# Performance Comparison of Adaptive and Robust Predictors for Long Range Dependent Signals *

R.K. Boel
SYSTEMS Group,
Electrical Engineering Department,
Universiteit Gent,
Technologiepark 9,
Gent, Belgium. †

I.M.Y. Mareels
Department of Electrical
and Electronic Engineering,
The University of Melbourne,
VIC3010, Australia. ‡

M.R. James
Department of Engineering,
Faculty of Engineering
and Information Technology,
The Australian National University,
ACT 0200, Australia. §

## Abstract

In the context of long range dependent processes we compare robust and adaptive prediction/filtering. The intuitive observation that adaptation achieves optimality asymptotically and outperforms in the long run robust filtering is quantified through the estimation of convergence rates and levels of achievable performance.

**Keywords: long range dependencies, adaptive filtering, robust filtering, Hurst parameter**

## 1 Introduction

Within the adaptive control community there has been an increasing interest, partly initiated by a seminal talk by G. Zames at the 1996 IFAC Congress, in the basic requirements on the accuracy of the predictor part of the feedback loop in an adaptive control system. Among the interesting questions is the following: when is a robust predictor better than an adaptive predictor. In general it is very difficult to give a fair comparison between these two possibilities because most signal models prejudice either one of the solutions. Recent research efforts in areas of research as diverse as telecommunications, finance, or electrical power systems have observed that the traces of measured signals show correlation over a wide range of time scales. It is known that this excludes the possibility of finding a finite dimensional Markov process as a model for such signals. Neither robust nor adaptive techniques can provide a good model tuned to this application.

The signals considered in this paper are asymptotically self-similar, but they also exhibit interesting short term dynamics obtained by passing the self-similar signal through a finite dimensional filter. These short term dynamics often represent the effects of various control laws acting at varous time scales (like flow control and policing in communication networks). Hence this problem seems a good case study for the comparison between robust and adaptive predictors, in view of understanding their quality as part of a feedback loop.

More specifically this paper compares the "performance" (measured by the probability with which large errors occur between the predicted value and the true value of the signal at the prediction horizon) of a robust filter on the one hand and of various adaptive filters on the other hand. Intuitively one feels that robust filters will give better performance initially, when the parameter estimates have not had time to converge because there are still insuf-

*Research supported in part by the Australian Research Council. Part of the results in this paper were based on research carried out by this author within the framework of the Belgian Program on Interuniversity Attraction Poles, Prime Minister's Office, Science Policy programming. The scientific responsibility rests with its authors.

†Senior Research Fellow, F.W.O.-Vlaanderen, E-mail: rene.boel@rug.ac.be.

‡E-mail: i.mareels@unimelb.edu.au

§E-mail: Matthew.James@anu.edu.au

ficient data. However after enough data have been observed to get good parameter estimates, one can expect that the adaptive predictor will give better performance. In the case where long range dependent components are present in the signal we find that the error bounds of robust predictors are large (the relative entropy rate between signal models with different Hurst parameters grows very fast). But at the same time the rate of convergence of the prediction error for an adaptive predictor is slower than for finite dimensional models. In this paper we provide some initial results in order to carry out a quantitative comparison.

Intuitively one can understand the difficulty of accurately predicting long range dependent signals as follows. In principle the long range memory allows prediction, assuming perfectly known signal models, over very long prediction horizons. However the predictors will have a much longer memory than for the usual short range dependent ARMA signals, and hence will require the summation of weighted measured values going far back into the past. The weighting coefficients will decay very slowly. One expects that in that case the results are very sensitive to parameter uncertainty. This explains why the robust filter has such a large error bound and similarly it explains the very slow convergence of adaptive predictors.

In order to analyze these phenomena quantitatively we introduce in the next section a mathematical model for signals exhibiting both short-range and long-range dynamics. In section 3 we analyse the rate of convergence of the error of an adaptive $L$-window predictor. The rate of convergence depends on the Hurst parameter of the signal, and becomes slower and slower as this Hurst parameter approaches 1. In section 4 on the other hand we consider the performance of risk-sensitive predictors for such signals. It turns out that the error bound is very sensitive to the Hurst parameter, especially for signals with a Hurst parameter close to 1. Finally in section 5 we give some preliminary comments on the implications of our results. This short paper does not contain any proofs of the results. The full details of the proofs will be published in a paper under preparation.

