SUN Yilin
520030910361

Homework 6

1

(a)

It is $max\{\gamma, 1\}$.
We know $\nabla f(\boldsymbol{x}) = \boldsymbol{Q}\boldsymbol{x}$, where $\boldsymbol{Q}$ is positive definite. So for any L such that f is L-smooth, L must satisfy that $||\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})|| = ||\boldsymbol{Q}(\boldsymbol{x} - \boldsymbol{y})|| \leq L(\boldsymbol{x} - \boldsymbol{y})$. Let $(\boldsymbol{x} - \boldsymbol{y}) = (d_1, d_2)$, the inequality is equivalent to $\gamma^2 d_1^2 + d_2^2 \leq L^2(d_1^2 + d_2^2)$, or $(\gamma^2 - L^2)d_1^2 + (1 - L^2)d_2^2 \leq 0$. Since this must hold for any (d_1, d_2), $\gamma^2 - L^2 \leq 0, 1 - L^2 \leq 0$. So the smallest L is $max\{\gamma, 1\}$.

(b)

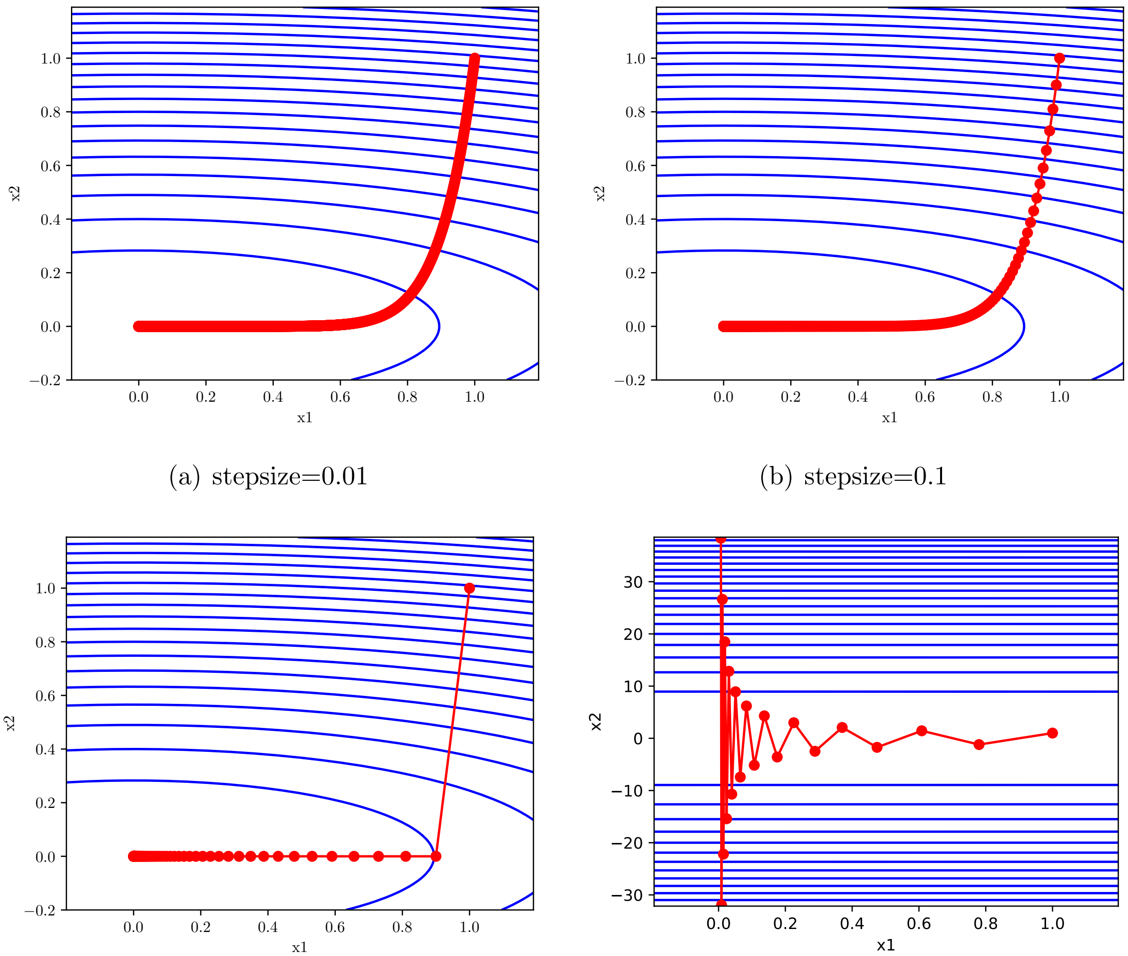
It is $min\{\gamma, 1\}$.
For any m such that f is m-strongly convex, m must satisfy that $f(\boldsymbol{x}) - \frac{m}{2}||\boldsymbol{x}||^2$ is convex, that is, $\frac{1}{2}\boldsymbol{x}^T(\boldsymbol{Q} - m\boldsymbol{I})\boldsymbol{x}$ is convex. So $\boldsymbol{Q} - m\boldsymbol{I}$ must be positive semidefinite. So its eigenvalues $\gamma - m$ and $1 - m$ must be nonnegative. So the largest m is $min\{\gamma, 1\}$.

