

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

Cristian R. Rojas<sup>a</sup> Tom Oomen<sup>b</sup> Håkan Hjalmarsson<sup>a</sup> Bo Wahlberg<sup>a</sup>

<sup>a</sup>*Automatic Control Lab and ACCESS Linnaeus Center, Electrical Engineering, KTH – Royal Institute of Technology, S-100 44 Stockholm, Sweden (e-mail: cristian.rojas@ee.kth.se, hakan.hjalmarsson@ee.kth.se, bo.wahlberg@ee.kth.se).*

<sup>b</sup>*Eindhoven University of Technology, Department of Mechanical Engineering, Control Systems Technology Group, PO Box 513, Building GEM-Z, 5600 MB Eindhoven, The Netherlands (e-mail: t.a.e.oomen@tue.nl).*

---

## Abstract

Many iterative approaches in the field of system identification for control have been developed. Although successful implementations have been reported, a solid analysis with respect to the convergence of these iterations has not been established. The aim of this paper is to present a thorough analysis of a specific iterative algorithm that involves nonparametric  $\mathcal{H}_\infty$ -norm estimation. The pursued methodology involves a novel frequency domain approach that addresses both additive stochastic disturbances and input normalization. The results of the convergence analysis are twofold: (1) the presence of additive disturbances introduces a bias in the estimation procedure, and (2) the iterative procedure can be interpreted as experiment design for  $\mathcal{H}_\infty$ -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.

*Key words:* Nonparametric methods, Iterative modelling and control design, Input and excitation design

---

## 1 Introduction

In the last decades, many iterative approaches have been proposed in the fields of system identification and control design. Relevant examples of such iterative approaches include iterative learning control (ILC) [8], iterative feedback tuning [14], and iterative identification and control [1]. 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 [16]. Furthermore, the iterations of these approaches may be divergent as is discussed in [1, Sec. 9.3]. Finally, the value of iterations of these approaches has been questioned in, e.g., [6].

Recently, an iterative approach for nonparametric  $\mathcal{H}_\infty$ -norm estimation has been proposed in [15, Sec. 12.2] and further extended in [27]. A relevant application of  $\mathcal{H}_\infty$ -norm estimation includes model error modeling, since re-

liable robust control design methodologies are available that consider model errors as  $\mathcal{H}_\infty$ -norm bounded operators. In contrast to most model error modeling techniques, including [12, 26, 18], the approach presented in [15, Sec. 12.2] does not require the estimation of an intermediate parametric model. Indeed, in [15, Sec. 12.2], the input to the system is iteratively determined, followed by a nonparametric estimation of the  $\mathcal{H}_\infty$ -norm from the measured data of two experiments. An essential property of the iterative procedure is that it is related to the so-called ‘power iterations’ method [11, Section 8.2], which known is to converge to the global optimum with an exponential rate of convergence in the finite-dimensional noise-free case.

Although several successful applications of iterative nonparametric  $\mathcal{H}_\infty$ -norm estimation have already been reported, including [2, 3], 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 observations. 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 specific iterative nonparametric

---

\* This work was supported in part by the Swedish Research Council under contract 621-2009-4017.

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 [17, 20], 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., [11], 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 stochastic, possibly unbounded, disturbances.

Other nonparametric approaches for the estimation of the  $\mathcal{H}_\infty$  norm have been proposed in the literature, see e.g. [13]. However, these methods do not include the design of the input signal to improve the statistical quality of the  $\mathcal{H}_\infty$  norm estimator.

The specific contributions of the paper are as follows:

- C1) In [27], a procedure for iterative nonparametric norm estimation is presented that requires two dedicated experiments for each iteration. In this paper, an iterative procedure is presented that only requires a single experiment per iteration.
- C2) In [17, 20, 11], analyses of power iterations in imperfect environments are performed, e.g., due to finite arithmetic precision in computer implementations. In this paper, power iterations that are subject to additive stochastic disturbances, a common scenario in system identification, are analyzed. A novel procedure in terms of limit power spectra is presented that appropriately deals with input normalization, which is required to account for, e.g., input power constraints.
- C3) The convergence of the spectra to the corresponding limit spectra, see C2, is established by employing the Hilbert projective metric, see [5, 9] for a definition. This analysis method appears novel in the context of convergence analysis of identification algorithms and has further potential in this field.
- C4) In [27], bias and variance analyses of nonparametric estimators are presented that are based on the assumption that the input converges to the optimal input corresponding to the noise-free case. However, C2 reveals that this is not the case in general. An extended bias analysis is presented, which quantifies the bias error and reveals that the iterative nonparametric norm estimation procedure leads to biased results in case of additive noise.
- C5) The value of iterations is investigated by means of the Fisher information matrix, revealing that the iterative algorithm can be interpreted as an optimal experiment design approach for  $\mathcal{H}_\infty$ -norm estimation for both parametric and nonparametric system identification methodologies.

The results are of significant practical importance in case

the iterative nonparametric norm estimation procedure is used for model error modeling. Indeed, measurement data from physical systems are always contaminated by measurement noise and disturbances. By virtue of Item C4, above, the estimate of the  $\mathcal{H}_\infty$ -norm is generally biased. In particular, it is shown that the  $\mathcal{H}_\infty$ -norm is underestimated, even though such bias is small in low noise scenarios. In case of model error modelling, this implies that the bound on the model error is smaller than what it should be. Consequently, in low noise scenarios the  $\mathcal{H}_\infty$  norm estimator gives a reasonable bound on the model error, while if the noise level is not small, this bound can underestimate the model error, which can be dangerous if used in robustness analysis or robust controller synthesis.

The outline of the paper is as follows. In Sec. 2, a nonparametric  $\mathcal{H}_\infty$ -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. For convenience of the reader, most of the proofs have been collected in Appendix A.

*Notation.* The variable  $z$  denotes either the forward shift operator  $zx_t = x_{t+1}$  in the time domain or the  $\mathcal{Z}$  transform variable in the frequency domain, depending on the context. For a square symmetric matrix  $X$ ,  $\lambda_{\max}(X) := \max_i |\lambda_i(X)|$ . Let  $\mathbf{E} := \{z \in \mathbb{C} : |z| > 1\}$  and  $\bar{\mathbf{E}} := \{z \in \mathbb{C} : |z| \geq 1\}$ .  $\mathcal{H}_\infty(\mathbf{E})$  and  $\mathcal{H}_\infty(\bar{\mathbf{E}})$  are the spaces of functions  $\sum_{k=0}^{\infty} c_k z^k$  which are analytic and bounded on  $\mathbf{E}$  and  $\bar{\mathbf{E}}$ , respectively, while  $\mathcal{H}_\infty^-(\mathbf{E})$  is the space of functions  $\sum_{k=1}^{\infty} c_k z^{-k}$  which are analytic and bounded on  $\{z \in \mathbb{C} : |z| < 1\}$ . In addition, the  $\ell_2$  norm of a signal  $\{x_t\}_{t \in \mathbb{N}}$  is defined as  $\|x\|_2 := (\sum_{t=1}^{\infty} x_t^2)^{1/2}$ . The  $\mathcal{H}_\infty$ -norm of an asymptotically stable linear time invariant (LTI) single input single output (SISO) discrete time system  $G$  is given by  $\|G\|_\infty := \sup_{|z| > 1} |G(z)| = \sup_{\omega \in (-\pi, \pi]} |G(e^{j\omega})|$ .  $\mathcal{L}_\infty(X, \mathbb{R})$  is the Banach space of bounded real-valued functions on  $X$ , endowed with the norm  $\|f\|_\infty := \sup_{x \in X} |f(x)|$ . Unless specified, integrals are assumed to be taken with respect to the Lebesgue measure in  $[-\pi, \pi]$ .  $o(f(N))$  and  $O(f(N))$ , where  $f : \mathbb{N} \rightarrow \mathbb{R}^+$ , denote functions from  $N \in \mathbb{N}$  to  $\mathbb{R}^+$  such that  $o(f(N))/f(N) \rightarrow 0$  as  $N \rightarrow \infty$ , and  $|O(f(N))| \leq C|f(N)|$  for all  $N \in \mathbb{N}$ , where  $C > 0$  is a pre-specified constant. The notation  $\ln$  denotes the natural logarithm.

## 2 Iterative Nonparametric Norm Estimation

The asymptotically stable SISO LTI system

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

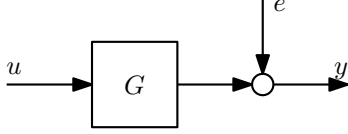


Fig. 1. Block diagram of the system.

is considered, see also Fig. 1, where  $u = \{u_t\}$  denotes a quasi-stationary input to the system [19],  $y = \{y_t\}$  is the output, and  $e = \{e_t\}$  is white noise with variance  $\lambda_e > 0$ , representing measurement noise or a disturbance term. In addition, it is assumed that  $e$  and  $u$  are independent.

One of the key properties of the  $\mathcal{H}_\infty$ -norm is that it equals the  $\ell_2$ -induced norm, i.e.,

$$\|G\|_\infty = \sup_{u \in \ell_2, u \neq 0} \frac{\|y\|_2}{\|u\|_2}. \quad (2)$$

The characterization of the  $\mathcal{H}_\infty$ -norm in (2) is useful for at least two reasons. Firstly, in the situation where (1) is the difference between the outputs of a system and a model, it enables the representation of model uncertainty by  $\mathcal{H}_\infty$ -norm bounded operators, since it is an induced norm. Secondly, (2) is at the basis of the nonparametric  $\mathcal{H}_\infty$ -norm estimation algorithm that is presented in this paper.

To motivate the development of the algorithm to be analyzed here, consider first the problem of estimating the  $\mathcal{H}_2$ -norm of  $G$ , i.e.,  $\|G\|_2 = \|Gu\|_2$ , where  $u$  is a Dirac impulse. By its definition, a natural estimator of  $\|G\|_2$  can be built without an explicit model for  $G$ , using simply an experiment where  $u$  is zero mean white noise with unit variance and taking  $\|y\|_2/\sqrt{N}$  as estimate of  $\|G\|_2$ .

Referring to (2), in order to estimate  $\|G\|_\infty$  without a model structure, it is necessary to choose the input of unit norm such that  $\|y\|_2$  is maximum. Such an input corresponds to a sinusoid whose frequency  $\omega$  is such that  $\|G(e^{j\omega})\|_2$  is maximum. The algorithm considered in this paper generates, via a series of iterative experiments, a sequence of inputs which approximates (in the limit) the optimal input signal, without requiring a model of the plant.

The following algorithm constitutes Contribution C1 of the paper and enables nonparametric estimation of the  $\mathcal{H}_\infty$ -norm. It generates a sequence of inputs of length  $N$  that can be used to estimate the  $\mathcal{H}_\infty$ -norm.

**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  $G$ , initially at rest (i.e., zero initial condition), obtaining  $y^{(n)} := [y_1^{(n)} \dots y_N^{(n)}]^T$ .
- (3) Time reverse the sequence  $y^{(n)}$ , i.e., determine  $\tilde{y}^{(n)} := [y_N^{(n)} \dots y_1^{(n)}]^T$ , and generate  $u^{(n+1)} =$

$\tilde{y}^{(n)}/\mu^{(n)}$ , where the normalization  $\mu^{(n)}$  is defined below.

- (4) Let  $n \mapsto n + 1$  and go to Step (2).

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. For a definition of the power norm, see, e.g., [30, Section 4.4]. Note that the normalization to unity (instead of an arbitrary positive constant) is nonrestrictive and introduced for notational convenience. Also, the normalization  $\mu^{(n)}$  can directly be adapted, e.g., to energy norm constraints or maximum amplitude constraints.

Algorithm 1 generates sequences  $\{u^{(n)}\}$  and  $\{y^{(n)}\}$ ,  $n = 1, \dots, N$ , from which the  $\ell_2$ -induced gain of  $G$  is estimated by

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

The attractive feature of the  $\mathcal{H}_\infty$ -norm estimation algorithm to be analyzed here is that it does not require a finite dimensional parametric model structure assumed to contain the true system; in fact notice that we impose no other assumption on the impulse response than stability. For this reason, and perhaps with some abuse of notation, the algorithm analyzed here can be cataloged as ‘non-parametric’, ‘model-free’, or ‘direct data-driven’, in contrast to other parametric or model-based techniques. On the other hand, it is also shown in Remark 10 that the method can be interpreted as estimating a single parameter model in the frequency domain (even though it is *not* assumed that such simple model structure contains the true system).

The ‘data-driven’ property of the analyzed algorithm is particularly useful for model error modelling [18], where the goal is to estimate how far is a given nominal parametric model  $\hat{G}$  from a true system  $G_0$ . This problem can be posed e.g. as an  $\mathcal{H}_\infty$ -norm estimation task, where instead of  $G_0$  we should work with the plant-model mismatch  $\hat{G} - G_0$  (or another transfer function reflecting the modelling error), and for this task we need to impose as little prior knowledge on  $\hat{G} - G_0$  as possible. This motivates the preference for data-driven/model-free approaches.

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

and  $\tilde{y}^{(n)} = \mathcal{T}y^{(n)}$ , hence  $\tilde{y}^{(n)} = \mathbf{G}_H u^{(n)}$ , where

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

and

$$\mathbf{G}_H := \mathcal{T}\mathbf{G} = \begin{bmatrix} g_{N-1} & g_{N-2} & \cdots & g_0 \\ g_{N-2} & g_{N-3} & & 0 \\ \vdots & & \ddots & \\ g_0 & 0 & \cdots & 0 \end{bmatrix}. \quad (5)$$

The following assumption is imposed for clarity of exposition.

**Assumption 1**  $\lambda_{\max}(\mathbf{G}_H)$  is unique. Furthermore,  $\omega \mapsto |G(e^{j\omega})|^2$  has a unique maximum in  $[0, \pi]$ .

