

Computing balanced realizations for nonlinear systems

Andrew J. Newman[†], P. S. Krishnaprasad[‡]

[†]ALPHATECH, Inc.

2101 Wilson Boulevard, Suite 402

Arlington, VA 22201

[‡]Institute for Systems Research and

Department of Electrical and Computer Engineering

University of Maryland

College Park, MD 20742

newman@dc.alphatech.com, krishna@isr.umd.edu

Keywords: model reduction, nonlinear balancing, computation, Morse-Palais, Fokker-Planck.

Abstract

In 1993, Scherpen generalized the balanced truncation method to the nonlinear setting. However, the Scherpen procedure is not easily computable and has not yet been applied in practice. We offer methods, tools, and algorithms for computing the energy functions and coordinate transformations involved in the Scherpen theory and procedure for nonlinear balancing. We then apply our approach to derive, for the first time, balanced representations of nonlinear state-space models.

1 Introduction

This paper addresses the problem of computability pertaining to the Scherpen theory and procedure for balancing of nonlinear systems [1, 2]. We offer methods, tools, and algorithms toward computing balanced realizations for asymptotically stable affine nonlinear control systems, i.e., state-space models of the form

$$\dot{x}(t) = f(x(t)) + \sum_{i=1}^m g_i(x(t)) u_i(t) \quad (1)$$

$$y(t) = h(x(t)) \quad (2)$$

where $u = (u_1, \dots, u_m) \in U \subset \mathbf{R}^m$, $y = (y_1, \dots, y_p) \in \mathbf{R}^p$, and $x = (x_1, \dots, x_n)$ are local coordinates for a smooth state-space manifold M . The maps f, g_1, \dots, g_m are smooth and we assume that $f(0) = 0$ and $h(0) = 0$.

We say that f is stable (asymptotically stable) if 0 is a stable (asymptotically stable) equilibrium for $\dot{x} = f(x)$, and normally assume asymptotic stability of f . We refer to the triple (f, g, h) as a realization of the nonlinear system.

The two main objects that we work with in nonlinear balancing are the controllability and observability energy functions of the system, defined as follows.

Definition 1.1 (Controllability Function) *The controllability function, $L_c : \mathbf{R}^n \rightarrow \mathbf{R}$, for system (1)-(2) is defined by*

$$L_c(x_0) = \min_{\substack{u \in L_2(-\infty, 0) \\ x(-\infty) = 0, x(0) = x_0}} \frac{1}{2} \int_{-\infty}^0 \|u(t)\|^2 dt \quad (3)$$

□

Definition 1.2 (Observability Function) *The observability function, $L_o : \mathbf{R}^n \rightarrow \mathbf{R}$, for system (1)-(2) is defined by*

$$L_o(x_0) = \frac{1}{2} \int_0^{\infty} \|y(t)\|^2 dt \quad (4)$$

$$x(0) = x_0, \quad u(t) \equiv 0, t \geq 0$$

□

We will use the following results pertaining to the controllability function of a nonlinear system.

Proposition 1.3 *In the case of an linear time-invariant (LTI) system the controllability function specializes to the quadratic form*

$$L_c(x_0) = \frac{1}{2} x_0^\top W_c^{-1} x_0 \quad (5)$$

where the symmetric positive-definite matrix W_c is the familiar controllability Gramian. □

Theorem 1.4 (Scherpen [1, 2]) *Suppose that 0 is an asymptotically stable equilibrium of $-(f + g g^\top [\frac{\partial L_c}{\partial x}]^\top)$ on a neighborhood V of 0. Then, for all $x \in V$, L_c is the unique smooth solution of*

$$0 = \frac{\partial L_c}{\partial x}(x) f(x) + \frac{1}{2} \frac{\partial L_c}{\partial x}(x) g(x) g^\top(x) \left[\frac{\partial L_c}{\partial x}(x) \right]^\top \quad (6)$$

with supplementary condition $L_c(0) = 0$ and under the assumption that (6) has a smooth solution on V .

In this paper we will consider two of the fundamental problems involved in computing balanced realizations for nonlinear systems. First, we consider the problem of computing the controllability energy function without solving the family of optimal control problems implied in its definition, or solving the associated Hamilton-Jacobi-Bellman (HJB) equation (6). Then, we consider the problem of computing a Morse coordinate transformation under which the controllability function is locally quadratic on a neighborhood of 0. We apply the methods developed in this paper to two example systems: a forced damped pendulum system, and a forced damped double pendulum system.

Methods for computing the observability function and the balancing coordinate transformation, algorithmic details, and additional numerical studies can be found in [3]. We note here that the authors have developed a MATLAB toolbox for implementing the methods described herein, which was used for performing all of the numerical studies.

2 Stochastic Methods for Computation

We seek a method for computing the controllability energy function without solving the family of optimal control problems implied in its definition, or solving the associated HJB equation (6). In this section we present an approach, based primarily on the theory of stochastically excited dynamical systems (see, e.g., [4, 5]), for computing an estimate of the controllability function. We show that in certain situations the method provides an exact solution. Below, for integer k , we let \underline{k} denote the set $\{1, \dots, k\}$.

By a stochastically excited dynamical system we mean a control system for which the m components of the input, u_i , $i \in \underline{m}$, have been replaced by the sample paths of m Gaussian white noises, $\{(\zeta_t)_i, t \in \mathbf{R}^+\}$, $i \in \underline{m}$. The state equation is given by

$$\frac{d}{dt} X_t = f(X_t) + \sum_{i=1}^m g_i(X_t) (\zeta_t)_i \quad (7)$$

The white noise driven system (7) is interpreted correctly via the stochastic differential equation (SDE) (given elementwise) for $i \in \underline{n}$

$$(dX_t)_i = \bar{f}_i(X_t) dt + \sum_{i=1}^m g_i(X_t) (dW_t)_i \quad (8)$$

where for $i \in \underline{n}$

$$\bar{f}_i(X_t) = \left[f_i(X_t) + \frac{1}{2} \sum_{j=1}^n \sum_{k=1}^m \frac{\partial g_{ik}}{\partial X_j}(X_t) g_{jk}(X_t) \right] \quad (9)$$

includes the so called *correction term*, $g_{ij} = (g_j)_i$, $i \in \underline{n}$, $j \in \underline{m}$, and where solutions to (8) are defined in terms of a corresponding stochastic integral.

