^2+L_s^2}{N_0}\right\}. \tag{6.93}$$

Now, with the threshold being z, we can calculate P_d as

$$P_d = \int\int_{\sqrt{L_c^2+L_s^2}\geq z}\frac{1}{\pi N_0}\exp\left\{-\frac{(L_c-1)^2+L_s^2}{N_0}\right\}dL_c dL_s, \tag{6.94}$$

with the substitutions $L_c = r\cos\theta$ $L_s = r\sin\theta$ in Eq.(6.94), we have

$$\begin{aligned} P_d &= \int_z^\infty\int_{-\pi}^{\pi}\frac{r}{\pi N_0}e^{-\frac{1}{N_0}(r^2-2r\cos\theta+1)}d\theta dr \\ &= \int_z^\infty\frac{2r}{N_0}e^{-\frac{r^2+1}{N_0}}I_0(\frac{2r}{N_0})dr, \end{aligned} \tag{6.95}$$

where $I_0(x)$ is zero order modified Bessel function denoted as:

$$I_0(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{x\cos\theta}d\theta. \tag{6.96}$$

In the same way, for the sidelobe level γ as defined in Eq.(6.87), the probability of false alarm is

$$\begin{aligned} P_f &= \int\int_{\sqrt{L_c^2+L_s^2}\geq z} \frac{1}{\pi N_0}\exp\left\{-\frac{(L_c-\gamma)^2+L_s^2}{N_0}\right\}dL_c dL_s \\ &= \int_z^{\infty}\int_{-\pi}^{\pi} \frac{r}{\pi N_0} e^{-\frac{1}{N_0}(r^2-2r\gamma\cos\theta+\gamma^2)}d\theta dr \\ &= \int_z^{\infty} \frac{2r}{N_0} e^{-\frac{r^2+\gamma^2}{N_0}} I_0(\frac{2r\gamma}{N_0})dr. \end{aligned} \tag{6.97}$$

With P_d and P_f given by Eqs.(6.95) and (6.97) respectively, ROC can be obtained, which present the relation between P_f and P_d under given SNR and γ while threshold varies.

Fig.6.7 to fig.6.9 show the ROC with different side lobe to main lobe ratio, at different SNR. The ROC indicates that with the false alarm probability P_f fixed, the lower the side lobe level is, the higher the detection probability P_d. If the side lobe level is zero, the high P_d and the low P_f can be obtained even when SNR (per Hz) = 0dB.

Comparing the two simulation examples given in Section 6.5.1, the WD of the Gaussian pulse is unimodal and the probability of having a spurious peak somewhere other than the main peak is much smaller than linear FM. Hence, we can conclude that a good signal for the joint estimation of TD and DS is the one which has a sharp peak in its WD along both time and frequency axes while having low sidelobes.

In this chapter, we have proposed to use the WD as a JTFR in the estimation of both time delay and Doppler shift. By examining the properties of the WD, we concluded that the joint estimation can be obtained by the location of the peak of the WD. The performance of such a method has been analyzed and has shown to be efficient in the sense that it is unbiased and its variance reaches the Cramér-Rao bound under high SNR. Computer simulations confirm that this is indeed the case. However, under low

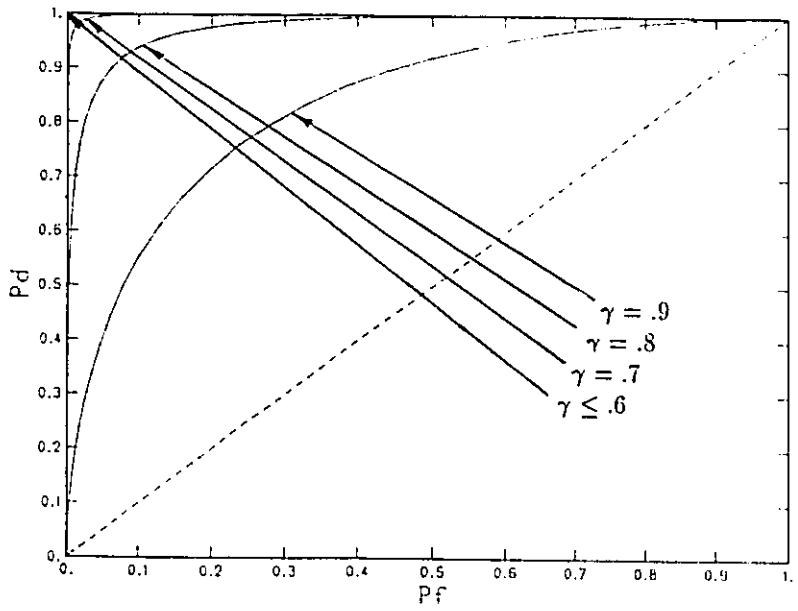


Figure 6.7: The ROC with Different Side Lobe to Main Lobe Ratio, SNR=20dB.

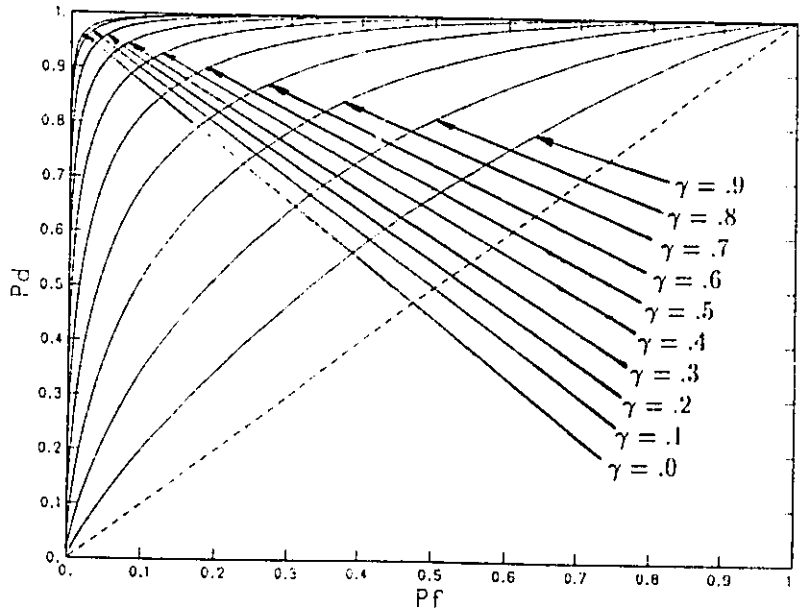


Figure 6.8: The ROC with Different Side Lobe to Main Lobe Ratio, SNR=10dB.

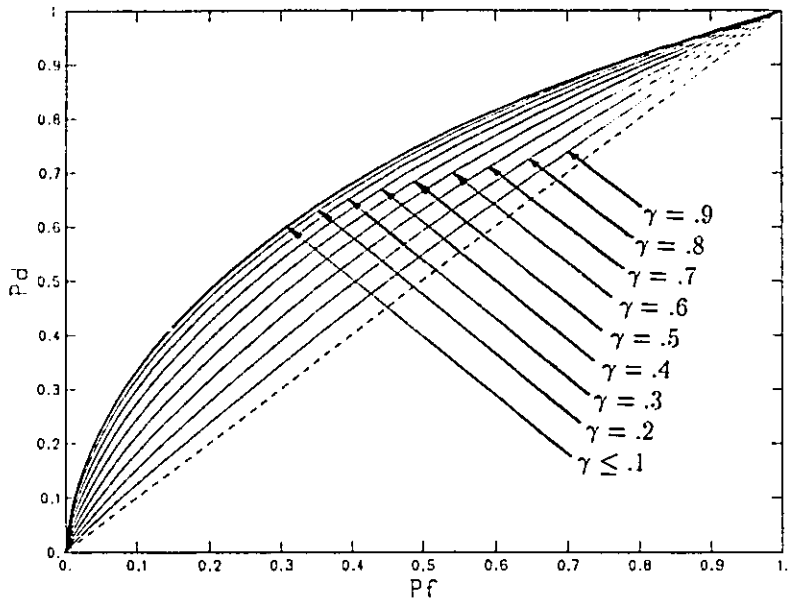


Figure 6.9: The ROC with Different Side Lobe to Main Lobe Ratio, SNR=0dB.

SNR, there exists a threshold below which the performance of the method deteriorates quickly.

For different types of signal, the performance of the joint estimation using the WD differs. This is due to the shape of the WD of the signal on the time-frequency plane. A good signal is one whose WD has a sharp and narrow peak on the time- frequency plane while having very low sidelobes. Naturally, the question "Is there an optimum signal such that the joint estimation error is the lowest?" arises. The answer is "Yes", and the method to design such a signal will be described in the next chapter.

Chapter 7

Optimum Signal Design and Uncertainty Principle

7.1 Selection of the Optimum Basis

7.1.1 Prolate Spheroidal Wave Function and Uncertainty in Fourier Analysis

In the study of communication systems, it is often convenient to expand a given signal in term of a set of basis functions. In this way, any signal can be uniquely identified with the set of coefficients in the expansion. The most commonly used communication signals are essentially time-limited as well as band-limited, because such signals often render the most efficient utilization of time slot and bandwidth. Although in theory one can expand such time- and band-limited signals over any set of basis functions, it is often desirable that the basis functions also have their energy concentrated in the given time duration and bandwidth. The reason is because when we expand a time-and band-limited signal over a set of basis functions which are also time- and band-limited, we usually have few coefficients, thus make signal processing very simple.

In 1961, D.Slepian, H.J.Landau and H.O.Pollak showed, in a series of papers [88] [62],

some important properties of time- and band-limited signals, which will be described next. Given any signal function $h(t)$ in $\mathcal{L}^2$, the space containing all square integrable function in $-\infty \leq t \leq \infty$, and given any time bound T and frequency bound F, the percentage of energy, E_t and E_f, in the designated time duration $(-\frac{T}{2}, \frac{T}{2})$ and bandwidth $(-F, F)$, are defined respectively as

$$E_t \triangleq \frac{\int_{-T/2}^{T/2} |h(t)|^2 dt}{\int_{-\infty}^{\infty} |h(t)|^2 dt}, \quad E_f \triangleq \frac{\int_{-F}^{F} |H(f)|^2 df}{\int_{-\infty}^{\infty} |H(f)|^2 df}, \tag{7.1}$$

where $H(f)$ is the Fourier transform of $h(t)$. Ideally we would like to have both E_t and E_f as large as possible. Unfortunately, this is generally impossible and it was proved [62] that the product $E_t E_f$ is bounded from above and the maximum of $E_t E_f$ is attained by the following function:

$$h(t) = p\psi_0(t) + qD[\psi_0(t)], \tag{7.2}$$

and the equality $E_t = E_f$ is also attained by this function. Here $\psi_0(t)$ is the first of a set of functions $\{\psi_i(t)\}$, called the Prolate Spheroidal Wave (PSW) functions satisfying the following integral equation [88] [62]:

$$\lambda_i \psi_i(t) = \int_{-T/2}^{T/2} \frac{\sin[2\pi F(t-\tau)]}{\pi(t-\tau)} \psi_i(\tau) d\tau, \tag{7.3}$$

with λ_0 being the largest eigenvalue of Eq. (7.3). Also $p = \sqrt{\frac{1-E_t}{1-\lambda_0}}$, $q = \sqrt{\frac{E_t}{\lambda_0}} - p$ and D is the time-limiting operator satisfying:

$$D[\psi_0(t)] = \begin{cases} \psi_0(t), & |t| \leq \frac{T}{2} \\ 0, & |t| > \frac{T}{2} \end{cases}.$$

Let B be the band-limiting operator with:

$$B[h(t)] = \int_{-F}^{F} H(f) e^{-j2\pi ft} df. \tag{7.4}$$

It is known that the PSW functions $\{\psi_i(t)\}$ are band-limited within the frequency range $[-F, F]$, i.e., $B[\psi_0(t)] = \psi_0(t)$. For the convenience of later development, we rewrite

Eq. (7.2) as

$$h(t) = pB[\psi_0(t)] + qD[\psi_0(t)].$$

The above equation implies that the function which maximizes $E_t E_f$ is a linear combination of the band-limited function $\psi_0(t)$ and its time-limited version $D[\psi_0(t)]$.

Many other properties of the prolate spheroidal wave function $\{\psi_i(t)\}$ were discussed in [88], [28], and [91]. Here, we list some of these properties which will be used later.

1. Let $\mathcal{B}$ be the subspace of $\mathcal{L}^2$ consisting of those functions $h(t)$, whose Fourier transform $H(f)$ is band-limited in the range $|f| < F$, i.e., $h(t) = B[h(t)]$. Then $\{\psi_i(t)\}$ are band-limited, orthonormal and complete in $\mathcal{B}$. Thus, we have

$$\psi_i(t) = B[\psi_i(t)] \quad \text{and} \quad \int_{-\infty}^{\infty} \psi_i(t)\psi_j(t)dt = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}. \tag{7.5}$$

2. $\{D[\psi_i(t)]\}$ are orthogonal and complete in $\mathcal{D}$, the space containing all square integrable functions which is time limited in the interval $-\frac{T}{2} \leq t \leq \frac{T}{2}$. In particular, there holds

$$\int_{-\frac{T}{2}}^{\frac{T}{2}} \psi_i(t)\psi_j(t)dt = \begin{cases} 0, & i \neq j \\ \lambda_i, & i = j \end{cases}. \tag{7.6}$$

3. The Fourier Transform of $\psi_i(t)$ is a constant multiple of the Fourier transform of $D[\psi_i(t)]$ at the given frequency band $[-F, F]$. that is,

$$\lambda_i \mathcal{F}[\psi_i(t)] = \mathcal{F}[D\{\psi_i(t)\}], \tag{7.7}$$

where $\mathcal{F}[\cdot]$ represents the Fourier transform.

Proof: Taking Fourier transform of both sides of Eq. (7.3) and using the property of sinc function

$$\text{sinc}(\cdot) = \frac{\sin(\cdot)}{(\cdot)},$$

we have:

$$\begin{aligned}
\lambda_i \mathcal{F}[\psi_i(t)] &= \int_{-\infty}^{\infty} \{\int_{-\frac{T}{2}}^{\frac{T}{2}} (2F)\mathrm{sinc}[2\pi F(t-\tau)]\psi_i(\tau)d\tau\} e^{-j2\pi ft} dt \\
&= \int_{-\frac{T}{2}}^{\frac{T}{2}} \{\int_{-\infty}^{\infty} (2F)\mathrm{sinc}[2\pi F(t-\tau)] e^{-j2\pi ft} dt\}\psi_i(\tau)d\tau \\
&= \int_{-\frac{T}{2}}^{\frac{T}{2}} \psi_i(\tau) H_F(f) e^{-j2\pi f\tau} d\tau \\
&= H_F(f)\mathcal{F}[D\psi_i(t)] \\
&= B\{\mathcal{F}[D\psi_i(t)]\}, \qquad (7.8)
\end{aligned}$$

where $H_F(f)$ is a low pass filter

$$H_F(f) = \begin{cases} 1, & |f| \le F \\ 0, & \text{otherwise.} \end{cases}$$

□

Property 3 is the most interesting one. It means that $\psi_i(t)$ can be uniquely determined from its time-limited version by using Fourier transform, reverse Fourier transform and normalization. Eq. (7.8) also gives the proof of property 1.

4. $\int_{-F}^{F} |\mathcal{F}[D\psi_i(t)]|^2 df = \lambda_i^2$.

Proof: From Eq. (7.7) and using the Parseval's Theorem, we have

$$\begin{aligned}
\int_{-F}^{F} |\mathcal{F}[D\psi_i(t)]|^2 df &= \int_{-F}^{F} \lambda_i^2 |\mathcal{F}[\psi_i(t)]|^2 df \\
&= \lambda_i^2
\end{aligned}$$

where the last step follows from the fact that $\psi_i(t)$ is band-limited and has unit energy. □

Recall that $\mathcal{B}$ denotes a band-limited space and $\mathcal{D}$ denotes a time-limited space. Now we consider the space $\mathcal{B}+\mathcal{D}$, the vector sum of $\mathcal{B}$ and $\mathcal{D}$. Any function $h(t)$ in $\mathcal{B}+\mathcal{D}$ can be written as the sum of a band limited function $b(t) \in \mathcal{B}$ and a time limited function

$d(t) \in \mathcal{D}$, i.e., $h(t) = b(t) + d(t)$. It was shown [89] that the space $\mathcal{B} + \mathcal{D}$ has the following properties:

1. $\mathcal{B} + \mathcal{D}$ is closed under addition.

   Proof: for any $\forall h_1(t),\ h_2(t) \in \mathcal{B} + \mathcal{D}$, we have

   $$h_1(t) = b_1(t) + d_1(t) \text{ and } h_2(t) = b_2(t) + d_2(t),$$

   where $b_1(t),\ b_2(t) \in \mathcal{B}$, and $b_1(t),\ b_2(t) \in \mathcal{D}$. Since $b_1(t) + b_2(t) \in \mathcal{B}$ and $d_1(t) + d_2(t) \in \mathcal{D}$, we obtain

   $$h_1(t) + h_2(t) = [b_1(t) + b_2(t)] + [d_1(t) + d_2(t)] \ \in \ \mathcal{B} + \mathcal{D}.$$

   □

2. If a function $h(t)$ in the space $\mathcal{L}^2$ is orthogonal to the space $\mathcal{B} + \mathcal{D}$, it must satisfy both $B[h(t)] = 0$ and $D[h(t)] = 0$.

   Proof: If $h(t) \in \mathcal{L}^2$ is orthogonal to the space $\mathcal{B} + \mathcal{D}$, it must be orthogonal to both $\mathcal{B}$ and $\mathcal{D}$. By the definition of $\mathcal{B}$ and $\mathcal{D}$, we have $B[h(t)] = D[h(t)] = 0$. □

3. There is no function in $\mathcal{B} + \mathcal{D}$, except 0, which can be orthogonal to both $\mathcal{B}$ and $\mathcal{D}$.

   Proof: Suppose $h(t) = b(t) + d(t) \in \mathcal{B} + \mathcal{D}$ satisfies $D[h(t)] = B[h(t)] = 0$, where $b(t) \in \mathcal{B}$ and $d(t) \in \mathcal{D}$. Then we have

   $$B[b(t)] + B[d(t)] = 0 \ \text{ and } \ D[b(t)] + D[d(t)] = 0.$$

   From the above equation, we obtain

   $$b(t) = -B[d(t)] = BD[b(t)].$$

   Since 1 is not an eigenvalue of the operator BD (see [89]), as a result we have $h(t) = d(t) = -D[b(t)] = 0$. □

In practice, the signal $h(t)$ may not lie exactly in $\mathcal{B}+\mathcal{D}$. However, $h(t)$ can usually be closely approximated by a signal in $\mathcal{B}+\mathcal{D}$. This is because the signals arising from most engineering applications have their energy concentrated in some finite interval and bandwidth. In other word, the energy lying in the complement of $\mathcal{B}+\mathcal{D}$ is negligible.

In the remainder of this section, we will use the properties of PSW discussed in this section to derive a new set of complete orthonormal basis in the space $\mathcal{B}+\mathcal{D}$. This set of basis has high energy concentration on a given time duration and bandwidth.

7.1.2 A New Basis for General Time- and Band-Limited Function

Let $\psi_i(t)$ be the eigenfunction of Eq. (7.3). Consider the class of functions $\rho_i(t) \in \mathcal{B}+\mathcal{D}$ given by

$$\rho_i(t) = p_i B[\psi_i(t)] + q_i D[\psi_i(t)], \qquad i = 0, 1, 2, \cdots,$$

where p_i, q_i are two parameters satisfying the unit energy constraint:

$$\int_{-\infty}^{\infty} \rho_i^2(t)dt = p_i^2 + (q_i^2 + 2p_iq_i)\lambda_i = 1. \tag{7.9}$$

Let the percentage of energy of $\rho_i(t)$, E_{ti} and E_{fi}, in the designated time duration $(-\frac{T}{2}, \frac{T}{2})$ and bandwidth $(-F, F)$, be defined respectively as

$$E_{ti} \triangleq \frac{\int_{-T/2}^{T/2} |\rho_i(t)|^2 dt}{\int_{-\infty}^{\infty} |\rho_i(t)|^2 dt}, \qquad E_{fi} \triangleq \frac{\int_{-F}^{F} |\mathcal{F}[\rho_i(t)]|^2 df}{\int_{-\infty}^{\infty} |\mathcal{F}[\rho_i(t)]|^2 df}.$$

For each i, we would like to maximize $E_{ti}E_{fi}$ over all possible (p_i, q_i), satisfying the constraint in Eq. (7.9). Simple computation reveals that there are two local maximums for $E_{ti}E_{fi}$ at respectively,

$$\begin{aligned}(p_{1i}, q_{1i}) &= \left((2+2\sqrt{\lambda_i})^{-\frac{1}{2}},\ (2\lambda_i + 2\lambda_i\sqrt{\lambda_i})^{-\frac{1}{2}}\right),\\ (p_{2i}, q_{2i}) &= \left((2-2\sqrt{\lambda_i})^{-\frac{1}{2}},\ -(2\lambda_i - 2\lambda_i\sqrt{\lambda_i})^{-\frac{1}{2}}\right),\end{aligned}$$

where λ_i is the eigenvalue of Eq. (7.3) corresponding to the eigenvector $\psi_i(t)$. Let

$$\rho_{1i}(t) = p_{1i}B[\psi_i(t)] + q_{1i}D[\psi_i(t)], \quad \rho_{2i}(t) = p_{2i}B[\psi_i(t)] + q_{2i}D[\psi_i(t)]. \tag{7.10}$$

Then the respective values of E_{ti} and E_{fi} at $\rho_{1i}(t)$, $\rho_{2i}(t)$ are

$$E_{t1i} = E_{f1i} = \frac{1+\sqrt{\lambda_i}}{2} \quad \text{and} \quad E_{t2i} = E_{f2i} = \frac{1-\sqrt{\lambda_i}}{2}. \tag{7.11}$$

Recall that the function given by Eq. (7.2) maximizes $E_t E_f$ with given T and F. Using Eq. (7.11), it is not hard to see that this function is a special case of Eq. (7.10) for $i = 0$, i.e., $h(t) = \rho_{10}(t)$. In the remainder of this section, we will show that $\{\rho_{1i}(t), \rho_{2i}(t)\}_{i=0}^{\infty}$ forms a complete basis in $\mathcal{B}+\mathcal{D}$. To do this, we use variational principle to the optimality condition of maximizing $E_t E_f$. The solution $h(t)$ can be regarded as the eigenvectors of certain integral equation.

Theorem 7.1 *For given T and F, the function $h(t)$ which maximizes $E_t E_f$ is the eigenvector, corresponding to the certain eigenvalue κ, of the following integral equation:*

$$\int_{-\infty}^{\infty} h(\tau) K_F(t,\tau) d\tau = \kappa h(t), \tag{7.12}$$

where

$$K_F(t,\tau) = \begin{cases} 4F\text{sinc}[2\pi F(t-\tau)], & |t|, |\tau| \le \frac{T}{2} \\ 2F\text{sinc}[2\pi F(t-\tau)], & |t| \le \frac{T}{2}, |\tau| > \frac{T}{2} \text{ or } |\tau| \le \frac{T}{2}, |t| > \frac{T}{2} \\ 0, & |t|, |\tau| > \frac{T}{2} \end{cases}. \tag{7.13}$$

Proof: Assume that the optimal solution $h(t)$, which maximizes $E_t E_f$, has been normalized

$$\int_{-\infty}^{\infty} |h(t)|^2 dt = 1, \tag{7.14}$$

which is equivalent to

$$\int_{-\infty}^{\infty} |H(f)|^2 df = 1, \tag{7.15}$$

where $H(f)$ is the Fourier transform of $h(t)$. The maximization of $E_t E_f$ is equivalent to maximizing

$$Y = \log(E_t E_f) = \log E_t + \log E_f, \tag{7.16}$$

since $E_t \geq 0, \quad E_f \geq 0$. Let $h(t,\varepsilon) = h(t) + \varepsilon\mu(t)$, then $H(f,\varepsilon) = H(f) + \varepsilon\mathcal{M}(f)$, where $\varepsilon\mu(t)$ represents a small perturbation of the optimum solution $h(t)$, ε is a small number, $\mu(t)$ is any square integrable function, and $H(f)$ and $\mathcal{M}(f)$ represent the Fourier transforms of $h(t)$ and $\mu(t)$ respectively.

Since $h(t)$ maximizes Y, we have $\frac{dY}{d\varepsilon}\Big|_{\varepsilon=0} = 0$, for all $\mu(t)$ and substituting Eq. (7.16) into it, we have

$$\frac{dY}{d\varepsilon}\Big|_{\varepsilon=0} = \left(E_t^{-1}(\varepsilon)\frac{dE_t(\varepsilon)}{d\varepsilon}\right)\Big|_{\varepsilon=0} + \left(E_f^{-1}(\varepsilon)\frac{dE_f(\varepsilon)}{d\varepsilon}\right)\Big|_{\varepsilon=0} = 0, \tag{7.17}$$

where

$$E_t(\varepsilon) = \frac{\int_{-T/2}^{T/2}|h(t)+\varepsilon\mu(t)|^2dt}{\int_{-\infty}^{\infty}|h(t)+\varepsilon\mu(t)|^2dt}, \tag{7.18}$$

$$E_f(\varepsilon) = \frac{\int_{-F}^{F}|H(f)+\varepsilon\mathcal{M}(f)|^2df}{\int_{-\infty}^{\infty}|H(f)+\varepsilon\mathcal{M}(f)|^2df}. \tag{7.19}$$

Hence

$$\frac{dE_t(\varepsilon)}{d\varepsilon}\Big|_{\varepsilon=0} = \frac{2\int_{-T/2}^{T/2}Re[h(t)\mu^*(t)]dt}{\int_{-\infty}^{\infty}|h(t)|^2dt} - 2E_t(0)\frac{\int_{-\infty}^{\infty}Re[h(t)\mu^*(t)]dt}{\int_{-\infty}^{\infty}|h(t)|^2dt}. \tag{7.20}$$

Substituting Eqs. (7.18) and (7.20) into the first term of Eq. (7.17), we have

$$\left(E_t^{-1}(\varepsilon)\,\frac{dE_t(\varepsilon)}{d\varepsilon}\right)\Big|_{\varepsilon=0} = \frac{2\int_{-T/2}^{T/2}Re[h(t)\mu^*(t)]dt}{\int_{-T/2}^{T/2}|h(t)|^2dt} - \frac{2\int_{-\infty}^{\infty}Re[h(t)\mu^*(t)]dt}{\int_{-\infty}^{\infty}|h(t)|^2dt}. \tag{7.21}$$

Similarly, we can obtain the second term of Eq. (7.17) as:

$$E_f^{-1}(\varepsilon)\left(\frac{dE_f(\varepsilon)}{d\varepsilon}\right)\Big|_{\varepsilon=0} = \frac{2\int_{-F}^{F}Re[H(f)\mathcal{M}^*(f)]df}{\int_{-F}^{F}|H(f)|^2df} - \frac{2\int_{-\infty}^{\infty}Re[H(f)\mathcal{M}^*(f)]df}{\int_{-\infty}^{\infty}|H(f)|^2df}. \tag{7.22}$$

Using the Parseval theorem $\int_{-\infty}^{\infty}h(t)\mu^*(t)dt = \int_{-\infty}^{\infty}H(f)\mathcal{M}^*(f)df$ and Eqs. (7.14), (7.15), we can rewrite Eq. (7.17) as

$$\begin{aligned}\frac{dY}{d\varepsilon}\Big|_{\varepsilon=0} &= \frac{2\int_{-T/2}^{T/2}Re[h(t)\mu^*(t)]dt}{\int_{-T/2}^{T/2}|h(t)|^2dt} + \frac{2\int_{-F}^{F}Re[H(f)\mathcal{M}^*(f)]df}{\int_{-F}^{F}|H(f)|^2df} \\ &\quad -4\int_{-\infty}^{\infty}Re[h(t)\mu^*(t)]dt \\ &= 0.\end{aligned} \tag{7.23}$$

Substituting $\int_{-F}^{F} H(f)\mathcal{M}^*(f)df = \int_{-\infty}^{\infty} \mu^*(t)\{\int_{-F}^{F} H(f)e^{j2\pi ft}df\}dt$ into Eq. (7.23) and noticing that $h(t)$ is normalized, we obtain

$$Re\{E_t^{-1}(0)\int_{-T/2}^{T/2} h(t)\mu^*(t)dt + E_f^{-1}(0)\int_{-\infty}^{\infty}\mu^*(t)\{\int_{-F}^{F} H(f)e^{j2\pi ft}df\}dt - 2\int_{-\infty}^{\infty} h(t)\mu^*(t)dt\} = 0. \tag{7.24}$$

Since Eq. (7.24) holds for any $\mu(t)$, we have

$$E_t^{-1}(0)h(t) + E_f^{-1}(0)\int_{-F}^{F} H(f)e^{j2\pi ft}df - 2h(t) = 0 \quad \text{for } |t| \le \tfrac{T}{2}, \tag{7.25}$$

$$E_f^{-1}(0)\int_{-F}^{F} H(f)e^{j2\pi ft}df - 2h(t) = 0 \quad \text{for } |t| > \tfrac{T}{2}. \tag{7.26}$$

So far, we have shown that any function maximizing $E_t E_f$ must be a solution of Eqs. (7.25) and (7.26). We next modify Eqs. (7.25) and (7.26) to show that $h(t)$ satisfies Eq. (7.12).

