Optimal Theory and Method

最优化理论与方法

翟春杰

杭州电子科技大学 自动化学院

科技馆512

Email: cjzhai@hdu.edu.cn

回顾

■ 无约束优化问题

$$\min_{x\in R^n} f(x)$$

■ 迭代算法的基本思想

- 给一个迭代的**初始点**$x^{(0)}$，
- 根据**迭代公式**$x^{(k+1)}=x^{(k)}+\lambda_k d_k$进行迭代，
- 迭代截止条件

若$f(x)$导数存在，截止条件一般为：$\left\|\nabla f(x^{(k)})\right\|\le\varepsilon$

其它情况，截止条件可为$\left\|x^{(k+1)}-x^{(k)}\right\|\le\varepsilon$

回顾

■ 下降方向

$$\nabla f(x^{(k)})^T d^{(k)} < 0$$

$$x^{(k+1)} = x^{(k)} + \lambda_k d_k$$

■ 最优性的必要条件

点$x^{(k)}$为目标函数$f(x)$局部最优点的必要条件：

点$x^{(k)}$处不存在可行下降方向

目录

- 最速下降法（考）
- 牛顿法（考）
- 共轭梯度法（考）
- 拟牛顿法
- 信赖域法
- 最小二乘法

目录

- 最速下降法（考）
- 牛顿法（考）
- 共轭梯度法（考）
- 拟牛顿法
- 信赖域法
- 最小二乘法

■ 最速下降法

- 最速下降法面向哪类优化问题
- 什么是最速下降方向
- 最速下降法的步长
- 最速下降法的步骤

■ 最速下降法

➢适用于的优化问题

无约束优化问题：

$$\min \quad f(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathbf{R}^n,$$

其中，$f(x)$具有一阶连续偏导数

适用的优化问题：

- **无约束**
- **目标函数一阶连续可导**

■ 最速下降法

➢最速下降方向

1874年法国著名数学家Cauchy提出：负梯度方向就是最速下降方向　　Why?

$$f(x^{(k)}+\lambda_k d_k)-f(x^{(k)})=\lambda_k \nabla f(x^{(k)})^T d^{(k)}+\|\lambda_k d_k\|\alpha(x^{(k)},\lambda_k d_k)<0$$

$$\nabla f(x^{(k)})^T d^{(k)}=\|\nabla f(x^{(k)})\|\|d^{(k)}\|\cos\theta$$

当$\theta=\pi$，$\nabla f(x^{(k)})^T d^{(k)}=-\|\nabla f(x^{(k)})\|\|d^{(k)}\|$

此时$\nabla f(x^{(k)})$与$d^{(k)}$方向相反

$$d^{(k)}=k\cdot(-\nabla f(x^{(k)}))$$

通常情况下$k=1$，$d^{(k)}=-\nabla f(x^{(k)})$

■ 最速下降法

➢最速下降法的最优步长

迭代算法的迭代公式：

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \lambda_k \boldsymbol{d}^{(k)}$$

最速下降法的迭代公式：

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \lambda_k \nabla f(\boldsymbol{x}^{(k)})$$

如何确定最优步长λ_k：

$$\lambda_k = \min_{\lambda>0} f(\boldsymbol{x}^{(k)} - \lambda \nabla f(\boldsymbol{x}^{(k)}))$$

■ 最速下降法

➢最速下降法的步骤

(1) 给定初点$x^{(1)} \in R^n$，允许误差$\varepsilon > 0$，置$k = 1$;

(2) 计算搜索方向$d^{(k)} = -\nabla f(x^{(k)})$;

(3) 若$\|d^{(k)}\| \le \varepsilon$,停止计算；否则，从$x^{(k)}$出发，沿$d^{(k)}$进行一维搜索，确定最优$\lambda_k$，使得

$$f(x^{(k)} + \lambda_k d_k) = \min_{\lambda>0} f(x^{(k)} + \lambda d^{(k)})$$

(4) 令$x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$,置$k := k + 1$,转到步骤(2).

■ 最速下降法

➢最速下降法

例题

$$\min \ f(x_1, x_2, x_3) = (x_1)^2 + x_1(1 - x_2) + (x_2)^2 - x_2x_3 + (x_3)^2 + x_3$$

初始值 $x^{(1)} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ 允许误差 $\varepsilon = 0.5$

■ 最速下降法

➢最速下降法

解：

$$\nabla f(x)=\begin{bmatrix}2x_1+(1-x_2)\\-x_1+2x_2-x_3\\-x_2+2x_3+1\end{bmatrix}$$

$$d^{(1)}=-\nabla f(x^{(1)})=\begin{bmatrix}-1\\0\\-1\end{bmatrix}$$

满足截止条件?

$$\|d^{(1)}\|=\sqrt{(-1)^2+(0)^2+(-1)^2}=\sqrt{2}>\varepsilon$$

$k=1$，进行第一次迭代

$$x^{(2)}=x^{(1)}+\lambda d^{(1)}=\begin{bmatrix}-\lambda\\0\\-\lambda\end{bmatrix}$$

确定最优λ_1

■ 最速下降法

➢最速下降法

$$f(x^{(1)}+\lambda d^{(1)})=2\lambda^2-2\lambda$$

$$\frac{df(x^{(1)}+\lambda d^{(1)})}{d\lambda}=4\lambda-2=0$$

$$\lambda_1=\frac{1}{2}$$

$$x^{(2)}=x^{(1)}+\lambda_1 d^{(1)}=\begin{bmatrix}-\frac{1}{2}\\0\\-\frac{1}{2}\end{bmatrix}$$

$$d^{(2)}=-\nabla f(x^{(2)})=\begin{bmatrix}0\\-1\\0\end{bmatrix}$$

■ 最速下降法

➢最速下降法

满足截止条件?