The state X_t is a Markov process with transition probability density $p(x, t; y, s)$. Time evolution of $p(x, t; y, s)$ is

governed by the Fokker-Planck equation, given by

$$\begin{aligned} \frac{\partial p}{\partial t}(x, t; y, s) = & \\ & - \sum_{i=1}^n \frac{\partial}{\partial x_i} (\bar{f}_i(x) p(x, t; y, s)) \\ & + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2}{\partial x_i \partial x_j} (b_{ij}(x) p(x, t; y, s)) \end{aligned} \quad (10)$$

with initial condition

$$p(x, s; y, s) = \delta(x - y)$$

and where

$$b_{ij}(x) = \sum_{k=1}^m g_{ik}(x) g_{jk}(x) = \left[[g(x)] [g(x)]^\top \right]_{ij}$$

Finally, in the steady-state, Equation (10) simplifies to the stationary Fokker-Planck equation

$$\begin{aligned} 0 = & - \sum_{i=1}^n \frac{\partial}{\partial x_i} (\bar{f}_i(x) p_\infty(x)) \\ & + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2}{\partial x_i \partial x_j} (b_{ij}(x) p_\infty(x)) \end{aligned} \quad (11)$$

where $p_\infty(x)$ denotes the stationary probability density (if it exists).

In the LTI case, the transition density function $p(x, t; y, s)$ describing the random properties of the state process X_t is Gaussian, and the stationary density p_∞ has zero mean and covariance equal to the controllability Gramian matrix W_c (see, e.g., [6]), i.e.,

$$p_\infty(x) = [(2\pi)^n \det(W_c)]^{-1/2} \exp\left(-\frac{1}{2} x^\top W_c^{-1} x\right) \quad (12)$$

Thus, in the LTI case, the controllability function L_c and the stationary density p_∞ are related exactly by

$$p_\infty(x) = [(2\pi)^n \det(W_c)]^{-1/2} \exp(-L_c(x)) \quad (13)$$

and

$$L_c(x) = -\log(p_\infty(x)) + \log\left([(2\pi)^n \det(W_c)]^{-1/2}\right) \quad (14)$$

In the nonlinear setting, the density $p(x, t; y, s)$, and in particular the stationary density $p_\infty(x)$, are not, in general, Gaussian, nor determined completely by their mean and covariance, i.e., higher order moments are involved. However, because the balancing coordinate transformation is local to a neighborhood of the origin, we are mainly interested in capturing a local characterization of the controllability function. In light of this, Equation (14) suggests that a useful approximation of L_c is defined by

$$L'_c(x) \triangleq -\log(p_\infty(x)) + C \quad (15)$$

where C is a normalizing constant, dependent on the particular system, such that $L'_c(0) = 0$. By Equation (14), L'_c specializes to the exact L_c in the LTI case.

There exist certain nonlinear systems for which Equation (15) provides an exact, rather than approximate, formula for the controllability function. As one simple example, consider the Langevin process X_t governed by the first-order SDE

$$dX_t = -\nabla \phi(X_t) + dW_t \quad (16)$$

where $\phi: \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a C^1 map such that $-\nabla \phi$ is asymptotically stable. The Maxwell-Boltzmann density

$$p_\infty^{\text{MB}}(x) = C \exp(-2\phi(x)) \quad (17)$$

satisfies the stationary Fokker-Planck equation where C is a constant such that $\int p_\infty^{\text{MB}} = 1$. Now, using (15), define

$$L_c^{\text{MB}}(x) = -\log(p_\infty^{\text{MB}}(x)) + \log(C) = 2\phi(x) \quad (18)$$

which is the unique controllability energy function for the Langevin system, i.e., stable affine nonlinear system with $f(x) = -\nabla \phi(x)$ and $g(x) = \mathbb{I}$.

Now we seek conditions under which a broader class of systems admits an exact relationship between the stationary density and the controllability function. In particular, we consider second-order mechanical systems with a Hamiltonian structure perturbed by dissipation and forcing. We determine conditions under which the controllability function for such a system can be expressed exactly in terms of the stationary density for the corresponding stochastically excited system. We adopt and modify somewhat the notation and framework of Fuller [7] and Zhu and Yang [8]. These authors have presented conditions under which exact solutions of the stationary Fokker-Planck equation can be derived. We show that in certain cases, the same conditions are sufficient for expressing the controllability function in terms of the stationary density, while in other cases, additional conditions are required.

We consider a forced, dissipatively perturbed, n -DOF Hamiltonian system. Let $q = (q_1, \dots, q_n) \in \mathbb{R}^n$ and $p = (p_1, \dots, p_n) \in \mathbb{R}^n$ denote, respectively, the generalized displacements and momenta. Let the Hamiltonian $H' = H'(q, p)$, i.e., the sum of the kinetic and potential energies of the system, be C^2 . Let $c'_{ij} = c'_{ij}(q, p)$ for $i, j \in \underline{n}$ be C^1 functions representing nonlinear dissipation coefficients. Let $d_{ij} = d_{ij}(q, p)$ for $i, j \in \underline{n}$ be C^2 . The system that we consider is governed by the equations of motion, for $i \in \underline{n}$

$$\dot{q}_i = \frac{\partial H'}{\partial p_i} \quad (19)$$

$$\dot{p}_i = -\frac{\partial H'}{\partial q_i} - \sum_{j=1}^n c'_{ij} \frac{\partial H'}{\partial p_j} + \sum_{k=1}^m d_{ik} u \quad (20)$$

The system is realized in standard state-space form with coordinates $x = (q, p) \in \mathbb{R}^{2n}$ and

$$f_i = \frac{\partial H'}{\partial p_i} \quad i = 1, \dots, n \quad (21)$$

$$f_i = -\frac{\partial H'}{\partial q_i} - \sum_{j=1}^n c'_{ij} \frac{\partial H'}{\partial p_j} \quad i = n+1, \dots, 2n$$

$$(g_k)_i = 0 \quad i = 1, \dots, n; \quad k = 1, \dots, m \quad (22)$$

$$(g_k)_i = d_{ik} \quad i = n+1, \dots, 2n; \quad k = 1, \dots, m$$

The output map h is irrelevant for purposes of the discussion here.

