1.

(1)

1. (1) $f(x) = \frac{1}{2}(Ax-b)^T(Ax-b)$

$= \frac{1}{2}(x^TA^TAx - \cancel{2x^TA^T}\, 2b^TAx + b^Tb)$

$\nabla f(x) = A^TAx - A^Tb$

$\forall x, y,$

$\nabla f(x) - \nabla f(y) = A^TA(x-y)$

不妨令 $x-y = v \cdot \|x-y\|_2$，显然有 $\|v\|=1$，

则 $\|\nabla f(x) - \nabla f(y)\| = \|x-y\|_2 \cdot \|A^TAv\|$

若 $\forall x, y$，$\|\nabla f(x) - \nabla f(y)\| \le M\|x-y\|$

则有 $\forall v$, s.t. $\|v\|=1$,

$\sup_{\|v\|=1} \|A^TAv\| \le M$

(求最小的 M)

当 v 为 A^TA 绝对值最大的特征值对应的特征向量时

$\|A^TAv\|$ 最大，为 $\rho(A^TA)$

取 $M = \rho(A^TA)$，即有

$\forall x, y$,

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

(2) $\mathrm{Prox}_{\alpha g}\circ(I - \alpha\nabla f)x_k$

~~$\arg\min \alpha g(x)$~~

$= (I + \alpha\partial g(x_k))^{-1}(I - \alpha\nabla f(x_k))x_k$

$\nabla f(x) = A^TAx - A^Tb$

$g(x) = \|x\|_1$

$\varphi(x)_i = \partial g(x)_i = \begin{cases} 1 & x_i > 0 \\ 0 & x_i = 0 \\ -1 & x_i < 0 \end{cases}$

∴ 对 $\forall x, y$，$\|y\|_1 - \|x\|_1 \ge (y-x)^T\varphi(x)$ 成立

证明如下：$\|y\|_1 = \sum_{i=1} |y_i|$，$\|x\|_1 = \sum_{i=1} |x_i|$，

$\|y\|_1 - \|x\|_1 = \sum_{i=1} |y_i| - |x_i|$，

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

当 $x_i \ne 0$ 时，$\varphi(x_i) = \pm 1$，

有 $|y_i| \ge \varphi(x_i)y_i$，$-|x_i| = -x_i$，

右 ≥ 左

(2)

(2) $\mathrm{prox}_{\alpha g}(v) = \operatorname{argmin}_x \alpha g(x) + \frac{1}{2}\|x-v\|_2^2$

$$= \operatorname{argmin}_x \alpha\|x\|_1 + \frac{1}{2}\|x-v\|_2^2$$

$x^Tx - 2v^Tx + v^Tv$

若 $x^* = \mathrm{prox}_{\alpha g}(v)$,

则任取 $\phi(x) \in \partial g$，有 $\alpha \cdot \phi(x^*) + x^* - v = 0$

$\because g(x) = \|x\|_1$，可取 $\phi(x)_i = \begin{cases} 1 & x_i > 0 \\ 0 & x_i = 0 \\ -1 & x_i < 0 \end{cases}$,

则有：$v_i = \begin{cases} x_i + \alpha & x_i > 0 \\ x_i & x_i = 0 \\ x_i - \alpha & x_i < 0 \end{cases}$

而 $v = (I - \alpha\nabla f)x_k$

$$= (x_k - \alpha A^TAx_k + \alpha A^Tb)$$

$\therefore$ 有 $x_{k+1,i} = \begin{cases} v_i - \alpha & v_i > \alpha \\ 0 & \text{others} \\ v_i + \alpha & v_i < -\alpha \end{cases}$

其中 $\alpha = \frac{1}{m}$

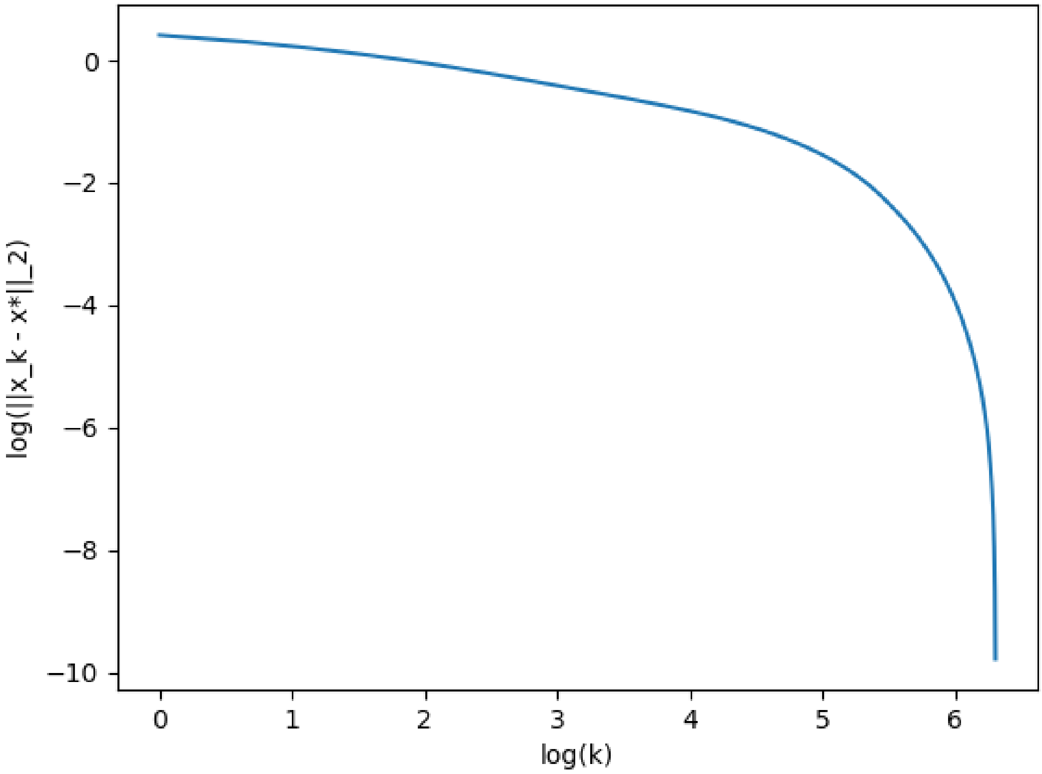
(3) ~~收敛有 $h(x_k) - h(x^*) \le \frac{m\|x_0 - x^*\|^2}{2k}$~~

