

1 Problem Formulation

Assume we are operating in discrete time with the following linear system:

$$x_{t+1} = A_t x_t + B_t^1 u_t^1 + B_t^2 u_t^2. \quad (1)$$

Furthermore, let us assume a feedback information structure where u_t^{i*} depends on t and x_t . This assumption is an oversimplification for many interesting problems (consider any system without full state observability), but it is frequently closer to reality than an open-loop information structure.

We have two players attempting to minimize quadratic costs over some horizon T . Let us assume that player 1 is able to predict what player 2 will do at time t in response to the input u_t^1 . This implies that when we optimize for player 1's objective, we assume that player 2's input $u_t^2(u_t^1)$ is a function of u_t^1 . For player 2's objective, we treat u_t^2 as a variable.

$$J^1 = \frac{1}{2} \sum_{t=1}^T x_t^\top Q_t^1 x_t + u_t^{1\top} R_t^{11} u_t^1 + u_t^{2\top}(u_t^1) R_t^{12} u_t^2(u_t^1) \quad (2)$$

$$J^2 = \frac{1}{2} \sum_{t=1}^T x_t^\top Q_t^2 x_t + u_t^{2\top} R_t^{22} u_t^2 + u_t^{1\top} R_t^{21} u_t^1 \quad (3)$$

TODO: Add the optimization problem for each player.

2 Solution

We are looking for a *feedback* solution, i.e. a set of functions γ_t^i such that $u_t^i \equiv \gamma_t^i(x_t)$.

TODO: To generalize, note that the leader and follower cases are different and proceed from there.

If every player is operating at Stackelberg, their value functions and associated optimal controls should be mutually consistent. Furthermore, we can infer that each value function V_t^i is quadratic because each is a sum of quadratic functions. Therefore, we see that the optimal controls for each player, at each time, will be an affine function of the state. Let us presuppose the form of the value function and derive the form of the control law and thereby the specific value functions themselves. Our derivation will be constructive in nature and we begin by supposing the value functions have quadratic forms, i.e.

$$V_t^i(x_t) = \frac{1}{2} x_t^\top L_t^i x_t, \quad (4)$$

where $L_t^i > 0, L_t^i = L_t^{i\top}, L_{T+1}^i = 0$.

Furthermore, we will soon see that the optimal choice for u_t^i will be a linear function of the state x_t , i.e.

$$u_t^{i*}(x_t) = -S_t^i x_t, \quad (5)$$

where $S_t^i > 0, S_t^i = S_t^{i\top}$

From here, we solve for the four unknowns $S_t^1, S_t^2, L_t^1, L_t^2$ to obtain recursions that let us solve this dynamic programming problem.

2.1 Solving for Player 2's Unknowns

We start by writing down the Hamilton-Jacobi equations for player 2's value function:

$$V_t^2(x_t) = \min_{u_t^2} \left\{ \frac{1}{2} \left[x_t^\top Q_t^2 x_t + u_t^{2\top} R_t^{22} u_t^2 + u_t^{1\top} R_t^{21} u_t^1 \right] + V_{t+1}^2(x_{t+1}) \right\}; \quad u_t^1 \text{ given} \quad (6)$$

$$= \min_{u_t^2} \left\{ \frac{1}{2} \left[x_t^\top Q_t^2 x_t + u_t^{2\top} R_t^{22} u_t^2 + u_t^{1\top} R_t^{21} u_t^1 \right] + V_{t+1}^2(A_t x_t + B_t^1 u_t^1 + B_t^2 u_t^2) \right\}; \quad u_t^1 \text{ given}, \quad (7)$$

with final value $V_{T+1}^i(x_{T+1}) = 0$.

At any time t , we can plug (4) into (7):

$$V_t^2(x_t) = \frac{1}{2} \min_{u_t^2} \left\{ x_t^\top Q_t^2 x_t + u_t^{2\top} R_t^{22} u_t^2 + u_t^{1\top} R_t^{21} u_t^1 + (A_t x_t + B_t^1 u_t^1 + B_t^2 u_t^2)^\top L_{t+1}^i (A_t x_t + B_t^1 u_t^1 + B_t^2 u_t^2) \right\} \quad (8)$$

We can find the feedback control law by finding the minimizer of (8). To do so, we assume strong convexity and set the gradient to 0 as follows:

$$0 = R_t^{22} u_t^{2*} + B_t^{2\top} L_{t+1}^2 (A_t x_t + B_t^1 u_t^1 + B_t^2 u_t^{2*}). \quad (9)$$

