A Stochastic Approach to Linear Estimation in H_∞.

Brett Ninness*

Abstract

This paper examines the problem of system identification from frequency response data. Recent approaches to this problem, known collectively as 'Estimation in H_∞', involve deterministic descriptions of noise corruptions to the data. In order to provide 'worst–case' convergence with respect to these deterministic noise descriptions, non–linear in the data algorithms are required. In contrast, this paper examines 'worst–case' estimation in H_∞ when the disturbances are subject to mild stochastic assumptions and linear in the data algorithms are employed. Issues of convergence, error bounds, and model order selection are considered.

Technical Report EE9552, Department of Electrical and Computer Engineering,
University of Newcastle,AUSTRALIA

1 Introduction

The bulk of system identification theory addresses the problem of estimating system models on the basis of observed time domain data [18, 32]. However, in many cases the available data involves measurements of a systems frequency response [31, 19]. Indeed, using frequency domain measurements together with filters to remove harmonics is an effective way to deal with static non–linearities which would otherwise obscure the estimation of an underlying linear system.

Estimation from frequency domain data has also been intensively studied as a means for providing a combination of an estimated model together with error bounds which are suitable for subsequent H_∞ robust control design. This latter line of work has become known as 'Estimation in H_∞'. It was pioneered by Parker and Bitmead [27] and later taken up by Helmicki, Jacobsen and Nett [11, 12, 13, 14, 16], Gu and Khargonekar [8, 9], Partington [29, 20, 28] and others. Two main themes are important in this latter literature:

1. Stochastic disturbance descriptions are considered too restrictive or require too much prior information to be specified and hence should be replaced by descriptions that involve assumptions only on the amplitude of any disturbances.

*This work was supported by the Australian Research Council and the Centre for Industrial Control Science. The author can be contacted at email:brett@ee.newcastle.edu.au or FAX: +61 49 21 69 93

2. Under these 'bounded error' descriptions simple linear algorithms are 'worst case' divergent. This motivates the design of various non–linear algorithms that avoid the divergence.

Motivated by these themes, the purpose of this paper is to examine the consequences and complexity of stochastic noise assumptions in an 'Estimation in H_∞' setting. Contrary to point 1 above, it is found that under very mild stochastic assumptions (which do not involve knowing any probability density functions) linear algorithms do not suffer the divergence problems of point 2, except for rare events of probability zero (they converge with probability one). As well (as will be detailed in §3) the stochastic assumptions employed do not involve significantly more prior information than deterministic assumptions involve. If one accepts that [1] a stochastic description can to some extent also characterise disturbance sequences satisfying deterministic 'bounded–magnitude' assumptions (see for example [25, 34]), then this analysis sheds light on the nature of the 'worst–case' disturbance required to provoke divergence in a linear algorithm. Namely, it exists but belongs to a set of measure zero.

To be more explicit, this paper shows that under mild stochastic assumptions and under certain conditions on how quickly the model order grows relative to the data length, then linear algorithms are robustly consistent[2] with probability one. The paper also derives probabilistic error bounds relating the true and estimated frequency responses. Because these error bounds account for undermodelling, they depend upon the model order. This leads to a new criterion for optimal model order selection by a strategy of maximising the probability that the worst–case error bound applies. As well, because of the probabilistic nature of the error bound, it is possible to gauge the conservativeness in the bound. If the purpose of the estimated model is for subsequent robust controller design, this allows for a tradeoff between safety and performance to be evaluated. Together, these results provide a stochastic approach to providing models suitable for robust control system design from observed frequency response data corrupted by random disturbances.

It should be emphasised at the outset that this paper is not adversarial to 'estimation in H_∞' ideas. Rather it is meant as a contribution to the better understanding of the performance of certain least–squares frequency domain estimation algorithms in a worst–case sense. In particular, the point of the paper is to show that (via a stochastic analysis) linear algorithms (as opposed to non–linear two–step ones) enjoy a worst–case performance better than might be expected from a purely deterministic analysis. This suggests that given their simplicity, linear algorithms may be a worthy candidate if one requires models for robust control design. This idea appears to have originally been put forward in [9] on the basis of empirical evidence; this paper can be considered as a theoretical contribution to the substantiation of this viewpoint.

[1] since any bounded magnitude sequence can be a realisation of a sequence of random variables with distribution functions of compact support.

[2] What is meant by 'robustly strongly consistent' will be defined later. It is a term borrowed from the 'estimation in H_∞' literature.

2 Identification from Frequency Response Data

The line of research known as 'Estimation in H_∞' aims at using frequency response data to find both a model in H_∞ and a hard bound on the error in this model. With rare exceptions [30, 21, 6, 10, 4, 3], the model structures used in this work are finite impulse response (FIR). As mentioned in the introduction, this line of research was begun by Parker and Bitmead [27, 26] in 1987. The multitude of papers [11, 12, 13, 14, 16, 15, 17, 29, 20, 28, 8, 9] following have concentrated mainly on improving the asymptotic in model order properties of Parker and Bitmead's ideas through the use of windowing functions and two–step algorithms, although there are other approaches [10, 3]. The problem attacked by Parker and Bitmead has been stated in canonical form by Helmicki, Jacobson and Nett in [11]:

- Assume the true system transfer function evaluated at z^{-1}, $G_T(z)$,is analytic on the domain $\mathbf{D}_\rho = \{z \in \mathbf{C} : |z| < \rho\}$, $\rho > 1$ and is bounded in magnitude by M on $\mathbf{D}_\rho$. That is, assume $G_T \in H_\infty(\mathbf{D}_\rho, M) \triangleq \mathcal{S}$.

- Assume the availability of n measurements of the frequency response of G_T at the n roots of unity. These measurements $\{f_0, \cdots, f_{n-1}\}$ are corrupted as follows

$$f_k = G_T(e^{-j\omega_s k}) + \nu_k; \qquad \omega_s \triangleq \frac{2\pi}{n}, \tag{1}$$

  where

$$|\nu_k| \leq \varepsilon. \tag{2}$$

  That is, the corruption is by the components of some element $\nu \in \ell_\infty$.

The problem then is to derive an estimate $\widehat{G}(z)$ such that $\widehat{G} \in H_\infty(\mathbf{D}_1)$ and

$$\lim_{\substack{\varepsilon \to 0 \\ n \to \infty}} \sup_{\substack{G_T \in \mathcal{S} \\ \|\nu\|_\infty \leq \varepsilon}} \left\| G_T - \widehat{G} \right\|_\infty = 0 \tag{3}$$

in which case the algorithm providing $\widehat{G}$ is termed 'robustly consistent' [16, 8]. Note that compared to stochastic estimation theory [18, 32], this definition of convergence is somewhat unusual since it involves an assumption of the magnitude of the noise disturbance going to zero. The purpose of §3 following will be to illustrate that in a stochastic setting, convergence can still be achieved without this assumption of the noise magnitude ε tending to zero.

