第二次大作业实验报告

王嘉禾 2193211079 计算机试验班 001

问题一

问题描述

例1 两个变量的二次规划问题：$\min f(x)=\frac{1}{2}(x_1^2+\gamma x_2^2)$，$\gamma>0$，最优解$x^*=(0,0)$。

$$x^k=\left(\frac{\gamma-1}{\gamma+1}\right)^k\begin{bmatrix}\gamma\\(-1)^k\end{bmatrix}$$

编程证明 ：

当γ=1 时，一次迭代可求得最优解

当γ离 1 不远时，收敛速度很快

当γ>>1 或γ<<1 时，收敛速度将会非常慢

解题思路

按照梯度下降的步骤，迭代路径已经给出，迭代过程中以梯度的二范数的规模作为迭代的终止条件，只需求得迭代路径上每个点的梯度的二范数，并对不同的γ记录迭代次数即可。γ的采样从 1/1000 到 1000，以 2 为公比进行采样，绘制迭代次数与γ值的关系图

程序设定

ε=1e-6 即迭代时到梯度的二范数小于等于 1e-6 时停止

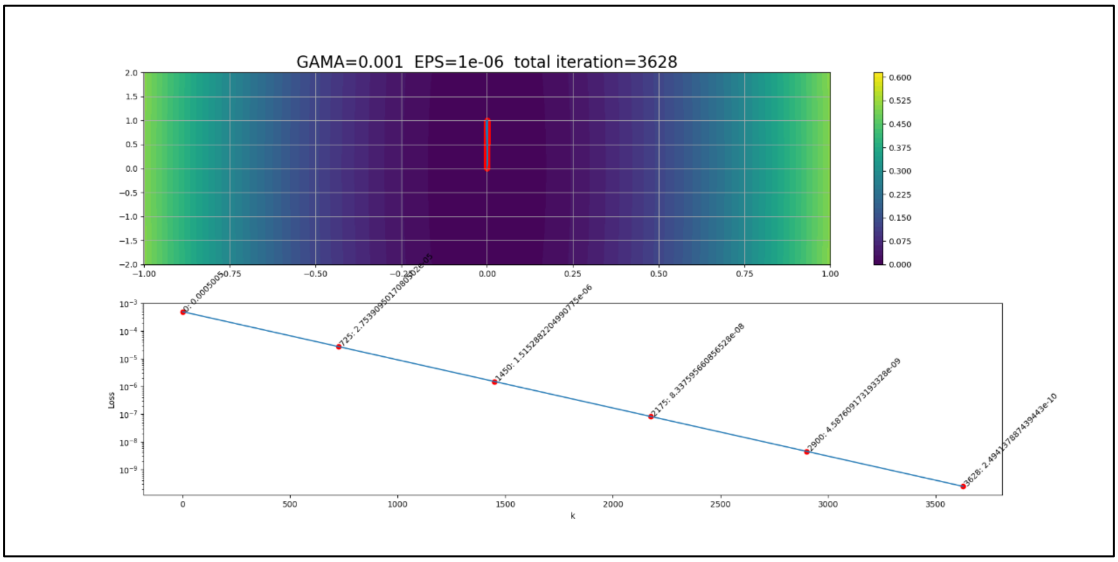
代码实现

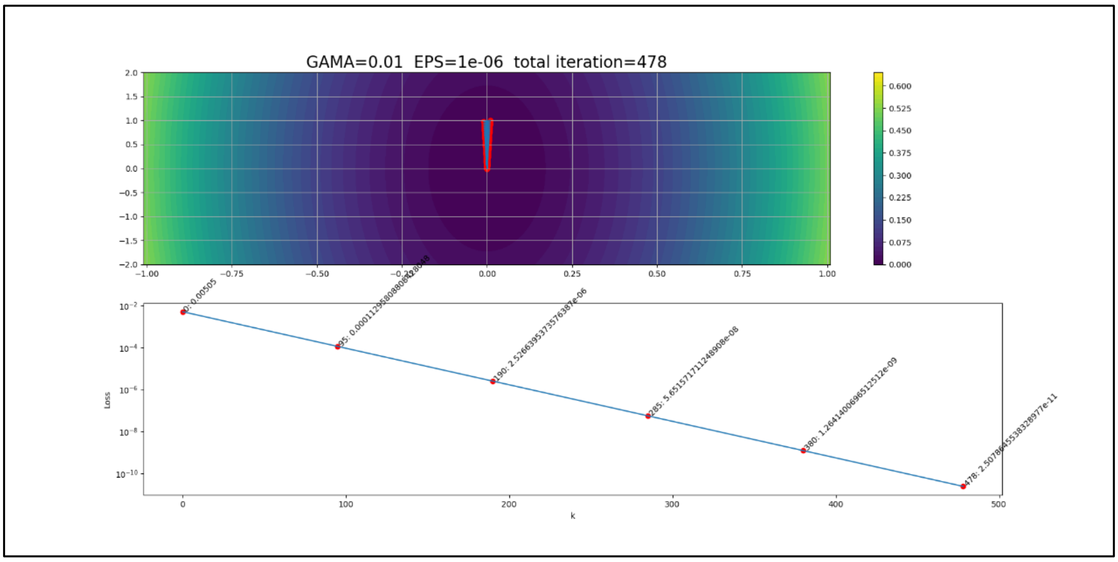
见 1_gradient.py

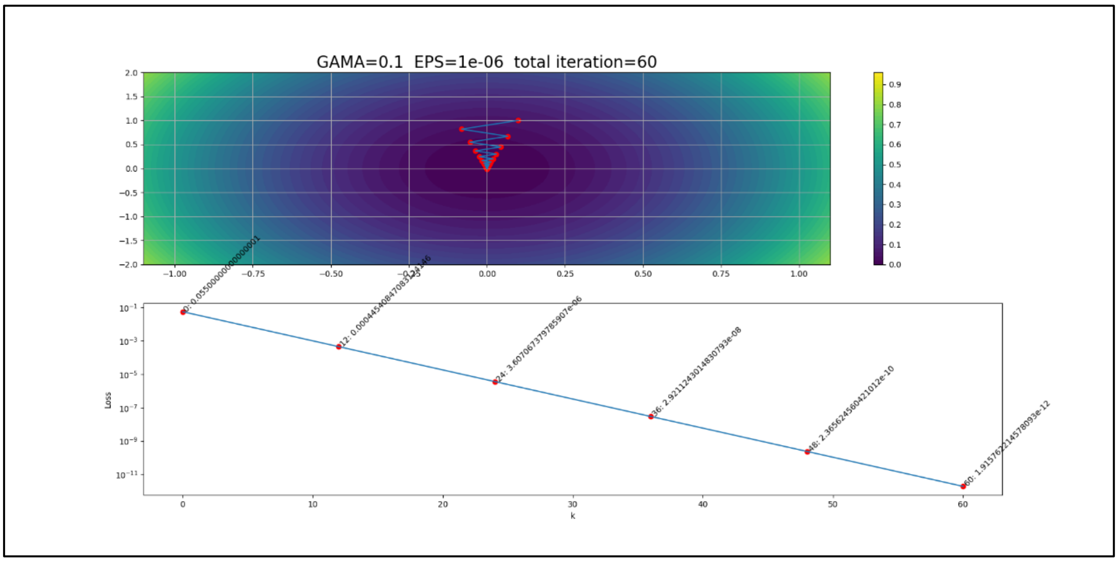
运行结果

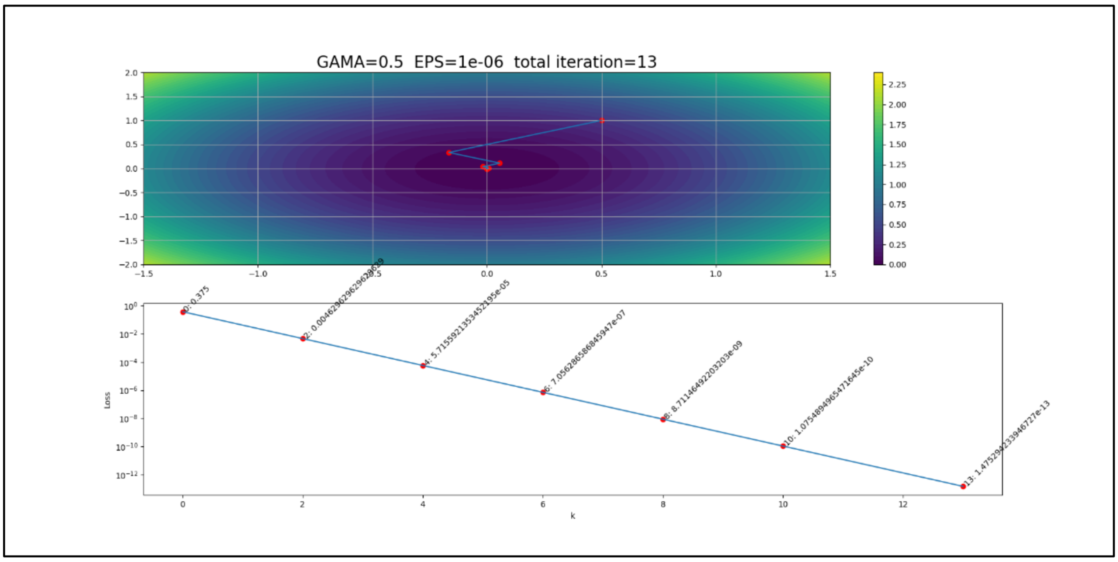
当γ分别取 [0.001, 0.01, 0.1, 0.5, 1, 2, 3, 10, 50, 100] 时，迭代次数情况如下表所示