## 2 Signal models

Various models for long range dependent and (asymptotically) self-similar stochastic processes have been proposed. A strictly self-similar process, which looks the same at all time scales, can be characterized by one additional parameter (besides its mean and its variance). This parameter is called the Hurst parameter $H$. It measures the decay rate of the correlation coefficients. A Hurst parameter of $H = 0.5$ corresponds to a finite-state Markov process, while a Hurst parameter of $H = 1$ corresponds to correlation which decays so slowly that future values of the process can be predicted perfectly (and infinitely far in the future).

In practice signals are only asymptotically self-similar. In this paper we obtain models for such long range dependent signals by passing a self similar process through a finite dimensional filter. This allows representation of signals where one believes on physical grounds that there is a long range dependent component, but where there is also significant short range dynamics influencing the signal. If the signal under study is the arrival rate at an access point in a broadband communication network (averaged over the smallest time interval which is of interest in the applications) then one expects this signal to exhibit long range dependence due to the heavy tailed distribution of arriving packets. However congestion and flow control protocols, shaping and policing, storage in finite buffers, all will modify this arrival process both at the edge and in the core of the network. Each of these control actions will modify the signal at a particular time scale. These control actions will behave approximately like finite dimensional filters, which introduce important short range dynamics in the signal. Similar arguments can be given justifying the use of this model for inflow of rain in a reservoir, for demand in a power system, for exchange rates between various currencies.

The goal of the predictors in this paper is to obtain an estimate $\hat{y}_{t+n}$ which depends only on the past observations $y_0, y_1, \ldots, y_t$ and which guarantees that the error $\hat{y}_{t+n} - y_{t+n}$ is small in some probabilitistic sense. To simplify the notation we only consider the case $n = 1$ in this paper. We assume that the observations available at time $t$ are generated as follows.

A self-similar signal $m_t$ is obtained by passing a Gaussian white noise signal $\epsilon_t$ through an infinite dimensional filter with $z$-transform

$$(1 - z^{-1})^{-d}, \quad 0 \le d < 0.5.$$

Beran [1] shows that the output $m_t$ of this filter,

$$m_t = (1 - z^{-1})^{-d} \epsilon_t = \sum_{k=0}^{\infty} \frac{d!}{k!(d-k)!} (-1)^k \epsilon_{t-k}$$

(where $\beta! = \Gamma(\beta + 1)$ for any $\beta \in R$), has Hurst parameter $H = d + 0.5$. More specifically $m_t$ is a zero mean Gaussian signal with covariance

$$\gamma(k) = E[m_t m_{t-k}] = \sigma_\epsilon^2 \frac{(-1)^k \Gamma(1-2d)}{\Gamma(k-d+1)\Gamma(1-k-d)}.$$

The autocorrelation coefficients $\rho(k) = \frac{\gamma(k)}{\gamma(0)}$ can be evaluated as:

$$\rho(k) = \frac{\Gamma(1-d)\Gamma(k+d)}{\Gamma(d)\Gamma(k+1-d)}$$

which behaves asymptotically as $k \to \infty$ like

$$\sim \frac{\Gamma(1-d)}{\Gamma(d)}|k|^{2d-1} = \frac{\Gamma(1-d)}{\Gamma(d)}|k|^{2H}.$$

The process $m_t$ is called a fractional ARIMA$(0, d, 0)$-process in Beran [1].

The long range dependent signal $x_t$ is obtained by passing the self-similar process $m_t$ through a simple short range ARX filter:

$$x_t = m_t - \sum_{n=1}^{N} \theta_n m_{t-n}.$$

The process $x_t$ is called an ARIMA$(0, d, N)$-process.

The observations available for prediction are noisy measurements $y_0, y_1, \ldots, y_t$ of $x_0, x_1, \ldots, x_t$, where

$$y_t = x_t + \nu_t$$

with $\nu_t$ an i.i.d. $\mathcal{N}(0, \sigma_\nu^2)$ Gaussian sequence.

