最优化理论
Optimality Theory

目 录 (CONTENT)

大连理工大学
Dalian University Of Technology

10

PART TEN

使用导数的最优化方法

Optimization Method Using Derivative

■ 主要内容

- 最速下降法
- 牛顿法
- 共轭梯度法
- 拟牛顿法
- 信赖域法

■ 概念

考虑无约束问题

$$\min f(x),\ x \in R^n \qquad (10.1.1)$$

其中 $f(x)$具有一阶连续偏导数。

在处理这类问题时，一般迭代策略是，希望从某一点出发，选择一个目标函数值下降最快的方向，沿此方向搜索以期尽快达到极小点，基于这一思想，Cauchy于1847年提出了**最速下降法**。这是无约束最优化中最简单的方法。

■ 概念

函数$f(x)$在点x处沿方向d的变化率可用方向导数表示

当函数可微时，有方向导数

$$Df(x,d)=\nabla f(x)^T d \quad (1.2)$$

求函数$f(x)$在点x处下降最快的方向,归结为求

$$\min \quad \nabla f(x)^T d$$

$$s.t \quad \|d\| \le 1 \qquad (1.3)$$

由*Schwartz*不等式,

$$\left|\nabla f(x)^T d\right| \le \|\nabla f(x)\|\|d\| \le \|\nabla f(x)\|$$

$$\Rightarrow \quad \nabla f(x)^T d \ge -\|\nabla f(x)\| \qquad (1.4)$$

■ 概念

由上式知. 当

$$d = -\frac{\nabla f(x)}{\|\nabla f(x)\|} \qquad (1.5)$$

时等号成立.故在点x处沿(1.5)所定义的方向变化率最小, 即负梯度方向为**最速下降方向.**

注意:在不同的尺度下最速下降方向是不同的.

■ 概念

最速下降算法的迭代公式为

$$x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)} \qquad (1.6)$$

其中$d^{(k)}$是从$x^{(k)}$出发的搜索方向, 此处取在点$x^{(k)}$的最速下降方向,即 $d^{(k)} = -\nabla f(x^{(k)})$.

λ_k是从$x^{(k)}$出发沿方向$d^{(k)}$进行一维搜索的步长, 即满足

$$f(x^{(k)} + \lambda_k d^{(k)}) = \min_{\lambda \geq 0} f(x^{(k)} + \lambda d^{(k)}) \qquad (1.7)$$

■ 算法

算法描述

$Step1,$ 给定初始点 $x^{(k)} \in E^n,$ 允许误差 $\varepsilon > 0,$ 置 $k = 1$

$Step2,$ 计算搜索方向 $d^{(k)} = -\nabla f(x^{(k)})$

$Step3,$ 若 $\left\| d^{(k)} \right\| \le \varepsilon,$ 停止, 否则, 从 $x^{(k)}$ 出发, 沿 $d^{(k)}$ 进行一维搜索,

求 λ_k, 使得 $\quad f(x^{(k)} + \lambda_k d^{(k)}) = \min_{\lambda \ge 0} f(x^{(k)} + \lambda d^{(k)})$

$Step4,$ 令 $x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)},$ 置 $k := k + 1,$ 转 $Step2$

■ 例子

例1.1 用最速下降法求解下列问题

$$\min \quad f(x)=2x_1^2+x_2^2$$

$$初始点 x^{(1)}=(1,1)^T, \varepsilon=\frac{1}{10}$$

第一次迭代

目标函数$f(x)$在点x处的梯度

$$\nabla f(x)=\begin{bmatrix}4x_1\\2x_2\end{bmatrix}$$

■ 例子

令搜索方向

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

$$\|d\| = \sqrt{16+4} = 2\sqrt{5} > 1/10$$

从$x^{(1)}$出发,沿方向$d^{(1)}$进行一维搜索,求步长λ_1,即

$$\min_{\lambda\geq 0} \varphi(\lambda) = f(x^{(1)} + \lambda d^{(1)})$$

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

$$\varphi(\lambda) = 2(1-4\lambda)^2 + (1-2\lambda)^2$$

■ 例子

令 $\varphi'(\lambda) = -16(1-4\lambda) - 4(1-2\lambda) = 0$

$\Rightarrow \quad \lambda_1 = 5/18$

在直线上的极小点

$$x^{(2)} = x^{(1)} + \lambda_1 d^{(1)} = \begin{bmatrix} -1/9 \\ 4/9 \end{bmatrix}$$