**Lemma 2** Let Assumption 1 hold and assume that  $u^{(1)}$  has a nonzero component in the direction of the eigenvector associated with  $\lambda_{\max}(\mathbf{G}_H)$ . Then,  $\hat{\beta}^{(n)} \rightarrow \lambda_{\max}(\mathbf{G}_H)$  as  $n \rightarrow \infty$ .

See [11, Section 7.3.1] for a proof of Lemma 2. The following theorem establishes that  $\hat{\beta}^{(n)}$  converges to  $\|G\|_\infty$  as  $N \rightarrow \infty$ .

**Theorem 3** Consider the systems  $G$  in (1) and  $\mathbf{G}_H$  in (5). Then,  $\|G\|_\infty = \lim_{N \rightarrow \infty} \lambda_{\max}(\mathbf{G}_H)$ .

Lemma 2 and Theorem 3 reveal that the estimate (4) is a reasonable estimate of  $\|G\|_\infty$  for a sufficiently long duration of each experiment and after a sufficient number of iterations.

The following remarks are appropriate.

**Remark 4** The requirement in Algorithm 1 that the system  $G$  should be at rest before the start of each iteration may seem restrictive. However, it is a standard requirement for other iterative techniques, such as Iterative Learning Control [8], which have been successfully applied in areas such as robotics and mechatronics. Nevertheless, this assumption only affects Lemma 2 and Theorem 3, because the remaining results (from Section 3 on) consider the number of samples to be  $N \rightarrow \infty$ , and hence the effect of the initial conditions are negligible.

**Remark 5** Algorithm 1 assumes that the number of samples is the same for each iteration. In case the user

would prefer to change  $N$ , Algorithm 1 could be modified by either removing part of the sequence (if the new  $N$  is smaller than the previous one) or by adding white noise or a periodic extension of the previous sequence. However, the analysis of these variants of Algorithm 1 is beyond the scope of the paper.

**Remark 6** The results of Lemma 2 and Theorem 3 can also be established when relaxing Assumption 1. However, Assumption 1 ensures uniqueness of the spectrum of  $\lim_{n \rightarrow \infty} u^{(n)}$ , which significantly simplifies the developments in the remainder of this paper.

**Remark 7** 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  $\ell_2$  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  $\ell_p$ , whose spectral radius is  $\|\mathbf{G}\|_\infty$ , see, e.g., [7, 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 8** To point out the relation between the presented approach, i.e., Algorithm 1 in conjunction with the estimator (4), and the  $\ell_2$ -induced norm characterization of the  $\mathcal{H}_\infty$ -norm, i.e., (2), as well as to clarify the relation with the algorithm in [27] that requires two experiments for nonparametric  $\mathcal{H}_\infty$ -norm estimation, observe that

$$\begin{aligned} \sup_{u^{(n)} \in \ell_2, u \neq 0} \frac{\|y^{(n)}\|_2^2}{\|u^{(n)}\|_2^2} &= \sup_{u^{(n)} \in \ell_2, u \neq 0} \frac{y^{(n)T} y^{(n)}}{u^{(n)T} u^{(n)}} \\ &= \sup_{u^{(n)} \in \ell_2, u \neq 0} \frac{u^{(n)T} \mathbf{G}^T \mathbf{G} u^{(n)}}{u^{(n)T} u^{(n)}} \end{aligned}$$

which is clearly maximal if  $u^{(n)}$  is in the eigenvector direction corresponding to  $\lambda_{\max}(\mathbf{G}^T \mathbf{G})$ . In [27], 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  $\mathbf{G}^T \mathbf{G} = \mathcal{T} \mathbf{G} \mathcal{T} \mathbf{G}$ . 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,

$$\mathbf{G}^T \mathbf{G} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^T \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^T = \mathbf{Q} \mathbf{\Lambda}^2 \mathbf{Q}^T,$$

hence  $\mathbf{G}^T \mathbf{G}$  has eigenvalues  $\Lambda^2$ .

**Remark 9** The eigenvalues of  $\mathbf{G}_H$ , unlike those of  $\mathbf{G}^T \mathbf{G}$ , see Remark 8, are not guaranteed to be positive.

As a result,  $\hat{\beta}^{(n)} = \frac{[u^{(n)}]^T \tilde{y}^{(n)}}{N}$  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. To avoid an oscillatory transient, the maximal eigenvalue of  $\mathbf{G}^T \mathbf{G}$  can be estimated directly, see also Remark 8. This estimator is similar to (4) but uses the signal  $u^{(n-1)}$ , from experiment  $n-1$ , instead  $u^{(n)}$ . In addition, the input normalization, i.e.,  $\mu^{(n-1)}$  is appropriately compensated for, leading to the estimator

$$\hat{\beta}_2^{(n)} = \sqrt{\frac{\mu^{(n-1)} [u^{(n-1)}]^T \tilde{y}^{(n)}}{N}}. \quad (6)$$

Estimator (6) may thus be preferable in the case where it is desired to determine the absolute value of the maximum eigenvalue. Specifically, this estimator corresponds to the square root of the largest eigenvalue of  $\mathbf{G}_H^2 = \mathbf{G}^T \mathbf{G}$ , i.e., to the largest eigenvalue of  $\mathbf{G}_H$ . Since  $\mathbf{G}_H^2 = \mathbf{G}^T \mathbf{G}$  is positive semidefinite, this last estimator does not suffer from a possible slow oscillatory transient. Note that the analysis in the remainder of the paper is largely independent of the considered estimator.

**Remark 10** The estimator  $\hat{\beta}^{(n)}$  in (4) can also be interpreted as coming from a one-parameter frequency-domain model:  $Y^{(n)} = \alpha U^{(n)} + E^{(n)}$ , where  $U^{(n)}$ ,  $Y^{(n)}$  and  $E^{(n)}$  are the discrete Fourier transform vectors of the input, output and noise at iteration  $n$ , respectively, based on  $N$  samples [19], and  $\alpha$  is a complex parameter. The maximum likelihood estimator of  $\alpha$ , assuming that  $E \sim \mathcal{N}_c(0, I)$ , is  $\hat{\alpha} = (U^{(n)})^H Y^{(n)} / (U^{(n)})^H U^{(n)} = (U^{(n)})^H Y^{(n)}$  (since the energy of the input has been normalized). Therefore, by the principle of invariance, the maximum likelihood estimator of  $\|G\|_\infty$  is  $|\hat{\alpha}| = |(U^{(n)})^H Y^{(n)}| = |(Y^{(n)})^H U^{(n)}|$ , which can be shown to correspond to  $\hat{\beta}^{(n)}$ .

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 (6) are analyzed in more detail.

### 3 Convergence of the Iterations

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, which constitutes Contribution C2 of the paper, followed by a convergence analysis in Section 3.2, which corresponds to Contribution C3 of the paper.

#### 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  $u$  at iteration  $n$ ; see [19, Chapter 2] for an appropriate definition.

In order to study the behavior of Algorithm 1 for  $N \rightarrow \infty$ , consider a quasi-stationary signal  $\{u_t^{(1)}\}$ , whose truncation  $[u_1^{(1)} \cdots u_N^{(1)}]^T$  is used as an initial input vector. As  $N$  is increased, the truncated signals such as  $u^{(n)}$  and  $y^{(n)}$  have discrete Fourier transforms whose (squared) magnitudes converge weakly (with probability 1) to well defined spectra [19, Eq. (6.47)]. Throughout, the results refer to the properties of these spectra.

Throughout this section, the time reversal operation is omitted in the frequency domain analysis of the power iterations method. Indeed, for a finite  $N$ , the effect of the time reversal operation  $\mathcal{T}$ , which can be interpreted 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, see also Remark 7.

**Lemma 11** Consider Algorithm 1 applied to the system  $G$  in (1). Then, for  $\lambda_e > 0$  and  $N \rightarrow \infty$ ,

$$\Phi_u^{(n+1)}(\omega) = \frac{|G(e^{j\omega})|^2 \Phi_u^{(n)}(\omega) + \lambda_e}{\frac{1}{2\pi} \int_{-\pi}^{\pi} |G(e^{j\omega})|^2 \Phi_u^{(n)}(\omega) d\omega + \lambda_e}. \quad (7)$$

**PROOF.** Combining (1) and Algorithm 1 yields

$$u^{(n+1)} = \frac{1}{\mu^{(n)}} (G u^{(n)} + e^{(n)}),$$

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

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

Finally, applying Parseval relation to  $\mu^{(n)}$  yields (7).  $\square$

Next, the fixed points of the function (7) are analyzed.

**Theorem 12** Consider the situation of Lemma 11. Then, the function (7) has a unique fixed point  $\Phi_u^{(\infty)} \in \mathcal{L}_1([- \pi, \pi], \mathbb{R}_0^+)$ , which is given by

$$\Phi_u^{(\infty)}(\omega) := \frac{\lambda_e}{\mu^2 - |G(e^{j\omega})|^2}, \quad (8)$$

where  $\mu > 0$  satisfies

$$\frac{1}{\lambda_e} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{\mu^2 - |G(e^{j\omega})|^2} d\omega. \quad (9)$$

**PROOF.** Every fixed point  $\Phi_u^{(\infty)}$  satisfies the equation

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

Denoting

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

and solving (10) for  $\Phi_u^{(\infty)}$  yields (8). In addition, for a given value of  $\mu$ , the resulting  $\Phi_u^{(\infty)}$  is unique. Next, to show (9), note that substitution of (8) into (11) yields

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

By strict positivity of  $\lambda_e$  and hence also of  $\mu^2$ , see (11), the latter equation directly implies (9). Finally, notice that the right side of (9) is strictly decreasing in  $\mu$ , which implies the uniqueness of  $\mu$ , and hence also of the fixed point  $\Phi_u^{(\infty)}$ .  $\square$

The following corollary, whose proof is omitted since it is immediate from (1) and Theorem 12, establishes an equivalent result in terms of the spectrum of the output  $\Phi_y^{(n)}(\omega)$ .

**Corollary 13** Consider Algorithm 1 applied to the system  $G$  in (1), where  $\lambda_e > 0$ . Then, the output spectrum has a unique fixed point  $\Phi_y^{(\infty)}(\omega)$  given by

$$\Phi_y^{(\infty)}(\omega) = \frac{\lambda_e \mu^2}{\mu^2 - |G(e^{j\omega})|^2},$$

where  $\mu > 0$  satisfies (9).

The result of Theorem 12 enables the derivation of the following properties of both  $\Phi_u^{(\infty)}$  and  $\mu$  as a function of  $\lambda_e$ .

**Theorem 14** Consider the iteration (7), where  $\lambda_e > 0$ ,  $\Phi_u^{(\infty)}$  satisfies the results in Theorem 12, and  $G \in \mathcal{H}_{\infty}(\bar{\mathbf{E}})$ . Then,

- (1)  $\mu > \|G\|_{\infty}$ .
- (2)  $\Phi_u^{(\infty)}(\omega)$  attains its (finite) maximum at the frequency 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  $|G(e^{j\bar{\omega}})|^2 = 0$  for some  $\bar{\omega}$ , then  $\Phi_u^{(\infty)}(\bar{\omega}) = \lambda_e / \mu^2$ .
- (4)  $\mu$  is a continuous and strictly increasing function of  $\lambda_e$ , such that  $\mu \rightarrow \|G\|_{\infty}$  as  $\lambda_e \rightarrow 0$ , and  $\mu \rightarrow \infty$  as  $\lambda_e \rightarrow \infty$ .

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

**Theorem 15** Let  $G \in \mathcal{H}_{\infty}(\bar{\mathbf{E}})$  and denote by  $\hat{\omega}$  the frequency in  $[0, \pi]$  at which  $|G(e^{j\omega})|^2$  is maximum. Also, let

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

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

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

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

Theorem 12 enables a qualitative analysis of  $\mu^2$  as a function of  $\lambda_e$ . For instance, in the case where the  $\mathcal{H}_{\infty}$ -norm is attained due to a dynamic phenomenon with low damping, then  $H_{\mu}$  is large and  $\mu^2$  significantly decreases as a function of  $\lambda_e$ .

### 3.2 Convergence to the Limit Spectrum

In the preceding analysis, it has been tacitly assumed that the iterative procedure (7) converges to  $\Phi_u^{(\infty)}$ . In this section, the convergence is thoroughly established and analyzed. This constitutes Contribution C3 of the paper.

The key issue in the convergence proof is the combined presence of additive noise, see (1), and the input normalization (3). To deal with these in the convergence of the sequence  $\{\Phi_u^{(n)}\}_{n \in \mathbb{N}}$ , consider the operator  $H : \mathcal{L}_{\infty}([-\pi, \pi], \mathbb{R}) \oplus \mathbb{R} \rightarrow \mathcal{L}_{\infty}([-\pi, \pi], \mathbb{R}) \oplus \mathbb{R}$  defined as

$$H \begin{bmatrix} \Phi \\ \alpha \end{bmatrix} = \begin{bmatrix} |G|^2 \Phi + \lambda_e \alpha \\ \frac{1}{2\pi} \int |G|^2 \Phi + \lambda_e \alpha \end{bmatrix}. \quad (13)$$

As a result, the iteration (7) can be interpreted as an application of  $H$ , since

$$H \begin{bmatrix} \Phi_u^{(n)} \\ 1 \end{bmatrix} = \begin{bmatrix} |G|^2 \Phi_u^{(n)} + \lambda_e \\ \frac{1}{2\pi} \int |G|^2 \Phi_u^{(n)} + \lambda_e \end{bmatrix}$$

$$= \left( \frac{1}{2\pi} \int |G|^2 \Phi_u^{(n)} + \lambda_e \right) \begin{bmatrix} \Phi_u^{(n+1)} \\ 1 \end{bmatrix},$$

and the fixed point  $\Phi_u^{(\infty)}$  of (7) satisfies the following eigenvalue equation of  $H$ :

$$H \begin{bmatrix} \Phi_u^{(\infty)} \\ 1 \end{bmatrix} = \left( \frac{1}{2\pi} \int |G|^2 \Phi_u^{(\infty)} + \lambda_e \right) \begin{bmatrix} \Phi_u^{(\infty)} \\ 1 \end{bmatrix} = \alpha \begin{bmatrix} \Phi_u^{(\infty)} \\ 1 \end{bmatrix}. \quad (14)$$

Therefore, the iteration (7) can be seen as the power iteration method applied to the operator  $H$  in the standard setting, which is noise-free, since the noise-term in (1) has been absorbed into  $H$ .

