Data-Driven Methods for $\mathcal{L}_2$ -Gain Estimation [★]

Märta Barenthin ^{*}, Bo Wahlberg ^{*}, Håkan Hjalmarsson ^{*},
Mathias Barkhagen ^{*}

^{*} Automatic Control and ACCESS, School of Electrical Engineering,
KTH, 100 44 Stockholm, Sweden

Abstract: In this paper we present and discuss some data-driven methods for estimation of the $\mathcal{L}_2$ -gain of dynamical systems. Partial results on convergence and statistical properties are provided. The methods are based on multiple experiments on the system. The main idea is to directly estimate the maximizing input signal by using iterative experiments on the true system. We study such a data-driven method based on a stochastic gradient method. We show that this method is very closely related to the so-called power iteration method based on the power method in numerical analysis. Furthermore, it is shown that this method is applicable for linear systems with noisy measurements. We will also study $\mathcal{L}_2$ -gain estimation of Hammerstein systems. The stochastic gradient method and the power iteration method are evaluated and compared in simulation examples.

Keywords: $\mathcal{L}_2$ -gain estimation; Small gain theorem; infinity norm; Power method; Input signal design.

1. INTRODUCTION

A very useful tool for validating closed-loop stability of a feedback system is the small gain theorem (see *e.g.* Khalil [2002]), which is based on an upper bound of the $\mathcal{L}_2$ -gain of the corresponding open loop system. Robust stability is often analyzed using the $\mathcal{L}_2$ -gain of the model error. Here the transfer function is decomposed into two parts $G_{nom} + G_{error}$. The nominal part, G_{nom} , can often be described by a parametric model, while the error part, G_{error} , should capture the modeling errors and is much more complex. Standard system identification methods can be used to find an accurate model of G_{nom} , while data-based model error modeling is more complicated and often of nonparametric or of high order nature, see *e.g.* Mosskull et al. [2003], Barenthin et al. [2005], Reinelt et al. [2002], Schoukens et al. [2002], Poolla et al. [1994] and Kosut [1995] for some results. However, in many robust control application it is enough to find an upper bound of the model error. The objective of this paper is to study methods for direct estimation of such a bound, namely the $\mathcal{L}_2$ -gain of a discrete time dynamical system. The approach is designed to be applied to the model error system G_{error} , and is based on iterative experiments rather than on standard system identification methods.

In this paper we consider discrete-time signals and systems. Consider. The $\mathcal{L}_2$ -norm of the signal $u(k)$, $k = 0, 1, 2, \dots$ is defined by

$$\|u\| = \sqrt{\sum_{k=0}^{\infty} u^2(k)}.$$

Consider a single input single output (SISO) discrete-time dynamical system

$$\begin{aligned} x(t+1) &= h_x(x(t), u(t)) \\ x(0) &= x_0 \\ y(t) &= h_y(x(t), u(t)), \quad t = 0, 1, 2, \dots, \end{aligned}$$

where $u(t)$ and $y(t)$ denote the input and output, respectively, and $x(t)$ is the state vector. The $\mathcal{L}_2$ -gain of such a system is here defined as ¹

$$\beta = \sup_{u \neq 0, u \in \mathcal{L}_2} \frac{\|y\|}{\|u\|},$$

where $\mathcal{L}_2 = \{u, \|u\| < \infty\}$. For linear time-invariant (LTI) systems it is also called the induced $\mathcal{L}_2$ -norm or the $\mathcal{H}_\infty$ -norm in the literature. In the sequel, when we speak of the gain we refer to the $\mathcal{L}_2$ -gain.

In many practical situations the true system is unknown, but experimental input/output data can be used in order to obtain information about the system dynamics. A trivial lower bound for β can be found by calculating $\|y\|/\|u\|$ using *any* pair of input/output signals. The paper van Heusden et al. [2007] presents a method to estimate the $\mathcal{H}_\infty$ -norm of LTI systems with noisy output measurements from a *given* set of input/output data. However, it is not obvious what input should be used in order to obtain a lower bound which is close to β . Thus the gain estimation problem corresponds to an input design problem: how should the input signal be chosen so that the lower bound is close to β ? It is harder to compute upper bounds of β from experiments, which of course is an even more relevant problem.

