PROJECT REPORT

MULTI-AGENT REINFORCEMENT LEARNING VIA ADAPTIVE KALMAN TEMPORAL DIFFERENCE AND SUCCESSOR REPRESENTATION

Mohammad Salimibeni | Arash Mohammadi | Parvin Malekzadeh and Konstantinos N. Plataniotis

RLinControl | Fall 2024

Professor: Dr. Saeed Shamaghdari
Seyede Setare Khosravi | Mobina Lashgari

TABLE OF CONTENTS

ABSTRACT

Reinforcement Learning Challenges

- Fixed Reward Function
- Overfitting, High sensitivity and …

Kalman an Idea to Solve The Challenges

- Online Learning and …

INTRODUCTION MARI

Questions!

INTRODUCTION MARI

Challenges!

Problem Formulation

SINGLE AGENT REINFORCEMENT LEARNING

$$Q_\pi(\boldsymbol{s}, a) = \mathbb{E}\left\{\sum_{k=0}^{T} \gamma^k r_k | \boldsymbol{s}_0 = \boldsymbol{s}, a_0 = a, a_k = \pi(\boldsymbol{s}_k)\right\}$$

SINGLE AGENT REINFORCEMENT LEARNING

$$Q_\pi(\mathbf{s},a) = \mathbb{E}\left\{\sum_{k=0}^{T} \gamma^k r_k | \mathbf{s}_0 = \mathbf{s}, a_0 = a, a_k = \pi(\mathbf{s}_k)\right\}$$

$$a_k = \arg\max_{a \in \mathcal{A}} Q_{\pi^*}(\mathbf{s}_k, a).$$

TD LEARNING

$$Q_{\pi^*}(\boldsymbol{s}_k, a_k) = Q_{\pi^*}(\boldsymbol{s}_k, a_k) + \alpha\left(r_k + \gamma \max_{a \in \mathcal{A}} Q_{\pi^*}(\boldsymbol{s}_{k+1}, a) - Q_{\pi^*}(\boldsymbol{s}_k, a_k)\right)$$

TD LEARNING

$$Q_{\pi^*}(\boldsymbol{s}_k, a_k) = Q_{\pi^*}(\boldsymbol{s}_k, a_k) + \alpha\left(r_k + \gamma \max_{a \in \mathcal{A}} Q_{\pi^*}(\boldsymbol{s}_{k+1}, a) - Q_{\pi^*}(\boldsymbol{s}_k, a_k)\right)$$

$$a_k = \arg\max_{a \in \mathcal{A}} Q_{\pi^*}(\boldsymbol{s}_k, a)$$

MULTI AGENT SETTING

Agent i, for $(1 \leq i \leq N)$

$$\mathbb{S} = \{\mathcal{S}^{(1)}, \ldots, \mathcal{S}^{(N)}\}$$

$$\mathbb{A} = \{\mathcal{A}^{(1)}, \ldots, \mathcal{A}^{(N)}\}$$

$$\mathbb{Z} = \{\mathcal{Z}^{(1)}, \ldots, \mathcal{Z}^{(N)}\}$$

$$r^{(i)} : \mathbb{S} \times \mathcal{A}^{(i)} \rightarrow \mathbb{R}$$

$$R^{(i)} = \sum_{t=0}^{T} \gamma^t (r^{(i)})^t$$

MULTI AGENT SR

$$\boldsymbol{M}_{\pi^{(i)}}(\boldsymbol{s}^{(i)}, \boldsymbol{s}'^{(i)}, a^{(i)}) = \mathbb{E}\left[\sum_{k=0}^{T} \gamma^k \mathbb{1}[\boldsymbol{s}_k^{(i)} = \boldsymbol{s}'^{(i)}] | \boldsymbol{s}_0^{(i)} = \boldsymbol{s}^{(i)}, a_0^{(i)} = a^{(i)}\right]$$

$$\begin{aligned}\boldsymbol{M}^{\text{new}}_{\pi^{(i)}}(\boldsymbol{s}_k^{(i)}, \boldsymbol{s}'^{(i)}, a_k^{(i)}) &= \boldsymbol{M}^{\text{old}}_{\pi^{(i)}}(\boldsymbol{s}_k^{(i)}, \boldsymbol{s}'^{(i)}, a_k^{(i)}) + \\ &\alpha\left(\mathbb{1}[\boldsymbol{s}_k^{(i)} = \boldsymbol{s}'^{(i)}] + \gamma \boldsymbol{M}_{\pi^{(i)}}(\boldsymbol{s}_{k+1}^{(i)}, \boldsymbol{s}'^{(i)}, a_{k+1}^{(i)}) - \boldsymbol{M}^{\text{old}}_{\pi^{(i)}}(\boldsymbol{s}_k^{(i)}, \boldsymbol{s}'^{(i)}, a_k^{(i)})\right)\end{aligned}$$

$$Q_{\pi^{(i)}}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) = \sum_{\boldsymbol{s}'^{(i)} \in \mathcal{S}^{(i)}} \boldsymbol{M}(\boldsymbol{s}_k^{(i)}, \boldsymbol{s}'^{(i)}, a_k^{(i)}) R^{(i)}(\boldsymbol{s}'^{(i)}, a_k^{(i)})$$

MULTI AGENT SR

$$\boldsymbol{M}_{\pi^{(i)}}(\boldsymbol{s}^{(i)}, \boldsymbol{s}'^{(i)}, a^{(i)}) = \mathbb{E}\left[\sum_{k=0}^{T} \gamma^k \mathbb{1}[\boldsymbol{s}_k^{(i)} = \boldsymbol{s}'^{(i)}] | \boldsymbol{s}_0^{(i)} = \boldsymbol{s}^{(i)}, a_0^{(i)} = a^{(i)}\right]$$

$$\begin{aligned}\boldsymbol{M}^{\text{new}}_{\pi^{(i)}}(\boldsymbol{s}_k^{(i)}, \boldsymbol{s}'^{(i)}, a_k^{(i)}) = \boldsymbol{M}^{\text{old}}_{\pi^{(i)}}(\boldsymbol{s}_k^{(i)}, \boldsymbol{s}'^{(i)}, a_k^{(i)}) + \\ \alpha\Big(\mathbb{1}[\boldsymbol{s}_k^{(i)} = \boldsymbol{s}'^{(i)}] + \gamma \boldsymbol{M}_{\pi^{(i)}}(\boldsymbol{s}_{k+1}^{(i)}, \boldsymbol{s}'^{(i)}, a_{k+1}^{(i)}) - \boldsymbol{M}^{\text{old}}_{\pi^{(i)}}(\boldsymbol{s}_k^{(i)}, \boldsymbol{s}'^{(i)}, a_k^{(i)})\Big)\end{aligned}$$

$$Q_{\pi^{(i)}}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) = \sum_{\boldsymbol{s}'^{(i)} \in \mathcal{S}^{(i)}} \boldsymbol{M}(\boldsymbol{s}_k^{(i)}, \boldsymbol{s}'^{(i)}, a_k^{(i)}) R^{(i)}(\boldsymbol{s}'^{(i)}, a_k^{(i)})$$

