

# LQGame, ILQGame Derivation

Mohamed Safwat

December 10, 2022

## 1 LQGame Derivation using value iteration

Suppose we have an  $N$  person discrete linear dynamic game:

$$x_{t+1} = f(x_t, u_t^1, u_t^2, \dots, u_t^N) = A_t x_t + \sum_{j=1}^N B_t^j u_t^j \quad (1)$$

Where  $x_t$  is the joint state of all agents, and  $u_t^i$  is the control input of each agent  $i$ .  
A quadratic cost function for agent  $i$  given by:

$$c_i(x_t, u_t^1, \dots, u_t^N) = x_t^\top Q_t^i x_t + \sum_{j=1}^N u_t^{j\top} R_t^{ij} u_t^j \quad (2)$$

Where  $Q_t$  penalizes state error and  $R_t$  penalizes actuation effort. Note that without loss of generality,  $R_t^{ij}$  is included and can be interpreted as the penalty of player  $j$ 's control onto player  $i$ 's cost function.

We can then formulate the Bellman Equation for player  $i$  as the current cost + cost to go,

$$J_i^*(x, t) = \min_{u_i} [c_i(x_t, u_t) + J^*(x_{t+1}, t+1)] \quad (3)$$

We then solve it in a backward induction fashion since we know the value function at the terminal state, which is given as,

$$J_i^*(x_{T-1}, T-1) = \min_{u_i} [c_i(x_{T-1}, u_{T-1})] \quad (4)$$

At the terminal state, the cost is only associated with the state (want to drive the actuation to 0),

$$J_i^*(x_{T-1}, T-1) = x_{T-1}^\top V_{T-1}^i x_{T-1} \quad (5)$$

Where the value matrix is defined as:

$$V_{T-1}^i = Q_{T-1}^i$$

At the second to last time step,  $T-2$ , the value function is,

$$J_i^*(x_{T-2}, T-2) = \min_{u_i} [c_i(x_{T-2}, u_{T-2}) + J_i^*(x_{T-1}, T-1)] \quad (6)$$

Substituting in the quadratic cost function

$$J_i^*(x_{T-2}, T-2) = \min_{u_i} [x_{T-2}^\top Q_{T-2}^i x_{T-2} + \sum_{j=1}^N u_{T-2}^{j\top} R_{T-2}^{ij} u_{T-2}^j + x_{T-1}^\top V_{T-1}^i x_{T-1}] \quad (7)$$

Substituting in the linear dynamics and setting  $t = T-2$  for notation simplicity,

$$J_i^*(x_t, t) = \min_{u_i} [x_t^\top Q_t^i x_t + \sum_{j=1}^N u_t^{j\top} R_t^{ij} u_t^j + (A_t x_t + \sum_{j=1}^N B_t^j u_t^j)^\top V_{t+1}^i (A_t x_t + \sum_{j=1}^N B_t^j u_t^j)] \quad (8)$$

We then take the derivative of the value function wrt the control input of player  $i$ , set to zero, and solve for the optimal control,

$$\frac{\partial}{\partial u^i} J_i^*(x_t, t) = 0 \quad (9)$$

$$0 = 2R_t^{ii}u_t^i + 2B_t^{i\top}V_{t+1}^iB_t^iu_t^i + 2B_t^{i\top}V_{t+1}^iA_tx_t + 2B_t^{i\top}V_{t+1}^i\sum_{j=1}^NB_t^ju_t^j \quad (10)$$

Simplifying and rearranging,

$$(R_t^{ii} + B_t^{i\top}V_{t+1}^iB_t^i)u_t^i + B_t^{i\top}V_{t+1}^i\sum_{j=1}^NB_t^ju_t^j = -B_t^{i\top}V_{t+1}^iA_tx_t \quad (11)$$

Notice the form of the control input that minimizes the objective, which is a linear function of the state,

$$u_t^{i*} = -P_t^i x_t \quad (12)$$

Where  $P_t^i$  is analogous to an optimal proportional gain

Substituting equation 12 into 11:

$$(R_t^{ii} + B_t^{i\top}V_{t+1}^iB_t^i)P_t^i x_t + B_t^{i\top}V_{t+1}^i\sum_{j=1}^NB_t^jP_t^j x_t = B_t^{i\top}V_{t+1}^iA_tx_t \quad (13)$$

