1.

先证 $f(y) \le f(x) + \nabla f(x)^T(y-x) + \frac{L}{2}\|y-x\|_2^2$

~~$f(y) = f(x) + \nabla f(z)^T(y-x)$~~

~~$f(y) - f(x) - \nabla f(x)^T(y-x) \le \frac{L}{2}\|y-x\|_2^2$~~

~~$= (\nabla f(z) - \nabla f(x))^T(y-x)$~~

令 $g(t) = f(x+t(y-x))$,

$g'(t) = (\nabla f(x+t(y-x)))^T(y-x)$

$g'(0) = \nabla f(x)^T(y-x)$

$g'(t) - g'(0) = (\nabla f(x+t(y-x)) - \nabla f(x))^T(y-x)$

$\le \|\nabla f(x+t(y-x)) - \nabla f(x)\|_2 \|y-x\|_2$

$\le tL\|y-x\|_2^2$

$\int_0^1 (g'(t) - g'(0))dt$

$= g(1) - g(0) - g'(0)$

$= f(y) - f(x) - \nabla f(x)^T(y-x)$

$\le \int_0^1 t\,dt \cdot L\|y-x\|_2^2$

$= \frac{L}{2}\|y-x\|_2^2$

得证

Convex Optimization
Week 12

经博21
[illegible]
2021312[illegible]

清華大學

12/5/2021

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

$\cancel{= f(x_k) - t\nabla^T f(x_k)\nabla f(x_k) + \frac{t^2}{2}\nabla^2 f(x_k)\ \nabla^2 f(\xi)}$

$\frac{Lt^2}{2}\|\nabla f(x_k)\|_2^2$

$= f(x_k) - t\nabla^T f(x_k)\cdot\nabla f(x_k) + \cancel{\frac{t^2}{2}\nabla f(x_k)\nabla f(x_k)}\ \nabla^2 f(\xi)$

$\forall\xi$

$\because \nabla^2 f(\xi) \preceq LI$

且 $t^2\ \|\nabla^T f(x_k)\|_2^2 \geq 0$,

$\Rightarrow$ 有 $f(x_{k+1}) \leq f(x_k) - (1-\frac{Lt}{2})t\|\nabla f(x_k)\|_2^2$

(2) $\cancel{\|\nabla f(x_k)\|_2^2 \leq}$

$(1-\frac{Lt}{2})\cancel{t\|\nabla f(x_k)\|_2^2 \geq f(x_{k+1}) - f(x_k)}$

$(2-Lt)t$

$\cancel{0<t\leq\frac{1}{L}\ \therefore 1-\frac{Lt}{2}>0,\ t>0,}$

$\cancel{\|\nabla f(x_k)\|_2^2} \geq$

$(1-\frac{Lt}{2})t\|\nabla f(x_k)\|_2^2 \leq f(x_k) - f(x_{k+1})$

$\because 0<t\leq\frac{1}{L}\quad 1-\frac{Lt}{2}>0,\ t>0,$

有 $\|\nabla f(x_k)\|_2^2 \leq \frac{2}{-Lt^2+2t}(f(x_k)-f(x_{k+1}))$

$\because t\leq\frac{1}{L},\quad \therefore -Lt^2 \geq -t$

$2t - Lt^2 \geq t.$

$\frac{2}{-Lt^2+2t} \leq \frac{2}{t},$

$\|\nabla f(x_k)\|_2^2 \leq \frac{2}{t}(f(x_k)-f(x_{k+1}))$

(3) 将(2)所有(k)里求和,有

$\sum_{i=0}^{k}\|\nabla f(x_i)\|_2^2 \leq \frac{2}{t}(f(x_0) - f(x_1)$
$+ f(x_1)$
$+ f(x_{k-1}) - f(x_k)$
$+ f(x_k)\ f(x_{k+1}))$

$\sum_{i=0}^{k}\|\nabla f(x_i)\|_2^2 \leq \frac{2}{t}(f(x_0) - f(x_{k+1}))$

$\because f(x_{k+1}) \geq f(x^*)$,

$\therefore$ 有 $\sum_{i=0}^{k}\|\nabla f(x_i)\|_2^2 \leq \frac{2}{t}(f(x_0) - f(x^*))$

