Aspects of Linear Estimation in H_∞.

Brett Ninness*

Abstract

A recent but rapidly maturing field in the area of system identification has been that of 'estimation in H_∞'. Greatly influencing this work has been the phenomenon that no linear (in-the-data) algorithm exists which is 'robustly convergent'. This paper conducts a study of the nature of this issue by combining specific new analysis together with existing results from the mathematics literature on the topic of polynomial approximation theory. Particular attention is paid in this paper to the role of model order, and this leads to the consideration of model order selection from a deterministic worst-case perspective.

Technical Report EE9833, Department of Electrical and Computer Engineering

University of Newcastle, Callaghan 2308, AUSTRALIA

1 Introduction

This paper is concerned with exploring certain aspects of using linear-in-the-data algorithms in an 'estimation in H_∞' context. Very early in the development of this field, it was recognised that such algorithms are not worst-case convergent, and since then the major focus has been on more complicated but effective nonlinear methods that are convergent (Helmicki *et al.* October 1991*a*, Gu *et al.* 1993, Giarré *et al.* 1997, Hakvoort and P.M.J. Van den Hof 1994, Partington 1997, Mäkilä *et al.* 1995).

However, almost equally early (Gu and Khargonekar July 1992*b*), it was also pointed out that linear methods can still perform well for finite data lengths, despite their poor asymptotic performance. The results presented in this paper are motivated by this, and the emphasis here is *not* on presenting new methods which compete with pre-existing ones. Rather, the interest here is three-fold;

1. To explore links between the linear estimation in H_∞ problem and the much older and better studied problem of real valued polynomial interpolation on the real line. To date, these connections appear not to have been fully exploited in the engineering literature.

*This work was supported by the Australian Research Council and the Centre for Integrated Dynamics and Control (CIDAC), University of Newcastle, Callaghan 2308, Australia. This author is with CIDAC and the Department of Electrical & Computer Engineering, University of Newcastle, Australia and can be contacted at email:brett@ee.newcastle.edu.au or FAX: +61 2 49 21 69 93.

2. To provide a better understanding of the nature of linear estimation in H_∞ by presenting analysis of both new and pre-existing methods; much of this is aimed at providing a better understanding of how the so-called 'noise error' behaves.

3. To quantify, as precisely as possible, the influence of model order on estimation error and via this, to introduce the idea of optimal model order selection in a deterministic context.

An overview of the content of this paper in terms of its contributions is as follows:

- In §3, a non-divergent linear algorithm is derived. To our knowledge, this is the first time that such an algorithm has been exhibited, and it provides a counter-example to any misplaced belief that the noise-induced error term in a linear algorithm is necessarily divergent.

- In §4, links to classical real-valued polynomial interpolation are explored, and results from that area are applied to establish the following:

  – For the special case of model order equal to number of data points, a precise expression for the noise induced error is obtained by using results on so-called 'Lebesgue constants' from the mathematical literature. This appears to be the first time such a result has been provided, with all similar results being only over-bounds on error, not expressions for the actual error.

    As well, in deriving the result, it would also appear to be the first time that the link between the Lebesgue constant of interpolation on the real line has been shown to be related to linear H_∞ estimation problem; the mathematical literature in the former area appears to be previously unreferenced in the engineering community concerned with estimation in H_∞.

  – In terms of choice of excitation frequencies, equally spaced points minimise worst case noise induced error. Again this is obtained by establishing how work from the mathematical literature aimed at quantifying Lebesgue constants is related to the linear 'estimation in H_∞' problem.

- In §5 the idea of constraining the derivatives of interpolating polynomial is investigated. A closed form formula for such an interpolant is provided, and worst case error bounds are derived for various norms and noise smoothness assumptions. The contribution here is to highlight how such things as worst-case error are affected by the norm used and the nature of the noise model.

- In §6 the manner in which model order affects worst case error is investigated. Previous studies in this area (Partington 1998) have derived divergence rate under-bounds, but less directly than is done in §6. As well, §6 also derives rate over-bounds, and via this establishes an actual expression for the worst case noise induced error.

- Finally, in §7, the results of §6 are used to suggest the idea of optimal model order selection. For the case of linear algorithms, the existence of an optimal order is established subject to a check involving the assumed rate of decay of true impulse response.

In order to present these new results, it is necessary to put them in context by defining the 'estimation in H_∞' framework being addressed, and also to provide some remarks of a historical nature that clarify the importance of the problems being considered.

2 Problem Setting and Historical Remarks

The problems considered in this paper are the identification ones addressed in the 'estimation in H_∞' literature (Helmicki *et al.* October 1991*a*, Gu and Khargonekar 1992*a*, Gu and Khargonekar July 1992*b*, Partington 1991*b*) wherein a set of n (complex valued) measurements $\{f_0, \cdots, f_{n-1}\}$ are available that represent evaluations at n regularly spaced frequencies $\{0, \omega_s, 2\omega_s, \cdots, (n-1)\omega_s\}$, $\omega_s \triangleq 2\pi/n$ of a linear time invariant system described by a discrete time transfer function $G(z)$ which is itself defined as

$$G(z) \triangleq \sum_{k=0}^{\infty} g_k z^k. \tag{1}$$

Here the sequence $\{g_k\}$ is the discrete time impulse response of the system under investigation. The definition (1) therefore represents the more usual transfer function definition (Oppenheim and Schafer 1989) evaluated at $z := 1/z$. It is assumed that $G(z) \in H_\infty(\mathbf{D}_\rho, M)$ where the latter is defined as the class

$$H_\infty(\mathbf{D}_\rho, M) \triangleq \{G : \ G(z) \text{ is analytic on } \mathbf{D}_\rho \text{ and } |G(z)| \le M \text{ on } \mathbf{D}_\rho\} \tag{2}$$

where M is some finite constant real number, and $\mathbf{D}_\rho \triangleq \{z \in \mathbf{C} : |z| < \rho\}$ is the radius ρ disk with $\mathbf{C}$ denoting the field of complex numbers. In the remainder of this paper it is always assumed (as in work by other authors) that $\rho > 1$.

A key component of the problem is that it is also assumed that the frequency response measurements $\{f_k\}$ are possibly additively corrupted as

$$f_k = G(e^{-j\omega_s k}) + \nu_k \tag{3}$$

where $\{\nu_k\}$ is some complex valued and unknown 'noise' sequence that is assumed uniformly bounded as

$$|\nu_k| \le \varepsilon$$

for some known bound $\varepsilon < \infty$.

The engineering-relevant problem to be solved is that of using the available data $\{f_0, \cdots, f_{n-1}\}$ in order to construct an estimate $\widehat{G}_n(z)$ of $G(z)$ which is of the form

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

Here the notation $d(n)$ in (4) for the order of the estimate $\widehat{G}_n(z)$ is meant to emphasise (as will become clear in the sequel) the possible dependence of model order d on the number of frequency response measurements n.

In the context of system identification, the consideration of this estimation problem (in the special case of $d(n) = n$) appears to have begun with Parker and Bitmead (Parker and Bitmead 1987), and thereafter a large number of other workers have also contributed to the problem, of which a small sample is (Helmicki *et al.* 1989, Helmicki *et al.* 1990*a*, Helmicki *et al.* 1990*b*, Helmicki *et al.* 1990*c*, Helmicki *et al.* October 1991*a*, Helmicki *et al.* 1991*b*, Jacobson and Nett 1991, Partington 1991*b*, Mäkilä and Partington 1991, Partington 1991*a*, Gu and Khargonekar 1992*a*, Gu and Khargonekar July 1992*b*).

This latter intense activity is strongly influenced by the surprising behaviour, first exposed in (Parker and Bitmead 1987), that (least squares calculated) solutions for the estimate (4) that interpolate the frequency response measurements $\{f_k\}$ are not convergent in a worst-case sense.

To be more precise on this point, perhaps the most obvious way of constructing $\widehat{G}_n(z)$ is to try to make the frequency response of $\widehat{G}_n(e^{-j\omega})$ the same as the measurements:

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

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

$$\widehat{\theta}_n \triangleq \left[\widehat{g}_0^n, \cdots, \widehat{g}_{d(n)-1}^n\right]^T, \qquad F \triangleq [f_0, \cdots, f_{n-1}]^T$$

$$[\Omega_n]_{k,\ell} = e^{-jk\ell\omega_s}, \qquad k = 0, \cdots, n-1, \qquad \ell = 0, \cdots, d(n)-1$$

so that (5) becomes $\Omega_n\widehat{\theta}_n = F$ with solution minimising the squared error as

$$\widehat{\theta}_n = \left(\Omega_n^\star \Omega_n\right)^{-1} \Omega_n^\star F = n^{-1}\Omega_n^\star F$$

($\cdot^\star$ denotes 'conjugate transpose') which implies

$$\widehat{g}_k^n = \frac{1}{n}\sum_{m=0}^{d(n)-1} f_m e^{jk\omega_s m}. \tag{6}$$

If, in fact, the choice made in (Parker and Bitmead 1987) of $d(n) = n$ is used, then (6) leads to

$$\widehat{G}_n(z) = \frac{1}{n}\sum_{k=0}^{n-1}\sum_{r=0}^{n-1} f_r e^{jkr\omega_s} z^k = \frac{1}{n}\sum_{r=0}^{n-1} f_r \sum_{k=0}^{n-1}\left[e^{jr\omega_s}z\right]^k = \frac{1}{n}\sum_{r=0}^{n-1} f_r \left(\frac{1-z^n}{1-e^{jr\omega_s}z}\right) \tag{7}$$

where in progressing to the last expression the fact that $n\omega_s = 2\pi$ was used.

Unfortunately, as calculated in (Parker and Bitmead 1987, Parker 1988) and later extensively corroborated (Helmicki *et al.* October 1991*a*, Gu and Khargonekar July 1992*b*), the tightest upper bound available on the worst-case estimation error involved with (7) is of the form

$$e_n \triangleq \sup_{\substack{G \in H_\infty(\mathbf{D}_\rho, M) \\ \|\nu\|_\infty \le \varepsilon}} \left\| G - \widehat{G}_n \right\|_\infty \le \frac{2M}{\rho^n(\rho - 1)} + \varepsilon\left(1 + \frac{2}{\pi}\log n\right) \tag{8}$$

which clearly diverges with increasing n.