$$\|d^{(2)}\| = \sqrt{(0)^2+(-1)^2+(0)^2} = 1 > \varepsilon$$

$k=2$，进行第2次迭代

$$x^{(3)} = x^{(2)} + \lambda d^{(2)} = \begin{bmatrix} -\frac{1}{2} \\ -\lambda \\ -\frac{1}{2} \end{bmatrix}$$

确定最优λ_2

$$f(x^{(2)} + \lambda d^{(2)}) = \lambda^2 - \lambda - \frac{1}{2}$$

$$\frac{df(x^{(2)} + \lambda d^{(2)})}{d\lambda} = 2\lambda - 1 = 0$$

$\lambda_2 = \frac{1}{2}$ $\quad x^{(3)} = x^{(2)} + \lambda_2 d^{(2)} = \left[-\frac{1}{2} \quad -\frac{1}{2} \quad -\frac{1}{2}\right]^T$

■ 最速下降法

➢最速下降法

$$d^{(3)} = -\nabla f(x^{(3)}) = \begin{bmatrix} -\frac{1}{2} \\ 0 \\ -\frac{1}{2} \end{bmatrix}$$

满足截止条件?

$$\|d^{(3)}\| = \sqrt{(-\frac{1}{2})^2+(0)^2+(-\frac{1}{2})^2} = \frac{\sqrt{2}}{2} > \varepsilon$$

$k = 3$，进行第3次迭代

$$x^{(4)} = x^{(3)} + \lambda d^{(3)} = \begin{bmatrix} -\frac{1}{2} \\ -\frac{1}{2} \\ -\frac{1}{2} \end{bmatrix} + \begin{bmatrix} -\frac{1}{2}\lambda \\ 0 \\ -\frac{1}{2}\lambda \end{bmatrix} = \begin{bmatrix} -\frac{1}{2}\lambda - \frac{1}{2} \\ -\frac{1}{2} \\ -\frac{1}{2}\lambda - \frac{1}{2} \end{bmatrix}$$

■ 最速下降法

➢最速下降法

确定最优λ_3

$$f(x^{(3)}+\lambda d^{(3)})=\frac{1}{2}(\lambda+1)^2-\frac{3}{2}(\lambda+1)+\frac{1}{4}$$

$$\lambda_3=\frac{1}{2}\qquad \frac{df(x^{(3)}+\lambda d^{(3)})}{d\lambda}=\lambda+1-\frac{3}{2}=0$$

$$x^{(4)}=x^{(3)}+\lambda_3 d^{(3)}=\begin{bmatrix}-\frac{3}{4}\\-\frac{1}{2}\\-\frac{3}{4}\end{bmatrix}$$

■ 最速下降法

➢最速下降法

$$\boldsymbol{d}^{(4)} = -\nabla f(\boldsymbol{x}^{(4)}) = \begin{bmatrix} 0 \\ -\frac{1}{2} \\ 0 \end{bmatrix}$$

$$\|\boldsymbol{d}^{(4)}\| = \sqrt{(0)^2+(-\frac{1}{2})^2+(0)^2} = \frac{1}{2} \leq \varepsilon$$

满足迭代截止条件

$$\boldsymbol{x}^{(4)} = \begin{bmatrix} -\frac{3}{4} \\ -\frac{1}{2} \\ -\frac{3}{4} \end{bmatrix}$$ **为满足要求的解**

■ 最速下降法

$$x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$$

➢正交特性

最速下降法极小化目标函数时，相邻两个搜索方向是正交的

Why?

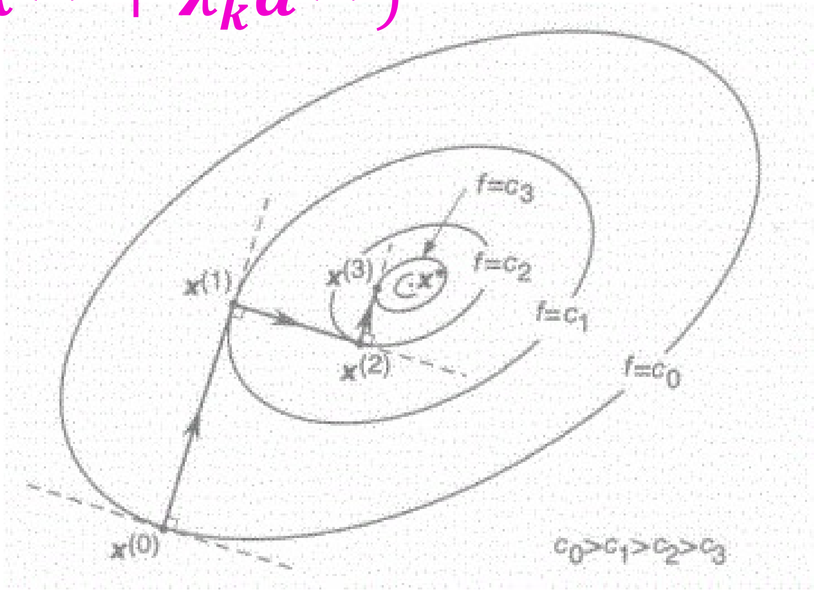
一维搜索函数

$$\varphi(\lambda) = f(x^{(k)} + \lambda d^{(k)}),\quad d^{(k)} = -\nabla f(x^{(k)})$$