Taking the Fourier transform of both sides of the combined equations, Eqs. (7.25) and (7.26), for $-\infty < t < \infty$, we have

$$E_t^{-1}(0)\int_{-\frac{T}{2}}^{\frac{T}{2}} h(t)e^{-j2\pi ft}dt + E_f^{-1}(0)\int_{-\infty}^{\infty}[\int_{-F}^{F} H(f_1)e^{j2\pi f_1 t}df_1]e^{-j2\pi ft}dt - 2\int_{-\infty}^{\infty} h(t)e^{-j2\pi ft}dt = 0. \tag{7.27}$$

Notice that the first term of Eq. (7.27) is equal to $E_t^{-1}(0)\mathcal{F}[Dh(t)]$, and the last term is just $2H(f)$, where D is the time-limiting operator. Exchanging the order of integration in the second term of Eq. (7.27), we have

$$E_f^{-1}(0)\int_{-F}^{F}\left[\int_{-\infty}^{\infty} e^{-j2\pi(f-f_1)t}df\right]H(f_1)df_1 = E_f^{-1}(0)\int_{-F}^{F} H(f_1)\delta(f-f_1)df_1 = E_f^{-1}(0)B[H(f)].$$

Thus, we have

$$E_t^{-1}(0)\mathcal{F}[Dh(t)] = 2H(f) - E_f^{-1}(0)B[H(f)]$$

or equivalent

$$H(f) = (2E_t(0) - \frac{E_t(0)}{E_f(0)})^{-1}\mathcal{F}[Dh(t)], \qquad |f| \le F. \tag{7.28}$$

Now, multiplying both sides of Eq. (7.25) with $h^*(t)$ and integrating in the interval of $(-\frac{T}{2}, \frac{T}{2})$, we have

$$1 + E_f^{-1}(0) \int_{-\frac{T}{2}}^{\frac{T}{2}} \int_{-F}^{F} h^*(t) H(f) e^{j2\pi ft} df dt - 2E_t(0) = 0,$$

or

$$\int_{-\frac{T}{2}}^{\frac{T}{2}} \int_{-F}^{F} h^*(t) H(f) e^{j2\pi ft} df dt = E_f(0)(2E_t(0) - 1). \tag{7.29}$$

By an identical argument, we can obtain the formula similar to Eqs. (7.28) and (7.29) in the time domain:

$$h(t) = (2E_f(0) - \frac{E_f(0)}{E_t(0)})^{-1} \mathcal{F}^{-1}[BH(f)], \quad |t| \leq \frac{T}{2}, \tag{7.30}$$

where $\mathcal{F}^{-1}(\cdot)$ represents the inverse Fourier transform, and

$$\int_{-\frac{T}{2}}^{\frac{T}{2}} \int_{-F}^{F} h(t) H^*(f) e^{j2\pi ft} df dt = E_t(0)(2E_f(0) - 1). \tag{7.31}$$

Comparing the left hand side of Eqs. (7.29) with that of Eq. (7.31), we find that they are complex conjugates of each other. Since the right hand sides of both equations are real, so they are equal:

$$E_t(0)(2E_f(0) - 1) = E_f(0)(2E_t(0) - 1).$$

Hence, we have

$$E_t(0) = E_f(0). \tag{7.32}$$

Substituting Eq. (7.32) into Eq. (7.28), taking inverse Fourier transform of both sides of Eq. (7.28) and multiplying with $(2E_t(0) - 1)$, we have

$$\begin{aligned}
(2E_t(0) - 1) \int_{-F}^{F} H(f) e^{j2\pi ft} df &= \int_{-F}^{F} \mathcal{F}[Dh(t)] e^{j2\pi ft} df \\
&= \int_{-F}^{F} \left[\int_{-\frac{T}{2}}^{\frac{T}{2}} h(t) e^{-j2\pi ft_1} dt_1 \right] e^{j2\pi ft} df \\
&= \int_{-\frac{T}{2}}^{\frac{T}{2}} h(t_1) \{2F \text{sinc}[2\pi F(t - t_1)]\} dt_1. \qquad (7.33)
\end{aligned}$$

Combining Eq. (7.33) with Eq. (7.26) and using Eq. (7.32), we have, for $|t| > \frac{T}{2}$

$$\int_{-\frac{T}{2}}^{\frac{T}{2}} h(t_1)\{2F\text{sinc}[2\pi F(t-t_1)]\}dt_1 = 2E_t(0)(2E_t(0)-1)h(t). \tag{7.34}$$

As for $|t| \leq \frac{T}{2}$, the second term in Eq. (7.25) is

$$E_f^{-1}(0)\int_{-F}^{F}\left[\int_{-\infty}^{\infty} h(t_1)e^{-j2\pi f t_1}dt_1\right]e^{j2\pi ft}df = E_t^{-1}(0)\int_{-\infty}^{\infty} h(t_1)\{2F\text{sinc}[2\pi F(t-t_1)]\}dt_1.$$

Hence Eq. (7.25) becomes

$$\int_{-\infty}^{\infty} h(t_1)\{2F\text{sinc}[2\pi F(t-t_1)]\}dt_1 = (2E_t(0)-1)h(t), \text{ for } |t| \leq \tfrac{T}{2}. \tag{7.35}$$

The left hand side of Eq. (7.33) is just $(2E_t(0)-1)\mathcal{F}^{-1}[BH(f)]$. Using Eq. (7.30), we have

$$\int_{-\frac{T}{2}}^{\frac{T}{2}} h(t_1)\{2F\text{sinc}[2F\pi(t-t_1)]\}dt_1 = (2E_t(0)-1)^2h(t), \quad |t| \leq \frac{T}{2}. \tag{7.36}$$

Adding Eq. (7.35) to Eq. (7.36) yields

$$\begin{aligned}&\int_{-\frac{T}{2}}^{\frac{T}{2}} h(t)\{2F\text{sinc}[2\pi F(t-t_1)]\}dt_1 + \int_{-\infty}^{\infty} h(t_1)\{2F\text{sinc}[2\pi f(t-t_1)]\}dt_1\\ =\ & 2E_t(0)(2E_t(0)-1)h(t), \qquad |t| \leq \frac{T}{2}.\end{aligned}$$

This together with (7.34) implies Eq. (7.12) and $\kappa = 2E_t(0)(2E_t(0)-1)$. □

Theorem 7.1 shows that every function maximizing E_tE_f must satisfy Eq. (7.12). Next, we shall determine all the solutions of Eq. (7.12) and prove that all these solutions form a complete basis. To begin, we first state a result from functional analysis [81] and [95].

Theorem 7.2 *Let $J(t,\tau)$ be a symmetric and square integrable function, i.e., $J(t,\tau) = J(\tau,t)$, $\forall(t,\tau) \in (-\infty,+\infty)^2$, and $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} J^2(t,\tau)dtd\tau < +\infty$. Consider the integral operator given by*

$$\mathcal{J}[h(t)] = \int_{-\infty}^{\infty} h(\tau)J(t,\tau)d\tau$$

which maps $\mathcal{B}+\mathcal{D} \to \mathcal{B}+\mathcal{D}$. Then $\mathcal{J}$ has at most countable number of eigenvalues, and the eigenvectors corresponding to nonzero eigenvalues form an orthonormal basis of the space $\mathcal{B}+\mathcal{D}$ if and only if 0 is not an eigenvalue of $\mathcal{J}$ in that space.

Proof: See page 103 in [81]. □

Consider the integral kernel $K_F(t,\tau)$ given by Eq. (7.13). We can prove that it has the following properties

1. $K_F(t,\tau)$ is symmetric, i.e. $K_F(t,\tau) = K_F(\tau,t)$.

   Proof: Since the function $\mathrm{sinc}(\cdot)$ is even symmetric, it can be easily extended that $K_F(t,\tau)$ is a symmetric function. □

2. $K_F(t,\tau)$ is square integrable, i.e.,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |K_F(t,\tau)|^2 dt d\tau < \infty \tag{7.37}$$

   Proof: Separating the above integration into two parts and using the symmetric property of $K_F(t,\tau)$, we have

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |K_F(t,\tau)|^2 dt d\tau = \int_{-T}^{T}\int_{-T}^{T} |K_F(t,\tau)|^2 dt d\tau + 4\int_{-\infty}^{\infty}\int_{T}^{\infty} |K_F(t,\tau)|^2 dt d\tau. \tag{7.38}$$

   Since $K_F(t,\tau)$ is bounded over $[-T,T]^2$, the first term of Eq. (7.38) is finite. By the definition of $K_F(t,\tau)$ in Eq. (7.13) and the fact that $K_F(t,\tau) = 0$ for all (t,τ) when both $|t| > \frac{T}{2}$ and $|\tau| > \frac{T}{2}$ are satisfied, the second term of Eq. (7.38) can be written as

$$\begin{aligned} 4\int_{-\infty}^{\infty}\int_{T}^{\infty} |K_F(t,\tau)|^2 dt d\tau &= 16F^2 \int_{-\frac{T}{2}}^{\frac{T}{2}}\int_{T}^{\infty} |\mathrm{sinc}[2\pi F(t-\tau)]|^2 dt d\tau \\ &\le \frac{4}{\pi^2}\int_{-\frac{T}{2}}^{\frac{T}{2}}\int_{T}^{\infty} \frac{1}{(t-\tau)^2} dt d\tau \\ &= \frac{4}{\pi^2}\int_{-\frac{T}{2}}^{\frac{T}{2}} \frac{1}{(T-\tau)} d\tau = \frac{4}{\pi^2}\ln 3 < \infty, \end{aligned}$$

where the second step follow from $|\text{sinc}(x)| = |\frac{\sin(x)}{x}| \leq \frac{1}{|x|}$. Since both terms in the Eq. (7.38) are finite, the sum is finite as well. This completes the proof of Eq. (7.37). □

3. The operator $\mathcal{J}[h(t)] = \int_{-\infty}^{\infty} h(\tau)K_F(t,\tau)d\tau$ maps $\mathcal{B}+\mathcal{D}$ to $\mathcal{B}+\mathcal{D}$.

   Proof: Given $h(t) \in \mathcal{B}+\mathcal{D}$ and using Eq. (7.12), we have, for $|t| \leq \frac{T}{2}$,

$$\begin{aligned}\mathcal{J}[h(t)] &= \int_{-\frac{T}{2}}^{\frac{T}{2}} h(t)2F\text{sinc}[2\pi F(t-\tau)]d\tau + \int_{-\infty}^{\infty} h(t)2F\text{sinc}[2\pi F(t-\tau)]d\tau \\ &= \int_{-\infty}^{\infty} \{D[h(t)]+h(t)\}2F\text{sinc}[2\pi F(t-\tau)]d\tau.\end{aligned}$$

   Since $2F\text{sinc}[2\pi Ft]$ is the transfer function of the ideal low pass filter with frequency band $[-F,F]$, $2F\text{sinc}[2\pi Ft]$ convoluting with any function is equivalent to the band-limiting operator $B[\cdot]$ onto that function. Examining the right hand side of the above equation, we find that it is just the convolution of $D[h(t)]+h(t)$ and $2F\text{sinc}[2\pi Ft]$. Therefore The above equation can be written as

$$\mathcal{J}[h(t)] = B\{D[h(t)]+[h(t)]\}, \quad |t| \leq \frac{T}{2}.$$

   Similarly, we have

$$\mathcal{J}[h(t)] = \int_{-\frac{T}{2}}^{\frac{T}{2}} [h(t)]2F\text{sinc}[2\pi F(t-\tau)]d\tau = BD[h(t)], \quad |t| > \frac{T}{2}.$$

   Combining the results for both $|t| \leq \frac{T}{2}$ and $|t| > \frac{T}{2}$, we have

$$\mathcal{J}[h(t)] = BD[h(t)] + DB[h(t)] \in \mathcal{B}+\mathcal{D},$$

   since $BD[h(t)] \in \mathcal{B}$ and $DB[h(t)] \in \mathcal{D}$. □

   It can also be concluded from the above proof that the operator $\mathcal{J}$ is also a mapping of $\mathcal{L}^2$ to $\mathcal{B}+\mathcal{D}$

4. 0 is not an eigenvalue of the operator $\mathcal{J}$ in the space $\mathcal{B}+\mathcal{D}$.

Proof: Assume that $h(t) \in \mathcal{B} + \mathcal{D}$ is an eigenvector of $\mathcal{J}$ corresponding to the eigenvalue zero. Then we have $\mathcal{J}[h(t)] = 0$. Using the similar procedures as in the proof of property 3, we can rewrite Eq. (7.12) for $|t| > \frac{T}{2}$ as

$$\int_{-\frac{T}{2}}^{\frac{T}{2}} h(t_1)\{2F\text{sinc}[2\pi F(t - t_1)]\}dt_1 = B[D(h(t))] = 0. \quad |t| > \frac{T}{2}.$$

This means that $B[D(h(t))]$ is time-limited. There is a well known property in Fourier analysis [73] that there does not exist any non-zero function which can be both time-limited and band-limited. The only possible solution of the above equation is $B[D(h(t))] = 0$ for all t. By the same token, there does not exist any non-zero time-limited function which is band-limited, so we must have $D[h(t)] = 0$. Using this result and Eq. (7.12) for $|t| \leq \frac{T}{2}$, we have

$$B[h(t)] = \int_{-\infty}^{\infty} h(t_1)\{2F\text{sinc}[2\pi F(t - t_1)]\}dt_1 = 0 \quad |t| \leq \frac{T}{2}.$$

Since the above equation holds for $|t| \leq \frac{T}{2}$, we have $D[B(h(t))] = 0$. By the same argument above, we have $B[h(t)] = 0$. So far, we have shown that each eigenvector $h(t)$ of $\mathcal{J}$ corresponding to the eigenvalue zero must satisfy both $B[h(t)] = 0$ and $D[h(t)] = 0$. Using the properties of space $\mathcal{B} + \mathcal{D}$ discussed in section 7.1.1, we have that there is no function except 0 in $\mathcal{B} + \mathcal{D}$ which can simultaneously satisfy $B[h(t)] = 0$ and $D[h(t)] = 0$. Therefore, 0 is not an eigenvalue of the operator $\mathcal{J}$ in the space $\mathcal{B} + \mathcal{D}$. □

5. If $h(t)$ satisfies $D[h(t)] = B[h(t)] = 0$, then it can not be the eigenvector of J corresponding to the nonzero eigenvalue.

Proof: Assume that $h(t)$ is the eigenvector of $\mathcal{J}$ and it satisfies $D[h(t)] = B[h(t)] = 0$. Using the similar procedure as in the proof of property 3, we have

$$\mathcal{J}[h(t)] = BD[h(t)] + B[h(t)], \quad |t| \leq \frac{T}{2},$$

$$\mathcal{J}[h(t)] = BD[h(t)], \qquad |t| \le \frac{T}{2}.$$

Under the assumption that $D[h(t)] = B[h(t)] = 0$, we have $\mathcal{J}[h(t)] = 0$. If $h(t) \neq 0$, then the corresponding eigenvalue must be zero. □

The above properties of $K_F(t,\tau)$ show that the integral kernel of Eq. (7.12) satisfies the condition of Theorem 7.2. Therefore, all the eigenvectors of $\mathcal{J}$ which are the solutions of Eq. (7.12) are complete in the space of $\mathcal{B}+\mathcal{D}$ and any functions in $\mathcal{B}+\mathcal{D}$ can be expanded onto this set of eigenvectors.

Next, we develop analytic formulas for the eigenvectors of $\mathcal{J}$.

Theorem 7.3 *Given time duration T and bandwidth F, let*

$$\rho_{1i}(t) = p_{1i}B[\psi_i(t)] + q_{1i}D[\psi_i(t)],\ p_{1i} = (2+2\sqrt{\lambda_i})^{-\frac{1}{2}}, q_{1i} = (2\lambda_i + 2\lambda_i\sqrt{\lambda_i})^{-\frac{1}{2}}, \quad (7.39)$$

$$\rho_{2i}(t) = p_{2i}B[\psi_i(t)] + q_{2i}D[\psi_i(t)],\ p_{2i} = (2-2\sqrt{\lambda_i})^{-\frac{1}{2}}, q_{2i} = -(2\lambda_i - 2\lambda_i\sqrt{\lambda_i})^{-\frac{1}{2}}, \quad (7.40)$$

where $i = 0,1,2,\cdots$ and each $\psi_i(t)$ is the eigenvector of Eq. (7.3) corresponding to the ith eigenvalue λ_i. Then, $\{\rho_{1i},\rho_{2i}\}$ are the eigenvectors of the operator $\mathcal{J}$ and they form a complete orthonormal basis in $\mathcal{B}+\mathcal{D}$. The eigenvalues corresponding to $\rho_{1i}(t)$ and $\rho_{2i}(t)$ are given respectively by

$$\kappa_{1i} = 2E_{t1i}(2E_{t1i}-1) \qquad \text{and} \qquad \kappa_{2i} = 2E_{t2i}(2E_{t2i}-1). \quad (7.41)$$

To prove Theorem 7.3, we need the following lemma.

Lemma 7.1 *For the parameters given in Theorem 7.3, we have the following relations*

$$\frac{q_{1i}}{p_{1i}} = -\frac{q_{2i}}{p_{2i}}, \quad (7.42)$$

$$\frac{(p_{1i}+q_{1i})\lambda_i}{p_{1i}} = \frac{p_{1i}+q_{1i}\lambda_i}{q_{1i}} = 2E_{t1i}(2E_{t1i}-1), \quad (7.43)$$

$$\int_{-\infty}^{\infty} D[\rho_{ki}(\tau)]2F\mathrm{sinc}[2\pi F(t-\tau)]d\tau = (p_{ki}+q_{ki})\lambda_i\psi_i(t), \quad (7.44)$$

$$\int_{-\infty}^{\infty} \rho_{ki}(\tau) 2F\text{sinc}[2\pi F(t-\tau)]d\tau = (p_{ki} + q_{ki}\lambda_i)\psi_i(t), \tag{7.45}$$

where $k = 1, 2$, $i = 0, 1, 2, \cdots$.

Proof: Eq. (7.42) can be obtained by directly using Eqs. (7.39) and (7.40).

From Eq. (7.11), it is easily obtained that

$$E_{t1i} = E_{f1i} \quad \text{and} \quad \sqrt{\lambda_i} = 2E_{t1i} - 1.$$

With p_{1i}, q_{1i} given by Eq. (7.39) and λ_i above, we have

$$p_{1i} = \frac{1}{2\sqrt{E_{t1i}}}, \quad \text{and} \quad q_{1i} = \frac{1}{2\sqrt{E_{ti}}(2E_{ti} - 1)}.$$

Using simple algebra, we can derive Eq. (7.43).

Now we consider $\rho_{ki}(t), k = 1, 2$. Since $\psi_i(t)$ is a band-limited function, we have $D[\rho_{ki}(t)] = p_{ki}D[B(\psi_i(t))] + q_{ki}D[\psi_i(t)] = (p_{ki} + q_{ki})D[\psi_i(t)]$. Using Eq. (7.3), i.e., $\int_{-\infty}^{\infty} D[\psi_i(\tau)]2F\text{sinc}[2\pi F(t-\tau)]d\tau = \lambda_i\psi_i(t)$, we have

$$\begin{aligned}\int_{-\infty}^{\infty} D[\rho_{ki}(\tau)]2F\text{sinc}[2\pi F(t-\tau)]d\tau &= \int_{-T/2}^{T/2}(p_{ki} + q_{ki})\psi_i(\tau)2F\text{sinc}[2\pi F(t-\tau)]d\tau \\ &= (p_{ki} + q_{ki})\lambda_i\psi_i(t),\end{aligned}$$

which proves Eq. (7.44). Also, since $\psi_i(t)$ is band-limited so that

$$\int_{-\infty}^{\infty} \psi_i(\tau)2F\text{sinc}[2\pi F(t-\tau)]d\tau = \psi_i(t),$$

we have

$$\begin{aligned}&\int_{-\infty}^{\infty} \rho_{ki}(\tau)2F\text{sinc}[2\pi F(t-\tau)]d\tau \\ =\ &\int_{-\infty}^{\infty}[p_{ki}\psi_i(\tau) + q_{ki}D\psi_i(\tau)]2F\text{sinc}[2\pi F(t-\tau)]d\tau \\ =\ &(p_{ki} + q_{ki}\lambda_i)\psi_i(t),\end{aligned}$$

which proves Eq. (7.45). □

Now we are ready to prove theorem 7.2.

Proof: Our proof can be divided into two parts. We first prove that $\rho_{ki}(t), k = 1, 2$ are the solutions of Eq. (7.12) and then we prove that they are complete in $\mathcal{B} + \mathcal{D}$.

1. Substituting $\rho_{ki}(t)$ into the left side of Eq. (7.12), and using Eqs. (7.44) and (7.45), we have that when $|t| \leq \frac{T}{2}$,

$$\begin{aligned} & \int_{-\frac{T}{2}}^{\frac{T}{2}} \rho_{ki}(\tau) 2F\mathrm{sinc}[2\pi F(t-\tau)]d\tau + \int_{-\infty}^{\infty} \rho_{ki}(\tau) 2F\mathrm{sinc}[2\pi F(t-\tau)]d\tau \\ = \; & (p_{ki} + q_{ki})\lambda_i\psi_i(t) + (p_{ki} + q_{ki}\lambda_i)\psi_i(t). \end{aligned} \tag{7.46}$$

For $|t| > \frac{T}{2}$, we obtain

$$\int_{-\frac{T}{2}}^{\frac{T}{2}} \rho_{ki}(\tau) 2F\mathrm{sinc}[2\pi F(t-\tau)]d\tau = (p_{ki} + q_{ki})\lambda_i\psi_i(t). \tag{7.47}$$

We now prove that $\rho_{1i}(t)$ is an eigenvector of $\mathcal{J}$. Using the definition of $K_F(t,\tau)$, and noticing that the first term of right hand side of Eq. (7.46) is the same as the right hand side of Eq. (7.47), we obtain:

$$\int_{-\infty}^{\infty} \rho_{1i}(\tau) K_F(t,\tau)d\tau = (p_{1i} + q_{1i})\lambda_i\psi_i(t) + (p_{1i} + q_{1i}\lambda_i)D\psi_i(t)$$

Substituting Eq. (7.43) into the above equation, we have

$$\begin{aligned} \int_{-\infty}^{\infty} \rho_{1i}(\tau) K_F(t,\tau)d\tau &= 2E_{t1i}(2E_{t1i} - 1)[p_{1i}\psi_i(t) + q_{1i}D\psi_i(t)] \\ &= 2E_{t1i}(2E_{t1i} - 1)\rho_{1i}(t). \end{aligned} \tag{7.48}$$

Therefore each $\rho_{1i}(t)$ is an eigenvector of Eq. (7.12) corresponding to the eigenvalue $\kappa_{1i} = 2E_{t1i}(2E_{t1i} - 1)$. Similarly, we can prove that each $\rho_{2i}(t)$ is also the eigenvectors of Eq. (7.12) corresponding to the eigenvalue $\kappa_{2i} = 2E_{t2i}(2E_{t2i} - 1)$. Moreover, Eqs.(7.5) and (7.6) imply that, for any i, both $\psi_i(t)$ and $D[\psi_i(t)]$ are orthogonal to $\psi_n(t)$ and $D[\psi_n(t)]$ for $i \neq n$. This, together with the Eqs.(7.39) and (7.40) for the definition of $\{\rho_{1i}, \rho_{2i}\}$, implies that $\rho_{ki}(t)$ are orthogonal with $\rho_{ln}(t)$

for $n \neq i$, $l,k = 1,2$. Also, a simple calculation shows that $\rho_{1i}(t)$ is orthogonal to $\rho_{2i}(t)$. As a result, the eigenvectors $\{\rho_{1i}, \rho_{2i}\}$ are orthnormal with each other. This proves that $\rho_{ki}(t), k = 1,2$, given by Eqs. (7.10) are the eigenfunctions of Eq. (7.12), each corresponding to the eigenvalue given by Eq. (7.41).

2. Suppose $h(t)$ is an eigenfunction of Eq. (7.12) corresponding to the eigenvalue κ_h, which is orthogonal to all $\rho_{1i}(t)$ and $\rho_{2i}(t)$. Since from the properties of PSW functions, $\psi_i(t)$ are complete in $\mathcal{D}$ so that we can write $D[h(t)] = \sum_{n=0}^{\infty} \alpha_n D[\psi_n(t)]$, where α_n being the coefficients. Substituting $h(t)$ into Eq. (7.12) for $|t| > \frac{T}{2}$, we have

$$\int_{-\frac{T}{2}}^{\frac{T}{2}} D[h(\tau)]2F\text{sinc}[2\pi F(t-\tau)]d\tau = \kappa_h h(t). \tag{7.49}$$

Since from the properties of PSW function, $\psi_n(t)$ satisfies the eigen equation

$$\int_{-\frac{T}{2}}^{\frac{T}{2}} D[\psi_n(\tau)]2F\text{sinc}[2\pi F(t-\tau)]d\tau = \lambda_n \psi_n(t),$$

the left hand side of Eq. (7.49) will be

$$\sum_{n=0}^{\infty} \alpha_n \int_{-\frac{T}{2}}^{\frac{T}{2}} D[\psi_n(\tau)]2F\text{sinc}[2\pi F(t-\tau)]d\tau = \sum_{n=0}^{\infty} \alpha_n \lambda_n \psi_n(t).$$

Hence Eq. (7.49) can be written as

$$h(t) = \sum_{n=0}^{\infty} \alpha_n \kappa_h^{-1} \lambda_n \psi_n(t), \qquad |t| > \frac{T}{2}.$$

Combining the expression of $h(t)$ for both $|t| \leq \frac{T}{2}$ and $|t| > \frac{T}{2}$, we have

$$h(t) = \sum_{n=0}^{\infty} h_n(t) = \sum_{n=0}^{\infty} \left[\alpha_n \lambda_n \kappa_h^{-1} \psi_n(t) + \alpha_n (1 - \lambda_n \kappa_h^{-1}) D[\psi_n(t)]\right], \tag{7.50}$$

where $h_n(t) = \alpha_n \lambda_n \kappa_h^{-1} \psi_n(t) + \alpha_n (1 - \lambda_n \kappa_h^{-1}) D[\psi_n(t)]$. Because $h(t)$ is orthogonal to all $\rho_{1i}(t)$, i.e.

$$(t)\rho_{1i}(t) = 0.$$

Substituting Eqs. (7.39) and (7.50) (the expression of $\rho_{1i}(t)$ and $h(t)$ respectively) into the above relation and using the properties of $\psi_i(t)$ that $\int_{-\infty}^{\infty} \psi_i(t)\psi_n(t)dt = \delta_{in}$ and $\int_{-\infty}^{\infty} D[\psi_i(t)]D[\psi_n(t)]dt = \lambda_i\delta_{in}$, we have

$$\delta_{ni}[\alpha_n(1-\lambda_n\kappa_h^{-1})(p_{1i}+q_{1i})\lambda_i + \kappa_h^{-1}\alpha_n\lambda_n(q_{1i}\lambda_i + p_{1i})] = 0. \tag{7.51}$$

Using Eqs. (7.42) and (7.43), Eq. (7.51) can be written as (for $i = n$):

$$\frac{\alpha_n(1-\lambda_n\kappa_h^{-1})}{\alpha_n\kappa_h^{-1}\lambda_n} = -\frac{q_{1n}\lambda_n + p_{1n}}{(p_{1n}+q_{1n})\lambda_n} = -\frac{q_{1n}}{p_{1n}} = \frac{q_{2n}}{p_{2n}}. \tag{7.52}$$

It means that $h_n(t)$ is the constant multiple of $\rho_{2n}(t)$. Therefore $h(t) = \sum_{n=0}^{\infty} h_n(t)$ is the linear combination of $\{\rho_{2i}(t)\}$, contradicting the orthorgonality of $h(t)$ with all $\rho_{2i}(t)$. □

Theorem 7.3 states that $\{\rho_{1i}(t), \rho_{2i}(t)\}_{i=0}^{\infty}$ forms a complete orthonormal basis in $\mathcal{B}+\mathcal{D}$. Any signal $s(t)$ in $\mathcal{B}+\mathcal{D}$ can be expanded as

$$s(t) = \sum_{i=0}^{\infty} \alpha_i\rho_{1i}(t) + \sum_{i=0}^{\infty} \beta_i\rho_{2i}(t), \tag{7.53}$$

with α_i and β_i being the coefficients.

