

# Bounding L2 Gain System Error Generated by Approximations of the Nonlinear Vector Field

Kin Cheong Sou

Department of Electrical Engineering  
and Computer Science  
Massachusetts Institute of Technology  
Cambridge MA, 02139  
Email: kcsou@mit.edu

Alexandre Megretski

Department of Electrical Engineering  
and Computer Science  
Massachusetts Institute of Technology  
Cambridge MA, 02139  
Email: ameg@mit.edu

Luca Daniel

Department of Electrical Engineering  
and Computer Science  
Massachusetts Institute of Technology  
Cambridge MA, 02139  
Email: luca@mit.edu

**Abstract**—Typical nonlinear model order reduction approaches need to address two issues: reducing the order of the model, and approximating the vector field. In this paper we focus exclusively on the second issue, and present results characterizing the repercussions at the system level of vector field approximations. The error assessment problem is formulated as the L2 gain upper bounding problem of a scaled feedback interconnection. Applying the small gain theorem in the proposed setup, we prove that the L2 gain of the error system is upper bounded by the L2 gain of the vector field approximation error, provided it is small. In addition, the paper also presents a numerical procedure, based on the IQC/LMI approach, to perform the error estimation task with less conservatism. A numerical example is given in this paper to demonstrate the practical implications of the presented results.

## I. INTRODUCTION

A growing number of results can be found in the literature addressing the problem of nonlinear model order reduction. For example, [1]–[6] employ Volterra series and moment matching techniques to solve the “weakly nonlinear” model order reduction problem. Another class of methods based on piecewise approximations address strongly nonlinear problems [7]–[13]. Most of such methods can be viewed as a two steps procedure: first some projection is performed, and then (optionally) the reduced nonlinear vector field is further approximated to facilitate simulation of the reduced model. However, to the best of our knowledge, there has been no published result on the approximation quality of the second step above. This paper presents an effort in this direction for some specific dynamical system settings. Specifically, consider the following system:

$$\begin{aligned}\dot{x}(t) &= Ax(t) - \Phi(x(t)) + Bu(t) \\ y(t) &= Cx(t)\end{aligned}\quad (1)$$

where  $A \in \mathbb{R}^{q \times q}$ ,  $B \in \mathbb{R}^{q \times 1}$ ,  $C \in \mathbb{R}^{1 \times q}$  and  $\Phi: \mathbb{R}^q \mapsto \mathbb{R}^q$  is a general “reduced” vector field. For example,  $\Phi(\cdot) = V' \Phi_f(V \cdot)$  for some projection matrix  $V \in \mathbb{R}^{n \times q}$  (e.g. see [14]) and  $\Phi_f: \mathbb{R}^n \mapsto \mathbb{R}^n$  is the original (full order i.e.,  $n \gg q$ ) nonlinear vector field. As mentioned before, when the reduced nonlinear vector field  $\Phi$  is replaced by an approximation  $\tilde{\Phi}$ , the system (1) becomes

$$\begin{aligned}\dot{x}(t) &= Ax(t) - \tilde{\Phi}(x(t)) + Bu(t) \\ y(t) &= Cx(t).\end{aligned}\quad (2)$$

The objective of this paper is to relate the error between nonlinear functions  $\Phi$  and  $\tilde{\Phi}$ , to the error between systems (1) and (2), in the sense of the worst case system output difference when both systems are subjected to the same input.

### A. A motivating application

This subsection presents a specific (but more restrictive) application to illustrate why an approximation such as (2) is useful, and why it would be interesting to provide a bound for the induced system error. Consider the more specific setup

$$\begin{aligned}\dot{x}(t) &= Ax(t) - V' \Phi_f(Vx(t)) + Bu(t) \\ y(t) &= Cx(t)\end{aligned}\quad (3)$$

where  $A \in \mathbb{R}^{q \times q}$ ,  $V \in \mathbb{R}^{n \times q}$ ,  $B \in \mathbb{R}^{q \times 1}$ ,  $C \in \mathbb{R}^{1 \times q}$ ,  $\Phi_f: \mathbb{R}^n \mapsto \mathbb{R}^n$ :

$$\Phi_f(v) = [\phi_f(v_1) \quad \phi_f(v_2) \quad \cdots \quad \phi_f(v_n)]',$$

where  $\phi_f: \mathbb{R} \mapsto \mathbb{R}$  is any nonlinear function. Note that system (3) has repeated nonlinearities, and it can model for instance any circuit with repeated nonlinear elements, such as the diode transmission line to be discussed in Section VI. Furthermore, the method in this example can be modified by appending the nonlinear function  $\Phi_f$  with different nonlinearities, at the expense of a more complicated derivation and computation. However, it should be emphasized that the mentioned restriction in system (3) pertains only to this example, and not to the main result of this paper.

System (3) can be considered as the result of applying for instance a congruence transformation on a model of order  $n$  using projection matrix  $V$ , where  $n$  and  $q$  are the orders of the full and reduced models respectively,  $n \gg q$ . A common complaint about the applicability of such a model is that when using the model in simulation, the nonlinear function  $\phi_f$  must be evaluated  $n$  times for every reduced vector field evaluation. Therefore finding an approximation function  $g: \mathbb{R}^q \mapsto \mathbb{R}^q$ , such that  $g(w) \approx V' \Phi_f(Vw)$ ,  $\forall w \in \mathbb{R}^q$ , with an evaluation cost much cheaper than  $O(n)$ , would be of great

interest for most nonlinear model order reduction techniques. A few results can be found about this topic. For example, [15] investigated the possibility of using Kernel methods for such a construction, while [16], [17] proposed methods based on polynomial (Taylor series) approximation of  $V'\Phi_f(V\cdot)$ .