As previously alluded to, this phenomenon has strongly influenced the large body of work (Helmicki *et al.* October 1991*a*, Helmicki *et al.* 1991*b*, Jacobson and Nett 1991, Partington 1991*b*, Mäkilä and Partington 1991, Partington 1991*a*, Gu and Khargonekar 1992*a*, Gu and Khargonekar July 1992*b*) that seeks to avoid the divergent behaviour by developing so-called 'robustly convergent' estimates $\widehat{G}_n(z)$ whose defining property is that

$$\lim_{\substack{\varepsilon \to 0 \\ n \to \infty}} \sup_{\substack{G \in H_\infty(\mathbf{D}_\rho, M) \\ \|\nu\|_\infty \le \varepsilon}} \left\| G - \widehat{G}_n \right\|_\infty = 0 \tag{9}$$

is satisfied.

Germane to this latter study is the major result of Partington (Partington 1991*b*) which exposed that there was no algorithm that was a linear operation on the data $\{f_0, \cdots, f_{n-1}\}$ which would satisfy the criterion (9). Consequently, there is no hope that by regulating the model order $d(n)$ in (6), (4) that a divergent worst case bound like (8) might be avoided and (9) satisfied.

The purpose of this paper is not to contribute to the already rich literature that does in fact satisfy (9) by various non-linear in the data estimation strategies (Helmicki *et al.* October 1991*a*, Jacobson and Nett 1991, Partington 1991*b*, Mäkilä and Partington 1991, Partington 1991*a*, Gu and Khargonekar 1992*a*, Gu and Khargonekar July 1992*b*).

Rather, the aim here is to delve more deeply into the genesis and precise nature of the divergence phenomenon underlying (8). Part of this study has already been undertaken in (Ninness 1998) where a main result is that provided $d(n) = \mathsf{o}\{\sqrt{n}\}$ as $n \to \infty$, then in fact the estimate (6), (4) does satisfy the robust estimation criterion (9) except on a set of measure zero, the measure being the probability measure used to characterise $\{\nu_k\}$ as a (possible non-stationary) zero mean stochastic process with some constraints on its 'memory'; see (Ninness 1998) for more detail.

However, in examining the performance of linear in the data methods, this paper has a different scope to (Ninness 1998) in that it examines schemes other than (6), (4) and also completely eschews any probabilistic formulation of the assumptions on $\{\nu_k\}$.

3 A Non-Divergent Linear Scheme

The section derives a linear estimation algorithm which is *not* worst-case divergent. It is believed that this is the first time that such a scheme has been exhibited. It's engineering utility as compared to pre-existing schemes is problem dependent, but quite apart from this it is of interest as a contribution to the understanding of the nature of linear H_∞ estimation methods. For example, it appears to be a widely held perception that any linear algorithm inherits its divergence from the noise induced error. By way of contrast, the algorithm presented via Theorem 3.2 below has a noise induced error which is worst-case convergent.

Before presenting the new algorithm, some necessary background involves recognising that although linear estimators such as (4), (6) start by taking noise corrupted measurements $\{f_0, \cdots, f_{n-1}\}$ and delivering an estimate $G_n(z) \in H_\infty(\mathbf{T})$ [1] they could, by assuming (as in (Partington 1991*b*)) that the $\{\nu_k\}$ are evaluations at the points $\{e^{jk\omega_s}\}$ of a continuous function [2] $\nu \in C(\mathbf{T})$, be equally well considered as particular linear operators, call them V_n, such that

$$V_n : C(\mathbf{T}) \rightarrow H_\infty(\mathbf{T}), \tag{10}$$

$$G(z) + \nu(z) \mapsto \widehat{G}_n(z). \tag{11}$$

With this notation in hand, the key result of (Partington 1991*b*) is [3]

Theorem 3.1. *Partington (Partington 1991*b*). There does not exist a sequence of uniformly bounded linear maps $V_n : C(\mathbf{T}) \rightarrow H_\infty(\mathbf{T})$ such that for every polynomial $g \in A(\mathbf{D})$*

$$\lim_{n\to\infty} \|V_n g - g\|_\infty = 0. \tag{12}$$

Now, since the operator V_n is linear, then

$$\|G - \widehat{G}_n\|_\infty \leq \|G - V_n G\|_\infty + \|V_n \nu\|_\infty. \tag{13}$$

Therefore, if an algorithm is to be robustly convergent then for every disturbance sequence ν in some $\|\cdot\|_\infty$ constrained ball, the component $V_n(\nu)$ must also be constrained to a ball for all n greater than some N. Also, in a complete space point-wise boundedness of an operator implies boundedness of operator norm so that there must exist some $K < \infty$ such that $\|V_n\| < K$ for all $n > N$. However Theorem 3.1 asserts that no such uniformly bounded operator sequence $\{V_n\}$ exists such that (since the polynomials are dense in $C(\mathbf{T})$) the 'undermodelling' error $\|G - V_n G\|_\infty$ tends to zero. Hence, no robustly convergent algorithm exists.

On an intuitive level, the reason that the noise term $\|V_n(\nu)\|_\infty$ grows with increasing n for the particular estimator (4), (6) is that via its formulation (7) for the case of $d(n) = n$, the estimate is formed as a convolution between the measurements $\{f_n\}$ and an interpolating function $(1 - e^{-j\omega n})/(1 - e^{j(r\omega_s - \omega)})$. The magnitude of this interpolating function is $|(\sin \omega n/2)(\sin(\omega - r\omega_s)/2)^{-1}|$, and the sum of the magnitudes of the maxima of this latter function grows with n at the rate $\log n$. It is therefore possible to find a bounded disturbance which gives an estimation error that grows like $\log n$. Hence the bound (8).

This heuristic explanation leads to the first result of the paper. Namely, that an obvious solution to avoiding the $\log n$ growth in the noise term $\|V_n \nu\|_\infty$ is to make the interpolating

---

[1] where $\mathbf{T} = \{z \in \mathbf{C} : |z| = 1\}$ is the complex unit circle, and $H_\infty(\mathbf{T})$ is the set of functions that are of bounded essential supremum on $\mathbf{T}$ and are analytic on $\mathbf{D}$, the open unit disc that is interior to $\mathbf{T}$.

[2] $C(\mathbf{T})$ is the class of functions that are continuous on $\mathbf{T}$.

[3] $A(\mathbf{D})$ is the so-called disc algebra of functions continuous on $\mathbf{T}$ and analytic on $\mathbf{D}$ so that $A(\mathbf{D}) = H_\infty(\mathbf{D}) \cap C(\mathbf{T})$.

function die in magnitude more quickly, and this can easily be achieved by simply squaring the original interpolating function (7). The analysis of this modification is presented in the following theorem.

Theorem 3.2. *Consider the linear in-the-data estimation scheme*

$$\widehat{G}_n(z) = \frac{1}{n^2}\sum_{r=0}^{n-1} f_r \left[\sum_{k=0}^{n-1}[e^{jr\omega_s}z]^k\right]^2 \tag{14}$$

then

$$\begin{aligned}\sup_{\substack{G\in H_\infty(\mathbf{D}_\rho,M)\\ \|\nu\|_\infty\le\varepsilon}} \left\|G-\widehat{G}_n\right\|_\infty \quad &\le \quad \left(\frac{M\rho}{\rho-1}\right)\left\{2\left|\sin\frac{\omega n}{2}\right| + \frac{1}{n\rho^n} + \frac{2}{\rho^{2n}}\right\} \\ &\qquad +\varepsilon\left(1+\frac{5}{n}+\frac{4}{\pi}\frac{\log n}{n}\right).\end{aligned} \tag{15}$$

Proof. Write $f_r = \eta_r + \nu_r$ where $\eta_r \triangleq G(e^{-j\omega_s r})$ and use the linear operator notation as in (10), (11) and (13). Consider the noise induced error component first. Then with the definition $2\phi = r\omega_s - \omega$

$$\begin{aligned}
|V_n\nu| &= \frac{1}{n^2}\left|\sum_{r=0}^{n-1}\nu_r\left[\sum_{k=0}^{n-1}e^{jk(r\omega_s-\omega)}\right]^2\right| \\
&\le \frac{\varepsilon}{n^2}\sum_{r=0}^{n-1}\left|\left[\sum_{k=0}^{n-1}e^{j2\phi k}\right]^2\right| \\
&= \frac{\varepsilon}{n^2}\sum_{r=0}^{n-1}\left|(\cos 2\phi - j\sin 2\phi)\left(\frac{\sin^2 n\phi}{\sin^2\phi}\right)e^{j2\phi n}\right| \\
&= \frac{\varepsilon}{n^2}\sum_{r=0}^{n-1}\left|\sin^2 n\phi\left(\frac{(1-2\sin^2\phi)}{\sin^2\phi} - j\frac{2\sin\phi\cos\phi}{\sin^2\phi}\right)e^{j2n\phi}\right| \\
&\le \frac{\varepsilon}{n^2}\sum_{r=0}^{n-1}\left|\operatorname{cosec}^2\phi - 2 - 2j\cot\phi\right|\sin^2 n\phi \\
&\le \frac{\varepsilon}{n^2}\sin^2\frac{\omega n}{2}\sum_{r=0}^{n-1}\left\{\operatorname{cosec}^2\left(\frac{\omega}{2}-\frac{\pi r}{n}\right) + 2 + 2\left|\cot\left(\frac{\omega}{2}-\frac{\pi r}{n}\right)\right|\right\}.
\end{aligned}$$

However, by equation 14.1.1 on page 259 of (Hanson 1975)

$$\sum_{r=0}^{n-1}\operatorname{cosec}^2\left(\frac{\omega}{2}-\frac{\pi r}{n}\right) = n^2\operatorname{cosec}^2\frac{\omega n}{2}.$$

Using this and Lemma A.2 then provides

$$|V_n\nu| \le \varepsilon + \frac{2\varepsilon}{n} + \frac{2\varepsilon}{n}\left(1.5 + \frac{2}{\pi}\log n\right) = \varepsilon\left(1 + \frac{5}{n} + \frac{4}{\pi}\frac{\log n}{n}\right).$$