Standard results for convergence of the power iterations cannot be applied directly for the operator  $H$ . For example, while the analysis in [11, Section 7.3.1] requires finite dimensionality of the underlying operator, and the results in [10] require compactness, the operator  $H$  is infinite dimensional and non-compact. This is shown in the following theorem.

**Theorem 16** *Let  $\mathcal{L}_\infty([-\pi, \pi], \mathbb{R}) \oplus \mathbb{R}$  be endorsed with the norm  $\|[\Phi \ \alpha]^T\| := \max\{\sup|\Phi|, |\alpha|\}$ . Then, the operator  $H$  is not compact.*

**PROOF.** Take  $\lambda \in (\min|G|^2, \max|G|^2)$  and define, for every  $n \in \mathbb{N}$ ,  $\Phi_n(\omega) = 1$  for  $\omega \in (\omega^* - 1/n, \omega^* + 1/n)$ , where  $\omega^*$  is such that  $|G(e^{j\omega^*})|^2 = \lambda$ , and  $\Phi_n(\omega) = 0$  elsewhere. Then  $(H - \lambda I)[\Phi_n \ 0]^T \rightarrow 0$  as  $n \rightarrow \infty$ , even though  $\|[\Phi_n \ 0]^T\| = 1$ . This shows that  $(\min|G|^2, \max|G|^2)$  belongs to the spectrum of  $H$ , hence  $H$  is not compact [21, Proposition A.2].  $\square$

The pursued strategy to analyze (7) consists in viewing the operator  $H$  as a nonexpansive mapping in an appropriate metric space that is equipped with the so-called Hilbert projective metric, using the setup described in [5]. The need for using a special metric, instead of a standard one that is derived from a norm, stems from the fact that the iterations involve a normalization step. Indeed, the normalization step in general results in a more difficult analysis; on the other hand, the Hilbert metric is invariant under such a normalization, as is shown in Lemma 18 below.

In the sequel, the input spectrum  $\Phi_u^{(n)}$ , corresponding to the  $n$ -th iteration of (7), is associated with the vector  $[\Phi_n \ \alpha_n]^T$ , related to the  $n - 1$ -th application of the operator  $H$  to an element  $[\Phi_1 \ \alpha_1]^T$ , i.e.,

$$\Phi_u^{(n)} = \frac{\Phi_n}{\alpha_n} \sim \begin{bmatrix} \Phi_n \\ \alpha_n \end{bmatrix} = H^{n-1} \begin{bmatrix} \Phi_1 \\ \alpha_1 \end{bmatrix}.$$

To study the convergence of  $\{\Phi_u^{(n)}\}$ , first define the following positive cone in  $L := \mathcal{L}_\infty([-\pi, \pi], \mathbb{R}) \oplus \mathbb{R}$ :

$$L^+ := \left\{ \begin{bmatrix} \Phi \\ \alpha \end{bmatrix} \in L : \alpha \geq 0 \text{ and } \Phi(\omega) \geq 0, \omega \in [-\pi, \pi] \right\}.$$

The order relation defined by  $L^+$  turns  $L$  into an *ordered vector space*<sup>1</sup>. The Hilbert projective metric is now defined as follows, see also [9].

**Definition 17** *The Hilbert metric on  $L$  is the function  $d : L \times L \rightarrow \mathbb{R}_0^+$  satisfying*

$$d(x, y) := \ln \left( \frac{\inf\{\lambda \geq 0 : x \leq \lambda y\}}{\sup\{\lambda \geq 0 : \lambda y \leq x\}} \right), \quad x, y \in L.$$

The following lemma contains several properties of the Hilbert projective metric that are essential in the subsequent analysis.

**Lemma 18** *Let  $[\Phi_1 \ \alpha_1]^T, [\Phi_2 \ \alpha_2]^T \in \text{int } L^+$ , and  $c_1, c_2 > 0$ . The Hilbert metric on  $L$  has the following properties:*

(1)

$$d \left( \begin{bmatrix} \Phi_1 \\ \alpha_1 \end{bmatrix}, \begin{bmatrix} \Phi_2 \\ \alpha_2 \end{bmatrix} \right) = \ln \max \left\{ 1, \sup_{\omega} \frac{\Phi_1(\omega)/\alpha_1}{\Phi_2(\omega)/\alpha_2} \right\} + \ln \max \left\{ 1, \sup_{\omega} \frac{\Phi_2(\omega)/\alpha_2}{\Phi_1(\omega)/\alpha_1} \right\}.$$

$$(2) \quad d \left( c_1 \begin{bmatrix} \Phi_1 \\ \alpha_1 \end{bmatrix}, c_2 \begin{bmatrix} \Phi_2 \\ \alpha_2 \end{bmatrix} \right) = d \left( \begin{bmatrix} \Phi_1 \\ \alpha_1 \end{bmatrix}, \begin{bmatrix} \Phi_2 \\ \alpha_2 \end{bmatrix} \right).$$

$$(3) \quad \left\| \frac{\Phi_1}{\alpha_1} - \frac{\Phi_2}{\alpha_2} \right\|_{\infty} \leq (e^{\gamma} - 1)e^{\gamma} \frac{\sup \Phi_1}{\alpha_1}, \text{ where } \gamma = d \left( \begin{bmatrix} \Phi_1 \\ \alpha_1 \end{bmatrix}, \begin{bmatrix} \Phi_2 \\ \alpha_2 \end{bmatrix} \right).$$

Notice that property (2) of Lemma 18 establishes the invariance of the Hilbert metric with respect to the (individual) scaling of its arguments.

The Hilbert projective metric now enables the following result, which constitutes Contribution C3 of this paper. See Appendix A.5 for a proof of Theorem 19.

<sup>1</sup> An *ordered vector space*  $L$  [25] is a real vector space with an order relation  $\leq$  given by  $a \leq b \Leftrightarrow b - a \in L^+$ , where  $L^+$  is a proper convex cone in  $L$  (i.e.,  $L^+ + L^+ \subseteq L^+$ ,  $\lambda L^+ \subseteq L^+$  for every  $\lambda > 0$ , and  $L^+ \cap -L^+ = \{0\}$ ).

**Theorem 19** 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 (7) converges in the  $\mathcal{L}_\infty$  norm to  $\Phi_u^{(\infty)}$  as defined in (8).

**Remark 20** The assumption from Section 2 that  $e$  is white noise has been imposed mainly for simplicity of presentation. In fact, the proof of Theorem 19 can in principle be extended to the case where the spectrum of  $e$ , say,  $\Phi_e$ , belongs to  $\mathcal{L}_\infty([- \pi, \pi], \mathbb{R})$  and satisfies  $0 < a \leq \Phi_e(\omega) \leq b < \infty$  for all  $\omega \in [- \pi, \pi]$ , where  $a, b$  are constants. The lower bound,  $a$ , is needed to ensure that  $\Phi_u^{(\infty)} \in \mathcal{L}_\infty([- \pi, \pi], \mathbb{R})$ . This includes the case when  $e$  is filtered white noise, i.e., when the noise is of the form  $e_t = H(q)w_t$ , where  $w$  is white noise and  $H(q)$  is a stable and minimum phase rational transfer function.

**Remark 21** Theorem 19 does not provide a rate of convergence for  $\Phi_u^{(n)}$ . The reason is that the convergence of this spectrum is not monotonic in the  $\mathcal{L}_\infty$  norm nor in the Hilbert metric, because the cones  $C_{m,M}$  (defined in Appendix A.5) are not invariant under the operator  $H$ . This is in fact the main obstacle towards establishing the limit behavior of  $\Phi_u^{(n)}$ . On the other hand, from the proof of Theorem 19 we can derive the following asymptotic rate of convergence result: Given an  $\varepsilon > 0$  sufficiently small, there is an  $n_0 \in \mathbb{N}$  such that for all  $n \geq n_0$ ,

$$\|\Phi_u^{(n)} - \Phi_u^{(\infty)}\|_\infty \leq C \left[ \frac{\ln \left\{ \frac{\|G\|_\infty^2 (M+\varepsilon)^2 / (m-\varepsilon) + \lambda_e}{\|G\|_\infty^2 (M+\varepsilon) + \lambda_e} \right\}}{\ln \{(M+\varepsilon)/(m-\varepsilon)\}} \right]^n,$$

for some  $C > 0$ . Here,  $M$  and  $m$  correspond to the essential supremum and infimum of  $\Phi_u^{(\infty)}$ , respectively.

In the noise free case, when the number of samples  $N$  is kept fixed and finite, the method possesses an exponential rate of convergence in the number of iterations, since, as it is explained in the Introduction, it coincides with the standard power iterations method from numerical linear algebra, whose behavior under these conditions has already been studied (see e.g. [11, Section 8.2]). From the proof of Theorem 3, it also follows that the limit estimate (in the number of iterations  $n$ ) is monotonically non-decreasing in  $N$ , and it tends to  $\|G\|_\infty$  as  $N \rightarrow \infty$ . These nice properties, however, are lost in the presence of noise.

**Remark 22** The convergence result of Theorem 19 is with respect to the number of iterations, where it has been considered that the number of samples per iterations has tended to infinity. This means that Theorem 19 is essentially deterministic in nature (even though it considers the effect of noise in the observations). The convergence with respect to the number of samples, which is beyond the scope of the paper, has to be studied in a stochastic sense.

## 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 (6) are 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 norm estimate is analyzed. This analysis constitutes Contribution C4 of this paper. Throughout this section, the emphasis is on the estimator in (6), since this enables a comparison with the results in [27], where a similar estimator modulo the normalization is considered. In [27], 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 3 and the discussion in Remark 8.

However, the results in Theorem 14 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 [27] of the power iteration procedure in a stochastic framework is incomplete. The following result enables a more detailed analysis of the estimator  $\hat{\beta}_2^{(n)}$ .

**Lemma 23** Consider the estimator (6). Then, for  $N \rightarrow \infty$ ,

$$\mathbb{E} \left\{ [\hat{\beta}_2^{(n)}]^2 \right\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} |G(e^{j\omega})|^2 \Phi_u^{(n-1)}(\omega) d\omega. \quad (15)$$

**PROOF.** Observe that for  $N \rightarrow \infty$ , the expression for  $\hat{\beta}_2^{(n)}$  in (6) can be recast as a sample cross-spectrum. Hence, taking expectations,

$$\begin{aligned} \mathbb{E} \left\{ [\hat{\beta}_2^{(n)}]^2 \right\} &= \mu^{(n-1)} \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_{y^{(n)}, u^{(n-1)}}(\omega) d\omega \\ &= \mu^{(n-1)} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{\mu^{(n-1)}} |G(e^{j\omega})|^2 \Phi_u^{(n-1)}(\omega) d\omega, \end{aligned}$$

which equals (15).  $\square$

Lemma 23 in conjunction with Theorem 14 reveals several qualitative results with respect to the bias of the limit estimator  $\hat{\beta}_2^{(\infty)}$ . Indeed, note that for  $\lambda_e > 0$ ,  $\Phi_u^{(\infty)}$  is a smoothed Dirac delta function. Clearly, this implies that

$$\mathbb{E} \left\{ [\hat{\beta}_2^{(\infty)}]^2 \right\} < \|G\|_\infty^2 \quad \text{for } \lambda_e > 0, \quad (16)$$



hence the power iterations result in a biased estimate of the  $\mathcal{H}_\infty$ -norm if  $\lambda_e > 0$ . Similarly, observe that if  $\lambda_e \rightarrow 0$ , then  $\Phi_u^{(\infty)}$  tends to a Dirac delta function. From the sampling property of the Dirac delta function, and the fact that the variance of  $\hat{\beta}_2^{(n)}$  tends to zero as  $\lambda_e \rightarrow 0$ , it follows that

$$\hat{\beta}_2^{(\infty)} \rightarrow \|G\|_\infty \quad \text{in mean, as } \lambda_e \rightarrow 0,$$

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 24** *Consider the estimator (6). Then, for  $N \rightarrow \infty$ ,*

$$\mathbb{E} \left\{ [\hat{\beta}_2^{(\infty)}]^2 \right\} = \|G\|_\infty^2 - \lambda_e + \frac{2}{H_\mu} \lambda_e^2 + o(\lambda_e^2).$$

**PROOF.** Lemma 23 reveals that for  $n \rightarrow \infty$ ,

$$\mathbb{E} \left\{ [\hat{\beta}_2^{(\infty)}]^2 \right\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} |G(e^{j\omega})|^2 \Phi_u^{(\infty)}(\omega) d\omega. \quad (17)$$

Combining (17) and (11) yields

$$\mu^2 = \mathbb{E} \left\{ [\hat{\beta}_2^{(\infty)}]^2 \right\} + \lambda_e.$$

Next, rearranging and applying the result of Theorem 15 gives the following asymptotic expression for the bias:

$$\mathbb{E} \left\{ [\hat{\beta}_2^{(\infty)}]^2 \right\} = \mu^2 - \lambda_e = \|G\|_\infty^2 - \lambda_e + \frac{2}{H_\mu} \lambda_e^2 + o(\lambda_e^2),$$

which concludes the proof.  $\square$

Theorem 24 reveals that the asymptotic bias, i.e.,  $\mathbb{E} \left\{ [\hat{\beta}_2^{(\infty)}]^2 \right\} - \|G\|_\infty^2$ , is dominated by  $-\lambda_e$  in the small noise regime. Equivalently, by considering a Taylor expansion of the square root of  $[\hat{\beta}_2^{(\infty)}]^2$  around  $\lambda_e = 0$ , it follows that

$$\mathbb{E} \left\{ \hat{\beta}_2^{(\infty)} \right\} - \|G\|_\infty = -\frac{\lambda_e}{2\|G\|_\infty} + o(\lambda_e),$$

which corroborates the previous analysis.

