最优化第一次作业

2115102001 安炳旭

2023 年 10 月 16 日

目录

1 证明

作业1的详细内容见压缩文件中的作业1.pdf

2 三种基础的梯度下降法

2.1 Constant Stepsize

Constant Stepsize的策略是:对于给定的函数$f(\mathbf{x})$而言

1.选定一个初始值$\mathbf{x_0} \in \mathbb{R}^n$.

2.计算$f(\mathbf{x_0})\nabla f_0$,用来表示梯度下降的方向.

3.设置一个参数t用来表示在该方向上移动的长度.

4.选定$\mathbf{x_1} = \mathbf{x_0} - t\nabla f_0$进行迭代,直到第$k$次时,有$||\nabla f_{k-1}|| \leq \varepsilon$,$\varepsilon$同样是自己设置的参数

对于给定的第一个函数函数$min\ x^2 + 2y^2$而言,在该实验中选取初始点$x_0 = y_0 = 100$,迭代所产生的结果如下所示:

1.当$\varepsilon = 10^{-5}, t = 0.05$时,迭代在第162次结束,此时$x = y = 3.48 \times 10^{-6}$

2.当$\varepsilon = 10^{-5}, t = 0.01$时,迭代在第76次结束,此时$x = y = 3.45 \times 10^{-6}$

3.当$\varepsilon = 10^{-5}, t = 10$时,由于t值过大,迭代无法收敛

而对于给定的第二个函数$min\ x^2 + 100y^2$而言,当$\varepsilon = 10^{-5}, t = 0.05$时,由于$t$值过大,迭代无法收敛.因此$t = 0.1, 10$时,函数也必然不会收敛

此时为了让第二个函数收敛，我们选取$t = 0.005$时,迭代在第1672次结束,此时$x = 9.97 \times 10^{-6}, y = 0.0$

将第一个函数的x,y与梯度值进行可视化如图1与图2所示:

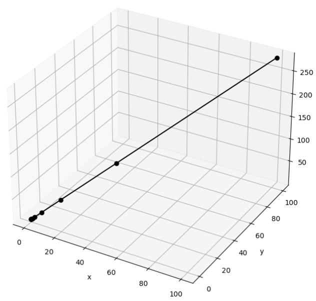

图 1: $t = 0.05, x^2 + y^2$

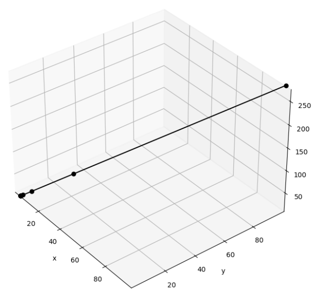

图 2: $t = 0.1, x^2 + y^2$

将第二个函数的x,y与梯度值进行可视化如图3所示:

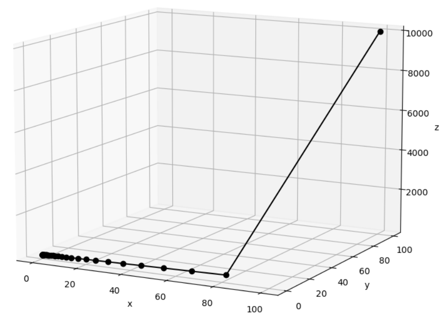

图 3: $t = 0.05, x^2 + 100y^2$

对于这种方法而言，由于调整步长是手动决定的，因此可能会带来些许不便。若对两个系数不同的变量共享同一个学习率，也会出现其中一个变量收敛较快，而另一个收敛较慢的问题。

2.2 Exact Line Search

Exact Line Search的策略是:对于给定的函数而言,寻求最优解时进行梯度下降的方向和步长值都是固定的,无需人为进行干预,保证在每一次迭代时所选取的下降方向均是最优的.

也就是对于

$$min\ f(\mathbf{x})$$

等价于

$$\min_{\mathbf{x}\in\mathbb{R}^n}\left\{\mathbf{x^TAx+2b^Tx}\right\}$$

其中$\mathbf{A}$是正定方阵,在迭代第k步时,自动选取的步长为$t_k = \frac{\|\nabla f(\mathbf{x}_k)\|^2}{2\nabla f(\mathbf{x}_k)^T\mathbf{A}\nabla f(\mathbf{x}_k)}$

因此,对于给定的两个函数$min\ x^2+y^2$与$min\ x^2+100y^2$而言,分别使用Exact Line Search，得到的结果如图4与图5所示

从图中可以观察到，每一步所走的路径方向与前一步都是垂直的，并且这种方法每迭代一次就会计算一次最优路径，会产生更多的计算开销。

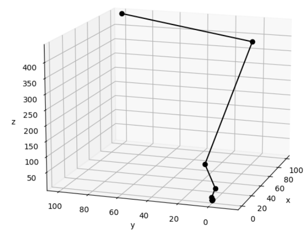

图 4: $t = 0.05, x^2 + y^2$

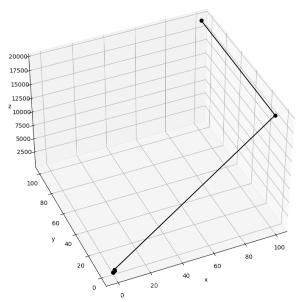

图 5: $t = 0.1, x^2 + 100y^2$

2.3 Backracking

Backtracking(回溯)是一种优化算法中的搜索策略，通常用于确定合适的步长或学习率以便在梯度下降等迭代过程中进行参数更新。其主要目标是在每次迭代中找到适当的步长，以确保算法收敛到目标函数的局部最小值。通常可以用以下公式进行描述:

$$f(\mathbf{x}_k) - f(\mathbf{x}_k + t_k\mathbf{d}_k) \geq -\alpha t_k \nabla f(\mathbf{x}_k)^T \mathbf{d}_k.$$

其中，$s > 0, \alpha \in (0, 1), \beta \in (0, 1)$. 初始值$t_0 = s$

之后的每次迭代都有$t_k := \beta t_{k-1}$，其中$\alpha, \beta\, s$都是需要手动调整的参数

对于给定的第一个函数$min\ x^2+2y^2$而言,当$\varepsilon=10^{-5},s=1,\alpha=0.5,\beta=0.5$时,初始点$x_0,y_0$选为$(100,100)$,只需要两步迭代即可收敛

对于给定的第二个函数$min\ x^2+100y^2$而言,当$\varepsilon=10^{-5},s=1,\alpha=0.5,\beta=0.5$时,初始点$x_0,y_0$选为$(100,100)$，最终收敛需要620步.

将第一个函数的x,y与梯度值进行可视化如图5与图6所示:

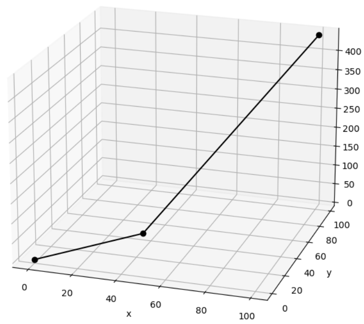

图 6: $t=0.05,x^2+y^2$

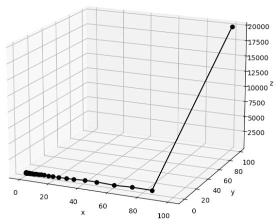

图 7: $t=0.1,x^2+100y^2$

该方法是目前梯度下降优化算法中较为常用的一种.

三个方法的代码文件在此不一一列出，详细代码见压缩文件.

3 信号去噪