This model is graphically represented in fig. 3.1.

# 3 Adaptive Predictors

If the model generating the signal were known exactly - if the parameters $d, \sigma_\epsilon^2, \sigma_\nu^2$, and $\theta_n, n = 1, \ldots, N$ were known exactly - then it would be possible to apply the Durbin-Levinson algorithm to the transfer function

$$(1 - z^{-1})^{-d}(1 + \sum_{i=1}^{N} \theta_i z^{-i})$$

in order to derive the parameters $\beta_j^{*L}$ of the best possible $L$-window predictor

$$y_{t+1}^{*L} = \sum_{j=0}^{L} \beta_j^{*L} y_{t-j} = \beta^{*L}\phi_t^L = \beta^{*L}\phi_t,$$

minimizing the mean square value of the error $\epsilon_{t,L} = y_{t+1}^{*L} - y_{t+1}$. Here $\phi_t$ is the infinite vector of past observations $(y_t, y_{t-1}, \ldots y_{t-k} \ldots)^T$ and with a slight abuse of notation we extended the vector $\beta^{*L}$ with zeros ($\beta_j^{*L} = 0, j > L$).

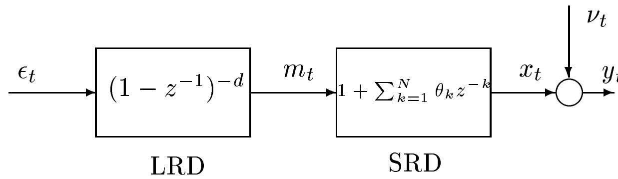


Figure 3.1: Block diagram of signal model.

In a direct adaptive predictor the optimal coefficients $\beta_j^{*L}$ are replaced by their least squares estimate $\hat{\beta}_j^L(t)$ minimizing the average prediction error

$$\begin{aligned} J(\hat{\beta}^L(t)) &= \tfrac{1}{t}\textstyle\sum_{s=0}^{t-1} \hat{\epsilon}_{s+1}^2 \\ &= \tfrac{1}{t}\textstyle\sum_{s\le t}((\phi_s^L)^T(\hat{\beta}^L(t) - \beta^{*L}) - \epsilon_{s+1,L}^*)^2 \end{aligned} \qquad (3.1)$$

This optimal parameter vector $\hat{\beta}^L(t)$ is usually obtained recursively, using:

$$\hat{\beta}^L(t+1) = \hat{\beta}^L(t) + P_{t+1}\phi_t^L(y_{t+1} - (\phi_t^L)^T\hat{\beta}^L(t))$$

$$P_{t+1} = P_t - \frac{P_t\phi_t(\phi^L)_t^T P_t}{1 + (\phi^L)_t^T P_t \phi_t^L}$$

The properties established below are proven directly for the off-line estimator, as obtained by minimizing the observed squared error as defined in (3.1). The recursive, on-line parameter estimates will then have the same asymptotic properties because the arguments in classical adaptive prediction [Goodwin and Sin] which show that the on-line estimator enjoys the same asymptotic properites as the off-line estimator remain valid for the model under study. The reason why methods directly analysing recursive estimators are not applicable here is that these methods require analysis of moments of the form $E[y_t^2 y_{t-j}^2]$, and these moments turn out to be $\infty$ due to the long range dependence of the signal.

The quadratic forms involved in the mean square cost $J(\hat{\beta}^L(t))$ can be brought into a form such that generalisations to the central limit theorem [5], [9] are applicable. This leads to the conclusion that the error between the adaptive predictor parameter vector $\hat{\beta}^L(t)$ and the optimal predictor parameter vector $\beta^{*L}$ converges in distribution to zero like $\frac{1}{t^{\max(\frac{1}{2}, 1-2d)}}.Z$ The random variable $Z$ has zero mean, and for $0 < d < \frac{1}{4}$ it has a Gaussian distribution, but for $\frac{1}{4} \le d < \frac{1}{2}$ it has another zero mean, finite variance distribution with heavier tails than the Gaussian. Moreover the rate of convergence $t^{-\max(\frac{1}{2}, 1-2d)}$ is much slower than in the case of short range dependent signals (where convergence is like $\frac{1}{t}$ provided there is enough excitation). Note the slowing down of the rate of convergence as soon as the Hurst parameter reaches 0.75, and the fact that the tail of the error distribution is then heavier than for a Gaussian.

