Galerkin Approximations of the Generalized Hamilton–Jacobi–Bellman Equation*

RANDAL W. BEARD,† GEORGE N. SARIDIS‡ and JOHN T. WEN†

The application of Galerkin’s method yields an efficient algorithm for the approximation of the generalized Hamilton–Jacobi–Bellman equation on a compact set. The result is a feedback control with a pre-defined set of attraction.

Key Words—Nonlinear control; optimal control; Galerkin approximation; feedback synthesis; generalized Hamilton–Jacobi–Bellman equation.

Abstract—In this paper we study the convergence of the Galerkin approximation method applied to the generalized Hamilton–Jacobi–Bellman (GHJB) equation over a compact set containing the origin. The GHJB equation gives the cost of an arbitrary control law and can be used to improve the performance of this control. The GHJB equation can also be used to successively approximate the Hamilton–Jacobi–Bellman equation. We state sufficient conditions that guarantee that the Galerkin approximation converges to the solution of the GHJB equation and that the resulting approximate control is stabilizing on the same region as the initial control. The method is demonstrated on a simple nonlinear system and is compared to a result obtained by using exact feedback linearization in conjunction with the LQR design method. © 1997 Elsevier Science Ltd. All rights reserved.

1. INTRODUCTION

The basic mathematical theory for classical optimal control is well established (Anderson and Moore, 1971; Bryson and Ho, 1975; Kirk, 1970; Lewis, 1986; Sage and White III, 1977). If we assume full-state knowledge and if the dynamics of the system are modeled by linear dynamics and the cost functional to be optimized is quadratic in the state and control, then the optimal control is a linear feedback of the states, where the control gains are obtained by solving a standard Riccati equation. The practical success of linear quadratic (LQ) methods in optimal control is directly linked to successful algorithms for solving the Riccati equation. However, if the system is modeled by nonlinear dynamics or the cost functional to be optimized is not quadratic, then the optimal control is a state feedback function that depends on the solution to the Hamilton–Jacobi–Bellman (HJB) equation. The HJB equation is extremely difficult to solve in general rendering optimal control techniques of limited use for nonlinear systems.

Motivated by the success of optimal control methods for linear systems, there has been a great deal of research devoted to approximating the HJB equation. If an open-loop solution is acceptable then there are a number of methods to solve the optimal control problem. A common approach is to numerically solve for the state and co-state equations obtained from a Hamiltonian formulation of the optimal control problem. The problem can be reduced to a two-point boundary-value problem which can be solved by various methods (Kirk, 1970; Sage and White III, 1977). In Bosarge et al. (1973) the two-point boundary-value problem is solved using the Ritz–Galerkin approximation theory that is employed (in a different context) in this paper. In Hofer and Tibken (1988) the authors reduce the optimal bilinear control problem to successive iterations of a sequence of Riccati equations. In Aganovic and Gajic (1994) the same problem is further reduced to successive approximations of a sequence of Lyapunov equations. In Cebuhar and Costanza (1984) the bilinear control problem is reduced to a sequence of linear control problems that converge uniformly to the optimal bilinear control. Another approach taken in Rosen and Luus (1992) is to cast the nonlinear optimal control problem in the form of a nonlinear programming problem. The method is formulated in path space so the point that solves the nonlinear programming problem is the optimal path of the system.

*Received 22 June 1995; revised 3 June 1996; received in final form 23 June 1997. An earlier version of this paper was presented at the 13th IFAC World Congress, held in San Francisco, CA, USA, July 1–5, 1996. This paper was recommended for publication in revised form by Associate Editor Matt James under the direction of Editor Tamer Başar. Corresponding Author: Professor Randal Beard. Tel. + 801 378-8392; Fax 001 801 378 6586; E-mail: beard@ee.byu.edu.

†Department of Electrical and Computer Engineering, Brigham Young University, Provo, UT 84602, U.S.A.
‡Department of Electrical, Computer and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180, U.S.A.

159
Since open-loop control is undesirable for practical systems, various approaches have been investigated to generate closed-loop solutions to the HJB equation. One technique is the perturbation method described in Albrecht (1961), Lukes (1969), Garrard and Jordan (1977) and Nishikawa et al. (1971). In this method, the nonlinear system is assumed to be a perturbation of a linear system. The optimal cost and control are assumed to be analytic and are expanded in a Taylor series. Various techniques are then employed to find the first few terms in the series. The first term corresponds to the solution of the matrix Riccati equation obtained by linearizing the system about the origin. The next term is a third-order approximation to the control and can also be expressed as matrix equations. Higher-order terms require the solution of a linear partial differential equation and so the algorithm is usually terminated after the first two terms. In Lukes (1969), the authors present a definitive study of analytic, infinite-time regulators. In Werner and Cruz (1968), it is shown that an nth-order Taylor series expansion of the optimal control gives a $(2n + 1)$th-order approximation of the performance index. The difficulty with perturbation methods is that they are limited to a small class of systems, i.e. systems that are small perturbations of a linear system and that have analytic optimal cost and control. In addition, these methods are inherently tied to the convergence of a power series for which it is difficult if not impossible to estimate the region of convergence. Consequently, it is equally difficult to estimate the stability region of a control calculated from a truncated power series. For bilinear systems, however, it appears that the region of attraction can be estimated, as reported in Cebuhar and Costanza (1984). In Halm and Hamalainen (1975), the authors present a method that is similar to perturbation methods. The basic idea is to represent the integral of the solution (via Green's functions) as a basic linear operator and then invert the operator. The method has several advantages over perturbation methods. Namely, it is possible to estimate the region of convergence of the power series that makes up the final control; therefore, it is possible to estimate the stability region of a truncated control law. One of the main advantages to the method presented in this paper is that there is a clearly defined estimate of the stability region of the approximate control. In addition, the designer has explicit control over this region.

Another approach to approximating the HJB equation is to "regularize" the cost function so that an analytic expression for the control can be obtained. The basic idea is to consider a cost function of the form

$$J = \int_0^\infty [x^TQx + u^TRu + \mathcal{H}(x)] \, dt,$$

(1)

where $\mathcal{H}(x)$ is a function chosen so that the Hamilton-Jacobi-Bellman equation reduces to a form similar to the Riccati equation. This method produces a control that minimizes the integral (1) and is therefore suboptimal with respect to the cost function that is really of interest [i.e. (1) without the $\mathcal{H}$ term]. The approach results in solutions that stabilize the system but it is difficult to estimate how far the control deviates from the optimal since $\mathcal{H}$ must be carefully selected. An example of this approach are Ryan (1984), Tzafestas et al. (1984), Lu (1993) and Freeman and Kokotovic (1995). When the method presented in this paper is iterated, the optimal cost and control can be approximated arbitrarily closely when the order of approximation is sufficiently large (see Beard, 1995 for a discussion of this algorithm).

A feedback design method that has been popular in recent years is feedback linearization (Hunt et al., 1983a,b; Isidori, 1989; Nijmeijer and van der Schaft, 1990). The basic idea is to use feedback to cancel out the nonlinearities in the system. The resulting system is linear and so a feedback control can be designed using any linear synthesis method. There are numerous examples in the literature where LQR methods have been used to optimize the resulting linear system (Lee and Chen, 1983; Gao et al., 1992; Chapman et al., 1993, Wang et al., 1993, 1994; Marino, 1984). Feedback linearization has several disadvantages. First it is difficult to quantify the robustness of the control. Second, the control sometimes cancels out nonlinearities that enhance stability and performance (Freeman and Kokotovic, 1995). Third, to cancel the nonlinearities, the control effort can be unreasonably large. Finally, even if the resulting linear system is optimized, it is not possible to determine how close the original nonlinear system is to optimal.

An important recent development in optimal control is the theory of viscosity solutions. For the solution of the HJB equation to be defined in a classical sense, it must be differentiable over the domain of interest. This restriction excludes a large class of nonlinear systems. For example, the system $\dot{x} = xu, J = \int_0^\infty (x^2 + u^2) \, dt$ has an optimal cost given by $V(x) = |x|$: clearly, the HJB is not defined in a classical sense at $x = 0$. The theory of viscosity solutions allows a continuous function to be defined as the unique solution to Hamilton-Jacobi equations that do not admit continuous solutions in the classical sense. The viscosity method is a fairly general theoretical tool for dealing with existence and uniqueness issues of nonlinear partial differential equations. For an introduction to viscosity solutions see Crandall et al. (1992). An important result is that the viscosity solution of the HJB equation is identical to the value function of the associated (stochastic or deterministic) optimal control problem.
control problem. If the solution of the HJB equation is differentiable, then the viscosity solution and the classical solution are identical.

We should note that the algorithm obtained by iterating the method presented in this paper can also be applied to systems that do not admit continuous solutions. However, in its present formulation, it is restricted to systems which admit continuous stabilizing (but not necessarily optimal) controls. Using the viscosity method, it should be relatively straightforward to extend the results in this paper to systems that admit stabilizing (but not continuous stabilizing) controls.

The fact that the viscosity solution is equivalent to the value function has inspired work in numerically approximating viscosity solutions for HJ equations. Two broad categories of solutions have been reported: finite-difference solutions and finite-element solutions. Capuzzo Dolcetta (1983) introduces a discrete approximation to the continuous-time, deterministic optimal control problem. The basic idea is to approximate the system equations by an Euler formula with step size \( h \) and then to discretize the solution according to \( h \). It is shown that as \( h \to 0 \) the discrete solution \( V_n \) converges (locally uniformly) to the viscosity solution of the HJB equation. The optimal control problem can then be approximated by solving the discrete dynamic programming problem. The convergence rate of the scheme is studied in Capuzzo Dolcetta and Ishii (1984), where the approximate controls are shown to converge in a relaxed sense. Falcone and Ferretti (1994) show that convergence can be improved by using a Runge–Kutta formula instead of Euler’s method to approximate the dynamics of the system. Other convergence acceleration methods are discussed in Capuzzo Dolcetta and Falcone (1989). Using the finite-element method, González and Roffman (1985a) lists an algorithm for finite-time problems with continuous and impulse controls for stationary systems. The non-stationary case is studied in González and Roffman (1985b). A similar approach is considered for infinite-time horizon problems in Falcone (1987). The method is shown to converge to the viscosity solution and error estimates for the convergence of the algorithm are also given. Variations on these two approaches can be found in Kushner (1990) and Fleming and Soner (1993, Chap. 9).

The main disadvantage of both finite-element and finite-difference methods is that they require a discretization of the state space which is particularly problematic in two ways. First, the computational burden scales exponentially in the dimension of the state space, i.e. the methods are \( O(M^n) \) where \( M \) is the number of partitions of each variable, i.e. Bellman’s “curse of dimensionality”. Second, an enormous amount of data must be stored and then recalled in real time to produce the feedback control.

The major motivation of using Galerkin’s spectral method in this paper was to avoid discretizing the time and space variable. Of course, the “curse of dimensionality” still exists but it shows up as weighted averages of the dynamics over a compact set \( \Omega \). The advantage is that this provides numerous options for dealing with the dimensionality problem. For example, if the system equation are separable, and \( \Omega \) is rectangular, then the multi-dimensional integrals reduce to iterated one-dimensional integrals which can be computed numerically or symbolically. Another major advantage of our method is that the feedback controls are determined by a small number of coefficients and can be implemented in a variety of ways. In fact, the computer can be completely taken out of the loop by implementing each of the basis functions in hardware, and the coefficients with amplifiers.

