AA203: Optimal and Learning-based Control Course Notes

James Harrison*

April 20, 2019

3 The HJB and HJI Equations

In this section, we will extend the ideas of dynamic programming to the continuous time setting. Restating the continuous time optimal control problem, we assume dynamics

$$\dot{\boldsymbol{x}}(t) = f(\boldsymbol{x}(t), \boldsymbol{u}(t), t) \tag{1}$$

and cost

$$J(\boldsymbol{x}(0)) = c_f(\boldsymbol{x}(t_f), t_f) + \int_0^{t_f} c(\boldsymbol{x}(\tau), \boldsymbol{u}(\tau), \tau) d\tau. \tag{2}$$

where t_f is fixed.

3.1 The Principle of Optimality in Continuous Time

3.1.1 Hamilton-Jacobi-Bellman

As in the discrete time principle of optimality, consider the tail problem

$$J(\boldsymbol{x}(t), \{\boldsymbol{u}(\tau)\}_{\tau=t}^{t_f}, t) = c_f(\boldsymbol{x}(t_f), t_f) + \int_t^{t_f} c(\boldsymbol{x}(\tau), \boldsymbol{u}(\tau), \tau) d\tau \tag{3}$$

where $t \leq t_f$ and $\boldsymbol{x}(t)$ is an admissible state value. The optimal solution to this tail problem comes from the functional minimization

$$J^*(\boldsymbol{x}(t), t) = \min_{\{\boldsymbol{u}(\tau)\}_{\tau=t}^{t_f}} \left\{ c_f(\boldsymbol{x}(t_f), t_f) + \int_t^{t_f} c(\boldsymbol{x}(\tau), \boldsymbol{u}(\tau), \tau) d\tau \right\}. \tag{4}$$

*Contact: jharrison@stanford.edu

Note, then, that due to the additivity of cost we can split the problem up over time,

$$J^*(\boldsymbol{x}(t),t) = \min_{\{\boldsymbol{u}(\tau)\}_{\tau=t}^{t_f}} \left\{ \int_t^{t+\Delta t} c(\boldsymbol{x}(\tau),\boldsymbol{u}(\tau),\tau)d\tau + c_f(\boldsymbol{x}(t_f),t_f) + \int_{t+\Delta t}^{t_f} c(\boldsymbol{x}(\tau),\boldsymbol{u}(\tau),\tau)d\tau \right\} \tag{5}$$

which by applying the principle of optimality to the tail cost,

$$J^*(\boldsymbol{x}(t),t) = \min_{\{\boldsymbol{u}(\tau)\}_{\tau=t}^{t+\Delta t}} \left\{ \int_t^{t+\Delta t} c(\boldsymbol{x}(\tau),\boldsymbol{u}(\tau),\tau)d\tau + J^*(\boldsymbol{x}(t+\Delta t),t+\Delta t) \right\}. \tag{6}$$

Let $J_t^*(\boldsymbol{x}(t),t) = \nabla_t J^*(\boldsymbol{x}(t),t)$ and $J_{\boldsymbol{x}}^*(\boldsymbol{x}(t),t) = \nabla_{\boldsymbol{x}} J^*(\boldsymbol{x}(t),t)$. Taylor expanding, we have

$$\begin{aligned} J^*(\boldsymbol{x}(t),t) = \min_{\{\boldsymbol{u}(\tau)\}_{\tau=t}^{t+\Delta t}} \{ & c(\boldsymbol{x}(t),\boldsymbol{u}(t),t)\Delta t + J^*(\boldsymbol{x}(t),t) + (J_t^*(\boldsymbol{x}(t),t))\Delta t \\ & + (J_{\boldsymbol{x}}^*(\boldsymbol{x}(t),t))^T(\boldsymbol{x}(t+\Delta t) - \boldsymbol{x}(t)) + o(\Delta t)\} \end{aligned} \tag{7}$$