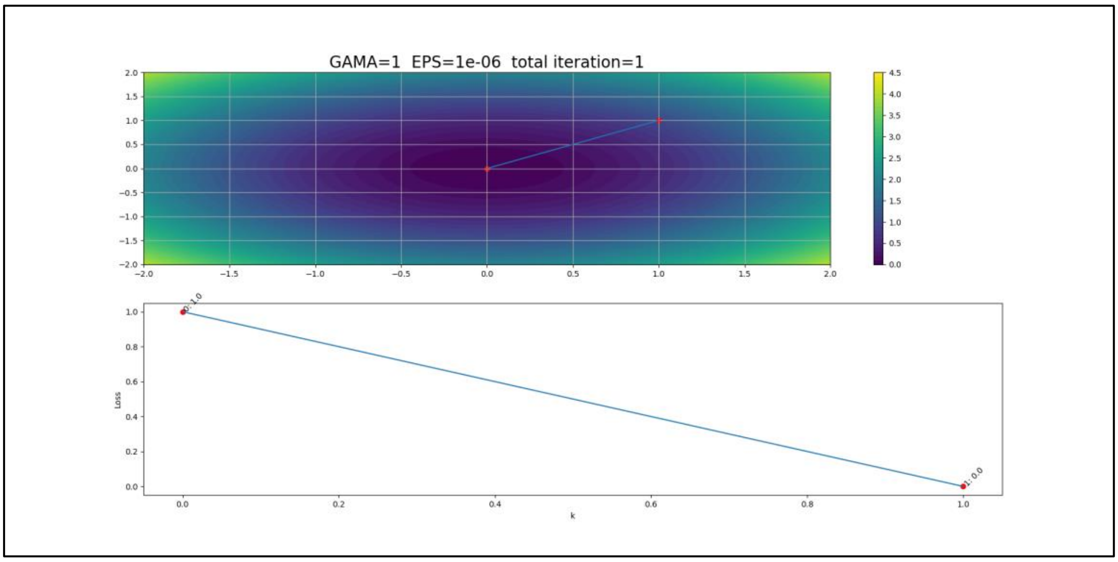
γ	0.001	0.01	0.1	0.5	1	2	3	10	50	100
次数	3628	478	60	13	1	14	23	83	452	939