Crossing out the  $x_t$ , we obtain the system of linear equations,

$$(R_t^{ii} + B_t^{i\top}V_{t+1}^iB_t^i)P_t^i + B_t^{i\top}V_{t+1}^i\sum_{j=1}^NB_t^jP_t^j = B_t^{i\top}V_{t+1}^iA_t \quad (14)$$

For instance, in an  $N$  player game, equation 14 can be written in matrix form as:

$$\begin{bmatrix} R_t^{11} + B_t^{1\top}V_{t+1}^1B_t^1 & B_t^{1\top}V_{t+1}^1B_t^2 & \cdots & B_t^{1\top}V_{t+1}^1B_t^N \\ B_t^{2\top}V_{t+1}^2B_t^1 & R_t^{22} + B_t^{2\top}V_{t+1}^2B_t^2 & \cdots & B_t^{2\top}V_{t+1}^2B_t^N \\ \vdots & \vdots & \ddots & \vdots \\ B_t^{N\top}V_{t+1}^NB_t^1 & B_t^{N\top}V_{t+1}^NB_t^2 & \cdots & R_t^{NN} + B_t^{N\top}V_{t+1}^NB_t^N \end{bmatrix} \begin{bmatrix} P_t^1 \\ P_t^2 \\ \vdots \\ P_t^N \end{bmatrix} = \begin{bmatrix} B_t^{1\top}V_{t+1}^1A_t \\ B_t^{2\top}V_{t+1}^2A_t \\ \vdots \\ B_t^{N\top}V_{t+1}^NA_t \end{bmatrix}$$

To solve for  $P_t^i$ , we substitute the optimal control input,  $u_t^{i*}$ , into the value function (equation 8) to obtain a recursive relation for the value matrix  $V_t$ ,

$$J_i^*(x_t, t) = x_t^\top Q_t^i x_t + \sum_{j=1}^N (P_t^j x_t)^\top R_t^{ij} (P_t^j x_t) + (A_t x_t - \sum_{j=1}^N B_t^j (P_t^j x_t))^\top V_{t+1}^i (A_t x_t - \sum_{j=1}^N B_t^j (P_t^j x_t)) \quad (15)$$

$$J_i^*(x_t, t) = x_t^\top (Q_t^i + \sum_{j=1}^N P_t^{j\top} R_t^{ij} P_t^j + (A_t - \sum_{j=1}^N B_t^j P_t^j)^\top V_{t+1}^i (A_t - \sum_{j=1}^N B_t^j P_t^j) x_t \quad (16)$$

$$J_i^*(x_t, t) = x_t^\top V_t^i x_t \quad (17)$$

The value matrix is updated recursively as,

$$V_t^i \leftarrow Q_t^i + \sum_{j=1}^N P_t^{j\top} R_t^{ij} P_t^j + (A_t - \sum_{j=1}^N B_t^j P_t^j)^\top V_{t+1}^i (A_t - \sum_{j=1}^N B_t^j P_t^j) \quad (18)$$

## 2 ILQgame Backup derivation

Still work in progress

Suppose we have an  $N$  person discrete nonlinear dynamic game:

$$x_{t+1} = f(x_t, u_t^1, u_t^2, \dots, u_t^N) \quad (19)$$

Using a first order Taylor series expansion around a proposed trajectory  $(\hat{x}_t, \hat{u}_t)$ :

$$\begin{aligned} f(\hat{x}_t + \delta \hat{x}_t, \hat{u}_t^1 + \delta \hat{u}_t^1, \dots, \hat{u}_t^N + \delta \hat{u}_t^N) &\approx f(\hat{x}_t, \hat{u}_t^1, \dots, \hat{u}_t^N) + \frac{\partial f}{\partial x} \big|_{\hat{x}_t, \hat{u}_t^1, \dots, \hat{u}_t^N} (x_t - \hat{x}_t) + \frac{\partial f}{\partial u^1} \big|_{\hat{x}_t, \hat{u}_t^1, \dots, \hat{u}_t^N} (u_t^1 - \hat{u}_t^1) \\ &\quad \dots + \frac{\partial f}{\partial u^N} \big|_{\hat{x}_t, \hat{u}_t^1, \dots, \hat{u}_t^N} (u_t^N - \hat{u}_t^N) \\ x_{t+1} - \hat{x}_{t+1} &= A(\hat{x}_t, \hat{u}_t)(x_t - \hat{x}_t) + \sum_{j=1}^N B_t^j (u_t^j - \hat{u}_t^j) \\ \delta x_{t+1} &= A_t \delta x_t + \sum_{j=1}^N B_t^j \delta u_t^j \end{aligned} \quad (20)$$