MULTI AGENT SR

$$\boldsymbol{M}_{\pi^{(i)}}(\boldsymbol{s}^{(i)}, \boldsymbol{s}'^{(i)}, a^{(i)}) = \mathbb{E}\left[\sum_{k=0}^{T} \gamma^k \mathbb{1}[\boldsymbol{s}_k^{(i)} = \boldsymbol{s}'^{(i)}] | \boldsymbol{s}_0^{(i)} = \boldsymbol{s}^{(i)}, a_0^{(i)} = a^{(i)}\right]$$

$$\begin{aligned}\boldsymbol{M}^{\text{new}}_{\pi^{(i)}}(\boldsymbol{s}_k^{(i)}, \boldsymbol{s}'^{(i)}, a_k^{(i)}) &= \boldsymbol{M}^{\text{old}}_{\pi^{(i)}}(\boldsymbol{s}_k^{(i)}, \boldsymbol{s}'^{(i)}, a_k^{(i)}) + \\ &\alpha\left(\mathbb{1}[\boldsymbol{s}_k^{(i)} = \boldsymbol{s}'^{(i)}] + \gamma \boldsymbol{M}_{\pi^{(i)}}(\boldsymbol{s}_{k+1}^{(i)}, \boldsymbol{s}'^{(i)}, a_{k+1}^{(i)}) - \boldsymbol{M}^{\text{old}}_{\pi^{(i)}}(\boldsymbol{s}_k^{(i)}, \boldsymbol{s}'^{(i)}, a_k^{(i)})\right)\end{aligned}$$

$$Q_{\pi^{(i)}}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) = \sum_{\boldsymbol{s}'^{(i)} \in \mathcal{S}^{(i)}} \boldsymbol{M}(\boldsymbol{s}_k^{(i)}, \boldsymbol{s}'^{(i)}, a_k^{(i)}) R^{(i)}(\boldsymbol{s}'^{(i)}, a_k^{(i)})$$

MAK-TD FRAMEWORK

$$Q_{\pi^{(i)*}}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) \approx r_k^{(i)} + \gamma \max_{a^{(i)} \in \mathcal{A}} Q_{\pi^{(i)*}}(\boldsymbol{s}_{k+1}^{(i)}, a^{(i)})$$

$$r_k^{(i)} = Q_{\pi^{(i)*}}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) - \gamma \max_{a^{(i)} \in \mathcal{A}} Q_{\pi^*}(\boldsymbol{s}_{k+1}^{(i)}, a^{(i)}) + v_k^{(i)}$$

v_k is modeled as a zero-mean normal distribution with variance of $R^{(i)}$

MAK-TD FRAMEWORK

$$Q_{\pi^{(i)*}}(\mathbf{s}_k^{(i)}, a_k^{(i)}) \approx r_k^{(i)} + \gamma \max_{a^{(i)} \in \mathcal{A}} Q_{\pi^{(i)*}}(\mathbf{s}_{k+1}^{(i)}, a^{(i)})$$

$$r_k^{(i)} = Q_{\pi^{(i)*}}(\mathbf{s}_k^{(i)}, a_k^{(i)}) - \gamma \max_{a^{(i)} \in \mathcal{A}} Q_{\pi^*}(\mathbf{s}_{k+1}^{(i)}, a^{(i)}) + v_k^{(i)}$$

v_k is modeled as a zero-mean normal distribution with variance of $R^{(i)}$

MAK-TD FRAMEWORK

$$Q_{\pi^{(i)}}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) = \boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)})^T \boldsymbol{\theta}_k^{(i)}$$

$$r_k^{(i)} = \left[\boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)})^T - \gamma \max_{a^{(i)} \in \mathcal{A}} \boldsymbol{\phi}(\boldsymbol{s}_{k+1}^{(i)}, a^{(i)})^T\right] \boldsymbol{\theta}_k^{(i)} + v_k^{(i)}$$

$$\boldsymbol{h}_k^{(i)} = \boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) - \gamma \max_{a^{(i)} \in \mathcal{A}} \boldsymbol{\phi}(\boldsymbol{s}_{k+1}^{(i)}, a^{(i)})$$

$$r_k^{(i)} = [\boldsymbol{h}_k^{(i)}]^T \boldsymbol{\theta}_k^{(i)} + v_k^{(i)}$$

MAK-TD FRAMEWORK

$$Q_{\pi^{(i)}}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) = \boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)})^T \boldsymbol{\theta}_k^{(i)}$$

$$r_k^{(i)} = \left[\boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)})^T - \gamma \max_{a^{(i)} \in \mathcal{A}} \boldsymbol{\phi}(\boldsymbol{s}_{k+1}^{(i)}, a^{(i)})^T\right] \boldsymbol{\theta}_k^{(i)} + v_k^{(i)}$$

$$\boldsymbol{h}_k^{(i)} = \boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) - \gamma \max_{a^{(i)} \in \mathcal{A}} \boldsymbol{\phi}(\boldsymbol{s}_{k+1}^{(i)}, a^{(i)})$$

$$r_k^{(i)} = [\boldsymbol{h}_k^{(i)}]^T \boldsymbol{\theta}_k^{(i)} + v_k^{(i)}$$

MAK-TD FRAMEWORK

$$Q_{\pi^{(i)}}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) = \boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)})^T \boldsymbol{\theta}_k^{(i)}$$

$$r_k^{(i)} = \left[\boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)})^T - \gamma \max_{a^{(i)} \in \mathcal{A}} \boldsymbol{\phi}(\boldsymbol{s}_{k+1}^{(i)}, a^{(i)})^T\right] \boldsymbol{\theta}_k^{(i)} + v_k^{(i)}$$

$$\boldsymbol{h}_k^{(i)} = \boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) - \gamma \max_{a^{(i)} \in \mathcal{A}} \boldsymbol{\phi}(\boldsymbol{s}_{k+1}^{(i)}, a^{(i)})$$

$$r_k^{(i)} = [\boldsymbol{h}_k^{(i)}]^T \boldsymbol{\theta}_k^{(i)} + v_k^{(i)}$$

MAK-TD FRAMEWORK

$$Q_{\pi^{(i)}}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) = \boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)})^T \boldsymbol{\theta}_k^{(i)}$$

$$r_k^{(i)} = \left[\boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)})^T - \gamma \max_{a^{(i)} \in \mathcal{A}} \boldsymbol{\phi}(\boldsymbol{s}_{k+1}^{(i)}, a^{(i)})^T\right] \boldsymbol{\theta}_k^{(i)} + v_k^{(i)}$$

$$\boldsymbol{h}_k^{(i)} = \boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) - \gamma \max_{a^{(i)} \in \mathcal{A}} \boldsymbol{\phi}(\boldsymbol{s}_{k+1}^{(i)}, a^{(i)})$$

$$r_k^{(i)} = [\boldsymbol{h}_k^{(i)}]^T \boldsymbol{\theta}_k^{(i)} + v_k^{(i)}$$