The objective of this paper is to present and discuss some data-driven methods for computing accurate estimates of β . One way of handling the fact that the system is unknown is by first identifying a model of the system, using *e.g.* prediction error methods, see *e.g.* Ljung [1999]. Then, the gain of the model can be used as an approximation of the gain of the true system. In fact, the invariance principle

[★] This work was partially supported by the Swedish Research Council and the Linnaeus Center ACCESS at KTH

¹ Some small gain stability problems need a bound of the form $\|y\|_2 \leq \beta \|u\|_2 + \alpha$ for some constant α , *e.g.* if the system contains a relay.

of the maximum likelihood (ML) method, assuming the correct model structure, shows that ML estimation of the system parameters followed by computation of the gain of the corresponding estimated model gives asymptotically (in the sample size) optimal accuracy of the $\mathcal{L}_2$ -gain for the given data set, see *e.g.* Hjalmarsson [2005].

We will study two data-driven approaches, where the optimization is with respect to the whole input sequence. The first method is a stochastic gradient method, and the second method is the so-called power iteration method. A basic version of the power iteration method for data-driven gain estimation was first suggested in Hjalmarsson [2005], where it was also illustrated by means of a simulation example that this method can be useful also for certain nonlinear systems.

Outline: Section 2 concerns gain estimation by iterative experiments. In Section 3 a stochastic gradient method is derived, and in Section 4 this method is used to gain estimation for Hammerstein systems. Section 5 shows that the power iteration method is a special case of the stochastic gradient method. The derived algorithms are evaluated by means of simulation examples in Section 6. The paper is concluded in Section 7.

2. GAIN ESTIMATION BY ITERATIVE EXPERIMENTS

In this section we will derive a data-driven method for gain estimation based on iterative experiments.

2.1 Experimental set-up

Consider the stable LTI discrete-time system

$$y(t) = \sum_{k=1}^{\infty} g_k u(t-k), \quad (1)$$

where the sequence $\{g_k\}_{k=1}^{\infty}$ denotes the impulse response of the system. Assume that $\sum_{k=1}^{\infty} |g_k| < \infty$ and $u(t) = 0$, $t \leq 0$. Introduce the following definitions:

$$\begin{aligned} u^N &= [u(1), u(2) \dots u(N)]^T, \\ y^N &= [y(1), y(2) \dots y(N)]^T. \end{aligned}$$

Notice that

$$\|u^N\| = \sqrt{(u^N)^T (u^N)}.$$

Given zero initial conditions, the input to output relation is given by

$$y^N = \mathbf{G} u^N,$$

where $\mathbf{G}$ is the lower triangular Toeplitz matrix

$$\mathbf{G} = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ g_1 & 0 & 0 & \dots & 0 \\ g_2 & g_1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ g_{N-1} & \dots & g_2 & g_1 & 0 \end{bmatrix}.$$

Define the objective function

$$V(u^N) = \frac{(y^N)^T (y^N)}{(u^N)^T (u^N)} = \frac{(u^N)^T \mathbf{G}^T \mathbf{G} (u^N)}{(u^N)^T (u^N)}, \quad (2)$$

which should be maximized w.r.t. u^N to find the optimal gain. From eigenvalue analysis it is straightforward to verify that

$$V(u^N) \leq \lambda^{\max}[\mathbf{G}^T \mathbf{G}],$$

the maximum eigenvalue of $\mathbf{G}^T \mathbf{G}$, with equality when u^N is the corresponding eigenvector. Notice that the $\mathcal{L}_2$ -gain equals

$$\beta = \lim_{N \rightarrow \infty} \sqrt{\lambda^{\max}[\mathbf{G}^T \mathbf{G}]}. \quad (3)$$

The Toeplitz structure now implies that $\lambda^{\max}[\mathbf{G}^T \mathbf{G}]$ tends to $\max_{\omega} |\sum_{k=1}^{\infty} g_k e^{-j\omega k}|^2$ as $N \rightarrow \infty$ and we have the previously mentioned result for LTI systems (Massoumina et al. [1994]).

Key observation: For a given model we can, of course, construct the matrix $\mathbf{G}^T \mathbf{G}$ and numerically solve the corresponding eigenvalue problem. However, in order to use an iterative algorithm for eigenvalue calculation it is enough to be able to find the vector $\mathbf{G}^T \mathbf{G} u^N = \mathbf{G}^T y^N$. Let the reversed time output sequence

$$\tilde{y}^N = [y(N), y(N-1) \dots y(1)]^T$$

be used as input signal to the system, and denote the corresponding output signal and its time reversed version by

$$\begin{aligned} z^N &= [z(1), z(2) \dots z(N)]^T \\ \tilde{z}^N &= [z(N), z(N-1) \dots z(1)]^T. \end{aligned}$$

Then, due to the Toeplitz structure,

$$\tilde{z}^N = \mathbf{G}^T y^N = \mathbf{G}^T \mathbf{G} u^N. \quad (3)$$

Hence, $\mathbf{G}^T \mathbf{G} u^N$ can be determined from two experiments without explicitly relying on a model of $\mathbf{G}$. Also notice that if $u^N = \bar{u}^N$ is the eigenvector associated with $\lambda^{\max}[\mathbf{G}^T \mathbf{G}]$, we have $\tilde{z}^N = \lambda^{\max}[\mathbf{G}^T \mathbf{G}] \bar{u}^N$.