Where  $\delta x_t$  and  $\delta u_t$  are changes to the state and control input trajectories, respectively.

A non-quadratic cost function for each agent  $i$ , is approximated using a second order Taylor expansion:

$$\begin{aligned} c(\hat{x}_t + \delta \hat{x}_t, \hat{u}_t + \delta \hat{u}_t) &\approx c(\hat{x}_t, \hat{u}_t) + \nabla_{x_t, u_t} c(\hat{x}_t, \hat{u}_t) \begin{bmatrix} x_t - \hat{x}_t \\ u_t - \hat{u}_t \end{bmatrix} \\ &\quad + \frac{1}{2} \begin{bmatrix} x_t - \hat{x}_t \\ u_t - \hat{u}_t \end{bmatrix}^\top \nabla_{x_t, u_t}^2 c(\hat{x}_t, \hat{u}_t) \begin{bmatrix} x_t - \hat{x}_t \\ u_t - \hat{u}_t \end{bmatrix} \end{aligned} \quad (21)$$

Where the gradient of the cost function is defined as,

$$\nabla_{x_t, u_t} c(\hat{x}_t, \hat{u}_t) = \begin{bmatrix} \frac{\partial c}{\partial x} & \frac{\partial c}{\partial u} \end{bmatrix} \quad (22)$$

and the hessian,

$$\nabla_{x_t, u_t}^2 c(\hat{x}_t, \hat{u}_t) = \begin{bmatrix} \frac{\partial^2 c}{\partial x^2} & \frac{\partial^2 c}{\partial x \partial u} \\ \frac{\partial^2 c}{\partial u \partial x} & \frac{\partial^2 c}{\partial u^2} \end{bmatrix} \quad (23)$$

Rearranging:

$$c(\delta x_t, \delta u_t) = \frac{1}{2} (\delta x_t^\top Q_t + 2l_t^\top) x_t + \frac{1}{2} (\delta u_t^\top R_t + 2r_t^\top) \delta u_t + \frac{1}{2} \delta x_t^\top H_t \delta u_t + \frac{1}{2} \delta u_t^\top H_t \delta x_t + c \quad (24)$$

Where:

$$\begin{aligned} Q_t &= \frac{\partial^2 c}{\partial x^2} \\ l_t &= \nabla_x c \\ R_t &= \frac{\partial^2 c}{\partial u^2} \\ r_t &= \nabla_u c \\ H_t &= \frac{\partial^2 c}{\partial u \partial x} \end{aligned} \quad (25)$$

Ignoring the mixed partials and the constant term,

$$c(\delta x_t, \delta u_t) = \frac{1}{2}(\delta x_t^\top Q + 2l_t^\top)\delta x_t + \frac{1}{2}(\delta u_t^\top R + 2r_t^\top)\delta u_t \quad (26)$$

The problem can now be formulated as an LQGame problem with an affine term due to the gradients in the cost function. For a dynamic game, the cost function 26, can be rewritten as for agent  $i$ ,

$$c_i(\delta x_t, \delta u_t^1, \dots, \delta u_t^N) = \frac{1}{2}(\delta x_t^\top Q_t^i + 2l_t^{i\top})\delta x_t + \frac{1}{2} \sum_{j=1}^N (\delta u_t^{j\top} R_t^{ij} + 2r_t^{ij\top})\delta u_t^j \quad (27)$$

At  $T - 1$ ,

$$J_i^*(\delta x_{T-1}, T - 1) = \min_{\delta u_i} [c_i(\delta x_{T-1}, \delta u_{T-1})] \quad (28)$$

At the terminal state, the cost is only associated with the state,

$$J_i^*(\delta x_{T-1}, T - 1) = \frac{1}{2}(\delta x_{T-1}^\top Q_{T-1}^i + 2l_{T-1}^{i\top})\delta x_{T-1} \quad (29)$$

Therefore, the next state value function has the form of,

$$J_i^*(\delta x_{T-1}, T - 1) = \frac{1}{2}(\delta x_{T-1}^\top V_{T-1}^i + 2\zeta_{T-1}^{i\top})\delta x_{T-1} \quad (30)$$

Which is quadratic in state.