$$\acute{\varphi}(\lambda_k) = \nabla f(x^{(k)} + \lambda_k d^{(k)})^T d^{(k)} = 0$$

$$d^{(k+1)} = -\nabla f(x^{(k+1)}) = -\nabla f(x^{(k)} + \lambda_k d^{(k)})$$

因此，$d^{(k)}$与$d^{(k+1)}$正交

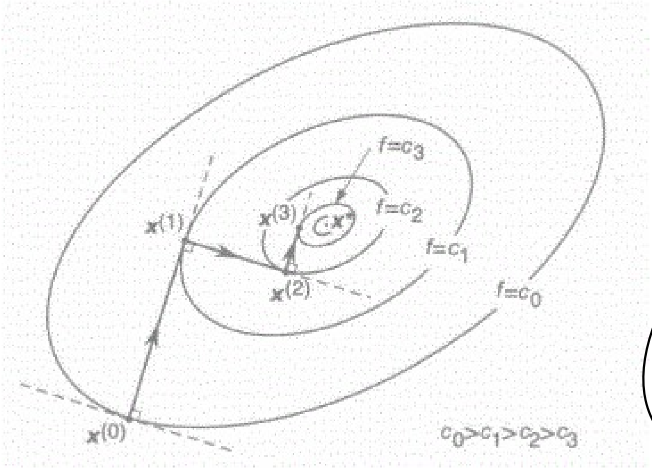


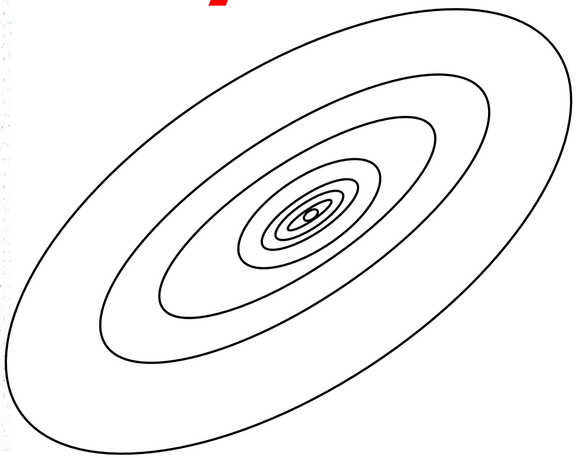
$$x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$$

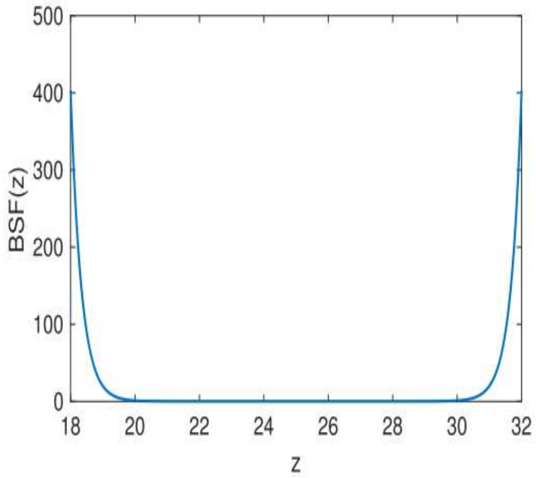
■ 最速下降法

➢锯齿现象

当迭代点接近极小点，每次搜索步长很小，出现**锯齿现象**，影响收敛速度 **Why?**







➢ **最速下降方向**反映了目标函数的一种**局部性质**，从全局看其**收敛是比较慢**的

➢ 最速下降法一般适用于计算过程的**前期迭代**。当接近极小值时，再使用其它算法。

目录

- 最速下降法（考）
- 牛顿法（考）
- 共轭梯度法（考）
- 拟牛顿法
- 信赖域法
- 最小二乘法

■ 牛顿法

- 牛顿法面向哪类优化问题
- 牛顿法的基本思想
- 牛顿法的步长
- 牛顿法的步骤

■ 牛顿法

➢适用于的优化问题

无约束优化问题：

$$\min \quad f(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathbf{R}^n,$$

其中，$f(x)$是二次可微实函数

适用的优化问题：

- **无约束**
- **目标函数二次可微**

■ 牛顿法

➢基本思想

获取函数$f(x)$在点$x^{(k)}$处二阶泰勒近似函数$\varphi(x)$，$\varphi(x)$的局部极小点就作为点$x^{(k+1)}$ 函数逼近法中 牛顿法向多维的推广

$$f(x) \approx f(x^{(k)}) + \nabla f(x^{(k)})^T (x - x^{(k)}) + \frac{1}{2}(x - x^{(k)})^T \nabla^2 f(x^{(k)})(x - x^{(k)})$$
$$= \varphi(x)$$

$$\nabla\varphi(x) = \nabla f(x^{(k)}) + \nabla^2 f(x^{(k)})(x - x^{(k)}) = 0$$

当海塞矩阵$\nabla^2 f(x^{(k)})$正定

局部极小点 $x = x^{(k)} - \nabla^2 f(x^{(k)})^{-1} \nabla f(x^{(k)})$

牛顿法的迭代公式：$x^{(k+1)} = x^{(k)} - \nabla^2 f(x^{(k)})^{-1} \nabla f(x^{(k)})$

■ 牛顿法

➢牛顿法的步长

牛顿法的迭代公式：

$$x^{(k+1)} = x^{(k)} - \nabla^2 f(x^{(k)})^{-1} \nabla f(x^{(k)})$$

牛顿法的方向：

$$d^{(k)} = -\nabla^2 f(x^{(k)})^{-1} \nabla f(x^{(k)})$$

牛顿法的迭代公式由目标函数的一阶导和二阶导确定，对应的步长为1

■ 牛顿法

➢牛顿法的步骤

(1) 给定初点$x^{(1)} \in R^n$，允许误差$\varepsilon > 0$，置$k = 1$;

