# Solutions for Mid-Term 1

ECE490 Introduction to Optimization    Fall 2018

## 1 Problem 1

- (25 points) Show that the function

$$f(x_1, x_2) = 8x_1 + 12x_2 + x_1^2 - 2x_2^2$$

has only one stationary point, and that it is neither a local min or a local max, but a saddle point.

(Hint: Stationary points are points whose gradients are 0.)

**Solution**: It is straightforward to obtain

$$\frac{\partial f}{\partial x_1} = 8 + 2x_1$$
$$\frac{\partial f}{\partial x_2} = 12 - 4x_2$$

Setting $8 + 2x_1 = 0$ and $12 - 4x_2 = 0$, we get $x_1 = -4$, and $x_2 = 3$. And this is the only stationary point.

Now we calculate the Hessian

$$\nabla^2 f = \begin{bmatrix} 2 & 0 \\ 0 & -4 \end{bmatrix}$$

This Hessian has two eigenvalues: 2 and $-4$. Since the Hessian has one negative eigenvalue and one positive eigenvalue, it is a (strict) saddle point.

## 2 Problem 2

Consider the unconstrained minimization problem $\min_{x \in \mathbb{R}^p} f(x)$ where $f : \mathbb{R}^p \to \mathbb{R}$ is the objective function. Suppose $x^*$ is a point satisfying $\nabla f(x^*) = 0$. Apply the gradient method with a constant stepsize $\alpha$.

1. (5 points) Please write down the gradient descent method;

2. (10 points) When $f$ is convex and $L$-smooth, the following inequality holds for all $x \in \mathbb{R}^p$

$$f(x^*) \geq f(x_k) + \nabla f(x_k)^\mathsf{T}(x^* - x_k) + \frac{1}{2L}\|\nabla f(x_k)\|^2$$

   Use the above inequality to show the gradient method with a stepsize $\alpha = \frac{1}{L}$ satisfies

$$f(x_k) - f(x^*) \leq \frac{L\|x_0 - x^*\|^2}{2(k+1)}$$

3. (10 points) Suppose $f$ is $m$-strongly convex and $L$-smooth. Now the following inequality holds for all $x \in \mathbb{R}^p$

$$\begin{bmatrix} x - x^* \\ \nabla f(x) \end{bmatrix}^\mathsf{T} \begin{bmatrix} 2mLI & -(m+L)I \\ -(m+L)I & 2I \end{bmatrix} \begin{bmatrix} x - x^* \\ \nabla f(x) \end{bmatrix} \leq 0$$

   Use the above inequality to show the gradient method with stepsize $\alpha = \frac{1}{L}$ satisfies

$$\|x_k - x^*\| \leq \left(1 - \frac{m}{L}\right)^k \|x_0 - x^*\|$$

**Solution**:

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

2. Based on the inequality in the problem statement, we can choose $X = \begin{bmatrix} 0 & -\frac{1}{2}I \\ -\frac{1}{2}I & \frac{1}{2L}I \end{bmatrix}$ such that the following inequality holds

$$\begin{bmatrix} x_k - x^* \\ \nabla f(x_k) \end{bmatrix}^\mathsf{T} X \begin{bmatrix} x_k - x^* \\ \nabla f(x_k) \end{bmatrix} \leq f(x^*) - f(x_k)$$

   Therefore, we will have $p\|x_{k+1} - x^*\|^2 - p\|x_k - x^*\|^2 \leq f(x^*) - f(x_k)$ if we can find $p > 0$ such that

$$\begin{bmatrix} 0 & -\alpha p \\ -\alpha p & \alpha^2 p \end{bmatrix} - \begin{bmatrix} 0 & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2L} \end{bmatrix} \leq 0$$

When $\alpha = \frac{1}{L}$, we can just choose $p = \frac{L}{2}$ and the above condition will be satisfied. Now we have proved $\frac{L}{2}\|x_{k+1} - x^*\|^2 - \frac{L}{2}\|x_k - x^*\|^2 \leq f(x^*) - f(x_k)$. Summing this inequality leads to

$$\sum_{t=0}^{k}(f(x_t) - f(x^*)) \leq \frac{L}{2}\|x_0 - x^*\|^2 - \frac{L}{2}\|x_{k+1} - x^*\|^2 \leq \frac{L}{2}\|x_0 - x^*\|^2$$

Finally, due to the smoothness and the gradient iteration, we know $f(x_{k+1}) \leq f(x_k)$ for all $k$. Hence we have $(k+1)(f(x_k) - f(x^*)) \leq \sum_{t=0}^{k}(f(x_t) - f(x^*)) \leq \frac{L}{2}\|x_0 - x^*\|^2$. We have reached the desired conclusion.