Furthermore, Theorem 15 shows that the normalization factor,  $\mu^{(n)}$ , which as  $n, N \rightarrow \infty$  converges to  $\mu$ , might be a better estimator in terms of bias of  $\|G\|_\infty$  for small  $\lambda_e$ . Indeed, observe that  $\mu^{(n)}$  can be interpreted as a direct estimator for (2) by forming the product  $[y^{(n)}]^T y^{(n)}$ .

**Remark 25** *It is important to notice that estimators of the  $\mathcal{H}_\infty$  norm based on finite-dimensional parametric models can be consistent (and asymptotically unbiased) under mild assumptions. In addition, ‘sieve’ estimators (i.e., those based on a parametric model whose number of parameters increases slowly with  $N$ ) can also share these properties. However, unless the input being applied to the system is the ‘right’ one (that is, a sinusoid at the peak frequency, whose power is concentrated in the ‘worst case’ frequency), additional prior knowledge, in the form of a fixed parametric model or some constraint on the degree of smoothness of the frequency response of  $G$ , is needed in order to achieve consistent estimation of  $\|G\|_\infty$ ; this is because, roughly speaking, the estimation of the  $\mathcal{H}_\infty$  norm relies implicitly on the determination of the peak frequency, which is only possible by using the prior knowledge in order to ‘interpolate’ the shape of  $G$ .*

*The essence of Algorithm 1 lies in the iterative redesign of the input signal being applied to the system. In fact, Algorithm 1 determines an input signal whose spectrum as  $n \rightarrow \infty$  tends to one very similar to the optimal input (a sinusoid with frequency at the peak of the system gain), thus delivering a low variance estimate (in comparison to estimators based on an arbitrarily selected input). Thanks to this input design mechanism, Algorithm 1 does not require an explicit parametric model, it is very simple to implement, and, as we have shown, in low noise scenarios it delivers a reasonable estimate.*

*On the other hand, in order to overcome the bias problem, it would be interesting to extend Algorithm 1 to make more efficient use of the data from all previous iterations, using for example stochastic approximation techniques. This, however, is beyond the scope of the present paper.*

#### 4.2 Fisher Information Matrix per Iteration

The value of iterations is immediate if the nonparametric estimator (6) is used in a noise-less environment, since Theorem 3 reveals that for an increasing number of iterations, the estimate converges to the  $\mathcal{H}_\infty$ -norm for  $\lambda_e = 0$ . However, the value of power iterations has not yet been investigated in the general case where possibly another estimator is used. Hereto, the limit of accuracy of the power iterations method is investigated through the asymptotic information matrix, which provides an analysis that is independent of the specific estimator. This leads to Contribution C5 of this paper.

To perform the analysis, an underlying parametric model is considered. Specializing to the prediction error framework, let  $G(z, \theta)$  be a parametric model structure. Under mild conditions, see [19, Chapter 9] and the assumption that there exists a *true parameter*  $\theta_o$  such that  $G(z, \theta_o) = G$ , where  $G$  denotes the true system, the result

$$\sqrt{N}(\hat{\theta}_N - \theta_o) \xrightarrow{d} \mathcal{N}(0, P_\theta), \quad N \rightarrow \infty$$

holds. The prediction error estimator turns out to be asymptotically efficient, i.e., the asymptotic covariance matrix  $P_\theta$  equals the inverse of the (Fisher) information matrix, which is given by

$$P_\theta = I_\theta^{-1} = (\mathbf{E} \{ \psi_t \Lambda^{-1} \psi_t^T \})^{-1}, \quad (18)$$

where, using  $\varepsilon_t = y_t - \hat{y}_{t|t-1}$ ,

$$\psi_t = - \left. \frac{\partial \varepsilon_t^T}{\partial \theta} \right|_{\theta=\theta_o}. \quad (19)$$

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

**Lemma 26** *Consider system (1) and the power iterations algorithm 1. Then,*

$$I_\theta = \sum_{k=1}^n I_\theta^{(k)}, \quad (20)$$

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 the definition of  $\varepsilon_t$ ,

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

Clearly,  $\frac{\partial \varepsilon_t^{(k)}}{\partial \theta} = -G'(q)u_t^{(k)}$ . Next, using (19),

$$\psi_t = \left[ \frac{\partial \varepsilon_t^{(1)}}{\partial \theta} \quad \frac{\partial \varepsilon_t^{(2)}}{\partial \theta} \quad \dots \quad \frac{\partial \varepsilon_t^{(k)}}{\partial \theta} \right] = -G'(q)U_t,$$

where  $U_t := [u_t^{(1)} \quad u_t^{(2)} \quad \dots \quad u_t^{(k)}]$ . Next, using (18),

$$I_\theta = \mathbf{E} \{ \psi_t \Lambda^{-1} \psi_t^T \} = \frac{1}{\lambda_e} \mathbf{E} \{ \psi_t \psi_t^T \}.$$

Subsequently applying Parseval's relation yields

$$I_\theta = \frac{1}{2\pi\lambda_e} \int_{-\pi}^{\pi} G'(e^{j\omega}) \Phi_U(\omega) (G'(e^{-j\omega}))^T d\omega.$$

Note that here [19, equation (2.63)]

$$\Phi_U(\omega) = \sum_{\tau=-\infty}^{\infty} R_U(\tau) e^{-j\tau\omega}, \quad (21)$$

where [19, Page 34]

$$\begin{aligned} R_U(\tau) &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N U_t U_{t-\tau} \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N \sum_{k=1}^n u_t^{(k)} u_{t-\tau}^{(k)} \\ &= \sum_{k=1}^n R_{u^{(k)}}(\tau). \end{aligned} \quad (22)$$

Next, combining (21) and (22) yields (20).  $\square$

Several interesting observations can be made with respect to Lemma 26. Firstly, it is observed that the Fisher information matrix for iterative schemes satisfies an additivity property, see (20), i.e., additional experiments can only increase information about the system. This is consistent with Fisher's original requirements for the definition of information [22, Page 59]. Secondly, due to the uniform convergence of  $\Phi_u^{(n)}$ , (20) divided by  $n$  corresponds to a Cesàro sum, hence in the asymptotic case where the number of experiments  $n \rightarrow \infty$  it holds that

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

If  $\lambda_e \rightarrow 0$ , then  $\Phi_u^{(\infty)}$  tends to a Dirac delta function at the peak frequency of  $|G(e^{j\omega})|$ . In [28] and [29], it is shown that for this input signal, among all signals of fixed power, the maximum likelihood (or PEM) estimator of  $\|G\|_\infty$ , for any parametric model structure, has the smallest possible asymptotic variance. In this respect, the power iterations method can be interpreted as an effective approach to iterative experiment design, as it generates a sequences of input signals whose spectra is, in the limit  $n \rightarrow \infty$ , close to the optimal input signal for  $\mathcal{H}_\infty$  norm estimation. Furthermore, notice that, due to the interpretation of  $\hat{\beta}^{(n)}$  as the estimator for one-parameter model from Remark 10, it can be concluded that for low noise scenarios ( $\lambda_e$  small), Algorithm 1 delivers a nearly unbiased estimator whose accuracy is close to the Cramér-Rao bound for large  $n$  and  $N$ .

## 5 Example

In this section, the results of the paper are illustrated in a simulation example. 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.

First, the results of Theorem 14 are illustrated. Hereto, use is made of the eigenvalue equation of the system  $H$ , see (13). Discretization of  $\omega$  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

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

is computed. The number of data points used in each iteration is  $N = 100$ . The resulting spectrum for  $\lambda_e \in \{1, 4, 9\}$  (using  $n_\omega = 1000$ ) is depicted in Fig. 3. Considering Fig. 2 in conjunction with Fig. 3 confirms Properties (1)-(4) in Theorem 14.

Next, to illustrate the behavior in the time domain and analyze the bias, the estimate  $\hat{\beta}_2^{(n)}$  is computed for a realization of the iterative algorithm 1 with  $\lambda_e = 9$ , see Fig. 4. Due to the presence of  $e_t$  in (1), the estimate  $\hat{\beta}_2^{(n)}$  is noisy as is expected. To analyze the bias, 1000 realizations of the iterative procedure are averaged, see Fig. 4. Clearly, the mean of  $\hat{\beta}_2^{(n)}$  over different realizations of the iterative algorithm is strictly below the finite time  $\ell_2$ -induced norm of  $G$ , i.e.,  $\sqrt{\lambda_{\max}(\mathbf{G}^T \mathbf{G})}$ . This confirms the result (16).

Finally, 1000 realizations of the iterative procedure are averaged for  $\lambda_e \in \{1, 4, 9\}$ , see Fig. 5 for the results. It can be observed that the bias increases for increasing  $\lambda_e$ . From Lemma 23, this can be understood when considering the smoothness of  $G$  in Fig. 2 in conjunction with the spectra in Fig. 3.

Notice that for  $N = 100$ , the finite time  $\ell_2$ -induced norm of  $G$  is 2.3202, which is noticeably smaller than the infinite time  $\ell_2$ -induced norm, 2.4116. This means that  $N = 100$  is a relatively small value for this lightly damped system. The conclusions of the previous sections, however, seem to apply even for such a small  $N$ .

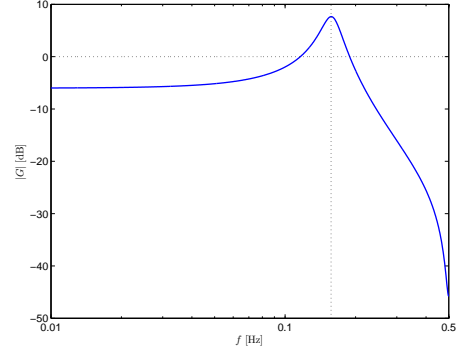


Fig. 2. Bode magnitude diagram of  $G$ .

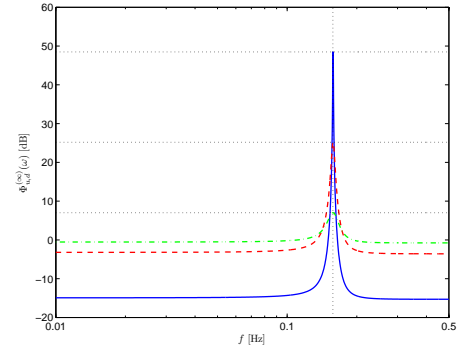


Fig. 3. Limit spectrum  $\Phi_{u,d}^{(\infty)}(\omega)$ ,  $\lambda_e = 1$  (solid blue),  $\lambda_e = 4$  (dashed red),  $\lambda_e = 9$  (dash-dotted green).

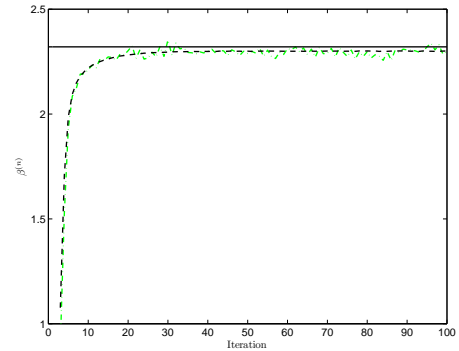


Fig. 4. Estimated norm during power iterations: single realization of the iterative procedure (dash-dotted green), average over 1000 realizations of the power iterations (dashed black), finite time  $\ell_2$ -induced norm (solid black).

## 6 Conclusion

The results presented in this paper contribute to the analysis of the role of iterations in system identification for control. In this paper, an approach for nonparametric  $\mathcal{H}_\infty$ -norm estimation is presented that requires only one experiment for the estimation procedure. In addition, it is shown that for nonparametric  $\mathcal{H}_\infty$ -norm estimation through iterative experiments, i) additive disturbances can introduce bias errors, and ii) iterative procedures can be interpreted as experiment design for  $\mathcal{H}_\infty$ -norm estimation, and the value of iterations has been

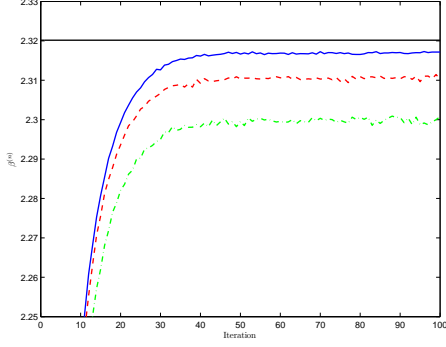


Fig. 5. Estimated norm during power iterations averaged over 1000 realizations,  $\lambda_e = 1$  (solid blue),  $\lambda_e = 4$  (dashed red),  $\lambda_e = 9$  (dash-dotted green). The finite time  $\ell_2$ -induced norm is also shown (solid black).

investigated by means of the Fisher information matrix. The analysis is based on a frequency domain approach, the novelty of which is that it addresses both i) additive stochastic disturbances that represent measurement errors, and ii) the normalization of the input signal to account for input power constraints, which involves a non-linear operation. In addition, convergence of the spectra is established through the use of the Hilbert projective metric.

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 (6) are being investigated to avoid these bias errors, e.g., by using the data from old experiments through stochastic approximation algorithms. Secondly, extensions of the power iteration algorithm are being investigated that enable the nonparametric estimation of other system properties.

