# HOMEWORK 6

## QUESTION 1

### (a.)

We know that a function is $L$-smooth if it is differentiable and its gradient is $L$-Lipschitz, i.e.

$$\|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\| \leq L\|\boldsymbol{x} - \boldsymbol{y}\|, \quad \forall \boldsymbol{x}, \boldsymbol{y} \tag{1}$$

$$\nabla f(\boldsymbol{x}) = \boldsymbol{Q}\boldsymbol{x} = \begin{pmatrix} 1 & 0 \\ 0 & \gamma \end{pmatrix} \boldsymbol{x} \tag{2}$$

Let $\boldsymbol{d} = \boldsymbol{x} - \boldsymbol{y}$ ,

$$\|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\| = \|\boldsymbol{Q}\boldsymbol{d}\| = \sqrt{\boldsymbol{d}^T\boldsymbol{Q}^2\boldsymbol{d}} \leq \sqrt{\lambda_{\max}\left(\boldsymbol{Q}^2\right)\|\boldsymbol{d}\|^2} = \lambda_{\max}(\boldsymbol{Q})\|\boldsymbol{x} - \boldsymbol{y}\| \tag{3}$$

The last equality uses the fact $\lambda_{\max}\left(\boldsymbol{Q}^2\right) = \lambda_{\max}^2(\boldsymbol{Q})$.

$\therefore$ the smallest $L$ that $f$ is $L$-smooth is $max(1, \gamma)$.

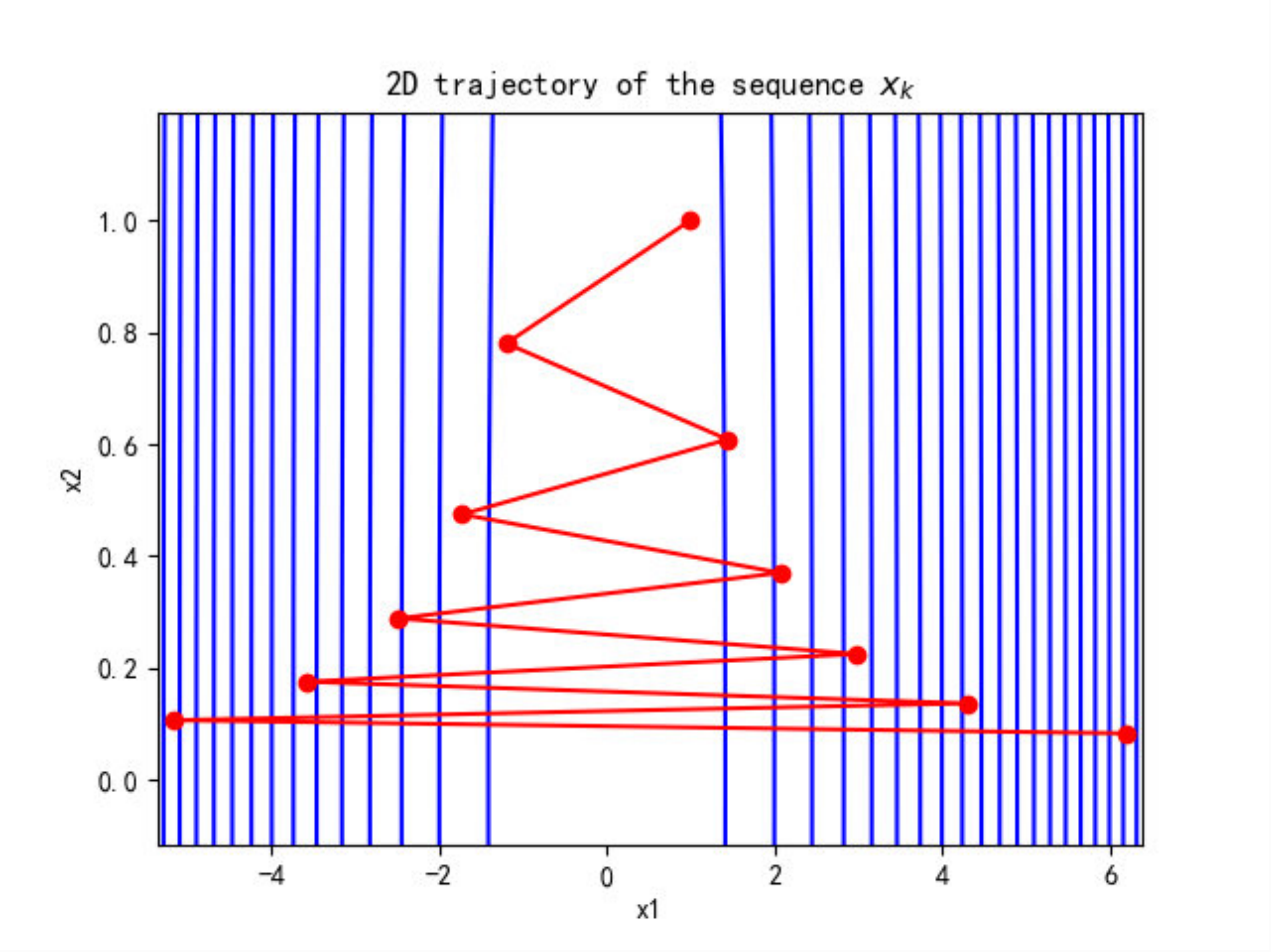
### (b.)