(4) 对(3)结果取平均,有

$\min_{i=0\sim k}\|\nabla f(x_i)\|_2^2 \leq \frac{2}{t(k+1)}(f(x_0)-f(x^*))$

$\therefore$ 有 $\min_{i=0\sim k}\|\nabla f(x_i)\| \leq \sqrt{\frac{2}{t(k+1)}(f(x_0)-f(x^*))}$

2.(1)

2.(1) $\mathrm{Prox}_f(v) = \arg\min_x \frac{\alpha}{2}(Ax-b)^T(Ax-b) + \frac{1}{2}(x-v)^T(x-v)$

$$= \arg\min_x \frac{\alpha}{2}x^TA^TAx - \alpha(A^Tb)^Tx + \frac{1}{2}x^Tx - v^Tx + \frac{1}{2}v^Tv = G(x)$$

$$\nabla G(x) = (\alpha A^TA + I)x - \alpha A^Tb - v$$

$$\nabla^2 G(x) = \alpha A^TA + I$$

$\because \alpha > 0,$

$\therefore \nabla^2 G(x) \succ 0$

$\therefore x = (\alpha A^TA + I)^{-1}(\alpha A^Tb + v)$

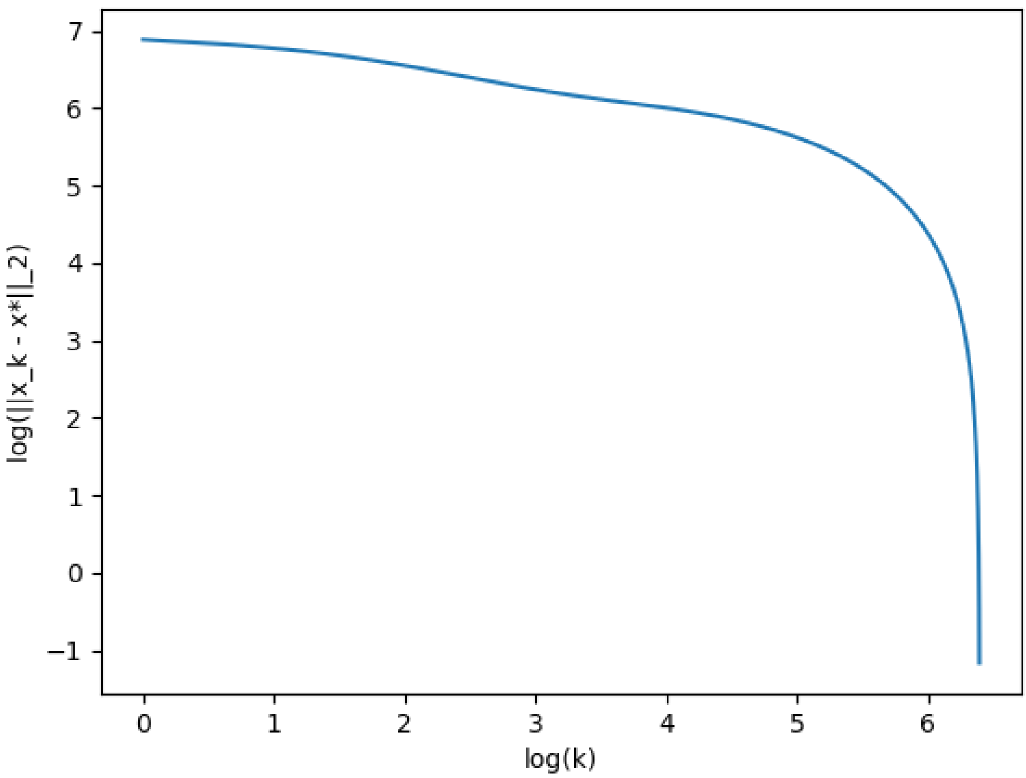
$$x_{k+1} = (\alpha A^TA + I)^{-1}(\alpha A^Tb + x_k)$$

