

Analyzing Iterations in Identification with Application to Nonparametric \mathcal{H}_∞ -norm Estimation

Tom Oomen* Cristian R. Rojas** Håkan Hjalmarsson**
Bo Wahlberg**

* *Eindhoven University of Technology, Department of Mechanical
Engineering, Eindhoven, The Netherlands (t.a.e.oomen@tue.nl).*

** *Automatic Control Lab and ACCESS Linnaeus Center, Electrical
Engineering, KTH – Royal Institute of Technology, Stockholm, Sweden
({cristian.rojas,hakan.hjalmarsson,bo.wahlberg} @kth.se)*

Abstract: In the last decades, many iterative approaches in the field of system identification for control have been proposed. Many successful implementations have been reported, despite the lack of a solid analysis with respect to the convergence and value of these iterations. The aim of this paper is to present a thorough analysis of a specific iterative algorithm that involves nonparametric \mathcal{H}_∞ -norm estimation. The pursued approach involves a novel frequency domain approach that appropriately deals with additive stochastic disturbances and input normalization. The results of the novel convergence analysis are twofold: i) the presence of additive disturbances introduces a bias in the estimation procedure, and ii) the iterative procedure can be interpreted as experiment design for \mathcal{H}_∞ -norm estimation, revealing the value of iterations and limits of accuracy in terms of the Fisher information matrix. The results are confirmed by means of a simulation example.

1. INTRODUCTION

In the last decades, many iterative approaches have been proposed in the fields of system identification and control design. Examples of such iterative approaches include iterative learning control (ILC) [Bristow et al., 2006], iterative feedback tuning [Hjalmarsson, 2002], and iterative identification and control [Albertos and Sala, 2002]. Although many successful implementations of these approaches have been reported in the literature, the application of these techniques has met mixed outcomes. Indeed, analyses of specific approaches have pointed out several shortcomings. For instance, in the case of iterative identification and control design, the stationary point of the iterative algorithm may not be a local minimum of the objective function as is pointed out in Hjalmarsson et al. [1995]. Furthermore, the iterations in these approaches may be divergent, see, e.g., Albertos and Sala [2002, Sec. 9.3]. Finally, the value of iterations in these approaches has been questioned in, e.g., Böling and Mäkilä [1998].

Recently, an iterative approach for nonparametric \mathcal{H}_∞ -norm estimation has been proposed in Hjalmarsson [2005, Sec. 12.2] and further extended in Wahlberg et al. [2010a]. A relevant application of \mathcal{H}_∞ -norm estimation includes model error modeling, since reliable robust control design methodologies are available that consider model errors as \mathcal{H}_∞ -norm bounded operators. In contrast to most model error modeling techniques, including Hakvoort and Van den Hof [1997], Ljung [1999a], the approach presented in [Hjalmarsson, 2005, Sec. 12.2] does not require the estimation of an intermediate parametric model. Indeed, in Hjalmarsson [2005, Sec. 12.2], the input to the system is iteratively determined, followed by a nonparametric estimation of the \mathcal{H}_∞ -norm from the measured data of two experiments. An essential property of the iterative

procedure is that it is known to converge to the global optimum with an exponential rate of convergence in the noise-free case, since in this case it coincides with a power iteration [Golub and Van Loan, 1996, Section 8.2].

Although several successful applications of iterative nonparametric \mathcal{H}_∞ -norm estimation have already been reported, including Barenthin et al. [2005], Barenthin et al. [2006], convergence of the considered algorithm has not been analyzed in a stochastic framework. Indeed, when performing experiments on any realistic system, measurement errors and unmeasured disturbances inevitably contaminate the observation. A suitable approach to model these measurement errors and unmeasured disturbances is to consider these in a stochastic framework. The aim of the present paper is to thoroughly analyze convergence, bias, accuracy, and the value of iterations of a certain iterative nonparametric norm estimation algorithm that is subject to additive stochastic disturbances.

Related analyses of power iterations that are implemented in an imperfect environment are reported in Krasulina [1970] and Oja and Karhunen [1985], where the involved matrices are considered random. However, the results for the case of random matrices cannot be directly extended to the case of additive stochastic disturbances that is relevant for the considered system identification problem. Similarly, in, e.g., Golub and Van Loan [1996], the effect of round-off errors in power iterations has been discussed. However, such round-off errors do not provide a suitable description for additive disturbances.

The main contribution of the present paper is a thorough stochastic analysis of a certain iterative nonparametric norm estimation procedure. As specific contributions of the paper: i) a novel algorithm is proposed for \mathcal{H}_∞ -norm

estimation based on a single experiment. In contrast, two experiments are required in Wahlberg et al. [2010a]. ii) a novel stochastic analysis of the iterative norm estimation procedure, which is based on signal spectra, is presented. The novelty of the presented analysis is that it addresses both additive stochastic signals that represent measurement errors and process disturbances as well as normalization of the input to deal with input power constraints. iii) it is shown that additive stochastic disturbances introduce a bias in the considered nonparametric norm estimation algorithm. iv) the value of iterations is established through the Fisher information matrix, revealing that the iterative algorithm can be interpreted as an optimal experiment design approach for \mathcal{H}_∞ -norm estimation for both parametric and nonparametric identification methodologies.