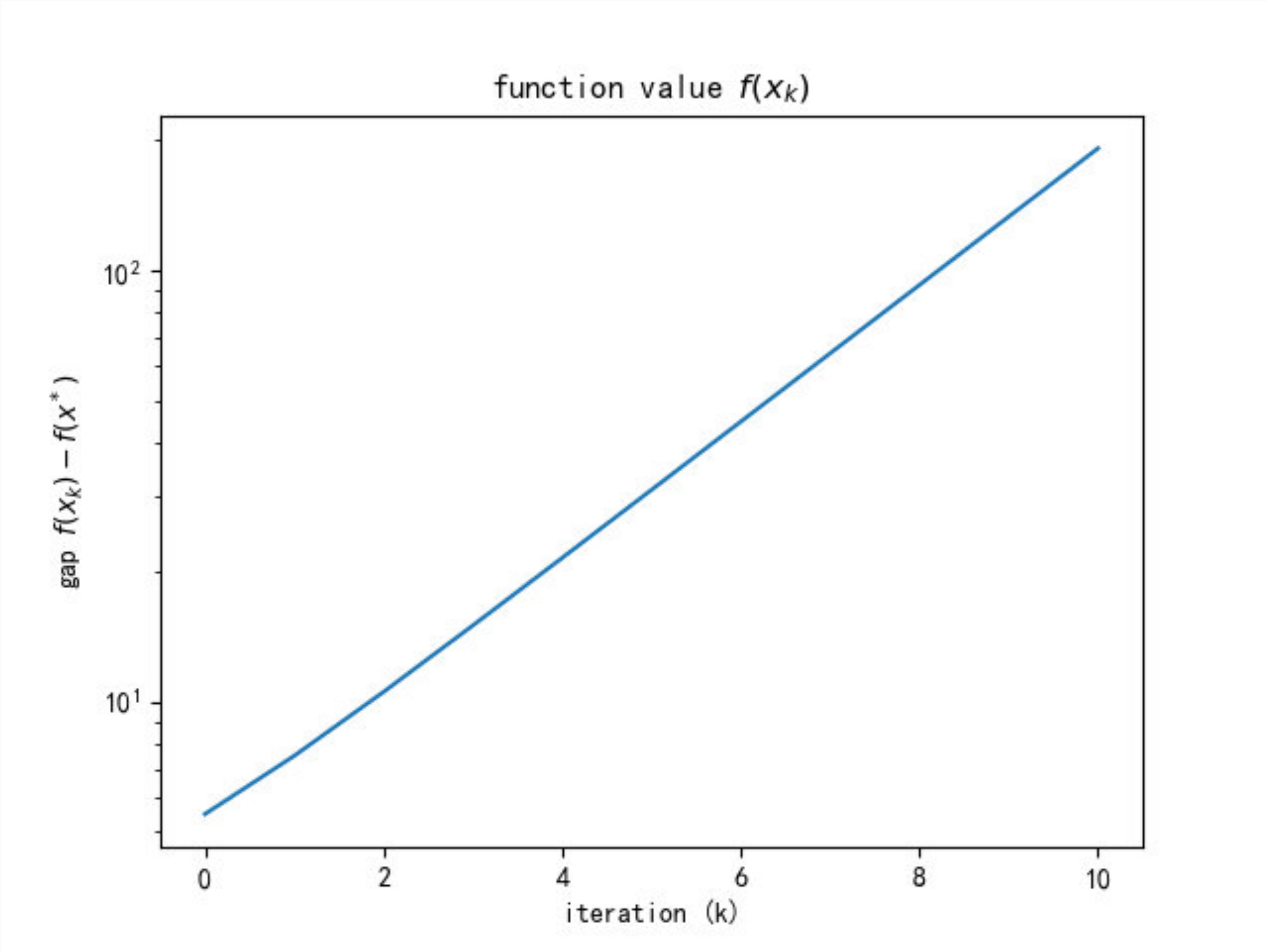
$$\boldsymbol{Q} = \begin{pmatrix} 1 & 0 \\ 0 & 10 \end{pmatrix} \tag{4}$$

```
stepsize = 0.22
```

