1996

Signal estimation using H [infinity sign] criteria

Sanjeev Tavathia
Iowa State University

Signal estimation using H_∞ criteria

by

Sanjeev Tavathia

A Dissertation Submitted to the

Graduate Faculty in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Department: Electrical and Computer Engineering
Major: Electrical Engineering (Communications and Signal Processing)

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Members of the Committee:

Signature was redacted for privacy.

Signature was redacted for privacy.

Signature was redacted for privacy.

Signature was redacted for privacy.

Iowa State University
Ames, Iowa

1996

UMI Number: 9620988

TABLE OF CONTENTS

ACKNOWLEDGMENTS v

CHAPTER 1. INTRODUCTION 1

CHAPTER 2. SCIENTIFIC BACKGROUND 8

CHAPTER 3. CHARACTERISTICS OF H_∞ FILTERS 13

3.1 Stochastic H_∞ Filters 13

3.2 H_∞ Filters for Finite Energy Signals 17

3.3 Advantages of H_∞ Filters 19

3.3.1 Remarks 21

CHAPTER 4. MATHEMATICAL SOLUTION 23

4.1 H_∞ Norm of Continuous Transfer Function 24

4.2 H_∞ Norm of Discrete Transfer Function 25

4.3 Frequency Domain Solutions 26

4.3.1 Model Matching Solution 26

4.3.1.1 Discrete Time 26

4.3.1.2 Continuous Time 28

4.4 Time Domain Solutions 30

4.4.1 Discrete Time 30

4.4.2 Continuous Time 33

4.5 Krein Space 41

4.6 Least Squares Solution 43

4.6.1 Wiener Solution 43

4.6.2 Kalman Filter 44

4.7 FIR Filter Solution 46

4.7.1 Wiener Filter 46

4.7.2 H_∞ Solution 49

CHAPTER 5. SIGNAL ESTIMATION 51

5.1 Continuous Time System 51

5.2 Discrete Time System 53

5.3 CDMA System 62

5.4 Robust Performance 68

CHAPTER 6. ADAPTIVE FILTERS 74

6.1 H_∞ Problem Formulation 74

6.2 Adaptive H_∞ Filters 77

6.3 Simulation Results 86

6.4 Acoustic Echo Canceller 94

CHAPTER 7. CONCLUSION 104

CHAPTER 8. FUTURE WORK 107

BIBLIOGRAPHY 109

APPENDIX A 113

APPENDIX B 116

ACKNOWLEDGMENTS

I would like to thank John Doherty, my advisor, for his excellent guidance and help through out my thesis work. I would like to thank John and Steven L. Gay for providing me with an opportunity to work at AT&T Bell Labs which besides excellent work at Bell Labs helped me financially to continue my Ph.D. work.

I would also like to thank all members of Ph.D. committee for supporting me in pursuing on H_∞ filtering idea.

Thanks to Dr. Horton for his guidance towards teaching assistantship at Department of Electrical Engineering, Iowa State which provided financial support for my Ph.D. program.

NOTATIONS AND DEFINITIONS

R^n : *n dimensional vector space of real numbers.*

C^n : *n dimensional vector space of complex numbers.*

$a, A, \mathbf{A}$: *scalar number (integer, real, or complex number) is represented by lower case, i.e., a. Vector with scalar numbers is represented by upper case, i.e., A. Matrix with scalar numbers is represented by upper case bold, i.e., $\mathbf{A}$.*

$r(t)$: *represent continuous time signal. t is a real number.*

$r(n)$: *represent discrete time signal. Let t_{sam} be the sampling interval used to obtain discrete time signal $r(n)$ from continuous time signal $r(t)$. Then $r(n) = r(t = n\, t_{sam})$. n and t_{sam} are integers.*

$x(t), x(n), X(t), X(n), \mathbf{X}(t), \mathbf{X}(n)$: *Time domain scalar signals are represented by lower case, i.e., $x(t)$ or $x(n)$. Vector valued signals are represented by upper case, i.e., $X(t)$ or $X(n)$. Matrix valued signals are represented by bold upper case, i.e., $\mathbf{X}(t)$ or $\mathbf{X}(n)$.*

$X'(t)$: *represent differential of $X(t)$ with respect to t, i.e., $X'(t) = \frac{d}{dt}(X(t))$.*

$G(s)$, $G(z)$, $\mathbf{G}(s)$, $\mathbf{G}(z)$: *Transfer function with upper case letter represent single-input single-output function, i.e., $G(s)$ or $G(z)$. Transfer function with upper case bold letter represent multi-input or multi-output transfer functions, i.e., $\mathbf{G}(s)$ or $\mathbf{G}(z)$. $s = \sigma + j\omega$,*

$-\infty < \omega < \infty$, $z = ke^{j\Omega}$, $-\pi \le \Omega \le \pi$. *For stable transfer functions* $1 \le k < \infty$ *and* $0 \le \sigma < \infty$.

G(s) with state space matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ *is represented as:*

$$G(s) = C(sI - A)^{-1}B + D$$

$$X^{'}(t) = \mathbf{A}\,X(t) + \mathbf{B}\,\mathbf{W}(t)$$
$$Y(t) = \mathbf{C}\,X(t) + \mathbf{D}\,\mathbf{W}(t)$$

where $X(t)$ *is a state variable vector,* $\mathbf{W}(t)$ *is input signal matrix to* $G(s)$*, and* $Y(t)$ *is output signal vector of* $G(s)$.

H_∞ *: Complex valued functions which are analytic and bounded in the open right half plane.*

L_∞ *: Complex valued functions which are analytic and bounded on* $j\omega$ *axis.*

RH_∞ *: Subset of* H_∞ *functions consisting of real-rational functions.*

RL_∞ *: Subset of* L_∞ *consisting of real-rational functions.*

$\|F(s)\|_\infty$ *:* H_∞ *norm of* F. $\|F\|_\infty = \sup\{|F(s)| : \mathrm{Re}(s) > 0\} = \sup\{|F(j\omega)| : \omega \in real\ \ number\}$

where Re(s) denotes the real part of complex number s, $s = \sigma + j\omega$, σ *is a real number.*

dis(R, RH_∞ *): Distance from* $R \in RL_\infty$ *to subspace* RH_∞.

$\|\Gamma_R\|$ *: Hankel norm of* R. $\|\Gamma_R\| = \inf_{X \in RH_\infty} \|R - X\|_\infty$; $R \in RL_\infty$.

$|x|$ *: represent absolute value of x.*

δ_{nk} *: represent dirac-delta function.* $\delta_{nk} = \begin{cases} 1 & if\ n = k \\ 0 & if\ n \ne k \end{cases}$.

Finite Energy Signal (FES): *$r(t)$ is continuous time finite energy signal if*

$\int_{-\infty}^{\infty} r(t)\,r^*(t)\,dt < \infty$. *$r(n)$ is discrete time finite energy signal if* $\sum_{n=-\infty}^{\infty} r(n)\,r^*(n) \;<\; \infty$.

Second norm of FES or 2-norm: *If $r(t)$ or $r(n)$ is FES then second norm of $r(t)$ or $r(n)$*

is defined as: $\|r\|_{2,[t_1,t_2]} \;=\; \int_{t_1}^{t_2} r(t)r^*(t)\,dt$ *or* $\|r\|_{2,[n_1,n_2]} \;=\; \sum_{n=n_1}^{n_2} r(n)r^*(n)$.

Zero mean wide-sense stationary signals: *$r(t)$ is zero mean wide-sense stationary (WSS) signal if $E[r(t)]=0$ and the autocorrelation function*

$E[r(t=t_1)\,r(t=t_2)] = ac_{rr}(t_1-t_2) = ac_{rr}(t=\tau_t)$, *i.e., the autocorrelation function depends only on the interval τ_t. $r(n)$ is zero mean WSS signal if $E[r(n)]=0$ and the autocorrelation function* $E[r(n=n_1)\,r(n=n_2)] = ac_{rr}(n_1-n_2) = ac_{rr}(n=\tau_n)$, *i.e., the autocorrelation function depends only on the interval τ_n.*

PSD: *Power spectral density of $e(t)$ is given by* $\Phi_{ee}(j\omega) = \int_{-\infty}^{\infty} ac_{ee}(t=\tau_t)\,e^{-j\omega\tau_t}\,d\tau_t$,

$-\infty \le \omega \le \infty$, *and PSD for $e(n)$ is* $\Phi_{ee}(z=e^{j\Omega}) = \sum_{\tau_n=-\infty}^{\infty} ac_{ee}(n)\;e^{-j\Omega n}, \quad -\pi \le \Omega \le \pi$.

Unity Variance Signal: *Signal $r(t)$ or $r(n)$ are unity variance if* $ac_{rr}(\,t=\tau_t=0\,)=1$ *or* $ac_{rr}(\,n=\tau_n=0\,)=1$.

SNR: *Signal-to-noise ratio. If $s(n)$ is the input signal and $w(n)$ is the noise then*

$$SNR = 10\log_{10}\left(\frac{E\left[s^2(n)\right]}{E\left[w^2(n)\right]}\right).$$

Zero mean WSS White Noise: *$r(t)$ is continuous time zero mean WSS white noise if* $E[r(t)] = 0$, $ac_{rr}(t=\tau_t)=0$ *for* $\tau_t \neq 0$, *and* $ac_{rr}(t=\tau_t=0) = \delta_{t0}\,\upsilon^2$. *Similarly, $r(n)$ is discrete time zero mean WSS white noise if* $E[r(n)] = 0$, $ac_{rr}(n=\tau_n)=0$ *for* $\tau_n \neq 0$, *and* $ac_{rr}(n=\tau_n=0) = \upsilon^2$. *Where υ^2 is the variance of white noise.*

Uncorrelated Signals: *$s(n)$ and $w(n)$ are uncorrelated signals if* $E[s(n)\,w(n)] = E[s(n)]\,E[w(n)]$. *If $s(n)$ and $w(n)$ are zero mean then uncorrelated signals* $\Rightarrow E\left[s(n)\,w(n)\right]=0$. *Similarly, if $s(t)$ and $w(t)$ are uncorrelated signals with zero mean then* $E\left[s(t)\,w(t)\right]=0$.

CHAPTER 1

INTRODUCTION

Design of optimal linear filters, predictors and state estimators is required in various signal processing and communications applications. In the past Kalman and Wiener filters have been used extensively for the design of optimal filters. These filters are based on l_2 (minimum variance) estimation, i.e., minimizing $E[e^2(n)]$, where $e(n)$ is the error between the actual signal and its estimate at time n and E is expectation operator. These filters minimize the variance or the power of the error signal at every point in time. The optimality of Wiener filters is based on exact knowledge of the input signal and noise power spectral densities (PSD). The performance of these filters will degrade if the input noise statistics or signal-to-noise ratio (SNR) is changing with time and is not known *a priori* [21]. In many applications there is no exact knowledge of the input signals, noise, or SNR's. One solution to this is to use filters which adapt to changing input signals and noise statistics. Often, convergence speed determines the performance as it is assumed that the convergence speed is fast enough to track the changes in the input signal and noise statistics. On average, one can expect variation in output error power. This variation will be large if the input signal and noise statistics are changing faster than the convergence speed of the adaptive algorithm. Another approach to overcome unknown input signals (noise in case of filtering problem) and model uncertainties is to use mini-max estimation. This will lead to a conservative (minimize the worst case input signals) design that is more robust to the unknown input noise. Performance of these filters will be conservative with less variation in the output error than the Wiener filter.

The motivation behind this thesis work is to study minimization criteria which provides inherent robustness with respect to the unknown noise and model uncertainties and use it for filtering applications encountered in signal processing and communications. Recently there is some work reported in the area of H_∞ minimization which provide robust stability and performance for control system applications [3][6][7][30]. The idea is extended to the filtering applications where optimal filter is computed by minimizing H_∞ norm of the error. It is reported [5][12][19][25] that H_∞ filters should be better suited to applications with the unknown input noise and model uncertainties. This provided the motivation to study the usefulness of H_∞ filters for signal processing and communications applications and compare it with the well known least squares based Wiener filters.

Let $\mathbf{H}(s)$ and $\mathbf{H}(z)$ be multiple input-output continuous and discrete time stable transfer functions respectively as shown in Figure 1. $\{I_k(t)\}_{k=1}^{p}$ or $\{I_k(n)\}_{k=1}^{p}$ are the p inputs and $\{O_k(t)\}_{k=1}^{m}$ or $\{O_k(n)\}_{k=1}^{m}$ are the m outputs of the continuous or discrete time transfer function $\mathbf{H}(s)$ or $\mathbf{H}(z)$ respectively.

The H_∞ norm is defined as:

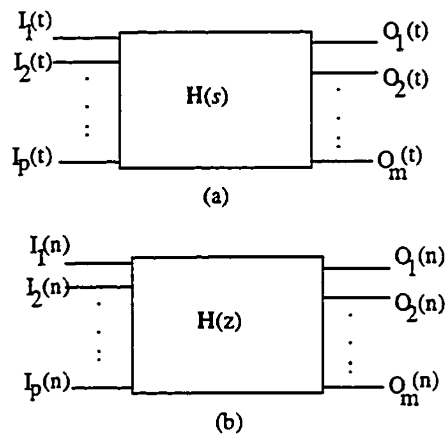

Figure 1: (a) Multiple p input and m output continuous transfer function. (b) Multiple p inputs and m outputs discrete time transfer function.

$$\|\mathbf{H}(s)\|_{\infty}^{2} = \sup_{\omega \in R^1} \sigma^2 \ [\mathbf{H}(j\omega)] = \sup_{\omega \in R^1} \lambda \left[\mathbf{H}(j\omega)\mathbf{H}^*(j\omega)\right] \tag{1}$$

$$\|\mathbf{H}(z)\|_{\infty}^{2} = \sup_{-\pi \le \Omega \le \pi} \sigma^2 \ \left[\mathbf{H}(e^{j\Omega})\right] = \sup_{-\pi \le \Omega \le \pi} \lambda \left[\mathbf{H}(e^{j\Omega})\mathbf{H}^*(e^{j\Omega})\right] \tag{2}$$

Where σ [.] and λ[.] are the largest singular and eigenvalues of the matrix respectively. Where $[.]^*$ is the conjugate transpose of [.]. For stable transfer functions H_∞ norm always exists [6]. The Problem of minimizing error using H_∞ criterion is equivalent to minimizing H_∞ norm of the transfer function between input signals to output error signals [24]. In case of Wiener filters l_2 norm of the transfer function is minimized. l_2 norm for $\mathbf{H}(z)$ or $\mathbf{H}(s)$ is defined as:

$$\|\mathbf{H}(z)\|_2^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} tr \ \left[\mathbf{H}(e^{j\Omega}) \ \mathbf{H}^*(e^{j\Omega})\right] \ d\Omega.$$

$$\|\mathbf{H}(s)\|_2^2 = \frac{1}{2\pi}\int_{-\infty}^{\infty} tr \ \left[\mathbf{H}(j\omega) \ \mathbf{H}^*(j\omega)\right] \ d\omega. \tag{3}$$

Where tr [.] is the trace of a matrix [.]. H_∞ filters in the time domain use the H_∞ performance criteria to minimize the worst case output error with respect to the input signals present in the system. In the frequency domain, the H_∞ criterion minimizes the maximum singular value of the error PSD over a specified frequency range. From (2) it is seen that in case of scalar problem (multiple input-single output) with inputs as white noises it minimizes the maximum value of the error PSD over all frequencies [9]. The criterion has the inherent property of being robust to the variation of the input signal and noise in the system [12]. The optimal solution forces the error to be white noise [9]. For scalar case let $\Phi_{ee}(s)$ be the PSD of the error and $W(s)$ be the weighting transfer function to shape the error PSD. Therefore, from white noise property [9](γ_{opt} is a constant):

$$W(s)\Phi_{ee}(s)W^*(s) = \gamma_{opt}^2 \;\Rightarrow \Phi_{ee}(s) = \frac{\gamma_{opt}^2}{W(s)W^*(s)}. \tag{4}$$

Where γ_{opt}^2 is the error variance when optimal solution is achieved. From (4) the error in a prescribed frequency band can be minimized by using an appropriate weighting function. Also, conservative filters can be designed to account for system and noise uncertainties using H_∞ criterion (see chapter 5).

This dissertation discusses the characteristics of H_∞ filters and performance benefits for applications encountered in signal processing and communications applications. The study of H_∞ filtering has lead to following contributions found in this dissertation:

* Most of the applications in signal processing and communications (SPC) assume zero mean wide sense stationary signals. There is a need to define the vector space for wide sense stationary input signals to give time and frequency domain representation of H_∞ minimization criteria. Most of the work reported assumes a deterministic signal and therefore this issue was not addressed properly. This thesis is able to provide

the stochastic interpretation of H_∞ filters which helps in understanding performance benefits of stochastic H_∞ filtering. Various performance advantages over minimum error variance filters are interpreted (see Chapter 3).

* In the applications area, a signal estimation problem of a Code Division Multiple Access (CDMA) system is studied. It is observed that mean square error (MSE) performance of H_∞ filters show robust characteristics. The performance of H_∞ filters is found to be better at low SNR when compared with minimum error variance filters. This performance benefit comes from the fact that H_∞ filters try to minimize the maximum eigenvalue of the matrix which maps input signals to the output error. This eigenvalue is an upper bound to the error variance. Therefore, it gives better performance when SNR is low (worst case). See Chapter 4 and 5 for details on this.

* Recently, it is shown that the well known Least-Mean-Square (LMS) algorithm minimizes the H_∞ criteria [11]. It is argued that the robustness of LMS algorithm compared to recursive least-squares (RLS), which satisfies least squares solution, comes from the fact that LMS satisfies H_∞ criteria. The advantage of RLS is its fast convergence compared to LMS or Normalized Least-Mean-Square (NLMS) but have poor robustness properties. Using the connection between H_∞ theory and LMS it is shown in this dissertation that all the three algorithms (LMS, RLS, NLMS) are connected by a Kalman

gain which estimates the gradient of the error in the direction of minimum error. From this it is shown that NLMS achieves the RLS solution when H_∞ upper bound is made large. Therefore, the upper bound controls the robustness of the algorithm and hence can be used to trade off robustness with convergence speed. This generates new class of algorithms which is named as sub-optimal NLMS (see Chapter 6 for details). Acoustic Echo Cancellation (AEC) example is taken to show that NLMS and RLS trade off robustness with performance.

For both signal estimation and adaptive filter applications it is observed that H_∞ filters have better robust performance (achieve minimum mean square error (MSE)) with respect to input noises as compared to minimum variance filters. However, the performance of minimum variance filters is better when the SNR is high or exact knowledge of input signal PSD is known. Therefore, there is a trade off between robustness and performance when using H_∞ and l_2 type filters respectively.

CHAPTER 2

SCIENTIFIC BACKGROUND

In the last few years, the H_∞ control problem has received considerable attention [1][3][30]. H_∞ criterion is used to compute controllers which provide robust stability and performance. Most of the control problems assume finite energy input signals which lead to early work on H_∞ minimization with finite energy signals. The finite energy signals taken were square summable. For example if $r(t)$ belongs to finite energy signal then it should satisfy:

$$\int_{-\infty}^{\infty} r(t)\ r^*(t)\ dt < \infty$$

For the finite energy input signal case, H_∞ minimization is a mini-max problem where the maximum energy in the error over all input signals is minimized. In the frequency domain it minimizes the maximum singular value of the transfer function from input signals to output error signal over all frequencies.

The concept of H_∞ norm was introduced and studied by Zames [30]. Later the concept was developed for single-input single-output systems by Zames and Francis [7]. The motivating factor for using H_∞ norm was the realization that classical Wiener or Kalman (WK) theory is concerned with a different category of mathematical problems. In a typical WK problem, the quadratic norm of the response to a interference or noise

w ($w(n)$ or $w(t)$) is minimized. In a deterministic version the frequency spectrum of w is known, in stochastic versions, w belongs to a single random process with known covariance properties. However, there are many practical problems for which w belongs to a class of random processes whose covariances are uncertain but belong to a prescribed set. For example, in audio design, w is often one of a set of narrow-band signals in the 20-20K Hz interval, as opposed to a single, wide-band signal in the same interval. Problems involving such more general interference sets are not tractable by WK [21]. It was shown in [30] that the H_∞ norm, as opposed to l_2 norm of the WK approach, is ideal for handling uncertainty in the system's model frequency response or in the frequency spectrum of the exogenous signals.

Subsequently, the H_∞ criterion was found to be useful for analyzing robust stability of closed loop systems with unknown disturbances [4][6]. It was used to find parametric controllers which achieve robust stability for a control system with uncertainty belong to certain class.

H_∞ optimality criterion for the filtering application was studied by some researchers. The design of optimal linear filters and predictors using H_∞ criterion was first considered by Grimble and Sayed [10]. The motivation behind their work was to keep the estimation error spectrum small over a range of frequencies. It was concluded that by introducing frequency weighting, the error in particular frequency ranges can be reduced to lower values than is possible with a minimum variance type of filter. Moreover, with an H_∞ filter the shape of the error spectrum is completely determined *a priori* by the inverse of the weighting function, whatever the signal and noise model descriptions. Finally, it was reported that the H_∞ filter is suitable for applications which require the error spectrum to be made uniformly small, or the error in a particular frequency range to be made as low as possible.

Shaked [24] studied the H_∞ filtering problem in a frequency domain setting for stochastic continuous time signals. He concluded that the estimation of a signal embedded in white noise is better (subjective visual comparison of the estimated signal) using an H_∞ than a Wiener filter. Later he provided a state space solution for continuous time H_∞ filters under some restrictive conditions [25].

In another paper by Nagpal and Khargonekar [19], filtering and smoothing was considered in an H_∞ setting for finite energy continuous time case. They derived necessary and sufficient conditions for the existence of estimators (both filters and smothers) that achieve a prescribed H_∞ performance bound for situations when the plant and measurement noises have uncertainties.

In a recent paper by Shaked [25], properties of the minimum H_∞ norm filtering estimation error have been investigated. The relation between the optimal estimator and the equalizing solution (error is white noise) to the standard H_∞ minimization problem is discussed. The optimal estimation method is applied in the multivariable deconvolution problem. A simple deconvolution filter of minimum order is obtained which minimizes the H_∞ norm of the deconvolution error. It was stated that the H_∞ methods of optimal estimation and deconvolution are useful in cases where the statistics of the noise signals are not known completely, or in cases where it is required to minimize the maximum singular value of the estimation, or the deconvolution, error spectrum.

In all of the above work, the H_∞ filters are derived in frequency domain (by assuming finite energy or stochastic input signals) or in time domain (by assuming finite energy input signals). Recent attempts have been made towards the time domain stochastic interpretation of these filters assuming the inputs as WSS [9][27]. Also, in [5][23] it was claimed that the H_∞ criterion was more suited to cancel Inter Symbol Interference (ISI) compared to Wiener filter.

In a recent work by Hassibi et al. [11], it was shown that LMS and NLMS adaptive algorithms are H_∞ optimal and therefore are more robust to the finite precision effects (intermediate variables in the algorithm accumulate error when computed on finite precision machines) and input noises (achieve lower MSE when the filter coefficients are updated in the presence of noise) compared to RLS which is l_2 (minimum variance) optimal.

In summary, it is been reported that H_∞ filters are useful in the following situations: the estimation error is required to be small in a specific frequency band, the error spectrum is to be made uniformly small for colored noise, the peak of the error spectrum is to be minimized, or the input signals have unknown statistics.

CHAPTER 3

CHARACTERISTICS OF H_∞ FILTERS

This section will discuss mathematical formulation of H_∞ optimality criterion and look into its main characteristics. The main emphasis is on the stochastic interpretation of H_∞ filters. Most of the earlier work done in H_∞ filtering area assumes input signals to be finite energy in nature. We provide mathematical frame work to define H_∞ filters when input signals involved are wide sense stationary signals. From this various advantages of H_∞ filter over Wiener filter are identified.

3.1 Stochastic H_∞ Filters

Consider the general estimation problem of Figure 2 encountered in most of the signal processing and communications problems. $G(z)$ and $N(z)$ are the signal and noise models respectively. $C(z)$ is the channel transfer function which corrupts the signal and $W(z)$ is the weighting function that weights the errors. $H(z)$ is the optimal filter obtained by minimizing the desired optimality criterion. Inputs to the system are random signals with zero mean and second order stationarity. Let $\mathbf{T}(z)$ be the transfer function from inputs ($I(n) = [s(n) \quad w(n)]$) and output ($e_1(n)$). Then

$$\mathbf{T}_1(z) = W(z)\left[G(z) - G(z)C(z)H(z) \quad - H(z)N(z)\right] = W(z)\mathbf{T}(z) \tag{5}$$

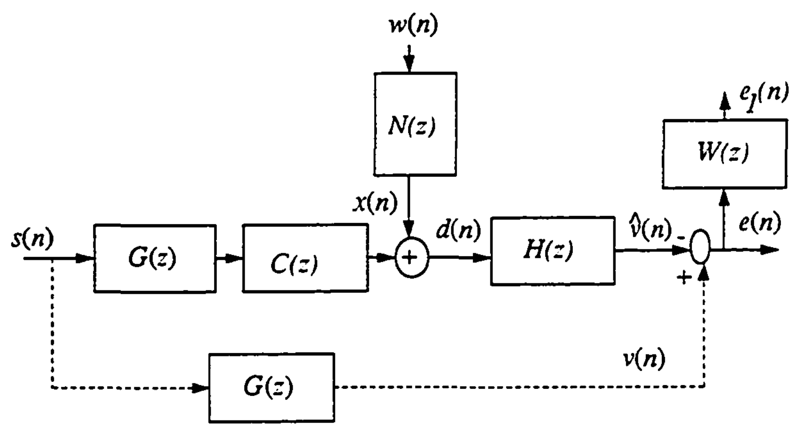

Figure 2: Signal estimation problem for discrete time system.