The outline of the paper is as follows. In Sec. 2, a nonparametric \mathcal{H}_∞ -norm estimation algorithm is presented, where the estimation of the norm is based on a single experiment. In Sec. 3, the limit spectra are derived and their convergence is analyzed. Then, in Sec. 4, the derived spectra are employed for a bias analysis of the resulting estimator and for a derivation of the information matrix. The derived results are illustrated by means of an example in Sec. 5. In Sec. 6, concluding remarks are presented. Several proofs are omitted due to space limitations. These proofs are available in Oomen et al. [2011].

2. ITERATIVE NONPARAMETRIC \mathcal{H}_∞ -NORM ESTIMATION

Throughout, the asymptotically stable SISO LTI system

$$y_t = G(z)u_t + e_t = \sum_{k=0}^{\infty} g_k u_{t-k} + e_t, \quad t = 1, 2, \dots, \quad (1)$$

is considered, see also Fig. 1, where u_t denotes the quasi stationary input to the system [Ljung, 1999b], y_t is the output, and e_t is white noise of variance $\lambda_e > 0$, representing measurement noise or a disturbance term, with e_t and u_t independent.

One of the key properties of the \mathcal{H}_∞ -norm, i.e., $\|G\|_\infty = \sup_{\omega \in (-\pi, \pi]} |G(e^{j\omega})|$, is that it equals the ℓ_2 -induced norm, since

$$\|G(z)\|_\infty = \sup_{u \in \ell_2} \frac{\|y\|_2}{\|u\|_2}, \quad (2)$$

where $\|x\|_2 := (\sum_{t=1}^{\infty} x_t^2)^{\frac{1}{2}}$. The characterization of the \mathcal{H}_∞ -norm in (2) is useful for several reasons. Firstly, it is an induced norm and thus enables the representation of model uncertainty by \mathcal{H}_∞ -norm bounded operators. Secondly, (2) is at the basis of the nonparametric \mathcal{H}_∞ -norm estimation algorithms that are presented in this paper.

The following algorithm is the main result of this section and enables nonparametric \mathcal{H}_∞ -norm estimation.

Algorithm 1. Apply the following sequence of steps:

- (1) Let $n = 1$ and generate an input sequence $u^{(1)} := [u_1^{(1)} \dots u_N^{(1)}]^T$ such that $\|u^{(1)}\|_2 / \sqrt{N} = 1$.
- (2) Apply $u^{(n)}$ to the system, yielding $y^{(n)} := [y_1^{(n)} \dots y_N^{(n)}]^T$.
- (3) Time reverse the sequence $y^{(n)}$, i.e., determine $\tilde{y}^{(n)} := [y_N^{(n)} \dots y_1^{(n)}]^T$, and generate $u^{(n+1)} = \frac{\tilde{y}^{(n)}}{\mu^{(n)}}$, where the normalization $\mu^{(n)}$ is defined below.
- (4) Let $n \mapsto n + 1$ and go to step (2).

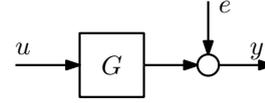


Fig. 1. Considered system.

In Step 3 of Algorithm 1, the normalization

$$\mu^{(n)} := \frac{\|\tilde{y}^{(n)}\|_2}{\sqrt{N}}. \quad (3)$$

is applied. The normalization (3) constrains the input power to unity and is essential in practical applications, e.g., for physical limitations or security reasons. Note that the normalization to unity is nonrestrictive and introduced for notational convenience. Also, the normalization $\mu^{(n)}$ can easily be adapted, e.g., to energy norm constraints or maximum amplitude constraints.

Algorithm 1 generates sequences $\{u^{(n)}\}_{n \in \mathbb{N}}$ and $\{y^{(n)}\}_{n \in \mathbb{N}}$, from which the ℓ_2 -induced gain of G is estimated by

$$\hat{\beta}^{(n)} := \frac{[u^{(n)}]^T \tilde{y}^{(n)}}{\sqrt{N}}. \quad (4)$$

To show convergence of the estimate (4) to the \mathcal{H}_∞ -norm of the underlying system, i.e., $\|G\|_\infty$, observe that in the noise-free case, i.e., $\lambda_e = 0$, $y^{(n)} = \mathbf{G}u^{(n)}$ and $\tilde{y}^{(n)} = \mathcal{T}y^{(n)}$, hence

$$\tilde{y}^{(n)} = \mathbf{G}_H u^{(n)}, \quad (5)$$

where

$$\mathbf{G} := \begin{bmatrix} g_0 & 0 & 0 & \dots & 0 \\ g_1 & g_0 & 0 & \dots & 0 \\ \vdots & & & \ddots & \\ g_{N-1} & g_{N-2} & \dots & g_0 & \end{bmatrix}, \quad \mathcal{T} := \begin{bmatrix} 0 & \dots & 0 & 1 \\ 0 & \dots & 1 & 0 \\ \vdots & \ddots & \vdots & \\ 1 & \dots & 0 & 0 \end{bmatrix},$$

and

$$\mathbf{G}_H := \mathcal{T}\mathbf{G} = \begin{bmatrix} g_{N-1} & g_{N-2} & \dots & g_0 \\ g_{N-2} & g_{N-3} & & 0 \\ \vdots & & \ddots & \\ g_0 & 0 & \dots & 0 \end{bmatrix}. \quad (6)$$

Equation (5) leads to the following result.

Lemma 1. Assume that $\lambda^{\max}(\mathbf{G}_H)$ is unique and that $u^{(1)}$ has a nonzero component in the direction of the eigenvector associated with $\lambda^{\max}(\mathbf{G}_H)$. Then, $\hat{\beta}^{(n)} \rightarrow \lambda^{\max}(\mathbf{G}_H)$ for $n \rightarrow \infty$.