However, when considering the special case (3), it would be much more convenient to find an approximation to the scalar nonlinear function  $\phi_f$ , instead of the entire vector field. For example, if  $\phi_f$  is approximated by a scalar polynomial of degree  $d$ ,

$$\phi_f(z) \approx \tilde{\phi}_f(z) = \sum_{k=0}^d p_k z^k, \quad (4)$$

and accordingly

$$\Phi_f(v) \approx \tilde{\Phi}_f(v) := \begin{bmatrix} \tilde{\phi}_f(v_1) \\ \tilde{\phi}_f(v_2) \\ \vdots \\ \tilde{\phi}_f(v_n) \end{bmatrix}, \quad (5)$$

then the corresponding vector field approximation is a  $q$ -variate polynomial of degree  $d$

$$V'\Phi_f(Vx) \approx V'\tilde{\Phi}_f(Vx) = \sum_{\beta} c_{\beta} x^{\beta}, \quad (6)$$

where  $\beta \in \mathbb{Z}_+^q$ ,  $\beta = (\beta_1, \beta_2, \dots, \beta_q)$ ,  $\sum_j \beta_j \leq d$ ,  $c_{\beta} \in \mathbb{R}^q$  and  $x^{\beta}$  is shorthand for  $\prod_j x_j^{\beta_j}$ . The approximated system becomes

$$\begin{aligned} \dot{x}(t) &= Ax(t) - V'\tilde{\Phi}_f(Vx(t)) + Bu(t) \\ y(t) &= Cx(t) \end{aligned} \quad (7)$$

The above polynomial approximation scheme has the following benefits:

- 1) Approximating a scalar nonlinear function  $\phi_f$  is much easier than approximating the vector-valued nonlinear function  $V'\Phi_f(V\cdot)$ .
- 2) It can be verified that the coefficient vectors  $c_{\beta}$  can be computed efficiently.
- 3) The Jacobian of the approximated vector field is

$$A - V' \text{diag} \left( \frac{d\tilde{\phi}_f}{dz}, \dots, \frac{d\tilde{\phi}_f}{dz} \right) V. \quad (8)$$

If  $A$  is symmetric and Hurwitz, the Jacobian can be constrained to be Hurwitz simply by constraining the univariate polynomial  $\frac{d\tilde{\phi}_f}{dz}$  to be nonnegative, which is true if and only if it is a sum of squares of polynomials, and this condition can in turn be efficiently enforced using linear matrix inequalities (LMI) [18].

However, there are two issues that are worth considering:

- Estimating and controlling the cost of evaluating the polynomial approximated vector field.
- Providing precise statements about the accuracy of the approximation quality in terms of quantifiable system measures such as the L2 gain of the difference of systems (3) and (7).

The answer to the first question depends on the specific application. The computation cost for evaluating nonlinear vector field  $V'\tilde{\Phi}_f(V\cdot)$  is  $O\left(q \binom{q+d}{d}\right)$ . Since such cost is independent of  $n$ , and since typically  $n \gg \max\{q, d\}$ , computation efficiency is greatly improved. However, as also pointed out in [15],  $\binom{q+d}{d}$  is admittedly still a large number even for not excessively large  $q$  and  $d$ . Measures should be taken to control computational complexity, but this will not be discussed here, as it is not the main focus of this paper.

Instead, this paper presents results that address the second issue: providing statements about the accuracy of the approximation. In particular, under the assumptions that system (3) has finite incremental L2 gain from  $u$  to  $y$ , and stability, it will be shown that the L2 gain from input  $u$  to the difference of output  $y$  of systems (3) and (7) is bounded by a linear function of the L2 gain of the difference of the scalar nonlinear functions  $\phi_f(\cdot) - \tilde{\phi}_f(\cdot)$ , if the difference is small enough. It should be emphasized that the result is true regardless of whether  $\tilde{\phi}_f$  is polynomial or not. In addition, the paper also presents a framework for numerically calculating an a priori (i.e. before simulation) error bound of the L2 gain of the difference system, again based on the L2 gain of  $\phi_f(\cdot) - \tilde{\phi}_f(\cdot)$ . Finally, it should be noted that the results of this paper are valid for a more general framework (1) than discussed in this motivating application subsection. Namely, the system error is presented in terms of (1) and (2), and the vector field approximation error is between  $\Phi$  and  $\tilde{\Phi}$ .

## B. The organization of the paper

The rest of the paper is organized as follows: Section II summarizes background results. In Section III the small gain theorem, the main theoretical apparatus in the paper, is employed, and a framework is proposed to broaden its use for the specific problem considered. Section IV presents the main theoretical contribution: under some assumptions, the L2 gain of the error system is upper bounded by the L2 gain of  $\Phi(\cdot) - \tilde{\Phi}(\cdot)$  with a positive multiplicative constant. Section V applies again the small gain theorem, and proposes a numerical procedure for computing an upper bound of the L2 gain of the error system (tighter than the theoretical bound in Section IV) based on the L2 gain of  $\Phi(\cdot) - \tilde{\Phi}(\cdot)$ . In Section V, the numerical procedure from Section V is applied to some nonlinear system model reduction problem to validate the statements.

## II. BACKGROUND

### A. L2 gain of a memoryless nonlinearity

Let  $u \in \mathbb{R}^m$  and  $y \in \mathbb{R}^p$  be the input and output to a memoryless nonlinearity  $F$ . That is  $y = F(u)$ . The L2 gain  $\gamma_F \geq 0$  of a memoryless nonlinearity  $F$  is defined as

$$\gamma_F := \sup_{u \neq 0} \frac{|F(u)|}{|u|} \quad (9)$$

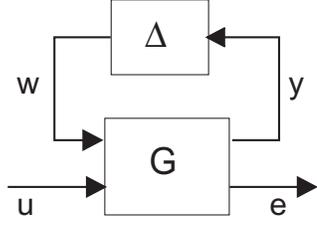


Fig. 1. Feedback interconnection of a nominal plant  $G$  and disturbance  $\Delta$ .

### B. L2 gain of a dynamical system

Let  $u : \mathbb{R}_+ \mapsto \mathbb{R}^m$  and  $y : \mathbb{R}_+ \mapsto \mathbb{R}^p$  denote the input and output of a dynamical system. The L2 gain  $\gamma \geq 0$  of a system is defined as

$$\gamma = \inf_{r \geq 0, T \geq 0} r : \int_0^T (r^2 |u(\tau)|^2 - |y(\tau)|^2) d\tau > -\infty. \quad (10)$$

for all valid input/output pairs  $(u, y)$ . For the rest of the paper, unless noted otherwise, L2 gain related integrals inequalities are assumed to hold for *all* valid input/output pairs. Also, for ease of notion, the upper limit of L2 gain related integral such as (10) will be shorthanded by  $+\infty$ , even though what we really mean with such notation is the infimum with respect to  $T$  as in (10).