## A Proofs

### A.1 Proof of Theorem 3

Firstly, by symmetry of  $G_H$ ,

$$\lambda_{\max}(G_H) = \left\| \begin{bmatrix} g_{N-1} & g_{N-2} & \cdots & g_0 \\ g_{N-2} & g_{N-3} & & 0 \\ \vdots & & \ddots & \\ g_0 & 0 & \cdots & 0 \end{bmatrix} \right\|,$$

where the latter expression equals the Hankel norm of the system  $\tilde{G}_N(z) := g_{N-1} + g_{N-2}z^{-1} + \cdots + g_0z^{-N+1}$ , i.e.,  $\|\tilde{G}_N\|_H$ . Next, by Nehari's theorem [21],

$$\|\tilde{G}_N\|_H \quad (\text{A.1})$$

$$\begin{aligned} &= \inf_{Q \in \mathcal{H}_{\infty}(\mathbf{E})} \|\tilde{G}_N + Q\|_{\infty} \\ &= \inf_{Q \in \mathcal{H}_{\infty}(\mathbf{E})} \|g_{N-1} + g_{N-2}z^{-1} + \cdots + g_0z^{-N+1} + Q(z)\|_{\infty} \\ &= \inf_{Q \in \mathcal{H}_{\infty}(\mathbf{E})} \|g_0 + g_1z + \cdots + g_{N-1}z^{N-1} + z^{N-1}Q(z)\|_{\infty} \\ &= \inf_{Q \in \mathcal{H}_{\infty}(\mathbf{E})} \|g_0 + g_1z^{-1} + \cdots + g_{N-1}z^{-N+1} + z^{-N}Q(z)\|_{\infty} \\ &= \inf_{Q \in \mathcal{H}_{\infty}(\mathbf{E})} \|G(z) + z^{-N}Q(z)\|_{\infty}, \end{aligned}$$

where the infimum is attained for some  $Q = Q_N^{opt}$ .

The last expression of (A.1) reveals that  $\|\tilde{G}_N\|_H$  is monotonically non-decreasing and that  $\|\tilde{G}_N\|_H \leq \|G\|_{\infty}$  for every  $N$ . Therefore,  $\lim_{N \rightarrow \infty} \|G_H\| = \lim_{N \rightarrow \infty} \|\tilde{G}_N\|_H$  exists, and it satisfies  $\lim_{N \rightarrow \infty} \|G_H\| \leq \|G\|_{\infty}$ .

To conclude the proof, it is shown that  $\lim_{N \rightarrow \infty} \|G_H\| < \|G\|_{\infty}$  leads to a contradiction.

Since the infimum in the last expression of (A.1) is always attained, let  $H_N$  denote the expression inside the norm for which the minimum norm is attained, i.e.,  $H_N := G(z) + z^{-N}Q_N^{opt}(z)$ . Since  $H_N \in \mathcal{H}_{\infty}(\mathbf{E})$  and  $\|H_N\|_{\infty} \leq \|G\|_{\infty}$ , it is concluded that  $\{H_N\}_{N \in \mathbb{N}}$  is a *normal family* [24, Theorem 14.6], which means that there is a subsequence  $\{H_{N_k}\}_{k \in \mathbb{N}}$  uniformly convergent on compact subsets of  $\mathbf{E}$ . Denote by  $H_{\infty}$  the limit of  $H_{N_k}$  as  $k \rightarrow \infty$ . Notice that  $H_{\infty} \in \mathcal{H}_{\infty}(\mathbf{E})$ , hence it has a Laurent expansion of the form  $H_{\infty}(z) = h_0 + h_1z^{-1} + h_2z^{-2} + \cdots$ . Due the uniform convergence of the subsequence, it follows that, for an arbitrary  $\epsilon > 0$ ,

$$\begin{aligned} h_i &= \oint_{|z|=1+\epsilon} H_{\infty}(z)z^{i-1}dz \\ &= \oint_{|z|=1+\epsilon} \left[ \lim_{k \rightarrow \infty} H_{N_k}(z) \right] z^{i-1}dz \\ &= \lim_{k \rightarrow \infty} \oint_{|z|=1+\epsilon} H_{N_k}(z)z^{i-1}dz \\ &= \lim_{k \rightarrow \infty} \oint_{|z|=1+\epsilon} [G(z) + z^{-N_k}Q_{N_k}^{opt}(z)]z^{i-1}dz \\ &= \oint_{|z|=1+\epsilon} G(z)z^{i-1}dz \\ &= g_i, \quad i \in \mathbb{N}. \end{aligned}$$

This means that  $H_{\infty} = G$  almost everywhere in  $\mathbf{E}$ , which leads to a contradiction:  $\|G\|_{\infty} = \|H_{\infty}\|_{\infty} = \lim_{N \rightarrow \infty} \|H_N\|_{\infty} = \lim_{N \rightarrow \infty} \|G_H\| < \|G\|_{\infty}$ . This proves the theorem.  $\square$

### A.2 Proof of Theorem 14

To establish Property 1, observe that  $\Phi_u^{(\infty)} \geq 0$  and (8) imply that  $\mu \geq |G(e^{j\omega})|$  for all  $\omega \in (-\pi, \pi]$ . By virtue of stability of  $G$ , this implies  $\mu \geq \|G\|_{\infty}$ . Moreover, since

$G \in \mathcal{H}(\bar{\mathbf{E}})$ , it follows that  $\mu > \|G\|_\infty$ , since otherwise the integral (9) becomes divergent (as  $|G(e^{j\omega})|^2 = \|G\|_\infty^2 + O(\omega - \hat{\omega})^2$ , where  $\hat{\omega} \in [-\pi, \pi]$  is such that  $|G(e^{j\hat{\omega}})| = \|G\|_\infty$ ), which would imply that  $\lambda_e = 0$ .

Next, Property 2 and Property 3 follow from (8) and  $\mu^2 > |G(e^{j\omega})|^2$  for all  $\omega \in (-\pi, \pi]$ .

To prove Property 4, observe that (9) constitutes a mapping from  $\mu \in (\|G\|_\infty, \infty)$  into  $\lambda_e \in (0, \infty)$ . By virtue of [4, Corollary 5.8], the continuity property is established. In addition, due to [4, Corollary 5.9], the mapping is differentiable. Hence, the mapping from  $\mu$  into  $\lambda_e$  is strictly increasing, since

$$\begin{aligned} \frac{d\lambda_e}{d\mu^2} &= \lim_{\Delta \rightarrow 0} \frac{2\pi}{\Delta \int_{-\pi}^{\pi} \frac{d\omega}{\mu^2 + \Delta - |G(e^{j\omega})|^2}} - \frac{2\pi}{\Delta \int_{-\pi}^{\pi} \frac{d\omega}{\mu^2 - |G(e^{j\omega})|^2}} \\ &= 2\pi \lim_{\Delta \rightarrow 0} \frac{\int_{-\pi}^{\pi} \frac{d\omega}{\mu^2 - |G(e^{j\omega})|^2} - \int_{-\pi}^{\pi} \frac{d\omega}{\mu^2 + \Delta - |G(e^{j\omega})|^2}}{\Delta \int_{-\pi}^{\pi} \frac{d\omega}{\mu^2 + \Delta - |G(e^{j\omega})|^2} \int_{-\pi}^{\pi} \frac{d\omega}{\mu^2 - |G(e^{j\omega})|^2}} \\ &\geq \lim_{\Delta \rightarrow 0} \frac{\int_{-\pi}^{\pi} \left[ \frac{1}{\mu^2 - |G(e^{j\omega})|^2} - \frac{1}{\mu^2 + \Delta - |G(e^{j\omega})|^2} \right] d\omega}{2\pi \Delta \frac{1}{(\mu^2 + \Delta - \|G\|_\infty^2)(\mu^2 - \|G\|_\infty^2)}} \\ &= \lim_{\Delta \rightarrow 0} \frac{\int_{-\pi}^{\pi} \frac{d\omega}{(\mu^2 - |G(e^{j\omega})|^2)(\mu^2 + \Delta - |G(e^{j\omega})|^2)}}{2\pi \frac{1}{(\mu^2 + \Delta - \|G\|_\infty^2)(\mu^2 - \|G\|_\infty^2)}} \\ &\geq \left[ \frac{\mu^2 - \|G\|_\infty^2}{\mu^2} \right]^2 > 0. \end{aligned}$$

Next, by the inverse function theorem, see [23, Theorem 9.24], the mapping  $\lambda_e$  into  $\mu$  also is strictly increasing. Finally, to show the limiting cases, note that for  $\mu = \|G\|_\infty$  the integral in (9) is divergent (as  $|G(e^{j\omega})|^2 = \|G\|_\infty^2 + O(\omega - \hat{\omega})^2$ ), and for  $\mu \rightarrow \infty$  it thus holds that  $\lambda_e \rightarrow \infty$ .  $\square$

### A.3 Proof of Theorem 15

First, (12) is rewritten as

$$\pi = \int_0^\pi \frac{\lambda_e d\omega}{\mu^2(\lambda_e) - |G(e^{j\omega})|^2}.$$

Now, let  $\delta \in (0, \|G\|_\infty^2)$ , hence  $\pi = I_1(\lambda_e) + I_2(\lambda_e)$ , where

$$\begin{aligned} I_1(\lambda_e) &= \int_{\{\omega \in [0, \pi] : |G(e^{j\omega})|^2 < \|G\|_\infty^2 - \delta\}} \frac{\lambda_e d\omega}{\mu^2(\lambda_e) - |G(e^{j\omega})|^2} \\ I_2(\lambda_e) &= \int_{\{\omega \in [0, \pi] : |G(e^{j\omega})|^2 \geq \|G\|_\infty^2 - \delta\}} \frac{\lambda_e d\omega}{\mu^2(\lambda_e) - |G(e^{j\omega})|^2}. \end{aligned}$$

With respect to  $I_1(\lambda_e)$ , the bound  $0 \leq I_1(\lambda_e) \leq \pi \frac{\lambda_e}{\delta}$ , holds. This lower bound of  $I_1(\lambda_e)$  comes from Property 1 of Theorem 14. The upper bound, on the other hand, can be established as follows:

$$\begin{aligned} &\int_{\{\omega \in [0, \pi] : |G(e^{j\omega})|^2 < \|G\|_\infty^2 - \delta\}} \frac{\lambda_e d\omega}{\mu^2(\lambda_e) - |G(e^{j\omega})|^2} \\ &\leq \lambda_e \int_{\{\omega \in [0, \pi] : |G(e^{j\omega})|^2 < \|G\|_\infty^2 - \delta\}} \frac{d\omega}{\mu^2(\lambda_e) - \|G\|_\infty^2 + \delta} \\ &\leq \lambda_e \int_{\{\omega \in [0, \pi] : |G(e^{j\omega})|^2 < \|G\|_\infty^2 - \delta\}} \frac{d\omega}{\delta} \leq \frac{\lambda_e}{\delta} \pi, \end{aligned}$$

where Property 1 of Theorem 14 has been used again. Hence,  $I_1(\lambda_e) \rightarrow 0$  if  $\delta$  is chosen such that  $\lambda_e/\delta = o(1)$ . Take for this purpose  $\delta = \sqrt{\lambda_e}$  (which is possible if  $\lambda_e$  is small enough). To obtain an approximation for  $I_2(\lambda_e)$ , first take  $\lambda_e$  small enough such that  $\{\omega \in [0, \pi] : |G(e^{j\omega})|^2 \geq \|G\|_\infty^2 - \delta\}$  is an interval, whose extremes satisfy

$$\delta = \|G\|_\infty^2 - |G(e^{j\omega})|^2 = \frac{1}{2} H_\mu (\omega - \hat{\omega})^2 + O([\omega - \hat{\omega}]^3),$$

or, equivalently,  $|\omega - \hat{\omega}| = \sqrt{\frac{2\delta}{H_\mu}} + O(\delta^{3/2})$ . Therefore, if  $\tilde{\mu}^2(\lambda_e) := \mu^2(\lambda_e) - \|G\|_\infty^2$ , then

$$\begin{aligned} I_2(\lambda_e) &= \int_{\hat{\omega} - \sqrt{\frac{2\delta}{H_\mu} + O(\delta^{3/2})}}^{\hat{\omega} + \sqrt{\frac{2\delta}{H_\mu} + O(\delta^{3/2})}} \frac{\lambda_e d\omega}{\mu^2(\lambda_e) - |G(e^{j\omega})|^2} \\ &= \int_{\hat{\omega} - \sqrt{\frac{2\delta}{H_\mu} + O(\delta^{3/2})}}^{\hat{\omega} + \sqrt{\frac{2\delta}{H_\mu} + O(\delta^{3/2})}} \frac{\lambda_e d\omega}{\tilde{\mu}^2(\lambda_e) + \frac{1}{2} H_\mu (\omega - \hat{\omega})^2 + O([\omega - \hat{\omega}]^3)} \\ &= \lambda_e \int_{-\sqrt{\frac{2\delta}{H_\mu} + O(\delta^{3/2})}}^{\sqrt{\frac{2\delta}{H_\mu} + O(\delta^{3/2})}} \left[ \frac{1}{\tilde{\mu}^2(\lambda_e) + \frac{1}{2} H_\mu \omega^2} + O(\omega^3) \right] d\omega \\ &= \lambda_e \int_{-\sqrt{\frac{2\delta}{H_\mu} + O(\delta^{3/2})}}^{\sqrt{\frac{2\delta}{H_\mu} + O(\delta^{3/2})}} \frac{d\omega}{\tilde{\mu}^2(\lambda_e) + \frac{1}{2} H_\mu \omega^2} + O(\lambda_e \delta^2) \\ &= \frac{\lambda_e}{\tilde{\mu}^2(\lambda_e)} \int_{-\sqrt{\frac{2\delta}{H_\mu} + O(\delta^{3/2})}}^{\sqrt{\frac{2\delta}{H_\mu} + O(\delta^{3/2})}} \frac{d\omega}{1 + \frac{H_\mu}{2\tilde{\mu}^2(\lambda_e)} \omega^2} + O(\lambda_e \delta^2) \\ &= \sqrt{\frac{2\lambda_e^2}{H_\mu \tilde{\mu}^2(\lambda_e)}} \int_{\sqrt{\frac{H_\mu}{2\tilde{\mu}^2(\lambda_e)}} \left\{ \sqrt{\frac{2\delta}{H_\mu} + O(\delta^{3/2})} \right\}}^{\sqrt{\frac{H_\mu}{2\tilde{\mu}^2(\lambda_e)}} \left\{ -\sqrt{\frac{2\delta}{H_\mu} + O(\delta^{3/2})} \right\}} \frac{d\omega}{1 + \omega^2} + O(\lambda_e \delta^2) \\ &= \sqrt{\frac{2\lambda_e^2}{H_\mu \tilde{\mu}^2(\lambda_e)}} \left[ \arctan \left( \sqrt{\frac{\delta}{\tilde{\mu}^2(\lambda_e)}} + O \left( \sqrt{\frac{\delta^3}{\tilde{\mu}^2(\lambda_e)}} \right) \right) \right] \end{aligned}$$

$$\begin{aligned}
& -\arctan\left(-\sqrt{\frac{\delta}{\tilde{\mu}^2(\lambda_e)}} + O\left(\sqrt{\frac{\delta^3}{\tilde{\mu}^2(\lambda_e)}}\right)\right) + O(\lambda_e \delta^2) \\
& -\arctan\left(-\sqrt{\frac{\lambda_e^{1/2}}{\tilde{\mu}^2(\lambda_e)}} + O\left(\sqrt{\frac{\lambda_e^{3/2}}{\tilde{\mu}^2(\lambda_e)}}\right)\right) + O(\lambda_e^2).
\end{aligned}$$