$\therefore$ m ~~越大，收敛越~~

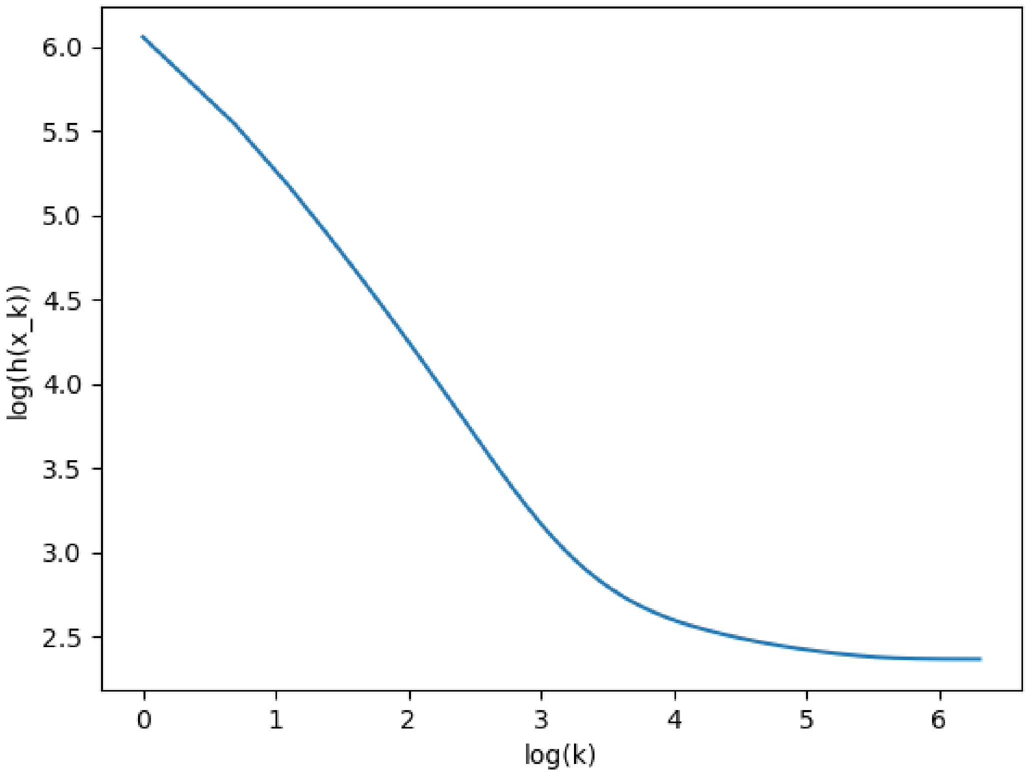
$y^Ty - (2x_{k+1} - 2x_k)y$

(3)使用A1,b1求解，迭代547次后，得到最优值为10.640556454308745，x结果储存在根目录下result/Q1/x1.csv中

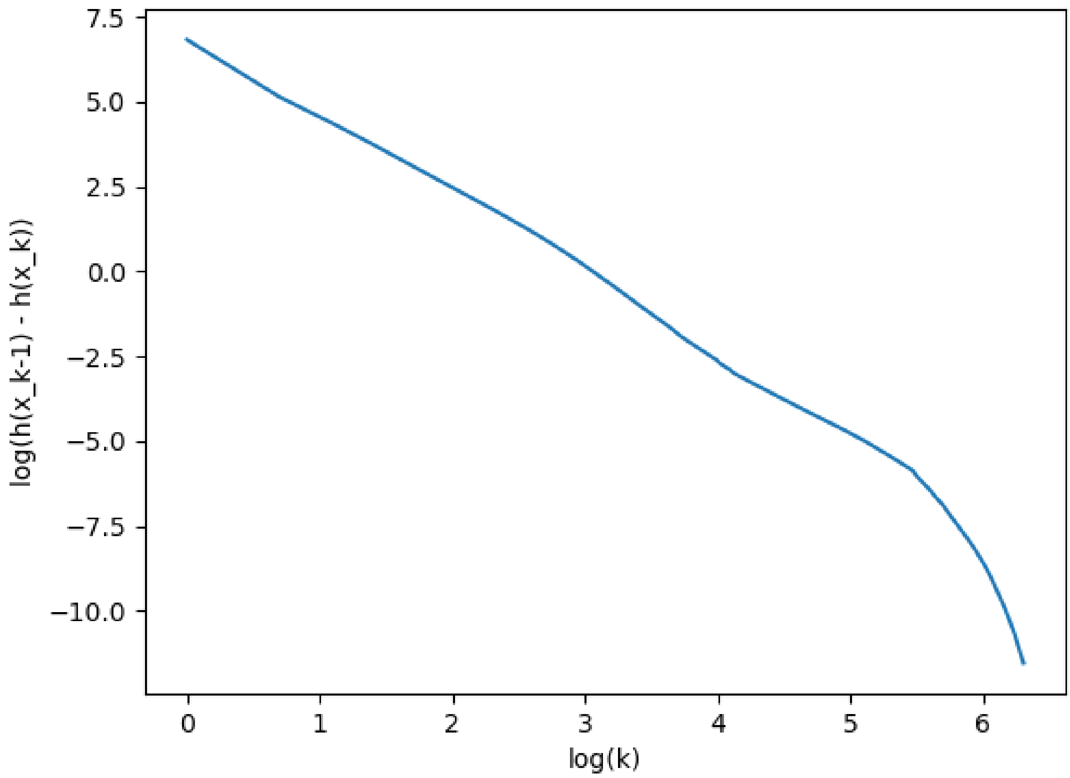
