深蓝学院
shenlanxueyuan.com

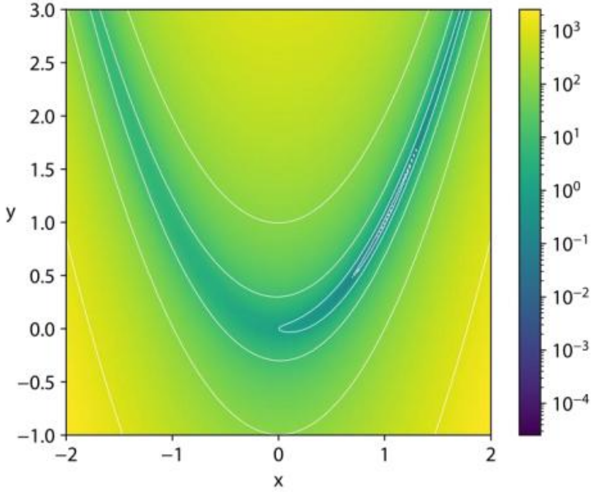
第一章作业思路

主讲人 ******

问题描述

深蓝学院
shenlanxueyuan.com

Implement the line-search steepest gradient descent with Armijo condition.



$$f(\mathbf{x}) = f(x_1, x_2, \ldots, x_N) = \sum_{i=1}^{N/2} \left[100\left(x_{2i-1}^2 - x_{2i}\right)^2 + \left(x_{2i-1} - 1\right)^2 \right]$$

Your program should properly minimize this Rosenbrock function.

思路

在优化问题中，我们常用的架构无非是：

- 搜索下降方向
- 搜索确定下降步长
- 迭代确定新位置

Choose search direction: $d = -\nabla f(x^k)$

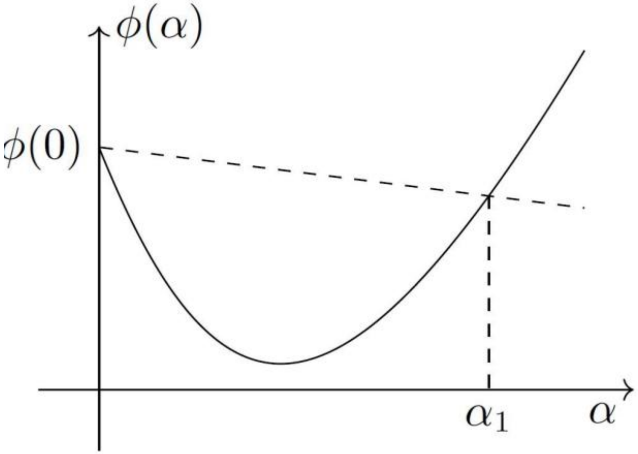
Backtracking/Armijo line search

$$x^{k+1} = x^k + \tau d$$

Armijo condition

最速下降法Steepest Gradient Descend采用Armijo Condition作为判断

Backtracking/Armijo line search



Choose search direction: $d = -\nabla f(x^k)$

While $f(x^k + \tau d) > f(x^k) + c \cdot \tau d^T \nabla f(x^k)$

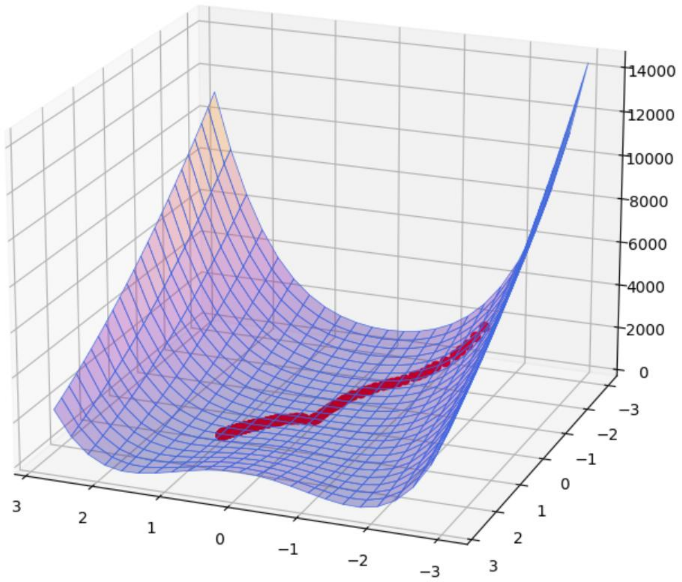
$\tau \leftarrow \tau/2$

Update iterate $x^{k+1} = x^k + \tau d$

Repeat this until gradient is small or subdifferential contains zero.