Since $\mathbf{T}_1(z)\ \mathbf{T}_1^*(z)$ ($[.]^*$ represents the conjugate transpose of $[.]$) is scalar (single input-output) the H_∞ norm from input to output (from (2)) will be [24]:

$$\|\mathbf{T}_1(z)\|_\infty^2 = \|W(z)\mathbf{T}(z)\|_\infty^2 = \sup_\Omega \left|W(e^{j\Omega})\mathbf{T}(e^{j\Omega})\ \mathbf{T}^*(e^{j\Omega})W^*(e^{j\Omega})\right| \qquad (6)$$

If input and output signals are vectors then the transfer function $\mathbf{T}_1(z)$ will be a matrix. Since we are considering signal estimation problem of Figure 2 the analysis will focus on the scalar H_∞ norm of (6). In order to see H_∞ norm in time domain let us define input signals belong to a two-dimensional ρ_2 vector space. Where $\rho_2 = \{[s(n) \quad w(n)]\ /\ s(n), w(n) \in WSS\}$. Inner product and norm on ρ_2 is defined as (assuming $s(n)$ and $w(n)$ are uncorrelated):

$$\langle [s_1(n) \quad w_1(n)], [s_2(n) \quad w_2(n)]^* \rangle = E\left[s_1(n)\ s_2^*(n)\right] + E\left[w_1(n) w_2^*(n)\right] \qquad (7)$$

$$\|[s_1(n) \quad w_1(n)]\|_2^2 = E\left[s_1^2(n) + w_1^2(n)\right]. \qquad (8)$$

The vector space with inner product and norm is a Hilbert space. Therefore, ρ_2 is a Hilbert space. PSD of the output signal ($\Phi_{e_1e_1}(z)$) is given by:

$\Phi_{e_1e_1}(z) = |\mathbf{T}_1(z)|^2\, \Phi_{II}(z) \qquad ;\ |[.]|^2 = [.]\,[.]^*$

where $\Phi_{II}(z) = \Phi_{ss}(z) + \Phi_{ww}(z)$ is the PSD of the input signals. Also,

$\|e_1(n)\|_2^2 = E\left[e_1^2(n)\right] = \frac{1}{2\pi}\int_{-\pi}^{+\pi} \left|\mathbf{T}_1(e^{j\Omega})\right|^2 \Phi_{II}(e^{j\Omega}) d\Omega$

$\leq \|\mathbf{T}_1(z)\|_\infty^2 \frac{1}{2\pi}\int_{-\pi}^{+\pi} \Phi_{II}(e^{j\Omega}) d\Omega = \|\mathbf{T}_1(z)\|_\infty^2 E\left[I(n) I^*(n)\right] =$

$\|\mathbf{T}_1(z)\|_\infty^2 E\left[s^2(n) + w^2(n)\right]$

$= \|\mathbf{T}_1(z)\|_\infty^2 \, \|I(n)\|_2^2 \; ; \; \|I(n)\|_2^2 = E\,[s^2(n) + w^2(n)]$.

Therefore,

$$\frac{\|LI(n)\|_2^2}{\|I(n)\|_2^2} = \frac{\|e_1(n)\|_2^2}{\|I(n)\|_2^2} \leq \|\mathbf{T}_1(z)\|_\infty^2 \tag{9}$$

where L is the linear time domain operator which maps the input signal $I(n)$ to the output signal $e_1(n)$. To show that $\|\mathbf{T}_1(z)\|_\infty^2$ is the least upper bound of $\frac{\|LI(n)\|_2^2}{\|I(n)\|_2^2}$ consider input signal $I_1(n)$ such that

$$\Phi_{I_1 I_1}(e^{j\Omega}) = 2\pi\delta_{\Omega\Omega_0} \quad ; \|I_1(n)\|_2^2 = 1; \; I_1(n) \in \rho_2 \tag{10}$$

where Ω_0 is the frequency such that $\|\mathbf{T}_1(z)\|_\infty^2 = \mathbf{T}_1(z = e^{j\Omega_0})\mathbf{T}_1^*(z = e^{j\Omega_0})$. Therefore,

$$E\left[e_1^2(n)\right] = \frac{1}{2\pi}\int_{-\pi}^{+\pi} \left|\mathbf{T}_1(e^{j\Omega})\right|^2 \; 2\pi \; \delta_{\Omega\Omega_0} \; d\Omega = \|\mathbf{T}_1(z)\|_\infty^2 \tag{11}$$

since $\|I_1(n)\|_2^2 = 1$, therefore,

$$\frac{\|LI_1(n)\|_2^2}{\|I_1(n)\|_2^2} = \frac{\|e_1(n)\|_2^2}{\|I_1(n)\|_2^2} = \|\mathbf{T}_1(z)\|_\infty^2. \tag{12}$$

from (9) and (12)

$$\|L\|_2 = \sup_{\|I(n)\|_2^2 \neq 0} \frac{\|LI(n)\|_2^2}{\|I(n)\|_2^2} = \sup_{\|I(n)\|_2^2 \neq 0} \frac{\|e_1(n)\|_2^2}{\|I(n)\|_2^2} = \|\mathbf{T}_1(z)\|_\infty^2 \tag{13}$$

where $\|L\|_2$ is called induced second norm of operator L and sup is supremum over all $I(n)$. Also, it is easy to see

$$\sup_{\|I(n)\|_2 \neq 0} \frac{\|e_1(n)\|_2^2}{\|I(n)\|_2^2} = \sup_{\|I(n)\|_2^2 \leq 1} \|e_1(n)\|_2^2 = \sup_{\|I(n)\|_2^2 = 1} \|e_1(n)\|_2^2 \tag{14}$$

From (14) this means that if the H_∞ norm of the transfer function between the input signals and the error is minimized then the time domain solution is

$$\min_{H(z)} \sup_{\|I(n)\|_2^2 \neq 0} \frac{\|e_1(n)\|_2^2}{\|I(n)\|_2^2} = \min_{H(z)} \sup_{\|I(n)\|_2^2 \leq 1} \|e_1(n)\|_2^2 = \min_{H(z)} \sup_{\|I(n)\|_2^2 = 1} \|e_1(n)\|_2^2 \tag{15}$$

and the frequency domain solution is [24]

$$\min_{H(z)} \sup_{\Omega} \left| W(e^{j\Omega})\mathbf{T}(e^{j\Omega})\ \mathbf{T}^*(e^{j\Omega})W^*(e^{j\Omega}) \right|. \tag{16}$$

$H(z)$ is the optimal filter which minimizes the error in H_∞ norm sense. Following the above methodology one can derive a similar relationship [9][27] between time domain and frequency domain H_∞ minimization criteria for continuous time system of Figure 3

$$\min_{H(s)} \sup_{\|I(t)\|_2^2 \neq 0} \frac{\|e_1(t)\|_2^2}{\|I(t)\|_2^2} = \min_{H(s)} \sup_{\omega} \left| W(j\omega)\mathbf{T}(j\omega)\ \mathbf{T}^*(j\omega)W^*(j\omega) \right|. \tag{17}$$

3.2 H_∞ Filters for Finite Energy Signals

Similar results as derived above for stochastic signals are shown for finite energy continuous time signals in [19], i.e.,

$$\min_{H(s)} \sup_{\|I(t)\|_2^2 \neq 0} \frac{\|e_1(t)\|_{2,[0,\infty]}}{\|I(t)\|_{2,[0,\infty]}} = \min_{H(s)} \sup_{\omega} \left| W(j\omega)\mathbf{T}(j\omega)\ \mathbf{T}^*(j\omega)W^*(j\omega) \right| \tag{18}$$

where $I(t)$ and $e_1(t)$ belong to space of finite energy signals with inner product and norm defined as:

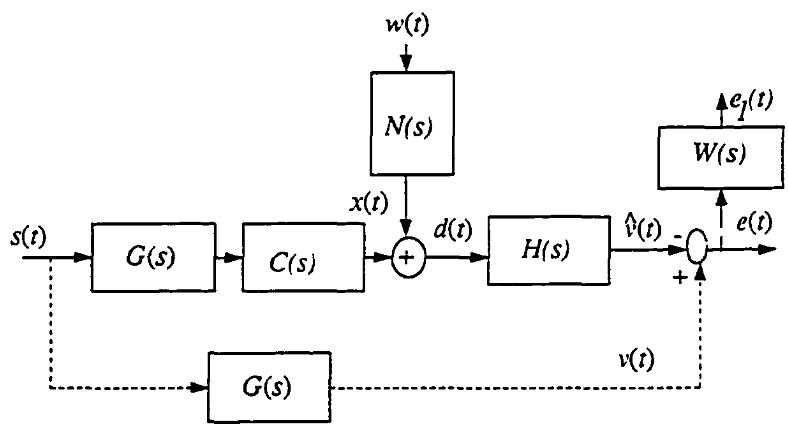

Figure 3: Signal estimation problem for continuos time system.

$\langle f(t), g(t) \rangle = \int_0^\infty f(t) g^*(t) dt$

$\|f(t)\|_{2,[0,\infty]} = \int_0^\infty f(t) f^*(t) dt$.

From above results it is clear that the optimal filter obtained by minimizing H_∞ criteria will be same for both finite energy and stochastic signals as the minimization depends on the transfer function from inputs to the output. However, the interpretation of the results in time domain will be different as minimization is dependent on input signals (vector space of the input signals are different for finite energy and stochastic signals).

3.3 Advantages of H_∞ Filters

The preceding mathematical formulation can have various interpretations leading to advantages to the H_∞ minimization. Following are some of the unique features of H_∞ minimization:

* The optimal solution (provided it exists) to the minimization problem of (13) is a constant [9]. Let this constant be γ_{opt}^2, then the optimal $H(z)$ will yield:

$$|W(z)\mathbf{T}(z)\ \mathbf{T}^*(z)W^*(z)| = \gamma_{opt}^2 \tag{19}$$

If inputs to the system are white noises with unity variances, then the PSD of the error

$(e_1(n))$, $\Phi_{e_1e_1}(z)$, is given by [10]:

$$\Phi_{e_1e_1}(z) = W(z)\Phi_{ee}(z)W^*(z) = |W(z)\mathbf{T}(z)\ \mathbf{T}^*(z)W^*(z)| = \gamma^2_{opt} \tag{20}$$

This shows that for scalar case the PSD of the error $e(n)$ ($\Phi_{ee}(z)$), is given by the inverse of the weighting function, no matter what the signal and noise models are used to compute the optimal filter. In other words the H_∞ criterion provides a whitening filter from the input signals to the output error provided the inputs are white noises of unit variance. This is often the assumption on inputs when considering signal and noise models. One advantage of this is that if the requirement exists to minimize the error in specific frequency bands, then it can be done exactly using H_∞ criterion. In case of Wiener filter we cannot guarantee output error to be white. Therefore, minimization of error in specific frequency bands cannot be done exactly. In case of sub-optimal solution:

$$\Phi_{e_1e_1}(z) = W(z)\Phi_{ee}(z)W^*(z) = |W(z)\mathbf{T}(z)\ \mathbf{T}^*(z)W^*(z)| < \gamma^2 \tag{21}$$

$$\Rightarrow \frac{\|e_1(n)\|_2^2}{\|I(n)\|_2^2} < \gamma^2 \tag{22}$$

where γ^2 is a constant.

* From (13) it can be interpreted that in the time domain optimal H_∞ filter looks for the worst case input random signal and minimizes over $H(z)$ assuming these worst case

input signal variations. This says that H_∞ filters are robust to input signal variations. When exact knowledge of SNR is not known or signal statistics are changing with time then these filters should provide a better performance compared to Wiener filters.

- From (22) if the input signals belong to space of stationary random signals with $\| I(n) \|_2 \leq 1$ then γ^2 places an upper bound to the error variance:

$$\|e_1(n)\|_2^2 < \gamma^2 \quad ; \| I(n) \|_2 \leq 1 \tag{23}$$

  This suggests that it also provides robust behavior with respect to minimum variance criterion under the restrictive class of input signals.

3.3.1 Remarks

For finite energy input signals the H_∞ norm is the maximum energy gain from inputs to the outputs. It guarantees the smallest estimation error energy over all possible input signals of fixed energy. H_∞ filters are thus overly conservative, which reflects in the better robust behavior to noise variation (see applications section for this point where comparison is made between LMS and RLS algorithms). Consider the estimation problem in state space

$$\begin{aligned} X(n+1) &= \mathbf{A}X(n) + \mathbf{B}U(n) \\ Y(n) &= \mathbf{C}X(n) + \mathbf{D}U(n) \\ O(n) &= \mathbf{L}X(n) \end{aligned} \tag{24}$$

where the output signal is given by $O(n) = \mathbf{L}X(n)$, $X(n)$ are the system states, and $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ are constant matrices. H_∞ estimators are observed to be dependent on $\mathbf{L}$ unlike Kalman filter estimator which is dependent on $\mathbf{C}$ [13][25]. Also, it is shown for both continuous [19][25] and discrete time cases [12] that $\gamma \to \infty$ gives the minimum variance solution, i.e., if we pose the minimization problem as:

$$\begin{aligned} \min_{H(z)} \sup_{\|I(n)\|_2^2 \neq 0} \frac{\|e_1(n)\|_2^2}{\|I(n)\|_2^2} \\ subject\ to\ :\ \frac{\|e_1(n)\|_2^2}{\|I(n)\|_2^2} < \gamma^2 \end{aligned} \tag{25}$$

then $\gamma \to \infty$ will converge to minimum variance solution. This suggests that the H_∞ norm of the minimum variance filters may be quite large, indicating that it may have poor robustness properties.

CHAPTER 4

MATHEMATICAL SOLUTION

As with any other mini-max design, the solution to H_∞ is more complex compared to minimum variance filters. Unlike the l_2 norm H_∞ does not have closed form solution. H_∞ norm is computed through iterative procedure (see section 4.1). The solution to the H_∞ problem was earlier used to compute H_∞ controllers. Most of the earlier solutions were available in control problem framework. Among these the classical approach to solving H_∞ control problem has been via analytic functions (Nevanlinna-Pick interpolation) or operator-theoretic methods [7], and frequency domain solution using model matching and Hankel approximation methods [6][16]. Then a procedure was designed requiring the solution of a Riccati equation [20]. This has led to the development of the "standard problem formulation" [3]. Recently some solutions are available using game theoretic approaches [1][26].

The first attempt to solve H_∞ filtering problem is made by Grimble [10] using polynomial systems approach in frequency domain for discrete time systems. Shaked have shown the solution for continuous time case in frequency domain setting [23][24]. Model matching solution for continuous time case is shown in [27] and solution using game theoretic approach for both continuous and discrete time systems is discussed in [1].

Recent contribution is by Babak et. al. [13] where discrete time solution is shown under krein space setting. It is reported that H_∞ filtering is Kalman filtering in krein space. Given a deterministic quadratic form in krein space, one can relate it to a corresponding stochastic problem for which the Kalman filter solution can be computed. Moreover, the condition for a minimum can also be expressed in terms of quantities easily related to the basic Riccati equations of the Kalman filter. All approaches require minimum of solving one Riccati equation to check the condition if a filter can be obtained so that $\|\mathbf{T}_1(z)\|_\infty < 1$.

This section will cover the methodology used to compute H_∞ filters available in literature, problems associated with the current solutions, and FIR filter design which minimize H_∞ criteria useful for applications in signal processing and communications.

4.1 H_∞ Norm of Continuous Transfer Function

Consider finding H_∞ norm of a scalar transfer function $G(s)$ represented with state space quantities A, B, C, and d as

$$G(s) = C\left(s\mathrm{I} - \mathrm{A}^{-1}B\right) + d \tag{26}$$

assume $G(s)$ is strictly proper, i.e., $d = 0$ (degree of the numerator is less than degree of the denominator) then following conditions are equivalent [3]

a. $\|G(s)\|_\infty < \gamma \; ; \gamma > 0$

b. $\mathbf{H}$ has no eigenvalues on the imaginary axis. Where $\mathbf{H} = \begin{bmatrix} \mathbf{A} & \gamma^{-2}BB^t \\ -C^tC & -\mathbf{A}^t \end{bmatrix}$ and $[.]^t$ is transpose of $[.]$. The H_∞ norm can be computed as follows: Select a positive number γ ; test if $\|G(s)\|_\infty < \gamma$ by calculating eigenvalues of $\mathbf{H}$; increase or decrease γ accordingly. Thus, H_∞ norm computation requires a search over γ, in contrast to minimum variance, which does not. If $d \neq 0$ then H_∞ norm will be

$$\left\|G'(s)\right\|_\infty = \|d + G(s)\|_\infty \leq d + \|G(s)\|_\infty < d + \gamma \tag{27}$$

where γ can be obtained for $G(s) = C\,(s\mathbf{I} - \mathbf{A})^{-1}\,B$.

4.2 H_∞ Norm of Discrete Transfer Function

Assume $G(z)$ is strictly proper. It can be shown [3] that if the transfer function given by

$$\frac{1}{\left[1 - \frac{1}{\gamma^2}G(z)\;G^*(z)\right]} \tag{28}$$

is stable (all poles inside unit circle) then $\left\|\frac{1}{\gamma}G(z)\right\|_\infty < 1$. H_∞ norm is the minimum value of γ for which $\left[1 - \frac{1}{\gamma^2}G(z)\;G^*(z)\right]^{-1}$ is stable. Therefore, a search over γ is done to obtain H_∞ minimum norm value as described for continuous time transfer function. In the case of minimum variance we can compute the l_2 norm in a closed form as:

$$\|G(z)\|_2^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} G(e^{j\Omega})\ G^*(e^{j\Omega}) d\Omega\ . \tag{29}$$

We can see complexity involved with computing H_∞ norm. Similarly, for computing optimal filter a search is to be made over γ.

4.3 Frequency Domain Solutions

There are many solutions present to solve H_∞ optimization problem for control system problems in frequency domain. Most of the earlier solutions are based on the Model Matching approach described below. Earlier attempt made towards H_∞ filter solution for discrete time system is the polynomial system approach [9].

4.3.1 Model Matching Solution

4.3.1.1 Discrete Time

The basic idea behind solving H_∞ optimization problem in frequency domain under model matching solution is to reduce the transfer function from input to output signals to Hankel approximation or Nehari extension problem. For example $\| \mathbf{T}_1(z) \|_\infty$ can be represented as [27]:

$$\|\mathbf{T}_{11}(z) - \mathbf{T}_{21}(z)H(z)\|_\infty = \|W(z)\,[G(z) \quad 0] - W(z)\,[C(z)G(z) \quad N(z)]\,H(z)\|_\infty \tag{30}$$

where $\mathbf{T}_{11}(z)$ and $\mathbf{T}_{21}(z)$ are stable transfer functions. Through series of operations (30) can be converted to $\|\mathbf{R}_-(z) + \mathbf{R}(z)\|_\infty$ [16]. Where $\mathbf{R}_-(z)$ describes a given anticausal linear time-invariant discrete time system and $\mathbf{R}(z)$ is a causal linear time-invariant system. $\mathbf{R}_-(z)$ is a function of $\mathbf{T}_{11}(z)$ and $\mathbf{T}_{21}(z)$ while $\mathbf{R}(z)$ is a function of $H(z)$. In order to obtain optimal $H(z)$ we have to solve:

$$\min_{H(z)} \|\mathbf{R}_-(z) + \mathbf{R}(z)\|_\infty . \tag{31}$$

Alternate to (31) we can solve

$$\left\| \frac{1}{\gamma} \left(\mathbf{R}_-(z) + \mathbf{R}(z)\right) \right\|_\infty < 1 \tag{32}$$

for minimum value of γ. So the problem is reduced to approximating a causal transfer function $\mathbf{R}(z)$ with anticausal transfer function $\mathbf{R}_-(z)$. Let $\mathbf{R}_-(z) = \mathbf{D} + \mathbf{C}\,(z\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}$ be the transfer matrix of an anticausal discrete time system and let $\mathbf{R}(z) = \mathbf{D}_1 + \mathbf{C}_1\,(z\mathbf{I} - \mathbf{A}_1)^{-1}\mathbf{B}_1$ be the transfer matrix of an causal discrete time system. If there exists an $\mathbf{R}(z)$ satisfying (32) for some value of γ then $\mathbf{A}_1$, $\mathbf{B}_1$, $\mathbf{C}_1$, and $\mathbf{D}_1$ are given by:

$$\mathbf{A}_1 = \left[\mathbf{A}^* + \mathbf{D}^{-1}\,(\mathbf{I} - \mathbf{A})^{-1}\,\mathbf{B}\mathbf{B}^*\right]\left[\mathbf{I} + \mathbf{D}^{-1}\,(\mathbf{I} - \mathbf{A})^{-1}\,\mathbf{B}\mathbf{B}^*\right]^{-1}$$

$\mathbf{B}_1 = (\mathbf{I}-\mathbf{A}^*)\left[\mathbf{I}+\mathbf{D}^{-1}(\mathbf{I}-\mathbf{A})^{-1}\mathbf{B}\mathbf{B}^*\right]^{-1}\mathbf{D}^{-1}(\mathbf{I}-\mathbf{A})^{-1}\mathbf{B}$

$\mathbf{C}_1 = \mathbf{C}(\mathbf{I}-\mathbf{A})^{-1}\mathbf{P}(\mathbf{I}-\mathbf{A}^*)\left[\mathbf{I}+\mathbf{D}^{-1}(\mathbf{I}-\mathbf{A})^{-1}\mathbf{B}\mathbf{B}^*\right]^{-1}$

$\mathbf{D}_1 = \mathbf{C}(\mathbf{I}-\mathbf{A})^{-1}\mathbf{B} - \mathbf{C}(\mathbf{I}-\mathbf{A})^{-1}\mathbf{P}(\mathbf{I}-\mathbf{A}^*)$

$\left[\mathbf{I}+\mathbf{D}^{-1}(\mathbf{I}-\mathbf{A})^{-1}\mathbf{B}\mathbf{B}^*\right]^{-1}\mathbf{D}^{-1}(\mathbf{I}-\mathbf{A})^{-1}\mathbf{B}$. Where $\mathbf{P}$ and $\mathbf{Q}$ for the discrete time system $\mathbf{R}_-(z)$ are given by:

$\mathbf{P} = \mathbf{A}\mathbf{P}\mathbf{A}^* + \mathbf{B}\mathbf{B}^*$; $\mathbf{Q} = \mathbf{A}^*\mathbf{Q}\mathbf{A} + \mathbf{C}^*\mathbf{C}$.

4.3.1.2 Continuous Time

Let $\mathbf{T}_1(s)$ be the transfer function between input signals $w(t)$ and $s(t)$ and output signal $e_1(t)$, as shown in Figure 3. Then the transfer function can be written as

$$\mathbf{T}_1(s) = [\mathbf{T}_{11}(s) - \mathbf{T}_{21}(s)H(s)] \tag{33}$$

where $\mathbf{T}_{11}(s) = [G(s) \quad 0]\,W(s)$ and $\mathbf{T}_{21}(s) = [G(s)C(s) \quad N(s)]\,W(s)$. If the input signals are white noises with unit variances, then the PSD of the error $e_1(t)$ is given by [10]

$$\Phi_{e_1e_1}(s) \;=\; \mathbf{T}_1(s)\,\mathbf{T}_1^*(s). \tag{34}$$

The H_∞ optimization criterion implies minimization of the largest singular value of the error PSD matrix, which can be represented as [10]

$$\min_{H(s)} \sup_{\omega} |\Phi_{e_1e_1}(j\omega)| \;=\; \min_{H(s)} \|\Phi_{e_1e_1}(s)\|_\infty \;=\; \min_{H(s)} \|\mathbf{T}_1(s)\|_\infty \tag{35}$$

where sup is the supremum over all ω . The optimal filter is the solution to

$$\min_{H(s)} \|\mathbf{T}_{11}(s) - \mathbf{T}_{21}(s)\, H(s)\|_\infty \tag{36}$$

where $H(s) \in RH_\infty$. This can be solved using Nehari's theorem [6] to obtain sub-optimal H_∞ filter. Let (representing transfer function $[.]\,(s)$ with $[.]$)

$$\gamma_\infty = \inf(\gamma : \|\mathbf{T}_{11} - \mathbf{T}_{21} H\|_\infty \le \gamma) \tag{37}$$

Theorem 1 *[6]: (i)* $\gamma_\infty = \inf\,(\gamma : \|\mathbf{Y}\|_\infty < \gamma, dis(Q, RH_\infty) < 1)$ *. Q and $\mathbf{Y}$ represent the same quantity as in step 3 below. (ii) Suppose* $\gamma > \gamma_\infty$ *;* $H, X \in RH_\infty$*.* $\|Q - X\|_\infty < 1$ *and* $X = T_{2o} H Y_o^{-1}$ *then* $\|\mathbf{T}_{11} - \mathbf{T}_{21} H\|_\infty \le \gamma$*.*

From this theorem following steps can be followed to compute sub-optimal H_∞ filter using γ iteration.

Step 1: Get inner-outer factorization of transfer function $\mathbf{T}_{21}$ such that $\mathbf{T}_{21} = \mathbf{T}_{21i} T_{21o}$. $\mathbf{T}_{21i}$ is inner of $\mathbf{T}_{21}$ and T_{21o} is outer of $\mathbf{T}_{21}$.