Now, since  $\arctan$  is a bounded function, and  $I_1(\lambda_e) + I_2(\lambda_e) = \pi$ , where  $I_1(\lambda_e) \rightarrow 0$  as  $\lambda_e \rightarrow 0$  and the same applies for the  $O(\lambda_e^2)$  term of  $I_2(\lambda_e)$ , then for all  $\lambda_e$  in a neighborhood of 0, the bound

$$\sqrt{\frac{\lambda_e^2}{\tilde{\mu}^2(\lambda_e)}} \geq M > 0 \quad (\text{A.2})$$

should hold for some constant  $M$ . This means that  $\tilde{\mu}^2(\lambda_e) = O(\lambda_e^2)$ , or, equivalently,  $\mu^2(\lambda_e) = \|G\|_\infty^2 + O(\lambda_e^2)$ . In addition, using (A.2) reveals that

$$\sqrt{\frac{\lambda_e^{1/2}}{\tilde{\mu}^2(\lambda_e)}} = \lambda_e^{-3/4} \sqrt{\frac{\lambda_e^2}{\tilde{\mu}^2(\lambda_e)}} \xrightarrow{\lambda_e \rightarrow 0} \infty.$$

Hence, by considering the limit  $\lambda_e \rightarrow 0$ , the equality  $I_1(\lambda_e) + I_2(\lambda_e) = \pi$  implies that

$$\begin{aligned}
\pi &= \lim_{\lambda_e \rightarrow 0} \sqrt{\frac{2\lambda_e^2}{H_\mu \tilde{\mu}^2(\lambda_e)}} \left[ \arctan\left(\sqrt{\frac{\lambda_e^{1/2}}{\tilde{\mu}^2(\lambda_e)}} + O\left(\sqrt{\frac{\lambda_e^{3/2}}{\tilde{\mu}^2(\lambda_e)}}\right)\right) \right. \\
&\quad \left. - \arctan\left(-\sqrt{\frac{\lambda_e^{1/2}}{\tilde{\mu}^2(\lambda_e)}} + O\left(\sqrt{\frac{\lambda_e^{3/2}}{\tilde{\mu}^2(\lambda_e)}}\right)\right) \right] \\
&= \lim_{\lambda_e \rightarrow 0} \pi \sqrt{\frac{2\lambda_e^2}{H_\mu \tilde{\mu}^2(\lambda_e)}},
\end{aligned}$$

thus

$$\tilde{\mu}^2(\lambda_e) = \frac{2\lambda_e^2}{H_\mu} + o(\lambda_e^2),$$

which is the desired result.  $\square$

#### A.4 Proof of Lemma 18

(1) Notice that

$$\begin{bmatrix} \Phi_1 \\ \alpha_1 \end{bmatrix} \leq \beta \begin{bmatrix} \Phi_2 \\ \alpha_2 \end{bmatrix} \Leftrightarrow \beta \geq \max\left\{\frac{\alpha_1}{\alpha_2}, \sup_\omega \frac{\Phi_1(\omega)}{\Phi_2(\omega)}\right\}$$

and

$$\beta \begin{bmatrix} \Phi_2 \\ \alpha_2 \end{bmatrix} \leq \begin{bmatrix} \Phi_1 \\ \alpha_1 \end{bmatrix} \Leftrightarrow \beta \leq \min\left\{\frac{\alpha_1}{\alpha_2}, \inf_\omega \frac{\Phi_1(\omega)}{\Phi_2(\omega)}\right\},$$

$$\begin{aligned}
&= \sqrt{\frac{2\lambda_e^2}{H_\mu \tilde{\mu}^2(\lambda_e)}} \left[ \arctan\left(\sqrt{\frac{\lambda_e^{1/2}}{\tilde{\mu}^2(\lambda_e)}} + O\left(\sqrt{\frac{\lambda_e^{3/2}}{\tilde{\mu}^2(\lambda_e)}}\right)\right) \right. \\
&\quad \left. d\left(\begin{bmatrix} \Phi_1 \\ \alpha_1 \end{bmatrix}, \begin{bmatrix} \Phi_2 \\ \alpha_2 \end{bmatrix}\right) = \ln \frac{\max\left\{\frac{\alpha_1}{\alpha_2}, \sup_\omega \frac{\Phi_1(\omega)}{\Phi_2(\omega)}\right\}}{\min\left\{\frac{\alpha_1}{\alpha_2}, \inf_\omega \frac{\Phi_1(\omega)}{\Phi_2(\omega)}\right\}} \right. \\
&= \ln \frac{\frac{\alpha_1}{\alpha_2} \max\left\{1, \sup_\omega \frac{\Phi_1(\omega)/\alpha_1}{\Phi_2(\omega)/\alpha_2}\right\}}{\frac{\alpha_1}{\alpha_2} \min\left\{1, \inf_\omega \frac{\Phi_1(\omega)/\alpha_1}{\Phi_2(\omega)/\alpha_2}\right\}} \\
&= \ln \max\left\{1, \sup_\omega \frac{\Phi_1(\omega)/\alpha_1}{\Phi_2(\omega)/\alpha_2}\right\} \\
&\quad \left. + \ln \max\left\{1, \sup_\omega \frac{\Phi_2(\omega)/\alpha_2}{\Phi_1(\omega)/\alpha_1}\right\}\right].
\end{aligned}$$

(2) This is immediate from property (1).

(3) By property (1),

$$\begin{aligned}
&\ln \max\left\{1, \sup_\omega \frac{\Phi_1(\omega)/\alpha_1}{\Phi_2(\omega)/\alpha_2}\right\} \\
&\quad + \ln \max\left\{1, \sup_\omega \frac{\Phi_2(\omega)/\alpha_2}{\Phi_1(\omega)/\alpha_1}\right\} = \gamma \\
&\Rightarrow \sup_\omega \frac{\Phi_1(\omega)/\alpha_1}{\Phi_2(\omega)/\alpha_2} \leq e^\gamma \text{ and } \sup_\omega \frac{\Phi_2(\omega)/\alpha_2}{\Phi_1(\omega)/\alpha_1} \leq e^\gamma \\
&\Rightarrow \begin{cases} \frac{\Phi_1(\omega)}{\alpha_1} - \frac{\Phi_2(\omega)}{\alpha_2} \leq (e^\gamma - 1) \frac{\Phi_2(\omega)}{\alpha_2}, \\ \frac{\Phi_2(\omega)}{\alpha_2} - \frac{\Phi_1(\omega)}{\alpha_1} \leq (e^\gamma - 1) \frac{\Phi_1(\omega)}{\alpha_1}, & \omega \in [-\pi, \pi] \\ \frac{\Phi_2(\omega)}{\alpha_2} \leq e^\gamma \frac{\Phi_1(\omega)}{\alpha_1}, \end{cases} \\
&\Rightarrow \begin{cases} \frac{\Phi_1(\omega)}{\alpha_1} - \frac{\Phi_2(\omega)}{\alpha_2} \leq (e^\gamma - 1) \frac{\sup \Phi_2}{\alpha_2}, \\ \frac{\Phi_2(\omega)}{\alpha_2} - \frac{\Phi_1(\omega)}{\alpha_1} \leq (e^\gamma - 1) \frac{\sup \Phi_1}{\alpha_1}, & \omega \in [-\pi, \pi]. \\ \frac{\sup \Phi_2}{\alpha_2} \leq e^\gamma \frac{\sup \Phi_1}{\alpha_1}, \end{cases}
\end{aligned}$$

These expressions imply that

$$\begin{aligned}
\left\| \frac{\Phi_1}{\alpha_1} - \frac{\Phi_2}{\alpha_2} \right\|_\infty &\leq (e^\gamma - 1) \max\left\{\frac{\sup \Phi_1}{\alpha_1}, \frac{\sup \Phi_2}{\alpha_2}\right\} \\
&\leq (e^\gamma - 1) \max\left\{\frac{\sup \Phi_1}{\alpha_1}, e^\gamma \frac{\sup \Phi_1}{\alpha_1}\right\} \\
&= (e^\gamma - 1) e^\gamma \frac{\sup \Phi_1}{\alpha_1}.
\end{aligned}$$

$\square$

#### A.5 Proof of Theorem 19

The proof of Theorem 19 is split into several smaller steps to facilitate the exposition. The proofs of the lemmas are given in Appendix B.

The following simple technical result will be used in the proof of several of the intermediate lemmas required to establish Theorem 19.

**Lemma 27** Let  $a, b, c, d > 0$ . Then,

$$\frac{a+b}{c+d} \leq \max\left(\frac{a}{c}, \frac{b}{d}\right). \quad (\text{A.3})$$

Moreover, if  $a/c \neq b/d$ , then (A.3) is a strict inequality.

Firstly, the results of Lemma 28 enable the derivation of the contraction ratio of  $H$ .

**Lemma 28** Let  $[\Phi_1 \ \alpha_1]^T, [\Phi_2 \ \alpha_2]^T \in \text{int } L^+$ . Then,

$$\frac{d\left(H \begin{bmatrix} \Phi_1 \\ \alpha_1 \end{bmatrix}, H \begin{bmatrix} \Phi_2 \\ \alpha_2 \end{bmatrix}\right)}{d\left(\begin{bmatrix} \Phi_1 \\ \alpha_1 \end{bmatrix}, \begin{bmatrix} \Phi_2 \\ \alpha_2 \end{bmatrix}\right)} = \quad (\text{A.4})$$

$$\frac{\ln \sup_{\omega} \frac{|G(e^{j\omega})|^2 \Phi_1(\omega)/\alpha_1 + \lambda_e}{|G(e^{j\omega})|^2 \Phi_2(\omega)/\alpha_2 + \lambda_e} \cdot \sup_{\omega} \frac{|G(e^{j\omega})|^2 \Phi_2(\omega)/\alpha_2 + \lambda_e}{|G(e^{j\omega})|^2 \Phi_1(\omega)/\alpha_1 + \lambda_e}}{\ln \max\left\{1, \sup_{\omega} \frac{\Phi_1(\omega)/\alpha_1}{\Phi_2(\omega)/\alpha_2}\right\} \cdot \max\left\{1, \sup_{\omega} \frac{\Phi_2(\omega)/\alpha_2}{\Phi_1(\omega)/\alpha_1}\right\}}.$$

Now, consider the cone  $C_{m,M} \subseteq L^+$  ( $0 < m \leq M$ ):

$$C_{m,M} := \left\{ \begin{bmatrix} \Phi \\ \alpha \end{bmatrix} \in L : m < \inf_{\omega} \frac{\Phi(\omega)}{\alpha} \leq \sup_{\omega} \frac{\Phi(\omega)}{\alpha} \leq M, \alpha > 0 \right\}.$$

Next, the contraction ratio is bounded in (A.4) in  $C_{m,M}$ .

**Lemma 29** Consider  $[\Phi_1 \ \alpha_1]^T, [\Phi_2 \ \alpha_2]^T \in C_{m,M}$ , such that  $\Phi_1/\alpha_1 \neq \Phi_2/\alpha_2$ . Then,

$$\frac{d\left(H \begin{bmatrix} \Phi_1 \\ \alpha_1 \end{bmatrix}, H \begin{bmatrix} \Phi_2 \\ \alpha_2 \end{bmatrix}\right)}{d\left(\begin{bmatrix} \Phi_1 \\ \alpha_1 \end{bmatrix}, \begin{bmatrix} \Phi_2 \\ \alpha_2 \end{bmatrix}\right)} \leq \frac{\ln \frac{\|G\|_{\infty}^2 M^2/m + \lambda_e}{\|G\|_{\infty}^2 M + \lambda_e}}{\ln M/m}.$$

The contractiveness of  $H$  is established next.

**Lemma 30**  $H$  is strictly contractive in  $C_{m,M}$ .

**Lemma 31** Let  $[\Phi_1 \ \alpha_1]^T \in C_{m,M}$  and define inductively  $[\Phi_{n+1} \ \alpha_{n+1}]^T = H[\Phi_n \ \alpha_n]^T$  for all  $n \in \mathbb{N}$ . Then there are constants  $c, C, K > 0$  such that  $[\Phi_n \ \alpha_n]^T \in C_{c,CK^n}$  for every  $n \in \mathbb{N}$ . Furthermore, these constants can be taken to satisfy  $c < 1 < K$ .

The final auxiliary result that is required is an asymptotic bound on the Hilbert distance between the spectrum at the  $n$ -th iteration and the limit spectrum.

**Lemma 32** Let  $[\Phi_1 \ \alpha_1]^T, [\Phi_{\infty} \ 1]^T \in C_{m,M}$ , where  $\Phi_{\infty}$  satisfies (14), and define inductively  $[\Phi_{n+1} \ \alpha_{n+1}]^T = H[\Phi_n \ \alpha_n]^T$  for all  $n \in \mathbb{N}$ . Then,

$$\frac{d\left(H \begin{bmatrix} \Phi_n \\ \alpha_n \end{bmatrix}, H \begin{bmatrix} \Phi_{\infty} \\ 1 \end{bmatrix}\right)}{d\left(\begin{bmatrix} \Phi_n \\ \alpha_n \end{bmatrix}, \begin{bmatrix} \Phi_{\infty} \\ 1 \end{bmatrix}\right)} \leq$$

$$1 + \frac{1}{n} \frac{\ln c}{\ln K} - \frac{1}{n^2} \frac{\ln(c) \ln(C/c)}{(\ln K)^2} + O(n^{-3}),$$

where  $c, C, K$  are given by Lemma 31.