This takes care of the 'noise' term $V_n\nu$. For the 'undermodelling' term $G - V_nG$ use of the notation $\{g_k\}$ for the impulse response of $G(z)$ provides

$$\begin{aligned}
\frac{1}{n^2}\sum_{r=0}^{n-1}\eta_r\sum_{k=0}^{n-1}e^{j(r\omega_s-\omega)k}\sum_{m=0}^{n-1}e^{j(r\omega_s-\omega)m} &= \frac{1}{n^2}\sum_{k=0}^{n-1}e^{-j\omega k}\sum_{m=0}^{n-1}e^{-j\omega m}\sum_{r=0}^{n-1}\eta_r e^{jr\omega_s(k+m)}\\
&= \frac{1}{n^2}\sum_{k=0}^{n-1}e^{-j\omega k}\sum_{m=0}^{n-1}e^{-j\omega m}\sum_{r=0}^{n-1}\sum_{\ell=0}^{\infty}g_\ell e^{-jr\omega_s\ell}e^{jr\omega_s(k+m)}\\
&= \frac{1}{n^2}\sum_{k=0}^{n-1}e^{-j\omega k}\sum_{m=0}^{n-1}e^{-j\omega m}\sum_{\ell=0}^{\infty}g_\ell\sum_{r=0}^{n-1}e^{j(k+m-\ell)2\pi r/n}\\
&= \frac{1}{n^2}\sum_{k=0}^{n-1}e^{-j\omega k}\sum_{m=0}^{n-1}e^{-j\omega m}\sum_{r=0}^{n-1}\sum_{\ell=0}^{\infty}g_{(m+k)\ (\mathrm{mod}\ n+\ell n)}.
\end{aligned}$$

Now, $(m+k)\,(\mathrm{mod}\ n) = 0$ for $n-1$ different permutations of m and k for $m,k \le n-1$. Therefore

$$\begin{aligned}
\sum_{m=0}^{n-1}\sum_{k=0}^{n-1}g_{(m+k)\ (\mathrm{mod}\ n)}e^{-j\omega(m+k)} &= g_0\left(1 + \underbrace{e^{-j\omega n} + e^{-j\omega n} + \cdots + e^{-j\omega n}}_{n-1\text{ terms}}\right) +\\
&\quad g_1\left(e^{-j\omega} + \underbrace{e^{-j\omega(n+1)} + e^{-j\omega(n+1)} + \cdots + e^{-j\omega(n+1)}}_{n-1\text{ terms}}\right) + \cdots +\\
&\quad g_{n-1}\left(e^{-j\omega(n-1)} + \underbrace{e^{-j\omega(2n-1)} + e^{-j\omega(2n-1)} + \cdots + e^{-j\omega(2n-1)}}_{n-1\text{ terms}}\right)\\
&= \left(1 + (n-1)e^{-j\omega n}\right)\sum_{m=0}^{n-1}g_m e^{-j\omega m}.
\end{aligned}$$

Similarly

$$\sum_{m=0}^{n-1}\sum_{k=0}^{n-1}g_{(m+k)\ (\mathrm{mod}\ n)+n}e^{-j\omega(m+k)} = \left(1 + (n-1)e^{-j\omega n}\right)\sum_{m=0}^{n-1}g_{m+n}e^{-j\omega m}$$

and so on, so that

$$V_nG = \left[\frac{1 + (n-1)e^{-j\omega n}}{n}\right]\sum_{\ell=0}^{\infty}\sum_{m=0}^{n-1}g_{m+\ell n}e^{-j\omega m}. \tag{16}$$

This leads to

$$
\begin{aligned}
|G - V_n G| &= \left| \sum_{m=0}^{\infty} g_m e^{-j\omega m} - \left[\frac{1+(n-1)e^{-j\omega n}}{n}\right] \sum_{\ell=0}^{\infty} \sum_{m=0}^{n-1} g_{m+\ell n} e^{-j\omega m} \right| \\
&= \left| \left(1 - \left[\frac{1+(n-1)e^{-j\omega n}}{n}\right]\right) \sum_{m=0}^{n-1} g_m e^{-j\omega m} + \right. \\
&\quad \left(1 - e^{j\omega n}\left[\frac{1+(n-1)e^{-j\omega n}}{n}\right]\right) \sum_{m=n}^{2n-1} g_m e^{-j\omega m} + \\
&\quad \left. \left(1 - e^{j2\omega n}\left[\frac{1+(n-1)e^{-j\omega n}}{n}\right]\right) \sum_{m=2n}^{3n-1} g_m e^{-j\omega m} + \cdots \right| \\
&= \left| \left(\frac{n-1}{n}\right)\left(1 - e^{-j\omega n}\right) \sum_{m=0}^{n-1} g_m e^{-j\omega m} + \frac{\left(1 - e^{j\omega n}\right)}{n} \sum_{m=n}^{2n-1} g_m e^{-j\omega n} + \right. \\
&\quad \left. \sum_{\ell=2}^{\infty} \left[\left(1 - e^{j\omega(\ell-1)n}\right) - \frac{e^{j\omega \ell n}\left(1 - e^{-j\omega n}\right)}{n}\right] \sum_{m=\ell n}^{(\ell+1)n-1} g_m e^{-j\omega m} \right| \\
&\leq M\left(\frac{n-1}{n}\right)\left|2\sin\frac{\omega n}{2}\right|\left(\frac{\rho}{\rho-1}\right)\left(\frac{\rho^n-1}{\rho^n}\right) + \frac{M}{n\rho^n}\left(\frac{\rho}{\rho-1}\right)\left(\frac{\rho^n-1}{\rho^n}\right) + \\
&\quad \frac{2M}{\rho^{2n}}\left(\frac{\rho}{\rho-1}\right)\left(\frac{\rho^n-1}{\rho^n}\right).
\end{aligned}
$$

$\square$

As far as the authors are aware, this is the first time that a non-divergent linear algorithm for estimation in H_∞ has been derived, and demonstration of its existence provides further understanding of the nature of the divergence exposed via Theorem 3.1.

Specifically, when the Parker and Bitmead estimate is modified as in (14) to prevent the component in the estimate due to noise diverging, then Theorem 3.2 shows that the 'undermodelling term', although interpolating the response $\widehat{G}_n(z)$, does not converge uniformly to it; although it doesn't diverge either. This highlights that there is a tradeoff between 'noise error' and 'undermodelling error' in the construction of linear algorithms.

Certainly, depending on M, ρ and ε, then for certain values of n one will obtain estimates with smaller error bounds using the linear estimator (14) rather than the more commonly employed linear estimator (Parker and Bitmead 1987, Helmicki *et al.* October 1991*a*, Gu and Khargonekar July 1992*b*) (4), (6).

As a final comment substantiating the possible lack of smoothness of $\nu(z)$ being responsible for divergence phenomena, note that if it is appropriate to model $\nu(z)$ as being analytic on $\mathbf{D} \cup \mathbf{T}$ then the simple linear method (7) is worst-case convergent as established by the following lemma.

Lemma 3.1. *Suppose that $\nu(z)$ is analytic on $\mathbf{D} \cup \mathbf{T}$. Then with $\widehat{G}_n(z)$ being the linear interpolant defined by (7) the robust convergence criterion*

$$\lim_{\substack{\varepsilon \to 0 \\ n \to \infty}} \sup_{\substack{G \in H_\infty(\mathbf{D}_\rho, M) \\ \|\nu\|_\infty \le \varepsilon}} \left\| G - \widehat{G}_n \right\|_\infty = 0$$

is satisfied.

Proof. Using the V_n operator notation defined in (10),(11) then via the definition (7) V_n is linear so that

$$\|G - \widehat{G}_n\|_\infty \le \|G - V_n G\|_\infty + \|V_n \nu - \nu\|_\infty + \|\nu\|_\infty.$$

But Theorem A in (Chui *et al.* 1993) asserts that under the given assumptions $\lim_{n\to\infty} \|V_n \nu - \nu\|_\infty = 0$ and $\lim_{n\to\infty} \|G - V_n G\|_\infty = 0$ so that since by assumption $\|\nu\|_\infty < \varepsilon$ the result follows. □

4 Links to Polynomial Interpolation

This section explores some aspects of the parallel relationship between the divergence phenomenon (8) and the 'classical' problem of possible divergence of Lagrange interpolants of real valued continuous functions. As already detailed in §3, the former problem of divergence of (8) has had a strong impact on the 'estimation in H_∞' literature (Helmicki *et al.* October 1991*a*, Gu and Khargonekar July 1992*b*, Partington 1991*b*). As well, in a manner that resonates strongly with this history of work in estimation in H_∞, the latter problem of Lagrange interpolation divergence has also had, over many decades, a profound effect on classical polynomial approximation theory (Cheney 1966, Erdös 1961*b*, Turán 1980), with the divergence problem being solved by constraining the derivative of the interpolating polynomial.

Given the close connections between the two linear problems (they are identical, save that in the context of this paper the data is complex valued), it would seem important to closely examine the importance of the (by now very old) derivative constrained linear solutions in an estimation in H_∞ context.

To explain these ideas, first it should be made clear that the classical Lagrange interpolation problem (Cheney 1966) is one wherein an $n-1$'th order polynomial

$$p_n(x) = p_0 + p_1 x + p_2 x^2 + \cdots + p_{n-2} x^{n-2} + p_{n-1} x^{n-1}$$

is sought such that for some real valued continuous function $g(x) : [-1, 1] \to \mathbf{R}$ and for some set of n interpolating points $\{x_0^n, \cdots, x_{n-1}^n\} \subset [-1, 1]$ then $p_n(x_k^n) = g(x_k^n)$ for all $k = 0, 1, \cdots, n-1$. Unfortunately, the worst-case error

$$e_n \triangleq \sup_{\substack{x \in [-1,1] \\ g \in C([-1,1])}} |p_n(x) - g(x)| \tag{17}$$

was discovered by Faber in 1914 (Erdös 1961*b*) to grow like $\log n$ as $n \to \infty$. This divergence is clearly reminiscent of the phenomenon (8), with the former being commented on by Erdös (Erdös 1961*a*) as being '*in contrary to everything what was expected since NEWTON*'(sic).

The problem of exactly quantifying the worst case noise error (17) has been intensively studied, with e_n defined in (17) being known as the 'Lebesgue constant' associated with a particular set of interpolating points $\{x_k^n\}$. To date, no exact quantification of the Lebesgue constant is available for any choice of interpolating points, but very precise asymptotic rate estimates are available for certain special cases. Application of these estimates then provides quite precise quantification of the worst-case noise induced error involved in the scheme (4), (6) as follows.

