回溯直线搜索不同参数对迭代次数的影响

一、问题背景

考虑以下二次规划问题:

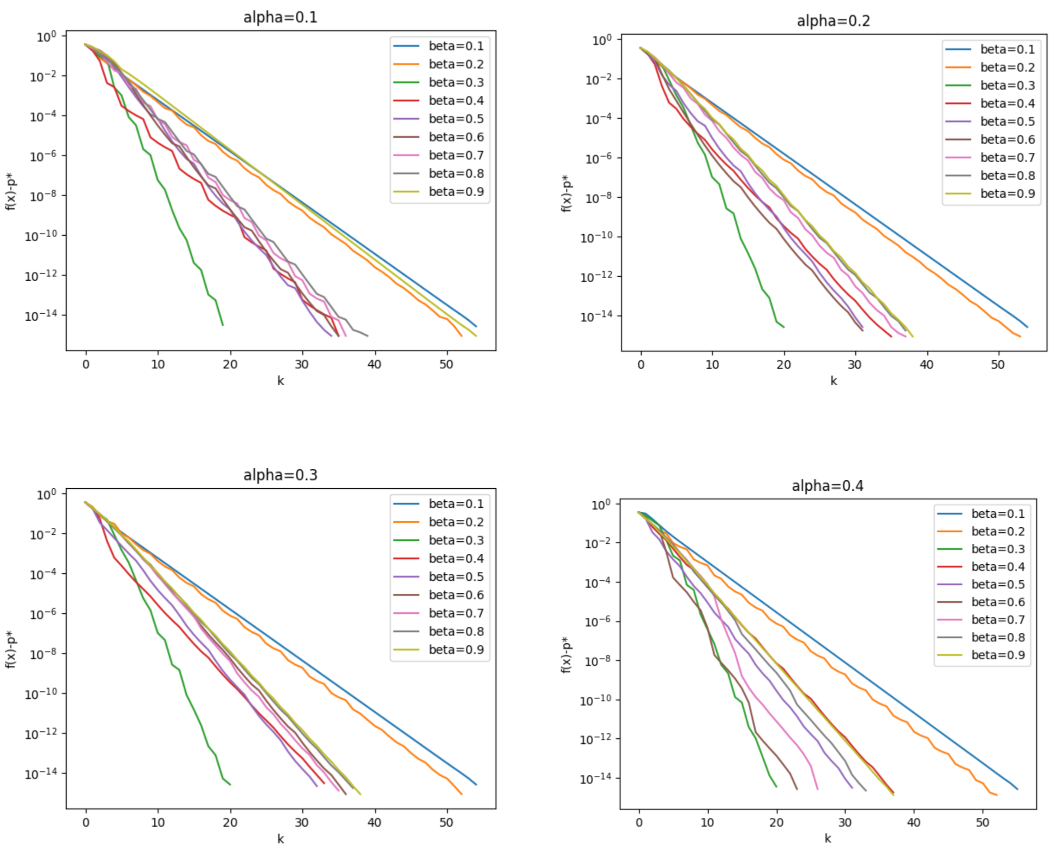
$$minf(x)=e^{x_1+3x_2-0.1}+e^{x_1-3x_2-0.1}+e^{-x_1-0.1}$$

分析可知，问题的最优解$x^*=\left(-\frac{ln2}{2},0\right)$，最优值$p^*=f(x^*)$，考虑使用**回溯直线搜索的梯度下降算法**对此问题进行求解，本实验设置迭代的停止条件为$\left\|\nabla f(x^k)\right\|_2<\varepsilon$，其中设置$\varepsilon=1e-7$，迭代的初始点$x_0=(0.1,0,1)$。本实验讨论采用不同的$\alpha,\beta$的取值时，误差随迭代次数改变的情况，以及不同$\alpha,\beta$对需要的收敛次数（迭代到收敛的次数）的影响。

