2007

Reduced Order Models for Nonlinear Control Systems

Krener, A.J.

Reduced Order Models for Nonlinear Control Systems, in Analysis and Design of Nonlinear Control Systems, In Honor of Alberto Isidori, A. Astolfi and L. Marconi, Eds. Springer Verlag

https://hdl.handle.net/10945/52029

Reduced Order Modeling of Nonlinear Control Systems

Arthur J. Krener

Department of Mathematics
University of California
Davis, CA 95616-8633, USA
and
Department of Applied Mathematics
Naval Postgraduate School
Monterey, CA 93943-5216, USA
ajkrener@ucdavis.edu

1 Introduction

The theory of model reduction for linear control systems was initiated by B. C. Moore [9]. His method, called balanced truncation, is applicable to controllable, observable and exponentially stable linear systems in state space form. The reduction is accomplished by making a linear change of state coordinates to simultaneously diagonalize the controllability and observability gramians and make them equal. Such a state space realization is said to be balanced. The diagonal entries of the gramians are the singular values of the Hankel map from past inputs to future outputs. The balanced reduction is accomplished by Galerkin projection onto the states associated to the largest singular values. The method is intrinsic in that the reduced order model depends only on the dimension of the reduced state space.

Scherpen [12] extended Moore's method to locally asymptotically stable nonlinear systems. She defined the controllability and observability functions which are the nonlinear analogs of the controllability and observability gramians. Scherpen made a change of state coordinates that took the system into input normal form where the controllability function is one half of the sum of squares of the state coordinates. She then made additional changes of state coordinates that preserved the input normal form while diagonalizing the observability function where the diagonal entries, which she called the singular value functions, are state dependent. She reduced the system by Galerkin projection onto coordinates with the largest singular value functions.

The reduction technique of Scherpen is not intrinsic. The singular value functions themselves are not unique [6]. Moreover the resulting reduced order

system depends on the changes of coordinates that are used to achieve it and these are not unique.

The goal of this paper is to present a more intrinsic method of nonlinear model reduction. Our approach differs from Scherpen in that we analyze the changes of coordinates degree by degree and give a normal form of the controllability and observability functions for each degree. Generically this normal form of the controllability and observability functions is unique up through terms of degree 7 and it is diagonalized in some sense. There are many changes of coordinates that achieve the normal form and this choice can affect the lower order model.

2 Input Normal Form of Degree d

Suppose we have an n dimensional system of the form

$$\begin{array}{lllll} \dot{x} = f(x,u) = Fx + Gu & +f^{[2]}(x,u) & + \ldots & +f^{[d]}(x,u) & +O(x,u)^{d+1} \\ y = h(x) \quad = Hx & +h^{[2]}(x) & + \ldots & +h^{[d]}(x) & +O(x)^{d+1} \end{array} \tag{1}$$

where $f(x,u),\ h(x)$ are $C^{d+1}, d \geq 1$ in some neighborhood of the equilibrium $x = 0, u = 0$. The superscript $^{[j]}$ denotes a function that is homogeneous and polynomial of degree j in its arguments so the right sides of the above are the Taylor series expansions of $f,\ h$ around $x = 0,\ u = 0$ with remainders of degree $d + 1$.

Following Moore [9] we assume that F is Hurwitz, i.e., all its eigenvalues lie in the open left half plane, F, G is a controllable pair and and H, F is an observable pair. Scherpen [12] defined the controllability and observability functions of the system. The controllability function $\pi_c(x)$ is the solution of the optimal control problem

$$\pi_c(x^0) = \inf_{u(-\infty:0)} \frac{1}{2} \int_{-\infty}^{0} |u|^2 dt \tag{2}$$

subject to the system (1) and the boundary conditions

$$\begin{aligned} x(-\infty) &= 0 \\ x(0) &= x^0. \end{aligned}$$

The notation $u(-\infty : 0)$ denotes a function in $L^2((-\infty, 0), I\!R^m)$. Loosely speaking $\pi_c(x)$ is the minimal "input energy" needed to excite the system from the zero state to x over the time interval $(-\infty, 0]$.

If $\pi_c(x)$ exists and is smooth then it and the optimal control $u = \kappa(x)$ satisfy the Hamilton-Jacobi-Bellman equations

$$0 = \frac{\partial \pi_c}{\partial x}(x) f(x, \kappa(x)) - \frac{1}{2}|\kappa(x)|^2, \qquad 0 = \frac{\partial \pi_c}{\partial x}(x) \frac{\partial f}{\partial u}(x, \kappa(x)) - \kappa'(x). \tag{3}$$

locally around $x = 0$ where $'$ denotes transpose. The negative signs in front of second terms in the above equations occur because we are considering an optimal control problem on $(-\infty, 0]$ rather than the more usual $[0, \infty)$.

Because F is Hurwitz and F, G is a controllable pair then from [1], [8] we know there there exists a unique local solution of these equations around $x = 0$ where $\pi_c(x)$ is C^{d+2} and $\kappa(x)$ is C^{d+1}. Moreover the Taylor series of this solution can be computed term by term,

$$\begin{array}{l} \pi_c(x) = \frac{1}{2}x' P_c^{-1} x + \pi_c^{[3]}(x) + \ldots + \pi_c^{[d+1]}(x) + O(x)^{d+2} \\ \kappa(x) = Kx + \kappa^{[2]}(x) + \ldots + \kappa^{[d]}(x) + O(x)^{d+1} \end{array} \tag{4}$$

where $P_c > 0$ is controllability gramian, i.e., the unique solution to linear Lyapunov equation

$$0 = P_c F + F' P_c + GG' \tag{5}$$

and the linear part of the feedback is

$$K = G' P_c^{-1} \tag{6}$$

The controllability gramian is finite because $F,\ G$ is a controllable pair and positive definite because F is Hurwitz. The higher degree terms of $\pi(x),\ \kappa(x)$ can be computed degree by degree following the method of Al'brecht [1].

The observability function $\pi_o(x)$ is defined by

$$\pi_o(x^0) = \frac{1}{2}\int_0^\infty |y(t)|^2 dt$$

subject to the system (1) and the initial condition

$$x(0) = x^0$$

Since F is Hurwitz we are assured that if x^0 is small enough then $x(t) \to 0$ exponentially fast as $t \to \infty$ so $y(0:\infty) \in L^2((0,\infty), I\!R^p)$. Again speaking loosely $\pi_c(x)$ is the "output energy" that is released by the system over the time interval $[0, \infty)$ when it is started at $x(0) = x$ and the input is zero.

The observability function satisfies the nonlinear Lyapunov equation

$$0 = \frac{\partial \pi_o}{\partial x}(x) f(x, 0) + \frac{1}{2}|h(x)|^2. \tag{7}$$

Because F is Hurwitz and H, F is an observable pair then there exists a unique C^{d+2} solution of this equation defined locally around $x = 0$. The Taylor series of this solution can also be computed term by term,

$$\pi_o(x) = \frac{1}{2}x' P_o x + \pi_o^{[3]}(x) + \ldots + \pi_o^{[d+1]}(x) + O(x)^{d+2}$$

where $P_o > 0$ is the observability gramian, i.e., the unique solution to the linear Lyapunov equation

$$P_oF' + FP_o + H'H = 0.$$

The observability gramian is finite because F is Hurwitz and positive definite because $H,\ F$ is an observable pair.