MAK-TD FRAMEWORK

$$r_k^{(i)} = [\boldsymbol{h}_k^{(i)}]^T \boldsymbol{\theta}_k^{(i)} + v_k^{(i)}$$

$$x_k = F_{k-1} x_{k-1} + G_{k-1} u_{k-1} + w_{k-1}$$

$$y_k = H_k x_k + v_k$$

$$E(w_k) = E(v_k) = 0 \quad E(w_k w_j^T) = Q_k \delta_{k-j} \quad E(v_k v_j^T) = R_k \delta_{k-j} \quad E(v_k w_j^T) = 0$$

MAK-TD FRAMEWORK

$$x_k = F_{k-1} x_{k-1} + G_{k-1} u_{k-1} + w_{k-1}$$

$$y_k = H_k x_k + v_k$$

$$r_k^{(i)} = [\boldsymbol{h}_k^{(i)}]^T \boldsymbol{\theta}_k^{(i)} + v_k^{(i)}$$

$$E(w_k) = E(v_k) = 0 \quad E(w_k w_j^T) = Q_k \delta_{k-j} \quad E(v_k v_j^T) = R_k \delta_{k-j} \quad E(v_k w_j^T) = 0$$

MAK-TD FRAMEWORK

$$x_k = F_{k-1}x_{k-1} + G_{k-1}u_{k-1} + w_{k-1}$$

$$y_k = H_k x_k + v_k \qquad r_k^{(i)} = [\boldsymbol{h}_k^{(i)}]^T \boldsymbol{\theta}_k^{(i)} + v_k^{(i)}$$

$$E(w_k) = E(v_k) = 0 \quad E(w_k w_j^T) = Q_k \delta_{k-j} \quad E(v_k v_j^T) = R_k \delta_{k-j} \quad E(v_k w_j^T) = 0$$

$$K_k = P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1} = P_k^+ H_k^T R_k^{-1}$$

$$\hat{x}_k^+ = \hat{x}_k^- + K_k (y_k - H_k \hat{x}_k^-)$$

$$P_k^+ = (I - K_k H_k) P_k^- = (I - K_k H_k) P_k^- (I - K_k H_k)^T + K_k R_k K_k^T$$

MAK-TD FRAMEWORK

$$r_k^{(i)} = [\boldsymbol{h}_k^{(i)}]^T \boldsymbol{\theta}_k^{(i)} + v_k^{(i)}$$

$$K_k = P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1} \quad = P_k^+ H_k^T R_k^{-1}$$

$$\hat{x}_k^+ = \hat{x}_k^- + K_k (y_k - H_k \hat{x}_k^-)$$

$$P_k^+ = (I - K_k H_k) P_k^- = (I - K_k H_k) P_k^- (I - K_k H_k)^T + K_k R_k K_k^T$$

$$\begin{aligned}
\boldsymbol{K}_k^{j\,(i)} &= \boldsymbol{P}_{(\boldsymbol{\theta},k|k-1)}^{(i)} \boldsymbol{h}_k^{(i)} \left(\boldsymbol{h}_k^{T\,(i)} \boldsymbol{P}_{(\boldsymbol{\theta},k|k-1)}^{(i)} \boldsymbol{h}_k^{(i)} + R^{j\,(i)}\right)^{-1} \\
\hat{\boldsymbol{\theta}}_k^{j\,(i)} &= \hat{\boldsymbol{\theta}}_{(k|k-1)}^{(i)} + \boldsymbol{K}_k^{j\,(i)} \left(r_k^{(i)} - \boldsymbol{h}_k^{T\,(i)} \hat{\boldsymbol{\theta}}_{(k|k-1)}^{(i)}\right) \\
\boldsymbol{P}_{\boldsymbol{\theta},k}^{j\,(i)} &= \left(\boldsymbol{I} - \boldsymbol{K}_k^{j\,(i)} \boldsymbol{h}_k^{T\,(i)}\right) \boldsymbol{P}_{(\boldsymbol{\theta},k|k-1)}^{T\,(i)} \left(\boldsymbol{I} - \boldsymbol{K}_k^{j\,(i)} \boldsymbol{h}_k^{T\,(i)}\right) + \boldsymbol{K}_k^{j\,(i)} R^{j\,(i)} \boldsymbol{K}_k^{j\,T\,(i)}
\end{aligned}$$

MAK-TD FRAMEWORK

$$x_k = F_{k-1}x_{k-1} + G_{k-1}u_{k-1} + w_{k-1}$$

$$\bar{x}_k = F_{k-1}\bar{x}_{k-1} + G_{k-1}u_{k-1}$$

$$P_k = F_{k-1}P_{k-1}F_{k-1}^T + Q_{k-1}$$

MAK-TD FRAMEWORK

$$\hat{\boldsymbol{\theta}}_k^{(i)} = \sum_{j=1}^{M} \omega^{j\,(i)} \hat{\boldsymbol{\theta}}_k^{j\,(i)}$$

$$\boldsymbol{P}_{\boldsymbol{\theta},k}^{(i)} = \sum_{j=1}^{M} \omega^{j\,(i)} \left(\boldsymbol{P}_{\boldsymbol{\theta},k}^{j\;\,(i)} + (\hat{\boldsymbol{\theta}}^{j\,(i)} - \hat{\boldsymbol{\theta}}^{(i)})(\hat{\boldsymbol{\theta}}^{j\,(i)} - \hat{\boldsymbol{\theta}}^{(i)})^T \right)$$

MAK-TD FRAMEWORK

$$\boldsymbol{\phi}(\mathbf{s}_k^{(i)}) = \left[\phi_1(\mathbf{s}_k^{(i)}), \phi_2(\mathbf{s}_k^{(i)}), \ldots, \phi_{N_b-1}(\mathbf{s}_k^{(i)}), \phi_{N_b}(\mathbf{s}_k^{(i)})\right]^T$$

$$\phi_n(\mathbf{s}_k^{(i)}) = \exp\left\{\frac{-1}{2}(\mathbf{s}_k^{(i)} - \boldsymbol{\mu}_n^{(i)})^T \boldsymbol{\Sigma}_n^{(i)^{-1}} (\mathbf{s}_k^{(i)} - \boldsymbol{\mu}_n^{(i)})\right\}$$

where $\boldsymbol{\mu}_n^{(i)}$ and $\boldsymbol{\Sigma}_n^{(i)}$ are the mean and covariance of $\phi_n(\mathbf{s}_k^{(i)})$, for $(1 \leq n \leq N_b)$

MAK-TD FRAMEWORK

$$\boldsymbol{\phi}(\mathbf{s}_k^{(i)}) = [\phi_1(\mathbf{s}_k^{(i)}), \phi_2(\mathbf{s}_k^{(i)}), \ldots, \phi_{N_b-1}(\mathbf{s}_k^{(i)}), \phi_{N_b}(\mathbf{s}_k^{(i)})]^T$$

$$\phi_n(\mathbf{s}_k^{(i)}) = \exp\{\frac{-1}{2}(\mathbf{s}_k^{(i)} - \boldsymbol{\mu}_n^{(i)})^T \boldsymbol{\Sigma}_n^{(i)^{-1}}(\mathbf{s}_k^{(i)} - \boldsymbol{\mu}_n^{(i)})\}$$

where $\boldsymbol{\mu}_n^{(i)}$ and $\boldsymbol{\Sigma}_n^{(i)}$ are the mean and covariance of $\phi_n(\mathbf{s}_k^{(i)})$, for $(1 \leq n \leq N_b)$

