ECE586BH: Interplay between Control and Machine Learning **Fall 2023**

Solutions for Homework 1

1. The matrix $\begin{bmatrix} 1-\rho^2 & -\alpha \\ -\alpha & \alpha^2 \end{bmatrix} + \lambda \begin{bmatrix} -2mL & m+L \\ m+L & -2 \end{bmatrix}$ is negative semidefinite if and only if

$$\begin{aligned}
&1-\rho^2-2mL\lambda \leq 0 \\
&\alpha^2-2\lambda \leq 0 \\
&(1-\rho^2-2mL\lambda)(\alpha^2-2\lambda)-(-\alpha+(m+L)\lambda)^2 \geq 0
\end{aligned}$$

Therefore we have $\|x_k - x^*\| \leq \rho^k \|x_0 - x^*\|$ if we can find $0<\rho<1$ and $\lambda \geq 0$ satisfying the above inequalities.

There are two cases. In the first case, we assume $\alpha^2 - 2\lambda = 0$. Then we must have $-\alpha + (m+L)\lambda = 0$ and $\lambda = \frac{1}{2}\alpha^2$. Hence we have $-2\alpha + (m+L)\alpha^2$. This gives $\alpha = \frac{2}{m+L}$ and $\lambda = \frac{2}{(m+L)^2}$. The smallest ρ satisfying $1-\rho^2-2mL\lambda \leq 0$ is given by $\rho = \sqrt{1-2mL\lambda} = \frac{L-m}{m+L}$, which satisfies the formula in the problem statement.

Now we discuss the second case and assume $\alpha^2 - 2\lambda \neq 0$. Then the above inequalities are equivalent to

$$\begin{aligned}
&\rho^2 \geq 1 - 2mL\lambda - \frac{(\lambda(m+L)-\alpha)^2}{\alpha^2-2\lambda} \\
&\lambda > \frac{\alpha^2}{2}
\end{aligned}$$

Now set $\lambda = \frac{1+t}{2}\alpha^2$ with some $t>0$. Clearly $\lambda > \frac{\alpha^2}{2}$. Substituting $\lambda = \frac{1+t}{2}\alpha^2$ to the first inequality $\rho^2 \geq 1 - 2mL\lambda - \frac{(\lambda(m+L)-\alpha)^2}{\alpha^2-2\lambda}$ leads to the following inequality

$$\begin{aligned}
\rho^2 &\geq 1 - mL(1+t)\alpha^2 + \frac{((1+t)\alpha(m+L)-2)^2}{4t} \\
&= 1 - mL\alpha^2 - mL\alpha^2 t + \frac{(t\alpha(m+L)+\alpha(m+L)-2)^2}{4t} \\
&= 1 - mL\alpha^2 - mL\alpha^2 t + \frac{\alpha^2(m+L)^2t^2 + 2(\alpha(m+L)-2)(m+L)\alpha t + (\alpha(m+L)-2)^2}{4t} \\
&= 1 + \frac{\alpha^2(m^2+L^2)}{2} - (m+L)\alpha + \frac{(L-m)^2\alpha^2 t}{4} + \frac{(\alpha(m+L)-2)^2}{4t}
\end{aligned}$$

We want to choose the smallest ρ and associated λ that satisfy the above inequality. As long as $\alpha \neq \frac{2}{m+L}$, we can choose positive t satisfying $\frac{(L-m)^2\alpha^2 t}{4} = \frac{(\alpha(m+L)-2)^2}{4t}$ and ρ satisfying

$$\rho^2 = 1 + \frac{\alpha^2(m^2+L^2)}{2} - (m+L)\alpha + \frac{1}{2}((L-m)\alpha)\sqrt{(\alpha(m+L)-2)^2}$$

When $\alpha < \frac{2}{m+L}$, we have

$$\begin{aligned}\rho^2 &= 1 + \frac{\alpha^2(m^2+L^2)}{2} - (m+L)\alpha + \frac{1}{2}(L-m)\alpha(2-\alpha(m+L)) \\ &= 1 - 2m\alpha + m^2\alpha^2 \\ &= (1-m\alpha)^2\end{aligned}$$

Similarly, we have $\rho^2 = (1-L\alpha)^2$ when $\alpha > \frac{2}{m+L}$. This is equivalent to $\rho = \max\{|1-m\alpha|, |1-L\alpha|\}$ for $\alpha \neq \frac{2}{m+L}$.

