Supplement to the paper: Distributed Reinforcement Learning via Aggregative Actor-Critic

Andrea Camisa, Lorenzo Sforni, Giuseppe Notarstefano

Appendix

A. Proof of Proposition 3.2

Fix the initial states to $\bar{x}_i$ and the policy parameters of π_i to some K_i, v_i for all $i \in \mathbb{I}$. Since each system follows the policy π_i such that $u_{i,k} = K_i x_{i,k} + v_i + \eta_{i,k}$, it holds

$$\begin{aligned} x_{i,k+1} &= A_i x_{i,k} + B_i(K_i x_{i,k} + v_i + \eta_{i,k}) + w_{i,k} \\ &= (A_i + B_i K_i) x_{i,k} + B_i v_i + B\eta_{i,k} + w_{i,k}. \end{aligned} \tag{35}$$

The evolution of each system i can be thus written in closed form as

$$x_{i,k} = \underbrace{(A_i+B_iK_i)^k}_{:=\Phi_{i,k}} \bar{x}_i + \underbrace{\sum_{\tau=0}^{k-1} A_i^{k-\tau-1}(B_i v_i + B\eta_{i,\tau} + w_{i,\tau})}_{:=\xi_{i,k}}$$

Similarly we can also express the input as a function of the initial state and of the noise realizations:

$$u_{i,k} = K_i x_{i,k} + v_i + \eta_{i,k} = K_i \Phi_{i,k} \bar{x}_i + \underbrace{K_i \xi_{i,k} + \eta_{i,k}}_{:=\psi_{i,k}}$$

Notice that, since we suppose $\mathbb{E}[w_{i,k}] = 0, \mathbb{E}[\eta_{i,k}] = 0$ and both of them i.i.d. we have

$$\mathbb{E}\left[x_{i,k}\right] = \mathbb{E}\left[\Phi_{i,k}\bar{x}_i\right] + \mathbb{E}\left[\xi_{i,k}\right] = \Phi_{i,k}\bar{x}_i \tag{36}$$

and, similarly,

$$\mathbb{E}\left[u_{i,t}\right] = \mathbb{E}\left[K_i\Phi_{i,k}\bar{x}_i\right] + \mathbb{E}\left[\psi_{i,k}\right] = K_i\Phi_{i,k}\bar{x}_i. \tag{37}$$

For ease of exposition, let us assume that the linear terms in the cost are zero, i.e., that $q_i = 0, r_i = 0, f_i = 0$ (the derivations that follow are similar for the case in which the linear terms are nonzero). Thus we must consider

$$\begin{aligned} J_\pi(x) = \sum_{i=1}^{N} \mathbb{E}\bigg[\frac{1}{2}\sum_{k=0}^{\infty} \alpha^k \big(& x_{i,k}^\top Q_i x_{i,k} + u_{i,k}^\top R_i u_{i,k} \\ &+ \sigma(x_k)^\top F_i \sigma(x_k)\big)\bigg]. \end{aligned} \tag{38}$$

Considering the closed form evolution of each system, exploiting the linearity of the expected value and using the definition of $\sigma(x)$, we obtain

$$\begin{aligned} J_\pi(x) = \sum_{i=1}^{N} \frac{1}{2} \sum_{k=0}^{\infty} \Bigg\{ & \mathbb{E}\bigg[\bar{x}_i^\top \alpha^k \Big(\Phi_{i,k}^\top Q_i \Phi_{i,k} \\ &\quad + \Phi_{i,k}^\top K_i^\top R_i K_i \Phi_{i,k}\Big)\bar{x}_i\bigg] \\ &+ \mathbb{E}\bigg[2\alpha^k\Big(\xi_{i,k}^\top Q_i \Phi_{i,k} + \psi_{i,k}^\top R_i K_i \Phi_{i,k}\Big)\bar{x}_i\bigg] \\ &+ \mathbb{E}\bigg[\alpha^k\Big(\xi_{i,k}^\top Q_i \xi_{i,k} + \psi_{i,k}^\top R_i \psi_{i,k}\Big)\bigg] \\ &+ \frac{1}{N^2}\sum_{j=1}^{N}\sum_{\ell=1}^{N}\Bigg(\mathbb{E}\bigg[\bar{x}_\ell^\top \alpha^k \Big(\Phi_{\ell,k}^\top H_\ell^\top F_i H_j \Phi_{j,k}\Big)\bar{x}_\ell\bigg] \\ &\quad + \mathbb{E}\bigg[2\alpha^k\Big(\xi_{\ell,k}^\top H_\ell^\top F_i H_j \Phi_{j,k}\Big)\bar{x}_j\bigg] \\ &\quad + \mathbb{E}\bigg[\alpha^k\Big(\xi_{\ell,k}^\top H_\ell^\top F_i H_j \xi_{j,k}\Big)\bigg]\Bigg)\Bigg\}. \end{aligned}$$

Then, in light of (36) and (37) and defining

$$\begin{aligned} \tilde{P}_i &:= \sum_{k=0}^{\infty} \alpha^k \Big(\tilde{\Phi}_{i,k}^\top Q_i \tilde{\Phi}_{i,k} + \tilde{\Phi}_{i,k}^\top K_i^\top R_i K_i \tilde{\Phi}_{i,k}\Big) \\ \tilde{S}_i &:= F_i \\ \tilde{\sigma}(\bar{x}) &:= \frac{1}{N}\sum_{i=0}^{N} \underbrace{\sum_{k=0}^{\infty} \sqrt{\alpha^k} H_i \Phi_{i,k}}_{:=\tilde{H}_i} \bar{x}_i \\ \zeta_i &:= \sum_{k=0}^{\infty} \alpha^k \Big(\xi_{i,k}^\top Q_i \xi_{i,k} + \psi_{i,k}^\top R_i \psi_{i,k}\Big) \\ \varsigma &:= \frac{1}{N}\sum_{i=0}^{N}\sum_{k=0}^{\infty} \sqrt{\alpha^k} H_i \xi_{i,k}, \end{aligned}$$

we can finally write:

$$J_\pi(x) = \frac{1}{2}\sum_{i=1}^{N} \big(\bar{x}_i^\top \tilde{P}_i \bar{x}_i + \tilde{\sigma}(\bar{x})^\top \tilde{S}_i \tilde{\sigma}(\bar{x}) + \tilde{\rho}_i\big), \tag{39}$$

with $\tilde{\rho}_i = \mathbb{E}[\zeta_i] + \mathbb{E}[\varsigma^\top \tilde{S}_i \varsigma]$. For the case in which the linear terms are nonzero, there will be additional linear terms in (39). The proof follows. $\square$