Lemma 4.1. *Consider the linear estimation scheme (4), (6) with the model order choice $d(n) = n$. Then with the V_n operator notation introduced in (10), (11) and assuming that $\nu \in C(\mathbf{T})$*

$$\sup_{\|\nu\|_\infty \le \varepsilon} \|V_n \nu\|_\infty = \frac{2}{\pi}\left(\log n + \gamma + \log\frac{8}{\pi}\right) + \sum_{k=1}^{\infty}\frac{a_k}{n^{2k}} + O\left(\left[\frac{\log\log n}{\log n}\right]^2\right)$$

as $n \to \infty$ where

$$a_k \triangleq (-1)^{k+1}(2^{2k-1}-1)^2 \pi^{2k-1}\frac{B_{2k}^2}{4^{k-1}k(2k)!},$$

with γ being Euler's constant ≈ 0.57722, and the $\{B_k\}$ are the Bernoulli numbers defined by $t(e^t-1)^{-1} = \sum_{n=0}^{\infty} B_n t^n/n!$ for $|t| < 2\pi$ so that it is possible to establish that

$$0 < \sum_{k=1}^{\infty}\frac{a_k}{n^{2k}} < \frac{\pi}{72n^2}.$$

Proof. Clearly the quantity $\sup_{\|\nu\|_\infty \le \varepsilon}\|V_n\nu\|_\infty$ in question is the Lebesgue constant of polynomial interpolation on $\mathbf{T}$ with equidistant spaced nodes $z_k^n = e^{j2\pi k/n}, k = 0,1,\cdots,n-1$. However, Brutman has established (Brutman 1980) that this is identical to the Lebesgue constant associated with trigonometric interpolation on $[0, 2\pi]$ also at equidistant spaced points $\theta_k^n = 2\pi k/n, k = 0,1,\cdots,n-1$. Continuing in this theme, Ehlich and Zeller established (Günttner 1980) that in turn, this is identical to the Lebesgue constant associated with real-valued interpolation on $[-1,1]$ at nodes x_k^n equal to the zeros of the n'th order Tchebychev polynomial $\cos(n\cos^{-1}x)$. Finally, combining the results of Vértesi (Vértesi 1990) and Günttner (Günttner 1980) provides the currently best available bounds on the Lebesgue constant associated with this latter Tchebychev node real valued polynomial interpolation problem; these are the bounds quoted in the lemma. □

The significance of this result is that it provides (it would appear for the first time) rather precise information on the rate (in n) of divergence of the linear scheme (6), (4) with $d(n) = n$. Previous work (Parker and Bitmead 1987, Helmicki *et al.* October 1991*a*, Gu and Khargonekar July 1992*b*) has separately provided over-bounds (such as (8)) or under-bounds (Partington 1998) but has not established their tightness. A key point is that the result in Lemma 4.1 has been obtained quite cheaply by drawing on deep analysis in the work of other authors.

In relation to this strategy, it may also be applied to the prima-facie difficult question as to how $\sup_{\|\nu\|_\infty<\varepsilon}\|V_n\nu\|_\infty$ is affected by the choice of frequencies $\{\omega_k\}$ at which the measurements $\{\nu_k = \nu(e^{-j\omega_s k})\}$ are obtained. In fact, Erdös has already considered this problem by conjecturing that it was optimal with respect to sup norm to interpolate at evenly spaced points (Erdös 1968, Turán 1980) such as specified in (3). It turns out that this is correct.

Lemma 4.2. *Consider, when using the estimation algorithm (6), (4) with $d(n) = n$, the problem of minimising*

$$\sup_{\|\nu\|_\infty<\varepsilon}\|V_n\nu\|_\infty$$

by choice of the frequencies $\{\omega_0, \cdots, \omega_{n-1}\}$ at which the measurements $\{f_0, \cdots, f_{n-1}\}$ are taken. The the solution providing this minimum worst case error is $\omega_t = 2\pi t/n = t\omega_s$.

Proof. This is equivalent to the question of optimal nodes for complex valued polynomial interpolation of continuous function on the unit circle, and it has been proven in (De Boor and Pinkus 1978, Brutman 1980, Brutman and Pinkus 1980) that in this case equidistant spaced nodes produce the smallest Lebesgue constant. □

5 Birkhoff Interpolation

Continuing now to consider how the divergence (8) might be mollified by a more sophisticated linear algorithm than that of (6), (4) it is important to note that Fejér solved the worst case divergence problem associated with real valued polynomial interpolation (together with Hermite) by constraining the polynomial approximation to not only interpolate the measurements, but also to have 'small' derivative at the interpolating points - these were constrained to be at the zeroes of the nth order Tchebychev polynomial $\cos(n\cos^{-1}x)$(Turán 1980). This amounts to a constructive proof of the famous Weierstrass approximation Theorem.

Unfortunately, this solution in the real-valued case is not directly transportable to the complex-valued setting because of the added requirement of analyticity. That is, although the real and imaginary parts of the measurements could be separately interpolated, the interpolants would not necessarily satisfy the Cauchy–Riemann equations, and hence their complex valued sum would not necessarily represent an analytic function.

So direct application of Fejér-Hermite interpolation will not be fruitful, but using it as motivation it is natural to explore the basic theme of constraining derivatives while at the same time imposing analyticity of interpolant.

Specifically, suppose that the derivatives of $\widehat{G}_n(z)$ are constrained to be zero at the n roots of unity and suppose it is also elected to specify that $\widehat{G}_n(z)$ interpolate the measurements $\{f_k\}$ at these same points. Finally, suppose that $\widehat{G}_n(z)$ is required to be a polynomial in z. Interpolation under these requirements is known in the mathematics literature as Birkhoff interpolation (Turán 1980). The construction of the required $\widehat{G}_n(z)$ of order $2n-1$ is unique and is given by

$$\widehat{G}_n(z) = V_n(G+\nu) \triangleq \frac{1}{n}\sum_{r=0}^{n-1} f_r \sum_{k=0}^{n-1}\left[e^{j\omega_s r}z\right]^k\left[1-\frac{k(z^n-1)}{n}\right]. \tag{18}$$

Note the similarity to the Lagrange interpolant (7). The convergence properties of the estimate (18) are examined in the following three theorems.

Theorem 5.1. *Suppose $\{\nu_k\}$ is obtained by sampling a function ν as $\nu_k = \nu(e^{-j\omega_s k})$ where $\omega_s \triangleq 2\pi/n$ and $\nu(z)$ is analytic on $\mathbf{D}$ and Lipschitz continuous of order $\alpha > 0$ on $\mathbf{T}$. Then for $\widehat{G}_n(z)$ given by (18)*

$$\|G - \widehat{G}_n\|_p \leq \varepsilon + \frac{C}{n^\alpha} \tag{19}$$

for any $0 < p < \infty$ and some $C < \infty$ that is independent of n.

Proof. As defined in (10), (11), the operator V_n is linear so in analogy with the decomposition (13)

$$\|G - \widehat{G}_n\|_p \leq \|G - V_n G\|_p + \|V_n \nu - \nu\|_p + \|\nu\|_p. \tag{20}$$

Now, since $G \in H_\infty(\mathbf{D}_\rho, M)$ for $\rho > 1$ then $G(e^{-j\omega})$ is Lipschitz continuous on $\mathbf{T}$ (Davis 1963). However, Theorem 1 in (Varma 1988) establishes that for any $f \in \mathrm{Lip}\,\alpha$ on $\mathbf{T}$, there exists a $C < \infty$ which is independent of n and such that $\|f - V_n f\|_p \leq C n^{-\alpha}$, hence the result. $\square$

Therefore, for a class of disturbances that come from a function more smooth than those in $C(\mathbf{T})$, then via the classical ideas of Birkhoff interpolation it is possible to construct a linear robustly convergent estimator via (18) that is worst case convergent in the sense (derived from (8)) that

$$\lim_{\substack{\varepsilon \to 0 \\ n \to \infty}} \sup_{\substack{G \in H_\infty(\mathbf{D}_\rho, M) \\ \|\nu\|_\infty \leq \varepsilon}} \left\| G - \widehat{G}_n \right\|_p = 0$$

for $p < \infty$. Again, this provides insight into the nature of the worst-case divergence properties of linear algorithms - the divergence is crucially dependent on the lack of Lipschitz continuity in $\nu(e^{-j\omega})$ and on the p-norm strictly being one where $p = \infty$.

Given these results, an obvious question is how the Birkhoff interpolant (18) behaves under the stricter conditions considered in the estimation in H_∞ literature wherein $p = \infty$ and there is no smoothness constraint imposed on the measurement corrupting sequence $\{\nu_k\}$. The following theorem addresses this question.

Theorem 5.2. *For the Birkhoff interpolant solution (18)*

$$\begin{aligned}\sup_{\substack{G \in H_\infty(\mathbf{D}_\rho, M) \\ \|\nu\|_\infty \leq \varepsilon}} \left\| G - \widehat{G}_n \right\|_\infty \quad \leq \quad & \frac{2M\rho}{(\rho - 1)} \left\{ \frac{1}{n(\rho - 1)} - \frac{1}{\rho^n(\rho^n - 1)} \right\} + \\ & 2\varepsilon \left(3 + \frac{2}{\pi} \log n \right).\end{aligned}$$

Proof. Direct application of simple trigonometric identities provides the useful formulation

$$\sum_{k=0}^{n-1} e^{j(r\omega_s-\omega)k} = \sin\frac{\omega n}{2} e^{-j\omega n/2}\left[\cot\left(\frac{\omega}{2}-\frac{\pi r}{n}\right)+j\right] \triangleq L_r(\omega,n).$$

Note that

$$j\frac{\mathrm{d}L_r(\omega,n)}{\mathrm{d}\omega} = j\sum_{k=0}^{n-1}(-jk)e^{j(r\omega_s-\omega)k} = \sum_{k=0}^{n-1} k e^{j(r\omega_s-\omega)k}.$$

