**ECE 586 BH: Interplay between Control and Machine Learning** **Fall 2023**

# Homework 1

Instructor: Bin Hu  Due date: September 21, 2023

**1**. (20 points) Consider the gradient method $x_{k+1} = x_k - \alpha \nabla f(x_k)$ and assume $f$ is $L$-smooth and $m$-strongly convex. In the class, we have shown that the gradient method converges at the rate $\rho$ if there exists non-negative $\lambda$ such that

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

Your task is to apply the above condition to show that the gradient descent method with any constant stepsize $0 < \alpha < \frac{2}{L}$ converges at the rate $\rho = \max\{|1 - m\alpha|, |1 - L\alpha|\}$. Write out the proof. What is the choice of $\lambda$?

**2**. (30 points) In this problem, you will be asked to perform several calculations, and these calculations eventually lead to the convergence rate proof for Nesterov's accelerated method applied to smooth strongly-convex objective functions. Recall Nesterov's method is defined by the following recursion:

$$x_{k+1} = x_k + \beta(x_k - x_{k-1}) - \alpha \nabla f((1+\beta)x_k - \beta x_{k-1})$$

which can also be written as

$$\begin{aligned} \xi_{k+1} &= A\xi_k + Bw_k \\ v_k &= C\xi_k \\ w_k &= \nabla f(v_k) \end{aligned}$$

where $A = \begin{bmatrix} (1+\beta)I & -\beta I \\ I & 0 \end{bmatrix}$, $B = \begin{bmatrix} -\alpha I \\ 0 \end{bmatrix}$, $C = \begin{bmatrix} (1+\beta)I & -\beta I \end{bmatrix}$, and $\xi_k = \begin{bmatrix} x_k \\ x_{k-1} \end{bmatrix}$.

(a) Assume $f$ is $L$-smooth and $m$-strongly convex. By $L$-smoothness and $m$-strong convexity, we have

$$\begin{aligned} f(x_k) - f(x_{k+1}) &= f(x_k) - f(v_k) + f(v_k) - f(x_{k+1}) \\ &\geq \nabla f(v_k)^\mathsf{T}(x_k - v_k) + \frac{m}{2}\|x_k - v_k\|^2 + \nabla f(v_k)^\mathsf{T}(v_k - x_{k+1}) - \frac{L}{2}\|v_k - x_{k+1}\|^2 \\ &= \begin{bmatrix} x_k - x^* \\ x_{k-1} - x^* \\ \nabla f(v_k) \end{bmatrix}^\mathsf{T} X_1 \begin{bmatrix} x_k - x^* \\ x_{k-1} - x^* \\ \nabla f(v_k) \end{bmatrix} \end{aligned}$$

The last step in the above derivation requires substituting $x_{k+1} = (1+\beta)x_k - \beta x_{k-1} - \alpha\nabla f(v_k)$ and $v_k = C\xi_k$ into the second-to-last line $\nabla f(v_k)^\mathsf{T}(x_k - v_k) + \frac{m}{2}\|x_k - v_k\|^2 + \nabla f(v_k)^\mathsf{T}(v_k - x_{k+1}) - \frac{L}{2}\|v_k - x_{k+1}\|^2$ and rewriting the resultant quadratic function. You task is figuring out this symmetric matrix $X_1$ (actually this matrix has already been given in the lecture note).

(b) Similarly, by $L$-smoothness and $m$-strong convexity, we have

$$\begin{aligned} f(x^*) - f(x_{k+1}) &= f(x^*) - f(v_k) + f(v_k) - f(x_{k+1}) \\ &\geq \nabla f(v_k)^\mathsf{T}(x^* - v_k) + \frac{m}{2}\|x^* - v_k\|^2 + \nabla f(v_k)^\mathsf{T}(v_k - x_{k+1}) - \frac{L}{2}\|v_k - x_{k+1}\|^2 \\ &= \begin{bmatrix} x_k - x^* \\ x_{k-1} - x^* \\ \nabla f(v_k) \end{bmatrix}^\mathsf{T} X_2 \begin{bmatrix} x_k - x^* \\ x_{k-1} - x^* \\ \nabla f(v_k) \end{bmatrix} \end{aligned}$$

The last step in the above derivation requires substituting $x_{k+1} = (1+\beta)x_k - \beta x_{k-1} - \alpha\nabla f(v_k)$ and $v_k = C\xi_k$ into the second-to-last line $\nabla f(v_k)^\mathsf{T}(x^* - v_k) + \frac{m}{2}\|x^* - v_k\|^2 + \nabla f(v_k)^\mathsf{T}(v_k - x_{k+1}) - \frac{L}{2}\|v_k - x_{k+1}\|^2$ and rewriting the resultant quadratic function. You task is figuring out this symmetric matrix $X_2$.

(c) Now based on the inequalities in (a) and (b), you can simply choose $X = \rho^2 X_1 + (1-\rho^2)X_2$ for any $0 < \rho < 1$, and we have

$$\begin{bmatrix} x_k - x^* \\ x_{k-1} - x^* \\ \nabla f(v_k) \end{bmatrix}^\mathsf{T} X \begin{bmatrix} x_k - x^* \\ x_{k-1} - x^* \\ \nabla f(v_k) \end{bmatrix} \le -(f(x_{k+1}) - f(x^*)) + \rho^2(f(x_k) - f(x^*))$$

Based on the testing condition presented in the class, if there exists $P \ge 0$ such that