See, e.g., [Golub and Van Loan, 1996, Section 7.3.1] for a proof of Lemma 1. The following theorem establishes that $\hat{\beta}^{(n)}$ converges to $\|G\|_\infty$ for $N \rightarrow \infty$.

Theorem 2. Consider the systems G in (1) and \mathbf{G}_H in (6). Then, $\|G\|_\infty = \lim_{N \rightarrow \infty} |\lambda_{\max}(\mathbf{G}_H)|$.

Lemma 1 and Theorem 2 imply that the estimate (4) can be expected to be a reasonable estimate of $\|G\|_\infty$ in the noise-free case for a sufficiently long duration of each experiment and after a sufficient number of iterations.

The following remarks are appropriate.

Remark 3. In Algorithm 1, a time reversal operator \mathcal{T} is introduced to enable the use of finite time experiments. Specifically, if Algorithm 1 is implemented without the time reversal operator in Step 3, then the procedure estimates the first Markov coefficient g_0 instead of the ℓ_2 gain. Indeed, since \mathbf{G} is a lower triangular Toeplitz matrix, the eigenvalues of \mathbf{G} equal the values on the diagonal.

This phenomenon is introduced solely by the finite sample effect. Indeed, for infinite N , \mathbf{G} is an infinite matrix representing a Toeplitz operator on ℓ_p , whose spectral radius is $\|\mathbf{G}\|_\infty$, see, e.g., [Böttcher and Grudsky, 2005, Corollary 1.12]. Note that the analysis in Section 3 is performed in the frequency domain in terms of power spectra, and leads to the same conclusions regardless of whether the power iterations are applied to $\mathbf{G}^T\mathbf{G}$ or \mathbf{G} .

Remark 4. To point out the relation between the presented approach, i.e., Algorithm 1 in conjunction with the estimator (4), and the ℓ_2 -induced norm characterization of the \mathcal{H}_∞ -norm, i.e., (2), as well as to clarify the relation with the algorithm in Wahlberg et al. [2010a] that requires two experiments for nonparametric \mathcal{H}_∞ -norm estimation, observe that

$$\sup_{u^{(n)} \in \ell_2} \frac{\|y^{(n)}\|_2}{\|u^{(n)}\|_2} = \sup_{u^{(n)} \in \ell_2} \frac{y^{(n)T} y^{(n)}}{u^{(n)T} u^{(n)}} = \sup_{u^{(n)} \in \ell_2} \frac{u^{(n)T} \mathbf{G}^T \mathbf{G} u^{(n)}}{u^{(n)T} u^{(n)}}$$

which is clearly maximal if $u^{(n)}$ is in the eigenvector direction corresponding to $\lambda^{\max}(\mathbf{G}^T\mathbf{G})$. In Wahlberg et al. [2010a], a power iteration is applied to $\mathbf{G}^T\mathbf{G}$ to estimate the maximum gain, which requires two experiments and two time reversal operations. Specifically, due to the Toeplitz structure of \mathbf{G} , $\mathbf{G}^T = \mathcal{T}\mathbf{G}\mathcal{T}$, hence $\mathbf{G}^T\mathbf{G} = \mathcal{T}\mathbf{G}^T\mathbf{G}$. To show that the estimated gain for $n \rightarrow \infty$ is equivalent in the noise-free case, observe that since \mathbf{G}_H is symmetric, it can be factorized as $\mathbf{G}_H = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$, where \mathbf{Q} is orthonormal and $\mathbf{\Lambda}$ is a diagonal matrix containing the eigenvalues of \mathbf{G}_H . As a result, $\mathbf{G}^T\mathbf{G} = \mathbf{Q}\mathbf{\Lambda}^2\mathbf{Q}^T$, hence $\mathbf{G}^T\mathbf{G}$ has eigenvalues $\mathbf{\Lambda}^2$.

Remark 5. The eigenvalues of \mathbf{G}_H , unlike those of $\mathbf{G}^T\mathbf{G}$, see Remark 4, are not guaranteed to be positive. As a result, $\hat{\beta}^{(n)} = \frac{[u^{(n)T} \tilde{y}^{(n)}]}{\sqrt{N}}$ may have a slow transient. Specifically, an oscillatory transient may arise if the first and second largest eigenvalues of \mathbf{G}_H have different sign. For this reason, the estimator in Wahlberg et al. [2010a] can be adapted to deal with the considered normalization in the present paper that is applied after each experiment, see Algorithm 1, in which case

$$\hat{\beta}_2^{(n)} = \sqrt{\mu^{(n-1)} [u^{(n-1)T} \tilde{y}^{(n)}]} \quad (7)$$

Estimator (7) may thus be preferable in the case where it is desired to determine the *absolute* value of the maximum eigenvalue. Specifically, the estimator $\hat{\beta}_2^{(n)}$ corresponds to the square root of $\lambda^{\max}(\mathbf{G}_H^2) = \lambda^{\max}(\mathbf{G}^T\mathbf{G})$ and thus to $\lambda^{\max}(\mathbf{G}_H)$. Since $\mathbf{G}_H^2 = \mathbf{G}^T\mathbf{G}$ is positive semidefinite, $\hat{\beta}_2^{(n)}$ does not suffer from a possible slow oscillatory transient. Note that the analysis in this paper also applies to $\hat{\beta}_2^{(n)}$.

In the preceding analysis of Algorithm 1, it is assumed that $\lambda_e = 0$, i.e., the noise-free situation. In the next section, a stochastic analysis is performed for $\lambda_e > 0$. Then, in Sec. 4, the estimators (4) and (7) are analyzed in detail.

