Appendix A: Proof of Theorem 1

Theorem 1 (Multi-Agent Model-based Policy Gradient)**.** $\forall i \in \{1, 2, \dots, N\}$, *we have:*

$$\boldsymbol{\nabla}_{\boldsymbol{\theta}_i} J = \sum_{s\in\mathcal{V}} (\log\mu(s)+1)(1-\mu(s))\boldsymbol{\nabla}_{\boldsymbol{\theta}_i}\log(1-\mu_i(s)).$$

The proof process of Theorem 1 as depicted below is to make use of the chain rule of multi-variable calculus, and take advantage of the fact that $\forall j \neq i$, $\boldsymbol{\nabla}_{\boldsymbol{\theta}_i}\mu_j(s) = \mathbf{0}$. Defining $J(s) = -\mu(s)\log(\mu(s))$, we have $J = \sum_{s\in\mathcal{V}} J(s)$, and the proof process of multi-agent model-based policy gradient is as follows:

Proof.

$$\begin{aligned}
&\boldsymbol{\nabla}_{\boldsymbol{\theta}_i} J \\
=&\boldsymbol{\nabla}_{\boldsymbol{\theta}_i} \sum_{s\in\mathcal{V}} J(s) \\
=&\sum_{s\in\mathcal{V}} \Big(\frac{\partial J(s)}{\partial\mu(s)} \times \sum_{j=1}^{N} \big(\frac{\partial\mu(s)}{\partial\mu_j(s)} \times \frac{\partial\mu_j(s)}{\partial\boldsymbol{\theta}_i}\big)\Big) \\
=&\sum_{s\in\mathcal{V}} \sum_{j=1}^{N} \frac{\partial J(s)}{\partial\mu(s)} \times \frac{\partial\mu(s)}{\partial\mu_j(s)} \times \frac{\partial\mu_j(s)}{\partial\boldsymbol{\theta}_i} \\
=&\sum_{s\in\mathcal{V}} \sum_{j\neq i} \frac{\partial J(s)}{\partial\mu(s)} \frac{\partial\mu(s)}{\partial\mu_j(s)} \frac{\partial\mu_j(s)}{\partial\boldsymbol{\theta}_i} + \sum_{s\in\mathcal{V}} \frac{\partial J(s)}{\partial\mu(s)} \frac{\partial\mu(s)}{\partial\mu_i(s)} \frac{\partial\mu_i(s)}{\partial\boldsymbol{\theta}_i} \\
=&\sum_{s\in\mathcal{V}} \frac{\partial J(s)}{\partial\mu(s)} \frac{\partial\mu(s)}{\partial\mu_i(s)} \frac{\partial\mu_i(s)}{\partial\boldsymbol{\theta}_i} \\
=&\sum_{s\in\mathcal{V}} -(\log\mu(s)+1) \times \prod_{j=1, j\neq i}^{N} (1-\mu_j(s)) \times \frac{\partial\mu_i(s)}{\partial\boldsymbol{\theta}_i} \\
=&\sum_{s\in\mathcal{V}} -(\log\mu(s)+1) \times \frac{\prod_{j=1}^{N}(1-\mu_j(s))}{1-\mu_i(s)} \times \frac{\partial\mu_i(s)}{\partial\boldsymbol{\theta}_i} \\
=&\sum_{s\in\mathcal{V}} -(\log\mu(s)+1) \times \frac{1-\mu(s)}{1-\mu_i(s)} \times \frac{\partial\mu_i(s)}{\partial\boldsymbol{\theta}_i} \\
=&\sum_{s\in\mathcal{V}} (\log\mu(s)+1)(1-\mu(s))\boldsymbol{\nabla}_{\boldsymbol{\theta}_i}\log(1-\mu_i(s))
\end{aligned}$$

□