From Eqs. (7.11), it could be easily observed that since $1 > \lambda_i > 0$, each function $\rho_{1i}(t)$ satisfies $E_{t1i} \geq 50\%$, $E_{f1i} \geq 50\%$ and $E_{t1i}E_{f1i}$ decreases with i increasing, while each $\rho_{2i}(t)$ satisfies $E_{t2i} \leq 50\%$, $E_{f2i} \leq 50\%$ and $E_{t2i}E_{f2i}$ increases with i increasing. Hence, the basis functions can be ordered according to its energy concentration, i.e., $\rho_{10}(t)$ has the maximum E_tE_f in $\mathcal{B}+\mathcal{D}$, $\rho_{11}(t)$ has the next maximum value, then $\rho_{12}(t)$, $\rho_{13}(t)$, $\cdots$, $\rho_{1\infty}(t)$, $\rho_{2\infty}(t)$, $\cdots$, $\rho_{21}(t)$, $\rho_{20}(t)$. Since the integral equation Eq. (7.12) is derived based on the condition that E_tE_f is maximized, it could be expected that this set of basis functions are optimum in the sense that when the basis functions have been ordered according to the energy concentration, every eigenvector yields the maximum E_tE_f in a subspace of $\mathcal{B}+\mathcal{D}$ which is complement to the subspace spanned by all the

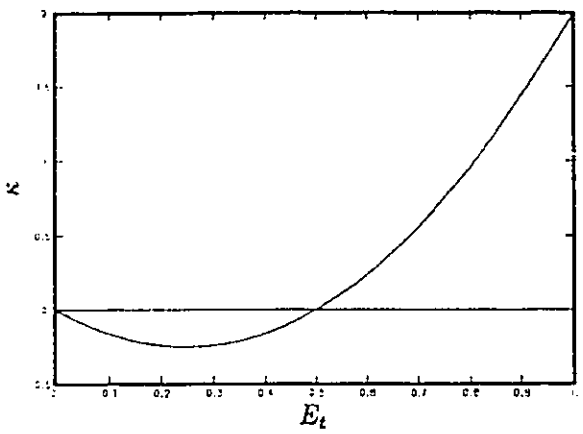


Figure 7.1: The relation between κ and E_t.

eigenvectors before it [49] and [50]. Therefore, if we span a time- and band-limited function over this basis, the coefficients should quickly vanish to zero.

To simplify the computation, the signal is often expanded in terms of finite number of basis functions. Also, in practice, we usually demand the concentration of the signal energy to be within a given duration and bandwidth. If $\rho_{1i}(t)$ and $\rho_{2i}(t)$ are ranked according to its energy concentration and if a signal $s(t)$ is expanded onto first N terms of basis functions, Eq. (7.53) becomes

$$s(t) \approx \sum_{i=0}^{N-1} \alpha_i \rho_{1i}(t). \tag{7.54}$$

Hence, in the subsequent applications, only $\{\rho_{1i}(t)\}$ is selected as the basis functions.

From Eq. (7.11), we know that each eigenvector $h(t)$ of Eq. (7.12) satisfies $E_t = E_f$ and from Eq. (7.41) the eigenvalues κ are related with the energy concentration of $h(t)$ in given time duration and bandwidth through $\kappa = 2E_t(2E_t - 1)$. The relation of κ and E_t are shown in Fig.7.1. If we choose the basis functions $\{\rho_{1i}\}$ which satisfy $E_t > 0.5$ (at least 50% energy should be concentrated in the given time duration and bandwidth), then κ increases monotonically with E_t. Therefore the eigenvector corresponding to the largest eigenvalues κ_{max} is the function which maximizes $E_t E_f$, i.e, ρ_{10}. Other eigenvectors,

ordered according to the eigenvalues κ_i, will be $\rho_{1i}(t)$ which are also ordered according to their respective energy concentration in given $\mathcal{B}$ and $\mathcal{D}$.

7.1.3 Conclusion

In this section, we have investigated the uncertainty principle existing in the time and frequency domain and derived a new set of complete and orthonormal basis in the space $\mathcal{B} + \mathcal{D}$. We have also proved that the basis functions can be ordered according to its energy concentration. Most communication signals can be expanded onto this base with few coefficients or fast convergence rate [49] and [50], and it can be applied in the general signal design problems[46].

7.2 Lower Bound of $Var(t)Var(f)$ and Optimum Signal

After establishing the optimum basis for signal representation, we return to the consideration of designing optimum signals for the estimation of TD and DS. From Eqs.(6.31) and (6.68), it can easily be seen that the performance of estimating the time delay (TD) and the Doppler shift (DS) depends on the second derivatives at the peak of the WD of the signal envelope. In general, the sharper is rise of the peak in both time and frequency axes, the smaller are the estimation errors in TD and DS. Unfortunately, due to the fact that signals of the same energy have equal volumes under their respective WD [20] [21], a sharp and narrow peak implies that the volume has to be distributed elsewhere giving rise to high sidelobes which would increase the difficulty of locating the true peak. Thus, the problem with which we are confronted is how to design a signal with its WD having the narrowest peak and yet having sidelobes below a certain level.

7.2.1 Condition for a Minimum $Var(\hat{t}_0)Var(\hat{f}_\Delta)$

From Eq.(6.68), it follows immediately that the product of the variances of $\hat{t}_0$ and $\hat{f}_\Delta$ is inversely proportional to the product $W_{11}W_{22}$, and using Eqs.(6.38) and (6.39) we can write

$$Var(\hat{t}_0)Var(\hat{f}_\Delta) \propto \frac{1}{F_a^2 T_a^2}, \tag{7.55}$$

where F_a^2 and T_a^2 are the normalized second moments of $|A(f)|^2$ and $a^2(t)$ respectively, i.e.,

$$F_a^2 = \frac{\int_{-\infty}^{\infty} f^2 \mid A(f) \mid^2 df}{\int_{-\infty}^{\infty} \mid A(f) \mid^2 df} \tag{7.56}$$

$$T_a^2 = \frac{\int_{-\infty}^{\infty} t^2 \mid a(t) \mid^2 dt}{\int_{-\infty}^{\infty} \mid a(t) \mid^2 dt}. \tag{7.57}$$

Without loss of generality, we can assume that the total energy of the signal is unity, i.e.,

$$\int_{-\infty}^{\infty} a^2(t)dt = \int_{-\infty}^{\infty} |A(f)|^2 df = 1. \tag{7.58}$$

Hence, if we want to minimize the joint estimation error represented by the product $Var(\hat{t}_0)Var(\hat{f}_\Delta)$, we must maximize the duration-bandwidth product $T_a^2 F_a^2$ of the signal $a(t)$. Without putting further constraints, we can easily obtain an $a(t)$ whose $T_a^2 F_a^2$ will approach infinity. However, in practice, we only use signals which are *essentially time-limited* and *essentially band-limited*. This implies that the signal outside a certain duration, outside a certain bandwidth is negligible, i.e.,

$$a(t) \approx 0 \quad |t| > \frac{T}{2} \qquad \text{and} \qquad |A(f)| \approx 0 \quad |f| > F, \tag{7.59}$$

where T and F are respectively the essential time duration and essential bandwidth. Under these conditions, we can rewrite Eq. (7.55) as

$$Var(\hat{t}_0)Var(\hat{f}_\Delta) \propto \frac{1}{\Delta F^2 \cdot \Delta T^2}, \tag{7.60}$$

where

$$\Delta F^2 = \int_{-F}^{F} f^2 \mid A(f) \mid^2 df \tag{7.61}$$

and

$$\Delta T^2 = \int_{-T/2}^{T/2} t^2 a^2(t) dt. \tag{7.62}$$

The problem of designing an optimum signal $a_0(t)$ so that minimum joint estimation error is achieved can then be stated as

$$\max_{a(t)} \log(\Delta F^2 \cdot \Delta T^2), \tag{7.63}$$

subject to the constraints of Eq. (7.58). The objective functional for the problem is thus

$$F = \log(\Delta T^2) + \log(\Delta F^2) + \lambda(\int_{-\infty}^{\infty} a^2(t) dt - 1), \tag{7.64}$$

where λ is the Lagrange multiplier. Let

$$a(t) = a_0(t) + \varepsilon \mu(t), \tag{7.65}$$

where $a_0(t)$ is the optimum signal, $\mu(t)$ is a real signal with finite energy, and ε is a small real constant. Thus, $\varepsilon\mu(t)$ represents a perturbation signal. Then, we have

$$\left.\frac{\partial F}{\partial \varepsilon}\right|_{\varepsilon=0} = 0. \tag{7.66}$$

Differentiating F with respect to ε and setting $\varepsilon = 0$, after simplifications (details shown in Appendix H), Eq. (7.66) can be written as

$$\begin{aligned} \left.\frac{\partial F}{\partial \varepsilon}\right|_{\varepsilon=0} &= \int_{-T/2}^{T/2} \mu(t) \left\{ \frac{2t^2 a_0(t)}{\Delta T^2} \right\} dt + \int_{-\infty}^{\infty} \mu^*(t) \left\{ \int_{-F}^{F} \frac{2f^2 A_0(f) e^{j2\pi ft}}{\Delta F^2} df \right\} dt \\ &\quad + \lambda \int_{-\infty}^{\infty} \mu(t)\{2a_0(t)\} dt \\ &= 0. \end{aligned} \tag{7.67}$$

Eq. (7.67) should be satisfied by all finite-energy, essentially band-limited and time-limited signals $\mu(t)$ if $a_0(t)$ is the optimum signal. Therefore, we have

$$\frac{t^2 a_0(t)}{\Delta T^2} + \frac{1}{\Delta F^2} \int_{-F}^{F} f^2 A_0(f) e^{j2\pi ft} df + \lambda a_0(t) = 0 \quad \text{for } \mid t \mid \leq T/2, \tag{7.68}$$

and

$$\frac{1}{\Delta F^2}\int_{-F}^{F} f^2 A_0(f)e^{j2\pi ft}df + \lambda a_0(t) = 0 \quad \text{for } |t| > T/2. \tag{7.69}$$

Eqs. (7.68) and (7.69) are the conditions that an optimum signal has to satisfy. Unfortunately, the exact solution of these equation is difficult to obtain. In the next section, we will discuss some properties of the objective function $\Delta F^2 \Delta T^2$ and show a way to obtain an approximate solution. Note that the optimum signal described above is focused on minimizing the joint estimation error $Var(\hat{t}_0) \cdot Var(\hat{f}_\Delta)$ only. No consideration has yet been given on the constraints of sidelobe magnitude, or on the amount of energy concentration.

7.2.2 The Signal Design Region

The signal design problem described in the previous section requires us to select from the class of signals which have unity energy and are essentially time- and band-limited, one which has the largest product of $\Delta F^2 \Delta T^2$. The signal $a(t)$ has infinitely many dimensions and the sum of the square of the coefficients in these dimensions is unity. Thus, the unity energy equation can geometrically be interpreted as a multi-dimensional sphere of unit radius. The design problem requires that this multi-dimensional sphere be mapped onto the two-dimensional $\Delta T^2 \sim \Delta F^2$ plane, forming a design region in which we can search for the optimum signal. Since ΔT^2 and ΔF^2 are always positive, the design region lies only in the positive quadrant of the $\Delta T^2 \sim \Delta F^2$ plane. Furthermore, the design region includes both the axes $\Delta F^2 = 0$ and $\Delta T^2 = 0$. This is because we can find a signal $a(t)$ which has unit energy but also satisfies

$$a(t) = 0 \qquad \text{for } |t| \leq \frac{T}{2}, \tag{7.70}$$

in which case $\Delta T^2 = 0$ and $\Delta F^2 > 0$. Similarly, we can also find signals with unit energy but

$$|A(f)| = 0 \qquad \text{for } |f| \leq F. \tag{7.71}$$

which maps onto the ΔT^2-axis. Now, let us examine some other properties of the design region.

Theorem 7.4 *Let $a_1(t)$ be a point on the surface of the infinite-dimensional unit sphere and let $(\Delta T_1^2, \Delta F_1^2)$ be the mapping of this point onto the $\Delta T^2 \sim \Delta F^2$ plane. Then $(k\Delta T_1^2, k\Delta F_1^2)$ is the mapping of $\sqrt{k}a_1(t)$ onto the $\Delta T^2 \sim \Delta F^2$ plane.*

Proof:

$$\begin{aligned} \Delta T_1^2 &= \int_{-T/2}^{T/2} t^2 a_1^2(t)dt \\ \Delta F_1^2 &= \int_{-F}^{F} f^2 \mid A_1(f) \mid^2 df \end{aligned}$$

Let $a_2(t) = \sqrt{k}a_1(t)$, then

$$\begin{aligned} \Delta T_2^2 = \int_{-T/2}^{T/2} t^2 a_2^2(t)dt &= k\Delta T_1^2 \\ \Delta F_2^2 = \int_{-F}^{F} f^2 \mid A_2(f) \mid^2 df &= k\Delta F_1^2 \end{aligned}$$

The geometric interpretation of this theorem is shown in Fig.7.2. □

Corollary 7.1 *If $\Re_u$ and $\Re_s$ denote, respectively, the region on the $\Delta T^2 \sim \Delta F^2$ plane onto which the point of the multi- dimensional sphere $\int_{-\infty}^{\infty} a^2(t) \leq 1$ and the points on the sphere surface $\int_{-\infty}^{\infty} a^2(t) = 1$ map. then (i) $\Re_s \subseteq \Re_u$, and (ii) $\Re_s$ includes the extreme points of $\Re_u$ from the origin.*

Proof: Part (i) that $\Re_s \subseteq \Re_u$ is obvious from their respective definitions.

To prove part (ii), we make use of Theorem 7.4. $\int_{-\infty}^{\infty} a^2(t) \leq 1$ represents a unit radius infinite-dimensional sphere, the surface of which is represented by the case when

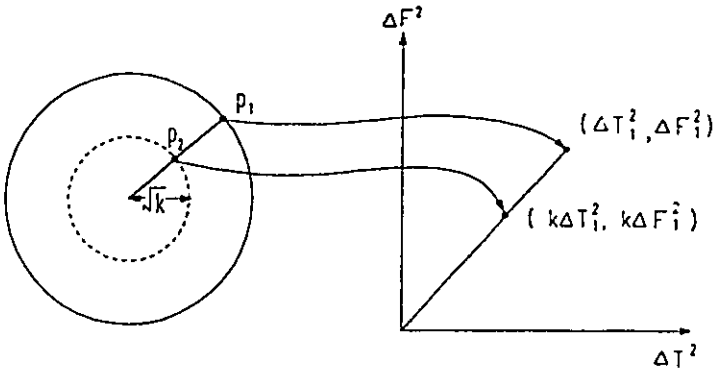


Figure 7.2: Geometric Interpretation of Theorem 7.4.

the equality holds. For $0 \leq k < 1$, suppose we can find a point $a_2(t)$ inside the sphere distance $\sqrt{k}$ from the origin, and suppose this point maps onto a point $(\Delta T_2^2, \Delta F_2^2)$ which is an extreme point (farthest away) in $\Re_u$ from the origin in that direction. Now, let

$$a_1(t) = \frac{1}{\sqrt{k}} a_2(t)$$

Then $a_1(t)$ is on the surface of the unit sphere and by Theorem 7.4, it maps onto the point $(\frac{1}{k}\Delta T_2^2, \frac{1}{k}\Delta F_2^2)$ which is the same direction as $(\Delta T_2^2, \Delta F_2^2)$. Since $k < 1$, this point is further away from the origin than $(\Delta T_2^2, \Delta F_2^2)$. Thus, the supposition must be wrong and part (ii) is proven. □

Theorem 7.4 and the corollary have some interesting implications on the design region. The multi-dimensional unit sphere maps onto the $\Delta T^2 \sim \Delta F^2$ plane. The region of mapping $\Re_u$ is bounded by the two axes $\Delta T^2 = 0$ and $\Delta F^2 = 0$, together with a continuous (since the mapping is continuous) curve C. The complete boundary of $\Re_u$ belongs to the sub-region $\Re_s$ which is mapped onto from points on the surface of the multi- dimensional unit sphere. Now, let us examine the shape of this curve C resulted by the mapping so that we can have a better understanding of the mapping region.

Consider a two-dimensional plane through the centre of the multi-dimensional unit sphere. This plane will intersect the sphere surface in a two- dimensional unit circle.

The signals on this unit circle of intersection can be expressed as

$$a(t) = a_1\varphi_1(t) + a_2\varphi_2(t), \tag{7.72}$$

where a_1 and a_2 are real constants and $\varphi_1(t), \varphi_2(t)$ are the two orthonormal basis functions spanning the unit circle in the two-dimensional space.

Theorem 7.5 *The mapping of the two-dimensional unit circle of intersection given by*

$$\int_{-\infty}^{\infty} a^2(t)dt = a_1^2 + a_2^2 = 1$$

onto the $\Delta T^2 \sim \Delta F^2$ *plane is an ellipse.*

Proof: The proof of this theorem is given in Appendix I. □

With the help of the above theorem, we come to a very important property of the design region. This is stated in the following theorem:

Theorem 7.6 $\Re_u$ *is a convex region.*

Proof: The statement of the theorem is depicted in Fig.7.3. It can be easily understood by Theorem 7.4 and its Corollary, that the total mapping region $\Re_u$ in the $\Delta T^2 \sim \Delta F^2$ plane is simple connected, bounded by two axes $\Delta T^2 = 0$, $\Delta F^2 = 0$ and a continuous curve C. Hence, in order to prove that $\Re_u$ is convex, we only need to prove that the segment of a straight line between any two points on the curve C belongs to $\Re_u$.

Select any two points $(\Delta T_1^2, \Delta F_1^2)$ and $(\Delta T_2^2, \Delta F_2^2)$ on the boundary curve C. By Theorem 7.4 and its Corollary, these two points are results of mapping of two points $a_1(t)$ and $a_2(t)$ on the surface of the multi-dimensional unit sphere. Now, we can construct a two- dimensional unit disc through these two points on the unit sphere surface and the origin. The boundary of the unit disc coincides the multi-dimensional sphere at a two-dimensional unit circle. By Theorem 7.5, the mapping of this two-dimensional unit circle is an ellipse on the $\Delta T^2 \sim \Delta F^2$ plane and according to Theorem 7.4 and its corollary,

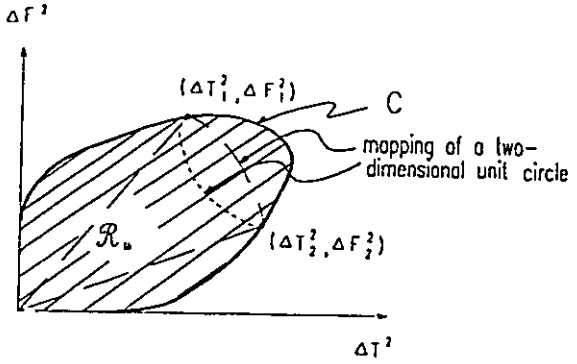


Figure 7.3: The Mapping of all Real Signals $a(t)$ onto $\Delta T^2 \sim \Delta F^2$ plane.

the mapping of the unit disc is a cone shape in the $\Delta T^2 \sim \Delta F^2$ plane. This cone is formed by two straight lines starting from the origin and tangent to the ellipse, and uses the farther side of the ellipse as its base. The shape of this cone is depicted in Fig.7.3, which is bounded by the dash line in the Fig.7.3. It is obvious that this cone belongs to $\Re_u$ and it is convex since the ellipse is convex. Therefore, if we form the straight line which connects two points $(\Delta T_1^2, \Delta F_1^2)$ and $(\Delta T_2^2, \Delta F_2^2)$, the segment of this straight line between $(\Delta T_1^2, \Delta F_1^2)$ and $(\Delta T_2^2, \Delta F_2^2)$ must lie inside the cone and hence belongs to $\Re_u$. Now, the fact that the total mapping region is convex is proved. □

7.2.3 Optimum Signal in the Unconstrained Design Region

The convexity of the design region $\Re_u$ shown by Theorem 7.6 has profound implications. In general, we do not know the exact shape of the design region, therefore, we cannot locate the point on the boundary of $\Re_u$ that yields the maximum value of $\Delta T^2 \Delta F^2$, and hence we cannot obtain the optimum signal $a_0(t)$. However, since we know the design region is convex, we can approximate the curve by a piecewise linear curve with any accuracy, i.e., we can approximate the boundary of $\Re_u$ by a polygon. From this polygon we can locate very efficiently the point on the boundary which yields the maximum value

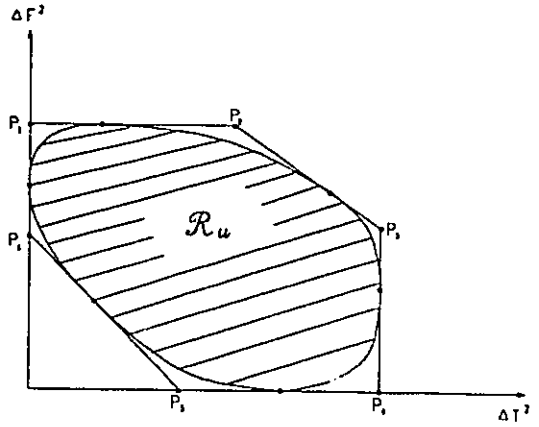


Figure 7.4: Initial Approximation to the Convex Region $\Re_u$ with an Octagon.

of $\Delta T^2 \Delta F^2$ and hence obtain the corresponding optimum signal $a_0(t)$. The larger is the number of sides of the polygon, the more accurate is the approximation. Let us suppose that our first step of approximation to the convex region $\Re_u$ is an octagon. Since we are searching for the maximum value of $\Delta T^2 \Delta F^2$ in this region, we only need to concentrate on the sides of the octagon furthest away from the region. These sides of the octagon are represented by the solid lines P_1P_2, P_2P_3, and P_3P_4 in Fig.7.4. In this example we have chosen P_1P_2 to be parallel to the ΔT^2-axis, P_2P_3 to be inclined at an angle $-\frac{\pi}{4}$ to the horizontal axis, and P_3P_4 to be vertical to the ΔT^2-axis. All these straight lines are tangential to the boundary of $\Re_u$. Thus the respective equations of these lines are

$$P_1P_2: \quad \alpha_0 \Delta T^2 + \beta_0 \Delta F^2 = k_0 \tag{7.73}$$

$$P_2P_3: \quad \alpha_1 \Delta T^2 + \beta_1 \Delta F^2 = k_{m1} \tag{7.74}$$

$$P_1P_2: \quad \alpha_\infty \Delta T^2 + \beta_\infty \Delta F^2 = k_\infty, \tag{7.75}$$

where $\alpha_0 = 0$, $\beta_0 = 1$; $\alpha_1 = 0.5$, $\beta_1 = 0.5$; $\alpha_\infty = 1$, $\beta_\infty = 0$; and the constants k_0, k_{m1}, and k_∞ ensures that these lines are tangents to the boundary of $\Re_u$.

To obtain the constant k_0, k_{m1}, and k_∞, let us first select a set of orthonormal basis functions $\{\phi_n(t)\}$ and let $a(t)$ be approximated to a finite $(N+1)$-dimensional space

such that

$$a(t) \approx \sum_{n=0}^{N} a_n \phi_n(t), \tag{7.76}$$

where

$$\int_{-\infty}^{\infty} \phi_m(t)\phi_n(t)dt = \delta_{mn}. \tag{7.77}$$

We defined the $(N+1)$-dimensional vectors

$$\boldsymbol{a} \triangleq [a_0, \cdots, a_N]^T \tag{7.78}$$

$$\boldsymbol{\phi}(t) \triangleq [\phi_0(t), \cdots, \phi_N(t)]^T. \tag{7.79}$$

Note that due to the unit energy constraint on $a(t)$,

$$\boldsymbol{a}^T\boldsymbol{a} = 1, \tag{7.80}$$

where the superscript ^{T} denotes the conjugate of a vector or matrix.

Then, by substituting Eqs. (7.76), (7.78) and (7.79) into the definitions of ΔF^2 and ΔT^2 as expressed in Eqs. (7.61) and (7.62), we obtain

$$\Delta F^2 = \boldsymbol{a}^T \boldsymbol{\Delta}_F \boldsymbol{a} \tag{7.81}$$

$$\Delta T^2 = \boldsymbol{a}^T \boldsymbol{\Delta}_T \boldsymbol{a} \tag{7.82}$$

where

$$\boldsymbol{\Delta}_F \triangleq \int_{-F}^{F} f^2 \boldsymbol{\Phi}(f)\boldsymbol{\Phi}^T(f)df \tag{7.83}$$

and

$$\boldsymbol{\Delta}_T \triangleq \int_{-\frac{T}{2}}^{\frac{T}{2}} t^2 \boldsymbol{\phi}(t)\boldsymbol{\phi}^T(t)dt, \tag{7.84}$$

with $\boldsymbol{\Phi}(f)$ being the Fourier transform of $\boldsymbol{\phi}(t)$. Thus, the three tangential lines of Eqs. (7.73) to (7.75) can be written as

$$\max_{\boldsymbol{a}} \Delta F^2 = \max_{\boldsymbol{a}} \boldsymbol{a}^T \boldsymbol{\Delta}_F \boldsymbol{a} = k_0 \tag{7.85}$$

$$\max_{\boldsymbol{a}} (.5\Delta T^2 + .5\Delta F^2) = \max_{\boldsymbol{a}} \boldsymbol{a}^T \boldsymbol{\Delta}_1 \boldsymbol{a} = k_{m1} \tag{7.86}$$

$$\max_{\boldsymbol{a}} \Delta T^2 = \max_{\boldsymbol{a}} \boldsymbol{a}^T \boldsymbol{\Delta}_T \boldsymbol{a} = k_\infty, \tag{7.87}$$

where

$$\Delta_1 = .5\Delta_F + .5\Delta_T \tag{7.88}$$

Eqs. (7.85) to (7.87) are separate eigenvalue-eigenvector equations for the symmetric matrices Δ_F, Δ_1, and Δ_T respectively, and can be solved separately. The maximization of the quantities implies that for each of the matrices, we have to find its largest eigenvalue k and the corresponding eigenvector ($(N+1)$-tuple) a. There are standard numerical methods for finding the eigenvalues and eigenvectors of a matrix. The details of these methods are described in [37]. Having solved for the largest eigenvalues k_0, k_{m1}, and k_∞ of the matrices Δ_F, Δ_1, and Δ_T respectively, we can then proceed to locate the point on the polygon yielding the maximum $\Delta T^2 \Delta F^2$ which can be equivalently described as

$$\max_{a(t)} \ \Delta T^2 \Delta F^2, \tag{7.89}$$

subject to the constraints

$$\Delta F^2 \ \leq \ k_0 \qquad .5\Delta T^2 + .5\Delta F^2 \ \leq \ k_{m1} \qquad \Delta T^2 \ \leq \ k_\infty. \tag{7.90}$$

This is a two-dimensional optimization problem which can be easily solved by first forming the objective function

$$F = \Delta T^2 \Delta F^2 + \lambda(.5\Delta T^2 + .5\Delta F^2 - k_{m1}), \tag{7.91}$$

where λ is a Lagrange multiplier. From Eq. (7.91), we obtain

$$\frac{\partial F}{\partial(\Delta T^2)} \ = \ \Delta F^2 + .5\lambda = 0 \tag{7.92}$$

$$\frac{\partial F}{\partial(\Delta F^2)} \ = \ \Delta T^2 + .5\lambda = 0 \tag{7.93}$$

$$\frac{\partial F}{\partial \lambda} \ = \ .5\Delta T^2 + .5\Delta F^2 - k_{m1} = 0. \tag{7.94}$$

Eqs. (7.92) to (7.94) lead to an optimum solution such that

$$\Delta T^2 = \Delta F^2 = k_{m1}. \tag{7.95}$$

This solution may or may not contradictory with the constraints of Eqs. (7.90). the following are the three cases to be considered:

1. If $k_0 \geq k_{m1}$ and $k_\infty \geq k_{m1}$, then there is no contradiction and we choose Eq. (7.95) as the optimum point.

2. If $k_0 < k_{m1}$, then since $k_0 + k_\infty \geq 2k_{m1}$, we have

$$k_\infty \geq 2k_{m1} - k_0 > k_{m1}$$

and the optimum point is chosen to be

$$\Delta F^2 = k_0, \quad \Delta T^2 = 2k_{m1} - k_0. \tag{7.96}$$

3. Similarly, if $k_\infty < k_{m1}$, the optimum point will be

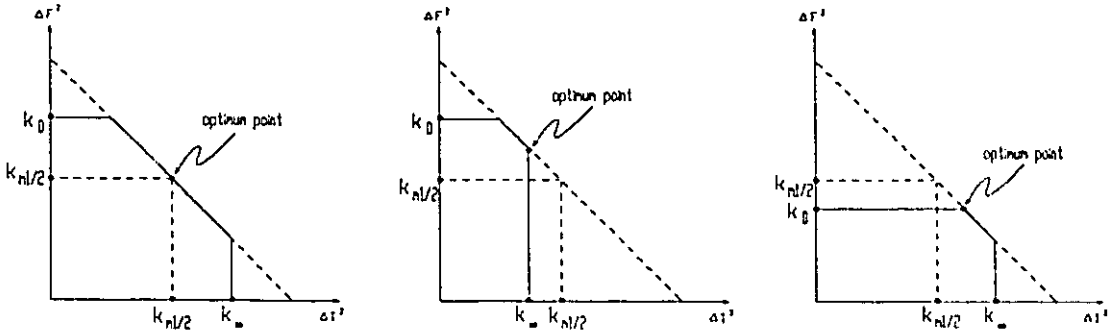
$$\Delta T^2 = k_\infty, \quad \Delta F^2 = 2k_{m1} - k_\infty. \tag{7.97}$$

These three cases are depicted in Figs.7.5(a), (b), and (c) respectively. Consideration of above three cases leads to the following inequality,

$$\Delta T^2 \Delta F^2 \leq \min(k_0, k_{m1}, k_\infty)\{2k_{m1} - \min(k_0, k_{m1}, k_\infty)\}. \tag{7.98}$$

Very often, we find that the first approximation may not be sufficiently accurate. Under such circumstances, we iterate the procedure by putting on an additional side of the polygon and re-establish another set of constraint equations similar to Eq. (7.91) taking into account the new additional side of the polygon. The details of the procedures are outlined below:

1. Obtain the matrices $\boldsymbol{\Delta}_F$, $\boldsymbol{\Delta}_1$, and $\boldsymbol{\Delta}_T$ using Eqs. (7.83), (7.88) and (7.94) respectively, and solve for their respective largest eigenvalues k_0, k_{m1}, and k_∞ together with the corresponding eigenvectors $\boldsymbol{a}_0$, $\boldsymbol{a}_{m1}$, and $\boldsymbol{a}_\infty$.