Remark 1. The operator $\mathbf{G}^T$ is the so-called adjoint operator, see *e.g.* Kailath [1980]. $\square$

Now assume that the measured output is corrupted with additive noise

$$y(t) = \sum_{k=1}^{\infty} g_k u(t-k) + e(t).$$

For simplicity assume $\{e(t)\}$ to be white noise with variance λ_e . The noise corrupted version of the signal (3) then equals

$$\tilde{z}^N = \mathbf{G}^T \mathbf{G} u^N + \mathbf{G}^T e_1^N + \tilde{e}_2^N,$$

where e_1^N is the noise vector from the first experiment and $\tilde{e}_2^N$ comes from the second one.

The maximum eigenvalue estimation problem to be studied consists of two parts. First we need to find the optimal input sequence, which in this case corresponds to u^N being an eigenvector associated with $\lambda^{\max}[\mathbf{G}^T \mathbf{G}]$. The second issue is how to compensate for noisy measurements. We will, however, do the analysis in the reverse order.

2.2 Statistical analysis

Assume that we know the corresponding dominating eigenvector $\bar{u}^N$ and that we use the corresponding sequence as input signal. This gives

$$\tilde{z}^N = \lambda^{\max} \bar{u}^N + \mathbf{G}^T e_1^N + \tilde{e}_2^N, \quad (4)$$

where we have introduced the short hand notation $\lambda^{\max} = \lambda^{\max}[\mathbf{G}^T \mathbf{G}]$. The variance of the noise part in (4) is $\lambda_e (\mathbf{G}^T \mathbf{G} + I_N)$. The best linear unbiased estimate of $\lambda^{\max}$ is then given by the weighted least squares problem

$$\min_{\lambda} [\tilde{z}^N - \lambda \bar{u}^N]^T (\mathbf{G}^T \mathbf{G} + I_N)^{-1} [\tilde{z}^N - \lambda \bar{u}^N]. \quad (5)$$

A difficulty here is that the weighting matrix is unknown, but since $\bar{u}^N$ is an eigenvector to $\mathbf{G}^T \mathbf{G}$ we have

$$(\mathbf{G}^T \mathbf{G} + I_N)^{-1} \bar{u}^N = \frac{1}{1 + \lambda^{max}} \bar{u}^N.$$

Using this result, it follows that the minimizing solution to (5) equals the standard (un-weighted) least squares solution

$$\hat{\lambda}^{max} = \frac{(\bar{u}^N)^T (\tilde{z}^N)}{(\bar{u}^N)^T (\bar{u}^N)}. \quad (6)$$

Hence

$$\hat{\lambda}^{max} = \lambda^{max} + \frac{(\bar{u}^N)^T (\mathbf{G}^T e_1^N + \tilde{e}_2^N)}{(\bar{u}^N)^T (\bar{u}^N)}.$$

It is possible to verify that the error term

$$\frac{(\bar{u}^N)^T (\mathbf{G}^T e_1^N + \tilde{e}_2^N)}{(\bar{u}^N)^T (\bar{u}^N)}$$

has zero mean value and variance

$$\mathbb{E} \left\{ \left(\frac{(\bar{u}^N)^T (\mathbf{G}^T e_1^N + \tilde{e}_2^N)}{(\bar{u}^N)^T (\bar{u}^N)} \right)^2 \right\} = \frac{\lambda_e (\lambda^{max} + 1)}{\|\bar{u}^N\|^2}$$

Here we have used that $e(t)$ is white noise and that $\bar{u}_N$ is an eigenvector of $\mathbf{G}^T \mathbf{G}$. This expression very nicely captures the dependence on the input signal energy $\|\bar{u}^N\|$. Also, we conclude that the error is proportional to the noise variance and to the gain λ^{max} due to possible amplification of noise in the second experiment.