The corresponding stochastically excited system is governed by the SDEs, for $i \in \underline{n}$

$$dQ_i = \frac{\partial H'}{\partial P_i} dt \quad (23)$$

$$dP_i \quad (24)$$

$$= -\left(\frac{\partial H'}{\partial Q_i} + \sum_{j=1}^n c'_{ij} \frac{\partial H'}{\partial P_j} + \frac{1}{2} \sum_{j=1}^n \sum_{k=1}^m \frac{\partial d_{ik}}{\partial P_j} d_{jk} \right) dt + \sum_{k=1}^m d_{ik} (dW_t)_i$$

where we have adopted the usual notation by substituting Q for q and P for p when dealing with the corresponding random variables.

It is usually the case that the correction terms $\frac{1}{2} \sum_{j=1}^n \sum_{k=1}^m \frac{\partial d_{ik}}{\partial P_j} d_{jk}$, $i \in \underline{n}$, can be split into two parts: one which modifies the conservative forces and the other that modifies the damping forces (see [8]). This allows Equations (23) and (24) to be rewritten, for $i \in \underline{n}$

$$dQ_i = \frac{\partial H}{\partial P_i} dt \quad (25)$$

$$dP_i = -\left(\frac{\partial H}{\partial Q_i} + \sum_{j=1}^n c_{ij} \frac{\partial H}{\partial P_j} \right) dt + \sum_{k=1}^m d_{ik} (dW_t)_i \quad (26)$$

for appropriately defined H and c_{ij} .

The stationary Fokker-Planck equation governing the stationary transition density $p_\infty = p_\infty(q, p)$ associated with the SDEs (25) and (26) is given by

$$0 = \sum_{i=1}^n \left[-\frac{\partial}{\partial Q_i} \left(\frac{\partial H}{\partial P_i} p_\infty \right) + \frac{\partial}{\partial P_i} \left(\frac{\partial H}{\partial Q_i} p_\infty \right) \right] + \sum_{i=1}^n \left[\frac{\partial}{\partial P_i} \left(\sum_{j=1}^n c_{ij} \frac{\partial H}{\partial P_j} p_\infty \right) + \frac{1}{2} \sum_{j=1}^n \frac{\partial^2}{\partial P_i \partial P_j} (b_{ij} p_\infty) \right] \quad (27)$$

where

$$b_{ij} = \sum_{k=1}^m d_{ik} d_{jk} = [d] [d]^\top_{ij}$$

and subject to boundary conditions (vanishing probability flow)

$$\lim_{\|(q,p)\| \rightarrow \infty} \frac{\partial H}{\partial P_i} p_\infty = 0 \quad (28)$$

and

$$\lim_{\|(q,p)\| \rightarrow \infty} \left(\frac{\partial H}{\partial Q_i} + \sum_{j=1}^n c_{ij} \frac{\partial H}{\partial P_j} \right) p_\infty \quad (29)$$

$$+ \frac{1}{2} \sum_{j=1}^n \frac{\partial}{\partial P_j} (b_{ij} p_\infty) = 0$$

Observe that the first summation term on the right-hand-side of (27) is equal to the Poisson bracket of p_∞ and H , i.e.,

$$\begin{aligned} \{p_\infty, H\} & \quad (30) \\ &= \sum_{i=1}^n \left[-\frac{\partial H}{\partial P_i} \frac{\partial p_\infty}{\partial Q_i} + \frac{\partial H}{\partial Q_i} \frac{\partial p_\infty}{\partial P_i} \right] \\ &= \sum_{i=1}^n \left[-\frac{\partial}{\partial Q_i} \left(\frac{\partial H}{\partial P_i} p_\infty \right) + \frac{\partial}{\partial P_i} \left(\frac{\partial H}{\partial Q_i} p_\infty \right) \right] \end{aligned}$$

Thus, we can rewrite the stationary Fokker-Planck equation (27) as

$$\begin{aligned} 0 & \quad (31) \\ &= \{p_\infty, H\} \\ &+ \sum_{i=1}^n \left[\frac{\partial}{\partial P_i} \sum_{j=1}^n \left(c_{ij} \frac{\partial H}{\partial P_j} p_\infty \right) + \frac{1}{2} \sum_{j=1}^n \frac{\partial^2}{\partial P_i \partial P_j} (b_{ij} p_\infty) \right] \end{aligned}$$

The relationship between the stationary density and the controllability function, and an exact formula for the latter in terms of the Hamiltonian, are given in the following results. We first consider the case where the parameters c'_{ij} and d_{ik} are independent of q and p , i.e., are constants. In this situation, the correction term vanishes, $H = H'$, and $c_{ij} = c'_{ij}$. The following is modified from Fuller [7].

Theorem 2.1 (Fuller [7]) Consider the stochastically excited system corresponding to the forced, dissipatively perturbed, n -DOF Hamiltonian system governed by the SDEs (25)-(26) where H is the Hamiltonian. Suppose that the coefficients c_{ij} and d_{ik} are independent of q and p . Furthermore, suppose that the following constant ratio holds for all $i, j \in \underline{n}$:

$$\frac{c_{ij}}{b_{ij}} = \ell = \text{constant} \quad (32)$$

Then the unique stationary density p_∞ that satisfies Equation (31) is

$$p_\infty(q, p) = C \exp(-2\ell H(q, p)) \quad (33)$$

where C is a constant such that $\int p_\infty = 1$.

Remark 2.2 The density (33) is of Maxwell-Boltzmann form. \square

Remark 2.3 The condition (32) is referred to as the equipartition of energy condition. The terminology derives from the situation in statistical mechanics where each DOF of a multi-particle system is associated with the same mean energy. \square

Remark 2.4 The equipartition of energy condition imposes a severe restriction on the class of systems for which (33) is the stationary density. \square

The controllability function L_c uniquely satisfies the HJB equation (6), which, for realization (f, g) given in Equations (21) and (22), takes the form

$$0 = \{L_c, H\} + \sum_{i=1}^n \left[\frac{\partial L_c}{\partial p_i} \sum_{j=1}^n \left(-c_{ij} \frac{\partial H}{\partial p_j} + \frac{1}{2} b_{ij} \frac{\partial L_c}{\partial p_j} \right) \right] \quad (34)$$

The relationship between the stationary density and the controllability function, and an exact formula for the latter in terms of the Hamiltonian, are given in the following result.