In the literature there are many other approaches to approximating optimal control laws (cf. Baumann and Rugh, 1986; Cloutier et al., 1996; Goh, 1993; Johansson, 1990). In approximating optimal controls there are three things that we would like an approximate control to satisfy:

1. We would like the approximate control to be in an explicit feedback form that is easy to implement.
2. We would like the approximation to converge uniformly to the optimal control, if it exists, as the complexity of the approximation is increased.
3. We would like the control to remain stable when the approximation is truncated at a finite degree of complexity.

To our knowledge, the technique presented in this paper, coupled with the successive approximation method as outlined in Beard (1995) is the only method that accomplishes all three of these objectives.

2. PROBLEM STATEMENT

In this paper we restrict ourselves to the state-feedback control problem for the class of nonlinear time-invariant systems described by ordinary differential equations that are affine in the control:

\[
\dot{x} = f(x) + g(x)u(x),
\]  
\[
\text{where } x \in \Omega \subset \mathbb{R}^n, \quad f: \Omega \to \mathbb{R}^n, \quad g: \Omega \to \mathbb{R}^{n \times m} \text{ and } u: \Omega \to \mathbb{R}^m \text{ is the control. To ensure that the control problem is well posed we assume that } f \text{ and } g \text{ are Lipschitz continuous on a set } \Omega \text{ that contains the origin as an interior point. We also assume that}
\]
Define \( \phi(t; x_0, u) \) to be the solution at time \( t \) to equation (2) with initial conditions \( x_0 \) and control \( u \). To simplify the notation we write \( \phi(t) \) for \( \phi(t; x_0, u) \) when \( x_0 \) and \( u \) are understood. For the system \( \text{[equation (2)]} \), we say that \( u \) is a stabilizing control on \( \Omega \) if the resulting closed-loop system is asymptotically stable in the sense of Lyapunov (Khalil, 1992) for all initial conditions in \( \Omega \).

To quantify the performance of the control we use the standard integral performance measure

\[
J(x_0, u) = \int_{0}^{\infty} l(\phi(t)) + \|u(\phi(t))\|_2^2 \, dt, \tag{3}
\]

where \( l: \Omega \to \mathbb{R} \) is a positive-definite function on \( \Omega \) chosen such that the system is zero state observable through \( l \), \( R \in \mathbb{R}^{n \times n} \) is a symmetric, positive-definite matrix, \( \|u\|_2^2 = u^T R u \) and \( x_0 \in \Omega \subset \mathbb{R}^n \). \( l \) is called the state penalty function and \( \|u\|_2^2 \) is the control penalty function. Typically, \( l \) is a quadratic weighting of the states, i.e. \( l = x^T Q x \) where \( Q \) is a positive-definite matrix.

For equation (3) to give any indication of the performance of the system, the integral must converge. Unfortunately, stability of \( \dot{x} = f + gu \) is not sufficient for the integral to be finite. For example, the solution to the system

\[
\dot{x} = xu, \quad u = -|x|
\]

\( \phi(t) = \frac{x_0}{1 + |x_0| t} \)

The control \( u \) asymptotically stabilizes the system but if \( l(x) \triangleq |x| \) then

\[
\int_{0}^{\infty} l(\phi(t)) \, dt = \int_{0}^{\infty} \frac{|x_0|}{1 + |x_0| t} \, dt = |x_0| \int_{0}^{\infty} \frac{du}{u} = \infty.
\]

However, if \( l(x) \triangleq |x|^\alpha \) with \( \alpha > 1 \) then the integral is finite.

This necessitates the restriction of stabilizing controls to those controls that render the cost function \( \text{[equation (3)]} \) finite with respect to a certain penalty on the states.

**Definition 1 (Admissible Control).** Given the system \( (f, g) \), a control \( u: \mathbb{R}^n \to \mathbb{R}^n \) is defined to be admissible with respect to the state penalty function \( l \) on \( \Omega \), written \( u \in \mathcal{A}(\Omega) \), if

- \( u \) is continuous on \( \Omega \),
- \( u(0) = 0 \),
- \( u \) stabilizes \( (f, g) \) on \( \Omega \),
- \( \int_{0}^{\infty} l(\phi(t; x, u)) + \|u(\phi(t; x, u))\|_2^2 \, dt < \infty, \forall x \in \Omega \).

When \( u \in \mathcal{A}(\Omega) \), a Lyapunov function for the systems on \( \Omega \) is given by

\[
V(x) = \int_{0}^{\infty} [l(\phi(t; x, u)) + \|u(\phi(t; x, u))\|_2^2] \, dt.
\]

We would like to know when asymptotic stability implies that there exist an \( l: \Omega \to \mathbb{R} \) such that the integral \( \text{[equation (3)]} \) is finite. The limitation is the quadratic penalty on \( u \); in short, we are restricted to systems whose control function has finite energy.

**Lemma 2.** If \( u \) is continuous on \( \Omega \), \( u(0) = 0 \), and \( u \) stabilizes the system \( (f, g) \) on \( \Omega \), then there exists a continuously differentiable, positive-definite state penalty function \( l: \Omega \to \mathbb{R} \) such that \( u \in \mathcal{A}(\Omega) \) if and only if \( u \) has finite energy for all \( x_0 \in \Omega \), i.e.

\[
\int_{0}^{\infty} \|u(\phi(t; x_0, u))\|^2 \, dt < \infty.
\]

**Remark 3.** In general it is difficult to derive specific conditions under which the control can be made to have finite energy. However, an important case when this is always true is when the linearization of \( (f, g) \) at \( x = 0 \), i.e.

\[
\left( \frac{\partial f}{\partial x}(0), g(0) \right)
\]

is stabilizable. In this case the origin can be made exponentially stable by an appropriate linear state feedback. Therefore, there exists a nonlinear state feedback, \( u \), such that the real parts of the eigenvalues of \( \partial f + g u \) are all negative, i.e. the origin is exponentially stable.

**Remark 4.** While not all systems can be stabilized by continuous state feedback, there are large classes of systems that can be. For example, feedback linearizable systems can be stabilized via continuous controls (Nijmeijer and van der Schaft, 1990, Chap. 6). In addition, If the linearization of the system at the equilibrium is controllable, then we can find a set \( \Omega \) such that there exists a continuous state feedback that stabilizes the system. See Aeyels (1985, 1986) and Battilotti (1996) for additional classes of systems that can be stabilized by continuous state feedback.

In the above discussion, the specification of the set \( \Omega \) has been somewhat arbitrary. To be precise, \( \Omega \) can be made as large as the domain of attraction of the system under the closed-loop control of \( u \).

**Lemma 5.** Given a system \( (f, g) \). If \( u \in \mathcal{A}(\Omega) \) and the region of stability of the system \( \dot{x} = f + gu \) is \( \mathcal{Y} \subset \mathbb{R}^n \) where \( \mathcal{Y} \supset \Omega \), then \( u \in \mathcal{A}(\mathcal{Y}) \).

**Proof.** Since \( u \) is asymptotically stabilizing on \( \Omega \), there exists a \( t' < \infty \) such that

\[
t > t' \Rightarrow \{ y : y = \phi(t; x, u), x \in \mathcal{Y} \} \subseteq \Omega.
\]
Then \( \forall x \in \Gamma \)
\[
\int_0^\infty [l(\varphi(t); x) + \|u(\varphi(t); x)\|^2] \, dt
\leq \int_0^\infty [l(\varphi(t); x) + \|u(\varphi(t); x)\|^2] \, dt
+ \int_{t'}^\infty [l(\varphi(t); \varphi(t')) + \|u(\varphi(t); \varphi(t'))\|^2] \, dt.
\]

The first integral is finite since it is over a finite time period and the second integral is finite since \( \varphi(t'; x, u) \in \Omega \) and \( u \in \mathcal{A}_f(\Omega) \).

In general, it is difficult to find the largest stability region, \( \Gamma \), corresponding to \( u \) (Genesio et al., 1985; Loparo and Blankenship, 1978). However, it is usually possible to find a region \( \Omega \subset \Gamma \), or to verify (by Lyapunov methods) that a subset \( \Omega \) is contained in \( \Gamma \). Since our method will require some set \( \Omega \) over which \( u \) is stabilizing, but does not require the entire stability region \( \Gamma \), we will retain the notation \( u \in \mathcal{A}_f(\Omega) \).

We will assume throughout the paper that system [equation (2)] is controllable on \( \Omega \), in that for an appropriate choice of \( l \), there exists at least one admissible control, \( u \in \mathcal{A}_f(\Omega) \).

The standard optimal control problem is to find a control to minimize the cost function given in equation (3). For the problem to be well posed mathematically, a unique optimal control must exist. This requirement places limitations on the applicability of optimal control theory. In addition, the optimal control is very difficult to find, while many controls close to optimal may be much easier to compute. In this section we generalize optimal control by considering the problem of improving the performance of an arbitrary admissible control.

Given an arbitrary control \( u(x) \in \mathcal{A}_f(\Omega) \), the performance of the control at \( x \in \Omega \) is given by the formula
\[
V(x) = \sum_{t=0}^{\infty} \left[ [l(\varphi(t)) + \|u(\varphi(t))\|^2] \right] \, dt.
\]  
(4)

However, this expression depends on the solution of the system \( \dot{x} = f + gu \) which is generally not available. To obtain an expression that is independent of the solution of the system, we differentiate \( V \) along the system trajectories to obtain
\[
\frac{\partial V}{\partial x} \cdot (f + gu) + l + \|u\|^2 = 0.
\]

The boundary condition is easily seen to be \( V(0) = 0 \). The equation is valid over \( \Omega \subset \mathbb{R}^n \). This partial differential equation is an incremental expression of the cost of an arbitrary control \( u \). If we can solve this equation then we have a compact expression for equation (4) that does not depend on the solution \( \varphi(t; x, u) \). This equation will be extremely important throughout the paper and is termed the generalized Hamilton–Jacobi–Bellman equation.

**Definition 6 (GHJB equation).** Given an admissible control \( u \in \mathcal{A}_f(\Omega) \), the function \( V: \Omega \to \mathbb{R} \) satisfies the *generalized Hamilton–Jacobi–Bellman equation*, written \( \text{GHJB}(V; u) = 0, \) if
\[
\frac{\partial V}{\partial x} \cdot (f + gu) + l + \|u\|^2 = 0, \quad V(0) = 0.
\]

(5)

To improve the performance of an arbitrary control \( u \in \mathcal{A}_f(\Omega) \) we fix \( V \) and minimize the pre-Hamiltonian, i.e.,
\[
\hat{u}(x) = \arg \min_{u \in \mathcal{A}_f(\Omega)} \left\{ \frac{\partial V}{\partial x} (f + gu) + l + \|u\|^2 \right\}
= \frac{1}{2} R^{-1} g^{T} \frac{\partial V}{\partial x}.
\]

(6)

The cost of \( \hat{u} \) is given by the solution of the equation \( \text{GHJB}(\hat{V}; \hat{u}) = 0 \). In Saridis and Lee (1979), Saridis and Wang (1994), Vaisbord (1963), Mil'shtein (1964) and Leake and Liu (1967) it is shown that \( \hat{V}(x) \leq V(x) \) for each \( x \in \Omega \) and that when the process is iterated, the value functions converges uniformly (on compact \( \Omega \)) to the solution of the Hamilton–Jacobi–Bellman equation
\[
\frac{\partial V}{\partial x} f + l - \frac{1}{4} \frac{\partial V}{\partial x} g R^{-1} g^{T} \frac{\partial V}{\partial x} = 0.
\]

The GHJB equation answers three fundamental questions. First, its solution is the performance of any admissible control. Second, its solution allows us to find a control law that improves the performance of the original control. Finally, by iterating the process we converge uniformly to the solution of the HJB equation. Since the GHJB equation is linear, it is theoretically easier to solve than the, nonlinear, HJB equation. Unfortunately however, there is no general closed-form solution to this equation. In Section 3 we will show how to approximate the GHJB equation and give sufficient conditions that guarantee that the approximation converges to the actual solution. In Section 4 the method will be used to compute the approximate cost of a feedback linearizing control and to compute a control that improves its performance.