for small Δt. The first term is a result of Taylor expanding the integral and applying the fundamental theorem of calculus. Note that we can pull $J^*(\boldsymbol{x}(t),t)$ out of the minimization over cost, as this quantity will not vary under different choices of future actions. Dividing through by Δt and taking the limit $\Delta t \to 0$, we obtain the *Hamilton-Jacobi-Bellman* equation

$$0 = J_t^*(\boldsymbol{x}(t),t) + \min_{\boldsymbol{u}(t)} \left\{ c(\boldsymbol{x}(t),\boldsymbol{u}(t),t) + (J^*(\boldsymbol{x}(t),t))^T f(\boldsymbol{x}(t),\boldsymbol{u}(t),t) \right\} \tag{8}$$

with terminal condition

$$J^*(\boldsymbol{x}(t_f),t_f) = c_f(\boldsymbol{x}(t_f),t_f). \tag{9}$$

For convenience, we will define the Hamiltonian

$$\mathcal{H}(\boldsymbol{x}(t),\boldsymbol{u}(t),J_{\boldsymbol{x}}^*,t) := c(\boldsymbol{x}(t),\boldsymbol{u}(t),t) + (J_{\boldsymbol{x}}^*(\boldsymbol{x}(t),t))^T f(\boldsymbol{x}(t),\boldsymbol{u}(t),t) \tag{10}$$

which allow us to compactly write the HJB equation as

$$0 = J_t^*(\boldsymbol{x}(t),t) + \min_{\boldsymbol{u}(t)} \left\{ \mathcal{H}(\boldsymbol{x}(t),\boldsymbol{u}(t),J_{\boldsymbol{x}}^*,t) \right\}. \tag{11}$$

The HJB equation is a partial differential equation that, for cost-to-go $J^*(\boldsymbol{x}(t),t)$, will satisfy all time-state pairs $(\boldsymbol{x}(t),t)$. The previous informal derivation assumed differentiability of $J^*(\boldsymbol{x}(t),t)$, which we do not know a priori. This assumption is rectified by the following theorem on solutions to the HJB equation.

Theorem 3.1 (Sufficiency Theorem)**.** *Suppose $V(\boldsymbol{x},t)$ is a solution to the HJB equation, that $V(\boldsymbol{x},t)$ is C^1 in $\boldsymbol{x}$ and t, and that*

$$\begin{aligned} 0 &= V_t(\boldsymbol{x},t) + \min_{\boldsymbol{u}\in\mathcal{U}} \left\{ c(\boldsymbol{x},\boldsymbol{u},t) + (V_{\boldsymbol{x}}(\boldsymbol{x},t))^T f(\boldsymbol{x},\boldsymbol{u},t) \right\} \\ V(\boldsymbol{x},t_f) &= c_f(\boldsymbol{x},t_f) \ \forall\, \boldsymbol{x} \end{aligned}$$

Suppose also that $\pi^*(\boldsymbol{x}, t)$ attains the minimum in this equation for all t and $\boldsymbol{x}$. Let $\{\boldsymbol{x}^*(t) \mid t \in [t_0, t_f]\}$ be the state trajectory obtained from the given initial condition $\boldsymbol{x}(0)$ when the control trajectory $\boldsymbol{u}^*(t) = \pi^*(\boldsymbol{x}^*(t), t), t \in [t_0, t_f]$ is used. Then V is equal to the optimal cost-to-go function, i.e.,

$$V(\boldsymbol{x}, t) = J^*(\boldsymbol{x}, t) \; \forall \, \boldsymbol{x}, t. \tag{12}$$

Furthermore, the control trajectory $\{\boldsymbol{u}^*(t) \mid t \in [t_0, t_f]\}$ is optimal..