The above auxiliary results enable the proof of Theorem 19 as is presented next.

**PROOF OF THEOREM 19.** Let  $m, M$  be such that  $[\Phi_1 \ \alpha_1]^T, [\Phi_{\infty} \ 1]^T \in C_{m,M}$ . Consider the Hilbert distance between  $[\Phi_n \ 1]^T$  and  $[\Phi_{\infty} \ 1]^T$ . By Lemma 32,

$$d\left(\begin{bmatrix} \Phi_n \\ \alpha_n \end{bmatrix}, \begin{bmatrix} \Phi_{\infty} \\ 1 \end{bmatrix}\right)$$

$$= d\left(H \begin{bmatrix} \Phi_{n-1} \\ \alpha_{n-1} \end{bmatrix}, H \begin{bmatrix} \Phi_{\infty} \\ 1 \end{bmatrix}\right) \quad (\text{A.5})$$

$$= d\left(\begin{bmatrix} \Phi_1 \\ \alpha_1 \end{bmatrix}, \begin{bmatrix} \Phi_{\infty} \\ 1 \end{bmatrix}\right) \prod_{k=1}^{n-1} \frac{d\left(H \begin{bmatrix} \Phi_k \\ \alpha_k \end{bmatrix}, H \begin{bmatrix} \Phi_{\infty} \\ 1 \end{bmatrix}\right)}{d\left(\begin{bmatrix} \Phi_k \\ \alpha_k \end{bmatrix}, \begin{bmatrix} \Phi_{\infty} \\ 1 \end{bmatrix}\right)}$$

$$\leq d\left(\begin{bmatrix} \Phi_1 \\ \alpha_1 \end{bmatrix}, \begin{bmatrix} \Phi_{\infty} \\ 1 \end{bmatrix}\right) \prod_{k=1}^{n-1} \left(1 + \frac{1}{k} \frac{\ln c}{\ln K} + O(k^{-2})\right),$$

where  $c, C, K$  are given by Lemma 31. Now,  $\ln c / \ln K$  is negative from Lemma 31 (since  $c < 1 < K$ ), and by the calculation in (B.1), the factors of the product in (A.5) lie between 0 and 1. Therefore,

$$\lim_{n \rightarrow \infty} d\left(\begin{bmatrix} \Phi_n \\ \alpha_n \end{bmatrix}, \begin{bmatrix} \Phi_{\infty} \\ 1 \end{bmatrix}\right)$$

$$\leq d\left(\begin{bmatrix} \Phi_1 \\ \alpha_1 \end{bmatrix}, \begin{bmatrix} \Phi_{\infty} \\ 1 \end{bmatrix}\right) \prod_{k=1}^{\infty} \left(1 - \frac{1}{k} \left|\frac{\ln c}{\ln K}\right| + O(k^{-2})\right)$$

$$\leq d\left(\begin{bmatrix} \Phi_1 \\ \alpha_1 \end{bmatrix}, \begin{bmatrix} \Phi_{\infty} \\ 1 \end{bmatrix}\right) \prod_{k=1}^{\infty} \exp\left(-\frac{1}{k} \left|\frac{\ln c}{\ln K}\right| + O(k^{-2})\right)$$

$$\leq d\left(\begin{bmatrix} \Phi_1 \\ \alpha_1 \end{bmatrix}, \begin{bmatrix} \Phi_{\infty} \\ 1 \end{bmatrix}\right) \exp\sum_{k=1}^{\infty} \left(-\frac{1}{k} \left|\frac{\ln c}{\ln K}\right| + O(k^{-2})\right)$$

$$= 0,$$

since  $1 + x \leq \exp x$  for every  $x \in \mathbb{R}$ . This means that  $[\Phi_n \ \alpha_n]^T \rightarrow [\Phi_\infty \ 1]^T$  in the Hilbert metric. To conclude the proof, it needs to be shown that  $\Phi_n/\alpha_n \rightarrow \Phi_\infty$  in the  $\mathcal{L}_\infty$  norm. Using Lemma 18,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \left\| \frac{\Phi_n}{\alpha_n} - \Phi_\infty \right\|_\infty \\ & \leq \lim_{n \rightarrow \infty} \left( \exp \left\{ d \left( \begin{bmatrix} \Phi_n \\ \alpha_n \end{bmatrix}, \begin{bmatrix} \Phi_\infty \\ 1 \end{bmatrix} \right) \right\} - 1 \right) \\ & \quad \cdot \exp \left\{ d \left( \begin{bmatrix} \Phi_n \\ \alpha_n \end{bmatrix}, \begin{bmatrix} \Phi_\infty \\ 1 \end{bmatrix} \right) \right\} (\sup \Phi_\infty) \\ & = 0. \end{aligned}$$

This concludes the proof.  $\square$

## B Proofs of Technical Lemmas

### B.1 Proof of Lemma 27

Without loss of generality, let  $a/c \leq b/d$ . Then,

$$\begin{aligned} \frac{a}{c} \leq \frac{b}{d} & \Leftrightarrow ad \leq bc \\ & \Leftrightarrow ad + bd \leq bc + bd \\ & \Leftrightarrow \frac{a+b}{c+d} \leq \frac{b}{d} = \max \left( \frac{a}{c}, \frac{b}{d} \right). \end{aligned}$$

The strict inequality case follows similarly.  $\square$

### B.2 Proof of Lemma 28

Notice that

$$\begin{aligned} & H \begin{bmatrix} \Phi_1 \\ \alpha_1 \end{bmatrix} \leq \beta H \begin{bmatrix} \Phi_2 \\ \alpha_2 \end{bmatrix} \\ & \Leftrightarrow \begin{bmatrix} |G|^2 \Phi_1 + \lambda_e \alpha_1 \\ \frac{1}{2\pi} \int |G|^2 \Phi_1 + \lambda_e \alpha_1 \end{bmatrix} \leq \beta \begin{bmatrix} |G|^2 \Phi_2 + \lambda_e \alpha_2 \\ \frac{1}{2\pi} \int |G|^2 \Phi_2 + \lambda_e \alpha_2 \end{bmatrix} \\ & \Leftrightarrow \begin{bmatrix} |G|^2 (\beta \Phi_2 - \Phi_1) + \lambda_e (\beta \alpha_2 - \alpha_1) \\ \frac{1}{2\pi} \int |G|^2 (\beta \Phi_2 - \Phi_1) + \lambda_e (\beta \alpha_2 - \alpha_1) \end{bmatrix} \geq 0 \\ & \Leftrightarrow |G(e^{j\omega})|^2 [\beta \Phi_2(\omega) - \Phi_1(\omega)] \geq -\lambda_e (\beta \alpha_2 - \alpha_1), \\ & \quad \omega \in [-\pi, \pi] \\ & \Leftrightarrow \beta \geq \frac{|G(e^{j\omega})|^2 \Phi_1(\omega) + \lambda_e \alpha_1}{|G(e^{j\omega})|^2 \Phi_2(\omega) + \lambda_e \alpha_2}, \quad \omega \in [-\pi, \pi] \end{aligned}$$

$$\Leftrightarrow \beta \geq \sup_\omega \frac{|G(e^{j\omega})|^2 \Phi_1(\omega) + \lambda_e \alpha_1}{|G(e^{j\omega})|^2 \Phi_2(\omega) + \lambda_e \alpha_2},$$

and, similarly,

$$\beta H \begin{bmatrix} \Phi_2 \\ \alpha_2 \end{bmatrix} \leq H \begin{bmatrix} \Phi_1 \\ \alpha_1 \end{bmatrix} \Leftrightarrow \beta \leq \inf_\omega \frac{|G(e^{j\omega})|^2 \Phi_1(\omega) + \lambda_e \alpha_1}{|G(e^{j\omega})|^2 \Phi_2(\omega) + \lambda_e \alpha_2}.$$

Therefore, the Hilbert metric between  $H[\Phi_1 \ \alpha_1]^T$  and  $H[\Phi_2 \ \alpha_2]^T$  is

$$\begin{aligned} d \left( H \begin{bmatrix} \Phi_1 \\ \alpha_1 \end{bmatrix}, H \begin{bmatrix} \Phi_2 \\ \alpha_2 \end{bmatrix} \right) &= \ln \frac{\sup_\omega \frac{|G(e^{j\omega})|^2 \Phi_1(\omega) + \lambda_e \alpha_1}{|G(e^{j\omega})|^2 \Phi_2(\omega) + \lambda_e \alpha_2}}{\inf_\omega \frac{|G(e^{j\omega})|^2 \Phi_1(\omega) + \lambda_e \alpha_1}{|G(e^{j\omega})|^2 \Phi_2(\omega) + \lambda_e \alpha_2}} \\ &= \ln \sup_\omega \frac{|G(e^{j\omega})|^2 \Phi_1(\omega) / \alpha_1 + \lambda_e}{|G(e^{j\omega})|^2 \Phi_2(\omega) / \alpha_2 + \lambda_e} + \ln \sup_\omega \frac{|G(e^{j\omega})|^2 \Phi_2(\omega) / \alpha_2 + \lambda_e}{|G(e^{j\omega})|^2 \Phi_1(\omega) / \alpha_1 + \lambda_e}. \end{aligned}$$

Combining the latter expression with Lemma 18 yields the desired result.  $\square$

### B.3 Proof of Lemma 29

The following four cases are distinguished:

- (1)  $\sup_\omega \frac{\Phi_1(\omega)/\alpha_1}{\Phi_2(\omega)/\alpha_2} \leq 1$  and  $\sup_\omega \frac{\Phi_2(\omega)/\alpha_2}{\Phi_1(\omega)/\alpha_1} \leq 1$ .
- (2)  $\sup_\omega \frac{\Phi_1(\omega)/\alpha_1}{\Phi_2(\omega)/\alpha_2} = k > 1$  and  $\sup_\omega \frac{\Phi_2(\omega)/\alpha_2}{\Phi_1(\omega)/\alpha_1} \leq 1$ .
- (3)  $\sup_\omega \frac{\Phi_1(\omega)/\alpha_1}{\Phi_2(\omega)/\alpha_2} \leq 1$  and  $\sup_\omega \frac{\Phi_2(\omega)/\alpha_2}{\Phi_1(\omega)/\alpha_1} = k > 1$ .
- (4)  $\sup_\omega \frac{\Phi_1(\omega)/\alpha_1}{\Phi_2(\omega)/\alpha_2} = k > 1$  and  $\sup_\omega \frac{\Phi_2(\omega)/\alpha_2}{\Phi_1(\omega)/\alpha_1} = l > 1$ .

In case 1, it turns out that  $\Phi_1/\alpha_1 = \Phi_2/\alpha_2$ , hence this case is omitted. Cases 2 and 3 are symmetrical, so only case 2 is considered. Here,

$$\begin{aligned} & \frac{d \left( H \begin{bmatrix} \Phi_1 \\ \alpha_1 \end{bmatrix}, H \begin{bmatrix} \Phi_2 \\ \alpha_2 \end{bmatrix} \right)}{d \left( \begin{bmatrix} \Phi_1 \\ \alpha_1 \end{bmatrix}, \begin{bmatrix} \Phi_2 \\ \alpha_2 \end{bmatrix} \right)} \leq \frac{\ln \sup_\omega \frac{|G(e^{j\omega})|^2 \Phi_1(\omega) / \alpha_1 + \lambda_e}{|G(e^{j\omega})|^2 \Phi_2(\omega) / \alpha_2 + \lambda_e}}{\ln k} \\ & \leq \frac{\ln \frac{k \|G\|_\infty^2 M + \lambda_e}{\|G\|_\infty^2 M + \lambda_e}}{\ln k} \leq \frac{\ln \frac{\|G\|_\infty^2 M^2 / m + \lambda_e}{\|G\|_\infty^2 M + \lambda_e}}{\ln M/m}, \end{aligned}$$

since  $k \leq M/m$ . Finally, in case 4, using Lemma 27 results in

$$\frac{d \left( H \begin{bmatrix} \Phi_1 \\ \alpha_1 \end{bmatrix}, H \begin{bmatrix} \Phi_2 \\ \alpha_2 \end{bmatrix} \right)}{d \left( \begin{bmatrix} \Phi_1 \\ \alpha_1 \end{bmatrix}, \begin{bmatrix} \Phi_2 \\ \alpha_2 \end{bmatrix} \right)}$$



$$\begin{aligned}
&= \frac{\ln \sup_{\omega} \frac{|G(e^{j\omega})|^2 \Phi_1(\omega)/\alpha_1 + \lambda_e}{|G(e^{j\omega})|^2 \Phi_2(\omega)/\alpha_2 + \lambda_e} \cdot \sup_{\omega} \frac{|G(e^{j\omega})|^2 \Phi_2(\omega)/\alpha_2 + \lambda_e}{|G(e^{j\omega})|^2 \Phi_1(\omega)/\alpha_1 + \lambda_e}}{\ln kl} \\
&\leq \frac{\ln \frac{k \|G\|_{\infty}^2 M + \lambda_e}{\|G\|_{\infty}^2 M + \lambda_e} + \ln \frac{l \|G\|_{\infty}^2 M + \lambda_e}{\|G\|_{\infty}^2 M + \lambda_e}}{\ln k + \ln l} \\
&\leq \max \left\{ \frac{\ln \frac{k \|G\|_{\infty}^2 M + \lambda_e}{\|G\|_{\infty}^2 M + \lambda_e}}{\ln k}, \frac{\ln \frac{l \|G\|_{\infty}^2 M + \lambda_e}{\|G\|_{\infty}^2 M + \lambda_e}}{\ln l} \right\} \\
&\leq \frac{\ln \frac{\|G\|_{\infty}^2 M^2/m + \lambda_e}{\|G\|_{\infty}^2 M + \lambda_e}}{\ln M/m}.
\end{aligned}$$