3. CONVERGENCE ANALYSIS

In this section, a convergence analysis of Algorithm 1 is presented in the case where additive stochastic disturbances are present, i.e., $\lambda_e > 0$ in (1). First, expressions for the limit spectra are derived in Section 3.1, followed by a convergence analysis in Section 3.2.

3.1 Limit Spectrum

In this section, Algorithm 1 is analyzed in the presence of noise, i.e., in the case where $\lambda_e > 0$. The first step in the analysis is to assume that $N \rightarrow \infty$, i.e., the number of data samples *at each iteration* tends to infinity. This enables an analysis in the frequency domain in terms of $\Phi_u^{(n)} \in \mathcal{L}_1([- \pi, \pi], \mathbb{R}_0^+)$, i.e., the spectrum of u_t at iteration n , see [Ljung, 1999b, Chapter 2] for an appropriate definition.

It is important to notice that, for a finite N , the effect of the time reversal operation \mathcal{T} , which can be described as a combination of a time shift (by N samples) plus a time inversion, $t \mapsto -t$, has no effect on the spectrum of a quasi-stationary signal. This means that the time reversal operation can be omitted in a frequency domain analysis of the power iterations method, see also Remark 3.

Lemma 6. Consider Algorithm 1 applied to the system \mathbf{G} in (1), where $\lambda_e > 0$. Then, for $N \rightarrow \infty$,

$$\Phi_u^{(n+1)}(\omega) = \frac{|G(e^{j\omega})|^2 \Phi_u^{(n)}(\omega) + \lambda_e}{\frac{1}{2\pi} \int_{-\pi}^{\pi} |G(e^{j\omega})|^2 \Phi_u^{(n)}(\omega) d\omega + \lambda_e} \quad (8)$$

Proof. Combining (1) and Algorithm 1 yields

$$u^{(n+1)} = \frac{1}{\mu^{(k)}} (Gu^{(n)} + e^{(n)})$$

where $\mu^{(k)}$ is defined in (3). Independence of $u^{(n)}$ and $e^{(n)}$ implies that

$$\Phi_u^{(n+1)}(\omega) = \frac{1}{(\mu^{(k)})^2} (|G(e^{j\omega})|^2 \Phi_u^{(n)}(\omega) + \lambda_e)$$

Finally, applying Parseval's relation to $\mu^{(k)}$ yields (8). \square

Next, fixed points of the function (8) are analyzed.

Theorem 7. The function (8) has a unique fixed point $\Phi_u^{(\infty)}$, which is given by

$$\Phi_u^{(\infty)}(\omega) = \frac{\lambda_e}{\mu^2 - |G(e^{j\omega})|^2} \quad (9)$$

where $\mu > 0$ satisfies

$$\frac{1}{\lambda_e} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{\mu^2 - |G(e^{j\omega})|^2} d\omega \quad (10)$$

Proof. Every such fixed point $\Phi_u^{(\infty)} \in \mathcal{L}_1([- \pi, \pi], \mathbb{R}_0^+)$ satisfies the equation

$$\Phi_u^{(\infty)}(\omega) = \frac{|G(e^{j\omega})|^2 \Phi_u^{(\infty)}(\omega) + \lambda_e}{\frac{1}{2\pi} \int_{-\pi}^{\pi} |G(e^{j\omega})|^2 \Phi_u^{(\infty)}(\omega) d\omega + \lambda_e} \quad (11)$$

Denoting

$$\mu^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |G(e^{j\omega})|^2 \Phi_u^{(\infty)}(\omega) d\omega + \lambda_e \quad (12)$$

and solving (11) for $\Phi_u^{(\infty)}$ yields (9). In addition, the resulting $\Phi_u^{(\infty)}$ is unique. Next, to show (10), note that substitution of (9) into (12) yields

$$\begin{aligned} \mu^2 &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\lambda_e |G(e^{j\omega})|^2 + \lambda_e (\mu^2 - |G(e^{j\omega})|^2)}{\mu^2 - |G(e^{j\omega})|^2} d\omega \\ &= \mu^2 \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\lambda_e}{\mu^2 - |G(e^{j\omega})|^2} d\omega \end{aligned}$$

Next, by nonnegativity of $\Phi_u^{(\infty)}(\omega)$, the latter equation directly implies the desired result (10). \square

The result of Theorem 7 enables the derivation of the following properties of $\Phi_u^{(\infty)}$ and μ as a function of λ_e .

Theorem 8. Consider the iteration (8), where the fixed points satisfy the results in Theorem 7. Then,

- (1) $\mu \geq \|G\|_\infty$.
- (2) $\Phi_u^{(\infty)}(\omega)$ attains its (finite) maximum at the frequencies where $|G(e^{j\omega})|^2$ is maximum.
- (3) $\Phi_u^{(\infty)}(\omega)$ attains its (non-zero) minimum at the frequencies where $|G(e^{j\omega})|^2$ is minimum. Furthermore, if $|G(e^{j\omega})|^2 \rightarrow 0$ at some ω , then $\Phi_u^{(\infty)}(\omega) \rightarrow \lambda_e/\mu^2$.
- (4) μ is a continuous and strictly increasing function of λ_e , such that $\mu \rightarrow \|G\|_\infty$ as $\lambda_e \rightarrow 0$, and $\mu \rightarrow \infty$ as $\lambda_e \rightarrow \infty$.

The behavior of μ for $\lambda_e \ll 1$ is analyzed next.