Proof. [Ber12], Volume 1, Section 3.2. □

3.1.2 Continuous-Time LQR

As a useful result of the HJB equations, we will derive LQR in continuous time. We aim to minimize

$$J(\boldsymbol{x}(0)) = \frac{1}{2}\boldsymbol{x}^T(t_f)Q_f\boldsymbol{x}(t_f) + \frac{1}{2}\int_0^{t_f} \boldsymbol{x}^T(t)Q(t)\boldsymbol{x}(t) + \boldsymbol{u}^T(t)R(t)\boldsymbol{u}(t)dt \tag{13}$$

subject to dynamics

$$\dot{\boldsymbol{x}}(t) = A(t)\boldsymbol{x}(t) + B(t)\boldsymbol{u}(t). \tag{14}$$

As in discrete LQR, we will assume $Q_f, Q(t)$ are positive semidefinite, and $R(t)$ is positive definite. We will also assume t_f is fixed, and the state and action are unconstrained.

We will write the Hamiltonian,

$$\mathcal{H} = \frac{1}{2}\boldsymbol{x}^T(t)Q(t)\boldsymbol{x}(t) + \frac{1}{2}\boldsymbol{u}^T(t)R(t)\boldsymbol{u}(t) + J^*_{\boldsymbol{x}}(\boldsymbol{x}(t), t)^T(A(t)\boldsymbol{x}(t) + B(t)\boldsymbol{u}(t)) \tag{15}$$

which yeilds necessary optimality conditions

$$0 = \nabla_{\boldsymbol{u}}\mathcal{H} = R(t)\boldsymbol{u}(t) + B^T(t)J^*_{\boldsymbol{x}}(\boldsymbol{x}(t), t). \tag{16}$$

Since $\nabla^2_{\boldsymbol{uu}}\mathcal{H} = R(t) > 0$, the control that satisfies the necessary conditions is the global minimizer. Rearranging, we have

$$\boldsymbol{u}^*(t) = -R^{-1}(t)B^T(t)J^*_{\boldsymbol{x}}(\boldsymbol{x}(t), t) \tag{17}$$

which we can plug back into the Hamiltonian to yield

$$\begin{aligned}\mathcal{H} &= \frac{1}{2}\boldsymbol{x}^T(t)Q(t)\boldsymbol{x}(t) + \frac{1}{2}J^*_{\boldsymbol{x}}(\boldsymbol{x}(t), t)^T B(t)R^{-1}(t)B^T(t)J^*_{\boldsymbol{x}}(\boldsymbol{x}(t), t) \qquad (18)\\ &\quad + J^*_{\boldsymbol{x}}(\boldsymbol{x}(t), t)^T A(t)\boldsymbol{x}(t) - J^*_{\boldsymbol{x}}(\boldsymbol{x}(t), t)^T B(t)R^{-1}(t)B^T(t)J^*_{\boldsymbol{x}}(\boldsymbol{x}(t), t)\\ &= \frac{1}{2}\boldsymbol{x}^T(t)Q(t)\boldsymbol{x}(t) - \frac{1}{2}J^*_{\boldsymbol{x}}(\boldsymbol{x}(t), t)^T B(t)R^{-1}(t)B^T(t)J^*_{\boldsymbol{x}}(\boldsymbol{x}(t), t) + J^*_{\boldsymbol{x}}(\boldsymbol{x}(t), t)^T A(t)\boldsymbol{x}(t). \qquad (19)\end{aligned}$$

This gives the HJB equation

$$0 = J_t^*(\boldsymbol{x}(t), t) + \frac{1}{2}\boldsymbol{x}^T(t)Q(t)\boldsymbol{x}(t) - \frac{1}{2}J_{\boldsymbol{x}}^*(\boldsymbol{x}(t), t)^T B(t)R^{-1}(t)B^T(t)J_{\boldsymbol{x}}^*(\boldsymbol{x}(t), t) + J_{\boldsymbol{x}}^*(\boldsymbol{x}(t), t)^T A(t)\boldsymbol{x}(t) \tag{20}$$

with boundary condition

$$J^*(\boldsymbol{x}(t_f), t_f) = \frac{1}{2}\boldsymbol{x}^T(t_f)Q_f\boldsymbol{x}(t_f). \tag{21}$$