MAK-TD FRAMEWORK

$$\boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) = [\phi_{1,a_1}(\boldsymbol{s}_k^{(i)}), \ldots \phi_{N_b,a_1}(\boldsymbol{s}_k^{(i)}), \phi_{1,a_2}(\boldsymbol{s}_k^{(i)}), \ldots \phi_{N_b,a_{D(i)}}(\boldsymbol{s}_k^{(i)})]^T$$

$$\boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) = [0, \ldots 0, \phi_1(\boldsymbol{s}_k^{(i)}), \ldots, \phi_N(\boldsymbol{s}_k^{(i)}), 0, \ldots 0,]^T$$

MAK-TD FRAMEWORK

$$L_k^{(i)} = \left(\boldsymbol{\phi}^T(\boldsymbol{s}_k^{(i)}, a_k)\,\boldsymbol{\theta}_k^{(i)} - r_k^{(i)}\right)^2$$

$$\begin{aligned}
\Delta\boldsymbol{\mu}^{(i)} &= -\frac{\partial L_k^{(i)}}{\partial \boldsymbol{\mu}^{(i)}} = -\frac{\partial L_k^{(i)}}{\partial Q_{\pi^*(i)}} \frac{\partial Q_{\pi^*(i)}}{\partial \boldsymbol{\phi}^{(i)}} \frac{\partial \boldsymbol{\phi}^{(i)}}{\partial \boldsymbol{\mu}^{(i)}} \\
\text{and } \Delta\boldsymbol{\Sigma}^{(i)} &= -\frac{\partial \boldsymbol{\Sigma}_k^{(i)}}{\partial \boldsymbol{\mu}^{(i)}} = -\frac{\partial L_k^{(i)}}{\partial Q_{\pi^*(i)}} \frac{\partial Q_{\pi^*(i)}}{\partial \boldsymbol{\phi}^{(i)}} \frac{\partial \boldsymbol{\phi}^{(i)}}{\partial \boldsymbol{\Sigma}^{(i)}}
\end{aligned}$$

MAK-TD FRAMEWORK

$$L_k^{(i)} = (\boldsymbol{\phi}^T(\boldsymbol{s}_k^{(i)}, a_k)\,\boldsymbol{\theta}_k^{(i)} - r_k^{(i)})^2$$

$$\begin{aligned}
\Delta\boldsymbol{\mu}^{(i)} &= -\frac{\partial L_k^{(i)}}{\partial \boldsymbol{\mu}^{(i)}} = -\frac{\partial L_k^{(i)}}{\partial Q_{\pi^*(i)}}\,\frac{\partial Q_{\pi^*(i)}}{\partial \boldsymbol{\phi}^{(i)}}\,\frac{\partial \boldsymbol{\phi}^{(i)}}{\partial \boldsymbol{\mu}^{(i)}} \\
\text{and } \Delta\boldsymbol{\Sigma}^{(i)} &= -\frac{\partial \boldsymbol{\Sigma}_k^{(i)}}{\partial \boldsymbol{\mu}^{(i)}} = -\frac{\partial L_k^{(i)}}{\partial Q_{\pi^*(i)}}\,\frac{\partial Q_{\pi^*(i)}}{\partial \boldsymbol{\phi}^{(i)}}\,\frac{\partial \boldsymbol{\phi}^{(i)}}{\partial \boldsymbol{\Sigma}^{(i)}}
\end{aligned}$$

MAK-TD FRAMEWORK

$$
\begin{aligned}
a_k^{(i)} &= \arg\max_a \left(\boldsymbol{h}_k^{(i)}(\boldsymbol{s}_k^{(i)}, a^{(i)}) R^{-1\,(i)} \boldsymbol{h}_k^{T\,(i)}(\boldsymbol{s}_k^{(i)}, a^{(i)}) \right) \\
&= \arg\max_a \left(\boldsymbol{h}_k^{(i)}(\boldsymbol{s}_k^{(i)}, a^{(i)}) \boldsymbol{h}_k^{T\,(i)}(\boldsymbol{s}_k^{(i)}, a^{(i)}) \right).
\end{aligned}
$$

Algorithm 1 THE PROPOSED MAK-TD FRAMEWORK

Learning Phase:

Set $\boldsymbol{\theta}_0, \boldsymbol{P}_{\theta,0}, \boldsymbol{F}, \boldsymbol{\mu}_{n,i_d}, \boldsymbol{\Sigma}_{n,i_d}$ for $n = 1, 2, \ldots, N$ and $i_d = 1, 2, \ldots, D$

Repeat (for each episode):

  Initialize $\boldsymbol{s}_k$

  Repeat (for each agent i):

    While $\boldsymbol{s}_k^{(i)} \neq \boldsymbol{s}_T$ **do**:

      $a_k^{(i)} = \arg\max_a \left(\boldsymbol{h}_k^{(i)}(\boldsymbol{s}_k^{(i)}, a^{(i)}) \boldsymbol{h}_k^{T(i)}(\boldsymbol{s}_k^{(i)}, a^{(i)}) \right)$

      Take action $a_k^{(i)}$, observe $\boldsymbol{s}_{k+1}^{(i)}, r_k^{(i)}$

      Calculate $\boldsymbol{\phi}^{(i)}(\boldsymbol{s}^{(i)}, a^{(i)})$ via Equations (22) and (23)

      $\boldsymbol{h}_k^{(i)}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) = \boldsymbol{\phi}^{(i)}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) - \gamma \arg\max_a \boldsymbol{\phi}^{(i)}(\boldsymbol{s}_{k+1}^{(i)}, a^{(i)})$

      $\hat{\boldsymbol{\theta}}_{(k|k-1)}^{(i)} = \boldsymbol{F}^{(i)} \hat{\boldsymbol{\theta}}_k^{(i)}$

      $\boldsymbol{P}_{(\theta,k|k-1)}^{(i)} = \boldsymbol{F}^{(i)} \boldsymbol{P}_{\theta,k-1}^{(i)} \boldsymbol{F}^{T(i)} + \boldsymbol{Q}^{(i)}$

      for $j = 1 : M$ **do**:

        $\boldsymbol{k}_k^{j\,(i)} = \boldsymbol{P}_{(\theta,k|k-1)}^{(i)} \boldsymbol{h}_k^{(i)} (\boldsymbol{h}_k^{T(i)} \boldsymbol{P}_{(\theta,k|k-1)}^{(i)} \boldsymbol{h}_k^{(i)} + R^{j\,(i)})^{-1}$

        $\hat{\boldsymbol{\theta}}_k^{j\,(i)} = \hat{\boldsymbol{\theta}}_{(\theta,k|k-1)}^{(i)} + \boldsymbol{k}_k^{j\,(i)} (r_k^j - \boldsymbol{h}_k^{T(i)} \hat{\boldsymbol{\theta}}_{(k|k-1)}^{(i)})$