Theorem 2.5 Consider the forced, dissipatively perturbed, n -DOF Hamiltonian control system, governed by the evolution equations (19) and (20), and realized by (f, g) given in Equations (21) and (22). Under the conditions stated in Theorem 2.1, the unique controllability energy function for the system is given by

$$\begin{aligned} L_c(q, p) &= -\log(p_\infty(q, p)) + C' \\ &= 2\ell H(q, p) + C' \end{aligned} \quad (35)$$

where p_∞ is the stationary density of the corresponding stochastically excited system and C' is a constant such that $L_c(0, 0) = 0$.

Proof It is necessary and sufficient to show that L_c satisfies Equation (34). Since L_c is a functional of H , we have that $\{L_c, H\} = 0$. Furthermore, $\frac{\partial L_c}{\partial p_j} = 2\ell \frac{\partial H}{\partial p_j}$, so that Equation (34) becomes

$$\begin{aligned} 0 &= \sum_{i=1}^n \left[2\ell \frac{\partial H}{\partial p_i} \sum_{j=1}^n \left(-c_{ij} \frac{\partial H}{\partial p_j} + \ell b_{ij} \frac{\partial H}{\partial p_j} \right) \right] \quad (36) \\ &= \sum_{i=1}^n \left[2\ell \frac{\partial H}{\partial p_i} \sum_{j=1}^n \frac{\partial H}{\partial p_j} (\ell b_{ij} - c_{ij}) \right] \quad (37) \end{aligned}$$

which is clearly satisfied given the equipartition of energy condition (32). \blacksquare

We now consider the more general situation where the parameters c'_{ij} and d_{ik} are permitted to be functions of q and p . The following is modified from Zhu and Yang [8].

Theorem 2.6 (Zhu and Yang [8]) Consider the stochastically excited system corresponding to the forced, dissipatively perturbed, n -DOF Hamiltonian system governed by the SDEs (25)-(26) where H is the Hamiltonian. Suppose that the following ratio holds for all $i \in \underline{n}$ and for some functional h of H :

$$\frac{\sum_{j=1}^n \left(2c_{ij} \frac{\partial H}{\partial P_j} + \frac{\partial b_{ij}}{\partial P_j} \right)}{\sum_{j=1}^n b_{ij} \frac{\partial H}{\partial P_j}} = h(H) \quad (38)$$

Then the unique stationary density p_∞ that satisfies Equation (31) is

$$p_\infty(q, p) = C \exp\left(-\int_0^{H(q,p)} h(u) du\right) \quad (39)$$

where C is a constant such that $\int p_\infty = 1$.

Remark 2.7 The density (39) is of Maxwell-Boltzmann form. \square

Remark 2.8 The condition (38) is analogous to an equipartition of energy condition, again imposing a severe restriction on the class of systems for which (39) is the stationary density. \square

The relationship between the stationary density and the controllability function, and an exact formula for the latter in terms of the Hamiltonian, are given in the following result.

Theorem 2.9 Consider the forced, dissipatively perturbed, n -DOF Hamiltonian control system, governed by the evolution equations (19) and (20), and realized by (f, g) given in Equations (21) and (22). Suppose that the following ratio holds for all $i \in \underline{n}$ and for some functional r of H :

$$\frac{\sum_{j=1}^n c_{ij} \frac{\partial H}{\partial p_j}}{\sum_{j=1}^n b_{ij} \frac{\partial H}{\partial p_j}} = r(H) \quad (40)$$

Then the unique controllability energy function for the system is given by

$$L_c(q, p) = 2 \int_0^{H(q,p)} r(u) du + C' \quad (41)$$

where C' is a constant such that $L_c(0, 0) = 0$. Furthermore, if the b_{ij} are independent of p then

$$L_c(q, p) = -\log(p_\infty(q, p)) + C' \quad (42)$$

where p_∞ is the stationary density of the corresponding stochastically excited system.

Proof Assume that $L_c(q, p) = \phi(H(q, p))$ for some functional ϕ of H . Then $\{L_c, H\} = 0$. The HJB equation (34) can be written

$$0 = \sum_{i=1}^n \frac{\partial H}{\partial p_i} \frac{\partial \phi}{\partial H} \left(b_{ij} \frac{\partial H}{\partial p_j} \frac{\partial \phi}{\partial H} - 2 c_{ij} \frac{\partial H}{\partial p_j} \right) \quad (43)$$

which is clearly satisfied if we assign $\frac{\partial \phi}{\partial H} = 2r(H)$ where $r(H)$ is defined by Equation (40). Thus, the desired functional ϕ is obtained through integration yielding Equation (41). Moreover, if the b_{ij} are independent of p then $2r(H) = h(H)$ where $h(H)$ is defined in Equation (38). In that case Equation (42) holds. \blacksquare

Additional details appear in [3].

3 Computing the Morse Coordinate Transformation

Recall that for an LTI system, the energy functions L_c and L_o globally take the form of quadratic functions given, for example, by (5). We wish to generalize the linear balancing procedure to the nonlinear setting, but the functions L_c and L_o are not, in general, quadratic. However, we can appeal to some important results from critical point theory (see, e.g., [9]) in order to find a change of coordinates under which a smooth function takes a quadratic form locally around a non-degenerate critical point. The key result is the Morse lemma [10], which guarantees the existence of the desired canonical form for functions with a non-degenerate critical point (called Morse functions) defined on a finite-dimensional manifold, and an analogous result of Palais [11], which generalizes the notion to functions defined on a Hilbert space. Our presentation of the established results are based mainly on those found in [9, 12].