Combining the results for the above two cases leads to the desired conclusion.

(Some extra explanation: Also notice that ρ^2 is required to be greater than 0 and smaller than 1, hence the formulas for ρ^2 only work for $\alpha < \frac{2}{L}$. Otherwise $(1-L\alpha)^2 \geq 1$ when $L\alpha \geq 2$. This explains why we require $\alpha < \frac{2}{L}$ in the problem statement.)

2

(a) Substituting $v_k = (1+\beta)x_k - \beta x_{k-1}$ and $x_{k+1} = (1+\beta)x_k - \beta x_{k-1} - \alpha\nabla f(v_k)$, we have

$$\begin{aligned}&\nabla f(v_k)^\mathsf{T}(x_k - v_k) + \frac{m}{2}\|x_k - v_k\|^2 + \nabla f(v_k)^\mathsf{T}(v_k - x_{k+1}) - \frac{L}{2}\|v_k - x_{k+1}\|^2 \\ =&\beta\nabla f(v_k)^\mathsf{T}(x_{k-1} - x_k) + \frac{m\beta^2}{2}\|x_{k-1} - x_k\|^2 + \alpha\|\nabla f(v_k)\|^2 - \frac{L\alpha^2}{2}\|\nabla f(v_k)\|^2 \\ =&\begin{bmatrix} x_k - x^* \\ x_{k-1} - x^* \\ \nabla f(v_k)\end{bmatrix}^\mathsf{T}\left(\frac{1}{2}\begin{bmatrix}\beta^2 m & -\beta^2 m & -\beta \\ -\beta^2 m & \beta^2 m & \beta \\ -\beta & \beta & \alpha(2-L\alpha)\end{bmatrix}\otimes I\right)\begin{bmatrix} x_k - x^* \\ x_{k-1} - x^* \\ \nabla f(v_k)\end{bmatrix}\end{aligned}$$

Therefore, we have

$$X_1 = \frac{1}{2}\begin{bmatrix}\beta^2 m & -\beta^2 m & -\beta \\ -\beta^2 m & \beta^2 m & \beta \\ -\beta & \beta & \alpha(2-L\alpha)\end{bmatrix}\otimes I.$$

(b) Substituting $v_k = (1+\beta)x_k - \beta x_{k-1}$ and $x_{k+1} = (1+\beta)x_k - \beta x_{k-1} - \alpha\nabla f(v_k)$, we have

$$\begin{aligned}&\nabla f(v_k)^\mathsf{T}(x^* - v_k) + \frac{m}{2}\|x^* - v_k\|^2 + \nabla f(v_k)^\mathsf{T}(v_k - x_{k+1}) - \frac{L}{2}\|v_k - x_{k+1}\|^2 \\ =& -\nabla f(v_k)^\mathsf{T}((1+\beta)(x_k - x^*) - \beta(x_{k-1} - x^*)) + \frac{m}{2}\|(1+\beta)(x_k - x^*) - \beta(x_{k-1} - x^*)\|^2 \\ &+ \alpha\|\nabla f(v_k)\|^2 - \frac{L\alpha^2}{2}\|\nabla f(v_k)\|^2 \\ =&\begin{bmatrix} x_k - x^* \\ x_{k-1} - x^* \\ \nabla f(v_k)\end{bmatrix}^\mathsf{T}\left(\frac{1}{2}\begin{bmatrix}(1+\beta)^2 m & -\beta(1+\beta)m & -(1+\beta) \\ -\beta(1+\beta)m & \beta^2 m & \beta \\ -(1+\beta) & \beta & \alpha(2-L\alpha)\end{bmatrix}\otimes I\right)\begin{bmatrix} x_k - x^* \\ x_{k-1} - x^* \\ \nabla f(v_k)\end{bmatrix}\end{aligned}$$

Therefore, we have

$$X_2 = \frac{1}{2}\begin{bmatrix} (1+\beta)^2 m & -\beta(1+\beta)m & -(1+\beta) \\ -\beta(1+\beta)m & \beta^2 m & \beta \\ -(1+\beta) & \beta & \alpha(2-L\alpha) \end{bmatrix} \otimes I.$$

(c) Now it is straightforward to verify that the following holds

$$\begin{bmatrix} A^\mathsf{T} PA - \rho^2 P & A^\mathsf{T} PB \\ B^\mathsf{T} PA & B^\mathsf{T} PB \end{bmatrix} - X = \frac{\sqrt{m}(\sqrt{L}-\sqrt{m})^3}{2(L+\sqrt{Lm})}\begin{bmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \otimes I \leq 0$$

This above fact can be verified using Matlab symbolic toolbox.

