Theorem 1

For a dynamic game:

Dynamics: $\mathbf{x}_{t+1} = f(\mathbf{x}_t, \mathbf{u}_t).$ **where:** $\mathbf{x}_t = [\mathbf{x}_t^{1,\top} \dots \mathbf{x}_t^{N,\top}]^\top \in \mathbb{R}^m$ **(1)**

$c_t^i(\mathbf{x}_t, \mathbf{u}_t, \boldsymbol{\omega})$ $\mathbf{u}_t = [\mathbf{u}_t^{1,\top} \dots \mathbf{u}_t^{N,\top}]^\top \in \mathbb{R}^n$

First, we can use linearization around a trajectory :

$$\begin{cases} \delta\mathbf{x}_{t+1} \approx A_t \delta\mathbf{x}_t + \sum_{j\in[N]} B_t^j \delta\mathbf{u}_t^j, & \textbf{(2)} \\ c_t^i(\mathbf{x}_t, \mathbf{u}_t, \boldsymbol{\omega}) \approx c_t^i(\bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t, \boldsymbol{\omega}) + \frac{1}{2}\omega^i(\delta\mathbf{x}_t^\top Q_t^i + 2l_t^{i,\top})\delta\mathbf{x}_t + \frac{1}{2}\sum_{j\in N}\omega^{ij}(\delta\mathbf{u}_t^{j,\top} R_t^{ij} + 2r_t^{ij,\top})\ \delta\mathbf{u}_t^j, & \textbf{(3)} \end{cases}$$

Problem:

$$J\left(\pi^{i^*}(\mathbf{x}_t), \pi^{\neg i^*}(\mathbf{x}_t)\right) \leqslant J\left(\pi^i(\mathbf{x}_t), \pi^{\neg i^*}(\mathbf{x}_t)\right) \quad \textbf{(6)}$$

Cost function: $$J^i(\pi^i) = E_\pi[\sum_{t=1}^{T} c_t^i(\mathbf{x}_t, \mathbf{u}_t, \boldsymbol{\omega}^i) + \gamma^i D_{KL}(\pi_t^i(\cdot|\mathbf{x}_t) | \tilde{\pi}_t^i(\cdot|\mathbf{x}_t)] \quad \textbf{(4)}$$

Expert stratery: $$\tilde{\pi}_t^{i,*} = N(\tilde{\mu}_t^i, \tilde{\Sigma}_t^i) \quad \textbf{(5)}$$

The mixed-strategy Nash equilibrium is given by: $\pi_t^{i,*} = N(\mu_t^{i,*}, \sum_t^*)$

where: $$\mu_t^{i*} = -P_t^i \delta\mathbf{x}_t - \alpha_t^i, \quad \Sigma_t^{i*} = \left[\frac{1}{\lambda^i}\left(R_t^{ii} + B_t^{i^T} Z_{t+1}^i B_t^i\right) + \left(\tilde{\Sigma}_t^i\right)^{-1}\right]^{-1}. \quad \textbf{(7)}$$

And the P_t^i α_t^i **are given by :**

$$\left[\omega^{ii} R_t^{ii} + \lambda^i \left(\tilde{\Sigma}_t^i\right)^{-1} + B_t^{i\top} Z_{t+1}^i B_t^i\right] P_t^i + B_t^{i\top} Z_{t+1}^i \sum_{j\neq i\in[N]} B_t^j P_t^j = B_t^{i\top} Z_{t+1}^i A_t \quad \textbf{(8)}$$

$$\left[\omega^{ii} R_t^{ii} + \lambda^i \left(\tilde{\Sigma}_t^i\right)^{-1} + B_t^{i\top} Z_{t+1} B_t^i\right] \alpha_t^i + B_t^{i\top} Z_{t+1}^i \sum_{j\neq i\in[N]} B_t^j \alpha_t^j = B_t^{i\top} z_{t+1}^i + \omega^{ii} r_t^{ii} - \lambda^i \left(\tilde{\Sigma}_t^i\right)^{-1} \tilde{\mu}_t^j$$

And the Z_t^i z_t^i **are given by :**

$$Z_t^i = \omega^i Q_t^i + \sum_{j\in[N]} \omega^{ij} \left(P_t^j\right)^\top R_t^{ij} P_t^j + F_t^\top Z_{t+1}^i F_t + \lambda^i P_t^{i\top} \left(\tilde{\Sigma}_t^i\right)^{-1} P_t^i, \quad \textbf{(9)}$$

$$z_t^i = l_t^i + \sum_{j\in[N]} \omega^{ij} \left(P_t^j\right)^\top R_t^{ij} \alpha_t^j + F_t^\top \left(z_{t+1}^i + Z_{t+1}^i \beta_t\right) - P_t^i r_t^{ij} + \lambda^i P_t^{i,\top} \left(\tilde{\Sigma}_t^i\right)^{-1} \left(\alpha_t^i - \tilde{\mu}_t^i\right),$$