The original estimation approach of Parker and Bitmead is as follows. Select the model to be of the form

$$\widehat{G}(z) = \sum_{k=0}^{d(n)-1} \widehat{g}_k z^k \tag{4}$$

and try to make its frequency response the same as the measurements:

$$\sum_{k=0}^{d(n)-1} \widehat{g}_k e^{-j\omega_s mk} = f_m; \qquad m = 0, \cdots, n-1. \tag{5}$$

Parker and Bitmead only consider the case of the model order chosen as $d(n) = n$. If the choice of the model order is $d(n) < n$, then the equations (5) are over-determined and a natural choice for $\widehat{g}_k$ would be one that minimised the total squared error in (5). This solution is most easily seen by vectorising:

$$\widehat{\theta}_n^T = \left[\widehat{g}_0, \cdots, \widehat{g}_{d(n)-1}\right], \tag{6}$$

$$\Lambda_n = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & e^{-j\omega_s} & e^{-j2\omega_s} & \cdots & e^{-j(d(n)-1)\omega_s} \\ 1 & e^{-j2\omega_s} & e^{-j4\omega_s} & \cdots & e^{-j2(d(n)-1)\omega_s} \\ \vdots & & & & \vdots \\ 1 & e^{-j(n-1)\omega_s} & e^{-j2(n-1)\omega_s} & \cdots & e^{-j(n-1)(d(n)-1)\omega_s} \end{bmatrix}, \tag{7}$$

$$F^T = [f_0, \cdots, f_{n-1}], \tag{8}$$

so that (5) becomes

$$\Lambda_n \widehat{\theta}_n = F \tag{9}$$

with solution minimising the squared error as:

$$\widehat{\theta}_n = \left(\Lambda_n^\star \Lambda_n\right)^{-1} \Lambda_n^\star F = \frac{1}{n} \Lambda_n^\star F. \tag{10}$$

In the case analysed by Parker and Bitmead ($d(n) = n$) they note that in the mathematics literature the solution is then known as Lagrange interpolation. In any case using (10) gives

$$\widehat{g}_m = \frac{1}{n} \sum_{k=0}^{n-1} f_k e^{jmk\omega_s}, \qquad m = 0, \cdots, d(n)-1. \tag{11}$$

That is, the model co–efficients are found by taking the inverse DFT of the frequency measurements.

By assuming the deterministic bounded noise corruption of (1), Parker and Bitmead then arrive at the following 'worst–case' bound on the estimation error

$$e_n(\varepsilon) \triangleq \sup_{\substack{G_T \in \mathcal{S} \\ \|\nu\|_\infty \leq \varepsilon}} \left\| G_T - \widehat{G} \right\|_\infty \leq \frac{K}{R^n} + \varepsilon \left(1 + \frac{2}{\pi} \ln n\right) \tag{12}$$

where $K \in \mathbf{R}^+$ and $R \in (1, \rho)$ are $G_T(z)$ dependent constants and (12) is only proved true by Parker in [26] when n is some power of 2; that is $\exists m \in \mathbf{Z}$ such that $n = 2^m$. The first term in the bound [3] (12) is due to the undermodelling in representing $G_T(z)$ with

[3] In subsequent work by other authors it has been established that this term is given by $\frac{M\rho}{\rho^n(\rho-1)}$.

an nth order model. The second term in the bound (12) is due to the corruption in the measurements by the noise sequence $\{\nu_k\}$.

In post Parker and Bitmead work on the subject, great import has been placed on the fact that the 'noise' term in (12) grows with model order n so that (3) is not satisfied. The aim has been to provide 'robustly convergent' algorithms for which (3) holds for any $G_T \in \mathcal{S}$. In [11] it was conjectured that no such algorithm existed which was linear in the data, and in [29] Partington proved this was so.

The first workers to derive a robustly convergent algorithm where Helmicki, Jacobson and Nett in [12, 16]. They avoided divergent behavior by using a low order (namely linear), interpolating polynomial between the frequency response measurement in order to impose a smoothness constraint on the estimate. A problem with this solution is that the spline interpolant usually will not have a frequency response that corresponds to the evaluation on the unit circle of some $\widehat{G} \in H_\infty$. In this case, the FIR model obtained is non–causal. This was overcome in [12, 16] by obtaining the final estimate $\widehat{G}(z)$ as the element in H_∞ closest in L_∞ norm to the L_∞ system

$$\widetilde{G_N}(z) = \sum_{k=-N}^{N} \tau_k \widehat{g}_k z^k \tag{13}$$

with $\widehat{g}_k$ given by (11), the model order $d(n)$ fixed at $d(n) = n$, and $\{\tau_k\}$ a user–chosen windowing sequence. It is at this H_∞ approximation stage that the algorithm becomes non–linear in the observation data. The resultant estimation scheme that Helmicki, Jacobson and Nett invented can be represented in the following two–stage format

$$\underbrace{\text{Data} \longrightarrow L_\infty}_{\substack{\text{Reconstruction of} \\ \text{function as Fourier Series}}} \qquad \underbrace{\longrightarrow H_\infty.}_{\substack{\text{Nonlinear Step} \\ \text{Nehari extension}}} \tag{14}$$

Until recently [10, 3], the majority of the 'robustly convergent' schemes that have been developed can be represented in this two stage way and differ only in the choice of the windowing sequence $\{\tau_k\}$ used in (13) which is the first step of (14) .

Given the added complexity required to arrive at robustly convergent estimation schemes, a question which must naturally arise is the necessity of the added computational burden and analytical difficulties. For example, even though linear schemes are worst case divergent, how likely is this worst case? Are there many disturbance sequences $\{\nu_k\}$ that can lead to divergence, or are such sequences very rare?

One way to approach this problem is to break the set of possible disturbances into subsets, one of which is the subset that leads to the worst–case error being achieved. The significance of the worst case disturbance can then be judged by measuring the size of this latter subset relative to the size of the whole set of disturbances. In the next section it is shown that for the simple linear scheme (4),(11) then for a very large set of possible disturbances, which is normalised to have size one, the subset of disturbances inducing worst–case divergence has size zero.

This method of analysis is, of course, that of employing the tools of probability theory to embed the disturbance descriptions within a stochastic framework. The point of the previous paragraph was to emphasise that such a framework can be given a very deterministic interpretation if the subjective idea of randomness is considered inappropriate.

3 Stochastic Analysis

The key issue in the deterministic 'worst–case' convergence analysis of H_∞ schemes in the literature [16, 8, 29] is the bounded noise assumption (2). This deterministic description is meant to be such as to pre–suppose as little as possible about the disturbance process.

The rationale for this is commonly argued [8, 11, 12, 14] along lines that stochastic descriptions are unwelcome because they require too much prior structure to be added to the problem assumptions; for example, knowledge of a probability density function. The purpose of this section is to examine this issue. What really is the amount of extra structure required for a stochastic description to be a realistic and tractable model for the disturbance, and what price is paid by not imposing it when it is valid? In fact, the argument here is that the extra structural impositions are minor, and if they are satisfied, then the worst case divergence occurs only for very rare disturbance sequences.