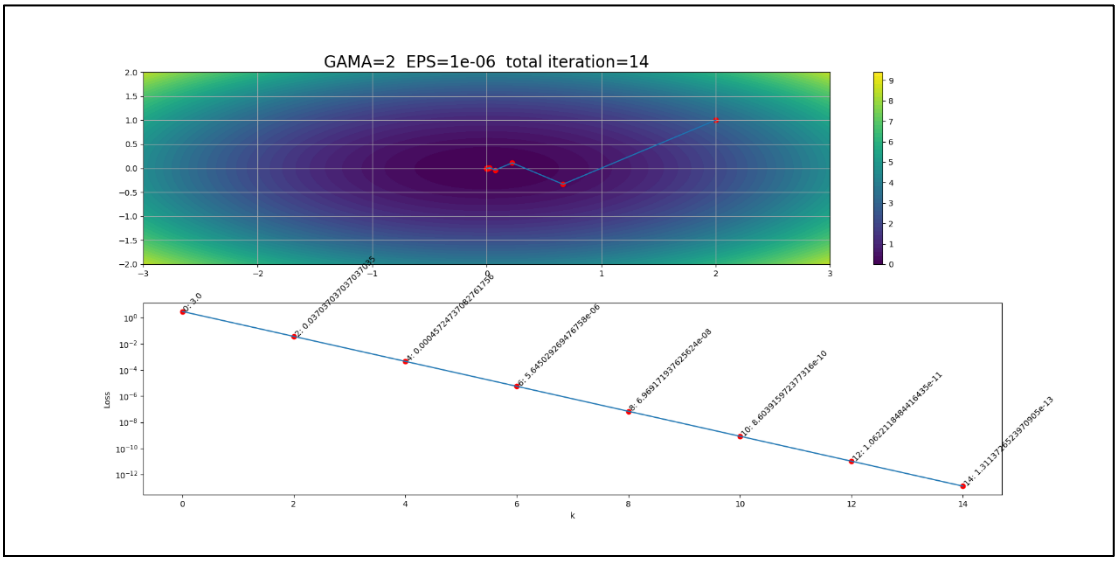


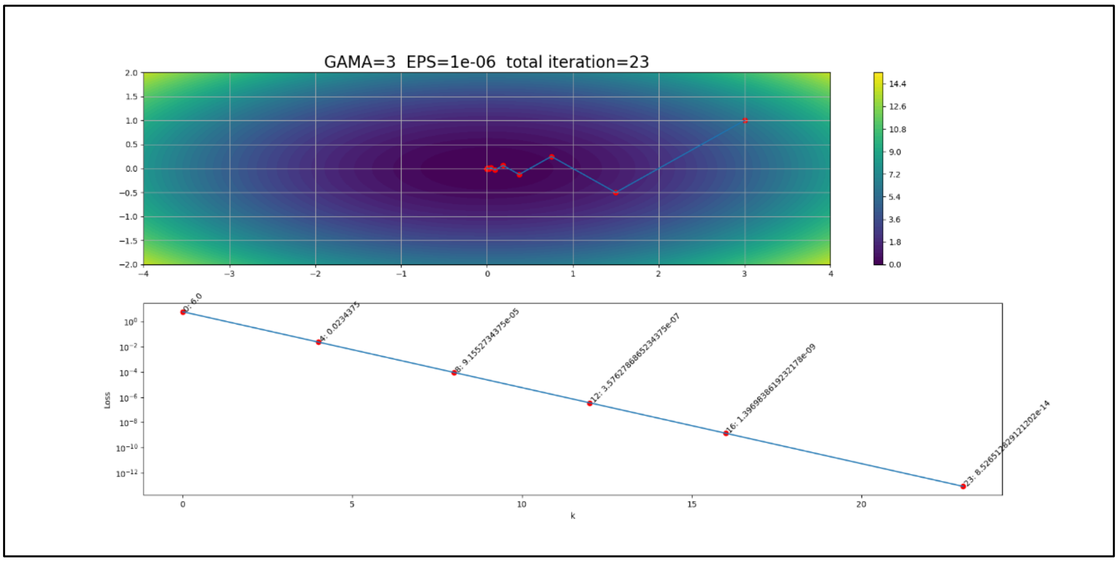

GAMA=0.01 EPS=1e-06 total iteration=478
Loss
k

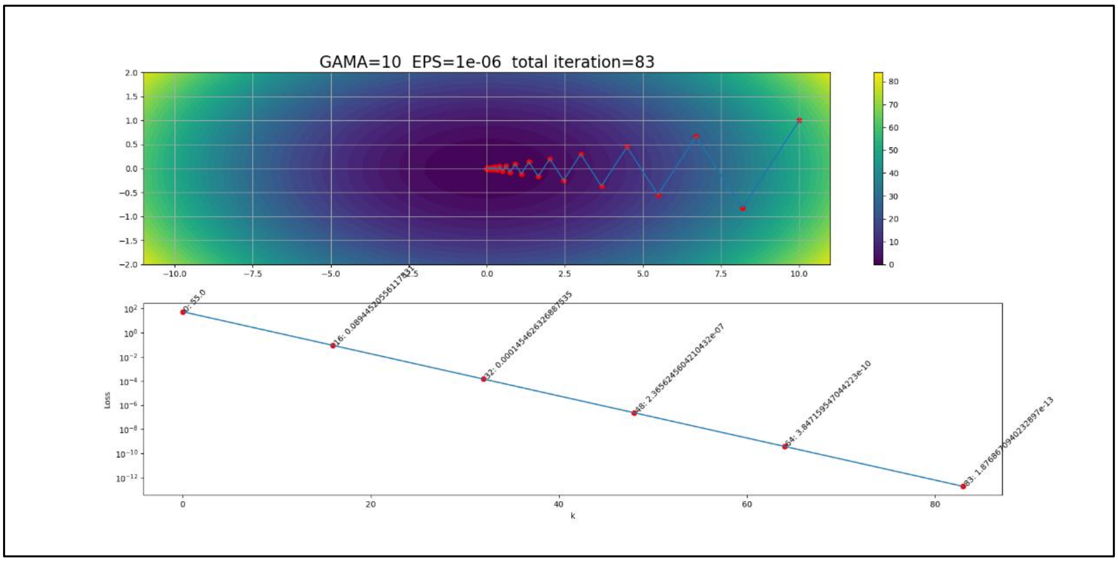

GAMA=0.1 EPS=1e-06 total iteration=60
Loss
k

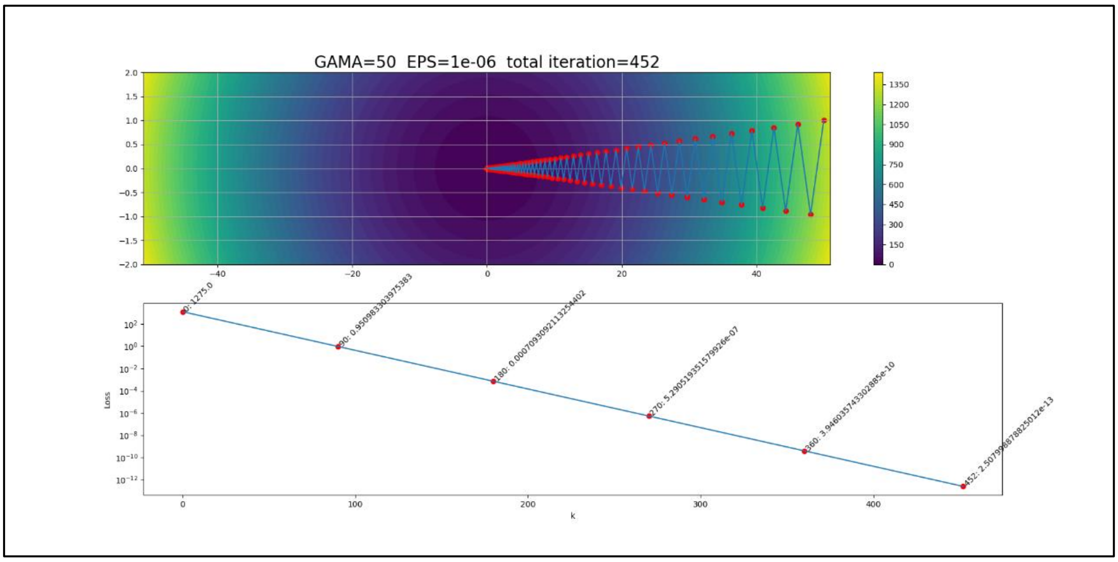

GAMA=0.5 EPS=1e-06 total iteration=13
Loss
k

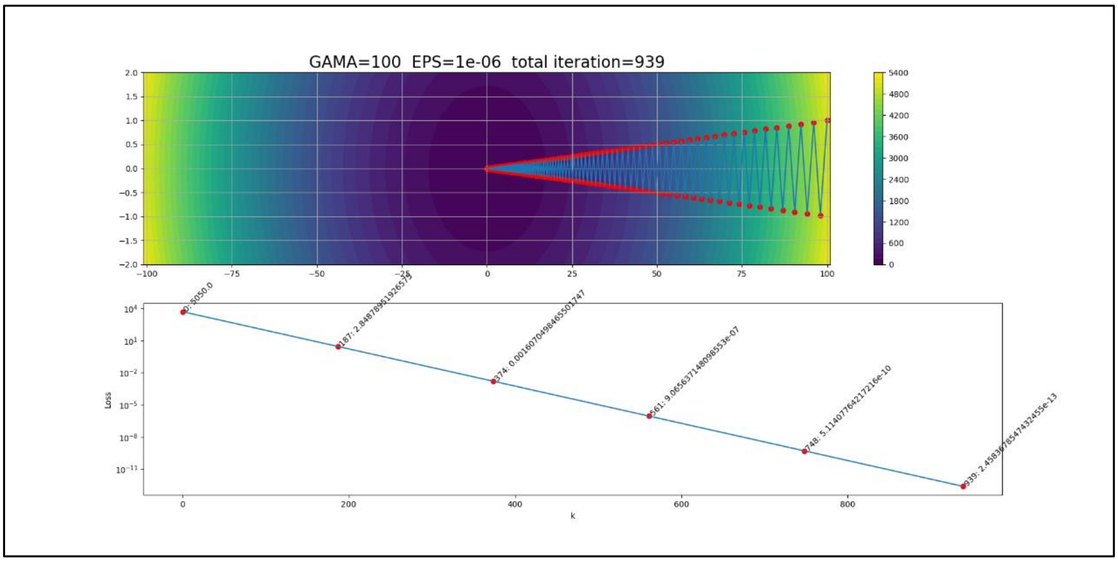