Intuitively, finiteness of the L2 gain of a system means that the output energy is no more than a constant times the input energy, and hence the L2 gain can serve as a notion for stability. In addition, if the L2 gain is small, then the system can be considered “small”, in the sense that it needs a very strong input to excite any non-negligible output. In particular, it is desirable that an error system has very small L2 gain.

### C. Incremental L2 gain of a system

Let  $(u, y)$  be any input/output pair of a system. Then the system has a finite incremental L2 gain if  $\exists \gamma \geq 0$ :

$$\int_0^\infty (\gamma^2 (|u_1(\tau) - u_2(\tau)|^2) - (|y_1(\tau) - y_2(\tau)|^2)) d\tau \geq 0, \quad (11)$$

for every  $(u_1, y_1)$  and  $(u_2, y_2)$  satisfying

$$\int_0^\infty |u_1(\tau) - u_2(\tau)|^2 d\tau < \infty. \quad (12)$$

Incremental L2 gain of a system can be used to quantify the sensitivity of the output to a perturbation in the input. In particular, having a finite incremental L2 gain means that distinct inputs will produce distinct outputs.

### D. Small gain theorem

The small gain theorem is a collection of statements bounding the L2 gain of the feedback interconnection of a nominal model  $G$  and a disturbance  $\Delta$ , using the L2 gains of the individual constituents. See for example [19], for a more detailed account of these statements. The statement relevant to the discussion of the paper is the following.

*Theorem 2.1:* Consider the feedback connection in Figure 1. Let  $\gamma_G$  be the L2 gain of  $G$  (from  $[w; u]$  to  $[y; e]$ ), and  $\gamma_\Delta$  be the L2 gain of  $\Delta$  (from  $y$  to  $w$ ). If  $\gamma_G \gamma_\Delta \leq 1$  then the L2

gain of the feedback connection (from  $u$  to  $e$ ) is less than or equal to  $\gamma_G$ .

See, for example [19], for a proof. The small gain theorem is the fundamental tool upon which the main results of the paper are based. The discussion of how to apply the theorem in the context of the paper will be presented in Section III.

## III. APPLICATION OF THE SMALL GAIN THEOREM

This section first setups the L2 gain error bound problem as the L2 gain upper bounding problem of the difference system. The difference system is formulated as a feedback connection between a “nominal” plant that does not contain any approximation vector field, and the “disturbance” part consisting of the error of the vector fields. Then the first contribution of the paper is presented: a scaling parameter is introduced in the feedback loop to allow a more general use of the small gain theorem. Finally the ramification of the reformulations will be discussed.

### A. Feedback interconnection of the difference system

One way to characterize the error between systems (1) and (2) is to upper bound the L2 gain of the difference system (from  $u$  to  $e$ )

$$\begin{aligned} \dot{x}_1 &= Ax_1 + \Phi(x_1) + Bu \\ \dot{x}_2 &= Ax_2 + \tilde{\Phi}(x_2) + Bu \\ e &= C(x_1 - x_2). \end{aligned} \quad (13)$$

System (13) can equivalently be written as

$$\begin{aligned} \dot{x}_1 &= Ax_1 + \Phi(x_1) + Bu \\ \dot{x}_2 &= Ax_2 + \Phi(x_2) + Bu + w \\ e &= C(x_1 - x_2) \\ y &= x_2 \\ w &= \tilde{\Phi}(y) - \Phi(y). \end{aligned} \quad (14)$$

It can be seen that system (14) fits in the small gain theorem framework in Figure 1. In particular, system  $G$  in the figure corresponds to the part of system (14) with input/output  $[w; u]$  and  $[y; e]$  and the disturbance in the figure being  $\Delta(y) = \tilde{\Phi}(y) - \Phi(y)$ . Unfortunately the small gain theorem cannot be readily applied because the assumption  $\gamma_G \gamma_\Delta \leq 1$  might not be satisfied. More importantly, even if the assumption  $\gamma_G \gamma_\Delta \leq 1$  is satisfied, the L2 gain upper bound  $\gamma$  of the feedback connection (14) is lower bounded by  $\gamma_G$ , which is independent of  $\gamma_\Delta$ . In reality, it would be desirable if  $\lim_{\gamma_\Delta \rightarrow 0} \gamma = 0$ , since the L2 gain of the difference of two identical systems should be zero. This latter difficulty can be resolved through the use of a scaling parameter discussed in the next subsection.

### B. Small gain theorem applied to a scaled feedback

Consider Figure 2, which is equivalent to Figure 1. For the rest of the paper  $a$  is assumed to be nonnegative. System  $G_a$  in the figure has the form

$$\begin{aligned} \dot{x}_1 &= Ax_1 + \Phi(x_1) + Bu \\ \dot{x}_2 &= Ax_2 + \Phi(x_2) + Bu + \sqrt{a}w \\ e &= C(x_1 - x_2) \\ y &= \sqrt{a}x_2. \end{aligned} \quad (15)$$

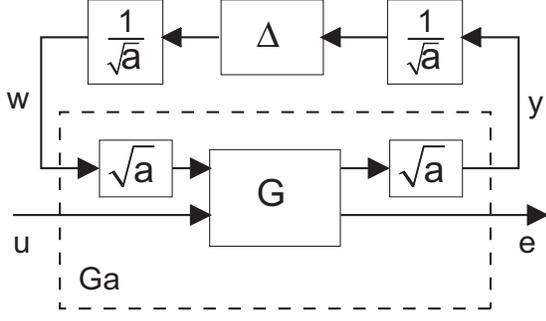


Fig. 2. Feedback interconnection of a nominal plant  $G$  and disturbance  $\Delta$  with self-cancelling parameter  $\sqrt{a}$  and  $\frac{1}{\sqrt{a}}$ .  $G_a$  is the original plant parameterized by the scalar  $a$ .

The statement of the small gain theorem becomes the following. Let  $\gamma_{G_a}$  be the L2 gain of system (15) (from  $[u; w]$  to  $[e; y]$ ). If  $\frac{\gamma_{G_a}\gamma_\Delta}{a} \leq 1$ , then the L2 gain of the feedback interconnection (14), from  $u$  to  $e$ , is  $\gamma \leq \gamma_{G_a}$ .