For convergence the size of $\|\bar{u}^N\|$ should tend to infinity when N tends to infinity. This corresponds to a standard persistence of excitation condition in system identification (Ljung [1999]). For a LTI system $G(e^{j\omega})$, the maximizing input signal is a sinusoidal signal with frequency corresponding to $\arg \max_{\omega} |G(e^{j\omega})|$. Thus for LTI systems, it holds that $\|\bar{u}^N\|^2$ is proportional to N . Hence, $\hat{\lambda}^{max}$ is a consistent estimate of the $\mathcal{L}_2$ -gain. The more direct estimate

$$\hat{\lambda}^{max,d} = \frac{(y^N)^T (y^N)}{(\bar{u}^N)^T (\bar{u}^N)} \quad (7)$$

has the dominating noise error term $(e_1^N)^T (e_1^N) / (\bar{u}^N)^T (\bar{u}^N)$, with mean value $N \lambda_e / \|\bar{u}^N\|^2$, which typically do not tend to zero as $N \rightarrow \infty$. Hence, one should be careful when using this estimate for noisy measurements.

Next, we will analyze a method for estimation of the dominating eigenvector $\bar{u}^N$ corresponding to the largest eigenvalue λ^{max} .

3. A STOCHASTIC GRADIENT METHOD (SGM)

Calculating the maximum of $V(u^N)$ defined by (2) with respect to u^N using gradient based numerical optimization methods directly leads to algorithms of the type

$$\begin{aligned} u_{k+1}^N &= u_k^N + \frac{\gamma_k}{2} \frac{dV(u^N)}{du^N} \Big|_{u^N=u_k^N} \\ &= u_k^N + \gamma_k \left(\frac{\left(\frac{dy_k^N}{du_k^N} \right)^T (y_k^N)}{(u_k^N)^T (u_k^N)} - \frac{(y_k^N)^T (y_k^N) (u_k^N)}{[(u_k^N)^T (u_k^N)]^2} \right), \end{aligned} \quad (8)$$

where $\gamma_k > 0$ is a scalar corresponding to the step length of the algorithm. As explained in Markusson [2002], this

approach is closely related to Iterative Learning Control, see *e.g.* Moore [1999] and Norrlöf [2000]. Define

$$\hat{\lambda}_k^{max,d} = \frac{(y_k^N)^T (y_k^N)}{(u_k^N)^T (u_k^N)}. \quad (9)$$

Consider the LTI system given by (1), for which $y^N = \mathbf{G} u^N$ and hence $dy^N/du^N = \mathbf{G}$. This means that $(dy_k^N/du_k^N)^T (y_k^N)$ in (8) equals $\mathbf{G}^T y_k^N$. This term can thus be determined experimentally by performing a second experiment on the system with the input $\tilde{y}_k^N$.

This leads to the update

$$u_{k+1}^N = u_k^N + \frac{\gamma_k}{(u_k^N)^T (u_k^N)} \left(\tilde{z}_k - \hat{\lambda}_k^{max,d} u_k^N \right).$$

Remark 2. For nonlinear systems $(dy_k^N/du_k^N)^T (y_k^N)$ does in general not equal $\mathbf{G}^T y_k^N$. $\square$

Recall from the previous section that $\hat{\lambda}^{max,d}$ is sensitive to noise. Therefore $\hat{\lambda}^{max,d}$ should be replaced by $\hat{\lambda}^{max}$ defined by (6). To get some insights in the convergence properties of this algorithm, consider noise free measurements and let the following properties be fixed, *i.e.* independent of k :

$$\begin{aligned} \bar{\gamma} &= \frac{\gamma_k}{(u_k^N)^T (u_k^N)} \\ \hat{\lambda}^{max,d} &= \hat{\lambda}_k^{max,d}. \end{aligned}$$

Then

$$\begin{aligned} u_{k+1}^N &= u_k^N + \bar{\gamma} (\tilde{z}_k - \hat{\lambda}^{max,d} u_k^N) \\ &= (I + \bar{\gamma} (\mathbf{G}^T \mathbf{G} - \hat{\lambda}^{max,d} I)) u_k^N. \end{aligned} \quad (10)$$

The eigenvalues of $I + \bar{\gamma} (\mathbf{G}^T \mathbf{G} - \hat{\lambda}^{max,d} I)$ equal

$$1 + \bar{\gamma} (\lambda_j - \hat{\lambda}^{max,d}), \quad j = 1 \dots N,$$

where $\lambda^{max} = \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N \geq 0$. Following the proof of the power method given in *e.g.* Golub et al. [1983], it is possible to show that

$$u_{k+1}^N = C \bar{u}^N + O \left(\frac{\max_{j \neq 1} |1 + \bar{\gamma} (\lambda_j - \hat{\lambda}^{max,d})|^k}{|1 + \bar{\gamma} (\lambda^{max} - \hat{\lambda}^{max,d})|^k} \right),$$