| Convergence | Num of Iterations |
|---|---|
| *False* | *10(Max)* |

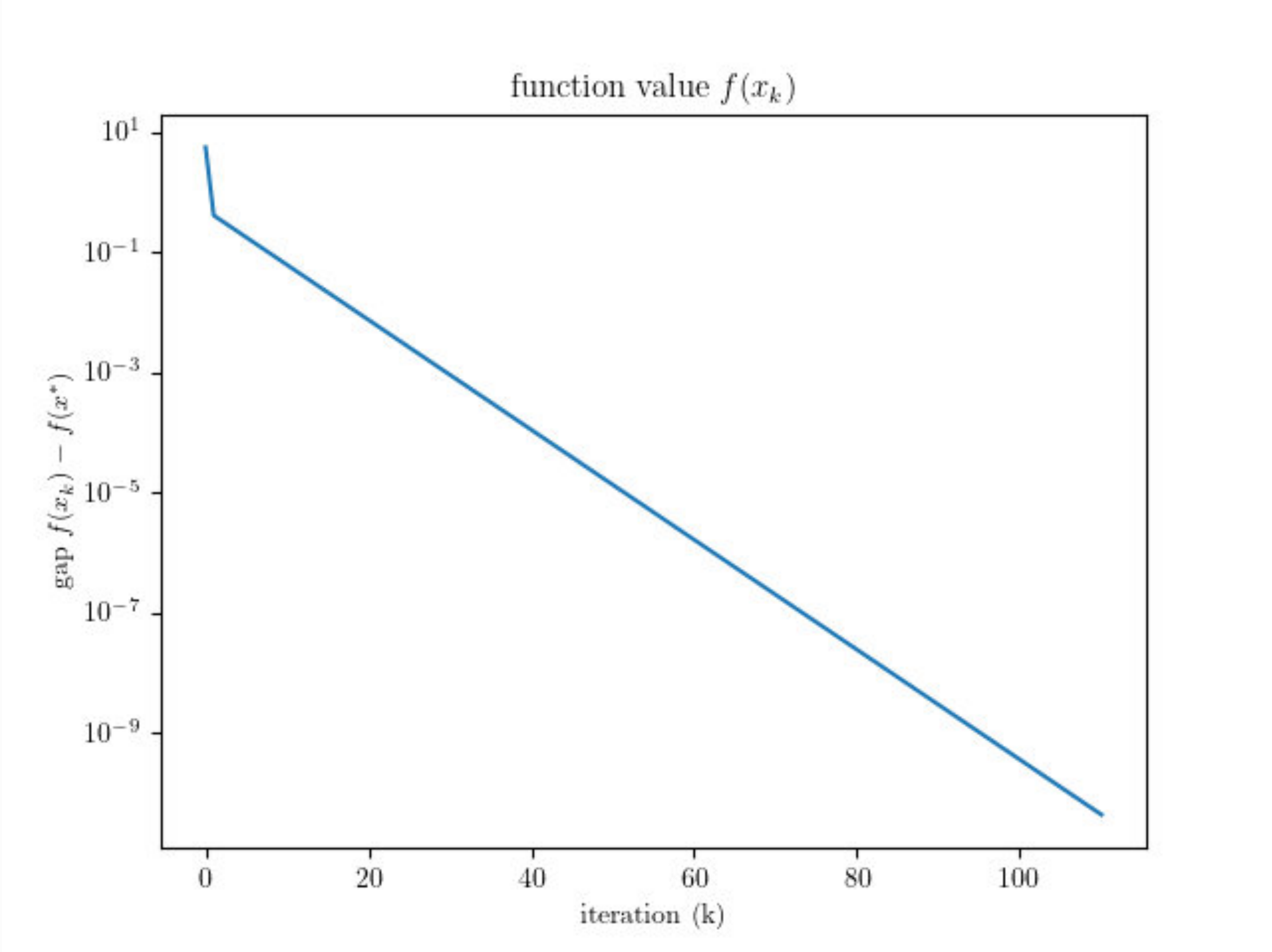
function value $f(\boldsymbol{x_k})$ :