In support of this argument, convergence analysis is performed under a stochastic model that supposes three features of the disturbance process $\{\nu_k\}$; zero mean, stationarity [4], and boundedness of spectral density (or bounded second moment).

These stochastic assumptions are meant to be very unrestrictive. No knowledge of a probability density function is required. It is believed that almost any reasonable disturbance sequence $\{\nu_k\}$ could be well modelled as a realisation from such a process. For example, assuming a bounded spectral density is, via the Paley–Wiener theorem, really assuming that $\{\nu_k\}$ is not a rank deficient process. That is, it is an assumption that there is no purely deterministic law that could describe a component of the measurement errors. If such structure did exist in the disturbance, it would make more sense to estimate it than include it as an unknown. The assumption of zero mean–ness is also mild since again, if this were not the case, it would make more sense add an extra scalar variable to the problem so as to estimate the mean. Finally, the bounded noise assumption (2) already implies bounded second moment, since if $\mathbf{P}$ denotes the measure on the probability space Ω that ν_k is defined on, and if $\mathbf{E}\{\cdot\}$ denotes expectation over Ω with respect to $\mathbf{P}$ then

$$\sigma_\nu^2 \triangleq \mathbf{E}\left\{|\nu_k|^2\right\} = \int_\Omega |\nu_k|^2 \, \mathrm{d}\mathbf{P} \leq \varepsilon^2 \int_\Omega \mathrm{d}\mathbf{P} = \varepsilon^2.$$

Of course, in passing, if one does know the probability density associated with ν_k, then this bound can be refined. For example, if $\mathbf{P}$ induces a uniform density on $\mathbf{R}$ then obviously $\sigma_\nu^2 = \varepsilon^2/3$.

Given these justifications for the stochastic assumptions, the following convergence result is available.

[4] This assumption can be dropped later.

Theorem 3.1. *Suppose $\{\nu_k\}$ is a realisation of a zero mean stationary stochastic process with bounded spectral density. Then the linear algorithm given by (4) and (11) is robustly strongly consistent:*

$$\sup_{G_T \in \mathcal{S}} \left\| G_T - \widehat{G} \right\|_\infty \xrightarrow{a.s.} 0 \quad \text{as } n \to \infty \tag{15}$$

provided that

$$\lim_{n\to\infty} \frac{d^2(n)}{n} = 0. \tag{16}$$

Proof. See appendix A. □

This theorem may be compared with theorem 6.1 of [28] where the special case of Gaussian disturbances, independence for $\{\nu_k\}$ and independence of the real and imaginary parts of $\{\nu_k\}$ are assumed. A similar result is obtained for the first L_∞ approximation stage of 2 step robustly convergent algorithms.

Aside from this discussion, the key point of theorem 3.1 is that it establishes that even when the disturbance sequence $\{\nu_k\}$ does not go to zero, the worst case estimation error still goes to zero for a very large class of disturbance sequences. In fact, this class can be enlarged.

Corollary 3.1. *Theorem 3.1 remains true with the stationarity condition dropped provided $\{\nu_k\}$ is such that the $n \times n$ covariance matrix R_n given by $[R_n]_{m,k} = \mathbf{E}\{\nu_m \overline{\nu_k}\}; m, k = 0, 1, \cdots, n-1$ can have its maximum eigenvalue upper–bounded by a finite number that is independent of n.*

Proof. This is not so much a corollary, as a result with almost identical proof. Examination of the proof of theorem 3.1 in appendix A shows that the only use of the stationarity condition is to bound the maximum eigenvalue of the covariance matrix R_n by the maximum of the spectral density in order to arrive at (B.2). If $\{\nu_k\}$ is not stationary, then it is still possible that R_n has its maximum eigenvalue bounded by some finite n–independent number μ, in which case since R_n is symmetric and positive definite, Rayleigh's principle leaves (B.2) invariant from the stationary noise case, and the rest of the proof goes through unchanged. □

This corollary significantly expands the set of possible disturbance sequences for which robust strong consistency holds. Nevertheless, the stationary case has been formulated in theorem 3.1 as the main result because prescribing conditions on a spectral density is more intuitive than prescribing conditions on the eigenvalues of a covariance matrix. However, this can be simplified as well. For example, if $\{\nu_k\}$ is non–stationary but satisfies

$$|\mathbf{E}\{\nu_n \overline{\nu}_m\}| \leq \frac{\kappa}{1 + |n-m|^{1+\delta}} \quad ; \delta > 0, \kappa < \infty$$

then Lemma A.4 gives that some finite bound on the eigenvalues exists and hence robust strong consistency occurs. As a side comment, it may arise that restrictions on the growth of model order are deemed undesirable. If this is the case, then (provided the assumption of

deterministic noise bound $\varepsilon \to 0$ is considered appropriate) perhaps the use of a non–linear scheme [8, 9, 16] is more suitable.

As already mentioned, in [29] Partington has proven that there exist functions in $C(\mathbf{T})$, such that linear algorithms based on data corrupted by samples from these functions are not robustly convergent. Theorem 3.1 and Corollary 3.1 do not contradict this result. They merely highlight that, with appropriate choice of model order $d(n)$, then there are not many functions in $C(\mathbf{T})$ to pick from in order to exhibit divergence. For example, it could be conjectured that samples of almost any element of $C(\mathbf{T})$ could be well modelled as being a particular realisation of a stationary stochastic process of bounded second moment.

Given this convergence result, further questions pose themselves. What is the rate of convergence with increasing data? Is it possible to bound the error in the estimate to a prescribed degree of confidence, and in a way that accounts for undermodelling? How does this error depend on the chosen model order? An answer to these questions (which is valid for finite data) is provided by the following theorem.

Theorem 3.2. *Suppose $\{\nu_k\}$ is a realisation of a zero mean stationary stochastic process with associated spectral density $f(\lambda)$ bounded as $f(\lambda) < \mu < \infty$. Then for the linear algorithm given by (4) and (11)*

$$\mathbf{P}\left\{\left|\widehat{G}(e^{-j\omega}) - G_T(e^{-j\omega})\right| \geq \kappa\right\} \leq \frac{1}{\kappa^2}\left\{\left[\frac{M\rho}{\rho^{d(n)}(\rho-1)}\left(1+\frac{\rho^{d(n)}-1}{\rho^n-1}\right)\right]^2 + \mu\frac{d(n)}{n}\right\}. \quad (17)$$

Proof. See Appendix B. □

Note that this theorem also allows us to conclude that the convergence in (15) occurs in probability provided that

$$\lim_{n\to\infty}\frac{d(n)}{n} = 0. \quad (18)$$

This condition is weaker than (16), but concomitantly it implies a weaker mode of convergence. Note as well, that as per the relationship between theorem 3.1 and Corollary 3.1, the stationarity condition can also be dropped in theorem 3.2 if μ is replaced by an n independent upper bound on the covariance matrix of the $\{\nu_k\}$ process.