As will be shown in Section IV, under some assumptions,

$$\exists c \geq 0 : \gamma_{G_a} \leq c\sqrt{a}, \quad \forall a \leq 1, \quad (16)$$

On the other hand,  $\gamma_{G_a}$  increases with  $a$  since if  $\gamma_{G_a}$  is the infimum:

$$\int_0^\infty (\gamma_{G_a}^2 (|w(\tau)|^2 + |u(\tau)|^2) - (|y(\tau)|^2 + |e(\tau)|^2)) d\tau > -\infty,$$

then  $\forall r \in (0, 1)$ ,

$$\int_0^\infty \left( \gamma_{G_a}^2 \left( \frac{|w(\tau)|^2}{r^2} - |u(\tau)|^2 \right) - (r^2|y(\tau)|^2 + |e(\tau)|^2) \right) d\tau > -\infty,$$

which means that  $\gamma_{G_a} \geq \gamma_{G_{\tilde{a}}}$  whenever  $a > \tilde{a}$  and this in turn shows that  $\gamma_{G_a}$  increases with  $a$ . The nonlinear functional dependency of  $\gamma_{G_a}$  on  $a$  has the following ramifications:

- If  $a > 1$ , then inequality (16) does not hold, hence in this case unfortunately the theory proposed does not guarantee the application of the small gain theorem.
- If  $a \leq 1$ , inequality (16) holds, it would be desirable (as guided by the small gain theorem) that  $a$  be as small as possible for the purpose of obtaining the tightest upper bound for the L2 gain of the error system. However, since it can only be established that  $\gamma_{G_a} < c\sqrt{a}$  (in fact, it is reasonable that  $\lim_{a \rightarrow 0} \frac{\gamma_{G_a}}{a} = \infty$ ), the upper bound  $\frac{c\gamma_\Delta}{\sqrt{a}}$  goes to infinity as  $a \rightarrow 0$ . Therefore it cannot be certified whether  $\frac{\gamma_{G_a}\gamma_\Delta}{a} \leq 1$  holds, and the small gain theorem fails to apply. Therefore, the minimum error system L2 gain upper bound that can be asserted is

$$\gamma_{G_a} : a = \inf_{\alpha > 0} \alpha : \frac{\gamma_{G_a}\gamma_\Delta}{\alpha} \leq 1 \quad (17)$$

In other words, there is a tradeoff between making  $a$  small to get tighter L2 gain upper bound, and making  $a$  large enough so that small gain theorem still applies.

- When  $a = 0$ , system  $G_a$  in (15) has zero L2 gain as it is the difference of two identical systems. Additionally, in the extreme case where  $\gamma_\Delta = 0$  (or  $\tilde{\Phi} = \Phi$ ), the infimum

$\alpha$  in (17) is 0. Hence the upper bound for the L2 gain of the error system is as expected zero.

The arguments given in this section can be summarized to formulate the main theorem in the next section.

#### IV. A THEORETICAL LINEAR ERROR BOUND IN THE LIMIT

This section presents the paper's main contribution: a relation between the L2 gain of the error in nonlinear vector field approximation, and the L2 gain of the error system. The development of the result can be divided into two parts: first, under some assumptions, the inequality

$$\exists c \geq 0 : \gamma_{G_a} \leq c\sqrt{a}, \quad \forall a \leq 1 \quad (18)$$

will be proved. Then the main theorem is given as a direct consequence of (18).

##### A. A preliminary lemma

To prove the first part, consider the system with input  $g$  and output  $z$

$$\begin{aligned} \dot{x} &= Ax + \Phi(x) + g \\ z &= Cx, \end{aligned} \quad (19)$$

where the matrices and functions are as defined in (13), except for the arbitrary function  $g$ . Define

$$\begin{aligned} \gamma_1 &\text{ as the incremental L2 gain of (19) from } g \text{ to } z, \\ \gamma_2 &\text{ as L2 gain of (19) from } [u; w] \text{ to } x \text{ when } g \equiv Bu + w, \end{aligned} \quad (20)$$

*Lemma 4.1:* Let  $\gamma_1$  and  $\gamma_2$  be the quantities defined in (20). Denote  $\gamma_{G_a}$  as the L2 gain of system  $G_a$  (15), from  $[u; w]$  to  $[e; y]$ .

If  $\gamma_1 < \infty$  and  $\gamma_2 < \infty$ , then

$$\gamma_{G_a} \leq \sqrt{2a} \max\{\gamma_1, \gamma_2\}, \quad \forall a \in [0, 1] \quad (21)$$

**Proof:** First let

$$\begin{aligned} g_1 &:= B\tilde{u}, \\ g_2 &:= B\tilde{u} + \tilde{w} \end{aligned}$$

be two inputs to system (19) and  $z_1$  and  $z_2$  be the corresponding outputs.  $\gamma_1 < \infty$  implies that for the system

$$\begin{aligned} \dot{x}_1 &= Ax_1 + \Phi(x_1) + B\tilde{u} \\ \dot{x}_2 &= Ax_2 + \Phi(x_2) + B\tilde{u} + \tilde{w} \\ \tilde{e} &= C(x_1 - x_2) \end{aligned}$$

the following integral inequality holds

$$\int_0^\infty (\gamma_1^2 |B\tilde{u} - B\tilde{u} - \tilde{w}|^2 - |\tilde{e}|^2) d\tau > -\infty,$$

which implies,  $\forall a > 0$ ,

$$\int_0^\infty \left( a\gamma_1^2 \left( |\tilde{u}|^2 + \frac{1}{a} |\tilde{w}|^2 \right) - |\tilde{e}|^2 \right) d\tau > -\infty,$$

or

$$\int_0^\infty (a\gamma_1^2 (|u|^2 + |w|^2) - |e|^2) d\tau > -\infty, \quad (22)$$

when  $u = \tilde{u}$ ,  $w = \frac{1}{\sqrt{a}}\tilde{w}$ , and  $e = \tilde{e}$ . That shows that the system

$$\begin{aligned} \dot{x}_1 &= Ax_1 + \Phi(x_1) + Bu \\ \dot{x}_2 &= Ax_2 + \Phi(x_2) + Bu + \sqrt{a}w \\ e &= C(x_1 - x_2) \end{aligned}$$

has L2 gain from  $[u; w]$  to  $e$  less than or equal to  $\sqrt{a}\gamma_1$ . This means that system  $G_a$  (15) has L2 gain from  $[u; w]$  to  $e$  is less than or equal to  $\sqrt{a}\gamma_1$ .

