梯度下降法

作者：周铁军

2019年7月1日

0.1 描述

梯度下降法基于以下的观察：如果实值函数 $F(x)$ 在点 a 处可微且有定义，那么函数 $F(x)$ 在 a 点沿着梯度相反的方向 $-\nabla F(a)$ 下降最快。

因而，如果

$$b = a - \gamma \nabla F(a) \tag{1}$$

对于 $\gamma > 0$ 为一个够小数值时成立，那么 $F(a) \geq F(b)$。

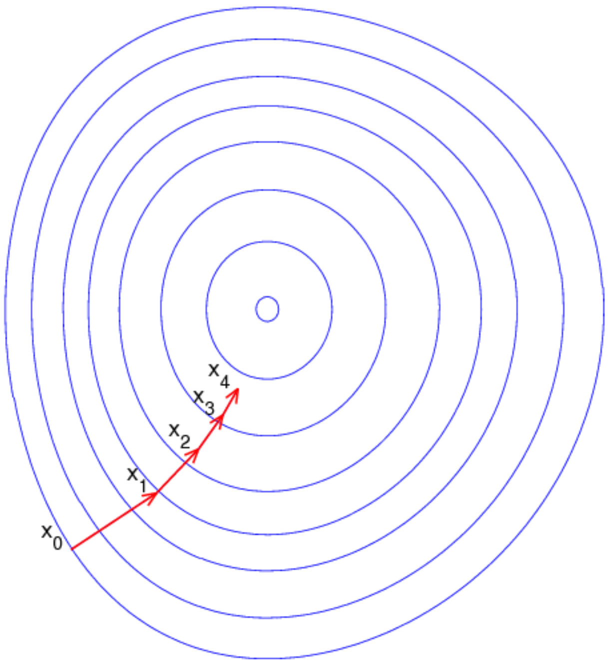
考虑到这一点，我们可以从函数 F 的局部极小值的初始估计 x_0 出发，并考虑如下序列 $x_0, x_1, x_2, \ldots$ 使得

$$x_{n+1} = x_n - \gamma_n \nabla F(x_n), n \geq 0 \tag{2}$$

因此可得到

$$F(x_0) \geq F(x_1) \geq F(x_2) \geq ..., \tag{3}$$

如果顺利的话序列 (x_n) 收敛到期望的极值。注意每次迭代步长 γ 可以改变。



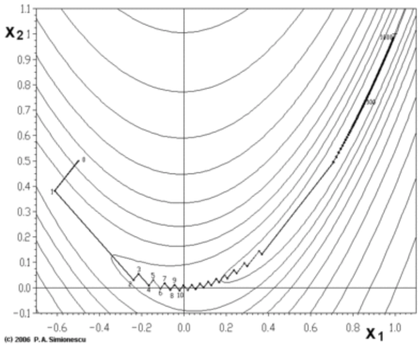
上图示例了这一过程，这里假设 F 定义在平面上，并且函数图像是一个碗形。蓝色的曲线是等高线（水平集），即函数 F 为常数的集合构成的曲线。红色的箭头指向该点梯度的反方向。（一点处的梯度方向与通过该点的等高线垂直）。沿着梯度下降方向，将最终到达碗底，即函数 F 值最小的点。

0.2 例子

梯度下降法处理一些复杂的非线性函数会出现问题，例如 Rosenbrock 函数

$$f(x, y) = (1 - x)^2 + 100(y - x^2)^2 \tag{4}$$

其最小值在（x,y)=(1,1) 处，数值为 f(x,y)=0. 但是此函数具有狭窄弯的山谷，最小值 (x,y)=(1,1) 就在这些山谷之中，并且谷底很平。优化过程是之字形的向极小值点靠近，速度非常缓慢。



下面这个例子也鲜明的示例了”之字”的上升（非下降），这个例子用梯度上升（非梯度下降）法求 $F(x,y) = \sin\left(\frac{1}{2}x^2 - \frac{1}{4}y^2 + 3\right)\cos(2x + 1 - e^y)$ 的极大值（非极小值，实际是局部极大值）。

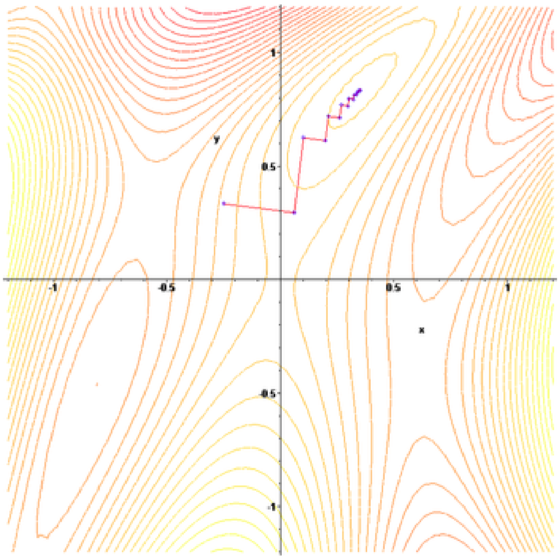


图 1: this is a figure4.

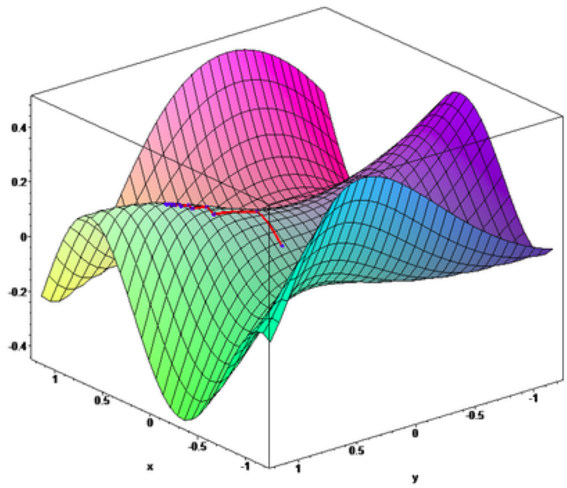


图 2: this is a figure5.

[1]

参考文献

[1] TKH Tam and Cecil G Armstrong. 2d finite element mesh generation by medial axis subdivision. *Advances in engineering software and workstations*, 13(5):313–324, 1991.