Step 2: Define RL_∞ function $\mathbf{Y} = (I - \mathbf{T}_{21i}\mathbf{T}_{21i}^{*}\mathbf{T}_{11})$.

Step 3: If γ is a real number greater than $\|\mathbf{Y}\|_\infty$ then $[\gamma^2 - \mathbf{Y}^{*}\mathbf{Y}]$ has a special factor Y_o . Define $Q = \mathbf{T}_{21i}^{*}\mathbf{T}_{11} Y_o^{-1}$. $Q \in RL_\infty$.

Step 4: Select trial value for γ in interval $(\|\mathbf{Y}\|_\infty, \alpha_1)$, $\alpha_1 > \|\mathbf{Y}\|_\infty$.

Step 5: Compute hankel norm $\|\Gamma_Q\|$. $\|\Gamma_Q\| < 1 \Longrightarrow \gamma_\infty < \gamma$. Reduce value of γ and go to step 4. When a sufficiently accurate upper bound for γ_∞ is obtained go to step 6.

Step 6: Find matrix X in RH_∞ such that $\|Q - X\|_\infty < 1$.

Step 7: Solve $H_{sub-optimal} = T_{21o}^{-1} X Y_o$.

Frequency domain solutions can have numerical inaccuracies when order of the polynomials involved are large. We have experienced this problem while working with model matching solution and also been reported in [5].

4.4 Time Domain Solutions

4.4.1 Discrete Time

Time domain solutions are present under state space settings. Consider state space representation of transfer function $\mathbf{M}(z)$ from inputs ($s(n)$ and $w(n)$) to outputs ($d(n)$ and $v(n)$) as shown in Figure 4:

$$\begin{aligned} X(n+1) &= \mathbf{A}X(n) + B_1 s(n) + B_2 w(n) \\ d(n) &= C_2 X(n) + d_1 s(n) + d_2 w(n) \\ v(n) &= C_1 X(n) + d_o s(n) \end{aligned} \tag{38}$$

where $s(n), w(n)$ are the input signals to $\mathbf{M}(z)$, $d(n)$ and $v(n)$ are the observation and desired signal value as shown in the Figure 4. $G(z), N(z)$, and $C(z)$ are signal, noise, and channel models respectively. $W(z)$ is the whitening filter. $X(n)$ is the state vector and $\mathbf{A}, B_1, B_2, C_1, C_2, d_1, d_2, d_o$ are constant matrices, vectors, and scalars. To obtain a filter $H(z)$ represented in state space as

$$X_1(n+1) = \mathbf{A}_f X_1(n) + B_f d(n) \tag{39}$$

$$\hat{v}(n) = C_f X_1(n) + d_f d(n)$$

which minimizes H_∞ norm of the error $e(n) = v(n) - \hat{v}(n)$ requires solution (matrix $\mathbf{P}$) to the discrete time Riccati equation

$$\mathbf{G} - \mathbf{P} + \mathbf{A}^*\mathbf{P}\mathbf{A} = (\mathbf{F} + \mathbf{A}^*\mathbf{P}\mathbf{B})(\mathbf{R} + \mathbf{B}^*\mathbf{P}\mathbf{B})^{-1}(\mathbf{F}^* + \mathbf{B}^*\mathbf{P}\mathbf{A}) \tag{40}$$

where $\mathbf{A}_f, B_f, C_f, d_f$ in (39) are constant matrices obtained for optimal $H(z)$, $\mathbf{B} = [B_1 \;\; B_2]$, and $\hat{v}(n)$ is the estimate of $v(n)$. The matrices $\mathbf{G}, \mathbf{F}, \mathbf{R}$ are coefficients of the quadratic form :

$$\varsigma(L, U) = L^*(n)\mathbf{G}L(n) + 2L^*(n)\mathbf{F}L(n) + U^*(n)\mathbf{R}U(n) \tag{41}$$

where $L(n)$ and $U(n)$ are function of $s(n), w(n)$, and $X(n)$ [17]. The optimum H_∞ filter is given by [17]:

$\mathbf{A}_f = \mathbf{A} + B_f C_1$, $C_f = -C_2 + d_f C_1$

$B_f L - d_f f = \min_U \Delta_+ [\mathbf{P}] (L, U, f)$

where

$\Delta_+ [\mathbf{P}] (L, U, f) = \varsigma(L, U) + |\mathbf{A}^* L(n) + C_1^* f(n) + C_2^* f(n)|^2_{\mathbf{P}} - |L(n)|^2_{\mathbf{P}}$. Where $|[.]|_{\mathbf{P}} = [.]^* \mathbf{P} [.]$.

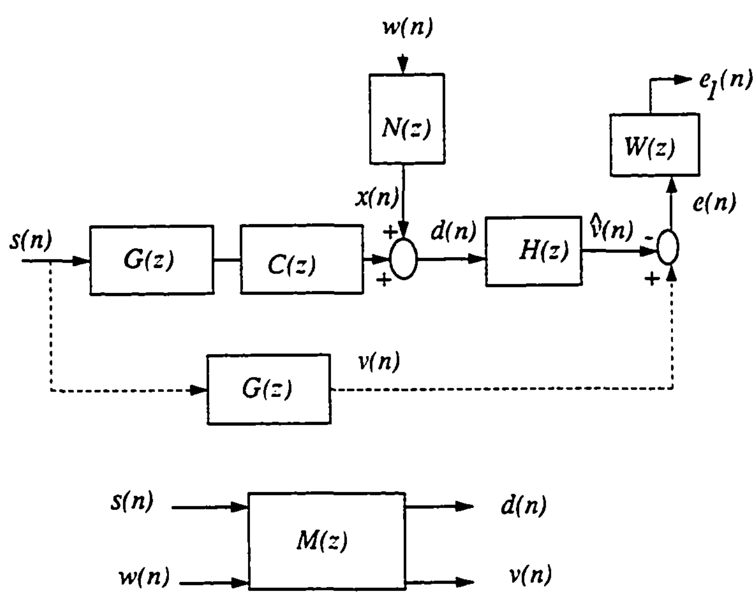

Figure 4: Signal estimation using filter $H(z)$ from observed signal $d(n)$.

4.4.2 Continuous Time

First we will solve a simple filtering problem (represented by (42)) and then extend the same idea to solve the general filtering problem (represented by (56)). This will show the problems associated with the solution to general filtering problem. Consider a filtering problem in state space form:

$$\begin{aligned} X'(t) &= \mathbf{A}X(t) + B_1 s(t) \\ d(t) &= C_2 X(t) + w(t) \\ v(t) &= C_1 X(t) \end{aligned} \tag{42}$$

where $s(t), w(t)$ are the finite energy input signals. $d(t)$ and $v(t)$ are the observation and desired signal value. $X(t)$ is the state vector, and $\mathbf{A}$,B_1, C_1,and C_2 are constant matrices. The idea is to estimate $v(t)$ as $\hat{v}(t)$ from observed signal $d(t)$ using a state space filter as:

$$\begin{aligned} X_1'(t) &= \mathbf{A}_f X_1(t) + B_f d(t) \\ \hat{v}(t) &= C_f X_1(t). \end{aligned} \tag{43}$$

We have to find optimal filter which achieves minimum H_∞ norm. First a filter is derived which keeps a unity upper bound on the H_∞ norm from input signals to the output error. Then through iteration minimum value of the H_∞ norm can be achieved as discussed in section 4.3. To find a filter which achieves H_∞ norm less than unity it should satisfy:

$$\frac{\int_{t_0}^{t} |v(t) - \hat{v}(t)|^2 \, dt}{\int_{t_0}^{t} |s(t)|^2 \, dt + \int_{t_0}^{t} |w(t)|^2 \, dt} < 1 \tag{44}$$

$$\Longrightarrow \int_{t_0}^{t} |s(t)|^2 + |w(t)|^2 - |v(t) - \hat{v}(t)|^2 \, dt > 0.$$

Let $t_0 = -\infty$ and $t = 0$. By changing the time axis the above optimization problem can be stated as [17]:

$$X'(t) = -\mathbf{A}X(t) - B_1 s(t) \tag{45}$$

$$\int_0^\infty |s(t)|^2 + |w(t)|^2 - |v(t) - \hat{v}(t)|^2 > 0.$$

For worst case input signals $d(t) = 0$ when $v(t) \neq 0$. For unbiased estimator $d(t) = 0$ $\Rightarrow \hat{v}(t) = 0$. Therefore, (45) can be modified as

$$X'(t) = -\mathbf{A}X(t) - B_1 s(t) \tag{46}$$

$$\int_0^\infty |s(t)|^2 + |C_2 X(t)|^2 - |C_1 X(t)|^2 > 0. \tag{47}$$

From Kalman Yakubovich theorem [29] (see Appendix A) it is shown in [17] that the solution to (46) with constraint (47) exists if $\mathbf{P}$ satisfy the Riccati equations

$$C_2^* C_2 - C_1^* C_1 - \mathbf{PA}^* - \mathbf{AP}^* = \mathbf{P}B_1 B_1^* \mathbf{P}^* \quad ; \mathbf{P} > 0 \tag{48}$$

$$B_1 B_1^* + \mathbf{AQ}^* + \mathbf{QA}^* = \mathbf{Q}(-C_1^* C_1 + C_2^* C_2); \mathbf{Q} = \mathbf{P}^{-1} > 0 \tag{49}$$

$(-\mathbf{A} - B_1 B_1^* \mathbf{P})$ and $(\mathbf{A}^* - C_1^* C_1 \mathbf{Q} + C_2^* C_2 \mathbf{Q})$ is Hurwitz.

multiplying (48) from the right hand side with $\mathbf{P}^{-1}$

$$\mathbf{A} + C_1^* C_1 \mathbf{Q} - C_2^* C_2 \mathbf{Q} = \mathbf{Q}^{-1}\left[-B_1 B_1^* - \mathbf{A}^* \mathbf{Q}\right] = \mathbf{Q}^{-1}\left[-\mathbf{A}^* - B_1 B_1^* \mathbf{Q}^{-1}\right]\mathbf{Q} \quad (50)$$

The same result has been derived in [19] in a different way.

From (42) and (43)

$$\begin{aligned} X'(t) &= \mathbf{A}X(t) + B_1 s(t) \\ X_1'(t) &= \mathbf{A}_f X_1(t) + B_f\left(C_2 X(t) + w(t)\right) \end{aligned} \quad (51)$$

or

$$\begin{bmatrix} X'(t) \\ X_1'(t) \end{bmatrix} = \begin{bmatrix} \mathbf{A} & 0 \\ B_f C_2 & \mathbf{A}_f \end{bmatrix} \begin{bmatrix} X(t) \\ X_1(t) \end{bmatrix} + \begin{bmatrix} B_1 & 0 \\ 0 & B_f \end{bmatrix} \begin{bmatrix} s(t) \\ w(t) \end{bmatrix} \quad (52)$$

$$e(t) = \begin{bmatrix} C_1 & -C_f \end{bmatrix} \begin{bmatrix} X(t) \\ X_1(t) \end{bmatrix}$$

Transfer function from inputs $s(t)$ and $w(t)$ to the output $e(t)$ is

$$\mathbf{T} = \tilde{C}\left(s\mathbf{I} - \widetilde{\mathbf{A}}\right)^{-1}\tilde{\mathbf{B}}$$

$$\Rightarrow \mathbf{T}^* = \tilde{\mathbf{B}}^*\left(-s\mathbf{I} - \widetilde{\mathbf{A}^*}\right)^{-1}\widetilde{C^*}.$$

State space representation of $\mathbf{T}^*$

$$\begin{bmatrix} P_0'(t) \\ P_1'(t) \end{bmatrix} = \begin{bmatrix} \mathbf{A}^* & C_2^* B_f^* \\ 0 & \mathbf{A}_f^* \end{bmatrix} \begin{bmatrix} P_0(t) \\ P_1(t) \end{bmatrix} + \begin{bmatrix} C_1^* \\ -C_f^* \end{bmatrix} q(t) \quad (53)$$

$$\Xi(t) = \begin{bmatrix} B_1^* & 0 \\ 0 & B_f^* \end{bmatrix} \begin{bmatrix} P_0(t) \\ P_1(t) \end{bmatrix}$$

since H_∞ norm of a transfer function and its conjugate is the same, therefore the minimization problem (45) can be formulated as

$$P_0'(t) = \mathbf{A}^* P_0(t) + C_2^* g(t) + C_1^* q(t) \quad (54)$$

$$\int_0^\infty -|q(t)|^2 + |B_1^* P_0(t)|^2 + |g(t)|^2 dt < 0 \qquad ; B_f^* P_1 = g(t)$$

since $P_0(t)$ is finite energy, therefore $P_0^*(t)\mathbf{Q}P_0(t)|_0^\infty = 0$ assuming $P_0(0) = 0$.

$\int_0^\infty -|q(t)|^2 + |B_1^* P_0(t)|^2 + |g(t)|^2 + (P_0^*(t)\mathbf{Q}P_0(t))' \, dt$

$= \int_0^\infty -|q(t)|^2 + |B_1^* P_0(t)|^2 + |g(t)|^2 + 2Re\,(P_0^*(t)\mathbf{Q}\,(\mathbf{A}^* P_0(t) + C_2^* g(t) + C_1^* q(t)))\,dt$

from (49)

$$= \int_0^\infty \left\{ |g(t) + C_2 \mathbf{Q} P_0(t)|^2 - |q(t) - C_1 \mathbf{Q} P_0(t)|^2 \right\} dt \quad (55)$$

Therefore,

$g(t) = -C_2 \mathbf{Q} P_0(t) = B_f^* P_1(t) \Rightarrow P_0(t) = P_1(t)$

$\Rightarrow P_0'(t) = \mathbf{A}^* P_0(t) + C_2^* g(t) + C_1^* q(t) \qquad ; P_1'(t) = (\mathbf{A}^* - C_2^* C_2 \mathbf{Q})\, P_1(t) + C_1^* q(t)$

$\Rightarrow \mathbf{A}_f = \mathbf{A} - \mathbf{Q} C_2^* C_2 \qquad ; C_f = -C_1 \qquad ; B_f = -\mathbf{Q}^* C_2^*$

Following the above methodology which is the basis for the state space solutions to H_∞ filtering problem in [19][3], we will compute the solution to a general filtering problem of Figure 3 with state space description:

$$X'(t) = \mathbf{A}X(t) + B_1 s(t) + B_2 w(t) \quad (56)$$
$$d(t) = C_2 X(t) + d_1 s(t) + d_2 w(t)$$
$$v(t) = C_1 X(t) + d_o s(t)$$

where $s(t), w(t)$ are the input signals. $d(t)$ and $v(t)$ are the observation and desired signal value as shown in the Figure 3. $X(t)$ is the state vector, and $\mathbf{A}, B_1, B_2, C_1, C_2, d_1, d_2, d_o$ are constant matrix, vectors, and scalars.

Most of the solutions [19][25] work under some restrictive conditions on matrices B_1, B_2, d_1, d_2. For example in [19] it is assumed that

$$D \begin{bmatrix} \mathbf{B}_0^* \\ D^* \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix} \quad ; \; D = [d_1 \quad d_2]; \; \mathbf{B}_0 = [B_1 \quad B_2]; \; d_0 = 0. \quad (57)$$

These conditions are found to be true in general for problems in control systems. But are not generally true for filtering problems (for example $d_0 \neq 0$ and $D\mathbf{B}_0^* \neq \mathbf{0}$ for filtering problem of Figure 3). One should be very careful while working with time domain solutions for a general filtering problem represented by (56) to obtain a general filter like (58). From the solution discussed above we will get an inside on why these restrictions are made and how to get around it.

Idea is to obtain a filter $H(s)$ represented in state space as

$$X_1'(t) = \mathbf{A}_f X_1(t) + B_f d(t) \quad (58)$$

$$\widehat{v}(t) \;=\; C_f X_1(t)$$

which minimizes H_∞ norm of the error $e(t) = v(t) - \widehat{v}(t)$. The minimization condition (46) and (47) for the general filtering problem is

$$X'(t) = -\mathbf{A}X(t) - B_1 s(t) - B_2\left(-d_2^{-1}C_2X(t) - d_2^{-1}d_1 s(t)\right)$$

$$\int_0^\infty |s(t)|^2 + \left|-d_2^{-1}C_2X(t) - d_2^{-1}d_1s(t)\right|^2 - |v(t) - \widehat{v}(t)|^2\,dt > 0$$

$$= \int_0^\infty |s(t)|^2 + \left|-d_2^{-1}C_2X(t) - d_2^{-1}d_1s(t)\right|^2 - |C_1X(t) + d_0s(t)|^2\,dt > 0$$

From Kalman Yakubovich theorem [29] it is shown in [17] that $\mathbf{P}$ should satisfy the following Riccati equation:

$$\mathbf{Q}_x + \mathbf{P}\mathbf{A}_1^* + \mathbf{A}_1\mathbf{P}^* = (\mathbf{F} + B\mathbf{P})\,r^{-1}\,(\mathbf{F} + \mathbf{P}B)^* \tag{59}$$

where

$B = -\left(B_1 - B_2 d_2^{-1} d_1\right)$

F is the coefficient of $X(t)$ and $s(t)$

r is the coefficient of $s^2(t)$

$\mathbf{Q}_x$ is the coefficient of $X^2(t)$

$\mathbf{Q}_x = C_2^*\left(d_2^{-1}\right)^* d_2^{-1}C_2 - C_1^*C_1$

$F = 2\,\mathrm{Re}\left(C_2^*\left(d_2^{-1}\right)^* d_2^{-1}d_1 + C_1^* d_0\right)$

$r = 1 + \left(d_1 d_2^{-1} d_0\right)^2$

$\mathbf{A}_1 = -\left(\mathbf{A} - B_2 d_2^{-1} C_2\right)$

Let $\mathbf{Q} = \mathbf{P}^{-1}$ and multiplying (59) both sides with $\mathbf{P}^{-1}$

$$\mathbf{Q}\left[\left(d_2^{-1}C_2\right)^*\left(d_2^{-1}C_2\right) - C_1^*C_1\right]\mathbf{Q} =$$

$$\mathbf{Q}\left\{\begin{array}{c}\left[2\,\mathrm{Re}\left[C_2^*\left(d_2^{-1}\right)^* d_2^{-1}d_1 + C_1^*d_0\right] + \mathbf{P}B\right]\left[1 + \left(d_1 d_2^{-1} d_0\right)^2\right]^{-1} \\ \left[2\,\mathrm{Re}\left[C_2^*\left(d_2^{-1}\right)^* d_2^{-1}d_1 + C_1^*d_0\right] + \mathbf{P}B\right]^*\end{array}\right\}$$

$$\mathbf{Q} + \mathbf{A}_1\mathbf{Q}^* + \mathbf{Q}\mathbf{A}_1^*$$

now, from (56) and (58)

$$X'(t) = \mathbf{A}X(t) + B_1 s(t) + B_2 w(t)$$

$$X_1'(t) = \mathbf{A}_f X_1(t) + B_f\left(C_2X(t) + d_1 s(t) + d_2 w(t)\right)$$

$$e(t) = C_1X(t) - C_fX_1(t) \qquad \text{(assume } d_0 = 0\text{)}$$

$$\begin{bmatrix} X'(t) \\ X_1'(t) \end{bmatrix} = \begin{bmatrix} \mathbf{A} & 0 \\ B_fC_2 & \mathbf{A}_f \end{bmatrix}\begin{bmatrix} X(t) \\ X_1(t) \end{bmatrix} + \begin{bmatrix} B_1 & B_2 \\ B_fd_1 & B_fd_2 \end{bmatrix}\begin{bmatrix} s(t) \\ w(t) \end{bmatrix}$$

$$e(t) = \begin{bmatrix} C_1 & -C_f \end{bmatrix}\begin{bmatrix} X(t) \\ X_1(t) \end{bmatrix}.$$

Transfer function from inputs $s(t)$ and $w(t)$ to the output $e(t)$ is

$$\mathbf{T} = \widetilde{C}\left(s\mathbf{I} - \widetilde{\mathbf{A}}\right)^{-1}\widetilde{\mathbf{B}} \Rightarrow \mathbf{T}^* = \widetilde{\mathbf{B}}^*\left(-s\mathbf{I} - \widetilde{\mathbf{A}}^*\right)^{-1}\widetilde{C}^*$$

state space representation of $\mathbf{T}^*$

$$\begin{bmatrix} P_0'(t) \\ P_1'(t) \end{bmatrix} = \begin{bmatrix} \mathbf{A}^* & C_2^*B_f^* \\ 0 & \mathbf{A}_f^* \end{bmatrix}\begin{bmatrix} P_0(t) \\ P_1(t) \end{bmatrix} + \begin{bmatrix} C_1^* \\ -C_f^* \end{bmatrix} q(t)$$

$$\Xi(t) = \begin{bmatrix} B_1^* & B_fd_1^* \\ B_2^* & B_fd_2^* \end{bmatrix}\begin{bmatrix} P_0(t) \\ P_1(t) \end{bmatrix}$$

Therefore,

$P_0'(t) = \mathbf{A}^* P_0(t) + C_2^* B_f^* P_1(t) + C_1^* q(t)$

Subjected to:

$\int_0^\infty -|q(t)|^2 + |B_1^* P_0(t) + B_f d_1^* P_1|^2 + |B_2^* P_0(t) + B_f d_2^* P_1|^2\, dt < 0$

$= \int_0^\infty -|q(t)|^2 + |B_1^* P_0(t) + B_f d_1^* P_1|^2 + |B_2^* P_0(t) + B_f d_2^* P_1|^2 + (P_0^*(t)\mathbf{Q}P_0(t))'\, dt$

$= \int_0^\infty -|q(t)|^2 + |B_1^* P_0(t) + B_f d_1^* P_1|^2 + |B_2^* P_0(t) + B_f d_2^* P_1|^2$

$+2Re\left(P_0^*(t)\mathbf{Q}\left(\mathbf{A}^* P_0(t) + C_2^* B_f^* P_1(t) + C_1^* q(t)\right)\right) dt$

To make above integral into components of perfect square to get equation similar to (55) is not easy. This is the reason why the assumptions in the earlier reported state space solutions are made on state space matrices to eliminate dependence on d_1, d_2, B_2, and d_0 matrices. One possible way to get around this problem is to transform the original system so that dependence on d_1 and d_2 is eliminated which can be done as follows:

$$\begin{bmatrix} s_{new}(t) \\ w_{new}(t) \end{bmatrix} = \begin{bmatrix} s(t) \\ d_1 s(t) + d_2 w(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ d_1 & d_2 \end{bmatrix} \begin{bmatrix} s(t) \\ w(t) \end{bmatrix}$$

therefore,

$$\begin{bmatrix} s(t) \\ w(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -d_2^{-1} d_1 & d_2^{-1} \end{bmatrix} \begin{bmatrix} s_{new}(t) \\ w_{new}(t) \end{bmatrix}$$

replacing $s(t)$ and $w(t)$ in (56) to eliminate d_1 and d_2 as follows:

$$\begin{aligned} X'(t) &= \mathbf{A}X(t) + \left(B_1 - B_2 d_2^{-1} d_1\right) s(t) + B_2 d_2^{-1} w(t) \qquad (60) \\ d(t) &= C_2 X(t) + 0 s(t) + w(t) \end{aligned}$$

$$v(t) = C_1 X(t) + d_0 s(t)$$

let ($d_2^{-1} \neq 0$)

$$B_{1n} = B_1 - B_2 d_2^{-1} d_1 \quad (61)$$

$$B_{2n} = B_2 d_2^{-1}$$

from (61), (60) can be written as (assume $d_0 = 0$)

$$X'(t) = \mathbf{A}X(t) + B_{1n}s(t) + B_{2n}w(t) \quad (62)$$

$$d(t) = C_2 X(t) + w(t)$$

$$v(t) = C_1 X(t)$$

even with this transformation we are restricted with the condition $d_2^{-1} \neq 0$ and $d_0 = 0$. Solution to (62) can be computed with the same procedure used for (42) and is given by:

$\mathbf{A}_f = \mathbf{A} - C_2^* C_2 - C_2^* B_{2n}^*$; $B_f = -(\mathbf{Q}C_2^* + B_{2n})$

$C_f = -C_1$

4.5 Krein Space

In recent work by Babak et. al. [12][13] H_∞ filtering problem is solved in krein space. There approach gives a recursive solution to H_∞ filtering problem. It is shown that H_∞ filter is Kalman filter in krein space. Consider the state space equations

$$X(n+1) = \mathbf{A}(n)X(n) + B_1(n)s(n) \tag{63}$$

$$d(n) = C_2(n)X(n) + w(n)$$

$$v(n) = C_2(n)X(n) \tag{64}$$

with

$$\left\langle \begin{bmatrix} s(n) \\ w(n) \\ X(0) \end{bmatrix}, \begin{bmatrix} s(k) \\ w(k) \\ X(0) \end{bmatrix} \right\rangle_k = \begin{bmatrix} q(n)\delta_{nk} & 0 & 0 \\ & r(n)\delta_{nk} & 0 \\ & & \mathbf{\Pi}_0 \end{bmatrix} \tag{65}$$

where $\langle . \rangle_k$ is the inner product in krein space. See Appendix B for krein space definition. Then the estimate of $v(n)$ from observed signal $d(n)$ is given by $\widehat{v}(n) = C_2(n)\widehat{X}(n)$ where

$$\widehat{X}(n+1) = \mathbf{A}(n)\widehat{X}(n) + K_{pn}\left(d(n) - C_2(n)\widehat{X}(n)\right) \qquad ; 0 \leq n \leq n_1 \qquad ; \widehat{X}(0) = 0$$

$$K_{pn} = \mathbf{A}(n)\mathbf{P}(n)C_2^*(n)r_{en}^{-1} \qquad ; r_{en}^{-1} = \langle e(n), e(n) \rangle_k = r(n) + C_2(n)\mathbf{P}(n)C_2^*(n)$$

and the $\mathbf{P}(n)$ can be recursively computed via the Riccati recursion

$$\mathbf{P}(n+1) = \mathbf{A}(n)\mathbf{P}(n)\mathbf{A}^*(n) - K_{pn}r_{en}K_{pn}^* + B_1(n)q(n)B_1^*(n) \qquad ; \mathbf{P}(0) = \mathbf{\Pi}_0 \tag{66}$$

4.6 Least Squares Solution

In least squares solution the average squared error is minimized. More appropriate name for this is linear minimum MSE. This is bit wordy so the name is usually shortened and the theory is included as part of the general theory of least squares. Simply stated, the linear least squares filter problem is this: given the spectral characteristics of an additive combination of signal, channel and noise, what linear operation on the input combination will yield the best estimate of the signal. Best in this case means minimum mean-squared error. This branch of filtering began with N. Wiener's work in the 1940s [28]. R. E. Kalman then made an important contribution in the early 1960s by providing an alternative approach to the same problem using state space methods [2][15]. Kalman's contribution has been especially significant in applied work, because his solution is readily implemented with modern digital methods.