GAMA=1 EPS=1e-06 total iteration=1
Loss
k


GAMA=2 EPS=1e-06 total iteration=14
Loss
k

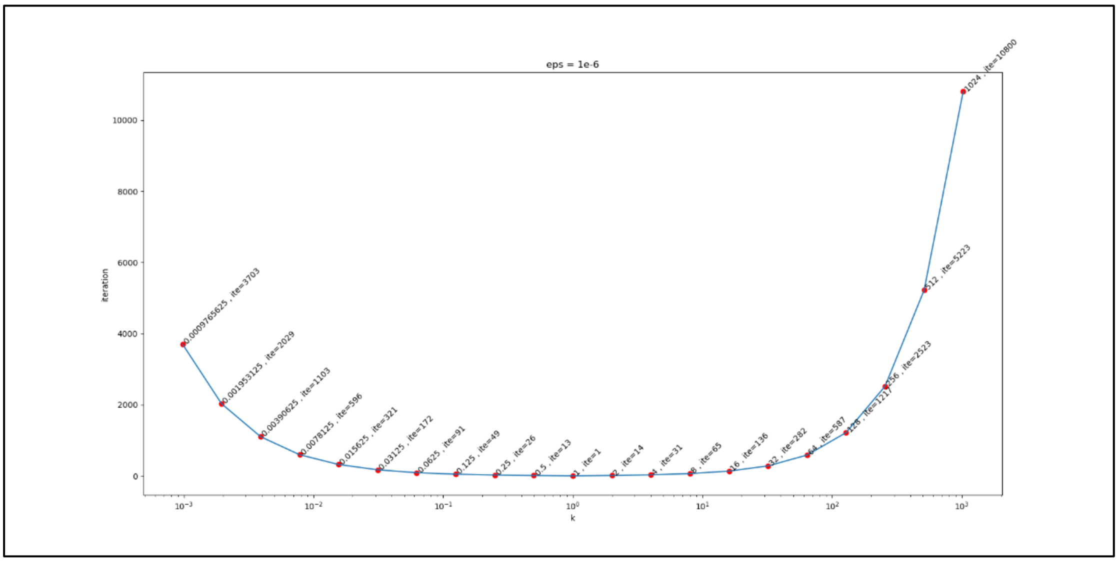

GAMA=3 EPS=1e-06 total iteration=23
Loss
k







迭代次数与γ的关系如下图所示



实验结论

可见，当γ=1 时，一次迭代即可达到最优值；当γ距离 1 较近时，收敛速度很快，迭代次数大约在 10^1 的数量级；当γ>>1 或γ<<1 时，收敛速度很慢，当γ达到 1000 或 1/1000 时，迭代次数也达到 10^3 甚至 10^4 的量级