where C is a constant and $\bar{u}^N$ is the dominating eigenvector corresponding to λ^{max} . Hence we need

$$\max_{j \neq 1} |1 + \bar{\gamma} (\lambda_j - \hat{\lambda}^{max,d})| < |1 + \bar{\gamma} (\lambda^{max} - \hat{\lambda}^{max,d})|$$

for the error to converge to zero. The rate of convergence depends on $\bar{\gamma}$. The choice $\bar{\gamma} = 1/\hat{\lambda}^{max,d}$ will correspond to the power method and has some good converge properties. For this case

$$\frac{\max_{j \neq 1} |1 + \bar{\gamma} (\lambda_j - \hat{\lambda}^{max,d})|}{|1 + \bar{\gamma} (\lambda^{max} - \hat{\lambda}^{max,d})|} = \frac{|\lambda_2|}{|\lambda^{max}|},$$

and hence the ratio between the second largest and largest eigenvalue determines the rate of convergence. This choice of step length is, however, not optimal from a stochastic perspective. For standard stochastic gradient algorithms, the normalized step length $\gamma_k / [(u_k^N)^T (u_k^N)]$ should tend to zero as $k \rightarrow \infty$ for convergence. Another extra difficulty in our case is the normalization of u_{k+1}^N described below.

Since it can happen that $\lambda^{max} - \hat{\lambda}^{max,d} \geq 0$ (we are using a lower bound of the maximum eigenvector), the recursion (10) is typically unstable. We are, however, only interested in the eigenvector direction. The practical way to solve this stability problem is to use normalization

$$v_{k+1}^N = u_k^N + \frac{\gamma_k}{(u_k^N)^T (u_k^N)} \left(\tilde{z}_k - \hat{\lambda}_k^{\max, d} u_k^N \right)$$

$$u_{k+1}^N = \alpha \frac{v_{k+1}^N}{\|v_{k+1}^N\|}$$

and use the normalized input sequence u_{k+1}^N in the experiment.

To conclude, the stochastic gradient method (SGM) to be evaluated is:

- (1) Let $k = 1$ and select a non-zero input sequence $u_1^N = [u_1(1), u_1(2) \dots u_1(N)]^T$, such that $\|u_1^N\| = \alpha$.
- (2) Perform an experiment where the input sequence u_k^N is applied to the system and the output sequence $y_k^N = [y_k(1), y_k(2) \dots y_k(N)]^T$ is measured.
- (3) Perform a second experiment where the input sequence $\tilde{y}_k^N = [y_k(N), y_k(N-1) \dots y_k(1)]^T$ is applied to the system and the output sequence $z_k^N = [z_k(1), z_k(2), \dots, z_k(N)]^T$ is measured. Let $\tilde{z}_k^N = [z_k(N), z_k(N-1), \dots, z_k(1)]^T$.
- (4) Calculate

$$\hat{\lambda}_k^{\max} = \frac{(u_k^N)^T (\tilde{z}_k^N)}{\alpha^2}$$

to obtain the gain estimate

$$\hat{\beta}_k = \sqrt{\hat{\lambda}_k^{\max}}.$$

- (5) Update the input signal

$$v_{k+1}^N = u_k^N + \frac{\gamma_k}{\alpha^2} \left(\tilde{z}_k^N - \hat{\lambda}_k^{\max} u_k^N \right)$$

$$u_{k+1}^N = \alpha \frac{v_{k+1}^N}{\|v_{k+1}^N\|}$$

- (6) Let $k = k + 1$ and go to Step 2.

Note that u_1^N must have a component in the direction of the maximum eigenvector in order for the method to converge. The design parameters are u_1^N , the step size γ_k and the normalization constant α . In case of high signal to noise ratio and slightly nonlinear systems $\lambda_k^{\max, d}$ defined by (9) can be used instead of $\lambda_k^{\max}$ in Steps 4 and 5. For nonlinear systems one may also have to test several values of α , since the gain may depend on the allowed size of the input signal.

We have presented only partial results about convergence and statistical properties of the SGM method. One way to further robustify the method against the presence of measurement noise is to apply a periodic input, where the sequence u_k^N is repeated l times, and compute an average gain estimate over the periods.