Aside from the added convergence in probability result, the main use of theorem 3.2 is in providing error bounds on estimated models. An acceptable error bound κ can be specified, and then depending depending on the number n of measurements, the model order $d(n)$, the noise size μ and the true system class (M, ρ), a probability of the estimated model having error less than this bound can be calculated. If this probability does not provide sufficient confidence in the model, then theorem 3.2 can be used to calculate the number of extra frequency measurements that will have to be made to remedy this.

The expression in theorem 3.2 therefore makes it possible to gauge the conservativeness of any hard bound on estimation error that is applied. If the purpose of the estimated model is for subsequent robust controller design, this allows for a tradeoff between safety and performance to be evaluated. For example, note that the bound in (17) may be greater

than 1. This indicates great conservativeness in the error bound κ, and suggests that a smaller error bound may be sensibly taken without unduly courting disaster in any robust control designs based on the identified model.

It could be argued that the quantification in theorem 3.2 is nebulous since it involves knowledge of μ, a bound on a spectral density, which would be difficult to define in practice. However, in line with the worst case thinking of 'Estimation in H_∞' literature, this can be circumvented by assuming that $\{\nu_k\}$ is in fact white noise, so that μ becomes the variance σ_ν^2, and it is reasonable to expect that knowledge of this quantity could be provided since it is easily related to the average magnitude of the error.

The reason why this approach is in line with 'worst–case' thinking is that disturbance realisations $\{\nu_k\}$ of non–smooth (uncorrelated) character are the ones producing the worst–case error. More explicitly, the reason that noise term in (12) grows with n in the worst case is that with no smoothness assumption on $\{\nu_k\}$; the measured responses can differ in magnitude by ε no matter how closely together the measuring frequencies are spaced. Divergence then occurs since any smooth function interpolating these arbitrarily non–smooth measurements must have 'overshoot' between the interpolating points, and this can grow to be arbitrarily large.

3.1 Model Order Selection

The final application of theorem 3.2 is to the model order selection problem. Although a fundamental issue in stochastic estimation theory [18], it does not seem to have been addressed by current thinking in the 'Identification in H_∞' literature.

The problem to be addressed is this. For given noise corruption and data length n, what is the best model order $d(n)$ to choose? Stochastic estimation theory addresses this issue by recognising that there is a bias versus variance tradeoff in the choice of model order; as the model order grows the bias error due to undermodelling drops but the variability in the estimates due to noise increases. The analogue in H_∞ estimation would be to choose the model order giving the smallest worst–case bound. In our stochastic setting, it is possible to choose a model order that minimises the probability of the worst–case bound given in theorem 3.2 being violated.

As an example, for the particular case of $n = 100$ frequency measurements being taken, a noise corruption of magnitude $\varepsilon = 0.1$ being present, the true system G_T being in the class parameterised by $M = 1, \rho = 1.4$, and error bounds of size $\kappa = 0.1$ being required, the probability given by theorem 3.2 of satisfying this $\kappa = 0.1$ bound is plotted against model order $d(n)$ in the left hand diagram of figure 1. As can be seen, for this case the optimal model order to choose to minimise the probability of exceeding the $\kappa = 0.1$ error bound is $d = 18$, in which case the bounds are 94% confidence bounds. The resulting estimates for the cases of $d(n) = 100$ and $d(n) = 18$ are shown in the right hand plot of figure 1. The utility of using theorem 3.2 is apparent here in the marked difference in the quality of the estimates

The optimal model order $d(n)$ can also be calculated directly without resort to plotting via the following simple Lemma.

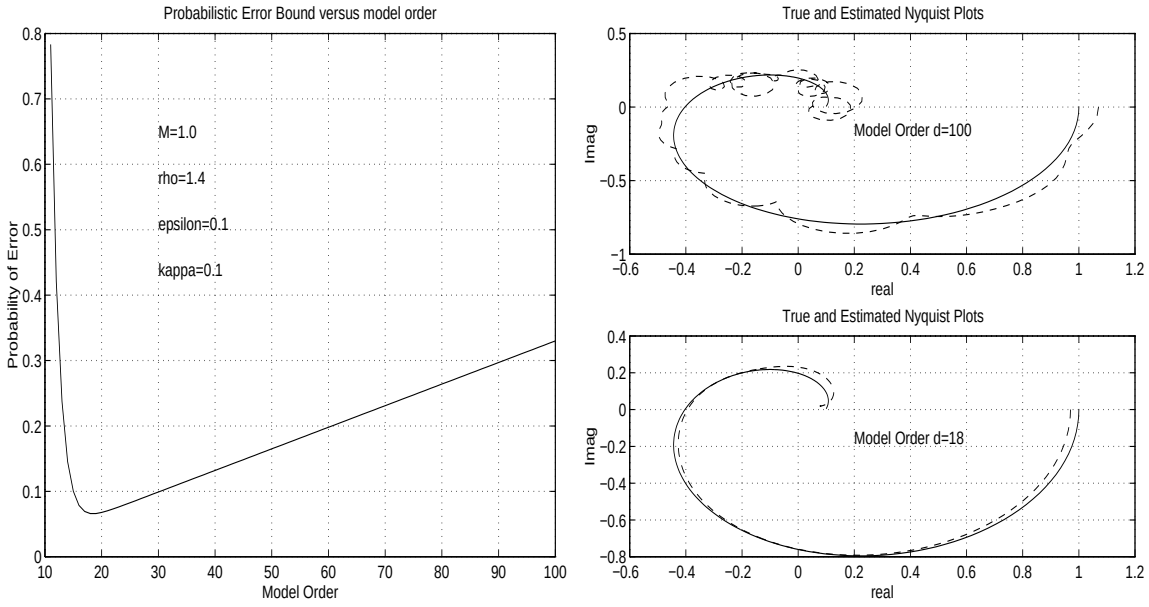

Figure 1: *Left figure shows example of how probabilistic error bound is minimised for model order $d(n) < n$. Right figures show estimates (dashed lines) and true systems (solid lines) represented as Nyquist plots for the cases of model orders $d(n) = 100$ (top) and $d(n) = 18$ (bottom) - the latter is the optimal model order as shown by the left plot.*

Lemma 3.1. *The probabilistic bound in theorem 3.2 is minimised by $d(n)$ chosen as the integer nearest*

$$d(n) = \ln\left[\frac{Kn + \sqrt{K^2n^2 - 4\mu Kn(2-\rho^n)}}{2\mu}\right]\frac{1}{\ln\rho} \tag{19}$$

where

$$K \triangleq \frac{2M^2\rho^2(\rho^n - 2)\ln\rho}{(\rho-1)^2(\rho^n-1)^2}. \tag{20}$$

Proof. See appendix C. □

Although this section has concentrated on minimising errors via the choice of model order for an FIR model structure, one could also be led to try to minimise errors by changing the model structure. At the same time, perusal of the proofs in this paper and in the 'Estimation in H_∞' literature in general shows that the orthogonality properties of the $\{e^{j\omega n}\}$ 'basis functions' used in an FIR structure are of great utility in analysis. Therefore, if the basis functions are changed to obtain a new model structure, it is desirable for this orthogonality to be maintained. In [24] this is done in such a way that prior information about likely pole locations for the true system can be encoded in the 'basis functions' chosen; the FIR structure becomes a special case when the poles are chosen at the origin. In [33, 23] the use of these new basis functions is examined in the same context as the work in this paper. A significant improvement in estimation accuracy is shown to be possible.