0 1 2 3

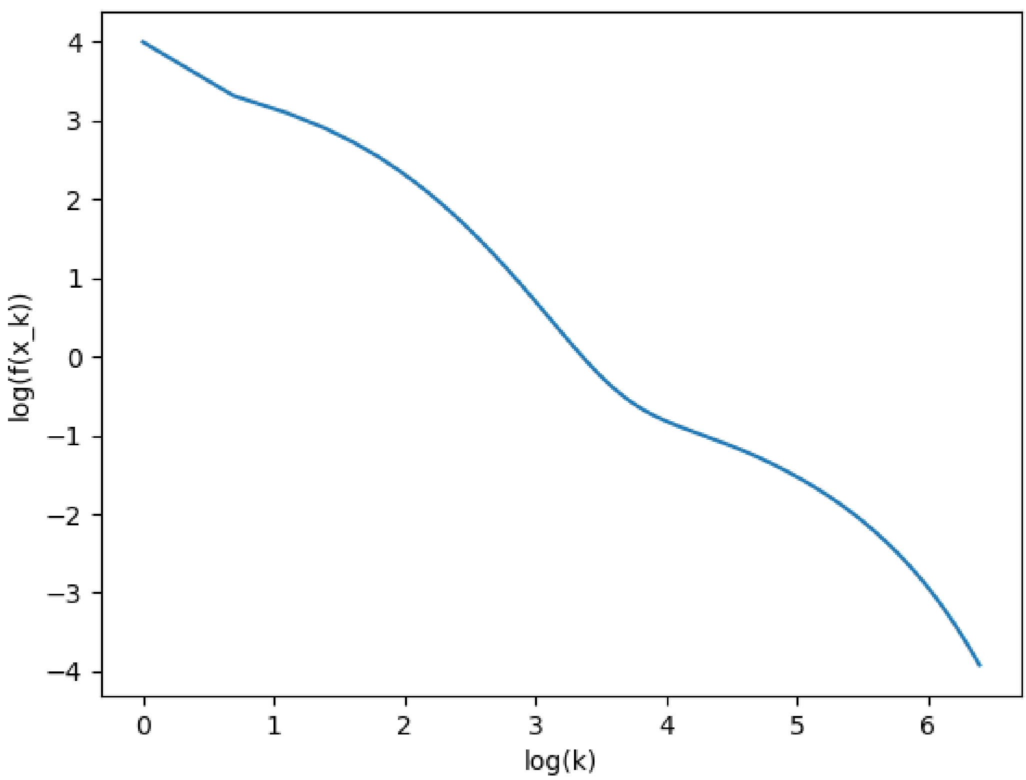
(2)

采用alpha=1000时，收敛到f(x)=0.01991554109994043，此时x结果见根目录下result/Q2/x.csv文件

log(k)和log(||x_k - x*||^2)的图像如下：



log(k)和log(f(x_k))的图像如下：



3.(1)

$$min_x \frac{1}{2}||Ax-b||_2^2 + ||x||_1 \text{的次梯度} g(x) \text{为}$$

$$A^TAx - A^Tb + sign(x), \text{其中} sign(x)_i = \begin{cases} 1 & x_i > 0 \\ 0 & x_i = 0 \\ -1 & x_i < 0 \end{cases}$$

具体推导如下：

3.1 $h(x) = ||Ax-b||_2^2 + ||x||_1$

$\because ||Ax-b||_2^2 = f(x)$ 在 R^n 上可微，

$\nabla f(x) = A^TAx - A^Tb$，

$\therefore$ 只需求 $||x||_1$ 在 R^n 上的次梯度 $\varphi(x)$，

$g(x) = \nabla f(x) + \varphi(x)$

满足 $\forall y, x$，$||y||_1 - ||x||_1 \geq (y-x)^T\varphi(x)$

$\because ||y||_1 = \sum_{i=1}^{n} |y_i|$，$||x||_1 = \sum_{i=1}^{n} |x_i|$，

$\therefore ||y||_1 - ||x||_1 = \sum_{i=1}^{n} |y_i| - |x_i|$

$(y-x)^T\varphi(x) = \sum_{i=1}^{n} \varphi(x_i)y_i - \varphi(x_i)x_i$

要对 $\forall y_i, x_i$，

有 $|y_i| - |x_i| \geq \varphi(x_i)y_i - \varphi(x_i)x_i$，

即要 $|y_i| \geq \varphi(x_i)y_i$

$\varphi(x_i)x_i \geq |x_i|$

取 $\varphi(x_i) = \begin{cases} 1 & x_i > 0 \\ 0 & x_i = 0 \\ -1 & x_i < 0 \end{cases}$ 即可

