10

DIFFERENTIAL GAMES

This chapter presents some basic ideas of differential games. Modern day society relies on the operation of complex systems, including aircraft, automobiles, electric power systems, economic entities, business organizations, banking and finance systems, computer networks, manufacturing systems, and industrial processes. Decision and control are responsible for ensuring that these systems perform properly and meet prescribed performance objectives. Networked dynamical agents have cooperative team-based goals as well as individual selfish goals, and their interplay can be complex and yield unexpected results in terms of emergent teams. Cooperation and conflict of multiple decision-makers for such systems can be studied within the field of cooperative and noncooperative game theory.

We discussed linear quadratic Nash games and Stackelberg games in Section 8.6. There, a unified treatment was given for linear games and decentralized control using structured design methods that are familiar from output feedback design. Controller designs were given based on solving coupled matrix equations. Here, we discuss nonlinear and linear differential games using a Hamiltonian function formulation.

In the first part of the book we used the calculus of variations to develop optimal control results such as those in Tables 3.2-1 and 3.3-1. In this chapter we use an approach based on the Hamiltonian function, Pontryagin's minimum principle (PMP) (Pontryagin et al. 1962), and the Bellman equation to develop optimal game theoretic solutions. PMP was discussed in Section 5.2 in connection with constrained input problems. First we show how to use PMP to derive the solutions to the nonlinear and linear optimal control problems for a single agent given in Tables 3.2-1 and 3.3-1. The Bellman equation is defined in terms of the Hamiltonian function and appears naturally in that development. Then we use PMP and the Bellman equation to solve multi agent decision problems. First we study 2-player noncooperative Nash zero-sum games, which have applications to H-infinity robust control. Finally, we use PMP and the Bellman equation to solve multiplayer nonzero sum games, which can have a cooperative team component as well as noncooperative components. In this chapter we include some proofs that show the relationship between performance criteria, value functions, and Lyapunov functions. The proofs also show the way of thinking in this field of research.

10.1 OPTIMAL CONTROL DERIVED USING PONTRYAGIN'S MINIMUM PRINCIPLE AND THE BELLMAN EQUATION

Q1 In this section we rederive the Hamilton-Jacobi-Bellman (HJB) equation from Section 6.3 and the Riccati equation solution of Table 3.3-1 using PMP and the Bellman equation.

Optimal Control for Nonlinear Systems

Consider the continuous-time nonlinear affine system

$$\dot{x} = F(x, u) = f(x) + g(x)u, \tag{10.1-1}$$

with state $x(t) \in R^n$ and control $u(t) \in R^m$. Let the drift dynamics $f(x)$ be locally Lipschitz and $f(0) = 0$ so that $x = 0$ is an equilibrium point. With this system, associate the performance index or value integral

$$V(x(t)) = \int_t^\infty r(x, u)d\tau = \int_t^\infty (Q(x) + u^{\mathrm{T}}Ru)d\tau \tag{10.1-2}$$

with $Q(x) > 0$, $R > 0$. The quadratic form in the control makes the development here simpler, but similar results hold for more general positive definite control weightings. A control $u(t)$ is said to be *admissible* if it is continuous, stabilizes the system (10.1-1), and makes the value (10.1-2) finite.

The optimal control problem is to select the control input $u(t)$ in (10.1-1) to minimize the value (10.1-2). Differentiating $V(x(t))$ using Leibniz's formula, one sees that a differential equivalent to the value integral is given in terms of the Hamiltonian function $H(\cdot)$ by the *Bellman equation*

$$0 = r(x, u) + \dot{V} = r(x, u) + \left(\frac{\partial V}{\partial x}\right)^{\mathrm{T}} \dot{x}$$

or

$$0 = Q(x) + u^{\mathrm{T}}Ru + \left(\frac{\partial V}{\partial x}\right)^{\mathrm{T}} (f(x) + g(x)u) \equiv H\left(x, \frac{\partial V}{\partial x}, u\right). \tag{10.1-3}$$

That is, given any stabilizing feedback control policy $u(x)$ yielding finite value, the positive definite solution to the Bellman equation (10.1-3) is the value given by (10.1-2). The Bellman equation is a partial differential equation for the value. The initial condition is $V(0) = 0$.

It is desired to select the control input to minimize the value. A necessary condition for this is Pontryagin's minimum principle

$$H(x^*, \frac{\partial V^*}{\partial x}, u^*) \leq H(x^*, \frac{\partial V^*}{\partial x}, u) \tag{10.1-4}$$

for all admissible controls *u(t)*, where * denotes the optimal (minimizing) state, value, and control. Since the control *u(t)* is unconstrained this is equivalent to the stationarity condition

$$\frac{\partial H}{\partial u} = 0. \tag{10.1-5}$$

Applying PMP to (10.1-3) yields

$$u(x) = u(V(x)) \equiv -\frac{1}{2}R^{-1}g^{\mathrm{T}}(x)\frac{\partial V}{\partial x} = -\frac{1}{2}R^{-1}g^{\mathrm{T}}(x)\nabla V(x), \tag{10.1-6}$$

where $\nabla V(x) = \partial V/\partial x \in R^n$ is the gradient, taken as a vector. Substitute this into the Bellman equation (10.1-3) to obtain

$$0 = Q(x) + \nabla V^{\mathrm{T}}(x)f(x) - \tfrac{1}{4}\nabla V^{\mathrm{T}}(x)g(x)R^{-1}g^{\mathrm{T}}(x)\nabla V(x). \qquad V(0) = 0 \tag{10.1-7}$$

This is the Hamilton-Jacobi-Bellman (HJB) equation from Section 6.3. It can also be written as

$$0 = \min_u [H(x, \nabla V^*, \ u)]. \tag{10.1-8}$$

This can be written in terms of the optimal Hamiltonian as

$$0 = H(x^*, \nabla V^*, u^*), \tag{10.1-9}$$

where $u^* = u(V^*(x)) = -\frac{1}{2}R^{-1}g^{\mathrm{T}}(x)\nabla V^*(x)$.

To solve the optimal control problem, one solves the HJB equation (10.1-7) for the optimal value $V^* > 0$, then the optimal control is given as a state variable feedback $u(V^*(x))$ in terms of the HJB solution by (10.1-6).

Note that in the calculus of variations derivation of Table 3.2-1, the Hamiltonian function $H(x, \nabla V, u)$ plays a central role, and its various partial derivatives define the flows of the state equation and the costate equation, as well as the optimal control through the stationarity condition. Yet, there is no mention there that the Hamiltonian should be equal to zero. The Hamiltonian is set to zero in the Bellman equation (10.1-3), which arises from differentiation of the value function (10.1-2), to get a differential equivalent to (10.1-2). Note further that the costate of Table 3.2-1 is the gradient vector $\nabla V(x)$.

Now we give a formal proof that the solution to the HJB equation provides the optimal control solution. The following key fact is instrumental. It shows that the Hamiltonian function is quadratic in the control deviations from a certain key control value.

