最优化算法/实用优化算法
第三章　无约束最优化方法

教师　邱松强

中国矿业大学　数学学院

June 10, 2021

目录

本章介绍无约束优化问题

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

的计算方法.

- 本章介绍的方法基本都属于下降算法;
- 本章介绍算法的关键在于如何选取搜索方向.

一元函数的最优性条件

回顾

性质1

设$\varphi(\alpha)$ 为定义在R 上的一元函数，则

(1) 若α^* 为$\varphi(\alpha)$ 的局部极小点，则$\varphi'(x^*) = 0$;

(2) 若$\varphi'(\alpha^*) = 0$, $\varphi''(\alpha^*) > 0$, 则α^* 为$\varphi(\alpha)$ 的严格局部极小点;

(3) 若α^* 为$\varphi(\alpha)$ 的严格局部极小点，则$\varphi'(\alpha^*) = 0$, $\varphi''(\alpha^*) \geq 0$.

最优性条件

定理1 (一阶必要条件)

若$\boldsymbol{x}^*$ 为$f(\boldsymbol{x})$ 的局部极小点，且在$\boldsymbol{x}^*$ 的某邻域内$f(\boldsymbol{x})$ 具有一阶连续偏导数，则

$$\nabla f(\boldsymbol{x}^*) = 0.$$

注：

最优性条件

定理1 (一阶必要条件)

若$\boldsymbol{x}^*$ 为$f(\boldsymbol{x})$ 的局部极小点，且在$\boldsymbol{x}^*$ 的某邻域内$f(\boldsymbol{x})$ 具有一阶连续偏导数，则

$$\nabla f(\boldsymbol{x}^*) = 0.$$

注：

- 满足上面条件的点称为驻点. 驻点有三种类型：极小点、极大点和鞍点.

最优性条件

定理1 (一阶必要条件)

若$\boldsymbol{x}^*$ 为$f(\boldsymbol{x})$ 的局部极小点，且在$\boldsymbol{x}^*$ 的某邻域内$f(\boldsymbol{x})$ 具有一阶连续偏导数，则

$$\nabla f(\boldsymbol{x}^*) = 0.$$

注：

- 满足上面条件的点称为驻点．驻点有三种类型：极小点、极大点和鞍点．
- 鞍点：沿某些方向是极大点，沿另一些方向是极大点．

鞍点

如图

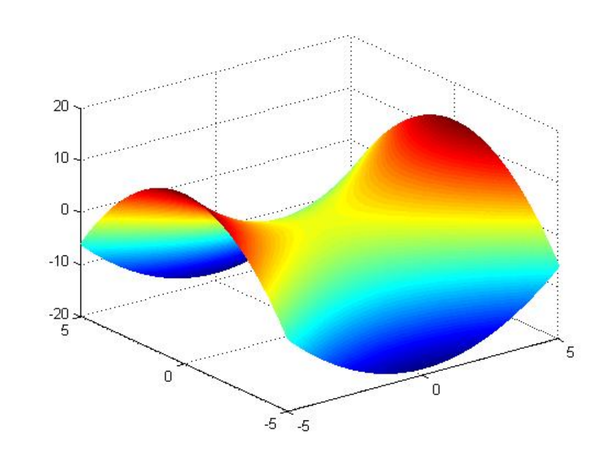


图: 鞍点

一阶必要性条件的证明

证明： 若$\nabla f(\boldsymbol{x}^*) \neq 0$，则存在方向$p \in R^n$ (例如，$p = -\nabla f(\boldsymbol{x}^*)$) 使得$p^T \nabla f(\boldsymbol{x}^*) < 0$. 由微分中值定理，存在$\alpha_1 \in (0, \alpha)$ 使得

$$f(\boldsymbol{x}^* + \alpha p) = f(\boldsymbol{x}^*) + \alpha p^T \nabla f(\boldsymbol{x}^* + \alpha_1 p)$$

成立. 由于∇f 在$\boldsymbol{x}^*$ 的某邻域内连续，故存在$\delta > 0$，使得$\forall \alpha \in [0, \delta]$，有$p^T \nabla f(\boldsymbol{x}^* + \alpha_1 p) < 0$. 所以，对$\forall \alpha \in (0, \delta)$ 有

$$f(\boldsymbol{x}^* + \alpha p) < f(\boldsymbol{x}^*).$$

这与x^* 是f 的局部极小点矛盾.

例2.1

计算 $Rosenbrock$ 函数的梯度 $\nabla f(x)$ 和 $Hesse$ 矩阵 $\nabla^2 f(x)$

$$f(\boldsymbol{x}) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2.$$

证明 $\boldsymbol{x}^* = (1,1)^T$ 是 $f(\boldsymbol{x})$ 的唯一局部极小点，且在此点处 $Hesse$ 矩阵是正定的.

例2.1

计算*Rosenbrock* 函数的梯度$\nabla f(x)$ 和*Hesse* 矩阵$\nabla^2 f(x)$

$$f(\boldsymbol{x}) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2.$$

证明$\boldsymbol{x}^* = (1,1)^T$ 是$f(\boldsymbol{x})$ 的唯一局部极小点，且在此点处*Hesse* 矩阵是正定的.

【解：】该函数的梯度为

$$\nabla f(\boldsymbol{x}) = \begin{pmatrix} -400\left(x_2 - x_1^2\right)x_1 - 2\left(1 - x_1\right) \\ 200\left(x_2 - x_1^2\right) \end{pmatrix},$$

例2.1

计算*Rosenbrock* 函数的梯度$\nabla f(x)$ 和*Hesse* 矩阵$\nabla^2 f(x)$

$$f(\boldsymbol{x}) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2.$$

证明$\boldsymbol{x}^* = (1,1)^T$ 是$f(\boldsymbol{x})$ 的唯一局部极小点，且在此点处*Hesse* 矩阵是正定的.

【解：】该函数的梯度为

$$\nabla f(\boldsymbol{x}) = \begin{pmatrix} -400\left(x_2 - x_1^2\right)x_1 - 2\left(1 - x_1\right) \\ 200\left(x_2 - x_1^2\right) \end{pmatrix},$$

Hesse 矩阵为

$$G(\boldsymbol{x}) = \begin{pmatrix} 1200x_1^2 - 400x_2 + 2 & -400x_1 \\ -400x_1 & 200 \end{pmatrix}.$$

令$\nabla f(\boldsymbol{x})=0$, 得$x^*=(1,1)^T$. 又当$\boldsymbol{x}\neq\boldsymbol{x}^*$ 时, $f(\boldsymbol{x})>f(\boldsymbol{x}^*)$. 所以$\boldsymbol{x}^*$ 是f 的唯一的极小值点.

令$\nabla f(\boldsymbol{x})=0$, 得$x^*=(1,1)^T$. 又当$\boldsymbol{x}\neq\boldsymbol{x}^*$ 时, $f(\boldsymbol{x})>f(\boldsymbol{x}^*)$. 所以$\boldsymbol{x}^*$ 是f 的唯一的极小值点. 由

$$G(\boldsymbol{x}^*)=\begin{pmatrix}802 & -400\\ -400 & 200\end{pmatrix}.$$

由于$802>0$, $\det(G(\boldsymbol{x}^*))=400>0$, 故而$G(\boldsymbol{x}^*)$ 正定.

例2.2

证明函数

$$f(\boldsymbol{x}) = 8x_1 + 12x_2 + x_1^2 - 2x_2^2$$

有唯一的稳定点，且该点既非极小点也非极大点，而是一个鞍点．画出 $f(\boldsymbol{x})$ 的等高线图．

【解:】

$$\nabla f(\boldsymbol{x}) = \begin{pmatrix} 8 + 2x_1 \\ 12 - 4x_2 \end{pmatrix}.$$

例2.2

证明函数

$$f(\boldsymbol{x}) = 8x_1 + 12x_2 + x_1^2 - 2x_2^2$$

有唯一的稳定点，且该点既非极小点也非极大点，而是一个鞍点．画出$f(\boldsymbol{x})$ 的等高线图．

【解:】

$$\nabla f(\boldsymbol{x}) = \begin{pmatrix} 8 + 2x_1 \\ 12 - 4x_2 \end{pmatrix}.$$

令$\nabla f(\boldsymbol{x}) = 0$，得唯一稳定点$\boldsymbol{x}^* = (-4, 3)^T$．

例2.2

证明函数

$$f(\boldsymbol{x}) = 8x_1 + 12x_2 + x_1^2 - 2x_2^2$$

有唯一的稳定点，且该点既非极小点也非极大点，而是一个鞍点．画出$f(\boldsymbol{x})$ 的等高线图．

【解:】

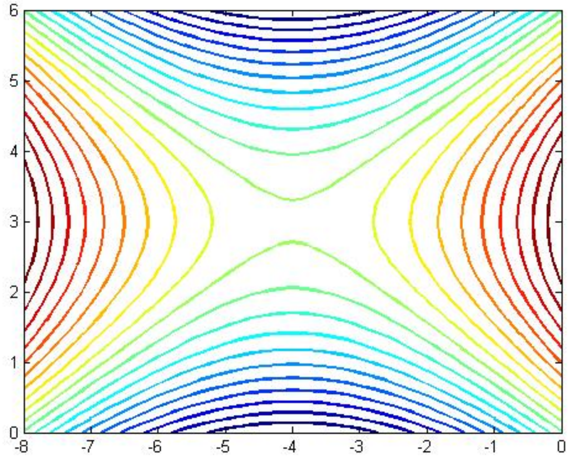
$$\nabla f(\boldsymbol{x}) = \begin{pmatrix} 8+2x_1 \\ 12-4x_2 \end{pmatrix}.$$

令$\nabla f(\boldsymbol{x}) = 0$，得唯一稳定点$\boldsymbol{x}^* = (-4, 3)^T$．又

$$G = \begin{pmatrix} 2 & \\ & -4 \end{pmatrix}.$$

故而$\boldsymbol{x}^*$ 是鞍点．

该函数的等高线图如下图所示



最优性条件

定理2 (二阶充分条件)

若$\boldsymbol{x}^*$ 为$f(\boldsymbol{x})$ 的局部极小点，且在$\boldsymbol{x}^*$ 的某邻域内$f(\boldsymbol{x})$ 具有二阶连续偏导数，

$$\nabla f(\boldsymbol{x}^*) = 0,\ G^* = \nabla_{\boldsymbol{xx}} f(\boldsymbol{x}^*)\text{正定},$$

则$\boldsymbol{x}^*$ 为无约束优化问题的严格局部极小点.

注

最优性条件

定理2 (二阶充分条件)

若$\boldsymbol{x}^*$ 为$f(\boldsymbol{x})$ 的局部极小点，且在$\boldsymbol{x}^*$ 的某邻域内$f(\boldsymbol{x})$ 具有二阶连续偏导数，

$$\nabla f(\boldsymbol{x}^*) = 0,\ \ G^* = \nabla_{\boldsymbol{x}\boldsymbol{x}} f(\boldsymbol{x}^*)\text{正定},$$

则$\boldsymbol{x}^*$ 为无约束优化问题的严格局部极小点.

注

- 对于驻点$\boldsymbol{x}^*$, 如果又有G^* 正定, 则$\boldsymbol{x}^*$ 为局部极小点;

最优性条件

定理2 (二阶充分条件)

若$\boldsymbol{x}^*$ 为$f(\boldsymbol{x})$ 的局部极小点，且在$\boldsymbol{x}^*$ 的某邻域内$f(\boldsymbol{x})$ 具有二阶连续偏导数，

$$\nabla f(\boldsymbol{x}^*) = 0,\ G^* = \nabla_{\boldsymbol{xx}} f(\boldsymbol{x}^*)\text{正定},$$

则$\boldsymbol{x}^*$ 为无约束优化问题的严格局部极小点.

注

- 对于驻点$\boldsymbol{x}^*$, 如果又有G^* 正定, 则$\boldsymbol{x}^*$ 为局部极小点;
- 对于驻点$\boldsymbol{x}^*$, 如果又有G^* 负定, 则$\boldsymbol{x}^*$ 为局部极大点;

证明： 因为G^* 正定, 故对$\forall p \in R^n$ 有

$$p^T G^* p \geq \lambda \|p\|^2,$$

其中, λ 为G^* 的最小特征值. 现将$f(\boldsymbol{x})$ 在$\boldsymbol{x}^*$ 点用Taylor 公式展开, 并注意到$\nabla f(x^*) = 0$, 有

$$f(\boldsymbol{x}) = f(\boldsymbol{x}^*) + \frac{1}{2}(x - x^*)^T G^* (x - x^*) + o(\|x - x^*\|^2).$$

于是

$$f(\boldsymbol{x}) - f(\boldsymbol{x}^*) \geq \left[\frac{1}{2}\lambda + o(1)\right] \|x - x^*\|^2.$$

当$\boldsymbol{x}$ 充分接近$\boldsymbol{x}^*$(但$\boldsymbol{x} \neq \boldsymbol{x}^*$) 时, 上式右端大于0, 故$f(\boldsymbol{x}) > f(\boldsymbol{x}^*)$, 即$\boldsymbol{x}^*$ 为$f(\boldsymbol{x})$ 的严格局部极小点. 证毕.

注: 如果G^* 负定, 则可证明$\boldsymbol{x}^*$ 为局部极大点.

最优性条件

定理3 (二阶必要条件)

若$\boldsymbol{x}^*$ 为$f(\boldsymbol{x})$ 的局部极小点，且在$\boldsymbol{x}^*$ 的某邻域内$f(\boldsymbol{x})$ 有二阶连续偏导数，则

$$\nabla f(\boldsymbol{x}^*) = 0,\ G^* = \nabla_{\boldsymbol{xx}} f(\boldsymbol{x}^*)\text{半正定}.$$

证明:任取非零向量$p \in R^n$，对于$\alpha \in R$，定义函数

$$\varphi(\alpha) = f(\boldsymbol{x}^* + \alpha p),$$

则

$$\varphi'(\alpha) = \nabla f(\boldsymbol{x}^* + \alpha p)^T p,$$
$$\varphi''(\alpha) = p^T \nabla_{\boldsymbol{xx}} f(\boldsymbol{x}^* + \alpha p) p.$$

因$\boldsymbol{x}^*$ 为$f(\boldsymbol{x})$ 的局部极小点, 所以当α 充分小时, $\varphi(\alpha) \geq \varphi(0)$, 即$\alpha = 0$ 为$\varphi(\alpha)$ 的局部极小点. 从而有

$$\varphi'(0) = 0, \ \varphi''(0) \geq 0,$$

即

$$\nabla f(\boldsymbol{x}^*)^T p = 0, \ p^T G^* p \geq 0$$

成立. 由p 的任意性知, $\nabla f(\boldsymbol{x}*) = 0$, G^* 半正定.

定理4 (凸函数的最优性条件)

设$f(\boldsymbol{x})$ 在R^n 上是凸函数且有一阶连续偏导数，则$\boldsymbol{x}^*$ 为$f(\boldsymbol{x})$ 的整体极小点的**充要条件** 是$\nabla f(\boldsymbol{x}^*) = 0$.

作业

P.128. 3.1,3.4

搜索方向

要点：沿下降最快的速度的方向搜索.

搜索方向

要点：沿下降最快的速度的方向搜索.

由Taylor 公式

$$f(\boldsymbol{x}_k + \alpha d_k) = f(\boldsymbol{x}_k) + \alpha \nabla f_k^T d_k + o(\alpha \|d_k\|).$$

搜索方向

要点：沿下降最快的速度的方向搜索.

由Taylor 公式

$$f(\boldsymbol{x}_k + \alpha d_k) = f(\boldsymbol{x}_k) + \alpha \nabla f_k^T d_k + o(\alpha \|d_k\|).$$

由于

$$\nabla f_k^T d_k = -\|\nabla f_k\| \|d_k\| \cos\theta,$$

其中, θ 为 d_k 与 $-\nabla f_k$ 的夹角.

搜索方向

要点：沿下降最快的速度的方向搜索.

由Taylor 公式

$$f(\boldsymbol{x}_k + \alpha d_k) = f(\boldsymbol{x}_k) + \alpha \nabla f_k^T d_k + o(\alpha \|d_k\|).$$

由于

$$\nabla f_k^T d_k = -\|\nabla f_k\| \|d_k\| \cos\theta,$$

其中, θ 为d_k 与$-\nabla f_k$ 的夹角. 当α, $\|d_k\|$ 固定时, $\cos\theta = 1$ 使得$\nabla f_k^T d_k$ 最小.

搜索方向

要点：沿下降最快的速度的方向搜索.

由Taylor 公式

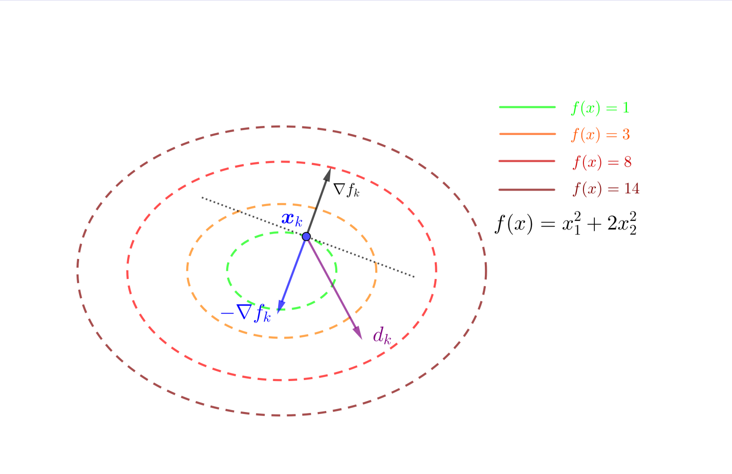
$$f(\boldsymbol{x}_k+\alpha d_k)=f(\boldsymbol{x}_k)+\alpha\nabla f_k^T d_k+o(\alpha\|d_k\|).$$

由于

$$\nabla f_k^T d_k=-\|\nabla f_k\|\|d_k\|\cos\theta,$$

其中, θ 为d_k 与$-\nabla f_k$ 的夹角. 当α, $\|d_k\|$ 固定时, $\cos\theta=1$ 使得$\nabla f_k^T d_k$ 最小. 也就是说, 当$d_k=-\nabla f_k$($\theta=0$) 时,即负梯度方向, $f(\boldsymbol{x})$ 下降速度最快.

如下图所示:


$f(x)=1$
$f(x)=3$
$f(x)=8$
$f(x)=14$
$f(x)=x_1^2+2x_2^2$
∇f_k
$\boldsymbol{x}_k$
$-\nabla f_k$
d_k

最速下降法

给定控制误差$\epsilon > 0$.

步 1 取初始点x_0, 令$k = 0$.

步 2 计算$\nabla f_k = \nabla f(\boldsymbol{x}_k)$.

步 3 若$\|\nabla f_k\| \le \epsilon$, 则令$\boldsymbol{x}^* = \boldsymbol{x}_k$; 否则, 令$d_k = -\nabla f_k$, 由一维搜索求步长$\alpha_k$, 使得

$$f(\boldsymbol{x}_k + \alpha_k d_k) = \min_{\alpha \ge 0} f(\boldsymbol{x}_k + \alpha d_k).$$

步 4 令$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k d_k$, $k = k + 1$, 转步2.

例题

例3.1

用最速下降法求解

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

设初始点为$(5,3)^T$.

例题

例3.1

用最速下降法求解

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

设初始点为$(5,3)^T$.

【解：】记$\boldsymbol{x}_0 = (5,3)^T$. 则

$$\nabla f(\boldsymbol{x}_0) = \begin{pmatrix} 2\boldsymbol{x}_0^{(1)} \\ 2\boldsymbol{x}_0^{(2)} \end{pmatrix} = \begin{pmatrix} 10 \\ 6 \end{pmatrix}.$$

例题

例3.1

用最速下降法求解

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

设初始点为$(5,3)^T$.

【解：】记$\boldsymbol{x}_0 = (5,3)^T$. 则

$$\nabla f(\boldsymbol{x}_0) = \begin{pmatrix} 2\boldsymbol{x}_0^{(1)} \\ 2\boldsymbol{x}_0^{(2)} \end{pmatrix} = \begin{pmatrix} 10 \\ 6 \end{pmatrix}.$$

$$d_0 = -\nabla f(\boldsymbol{x}_0) = \begin{pmatrix} -10 \\ -6 \end{pmatrix}, \boldsymbol{x}_0 + \alpha d_0 = \begin{pmatrix} 5 - 10\alpha \\ 3 - 6\alpha \end{pmatrix}.$$

于是

例题

$$
\begin{aligned}
\varphi(\alpha) &= f(\boldsymbol{x}_0 + \alpha d_0) \\
&= (5 - 10\alpha)^2 + (3 - 6\alpha)^2 \\
&= 136\alpha^2 - 136\alpha + 34.
\end{aligned}
$$

例题

$$\begin{aligned}\varphi(\alpha) &= f(\boldsymbol{x}_0 + \alpha d_0) \\ &= (5 - 10\alpha)^2 + (3 - 6\alpha)^2 \\ &= 136\alpha^2 - 136\alpha + 34.\end{aligned}$$

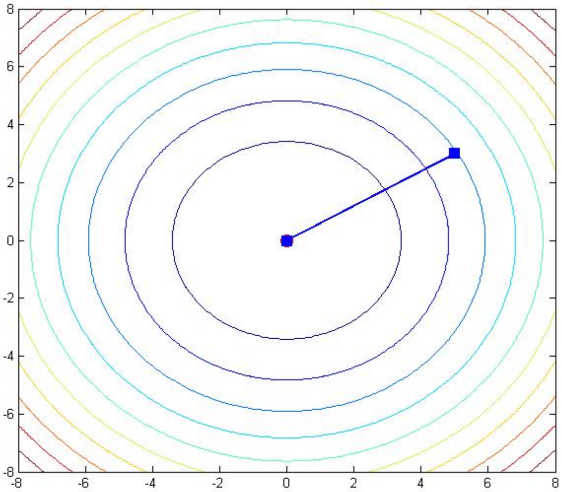
令$\varphi'(\alpha) = 0$, 得$\alpha_0 = \frac{1}{2}$. 于是

$$\boldsymbol{x}_1 = \boldsymbol{x}_0 + \alpha_0 d_0 = (0, 0)^T.$$

显然，$\boldsymbol{x}_1$ 是问题的解.