(2) 计算$\nabla f(x^{(k)})$和$\nabla^2 f(x^{(k)})^{-1}$;

(3) 若$\|\nabla f(x^{(k)})\| \leq \varepsilon$,停止计算；否则，

$$d^{(k)} = -\nabla^2 f(x^{(k)})^{-1}\nabla f(x^{(k)})$$

(4) 令$x^{(k+1)} = x^{(k)} + d^{(k)}$,置$k := k + 1$,转到步骤(2).

■ 牛顿法

➢牛顿法

例题

$$\min\ f(x_1, x_2, x_3) = (x_1)^2 + x_1(1 - x_2) + (x_2)^2 - x_2 x_3 + (x_3)^2 + x_3$$

初始值 $x^{(1)} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ 允许误差 $\varepsilon = 0.5$

■ 牛顿法

➢牛顿法

解：

梯度 $$\nabla f(x)=\begin{bmatrix}2x_1+(1-x_2)\\-x_1+2x_2-x_3\\-x_2+2x_3+1\end{bmatrix}$$

海塞矩阵

$$\nabla^2 f(x)=\begin{bmatrix}2&-1&0\\-1&2&-1\\0&-1&2\end{bmatrix}$$ **正定**

海塞矩阵的逆

$$\nabla^2 f(x)^{-1}=\begin{bmatrix}3/4&1/2&1/4\\1/2&1&1/2\\1/4&1/2&3/4\end{bmatrix}$$

■ 牛顿法

➢牛顿法

$$\nabla f(x^{(1)}) = \begin{bmatrix}1\\0\\1\end{bmatrix}$$

满足截止条件?

$$\|\nabla f(x^{(1)})\| = \sqrt{(1)^2+(0)^2+(1)^2} = \sqrt{2} > \varepsilon$$

$k = 1$，进行第一次迭代

$$d^{(1)} = -\nabla^2 f(x^{(1)})^{-1}\nabla f(x^{(1)})$$

$$= -\begin{bmatrix}3/4 & 1/2 & 1/4\\1/2 & 1 & 1/2\\1/4 & 1/2 & 3/4\end{bmatrix}\begin{bmatrix}1\\0\\1\end{bmatrix} = -\begin{bmatrix}1\\1\\1\end{bmatrix}$$

$$x^{(2)} = x^{(1)} + d^{(1)} = \begin{bmatrix}-1\\-1\\-1\end{bmatrix}$$

■ 牛顿法

➢牛顿法

$$\nabla f(x^{(2)}) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\|\nabla f(x^{(2)})\| = \sqrt{(0)^2+(0)^2+(0)^2} = 0 < \varepsilon$$

满足迭代截止条件

$$x^* = x^{(2)} = \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}$$

■ 牛顿法

$$x^{(k+1)} = x^{(k)} - \nabla^2 f(x^{(k)})^{-1}\nabla f(x^{(k)})$$

➢二次型函数

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} + \mathbf{c}$$

令 $\nabla f(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b} = 0$ ➡ $\bar{\mathbf{x}} = -\mathbf{A}^{-1}\mathbf{b}$ **最优**

牛顿法如何求解?

$$\begin{aligned}\mathbf{x}^{(1)} &= \mathbf{x}^{(0)} - \mathbf{A}^{-1}\nabla f(\mathbf{x}^{(0)}) \\ &= \mathbf{x}^{(0)} - \mathbf{A}^{-1}(\mathbf{A}\mathbf{x}^{(0)} + \mathbf{b}) \\ &= -\mathbf{A}^{-1}\mathbf{b}\end{aligned}$$