(c)

Below are the figures of 2D trajectory of $\boldsymbol{x}_k$ and function values $f(\boldsymbol{x}_k)$. Note that the suboptimality gap $f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*)$ is simply $f(\boldsymbol{x}_k)$ since $f(\boldsymbol{x}^*) = 0$.
As for convergence, I notice that when stepsize=2.2, it does not converge (I run 20 iterations), while the other three stepsizes all converge. When it does converge, the iterations it takes are as follows (results are from running q1.py):

Stepsize	Iterations
1	88
0.1	917
0.01	9206

Table 1: Number of Iterations for each stepsize

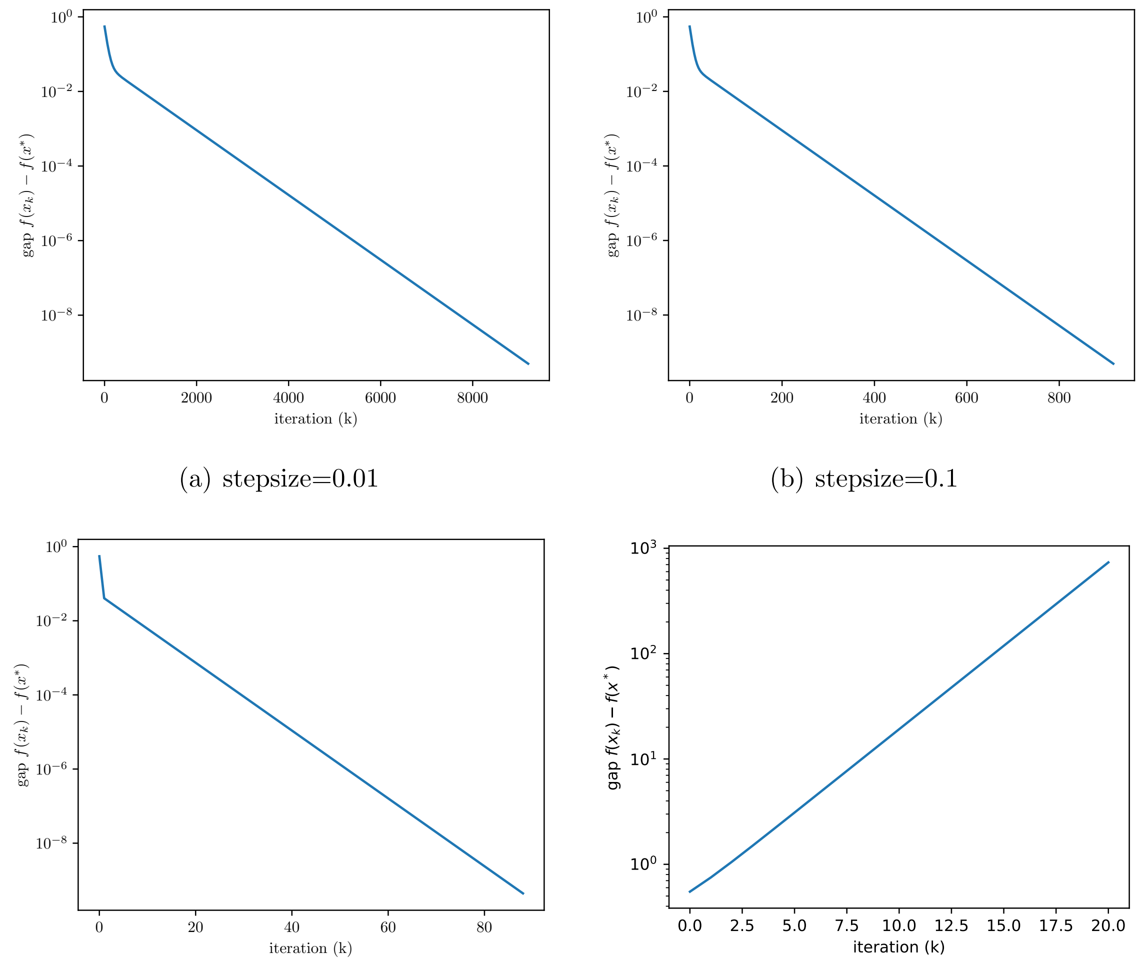


(a) stepsize=0.01 (b) stepsize=0.1

(c) stepsize=1 (d) stepsize=2.2

Figure 1: 2D trajectory of $\boldsymbol{x}_k$ ($\gamma = 0.1$)

From the figure of trajectory(above) we can actually see when stepsize=2.2, x_1 goes to 0 as we descend according to the direction of negative gradient, but x_2 oscillates as x_1 goes to 0, it does not converge.
From the figure of function value(below) we can see that when stepsize=1, the gradient method actually works well, we do not need a smaller stepsize. Because stepsize=1 only takes less than 100 iterations to get a same outcome as stepsize=0.1 or 0.01, which requires much more iterations.