问题二

问题描述

对于$\mathbb{R}^2$空间非二次规划问题 $\min f(x) = e^{x_1+3x_2-0.1} + e^{x_1-3x_2-0.1} + e^{-x_1-0.1}$，分析回溯直线搜索采用不同的 α, β 值时，误差随迭代次数改变的情况。注：初始值相同。

解题思路

采用梯度下降方法+回溯直线搜索。在每个循环中，首先获取负梯度方向 dk，判断终止条件即 dk 的二范数是否小于 eps。若未终止，则采用回溯直线搜索获取步长并进行迭代，在迭代结束时，记录下迭代次数。对于不同的α和β值，记录下迭代次数并画出热点图对比

程序设定

（1） 程序中对迭代次数作了两组对比，第一组以(-1,1) (0,1) (1,1) (-1,0) (0,0) (1,0) (-1,-1) (0,-1) (1,-1)九个点为起始点，ε=1e-4；第二组以(-2,2) (0,2) (2,2) (-2,0) (0,0) (2,0) (-2,-2) (0,-2) (2,-2)九个点为起始点，ε=1e-6；并分别记录下α从 0.1~0.4，β从 0.1~0.9 的迭代次数，以β为横坐标，α为纵坐标画出迭代次数的热点图。

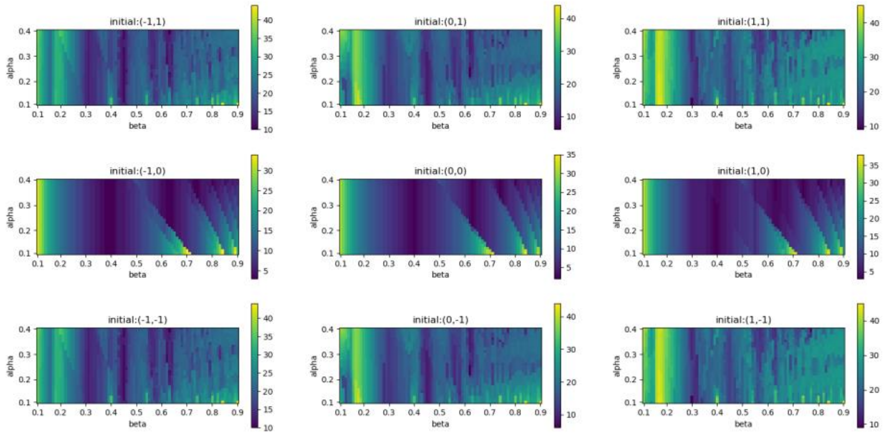
（2） 程序中对迭代过程也作了两组对比，分别以(2,2) (4,0)为起点，记录α=[0.1 0.2 0.3]，β=[0.3 0.5 0.7]时的迭代过程中误差的下降过程，其中（图一：ε=1e-4，图二：ε=1e-6）。