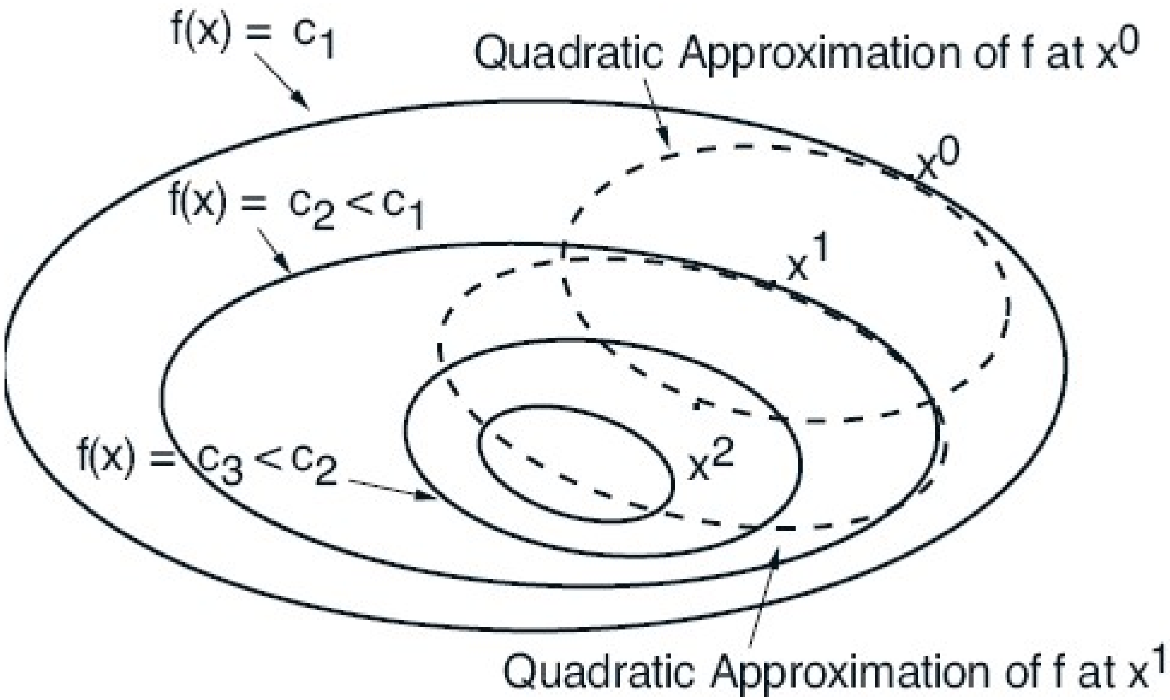
对于二次凸函数，牛顿法经过1次迭代达到最优值。

一些算法，把它们用于二次凸函数时，经有限次迭代必达到极小点．这种性质称为**二次终止性**

■ 牛顿法

➢牛顿法的特点

- ➢ 牛顿法利用目标函数的**一阶和二阶导数**
- ➢ 减小迭代次数，**但是**提高每次迭代计算量



■ 牛顿法

➢牛顿法的特点

- ➢ 当初始点远离极小点时，牛顿法可能不收敛 **Why?**

原因之一：牛顿方向不一定是下降方向

$$d^{(k)} = -\nabla^2 f(x^{(k)})^{-1} \nabla f(x^{(k)})$$

$$\nabla f(x^{(k)})^T d^{(k)} = -\nabla f(x^{(k)})^T \nabla^2 f(x^{(k)})^{-1} \nabla f(x^{(k)})$$

原因之二：即使目标函数值下降，得到的值也未必是沿牛顿方向的极小点。步长为1

- ➢ 改进：阻尼牛顿法—— 沿牛顿方向一维搜索

■ 牛顿法

➢阻尼牛顿法

■ 沿牛顿方向一维搜索

迭代公式：

$$x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$$

搜索方向：牛顿方向

$$d^{(k)} = -\nabla^2 f(x^{(k)})^{-1} \nabla f(x^{(k)})$$

搜索步长：

$$f(x^{(k)} + \lambda_k d_k) = \min_{\lambda>0} f(x^{(k)} + \lambda d^{(k)})$$ 一维搜索

每次迭代目标函数值有所下降（绝不会上升）

阻尼：就是阻碍目标函数值上升

■ 牛顿法

➢阻尼牛顿法

■ 计算步骤

(1) 给定初点$x^{(1)} \in R^n$，允许误差$\varepsilon > 0$，置$k = 1$;

(2) 计算$\nabla f(x^{(k)})$和$\nabla^2 f(x^{(k)})^{-1}$;

(3) 若$\left\|\nabla f(x^{(k)})\right\| \le \varepsilon$,停止计算；否则，

$$d^{(k)} = -\nabla^2 f(x^{(k)})^{-1} \nabla f(x^{(k)})$$

(4) 从$x^{(k)}$出发，沿$d^{(k)}$进行一维搜索，确定最优λ_k，使得$f(x^{(k)} + \lambda_k d_k) = \min_{\lambda>0} f(x^{(k)} + \lambda d^{(k)})$

(5) 令$x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$,置$k := k + 1$,转到步骤(2).

目录

■ 共轭梯度法

- 适用的优化问题
- 共轭梯度法的基本思想
- 什么是共轭方向
- 共轭梯度法的性质
- 典型共轭梯度法

■ 共轭梯度法

➢适用于的优化问题

无约束优化问题：

$$\min \quad f(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathbf{R}^n,$$