(a) stepsize=0.01

(b) stepsize=0.1

(c) stepsize=1

(d) stepsize=2.2

Figure 2: Function Values ($\gamma = 0.1$)

(d)

Gamma	Iterations
1	1
0.1	88
0.01	688
0.001	4603

Table 2: Number of Iterations for different gamma

From the table we can see that as γ decreases, the number of iterations increases. And the relation is approximately inversely proportional. The intuition is that the condition number of $\boldsymbol{Q}$ is $\frac{1}{\gamma}$ since $\gamma \leq 1$. So with a larger γ we need more iterations to get a same optimum if we are applying the same stepsize.

2

Codes for this problem are in file p2.py.
I notice that for a large stepsize it will not converge, for example 3. I choose stepsize=0.01 with a starting point $(1, 1)$and the number of iterations is 605.

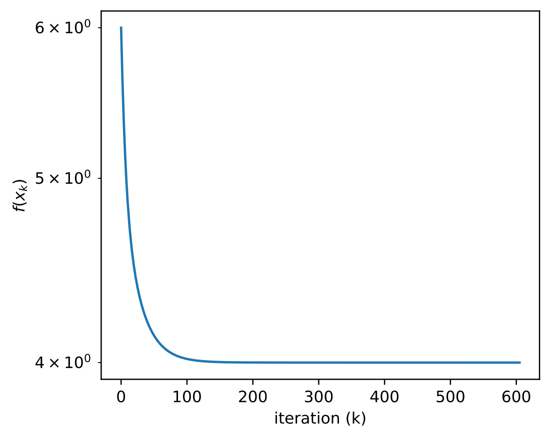


Figure 3: Least Squares gd with stepsize=0.01

The solution of closed form in HW5, running gradient descent and using np.linalg.solve are the same. The optimal variables are all $(w_1, w_2) = (1.5, 2)$ and the optimal value are all 4. Figure 3 shows the how the function value descends as we do iterations.