例题

如下图所示



例题

例3.2

用最速下降法求解$\min\ f(\boldsymbol{x}) = \frac{1}{2}x_1^2 + \frac{9}{2}x_2^2$, 设初始点为$(9,1)^T$.

例题

例题

【解：】

$$g(\boldsymbol{x}) = \nabla f(\boldsymbol{x}) = \begin{pmatrix} x_1 \\ 9x_2 \end{pmatrix}, G(\boldsymbol{x}) = \nabla_{\boldsymbol{xx}} f(\boldsymbol{x}) = \begin{pmatrix} 1 & 0 \\ 0 & 9 \end{pmatrix}.$$

例题

【解：】

$$g(\boldsymbol{x}) = \nabla f(\boldsymbol{x}) = \begin{pmatrix} x_1 \\ 9x_2 \end{pmatrix}, G(\boldsymbol{x}) = \nabla_{\boldsymbol{xx}} f(\boldsymbol{x}) = \begin{pmatrix} 1 & 0 \\ 0 & 9 \end{pmatrix}.$$

可以证明，如果$f(\boldsymbol{x})$ 是正定二次函数，则由精确一维搜索确定的步长α_k 为

$$\alpha_k = -\frac{g_k^T d_k}{d_k^T G d_k}.$$

例题

故对正定二次目标函数，最速下降法的迭代公式为:

$$x_{k+1} = x_k - \frac{g_k^T g_k}{g_k^T G g_k} g_k.$$

例题

故对正定二次目标函数，最速下降法的迭代公式为:

$$x_{k+1} = x_k - \frac{g_k^T g_k}{g_k^T G g_k} g_k.$$

由于$g_0 = g(x_0) = (9, 9)^T$, 所以由上式可得

$$x_1 = \begin{pmatrix} 9 \\ 1 \end{pmatrix} - \frac{(9\ 9)\begin{pmatrix} 9 \\ 9 \end{pmatrix}}{(9\ 9)\begin{pmatrix} 1 & 0 \\ 0 & 9 \end{pmatrix}\begin{pmatrix} 9 \\ 9 \end{pmatrix}} = \begin{pmatrix} 7.2 \\ -0.8 \end{pmatrix}.$$

例题

类似地计算下去，可用归纳法证明，最速下降法产生如下点列

$$x_k = \begin{pmatrix} 9 \\ (-1)^k \end{pmatrix} 0.8^k, \ \ k = 1, 2, \cdots$$

例题

类似地计算下去，可用归纳法证明，最速下降法产生如下点列

$$x_k = \begin{pmatrix} 9 \\ (-1)^k \end{pmatrix} 0.8^k, \ k = 1, 2, \cdots$$

显然$x_k \to x^* = (0, 0)^T$，且

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = 0.8.$$

例题

类似地计算下去, 可用归纳法证明, 最速下降法产生如下点列

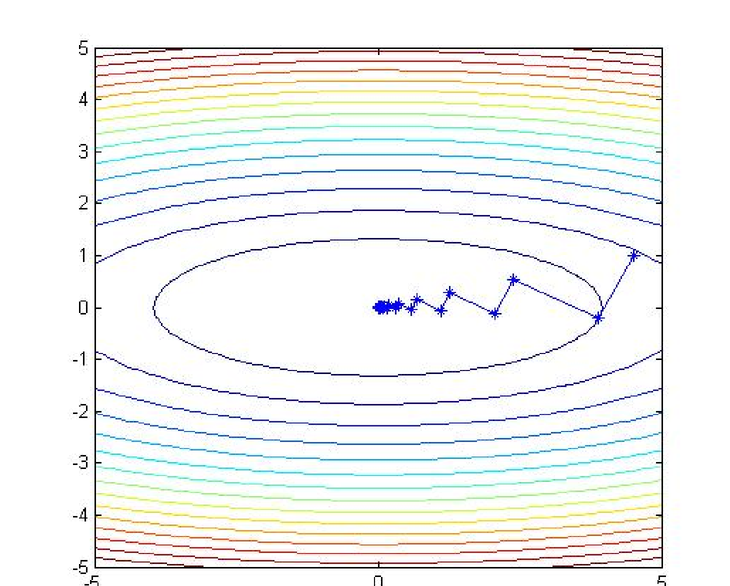
$$x_k = \begin{pmatrix} 9 \\ (-1)^k \end{pmatrix} 0.8^k, \ \ k = 1, 2, \cdots$$

显然$x_k \to x^* = (0,0)^T$, 且

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = 0.8.$$

可见对所给的目标函数, 算法是收敛的, 收敛速度是线性的.

迭代点序列如下图所示



收敛性

Zigzag(锯齿)现象

由步长α_k的定义知，

收敛性

Zigzag(锯齿)现象

由步长α_k的定义知,

$$\varphi'(\alpha_k) = \frac{d}{d\alpha}(f(x_k + \alpha_k d_k)) = \nabla f(x_k + \alpha_k d_k)^T d_k = 0.$$

收敛性

Zigzag(锯齿)现象

由步长α_k 的定义知,

$$\varphi'(\alpha_k) = \frac{d}{d\alpha}(f(x_k + \alpha_k d_k)) = \nabla f(x_k + \alpha_k d_k)^T d_k = 0.$$

在最速下降法中, $d_k = -\nabla f_k$,

$$d_{k+1} = -\nabla f(x_{k+1}) = -\nabla f(x_k + \alpha_k d_k).$$

收敛性

Zigzag(锯齿)现象

由步长α_k 的定义知,

$$\varphi'(\alpha_k) = \frac{d}{d\alpha}(f(x_k + \alpha_k d_k)) = \nabla f(x_k + \alpha_k d_k)^T d_k = 0.$$

在最速下降法中, $d_k = -\nabla f_k$,

$$d_{k+1} = -\nabla f(x_{k+1}) = -\nabla f(x_k + \alpha_k d_k).$$

则$d_{k+1}^T d_k = 0$. 也就是说, 最速下降法相邻的两个搜索方向互相垂直,

收敛性

Zigzag(锯齿)现象

由步长α_k 的定义知,

$$\varphi'(\alpha_k) = \frac{d}{d\alpha}(f(x_k + \alpha_k d_k)) = \nabla f(x_k + \alpha_k d_k)^T d_k = 0.$$

在最速下降法中, $d_k = -\nabla f_k$,

$$d_{k+1} = -\nabla f(x_{k+1}) = -\nabla f(x_k + \alpha_k d_k).$$

则$d_{k+1}^T d_k = 0$. 也就是说, 最速下降法相邻的两个搜索方向互相垂直,
于是整个迭代序列产生了**Zigzag(锯齿)现象**.

收敛性

定理5 (整体收敛性或全局收敛性)

设$f(x)$ 具有一阶连续函数偏导数，给定$x_0 \in R^n$，假定水平集$L = \{x \in R^n \mid f(x) \leq f(x_0)\}$ 有界，令$\{x_k\}$ 为由最速下降法产生的点列，则或者

(i) 对某个k_0，$\nabla f(x_{k_0}) = 0$；或者

(ii) 当$k \to \infty$ 时，$\nabla f_k \to 0$.

证明: 假设$\forall k$, $\nabla f_k \neq 0$. 因$\{f(x_k)\}$ 单调下降且水平集L 有界, 故而$\lim_{k\to+\infty} f(x_k)$ 存在. 所以有

$$f_k - f_{k+1} \to 0,\ k \to +\infty. \tag{1}$$

利用反证法. 假设$\nabla f_k \to 0$ 不成立. 则存在$\epsilon_0 > 0$ 及$\{1, 2, \cdots\}$ 的子列$\mathcal{K}$ 使得

$$\|\nabla f_k\| \geq \epsilon_0 > 0,\ \forall k \in \mathcal{K}.$$

对于这样的k, 有

$$-\frac{\nabla f_k^T d_k}{\|d_k\|} = \|\nabla f_k\| \geq \epsilon_0,\ \forall k \in \mathcal{K}.$$

于是, 由Taylor 公式,

$$
\begin{aligned}
f(\boldsymbol{x}_k+\alpha d_k) &= f(\boldsymbol{x}_k)+\alpha\nabla f(\xi_k)^T d_k \\
&= f(\boldsymbol{x}_k)+\alpha\nabla f_k^T d_k+\alpha[\nabla f(\xi_k)-\nabla f_k]^T d_k \\
&\le f(\boldsymbol{x}_k)+\alpha\|d_k\|[\nabla f_k^T d_k/\|d_k\|+\|\nabla f(\xi_k)-\nabla f_k\|], (2)
\end{aligned}
$$

其中, ξ_k 在$\boldsymbol{x}_k$ 与$\boldsymbol{x}_k+\alpha d_k$ 的连线线段上.
由于, $\nabla f(\boldsymbol{x})$ 连续且L 有界, 所以$\nabla f(\boldsymbol{x})$ 在L 上一致连续, 故存在$\bar{\alpha}>0$, 使当$0\le\alpha\|d_k\|\le\bar{\alpha}$ 时,

$$
\|\nabla f(\xi_k)-\nabla f_k\|\le\frac{1}{2}\epsilon_0,
$$

对所有k 成立.

在式(2) 中取$\alpha = \bar{\alpha}/\|d_k\|$, 则有

$$
\begin{aligned}
& f\left(\boldsymbol{x}_k + \frac{\bar{\alpha}}{\|d_k\|} d_k\right) \\
& \quad \le f(\boldsymbol{x}_k) + \bar{\alpha}\left[\frac{\nabla f_k^T d_k}{\|d_k\|} + \|\nabla f(\xi_k) - \nabla f_k\|\right] \\
& \quad \le f(\boldsymbol{x}_k) + \bar{\alpha}\left[-\epsilon_0 + \frac{1}{2}\epsilon_0\right] \\
& \quad \le f(\boldsymbol{x}_k) - \frac{1}{2}\epsilon_0.
\end{aligned}
$$

从而有

$$
\begin{aligned}
f(\boldsymbol{x}_{k+1}) &= \min_{\alpha_0>0} f(\boldsymbol{x}_k + \alpha d_k) \\
&\le f\left(\boldsymbol{x}_k + \frac{\bar{\alpha}}{\|d_k\|} d_k\right) \le f(\boldsymbol{x}_k) - \frac{1}{2}\bar{\alpha}\epsilon_0,
\end{aligned}
$$

对于无穷多个k 成立, 这与式(1) 矛盾. 故而$\nabla f_k \to 0$.

用于二次函数时的收敛速度

定理6 (收敛速度)

设$f(x)=\frac{1}{2}x^TGx$, 其中, G 为正定矩阵. 用λ_1, λ_n 表示G 的最小与最大特征值, 则由最速下降法产生的点列$\{x_k\}$ 满足

$$f(x_{k+1})\le\left(\frac{\lambda_n-\lambda_1}{\lambda_n+\lambda_1}\right)^2 f(x_k),\ \ k=0,1,2,\cdots,$$

$$\|x_k\|\le\sqrt{\frac{\lambda_n}{\lambda_1}}\left(\frac{\lambda_n-\lambda_1}{\lambda_n+\lambda_1}\right)^k\|x_0\|,\ \ k=0,1,2,\cdots$$

用于二次函数时的收敛速度

定理6 (收敛速度)

设$f(x)=\frac{1}{2}x^TGx$, 其中, G 为正定矩阵. 用λ_1, λ_n 表示G 的最小与最大特征值, 则由最速下降法产生的点列$\{x_k\}$ 满足

$$f(x_{k+1})\le\left(\frac{\lambda_n-\lambda_1}{\lambda_n+\lambda_1}\right)^2 f(x_k),\ \ k=0,1,2,\cdots,$$

$$\|x_k\|\le\sqrt{\frac{\lambda_n}{\lambda_1}}\left(\frac{\lambda_n-\lambda_1}{\lambda_n+\lambda_1}\right)^k\|x_0\|,\ \ k=0,1,2,\cdots$$

即: 线性收敛.

定理6 说明, 对于二次函数, 最速下降法至少是线性收敛的, 其收敛比$\beta \leq \frac{\lambda_n-\lambda_1}{\lambda_n+\lambda_1}$. 所以当$G$ 的特征值比较分散, 即$\lambda_n \gg \lambda_1$ 时, 收敛比接近1, 收敛速度很慢; 当G 的特征值比较集中, 即$\lambda_n \approx \lambda_1$ 时, 收敛比接近于0, 从而收敛速度接近超线性收敛.

例题

例3.3

用最速下降法求$Rosenbrock$ 函数

$$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

的最小值，初始点为$x_0 = (0, 0)^T$.

例题

例3.3

用最速下降法求*Rosenbrock* 函数

$$f(x)=100(x_2-x_1^2)^2+(1-x_1)^2$$

的最小值，初始点为$x_0=(0,0)^T$.

【解：】

$$\nabla f(x)=\begin{pmatrix}-400(x_2-x_1^2)x_1-2(1-x_1)\\200(x_2-x_1^2)\end{pmatrix},$$

$f_0=f(0,0)=1,\ \nabla f_0=(-2,0)^T.$

例题

$d_0 = (2,0)^T$. 对于任意的$\alpha > 0$,

$$f(x_0 + \alpha d_0) = f(2\alpha, 0) = 1600\alpha^4 + 4\alpha^2 - 4\alpha + 1.$$

例题

$d_0 = (2, 0)^T$. 对于任意的$\alpha > 0$,

$$f(x_0 + \alpha d_0) = f(2\alpha, 0) = 1600\alpha^4 + 4\alpha^2 - 4\alpha + 1.$$

令$f'(\alpha) = 0$, 得

$$6400\alpha^3 + 8\alpha - 4 = 0.$$

其近似解为$\alpha_0 \approx 0.0806$.

$$x_1 = x_0 + \alpha_0 d_0 = (0.1612, 0)^T.$$

……

例3.4 (思考题(固定步长梯度法))

求严格凸二次函数

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

的最小值点. 若以$d_k = -\nabla f_k$ 为搜索方向, 且对所有k, $\alpha_k \equiv \alpha \in (0, +\infty)$, 相应的迭代公式为

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha\nabla f_k.$$

1. 找几个正定二次函数, 看能否选一个合适的α 使得$\{x_k\}$ 收敛到f 的最小值点?
2. 和最速下降法比较一下.
3. 当α 满足什么条件时, $\{x_k\}$ 收敛到f 的最小值点?

P.129. 3.5,3.6.

引言

最速下降法的优点

引言

最速下降法的优点

- 原理简单, 容易实现;

引言

最速下降法的优点

- 原理简单，容易实现;
- 每次迭代的计算量小.

引言

最速下降法的优点

- 原理简单, 容易实现;
- 每次迭代的计算量小.

缺点

引言

最速下降法的优点

- 原理简单，容易实现;
- 每次迭代的计算量小.

缺点

- 收敛速度慢(有锯齿现象).

Newton 法：启发1

最速下降法用一次迭代就得到例3.1 的解，但即使是像例3.2 的二次函数，最速下降法的收敛速度也是非常慢.
能不能设计一个算法，使得它用一次迭代就能得到(凸)二次函数的最小值呢？

Newton 法：启发1

目标：一步求凸二次函数的最小值

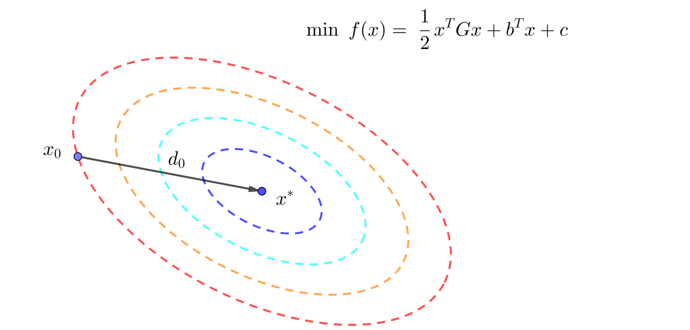
最小化凸二次函数：

$$\min \ f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T G\boldsymbol{x} + b^T\boldsymbol{x} + c,$$

其中G 是正定矩阵. 任取一个初始点$\boldsymbol{x}_0$, 找一个搜索方向d_0, 使得我们可以仅用一次线性搜索就得到问题的解（最好连线性搜索都不用）.

Newton 法：启发1

如图所示



Newton 法：启发1

二次函数 $f(x)$ 的梯度和Hesse 矩阵分别是

$$\nabla f(\boldsymbol{x}) = G\boldsymbol{x} + b,\ \nabla^2 f(\boldsymbol{x}) = G.$$

Newton 法：启发1

二次函数$f(x)$ 的梯度和Hesse 矩阵分别是

$$\nabla f(\boldsymbol{x}) = G\boldsymbol{x} + b,\ \nabla^2 f(\boldsymbol{x}) = G.$$

要使

$$\nabla f(\boldsymbol{x}_0 + d_0) = 0,$$

则必须有

$$G(\boldsymbol{x}_0 + d_0) + b = 0 \Rightarrow d_0 = -G^{-1}(G\boldsymbol{x}_0 + b).$$

Newton 法：启发1

二次函数$f(x)$的梯度和Hesse 矩阵分别是

$$\nabla f(\boldsymbol{x}) = G\boldsymbol{x} + b, \ \nabla^2 f(\boldsymbol{x}) = G.$$

要使

$$\nabla f(\boldsymbol{x}_0 + d_0) = 0,$$

则必须有

$$G(\boldsymbol{x}_0 + d_0) + b = 0 \Rightarrow d_0 = -G^{-1}(G\boldsymbol{x}_0 + b).$$

也就是

$$d_0 = -(\nabla^2 f(\boldsymbol{x}_0))^{-1}\nabla f(\boldsymbol{x}_0).$$

此为Newton 方向.

Newton 法：动机2

最速下降法只用到了目标函数的一阶导数（梯度）信息.

Newton 法：动机2

最速下降法只用到了目标函数的一阶导数（梯度）信息.

$$f(\boldsymbol{x}) \approx f_k + \nabla f_k^T(\boldsymbol{x} - \boldsymbol{x}_k) \overset{d=\boldsymbol{x}-\boldsymbol{x}_k}{=} f_k + \nabla f_k^T d$$

Newton 法：动机2

最速下降法只用到了目标函数的一阶导数（梯度）信息.

$$f(\boldsymbol{x}) \approx f_k + \nabla f_k^T(\boldsymbol{x} - \boldsymbol{x}_k) \overset{d=\boldsymbol{x}-\boldsymbol{x}_k}{=} f_k + \nabla f_k^T d$$

启发

Newton 法：动机2

最速下降法只用到了目标函数的一阶导数（梯度）信息.

$$f(\boldsymbol{x}) \approx f_k + \nabla f_k^T(\boldsymbol{x} - \boldsymbol{x}_k) \overset{d=\boldsymbol{x}-\boldsymbol{x}_k}{=} f_k + \nabla f_k^T d$$

启发

- 如果在迭代方法中引入高阶导数，其效率可能会提高.

Newton 法：动机2

最速下降法只用到了目标函数的一阶导数（梯度）信息.

$$f(\boldsymbol{x}) \approx f_k + \nabla f_k^T(\boldsymbol{x} - \boldsymbol{x}_k) \overset{d=\boldsymbol{x}-\boldsymbol{x}_k}{=} f_k + \nabla f_k^T d$$

启发

- 如果在迭代方法中引入高阶导数，其效率可能会提高.

Newton 法

Newton 法：动机2

最速下降法只用到了目标函数的一阶导数（梯度）信息.

$$f(\boldsymbol{x}) \approx f_k + \nabla f_k^T(\boldsymbol{x} - \boldsymbol{x}_k) \overset{d=\boldsymbol{x}-\boldsymbol{x}_k}{=} f_k + \nabla f_k^T d$$

启发

- 如果在迭代方法中引入高阶导数，其效率可能会提高.

Newton 法

- 同时使用一阶导数和二阶导数来确定搜索方向;

Newton 法：动机2

最速下降法只用到了目标函数的一阶导数（梯度）信息.

$$f(\boldsymbol{x}) \approx f_k + \nabla f_k^T(\boldsymbol{x} - \boldsymbol{x}_k) \overset{d=\boldsymbol{x}-\boldsymbol{x}_k}{=} f_k + \nabla f_k^T d$$

启发

- 如果在迭代方法中引入高阶导数，其效率可能会提高.

Newton 法

- 同时使用一阶导数和二阶导数来确定搜索方向;
- 当初始点接近目标函数的极小点时, Newton 法的效率要远高于最速下降法.

Newton 法：启发2

目标函数的二次近似

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

其中, G_k 为Hesse 矩阵.

Newton 法：启发2

目标函数的二次近似

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

其中, G_k 为Hesse 矩阵.