        $\boldsymbol{P}_{\theta,k}^{(i)} = (\boldsymbol{I} - \boldsymbol{K}_k^{j\,(i)} \boldsymbol{h}_k^{T(i)}) \boldsymbol{P}_{(\theta,k|k-1)}^{(i)} (\boldsymbol{I} - \boldsymbol{K}_k^{j\,(i)} \boldsymbol{h}_k^{T(i)})^T + \boldsymbol{K}_k^{j\,(i)} R^j \boldsymbol{K}_k^{j\,T(i)}$

      end for

```
Compute the value of c and w^{j(i)} by using Σ_{j=1}^{M} w^{j(i)} = 1 and Equation (19)
```

Compute the value of c and $w^{j^{(i)}}$ by using $\sum_{j=1}^{M} w^{j^{(i)}} = 1$ and Equation (19)

$\hat{\boldsymbol{\theta}}_k^{(i)} = \sum_{j=1}^{M} w^{j^{(i)}} \hat{\boldsymbol{\theta}}_k^{j^{(i)}}$

$\boldsymbol{P}_{\boldsymbol{\theta}_k}^{(i)} = \sum_{j=1}^{M} \omega^{j^{(i)}} \left(\boldsymbol{P}_{\boldsymbol{\theta},k}^{j^{(i)}} + (\hat{\boldsymbol{\theta}}^{j^{(i)}} - \hat{\boldsymbol{\theta}}^{(i)})(\hat{\boldsymbol{\theta}}^{j^{(i)}} - \hat{\boldsymbol{\theta}}^{(i)})^T \right)$

RBFs Parameters Update:

$L_k^{(i)} = (\boldsymbol{\phi}^T(\boldsymbol{s}_k^{(i)}, a_k)\, \boldsymbol{\theta}_k^{(i)} - r_k^{(i)})^2$

if $L_k^{(i)\frac{1}{2}} (\boldsymbol{\theta}_k^{(i)^T} \boldsymbol{\phi}(\cdot)) > 0$ **then**:

Update $\boldsymbol{\Sigma}_{n,a_d}$ via Equation (29)

else:

Update $\boldsymbol{\mu}_{n,a_d}$ via Equation (30)

end if

end while

Testing Phase:

Repeat (for each trial episode):

While $s_k \neq s_T$ **do**:

Repeat (for each agent):

$a_k = \arg\max_a \boldsymbol{\phi}(s_k, a)^T \boldsymbol{\theta}_k$

Take action a_k, and observe $\boldsymbol{s}_{k+1}, r_k$

Calculate Loss S_k for all agents

End While

MAK-SR FRAMEWORK

$$\boldsymbol{M}_{\pi^{(i)}}(\boldsymbol{s}^{(i)}, :, a^{(i)}) = \mathbb{E}\left[\sum_{k=0}^{T} \gamma^k \boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) | \boldsymbol{s}_0^{(i)} = \boldsymbol{s}^{(i)}, a_0^{(i)} = a^{(i)}\right]$$

$$r^{(i)}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) \approx \boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)})^T \boldsymbol{\theta}_k^{(i)}$$

$$Q(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) = \boldsymbol{\theta}_k^{(i)T} \boldsymbol{M}(\boldsymbol{s}_k^{(i)}, :, a_k^{(i)})$$

$$\boldsymbol{M}_{\pi^{(i)}}(\boldsymbol{s}_k^{(i)}, :, a_k^{(i)}) \approx \boldsymbol{M}_k \boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)})$$

$$\boldsymbol{M}_{\pi^{(i)}}^{\text{new}}(\boldsymbol{s}_k^{(i)}, :, a_k^{(i)}) = \boldsymbol{M}_{\pi^{(i)}}^{\text{old}}(\boldsymbol{s}_k^{(i)}, :, a_k^{(i)}) + \alpha\left(\boldsymbol{\phi}^{(i)}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) + \gamma \boldsymbol{M}_{\pi^{(i)}}(\boldsymbol{s}_{k+1}^{(i)}, :, a_{k+1}^{(i)}) - \boldsymbol{M}_{\pi^{(i)}}^{\text{old}}(\boldsymbol{s}_k^{(i)}, :, a_k^{(i)})\right)$$

MAK-SR FRAMEWORK

$$\boldsymbol{M}_{\pi^{(i)}}(\boldsymbol{s}^{(i)}, :, a^{(i)}) = \mathbb{E}\left[\sum_{k=0}^{T} \gamma^k \boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) | \boldsymbol{s}_0^{(i)} = \boldsymbol{s}^{(i)}, a_0^{(i)} = a^{(i)}\right]$$

$$r^{(i)}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) \approx \boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)})^T \boldsymbol{\theta}_k^{(i)}$$

$$Q(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) = \boldsymbol{\theta}_k^{(i)T} \boldsymbol{M}(\boldsymbol{s}_k^{(i)}, :, a_k^{(i)})$$

$$\boldsymbol{M}_{\pi^{(i)}}(\boldsymbol{s}_k^{(i)}, :, a_k^{(i)}) \approx \boldsymbol{M}_k \boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)})$$

$$\boldsymbol{M}_{\pi^{(i)}}^{\text{new}}(\boldsymbol{s}_k^{(i)}, :, a_k^{(i)}) = \boldsymbol{M}_{\pi^{(i)}}^{\text{old}}(\boldsymbol{s}_k^{(i)}, :, a_k^{(i)}) + \alpha\left(\boldsymbol{\phi}^{(i)}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) + \gamma \boldsymbol{M}_{\pi^{(i)}}(\boldsymbol{s}_{k+1}^{(i)}, :, a_{k+1}^{(i)}) - \boldsymbol{M}_{\pi^{(i)}}^{\text{old}}(\boldsymbol{s}_k^{(i)}, :, a_k^{(i)})\right)$$

MAK-SR FRAMEWORK

$$\boldsymbol{M}_{\pi^{(i)}}(\boldsymbol{s}^{(i)}, :, a^{(i)}) = \mathbb{E}\left[\sum_{k=0}^{T} \gamma^k \boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) | \boldsymbol{s}_0^{(i)} = \boldsymbol{s}^{(i)}, a_0^{(i)} = a^{(i)}\right]$$

$$r^{(i)}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) \approx \boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)})^T \boldsymbol{\theta}_k^{(i)}$$

$$Q(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) = \boldsymbol{\theta}_k^{(i)\,T} \boldsymbol{M}(\boldsymbol{s}_k^{(i)}, :, a_k^{(i)})$$

$$\boldsymbol{M}_{\pi^{(i)}}(\boldsymbol{s}_k^{(i)}, :, a_k^{(i)}) \approx \boldsymbol{M}_k\, \boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)})$$

$$\boldsymbol{M}_{\pi^{(i)}}^{\text{new}}(\boldsymbol{s}_k^{(i)}, :, a_k^{(i)}) = \boldsymbol{M}_{\pi^{(i)}}^{\text{old}}(\boldsymbol{s}_k^{(i)}, :, a_k^{(i)}) + \alpha\left(\boldsymbol{\phi}^{(\boldsymbol{i})}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) + \gamma \boldsymbol{M}_{\pi^{(i)}}(\boldsymbol{s}_{k+1}^{(i)}, :, a_{k+1}^{(i)}) - \boldsymbol{M}_{\pi^{(i)}}^{\text{old}}(\boldsymbol{s}_k^{(i)}, :, a_k^{(i)})\right)$$