Therefore, it is possible to express the frequency response of the Birkhoff interpolant (18) as

$$\widehat{G}_n(e^{-j\omega}) = \frac{1}{n}\sum_{r=0}^{n-1} f_r M_r(\omega,n) \tag{21}$$

where

$$\begin{aligned}
M_r(\omega,n) &\triangleq L_r(\omega,n) - \frac{j}{n}(e^{-j\omega n}-1)\frac{\mathrm{d}L_r(\omega,n)}{\mathrm{d}\omega}\\
&= \left\{\cot\left(\frac{\omega}{2}-\frac{\pi r}{n}\right)+j\right\}\sin\frac{\omega n}{2}e^{-j\omega n/2} - j(e^{-j\omega n}-1)\times\\
&\quad \left\{-\frac{1}{2n}\operatorname{cosec}^2\left(\frac{\omega}{2}-\frac{\pi r}{n}\right)\sin\frac{\omega n}{2}e^{-j\omega n/2} - \right.\\
&\quad \frac{j}{2}\left(\cot\left(\frac{\omega}{2}-\frac{\pi r}{n}\right)+je^{-j\omega n}\right)\sin\frac{\omega n}{2}e^{-j\omega n/2}+\\
&\quad \left.\frac{1}{2}\left(\cot\left(\frac{\omega}{2}-\frac{\pi r}{n}\right)+j\right)\cos\frac{\omega n}{2}e^{-j\omega n/2}\right\}\\
&\triangleq A_r(\omega,n)+B_r(\omega,n).
\end{aligned}$$

In this last line, the following definitions (with subsequent indicated simplifications) have evidently been made

$$\begin{aligned}
A_r(\omega,n) &\triangleq \sin\frac{\omega n}{2}e^{-j\omega n/2}\cot\left(\frac{\omega}{2}-\frac{\pi r}{n}\right) + j(e^{-j\omega n}-1)\times\\
&\quad \left\{\frac{j}{2}\sin\frac{\omega n}{2}e^{-j\omega n/2} - \frac{1}{2}\cos\frac{\omega n}{2}e^{-j\omega n/2}\right\}\cot\left(\frac{\omega}{2}-\frac{\pi r}{n}\right)+\\
&\quad \frac{j}{2n}(e^{-j\omega n}-1)\operatorname{cosec}^2\left(\frac{\omega}{2}-\frac{\pi r}{n}\right)\sin\frac{\omega n}{2}e^{-j\omega n/2}\\
&= \sin\frac{\omega n}{2}e^{-j\omega n/2}\cot\left(\frac{\omega}{2}-\frac{\pi r}{n}\right) - \frac{j}{2}(e^{-j\omega n}-1)\times\\
&\quad \left\{\cos\frac{\omega n}{2} - j\sin\frac{\omega n}{2}\right\}e^{-j\omega n/2}\cot\left(\frac{\omega}{2}-\frac{\pi r}{n}\right)+\\
&\quad \frac{1}{n}\sin^2\frac{\omega n}{2}e^{-j\omega n/2}\operatorname{cosec}^2\left(\frac{\omega}{2}-\frac{\pi r}{n}\right)
\end{aligned}$$

$$
\begin{aligned}
&= e^{-j\omega n/2}\cot\left(\frac{\omega}{2}-\frac{\pi r}{n}\right)\left\{\sin\frac{\omega n}{2}-\frac{j(e^{-j\omega n}-1)e^{j\omega n/2}}{2}\right\}+\\
&\quad \frac{1}{n}\sin^2\frac{\omega n}{2}e^{-j\omega n/2}\operatorname{cosec}^2\left(\frac{\omega}{2}-\frac{\pi r}{n}\right)\\
&= e^{-j\omega n/2}\cot\left(\frac{\omega}{2}-\frac{\pi r}{n}\right)\left\{\sin\frac{\omega n}{2}-\frac{j}{2}e^{-j\omega n}\left[-2j\sin\frac{\omega n}{2}\right]\right\}+\\
&\quad \frac{1}{n}\sin^2\frac{\omega n}{2}e^{-j\omega n/2}\operatorname{cosec}^2\left(\frac{\omega}{2}-\frac{\pi r}{n}\right)\\
&= e^{-j\omega n/2}\cot\left(\frac{\omega}{2}-\frac{\pi r}{n}\right)\sin\frac{\omega n}{2}(1-e^{-j\omega n})+\\
&\quad \frac{1}{n}\sin^2\frac{\omega n}{2}e^{-j\omega n/2}\operatorname{cosec}^2\left(\frac{\omega}{2}-\frac{\pi r}{n}\right)\\
&= \sin^2\frac{\omega n}{2}e^{-j\omega n}\left\{\frac{1}{n}\operatorname{cosec}^2\left(\frac{\omega}{2}-\frac{\pi r}{n}\right)+2j\cot\left(\frac{\omega}{2}-\frac{\pi r}{n}\right)\right\}.
\end{aligned}
$$

Similarly

$$
\begin{aligned}
B_r(\omega,n) &= j\sin\frac{\omega n}{2}e^{-j\omega n/2}-\frac{j}{2}(e^{-j\omega n}-1)\left\{j\cos\frac{\omega n}{2}-j\sin\frac{\omega n}{2}\right\}e^{-j\omega n/2}\\
&= j\sin\frac{\omega n}{2}e^{-j\omega n/2}(1-e^{-j\omega n})\\
&= -2e^{-j\omega n}\sin^2\frac{\omega n}{2}.
\end{aligned}
$$

Therefore, it is possible to write $M_r(\omega,n)$ more compactly as

$$
M_r(\omega,n)=e^{-j\omega n}\sin^2\frac{\omega n}{2}\left\{\frac{1}{n}\operatorname{cosec}^2\left(\frac{\omega}{2}-\frac{\pi r}{n}\right)+2j\cot\left(\frac{\omega}{2}-\frac{\pi r}{n}\right)-2\right\}. \tag{22}
$$

Now, by assumption $f_r = G(e^{-jr\omega_s})+\nu_r$ and so by the linearity of (21) it is possible to write the Birkhoff interpolant estimate $\widehat{G}_n(e^{-j\omega})$ as

$$
\widehat{G}_n(e^{-j\omega})=\frac{1}{n}\sum_{r=0}^{n-1}G(e^{-jr\omega_s})M_r(\omega,n)+\frac{1}{n}\sum_{r=0}^{n-1}\nu_r M_r(\omega,n). \tag{23}
$$

Therefore,

$$
\left|G(e^{-j\omega})-\widehat{G}_n(e^{-j\omega})\right|\le\left|G(e^{-j\omega})-\frac{1}{n}\sum_{r=0}^{n-1}G(e^{-jr\omega_s})M_r(\omega,n)\right|+\frac{1}{n}\left|\sum_{r=0}^{n-1}\nu_r M_r(\omega,n)\right|.
$$

Using (22) and the assumption that $|\nu_r| \leq \varepsilon$ then provides

$$
\begin{aligned}
\frac{1}{n}\left|\sum_{r=0}^{n-1}\nu_r M_r(\omega,n)\right| &= \frac{1}{n}\sin^2\frac{\omega n}{2}\left|\frac{1}{n}\sum_{r=0}^{n-1}\nu_r \operatorname{cosec}^2\left(\frac{\omega}{2}-\frac{\pi r}{n}\right)+\right. \\
&\qquad \left. 2j\sum_{r=0}^{n-1}\nu_r \cot\left(\frac{\omega}{2}-\frac{\pi r}{n}\right) - 2\sum_{r=0}^{n-1}\nu_r\right| \\
&\leq \frac{\varepsilon}{n^2}\sin^2\frac{\omega n}{2}\sum_{r=0}^{n-1}\operatorname{cosec}^2\left(\frac{\omega}{2}-\frac{\pi r}{n}\right)+ \\
&\qquad 2\frac{\varepsilon}{n}\sin^2\frac{\omega n}{2}\sum_{r=0}^{n-1}\left|\cot\left(\frac{\omega}{2}-\frac{\pi r}{n}\right)\right| + 2\varepsilon.
\end{aligned}
$$

However, by equation 14.1.1 on page 259 of (Hanson 1975)

$$
\sum_{r=0}^{n-1}\operatorname{cosec}^2\left(\frac{\omega}{2}-\frac{\pi r}{n}\right) = n^2\operatorname{cosec}^2\frac{\omega n}{2}.
$$

Using this and Lemma A.2 for the cotangent term then provides

$$
\frac{1}{n}\left|\sum_{r=0}^{n-1}\nu_r M_r(\omega,n)\right| \leq 2\varepsilon\left(3+\frac{2}{\pi}\log n\right). \tag{24}
$$

This takes care of the ‘noise term’. For the ‘undermodelling term’, then returning to the definition (18)

$$
\begin{aligned}
V_n G &= \frac{1}{n}\sum_{k=0}^{n-1}\left(\sum_{r=0}^{n-1}G(e^{-jr\omega_s})e^{j\omega_s rk}\right)e^{-j\omega k} - \\
&\qquad \frac{(e^{-j\omega n}-1)}{n^2}\sum_{k=0}^{n-1}\left(\sum_{r=0}^{n-1}G(e^{-jr\omega_s})e^{j\omega_s rk}\right)ke^{-j\omega k}.
\end{aligned}
$$

But as noted in (A.5)-(A.7)

$$
\sum_{k=0}^{n-1}\left(\sum_{r=0}^{n-1}G(e^{-jr\omega_s})e^{j\omega_s rk}\right) = n\sum_{m=0}^{\infty}g_{k+mn}
$$

where $\{g_k\}$ is the impulse response of G. Therefore, following the same method as in the proof of Lemma A.1

$$
V_n G = \sum_{k=0}^{n-1}\sum_{m=0}^{\infty}g_{k+mn}e^{-j\omega k} - \frac{(e^{-j\omega n}-1)}{n}\sum_{k=0}^{n-1}\sum_{m=0}^{\infty}kg_{k+mn}e^{-j\omega k}
$$

so

$$
\begin{aligned}
\left|G(e^{-j\omega}) - V_n G(e^{-j\omega})\right| &\leq \left|\sum_{k=0}^{\infty} g_k e^{-j\omega k} - \sum_{k=0}^{n-1}\sum_{m=0}^{\infty} g_{k+mn} e^{-j\omega k}\right| + \\
&\quad \frac{2}{n}\left|\sin\frac{\omega n}{2}\right| \sum_{k=0}^{n-1}\sum_{m=0}^{\infty} k|g_{k+mn}| \\
&\leq \frac{2M\rho}{\rho^n(\rho-1)} + \frac{2M}{n}\sum_{k=0}^{n-1}\frac{k}{\rho^k}\sum_{m=0}^{\infty}\frac{1}{\rho^{mn}} \\
&= \frac{2M\rho}{\rho^n(\rho-1)} + \frac{2M}{n(\rho-1)^2}\left\{\rho + \frac{n(1-\rho)-1}{\rho^{n-1}}\right\}\frac{\rho^n}{(\rho^n-1)} \quad (25)\\
&= \frac{2M\rho}{(\rho-1)}\left\{\frac{1}{\rho^n} + \frac{1}{n(\rho-1)} - \frac{1}{(\rho^n-1)}\right\} \quad (26)
\end{aligned}
$$

where equation (25) was obtained by differentiating the formula for the partial sum of a geometric series. Combining the bound (26) with (24) and the triangle inequality via (23) then completes the proof. □

6 Influence of Model Order

The paper up to this point has examined only interpolatory algorithms ($d(n) \geq n$). In the remainder of the work, this strategy will be replaced by one that takes advantage of the flexibility of the formulation (4),(6) which allows for a model order of $d(n) < n$. In making this choice, it is clearly of interest to examine its manifestation on the estimation error and to date, there appears to be no work directly addressing this issue.