From [9], [12] we know that we can choose a linear change of coordinates so that in the new coordinates also denoted by x

$$\pi_c(x) = \frac{1}{2}|x|^2 + \pi_c^{[3]}(x) + O(x)^4 \tag{8}$$

$$\pi_o(x) = \frac{1}{2}\sum \tau_i x_i^2 + \pi_o^{[3]}(x) + O(x)^4 \tag{9}$$

where the squared singular values $\tau_1 \geq \tau_2 \geq \ldots \geq \tau_n > 0$ are the ordered eigenvalues of P_oP_c. When (8) holds then we say that the system is in *input normal form of degree one.*

Instead we could have made a linear change of coordinates to make $P_o = I$ and P_c a diagonal matrix. Then we say that the system is in *output normal form of degree one.* Throughout this paper we shall concentrate on systems that are in input normal form but there are analogous results for systems that are in output normal form .

The linear part of the system is said to be balanced [9] if the state coordinates have been chosen so that P_c and P_o are diagonal and equal. The diagonal entries $\sigma_1 \geq \ldots \geq \sigma_n > 0$ are called the Hankel singular values of the linear part of the system and they are related to the squared singular values $\tau_i = \sigma_i^2$.

Definition 1. *A system with distinct squared singular values $\tau_1 > \tau_2 > \ldots > \tau_n$ is in input normal form of degree d if*

$$\pi_c(x) = \frac{1}{2}\sum_{i=1}^n x_i^2 + O(x)^{d+2}, \qquad \pi_o(x) = \frac{1}{2}\sum_{i=1}^n \tau_i^{[0:d-1]}(x_i)x_i^2 + O(x)^{d+2} \tag{10}$$

where $\tau_i^{[0:d-1]}(x_i) = \tau_i + \ldots$ is a polynomial in x_i with terms of degrees 0 through $d-1$. They are called the squared singular value polynomials of degree $d-1$.

Input normal form of degree d is similar to a normal form of Scherpen [12]. She showed that, for nonlinear systems with controllable, observable and exponentially stable linear part, state coordinates x can be found such that

$$\pi_c(x) + \frac{1}{2}\sum_{i=1}^n x_i^2, \qquad \pi_o(x) = \frac{1}{2}\sum_{i=1}^n \tau_i(x)x_i^2 \tag{11}$$

where Scherpen called $\tau_i(x)$ the singular value functions.

Grey and Scherpen [6] have shown that the singular value functions $\tau_i(x)$ are not unique except at $x = 0$ where they equal the squared singular values

of the linear part of the system $\tau_i(0) = \tau_i = \sigma_i^2$. For example, suppose $1 \leq i < j \leq n$ and we define for any $c \in I\!R$

$$\begin{aligned} \bar{\tau}_i(x) &= \tau_i(x) + cx_j^2 \\ \bar{\tau}_j(x) &= \tau_j(x) - cx_i^2 \\ \bar{\tau}_l(x) &= \tau_l(x) \qquad\qquad \text{otherwise} \end{aligned}$$

then

$$\pi_o(x) = \frac{1}{2}\sum_{l=1}^{n} \tau_l(x)x_l^2 = \frac{1}{2}\sum_{l=1}^{n} \bar{\tau}_l(x)x_l^2$$

Moreover there are many local coordinate systems around zero in which the controllability and observability functions of the system are in the normal form of Scherpen (11), see [6].

The differences between Scherpen's normal form and input normal form of degree d are threefold. First the former is exact while the latter is only approximate through terms of degree $d+1$. The second difference is that, in the former, the parameters $\tau_i(x)$ can depend on all the components of x, while, in the latter, when the Hankel singular values are distinct, the i^{th} parameter $\tau_i^{[0:d-1]}(x_i)$ only depends on x_i. Finally and most importantly, the singular value functions $\tau_i(x)$ of the former are not unique except at $x = 0$ while the squared singular value polynomials $\tau_i^{[0:d-1]}(x_i)$ of the latter are unique if $d \leq 6$ and the Hankel singular values are distinct. If the system is odd, i.e., $f(-x,-u) = -f(x,u),\ h(-x) = -h(x)$ then the squared singular value polynomials $\tau_i^{[0:d-1]}(x_i)$ are unique if $d \leq 12$.

Recently Fujimoto and Scherpen [3] have shown the existence of a normal form where π_c is one half the sum of squares of the state coordinates and

$$\frac{\partial \pi_o}{\partial x_i}(x) = 0 \ \text{ iff } \ x_i = 0. \tag{12}$$

But the normal form of Fujimoto and Scherpen [3] is not unique while the input normal form of degree $d \leq 6$ is unique.

While writing this paper we became aware of a earlier paper of Fujimoto and Scherpen [2] that claims the following. Suppose the linear part of the system is controllable, observable and Hurwitz and the Hankel singular values are distinct. Then there exists a local change of coordinates such that the controllability and observability functions are of the form

$$\pi_c(x) = \frac{1}{2}\sum_{i=1}^{n} x_i^2 \qquad \pi_o(x) = \frac{1}{2}\sum_{i=1}^{n} \left(\rho_i(x_i)x_i\right)^2 \tag{13}$$

Unfortunately there appears to be a gap in their proof. Such a result would be an extremely useful generalization of Morse's Lemma.

Notice that if a system with distinct squared singular values $\tau_i = \tau_i(0)$ is in input normal form of degree d then its controllability and observability functions are "diagonalized" through terms of degree $d+1$. They contain no cross terms of degree less than or equal to $d+1$ where one coordinate multiplies a different coordinate. This is reminiscent of the balancing of linear systems by B. C. Moore [9].

For linear systems the squared singular value τ_i is a measure of the importance of the coordinate x_i. The "input energy" in the state x is $\pi_c(x)$ and the "output energy" is $\pi_o(x)$. The states that are most important are those with the most "output energy" for fixed "input energy". Therefore in constructing the reduced order model, Moore kept the subspace of states with largest τ_i for they have the most "output energy" per unit "input energy".

In Scherpen's generalization [12] of Moore, the singular value functions $\tau_i(x)$ measure the importance of the state x_i. To obtain a reduced order model, she assumed $\tau_i(x) > \tau_j(x)$ whenever $1 \leq i \leq k < j \leq n$ and x is in a neighborhood of the origin. Then she kept the states $x_1, \ldots, x_k$ in the reduced order model. But the $\tau_i(x)$ are not unique so this approach is not uniquely defined.

For nonlinear systems in input normal form of degree d, the polynomial $\tau_i^{[0:d-1]}(x_i)$ is a measure of the importance of the coordinate x_i for moderate sized x. We shall show that if the τ_i are distinct and $d \leq 6$ then $\tau_i^{[0:d-1]}(x_i)$ is unique. The leading coefficient of this polynomial is the squared singular value τ_i so in constructing a reduced order model we will want to keep the states with the largest τ_i. But τ_i can be small yet $\tau_i^{[0:d-1]}(x_i)$ can be large for moderate sized x_i. If we are interested in capturing the behavior of the system for moderate sized inputs, we may also want to keep such states in the reduced order model. We shall return to this point when we discuss reduced order models in a Section 4.