### 3. THE MAIN RESULT

In this section we use Galerkin's spectral method to approximate the solution to the GHJB equation.
We also show conditions under which this method converges and show that the resulting feedback control stabilizes the system on $\Omega$ for a high enough order of approximation.

There is a vast literature on the use of Galerkin’s method to solve differential equations. Classical references include Kantorovich and Krylov (1958), Mikhlin (1964), Mikhlin and Smolitskii (1967) and Petryshyn (1965). A survey of the literature prior to 1972 is given in Finlayson (1972). A modern treatment is given in Zeidler (1990a) for linear operators and Zeidler (1990b) for nonlinear operators. These references derive a number of sufficient conditions that guarantee that Galerkin’s method converges as the number of basis functions increase to infinity. To apply these results, a linear operator must be bounded or symmetric or positive bounded below. In Beard (1995) it is shown that the linear operator associated with the GHJB equation does not satisfy any of these requirements. Therefore, a new convergence proof for the GHJB equation is necessary.

We use Galerkin’s method to derive an approximate solution to the GHJB equation. To apply Galerkin’s method it is first necessary to place the solution to the differential equation in a Hilbert space. To do so we restrict attention to a compact subset $\Omega$ of the stability region of a known stabilizing control $u$. When the solutions to the GHJB equation are restricted to this set they exist in the Hilbert space $L^2(\Omega)$. We assume that

C1: We can select a set (not necessarily linearly independent), $\Phi = \{\phi_j(x)\}_{j=1}^\infty$, where $\phi_j: \Omega \rightarrow \mathbb{R}$ and $\phi_j(0) = 0$ (to satisfy the boundary condition), such that $V \in L^2(\Omega)$.

This implies that there exists coefficients $b_j$ such that

$$
V(x) - \sum_{j=1}^\infty b_j \phi_j(x) \rightarrow 0.
$$

Additional conditions on the set $\{\phi_j(x)\}_{j=1}^\infty$ will be developed in subsequent lemmas.

We seek an approximate solution, $V_N$, to the equation $GHJB(V; u) = 0$ by letting

$$
V_N(x) = \sum_{j=1}^N c_j \phi_j(x). \tag{7}
$$

Substituting this expression into the GHJB equation results in an error

$$
\text{error}_N(x) = GHJB\left(\sum_{j=1}^N c_j \phi_j(x); u\right). \tag{8}
$$

The coefficients $c_j$ are determined by setting the projection of the error, (8), on the finite basis $\{\phi_j\}_1^N$ to zero $\forall x \in \Omega$:

$$
\langle GHJB\left(\sum_{j=1}^N c_j \phi_j(x), u\right), \phi_n \rangle = 0, \tag{9}
$$

$n = 1, \ldots, N$, where the inner product is defined as

$$
\langle f, g \rangle = \int_{\Omega} f(x)g(x)\,dx. \tag{10}
$$

Using equation (5), this expression reduces to the following $N$ equations in $N$ unknowns:

$$
\sum_{j=1}^N c_j \left(\frac{\partial^2}{\partial x^2} + (f + gu), \phi_n \right) + \left(1 + \|u\|_K^2, \phi_n \right) = 0,
$$

$n = 1, \ldots, N$.

To simplify the notation, define

$$
\Phi_N(x) \triangleq (\phi_1(x), \ldots, \phi_N(x))^T, \tag{11}
$$

and let $\nabla \Phi_N$ be the Jacobian of $\Phi_N$. If $\eta: \mathbb{R}^N \rightarrow \mathbb{R}$ is a real-valued function then we define the notation

$$
\langle \eta, \Phi_N \rangle_v \triangleq \langle \eta_1, \phi_1 \rangle _{v} \cdots \langle \eta_N, \phi_N \rangle _{v}. \tag{12}
$$

If $\eta: \mathbb{R}^N \rightarrow \mathbb{R}^N$ is a vector-valued function then we define the notation

$$
\langle \eta, \Phi_N \rangle_v \triangleq \begin{pmatrix}
\langle \eta_1, \phi_1 \rangle_v & \cdots & \langle \eta_N, \phi_1 \rangle_v \\
\vdots & \ddots & \vdots \\
\langle \eta_1, \phi_N \rangle_v & \cdots & \langle \eta_N, \phi_N \rangle_v
\end{pmatrix}. \tag{13}
$$

The key to the notation is that the jth row corresponds to integration weighted by $\phi_j$.

Using this notation we can write the Galerkin projection of the GHJB equation in the compact form

$$
\langle GHJB(V; u), \Phi_N \rangle_v = 0. \tag{14}
$$

We will also use bold face letters to denote the coefficients in the Galerkin approximation method, i.e.,

$$
c_N \triangleq (c_1, \ldots, c_N)^T. \tag{15}
$$

From equation (12), the coefficients are given by the expression

$$
\langle \nabla \Phi_N(f + gu), \Phi_N \rangle_v c_N = -\langle 1 + \|u\|_K^2, \Phi_N \rangle_v. \tag{16}
$$

To approximate the improved control $\hat{u}$ in equation (6), we let $\hat{u}_N$ depend on $V_N$ instead of $V$:

$$
\hat{u}_N(x) = -\frac{1}{2} R^{-1} g^T(x) \frac{\partial V_N}{\partial x}(x) \tag{17}
$$

$$
= -\frac{1}{2} R^{-1} g^T(x) \nabla \Phi_N c_N. \tag{18}
$$

Remark 7. It is, of course, possible to apply Galerkin’s method to the HJB equation directly. Assuming that the optimal cost function exists in $L^2(\Omega)$, we substitute the approximation $V_N^* \triangleq c_N^* \Phi_N$ into the HJB equation and set the projection of the error to zero

$$
\langle HJB(V), \Phi_N \rangle_v = 0. \tag{19}
$$
After some algebra we obtain the nonlinear algebraic equation
\[
\langle \nabla \Phi_{\mathcal{K}} f, \Phi_{\mathcal{K}} \rangle_{\mathcal{K}} + \frac{1}{4} \left( \sum_{i=1}^{N} \Phi_{\mathcal{K}} \right) \mathcal{K} = 0,
\]
where
\[
\mathcal{K} = \left( \nabla \Phi_{\mathcal{K}} gR^{-1} g^T g \frac{\partial \Phi_{\mathcal{K}}}{\partial \mathcal{K}} \Phi_{\mathcal{K}} \right)_m.
\]

The difficulty is that equation (16) is a nonlinear algebraic expression with multiple solutions, one of which corresponds to a stabilizing control law. [The situation is similar to the Riccati equation which may have several solutions but only one corresponds to a stabilizing control (Bittanti et al., 1991).] We are now faced with the question of how to solve this equation and how to guarantee that the solution produces a stabilizing control. The answer is given by iterating the procedure outlined in this paper to obtain a successive approximation algorithm with the initial condition $u$. A complete analysis of this algorithm is given in Beard (1995) and appears in Beard et al. (1998).

We will now develop conditions that guarantee that $V_{\mathcal{K}} \to V$ as $N \to \infty$. We will also show that for $N$ sufficiently large (but finite) the approximate control $\hat{u}_n$ stabilizes the system and is robust in the same sense as the optimal control. In the subsequent discussion, it is important to note that for $N$ sufficiently large, the stability region of $\hat{u}_n$ contains $\Omega$, the closed and bounded subset of the stability region of $u$. This is important since $\Omega$ is chosen by the designer and therefore gives a well-defined estimate of the stability region of $\hat{u}_n$.

We begin by stating three known results regarding the GHJB equation.

**Lemma 8.** If

- C2: $\Omega$ is compact,
- C3: $f$ and $g$ are Lipschitz continuous on $\Omega$ and $f(0) = 0$,
- C4: $I$ is a positive definite, monotonically increasing function on $\Omega$ and $R$ is a symmetric positive-definite matrix,
- C5: $u \in \mathcal{A}(\Omega),$

then:

- On $\Omega$, there exists a unique continuously differentiable solution $V(x)$ to the equation GHJB($V$; $u$) = 0 with boundary conditions $V(0) = 0$,
- $V(x)$ is a Lyapunov function for the system $(f, g, u)$ on $\Omega$,
- GHJB($V$; $u$) = 0 $\iff$ $V(x) = J(x)$, where $J(x)$ is the performance index given in equation (3).

**Proof.** See Saridis and Lee (1979).

The next lemma shows that if an updated control is chosen according to equation (6), that the new control is admissible and that it improves the performance of the system.

**Lemma 9.** Let the conditions of Lemma 8 hold. If $V$ satisfies the equation $\text{GHJB}(V; u) = 0$, and
\[
\hat{u}(x) = -\frac{1}{2} R^{-1} g^T g \frac{\partial V}{\partial x}(x),
\]
then $\hat{u} \in \mathcal{A}(\Omega)$. If $\hat{V}$ satisfies $\text{GHJB}((\hat{V}, \hat{u}) = 0$ with boundary condition $\hat{V}(0) = 0$, then $\hat{V}(x; \hat{u}) \leq V(x; u)$, for all $x \in \Omega$.

**Proof.** See Saridis and Lee (1979) and Beard (1995).

It has been shown in Glad (1985, 1987) and Tsitsiklis and Athans (1984) that the optimal control $u^*$ is robust in the sense that it has infinite gain margin and 50% gain reduction margin. A similar result has been shown for the control $\hat{u}$ obtained from the GHJB equation via equation (17).

**Lemma 10.** Let $\hat{u} \in \mathcal{A}(\Omega)$ be a control obtained from Lemma 9 and let the gain perturbation $D : \mathbb{R}^m \to \mathbb{R}^m$ satisfy
\[
z^T RD(z) \geq \frac{1 + \alpha}{2} ||z||^2_\mathcal{K}, \quad \alpha > 0;
\]
then $\hat{x} = f + gD(\hat{u})$ is asymptotically stable on $\Omega$.

**Proof.** See Saridis and Balaram (1986) and Beard (1995).

For one-dimensional systems, the situation is depicted geometrically in Fig. 1; the system remains stable as long as the perturbed control is bounded below by $\frac{1}{2} \hat{u}$.

The next two lemmas will be needed at several points in subsequent proofs.

**Lemma 11.** If the set $\{\phi_j\}_{j=1}^N$ is linearly independent and $u \in \mathcal{A}(\Omega)$ then the set
\[
\left\{ \frac{\partial \phi_j}{\partial x} (f + gu) \right\}_{j=1}^N
\]
is also linearly independent.