At the second to last time step,  $T - 2$ , the value function is,

$$J_i^*(\delta x_{T-2}, T - 2) = \min_{u_i} [c(\delta x_{T-2}, \delta u_{T-2}) + J_i^*(\delta x_{T-1}, T - 1)] \quad (31)$$

Substituting in the quadratic cost function and the next state value function,

$$J_i^*(\delta x_{T-2}, T - 2) = \min_{u_i} \left[ \frac{1}{2}(\delta x_{T-2}^\top Q_{T-2}^i + 2l_{T-2}^{i\top})\delta x_{T-2} + \frac{1}{2} \sum_{j=1}^N (\delta u_{T-2}^{j\top} R_{T-2}^{ij} + 2r_{T-2}^{ij\top})\delta u_{T-2}^j \right. \\ \left. + \frac{1}{2}(\delta x_{T-1}^\top V_{T-1}^i + 2\zeta_{T-1}^{i\top})\delta x_{T-1} \right] \quad (32)$$

Substituting in the linear dynamics and setting  $t = T - 2$  for notation simplicity,

$$J_i^*(x_t, t) = \min_{u_i} \left[ \frac{1}{2}(\delta x_t^\top Q_t^i + 2l_t^{i\top})\delta x_t + \frac{1}{2} \sum_{j=1}^N (\delta u_t^{j\top} R_t^{ij} + 2r_t^{ij\top})\delta u_t^j \right. \\ \left. + \frac{1}{2}((A_t \delta x_t + \sum_{j=1}^N B_t^j \delta u_t^j)^\top V_{t+1}^i + 2\zeta_{t+1}^{i\top})(A_t \delta x_t + \sum_{j=1}^N B_t^j \delta u_t^j) \right] \quad (33)$$

We then take the derivative of the value function wrt the control input of player  $i$ , set to zero, and solve for the optimal control,

$$\frac{\partial}{\partial \delta u^i} J_i^*(\delta x_t, t) = 0 \quad (34)$$

$$0 = R_t^{ii} \delta u_t^i + r_t^{ii} + B_t^{i\top} V_{t+1}^i B_t^i \delta u_t^i + B_t^{i\top} V_{t+1}^i A_t \delta x_t + B_t^{i\top} V_{t+1}^i \sum_{j=1}^N B_t^j \delta u_t^j + B_t^{i\top} \zeta_{t+1}^i \quad (35)$$

Simplifying and rearranging,

$$(R_t^{ii} + B_t^{i\top} V_{t+1}^i B_t^i) \delta u_t^i + B_t^{i\top} V_{t+1}^i \sum_{j=1}^N B_t^j \delta u_t^j = -B_t^{i\top} V_{t+1}^i A_t \delta x_t - B_t^{i\top} \zeta_{t+1}^i \quad (36)$$

Notice the form of the optimal control input that minimizes the objective,

$$\delta u_t^{i*} = -P_t^i \delta x_t - \alpha_t^i \quad (37)$$

Substituting 37 into 36,

$$(R_t^{ii} + B_t^{i\top} V_{t+1}^i B_t^i)(-P_t^i \delta x_t - \alpha_t^i) + B_t^{i\top} V_{t+1}^i \sum_{j=1}^N B_t^j (-P_t^j \delta x_t - \alpha_t^j) = -B_t^{i\top} V_{t+1}^i A_t \delta x_t - B_t^{i\top} \zeta_{t+1}^i \quad (38)$$

the system of linear equations in 38 can be rewritten as,

$$(R_t^{ii} + B_t^{i\top} V_{t+1}^i B_t^i) P_t^i + B_t^{i\top} V_{t+1}^i \sum_{j=1}^N B_t^j P_t^j = B_t^{i\top} V_{t+1}^i A_t \quad (39)$$

$$(R_t^{ii} + B_t^{i\top} V_{t+1}^i B_t^i) \alpha_t^i + B_t^{i\top} V_{t+1}^i \sum_{j=1}^N B_t^j \alpha_t^j = B_t^{i\top} \zeta_{t+1}^i \quad (40)$$

Where both system of equations can be written down in matrix form similar to 14.