4.6.1 Wiener Solution

Consider the filtering problem of Figure 3. Let $\mathbf{T}_1(s)$ be the transfer function from input signals ($s(t)$ and $w(t)$) and the output error ($e_1(t)$). Assume $W(s) = 1$. If input signals are white noises with unit variance then PSD of the error (Φ_{e1e1}) is given by [9]

$$\Phi_{e_1e_1}(s) = \mathbf{T}_1(s)\mathbf{T}_1^*(s). \tag{67}$$

least squares solution minimizes the cost function

$$J_2 = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \Phi_{e_1 e_1}(j\omega)d\omega \tag{68}$$

for minimization of J_2 one method is to convert $\mathbf{T}_1\mathbf{T}_1^*$ as (representing $[.]\,(s)$ with $[.]$)

$$\mathbf{T}_1\mathbf{T}_1^* = G_1 G_1 - G_1 G_2^* H^* - H G_2 G_1^* + H\Delta\Delta^* H^* \tag{69}$$

where Δ is the spectral factor of $(NN^* + G_2 G_2^*)$, $G_2 = GC$, $G_1 = G$. From (69):

$$\mathbf{T}_1\mathbf{T}_1^* = \left[H\Delta - G_1 G_2^* (\Delta^*)^{-1}\right]\left[\Delta^* H^* - \Delta^{-1} G_2 G_1^*\right] + \chi_1 \tag{70}$$

where

$$\chi_1 = G_1 \left(1 + G_2^* G_2\right)^{-1} G_1^*. \tag{71}$$

Since χ_1 does not depend on H therefore filter which minimizes cost function J_2 is

$$H_{weiner} = \left[G_1 G_2^* (\Delta^*)^{-1}\right]_+ \Delta^{-1} \tag{72}$$

where $[u]_+$ is the causal part of u. Similarly, the solution to Wiener filter for discrete time system of Figure 2 is given by (72) where all the transfer functions are in z domain and Δ is obtained by spectral factor of the $(N(z)N^*(z) + G_2(z)G_2^*(z))$.

4.6.2 Kalman Filter

Consider filtering problem in state space as

$$X(n+1) = \mathbf{A}(n)X(n) + B_1 s(n) \tag{73}$$

$$d(n) = C_2(n)X(n) + w(n)$$
$$v(n) = C_2(n)X(n) \tag{74}$$

where $d(n)$ and $v(n)$ are the observation and desired signal values. $\mathbf{A}(n)$ and $C_2(n)$ represent time variant state space matrix and vector respectively. The correlation values for input signals $s(n)$ and $w(n)$ are given by:

$$E\left[s(k)s^*(i)\right] = \left\{ \begin{array}{cc} q(k) & i = k \\ 0 & i \neq k \end{array} \right\}$$

$$E\left[w(k)w^*(i)\right] = \left\{ \begin{array}{cc} r(k) & i = k \\ 0 & i \neq k \end{array} \right\}$$

$E\left[s(k)w^*(i)\right] = 0$, for all i and k.

The recursive filter which estimates $X(n)$ as $\widehat{X}(n)$ and minimizes

$$\mathbf{P}(n) = E\left[\left(X(n) - \widehat{X}(n)\right)\left(X(n) - \widehat{X}(n)\right)^*\right] \tag{75}$$

is given by [2] (assume $\widehat{X}_1(n)$ and $\mathbf{P}^{-1}(n)$ are known for $n = 0$)

$$K(n) = \mathbf{P}^{-1}(n)C_2^*(n)\left(C_2(n)\mathbf{P}^{-1}(n)C_2^*(n) + r(n)\right)^{-1}$$

$$\widehat{X}(n) = \widehat{X}_1(n) + K(n)\left(d(n) - C_2(n)\widehat{X}_1(n)\right)$$

$$\mathbf{P}(n) = \left(1 - K(n)C_2(n)\right)\mathbf{P}^{-1}(n)$$

$$\widehat{X}_1(n+1) = \mathbf{A}(n)\widehat{X}_1(n)$$

$$\mathbf{P}^{-1}(n+1) = \mathbf{A}(n)\mathbf{P}(n)\mathbf{A}^*(n) + B_1B_1^*q(n)$$

The above equations are called Kalman filter equations and the recursive solution of

Kalman filter converges to Wiener filter solution in steady state as both minimizes MSE.

4.7 FIR Filter Solution

There are lot of applications in signal processing and communications where finite impulse response (FIR) filters are preferred over infinite impulse response (IIR) filters. For example applications which require real time implementation, FIR filters are preferred. The reason is that there are fast digital signal processing (DSP) chips which are available for real time implementation of the FIR filters. Also, in adaptive FIR filter applications stability is not an issue and therefore are preferred over IIR filters. IIR filters have problem with stability and hence require more computing complexity due to requirement of stability checks during filter coefficient update.

This section looks into the FIR filter solutions which minimizes least squares and H_∞ error criteria.

4.7.1 Wiener Filter

Consider the filtering problem of Figure 2. Assume all the transfer functions are FIR (ARMA model can be approximated with FIR). Observed signal $d(n)$ is given by

$$d(n) = \left(G_{imp} * \sum_{k=0}^{r_1-1} c_{imp}(k) s(n-k) \right) + \left(\sum_{l=0}^{g_1-1} n_{imp}(l) w(n-l) \right) \quad (76)$$

$$= \left(\sum_{m=0}^{z_1-1}\left(\sum_{k=0}^{r_1-1} c_{imp}(k)s(n-m-k)\right) g_{imp}(m)\right) + \left(\sum_{l=0}^{g_1-1} n_{imp}(l)w(n-l)\right)$$

G_{imp}, C_{imp}, and N_{imp} are the signal, channel, and noise impulse response respectively and are represented as:

$$\begin{aligned} G_{imp} &= [g_{imp}(0)\ \ g_{imp}(1)\ \ \ldots\ \ g_{imp}(z_1-1)] \\ C_{imp} &= [c_{imp}(0)\ \ c_{imp}(1)\ \ \ldots\ \ c_{imp}(r_1-1)] \\ N_{imp} &= [n_{imp}(0)\ \ n_{imp}(1)\ \ \ldots\ \ n_{imp}(g_1-1)] \end{aligned} \tag{77}$$

Where r_1, g_1, and z_1 are the lengths of channel, noise, and signal FIR models. Estimate of the signal is obtained by passing $d(n)$ through a filter obtained by minimizing $e_1(n)$ using Wiener or H_∞ criterion and is given by

$$\begin{aligned} \hat{v}(n) &= H_{imp} * d(n) \\ &= \sum_{o=0}^{u_1-1}\left(\begin{array}{c} \left(\sum_{m=0}^{z_1-1}\left(\sum_{k=0}^{r_1-1} c_{imp}(k)s(n-o-m-k)\right) g_{imp}(m)\right) + \\ \left(\sum_{l=0}^{g_1-1} n_{imp}(l)w(n-o-l)\right) \end{array}\right) h_{imp}(o) \end{aligned} \tag{78}$$

The signal to be estimated is given by

$$\begin{aligned} v(n) &= G_{imp} * s(n) \\ &= \sum_{p=0}^{z_1-1} g_{imp}(p)s(n-p). \end{aligned} \tag{79}$$

The error is given by

$$e(n) = v(n) - \hat{v}(n) \tag{80}$$

$$\begin{aligned} e_1(n) &= W_{imp} * e(n) \\ &= \sum_{q=0}^{v_1-1} w_{imp}(q)e(n-q) \end{aligned} \quad (81)$$

where H_{imp} and W_{imp} is the impulse response of the filter $H(z)$ and weighting filter $W(z)$ respectively and given by:

$$\begin{aligned} H_{imp} &= [h_{imp}(0) \;\; h_{imp}(1) \;\; \ldots \;\; h_{imp}(u_1-1)] \\ W_{imp} &= [w_{imp}(0) \;\; w_{imp}(1) \;\; \ldots \;\; w_{imp}(v_1-1)] \end{aligned} \quad (82)$$

From (78), (79), and (80) $e(n)$ is given by

$$\sum_{p=0}^{z_1-1} g_{imp}(p)s(n-p) - \sum_{o=0}^{u_1-1} \left(\begin{array}{c} \left(\sum_{m=0}^{z_1-1} \left(\sum_{k=0}^{r_1-1} c_{imp}(k)s(n-o-m-k)\right) g_{imp}(o)\right) + \\ \left(\sum_{l=0}^{g_1-1} n_{imp}(l)w(n-o-l)\right) \end{array} \right) h_{imp}(o) \quad (83)$$

To find $H(z)$ which minimizes $E\left[e^2(n)\right]$, error and observation signal should satisfy [14]

$$E\left[e(n)d^*(n-k)\right] = 0 \qquad k = 0, 1, 2, .., u_1 - 1. \quad (84)$$

From (84) we can obtain u_1 equations with u_1 unknowns (filter coefficients) [14]

$$\mathbf{R}_{vv} H_{imp}^t = R_{vd} \quad (85)$$

where

$R_{vv} =$

$$\begin{bmatrix} E[v(n)v^*(n)] & E[v(n)v^*(n-1)] & E[v(n)v^*(n-u_1-1)] \\ E[v(n-1)v^*(n)] & E[v(n-1)v^*(n-1)] & E[v(n-1)v^*(n-u_1-1)] \\ . & . & \\ . & . & . \\ . & . & \\ E[v(n-u_1-1)v^*(n)] & \ldots & ..E[v(n-u_1-1)v^*(n-u_1-1)] \end{bmatrix}$$

$$R_{vd} = \begin{bmatrix} E[v(n)d^*(n)] \\ E[v(n)d^*(n-1)] \\ E[v(n)d^*(n-2)] \\ . \\ . \\ . \\ E[v(n)d^*(n-u_1-1)] \end{bmatrix}$$

4.7.2 H_∞ Solution

Unlike Wiener solution H_∞ filter is more complex because the error surface with respect to filter coefficients is not smooth. This makes it difficult to estimate the gradient on the error surface in the direction of the minima. In case of Wiener solution the error surface has a quadratic form and so the solution is computed in closed form.

First it will be shown that the error surface in case of H_∞ filter solution is convex. Since $H_{imp} \epsilon R^n$ (set of real numbers), therefore an affine function $\mathbf{T}_1(z) = \Im_1(H_{imp})$ is convex. Where $\mathbf{T}_1(z) = [G(z) - G(z)C(z)H(z) \quad - H(z)N(z)]\,W(z)$ is the transfer function from input signals $(s(n)\ ,\ w(n))$ to output error $(e_1(n))$ as in Figure 2. $H(z) = \sum_{k=0}^{u_1-1} h_{imp}(k) z^{-k}$. Also, function $\|\mathbf{T}_1\|_\infty = \Im_2(\mathbf{T}_1)$ is convex because norm is a convex function. Therefore, the function $\|\mathbf{T}_1\|_\infty = \Im_3(H_{imp})$ is convex.

Using convex nature any minimization techniques can be used to find minima with respect to the filter coefficients $([h_{imp}(0) \quad h_{imp}(1) \quad \ldots \quad h_{imp}(u_1 - 1)])$. The FIR filter solutions computed in Chapter 5 have used simplex method (function FMINS of MATLAB) to obtain the filter which minimizes the H_∞ criteria.

CHAPTER 5

SIGNAL ESTIMATION

This section discusses the signal estimation problem encountered in signal processing and communication applications. In many applications we are faced with the problem of estimating signal embedded in noise or corrupted by channel. Consider the signal estimation problem of Figure 3 for a continuous time system and Figure 2 for discrete time system. $w(t)$, $s(t), w(n)$, and $s(n)$ are continuous and discrete time white noises of unit variance which are inputs to the system. $G(s)$, $C(s)$, $N(s)$, and $G(z)$, $C(z)$, $N(z)$ are signal, noise, channel models for continuous and discrete time system respectively. $H(s)$ or $H(z)$ is the filter obtained by minimizing error using H_∞ or minimum variance criterion. $W(s)$ or $W(z)$ is the model for weighing the error in specific frequency bands. The signal estimation problem is to estimate $v(t)$ or $v(n)$ from given corrupted signal $d(t)$ or $d(n)$ by minimizing error $e_1(t)$ or $e_1(n)$ for continuous or discrete time system respectively.

5.1 Continuous Time System

Consider an example of signal estimation problem for continuous time system. Let

$$G(s) = \frac{50}{s+50} \tag{86}$$

$$C(s) = \frac{2s + 100}{s^2 + 0.4s + 100} \tag{87}$$

$$N(s) = k_{po} \qquad ; W(s) = 1 \tag{88}$$

where k_{po} is a constant value and controls the power of the noise added to the observed signal $d(n)$. Frequency magnitude plot for signal ($G(s)$), channel ($C(s)$), and noise ($N(s)$) models are shown in Figure 5. As seen from the Figure 5 signal is low pass in nature corrupted by low pass channel and additive white noise. To estimate the signal $v(t)$ from observed signal, $d(t)$ is passed through a filter $H(s)$ obtained by minimizing $e_1(t)$ using Wiener solution or sub-optimal H_∞ state space solution as discussed in chapter on mathematical solutions. $k_{po} = 0.1$. Frequency magnitude plots of the filters obtained using minimum variance and H_∞ criteria are shown in Figure 6. Both filters are low pass in nature with a notch around 10 rad/sec frequency to compensate for the channel. The roll off for Wiener filter is much higher than H_∞ filter suggesting better rejection of white noise in the high frequency range. Frequency magnitude of Wiener and H_∞ filters for low SNR ($k_{po} = 200$) is shown in Figure 7. Both filters have similar shape with the difference that H_∞ filter have deeper notch to combat channel effects compared to Wiener filter. This suggests that H_∞ filter can recover signal corrupted by channel much better than Wiener filter at low SNR. To see the effects of parameter γ on sub-optimal H_∞ filter solution plots of MSE are generated for varying value of γ for high ($k_{po} = .1$) and low ($k_{po} = 200$) SNR as shown in Figure 8 and Figure 9 respectively. Also, plots of power spectrum density with varying value of γ are shown in Figure 10. It is clear from Figures 8, 9, and 10 that as γ is increased the H_∞ filter solution converges to Wiener solution. MSE obtained with H_∞ filter ($\gamma = 10$) converges to MSE using Wiener filter for both vales of k_{po} ($k_{po} = .1$ and $k_{po} = 200$). This coincides with the observation of the fact that as $\gamma \rightarrow \infty$, H_∞ solution converges to Wiener solution (see Chapter 4 on Mathematical Solution). Also, it is clear from the plots that at low SNR the effect of parameter γ on the error variance is reduced or in other words variation of the error variance with respect to γ is reduced at low SNR. From Figures 8 and 9 it is clear that sub-optimal H_∞ filters are useful in situation where trade-off is to be made between performance (in minimum MSE sense) and robustness (achieving lowest upper bound on the MSE in the worst case). For example, from Figure 10 $\gamma = 1.2$ is a good choice for trade-off between performance and robustness.

5.2 Discrete Time System

Signal estimation problem for a discrete time system is considered with following transfer functions for signal and noise:

$$N(z) = \frac{k_{po}}{1 + 1.4z^{-1} + .85z^{-2}} \tag{89}$$

$$C(z) = 1 \qquad ; G(z) = 1; \qquad W(z) = 1 \tag{90}$$

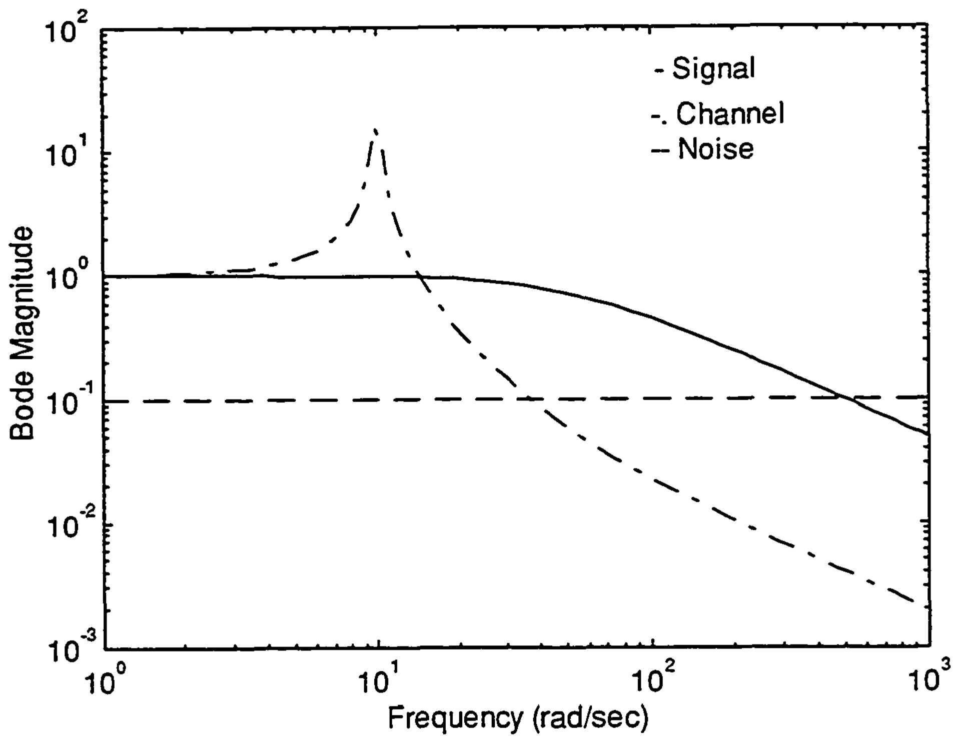

Figure 5: Frequency magnitude plot for Signal, Channel, and Noise transfer functions.

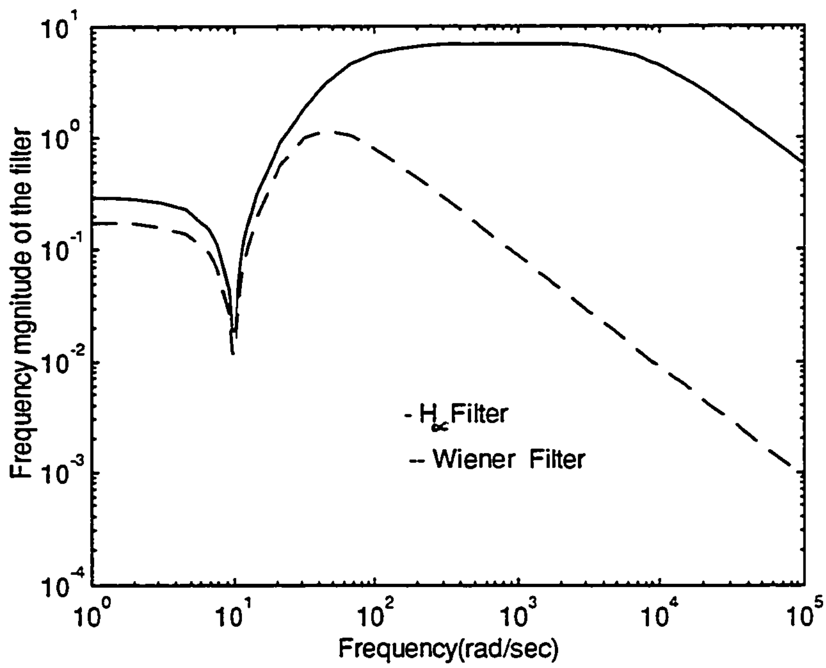

Figure 6: Frequency magnitude plot of filters obtained using Wiener and H_∞ filter.

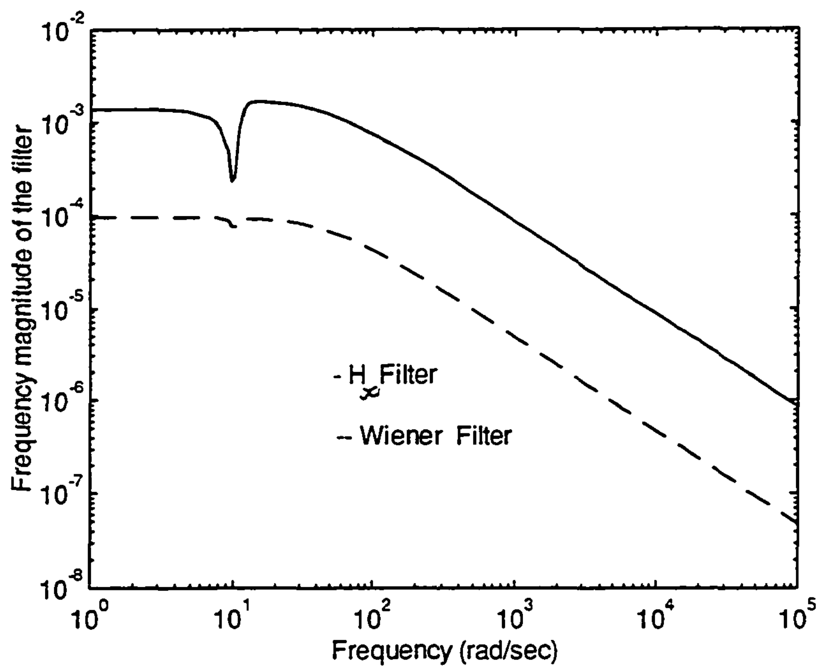

Figure 7: Frequency magnitude plot of Wiener and H_∞ filters at low SNR.

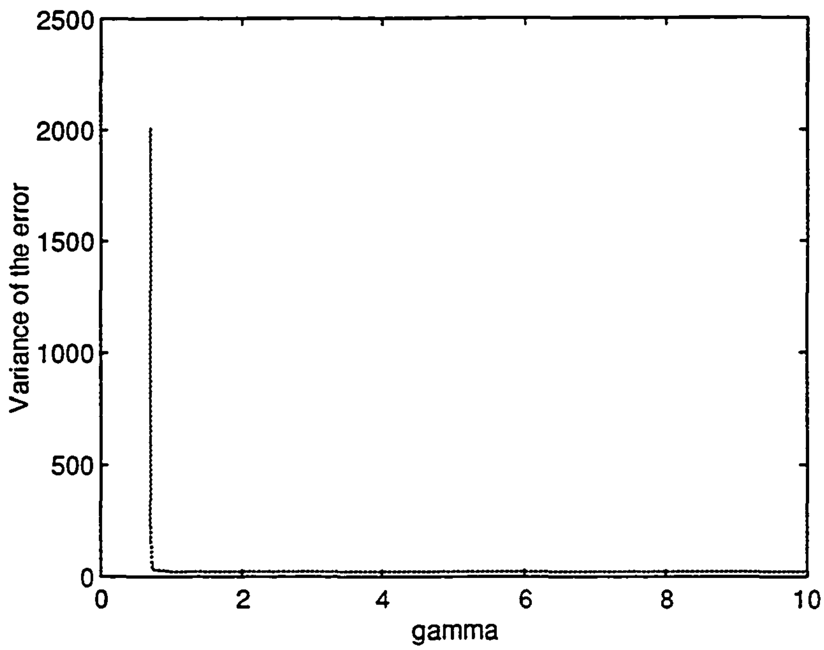

Figure 8: Plot of MSE with varying value of gamma (γ) for high SNR ($k = .1$).