如下图所示

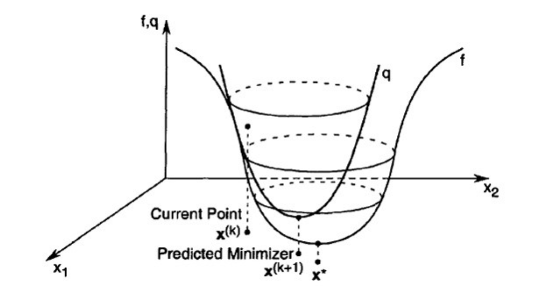


图: 目标函数的二次型近似函数

Newton 法：启发2

若G_k 正定，则$q_k(\boldsymbol{x})$ 有唯一的极小点．由一阶必要条件知，$\boldsymbol{x}_{k+1}$ 满足

$$\nabla q_k(\boldsymbol{x}_{k+1}) = 0.$$

也就是

Newton 法：启发2

若G_k 正定，则$q_k(\boldsymbol{x})$ 有唯一的极小点. 由一阶必要条件知，$\boldsymbol{x}_{k+1}$ 满足

$$\nabla q_k(\boldsymbol{x}_{k+1}) = 0.$$

也就是

$$G_k(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k) + \nabla f_k = 0.$$

Newton 法：启发2

若G_k 正定，则$q_k(\boldsymbol{x})$ 有唯一的极小点．由一阶必要条件知，$\boldsymbol{x}_{k+1}$ 满足

$$\nabla q_k(\boldsymbol{x}_{k+1}) = 0.$$

也就是

$$G_k(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k) + \nabla f_k = 0.$$

解得

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - G_k^{-1}\nabla f_k.$$

Newton 法：启发2

若G_k 正定，则$q_k(\boldsymbol{x})$ 有唯一的极小点．由一阶必要条件知，$\boldsymbol{x}_{k+1}$ 满足

$$\nabla q_k(\boldsymbol{x}_{k+1}) = 0.$$

也就是

$$G_k(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k) + \nabla f_k = 0.$$

解得

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - G_k^{-1}\nabla f_k.$$

这就是Newton 迭代公式.

算法

Newton 法

给定误差控制参数$\epsilon > 0$.

步 1　取初始点x_0, 令$k = 0$.

步 2　计算∇f_k.

步 3　若$\|\nabla f_k\| \leq \epsilon$, 则令$\boldsymbol{x}^* = \boldsymbol{x}_k$, 停; 否则计算$G_k$, 求解

$$G_k d_k = -\nabla f_k$$

得d_k.

步 4　令$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + d_k$, $k = k + 1$, 转步2.

例题

例4.1

用 *Newton* 法求解

$$\min \ f(\boldsymbol{x}) = \frac{1}{2}x_1^2 + \frac{9}{2}x_2^2$$

例题

例4.1

用 *Newton* 法求解

$$\min \ f(\boldsymbol{x}) = \frac{1}{2}x_1^2 + \frac{9}{2}x_2^2$$

【解：】本题可取任意初始点，这里取$\boldsymbol{x}_0 = (9,9)^T$. 由$\nabla f(\boldsymbol{x}) = \begin{pmatrix} x_1 \\ 9x_2 \end{pmatrix}$, $G(\boldsymbol{x}) = \begin{pmatrix} 1 & 0 \\ 0 & 9 \end{pmatrix}$ 有

例题

例4.1

用 *Newton* 法求解

$$\min \ f(\boldsymbol{x}) = \frac{1}{2}x_1^2 + \frac{9}{2}x_2^2$$

【解：】本题可取任意初始点，这里取$\boldsymbol{x}_0 = (9,9)^T$. 由$\nabla f(\boldsymbol{x}) = \begin{pmatrix} x_1 \\ 9x_2 \end{pmatrix}$, $G(\boldsymbol{x}) = \begin{pmatrix} 1 & 0 \\ 0 & 9 \end{pmatrix}$ 有

$$\begin{aligned}
\boldsymbol{x}_1 &= \boldsymbol{x}_0 - G_0^{-1}\nabla f_0 \\
&= \begin{pmatrix} 9 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 9 \end{pmatrix}^{-1} \begin{pmatrix} 9 \\ 9 \end{pmatrix} \\
&= \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \boldsymbol{x}^*.
\end{aligned}$$

例题

例4.2

用$Newton$法求$Rosenbrock$ 函数

$$f(\boldsymbol{x}) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

的最小值，初始点为$\boldsymbol{x}_0 = (0, 0)^T$.

例题

例4.2

用*Newton*法求*Rosenbrock* 函数

$$f(\boldsymbol{x}) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

的最小值，初始点为$\boldsymbol{x}_0 = (0, 0)^T$.

【解：】该函数的梯度为

$$\nabla f = \begin{pmatrix} -400\left(x_2 - x_1^2\right)x_1 - 2\left(1 - x_1\right) \\ 200\left(x_2 - x_1^2\right) \end{pmatrix},$$

例题

例4.2

用*Newton*法求*Rosenbrock* 函数

$$f(\boldsymbol{x}) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

的最小值，初始点为$\boldsymbol{x}_0 = (0,0)^T$.

【解：】该函数的梯度为

$$\nabla f = \begin{pmatrix} -400\left(x_2 - x_1^2\right)x_1 - 2\left(1 - x_1\right) \\ 200\left(x_2 - x_1^2\right) \end{pmatrix},$$

Hesse 矩阵为

$$G = \begin{pmatrix} 1200x_1^2 - 400x_2 + 2 & -400x_1 \\ -400x_1 & 200 \end{pmatrix}.$$

例题

在初始点处

$$\nabla f_0 = \begin{pmatrix} -2 \\ 0 \end{pmatrix}, \ G_0 = \begin{pmatrix} 2 & 0 \\ 0 & 200 \end{pmatrix}.$$

例题

在初始点处

$$\nabla f_0 = \begin{pmatrix} -2 \\ 0 \end{pmatrix}, \ G_0 = \begin{pmatrix} 2 & 0 \\ 0 & 200 \end{pmatrix}.$$

则

$$d_0 = -G_0^{-1} \nabla f_0 = -\begin{pmatrix} 1/2 & 0 \\ 0 & 1/200 \end{pmatrix} \begin{pmatrix} -2 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

例题

在初始点处

$$\nabla f_0 = \begin{pmatrix} -2 \\ 0 \end{pmatrix}, \; G_0 = \begin{pmatrix} 2 & 0 \\ 0 & 200 \end{pmatrix}.$$

则

$$d_0 = -G_0^{-1}\nabla f_0 = -\begin{pmatrix} 1/2 & 0 \\ 0 & 1/200 \end{pmatrix}\begin{pmatrix} -2 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

于是

$$\boldsymbol{x}_1 = \boldsymbol{x}_0 + d_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

例题

在初始点处

$$\nabla f_0 = \begin{pmatrix} -2 \\ 0 \end{pmatrix},\ G_0 = \begin{pmatrix} 2 & 0 \\ 0 & 200 \end{pmatrix}.$$

则

$$d_0 = -G_0^{-1}\nabla f_0 = -\begin{pmatrix} 1/2 & 0 \\ 0 & 1/200 \end{pmatrix}\begin{pmatrix} -2 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

于是

$$\boldsymbol{x}_1 = \boldsymbol{x}_0 + d_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

因$\nabla f_1 = (400, -200)^T \neq 0$, 继续迭代.

例题

在x_1处

$$\nabla f_1 = \begin{pmatrix} 400 \\ -200 \end{pmatrix},\ G_1 = \begin{pmatrix} 1202 & -400 \\ -400 & 200 \end{pmatrix}.$$

例题

在x_1处

$$\nabla f_1 = \begin{pmatrix} 400 \\ -200 \end{pmatrix}, \quad G_1 = \begin{pmatrix} 1202 & -400 \\ -400 & 200 \end{pmatrix}.$$

则

$$d_1 = -G_1^{-1}\nabla f_1 = -\frac{1}{80400}\begin{pmatrix} 200 & 400 \\ 400 & 1202 \end{pmatrix}\begin{pmatrix} 400 \\ -200 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

例题

在x_1处

$$\nabla f_1 = \begin{pmatrix} 400 \\ -200 \end{pmatrix}, \; G_1 = \begin{pmatrix} 1202 & -400 \\ -400 & 200 \end{pmatrix}.$$

则

$$d_1 = -G_1^{-1} \nabla f_1 = -\frac{1}{80400} \begin{pmatrix} 200 & 400 \\ 400 & 1202 \end{pmatrix} \begin{pmatrix} 400 \\ -200 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

于是

$$\boldsymbol{x}_2 = \boldsymbol{x}_1 + d_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

例题

在x_1处

$$\nabla f_1 = \begin{pmatrix} 400 \\ -200 \end{pmatrix}, \quad G_1 = \begin{pmatrix} 1202 & -400 \\ -400 & 200 \end{pmatrix}.$$

则

$$d_1 = -G_1^{-1}\nabla f_1 = -\frac{1}{80400}\begin{pmatrix} 200 & 400 \\ 400 & 1202 \end{pmatrix}\begin{pmatrix} 400 \\ -200 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

于是

$$\boldsymbol{x}_2 = \boldsymbol{x}_1 + d_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

由于$\nabla f_2 = (0,0)^T$, 算法终止.

例题

在x_1处

$$\nabla f_1 = \begin{pmatrix} 400 \\ -200 \end{pmatrix}, \ G_1 = \begin{pmatrix} 1202 & -400 \\ -400 & 200 \end{pmatrix}.$$

则

$$d_1 = -G_1^{-1}\nabla f_1 = -\frac{1}{80400}\begin{pmatrix} 200 & 400 \\ 400 & 1202 \end{pmatrix}\begin{pmatrix} 400 \\ -200 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

于是

$$\boldsymbol{x}_2 = \boldsymbol{x}_1 + d_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

由于$\nabla f_2 = (0,0)^T$, 算法终止.

数值实验: Rosenbrock

例题

例4.3

问题

$$\min \ f(\boldsymbol{x}) = (x_1-2)^4 + (x_1-2)^2 x_2^2 + (x_2+1)^2$$

具有极小点$(2,-1)^T$. 若取初始点为$(1,1)^T$, 用*Newton* 法求解此问题.

例题

例4.3

问题

$$\min\ f(\boldsymbol{x}) = (x_1-2)^4 + (x_1-2)^2 x_2^2 + (x_2+1)^2$$

具有极小点$(2,-1)^T$. 若取初始点为$(1,1)^T$, 用*Newton* 法求解此问题.

【解：】计算

$$\nabla f(\boldsymbol{x}) = \begin{pmatrix} 4(x_1-2)^3 + 2x_2^2(x_1-2) \\ 2x_2 + 2x_2(x_1-2)^2 + 2 \end{pmatrix}$$

例题

例4.3

问题

$$\min \ f(\boldsymbol{x}) = (x_1-2)^4 + (x_1-2)^2 x_2^2 + (x_2+1)^2$$

具有极小点$(2,-1)^T$. 若取初始点为$(1,1)^T$, 用*Newton* 法求解此问题.

【解：】计算

$$\nabla f(\boldsymbol{x}) = \begin{pmatrix} 4(x_1-2)^3 + 2x_2^2(x_1-2) \\ 2x_2 + 2x_2(x_1-2)^2 + 2 \end{pmatrix}$$

$$G(\boldsymbol{x}) = \begin{pmatrix} 12(x_1-2)^2 + 2x_2^2 & 4x_2(x_1-2) \\ 4x_2(x_1-2) & 2(x_1-2)^2 + 2 \end{pmatrix}$$

例题

在初始点$(1,1)^T$ 处,

$$f_0 = 6,\ \nabla f_0 = \begin{pmatrix} -6 \\ 6 \end{pmatrix},\ G_0 = \begin{pmatrix} 14 & -4 \\ -4 & 4 \end{pmatrix}.$$

例题

在初始点$(1,1)^T$ 处,

$$f_0=6,\ \nabla f_0=\begin{pmatrix}-6\\6\end{pmatrix},\ G_0=\begin{pmatrix}14&-4\\-4&4\end{pmatrix}.$$

由Newton 迭代公式知

$$\boldsymbol{x}_1=\boldsymbol{x}_0-G_0^{-1}\nabla f_0=\begin{pmatrix}1\\-0.5\end{pmatrix}.$$

例题

在初始点$(1,1)^T$ 处,

$$f_0=6,\ \nabla f_0=\begin{pmatrix}-6\\6\end{pmatrix},\ G_0=\begin{pmatrix}14&-4\\-4&4\end{pmatrix}.$$

由Newton 迭代公式知

$$\boldsymbol{x}_1=\boldsymbol{x}_0-G_0^{-1}\nabla f_0=\begin{pmatrix}1\\-0.5\end{pmatrix}.$$

此时$\nabla f_1=\begin{pmatrix}-4.5\\0\end{pmatrix}$.

例题

继续迭代，得如下迭代过程

k	$\boldsymbol{x}_k$	$\Vert\boldsymbol{x}_k - \boldsymbol{x}^*\Vert$
0	$(1, 1)^T$	2.0
1	$(1, -0.5)^T$	1.49
2	$(1.39130, -0.69565)^T$	5.23×10^{-1}
3	$(1.74594, -0.94880)^T$	1.01×10^{-1}
4	$(1.98628, -1.04821)^T$	2.55×10^{-3}
5	$(1.99873, -1.00017)^T$	3.32×10^{-6}
6	$(1.9999996, -1.0000016)^T$	2.81×10^{-12}

例题

继续迭代，得如下迭代过程

k	$\boldsymbol{x}_k$	$\|\boldsymbol{x}_k - \boldsymbol{x}^*\|$
0	$(1, 1)^T$	2.0
1	$(1, -0.5)^T$	1.49
2	$(1.39130, -0.69565)^T$	5.23×10^{-1}
3	$(1.74594, -0.94880)^T$	1.01×10^{-1}
4	$(1.98628, -1.04821)^T$	2.55×10^{-3}
5	$(1.99873, -1.00017)^T$	3.32×10^{-6}
6	$(1.9999996, -1.0000016)^T$	2.81×10^{-12}

数值实验：Prob2

收敛性

- 对于二次函数，Newton 法只需要一次迭代就可以得到极小点.

收敛性

- 对于二次函数，Newton 法只需要一次迭代就可以得到极小点.
- 对于一般函数

定理7 (Newton 法的收敛性)

设$f(\boldsymbol{x})$ 是某一开域内的三阶连续可微函数，且它在该开域内有极小点$\boldsymbol{x}^*$，设$G^* = G(\boldsymbol{x}^*)$ 正定，则当$\boldsymbol{x}_0$ 与$\boldsymbol{x}^*$ 充分接近时，对一切k，*Newton* 法有定义，且当$\{\boldsymbol{x}_k\}$ 为无穷点列时，$\{\boldsymbol{x}_k\}$ 二阶收敛于$\boldsymbol{x}^*$，即$\|\boldsymbol{x}_k - \boldsymbol{x}^*\| \to 0$ 且

$$\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\| = O(\|\boldsymbol{x} - \boldsymbol{x}^*\|^2).$$

收敛性

证明: 为简便起见, 令$h_k = \boldsymbol{x}_k - \boldsymbol{x}^*$. 因$f(\boldsymbol{x})$ 是三阶连续可微的, 故而由Taylor 公式有

$$\nabla f(\boldsymbol{x}_k + h) = \nabla f_k + G_k h + O(\|h\|^2).$$

取$h = -h_k$, 即$\boldsymbol{x}_k + h = \boldsymbol{x}^*$, 得

$$0 = \nabla f(\boldsymbol{x}^*) = \nabla f_k - G_k h_k + O(\|h_k\|^2). \tag{3}$$

由于G^* 正定, 所以存在$\boldsymbol{x}^*$ 的一个β 邻域

$$N(\boldsymbol{x}^*) = \{\boldsymbol{x} \mid \|\boldsymbol{x} - \boldsymbol{x}^*\| \leq \beta\},$$

使得$\forall \boldsymbol{x} \in N(\boldsymbol{x}^*)$, $G(\boldsymbol{x})$ 正定, 且$\|G(\boldsymbol{x})^{-1}\|$ 有上界.

收敛性

于是，若$\boldsymbol{x}_k \in N(\boldsymbol{x}^*)$，对(3) 两边同乘以$G_k^{-1}$，再由$h_{k+1}$ 的定义知

$$0 = -G_k^{-1}\nabla f_k - h_k + O(\|h_k\|^2) = -h_{k+1} + O(\|h_k\|^2). \quad (4)$$

由此，由$O(\cdot)$ 的定义，存在常数γ 使得

$$\|h_{k+1}\| \leq \gamma\|h_k\|^2$$

成立. 若取β 充分小使之满足$\beta\gamma < 1$，则有

$$\|h_{k+1}\| \leq \gamma\|h_k\|^2 \leq \beta\gamma\|h_k\| < \|h_k\|.$$

因此$\boldsymbol{x}_k \in N(\boldsymbol{x}^*)$.

收敛性

由归纳法，如果$\boldsymbol{x}_0 \in N(\boldsymbol{x}^*)$，则对所有的$k$，Newton 法有定义,且

$$\|h_k\| \leq (\gamma\beta)^k \|h_0\| \to 0(k \to) + \infty.$$

而由式(4)得$\|h_{k+1}\| = O(\|h_k\|^2)$.

例题

例4.4

考虑问题

$$\min f(\boldsymbol{x}) = 3x_1^2 + 3x_2^2 - x_1^2 x_2.$$

分别从初始点$(1.5, 1.5)^T$,$(-2, 4)^T$,$(0, 3)^T$ 出发，用*Newton* 法求解该问题.

例题

例4.4

考虑问题

$$\min f(\boldsymbol{x}) = 3x_1^2 + 3x_2^2 - x_1^2 x_2.$$

分别从初始点$(1.5, 1.5)^T$,$(-2, 4)^T$,$(0, 3)^T$ 出发，用*Newton* 法求解该问题.

【解：】$f(x)$ 的一，二阶导数分别为

$$g(\boldsymbol{x}) = \left(6x_1 - 2x_1x_2, 6x_2 - x_1^2\right)^{\mathrm{T}},$$

$$G(x) = \begin{bmatrix} 6 - 2x_2 & -2x_1 \\ -2x_1 & 6 \end{bmatrix}$$

例题

$f(\boldsymbol{x})$ 有三个稳定点（驻点）：极小点$(0,0)^T$, 鞍点$(3\sqrt{2},3)^T$ 和$(-3\sqrt{2},3)^T$.

例题

$f(x)$ 有三个稳定点（驻点）：极小点$(0,0)^T$, 鞍点$(3\sqrt{2},3)^T$和$(-3\sqrt{2},3)^T$.

在这三个点处的Hesse矩阵分别为

$$G_1=\begin{bmatrix}6&0\\0&6\end{bmatrix},\quad G_2=\begin{bmatrix}0&-6\sqrt{2}\\-6\sqrt{2}&6\end{bmatrix}$$

$$G_3=\begin{bmatrix}0&6\sqrt{2}\\6\sqrt{2}&6\end{bmatrix}$$

例题

$f(x)$ 有三个稳定点（驻点）：极小点$(0,0)^T$, 鞍点$(3\sqrt{2},3)^T$和$(-3\sqrt{2},3)^T$.
在这三个点处的Hesse矩阵分别为

$$G_1 = \begin{bmatrix} 6 & 0 \\ 0 & 6 \end{bmatrix}, \quad G_2 = \begin{bmatrix} 0 & -6\sqrt{2} \\ -6\sqrt{2} & 6 \end{bmatrix}$$
$$G_3 = \begin{bmatrix} 0 & 6\sqrt{2} \\ 6\sqrt{2} & 6 \end{bmatrix}$$

- 初始点$x_0 = (1.5, 1.5)^T$, 收敛到极小点$(0,0)^T$.

例题

$f(\boldsymbol{x})$ 有三个稳定点（驻点）：极小点$(0,0)^T$, 鞍点$(3\sqrt{2},3)^T$和$(-3\sqrt{2},3)^T$.
在这三个点处的Hesse矩阵分别为

$$G_1 = \begin{bmatrix} 6 & 0 \\ 0 & 6 \end{bmatrix}, \quad G_2 = \begin{bmatrix} 0 & -6\sqrt{2} \\ -6\sqrt{2} & 6 \end{bmatrix}$$
$$G_3 = \begin{bmatrix} 0 & 6\sqrt{2} \\ 6\sqrt{2} & 6 \end{bmatrix}$$

- 初始点$\boldsymbol{x}_0 = (1.5, 1.5)^T$, 收敛到极小点$(0,0)^T$.
- 初始点$\boldsymbol{x}_0 = (-2, 4)^T$, 收敛到鞍点$(-3\sqrt{2},3)^T$.

例题

$f(\boldsymbol{x})$ 有三个稳定点（驻点）：极小点$(0,0)^T$, 鞍点$(3\sqrt{2},3)^T$和$(-3\sqrt{2},3)^T$.
在这三个点处的Hesse矩阵分别为

$$G_1 = \begin{bmatrix} 6 & 0 \\ 0 & 6 \end{bmatrix}, \quad G_2 = \begin{bmatrix} 0 & -6\sqrt{2} \\ -6\sqrt{2} & 6 \end{bmatrix}$$
$$G_3 = \begin{bmatrix} 0 & 6\sqrt{2} \\ 6\sqrt{2} & 6 \end{bmatrix}$$

- 初始点$\boldsymbol{x}_0 = (1.5, 1.5)^T$, 收敛到极小点$(0,0)^T$.
- 初始点$\boldsymbol{x}_0 = (-2, 4)^T$, 收敛到鞍点$(-3\sqrt{2},3)^T$.
- 初始点$\boldsymbol{x}_0 = (0, 3)^T$, 这时$G(x_0)$ 奇异, 方法失败.