MAK-SR FRAMEWORK

$$\boldsymbol{M}_{\pi^{(i)}}(\boldsymbol{s}^{(i)}, :, a^{(i)}) = \mathbb{E}\left[\sum_{k=0}^{T} \gamma^k \boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) | \boldsymbol{s}_0^{(i)} = \boldsymbol{s}^{(i)}, a_0^{(i)} = a^{(i)}\right]$$

$$r^{(i)}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) \approx \boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)})^T \boldsymbol{\theta}_k^{(i)}$$

$$Q(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) = \boldsymbol{\theta}_k^{(i)T} \boldsymbol{M}(\boldsymbol{s}_k^{(i)}, :, a_k^{(i)})$$

$$\boldsymbol{M}_{\pi^{(i)}}(\boldsymbol{s}_k^{(i)}, :, a_k^{(i)}) \approx \boldsymbol{M}_k \boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)})$$

$$\boldsymbol{M}_{\pi^{(i)}}^{\text{new}}(\boldsymbol{s}_k^{(i)}, :, a_k^{(i)}) = \boldsymbol{M}_{\pi^{(i)}}^{\text{old}}(\boldsymbol{s}_k^{(i)}, :, a_k^{(i)}) + \alpha(\boldsymbol{\phi}^{(i)}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) + \gamma \boldsymbol{M}_{\pi^{(i)}}(\boldsymbol{s}_{k+1}^{(i)}, :, a_{k+1}^{(i)}) - \boldsymbol{M}_{\pi^{(i)}}^{\text{old}}(\boldsymbol{s}_k^{(i)}, :, a_k^{(i)}))$$

MAK-SR FRAMEWORK

$$\boldsymbol{M}_{\pi^{(i)}}(\boldsymbol{s}^{(i)}, :, a^{(i)}) = \mathbb{E}\left[\sum_{k=0}^{T} \gamma^k \boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) | \boldsymbol{s}_0^{(i)} = \boldsymbol{s}^{(i)}, a_0^{(i)} = a^{(i)}\right]$$

$$r^{(i)}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) \approx \boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)})^T \boldsymbol{\theta}_k^{(i)}$$

$$Q(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) = \boldsymbol{\theta}_k^{(i)T} \boldsymbol{M}(\boldsymbol{s}_k^{(i)}, :, a_k^{(i)})$$

$$\boldsymbol{M}_{\pi^{(i)}}(\boldsymbol{s}_k^{(i)}, :, a_k^{(i)}) \approx \boldsymbol{M}_k \, \boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)})$$

$$\boldsymbol{M}_{\pi^{(i)}}^{\text{new}}(\boldsymbol{s}_k^{(i)}, :, a_k^{(i)}) = \boldsymbol{M}_{\pi^{(i)}}^{\text{old}}(\boldsymbol{s}_k^{(i)}, :, a_k^{(i)}) + \alpha\left(\boldsymbol{\phi}^{(i)}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) + \gamma \boldsymbol{M}_{\pi^{(i)}}(\boldsymbol{s}_{k+1}^{(i)}, :, a_{k+1}^{(i)}) - \boldsymbol{M}_{\pi^{(i)}}^{\text{old}}(\boldsymbol{s}_k^{(i)}, :, a_k^{(i)})\right)$$

MAK-SR FRAMEWORK

$$\hat{\boldsymbol{\phi}}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) = \boldsymbol{M}^{\text{new}}(\boldsymbol{s}_k^{(i)}, :, a_k^{(i)}) - \gamma \boldsymbol{M}(\boldsymbol{s}_{k+1}^{(i)}, :, a_{k+1}^{(i)}) + \boldsymbol{n}_k^{(i)}$$

$$\hat{\boldsymbol{\phi}}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) = \boldsymbol{M}_k \underbrace{\left[\boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) - \gamma \boldsymbol{\phi}(\boldsymbol{s}_{k+1}^{(i)}, a_{k+1}^{(i)})\right]}_{g_k^{(i)}} + \boldsymbol{n}_k^{(i)}$$

$$\hat{\boldsymbol{\phi}}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) = (\boldsymbol{g}^{(i)}{}_k^T \otimes \boldsymbol{I}) \boldsymbol{m}_k^{(i)} + \boldsymbol{n}_k^{(i)}$$

$$\boldsymbol{m}_{k+1}^{(i)} = \boldsymbol{m}_k^{(i)} + \boldsymbol{\mu}_k^{(i)}$$

MAK-SR FRAMEWORK

$$\hat{\boldsymbol{\phi}}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) = \boldsymbol{M}^{\text{new}}(\boldsymbol{s}_k^{(i)}, :, a_k^{(i)}) - \gamma \boldsymbol{M}(\boldsymbol{s}_{k+1}^{(i)}, :, a_{k+1}^{(i)}) + \boldsymbol{n}_k^{(i)}$$

$$\hat{\boldsymbol{\phi}}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) = \boldsymbol{M}_k \underbrace{\left[\boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) - \gamma \boldsymbol{\phi}(\boldsymbol{s}_{k+1}^{(i)}, a_{k+1}^{(i)})\right]}_{g_k^{(i)}} + \boldsymbol{n}_k^{(i)}$$

$$\hat{\boldsymbol{\phi}}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) = (\boldsymbol{g}^{(i)}{}_k^T \otimes \boldsymbol{I})\boldsymbol{m}_k^{(i)} + \boldsymbol{n}_k^{(i)}$$

$$\boldsymbol{m}_{k+1}^{(i)} = \boldsymbol{m}_k^{(i)} + \boldsymbol{\mu}_k^{(i)}$$

MAK-SR FRAMEWORK

$$\hat{\boldsymbol{\phi}}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) = \boldsymbol{M}^{\text{new}}(\boldsymbol{s}_k^{(i)}, :, a_k^{(i)}) - \gamma \boldsymbol{M}(\boldsymbol{s}_{k+1}^{(i)}, :, a_{k+1}^{(i)}) + \boldsymbol{n}_k^{(i)}$$

$$\hat{\boldsymbol{\phi}}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) = \boldsymbol{M}_k \underbrace{\left[\boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) - \gamma \boldsymbol{\phi}(\boldsymbol{s}_{k+1}^{(i)}, a_{k+1}^{(i)})\right]}_{g_k^{(i)}} + \boldsymbol{n}_k^{(i)}$$

$$\hat{\boldsymbol{\phi}}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) = (\boldsymbol{g}^{(i)}{}_k^T \otimes \boldsymbol{I}) \boldsymbol{m}_k^{(i)} + \boldsymbol{n}_k^{(i)}$$

$$\boldsymbol{m}_{k+1}^{(i)} = \boldsymbol{m}_k^{(i)} + \boldsymbol{\mu}_k^{(i)}$$