3

The optimal $\boldsymbol{w}^*$ given by running p3.py is (-1.47020052, 4.44377575, -4.37548225). The accuracy is 0.8667. And the classification result is given in the figure below.

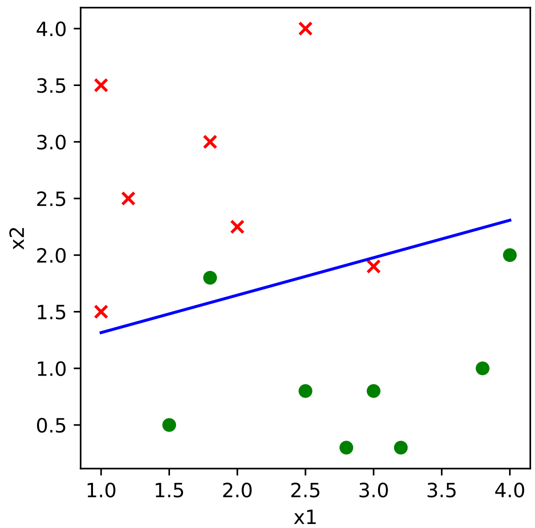


Figure 4: Classification Result

4

From f is differentiable and α-strongly convex we know $f(\boldsymbol{x})-\frac{\alpha}{2}||\boldsymbol{x}||^2$ is convex. By the first order condition for convexity we get

$$f(\boldsymbol{y})-\frac{\alpha}{2}||\boldsymbol{y}||^2-f(\boldsymbol{x})+\frac{\alpha}{2}||\boldsymbol{x}||^2 \geq (\nabla f(\boldsymbol{x})-\alpha\boldsymbol{x})^T(\boldsymbol{y}-\boldsymbol{x}), \forall \boldsymbol{x}, \boldsymbol{y}$$

From g is β-smooth we know

$$g(\boldsymbol{y})-g(\boldsymbol{x})-\nabla g(\boldsymbol{x})^T(\boldsymbol{y}-\boldsymbol{x}) \leq \frac{\beta}{2}||\boldsymbol{y}-\boldsymbol{x}||^2, \forall \boldsymbol{x}, \boldsymbol{y}$$

Adding these two inequalities together gives us

$$[f(\boldsymbol{y})-g(\boldsymbol{y})]-[f(\boldsymbol{x})-g(\boldsymbol{x})]-[\nabla f(\boldsymbol{x})-\nabla g(\boldsymbol{x})]^T(\boldsymbol{y}-\boldsymbol{x}) \geq$$
$$\frac{\alpha}{2}(||\boldsymbol{y}||^2-||\boldsymbol{x}||^2)-\alpha\boldsymbol{x}^T\boldsymbol{y}+\alpha||\boldsymbol{x}||^2-\frac{\beta}{2}(||\boldsymbol{y}||^2+||\boldsymbol{x}||^2)+\beta\boldsymbol{x}^T\boldsymbol{y}, \forall \boldsymbol{x}, \boldsymbol{y}$$

RHS is simply

$$\frac{\alpha-\beta}{2}||\boldsymbol{y}-\boldsymbol{x}||^2$$

Which is greater or equal to 0. So

$$[f(\boldsymbol{y})-g(\boldsymbol{y})] \geq [f(\boldsymbol{x})-g(\boldsymbol{x})]+[\nabla f(\boldsymbol{x})-\nabla g(\boldsymbol{x})]^T(\boldsymbol{y}-\boldsymbol{x}), \forall \boldsymbol{x}, \boldsymbol{y}$$

Or equivalently

$$h(\boldsymbol{y}) \geq h(\boldsymbol{x})+\nabla h(\boldsymbol{x})^T(\boldsymbol{y}-\boldsymbol{x}), \forall \boldsymbol{x}, \boldsymbol{y}$$

Then by first order condition for convexity we know h is convex.