ECE598ICM: Interplay between Control and Machine Learning **Fall 2020**

Solutions for Homework 1

1.

(a) A sample code is provided. If we enforce $P \geq 10^{-3}I$ and $A^\mathsf{T}PA - P \leq -10^{-3}I$, we get

$$P = \begin{bmatrix} 7.9754 & 1.6911 & -1.4204 \\ 1.6911 & 11.2306 & 3.2385 \\ -1.4204 & 3.2385 & 14.2761 \end{bmatrix}$$

If we enforce the trace of P to be 1, we get

$$P = \begin{bmatrix} 0.2606 & 0.0654 & -0.0917 \\ 0.0654 & 0.3146 & 0.0661 \\ -0.0917 & 0.0661 & 0.4249 \end{bmatrix}$$

You can double check that the above values of P are indeed two feasible solutions for the LMI in the problem statement.

(b) The spectral radius of A is 0.97293. Now if we test the LMI with $\rho = 0.97293$ and break the homogeneity by setting $\varepsilon = 0.001$, we can get

$$P = \begin{bmatrix} 173.3426 & 74.6824 & -147.0031 \\ 74.6824 & 60.9757 & -53.3249 \\ -147.0031 & -53.3249 & 161.1627 \end{bmatrix}$$

You can double check that the above P is indeed a feasible solution for the original LMI. If we test the LMI with $\rho = 0.97292$, the LMI becomes infeasible. Hence the smallest value of ρ for the LMI is the same as the spectral radius of A.

(c) A sample code is provided. Since the LMI condition is linear in both ρ^2 and λ, we can choose a new variable $r_2 = \rho^2$ and just minimizes the LMI over r_2. We can find the value of r_2 is always extremely closed to $\max\{|1 - m\alpha\|, \|1 - L\alpha\|\}$. When L/m is large, the problem becomes ill-conditioned and the value of ρ is extremely close to 1.

2

(a) For any matrix M, we have $M \leq 0$ if and only if $M \otimes I \leq 0$. Therefore, the LMI condition (1) in the problem statement is feasible if and only if the following condition is feasible

$$\begin{bmatrix} (1-\rho^2)I & -\alpha I \\ -\alpha I & \alpha^2 I \end{bmatrix} - \lambda_1 \begin{bmatrix} -2L^2 I & 0 \\ 0 & I \end{bmatrix} - \lambda_2 \begin{bmatrix} 2mI & -I \\ -I & 0 \end{bmatrix} \leq 0$$

We can left and right multiply the above condition with $\begin{bmatrix} x_k - x^* \\ w_k \end{bmatrix}^\mathsf{T}$ and $\begin{bmatrix} x_k - x^* \\ w_k \end{bmatrix}$. This leads to

$$\begin{bmatrix} x_k - x^* \\ w_k \end{bmatrix}^\mathsf{T} \left(\begin{bmatrix} (1-\rho^2)I & -\alpha I \\ -\alpha I & \alpha^2 I \end{bmatrix} - \lambda_1 \begin{bmatrix} -2L^2 I & 0 \\ 0 & I \end{bmatrix} - \lambda_2 \begin{bmatrix} 2mI & -I \\ -I & 0 \end{bmatrix} \right) \begin{bmatrix} x_k - x^* \\ w_k \end{bmatrix} \leq 0$$

Substituting the fact $\|x_{k+1} - x^*\|^2 - \rho^2 \|x_k - x^*\|^2 = \begin{bmatrix} x_k - x^* \\ w_k \end{bmatrix}^\mathsf{T} \begin{bmatrix} (1-\rho^2)I & -\alpha I \\ -\alpha I & \alpha^2 I \end{bmatrix} \begin{bmatrix} x_k - x^* \\ w_k \end{bmatrix}$ into the above inequality, we get

$$\begin{aligned} &\|x_{k+1} - x^*\|^2 - \rho^2 \|x_k - x^*\|^2 \leq \\ &\lambda_1 \begin{bmatrix} x_k - x^* \\ w_k \end{bmatrix}^\mathsf{T} \begin{bmatrix} -2L^2 I & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} x_k - x^* \\ w_k \end{bmatrix} + \lambda_2 \begin{bmatrix} x_k - x^* \\ w_k \end{bmatrix}^\mathsf{T} \begin{bmatrix} 2mI & -I \\ -I & 0 \end{bmatrix} \begin{bmatrix} x_k - x^* \\ w_k \end{bmatrix} \end{aligned}$$

Now we can take expectation of the above inequality and apply the two supply rate conditions given in the problem statement to show

$$\begin{aligned} \mathbb{E}\|x_{k+1} - x^*\|^2 &\leq \rho^2 \mathbb{E}\|x_k - x^*\|^2 + \lambda_1 M \\ &\leq \rho^4 \mathbb{E}\|x_{k-1} - x^*\|^2 + (1+\rho^2)\lambda_1 M \\ &\leq \rho^{2k} \mathbb{E}\|x_0 - x^*\| + \left(\sum_{t=0}^{\infty} \rho^{2t} \right) \lambda_1 M \\ &= \rho^{2k} \mathbb{E}\|x_0 - x^*\| + \frac{\lambda_1 M}{1-\rho^2} \end{aligned}$$

This completes the proof.

