ELECTRICAL AND COMPUTER ENGINEERING
DEPARTMENT

CLEMSON UNIVERSITY

CLEMSON, SC 29634-0915

TR-112894-3570P

MAXIMUM-LIKELIHOOD SPECTRAL ESTIMATION AND ADAPTIVE FILTERING TECHNIQUES WITH APPLICATION TO AIRBORNE DOPPLER WEATHER RADAR

by

Jonathan Y. Lai

Radar Systems Laboratory
Technical Report No. 20

Maximum-Likelihood Spectral Estimation and Adaptive Filtering Techniques with Application to Airborne Doppler Weather Radar

by

Jonathan Y. Lai

Technical Report #20
November 28, 1994

Radar Systems Laboratory
Electrical and Computer Engineering Department
Clemson University
Clemson, SC 29634-0915

Windshear Detection Radar Signal Processing Studies
Grant NAG-1-928
National Aeronautics and Space Administration
Langley Research Center
Hampton, VA 23665

ABSTRACT

This dissertation focuses on the signal processing problems associated with the detection of hazardous windshears using airborne Doppler radar when weak weather returns are in the presence of strong clutter returns. In light of the frequent inadequacy of spectral-processing oriented clutter suppression methods, we model a clutter signal as multiple sinusoids plus Gaussian noise, and propose adaptive filtering approaches that better capture the temporal characteristics of the signal process. This idea leads to two research topics in signal processing: (1) signal modeling and parameter estimation, and (2) adaptive filtering in this particular signal environment. A high-resolution, low SNR threshold ML frequency estimation and signal modeling algorithm is devised and proves capable of delineating both the spectral and temporal nature of the clutter return. Furthermore, the LMS-based adaptive filter's performance for the proposed signal model is investigated, and promising simulation results have testified to its potential for clutter rejection leading to more accurate estimation of windspeed thus obtaining a better assessment of the windshear hazard.

ACKNOWLEDGMENTS

The author would like to express his sincere gratitude to all of those whose help and encouragement made the completion of this thesis possible. First of all, he would like to thank his advisor, Dr. Ernest G. Baxa, Jr., for his support and mentorship. Also to be acknowledged are the Ph.D. committee members, Dr. John Komo, Dr. Kenneth Wallenius, and Dr. Robert Snelsire, for their time in reviewing this thesis and their classes that led the author up to this point. NASA Langley Research Center is also recognized for their financial support under grant No. NAG-1-928 and for their technical assistance specially from the Antenna and Microwave Research Branch.

Thanks are also due to the students of the Radar System Laboratory at Clemson University for their help and friendship. Particularly, the author would like to thank Mr. David Aalfs, whose expertise and acumen in computer systems has been a blessing to this thesis work.

Precious friendship and hospitality the author received from the Clemson community made his graduate life a pleasant journey. In particular, constant love and encouragement from his host family, Dr. and Mrs. Hugh (Mary) Macaulay, is deeply appreciated. Many brothers and sisters in the body of Christ are also acknowledged for their loving fellowship and faithful prayers.

Finally, the author would like to honor his parents for their love, support, and encouragement throughout his many years in school, without which the completion of his graduate studies would not be possible.

TABLE OF CONTENTS

Table of Contents (Continued)

LIST OF FIGURES

List of Figures (Continued)

CHAPTER 1

INTRODUCTION

1.1 Modeling Discrete-Time Signal

Though conventional time series (signal) analysis is heavily dependent on the twin assumptions of *linearity* and *stationarity*, since the late 1960s, parametric modeling of nonstationary signals has received a great deal of attention [1]. However, the definitions of signal nonstationarity are still diversified among engineering researchers and mathematical statisticians. As opposed to statisticians' interest in the temporal characteristics of an observed data record, signal processors are generally more absorbed in analyzing its spectral content [2, 3]. In Section 1.1.1, we give a brief introduction of some commonly used time-domain approaches for modeling nonstationary signals.

1.1.1 Time Series Analysis Approach

It is generally recognized by statisticians that nonstationary processes may arise in several ways [1, 4].

The first is the "trend plus stationary residual" model

$$x(n) = \mu(n) + \alpha(n)$$

where $\mu(n)$ is a *deterministic* function, and $\alpha(n)$ is a zero-mean stationary processes. An extension of this model, i.e., the "trend–seasonal–irregular" model,

$$x(n) = T(n) + S(n) + R(n)$$

with $T(n)$ as the trend, $S(n)$ the seasonal term, and $R(n)$ an irregular component, is often encountered in econometric time series analysis [5].

The second is the the autoregressive integrated moving average (ARIMA) process

$$\phi_p(B)(1 - B)^d x(n) = \theta_q(B)e(n)$$

where B is a backshift operator such that $B^k x(n) = x(n-k)$, the stationary AR operator $\phi_p(B) = 1 - a_1 B - \cdots - a_p B^p$ and the invertible MA operator $\theta_q(B) = 1 - b_1 B - \cdots - b_q B^q$ share no common factor, and $e(n)$ is a white noise process of zero mean and constant variance. This model postulates that differencing the $x(n)$ process d times will result in a stationary ARMA process. Furthermore, one should notice that the nonstationary characteristics of this model may be seen by interpreting $x(n)$ as the result of passing the white noise process $e(n)$ through an "unstable" IIR filter whose transfer function $\frac{\theta_q(z^{-1})}{\phi_p(z^{-1})(1-z^{-1})^d}$ has d overlapping poles on the unit circle.

The third is the ARMA process with freely varying time-dependent parameters which is a more general class of nonstationary processes in which the time-dependent nature of the processes can be delineated in various ways [1, 6].

Many common signals analyzed in practice are indeed not stationary, and their time-domain characteristics may be captured by the models proposed above. However, in the development of various spectrum estimation algorithms, as addressed by Marple [2] and Kay [3], short data segments from the longer data record may be considered to be *locally stationary*, availing those estimation methods to many real world applications. Assuming the data record $x(n)$ is from a stationary process, the formal definition of the spectrum shows that it is a function strictly of the second-order statistics [2]. The second-order statistics are also assumed to remain unchanged, or *stationary*, over time. Thus, the spectrum is not a complete statistical picture of a random process that may have other information in third- and higher-order statistics. Section 1.1.2 deals with another modeling approach particularly suitable for a narrowband signal process, where the signal's features over both *the time-domain and the frequency-domain* are simultaneously retained.

1.1.2 Sinusoidal Parameter Estimation Approach

The estimation of the frequencies of sinusoidal components embedded in white noise is a problem that arises in many fields such as communications, radar, and sonar

applications, where enhancement (suppression) of *narrowband* signals (interferences) is among the signal processors' major concerns [3, 7]. In this model, given as

$$x(n) = \sum_{i=1}^{p} A_i \exp(j2\pi f_i n + \phi_i) + w(n)$$

where $w(n)$ is complex white Gaussian noise of zero mean and constant variance [8], the signal amplitudes are usually assumed to be constants, whereas the phases are either constants or random variables dependent on the frequency estimation approaches: the approximate maximum likelihood techniques regard the sinusoidal process as a *deterministic signal* with unknown frequencies, while the eigenanalysis approaches employ a WSS *random process* model so that the frequencies appear as unknown parameters in an autocorrelation matrix. It should be pointed out that when the sinusoids are of constant phase, this model belongs to the first category of nonstationary processes discussed in the previous section.

As will be addressed in next two chapters, this dissertation focuses on the maximum likelihood frequency estimation problems with the idea of modeling the observed data as a *deterministic* sinusoidal signal with unknown parameters plus noise, though the data itself may be a partial realization of a nonstationary process [3, 9].

1.2 Motivation for Adaptive Processing Scheme

One of the most challenging problems facing modern signal processors is the detection and estimation of weather information using airborne Doppler radar [10, 11]. This application is associated with the development of a new generation airborne Doppler weather radar for detecting windshear which is particularly hazardous to aircraft operating at low altitudes [11, 12, 13]. For an aircraft in a landing mode, with an antenna scanning the airspace along the intended flight path, the radar return represents a sample function of a nonstationary process which in the presence of weather can consist of two slowly time-varying components. A signal component is due to the weather return and a noise component is due to the ground clutter return. Furthermore, in situations where the signal-to-clutter ratio (SCR) is extremely

low, it becomes a very difficult task to suppress the clutter component so that the radar return can be processed to determine windfield characteristics within the search volume.

The general approach in reducing the effects of ground clutter on windspeed estimates made from the Doppler radar return sequences is to use some form of frequency selective filtering [10]. Fixed-notch clutter rejection filters are commonly used to separate the two processes on the basis of *spectral content.* Moreover, in most spectral analysis oriented approaches, two inherent problems are unavoidable. First, good spectrum estimation can only be attained when the signal-to-noise ratio (SNR) is above a certain threshold; Secondly, the signal's *temporal information* (e.g., phase) is often discarded. Accordingly, approaches of this sort will fail to enhance the weak weather return in the presence of strong clutter returns.

An adaptive filter, on the other hand, is an active system that will make appropriate parametric adjustments to the unknown signal environment or its time-varying temporal statistics, by following some type of optimization criterion. For the application of clutter rejection and weather information extraction associated with airborne Doppler radar, some promising results obtained by LMS-based adaptive filtering have been reported by Lai and Baxa [14]. This dissertation serves as an in-depth study of their work. Major contributions are summarized in the following section.

1.3 Contribution and Organization of This Dissertation

The contributions of this dissertation include three aspects. First, as noted in Chapter 2, the major difficulty in ML frequency estimation is that the likelihood function is highly nonlinear with respect to the frequency parameters. Moreover, given a short data record under low SNR conditions, the presence of spectrally close signals makes correct detection a very tough task. Chapter 2 formulates this ML frequency estimation problem, briefly surveys its solutions proposed by the signal processing community, and provides the basic ideas in Wiener filtering and the adaptive LMS

algorithm. All the discussions therein serve as a theoretical basis for the chapters to follow.

In Chapter 3, we devise a high-resolution frequency estimation and signal detection scheme utilizing the expectation-maximization (EM) algorithm. This method efficiently solves the problem of multi-dimensional optimization, and through the inclusion of MDL criterion or Fisher's T_o statistic, attains high probability of correct detection in low SNR situations.

Secondly, convergence responses of the LMS-based adaptive noise canceler to sinusoidal signal processes are investigated in Chapter 4, through extensive simulation. Though the approach taken is somewhat more experimental than analytical, our simulation results provide valuable insights regarding different adaptive filter configurations and design considerations, under a very realistic signal environment encountered in many applications.

Thirdly in Chapter 5 the adaptive filter is investigated through simulation as a process decorrelator to suppress the clutter return and thus enhance the weather return. Results from Chapter 4 have been used as a design guideline for these adaptive filtering schemes, whereas the algorithms in Chapter 3 are employed to model the Doppler weather radar clutter return process. As our simulation results show, even in a low SCR signal scenario, an adaptive signal processing scheme can significantly outperform traditional fixed filtering approaches. In light of all these analysis and simulation results, a concise conclusion and some suggestions for future work are provided in Chapter 6.

CHAPTER 2

ML FREQUENCY ESTIMATION AND ADAPTIVE FILTER

In this chapter, we give a brief overview of two important research topics in signal processing: spectral estimation and adaptive filtering. All the discussions here will serve as the theoretical basis for Chapters 3 and 4 and their applications in Chapter 5. Traditionally, power spectral density estimation has very much relied upon the Fourier transform, the theories of random processes and filter theory. Specifically, classical estimation methods like periodogram and the Blackman-Tukey spectral estimator are Fourier-transform-based, whereas "modern" spectral estimation, which is a *parametric* modeling approach, mainly depends on random process (e.g., ARMA) and filter theories [2, 3]. Our attention will be directed to the problem of parameter estimation of sinusoids corrupted by noise, a research issue still challenging many signal processors, especially when a short data record is available and high frequency resolution is required [3, 15, 16]. An adaptive filter is an active system that will make appropriate *parametric* adjustments to the unknown signal environment or its time-varying statistics, by following some type of optimization criterion [7, 17]. According to Haykin [7], to derive the recursive algorithms for the operation of adaptive filters, one can identify three distinct methods: (1) an approach based on Wiener filter theory, (2) an approach based on Kalman filter theory, and (3) the method of least squares. We will focus on the first approach, Wiener filtering.

2.1 Basic Assumptions in Signal Model

To begin the discussion of maximum likelihood estimation (MLE) of frequency parameters, specifically assume that the received data vector $\mathbf{y} = [y(0)y(1)\cdots y(N-1)]^T$ consists of p complex sinusoids in complex white Gaussian noise (CWGN)

$$y(n) = \sum_{i=1}^{p} A_i \exp(j2\pi f_i n) + w(n) \tag{2.1}$$

where $A_i = |A_i|e^{j\phi_i}$ is the complex amplitude of the ith sinusoidal component, and $w(n)$ is complex white Gaussian noise with zero mean and variance σ^2. The sinusoidal parameters $\{|A_1|, |A_2|, \cdots, |A_p|, \phi_1, \phi_2, \cdots, \phi_p, f_1, f_2, \cdots, f_p\}$, which consists of amplitudes ($0 < |A_i| < \infty$), phases ($0 \leq \phi_i < 2\pi$), and frequencies ($0 \leq f_i < 1$), are assumed to be constant but unknown and are to be estimated. The idea here is to model the observed data as a *deterministic* sinusoidal signal with unknown parameters plus white noise, though the data itself may be a snapshot of a nonstationary process [3, 9]. It is also important to remember that the discussion to follow assumes that the number of sinusoids, p, is known. Methods to estimate p are discussed in Section 2.3.

2.2 Problem Formulation and Implementation of MLE

2.2.1 Single Sinusoid

To help formulate the ML estimation problem, let us first look at the case of a single sinusoid. Let $\mathbf{e}_1 = [1\ e^{j2\pi f_1}\ e^{j2\pi f_1 2} \cdots e^{j2\pi f_1 (N-1)}]^T$. It is well known in the literature that the MLE of the frequency and amplitude of a single complex sinusoid in complex white Gaussian noise is found by minimizing the *scoring function*, $S(A_1, f_1) = \|\mathbf{y} - A_1\mathbf{e}_1\|^2 = (\mathbf{y} - A_1\mathbf{e}_1)^H(\mathbf{y} - A_1\mathbf{e}_1)$, where H denotes the Hermitian transposition and $\|\cdot\|$ the Euclidean norm. This is equivalent to choosing the frequency at which the periodogram attains its maximum [3]. The MLE of $\{|A_1|, \phi_1, f_1\}$ is given as

$$\begin{aligned}
\hat{f}_1 &= \arg\max_{f_1} \frac{1}{N}\left|\mathbf{e}_1^H\mathbf{y}\right|^2 = \arg\max_{f_1} \frac{1}{N}\left|\sum_{n=0}^{N-1} y(n)\exp(-j2\pi f_1 n)\right|^2 \\
|\hat{A}_1| &= \left|\frac{1}{N}\sum_{n=0}^{N-1} y(n)\exp(-j2\pi \hat{f}_1 n)\right| \\
\hat{\phi}_1 &= \arctan\left[\frac{\mathrm{Im}\left(\sum_{n=0}^{N-1} y(n)\exp(-j2\pi \hat{f}_1 n)\right)}{\mathrm{Re}\left(\sum_{n=0}^{N-1} y(n)\exp(-j2\pi \hat{f}_1 n)\right)}\right].
\end{aligned} \tag{2.2}$$

Note that the frequency estimate is found as the result of a one-dimensional search. To achieve this goal, a coarse FFT followed by a Newton-Raphson search can be implemented. Please refer to the Appendix for details.

2.2.2 Multiple Sinusoids

In vector form, the observed data $\mathbf{y} = [y(0)y(1)\cdots y(N-1)]^T$ can be represented as

$$\mathbf{y} = \sum_{i=1}^{p} A_i \mathbf{e}_i + \mathbf{w} = \mathbf{E}\mathbf{A} + \mathbf{w} \tag{2.3}$$

where $\mathbf{e}_i = [1\ e^{j2\pi f_i}\ e^{j2\pi f_i 2}\ \cdots\ e^{j2\pi f_i(N-1)}]^T$, $\mathbf{E} = [\mathbf{e}_1\ \mathbf{e}_2\cdots\mathbf{e}_p]$, $\mathbf{A} = [A_1\ A_2\cdots A_p]^T$, and $\mathbf{w} = [w(0)\ w(2)\cdots w(N-1)]^T$. For notational simplicity we have concealed the dependence of matrix $\mathbf{E}$ on the frequency parameter vector $\mathbf{f} = [f_1\ f_2\cdots f_p]$.

Based on the white Gaussian assumption, it has been shown that the MLE of $\mathbf{A}$ and $\mathbf{f}$, denoted as $\hat{\mathbf{A}}$ and $\hat{\mathbf{f}}$, can be found by solving the following nonlinear least square problem [3]:

$$\{\hat{\mathbf{A}}, \hat{\mathbf{f}}\} = \arg\min_{\mathbf{A},\mathbf{f}} S(\mathbf{A}, \mathbf{f}) \tag{2.4}$$

$$\begin{aligned} S(\mathbf{A}, \mathbf{f}) &= \|\mathbf{y} - \mathbf{E}\mathbf{A}\|^2 = (\mathbf{y} - \mathbf{E}\mathbf{A})^H(\mathbf{y} - \mathbf{E}\mathbf{A}) \\ &= \sum_{n=0}^{N-1} \left| y(n) - \sum_{i=1}^{p} A_i \exp(j2\pi f_i n) \right|^2 . \end{aligned} \tag{2.5}$$

Fixing $\mathbf{f}$ and thus assuming $\mathbf{E}$ to be a known matrix, $S(\mathbf{A}, \mathbf{f})$ is minimized over $\mathbf{A}$ by

$$\hat{\mathbf{A}} = (\mathbf{E}^H\mathbf{E})^{-1}\mathbf{E}^H\mathbf{y}. \tag{2.6}$$

It is well known that if $\mathbf{f}$ is replaced by its MLE $\hat{\mathbf{f}}$, then $\hat{\mathbf{A}}$ will be the MLE of $\mathbf{A}$ [3]. Substituting (2.6) into (2.5) yields the scoring function

$$\begin{aligned} S(\hat{\mathbf{A}}, \mathbf{f}) &= \mathbf{y}^H(\mathbf{y} - \mathbf{E}\hat{\mathbf{A}}) \\ &= \mathbf{y}^H\mathbf{y} - \mathbf{y}^H\mathbf{E}(\mathbf{E}^H\mathbf{E})^{-1}\mathbf{E}^H\mathbf{y}. \end{aligned} \tag{2.7}$$

Therefore, the MLE $\hat{\mathbf{f}}$ of the frequencies can be found as

$$\begin{aligned} \hat{\mathbf{f}} &= \arg\max_{\mathbf{f}} L(\mathbf{y}; \mathbf{f}) \\ \text{where} \quad L(\mathbf{y}; \mathbf{f}) &= \mathbf{y}^H\mathbf{E}(\mathbf{E}^H\mathbf{E})^{-1}\mathbf{E}^H\mathbf{y}. \end{aligned} \tag{2.8}$$

$L(\mathbf{y}; \mathbf{f})$ is a *highly nonlinear function of the unknown frequencies* and therein lies the central problem of ML estimation. We refer to $L(\mathbf{y}; \mathbf{f})$ as the log-likelihood of $\mathbf{y}$ given $\mathbf{f}$ since it is $\ln f_y(\mathbf{y}; \mathbf{f})$ with terms not associated with $\mathbf{f}$ discarded.

2.2.3 Implementation of MLE

Modern methods for ML frequency estimation generally fall within two classes: linear prediction based methods and eigenstructure-based methods [3, 16]. In the 1980s, the modified forward-backward linear prediction (MFBLP) approach developed by Tufts and Kumaresan [15] achieves great success in that it outperforms other eigendecomposition-based methods in terms of its good frequency resolution at low SNR conditions.

However, since Ziskind and Wax [9] proposed the alternating projection (AP) algorithm in the late 1980s, several AP-oriented approaches that show even better performances than that of the MFBLP method have been developed by signal processors [16, 18, 19]. In Chapter 3, we will devise a new class of high-resolution frequency estimation algorithm that not only possesses some nice computational and statistical features, but also demonstrates performances favorably compared with those advanced methods.

2.3 Order Selection

In this section, we consider two information theoretic criteria [20] found favorable in both time series analysis and signal processing for model order selection, and give their modified versions for deciding the number of sinusoids buried in complex white Gaussian noise. Let $f(\mathbf{y};\Theta)$ denote the PDF of data $\mathbf{y}$ given the *true* parameter vector Θ of p components. Assuming we have an MLE $\hat{\Theta}$ of Θ for each model of order k (i.e., the number of *independently* adjusted parameters in the model) in question, Akaike's information criterion (AIC) [21] is defined as

$$AIC(k) = -2\ln f(\mathbf{y};\hat{\Theta}) + 2k, \tag{2.9}$$

and according to Wax and Kailath [20], the minimum description length (MDL) principle originally proposed by Rissanen [22] is given as

$$MDL(k) = -\ln f(\mathbf{y};\hat{\Theta}) + \frac{1}{2}k\ln N. \tag{2.10}$$

To tailor AIC for modeling the signal described by (2.3), Kay [3] suggests that we can choose the number of sinusoids by minimizing

$$AIC(i) = 2N \ln S_i(\hat{\mathbf{A}}, \hat{\mathbf{f}}) + 6i \tag{2.11}$$

where i is the number of sinusoids, and k in (2.9) has been replaced by $3i$ to account for the three parameters (amplitude, phase, and frequency) associated with each sinusoid. Before using MDL, we need to remember that Rissanen's basic idea behind this order selection principle is to find the shortest code length (number of bits) to *encode* an observed data set $\mathbf{y}$, assuming the parameter $\hat{\Theta}$ is a vector of k *real-valued* components [22]. As expressed in (2.5), the parameters to be estimated are associated with the *complex-valued* multiple sinusoidal signal vector $\mathbf{EA}$. Therefore, replacing k in (2.10) by $2k$, and following the same rationale used by Kay to account for the three parameters associated with each sinusoid, the MDL criterion can be presented as

$$MDL(i) = N \ln S_i(\hat{\mathbf{A}}, \hat{\mathbf{f}}) + 3i \ln N \tag{2.12}$$

where i is the number of sinusoids. Notice that the first terms in (2.11) and (2.12) involve the negative of the log-likelihood of an ML estimate, and the second terms represent the "cost" for the selected model order (correspondingly the number of parameters). However, as observed in (2.12), the MDL criterion will "penalize" more for redundancy in modeling than the AIC. Performance comparison of these two model-order estimators will be given in Section 3.4 through simulation.

2.4 Wiener Filtering and Widrow-Hoff LMS Algorithm

Adaptive filters are mainly derived from linear optimum filter theory for wide-sense stationary stochastic processes [7, 17]. Regarding the filter specification, two choices have to be made clear. First, the choice of a finite impulse response (FIR) or an infinite impulse response (IIR) for the filter is dictated by practical considerations. Secondly, the type of *statistical criterion* used for the optimization is often influenced

by mathematical tractability [7]. Our attention will be confined to the more well-developed FIR filter theory and the *least-mean-square* (LMS) algorithm, a simplified criterion derived from the method of steepest descent.

2.4.1 Wiener Filter

This section begins with the discussion of a class of optimum linear discrete-time filters known collectively as *Wiener filters.* Consider the block diagram of Figure 2.1 for this specific filtering problem. The filter *input* consists of a *time series* $y(0)$, $y(1)$, $y(2)$ $\cdots$, and the filter is itself characterized by the *impulse response* w_0, w_1, w_2 $\cdots$. At some time instant n, the filter produces an *output* denoted by $v(n)$. This output is used to provide an *estimate* of a *desired response* $d(n)$. With the filter input and the desired response representing single realizations of respective stochastic processes, the estimation is accompanied by an error with statistical characteristics of its own. In particular, the *estimation error*, denoted by $e(n)$, is defined as the difference between the desired response $d(n)$ and the filter output $v(n)$. The basic requirement is to make the estimation error "as small as possible".

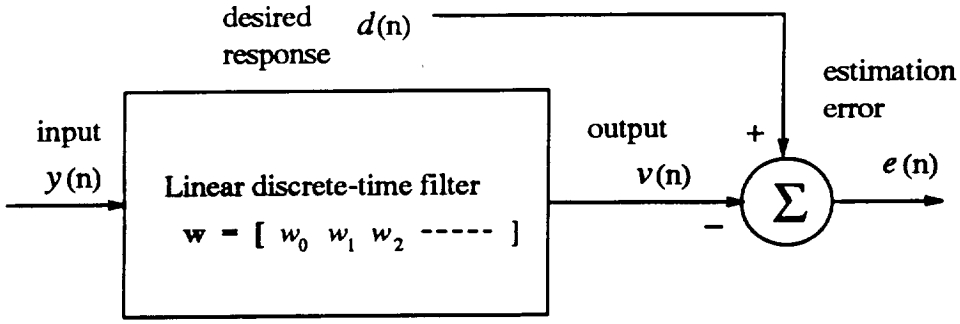


Figure 2.1. Structure for the Wiener filtering problem.

Generally speaking, the filter design can be *optimized* in the sense of minimizing a *cost function*, or *index of performance.* Among others, the possible choices are the

mean-square value, the expected absolute value, or the third (or higher) power of the absolute value of the estimation error [7, 17]. The choice of the minimum mean-square-error (MSE) criterion has attracted more research attention than the others, because it has the advantage of leading to tractable mathematics [7]. In particular, the MSE criterion results in a second-order dependence for the cost function on the unknown coefficients $\mathbf{w}$ in the impulse response of the filter. Moreover, the cost function has a distinct minimum that uniquely defines the statistically optimum design of the filter.