第二次迭代

$f(x)$在点$x^{(2)}$处的最速下降方向为

$$d^{(2)} = -\nabla f(x^{(2)}) = \begin{bmatrix} 4/9 \\ -8/9 \end{bmatrix}$$

$$\left\| d^{(2)} \right\| = \frac{4}{9}\sqrt{5} > 1/10$$

■ 例子

从$x^{(2)}$出发,沿方向$d^{(2)}$进行一维搜索:

$$\min_{\lambda\geq 0} \varphi(\lambda)=f(x^{(2)}+\lambda d^{(2)})$$

$$x^{(2)}+\lambda d^{(2)}=\begin{bmatrix}-1/9\\4/9\end{bmatrix}+\lambda\begin{bmatrix}4/9\\-8/9\end{bmatrix}=\begin{bmatrix}(-1+4\lambda)/9\\(4-8\lambda)/9\end{bmatrix}$$

$$\varphi(\lambda)=\frac{2}{81}(-1+4\lambda)^2+\frac{16}{81}(1-2\lambda)^2$$

令

$$\varphi'(\lambda)=\frac{16}{81}(-1+4\lambda)-\frac{64}{81}(1-2\lambda)=0$$

$$\Rightarrow \qquad \lambda_2=5/12$$

■ 例子

得到

$$x^{(3)} = x^{(2)} + \lambda_2 d^{(2)} = \frac{2}{27}\begin{bmatrix}1\\1\end{bmatrix}$$

第三次迭代

$f(x)$在点$x^{(3)}$处的最速下降方向为

$$d^{(3)} = -\nabla f(x^{(3)}) = \frac{4}{27}\begin{bmatrix}-2\\-1\end{bmatrix}$$

$$\left\| d^{(3)} \right\| = \frac{4}{27}\sqrt{5} > 1/10$$

■ 例子

从$x^{(3)}$出发,沿方向$d^{(3)}$进行一维搜索:

$$\min_{\lambda\geq 0}\varphi(\lambda)=f(x^{(3)}+\lambda d^{(3)})$$

$$x^{(3)}+\lambda d^{(3)}=\frac{2}{27}\begin{bmatrix}1\\1\end{bmatrix}+\lambda\frac{4}{27}\begin{bmatrix}-2\\-1\end{bmatrix}=\frac{2}{27}\begin{bmatrix}1-4\lambda\\1-2\lambda\end{bmatrix}$$

$$\varphi(\lambda)=\frac{8}{27^2}(1-4\lambda)^2+\frac{4}{27^2}(1-2\lambda)^2$$

令 $\varphi'(\lambda)=0$

$\Rightarrow \lambda_2=5/18$

■ 例子

$$x^{(4)} = x^{(3)} + \lambda_2 d^{(3)} = \frac{2}{27}\begin{bmatrix} -1/9 \\ 4/9 \end{bmatrix} = \frac{2}{243}\begin{bmatrix} -1 \\ 4 \end{bmatrix}$$

此时

$$\left\|\nabla f(x^{(4)})\right\| = \frac{8}{243}\sqrt{5} < \frac{1}{10}$$

已经满足精度要求, 得近似解

$$\tilde{x} = \frac{2}{243}\begin{bmatrix} -1 \\ 4 \end{bmatrix}$$

问题的最优解为x*=(0,0)

■ 算法的收敛性

*Theorem*1.1　设$f(x)$是连续可微的实函数，解集合$\Omega=\{\bar{x} \mid \nabla f(\bar{x})=0\}$，最速下降法产生的序列$\{x^{(k)}\}$含于某个紧集，则序列$\{x^{(k)}\}$的每个聚点$\hat{x}\in\Omega$.

*Theorem*1.2　设$f(x)$存在连续二阶偏导，$\bar{x}$是局部极小点Hessian矩阵$\nabla^2 f(\bar{x})$的最小特征值$a>0$，最大特征值为A，算法产生的序列$\{x^{(k)}\}$收敛于$\bar{x}$，则目标函数序列$\{f(x^{(k)})\}$以不大于$\left(\dfrac{A-a}{A+a}\right)^2$的收敛比线性的收敛于$f(\bar{x})$.