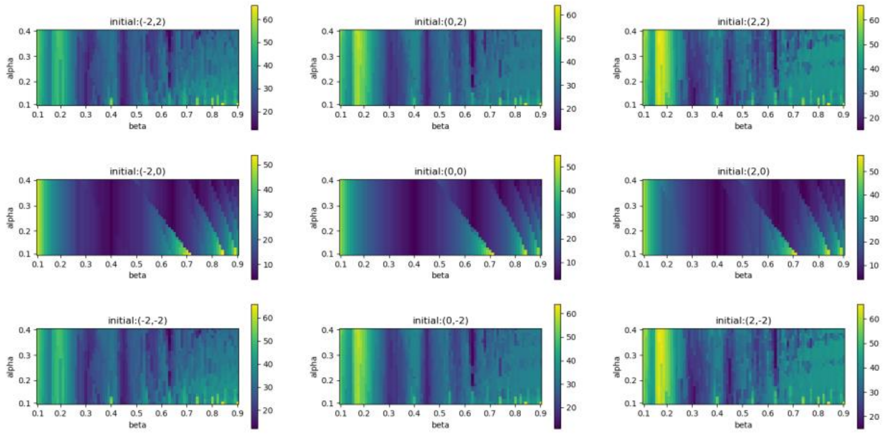
代码实现

见 2_gradient_exp.py

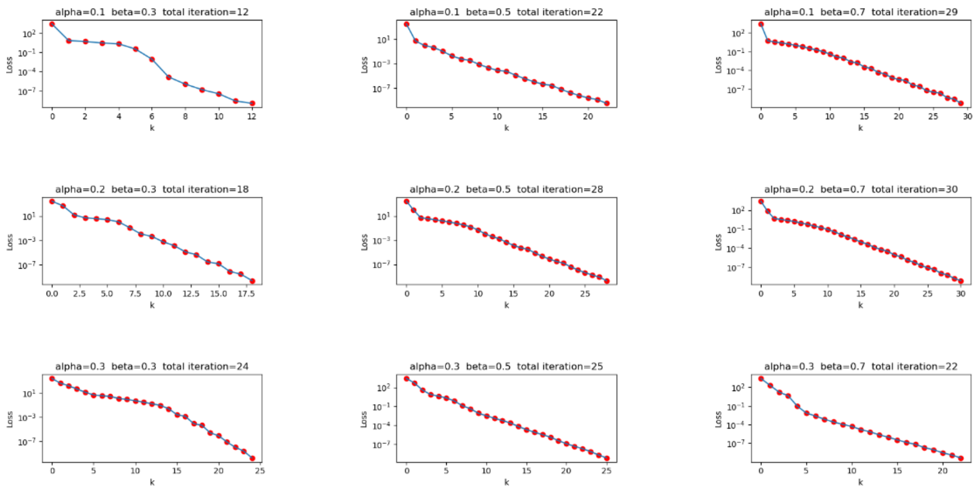
实验结果

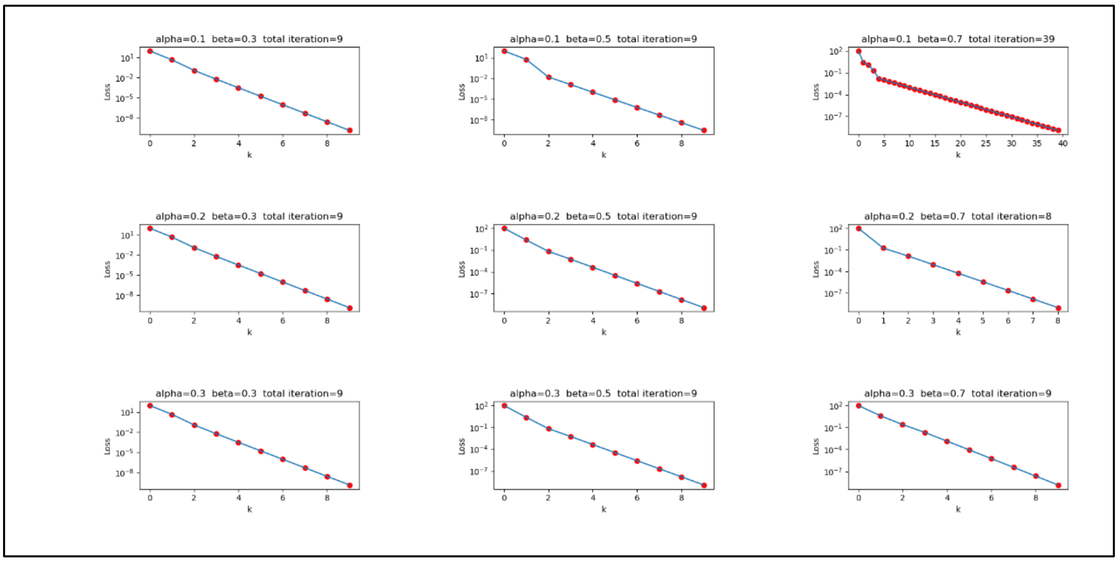
对于迭代次数，当α取遍[0.1,0.4]的值和β取遍[0.1,0.9]的值时，迭代次数如下图所示（图一：ε=1e-4 图二：ε=1e-6）





对于迭代过程，当α=[0.1 0.2 0.3]，β=[0.3 0.5 0.7]时，迭代中误差下降过程如下图所示（图一：起始点(2,2) eps=1e-4 图二：起始点(4,0) eps = 1e-6）





实验结论

(1) 对于第一组对比实验，不难发现，热力图大多成竖向条带分布，并且，有一条斜率为正的分割线。当β较小时（比如β<0.5）,不同的α对迭代次数几乎没有影响；而当β较大时，热力图呈现出“上暗下亮”的情况，这说明当α较大时迭代次数较小，而α较大时迭代次数较大。这在一定范围内是符合直觉的，因为当α过小时，对于一次迭代下降的差值要求很低，则迭代次数倾向于更多。

(2) 在第一组对比实验中，通过对比不同初始点的实验结果可看出，当β在 3.5 和 4.5 附近时，迭代次数均很少。特别的，当初始点离全局最优解较近时，β在 0.4 附近有大面积的迭代次数很少的区域。由此可猜测，对于该问题，从任何初始点开始迭代，令β为 0.35 或 0.45 可能可以使得迭代次数尽量少。