(a) (b) (c)

Figure 7.5: Three Different Cases for the Selection of Optimum Solutions on the Octagon Approach.

2. Set up the three sides of the polygon represented by Eq. (7.91). This is depicted in Fig.7.6. The tangential points $(\Delta T_0^2,\ \Delta F_0^2)$, $(\Delta T_{m1}^2,\ \Delta F_{m1}^2)$ and $(\Delta T_\infty^2,\ \Delta F_\infty^2)$ can be easily calculated by substituting the eigenvector of the corresponding subscripts into Eqs. (7.81) and (7.82). The first approximated optimum point $(\Delta T_{op1}^2,\ \Delta F_{op1}^2)$ is then calculated using Eq. (7.98), shown as the points on the solid straight lines in Fig.7.6.

3. Test the difference of the value $\Delta T^2 \Delta F^2$ between the point $(\Delta T_{op1}^2,\ \Delta F_{op1}^2)$ and the closest tangential point which in the example of Fig.7.6, is $(\Delta T_{m1}^2,\ \Delta F_{m1}^2)$. If the difference is smaller than a prescribed small constant ε, then assign the closest tangential point $(\Delta T_{m1}^2,\ \Delta F_{m1}^2)$ as the optimum point and employ the corresponding $(N+1)$−tuple a_{m1} in Eq. (7.76) to construct the approximate optimum signal $a(t)$.

4. If the difference of the value between $(\Delta T_{op1}^2,\ \Delta F_{op1}^2)$ and the closest tangential point is greater than ε, then we enter the second iteration of approximation by constructing an additional side (shown by the dashed line in Fig.7.6) of the polygon. Now, suppose that the first approximate optimum point $(\Delta T_{op1}^2,\ \Delta F_{op1}^2)$ is located on the side of the polygon between the two tangential points $(\Delta T_0^2,\ \Delta F_0^2)$ and

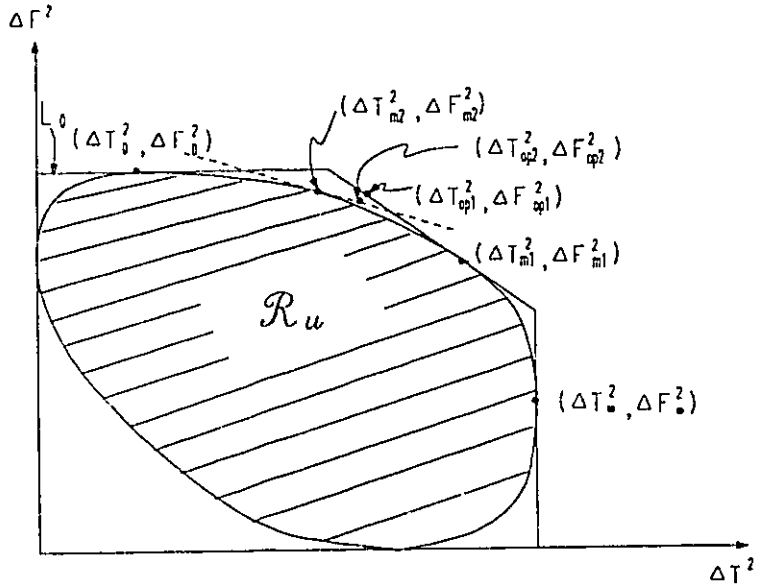


Figure 7.6: Selection of the Second Approximation to the Convex Region $\Re_u$ Based on the Octagon.

$(\Delta T_{m1}^2,\ \Delta F_{m1}^2)$, then it can be proved that the true optimum point lies between these two points as well (Appendix J). Thus, the additional side of the polygon should be constructed between these two points. The construction of this additional straight line is rather arbitrary. We employ the equation

$$\alpha_2 \Delta T^2 + \beta_2 \Delta F^2 = k_{m2}, \tag{7.99}$$

where the subscript denotes the second step of iteration. We also arbitrarily make

$$\alpha_i + \beta_i = 1 \quad \forall i. \tag{7.100}$$

Initially we chose the first side to be inclined at an angle of $-\frac{\pi}{4}$, and hence $\alpha_1 = \beta_1 = 0.5$. Now, since the additional side of the polygon line between $(\Delta T_0^2,\ \Delta F_0^2)$ and $(\Delta T_{m1}^2,\ \Delta F_{m1}^2)$, its gradient will be less negative than $-\frac{\pi}{4}$, i.e.,

$$\frac{\alpha_0}{\beta_0} < \frac{\alpha_2}{\beta_2} < \frac{\alpha_1}{\beta_1}. \tag{7.101}$$

We arbitrarily assign

$$\alpha_2 = \frac{(\alpha_0 + \alpha_1)}{2} \quad \text{and} \quad \beta_2 = \frac{(\beta_0 + \beta_1)}{2}, \tag{7.102}$$

resulting in a straight line whose gradient is somewhere between Eqs. (7.73) and (7.74).

5. We form the matrix

$$\Delta_2 = \alpha_2 \Delta_T + \beta_2 \Delta_F$$

and solve for its largest eigenvalue k_{m2} and the corresponding eigenvector a_{m2}.

6. Since the second approximate optimum point is to be located along the three sides represented by Eqs. (7.73), (7.99), and (7.74), we can disregard the vertical side of Eq. (7.75), and use only Eqs. (7.73), (7.99) and (7.74) as our constraints. The second approximate optimum point $(\Delta T^2_{op2},\ \Delta F^2_{op2})$ is obtained in a similar manner to step 2 and again a test of the difference of the value $\Delta T^2 \Delta F^2$ between $(\Delta T^2_{op2},\ \Delta F^2_{op2})$ and the closest tangential point is carried out. If the difference is smaller than ε, then the closest tangential point is chosen and an optimum signal constructed accordingly. If the difference is larger than ε, we construct yet another side of the polygon and the iteration continues until a sufficiently close point is located.

The above procedure leads to an approximate optimum point which converges to the true optimum point on the boundary of $\Re_u$. This is stated in the following theorem.

Theorem 7.7 *Let $(\Delta T^2_{mM},\ \Delta F^2_{mM})$ be the approximate optimum evaluated after M stages of iteration as described in the procedures above. Let $(\Delta T^2 \cdot \Delta F^2)_{max}$ be the upperbound of $\Delta T^2 \Delta F^2$ obtainable from the points on the boundary of the design region $\Re_u$. Then, given $\varepsilon > 0$ we can find an integer M_0 and a series of suitable α_i and β_i, $i = 1, \cdots, M$, such that for $M \geq M_0$,*

$$(\Delta T^2_{mM} \cdot \Delta F^2_{mM}) - (\Delta T^2 \cdot \Delta F^2)_{max} \leq \varepsilon. \tag{7.103}$$

Proof: See Appendix J. □

In Appendix J, we also shown that the polygon approximate method is linearly convergent making it very efficient.

As an example, we design an approximate signal $\hat{a}_0(t)$ which maximizes $\Delta T^2 \Delta F^2$. The basis function chosen is a linear combination of prolate spheroidal wave (PSW) function $\psi_n(t)$ and truncated PSW function $D[\psi_n(t)]$ which are given by Eq. (7.39)

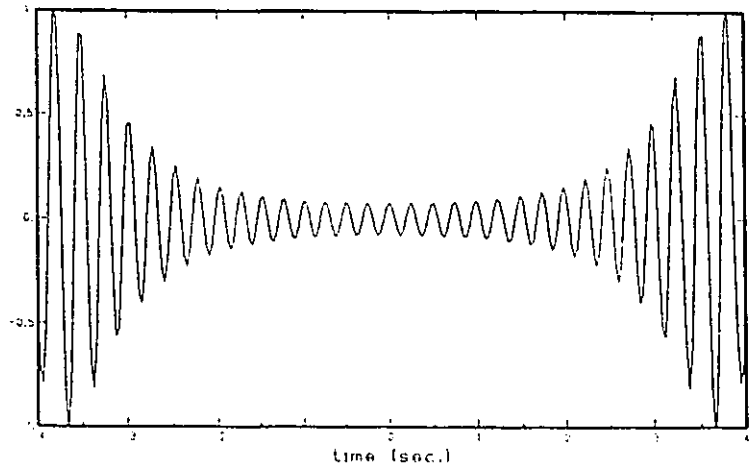
$$\{\phi_n(t)\} = \{p_{1n}\psi_n(t) + q_{1n}D\psi_n(t); n = 0, 1, \cdots\}. \tag{7.104}$$

These functions have been shown to be optimum as a basis because they have maximum "energy concentration" within a given duration and a given bandwidth respectively, as discussed in Section 7.1 and [49] and [50].

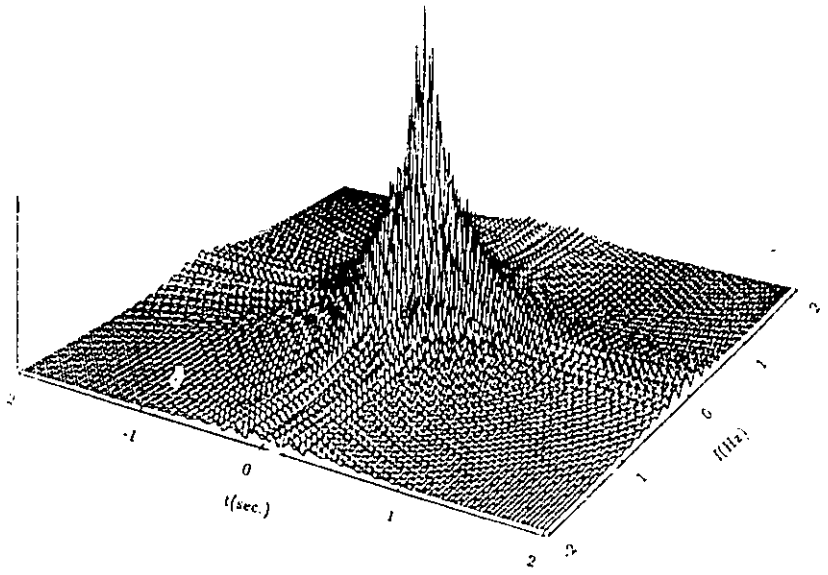
We have also decided that the signal $a(t)$ is expanded into 66 terms of the orthonormal basis functions, i.e., N=65 in Eq. (7.76). The essential duration and the essential bandwidth are chosen to be $T = 8$ seconds and $F = 4Hz$. The approximate optimum signal is obtained by the polygon approach described above and the search was terminated at $n = 10$ iterations when the difference of the values $\Delta T^2 \Delta F^2$ between the point $(\Delta T^2_{opn}, \Delta F^2_{opn})$ and the close tangential point is smaller than $.001\Delta T^2_{opn}\Delta F^2_{opn}$. Figs. 7.7 a) and b) show the final signal $\hat{a}_0(t)$ obtained and the amplitude of its corresponding Wigner distribution $W_{\hat{a}_0}(t, f)$ around $t = 0$, $f = 0$. As a comparison, we also performed a multi-dimensional global search for the vector $\boldsymbol{a}_0$ which maximizes $\Delta T^2 \Delta F^2$ yielding unit energy signal. The resulting signal $a_0(t)$ and the amplitude of its Wigner distribution $W_{a_0}(t, f)$ around $t = 0, f = 0$ are shown in Figs.7.8 a) and b) respectively. The mean square difference between the two signals is calculated to be

$$\int_{-T/2}^{T/2} \{a_0(t) - \hat{a}_0(t)\}^2 dt = 1.44 \times 10^{-4}, \tag{7.105}$$

certainly negligible compared to the energy of the signal. The computing time for the polygon approach, however, is over 300 times faster than that for the global search.

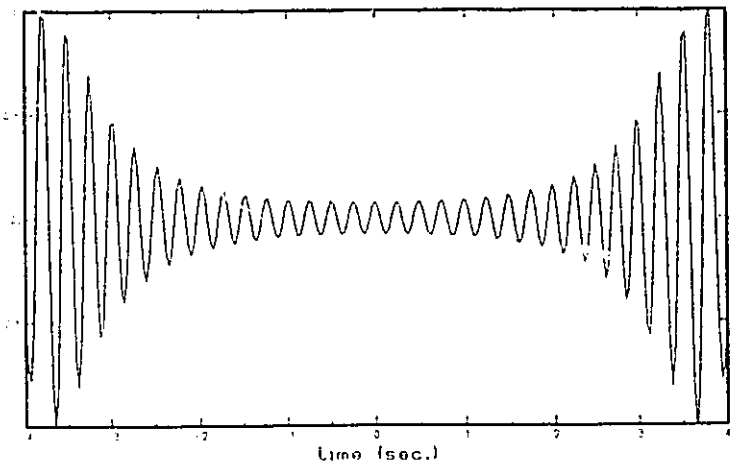


a)

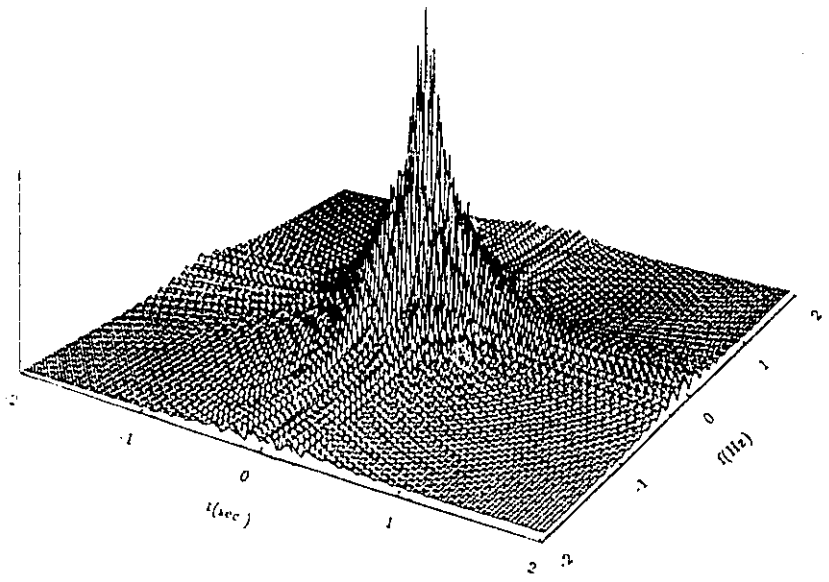


b)

Figure 7.7: a) Optimum Signal Obtained by Polygon Approach and b) the amplitude of its Corresponding WD around $t = 0$, $f = 0$.



a)



b)

Figure 7.8: a) Optimum Signal Obtained by Multi-dimensional Global Search and b) the amplitude of its Corresponding WD around $t = 0$, $f = 0$.

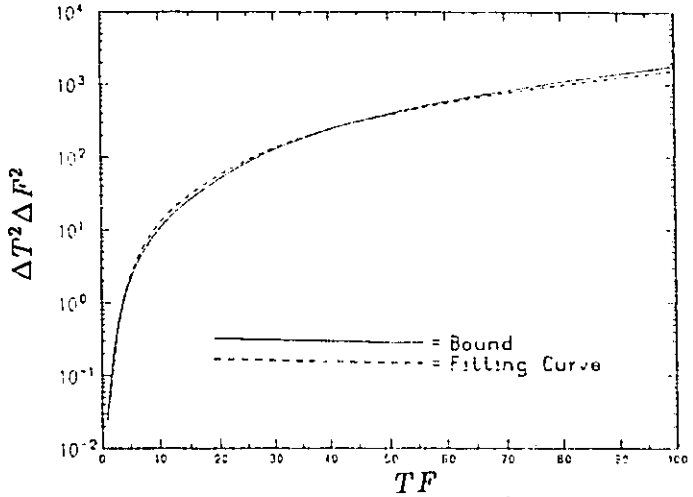


Figure 7.9: The Relationship Between maximum $\Delta T^2 \Delta F^2$ and the Essential Duration-Bandwidth Product FT.

In practice, it is found that the above method converges in about two or three iterations to within an accuracy of 1% of the product $(\Delta T^2 \cdot \Delta F^2)_{max}$ for most of situations, and this is a very efficient way of obtaining an optimum signal which minimizes the joint estimation error $Var(\hat{t}_0)Var(\hat{f}_\Delta)$.

In general, the value of $(\Delta T^2 \cdot \Delta F^2)_{max}$ for unit energy signals depends on the product of the essential duration and the essential bandwidth, i.e.,

$$(\Delta T^2 \cdot \Delta F^2)_{max} = y(FT). \tag{7.106}$$

This can be easily seen by time-scaling the unit energy signal and evaluating the product $\Delta T^2 \cdot \Delta F^2$ of the resulting signal. Fig.7.9 shows the relationship between the product $(\Delta T^2 \cdot \Delta F^2)_{max}$ and the essential duration-bandwidth product FT. Using a least square curve- fitting method, we can approximate this relationship by a quadratic function such that

$$\Delta T^2 \Delta F^2 = c_0 + c_1(FT) + c_2(FT)^2, \tag{7.107}$$

where c_0, c_1 and c_2 are obtained by solving following weighted least square problem:

$$\min_{\boldsymbol{c}} \epsilon^2 = \frac{1}{100} \min_{\boldsymbol{c}} \sum_{i=1}^{100} W(i)\{(\Delta T^2 \Delta F^2)_i - [c_0 + c_1(TF)_i + c_2(TF)_i^2]\}^2$$

where $\boldsymbol{c} = [c_0,\ c_1,\ c_2]$ and $(\Delta T^2 \Delta F^2)_i$ is the maximum value of $\Delta T^2 \Delta F^2$ when TF is equal to $(TF)_i$.

With $(TF)_i = i$ and $W(i) = (\Delta T^2 \Delta F^2)_i^{-2}$, the solution of above least square problem is:

$$c_0 = 0.255,\ c_1 = -0.388,\ c_2 = 0.159,$$

with the average normalized fitting error being $\epsilon^2 = 1.08\%$. The quadratic approximation function:

$$(\Delta T^2 \cdot \Delta F^2)_{max} = 0.255 - 0.388(FT) + 0.159(FT)^2, \tag{7.108}$$

is also shown in Fig.7.9, which is certainly a good approximation for the upperbound of $(\Delta T^2 \cdot \Delta F^2)$ for $1 \leq FT \leq 100$.

7.3 Optimum Signal Design under Practical Constrains

In the previous sections, we discussed the design of an essentially time limited and band-limited unit energy signal which minimizes the joint estimation error $Var(\hat{t}_0)Var(\hat{f}_\Delta)$. Equivalently it maximizes the product $\Delta T^2 \Delta F^2$. In practical applications, however, we often encounter situations which demand stricter constraints. For example, very often we may require more efficient use of time slot and bandwidth so that other channels can be accommodated with little interference. Under such circumstances, we would like to specify the energy contents within the prescribed time duration T and the prescribed bandwidth F such that

$$\int_{-T/2}^{T/2} a^2(t)dt \geq E_T \tag{7.109}$$

$$\int_{-F}^{F} |A(f)|^2 df \geq E_F, \tag{7.110}$$

where for unit energy signals, E_T and E_F are positive constants less than but close to unity. Another practical consideration, as has been discussed in section 6.5.2, is the level of the sidelobes in the WD of the optimum signal as obtained.

With these extra constraints, the design region on the $\Delta T^2 \sim \Delta F^2$ plane is, unfortunately, not convex any more. To obtain an optimum signal under these constraints is still possible. Here, we take two different approaches.

7.3.1 . Quasi-Simulated Annealing (QSA) Method of Optimization

The Simulated Annealing method, first introduced in 1983 [53] [40], is based on the annealing process in crystal growing. The idea is that if enough simulations are done with slowly decreasing "temperature", then the method will reach the global minimum point with probability 1. The theory behind the method is well explained in [53] [40]. Here, we give a brief outline of how it can be modified and applied to our constrained optimum problem.

Our problem is

$$\max_{\boldsymbol{a}} \Delta T^2 \Delta F^2$$

where

$$\boldsymbol{a} = [a_0\ a_1\ \cdots\ a_{N-1}\ a_N]^T. \tag{7.111}$$

Since the total energy of the signal is unity, then $-1 \leq a_n \leq 1, n = 0, \cdots, N$.

1. Select several "temperature" parameters θ_k such that

$$\theta_1 > \theta_2 \cdots > \theta_K, \qquad k = 1, \cdots, K$$

and

$$\lim_{K \to \infty} \theta_K = 0. \tag{7.112}$$

2. Discretize the range of the elements a_n in $\boldsymbol{a}$ with L-1 steps, i.e., the possible discrete values that a_n can assume are

$$a_{n1},\ a_{n2},\ \cdots,a_{nL}$$

where

$$a_{nl} = -1 + \frac{2(l-1)}{L-1} \qquad \begin{matrix} n = 0, & \cdots, N \\ l = 1, & \cdots \quad, L; \end{matrix} \quad . \tag{7.113}$$

3. For any point

$$\boldsymbol{a}_l = [a_{0l_0}\ a_{1l_1}\ \cdots\ a_{Nl_N}]^T, \tag{7.114}$$

a neighbour of $\boldsymbol{a}_l$ is defined as

$$\boldsymbol{a}_{l'} = [a_{0l'_0}\ a_{1l'_1}\ \cdots\ a_{Nl'_N}]^T, \tag{7.115}$$

where the element $a_{nl'_n}$ assumes the relationship with a_{nl_n} with certain probabilities in the following manner

$$a_{nl'_n} = \begin{cases} a_{n(l_n-1)}, & \text{with prob. } p_1 \\ a_{nl_n}, & \text{with prob. } p_2 \\ a_{n(l_n+1)}, & \text{with prob. } p_3 \end{cases} . \tag{7.116}$$

The probabilities p_1, p_2, p_3 ensure that the transition of elements are equally likely, i.e.

$$p_1 = \begin{cases} 0, & \text{if } a_{nl_n} = -1 \\ \frac{1}{2}, & \text{if } a_{nl_n} = 1 \\ \frac{1}{3}, & \text{otherwise} \end{cases} ; \quad p_2 = \begin{cases} \frac{1}{2}, & \text{if } a_{nl_n} = -1 \\ \frac{1}{2}, & \text{if } a_{nl_n} = 1 \\ \frac{1}{3}, & \text{otherwise} \end{cases} ; \quad p_3 = \begin{cases} \frac{1}{2}, & \text{if } a_{nl_n} = -1 \\ 0, & \text{if } a_{nl_n} = 1 \\ \frac{1}{3}, & \text{otherwise} \end{cases}$$

If we defined the neighbourhood of $\boldsymbol{a}_l$ by $\aleph(\boldsymbol{a}_l)$, to be the set of all the neighbours of $\boldsymbol{a}_l$, then the transition to any point in the neighbourhood is equally likely with a probability $\frac{1}{\#(\aleph(\boldsymbol{a}_l))}$ where $\#(\aleph(\boldsymbol{a}_l))$ denotes the number of elements of the neighbourhood $\aleph(\boldsymbol{a}_l)$.

4. Set the initial "temperature"parameter θ_k to θ_1, i.e. $k = 1$.

5. Set the number of simulation count $m = 0$.

6. Select an initial member $\boldsymbol{a}_l(\theta_k, m)$ which satisfies the constraints.

7. According to the transition probability $\frac{1}{\#[\aleph(\boldsymbol{a}_l(\theta_k,m))]}$, we arrive at a member $\boldsymbol{a}_{l'}(\theta_k, m)$ in the neighbourhood.

8. Test the constraints with $\boldsymbol{a}_{l'}(\theta_k, m)$.

9. If $\boldsymbol{a}_{l'}(\theta_k, m)$ satisfies the constraints, then go to step 11. Otherwise, set $\boldsymbol{a}_l(\theta_k, m + 1) = \boldsymbol{a}_l(\theta_k, m)$, and $m = m + 1$.

10. If $m \leq M_1$, go to step 7; otherwise go to step 13.

11. Determine $\boldsymbol{a}_l(\theta_k, m + 1)$ according to the following probabilities

$$\begin{aligned} \boldsymbol{a}_l(\theta_k, m+1) &= \begin{cases} \boldsymbol{a}_{l'}(\theta_k, m) & \text{with prob. } p_0 \quad \text{if } \boldsymbol{a}_{l'}(\theta_k, m) \neq \boldsymbol{a}_l(\theta_k, m) \\ \boldsymbol{a}_l(\theta_k, m) & \text{with prob. } 1 - p_0 \quad \text{if } \boldsymbol{a}_{l'}(\theta_k, m) \neq \boldsymbol{a}_l(\theta_k, m) \end{cases} \\ \boldsymbol{a}_l(\theta_k, m+1) &= \boldsymbol{a}_l(\theta_k, m) \qquad \text{if } \boldsymbol{a}_{l'}(\theta_k, m) = \boldsymbol{a}_l(\theta_k, m) \end{aligned}$$

where

$$p_0 = exp\left\{ \frac{1}{\theta_k} \left[\Delta T'^2 \Delta F'^2 \Big|_{\boldsymbol{a}'_l} - \Delta T'^2 \Delta F'^2 \Big|_{\boldsymbol{a}_l} \right]^- \right\}, \tag{7.117}$$

with the superscript notation being

$$[X]^- = \begin{cases} X, & \text{if } X < 0 \\ 0, & \text{otherwise} \end{cases}. \tag{7.118}$$

12. Set $m = m + 1$; if $m \leq M_1$, go to step 7, otherwise go to step 13.

13. Set $\theta_k = \theta_{k+1}$, if $k = K$ then stop; otherwise go to step 5.

The above is an outline of the quasi-simulated annealing (QSA) method. It has been proved [40] that if $\Delta T^2 \Delta F^2 \leq C$, and $\{\boldsymbol{a}_{op}\}$ is the set of global optimum vectors, then with

$$\theta_k = \frac{C}{\ln(k+1)} \tag{7.119}$$

and large enough M_1, the method will converge to the global optimum with probability 1, i.e.,

$$\lim_{k \to \infty} P(\boldsymbol{a}_l(\theta_k, M_1) \in \{\boldsymbol{a}_{op}\}) = 1. \tag{7.120}$$

7.3.2 Iterative Signal Synthesis (ISS) Using WD

In Chapter 6, we have seen how the Cramér-Rao bound of the joint estimates of t_0 and f_Δ relates to the sharpness of the peak of the WD $W_a(t,f)|_{\substack{t=0 \\ f=0}}$. Thus, we may use this fact to impose specifications on the WD of the signal such that it has very narrow mainlobe and low sidelobe amplitudes, and then try to synthesize the signal from the specified WD. The method proposed here is based on the discrete WD synthesis algorithm proposed in [13] [14]. Various researches [32] [76] have examined similar problems from different perspectives. The method used here can be summarized as follows:

1. Set n, the number of iterative synthesis to unity.

2. Specify an initial WD $W_a^{(n)}(m,k)$ so that the area of the mainlobe of the WD contains L_0 discrete points in the immediate neighbourhood $\aleph_{00}$ of $t=0,\ f=0$ on the time-frequency plane, and specify the amplitude of $W_a^{(n)}(0,0)$ to be constant for all iterations.

3. The WD of the points outside the mainlobe area are specified such that

$$\frac{W_a^{(n)}(m,k)}{W_a^{(n)}(0,0)} \leq \gamma \ , \ (m,k) \notin \aleph_{00}. \tag{7.121}$$

4. Use the WD synthesis algorithm described in [13] [14] to obtain a discrete signal $\hat{a}^{(n)}(i)$ whose discrete WD is the least square approximation to the specification $W_a^{(n)}(m,k)$ in the following way:

   (a) Compute $Y^{(n)}(m,j) = IFFT_k[W_a^{(n)}(m,k)]$; $m = 0,\cdots,N-1; j,k = -N/2+1,\cdots,N/2$, where $IFFT_k[\cdot]$ denotes the inverse fast Fourier transform with respect to k and returns the result with the sequence number j.

   (b) Form the following matrices

$$\begin{aligned} C_e^{(n)}(p,q) &= 2Y^{(n)}(p+q,p-q) \\ C_o^{(n)}(p,q) &= 2Y^{(n)}(p+q+1,p-q) \\ & \qquad p,q = 0,\cdots,\frac{N}{2}-1. \end{aligned} \tag{7.122}$$

   (c) Calculate the maximum eigenvalues $\lambda_{me}^{(n)}$ and $\lambda_{mo}^{(n)}$ and the corresponding eigenvectors $\boldsymbol{v}_{me}^{(n)}$ and $\boldsymbol{v}_{mo}^{(n)}$ of the matrices $\boldsymbol{C}_e$ and $\boldsymbol{C}_o$ respectively.

   (d) Form the discrete synthesized signal such that

$$\begin{aligned} \hat{a}^{(n)}(2i) &= \sqrt{\frac{\lambda_e^{(n)}}{2}}v_{me}^{(n)}(i) \\ \hat{a}^{(n)}(2i-1) &= \sqrt{\frac{\lambda_o^{(n)}}{2}}v_{mo}^{(n)}(i), \end{aligned} \tag{7.123}$$

   where $v_{me}^{(n)}(i)$ and $v_{mo}^{(n)}(i)$ are the ith elements of $\boldsymbol{v}_{me}^{(n)}$ and $\boldsymbol{v}_{mo}^{(n)}$ respectively.

