习题课

张帆

2020 年 12 月 3 日

Outline

- 非线性规划和求根问题
  - 牛顿法
  - 割线法
- 全局收敛方法
  - 线搜索
  - 信赖域
- 共轭梯度法 (二次规划子问题)

非线性规划

考虑如下无约束非线性规划问题

$$\min_{x \in \mathbb{R}^n} \quad f(x) \tag{1}$$

其中函数 f 是一阶且二阶连续可微的。

全局解和局部解

Definition

全局最小值点：若一点 $\boldsymbol{x}^*$，对于任意 $\boldsymbol{x}^* \in \mathbb{R}^n$ 都有 $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$，则称 $\boldsymbol{x}^*$ 为该问题的全局最小值点。

Definition

局部最小值点：若有一点 $\boldsymbol{x}^*$，存在 $\epsilon > 0$，对于任意

$$\boldsymbol{x} \in \mathbb{B}(\boldsymbol{x}^*, \epsilon) := \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x} - \boldsymbol{x}^*\|_2 \leq \epsilon\}$$

都有 $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$，则称 $\boldsymbol{x}^*$ 为该问题的局部最小值点。

最优性条件

Theorem (一阶必要性条件)

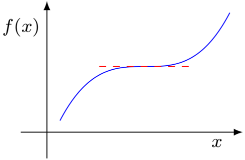
如果 $\boldsymbol{x}^*$ 是一个局部极小点，那么 $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$。

Theorem (二阶必要性条件)

如果 $\boldsymbol{x}^*$ 是一个局部极小点，那么 $\nabla^2 f(\boldsymbol{x}^*) \succeq \mathbf{0}$。

Theorem (二阶充分性条件)

如果 $\nabla f(\boldsymbol{x}^*) = 0$ 且 $\nabla^2 f(\boldsymbol{x}^*) \succ \mathbf{0}$，那么 $\boldsymbol{x}^*$ 是一个局部极小点。



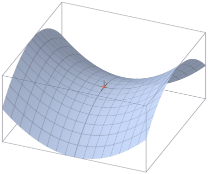


图: (a) $f(x) = 1 + (x-4)^3$ (b) $f(x) = x_1^4 - x_2^4$

$$\nabla f(x)|_{x=4} = 3(x-4)^2|_{x=4} = 0 \qquad \nabla f(\boldsymbol{x})|_{\boldsymbol{x}=\boldsymbol{0}} = \left.\begin{bmatrix} 4x_1^3 \\ -4x_2^3 \end{bmatrix}\right|_{\boldsymbol{x}=\boldsymbol{0}} = \boldsymbol{0}$$

$$\nabla^2 f(x)|_{x=4} = 6(x-4)|_{x=4} = 0 \qquad \nabla^2 f(\boldsymbol{x})|_{\boldsymbol{x}=\boldsymbol{0}} = \left.\begin{bmatrix} 12x_1^2 & 0 \\ 0 & -12x_2^2 \end{bmatrix}\right|_{\boldsymbol{x}=\boldsymbol{0}} = \boldsymbol{0}$$

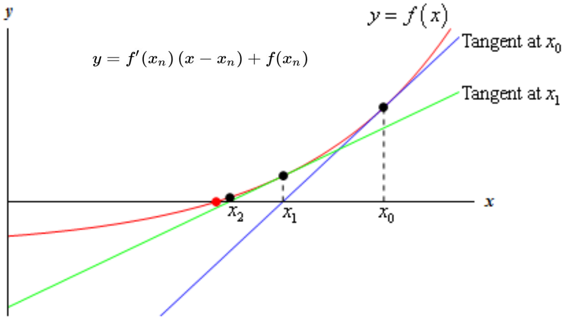
等式系统

求解非线性规划的算法一般都是基于一阶必要性条件的，即找到 $F(\boldsymbol{x}) := \nabla f(\boldsymbol{x}) = \mathbf{0}$ 的根。而对于一般的等式方程 (组):

1. 很可能没有显式解
2. 不知道是否有解，有几个解

$\Longrightarrow$ 迭代类算法求根

单变量牛顿法



多变量牛顿法

对于函数 $F(\boldsymbol{x}):\mathbb{R}^n \to \mathbb{R}^n$，其在距离 $\boldsymbol{x}$ 充分近一点 $\boldsymbol{x}^k$ 处的一阶 Taylor 逼近为:

$$F(\boldsymbol{x}) \approx F(\boldsymbol{x}^k) + \nabla F(\boldsymbol{x}^k)^T(\boldsymbol{x} - \boldsymbol{x}^k).$$