4. SGM FOR HAMMERSTEIN SYSTEMS

A Hammerstein system with input $u(t)$ and output $y(t)$ is given by

$$y(t) = G(q)v(t) \quad (11)$$

$$v(t) = f(u(t)),$$

where q denotes the shift operator $qu(t) = u(t+1)$. Furthermore, $G(q)$ represents an LTI system and f denotes a static nonlinearity. The system is depicted in Figure 1. Let us now study the expression (8) for a Hammerstein system. We have that

$$(dy_k^N / du_k^N)^T y_k^N = \mathbf{f}_k \tilde{z}_k^N,$$

Fig. 1. Hammerstein system.

where

$$\mathbf{f}_k = \text{diag} [f'(u_k(1)), f'(u_k(2)), \dots, f'(u_k(N))]$$

and $\tilde{z}_k^N = \mathbf{G}^T y_k^N$. Note that calculating this derivative requires f to be known and invertible. However, if f contains a saturation then it can happen that $f^{-1}(a) = f^{-1}(b)$, $a \neq b$. This can be easily solved by noting that scaling the signal does not affect the gain of the LTI part of the Hammerstein system. The scaling procedure is as follows:

Assume that $f(a) = f(b) \Rightarrow a = b$, $\forall a, b \in [-x_f, x_f]$ and define $c_f = \sup_t \tilde{y}(t)/x_f$. Then the experiment on the LTI part is performed with the scaled input

$$\left[f^{-1} \left(\frac{\tilde{y}_k(1)}{c_f} \right), f^{-1} \left(\frac{\tilde{y}_k(2)}{c_f} \right), \dots, f^{-1} \left(\frac{\tilde{y}_k(N)}{c_f} \right) \right] \quad (12)$$

and the output is multiplied by c_f .

Since the Hammerstein system is nonlinear there is no guarantee that the gain estimate will increase monotonically with the iterations. Therefore, we add a condition such that signal updates are avoided if the gain decreases compared to the previous input signal. To summarize, Steps 3-5 in SGM are modified as follows:

- (3) Perform a second experiment where the input sequence (12) is applied to the system and the output sequence $z_k^N = [z_k(1), z_k(2), \dots, z_k(N)]^T$ is measured. Let $\tilde{z}_k^N = c_f [z_k(N), z_k(N-1), \dots, z_k(1)]^T$.
- (4) Calculate

$$\hat{\lambda}_k^{\max} = \frac{(y_k^N)^T y_k^N}{\alpha^2}$$

to obtain the gain estimate

$$\hat{\beta}_k = \sqrt{\hat{\lambda}_k^{\max}}.$$

- (5) Update the signal: if $\hat{\beta}_k < \hat{\beta}_{k-1}$ then

$$v_{k+1}^N = \frac{1}{2} (u_k^N + u_{k-1}^N)$$

$$u_{k+1}^N = \alpha \frac{v_{k+1}^N}{\|v_{k+1}^N\|}.$$

else

$$v_{k+1}^N = u_k^N + \frac{\gamma_k}{\alpha^2} (\mathbf{f}_k \tilde{z}_k^N - \hat{\lambda}_k^{\max} u_k^N)$$

$$u_{k+1}^N = \alpha \frac{v_{k+1}^N}{\|v_{k+1}^N\|}$$

This method can be used only for the case when f is known and invertible.

5. THE POWER ITERATION METHOD (PIM)

A simple way to calculate the maximum eigenvalue of a matrix $\mathbf{A}$ is the classical power method, which appears in many standard books on matrix analysis, see *e.g.* Golub et al. [1983]. To be more specific define the positive semi-definite symmetric matrix

$$\mathbf{A} = \mathbf{G}^T \mathbf{G} \in \mathbb{R}^{N \times N}$$

and denote by $\lambda_1, \lambda_2, \dots, \lambda_N$ the eigenvalues of $\mathbf{A}$ with corresponding normalized eigenvectors $x_1, x_2, \dots, x_N$ where $\|x_i\| = 1, \forall i$. Assume that $\lambda_1 > \lambda_2 \geq \lambda_3, \dots, \geq \lambda_N \geq 0$. Given $u_1, \|u_1\| = 1$, the power method is as follows for $k = 1, 2, \dots$:

$$\begin{aligned} v_{k+1}^N &= \mathbf{A}u_k \\ \hat{\lambda}_k^{\max} &= \|v_{k+1}\| \\ u_{k+1}^N &= \frac{v_{k+1}^N}{\|v_{k+1}^N\|}, \end{aligned}$$

where $\hat{\lambda}_k^{\max}$ denotes the estimate of λ_1 after k iterations. Assume that we can write $u_1^N = a_1x_1 + a_2x_2 + \dots + a_Nx_N$. For convergence it must hold that $a_1 \neq 0$, *i.e.* u_1^N must have a component in the direction of x_1 . The convergence rate of the power method is $O((\lambda_2/\lambda_1)^k)$, see Golub et al. [1983].