These results are not sufficient to calculate accurate error bounds for

$$\begin{aligned} E[\ \hat{y}^L\ (t+1) - y_{t+1}]^2 \\ = \ E[\hat{y}^L(t+1) - y^{*L}(t+1)] \\ +E[y^{*L}(t+1) - y_{t+1}]^2 \end{aligned}$$

One can however get insight in the asymptotic behaviour of the adaptive prediction error by considering the average error which one would obtain if one kept the parameter estimate $\hat{\beta}^L(t)$ fixed during the time interval $t+1, \ldots, 2t-1$, $2t$. Given that the parameter estimates converge in distribution to the optimal parameter vector the following expression provides an upper bound for the average error over the time interval $t$ up to $2t$ if one kept updating the predictor parameter estimate:

$$\begin{aligned} &\frac{1}{t} \quad \Sigma_{s=t+1}^{2t} \|\hat{y}^L_{\hat{\beta}^L(t)}(s+1) \quad -y^{*L}(s+1)\|^2 \\ &\to \left(\tfrac{1}{t}\Sigma_{s=0}^{t-1}\phi_s^L(y_{s+1} - y^{*L}_{s+1})\right) \left(\frac{1}{t}\Sigma_{s=0}^{t-1}\phi_s^L(\phi_s^L)^T\right)^{-1} \\ &\qquad \left(\frac{1}{t}\Sigma_{s=0}^{t-1}\phi_s^L(y_{s+1} - y^{*L}_{s+1})\right) \end{aligned}$$

Again the limit theorems in [5] and [9] are applicable to the individual terms in the last expression. Its limit behaves asymptotically like

$$\frac{1}{t^{\max(1,2-4d)}}(Z_0\, Z_1 \ldots Z_L)\Sigma_L^{-1}(Z_0\, Z_1 \ldots Z_L)^T$$

where the vector $(Z_0\, Z_1 \ldots Z_L)$ contains zero mean, finite variance random variables which are correlated. Because of this correlation it is not clear how one could get insight in the behaviour of this error term as a function of $L$. Finding the optimal window size $L$ (as a function of $t$, the observation window) for the adaptive predictors will therefore require simulation experiments. Intuitively we expect that $L(t) \sim t^{\max(\frac{1}{2}, 1-2d)}$ which means that the allowable growth of the prediction window will have to be very slow.

Similar conclusions can be obtained for indirect adaptive predictors, where estimates $\hat{d}, \hat{\sigma_\nu^2}, \hat{\sigma_\epsilon^2}$ and $\hat{\theta_k}$ of the parameters $d, \sigma_\nu^2, \sigma_\epsilon^2$ and $\theta_k$ are introduced in the explicit expression for the correlations $\gamma_k, k \leq K+3$. The estimates $\hat{d}, \hat{\sigma_\nu^2}, \hat{\sigma_\epsilon^2}$ and $\hat{\theta_k}$ can be obtained efficiently e.g. solving the equations $\gamma_k(d, \sigma_\nu^2, \sigma_\epsilon^2, \theta_k) = \hat{\gamma}_k$ for $k = 0, 1, \ldots, K+2$ where $\hat{\gamma}_k$ is the empirical correlation function based on the past measurements $y_s, s \leq t$. Consequently it is possible to write down an approximate version of the Durbin-Levinson equations, and obtain estimates of the optimal coefficients $\beta_k^{*,L}$ of an $L$-window predictor.