例题

$f(x)$ 有三个稳定点（驻点）：极小点$(0,0)^T$，鞍点$(3\sqrt{2},3)^T$和$(-3\sqrt{2},3)^T$.
在这三个点处的Hesse矩阵分别为

$$G_1 = \begin{bmatrix} 6 & 0 \\ 0 & 6 \end{bmatrix}, \quad G_2 = \begin{bmatrix} 0 & -6\sqrt{2} \\ -6\sqrt{2} & 6 \end{bmatrix}$$

$$G_3 = \begin{bmatrix} 0 & 6\sqrt{2} \\ 6\sqrt{2} & 6 \end{bmatrix}$$

- 初始点$\boldsymbol{x}_0 = (1.5, 1.5)^T$，收敛到极小点$(0,0)^T$.
- 初始点$\boldsymbol{x}_0 = (-2, 4)^T$，收敛到鞍点$(-3\sqrt{2},3)^T$.
- 初始点$\boldsymbol{x}_0 = (0, 3)^T$，这时$G(x_0)$ 奇异，方法失败.

数值实验：Prob 9.

例4.5 (课后习题3.8)

考虑函数

$$f(\boldsymbol{x}) = 2x_1^2 + x_2^2 - 2x_1x_2 + 2x_1^3 + x_1^4.$$

找出$\boldsymbol{x}^* = 0$ 的最大开球B, 使得$G(\boldsymbol{x})$ 在其中正定. 对此球中怎样的点$\boldsymbol{x}_0$ (其中$\boldsymbol{x}_0^{(1)} = \boldsymbol{x}_0^{(2)}$), Newton 法收敛到$\boldsymbol{x}^*$.

【解: 】

$$\nabla f(\boldsymbol{x}) = \begin{pmatrix} 4x_1 - 2x_2 + 6x_1^2 + 4x_1^3 \\ 2x_2 - 2x_1 \end{pmatrix},$$

$$G(\boldsymbol{x}) = \begin{pmatrix} 4 + 12x_1 + 12x_1^2 & -2 \\ -2 & 2 \end{pmatrix}.$$

Hesse 矩阵$G(\boldsymbol{x})$ 正定当且仅当

$$\begin{array}{l} 4+12x_1+12x_1^2>0, (\text{该式恒成立}) \\ \det(G(\boldsymbol{x}))=4+24x_1+24x_1^2>0. \end{array}$$

解得

$$x_1>\frac{-3+\sqrt{3}}{6} \text{ 或 } x_1<\frac{-3-\sqrt{3}}{6}.$$

令$r=\frac{-3+\sqrt{3}}{6}$. 则当$\boldsymbol{x}$ 位于如下开球

$$B(\boldsymbol{x}^*,r)=\{\boldsymbol{x}\mid \|\boldsymbol{x}-\boldsymbol{x}^*\|<r\}$$

时, $G(\boldsymbol{x})$ 正定.

为简便起见，记$\boldsymbol{x} = \boldsymbol{x}_0$. 当$x_1 = x_2$ 时，

$$\nabla f = \begin{pmatrix} 2x_1 + 6x_1^2 + 4x_1^3 \\ 0 \end{pmatrix},$$

$$G^{-1} = \frac{1}{4 + 24x_1 + 24x_1^2} \begin{pmatrix} 2 & 2 \\ 2 & 4 + 12x_1 + 12x_1^2 \end{pmatrix},$$

$$-G^{-1}\nabla f = -\frac{1}{4 + 24x_1 + 24x_1^2} \begin{pmatrix} 4x_1 + 12x_1^2 + 8x_1^3 \\ 4x_1 + 12x_1^2 + 8x_1^3 \end{pmatrix}.$$

所以下一个迭代点为

$$\boldsymbol{x}^+ = \boldsymbol{x} - G^{-1}\nabla f = \begin{pmatrix} \frac{12x_1^2+16x_1^3}{4+24x_1+24x_1^2} \\ \frac{12x_1^2+16x_1^3}{4+24x_1+24x_1^2} \end{pmatrix}$$

$$= \frac{12x_1+16x_1^2}{4+24x_1+24x_1^2}\begin{pmatrix} x_1 \\ x_1 \end{pmatrix}$$

所以当

$$\left|\frac{12x_1+16x_1^2}{4+24x_1+24x_1^2}\right| < 1 \tag{5}$$

时，可使迭代点列收敛到$\boldsymbol{x}^*$.

求解(5), 结合$B(\boldsymbol{x}^*, r)$ 可知当初始点$\boldsymbol{x}_0(\boldsymbol{x}_0^{(1)} = \boldsymbol{x}_0^{(2)} = x_1)$ 满足

$$|\boldsymbol{x}_0^{(1)}| < \frac{-9+\sqrt{41}}{20}$$

时, 迭代点列收敛到$\boldsymbol{x}^*$.

求解(5), 结合$B(\boldsymbol{x}^*, r)$ 可知当初始点$\boldsymbol{x}_0(\boldsymbol{x}_0^{(1)} = \boldsymbol{x}_0^{(2)} = x_1)$ 满足

$$|\boldsymbol{x}_0^{(1)}| < \frac{-9+\sqrt{41}}{20}$$

时, 迭代点列收敛到$\boldsymbol{x}^*$.

数值实验: Prob 3.

优缺点

Newton 法的优缺点

优点

优缺点

Newton 法的优缺点

优点

- 如果G^* 正定且初始点合适, 算法是二阶收敛的;

优缺点

Newton 法的优缺点

优点

- 如果G^* 正定且初始点合适, 算法是二阶收敛的;
- 对于二次函数，迭代一次就可得到极小点.

优缺点

Newton 法的优缺点

优点

- 如果G^* 正定且初始点合适, 算法是二阶收敛的;
- 对于二次函数，迭代一次就可得到极小点.

缺点

优缺点

Newton 法的优缺点

优点

- 如果G^* 正定且初始点合适, 算法是二阶收敛的;
- 对于二次函数，迭代一次就可得到极小点.

缺点

- 对多数问题并不是整体收敛（或全局收敛）的;

优缺点

Newton 法的优缺点

优点

- 如果G^* 正定且初始点合适，算法是二阶收敛的；
- 对于二次函数，迭代一次就可得到极小点.

缺点

- 对多数问题并不是整体收敛（或全局收敛）的；
- 在每次迭代中需要计算G_k;

优缺点

Newton 法的优缺点

优点

- 如果G^* 正定且初始点合适, 算法是二阶收敛的;
- 对于二次函数，迭代一次就可得到极小点.

缺点

- 对多数问题并不是整体收敛（或全局收敛）的;
- 在每次迭代中需要计算G_k;
- 每次迭代需要求解线性方程组$G_kd_k = -\nabla f_k$;

优缺点

Newton 法的优缺点

优点

- 如果G^* 正定且初始点合适，算法是二阶收敛的;
- 对于二次函数，迭代一次就可得到极小点.

缺点

- 对多数问题并不是整体收敛（或全局收敛）的;
- 在每次迭代中需要计算G_k;
- 每次迭代需要求解线性方程组$G_k d_k = -\nabla f_k$;
- 收敛于鞍点或极大点的可能性并不小.

定理8

设$\{\boldsymbol{x}_k\}$ 为利用$Newton$ 法求解$\min\ f(\boldsymbol{x})$ 时得到迭代点序列. 如果$G_k = \nabla_{xx} f(\boldsymbol{x}_k)$ 正定且$\nabla f_k \neq 0$, 则$Newton$ 方向

$$d_k = -G_k^{-1}\nabla f_k$$

是一个下降方向.

证明: 只需证明$\nabla f_k^T d_k < 0$ 即可. 设λ_k, μ_k 分别是G_k 的最小、最大特征值. 由于G_k 正定, 故而有

$$0 < \lambda_k \|p\|^2 \le p^T G_k p \le \mu_k \|p\|^2, \ \forall p \in R^n.$$

从而

$$0 < \mu_k^{-1} \|p\|^2 \le p^T G_k^{-1} p \le \lambda_k^{-1} \|p\|^2, \ \forall p \in R^n.$$

于是有

$$\nabla f_k^T d_k = -\nabla f_k^T G_k^{-1} \nabla f_k \le -\mu_k^{-1} \|\nabla f_k\|^2 < 0.$$

证毕.

对Newton 法的改进

改进1: 阻尼Newton 法

当得到Newton 方向d_k 后，沿d_k 进行一维搜索求得步长α_k，例如

对Newton 法的改进

改进1: 阻尼Newton 法

当得到Newton 方向d_k 后，沿d_k 进行一维搜索求得步长α_k，例如

- 用精确搜索确定α_k，使其满足

$$f(\boldsymbol{x}_k + \alpha_k d_k) = \min_{\alpha \geq 0} f(\boldsymbol{x}_k + \alpha d_k).$$

或者

对Newton 法的改进

改进1: 阻尼Newton 法

当得到Newton 方向d_k 后，沿d_k 进行一维搜索求得步长α_k，例如

- 用精确搜索确定α_k，使其满足

$$f(\boldsymbol{x}_k + \alpha_k d_k) = \min_{\alpha \geq 0} f(\boldsymbol{x}_k + \alpha d_k).$$

或者

- 用非精确搜索α_k, 使其满足Wolfe 准则或Armijo 准则.

对Newton 法的改进

改进1: 阻尼Newton 法

当得到Newton 方向d_k 后，沿d_k 进行一维搜索求得步长α_k，例如

- 用精确搜索确定α_k，使其满足

$$f(\boldsymbol{x}_k + \alpha_k d_k) = \min_{\alpha \geq 0} f(\boldsymbol{x}_k + \alpha d_k).$$

  或者

- 用非精确搜索α_k, 使其满足Wolfe 准则或Armijo 准则.

这种改进可以克服第一和第四个缺点.

阻尼Newton 法

阻尼Newton 法(使用精确搜索)

给定误差控制参数$\epsilon > 0$.

步 1 取初始点$\boldsymbol{x}_0$, 令$k = 0$.

步 2 计算∇f_k. 若$\|\nabla f_k\| \le \epsilon$, 则令$\boldsymbol{x}^* = \boldsymbol{x}_k$, 停; 否则计算$G_k$, 求解

$$G_k d_k = -\nabla f_k$$

得d_k.

步 3 用精确搜索确定α_k, 使其满足

$$f(\boldsymbol{x}_k + \alpha_k d_k) = \min_{\alpha \ge 0} f(\boldsymbol{x}_k + \alpha d_k).$$

步 4 令$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k d_k$, $k = k + 1$, 转步2.

阻尼Newton 法

阻尼Newton 法(使用Armijo 非精确搜索)

给定误差控制参数$\epsilon > 0$.

步 1 取参数$0 < \rho, \beta < 1$ 和初始点$\boldsymbol{x}_0$, 令$k = 0$.

步 2 计算∇f_k. 若$\|\nabla f_k\| \leq \epsilon$, 则令$\boldsymbol{x}^* = \boldsymbol{x}_k$, 停; 否则计算$G_k$, 求解

$$G_k d_k = -\nabla f_k$$

得d_k. 令$m = 0$.

步 3 若

$$f(\boldsymbol{x}_k + \beta^m d_k) \leq f_k + \rho\beta^m \nabla f_k^T d_k,$$

令$\alpha_k = \beta^m$; 转步4; 否则, 转步5.

步 4 令$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k d_k$, $k = k + 1$, 转步2.

步 5 令$m = m + 1$, 转步3.

例题

例4.6

求解

$$\min \ \frac{1}{2}\boldsymbol{x}^T\boldsymbol{x} + \sigma(\boldsymbol{x}^T A\boldsymbol{x})^2,$$

其中，$\sigma = 10^4$，$A = \begin{pmatrix} 1 & 1 & 0 & \frac{1}{2} \\ 2 & 4 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 3 & 0 \\ \frac{1}{2} & 0 & 0 & 2 \end{pmatrix}$.

初始点为$(\cos 70^o, \sin 70^o, \cos 70^o, \sin 70^o)$.

例题

例4.6

求解

$$\min \ \frac{1}{2}\boldsymbol{x}^T\boldsymbol{x} + \sigma(\boldsymbol{x}^T A\boldsymbol{x})^2,$$

其中, $\sigma = 10^4$, $A = \begin{pmatrix} 1 & 1 & 0 & \frac{1}{2} \\ 2 & 4 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 3 & 0 \\ \frac{1}{2} & 0 & 0 & 2 \end{pmatrix}$.

初始点为 $(\cos 70^o, \sin 70^o, \cos 70^o, \sin 70^o)$.

本题Newton法失败，阻尼Newton法可以高效求解.

例题

例4.6

求解

$$\min\ \frac{1}{2}\boldsymbol{x}^T\boldsymbol{x}+\sigma(\boldsymbol{x}^TA\boldsymbol{x})^2,$$

其中, $\sigma=10^4$, $A=\begin{pmatrix} 1 & 1 & 0 & \frac{1}{2} \\ 2 & 4 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 3 & 0 \\ \frac{1}{2} & 0 & 0 & 2 \end{pmatrix}.$

初始点为$(\cos 70^o, \sin 70^o, \cos 70^o, \sin 70^o)$.

本题Newton法失败，阻尼Newton法可以高效求解.

数值实验: Prob 5.

对Newton 法的改进

注意到当G_k 正定时, Newton 方向$d_k = -G_k^{-1}\nabla f_k$ 是下降方向. 受此启发, 当G_k 非正定甚至奇异时:

改进2: Levenberg-Marquardt 校正

对Newton 法的改进

注意到当G_k 正定时, Newton 方向$d_k = -G_k^{-1}\nabla f_k$ 是下降方向. 受此启发, 当G_k 非正定甚至奇异时:

改进2: Levenberg-Marquardt 校正

选择参数$\mu_k \geq 0$, 使得$G_k + \mu_k I$ 正定. 令

$$d_k = -(G_k + \mu_k I)^{-1}\nabla f_k,$$

对Newton 法的改进

注意到当G_k 正定时, Newton 方向$d_k = -G_k^{-1}\nabla f_k$ 是下降方向. 受此启发, 当G_k 非正定甚至奇异时:

改进2: Levenberg-Marquardt 校正

选择参数$\mu_k \geq 0$, 使得$G_k + \mu_k I$ 正定. 令

$$d_k = -(G_k + \mu_k I)^{-1}\nabla f_k,$$

再沿d_k 进行线性搜索，确定步长α_k. 令

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k d_k.$$

算例

例4.7

求解问题

$$\min \ f(\boldsymbol{x}) = x_1^4 + x_1 x_2 + (1 + x_2)^2.$$

初始点为$(0, 0)$.

算例

例4.7

求解问题

$$\min \ f(\boldsymbol{x}) = x_1^4 + x_1 x_2 + (1 + x_2)^2.$$

初始点为$(0, 0)$.

Newton 法求解该问题失败.

例题

使用阻尼牛顿法. 在初始点处,

$$\nabla f_0 = \begin{pmatrix} 0 \\ 2 \end{pmatrix},\ G_0 = \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix}, d_0 = -G_0^{-1}\nabla f_0 = \begin{pmatrix} -2 \\ 0 \end{pmatrix}.$$

例题

使用阻尼牛顿法. 在初始点处,

$$\nabla f_0 = \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \; G_0 = \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix}, d_0 = -G_0^{-1}\nabla f_0 = \begin{pmatrix} -2 \\ 0 \end{pmatrix}.$$

由于$\nabla f_0^T d_0 = 0$, 该Newton 方向不是下降方向. 沿方向d_0 进行线性搜索,

$$f(\boldsymbol{x}_0 + \alpha d_0) = (-2\alpha)^4 + 1,$$

其极小点为$\alpha_0 = 0$. 故而,迭代无法继续下去. 阻尼Newton法无法找到下一个迭代点.

例题

使用阻尼牛顿法. 在初始点处,

$$\nabla f_0 = \begin{pmatrix} 0 \\ 2 \end{pmatrix},\ G_0 = \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix}, d_0 = -G_0^{-1}\nabla f_0 = \begin{pmatrix} -2 \\ 0 \end{pmatrix}.$$

由于$\nabla f_0^T d_0 = 0$, 该Newton 方向不是下降方向. 沿方向d_0 进行线性搜索,

$$f(\boldsymbol{x}_0 + \alpha d_0) = (-2\alpha)^4 + 1,$$

其极小点为$\alpha_0 = 0$. 故而,迭代无法继续下去. 阻尼Newton法无法找到下一个迭代点.

但使用L-M 方法可以快速求解该问题. 当G 不是正定矩阵时, 取

$$\mu_k = -\lambda^{\min} + 10^{-3},$$

其中, $\lambda^{\min}$ 是G 的最小特征值.

例题

使用阻尼牛顿法. 在初始点处,

$$\nabla f_0 = \begin{pmatrix} 0 \\ 2 \end{pmatrix},\ G_0 = \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix}, d_0 = -G_0^{-1}\nabla f_0 = \begin{pmatrix} -2 \\ 0 \end{pmatrix}.$$

由于$\nabla f_0^T d_0 = 0$, 该Newton 方向不是下降方向. 沿方向d_0 进行线性搜索,

$$f(\boldsymbol{x}_0 + \alpha d_0) = (-2\alpha)^4 + 1,$$

其极小点为$\alpha_0 = 0$. 故而,迭代无法继续下去. 阻尼Newton法无法找到下一个迭代点.

但使用L-M 方法可以快速求解该问题. 当G 不是正定矩阵时, 取

$$\mu_k = -\lambda^{\min} + 10^{-3},$$

其中, $\lambda^{\min}$ 是G 的最小特征值.

数值实验: Prob 7.

作业

作业. P129. 3.7

思想

取最速下降法和Newton法的优点而舍它们的缺点.

思想

取最速下降法和Newton法的优点而舍它们的缺点.

对于二次函数

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T G\boldsymbol{x} + b^T\boldsymbol{x} + c, \tag{6}$$

其中, G 是正定的.

思想

取最速下降法和Newton法的优点而舍它们的缺点.

对于二次函数

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T G\boldsymbol{x} + b^T\boldsymbol{x} + c, \tag{6}$$

其中, G 是正定的.

- Newton 法只需一次迭代，最速下降法一般需要迭代无穷多次;

思想

取最速下降法和Newton法的优点而舍它们的缺点.

对于二次函数

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T G\boldsymbol{x} + b^T\boldsymbol{x} + c, \tag{6}$$

其中, G 是正定的.

- Newton 法只需一次迭代，最速下降法一般需要迭代无穷多次;
- Newton 法每一步迭代的计算量大(需要计算Hesse 矩阵，解一个线性方程组)，最速下降法每一步迭代的计算量非常小.

思想

取最速下降法和Newton法的优点而舍它们的缺点.

对于二次函数

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T G\boldsymbol{x} + b^T\boldsymbol{x} + c, \tag{6}$$

其中, G 是正定的.

- Newton 法只需一次迭代，最速下降法一般需要迭代无穷多次;
- Newton 法每一步迭代的计算量大(需要计算Hesse 矩阵，解一个线性方程组)，最速下降法每一步迭代的计算量非常小.

折中一下, 希望算法能在有限步内找到二次函数的最小值.

共轭方向

比如，对于二维的二次函数的最优化问题

$$\min f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T G\boldsymbol{x} + b^T\boldsymbol{x} + c,$$

其中, $G \in R^{2\times 2}$ 正定, $b \in R^2$, $c \in R$. 我们希望能够在两步内找到最优解. 如下图

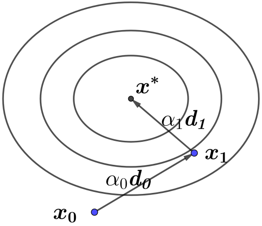


图: 两步解二次函数

思想

设有

迭代点	搜索方向	最优步长
$\boldsymbol{x}_0$	d_0	α_0
$\boldsymbol{x}_1$	d_1	α_1
$\boldsymbol{x}_2 = \boldsymbol{x}^*$	-	-

思想

设有

迭代点	搜索方向	最优步长
x_0	d_0	α_0
x_1	d_1	α_1
$x_2 = x^*$	-	-

其中x_2 是问题的解. 即

思想

设有

迭代点	搜索方向	最优步长
$\boldsymbol{x}_0$	d_0	α_0
$\boldsymbol{x}_1$	d_1	α_1
$\boldsymbol{x}_2 = \boldsymbol{x}^*$	-	-

其中$\boldsymbol{x}_2$ 是问题的解. 即

$$\begin{aligned}
&\nabla f(\boldsymbol{x}_2) = G\boldsymbol{x}_2 + b \\
=&G(\boldsymbol{x}_1 + \alpha_1 d_1) + b \\
=&(G\boldsymbol{x}_1 + b) + \alpha_1 G d_1 = 0.
\end{aligned}$$

思想

$$\boxed{(G\boldsymbol{x}_1 + b) + \alpha_1 G d_1 = 0}.$$

思想

$$\boxed{(G\boldsymbol{x}_1 + b) + \alpha_1 G d_1 = 0}.$$

又由精确搜索的性质

$$\nabla f(\boldsymbol{x}_{k+1})^T d_k = 0$$

知,

$$(G\boldsymbol{x}_1 + b)^T d_0 = 0.$$

思想

$$\boxed{(G\boldsymbol{x}_1+b)+\alpha_1 G d_1=0}.$$

又由精确搜索的性质

$$\nabla f(\boldsymbol{x}_{k+1})^T d_k=0$$

知,

$$(G\boldsymbol{x}_1+b)^T d_0=0.$$

所以有

$$\begin{aligned}0=&d_0^T((G\boldsymbol{x}_1+b)+\alpha_1 G d_1)\\ =&d_0^T(G\boldsymbol{x}_1+b)+\alpha_1 d_0^T G d_1\\ =&0+\alpha_1 d_0^T G d_1.\end{aligned}$$

思想

得到关系式

$$d_0^T G d_1 = 0.$$

思想

得到关系式

$$d_0^T G d_1 = 0.$$

满足这个条件的两个方向称为共轭方向.