log(k)与log(||x_k-x*||_2)的图像如下



log(k)与log(h(x_k))的图像如下

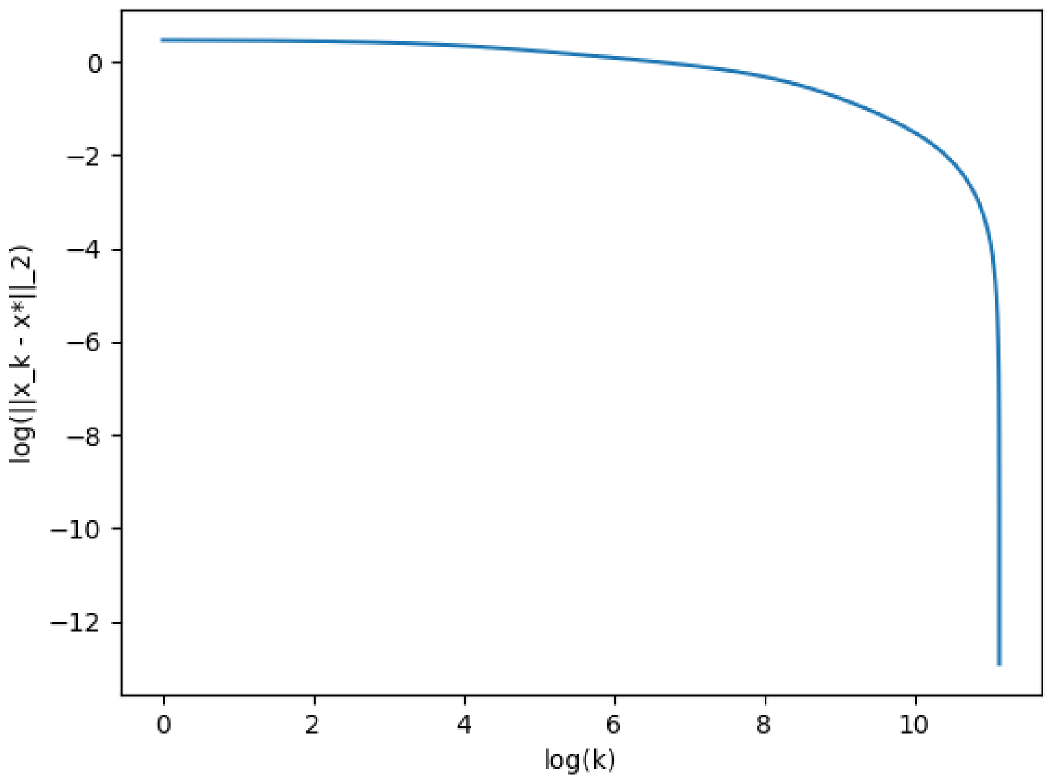


log(k)与log(h(x_k-1) - h(x_k))的图像如下

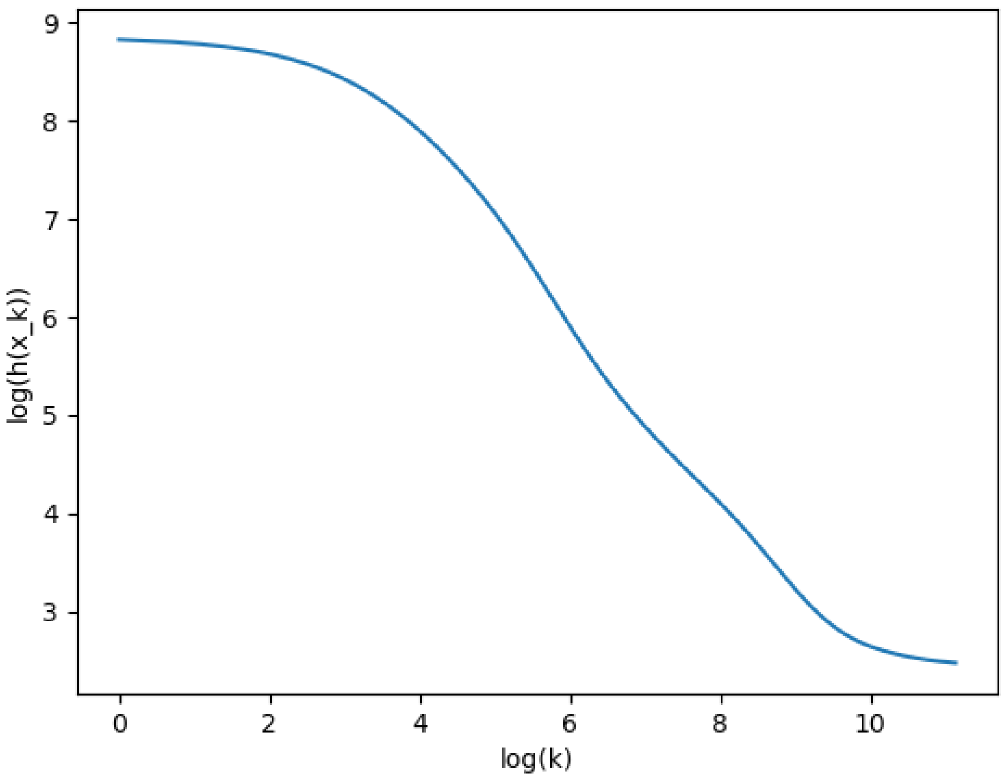