Firstly, it is possible to establish that the estimation error component $V_n\nu$ due to the measurement corruption $\{\nu_k\}$ grows at least as fast as $\log d(n)$.

Theorem 6.1. *For the problem setup described in (3),(2) and the estimator V_n formed via (4),(6)*

$$\sup_{\|\nu\|_\infty \leq \varepsilon} \|V_n\nu\|_\infty \geq \frac{4\varepsilon}{\pi^2}\log[d(n)+1]. \quad (27)$$

Proof. Consider the noise induced error term given by (4) and (6) as:

$$
\begin{aligned}
(V_n\nu)(e^{-j\omega}) &= \frac{1}{n}\sum_{r=0}^{n-1}\nu_r\sum_{m=0}^{d(n)-1} e^{-j(\omega - r\omega_s)m} \\
&= \frac{1}{n}\sum_{r=0}^{n-1}\nu_r\sum_{m=0}^{d(n)-1}\cos(\omega - r\omega_s)m - j\sin(\omega - r\omega_s)m \\
&= \frac{1}{n}\sum_{r=0}^{n-1}\left(\frac{\sin d(n)\left(\frac{\omega}{2} - \frac{r\pi}{n}\right)}{\sin\left(\frac{\omega}{2} - \frac{r\pi}{n}\right)}\right)|\nu_r| e^{j\varphi_r} e^{-j(d(n)-1)(\omega/2 - r\pi/n)}
\end{aligned}
$$

Here φ_r is the phase of the frequency measurement error at the rth measurement. The worst case measurement error sequence is clearly one where at some worst case frequency ω, $|\nu_r| = \varepsilon$ and $\varphi_r - (d(n)-1)(\omega/2 - r\pi/n) = 0$ or $\pi \pmod{2\pi}$ so that

$$\chi(\omega) \triangleq \sup_{\|\nu\|_\infty \leq \varepsilon} \left|(V_n\nu)(e^{-j\omega})\right| = \frac{\varepsilon}{n}\sum_{r=0}^{n-1}\left|\frac{\sin d(n)(\omega/2 - r\pi/n)}{\sin(\omega/2 - r\pi/n)}\right| \tag{28}$$

$$= \frac{\varepsilon}{n}\sum_{r=0}^{n-1}\left|D_k\left(\omega - \frac{2\pi r}{n}\right)\right| \tag{29}$$

where $d(n) = 2k(n)+1$ and $D_k(\theta) = \frac{\sin(2k+1)\theta/2}{\sin\theta/2}$ is the kth Dirichlet kernel (Körner 1988). In this case

$$\begin{aligned}\sup_{\|\nu\|_\infty \leq \varepsilon} \|V_n\nu\|_1 &= \frac{\varepsilon}{2\pi n}\sum_{r=0}^{n-1}\int_{-\pi}^{\pi}\left|D_k\left(\omega - \frac{2\pi r}{n}\right)\right| \mathrm{d}\omega \\ &\geq \frac{4\varepsilon}{\pi^2}\log 2(k(n)+1).\end{aligned}$$

This last inequality is found on page 69 of (Körner 1988). Finally, the fact that $\|V_n\nu\|_1 \leq \|V_n\nu\|_\infty$ completes the proof. Note, in passing, that it is possible to exactly calculate the worst case L_1 norm for the noise error by using the results of Theorem 2.4 of (Geddes and Mason 1975).

$$\sup_{\|\nu\|_\infty \leq \varepsilon} \|V_n\nu\|_1 = 2\pi\varepsilon\left(\frac{1}{d(n)} + \sum_{r=1}^{k(n)}\frac{1}{2r}\tan\frac{\pi r}{d}\right)$$

where $d(n) = 2k(n)+1$. □

Note that this result may also be obtained as a special case of Theorem 2.1 of (Partington 1998). However, by exploiting the special properties of the frequency domain case of interest here, the proof of Theorem 6.1 is much more direct than that of Theorem 2.1 of (Partington 1998) where a more general setting is considered.

As well as this lower bound that grows like $\log d(n)$ it is also possible to establish that the worst case noise induced error asymptotically grows no faster than $\log d(n)$.

Theorem 6.2. *For the problem setup described in (3),(2) and the estimate formed via (4) and (6) then for any sequence of positive numbers $\{\alpha_n\}$ such that*

$$\lim_{n\to\infty} \alpha_n^{-1}\log d(n) = 0 \tag{30}$$

it holds that

$$\lim_{n\to\infty} \sup_{\substack{G\in H_\infty(\mathbf{D}_\rho, M)\\ \|\nu\|_\infty \leq \varepsilon}} \frac{1}{\alpha_n}\left\|G(z) - \widehat{G}_n(z)\right\|_\infty = 0.$$

Proof. Consider the worst case noise induced error term given by (29) as

$$\chi(\omega) \triangleq \sup_{\|\nu\|_\infty \le \varepsilon} \left|(V_n\nu)(e^{-j\omega})\right| = \frac{\varepsilon}{n}\sum_{r=0}^{n-1}\left|D_k\left(\omega - \frac{2\pi r}{n}\right)\right|$$

where $d(n) = 2k(n)+1$ and $D_k(\theta)$ is the Dirichlet kernel (Körner 1988). Let μ be Lebesgue measure on $\mathbf{T}$. Then for any $\beta > 0$;

$$\begin{aligned} \int_{\mathbf{T}} \chi(\omega)\,\mathrm{d}\mu(\omega) &\ge \int_{\omega:\chi\ge\beta\alpha_n} \chi(\omega)\,\mathrm{d}\mu(\omega) \qquad ;\beta\in\mathbf{R}^+ \\ &\ge \beta\alpha_n \int_{\omega:\chi\ge\beta\alpha_n} \mathrm{d}\mu(\omega) \\ &= \beta\alpha_n\mu\left\{\omega\in[-\pi,\pi] : \chi(\omega)\ge\beta\alpha_n\right\}. \end{aligned}$$

However

$$\int_{\mathbf{T}} \chi(\omega)\,\mathrm{d}\mu(\omega) = \frac{\varepsilon}{n}\sum_{r=0}^{n-1}\int_{\mathbf{T}}\left|D_k\left(\omega-\frac{2\pi r}{n}\right)\right|\,\mathrm{d}\mu(\omega) \le \varepsilon c_1 \log d(n)$$

for some $c_1 \le \infty$ by (Körner 1988). Therefore, by the assumption (30) on $\{\alpha_n\}$, by the linearity of the estimation scheme, and by Lemma A.1 there exists a $c_2 < \infty$ such that

$$\begin{aligned} \mu\left\{\omega\in[-\pi,\pi] : \frac{1}{\alpha_n}\left|G(e^{-j\omega}) - \widehat{G}_n(e^{-j\omega})\right| \ge \beta\right\} &\le \frac{1}{\beta\alpha_n}\int_{\mathbf{T}} |G(e^{j\omega}) - V_nG(e^{j\omega})|\mathrm{d}\mu(\omega) + \\ &\quad \frac{1}{\beta\alpha_n}\int_{\mathbf{T}} |(V_n\nu)(e^{j\omega})|\mathrm{d}\mu(\omega) \\ &\le \frac{2\pi}{\beta\alpha_n}\|G - V_nG\|_\infty + \frac{1}{\beta\alpha_n}\int_{\mathbf{T}}\chi(\omega)\mathrm{d}\mu(\omega) \\ &\le \frac{c_2}{\rho^{d(n)}\beta\alpha_n} + \frac{c_1\varepsilon\log d(n)}{\beta\alpha_n} \\ &\to 0 \quad \text{as } n\to\infty. \end{aligned}$$

□

So the $\log d(n)$ divergence rate due to noise seems tight, but what is its exact formulation for finite n? The answer for the specific case of $d(n) = n$ has eluded mathematicians for many years, although as already mentioned, precise asymptotic results are known for the case of $d(n) = n$; see Lemma 4.1. Any hope of exactly quantifying it for the case of general $d(n)$ would therefore seem to hopelessly intractable. However, note that the expression that requires clarification is obtained from (28) as

$$\sup_{\|\nu\|_\infty\le\varepsilon}\|V_n\nu\|_\infty = \sup_{\omega\in[-\pi,\pi]}\frac{\varepsilon}{n}\sum_{r=0}^{n-1}\left|D_k\left(\omega-\frac{2\pi r}{n}\right)\right| \tag{31}$$

where $D_k(\theta)$ is the kth Dirichlet kernel (Körner 1988) and $d(n) = 2k(n) + 1$. Since it is known from Theorems 6.1 and 6.2 that (31) grows like $\log d(n)$ is makes sense to fit a $\log d(n)$ dependent function to (31) in order to generate an empirical answer to finding finite bounds. The result arrived at via computer simulation is

$$\sup_{\|\nu\|_\infty \leq \varepsilon} \|V_n\nu\|_\infty \approx \varepsilon \begin{cases} \dfrac{2}{\pi}\left[\gamma + \log\dfrac{8}{\pi} + \dfrac{2}{\pi}\log d(n)\right] & ; d(n) \leq \dfrac{n}{2} \\ \dfrac{2}{\pi}\left[\gamma + \log\dfrac{8}{\pi} + \log\dfrac{n}{2} + \left(\dfrac{2}{\pi} - 1\right)\log(n - d(n))\right] & ; \dfrac{n}{2} < d(n) \leq n - 1 \end{cases} \tag{32}$$

where $\gamma \approx 0.57722$, as in Lemma 4.1, is Euler's constant. This empirically derived formula (32) is compared to the right hand side of (31) for $n = 50, 100, 200$ in figure 1. The agreement