# 4 Robust Predictors

The uncertainty about the parameters $d, \sigma_\nu^2, \sigma_\epsilon^2$ and $\theta_k$ specifying the signal model can also be taken into account by assuming that these parameters lie in prespecified intervals, and by finding a predictor which in some sense makes the quadratic prediction error small for all the parameter values within these intervals. This is the robust prediction approach. Specifically we study *risk sensitive* filters, considered in [2], for which explicit error bounds are available. The idea is to define the filter by minimizing an average-of-exponential error criterion (in place of the usual average error criterion), and exploit a well-known duality relationship to obtain error bounds in terms of the relative entropy between the true and design distributions.

As in the preceeding section, consider the prediction of $y_{t+1}$ using an $L$-window linear predictor of the form

$$\hat{y}^{L,T}_{t+1} = \sum_{j=0}^{L} \beta_j^{L,T} y_{t-j} \qquad (4.1)$$

using an error criterion that includes $T+1$ cumulative error terms:

$$J_{L,T}(\beta^{L,T}) = E_{\alpha_d} Z \qquad (4.2)$$

with

$$Z = \left[\exp\left(\mu_1 \sum_{j=0}^{T-1} |\hat{y}^{L,T}_{t-j} - y_{t-j}|^2 + \mu_2 |\hat{y}^{L,T}_{t+1} - y_{t+1}|^2\right)\right].$$

The expectation $E_{\alpha_d}$ is taken with respect to the design probability distribution $\mathcal{P}_{\alpha_d}$ for the model of section 2 describing the nominal values of the parameters used for predictor design. Let $\hat{\beta}^{L,T}$ denote a minimizer of the criterion (4.2); this defines the *risk-sensitive filter*. In other words the coefficients $\hat{\beta}^{L,T}$ are such that the expression (4.2) is minimized if $\hat{y}_s^{L,T}$ is replaced by its expression (4.1), and if the expectation is calculated explicitly under the assumption that $\mathcal{P}_{\alpha_d}$ correctly describes the signal model.

In reality the observations are generated by another signal model. Assume that the probability distrbution $\mathcal{P}_{\alpha_t}$ describes the "true" probability distribution from which the observation samples $y_s$ are drawn, which are input into the predictor. This model also specifies the distribution of the value $y_{t+1}$ to be predicted. We assume that $\mathcal{P}_{\alpha_t}$ is absolutely continuous with respect to $\mathcal{P}_{\alpha_d}$. The term "robust predictor" is justified by the error bound (based on the duality between free energy and relative entropy):

$$\begin{aligned} E_{\alpha_t} \ \left[\mu_1 \textstyle\sum_{j=0}^{T-1} |\hat{y}^{L,T}_{t-j} - y_{t-j}|^2 + \mu_2 |\hat{y}^{L,T}_{t+1} - y_{t+1}|^2\right] \\ \leq \log J_{L,T}(\hat{\beta}^{L,T}) + R^M(\mathcal{P}^M_{\alpha_t}, \mathcal{P}^M_{\alpha_d}). \quad (4.3) \end{aligned}$$

where

$$R(Q|P) = \begin{cases} \int \log \frac{dQ}{dP} dQ & \text{if } Q << P \\ +\infty & \text{otherwise} \end{cases}$$

is the relative entropy of the probability measure $Q$ with respect to the probability measure $P$, provided $\log \frac{dQ}{dP}$ is well defined and sufficiently integrable.

The bound (4.3) limits the degradation in performance of the filter when the true distribution $\mathcal{P}_{\alpha_t}$ is not known. The error bound consists of two parts. The first term $\log J_{L,T}(\hat{\beta}^{L,T})$ bounds the prediction error in the case $\mathcal{P}_{\alpha_t} = \mathcal{P}_{\alpha_d}$, that is when the true model coincides with the model used in the design of the predictor. The second term $R^M(\mathcal{P}^M_{\alpha_t}, \mathcal{P}^M_{\alpha_d})$ describes how fast the error bound grows when the true model is different from the design model. The relative entropy can be thought of as a distance between measures. This distance $R^M(\mathcal{P}^M_{\alpha_t}, \mathcal{P}^M_{\alpha_d})$ vanishes when the probability measures are equal.