使用A2,b2求解，迭代68188次后，得到最优值为11.938330644320333，x结果储存在根目录下result/Q1/x2.csv中

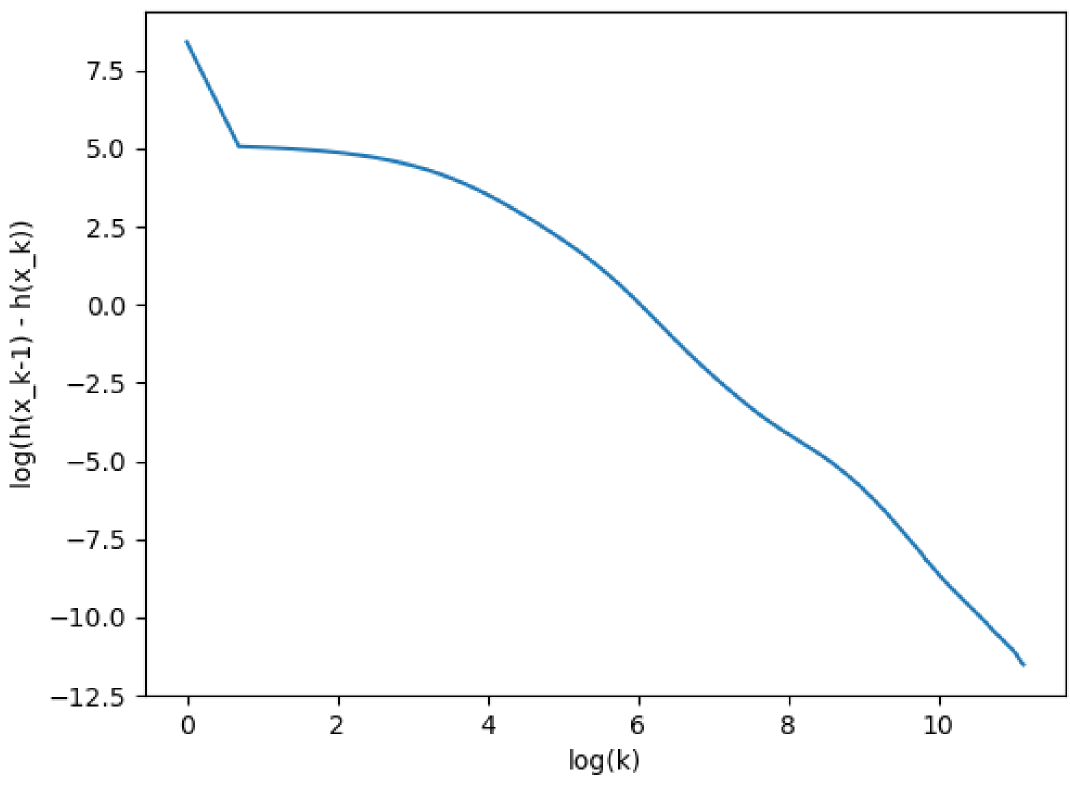
log(k)与log(||x_k-x*||_2)的图像如下



log(k)与log(h(x_k))的图像如下



log(k)与log(h(x_k-1) - h(x_k))的图像如下



(4)