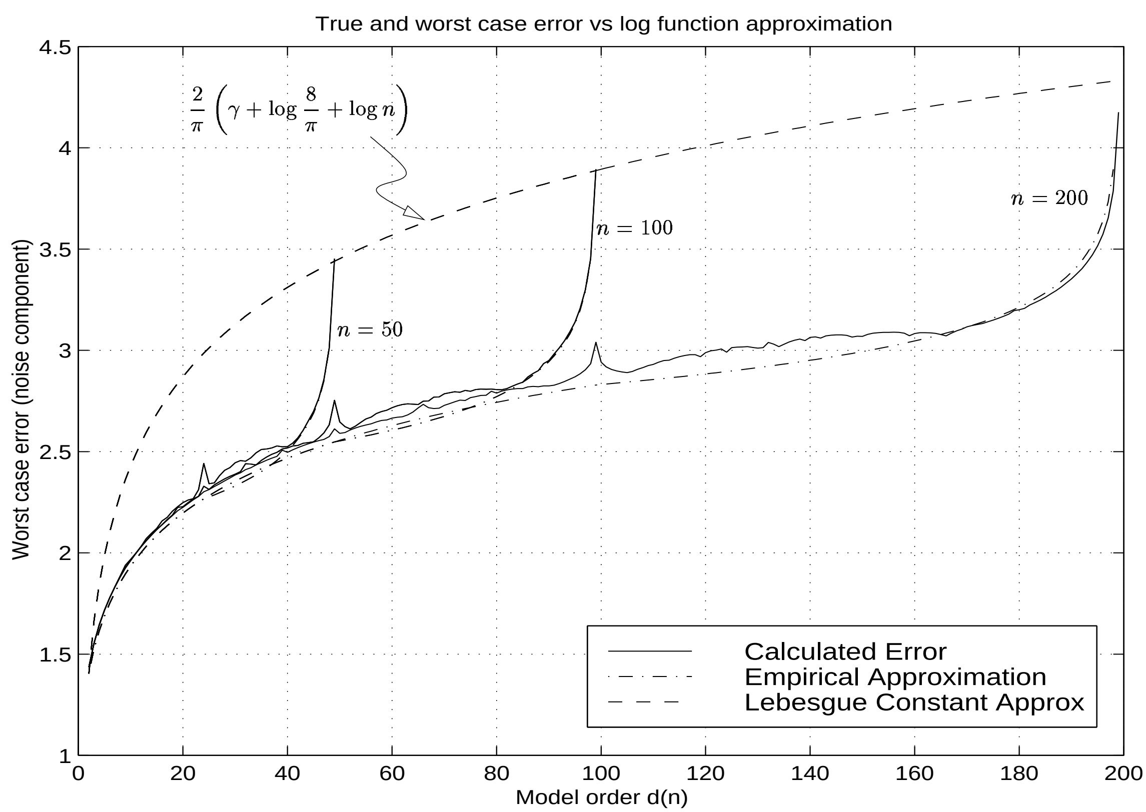

Figure 1: *Performance of the empirically derived noise bound for $n = 50, 100, 200$. Dash-dot line is the empirically derived approximation, solid line is the numerically calculated true value, dashed line is Lebesgue constant approximation from Lemma 4.1 that applies only for $d(n) = n$.*

is close. Note in particular that as $d(n) \to n$ the expression (32) matches the quantity derived theoretically for $d(n) = n$ in Lemma 4.1 (the slight gap for the $n = 200$ case is due to errors involved with finite precision calculations).

Using Lemma A.1 then provides the empirical bound for the worst case error:

$$\sup_{\substack{G\in H_\infty(\mathbf{D}_\rho,M)\\ \|\nu\|_\infty\leq\varepsilon}} \left\|G(z)-\widehat{G}_n(z)\right\|_\infty \leq \frac{M\rho}{\rho^{d(n)}(\rho-1)}\left[1+\frac{\rho^{d(n)}-1}{\rho^n-1}\right]+ \varepsilon\begin{cases} 0.96+0.42\log d(n) & ; d(n)\leq \dfrac{n}{2} \\ 0.96+0.62\log\dfrac{n}{2}-0.2\log(n-d(n)) & ; \dfrac{n}{2}<d(n)\leq n-1 \end{cases} \tag{33}$$

Note that for large enough n this provides a tighter bound than provided by Theorem 3.4 of (Gu and Khargonekar July 1992*b*). However, balancing this, the bound here is obtained empirically as opposed to the theoretical derivation in (Gu and Khargonekar July 1992*b*).

7 Model Order Selection

The final contribution of this paper is to use the results of the previous section in order to examine the issue of model order selection. Specifically, the tightly quantified noise error bound in (33) highlights that for given noise and given data length n there is an model order $d(n)$ which provides that smallest overall error.

Stochastic estimation has a long history of also addressing this issue by recognising that there is a so-called 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 (Ljung 1987).

The purpose of this section is to suggest that an analogous idea be employed when using H_∞ estimation methods. Namely, one should choose the model order giving the smallest worst case bound that balances the undermodelling and measurement corruption components of the error bound in an optimal way so as to minimise the total bound.

For the bound given by (32), since explicit error bounds (and not just rate bounds) are given for both the noise and undermodelling induced errors, the model order choice $d(n)$ providing the optimum tradeoff maybe analytically calculated as follows.

Lemma 7.1. *An optimal model order $d(n)$ exists in the sense of minimising the worst case bound in (32) if when $d(n)$ is chosen as the integer nearest the solution of*

$$\frac{\rho^{d(n)}}{d(n)}=\frac{2.381}{\varepsilon}\frac{M\rho}{(\rho-1)}\left(\frac{\rho^n-2}{\rho^n-1}\right)\log\rho \quad ; d(n)\leq\frac{n}{2} \tag{34}$$

or

$$\frac{\rho^{d(n)}}{(n-d(n))}=\frac{5}{\varepsilon}\frac{M\rho}{(\rho-1)}\left(\frac{\rho^n-2}{\rho^n-1}\right)\log\rho \quad ; d(n)>\frac{n}{2} \tag{35}$$

then that solution satisfies

$$d(n)>\frac{1}{\log\rho}. \tag{36}$$

Proof. The bound to be minimised is

$$B_n(d) \triangleq \frac{M\rho}{\rho^{d(n)}(\rho-1)}\left[1+\frac{\rho^{d(n)}-1}{\rho^n-1}\right]+\varepsilon\begin{cases} 0.96+0.42\log d(n) & ; d(n)\le \dfrac{n}{2} \\ 0.96+0.62\log\dfrac{n}{2}-0.2\log(n-d(n)) & ; d(n) > \dfrac{n}{2}\end{cases}$$

therefore

$$\frac{\mathrm{d}B_n(d)}{\mathrm{d}d} = -\frac{M\rho}{\rho^{d(n)}(\rho-1)}\left(\frac{\rho^n-2}{\rho^n-1}\right)\log\rho+\varepsilon\begin{cases} \dfrac{0.42}{d(n)} & ; d(n)\le \dfrac{n}{2} \\ \dfrac{0.2}{n-d(n)} & ; d(n) > \dfrac{n}{2}\end{cases}$$

$$\frac{\mathrm{d}^2B_n(d)}{\mathrm{d}d^2} = \frac{M\rho}{\rho^{d(n)}(\rho-1)}\left(\frac{\rho^n-2}{\rho^n-1}\right)\log^2\rho+\varepsilon\begin{cases} -\dfrac{0.42}{d^2(n)} & ; d(n)\le \dfrac{n}{2} \\ \dfrac{0.2}{(n-d(n))^2} & ; d(n) > \dfrac{n}{2}\end{cases}$$

Choosing $d(n)$ to make the first derivative zero and the second derivative positive then ensures that $B_n(d)$ is minimised. □

Clearly, by establishing the existence of an optimal model order only for linear algorithms, there is still great scope for examining this issue with the effective more general classes of nonlinear algorithms.

8 Conclusions

The purpose of this paper was to conduct a detailed study of the worst case divergence behaviour of linear algorithms for ‘estimation in H_∞’. A central theme of the paper was to draw on relevant work in the field of polynomial approximation theory where parallel issues in the field of real valued Lagrange interpolation have been addressed since the start of this century. In relation to this, it was shown by constructive analysis that the divergence of linear estimation in H_∞ schemes is critically dependent on the p-norm used to measure divergence being strictly $p=\infty$ (and no value of $p<\infty$) and on the measurement corrupting component $\{\nu_k\}$ having no smoothness constraint placed on it that allows their rate of variation between samples to be limited either in phase or in magnitude. The final topic of the paper was to conduct a study as to how the worst case estimation error is affected by the choice of model order. This consideration allowed the issue of model order selection to be addressed in a deterministic worst case setting.

A Auxiliary Results

Lemma A.1. *Suppose $\{f_k\}$ is given by $f_k = G(e^{-j\omega_s k})$ where $\omega_s = 2\pi/n$ and that the operator V_n introduced in (10), (11) is defined via:*

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

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

$$\|G(z) - V_n G(z)\|_\infty \leq \frac{M\rho}{\rho^{d(n)}(\rho-1)}\left[1 + \frac{\rho^{d(n)}-1}{\rho^n - 1}\right]. \tag{A.2}$$

Proof. Since $G(z) \in H_\infty(\mathbf{D}_\rho, M)$ it is possible to write $G(z)$ as a Taylor series valid on $z \in \mathbf{D}_\rho$:

$$G(z) = \sum_{\ell=0}^{\infty} g_\ell z^\ell \tag{A.3}$$

and by Cauchy's estimate in Lemma A.3

$$|g_k| \leq M\rho^{-\ell}. \tag{A.4}$$

Substituting (A.3) into (A.1) provides

$$\begin{aligned} \widehat{f}_k &= \frac{1}{n}\sum_{r=0}^{n-1}\left(\sum_{\ell=0}^{\infty} g_\ell e^{-j\omega_s \ell r}\right) e^{j\omega_s kr} && \text{(A.5)} \\ &= \frac{1}{n}\sum_{\ell=0}^{\infty} g_\ell \sum_{r=0}^{n-1} e^{j2\pi r(k-\ell)/n}. && \text{(A.6)} \end{aligned}$$

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

$$\widehat{f}_k = \sum_{m=0}^{\infty} g_{k+mn}. \tag{A.7}$$

This leads to

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

□

Lemma A.2.