当 $x_i > 0$ 时，$\varphi(x_i) = 1$，有

$|y_i| \geq \varphi(x_i)y_i$，

$-|x_i| = -x_i$，

左 ≥ 右

当 $x_i < 0$ 时，$\varphi(x_i) = -1$，

有 $|y_i| \geq \varphi(x_i)y_i$，

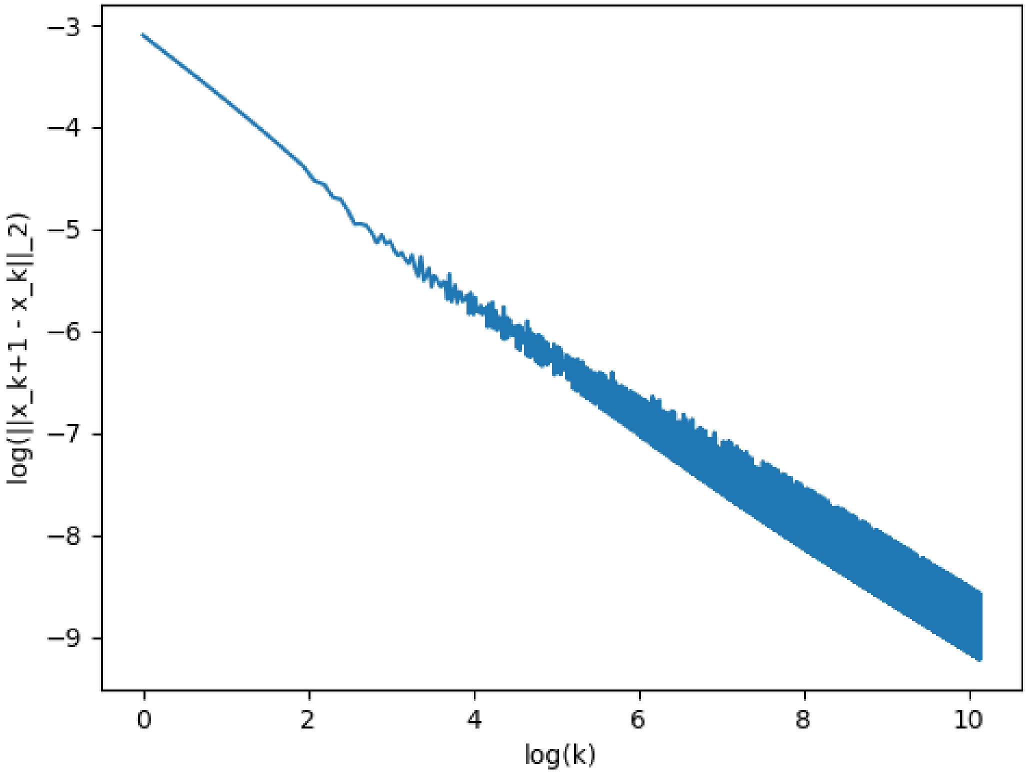
$-|x_i| = x_i$，左 ≥ 右

当 $x_i = 0$ 时，$\varphi(x_i) = 0$

