Supplementary Material for *Optimization Landscape of Policy Gradient Methods for Discrete-time Static Output Feedback*

Jingliang Duan, Jie Li, Xuyang Chen, Kai Zhao, Shengbo Eben Li, Lin Zhao

S.I. Proof of Lemma 6

The performance difference lemma, also known as almost smoothness, is the basis for deriving the gradient domination condition.

Lemma 1 (Performance difference lemma)**.** Suppose $K, K' \in \mathbb{K}$. It holds that:

$$\begin{aligned}J(K') - J(K) = 2\mathrm{Tr}\big(\Sigma_{K'}(K'C-KC)^\top E_K\big) + \\ + \mathrm{Tr}\big(\Sigma_{K'}(K'C-KC)^\top(R+B^\top P_K B)(K'C-KC)\big).\end{aligned}$$

□

Proof. Let x'_t and u'_t be the state and action sequences generated by K', and $c'_t = x'^\top_t Q x'_t + u'^\top_t R u'_t$. Then, one has

$$\begin{aligned}J(K')-&J(K)\\ &=\mathbb{E}_{x_0\sim\mathcal{D}}\Big[\sum_{t=0}^{\infty} c'_t - V_K(x_0)\Big]\\ &=\mathbb{E}_{x_0\sim\mathcal{D}}\Big[\sum_{t=0}^{\infty}(c'_t + V_K(x'_t) - V_K(x'_t)) - V_K(x_0)\Big]\\ &=\mathbb{E}_{x_0\sim\mathcal{D}}\Big[\sum_{t=0}^{\infty}(c'_t + V_K(x'_{t+1}) - V_K(x'_t))\Big],\end{aligned}$$

where the last step utilizes the fact that $x_0 = x'_0$.

Let $A_K(x_t, K') = c_t + V_K(x_{t+1}) - V_K(x_t)|_{u_t=-K'Cx_t}$, which can be expanded as

$$\begin{aligned}&A_K(x_t,K')\\ &=x_t^\top(Q + C^\top K'^\top R K' C)x_t + x_t^\top \mathcal{A}_{K'}^\top P_K \mathcal{A}_{K'} x_t - V_K(x_t)\\ &=x_t^\top(Q + (K'C-KC+KC)^\top R(K'C-KC+KC))x_t\\ &\quad + x_t^\top(A - B(K'C-KC+KC))^\top P_K(A\\ &\qquad\qquad - B(K'C-KC+KC))x_t - V_K(x_t)\\ &=2x_t^\top(K'C-KC)^\top((R+B^\top P_K B)KC - B^\top P_K A)x_t\\ &\quad + x_t^\top(K'C-KC)^\top(R+B^\top P_K B)(K'C-KC)x_t\\ &=2x_t^\top(K'C-KC)^\top E_K x_t\\ &\quad + x_t^\top(K'C-KC)^\top(R+B^\top P_K B)(K'C-KC)x_t.\end{aligned}$$

Then, we get that

$$\begin{aligned}&J(K') - J(K)\\ &=\mathbb{E}_{x_0\sim\mathcal{D}}\Big[\sum_{t=0}^{\infty} A_K(x'_t, K')\Big]\\ &=\mathbb{E}_{x_0\sim\mathcal{D}}\Big[\sum_{t=0}^{\infty}\Big(2\mathrm{Tr}\big(x'_t x'^\top_t (K'C-KC)^\top E_K\big) + \\ &\quad \mathrm{Tr}\big(x'_t x'^\top_t(K'C-KC)^\top(R+B^\top P_K B)(K'C-KC)\big)\Big)\Big]\\ &=2\mathrm{Tr}\big(\Sigma_{K'}(K'C-KC)^\top E_K\big) + \\ &\quad \mathrm{Tr}\big(\Sigma_{K'}(K'C-KC)^\top(R+B^\top P_K B)(K'C-KC)\big).\end{aligned}$$

□

Next, we show the main proof of Lemma 6.

Proof. Let $X = (R + B^\top P_K B)^{-1} E_K \Sigma_{K'} C^\top \mathcal{L}_{K'}^{-1}$. From Lemma 1, we find that

$$\begin{aligned}&J(K') - J(K)\\ &=2\mathrm{Tr}\big(\Sigma_{K'}(K'C-KC)^\top E_K\big)\\ &\quad + \mathrm{Tr}\big(\Sigma_{K'}(K'C-KC)^\top(R+B^\top P_K B)(K'C-KC)\big)\\ &=\mathrm{Tr}\big(\Sigma_{K'}C^\top(\Delta K + X)^\top(R+B^\top P_K B)(\Delta K + X)C\big)\\ &\quad - \mathrm{Tr}\big(\Sigma_{K'}C^\top \mathcal{L}_{K'}^{-1} C\Sigma_{K'} E_K^\top (R+\\ &\qquad\qquad B^\top P_K B)^{-1} E_K \Sigma_{K'} C^\top \mathcal{L}_{K'}^{-1} C\big)\\ &\geq -\mathrm{Tr}\big(\mathcal{L}_{K'}^{-1} C\Sigma_{K'} E_K^\top (R+B^\top P_K B)^{-1} E_K \Sigma_{K'} C^\top\big),\end{aligned}\tag{S1}$$

where $\Delta K = K' - K$ and the equality holds when $K' = K - X$.