It may appear as if we are stuck here, as this form of the HJB doesn't immediately yield $J^*(\boldsymbol{x}(t), t)$. Armed with the knowledge that the discrete time LQR problem has a quadratic cost-to-go, we will cross our fingers and guess a solution of the form

$$J^*(\boldsymbol{x}(t), t) = \frac{1}{2}\boldsymbol{x}^T(t)V(t)\boldsymbol{x}(t). \tag{22}$$

Substituting, we have

$$0 = \frac{1}{2}\boldsymbol{x}^T(t)\dot{V}(t)\boldsymbol{x}(t) + \frac{1}{2}\boldsymbol{x}^T(t)Q(t)\boldsymbol{x}(t) - \frac{1}{2}\boldsymbol{x}^T(t)V(t)B(t)R^{-1}(t)B^T(t)V(t)\boldsymbol{x}(t) + \boldsymbol{x}^T(t)V(t)A(t)\boldsymbol{x}(t) \tag{23}$$

Note that we will decompose

$$\boldsymbol{x}^T(t)V(t)A(t)\boldsymbol{x}(t) = \frac{1}{2}\boldsymbol{x}^T(t)V(t)A(t)\boldsymbol{x}(t) + \frac{1}{2}\boldsymbol{x}^T(t)A^T(t)V(t)\boldsymbol{x}(t) \tag{24}$$

which yields

$$0 = \frac{1}{2}\boldsymbol{x}^T(t)\left(\dot{V}(t) + Q(t) - V(t)B(t)R^{-1}(t)B^T(t)V(t) + V(t)A(t) + A^T(t)V(t)\right)\boldsymbol{x}(t). \tag{25}$$

This equation must hold for all $\boldsymbol{x}(t)$, so

$$-\dot{V}(t) = Q(t) - V(t)B(t)R^{-1}(t)B^T(t)V(t) + V(t)A(t) + A^T(t)V(t) \tag{26}$$

with boundary condition $V(t_f) = Q_f$.

Therefore, the HJB PDE has been reduced to a set of matrix ordinary differential equations (the Riccati equation). This is integrated backwards in time to find the full control policy as a function of time. One we have found $V(t)$, the control policy is

$$\boldsymbol{u}^*(t) = -R^{-1}(t)B^T(t)V(t)\boldsymbol{x}(t). \tag{27}$$

Similarly to the discrete case, the feedback gains tend toward constant in the limit of the infinite horizon problem, under some technical assumptions.

3.2 Differential Games

We have so far addressed the case in which we aim to solve the optimal control problem for a single agent. We will now consider an adversarial game setting, in which there exists another player that aims to maximally harm the first agent. In particular, we will consider zero sum games in which the second agent aims to maximize the cost of the first agent. While the differential game setting is not restricted to this case — agents may have separate cost functions that partially interfere or aid each other — the zero-sum case lends itself to useful analytical tools.

3.2.1 Differential Games and Information Patterns

We consider the two player differential game with dynamics

$$\dot{\boldsymbol{x}}(t) = f(\boldsymbol{x}(t), \boldsymbol{u}(t), \boldsymbol{d}(t)) \tag{28}$$

where the first player takes action $\boldsymbol{u}(t)$ at time t, and the second player takes action $\boldsymbol{d}(t)$. The state $\boldsymbol{x}(t)$ is the joint state of both players. We write the cost as

$$J(\boldsymbol{x}(t)) = c_f(\boldsymbol{x}(0)) + \int_t^0 c(\boldsymbol{x}(\tau), \boldsymbol{u}(\tau), \boldsymbol{d}(\tau))d\tau \tag{29}$$

which the first agent aims to maximize, and the second agent aims to minimize.

To fully specify the differential game, we must specify what each agent knows, and when. This is referred to as the *information pattern* of the game. In addition to capturing the knowledge of the state available to each agent, the information pattern also captures the knowledge of each other agents' strategies available to each agent.