二、固定α，讨论β对收敛性能的影响

1.误差随迭代次数的改变

固定α取值（为 0.1/0.2/0.3/0.4），改变β的取值（从 0.1 到 0.9），绘制误差（$f(x^k)-p^*$）随迭代次数k的变化曲线，如下所示。

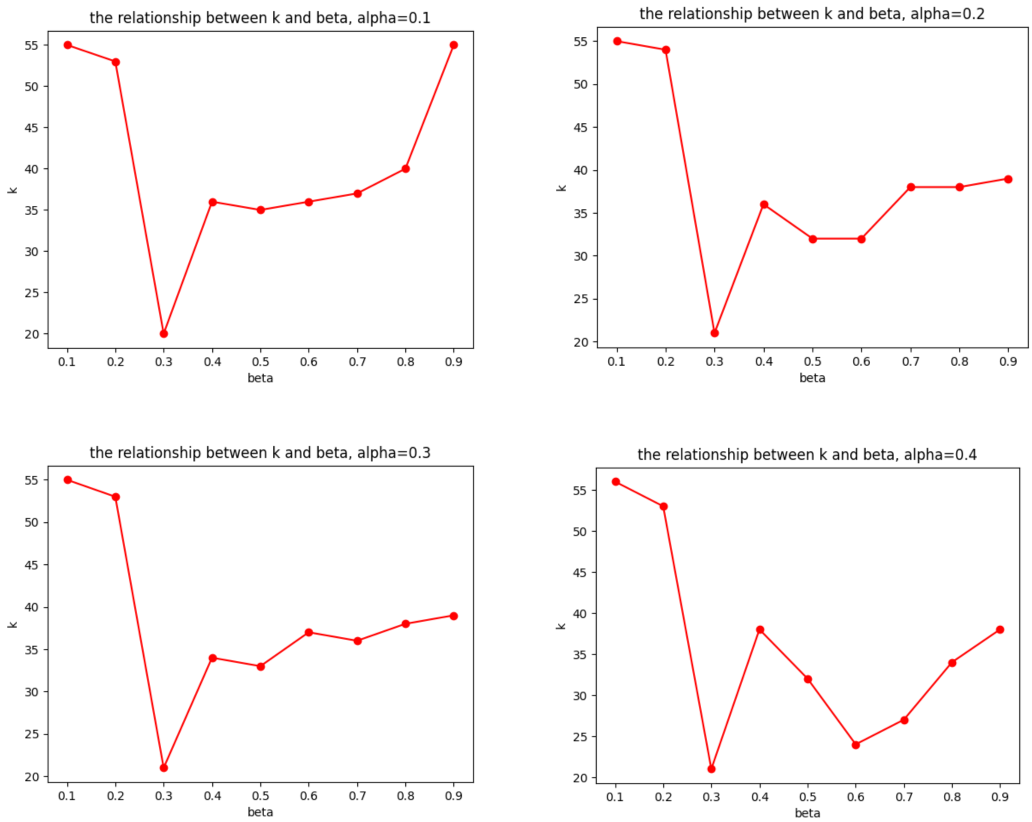


图一：固定α，不同β下的误差随迭代次数变化的曲线

由图可知，同一α下的不同β取值对收敛速度有一定的影响，从最小的需要 20 次左右的迭代到最大需要 50 次左右，且**β取值适中（比如，0.3）时收敛效果较好。**