The proof

With the help of the idea of Bellman's equation, we can denote state value function and state-action value function:

$$V^{i}\left(\mathbf{x_t}\right)=\inf_{\pi_t^i}\left\{E^{\pi}\left[Q^{i}\left(\mathbf{x_t},\mathbf{u_t}\right)\right]+\lambda^{i}D_{KL}\left[\pi^{i}\left(\cdot\mid \mathbf{x_t}\right)\|\ \tilde{\pi}^{i}\left(\cdot\mid \mathbf{x_t}\right)\right]\right\} \tag{10}$$

$$Q^{i}(\mathbf{x}_t,\mathbf{u}_t,\boldsymbol{\omega})=c_t^{i}(\mathbf{x_t},\mathbf{u_t},\boldsymbol{\omega})+V^{i}(\mathbf{x}_{t+1}) \tag{11}$$

Based on the assumption of the cost-generalised form in the statement of the (t = T) theorem, we assume by induction that the state-value function is a quadratic form at time t + 1.

$$V_{t+1}^{i*}=\frac{1}{2}\mathbf{x}_{t+1}^{\top}Z_{t+1}^{\top}\mathbf{x}_{t+1}+z_{t+1}^{\mathrm{i}\top}\mathbf{x}_{t}+\eta_{t+1}^{\mathrm{i}} \tag{12}$$

And we can denote equation (9) as:

$$\begin{aligned}V_t^{i*}\left(x_t\right)&=\min_{\pi_t^i}E^{\pi_t}\left[c_t^{i}\left(\bar{\mathbf{x}}_t,\bar{\mathbf{u}}_t,\boldsymbol{\omega}\right)+\frac{1}{2}\omega^{i}(\delta\mathbf{x}_t^{\top}Q_t^{i}+2l_t^{i,\top})\delta\mathbf{x}_t+\frac{1}{2}\sum_{j\in N}\omega^{ij}(\delta\mathbf{u}_t^{j,\top}R_t^{ij}+2r_t^{ij,\top})\ \delta\mathbf{u}_t^{j}+\lambda^{i}D_{KL}\left(\pi_t^{i}\|\ \tilde{\pi}_t^{i}\right)+E^{x_{t+1}}\left[V_{t+1}^{i*}\left(x_{t+1}\right)\right]\right]\\&=\min_{\pi_t^i}E^{\pi_t}[\underbrace{c_t^{i}\left(\bar{\mathbf{x}}_t,\bar{\mathbf{u}}_t,\boldsymbol{\omega}\right)+\frac{1}{2}\omega^{i}(\delta\mathbf{x}_t^{\top}Q_t^{i}+2l_t^{i,\top})\delta\mathbf{x}_t+\frac{1}{2}\sum_{j\in N}\omega^{ij}(\delta\mathbf{u}_t^{j,\top}R_t^{ij}+2r_t^{ij,\top})\ \delta\mathbf{u}_t^{j}}_{(1)}+\underbrace{\lambda^{i}D_{KL}\left(\pi_t^{i}\|\ \tilde{\pi}_t^{i}\right)}_{(2)}+\underbrace{E^{x_{t+1}}\left[V_{t+1}^{i*}\left(A_t\delta\mathbf{x}_t+\sum_{j\in[N]}B_t^{j}\delta\mathbf{u}_t^{j}\right)\right]}_{(3)}.\end{aligned} \tag{13}$$

where:

$$\begin{cases}(1)=c_t^{i}\left(\bar{\mathbf{x}}_t,\bar{\mathbf{u}}_t,\boldsymbol{\omega}\right)+\frac{1}{2}\omega^{i}(\delta\mathbf{x}_t^{\top}Q_t^{i}+2l_t^{i,\top})\delta\mathbf{x}_t+\frac{1}{2}\left(\sum_{j\in[N]}\omega^{ij}\mu_t^{j\top}R_t^{ij}\mu_t^{j}+\omega^{ij}\operatorname{tr}\left(R_t^{ij}\Sigma_t^{j}\right)+2\omega^{ij}r_t^{ij,\top}\mu_t^{j}\right) & (14)\\ (2)=E^{\pi_t}\left[\lambda^{i}D_{KL}\left(\pi_t^{i}\|\ \tilde{\pi}_t^{i}\right)\right]=\frac{\lambda^{i}}{2}\left(n_{u^i}-\log\det\left(\Sigma_t^{i}\right)+\log\det\left(\tilde{\Sigma}_t^{i}\right)+\operatorname{tr}\left(\left(\tilde{\Sigma}_t^{i}\right)^{-1}\Sigma_t^{i}\right)+\left(\mu_t^{i}-\tilde{\mu}_t^{i}\right)^{\top}\left(\tilde{\Sigma}_t^{i}\right)^{-1}\left(\mu_t^{i}-\tilde{\mu}_t^{i}\right)\right) & (15)\\ (3)=E^{\pi_t}\left[E^{\mathbf{x}_{t+1}}\left[\frac{1}{2}\left(A_t\delta\mathbf{x}_t+\sum_{j\in[N]}B_t^{j}\delta\mathbf{u}_t^{j}\right)^{\top}Z_{t+1}^{i}\left(A_t\delta\mathbf{x}_t+\sum_{j\in[N]}B_t^{j}\delta\mathbf{u}_t^{j}\right)+z_{t+1}^{i\top}\left(A_t\delta\mathbf{x}_t+\sum_{j\in[N]}B_t^{j}\delta\mathbf{u}_t^{j}\right)+\eta_{t+1}^{i}\right]\right] & (16)\end{cases}$$

$$(16) \quad \text{dropping terms not affecting:} \Rightarrow \frac{1}{2}\left(\sum_{j\in[N]} B_t^j \mu_t^j\right)^\top Z_{t+1}^i \left(\sum_{j\in[N]} B_t^j \mu_t^j\right) + \sum_{j\in[N]} \left(A_t \delta \mathbf{x}_t\right)^\top Z_{t+1}^i B_t^j \mu_t^j + \frac{1}{2}\mathrm{tr}\left(Z_{t+1}^i \left(\Sigma_d + \sum_{j\in[N]} B_t^{j\top} \Sigma_t^j B_t^j\right)\right) + \sum_{j\in[N]} z_{t+1}^{i\top} B_t^j \mu_t^j$$

In addition, combining Equations (27), (28), (30) and dropping terms not affecting the minimum, we have:

$$V_t^{i*}(x_t) = \frac{1}{2}\left(\sum_{j\in[N]} \omega^{ij} \mu_t^{j\top} R_t^{ij} \mu_t^j + \omega^{ij}\,\mathrm{tr}\left(R_t^{ij}\Sigma_t^j\right) + 2\omega^{ij} r_t^{ij,\top} \mu_t^j\right) + \frac{\lambda^i}{2}\left(-\log\det\left(\Sigma_t^i\right) + \log\det\left(\tilde{\Sigma}_t^i\right) + \mathrm{tr}\left(\left(\tilde{\Sigma}_t^i\right)^{-1}\Sigma_t^i\right) + \left(\mu_t^i - \tilde{\mu}_t^i\right)^\top \left(\tilde{\Sigma}_t^i\right)^{-1}\left(\mu_t^i - \tilde{\mu}_t^i\right)\right) +$$

$$\frac{1}{2}\left(\sum_{j\in[N]} B_t^j \mu_t^j\right)^\top Z_{t+1}^i \left(\sum_{j\in[N]} B_t^j \mu_t^j\right) + \sum_{j\in[N]} \left(A_t \delta \mathbf{x}_t\right)^\top Z_{t+1}^i B_t^j \mu_t^j + \frac{1}{2}\mathrm{tr}\left(Z_{t+1}^i \left(\Sigma_d + \sum_{j\in[N]} B_t^{j\top} \Sigma_t^j B_t^j\right)\right) + \sum_{j\in[N]} z_{t+1}^{i\top} B_l^j \mu_t^j \qquad \textbf{(17)}$$

We may take the gradient of Equation (31) and equate it to zero to find the optimal mean controls and the corresponding covariances. The optimal control is adopted by linear state-feedback in the form of $\mu_t^{i*} = -P_t^i \delta \mathbf{x}_t - \alpha_t^i$, and we can get Riccati equations

$$\left[\omega^{ii} R_t^{ii} + \lambda^i \left(\tilde{\Sigma}_t^i\right)^{-1} + B_t^{i\top} Z_{t+1}^i B_t^i\right] P_t^i + B_t^{i\top} Z_{t+1}^i \sum_{j\neq i\in[N]} B_t^j P_t^j = B_t^{i\top} Z_{t+1}^i A_t \qquad \textbf{(8)}$$

$$\left[\omega^{ii} R_t^{ii} + \lambda^i \left(\tilde{\Sigma}_t^i\right)^{-1} + B_t^{i\top} Z_{t+1}^i B_t^i\right] \alpha_t^i + B_t^{i\top} Z_{t+1}^i \sum_{j\neq i\in[N]} B_t^j \alpha_t^j = B_t^{i\top} z_{t+1}^i + \omega^{ii} r_t^{ii} - \lambda^i \left(\tilde{\Sigma}_t^i\right)^{-1} \tilde{\mu}_t^j$$