$$\frac{1}{n}\sin^2\frac{\omega n}{2} \sum_{r=0}^{n-1} \left| \cot\left(\frac{\omega}{2} - \frac{\pi r}{n} \right) \right| \le 1.5 + \frac{2}{\pi}\log n \tag{A.8}$$

$$\frac{1}{n}\left|\sin\frac{\omega n}{2}\right| \sum_{r=0}^{n-1} \left| \cot\left(\frac{\omega}{2} - \frac{\pi r}{n} \right) \right| \le 2.0 + \frac{2}{\pi}\log n. \tag{A.9}$$

Proof. Define m and Δ by taking m as the integer such that for some $\Delta \in [0, \pi/n)$ it is possible to write

$$\frac{\omega}{2} = \frac{m\pi}{n} + \Delta. \tag{A.10}$$

In this case, putting $\ell = m - r$, noting that $|\cot(\Delta + (m-r)\pi/n)| = |\cot(\Delta + (m+n-r)\pi/n)|$, and putting $k = m + n - r$ leads to

$$\begin{aligned}
\sum_{r=0}^{n-1} \left| \cot\left(\frac{\omega}{2} - \frac{\pi r}{n} \right) \right| &= \sum_{r=0}^{n-1} \left| \cot\left(\Delta + (m-r)\frac{\pi}{n} \right) \right| \\
&= |\cot\Delta| + |\cot\delta| + \sum_{\ell=1}^{n-2} \left| \cot\left(\Delta + \ell\frac{\pi}{n} \right) \right|.
\end{aligned}$$

However, by the definition of Δ

$$\sin^2\frac{\omega n}{2} = \sin^2\left(\frac{\omega n}{2} - m\pi\right) = \sin^2 n\left(\frac{\omega}{2} - m\frac{\pi}{n}\right) = \sin^2 n\Delta$$

so that using a strategy of differentiating and setting to zero to find the maximum yields

$$\frac{1}{n}\sin^2\frac{\omega n}{2}|\cot\Delta| = \frac{1}{n}\left|\frac{\sin^2 n\Delta\cos\Delta}{\sin\Delta}\right| \leq \frac{3}{4}.$$

Similarly, defining δ by $\delta + \Delta = \pi/n$ yields

$$\sin^2\frac{\omega n}{2} = \sin^2\left(\frac{\omega n}{2} - (m+1)\pi\right) = \sin^2 n\left(\frac{\omega}{2} - (m+1)\frac{\pi}{n}\right) = \sin^2 n\delta$$

so that by an argument identical to that just employed

$$\frac{1}{n}\sin^2\frac{\omega n}{2}|\cot\delta| \leq \frac{3}{4}.$$

Finally, exploiting the monotonicity properties of $|\cot\theta|$ provides

$$\begin{aligned}
\frac{\pi}{n}\sum_{\ell=1}^{n-2}\left|\cot\left(\Delta + \ell\frac{\pi}{n}\right)\right| &\leq \int_{\Delta}^{\pi-\delta}|\cot\theta|\,\mathrm{d}\theta \\
&= \int_{\Delta}^{\pi/2}\cot\theta\,\mathrm{d}\theta - \int_{\pi/2}^{\pi-\delta}\cot\theta\,\mathrm{d}\theta \\
&= \log\left(\frac{1}{\sin\Delta\sin\delta}\right)
\end{aligned}$$

so that

$$\begin{aligned}
\left(\frac{1}{n}\sin^2\frac{\omega n}{2}\right)\sum_{\ell=1}^{n-1}\left|\cot\left(\Delta + \ell\frac{\pi}{n}\right)\right| &\leq \frac{1}{\pi}\sin^2 n\Delta\log\left(\frac{1}{\sin\Delta\sin(\pi/n - \Delta)}\right) \\
&\leq \frac{1}{\pi}\sin^2\frac{\pi}{2}\left|\log\sin^2\frac{\pi}{2n}\right| \\
&\leq \frac{2}{\pi}\log n
\end{aligned}$$

where the second last line follows from the third last line by setting the derivative with respect to Δ equal to zero. Hence (A.8) is established, and (A.9) follows using an identical line of argument save that the bound

$$\frac{1}{n}\left|\sin\frac{\omega n}{2}\cot\Delta\right| = \frac{1}{n}\left|\frac{\sin n\Delta\cos\Delta}{\sin\Delta}\right| \leq 1$$

(which follows since $n^{-1}|\sin n\Delta/\sin\Delta| \leq 1$) is used. □

Lemma A.3. *Cauchy's Estimates Set $\mathbf{D}(a,R) = \{z \in \mathbf{C} : |z-a| < R\}$ and suppose f is analytic in $\overline{\mathbf{D}(a,R)}$ and $\exists M < \infty$ such that $|f(z)| \leq M$ on $\mathbf{D}(a,R)$. Then*

$$\left|\left.\frac{\partial^n f(z)}{\partial z^n}\right|_{z=a}\right| \leq \frac{n!M}{R^n} \tag{A.11}$$

Proof. See (Rudin 1966). □

References

Brutman, L. (1980). 'On the polynomial and rational projections in the complex plane'. *SIAM Journal of Numerical Analysis* **17**(3), 366–372.

Brutman, L. and A. Pinkus (1980). 'On the Erdös conjecture concerning minimal norm interpolation on the unit circle'. *SIAM Journal on Numerical Analysis* **17**(3), 373–375.

Cheney, E. (1966). *Introduction to Approximation Theory*. McGraw Hill. New York.

Chui, C. K., Xie-Chang Shen and Lefan Zhong (1993). 'On Lagrange interpolation at disturbed roots of unity'. *Transactions of the American Mathematical Society* **336**(2), 817–829.

Davis, P. (1963). *Interpolation and Approximation*. Blaisdell Publishing Company.

De Boor, C. and A. Pinkus (1978). 'Proof of the conjectures of Bernstein and Erdös concerning the optimal nodes for polynomial interplation'. *Journal of Approximation Theory* **24**, 289–303.

Erdös, P. (1961*a*). 'An extremal problem in the theory of interpolation'. *Acta Mathematica Hungarica* **12**, 222–234.

Erdös, P. (1961*b*). 'Problems and results on the theory of interpolation II'. *Acta Mathematica Hungarica* **12**, 235–244.

Erdös, P. (1968). 'Problems and results on the convergence and divergence properties of the Lagrange interpolations polynomials and some extremal problems'. *Mathematica(Cluj.)* **10**, 65–73.

Geddes, K. and J.C. Mason (1975). 'Polynomial approximations by projections on the unit circle'. *SIAM Journal of Numerical Analysis* **12**(1), 111–120.

Giarré, L., Mario Milanese and M. Taragna (1997). 'H_∞ identification and model quality evaluation'. *IEEE Transactions on Automatic Control* **42**(2), 188–199.

Gu, G. and P. Khargonekar (1992*a*). 'A class of algorithms for identification in H_∞'. *Automatica* **28**(2), 299–312.

Gu, G. and P. Khargonekar (July 1992*b*). 'Linear and nonlinear algorithms for identification in H_∞ with error bounds'. *IEEE Transactions on Automatic Control* **37**(7), 953–963.

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

Günttner, R. (1980). 'Evaluation of lebesgue constants'. *Siam Journal of Numerical Analysis* **17**(4), 512–520.

Hakvoort, R. and P.M.J. Van den Hof (1994). 'Frequency domain curve fitting with maximum amplitude criterion and guaranteed stability'. *International Journal of Control* **60**(5), 809–825.

Hanson, E. (1975). *A Table of Series and Products*. Prentice Hall Inc.. Englewood Cliffs N.J.

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

Helmicki, A., C.A. Jacobson and C.N. Nett (1990*a*). 'Identification in H_∞: A robustly convergent, nonlinear algorithm'. *Proceedings of the American Control Conference* pp. 386–408.

Helmicki, A., C.A. Jacobson and C.N. Nett (1990*b*). 'Identification in H_∞:Linear algorithms'. *Proceedings of the American Control Conference* pp. 2418–2423.

Helmicki, A., C.A. Jacobson and C.N. Nett (1990*c*). 'Identification in H_∞:The continuous time case'. *Proceedings of the American Control Conference* pp. 1893–1898.

Helmicki, A., C.A. Jacobson and C.N. Nett (1991*b*). 'Fundamentals of control oriented system identification and their application for identification in H_∞'. *Proceedings of the American Control Conference* pp. 89–99.

Helmicki, A., C.A. Jacobson and C.N. Nett (October 1991*a*). 'Control oriented system identification:A worst case/deterministic approach in H_∞'. *IEEE Transactions on Automatic Control* **36**(10), 1163–1176.

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

Körner, T. (1988). *Fourier Analysis*. Cambridge University Press.

Ljung, L. (1987). *System Identification: Theory for the User*. Prentice-Hall, Inc.. New Jersey.

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

Mäkilä, P., J.R. Partington and T.K. Gustafsson (1995). 'Worst-case control-relevant identification'. *Automatica* **31**, 1799–1819.

Ninness, B. (1998). 'A stochastic approach to linear estimation in H_∞'. *Automatica* **34**(1), 405–414.

Oppenheim, A. and R.W. Schafer (1989). *Discrete Time Signal Processing*. Prentice Hall. New Jersey.

Parker, P. (1988). Frequency Domain Descriptions of Linear Systems. PhD thesis. Australian National University.

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

Partington, J. (1991*a*). 'Robust identification and interpolation in H_∞'. *International Journal of Control* **54**, 1281–1290.

Partington, J. (1991*b*). 'Robust identification in H_∞'. *Journal of Mathematical Analysis and Application* **166**, 428–441.

Partington, J. (1998). 'Worst-case errors of linear algorithms for identification in H_∞'. *International Journal of Control* **69**(2), 347–352.

Partington, J. R. (1997). *Interpolation, Identification, and Sampling*. Oxford University Press.

Rudin, W. (1966). *Real and Complex Analysis*. McGraw Hill. New York.

Turán, P. (1980). 'On some open problems of approximation theory'. *Journal of Approximation Theory* **29**, 23–85.

Varma, A. (1988). 'Complex interpolating polynomials'. *Proceedings of the American Mathematical Society* **103**(1), 125–130.

Vértesi, P. (1990). 'Optimal Lebesgue constant for Lagrange interpolation'. *SIAM Journal on Numerical Analysis* **27**(5), 1322–1331.