Assume that the filter input and the desired response are single realizations of jointly wide-sense stationary stochastic processes, and denote the FIR coefficient vector of the filter as $\mathbf{w} = [w_0\, w_1 \cdots w_{M-1}]^T$. The filter output $v(n)$ and the estimation error $e(n)$ at discrete time n are defined as

$$\begin{aligned} v(n) &= \sum_{m=0}^{M-1} w_m y(n-m) = \mathbf{w}^T \mathbf{y}(n), \\ \text{and} \quad e(n) &= d(n) - v(n) \end{aligned} \tag{2.13}$$

where $\mathbf{y}(n) = [y(n)\, y(n-1) \cdots y(n-M+1)]^T$. To optimize the filter design, we choose to minimize the mean-square value of the estimation error, $J = E[e(n)e^*(n)] = E[|e(n)|^2]$. For the cost function $J(\mathbf{w})$ to attain its minimum value, all the elements of the complex gradient vector ∇J must be simultaneously equal to zero. (For a detail treatment of the concept of a complex gradient operator, please refer to Kay [23], Haykin [7], and Brandwood [24].) Under this condition, the filter is said to be *optimum in the mean-square-error sense.*

Let $\mathbf{R}$ denote the M-by-M correlation matrix of the tap input vector $\mathbf{y}(n) = [y(n)\, y(n-1) \cdots y(n-M+1)]^T$ in the transversal (FIR) filter of Figure 2.1: $\mathbf{R} = E[\mathbf{y}(n)\mathbf{y}^H(n)]$. Correspondingly, let $\mathbf{P}$ denote the M-by-1 cross-correlation vector between the tap input vector and the desired response $d(n)$: $\mathbf{P} = E[\mathbf{y}(n)d^*(n)]$. Let $v_o(n)$ denote the output produced by the filter optimized in the MSE sense, with $e_o(n)$ as the corresponding estimation error. The essential idea of Wiener filtering hinges

on two important results. First, the necessary and sufficient condition for minimizing the cost function J is that the corresponding value of the estimation error $e_o(n)$ is *orthogonal* to each input sample $y(n)$ contributing to the estimation of the desired response at time n. Furthermore, it can be shown that under the same optimum condition, $e_o(n)$ and the estimate of the desired response $v_o(n)$ are *orthogonal* to each other. These statements constitute the *principle of orthogonality* and its corollary [7], which in mathematical terms are given as

$$\begin{aligned} E[y(n-m)e_o^*(n)] &= 0, \quad m = 0, 1, 2, \cdots, M-1, \\ \text{and} \quad E[v_o(n)e_o^*(n)] &= 0. \end{aligned} \tag{2.14}$$

Secondly, the *optimum tap-weight vector* (impulse response) of the transversal filter, denoted as $\mathbf{w}_o = [w_{o0}\, w_{o1} \cdots w_{oM-1}]^T$, can be obtained by solving the Wiener-Hopf equations

$$\mathbf{R}\mathbf{w}_o^* = \mathbf{P}. \tag{2.15}$$

Therefore, $\mathbf{w}_o^* = \mathbf{R}^{-1}\mathbf{P}$, assuming the correlation matrix $\mathbf{R}$ is nonsingular.

2.4.2 LMS Algorithm

When the filter's performance index $J = E[|e(n)|^2]$ is a known function of $\mathbf{w}$, Newton's search method can be applied to minimize the required number of iterations. However, in many practical adaptive system applications the cost function $J(\cdot)$ is unknown and must be measured or estimated on the basis of stochastic input data. Among others, the method of steepest descent, which adjusts the filter weight vector in the direction of the gradient ∇J at each iteration step, has thus far proven to be the most widely applicable. If it were possible to make exact measurements of the gradient vector $\nabla J(n)$ at each iteration, and if the step size μ is suitably chosen, then the tap-weight vector computed by using the steepest-descent algorithm would indeed converge to the optimum Wiener solution. In realty, however, exact measurements of the gradient vector are not possible, since this would require prior knowledge of

both the correlation matrix $\mathbf{R}$ of the tap inputs and the cross-correlation vector $\mathbf{P}$ [7]. Consequently, the gradient vector of $J = E[|e(n)|^2]$ must be *estimated* from the available data. To simplify this problem, Widrow *et al.* [17, 25] use $|e(n)|^2$ itself as an estimate of J, and suggest that the required direction in which the weight vector should be changed is the opposite (or negative) direction of the maximum instantaneous rate of increase of the error power $|e(n)|^2$ with the weight vector. They come up with the LMS algorithm, which updates the filter coefficients according to

$$\begin{aligned} e(n) &= d(n) - \mathbf{w}^T(n)\mathbf{y}(n) \\ \mathbf{w}(n+1) &= \mathbf{w}(n) + \mu e(n)\mathbf{y}^*(n) \end{aligned} \tag{2.16}$$

where $\mathbf{w}(n) = [w_0(n)\ w_1(n)\ \cdots\ w_{M-1}(n)]^T$ is the filter coefficients estimate at time instant n and μ the adaptation gain or step size.

In the LMS algorithm, the correction term $\mu e(n)\mathbf{y}^*(n)$ applied to the tap-weight vector $\mathbf{w}(n)$ at time $n+1$ is directly proportional to the tap-input vector $\mathbf{y}(n)$. Therefore, when $\mathbf{y}(n)$ is large, the LMS algorithm experiences a *gradient noise amplification* problem [7]. To overcome this problem, we may use the *normalized least-mean-square* (NLMS) algorithm [26], a specific form of the LMS algorithm with a reparameterized step size, viz.,

$$\begin{aligned} e(n) &= d(n) - \mathbf{w}^T(n)\mathbf{y}(n) \\ \mathbf{w}(n+1) &= \mathbf{w}(n) + \frac{\tilde{\mu}}{\|\mathbf{y}(n)\|^2} e(n)\mathbf{y}^*(n) \end{aligned} \tag{2.17}$$

where $\tilde{\mu}$ is the real positive step size. Using the idea of the projection algorithm [27] in the control literature, Slock [28] indicates that the NLMS algorithm is a potentially faster converging algorithm compared to the LMS algorithm, when the design of the adaptive filter is based on the usually quite limited knowledge of its input statistics. This performance advantage of the NLMS algorithm over its LMS precedent is also observed in the area of adaptive radar signal processing [29].

Although the applications of adaptive filtering are quite different in nature, nevertheless, they have one basic common feature, i.e., an input vector and a desired response are used to compute an estimation error, which is in turn employed to control the values of a set of adjustable filter coefficients. Dependent on the manner in which the desired response is extracted, the functions of the four basic classes of adaptive filtering applications can be categorized as: (1) system identification; (2) inverse modeling; (3) linear prediction; (4) interference (noise) canceling [7, 17].

Figure 2.2 shows the block diagram of a system identification configuration where an adaptive filter is used to provide a linear model that represents the best fit (in some sense) to an *unknown plant* characterized by its impulse response **h**. The plant and the adaptive filter are driven by the same input. The plant output supplies the desired response for the adaptive filter. If the plant is dynamic in nature, the model will be time-varying. Since the 1970s, many published papers have contributed to the understanding and confirmation of the LMS algorithm's performance in tracking an unknown system, especially in the nonstationary signal environment [30, 31, 32]. In Chapter 4, through extensive simulations, we will evaluate the capability of adaptive noise canceler (ANC) for the rejection of sinusoidal interference, following the configuration in Figure 2.2. Performance comparison between the LMS and NLMS algorithms will also be presented.

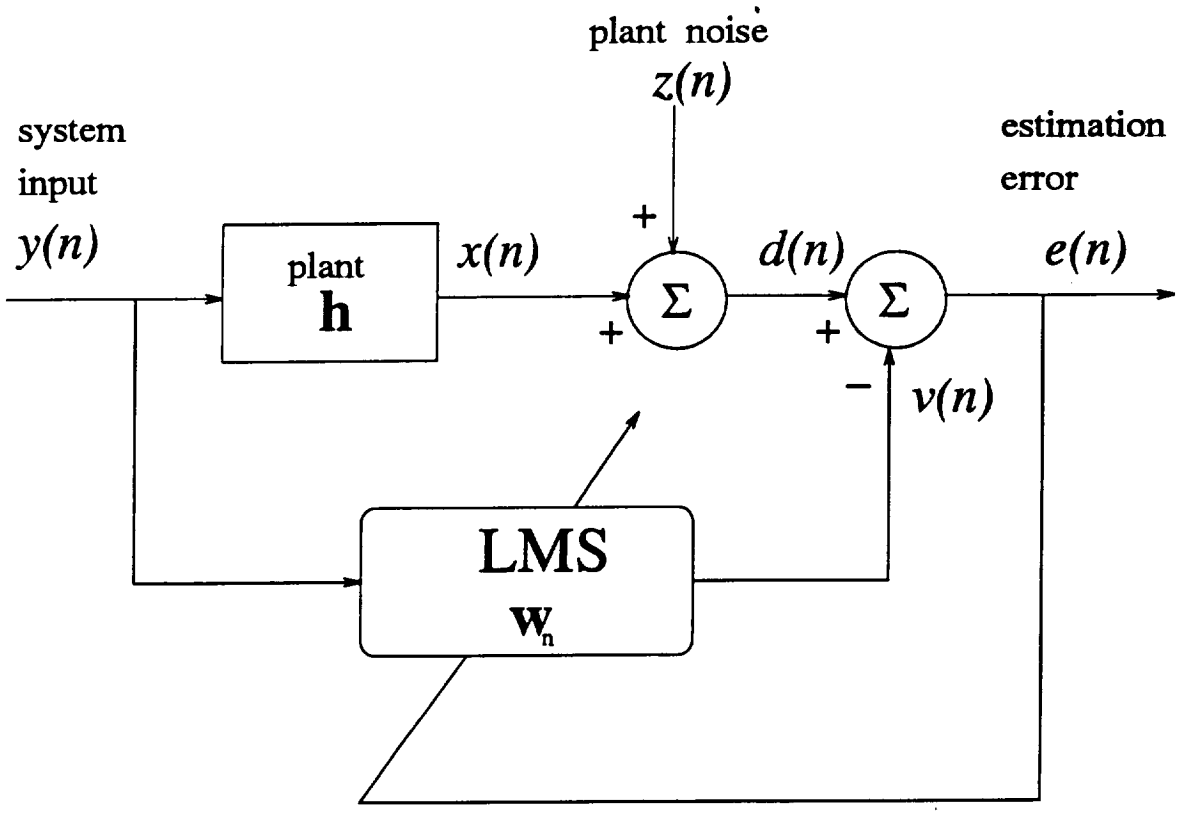


Figure 2.2. System identification via LMS adaptive algorithm.

CHAPTER 3

HIGH-RESOLUTION FREQUENCY ESTIMATION VIA EM

The essential ideas underlying the expectation maximization (EM) algorithm have been presented in special cases by many authors [33, 34, 35]. Dempster, Laird and Rubin [36] first recognized the expectation step (E-step) and the maximization step (M-step) in their general forms, and introduced it for computing maximum likelihood estimates from incomplete data. Since the late 1980s, the EM algorithm has attracted signal processors' attention, and its significant contributions particularly to the area of sensor-array signal processing have been found in the literature [37, 38, 39].

In light of the frequency estimation problems addressed in the previous chapter, we develop a computationally efficient scheme for joint ML estimation of the signal spectral parameters, based on the iterative EM algorithm. This chapter begins with an overview of the data model and the basic ideas behind the EM algorithm, and then shows how this algorithm can be utilized to implement the ML frequency estimator. The important problem of determining the number of signals (order selection) is also addressed. To achieve the goal of correct order selection, we come up with an order recursive combined signal detection and estimation scheme via the EM algorithm. Finally, through intensive simulation, we show how this algorithm attains high spectral resolution capability, low SNR threshold[1], and high correct detection probability, a set of performance indices required of most modern signal processors.

---

[1]SNR threshold is the lowest SNR level above which a frequency estimator approximately attains the performance of an ML estimator.

3.1 Signal Model for the EM Algorithm

3.1.1 Complete and Incomplete Data

The EM algorithm, developed in [36], is a general approach to iterative computation of maximum-likelihood estimates when the observed signal samples can be viewed as *incomplete data.* The term "incomplete data" in its general form implies the existence of two sample spaces $\mathcal{Y}$ and $\mathcal{X}$ and a many-to-one mapping $\mathbf{x} \rightarrow \mathbf{y}(\mathbf{x})$ from $\mathcal{X}$ to $\mathcal{Y}$. Instead of observing the "complete data" $\mathbf{x}$ in $\mathcal{X}$, we observe the "incomplete data" $\mathbf{y}$ in $\mathcal{Y}$. Let the density function of $\mathbf{x}$ be $f_x(\mathbf{x};\theta)$ with parameters $\theta \in \Omega$ and let the density function of $\mathbf{y}$ be given by

$$f_y(\mathbf{y};\theta) = \int_{\mathcal{X}(\mathbf{y})} f_x(\mathbf{x};\theta)d\mathbf{x} \tag{3.1}$$

where $\mathcal{X}(\mathbf{y}) = \{\mathbf{x}\colon\ \mathbf{y}(\mathbf{x}) = \mathbf{y}\}$.

3.1.2 Definition of EM Algorithm

Given the observations $\mathbf{Y} = \mathbf{y}$, the MLE of θ can be obtained by maximizing $f_y(\mathbf{y};\theta)$. However, in many statistical problems, maximization of the complete-data specification $f_x(\mathbf{x};\theta)$ is simpler than that of the incomplete-data specification $f_y(\mathbf{y};\theta)$. Following Dempster *et al.* [36], take logarithms of each side of

$$f_y(\mathbf{y};\theta) = f_x(\mathbf{x};\theta)/f_{x|y}(\mathbf{x};\theta|\mathbf{y})$$

and then take conditional expectations given $\mathbf{Y} = \mathbf{y}$, under a parameter value θ_k, to obtain

$$\begin{aligned} L(\mathbf{y};\theta) &= \ln f_y(\mathbf{y};\theta) \\ &= E_x[\ln f_x(\mathbf{x};\theta)|\mathbf{Y} = \mathbf{y};\theta_k] - E_{x|y}[\ln f_{x|Y=y}(\mathbf{x};\theta)|\mathbf{Y} = \mathbf{y};\theta_k] \\ &= Q(\theta|\theta_k) - H(\theta|\theta_k). \end{aligned} \tag{3.2}$$

It is a well-known consequence of Jensen's inequality that $H(\theta|\theta_k) \leq H(\theta_k|\theta_k)$ [36, 40], and thus implies

$$\forall\theta \rightarrow Q(\theta|\theta_k) > Q(\theta_k|\theta_k), \quad L(\mathbf{y};\theta) > L(\mathbf{y};\theta_k). \tag{3.3}$$

Letting θ_k denote the kth guess of the MLE of θ, we then have this iterative EM algorithm.

Expectation (E-step): Determine the average log-likelihood of the complete data

$$\begin{aligned} Q(\theta|\theta_k) &= E_x[\ln f_x(\mathbf{x};\theta)|\mathbf{Y}=\mathbf{y};\theta_k] \\ &= \int \ln f_x(\mathbf{x};\theta) f_{x|y}(\mathbf{x}|\mathbf{y};\theta_k) d\mathbf{x}. \end{aligned} \tag{3.4}$$

Maximization (M-step): Maximize the average log-likelihood of the complete data

$$\theta_{k+1} = \arg\max_{\theta} Q(\theta|\theta_k). \tag{3.5}$$

At convergence we hopefully will have the MLE. This issue will be discussed later in Section 3.4.

3.2 ML Frequency Estimator via EM Algorithm

3.2.1 A Model for Signal Decompositions

Consider again the observed signal $y(n)$ in Section 2.1 which consists of p complex sinusoids corrupted by complex white Gaussian noise

$$y(n) = \sum_{i=1}^{p} A_i \exp(j2\pi f_i n) + w(n), \quad n = 0, 1, \cdots, N-1 \tag{3.6}$$

where $A_i = |A_i|e^{j\phi_i}$ is the complex amplitude of the ith sinusoidal component, and $w(n)$ is complex white Gaussian noise with zero mean and variance σ^2. In the EM formulation, we call $\{y(n)\}$ *incomplete data*. Suppose there exists the unobservable *complete data* $\{x_1(n), x_2(n), \cdots, x_p(n)\}$

$$x_i(n) = A_i \exp(j2\pi f_i n) + v_i(n), \quad n = 0, 1, \cdots, N-1;\ i = 1 \cdots p \tag{3.7}$$

where $v_i(n)$ is complex white Gaussian noise with zero mean and variance ${\sigma_i}^2$, and $v_i(n)$ is independent of $v_j(n)$ whenever $i \neq j$. Let $\{y(n)\}$ and $\{x_1(n), x_2(n), \cdots, x_p(n)\}$ be related through the following noninvertible many-to-one mapping

$$y(n) = \sum_{i=1}^{p} x_i(n)\ , \tag{3.8}$$

which implies

$$w(n) = \sum_{i=1}^{p} v_i(n) \quad \text{and} \quad \sigma^2 = \sum_{i=1}^{p} {\sigma_i}^2,$$
$$\text{or equivalently,} \quad \sum_{i=1}^{p} \beta_i = \sum_{i=1}^{p} \frac{\sigma_i^2}{\sigma^2} = 1. \tag{3.9}$$

3.2.2 Frequency Estimation via EM: a Brief Derivation

Let $\mathbf{x}_i = [x_i(0)\, x_i(1) \cdots x_i(N-1)]^T$, $\mathbf{e}_i = \mathbf{e}_i(f_i) = [1\, e^{j2\pi f_i}\, e^{j2\pi f_i 2} \cdots e^{j2\pi f_i(N-1)}]^T$, $\theta_i = \{f_i, A_i\}$, $\mathbf{X} = [\mathbf{x}_1\, \mathbf{x}_2 \cdots \mathbf{x}_p]$, $\Theta = \{\theta_1, \theta_2, \cdots \theta_p\}$, and $\Theta_k = \{\theta_{1_k}, \theta_{2_k}, \cdots \theta_{p_k}\}$, the kth guess of the MLE of Θ. From (3.7) we have, upon noting the independent data set assumption,

$$\begin{aligned}
\ln f_x(\mathbf{X};\Theta) &= \sum_{i=1}^{p} \ln f(\mathbf{x}_i, \theta_i) \\
&= \sum_{i=1}^{p} \ln \left\{ \frac{1}{(\pi\sigma_i^2)^N} \exp\left[-\frac{1}{\sigma_i^2} \sum_{n=0}^{N-1} |x_i(n) - A_i \exp(j2\pi f_i n)|^2 \right] \right\} \\
&= g - \sum_{i=1}^{p} \frac{1}{\sigma_i^2} \sum_{n=0}^{N-1} |x_i(n) - A_i \exp(j2\pi f_i n)|^2 \\
&= g - \sum_{i=1}^{p} \frac{1}{\sigma_i^2} \|\mathbf{x}_i - A_i \mathbf{e}_i\|^2
\end{aligned} \tag{3.10}$$

where g is a constant independent of the parameter set Θ.

Given the observed signal $\mathbf{y}$ and previous parameter estimates Θ_k, by taking the expectation of (3.10) we find

$$\begin{aligned}
Q(\Theta|\Theta_k) &= E_x[\ln f_x(\mathbf{X};\Theta)|\mathbf{y};\Theta_k] \\
&= E(g|\mathbf{y};\Theta_k) - \sum_{i=1}^{p} \frac{1}{\sigma_i^2} \|E(\mathbf{x}_i|\mathbf{y};\Theta_k) - A_i\mathbf{e}_i\|^2 \\
&= E(g|\mathbf{y};\Theta_k) - \sum_{i=1}^{p} \frac{1}{\sigma_i^2} S_i(A_i, f_i)
\end{aligned} \tag{3.11}$$

where $\theta_{i_k} = \{f_{i_k}, A_{i_k}\}$ is the kth MLE guess of θ_i. Using the standard result for conditional expectations of jointly Gaussian random vectors, and following the similar derivation by Kay [23] for the case of real sinusoidal signal, it can be shown that

$$\begin{aligned}
\hat{\mathbf{x}}_i &= E(\mathbf{x}_i|\mathbf{y};\Theta_k) \\
&= A_{i_k}\mathbf{e}_{i_k} + \frac{\sigma_i^2}{\sigma^2}\left(\mathbf{y} - \sum_{i=1}^{p} A_{i_k}\mathbf{e}_{i_k}\right)
\end{aligned} \tag{3.12}$$

where $\mathbf{e}_{i_k} = \mathbf{e}_{i_k}(f_{i_k}) = [1 e^{j2\pi f_{i_k}} e^{j2\pi f_{i_k}2} \cdots e^{j2\pi f_{i_k}(N-1)}]^T$, and $\hat{\mathbf{x}}_i = [\hat{x}_i(0)\hat{x}_i(1) \cdots \hat{x}_i(N-1)]^T$. Note that $E(\mathbf{x}_i|\mathbf{y};\Theta_k)$ can be thought of as an estimate of $x_i(n)$ in the complete data set since from (3.12),

$$\hat{x}_i(n) = A_{i_k} \exp(j2\pi f_{i_k} n) + \frac{\sigma_i^2}{\sigma^2}\left(y(n) - \sum_{i=1}^{p} A_{i_k} \exp(j2\pi f_{i_k} n)\right). \tag{3.13}$$

Obviously, in (3.11), to maximize $Q(\Theta|\Theta_k)$ with respect to Θ is to minimize each $S_i(A_i, f_i)$ individually, knowing that $E(g|\mathbf{y};\Theta_k)$ is independent of Θ. According to Section 2.2.1, minimization of $S_i(A_i, f_i)$, the scoring function for single sinusoidal parameter estimation, can be achieved by choosing

$$\begin{aligned} f_{i_{k+1}} &= \arg\max_{f_i} \frac{1}{N}\left|\mathbf{e}_i^H \hat{\mathbf{x}}_i\right|^2 \\ \text{and} \quad A_{i_{k+1}} &= \frac{1}{N}\mathbf{e}_{i_{k+1}}^H \hat{\mathbf{x}}_i, \quad \text{for } i = 1, \cdots, p. \end{aligned} \tag{3.14}$$

Up to this point, it can be clearly recognized that (3.13) (E-step) and (3.14) (M-step) constitute the *iterative* EM algorithm for ML frequency estimation.

Notice that (3.14) represents a *one-dimensional MLE processor* producing its by-product $A_{i_{k+1}}$ as the ML amplitude estimate of the *single-sinusoid* in $\hat{x}_i(n)$. According to Section 2.1.2, given the current frequency estimate $\mathbf{f}_{k+1} = [f_{1_{k+1}}\ f_{2_{k+1}} \cdots f_{p_{k+1}}]$,

$$\mathbf{A}_{k+1} = (\mathbf{E}_{k+1}^H \mathbf{E}_{k+1})^{-1} \mathbf{E}_{k+1}^H \mathbf{y}$$

is the amplitude estimate that will maximize the *joint* log-likelihood $L(\mathbf{y};\mathbf{f}_{k+1})$, where $\mathbf{E}_{k+1} = [\mathbf{e}_{1_{k+1}}\ \mathbf{e}_{2_{k+1}} \cdots \mathbf{e}_{p_{k+1}}]$. In other words, using this amplitude estimate instead of that produced by (3.14) will lead to a more *generalized* EM (GEM) algorithm as described in Dempster *et al.* [36].

Finally, since the σ_i^2's are not unique, they can be chosen arbitrarily as long as (3.9) is satisfied. Recognizing that after the kth iteration, $\eta_{i_k} = \frac{|A_{i_k}|^2}{\sigma_i^2}$ is the SNR estimate of the expected ith sinusoidal component $\hat{x}_i(n)$, it is quite reasonable to choose σ_i^2 such that

$$\eta_{1_k} = \eta_{2_k} = \cdots = \eta_{p_k} = \text{constant}$$

for each MLE processor in (3.14). Considering this choice of σ_i^2 in (3.9), it is very straightforward to show

$$\beta_i = \frac{\sigma_i^2}{\sigma^2} = \frac{|A_{i_k}|}{\sum_{i=1}^{p} |A_{i_k}|}, \quad \text{for } i = 1, 2, \cdots, p.$$

Now, the *generalized* EM frequency estimation algorithm is summarized below.

• **Initialization**: Given the initial frequency estimate $\mathbf{f}_0 = [f_{1_0}\ f_{2_0}\ \cdots\ f_{p_0}]$, and thus the initial amplitude estimate $\mathbf{A}_0 = (\mathbf{E}_0^H \mathbf{E}_0)^{-1} \mathbf{E}_0^H \mathbf{y}$, continue the following EM iteration until $\|\mathbf{f}_{k+1} - \mathbf{f}_k\|_\infty \leq \epsilon$, where $\|\mathbf{h}\|_\infty = \max_i |h_i|$.