其中， $f(x)$**具有一阶连续偏导数**

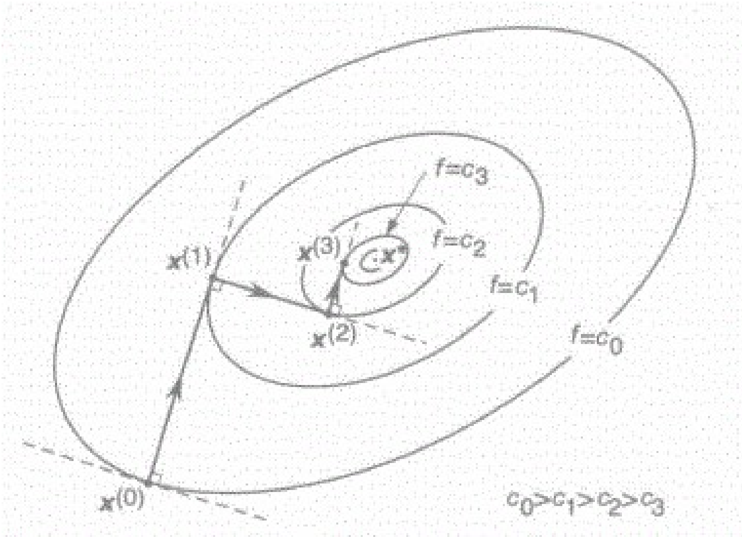
适用的优化问题：

- **无约束**
- **目标函数一阶连续偏导**

■ 共轭梯度法

➢基本思想

最速下降法极小化目标函数时，**相邻两个搜索方向**是**正交**的



把**共轭性**和**最速下降法**相结合，利用已知点$x^{(k)}$处的梯度**构造**一组**共轭方向**，并沿着这组方向进行搜索，求出目标函数的极小点。

需要了解：**共轭方向**　　　需要掌握：**共轭方向的构造方法**

■ 共轭梯度法

➢共轭方向

◆定义10．3．1 设A是n×n对称正定矩阵，若E^n中的两个方向$d^{(1)}$和$d^{(2)}$满足

$$d^{(1)T}Ad^{(2)}=0 \qquad (10.3.1)$$

则称这两个方向关于A共轭，或称它们关于A正交。

◆共轭是正交概念的推广

如A为单位矩阵，则两个方向关于A共轭等价于两个方向正交

$$\mathbf{d}^{(i)}\mathbf{A}\,\mathbf{d}^{(j)}=0$$

正交

■ 共轭梯度法

➢共轭方向的几何意义

设有二次函数

$$f(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}})^{\mathrm{T}}\mathbf{A}(\mathbf{x}-\bar{\mathbf{x}})$$

A是 $f(x)$**对称的正定矩阵**, $\bar{x}$为定点。

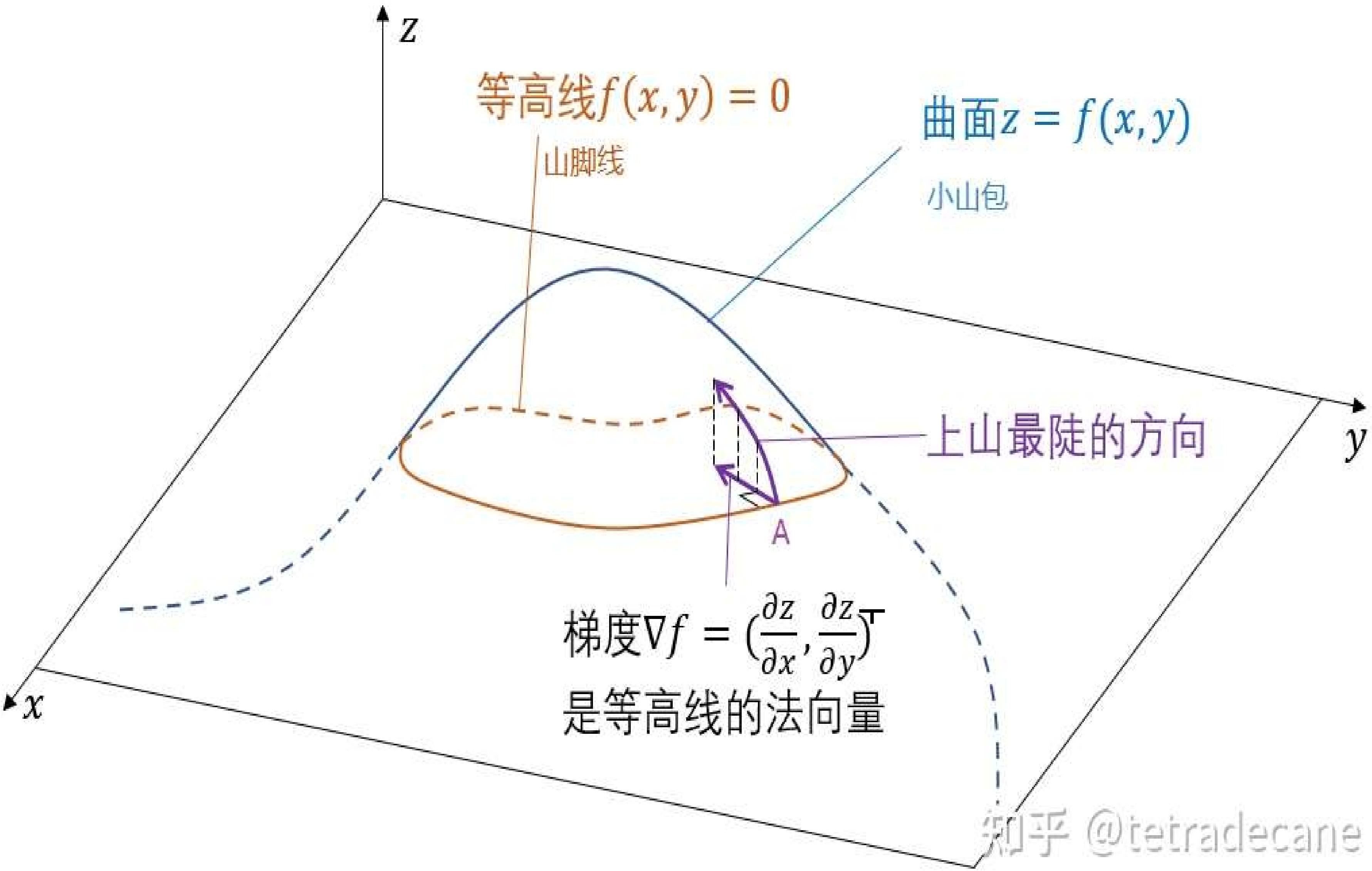
函数f(x)**等值面**为

$$\frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}})^{\mathrm{T}}\mathbf{A}(\mathbf{x}-\bar{\mathbf{x}}) = c$$ ⇨ 以 $\bar{x}$ 为中心的椭球面