可以看出，$A1$收敛明显比$A2$要快很多，快了100多倍该方法收敛性满足如下条件：

$$h(x_k)-h(x^*)\leq\frac{M||x_0-x^*||_2^2}{2k}$$

求得A_1的M_1为3040.0792115057775, A_2的M_2为1638363.7463906936

A_1迭代$k_1=547$次收敛，A_2迭代$k_2=68188$次收敛

$\frac{M_2}{M_1}=538.9214005312704, \frac{k_2}{k_1}=124.6581352833638$，两者在同一数量级，与理论相符合。

2.

(1)推导如下

清華大學

2.

$\mathrm{Prox}_{\alpha g}(z_k) = \arg\min_x \frac{1}{2}\|x-z_k\|_2^2 + \alpha I_{Ax=b}(x)$

$= \arg\min_x \frac{1}{2}\|x-z_k\|_2^2,$

s.t. $Ax=b$,

∵为凸问题，满足Slater条件，∴可以用对偶问题求

$L(x,v) = \frac{1}{2}\|x-z_k\|_2^2 + v^T(Ax-b)$

$\frac{\partial L}{\partial x} = x - z_k + A^T v$

$\frac{\partial^2 L}{\partial x^2} = I$，正定

$x = z_k - A^T v$ 时，取得最大值

$g(v) = \sup_x L(x,v)$

$= -\frac{1}{2} v^T A A^T v + (Az_k-b)^T v$

∴ 求解 $\min_v \frac{1}{2} v^T A A^T v - (Az_k-b)^T v$

$\frac{\partial -g(v)}{\partial v} = AA^T v - (Az_k-b)$

$\frac{\partial^2 g(v)}{\partial v} = AA^T$ 正定

∴ $v = (AA^T)^{-1}(Az_k-b)$

$x_{k+1} = z_k - A^T v$

$= z_k - A^T(AA^T)^{-1}(Az_k-b)$

$\mathrm{Prox}_{\alpha f}(2x_{k+1}-z_k)$

$= \arg\min_y \frac{1}{2}\|y-2x_{k+1}+z_k\|_2^2 + \alpha\|y\|_1$

为 L_1 范数问题，

有 $y_{k+1,i} = \begin{cases} v_i-\alpha & v_i>\alpha \\ 0 & \text{others} \\ v_i+\alpha & v_i\le -\alpha \end{cases}$