2.需要的迭代次数随β的变化

固定α取值（为 0.1/0.2/0.3/0.4），改变β的取值（从 0.1 到 0.9），绘制迭代到收敛时需要的次数k随β的变化曲线，如下所示。

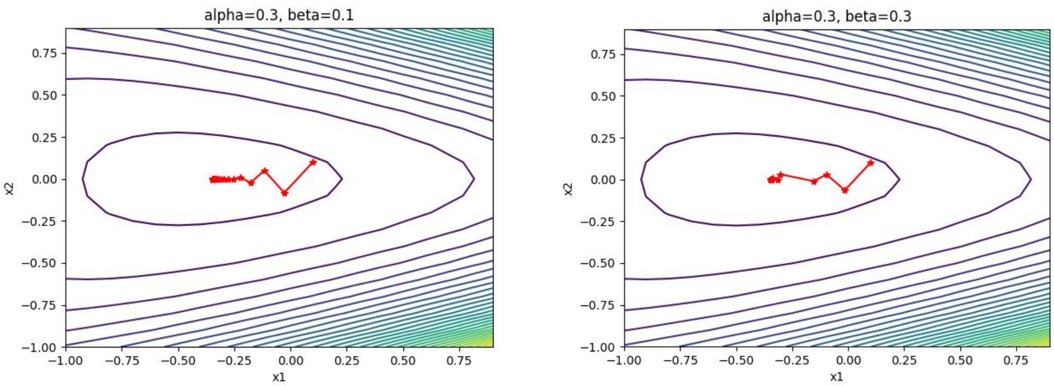


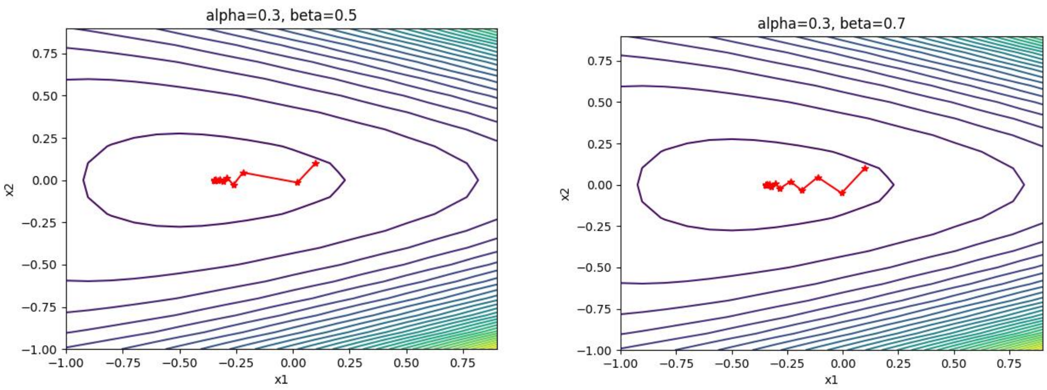
图二：固定α，不同β下所需的迭代次数曲线

从图中可以看出，在某一个α下，**需要的迭代次数随β的变大，有先减小后增大的趋势，**其中β取值在 0.3 左右时，收敛需要的次数较少，为 20 次左右，在 0.1 左右时收敛次数较大在 50 次左右。

3.收敛过程绘图

本节选取了一些典型的收敛过程进行绘图。选取$\alpha = 0.3$，β为 0.1/0.3/0.5/0.7 时的收敛过程进行图像绘制。



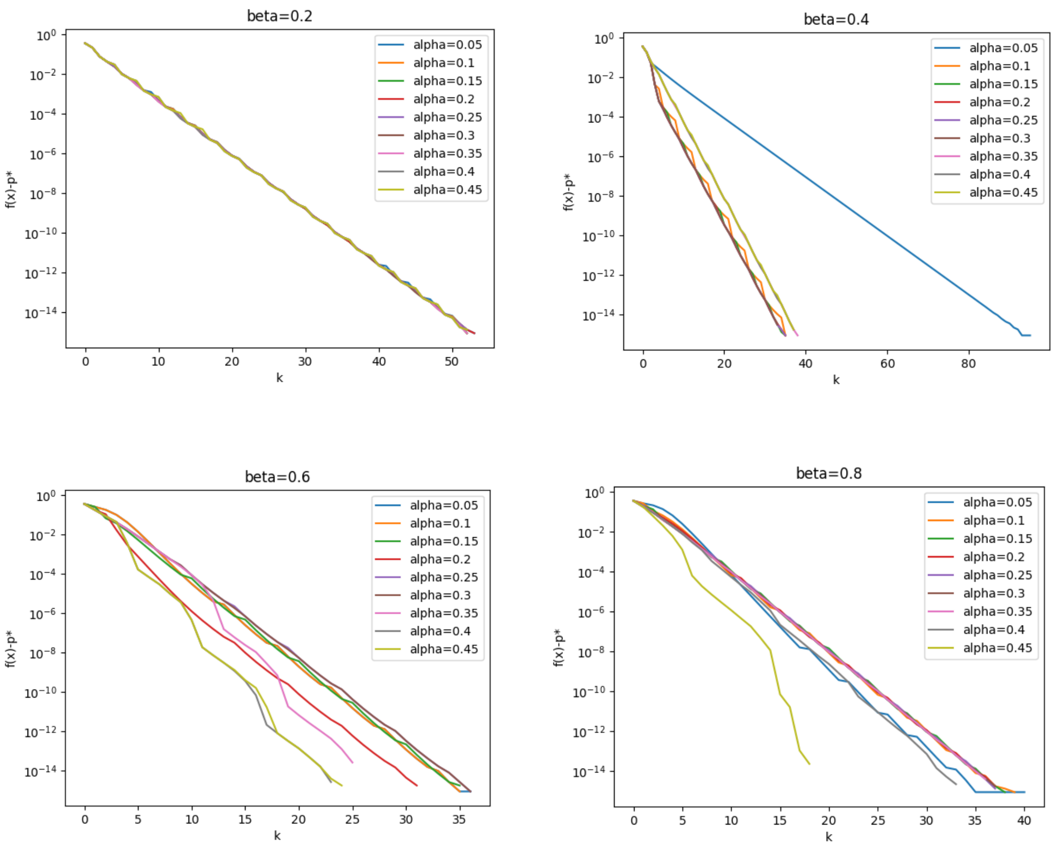


图三：固定α下的一些典型收敛过程

三、固定β，讨论α对收敛性能的影响

1.误差随迭代次数的改变

固定β取值（为 0.2/0.4/0.6/0.8），改变α的取值（从 0.05 到 0.45），绘制误差（$f(x^k)-p^*$）随迭代次数k的变化曲线，如下所示。