As a concluding comment, note that in the context of stochastic analysis of frequency domain estimation a natural mode of analysis that has not been pursued in this paper is to examine the asymptotic distributional properties of the estimates. This has been pursued in [5, 33] by employing the work of Brillinger [2].

4 Conclusion

This paper has examined the performance of the scheme of least–squares estimation of FIR model structures from frequency domain data. This technique happens to be the first step of a large class of two–stage and non–linear in the data algorithms that are known under the title of 'Estimation in H_∞'. In this latter work much concern is placed on the 'worst–case' divergence of this first least–squares step. The purpose of this paper was to investigate the nature of the disturbances that might lead to divergence. These were shown to have measure zero within a large class of possible disturbances, namely those that can be modelled as a stationary stochastic process of bounded spectrum. In fact, it was also shown to be possible to relax the stationarity assumption while still maintaining convergence.

The paper also established a new expression that allows the calculation of the probability that a given hard bound on the estimation error applies. Again, the validity of the bound depends only on very mild assumptions on the disturbance process. The utility of this probability expression is that it gives a criterion for model order selection by choosing the model order that maximises the probability of error being smaller than a given limit. Finally, in the context of using the estimated model for control design,the probability expression makes it possible to evaluate the tradeoff between conservativeness and performance in a robust design.

A Proof of Theorem 3.1

Proof. From the proof of Lemma B.1

$$\widehat{G}(e^{-j\omega}) = \sum_{m=0}^{d(n)-1} \widehat{f}_m e^{-j\omega m}$$

where

$$\widehat{f}_m = \sum_{\ell=0}^{\infty} g_{m+\ell n} + \frac{1}{n}\sum_{r=0}^{n-1} \nu_r e^{j\omega_s m r}.$$

So

$$\widehat{G}(e^{-j\omega}) = H_n(\omega) + \frac{1}{n}V_n(\omega) \tag{A.1}$$

where

$$H_n(\omega) \triangleq \sum_{m=0}^{d(n)-1} \left(g_m + \sum_{\ell=1}^{\infty} g_{m+\ell n} \right) e^{-j\omega m}, \tag{A.2}$$

$$V_n(\omega) \triangleq \sum_{r=0}^{n-1} \nu_r \gamma_r^n(\omega),$$

$$\gamma_r^n(\omega) \triangleq \sum_{m=0}^{d(n)-1} e^{-j(\omega - r\omega_s)m}; \quad \omega_s \triangleq \frac{2\pi}{n}. \tag{A.3}$$

Dealing first with the noise induced error $V_n(\omega)$, note that via Abel's partial summation formula

$$\sum_{r=0}^{n-1} \nu_r \gamma_r^n(\omega) = \sum_{r=0}^{n-2} \left(\sum_{k=0}^{r} \nu_k \right) [\gamma_r^n(\omega) - \gamma_{r+1}^n(\omega)] + \gamma_{n-1}(\omega) \sum_{r=0}^{n-1} \nu_r.$$

Therefore, setting $\gamma_n^n(\omega) = 0$ and denoting $S_r \triangleq \sum_{k=0}^{r} \nu_k$ gives

$$\max_{\omega\in[-\pi,\pi]} |V_n(\omega)| \le \max_{0\le k<n} |S_k| \max_{\omega\in[-\pi,\pi]} \left(\sum_{r=0}^{n-1} |\gamma_r^n(\omega) - \gamma_{r+1}^n(\omega)| \right),$$

so that via Lemma A.2 and by choosing an integer m such that $2^m \le n \le 2^{m+1}$

$$\max_{\omega\in[-\pi,\pi]} \frac{1}{n} |V_n(\omega)| \le \frac{4\pi d(n) \log n}{n} \max_{0\le k<n} |S_k| \le \frac{4\pi(m+1) d(2^{m+1}) \log 2}{2^m} \max_{0\le k<2^{m+1}} |S_k|,$$

so that

$$\limsup_{n\to\infty} \max_{\omega\in[-\pi,\pi]} \frac{1}{n} |V_n(\omega)| \le \limsup_{m\to\infty} \frac{4\pi(m+1) d(2^{m+1}) \log 2}{2^m} \max_{0\le k<2^{m+1}} |S_k|. \tag{A.4}$$

However, if $f(\lambda)$ is the spectral density associated with $\{\nu_k\}$, then denoting $\mu = \max_\lambda f(\lambda)$ and using the same method as employed to derive (B.2),

$$\mathbf{E} \left\{ \left| \sum_{k=m+1}^{n} \nu_k \right|^2 \right\} \le \mu(n-m)$$

so that Lemma A.1 may be applied with $u_k = 1$ and any $\alpha > 1$. Therefore, for any $\varepsilon > 0$, $\alpha > 1$, there exists $K < \infty$ such that

$$\sum_{m=0}^{\infty} \mathbf{P} \left\{ \max_{0\le k<2^{m+1}} |S_k| > \varepsilon \frac{2^m}{(4\pi \log 2)(m+1) d(2^{m+1})} \right\} \le \frac{K}{\varepsilon^2} \sum_{m=0}^{\infty} \frac{m^2 d^2(2^{m+1})}{2^{(2-\alpha)m}}.$$

Since α may be arbitrarily close to 1, this latter sum is convergent for $d(n) = o(\sqrt{n})$ so that under the conditions of the theorem, by the Borel–Cantelli Lemma and (A.4)

$$\max_{\omega\in[-\pi,\pi]} \frac{1}{n} |V_n(\omega)| \xrightarrow{a.s.} 0 \quad \text{as } n \to \infty.$$

This takes care of the convergence of the 'variance' term. For the 'bias' term note that by (A.2)

$$H_n(\omega) = G_T(e^{-j\omega}) - \sum_{m=d(n)}^{\infty} g_m e^{-j\omega m} + \sum_{m=0}^{d(n)-1} \sum_{\ell=1}^{\infty} g_{m+\ell n} e^{-j\omega m}.$$