3.2.2 Hamilton-Jacobi-Isaacs

The key idea in building the multi-agent equivalent of the HJB equation will again be to apply the principle of optimality. We consider the information pattern in which the adversary has access to the instantaneous control action of the first agent, so the cost takes the form

$$J(\boldsymbol{x}(t), t) = \min_{\Gamma(\boldsymbol{u})(\cdot)} \max_{\boldsymbol{u}(\cdot)} \left\{ \int_t^0 c(\boldsymbol{x}(\tau), \boldsymbol{u}(\tau), \boldsymbol{d}(\tau))d\tau + c_f(\boldsymbol{x}(0)) \right\}. \tag{30}$$

Applying the dynamic programming principle, we have

$$J(\boldsymbol{x}(t), t) = \min_{\Gamma(\boldsymbol{u})(\cdot)} \max_{\boldsymbol{u}(\cdot)} \left\{ \int_t^{t+\Delta t} c(\boldsymbol{x}(\tau), \boldsymbol{u}(\tau), \boldsymbol{d}(\tau))d\tau + J(\boldsymbol{x}(t+\Delta t), t+\Delta t) \right\}. \tag{31}$$

We can take the same strategy as with the informal derivation of the HJB equation, and Taylor expand both terms to yield

$$\begin{aligned} J(\boldsymbol{x}(t), t) = \min_{\Gamma(\boldsymbol{u})(\cdot)} \max_{\boldsymbol{u}(\cdot)} \{ & c(\boldsymbol{x}(\tau), \boldsymbol{u}(\tau), \boldsymbol{d}(\tau))\Delta t + J(\boldsymbol{x}(t), t) \\ & + (J_{\boldsymbol{x}}(\boldsymbol{x}(t), t))^T f(\boldsymbol{x}(t), \boldsymbol{u}(t), \boldsymbol{d}(t))\Delta t + J_t(\boldsymbol{x}(t), t)\Delta t \}. \end{aligned} \tag{32}$$

Note that we are optimizing over instantaneous actions, and so we optimizing over finite dimensional quantities as opposed to functions. Dividing through by Δt and removing redundant terms, we get the *Hamilton-Jacobi-Isaacs* (HJI) equation

$$0 = J_t(\boldsymbol{x}, t) + \max_{\boldsymbol{u}} \min_{\boldsymbol{d}} \left\{ c(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{d}) + (J_{\boldsymbol{x}}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{d}))^T f(\boldsymbol{x}, \boldsymbol{d}, \boldsymbol{u}) \right\} \tag{33}$$

with boundary condition

$$J(\boldsymbol{x}, 0) = c_f(\boldsymbol{x}). \tag{34}$$

Note that we have switched the order of the min/max.

3.2.3 Reachability

Differential games have applications in multi-agent modeling (both in the context of autonomous systems engineering and, e.g., economics and operations research). One concrete application in engineering is reachability analysis. In this setting, an agent aims to compute the set of states in which there exists a policy that either avoids a target set or enters a target set, subject to adversarial disturbances. The former case, in which we would like to avoid a target set, is useful for safety verification. If we are able to, even in the worst case, guarantee e.g. collision avoidance, we have guarantees on safety (subject of course to our system assumptions). The latter case is useful for task satisfaction. For example, we would like a quadrotor to reach a set of safe hovering poses, even under adversarial disturbances. Finding the backward reachable set in this case would find all states such that there exists a policy that succeeds in reaching the target set.

More concretely, the first case aims to find a set

$$\mathcal{A}(t) = \{\bar{\boldsymbol{x}} : \exists \Gamma(\boldsymbol{u})(\cdot), \forall \boldsymbol{u}(\cdot), \dot{\boldsymbol{x}} = f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{d}), \boldsymbol{x}(t) = \bar{\boldsymbol{x}}, \boldsymbol{x}(0) \in \mathcal{T}\} \tag{35}$$

where $\mathcal{T}$ is the unsafe set which we aim to avoid. Breaking this down, $\mathcal{A}(t)$ is the set of states at time t such that there exists $\Gamma(\boldsymbol{u})$ that maps action $\boldsymbol{u}$ to a disturbance such that, following the dynamics induced by the disturbance and the action sequence, the state is in $\mathcal{T}$ at time 0 (note that we are considering $0 \leq t$).