The power iteration method (PIM) for $\mathcal{L}_2$ -gain estimation is just an implementation of the power method using the two-experiments approach described previously. As noted in the previous section this corresponds to setting

$$\gamma_k = \alpha^2 / \hat{\lambda}_k^{\max}$$

in Step 5 in the basic data-driven algorithm presented in the previous section. Thus PIM is a special case of SGM. This gives the simple updates

$$\begin{aligned} v_{k+1}^N &= \tilde{z}_k^N \\ u_{k+1}^N &= \alpha \frac{v_{k+1}^N}{\|v_{k+1}^N\|} \end{aligned}$$

The results and analysis for the power method now directly apply to PIM. Hence, without measurement errors the iteration produce monotonically increasing gain estimates and the convergence point can be made arbitrarily close to the system gain by using long enough experiments (large N). A simplified version of this algorithm was presented in Hjalmarsson [2005], where it is illustrated, by means of a simulation example, that PIM may also be useful for certain nonlinear systems.

The properties of PIM have been established theoretically for LTI systems only. For nonlinear systems there is at present no theoretical support for PIM. This method produces, of course, lower bound to the gain but nothing is known about the accuracy. An interesting observation is that if the iterations do not converge to a sinusoid, the system must be nonlinear. In the next section the method is implemented also for a nonlinear system.

In particular, nonlinear systems have the property that the gain may be dependent on the amplitude of the input signal. Therefore, different types of initial inputs and different norms α may give rise to more or less accurate gain estimates and it is often a good idea to try several different initial inputs. An extension to PIM is to allow α to vary from one iteration to the other. Another possibility is to run several parallel PIMs with different α , and check for the one that gives the highest gain.

6. NUMERICAL EXAMPLES

In this section the approaches presented in this paper are illustrated with numerical examples.

6.1 LTI system

In this example SGM is implemented on an LTI system, given by

$$G(q) = \frac{0.445q^{-1} + 0.4299q^{-2}}{1 - 1.03q^{-1} + 0.9048q^{-2}}. \quad (13)$$

This system has $\mathcal{L}_2$ -gain given by 9.6 which corresponds to the frequency 1 rad/s.

SGM: Consider the system (13) and let $\alpha = 10$. Here the output measurements are corrupted with white noise with variance $\lambda_e = 20$. SGM is implemented to estimate the gain of the system. The step length is chosen as $\gamma_k = 1/\sqrt{k}$ and 40 iterations are performed. In order to study the average performance, 1000 realizations of the initial input signal are made. The result is illustrated in Table 1. The result from the last experiment is plotted in Figure 2. Also, SGM is compared to the approach where $\hat{\lambda}^{\max, d}$ is computed as in (7). The result is shown in Table 1.

	SGM (no noise)	SGM	using (7)
mean	9.3246	9.3086	13.6426
std dev	10^{-13}	0.2342	0.3783

Table 1. The mean and standard deviation of the gain estimate after 40 iterations resulting from 1000 different realizations of the initial input signal.

Note that the gain of the true system, *i.e.* 9.6 cannot be obtained with a finite data set. It is clear from Table 1 that (7) is not suitable in the case of noisy data, since it results in an overestimation of the gain. SGM gives an accurate estimate of the gain.

Fig. 2. Example in Section 6.1: SGM subject to noisy measurements. Gain estimate versus the number of iterations from one experiment. Noisy measurements: dashed line. The approach where (7) is used to estimate the gain: $-\ast-$. Noise free measurements: solid line. True gain of the system: dash-dotted line.

6.2 Hammerstein system

Consider the Hammerstein system (11) where the transfer function $G(q)$ is given by (13) and the static nonlinear system f is given by

$$f(u) = \begin{cases} 2u - 0.1u^5 + 2u^3, & |u| \leq 4.4271, \\ 12.3317, & u > 4.4271, \\ -12.3317, & u < -4.4271. \end{cases} \quad (14)$$

This nonlinearity is depicted in Figure 3. An upper bound for the gain of the Hammerstein system can be constructed as following. For the system (11) we have $\beta_1 = 12$ (see Figure 3) and $\beta_2 = 9.6$ and thus the upper bound is $\beta_1\beta_2 = 115.2$.

Fig. 3. Example in Section 6.2. Solid line: the nonlinear function (14). Dashed line: upper/lower bound for the gain.