■ 算法的收敛性

在上述定理中, 若令$r=A/a$, 则

$$\left(\frac{A-a}{A+a}\right)^2=\left(\frac{r-1}{r+1}\right)^2<1$$

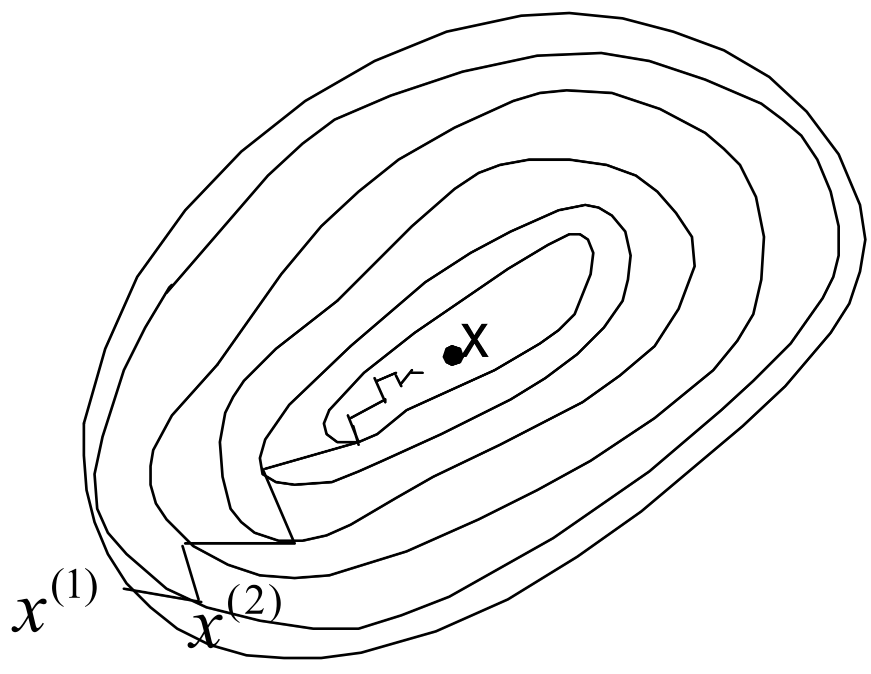
r是对称正定矩阵$\nabla^2 f(x)$的条件数.

定理表明:

条件数越小, 收敛越快; 条件数越大, 收敛越慢.

■ 性质1

最速下降法存在锯齿现象



■ 性质2

容易证明, 用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.

令

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

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

则

$$\varphi'(\lambda) = \nabla f(x^{(k)} + \lambda_k d^{(k)})^{\mathrm{T}} d^{(k)} = 0$$

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

$$\Rightarrow d^{(k+1)} = -\nabla f(x^{(k+1)}) \text{与} d^{(k)} = -\nabla f(x^{(k)}) \text{正交}$$

■ 概念

设$f(x)$是二次可微函数,$x \in R^n$.又设$x^{(k)}$是$f(x)$的极小点的一个估计,将$f(x)$在$x^{(k)}$点Taylor展开,取二阶近似:

$$
\begin{aligned}
f(x) &\approx \phi(x) \\
&= f(x^{(k)}) + \nabla f(x^{(k)})^T (x - x^{(k)}) + \\
&\quad \frac{1}{2}(x - x^{(k)})^T \nabla^2 f(x^{(k)})(x - x^{(k)})
\end{aligned}
$$

其中$\nabla^2 f(x^{(k)})$是$f(x)$在点$x^{(k)}$处的Hessian矩阵.

■ 概念

为求ϕ (x)的驻点,令 $\nabla\phi(x)=0$

即

$$\nabla f(x^{(k)})+\nabla^2 f(x^{(k)})(x-x^{(k)})=0 \qquad (10.2.1)$$

设$\nabla^2 f(x^{(k)})$可逆, 则得牛顿法的迭代公式

$$x^{(k+1)}=x^{(k)}-\nabla^2 f(x^{(k)})^{-1}\nabla f(x^{(k)}) \qquad (10.2.2)$$

算法(Newton法)

$Step1$,给定初始点$x^{(0)}$,允许误差$\varepsilon>0$, 置$k=1$;

$Step2$,若$\left\|\nabla f(x^{(k)})\right\|\le\varepsilon$,停止,得解$x^{(k)}$;否则,令

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

■ 算法的收敛性

注意:牛顿法的迭代格式也可以从最速下降方向的角度来理解.