Theorem 1. *Suppose the system (1) is C^r, $r \geq 2$ with controllable, observable and exponentially stable linear part . If the squared singular values $\tau_1, \ldots, \tau_n$ are distinct and if $2 \leq d < r-1$ then there is at least one change of state coordinates that takes the system into input normal form of degree d (10). The controllability and observability functions of a system in input normal form of degree $d \leq 6$ are unique. But the system and a change of coordinates that achieves input normal form are not necessarily unique even to degree d. If the system is odd then the controllability and observability functions of a system in input normal form of degree $d \leq 12$ are unique but again the system and a change of coordinates that achieves it are not necessarily unique.*

Proof. We shall prove the first part by induction. Moore has shown the existence of input normal form of degree $d = 1$ so assume that we have shown the existence of input normal form of degree $d-1$. Then there are state coordinates x and polynomials $\tau_i^{[0:d-2]}(x_i)$ of degree 0 through $d-2$ such that

$$\pi_c(x) = \frac{1}{2}\sum_{i=1}^{n} x_i^2 + \pi_c^{[d+1]}(x) + O(x)^{[d+2]}$$
$$\pi_o(x) = \frac{1}{2}\sum_{i=1}^{n} \tau_i^{[0:d-2]}(x_i)x_i^2 + \pi_o^{[d+1]}(x) + O(x)^{[d+2]}.$$

A *near identity change of coordinates of degree* $d > 1$ is one of the form $x = z + \phi^{[d]}(z)$.. For brevity we refer to it as a change of coordinates of degree d. Notice that a change of coordinates of degree d does not change the expansions of π_c and π_o through terms of degree d but it can change terms of degrees greater than d. We shall show that there is a degree d change of coordinates that will bring a system from input normal form of degree $d-1$ to input normal form of degree d. In fact there may be several such degree d changes of coordinates.

Suppose we have a degree $d+1$ monomial

$$x_i x_j x_{k_1} \cdots x_{k_{d-1}} \tag{14}$$

with at least two distinct indices, say $i \neq j$. Let γ_c and γ_o be the coefficients of this monomial in $\pi_c^{[d+1]}(x)$ and $\pi_o^{[d+1]}(x)$

After the degree d change of coordinates

$$\begin{array}{ll} \phi_i^{[d]}(z) = a_i z_j z_{k_1} \cdots z_{k_{d-1}} & \\ \phi_j^{[d]}(z) = a_j z_i z_{k_1} \cdots z_{k_{d-1}} & \\ \phi_l^{[d]}(z) = 0 & \text{otherwise} \end{array} \tag{15}$$

we have

$$\begin{aligned} \pi_c(z) = \frac{1}{2}\sum_{i=1}^{n} z_i^2 + \pi_c^{[d+1]}(z) + (a_i + a_j) z_i z_j z_{k_1} \cdots z_{k_{d-1}} \\ + O(z)^{[d+2]} \end{aligned}$$

$$\begin{aligned} \pi_o(z) = \frac{1}{2}\sum_{i=1}^{n} \tau_i^{[0:d-2]}(z_i) z_i^2 \\ + \pi_o^{[d+1]}(z) + (\tau_i a_i + \tau_j a_j) z_i z_j z_{k_1} \cdots z_{k_{d-1}} \\ + O(z)^{[d+2]}. \end{aligned}$$

We would like to choose a_i and a_j so as to cancel the monomial $z_i z_j z_{k_1} \cdots z_{k_{d-1}}$ from both $\pi_c^{[d+1]}(z)$ and $\pi_o^{[d+1]}(z)$ so they must satisfy

$$\begin{bmatrix} 1 & 1 \\ \tau_i & \tau_j \end{bmatrix} \begin{bmatrix} a_i \\ a_j \end{bmatrix} = - \begin{bmatrix} \gamma_c \\ \gamma_o \end{bmatrix} \tag{16}$$

Since $i \neq j$ then $\tau_i \neq \tau_j$ and this is always possible .

We proceed in this way to cancel all monomials in $\pi_c^{[d+1]}(z)$ and $\pi_o^{[d+1]}(z)$ with at least two distinct indices and so all that are left are monomials with all indices the same $i = j = k_1 = \cdots = k_{d-1}$. For such a monomial the degree d change of coordinates

$$\begin{aligned} \phi_i^{[d]}(z) &= -\gamma_c z_i^{d+1} \\ \phi_l^{[d]}(z) &= 0 \qquad\quad \text{otherwise} \end{aligned} \tag{17}$$

can be used to cancel the monomial z_i^{d+1} from $\pi_c^{[d+1]}(z)$ but nothing can be done about the same monomial in $\pi_o^{[d+1]}(z)$. Hence it is added to $\tau_i^{[0:d-2]}(z_i)$ to form $\tau_i^{[0:d-1]}(z_i)$

Next we show that if $d \leq 6$ the normal form is unique. Let γ_c and γ_o be the coefficients the monomial $x_i x_j x_k$ in $\pi_c^{[3]}$ and $\pi_o^{[3]}$. If $i = j = k$ then there is only one change of coordinates (17) that cancels the monomial from $\pi_c^{[3]}$. If there are only two distinct indices among $i,\ j,\ k$ then there is only one change of coordinates (15, 16) that cancels the monomial from $\pi_c^{[3]}$ and $\pi_o^{[3]}$.

Assume that the indices are distinct, $i < j < k$. Then there is a one parameter family of degree two transformations that cancel this monomial from $\pi_c^{[3]},\ \pi_o^{[3]}$,

$$\begin{aligned} x_i &= z_i + a_i z_j z_k \\ x_j &= z_j + a_j z_i z_k \\ x_k &= z_k + a_k z_i z_j \\ x_l &= z_l, \qquad \text{otherwise.} \end{aligned} \tag{18}$$

The coefficients $a_i,\ a_j,\ a_k$ must satisfy

$$\begin{bmatrix} 1 & 1 & 1 \\ \tau_i & \tau_j & \tau_k \end{bmatrix} \begin{bmatrix} a_i \\ a_j \\ a_k \end{bmatrix} = - \begin{bmatrix} \gamma_c \\ \gamma_o \end{bmatrix}$$

Since $\tau_i > \tau_j > \tau_k$ we can choose any a_i and adjust $a_j,\ a_k$ to satisfy this constraint. This freedom propagates to the higher order remainders of π_c and π_c in three ways.

The first way is that it introduces terms like $z_j^2 z_k^2,\ z_i^2 z_k^2,\ z_i^2 z_j^2$ with coefficients that are not unique because they depend on a_i. But all these contain two distinct indices and so all can be cancelled. For example we would cancel the $z_j^2 z_k^2$ terms with a degree three change of coordinates of the form

$$\begin{aligned} z_j &= \xi_j + b_j \xi_j \xi_k^2 \\ z_k &= \xi_k + b_k \xi_j^2 \xi_k. \end{aligned}$$

This introduces nonunique terms like $\xi_j^2 \xi_k^4$ and $\xi_j^4 \xi_k^2$ but these are easily cancelled because they contain two distinct indices. The coordinate transformations that cancel them introduce nonunique terms of degree 12 that we don't care about.

Here is another way that (18) can nonuniquely change the higher remainders. Suppose the monomial x_i^3 appears in π_o in the input normal form of degree $d > 2$. Then after (18) it is replaced by

$$z_i^3 + 3a_i z_i^2 z_j z_k + 3a_i^2 z_i z_j^2 z_k^2 + a_i^3 z_j^3 z_k^3.$$

The first nonunique term $3a_i z_i^2 z_j z_k$ contains three distinct indices so it is easily cancelled by a change of coordinates of degree 3 that introduce nonunique terms of degree 6 with at least two distinct indices which in turn are easily cancelled by a changes of coordinates of degree 5 which introduce nonunique terms of degree 10 that we don't care about. The second nonunique term $3a_i^2 z_i z_j^2 z_k^2$ also contains three distinct indices so it is easily cancelled by a change of coordinates of degree 4 that introduce extra terms of degree 8 that we don't care about. The last nonunique term $a_i^3 z_j^3 z_k^3$ has two distinct indices so it can be cancelled by a change of coordinates that introduce nonunique terms of degree 12 that we don't care about.