Theorem 9. Let $G \in \mathcal{H}_\infty(\mathbb{E})$, where $\mathbb{E} := \{z \in \mathbb{C} : |z| \geq 1\}$, be such that $|G(e^{j\omega})|^2$ has a single global maximum in $[0, \pi]$ at, say, $\hat{\omega}$. Let

$$\frac{1}{\lambda_e} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{\mu^2(\lambda_e) - |G(e^{j\omega})|^2} d\omega,$$

where $\lambda_e > 0$ and $\mu : \mathbb{R}_0^+ \rightarrow (\|G\|_\infty, \infty)$. Then,

$$\mu^2(\lambda_e) = \|G\|_\infty^2 + \frac{2}{H_\mu} \lambda_e^2 + o(\lambda_e^2),$$

where $H_\mu := -\partial^2 |G(e^{j\omega})|^2 / \partial \omega^2|_{\omega=\hat{\omega}}$.

Theorem 9 enables a qualitative analysis of μ^2 as a function of λ_e . For instance, in the case where the \mathcal{H}_∞ -norm is attained due to a dynamic phenomenon with low damping, then H_μ is large and μ^2 significantly decreases as a function of λ_e .

3.2 Convergence to the Limit Spectrum

A remaining question is under which conditions the iterations (8) converge to the fixed point $\Phi_u^{(\infty)}$. The main result of this section is given by the following theorem.

Theorem 10. Let $\Phi_u^{(1)} \in \mathcal{L}_\infty([-\pi, \pi], \mathbb{R}_0^+)$ be such that $\inf_\omega \{\Phi_u^{(1)}(\omega)\} > 0$. Then, the sequence $\{\Phi_u^{(n)}\}_{n \in \mathbb{N}}$ generated by (8) converges in the \mathcal{L}_∞ norm to $\Phi_u^{(\infty)}$ in (9).

4. PROPERTIES OF THE ESTIMATOR

In this section, the results of Section 3 are employed to analyze Algorithm 1 in Section 2. Specifically, in Section 4.1, the bias of estimators 4 and (7) is analyzed, followed by an analysis of the value of iterations in terms of the Fisher information matrix in Section 4.2.

4.1 Bias Analysis

In this section, the bias of the nonparametric gain estimate is analyzed. Throughout this section, the emphasis is on the estimator in (7), since this enables a comparison with the results in Wahlberg et al. [2010a], where a similar estimator modulo the normalization is considered. In Wahlberg et al. [2010a], it is shown that this estimator is unbiased, provided that $u^{(n-1)}$ is in the eigenvector direction corresponding to the largest eigenvalue of $\mathbf{G}^T \mathbf{G}$, which in the case $N \rightarrow \infty$ corresponds to a sinusoidal

signal. These results are thus in line with the result of Theorem 2 and the discussion in Remark 4.

However, the results in Theorem 8 reveal that the input does not converge to the eigenvector direction corresponding to the largest eigenvalue of $\mathbf{G}^T \mathbf{G}$ if $\lambda_e > 0$, since even for $N \rightarrow \infty$ the input $u^{(n)}$ does not converge to a sinusoid. Hence the bias analysis in Wahlberg et al. [2010a] of the power iteration procedure in a stochastic framework is incomplete. The following result enables a more detailed analysis of the estimator $\hat{\beta}^{(n)}$.

Lemma 11. Consider the estimator (7). Then, for $N \rightarrow \infty$,

$$\mathbb{E} \left\{ [\hat{\beta}^{(n)}]^2 \right\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} |G(e^{j\omega})|^2 \Phi_u^{(n-1)}(\omega) d\omega. \quad (13)$$

Proof. Observe that for $N \rightarrow \infty$, the numerator and denominator in (7) can be recast as a sample cross-spectrum and a sample spectrum, respectively. Hence, taking expectations,

$$\begin{aligned} \mathbb{E} \left\{ [\hat{\beta}^{(n)}]^2 \right\} &= \mu^{(n-1)} \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_{y^{(n)}, u^{(n-1)}}(\omega) d\omega \\ &= \mu^{(n-1)} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{\mu^{(n-1)}} |G(e^{j\omega})|^2 \Phi_u^{(n-1)}(\omega) d\omega, \end{aligned}$$

which equals (13). \square

Lemma 11 in conjunction with Theorem 8 reveals several qualitative results with respect to the bias of the limit estimator $\hat{\beta}^{(\infty)}$. Indeed, note that for $\lambda_e > 0$, $\Phi_u^{(\infty)}$ is a smoothed Dirac delta function. Clearly, this implies that

$$\mathbb{E} \left\{ [\hat{\beta}^{(\infty)}]^2 \right\} < \|G\|_\infty^2 \quad \text{for } \lambda_e > 0, \quad (14)$$

hence the power iterations result in a biased estimate of the \mathcal{H}_∞ -norm if $\lambda_e > 0$. Similarly, observe that if $\lambda_e \rightarrow 0$, then $\Phi_u^{(\infty)}$ tends to a Dirac delta function. Exploiting the sifting property of the Dirac delta function, and the fact that the variance of $\hat{\beta}^{(n)}$ goes to zero as $\lambda_e \rightarrow 0$, reveals that

$$\hat{\beta}^{(\infty)} \rightarrow \|G\|_\infty \quad \text{in mean, as } \lambda_e \rightarrow 0,$$

hence the estimator is unbiased if $\lambda_e \rightarrow 0$.

A quantitative expression of the asymptotic bias of the power method is given in the following theorem.