**Proposition 4.1** *Assume $T = M = L + 1$ and assume that the parameters $\mu_1$ and $\mu_2$ are sufficiently small so that the right hand side in (4.3) is finite. Then a stationary robust filter is well-defined and the following functions exist:*

$$\limsup_{T\to\infty} \tfrac{1}{T} \log J_{L,T}(\hat{\beta}^{L,T}) = c(\mu_1, \mu_2) < +\infty$$
$$\limsup_{T\to\infty} \tfrac{1}{T} R^T(\mathcal{P}^T_{\alpha_t}, \mathcal{P}^T_{\alpha_t}) = r(\alpha_t, \alpha_d) < +\infty$$

*Then the stationary robust filter enjoys the bound*

$$\limsup_{T\to\infty} \frac{1}{T} \mathbf{E}_{\alpha_t} Z \le c(\mu_1, \mu_2) + r(\alpha_t, \alpha_d). \quad (4.4)$$

*where*

$$Z = \left[ \mu_1 \sum_{j=0}^{T-1} |\hat{y}^{L,T}_{t-j} - y_{t-j}|^2 + \mu_2 |\hat{y}^{L,T}_{t+1} - y_{t+1}|^2 \right]$$

In order to make the above results practically useful it is necessary to obtain the predictor in recursive form. The equivalent recursive filter is obtained by using an infinite dimensional state space representation for the signal model, and by introducing this into (4.2). The details will be given in [7].

In order to understand the implications of the bound (4.5) it is necessary to be able to explicitly calculate bounds for the error terms $c(\mu_1, \mu_2)$ and $r(\alpha_t, \alpha_d)$. This leads to quite tedious calculations. The expression for the relative entropy is quite complicated even for the case under consideration here where all the signals are Gaussian. Initial experiments using approximations and rough bounds do indicate in any case that the relative entropy is very sensitive as a function of $d$, especially for $d$ close to 0.5. A small change in $d$ causes a large relative entropy. This is intuitively explained by the fact that th long memory in the signal model causes the optimal coefficients $\beta^{L,T}$ to decay very slowly. The error bound diverges as the Hurst parameter $H$ approaches 1. This confirms what one expects intuitively: underestimating the long range dependence can seriously degrade the performance of the predictor.

# 5 Some preliminary comments on the quality of adaptive and robust predictors

We now make some comments on the expected behavior of our adaptive and robust filters based on our quantitative error estimates. Figure 5.2 shows typical plots of MSE as a function of the design parameter $\alpha_d$ for a standard MMSE filter and for a robust filter, cf. [2]. It can be seen that when $\alpha_t \neq \alpha_d$, and when the difference is significant then the robust filter error (at point $C$) is less than that of the MMSE filter (at point $A$); indeed, as the difference between $\alpha_d$ and $\alpha_t$ increases, we see that the performance of the robust filter degrades less rapidly relative to the MMSE filter. This illustrates the benefit of the robust filter. Of course, if $\alpha_d$ equals or is close to $\alpha_t$ the robust filter pays a price in loss of optimality since the robust error (at point $D$ if $\alpha_t = \alpha_d$) is greater than the MMSE filter (the error at $A$ is the smallest possible among all predictors, and assuming perfect knowledge about the model). The behavior of the adaptive filter is shown by the arrowed line going from $A$ to $B$. Initially the parameter estimates $\hat{d}, \hat{\sigma^2_\nu}, \hat{\sigma^2_\epsilon}$ and $\hat{\theta}_k$ differ strongly from the true values, and the adaptive predictor uses a model $\alpha_d(\hat{d}, \hat{\sigma^2_\nu}, \hat{\sigma^2_\epsilon}, \hat{\theta}_k)$ which is very different $\alpha_t$. Since the MMSE predictor is not robust the error may be substantial at $A$. As time progresses and more is learned about the signal model the error decreases, moving ultimately to $B$ as indicated.