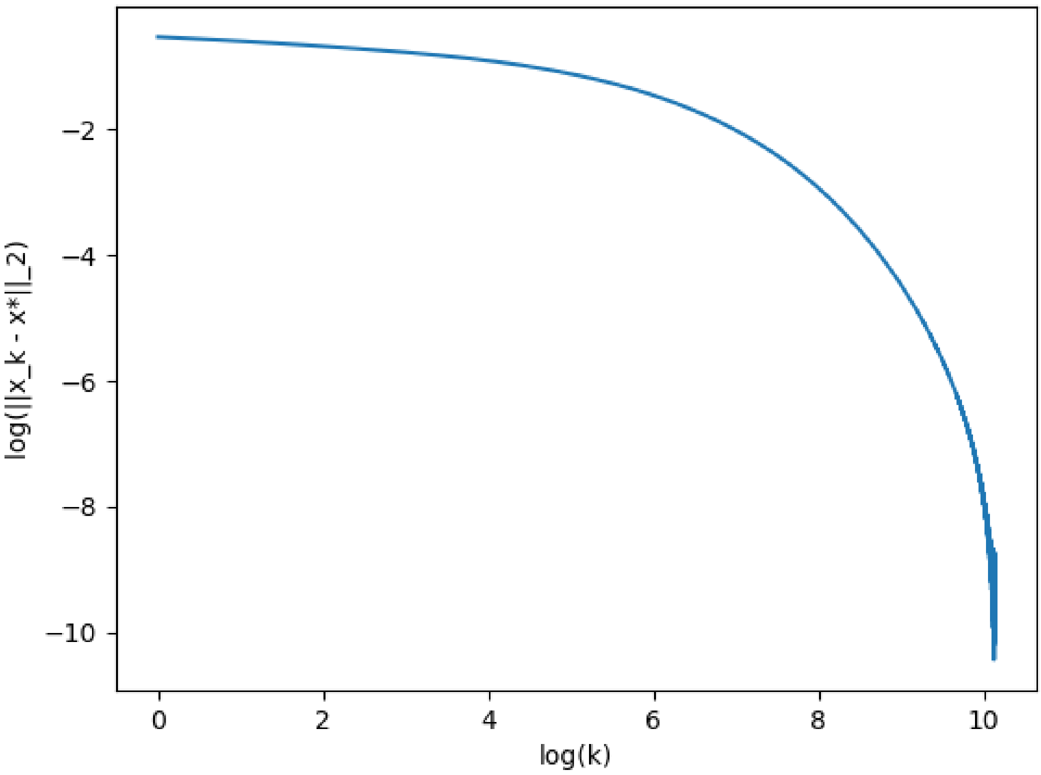
有 左 $= |y_i| \geq 0$

解得最优的h(x)=12.522447423125461,x存储在根目录下的result/Q3/x1.csv中

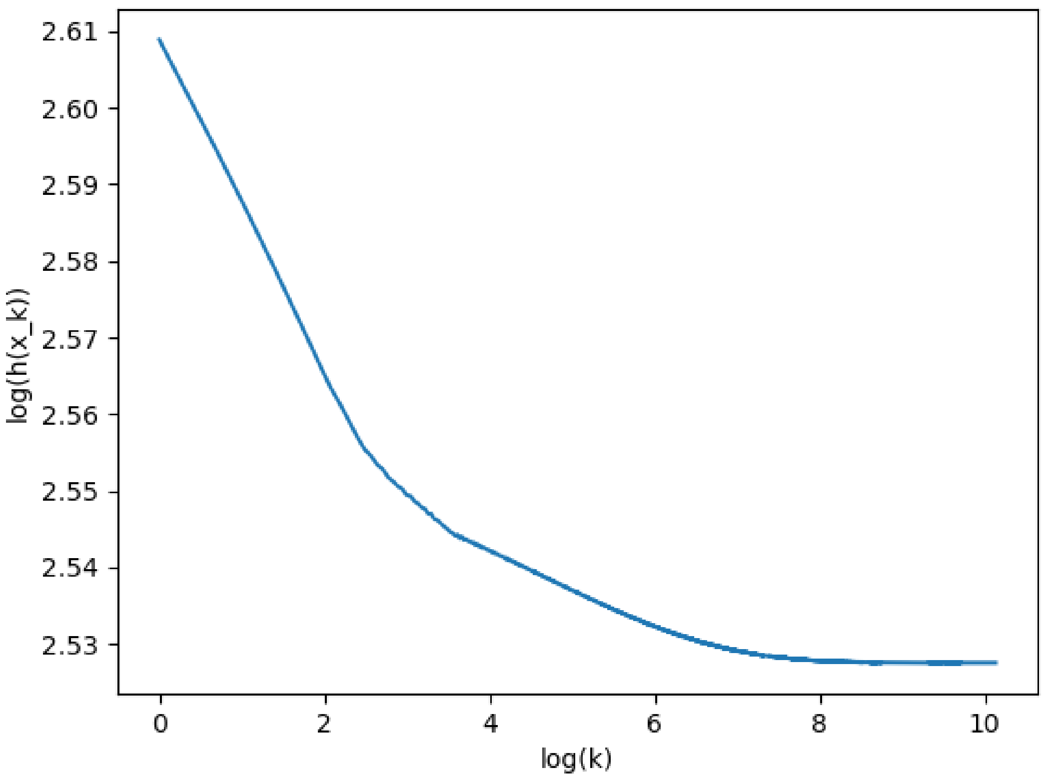
log(k)和log(||x_k+1 - x_k||_2)的图像如下：



log(k)和log(||x_k-x*||_2)的图像如下:



log(x)和log(h(x))的图像如下:

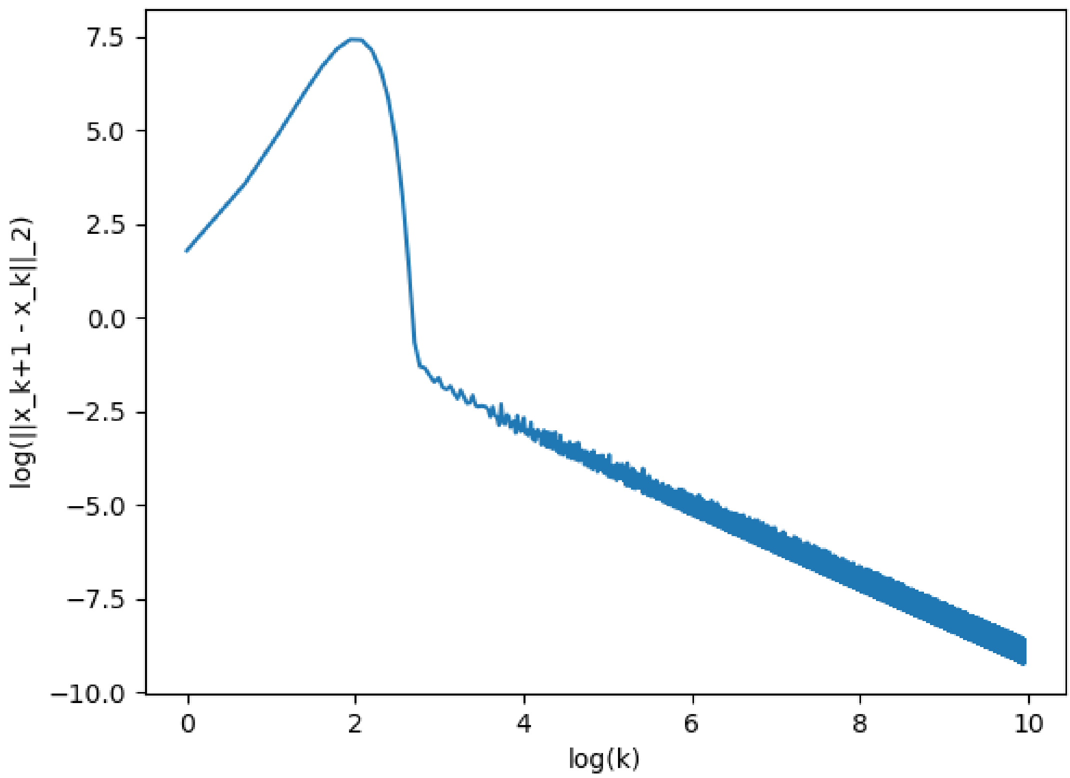


(2)

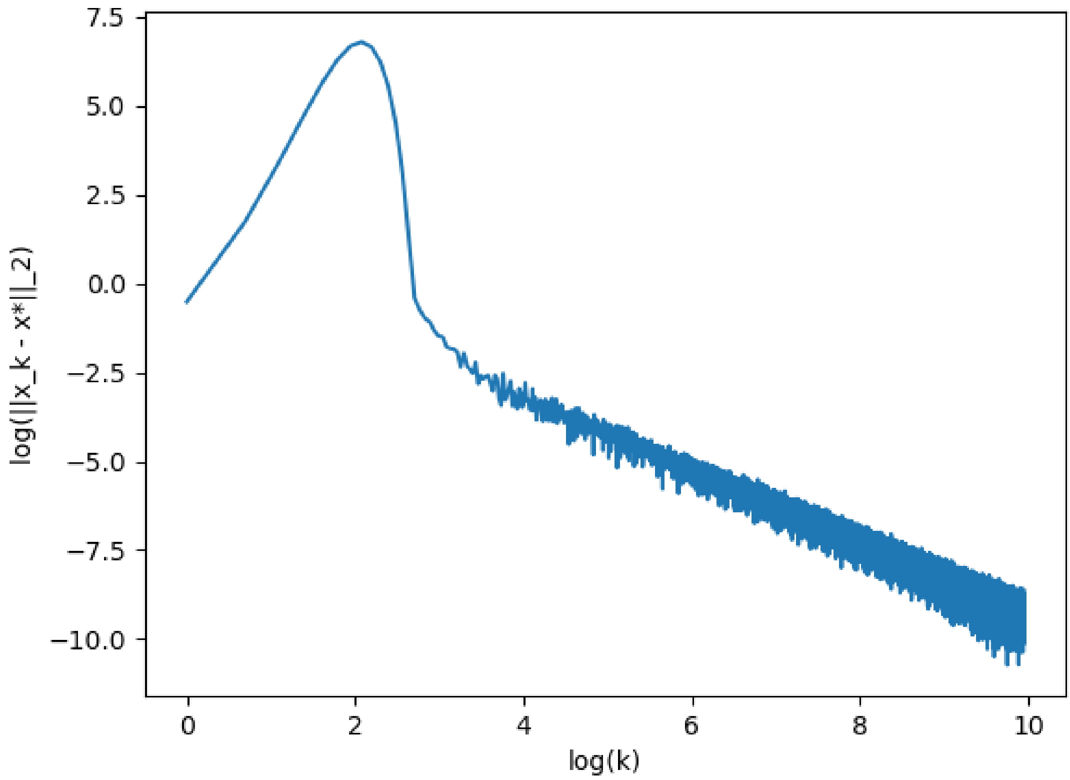
$$\text{对}f\text{做泰勒展开，有}f(y)=f(x)+\nabla f(x)^T(y-x)+\frac{1}{2}(y-x)^T\nabla^2 f(x)(y-x)$$
$$\geq f(x)+\nabla f(x)^T(y-x)+\frac{1}{2}m(y-x)^T(y-x),\text{其中}\nabla f(x)^2=A^TA$$
$$\text{因此有对于任意}x\text{，}y\in dom(f)\text{，有}(y-x)^T(A^TA-mI)(y-x)\geq 0,\text{也就是}A^TA-mI\text{半正定}$$
$$m\text{最大值为}A^TA\text{最小特征值}$$