• **Expectation (E-step)**: Decompose y(n) into a set of *expected* sinusoidal components $\{\hat{x}_i(n),\ i = 1 \cdots p\}$

$$\begin{aligned} \hat{x}_i(n) &= A_{i_k} \exp(j2\pi f_{i_k} n) + \frac{|A_{i_k}|}{\sum_{i=1}^{p} |A_{i_k}|} \left(y(n) - \sum_{i=1}^{p} A_{i_k} \exp(j2\pi f_{i_k} n) \right), \\ \text{or } \hat{\mathbf{x}}_i &= A_{i_k} \mathbf{e}_{i_k} + \frac{|A_{i_k}|}{\sum_{i=1}^{p} |A_{i_k}|} \left(\mathbf{y} - \sum_{i=1}^{p} A_{i_k} \mathbf{e}_{i_k} \right). \end{aligned} \tag{3.15}$$

• **Maximization (M-step)**: Maximize the log-likelihood of each expected sinusoidal component separately by finding

$$f_{i_{k+1}} = \arg\max_{f_i} \frac{1}{N} \left| \mathbf{e}_i^H \hat{\mathbf{x}}_i \right|^2, \quad i = 1, 2, \cdots, p \tag{3.16}$$

$$\text{and} \quad \mathbf{A}_{k+1} = (\mathbf{E}_{k+1}^H \mathbf{E}_{k+1})^{-1} \mathbf{E}_{k+1}^H \mathbf{y}. \tag{3.17}$$

At its convergence, denote the final frequency and amplitude estimates as $\hat{\mathbf{f}}_p = [\hat{f}_1\ \hat{f}_2\ \cdots\ \hat{f}_p]$ and $\hat{\mathbf{A}}_p = [\hat{A}_1\ \hat{A}_2\ \cdots\ \hat{A}_p]^T$. Define the pth-order model *residual* $\mathbf{z}_p$ as

$$\mathbf{z}_p \triangleq \mathbf{y} - \sum_{i=1}^{p} \hat{A}_i \hat{\mathbf{e}}_i \tag{3.18}$$

where $\hat{\mathbf{e}}_i = \hat{\mathbf{e}}_i(\hat{f}_i) = [1\ e^{j2\pi \hat{f}_i}\ e^{j2\pi \hat{f}_i 2} \ \cdots\ e^{j2\pi \hat{f}_i (N-1)}]^T$. From now on, for the sake of notational clarity, we will refer to this *EM frequency estimator* by the following formula:

$$\{\hat{\mathbf{f}}_p, \hat{\mathbf{A}}_p, \mathbf{z}_p\} = \mathcal{EM}(\mathbf{y}, p, \{\mathbf{f}_0, \mathbf{A}_0\}).$$

3.3 Combined Signal Detection and Estimation

Like most iterative frequency estimation algorithms, the EM algorithm requires a reasonably accurate initial parameter estimate Θ_0 to begin with. Taking a closer look at (3.15), (3.16) and (3.18), it is not difficult to observe some inherent "order-recursive" features associated with this algorithm. Let p be the true model order (number of sinusoidal signals), and m be the order we pick. First notice that when $m = 1$, $\hat{\mathbf{x}}_1 = \mathbf{y}$ according to (3.15), and $\hat{f}_1$ is the frequency at which the periodogram of $\mathbf{y}$ attains its maximum. Secondly, the model residual $\mathbf{z}_m$ defined by (3.18) reveals to us the signal's structural remnant yet to be modeled, when $m < p$. Based upon the above observations, we present the *order-recursive EM algorithm for ML frequency estimation* as follows:

- Let $m = 1$, and $\hat{\mathbf{x}}_m = \mathbf{y}$. Find $\hat{f}_1$, $\hat{A}_1$ and $\mathbf{z}_1$ using (3.16)~(3.18). $\hat{\mathbf{f}}_1 = [\hat{f}_1]$.
- For $m = 2$ to p
  do
  1. *initialization*:

$$f_{m_0} = \arg\max_{f_m} \frac{1}{N} \left| \mathbf{e}_m^H \mathbf{z}_{m-1} \right|^2. \quad \text{Let } \mathbf{f}_0 = [\hat{\mathbf{f}}_{m-1}\ f_{m_0}].$$

$$\text{Calculate } \mathbf{A}_0 = (\mathbf{E}_0^H \mathbf{E}_0)^{-1} \mathbf{E}_0^H \mathbf{y}, \quad \text{where } \mathbf{E}_0 = \mathbf{E}_0(\mathbf{f}_0).$$

  2. EM *iteration*:

$$\{\hat{\mathbf{f}}_m, \hat{\mathbf{A}}_m, \mathbf{z}_m\} = \mathcal{EM}(\mathbf{y}, m, \{\mathbf{f}_0, \mathbf{A}_0\}).$$

  end.

At this point, one can see that an interesting feature this algorithm possesses is its built-in initialization procedure. For each model order $m \leq p$, the initial frequency estimate is the *optimal* (in the ML sense) estimate of previous order $(m-1)$ *plus* one additional estimate extracted from the model residual $\mathbf{z}_{m-1}$.

When the true model order p is unknown, some sort of "stopping rule" has to be applied to the order-recursive EM algorithm. Apart from the information theoretic criteria mentioned in Section 2.3 for order selection, a natural way to stop modeling

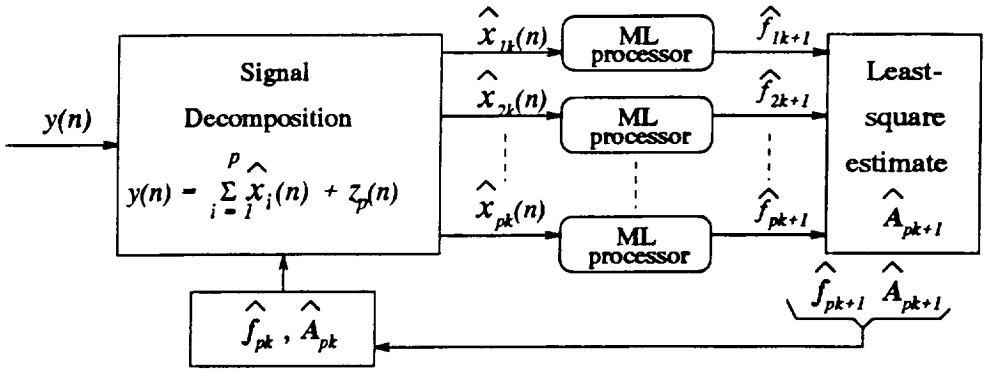
the sinusoids-plus-noise time series is to test the whiteness of the model residual $\mathbf{z}_m$. Since order increment and frequency estimation in this algorithm hinge on the periodogram of the model residual, periodogram-based methods might be used to construct statistical tests for departure from white noise. If the suspected source of departure is the presence of a single sinusoidal component at unknown frequency f, a natural test statistic is the *maximum periodogram ordinate* [41], $T = \max_f \frac{1}{N}|\mathbf{e}^H\mathbf{y}|^2$. Let $\{I_k\}$ be the periodogram ordinates, where $I_k = \frac{1}{N}|\mathbf{e}_k^H\mathbf{y}|^2$ for $k = 1, 2, \cdots, N$. Under the hypothesis that $\mathbf{y}$ is white noise, the distribution function of T requires the knowledge of the noise variance σ^2. However, in most real world applications, σ^2 is unknown. To solve this problem, Fisher [42] devised the exact distribution function of $T_0 = \frac{T}{\frac{1}{N}\sum_{k=1}^{N} I_k}$, the ratio of the maximum to average periodogram ordinate. In particular, he showed that

$$P(T_0 > Nq) = \sum_{k=1}^{R} \left[\frac{N!}{k!(N-k)!}\right] (-1)^{k-1}(1-kq)^{N-1} \tag{3.19}$$

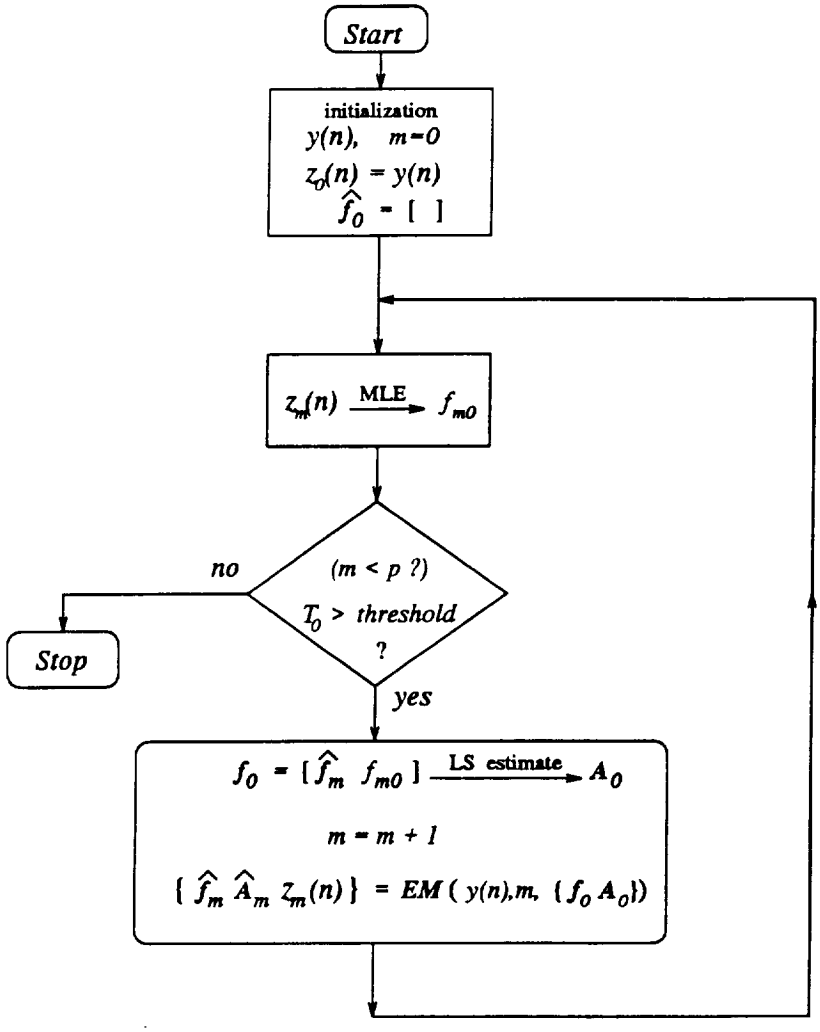
where R is the largest integer less than q^{-1}. Based upon the above arguments, a reasonable way to decide whether to increase the model order is to test the whiteness of model residual $\mathbf{z}_m$ using (3.19). In other words, if the model order we pick is correct, the residual series $\mathbf{z}_p$ is approximately a partial realization of white noise, and thus passes the whiteness test. Figure 3.1(a) shows the block diagram of the EM frequency estimator for a given order p, and Figure 3.1(b) provides a flow-chart representation of the order-recursive EM frequency estimator with Fisher's T_0 statistic as the model order selector. The comparison of order decision performances for Fisher's T_0 statistic and information theoretic criteria like AIC and MDL will be given in the next section.

3.4 EM Convergence Properties and Simulation Results

In general, if the log-likelihood function $L(\Theta)$ (i.e., $L(\mathbf{y};\Theta)$) has several (local or global) maxima and stationary points, convergence of the EM sequence $\{\Theta_k\}$ to either type of point depends on the choice of starting points. This phenomenon has



(a)



(b)

Figure 3.1. (a) Block diagram of EM frequency estimator, and (b) flow chart representation of the combined order-recursive frequency estimation and signal detection algorithm.

been well recognized by statisticians and signal processors [36, 37, 38, 43]. The most striking characteristic of the EM algorithm is that $L(\Theta_{k+1}) \geq L(\Theta_k)$ for any EM sequence $\{\Theta_k\}$. This implies that if $L(\Theta_k)$ is bounded above, then it converges to some L^*. Miller and Fuhrmann [38] suggest that in real world applications, $L(\Theta_k)$ can be upper bounded by properly choosing a closed and bounded parameter space Ω for Θ. More specifically, Wu [43] shows that, if $Q(\Theta|\Phi)$ in (3.11) is continuous in both Θ and Φ, a condition satisfied in most practical situations, then all the limit points of any instance $\{\Theta_k\}$ of an EM algorithm are stationary points of L. Furthermore, $L(\Theta_k)$ converges monotonically to $L^* = L(\Theta^*)$ for some stationary point Θ^*. For a rigorous treatment of EM convergence properties, please refer to Wu [43] and Boyles [44].

In this section, we demonstrate via simulation that the EM algorithm for ML frequency estimation will resolve signal components in situations of small sample size and low SNR which cause other high resolution estimators to fail. We consider two signal scenarios that consist of two or three sinusoids with different frequency separation and SNR level. For each SNR level of interest, 200 Monte Carlo simulations are undertaken. In the first case, the observed signal is generated as

$$y(n) = |A|e^{j2\pi f_1 n} + |A|e^{j0.25\pi}e^{j2\pi f_2 n} + w(n), \quad n = 0, 1, \cdots, 24$$

where $|A|$ is the scalar amplitude, $f_1 = 0.19$, $f_2 = 0.21$, and $w(n)$ is a complex white Gaussian sequence. To help visualize the convergence behavior of EM iterations, Figures 3.2 and 3.3 show the log-likelihood surfaces, surface contours, and trajectories of EM frequency estimation starting from different initial estimates. It is the result of processing a snapshot of $y(n)$ at both low SNR (0 dB) and high SNR (20 dB) individually.

Similarly, in the second case, the observed signal is

$$y(n) = |A|e^{j2\pi f_1 n} + |A|e^{j0.25\pi}e^{j2\pi f_2 n} + |A|e^{j0.85\pi}e^{j2\pi f_3 n} + w(n), \quad n = 0, 1, \cdots, 24$$

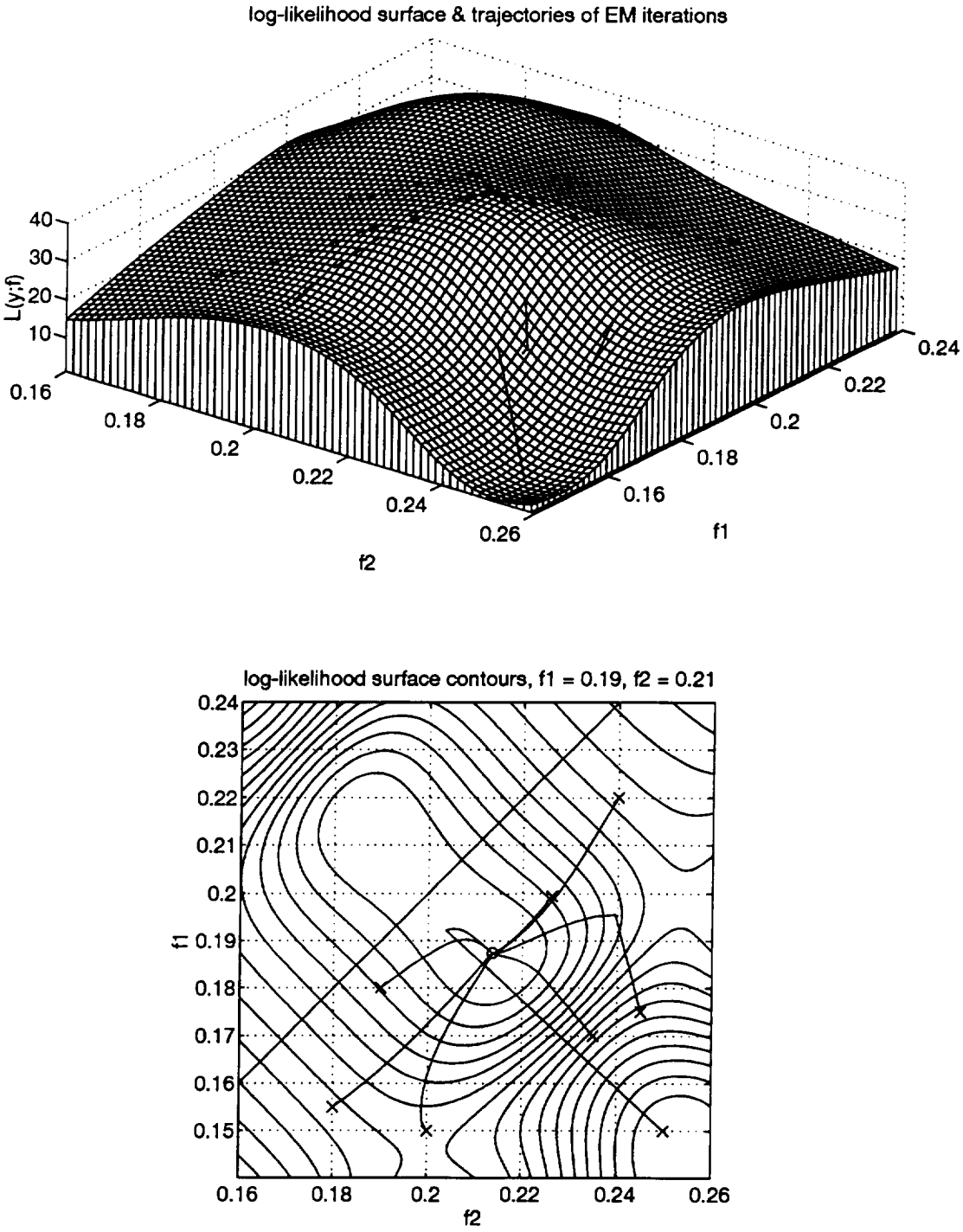


Figure 3.2. Log-likelihood surface, surface contours, and trajectories of EM iterations with different initial frequency estimates and $f_1 = 0.19$, $f_2 = 0.21$, $SNR = 0$ dB, $N = 25$, EM estimate : $\hat{f}_1 = 0.1874$, $\hat{f}_2 = 0.2136$.

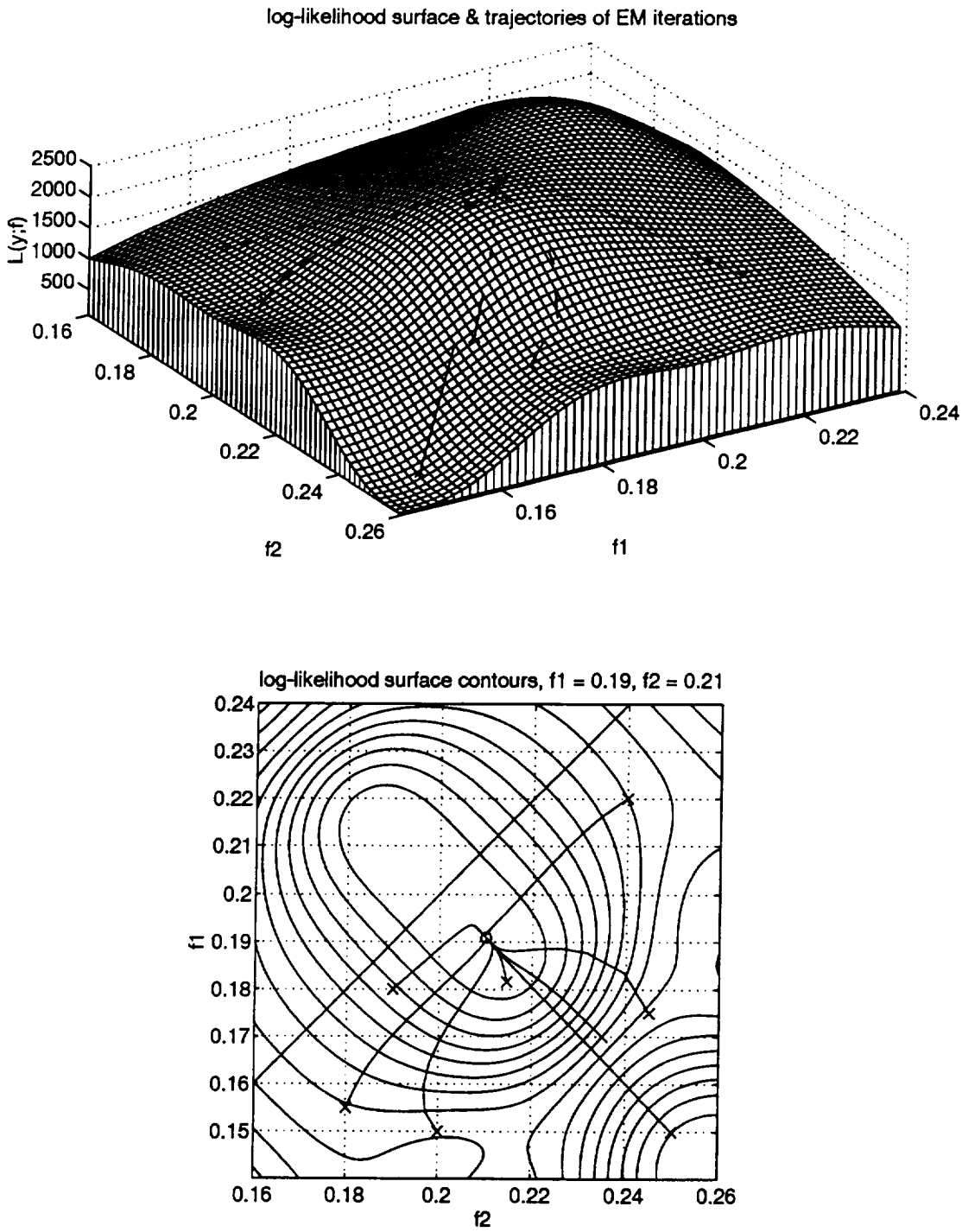


Figure 3.3. Log-likelihood surface, surface contours, and trajectories of EM iterations with different initial frequency estimates and $f_1 = 0.19$, $f_2 = 0.21$, $SNR = 20$ dB, $N = 25$, EM estimate : $\hat{f}_1 = 0.1909$, $\hat{f}_2 = 0.2100$.

where $f_3 = 0.365$, a frequency further away from the first two. Notice that with a sample length of 25, the Fourier resolution bandwidth is $1/25 = 0.04$, which is larger than the frequency spacing of the first two sinusoids. Also, none of the selected signal frequencies is a multiple of this Fourier resolution bandwidth. In terms of being unbiased, small mean square error (MSE), and low SNR threshold (about 2 dB), Figures 3.4 and 3.5 reveal the excellent performance of our algorithm, where the Cramér-Rao bound [3, 45] is used as the benchmark for performance assessment. Although global convergence is still an issue, we show through simulation that when the initial frequency estimate of each component is within approximately one resolution bandwidth of of the global maximum, convergence can be achieved. In fact, the *optimal* built-in initialization procedure of our algorithm, as pointed out in Section 3.3, has successfully carried out this task.

Finally, we demonstrate via simulation the modeling (detection) performances of Fisher's T_0 statistic, MDL and AIC, operating in conjunction with the EM frequency estimation algorithm. The simulated signal is the *three*-sinusoids-plus-noise described above. For the information theoretic criteria (MDL and AIC), model order selected for examination ranges from 1 to 5. Therefore, we count orders 1 and 2 picked by these criteria as *underfitting*, and orders 4 and 5 as *overfitting*. In the use of Fisher's T_0 statistic, H_0 (noise hypothesis) is rejected at the significance level $\alpha = 0.01$, which equivalently thresholds T_0 at 6.9547 according to (3.19) with $N = 25$. Figure 3.6 shows the model selection capabilities of these criteria, based upon 100 Monte Carlo simulations at each SNR level of 0 dB, 3 dB, 5 dB and 10 dB. Notice that at all SNR levels of interest, MDL consistently outperforms Fisher's T_0 statistic with higher probability of correct modeling, and AIC completely overfits the signal model in all simulations.

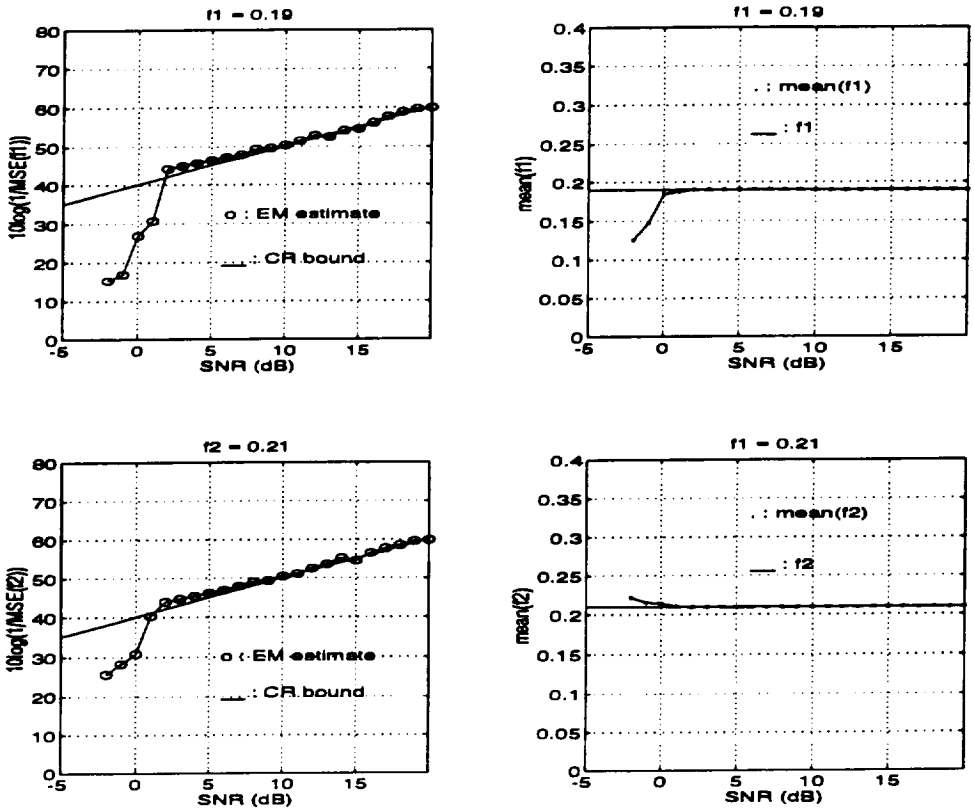


Figure 3.4. Performance of the EM algorithm for ML frequency estimation, $p = 2$, $f_1 = 0.19$, $f_2 = 0.21$, $N = 25$.