Figure 5.1 illustrates filter errors as time progresses. The three horizontal lines correspond to non-adaptive filters: the MMSE filter when $\alpha_d \neq \alpha_t$ with error level at $A$, the robust filter when $\alpha_d \neq \alpha_t$ with error level at $C$, and the MMSE filter when $\alpha_d = \alpha_t$ with error level at $B$. The error for the adaptive MMSE starts at $A$, and decreases to level $B$ as time progresses as mentioned. It would be desirable to construct an *adaptively robust* filter based on the robust filter. This can be obtained by using a design model $\mathcal{P}_{\alpha_d(\hat{d}, \hat{\sigma^2_\nu}, \hat{\sigma^2_\epsilon}, \hat{\theta}_k)}$ which depends on the present best estimate of the true model parameters. Using this expectation in the minimization of (4.2) will lead to robust optimal parameters $\beta^{L,T}$ which give error bounds which decrease to $B$. It is also possible then to decrease the $\mu$ parameters as time progresses. The corresponding error bound curve is shown beginning at $C$ and decreasing to level $B$. It is clear that such an adaptively robust predictor will give good convergence provided that one can achieve:

$$\frac{r_t(\alpha, \alpha_d(t))}{\mu(t)} \to 0 \quad \text{as} \quad t \to \infty. \tag{5.1}$$

## References


[1] J. Beran, *Statistics for Long-Memory Processes*, Chapman and Hall, 1994

[2] R. Boel, M. James and I. Petersen, *Robustness and Risk-Sensitive Filtering*, Proceedings of the 36th Conference on Decision and Control, San Diego, Dec. 1997, pp.2273-2279.

[3] P. Dupuis and R. Ellis, "A Weak Convergence Approach to the Theory of Large Deviations", Wiley, New York, 1997.

[4] G. Goodwin and K. Sin, *Adaptive filtering prediction and control*, Prentice Hall, 1984.

[5] R. Fox and M. Taqqu, *Central limit theorems for quadratic forms for non-linear functions of Gaussian fields*, Prob. Theory Rel. Fields, vol. 74, pp. 213-240, 1987

[6] D.H. Jacobson, *Optimal Stochastic Linear Systems with Exponential Performance Criteria and Their Relation to Deterministic Differential Games*, IEEE Transactions on Automatic Control, **18**(2), 124-131, 1973.

[7] R. Boel, I. Mareels and M. James, *Performance evaluation of Adaptive and Robust Predictors for Signals with Long Range Dependent Components*, in preparation.

[8] J.B. Moore, R.J Elliott and S. Dey, *Risk Sensitive Generalizations of Minimum Variance Estimation and Control*, J. Math. Systems, Estimation and Control, 1996.

[9] N. Terrin and M. Taqqu, *A Noncentral Limit Theorem for Quadratic Forms of Gaussian Stationary Sequences*, Prob. Theory Related Fields, 1990, pp. 449-475


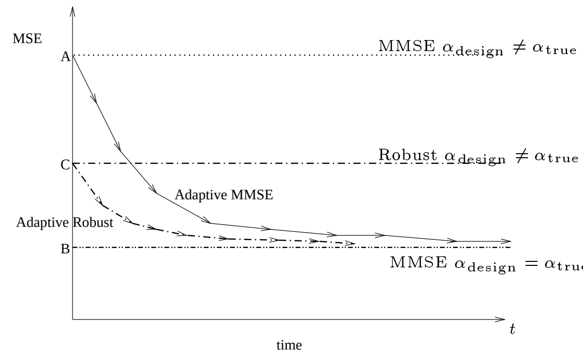


Figure 5.1: Typical error curves for non-adaptive and adaptive filters.

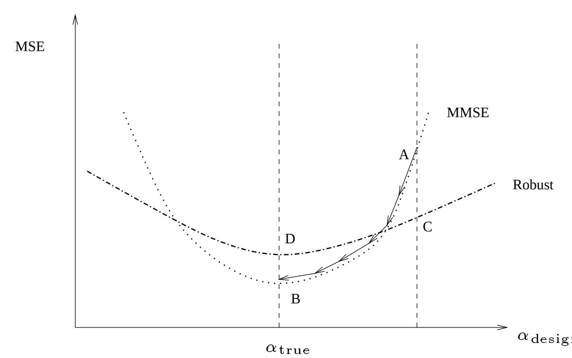


Figure 5.2: Typical error curves MSE vs. $\alpha_{\text{design}}$ for MMSE and robust filters.