Theorem 3.1 (Morse-Palais) Let f be a smooth real-valued function defined on an open neighborhood \mathcal{O} of 0 in the Hilbert space \mathcal{E} . Assume that $f(0) = 0$ and that 0 is a non-degenerate critical point of f . Then there exists a neighborhood $U \subset \mathcal{O}$ of 0, a local change of coordinates ϕ on U , and an invertible symmetric operator \mathcal{A} such that

$$f(x) = \langle \mathcal{A}\phi(x), \phi(x) \rangle_{\mathcal{E}} \quad x \in U \quad (44)$$

\square

Corollary 3.2 Let f be a smooth real-valued function defined on an open neighborhood \mathcal{O} of 0 in the Hilbert space \mathcal{E} . Assume that $f(0) = 0$ and that 0 is a non-degenerate critical point of f . Then there exists a neighborhood $U \subset \mathcal{O}$ of 0, a local change of coordinates $z = \xi(x)$ on U , and an orthogonal decomposition $\mathcal{E} = \mathcal{F} + \mathcal{F}^\perp$ such that if we write $z = \xi(x) = u + v$ with $u \in \mathcal{F}$ and $v \in \mathcal{F}^\perp$ then

$$f(z) = f(\xi(x)) = \langle u, u \rangle_{\mathcal{E}} - \langle v, v \rangle_{\mathcal{E}} \quad x \in U \quad (45)$$

\square

Remark 3.3 Consider the special case where $\mathcal{E} = \mathbf{R}^n$ and critical point 0 has index r . Define $z = \xi(x)$ for $x \in U$ and $\psi = \xi^{-1}$ on $\xi(U)$. Then Corollary 3.2 implies that

$$f(x) = f(\psi(z)) = -\sum_{i=1}^r z_i^2 + \sum_{i=r+1}^n z_i^2 \quad (46)$$

In the new coordinates, the function f is said to be in spherical quadratic form. The transformation is illustrated in Figure 1. \square

Definition 3.4 (Morse Coordinate Transformation) A change of coordinates ψ satisfying (46) is said to be a Morse coordinate transformation for f around 0. \square

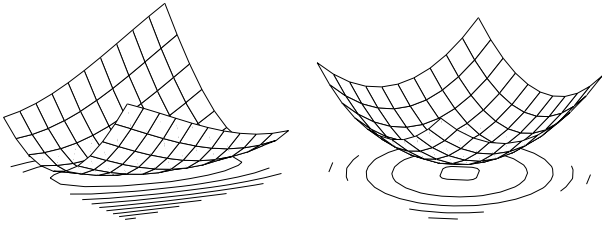


Figure 1: An example of a Morse function on \mathbb{R}^2 (with level contours) before and after transformation to spherical quadratic form.

The original proof by Morse uses the Gram-Schmidt orthogonalization process which is essentially a coordinate-by-coordinate induction argument. The generalization by Palais is proved without a coordinate-wise procedure, which we use to our advantage for purposes of computation. For purposes of brevity, the proofs do not appear here. We proceed directly to the algorithms.

We present here, somewhat loosely, an algorithm for numerical implementation of Theorem 3.1 and Corollary 3.2. The algorithm is presented more rigorously in [3] where a computational framework is introduced. The algorithm takes a Morse function f and returns a neighborhood U , Morse coordinate transformation ϕ , and invertible symmetric matrix A under which f takes the desired form (44) on U . An additional algorithm takes ϕ , A , and U and returns a coordinate transformation ξ under which f takes the spherical quadratic form (45). The main building blocks are as follows.

smooth function decomposition Given smooth real-valued function f with critical point at 0, return smooth functions g_i , $i \in \underline{n}$, such that

$$f(x) = \sum_{i=1}^n g_i(x) x_i \quad i \in \underline{n}, \quad x \in \mathcal{O} \quad (47)$$

and

$$g_i(0) = \frac{\partial f}{\partial x_i}(0) \quad i \in \underline{n} \quad (48)$$

1. Compute (approximate) partial derivatives $\frac{\partial f}{\partial x_i}$, $i \in \underline{n}$.

2. For each point x in the domain of definition of f , compute (approximate) integrals

$$g_i(x) = \int_0^1 \frac{\partial f}{\partial x_i}(tx) dt, \quad i \in \underline{n}.$$

Morse function decomposition Given Morse function f , i.e., f has non-degenerate critical point at 0, return smooth functions h_{ij} , $i, j \in \underline{n}$ such that

$$f(x) = \sum_{i=1}^n \sum_{j=1}^n h_{ij}(x) x_i x_j \quad i, j \in \underline{n}, \quad x \in \mathcal{O} \quad (49)$$

with symmetry property

$$h_{ij}(x) = h_{ji}(x) \quad i, j \in \underline{n}, \quad x \in \mathcal{O} \quad (50)$$

and such that

$$h_{ij}(0) = \frac{1}{2} \frac{\partial^2 f}{\partial x_i \partial x_j}(0) = \frac{1}{2} D^2 f(0) \quad (51)$$

This is accomplished via $n+1$ smooth function decompositions.

1. Apply the smooth function decomposition to f yielding g_i , $i \in \underline{n}$.
2. Apply the smooth function decomposition to each of the g_i yielding h_{ij} , $i, j \in \underline{n}$.

matrix square root Given matrix B close to the identity, return its square root C , i.e., $B = C^2$. The matrix B must satisfy

$$\|\mathbb{I} - B\| < 1 \quad (52)$$

In that case, the following algorithm converges to a fixed point corresponding to the desired matrix $C = B^{1/2}$.

$$\begin{aligned} C_{k+1} &= C_k + \frac{1}{2} (B - C_k^2) & k = 0, 1, \dots \\ C_0 &= \mathbb{I} \end{aligned} \quad (53)$$

The convergence of the sequence $\{C_k\}$ to the fixed point $B^{1/2}$ can be shown to be a consequence of the contraction mapping principle.

Morse-Palais transformation Given Morse function f , return neighborhood U , coordinate transformation ϕ , and invertible symmetric matrix A such that Equation (44) holds.

1. Apply the Morse function decomposition to f yielding h_{ij} , $i, j \in \underline{n}$. Let $H(x) = [h_{ij}(x)]$ and $A = H(0)$. Note that the nondegeneracy of 0 ensures the invertibility of A .
2. For each point x in the domain of definition of f :
 - (a) Compute the solution B of the matrix equation $AB = H(x)$. Note that $B = \mathbb{I}$ at $x = 0$.
 - (b) If $\|\mathbb{I} - B\| < 1$ then:
 - i. Apply the matrix square root algorithm to compute $C = B^{1/2}$.
 - ii. Let $\phi(x) = Cx$.
 - iii. Include the point x in the neighborhood U .
 - (c) Otherwise, do not include the point x in the neighborhood U and no further calculations apply.

This procedure provides an estimate of the neighborhood U for which the function can be transformed to the canonical quadratic form. It is possible that the maximal neighborhood is larger.

spherical transformation Given transformation ϕ and invertible symmetric matrix A such that Equation (44) holds, return index r and coordinate transformation ψ such that Equation (46) holds.