**Proof.** If the vector field $f + gu$ is asymptotically stable then along the trajectories $\phi(t; x_0, u)$, $x_0 \in \Omega$, we have that
\[
\phi(x_0) = -\int_0^\omega \frac{d\phi}{dt}(\phi(t; x_0, u)) dt = -\int_0^\omega \frac{\partial \phi}{\partial x}(f + gu)(\phi(t; x_0, u)) dt.
\]
Now suppose that the lemma is not true. Then there exists a nonzero \( c \in \mathbb{R}^N \) such that
\[
c^T \nabla \Phi(x(f + gu)) = 0.
\]
This implies that for all \( x_0 \in \Omega \)
\[
\int_0^\infty c^T \nabla \Phi(x_0)(\phi(t; x_0, u)) \, dt = 0,
\]
\[
\Rightarrow c^T \int_0^\infty \nabla \Phi(x_0)(\phi(t; x_0, u)) \, dt = 0,
\]
\[
\Rightarrow c^T \Phi(x_0) = 0,
\]
which contradicts the linear independence of \( \{ \phi_j \}^N_{j=1} \).

**Corollary 12.** Suppose that for all \( j \in [1, N] \), \( \phi_j \) is continuously differentiable and \( 3x_0 \) such that \( (\partial \phi_j / \partial x)(x_0) \neq 0 \). If the set \( \{ \phi_j \}^N_{j=1} \) is linearly independent then so is the set \( \{ \partial \phi_j / \partial x \}^N_{j=1} \).

**Proof.** Suppose not, then \( \exists \varepsilon \in \mathbb{R}^N \) such that
\[
c^T \nabla \Phi_N = 0 \Rightarrow c^T \nabla \Phi_N(\phi_j = 0)
\]
which contradicts Lemma 11.

**Definition 13.** Given a countable set of functions \( \Phi = \{ \phi_j \}^N_{j=1} \) where \( \phi_j : \Omega \rightarrow \mathbb{R}^n \), define \( \delta(\phi, \Omega) \) to be the set of all linear combinations of elements in \( \Phi \) that converge pointwise at each \( x \in \Omega \).

**Lemma 14.** If the set \( \{ \phi_j \}^N_{j=1} \) is linearly independent, \( u \in \delta(\Omega) \), and \( \sum_{j=1}^N (\partial \phi_j / \partial x) (f + gu) \in P(\Phi, \Omega) \), then for all \( N \)
\[
\text{rank}(\langle \nabla \Phi_N(f + gu), \Phi_N \rangle) = N.
\]

**Proof.** Define \( \Phi = (\phi_1, \phi_2, \ldots)^T \) and \( d_j = (d_{j, 1}, d_{j, 2}, \ldots)^T \). Then by the hypothesis
\[
\frac{\partial \phi_j}{\partial x}(f + gu) = \sum_{k=1}^N d_{j, k} \phi_k = d_j^T \Phi.
\]
So
\[
\langle \nabla \Phi_N(f + gu), \Phi_N \rangle = \langle d_1^T \Phi, \phi_1 \rangle \ldots \langle d_N^T \Phi, \phi_N \rangle
\]
\[
= \langle \Phi, \Phi_N \rangle d_1 \ldots \langle \Phi, \Phi_N \rangle d_N
\]
\[
= \langle \Phi, \Phi_N \rangle d \cdot \langle \Phi, \Phi_N \rangle = \langle \Phi, \Phi_N \rangle d_N,
\]
where \( D \triangleq [d_1, \ldots, d_N] \). Therefore, \( \text{rank}(\langle \nabla \Phi_N(f + gu), \Phi_N \rangle) = \text{rank}(\langle \Phi, \Phi_N \rangle D) \), where \( \langle \Phi, \Phi_N \rangle \) has rank \( N \) since the set \( \{ \phi_j \} \) are linearly independent. To prove the result we need to show that \( D \) has rank \( N \).

**Lemma 15.** Suppose that the set \( \{ \phi_j \}^N_{j=1} \) is linearly independent but the set \( \{ \phi_j \}^{N+1}_{j=1} \) is linearly dependent. Let \( V_N = c_N \Phi_N \) and \( W_{N+1} = b_{N+1} \Phi_{N+1} \) satisfy the equations
\[
\langle \text{GHJB}(V_N, u), \Phi_N \rangle = 0
\]
and
\[
\langle \text{GHJB}(W_{N+1}, \Phi_{N+1}) \rangle = 0
\]
respectively, then \( V_N \equiv W_{N+1} \).

**Proof.** From the hypothesis we know that there exist a nonzero \( c_N \in \mathbb{R}^N \), such that \( \phi_{N+1} = \beta_N \Phi_N \) so
\[
W_{N+1} = b_{N+1} \Phi_N + b_{N+1} \phi_{N+1}
\]
\[
= b_{N+1} \Phi_N + b_{N+1} \beta_N \Phi_N = (b_n + b_{N+1} \beta_N)^T \Phi_N.
\]
Therefore, \( V_N \equiv W_{N+1} \iff c_N = b_n + b_{N+1} \beta_N \). From the hypothesis we know that \( c_n \) satisfies
\[
\langle \nabla \Phi_N(f + gu), \Phi_N \rangle c_N + \langle l + ||u||^2 \Phi_N \rangle = 0.
\]
We also know the \( b_{N+1} \) satisfies
\[
\langle \nabla \Phi_{N+1}(f + gu), \Phi_{N+1} \rangle b_{N+1} + \langle l + ||u||^2 \Phi_{N+1} \rangle = 0.
\]
This implies that
\[
\langle \nabla \Phi_N(f + gu), \Phi_{N+1} \rangle b_{N+1} + \langle l + ||u||^2 \Phi_{N+1} \rangle = 0.
\]

Lemma 11 shows that the set
\[
\left\{ \frac{\partial \phi_j}{\partial x}(f + gu) \right\}_{j=1}^N = \left\{ d_j^T \Phi \right\}_{j=1}^N
\]
is linearly independent, therefore the Gram matrix of \( N \) of these vectors has rank equal to \( N \), i.e.
\[
\text{rank}\left( \begin{pmatrix} \langle d_1^T \Phi, d_1^T \Phi \rangle & \ldots & \langle d_1^T \Phi, d_N^T \Phi \rangle \\ \vdots & \ddots & \vdots \\ \langle d_N^T \Phi, d_1^T \Phi \rangle & \ldots & \langle d_N^T \Phi, d_N^T \Phi \rangle \end{pmatrix} \right) = \text{rank}(D^T \Phi_{m} D) = N.
\]
Since the set \( \{ \phi_j \}_{j=1}^N \) is linearly independent, \( \langle \Phi, \Phi \rangle m \) has full rank, which implies that \( \text{rank}(D) = N \).

The analysis of Galerkin's method is greatly simplified by assuming that the basis functions \( \{ \phi_j \}_{j=1}^n \) are orthonormal. However, from a practical point of view, orthonormalizing a set of functions can require extensive computational effort: therefore, we do not want to require orthonormality. The next two lemmas show that we can analyze Galerkin's method using orthonormal basis functions, while not requiring that they be used in practice.
After some algebraic manipulation we obtain
\[
\left\langle \frac{\partial \Phi_{n+1}}{\partial x}(f + gu), \Phi_n \right\rangle_v = \left\langle \nabla \Phi_{n+1}(f + gu), \Phi_n \right\rangle_m \Phi_n.
\]
Therefore, \(c_n\) and \(b_n + b_{n+1}\) both satisfy the linear equation
\[
\left\langle \nabla \Phi_{n+1}(f + gu), \Phi_n \right\rangle_m \Phi_n = \langle l - \|u\|^2, \Phi_n \rangle_v
\]
and are therefore equivalent since \(\langle \nabla \Phi(f + gu), \Phi_n \rangle_m\) is invertible by Lemma 14.

**Lemma 16.** Given a set of linearly independent functions \(\{\phi_j\}_1^N\). Suppose that these functions are orthonormalized to form the set \(\{\bar{\phi}_j\}_1^N\); then there exists constants \(\beta_{ij}\) such that
\[
\bar{\phi}_1 = \beta_{11} \phi_1,
\bar{\phi}_2 = \beta_{21} \phi_1 + \beta_{22} \phi_2,
\vdots
\bar{\phi}_N = \beta_{N1} \phi_1 + \cdots + \beta_{NN} \phi_N.
\]
(18)
Let \(V_N = c_N \Phi_N\) solve
\[
\left\langle GHJB(V_N; u), \Phi_N \right\rangle_v = 0,
\]
and let \(W_N = c_N \Phi_N\) solve
\[
\left\langle GHJB(W_N; u), \Phi_N \right\rangle_v = 0;
\]
then \(V_N \equiv W_N\).

**Proof.** First note that we can write \(\bar{\phi}_N = B_N \Phi_N\) where \((B_N)_{ij} = \beta_{ij}\), \(W_N = c_N \Phi_N = c_N B_N \Phi_N\), so \(c_N = B_N^{\dagger} c_N \Rightarrow V_N \equiv W_N\). But since \(B_N\) is lower diagonal and invertible, we have that
\[
\left\langle GHJB(W_N; u), \Phi_N \right\rangle_v = 0,
\]
\[
\Rightarrow B_N \left\langle GHJB(W_N; u), \Phi_N \right\rangle_v = 0,
\]
\[
\Rightarrow \left\langle GHJB(W_N; u), \Phi_N \right\rangle_v = 0.
\]
So \(B_N^{\dagger} c_N\) satisfies the same linear equation as \(c_N\). The lemma is proved since the invertibility (by Lemma 14) of \(\langle \nabla \Phi(f + gu), \Phi_n \rangle_m\) implies that the equations solution is unique.

Throughout the rest of this paper, we will use the following notation: the set \(\{\phi_j\}_1^N\) will be the original basis functions. The set \(\{\bar{\phi}_j\}_1^N\) is obtained by first removing linearly dependent functions and then orthonormalizing the set \(\{\phi_j\}_1^N\) according to equation (18). The "tilde" will indicate orthonormalization. Similar notation will be used for the coefficients weighting the vector. Hence, we will write
\[
V = \sum_{j=1}^N c_j \phi_j = \sum_{j=1}^N \bar{c}_j \bar{\phi}_j.
\]
where \(c_j\) and \(\bar{c}_j\) are related through the coefficients \(\beta_{ij}\) defined above. Since orthonormality does not affect convergence, we will use \(\phi_j\) when orthonormality is not used in the argument and \(\bar{\phi}_j\) when it is.

The orthonormality of the set \(\{\bar{\phi}_j\}_1^N\) on \(\Omega\) implies that if a function \(\psi(x) \in \mathcal{H}(\Omega)\) and \(\psi \in L^2(\Omega)\) then
\[
\psi(x) = \sum_{j=1}^\infty \langle \psi, \bar{\phi}_j \rangle \bar{\phi}_j(x),
\]
where the series converges pointwise, i.e. for any \(\varepsilon > 0\) and \(x \in \Omega\), we can choose \(N(x)\) sufficiently large to guarantee that
\[
\left| \sum_{j=N(x) + 1}^\infty \langle \psi, \bar{\phi}_j \rangle \bar{\phi}_j(x) \right| < \varepsilon.
\]
(20)

To show stability of the approximate control obtained by Galerkin's method we will need conditions under which a pointwise convergent series converges uniformly.