The last way that (18) can nonuniquely change the higher remainders is as follows. Suppose the monomial $x_i x_{l_1} x_{l_2} x_{l_3}$ appears in the quartic remainders of $\pi_c,\ \pi_o$ Then after (18) it is replaced by

$$z_i z_{l_1} z_{l_2} z_{l_3} + a_i z_j z_k z_{l_1} z_{l_2} z_{l_3} + \dots$$

The nonunique term $a_i z_j z_k z_{l_1} z_{l_2} z_{l_3}$ contains at least two distinct indices $j \neq k$ so it can be cancelled by a change of coordinates of degree 4 which introduces nonunique terms of degree 8 that we don't care about.

But if if $l_1 = l_2 = l_3 = k$ then the change of coordinates that cancels the nonunique term $a_i z_j z_k^4$ is of the form

$$\begin{aligned} z_j &= \xi_j + b_j \xi_k^4 \\ z_k &= \xi_k + b_k \xi_j \xi_k^3 \end{aligned}$$

and the first of these introduces a nonunique term $b_j^2 \xi_k^8$ that contains only one distinct index and so it cannot be cancelled from $\pi_o^{[8]}$. This is why input normal form is not unique for $d \geq 7$.

If the system is odd then it is easy to see that $\pi_c,\ \pi_o$ are even functions

$$\pi_c(x) = \pi_c(-x), \qquad \pi_o(x) = \pi_o(-x)$$

so there Taylor expansions contain only even terms. A slight modification of the above argument shows that input normal of degree $d \leq 12$ is unique.

The *normal change of coordinates of degree d* that achieves input normal form of degree d is constructed as follows. For each monomial (14) let $i,\ j$ be the pair of distinct indices that are furthest apart. Then we choose $a_i,\ a_j$ in the change of coordinates (15) to cancel this monomial in $\pi_c^{[d+1]}(z)$ and $\pi_o^{[d+1]}(z)$. If there are not two distinct indices then we choose the change of coordinates

(17) to cancel the monomial in $\pi_c^{[d+1]}(z)$. Then form the composition of all such changes of coordinates as one ranges over all monomials of degree $d+1$ and throw away the part of composition of degree greater than d. The result does not depend on the order of the composition and it is *the unique normal change of coordinates of degree d*. The rational behind using the normal change of coordinates of degree d is that if $i,\ j$ are as far apart as possible then so are $\tau_i,\ \tau_j$. The coefficients $a_i,\ a_j$ that are used to cancel the monomial (14) in both $\pi_c^{[d+1]}$ and $\pi_o^{[d+1]}$ satisfy the pair of linear equations (16). The determinant of the matrix on the left is $\tau_j - \tau_i$ and we would like to make its magnitude as large as possible to minimize the effect of numerical errors in solving these linear equations. Hence we choose $i,\ j$ as far apart as possible.

While writing this paper we became aware of a paper of Fujimoto and Tsubakino [4] that discusses the term by term computation of a change of coordinates that takes a system into input normal form of degree d. They show that at each degree the coefficients of the change of coordinates must satisfy a set of linear equations that is underdetermined, there are more coordinates than there are constraints in the normal form. But they don't show that the set of linear equations is always solvable as we have above.

Next we drop the assumption that the squared singular values are distinct.

Definition 2. *The system is in input normal form of degree d if*

$$\begin{aligned} \pi_c(x) &= \tfrac{1}{2}\textstyle\sum_{i=1}^n x_i^2 + O(x)^{d+2} \\ \pi_o(x) &= \tfrac{1}{2}\textstyle\sum_{i=1}^n \tau_i^{[0:d-1]}(x) x_i^2 + O(x)^{d+2} \end{aligned} \tag{19}$$

where $\tau_i^{[0:d-1]}(x)$ *is a polynomial of degrees* 0 *through* $d-1$ *in the variables* $\{x_j : \tau_j = \tau_i\}$ *and with constant term*

$$\tau_i^{[0:d-1]}(0) = \tau_i$$

Theorem 2. *Suppose the system (1) is* C^r $r \geq 2$ *with controllable, observable and exponentially stable linear part . If* $2 \leq d < r-1$ *then there is a change of state coordinates that takes the system into input normal form of degree d (19).*

Proof. A slight extension of the proof of the previous theorem yields the existence of the input normal form of degree d. One uses changes of coordinates of the form (15) where $\tau_i \neq \tau_j$ to cancel the monomial $z_i z_j z_{k_1} \cdots z_{k_{d-1}}$ from $\pi_c^{[d+1]}(z)$ and $\pi_c^{[d+1]}(z)$.

If $\tau_i = \tau_j = \tau_{k_1} = \ldots = \tau_{k_{d-1}}$ then the monomial $z_i z_j z_{k_1} \cdots z_{k_{d-1}}$ can be canceled from $\pi_c^{[d+1]}(z)$ using the change of coordinates

$$\begin{aligned} \phi_i^{[d]}(z) &= c_i z_i z_j z_{k_1} \cdots z_{k_{d-1}} \\ \phi_l^{[d]}(z) &= 0 \qquad\qquad \text{otherwise} \end{aligned} \tag{20}$$

But this change of coordinates is not uniquely determined unless $i = j = k_1 = \ldots, k_{d-1}$. For example if $i \neq j$ we could as well use the change of coordinates

$$\begin{aligned} \bar{\phi}_j^{[d]}(z) &= c_j z_i z_j z_{k_1} \cdots z_{k_{d-1}} \\ \bar{\phi}_l^{[d]}(z) &= 0 \qquad\qquad \text{otherwise} \end{aligned} \tag{21}$$

to cancel the monomial $z_i z_j z_{k_1} \cdots z_{k_{d-1}}$ from $\pi_c^{[d+1]}(z)$. Or we could use a combination of them both

$$\begin{aligned} \tilde{\phi}_i^{[d]}(z) &= c_i z_i z_j z_{k_1} \cdots z_{k_{d-1}} \\ \tilde{\phi}_j^{[d]}(z) &= c_j z_i z_j z_{k_1} \cdots z_{k_{d-1}} \\ \tilde{\phi}_l^{[d]}(z) &= 0 \qquad\qquad \text{otherwise.} \end{aligned} \tag{22}$$

If the squared singular values $\tau_1, \ldots, \tau_n$ are not distinct then input normal form of any degree $d > 1$ is not unique. For example we could make a block diagonal change of coordinates

$$x_i = z_i + \phi_i(z)$$

where $\phi_i(z)$ only depends on those z_j such that $\tau_j = \tau_i$. By the an argument similar to the above we see that input normal form of degree $d \leq 6$ is unique up to such block diagonal changes of coordinates.