1. Compute the spectral decomposition of matrix A , i.e., $A = V \Lambda V^T$.
2. Let $E = \text{diag}(|\lambda_1|, \dots, |\lambda_n|)$.
3. Let r equal the number of λ_i such that $\lambda_i < 0$.
4. Let $R = E V^T$.
5. For each point x in the domain of definition of f , let $\psi(x) = R \phi(x)$.

Remark 3.5 *The terminology ‘‘Morse function decomposition’’ is somewhat misleading since the decomposition (49) merely requires a critical point that is not necessarily non-degenerate. However, we adopt the terminology for lack of a better name and because we are applying the decomposition to Morse functions.* \square

4 Applications

In this section we illustrate our methods and algorithms by applying them to two examples of rigid link mechanical systems. We compute a balanced realization for a forced damped pendulum system, and take steps toward balancing a forced damped double pendulum system.

The first example that we consider is a simple pendulum system as illustrated in Figure 2. The system incorporates linear torsional damping, linear torsional stiffness, and a torque input at the rotary joint. We assume that the shaft is massless and that the pendulum moves only in the plane. We consider the case where the joint angle is measured (position read-out). (See [3] for the case of velocity read-out.)

It is beneficial to study the pendulum as an example because

- it is nearly linear, so we can use linear theory to obtain a good estimate of the correct results for comparison; and
- in previous sections we have studied second-order mechanical systems and obtained an exact formula for the controllability function.

We obtain a state-space realization (f, g, h) for the pendulum system via Lagrangian mechanics. Let the generalized position q and velocity \dot{q} correspond to the joint angle θ and angular velocity $\dot{\theta}$, respectively. The affine nonlinear control system is realized in coordinates $x = (x_1, x_2) = (q, \dot{q})$ by

$$f(x) = \begin{bmatrix} x_2 \\ -\frac{G}{L} \sin(x_1) - \frac{k}{m L^2} x_1 - \frac{b}{m L^2} x_2 \end{bmatrix} \quad (54)$$

$$g(x) = \begin{bmatrix} 0 \\ \frac{1}{m L^2} \end{bmatrix} \quad (55)$$

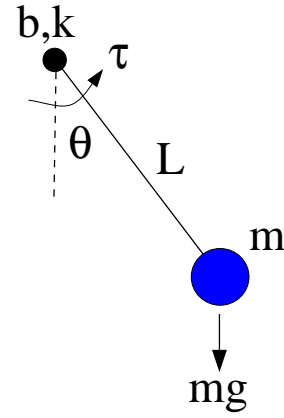


Figure 2: Planar pendulum system with massless shaft, linear torsional damping, linear torsional stiffness, and torque input applied at the rotary joint. Parameter values used for numerical studies: $m = 1/40$, $b = 2$, $k = 1$, $L = 20$, $G = 10$.

and $h(x) = x_1$.

The Hamiltonian H for the pendulum system is given by

$$\begin{aligned} H(x) &= K(x) + U(x) \\ &= \frac{1}{2} m L^2 x_2^2 + \frac{1}{2} k x_1^2 - m G L \cos(x_1) \end{aligned} \quad (56)$$

Furthermore, the equipartition of energy condition (32) is satisfied trivially for the 1-DOF system with ratio $\ell = b$. Applying Theorem 2.5, the controllability function L_c is given, exactly, by

$$\begin{aligned} L_c(x) &= -2 b m G L \cos(x_1) + b k x_1^2 \\ &\quad + b m L^2 x_2^2 + 2 b m G L \end{aligned} \quad (57)$$

Numerical studies were conducted via application of our algorithms to the pendulum system with parameter values given in Figure 2. The observability function L_o was computed using an algorithm based directly on Equation (4). The controllability and observability functions for the pendulum system with position read-out are shown in Figure 3.

We also computed an approximation to the controllability function via Equation (15) and a Monte-Carlo approach. We simulated 50,000 sample paths for the pendulum system with approximate Gaussian white noise injected as the torque input. We assumed that steady-state was reached after 60 time units, 6 times the largest time constant of the system.

The results of the Monte-Carlo experiments are presented in Figure 4. In this case, we used a relatively coarse grid, which is roughly the highest resolution that provides a smooth approximation. By generating additional sample paths, we can increase the grid resolution while maintaining a smooth approximation.

Even though we have an exact expression for L_c , we use the approximation L'_c generated by the Monte-Carlo experiments in order to compute a balanced realization for the pendulum system. We do this to demonstrate that L'_c can be a

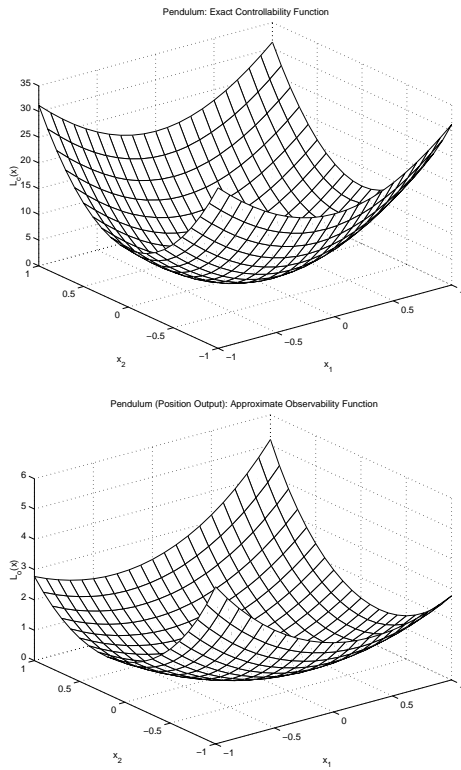


Figure 3: The exact controllability function (top) and observability function (bottom) for the pendulum system with position read-out.