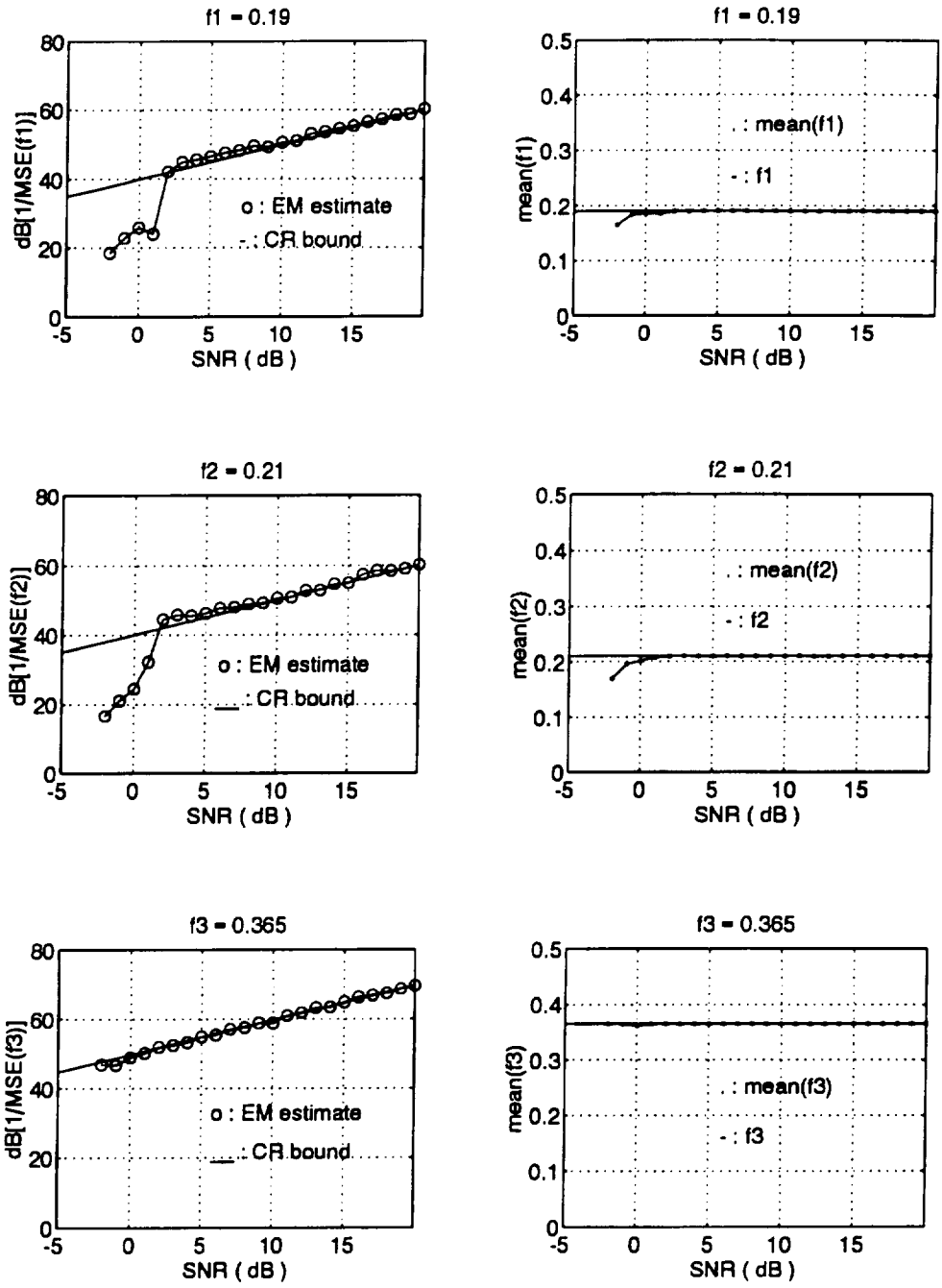


Figure 3.5. Performance of the EM algorithm for ML frequency estimation, $p = 3$, $f_1 = 0.19$, $f_2 = 0.21$, $f_3 = 0.365$, $N = 25$.

SNR = 0 dB			
criterion	underfitting	correct fit	overfitting
Fisher's T_0	37	58	5
MDL	27	70	3
AIC	0	0	100

SNR = 3 dB			
criterion	underfitting	correct fit	overfitting
Fisher's T_0	6	86	8
MDL	1	95	4
AIC	0	0	100

SNR = 5 dB			
criterion	underfitting	correct fit	overfitting
Fisher's T_0	0	88	12
MDL	0	96	4
AIC	0	0	100

SNR = 10 dB			
criterion	underfitting	correct fit	overfitting
Fisher's T_0	0	91	9
MDL	0	97	3
AIC	0	0	100

Figure 3.6. Comparison of detection performances for Fisher's T_0 statistic (level of significance $\alpha = 0.01$), MDL, and AIC.

CHAPTER 4

LMS ADAPTIVE FILTER FOR SINUSOIDAL PROCESS

The suppression of a sinusoidal interference corrupting an information-bearing signal is a problem often encountered in many signal processing applications. The traditional way of dealing with this problem is to design a fixed *notch filter* tuned to the frequency of the interference. To design the filter, precise knowledge of the interfering signal's frequency is always required. When the notch is desired to be very sharp and the sinusoidal interference is known to *drift* slowly, a fixed filtering approach may have difficulty solving the problem. It has been shown by Widrow *et al.* [46] and Glover [47] that a notch filter realized by an adaptive noise canceler can offer advantages such as easy control of notch bandwidth, an infinite null, and the capability of adaptively tracking the exact frequency and phase of the interference.

In this chapter, we will investigate the performance of the adaptive noise canceler in the suppression of *multiple complex* sinusoidal interferences. Different functions of the adaptive noise canceler, e.g., notch filter, a process decorrelator and a line enhancer, as well as the convergence properties associated with various model orders of both the signal and system will also be addressed.

4.1 Notch Filter Realized by Adaptive Noise Canceler

Figure 4.1(a) shows the block diagram of a dual-input adaptive noise canceler (ANC). The primary input supplies an information-bearing *signal* and an interfering *noise* of multiple sinusoids that are uncorrelated from each other. The reference input consists of a *correlated* version of the sinusoidal interferences. For the adaptive filter, we use a transversal filter whose weights are adapted by means of the LMS algorithm. The reference input and the primary input are given respectively as

$$y(n) = \sum_{i=1}^{p} A_i \exp(j\omega_i n), \tag{4.1}$$

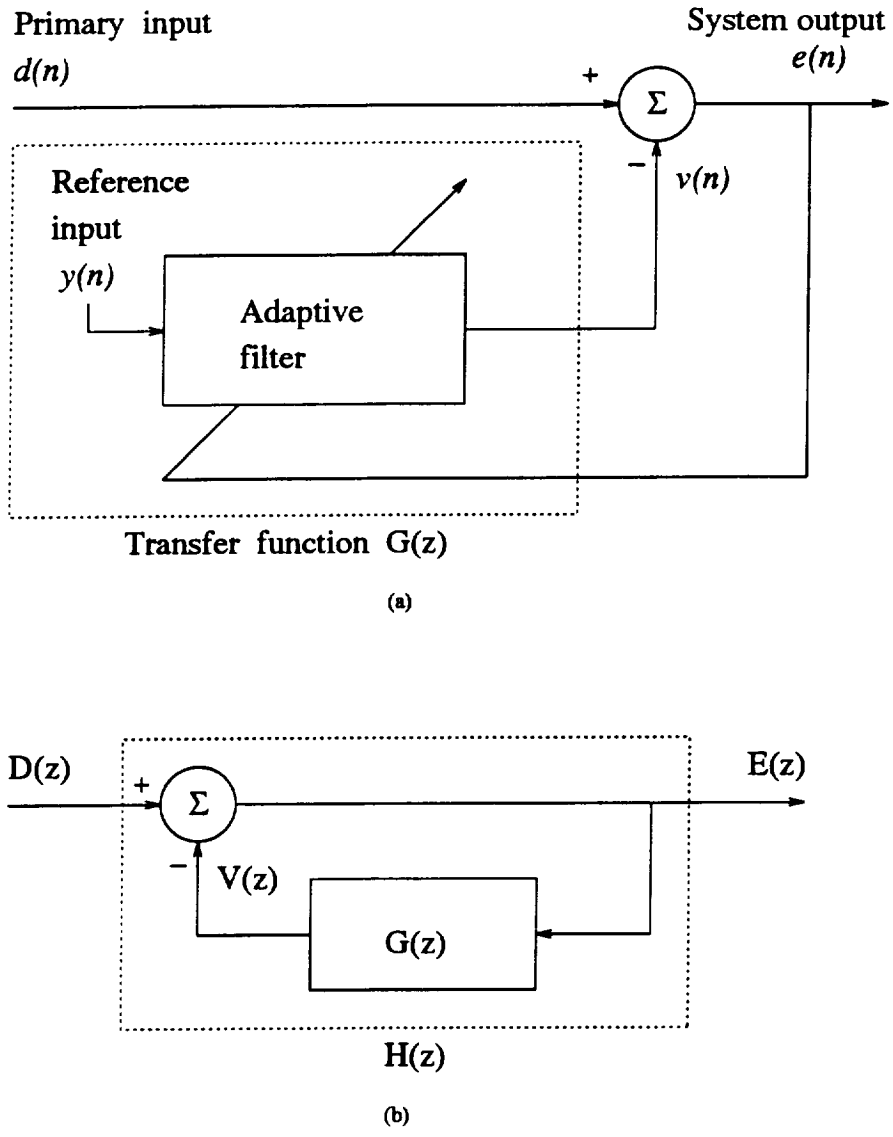


Figure 4.1. (a) Schematic representation of adaptive noise canceler and (b) equivalent model in the z-domain.

$$\text{and } d(n) = s(n) + \sum_{i=1}^{p} B_i \exp(j\omega_i n) \quad n = 0, 1, 2, \cdots \tag{4.2}$$

where $s(n)$ is the information-bearing signal, A_i and B_i are complex amplitudes, and $\omega_i = 2\pi f_i$ $(0 \leq f_i < 1)$.

The filter will use the reference input to provide (at its output) an estimate of the sinusoidal interfering signal contained in the primary input. Thus, by subtracting the adaptive filter output, $v(n)$, from the primary input $d(n)$, the effect of the sinusoidal interference is diminished. According to Section 2.4.2, the LMS algorithm updates the tap-weight as follows:

$$v(n) = \sum_{m=0}^{M-1} w_m(n) y(n-m) \tag{4.3}$$

$$e(n) = d(n) - v(n) \tag{4.4}$$

$$w_m(n+1) = w_m(n) + \mu y^*(n-m) e(n), \quad m = 0, 1, \cdots, M-1 \tag{4.5}$$

where M is the total number of tap weights (order of filter) in the transversal filter, and μ the constant step-size parameter. Let $V(z)$ and $E(z)$ denote the z-transform of the filter output $v(n)$ and the estimation error $e(n)$, respectively. Following the schematic representation in Figure 4.1(a), we may lump the sinusoidal reference input $y(n)$, the transversal filter, and the weight-update equation of the LMS algorithm into an open-loop system defined by a transfer function $G(z) = \frac{V(z)}{E(z)}$, as in the equivalent model of Figure 4.1(b). Our goal is to find $V(z)$, and thus $G(z)$, given $E(z)$. Starting from the weight-updating formula in (4.5), by taking the z-transform of both sides, we get

$$zW_m(z) = W_m(z) + \mu Z\{y^*(n-m)e(n)\}, \quad m = 0, 1, \cdots, M-1 \tag{4.6}$$

where $W_m(z)$ is the z-transform of $w_m(n)$. Substituting (4.1) into (4.6), we find

$$W_m(z) = \frac{\mu z^{-1}}{1 - z^{-1}} \left[\sum_{k=1}^{p} A_k^* e^{j\omega_k m} E(z e^{j\omega_k}) \right] \tag{4.7}$$

where $E(z)$ is the z-transform of $e(n)$, and thus $E(z\, e^{j\omega_i})$ is $E(z)$ rotated clockwise around the unit circle through the angle ω_i. Furthermore, taking the z-transform of

$v(n)$ in (4.3), with $y(n)$ and $W_m(z)$ replaced by (4.1) and (4.3) respectively, it can be shown that

$$\begin{aligned}
V(z) &= Z\left\{\sum_{m=0}^{M-1} w_m(n) \sum_{i=1}^{p} A_i e^{-j\omega_i m} e^{j\omega_i n}\right\} \\
&= \sum_{m=0}^{M-1}\sum_{i=1}^{p}(A_i e^{-j\omega_i m}) Z\left\{w_m(n) e^{j\omega_i n}\right\} \\
&= \sum_{m=0}^{M-1}\sum_{i=1}^{p}(A_i e^{-j\omega_i m}) W_m(z e^{-j\omega_i}) \\
&= \sum_{m=0}^{M-1}\sum_{i=1}^{p}(A_i e^{-j\omega_i m})\left[\frac{\mu z^{-1} e^{j\omega_i}}{1 - z^{-1} e^{j\omega_i}} \sum_{k=1}^{p} A_k^* e^{j\omega_k m} E\left(z e^{j(\omega_k - \omega_i)}\right)\right] \\
&= \underbrace{\sum_{i=1}^{p} \frac{\mu M |A_i|^2 z^{-1} e^{j\omega_i}}{1 - z^{-1} e^{j\omega_i}} E(z)}_{\text{TI component}} \\
&\quad + \underbrace{\sum_{i=1}^{p}\sum_{\substack{k=1 \\ k\neq i}}^{p} \frac{\mu \beta_{ik}(M) A_i A_k^* z^{-1} e^{j\omega_i}}{1 - z^{-1} e^{j\omega_i}} E\left(z e^{j(\omega_k - \omega_i)}\right)}_{\text{TV component}}
\end{aligned} \tag{4.8}$$

where

$$\begin{aligned}
\beta_{ik}(M) &= \sum_{m=0}^{M-1} e^{j(\omega_k - \omega_i)m} = \sum_{m=0}^{M-1} e^{j2\pi(f_k - f_i)m} \\
&= \frac{\sin((f_k - f_i)M\pi)}{\sin((f_k - f_i)\pi)} e^{j(f_k - f_i)(M-1)\pi}, \quad i, k = 1, 2, \cdots, M.
\end{aligned} \tag{4.9}$$

Taking a closer look at the expression for $V(z)$ in (4.8), we consider the first term as a *time-invariant (TI) component*, and the second term as a *time-varying (TV) component* [47]. According to (4.8), the effect of the time-varying component depends on the factor $\beta_{ik}(M)$ defined by (4.9). Particularly, when $\frac{|\beta_{ik}(M)|}{M} \approx 0$, $G(z)$ is determined by retaining only the time-invariant component of $V(z)$, and the ANC behaves like a fixed multiple-notch filter. Letting $\delta_f = \min_{k\neq i} |f_k - f_i|$, the minimum frequency spacing between the interfering sinusoids in the reference input, the time-invariance condition of the ANC can be satisfied by choosing $M > \frac{1}{\delta_f}$. The open-loop transfer function $G(z)$ is therefore

$$G(z) = \frac{V(z)}{E(z)} \approx \sum_{i=1}^{p} \frac{\mu M |A_i|^2 z^{-1} e^{j\omega_i}}{1 - z^{-1} e^{j\omega_i}}, \tag{4.10}$$

and the notch filter realized by Figure 4.1(a) can be recognized as a *closed-loop feedback system* with transfer function

$$H(z) = \frac{E(z)}{D(z)} = \frac{1}{1+G(z)} \approx \frac{1}{1+\sum_{i=1}^{p} \frac{\mu M |A_i|^2 z^{-1} e^{j\omega_i}}{1 - z^{-1} e^{j\omega_i}}}. \tag{4.11}$$

From Equation (4.11), the zeros (notches) of $H(z)$ are at the poles of $G(z)$; that is, they are located on the unit circle at $e^{j\omega_i}$. Furthermore, if a *small* value of the step-size parameter μ is chosen, such that $\mu M |A_i|^2 \ll 1$, $i = 1, 2, \cdots, p$, the poles of $H(z)$ can be approximately located at

$$z_i \approx (1 - \mu M |A_i|^2) e^{j\omega_i}.$$

This fact implies that the poles of $H(z)$ lie inside the unit circle, and thus that the ANC is stable, as it should be for practical use in real time [7].

Following Equation (4.11), Figures 4.2(a) and 4.2(b) give the time-invariant portion of an adaptive notch filter's frequency response and its corresponding pole-zero plot, with $|A_1| = |A_2| = |A_3| = 1$, $[f_1\ f_2\ f_3] = [0.38\ 0.45\ 0.76]$, $M = 32$, and a very small adaptation rate $\mu = 0.0002$ to narrow the notch bandwidths. Due to this choice of $\mu M |A_i|^2$, as can be seen in Figure 4.2(b), poles and zeros of $H(z)$ tend to overlap each other on the unit circle. Furthermore, Figure 4.2(c) provides the result of an experiment performed to characterize the adaptive notch filter's response to a complex unit impulse function as the primary input, i.e., $d(n) = \frac{1}{\sqrt{2}}(1+j)\delta(n)$. In Figures 4.2(a) and 4.2(c), the frequency response of $H(z)$ and the spectrum of $e(n)$ are evaluated at 500 normalized digital frequencies from 0 to 1.

4.2 Adaptive Noise Canceler as Process Decorrelator

In Section 4.1, we assume that the reference signal is *deterministic* while trying to implement a multiple-notch filter via ANC. However, to deal with signals encountered in real world applications, it is more often the *random* nature associated with

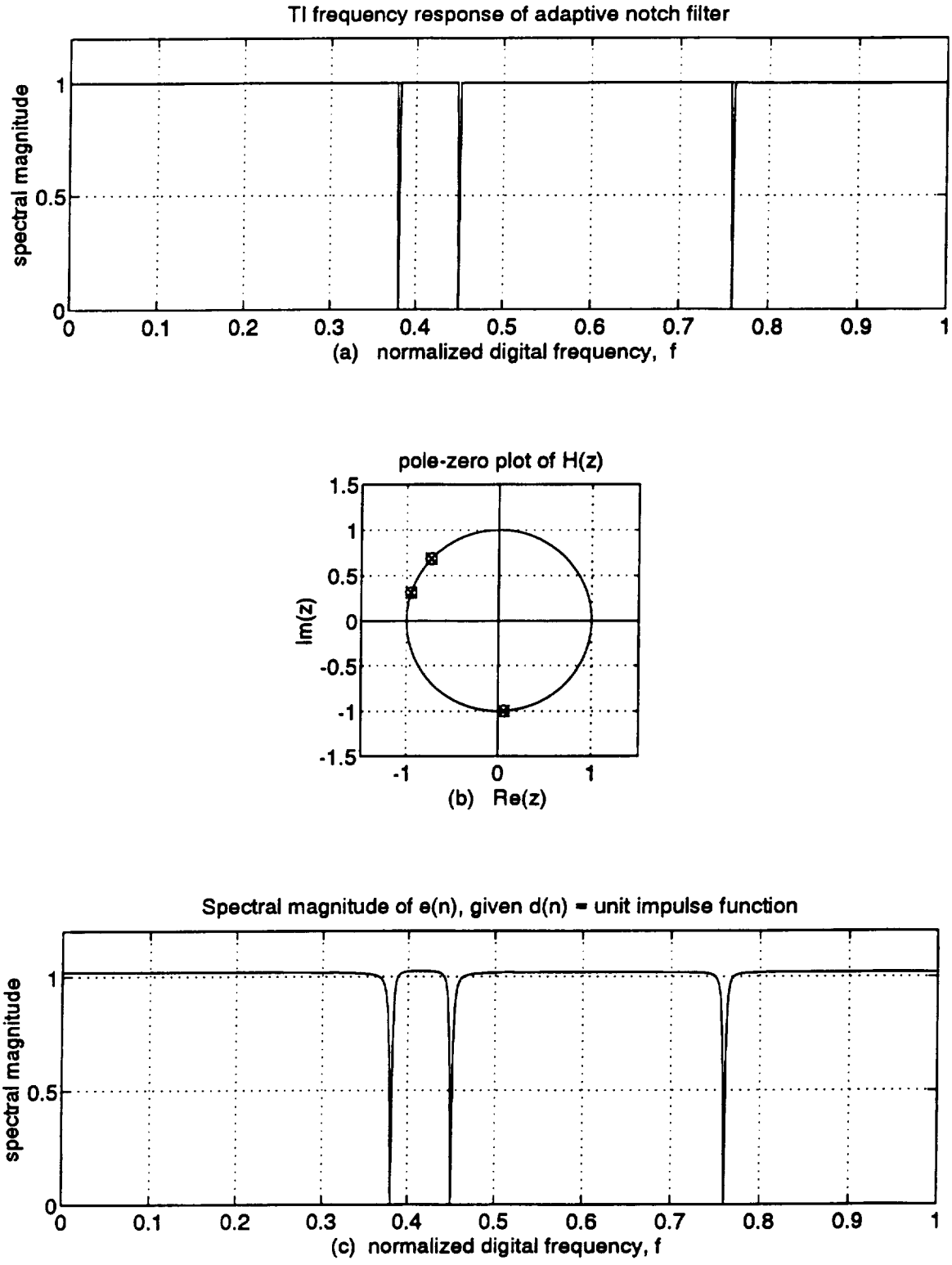


Figure 4.2. (a) Time-invariant frequency response of a notch filter realized by ANC; (b) pole-zero plot of $H(z)$; (c) spectrum of $e(n)$ when $d(n)$ is a unit impulse function.

the signals that necessitates adaptive approaches. Assuming that all the signals considered are realizations of some random processes, and only limited knowledge of the correlation properties among the signals is available, we can use the ANC as a *process decorrelator* to extract the information-bearing signal. Let the reference signal $y(n)$ in Figure 4.1(a) be represented as

$$y(n) = \sum_{i=1}^{p} A_i \exp(j2\pi f_i n + \phi_i) + a(n), \quad n = 0, 1, \cdots, N-1 \tag{4.12}$$

where A_i are constant amplitudes, phases ϕ_i are independent random variables uniformly distributed over $[0\ 2\pi)$, and $a(n)$ is an independent CWGN with zero mean and variance σ_a^2. It has been shown that $y(n)$ is a wide-sense stationary (WSS) process with autocorrelation function (ACF) given as [3]

$$r_{yy}(k) = \sum_{i=1}^{p} A_i^2 \exp(j2\pi f_i k) + \sigma_a^2 \delta(k). \tag{4.13}$$

Denote $\mathbf{S}$ as the diagonal matrix with the power of the ith sinusoid, $S_i = A_i^2$, as its ith diagonal element so that the $M \times M$ autocorrelation matrix for $y(n)$ is

$$\mathbf{R}_{yy} = \sum_{i=1}^{p} S_i \mathbf{e}_i \mathbf{e}_i^H + \sigma_a^2 \mathbf{I} = \mathbf{E}\mathbf{S}\mathbf{E}^H + \sigma_a^2 \mathbf{I} \tag{4.14}$$

where $\mathbf{E} = [\mathbf{e}_1\, \mathbf{e}_2 \cdots \mathbf{e}_p]$ with $\mathbf{e}_i = [1\ e^{j2\pi f_i}\ e^{j2\pi f_i 2} \ \cdots\ e^{j2\pi f_i (M-1)}]^T$. Notice also that $\mathbf{R}_{yy}$ is Hermitian.

Suppose that the primary input $d(n)$ and the reference input $y(n)$ are related as follows:

$$\begin{aligned} d(n) &= s(n) + x(n) \\ x(n) &= \sum_{k=0}^{L-1} h_k y(n-k) = \mathbf{h}^T \mathbf{y}_L(n) \end{aligned} \tag{4.15}$$

where $\mathbf{h} = [h_0\, h_1 \cdots h_{L-1}]^T$ characterizes the linear correlation between $x(n)$ and $y(n)$, $s(n)$ is an information-bearing signal process uncorrelated with $x(n)$, and $\mathbf{y}_L(n) = [y(n)\, y(n-1) \cdots y(n-L+1)]^T$. We call L the *order of correlation* between the

primary input and the reference signal. Letting $\mathbf{P} = E[\mathbf{y}_M(n)d^*(n)]$, the $M \times 1$ cross-correlation vector between the tap input vector and the primary input, and based upon the above assumptions, it is straightforward to show that

$$\mathbf{P} = \mathbf{R}_{ML}\mathbf{h}^*$$

where $\mathbf{R}_{ML} = E[\mathbf{y}_M(n)\mathbf{y}_L^H(n)]$. Therefore, assuming that the autocorrelation matrix $\mathbf{R}_{yy}$ is nonsingular, the optimum Wiener solution described by (2.15) can be obtained as

$$\mathbf{w}_o^* = \mathbf{R}_{yy}^{-1}\mathbf{R}_{ML}\mathbf{h}^*. \tag{4.16}$$

Up to this point, it is quite obvious that when $M = L$, $\mathbf{R}_{yy} = \mathbf{R}_{ML}$, and thus $\mathbf{w}_o = \mathbf{h}$.

To show the performance of the ANC functioning as a process decorrelator, we undertake another experiment of 1000 Monte Carlo simulations. The reference signal is generated according to (4.12) with $[f_1\ f_2\ f_3] = [0.38\ 0.56\ 0.75]$ and 10 dB of sinusoids-to-noise power ratio. For each realization of $y(n)$, $d(n)$ is correspondingly generated by (4.15), where $s(n)$ is replaced by a white Gaussian noise $z(n)$ with power 30 dB below that of $x(n)$. We choose the correlation vector $\mathbf{h}$ as

$$\begin{aligned}\mathbf{h} &= [e^{(-0.2+j0.6048\pi)0}\ \ e^{(-0.2+j0.6048\pi)1}\ \ e^{(-0.2+j0.6048\pi)2}]^T \\ &= [1.0000 + j0 \quad -0.2647 + j0.7748 \quad -0.5302 - j0.4102]^T,\end{aligned}$$

and assume that the number of filter taps is equal to the order of correlation, i.e., $M = L = 3$. If $z(n)$ is taken as the "plant noise", we are equivalently facing a system identification problem as depicted in Figure 2.2.