下面介绍一下A度量及其意义下的最速下降方向.
设A为对称正定矩阵,向量d的A范数定义为

$$\|d\|_A = \sqrt{d^T A d}$$

下求A度量下的最速下降方向,为此,考虑

$$\begin{aligned} \min \quad & \nabla f(x)^T d \\ s.t \quad & d^T A d \le 1 \end{aligned} \qquad (10.2.3)$$

■ 算法的收敛性

由A, A^{-1}对称正定,故存在对称平方根$A^{1/2}$, $A^{-1/2}$,使得

$$A = A^{\frac{1}{2}}A^{\frac{1}{2}}, \qquad A^{-1} = A^{-\frac{1}{2}}A^{-\frac{1}{2}}$$

于是

$$d^T A d = d^T A^{\frac{1}{2}} A^{\frac{1}{2}} d = (A^{\frac{1}{2}} d)^T A^{\frac{1}{2}} d$$

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

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

■ 算法的收敛性

令$y = A^{\frac{1}{2}}d$,则(10.2.3)可写成

$$\min \quad (A^{-\frac{1}{2}}\nabla f(x))^T y$$
$$s.t \quad y^T y \le 1 \qquad (10.2.4)$$

根据Schwartz不等式,得到

$$\left|(A^{-\frac{1}{2}}\nabla f(x))^T y\right| \le \left\|A^{-\frac{1}{2}}\nabla f(x)\right\| \|y\| \le \left\|A^{-\frac{1}{2}}\nabla f(x)\right\|$$

去掉绝对值符号,有

$$(A^{-\frac{1}{2}}\nabla f(x))^T y \ge -\left\|A^{-\frac{1}{2}}\nabla f(x)\right\|$$

■ 算法的收敛性

即

$$\nabla f(x)^T A^{-\frac{1}{2}} A^{\frac{1}{2}} d \geq -\left\| A^{-\frac{1}{2}} \nabla f(x) \right\|$$

$$\Rightarrow \nabla f(x)^T d \geq -\left\| A^{-\frac{1}{2}} \nabla f(x) \right\|$$

为得到在点x处下降最快的方向,按下式选取d

$$d = \frac{-A^{-1}\nabla f(x)}{(\nabla f(x)^T A^{-1} \nabla f(x))^{\frac{1}{2}}} \qquad (10.2.5)$$

这时上式等号成立,由此确定的方向即**度量A意义下的最速下降方向**

■ 算法的收敛性

若取

$$A = G_k = \nabla^2 f(x^k)$$

则牛顿法的搜索方向实际上是关于向量椭球范数 $\|\cdot\|_{G_k}$ 的最速下降方向,其步长为$\left\|\nabla f(x^k)\right\|_{G_k}$.

■ 例子

例 用牛顿法求解下列问题

$$\min \quad (x_1-1)^4+x_2^2$$

$$取初点 x^{(1)}=(0,1)'$$

$$\nabla f(x)=\begin{bmatrix} 4(x_1-1)^3 \\ 2x_2 \end{bmatrix}, \nabla^2 f(x)=\begin{bmatrix} 12(x_1-1)^2 & 0 \\ 0 & 2 \end{bmatrix}$$

第1次迭代

$$\nabla f(x^{(1)})=\begin{bmatrix} -4 \\ 2 \end{bmatrix}, \nabla^2 f(x^{(1)})=\begin{bmatrix} 12 & 0 \\ 0 & 2 \end{bmatrix}$$

■ 例子

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

$$= \begin{bmatrix} 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 12 & 0 \\ 0 & 2 \end{bmatrix}^{-1} \begin{bmatrix} -4 \\ 2 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 0 \end{bmatrix}$$

第2次迭代

$$\nabla f(x^{(2)}) = \begin{bmatrix} -32/27 \\ 0 \end{bmatrix}, \nabla^2 f(x^{(2)}) = \begin{bmatrix} 48/9 & 0 \\ 0 & 2 \end{bmatrix}$$

$$x^{(3)} = x^{(2)} - \nabla^2 f(x^{(2)})^{-1} \nabla f(x^{(2)})$$

$$= \begin{bmatrix} 1/3 \\ 0 \end{bmatrix} - \begin{bmatrix} 48/9 & 0 \\ 0 & 2 \end{bmatrix}^{-1} \begin{bmatrix} -32/27 \\ 0 \end{bmatrix} = \begin{bmatrix} 5/9 \\ 0 \end{bmatrix}$$

■ 例子

第3次迭代

$$\nabla f(x^{(3)})=\begin{bmatrix}-256/729\\0\end{bmatrix},\nabla^2 f(x^{(2)})=\begin{bmatrix}64/27&0\\0&2\end{bmatrix}$$

$$x^{(4)}=x^{(3)}-\nabla^2 f(x^{(3)})^{-1}\nabla f(x^{(4)})$$

$$=\begin{bmatrix}5/9\\0\end{bmatrix}-\begin{bmatrix}64/27&0\\0&2\end{bmatrix}^{-1}\begin{bmatrix}-256/729\\0\end{bmatrix}=\begin{bmatrix}19/27\\0\end{bmatrix}$$

继续下去,第4次迭代,…得到点列收敛于(1,0),此为最优解.