若 $$\left.\begin{array}{c}\nabla f(\mathbf{x}) = \mathbf{A}(\mathbf{x}-\bar{\mathbf{x}}) = 0 \\ \mathbf{A}\text{正定}\end{array}\right\} \Rightarrow$$ $\bar{x}$ 为$f(x)$的局部极小值

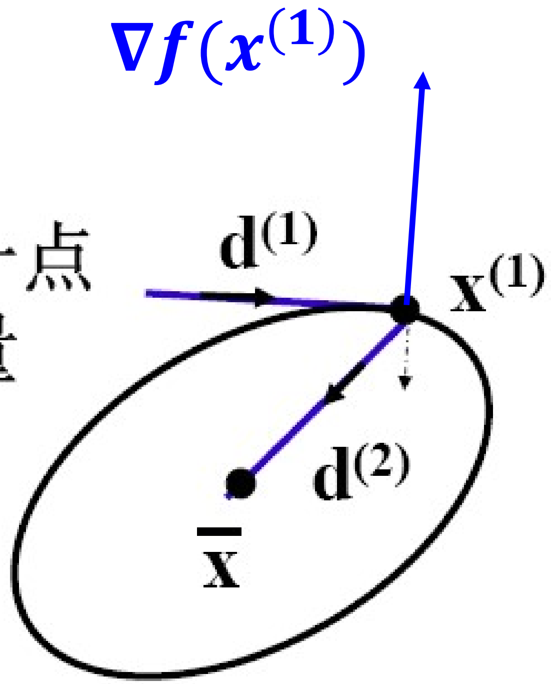
■ 共轭梯度法

➢共轭方向的几何意义



■ 共轭梯度法

➢共轭方向的几何意义



设$x^{(1)}$是在某个等值面上的一点

该等值面在点$x^{(1)}$处的法向量

$$\nabla f(x^{(1)}) = A(x^{(1)} - \bar{x})$$

$$d^{(2)} = \bar{x} - x^{(1)}$$

$$\nabla f(x^{(1)}) = -Ad^{(2)}$$

又设$d^{(1)}$是这个等值面在$x^{(1)}$处的一个切向

自 $d^{(1)}$ 与 $\nabla f(x^{(1)})$ 正交

因此有: $d^{(1)T}Ad^{(2)} = 0$

等值面上一点处的切向量与由这一点指向极小点的向量关于A共轭

■ 共轭梯度法

➢共轭梯度的性质

- 提高最速下降法的收敛速度，无需牛顿法计算海塞矩阵

- 目标函数为二次型，有限次迭代收敛至最优值

- 影响迭代搜索效率关键是搜索方向，对于二次型目标函数最好的搜索方向为共轭方向

- 也可求解非二次型目标函数

■ 共轭梯度法

➢典型共轭梯度法

$$f(\mathbf{x})=\frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}+\mathbf{b}^{\mathrm{T}}\mathbf{x}+\mathbf{c}$$

➢ **迭代公式** $x^{(k+1)}=x^{(k)}+\lambda_k d^{(k)}$

➢ **搜索方向** $d^{(1)},\cdots,d^{(k)}$ **关于A共轭**

$$d^{(i)}\mathbf{A}d^{(j)}=0 \quad i\neq j$$

➢ **搜索步长：** $\lambda_k=\arg\min\limits_{\lambda>0} f(x^{(k)}+\lambda d^{(k)})$ **一维搜索**

需讨论如何构造共轭方向，确定最优步长

■ 共轭梯度法

➢典型共轭梯度法

➢ 第k次迭代结束后，利用$d^{(k)}$ 和负梯度$-\nabla f(x^{(k+1)})$构造下一个搜索方向 $d^{(k+1)}$

$$d^{(k+1)} = -\nabla f(x^{(k+1)}) + \beta_k d^{(k)}$$

如何确定β_k?

正交 $(d^{(k)})^T A d^{(k+1)} = 0$

$$(d^{(k)})^T A(-\nabla f(x^{(k+1)}) + \beta_k d^{(k)}) = 0$$

$$-(d^{(k)})^T A \nabla f(x^{(k+1)}) + \beta_k (d^{(k)})^T A d^{(k)} = 0$$

$$\beta_k = \frac{(d^{(k)})^T A \nabla f(x^{(k+1)})}{(d^{(k)})^T A d^{(k)}} = \frac{(d^{(k)})^T A g_{k+1}}{(d^{(k)})^T A d^{(k)})}$$

$$g_{k+1} = \nabla f(x^{(k+1)})$$

■ 共轭梯度法

➢典型共轭梯度法

➢ **确定最优步长λ_k**

$$f(x^{(k)}+\lambda_k d_k)=\min_{\lambda>0} f(x^{(k)}+\lambda d^{(k)})$$

$$\varphi(\lambda)=f(x^{(k)}+\lambda d^{(k)})$$

$$\acute{\varphi}(\lambda)=\frac{df(x^{(k)}+\lambda d^{(k)})}{d\lambda}=\nabla f(x^{(k)}+\lambda d^{(k)})^T d^{(k)}$$

$$=(Ax^{(k)}+\lambda Ad^{(k)}+b)^T d^{(k)}$$

$$=(g_k+\lambda Ad^{(k)})^T d^{(k)}$$

$$=g_k{}^T d^{(k)}+\lambda d^{(k)T} Ad^{(k)}=0$$

$$\lambda_k=-\frac{g_k{}^T d^{(k)}}{d^{(k)T} Ad^{(k)}}$$

■ 共轭梯度法

➢典型共轭梯度法的步骤

$$f(\mathbf{x})=\frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}+\mathbf{b}^{\mathrm{T}}\mathbf{x}+\mathbf{c}$$