Secondly, for system (19), let  $g = B\tilde{u} + \tilde{w}$ . Then  $\gamma_2 < \infty$  implies in the following system

$$\begin{aligned} \dot{x}_1 &= Ax_1 + \Phi(x_1) + B\tilde{u} \\ \dot{x}_2 &= Ax_2 + \Phi(x_2) + B\tilde{u} + \tilde{w} \\ \tilde{y} &= x_2 \end{aligned}$$

the following inequality holds

$$\int_0^\infty (\gamma_2^2 (|\tilde{u}|^2 + |\tilde{w}|^2) - |\tilde{y}|^2) d\tau > -\infty,$$

which implies,  $\forall a \in (0, 1]$ ,

$$\int_0^\infty \left( a\gamma_2^2 \left( |\tilde{u}|^2 + \frac{1}{a}|\tilde{w}|^2 \right) - a|\tilde{y}|^2 \right) d\tau > -\infty, \quad (23)$$

Note that the fact that  $\frac{1}{a} \geq 1$  for  $a \leq 1$  was indeed used. Rewrite the signals in (23) in terms of the signals in (15). That is,  $u = \tilde{u}$ ,  $w = \frac{1}{\sqrt{a}}\tilde{w}$  and  $y = \sqrt{a}\tilde{y}$ . This results in the following inequality:

$$\int_0^\infty (a\gamma_2^2 (|u|^2 + |w|^2) - |y|^2) d\tau > -\infty, \quad (24)$$

which means that the L2 gain of system  $G_a$  in (15) from  $[u; w]$  to  $y$  has L2 gain less than or equal to  $\sqrt{a}\gamma_2$ .

(22) together with (24) implies that, in terms of the quantities associated with  $G_a$  in (15), the following integral

$$\int_0^\infty (2a(\max\{\gamma_1, \gamma_2\})^2 (|w|^2 + |u|^2) - (|y|^2 + |e|^2)) d\tau \quad (25)$$

is bounded from below for all input/output pair of  $G_a$  and this proves (21)  $\forall a \in (0, 1]$ . For the case of  $a = 0$ ,  $\gamma_1 < \infty$  implies  $\gamma_{G_a}|_{a=0} = 0$ , so (21) also holds in this case.

Q.E.D.

### B. The main result

Using Lemma 4.1, the main result is now presented.

*Theorem 4.1:* Let  $\gamma_1$  and  $\gamma_2$  be the quantities defined in (20). Also let  $\gamma_\Delta$  be the L2 gain of  $\Phi - \tilde{\Phi}$  in (14). That is,

$$\gamma_\Delta := \sup_{v \neq 0} \frac{|\Phi(v) - \tilde{\Phi}(v)|}{|v|}$$

Denote  $\gamma$  as the L2 gain from  $u$  to  $e$  in system (13).

If  $\gamma_1 < \infty$ ,  $\gamma_2 < \infty$  and  $\sqrt{2}\max\{\gamma_1, \gamma_2\}\gamma_\Delta \leq 1$ , then

$$\gamma \leq 2(\max\{\gamma_1, \gamma_2\})^2 \gamma_\Delta. \quad (26)$$

**Proof:** If  $\gamma_\Delta = 0$ , then by the finiteness of  $\gamma_1$ ,  $\gamma = 0$  and hence (26) holds. Now consider the case when  $\gamma_\Delta > 0$ , the small gain theorem states that

$$\gamma \leq \gamma_{G_a}, \quad \forall a : \frac{\gamma_{G_a}\gamma_\Delta}{a} \leq 1.$$

Therefore,

$$\gamma \leq \min_{a: \frac{\gamma_{G_a}\gamma_\Delta}{a} \leq 1} \gamma_{G_a}. \quad (27)$$

Denote  $c := \sqrt{2}\max\{\gamma_1, \gamma_2\}$ . Since  $\gamma_1 < \infty$  and  $\gamma_2 < \infty$  by statement assumption, Lemma 4.1 states that  $\forall a \in (0, 1]$ ,

$$\gamma_{G_a} \leq c\sqrt{a} \quad \text{and hence} \quad \frac{\gamma_{G_a}\gamma_\Delta}{a} \leq \frac{c\gamma_\Delta}{\sqrt{a}}.$$

Since  $c\gamma_\Delta \leq 1$  by statement assumption, the set  $[c\gamma_\Delta, 1] \neq \emptyset$ .  $\exists a \in [c\gamma_\Delta, 1]$ :

$$(c\gamma_\Delta)^2 \leq a \leq 1$$

and hence

$$1 \geq \frac{c\gamma_\Delta}{\sqrt{a}} \geq \frac{\gamma_{G_a}\gamma_\Delta}{a}.$$

Therefore,

$$\gamma \leq \min_{a: \frac{\gamma_{G_a}\gamma_\Delta}{a} < 1} \gamma_{G_a} \leq \min_{a \geq c^2\gamma_\Delta^2} c\sqrt{a} = c^2\gamma_\Delta.$$

Q.E.D.

Intuitively, Theorem 4.1 asserts that if  $\gamma_\Delta$ , the L2 gain of the difference  $\Phi - \tilde{\Phi}$  (and also  $\phi - \tilde{\phi}$ ) is *sufficiently small*, then the approximation quality in terms of the L2 gain of the error system (13) is also small. In particular, it provides a guideline for designing the approximation system (2). It states that searching for a  $\tilde{\phi}$  that is close to  $\phi$  in L2 gain sense, should be a reasonable choice, as opposed to other methods such as Taylor Series, for which the accuracy has not been rigorously established. In addition, the linear error bound (27) can be used to guide the design of the vector field approximation in the following sense:

- Pick a desired system error  $\epsilon$ .
- Choose any available vector field approximation technique (not discussed in this paper).
- Obtain an approximated reduced system; compute the vector field L2 gain error, and the error system L2 gain, denoted as  $\epsilon_\Delta$  and  $\epsilon_1$  respectively.
- If  $\epsilon_1 < \epsilon$  then the desired approximated reduced system has already been obtained. Otherwise, obtain a better approximated system (e.g. by increasing polynomial order) so that the new vector field L2 gain error is less than  $\frac{\epsilon_\Delta \epsilon}{\epsilon_1}$ , then under the assumptions of Theorem 4.1, the new reduced model will satisfy the desired system error tolerance.