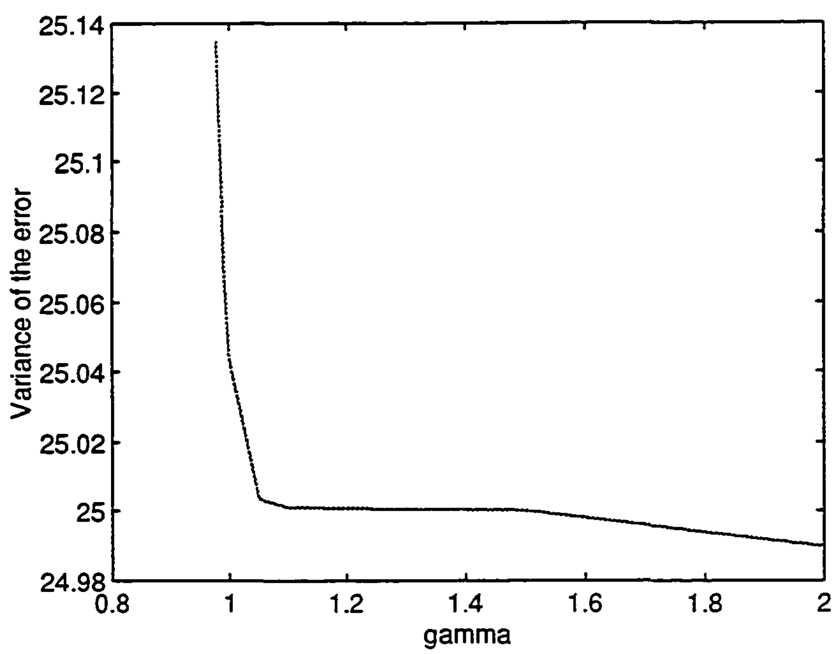

Figure 9: Plot of MSE with varying value of gamma (γ) for low SNR ($k = 200$).

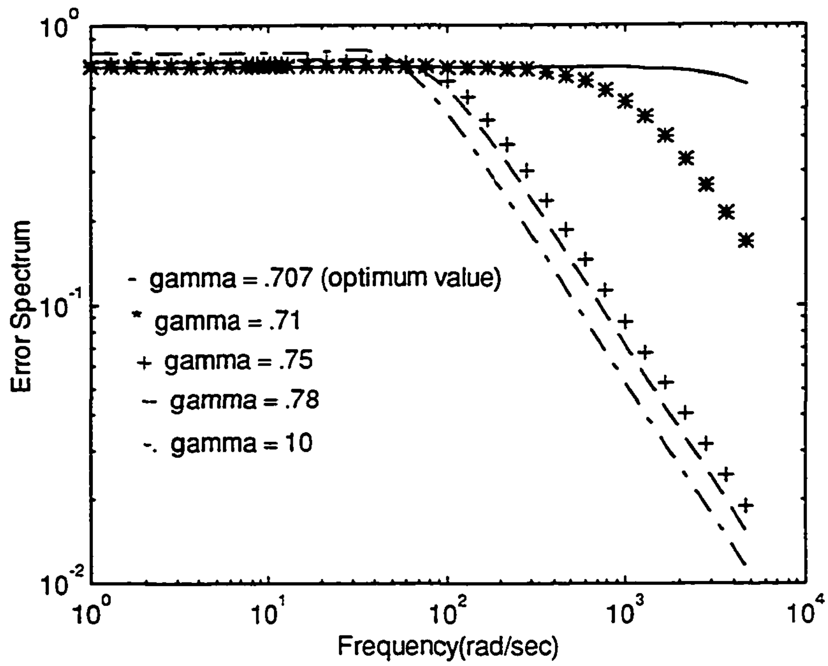

Figure 10: Error spectrum plot with varying value of γ for high SNR ($k = .1$).

where k_{po} is a constant and controls the SNR at the input (in dB) defined as:

$$10\log\left(\frac{E\left[v(n)v^{*}(n)\right]}{E\left[x(n)x^{*}(n)\right]}\right). \tag{91}$$

The signal is wide bandwidth with noise as narrow band. Idea is to reject narrow band interference and estimate the signal from given observed signal $d(n)$ as shown in Figure 2. This kind of signal and noise models are encountered in CDMA system as discussed in the next section. FIR filter is obtained using Wiener and H_{∞} minimization criterion (see chapter 4 on Mathematical Solutions). Performance of filters using Wiener and H_{∞} criterion is compared by computing the SNR at the output (in dB),i.e., computing signal and error power ratio as follows:

$$10\log\left(\frac{E\left[v(n)v^{*}(n)\right]}{E\left[e1(n)e1^{*}(n)\right]}\right). \tag{92}$$

The performance benefits of H_{∞} and Wiener filter is shown in Figure 11. Since Wiener filter needs a priori knowledge of SNR therefore, $SNR = -10dB$ is used to compute optimal filter. It is clear from the Figure 11 that H_{∞} filters are more robust to changes in input SNR. There performance degrades uniformly when input SNR is reduced away from $-10dB$. On the other hand the Wiener filter performance degrades drastically when input SNR is below $-10dB$. However, the performance of Wiener filter is superior at high SNR. On an average if the SNR at the input is changing over time then H_{∞} filter should give better performance in terms of output SNR as seen from Figure 11.

To generate theoretical upper and lower bounds to the experimental curves obtained from Wiener and H_∞ solution, theoretical MSE is computed at the output as follows:

$$d(n) = (G_{imp} * C_{imp} * s(n)) + (N_{imp} * w(n)) \tag{93}$$

where G_{imp}, C_{imp}, and N_{imp} are the time domain impulse response of the signal, channel, and noise model transfer functions. For simplicity we assume FIR models for all the transfer functions (as described in section 4.7.1). For the example considered above $c_{imp}(0) = 1, g_{imp}(0) = 1$, and $W(z) = 1 \Rightarrow e(n) = e_1(n)$. $c_{imp}(k) = 0$, and $g_{imp}(k) = 0$ for $k \neq 0$. Assuming these values for C_{imp} and G_{imp} we can simplify (83) as

$$e(n) = s(n) - \sum_{o=0}^{u_1-1} s(n-o)h_{imp}(o) + \sum_{o=0}^{u_1-1} h_{imp}(o)\left(\sum_{l=0}^{g_1-1} n_{imp}(l)w(n-o-l)\right). \tag{94}$$

To compute MSE we have to find $E\left[e(n)e^*(n)\right]$. From (94) $E\left[e(n)e^*(n)\right]$ is given by ($s(n)$ and $w(n)$ are zero mean and uncorrelated)

$$\begin{aligned}
&E\left[s(n)s^*(n)\right] + &(95)\\
&\sum_{k=0}^{u_1-1}\sum_{l=0}^{u_1-1} h_{imp}(l)h^*_{imp}(k)E\left[s(n-l)s^*(n-k)\right] + &\\
&\sum_{o=0}^{u_1-1}\sum_{m=0}^{u_1-1} h_{imp}(o)h^*_{imp}(m) &\\
&\left(\sum_{l=0}^{g_1-1}\sum_{p=0}^{g_1-1} n_{imp}(l)n^*_{imp}(p)E\left[w(n-o-l)w^*(n-m-p)\right]\right) &(96)\\
&-2\sum_{q=0}^{u_1-1} h^*_{imp}(q)E\left[s(n)s^*(n-q\right] &
\end{aligned}$$

To obtain an upper bound on experimental values, filter $H(z)$ is computed using Wiener solution as discussed in chapter 4 with $u_1 = 6$ for varying value of input SNR. This value of $H(z)$ is used to compute MSE from (95) for varying value of input SNR and is plotted in Figure 11. The lower bound is obtained from the fact that H_∞ solution in time domain satisfies (Chapter 3)

$$E\left[e(n)e^*(n)\right] < \gamma^2 \left(E\left[s(n)s^*(n)\right] + E\left[w(n)w^*(n)\right]\right). \tag{97}$$

From (97) and the fact that SNR at the output is given by (92) a lower bound on the output SNR is plotted in Figure 11. It is seen from Figure 11 that the SNR plots obtained using Wiener and H_∞ filters lie between theoretical upper and lower bounds.

5.3 CDMA System

Consider an example of code division multiple access (CDMA) system. Binary information is to be transmitted using binary phase shift keying through an additive noisy channel. Binary bits are correlated with a pseudo noise (PN) sequence before transmission as shown in Figure 12. Signal at the receiver is passed through the filter, and correlator. Output of the correlator is passed to the decision making device. When the noise is narrow band, then the purpose of the filter is to reduce noise level (increase SNR) at the input of the decision device.

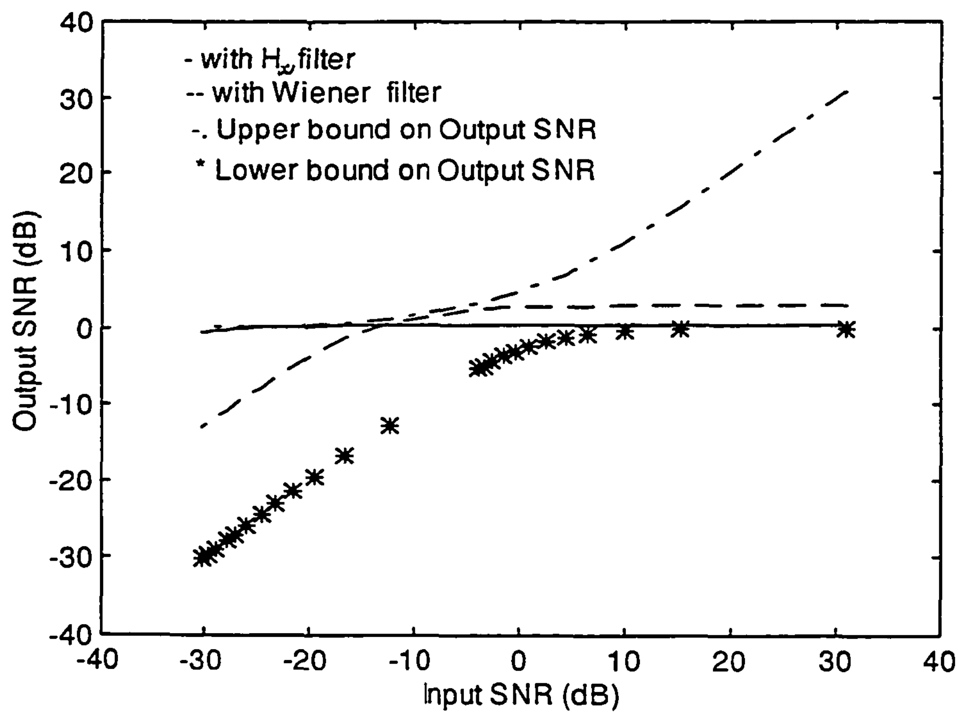

Figure 11: Plot of output SNR with varying input SNR using Wiener and H_∞ filters.

To compare the performance benefits of minimum variance and H_∞ criterion, SNR input to the decision device is computed using Wiener and H_∞ filters. Binary phase shift keying is used for the modulation [22] and pseudo noise sequence of length 31 is taken with chip to be $+/-1$. $C(z) = 1$.

$$N(z) = 1.08/\left[1 - 0.94880z^{-1} + .614z^{-2} + 0.1416z^{-3} + 0.1291z^{-4} + 0.0743z^{-5}\right] \quad (98)$$

Where $s(n)$ is a white noise of unit variance. Frequency plot of noise transfer function, Wiener filter, and H_∞ filter is shown in Figure 13. It is clear from the plot that H_∞ filter has a deeper notch and high attenuation compared to Wiener filter. This reflects the conservative behavior of H_∞ criterion which is tuned to worst case noises. To see this effect further, SNR at the input of the decision device when no filtering is used (SNRi), is plotted against SNR when Wiener filter (SNRw) and H_∞ filter (SNRh) is used to reduce the noise effects as shown in Figure 14. Optimum filters are computed using SNRi=15dB. Keeping the filters same the SNRi is varied by changing input signal and noise levels and SNRw and SNRh is plotted for this varying SNRi. The performance of the H_∞ filter shows improvement of 5dB on an average over different SNRi while the Wiener filter performance degrades drastically after SNRi=0dB. This again reflects the robust behavior of H_∞ filters with superior performance when input SNR is not known and is changing in an random fashion over a range of -4dB to 15dB.

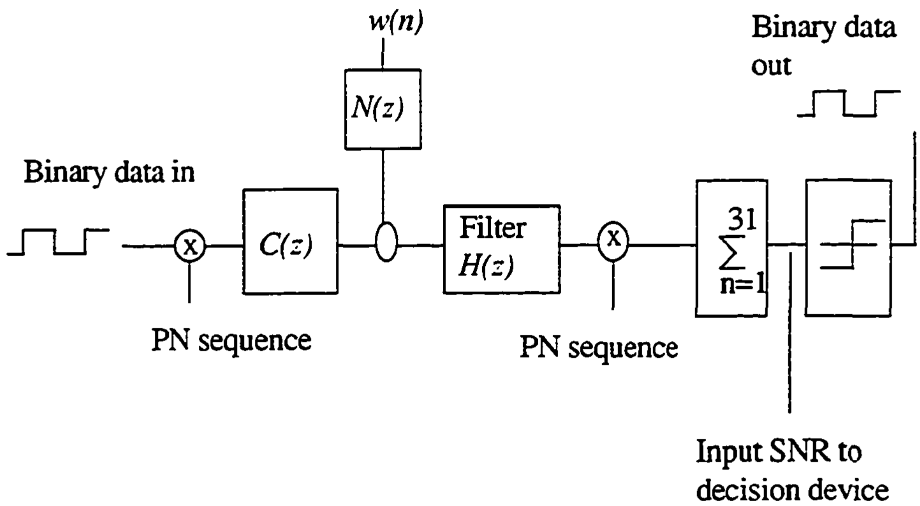

Figure 12: CDMA system with narrow band interference with transfer function $N(z)$.

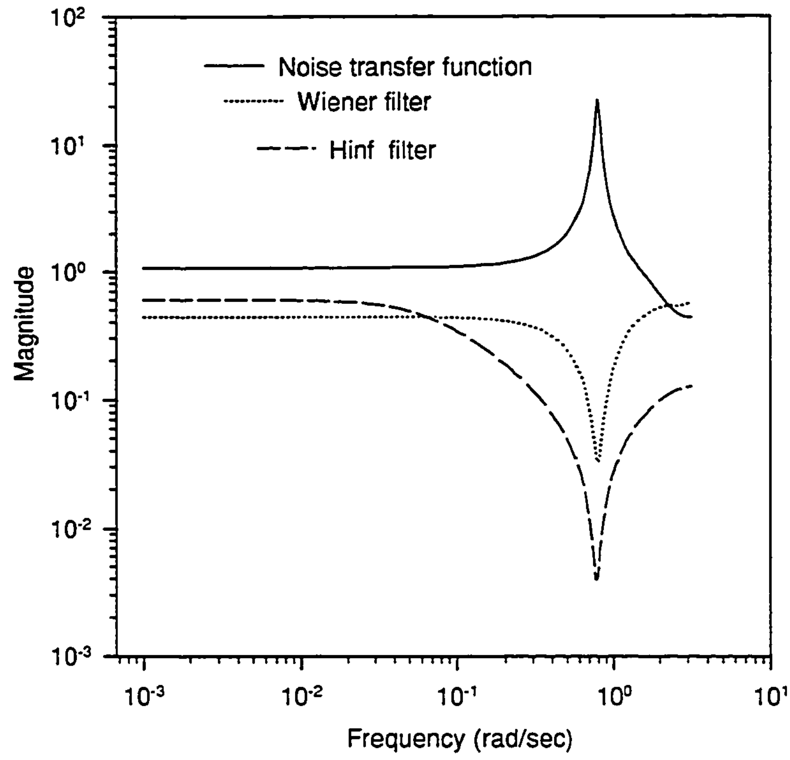

Figure 13: Frequency response of noise transfer function ($N(z)$), Wiener, and H_∞ filters.

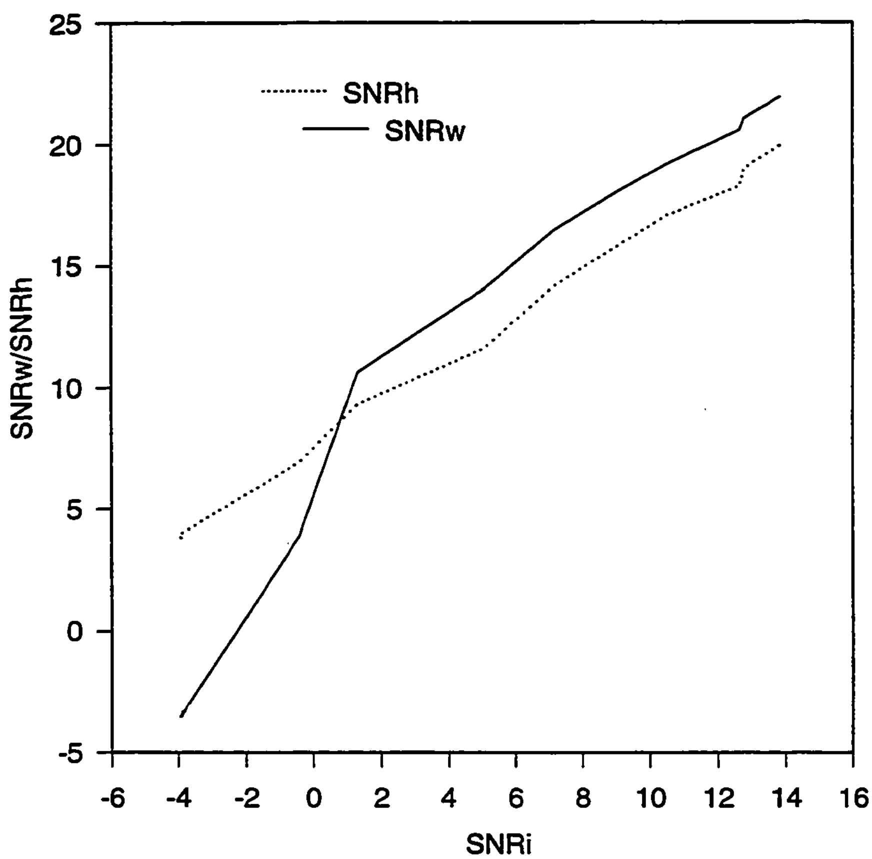

Figure 14: Input SNR in dB (SNRi) versus SNR in dB obtained using Wiener and H_∞ (SNRw/ SNRh) filters.

5.4 Robust Performance

Assume $C(s) = 1$ and let $\mathbf{T}(s)$ be the transfer function between input signals $s(t)$ and $w(t)$ and output signal $e(t)$, as shown in Figure 3. Then the transfer function can be written as

$$\mathbf{T} = [\mathbf{T}_1(s) - \mathbf{T}_2(s)H(s)] \tag{99}$$

where $\mathbf{T}_1(s) = [G(s) \quad 0]$ and $\mathbf{T}_2(s) = [G(s) \quad N(s)]$. If the input signals are white noises with unit variances, then the PSD of the error $e(t)$ is given by [10]

$$\Phi_{ee}(s) = \mathbf{T}(s)\, \mathbf{T}^*(s) \tag{100}$$

The H_∞ optimization criterion implies minimization of the largest singular value of the error PSD matrix, which can be represented as [10]

$$\min_{H(s)} \sup_{\omega} |\Phi_{ee}(j\omega)| = \min_{H(s)} \|\Phi_{ee}(s)\|_\infty = \min_{H(s)} \|\mathbf{T}(s)\|_\infty \tag{101}$$

where sup is the supremum over all ω . The optimal filter is the solution to

$$\min_{H(s)} \|\mathbf{T}_1(s) - \mathbf{T}_2(s)\, H(s)\|_\infty \tag{102}$$

where $H(s) \in RH_\infty$.

Theorem 2 *Let H' be the optimal H_∞ filter for (representing transfer function* [.] (s) *with* [.]*)*

$$\mathbf{T} = [G - G\,H \quad -H\,N] \tag{103}$$

with $\|\mathbf{T}\|_\infty \leq \gamma$ *. Then for any noise transfer function with* $\left\|N'(s)\right\|_\infty \leq p_1$ *and*

$$\mathbf{T}' = \left[G - G\,H' \quad - H'N'\right] \tag{104}$$

the error variance is bounded as

$$\sup_{\|N'\|_\infty \leq p1} \|e(t)\|_2^2 \leq \gamma^2 \tag{105}$$

where

$$p_1 = \frac{\sqrt{\gamma^2 - \delta^2}}{\|H'\|_\infty} \tag{106}$$

and

$$\delta = \left\|G - G\,H'\right\|_\infty \tag{107}$$

Proof.

$$\begin{aligned} \Phi_{ee}(s) = & \left\|G - G\,H'\right\|^2 + \left\|N\,H'\right\|^2 \leq \gamma^2 \\ \Rightarrow \quad & \|N\|^2 \left\|H'\right\|^2 \leq \gamma^2 - \left\|G - G\,H'\right\|^2 \\ \Rightarrow \quad & \|N\|^2 \leq \frac{\gamma^2 - \left\|G - G\,H'\right\|^2}{\|H'\|^2} \end{aligned} \tag{108}$$

where $\|[.]\|^2 = [.]\,[.]^*$*. This means any noise transfer function satisfying*

$$\left\|N'\right\| \leq \sqrt{\frac{\gamma^2 - \|G - G\,H'\|^2}{\|H'\|^2}} \tag{109}$$

also satisfies

$$\Phi_{ee}(s) = \left\|G - G\,H'\right\|^2 + \left\|N'\,H'\right\|^2 \leq \gamma^2 \tag{110}$$

∎

A more relaxed bound can be represented using the infinity norm on N' satisfying

$$\left\|N'\right\|_\infty \le \frac{\sqrt{\gamma^2 - \|G - GH'\|_\infty}}{\|H'\|_\infty} \tag{111}$$

This means that for all noise transfer functions belonging to the class $\left\|N'\right\|_\infty \le p_1$ there will be an upper bound on the error variance and the error PSD. This gives robust l_2 (minimum variance) and H_∞ performance where upper bound is given by γ^2 .

The following example is presented to demonstrate robust performance of an H_∞ filter and compare it with the Wiener filter. Our approach to designing an H_∞ filter is to use the γ -iteration algorithm, which is based on Nehari's theorem as discussed in Chapter 4 on Mathematical Solution. Let

$$G(s) = \frac{1.5s^2 + 4s + 2}{s^2 + 2.4s + 7.2} \tag{112}$$

$G = G_1$, $E\left[s(t)s^*(t)\right] = 10$, $E\left[w(t)w^*(t)\right] = 1$, and $N(s) = 1$. The estimated signals PSD using Wiener and H_∞ filters are shown in Figure 15. The constants of the design are found to be $\gamma^2 = 1$ and $p1 = 0.98$. It is seen from Figure 15 that the estimate of the signal PSD using H_∞ follows closely with true spectra compared to Wiener filter. This is due the large bandwidth of H_∞ filters [25] and so can estimate large bandwidth signal much better than Wiener filter. Also, the PSD of the error using two different noise transfer functions

$$N1' = \frac{(s+1)}{(s+2)} \tag{113}$$

and

$$N2' = \frac{0.6\ (s + 1)}{s^4 + s^3 + s^2 + s + 1} \tag{114}$$

were obtained with the restriction of $\left\| N1' \right\|_\infty \leq p1$ and $\left\| N2' \right\|_\infty \leq p1$ as shown in Figure 16. It is seen from Figure 16 that the PSD is always less than γ^2 for H_∞ case.

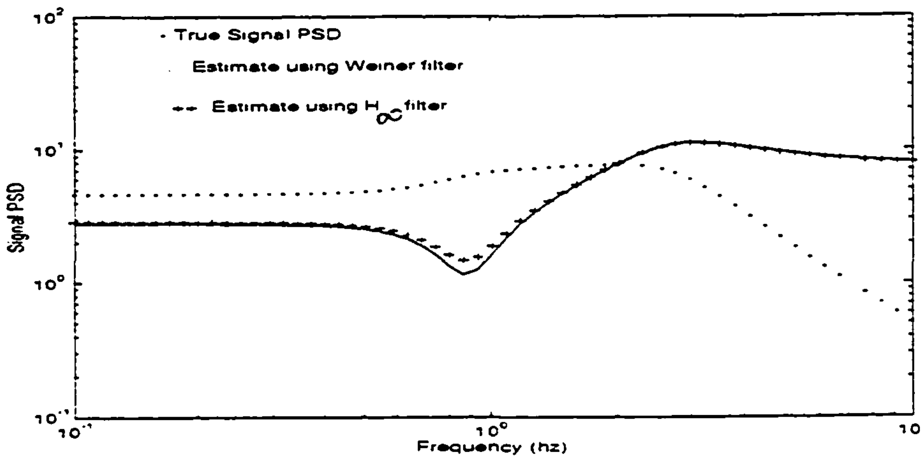

Figure 15: Estimated signal PSD using Weiner and H_∞ filter.

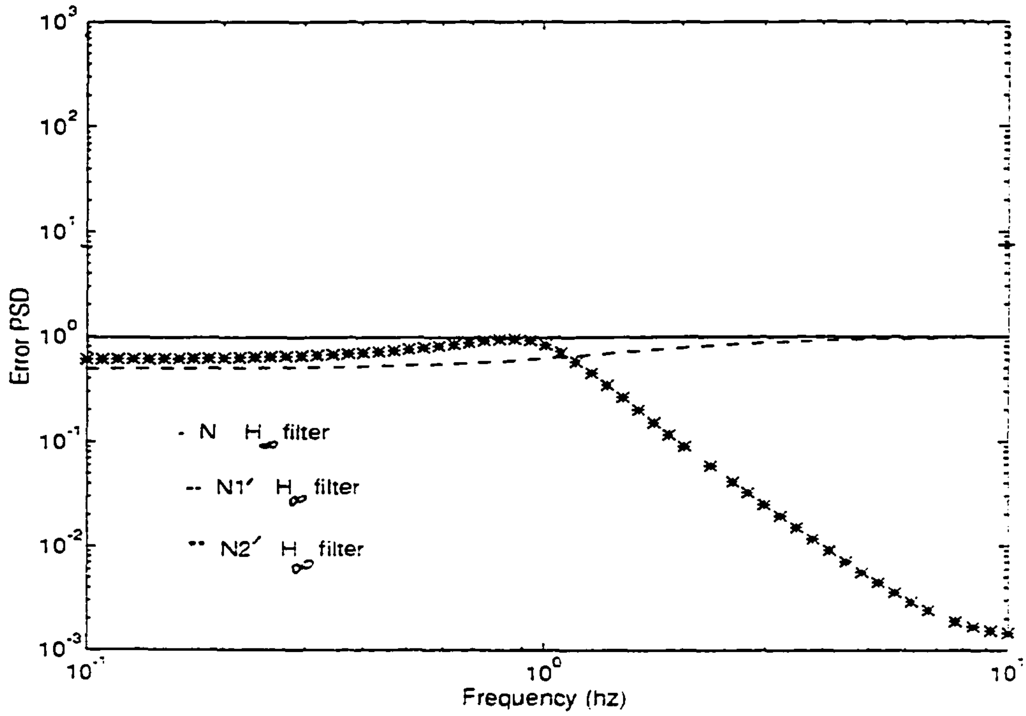

Figure 16: Error PSD plot using three noise transfer functions. $N = 1$ is used to compute Weiner and H_∞ filter.