MAK-SR FRAMEWORK

$$\hat{\boldsymbol{\phi}}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) = \boldsymbol{M}^{\text{new}}(\boldsymbol{s}_k^{(i)}, :, a_k^{(i)}) - \gamma \boldsymbol{M}(\boldsymbol{s}_{k+1}^{(i)}, :, a_{k+1}^{(i)}) + \boldsymbol{n}_k^{(i)}$$

$$\hat{\boldsymbol{\phi}}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) = \boldsymbol{M}_k \underbrace{\left[\boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) - \gamma \boldsymbol{\phi}(\boldsymbol{s}_{k+1}^{(i)}, a_{k+1}^{(i)})\right]}_{g_k^{(i)}} + \boldsymbol{n}_k^{(i)}$$

$$\hat{\boldsymbol{\phi}}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) = (\boldsymbol{g}^{(i)}{}_k^T \otimes \boldsymbol{I})\boldsymbol{m}_k^{(i)} + \boldsymbol{n}_k^{(i)}$$

$$\boldsymbol{m}_{k+1}^{(i)} = \boldsymbol{m}_k^{(i)} + \boldsymbol{\mu}_k^{(i)}$$

Algorithm 2 THE PROPOSED MAK-SR FRAMEWORK

Learning Phase:
Initialize: $\theta_0, \boldsymbol{P}_{\theta,0}, \boldsymbol{m}_0, \boldsymbol{P}_{M,0}, \boldsymbol{\mu}_n$, and $\boldsymbol{\Sigma}_n$ for $n = 1, 2, \ldots, N$
Parameters: $Q_\theta, Q_M, \lambda_\mu, \lambda_\Sigma$, and $\{R_\theta^j, R_M^j\}$ for $j = 1, 2, \ldots, M$
Repeat (for each episode):
  Initialize $\boldsymbol{s}_k$
  Repeat (for each agent i):
    While $\boldsymbol{s}_k^{(i)} \neq \boldsymbol{s}_T$ **do**:
      Reshape $\boldsymbol{m}_k$ into $L \times L$ to construct 2-D matrix $\boldsymbol{M}_k$.
      $a_k^{(i)} = \arg\max_a \left(\boldsymbol{g}_k^{(i)}(\boldsymbol{s}_k^{(i)}, a) {\boldsymbol{g}_k^{(i)}}^T (\boldsymbol{s}_k^{(i)}, a^{(i)}) \right)$
      Take action $a_k^{(i)}$, observe $\boldsymbol{s}_{k+1}^{(i)}$ and $r_k^{(i)}$.
      Calculate $\boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)})$ via Equations (23) and (25).
      Update reward weights vector: Perform MMAE to update $\boldsymbol{\theta}_k^{(i)}$.
      Update SR weights vector: Perform KF on Equations (40) and (41) to update $\boldsymbol{m}_k^{(i)}$.
      Update RBFs parameters: Perform RGD on the loss function L_k to update $\boldsymbol{\Sigma}_n$ and $\boldsymbol{\mu}_n$.
    end while

Experimental Results

Experimental Results

Experimental Results

Table 1. Total loss averaged across all the episodes and for all the four implemented scenarios.

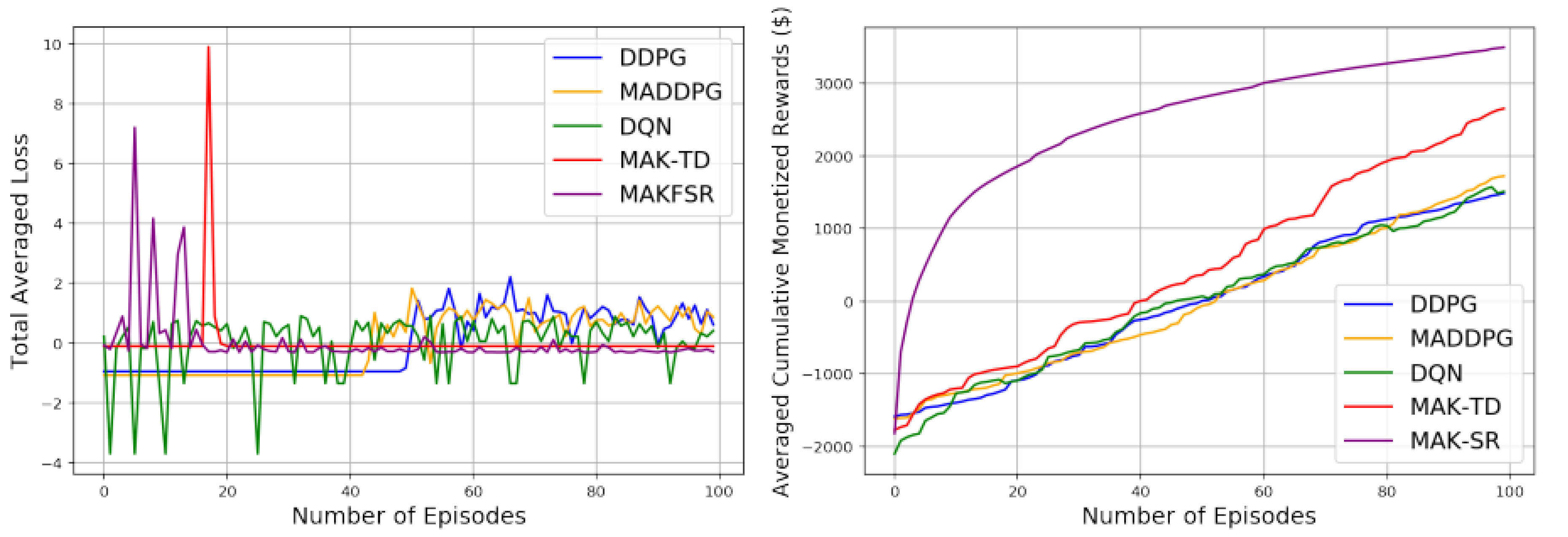
Environment	MAK-SR	MAK-TD	MADDPG	DDPG	DQN
Cooperation	8.93	2.4088	9649.84	10,561.16	10.93
Competition	0.43	4.9301	10,158.18	10,710.37	107.39
Predator–Prey 1v2	0.005	1.9374	6816.34	6884.33	8.21
Predator–Prey 2v1	8.87	1.2421	7390.18	6882.2	10.24

Experimental Results

Table 2. Total received reward by the agents averaged for all the four implemented scenarios.