PIM: Now PIM is applied to the system. The initial input is uniformly distributed white noise with length $N = 1000$. We have tried different norms and realizations for this input signal. The result is that the gain estimates depend heavily on the realization and the norm of the initial input signal. The best gain estimate obtained is 37, which is far from the upper bound 115.2. Also, the gain estimates are not monotonically increasing as a function of the number of iterations. It is concluded that PIM is not suitable for this Hammerstein system.

SGM for Hammerstein systems: Let us instead use SGM tailored for Hammerstein systems, given in Section 4. The step length is chosen as $\gamma_k = 0.1/\sqrt{k}$. The resulting gain estimate is plotted in Figure 4 for 2 different initial input signals. From this figure it is concluded that the gain estimate depends on the initial input signal.

Fig. 4. Example in Section 6.2. Gain estimate versus the number of iterations using SGM. Solid line: $\alpha = 77.4$ (white noise initial input). Dashed line: $\alpha = 50$ (white noise initial input). Dash-dotted line: Upper bound ($\beta_1\beta_2 = 115.2$).

7. CONCLUSION

In this paper different non-standard approaches to $\mathcal{L}_2$ -gain estimation have been presented. Since the true system is in practise unknown the gain estimation procedure is data based in the sense that input/output measurements are used to either identify a model of the system or to draw conclusions about suitable inputs for gain estimation. If a model of the system is available then state of the art methods can be used to obtain the gain of the model, which is an approximation of the system gain. If no model of the system is available, we present methods where the system itself can generate suitable inputs for gain

estimation. We have presented a data-driven method for input sequence estimation. The derivation was based on linear system properties, and there is no theoretical proof for convergence for nonlinear systems. The simulation examples indicated, however, that the approach could be used on nonlinear systems. Gain estimation is only one problem where the estimation problem can be considerably simplified if the proper input sequence is used. The recent PhD thesis, Mårtensson [2007], contains an interesting discussion on how the input signal can be used to improve estimation of certain general properties of the system.

REFERENCES

- Barenthin, M., H. Moskuil, H. Hjalmarsson, and B. Wahlberg. Validity of stability for an induction machine drive using power iterations. In: *Proceedings of the 16th IFAC World Congress*. Prague, Czech Republic, 2005.
- Golub, G. H. and C. F. van Loan. *Matrix Computations*. The Jon Hopkins University Press. Baltimore, 1983.
- Hjalmarsson, H. From experiment design to closed-loop control. *Automatica* 41(3), pages 393-438, 2005.
- Kailath, T. *Linear systems*. Prentice Hall. Englewood Cliffs, New Jersey, 1980.
- Khalil, H.K. *Nonlinear systems, 3d edition*. Prentice Hall. Upper Sadle River, New Jersey, 2002.
- Kosut, R.L. Uncertainty model satisfaction: a system identification paradigm compatible with robust control design. In: *Proceedings of the 34th IEEE Conference on Decision and Control*, 1995.
- Ljung, L. *System Identification - Theory for the User, 2nd ed.* Prentice Hall. Upper Sadle River, New Jersey, 1999.
- Markusson, O. Model and System Inversion with Applications in Nonlinear System Identification and Control. PHD Thesis. KTH Royal Institute of Technology. TRISTA-S3-REG-0201, 2002.
- Mårtensson J. Geometric analysis of stochastic model errors in system identification. PHD Thesis. KTH Royal Institute of Technology. TRITA-EE 2007:061, 2007.
- Massoumina, M.A. and R.L. Kosut. A family of norms for system identification problems. *IEEE Transactions on Automatic Control* 39(5), pages 1027-1031, 1994.
- Moskuil H., B. Wahlberg and J. Galic. Validation of stability for an induction machine drive. In: *13th IFAC Symposium on System Identification, Preprints*. Rotterdam, The Netherlands, pages 1460-1465 2003.
- Poolla K., P. P. Khargonnekar, A. Tikku and K. Nagpal. A time-domain approach to model validation. *IEEE Transactions on Automatic Control* 39(5), pages 951-959, 1994.
- Reinelt W., A. Garulli and L. Ljung. Comparing different approaches to model error modelling in robust identification. *Automatica* 38(4), pages 787-803, 2002.
- Schoukens, J., R. Pintelon and T. Dobrowiecki. Identification of the stability of feedback systems in the presence of nonlinear distortion. In: *Proceedings of the 15th IFAC World Congress*. Barcelona, Spain, 2002.
- van Heusden, K., 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*, New Orleans, USA, 2007.
- Norrlof, M. Iterative Learning Control Analysis, Design and Experiments Ph.D. thesis, Linköping University, 2000
- Moore, K. Applied and computational controls, signal processing and circuits Springer-Verlag, 1999