■ 局部收敛性

定理10.2.1 设$f(x)$是二次可微函数，$x\in E^n$，$\bar{x}$满足$\nabla f(\bar{x})=0$且$\nabla^2 f(x^{(k)})^{-1}$存在，又设初始点$x^{(1)}$充分接近$\bar{x}$，使得存在$k_1,k_2>0$，满足$k_1k_2<1$，且对每一个

$$x\in X=\left\{x \,\middle|\, \|x-\bar{x}\|\le\|x^{(1)}-\bar{x}\|\right\}\text{成立:}$$

$$(1)\ \|\nabla^2 f(x^{(k)})^{-1}\|\le k_1,$$

$$(2)\ \frac{\|\nabla f(\bar{x})-\nabla f(x)-\nabla^2 f(x)(\bar{x}-x)\|}{\|\bar{x}-x\|}\le k_2$$

则牛顿法产生的序列收敛于$\bar{x}$.

证明:根据(10.2.2)，牛顿算法映射定义为

$$A(x)=x-\nabla^2 f(x)^{-1}\nabla f(x) \qquad \text{(a)}$$

定义解集合$\Omega=\{\bar{x}\}$,令函数$\alpha(x)=\|x-\bar{x}\|$

下证α(x) 是关于解集合Ω和算法A的下降函数.

令$x\in X$,且$x\neq\bar{x}$,又令$y\in A(x)$.

根据算法A的定义及$\nabla f(\bar{x})=0$的假设,有

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

$$=(x-\bar{x})-\nabla^2 f(x)^{-1}[\nabla f(x)-\nabla(\bar{x})]$$

$$=\nabla f^2 f(x)^{-1}[\nabla(\bar{x})-\nabla f(x)-\nabla^2 f(x)(\bar{x}-x)] \qquad \text{(b)}$$

于是可得

$$\|y-\overline{x}\| \le \left\|\nabla^2 f(x)^{-1}\right\| \left\|\nabla(\overline{x}) - \nabla f(x) - \nabla^2 f(x)(\overline{x}-x)\right\|$$

$$\le k_1 k_2 \|x-\overline{x}\| < \|\overline{x}-x\| \qquad \text{(c)}$$

从而α(x) 是关于解集合Ω和算法A的下降函数

由(c)可知, $y \in X$, 故迭代产生的序列$\{x^{(k)}\} \subset X$根据定义知X是紧集, 故迭代产生的序列含于紧集. 此外,算法映射A在紧集X上是闭的.

综上, 迭代产生的序列$\{x^{(k)}\}$必收敛于$\overline{x}$.

■ 局部收敛性

定理(局部收敛定理)设函数$f \in C^2(R^n)$, 它在x^*的梯度$\nabla f(x^*) = 0$, Hesse矩阵$\nabla^2 f(x^*)$正定. 若初始点x^0充分靠近x^*, 并且Hesse矩阵$G(x) = \nabla^2 f(x)$满足Lipschitz条件, 即存在$L > 0$, 使得$\forall x, y \in R^n$, 有

$$\|G(x) - G(y)\| \le L\|x - y\|,$$

则对$\forall$k, 迭代格式(10.2.2)有意义, 且迭代点序列$\{x^k\}$以二阶的收敛速度收敛到x^*.

大连理工大学 Dalian University Of Technology

■ 局部收敛性

当牛顿法收敛时, 有下列关系

$$\left\|x^{(k+1)}-\bar{x}\right\| \leq c\left\|x^{(k)}-\bar{x}\right\|^{2}$$

c是某个常数, 因此算法至少是2阶收敛的.

特别的, 对于二次凸函数, 用牛顿法求解, 经一次迭代即达到极小点. 设有二次凸函数

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

其中A是对称正定矩阵

牛顿法

- 注意，当初始点远离极小点时，牛顿法可能不收敛

- **阻尼牛顿法**

基本思想：增加了沿牛顿方向的一维搜索.