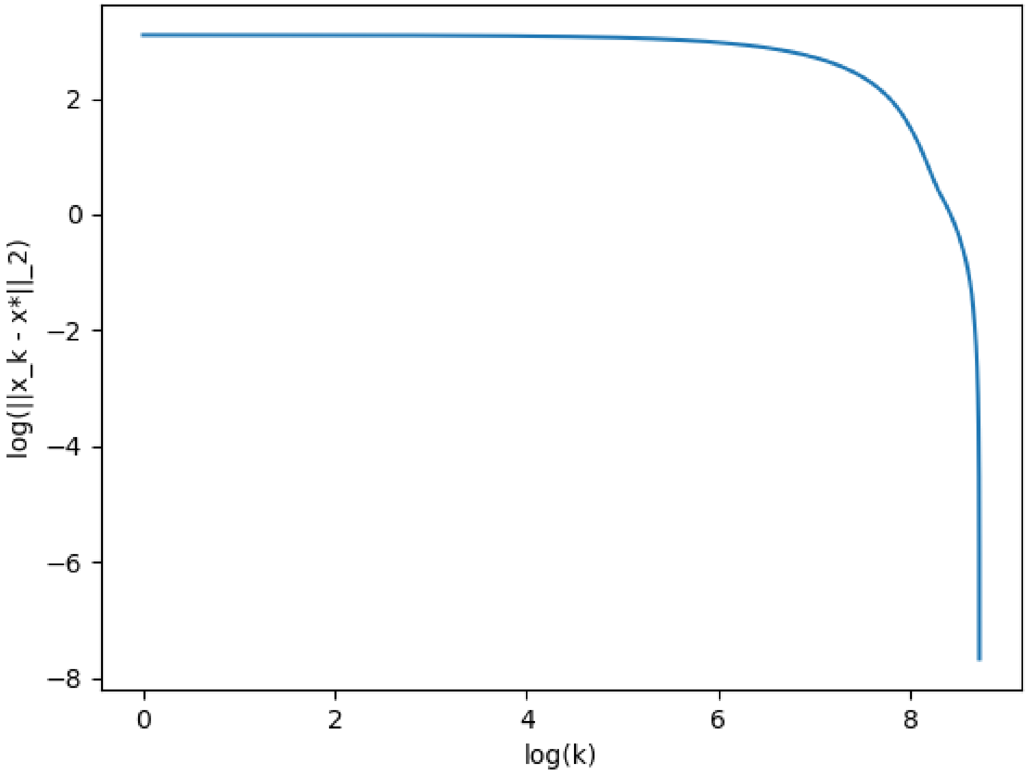
其中 $v = 2x_{k+1}-z_k$

~~则有更新~~ $\cancel{x_{k+1} = z_k - A^T(Az_k-b)}$

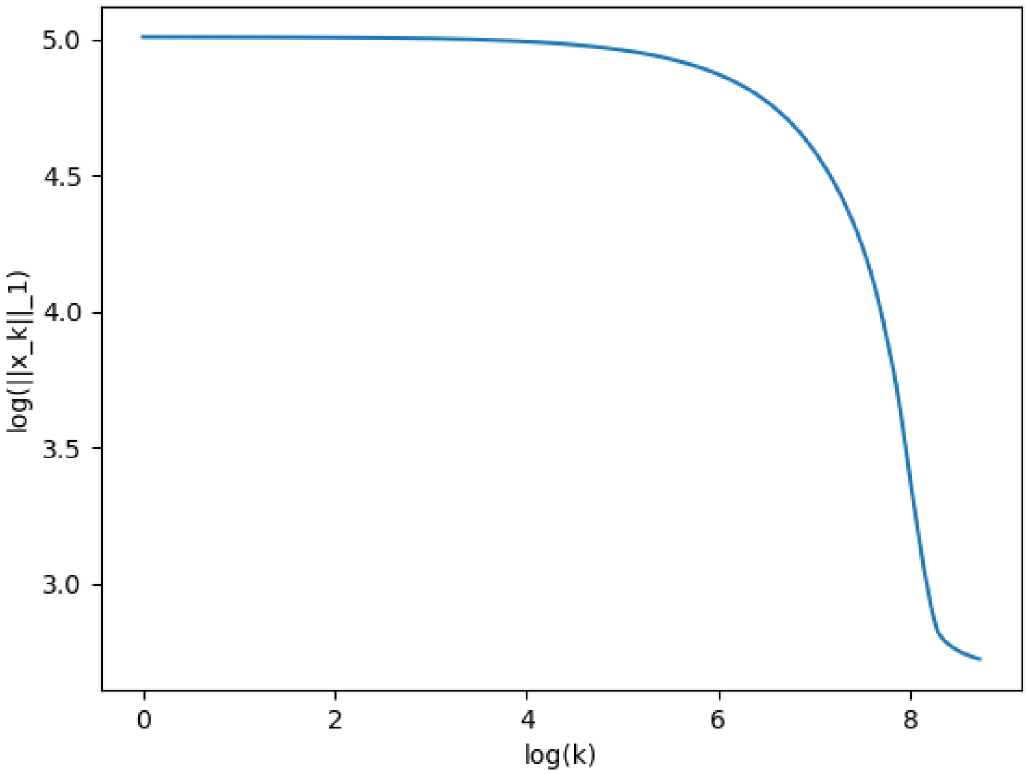
$\cancel{y_{k+1} =}$

取alpha=1e-3,迭代得到最优值为15.245250663531616，最优解存储在根目录下result/Q2/x1.csv中

log(k)与log(||x_k-x*||_2)的图像如下：



log(k)与log(||x_k||_)的图像如下：



(2)推导如下

② $x_{k+1} = \arg\min_x L_\alpha(x, y_k, u_k, v_k)$:

取 $\|x\|_1$ 次梯度为 $\phi(x)_i = \begin{cases} 1 & x_i > 0 \\ 0 & x_i = 0 \\ -1 & x_i < 0 \end{cases}$

$$\frac{\partial L_\alpha(x, y_k, u_k, v_k)}{\partial x} = \phi(x) + u + \alpha x - \alpha y$$

$$\frac{\partial^2 L_\alpha}{\partial x^2} = \alpha I \text{ 正定}$$

$\therefore$ 当 $\alpha x + \phi(x) = \alpha y - u$ 时取极值

x_{k+1} 有 $x_i = \begin{cases} y_{ki} - \frac{u_{ki}}{\alpha} - \frac{1}{\alpha} & (y_{ki} - \frac{u_{ki}}{\alpha} > \frac{1}{\alpha}) \\ 0 & \text{others} \\ y_{ki} - \frac{u_{ki}}{\alpha} + \frac{1}{\alpha} & (y_{ki} - \frac{u_{ki}}{\alpha} < -\frac{1}{\alpha}) \end{cases}$