The next lemma, found in Apostol (1974), Exercise 9.8, states necessary and sufficient conditions for pointwise convergence of a series to imply uniform convergence on a compact set.

**Definition 17.** Let \(\Psi(\Omega)\) be the set of infinite series, \(\sum_{j=1}^\infty c_j \phi_j(x)\), that converge pointwise \(\Omega\), such that \(\forall k \geq 1, x \in \Omega\), and \(\forall \varepsilon > 0\), there exists \(\rho > 0\) and \(m > 0\) such that \(\forall x \in \Omega\), the conditions \(n > m\) and \(\sum_{j=m+1}^\infty c_j \phi_j(x) < \rho\) imply that
\[
\sum_{j=m+1}^\infty c_j \phi_j(x) < \varepsilon.
\]
(20)

**Remark 18.** This condition implies that if the tail of a sequence at some point \(x \in \Omega\) is small, then after removing \(n > m\) terms, it is still small, where \(m\) is a uniform number for all \(x \in \Omega\). For example, if a series is monotonically decreasing on \(\Omega\), then \(m = 1\) and \(\varepsilon = \varepsilon\) and so \(\sum_{j=m+1}^\infty c_j \phi_j(x) \in \Psi(\Omega)\).

**Lemma 19.** If \(\Omega \subset \mathbb{R}^n\) is a compact set and \(W(x) = \sum_{j=1}^\infty c_j \phi_j(x)\), \(\forall x \in \Omega\) and the basis functions \(\phi_j(x)\) are continuous on \(\Omega\), then \(\sum_{j=m+1}^\infty c_j \phi_j(x)\) converges to zero uniformly on \(\Omega\) if
(i) \(W(x)\) is continuous on \(\Omega\),
(ii) \(\sum_{j=1}^\infty c_j \phi_j(x) \in \Psi(\Omega)\).

**Proof.** See Apostol (1974, Exercise 9.8) and Beard (1995).

In the remainder of this section, we use the following notation: \(u \in \mathcal{A}(\Omega)\) is an arbitrary admissible control, \(V_N = \sum_{j=1}^N c_j \phi_j\) satisfies the algebraic equation \(\langle GHJB(V_N; u), \Phi_N \rangle_v = 0\), \(u_N = -\frac{1}{\beta^2} g^T \partial V_N / \partial x\), \(V = \sum_{j=1}^N b_j \phi_j\) satisfies the differential equation \(\text{GHJB}(V; u) = 0\), \(\dot{u} = -\frac{1}{\beta^2} g^T \partial V / \partial x\), \(\mathcal{P} = \sum_{j=1}^N b_j \phi_j\) satisfies the differential equation \(\text{GHJB}(\mathcal{P}; \dot{u}) = 0\), \(c_N = (c_1, \ldots, c_N)^T\),
\( \mathbf{b}_N = (b_1, \ldots, b_N)^T, \mathbf{b}_N = (\hat{b}_1, \ldots, \hat{b}_N)^T \). The key to the notation is that \( b \) and \( c \), respectively, denote the coefficients of the actual and approximate solutions to the GHJB equation. The "hat" notation is used to denote the updated control and value functions.

When \( u \) is admissible, the equation \( \langle \text{GHJB}(V_N; u) \Phi_N \rangle_v = 0 \) forces the error caused by approximating the actual solution \( V \), projected on the linear space spanned by \( \{ \Phi_j \}_1^N \) to be zero. Lemma 20 shows that the residual error tends to zero as \( N \to \infty \). We then use this result to show in Lemma 22 that the coefficients for the approximate \( (V_N) \) solution to the GHJB equation converges to the coefficients of the actual \( (V) \) solution. This implies that \( V_N \) converges to \( V \) in the \( L^2(\Omega) \) norm, which is shown in Corollary 23.

**Lemma 20.** If the hypothesis of Lemmas 8 and 14 are satisfied and

\[ C6: \partial \phi / \partial x \cdot (f + gu), \| u \|_\Omega, l, \] are continuous, in the space \( \mathcal{H}(\Phi, \Omega) \cap L^2(\Omega) \), then

\[ |\text{GHJB}(V_N; u)| \to 0 \]

pointwise on \( \Omega \) as \( N \to \infty \).

**Proof.** Condition C6 implies that \( \text{GHJB}(V_N; u) \in \mathcal{H}(\Phi, \Omega) \), so

\[ |\text{GHJB}(V_N; u)(x)| = \left| \sum_{j=1}^{\infty} \langle \text{GHJB}(V_N; u), \Phi_j \rangle \phi_j(x) \right| \]

\[ \leq \sum_{j=N+1}^{\infty} \left| \phi_j \right| \left( f + gu \right) \| \phi_j \| \phi_j(x) \]

\[ + \sum_{j=N+1}^{\infty} \left( l + \| u \|_\Omega \right) \| \phi_j \| \phi_j(x) \]

\[ \leq AB \]

\[ + \sum_{j=N+1}^{\infty} \left( l + \| u \|_\Omega \right) \| \phi_j \| \phi_j(x) \]

where

\[ A \triangleq \sup_{N=1, 2, \ldots} \| \epsilon_N \|_2, \]

\[ B \triangleq \sup_{k=1, 2, \ldots} \left| \sum_{j=N+1}^{\infty} \left( \phi_j (f + gu), \partial \phi_j / \partial x \right) \right|. \]

Since \( \partial \phi_j / \partial x (f + gu) \) and \( l + \| u \|_\Omega \) are in \( L^2(\Omega) \), equation (20) implies that \( B \) and the second term on the right-hand side can be made arbitrarily small by an appropriate choice of \( N \), which gives pointwise convergence on \( \Omega \) if \( \| \epsilon_N \|_2 \) is uniformly bounded for all \( N \).

To show this define

\[ A_N \triangleq \langle \nabla \Phi_N (f + gu), \Phi_N \rangle_v, \]

\[ b_N \triangleq -\langle l + \| u \|_\Omega, \Phi_N \rangle_v, \]

i.e. \( A_N \epsilon_N = b_N \), and let

\[ \sigma(A_N) = \min_{\| u \|_2 = 1, u \in \mathcal{H}} \| A_N u \|_2 = \min_{u \in \mathcal{H}(\Phi, \Omega)} \| A_N u \|_2 \]

be the minimum singular value of the matrix \( A_N \). Lemma 14 guarantees that \( A_N \) is nonsingular, therefore

\[ \| b_N \|_2^2 = \| A_N \epsilon_N \|_2^2 \geq \sigma(A_N) \| \epsilon_N \|_2^2 \Rightarrow \| \epsilon_N \|_2^2 \leq \frac{\| b_N \|_2^2}{\sigma(A_N)}. \]

To prove that \( \| \epsilon_N \|_2 \) is uniformly bounded we need to show that there exist constants \( M_1 \) and \( M_2 \) such that (1) \( \| b_N \|_2 \leq M_1 < \infty \) and (2) \( \sigma(A_N) \geq M_2 > 0 \) for all \( N = 1, 2, \ldots \).

To show item (1) we note that condition C6 implies that

\[ \| b_N \|_2^2 = \sum_{j=1}^{N} \langle \| f + gu \|_\Omega, \phi_j \rangle^2 \]

\[ \leq \sum_{j=1}^{\infty} \langle \| l + \| u \|_\Omega \|, \phi_j \rangle^2 \]

\[ = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \langle \| l + \| u \|_\Omega \|, \phi_j \rangle \langle \phi_j, \phi_k \rangle \]

\[ = \int_{\Omega} \left( \sum_{j=1}^{\infty} \langle \| l + \| u \|_\Omega \|, \phi_j \rangle \langle \phi_j, \phi_j \rangle \right) dx \]

\[ = \| l + \| u \|_\Omega \|_2 \| \Omega \leq M_1 < \infty. \]

To show item (2), define the operator \( A_w \) by

\[ (A_w u)_j = \sum_{k=1}^{\infty} \left( \int_{\Omega} \frac{\partial \phi_k^T (f + gu) \phi_j(x)}{\partial x} \right) u_k, \]

and the minimum singular value of \( A_w \) as

\[ \sigma(A_w) = \inf_{\| u \|_2 = 1, u \in \mathcal{H}} \| A_w u \|_2 \]

Item (2) follows by showing that for all \( N \)

\[ \sigma(A_N) \geq \sigma(A_w) \]

which will imply that

\[ \| b_N \|_2^2 \geq \sigma(A_N) \geq \sigma(A_w) \triangleq M_2 > 0. \]

To show that \( \sigma(A_N) \geq \sigma(A_w) \) for all \( N \), let \( \hat{u} = \arg \min_{\| u \|_2 = 1, u \in \mathcal{H}} \| A_N u \|_2 \), and define \( w \) as

\[ w_j = \begin{cases} \hat{u}_j, & j \leq [1, N], \\ 0, & j > N + 1. \end{cases} \]

Then \( \| w \|_2 \leq 1 \) and

\[ \sigma(A_w) = \inf_{\| u \|_2 = 1, u \in \mathcal{H}} \| A_w u \|_2 \]

\[ = \| A_N \hat{u} \|_2 = \min_{\| u \|_2 = 1, u \in \mathcal{H}} \| A_N u \|_2 = \sigma(A_N). \]
To show that \( g(A_{\infty}) > 0 \), notice that

\[
\| A_{\infty} c_{\infty} \|^2 = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \left( \frac{\partial \phi_j^x}{\partial x}(f + gu) \cdot \phi_j \right) c_k^2,
\]

and that

\[
\left( \frac{\partial V}{\partial x}(x) \cdot (f + gu) \right)^2 = \int_{\Omega} \left( \frac{\partial V}{\partial x}(x) \cdot (f + gu) \right)^2 \, dx
\]

\[
= \int_{\Omega} \left( \sum_{j=1}^{\infty} \frac{\partial V}{\partial x}(x) \cdot (f + gu) \cdot \phi_j \right)^2 \, dx
\]

\[
\leq \sum_{j=1}^{\infty} \left( \frac{\partial V}{\partial x}(x) \cdot (f + gu) \cdot \phi_j \right)^2 \int_{\Omega} |\phi_j|^2 \, dx
\]

\[
\leq \sum_{j=1}^{\infty} \left( \frac{\partial V}{\partial x}(x) \cdot (f + gu) \cdot \phi_j \right)^2 = \| A_{\infty} c_{\infty} \|^2.
\]

In addition, Bessel's inequality implies that

\[
\| c_{\infty} \|^2 = \sum_{j=1}^{\infty} |c_k|^2 \leq \int_{\Omega} |V(x)|^2 \, dx = \| V \|^2_{L^2},
\]

therefore

\[
\inf_{\| V \|^2_{L^2} \neq 0} \left\{ \left( \frac{\| V \|^2_{L^2}}{\| V \|^2_{L^2}} \right) \right\}
\]

\[
\leq \inf_{\| V \|^2_{L^2} \neq 0} \left\{ \left( \frac{\| A_{\infty} c_{\infty} \|^2}{\| c_{\infty} \|^2} \right) \right\}.
\]

We claim that

\[
\inf_{\| V \|^2_{L^2} \neq 0} \left\{ \left( \frac{\| V \|^2_{L^2}}{\| V \|^2_{L^2}} \right) \right\} > 0 \quad (21)
\]

iff the GHJB equation has a unique solution. To show the necessity of this claim assume that equation (21) holds and let \( V_1 \) solve \((\partial V_1^T/\partial x)(f + gu) = -l \cdot \| u \|^2_x\). Suppose that \( V_2 \neq V_1 \) is also a solution on \( \Omega \), then \((\partial V_2^T/\partial x)(f + gu) = -l \cdot \| u \|^2_x\), which implies that \((\partial V_1 - V_2)(f + gu) = 0\).