CHAPTER 6

ADAPTIVE FILTERS

This section will describe adaptive filters satisfying H_∞ criteria. Adaptive filters are useful in estimating an unknown impulse response of a given system. The filter coefficients adapt to the changing impulse response and input signal non-stationarity. First H_∞ problem is formulated in state space form and reduced to an adaptive filter problem. From this it is shown that well known NLMS and LMS algorithms satisfy optimal H_∞ criteria. New class of sub-optimal H_∞ algorithms are originally derived and compared with RLS algorithm.

6.1 H_∞ Problem Formulation

Let us consider a state space signal and noise model for a filtering problem (single input output case) as follows:

$$\begin{aligned} X(n+1) &= \mathrm{A}(n)X(n) + B_1(n)s(n) \quad , \quad X(0) \qquad (115) \\ d(n) &= C_2(n)X(n) + w(n) \end{aligned}$$

where $\mathrm{A}(n)$, $B_1(n)$, and $C_2(n)$ are the time variant state space matrix and vectors describing the system. $X(n)$ is the state vector. $d(n)$ is the observation vector. $s(n)$, and $w(n)$ are unknown quantities and correspond to input signal and noise respectively. The signal to be estimated is given by the linear combination ($C_1(n)$) of state vector:

$$v(n) = C_1(n)X(n). \tag{116}$$

We would like to find optimal estimators ϱ_f and ϱ_p such that

$$\hat{v}_f(n) = \varrho_f(d(0), d(1), d(2), ..., d(n)) \tag{117}$$

$$\hat{v}_p(n) = \varrho_p(d(0), d(1), d(2), ..., d(n-1))$$

$$e_f(n) = v(n) - \hat{v}_f(n) \qquad ; e_p(n) = v(n) - \hat{v}_p(n) \tag{118}$$

where $\hat{v}_f(n)$ and $\hat{v}_p(n)$ denote the estimate of $v(n)$ given observations from time 0 up to n and from time 0 up to time $n-1$ respectively. $e_f(n)$ and $e_p(n)$ are filter and prediction errors respectively. Let T_f (T_p) denote the transfer operators that maps the unknown disturbances $\{X(0) - \widehat{X}(0), s(n), w(n)\}$ ($\widehat{X}(0)$ denote the initial guess of $X(0)$) to the filtered (predicted) error $e_f(n)$ ($e_p(n)$).

To find optimal estimators which satisfy H_∞ criteria we have to find solution to the following:

$$\inf_{\varrho_f} \|T_f\|_\infty^2 =$$

$$\inf_{e_f} \sup_{X(0),s(n)\in h_2,w(n)\in h_2} \frac{\|e_f\|_{2,[0,n]}}{(X(0)-\widehat{X}(0))^*\Pi_0(X(0)-\widehat{X}(0)) + \|s\|_{2,[0,n]} + \|w\|_{2,[0,n]}} \tag{119}$$

$$\inf_{e_p} \|T_p\|_\infty^2 =$$

$$\inf_{e_p} \sup_{X(0),s\in h_2,w\in h_2} \frac{\|e_p\|_{2,[0,n-1]}}{(X(0)-\widehat{X}(0))^*\Pi_0(X(0)-\widehat{X}(0)) + \|s\|_{2,[0,n-1]} + \|w\|_{2,[0,n-1]}} \tag{120}$$

where h_2 is the space of finite energy signals and Π_0 is a positive definite matrix that reflects *a priori* knowledge as to how close $X(0)$ is to the initial guess $\widehat{X}(0)$.

Theorem 3 *[11]For a given $\gamma > 0$, if the $\mathbf{A}(k)$ are nonsingular then an estimator with $\|T_f\|_\infty \leq \gamma$ exists if, and only if,*

$$\mathbf{P}^{-1}(k) + C_2^*(k)C_2(k) - \gamma^{-2}C_1^*(k)C_1(k) > 0 \qquad ; k = 0, 1, ..., n \tag{121}$$

where $\mathbf{P}(0) = \Pi_0$ and $\mathbf{P}(k)$ satisfies the Riccati recursion

$$\mathbf{P}(k+1) = \mathbf{A}(k)\mathbf{P}(k)\mathbf{A}^*(k) + B_1(k)B_1^*(k) - \tag{122}$$

$$\mathbf{A}(k)\mathbf{P}^*(k)\left[C_1^*(k) \quad C_2^*(k)\right] \left\{ \begin{bmatrix} -\gamma^2 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} C_1(k) \\ C_2(k) \end{bmatrix} \mathbf{P}(k)\left[C_1^*(k) \quad C_2^*(k)\right] \right\}^{-1} \begin{bmatrix} C_1(k) \\ C_2(k) \end{bmatrix} \mathbf{P}(k)\mathbf{A}^*(k) \tag{123}$$

Then one possible H_∞ filter with level γ is given by $\hat{v}_f(n/n) = C_1(n)\widehat{X}(n/n)$ where

$$\widehat{X}(k+1/k+1) = \mathbf{A}(k)\widehat{X}(k/k)+K_f(k)(d(k+1)-C_2(k+1)\mathbf{A}(k)\widehat{X}(k/k)); \widehat{X}(0/0) \quad (124)$$

$$K_f(k) = \mathbf{P}(k+1)C_2^*(k+1)\left(1 + C_2(k+1)\mathbf{P}(k+1)C_2^*(k+1)\right)^{-1} \quad (125)$$

Theorem 4 *[11]For a given $\gamma > 0$, if the $\mathbf{A}(k)$ are nonsingular then an estimator with $\|T_p\|_\infty \leq \gamma$ exists if, and only if,*

$$\tilde{\mathbf{P}}^{-1}(k) = \mathbf{P}^{-1}(k) - \gamma^{-2}C_1^*(k)C_1(k) > 0 \qquad ;k = 0,1,...,n \quad (126)$$

where $\mathbf{P}(k)$ is same as in Theorem 3. One possible H_∞ estimator with level γ is given by $\hat{v}_p(n) = C_1(n)\widehat{X}(n)$ where

$$\widehat{X}(k+1) = \mathbf{A}(k)\widehat{X}(k) + K_p(k)\left(d(k) - C_2(k)\widehat{X}(k)\right) \qquad ;\widehat{X}(0) \quad (127)$$

$$K_p(k) = \mathbf{A}(k)\tilde{\mathbf{P}}(k)C_2^*(k)\left(1 + \left(C_2(k)\tilde{\mathbf{P}}(k)C_2^*(k)\right)\right)^{-1}$$

6.2 Adaptive H_∞ Filters

This section uses Theorems 3 and 4 to derive the optimal and sub-optimal adaptive filters which satisfy H_∞ criteria. Consider an adaptive filtering problem of Figure 17.

A given known signal $S_{in}(n) = [s(0) .. s(n-l)\ s(n-1) ... s(n)]$ with only past l samples at time n represented as $S(n) = [s(n-l)\ s(n-1) ... s(n)]$ is passed through an unknown system $W_{wig} = [w_0 w_1 ... w_l]$ and we observe the output $D_{out}(n) = [d(0)\ d(1)\ ...\ d(n)]$

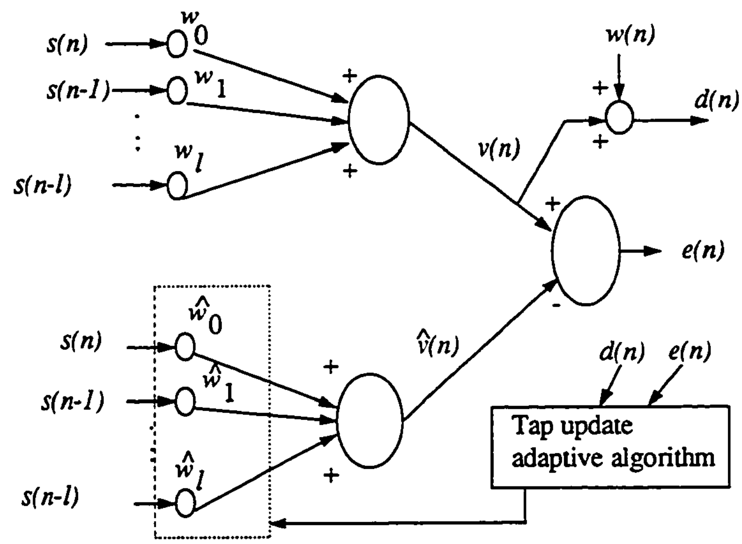

Figure 17: Example of adaptive filter problem.

system $W_{wig} = [w_0 w_1 \, ... w_l]$ and we observe the output $D_{out}(n) = [d(0)\ d(1)\ ...\ d(n)]$ corrupted by noise $W_{noise}(n) = [w(0)\ w(1)\ ...\ w(n)]$. It is desired to estimate W_{wig} with $\widehat{W}_{wig}(n)$ by minimizing the output error sequence $E_{err}(n) = [e(0)\ e(1)\ ...\ e(n)]$ where $v(n) = S(n)W^*_{wig}, \widehat{v}(n) = S(n)\widehat{W}^*_{wig}(n)$, and $e(n) = v(n) - \widehat{v}(n)$. Alternately, the problem can be formulated in the state space form as follows:

$$X(k+1) = X(k) \qquad ; X(0) = W^t_{wig} \tag{128}$$

$$d(k) = S(k)X(k) + w(k) \qquad k = 0, 1, ..., n$$

$$\widehat{v}(k) = S(k)\widehat{W}^*_{wig}(k) \tag{129}$$

To minimize the H_∞ norm of the transfer function from inputs $W_{wig} - \widehat{W}_{wig}(0)$ and $W_{noise}(n)$ to the output error $(e(n))$ we should find solution to the following:

$$\min_F \|\mathbf{T}\|_\infty = \min_F \sup_{W_{wig}, w(n) \in h_2} \frac{\|e\|_{2,[0,n]}}{(W_{wig} - \widehat{W}_{wig}(0))\mathbf{P}(0)^{-1}(W_{wig} - \widehat{W}_{wig}(0))^* + \|w\|_{2,[0,n]}} \tag{130}$$

where $\widehat{v}(n) = F(d(0)\ d(1)\ ...\ d(n))$ is the optimal H_∞ estimation strategy and $\mathbf{P}(0)$ is a positive definite matrix that reflects *a priori* knowledge as to how close W_{wig} is to initial guess $\widehat{W}_{wig}(0)$. In other words it weights the difference in the estimate at $n = 0$ and W_{wig}.

Theorem 5 *From (115), (128) and Theorem 3 it can be shown that solution to (130) is* *given by:*

$$\mathbf{P}(k+1) = \mathbf{P}(k) - \frac{\mathbf{P}(k)S^*(k)S(k)\mathbf{P}(k)}{(1+S(k)\mathbf{P}(k)S^*(k))} \left[\frac{1-\gamma^2}{(-\gamma^2+S(k)\mathbf{P}(k)S^*(k)) - (1+S(k)\mathbf{P}(k)S^*(k))^{-1}(S(k)\mathbf{P}(k)S^*(k))^2}\right] \quad (131)$$

where $\mathbf{P}(0) = \mu\mathbf{I}$ *and* μ *is a constant.*

$$\widehat{W}^*_{wig}(k+1) = \widehat{W}^*_{wig}(k) + \mathbf{P}(k+1)S^*(k+1) (1+S(k+1)\mathbf{P}(k+1)S^*(k+1))^{-1}\left(d(k+1) - S(k+1)\widehat{W}^*_{wig}(k)\right) \quad (132)$$

Proof:

from (128) and Theorem 3 $\mathbf{A}(k) = \mathbf{I}$, $B_1(k) = 0$, $C_2(k) = S(k)$, and $C_1(k) = S(k)$.

Therefore, (122) reduces to:

$$\mathbf{P}(k+1) = \mathbf{P}(k) - \mathbf{P}(k)\left[S^*(k) \quad S^*(k)\right] \quad (133)$$

$$\left\{\begin{bmatrix} -\gamma^2 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} S(k) \\ S(k) \end{bmatrix} \mathbf{P}(k)\left[S^*(k) \quad S^*(k)\right]\right\}^{-1} \begin{bmatrix} S(k) \\ S(k) \end{bmatrix} \mathbf{P}(k)$$

$$\mathbf{TM}(k) = \left\{\begin{bmatrix} -\gamma^2 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} S(k) \\ S(k) \end{bmatrix} \mathbf{P}(k)\left[S^*(k) \quad S^*(k)\right]\right\}^{-1}$$

$$= \begin{bmatrix} -\gamma^2 + S(k)\mathbf{P}(k)S^*(k) & S(k)\mathbf{P}(k)S^*(k) \\ S(k)\mathbf{P}(k)S^*(k) & 1+S(k)\mathbf{P}(k)S^*(k) \end{bmatrix}^{-1}$$

$$= (1/te(k)) \begin{bmatrix} 1+S(k)\mathbf{P}(k)S^*(k) & -S(k)\mathbf{P}(k)S^*(k) \\ -S(k)\mathbf{P}(k)S^*(k) & -\gamma^2+S(k)\mathbf{P}(k)S^*(k) \end{bmatrix}^*$$

where $te(k) = (-\gamma^2 + S(k)\,\mathbf{P}(k)\,S^*(k))(1 + S(k)\,\mathbf{P}(k)\,S^*(k)) - (\,S(k)\,\mathbf{P}(k)\,S^*(k))^2$

$$= (1/te(k)) \begin{bmatrix} 1+S(k)\mathbf{P}^*(k)S^*(k) & -S(k)\mathbf{P}^*(k)S^*(k) \\ -S(k)\mathbf{P}^*(k)S^*(k) & -\gamma^2+S(k)\mathbf{P}^*(k)S^*(k) \end{bmatrix}$$

Therefore,

$$\mathbf{P}(k)\,[S^*(k) \quad S^*(k)]\,\mathbf{TM}(k) \begin{bmatrix} S(k) \\ S(k) \end{bmatrix} \mathbf{P}(k)$$

$$= \mathbf{P}(k)\,[S^*(k) \quad S^*(k)]\,(1/te(k))$$

$$\begin{bmatrix} S(k)\mathbf{P}(k)+S(k)\mathbf{P}^*(k)S^*(k)S(k)\mathbf{P}(k)-S(k)\mathbf{P}^*(k)S^*(k)S(k)\mathbf{P}(k) \\ -S(k)\mathbf{P}^*(k)S^*(k)S(k)\mathbf{P}(k)-\gamma^2S(k)P(k)+S(k)\mathbf{P}^*(k)S^*(k)S(k)\mathbf{P}(k) \end{bmatrix}$$

$$= \frac{\mathbf{P}(k)}{te(k)}[\,S^*(k)S(k) + \,S^*(k)\,S(k)\,\mathbf{P}^*(k)\,S^*(k)\,S(k) -$$

$$S^*(k)\,S(k)\,\mathbf{P}^*(k)\,S^*(k)\,S(k) -$$

$$S^*(k)\,S(k)\,\mathbf{P}^*(k)\,S^*(k)\,S(k) - \,S^*(k)\gamma^2S(k) +$$

$$S^*(k)\,S(k)\,\mathbf{P}^*(k)\,S^*(k)\,S(k)]\,\mathbf{P}(k)$$

$$= \frac{\mathbf{P}(k)}{te(k)}\,S^*(k)[1 + \,S(k)\,\mathbf{P}^*(k)\,S^*(k) - \,S(k)\,\mathbf{P}^*(k)\,S^*(k) - \,S(k)\,\mathbf{P}^*(k)$$

$$S^*(k) - \gamma^2 + \,S(k)\,\mathbf{P}^*(k)\,S^*(k)]\,S(k)\,\mathbf{P}(k)$$

$$= \frac{\mathbf{P}(k)S^*(k)S(k)\mathbf{P}(k)}{te(k)}[1-\gamma^2] = \frac{\mathbf{P}(k)S^*(k)S(k)\mathbf{P}(k)}{(1+S(k)\mathbf{P}(k)S^*(k))} \tag{134}$$

$$\left[\frac{1-\gamma^2}{(-\gamma^2+S(k)\mathbf{P}(k)S^*(k))-(1+S(k)\mathbf{P}(k)S^*(k))^{-1}(S(k)\mathbf{P}(k)S^*(k))^2}\right]$$

From (134) and (133)

$$\mathbf{P}(k+1) = \mathbf{P}(k) - \frac{\mathbf{P}(k)S^*(k)S(k)\mathbf{P}(k)}{(1+S(k)\mathbf{P}(k)S^*(k))} \left[\frac{1-\gamma^2}{(-\gamma^2+S(k)\mathbf{P}(k)S^*(k))-(1+S(k)\mathbf{P}(k)S^*(k))^{-1}(S(k)\mathbf{P}(k)S^*(k))^2}\right] \tag{135}$$

∎

From (135) following interpretations can be drawn:

* $\gamma = 1 \Rightarrow \mathbf{P}(k+1) = \mathbf{P}(k) = \mu\mathbf{I}$, $(\mathbf{P}(0) = \mu\mathbf{I})$ and the recursive form of $\widehat{W}_{wig}(k+1)$ in (132) collapses to update form similar to NLMS:

$$\widehat{W}^*_{wig}(k+1) = \widehat{W}^*_{wig}(k) + \frac{\mu S^*(k+1)}{1+S(k+1)\mu S^*(k+1)}\left[d(k+1) - S(k+1)\widehat{W}^*_{wig}(k)\right] \tag{136}$$

* $\gamma \to \infty \Rightarrow \mathbf{P}(k+1) = \mathbf{P}(k) - \frac{\mathbf{P}(k)S^*(k)S(k)\ \mathbf{P}(k)}{1+S^*(k)\ \mathbf{P}(k)S(k)}$ and the recursive form of $\widehat{W}_{wig}(k+1)$ in (132) collapses exactly to update form of RLS [14]:

$$\widehat{W}^*_{wig}(k+1) = \widehat{W}^*_{wig}(k) + \frac{\mathbf{P}(k)S^*(k+1)}{1+S(k+1)\mathbf{P}(k)S^*(k+1)}\left[d(k+1) - S(k+1)\widehat{W}^*_{wig}(k)\right] \tag{137}$$

* $1 < \gamma < \infty$ recursive form of $\widehat{W}_{wig}(k+1)$ in (132) satisfy sub-optimal H_∞ criteria

and fall into new category of adaptive algorithms which we call it sub-optimal NLMS adaptive algorithms.

* From (115), (128), Theorem 4, $\gamma = 1$, and $e(n) = v(n) - \widehat{v}(n)$ it can be shown that recursion $\widehat{W}_{wig}(k+1)$ (represented as $\widehat{X}(k+1)$ in Theorem 4) collapses exactly to update equation of LMS [11]:

$$\widehat{W}^*_{wig}(k+1) = \widehat{W}^*_{wig}(k) + \mu S^*(k)\left[d(k) - S(k)\widehat{W}^*_{wig}(k)\right] \tag{138}$$

Also, in case of LMS when $1 < \gamma < \infty$ recursive form of $\widehat{W}_{wig}(k+1)$ in (138) satisfy sub-optimal H_∞ criteria and fall into new category of adaptive algorithms which we call it sub-optimal LMS adaptive algorithms.

It is shown in [11] that $\gamma = 1$ is the optimal H_∞ solution to both filtering and estimation problem of (130). From above interpretations it is clear that both NLMS and LMS are optimal H_∞ solution to adaptive filtering problem of (128) and RLS is a special case of NLMS when $\gamma \to \infty$. This implies that there is no upper bound guaranteed for RLS or the H_∞ bound is quite large suggesting poor robust properties with respect to inputs,i.e., noise and unknown weights. On the other hand LMS and NLMS have a finite H_∞ norm equal to one. This guarantees that the energy of mean square error in LMS and NLMS cannot exceed the energy of inputs suggesting superior robustness properties of these algorithms with respect to tap weights and input noise.

Now we connect the above state space solution to the RLS solution [14] to study the similarities and dissimilarities in the above adaptive algorithms. We do this by simplifying $\mathbf{P}(k)$ further. From (133) and using inversion lemma:

$$\begin{aligned}
\mathbf{P}^{-1}(k+1) &= \mathbf{P}^{-1}(k) + [S^*(k) \qquad S^*(k)] \begin{bmatrix} -\gamma^2 I & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} S(k) \\ S(k) \end{bmatrix} \\
&= \mathbf{P}^{-1}(k) + (1-\gamma^{-2}) S^*(k) S(k) \\
&= \mu^{-1}\mathbf{I} + (1-\gamma^{-2}) \sum_{k=0}^{n} S^*(k) S(k) \qquad ; \mathbf{P}(0) = \mu \mathbf{I} \qquad (139)
\end{aligned}$$

for LMS the $\tilde{\mathbf{P}}(k)$ from equation (126) and (139) is given by

$$\tilde{\mathbf{P}}^{-1}(k) = \mu^{-1}\mathbf{I} + (1-\gamma^{-2}) \sum_{j=0}^{k-1} S^*(j) S(j) - \gamma^{-2} S^*(k) S(k) \qquad ; \tilde{\mathbf{P}}(0) = \mu \mathbf{I}. \qquad (140)$$

The update equation can be made the same for all the algorithms by just varying $\mathbf{P}'(k)$ as follows:

$$\widehat{W}^*_{wig}(k+1) = \widehat{W}^*_{wig}(k) + \frac{\mathbf{P}'(k) S^*(k+1)}{1 + S(k+1)\mathbf{P}'(k) S^*(k+1)} \left[d(k+1) - S(k+1)\widehat{W}^*_{wig}(k) \right]. \qquad (141)$$

Where, $\mathbf{P}'(k) = \mathbf{P}(k)$ and $\gamma = 1$ for NLMS, $\mathbf{P}'(k) = \mathbf{P}(k)$ and $\gamma = \infty$ for RLS. Since LMS is a prediction filter therefore, (141) is modified as:

$$\widehat{W}^*_{wig}(k+1) = \widehat{W}^*_{wig}(k) + \frac{\mathbf{P}'(k) S^*(k)}{1 + S(k)\mathbf{P}'(k) S^*(k)} \left[d(k) - S(k)\widehat{W}^*_{wig}(k) \right]. \qquad (142)$$

In (142) if $\mathbf{P}'(k) = \tilde{\mathbf{P}}(k)$ and $\gamma = 1$ it becomes a LMS update equation. $\mathbf{P}'(k) = \mathbf{P}(k)$ in (141) for RLS is the correlation matrix of input data with added small μ^{-1}. This is same as the correlation matrix equation obtained in [14] by minimizing mean squared error. It is interesting to observe that for NLMS the $\mathbf{P}'(k)$ in (141) is always a constant while in (142) for LMS it is dependent on instantaneous value of autocorrelation estimate from single observation. The positive definite criteria for $\tilde{\mathbf{P}}(k)$ puts a restriction on value of μ which from (140) is given by

$$\mu < \inf_k \frac{1}{S(k)S^*(k)}. \tag{143}$$

Role of μ^{-1} is very well established for RLS and acts as the noise added to the autocorrelation estimate [14]. Larger the value of μ smaller is the noise and better is the estimate making RLS converge fast but less stable due to singularity of correlation matrix. It is interesting to see that same role is played by μ in NLMS (optimal and sub-optimal) and LMS (optimal and sub-optimal). It acts as noise to the correlation estimate of $\mathbf{P}'(k)$. Therefore, large value of μ should give faster convergence as is the case observed for both LMS and NLMS [14]. The slow convergence rate of LMS and NLMS compared to RLS is because they try to minimize an upper bound on the MSE at every point in time. On the other hand RLS try to minimize MSE and hence converge faster to minimum MSE compared to LMS and NLMS. Also, it is interesting to see that from (139) we can make RLS have finite upper bound on error variance,i.e., upper bound on H_∞ norm by weighting the correlation matrix with $(1-\gamma^{-2})$. Therefore, sub-optimal NLMS algorithms should have better robustness properties (finite upper bound on the error variance) and performance (fast convergence) close to RLS. Robustness is related to the stability of the algorithm and better (minimum MSE) estimate of unknown weights in presence of added noise and noise due to finite precision effects. This will allow us to trade off robustness with performance. We have found in our simulations that $\gamma \approx 2$ produces convergence rate similar to RLS and is more stable compared to RLS..