The second case aims to find a set

$$\mathcal{R}(t) = \{\bar{st} : \forall \Gamma(\boldsymbol{u})(\cdot), \exists \boldsymbol{u}(\cdot), \dot{\boldsymbol{x}} = f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{d}), \boldsymbol{x}(t) = \bar{\boldsymbol{x}}, \boldsymbol{x}(0) \in \mathcal{T}\}, \tag{36}$$

where in this case $\mathcal{T}$ is the set that we wish to reach. In this setting, we wish to find all states that, no matter what strategy the disturbance takes, there exist control actions that can steer the system to the goal state. Because the disturbance is adversarial (we reason over all adversary strategies), this is an extremely conservative form of safety analysis.

Computation of the backward reachable set results from solving a differential *game of kind* in which the outcome is Boolean (i.e. whether or not $\boldsymbol{x}(0) \in \mathcal{T}$). This boolean outcome can be encoded by removing the running cost and choosing a particular form for the final cost. In particular, we can choose a final cost where

$$\boldsymbol{x} \in \mathcal{T} \iff c_f(\boldsymbol{x}) \leq 0. \tag{37}$$

As a result, the agent should aim to maximize c_f to avoid $\mathcal{T}$, whereas the disturbance should aim to minimize it. The two settings then take the following forms:

- Set avoidance: $J(\boldsymbol{x}, t) = \min_{\Gamma(\boldsymbol{u})} \max_{\boldsymbol{u}} c_f(\boldsymbol{x}(0))$
- Set reaching: $J(\boldsymbol{x}, t) = \max_{\Gamma(\boldsymbol{u})} \min_{\boldsymbol{u}} c_f(\boldsymbol{x}(0))$

Sets vs. Tubes. We have so far considered avoidance or reachability problems for which we care about set membership at time $t = 0$. However, for something like collision avoidance, we would like to stay collision free at every time as opposed to a particular time. *Backward reachable sets* capture the case in which only the final time set membership matters, and states for times $t < 0$ do not matter. *Backward reachable tubes* capture the entire time duration of the problem. Any state that passes through the target at any time in the problem duration is included. This yields a modified value function of the form

$$J(\boldsymbol{x}, t) = \min_{\Gamma(\boldsymbol{u})} \max_{\boldsymbol{u}} \min_{\tau \in [t,0]} c_f(\boldsymbol{x}(\tau)). \tag{38}$$

If the target set membership holds at any time τ', then $\min_{\tau \in [t,0]} c_f(\boldsymbol{x}(\tau)) \leq c_f(\boldsymbol{x}(\tau')) \leq 0$.

3.3 Further Reading

Our coverage of reachability analysis is based on the [MBT05], which is an important early work in the field, in addition to being a relatively comprehensive coverage of the method. For a review of differential games with a (slight) emphasis on economics and management science, we refer the reader to [Bre10]. For a review of HJB and continuous time LQR, we refer the reader to [Ber12] and [Kir12].

References

[Ber12] Dimitri P Bertsekas. *Dynamic programming and optimal control*. Number 1. 4 edition, 2012.

[Bre10] Alberto Bressan. Noncooperative differential games. a tutorial. 2010.

[Kir12] Donald E Kirk. *Optimal control theory: an introduction*. Courier Corporation, 2012.

[MBT05] Ian M Mitchell, Alexandre M Bayen, and Claire J Tomlin. A time-dependent hamilton-jacobi formulation of reachable sets for continuous dynamic games. *IEEE Transactions on Automatic Control*, 2005.