Practical Optimization Algorithms
实用优化算法

徐 翔

数学科学学院
浙江大学

Mar 18, 2022

第三讲: 最速下降法和牛顿法

最速下降法

- $f(x)$在x_k处附近连续可微, $g_k = \nabla f(x_k) \neq 0$, Taylor展开为

$$f(x) = f(x_k) + \nabla^T f(x_k)(x - x_k) + \mathcal{O}(\|x - x_k\|^2)$$

最速下降法

- $f(x)$在x_k处附近连续可微，$g_k = \nabla f(x_k) \neq 0$，Taylor展开为

$$f(x) = f(x_k) + \nabla^T f(x_k)(x - x_k) + \mathcal{O}(\|x - x_k\|^2)$$

- 记$x - x_k = \alpha p_k$，满足$g_k^T p_k < 0$的方向是下降方向

最速下降法

- $f(x)$在x_k处附近连续可微, $g_k = \nabla f(x_k) \neq 0$, Taylor展开为

$$f(x) = f(x_k) + \nabla^T f(x_k)(x - x_k) + \mathcal{O}(\|x - x_k\|^2)$$

- 记$x - x_k = \alpha p_k$, 满足$g_k^T p_k < 0$的方向是下降方向
- 如果α相同, 在p_k取到$-g_k$时， $g_k^T p_k$取到最小, 即能使目标函数下降最多

最速下降法

算法1

给定$x_0 \in R^n$, $0 \le \varepsilon \ll 1$;

最速下降法

算法1

给定$x_0 \in R^n$, $0 \leq \varepsilon \ll 1$;
for $k = 0,\ 1,\ \cdots$

最速下降法

算法1

给定$x_0 \in R^n$, $0 \le \varepsilon \ll 1$;

for $k = 0,\ 1,\ \cdots$

计算搜索方向$p_k = -\nabla f(x_k)$;

最速下降法

算法1

给定$x_0 \in R^n$, $0 \le \varepsilon \ll 1$;

for $k = 0,\ 1,\ \cdots$

计算搜索方向$p_k = -\nabla f(x_k)$;

计算步长α_k, 使得 $f(x_k + \alpha_k p_k) = \min\limits_{\alpha \ge 0} f(x_k + \alpha p_k)$;

最速下降法

算法1

给定$x_0 \in R^n$, $0 \leq \varepsilon \ll 1$;

for $k = 0,\ 1,\ \cdots$

　　计算搜索方向$p_k = -\nabla f(x_k)$;

　　计算步长α_k, 使得 $f(x_k + \alpha_k p_k) = \min\limits_{\alpha \geq 0} f(x_k + \alpha p_k)$;

　　$x_{k+1} = x_k + \alpha_k p_k$;

最速下降法

算法1

给定$x_0 \in R^n$, $0 \leq \varepsilon \ll 1$;

for $k = 0,\ 1,\ \cdots$

　　计算搜索方向$p_k = -\nabla f(x_k)$;

　　计算步长α_k, 使得 $f(x_k + \alpha_k p_k) = \min\limits_{\alpha \geq 0} f(x_k + \alpha p_k)$;

　　$x_{k+1} = x_k + \alpha_k p_k$;

　　if $\|\nabla f(x_{k+1})\| \leq \varepsilon$

最速下降法

算法1

给定$x_0 \in R^n$, $0 \le \varepsilon \ll 1$;

for $k = 0,\ 1,\ \cdots$

  计算搜索方向$p_k = -\nabla f(x_k)$;

  计算步长α_k, 使得 $f(x_k + \alpha_k p_k) = \min_{\alpha \ge 0} f(x_k + \alpha p_k)$;

  $x_{k+1} = x_k + \alpha_k p_k$;

  if $\|\nabla f(x_{k+1})\| \le \varepsilon$

    stop;

最速下降法

算法1

给定$x_0 \in R^n$, $0 \le \varepsilon \ll 1$;

for $k = 0,\ 1,\ \cdots$

计算搜索方向$p_k = -\nabla f(x_k)$;

计算步长α_k, 使得 $f(x_k + \alpha_k p_k) = \min\limits_{\alpha \ge 0} f(x_k + \alpha p_k)$;

$x_{k+1} = x_k + \alpha_k p_k$;

if $\|\nabla f(x_{k+1})\| \le \varepsilon$

stop;

end (if)

end (for)

最速下降法

性质

最速下降法

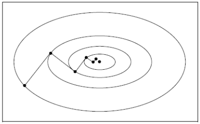
性质

- 负梯度方向是下降最快的方向，但仅是局部性质 。

最速下降法

性质

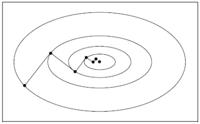
- 负梯度方向是下降最快的方向，但仅是局部性质。
- 如果采用精确线搜索方法，由于$\varphi'(\alpha_k) = 0$, 则会出现$p_{k+1}^T p_k = 0$. (当目标函数的等值线是一个扁长椭球时，会出现"锯齿现象"，下降十分缓慢)



最速下降法

性质

- 负梯度方向是下降最快的方向，但仅是局部性质。
- 如果采用精确线搜索方法，由于$\varphi'(\alpha_k) = 0$, 则会出现$p_{k+1}^T p_k = 0$. (当目标函数的等值线是一个扁长椭球时，会出现"锯齿现象"，下降十分缓慢)



- 总体是线性收敛的.