5. Since the signal is synthesized from a least-square fit of the specified WD, often, it does not satisfy the specifications at this point. We compute the discrete WD $W_{\hat{a}}^{(n)}(m,k)$ of the synthesized signal $\hat{a}^{(n)}(i)$. We examine all the points in $W_{\hat{a}}^{(n)}(m,k)$ and select those points which do not satisfy the specification. Force these points to conform to the specifications resulting in another specified WD $W_a^{(n+1)}(m,k)$. Set $n = n+1$. Go back to step 4.

4. Use the WD synthesis algorithm described in [13] [14] to obtain a discrete signal $\hat{a}^{(n)}(i)$ whose discrete WD is the least square approximation to the specification $W_a^{(n)}(m,k)$ in the following way:

   (a) Compute $Y^{(n)}(m,j) = IFFT_k[W_a^{(n)}(m,k)]$; $m = 0,\cdots,N-1; j,k = -N/2+1,\cdots,N/2$, where $IFFT_k[\cdot]$ denotes the inverse fast Fourier transform with respect to k and returns the result with the sequence number j.

   (b) Form the following matrices

$$\begin{aligned} C_e^{(n)}(p,q) &= 2Y^{(n)}(p+q,p-q) \\ C_o^{(n)}(p,q) &= 2Y^{(n)}(p+q+1,p-q) \\ & \qquad p,q = 0,\cdots,\frac{N}{2}-1. \end{aligned} \tag{7.122}$$

   (c) Calculate the maximum eigenvalues $\lambda_{me}^{(n)}$ and $\lambda_{mo}^{(n)}$ and the corresponding eigenvectors $\boldsymbol{v}_{me}^{(n)}$ and $\boldsymbol{v}_{mo}^{(n)}$ of the matrices $\boldsymbol{C}_e$ and $\boldsymbol{C}_o$ respectively.

   (d) Form the discrete synthesized signal such that

$$\begin{aligned} \hat{a}^{(n)}(2i) &= \sqrt{\frac{\lambda_e^{(n)}}{2}}\, v_{me}^{(n)}(i) \\ \hat{a}^{(n)}(2i-1) &= \sqrt{\frac{\lambda_o^{(n)}}{2}}\, v_{mo}^{(n)}(i), \end{aligned} \tag{7.123}$$

   where $v_{me}^{(n)}(i)$ and $v_{mo}^{(n)}(i)$ are the ith elements of $\boldsymbol{v}_{me}^{(n)}$ and $\boldsymbol{v}_{mo}^{(n)}$ respectively.

5. Since the signal is synthesized from a least-square fit of the specified WD, often, it does not satisfy the specifications at this point. We compute the discrete WD $W_{\hat{a}}^{(n)}(m,k)$ of the synthesized signal $\hat{a}^{(n)}(i)$. We examine all the points in $W_{\hat{a}}^{(n)}(m,k)$ and select those points which do not satisfy the specification. Force these points to conform to the specifications resulting in another specified WD $W_a^{(n+1)}(m,k)$. Set $n = n+1$. Go back to step 4.

6. Repeat until $n > L$, then $\hat{a}^{(L)}(i)$ is the discrete form of the approximate optimum signal.

The two methods of obtaining an optimum signal under constraints have the following essential similarities and differences:

1. Both methods operate on the discrete WD of the signal to obtain a measure of the allowable sidelobe/mainlobe ratio.

2. While the ISS method tries to obtain a set of samples representing the optimum signal and thus implying that the optimum signal is expanded in a set of time-shifted band-limited sinc functions, the QSA approach tries to obtain a set of coefficients of the orthogonal expansion of the signal in which we are not restricted to any one set of basis functions.

3. While the QSA approach directly optimizes the objective function $\Delta T^2 \Delta F^2$, the ISS method can only stipulate the number of samples in the mainlobe of the discrete WD thereby indirectly controlling the sharpness of mainlobe.

4. While the QSA approach can take into account the energy concentration constraints in the essential time duration and the essential bandwidth of the signal, we cannot build these constraints into the ISS method.

5. ISS method is more than 100 times faster than QSA method.

7.3.3 Design Examples and performance of the Optimum Signals

In this section, we show some examples of constrained optimum signal design and compare the results obtained using the two methods described in the previous section. In all the examples, we have specified $T = 8sec.$, and $F = 4Hz$.

In all simulations of the QSA method, our "temperature" parameters are chosen heuristically to be

$$\theta_k = \frac{150}{50 + 100\ln(k+1)} \tag{7.124}$$

where $k = 1, \cdots, 100$, and for each "temperature", we perform $M_1 = 100,000$ simulations. In each case, we employ the new functions given by Eq. (7.104) as the basis function for the signal. We choose $N = 65$ in Eq. (7.76), and divide the range of elements a_n in 2000 steps, i.e., $L = 2001$ in Eq. (7.113). In all our examples, we specified the minimum energy concentrations in both time and frequency to be 97%, i.e.,

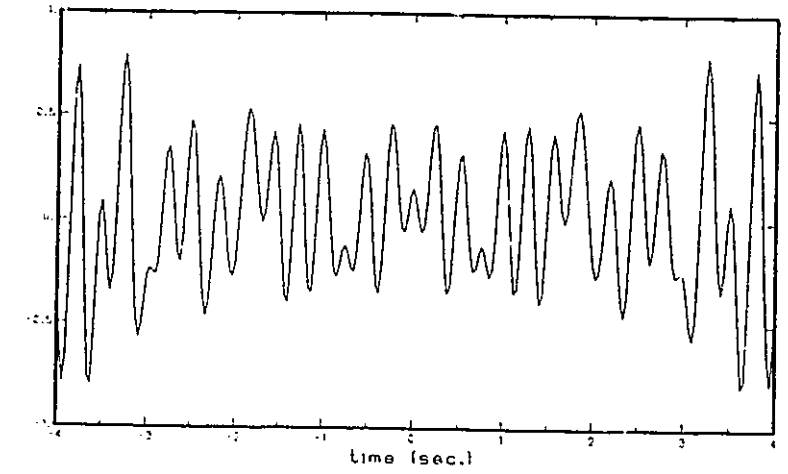
$$E_T = E_F = 0.97 \tag{7.125}$$

where E_T and E_F are defined in Eqs. (7.109) and (7.110) respectively. In the testing of the constraints, we also confine the mainlobe in the discrete WD to a single point at (0,0). The optimum signals obtained by the QSA approach is designated $\hat{a}_{o\gamma}(t)$ in which the subscript γ represents the allowable sidelobe/mainlobe ratio constraint under which the signal is designed.

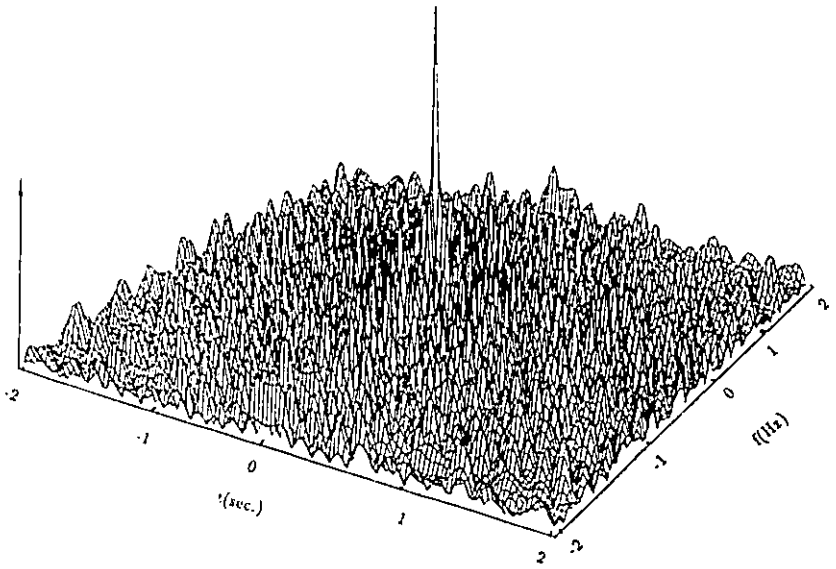
Fig. 7.10 a) shows the optimum time signal $\hat{a}_{o.3}(t)$ obtained by the simulated annealing method in which we choose $\gamma = 0.3$. The magnitude of its WD around $t = 0, f = 0$ is shown in Fig. 7.10b).

Fig.7.11 a) and b) show respectively the optimum signal $\hat{a}_{o.2}(t)$ in the time domain and the magnitude of its WD around $t = 0, f = 0$ when $\gamma = 0.20$. Table 7.1 shows a comparison of the parameters of these two signals.

In practice, we cannot lower the maximum sidelobe to mainlobe ratio γ in the WD any further than 0.2 without the mainlobe occupying more than one point in the time-frequency plane. Thus $\gamma = 0.2$ is the minimum under the conditions of our signal design. In general, the higher is the allowable ratio γ, the larger is the product $\Delta T^2 \Delta F^2$ of the optimum signal. As well, the higher is the allowable ratio γ, the smaller is the

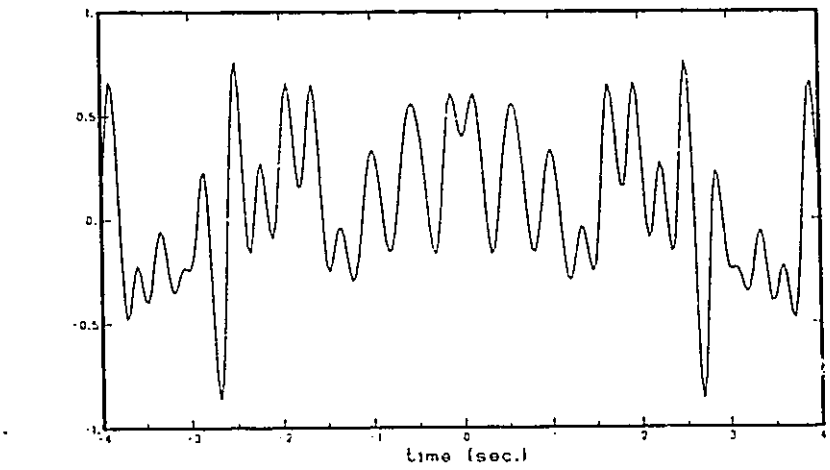


a)

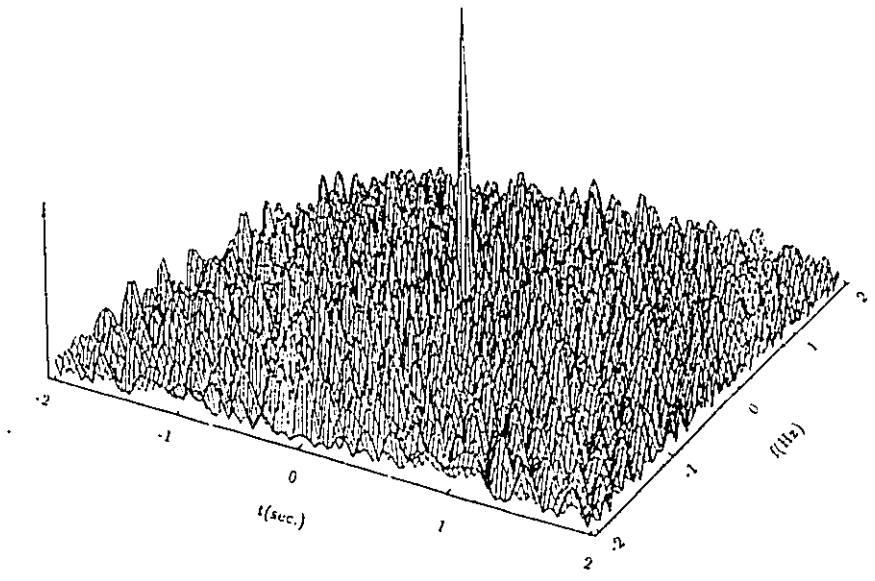


b)

Figure 7.10: a) The Optimum Signal $\hat{a}_{0.3}(t)$ Obtained by Simulated Annealing Method in Which We Choose $\gamma = 0.3$ and b) its Corresponding WD.



a)



b)

Figure 7.11: a) The Optimum Signal $\hat{a}_{0.2}(t)$ Obtained by Simulated Annealing Method in Which We Choose $\gamma = 0.2$ and b) its Corresponding WD.

Table 7.1: **Comparison of the Parameters of the Two Optimum Signals $\hat{a}_{o.2}(t)$ and $\hat{a}_{o.3}(t)$ Obtained by the QSA**

Allowable sidelobe/mainlobe ratio γ	Energy Concentration in essential duration E_T	Energy concentration in essential bandwidth E_F	$\Delta T^2 \Delta F^2$
0.2	99.4%	99.9%	25.4
0.3	99.2%	98.5 %	70.6

energy concentrations within the essential time duration and the essential bandwidth. Fig.7.12 shows the relationship between the values of $\Delta T^2 \Delta F^2$ of the optimum signal and allowable sidelobe/mainlobe ratio γ. All the signals designed with various values of γ satisfy the energy concentration of Eq. (7.125). However, while the optimum signal with $\gamma = 0.2$ over satisfies the energy concentration constraints, having over 99% of the energies within the essential time duration and essential bandwidth, the optimum signal with $\gamma \approx 1.0$ barely satisfies the constraints of Eq. (7.125), which can be explain as follows. While $\gamma \approx 1$, it is equivalent to that the sidelobe constraint is gradually lifted. and the maximum value $\Delta T^2 \Delta F^2$ gradually increases. The increase of $\Delta T^2 \Delta F^2$ means that the energy of $s(t)$ will concentrate more to the constraint boundary $|t| = \frac{T}{2}$ and $|f| = F$. Therefore, it becomes more difficult to constrain the energy into the given time duration and bandwidth.

From Fig.7.12, we can also see that the slope of the curve is very steep in the region of $0.2 \leq \gamma \leq 0.3$, and levels off beyond this region. This indicates that by relaxing the allowable sidelobe/mainlobe ratio from 0.2 to 0.3, we can gain a large increase in the estimation accuracy. This is done at the expense of increasing P_e, the probability of making an detection error. But from Figs.6.6, we can see that for SNR $\leq 15dB$, the difference between P_e at $\gamma = 0.2$ and P_e at $\gamma = 0.3$ is relative small, which implies this trade-off is quite reasonable at lower SNR.

Now, let us examine the results of the optimum signal design using the Iterative

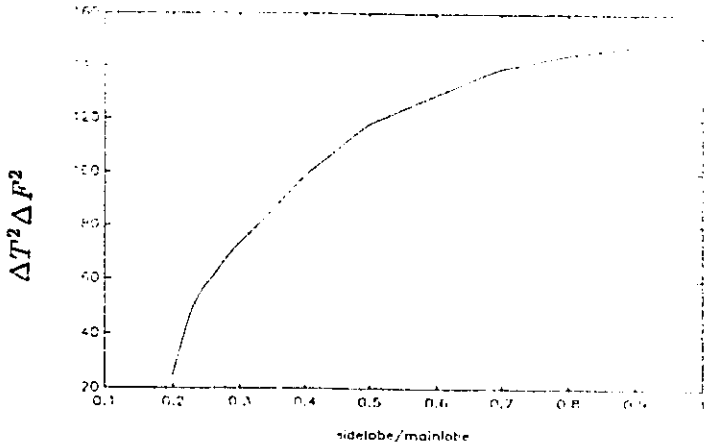
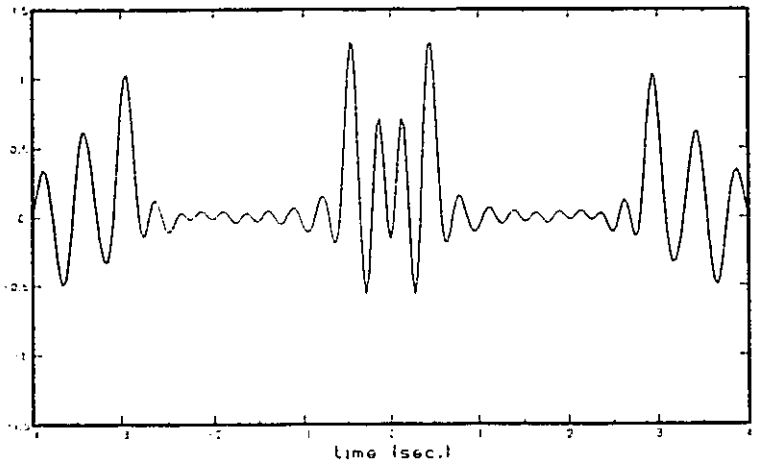


Figure 7.12: The Relationship Between the Value $\Delta T^2 \Delta F^2$ of the Optimum Signal and Allowable Sidelobe to Mainlobe Ratio γ.

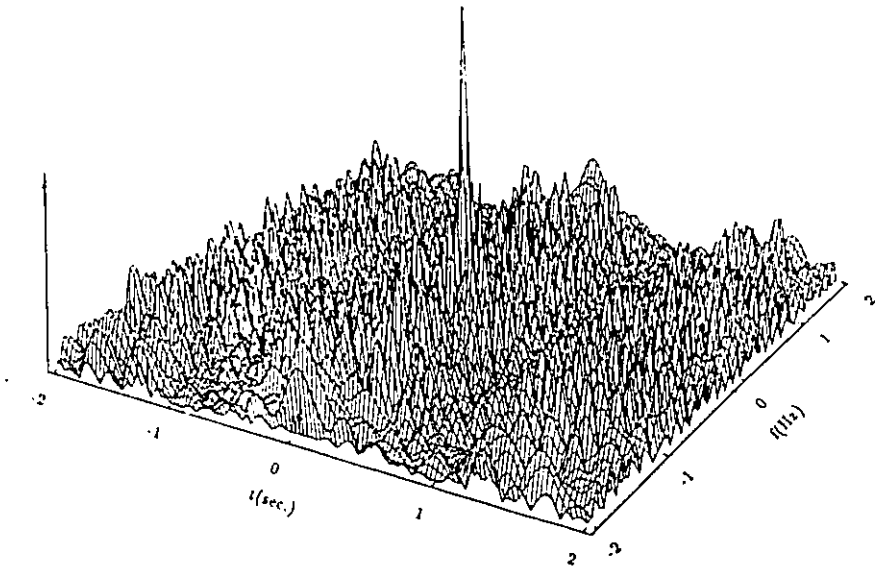
Table 7.2: **Comparison of the Parameters of the Two Optimum Signals $\tilde{a}_{o,2}(t)$ and $\tilde{a}_{o,3}(t)$ Obtained by the ISS Method**

Allowable sidelobe/mainlobe ratio γ	Energy Concentration in essential duration E_T	Energy concentration in essential bandwidth E_F	$\Delta T^2 \Delta F^2$
0.2	99.4%	100%	25.4
0.3	99.6%	100 %	25.9

Signal Synthesis Method. In all the examples here, we stipulate the mainlobe in the discrete WD of the signal to one single sample point at $t = 0$, $f = 0$. We have 64 time samples for the signal and we choose $L = 500$, i.e., the synthesis procedure is stopped after 500 iterations. We designate the optimum signal obtained by the ISS method $\tilde{a}_{o\gamma}(t)$. Fig.7.13a) shows the optimum signal $\tilde{a}_{o,3}(t)$ obtained by interpolating the time samples in the time domain with $\gamma = 0.30$ whereas Fig.7.13b) shows the magnitude of its WD around $t = 0, f = 0$. Fig. 7.14a) and b) show the optimum signal $\tilde{a}_{o,2}(t)$ in the time domain and the magnitude of its WD around $t = 0, f = 0$ respectively when $\gamma = 0.20$. Table 7.2 shows a comparison of the parameters of these two signals. Since the signals are obtained by interpreting the time samples with a sinc function which is band-limited to $\pm F$, the energy concentration is $E_F = 100\%$.

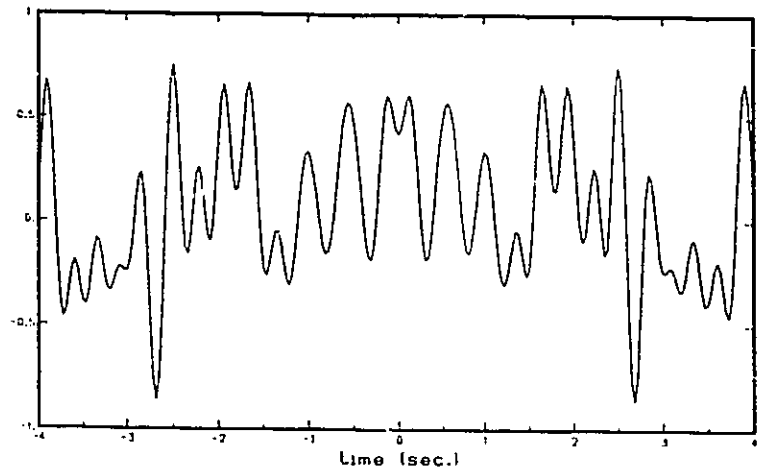


a)

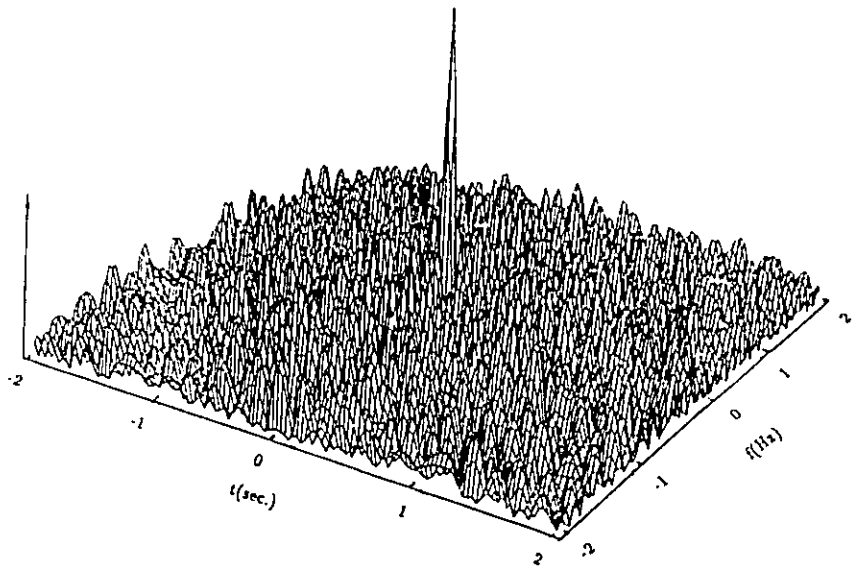


b)

Figure 7.13: a) The Optimum Signal $\tilde{a}_{0.3}(t)$ Obtained by ISS Method in Which We Choose $\gamma = 0.3$ and b) its Corresponding WD.



a)



b)

Figure 7.14: a) The Optimum Signal $\tilde{a}_{0.2}(t)$ Obtained by ISS Method in Which We Choose $\gamma = 0.2$ and b) its Corresponding WD.

In practice, it is very difficult to make γ to be exactly equal to a particular quantity since the method may not yield a signal that converges to a specified γ. In the examples above, we obtain the signal by deliberately lowering the specified value γ to be slightly less than the desired values of 0.3 and 0.2 and after 500 iterations, then the signals obtained have allowable sidelobe/mainlobe ratios very close to these desired values. Also, as in the case of using the QSA, we find that we cannot obtain a signal with γ less than 0.20 without increasing the number of points in the mainlobe.

Now, let us compare the signals obtained by the two different methods. Fig.7.10a) and Fig.7.13a) are two signals with the same constraint $\gamma = 0.3$ but designed by the QSA and the ISS method respectively. These two, as can be observed, are entirely different signals. Their mean square difference is calculated to be

$$\int_{-T/2}^{T/2} [\tilde{a}_{o.3}(t) - \tilde{a}_{o.3}(t)]^2 dt = 2.77. \tag{7.126}$$

The reason for the methods to end in entirely different signals is quite easy to understand. While the ratio γ only specifies the maximum allowable sidelobe level for a given mainlobe amplitude, the levels of the other sidelobes are unspecified and hence there exist many degrees of freedom in the design procedure. Thus, if the two approaches are basically different in several way as has been pointed out in the last section, it is not surprising that they end in two different results. Since the QSA approach seeks for the maximization of $\Delta T^2 \Delta F^2$, it can be assumed that this parameter of the signal obtained by the QSA is larger than that obtained by the ISS method as is evidenced by comparing the last columns of Table 7.1 and Table 7.2. On the other hand, Fig.7.11a) and Fig.7.14a) are two almost identical signals with the same constraint $\gamma = 0.20$ but designed by the two methods respectively. Their mean square difference is calculated to be

$$\int_{-T/2}^{T/2} [\tilde{a}_{o.2}(t) - \tilde{a}_{o.2}(t)]^2 dt = 7.322 \times 10^{-8}. \tag{7.127}$$

The almost identical results so obtained can be explained as follows: In both cases, we stipulate the mainlobe of the WD to contain only a single point at $t = 0, \quad f = 0$. However, in both cases, we find it impossible to lower the ratio γ any further without increasing the number of points in the mainlobe. Recalling that the volume under the WD is the total energy of the signal [20], and since $W_{aa}(0,0)$ is a constant, $\gamma = 0.20$ represents the minimum level of the sidelobes that we can attain so that the signal energy outside the mainlobe is distributed evenly between the sidelobes. When this is achieved, the optimum signal is resulted.

Let us now examine the performance of the various signals when applied to the joint estimation of time delay and Doppler shift. The signal processing system which we examine is as shown in Fig.6.1. We employ the peak location strategy on the WD $|W_{xa}(t,f)|^2$ as described in Chapter 6 in which the signal $x(t)$ is the delayed and Doppler shifted version of the signal $a(t)$ mixed with white Gaussian noise as expressed in Eq.(6.12). Since we limited our consideration to $a(t)$ being real and since the peak location method reaches the Cramér-Rao bound under high SNR, then from Eq.(6.67), the joint error of estimation is given by

$$Var(\hat{t}_0)Var(\hat{f}_\Delta) = \frac{N_0^2}{64\pi^4 \Delta T^2 \Delta F^2}. \tag{7.128}$$

Eq. (7.128) expresses the theoretical performance of a signal $a(t)$ when used to jointly estimate t_0 and f_Δ by the location of the peak of the WD $|W_a(t,f)|^2$. Fig.7.15 shows the theoretical performance of various signals under different SNR which is defined in the same form as Eq.(6.84). We compare the performance of seven signals: Gaussian signal $a_1(t)$ given by Eq.(6.85), Linear FM pulse $a_2(t)$ given by Eq.(6.86), $\hat{a}_{o.2}(t)$, $\tilde{a}_{o.2}(t)$, $\hat{a}_{o.3}(t)$, $\tilde{a}_{o.3}(t)$, as well as $\hat{a}_o(t)$ which is the optimum signal obtained with no energy concentration or sidelobe concentrations as in Fig.7.7a). As expected the unconstrained optimum signal $\hat{a}_o(t)$ yields the best performance followed by $\hat{a}_{o.3}(t)$ and $\tilde{a}_{o.3}(t)$. $\hat{a}_{o.2}(t)$ and $\tilde{a}_{o.2}(t)$ have identical performance as expected. All these signals yield better performance than $a_2(t)$

and far better than $a_1(t)$. In the evaluation of the performance here, we have ignored the finite probability P_e of making an erroneous estimation. Among the optimum signals, $\hat{a}_o(t)$ has the highest P_e whereas $\hat{a}_{o.3}(t)$ and $\tilde{a}_{o.3}(t)$ will have equal P_e being larger than that of $\hat{a}_{o.2}(t)$ and $\tilde{a}_{o.2}(t)$. We now test the performance of these signals using computer simulations. The different signals $a_1(t)$, $a_2(t)$, $\hat{a}_{0.2}(t)$, $\tilde{a}_{0.2}(t)$, $\hat{a}_{0.3}(t)$, $\tilde{a}_{0.3}(t)$ and $\hat{a}_0(t)$ are used to jointly estimate the TD and DS in white Gaussian noise. In each of the cases, the peak location strategy on the WD $|W_{za}(t,f)|$ is employed and the TD and DS are then determined accordingly as described in Section 6.5.1. The procedure is repeated 100 times under the same SNR for each signal, and the average $var(\hat{t}_0)var(\hat{f}_\Delta)$ is evaluated. Fig.7.16 shows the performance of these signals by computer simulations. At high SNR, the performance of the various signals are all in accordance to theoretical evaluation with $\hat{a}_0(t)$ having the best performance followed by $\hat{a}_{0.3}(t)$, $\tilde{a}_{0.3}(t)$, $\hat{a}_{0.2}(t)(=\tilde{a}_{0.2}(t))$, $a_2(t)$, and $a_1(t)$. At SNR$\approx 20dB$, a threshold effect occurs for $\hat{a}_0(t)$ while the rest of the signals follow more or less the theoretical predictions. This is due to the effect of the high sidelobes associated with the WD of $\hat{a}_0(t)$ such that a sidelobe has been erroneously identified as the mainlobe. At SNR$\approx 15dB$, all the signals experience the threshold effect. This is because the non-linear terms in the expansion of $W_{aa}(t,f)$ around $t=0$ and $f=0$ are no longer negligible.