Theorem 12. Consider the estimator (7). Then, for $N \rightarrow \infty$,

$$\mathbb{E} \left\{ [\beta^{(\infty)}]^2 \right\} = \|G\|_\infty^2 - \lambda_e + \frac{2}{H} \lambda_e^2 + o(\lambda_e^2).$$

Proof. Lemma 11 reveals that for $n \rightarrow \infty$,

$$\mathbb{E} \left\{ [\beta^{(\infty)}]^2 \right\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} |G(e^{j\omega})|^2 \Phi_u^{(\infty)}(\omega) d\omega. \quad (15)$$

Combining (15) and (12) yields

$$\mu^2 = \mathbb{E} \left\{ [\beta^{(\infty)}]^2 \right\} + \lambda_e.$$

Next, rearranging and applying the result of Theorem 9 gives the following asymptotic expression for the bias:

$$\mathbb{E} \left\{ [\beta^{(\infty)}]^2 \right\} = \mu^2 - \lambda_e = \|G\|_\infty^2 - \lambda_e + \frac{2}{H} \lambda_e^2 + o(\lambda_e^2),$$

which concludes the proof. \square

Theorem 12 shows that the asymptotic bias, $\mathbb{E} \left\{ [\beta^{(\infty)}]^2 \right\} - \|G\|_\infty^2$, is dominated by $-\lambda_e$ in the small noise regime.

Equivalently, $E\{\beta^{(\infty)}\} - \|G\|_\infty = -(1/2)\lambda_e \|G\|_\infty^{-1} + o(\lambda_e)$, which corroborates the previous analysis.

Furthermore, Theorem 12 shows that the normalization factor, $\mu^{(n)}$, which as $n, N \rightarrow \infty$ is equal to μ , might be a better estimator of $\|G\|_\infty$ for small λ_e in terms of bias. Indeed, observe that $\mu^{(n)}$ can be interpreted as a direct estimator for (2) by forming the product $[y^{(n)}]^T y^{(n)}$.

4.2 Fisher Information Matrix per Iteration

The value of iterations is immediate if the nonparametric estimator (7) is used, since Theorem 2 reveals that the estimate converges to the \mathcal{H}_∞ -norm for $\lambda_e = 0$ and an increasing number of iterations. However, the value of power iterations has not yet been investigated in the general case where possibly another estimator is used. Independent of the specific estimator, the limit of accuracy of the power iterations method can be analyzed through the asymptotic information matrix. Hereto, an analysis is performed where an underlying parametric model is considered.

Specializing to the prediction error framework, let $G(z, \theta)$ be a parametric model structure. Assuming that there exists a θ_o , called the *true parameter*, such that $G(z, \theta_o) = G$, where G denotes the true system, then under mild conditions

$$\sqrt{N}(\hat{\theta}_N - \theta^*) \xrightarrow{d} \mathcal{N}(0, P_\theta),$$

see, e.g., [Ljung, 1999b, Chapter 9]. The prediction error estimator turns out to be asymptotically efficient, i.e., the asymptotic covariance matrix P_θ equals the inverse of the (Fisher) information matrix, which is given by

$$I_\theta = E\{\psi_t \Lambda^{-1} \psi_t^T\}. \quad (16)$$

In addition,

$$\psi_t = - \left. \frac{\partial \varepsilon_t^T}{\partial \theta} \right|_{\theta=\theta_o}, \quad (17)$$

where ε_t is the prediction error

$$\varepsilon_t = y_t - \hat{y}_{t|t-1}. \quad (18)$$

Evaluating the Fisher information matrix for the system (1) and Algorithm (1) yields the following result.

Lemma 13. Consider system (1) and the power iterations algorithm 1. Then,

$$I_\theta = \sum_{k=1}^n I_\theta^{(k)}, \quad (19)$$

where $I_\theta^{(k)} = \frac{1}{2\pi\lambda_e} \int_{-\pi}^{\pi} G'(e^{j\omega}) \Phi_u^{(k)}(\omega) (G'(e^{-j\omega}))^T d\omega$ and $G'(z) := \partial G(z, \theta) / \partial \theta|_{\theta=\theta_o}$.

Proof. Using (1) and (18), $\varepsilon_t^{(k)} = y_t^{(k)} - \hat{y}_{t|t-1}^{(k)}$. Clearly, $\frac{\partial \varepsilon_t^{(k)}}{\partial \theta} = -G'(q)u_t^{(k)}$. Next, using (17),

$$\psi_t = \begin{bmatrix} \frac{\partial \varepsilon_t^{(1)}}{\partial \theta} & \frac{\partial \varepsilon_t^{(2)}}{\partial \theta} & \dots & \frac{\partial \varepsilon_t^{(k)}}{\partial \theta} \end{bmatrix} = G'(q)U_t,$$

where $U_t := \begin{bmatrix} u_t^{(1)} & u_t^{(2)} & \dots & u_t^{(k)} \end{bmatrix}$. Next, using (16) and Parseval's relation,

$$\begin{aligned} I_\theta &= E\{\psi_t \Lambda^{-1} \psi_t^T\} = \frac{1}{\lambda_e} E\{\psi_t \psi_t^T\} \\ &= \frac{1}{2\pi\lambda_e} \int_{-\pi}^{\pi} G'(e^{j\omega}) \Phi_U(\omega) (G'(e^{-j\omega}))^T d\omega. \end{aligned}$$