3. If we can find $p > 0$ such that

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

then we have $p\|x_{k+1} - x^*\|^2 \leq \rho^2 p\|x_k - x^*\|^2$. When $\alpha = \frac{1}{L}$, we can just choose $\rho^2 = 1 - \frac{2m}{L} + \frac{m^2}{L^2}$ and $p = L^2$ such that the left matrix in the above inequality becomes $\begin{bmatrix} -m^2 & m \\ m & -1 \end{bmatrix}$ which is clearly negative semidefinite. This proves $\|x_{k+1} - x^*\| \leq \left(1 - \frac{m}{L}\right)\|x_k - x^*\|$. Iterating this inequality leads to the desired conclusion.

## 3 Problems 3

1. (10 points) Consider the positive definite quadratic minimization problem $\min_{x\in\mathbb{R}^p} \frac{1}{2}x^\top Qx$ where $Q$ is a positive definite matrix. Apply the gradient method with the stepsize $\alpha_k$ chosen by the direct line search. What is $\alpha_k$? Write out $\alpha_k$ as a function of $Q$ and $x_k$.

2. (15 points) Consider the problem of minimizing the function of two variables $f(x, y) = 3x^2+y^4$. Calculate one iteration of the gradient method with $(1, 2)$ as the starting point and with the stepsize chosen by the Armijo rule with $\alpha_0 = 1$, $\sigma = 0.1$, and $\beta = 0.25$.

**Solution**:

1. Notice $\nabla f(x_k) = Qx_k$. Hence we have

$$\alpha_k = \arg\min_{\alpha} f(x_k - \alpha\nabla f(x_k)) = \arg\min_{\alpha} (x_k - \alpha Qx_k)^\top Q (x_k - \alpha Qx_k)$$

Then we get

$$f(x_k - \alpha\nabla f(x_k)) = \frac{1}{2}(x_k - \alpha Qx_k)^\top Q (x_k - \alpha Qx_k) = \frac{1}{2}\left(x_k^\top Qx_k - 2(x_k^\top Q^2x_k)\alpha + (x_k^\top Q^3x_k)\alpha^2\right)$$

In the above expression, $x_k^\top Q^2x_k$ and $x_k^\top Q^3x_k$ are both just scalars. Due to the positive definiteness of $Q$, we know the above problem is a one-dimensional positive definite quadratic minimization problem and we can just set its derivative with respect to $\alpha$ to be 0. We get

$$\alpha_k = \frac{x_k^\top Q^2x_k}{x_k^\top Q^3x_k}$$

2. The function value at$(x_0, y_0)$ is $3 + 16 = 19$. The gradient at $(x, y)$ is $\begin{bmatrix}6x\\4y^3\end{bmatrix}$. The gradient evaluated at $(x_0, y_0)$ is $\begin{bmatrix}6\\32\end{bmatrix}$.

We have

$$\begin{bmatrix}x_0\\y_0\end{bmatrix} - \alpha_0\beta^m\begin{bmatrix}6x_0\\4y_0^3\end{bmatrix} = \begin{bmatrix}(1-6\beta^m)x_0\\y_0 - 4\beta^m y_0^3\end{bmatrix}$$

and

$$f((1-6\beta^m)x_0, y_0 - 4\beta^m y_0^3) = 3(1-6\beta^m)^2x_0^2 + (y_0 - 4\beta^m y_0^3)^4$$

For $m = 0$, the above term is $810075 > 19 - 106 = -87$. So the Arimjo condition is not met. When $m = 1$, the above term is $1296.75 > 19 - 26.5 = -7.5$. The Armijo condition is not met. When $m = 2$, the above term is $1.17 < 19 - 6.625 = 12.375$. The Armijo condition is met. So we should choose $m = 2$ and we have

$$\begin{bmatrix}x_1\\y_1\end{bmatrix} - \alpha_0\beta^2\begin{bmatrix}6x_0\\4y_0^3\end{bmatrix} = \begin{bmatrix}(1-6\beta^2)x_0\\y_0 - 4\beta^2 y_0^3\end{bmatrix} = \begin{bmatrix}0.625\\0\end{bmatrix}$$