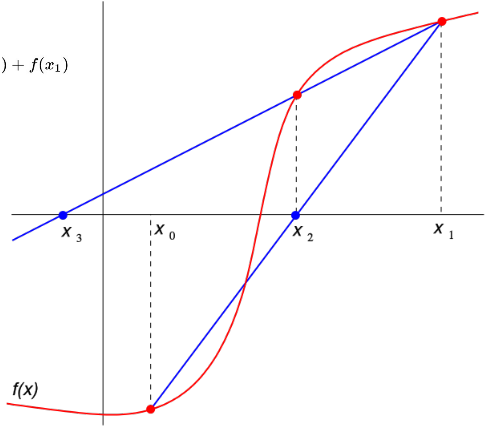
当 $F(\boldsymbol{x}) = \mathbf{0}$ 时，令 $\boldsymbol{d}^k = \boldsymbol{x} - \boldsymbol{x}^k$，则

$$\nabla F(\boldsymbol{x}^k)^T \boldsymbol{d}^k = -F(\boldsymbol{x}^k).$$

更新 $\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \boldsymbol{d}^k$。

割线法

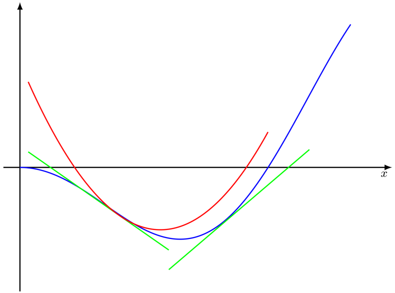
$$y = \frac{f(x_1) - f(x_0)}{x_1 - x_0}(x - x_1) + f(x_1)$$



割线法

Example: Quadratic model at x_{k+1}

We create a quadratic model that yields the same gradient value at the last iterate:



拟牛顿法

Suppose we choose $m_{k+1}(d)$ to satisfy the following conditions:

$$\begin{aligned} m_{k+1}(0) &= f(x_{k+1}); \\ \nabla m_{k+1}(0) &= \nabla f(x_{k+1}); \\ \nabla m_{k+1}(-\alpha_k d_k) &= \nabla f(x_k). \end{aligned}$$

That is,

- function value should match at x_{k+1};
- gradient values should match at x_{k+1} and x_k.

(Note that with respect to x_{k+1}, we get to x_k with the step $-\alpha_k d_k$.)

- The last is the only condition that depends on H_{k+1}. Rearranging

$$\nabla f(x_k) = \nabla m_{k+1}(-\alpha_k d_k) = \nabla f(x_{k+1}) + H_{k+1}(-\alpha_k d_k),$$

we obtain

$$H_{k+1} s_k = y_k \quad \text{(the “secant equation”)},$$

where

$$s_k = x_{k+1} - x_k = \alpha_k d_k \text{ and } y_k = \nabla f(x_{k+1}) - \nabla f(x_k).$$

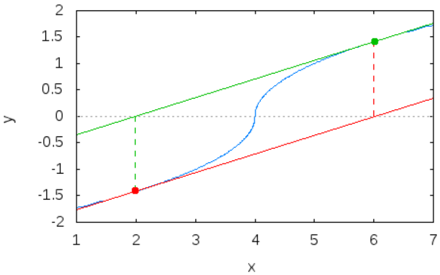
局部二次收敛

Theorem

假设 x^* 为满足二阶充分性条件的一个局部极小值点。$\nabla F(x)$ 在 x^* 的邻域内是 *Lipschitz* 连续的。利用牛顿法求解该问题，若我们选取的初始点 x^0 与 x^* 充分接近，则:

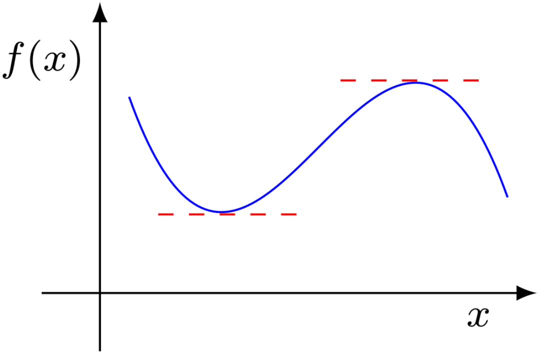
1. 序列 $\{x^k\}$ 二次收敛到局部极小值点 x^*;
2. $\{\|F(x^k)\|\}$ 以平方收敛速率收敛到 $\mathbf{0}$。

牛顿法失效



$$f(x) = \begin{cases} \sqrt{x-4}, & \text{for } x \geq 4 \\ -\sqrt{4-x}, & \text{for } x < 4 \end{cases}$$

可能找到极大值点



线搜索方法

出发点：至少保证新迭代点单调不增

步骤：

1. 选取**下降**方向
2. (根据某种线搜索技巧) 确定步长

线搜索条件:

1. **Sufficient descent (Armijo condition)**
2. Curvature condition

收敛性证明大体步骤:

1. 找到步长下界
2. 用 Armijo 条件得到相邻两步目标函数值下降量 (与梯度范数相关的) 下界
3. 累和证明梯度范数极限 $\to \mathbf{0}$

出发点: 原问题的局部近似二次逼近

子问题求解:

1. 精确求解
2. 非精确:Cauchy point method　dogleg method

信赖域半径更新准则 (比值)

Consider the ratio (with positive denominator):

$$\rho_k(d_k) := \frac{f(x_k) - f(x_k + d_k)}{m_k(0) - m_k(d_k)}.$$

特定问题 (本课程所重点要求): 无约束凸二次规划

出发点: "坐标下降法"

重点掌握:

1. 更新步推导
2. 每次迭代的计算复杂度