However, it should also be noted that Theorem 4.1 is somewhat conservative in the sense that (26) is not universally true for all  $\gamma_\Delta$ . Therefore, it would be interesting to see if there exists a less restrictive statement or a numerical procedure to compute a tighter bound. The result in the next section is an attempt to do so.

## V. A NUMERICAL ERROR BOUND WITH IQC

This section presents a numerical result for the following problem: given a system in the form of (1) and given any approximation  $\tilde{\Phi}$  for the nonlinear vector field, find an upper bound  $\gamma$  for the L2 gain of the difference system (13). This problem can partially be addressed by Theorem 4.1 in Section IV, provided that the coefficient in (26) can be estimated. However, the linear error bound obtained there can be conservative

because of the use of inequality (21). Therefore, a numerical result that can potentially provide a tighter upper bound is of interest.

The numerical procedure proposed in this section is also based on the application of the small gain theorem described in Subsection III-B. The idea is straightforward: instead of using inequality (21), we try to find a tighter upper bound for the L2 gain of  $G_a$  (15) using an IQC analysis procedure (35) to be discussed. The rest of this section is organized as follows: first the standard IQC analysis procedure is reviewed in Subsection V-A. Then the proposed numerical procedure is described in Subsection V-B.

#### A. Nonlinear system L2 gain upper bounding using integral quadratic constraints (IQC)

It should be noted that the materials in this subsection are standard. However, they have been modified as appropriate for the problem in the paper, and they are presented here in this subsection, instead of Section II, in order to maintain the flow of the paper.

Consider the system in (1). If  $\exists \gamma \geq 0$  and a continuously differentiable function  $W : \mathbb{R}^q \mapsto \mathbb{R}_+ : W(x) \geq 0, \forall x \in \mathbb{R}^q$  and the following inequality holds

$$\gamma^2 |u|^2 - |y|^2 - (\nabla_x W)' \dot{x} \geq 0, \quad \forall (x, u) \in \mathbb{R}^q \times \mathbb{R}, \quad (28)$$

then  $\forall T > 0$

$$\int_0^T (\gamma^2 |u|^2 - |y|^2) d\tau \geq W(x(T)) - W(x(0)) > -\infty, \quad (29)$$

and therefore  $\gamma$  is an upper bound for the L2 gain of system (1) and  $W$  is a certificate for proving the L2 gain upper bound. A class of nonnegative functions  $W(x)$  that is particularly useful is the quadratic function  $W(x) = x'Px$  for some symmetric positive semidefinite matrix  $P \in \mathbb{R}^{q \times q}$ . Quadratic certificate is interesting because the search of matrix  $P \geq 0$  can be performed (in principle) efficiently through the use of LMI optimization. That is, if  $W(x) = x'Px$  then (28) becomes

$$\gamma^2 |u|^2 - |Cx|^2 - 2x'P(Ax - V'w + Bu) \geq 0, \quad (30)$$

required to be true  $\forall (x, u) \in \mathbb{R}^q \times \mathbb{R}$  and  $w = \Phi(Vx)$ . For a general nonlinear vector field  $\Phi$ , showing the existence of  $P \geq 0$  and  $\gamma$  that satisfy inequality (30) is difficult. However, the technique of integral quadratic constraints [20] can be employed here: first we introduce a quadratic functional  $\sigma(x, w)$  that satisfies the following property

$$\sigma(x, w) = \begin{bmatrix} x \\ w \end{bmatrix}' \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix} \geq 0 \quad \text{if } w = \Phi(Vx), \quad (31)$$

then we remove the constraint  $w = \Phi(Vx)$  in (30) and solve instead the following:

$$\gamma^2 |u|^2 - |Cx|^2 - 2x'P(Ax - V'w + Bu) - \sigma(x, w) \geq 0, \quad \forall x, w, u \quad (32)$$

Note that if  $\gamma$  and  $P$  satisfy (32) then they automatically satisfy (30) by the definition of  $\sigma$  (31). But the converse is not necessarily true, therefore searching for  $\gamma$  and  $P$  through

(32) results in fewer options. However, (32) has the advantage that it is equivalent to the following LMI feasibility problem (with respect to  $P$  and  $r := \gamma^2$ )

$$\begin{bmatrix} -C'C - PA - A'P - \Sigma_{11} & PV' - 0.5\Sigma_{12} & -PB \\ VP - 0.5\Sigma_{12}' & -\Sigma_{22} & 0 \\ -B'P & 0 & rI \end{bmatrix} \geq 0. \quad (33)$$

More generally, if there exist more quadratic functionals  $\sigma_1, \sigma_2, \dots$  such that

$$w = \Phi(Vx) \Rightarrow \sigma_i(x, w) \geq 0, \quad \forall i,$$

then optimizing the LMI problem (with decision variables  $P, r, \tau_i \geq 0$ )

$$\gamma^2 |u|^2 - |Cx|^2 - 2x'P(Ax - V'w + Bu) - \sum_i \tau_i \sigma_i(x, w) \geq 0, \quad \forall x, w, u \quad (34)$$

would result in a less conservative search than (32) because if  $(r, P)$  satisfy (32) then they also satisfy (34), simply by picking  $\tau_j = 0, j \geq 2$ . However, the converse is not necessarily true. Note also that (34) is more restrictive than (30) for the same reason mentioned in the case of a single  $\sigma$ .