# 4 Problems 4

1. (25 points) Suppose we apply several first-order optimization methods to solve an unconstrained minimization problem where the objective function $f$ is $L$-smooth and $m$-strongly convex. Suppose we have tried i) the gradient method with stepsize $\alpha = \frac{1}{L}$; ii) Nesterov's method with $\alpha = \frac{1}{L}$ and $\beta = \frac{\sqrt{L}-\sqrt{m}}{\sqrt{L}+\sqrt{m}}$; iii) Heavy-ball method with $\alpha = \frac{4}{(\sqrt{L}+\sqrt{m})^2}$ and $\beta = \left(\frac{\sqrt{L}-\sqrt{m}}{\sqrt{L}+\sqrt{m}}\right)^2$; iv) a new method called triple momentum method which achieves a convergence rate $f(x_k) - f(x^*) \leq \left(1 - 2\sqrt{\frac{m}{L}}\right)^k C$ (where $C$ is some constant). Figure 1 plots the iteration trajectories of $\|x_k - x^*\|$ for all these four methods. Suppose the trajectories in Figure 1 are consistent with the convergence rate theory. There are four trajectories in Figure 1, and each trajectory corresponds to one method. Your task is to label the trajectory for each method. And briefly explain how you pair the trajectory with the optimization method based on the convergence rate theory.

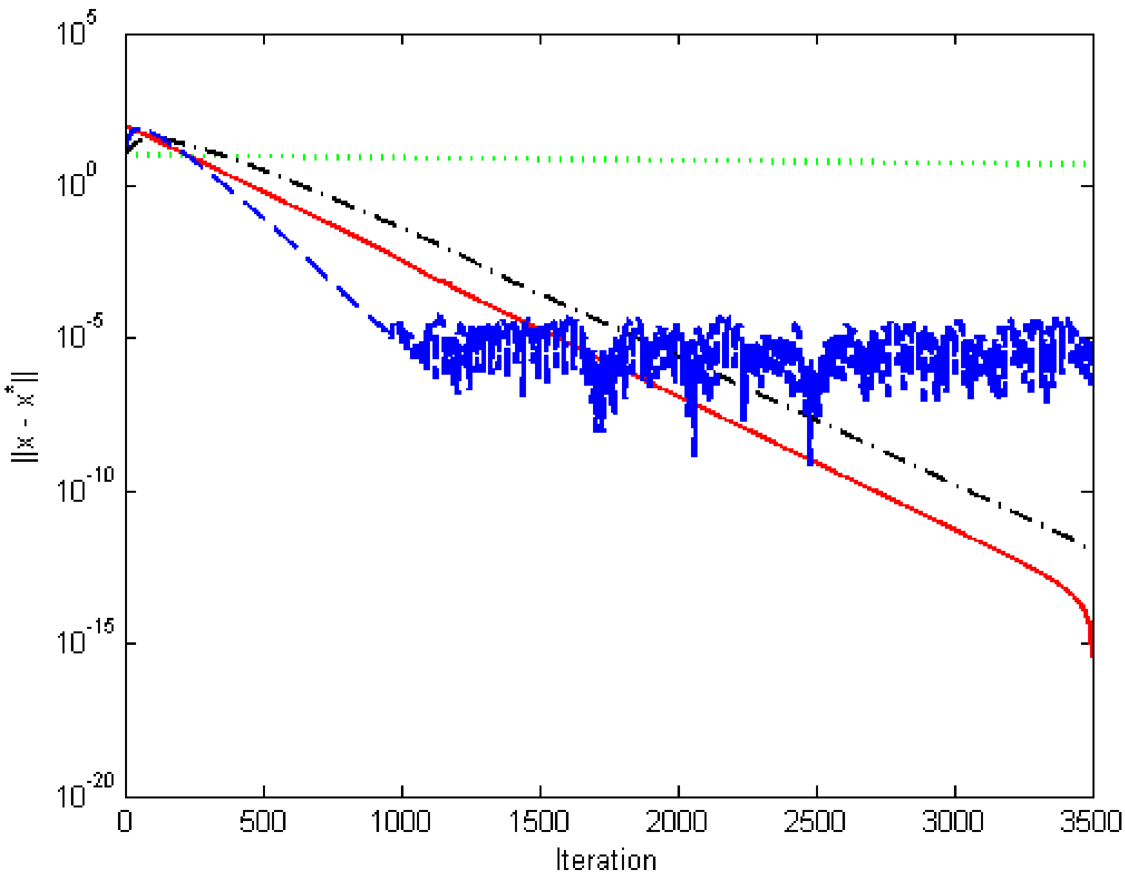


Figure 1:

**Solution**: The green curve is the gradient method since its guaranteed convergence rate is much slower than the rates of Nesterov's method and the triple momentum method. The convergence rate of the triple momentum method is slightly faster than Nesterov's method (by a constant factor). Heavy-ball method is not guaranteed to converge for general smooth strongly-convex functions and may oscillate. Therefore, the black curve is Nesterov's method, and the red curve is the triple momentum method. Then the blue curve is the Heavy-ball method.