(3) 在第二组对比实验中可看到。对于第一种起始点，当β较低时，误差的下降过程呈一个先慢后快的趋势，而随着β的增大，误差的下降呈平稳的指数下降。对于第二种起始点，大多数情况下误差的下降均呈平稳的指数下降。但当α=0.1 β=0.7 时，迭代次数大幅增加。

(4) 由此可见，如果选择了较差的α和β，可能使得迭代过程大幅增加，程序效率降低，并且，对于该函数的梯度下降，参数β对于下降过程起到主导作用，当β较大时，α才开始影响迭代过程。

问题三

问题描述

$$\min_{x} f(x) = e^{x_1+3x_2-0.1} + e^{x_1-3x_2-0.1} + e^{-x_1-0.1}$$

- ☐ 使用Newton下降方法，令回溯直线搜索时$\alpha = 0.1$，$\beta = 0.7$。

画出误差和迭代次数的关系

解题思路

和之前的梯度下降方法类似，在获取下降方向和进行回溯直线搜索时有所改动。计算下降方向时需计算牛顿下降方向，因此需计算 Hessian 矩阵的逆矩阵和梯度的乘积，并且可提前算出牛顿减少量进行回溯直线搜索的简化

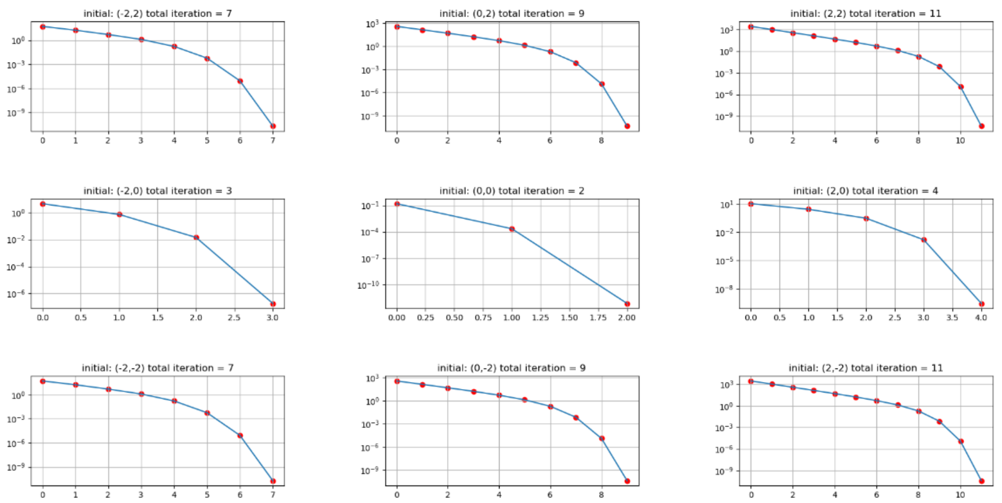
程序设定

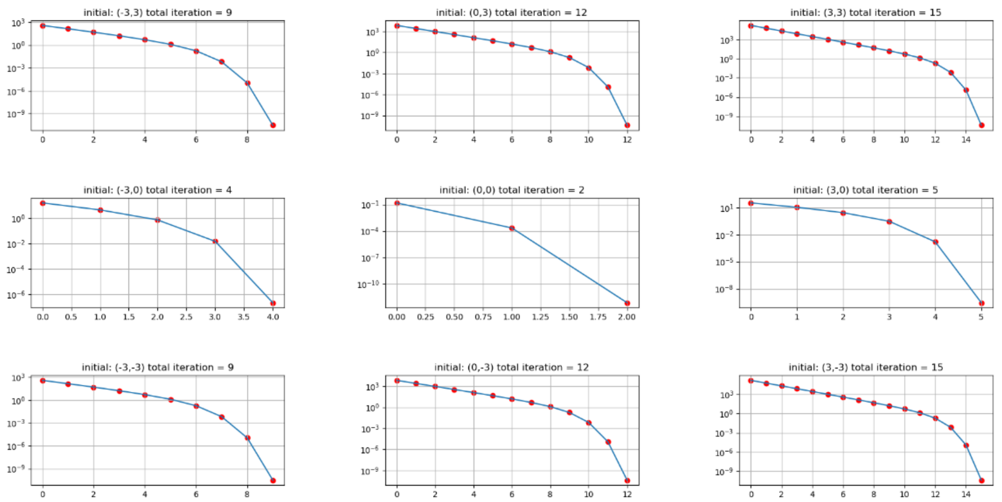
本实验一共计算了 18 个初始点的迭代过程，分别为(-2,2) (0,2) (2,2) (-2,0) (0,0) (2,0) (-2,-2) (0,-2) (2,-2)九个点和(-3,3) (0,3) (3,3) (-3,0) (0,0) (3,0) (-3,-3) (0,-3) (3,-3)九个点。ε=1e-6，即当牛顿减少量的 1/2 小于 1e-6 时终止迭代。

程序实现

见 3_newton_exp.py

实验结果





实验结论

可见迭代过程相比之前的梯度下降更加快速，平均在个位数次数的迭代就可以达到很高的精度。