6.3 Simulation Results

The convergence characteristics of the LMS, NLMS, and RLS algorithms is very well understood [14]. However, from the above analysis it is much clear now that LMS and NLMS should be more robust with respect to input noise variations compared to RLS. This is because LMS and NLMS are solution to optimal H_∞ minimization and have upper bound on the error variance while RLS does not guarantee any such bounds. First, we show that the upper bound achieved by LMS is one while for RLS it is greater than one. Consider the adaptive problem as follows:

$$X(k+1) = X(k) \qquad ; X(0) = W_{wig} \tag{144}$$

$$d(k) = S(k)X^*(k) + w(k) \qquad k = 0, 1, ..., n$$

$$e(k) = S(k)\left(W_{wig} - \widehat{W}_{wig}(k)\right)^* \tag{145}$$

Computing the matrix which maps inputs $(w(k), \mu^{\frac{1}{2}}\left(W_{wig} - \widehat{W}_{wig}(k)\right))$ to the output error

$(e_{lms}(k))$ using LMS equation (138):

$$\widetilde{W}_{wig}(k) = W_{wig} - \widehat{W}_{wig}(k) = W_{wig} -$$

$$\left[\widehat{W}_{wig}(k-1) + \mu\left(d(k) - S(k)\widehat{W}^{*}_{wig}(k-1)\right)S(k)\right]$$

$$= \widetilde{W}_{wig}(k-1) - \mu\ d(k)S(k) + \mu S(k)\widehat{W}^{*}_{wig}(k-1)S(k)$$

$$= \widetilde{W}_{wig}(k-1) - \mu\ S(k)W^{*}_{wig}(k)S(k) - \mu w(k)S(k) + \mu S(k)\widehat{W}^{*}_{wig}(k-1)S(k)$$

$$= \widetilde{W}_{wig}(k-1) - \mu S(k)\widetilde{W}_{wig}(k-1)S(k) - \mu w(k)S(k)$$

$$\widetilde{W}_{wig}(1) = \widetilde{W}_{wig}(0) - \mu S(1)\widetilde{W}^{*}_{wig}(0)S(1) - \mu w(1)S(1)$$

$$\widetilde{W}_{wig}(2) = \widetilde{W}_{wig}(1) - \mu S(2)\widetilde{W}^{*}_{wig}(1)S(2) - \mu w(2)S(2)$$

$$= \left(\widetilde{W}_{wig}(0) - \mu S(1)\widetilde{W}^{*}_{wig}(0)S(1) - \mu w(1)S(1)\right) - \mu S(2)$$

$$\left(\widetilde{W}_{wig}(0) - \mu S(1)\widetilde{W}^{*}_{wig}(0)S(1) - \mu w(1)S(1)\right)S(2) - \mu w(2)S(2)$$

The equation becomes very messy with $l > 2$ so we analyze the matrix for $l = 1$ case.

Since $l = 1$ is a scalar case therefore, $w_{wig} = W_{wig}$, $w_{wig} \in R^1$ and $s(n) = S(n)$.

$$\tilde{w}_{wig}(2) = \tilde{w}_{wig}(0) - \mu s^2(1)\tilde{w}_{wig}(0) - \mu s^2(2)\tilde{w}_{wig}(0) + \mu^2 s^2(2)s^2(1)\tilde{w}_{wig}(0) -$$

$$\mu w(1)s(1) - \mu^2 s^2(2)s(1)w(1) - \mu w(2)s(2)$$

for $k = n$ and $s(k) = \pm 1$

$$\tilde{w}_{wig}(n) = (1-\mu)^n\,\tilde{w}_{wig}(0) - \mu(1-\mu)^{n-1}s(1)w(1) -$$

$$\mu(1-\mu)^{n-2}s(2)w(2)... - \mu s(n)w(n)$$

$$e_{lms}(n) = \tilde{w}_{wig}(n)s(n)$$

If we observe n points of data than the transfer matrix which maps inputs

$(w(k), \mu^{\frac{1}{2}}\tilde{w}_{wig}(k))$ to output error $e_{lms}(k)$ is given as:

$$\begin{bmatrix} \mu^{\frac{1}{2}}(1-\mu)\,s(1) & -\mu & 0 & 0 \\ \mu^{\frac{1}{2}}(1-\mu)^2\,s(2) & -\mu\,(1-\mu)\,s(1)s(2) & -\mu & 0 \\ \mu^{\frac{1}{2}}(1-\mu)^3\,s(3) & -\mu\,(1-\mu)^2\,s(1)s(3) & -\mu\,(1-\mu)\,s(2)s(3) & 0 \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ \mu^{\frac{1}{2}}(1-\mu)^n\,s(n) & -\mu\,(1-\mu)^{n-1}\,s(1)s(n-1) & -\mu\,(1-\mu)^{n-2}\,s(2)s(n-1) & -\mu \end{bmatrix} \tag{146}$$

.

similarly we can obtain transfer matrix from RLS update equations ($l = 1$, $s(k) = \pm 1$).

From (137) ($p(k) = \mathbf{P}(k)$)

$\widehat{w}_{wig}(k+1) = \widehat{w}_{wig}(k) + \frac{p(k)s(k+1)}{1+s(k+1)p(k)s(k+1)}\left[d(k+1) - s(k+1)\widehat{w}_{wig}(k)\right]$

$= \widehat{w}_{wig}(k) + \frac{p(k)s(k+1)}{1+p(k)}\left[d(k+1) - s(k+1)\widehat{w}_{wig}(k)\right]$

$p(k+1) = p(k) - \frac{p(k)s(k)s(k)\ p(k)}{1+s(k)\ p(k)s(k)} = p(k) - \frac{p^2(k)s^2(k)}{1+p(k)s^2(k)} = \frac{p(k)}{1+p(k)} = \frac{p(0)}{1+(k+1)p(0)}$

$\widehat{w}_{wig}(k+1) = \widehat{w}_{wig}(k) + p(k+1)s(k+1)\left[d(k+1) - s(k+1)\widehat{w}_{wig}(k)\right]$

$\tilde{w}_{wig}(k) = w_{wig} - \widehat{w}_{wig}(k) = w_{wig} - \widehat{w}_{wig}(k-1) - p(k)s(k)$

$\left[s(k)w_{wig}(k) + w(k) - s(k)\widehat{w}_{wig}(k-1)\right]$

$= \tilde{w}_{wig}(k-1) - p(k)w_{wig} - p(k)s(k)w(k) + p(k)\widehat{w}_{wig}(k-1)$

$= \widetilde{w}_{wig}(k-1) - p(k)\widetilde{w}_{wig}(k-1) - p(k)s(k)w(k) = (1-p(k))\,\widetilde{w}_{wig}(k-1) - p(k)s(k)w(k)$

$= \left(\frac{1+(k-1)p(0)}{1+kp(0)}\right)\widetilde{w}_{wig}(k-1) - \frac{p(0)}{1+kp(0)}s(k)w(k)$

$\widetilde{w}_{wig}(1) = \left(\frac{1}{1+p(0)}\right)\widetilde{w}_{wig}(0) - \frac{p(0)}{1+p(0)}s(1)w(1)$

$\widetilde{w}_{wig}(2) = \left(\frac{1+p(0)}{(1+2p(0))}\right)\widetilde{w}_{wig}(1) - \frac{p(0)}{(1+2p(0))}s(2)w(2)$

$= \left(\frac{1+p(0)}{(1+2p(0))}\right)\left(\left(\frac{1}{1+p(0)}\right)\widetilde{w}_{wig}(0) - \frac{p(0)}{1+p(0)}s(1)w(1)\right) - \frac{p(0)}{(1+2p(0))}s(2)w(2)$

$= \frac{1}{1+2p(0)}\widetilde{w}_{wig}(0) - \frac{p(0)}{1+2p(0)}s(1)w(1) - \frac{p(0)}{1+2p(0)}s(2)w(2)$

$\widetilde{w}_{wig}(n) = \frac{1}{1+nP(0)}\widetilde{w}_{wig}(0) - \frac{p(0)}{1+np(0)}s(1)w(1) - \frac{p(0)}{1+np(0)}s(2)w(2)... - \frac{p(0)}{1+np(0)}s(n)w(n)$

$e_{rls}(n) = \widetilde{w}_{wig}(n)s(n) = \frac{s(n)}{1+np(0)}\widetilde{w}_{wig}(0) - \frac{p(0)}{1+np(0)}s(1)s(n)w(1)$

$-\frac{p(0)}{1+np(0)}s(2)s(n)w(2)... - \frac{p(0)}{1+np(0)}w(n)$

Therefore, the transfer matrix for RLS which maps inputs $(w(k), \mu^{\frac{1}{2}}\widetilde{w}_{wig}(k))$ to the output error $e_{rls}(k)$ is given by:

$$\begin{bmatrix} \mu^{\frac{1}{2}}\frac{s(1)}{1+p(0)} & -\frac{p(0)}{1+p(0)} & 0 & & \cdots & 0 \\ \mu^{\frac{1}{2}}\frac{s(2)}{1+2p(0)} & -\frac{p(0)}{1+2p(0)}s(1)s(2) & -\frac{p(0)}{1+2p(0)} & & \cdots & 0 \\ \mu^{\frac{1}{2}}\frac{s(3)}{1+3p(0)} & -\frac{p(0)}{1+3p(0)}s(1)s(3) & -\frac{p(0)}{1+3p(0)}s(2)s(3) & & \cdots & 0 \\ . & . & . & & \cdots & 0 \\ . & . & . & & \cdots & 0 \\ . & . & . & & \cdots & 0 \\ \mu^{\frac{1}{2}}\frac{s(n)}{1+np(0)} & -\frac{p(0)}{1+np(0)}s(1)s(n) & -\frac{p(0)}{1+np(0)}s(2)s(n) & ...... \quad .. & ..... & -\frac{p(0)}{1+np(0)} \end{bmatrix} \tag{147}$$

Maximum singular values of the transfer matrix for LMS and RLS are plotted in Figure 18, Figure 19, and Figure 20. From (143) value of μ permissible is close to one. Therefore, $\mu = .95$ and $p(0) = 10, 100$. It is seen that maximum singular values for LMS does not exceed one while RLS is greater than one. This is expected as LMS is guaranteed to be always below one as it satisfies H_∞ criteria (with upper bound on singular value of one) while RLS has no such upper bound. It is interesting to observe that upper bound on the maximum singular value increases with increase in $p(0)$ as shown in Figure 19 and Figure 20 for $p(0) = 10$ and $p(0) = 100$ respectively. This means that with increase in $p(0)$ upper bound on the maximum singular value is increased making RLS less robust. In [14] it is shown that $\frac{1}{p(0)}$ is the noise added to the correlation matrix. Therefore, increasing the noise level makes RLS more robust which agrees with the result of [14].

For performance comparison of sub-optimal algorithms with RLS following simulation parameters were chosen:

$$l = 15, \qquad w_i = i; i = 1, 2, ..., 15, \qquad SNR = 30dB, \qquad \mathbf{P}(0) = 100\mathbf{I}. \qquad (148)$$

Input $S(n)$ is a colored signal obtained by passing unit variance gaussian distributed white noise through a second order autoregressive (AR) process with transfer function $H(z) = \frac{1}{1-1.4z^{-1}+.85z^{-2}}$. Noise $W_{noise}(n)$ is white gaussian and $\gamma = 2$ for sub-optimal NLMS. $\mathbf{P}(0)$ value is kept same for NLMS, sub-optimal NLMS, and RLS. It is observed that with $\gamma \approx 2$ the MSE convergence rate of RLS and sub-optimal NLMS is almost same

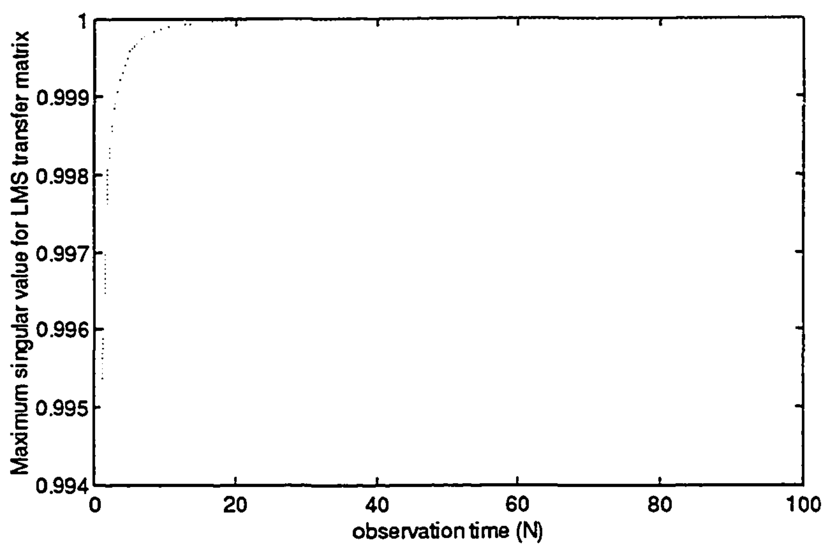

Figure 18: Maximum singular value plot of LMS equation transfer matrix (μ = .99).

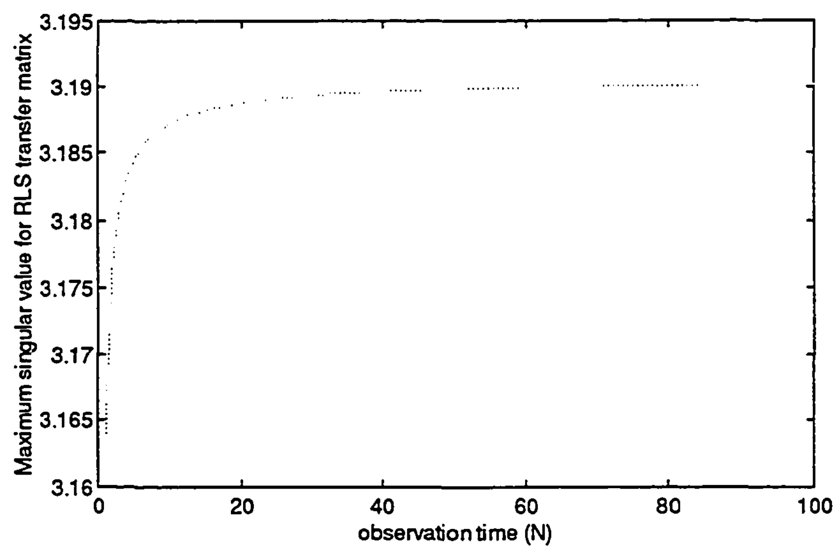

Figure 19: Maximum singular value plot of RLS equation transfer matrix ($p(0) = 10$).

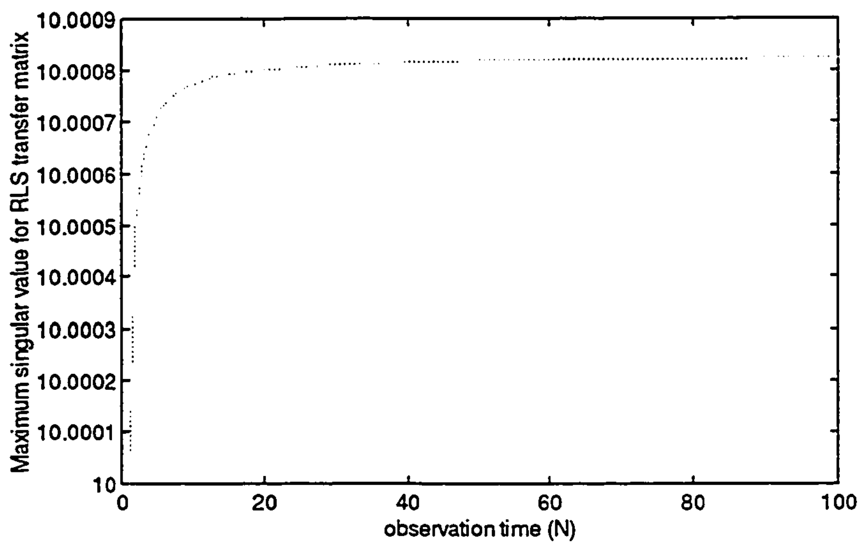

Figure 20: Maximum singular value plot of RLS equation transfer matrix ($p(0) = 100$).

as shown in Figure 21. It is clear from the convergence plots that sub-optimal NLMS have less variance in the MSE compared to RLS. The convergence rate of NLMS ($\gamma = 1$) is shown in Figure 22 and is inferior compared to RLS and sub-optimal NLMS.

6.4 Acoustic Echo Canceller

Acoustic echo produced in a teleconferencing system degrades the quality of voice and in some cases make it impossible to achieve reliable communication. The echo is produced due to acoustical coupling between speaker and microphone. Figure 23 shows a typical teleconferencing system. Speech $s(n)$ from far end speaker is passed through a loud speaker (LS). Some part of $s(n)$ is picked up by microphone (MIC) due to acoustic echo path between LS and MIC. Therefore, the outgoing speech $d(n)$ not only contain near end speech $p(n)$ and added noise $w(n)$ but also has portion of $s(n)$ passed through an acoustical echo path. $d(n)$ can be represented as:

$$d(n) = W_{wig}(n)S(n) + p(n) + w(n) \tag{149}$$

where $W_{wig}(n) = [w_0(n)w_1(n) \ ...w_l(n)]$ is the FIR filter model for the acoustical echo path at time n and $S(n) = [s(n-l) \ s(n-1) \ ...s(n)]^t$. In order to have reliable communication a portion of $d(n)$ represented by $W_{wig}(n)S(n)$ should be removed from $d(n)$ before it is delivered to the far end speaker. This can be achieved by Acoustic Echo Canceller (AEC).

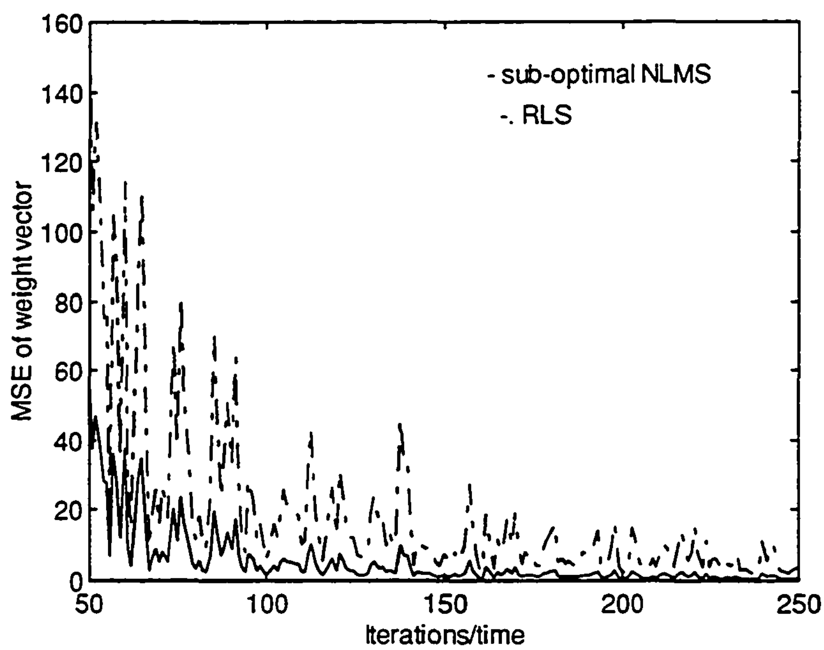

Figure 21: Convergence plots of MSE using RLS and suboptimal NLMS.

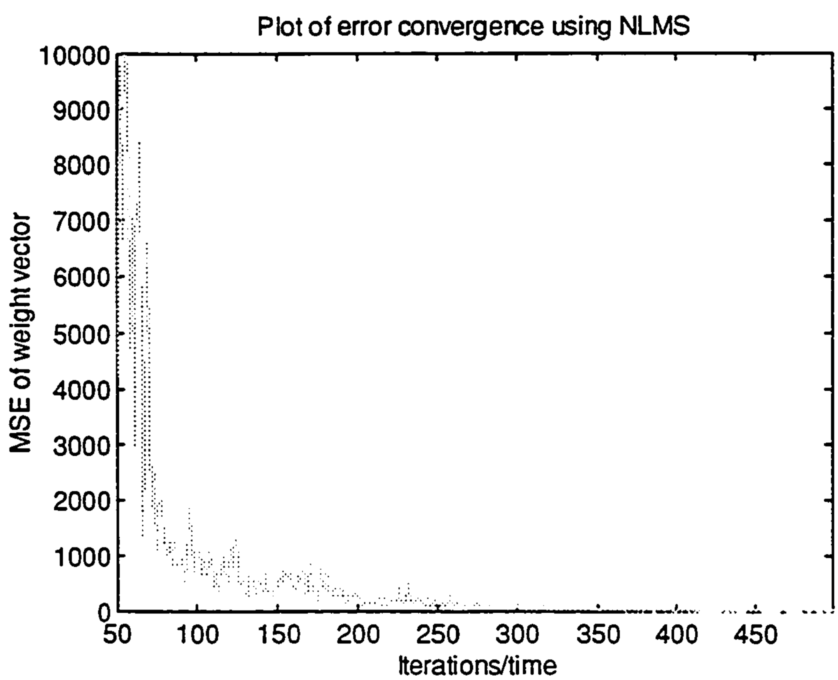

Figure 22: Convergence plot of MSE using NLMS.

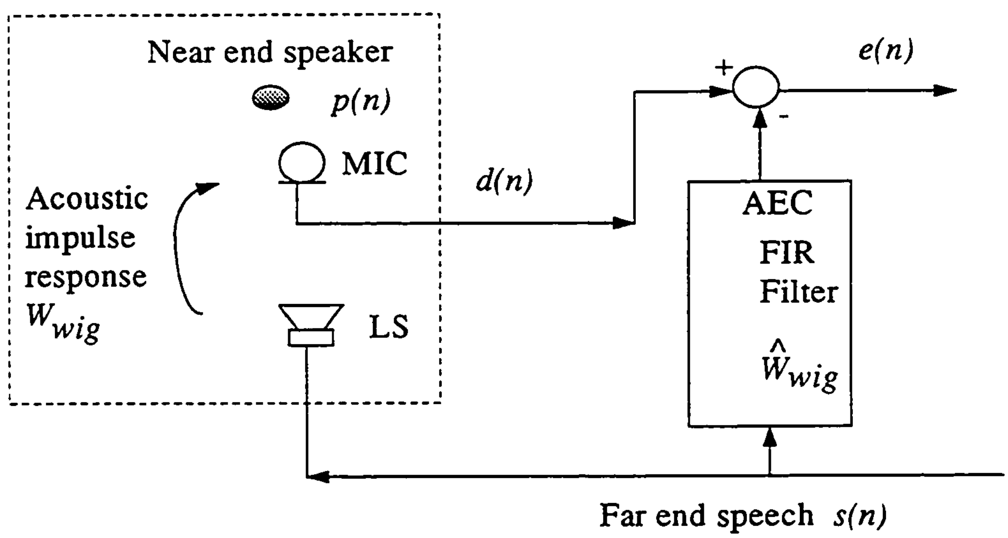

Figure 23: Echo produced in a teleconferencing system. AEC is used to cancel echo produced due to acoustic coupling between MIC and speaker.

The task of AEC is to estimate $W_{wig}(n)$ from input signal $s(n)$ and observed signal $d(n)$ as shown in Figure 23.

Adaptive filters (LMS, NLMS, RLS) are widely used as AEC [8][14]. The two desirable properties of an adaptive filter for AEC are fast convergence (performance) and good stability (robustness). As discussed in section 6.2 LMS and NLMS satisfy H_∞ criteria and therefore, have slow convergence but good stability. On the other hand RLS which satisfy minimum variance criteria have fast convergence but poor stability. Also, sub-optimal NLMS can be used to trade off convergence with stability.

Performance of AEC is evaluated with respect to adaptive algorithms satisfying H_∞ and minimum variance criteria. Far end speech $s(n)$ is passed through a 256 tap FIR filter model of an acoustical impulse response. This impulse response is obtained by passing white noise through a room and observing the output signal. Far end speech is obtained by passing an analog speech through analog-to-digital (A/D) converter with 16 bits of resolution and 8khz sampling rate. $d(n)$ is obtained by passing $s(n)$ through acoustical impulse response and adding white noise $w(n)$ to it, i.e., $d(n) = W_{wig}S(n) + w(n)$. Adaptive algorithms are used as AEC to estimate the unknown impulse response W_{wig} as $\widehat{W}_{wig}$ with $l = 256$. Most of the applications require real time implementation of RLS and NLMS. High speed DSP's are low cost solution for real time implementation of AEC. Using DSP's it is possible to implement NLMS in real time. However, due to high computational requirement (order l^3) of RLS it cannot be implemented in real time on DSP's. There are many algorithms present which reduces the computation complexity of RLS to order $2l$ [8][18]. In our simulations Affine Projection Algorithm (APA) is been used [8]. For RLS $\mathbf{P}(0) = .01\ eng(s(n))$, where $eng(s(n)) = \sum_{n=0}^{k} s(n)s^*(n)$ is the energy estimate of the input speech signal based on the window of size $k = 40,000$. Convergence plots of MSE is obtained by taking ensemble average over 10 different speech signals. Figure 24 shows the convergence rate of MSE using RLS, NLMS, and sub-optimal NLMS ($\gamma = 2$) operating with $SNR = 30\ dB$. The inferior convergence rate of NLMS compared to RLS is reflected in Figure 24. This reflects the inferior performance of NLMS compared to RLS under high SNR condition. RLS showed some unstable behavior after 15000 iterations as shown in Figure 25. This shows the poor stability property of RLS compared to NLMS. Sub-optimal NLMS has a faster convergence than NLMS and a better stability than RLS as seen in Figure 24 and 25. This observation is in accordance with the theory discussed in Section 6.2. Figure 26 shows the convergence plots for RLS, NLMS, and sub-optimal NLMS under low SNR condition ($SNR = 0dB$). Convergence rates for all the three algorithms under low SNR are similar to the ones observed for high SNR case. However, the steady state MSE achieved is increased under low SNR case for all the three algorithms.