思想

得到关系式

$$d_0^T G d_1 = 0.$$

满足这个条件的两个方向称为共轭方向.

由于$d_1 \in R^2$, 故而d_1 可以表示为$-g_1$ 和d_0 的线性组合.

思想

得到关系式

$$d_0^T G d_1 = 0.$$

满足这个条件的两个方向称为共轭方向.
由于$d_1 \in R^2$, 故而d_1 可以表示为$-g_1$ 和d_0 的线性组合.
容易验证

$$d_1 = -g_1 + \frac{d_0^T G g_1}{d_0^T G d_0} d_0.$$

满足共轭性. 这就是共轭梯度的搜索方向格式.

共轭方向及其性质

定义1 (共轭向量)

设G 为n 阶正定矩阵, $d_1, d_2, \cdots, d_k$ 为n 维向量组, 如果

$$d_i^T G d_j = 0,\ \ i, j = 1, 2, \cdots, k,\ \ i \neq j,$$

则称向量组$d_1, d_2, \cdots, d_k$ 关于G 共轭.

共轭方向及其性质

定义1 (共轭向量)

设G 为n 阶正定矩阵, d_1, d_2, $\cdots$, d_k 为n 维向量组, 如果

$$d_i^T G d_j = 0,\ \ i, j = 1, 2, \cdots, k,\ \ i \neq j,$$

则称向量组d_1, d_2, $\cdots$, d_k 关于G 共轭.

【注: 】如果$G = I$, 则$d_i^T G d_j = 0$ 变成$d_i^T d_j = 0$. 所以, 共轭概念是正交概念的推广.

例题

例5.1

设$f(x) = x^T Gx + b^T x$, 其中$G = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$, $b = \begin{pmatrix} -3 \\ -3 \end{pmatrix}$.

(1) 证明$d_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ 与$d_1 = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$ 关于G 共轭;

(2) 设$x_0 = (0,0)^T$, 以d_0 和d_1 为搜索方向, 用精确搜索求f 的极小点.

例题

例5.1

设$f(x) = x^T G x + b^T x$, 其中$G = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$, $b = \begin{pmatrix} -3 \\ -3 \end{pmatrix}$.

(1) 证明$d_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ 与$d_1 = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$ 关于G 共轭;

(2) 设$x_0 = (0, 0)^T$, 以d_0 和d_1 为搜索方向, 用精确搜索求f 的极小点.

【解: 】(1) 验证

$$
\begin{aligned}
d_0^\top G d_1 &= (1, 0) \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} -1 \\ 2 \end{pmatrix} \\
&= (2, 1) \begin{pmatrix} -1 \\ 2 \end{pmatrix} \\
&= 0
\end{aligned}
$$

例题

(2) $x_0 + \alpha d_0 = (\alpha, 0)^T$. 则

$$\begin{aligned}\varphi(x) = f\left(x_0 + \alpha d_0\right) &= (\alpha, 0)\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} \alpha \\ 0 \end{pmatrix} - 3\alpha \\ &= 2\alpha^2 - 3\alpha\end{aligned}$$

例题

(2) $x_0 + \alpha d_0 = (\alpha, 0)^T$. 则

$$\varphi(x) = f(x_0 + \alpha d_0) = (\alpha, 0)\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} \alpha \\ 0 \end{pmatrix} - 3\alpha$$
$$= 2\alpha^2 - 3\alpha$$

令$\varphi'(\alpha) = 4\alpha - 3 = 0$ 得步长$\alpha_0 = \frac{3}{4}$. 故

$$x_1 = x_0 + \alpha_0 d_0 = \left(\frac{3}{4}, 0\right)^T.$$

例题

(2) $x_0 + \alpha d_0 = (\alpha, 0)^T$. 则

$$\varphi(x) = f(x_0 + \alpha d_0) = (\alpha, 0)\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} \alpha \\ 0 \end{pmatrix} - 3\alpha$$
$$= 2\alpha^2 - 3\alpha$$

令$\varphi'(\alpha) = 4\alpha - 3 = 0$ 得步长$\alpha_0 = \frac{3}{4}$. 故

$$x_1 = x_0 + \alpha_0 d_0 = \left(\frac{3}{4}, 0\right)^T.$$

定义$\varphi(\alpha) = f(x_1 + \alpha d_1) = 6\alpha^2 - 3\alpha - \frac{9}{8}$.

令$\varphi'(\alpha) = 12\alpha - 3 = 0$ 得步长$\alpha_1 = \frac{1}{4}$.

令$\varphi'(\alpha) = 12\alpha - 3 = 0$ 得步长$\alpha_1 = \frac{1}{4}$.
于是$x_2 = x_1 + \alpha_1 d_1 = \left(\frac{1}{2}, \frac{1}{2}\right)^T$. 这是问题的解.

定理9

设G 为n 阶正定矩阵, 非零向量组$p_1, p_2, \cdots, p_k$ 关于G 共轭, 则此向量组线性无关.

证明:设存在常数$\lambda_1, \lambda_2, \cdots, \lambda_k$, 使得

$$\lambda_1 p_1 + \lambda_2 p_2 + \cdots + \lambda_k p_k = 0.$$

用$p_i^T G$ 左乘上式, 根据假设得

$$\lambda_i p_i^T G p_i = 0, i = 1, 2, \cdots, k.$$

由于G 是正定的, $p_i \neq 0$, 故而$\lambda_i = 0$, $i = 1, 2, \cdots, k$. 所以向量组$p_1, p_2, \cdots, p_k$ 线性无关.

推论1

设G 为n 阶正定矩阵，非零向量组$p_1, p_2, \cdots, p_n$ 关于G 共轭，则此向量组构成n 维向量空间R^n 的一组基.

推论2

设G 为n 阶正定矩阵，非零向量组$p_1, p_2, \cdots, p_n$ 关于G 共轭. 若向量v 与$p_1, p_2, \cdots, p_n$ 关于G 共轭，则$v = 0$.

共轭方向法

二次函数的共轭方向法框架

设目标函数为二次函数(6)，其中，G 是正定的，给定$\epsilon > 0$.

【步 1】 给定初始点x_0 及初始下降方向d_0，令$k = 0$.

【步 2】 进行精确一维搜索，求步长α_k，

$$f(x_k + \alpha d_k) = \min_{\alpha > 0} f(x_k + \alpha d_k).$$

【步 3】 令$x_{k+1} = x_k + \alpha_k d_k$.

【步 4】 若$\|g_{k+1}\| \le \epsilon$，则令$x^* = x_{k+1}$，否则，转步5.

【步 5】 取共轭方向d_{k+1}（有无穷多种取法）使得

$$d_{k+1}^T G d_i = 0,\ \ i = 0, 1, \cdots, k.$$

【步 6】 令$k = k + 1$，转步2.

定义2

设n 维向量组p_1, p_2, p_k 线性无关, $\boldsymbol{x}_1 \in R^n$. 称向量集合

$$H_k = \{\boldsymbol{x}_1 + \sum_{i=1}^{k} \alpha_i p_i \mid \alpha_i \in R^1, i = 1, 2, \cdots, k\}$$

为由点$\boldsymbol{x}_1$ 和p_1, p_2, p_k 生成的k 为超平面.

引理1

设$f(\boldsymbol{x})$ 为连续可微的严格凸函数，又p_1，p_2，p_k 为线性无关的n 维向量组，$\boldsymbol{x}_1 \in R^n$. 则

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_1 + \sum_{i=1}^{k} \bar{\alpha}_i p_i$$

是$f(\boldsymbol{x})$ 在$\boldsymbol{x}_1$ 和p_1，p_2，p_k 生成的k 为超平面H_k 上的唯一极小点的充分必要条件是

$$\nabla f_{k+1}^T p_i = 0, \; i = 1, 2, \cdots, k. \tag{7}$$

证明:定义函数

$$h(\alpha_1, \alpha_2, \cdots, \alpha_k) = f(\boldsymbol{x}_1 + \sum_{k=1}^{k} \alpha_i p_i).$$

则h 是k 维严格凸函数, 且$\boldsymbol{x}_{k+1}$ 是$f(\boldsymbol{x})$ 在H_k 上的极小点当且仅当$(\bar{\alpha}_1, \bar{\alpha}_2, \cdots, \bar{\alpha}_k)^T$ 是$h(\alpha_1, \alpha_2, \cdots, \alpha_k)$ 在R^k 上的极小点.

若$\boldsymbol{x}_{k+1}$ 是$f(\boldsymbol{x})$ 在H_k 上的极小点, 则

$$\nabla h(\bar{\alpha}_1, \bar{\alpha}_2, \cdots, \bar{\alpha}_k) = 0.$$

又因

$$\nabla h(\bar{\alpha}_1, \bar{\alpha}_2, \cdots, \bar{\alpha}_k) = (g_{k+1}^T p_1, g_{k+1}^T p_2, \cdots, g_{k+1}^T p_k)^T,$$

所以式(7) 成立.

反之, 设式(7) 成立, 则又上面的推导过程知

$$\nabla h(\bar{\alpha}_1, \bar{\alpha}_2, \cdots, \bar{\alpha}_k) = 0,$$

又因h 是严格凸函数, 所以$(\bar{\alpha}_1, \bar{\alpha}_2, \cdots, \bar{\alpha}_k)^T$ 是h 的唯一极小点. 从而知$\boldsymbol{x}_{k+1}$ 是$f(\boldsymbol{x})$ 在H_k 上的唯一极小点.

共轭方向法

定理10

设G 为n 阶正定矩阵，向量组$p_1, p_2, \cdots, p_k$ 关于G 共轭，对正定二次函数

$$f(x) = \frac{1}{2}\boldsymbol{x}^T G\boldsymbol{x} + b^T\boldsymbol{x} + c, \tag{8}$$

由任意初始点$\boldsymbol{x}_1$ 开始，依次进行k 次精确一维搜索

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i + \alpha_i p_i, \ i = 1, 2, \cdots, k, \tag{9}$$

则

(i) $\nabla f_{k+1}^T p_i = 0, \ i = 1, 2, \cdots, k;$

(ii) $\boldsymbol{x}_{k+1}$ 是二次函数(8) 在k 维超平面H_k 上的极小点.

共轭方向法

证明:由引理1, 只需证明(i), 因为

$$\nabla f_{k+1}^T p_i = \nabla f_{i+1}^T p_i + \sum_{j=i+1}^{k} (\nabla f_{j+1} - \nabla f_j)^T p_i.$$

对于二次函数

$$\nabla f_{j+1} - \nabla f_j = G(x_{j+1} - x_j) = \alpha_j G p_j.$$

所以

$$\nabla f_{k+1}^T p_i = \nabla f_{i+1}^T p_i + \sum_{j=i+1}^{k} \alpha_j p_j^T G p_i.$$

由于采用精确一维搜索, 故$\nabla f_{i+1}^T p_i = 0$, 又由共轭性, $p_j^T G p_i = 0$, $j = i+1, \cdots, k$, 因此(i) 成立.

共轭梯度法

推论3 (共轭方向法的有限终止性)

在上述定理中，当$k=n$ 时，$\boldsymbol{x}_{n+1}$ 为正定二次函数(8) 在R^n 上的极小点.

共轭方向法

分析:

共轭方向法

分析:

- 共轭方向法的计算效率很高;

共轭方向法

分析：

- 共轭方向法的计算效率很高;
- 在应用该算法时，应当给出具体的计算共轭方向的方法.

共轭梯度法

共轭梯度法-共轭方向的构造

共轭梯度法

共轭梯度法-共轭方向的构造

思想：利用负梯度方向$-g_k$ 与上一次迭代的搜索方向d_{k-1}构造搜索方向.

共轭梯度法

共轭梯度法-共轭方向的构造

思想：利用负梯度方向$-g_k$ 与上一次迭代的搜索方向d_{k-1}构造搜索方向.

- $d_0 = -g_0$;

共轭梯度法

共轭梯度法-共轭方向的构造

思想：利用负梯度方向$-g_k$ 与上一次迭代的搜索方向d_{k-1}构造搜索方向.

- $d_0 = -g_0$;
- $d_k = -g_k + \beta_{k-1} d_{k-1}$.

共轭梯度法

共轭梯度法-共轭方向的构造

思想：利用负梯度方向$-g_k$与上一次迭代的搜索方向d_{k-1}构造搜索方向.

- $d_0 = -g_0$;
- $d_k = -g_k + \beta_{k-1} d_{k-1}$.
- 由于$d_k^T G d_{k-1} = 0$, 可得

$$\beta_{k-1} = \frac{g_k^T G d_{k-1}}{d_{k-1}^T G d_{k-1}}, \quad \text{Hestenes-Stiefel 公式.}$$

共轭梯度法

共轭梯度法的一般框架

给定误差$\epsilon > 0$.

【步 1】 给出$\boldsymbol{x}_0 \in R^n$, $k := 0$;

【步 2】 若$\|g_k\| \leq \epsilon$, 则停止迭代;

【步 3】 若$k = 0$, 则$\beta_{-1} = 0$, 否则计算β_{k-1}, $d_k = -g_k + \beta_{k-1} d_{k-1}$,

【步 4】 做一维线性搜索求α_k;

【步 5】 计算$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k d_k$;

【步 6】 令$k := k + 1$. 转步2.

共轭梯度法

定理11 (共轭梯度法的有限终止性)

对正定二次函数(6)采用共轭梯度法确定共轭方向，并采用精确一维搜索得到步长，在m $(m \leq n)$ 次迭代后可求得(6) 的极小点，并对所有i, $(1 \leq i \leq m)$，有

(i) $d_i^T G d_j = 0,\ j = 1, 2, \cdots, i-1;$（共轭性）

共轭梯度法

定理11 (共轭梯度法的有限终止性)

对正定二次函数(6)采用共轭梯度法确定共轭方向，并采用精确一维搜索得到步长，在m $(m \le n)$ 次迭代后可求得(6) 的极小点，并对所有i, $(1 \le i \le m)$，有

(i) $d_i^T G d_j = 0,\ j = 1, 2, \cdots, i-1;$（共轭性）

(ii) $g_i^T g_j = 0,\ j = 1, 2, \cdots, i-1;$（正交性）

共轭梯度法

定理11 (共轭梯度法的有限终止性)

对正定二次函数(6)采用共轭梯度法确定共轭方向，并采用精确一维搜索得到步长，在m $(m \leq n)$ 次迭代后可求得(6) 的极小点，并对所有i，$(1 \leq i \leq m)$，有

(i) $d_i^T G d_j = 0,\ j = 1, 2, \cdots, i-1;$（共轭性）

(ii) $g_i^T g_j = 0,\ j = 1, 2, \cdots, i-1;$（正交性）

(iii) $g_i^T d_j = 0,\ j = 1, 2, \cdots, i-1;$

共轭梯度法

定理11 (共轭梯度法的有限终止性)

对正定二次函数(6)采用共轭梯度法确定共轭方向，并采用精确一维搜索得到步长，在m $(m \leq n)$ 次迭代后可求得(6) 的极小点，并对所有i, $(1 \leq i \leq m)$，有

(i) $d_i^T G d_j = 0,\ j = 1, 2, \cdots, i-1;$（共轭性）

(ii) $g_i^T g_j = 0,\ j = 1, 2, \cdots, i-1;$（正交性）

(iii) $g_i^T d_j = 0,\ j = 1, 2, \cdots, i-1;$

(iv) $d_i^T g_i = -g_i^T g_i.$（下降性）

共轭梯度法

系数β_{k-1} 的其它形式

系数β_{k-1} 的其它形式

- Fletcher-Reeves (FR) 公式:

$$\beta_{k-1} = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}.$$

共轭梯度法

系数β_{k-1} 的其它形式

- Fletcher-Reeves (FR) 公式:

$$\beta_{k-1} = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}.$$

- Polak-Ribiere-Polyak (PRP) 公式

$$\beta_{k-1} = \frac{g_k^T (g_k - g_{k-1})}{g_{k-1}^T g_{k-1}}.$$

FR共轭梯度法

在共轭梯度法中采用FR 公式即得FR 共轭梯度法

FR 共轭梯度法

给定误差ϵ.

步 1 给定初始点$\boldsymbol{x}_0$, $k=0$.

步 2 计算$g_k = g(\boldsymbol{x}_k)$.

步 3 若$\|g_k\| \le \epsilon$, 则令$\boldsymbol{x}^* = \boldsymbol{x}_k$, 停; 否则令

$$d_k = -g_k + \beta_{k-1} d_{k-1}, \beta_{k-1} = \begin{cases} \dfrac{g_k^T g_k}{g_{k-1}^T g_{k-1}}, & \text{当}k>0\text{时}, \\ 0, & \text{当}k=0\text{时}. \end{cases}$$

步 4 由精确一维搜索确定步长α_k, 满足

$$f(\boldsymbol{x}_k + \alpha_k d_k) = \min_{\alpha > 0} f(\boldsymbol{x}_k + \alpha d_k).$$

步 5 令$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k d_k$, $k = k+1$, 转步2.

FR 共轭梯度法

例5.2

用FR 共轭梯度法求解

$$\min \ f(\boldsymbol{x}) = \frac{3}{2}x_1^2 + \frac{1}{2}x_2^2 - x_1x_2 - 2x_1,$$

取初始点$\boldsymbol{x}_0 = (0,0)^T$.

FR 共轭梯度法

例5.2

用 FR 共轭梯度法求解

$$\min \ f(\boldsymbol{x}) = \frac{3}{2}x_1^2 + \frac{1}{2}x_2^2 - x_1x_2 - 2x_1,$$

取初始点 $\boldsymbol{x}_0 = (0,0)^T$.

【解：】 $g(\boldsymbol{x}) = \begin{pmatrix} 3x_1 - x_2 - 2 \\ x_2 - x_1 \end{pmatrix}, G = \begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix}.$

FR 共轭梯度法

例5.2

用 FR 共轭梯度法求解

$$\min\ f(\boldsymbol{x}) = \frac{3}{2}x_1^2 + \frac{1}{2}x_2^2 - x_1x_2 - 2x_1,$$

取初始点 $\boldsymbol{x}_0 = (0,0)^T$.

【解：】$g(\boldsymbol{x}) = \begin{pmatrix} 3x_1 - x_2 - 2 \\ x_2 - x_1 \end{pmatrix}$, $G = \begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix}$.

因 $g_0 = (-2,0)^T \neq 0$, 故取 $d_0 = (2,0)^T$, 从 $\boldsymbol{x}_0$ 出发, 沿 d_0 进行一维搜索, 即求

$$\min\ f(\boldsymbol{x}_0 + \alpha d_0) = 6\alpha^2 - 4\alpha$$

的极小点, 得步长 $\alpha = \frac{1}{3}$.

FR共轭梯度法

于是得到$\boldsymbol{x}_1 = \boldsymbol{x}_0 + \alpha_0 d_0 = \left(\frac{2}{3}, 0\right)^T$, 则$g_1 = \left(0, -\frac{2}{3}\right)^T$.

FR共轭梯度法

于是得到$\boldsymbol{x}_1 = \boldsymbol{x}_0 + \alpha_0 d_0 = \left(\frac{2}{3}, 0\right)^T$, 则$g_1 = \left(0, -\frac{2}{3}\right)^T$. 由FR公式得

$$\beta_0 = \frac{g_1^T g_1}{g_0^T g_0} = \frac{1}{9},$$

故$d_1 = -g_1 + \beta_0 d_0 = \left(\frac{2}{9}, \frac{2}{3}\right)^T$.

FR共轭梯度法

于是得到$\boldsymbol{x}_1 = \boldsymbol{x}_0 + \alpha_0 d_0 = \left(\frac{2}{3}, 0\right)^T$, 则$g_1 = \left(0, -\frac{2}{3}\right)^T$. 由FR公式得

$$\beta_0 = \frac{g_1^T g_1}{g_0^T g_0} = \frac{1}{9},$$

故$d_1 = -g_1 + \beta_0 d_0 = \left(\frac{2}{9}, \frac{2}{3}\right)^T$.

从$\boldsymbol{x}_1$ 出发, 沿d_1 进行一维搜索, 求

$$\min\ f(\boldsymbol{x}_1 + \alpha d_1) = \frac{4}{27}\alpha^2 - \frac{4}{9}\alpha + \frac{2}{3}$$

的极小点. 解之得$\alpha_1 = \frac{3}{2}$, 于是$\boldsymbol{x}_2 = \boldsymbol{x}_1 + \alpha_1 d_1 = (1, 1)^T$.

于是得到$\boldsymbol{x}_1 = \boldsymbol{x}_0 + \alpha_0 d_0 = \left(\frac{2}{3}, 0\right)^T$, 则$g_1 = \left(0, -\frac{2}{3}\right)^T$. 由FR公式得

$$\beta_0 = \frac{g_1^T g_1}{g_0^T g_0} = \frac{1}{9},$$

故$d_1 = -g_1 + \beta_0 d_0 = \left(\frac{2}{9}, \frac{2}{3}\right)^T$.
从$\boldsymbol{x}_1$ 出发, 沿d_1 进行一维搜索, 求

$$\min \ f(\boldsymbol{x}_1 + \alpha d_1) = \frac{4}{27}\alpha^2 - \frac{4}{9}\alpha + \frac{2}{3}$$

的极小点. 解之得$\alpha_1 = \frac{3}{2}$, 于是$\boldsymbol{x}_2 = \boldsymbol{x}_1 + \alpha_1 d_1 = (1,1)^T$. 此时$g_2 = (0,0)^T$, 故$\boldsymbol{x}^* = \boldsymbol{x}_2 = (1,1)^T$, $f^* = -1$.