Repeating the same procedures, we can get:

$$\Sigma_t^{i*} = \left[\frac{1}{\lambda^i}\left(\omega^{ii} R_t^{ii} + B_t^{iT} Z_{t+1}^i B_t^i\right) + \left(\tilde{\Sigma}_t^i\right)^{-1}\right]^{-1}. \qquad \textbf{(18)}$$

To solve the Riccati equations, it is necessary to derive expressions for Z_t^i and z_t^i, we can take the mean and covariance into value equation:

$$
\begin{aligned}
V_t^{i*}(x_t) = E^{\pi_t^*}[c_t^i(\bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t, \boldsymbol{\omega}) + \frac{1}{2}\omega^i(\delta\mathbf{x}_t^\top Q_t^i + 2l_t^{i,\top})\delta\mathbf{x}_t + \frac{1}{2}\sum_{j\in N}\omega^{ij}((-P_t^j\delta\mathbf{x}_t - \alpha_t^j)^\top R_t^{ij} + 2r_t^{ij,\top})\ (-P_t^j\delta\mathbf{x}_t - \alpha_t^j) \\
+\lambda^i D_{KL}\left(\pi_t^i \| \ \tilde{\pi}_t^i\right) + E^{\delta\mathbf{x}_{t+1}}\frac{1}{2}\left(A_t\delta\mathbf{x}_t + \sum_{j\in[N]}B_t^j\left(-P_t^j\delta\mathbf{x}_t - \alpha_t^j\right) + d_t\right)Z_{t+1}^i\left(A_t\delta\mathbf{x}_t + \sum_{j\in[N]}B_t^j\left(-P_t^j\delta\mathbf{x}_t - \alpha_t^j\right)\right) \\
+E^{\delta\mathbf{x}_{t+1}} z_{t+1}^{i\top}\left(A_t\delta\mathbf{x}_t + \sum_{j\in[N]}B_t^j\left(-P_t^j\delta\mathbf{x}_t - \alpha_t^j\right)\right) + \eta_{t+1}^i].
\end{aligned}
\tag{19}
$$

Taking expectations and rearranging yields the recursion parameters for the quadratic value function, which we now show. Firstly, let $\begin{cases} F_t = A_t - \sum_{j\in[N]} B_t^j P_t^j \\ \beta_t = -\sum_{j\in[N]} B_t^j \alpha_t^j \end{cases}$

$$
\begin{aligned}
V_t^{i*}(x_t) = \frac{1}{2}\delta\mathbf{x}_t^\top\left[\omega^i Q_t^i + \sum_{j\in[N]}\omega^{ij}P_t^{j\top}R_t^{ij}P_t^j + F_t^\top Z_{t+1}^i F_t + \lambda^i P_t^{i\top}\left(\tilde{\Sigma}_t^i\right)^{-1}P_t^i\right]\delta\mathbf{x}_t \\
+\left[l_t^i + \sum_{j\in[N]}\omega^{ij}P_t^{j\top}R_t^{ij}\alpha_t^j + F_t^\top\left(z_{t+1}^i + Z_{t+1}^i\beta_t\right)^\top + \lambda^i P_t^{i\top}\left(\tilde{\Sigma}_t^i\right)^{-1}\left(\alpha_t^i - \tilde{\mu}_t^i\right)\right]\delta\mathbf{x}_t + \cdots,
\end{aligned}
\tag{20}
$$

We then compare the above equation with the form of the value function assumed at the outset by comparing

$$
\begin{aligned}
Z_t^i &= \omega^i Q_t^i + \sum_{j\in[N]}\omega^{ij}\left(P_t^j\right)^\top R_t^{ij}P_t^j + F_t^\top Z_{t+1}^i F_t + \lambda^i P_t^{i\top}\left(\tilde{\Sigma}_t^i\right)^{-1}P_t^i, \\
z_t^i &= l_t^i + \sum_{j\in[N]}\omega^{ij}\left(P_t^j\right)^\top R_t^{ij}\alpha_t^j + F_t^\top\left(z_{t+1}^i + Z_{t+1}^i\beta_t\right) - P_t^i r_t^{ij} + \lambda^i P_t^{i,\top}\left(\tilde{\Sigma}_t^i\right)^{-1}\left(\alpha_t^i - \tilde{\mu}_t^i\right),
\end{aligned}
\tag{9}
$$

Finally, we have completed the proof of Theorem I