3

(a) The policy gradient theorem states the following result:

$$\nabla \mathcal{C}(\theta) = \mathbb{E}\left[\Big(\sum_{t=0}^{N} \gamma^t c(x_t,u_t)\Big)\Big(\sum_{k=0}^{N} \nabla_\theta \log \pi(u_k|x_k)\Big)\right] \quad \text{with large } N$$

Or

$$\nabla \mathcal{C}(\theta) = \mathbb{E}\sum_{t=0}^{\infty}\left[\gamma^t \Big(\sum_{k=t}^{\infty} \gamma^{k-t} c(x_k,u_k)\Big)\nabla_\theta \log \pi_\theta(u_t|x_t)\right]$$

Or

$$\nabla \mathcal{C}(\theta) = \mathbb{E}\sum_{t=0}^{\infty}[\gamma^t \Psi_t \nabla_\theta \log \pi_\theta(u_t|x_t)]$$

where Ψ_t can be chosen as Monte Carlo estimation $\sum_{t'=t}^{\infty} \gamma^{t'-t} c_{t'}$ or Baseline variant $\sum_{t'=t}^{\infty}(\gamma^{t'-t} c_{t'} - b(x_t))$ or other variants.

It is OK to provide any one of the above answers.

(b) We have

$$\begin{aligned} \mathbb{E}_{\varepsilon\sim\mathcal{N}(0,I)}\left(\lim_{\sigma\to 0}\frac{\mathcal{C}(K+\sigma\varepsilon)-\mathcal{C}(K)}{\sigma}\right)\varepsilon &= \mathbb{E}_{\varepsilon\sim\mathcal{N}(0,I)}(\varepsilon^\mathsf{T}\nabla\mathcal{C}(K))\varepsilon \\ &= \mathbb{E}_{\varepsilon\sim\mathcal{N}(0,I)}\varepsilon(\varepsilon^\mathsf{T}\nabla(K)) \\ &= \mathbb{E}_{\varepsilon\sim\mathcal{N}(0,I)}(\varepsilon\varepsilon^\mathsf{T})\nabla(K) \\ &= \nabla C(K) \end{aligned}$$

4

(a) Under the state feedback policy $u(t) = Kx(t)$, the closed-loop system becomes $\dot{x}(t) = (A+BK)x(t)$. Therefore, we have:

$$x(t) = e^{(A+BK)t}x(0).$$

Substituting $u(t) = Kx(t)$ into the cost function gives:

$$\begin{aligned}\mathcal{C}(K) &= \mathbb{E}_{x(0)\sim\mathcal{D}} \int_0^\infty x(t)^\mathsf{T}(Q + K^\mathsf{T}RK)x(t)dt \\ &= \mathbb{E}_{x(0)\sim\mathcal{D}} \int_0^\infty x(0)^\mathsf{T}(e^{(A+BK)t})^\mathsf{T}(Q + K^\mathsf{T}RK)e^{(A+BK)t}x(0)dt \\ &= \mathbb{E}_{x(0)\sim\mathcal{D}}\, x(0)^\mathsf{T}\left[\int_0^\infty (e^{(A+BK)t})^\mathsf{T}(Q + K^\mathsf{T}RK)e^{(A+BK)t}dt\right]x(0).\end{aligned}$$

Denote $P_K := \int_0^\infty (e^{(A+BK)t})^\mathsf{T}(Q + K^\mathsf{T}RK)e^{(A+BK)t}dt$. Since $A + BK$ is Hurwitz, P_K solves the following Lyapunov equation:

$$(A+BK)^\mathsf{T}P_K + P_K(A+BK) + Q + K^\mathsf{T}RK = 0. \tag{1}$$

Therefore, we have:

$$\mathcal{C}(K) = \mathbb{E}_{x(0)\sim\mathcal{D}}\, x(0)^\mathsf{T}P_Kx(0) = \mathrm{trace}(P_K\Sigma_0).$$