例题

例5.3

利用例*5.2* 的结果验证定理*11* 的各个结论.

例题

例5.3

利用例5.2 的结果验证定理11 的各个结论.

【解：】由例5.2 得

$$G=\begin{pmatrix}3 & -1\\ -1 & 1\end{pmatrix}\quad g_0=\begin{pmatrix}-2\\ 0\end{pmatrix}\quad d_0=\begin{pmatrix}2\\ 0\end{pmatrix},$$

$$g_1=\begin{pmatrix}0\\ -\frac{2}{3}\end{pmatrix}\quad d_1=\begin{pmatrix}\frac{2}{9}\\ \frac{2}{3}\end{pmatrix}$$

例题

例5.3

利用例*5.2* 的结果验证定理*11* 的各个结论.

【解：】由例5.2 得

$$G=\begin{pmatrix}3 & -1\\ -1 & 1\end{pmatrix}\quad g_0=\begin{pmatrix}-2\\ 0\end{pmatrix}\quad d_0=\begin{pmatrix}2\\ 0\end{pmatrix},$$

$$g_1=\begin{pmatrix}0\\ -\frac{2}{3}\end{pmatrix}\quad d_1=\begin{pmatrix}\frac{2}{9}\\ \frac{2}{3}\end{pmatrix}$$

验证

$$d_0^TGd_1=0,\ g_1^Td_0=0, g_0^Tg_1=0,$$
$$g_1^Td_1=-g_1^Tg_1, g_0^Td_0=-g_0^Tg_0.$$

算法的下降性

- 对于正定二次函数, FR 共轭梯度法与PRP共轭梯度法等价.
- 对于一般函数
  - 二者是不同的;
  - 且由于目标函数的Hesse 矩阵不是常数矩阵, 因而迭代过程中产生的方向不再是共轭方向;
  - 在解$\boldsymbol{x}^*$ 的附近, Hesse 矩阵的近似于$G(\boldsymbol{x}^*)$, 因此方向接近共轭方向;
  - 算法产生的方向都满足

$$g_k^T d_k = (-g_k + \beta_{k-1} d_{k-1})^T g_k = -g_k^T g_k < 0,$$

故d_k 都是下降方向.

问题：如果初始方向不是负梯度方向, 那么按FR 或PRP 方法产生的方向还是共轭梯度方向吗?

例题

例5.4

问题

$$\min f(x,y) = x^2 + y^2.$$

初始点为$x_0 = (1,0)^T$, $g_0 = (2,0)^T$, 取$d_0 = (-1,-1)^T$. 用FR共轭梯度公式计算d_1, 并验证d_0, d_1 是否共轭.

例题

例5.4

问题

$$\min f(x,y) = x^2 + y^2.$$

初始点为$x_0 = (1,0)^T$, $g_0 = (2,0)^T$, 取$d_0 = (-1,-1)^T$. 用FR共轭梯度公式计算d_1, 并验证d_0, d_1 是否共轭.

【解: 】记

$$G = \begin{pmatrix} 2 & \\ & 2 \end{pmatrix}.$$

由于$g_0^T d_0 = -2 < 0$, 故而d_0 是下降方向.

例题

例5.4

问题

$$\min f(x,y) = x^2 + y^2.$$

初始点为$x_0 = (1,0)^T$, $g_0 = (2,0)^T$, 取$d_0 = (-1,-1)^T$. 用FR共轭梯度公式计算d_1, 并验证d_0, d_1 是否共轭.

【解：】记

$$G = \begin{pmatrix} 2 & \\ & 2 \end{pmatrix}.$$

由于$g_0^T d_0 = -2 < 0$, 故而d_0 是下降方向. 按精确一维搜索

$$\alpha_0 = \arg\min f(x_0 + \alpha d_0) = \frac{1}{2}.$$

例题

例5.4

问题

$$\min f(x,y) = x^2 + y^2.$$

初始点为$x_0 = (1,0)^T$, $g_0 = (2,0)^T$, 取$d_0 = (-1,-1)^T$. 用FR共轭梯度公式计算d_1, 并验证d_0, d_1 是否共轭.

【解: 】记

$$G = \begin{pmatrix} 2 & \\ & 2 \end{pmatrix}.$$

由于$g_0^T d_0 = -2 < 0$, 故而d_0 是下降方向. 按精确一维搜索

$$\alpha_0 = \arg\min f(x_0 + \alpha d_0) = \frac{1}{2}.$$

从而, $x_1 = x_0 + \alpha_0 d_0 = \left(\frac{1}{2}, -\frac{1}{2}\right)^T$.

例题

计算

$$g_1 = (1, -1)^T,\ \beta_0 = \frac{1}{2},\ d_1 = -g_1 + \beta_0 d_0 = \left(-\frac{3}{2}, \frac{1}{2}\right)^T.$$

例题

计算

$$g_1 = (1, -1)^T,\ \beta_0 = \frac{1}{2},\ d_1 = -g_1 + \beta_0 d_0 = \left(-\frac{3}{2}, \frac{1}{2}\right)^T.$$

但是

$$d_1^T G d_0 = 2 \neq 0,$$

即d_0, d_1 不是关于G 共轭的.

例题

计算

$$g_1 = (1, -1)^T,\ \beta_0 = \frac{1}{2},\ d_1 = -g_1 + \beta_0 d_0 = \left(-\frac{3}{2}, \frac{1}{2}\right)^T.$$

但是

$$d_1^T G d_0 = 2 \neq 0,$$

即d_0, d_1 不是关于G 共轭的.

本题说明：若$d_0 \neq -g_0$, 那么用FR (或PRP) 共轭梯度公式得到的搜索方向不具有共轭性.

PRP 共轭梯度法

将FR 共轭梯度法步3 中，用PRP 公式

$$\beta_{k-1} = \begin{cases} \frac{g_k^T(g_k - g_{k-1})}{g_{k-1}^T g_{k-1}}, & 当k > 0时, \\ 0, & 当k = 0时. \end{cases}$$

就得到PRP 共轭梯度法.

非二次函数的共轭梯度法

对正定二次函数, FR 共轭梯度法和PRP 共轭梯度法是等价的. 对于一般函数, 二者是不相同的.

非二次函数的共轭梯度法

对正定二次函数, FR 共轭梯度法和PRP 共轭梯度法是等价的. 对于一般函数, 二者是不相同的.

对一般函数, 迭代过程中所产生的方向不再是共轭方向. 使用精确搜索时, 有

$$g_k^T d_k = g_k^T(-g_k + \beta_{k-1} d_{k-1}) = -g_k^T g_k < 0,$$

故二者都是下降算法.

非二次函数的共轭梯度法

若采用非精确搜索，则FR 方法和PRP 方法都可能产生上升方向. 对于FR 方法，只有使用如下的强Wolfe 搜索

非二次函数的共轭梯度法

若采用非精确搜索，则FR 方法和PRP 方法都可能产生上升方向. 对于FR 方法，只有使用如下的强Wolfe 搜索

- $f(x_k) - f(x_k + \alpha_k d_k) \geq -\mu\alpha_k \nabla f_k^T d_k, \quad \text{(sw.a)}$
- $|\nabla f(x_k + \alpha_k d_k)^T d_k| \leq -\sigma \nabla f_k^T d_k \quad \text{(sw.b)}$
- $\mu \in \left(0, \frac{1}{2}\right), \sigma \in (\mu, 1),$

且保证$0 < \sigma < 1/2$ 时，得到的方向是下降方向.

非二次函数的共轭梯度法

若采用非精确搜索，则FR 方法和PRP 方法都可能产生上升方向. 对于FR 方法，只有使用如下的强Wolfe 搜索

- $f(x_k) - f(x_k + \alpha_k d_k) \geq -\mu\alpha_k \nabla f_k^T d_k, \quad \text{(sw.a)}$
- $|\nabla f(x_k + \alpha_k d_k)^T d_k| \leq -\sigma \nabla f_k^T d_k \quad \text{(sw.b)}$
- $\mu \in \left(0, \frac{1}{2}\right), \sigma \in (\mu, 1),$

且保证$0 < \sigma < 1/2$ 时，得到的方向是下降方向.

实际计算表明，PRP 算法一般优于FR 算法.

例题

例5.5

分别用FR、PRP共轭梯度法求$Rosenbrock$ 函数

$$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

的最小值，初始点为$x_0 = (0,0)^T$.

例题

例5.5

分别用FR、PRP共轭梯度法求$Rosenbrock$ 函数

$$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

的最小值，初始点为$x_0 = (0,0)^T$.

【解：】第一步与例3.3 相同.

$$\nabla f(x) = \begin{pmatrix} -400(x_2 - x_1^2)x_1 - 2(1 - x_1) \\ 200(x_2 - x_1^2) \end{pmatrix},$$

$f_0 = f(0,0) = 1$, $\nabla f_0 = (-2, 0)^T$.

例题

$d_0 = (2,0)^T$. 对于任意的$\alpha > 0$,

$$f(x_0 + \alpha d_0) = f(2\alpha, 0) = 1600\alpha^4 + 4\alpha^2 - 4\alpha + 1.$$

例题

$d_0 = (2, 0)^T$. 对于任意的$\alpha > 0$,

$$f(x_0 + \alpha d_0) = f(2\alpha, 0) = 1600\alpha^4 + 4\alpha^2 - 4\alpha + 1.$$

令$f'(\alpha) = 0$, 得

$$6400\alpha^3 + 8\alpha - 4 = 0.$$

其近似解为$\alpha_0 \approx 0.0806$.

$$x_1 = x_0 + \alpha_0 d_0 = (0.1612, 0)^T.$$

例题

【FR方法】则$\nabla f_1 = (0.0007, -5.2024)^T$,

$$\beta_0 = \frac{\|g_1\|^2}{\|g_0\|^2} = 6.7662.$$

$$d_1 = -\nabla f_1 + \beta_0 d_0 = \begin{pmatrix} 13.5317 \\ 5.2024 \end{pmatrix}.$$

做精确搜索得$\alpha_1 = 0.0097$. 故

$$x_2 = x_1 + \alpha_1 d_1 = (0.2927, 0.0505)^T.$$

例题

$$\nabla f_2 = (2.7021, -7.0312)^T,$$

$$\beta_1 = \frac{\|g_2\|^2}{\|g_1\|^2} = 2.0964.$$

例题

$$\nabla f_2 = (2.7021, -7.0312)^T,$$

$$\beta_1 = \frac{\|g_2\|^2}{\|g_1\|^2} = 2.0964.$$

$$d_2 = -\nabla f_2 + \beta_1 d_1 = \begin{pmatrix} 25.6664 \\ 17.9378 \end{pmatrix}.$$

做精确搜索得$\alpha_2 = 0.0052.$

例题

$\nabla f_2 = (2.7021, -7.0312)^T$,

$$\beta_1 = \frac{\|g_2\|^2}{\|g_1\|^2} = 2.0964.$$

$$d_2 = -\nabla f_2 + \beta_1 d_1 = \begin{pmatrix} 25.6664 \\ 17.9378 \end{pmatrix}.$$

做精确搜索得$\alpha_2 = 0.0052$. 故

$$x_3 = x_2 + \alpha_2 d_2 = (0.4252, 0.1431)^T, \ \nabla f_3 = \begin{pmatrix} 5.2576 \\ -7.5351 \end{pmatrix}$$

……

例题

【PRP方法】则$\nabla f_1 = (0.0007, -5.2024)^T$,

$$\beta_0 = \frac{g_1^T(g_1 - g_0)}{\|g_0\|^2} = 6.7662.$$

$$d_1 = -\nabla f_1 + \beta_0 d_0 = \begin{pmatrix} 13.5317 \\ 5.2024 \end{pmatrix}.$$

做精确搜索得$\alpha_1 = 0.0097$. 故

$$x_2 = x_1 + \alpha_1 d_1 = (0.2927, 0.0505)^T.$$

例题

$\nabla f_2 = (2.7021, -7.0312)^T,$

$$\beta_1 = \frac{g_2^T(g_2 - g_1)}{\|g_1\|^2} = 0.7449.$$

例题

$$\nabla f_2 = (2.7021, -7.0312)^T,$$

$$\beta_1 = \frac{g_2^T(g_2 - g_1)}{\|g_1\|^2} = 0.7449.$$

$$d_2 = -\nabla f_2 + \beta_1 d_1 = \begin{pmatrix} 7.3776 \\ 10.9064 \end{pmatrix}.$$

做精确搜索得$\alpha_2 = 0.1151.$

例题

$$\nabla f_2 = (2.7021, -7.0312)^T,$$

$$\beta_1 = \frac{g_2^T(g_2 - g_1)}{\|g_1\|^2} = 0.7449.$$

$$d_2 = -\nabla f_2 + \beta_1 d_1 = \begin{pmatrix} 7.3776 \\ 10.9064 \end{pmatrix}.$$

做精确搜索得$\alpha_2 = 0.1151$. 故

$$x_3 = x_2 + \alpha_2 d_2 = (1.1420, 1.3061)^T. \ \nabla f_3 = \begin{pmatrix} -0.5491 \\ 0.3648 \end{pmatrix}.$$

重开始的共轭梯度法

改进思路

重开始的共轭梯度法

改进思路

- 共轭梯度法对正定二次函数可以有限终止;

重开始的共轭梯度法

改进思路

- 共轭梯度法对正定二次函数可以有限终止;
- 在最优解的附近, 目标函数与一个正定二次函数很接近;

重开始的共轭梯度法

改进思路

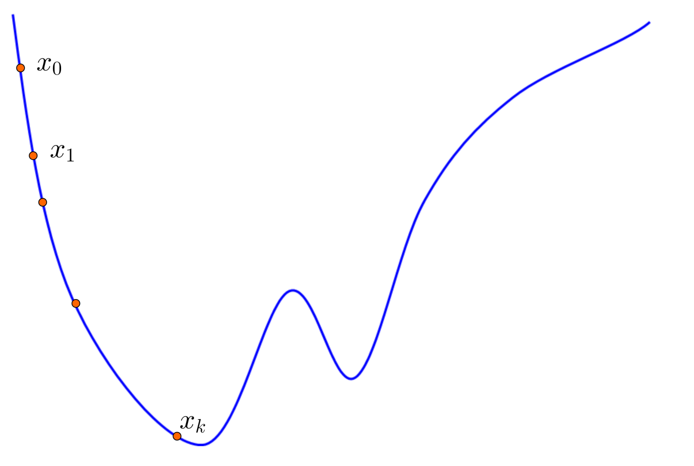
- 共轭梯度法对正定二次函数可以有限终止;
- 在最优解的附近，目标函数与一个正定二次函数很接近;
- 初始方向如果不取负梯度方向，即便应用于二次函数，也往往不能产生n 个共轭方向;

重开始的共轭梯度法

改进思路

- 共轭梯度法对正定二次函数可以有限终止;
- 在最优解的附近, 目标函数与一个正定二次函数很接近;
- 初始方向如果不取负梯度方向, 即便应用于二次函数, 也往往不能产生n 个共轭方向;
- 当迭代点接近解时, 重新取负梯度方向, 则接下来将产生近似共轭方向, 从而提高算法的效率.

重开始的共轭梯度法



重开始的共轭梯度法

对共轭梯度法进行修改:
每迭代n 次或$n+1$ 次, 就重新取共轭梯度方向为搜索方向.
改进后的算法称为n 步重新开始的共轭梯度法.

重开始的共轭梯度法

n 步重开始的FR 共轭梯度法

给定误差ϵ

【步 1】 给定初始点x_0, $k = 0$.

重开始的共轭梯度法

n 步重开始的FR 共轭梯度法

给定误差ϵ

【步 1】 给定初始点x_0, $k = 0$.

【步 2】 计算$g_k = g(x_k)$, 若$\|g_k\| \le \epsilon$, 则令$x^* = x_k$, 停; 否则, 转步3.

重开始的共轭梯度法

n 步重开始的FR 共轭梯度法

给定误差ϵ

【步 1】 给定初始点x_0, $k=0$.

【步 2】 计算$g_k=g(x_k)$, 若$\|g_k\|\le\epsilon$, 则令$x^*=x_k$, 停; 否则, 转步3.

【步 3】 若k 是n 的倍数, 则$d_k=-g_k$, 否则, 令

$$d_k=-g_k+\frac{g_k^Tg_k}{g_{k-1}^Tg_{k-1}}d_{k-1}.$$

重开始的共轭梯度法

n 步重开始的FR 共轭梯度法

给定误差ϵ

【步 1】 给定初始点x_0, $k=0$.

【步 2】 计算$g_k=g(x_k)$, 若$\|g_k\|\le\epsilon$, 则令$x^*=x_k$, 停; 否则, 转步3.

【步 3】 若k 是n 的倍数, 则$d_k=-g_k$, 否则, 令

$$d_k=-g_k+\frac{g_k^Tg_k}{g_{k-1}^Tg_{k-1}}d_{k-1}.$$

【步 4】 由精确一维搜索确定步长α_k, 满足

$$f(x_k+\alpha_kd_k)=\min_{\alpha>0}f(x_k+\alpha d_k).$$

重开始的共轭梯度法

n 步重开始的FR 共轭梯度法

给定误差ϵ

【步 1】 给定初始点x_0, $k=0$.

【步 2】 计算$g_k = g(x_k)$, 若$\|g_k\| \le \epsilon$, 则令$x^* = x_k$, 停; 否则, 转步3.

【步 3】 若k 是n 的倍数, 则$d_k = -g_k$, 否则, 令

$$d_k = -g_k + \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}} d_{k-1}.$$

【步 4】 由精确一维搜索确定步长α_k, 满足

$$f(x_k + \alpha_k d_k) = \min_{\alpha > 0} f(x_k + \alpha d_k).$$

【步 5】 令$x_{k+1} = x_k + \alpha_k d_k$, $k = k+1$, 转步2.

例题

例5.6

分别用PRP 和重开始PRP 共轭梯度法求四维$Rosenbrock$ 函数

$$
\begin{aligned}
f(x) =&100\left(x_1^2-x_2\right)^2+(1-x_1)^2+90\left(x_3^2-x_4\right)^2+(1-x_3)^2 \\
&+10.1\left[(x_2-1)^2+(x_4-1)^2\right]+19.8\,(x_2-1)\,(x_4-1)
\end{aligned}
$$

的最小值，初始点为$x_0=(-3,-1,-3,-1)^T$.

引言

引入拟Newton 法的目的：

引言

引入拟Newton 法的目的:

- 克服Newton 法的以下两个缺点:

引言

引入拟Newton 法的目的:

- 克服Newton 法的以下两个缺点:
  - 计算量大的问题(计算Hesse 矩阵);

引言

引入拟Newton 法的目的:

- 克服Newton 法的以下两个缺点:
  - 计算量大的问题(计算Hesse 矩阵);
  - Hesse 矩阵可能不正定

引言

引入拟Newton 法的目的:

- 克服Newton 法的以下两个缺点:
    - 计算量大的问题（计算Hesse 矩阵）;
    - Hesse 矩阵可能不正定
- 保持Newton法收敛速度快的优点.

引言

引入拟Newton 法的目的:

- 克服Newton 法的以下两个缺点:
  - 计算量大的问题（计算Hesse 矩阵）;
  - Hesse 矩阵可能不正定
- 保持Newton法收敛速度快的优点.

解决方案: Davidon(1959): 近似Hesse 矩阵.

拟Newton 法的基本思想

统一公式

最速下降法和阻尼Newton 法的迭代公式可以统一写成

$$x_{k+1} = x_k - \alpha_k H_k g_k,$$

其中, α_k 为步长, $g_k = \nabla f_k$, H_k 为n 阶对称矩阵.

拟Newton 法的基本思想

统一公式

最速下降法和阻尼Newton 法的迭代公式可以统一写成

$$x_{k+1} = x_k - \alpha_k H_k g_k,$$

其中, α_k 为步长, $g_k = \nabla f_k$, H_k 为n 阶对称矩阵.

- 最速下降法: $H_k = I$;

拟Newton 法的基本思想

统一公式

最速下降法和阻尼Newton 法的迭代公式可以统一写成

$$x_{k+1} = x_k - \alpha_k H_k g_k,$$

其中, α_k 为步长, $g_k = \nabla f_k$, H_k 为n 阶对称矩阵.

- 最速下降法: $H_k = I$;
- 阻尼Newton 法: $H_k = G_k^{-1}$;

拟Newton 法的基本思想

统一公式

最速下降法和阻尼Newton 法的迭代公式可以统一写成

$$x_{k+1} = x_k - \alpha_k H_k g_k,$$

其中, α_k 为步长, $g_k = \nabla f_k$, H_k 为n 阶对称矩阵.

- 最速下降法: $H_k = I$;
- 阻尼Newton 法: $H_k = G_k^{-1}$;
- L-M 方法: $H_k = (G_k + \mu_k I)^{-1}$.