Finally, the desired result is obtained by combining all the considered cases.  $\square$

#### B.4 Proof of Lemma 30

From Lemmas 29 and 27, it follows that

$$\begin{aligned}
&\frac{d \left( H \begin{bmatrix} \Phi_1 \\ \alpha_1 \end{bmatrix}, H \begin{bmatrix} \Phi_2 \\ \alpha_2 \end{bmatrix} \right)}{d \left( \begin{bmatrix} \Phi_1 \\ \alpha_1 \end{bmatrix}, \begin{bmatrix} \Phi_2 \\ \alpha_2 \end{bmatrix} \right)} \leq \frac{\ln \frac{\|G\|_{\infty}^2 M^2/m + \lambda_e}{\|G\|_{\infty}^2 M + \lambda_e}}{\ln M/m} \\
&< \frac{\ln \frac{\|G\|_{\infty}^2 M^2/m}{\|G\|_{\infty}^2 M}}{\ln M/m} = \frac{\ln M/m}{\ln M/m} = 1. \quad (\text{B.1})
\end{aligned}$$

This shows that  $H$  is strictly contractive in  $C_{m,M}$ .  $\square$

#### B.5 Proof of Lemma 31

First notice that, from the definition of  $H$ ,

$$\frac{\Phi_{n+1}(\omega)}{\alpha_{n+1}} = \frac{|G(e^{j\omega})|^2 \Phi_n(\omega)/\alpha_n + \lambda_e}{\frac{1}{2\pi} \int_{-\pi}^{\pi} |G(e^{j\tau})|^2 \Phi_n(\tau)/\alpha_n d\tau + \lambda_e}. \quad (\text{B.2})$$

From (B.2), for every  $n > 1$ ,

$$\begin{aligned}
&\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Phi_n(\omega)}{\alpha_n} d\omega \\
&= \frac{\frac{1}{2\pi} \int_{-\pi}^{\pi} |G(e^{j\omega})|^2 \Phi_{n-1}(\omega)/\alpha_{n-1} d\omega + \lambda_e}{\frac{1}{2\pi} \int_{-\pi}^{\pi} |G(e^{j\tau})|^2 \Phi_{n-1}(\tau)/\alpha_{n-1} d\tau + \lambda_e} \\
&= 1.
\end{aligned}$$

Therefore, by the mean value theorem,

$$\inf \frac{\Phi_n}{\alpha_n} \geq \frac{\lambda_e}{\|G\|_{\infty}^2 + \lambda_e} \geq c,$$

where

$$c := \min \left\{ \frac{\lambda_e}{\|G\|_{\infty}^2 + \lambda_e}, m \right\}.$$

For the supremum of  $\Phi_n$ , (B.2) gives that

$$\frac{\sup \Phi_{n+1}}{\alpha_{n+1}} \leq \frac{\|G\|_{\infty}^2 \sup \Phi_n / \alpha_n + \lambda_e}{\lambda_e} \leq \frac{\|G\|_{\infty}^2}{\lambda_e} \frac{\sup \Phi_n}{\alpha_n} + 1. \quad (\text{B.3})$$

By iterating (B.3), for  $n \geq 2$

$$\begin{aligned}
\frac{\sup \Phi_n}{\alpha_n} &\leq \left( \frac{\|G\|_{\infty}^2}{\lambda_e} \right)^{n-1} \frac{\sup \Phi_1}{\alpha_1} + \sum_{k=0}^{n-2} \left( \frac{\|G\|_{\infty}^2}{\lambda_e} \right)^k \\
&= \left( \frac{\|G\|_{\infty}^2}{\lambda_e} \right)^{n-1} \frac{\sup \Phi_1}{\alpha_1} + \frac{\left( \frac{\|G\|_{\infty}^2}{\lambda_e} \right)^{n-1} - 1}{\left( \frac{\|G\|_{\infty}^2}{\lambda_e} \right) - 1}. \quad (\text{B.4})
\end{aligned}$$

Notice that (B.4) also holds for  $n = 1$ . Three cases have to be distinguished:  $\|G\|_{\infty}^2 > \lambda_e$ ,  $\|G\|_{\infty}^2 = \lambda_e$  and  $\|G\|_{\infty}^2 < \lambda_e$ . For the first case,

$$\begin{aligned}
\frac{\sup \Phi_n}{\alpha_n} &\leq \left( \frac{\|G\|_{\infty}^2}{\lambda_e} \right)^n \left[ \frac{\lambda_e}{\|G\|_{\infty}^2} M + \frac{\lambda_e^2}{\|G\|_{\infty}^4 - \lambda_e \|G\|_{\infty}^2} \right] \\
&= CK^n,
\end{aligned}$$

where

$$\begin{aligned}
C &:= \frac{\lambda_e}{\|G\|_{\infty}^2} M + \frac{\lambda_e^2}{\|G\|_{\infty}^4 - \lambda_e \|G\|_{\infty}^2} \\
K &:= \frac{\|G\|_{\infty}^2}{\lambda_e}.
\end{aligned}$$

For the case  $\|G\|_{\infty}^2 = \lambda_e$ , iterating (B.3), reveals that

$$\begin{aligned}
\frac{\sup \Phi_n}{\alpha_n} &\leq M + n - 1 \\
&= M \left( 1 + \frac{n-1}{M} \right) \leq M \exp \left( \frac{n-1}{M} \right) \\
&= M e^{-1/M} e^{n/M} = CK^n,
\end{aligned}$$

since  $1 + x \leq \exp x$  for every  $x \in \mathbb{R}$ , where  $C := M \exp(-1/M)$  and  $K := \exp(1/M)$ . Finally, if  $\|G\|_{\infty}^2 < \lambda_e$ , (B.4) implies that

$$\begin{aligned}
\frac{\sup \Phi_n}{\alpha_n} &\leq \left( \frac{\|G\|_{\infty}^2}{\lambda_e} \right)^{n-1} \frac{\sup \Phi_1}{\alpha_1} + \frac{1 - \left( \frac{\|G\|_{\infty}^2}{\lambda_e} \right)^{n-1}}{1 - \left( \frac{\|G\|_{\infty}^2}{\lambda_e} \right)} \\
&\leq \frac{\sup \Phi_1}{\alpha_1} + \frac{1}{1 - \left( \frac{\|G\|_{\infty}^2}{\lambda_e} \right)} \leq CK^n,
\end{aligned}$$

where

$$C := \frac{\sup \Phi_1}{\alpha_1} + \frac{\lambda_e}{\lambda_e - \|G\|_{\infty}^2}$$

and  $K > 1$  is arbitrary. This concludes the proof.  $\square$

## B.6 Proof of Lemma 32

By direct calculation from Lemmas 29 and 31,

$$\begin{aligned}
& \frac{d\left(H \begin{bmatrix} \Phi_n \\ \alpha_n \end{bmatrix}, H \begin{bmatrix} \Phi_\infty \\ 1 \end{bmatrix}\right)}{d\left(\begin{bmatrix} \Phi_n \\ \alpha_n \end{bmatrix}, \begin{bmatrix} \Phi_\infty \\ 1 \end{bmatrix}\right)} \\
& \leq \frac{\ln \frac{\|G\|_\infty^2 C^2 K^{2n}/c + \lambda_e}{\|G\|_\infty^2 C K^n/c + \lambda_e}}{\ln C K^n/c} \\
& = \frac{n \ln K + \ln C + \ln \frac{1 + \lambda_e c \|G\|_\infty^{-2} C^{-2} K^{-2n}}{1 + \lambda_e c \|G\|_\infty^{-2} C^{-1} K^{-n}}}{n \ln K + \ln C - \ln c} \\
& = 1 + \frac{1}{n \ln K} \frac{\ln c + \ln \frac{1 + \lambda_e c \|G\|_\infty^{-2} C^{-2} K^{-2n}}{1 + \lambda_e c \|G\|_\infty^{-2} C^{-1} K^{-n}}}{1 + \frac{\ln C/c}{n \ln K}} \\
& = 1 + \frac{1}{n \ln K} \left( \ln c + \lambda_e c \|G\|_\infty^{-2} C^{-2} K^{-2n} \right. \\
& \quad \left. - \lambda_e c \|G\|_\infty^{-2} C^{-1} K^{-n} + O(K^{-2n}) \right) \\
& \quad \cdot \left( 1 - \frac{1}{n} \frac{\ln C/c}{\ln K} + O(n^{-2}) \right) \\
& = 1 + \frac{1}{n \ln K} \left( \ln c - \frac{1}{n} \frac{\ln c \ln C/c}{\ln K} + O(n^{-2}) \right) \\
& = 1 + \frac{1}{n \ln K} - \frac{1}{n^2} \frac{\ln(c) \ln(C/c)}{(\ln K)^2} + O(n^{-3}).
\end{aligned}$$

□

## References

- [1] P. Albertos and A. Sala, editors. *Iterative Identification and Control*. Springer, 2002.
- [2] M. Barenthin, M. Enqvist, B. Wahlberg, and H. Hjalmarsson. Gain estimation for Hammerstein systems. In *Proceedings of the 14th IFAC Symposium on System Identification*, pages 784–789, Newcastle, Australia, 2006.
- [3] M. Barenthin, H. Mosskull, H. Hjalmarsson, and B. Wahlberg. Validation of stability for an induction machine drive using power iterations. In *Proceedings of the 16th IFAC World Congress*, Prague, Czech Republic, 2005.
- [4] R. G. Bartle. *The Elements of Integration*. John Wiley & Sons, 1966.
- [5] G. Birkhoff. Extensions of Jentzsch’s theorem. *Transactions of the American Mathematical Society*, 85:219–227, 1957.
- [6] J. M. Böling and P. M. Mäkilä. On control relevant criteria in  $\mathcal{H}_\infty$  identification. *IEEE Transactions on Automatic Control*, 43(5):694–700, 1998.
- [7] A. Böttcher and S. M. Grudsky. *Spectral properties of banded Toeplitz matrices*. SIAM, 2005.
- [8] 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 Control Systems Magazine*, 26(3):96–114, 2006.
- [9] P. J. Bushell. Hilbert’s metric and positive contraction mappings in a Banach space. *Archive for Rational Mechanics and Analysis*, 52(2):330–338, 1973.
- [10] M. A. Erickson, R. S. Smith, and A. J. Laub. Power methods for calculating eigenvalues and eigenvectors of spectral operators on Hilbert spaces. *International Journal of Control*, 62(5):1117–1128, 1995.
- [11] G. Golub and C. Van Loan. *Matrix Computations*, 3rd Edition. Johns Hopkins University Press, 1996.
- [12] 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 Transactions on Automatic Control*, 42(11):1516–1528, 1997.
- [13] K. van Heusden, A. Karimi, and D. Bonvin. Data-driven estimation of the infinity norm of a dynamical system. In *Proceedings of the 46th IEEE Conference on Decision and Control (CDC’07)*, New Orleans, USA, December 12–14 2007.
- [14] H. Hjalmarsson. Iterative feedback tuning – an overview. *International Journal of Adaptive Control and Signal Processing*, 16(5):373–395, 2002.
- [15] H. Hjalmarsson. From experiment design to closed-loop control. *Automatica*, 41:393–438, 2005.
- [16] H. Hjalmarsson, S. Gunnarsson, and M. Gevers. Optimality and sub-optimality of iterative identification and control design schemes. In *Proceedings of the 15th American Control Conference (ACC’95)*, pages 2559–2563, Seattle, 1995.
- [17] T. P. Krasulina. Method of stochastic approximation in the determination of the largest eigenvalue of the mathematical expectation of random matrices. *Automation and Remote Control*, 2:215–221, 1970.
- [18] L. Ljung. Model validation and model error modeling. In B. Wittenmark and A. Rantzer, editors, *The Åström Symposium on Control*, pages 15–42, Lund, Sweden, 1999. Studentlitteratur.
- [19] L. Ljung. *System Identification: Theory for the User*, 2nd Edition. Prentice Hall, 1999.
- [20] E. Oja and J. Karhunen. On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix. *Journal of Mathematical Analysis and Applications*, 106(1):69–84, 1985.
- [21] J. R. Partington. *An Introduction to Hankel Operators*. Cambridge University Press, 1988.
- [22] B. Porat. *Digital Processing of Random Signals*. Prentice-Hall, 1994.
- [23] W. Rudin. *Principles of Mathematical Analysis*, 3rd Edition. McGraw-Hill, 1976.
- [24] W. Rudin. *Real and Complex Analysis*, 3rd Edition. McGraw-Hill, 1987.
- [25] H. H. Schaefer. *Topological Vector Spaces*. Springer-Verlag, 1971.
- [26] D. K. de Vries and P. M. J. Van den Hof. Quantification of uncertainty in transfer function estimation: a mixed probabilistic-worst-case approach. *Automatica*, 31(4):543–557, 1995.
- [27] B. Wahlberg, M. Barenthin, and H. Hjalmarsson. Non-parametric methods for  $\mathcal{L}_2$ -gain estimation using iterative experiments. *Automatica*, 46(8):1376–1381, 2010.
- [28] B. Wahlberg, H. Hjalmarsson, and P. Stoica. On optimal input signal design for frequency response estimation. In *Proceedings of the 49th Conference on Decision and Control (CDC’10)*, Atlanta, 2010.
- [29] B. Wahlberg, H. Hjalmarsson, and P. Stoica. On estimation of the gain of a dynamical system. In *Proceedings of the IEEE Digital Signal Processing Workshop and IEEE Signal Processing Education Workshop (DSE/SPE)*, pages 784–789, Sedona, AZ, USA, 2011.
- [30] K. Zhou, J. C. Doyle, and K. Glover. *Robust and Optimal Control*. Prentice Hall, 1996.