From the simulation results on AEC it is clear that NLMS show poor convergence and performance compared to RLS under high SNR condition. From stability (robustness) point of view NLMS is better then RLS. Sub-optimal NLMS is a trade off between stability and convergence rate. It is seen to have better stability than RLS and better convergence rate than NLMS.

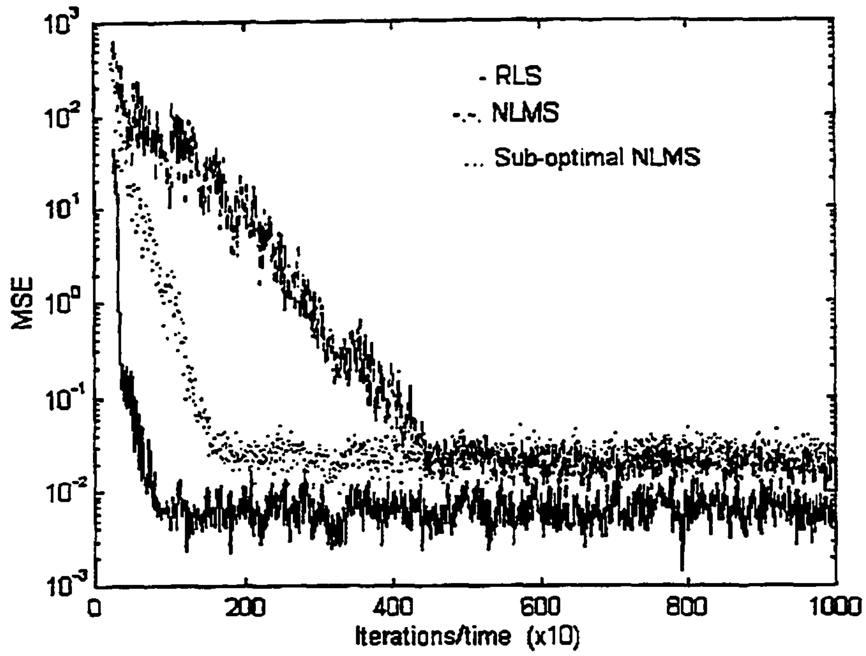

Figure 24: Convergence plots of MSE using RLS, NLMS, and sub-optimal NLMS algorithm. $SNR = 30dB$.

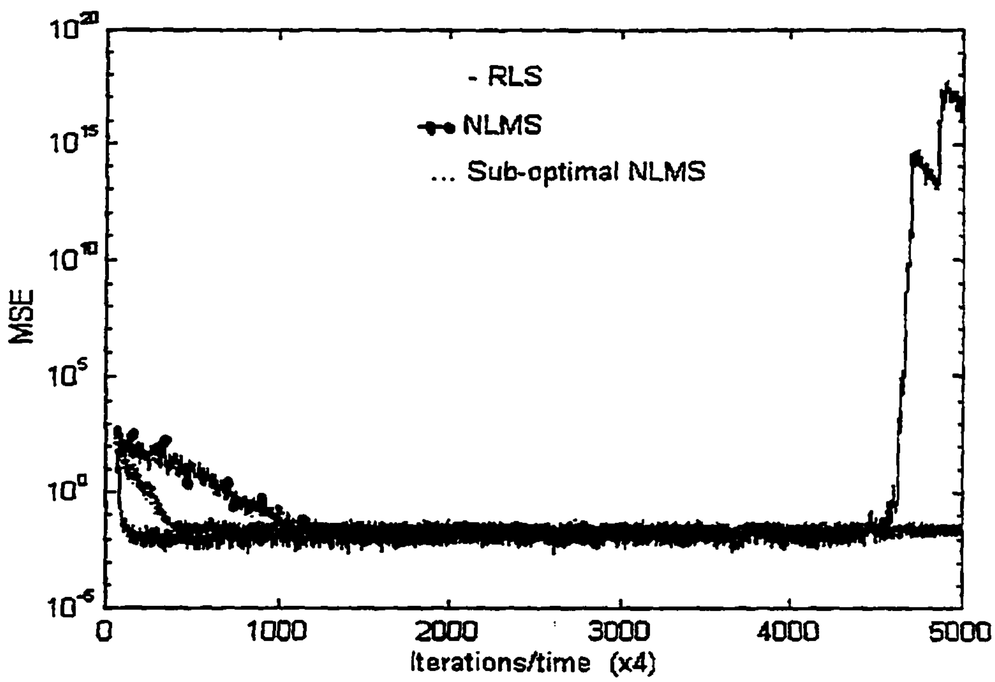

Figure 25: Convergence plots of MSE using RLS, NLMS, and sub-optimal NLMS. RLS shows unstable behavior after 15000 iterations. $SNR = 30dB$.

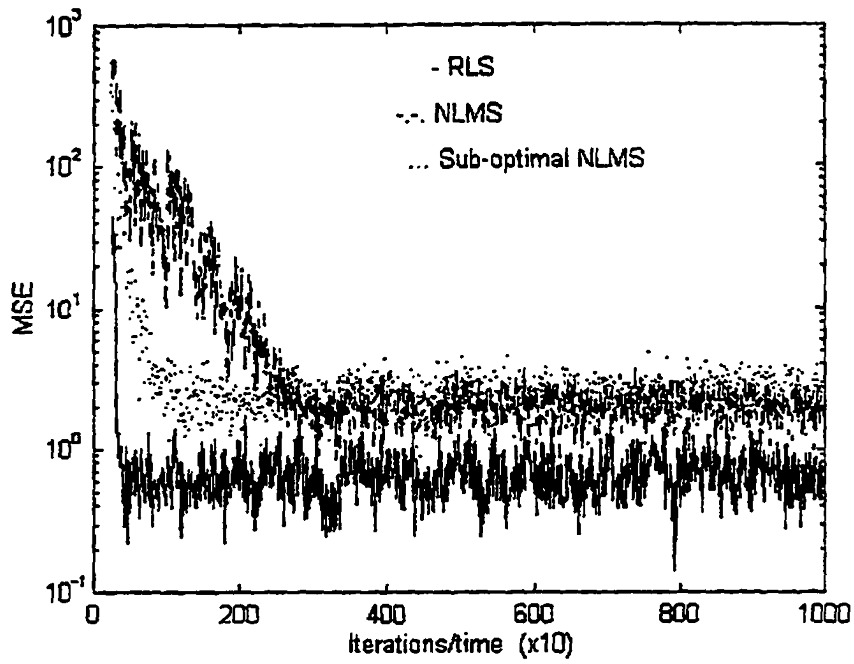

Figure 26: Convergence plots of MSE using RLS, NLMS, and sub-optimal NLMS algorithm. $SNR = 0dB$.

CHAPTER 7

CONCLUSION

Filters based on H_∞ criteria are useful in situations when input signals are not known completely. In situations where error is to be minimized in specific frequency bands then H_∞ filters are more suitable compared to minimum variance (minimizes MSE) filters. H_∞ filters provide more robust performance than minimum variance filters when the underlying statistics of the input signals are not known *a priori*. However, minimum variance filters have better performance compared to H_∞ filters when complete knowledge of input signal statistics is known.

It is shown that for a general filtering problem of Figure 2 that H_∞ criteria for stochastic input signals can be represented in time and frequency domain as:

$$\|L\|_2 = \sup_{\|I(n)\|_2^2 \neq 0} \frac{\|LI(n)\|_2^2}{\|I(n)\|_2^2} = \sup_{\|I(n)\|_2^2 \neq 0} \frac{\|e_1(n)\|_2^2}{\|I(n)\|_2^2} = \|\mathbf{T}_1(z)\|_\infty^2 = \gamma_{opt}^2 \leq \gamma^2 \tag{150}$$

where $I(n) = [s(n) \quad w(n)]\,, \mathbf{T}_1(z) = W(z)\,[G(z) - G(z)C(z)H(z) \quad - H(z)N(z)]$, and L is the time domain operator which maps input WSS signal $I(n)$ to output error $e_1(n)$ and $\|X(n)\|_2^2 = E\,[X(n)X^*(n)]$. From this it is clear that the H_∞ filters provide an upper bound to the error variance for input signals with second order stationarity. Error variance is guaranteed to be below γ^2 times power of the input signals no matter what specific set of input signals are present. In other words H_∞ filters try to minimize maximum upper bound on the error variance as opposed to minimum variance filters which tries to achieve minimum MSE. Therefore, performance of H_∞ should be more robust compared to minimum variance filters when input signals are not known *a priori*. Also, when input noises are white then using proper weighting functions the error in desired frequency band can be minimized more efficiently compared to minimum variance filters.

For continuous time filtering problem it is experimentally observed that as $\gamma \rightarrow \infty$ H_∞ filters satisfy minimum variance criteria. This is in accordance with theoretical results. This suggests that γ can be adjusted to trade off robustness with performance. It is observed that at very low SNR the variation in the error variance with respect to γ is reduced .

For discrete time system it is concluded that if SNR is unknown and is varying with time then the performance of H_∞ filter is much better than Weiner filter. The degradation in performance when SNR is reduced below $0dB$ is steady for H_∞ filter but drops drastically for Wiener filter. On an average H_∞ filter show superior performance compared to Wiener filter when input SNR is time varying.

It is shown that well known LMS and NLMS algorithm satisfy H_∞ criteria as opposed to RLS which satisfies minimum variance criteria. Therefore, LMS and NLMS should be more robust with respect to the input signals and the update error in filter coefficients compared to RLS. It is shown that for one dimensional system the gain matrix of LMS and NLMS has an upper bound of one while no upper bound is guaranteed for RLS. The superior robustness of LMS and NLMS with respect to input noise and noise due to finite precision effects compared to RLS is been reported by many researchers [8][14][18]. It is been reported in [14] that minimum error achieved by LMS and NLMS is much higher than RLS when SNR is high. This fact is confirmed by our observation on better performance of Wiener filter compared to H_∞ when SNR is high. Simulation results on AEC show that NLMS has slower convergence rate (performance) compared to RLS but better stability (robustness) compared to RLS. It is shown that when $\gamma \rightarrow \infty$ NLMS converges to RLS solution and therefore γ value can be set to trade off robustness with performance. The simulation results on AEC have shown that sub-optimal NLMS have better stability than RLS and better convergence rate than NLMS and therefore is useful in situation where trade off between robustness and performance is desired.

CHAPTER 8

FUTURE WORK

This thesis has looked into the advantages of H_∞ filters when input signal statistics in not known completely. There are applications where we are faced with problems of model uncertainties,i.e., models describing channel, signal, or noise have some uncertainties associated with them as shown in the Figure 27. These uncertainties can be modelled with an H_∞ upper bound as:

$$\|\triangle_1(z)\|_\infty < \gamma_1 \quad ; \quad \|\triangle_2(z)\|_\infty < \gamma_2 \quad ; \quad \|\triangle_3(z)\|_\infty < \gamma_3 \tag{151}$$

One of the solutions to handle uncertainties is to find a filter which minimizes the following cost function:

$$\sup_{\|\triangle's\|_\infty < \gamma's} \ \sup_{\|I(n)\|_2^2 \neq 0} \frac{\|e_1(n)\|_2^2}{\|I(n)\|_2^2} = \sup_{\|\triangle's\|_\infty < \gamma's} \|\mathbf{T}_1(z)\|_\infty^2 = \gamma^2 \tag{152}$$

where $\triangle's = \triangle_1, \triangle_2$, or $\triangle_3$ and $\gamma's = \gamma_1, \gamma_2$, or γ_3 . Other quantities $\mathbf{T}_1, I(n), e_1(n)$, and $\|.\|_2^2$ have there usual meaning as in Chapter 3.

Filter obtained by minimizing (152) will be robust to both input signal and model uncertainties and therefore should provide better performance in situations where uncertainties can be modelled.

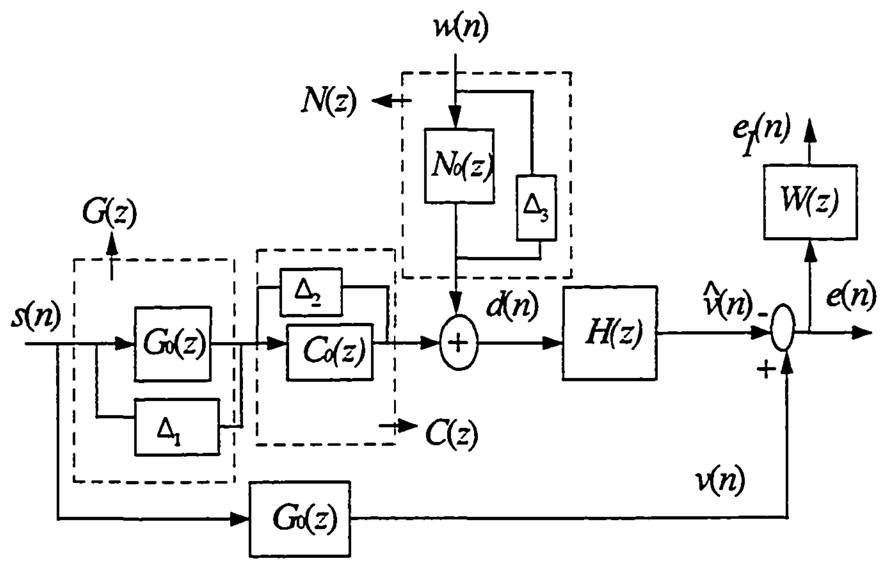

Figure 27: Filtering problem of Figure 2 with uncertainities.

REFERENCES

[1] T. Basar and P. Bernhard. *H_∞ - optimal control and related minimax design problems*. Birkhauser, New York, 1991.

[2] R. G. Brown and P. Y. C. Hwang. *Introduction to Random Signals and Applied Kalman Filtering*. John Wiely and Sons Inc., New York, 1992.

[3] J. Doyle, K. Grover, P. Khargoenkar, and B. Francis. State space solutions to standard H_2 and H_∞ control problems. *IEEE Transactions on Automatic Control*, 34:831–847, August 1989.

[4] J. C. Doyle, B. A. Francis, and A. R. Tannenbaum. *Feedback control theory*. Macmillan publishing company, New York, 1992.

[5] G. A. Einicke and L. B. White. The use of H_∞ filters for linear demodulation and equalization. *IEEE International Conference on Cicuits and Systems*, 1992.

[6] B. A. Francis. *A course in H_∞ control theory*. Springer-Verlag, New York, 1987.

[7] B. A. Francis and G. Zames. On H_∞ -optimal sensitivity theory for SISO feedback systems. *IEEE Transactions on Automatic Control*, 29:9–16, 1984.

[8] S. L. Gay and S. Tavathia. The affine projection algorithm for acoustic echo cancellation. *IEEE International Conference on Acoustic Speech and Signal Processing, Detriot, USA*, 1995.

[9] M. Grimble. Polynomial matrix solution of the H_∞ filtering problem and relationship to Riccati equation state-space results. *IEEE Transactions on Signal Processing*, 41:67–81, 1993.

[10] M. Grimble and A. Elsayed. Solution of the H_∞ optimal linear filtering problem for discrete-time systems. *IEEE Transactions on Acoustic Speech and Signal Processing (ASSP)*, 37:1092–1104, 1990.

[11] B. Hassibi, A. H. Sayed, and T. Kailath. H_∞ optimality of the LMS algorithm. *Submitted to IEEE Transations on Signal Processing*, 1993.

[12] B. Hassibi, A. H. Sayed, and T. Kailath. Kalman filtering in Krein space - part II: applications. *Submitted to IEEE Transactions on Automatic Control*, 1994.

[13] B. Hassibi, A. H. Sayed, and T. Kailath. Kalman filtering in krein spaces- part I: Theory. *Submitted to IEEE Transactions on Automatic Control*, 1994.

[14] S. Haykin. *Adaptive filter theory*. Prentice Hall, Englewood Cliffs, New Jersey, second edition, 1991.

[15] C. W. Helstrom. *Probability and Stochastic Processes for Engineers*. Macmillan Publishing Company, New York, 1991.

[16] J. M. Maciejowski. *Multivariable feedback design*. Addison-Wesley Publishing Company Inc., New York, 1989.

[17] A. Megretsky. Technical report trita/mat-92-0015. Technical report, Department of Mathematics, Royal Institute of Technology, S-100 44 Stockholm, March, 1992.

[18] G. V. Moustakides and S. Theordoridis. Fast newton transversal filters - a new class of adaptive estimation algorithms. *IEEE Transactions on Signal Processing*, 39(10), october, 1991.

[19] K. Nagpal and P. Khargoenkar. Filtering and smoothing in H_∞ setting. *IEEE Transactions on Automatic Control*, 36:153–166, 1991.

[20] I. R. Peterson. Disturbance attenuation and H_∞ optimization: A design method based on algebraic Riccati equation. *IEEE Transactions on Automatic Control*, 32:427–429, 1987.

[21] V. H. Poor. On robust Wiener filtering. *IEEE Transactions on Automatic Control*, 25(3):531–537, June 1980.

[22] J. G. Proakis. *Digital communications*. McGRAW-HILL International Editions, New York, 1989.

[23] U. Shaked. The explicit structure of inner matrices and its applications in H_∞. *IEEE Transactions on Automatic Control*, 34:734–738, 1989.

[24] U. Shaked. H_∞ minimum error state estimation of linear stationary processes. *IEEE Transactions on Automatic Control*, 35:554–558, May 1990.

[25] U. Shaked and Y. Theodor. A frequency domain approach to the problems of H_∞ -minimum error state estimation and deconvolution. *IEEE Transactions on Signal Processing*, 40:3001–3011, December 1992.

[26] A. Stoorvogel. *The H_∞ control problem, A state space approach.* Prentice Hall, New York, 1992.

[27] S. Tavathia and J. F. Doherty. Robust signal estimation using H_∞ criteria. *IEEE International Conference on Acoustic Speech and Signal Processing (ICASSP), Adelaide, Australia*, 1994.

[28] N. Wiener. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series.* Wiley, New York, 1949.

[29] V. A. Yakubovich. A linear-quadratic optimization problem and the frequency theorem for periodic systems. *Siberian Math. J.*, pages 614–630, 1986.

[30] G. Zames. Feedback and optimal sensitivity: Model reference transformations, multiplicative seminorms, and approximate inverses. *IEEE Transactions on Automatic Control*, 26:301–320, April 1981.

APPENDIX A

The Basic LQ Problem: Let **A**, **B**, **Q**, **F**, **R** be given matrices of sizes $n \times n$, $n \times m$, $n \times n$, $n \times m$, and $m \times m$ respectively, such that $\mathbf{Q} = \mathbf{Q}^*$ and $\mathbf{R} = \mathbf{R}^*$. Let $X \in C^n, U \in C^m$ and

$$\sigma(X, U) = X^*\mathbf{Q}X + 2Re\,(X^*\mathbf{F}U) + U^*\mathbf{R}U. \tag{153}$$

The following problem will be called the *basic linear quadratic optimization problem* (basic LQ problem):

given $A_0 \in C^n$, find functional $X(t)$ and $U(t)$, defined for $t \geq 0$, such that

$$X'(t) = \mathbf{A}X(t) + \mathbf{B}U(t), \quad X(0) = A_0 \tag{154}$$

$$\int_0^\infty \left\{|X(t)|^2 + |U(t)|^2\right\} dt < \infty \tag{155}$$

and the value of the integral

$$\Phi\,(X(.), U(.)) = \int_0^\infty \sigma\,(X(t), U(t))\, dt \tag{156}$$

is minimal.

The Kalman-Yakubovich Lemma: The following conditions are equivalent

(i) For any $A_0 \in C^n$ there exists a unique pair of functional $(X(.), U(.))$ which minimizes functional (156) under the constraints (154), and (155);

(ii) The pair $(\mathbf{A}, \mathbf{B})$ is stabilizable, and there exists $\epsilon > 0$ such that

$$\sigma(X, U) \geq \epsilon \left(|X|^2 + |U|^2\right) \tag{157}$$

for any $X \in C^n$, $U \in C^m$ such that

$$j\omega X = \mathbf{A}X + \mathbf{B}U \; ; \tag{158}$$

(iii) The pair $(\mathbf{A}, \mathbf{B})$ is stabilizable, and there exists $n \times n$ matrix $\mathbf{P}_0 = \mathbf{P}_0^*$ such that

$$\begin{bmatrix} \mathbf{Q} & \mathbf{F} \\ \mathbf{F}^* & \mathbf{R} \end{bmatrix} + \begin{bmatrix} \mathbf{P}_0\mathbf{A} + \mathbf{A}^*\mathbf{P}_0 & \mathbf{P}_0\mathbf{B} \\ \mathbf{B}^*\mathbf{P}_0 & 0 \end{bmatrix} > 0 \; ; \tag{159}$$

(iv) $\mathbf{R} > 0$, and there exist $n \times n$ matrix $\mathbf{P} = \mathbf{P}^*$ which is a stabilizing solution of the riccati equation

$$\mathbf{Q} + \mathbf{P}\mathbf{A} + \mathbf{A}^*\mathbf{P} = (\mathbf{P}\mathbf{B} + \mathbf{F})\,\mathbf{R}^{-1}\,(\mathbf{P}\mathbf{B} + \mathbf{F})^*, \tag{160}$$

and the matrix $\widetilde{\mathbf{A}} = \mathbf{A} - \mathbf{B}\mathbf{R}^{-1}(\mathbf{P}\mathbf{B} + \mathbf{F})^*$ is a Hurwitz matrix ;

(v) $\mathbf{R} > 0$, and there exists $n \times n$ matrix $\mathbf{P} = \mathbf{P}^*$ and a $m \times n$ matrix $\mathbf{K}$ such that

$$\sigma(X, U) + 2Re\{X^*\mathbf{P}(\mathbf{A}X + \mathbf{B}U)\} = (U - \mathbf{K}X)^*\,\mathbf{R}\,(U - \mathbf{K}X) \tag{161}$$

for any $X \in C^n, U \in C^m$, and $\mathbf{A} + \mathbf{B}\mathbf{K}$ is Hurwitz matrix;

(vi) $\mathbf{R} > 0$, the pair $(\mathbf{A}, \mathbf{B})$ is stabilizable, and the $2n \times 2n$ Hamiltonian matrix

$$\mathbf{H} = \begin{bmatrix} \mathbf{A} - \mathbf{B}\mathbf{R}^{-1}\mathbf{F}^* & \mathbf{B}\mathbf{R}^{-1}\mathbf{B}^* \\ \mathbf{Q} - \mathbf{F}\mathbf{R}^{-1}\mathbf{F}^* & -\mathbf{A}^* + \mathbf{F}\mathbf{R}^{-1}\mathbf{B}^* \end{bmatrix} \tag{162}$$

does not have eigenvalues on the imaginary axis ;

(vii) The minimal of functional (156) under conditions (154), (155) is equal to $A_0^*\mathbf{P}A_0$, and the optimal pair (X, U) is defined by (154) and by

$$U(t) = \mathbf{K}X(t) , \tag{163}$$

$$X(t) = e^{\widetilde{\mathbf{A}}t}A_0,\ U(t) = \mathbf{K}e^{\widetilde{\mathbf{A}}t}A_0 ;$$

(vii) $\mathbf{P} > \mathbf{P}_0$ for any $\mathbf{P}_0$ such that

$$\begin{bmatrix} \mathbf{Q} & \mathbf{F} \\ \mathbf{F}^* & \mathbf{R} \end{bmatrix} + \begin{bmatrix} \mathbf{P}_0\mathbf{A} + \mathbf{A}^*\mathbf{P}_0 & \mathbf{P}_0\mathbf{B} \\ \mathbf{B}^*\mathbf{P}_0 & 0 \end{bmatrix} \geq 0. \tag{164}$$

APPENDIX B

Krein Space

An abstract vector space $\{\kappa, \langle .,. \rangle_\kappa\}$ that satisfies the following requirements is called a Krein Space:

(i) κ is a linear space over C, the complex numbers.

(ii) There exists a bilinear form $\langle .,. \rangle_\kappa$ on κ such that

(a) $\langle Y, X \rangle_\kappa = \langle X, Y \rangle_\kappa^*$

(b) $\langle aX + bY, Z \rangle_\kappa = a \langle X, Z \rangle_\kappa + b \langle Y, Z \rangle_\kappa$

for any $X, Y, Z \in \kappa$, $a, b \in C$.

(iii) The vector space κ admits a direct orthogonal sum decomposition

$$\kappa = \kappa_+ + \kappa_- \tag{165}$$

such that $\{\kappa_+, \langle .,. \rangle_\kappa\}$ and $\{\kappa_-, -\langle .,. \rangle_\kappa\}$ are Hilbert spaces, and $\langle X, Y \rangle_\kappa = 0$ for any $X \in \kappa_+$ and $Y \in \kappa_-$.

In view of the above, a vector $X \in \kappa$ can be positive ($\langle X, X \rangle_\kappa > 0$), neutral ($\langle X, X \rangle_\kappa = 0$) or negative ($\langle X, X \rangle_\kappa < 0$). Correspondingly, a subspace $\iota \subset \kappa$ can be positive, neutral, or negative.