Therefore

\[
0 = \frac{\| V_1 - V_2 \|^2_{L^2(\Omega)}}{\| V_2 \|^2_{L^2(\Omega)}}
\]

\[
\geq \inf_{\| V \|^2_{L^2} \neq 0} \left\{ \left( \frac{\| V \|^2_{L^2}}{\| V \|^2_{L^2}} \right) \right\} > 0
\]

which is a contradiction.

To show the sufficiency let \( V \) be the unique solution to \((\partial V/\partial x)(f + gu) = -l \cdot \| u \|^2_x\). Suppose that

\[
\inf_{\| V \|^2_{L^2} \neq 0} \left\{ \left( \frac{\| V \|^2_{L^2}}{\| V \|^2_{L^2}} \right) \right\} = 0.
\]

This implies that

\[
\inf_{\| V \|^2_{L^2} \neq 0} \left\{ \left( \frac{\| V \|^2_{L^2}}{\| V \|^2_{L^2}} \right) \right\} = 0.
\]

Since \( V \) satisfies the GHJB equation we get that

\[
\inf_{\| V \|^2_{L^2} \neq 0} \left\{ \left( \frac{\| V \|^2_{L^2}}{\| V \|^2_{L^2}} \right) \right\} = 0.
\]

This implies that there exists a sequence \( \{V_k\} \) such that \( \lim_{k \to \infty} V_k \neq V \) but \( \lim_{k \to \infty} V_k \) is a solution to the GHJB equation. This statement however, contradicts the assumption on uniqueness.

The proof is complete since Lemma 8 guarantees that the solution to the GHJB equation is unique.

**Corollary 21.** Under the hypothesis of Lemma 20, if

- C7: \( \sum_{j=1}^{\infty} \| l \|_{L^2(\Omega)} \| \phi_j \|^2 < \infty \),
- C8: \( \sum_{j=1}^{\infty} \| \left( \frac{\partial \phi_j}{\partial x} \right)(f + gu) \cdot \phi_j \|^2 < \infty \),

then convergence is uniform on \( \Omega \).

**Proof.** Immediate from Lemma 19.

In the next lemma we show that the GHJB equation is bounded below so that the previous lemma imply convergence of the approximation to the solution.

**Lemma 22.** If conditions C1–C6 hold, then

\[
\| c_N - b_N \|_2 \to 0.
\]

**Proof.** Define

\[
\eta(x) \triangleq \text{GHJB}(V_N; u(x)),
\]

then for all \( x \in \Omega \),

\[
\text{GHJB}(V_N; u(x)) = \eta(x).
\]

Substituting the series expansion for \( V_N \) and \( V_W \), and moving the terms in the series that are greater than \( N \) to the right-hand side we obtain

\[
(c_N - b_N) \frac{\partial \phi_i(x)}{\partial x}(f + gu)
\]

\[
= \eta(x) + \sum_{j=N+1}^{\infty} b_j \frac{\partial \phi_j}{\partial x}(f + gu) \triangleq \hat{\eta}_N(x).
\]

Since \( \left( \frac{\partial \phi_j}{\partial x} \cdot (f + gu) \right)_j \) are continuous and linearly independent,

\[
\| (c_N - b_N) \frac{\partial \phi_i}{\partial x}(f + gu) \|_{L^2(\Omega)} = 0 \iff c_N = b_N.
\]
If \( c_N = b_N \) for all \( N \), then the theorem is proved. Assume that \( c_N \neq b_N \). Define \( W \) as

\[
W = \int_{\Omega} (\nabla \Phi_N(x) + gu)[\nabla \Phi_N(x) + gu]^T dx.
\]

Since the set \( \{ (\partial \phi_j / \partial x) \}^\infty_{j=1} \) is linearly independent, \( W \) is positive definite. Therefore,

\[
\| (c_N - b_N)^T \nabla \Phi_N(x) + gu \|_{L_2(\Omega)}^2 = (c_N - b_N)^T W (c_N - b_N) = \int_{\Omega} |\eta_N(x)|^2 dx \geq \lambda_{\text{min}}(W) \| c_N - b_N \|_2^2 > 0,
\]

where \( \lambda_{\text{min}}(W) \) is the minimum eigenvalue of \( W \). Therefore,

\[
\int_{\Omega} |\eta_N(x)|^2 dx \to 0 \Rightarrow \| c_N - b_N \|_2 \to 0.
\]

But by the mean value theorem, \( \exists \xi \in \Omega \) such that

\[
\int_{\Omega} |\eta_N(x) + \sum_{j=N+1}^{\infty} b_j \frac{\partial \phi_j}{\partial x}(f + gu)(\xi)|^2 dx \leq \mu(\Omega) \left( |\eta_N(\xi)|^2 + 2|\eta_N(\xi)| \sum_{j=N+1}^{\infty} b_j \frac{\partial \phi_j}{\partial x}(f + gu)(\xi) \right) + \int_{\Omega} \left| \sum_{j=N+1}^{\infty} b_j \frac{\partial \phi_j}{\partial x}(f + gu)(\xi) \right|^2 dx,
\]

where \( \mu(\Omega) \) is the Lebesgue measure of \( \Omega \). Lemma 20 implies the pointwise convergence of \( \eta_N(x) \), so \( \forall \varepsilon > 0, \exists K_1(\varepsilon) \) such that

\[
N > K_1(\varepsilon) \Rightarrow |\eta_N(\xi)| < \frac{\varepsilon}{\sqrt{\mu(\Omega)}}.
\]

Since \( \nabla \Phi_N(x; u) = 0 \),

\[
\sum_{j=1}^{\infty} b_j \frac{\partial \phi_j}{\partial x}(f + gu)(x) = -l(x) - \| u(x) \|_K
\]

converges pointwise so \( \forall \varepsilon > 0, \exists K_2(\varepsilon) \) such that \( N > K_2(\varepsilon) \) implies that

\[
\sum_{j=1}^{\infty} b_j \frac{\partial \phi_j}{\partial x}(f + gu)(\xi) < \frac{\varepsilon}{\sqrt{\mu(\Omega)}},
\]

which proves the lemma.

**Corollary 23.** Under the hypothesis of Lemma 22,

\[
\| V_N - V \|_{L_2(\Omega)} \to 0.
\]

**Proof.**

\[
\| V_N - V \|_{L_2(\Omega)}^2 = \int_{\Omega} |V_N - V|^2 dx = \int_{\Omega} |(c_N - b_N)^T \nabla \Phi_N(x)|^2 dx + \int_{\Omega} \sum_{j=N+1}^{\infty} b_j \phi_j(x)^2 dx = (c_N - b_N)^T (\nabla \Phi_N(x)) \cdot (c_N - b_N) + \int_{\Omega} \sum_{j=N+1}^{\infty} b_j \phi_j(x)^2 dx.
\]

By the mean value theorem, \( \exists \xi \in \Omega \) such that

\[
\| V_N - V \|_{L_2(\Omega)}^2 = \int_{\Omega} |(c_N - b_N)^T \nabla \Phi_N(x)|^2 dx + \mu(\Omega) \sum_{j=N+1}^{\infty} b_j \phi_j(x)^2 \to 0.
\]

It is not sufficient to know that the value function converges, we also need to know that the approximate control converges. We show in Lemma 24 that \( u_N \) converges uniformly to \( \hat{u} \), where \( \hat{u} \) is derived from \( V \) according to equation (6). Since Lemma 9 implies that \( \hat{u} \) is admissible, we are able to show in Lemma 25 that \( u_N \) is also admissible.

**Lemma 24.** If conditions C1–C6 are satisfied then \( \| u_N(x)(x) - \hat{u}(x) \|_K \to 0 \) pointwise on \( \Omega \). If in addition C7 and C8 are satisfied and

\[
C9: \partial V / \partial x \text{ can be approximated uniformly close on } \Omega \text{ by } \{ \partial \phi_j / \partial x \}^\infty_{j=1},
\]

then \( \| u_N(x) - \hat{u}(x) \|_K \to 0 \) uniformly on \( \Omega \).

**Proof.**

\[
\| u_N(x) - \hat{u}(x) \|_K \leq \| -\frac{1}{2} R^{-1} g^T(x) \nabla \Phi_N(x)(c_N - b_N) \|_K + \int_{\Omega} \sum_{j=N+1}^{\infty} b_j R^{-1} g^T(x) \frac{\partial \phi_j}{\partial x}(f + gu)(x) dx.
\]

\( \hat{u} = -\frac{1}{2} \sum_{j=1}^{\infty} b_j R^{-1} g^T \frac{\partial \phi_j}{\partial x}(f + gu) \) implies that the second term on the right-hand side converges pointwise to 0 and uniformly if condition C9 is satisfied. By Lemma 20 we know that

\[
\| (c_N - b_N)^T \nabla \Phi_N(x) \| \to 0 \text{ uniformly if conditions C7 and C8 are satisfied and } |\phi_j|^2 \text{ are linearly independent and } (f + gu) \text{ is admissible},
\]

we have from Lemma 11 that

\[
\| (c_N - b_N)^T \nabla \Phi_N(f + gu) \| \to 0 \Leftrightarrow \| c_N - b_N \| \to 0.
\]

In addition, from Corollary 12 we have that

\[
\| c_N - b_N \| \to 0 \Leftrightarrow \| \nabla \Phi_N(x)(c_N - b_N) \|_2 \to 0.
\]

Therefore, \( \| \nabla \Phi_N(x)(c_N - b_N) \|_2 \) converges in the same sense as \( \| (c_N - b_N)^T \nabla \Phi_N(f + gu) \| \). Since
$R^{-1}y^{T}(x)$ in continuous on $\Omega$ and hence uniformly bounded, we have that
\[ \|R^{-1}y^{T}(x)\nabla \Phi_{0}(x) (c_{N} - b_{N})\|_{R} \to 0 \]
in the same sense as $\| (c_{N} - b_{N})^{T} \nabla \Phi_{0} (f + gu(x)) \|$. 

The next lemma shows that for $N$ sufficiently large, $u_{N}$ is admissible.

**Lemma 25.** Under the conditions of Lemma 24, if

\( C_{10} \): The set $\{\|y(0)R^{-1}y^{T}(0) \nabla x_{k}(0)\|_{2} \}_{j=1}^{\infty}$ is uniformly bounded for all $N$, then for $N$ sufficiently large, $u_{N} \in \mathcal{A}(\Omega)$.