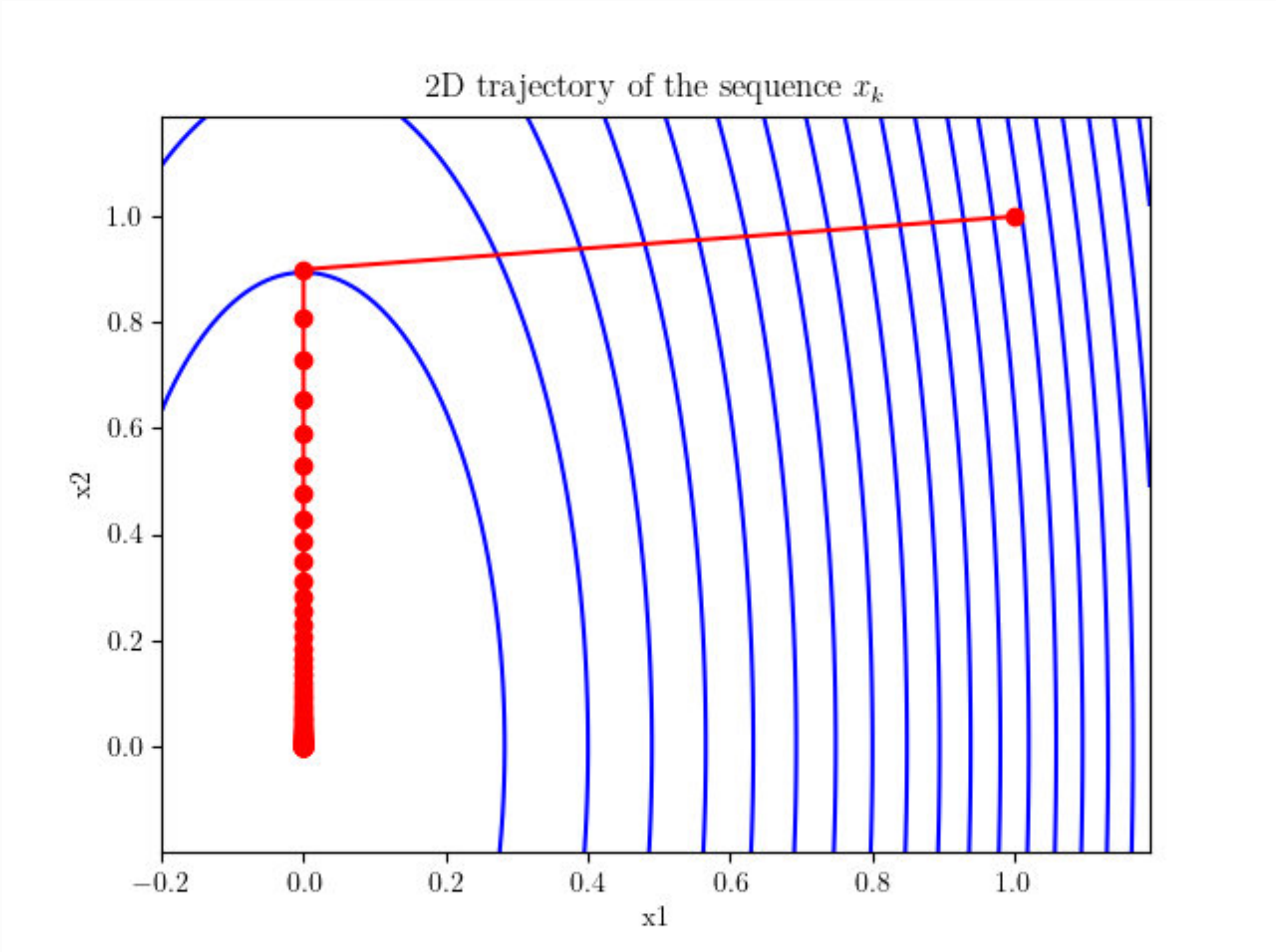





`stepsize = 0.1`

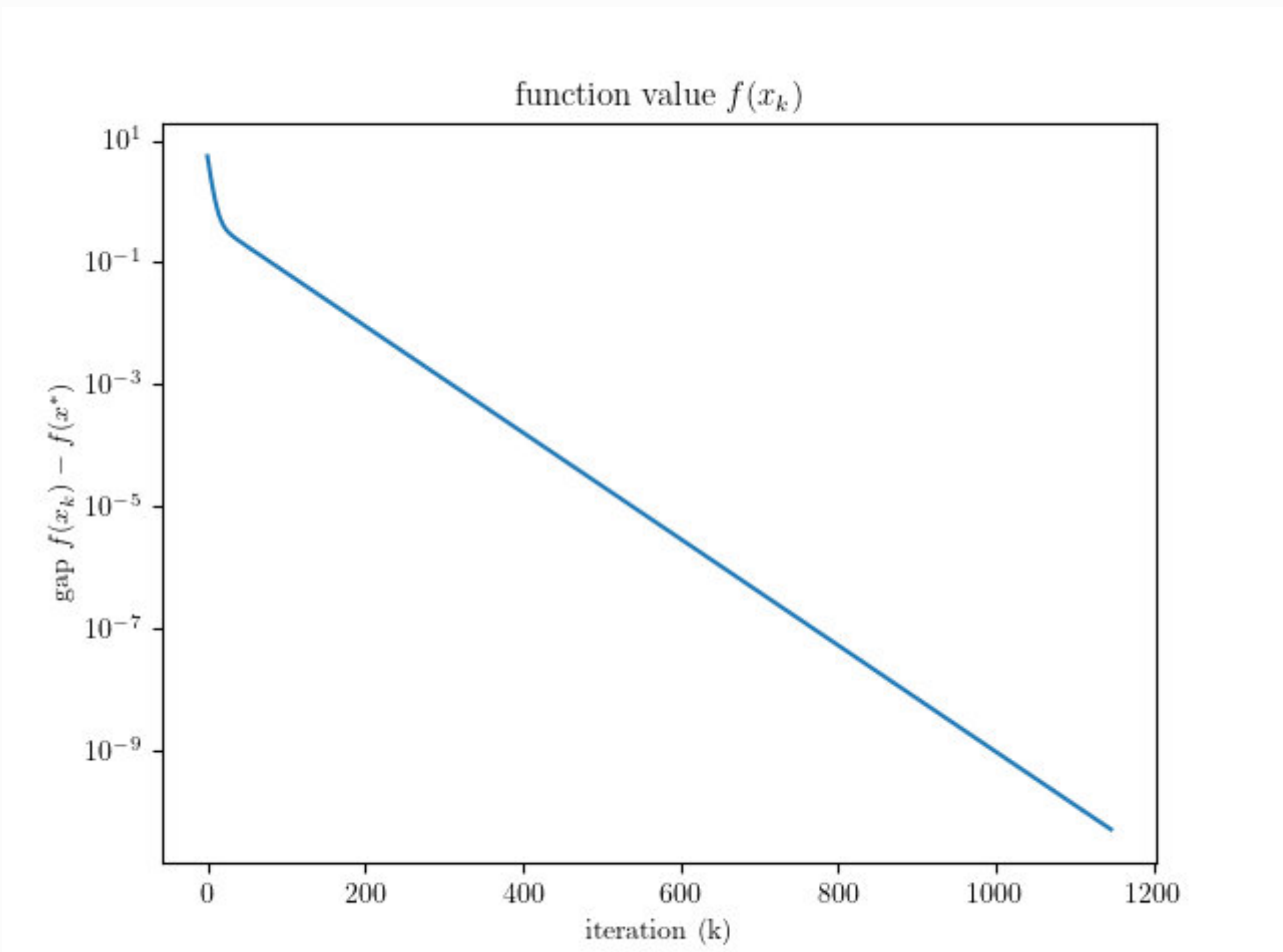
| Convergence | Num of Iterations |
| --- | --- |
| *True* | *110* |