Figure 4.3 demonstrates the ensemble averages of both the filter's tap weight vectors and the power of the estimation error $e(n)$. At the end of all simulations, we obtain the filter's tap weight vector estimate

$$\hat{\mathbf{w}}_o = [0.9979 - j0.0005 \quad -0.2679 + j0.7746 \quad -0.5324 - j0.4096]^T,$$

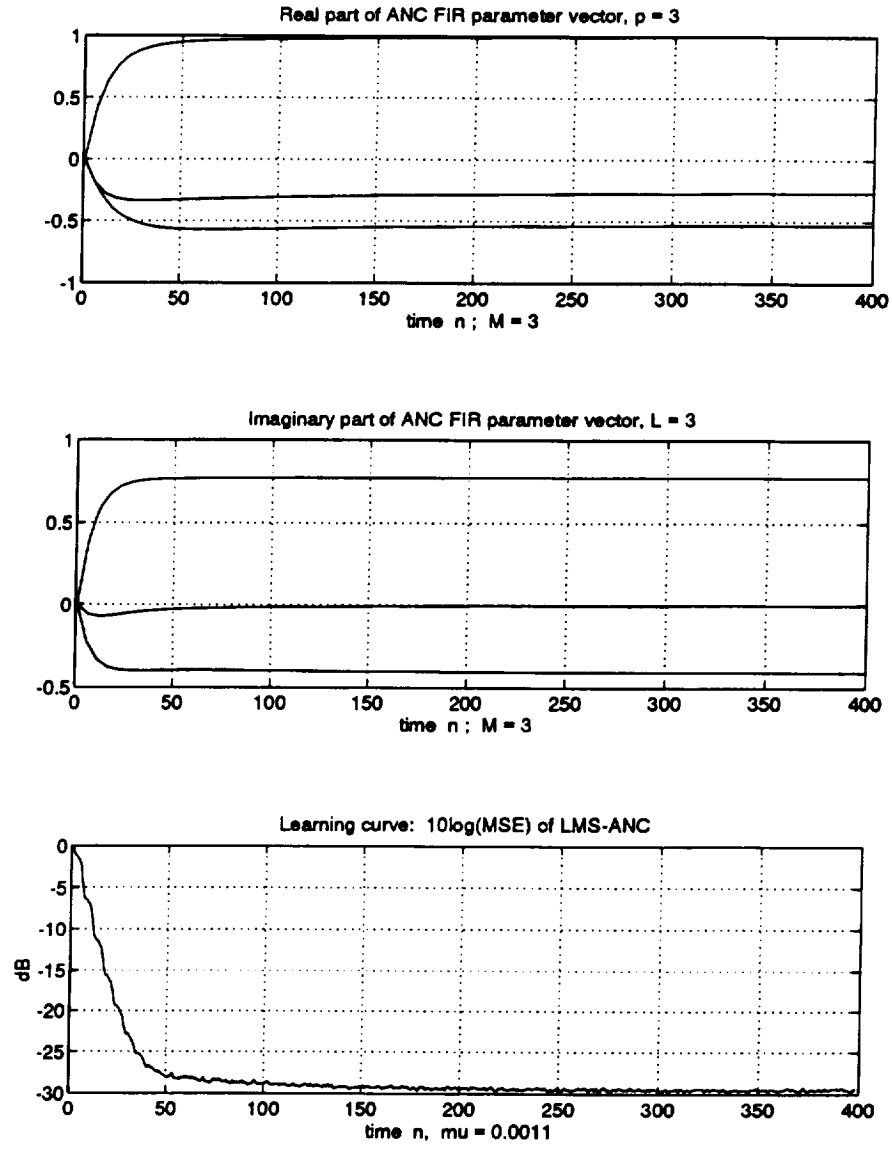


Figure 4.3. Process decorrelator via ANC, $M = L = 3$, $\mu = 0.0011$: evolution of tap weight vector: $\mathrm{Re}\{\mathbf{w}(n)\}$ and $\mathrm{Im}\{\mathbf{w}(n)\}$; learning curve: $10\ \log(J(n))$ vs. time.

which is almost equal to $\mathbf{w}_o = \mathbf{h}$, the Wiener solution of this case. Under such an optimum condition, only the plant noise is expected to be left at the process decorrelator's output, i.e., $e(n) \approx z(n)$. In other words, by examining the "learning curve" at the bottom of Figure 4.3, one can find that the mean-squared error sequence, $E[|e(n)|^2]$, converges to about -30 dB, a phenomenon in agreement with the assumption made for the primary input signal.

4.3 Adaptive Filter Design Considerations

In Section 4.2, we assume that the reference signal $y(n)$ and its linearly correlated signal $x(n)$ are random-phased sinusoids mixed with white Gaussian noise, as modeled by (4.12) and (4.15). Due to the presence of the noise's autocorrelation matrix, $\sigma_w^2\mathbf{I}$, at the right-hand side of (4.14), $\mathbf{R}_{yy}$ is of full rank and nonsingular. These facts justify the existence of a nontrivial solution to the Wiener-Hopf equation, as expressed by (4.16): $\mathbf{w}_o^* = \mathbf{R}_{yy}^{-1}\mathbf{R}_{ML}\mathbf{h}^*$. Moreover, an adaptive noise canceler is expected to suppress the unwanted interference $x(n)$ by adaptively tracking its tap weight vector $\mathbf{w}(n)$ to this optimum $\mathbf{w}_o$. As pointed out before, when $M = L$, we have $\mathbf{w}_o = \mathbf{h}$. However, in most applications, the order of correlation L is unknown to the signal processors. In this section, we will discuss the issue regarding the choices of two important design parameters: the filter's number of taps, M, and the LMS (NLMS) algorithm's adaptation rate μ.

4.3.1 Choice of Filter's Number of Taps M

According to Orfanidis [48], the rule of choosing M with respect to L is that the adaptive filter must have at least as many delays as that part of $d(n)$ which is correlated with $y(n)$, in other words, $M \geq L$. Taking a closer look at Equation (4.16) can help us see why it is so. First, consider the case of *overmodeling*, i.e., $M > L$, then partition $\mathbf{R}_{yy}$ and its inverse matrix $\mathbf{R}_{yy}^{-1}$ as

$$\mathbf{R}_{yy} = [\mathbf{R}_{ML} \;\; \mathbf{B}] \quad \text{and} \quad \mathbf{R}_{yy}^{-1} = \begin{bmatrix} \mathbf{R}_{iu} \\ \mathbf{R}_{il} \end{bmatrix}$$

where $\mathbf{B}$ is an $M \times (M-L)$ matrix, $\mathbf{R}_{iu}$ an $L \times M$ matrix, and $\mathbf{R}_{il}$ an $(M-L) \times M$ matrix. Since $\mathbf{R}_{yy}^{-1}\mathbf{R}_{yy} = \mathbf{I}$, it is straightforward to show that $\mathbf{R}_{iu}\mathbf{R}_{ML} = \mathbf{I}$ and $\mathbf{R}_{il}\mathbf{R}_{ML} = \mathbf{0}$, thus the Wiener solution can be found as

$$\mathbf{w}_o^* = \mathbf{R}_{yy}^{-1}\mathbf{R}_{ML}\mathbf{h}^* = \begin{bmatrix} \mathbf{I}_{L\times L} \\ \mathbf{0}_{(M-L)\times L} \end{bmatrix} \mathbf{h}^* = \begin{bmatrix} \mathbf{h}^* \\ \mathbf{0} \end{bmatrix} \tag{4.17}$$

where $\mathbf{0}$ is the $(M-L) \times 1$ zero vector. Therefore, according to (4.3) ,(4.4), (4.15), and (4.17), if $\mathbf{w}(n) = \mathbf{w}_o = [h_0\ h_1 \cdots h_{L-1}\ 0 \cdots 0]^T$,

$$e(n) = d(n) - \mathbf{w}_o^T\mathbf{y}(n) = d(n) - \sum_{k=0}^{L-1} h_k y(n-k) = d(n) - x(n) = s(n).$$

This implies the complete cancelation of the $y(n)$-dependent part of $d(n)$.

When the adaptive filter is *undermodeling* the correlation between $y(n)$ and $d(n)$, i.e., $M < L$, following (4.15) we can express $d(n)$ as

$$d(n) = \mathbf{h}_1^T\mathbf{y}_1(n) + \mathbf{h}_2^T\mathbf{y}_2(n) + s(n)$$

where $\mathbf{h}_1 = [h_0\, h_1 \cdots h_{M-1}]^T$, $\mathbf{h}_2 = [h_M\, h_{M+1} \cdots h_{L-1}]^T$, $\mathbf{y}_1(n) = [y(n)\ y(n-1) \cdots y(n-M+1)]^T$, and $\mathbf{y}_2(n) = [y(n-M)\ y(n-M-1) \cdots y(n-L+1)]^T$. Furthermore, it can be shown that

$$\mathbf{w}_o^* = \mathbf{R}_{yy}^{-1}\mathbf{R}_{ML}\mathbf{h}^* = \mathbf{R}_{11}^{-1}[\mathbf{R}_{11}\ \mathbf{R}_{12}] \begin{bmatrix} \mathbf{h}_1^* \\ \mathbf{h}_2^* \end{bmatrix} = \mathbf{h}_1^* + \mathbf{R}_{11}^{-1}\mathbf{R}_{12}\mathbf{h}_2^*, \tag{4.18}$$

and the optimum estimate of $d(n)$ given $\mathbf{y}_1(n)$ is

$$\hat{d}(n) = E[d(n)|\mathbf{y}_1(n)] = E[d(n)\mathbf{y}_1^H(n)]\mathbf{R}_{11}^{-1}\mathbf{y}_1(n) = \mathbf{w}_o^T\mathbf{y}_1(n)$$

where $\mathbf{R}_{yy} = \mathbf{R}_{11} = E[\mathbf{y}_1(n)\mathbf{y}_1^H]$ and $\mathbf{R}_{12} = E[\mathbf{y}_1(n)\mathbf{y}_2^H]$. More specifically, as $\mathbf{w}(n)$ converges to $\mathbf{w}_o$,

$$\begin{aligned} \hat{d}(n) &= v(n) = \mathbf{w}_o^T\mathbf{y}_1(n) \\ &= \mathbf{h}_1^T\mathbf{y}_1(n) + \mathbf{h}_2^T\mathbf{R}_{12}^H(\mathbf{R}_{11}^{-1})^H\mathbf{y}_1(n) = \mathbf{h}_1^T\mathbf{y}_1(n) + \mathbf{h}_2^T\mathbf{R}_{21}\mathbf{R}_{11}^{-1}\mathbf{y}_1(n), \end{aligned}$$

and thus the estimated information-bearing signal

$$e(n) = d(n) - \hat{d}(n) = \mathbf{h}_2^T[\mathbf{y}_2(n) - \hat{\mathbf{y}}_{2/1}(n)] + s(n)$$

where $\hat{\mathbf{y}}_{2/1}(n) = E[\mathbf{y}_2(n)\mathbf{y}_1^H(n)]E[\mathbf{y}_1(n)\mathbf{y}_1^H(n)]^{-1}\mathbf{y}_1(n) = \mathbf{R}_{21}\mathbf{R}_{11}^{-1}\mathbf{y}_1(n)$ is the optimum estimate of $\mathbf{y}_2(n)$ based on $\mathbf{y}_1(n)$. This analysis shows that the $\mathbf{y}_1(n)$ part is completely removed from the primary input, and the $\mathbf{y}_2(n)$ part is suppressed as much as possible.

4.3.2 Choice of Adaptation Rate μ and Signal Statistics

In designing adaptive filters via the LMS algorithm, a problem of many concerns is the convergence behavior in tracking the optimum Wiener solution, where the adaptation rate μ plays a very critical role. Before delving into this issue, we need to know that the LMS algorithm is an example of a *multivariable nonlinear stochastic feedback system*, and such combined presence of nonlinearity and randomness makes its convergence (stability) analysis a difficult mathematical task [7, 17]. To alleviate the mathematical intractability in the convergence analysis of the LMS algorithm, a set of *fundamental assumptions* needs to be followed:

1. The tap-input vectors $\mathbf{y}(1), \mathbf{y}(2), \cdots, \mathbf{y}(n)$ are statistically independent.
2. At time n, the tap-input vector $\mathbf{y}(n)$ and the desired response $d(n)$ are statistically dependent, but independent of their *previous* counterparts.
3. $\mathbf{y}(n)$ and $d(n)$ are *jointly Gaussian-distributed* random variables for all n.

The statistical analysis of the LMS algorithm based on the fundamental assumptions is called the *independence theory.* Please refer to Gardner [49] and Haykin [7] for a detail account. Our discussion regarding the choice of μ and its corresponding mean-squared error $J(n) = E[|e(n)|^2]$ will rely on the independence theory.

Letting λ_i, $i = 1, 2, \cdots, M$, denote the eigenvalues of the correlation matrix $\mathbf{R}_{yy}$, the mean-squared error $J(n)$ converges to a steady-state value $J(\infty)$ if, and only if, the adaptation rate μ satisfies

$$0 < \mu < \frac{2}{\lambda_{\max}} \tag{4.19}$$

and

$$\sum_{i=1}^{M} \frac{\mu\lambda_i}{2 - \mu\lambda_i} < 1 \tag{4.20}$$

where M is the number of filter taps, and $\lambda_{\max}$ is the largest eigenvalue of $\mathbf{R}_{yy}$. Since $\mathbf{R}_{yy}$ is usually unavailable to the ANC designers, when μ is small compared to $\frac{2}{\lambda_{\max}}$, (4.20) can be simplified to a rule of thumb as

$$0 < \mu < \frac{2}{\text{total input power}} \tag{4.21}$$

where the *total input power* is an estimate of $E[\mathbf{y}^H(n)\mathbf{y}(n)] = \text{tr}[\mathbf{R}_{yy}]$. When these conditions are satisfied, the LMS algorithm is expected to *converge in the mean-square* sense. Furthermore, define the *minimum mean-squared error* $J_{\min}$ as the MSE produced by the optimum Wiener filter, i.e., $J_{\min} = E[|d(n) - \mathbf{w}_o^T\mathbf{y}(n)|^2]$. According to the independence theory, the mean-squared error produced by the LMS algorithm has the final value

$$J(\infty) = \frac{J_{\min}}{1 - \sum_{i=1}^{M} \frac{\mu\lambda_i}{2 - \mu\lambda_i}}, \tag{4.22}$$

which is always in excess of the minimum value $J_{\min}$ due to the variance of $\mathbf{w}(n)$ with respect to the optimum Wiener solution $\mathbf{w}_o$. A quantitative measure of this *cost of adaptability* is the *misadjustment* $\mathcal{M}$, defined as

$$\mathcal{M} \triangleq \frac{J(\infty) - J_{\min}}{J_{\min}} = \frac{\sum_{i=1}^{M} \mu\lambda_i/(2 - \mu\lambda_i)}{1 - \sum_{i=1}^{M} \mu\lambda_i/(2 - \mu\lambda_i)}. \tag{4.23}$$

In Section 4.3.1 we see the effect of choosing M in general Wiener filtering. So far as the LMS algorithm is concerned, Equations (4.20) and (4.23) show that three principal factors affect the convergence behavior of this algorithm: the step-size parameter μ, the number of filter taps M, and the eigenvalue distribution of the correlation matrix $\mathbf{R}_{yy}$. The *condition number* of $\mathbf{R}_{yy}$, defined as

$$\chi(\mathbf{R}_{yy}) = \frac{\lambda_{\max}}{\lambda_{\min}},$$

is a good indicator of the signal statistics. This ratio is commonly referred to as the *eigenvalue spread (ES)* or the *eigenvalue disparity*, a factor that controls the LMS

algorithm's convergence speed [7, 48]. A large eigenvalue spread in the correlation matrix $\mathbf{R}_{yy}$ corresponds to a *highly self-correlated* signal $y(n)$. While a reference signal $y(n)$ of such nature tends to slow down the LMS algorithm's convergence process, the tracking dynamics of the NLMS algorithm appear to be significantly less sensitive to a variety of input signal distribution aspects than holds for its precedent [28, 50]. This point can be partially validated by the principle that the NLMS algorithm converges if and only if

$$0 < \tilde{\mu} < 2 \tag{4.24}$$

which is a condition on the step size parameter that is independent of the signal statistics. The fastest convergence occurs for

$$\tilde{\mu} = 1,$$

corresponding to the projection interpretation discussed in [27]. In next section, we will demonstrate the effect of these design parameters, and the performance comparison for the LMS and NLMS algorithms through simulation.

4.3.3 Simulation Results

Clearly, there are many practical problems for which the reference input process and the desired response do not satisfy the fundamental assumptions. An example will be the signal model assumed in Section 4.2 for the process decorrelator. Nevertheless, experience with the LMS algorithm has shown that the independence theory retains sufficient information about the structure of the adaptive process for the results of the theory to serve as reliable design guidelines, even for some problems having highly dependent data samples [7, 51].

To demonstrate the effects of signal statistics, correlation modeling, and the choice of μ, we perform three experiments of 1000 Monte Carlo simulations. In each experiment, the reference input and the primary input are generated according to (4.12) and (4.15) with $L = 4$, $N = 400$, and $s(n)$ being white Gaussian noise of power

level 30 dB below that of the sinusoidal interference. The sinusoid-to-noise ratio at the reference input is 5 dB. Performances of both the LMS and NLMS algorithms are examined in every simulation. Moreover, we choose the filter's number of taps as $M = 3$, $M = 4$, and $M = 5$ for *undermodeling, correct-modeling*, and *overmodeling* the correlation between both inputs, respectively. To expedite the performance comparison, step size $\tilde{\mu}$ of the NLMS algorithm is fixed at 1 for all cases.

In the first experiment, we have $\mathbf{h} = [h_0 \ h_1 \ h_2 \ h_3]^T$ with $h_k = e^{-ak}$ where $a = 0.2 - j0.6048\pi$, and three interfering sinusoids at frequencies $[f_1 \ f_2 \ f_3] = [0.23 \ 0.45 \ 0.78]$. Figure 4.4 shows the LMS algorithm's convergence behavior in terms of its time evolution of filter's weight vectors and output mean-squared errors. Notice that in the undermodeling case ($M < L$), $\mathbf{w}(n)$ very quickly converges to the Wiener solution, but $J(n)$ converges to a steady level of -10 dB due to the reason addressed in Section 4.3.1 and the inherent misadjustment associated with the LMS algorithm. However, as M increases, the convergences of $\mathbf{w}(n)$ and $J(n)$ are significantly slowed down. When $M \geq L$, the observed $J(\infty)$ is about -28.55 dB, very close to -28.66 dB as predicted by (4.22). Similar results for the NLMS algorithm are given in Figure 4.5.

The second and third experiments are devised to compare the convergence performances of the LMS and NLMS algorithms. The test scenarios for both experiments are very similar to the first one except that $h_k = 0.5^k$ with $k = 0, 1, 2, 3$, $[f_1 \ f_2 \ f_3] = [0.38 \ 0.56 \ 0.75]$ for the second one, $[f_1 \ f_2 \ f_3 \ f_4] = [0.38 \ 0.56 \ 0.75 \ 0.57]$ for the third one, and each one has two choices of μ (LMS) for comparison: $\mu_1 = \frac{0.4}{\mathrm{tr}[\mathbf{R}_{yy}]}$ and $\mu_2 = \frac{0.5}{\mathrm{tr}[\mathbf{R}_{yy}]}$, where the total input power $\mathrm{tr}[\mathbf{R}_{yy}]$ is estimated assuming $M = L$. Notice that in the third experiment, two spectrally close sinusoids ($f_2 = 0.56$ and $f_4 = 0.57$) are included in the reference input to create larger eigenvalue disparity than that of the second experiment. Results of these two experiments are given in Figures 4.6 and 4.7, where we can see that the difference between both algorithms'

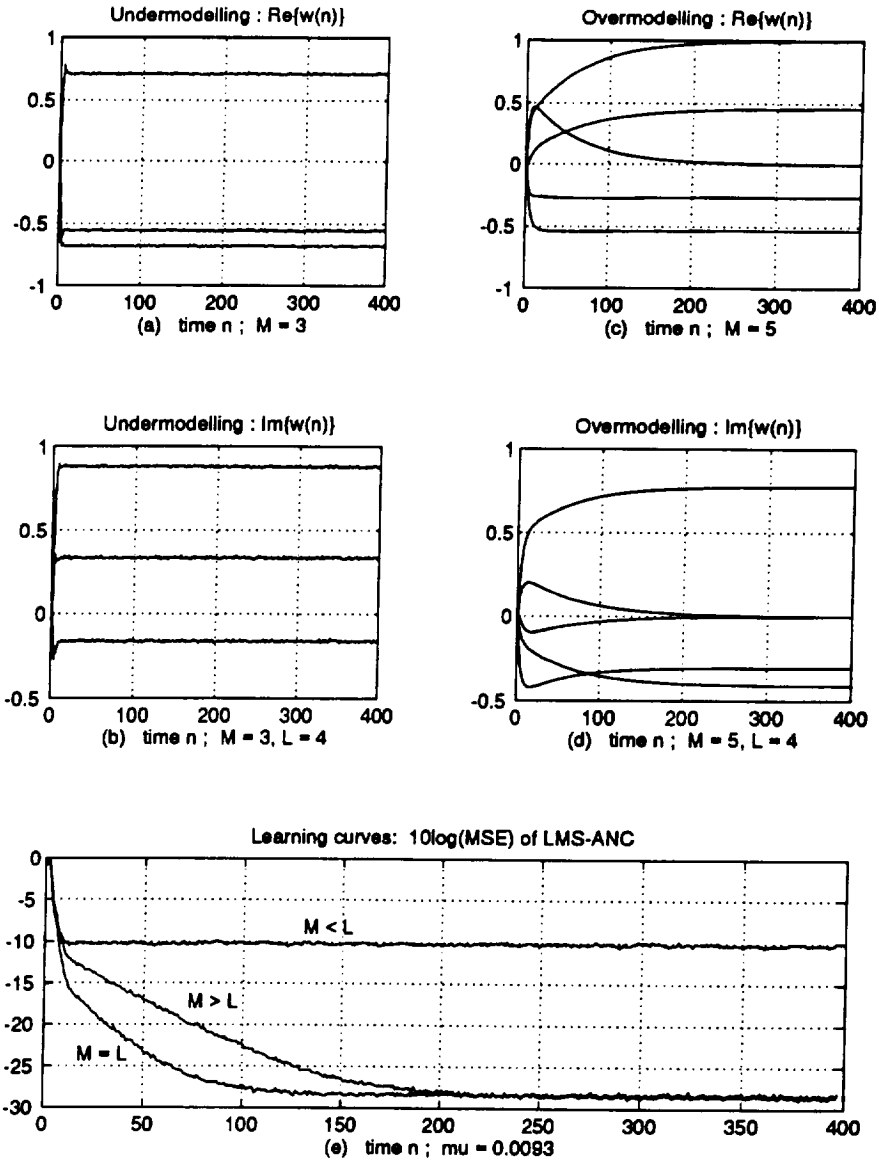


Figure 4.4. LMS-ANC convergence behavior: $\mu = 0.0093$, $L = 4$; undermodeling $M = 3 < L$, $ES = 2.98$: (a) $\mathrm{Re}\{\mathbf{w}(n)\}$, (b) $\mathrm{Im}\{\mathbf{w}(n)\}$; overmodeling $M = 5 > L$, $ES = 21.60$: (c) $\mathrm{Re}\{\mathbf{w}(n)\}$, (d) $\mathrm{Im}\{\mathbf{w}(n)\}$; (e) learning curves: $10\log(J(n))$ vs. time ($ES = 16.75$ when $M = L$).

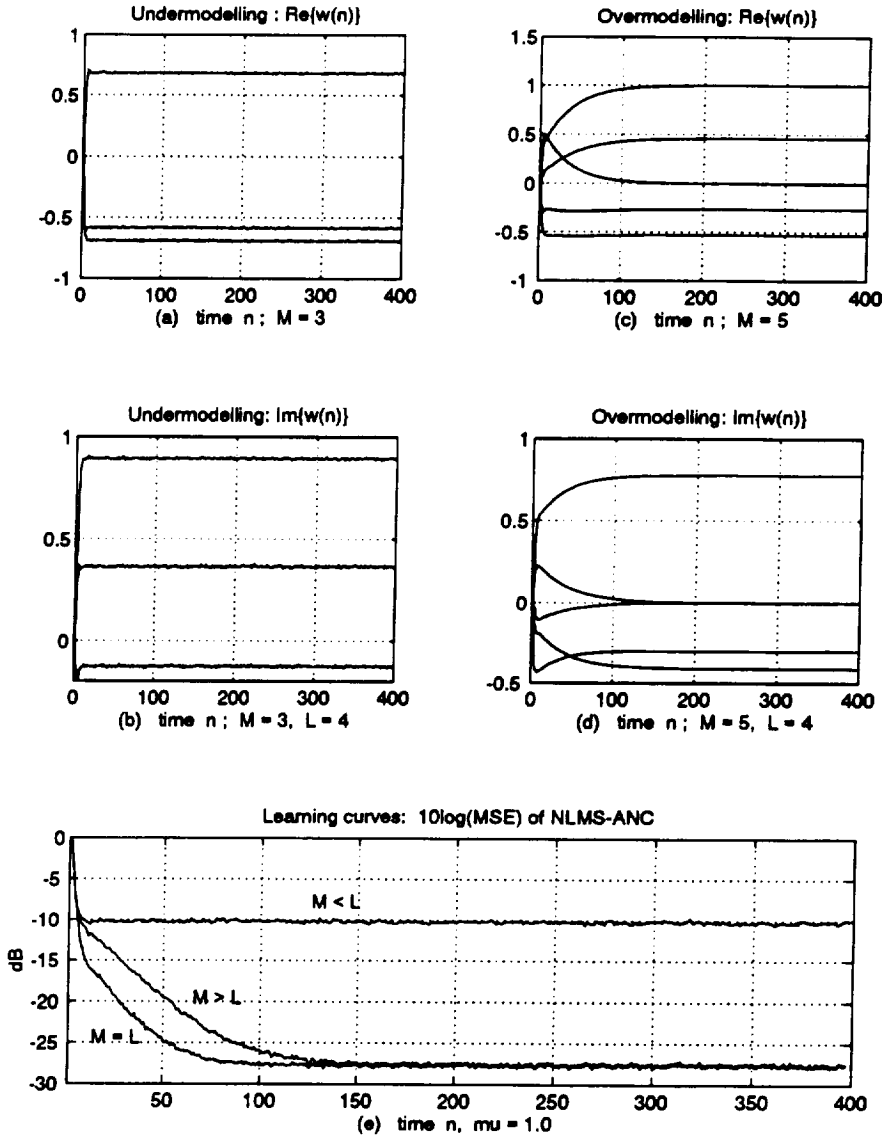


Figure 4.5. NLMS-ANC convergence behavior: $\tilde{\mu} = 1.0$, $L = 4$; undermodeling $M = 3 < L$, $ES = 2.98$: (a) $\mathrm{Re}\{\mathbf{w}(n)\}$, (b) $\mathrm{Im}\{\mathbf{w}(n)\}$; overmodeling $M = 5 > L$, $ES = 21.60$: (c) $\mathrm{Re}\{\mathbf{w}(n)\}$, (d) $\mathrm{Im}\{\mathbf{w}(n)\}$; (e) learning curves: $10\log(J(n))$ vs. time ($ES = 16.75$ when $M = L$).