**Proof.** From Lemma 9 we know that $\hat{u} \in \mathcal{A}(\Omega)$. Therefore, from Lemma 10, $u_{N}$ is stabilizing on $\Omega$ if

\[ \hat{u}^{T}(x)Ru_{N}(x) > \frac{1}{2} \hat{u}^{T}(x)R\hat{u}(x) \leftrightarrow \hat{u}^{T}(x)R(2u_{N}(x)) - \hat{u}(x) > 0 \]

for all $x \in \Omega$. In the 1D case, the situation is shown in Fig. 1: if $\hat{u}(x) > 0$ then $u_{N}(x) > \frac{1}{2} \hat{u}(x)$ is stable, if $\hat{u}(x) < 0$ then $u_{N}(x) < \frac{1}{2} \hat{u}(x)$ is stable. From Lemma 24 we know that $u_{N}$ is within a uniform $\epsilon$ ball of $\hat{u}$, where $\epsilon$ can be made arbitrarily small by making $N$ large enough. Therefore, $u_{N}$ is guaranteed to be stabilizing everywhere but some ball $B(0; \rho_{N})$ centered at the origin, where $\rho_{N} \to 0$ as $N \to \infty$ (see Fig. 1). By Lyapunov's first theorem, $u_{N}$ will be stabilizing on a small region $\Omega \supseteq B(0; \rho_{N})$ (for $N$ sufficiently large) if and only if the real parts of the eigenvalues of the linearized system are less than zero. So we need to examine the eigenvalues of the matrix $(\partial f/\partial x)(f + gu(x))$. Define

\[ F \triangleq \frac{\partial f}{\partial x}(0), \]

\[ G_{j} \triangleq -\frac{1}{2} g(0)R^{-1}y^{T}(0) \frac{\partial^{2} \Phi_{j}}{\partial x_{k}^{2}}(0). \]

Since $\hat{u}(x)$ is stabilizing on $\Omega$, we know that

\[ \Re \left\{ \lambda \left( F + \sum_{j=1}^{N} b_{j}G_{j} \right) \right\} < 0, \]

where $\lambda(M)$ are the eigenvalues of $M$. So for $N$ sufficiently large, $u_{N}$ will be stabilizing if

\[ \Re \left\{ \lambda \left( F + \sum_{j=1}^{N} c_{j}G_{j} \right) \right\} < 0. \]

Note that

\[ \left\| F + \sum_{j=1}^{\infty} b_{j}G_{j} - F - \sum_{j=1}^{N} c_{j}G_{j} \right\|_{2} \leq \left\| \sum_{j=N+1}^{\infty} (b_{j} - c_{j})G_{j} \right\|_{2} + \left\| \sum_{j=N+1}^{\infty} b_{j}G_{j} \right\|_{2}. \]

Since we know that the series $\sum_{j=1}^{\infty} b_{j}G_{j}$ converges, $\forall \epsilon > 0, \exists K_{1}$ such that

\[ N > K_{1} \Rightarrow \left\| \sum_{j=N+1}^{\infty} b_{j}G_{j} \right\|_{2} < \epsilon/2. \]

Also, since $\| G_{j} \|_{2}$ are uniformly bounded, Lemma 22 implies that $\exists K_{2}$ such that $N > K_{2}$ implies

\[ \left\| \sum_{j=1}^{N} (b_{j} - c_{j})G_{j} \right\|_{2} \leq \| b_{N} - c_{N} \|_{2} \| G_{j} \|_{2}. \]

![Fig. 1. Gain margins for updated control $\hat{u}$.](image-url)
is less than $\varepsilon/2$, which proves that as $N \to \infty$,
\[
\lambda\left(F + \sum_{j=1}^{N} c_j G_j\right) \to \lambda\left(F + \sum_{j=1}^{\infty} b_j G_j\right).
\]
Since all of the eigenvalues of $F + \sum_{j=1}^{\infty} b_j G_j$ are strictly less than zero, there exists some $K$ after which all of the eigenvalues of $F + \sum_{j=1}^{N} c_j G_j$ are strictly less than zero. So for some finite $K$, $N > K$ implies that $u_N$ is stabilizing on $\Omega$.

To show that $u_N \in \mathcal{A}_d(\Omega)$ we must show that for all $x \in \Omega$
\[
J(x; u_N) \triangleq \int_{0}^{\infty} [\ell(x(t); x, u_N)]dt < \infty.
\]

But since the eigenvalues of $(\partial / \partial x)(f + gu_N)(0)$ can be made arbitrarily close to the eigenvalues of $(\partial / \partial x)(f + gu_N)(0)$, the decay rates of $f + gu_N$ and $f + gu$ are of the same order in a region close to zero. So (see Remark 3)
\[
J(x; u) < \infty \Rightarrow J(x; u_N) < \infty \Rightarrow u_N \in \mathcal{A}_d(\Omega). \quad \Box
\]

We have therefore established the following main result.

**Theorem 26.** Let
- $\Omega$ satisfy C2,
- $u$ satisfy C5,
- $(f, g, l)$ satisfy C3 and C4, and
- the set $\{\phi_j\}_{1}^{m}$ satisfy C1, C6–C10.

Under these conditions, $\forall \varepsilon > 0, \exists K$ such that $N > K$ implies that

(a) $\|V - V_N\|_{L^2(\Omega)} < \varepsilon$,
(b) $u_N(x) \in \mathcal{A}_d(\Omega)$.

4. ILLUSTRATIVE EXAMPLE

In this section we will use the method described in the previous section to solve the GHJB equation associated with a feedback linearizing control. We will show how this solution is used to obtain a control law that improves the closed-loop performance of the original control.

Consider the following nonlinear system:
\[
\dot{x} = \begin{pmatrix} -x_1^2 - x_2 \\ x_1 + x_2 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} u.
\]
(23)
The control objective is to regulate the system while minimizing the quadratic functional of the states and control
\[
J = \int_{0}^{\infty} x^T(t)x(t) + u^2(x(t))dt.
\]

To apply the method introduced in this paper we must choose
1. $u$: a stabilizing control,
2. $\Omega$: an estimate of the stability region of $u$, and
3. $\{\phi_j\}_{1}^{m}$: a set of complete basis functions.

We will first select these quantities and then show that our selections satisfy conditions C1–C10.

System [equation (23)] can be stabilized using feedback linearization. Using the method outlined in Isidori (1989) the system [equation (23)] is linearized with the state feedback
\[
u(x) = 3x_1^3 + 3x_1^2x_2 - x_2 + v
\]
(24)
and the coordinate transformation
\[
\begin{align*}
    z_1 &= x_1, \\
    z_2 &= x_1^2 + x_2.
\end{align*}
\]
(25)

In the new coordinates, system [equation (23)] becomes
\[
\dot{z} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} z + \begin{pmatrix} 0 \\ 1 \end{pmatrix} v.
\]

We use standard LQR theory to optimize the transformed system with respect to the cost function
\[
J = \int_{0}^{\infty} z^T(t)z(t) + v^2(z(t))dt
\]
to obtain the control
\[
u(z) = 0.4142z_1 - 1.3522z_2.
\]
(26)

A continuous stabilizing control for the original system is given by substituting equation (25) into equation (26) to obtain a $v$ as a function of $z$, and then using the control given in equation (24) to obtain
\[
u(x) = 3x_1^3 + 3x_1^2x_2 - x_2 + 0.4142x_1 - 1.3522(x_1^2 + x_2).
\]
(27)
The above control is stabilizing on $\mathbb{R}^2$ and so we are free to choose $\Omega$ as any compact set containing the origin. For simplicity, let $\Omega = [-1, 1] \times [-1, 1]$.

In this example we use polynomials as our approximating functions:
\[
\{\phi_j\}_{1}^{m} \triangleq \{x_1^2, x_1x_2, x_2^2, x_1^3, x_1^2x_2, x_1x_2^2, x_1^3, x_1^2x_2, \ldots\}.
\]

In the appendix we show that if $\Omega$ is a rectangle, symmetrically centered at the origin, $f + gu$ is a separable, odd-symmetric function, $l + \|u\|_2^2$ is a separable, even-symmetric function, and the set $\{\phi_j\}_{1}^{m}$ is composed of separable, even and odd-symmetric
functions then the coefficients associated with the odd-symmetric basis functions are zero. Since the system in this example satisfies these constraints, we will extract all \( \phi_j \)'s that are odd-symmetric. Therefore, the basis functions become

\[
\{ \phi_j \} \triangleq \left\{ x_1^2, x_1 x_2, x_2^2, x_1^3, x_1^2 x_2, x_1^3 x_2, x_1^2 x_3, x_1 x_2 x_3, x_2 x_3, x_1^4, x_1^3 x_2, x_1^2 x_2^2, x_1^2 x_2 x_3, x_1 x_2 x_3^2, x_2 x_3^2, \ldots \right\}. \tag{28}
\]

We will now show that \( \{ f, g, I, u, \{ \phi_j \} \} \) satisfy conditions C1–C10.

C2: \( \Omega \) is compact by definition.

C3: By the mean value theorem, any differentiable function on a compact set is Lipschitz continuous on the same set, therefore \( f \) and \( g \) are Lipschitz continuous on \( \Omega \) and \( f(0) = 0 \).

C4: \( I \triangleq x^T x \) and \( R = 1 \) are positive definite by definition.

C5: We must show that \( u \) defined in equation (27) satisfies Definition 1. Clearly, \( u \) is continuous and satisfies \( u(0) = 0 \). \( u \) is shown to be stabilizing on \( \Omega \) by using the transformed system to find a Lyapunov function

\[
V_{lyap}(x) = \begin{pmatrix} x_1 \\ x_3^2 + x_2 \end{pmatrix}^T P \begin{pmatrix} x_1 \\ x_3^2 + x_2 \end{pmatrix}
\]

where \( P \) satisfies the Lyapunov equation \( AP + PA^T = -I \) and

\[
A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \begin{pmatrix} 0.4142 & -1.3522 \end{pmatrix}.
\]

The finiteness of the integral \( \int_\Omega (l + \|u\|_2) \) \( dr \) for all \( x \in \Omega \) follows from Remark 3 by noting that

\[
\text{eig}\left\{ \frac{\partial(f + gu)}{\partial x}(0) \right\} = -0.6761 \pm j0.9783.
\]

C1: We must show that \( V \in \mathcal{P}(\Phi, \Omega) \cap L^2(\Omega) \). The Weierstrass approximation theorem guarantees that if \( V(x) \) is continuous, then \( V \in \mathcal{P}(\Phi, \Omega) \). If \( V(x) \) is continuous and finite on compact \( \Omega \) then \( V \in L^2(\Omega) \). C2–C5 and Lemma 8 show that \( V \) is both continuous and finite on \( \Omega \).

C6: Since \( (\partial \phi_j / \partial x) \cdot (f + gu), \|u\|_2 \) and \( l \) are all polynomials of degree greater than 2, they are in \( \mathcal{P}(\Phi, \Omega) \).

C7: Since \( l + \|u\|_2 \) are polynomials we have that there exists some \( K \) such that \( l + \|u\|_2 \in \text{span}\{\phi_j\}_1^K \), i.e.

\[
l(x) + \|u(x)\|_2^2 = \sum_{j=1}^{K} d_j \phi_j(x).
\]

Therefore

\[
\sum_{j=1}^{K} \langle l + \|u\|_2^2, \phi_j \rangle \phi_j = \sum_{j=1}^{K} \sum_{k=1}^{K} d_k \langle \phi_k, \phi_j \rangle \phi_j = \sum_{k=1}^{K} \sum_{k=1}^{K} d_k \phi_k(x)
\]

which clearly satisfies Definition 17 since the tail of the series is zero after \( K \) terms.