启发:

启发：

若H_k 的选取既能逐步逼近G_k^{-1} 又不需要计算二阶导数，则算法有可能

启发：

若H_k 的选取既能逐步逼近G_k^{-1} 又不需要计算二阶导数，则算法有可能

- 比最速下降法快;

启发：

若H_k 的选取既能逐步逼近G_k^{-1} 又不需要计算二阶导数，则算法有可能

- 比最速下降法快；
- 比Newton 法计算简单，且整体（全局）收敛性好.

拟Newton 方程

模拟Newton 方向生成的途径. 将$f(x)$ 在x_{k+1} 处展开为二次函数

$$
\begin{aligned}
f(x) \approx & f\left(x_{k+1}\right)+g_{k+1}^{\mathrm{T}}\left(x-x_{k+1}\right) \\
& +\frac{1}{2}\left(x-x_{k+1}\right)^{\mathrm{T}} G_{k+1}\left(x-x_{k+1}\right) .
\end{aligned}
$$

拟Newton 方程

模拟Newton 方向生成的途径. 将$f(x)$ 在x_{k+1} 处展开为二次函数

$$f(x) \approx f\left(x_{k+1}\right)+g_{k+1}^{\mathrm{T}}\left(x-x_{k+1}\right) + \frac{1}{2}\left(x-x_{k+1}\right)^{\mathrm{T}} G_{k+1}\left(x-x_{k+1}\right).$$

对两边求导, 得

$$g(x) \approx g_{k+1}+G_{k+1}^{-1}\left(x-x_{k+1}\right).$$

拟Newton 方程

模拟Newton 方向生成的途径. 将$f(x)$ 在x_{k+1} 处展开为二次函数

$$\begin{aligned} f(x) \approx & f\left(x_{k+1}\right)+g_{k+1}^{\mathrm{T}}\left(x-x_{k+1}\right) \\ & +\frac{1}{2}\left(x-x_{k+1}\right)^{\mathrm{T}} G_{k+1}\left(x-x_{k+1}\right). \end{aligned}$$

对两边求导, 得

$$g(x) \approx g_{k+1}+G_{k+1}^{-1}\left(x-x_{k+1}\right).$$

令$x=x_k, s_k=x_{k+1}-x_k, y_k=g_{k+1}-g_k$, 得

$$G_{k+1}^{-1} y_k \approx s_k.$$

于是要求Hesse 矩阵的逆矩阵H_{k+1} 满足关系式

$$H_{k+1}y_k = s_k. \quad (\text{拟Newton 方程})$$

于是要求Hesse 矩阵的逆矩阵H_{k+1} 满足关系式

$$H_{k+1}y_k = s_k. \quad (\text{拟Newton 方程})$$

令$B_{k+1} = H_{k+1}^{-1}$, 则得到

$$B_{k+1}s_k = y_k,$$

其中, B_{k+1} 为Hesse 矩阵的近似.

H_k 须满足的条件

H_k 须满足的条件

(1) $\{H_k\}$ 都是对称正定矩阵;

H_k 须满足的条件

(1) $\{H_k\}$ 都是对称正定矩阵;

(2) H_{k+1} 由H_k 经简单形式修正而得,

$$H_{k+1} = H_k + E_k, \quad (\text{修正公式})$$

其中E_k 称为修正矩阵.

H_k 须满足的条件

(1) $\{H_k\}$ 都是对称正定矩阵;

(2) H_{k+1} 由H_k 经简单形式修正而得,

$$H_{k+1} = H_k + E_k, \qquad (\text{修正公式})$$

其中E_k 称为修正矩阵.

(3) H_{k+1} 满足拟Newton 方程

$$H_{k+1} y_k = s_k,$$

其中, $s_k = \alpha_k d_k = x_{k+1} - x_k$, $y_k = g_{k+1} - g_k$.

H_k 须满足的条件

(1) $\{H_k\}$ 都是对称正定矩阵;

(2) H_{k+1} 由H_k 经简单形式修正而得,

$$H_{k+1} = H_k + E_k, \qquad (\text{修正公式})$$

其中E_k 称为修正矩阵.

(3) H_{k+1} 满足拟Newton 方程

$$H_{k+1}y_k = s_k,$$

其中, $s_k = \alpha_k d_k = x_{k+1} - x_k$, $y_k = g_{k+1} - g_k$.

显然，拟Newton 方程有无穷多个解.

秩1 校正

- 一般形式

$$H_{k+1} = H_k + \sigma v v^T.$$

秩1 校正

- 一般形式

$$H_{k+1} = H_k + \sigma v v^T.$$

- 将它代入拟Newton 方程，有

$$H_{k+1} = H_k + \frac{(s_k - H_k y_k)(s_k - H_k y_k)^T}{(s_k - H_k y_k)^T y_k}$$

秩1 校正

- 一般形式
$$H_{k+1} = H_k + \sigma v v^T.$$
- 将它代入拟Newton 方程，有
$$H_{k+1} = H_k + \frac{(s_k - H_k y_k)(s_k - H_k y_k)^T}{(s_k - H_k y_k)^T y_k}$$
- 对于B_{k+1} 有校正公式
$$B_{k+1} = B_k + \frac{(y_k - B_k s_k)(y_k - B_k s_k)^T}{(y_k - B_k s_k)^T s_k}$$

秩1 校正算法

算法描述

给定控制误差$\epsilon > 0$.

步 1 给定初始点x_0, 初始矩阵H_0(通常取单位矩阵), 计算g_0, 令$k = 0$.

步 2 令$d_k = -H_k g_k$.

步 3 由精确一维搜索确定步长α_k,

$$f(x_k + \alpha_k d_k) = \min_{\alpha \geq 0} f(x_k + \alpha d_k),$$

步 4 令$x_{k+1} = x_k + \alpha_k d_k$.

步 5 若$\|g_{k+1}\| \leq \epsilon$, 则令$x^* = x_{k+1}$, 停; 否则, 令

$$s_k = x_{k+1} - x_k,\ y_k = g_{k+1} - g_k.$$

步 6 由秩1 校正公式得H_{k+1}. 令$k = k + 1$, 转步2.

秩1 校正

秩1 校正不能保证H_{k+1} 的正定性, (即使H_k 正定). 这是它的主要缺点. 但是它与信赖域方法结合有较好的数值表现. 对Hesse 矩阵及其逆矩阵的逼近程度较好.

例题

例6.1

用秩*1* 校正算法求

$$f(x_1, x_2) = x_1^2 + \frac{1}{2}x_2^2 + 3$$

的极小点. 初始点为$x_0 = (1, 2)^T$, $H_0 = I_2$.

例题

例6.1

用秩1 校正算法求

$$f(x_1, x_2) = x_1^2 + \frac{1}{2}x_2^2 + 3$$

的极小点. 初始点为$x_0 = (1, 2)^T$, $H_0 = I_2$.

f 可写为

$$f(x) = \frac{1}{2}x^T \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} x + 3.$$

例题

例6.1

用秩1 校正算法求

$$f(x_1, x_2) = x_1^2 + \frac{1}{2}x_2^2 + 3$$

的极小点. 初始点为$x_0 = (1, 2)^T$, $H_0 = I_2$.

f 可写为

$$f(x) = \frac{1}{2}x^T \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} x + 3.$$

则$G = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$. $g_k = \nabla f(x_k) = Gx_k = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} x_k$.

例题

第一次迭代.

例题

第一次迭代.

$$d_0 = -g_0 = \begin{pmatrix} -2 \\ -2 \end{pmatrix}.$$

例题

第一次迭代.

$$d_0 = -g_0 = \begin{pmatrix} -2 \\ -2 \end{pmatrix}.$$

由于f 是二次函数, 故而

$$\alpha_0 = -\frac{g_0^T d_0}{d_0^T G d_0} = \frac{(2,2)\begin{pmatrix} 2 \\ 2 \end{pmatrix}}{(2,2)\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 2 \\ 2 \end{pmatrix}} = \frac{2}{3}.$$

例题

第一次迭代.

$$d_0 = -g_0 = \begin{pmatrix} -2 \\ -2 \end{pmatrix}.$$

由于f 是二次函数, 故而

$$\alpha_0 = -\frac{g_0^T d_0}{d_0^T G d_0} = \frac{(2,2)\begin{pmatrix} 2 \\ 2 \end{pmatrix}}{(2,2)\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 2 \\ 2 \end{pmatrix}} = \frac{2}{3}.$$

则$x_1 = x_0 + \alpha d_0 = \begin{pmatrix} -\frac{1}{3} \\ \frac{2}{3} \end{pmatrix}$.

例题

秩1 校正.

例题

秩1 校正.

$$s_0 = \alpha_0 d_0 = \begin{pmatrix} -\frac{4}{3} \\ -\frac{4}{3} \end{pmatrix}, g_1 = Gx_1 = \begin{pmatrix} -\frac{2}{3} \\ \frac{2}{3} \end{pmatrix}$$

$$y_0 = g_1 - g_0 = \begin{pmatrix} -\frac{8}{3} \\ -\frac{4}{3} \end{pmatrix},$$

例题

秩1 校正.

$$s_0 = \alpha_0 d_0 = \begin{pmatrix} -\frac{4}{3} \\ -\frac{4}{3} \end{pmatrix}, g_1 = Gx_1 = \begin{pmatrix} -\frac{2}{3} \\ \frac{2}{3} \end{pmatrix}$$

$$y_0 = g_1 - g_0 = \begin{pmatrix} -\frac{8}{3} \\ -\frac{4}{3} \end{pmatrix},$$

$$H_1 = H_0 + \frac{(s_0 - H_0 y_0)(s_0 - H_0 y_0)^T}{(s_0 - H_0 y_0)^T y_0} = \begin{pmatrix} \frac{1}{2} & \\ & 1 \end{pmatrix} (= G^{-1}).$$

例题

第二次迭代.

例题

第二次迭代. 搜索方向为

$$d_1 = -H_1 g_1 = \begin{pmatrix} \frac{1}{3} \\ -\frac{2}{3} \end{pmatrix}$$

例题

第二次迭代. 搜索方向为

$$d_1 = -H_1 g_1 = \begin{pmatrix} \frac{1}{3} \\ -\frac{2}{3} \end{pmatrix}$$

步长为

$$\alpha_1 = -\frac{g_1^T d_1}{d_1^T G d_1} = 1.$$

例题

第二次迭代. 搜索方向为

$$d_1 = -H_1 g_1 = \begin{pmatrix} \frac{1}{3} \\ -\frac{2}{3} \end{pmatrix}$$

步长为

$$\alpha_1 = -\frac{g_1^T d_1}{d_1^T G d_1} = 1.$$

新的迭代点

$$x_2 = x_1 + \alpha_1 d_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

例题

第二次迭代. 搜索方向为

$$d_1 = -H_1 g_1 = \begin{pmatrix} \frac{1}{3} \\ -\frac{2}{3} \end{pmatrix}$$

步长为

$$\alpha_1 = -\frac{g_1^T d_1}{d_1^T G d_1} = 1.$$

新的迭代点

$$x_2 = x_1 + \alpha_1 d_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

此时$g_2 = (0,0)^T$, 也就是说$x^* = x_2$.

例题

例6.2

考虑目标函数

$$f(x) = (x_2 - x_1)^4 + 12x_1x_2 - x_1 + x_2 - 3.$$

取初始点$x_0 = (-0.5262, 0.6014)^T$, 初始矩阵为

$$H_0 = \begin{pmatrix} 0.1186 & -0.0376 \\ -0.0376 & 0.1191 \end{pmatrix}.$$

用秩*1* 校正计算H_1.

例题

例6.2

考虑目标函数

$$f(x) = (x_2 - x_1)^4 + 12x_1x_2 - x_1 + x_2 - 3.$$

取初始点$x_0 = (-0.5262, 0.6014)^T$, 初始矩阵为

$$H_0 = \begin{pmatrix} 0.1186 & -0.0376 \\ -0.0376 & 0.1191 \end{pmatrix}.$$

用秩1 校正计算H_1.

【解：】搜索方向

$$d_0 = -H_0 g_0 = \begin{pmatrix} -0.0413 \\ -0.0320 \end{pmatrix}$$

最优步长为

$$\alpha_0 = \arg\min_{\alpha \geq 0} f(x_0 + \alpha d_0) = 1.009.$$

最优步长为

$$\alpha_0 = \arg\min_{\alpha \geq 0} f(x_0 + \alpha d_0) = 1.009.$$

新的迭代点

$$x_1 = x_0 + \alpha_0 d_0 = \begin{pmatrix} -0.5679 \\ 0.5679 \end{pmatrix}. g_1 = \nabla f(x_1) = \begin{pmatrix} -0.0507 \\ 0.0656 \end{pmatrix}.$$

例题

秩1 校正

$$s_0 = \alpha_0 d_0 = \begin{pmatrix} -0.0417 \\ -0.0322 \end{pmatrix}, y_0 = \begin{pmatrix} -0.5326 \\ -0.3549 \end{pmatrix}.$$

例题

秩1 校正

$$s_0 = \alpha_0 d_0 = \begin{pmatrix} -0.0417 \\ -0.0322 \end{pmatrix}, y_0 = \begin{pmatrix} -0.5326 \\ -0.3549 \end{pmatrix}.$$

$$H_1 = H_0 + \frac{(s_0 - H_0 y_0)(s_0 - H_0 y_0)^T}{(s_0 - H_0 y_0)^T y_0} = \begin{pmatrix} 0.0331 & 0.0679 \\ 0.0679 & -0.0110 \end{pmatrix}.$$

例题

秩1 校正

$$s_0 = \alpha_0 d_0 = \begin{pmatrix} -0.0417 \\ -0.0322 \end{pmatrix}, y_0 = \begin{pmatrix} -0.5326 \\ -0.3549 \end{pmatrix}.$$

$$H_1 = H_0 + \frac{(s_0 - H_0 y_0)(s_0 - H_0 y_0)^T}{(s_0 - H_0 y_0)^T y_0} = \begin{pmatrix} 0.0331 & 0.0679 \\ 0.0679 & -0.0110 \end{pmatrix}.$$

由于H_1 的$(2,2)$ 元素为负数-0.0110, 故H_1 不是正定的.

例题

秩1 校正

$$s_0 = \alpha_0 d_0 = \begin{pmatrix} -0.0417 \\ -0.0322 \end{pmatrix}, y_0 = \begin{pmatrix} -0.5326 \\ -0.3549 \end{pmatrix}.$$

$$H_1 = H_0 + \frac{(s_0 - H_0 y_0)(s_0 - H_0 y_0)^T}{(s_0 - H_0 y_0)^T y_0} = \begin{pmatrix} 0.0331 & 0.0679 \\ 0.0679 & -0.0110 \end{pmatrix}.$$

由于H_1 的$(2,2)$ 元素为负数-0.0110, 故H_1 不是正定的.

【注：】本例印证了, 秩1 校正不能保持H_k 的正定性.

DFP 算法

DFP 是Davidon(1959)提出，后来Fletcher 和Powell(1963) 年做出改进，形成了Davidon-Fletcher-Powell 算法，简称DFP 算法. 它是第一个被提出的拟Newton 法，也是无约束最优化问题的最有效的算法之一.

算法

DFP 修正公式

思想:

算法

DFP 修正公式

思想：

- H_k 满足条件(1),(2),(3)

算法

DFP 修正公式

思想：

- H_k 满足条件(1),(2),(3)
- 修正项E_k 应有简单的形式. 考虑

$$E_k = \alpha u u^T + \beta v v^T.$$

其中, u, v 为待定向量.

算法

DFP 修正公式

思想：

- H_k 满足条件(1),(2),(3)
- 修正项E_k 应有简单的形式. 考虑

$$E_k = \alpha uu^T + \beta vv^T.$$

其中, u, v 为待定向量.

- 代入拟Newton 方程有

$$s_k = H_k y_k + \alpha uu^T y_k + \beta vv^T y_k.$$

DFP 算法

DFP 修正公式

- 取$u = s_k$, $v = H_k y_k$, $\alpha u^T y_k = 1$, $\beta v^T y_k = -1$, 则导出公式

DFP 算法

DFP 修正公式

- 取$u = s_k$, $v = H_k y_k$, $\alpha u^T y_k = 1$, $\beta v^T y_k = -1$, 则导出公式

$$H_{k+1} = H_k - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} + \frac{s_k s_k^T}{y_k^T s_k}, \quad \text{(DFP修正公式)}$$

DFP 算法

算法描述

给定控制误差$\epsilon > 0$.

步 1　给定初始点x_0, 初始矩阵H_0(通常取单位矩阵), 计算g_0, 令$k = 0$.

步 2　令$d_k = -H_k g_k$.

步 3　由精确一维搜索确定步长α_k,

$$f(x_k + \alpha_k d_k) = \min_{\alpha \geq 0} f(x_k + \alpha d_k),$$

步 4　令$x_{k+1} = x_k + \alpha_k d_k$.

步 5　若$\|g_{k+1}\| \leq \epsilon$, 则令$x^* = x_{k+1}$, 停; 否则, 令

$$s_k = x_{k+1} - x_k, \quad y_k = g_{k+1} - g_k.$$

步 6　由DFP 修正公式得H_{k+1}. 令$k = k + 1$, 转步2.

例题

例6.3

用 *DFP* 算法求解

$$\min\ \ f(x) = x_1^2 + 2x_2^2 - 2x_1x_2 - 4x_1,$$

取$x_0 = (1,1)^T$, $H_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

例题

例6.3

用 DFP 算法求解

$$\min \ \ f(x) = x_1^2 + 2x_2^2 - 2x_1x_2 - 4x_1,$$

取 $x_0 = (1,1)^T$, $H_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

【解：】 $g(x) = \begin{pmatrix} 2x_1 - 2x_2 - 4 \\ -2x_1 + 4x_2 \end{pmatrix}$, $g_0 = \begin{pmatrix} -4 \\ 2 \end{pmatrix}$.

$$d_0 = -H_0 g_0 = \begin{pmatrix} 4 \\ -2 \end{pmatrix}.$$

例题

(i) 求迭代点x_1, 令

$$\varphi_0(\alpha) = f(x_0 + \alpha d_0) = 40\alpha^2 - 20\alpha - 3,$$

例题

(i) 求迭代点x_1, 令

$$\varphi_0(\alpha) = f(x_0 + \alpha d_0) = 40\alpha^2 - 20\alpha - 3,$$

得$\varphi_0(\alpha)$ 的极小点为$\alpha_0 = \frac{1}{4}$(令$\varphi_0'(\alpha)=0$ 即可),

例题

(i) 求迭代点x_1, 令

$$\varphi_0(\alpha)=f(x_0+\alpha d_0)=40\alpha^2-20\alpha-3,$$

得$\varphi_0(\alpha)$ 的极小点为$\alpha_0=\frac{1}{4}$(令$\varphi_0'(\alpha)$=0 即可),所以

$$x_1=x_0+\alpha_0 d_0=(2,0.5)^T, g_1=(-1,-2)^T,$$
$$s_0=x_1-x_0=(1,-0.5)^T, y_0=g_1-g_0=(3,-4)^T.$$

例题

(i) 求迭代点x_1, 令

$$\varphi_0(\alpha) = f(x_0 + \alpha d_0) = 40\alpha^2 - 20\alpha - 3,$$

得$\varphi_0(\alpha)$ 的极小点为$\alpha_0 = \frac{1}{4}$(令$\varphi_0'(\alpha)$=0 即可),所以

$$x_1 = x_0 + \alpha_0 d_0 = (2, 0.5)^T, g_1 = (-1, -2)^T,$$
$$s_0 = x_1 - x_0 = (1, -0.5)^T, y_0 = g_1 - g_0 = (3, -4)^T.$$

于是，有DFP 修正公式有

$$H_1 = H_0 - \frac{H_0 y_0 y_0^T H_0}{y_0^T H_0 y_0} + \frac{s_0 s_0^T}{y_0^T s_0} = \frac{1}{100}\begin{pmatrix} 84 & 38 \\ 38 & 41 \end{pmatrix}.$$

例题

下一个搜索方向为

$$d_1 = -H_1 g_1 = \frac{1}{5}\begin{pmatrix} 8 \\ 6 \end{pmatrix}.$$

例题

下一个搜索方向为

$$d_1 = -H_1 g_1 = \frac{1}{5}\begin{pmatrix} 8 \\ 6 \end{pmatrix}.$$

求迭代点x_2，令

$$\varphi_1(\alpha) = f(x_1 + \alpha d_1) = \frac{8}{5}\alpha^2 - 4\alpha - 5.5,$$

例题

下一个搜索方向为

$$d_1 = -H_1 g_1 = \frac{1}{5}\begin{pmatrix} 8 \\ 6 \end{pmatrix}.$$

求迭代点x_2, 令

$$\varphi_1(\alpha) = f(x_1 + \alpha d_1) = \frac{8}{5}\alpha^2 - 4\alpha - 5.5,$$

其极小点为$\alpha_1 = \frac{5}{4}$, 于是

$$x_2 = x_1 + \alpha_1 d_1 = (4,2)^T \; g_2 = (0,0)^T.$$

例题

下一个搜索方向为

$$d_1 = -H_1 g_1 = \frac{1}{5}\begin{pmatrix} 8 \\ 6 \end{pmatrix}.$$

求迭代点x_2, 令

$$\varphi_1(\alpha) = f(x_1 + \alpha d_1) = \frac{8}{5}\alpha^2 - 4\alpha - 5.5,$$

其极小点为$\alpha_1 = \frac{5}{4}$, 于是

$$x_2 = x_1 + \alpha_1 d_1 = (4,2)^T \; g_2 = (0,0)^T.$$

所以, $x^* = x_2 = (4,2)^T$, 此时$f^* = -8$.