To solve for  $P_t^i$  and  $\alpha_t^i$ , we substitute the optimal control input,  $u_t^{i*}$ , into the value function (equation 33) to obtain a recursive relation for  $V_t$  and  $\zeta_t$ ,

$$\begin{aligned} J_i^*(x_t, t) = & \frac{1}{2}(\delta x_t^\top Q_t^i + 2l_t^{i\top})\delta x_t + \frac{1}{2} \sum_{j=1}^N ((-P_t^j \delta x_t - \alpha_t^j)^\top R_t^{ij} + 2r_t^{ij\top})(-P_t^j \delta x_t - \alpha_t^j) \\ & + \frac{1}{2}((A_t \delta x_t + \sum_{j=1}^N B_t^j (-P_t^j \delta x_t - \alpha_t^j))^\top V_{t+1}^i + 2\zeta_{t+1}^{i\top})(A_t \delta x_t + \sum_{j=1}^N B_t^j (-P_t^j \delta x_t - \alpha_t^j)) \end{aligned} \quad (41)$$

$$\begin{aligned} J_i^*(x_t, t) = & \frac{1}{2}(\delta x_t^\top Q_t^i + 2l_t^{i\top})\delta x_t + \frac{1}{2} \sum_{j=1}^N ((P_t^j \delta x_t + \alpha_t^j)^\top R_t^{ij} - 2r_t^{ij\top})(P_t^j \delta x_t + \alpha_t^j) \\ & + \frac{1}{2}((A_t \delta x_t - \sum_{j=1}^N B_t^j (P_t^j \delta x_t + \alpha_t^j))^\top V_{t+1}^i + 2\zeta_{t+1}^{i\top})(A_t \delta x_t - \sum_{j=1}^N B_t^j (P_t^j \delta x_t + \alpha_t^j)) \end{aligned} \quad (42)$$

Simplifying further to match the form of the value function (equation 30) and neglecting the constant term,

$$\begin{aligned} J_i^*(x_t, t) = & \frac{1}{2} \delta x_t^\top (Q_t^i + \sum_{j=1}^N P_t^{j\top} R_t^{ij} P_t^j + (A_t \delta x_t - \sum_{j=1}^N B_t^j P_t^j)^\top V_{t+1}^i (A_t \delta x_t - \sum_{j=1}^N B_t^j P_t^j)) \delta x_t \\ & + (l_t^i + \sum_{j=1}^N (P_t^{j\top} R_t^{ij} \alpha_t^j - P_t^{j\top} r_t^{ij}) + (A_t \delta x_t - \sum_{j=1}^N B_t^j P_t^j)^\top (\zeta_{t+1}^i - V_{t+1}^i \sum_{j=1}^N B_t^j \alpha_t^j))^\top \delta x_t \end{aligned} \quad (43)$$

The value function can be updated recursively as,

$$V_t^i \leftarrow Q_t^i + \sum_{j=1}^N P_t^{j\top} R_t^{ij} P_t^j + (A_t \delta x_t - \sum_{j=1}^N B_t^j P_t^j)^\top V_{t+1}^i (A_t \delta x_t - \sum_{j=1}^N B_t^j P_t^j) \quad (44)$$

$$\zeta_t^i \leftarrow l_t^i + \sum_{j=1}^N (P_t^{j\top} R_t^{ij} \alpha_t^j - P_t^{j\top} r_t^{ij}) + (A_t \delta x_t - \sum_{j=1}^N B_t^j P_t^j)^\top (\zeta_{t+1}^i - V_{t+1}^i \sum_{j=1}^N B_t^j \alpha_t^j) \quad (45)$$