C8: Follows from a similar argument as C7.

C9: Lemma (8) guarantees that \( \partial V / \partial x \) is continuous, so the Weierstrass approximation theorem guarantees that this function can be approximated uniformly close by polynomials.

C10: From equation (28) we see that each basis function can be written as \( \phi_j = x_1^j x_2^k \) where \( p_j + q_j \geq 2 \). It can be easily seen that the matrix \( (\partial^2 \phi_j / \partial x^2)(0) \) is nonzero only when

\[
(p_j, q_j) \in \{(2,0),(0,2),(1,1)\}.
\]

![Fig. 2. Phase portrait for \( u, u_s (x^2), u_{s1} (x^2), u_{15} (x^2) \).](image-url)
in which case it is constant. Therefore, the set
\( \{\|g(0)R^{-1}g^T(0)\|_{\infty}^2\} \) is uniformly bounded for all \( N \).

In this example we will use \( N = 0, 3, 8, 15 \) basis functions, where \( N = 0 \) corresponds to the initial control [equation (27)], and \( N = 3 \) (respectively, 8, 15) corresponds to basis functions with terms of order up to \( x^2 \) (respectively, \( x^4, x^6 \)). An approximate control with \( N \) terms is given by the formula

\[
u_N(x) = -\frac{1}{2}R^{-1}g^T(x)\left(\sum_{j=1}^{N} c_j \frac{\partial \phi_j}{\partial x}(x)\right).
\]

For example, when \( N = 3, 8, 15 \) we obtain the following feedback control laws:
\[
u_3(x) = -0.5484x_1 - 2.5258x_2
\]
\[
u_8(x) = -0.4215x_1 - 2.2225x_2 - 0.4784x_1^3 \\
+ 0.2719x_1^2x_2 + 0.6494x_1x_2^2 + 0.0588x_2^3,
\]
\[
u_{15}(x) = 0.2652x_1 - 2.6857x_2 - 1.0960x_1^3 \\
+ 1.2144x_1^2x_2 - 1.0805x_1x_2^2 + 0.4505x_2^3 \\
- 0.2191x_1^4 - 0.9098x_1^4x_2 + 1.0508x_1^3x_2^2 \\
- 0.4009x_1^3x_2^3 + 0.2669x_1x_2^4 - 0.1525x_2^5.
\]

Figure 2 shows the phase portrait to the system under the control \( u_0, u_3, u_8, \) and \( u_{15} \). Figure 3 shows the cost of these controls for initial conditions \( (x_10)^T, x_1 \in [-1, 1] \). Increasing \( N \) larger than 15 does not result in noticeable improvement in the cost. This plot shows the effectiveness of the method proposed in this paper for improving the closed-loop performance of a control obtained via feedback linearization and the LQR method.

5. CONCLUSIONS

In this paper we posed the problem of finding a practical method to improve the closed-loop performance of a stabilizing feedback control laws for nonlinear systems and showed that the problem reduces to solving the generalized Hamilton–Jacobi–Bellman (GHJB) equation. We showed that Galerkin’s spectral method could be used to approximate the GHJB equation such that the resulting control is in feedback form and stabilizes the closed-loop system.

There are several advantages of the method. First, the method produces feedback control laws. Second, the controls are robust in the same sense as the optimal control. Third, all computations are performed off-line and once a solution is found, the control can be implemented in hardware and run in real time. Fourth, the stability region contains the set \( \Omega \) which is specified by the designer and is only restricted to be contained in the stability region of the initial admissible control. Finally, coefficients for the state and control weighting functions can be taken outside the integral so tuning the control through the penalty function becomes computationally fast. The disadvantages are that \( O(N^2) \) \( n \)-dimensional integrals need to be computed and...
that, since the control laws are given as a series of basis functions, they are inherently complex. Implementation issues associated with the method are discussed in Beard (1995) and Beard et al. (1996). In addition, it should be pointed out that the results of this paper only guarantee that for $N$ sufficiently large, the method converges. However, for a given $\varepsilon$, we have not given an estimate of how large $N$ might have to be. An additional disadvantage is that it is not clear how conditions C7 and C8 might be checked for a particular system.

The method introduced in this paper can be iterated to produce a procedure similar to Bellman’s approximation in policy space (Bellman and Dreyfus, 1962) Full implementation of this combined algorithm can be found in Beard (1995) and is the subject of Beard et al. (1998). The method has been extended to finite-time horizon problems for stationary and non-stationary plants as reported in Beard (1995). It has also been applied to the Hamilton–Jacobi–Isaacs equations that arise in nonlinear $H_{\infty}$ optimal control. The results on topic will be the subject of forthcoming papers.

Acknowledgements—Partially supported by NASA grant NAGW-1333.

REFERENCES


**APPENDIX A. SYMMETRIC SYSTEMS OVER HYPERCUBES**

In this appendix we will show that under certain conditions, we can reduce the number of basis functions used to approximate $V$. Our motivation is from Lemma 8 where it was shown that $V$ is positive definite. Therefore, it should be sufficient to approximate $V$ with positive-definite basis functions.

We begin by making a number of definitions.

$f: \mathbb{R}^n \to \mathbb{R}$ is called "separable on $\Omega$" if $f(x) = \Pi_{i=1}^n f_i(x_i)$ for all $x \in \Omega$. $f: \mathbb{R}^n \to \mathbb{R}$ is called "even-symmetric on $\Omega$" if $f(-x) = f(x)$ for all $x \in \Omega$. $f: \mathbb{R}^n \to \mathbb{R}$ is called "odd-symmetric on $\Omega$" if $f(-x) = -f(x)$ for all $x \in \Omega$. Define $\mathcal{S}_e \{ f: \mathbb{R}^n \to \mathbb{R} \}$ is separable and odd-symmetric on $\Omega$ and $\mathcal{S}_o \{ f: \mathbb{R}^n \to \mathbb{R} \}$ is separable and even-symmetric on $\Omega$.

Let $\mathcal{S}_e \{ $ (resp. $\mathcal{S}_o \{ $) be the set of vector-valued functions whose elements are in $\mathcal{S}_e \{ $ (resp. $\mathcal{S}_o \{ $). We will first state a number of facts that follow directly from the definitions.

**Fact 1**: If $\eta_1, \eta_2 \in \mathcal{S}_e \{ $ and $\sigma_1, \sigma_2 \in \mathcal{S}_o \{ $, then $\eta_1 \eta_2 \in \mathcal{S}_e \{ $, $\sigma_1 \sigma_2 \in \mathcal{S}_o \{ $, $\eta_1 \sigma_2 \in \mathcal{S}_e \{ $.

**Fact 2**: If $\eta: \mathbb{R}^n \to \mathbb{R}$ is separable then $\eta = \Pi_{i=1}^n \eta_i(x_i) \in \mathcal{S}_e \{ $ if there are an even number (possibly zero) of $\eta_i$'s that are odd-symmetric. Similarly, $\eta = \Pi_{i=1}^n \eta_i(x_i) \in \mathcal{S}_o \{ $ if there are an odd number (at least one) of $\eta_i$'s that are odd-symmetric.

**Fact 3**: If $\Omega = \{ -a_1, a_1 \} \times \cdots \times \{ -a_n, a_n \}$ and $\eta \in \mathcal{S}_o \{ $, then $\int_{\Omega} \eta(x) \, dx = 0$. (Proof. Recoder $\Pi_{i=1}^n \eta_i(x)$ such that $\eta_1$ is odd, which fact 2 guarantees is possible, then $\int_{\Omega} \eta = \int_{-a_1}^{a_1} \eta_1(x) \Pi_{i=2}^n \sigma_i(x) \, dx = 0$.)

**Fact 4**: If $\eta_i: \mathbb{R} \to \mathbb{R}$ is even (resp. odd) on $\Omega$, then $\partial \eta_i / \partial x_j$ is odd (resp. even) on $\Omega$. (Proof. $\eta_i$, even $\Rightarrow \eta_i(x_j) = \eta_i(-x_j) \Rightarrow (\partial \eta_i / \partial x_j)(x_j) = (\partial \eta_i / \partial x_j)(-x_j)$).

**Fact 5**: If $\eta \in \mathcal{S}_e \{ $ (resp. $\mathcal{S}_o \{ $) then $\partial \eta / \partial x \in \mathcal{S}_e \{ $ (resp. $\mathcal{S}_o \{ $). (Proof.)
Theorem A.1. Assume that the conditions of Lemma 14 are satisfied. Let $V_N = \sum_{j=1}^{N} c_j\phi_j$ satisfy
\[
\left\langle \frac{\partial V_N}{\partial x}(f + gu) + l + \|u\|_R^2, \Phi_N \right\rangle_v = 0,
\]
and suppose that

T1: $\Omega$ is a rectangle, symmetrically centered at the origin, i.e. $\Omega = [-a_1, a_1] \times \cdots \times [-a_n, a_n]$.
T2: $f + gu \in S_o^m$.
T3: $l + \|u\|_R^2 \in S_o$.
T4: $\{\phi_j\}^T \subseteq S_o \cup S_e$.

If $\phi_j \in S_o$, then $c_j = 0$ (i.e. $V_N \in S_o$).

Proof. Define
\[
K_o = \{k_1, \ldots, k_{[K_o]}\} \triangleq \{k \in [1, N]: \phi_k \in S_o\},
K_e = \{k_1, \ldots, k_{[K_e]}\} \triangleq \{k \in [1, N]: \phi_k \in S_e\}.
\]

Define $\Phi_K \triangleq (\phi_{k_1}, \ldots, \phi_{k_{[K_e]}})^T$ and $c_K \triangleq (c_1, \ldots, c_{[K_e]})^T$ and $\Phi_K$ and $c_K$ accordingly. Then
\[
\left\langle \frac{\partial V_N}{\partial x}(f + gu) + l + \|u\|_R^2, \Phi_N \right\rangle_v = 0
\]
can be written as
\[
M\begin{pmatrix} c_K \\ e_K \end{pmatrix} = \left( \begin{pmatrix} -\langle l + \|u\|_R, \Phi_K \rangle_v \\ -\langle l + \|u\|_R^2, \Phi_K \rangle_v \end{pmatrix} \right),
\]
where
\[
M \triangleq \begin{pmatrix} \langle \nabla \Phi_K(f + gu), \Phi_K \rangle_m & \langle \nabla \Phi_K(f + gu), \Phi_K \rangle_m \\ \langle \nabla \Phi_K(f + gu), \Phi_K \rangle_m & \langle \nabla \Phi_K(f + gu), \Phi_K \rangle_m \end{pmatrix}.
\]
The previously stated facts show that
\[
\langle \nabla \Phi_K(f + gu), \Phi_K \rangle_m = 0,
\]
\[
\langle \nabla \Phi_K(f + gu), \Phi_K \rangle_m = 0,
\]
\[
\langle l + \|u\|_R^2, \Phi_K \rangle_v = 0.
\]
Therefore, $c_K$ satisfies the equation
\[
\langle \nabla \Phi_K(f + gu), \Phi_K \rangle_m c_K = 0.
\]
The theorem follows from Lemma 14 which shows that the matrix $\langle \nabla \Phi_K(f + gu), \Phi_K \rangle_m$ is full rank. \qed