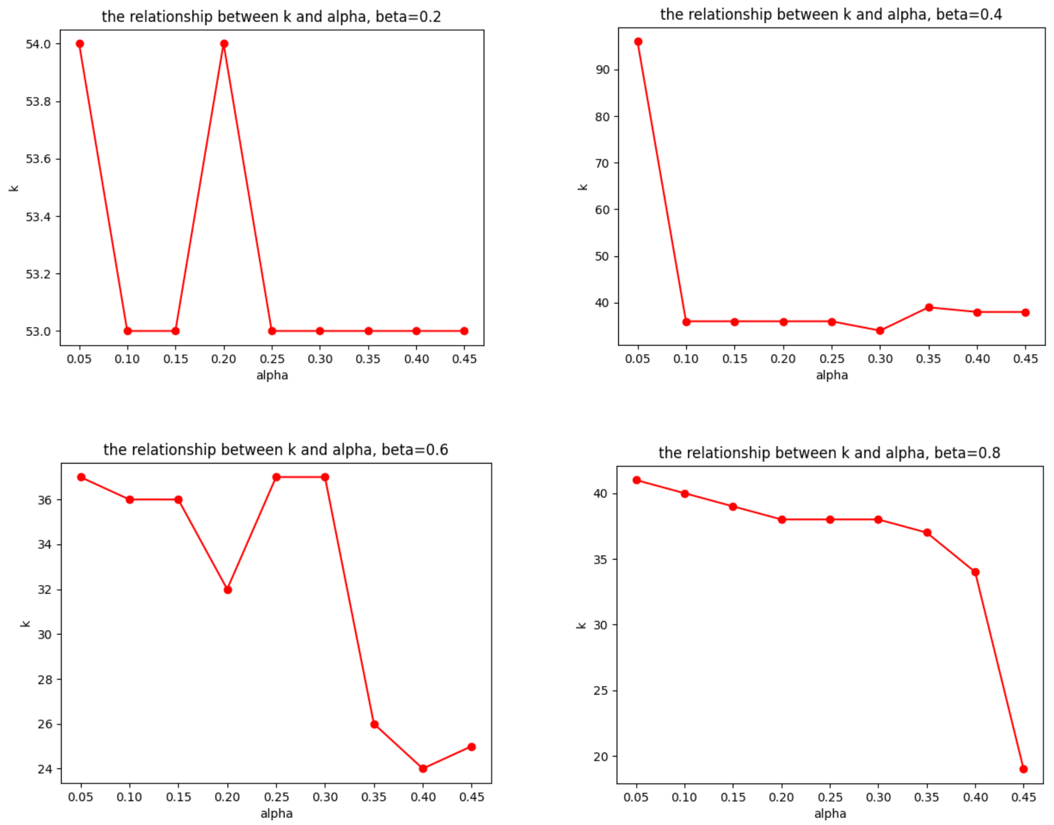


图四：固定β，不同α下的误差随迭代次数变化的曲线

从图中可以看出，在某些β下（比如，0.2），α对迭代次数没有较大的影响，一般而言，**选取较大的α（比如，0.3-0.5），有利于加快收敛的速度。**

2.需要的迭代次数随β的变化

固定β取值（为 0.2/0.4/0.6/0.8），改变α的取值（从 0.05 到 0.45），绘制迭代到收敛时需要的次数k随α的变化曲线，如下所示。

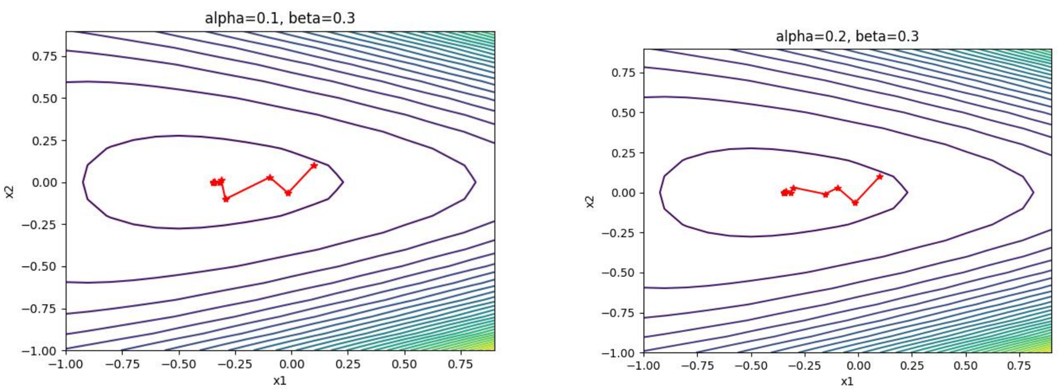


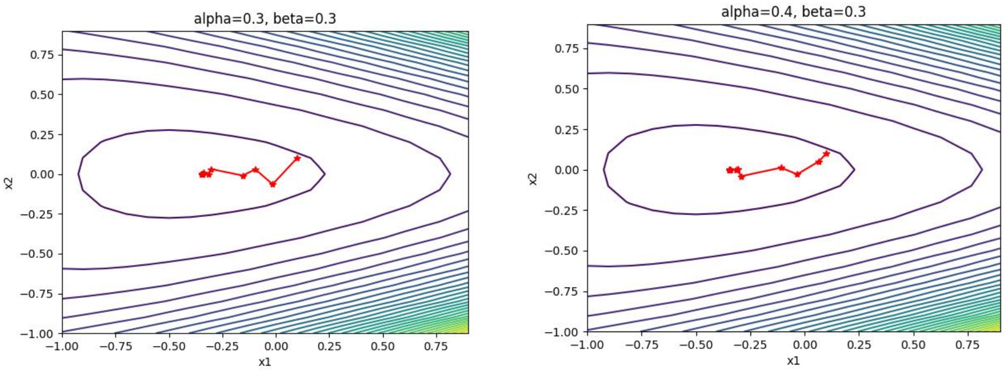
图五：固定β，不同α下所需的迭代次数曲线

从图中可以看出，**随着α的增大，需要的收敛次数有下降的趋势，**有的从需要 90 次以上到需要 40 次左右，有的从需要 40 次左右到需要 20 次左右，这和β的具体取值有关。**α取较大的值，有利于减少需要的迭代次数。**

3.收敛过程绘图

本节选取了一些典型的收敛过程进行绘图。选取$\beta = 0.3$，α为 0.1/0.2/0.3/0.4 时的收敛过程进行图像绘制。