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

then the following inequality holds

$$\begin{aligned}(\xi_{k+1} - \xi^*)^\mathsf{T} P (\xi_{k+1} - \xi^*) - \rho^2 (\xi_k - \xi^*)^\mathsf{T} P (\xi_k - \xi^*) &\le \begin{bmatrix} x_k - x^* \\ x_{k-1} - x^* \\ \nabla f(v_k) \end{bmatrix}^\mathsf{T} X \begin{bmatrix} x_k - x^* \\ x_{k-1} - x^* \\ \nabla f(v_k) \end{bmatrix} \\ &\le -(f(x_{k+1}) - f(x^*)) + \rho^2(f(x_k) - f(x^*))\end{aligned}$$

which directly leads to the linear convergence rate for Nesterov's method:

$$(\xi_{k+1} - \xi^*)^\mathsf{T} P (\xi_{k+1} - \xi^*) + f(x_{k+1}) - f(x^*) \le \rho^2 \left( (\xi_k - \xi^*)^\mathsf{T} P (\xi_k - \xi^*) + f(x_k) - f(x^*) \right).$$

Finding $P$ to satisfy (1) is not trivial. Your task is to verify (1) holds with $P = \frac{1}{2} \begin{bmatrix} \sqrt{L} I \\ (\sqrt{m} - \sqrt{L}) I \end{bmatrix} \begin{bmatrix} \sqrt{L} I & (\sqrt{m} - \sqrt{L}) I \end{bmatrix} \ge 0$, $\rho^2 = 1 - \sqrt{\frac{m}{L}}$, $\alpha = \frac{1}{L}$, and $\beta = \frac{\sqrt{L} - \sqrt{m}}{\sqrt{L} + \sqrt{m}}$. This gives a complete proof for the accelerated rate of Nesterov's method.

(Hint: The calculation here can be lengthy. So you are allowed to use some symbolic toolbox to help as long as you turn in the code.)

**3**. (20 points) Consider a nonlinear system $x_{k+1} = f(x_k, u_k, w_k)$ where $x_k$ is the state, $u_k$ is the action, and $w_k$ is the process noised sampled from an IID Gaussian distribution. The objective is to choose $u_k$ to minimize the cost

$$\mathcal{C} = \mathbb{E} \sum_{k=0}^{\infty} \gamma^k c(x_k, u_k) \tag{2}$$

where $0 < \gamma < 1$ is the discount factor. Suppose $f$ is unknown, and we want to learn policy from sampled trajectories of $\{(x_k, u_k)\}$.

(a) What is the policy gradient theorem? Write out the statement.

(b) When using the policy gradient theorem, we inject noise into the control actions for exploration purposes. Suppose now we want to directly learn a deterministic policy using evolution strategies which are based on the following estimation of the policy gradient:

$$\nabla \mathcal{C}(K) \approx \frac{\mathbb{E}_{\varepsilon \sim \mathcal{N}(0,\sigma^2 I)} \mathcal{C}(K+\varepsilon)\varepsilon}{\sigma^2}$$

To understand this update rule, we analyze a shifted variant of the above update:

$$g = \frac{\mathbb{E}_{\varepsilon \sim \mathcal{N}(0,\sigma^2 I)} (\mathcal{C}(K+\varepsilon) - \mathcal{C}(K))\varepsilon}{\sigma^2} = \frac{\mathbb{E}_{\varepsilon \sim \mathcal{N}(0,I)} (\mathcal{C}(K+\sigma\varepsilon) - \mathcal{C}(K))\varepsilon}{\sigma}$$

Roughly speaking, the above estimation shifts the original zeroth-order gradient estimate with a zero mean vector and should not change the mean of the gradient estimator. Your task is to apply the fact $\lim_{\sigma \to 0} \frac{\mathcal{C}(K+\sigma\varepsilon) - \mathcal{C}(K)}{\sigma} = (\nabla \mathcal{C}(K))^\mathsf{T} \varepsilon$ to show the following equation:

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

(Remark: Now you can see $\sigma$ serves as a stepsize for the stochastic finite difference estimation. In the setting of data-driven control, typically we only have samples of $\mathcal{C}$. Choosing $\sigma$ to be too small can amplify this error, and hence one has to tune $\sigma$ carefully.)

(Hint: Relevant derivations can be found in the lecture note.)

**4**. (30 points) Consider a continuous-time LQR problem. We have $\dot{x}(t) = Ax(t) + Bu(t)$ with $x(0) \sim \mathcal{D}$. The quadratic cost is defined as

$$\mathcal{C} = \mathbb{E}_{x(0)\sim\mathcal{D}} \int_0^\infty [x(t)^\mathsf{T} Q x(t) + u(t)^\mathsf{T} R u(t)] dt$$

(a) Suppose we are using the state-feedback policy $u(t) = Kx(t)$. Suppose $K$ is stabilizing, i.e. $(A+BK)$ is Hurwitz. What is the cost $\mathcal{C}(K)$? Derive a cost formula from Lyapunov equation.

(b) Given $(A, B, Q, R)$, what is the gradient formula $\nabla\mathcal{C}(K)$? (Hint: Some relevant derivations can be found in Section 3 of the paper "Computational design of optimal output feedback controllers.")

(c) Now apply the gradient descent method $K_{l+t1} = K_l - \alpha\nabla\mathcal{C}(K_l)$. Suppose we have already proved that $\{K_l\}$ will stay in a compact sublevel set where $\mathcal{C}(K)$ is $L$-smooth and also satisfies the following gradient dominance property:

$$\mathcal{C}(K) - \mathcal{C}(K^*) \leq \frac{1}{2\mu}\|\nabla\mathcal{C}(K)\|_F^2 \tag{3}$$

where $K^*$ is the optimal policy, and $\mu$ is some positive constant. Your task is to further prove $\mathcal{C}(K_l) - \mathcal{C}(K^*) \leq (1 - 2\mu\alpha + \mu L\alpha^2)^l(\mathcal{C}(K_0) - \mathcal{C}(K^*))$ for any $0 < \alpha < \frac{2}{L}$.