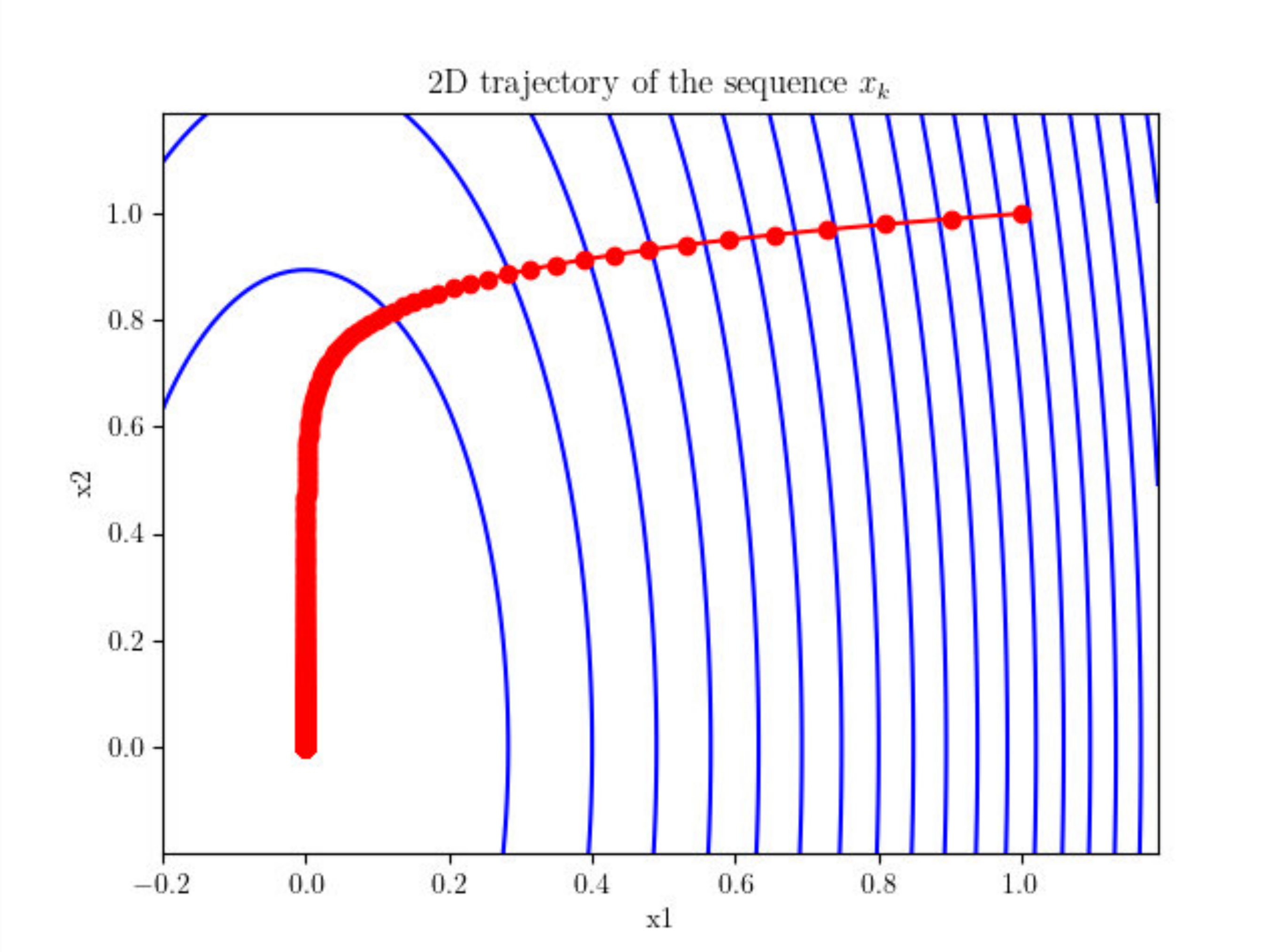





`stepsize = 0.01`

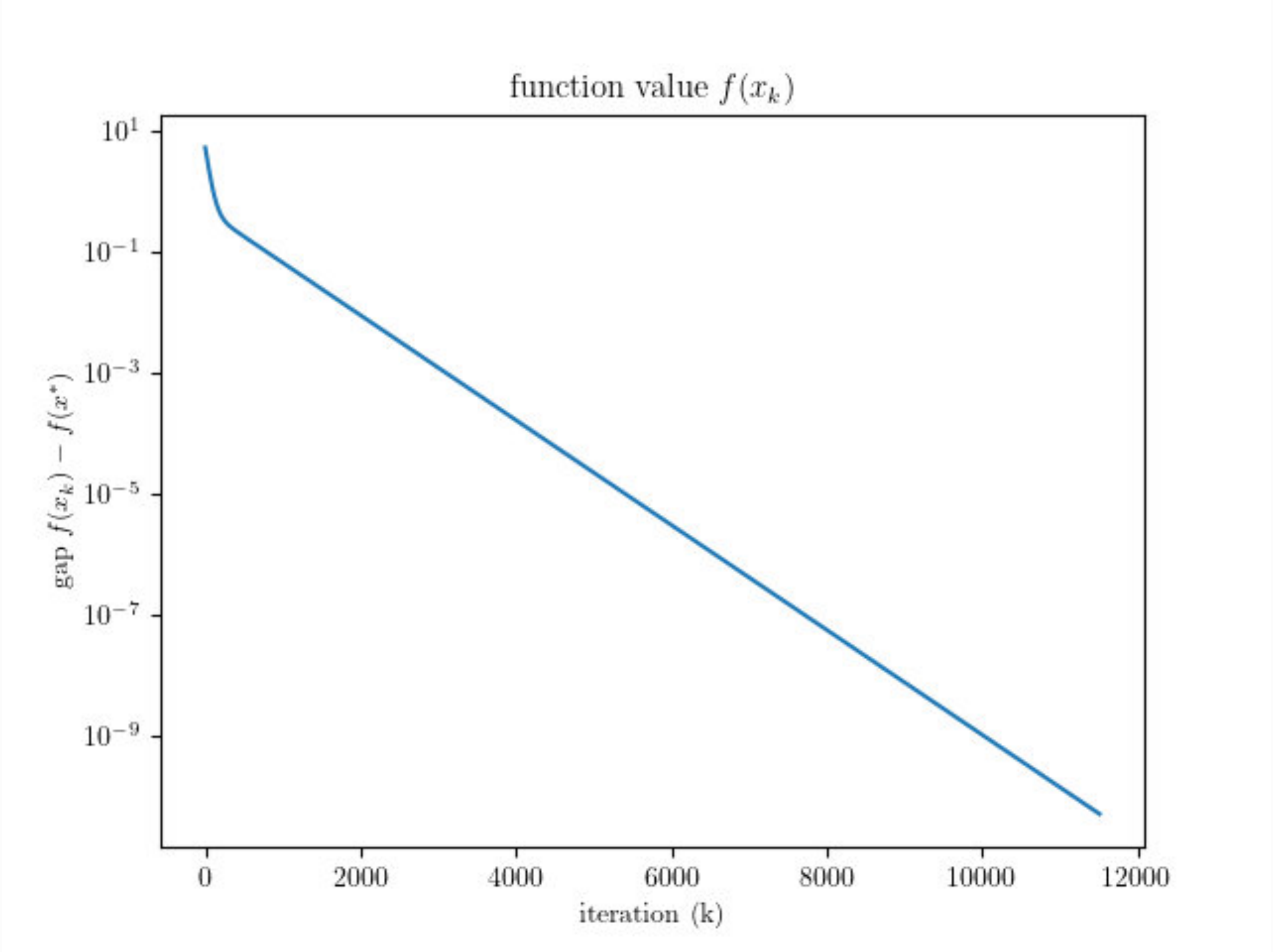
| Convergence | Num of Iterations |
| --- | --- |
| *True* | *1146* |