Then, one has

$$\begin{aligned}&J(K) - J(K^*)\\ &\leq \mathrm{Tr}\big(\mathcal{L}_{K^*}^{-1} C\Sigma_{K^*} E_K^\top (R+B^\top P_K B)^{-1} E_K \Sigma_{K^*} C^\top\big)\\ &\leq \|\Sigma_{K^*} C^\top \mathcal{L}_{K^*}^{-1} C\Sigma_{K^*}\| \mathrm{Tr}\big(E_K^\top (R+B^\top P_K B)^{-1} E_K\big)\\ &\leq \|\Sigma_{K^*} C^\top \mathcal{L}_{K^*}^{-1} C\| \|\Sigma_{K^*}\| \mathrm{Tr}\big(E_K^\top (R+B^\top P_K B)^{-1} E_K\big)\\ &\leq \|\Sigma_{K^*}\| \mathrm{Tr}\big(E_K^\top (R+B^\top P_K B)^{-1} E_K\big)\\ &\leq \frac{\|\Sigma_{K^*}\| \mathrm{Tr}\big(E_K^\top E_K\big)}{\sigma_{\min}(R)}.\end{aligned}\tag{S2}$$

From (15), it follows that

$$\begin{aligned}\|\nabla J(K))\|_F^2 &= 4\mathrm{Tr}(C\Sigma_K E_K^\top E_K \Sigma_K C^\top)\\ &\geq 4\mu^2 \sigma_{\min}(C)^2 \mathrm{Tr}(E_K^\top E_K),\ \forall C\in\mathbb{C}.\end{aligned}\tag{S3}$$

By (S2) and (S3), it holds that

$$J(K) - J(K^*) \leq \frac{\|\Sigma_{K^*}\| \|\nabla J(K))\|_F^2}{4\mu^2 \sigma_{\min}(C)^2 \sigma_{\min}(R)}, \ \forall C \in \mathbb{C}. \quad \text{(S4)}$$

Suppose K' satisfies that $K' = K - X$. According to (S1), we get

$$\begin{aligned} J(K) &- J(K^*) \\ &\geq J(K) - J(K') \\ &= \mathrm{Tr}\big(\mathcal{L}_{K'}^{-1} C \Sigma_{K'} E_K^\top (R + B^\top P_K B)^{-1} E_K \Sigma_{K'} C^\top\big) \\ &\geq \frac{\mu \mathrm{Tr}\big(E_K^\top E_K\big)}{\|R + B^\top P_K B\|}, \ \forall C \in \mathbb{C}. \end{aligned} \quad \text{(S5)}$$

In addition, when $C \in \mathbb{C}$, since we can always identity the state x by $x = (C^\top C)^{-1} C y$, it is clear that $J(K^*) = J_s^*$ for every $C \in \mathbb{C}$. By replacing $J(K^*)$ in (S4) and (S5) with J_s^*, we finally complete the proof. $\square$

S.II. Example: Four-dimensional Linear System

Consider a circuit system given in [*S1*] with

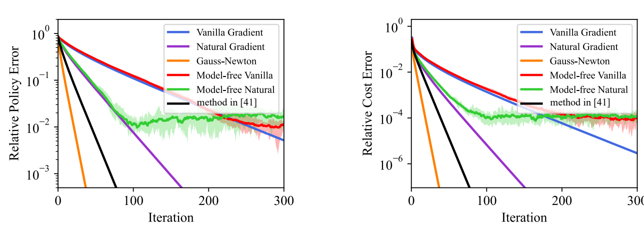
$$A = \begin{bmatrix} 0.90031 & -0.00015 & 0.09048 & -0.00452 \\ -0.00015 & 0.90031 & 0.00452 & -0.09048 \\ -0.09048 & -0.00452 & 0.90483 & -0.09033 \\ 0.00452 & 0.09048 & -0.09033 & 0.90483 \end{bmatrix},$$

$$B = \begin{bmatrix} 0.00468 & -0.00015 \\ 0.00015 & -0.00468 \\ 0.09516 & -0.00467 \\ -0.00467 & 0.09516 \end{bmatrix}, C = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix},$$

where $Q = \mathrm{diag}([0.1, 0.2, 0, 0])$, $R = \mathrm{diag}([10^{-6}, 10^{-4}])$, and $X_0 = 10 I_4$. According to [41, Theorem 1] , the optimal gain is

$$K^* = \begin{bmatrix} 2.9738 & -7.2907 \\ 2.1067 & -12.5384 \end{bmatrix}.$$

We set $K_0 = \begin{bmatrix} 0 & -1 \\ 0 & -2 \end{bmatrix}$ for all methods and adopt the same hyperparameters as in Section VI-B. The relative errors of the control gain and the cost function of different methods are shown in Fig. S1. The observed trend of this example is quite similar to the example given in Section VI-B. Overall, the numerical results are consistent with our convergence analysis.



(a) policy error

(b) cost error

Fig. S1: Learning curves of different methods. The solid lines correspond to the mean and the shaded regions correspond to interval between maximum and minimum values over 10 runs.

References

[S1] F. Lewis, *Applied Optimal Control & Estimation: Digital Design & Implementation*. Prentice Hall, 1992.