3 Linear Model Reduction

Moore's method [9] of obtaining a reduced order model of a linear system is called balanced truncation. One chooses so-called balanced linear state coordinates z where the controllability and observability gramians are diagonal and equal,

$$P_c = P_o = \text{diagonal}(\sigma_1, \ldots, \sigma_n)$$

If $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_k >> \sigma_{k+1} \geq \ldots \geq \sigma_n > 0$ then a k dimensional reduced order model is obtained by Galerkin projection onto the subspace of the first k balanced coordinates.

An equivalent method is to chose linear coordinates x so that the system is in input normal form of degree 1,

$$P_c = I \qquad P_o = \text{diagonal}(\tau_1, \ldots, \tau_n)$$

where $\tau_i = \sigma_i^2$. A k dimensional reduced order model is obtained by Galerkin projection onto the subspace of the first k input normal coordinates. The two sets of coordinates and the reduced order models are related by $z_i = \pm \sigma_i^{\frac{1}{2}} x_i$.

It is convenient to let $\mathbf{x}_1 = (x_1, x_2, \ldots, x_k)$ and $\mathbf{x}_2 = (x_{k+1}, \ldots, x_n)$ then in these coordinates the full order linear system is

$$\begin{aligned} \dot{\mathbf{x}_1} &= F_{11}\mathbf{x}_1 + F_{12}\mathbf{x}_2 + G_1 u \\ \dot{\mathbf{x}_2} &= F_{21}\mathbf{x}_1 + F_{22}\mathbf{x}_2 + G_2 u \\ y &= H_1\mathbf{x}_1 + H_2\mathbf{x}_2 \end{aligned} \tag{23}$$

and the reduced order model is

$$\begin{aligned} \dot{\mathbf{x}_1} &= F_{11}\mathbf{x}_1 + G_1 u \\ y &= H_1\mathbf{x}_1 \end{aligned} \tag{24}$$

How does one justify balanced truncation? Consider a linear system

$$\begin{aligned} \dot{x} &= Fx + Gu \\ y &= Hx \end{aligned} \tag{25}$$

If F is Hurwitz then it defines an input-output map

$$\begin{gathered} \mathcal{IO} : L^2((-\infty, \infty), I\!R^m) \to L^2(-\infty, \infty, I\!R^p) \\ \mathcal{IO} : u(-\infty : \infty) \mapsto y(-\infty : \infty) \end{gathered}$$

given by

$$y(t) = \int_{-\infty}^{t} He^{F(t-s)}Gu(s)\, ds$$

Ideally want would like to choose the reduced order model to minimize over all models of state dimension k the norm of the difference between the input-output maps of the full and reduced models. Balanced truncation does not achieve this goal.

The input-output map of a linear system is not a compact operator which causes mathematical difficulties. There is a closely related map which is of finite rank hence compact. It is the Hankel map from past inputs to future outputs which factors through the current state

$$\begin{gathered} \mathcal{H} : L^2(-\infty, 0], I\!R^m) \to L^2([0, \infty), I\!R^p) \\ \mathcal{H} : u(-\infty : 0) \mapsto y(0 : \infty) \end{gathered}$$

given by

$$\begin{aligned} x(0) &= \int_{-\infty}^{0} e^{-Fs}Gu(s)\, ds \\ y(t) &= He^{Ft}x(0) \end{aligned}$$

Unfortunately balanced truncation does not minimize the difference of norm between the Hankel maps of the full and reduced models over all reduced models of state dimension k.

So how does one justify balanced truncation and how can it be generalized to nonlinear systems? Newman and Krishnaprasad [11] have given a stochastic way. Here is another way. We start by restricting our attention to reduced order models that can be obtained by Petrov Galerkin projection of (25). A Petrov Galerkin requires two linear maps

$$\begin{aligned} \Psi &: I\!R^k \to I\!R^n, \qquad \Psi : z \mapsto x = \Psi z \\ \Phi &: I\!R^n \to I\!R^k, \qquad \Phi : x \mapsto z = \Phi x \end{aligned}$$

such that $\Phi\Psi z = z$ and$(\Psi\Phi)^2 = \Psi\Phi$. The reduced order model of (25) is then

$$\begin{aligned} \dot{z} &= \Phi\left(F\Psi z + Gu\right) \\ y &= H\Psi z \end{aligned}$$

Balanced truncation is a Galerkin projection in balanced coordinates where

$$\Phi = \Psi' = \begin{bmatrix} I & 0 \end{bmatrix} \tag{26}$$

But in the original coordinates it is a Petrov Galerkin projection. How were Ψ and Φ chosen?

Intuitively to obtain a reduced order model of dimension k of the linear system (25) we should Ψ so that the states in its range have the largest output energy $\pi_o(x)$ for given input energy $\pi_c(x)$. More precisely, the range of Ψ should be the k dimensional subspace through the origin where $\pi_o(x)$ is maximized for given $\pi_c(x)$. If the linear system is in input normal coordinates the clearly this subspace is given by $x_{k+1} = \cdots = x_n = 0$ and a convenient choice of Ψ is (26).

We choose Φx so that the norm of the difference in the outputs starting from x and $\Psi\Phi x$ is as small as possible. To do this we define the coobservability function

$$\pi_{oo}(x, \bar{x}) = \frac{1}{2}\int_0^\infty |y(t) - \bar{y}(t)|^2 \, dt \tag{27}$$

where $y(0:\infty),\ \bar{y}(0:\infty)$ are the outputs of the linear system (25) starting from $x,\ \bar{x}$ at $t = 0$ with $u(0:\infty) = 0$. Then we choose Φx to minimize

$$\pi_{oo}(x, \Psi\Phi x)$$

Because of the system is linear, $\pi_{oo}(x, \bar{x})$ is a quadratic form in $(x, \bar{x})$ and

$$\pi_{oo}(x, \bar{x}) = \pi_o(x - \bar{x}) = \frac{1}{2}\sum \tau_i (x_i - \bar{x}_i)^2$$

If the system is in input normal coordinates then the minimizing Φ is given by (26).This explains choices of $\Psi,\ \Phi$ that are made in balanced truncation.

4 Nonlinear Model Reduction

We would like to generalize linear balanced truncation to nonlinear systems of the form

$$\begin{aligned} \dot{x} &= f(x,u) \\ y &= h(x) \end{aligned} \tag{28}$$

To do so we restrict our attention to reduced order models that are constructed by a nonlinear Galerkin projection. A nonlinear Galerkin projection of (28) is defined by two maps, an embedding ψ from the lower dimensional state space into the higher one and a submersion ϕ of the higher dimensional state space onto the lower dimensional one. These state spaces could be manifolds but, since we will focus on local methods, we shall assume that they are neighborhoods of the origin in $\mathbb{R}^k,\ \mathbb{R}^n$. To simplify notation we just let $\mathbb{R}^k,\ \mathbb{R}^n$ stand for these neighborhoods.

So we seek

$$\begin{aligned} \psi &: \mathbb{R}^k \to \mathbb{R}^n, \qquad \psi : z \mapsto x \\ \phi &: \mathbb{R}^n \to \mathbb{R}^k, \qquad \phi : x \mapsto z \end{aligned}$$

such that $\phi(\psi(z)) = z$ and $(\psi \circ \phi)^2(x) = \psi \circ \phi(x)$

Motivated by the interpertation of linear balanced truncation given above we would like to choose ψ so that the submanifold that is its range maximizes output energy $\pi_o(x)$ for fixed input energy $\pi_c(x)$. But if $k > 1$ and the system is nonlinear then this submanifold is not well-defined.