In this chapter, we considered the problem of the joint estimation of time delay and Doppler shift. It has been shown that an essentially time limited and band-limited signal that has the minimum joint estimation error would have the largest time-bandwidth product $\Delta T^2 \Delta F^2$. The conditions for optimality were derived, and a very efficient method of obtaining the optimum signal was developed. Practical applications, however, demand the signal to have other constraints such as energy concentrations in the essential time duration and essential bandwidth as well as a maximum allowable

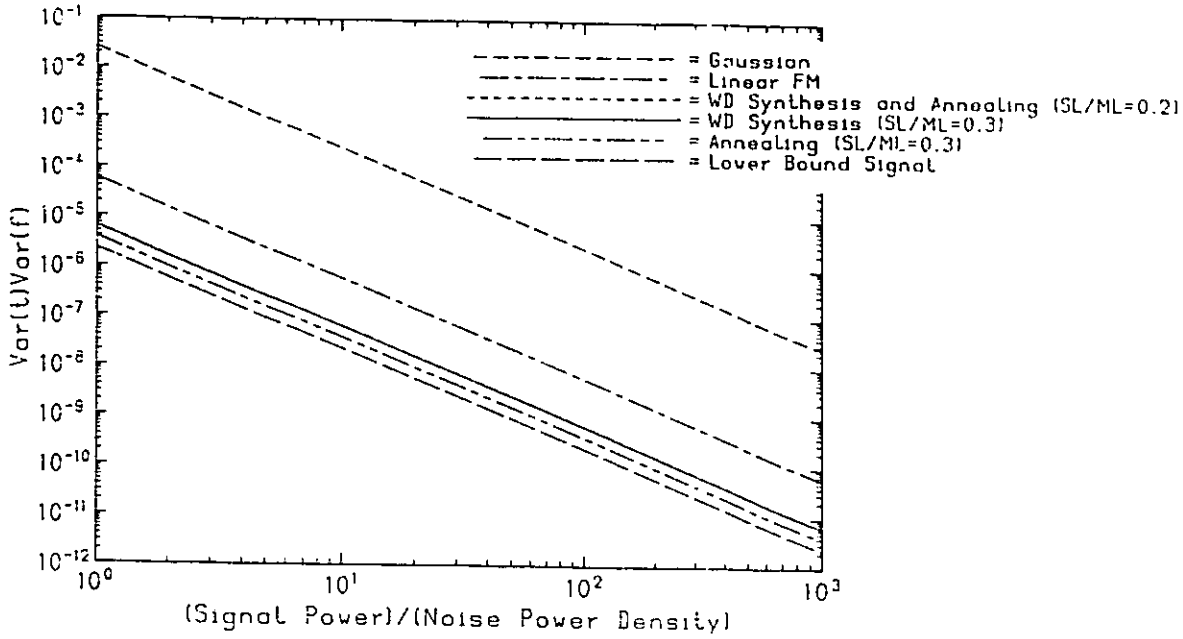


Figure 7.15: The Theoretical performance of Various Signals Under Different SNR.

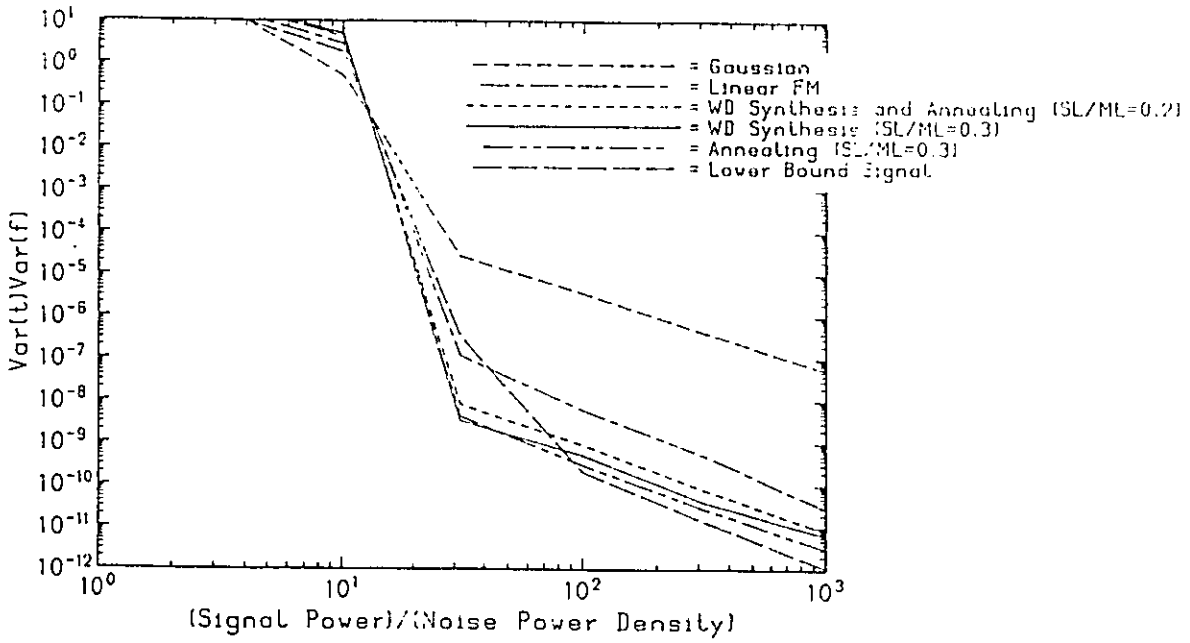


Figure 7.16: The Simulated performance of Various Signals Under Different SNR.

sidelobe/mainlobe level to maintain an acceptably low probability of erroneous detection. These constraints caused complications in obtaining an optimum signal. Two approaches, viz the Quasi-Simulated Annealing Method and the Iterative Signal Synthesis using the WD were proposed. The two approaches generally yield different signals with the QSA approach resulting in signals with lower joint estimation error (higher $\Delta T^2 \Delta F^2$). However, the QSA method is very time consuming while the ISS method is more than 100 times faster. At least in one case, the two methods yield identical results. In general, for the optimum signals so obtained, the more relaxed is the allowable sidelobe/mainlobe ratio, the smaller is the joint estimation error, but the higher is P_e, the probability of error detection. These optimum signals all showed superior performance to the commonly used linear FM pulse, and are far superior to the Gaussian pulse.

Chapter 8

Conclusion

In this thesis, we have discussed the properties and applications of time- frequency representations, which can process the nonstationary signal with time- dependent spectrum. The obvious advantages of time-frequency representations over classic spectrum can be seen from their respective application studies. Generally speaking, the WD possesses many advantageous properties and is widely used among all the representations. Hence, all over in this thesis, we concentrated mainly on the use of this method.

Although in this thesis, we are only concerned with some applications related with signal transmission and reception, their potential application area is far more than that.

One interesting research topic, which we have not mentioned, is whether time- frequency representations can be applied to the radar reflection signal, so that the signal can be expanded onto the two dimensional plane and more target parameters, such as size, structure and materials, etc., can be included. W.Martin and P. Flandrin [66] have used the WD to analyze the change of signal structure. In the research of bridge deck fault detection, the WD has been applied to detect the fault of bridge deck by observing the time-dependent spectrum of the impulse radar reflecting signal.

Another interesting question is whether there exists a representation which can represent the time-dependent spectrum better than all existing representations. Although

the WD is widely used, it has the main drawback that it is not a linear transform so that it contains the cross-interference terms when more than one signals present, as discussed in section 4.3.4. W.J. Williams and others [98] have developed a new distribution, which is called Reduced Interference Distribution (RID). By careful selection of the kernel function in Eq.(1.10), the RID can eliminate the cross interferences and still retain most properties of the WD. Recently, a new linear transform, wavelet transform, has attracted a lot of research interest. In some aspects, it may possess more advantageous properties than the WD. The potential applications are still under investigation and how to obtain the optimum wavelet is a still an unsolved problem [24].

Besides the time-frequency representations, in this thesis, we have also discussed the uncertainty principle in Chapter 7. We try to investigate one of the basic laws in nature, the uncertainty in time and frequency domain. The optimum basis derived in the Section 7.1 may have profound implications in the communication theory. It presents the maximum number of orthogonal signals when the product of percentage energy in the given time slot and bandwidth is fixed, and in another way, when the number of orthogonal signals is fixed, it yields the maximum product of percentage energy in the given time slot and bandwidth [49] and [50]. This extends the previous optimum channel coding theory [108] which deals only with the band-limited signals and considers their possible concentration in given time slot.

Appendix A

Gabor Function and the Suitable Window for the Spectrogram

In order that the definition of the spectrogram given in Eqs.(2.8) and (2.7) can truly reflect the instantaneous spectrum, the selection of suitable $h(t)$ is important. The reasons can be explained as followings.

If the signal $s(t)$ changes fast, the width of time window $h(t)$ should be shortened appropriately, so that $G_s(t,f)$ really gives out the instantaneous spectrum of $s(t)$ at time t. However, according to the well-known property of the Fourier transform, the short window width means the loss of frequency resolution. In order to have a trade off between the instantaneous time width and frequency resolution of the spectrogram, it is crucial to select a suitable window, which has minimum support area in both time and frequency. In general, the time and frequency support areas are defined, respectively, by its r.m.s. (root mean square) duration T_h and r.m.s. frequency F_h, where

$$T_h = \left[\frac{\int_{-\infty}^{\infty}(t-\bar{t})^2 \mid h(t) \mid^2 dt}{\int_{-\infty}^{\infty} \mid h(t) \mid^2 dt}\right]^{\frac{1}{2}}, \tag{A.1}$$

$$F_h = \left[\frac{\int_{-\infty}^{\infty}(f-\bar{f})^2 \mid H(f) \mid^2 df}{\int_{-\infty}^{\infty} \mid H(f) \mid^2 df}\right]^{\frac{1}{2}}, \tag{A.2}$$

and

$$\bar{t} = \int_{-\infty}^{\infty} t \mid h(t) \mid^2 dt,$$
$$\bar{f} = \int_{-\infty}^{\infty} f \mid H(f) \mid^2 df.$$

D. Gabor proved in 1946 [36] that the Gaussian function:

$$h(t) = (2\pi\sigma_g^2)^{-\frac{1}{4}} e^{-\frac{t^2}{4\sigma_g^2}}, \tag{A.3}$$

has the minimum product of r.m.s duration and r.m.s. frequency, i.e. minimum $T_h F_h$. The proof can be stated as followings:

Proof: Without loss of generality, we can choose $\bar{t} = 0$ and $\bar{f} = 0$, and the inner product is:

$$< h, h > = < H, H > = 1. \tag{A.4}$$

Therefore the r.m.s duration-frequency product of $h(t)$ is

$$\begin{aligned} T_h F_h &= \left[\frac{< th, th >< fH, fH >}{< h, h >< H, H >} \right]^{\frac{1}{2}} \\ &= \left[\frac{1}{4\pi^2} < th, th >< \dot{h}, \dot{h} > \right]^{\frac{1}{2}} \end{aligned} \tag{A.5}$$

where $\dot{h}$ represents the derivative of h with respect to time t and in the second equality of Eq.(A.5), the Parseval's theorem, $< 2\pi j f H, 2\pi j f H > = < \dot{h}, \dot{h} >$, is used.

Applying the Schwarz inequality to Eq.(A.5), we have

$$T_h F_h = \frac{1}{2\pi} \| th \| \| \dot{h} \| \geq \frac{1}{2\pi} \mid < th, \dot{h} > \mid. \tag{A.6}$$

The right-hand term in Eq.(A.6) can be evaluated by writing $h(t) = |h| e^{j\phi_h}$ so that

$$\begin{aligned} < th, \dot{h} > &= \int_{-\infty}^{\infty} th(t) \dot{h}^*(t) dt \\ &= \int_{-\infty}^{\infty} t|h| \frac{d|h|}{dt} dt - j \int_{-\infty}^{\infty} t|h|^2 \frac{d\phi_h}{dt} dt. \end{aligned} \tag{A.7}$$

If there is no phase modulation, or if the phase is linear with time t, then $\frac{d\phi_h}{dt}$ is a constant, and the second term in Eq.(A.7) is the first moment of $|h(t)|^2$ which is assumed to be zero. The first term in Eq.(A.7) can be evaluated by integrating by parts such that,

$$\int_{-\infty}^{\infty} t|h|\frac{d|h|}{dt}dt = \frac{1}{2}\int_{-\infty}^{\infty} t\frac{d|h|^2}{dt}dt = -\frac{1}{2}\int_{-\infty}^{\infty}|h(t)|^2dt = -\frac{1}{2} \tag{A.8}$$

where it is assumed that $t\frac{d|h(t)|^2}{dt} \rightarrow 0$ when $|t| \rightarrow \infty$. Hence Eq.(A.6) becomes

$$T_h F_h \geq \frac{1}{4\pi}. \tag{A.9}$$

Equality holds if $\dot{h}(t) = k_1 t h(t)$ with k_1 being a constant number, i.e., if

$$\frac{d}{dt}[\ln h(t)] = k_1 t \tag{A.10}$$

The solution of Eq.(A.10) is

$$h(t) = k_2 e^{k_1 \frac{t^2}{2}} \tag{A.11}$$

where k_2 is a constant number which makes $|h(t)| = 1$. Thus the Gaussian function is optimum in the sense that it minimizes the r.m.s duration- frequency product and obtain the lower bound $\frac{1}{4\pi}$. □

Due to this important property of the Gaussian function, in the spectrogram, it is usually chosen as a window function. Since Gabor [36] first proved the above relationship, the function in Eq.(A.11) is also called the Gabor function while used in the spectrogram as the optimum window function. The relation given in Eq.(A.9) is also a special form of the Heisenberg's uncertainty principle [25].

Appendix B

Proof of Eq.(3.45)

Since J^{ij} is the ijth element of $\boldsymbol{J}^{-1}$, it is sufficient to show that if we use the expression of J^{ij} given by Eq.(3.45), then

$$\sum_{k=1}^{K} J_{ik} J^{kj} = \delta_{ij}. \tag{B.1}$$

Now,

$$\begin{aligned}
\sum_{k=1}^{K} J_{ik} J^{kj} &= \sum_{k=1}^{K} \int\int\int\int_{T_S}^{T_F} J_K(t_1,t_2) J_K^{-1}(t_3,t_4) \cdot \\
&\quad \psi_{gi}(t_1)\psi_{gk}(t_2)\psi_{gk}(t_3)\psi_{gj}(t_4) dt_1 dt_2 dt_3 dt_4 \\
&= \int\int\int\int_{T_S}^{T_F} J_K(t_1,t_2) J_K^{-1}(t_3,t_4) \psi_{gi}(t_1)\psi_{gj}(t_4) \cdot \\
&\quad \left[\sum_{k=1}^{K} \psi_{gk}(t_2)\psi_{gk}(t_3)\right] dt_1 dt_2 dt_3 dt_4.
\end{aligned} \tag{B.2}$$

Using Eq.(3.46) on the bracketed term, we have

$$\begin{aligned}
\sum_{k=1}^{K} J_{ik} J^{kj} &= \int\int\int\int\int_{T_S}^{T_F} J_K(t_1,t_2) J_K^{-1}(t_3,t_4) J_K^{-1}(t_2,t_5) J_K(t_5,t_3) \cdot \\
&\quad \psi_{gi}(t_1)\psi_{gj}(t_4) dt_1 dt_2 dt_3 dt_4 dt_5 \\
&= \int\int\int \psi_{gi}(t_1)\psi_{gj}(t_4) \left[\int J_K(t_1,t_2) J_K^{-1}(t_2,t_5) dt_2\right] \cdot \\
&\quad \left[\int J_K(t_5,t_3) J_K^{-1}(t_3,t_4) dt_3\right] dt_1 dt_4 dt_5 \\
&= \int\int\int \psi_{gi}(t_1)\psi_{gj}(t_4) \left[\sum_{m=1}^{K} \psi_{gm}(t_1)\psi_{gm}(t_5)\right] \cdot
\end{aligned} \tag{B.3}$$

$$\left[\sum_{l=1}^{K} \psi_{gl}(t_4)\psi_{gl}(t_5)\right] dt_1 dt_4 dt_5, \tag{B.4}$$

where the last step is arrived by applying Eq.(3.46) on the two bracketed terms in Eq. (B.3). Integrating with respect to t_5 and using the orthonormality of $\psi_g(t_5)$, the double sum in Eq. (B.4) becomes a single sum, and repeating these steps again, we obtain

$$\begin{aligned} \sum_{k=1}^{K} J_{ik}J^{kj} &= \int\int \psi_{gi}(t_1)\psi_{gj}(t_4)\left[\sum_{m=1}^{K} \psi_{gm}(t_1)\psi_{gm}(t_4)\right] dt_1 dt_4 \\ &= \int \psi_{gi}(t_1)\psi_{gj}(t_1)dt_1 = \delta_{ij}. \end{aligned} \tag{B.5}$$

Appendix C

Simplification of Eq.(3.51)

To simplify the first double integral on the LHS of Eq.(3.51), we use the definitions of Eqs.(3.39) and (3.40) and employing the orthonormality of eigenfunctions, we have

$$\begin{aligned}
&\int\int J_K^{-1}(t,u)Q_{gK}(u,t_1)R_{gK}(t_1,\tau)dudt_1 \\
&= \int\int J_K^{-1}(t,u)\left[\sum_{k=1}^{K}\frac{1}{\lambda_{gk}}\psi_{gk}(u)\psi_{gk}(t_1)\right]\left[\sum_{l=1}^{K}\lambda_{gl}\psi_{gl}(t_1)\psi_{gl}(\tau)\right]dudt_1 \\
&= \int J_K^{-1}(t,u)\left[\sum_{k=1}^{K}\psi_{gk}(u)\psi_{gk}(\tau)\right]du. \qquad \text{(C.1)}
\end{aligned}$$

We want to show that the RHS of Eq. (C.1) is equal to $J_K^{-1}(t,\tau)$. To do that we multiply $J_K^{-1}(t,\tau)$ by $J_K(\tau,\nu)$ and integrate with respect to τ. We obtain

$$\int J_K^{-1}(t,\tau)J_K(\tau,\nu)d\tau = \sum_{i=1}^{K}\psi_{gi}(t)\psi_{gi}(\nu), \qquad \text{(C.2)}$$

where Eq.(3.46) has been used. Also multiply the RHS of Eq. (C.1) by $J_K(\tau,\nu)$ and integrate withe respect to τ, we obtain

$$\begin{aligned}
&\int\int J_K^{-1}(t,u)J_K(\tau,\nu)\left[\sum_{k=1}^{K}\psi_{gk}(u)\psi_{gk}(\tau)\right]dud\tau \\
&= \int\int J_K^{-1}(t,u)J_K(\tau,\nu)\int J_K(u,\omega)J_K^{-1}(\omega,\tau)d\omega dud\tau \\
&= \int\left[\int J_K^{-1}(t,u)J_K(u,\omega)du\right]\left[\int J_K^{-1}(\omega,\tau)J_K(\tau,\nu)d\tau\right]d\omega \\
&= \int\left[\sum_{i=1}^{K}\psi_{gi}(t)\psi_{gi}(\omega)\right]\left[\sum_{k=1}^{K}\psi_{gk}(\omega)\psi_{gk}(\nu)\right]d\omega
\end{aligned}$$

$$= \sum_{i=1}^{K} \psi_{gi}(t)\psi_{gi}(\nu). \tag{C.3}$$

The last step in Eq. (C.3) is arrived at using the orthonormality of the eigenfunctions. Therefor, Eq. (C.2) and Eq. (C.3) have identical RHS.

To further simplify the RHS of Eq.(3.51), we use Eq.(3.39), and obtain

$$\int \left[\sum_{k=1}^{K} \psi_{gk}(t)\psi_{gk}(t_1)\right] \left[\sum_{l=1}^{K} \lambda_{gl}\psi_{gl}(t_1)\psi_{gl}(\tau)\right] dt_1$$

$$= \sum_{k=1}^{K} \lambda_{gk}\psi_{gk}(t)\psi_{gk}(\tau) = R_{gK}(t,\tau). \tag{C.4}$$

Hence, the simplification of the RHS of Eq.(3.51) is proved.

Appendix D

Evaluation of the Variance of the Estimated Instantaneous Frequency Using WD

For complex white Gaussian noise, we have the following properties[35]:

$$E[n(t_1)n(t_2)] = 0, \tag{D.1}$$

$$E[n(t_1)n^*(t_2)] = N_0\delta(t_1 - t_2), \tag{D.2}$$

$$E[n(t_1)n^*(t_2)n(t_3)] = E[n^*(t_1)n(t_2)n^*(t_3)] = 0, \tag{D.3}$$

and [35]

$$\begin{aligned} E[n(t_1)n^*(t_2)n(t_3)n^*(t_4)] &= E[n(t_1)n^*(t_2)]E[n(t_3)n^*(t_4)] + E[n(t_1)n(t_3)]E[n^*(t_2)n^*(t_4)] \\ &\quad + E[n(t_1)n^*(t_4)]E[n(t_3)n^*(t_2)]. \end{aligned} \tag{D.4}$$

Using Eqs. (D.1) and (D.2), we have

$$E[n(t_1)n^*(t_2)n(t_3)n^*(t_4)] = N_0^2[\delta(t_1 - t_2)\delta(t_3 - t_4) + \delta(t_1 - t_4)\delta(t_3 - t_2)]. \tag{D.5}$$

Therefore, if the signal exists in the interval $[-\frac{T_I}{2}, \frac{T_I}{2}]$, we can write the following:

$$E\left[\left(\frac{\partial W_{sn}}{\partial f}\right)^2\right] = -4\pi^2 \int\int_{-\infty}^{\infty} \tau_1\tau_2 s(t_0 + \frac{\tau_1}{2})s(t_0 + \frac{\tau_2}{2})$$

$$\cdot E\left[n^*(t_0-\frac{\tau_1}{2})n^*(t_0-\frac{\tau_2}{2})\right]\cdot e^{-j2\pi f(\tau_1+\tau_2)}d\tau_1 d\tau_2$$

$$= 0, \tag{D.6}$$

where Eq. (D.1) has been used. similarly,

$$E\left[\left(\frac{\partial W_{sn}^*}{\partial f}\right)^2\right]=0. \tag{D.7}$$

Also,

$$\begin{aligned} & E\left[\left(\frac{\partial W_n}{\partial f}\right)\left(\frac{\partial W_{sn}}{\partial f}\right)\right] \\ = & -4\pi^2\int\int_{-\infty}^{\infty}\tau_1\tau_2 s(t_0+\frac{\tau_2}{2})\cdot E\left[n(t_0+\frac{\tau_1}{2})n^*(t_0-\frac{\tau_1}{2})n^*(t_0-\frac{\tau_2}{2})\right]\cdot e^{-j2\pi f(\tau_1+\tau_2)}d\tau_1 d\tau_2 \\ = & 0, \end{aligned} \tag{D.8}$$

where Eq. (D.3) has been used, and similarly,

$$E\left[\left(\frac{\partial W_n}{\partial f}\right)\left(\frac{\partial W_{sn}^*}{\partial f}\right)\right]=0. \tag{D.9}$$

Furthermore,

$$\begin{aligned} & E\left[\left|\frac{\partial W_{sn}}{\partial f}\right|^2\right] \\ = & 4\pi^2\int\int_{-\infty}^{\infty}\tau_1\tau_2 s(t_0+\frac{\tau_1}{2})s^*(t_0+\frac{\tau_2}{2})\cdot E\left[n^*(t_0-\frac{\tau_1}{2})n(t_0-\frac{\tau_2}{2})\right]\cdot e^{-j2\pi f(\tau_1-\tau_2)}d\tau_1 d\tau_2 \\ = & 4\pi^2\int\int_{-\infty}^{\infty}\tau_1\tau_2 s(t_0+\frac{\tau_1}{2})s^*(t_0+\frac{\tau_2}{2})\cdot\left[N_0\delta\left(\frac{\tau_1-\tau_2}{2}\right)\right]\cdot e^{-j2\pi(\tau_1-\tau_2)}d\tau_1 d\tau_2 \\ = & 8\pi^2 N_0\int_{-\infty}^{\infty}\tau_1^2\mid s(t_0+\frac{\tau_1}{2})\mid^2 d\tau_1. \end{aligned} \tag{D.10}$$

Notice that

$$\int\delta(\frac{\tau}{2})d\tau=2.$$

For $s(t)$ given by Eq.(4.8)

$$\mid s(t_0+\frac{\tau_1}{2})\mid^2= A^2,$$

and since the signal duration is finite, Eq. (D.10) becomes

$$E\left[\left|\frac{\partial W_{sn}}{\partial f}\right|^2\right]=8\pi^2A^2N_0\int_{-\frac{\tau_I}{2}}^{\frac{\tau_I}{2}}\tau_1^2d\tau_1=\frac{2\pi^2A^2N_0T_I^3}{3}. \tag{D.11}$$

Now, let us turn our attention to the evaluation of the first term in the numerator of Eq.(4.22). We have

$$
\begin{aligned}
& E\left[\left|\frac{\partial W_n}{\partial f}\right|^2\right] \\
= \; & 4\pi^2 \int\int_{-\frac{T_I}{2}}^{\frac{T_I}{2}} \tau_1\tau_2 E\left[n(t_0+\frac{\tau_1}{2})n^*(t_0-\frac{\tau_1}{2})\cdot n^*(t_0+\frac{\tau_2}{2})n(t_0-\frac{\tau_2}{2})\right]\cdot e^{-j2\pi f(\tau_1-\tau_2)}d\tau_1 d\tau_2 \\
= \; & 4\pi^2 N_0^2 \int\int_{-\frac{T_I}{2}}^{\frac{T_I}{2}} \tau_1\tau_2\left[\delta(\frac{\tau_1}{2})\delta(\frac{\tau_2}{2})+\delta(\frac{\tau_1-\tau_2}{2})\delta(\frac{\tau_1-\tau_2}{2})\right] e^{-j2\pi f(\tau_1-\tau_2)}d\tau_1 d\tau_2. \qquad \text{(D.12)}
\end{aligned}
$$

Since $\delta(\frac{\tau_2}{2})$ has value only at $\tau_2 = 0$, the integration of the first term with respect to τ_2 vanished so that Eq. (D.12) is reduced to

$$
E\left[\left|\frac{\partial W_n}{\partial f}\right|^2\right] = 4\pi^2 N_0^2 \int\int_{-\frac{T_I}{2}}^{\frac{T_I}{2}} \tau_1\tau_2\delta(\frac{\tau_1-\tau_2}{2})\delta(\frac{\tau_1-\tau_2}{2})e^{-j2\pi f(\tau_1-\tau_2)}d\tau_1 d\tau_2. \qquad \text{(D.13)}
$$

Strictly speaking, Eq. (D.13) cannot be evaluated since the product of two impulse functions is not defined. This difficulty is caused by the definition of the autocorrelation of white noise which implies that the mean power of white noise is infinite. However, since in most engineering applications, the bandwidth is finite, we can assign a finite bandwidth $[-B_n, B_n]$ to the system so that

$$
E[n(t_1)n^*(t_2)] = 2N_0B_n\frac{\sin 2\pi B_n(t_1-t_2)}{2\pi B_n(t_1-t_2)}. \qquad \text{(D.14)}
$$

Using Eq. (D.14) instead of Eq. (D.2), and realizing that

$$
2B_nN_0 = \sigma_n^2
$$

where σ_n^2 is the noise power, Eq. (D.13) becomes

$$
E\left[\left|\frac{\partial W_n}{\partial f}\right|^2\right] = 4\pi^2\sigma_n^4 \int\int_{-\frac{T_I}{2}}^{\frac{T_I}{2}} \tau_1\tau_2 e^{-j2\pi f(\tau_1-\tau_2)}\cdot\frac{\sin^2\pi B_n(\tau_1-\tau_2)}{\{\pi B_n(\tau_1-\tau_2)\}^2}d\tau_1 d\tau_2. \qquad \text{(D.15)}
$$

If B_n is sufficiently large, then the variation of $\tau_1\tau_2e^{-j2\pi(\tau_1-\tau_2)}$ is slow compared to

$$
\frac{\sin^2\pi B_n(\tau_1-\tau_2)}{\{\pi B_n(\tau_1-\tau_2)\}^2}.
$$

Therefore, the value of $\tau_1\tau_2 e^{-j2\pi(\tau_1-\tau_2)}$ can be considered to be concentrated at $\tau_2 = \tau_1$. Hence, Eq. (D.15) can be approximated to

$$E\left[\left|\frac{\partial W_n}{\partial f}\right|^2\right] \approx 4\pi^2\sigma_n^4 \int_{-\frac{T_I}{2}}^{\frac{T_I}{2}} \tau_1^2 \int_{-\frac{T_I}{2}}^{\frac{T_I}{2}} \frac{\sin^2 \pi B_n(\tau_1-\tau_2)}{\{\pi B_n(\tau_1-\tau_2)\}^2} d\tau_1 d\tau_2. \tag{D.16}$$

It is well known (see, e.g., [73]) that

$$\int_{-\infty}^{\infty} B_n \frac{\sin^2(\pi B_n t)}{(\pi B_n t)^2} dt = 1. \tag{D.17}$$

For B_n sufficiently large, the area under the inner integral in Eq. (D.16) can be considered to be completely covered within the range $[-\frac{T_I}{2}, \frac{T_I}{2}]$. Thus, Eq. (D.16) reduces to

$$E\left[\left|\frac{\partial W_n}{\partial f}\right|^2\right] = 8\pi^2 N_0 \sigma_n^2 \int_{-\frac{T_I}{2}}^{\frac{T_I}{2}} \tau_1^2 d\tau_1 = \frac{2\pi^2\sigma_n^2 N_0 T_I^3}{3}. \tag{D.18}$$

Substituting Eqs. (D.6), (D.7), (D.8), (D.9), (D.11), and (D.18) into (4.22), we have

$$\sigma^2_{\hat{F}_W}(t_0) = \frac{\frac{2\pi^2 N_0 T_I^3}{3}(2A^2+\sigma_n^2)}{\left(\frac{\partial^2 W_s}{\partial f^2}\Big|_{f=F}\right)^2}. \tag{D.19}$$

Appendix E

Evaluation of the Bias and Variance in the Application of Spectrogram to Estimate the Instantaneous Frequency of a Signal

The spectrogram is the most early representation of the time-varying spectrum. With a real and even window function $h(t)$, we examine the spectrogram given in Eq.(2.8) on the time-frequency plane. We postulate that the instantaneous frequency of the signal is the frequency which has maximum contribution to the spectrogram as function of time t on the time-frequency plane. Therefore, at a particular instant on the time axis, we locate the peak of the spectrogram and the corresponding frequency is the estimated instantaneous frequency of the signal. Here we will derive the performance of time-varying frequency estimation using the spectrogram and compare it with the PWD in Section 4.3.4.