Lemma 10.1-1. For any admissible control policy $u(x)$, let $V(x) \geq 0$ be the corresponding solution to the Bellman equation (10.1-3). Define $u^* = u(V(x))$ by (10.1-6) in terms of $V(x)$. Then

$$H(x, \nabla V, u) = H(x, \nabla V, u^*) + (u - u^*)^{\mathrm{T}} R(u - u^*). \tag{10.1-10}$$

Proof: The Hamiltonian function is

$$H(x, \nabla V, u) = Q(x) + u^{\mathrm{T}} Ru + \nabla V^{\mathrm{T}}(f + gu).$$

Complete the squares to write

$$\begin{aligned} H(x, \nabla V, u) = {} & \nabla V^{\mathrm{T}} f + Q(x) + (\tfrac{1}{2}\nabla V^{\mathrm{T}} g R^{-1} + u^{\mathrm{T}}) R (\tfrac{1}{2} R^{-1} g^{\mathrm{T}} \nabla V + u) \\ & - \tfrac{1}{4}\nabla V^{\mathrm{T}} g R^{-1} g^{\mathrm{T}} \nabla V. \end{aligned}$$

■

The next result shows that under certain conditions the HJB solution solves the optimal control problem. Set $Q(x) = h^{\mathrm{T}}(x)h(x)$. System

$$\dot{x} = f(x) + g(x)u(x), \quad y = h(x) \tag{10.1-11}$$

is said to be zero-state observable if $u(t) \equiv 0$, $y(t) \equiv 0 \Rightarrow x(t) = 0$.

Theorem 10.1-2. Solution to Optimal Control Problem. Consider the optimal control problem for (10.1-1), (10.1-2) with $Q(x) = h^{\mathrm{T}}(x)h(x)$. Suppose $V^*(x) \in C^1 : R^n \to R$ is a smooth positive definite solution to the HJB equation (10.1-7). Define control $u^* = u(V^*(x))$ as given by (10.1-6). Assume (10.1-11) is zero-state observable. Then the closed-loop system

$$\dot{x} = f(x) + g(x)u^* = f(x) - \tfrac{1}{2} g R^{-1} g^{\mathrm{T}} \nabla V^* \tag{10.1-12}$$

is locally asymptotically stable. Moreover, $u^* = u(V^*(x))$ minimizes the performance index (10.1-2) over all admissible controls, and the optimal value on $[0, \infty)$ is given by $V^*(x(0))$.

Proof:

a. Stability.

Note that for any C^1 function $V(x) : R^n \to R$ one has, along the system trajectories,

$$\frac{dV}{dt} = \frac{\partial V}{\partial t} + \frac{\partial V}{\partial x}^{\mathrm{T}} \dot{x} = \frac{\partial V}{\partial x}^{\mathrm{T}} (f + gu),$$

so that

$$\frac{dV}{dt} + h^{\mathrm{T}}h + u^{\mathrm{T}}Ru = H(x, \nabla V, u).$$

Suppose now that $V(x)$ satisfies the HJB equation $H(x^*, \nabla V^*, u^*) = 0$. Then according to (10.1-10) one has

$$\begin{aligned} H(x, \nabla V^*, u) &= H(x^*, \nabla V^*, u^*) + (u - u^*)^{\mathrm{T}} R(u - u^*) \\ &= (u - u^*)^{\mathrm{T}} R(u - u^*). \end{aligned}$$

Therefore,

$$\frac{dV}{dt} + h^{\mathrm{T}}h + u^{\mathrm{T}}Ru = (u - u^*)^{\mathrm{T}} R(u - u^*).$$

Selecting $u = u^*$ yields

$$\begin{aligned} &\frac{dV}{dt} + h^{\mathrm{T}}h + u^{\mathrm{T}}Ru = 0 \\ &\frac{dV}{dt} \le -(h^{\mathrm{T}}h + u^{\mathrm{T}}Ru). \end{aligned}$$

Now LaSalle's extension shows that the state goes to a region of R^n wherein $\dot{V} = 0$. However, zero-state observable means $u(t) \equiv 0, y(t) \equiv 0 \Rightarrow x(t) = 0$. Therefore, the system is locally asymptotically stable with Lyapunov function $V(x) > 0$.

b. Optimality.

For any C^1 smooth function $V(x) : R^n \to R$ and $T > 0$ one can write the performance index (10.1-2) as

$$\begin{aligned} V(x(0), u) &= \int_0^T \left(h^{\mathrm{T}}h + u^{\mathrm{T}}Ru\right) dt + \int_0^T \dot{V}\, dt - V(x(T)) + V(x(0)). \\ &= \int_0^T \left(h^{\mathrm{T}}h + u^{\mathrm{T}}Ru\right) dt + \int_0^T \nabla V^{\mathrm{T}}(f + gu)\, dt - V(x(T)) + V(x(0)) \\ &= \int_0^T H(x, \nabla V, u)\, dt - V(x(T)) + V(x(0)). \end{aligned}$$

Now, suppose $V(x)$ satisfies the HJB equation (10.1-7). Then $0 = H(x^*, \nabla V^*, u^*)$ and (10.1-10) yields

$$V(x(0), u) = \int_0^T \left((u - u^*)^{\mathrm{T}} R(u - u^*)\right) dt - V^*(x(T)) + V^*(x(0)).$$

Assuming $u(t)$ is admissible one has $V^*(x(\infty)) = 0$ and can write

$$V(x(0), u) = \int_0^\infty \left((u - u^*)^\mathrm{T} R(u - u^*)\right) dt + V^*(x(0)),$$

which shows that u^* is an optimal control and the optimal value is $V^*(x(0))$. ■

Solution of HJB Equations

The HJB equation may not have smooth solutions, but may have the so-called viscosity solutions (Bardi and Capuzzo-Dolcetla 1997). Under certain local reachability and observability assumptions, it has a local smooth solution (van der Schaft 1992). Various other assumptions guarantee existence of smooth solutions, such as that the dynamics not be bilinear and the value not contain cross-terms in the state and control input.

The HJB equation is difficult to solve for nonlinear systems. An algorithm for finding an approximate smooth solution to the HJB is given by Abu-Khalaf and Lewis (2005). The solution to optimal control problems is generally obtained by solving the HJB equation offline. This does not allow the objectives to change in real time. The work of Vamvoudakis and Lewis (2010a) shows how to solve the HJB equation online in real time using data measured along the system trajectories. This allows solution of the optimal control problem if $Q(x)$, R in (10.1-2) change slowly with time. Vrabie and Lewis (2009) has provided algorithms for solving the HJB equations online in real time without knowing the system drift dynamics $f(x)$.

In Chapter 11 we showed how to solve these equations online in real time using adaptive control techniques (Vrabie and Lewis 2009, Vamvoudakis and Lewis 2010a). We developed there a class of adaptive controllers that converge online to optimal control solutions by measuring data along the system trajectories. The approach is based on techniques of policy iteration from reinforcement learning and uses two approximator structures: one critic neural network to solve the Bellman equation (10.1-3), and one actor neural network to compute the control using (10.1-6). Note that if these two equations are solved simultaneously, then one has effectively solved the HJB equation (10.1-7).