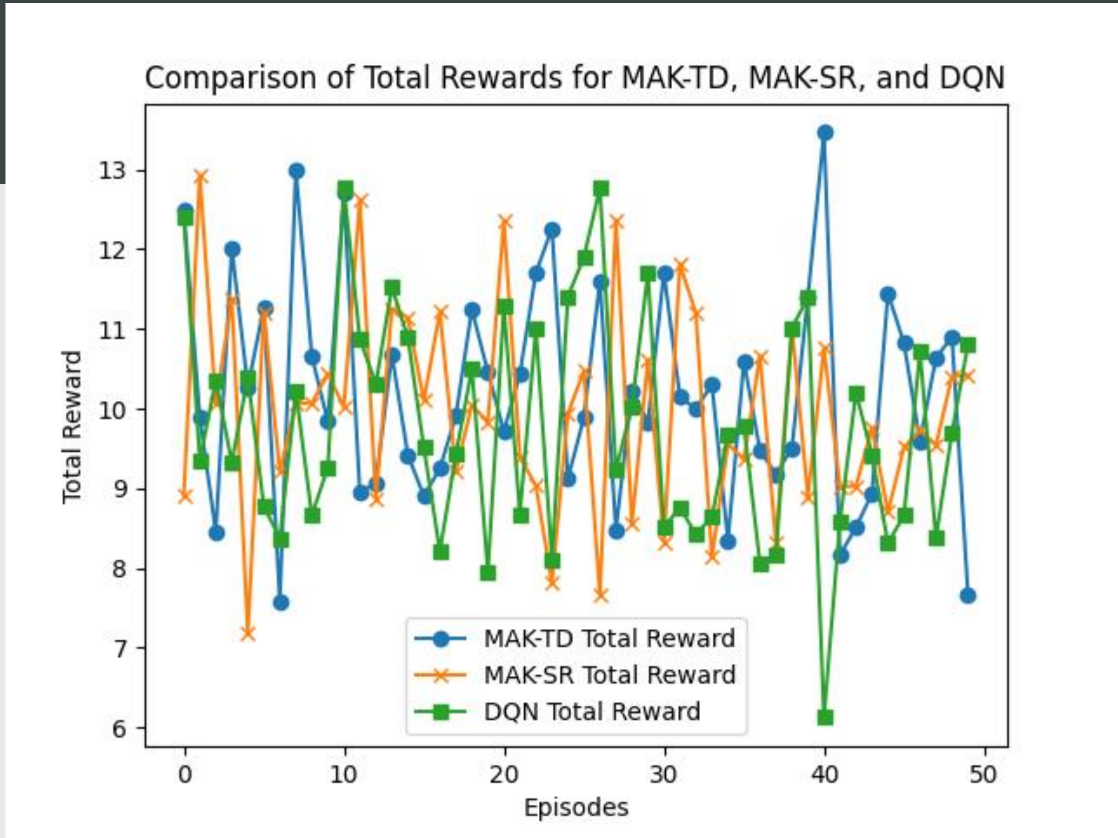
Environment	MAK-SR	MAK-TD	MADDPG	DDPG	DQN
Cooperation	−16.0113	−23.0113	−69.28	−66.29	−39.96
Competition	−0.778	−13.358	−63.30	−61.34	−14.49
Predator–Prey 1v2	−0.0916	−13.432	−46.17	−20.53	−23.451
Predator–Prey 2v1	−0.081	−17.0058	−55.69	−49.41	−44.32

Experimental Results



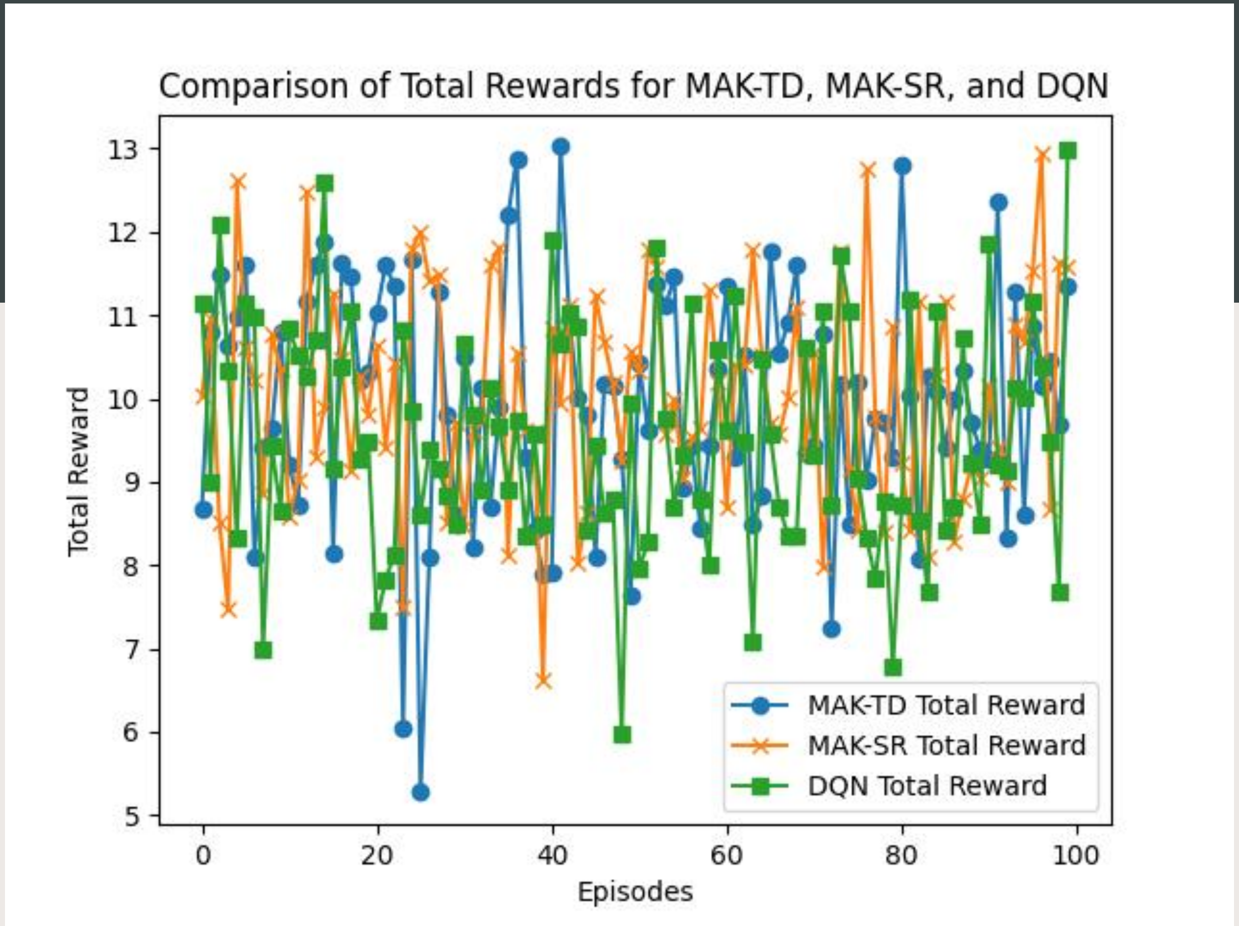
(a) (b)

Simulation Results




```
Final Total Rewards after 50 Episodes:
MAK-TD Total Reward: 510.10918848452957
MAK-SR Total Reward: 498.127397822994
DQN Total Reward: 488.5321211723451
```

Simulation Results



Final Total Rewards after 100 Episodes:
MAK-TD Total Reward: 996.1441199350996
MAK-SR Total Reward: 1003.7453232643612
DQN Total Reward: 955.8240584992651

Key Achievements:

- MAK-TD
- MAK-SR

CONCLUSIONS

THANK YOU!

Winter 2025