We use the symmetry of L_{t+1}^2 to rearrange terms in this expression. Note that we can remove $\min_{u_t^2}$ since we already satisfy it through the first-order conditions. We take note of one more relationship before proceeding - given the control law we assume in (5), we see that

$$x_{t+1} = A_t x_t + B_t^1 u_t^{1*} + B_t^2 u_t^{2*} = (A_t - B_t^1 S_t^1 - B_t^2 S_t^2) x_t \quad (10)$$

Plugging in equations (5) and (10) to (9), we then rewrite it in the form:

$$u_t^{2*} = - \left[R_t^{22} + B_t^{2\top} L_{t+1}^2 B_t^2 \right]^{-1} B_t^{2\top} L_{t+1}^2 (A_t - B_t^1 S_t^1) x_t, \quad (11)$$

which fits the form of the control law proposed in (5) and from which we can extract S_t^2 . Thus,

$$S_t^2 = \left[R_t^{22} + B_t^{2\top} L_{t+1}^2 B_t^2 \right]^{-1} B_t^{2\top} L_{t+1}^2 (A_t - B_t^1 S_t^1). \quad (12)$$

To identify L_t^2 , we then plug $u_t^{i*} = -S_t^i x_t$ into V_t^2 to put the expression in the form we propose in (4) and extract L_t^2 :

$$\begin{aligned} V_t^2(x_t) &= \frac{1}{2} \left[x_t^\top Q_t^2 x_t + x_t^\top S_t^{2\top} R_t^{22} S_t^2 x_t + x_t^\top S_t^{1\top} R_t^{21} S_t^1 x_t + x_t^\top (A_t - B_t^1 S_t^1 - B_t^2 S_t^2)^\top L_{t+1}^2 (A_t - B_t^1 S_t^1 - B_t^2 S_t^2) x_t \right] \\ &= \frac{1}{2} x_t^\top \left[Q_t^2 + S_t^{2\top} R_t^{22} S_t^2 + S_t^{1\top} R_t^{21} S_t^1 + (A_t - B_t^1 S_t^1 - B_t^2 S_t^2)^\top L_{t+1}^2 (A_t - B_t^1 S_t^1 - B_t^2 S_t^2) \right] x_t. \end{aligned}$$

Thus, we see that

$$L_t^2 = Q_t^2 + S_t^{2\top} R_t^{22} S_t^2 + S_t^{1\top} R_t^{21} S_t^1 + (A_t - B_t^1 S_t^1 - B_t^2 S_t^2)^\top L_{t+1}^2 (A_t - B_t^1 S_t^1 - B_t^2 S_t^2), \quad (13)$$

where $L_{T+1}^i = Q_{T+1}^i$.

2.2 Solving for Player 1's Unknowns

Recall for player 1's value function, player 2's input $u_t^2(u_t^1)$ is a function of u_t^1 . We can repeat the same process as in Section 2.1, with a change in the derivation of the recursion based on our assumptions about the function $u_t^2(u_t^1)$. We start by writing down the Hamilton-Jacobi equations for player 1's value function:

$$V_t^1(x_t) = \min_{u_t^1} \left\{ \frac{1}{2} \left[x_t^\top Q_t^1 x_t + u_t^{1\top} R_t^{11} u_t^1 + u_t^{2\top}(u_t^1) R_t^{12} u_t^2(u_t^1) \right] + V_{t+1}^1(x_{t+1}) \right\} \quad (14)$$

$$= \min_{u_t^1} \left\{ \frac{1}{2} \left[x_t^\top Q_t^1 x_t + u_t^{1\top} R_t^{11} u_t^1 + u_t^{2\top}(u_t^1) R_t^{12} u_t^2(u_t^1) \right] + V_{t+1}^1(A_t x_t + B_t^1 u_t^1 + B_t^2 u_t^2) \right\} \quad (15)$$

with final value $V_{T+1}^1(x_{T+1}) = 0$.