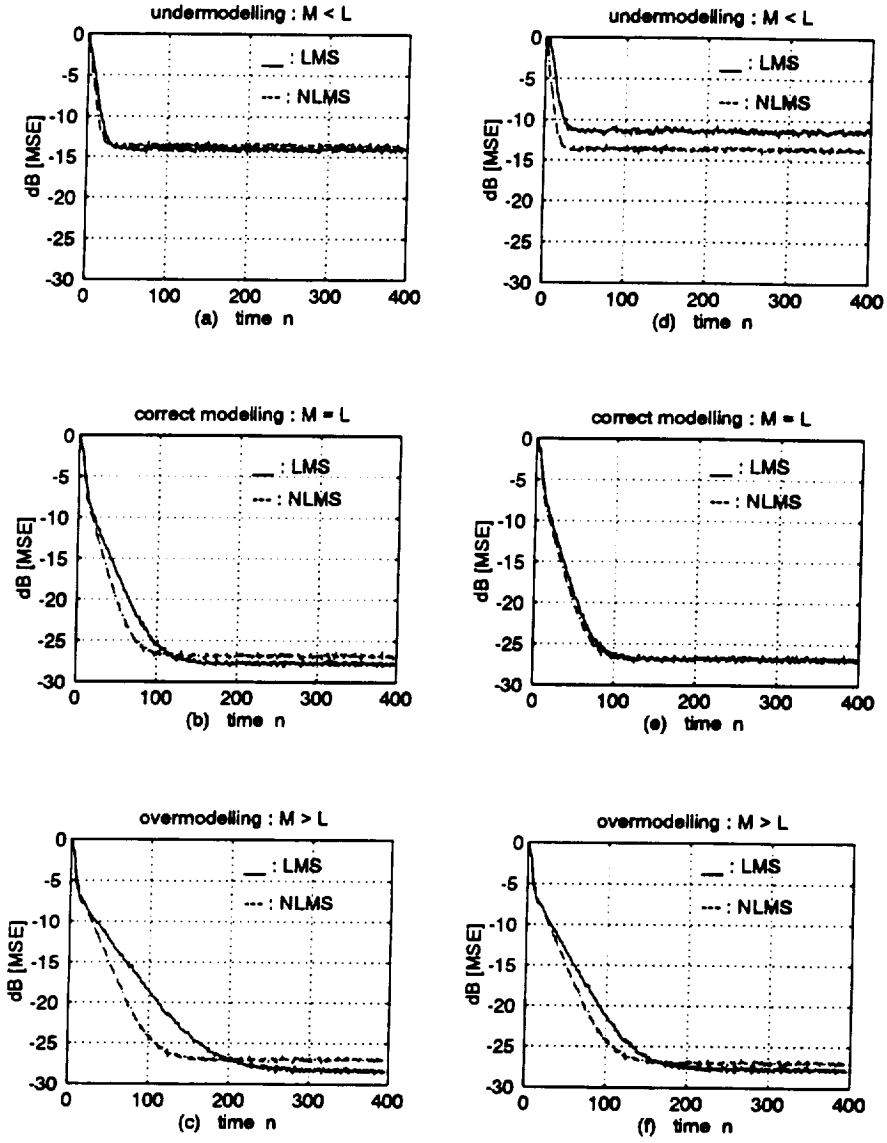


Figure 4.6. Performance comparison for LMS ((a),(b),(c): $\mu_1 = 0.0095$; (d),(e) ,(f): $\mu_2 = 0.0119$) and NLMS ($\tilde{\mu} = 1.0$), $L = 4$, reference input $SNR = 5$ dB, $[f_1\ f_2\ f_3] = [0.38\ 0.56\ 0.75]$; (a),(d): $M = 3$, $ES = 9.67$; (b),(e): $M = 4$, $ES = 18.25$; (c),(f): $M = 5$, $ES = 18.82$.

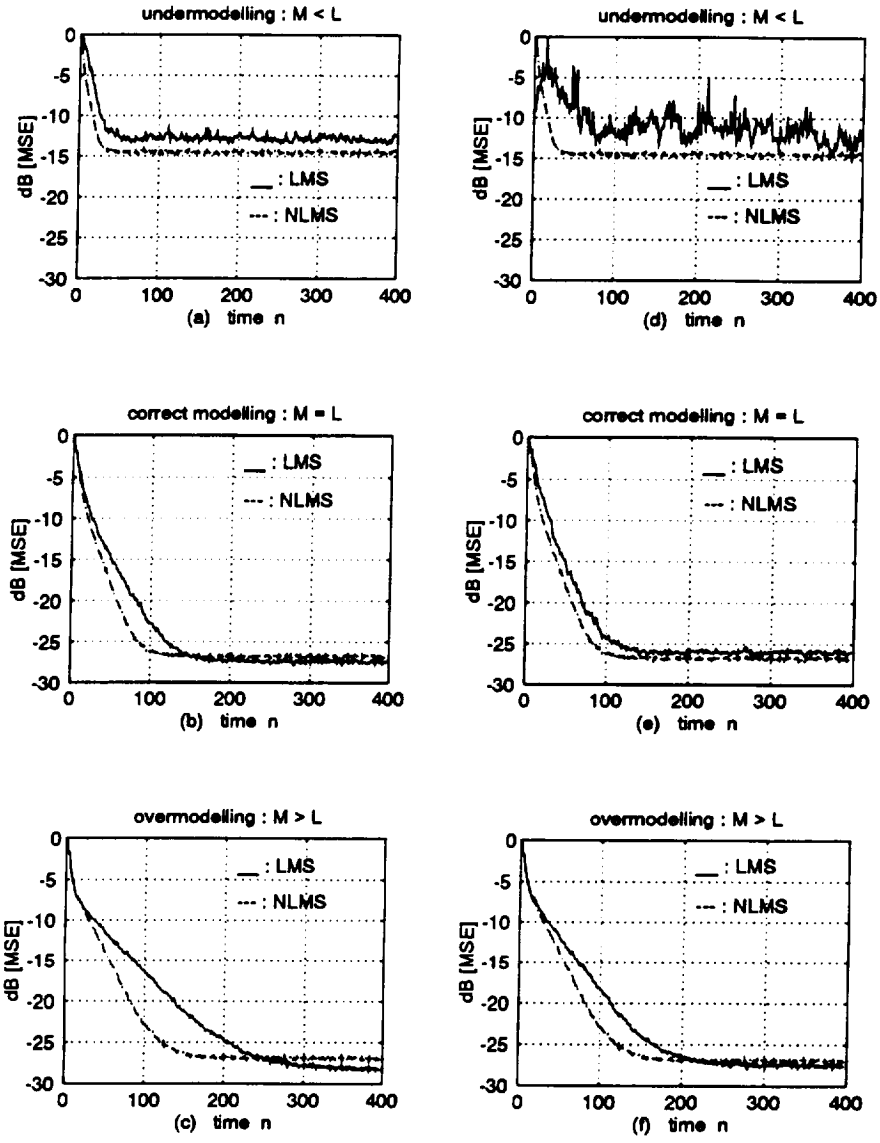


Figure 4.7. Performance comparison for LMS ((a),(b),(c): $\mu_1 = 0.0073$; (d),(e),(f): $\mu_2 = 0.0092$) and NLMS ($\tilde{\mu} = 1.0$), $L = 4$, reference input $SNR = 5$ dB, $[f_1 \ f_2 \ f_3 \ f_4] = [0.38 \ 0.56 \ 0.75 \ 0.57]$; (a),(d): $M = 3$, $ES = 12.88$; (b),(e): $M = 4$, $ES = 29.83$; (c),(f): $M = 5$, $ES = 32.97$.

convergence speeds increases as the eigenvalue spread of $\mathbf{R}_{yy}$ increases. This indicates that the NLMS algorithm's convergence response is less susceptible to the signal's eigenvalue disparity.

On the other hand, though the LMS algorithm shows slower convergence throughout the simulations, it is able to settle at a smaller value of $J(\infty)$, especially in the $M \geq L$ cases. However, increasing μ to boost up the LMS algorithm's convergence rate tends to induce larger misadjustment and possible divergence of $J(n)$ sequence when the adaptive filter is undermodeling the correlation between both filter inputs. Such phenomena can be observed in Figures 4.6(a), 4.7(a), 4.7(e), and 4.7(d). Based on these simulation results, with the input signal modeled as a stationary sinusoidal process as discussed in Section 4.2, the NLMS algorithm (with $\tilde{\mu} = 1$) generally outperforms the LMS algorithm in terms of a better tradeoff between convergence speed and misadjustment, and stability under a more ill-conditioned signal environment.

4.4 Adaptive Noise Canceler as Line Enhancer

As we have addressed before, adaptive noise cancelation requires the presence of a reference signal highly correlated with the noise component interfering with the signal of interest at the primary input. However, there are several circumstances where only *one* noise-contaminated signal $d(n)$ is available. This problem occurs particularly when a broadband signal $s(n)$ is corrupted by a sinusoidal interference $x(n)$, and no external reference input free of the signal is available. In such a case, the signal $d(n)$ provides its own reference signal $y(n)$, which is taken to be a *delayed* replica of $d(n)$, i.e., $y(n) = d(n - \Delta)$.

Figure 4.8 shows the block diagram of an *adaptive line enhancer* (ALE) configured via the ANC. Suppose the signal $d(n)$ consists of two signal components: a *narrowband* component $x(n)$ according to (4.12) that has *long-range* correlations, and a *broadband* component $s(n)$ which tends to have *short-range* correlations. To design an ALE as depicted in Figure 4.8, the delay Δ is usually selected so that

$$E[s(n)s^*(n-k)] \approx 0, \quad k \geq \Delta.$$

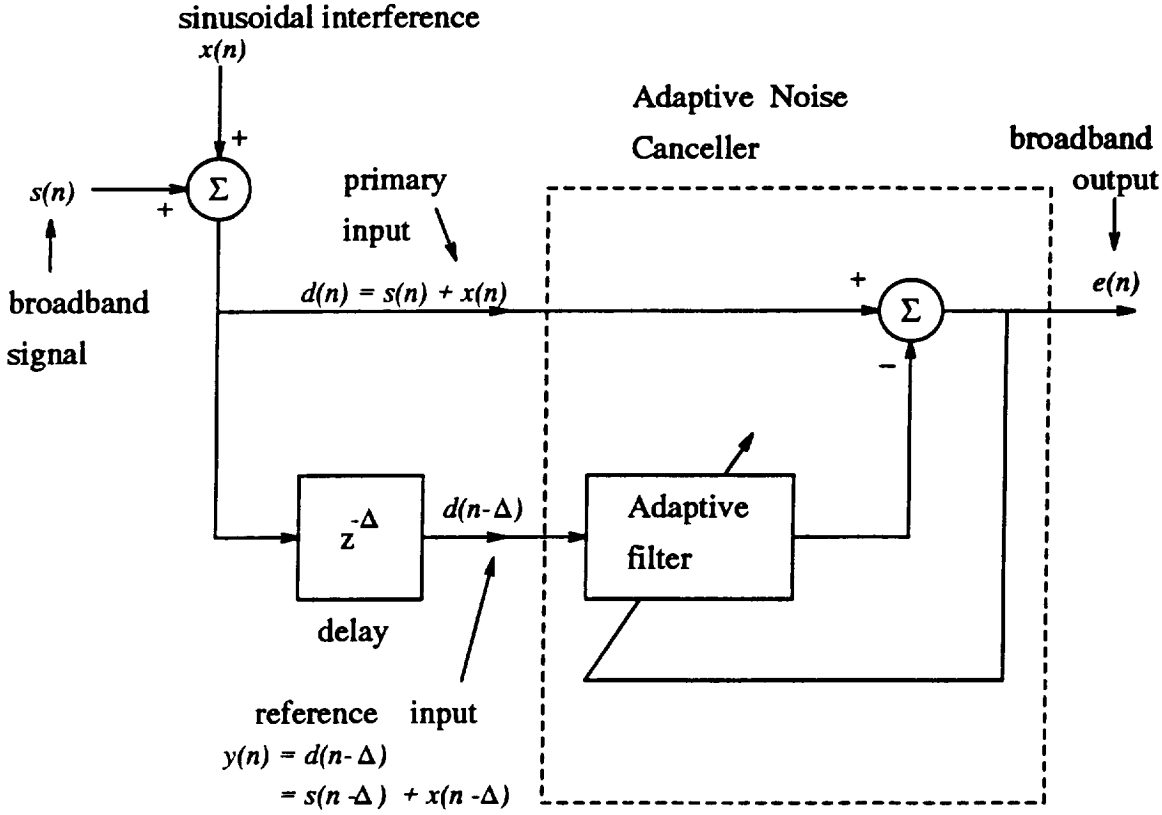


Figure 4.8. Line enhancer via adaptive noise canceler: to enhance the broadband input by suppressing the narrowband sinusoidal interference.

Since Δ is larger than the effective correlation length of the wideband component, the delayed replica $s(n-\Delta)$ will be uncorrelated with $s(n)$ in the primary signal, and thus the adaptive filter will not be able to respond to this component. On the other hand, since the autocorrelation of $x(n)$ does not taper off, the delayed replica $x(n-\Delta)$ that appears in the reference input will still be correlated with the narrowband part of the primary signal, and the filter will respond to cancel it.

From our discussion and simulation result in previous sections, the tap weight vector in the ALE using the LMS algorithm will approximately converge to the Wiener solution when the input signal consists of random-phased sinusoids plus white noise. However, the presence of a wideband signal can complicate the statistical characteristics of the input autocorrelation matrix $\mathbf{R}_{xx}$. In next chapter, we will investigate the performances of LMS-ANC (ALE) in Doppler weather radar applications, where the signal environment is very similar to what we propose here.

CHAPTER 5

APPLICATION TO WINDSHEAR DETECTION PROBLEM

5.1 Weather Detection via Airborne Doppler Radar

Pulse Doppler radar is used for remote detection of windshears. With appropriate signal and data processing, this radar can map wind velocity vs. range over an antenna scan sector, enabling it to look into storm systems that may contain hazardous windshears. A form of low altitude windshears known as a microburst has been identified as a particular hazard to aircrafts during takeoff and landing [12]. A microburst can be qualitatively considered as either wet or dry, dependent on its radar reflectivity [52]. Dry microbursts will call for more sophisticated signal processing, since the primary sources of reflections, particles of dust and insects, exhibit a much lower reflectivity of radar energy. Furthermore, when an aircraft is in low altitude flight, such as during takeoff or landing, a portion of the antenna beam is likely to illuminate objects on the ground, such as buildings, trees or cars on a freeway. This scenario is depicted in Figure 5.1.

Most of the clutter energy appears around zero Doppler with reference to the aircraft ground speed, resulting from strong returns from stationary objects through the main beam of the antenna. Additional discrete clutter due to returns from moving objects on the ground, or returns from large objects through antenna sidelobes may appear at frequencies shifted away from zero Doppler. The large radar cross-section of these reflectors on the ground can backscatter enough energy to mask out any returns due to weather [53]. This complicates the detection of low-reflectivity weather phenomena, where most spectrum-oriented signal processing schemes will operate poorly in the presence of a severely low SCR.

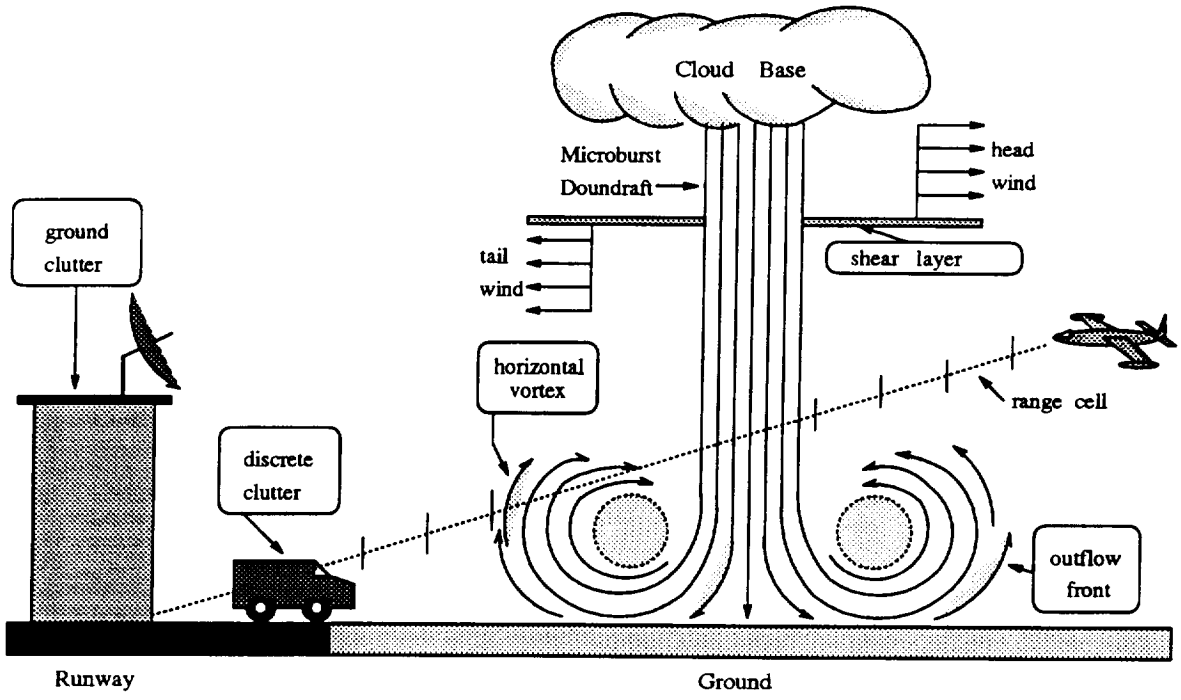


Figure 5.1. Microburst and wind speed profile.

Recently, researchers have suggested that the weather return and the clutter return be assumed statistically independent, and setting a signal processing goal of decorrelating a nonstationary wideband process (weather) from a nonstationary narrowband process (clutter) on a per-range cell basis, some preliminary analysis results have shown the potential applicability of LMS-based adaptive filters in solving this particular signal processing problem [14]. In the next two sections, we will try to further understand the nature of the interfering clutter return that deteriorates weather estimation, and to explore weather detection performances of various adaptive and passive approaches.

5.2 Modeling Airborne Doppler Radar Clutter Return

As elaborated in Chapter 1, the motivation for using adaptive filtering techniques in radar clutter rejection is that the separation of two signal processes can be better achieved by exploiting their temporal characteristics. Based on our discussions in Chapter 4 relative to the performance of adaptive filters, understanding of the interference (or noise) characteristics is more important than knowing the nature of the signal. Therefore, before the adaptive filtering techniques can be brought into this particular application, the statistical nature of the clutter processes needs some investigation.

Based upon examination of many clutter returns, especially those collected with the antenna scan angle kept between ± 5 degrees, in terms of their characteristics like dominant frequencies, spectral bandwidths, and time-domain fluctuations of average power level, we can model the clutter return from a specific range cell as multiple sinusoids of constant frequencies, amplitudes, and phases, mixed with complex zero mean white Gaussian noise, i.e.,

$$y(n) = \sum_{i=1}^{p} A_i \exp(j2\pi f_i n + \phi_i) + w(n).$$

The signal detection and parameter estimation problems associated with this signal model have been addressed in Chapter 2. As the simulation result in Chapter 3

testifies, the order-recursive EM frequency estimation algorithm combined with the MDL criterion demonstrates excellent performances including high spectral resolution, signal detection capability, low SNR threshold, and the statistical characteristics possessed by ML estimators. Therefore, this algorithm is used for modeling clutter returns.

Figures 5.2 and 5.3 show two case analyses of typical clutter returns collected from the Denver Stapleton airport, with the antenna main beam intercepting the ground during a level flight over an interstate highway. Figures 5.4 and 5.5 give similar analyses of clutter returns collected from the Philadelphia airport, where the aircraft is in a landing approach at about 300 ft above the ground. These spectra show the presence of dominant zero Doppler ground clutter and discrete clutter returns due to vehicles along the highway. Another two case analyses, as given in Figures 5.6 and 5.7, are associated with zero Doppler ground clutter returns collected from the Orlando airport. For these analyses, the model order (number of sinusoids) considered for the MDL order selection criterion ranges from 1 to 12, and the order p thus decided is indicated in each figure. Looking closely at these plots, one can see that our algorithm is able to provide a signal model which captures both the spectral mode and the temporal characteristics of the radar ground clutter returns.

5.3 Clutter Rejection via LMS-based Adaptive Filtering

5.3.1 Experiment for Performance Test

In order to independently measure the weather detection performances of different signal processing methods, *a priori* knowledge of both the weather and clutter returns should be available. To make this possible, we undertake an experiment as depicted in Figure 5.8, where a block diagram shows how *real* clutter returns can be merged with *simulated* weather returns to test clutter rejection capabilities of adaptive filters and fixed-notch filters. The windshear return is regenerated by Research Triangle Institute's "Airborne Windshear Doppler Radar Simulation (AWDRS)" program

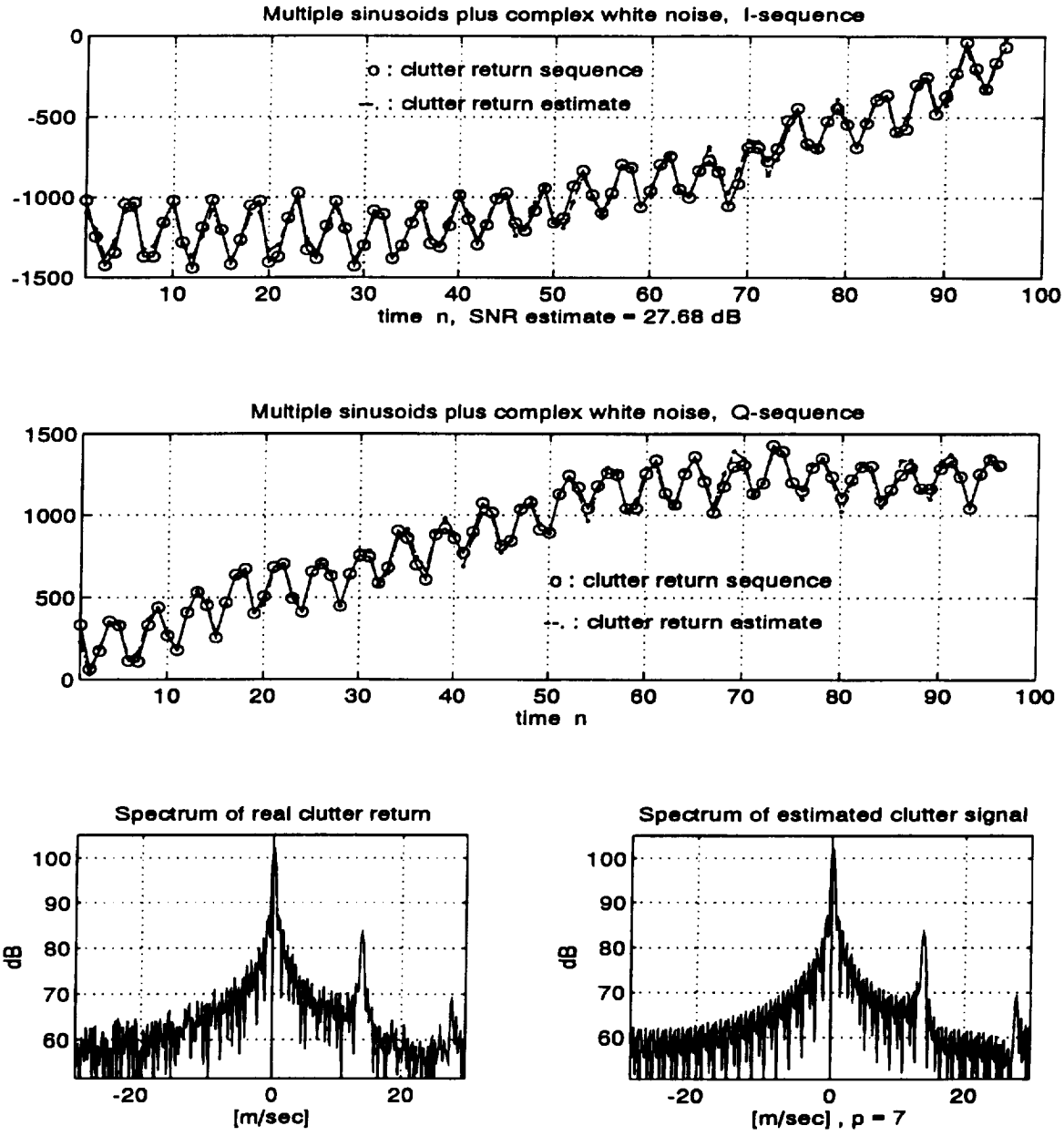


Figure 5.2. Modeling Denver ground clutter via EM algorithm and MDL criterion, $p = 7$, (NASA test flight-dn4c1s5.m18, frame-77, RC-65).