总体收敛性定理

Theorem (Zoutendijk)

总体收敛性定理

Theorem (Zoutendijk)

- 考虑任何一个线搜索算法，如果p_k是一个下降方向，步长因子α_k满足 Wolfe 条件.

总体收敛性定理

Theorem (Zoutendijk)

- 考虑任何一个线搜索算法，如果p_k是一个下降方向，步长因子α_k满足 Wolfe 条件.
- 假设$f(x)$ 在 $\mathcal{N} \equiv \{x| : f(x) \leq f(x_0)\}$上连续可微并且有下界，其中$x_0$ 是迭代起始点.

总体收敛性定理

Theorem (Zoutendijk)

- 考虑任何一个线搜索算法，如果p_k是一个下降方向，步长因子α_k满足 Wolfe 条件.
- 假设$f(x)$ 在 $\mathcal{N} \equiv \{x| : f(x) \leq f(x_0)\}$上连续可微并且有下界，其中$x_0$ 是迭代起始点.
- 假设 ∇f 在 $\mathcal{N}$ 上 Lipschitz 连续，即存在常数 $L > 0$ 使得

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \quad \forall x, y \in \mathcal{N}. \tag{2.1}$$

总体收敛性定理

Theorem (Zoutendijk)

- 考虑任何一个线搜索算法，如果p_k是一个下降方向，步长因子α_k满足 Wolfe 条件.
- 假设$f(x)$ 在 $\mathcal{N} \equiv \{x| : f(x) \le f(x_0)\}$上连续可微并且有下界，其中$x_0$ 是迭代起始点.
- 假设 ∇f 在 $\mathcal{N}$ 上 Lipschitz 连续，即存在常数 $L > 0$ 使得

$$\|\nabla f(x) - \nabla f(y)\| \le L\|x - y\|, \quad \forall x, y \in \mathcal{N}. \tag{2.1}$$

- **Then**

$$\sum_{k\ge 0} \cos^2(\theta_k)\|\nabla f(x_k)\|^2 < \infty \tag{2.2}$$

which is called Zoutendijk condition.

线搜索法收敛性

Remark

线搜索法收敛性

Remark

- 如果使用 Goldstein 条件 strong Wolfe 条件，类似结果也成立

线搜索法收敛性

Remark

- 如果使用 Goldstein 条件 strong Wolfe 条件，类似结果也成立
- The Zoutendijk condition (2.2) implies that

$$\cos^2(\theta_k)\|\nabla f(x_k)\|^2 \to 0. \tag{2.3}$$

线搜索法收敛性

Remark

- 如果使用 Goldstein 条件 strong Wolfe 条件，类似结果也成立
- The Zoutendijk condition (2.2) implies that

$$\cos^2(\theta_k)\|\nabla f(x_k)\|^2 \to 0. \tag{2.3}$$

- 可以推导 global convergence 结果.

总体收敛性定理

Remark

总体收敛性定理

REMARK

- 如果 p_k 满足与$-\nabla f(x_k)$的夹角 $\theta_k < 90°$，即存在正常数$\delta > 0$，

$$\cos\theta_k \geq \delta > 0, \forall k \tag{2.4}$$

It follows immediately from (2.3) that

$$\lim_{k\to\infty} \|\nabla f(x_k)\| = 0. \tag{2.5}$$

总体收敛性定理

Remark

- 如果 p_k 满足与$-\nabla f(x_k)$的夹角 $\theta_k < 90°$，即存在正常数$\delta > 0$，

$$\cos\theta_k \geq \delta > 0, \forall k \tag{2.4}$$

  It follows immediately from (2.3) that

$$\lim_{k\to\infty} \|\nabla f(x_k)\| = 0. \tag{2.5}$$

- 即$\|\nabla f(x_k)\| \to 0$，只要p_k不靠近 $-\nabla f(x_k)$的正交方向。

最速下降法收敛速度

首先考虑二次函数$f(x) = x^T Gx$, 其中G对称正定。

二次函数最速下降法的收敛速度定理

假设λ_1和λ_n是G的最大和最小特征值. 设x^*是问题的解, 则最速下降法的收敛,速度至少是线性的, 并且满足

$$\frac{f(x_{k+1}) - f(x^*)}{f(x_k) - f(x^*)} \le \frac{(\kappa-1)^2}{(\kappa+1)^2} = \frac{(\lambda_1-\lambda_n)^2}{(\lambda_1+\lambda_n)^2}, \tag{2.6}$$

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} \le \sqrt{\kappa}\frac{(\kappa-1)}{(\kappa+1)}, \tag{2.7}$$

其中$\kappa = \lambda_1/\lambda_n \ge 1$.