(1) 给定初点$x^{(1)} \in R^n$，允许误差$\varepsilon > 0$，置$k = 1$;

(2) 计算$g_k = \nabla f(x^{(k)})$,若$\|g_k\| = 0$,停止计算；否则进行步骤(3);

(3) 构造搜索方向

$$d^{(k)} = \begin{cases} -g_k, & k = 1 \\ -g_k + \beta_{k-1} d^{(k-1)}, & k > 1 \end{cases}$$

(4) 从$x^{(k)}$出发，沿$d^{(k)}$进行一维搜索，确定最优λ_k

$$\lambda_k = \arg\min_{\lambda} f(x^{(k)} + \lambda d^{(k)}) = -\frac{{g_k}^T d^{(k)}}{d^{(k)T} \mathbf{A} d^{(k)}}$$

(5) 若$k = n$，则停止计算；否则，$k := k + 1$,返回步骤(3)

■ 共轭梯度法

➢典型共轭梯度法

求解二次型目标函数

$$\min_{\boldsymbol{x}} x_1^2 + 2x_2^2$$

初始值 $\boldsymbol{x}^{(1)} = (5,5)^T$ ~~**允许误差**~~ $\cancel{\varepsilon = 0.5}$

■ 共轭梯度法

➢典型共轭法

解：梯度

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} 2x_1 \\ 4x_2 \end{bmatrix} \quad \boldsymbol{g}_1 = \nabla \boldsymbol{f}(\boldsymbol{x}^{(1)}) = \begin{bmatrix} \mathbf{10} \\ \mathbf{20} \end{bmatrix}$$

$$\|\boldsymbol{g}_1\| = \sqrt{(\mathbf{10})^2 + (\mathbf{20})^2} > \mathbf{0}$$

第1次迭代： $\boldsymbol{d}^{(1)} = -\boldsymbol{g}_1 = \begin{bmatrix} -10 \\ -20 \end{bmatrix}$

从$\boldsymbol{x}^{(1)}$出发，沿方向 $\boldsymbol{d}^{(1)}$ 做一维搜索，求λ_1

$$\lambda_1 = -\frac{\boldsymbol{g}_1^T \boldsymbol{d}^{(1)}}{\boldsymbol{d}^{(1)T} \mathbf{A} \boldsymbol{d}^{(1)}} = \frac{[-10 \ -20]\begin{bmatrix} -10 \\ -20 \end{bmatrix}}{[-10 \ -20]\begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}\begin{bmatrix} -10 \\ -20 \end{bmatrix}} = \frac{5}{18}$$

$$\boldsymbol{x}^{(2)} = \boldsymbol{x}^{(1)} + \lambda_1 \boldsymbol{d}^{(1)} = \begin{bmatrix} 5 \\ 5 \end{bmatrix} + \frac{5}{18}\begin{bmatrix} -10 \\ -20 \end{bmatrix} = \begin{bmatrix} \frac{20}{9} \\ -\frac{5}{9} \end{bmatrix}$$

■ 共轭梯度法

➢典型梯度法

$$g_2 = \nabla f\left(x^{(2)}\right) = \begin{bmatrix} \frac{40}{9} & -\frac{20}{9} \end{bmatrix}^{\mathrm{T}} \qquad \|g_2\| > 0$$

第2次迭代：

构造搜索方向：$d^{(2)} = -g_2 + \beta_1 d^{(1)}$，需求 β_1

$$\beta_1 = \frac{\left(d^{(1)}\right)^T A \nabla f(x^{(2)})}{\left(d^{(1)}\right)^T A\left(d^{(1)}\right)} = \frac{[-10\ -20]\begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}\begin{bmatrix} \frac{40}{9} \\ -\frac{20}{9} \end{bmatrix}}{[-10\ -20]\begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}\begin{bmatrix} -10 \\ -20 \end{bmatrix}} = \frac{4}{81}$$

从$x^{(2)}$出发，沿方向 $d^{(2)}$ 做一维搜索，求λ_2

$$\lambda_2 = -\frac{g_2^T d^{(2)}}{d^{(2)T} A d^{(2)}} = \frac{9}{20} \qquad x^{(3)} = x^{(2)} + \lambda_2 d^{(2)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$g_3 = \nabla f\left(x^{(3)}\right) = [0 \quad 0]^{\mathrm{T}} \qquad \|g_3\| = 0 \qquad x^* = x^{(3)}$$

■ 共轭梯度法

➢典型梯度法

■ 求解一般函数

➢ 步长无法采用公式，只能用一维搜索确定

$$\lambda_{k+1} = \arg\min_{\lambda>0} f(x^{(k+1)} + \lambda d^{(k+1)})$$

➢ 当前点的海塞矩阵替代A

$$\beta_k = \frac{(d^{(k)})^{\mathrm{T}} \mathbf{A} \nabla f(\mathbf{x}^{k+1})}{(d^{(k)})^{\mathrm{T}} \mathbf{A} d^{(k)}} \qquad \nabla^2 f(\mathbf{x}^{k+1}) \to \mathbf{A}$$

➢ 求解一般函数问题，有限步迭代达不到最优

■ 作业

P330：

2、14(偶数)

P331：

19（本章完结再做）