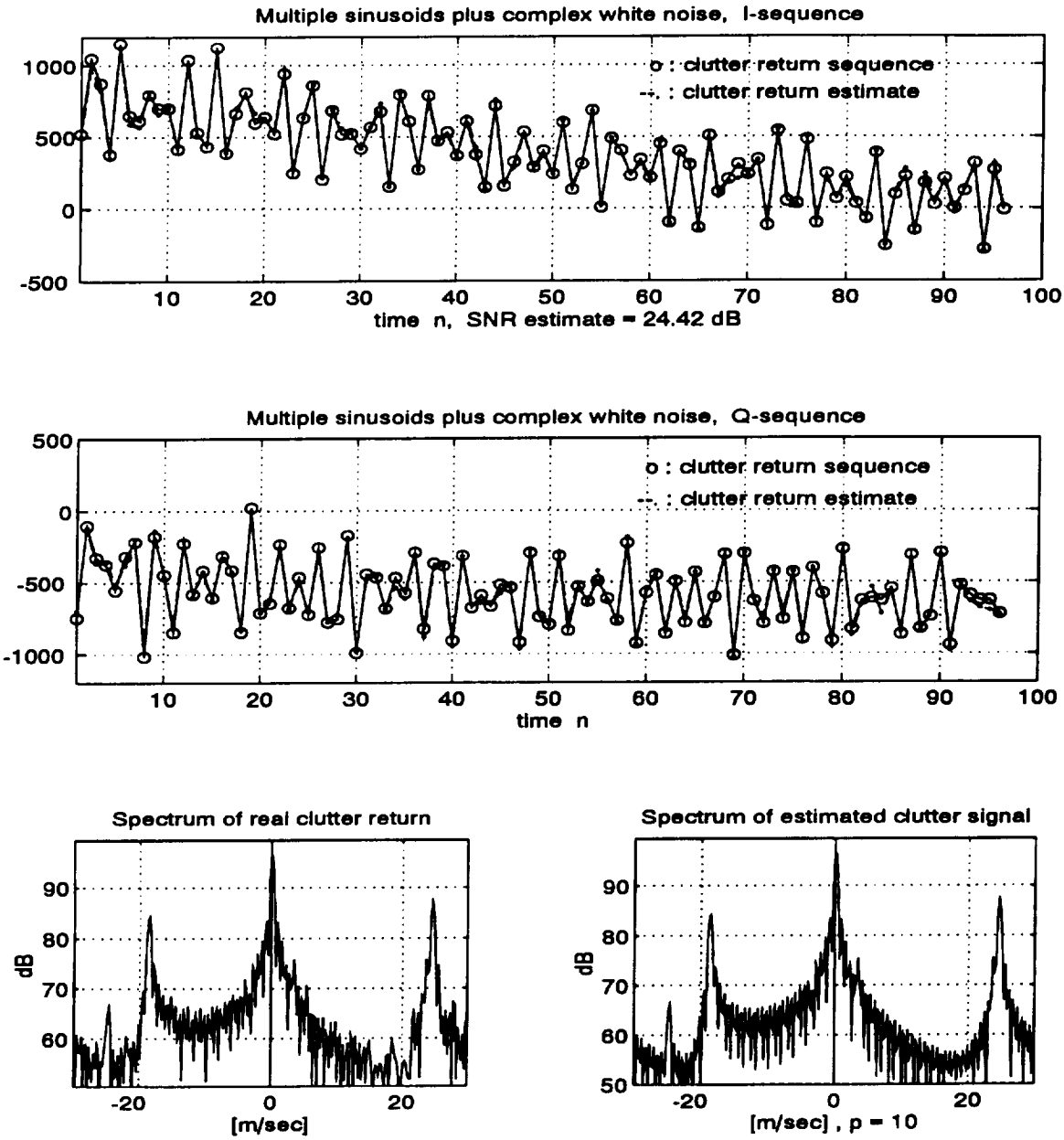


Figure 5.3. Modeling Denver ground clutter via EM algorithm and MDL criterion, $p = 10$, (NASA test flight-dn4c1s5.m18, frame-123, RC-50).

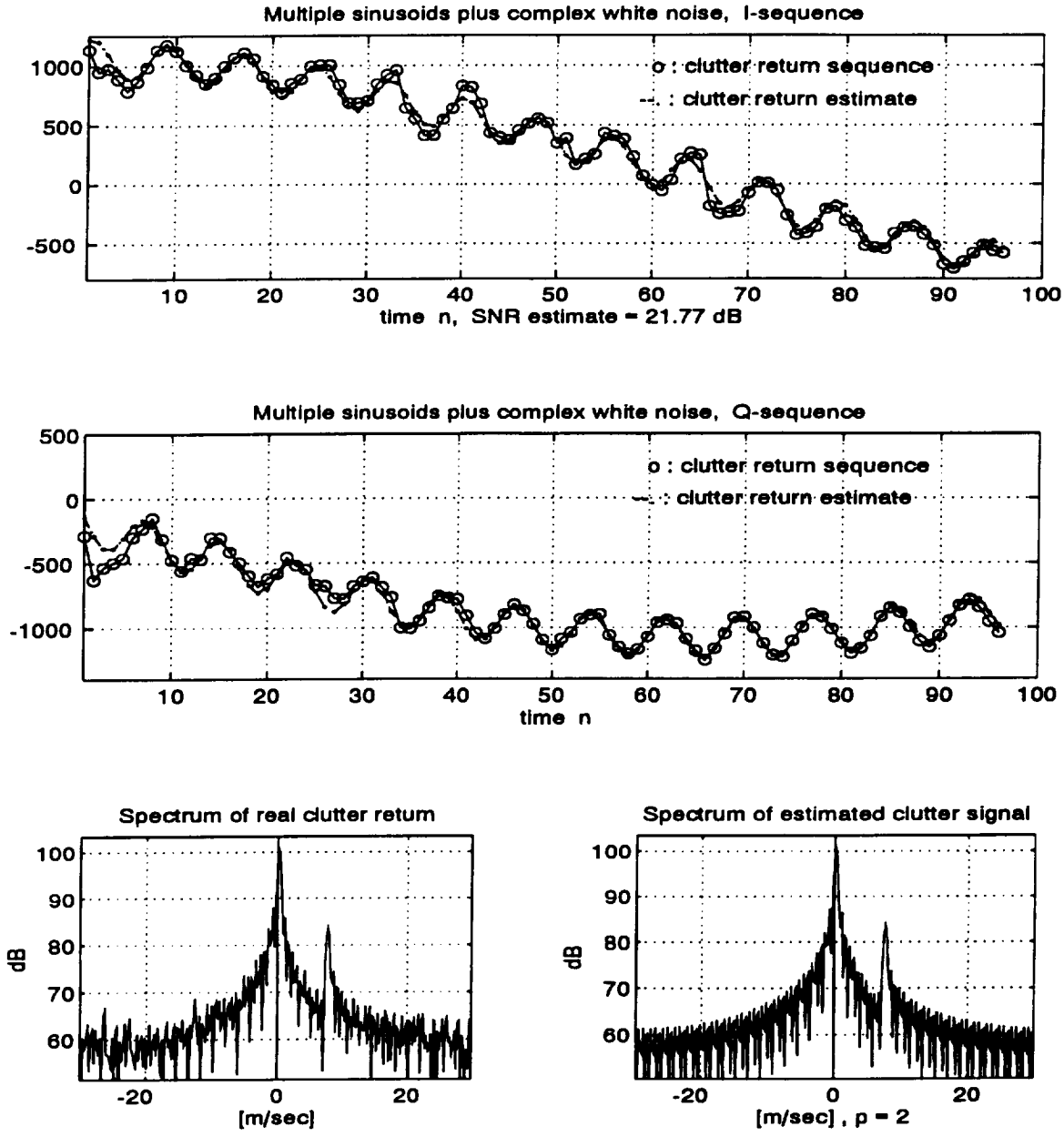


Figure 5.4. Modeling Philadelphia ground clutter via EM algorithm and MDL criterion, $p = 2$, (NASA test flight-f3r3s6.m14, frame-1036, RC-55).

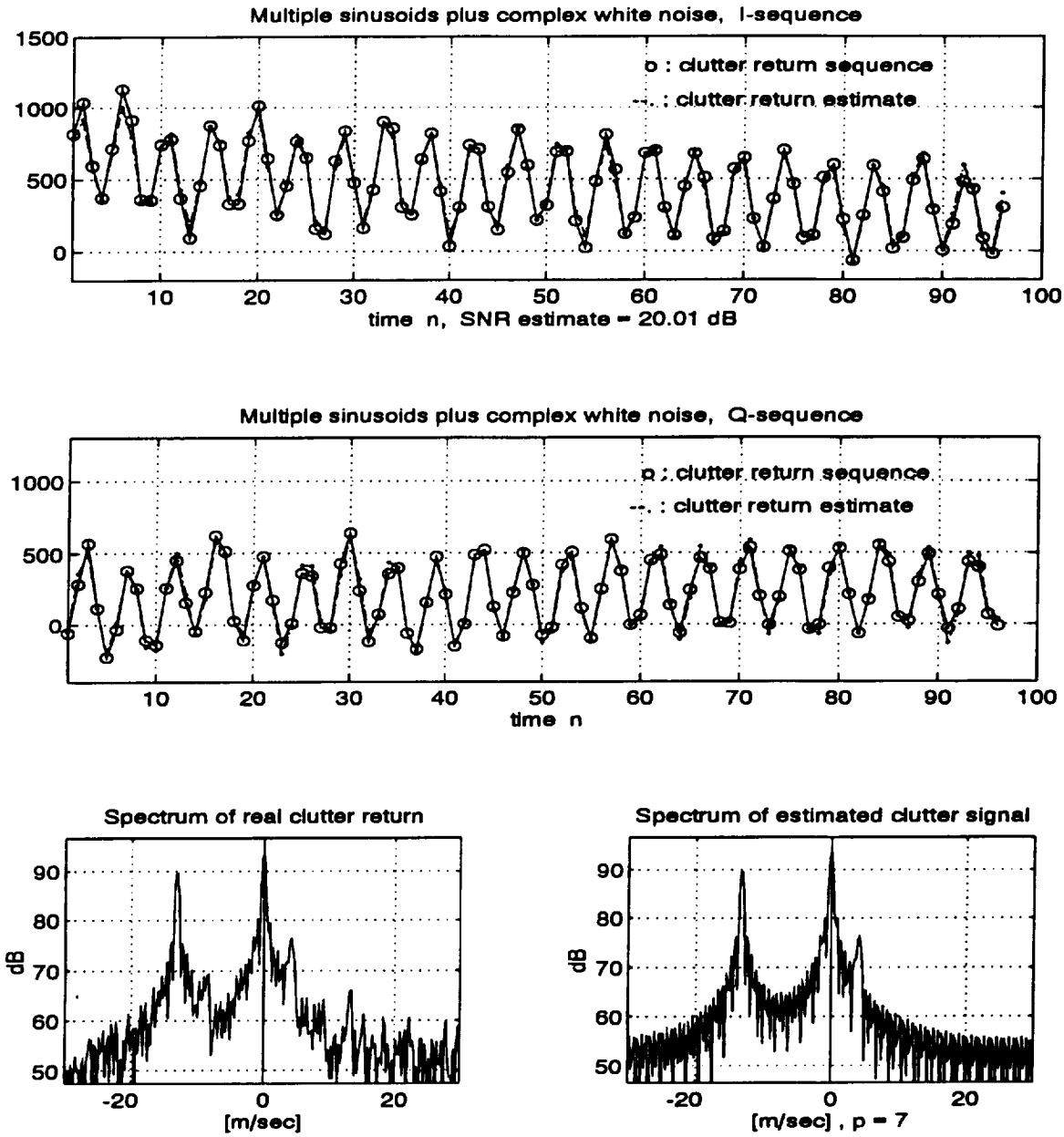


Figure 5.5. Modeling Philadelphia ground clutter via EM algorithm and MDL criterion, $p = 7$, (NASA test flight-f3r3s6.m14, frame-3496, RC-39).

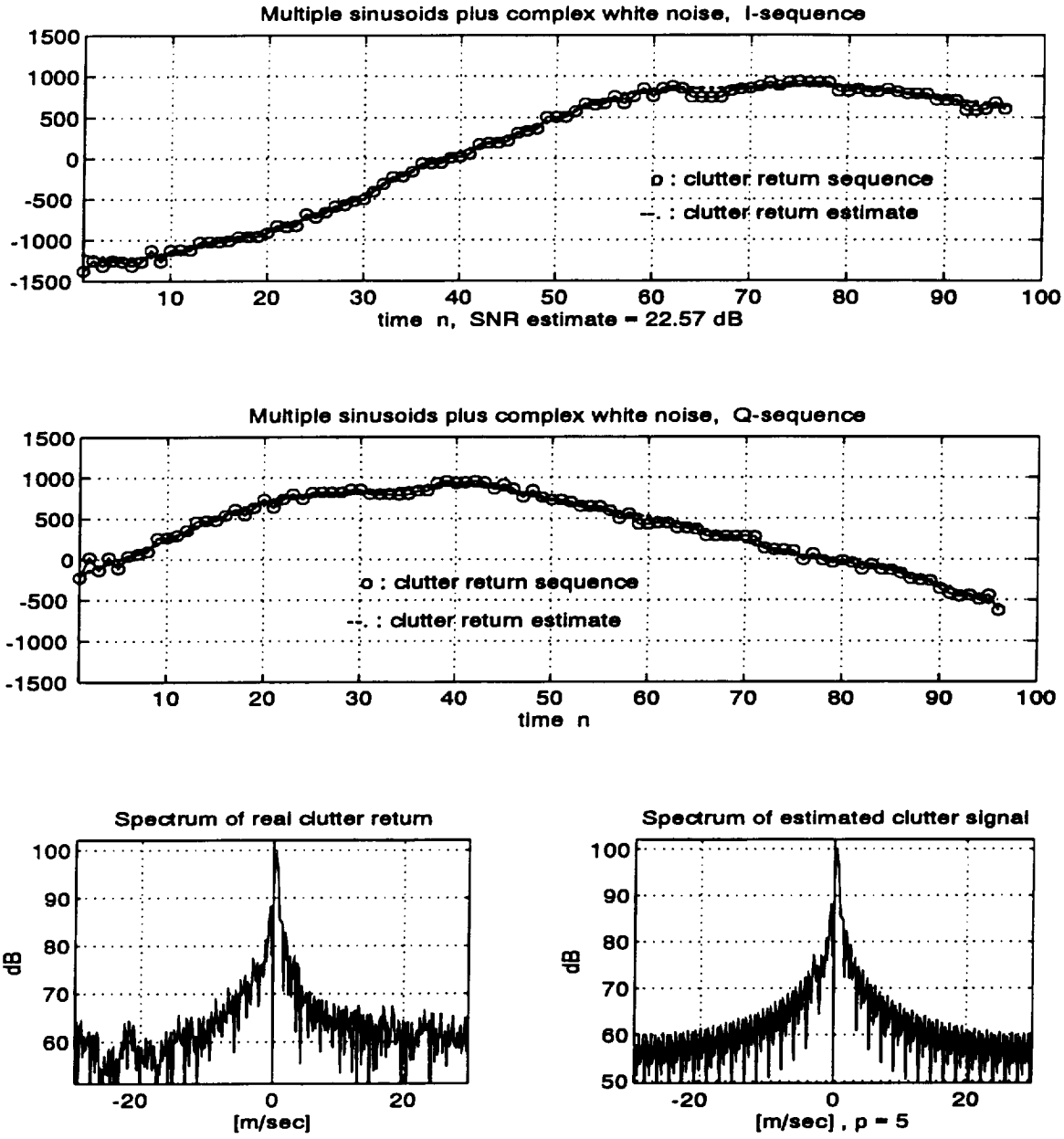


Figure 5.6. Modeling Orlando ground clutter via EM algorithm and MDL criterion, $p = 7$, (NASA test flight-o18c43s4.m6, frame-1232, RC-33).

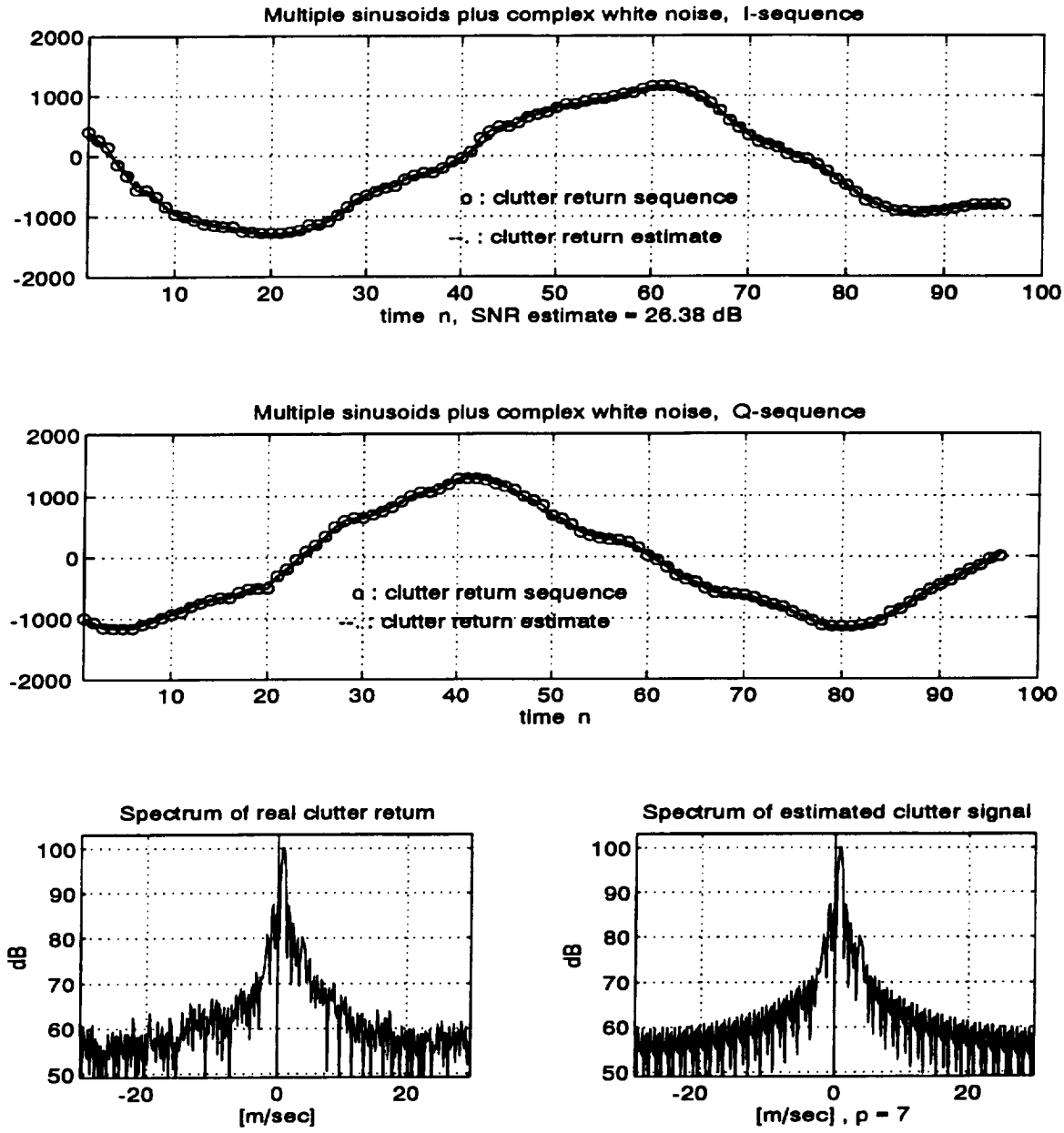


Figure 5.7. Modeling Orlando ground clutter via EM algorithm and MDL criterion, $p = 5$, (NASA test flight-o18c43s4.m6, frame-2394, RC-53).

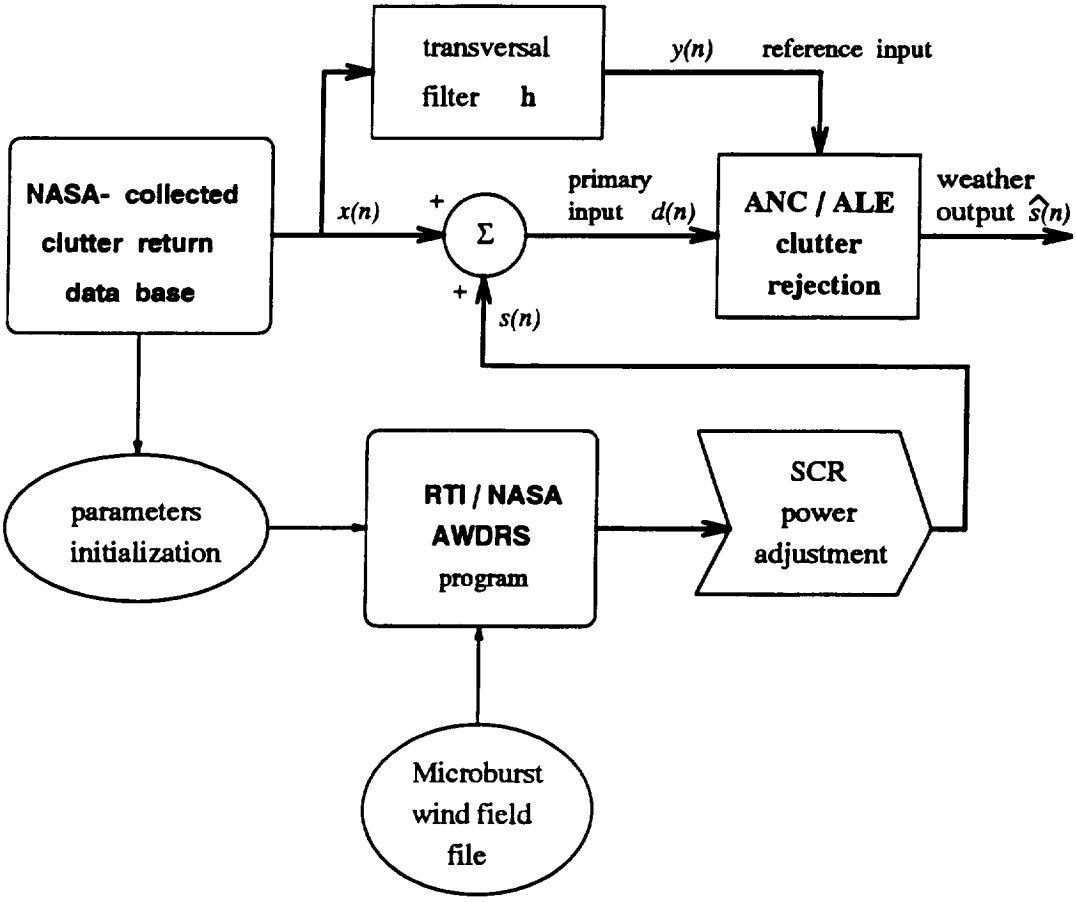


Figure 5.8. Schematic diagram for generation and processing of weather-plus-clutter return.

[54], assuming the same aircraft operation scenario where the near terminal ground clutter return is collected.

The data submitted for performance test comes from 156 range cells that contain clutter returns collected from the Denver Stapleton airport and the correspondingly simulated windshear returns, and each range cell has 96 signal samples. To create a specific SCR condition, the average power of the simulated weather return is adjusted with respect to that of the real clutter return on a range-cell by range-cell basis. After this SCR power adjustment, the weather return (signal) and the clutter return (interference) are designated as $s(n)$ and $x(n)$, respectively. The weather-plus-clutter return, $d(n) = s(n) + x(n)$, is then submitted for the performance test of different adaptive and fixed clutter rejection approaches.

In our simulation, two adaptive filter configurations, ANC (2nd order) and ALE (3rd order, $\Delta = 10$) implemented with either LMS or NLMS algorithms, are considered. We denote them as ANC-LMS, ANC-NLMS, ALE-LMS, and ALE-NLMS. For the ANC, the "reference clutter" input $y(n)$ is generated as a linear combination of the clutter return's delayed replicas, i.e.,

$$y(n) = \sum_{k=0}^{L-1} h_k x(n-k)$$

where denoting $H(z)$ and $W_o(z)$ respectively as the z-transforms of h_k and the optimum Wiener solution w_{ok}, $\mathbf{h} = [h_0\ h_1\ \cdots h_{L-1}]$ and the filter's number of taps M are chosen such that $H(z) \approx \dfrac{1}{W_o(z)}$. For the ALE, the number of delays Δ is chosen so as to decorrelate $s(n)$ and $s(n-\Delta)$, i.e., $E[s(n)s^*(n-\Delta)] \approx 0$. Remember that the ALE is the ANC with its reference input replaced by the delayed primary input, i.e.,

$$y(n) = d(n-\Delta) = s(n-\Delta) + x(n-\Delta).$$

Another clutter rejection filter brought into comparison is a Butterworth second order filter with 3 m/sec notch bandwidth centered at zero Doppler velocity.

Since the available data record in each range cell contains only 96 samples, these LMS-based adaptive filters can not be expected to converge (e.g., refer to Figure 4.4(e)), though a certain amount of clutter interference can still be suppressed. Furthermore, experience shows that to reduce the adaptation noise, smaller values of μ and $\tilde{\mu}$ (compared to the upper bounds given in (4.21) and (4.24)) are necessary. In order to alleviate this convergence problem and help stabilize the adaptive filters, both the ANC and ALE operate on the data twice by retaining the final filter tap weights obtained from the first processing as the initial tap weights for the second processing.

5.3.2 Simulation Results

Figures 5.9 and 5.10 demonstrate the weather detection performance of different clutter rejection schemes in terms of their standard deviations of post-processing wind velocity estimates, over a wide range of SCR (-25dB $\sim$ 10dB) conditions. In Figure 5.9, we consider 120 range cells that contain only near zero Doppler ground clutter return and weather return, whereas in Figure 5.10, 36 more range cells that have both zero Doppler and discrete clutter returns are also taken into account. Based on these simulation results, supposing the allowable standard deviation in wind velocity estimates is about 2 m/sec, several observations can be noted:

1. Basically, the presence of discrete clutter does not affect the performance of the ANC because the same amount of *a priori* knowledge regarding discrete clutter returns is available, though it significantly degrades that of the fixed Butterworth notch filter. However, when the SCR value is above a certain level (say, 0 dB), using the Butterworth filter is still an efficient method of clutter rejection.

2. On the other hand, the performance of both ALE-LMS and ALE-NLMS is also less sensitive to the discrete clutters than the Butterworth filter's. This can be seen in Figure 5.10 (in comparison with Figure 5.9), where the increase in the standard deviation of the ALE windspeed estimate caused by discrete clutter is less than 1 m/sec, whereas use of fixed notch filters increases the standard deviation by 2 m/sec.

3. When discrete clutter is absent, fixed notch Butterworth filters can efficiently suppress the zero Doppler ground clutter even with the SCR down to -10 dB. However, ANC-NLMS, ALE-LMS, and ALE-NLMS can increase the processing gain by at least 7 dB when zero Doppler ground clutter is the only hindrance to weather estimation. Considering the effects of the discrete clutter, adaptive filtering can obtain a processing gain of 7 to 15 dB.