图三：固定β下的一些典型收敛过程

四、总结

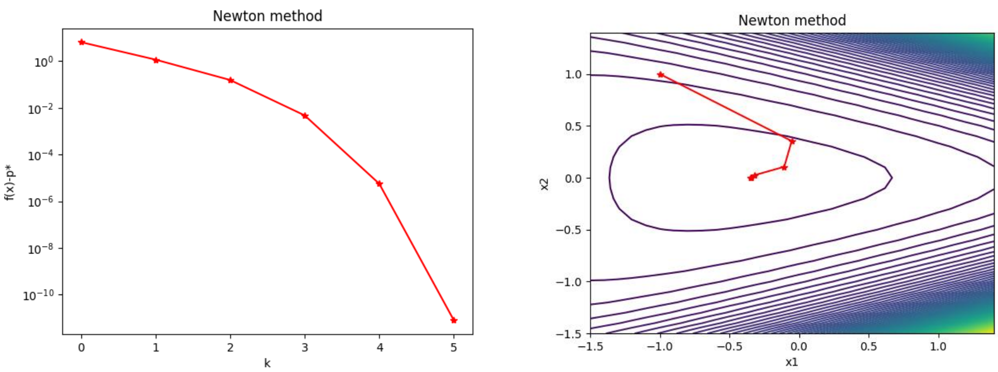
（1）α, β对回溯搜索的梯度下降算法性能有一定的影响，但是不会产生戏剧性的影响。

（2）α在取值较大的时候（比如 0.3-0.5），算法的性能较好。

（3）β在取值适中的时候（比如 0.3 左右），算法的性能较好。

五、牛顿法补充

对以上函数使用牛顿法进行最小值的求解，采用回溯直线搜索，设置$\alpha = 0.1, \beta = 0.7$，设置迭代的初始点$x_0 = (-1,1)$，设置迭代停止条件为$\lambda^2/2 \le \epsilon$，其中$\lambda^2 = \nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x)$，$\epsilon = 1e-8$，误差（$f(x^k) - p^*$）随迭代次数$k$的变化曲线和使用牛顿法迭代的收敛路径如下图所示。



可以看出，对于牛顿迭代法，只需要较少的次数（这里为 5 次左右），即可收敛到离最优值接近的点，作为对比，前几节讨论的使用回溯直线搜索的梯度下降方法，需要大于 10 次的迭代才能收敛到和牛顿法相同的精度。