In summary, in order to find an upper bound of the L2 gain of a system of the form (1). The following procedure can be used: first collect characterizations of the nonlinearity  $\Phi$  in the form of IQCs  $\sigma_1, \sigma_2, \dots$ , then setup and solve the following LMI

$$\begin{aligned} & \text{minimize } r \\ & \quad r, P, \tau_i \geq 0 \\ & \text{subject to } \text{LMI (34)} \\ & \quad r \geq 0 \\ & \quad P = P' \geq 0. \end{aligned} \quad (35)$$

Note that the L2 gain upper bound provided by such a procedure can be strictly greater than the true L2 gain because the class of certificates is restricted to quadratic (which is generally not rich enough except for the LTI case). Furthermore, inequalities such as (34) do not allow all the options (in terms of  $r$  and  $P$ ) that satisfy (30). Nevertheless, this is a practical method for nonlinear system L2 gain upper bounding because of its tractability.

#### B. The numerical procedure

Our proposed numerical procedure is as follows.

- For a discrete set of  $\{a_1, a_2, \dots\}$  (e.g.  $a_k := 10^{-k}$ ), use IQC analysis to find  $\gamma_1, \gamma_2, \dots$  as the L2 gain upper bounds for the parameterized systems  $G_{a_1}, G_{a_2}, \dots$
- For any approximation vector field  $\tilde{\Phi}$ , evaluate the L2 gain of  $\Phi - \tilde{\Phi}$ . Denote it as  $\gamma_\Delta$ . Find the index  $i$  such that

$$i = \underset{k}{\operatorname{argmin}} a_k : \frac{\gamma_k \gamma_\Delta}{a_k} < 1$$

- $\gamma_i$  is returned as the upper bound of the L2 gain of the difference system (13).

Since the order of system  $G_a$  (15) is  $2q$  and  $q$  is assumed to be small, solving the LMIs to obtain L2 gain upper bounds  $\gamma_k$  for all  $a_k$  is relatively cheap. Once the L2 gain upper bounds  $\gamma_1, \gamma_2, \dots$  have been found, the numerical procedure requires a

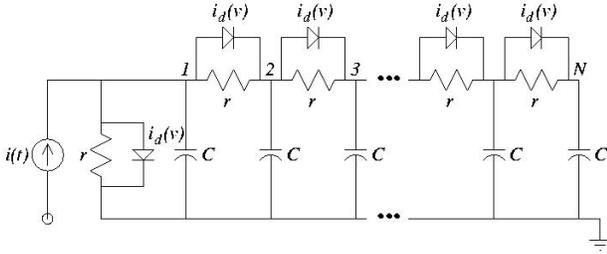


Fig. 3. A transmission line with diodes.

trivial amount of time to analyse the system L2 gain error for all  $\tilde{\Phi}$  such that  $\gamma_{\Delta}$  is small enough. As a final note, it should be pointed out that since the numerical procedure is based on the small gain theorem, it is possible that when  $\gamma_{\Delta}$  is large, the procedure fails to return any conclusive result.

## VI. NUMERICAL EXPERIMENT

In this section the numerical procedure described in Section V is applied to analyze the L2 gain of the error system due to approximation of the nonlinear vector field. The specific application example is a transmission line with diodes described in [14] and shown in Figure 3. Using nodal analysis, the model of the diode line has the form

$$\begin{aligned} \dot{x}_f &= A_f x_f - M' \Phi(M x_f) + B_f u \\ y_f &= C_f x_f, \end{aligned}$$

with

$$\begin{aligned} A_f &\in \mathbb{R}^{N \times N}, M \in \mathbb{R}^{N \times N}, B_f \in \mathbb{R}^{N \times 1}, C_f \in \mathbb{R}^{1 \times N}, \\ \Phi(v) &= \text{diag}(\phi(v_1), \phi(v_2), \dots) \quad \text{and} \quad \phi(v_k) = e^{-v_k} - 1, \end{aligned}$$

with  $M$  being a sparse matrix relating branch voltages to node voltages. Suppose there exists a projection matrix  $V \in \mathbb{R}^{N \times q}$  (e.g. dominant singular vectors of some matrix stacked by columns of trajectories), then the “reduced” model is

$$\begin{aligned} \dot{x} &= A_r x - V_r' \Phi(V_r x) + B_r u \\ y &= C_r x, \end{aligned} \quad (36)$$

with  $A_r = V' A_f V$ ,  $V_r = M V$ ,  $B_r = V' B_f$  and  $C_r = C V$ . System (36) is of the form of (1), hence the numerical procedure described in Section V can be applied.  $\gamma_{G_a}$  and  $\frac{a}{\gamma_{G_a}}$  are plotted in Figure 4 for a range of values of  $a$ .

In this figure, the uppermost line (dots) is  $\gamma_{G_a}|_{a=1} \sqrt{a}$  and represents a theoretical upper bound in the spirit of (26). The middle line (circles) is  $\gamma_{G_a}$  that is used as the upper bound for the L2 gain of error system (13). The bottom line (triangles) is the quantity  $\frac{a}{\gamma_{G_a}}$  used in determining the minimum  $a$ , for a specific  $\gamma_{\Delta}$ , such that  $\frac{\gamma_{\Delta} \gamma_{G_a}}{a} \leq 1$  (hence the small gain theorem applies).

As an example to illustrate how Figure 4 can be applied, let the desired system level error be 1% or less. By the small gain theorem, if  $\gamma_{G_a} < 1\%$  then the accuracy is achieved. According to Figure 4, the maximum allowable  $a$  for the small gain theorem to be applicable is about  $2 \times 10^{-5}$  (the  $x$  coordinate where horizontal  $y = 10^{-2}$  intersects the middle line). For  $a = 2 \times 10^{-5}$ , the corresponding value of  $\frac{a}{\gamma_{G_a}}$  is about  $10^{-3}$ ,

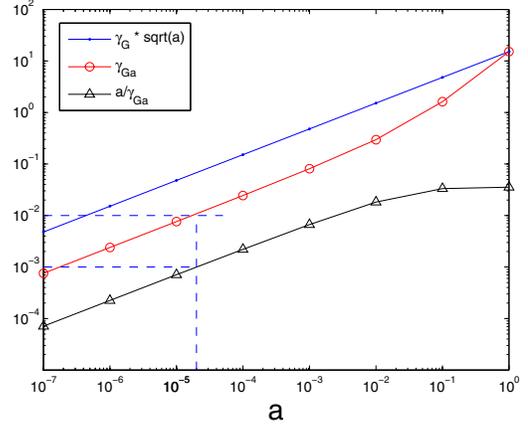


Fig. 4. Transmission line example. The uppermost line (dots) is the theoretical upper bound (26). The middle line (circles) is the numerical upper bound for the L2 gain of the error system. The lower line (triangles) is the minimum allowable  $a$  such that  $\frac{\gamma_{\Delta} \gamma_{G_a}}{a} < 1$ , and hence the small gain theorem still applies. For instance, if we want the system L2 gain error to be less than  $10^{-2}$ , then  $a$  should be at most  $2 \times 10^{-5}$ , corresponding to a maximum allowable vector field error  $\gamma_{\Delta}$  of about  $10^{-3}$ .

which means that the vector field L2 gain error  $\gamma_{\Delta}$  should be at most  $10^{-3}$ .