4. For the ANC, the NLMS algorithm consistently outperforms the LMS algorithm for the μ and $\tilde{\mu}$ chosen, whereas ALE-LMS and ALE-NLMS have about the same performance. Curiously, for some SCR values, the ALE is able to outperform ANC. It may indicate the potential applicability of the adaptive line enhancer in windshear detection, when the collection of reference clutter returns required of the ANC implementation is technically difficult.

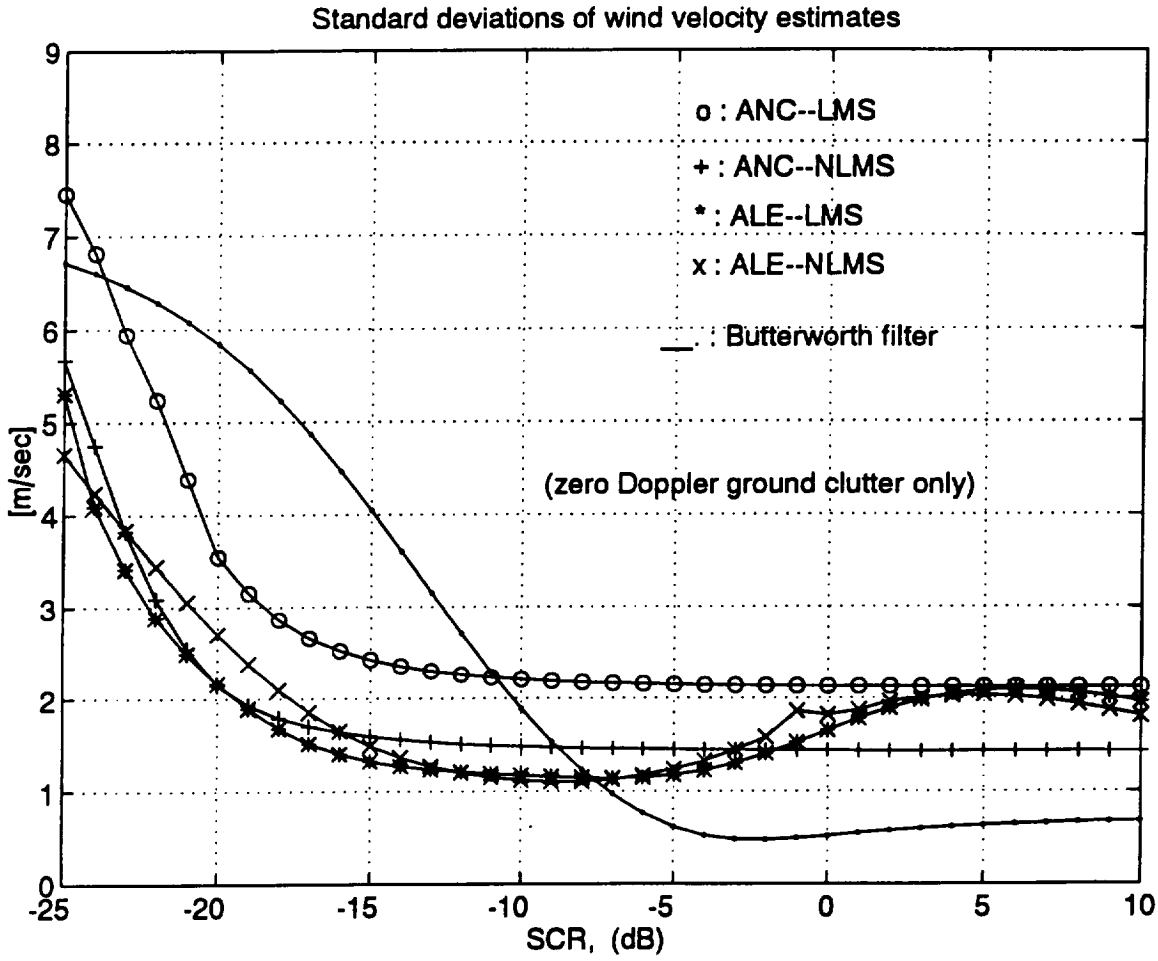


Figure 5.9. Performance comparisons for LMS-based adaptive filters and fixed-notch Butterworth filter, considering range cells that contain zero Doppler ground clutter and weather returns. (source of clutter: NASA test flight-dn4c1s5.m18).

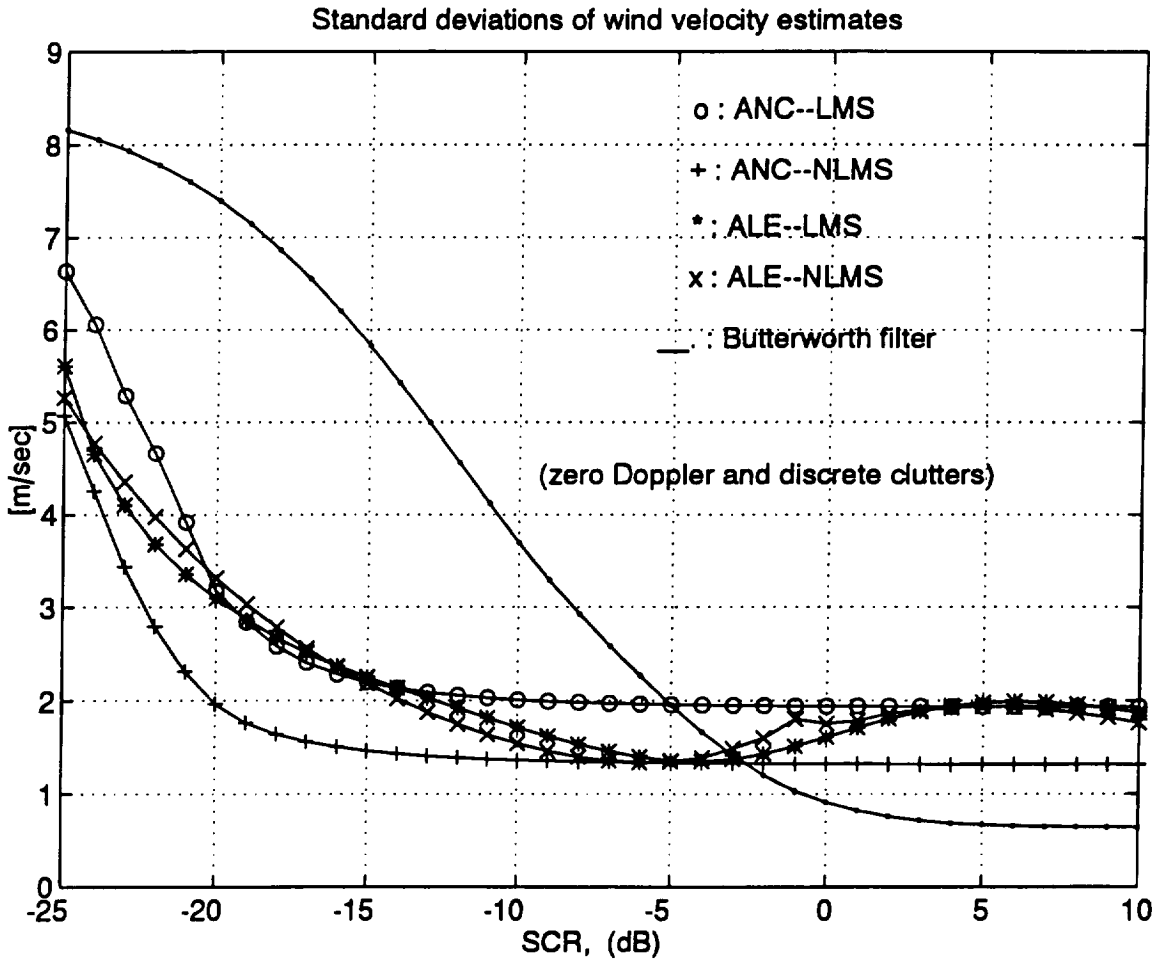


Figure 5.10. Performance comparisons for LMS-based adaptive filters and fixed-notch Butterworth filter, considering range cells that contain zero Doppler ground clutter, discrete clutter, and weather returns. (source of clutter: NASA test flight-dn4c1s5.m18).

CHAPTER 6

CONCLUSIONS

This work basically deals with three research topics in signal processing and its application: (1) signal modeling and parameter estimation, especially for narrow-band signals modeled as multiple sinusoids plus white Gaussian noise, (2) LMS-based adaptive filter convergence response in this particular signal environment, and (3) application of these signal processing techniques in airborne Doppler weather radar. Concluding remarks regarding these three areas are presented in the next section.

6.1 Discussion of Results

6.1.1 Frequency Estimation and Signal Modeling

Our focus of investigation is on signals that consist of multiple sinusoids plus white Gaussian noise, particularly when only a short record of observations is available. Thanks to the simplicity of the complete data model and the monotonically increasing conditional likelihood function, the EM algorithm is able to solve the problem of maximizing the highly nonlinear likelihood function of the signal frequencies.

Simulation results show that even when the SNR level is very low, our algorithm has high frequency resolution capability which most periodogram- or AR-oriented approaches fail to possess. The building block of this algorithm is a one-dimensional ML frequency estimator, which can be implemented via the Newton-Raphson method or FFT, depending on the required numerical accuracy. This indicates the computational flexibility of our algorithm. Taking a closer look at Figure 3.1(a), one can find another feature of computational parallelism in the sense that all the signal parameters (frequencies) can be *simultaneously* estimated. Further, according to its order-recursiveness, for each model order $m \leq p$, the initial frequency estimate is

the *optimal* (in the ML sense) estimate of previous order $(m-1)$ *plus* one additional estimate extracted from the model residual $\mathbf{z}_{m-1}$.

For order selection, we propose the use of Fisher's T_o statistic and the MDL criterion tailored for complex data. High probability of correct signal detection of these two methods has been observed in Monte Carlo simulation.

6.1.2 LMS Adaptive Filtering with the Sinusoidal Process

As an extension of Glover's work [47] for the case of the complex-valued signal, a notch filter implemented by the adaptive noise canceler for suppressing multiple sinusoidal interference is investigated through analysis and simulation. Results show that a notch filter with deep and narrow spectral nulls indeed can be realized by the ANC. Regarding the ANC's performance response to the sinusoids-plus-noise environment, several observations can be made based on our simulation results:

1. For the LMS algorithm, to maintain stability, the step size μ should be *small* compared to the upper bound specified by independence theory, whereas the NLMS algorithm with $\tilde{\mu} = 1$ generally offers improved performance in terms of a better tradeoff between convergence speed and misadjustment, and stability in a more ill-conditioned signal environment.

2. The LMS and NLMS algorithms are capable of approximately tracking the optimum Wiener solution and thus totally rejecting the reference-input-dependent component present in the primary input, when the filter's number of taps is greater than or equal to the order of correlation between both inputs. Therefore, in most applications where the correlation property associated with the signal and noise is unknown, filters of higher order should be employed.

3. The NLMS algorithm's convergence response is less sensitive to the eigenvalue disparity associated with the signal's autocorrelation matrix than that of the LMS algorithm, though not as significantly as claimed by Slock [28]. It should be noted that these nice features associated with the NLMS algorithm only come with extra computational cost.

6.1.3 Windshear Detection Application

Based upon our analysis and the simulation results, clutter returns have been modeled as multiple sinusoids plus Gaussian noise using the EM frequency estimation

algorithm and the MDL criterion. The goodness of fit, in terms of the clutter process's temporal and spectral characteristics, demonstrates the robustness of our algorithm.

Simulation results show that when the SCR level is sufficiently high and only zero Doppler clutter is present to bias the estimation of weather information, the fixed Butterworth filter is capable of removing the clutter return and thus enhancing the weather return. However, when both zero Doppler ground clutter and discrete clutter are present in a very low SCR situation, adaptive filtering approaches can significantly improve the windshear detection capability. Considering the performance of all clutter rejection methods, either adaptive or passive, Figure 5.10 showed that the ANC-NLMS stands out as the best approach. Nevertheless, the overall performance of the ALE demonstrates a potential applicability in this particular signal processing problem, when the implementation of the adaptive noise canceler is impractical.

6.2 Suggestions for Future Work

Though our EM frequency estimator produces simulation results that reveal the features of the MLE, an analytical result that guarantees the global maximization of the likelihood function via the EM algorithm is still not available. Since the EM algorithm is potentially capable of solving many MLE problems in signal processing, the problems of how to simplify the computation involved for real-time application and the analytical verification of its convergence properties in a case-specific fashion will be an area for future work. To model clutter returns using the EM frequency estimation algorithm and the MDL criterion, our focus is on the *narrowband* clutter data collected with the antenna scan angle kept between ± 5 degrees. Clutter collected otherwise may have broader bandwidths, and alternative modeling approaches may be necessary.

As for the adaptive filters, though the LMS (NLMS) algorithm is popular due to its computational efficiency and ease of implementation, its convergence is slow, and the statistical analysis of it for many practical applications, e.g., the realistic signal

model proposed in Section 4.2, can be mathematically intractable. The independence theory and most analysis results can only provide a picture of this algorithm's statistical behavior from a certain perspective. Furthermore, in a radar application, the data available in each range cell is usually quite limited, and the autocorrelation matrix of the clutter return can be very ill-conditioned. The convergence performance of the LMS-based adaptive filters in this type application is an open question. How to design adaptive filters with real-time potential, fast convergence, and numerical stability, indeed warrants further investigation.

APPENDIX A

ML FREQUENCY ESTIMATOR FOR A SINUSOID PLUS CWGN

Consider the signal $y(n)$ of a single sinusoid plus complex white Gaussian noise as described in Section 2.2.1, i.e. $y(n) = A\exp(j2\pi fn) + w(n),\ \ n = 0, 1, \cdots, N-1$. According to (2.2), the MLE of f is given as $\hat{f} = \arg\max_f \frac{1}{N}J(f)$ where

$$J(f) = \left|\sum_{n=0}^{N-1} y(n)\exp(-j2\pi fn)\right|^2 = \left|\mathbf{e}^H\mathbf{y}\right|^2 = \mathbf{y}^T\mathbf{e}^*\mathbf{e}^T\mathbf{y}^*.$$

Let

$$\mathbf{Q}(f) = \mathbf{e}^*\mathbf{e}^T = \begin{bmatrix} 1 & e^{j2\pi f} & e^{j2\pi f2} & \cdots & e^{j2\pi f(N-1)} \\ e^{-j2\pi f} & 1 & e^{j2\pi f} & \cdots & e^{j2\pi f(N-2)} \\ e^{-j2\pi f2} & e^{-j2\pi f} & 1 & \cdots & e^{j2\pi f(N-3)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ e^{-j2\pi f(N-1)} & e^{-j2\pi f(N-2)} & e^{-j2\pi f(N-3)} & \cdots & 1 \end{bmatrix},$$

then

$$J'(f) = \frac{dJ(f)}{df} = \mathbf{y}^T\frac{d\mathbf{Q}(f)}{df}\mathbf{y}^* \ \text{ and } \ J''(f) = \frac{d^2J(f)}{df^2} = \mathbf{y}^T\frac{d^2\mathbf{Q}(f)}{df^2}\mathbf{y}^*$$

where

$$\frac{d\mathbf{Q}(f)}{df} = -j2\pi\begin{bmatrix} 0 & -e^{j2\pi f} & \cdots & -(N-1)e^{j2\pi f(N-1)} \\ e^{-j2\pi f} & 0 & \cdots & -(N-2)e^{j2\pi f(N-2)} \\ 2e^{-j2\pi f2} & e^{-j2\pi f} & \cdots & \vdots \\ \vdots & \vdots & \ddots & -e^{j2\pi f} \\ (N-1)e^{-j2\pi f(N-1)} & \cdots & e^{-j2\pi f} & 0 \end{bmatrix},$$

and

$$\frac{d^2\mathbf{Q}(f)}{df^2} = -4\pi^2\begin{bmatrix} 0 & e^{j2\pi f} & \cdots & (N-1)^2e^{j2\pi f(N-1)} \\ e^{-j2\pi f} & 0 & \cdots & (N-2)^2e^{j2\pi f(N-2)} \\ 4e^{-j2\pi f2} & e^{-j2\pi f} & \cdots & \vdots \\ \vdots & \vdots & \ddots & e^{j2\pi f} \\ (N-1)^2e^{-j2\pi f(N-1)} & \cdots & e^{-j2\pi f} & 0 \end{bmatrix}.$$

Given an initial frequency estimate f_0 obtained from a coarse FFT, the Newton-

Raphson iterative scheme calculates

$$f_{k+1} = f_k + \frac{J'(f_k)}{J''(f_k)}$$

until $|f_{k+1} - f_k| \leq \epsilon$.

REFERENCES

1. M. B. Priestley. *"Non-linear and Non-stationary Time Series Analysis"*. Academic Press, San Diego, CA, 1988.

2. S. L. Marple Jr. *"Digital Spectral Analysis"*. Prentice Hall, Englewood Cliffs, NJ, 1987.

3. S. M. Kay. *"Modern Spectral Estimation"*. Prentice Hall, Englewood Cliffs, NJ, 1988.

4. G. Janacek and L. Swift. *"Time Series — Forecasting, Simulation, Applications"*. Ellis Horwood Limited, Chichester, England, 1993.

5. A. C. Harvey. *"The Econometric Analysis of Time Series"*. 2 edition, The MIT Press, Cambridge, MA, 1990.

6. Y. Grenier. "Time-Dependent ARMA Modeling of Nonstationary Signals". *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 31(4):899–911, 1983.

7. S. S. Haykin. *"Adaptive Filter Theory"*. 2 edition, Prentice Hall, Englewood Cliffs, NJ, 1991.

8. N. R. Goodman. "Statistical Analysis Based on a Certain Multivariate Complex Gaussian Distribution". *Annals of Mathematical Statistics*, 34:152–176, 1963.

9. I. Ziskind and M. Wax. "Maximum Likelihood Localization of Multiple Sources by Alternating Projection". *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(10):1553–1560, 1988.

10. E. G. Baxa, Jr. "Airborne Pulsed Doppler Radar Detection of Low-Altitude Windshear—A Signal Processing Problem". *Digital Signal Processing*, 1(4):115–140, October 1991.

11. E. G. Baxa, Jr., Y. Lai, and M. Kunkle. "New Signal Processing Developments in the Detection of Low-Altitude Windshear with Airborne Doppler Radar". In *Proceedings of IEEE National Radar Conference*, Boston, 1993.

12. T. T. Fujita. "The Downburst: Microburst and Macroburst". SMRP Research Paper No. 210, Univ. of Chicago, 1985. (available from NTIS as PB85-148-880).

13. P. R. Mahapatra and D. S. Zernić. "Sensors and Systems to Enhance Aviation Safety Against Weather Hazards". *Proceedings of the IEEE*, 79(9):1234–1267, 1991.

14. Y. Lai and E. G. Baxa, Jr. "On the Application of the LMS-Based Adaptive Noise Canceller in Nonstationary Environment Associated with Airborne Doppler Radar". In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Minneapolis, 1993.

15. D. W. Tufts and R. Kumaresan. "Estimation of Frequencies of Multiple Sinusoids: Making Linear Prediction Perform Like Maximum Likelihood". *Proceedings of The IEEE*, 70(9):975–989, 1982.

16. J.-K. Hwang and Y.-C. Chen. "Superresolution Frequency Estimation by Alternating Notch Periodogram". *IEEE Transactions on Signal Processing*, 41(2):727–741, 1993.

17. B. Widrow and S. D. Stearns. *"Adaptive Signal Processing"*. Prentice Hall, Englewood Cliffs, NJ, 1985.

18. L. S. Biradar and V. U. Reddy. "A Newton-Based Ziskind-Wax Alternating Projection Algorithm Applied to Spectral Estimation". *IEEE Transactions on Signal Processing*, 41(3):1435–1438, 1993.

19. H. Van hamme. "Maximum Likelihood Estimation of Superimposed Complex Sinusoids in White Gaussian Noise by Reduced Effort Coarse Search (RECS)". *IEEE Transactions on Signal Processing*, 39(2):536–538, 1991.

20. M. Wax and T. Kailath. "Detection of Signals by Information Theoretic Criteria". *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33:387–392, 1985.

21. H. Akaike. "A New Look at the Statistical Model Identification". *IEEE Transactions on Automatic Control*, AC-19(6):716–723, 1974.

22. J. Rissanen. "A Universal Prior for Integers and Estimation by Minimum Description Length". *The Annals of Statistics*, 11(2):416–431, 1983.

23. S. M. Kay. *"Fundamentals of Statistical Signal Processing—Estimation Theory"*. Prentice Hall, Englewood Cliffs, NJ, 1993.

24. D. H. Brandwood. "A Complex Gradient Operator and Its Application in Adaptive Array Theory". *IEE Proceedings*, 130(1):11–16, 1983.

25. B. Widrow, J. McCool, and M. Ball. "The Complex LMS Algorithm". *Proceedings of the IEEE*, 63:719–720, 1975.

26. R. R. Bitmead and B. D. O. Anderson. "Performance of Adaptive Estimation Algorithms in Dependent Random Environments". *IEEE Transactions on Automatic Control*, AC-25:788–794, 1980.

27. G. C. Goodwin and K. S. Sin. *"Adaptive Filtering Prediction and Control"*. Prentice Hall, Englewood Cliffs, NJ, 1984.

28. Dirk T. M. Slock. "On the Convergence Behavior of the LMS and the Normalized LMS Algorithms". *IEEE Transactions on Signal Processing*, 41(9):2811–2825, 1993.

29. R. Nitzberg. *"Adaptive Signal Processing for Radar"*. Artech House, Norwood, MA, 1991.

30. B. Widrow, J. M. McCool, M. G. Larimore, and C. R. Johnson, Jr. "Stationary and Nonstationary Characteristics of LMS Adaptive Filters". *Proceedings of the IEEE*, 64(8):1151–1162, 1976.

31. B. Widrow and E. Walach. "On the Statistical Efficiency of the LMS Algorithm with Nonstationary Inputs". *IEEE Transactions on Information Theory*, IT-30(2):211–221, 1984.

32. O. Macchi. "Optimization of Adaptive Identification for Time-Varying Filters". *IEEE Transactions on Automatic Control*, AC-31(3):283–287, 1986.

33. L. E. Baum, G. Petrie, G. Soules, and N. Weiss. "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains". *Annals of Mathematical Statistics*, 41:164–171, 1970.

34. H. O. Hartley and R. R. Hocking. "The Analysis of Incomplete Data". *Biometrics*, 27:783–808, 1971.

35. R. Sundberg. "Maximum Likelihood Theory for Incomplete Data from an Exponential Family". *Scandinavian Journal of Statistics*, 1:49–58, 1974.

36. A. P. Dempster, N. M. Laird, and D. B. Rubin. "Maximum Likelihood from Incomplete Data via the EM Algorithm". *Journal of the Royal Statistical Society*, B–39:1–38, 1977.

37. M. Feder and E. Weinstein. "Parameter Estimation of Superimposed Signals Using the EM Algorithm". *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(4):477–489, 1988.

38. M. I. Miller and D. R. Fuhrmann. "Maximum-Likelihood Narrow-Band Direction Finding and the EM Algorithm". *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(9):1560–1577, 1990.

39. M. Segal, E. Weinstein, and B. R. Musicus. "Estimate-Maximize Algorithms for Multichannel Time Delay and Signal Estimation". *IEEE Transactions on Signal Processing*, 39(1):1–16, 1991.

40. C. R. Rao. *"Linear Statistical Inference and Its Applications"*. Wiley, New York, 1965.

41. P. J. Diggle. *"Time Series — A Biostatistical Introduction"*. Oxford University Press, New York, 1990.

42. R. A. Fisher. "Tests of Significance in Harmonic Analysis". *Proceedings of the Royal Society*, A 125:54–59, 1929.

43. C. F. J. Wu. "On the Convergence Properties of the EM Algorithm". *The Annals of Statistics*, 11(1):95–103, 1983.

44. R. A. Boyles. "On the Convergence of the EM Algorithm". *Journal of the Royal Statistical Society*, B 45(1):47–50, 1983.

45. D. C. Rife and R. R. Boorstyn. "Multiple Tone Parameter Estimation from Discrete-Time Observations". *The Bell System Technical Journal*, 55:1389–1410, 1976.

46. B. Widrow, J. R. Glover, Jr., J. M. McCool, J. Kaunitz, C. S. Williams, R. H. Hearn, J. R. Zeidler, E. Dong, Jr., and R. C. Goodlin. "Adaptive Noise Cancelling: Principles and Applications". *Proceedings of the IEEE*, 63:1692–1716, 1975.

47. J. R. Glover, Jr. "Adaptive Noise Canceling Applied to Sinusoidal Interferences". *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 25(6):484–491, 1977.

48. S. J. Orfanidis. *"Optimum Signal Processing"*. 2 edition, McGraw-Hill, New York, 1988.

49. W. A. Gardner. "Learning Characteristics of Stochastic-Gradient-Descent Algorithms". *Signal Processing*, 6:113–133, 1984.

50. M. Tarrab and A. Feuer. "Convergence and Performance Analysis of the Normalized LMS Algorithm with Uncorrelated Gaussian Data". *IEEE Transactions on Information Theory*, IT-34(4):680–691, 1988.

51. J. E. Mazo. "On the Independence Theory on Equalizer Convergence". *The Bell System Technical Journal*, 58(5):963–993, 1979.

52. D. D. Aalfs, E. G. Baxa, Jr., and E. M. Bracalente. "Signal Processing Aspects of Windshear Detection". *Microwave Journal*, 36(9):76–96, 1993.

53. M. I. Skolnik (ed.). *"Radar Handbook"*. 2 edition, McGraw-Hill, New York, 1990.

54. C. L. Britt. "Users Guide for an Airborne Windshear Doppler Radar Simulation (AWDRS) Program". NASA CR-182025, DOT/FAA/RD-91/2, Research Triangle Institute, June 1990.