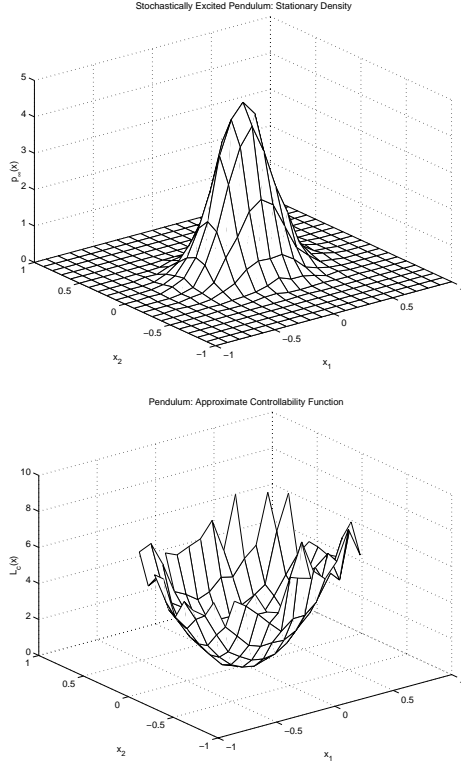


Figure 4: The stationary density (top) and derived approximate controllability function (bottom) for the pendulum system.

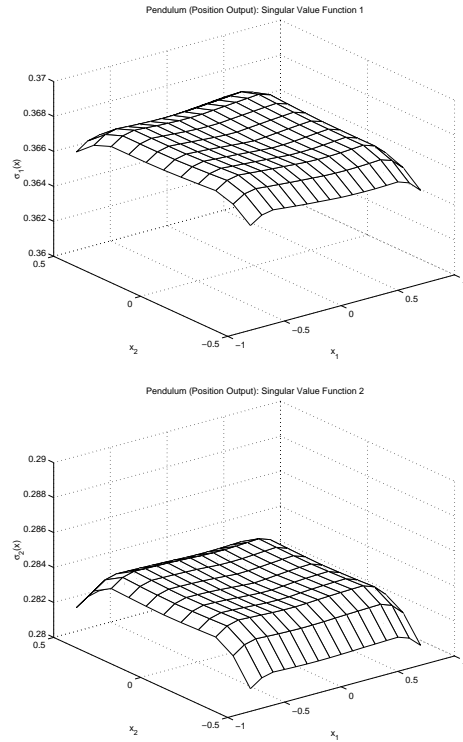


Figure 5: The singular value functions for the pendulum system with position output. Left: $\sigma_1(x)$ nearly constant 0.367; Right: $\sigma_2(x)$ nearly constant 0.284.

suitable surrogate for L_c in the balancing algorithms. The algorithm presented in Section 3 was applied to L'_c to compute a Morse coordinate transformation. Further computations (see [3]) yielded a balanced representation for the pendulum system with position read-out.

The computed singular value functions (see [1, 2]) are shown in Figure 5. Because the pendulum system is nearly linear, we expect the singular value functions to be nearly constant at the value of the corresponding Hankel singular values of the linearized system. This is reflected in the computations. The singular value functions are nearly constant at grid points close to the origin, taking values close to 0.367 and 0.284. This closely matches the Hankel singular values of the linearized system which are 0.3671 and 0.2838. One state is roughly 1.3 times as important to the input-to-output behavior of the system.

We simulated the pendulum system in the original and balanced coordinates using two input signals: $u \equiv 0$ (natural response) and $u(t) = 0.5 \sin(t/\pi)$. The output responses are shown in Figure 6. Theoretically, the output responses of the original and balanced systems should be identical, since they are merely different representations of the same physical system. However, the computations introduce numerical error.

In [3], it was demonstrated that by using the exact controllability function, the output responses of the original and balanced systems are virtually identical. Thus, the algorithms for computing the Morse, input-normal, and balanc-

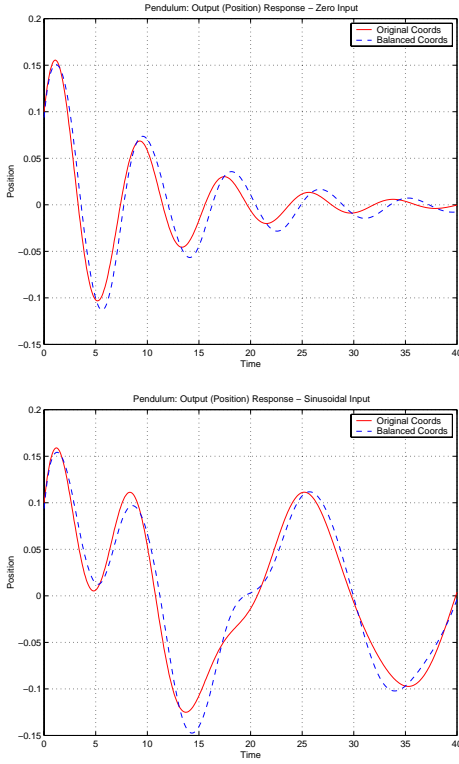


Figure 6: Output response for the pendulum system with position read-out: original coordinates (solid) vs. balanced coordinates (dashed). Top: zero input, approximate L_c ; Bottom: sinusoidal input, approximate L_c .

ing transformations introduced negligible error. On the other hand, when using the approximate controllability function generated using Monte-Carlo data, the output responses of the original and balanced systems deviate somewhat. Thus, a better approximation may be desirable, which can be achieved by generating additional Monte-Carlo data.

We now consider a double pendulum system as illustrated in Figure 7. As with the pendulum system, the system incorporates linear torsional damping, linear torsional stiffness, and torque inputs at the rotary joints. We assume that the shafts are massless and that the pendulum moves only in the plane. We measure the horizontal position of the end-effector as the system output (a nonlinear function of the state variables).

As before, we obtain a state-space realization (f, g, h) for the pendulum system via Lagrangian mechanics. The affine nonlinear control system is realized in coordinates $x = (x_1, x_2, x_3, x_4) = (q_1, q_2, \dot{q}_1, \dot{q}_2)$ by

$$f(x) = \begin{bmatrix} \dot{q} \\ -M^{-1}(q)(C(q, \dot{q}) + N(q, \dot{q})) \end{bmatrix} \quad (58)$$

$$g(x) = \begin{bmatrix} 0 \\ M^{-1}(q) \end{bmatrix} \quad (59)$$

where expressions for the matrices M , C , and N can be found in [3] and $h(x) = L_1 \sin(q_1) + L_2 \sin(q_1 + q_2)$.