## VII. CONCLUSION

This paper investigates the estimation of the L2 gain system error produced by the approximation of the nonlinear vector field within any nonlinear model order reduction algorithm for systems in the form of (1). This problem is formulated as an L2 gain upper bounding problem of a feedback interconnection of a “nominal” plant and a “disturbance” (i.e. vector field error). The paper proposes a framework for broadening the use of the small gain theorem by introducing the self-cancelling gains  $\sqrt{a}$  and  $\frac{1}{\sqrt{a}}$  in the feedback loop. While this modification fails exactly when the small gain theorem fails to apply, it is nevertheless able to tighten the L2 gain upper bound (by the use of  $\gamma_{G_a}$ ), and the bound is asymptotically tight. Based on the scaled feedback setup, we have shown that the error system L2 gain  $\gamma$  is upper bounded by a linear function of the vector field difference L2 gain  $\gamma_{\Delta}$ , provided  $\gamma_{\Delta}$  is sufficiently small. In an attempt to fight the conservatism of the bound, the paper also proposes a numerical procedure that combines IQC/LMI techniques and small gain theorem. Although the numerical procedure still does not apply for large errors in the vector field, it does produce a tighter bound than the theoretical linear bound. Finally, a numerical example was given to demonstrate the use of our numerical procedure.

## REFERENCES

- [1] M. Celik, A. Atalar, and M. A. Tan, “Transient analysis of nonlinear circuits by combining asymptotic waveform evaluation with Volterra series,” *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 42, no. 8, pp. 470–473, August 1995.
- [2] J. Chen and S. M. Kang, “An algorithm for automatic model-order reduction of nonlinear MEMS devices,” in *Proceedings of ISCAS 2000*, 2000, pp. 445–448.

- [3] J. R. Phillips, "Projection frameworks for model reduction of weakly nonlinear systems," in *37<sup>th</sup> ACM/IEEE Design Automation Conference*, 2000, pp. 184–189.
- [4] —, "Automated extraction of nonlinear circuit macromodels," in *Proceedings of the Custom Integrated Circuit Conference*, Orlando, FL, May 2000, pp. 451–454.
- [5] J. Phillips, "Projection-based approaches for model reduction of weakly nonlinear, time-varying systems," *IEEE Trans. Computer-Aided Design*, vol. 22, no. 2, pp. 171–87, 2003.
- [6] P. Li and L. T. Pileggi, "Norm: Compact model order reduction of weakly nonlinear systems," in *Proc. of the ACM/IEEE Design Automation Conference*, June 2003.
- [7] Y. Chen, "Model order reduction for nonlinear systems," M.S. Thesis, Massachusetts Institute of Technology, September 1999.
- [8] M. Rewienski and J. K. White, "A trajectory piecewise-linear approach to model order reduction and fast simulation of nonlinear circuits and micromachined devices," in *Proc. of IEEE/ACM International Conference on Computer Aided-Design*, San Jose, CA, USA, November 2001, pp. 252–7.
- [9] M. Rewienski and J. White, "A trajectory piecewise-linear approach to model order reduction and fast simulation of nonlinear circuits and micromachined devices," *IEEE Trans. Computer-Aided Design*, vol. 22, no. 2, pp. 155–70, Feb 2003.
- [10] D. Vasilyev, M. Rewienski, and J. White, "A tbr-based trajectory piecewise-linear algorithm for generating accurate low-order models for nonlinear analog circuits and mems," in *Proc. of the ACM/IEEE Design Automation Conference*, June 2003, pp. 490–5.
- [11] N. Dong and J. Roychowdhury, "Piecewise polynomial nonlinear model reduction," in *Proc. of the ACM/IEEE Design Automation Conference*, June 2003.
- [12] —, "Automated extraction of broadly applicable nonlinear analog macromodels from spice-level descriptions," in *Proc. of the IEEE Custom Integrated Circuits Conference*, Oct. 2004.
- [13] B. Bond and L. Daniel, "Parameterized model order reduction for nonlinear dynamical systems," in *Proc. of the IEEE/ACM International Conference on Computer-Aided Design*, San Jose, CA, 2005, pp. 487–494.
- [14] Rewienski M. and White J., "A Trajectory Piecewise-linear Approach to Model Order Reduction and Fast Simulation of Nonlinear circuits and Micromachined Devices," *IEEE Trans. Computer-Aided Design*, vol. 22, pp. 155–170, 2003.
- [15] Philips J. and Afonso J. and Oliveria A. and Silverira L., "Analog macromodeling using kernel methods," in *2003 IEEE/ACM ICCAD*, 2003, pp. 446–453.
- [16] Philips J., "Projection-based approaches for model reduction of weakly nonlinear, time-varying systems," *IEEE Trans. Computer-Aided Design*, vol. 22, pp. 171–187, 2003.
- [17] Li P. and Pileggi L., "NORM: Compact model order reduction of weakly nonlinear systems," in *40th ACM/IEEE Design Automation Conference*, June 2003, pp. 472–477.
- [18] Parrilo P., "Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization," Ph.D. dissertation, California Institute of Technology, 2000.
- [19] Zhou K., *Essentials of Robust Control*. Prentice Hall, 1998.
- [20] Megretski A. and Rantzer A., "System Analysis via Integral Quadratic Constraints," *IEEE Trans. Automatic Control*, vol. 42, no. 6, June 1997.