Suppose $k = 1$ then for each small positive constant c we can maximize $\pi_o(x)$ subject to $\pi_c(x) = c$. Since the quadratic parts of these functions are positive definite, for small $c > 0$, $\pi_o(x)$ will have two local maxima on each level set $\pi_c(x) = c$. The locus of these local maxima form a one dimensional submanifold through the origin which we can take as the state space of our one dimensional reduced order model.

But if $k > 1$ then the k dimensional submanifold that maximizes $\pi_o(x)$ for $\pi_c(x) = c$ is not well-defined. When the system is linear and $\pi_o(x),\ \pi_c(x)$ are quadratic forms then this submanifold is assumed to be a subspace and hence is well-defined. It is the subspace spanned by the k leading eigenvectors of P_o when the system is in input normal form $P_c = I$.

Suppose that the squared singular values are distinct and that the nonlinear system has been brought to input normal form (19) of degree d by changes of coordinates up to degree d. The minimum input energy necessary to excite the system to state x is

$$\pi_c(x) = \frac{1}{2}|x|^2 + O(x)^{d+2}.$$

The output energy generated by the system relaxing from the state x is

$$\pi_o(x) = \frac{1}{2}\sum_{j=1}^{n} \tau_j^{[0:d-1]}(x_j)x_j^2 + O(x)^{d+2}.$$

Suppose further that there is a gap in the squared singular value polynomials over the range of states of interest $|x| \leq c$,

$$\tau_i^{[0:d-1]}(x_i) >> \tau_j^{[0:d-1]}(x_j) \tag{29}$$

for $1 \leq i \leq k < j \leq n$ and $|x_i| \leq c,\ |x_j| \leq c$.

Then a k dimensional submanifold that "approximately maximizes" $\pi_o(x)$ for given $\pi_c(x)$ is given by $x_{k+1} = \cdots = x_n = 0$ and we define

$$\psi(z_1, \ldots, z_k) = x = (z_1, \ldots, z_k, 0, \ldots, 0) \tag{30}$$

We find the submersion ϕ as before. Define the co-observability function $\pi_{oo}(x, \bar{x})$ as before (27) except that $y(0:\infty),\ \bar{y}(0:\infty)$ are the outputs of the nonlinear system (28) starting from $x,\ \bar{x}$ with $u(0:\infty) = 0$. It is not hard to see that $\pi_{oo}(x, \bar{x})$ satisfies the Lyapunov PDE

$$0 = \left[\frac{\partial \pi_{oo}}{\partial x}(x, \bar{x})\ \frac{\partial \pi_{oo}}{\partial \bar{x}}(x, \bar{x}) \right] \left[\begin{matrix} f(x, 0) \\ f(\bar{x}, 0) \end{matrix} \right] + \frac{1}{2}|h(x) - h(\bar{x})|^2$$

and this can be easily solved term by term,

$$\pi_{oo}(x, \bar{x}) = \frac{1}{2} \sum \tau_i (x_i - \bar{x}_i)^2 + \pi_{oo}^{[3]}(x, \bar{x}) + \pi_{oo}^{[4]}(x, \bar{x}) + \ldots$$

We choose $\phi(x)$ to minimize $\pi_{oo}(x, \psi(\phi(x)))$. Assume that the system is in input normal form of degree d and ψ has been chosen as above . Then a straightforward calculation leads to

$$\begin{aligned} \phi_i(x) = x_i \\ + \tfrac{1}{\tau_i} \left(\tfrac{\partial \pi_{oo}^{[3]}}{\partial \bar{x}_i}(x, (\phi(x), 0)) + \tfrac{\partial \pi_{oo}^{[4]}}{\partial \bar{x}_i}(x, (\phi(x), 0)) + \ldots \right) \end{aligned} \tag{31}$$

for $i = 1, \ldots, k$ which can be solved by repeated substitution.

The reduced order nonlinear model is

$$\begin{aligned} \dot{z} = a(z, u) &= \frac{\partial \phi}{\partial x}(\psi(z)) f(\psi(z), u) \\ y = c(z) \quad &= h(\psi(z)) \end{aligned} \tag{32}$$

Here is our algorithm for nonlinear model reduction to degree d.

1. Compute the controllability and observability functions $\pi_c(x),\ \pi_o(x)$ to degree $d+1$ by solving the HJB and Lyapuov equations (3, 7) term by term.
2. Make normal changes of coordinates of degrees 1 through d to bring the system into input normal form of degree d, (10)
3. Examine the squared singular value polynomial $\tau_i^{[0:d-1]}(x_i)$ to see if there is a gap (29) for some k over the range of states of interest.
4. Define the embedding ψ by (30).
5. Find the submersion ϕ by solving (31) to degree d.
6. The degree d reduced order model is given by the truncation of (32) to terms of degree less than or equal to d.

5 Linear Error Estimates

K. Glover [5] has given an important error bound for the norm of the difference between the input-output map of the full linear system $\mathcal{IO}_n$ and the input-output map of its balanced truncation $\mathcal{IO}_k$,

$$\|\mathcal{IO}_n - \mathcal{IO}_k\| \leq 2 \sum_{j=k+1}^{n} \sigma_j$$

where the norm is the induced L^2 norm.

We know that the corresponding Hankel maps $\mathcal{H}_n,\ \mathcal{H}_k$ satisfy

$$\sigma_{k+1} \leq \|\mathcal{H}_n - \mathcal{H}_k\|$$

so we have for linear balanced truncation

$$\sigma_{k+1} \leq \|\mathcal{H}_n - \mathcal{H}_k\| \leq \|\mathcal{IO}_n - \mathcal{IO}_k\| \leq 2 \sum_{j=k+1}^{n} \sigma_j$$

Unfortunately we do not have similar estimates for nonlinear model reduction. But there is a new error estimate for linear systems that can be extended to nonlinear systems. To each state x of the full order model there is an optimal open loop control $u_x(-\infty : 0)$ that excites the system from state 0 at $t = -\infty$ to state x at $t = 0$. It is the solution of the optimal control problem (2). These optimal controls form a n dimensional subspace of the space of inputs $L^2([-\infty : 0], \mathbb{R}^m)$ to the Hankel map. We expect that these optimal controls are typical of those that are used in the full order model and hence we are interested in the size of errors when they are used in the reduced order model. The error of the Hankel maps can be readily bounded as follows.

The optimal open loop controls are generated by the optimal feedback (6). If the linear system is in input normal coordinates then the optimal gain is $K = G'$. We drive both the full model (23) and the reduced model (24) obtained by balanced truncation by this feedback to get the combined system

$$\begin{bmatrix} \dot{x} \\ \dot{z} \end{bmatrix} = \begin{bmatrix} F + GK & 0 \\ G_1 K & F_{11} \end{bmatrix}$$

For balanced truncation $-(F + GK)$ and F_{11} are Hurwitz so there is an unstable subspace $z = Tx$ where T satisfies the Sylvester equation

$$T(F + GK) - F_{11}T = G_1 K \tag{33}$$

The meaning of T is that if we excite the reduced order system with the optimal control $u_x(-\infty : 0)$ that excites the full order system to x then $z(0) = Tx$.