迭代公式为

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

其中$d^{(k)}=-\nabla^2 f(x^{(k)})^{-1}\nabla f(x^{(k)})$为牛顿方向，$\lambda_k$是由一维搜索所得的步长,即满足

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

- 算法(阻尼牛顿法)

$Step1$, 给定初始点 $x^{(0)}$, 允许误差 $\varepsilon>0$, 置 $k=1$;

$Step2$, 计算 $\nabla f(x^{(k)}), \nabla^2 f(x^{(k)})^{-1}$

$Step3$, 若 $\left\|\nabla f(x^{(k)})\right\| \le \varepsilon$, 停止, 得解 $x^{(k)}$; 否则, 令

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

$Step4$, 从 $x^{(k)}$ 出发, 沿方向 $d^{(k)}$ 作一维搜索:

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

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

$Step5$, 置 $k:=k+1$, 转2

■ 存在问题

显然可能存在：

1) Hessian 矩阵$\nabla^2 f(x^{(1)})$奇异-不可逆

2) 即使非奇异，Hessian 矩阵$\nabla^2 f(x^{(1)})$也可能非正定

牛顿方向不一定是下降方向，算法失效！

大连理工大学 Dalian University Of Technology

■ 例子

- 例 用阻尼牛顿法求解下列问题

$$\min f(x) = x_1^4 + x_1 x_2 + (1 + x_2)^2$$

取初点$x^{(1)} = (0,0)^T$.在该点函数的梯度和Hessian阵为

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

牛顿方向

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

$$= \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 2 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$$

大连理工大学 Dalian University Of Technology

■ 例子

从$x^{(1)}$出发,沿方向$d^{(1)}$进行一维搜索,令

$$\varphi(\lambda)=f(x^{(1)}+\lambda d^{(1)})=16\lambda^4+1$$

再令

$$\varphi'(\lambda)=0\Rightarrow\lambda_1=0$$

显然,用阻尼牛顿法不能产生新点,而点$x^{(1)}=(0,0)^T$并不是问题极小点.可见从$x^{(1)}$出发,用阻尼牛顿法求不出问题的极小点,原因在于 Hessian 矩阵$\nabla^2 f(x^{(1)})$非正定

■ 修正牛顿法

考虑 (10.2.2),记搜索方向$d^{(k)} = x - x^{(k)}$

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

阻尼牛顿法所用搜索方向是上述方程的解

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

此处假设逆矩阵 $\nabla^2 f(x^{(k)})^{-1}$ 存在

■ 修正牛顿法

解决$Hessian$矩阵$\nabla^2 f(x^{(k)})$非正定问题的基本思想

修正$\nabla^2 f(x^{(k)})$,构造一个对称正定矩阵G_k,在方程(d)

中用G_k取代矩阵$\nabla^2 f(x^{(k)})$:

$$G_k d^{(k)} = -\nabla f(x^{(k)}) \qquad \text{(f)}$$

$$\Rightarrow d^{(k)} = -G_k^{-1} \nabla f(x^{(k)}) \qquad \text{(g)}$$

再沿此方向作一维搜索

如何构造G_k?比如,可令

$$G_k = \nabla^2 f(x^{(k)}) + \varepsilon_k I \qquad \text{(h)}$$

其中I是单位阵,ε_k是一个适当的正数.

大连理工大学 Dalian University Of Technology

■ 修正牛顿法

$Step1$,给定初始点$x^{(0)}$,允许误差$\varepsilon > 0$,置 $k=0$;

$Step2$,计算梯度$g^k = \nabla f(x^{(k)})$,若$\|\nabla f(x^{(k)})\| \le \varepsilon$,停止,得解 $x^{(k)}$;

否则转Step3

$Step3$, 计算Hesse矩阵$G_k = \nabla^2 f(x^{(k)})$,置矩阵 $B_k = G_k + E_k$,其中E_k为修正矩阵(当G_k正定时,它取0),计算修正牛顿方向

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

$Step4$,从$x^{(k)}$出发,沿方向$d^{(k)}$作(精确或非精确)一维搜索:

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

令 $x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$,置$k := k+1$,转 2

■ 修正牛顿法

定理(全局收敛定理) 设$f:R^n\to \mathrm{R}$在某开集D上二阶连续可微,且修正牛顿法的初始点$x^0\in \mathrm{D}$使得f的水平集$S_{f(x^0)}=\{x\in D\mid f(x)\le f(x^0)\}$是紧集. 若矩阵序列满足有界分解特性,则$\lim\limits_{x\to\infty}\nabla f(x^k)=0.$

小结

- 使用导数基本概念
- 最速下降法
- 牛顿法

作业

- 习题 3、4、5、10、14、17、18、19