So from (A.1)

$$\begin{aligned}
\max_{\omega\in[-\pi,\pi]} |\widehat{G}(e^{-j\omega}) - G_T(e^{-j\omega})| &\leq \left| \sum_{m=0}^{d(n)-1} \sum_{\ell=1}^{\infty} g_{m+\ell n} \right| + \left| \sum_{m=d(n)}^{\infty} g_m \right| + \max_{\omega\in[-\pi,\pi]} \frac{1}{n} |V_n(\omega)|, \\
&\leq \sum_{m=0}^{d(n)-1} \frac{1}{\rho^m} \sum_{\ell=1}^{\infty} \frac{1}{\rho^{n\ell}} + \sum_{m=d(n)}^{\infty} \frac{1}{\rho^m} + \max_{\omega\in[-\pi,\pi]} \frac{1}{n} |V_n(\omega)|, \\
&= \frac{M\rho}{\rho^{d(n)}(\rho-1)} \left[1 + \frac{\rho^{d(n)} - 1}{\rho^n - 1} \right] + \max_{\omega\in[-\pi,\pi]} \frac{1}{n} |V_n(\omega)|, \\
&\xrightarrow{a.s.} 0 \quad \text{as } n \to \infty.
\end{aligned}$$

□

Lemma A.1. *Let $\{X_k\}$ be arbitrary random variables (not necessarily independent or identically distributed). Suppose that there exists a set of non–negative numbers $\{u_k\}$ such that for some $\alpha > 1$*

$$\mathbf{E}\left\{ \left| \sum_{k=n}^{m} X_k \right|^2 \right\} \leq \left(\sum_{k=n}^{m} u_k \right)^{\alpha}.$$

Then for any $\varepsilon > 0$, there exists $K < \infty$ that is independent of n such that

$$\mathbf{P}\left\{ \max_{0\leq\ell<n} \left| \sum_{k=0}^{\ell-1} X_k \right| \geq \varepsilon \right\} \leq \frac{K}{\varepsilon^2} \left(\sum_{k=0}^{n-1} u_k \right)^{\alpha}.$$

Proof. See theorem 12.2 of [1] or (under more general conditions) theorem 3.1 of [22]. □

Lemma A.2. *Using the notation of (A.3),*

$$\max_{\omega\in[-\pi,\pi]} \sum_{r=0}^{n-1} |\gamma_r^n(\omega) - \gamma_{r+1}^n(\omega)| \leq 4\pi d(n) \log n. \tag{A.5}$$

Proof. By the mean value theorem and the definition (A.3) for $\gamma_r^n(\omega)$

$$\sum_{r=0}^{n-1} |\gamma_r^n(\omega) - \gamma_{r+1}^n(\omega)| \leq \frac{2\pi}{n} \sum_{r=0}^{n-1} \max_{\frac{2\pi r}{n} \leq \theta < \frac{2\pi(r+1)}{n}} \left| \frac{\mathrm{d}}{\mathrm{d}\theta} \sum_{m=0}^{d(n)-1} e^{-j(\omega-\theta)m} \right|. \tag{A.6}$$

Now, since $\gamma_r^n(\omega)$ has period 2π and since (A.6) involves a sum over a 2π range, then (A.6) is invariant to the particular choice of ω (the ordering of the terms in the sum are affected, but not the terms themselves). Therefore, for simplicity $\omega = 0$ may as well be chosen, in which case we note that

$$\begin{aligned}\left|\sum_{m=0}^{d(n)-1} me^{j\theta m}\right| &= \frac{1}{2\sin^2\frac{\theta}{2}}\left|d\sin\frac{\theta}{2}\cos\frac{d\theta}{2} + j(d-1)\sin\frac{\theta}{2}\sin\frac{d\theta}{2} - \cos\frac{\theta}{2}\sin\frac{d\theta}{2}\right|, \\ &\leq \frac{1}{2|\sin\frac{\theta}{2}|}\left\{d\left|\cos\frac{d\theta}{2} - j\sin\frac{d\theta}{2}\right| + \left|\frac{\sin\frac{d\theta}{2}}{\sin\frac{\theta}{2}}\right|\right\} + \left|\frac{\sin\frac{d\theta}{2}}{2\sin\frac{\theta}{2}}\right| \leq \frac{2d}{|\sin\frac{\theta}{2}|},\end{aligned}$$

so that since $|\sin\theta/2| \geq |\theta|/\pi$ for $\theta \in [-\pi, \pi]$ then for $r \neq 0$

$$\max_{\frac{2\pi r}{n} \leq \theta < \frac{2\pi(r+1)}{n}} \left|\sum_{m=0}^{d(n)-1} me^{j\theta m}\right| \leq \frac{nd(n)}{r} \leq \frac{2nd(n)}{r+1}.$$

When $r = 0$, note that since $\sum_{m=0}^{d(n)-1} m = (d-1)(d-2)/2 < dn/(r+1)$ the bound applies for all $r \leq n-1$ so that substituting it into (A.6) gives

$$\max_{\omega\in[-\pi,\pi]} \sum_{r=0}^{n-1} |\gamma_r^n(\omega) - \gamma_{r+1}^n(\omega)| \leq 4\pi d(n) \sum_{r=0}^{n-1} \frac{1}{r+1} \leq 4\pi d(n) \int_0^{n-1} \frac{\mathrm{d}x}{x+1} = 4\pi d(n)\log n.$$

□

Lemma A.3.

$$\sum_{r=0}^{n-1}\left|\sum_{m=0}^{\ell-1} e^{-j(\omega - r\omega_s)m}\right|^2 = n\ell \qquad ;\omega_s \triangleq \frac{2\pi}{n}. \tag{A.7}$$

Proof.

$$\sum_{r=0}^{n-1}\left|\sum_{m=0}^{\ell-1} e^{-j(\omega - r\omega_s)m}\right|^2 = \sum_{m=0}^{\ell-1}\sum_{k=0}^{\ell-1} e^{j\omega(k-m)} \sum_{r=0}^{n-1} e^{jr\omega_s(m-k)} = n\sum_{m=0}^{\ell-1}\sum_{k=0}^{\ell-1} e^{j\omega(k-m)}\delta(m-k) = n\ell.$$

□

Lemma A.4. *Let $\{a_{ij}\}$ be given. If $\sum_i |a_{ij}| \leq M$ for all j and $\sum_j |a_{ij}| \leq N$ for all i then*

$$|\sum_j \sum_i a_{ij} x_i \overline{x}_j| \leq \sqrt{MN} \sum_i |x_i|^2.$$

Proof. Direct consequence of Corollary 2.3.2 in [7]. □

B Proof of Theorem 3.2

Proof. By (4) and (11) and the proof of Lemma B.1

$$\widehat{G}(e^{-j\omega}) - G_T(e^{-j\omega}) = \varphi_n(\omega) + \frac{1}{n}V_n(\omega)$$

where

$$\begin{aligned}
\varphi_n(\omega) &\triangleq \sum_{k=d(n)}^{\infty} g_k e^{-j\omega k} - \sum_{k=0}^{d(n)-1}\sum_{m=0}^{\infty} g_{k+mn} e^{-j\omega k}, \\
V_n(\omega) &\triangleq \sum_{r=0}^{n-1} \nu_r \sum_{m=0}^{\ell-1} e^{-j(\omega - r\omega_s)m}.
\end{aligned}$$