(b) We can choose $\lambda_1 = \alpha^2$ and $\lambda_2 = \alpha$ to make the LMI condition (1) feasible. In this case, the left side of the LMI condition (1) becomes a zero matrix. Then the desired conclusion directly follows.

(c) A matrix M is positive semidefinite if and only if $M \otimes I_p \geq 0$. Therefore, we can get rid of the Kronecker product with I_p in our LMI implementation. For SAGA, we can set the matrices as

$$A_i = \begin{bmatrix} I_n - e_i e_i^\mathsf{T} & 0_{n\times 1} \\ -\frac{\alpha}{n}(e - n e_i)^\mathsf{T} & 1 \end{bmatrix}, \; B_i = \begin{bmatrix} e_i e_i^\mathsf{T} \\ -\alpha e_i^\mathsf{T} \end{bmatrix}, \; C = \begin{bmatrix} 0_{1\times n} & 1 \end{bmatrix}$$

In addition, we choose X_0 as

$$X_0 = \begin{bmatrix} C & 0_{1\times 20} \\ 0_{20\times 21} & I_{n\times n} \end{bmatrix}^\mathsf{T} \begin{bmatrix} 1 & 0_{1\times n} \\ 0 & \frac{1}{n} e^\mathsf{T} \end{bmatrix}^\mathsf{T} \begin{bmatrix} 2mL & -(m+L) \\ -(m+L) & 2 \end{bmatrix} \begin{bmatrix} 1 & 0_{1\times n} \\ 0 & \frac{1}{n} e^\mathsf{T} \end{bmatrix} \begin{bmatrix} C & 0_{1\times 20} \\ 0_{20\times 21} & I_{n\times n} \end{bmatrix}$$

For $j = 1, \dots, n$, we choose X_j as

$$X_j = \begin{bmatrix} C & 0_{1\times 20} \\ 0_{20\times 21} & I_{n\times n} \end{bmatrix}^\mathsf{T} \begin{bmatrix} 1 & 0_{1\times n} \\ 0 & e_j^\mathsf{T} \end{bmatrix}^\mathsf{T} \begin{bmatrix} 2mL & -(m+L) \\ -(m+L) & 2 \end{bmatrix} \begin{bmatrix} 1 & 0_{1\times n} \\ 0 & e_j^\mathsf{T} \end{bmatrix} \begin{bmatrix} C & 0_{1\times 20} \\ 0_{20\times 21} & I_{n\times n} \end{bmatrix}$$

Next, we implement the following LMI with $\alpha = \frac{1}{3L}$ and $\rho^2 = 1 - \min\{\frac{1}{3n}, \frac{m}{3L}\}$:

$$\frac{1}{n}\sum_{i=1}^{n} \begin{bmatrix} A_i^\mathsf{T} P A_i - \rho^2 P & A_i^\mathsf{T} P B_i \\ B_i^\mathsf{T} P A_i & B_i^\mathsf{T} P B_i \end{bmatrix} \leq \sum_{j=0}^{n} X_j$$

We try both $(m, L) = (1, 10, 20)$ and $(m, L, n) = (1, 100, 20)$. The above LMI is always feasible. Then enforce P to be a diagonal matrix. Set $\lambda_0 = 0$ and $\lambda_j = \lambda$ for all $1 \leq j \leq n$. The LMI is still feasible.

Actually, one can enforce $P = \begin{bmatrix} \frac{2}{3L} I_{n\times n} & 0_{n\times 1} \\ 0_{1\times n} & \frac{1}{\alpha} \end{bmatrix}$ and $\lambda = \frac{1}{Ln}$. The LMI is still feasible. Based on these parameters, one can even get an analytical proof for the convergence rate of SAGA. A sample code is also provided for demonstrations.

3

(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) A sample code is provided. Notice $M \leq 0$ if and only if $M \otimes I \leq 0$. Hence we can get rid of the Kronecker product in the LMI implementation. The resultant LMI is 3×3. From the numerical solution, we can see that P looks like a matrix with rank 1. For example, if we choose $m = 1$ and $L = 100$ in the code, the value of P is

$$
P = \begin{bmatrix} 50 & -45 \\ -45 & 40.5 \end{bmatrix}
$$

The rank of this matrix is 1. The left side of the LMI also has a pattern. For $m = 1$ and $L = 100$, we have

$$
\begin{bmatrix} A^\mathsf{T} P A - \rho^2 P & A^\mathsf{T} P B \\ B^\mathsf{T} P A & B^\mathsf{T} P B \end{bmatrix} - X = 3.3136 \begin{bmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}
$$

Actually after trying different values of (m, L), we can always find the following pattern:

$$
\begin{bmatrix} A^\mathsf{T} P A - \rho^2 P & A^\mathsf{T} P B \\ B^\mathsf{T} P A & B^\mathsf{T} P B \end{bmatrix} - X = c \begin{bmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \leq 0
$$

where c is some positive constant. If we can figure P and c, then we are done with the convergence rate proof.

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

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

This above fact can be verified using Matlab symbolic toolbox.