We consider the received signal of the form given by Eq.(4.8) with the additive complex Gaussian noise the same as in the PWD estimation. Hence Eqs.(D.1) to (D.5)

hold. Furthermore, if we separate the real and imaginary part of $n(t)$, we have

$$n(t) = n_1(t) + jn_2(t), \tag{E.1}$$

such that

$$E[n_1(t)] = E[n_2(t)] = 0 \tag{E.2}$$

$$E[n_1(t_1)n_2(t_2)] = 0 \tag{E.3}$$

$$E[n_1(t_1)n_1^*(t_2)] = E[n_2(t_1)n_2^*(t_2)] = \frac{N_0}{2}\delta(t_1 - t_2). \tag{E.4}$$

Substituting Eq.(4.8) into Eq.(2.7), changing the variable in the integrant and draping the obvious dependence of the spectrogram on time t_0 and f_0, we obtain

$$S_x(t_0, f_0) = \int_{-\infty}^{\infty} h(t)e^{j2\pi f_0 t}\{Ae^{j\theta(t_0-t)} + n(t_0 - t)\}dt = S_{x1} + jS_{x2}, \tag{E.5}$$

where

$$S_{x1} = S_{s1} + S_{n1}, \tag{E.6}$$

$$S_{x2} = S_{s2} + S_{n2}, \tag{E.7}$$

with

$$S_{s1} = A\int_{-\infty}^{\infty} h(t)\cos\{\theta(t_0 - t) + 2\pi f_0 t\}dt \tag{E.8}$$

$$S_{s2} = A\int_{-\infty}^{\infty} h(t)\sin\{\theta(t_0 - t) + 2\pi f_0 t\}dt \tag{E.9}$$

$$S_{n1} = A\int_{-\infty}^{\infty} h(t)\{n_1(t_0 - t)\cos 2\pi f_0 t - n_2(t_0 - t)\sin 2\pi f_0 t\}dt \tag{E.10}$$

$$S_{n2} = A\int_{-\infty}^{\infty} h(t)\{n_1(t_0 - t)\sin 2\pi f_0 t + n_2(t_0 - t)\cos 2\pi f_0 t\}dt. \tag{E.11}$$

We now proceed on to analyze the statistical accuracy of the frequency estimate of a signal using the spectrogram. As mentioned before, the estimate is to locate the maximum point of $|S_x(t_0, f_0)|^2$ at a particular instant of t_0, and the corresponding frequency $\hat{F}_S(t_0)$

is the estimate of the instantaneous frequency of the signal. Let us expand $|S_x(t_0, f_0)|^2$ around the true instantaneous frequency $F(t_0) = \frac{1}{2\pi}\frac{d\theta}{dt}\,|_{t=t_0}$ in a Taylor series such that

$$\begin{aligned} |S_x(t_0, f_0)|^2 &= |S_x(t_0, F(t_0))|^2 + (f_0 - F(t_0))\frac{\partial}{\partial f_0}|S_x(t_0, f_0)|^2_{f_0=F(t_0)} \\ &\quad + \frac{1}{2}(f_0 - F(t_0))^2\frac{\partial^2}{\partial f_0^2}|S_x(t_0, f_0)|^2_{f_0=F(t_0)} + \cdots . \end{aligned} \tag{E.12}$$

Now the maximization of $|S_x(t_0, f_0)|^2$ occurs when

$$\frac{\partial}{\partial f_0}|S_x(t_0, f_0)|^2_{f_0=\hat{F}_s} = 0, \tag{E.13}$$

and assuming that $\hat{F}_S(t_0)$ is close to $F(t_0)$, we can differentiate Eq.(E.12) with respect to f_0 and equate it to zero for maximization. Ignoring the second order and higher order terms of $(f_0 - F(t_0))$, we obtain

$$\hat{F}_S(t_0) = F(t_0) - \left\{ \frac{\frac{\partial}{\partial f_0} \mid S_x(t_0, f_0) \mid^2_{f_0=F(t_0)}}{\frac{\partial^2}{\partial f_0^2} \mid S_x(t_0, f_0) \mid^2_{f_0=F(t_0)}} \right\}. \tag{E.14}$$

But

$$\frac{\partial}{\partial f_0} \mid S_x(t_0, f_0) \mid^2 = 2(S_{x1}S'_{x1} + S_{x2}S'_{x2}), \tag{E.15}$$

where $S'_{xi} = \frac{\partial S_{xi}}{\partial f_0}$, $i = 1, 2$, and

$$\frac{\partial^2}{\partial f_0^2} \mid S_x(t_0, f_0) \mid^2 = 2|S'_x|^2 + 2(S_{x1}S''_{x1} + S_{x2}S''_{x2}). \tag{E.16}$$

Hence, the frequency estimate in Eq.(E.14) can be rewritten as

$$\begin{aligned} \hat{F}_S(t_0) &= F(t_0) - \frac{S_{x1}S'_{x1} + S_{x2}S'_{x2}}{|S'_x|^2 + S_{x1}S''_{x1} + S_{x2}S''_{x2}}\bigg|_{f_0=F(t_0)} \\ &\approx F(t_0) - \frac{S_{x1}S'_{x1} + S_{x2}S'_{x2}}{|S'_s|^2 + S_{s1}S''_{s1} + S_{s2}S''_{s2}}\bigg|_{f_0=F(t_0)}, \end{aligned} \tag{E.17}$$

where in the last step we have assumed the condition of high signal to noise ratio (SNR) such that $S_x \approx S_s$. From Eq.(E.17), we can derive the bias and variance of the frequency estimate at any instant T using the spectrogram, which are defined respectively as

$$\beta_{\hat{F}_S}(T) = E[\hat{F}_S(T) - F(T)] \tag{E.18}$$

$$\sigma^2_{\hat{F}_S}(T) = E\{\hat{F}_S(T) - E[\hat{F}_S(T)]\}^2 \tag{E.19}$$

From the definitions of Eqs.(E.6) to (E.11), we can write

$$\begin{aligned} S'_{x1} &= S'_{s1} + S'_{n1} \\ &= -2\pi A \int_{-\infty}^{\infty} th(t) \sin\{\theta(t_0 - t) + 2\pi f_0 t\} dt \\ &\quad -2\pi \int_{-\infty}^{\infty} th(t)\{n_1(t_0 - t)\sin(2\pi f_0 t) + n_2(t_0 - t)\cos(2\pi f_0 t)\} dt, \quad \text{(E.20)} \\ S'_{x2} &= S'_{s2} + S'_{n2} \\ &= 2\pi A \int_{-\infty}^{\infty} th(t) \cos\{\theta(t_0 - t) + 2\pi f_0 t\} dt \\ &\quad +2\pi \int_{-\infty}^{\infty} th(t)\{n_1(t_0 - t)\cos(2\pi f_0 t) - n_2(t_0 - t)\sin(2\pi f_0 t)\} dt. \quad \text{(E.21)} \end{aligned}$$

Since S_{n1}, S_{n2}, S'_{n1} and S'_{n2} are all linear combinations of n_1, and n_2, they are Gaussian random variables with zero mean. The variances are given by

$$\begin{aligned} \sigma_N^2 &= E[S_{n1}^2] = E[S_{n2}^2] \\ &= \frac{N_0}{2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(t_1)h(t_2)\delta(t_1 - t_2) \\ &\quad \{\cos(2\pi f_0 t_1)\cos(2\pi f_0 t_2) + \sin(2\pi f_0 t_1)\sin(2\pi f_0 t_2)\} dt_1 dt_2 \\ &= \frac{N_0}{2} \int_{-\infty}^{\infty} h^2(t_1) dt_1, \quad \text{(E.22)} \end{aligned}$$

and

$$\sigma_N'^2 = E[S_{n1}'^2] = E[S_{n2}'^2] = 2\pi^2 N_0 \int_{-\infty}^{\infty} t^2 h^2(t_1) dt_1. \quad \text{(E.23)}$$

Furthermore,

$$\begin{aligned} E[S_{n1} S_{n2}] &= \frac{N_0}{2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(t_1)h(t_2)\delta(t_1 - t_2) \\ &\quad \{\cos(2\pi f_0 t_1)\sin(2\pi f_0 t_2) - \sin(2\pi f_0 t_1)\cos(2\pi f_0 t_2)\} dt_1 dt_2 \\ &= 0, \quad \text{(E.24)} \\ E[S_{n1} S'_{n1}] &= \frac{1}{2}\frac{\partial}{\partial f_0} E[S_{n1}^2] = 0, \quad \text{(E.25)} \\ E[S_{n2} S'_{n2}] &= 0, \quad \text{(E.26)} \\ E[S'_{n1} S'_{n2}] &= -2\pi^2 N_0 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} t_1 t_2 h(t_1)h(t_2)\delta(t_1 - t_2) \end{aligned}$$

$$\{\sin(2\pi f_0 t_1)\cos(2\pi f_0 t_2) - \cos(2\pi f_0 t_1)\sin(2\pi f_0 t_2)\}dt_1 dt_2$$

$$= \quad 0. \tag{E.27}$$

Also

$$E[S_{n1}S'_{n2}] = E[S_{n2}S'_{n1}] = \pi N_0 \int_{-\infty}^{\infty} th^2(t)dt = 0, \tag{E.28}$$

since $h^2(t)$ is an even function. Eqs.(E.24) through (E.28) state that S_{n1}, S_{n2}, S'_{n1} and S'_{n2} are all uncorrelated. Hence, we have

$$E[S_{x1}S'_{x1} + S_{x2}S'_{x2}] = S_{s1}S'_{s1} + S_{s2}S'_{s2} \tag{E.29}$$

Substituting Eq.(E.29) and (E.17) into the first term of Eq.(E.18), we obtain the bias of frequency estimation using the spectrogram as

$$\beta_{\hat{F}_S}(T) = -\left.\frac{S_{s1}S'_{s1} + S_{s2}S'_{s2}}{|S'_s|^2 + S_{s1}S''_{s1} + S_{s2}S''_{s2}}\right|_{\substack{t_0=T \\ f_0=F(T)}}. \tag{E.30}$$

Substituting Eq.(E.30) into Eq.(E.19) and using Eq.(E.17), we obtain

$$\sigma^2_{\hat{F}_S}(T) = \left.\frac{E[\{S_{n1}S'_{s1} + S_{s1}S'_{n1} + S_{n2}S'_{s2} + S_{s2}S'_{n2} + S_{n1}S'_{n1} + S_{n2}S'_{n2}\}^2]}{\{|S'_s|^2 + S_{s1}S'_{s1} + S_{s2}S'_{s2}]^2\}}\right|_{\substack{t_0=T \\ f_0=F}} \tag{E.31}$$

Expanding the numerator and using following facts

$$E[S_{n1}S'_{n1}S_{n2}S'_{n2}] = \quad E[S_{n1}]E[S'_{n1}]E[S_{n2}]E[S'_{n2}] \quad = 0, \tag{E.32}$$

$$E[(S_{n1}S'_{n1})^2] = \qquad E[(S_{n1})^2]E[(S'_{n1})^2] \qquad = \sigma^2_N\sigma'^2_N, \tag{E.33}$$

$$E[(S_{n2}S'_{n2})^2] = \qquad E[(S_{n2})^2]E[(S'_{n2})^2] \qquad = \sigma^2_N\sigma'^2_N, \tag{E.34}$$

on it, we obtain

$$\sigma^2_{\hat{F}_S}(T) = \left.\frac{\sigma^2_N|S'_s|^2 + \sigma^2_{N'}|S_s|^2 + 2\sigma^2_N\sigma^2_{N'}}{\{|S'_s|^2 + S_{s1}S''_{s1} + S_{s2}S''_{s2}\}^2}\right|_{\substack{t_0=T \\ f_0=F(T)}}, \tag{E.35}$$

where σ^2_N and $\sigma^2_{N'}$ are defined in Eqs.(E.22) and (E.23) respectively.

We can further measure the performance of the spectrogram in estimating the instantaneous frequency by the use of mean square error such that

$$e^2_{\hat{F}_S}(T) = E\left[\left\{\hat{F}(T) - F(T)\right\}^2\right] = \beta^2_{\hat{F}_S}(T) + \sigma^2_{\hat{F}_S}(T) \tag{E.36}$$

Appendix F

Auto-Regression (AR) Model and Burg Algorithm

A stationary process x_t which contains sinusoidal components can be well described by an Auto-Regression (AR) model:

$$x_t = \sum_{m=1}^{M_1} a_m x_{t-m} + n_t, \tag{F.1}$$

where n_t is a white Gaussian processing with mean zero and variance σ_n^2, M_1 is the order of AR model and x_{t-m} represents the delay of x_t. If the coefficients a_m, $m = 1, 2, ..., M_1$ are known, the power spectrum density of x_t can be obtained as:

$$S_x(f) = \frac{\sigma_n^2}{2B} \left| 1 - \sum_{m=1}^{M_1} a_m e^{-j2\pi m f \Delta t} \right|^{-2}, \tag{F.2}$$

where B is the bandwidth of the stochastic process, and Δt is the sampling period equal to $\frac{1}{2B}$.

Given input x_t, the coefficients a_m, $m = 1, 2, ..., M_1$, at stage M_1, noted as $a_m^{(M_1)}$, may be obtained by either minimizing the forward error power defined by:

$$P_{f,M_1} = E[(e_{f,t}^{(M_1)})(e_{f,t}^{(M_1)*})], \tag{F.3}$$

or minimizing the backward error power defined by

$$P_{b,M_1} = E[(e_{b,t}^{(M_1)})(e_{b,t}^{(M_1)*})] \tag{F.4}$$

where $e_{f,t}^{(M_1)}$ is the error that results at stage M_1 when prediction-error filter is operate in the forward direction. Using the model given by Eq.(F.1), $e_{f,t}^{(M_1)}$ can be expressed as

$$e_{f,t}^{(M_1)} = x_t - \sum_{m=1}^{M_1} a_m^{(M_1)} x_{t-m}, \tag{F.5}$$

and $e_{b,t}^{(M_1)}$ is the error that results at stage M_1, when prediction-error filter is operate in the backward direction. It was proved [15] that the coefficients of forward and backward prediction-error filter are related with each other and with $e_{f,t}^{(M_1)}$ given by Eq.(F.5), $e_{b,t}^{(M_1)}$ can be expressed as

$$e_{b,t}^{(M_1)} = x_{t-M_1} - \sum_{m=0}^{M_1-1} a_{M_1-m}^{(M_1)*} x_{t-m}. \tag{F.6}$$

Burg derived a recursive formula to calculate $a_m^{(M_1)}$ by minimizing the prediction-error power taken as the average of both forward and backward prediction error filters over the entire time interval, as shown by

$$P_{M_1} = \frac{1}{2}(P_{f,M_1} + P_{b,M_1}). \tag{F.7}$$

The procedures of Burg algorithm can be summarized as followings.

step 1: Initialization: $L = 0$ and given input x_t, $t = 1, \cdots, N_1$

$$P_0 = \frac{1}{N_1} \sum_{m=1}^{N_1} |x_t^2|^2 \tag{F.8}$$

$$e_{f,t}^{(0)} = e_{b,t}^{(0)} = x_t \tag{F.9}$$

$$a_0^{(0)} = 1, \qquad a_t^{(0)} = 0,$$

step 2: $L = L + 1$, calculate the reflection coefficients Γ_L, and update P_L and a_m:

$$\Gamma_L = \frac{-2\sum_{t=L+1}^{N_1} e_{f,t}^{(L-1)} e_{f,(t-1)}^{(L-1)*}}{\sum_{t=L+1}^{N_1} [|e_{f,t}^{(L-1)}|^2 + |e_{f,(t-1)}^{(L-1)*}|^2]} \tag{F.10}$$

$$P_L = P_{L-1}(1 - |\Gamma_L|^2) \tag{F.11}$$

$$a_m^{(L)} = a_m^{(L-1)} + \Gamma_L a_{L-m}^{(L-1)*}, \quad m = 0, \cdots, L-1 \tag{F.12}$$

$$a_L^{(L)} = -\Gamma_L \tag{F.13}$$

step 3: update the forward and backward error:

$$e_{f,m}^{(L)} = e_{f,m}^{(L-1)} + \Gamma_L e_{f,(m-1)}^{(L-1)}; \quad m = L+1, \cdots, N_1 \tag{F.14}$$

$$e_{b,m}^{(L)} = e_{b,m-1}^{(L-1)} + \Gamma_L^* e_{f,m}^{(L-1)}; \quad m = L, \cdots, N_1 - 1 \tag{F.15}$$

step 4: Go back to step 2, until $L = M_1$.

step 5:Calculate the estimate power spectrum density of x_t as

$$\hat{S}_x(f) = \frac{P_{M_1}}{2B} \left| 1 - \sum_{m=1}^{M_1} a_m^{(M_1)} e^{-j2\pi m f \Delta t} \right|^{-2}, \tag{F.16}$$

It was also proved that $\hat{S}_x(f)$ given by Eq. (F.16) is equivalent to the maximum-entropy spectral density estimate of the process [41].

As for the selection of suitable M_1 in the AR model, certain criterion is required. The commonly used criterions are AIC (information theory criterion) and FPE (Final Prediction Error) which can be briefly stated as follows [41]:

1. AIC is to select M_1 which minimizes the loglikelihood of the predicting error variance and is equivalent to minimize

$$AIC(M_1) = \ln(P_{M_1}) + \frac{2M_1}{N_1}. \tag{F.17}$$

2. FPE criterion is to select M_1 which minimizes the estimate of the mean square error in prediction, represented by

$$FPE(M_1) = \frac{N_1 + (M_1 + 1)}{N_1 - (M_1 + 1)} P_{M_1}. \tag{F.18}$$

Appendix G

Evaluation of the Probability of Erroneous Estimations

Using Eq.(6.12), we obtain the square magnitude of the WD

$$|W_{xa}(t,f)|^2 = W_a^2(t,f) + |W_{na}(t,f)|^2 + 2Re[W_a(t,f)W_{na}(t,f)]. \qquad \text{(G.1)}$$

The main lobe of $|W_{xa}(t,f)|^2$ occurs at $t = 0, \quad f = 0$. Now, let $t = t_1$ and $f = f_1$ be the location of the highest sidelobe of $|W_{xa}(t,f)|^2$. We are interested in the probability of error such that

$$P_e \approx P(d \geq 0), \qquad \text{(G.2)}$$

where

$$d = |W_{xa}(t_1,f_1)|^2 - |W_{xa}(0,0)|^2. \qquad \text{(G.3)}$$

Now if we normalize d such that $\bar{d} = \frac{d}{W_a^2(0,0)}$ and using Eq. (G.1), we obtain

$$\begin{aligned} \bar{d} &= [\gamma^2 - 1] + [|\frac{W_{na}^2(t_1,f_1)}{W_a(0,0)}|^2 - |\frac{W_{na}^2(0,0)}{W_a(0,0)}|^2] \\ &\quad +2Re[\gamma\frac{W_{na}(t_1,f_1)}{W_a(0,0)} - \frac{W_{na}(0,0)}{W_a(0,0)}], \end{aligned} \qquad \text{(G.4)}$$

where $\gamma = \frac{W_a(t_1,f_1)}{W_a(0,0)}$. Using the same derivation as Theorem 6.1, we can easily see that $\frac{W_{na}(t,f)}{W_a(0,0)}$ is Gaussian distributed with zero-mean and variance $\frac{2N_0}{W_a(0,0)}$. Furthermore,

$$E[|\frac{W_{na}(t,f)}{W_a(0,0)}|^2] = \frac{2N_0}{W_a(0,0)} \qquad \text{(G.5)}$$

$$
\begin{aligned}
var[|\frac{W_{na}(t,f)}{W_a(0,0)}|^2] &= E\left|\frac{W_{na}(t,f)}{W_a(0,0)}\right|^2 - \left[E\left|\frac{W_{na}(t,f)}{W_a(0,0)}\right|^2\right]^2 \\
&= 2\left[E\left|\frac{W_{na}(t,f)}{W_a(0,0)}\right|^2\right]^2 \\
&= \frac{8N_0^2}{W_a^2(0,0)}. \qquad \text{(G.6)}
\end{aligned}
$$

Under high SNR, $\frac{N_0^2}{W_a(0,0)^2} \ll \frac{N_0}{W_a(0,0)}$, and we can assume the $var[|\frac{W_{na}(t,f)}{W_a(0,0)}|^2] \ll E[|\frac{W_{na}(t,f)}{W_a(0,0)}|^2]$. Therefore, we may neglect the term in the second bracket of Eq. (G.4), which has a zero mean and a negligible variance. Under the above assumption, $\bar{d}$ becomes Gaussian variable such that

$$
\begin{aligned}
E[\bar{d}] &= E\left\{(\gamma^2-1) + \frac{2}{W_a(0,0)} Re[\gamma W_{na}(t_1,f_1) - W_{na}(0,0)]\right\} \\
&= (\gamma^2-1). \qquad \text{(G.7)}
\end{aligned}
$$

Using Eq.(6.60) and the same derivation as Theorem 6.3, we obtain:

$$
var(\bar{d}) = \frac{4N_0}{W_a(0,0)}\left[\gamma^2 - 2\frac{W_a(t_1,f_1)}{W_a(0,0)}\gamma + 1\right]. \qquad \text{(G.8)}
$$

But since the signal energy is unity, then $W_a(0,0) = 2$ and $W_a(t_1,f_1) = 2\gamma$. Thus,

$$
var(\bar{d}) = 2N_0(1-\gamma^2), \qquad \text{(G.9)}
$$

and

$$
\begin{aligned}
P_e &= \frac{1}{\sqrt{4\pi(1-\gamma^2)N_0}} \int_0^\infty e^{-\frac{(\bar{d}-\gamma^2+1)^2}{4N_0(1-\gamma^2)}} d\bar{d} \\
&= \frac{1}{2}\text{erfc}\left(\sqrt{\frac{1-\gamma^2}{2N_0}}\right). \qquad \text{(G.10)}
\end{aligned}
$$

Appendix H

Evaluation of $\frac{\partial F}{\partial \varepsilon}$ in Eq.(7.66)

Using the definition of ΔT^2 in Eq. (7.62) together with Eq. (7.65) and differentiating it with respect to ε, we have

$$\begin{aligned}\left.\frac{\partial[\log \Delta T^2]}{\partial \varepsilon}\right|_{\varepsilon=0} &= \frac{1}{\Delta T^2}\int_{-T/2}^{T/2} t^2 \frac{\partial}{\partial \varepsilon}\left[\mid a_0(t)\mid^2 + 2Re[\varepsilon a_0(t)\mu(t)] + \varepsilon^2 \mid \mu(t)\mid^2\right] dt\Big|_{\varepsilon=0} \\ &= \frac{2}{\Delta T^2}\int_{-T/2}^{T/2} t^2 a_0(t)\mu(t)dt. \end{aligned} \tag{H.1}$$

Let $A_0(f)$ and $\mathcal{M}(f)$ be the Fourier transform of $a_0(t)$ and $\mu(t)$ respectively. Then, using Eqs. (7.61) and (7.65), we have

$$\begin{aligned}\left.\frac{\partial[\log \Delta F^2]}{\partial \varepsilon}\right|_{\varepsilon=0} &= \frac{2}{\Delta F^2} Re\left[\int_{-F}^{F} f^2 A_0(f)\mathcal{M}^*(f)df\right] \\ &= \frac{2}{\Delta F^2} Re\left[\int_{-F}^{F} f^2 A_0(f)\left\{\int_{-\infty}^{\infty}\mu(t)e^{j2\pi ft}dt\right\}df\right] \\ &= \frac{2}{\Delta F^2} Re\left[\int_{-\infty}^{\infty}\mu(t)\left\{\int_{-F}^{F} f^2 A_0(f)e^{j2\pi ft}df\right\}dt\right] \\ &= \frac{2}{\Delta F^2}\int_{-\infty}^{\infty}\mu(t)\int_{-F}^{F} f^2 A_0(f)e^{j2\pi ft}df\,dt. \end{aligned} \tag{H.2}$$

Furthermore

$$\frac{\partial}{\partial \varepsilon}\left\{\int_{-\infty}^{\infty} a^2(t)dt\right\}\Big|_{\varepsilon=0} = 2\int_{-\infty}^{\infty} a_0(t)\mu(t)dt. \tag{H.3}$$

Substituting Eqs. (H.1) through (H.3) into Eq. (7.66), we obtain Eq. (7.67).

Appendix I

Proof of Theorem 7.5

Given that $a(t) = a_1\varphi_1(t) + a_2\varphi_2(t)$ and its Fourier transform $A(f) = a_1\Phi_1(f) + a_2\Phi_2(f)$, we have

$$\Delta T^2 = a_1^2 \int_{-T/2}^{T/2} t^2\varphi_1^2(t)dt + 2a_1c_2 \int_{-T/2}^{T/2} t^2\varphi_1(t)\varphi_2(t)dt + a_2^2 \int_{-T/2}^{T/2} t^2\varphi_2^2(t)dt, \tag{I.1}$$

$$\Delta F^2 = \left\{ a_1^2 \int_{-F}^{F} f^2 \mid \Phi_1(f) \mid^2 df + 2a_1a_2 Re\left[\int_{-F}^{F} f^2\Phi_1(f)\Phi_2^*(f)df\right] + a_2^2 \int_{-F}^{F} f^2 \mid \Phi_2(f) \mid^2 df. \right\} \tag{I.2}$$

Define

$$\boldsymbol{C}_0 = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \quad \boldsymbol{B}_0 = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \quad \boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \tag{I.3}$$

where

$$c_{ij} = \int_{-T/2}^{T/2} t^2\varphi_i(t)\varphi_j(t)dt, \tag{I.4}$$

and

$$b_{ij} = Re\left[\int_{-F}^{F} f^2\Phi_i(f)\Phi_j^*(f)df\right] \qquad i,j = 1,2. \tag{I.5}$$

Then

$$\Delta T^2 = \boldsymbol{a}^T\boldsymbol{C}_0\boldsymbol{a} \qquad \Delta F^2 = \boldsymbol{a}^T\boldsymbol{B}_0\boldsymbol{a}. \tag{I.6}$$

Using similarity transformation, we can obtain a diagonal matrix C_1 such that

$$C_1 = V^T C_0 V = diag(\gamma_{11}, \gamma_{12}), \tag{I.7}$$

where V is the transform matrix. Since C_0 is non-negative definite, we have $\gamma_{11} \geq \gamma_{22} \geq 0$. Defining

$$B_1 = V^T B_0 V = \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \text{ ; and } \overline{a} = V^T a = \begin{bmatrix} \overline{a}_1 \\ \overline{a}_2 \end{bmatrix}, \tag{I.8}$$

where $\beta_{12} = \beta_{21}$, then we have

$$\Delta T^2 = \overline{a}^T C_1 \overline{a} = \gamma_{11}\overline{a}_1^2 + \gamma_{22}\overline{a}_2^2, \tag{I.9}$$

$$\Delta F^2 = \overline{a}^T B_1 \overline{a} = \beta_{11}\overline{a}_1^2 + \beta_{22}\overline{a}_2^2 + 2\beta_{12}\overline{a}_1\overline{a}_2, \tag{I.10}$$

and $\overline{a}^T\overline{a} = \overline{a}_1^2 + \overline{a}_2^2 = 1$, which can be rewritten as

$$\overline{a}_2 = \pm\sqrt{1 - \overline{a}_1^2}. \tag{I.11}$$

Substituting Eq.(I.11) into Eqs.(I.9) and (I.10) we obtain

$$\Delta T^2 = (\gamma_{11} - \gamma_{22})\overline{a}_1^2 + \gamma_{22}, \tag{I.12}$$

$$\Delta F^2 = (\beta_{11} - \beta_{22})\overline{a}_1^2 + \beta_{22} \pm 2\beta_{12}\overline{a}_1\sqrt{1 - \overline{a}_1^2}. \tag{I.13}$$

Eliminating $\overline{a}_1$ between Eqs.(I.12) and (I.13), and rearranging, we obtain a general conic section equation in ΔT^2 and ΔF^2 such that

$$\begin{aligned} &(\gamma_{11} - \gamma_{22})^2(\Delta F^2)^2 - 2(\gamma_{11} - \gamma_{22})(\beta_{11} - \beta_{22})\Delta T^2\Delta F^2 + \{(\beta_{11} - \beta_{22})^2 + 4\beta_{12}^2\}(\Delta T^2)^2 \\ &+2\{\gamma_{22}(\beta_{11} - \beta_{22}) - \beta_{22}(\gamma_{11} - \gamma_{22})\}(\gamma_{11} - \gamma_{22})\Delta F^2 - 2\{(\beta_{11} - \beta_{22})[\gamma_{22}(\beta_{11} - \beta_{22}) \\ &-\beta_{22}(\gamma_{11} - \gamma_{22})] + 4\beta_{12}^2(\gamma_{11} + \gamma_{22})\}\Delta T^2 + \{[\gamma_{22}(\beta_{11} - \beta_{22}) - \beta_{22}(\gamma_{11} - \gamma_{22})]^2 + \\ &4\beta_{12}^2\gamma_{11}\gamma_{22}\} \\ = \ &0. \end{aligned} \tag{I.14}$$

This is the equation of an ellipse since the discriminant

$$\begin{aligned} D &= (\gamma_{11}-\gamma_{22})^2[(\beta_{11}-\beta_{22})^2+4\beta_{12}^2]-(\gamma_{11}-\gamma_{22})^2(\beta_{11}-\beta_{22})^2 \\ &= 4\beta_{12}^2(\gamma_{11}-\gamma_{22})^2 \geq 0. \end{aligned} \tag{I.15}$$

Appendix J

Proof of Theorem 7.7

First step, we argue that the optimum problem

$$\max_{(\Delta T^2, \Delta F^2) \in E} \Delta T^2 \cdot \Delta F^2 \tag{J.1}$$

has a unique optimal solution, where E is an arbitrary convex set in the positive quadrant of the $(\Delta T^2, \Delta F^2)$ plane. Suppose the contrary, that is, $(\Delta T_1^2, \Delta F_1^2) \in E$, $(\Delta T_2^2, \Delta F_2^2) \in E$ are two distinct optimal solutions of Eq.(J.1), then $\Delta T_1^2 \cdot \Delta F_1^2 = \Delta T_2^2, \Delta F_2^2$.