例题

因$Hesse$ 阵

$$G(x) = G = \begin{pmatrix} 2 & -2 \\ -2 & 4 \end{pmatrix}$$

为正定矩阵, $f(x)$ 为严格凸函数, 所以x^* 为整体（全局）极小点.

例题

因$Hesse$ 阵

$$G(x) = G = \begin{pmatrix} 2 & -2 \\ -2 & 4 \end{pmatrix}$$

为正定矩阵, $f(x)$ 为严格凸函数, 所以x^* 为整体（全局）极小点.

可以验证，如果再用一次DFP 修正公式, 则

$$H_2 = G^{-1} = \begin{pmatrix} 2 & -2 \\ -2 & 4 \end{pmatrix}^{-1}.$$

算法

据例题可知

算法

据例题可知

- 在更新H_k 的过程中，只需用到矩阵、向量之间的乘法，很容易计算;

算法

据例题可知

- 在更新H_k 的过程中，只需用到矩阵、向量之间的乘法，很容易计算；
- 计算搜索方向时，不需要像Newton 法一样求解线性方程组.

算法

据例题可知

- 在更新H_k 的过程中，只需用到矩阵、向量之间的乘法，很容易计算;
- 计算搜索方向时，不需要像Newton 法一样求解线性方程组.

所以DFP 算法(或拟Newton 法)相比Newton 法，每次迭代的计算量显著降低了.

算法

据例题可知

- 在更新H_k 的过程中，只需用到矩阵、向量之间的乘法，很容易计算；
- 计算搜索方向时，不需要像Newton 法一样求解线性方程组.

所以DFP 算法(或拟Newton 法)相比Newton 法，每次迭代的计算量显著降低了.

又由于H_k 逐渐逼近G_k^{-1}，从而在解的附近，算法的效率也近似于Newton 法

DFP 算法

DFP 算法的性质

DFP 算法

DFP 算法的性质

(1) 对于正定二次函数

DFP 算法

DFP 算法的性质

(1) 对于正定二次函数

- 至多经过n 次迭代即终止, 且$H_n = G^{-1}$;

DFP 算法

DFP 算法的性质

(1) 对于正定二次函数

- 至多经过n 次迭代即终止, 且$H_n = G^{-1}$;
- 保持拟Newton 方程

$$H_i y_j = s_j, \ j = 0, 1, \cdots, i-1$$

DFP 算法

DFP 算法的性质

(1) 对于正定二次函数

- 至多经过n 次迭代即终止，且$H_n = G^{-1}$;
- 保持拟Newton 方程

$$H_i y_j = s_j,\ \ j = 0, 1, \cdots, i-1$$

对于一般函数

DFP 算法

DFP 算法的性质

(1) 对于正定二次函数

- 至多经过n 次迭代即终止，且$H_n = G^{-1}$;
- 保持拟Newton 方程

$$H_i y_j = s_j, \ j = 0, 1, \cdots, i-1$$

对于一般函数

- 保持H_k 的正定性;

DFP 算法

DFP 算法的性质

(1) 对于正定二次函数

- 至多经过n 次迭代即终止, 且$H_n = G^{-1}$;
- 保持拟Newton 方程

$$H_i y_j = s_j, \ j = 0, 1, \cdots, i-1$$

对于一般函数

- 保持H_k 的正定性;
- 算法为超线性收敛的;

DFP 算法

DFP 算法的性质

(1) 对于正定二次函数

- 至多经过n 次迭代即终止，且$H_n = G^{-1}$;
- 保持拟Newton 方程

$$H_i y_j = s_j, \ j = 0, 1, \cdots, i-1$$

对于一般函数

- 保持H_k 的正定性;
- 算法为超线性收敛的;
- 对于凸函数是整体(全局)收敛的;

DFP 算法

DFP 算法的性质

(1) 对于正定二次函数

- 至多经过n 次迭代即终止, 且$H_n = G^{-1}$;
- 保持拟Newton 方程

$$H_i y_j = s_j, \ j = 0, 1, \cdots, i-1$$

对于一般函数

- 保持H_k 的正定性;
- 算法为超线性收敛的;
- 对于凸函数是整体(全局)收敛的;
- 计算量小.

正定继承性

引理2

设

$$H_+ = H - \frac{Hyy^T H}{y^T Hy} + \frac{ss^T}{y^T s},$$

H 为正定矩阵，则H_+ 正定的充分必要条件是

$$y^T s > 0.$$

证明:必要性.由H_+ 正定，$y \neq 0$ 知，$y^T H_+ y > 0$，又

$$y^T H_+ y = y^T \left[H - \frac{Hyy^T H}{y^T Hy} + \frac{ss^T}{y^T s} \right] y = y^T s,$$

故$y^T s > 0$.

充分性. 任取$0 \neq \boldsymbol{x} \in R^n$, 则

$$\begin{aligned} x^T H_+ x &= x^T H x - \frac{x^T H y y^T H x}{y^T H y} + \frac{(s^T x)^2}{y^T s} \\ &= \frac{x^T H x y^T H y - x^T H y y^T H x}{y^T H y} + \frac{(s^T x)^2}{y^T s}. \end{aligned}$$

因为H 正定, 所以存在可逆矩阵D, 使得$H = D^T D$. 令$u = Dx$, $v = Dy$,则

$$x^T H_+ x = \frac{u^T u v^T v - (u^T v)^2}{y^T H y} + \frac{(s^T x)^2}{y^T s}. \tag{10}$$

由Cauchy-schwartz 不等式

$$(u^T v)^2 \leq u^T u v^T v, \ \forall u, v \in R^n,$$

上式等号当且仅当$u = \beta v, \ \beta \neq 0$ 时成立. 因此(10) 右端第一项大于等于0, 且仅当$u = \beta v, \ \beta \neq 0$ 时等于0. 而此时(10) 右端第二项

$$\frac{(s^T x)^2}{y^T s} = \frac{(\beta s^T y)^2}{y^T s} = \beta^2 s^T y > 0.$$

故而总有$x^T H_+ x > 0$.

从而H_+ 正定.

定理12 (DFP 修正公式的正定继承性)

在DFP 算法中，如果初始矩阵H_0 正定，则整个矩阵列$\{H_k\}$ 都是正定的.

定理12 (DFP 修正公式的正定继承性)

在DFP 算法中，如果初始矩阵H_0 正定，则整个矩阵列$\{H_k\}$ 都是正定的.

证明: 用归纳法. 当$k=0$ 时，H_0 正定.
假设$k=i$ 时，H_i 是正定的，且$g_i\neq 0$(否则迭代终止)，要证H_{i+1} 也是正定的. 由引理2 知，只需证明$y_i^T s_i>0$. 由y_i, s_i 的定义

$$y_i^T s_i=(g_{i+1}-g_i)^T(-\alpha_i H_i g_i)=-\alpha_i g_{i+1}^T H_i g_i+\alpha_i g_i^T H_i g_i.$$

再由精确步长α_i 的性质知，$g_{i+1}^T H_i g_i=0$. 又因H_i 正定，$-H_i g_i$ 为下降方向，故$\alpha_i>0$，于是

$$y_i^T s_i=\alpha_i g_i^T H_i g_i>0.$$

从而H_{i+1} 正定.

二次终止性：DFP 的共轭性

定理13

将DFP 算法应用与二次函数

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

并设初始矩阵H_0 是正定的，产生的迭代点是互异的，并设搜索方向为$d_0, d_1, \cdots, d_k$，则

(i) $d_i^T G d_j = 0,\ 0 \le i < j \le k;$

(ii) $H_k y_i = s_i,\ 0 \le i \le k-1.$

DFP 方法的共轭性

证明:对k 用归纳法.
注意$s_i = x_{i+1} - x_i = \alpha_i d_i$, $y_i = g_{i+1} - g_i = G(x_{i+1} - x_i) = G s_i$.
当$k = 1$ 时, 由拟Newton 方程, 显然有$H_1 y_0 = s_0$, 且

$$\begin{aligned} d_0^T G d_1 &= (G d_0)^T d_1 = -\frac{1}{\alpha_0}(G s_0)^T H_1 g_1 \\ &= -\frac{1}{\alpha_0} y_0^T H_1 g_1 = -\frac{1}{\alpha_0} g_1^T s_0 = -g_1^T d_0 = 0. \end{aligned}$$

故$k = 1$ 时结论成立. 设$k = l$ 时结论成立, 要证$k = l + 1$ 时亦成立, 由归纳假设知

$$d_i^T G d_j = 0,\ 0 \le i < j \le l, \tag{11}$$

$$H_l y_i = s_i,\ 0 \le i \le l - 1. \tag{12}$$

当$k = l+1$ 时, 对于$0 \le i \le l-1$, 有

$$
\begin{aligned}
H_{l+1}y_i &= \left[H_l - \frac{H_l y_l y_l^T H_l}{y_l^T H_l y_l} + \frac{s_l s_l^T}{y_l^T s_l}\right] y_l \\
&= H_l y_l - \frac{H_l y_l y_l^T H_l y_i}{y_l^T H_l y_l} + \frac{s_l s_l^T y_i}{y_l^T s_l} \\
&= s_i - \frac{H_l y_l (y_l^T s_i)}{y_l^T H_l y_l} + \frac{s_l s_l^T G s_i}{y_l^T s_l} \\
&= s_i - \frac{H_l y_l (s_l^T G s_i)}{y_l^T H_l y_l} + \frac{s_l s_l^T G s_i}{y_l^T s_l}.
\end{aligned}
$$

由式(11) 及$s_i = \alpha_i d_i$ 知, 上式右端后两项和为0, 故而

$$H_{l+1}y_i = s_i \tag{13}$$

对于$0 \le i \le l-1$ 成立. 又由拟Newton 方程, $H_{l+1}y_l = s_l$. 所以(13) 对于$0 \le i \le l$ 成立.

下面再证

$$d_i^T G d_j,\ 0 \le i < j \le l+1. \tag{14}$$

由式(11) 知, 只需证明$d_i^T G d_{l+1} = 0$ 对$0 \le i \le l$ 成立. 事实上,

$$\begin{aligned} d_i^T G d_{l+1} = (Gd_i)^T d_{l+1} &= \frac{1}{\alpha_i} y_i^T (-H_{l+1} g_{l+1}) \\ &= -\frac{1}{\alpha_i} s_i^T g_{l+1} = -g_{l+1}^T d_i \\ &= -[g_{i+1} + \sum_{j=i+1}^{l} y_j]^T d_i \\ &= -g_{i+1}^T d_i - \sum_{j=i+1}^{l} (s_j^T G d_j) \\ &= 0. \end{aligned}$$

因此, (14) 对于$k = l+1$ 成立.

二次终止性

推论4 (二次终止性)

在上述定理的条件下，我们有

(i) 对于正定二次函数，*DFP* 算法至多迭代n 次就可得到极小点，即存在k_0, $(0 \leq k_0 \leq n)$，使得$x_{k_0} = x^*$;

(ii) 若$x_k \neq x^*$, $0 \leq k \leq n-1$，则$H_n = G^{-1}$.

证明:(i) 由定理13 知，DFP 算法是一种共轭方向法，因而结论(i) 成立.

(ii) 若对$\forall k \le n-1$, $x_k \ne x^*$, 则DFP 算法产生n 个共轭方向d_0, d_1, $\cdots$, d_{n-1}, 因而d_0, d_1, $\cdots$, d_{n-1} 线性无关, 故而s_0, s_1, $\cdots$, s_{n-1} 线性无关. 又由定理13, 有

$$H_n y_i = H_n G s_i = s_i,\ i = 0, 1, \cdots, n-1,$$

即

$$H_n G(s_0, s_1, \cdots, s_{n-1}) = (s_0, s_1, \cdots, s_{n-1}).$$

由$s_0, s_1, \cdots, s_{n-1}$ 线性无关知, $(s_0, s_1, \cdots, s_{n-1})$ 是非奇异矩阵, 所以$H_n G = I$, 即$H_n = G^{-1}$.

例题

例6.4

目标函数为

$$f(x)=\frac{x_1^4}{4}+\frac{x_2^2}{2}-x_1x_2+x_1-x_2.$$

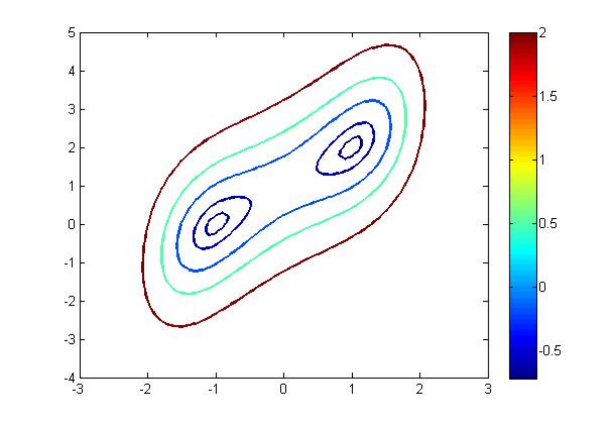
a. 利用$matlab$ 绘制函数f 在-0.72、-0.6、-0.2、0.5 和2 处的等高线图，根据图确定函数f 的极小点;
b. 用解析方法求$f(x)$ 的极小点;
c. 令起始点分别为$x^{(0)}=[0,0]^T$ 和$x^{(0)}=[1.5,1]^T$, $H_0=I_2$, 利用DFP 算法求函数f 的极小点. 观察在这两个起始点下，算法是否收敛到同一个点. 如果不是，试分析原因.

例题

【解：】

例题

【解：】 a. 如下图



例题

b. 令

$$\nabla f(x) = \begin{pmatrix} x_1^3 - x_2 + 1 \\ x_2 - x_1 - 1 \end{pmatrix} = 0,$$

例题

b. 令

$$\nabla f(x) = \begin{pmatrix} x_1^3 - x_2 + 1 \\ x_2 - x_1 - 1 \end{pmatrix} = 0,$$

即

$$\begin{cases} x_1^3 - x_2 + 1 = 0, \\ x_2 - x_1 - 1 = 0 \end{cases}$$

例题

b. 令

$$\nabla f(x) = \begin{pmatrix} x_1^3 - x_2 + 1 \\ x_2 - x_1 - 1 \end{pmatrix} = 0,$$

即

$$\begin{cases} x_1^3 - x_2 + 1 = 0, \\ x_2 - x_1 - 1 = 0 \end{cases}$$

解得

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \text{ 或 } \begin{pmatrix} -1 \\ 0 \end{pmatrix}$$

例题

f 的Hesse 矩阵为

$$H = \begin{pmatrix} 3x_1^2 & -1 \\ -1 & 1 \end{pmatrix}.$$

例题

f 的Hesse 矩阵为

$$H = \begin{pmatrix} 3x_1^2 & -1 \\ -1 & 1 \end{pmatrix}.$$

在这三点处, Hesse 矩阵分别为

$$H = \begin{pmatrix} 0 & -1 \\ -1 & 1 \end{pmatrix}, \begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix} \text{ 和 } \begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix}$$

例题

f 的Hesse 矩阵为

$$H = \begin{pmatrix} 3x_1^2 & -1 \\ -1 & 1 \end{pmatrix}.$$

在这三点处, Hesse 矩阵分别为

$$H = \begin{pmatrix} 0 & -1 \\ -1 & 1 \end{pmatrix}, \begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix} \text{ 和 } \begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix}$$

第一个矩阵不是正定的，后两者为正定矩阵，从而f 的极小点为

$$(1, 2)^T \text{ 和 } (-1, 0)^T.$$

例题

f 的Hesse 矩阵为

$$H = \begin{pmatrix} 3x_1^2 & -1 \\ -1 & 1 \end{pmatrix}.$$

在这三点处, Hesse 矩阵分别为

$$H = \begin{pmatrix} 0 & -1 \\ -1 & 1 \end{pmatrix}, \begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix} \text{ 和 } \begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix}$$

第一个矩阵不是正定的，后两者为正定矩阵，从而f 的极小点为

$$(1,2)^T \text{ 和 } (-1,0)^T.$$

而$(0,1)^T$ 为鞍点.

例题

c. （计算机演示）.

Broyden 族, BFGS 算法

Broyden 族修正公式(1967年)

$$H_{k+1}^{\varphi} = H_k - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} + \frac{s_k s_k^T}{y_k^T s_k} + \varphi w_k w_k^T,$$

其中, φ 为实数,

$$w_k = (y_k^T H_k y_k)^{\frac{1}{2}} \left(\frac{s_k}{y_k^T s_k} - \frac{H_k y_k}{y_k^T H_k y_k} \right).$$

Broyden 族, BFGS 算法

Broyden 族, BFGS 算法

- 显然, 取$\varphi = 0$ 时, 就是DFP 公式

$$H_{k+1}^{\varphi} = H_k - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} + \frac{s_k s_k^T}{y_k^T s_k}.$$

Broyden 族, BFGS 算法

- 显然, 取$\varphi = 0$ 时, 就是DFP 公式

$$H_{k+1}^{\varphi} = H_k - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} + \frac{s_k s_k^T}{y_k^T s_k}.$$

- 取$\varphi = y_k^T s_k / (s_k - H_k y_k)^T y_k$, 得

$$H_{k+1} = H_k - \frac{(s_k - H_k y_k)(s_k - H_k y_k)^T}{y_k^T (s_k - H_k y_k)} \quad (\text{秩1 修正公式}).$$

BFGS 公式

取$\varphi = 1$ 时，得

$$H_{k+1}^{\varphi} = H_k - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} + \frac{s_k s_k^T}{y_k^T s_k} + w_k w_k^T$$
$$= \left(I - \frac{s_k y_k^T}{s_k^T y_k}\right) H_k \left(I - \frac{y_k s_k^T}{s_k^T y_k}\right) + \frac{s_k s_k^T}{s_k^T y_k}.$$

称为BFGS (Broyden-Fletcher-Goldfarb-Shanno) 公式.

BFGS 公式

取$\varphi = 1$ 时, 得

$$
\begin{aligned}
H_{k+1}^{\varphi} &= H_k - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} + \frac{s_k s_k^T}{y_k^T s_k} + w_k w_k^T \\
&= \left(I - \frac{s_k y_k^T}{s_k^T y_k} \right) H_k \left(I - \frac{y_k s_k^T}{s_k^T y_k} \right) + \frac{s_k s_k^T}{s_k^T y_k}.
\end{aligned}
$$

称为BFGS (Broyden-Fletcher-Goldfarb-Shanno) 公式.

BFGS 算法是目前公认的最好的拟Newton 算法.

算例

例6.5

分别用*DFP* 方法，*BFGS* 方法求*Rosenbrock* 函数的极小点，初始点为$x_0 = (0, 0)^T$. 分别使用精确搜索和非精确搜索.

算例

例6.6

分别用*BFGS* 拟*Newton* 法，阻尼*Newton* 法，*FR*共轭梯度法，负梯度非精确搜索法极小化*Rosenbrock* 函数. 初始点为$x_0 = (-1.2, 1)^T$, 终止条件为$\|g(x_k)\| \leq 10^{-12}$.

【比较：】四种方法分别需要15，16，111，1334 次迭代.

表：最后十次迭代$\|x_k - x^*\|$

BFGS	阻尼Newton	FR	最速下降法
0.2976	1.2422	2.17E-10	2.42E-12
0.1329	0.9921	2.26E-10	2.42E-12
0.1307	0.8211	2.26E-10	2.42E-12
0.0365	0.5299	2.06E-10	2.42E-12
0.0098	0.3045	4.48E-12	2.42E-12
9.63E-04	0.1100	6.84E-12	6.13E-12
1.03E-05	0.0087	6.79E-11	2.41E-12
2.65E-08	6.24E-05	1.46E-12	2.41E-12
3.97E-13	1.90E-08	2.44E-13	2.41E-12
0	1.99E-15	2.36E-13	2.41E-12

算例及课堂练习

例6.7

求问题$f(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^T\boldsymbol{x}+\frac{\sigma}{4}(\boldsymbol{x}^TA\boldsymbol{x})^2$ 的极小值，其中$\sigma=10^4$,

$$A=\begin{pmatrix} 5 & 1 & 0 & 0.5 \\ 1 & 4 & 0.5 & 0 \\ 0 & 0.5 & 3 & 0 \\ 0.5 & 0 & 0 & 2 \end{pmatrix}$$

初始点为$x_0=(\cos 70°,\sin 70°,\cos 70°,\sin 70°)^T$.

(i) 证明：A 是正定矩阵;

(ii) 求$\nabla f(x)$;

(iii) 证明：f 有唯一极小点0.

一个神奇的性质

定理14 (Dixon 定理)

设$f(\boldsymbol{x})$ 为R^n 上连续可微函数，给定x_0，H_0，则由$Broyden$ 族算法产生的点列$\{\boldsymbol{x}_k\}$ 于参数φ 无关，即$Broyden$ 族产生相同的点列.

利用Dixon 定理，我们可以把DFP 算法的性质推广到Broyden 族算法. 比如具有二次终止性、整体收敛性和超线性收敛性等性质.

对Hesse 矩阵的近似B

- 将对H 的BFGS 校正公式中的s 和y 互换, 并将H 换成B, 得对Hesse 矩阵的DFP校正公式

$$B_{k+1} = \left(I - \frac{y_k s_k^T}{y_k^T s_k}\right) B_k \left(I - \frac{s_k y_k^T}{y_k^T s_k}\right) + \frac{y_k y_k^T}{y_k^T s_k}$$

- 将对H 的DFP 校正公式中的s 和y 互换, 并将H 换成B, 得对Hesse 矩阵的BFGS 校正公式

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{s_k^T y_k}.$$