$y_{k+1} = \arg\min_y L_\alpha(x_{k+1}, y, u_k, v_k)$

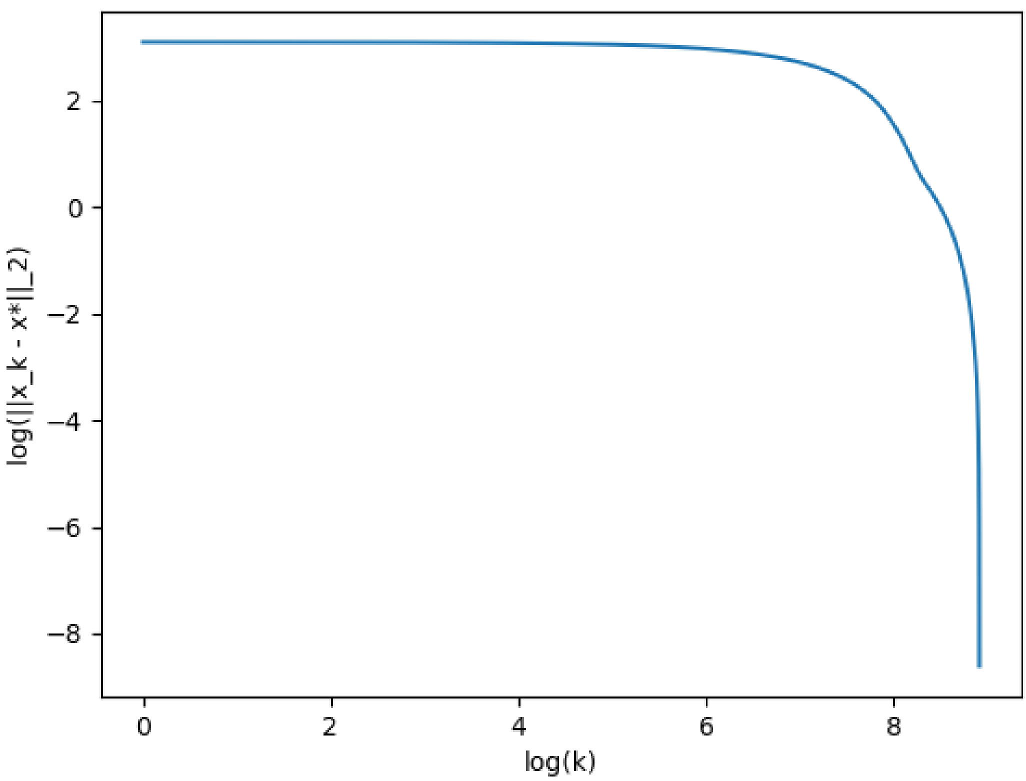
$$\frac{\partial L_\alpha}{\partial y} = -u + A^T v + \alpha(y - x_{k+1}) + \alpha(A^T A y - A^T b)$$

$$\frac{\partial^2 L_\alpha}{\partial y^2} = \alpha(A^T A + I) \text{ 正定}$$

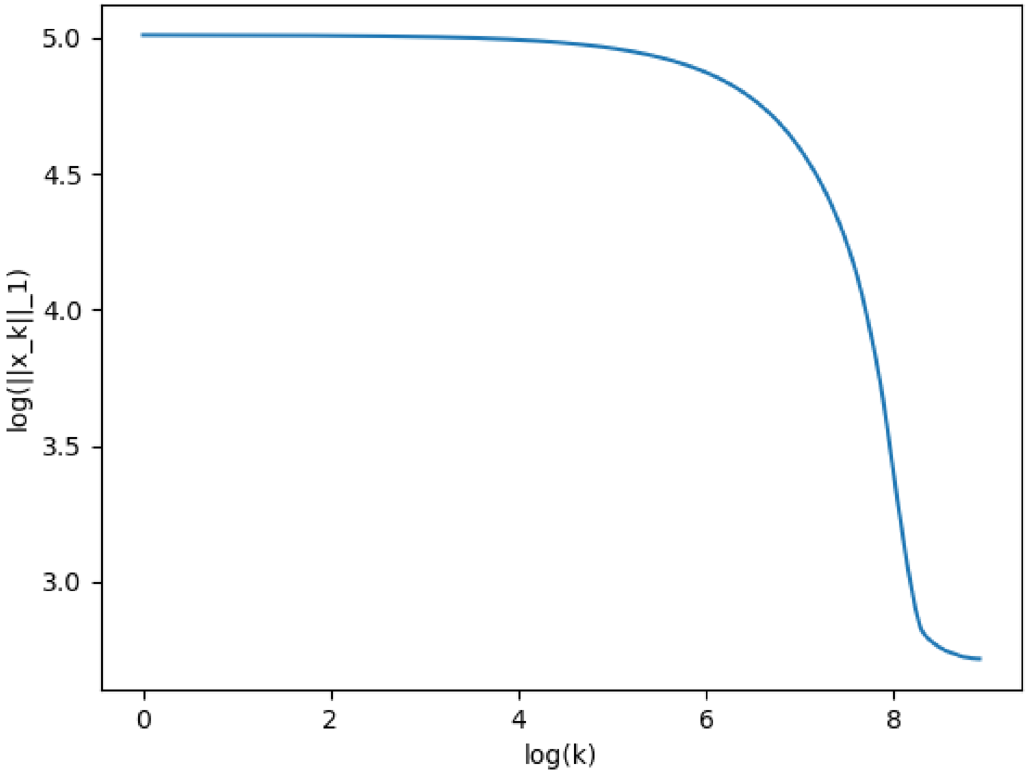
$$\therefore y_{k+1} = (A^T A + I)^{-1}\left(A^T b + x_{k+1} + \frac{u_k}{\alpha} - \frac{A^T v_k}{\alpha}\right)$$