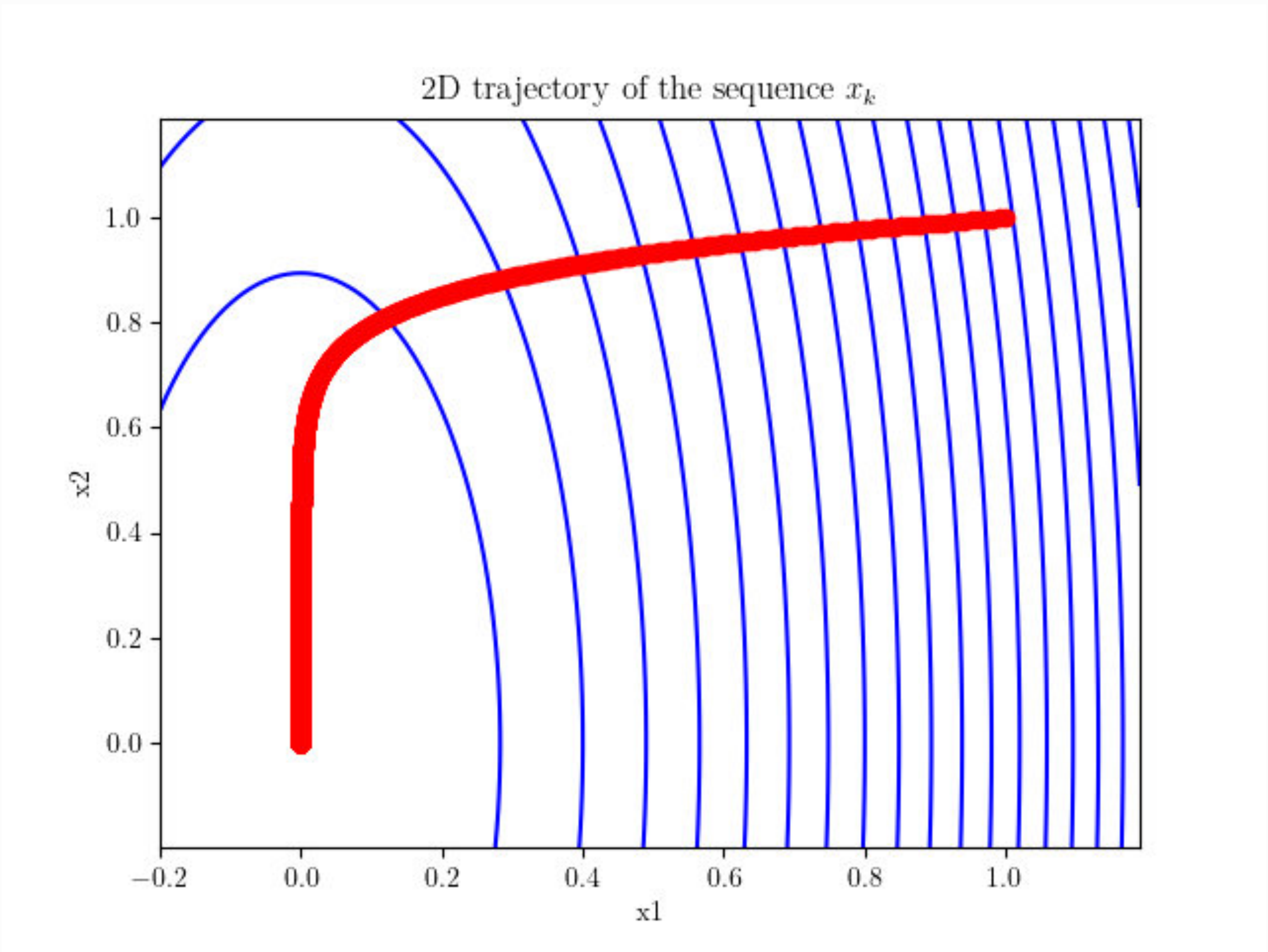


`stepsize = 0.001`

| Convergence | Num of Iterations |
| --- | --- |
| *True* | 11508 |




(c.)

```
gamma=1, stepsize=1, number of iterations=1
gamma=0.1, stepsize=1, number of iterations=88
gamma=0.01, stepsize=1, number of iterations=688
gamma=0.001, stepsize=1, number of iterations=4603
```

As $\gamma$ decreases, the number of iterations increases.

Since $f$ is 2D quadratic function, we notice that $\kappa(\boldsymbol{Q}) = \frac{\lambda_{\max}(\boldsymbol{Q})}{\lambda_{\min}(\boldsymbol{Q})}$ keeps increasing wot==ith the cases, thus the problem is turning from a well-conditioned problem to an ill-conditioned problem.

Generally, for the number of iteration $k$, if we want $|f(\boldsymbol{x}) - f(\boldsymbol{x}^*)| < \epsilon$ ,

$$k = O(log(\frac{1}{\epsilon})/log(\frac{\kappa(\boldsymbol{Q})+1}{\kappa(\boldsymbol{Q})-1})) \tag{5}$$

# QUESTION 2

```
stepsize = 0.001, w0 = [0, 0, 0, 0, 0, 0], maxiter = 100000, tolerance = 0.00001
```

```
(base) husky@Huskys-MacBook-Pro code % python -u "/Users/husky/大二上/线性优化与凸优化/Homeworks/hw6/code/p2.py"
stepsize=0.001, number of iterations=4179
[ 1.22170436 -0.21469164  0.1554913  -0.45867604  1.18537713  0.00613276]
```

**Solution found in HW5:**

```
status: optimal
optimal value: 13.295569218196668
optimal var:[ 1.22170662 -0.21469308  0.15549205 -0.4586777   1.18537705  0.00613318]
```

**Solution by solving the norm equation:**

The normal equation is:

$$\boldsymbol{X}^T \boldsymbol{X} \boldsymbol{w} = \boldsymbol{X}^T \boldsymbol{y} \tag{6}$$

Result:

```
[ 1.22170662 -0.21469307  0.15549204 -0.4586777   1.18537706  0.00613317]
```

Comparing the three solutions above, it seems that all three solutions can give us accurate enough answers to the least squares problem. However, gradient descent is slightly less accurate than the other two methods, as you can recognize the difference in the last digits of it from the other two results.

# QUESTION 3

*Result:*

```
stepsize=0.1, number of iterations=2133
the w found:[-1.73186234  5.05432758 -3.31093348]
accuracy = 0.9333333333333333
```

*The classification graph:*