Linear Quadratic Regulator

For the linear quadratic regulator (LQR) one has the linear dynamics

$$\dot{x} = F(x, u) = Ax + Bu, \tag{10.1-13}$$

with state $x(t) \in R^n$ and control $u(t) \in R^m$, and quadratic value integral

$$V(x(t)) = \int_t^\infty r(x, u)d\tau = \frac{1}{2}\int_t^\infty (x^\mathrm{T}Qx + u^\mathrm{T}Ru)d\tau, \tag{10.1-14}$$

with $Q \geq 0$, $R > 0$. Then, the value is quadratic in the state so that

$$V(x) = \tfrac{1}{2}x^{\mathrm{T}}Sx \tag{10.1-15}$$

for some matrix $S > 0$. Substituting this into (10.1-3) gives the Bellman equation

$$0 = \tfrac{1}{2}(x^{\mathrm{T}}Qx + u^{\mathrm{T}}Ru) + x^{\mathrm{T}}S(Ax + Bu) \equiv H(x, \nabla V, u). \tag{10.1-16}$$

Using now a linear state variable feedback (SVFB)

$$u = -Kx \tag{10.1-17}$$

gives

$$0 = (A - BK)^{\mathrm{T}}S + S(A - BK) + Q + K^{\mathrm{T}}RK. \tag{10.1-18}$$

It has been assumed that the Bellman equation holds for all initial conditions, and the state $x(t)$ has been cancelled in writing (10.1-18). It is seen that, for the LQR, the Bellman equation is equivalent to a Lyapunov equation for S in terms of the prescribed SVFB K. If $(A - BK)$ is stable and $(A, \sqrt{Q})$ observable, there is a positive definite solution $S > 0$, and then (10.1-15) is the value (10.1-14) for that selected feedback K.

The LQR optimal control that minimizes (10.1-14) is (10.1-6), or

$$u(x) = -R^{-1}B^{\mathrm{T}}Sx = -Kx. \tag{10.1-19}$$

Substituting this into (10.1-18) yields the algebraic Riccati equation (ARE)

$$0 = A^{\mathrm{T}}S + SA + Q - SBR^{-1}B^{\mathrm{T}}S \tag{10.1-20}$$

exactly as given in Table 3.3-1.

To solve the optimal control problem, one solves the ARE equation (10.1-20) for the optimal value kernel $S > 0$, then the optimal control is given as a state variable feedback in terms of the ARE solution by (10.1-19). There exists a solution $S > 0$ if $(A,\ B)$ is stabilizable and $(A, \sqrt{Q})$ is observable.

The work of Vamvoudakis and Lewis (2010a) shows how to solve the ARE equation online in real time using data measured along the system trajectories. This allows solution of the optimal control problem if the performance index weighting matrices Q, R change slowly with time. Vrabie and Lewis (2009) has provided algorithms for solving the ARE equation online in real time without knowing the system matrix A. In Chapter 11 we showed how to solve the ARE equation online in real time using adaptive control techniques based on this work.

10.2 TWO-PLAYER ZERO-SUM GAMES

In this section we use the Bellman equation approach to solve the 2-player zero-sum (ZS) games (Basar and Olsder 1999). ZS games refer to the fact that whatever one player gains, the other loses. The Nash solution of the 2-player ZS game is important in feedback control because it provides a solution to the bounded L_2-gain problem, and, hence, allows solution of the H-infinity disturbance rejection problem (Zames 1981, van der Schaft 1992, Knobloch et al. 1993).

Zero-sum Game

Consider the nonlinear time-invariant dynamical system given by

$$\dot{x} = f(x) + g(x)u + k(x)d, \tag{10.2-1}$$

where state $x(t) \in R^n$. This system has two inputs or players, known as the control input $u(t) \in R^m$ and the disturbance input $d(t) \in R^q$. Let $f(x)$ be locally Lipschitz and $f(0) = 0$. Define the performance index

$$J(x(0), u, d) = \int_0^\infty \left(Q(x) + u^{\mathrm{T}}Ru - \gamma^2 \|d\|^2\right) d\tau \equiv \int_t^\infty r(x, u, d)d\tau, \tag{10.2-2}$$

where $Q(x) = h^{\mathrm{T}}(x)h(x) \geq 0$ for some function $h(x)$, $R = R^{\mathrm{T}} > 0$, and $\gamma > 0$. This is a function of the initial state and the functions $u(\tau), d(\tau) : 0 \leq \tau$.

Define the 2-player zero-sum differential game

$$V^*(x(0)) = \min_u \max_d J(x, u, d) = \min_u \max_d \int_0^\infty \left(h^{\mathrm{T}}h + u^{\mathrm{T}}Ru - \gamma^2 \|d\|^2\right) dt, \tag{10.2-3}$$

where the control player seeks to minimize the value and the disturbance to maximize it. This game puts the control and disturbance at odds with one another so that anything one gains in its objective is lost by the other. Therefore, it is termed a zero-sum game. This is equivalent to defining two performance measures $J_1(x(0), u, d) = J(x(0), u, d) = -J_2(x(0), u, d)$ and solving the game where the control player seeks to minimize J_1 and the disturbance player seeks to minimize J_2.

This game has a unique solution if a game theoretic saddle point (u^*, d^*) exists; that is, if

$$V^*(x_0) = \min_u \max_d J(x(0), u, d) = \max_d \min_u J(x(0), u, d). \tag{10.2-4}$$

A saddle point is shown in Figure 10.2-1. The associated value V^* is called the value of the game. This is equivalent to the Nash equilibrium condition

$$J(x(0), u^*, d) \leq J(x(0), u^*, d^*) \leq J(x(0), u, d^*) \tag{10.2-5}$$

holding for all policies $u,\ d$. These conditions depend on the performance integral (10.2-2) and the dynamics (10.2-1). According to the Nash condition, if both

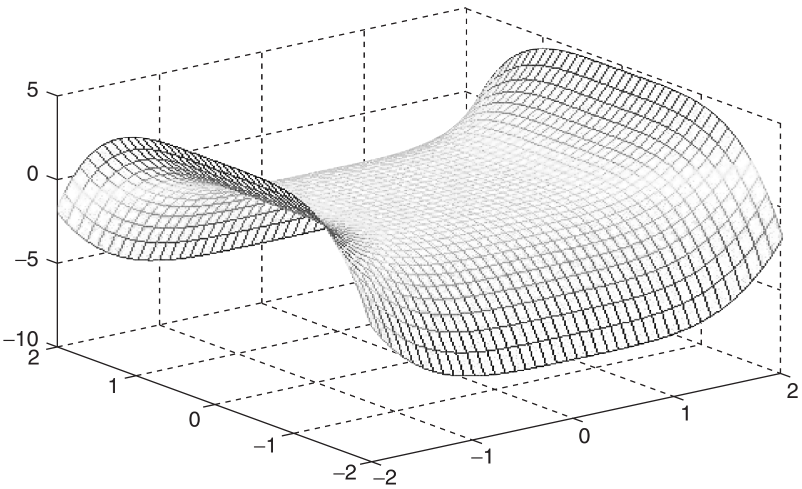

FIGURE 10.2-1 Saddle-point equilibrium.

players are at equilibrium, then neither has an incentive to change his policy unilaterally, since unilateral changes make one's performance worse.