取alpha=1000,迭代得到最优值为15.1189574103693，最优解存储在根目录下result/Q2/x2.csv中

log(k)与log(||x_k-x*||_2)的图像如下:



log(k)与log(||x_k||_)的图像如下:



(3)推导如下

② $x_{k+1} = \arg\min_x L_\alpha(x, y_k, u_k, v_k)$：

取 $\|x\|_1$ 次梯度为 $\phi(x)_i = \begin{cases} 1 & x_i > 0 \\ 0 & x_i = 0 \\ -1 & x_i < 0 \end{cases}$

因

$$\frac{\partial L_\alpha(x, y_k, u_k, v_k)}{\partial x} = \phi(x) + u + \alpha x - \alpha y$$

$$\frac{\partial^2 L_\alpha}{\partial x} = \alpha I \text{ 正定}$$

$\therefore$ 当 $\alpha x + \phi(x) = \alpha y - u$ 时取到最佳

x_{k+1} 有 $x_i = \begin{cases} y_{k_i} - \frac{u_{k_i}}{\alpha} - \frac{1}{\alpha} & (y_{k_i} - \frac{u_{k_i}}{\alpha} > \frac{1}{\alpha}) \\ 0 & \text{others} \\ y_{k_i} - \frac{u_{k_i}}{\alpha} + \frac{1}{\alpha} & (y_{k_i} - \frac{u_{k_i}}{\alpha} < -\frac{1}{\alpha}) \end{cases}$

$y_{k+1} = \arg\min_y L_\alpha(x_{k+1}, y, u_k, v_k)$

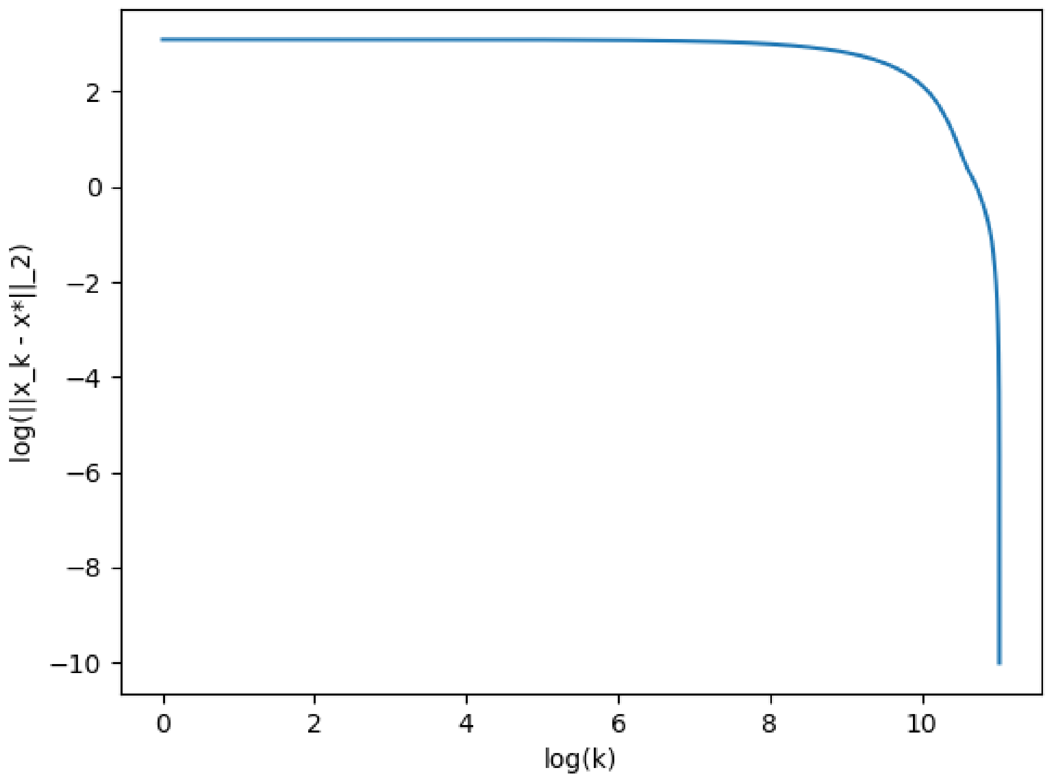
$$\frac{\partial L_\alpha}{\partial y} = -u + A^T v + \alpha(y - x_{k+1}) + \alpha(A^T A y - A^T b)$$

$$\frac{\partial^2 L_\alpha}{\partial y^2} = \alpha(A^T A + I) \text{ 正定}$$

$$\therefore y_{k+1} = (A^T A + I)^{-1}\left(A^T b + x_{k+1} + \frac{u_k}{\alpha} - \frac{A^T v_k}{\alpha}\right)$$

取alpha=10000/(M + 1),迭代得到最优值为15.245069566491548，最优解存储在根目录下result/Q2/x3.csv中

log(k)与log(||x_k-x*||_2)的图像如下：



log(k)与log(||x_k||_)的图像如下：