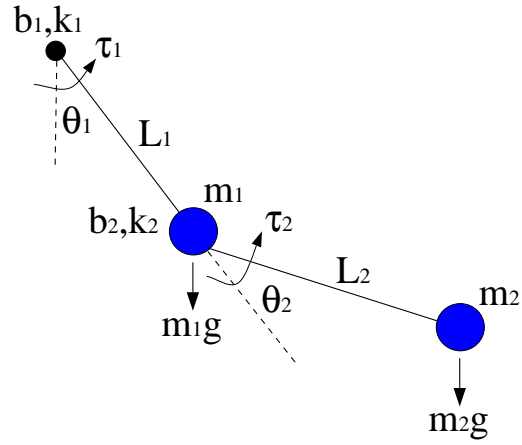


Figure 7: Planar double pendulum system with massless shafts, linear torsional damping, linear torsional stiffness, and torque input applied at the rotary joints. Parameters values used for numerical studies: $m_1 = m_2 = b_1 = b_2 = k_1 = k_2 = L_1 = L_2 = 1, G = 10$.

The double pendulum system is not integrable and does not, in general, satisfy the equipartition of energy condition. However, in the special case where $b_1 = b = b_2$ the equipartition of energy condition is satisfied with ratio $\ell = b$. Applying Theorem 2.5, the controllability function for the double pendulum system is given, exactly, by

$$L_c(x) = \begin{aligned} & b(m_1 + m_2)L_1^2 x_3^2 + b m_2 L_2^2 (x_3 + x_4)^2 + \\ & 2 b m_2 L_1 L_2 \cos(x_2) x_3 (x_3 + x_4) + \\ & b k_1 x_1^2 + k_2 x_2^2 - 2 b (m_1 + m_2) G L_1 \cos(x_1) - \\ & 2 b m_2 G L_2 \cos(x_1 + x_2) + \\ & 2 b G ((m_1 + m_2) L_1 + m_2 L_2) \end{aligned} \quad (60)$$

A numerical study was conducted via application of our algorithms to the double pendulum system with parameter values given in Figure 7. Graphical representations of the controllability and observability functions are presented in [3]. Figure 8 shows a 2-dimensional slice of each singular value function for the double pendulum system. At the origin, the singular value functions take the values, respectively, 0.487, 0.444, 0.135, and 0.050. These values are reasonable close to what we expect from the Hankel singular values of the linearization, which are 0.5029, 0.4702, 0.0249, and 0.0106. Two states of the balanced realization have considerably greater input-to-output importance than the other two states. We also observe that numerical errors are more prominent for the singular value functions of small magnitude, i.e., the oscillations that they display are likely caused by numerical error rather than being an accurate reflection of their actual behavior.

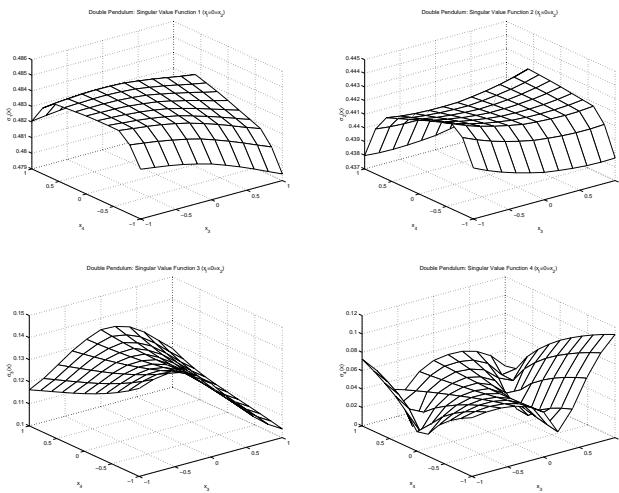


Figure 8: Singular value functions for double pendulum (x_3 - x_4 plane). Top left: $\sigma_1(x)$; Top right: $\sigma_2(x)$; Bottom left: $\sigma_3(x)$; Bottom right: $\sigma_4(x)$.

5 Concluding Remarks

Because of the computational complexity of our algorithms, this research merely represents a first step toward making the balancing procedure a practical model reduction tool. It is also clear that results with broader applicability will be of benefit.

Acknowledgments

The authors wish to acknowledge Prof. Jacqueline Scherpen of Delft University, The Netherlands, and Prof. W. Steven Gray of Old Dominion University, Virginia, USA, for many useful discussions and prior review of much of this work.

This research was supported by grants from the Northrop Grumman Foundation, the National Science Foundation's Engineering Research Centers Program: NSFD CDR 8803012 and NSF Grant EEC-9527576, and the Army Research Office under the ODDR&E MURI97 Program Grant No. DAAG55-97-1-0114 (through Harvard University).

References

- [1] Jacqueliën M. A. Scherpen, *Balancing for Nonlinear Systems*, *Systems and Control Letters*, 21:143-153, 1993
- [2] Jacqueliën M. A. Scherpen, *Balancing for Nonlinear Systems*, Ph.D. Thesis, University of Twente, The Netherlands, 1994
- [3] Andrew J. Newman, *Modeling and Reduction with Applications to Semiconductor Manufacturing*, Ph. D. Thesis, Institute for Systems Research Ph.D. 99-5 and Center for Dynamics and Control of Smart Structures Ph.D. 99-2, University of Maryland, College Park, 1999
- [4] Ludwig Arnold, *Stochastic Differential Equations: Theory and Applications*, Wiley, 1974
- [5] Eugene Wong, *Stochastic Processes in Information and Dynamical Systems*, McGraw-Hill, 1971
- [6] M. H. A. Davis, *Linear Estimation and Stochastic Control*, Halsted Press, 1977
- [7] A. T. Fuller, *Analysis of Nonlinear Stochastic Systems by Means of the Fokker-Planck Equation*, *International Journal of Control*, 9(6):603-655, 1969
- [8] W. Q. Zhu and Y. Q. Yang, *Exact Stationary Solutions of Stochastically Excited and Dissipated Integrable Hamiltonian Systems*, *Journal of Applied Mechanics*, 63:493-500, 1996
- [9] J. Milnor, *Morse Theory*, Princeton University Press, 1963
- [10] M. Morse, *The Calculus of Variations in the Large*, AMS, 1934
- [11] Richard S. Palais, *Morse Theory on Hilbert Manifolds*, *Topology*, 2:299-340, 1963
- [12] Serge Lang, *Differential Manifolds*, Springer-Verlag, 1985