Then by Chebychev's inequality

$$\begin{aligned}
\mathbf{P}\left\{\left|\widehat{G}(e^{-j\omega}) - G_T(e^{-j\omega})\right| \geq \kappa\right\} &\leq \frac{\mathbf{E}\left\{\left|\widehat{G}(e^{-j\omega}) - G_T(e^{-j\omega})\right|^2\right\}}{\kappa^2} \\
&= \frac{1}{\kappa^2}\left(|\varphi_n(\omega)|^2 + \frac{1}{n^2}\mathbf{E}\left\{|V_n(\omega)|^2\right\}\right). \qquad \text{(B.1)}
\end{aligned}$$

Now, denote by $f(\lambda)$ the spectral density associated with $\{\nu_k\}$. By assumption, this density is bounded as $f(\lambda) < \mu < \infty$. Note that even though $\{\nu_k\}$ is a complex valued sequence, its spectral density is real by virtue of the fact that the covariance function $R(\tau) = \mathbf{E}\{\nu_k \overline{\nu_{k-\tau}}\}$ is Hermitian and positive by virtue of Bochner's theorem. In this case, using Lemma A.3

$$\begin{aligned}
\mathbf{E}\left\{|V_n(\omega)|^2\right\} &= \mathbf{E}\left\{\left|\sum_{r=0}^{n-1} \nu_r \gamma_r^n(\omega)\right|^2\right\}, \\
&= \sum_{r=0}^{n-1}\sum_{m=0}^{n-1} \gamma_r^n(\omega)\overline{\gamma_m^n(\omega)}\mathbf{E}\left\{\nu_r \overline{\nu_m}\right\}, \\
&= \frac{1}{2\pi}\sum_{r=0}^{n-1}\sum_{m=0}^{n-1} \gamma_r^n(\omega)\overline{\gamma_m^n(\omega)}\int_{-\pi}^{\pi} f(\lambda) e^{j(r-m)\lambda}\,\mathrm{d}\lambda, \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\sum_{r=0}^{n-1} e^{jr\lambda}\gamma_r^n(\omega)\right|^2 f(\lambda)\,\mathrm{d}\lambda, \\
&\leq \frac{\mu}{2\pi}\sum_{r=0}^{n-1}\sum_{m=0}^{n-1} \gamma_r^n(\omega)\overline{\gamma_m^n(\omega)}\int_{-\pi}^{\pi} e^{j(r-m)\lambda}\,\mathrm{d}\lambda, \\
&= \mu\sum_{r=0}^{n-1}|\gamma_r^n(\omega)|^2 = \mu n d(n). \qquad \text{(B.2)}
\end{aligned}$$

Substituting this and (B.7), (B.8) into (B.1) then gives the result. □

Lemma B.1. *Suppose* $\{f_k\}$ *is given by* $f_k = G_T(e^{-j\omega_s k})$ *where* $k = 2\pi/n$. *Suppose* $V_n G_T$ *is constructed by:*

$$V_n G_T = \sum_{m=0}^{d(n)-1} \widehat{f}_m z^m, \quad \widehat{f}_k \triangleq \frac{1}{n} \sum_{r=0}^{n-1} f_r e^{j\omega_s kr} \tag{B.3}$$

Then provided $G_T \in H_\infty(\mathbf{D}_\rho, M)$

$$\|G_T(z) - V_n G_T(z)\|_\infty \leq \frac{M\rho}{\rho^{d(n)}(\rho-1)} \left[1 + \frac{\rho^{d(n)}-1}{\rho^n - 1}\right].$$

Proof. Since $G_T(z) \in H_\infty(\mathbf{D}_\rho, M)$ it can be written as a Taylor series valid on $z \in \mathbf{D}_\rho$:

$$G_T(z) = \sum_{\ell=0}^{\infty} g_\ell z^\ell \tag{B.4}$$

and by Cauchy's estimate $|g_k| \leq M\rho^{-\ell}$. Substituting (B.4) into (B.3) gives:

$$\widehat{f}_k = \frac{1}{n} \sum_{r=0}^{n-1} \left(\sum_{\ell=0}^{\infty} g_\ell e^{-j\omega_s \ell r} \right) e^{j\omega_s kr} = \frac{1}{n} \sum_{\ell=0}^{\infty} g_\ell \sum_{r=0}^{n-1} e^{j\frac{2\pi r(k-\ell)}{n}}.$$

But the complex roots of unity sum to zero so the right hand sum is zero unless $k - \ell = mn$ for $m = 0, 1, 2, \cdots$. Therefore

$$\widehat{f}_k = \sum_{m=0}^{\infty} g_{k+mn}. \tag{B.5}$$

This gives

$$\begin{aligned}
\|G_T(z) - V_n G_T(z)\|_\infty &= \max_{\omega\in[-\pi,\pi]} \left| \sum_{k=0}^{\infty} g_k e^{-j\omega k} - \sum_{k=0}^{d(n)-1} \widehat{f}_k e^{-j\omega k} \right|, && \text{(B.6)} \\
&= \max_{\omega\in[-\pi,\pi]} \left| \sum_{k=0}^{\infty} g_k e^{-j\omega k} - \sum_{k=0}^{d(n)-1} \left(g_k e^{-j\omega k} + \sum_{m=1}^{\infty} g_{k+mn} e^{-j\omega k} \right) \right|, \\
&= \max_{\omega\in[-\pi,\pi]} \left| \sum_{k=d(n)}^{\infty} g_k e^{-j\omega k} - \sum_{k=0}^{d(n)-1} \sum_{m=1}^{\infty} g_{k+mn} e^{-j\omega k} \right|, && \text{(B.7)} \\
&\leq \sum_{k=d(n)}^{\infty} |g_k| + \sum_{k=0}^{d(n)-1} \sum_{m=1}^{\infty} |g_{k+mn}|, \\
&\leq M \sum_{k=d(n)}^{\infty} \frac{1}{\rho^k} + M \sum_{k=0}^{d(n)-1} \frac{1}{\rho^k} \sum_{m=1}^{\infty} \frac{1}{\rho^{mn}}, \\
&= \frac{M\rho}{\rho^{d(n)}(\rho-1)} \left(1 + \frac{\rho^{d(n)}-1}{\rho^n - 1}\right). && \text{(B.8)}
\end{aligned}$$

□

C Proof of Lemma 3.1

Proof. The bound to be minimised is

$$B_n(d) \triangleq \left(\frac{M\rho}{\rho^{d(n)}(\rho-1)} \left[1 + \frac{\rho^{d(n)}-1}{\rho^n - 1} \right] \right)^2 + \mu \frac{d(n)}{n}.$$