For fixed control and disturbance feedback policies $u(x)$, $d(x)$define the value function

$$V(x(t), u, d) = \int_t^\infty \left(h^{\mathrm{T}}h + u^{\mathrm{T}}Ru - \gamma^2 \|d\|^2\right) d\tau \equiv \int_t^\infty r(x, u, d)d\tau, \quad (10.2\text{-}6)$$

which is only a function of the initial state $x(t)$. When the value is finite, a differential equivalent to this is found by differentiating using Leibniz's formula. The result is the nonlinear *ZS game Bellman equation* given in terms of the Hamiltonian function as

$$0 = r(x, u, d) + \dot{V} = r(x, u, d) + \nabla V^{\mathrm{T}}\dot{x}$$

or

$$0 = h^{\mathrm{T}}h + u^{\mathrm{T}}Ru - \gamma^2 \|d\|^2 + \nabla V^{\mathrm{T}}(f + gu + kd) \equiv H(x, \nabla V, u, d). \quad (10.2\text{-}7)$$

The boundary condition for this partial differential equation is $V(0) = 0$. A minimal solution such that $V(x) \geq 0$ to (10.2-7) is the value (10.2-6) for the given feedback policy $u(x)$ and disturbance policy $d(x)$.

A necessary condition for Nash condition (10.2-4) is Isaacs' condition

$$\min_u \max_d H(x, \nabla V, u, d) = \max_d \min_u H(x, \nabla V, u, d) \quad (10.2\text{-}8)$$

or, equivalently,

$$H(x, \nabla V, u^*, d) \leq H(x, \nabla V, u^*, d^*) \leq H(x, \nabla V, u, d^*) \quad (10.2\text{-}9)$$

for all policies $u,\ d$. These conditions hold pointwise in time. Under certain regularity conditions they guarantee (10.2-4). The Isaacs condition generalizes the PMP (10.1-4) to ZS games.

At equilibrium, one has the two stationarity conditions

$$\frac{\partial H}{\partial u} = 0, \quad \frac{\partial H}{\partial d} = 0. \tag{10.2-10}$$

Note from (10.2-7) that $\partial^2 H/\partial^2 u = 2R > 0$, $\partial^2 H/\partial^2 d = -2\gamma^2 < 0$, so that at the stationary point the Hamiltonian attains a minimum in u and a maximum in d, that is, a saddle point.

Applying (10.2-10) to the Hamiltonian (10.2-7) yields the control and disturbance policies

$$u = u(V(x)) \equiv -\tfrac{1}{2} R^{-1} g^{\mathrm{T}}(x) \nabla V \tag{10.2-11}$$

$$d = d(V(x)) \equiv \frac{1}{2\gamma^2} k^{\mathrm{T}}(x) \nabla V. \tag{10.2-12}$$

Substituting these into the Bellman equation (10.2-7) yields the Hamilton-Jacobi-Isaacs (HJI) equation

$$\begin{aligned} 0 = h^{\mathrm{T}} h + \nabla V^{\mathrm{T}}(x) f(x) - \frac{1}{4} \nabla V^{\mathrm{T}}(x) g(x) R^{-1} g^{\mathrm{T}}(x) \nabla V(x) \\ + \frac{1}{4\gamma^2} \nabla V^{\mathrm{T}}(x) k k^{\mathrm{T}} \nabla V(x), \quad V(0) = 0, \end{aligned} \tag{10.2-13}$$

which can be written

$$0 = H(x, \nabla V^*, u^*, d^*), \qquad V(0) = 0. \tag{10.2-14}$$

The minimum positive semi-definite (PSD) solution V^* of this equation gives the Nash value, and the Nash equilibrium solution $(u^*,\ d^*) = (u(V^*(x)),\ d(V^*(x)))$ is given by (10.2-11), (10.2-12) in terms of ∇V^*.

For the HJI equation to have a PSD solution, the gain γ should be chosen large enough (Başar and Olsder 1999, van der Schaft 1992). It is shown that there exists a critical gain γ^* such that the HJI has a PSD solution for any $\gamma > \gamma^*$. The critical value γ^* is known as the H-infinity gain for system (10.2-1). The H-infinity gain can be explicitly computed for linear systems (Chen et al. 2004).

Instrumental to the analysis of ZS games is the following key fact, which shows that the Hamiltonian is quadratic in deviations in the control and disturbance about some critical policies.

Lemma 10.2-1. For any policies $u(x), d(x)$ yielding a finite value, let $V(x) \geq 0$ be the corresponding solution to the Bellman equation 10.2-7. Define $(u^*,\ d^*) = (u(V(x)), d(V(x)))$ by (10.2-11), (10.2-12) in terms of $V(x)$. Then

$$H(x, \nabla V, u, d) = H(x, \nabla V, u^*, d^*) + (u - u^*)^{\mathrm{T}} R (u - u^*) - \gamma^2 \left\| d - d^* \right\|^2. \tag{10.2-15}$$

■

Exercise 10.2-1.

Prove (10.2-15) by completing the squares in (10.2-7). This result immediately shows that the Isaacs condition (10.2-9) holds for solutions V^* to the HJI equation (10.2-13), for according to (10.2-15), when (10.2-14) holds one has

$$H(x, \nabla V^*, u, d) = (u - u^*)^{\mathrm{T}} R(u - u^*) - \gamma^2 \left\| d - d^* \right\|^2, \tag{10.2-16}$$

which satisfies (10.2-9).

The next result shows that under certain conditions the HJI solution satisfies the Nash condition (10.2-4) and so solves the ZS game (Başar and Olsder 1999, van der Schaft 1992). System

$$\dot{x} = f(x) + g(x)u(x), \quad y = h(x) \tag{10.2-17}$$

is said to be zero-state observable if $u(t) \equiv 0$, $y(t) \equiv 0 \Rightarrow x(t) = 0$. ■

Theorem 10.2-2. Solution to 2-player ZS Game. Assume the game (10.2-3) has a finite value. Select $\gamma > \gamma^* > 0$. Suppose $V^*(x) \in C^1 : R^n \to R$ is a smooth positive semi-definite solution to the HJI equation (10.2-13) such that closed-loop system

$$\dot{x} = f(x) + g(x)u^* + k(x)d^* = f(x) - \frac{1}{2} g R^{-1} g^{\mathrm{T}} \nabla V^* + \frac{1}{2\gamma^2} k(x) k^{\mathrm{T}}(x) \nabla V^* \tag{10.2-18}$$

is locally asymptotically stable. Assume (10.2-17) is zero-state observable. The Nash condition (10.2-5) is satisfied for control $u^* = u(V^*(x))$ given by (10.2-11) and $d^* = d(V^*(x))$ given by (10.2-12) in terms of $V^*(x)$. Then the system is in Nash equilibrium, the game has a value, and (u^*, d^*) is a saddle-point equilibrium solution among policies in $L_2[0, \infty)$. Moreover, the value of the game is given by the HJI solution $V^*(x(0))$.