解得m=0.7569455027845506， 最优的h(x)=12.52240996025351,x存储在根目录下的result/Q3/x2.csv中

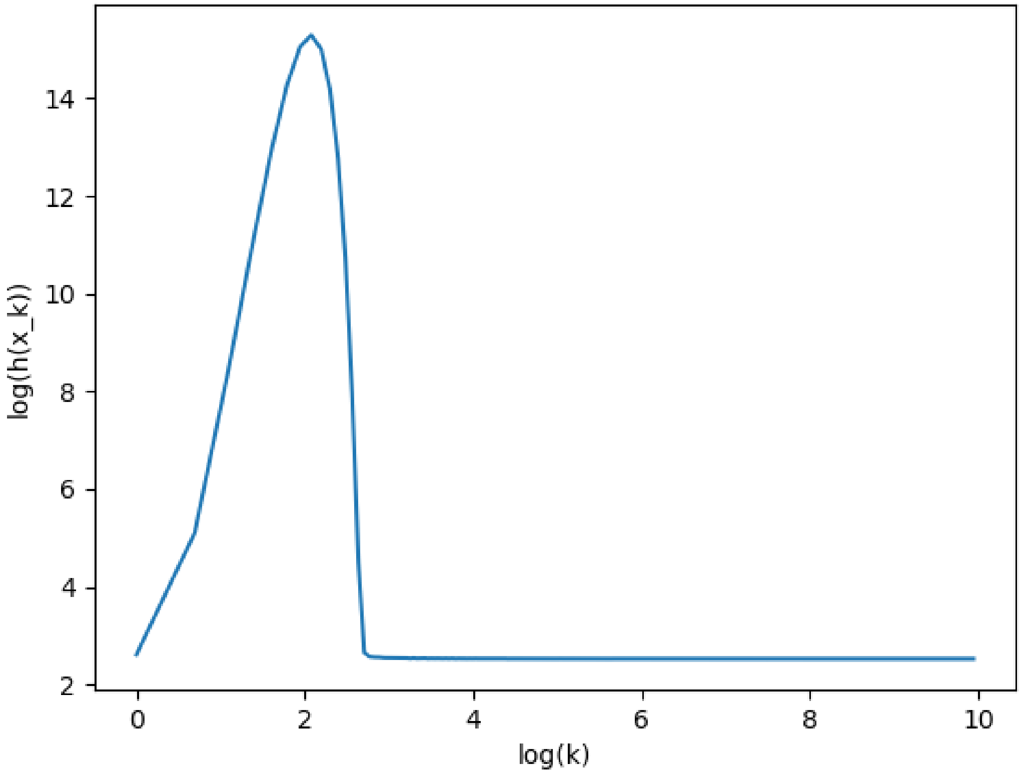
log(k)和log(||x_k+1 - x_k||_2)的图像如下：



log(k)和log(||x_k-x*||_2)的图像如下:



log(k)和log(h(x_k))的图像如下:


14
12
10
8
6
4
2
0
2
4
6
8
10
log(h(x_k))
log(k)