Since E is a strictly convex set, the line segment joining any two points of E lies strictly inside of E. Therefore, we have $\left(\frac{\Delta T_1^2 + \Delta T_2^2}{2}, \frac{\Delta F_1^2 + \Delta F_2^2}{2}\right)$ lies strictly inside E. Since $(\Delta T_1^2, \Delta F_1^2) \in E$ and $(\Delta T_2^2, \Delta F_2^2) \in E$ are two distinct points, satisfying $\Delta T_1^2 \cdot \Delta F_1^2 = \Delta T_2^2 \cdot \Delta F_2^2$, we can assume

$$\Delta T_1^2 > \Delta T_2^2 \text{ and } \Delta F_1^2 < \Delta F_2^2.$$

Now, we have

$$(\Delta T_1^2 - \Delta T_2^2)(\Delta F_2^2 - \Delta F_1^2) > 0,$$

which implies

$$\Delta T_1^2 \Delta F_2^2 + \Delta T_2^2 \Delta F_1^2 > \Delta T_1^2 \Delta F_1^2 + \Delta T_2^2 \Delta F_2^2.$$

Adding both sides of the above equation with $\Delta T_1^2 \Delta F_1^2 + \Delta T_2^2 \Delta F_2^2$ and using the fact

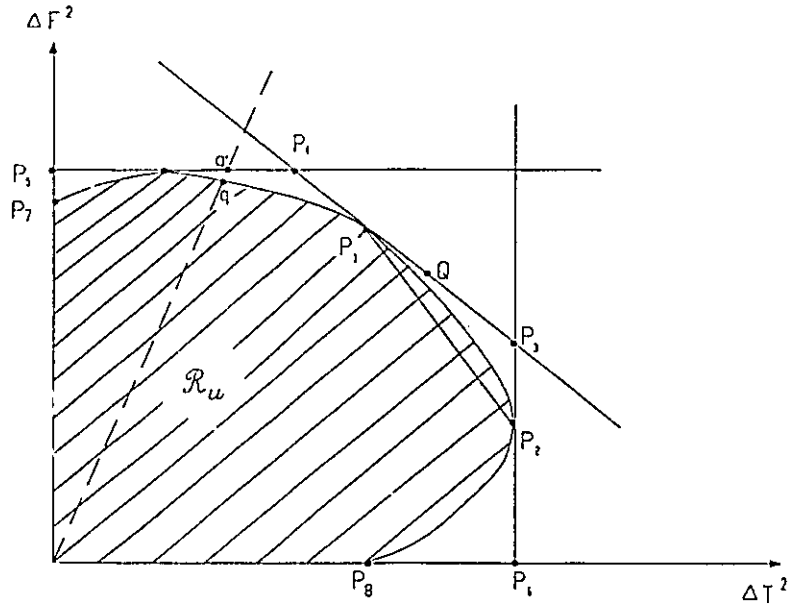


Figure J.1: Diagram of Expressing the Convergence of Polygon Approach.

that $\Delta T_1^2 \cdot \Delta F_1^2 = \Delta T_2^2 \cdot \Delta F_2^2$, we obtain

$$\left(\frac{\Delta T_1^2 + \Delta T_2^2}{2}\right) \cdot \left(\frac{\Delta F_1^2 + \Delta F_2^2}{2}\right) > \Delta T_1^2 \cdot \Delta F_1^2 \tag{J.2}$$

Eq.(J.2) implies $(\Delta T_1^2, \Delta F_1^2)$ is non-optimal for Eq.(J.1), a contradiction.

Secondly, suppose line P_3P_4 is tangent to $\Re_u$ at P_1, line P_3P_6 is tangent to $\Re_u$ at P_2. Let E denote the polygon sided by ΔT^2 axis, ΔF^2 axis and lines P_4P_3, P_3P_6 and P_4P_5 (see Fig.J.1). Then, $\Re_u \subseteq E$.

Suppose $\max_{(\Delta T^2, \Delta F^2) \in E} \Delta T^2 \Delta F^2$ attains its maximum at a point Q. We argue $\max_{(\Delta T^2, \Delta F^2) \in \Re_u} \Delta T^2 \Delta F^2$ attains its maximum on the arc $\overset{\frown}{P_1P_2}$ along the boundary of $\Re_u$. To see this, let $f(\Delta T^2, \Delta F^2) = \Delta T^2 \cdot \Delta F^2$. Then it can been seen that $f(Q) > f(P_1)$, $f(Q) > f(P_2)$. Moreover, $f(P)$ is monotonically decreasing along the line QP_4 and P_4P_5. Therefore,

$$f(P_1) \geq f(P) \quad \text{for all } P \text{ on the line segment } P_1P_4 \text{ and } P_4P_5. \tag{J.3}$$

Similarly, we have

$$f(P_2) \geq f(P) \quad \text{for all } P \text{ on the line segment } P_2P_6. \tag{J.4}$$

On the other hand, for any point q on the arc $\overset{\frown}{P_7P_1}$ along the boundary of set $\Re_u$, we can draw a line from the origin to q, and extend the line to intersect with the line P_4P_1 or P_4P_5, at some point q' (see Fig. J.1). Obviously, $f(q) \leq f(q')$, which, by Eq.(J.3), implies that

$$f(P_1) \geq f(q') \geq f(q) \quad \text{for all } q \text{ on the arc } \overset{\frown}{P_7P_1}. \tag{J.5}$$

Similarly, we can show

$$f(P_2) \geq f(q) \quad \text{for all } q \text{ on the arc } \overset{\frown}{P_2P_8}. \tag{J.6}$$

Thus, the optimal solution of

$$\max_{(\Delta T^2, \Delta F^2) \in \Re_u} \Delta T^2 \Delta F^2 \tag{J.7}$$

must lie on the arc $\overset{\frown}{P_1P_2}$. This means that we can concentrate on searching for the optimal solution along the arc $\overset{\frown}{P_1P_2}$.

Finally, suppose the convex polygon approximation algorithm has resulted in an N-sided polygon, and the two-dimensional optimization problem over this polygon attains the optimal solution at a point Q on the boundary of the polygon. Suppose the closest two tangent points to Q are P_1, P_2. Using the result from above we see that the optimal solution of Eq.(J.7) lies in the arc $\overset{\frown}{P_1P_2}$. Thus, as the algorithm tries to refine the polygon approximation of $\Re_u$ along the arc $\overset{\frown}{P_1P_2}$, we shall get better and better estimate of the location of the optimal solution to Eq.(C.7).

In particular, we chose the slope of new line to be $\frac{1}{2}$(slope of the line tangent at P_1 + slope of the line tangent at P_2). Suppose the new line is tangent to $\Re_u$ at P_{new}. Then, the optimal solution lies either in the arc $\overset{\frown}{P_1P_{new}}$ or $\overset{\frown}{P_{new}P_2}$. By our construction of P_{new},

the length of $P_1\widehat{P_{new}}$ and $P_{new}\widehat{P_2}$ is roughly $\frac{1}{2}$ of the length $\widehat{P_1P_2}$. Thus, we see that the search region is cut by half after each iteration. This implies that it takes $0(\log \frac{1}{\epsilon})$ iterations to generate an ϵ-optimum solution. In other words, the polygon algorithm has a linear rate of convergence.

Bibliography

[1] Abramowitz, M. and Stegun, I.A., "Handbook of Mathematical Functions", National Bureau of Standards, 1964.

[2] Ackroyd, M.H. "Instantaneous and Time-Varying Spectra- An Introduction", The Radio and Electronic Engineer, Vol. 39, No.3, March 1970. pp.145-152.

[3] Ackroyd, M.H. "Short-Time Spectra and Time- Frequency Energy Distributions", J. Acoust. Soc. Amer. 50(5) 1970, pp.1229-1231.

[4] Appleton, E.V. and Barnett, M.A.F., "On Some Direct Evidence for Downward Atmospheric Reflection of Electric Rays", Proce. Roy. Soc. Ser. A. 109, Dec.1, 1925, pp.621-641.

[5] Basseville, M. "Detecting Changes in Signals and Systems-A Survey", Automatic, Vol.24, No.3, 1988. pp.309-326.

[6] Boashash, B., Lovell, B and Kootsookos, P.J., "Time-Frequency Signal Analysis and Instantaneous Frequency Estimation: Their Inter-Relationship and Applications," IEEE Int. Conf. Ckts and Syst., Portland, Oregon, April 1989, pp. 1237-1242.

[7] Boashash, B., Jones, G., and O'shea, P.," Instantaneous Frequency Estimation: Implementation and Applications," Proc. of SPIE Conf. Ad. Alg. and Arch. for Sig. Proc., Vol. 1152, San Diego, CA, August 1989.

[8] Boashash, B., "Time-Frequency Signal Analysis," chapter in Advances in Spectrum Estimation, ed. S. Haykin, Prentice-Hall, 1990.

[9] Boashash, B. and White, L.B.," Instantaneous Frequency Estimation and Automatic Time-Varying Filtering," Int. Conf. Acoust., Speech, Sig. Proc., 1990, pp. 1221-1224.

[10] Boashash, B. and O'Shea, P.,"A Methodology for Detection and Classification of Underwater Acoustic Signals Using Time-Frequency Analysis Techniques," IEEE Trans. Acoust. Speech, Sig. Proc., Vol. 38, no. 11, Nov. 1990, pp. 1829-1841.

[11] Boashash, B., ed., "Time-Frequency Signal Analysis-Methods and Applications," Longman-Chesire, Melbourne, Australia, 1991.

[12] Bonzanigo, F., "An Improvement to Tribolet's Phase Unwrapping Algorithm", IEEE Trans. on ASSP, Vol. ASSP-26, Feb. 1978. pp. 104-105.

[13] Boudreaux-Bartels, G.F., "Time-Frequency Signal Processing Algorithm: Analysis and Synthesis Using Wigner Distribution", Ph.D. Dissertation, Rice University, Houston, Texas, U.S.A., Dec., 1983.

[14] Boudreaux-Bartels, G.F. and Parks, T.W., "Time-Varying Filtering and Signal Estimation Using Wigner Distribution Synthesis", IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. ASSP-34, No.3, pp.442-451, June, 1986.

[15] Burg, J.P. "Maximum Entropy Spectral Analysis," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1975.

[16] Cadzow, J. A. "Spectral Estimation: An Overdetermined Rational Model Equation Approach", Proceedings of the IEEE. Vol.70, No.9, Sep. 1982.

[17] Carter, C.R., Casas, E. and Chung, T. "Emergency Locator Transmitter (ELT) Signal Models for SARSAT", Can. Elec. Eng. J., Vol.11 No.2, 1986.

[18] Chung,T. and Carter, C.R. "Basic Concepts in the Process of SARSAT Signals", IEEE Trans. on Aerospace and Electronic System, vol. AES-23, No.2, March 1987.

[19] Chung, T. "Advanced Signal Processing Strategies for Search and Rescue Satellite Aided Tracking", Ph.D. Thesis, McMaster University, Mar., 1988

[20] Classen, T.A.C.M. and Mecklenbrauker, W.F.G., "The Wigner Distribution–A Tool for Time Frequency Signal Analysis, Part I: Continuous-Time Signals", Philips Journal of Research Vol.35, No.3, 1983.

[21] Classen, T.A.C.M. and Mecklenbrauker, W.F.G., "The Wigner Distribution–A Tool for Time Frequency Signal Analysis, Part II: Discrete-Time Signals", Philips Journal of Research Vol.35, No.4/5, 1983.

[22] Cohen, F.S., Boudreaux-Bartels, G.F. and Kadambe, S. "Tracking of Unknown Non-Stationary Chirp Signals Using Unsupervised Clustering in the Wigner Distribution Space", ICASSP April, 1988.

[23] Cohen, L., "Generalized phase-space distribution functions", J. of Math. Phys. 7, pp.781-786, 1966.

[24] Daubechies, I., "The Wavelet Transform, Time-Frequency Localization and Signal Analysis", IEEE Trans. on Information Theory, Vol. IT-36, No.5, Sept., 1990, pp. 961-1005.

[25] De Bruijn, N.G., "Uncertainty Principle in Fourier Analysis", in Inequalities, O.Shisha, Ed. New York, NY: Academic Press, 1967, pp.57-71.

[26] De Recheres, C., "A New Method for the Numerical Analysis of Non-Stationary Signals", Physics of the Earth and Planetary Interiors, Vol.12, 1974.

[27] Fettweis, A., "On the Significant of Group Delay in communication Engineering", AEU 33, 1977, pp.342-348.

[28] Flammer, C., "Spheroidal Wave Functions", Stanford, Calif.: Stanford University Press, 1957.

[29] Flanagan, J.L., "Speech Analysis and Perception", Springer Verlag, Ed., Berlin, 1955.

[30] Flandrin, P. "Time-Dependent Spectra of Non- Stationary Processes", Lecture given at the CISM Advanced School, Udine, Italy, Sept.21-25, 1987.

[31] Flandrin, P. and Martin, W., "Pseudo-Wigner Estimators", IEEE ASSP Spectrum Estimation Workshop II, Tampa, Florida Nov. 1983.

[32] Flandrin, P. "Maximum Signal Energy Concentration in a Time- Frequency Domain", Proc. ICASSP'88, Vol. IV, pp.2176-2179, April, 1988.

[33] Flandrin, P. "When Is the Wigner-Ville Spectrum Non-Negative?", Signal Processing III: Theories and Applications, I.T.Young et al. (editors), Elsevier Science Publishers B.V. (North-Holland), EURASIP, 1986, pp.239-242.

[34] Frank, L.E., "Signal Theory", Prentice-Hall, Inc. 1969. Revised edition published by Dowden Culver. Inc. 1981.

[35] Fudan University, "Probability, – Part III, Stochastic", People's Education Publication, P.R.China, Feb. 1982.

[36] Gabor, D., "Theory of Communication", J. IEE (London), Vol.93, 1946, pp.429-457.

[37] Golub, G.H. and Vanloan, C.F., "Matrix Computations", Johns Hopkins Unversity Press, 1983.

[38] Greenlay, T. and Haykin, S. "Application of the Forward and Backward Linear Predication Method to Angle of Arrival Estimation in a Multipath Environment", CRL Report No.126, McMaster Univ., Hamilton, Canada, April 1984.

[39] Gupta, M. S. "Definition of Instantaneous Frequency and Frequency Measurability", American Journal of Physics Vol.43, No. 12, Dec. 1975, pp.1087-1088.

[40] Hajek, B., "A Tutorial Survey of Theory and Applications of Simulated Annealing", Proceedings of 24th Conference on Decision and Control, FT. Lauderdale, FL. Dec., 1985, pp.755-759.

[41] Haykin, S., "Nonlinear Methods of Spectral Analysis", Springer- Verlay, 1979.

[42] Herring, R. W., "The cause of Line Splitting in Burg Maximum- Entropy Spectral", IEEE Trans. on ASSP, Vol. ASSP-28, No.6, Dec., 1980.

[43] Hlawatsch, F., "Interference Terms in the Wigner Distribution", in Digital Signal Processing - 84, V. Cappellini and A.G. Constantinides, Eds. Amsterdam, the Netherlands: North-Holland, 1984. pp.363-367.

[44] Jain, A.K. and Rangnanth, S. "Extrapolation Algorithms for Discrete Signals with Application in Spectral Estimation", IEEE Trans. on ASSP, Vol.ASSP-29, No.4, Aug., 1981.

[45] Janssen, A.J.E.M. and Claasen, T.A.C.M., "On the Positivity of Time-Frequency Distribution", IEEE Trans. on ASSP, Vol. ASSP-33, No.4, Aug., 1985. pp.1029-1032.

[46] Jin, Q, Luo, Z.Q. and Wong, K.M., "Optimum Signal Design in Time-Frequency Plane", International Conference on Signal Processing '90/Beijing, China, Oct., 1990.

[47] Jin, Q, Luo, Z.Q. and Wong, K.M.,"Uncertainty Principle and Optimum Signal Design for Simultaneous Time and Frequency Estimate", CRL Rep. 213, Commun. Res. Lab., McMaster Univ. Hamilton, Ont. Canada, Apr., 1990.

[48] Jin, Q. "Parameter Estimation of Time- and Band-limited Signal under Band-limited White Noise", Presented at 1986 IEEE workshop on ASSP, Beijing.

[49] Jin, Q., Luo, Z.Q., and Wong, K.M., "An Optimum Complete Orthonormal Basis for Signal Analysis and Design", CRL Rep. 240, Feb., 1992.

[50] Jin, Q., Luo, Z.Q., and Wong, K.M., "The Optimum Complete Orthonormal Basis with Maximum Energy Concentration", submitted to IEEE Trans. on Information Theory.

[51] Jordan, P., "Zur Quantenmechanik", Z. Phys., Dtsch. Vol 34, 1925, pp.858-888.

[52] Kay, S. and Marple, S. L. ,Jr., "Spectrum Analysis-A Modern Perspective," Proc. IEEE, Nov. 1981.

[53] Kirkpatric, S., Gelatt, C.D., and Vecchi, Jr.,M.P., "Optimization by Simulated Annealing", Science, Vol. 220, No. 4598, May 13, 1983, pp.671-680.

[54] Kenneth, W. C., and O'Reilly, J. J., "Mathematical Topics in Telecommunications Vol.1: Optimization Methods in Electronics and Communications", John Wiley and Sons Inc., 1984.

[55] Klauder, J.R., "The Design of Radar Signals Having Both High Range Resolution and High Velocity Resolution", The Bell System Technical Journal, July 1960, pp.809-819.

[56] Kodera, K., de Villedary, C. and Gendrin, R., "A new Method for the Numerical Analysis of Nonstationary Signals", Physics of the Earth and Planetary Interiors, Vol.12, 1976. pp.142-150.

[57] Kodera, K., Gendrin, R. and de Villedary, C., "Analysis of Time-Varying Signals with Small BT Values", IEEE Trans. on ASSP, Vol. ASSP-26, No.1, Feb. 1978. pp.64-76.

[58] Kodera, K., "Analyse Numerique de Signaux de Signaux Geophysiques Non-Stationnaires", Ph.D. Thesis, University de Paris VI, Jan. 1976.

[59] Koenig, R., Dunn, H.K. and Lacy, L.Y., "The sound Spectrograph", J.Acoust. Soc. Am., Vol. 18, 1976, pp.64-76

[60] Kumaresan, R. and Tufts, Donald W., "Estimating the Angles of Arrival of Multiple Plane Waves", IEEE Trans. on AES, Vol. AES-19, No.1, Jan. 1983.

[61] Lacoss, R.T. "Data Adaptive Spectral Analysis Methods", Geophysics, Vol.36, No.4, Aug. 1971, pp.661-675.

[62] Landau, H. J. and Pollak, H. O. "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty —- II", B.S.T.J., Jan.,1961.

[63] Landau, H. J. and Pollak, H. O. "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty —- III: The Dimension of the Space of Essentially Time- and Band-Limited Signals", B.S.T.J., July, 1962.

[64] Li, S., "On a Class of Nonstationary Signals", Proc. IEEE ICASSP, New York, 1988. pp.2192-2195.

[65] McMahon, D.R.A. and Barrett, R.F., "An Efficient Method for the Estimation of the Frequency of a Single Tone in Noise from the Phases of Discrete Fourier Transforms", Signal Processing, Vol. 11, No.2, Sept. 1986. pp.169-177.

[66] Martin, W. and Flandrin, P., "Detection of Changes of Signal Structure by Using the Wigner-Ville Spectrum", Signal Processing 8, North-Holland 1985, pp.215-233.

[67] Martin, W. and Flandrin, P., "Analysis of Non-Stationary Processes: Short-Time Periodograms vs a Pseudo-Wigner Estimator", Signal Processing II: Theories and Applications (H.W.Schüssler, ed.), North-Holland, Amsterdam, 1983, pp.455-458.

[68] Mohamed, A., "Time-Frequency Spectral Analysis", Draft Copy of Part of a Drea Report, 1984.

[69] Monzingo, R.A. and Miller, T.W., "Introduction to Adaptive Arrays", Wiley-Interscience, 1980.

[70] Oppenheim, A.V. and Schafer, R.W., "Digital Signal Processing", Englewood Cliffs, NJ: Prentice-Hall, 1960.

[71] Osyczka, A. and Davies, B. J., "Multicriterion Optimization in Engineering", John Wiley and Sons Inc., 1984.

[72] Page, C.H., "Instantaneous Power Spectra", J. Appl. Phys., Vol. 23, 1952, pp.103-106.

[73] Papoulis, A., "The Fourier Integral and Its Applications", McGraw-hill Book Company, Inc. 1962.

[74] Papoulis, A., "Ambiguity Function in Fourier Optics", J. Opt. Soc. Am. 64, 1974, pp.779-788.

[75] Papoulis, A. "A New Algorithm in Spectral Analysis and Band- Limited Extrapolation", IEEE Trans. on CAS, Vol.CAS-22, No.9, Sept.1975.

[76] Parks, T.W. and Shenoy, R.G., "Time-Frequency Concentrated Basis Functions", Proc. ICASSP'90, Vol.5, pp.2459-2462, April, 1990.

[77] Potter, R.K., Kopp, G. and Green, C., "Visible Speech", New York: NY: Academic Press, 1981.

[78] Price, R. and Green_ Jr., P., "Signal Processing in Radar Astronomy - Communication via Fluctuating Multipath Media", Tech. Rept. No. 234, MIT Lincoln Lab. Lexington, Mass. Oct., 1960.

[79] Price, R. and Turin, G.L., "Communications and Radar - Section A - General", IEEE Trans. Info. theory IT-9(4), Oct., 1963, pp.240-246.

[80] Priestley, M.B. "Spectral Analysis and Time Series", Academic Press, New York, 1981.

[81] Riesz, F. and Nagy, B.S.Z., "Leçons D'analyse Fonctionnelle" (Chinese translation), Akadémiai Kiadó, Budapest 1965, 4 ^{éme}édition.

[82] Rihaczek, A.W., "Signal Energy Distribution in Time and Frequency", IEEE Trans. Information Theory, Vol. IT-14, Jan.1964, pp.95-97.

[83] Robinson, S.R., "Problem of Phase from Intensity Measurements," J. Opt. Soc. 68. 1978, pp.87-92.

[84] Schmidt, R.O. "Multiple Emitter Location and Signal Parameter Estimation", in Proc. RADC Spectral Est. Workshop, Griffiss AFBS. NY, 1979.

[85] Shuchmann, R.A., Davis, C.F. and Jackson, P.L., "Contour Strip Mine Detection and Identification with Imaging Radar", IEEE Trans. AER EL, AES-11, 1975.

[86] Silverman, R.A. "Local Stationary Random Processes", IRE Trans. on IT, IT-3, 1957, pp.182-187.

[87] Skolnik, M.I., "Introduction to Radar Systems", Second Ed., McGraw-Hill Book Co., 1980.

[88] Slepian,D and Pollak,H.O. "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty —I", The Bell System Technical Journal, January 1961.

[89] Slepian, D., "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty —- IV: Extensions to many Dimensions; Generalized Prolate Spheroidal Functions", B.S.T.J., Nov.,1964.

[90] Slepian, D., "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty —- V: The Discrete Case", B.S.T.J., May-June, 1978.

[91] Stratton, J.A., Morse, P.M., Chu, L.J., LIttle, J.D.C., and Corbator, F.J., "Spheroidal Wave Functions", Cambridge, Mass.: Technology Press, 1956.

[92] Tribolet, J.M., "A New Phase Unwrapping Algorithm", IEEE Trans. on ASSP. Vol. ASSP-25, Apr. 1977. pp.170-177.

[93] Urkowitz, H., "Signal Theory and Random Processes", New York: Artech House, 1983.

[94] Vakman, D.Y. "On the Definition of Concepts of Amplitude, Phase and Instantaneous Frequency of a Signal", Radio Eng. Electron Phys., 17, 1972. pp.754-759.

[95] Van Trees, H.L. "Detection, Estimation, and Modulation Theory", Part I. New York: Wiley, 1968.

[96] Ville, J. "Théorie et applications de la notion de signal analytique", Cables et Transmission, Vol. 2A, 1948, pp.61-74.

[97] Wigner, E. P. "Quantum-Mechanical Distribution Function Revisited", Perspectives in Quantum Theory (w.Yourgrau, A. Van der Merwe, eds.) MIT Press, Cambridge, 1971, pp.25-36.

[98] Williams, W.J., Krishnamachari, S., Jeong, J., and Driscal, M.E., "Application of Reduced Interference Distribution", The 1990 Digital Signal Processing Workshop, New Paltz, New York, Sept. 16-19, 1990.

[99] White, L.B. and Boashash, B., "Estimation of the Instantaneous Frequency of a Gaussian Random Process by Use of the Wigner-Ville Spectrum", Proceeding 1987 Int. Symp. on Signal Processing and its Applications, ed., B.Boashash, pp.275-279, Brisbane, Australia, August 1987.

[100] White, L.B. and Boashash, B., "Estimating the Instantaneous Frequency of a Gaussian Random Process", IEEE Trans. on ASSP, Vol. 36,Mar., 1988 pp. 417-420.

[101] White, L.B. and Boashash, B.,"A Design Methodology for the Optimal Estimation of the Spectral Characteristics of Non-Stationary Random Signals," IEEE ASSP Workshop on Spectrum Estimation and Modeling, Minneapolis, MN, 1988, pp.65-70.

[102] White, L.B.,"Some Aspects of the Time-Frequency Analysis of Random Processes," thesis, University of Queensland, Australia, 1989.

[103] White, L.B. Qiu, L. and Boashash, B.,"Instantaneous Frequency Estimation: Statistical Properties; Application to Time-Varying Filtering," Int. Conf. Acoust., Speech, Sig. Proc., Albuquerque, NM, 1990.

[104] White, L.B.,"Robust Methods for Instantaneous Frequency Estimation," Proc. SPIE Int Conf. on Adv. Sig. Proc. Alg. and Arch., San Diego, CA, July 1990.

[105] Wong, K.M. and Jin, Q., "Estimation of the Time-Varying Frequency of a Signal: the Crámer-Rao Bound and the Application of Wigner Distribution" IEEE Trans. on ASSP, Vol. ASSP-38, No.3, March 1990, pp.519-536.

[106] Wong, K.M. and Jin, Q., "Techniques of Analyzing Signals Having Time-Varying Frequencies", CRL Rep. 189, Commun. Res. Lab., McMaster Univ., Hamilton, Ont., Canada, Feb. 1988.

[107] Wong, K.M., Luo, Z.Q. and Jin, Q., "Design of Optimum Signals for the Simultaneous Estimation of Time Delay and Doppler Shift", accepted by IEEE Trans. on Signal Processing.

[108] Wozencraft, J.M. and Jacobs, I.M., "Principles of Communication Engineering", John Wiley and Sons, Ney York, 1968

[109] Woodward, P.M., "Probability and Information Theory with Application to Radar", London, Pergamon, 1953.

[110] Ziomek, L.J., "Underwater Acoustics", Academic Press, Inc., 1985.