Note that here

$$\Phi_U(\omega) = \sum_{\tau=-\infty}^{\infty} R_U(\tau) e^{-j\tau\omega}, \quad (20)$$

$$R_U(\tau) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N \sum_{k=1}^n u_t^{(k)} u_{t-\tau}^{(k)} = \sum_{k=1}^n R_{u^{(k)}}(\tau). \quad (21)$$

Next, combining (20) and (21) yields (19). \square

Several interesting observations can be made with respect to Lemma 13. Firstly, it is observed that the Fisher information matrix for iterative schemes satisfies an additivity property, see (19), i.e., additional experiments can only increase information about the system. This is consistent with Fisher's original requirements for the definition of information [Porat, 1994, page 59]. Secondly, due to the uniform convergence of $\Phi_u^{(n)}$, (19) divided by n corresponds to a Cesàro sum, it holds in the asymptotic case where the number of experiments $n \rightarrow \infty$,

$$\lim_{n \rightarrow \infty} \frac{I_\theta}{n} = \frac{1}{2\pi\lambda_e} \int_{-\pi}^{\pi} G'(e^{j\omega}, \theta) \Phi_u^{(\infty)}(\omega) (G'(e^{-j\omega}, \theta))^T d\omega.$$

If $\lambda_e \rightarrow 0$, then $\Phi_u^{(\infty)}$ tends to a Dirac delta function. In Wahlberg et al. [2010b] and Wahlberg et al., this is shown to lead to optimal accuracy for estimating a peak of $|G(e^{j\omega})|$. In this respect, the power iterations are of value in the sense of iterative experiment design.

5. EXAMPLE

In this section, the results of the preceding sections are illustrated. Specifically, the limit spectra are computed by using the results of Sec. 3, followed by an analysis of the bias corresponding to the theoretical results in Sec. 4.1.

Consider the system G given by

$$G(z) = \frac{0.2155z^{-1} + 0.2012z^{-2}}{1 - 0.9854z^{-1} + 0.8187z^{-2}}.$$

A Bode magnitude diagram of G is depicted in Fig. 2 (left).

First, the results of Theorem 8 are illustrated. Hereto, use is made of an eigenvalue equation of an extended system H that appears in the proof of Theorem 10, which will be published elsewhere. Discretization of H in ω yields a discrete frequency set $\Omega = \{\omega_1, \omega_2, \dots, \omega_{n_\omega}\}$. This results in the discretized operator

$$H_d = \begin{bmatrix} |G(\omega_1)|^2 & & 0 & \lambda_e \\ & \ddots & & \vdots \\ 0 & & |G(\omega_{n_\omega})|^2 & \lambda_e \\ \frac{1}{n_\omega} |G(\omega_1)|^2 & \dots & \frac{1}{n_\omega} |G(\omega_{n_\omega})|^2 & \lambda_e \end{bmatrix}.$$

Next, to compute the discretized spectrum $\Phi_{u,d}^{(\infty)}(\omega), \omega \in \Omega$, a vector

$$\lambda_{\max}(H_d) \begin{bmatrix} \Phi_{u,d}^{(\infty)}(\omega_1) \\ \vdots \\ \Phi_{u,d}^{(\infty)}(\omega_{n_\omega}) \\ 1 \end{bmatrix} = H_d \begin{bmatrix} \Phi_{u,d}^{(\infty)}(\omega_1) \\ \vdots \\ \Phi_{u,d}^{(\infty)}(\omega_{n_\omega}) \\ 1 \end{bmatrix}$$

has to be computed. The resulting spectrum for $\lambda_e \in \{1, 4, 9\}$ (using $n_\omega = 1000$) is depicted in Fig. 2 (right). Considering Fig. 2 (left) in conjunction with Fig. 2 (right) confirms Properties (1)-(4) in Theorem 8.

Next, to illustrate the behavior in the time domain and analyze the bias, the estimate $\beta^{(n)}$ is computed for a

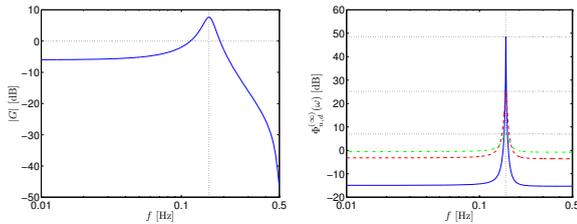


Fig. 2. Left: Bode magnitude diagram of G . Right: limit spectrum $\Phi_{u,d}^{(\infty)}(\omega)$, $\lambda_e = 1$ (solid blue), $\lambda_e = 4$ (dashed red), $\lambda_e = 9$ (dash-dotted green).

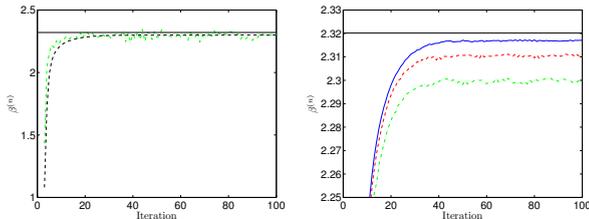


Fig. 3. Estimated gain $\beta^{(n)}$ during iterations. Left: noise variance $\lambda_e = 9$: single realization (dash-dotted green), average over 1000 realizations (dashed black), finite time ℓ_2 -induced norm (solid black). Right: Varying noise variance, averaged over 1000 realizations, $\lambda_e = 1$ (solid blue), $\lambda_e = 4$ (dashed red), $\lambda_e = 9$ (dash-dotted green), finite time ℓ_2 -induced norm (solid black).

realization of the iterative algorithm 1 with $\lambda_e = 9$, see Fig. 3 (left). Due to the presence of e_t in (1), the estimate $\beta^{(n)}$ is noisy as expected. To analyze the bias, 1000 realizations of the iterative procedure are averaged, see Fig. 3 (left). The mean of $\beta^{(n)}$ over different realizations of the iterative algorithm is strictly below the finite time ℓ_2 -induced norm of G , i.e., $\sqrt{\lambda_{\max}(\mathbf{G}^T \mathbf{G})}$, confirming (14).