结果

optimization of Rosenbrock function



二维Rosenbrock function优化结果

补充：可行牛顿法

深蓝学院 shenlanxueyuan.com

可行牛顿法Practical Newton’s Method

首先初始化x, $x \leftarrow x_0 \in \mathbb{R}^n$

当$||\nabla f(x)|| > \delta$时，进行如下计算

$do:$

$d \leftarrow -M^{-1}\nabla f(x)$

$t \leftarrow backtrackiong \quad line \quad search$

$x \leftarrow x + td$

$end \quad while$

$return$

其中，M是一个接近Hessian阵的正定矩阵，以此来替代线性搜索中的求梯度和求Hessian阵。

如果函数为凸函数，则有

避免hessian的逆不存在

$$M = \nabla^2 f(x) + \epsilon I, \epsilon = \min\left(1, \|\nabla f(x)\|_\infty\right)/10$$

因为M是正定的，因此可以使用Cholesky factorization

$$Md = -\nabla f(x), M = LL^{\mathrm{T}}$$

如果函数是非凸的，那么我们通过如下计算M

Bunch-Kaufman Factorization:

$$Md = -\nabla f(x), M = LBL^{\mathrm{T}}$$

核心思想：对函数进行二阶泰勒展开近似，每一次不按照梯度走，二是直接走到近似后的函数的最优位置

$$f(\boldsymbol{x}) \approx \hat{f}(\boldsymbol{x}) \triangleq f(\boldsymbol{x}_k) + \nabla f(\boldsymbol{x}_k)^T(\boldsymbol{x}-\boldsymbol{x}_k) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_k)^T\nabla^2 f(\boldsymbol{x}_k)(\boldsymbol{x}-\boldsymbol{x}_k)$$

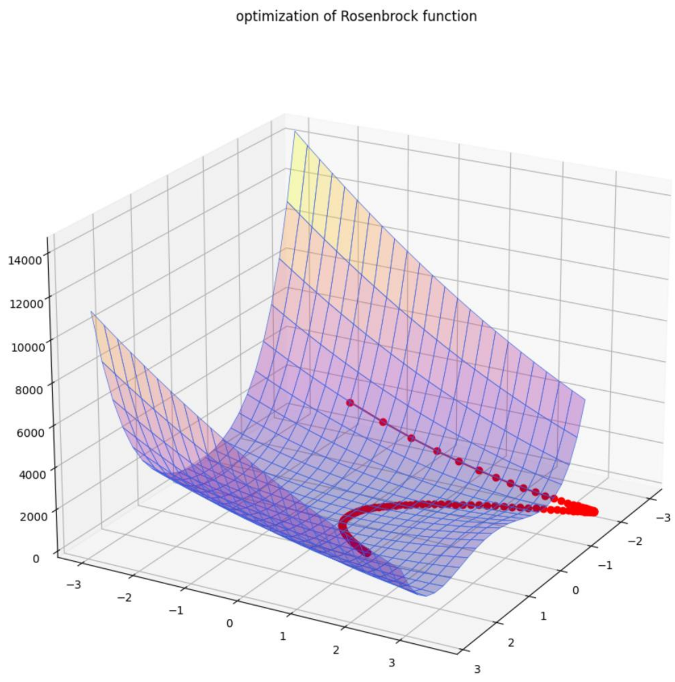
Minimizing quadratic approximation

$$\nabla \hat{f}(\boldsymbol{x}) = \nabla^2 f(\boldsymbol{x}_k)(\boldsymbol{x}-\boldsymbol{x}_k) + \nabla f(\boldsymbol{x}_k) = \mathbf{0}$$

$$\Longrightarrow \boldsymbol{x} = \boldsymbol{x}_k - \left[\nabla^2 f(\boldsymbol{x}_k)\right]^{-1}\nabla f(\boldsymbol{x}_k)$$