When we, under the same assumptions of strong convexity and set the gradient to 0, we find that we must apply the chain rule to compute the derivative because of the dependence of u_t^2 on u_t^1 . We can now differentiate and solve for the minimum cost control u_t^{1*} as we did for u_t^{2*} in (9):

$$0 = \frac{\partial V_t^1}{\partial u_t^1} = R_t^{11} u_t^{1*} + u_t^{2\top}(u_t^{1*}) R_t^{12} \frac{\partial u_t^2}{\partial u_t^1} + \left(B_t^1 + B_t^2 \frac{\partial u_t^2}{\partial u_t^1} \right)^\top L_{t+1}^1 (A_t x_t + B_t^1 u_t^{1*} + B_t^2 u_t^2(u_t^1)) \quad (16)$$

We will need the derivative of $u_t^{2*}(u_t^1)$ with respect to u_t^1 later, so we compute that here using the chain rule:

$$\frac{\partial u_t^2}{\partial u_t^1} = - \left[R_t^{22} + B_t^{2\top} L_{t+1}^2 B_t^2 \right]^{-1} B_t^{2\top} L_{t+1}^2 B_t^1. \quad (17)$$

Let us next define a variable

$$D_t = \left[R_t^{22} + B_t^{2\top} L_{t+1}^2 B_t^2 \right]^{-1} B_t^{2\top} L_{t+1}^2. \quad (18)$$

Then, for the sake of easier substitution, we can rewrite (11) as

$$u_t^{2*} = -D_t A_t x_t + \frac{\partial u_t^2}{\partial u_t^1} u_t^1 \quad (19)$$

and (17) as

$$\frac{\partial u_t^2}{\partial u_t^1} = -D_t B_t^1 \quad (20)$$

Finally, we can plug in $u_t^2(u_t^1)$ (19) and $\frac{\partial u_t^2}{\partial u_t^1}$ (20) into (16) and solve for u_t^{1*} , which will give us S_t^1 .

$$u_t^{1*} = -S_t^1 x_t = - \left[R_t^{11} + B_t^{1\top} D_t^T R_t^{12} D_t B_t^1 + (B_t^1 - B_t^2 D_t B_t^1)^\top L_{t+1}^1 (B_t^1 - B_t^2 D_t B_t^1) \right]^{-1} \cdot \left[B_t^{1\top} D_t^T R_t^{12} D_t + (B_t^1 - B_t^2 D_t B_t^1)^\top L_{t+1}^1 (I - D_t) \right] A_t x_t \quad (21)$$

Finally, we can plug in our substitution (18) to get the full expression for S_t^1 :

$$S_t^1 = \left[R_t^{11} + B_t^{1\top} D_t^T R_t^{12} D_t B_t^1 + (B_t^1 - B_t^2 D_t B_t^1)^\top L_{t+1}^1 (B_t^1 - B_t^2 D_t B_t^1) \right]^{-1} \quad (22)$$

$$\cdot \left[B_t^{1\top} D_t^T R_t^{12} D_t + (B_t^1 - B_t^2 D_t B_t^1)^\top L_{t+1}^1 (I - D_t) \right] A_t$$

Finally, we can solve for L_t^1 in the same manner as L_t^2 , and we find that the recurrence takes the same form.

$$L_t^1 = Q_t^1 + S_t^{1\top} R_t^{11} S_t^1 + S_t^{2\top} R_t^{12} S_t^2 + (A_t - B_t^1 S_t^1 - B_t^2 S_t^2)^\top L_{t+1}^1 (A_t - B_t^1 S_t^1 - B_t^2 S_t^2) \quad (23)$$

2.3 Recursions

$$D_t = \left[R_t^{22} + B_t^{2\top} L_{t+1}^2 B_t^2 \right]^{-1} B_t^{2\top} L_{t+1}^2 \quad (24)$$

$$S_t^1 = \left[R_t^{11} + B_t^{1\top} D_t^T R_t^{12} D_t B_t^1 + (B_t^1 - B_t^2 D_t B_t^1)^\top L_{t+1}^1 (B_t^1 - B_t^2 D_t B_t^1) \right]^{-1} \cdot \left[B_t^{1\top} D_t^T R_t^{12} D_t + (B_t^1 - B_t^2 D_t B_t^1)^\top L_{t+1}^1 (I - D_t) \right] A_t \quad (25)$$

$$S_t^2 = \left[R_t^{22} + B_t^{2\top} L_{t+1}^2 B_t^2 \right]^{-1} B_t^{2\top} L_{t+1}^2 (A_t - B_t^1 S_t^1) \quad (26)$$

$$L_t^i = Q_t^i + S_t^{i\top} R_t^{ii} S_t^i + S_t^{j\top} R_t^{ij} S_t^j + (A_t - B_t^1 S_t^1 - B_t^2 S_t^2)^\top L_{t+1}^i (A_t - B_t^1 S_t^1 - B_t^2 S_t^2) \quad (i, j \in [1, 2], i \neq j) \quad (27)$$

This concludes our derivation by construction.

3 Extension to Nonlinear Systems

TODO: Describe the iterative solver here.