Finally, 1000 realizations of the iterative procedure are averaged for $\lambda_e \in \{1, 4, 9\}$, see Fig. 3 (right). It is observed that the bias increases for increasing λ_e . From Lemma 11, this can be understood when considering the smoothness of G and the spectra in Fig. 2.

6. CONCLUSION

The results presented in this paper contribute to the analysis of the role of iterations in system identification for control. An approach for nonparametric \mathcal{H}_∞ -norm estimation is presented that requires only one experiment for the estimation procedure. In addition, it is shown that for a nonparametric \mathcal{H}_∞ -norm estimation through iterative experiments, (1) additive disturbances can introduce bias errors, and (2) iterative procedures can be interpreted as experiment design for \mathcal{H}_∞ -norm estimation, and the value of iterations has been investigated by means of the Fisher information matrix. The analysis is based on a frequency domain approach, the novelty of the presented approach is that it addresses both (1) additive stochastic disturbances that represent measurement errors, and (2) the normalization of the input signal to account for input power constraints, which involves a nonlinear operation.

Future research includes the extension of the presented results in several directions. Firstly, it is shown in the present paper that the implementation of power iterations in the presence of stochastic disturbances results in bias errors. Presently, modifications of the input signal update, i.e., Step 3 in Algorithm 1, as well as the estimators (4) and (7) are being investigated to avoid these bias errors, e.g.,

by using data from old experiments through stochastic approximation. Secondly, extensions of the power iteration algorithm are being investigated that enable the nonparametric estimation of other system properties.

REFERENCES

- P. Albertos and A. Sala, editors. *Iterative Identification and Control*. Springer, 2002.
- M. Barenthin, H. Mosskull, H. Hjalmarsson, and B. Wahlberg. Validation of stability for an induction machine drive using power iterations. In *IFAC 16th Triennial World Congress*, Prague, Czech Republic, 2005.
- M. Barenthin, M. Enqvist, B. Wahlberg, and H. Hjalmarsson. Gain estimation for Hammerstein systems. In *14th IFAC Symp. Sys. Id.*, Newcastle, Australia, 2006.
- J. M. Böling and P. M. Mäkilä. On control relevant criteria in \mathcal{H}_∞ identification. *IEEE Trans. Automat. Contr.*, 43(5):694–700, 1998.
- A. Böttcher and S. M. Grudsky. *Spectral properties of banded Toeplitz matrices*. SIAM, 2005.
- D. A. Bristow, M. Tharayil, and A. G. Alleyne. A survey of iterative learning control: A learning-based method for high-performance tracking control. *IEEE Contr. Syst. Mag.*, 26(3):96–114, 2006.
- G. Golub and C. Van Loan. *Matrix Computations, 3rd Edition*. Johns Hopkins University Press, 1996.
- R. G. Hakvoort and P. M. J. Van den Hof. Identification of probabilistic system uncertainty regions by explicit evaluation of bias and variance errors. *IEEE Trans. Automat. Contr.*, 42(11):1516–1528, 1997.
- H. Hjalmarsson. Iterative feedback tuning – an overview. *Int. J. Adapt. Contr. Sign. Proc.*, 16(5):373–395, 2002.
- H. Hjalmarsson. From experiment design to closed-loop control. *Automatica*, 41:393–438, 2005.
- H. Hjalmarsson, S. Gunnarsson, and M. Gevers. Optimality and sub-optimality of iterative identification and control design schemes. In *Proc. 15th Americ. Contr. Conf.*, pages 2559–2563, Seattle, WA, USA, 1995.
- T. P. Krasulina. Method of stochastic approximation in the determination of the largest eigenvalue of the mathematical expectation of random matrices. *Aut. Rem. Contr.*, 2:215–221, 1970.
- L. Ljung. Model validation and model error modeling. In *The Åström Symposium on Control*, pages 15–42, Lund, Sweden, 1999a.
- L. Ljung. *System Identification: Theory for the User, 2nd Edition*. Prentice Hall, 1999b.
- E. Oja and J. Karhunen. On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix. *J. Math. Anal. Appl.*, 106(1):69–84, 1985.
- T. Oomen, C. Rojas, H. Hjalmarsson, and B. Wahlberg. In preparation. *Technical Report available at <http://www.ee.kth.se/~crrro/techrep.pdf>*, 2011.
- B. Porat. *Digital Processing of Random Signals*. Prentice-Hall, 1994.
- B. Wahlberg, H. Hjalmarsson, and P. Stoica. On the performance of optimal input signals for frequency response estimation. *Submitted for journal publication*.
- B. Wahlberg, M. Barenthin, and H. Hjalmarsson. Nonparametric methods for \mathcal{L}_2 -gain estimation using iterative experiments. *Automatica*, 46(8):1376–1381, 2010a.
- B. Wahlberg, H. Hjalmarsson, and P. Stoica. On optimal input signal design for frequency response estimation. In *Proc. 49th Conf. Dec. Contr.*, Atlanta, GA, USA, 2010b.