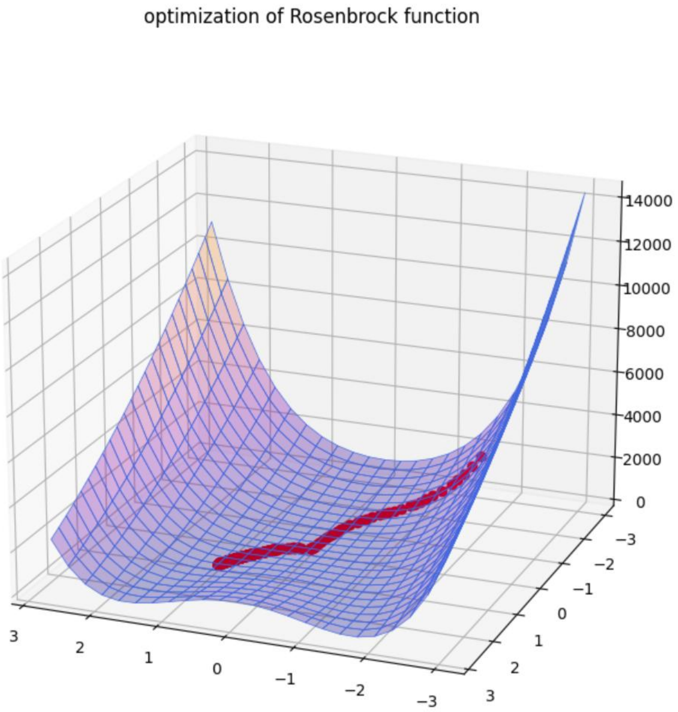
d 下降方向

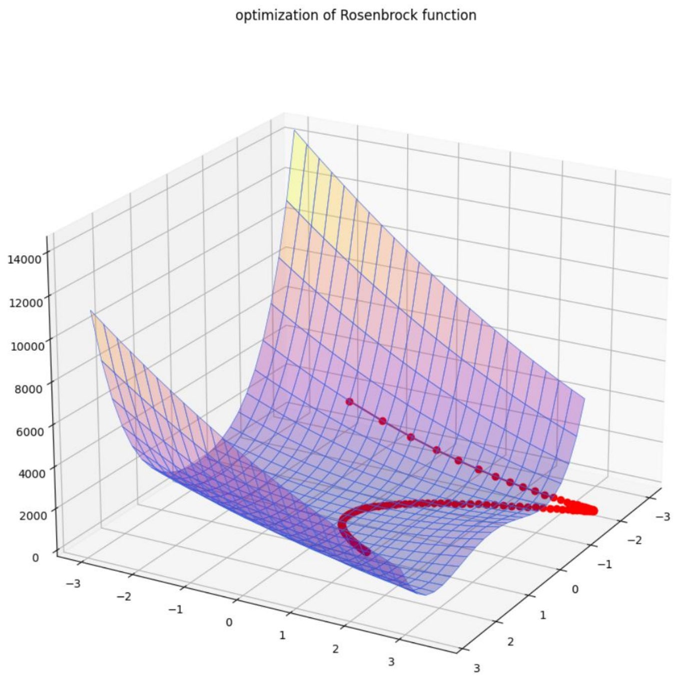
结果



二维Rosenbrock function优化结果

结果





我们可以发现PNM的迭代步长相比于SGD更大，因此每一步效率更高。而SGD严格按照梯度下降。PNM没有按照梯度下降的原因是PNM在迭代点处进行了二阶泰勒展开，这是一把双刃剑，虽然可以直接求出最优的迭代方向，但是又以后因为二阶展开产生了误差

深蓝学院
shenlanxueyuan.com

感谢各位聆听！

Thanks for Listening