Proof: One has for any C^1 smooth function $V(x) : R^n \to R$ and $T > 0$

$$\begin{aligned}
J_T(x(0), u, d) &\equiv \int_0^T \left(h^{\mathrm{T}} h + u^{\mathrm{T}} R u - \gamma^2 \|d\|^2\right) dt \\
&= \int_0^T \left(h^{\mathrm{T}} h + u^{\mathrm{T}} R u - \gamma^2 \|d\|^2\right) dt \quad + \int_0^T \dot{V} dt \quad - V(x(T)) + V(x(0)) \\
&= \int_0^T \left(h^{\mathrm{T}} h + u^{\mathrm{T}} R u - \gamma^2 \|d\|^2\right) dt \\
&\quad + \int_0^T \nabla V^{\mathrm{T}} (f + gu + kd)\, dt \quad - V(x(T)) + V(x(0)) \\
&= \int_0^T H(x, \nabla V, u, d)\, dt \quad - V(x(T)) + V(x(0)).
\end{aligned}$$

Now, suppose $V^*(x)$ satisfies the HJI equation (10.2-13). Then $0 = H(x, \nabla V^*, u^*, d^*)$ and (10.2-15) yields

$$J_T(x(0), u, d) = \int_0^T \left((u - u^*)^{\mathrm{T}} R(u - u^*) - \gamma^2 \left\| d - d^* \right\|^2 \right) dt - V^*(x(T)) + V^*(x(0)).$$

Since $u(t),\ d(t) \in L_2[0, \infty)$, and since the game has a finite value as $T \to \infty$, this implies that $x(t) \in L_2[0, \infty)$; therefore, zero state observability implies $x(t) \to 0$, $V^*(x(\infty)) = 0$ and

$$J(x(0), u, d) = \int_0^\infty \left((u - u^*)^{\mathrm{T}} R(u - u^*) - \gamma^2 \left\| d - d^* \right\|^2 \right) dt + V^*(x(0)),$$

which implies saddle-point equilibrium, e.g., (10.2-5). One obtains

$$J(x(0), u^*, d^*) = \min_u \max_d J(x(0), u, d) = V^*(x(0)),$$

which shows the Nash solution is $u(t) = u^*(t), d(t) = d^*(t)$ and the Nash value is

$$V^*(x(0)).$$

■

Solution of HJI Equations

A minimum positive semi-definite solution $V^*(x) \geq 0$ to the HJI is one for which there is no other solution $V(x) \geq 0$ such that $V^*(x) \geq V(x) \geq 0$. The minimum positive semi-definite HJI solution is the unique HJI solution such that the closed-loop system (10.2-18) is locally asymptotically stable (Başar and Olsder 1999, van der Schaft 1992). More information is provided below on the discussion on linear quadratic zero-sum games.

The HJI equation may not have smooth solutions, but may have the so-called viscosity solutions (Bardi and Capuzzo-Dolcetla 1997). Under certain local reachability and observability assumptions, it has a local smooth solution (van der Schaft 1992).

The HJI equation is difficult to solve for nonlinear systems. An algorithm for finding an approximate minimum PSD solution to the HJI is given in Abu-Khalaf et al. (2006, 2008). The solution to ZS games is generally obtained by solving the HJI equation offline. This does not allow the objectives to change in real time as the players learn from each other. The work of Vamvoudakis and Lewis (2010b) shows how to solve the HJI equation online in real time using data measured along the system trajectories. This allows the weights $R,\ \gamma$ in (10.2-2) to change slowly as the game develops. Vrabie and Lewis (2010a,b) has provided algorithms for solving the HJI equations online in real time without knowing the

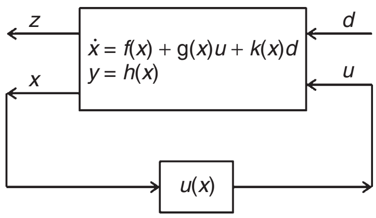

FIGURE 10.3-1 Bounded L_2-gain control.

system drift dynamics $f(x)$. These references extend the reinforcement learning techniques presented in Chapter 11 to online solution of zero-sum games.

10.3 APPLICATION OF ZERO-SUM GAMES TO H_∞ CONTROL

Consider the system (10.2.1) with output $y = h(x)$ and a performance output $z(t) = [u^T(t) \quad y^T(t)]^T$. This setup is shown in Figure 10.3-1. In the bounded L_2-gain problem (Zames 1981, van der Schaft 1992, Başar and Olsder 1999), one desires to find a feedback control policy $u(x)$ such that, when $x(0) = 0$ and for all disturbances $d(t) \in L_2[0, \infty)$ one has

$$\frac{\int_0^T \|z(t)\|^2\,dt}{\int_0^T \|d(t)\|^2\,dt} = \frac{\int_0^T (h^T h + u^T R u)dt}{\int_0^T \|d(t)\|^2\,dt} \le \gamma^2 \tag{10.3-1}$$

for a prescribed $\gamma > 0$ and $R = R^T > 0$ and for all $T > 0$. That is, the L_2-gain from the disturbance to the performance output is less than or equal to γ.

The H-infinity control problem is to find, if it exists, the smallest value $\gamma^* > 0$ such that for any $\gamma > \gamma^*$ the bounded L_2-gain problem has a solution. In the linear case an explicit expression can be provided for the H-infinity gain γ^* (Chen et al. 2004).

To solve the bounded L_2-gain problem, one may use the machinery of 2-player ZS games just developed. In the 2-player ZS game, both inputs can be controlled, with the control input seeking to minimize a performance index and the disturbance input seeking to maximize it. By contrast, here $d(t)$ is a disturbance that cannot be controlled, and $u(t)$ is the control input used to offset the deleterious effects of the disturbance.

The next result provides a solution to the bounded L_2-gain problem in terms of a solution to the HJI equation.

Theorem 10.3-1. Solution to Bounded L_2-gain Problem. Select $\gamma > \gamma^* > 0$. Suppose $V^*(x) > 0 \in C^1 : R^n \to R$ is a smooth positive definite solution to the HJI equation 10.2-13. Assume (10.2-17) is zero-state observable. Then the closed-loop system

$$\dot{x} = f(x) + g(x)u^* = f(x) - \tfrac{1}{2}gR^{-1}g^{\mathrm{T}}\nabla V^* \qquad (10.3\text{-}2)$$

is locally asymptotically stable with control input $u^* = u(V^*(x))$ given by (10.2-11) in terms of $V^*(x)$. Moreover, for this choice of control input, (10.3-1) holds for all disturbances $d(t) \in L_2[0, \infty)$.

Proof: Note that for any C^1 function $V(x) : R^n \to R$ one has, along the system trajectories,

$$\frac{dV}{dt} = \frac{\partial V}{\partial t} + \frac{\partial V}{\partial x}^{\mathrm{T}}\dot{x} = \frac{\partial V}{\partial x}^{\mathrm{T}}(f + gu + kd),$$

so that

$$\frac{dV}{dt} + h^{\mathrm{T}}h + u^{\mathrm{T}}Ru - \gamma^2 d^{\mathrm{T}}d = H(x, \nabla V, u, d).$$

Suppose now that $V(x)$ satisfies the Hamilton-Jacobi-Isaacs (HJI) equation (10.2-13), (10.2-14). Then according to (10.2-15) one has

$$H(x, \nabla V, u, d) = -\gamma^2 \|d - d*\|^2 + (u - u^*)^{\mathrm{T}}R(u - u^*).$$

Therefore,

$$\frac{dV}{dt} + h^{\mathrm{T}}h + u^{\mathrm{T}}Ru - \gamma^2 d^{\mathrm{T}}d = -\gamma^2 \|d - d*\|^2 + (u - u^*)^{\mathrm{T}}R(u - u^*).$$

Selecting $u = u^*$ yields

$$\frac{dV}{dt} + h^{\mathrm{T}}h + u^{\mathrm{T}}Ru - \gamma^2 d^{\mathrm{T}}d \le 0 \qquad (10.3\text{-}3)$$

for all $d(t)$. To show asymptotic stability of the closed-loop system (10.3-2) set $d = 0$ and note that

$$\frac{dV}{dt} \le -(h^{\mathrm{T}}h + u^{\mathrm{T}}Ru) = -\|z\|^2.$$

LaSalle's extension now shows that the state goes to a region of R^n wherein $\dot{V} = 0$. However, zero-state observable means $u(t) \equiv 0, y(t) \equiv 0 \Rightarrow x(t) = 0$. Therefore, the system is locally asymptotically stable with Lyapunov function $V(x) > 0$.