Next we define the cross-observability function

$$\rho(x,z) = \frac{1}{2}\int_0^\infty |y_f(t) - y_r(t)|^2 \, dt \tag{34}$$

where $y_f(0:\infty)$ is the output of the full order model starting at $x(0) = x$ and $y_r(0:\infty)$ is the output of the reduced order model starting at $z(0) = z$ with $u(0:\infty) = 0$. Because the systems are linear, ρ is a quadratic form

$$\rho(x,z) = \frac{1}{2}\begin{bmatrix} x \\ z \end{bmatrix}' \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix}$$

where Q satisfies the Sylvester equation

$$0 = \begin{bmatrix} F & 0 \\ 0 & F_{11} \end{bmatrix}' \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} + \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{bmatrix} F & 0 \\ 0 & F_{11} \end{bmatrix} + \begin{bmatrix} H'H & H'H_1 \\ H_1'H & H_1'H_1 \end{bmatrix} \tag{35}$$

Clearly Q_{11} is the observability gramian. If the system is in balanced or input normal coordinates then Q_{22} is the upper left $k \times k$ block of Q_{11}.

If we use the optimal control $u_x(-\infty:0)$ then the norm of error between the full and reduced Hankel maps is

$$2\rho(x,Tx) = x' \begin{bmatrix} I \\ T \end{bmatrix}' \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{bmatrix} I \\ T \end{bmatrix} x$$

If the system is in input normal form then the maximum squared norm of error between the Hankel maps restricted to optimal inputs of the full system is the largest eigenvalue of

$$\begin{bmatrix} I \\ T \end{bmatrix}' \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{bmatrix} I \\ T \end{bmatrix} \tag{36}$$

This error estimate is always greater than or equal to the largest neglected squared singular value τ_{k+1} because the right singular vectors of the Hankel map of the full system are a basis for the space of optimal controls. In the few examples that we have computed we found that this error estimate is much closer to τ_{k+1} than to the square of Glover's bound.

One can compute the maximum norm of error between the Hankel maps restricted to optimal inputs of the reduced system in a similar fashion.

6 Nonlinear Error Estimates

Unfortunately for nonlinear model reduction we don't have an error bound on the input-output maps similar to Glover. Furthermore we don't have a lower bound on the norm of the Hankel error like the first neglected singular value. But we can generalize the error bounds of the Hankel maps restricted to optimal inputs of the full or reduced system. We present the bound for the optimal inputs of the full system. The other is very similar.

Suppose we have a full order system (1) which is in input normal form of degree d and its reduced order model

$$\begin{array}{lll} \dot{z} = a(z,u) = F_{11}z + G_1 u & +a^{[2]}(z,u) + \ldots & +a^{[d]}(z,u) \\ y = c(z) \quad = H_1 z & +c^{[2]}(z) \quad + \ldots & +c^{[d]}(x) \end{array} \tag{37}$$

found by the method above or a similar method.

For each $x \in I\!R^n$ there is an optimal open loop control $u_x(-\infty : 0)$ that excites the system from state 0 at $t = -\infty$ to state x at $t = 0$. It is the solution of the optimal control problem (2). These optimal controls form a n dimensional submanifold of the space of inputs $L^2([-\infty : 0], I\!R^m)$ of the Hankel map. Again we expect that these optimal controls are typical of those that are used in the full order model and hence we are interested in the errors when they are used in the reduced order model.

The optimal controls are generated by the feedback (4) which can be computed term by term. We plug this feedback into the combined system

$$\begin{aligned} \dot{x} &= f(x, \kappa(x)) = (F + GK)x + \ldots \\ \dot{z} &= a(z, \kappa(x)) = F_{11}z + G_1 K x + \ldots \end{aligned}$$

Again $-(F + GK)$ and F_{11} are Hurwitz so there exists an unstable manifold $z = \theta(x)$ which satisfies the PDE

$$a(\theta(x), \kappa(x)) = \frac{\partial \theta}{\partial x}(x) f(x, \kappa(x))$$

This PDE can be solved term by term and the linear coefficient is the T satisfying (33).

The cross-obervability function $\rho(x, z)$ is defined as before (34) except now the full (1)and reduced (37) systems are nonlinear and the input is zero. The cross-obervability function satisfies the PDE

$$0 = \left[\begin{array}{cc} \frac{\partial \rho}{\partial x}(x,z) & \frac{\partial \rho}{\partial z}(x,z) \end{array} \right] \begin{bmatrix} f(x,0) \\ a(z,0) \end{bmatrix} + \frac{1}{2}|h(x) - c(z)|^2$$

This also has a series solution

$$\rho(x,z) = \frac{1}{2} \begin{bmatrix} x \\ z \end{bmatrix}' \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} + \rho^{[3]}(x,z) + \ldots$$

where Q satisfies the Sylvester equation (35).

Then the squared norm of the error between the full and reduced Hankel maps using the optimal control $u_x(-\infty : 0)$ is

$$2\rho(x, \theta(x))$$

and good estimate of the maximum relative squared norm of the error is

$$\sup \frac{\rho(x, \theta(x))}{\pi_c(x)}$$

Suppose the system is in input normal form of degree d, so that

$$\pi_c(x) = \frac{1}{2}|x|^2 + O(x)^{d+2} \tag{38}$$

Then we can make a linear orthogonal change of coordinates to diagonalize the quadratic part (36) of $\rho(x, \theta(x))$. If the diagonal entries are distinct, then as with the transformation to input normal form of degree d we can make further changes of state coordinates of degrees 2 through d that leave $\pi_c(x)$ as above (38) and bring $\rho(x, \theta(x))$ into the normal form

$$\rho(x, \theta(x)) = \frac{1}{2} \sum_i \epsilon_i^{[0:d-1]}(x_i) x_i^2$$

The $\epsilon_i^{[0:d-1]}(x_i)$ are polynomials of degrees 0 through $d-1$ and are called the squared error polynomials. They are unique if $d \leq 6$ ($d \leq 12$ for odd systems). They measure how fast the squared error between the Hankel maps grows restricted to optimal controls $u_x(-\infty : 0)$ as x grows.

7 Example

We consider three linked rods connected by planar rotary joints with springs and dampening hanging from the ceiling. The input is a torque applied to the top joint and the output is the horizontal displacement of the bottom. Each rod is uniform of length 2, mass 1, with spring constant 3, dampening constant 0.5 and gravity constant 0.5.

We approximated the nonlinear system by its Taylor series through terms of degree 5. The Taylor series of controllability and observability functions $\pi_c(x)$, $\pi_o(x)$ were computed through terms of degree 6. The system was brought into input normal form of degree 5 by a changes of state coordinates of degrees 1 through 5. The Hankel singular values of the linear part of the system are 15.3437, 14.9678, 0.3102, 0.2470, 0.0156, 0.0091. Apparently only two dimensions are linearly significant.

Figure 1 is a semilog plot of the squared singular value polynomials $\tau_i^{[0:4]}$ Notice the difference in scale and how flat they are. Apparently only two dimensions are nonlinearly significant.