Now

$$\frac{\mathrm{d}B_n(d)}{\mathrm{d}d} = \frac{\mu}{n} - \frac{2\,M^2\,(\rho^n-2)\,\left(\rho^n+\rho^d-2\right)\,\ln\rho}{(\rho-1)^2\,\rho^{2\,d-2}\,(\rho^n-1)^2}, \tag{C.1}$$

$$\frac{\mathrm{d}^2B_n(d)}{\mathrm{d}d^2} = \frac{2\,M^2\,(\rho^n-2)\,\left(2\,\rho^n+\rho^d-4\right)\,\ln^2\rho}{(\rho-1)^2\,\rho^{2\,d-2}\,(\rho^n-1)^2} \tag{C.2}$$

The only sign-indefinite term in (C.2) is the numerator component

$$(\rho^n-2)(2\rho^n+\rho^d-4) = (\rho^n-2)^2 + (\rho^n-2)(\rho^n+\rho^d-2).$$

However, since d is chosen to make $\mathrm{d}B_n(d)/\mathrm{d}d = 0$, then by (C.1)

$$(\rho^n-2)(\rho^n+\rho^d-2) = \frac{\mu}{n}\frac{(\rho-1)^2\rho^{2d-2}(\rho^n-1)^2}{2M^2\ln\rho} > 0.$$

Therefore, choosing $d(n)$ to make the first derivative zero ensures that $B_n(d)$ is minimised since at this point the second derivative is always positive. □

References

[1] P. Billingsley, *Convergence of Probability Measures*, John Wiley and Sons., New York, 1968.

[2] D. R. Brillinger, *Time Series: Data Analysis and Theory*, Holden-Day, 1981.

[3] J. Chen, G. Gu, and C. Nett, *Worst case identification of continuous time systems via interpolation*, in Proceedings of the American control conference, San Francisco, 1993, pp. 1544–1548.

[4] J. Chen, C. Nett, and M. Fan, *Worst case system identification in H_∞:Validation of apriori information, essentially optimal algorithms, and error bounds*, Proceedings of the American Control Conference, (1992), pp. 251–257.

[5] D. de Vries, *Identification of Model Uncertainty for Control Design*, PhD thesis, Delft University of Technology, September 1994.

[6] N. Dudley and J. Partington, *Robust identification in the disc algebra using rational wavelets and orthonormal basis functions*, International Journal of Control, 64 (1996), pp. 409–423.

[7] G. Golub and C. V. Loan, *Matrix Computations*, Johns Hopkins University Press, 1989.

[8] G. Gu and P. Khargonekar, *A class of algorithms for identification in H_∞*, Automatica, 28 (1992), pp. 299–312.

[9] ——, *Linear and nonlinear algorithms for identification in H_∞ with error bounds*, IEEE Transactions on Automatic Control, 37 (July 1992), pp. 953–963.

[10] G. Gu, D. Xiong, and K. Zhou, *Identification in H_∞ using Pick's interpolation*, Systems and Control Letters, 20 (1993), pp. 263–272.

[11] A. Helmicki, C. Jacobson, and C. Nett, *H_∞ identification of stable LSI systems: A scheme with direct application to controller design*, Proceedings of the American Control Conference, (1989), pp. 1428–1434.

[12] ——, *Identification in H_∞: A robustly convergent, nonlinear algorithm*, Proceedings of the American Control Conference, (1990), pp. 386–408.

[13] ——, *Identification in H_∞:Linear algorithms*, Proceedings of the American Control Conference, (1990), pp. 2418–2423.

[14] ——, *Identification in H_∞:The continuous time case*, Proceedings of the American Control Conference, (1990), pp. 1893–1898.

[15] ——, *Fundamentals of control oriented system identification and their application for identification in H_∞*, Proceedings of the American Control Conference, (1991), pp. 89–99.

[16] ——, *Control oriented system identification:A worst case/deterministic approach in H_∞*, IEEE Transactions on Automatic Control, 36 (October 1991), pp. 1163–1176.

[17] C. Jacobson and C. Nett, *Worst case system identification in ℓ_1:Optimal algorithms and error bounds*, Proceedings of the American Control Conference, (1991), pp. 3152–3157.

[18] L. Ljung, *System Identification: Theory for the User*, Prentice-Hall, Inc., New Jersey, 1987.

[19] ——, *Some results on identifying linear systems using frequency domain data*, in Proceedings of the 32nd Conference on Decision and Control, December 1993, pp. 3534–3538.

[20] P. Mäkilä and J. Partington, *Robust approximation and identification in H_∞*, Proceedings of the American Control Conference, (1991), pp. 70–76.

[21] ——, *Robust identification of strongly stabilizable systems*, IEEE Transactions on Automatic Control, 37 (1992), pp. 1709–1716.

[22] F. Móricz, R. Serfling, and W. Stout, *Moment and probability bounds with quasi–superadditive structure for the maximum partial sum*, The Annals of Probability, 10 (1982), pp. 1032–1040.

[23] B. Ninness, *Frequency domain estimation using orthonormal bases*, in Proceedings of the 13th IFAC World Congress, San Francisco, 1996.

[24] B. Ninness and F. Gustafsson, *A unifying construction of orthonormal bases for system identification*, IEEE Transactions on Automatic Control, (To Appear, March 1997).

[25] J. Norton and S. Veres, *Topics in Stochastic Systems: Modelling, Estimation and Adaptive Control*, vol. 161 of Lecture Notes in Control and Information Sciences, Springer-Verlag, 1991, ch. Developments in Parameter Bounding, pp. 137–158.

[26] P. Parker, *Frequency Domain Descriptions of Linear Systems*, PhD thesis, Australian National University, 1988.

[27] P. Parker and R. Bitmead, *Adaptive frequency response estimation*, Proceedings of the Conference on Decision and Control, (1987), pp. 348–353.

[28] J. Partington, *Robust identification and interpolation in H_∞*, International Journal of Control, 54 (1991), pp. 1281–1290.

[29] ——, *Robust identification in H_∞*, Journal of Mathematical Analysis and Application, 166 (1991), pp. 428–441.

[30] ——, *Interpolation in normed spaces from the values of linear functionals*, Bulletin of the London Mathematical Society, 26 (1994), pp. 165–170.

[31] R. Pintelon, P. Guillaume, Y. Rolain, J. Schoukens, and H. V. hamme, *Parametric identification of transfer functions in the frequency domain-a survey*, IEEE Transactions on Automatic Control, 39 (1994), pp. 2245–2260.

[32] T.Söderström and P.Stoica, *System Identification*, Prentice Hall, New York, 1989.

[33] P. Van den Hof and D. K. de Vries, *Frequency domain identification with generalized orthonormal basis functions*, in Proceedings of the 34th IEEE Conference on Decision and Control, New Orleans, 1995, pp. 1240–1245.

[34] S. Veres and J. Norton, *Structure selection for bounded-parameter models: Consistency conditions and selection criterion*, IEEE Transactions on Automatic Control, AC-36 (1991), pp. 474–481.