Now, integrating (10.3-3) yields

$$V(x(T)) - V(x(0)) + \int_0^T \left(h^{\mathrm{T}}h + u^{\mathrm{T}}Ru - \gamma^2 d^{\mathrm{T}}d\right)dt \le 0.$$

Select $x(0) = 0$. Noting that $V(0) = 0$ and $V(x(T)) > 0$ one has

$$\int_0^T \left(h^{\mathrm{T}}h + u^{\mathrm{T}}Ru\right) dt \le \gamma^2 \int_0^T (d^{\mathrm{T}}d)\, dt,$$

so the L_2 gain is less than γ. ■

The disturbance $d^* = d(V^*) = (1/2\gamma^2)k^{\mathrm{T}}(x)\nabla V^*$ provided by the ZS game solution (10.2-12) is known as the worst-case disturbance.

Linear Quadratic Zero-sum Game

In the LQ ZS game one has linear dynamics

$$\dot{x} = Ax + Bu + Dd, \tag{10.3-4}$$

with $x(t) \in R^n$. The value is the integral quadratic form

$$V(x(t), u, d) = \frac{1}{2}\int_t^\infty \left(x^{\mathrm{T}}H^{\mathrm{T}}Hx + u^{\mathrm{T}}Ru - \gamma^2 \|d\|^2\right) d\tau \equiv \int_t^\infty r(x, u, d)d\tau, \tag{10.3-5}$$

where $R = R^{\mathrm{T}} > 0$, and $\gamma > 0$. Assume that the value is quadratic in the state so that

$$V(x) = \tfrac{1}{2}x^{\mathrm{T}}Sx \tag{10.3-6}$$

for some matrix $S > 0$. Select state feedbacks for the control and disturbance so that

$$u = -Kx \tag{10.3-7}$$

$$d = Lx. \tag{10.3-8}$$

Substituting this into Bellman equation (10.2-7) gives

$$S(A - BK + DL) + (A - BK + DL)^{\mathrm{T}}S + H^{\mathrm{T}}H + K^{\mathrm{T}}RK - \gamma^2 L^{\mathrm{T}}L. \tag{10.3-9}$$

It has been assumed that the Bellman equation holds for all initial conditions, and the state $x(t)$ has been canceled. This is a Lyapunov equation for S in terms of the prescribed SVFB policies K and L. If $(A - BK + DL)$ is stable, (A, H) is observable, and $\gamma > \gamma^* > 0$, then there is a positive definite solution $S \ge 0$, and then (10.3-6) is the value (10.3-5) for the selected feedback policies K and L.

The stationary point control (10.2-11) and disturbance (10.2-12) are given by

$$u(x) = -R^{-1}B^{\mathrm{T}}Sx = -Kx \tag{10.3-10}$$

$$d(x) = \frac{1}{\gamma^2}D^{\mathrm{T}}Sx = Lx. \tag{10.3-11}$$

Substituting these into (10.3-9) yields the game algebraic Riccati equation (GARE)

$$0 = A^{\mathrm{T}}S + SA + H^{\mathrm{T}}H - SBR^{-1}B^{\mathrm{T}}S + \frac{1}{\gamma^2}SDD^{\mathrm{T}}S. \tag{10.3-12}$$

To solve the ZS game problem, one solves the GARE equation for the optimal value kernel $S \geq 0$, then the optimal control is given as a state-variable feedback in terms of the ARE solution by (10.3-10) and the worst-case disturbance by (10.3-11). There exists a solution $S > 0$ if $(A,\ B)$ is stabilizable, $(A,\ \sqrt{Q})$ is observable, and $\gamma > \gamma^*$, the H-infinity gain. Since we have developed a solution to the problem, it is verified that assumption (10.3-6) holds.

There may more than one PSD solution to the GARE. To solve the ZS game problem, one requires gains such that the poles of $(A - BK + DL)$ are in the open left-half plane. The minimum PSD solution of the GARE is the unique stabilizing solution. It is shown in van der Schaft (1992), and Başar and Olsder (1999) that the stabilizing solution of the GARE corresponds to the stable eigenspace of a certain Hamiltonian matrix, similarly to the discussion presented in Section 3.4 for the stabilizing ARE solution. A method for finding the minimum PSD solution is given in (Abu Khalaf et al. 2006).

The work of Vamvoudakis and Lewis (2010b) shows how to solve the GARE online in real time using data measured along the system trajectories. This allows the performance index weights R, γ to change slowly as the game develops. Vrabie and Lewis (2010a) has provided algorithms for solving the HJB equations online in real time without knowing the system matrix A. These references extend the reinforcement learning techniques presented in Chapter 11 to online solution of LQ zero-sum games.

10.4 MULTIPLAYER NON-ZERO-SUM GAMES

In zero-sum games, whatever one player gains, the rest lose. Cooperation and competition in nature and in human enterprises are fascinating topics. Most often, there is a balance between cooperation toward team goals and competition among players to achieve individual goals. This interplay is very nicely captured by the machinery of non-zero-sum (NZS) multiplayer games (Başar and Olsder 1999).

Nonlinear NZS Games

Consider the nonlinear time-invariant dynamical system given by

$$\dot{x} = f(x) + \sum_{j=1}^{N} g_j(x)u_j, \tag{10.4-1}$$

with state $x(t) \in R^n$ and controls $u_j(t) \in R^{m_j}$. This system has N inputs or players, all of whom influence each other through their joint effects on the overall system state dynamics. Let $f(x)$ be locally Lipschitz, $f(0) = 0$, and $g_j(x)$ be continuous. Define the performance index of player i as

$$J_i(x(0), u_1, u_2, \ldots, u_N) = \int_0^\infty (Q_i(x) + \sum_{j=1}^N u_j^{\mathrm{T}} R_{ij} u_j)\, dt$$

$$\equiv \int_0^\infty r_i(x(t), u_1, u_2, \ldots, u_N) dt, i \in N, \qquad (10.4\text{-}2)$$

where function $Q_i(x) \geq 0$ is generally nonlinear, and $R_{ii} > 0$, $R_{ij} \geq 0$ are symmetric matrices. The notation $i \in N$ means $i = 1, \ldots, N$.

Define the multiplayer differential game

$$V_i^*(x(t), u_1, u_2, \ldots, u_N) = \min_{u_i} \int_t^\infty (Q_i(x) + \sum_{j=1}^N u_j^{\mathrm{T}} R_{ij} u_j) d\tau, \quad \forall i \in N. \qquad (10.4\text{-}3)$$

This game implies that all the players have the same competitive hierarchical level and seek to attain a Nash equilibrium as given by the following definition.

Definition 10.4-1. Nash equilibrium.
Policies $\left\{u_1^*(x), u_2^*(x), \ldots, u_{\mathrm{N}}^*(x)\right\}$ are said to constitute a Nash equilibrium solution for the *N-player* game if

$$J_i^* \triangleq J_i(u_1^*, u_2^*, u_i^*, ..., u_{\mathrm{N}}^*) \leq J_1(u_1^*, u_2^*, u_i, ..., u_{\mathrm{N}}^*), \quad \forall u_i, \forall i \in N. \quad (10.4\text{-}4)$$

The *N-tuple* $\left\{J_1^*, J_2^*, ..., J_{\mathrm{N}}^*\right\}$is known as a Nash equilibrium value set or outcome of the N-player game.