Let $u_x(-\infty : 0)$ be the optimal input that excites the full system to x. Let $\mathcal{H}_n, \mathcal{H}_k$ be the Hankel of the full order model ($n = 6$) and the reduced order model ($k = 2$). The error between them satisfies

$$|\mathcal{H}_n(u_x(-\infty : 0)) - \mathcal{H}_k(u_x(-\infty : 0))|^2 \leq 0.0965|x|^2 - 0.0009|x|^4 + \ldots$$

By way of comparison, the square of the third Hankel singular value is 0.0962 so this estimate is tight. Figure 2 shows the outputs of the Hankel maps of

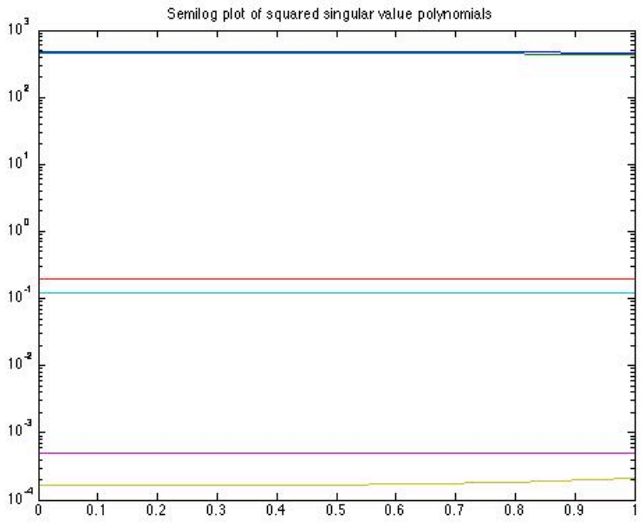

Fig. 1. Squared Singular Value Polynomials

the full and reduced systems excited by an optimal control $u_x(-\infty : 0)$ for random x.

Figure 3 shows the responses of the full nonlinear model, the reduced nonlinear model and the linear part of the full model to a sinusoidal input. The linear response has the largest amplitude and exceeds the total length of the three rods. The full nonlinear response has a small secondary oscillation that does not appear in the reduced nonlinear response.

8 Conclusion

We have presented a new normal form for the controllability and observability functions of a nonlinear control system. This normal form is valid through terms of degree $d+1$ where d is an integer chosen by the user and less than the degree of smoothness of the system. There are several advantages to this normal form. It can be computed term by term from the Taylor expansion of the nonlinear system. It is essentially unique for $d \leq 6$ and therefore gives an unambiguous measure of the relative importance of the different components of the state. A reduced order model can be constructed by projection onto the most important coordinates. One nice property of the reduced order model is that its controllability and observability functions are almost the restrictions of the controllability and observability functions of the full order model. Also the state space of the reduced order model almost achieves the intuitive

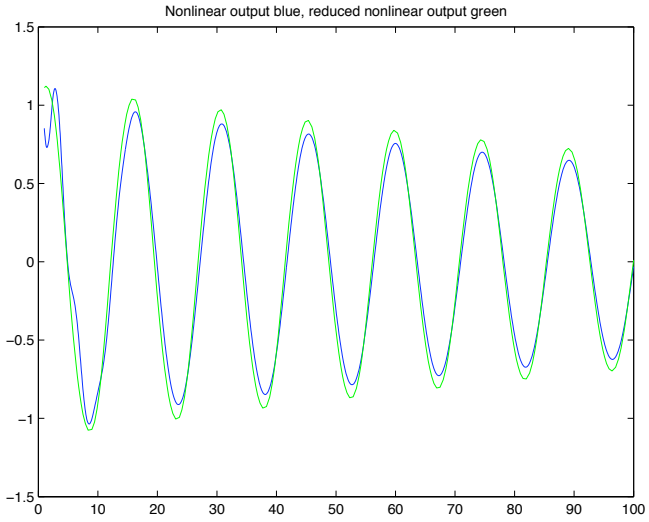

Fig. 2. Optimal Response

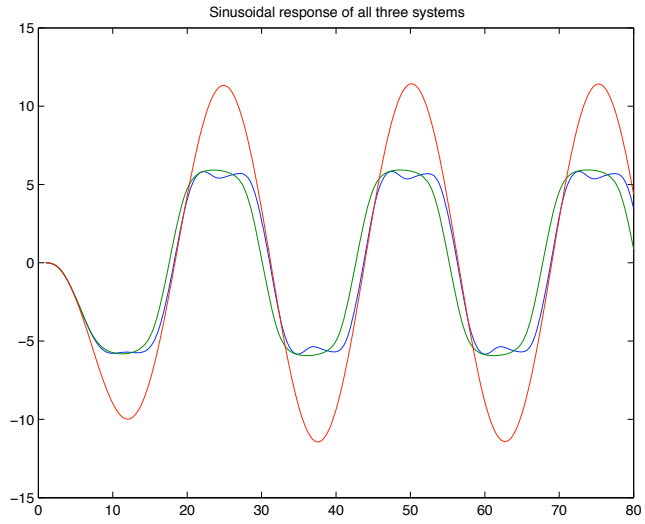

Fig. 3. Sinusoidal Response

goal of maximizing the observability function while holding the controllability function constant. Our methods readily extend to other forms of nonlinear model reduction such as LQG [7], [14] and H_∞ , [10], [13].

This paper is dedicated to my good friend and esteemed colleague Alberto Isidori on the occasion of his 65th birthday.

References

1. E. G. Al'brecht, *On the optimal stabilization of nonlinear systems*. PMM-J. Appl. Math. Mech., **25** pp. 1254–1266, 1961.
2. K. Fujimoto and J. M. A. Scherpen, *Nonlinear Balanced Realizations Based on Singular Value Analysis of Hankel Operators*. Proceeding of the IEEE CDC 2003.
3. K. Fujimoto and J. M. A. Scherpen, *Nonlinear Input-Normal Realizations Based on the Differential Eigenstructure of Hankel Operators*. IEEE Trans. Auto. Con., **50**, pp. 2-18, 2005.
4. K. Fujimoto and D. Tsubakino, *On Computation of Nonlinear Balanced Realization and Model Reduction*. Proceeding of the ACC, 2006.
5. K. Glover, *All Optimal Hankel-norm approximations of Linear Multivariable Systems and Their L^∞ error bounds*. Int. J. of Control, **39** pp.115-1193.
6. W. S. Grey and J. M. A. Scherpen, *On the Nonuniqueness of Singular Value Functions and Balanced Nonlinear Realizations*. Systems and Control Letters, **44** pp. 219-232, 2001.
7. E. A. Jonckheere and L. M. Silverman, *A New Set of Invariants for Linear Systems-Application to Reduced Order Compensator Design*. IEEE Trans. Auto. Con., **AC-28**, pp. 953-964, 1983.
8. D. L. Lukes. *Optimal regulation of nonlinear dynamical systems*, SIAM J. Contr., 7:75–100, 1969.
9. B. C. Moore, *Principle Component Analysis in Linear Systems: Controllability, Observability and Model Reduction*. IEEE Trans. Auto. Con., **AC-26**, pp. 17-32, 1981.
10. D. Mustafa and K. Glover, *Controller Reduction by H_∞ Balanced Truncation*. IEEE Trans. Auto. Con., **AC-36**, pp. 668-682, 1991.
11. A. J. Newman and P. S. Krishnaprasad, *Computation for Nonlinear Balancing*. Proceedings 37th IEEE CDC.,pp.4103-4104, 1998.
12. J. M. A. Scherpen, *Balancing for Nonlinear Systems*. Systems and Control Letters, **21** pp. 143-153, 1993.
13. J. M. A. Scherpen, *H_∞ Balancing for Nonlinear Systems*. Int. J. of Robust and Nonlinear Control, **6** pp. 645-668, 1996.
14. J. M. A. Scherpen and A. J. van der Schaft , *Normalized Coprime Factorizations and Balancing for Unstable Nonlinear Systems*. Int. J. of Control, **60** pp. 1193-1222, 1994.