(b) Applying chain rule on both sides of (1) gives:

$$\begin{aligned}&(BdK)^\mathsf{T}P_K + (A+BK)^\mathsf{T}dP_K + dP_K(A+BK) + P_K(BdK) + dK^\mathsf{T}RK + K^\mathsf{T}RdK = 0 \\ \iff & dP_K(A+BK) + (A+BK)^\mathsf{T}dP_K + dK^\mathsf{T}(B^\mathsf{T}P_K + RK) + (P_KB + K^\mathsf{T}R)dK^\mathsf{T} = 0.\end{aligned}$$

If we view dP_K as the variable, denote $E_K := B^\mathsf{T}P_K + RK$, the above equation is a Lyapunov equation which can be solved as

$$dP_K = \int_0^\infty (e^{(A+BK)t})^\mathsf{T}(dK^\mathsf{T}E_K + E_K^\mathsf{T}dK)e^{(A+BK)t}dt.$$

By definition, we have $d\mathcal{C}(K) = \mathrm{trace}(\nabla C(K)dK^\mathsf{T})$. On the other hand, we have

$$\begin{aligned}d\mathcal{C}(K) &= \mathrm{trace}(dP_K\Sigma_0) \\ &= \mathrm{trace}\left(\int_0^\infty (e^{(A+BK)t})^\mathsf{T}(dK^\mathsf{T}E_K + E_K^\mathsf{T}dK)e^{(A+BK)t}dt\Sigma_0\right) \\ &= \mathrm{trace}\left(\int_0^\infty x(0)^\mathsf{T}(e^{(A+BK)t})^\mathsf{T}(dK^\mathsf{T}E_K + E_K^\mathsf{T}dK)e^{(A+BK)t}x(0)dt\right) \\ &= \mathrm{trace}\left((dK^\mathsf{T}E_K + E_K^\mathsf{T}dK)\int_0^\infty x(0)^\mathsf{T}(e^{(A+BK)t})^\mathsf{T}e^{(A+BK)t}x(0)dt\right) \\ &= \mathrm{trace}(2E_K\Sigma_KdK^\mathsf{T}),\end{aligned}$$

where $\Sigma_K = \int_0^\infty x(0)^\mathsf{T} (e^{(A+BK)t})^\mathsf{T} e^{(A+BK)t} x(0) dt$. Therefore, we have $\nabla \mathcal{C}(K) = 2E_K \Sigma_K$.

(c) Since $\mathcal{C}(K)$ is L-smooth, we have:

$$\begin{aligned}
\mathcal{C}(K_l) &\leq \mathcal{C}(K_{l-1}) + \langle \nabla C(K_{l-1}), K_l - K_{l-1} \rangle + \frac{L}{2} \|K_l - K_{l-1}\|_F^2 \\
&\leq \mathcal{C}(K_{l-1}) + \langle \nabla C(K_{l-1}), -\alpha \nabla C(K_{l-1}) \rangle + \frac{L}{2} \|\alpha \nabla C(K_{l-1})\|_F^2 \\
&\leq \mathcal{C}(K_{l-1}) - (\alpha - \frac{L}{2}\alpha^2) \|\nabla C(K_{l-1})\|_F^2
\end{aligned}$$

Since $\alpha \in (0, \frac{2}{L})$, we have $\alpha - \frac{L}{2}\alpha^2 > 0$. By gradient dominance property, we have:

$$-\|\nabla C(K_{l-1})\|_F^2 \leq -2\mu(\mathcal{C}(K_{l-1}) - \mathcal{C}(K^*)).$$

Combining the above two inequalities yields:

$$\begin{aligned}
\mathcal{C}(K_l) - \mathcal{C}(K^*) &\leq \mathcal{C}(K_{l-1}) - \mathcal{C}(K^*) - \mu(2\alpha - L\alpha^2)(\mathcal{C}(K_{l-1}) - \mathcal{C}(K^*)) \\
&= (1 - 2\mu\alpha + \mu L \alpha^2)(\mathcal{C}(K_{l-1}) - \mathcal{C}(K^*)).
\end{aligned}$$

Applying the above inequality iteratively gives:

$$\mathcal{C}(K_l) - \mathcal{C}(K^*) \leq (1 - 2\mu\alpha + \mu L \alpha^2)^l (\mathcal{C}(K_0) - \mathcal{C}(K^*)).$$