The implication of this definition is that if any player unilaterally changes his control policy while the policies of all other players remain the same, then that player will obtain worse performance. For fixed stabilizing feedback control policies $u_j(x)$define the value function

$$V_i(x(t)) = \int_t^\infty (Q_i(x) + \sum_{j=1}^N u_j^{\mathrm{T}} R_{ij} u_j) d\tau = \int_t^\infty r(x, u_1, \ldots, u_N) d\tau, i \in N, \qquad (10.4\text{-}5)$$

which is only a function of the initial state $x(t)$. When the value is finite, a differential equivalent to this is found by differentiating using Leibniz's formula.

This yields the nonlinear NZ*S game Bellman equations* given in terms of the Hamiltonian functions as

$$0 = Q_i(x) + \sum_{j=1}^{N} u_j^{\mathrm{T}} R_{ij} u_j + (\nabla V_i)^{\mathrm{T}}\Big(f(x) + \sum_{j=1}^{N} g_j(x) u_j\Big)$$

$$\equiv H_i(x, \nabla V_i, u_1, \ldots, u_N),\, i \in N. \tag{10.4-6}$$

The initial conditions for these partial differential equations are $V_i(0) = 0$.

At equilibrium, one has the stationarity conditions

$$\frac{\partial H_i}{\partial u_i} = 0, \quad i \in N, \tag{10.4-7}$$

which yield the policies

$$u_i(x) = u_i(V_i(x)) = -\tfrac{1}{2} R_{ii}^{-1} g_i^{\mathrm{T}}(x) \nabla V_i, \quad i \in N. \tag{10.4-8}$$

Note that $\partial^2 H_i / \partial^2 u_i = 2R_{ii} > 0$, so that at the stationary point the Hamiltonian H_i attains a minimum in control policy $u_i(x)$. Using these control policies the closed-loop system is

$$\dot{x} = f(x) - \frac{1}{2} \sum_{j=1}^{N} g_j(x) R_{jj}^{-1} g_j^{\mathrm{T}}(x) \nabla V_j. \tag{10.4-9}$$

Substituting (10.4-8) into (10.4-6) one obtains the N-coupled Hamilton-Jacobi (HJ) equations

$$0 = (\nabla V_i)^{\mathrm{T}} \left(f(x) - \frac{1}{2} \sum_{j=1}^{N} g_j(x) R_{jj}^{-1} g_j^{\mathrm{T}}(x) \nabla V_j \right)$$

$$+ Q_i(x) + \frac{1}{4} \sum_{j=1}^{N} \nabla V_j^{\mathrm{T}} g_j(x) R_{jj}^{-\mathrm{T}} R_{ij} R_{jj}^{-1} g_j^{\mathrm{T}}(x) \nabla V_j, \tag{10.4-10}$$

with $V_i(0) = 0$. These coupled HJ equations are in closed-loop form. The equivalent open-loop form is

$$0 = \nabla V_i^{\mathrm{T}} f(x) + Q_i(x) - \frac{1}{2} \nabla V_i^{\mathrm{T}} \sum_{j=1}^{N} g_j(x) R_{jj}^{-1} g_j^{\mathrm{T}}(x) \nabla V_j$$

$$+ \frac{1}{4} \sum_{j=1}^{N} \nabla V_j^{\mathrm{T}} g_j(x) R_{jj}^{-\mathrm{T}} R_{ij} R_{jj}^{-1} g_j^{\mathrm{T}}(x) \nabla V_j. \tag{10.4-11}$$

These equations can be written as

$$H_i(x, \nabla V_i^*, u_1^*, \dots, u_N^*) = 0, \tag{10.4-12}$$

where $u_i^*(x) = u_i(V_i^*(x))$ as given by (10.4-8).

The next key result shows that the Hamiltonian function has a specific dependence on control deviations from certain key values.

Lemma 10.4-2. For any control policies $u_i(x), i = 1, N$ that yield finite values (10.4-5), let $V_i(x) \geq 0$ be the corresponding solutions to the Bellman equations (10.4-6). Define $u_i^*(x) = u_i(V_i(x))$ according to (10.4-8) in terms of $V_i(x)$. Then

$$\begin{aligned} &H_i(x, \nabla V_i, u_1, \dots, u_N x) \\ &\quad = H_i(x, \nabla V_i, u_1^*, \dots, u_N^*) + \sum_j (u_j - u_j^*)^{\mathrm{T}} R_{ij}(u_j - u_j^*) \\ &\qquad + \nabla V_i^{\mathrm{T}} \sum_j g_j(u_j - u_j^*) + 2\sum_j (u_j^*)^{\mathrm{T}} R_{ij}(u_j - u_j^*). \end{aligned} \tag{10.4-13}$$

■

Exercise 10.4-1.

Prove Lemma 10.4-1 by completing the squares on (10.4-6). This is simplified if one writes

$$H_i(x, \nabla V_i, u_1^*, \dots, u_N^*) = \nabla V_i^{\mathrm{T}} f(x) + Q_i(x) + \nabla V_i^{\mathrm{T}} \sum_{j=1}^{N} g_j(x)u_j^* + \sum_{j=1}^{N} (u_j^*)^{\mathrm{T}} R_{ij} u_j^*. \tag{10.4-14}$$

The next result shows when the coupled HJ solutions solve the multiplayer game (Basar and Olsder 1999). ■

Theorem 10.4-2. Stability and Solution of NZS Game Nash Equilibrium. Let $Q_i(x) > 0$ in the performance index (10.4-2). Let $V_i(x) > 0 \in C^1, i = 1, N$ be smooth solutions to HJ equations (10.4-11), and control policies $u_i^*, i \in N$ be given by (10.4-8) in terms of these solutions V_i. Then

a. System (10.4-9) is asymptotically stable and $V_i(x)$ serve as Lyapunov functions.
b. $\{u_i^*, i = 1, N\}$ are in Nash equilibrium and the corresponding game values are

$$J_i^*(x(0)) = V_i, i \in N. \tag{10.4-15}$$

Proof:

a. If $V_i > 0$ satisfies (10.4-11) then it also satisfies (10.4-6). Take the time derivative to obtain

$$\dot{V}_i = \nabla V_i^{\mathrm{T}} \dot{x} = \nabla V_i^{\mathrm{T}} \left(f(x) + \sum_j g_j(x) u_j \right)$$
$$= -\frac{1}{2} \left(Q_i(x) + \sum_j u_j^{\mathrm{T}} R_{ij} u_j \right), \tag{10.4-16}$$

which is negative definite since $Q_i > 0$. Therefore, $V_i(x)$ is a Lyapunov function for x and systems (10.4-9) are asymptotically stable.

b. According to part a, $x(t) \to 0$ for the selected control policies. Define u_{-i} as the set of control policies of all nodes other than node i. For any smooth functions $V_i(x), i \in N$, such that $V_i(0) = 0$, by setting $V_i(x(\infty)) = 0$ one can write (10.4-2) as

$$J_i(x(0), u_i, u_{-i}) = \frac{1}{2} \int_0^\infty (Q_i(x) + \sum_j u_j^{\mathrm{T}} R_{ij} u_j)\, dt + \int_0^\infty \dot{V}_i dt + V_i(x(0))$$

or

$$J_i(x(0), u_i, u_{-i}) = \frac{1}{2} \int_0^\infty (Q_i(x) + \sum_j u_j^{\mathrm{T}} R_{ij} u_j)\, dt$$
$$+ \int_0^\infty \nabla V_i^{\mathrm{T}} (f + \sum_j g_j u_j)\, dt + V_i(x(0))$$
$$J_i(x(0), u_i, u_{-i}) = \frac{1}{2} \int_0^\infty H_i(x, \nabla V_i, u_1, \ldots, u_N) + V_i(x(0)).$$

Now let V_i satisfy (10.4-11) and u_i^*, u_{-i}^* be the optimal controls given by (10.4-8). By Lemma 10.4-1 one has

$$J_i(x(0), u_i, u_{-i}) = V_i(x(0)) + \int_0^\infty (\sum_j (u_j - u_j^*)^{\mathrm{T}} R_{ij} (u_j - u_j^*)^{\mathrm{T}}$$
$$- \nabla V_i^{\mathrm{T}} \sum_j B_j (u_j - u_j^*) + \sum_j u_j^{*\mathrm{T}} R_{ij} (u_j - u_j^*))\, dt \tag{10.4-17}$$

$$
\begin{aligned}
J_i(x(0), u_i, u_{-i}) = V_i(x(0)) &+ \int_0^{\infty} (u_i - u_i^*)^{\mathrm{T}} R_{ii} (u_i - u_i^*)^{\mathrm{T}} dt \\
&+ \int_0^{\infty} \Big(\sum_{j \neq i} (u_j - u_j^*)^{\mathrm{T}} R_{ij} (u_j - u_j^*)^{\mathrm{T}} dt \\
&- \nabla V_i^{\mathrm{T}} \sum_j B_j (u_j - u_j^*) + \sum_j u_j^{*\mathrm{T}} R_{ij} (u_j - u_j^*)\Big)\, dt.
\end{aligned}
$$

At the equilibrium point $u_i = u_i^*$ and $u_j = u_j^*, \forall j$ so

$$J_i^*(x(0), u_i^*, u_{-i}^*) = V_i(x(0)).$$

Define

$$J_i(u_i, u_{-i}^*) = V_i(x(0)) + \frac{1}{2} \int_0^{\infty} (u_i - u_i^*)^{\mathrm{T}} R_{ii} (u_i - u_i^*)\, dt \tag{10.4-18}$$

and $J_i^* = V_i(x(0))$. Then clearly J_i^* and $J_i(u_i, u_{-i}^*)$ satisfy (10.4-4). ■

Note that Lemma 10.4-1 shows that the values are not quadratic in all the control policies. In the proof this carries over to the fact that the performance indices (10.4-17) are not quadratic in all the control policies. However, Definition 10.4-1 of Nash equilibrium requires that the optimal value J_i^* be quadratic in u_i when all other policies are held at Nash. This definition means that the nonquadratic terms in (10.4-13) are equal to zero at the equilibrium and do not cause trouble. Then, (10.4-18) is quadratic in u_i, as required.

Solution of the Coupled HJ Equations

The coupled HJ equations are difficult to solve for nonlinear systems and the solution to multiplayer games is generally obtained by solving these equations offline. This does not allow the objectives to change in real time as the players learn from each other. The work of Vamvoudakis (2011) shows how to solve the coupled HJ equations online in real time using data measured along the system trajectories. This allows the weight functions $Q_i(x)$, R_{ij} in (10.4-2) to change slowly as the game develops. Vrabie (2010) has provided algorithms for solving the coupled HJ equations online in real time without knowing the system drift dynamics $f(x)$. These references extend the reinforcement learning techniques presented in Chapter 11 to online solution of non-zero-sum games.

Cooperative and Competitive Aspects of NZS Games The multiplayer game formulation allows for considerable freedom of each agent within the framework of overall team Nash equilibrium. Each agent has a performance objective (10.4-2) that can embody team objectives as well as individual objectives. In fact, the performance objective of each agent can be written as

$$J_i = \frac{1}{N}\sum_{j=1}^{N} J_j + \frac{1}{N}\sum_{j=1}^{N}(J_i - J_j) \equiv J_{\text{team}} + J^i_{\text{conflict}}, \tag{10.4-19}$$

where J_{team} is the overall cooperative (center of gravity) performance objective of the networked team, and J^i_{conflict} is the conflict of interest or competitive objective. J_{team}measures how much the players are vested in common goals, and J^i_{conflict} expresses to what extent their objectives differ. The objective functions can be chosen by the individual players, or they may be assigned to yield some desired team behavior.

In the case of N-player zero-sum games, one has $J_{\text{team}} = 0$ and there are no common objectives. One case is the 2-player zero-sum game, where $J_2 = -J_1$, and one has $J_{\text{team}} = 0$ and $J^i_{\text{conflict}} = J_i$.

Linear Quadratic Multiplayer Games In linear systems of the form

$$\dot{x} = Ax + \sum_{j=1}^{N} B_j u_j, \tag{10.4-20}$$

with quadratic performance indices

$$J_i(x(0), u_1, u_2, \ldots, u_N) = \frac{1}{2}\int_0^\infty \left(x^T Q_i x + \sum_{j=1}^{N} u_j^{\mathrm{T}} R_{ij} u_j\right) dt, \tag{10.4-21}$$

the HJ equations (10.4-11) become the N-coupled generalized algebraic Riccati equations

$$0 = P_i A_c + A_c^{\mathrm{T}} P_i + Q_i + \sum_{j=1}^{N} P_j B_j R_{jj}^{-\mathrm{T}} R_{ij} R_{jj}^{-1} B_j^{\mathrm{T}} P_j, \quad i \in N, \tag{10.4-22}$$

where $A_c = A - \sum_{i=1}^{N} B_i R_{ii}^{-1} B_i^{\mathrm{T}} P_i$. It is shown in Başar and Olsder (1999) that if there exist solutions to (10.4.22) further satisfying the conditions that for each $i \in N$ the pairs

$$\left(A - \sum_{j\neq i} B_j R_{jj}^{-1} B_j^{\mathrm{T}} P_j, \quad B_i\right),$$

$$\left(A - \sum_{j\neq i} B_j R_{jj}^{-1} B_j^{\mathrm{T}} P_j, \sqrt{Q_i + \sum_{j\neq i} P_j B_j R_{jj}^{-\mathrm{T}} R_{ij} R_{jj}^{-1} B_j^{\mathrm{T}} P_j}\right)$$

are respectively stabilizable and detectable, then the N-tuple of stationary feedback policies $u_i^*(x) = -K_i x = -\frac{1}{2} R_{ii}^{-1} B_i^T P_i x, \quad i \in N$ provides a Nash equilibrium solution for the linear quadratic N-player differential game. Furthermore, the resulting system dynamics, described by $\dot{x} = A_c x$, are asymptotically stable.

Vamvoudakis and Lewis (2011) show how to solve the coupled HJ equations online in real time using data measured along the system trajectories. Vrabie and Lewis (2010a,b) have provided algorithms for solving these equations online in real time without knowing the system A matrix. These references use reinforcement learning techniques based on the development in Chapter 11.