最速下降法收敛速度

一般函数最速下降法的收敛速度定理

设$f(x)$在x^*附近二次连续可微，$\nabla f(x^*) = 0$，$\nabla^2 f(x^*)$正定，则

$$\frac{f(x_{k+1}) - f(x^*)}{f(x_k) - f(x^*)} = \beta_k < 1, \tag{2.8}$$

$$\limsup_{k\to+\infty} \beta_k \le \frac{M - m}{M} < 1, \tag{2.9}$$

其中M, m 满足$0 < m \le \lambda_n \le \lambda_1 \le M$，$\lambda_n$和$\lambda_1$是$\nabla^2 f(x)$ 的最小和最大特征值.

梯度法推广

两点步长梯度法

梯度法推广

两点步长梯度法

- 基本思想： $x_{k+1} = x_k - D_k \nabla f(x_k)$ 其中 $D_k = \alpha_k I$.

梯度法推广

两点步长梯度法

- 基本思想： $x_{k+1} = x_k - D_k \nabla f(x_k)$ 其中 $D_k = \alpha_k I$.
- 可以使D_k具有拟牛顿性质，即计算α_k使得

$$\min \ \|s_{k-1} - D_k y_{k-1}\| \text{ 或者 } \min \ \|D_k^{-1} s_{k-1} - y_{k-1}\| \tag{2.10}$$

其中 $s_{k-1} = x_k - x_{k-1}$, $y_{k-1} = \nabla f(x_k) - \nabla f(x_{k-1})$

梯度法推广

两点步长梯度法

- 基本思想： $x_{k+1} = x_k - D_k \nabla f(x_k)$ 其中 $D_k = \alpha_k I$.
- 可以使D_k具有拟牛顿性质，即计算α_k使得

$$\min \ \|s_{k-1} - D_k y_{k-1}\| \ \text{或者} \ \min \ \|D_k^{-1} s_{k-1} - y_{k-1}\| \tag{2.10}$$

其中 $s_{k-1} = x_k - x_{k-1}$, $y_{k-1} = \nabla f(x_k) - \nabla f(x_{k-1})$

- 可以求出α_k，

$$\alpha_k = \frac{s_{k-1}^T y_{k-1}}{\|y_{k-1}\|^2}, \ \text{或者} \ \alpha_k = \frac{\|s_{k-1}\|^2}{s_{k-1}^T y_{k-1}}$$

两点步长梯度法

The Barzilai-Borwein gradient method

给定$x_0 \in R^n$, $0 \leq \varepsilon \ll 1$;

两点步长梯度法

The Barzilai-Borwein gradient method

给定$x_0 \in R^n$, $0 \leq \varepsilon \ll 1$;

for $k = 0,\ 1,\ \cdots$

两点步长梯度法

The Barzilai-Borwein gradient method

给定$x_0 \in R^n$, $0 \leq \varepsilon \ll 1$;

for $k = 0,\ 1,\ \cdots$

计算搜索方向$p_k = -\nabla f(x_k)$;

两点步长梯度法

The Barzilai-Borwein gradient method

给定$x_0 \in R^n$, $0 \le \varepsilon \ll 1$;

for $k = 0,\ 1,\ \cdots$

计算搜索方向$p_k = -\nabla f(x_k)$;

计算步长α_k，如果$k = 0$, 进行线搜索,

如果$k \neq 0$, $\alpha_k = \frac{s_{k-1}^T y_{k-1}}{\|y_{k-1}\|^2}$, 或者 $\alpha_k = \frac{\|s_{k-1}\|^2}{s_{k-1}^T y_{k-1}}$;

两点步长梯度法

The Barzilai-Borwein gradient method

给定$x_0 \in R^n$, $0 \le \varepsilon \ll 1$;

for $k = 0,\ 1,\ \cdots$

计算搜索方向$p_k = -\nabla f(x_k)$;

计算步长α_k，如果$k = 0$, 进行线搜索,

如果$k \neq 0$, $\alpha_k = \frac{s_{k-1}^T y_{k-1}}{\|y_{k-1}\|^2}$, 或者 $\alpha_k = \frac{\|s_{k-1}\|^2}{s_{k-1}^T y_{k-1}}$;

$x_{k+1} = x_k + \alpha_k p_k$;

两点步长梯度法

The Barzilai-Borwein gradient method

给定$x_0 \in R^n$, $0 \le \varepsilon \ll 1$;

for $k = 0,\ 1,\ \cdots$

计算搜索方向$p_k = -\nabla f(x_k)$;

计算步长α_k，如果$k = 0$, 进行线搜索,

如果$k \neq 0$, $\alpha_k = \frac{s_{k-1}^T y_{k-1}}{\|y_{k-1}\|^2}$, 或者 $\alpha_k = \frac{\|s_{k-1}\|^2}{s_{k-1}^T y_{k-1}}$;

$x_{k+1} = x_k + \alpha_k p_k$;

if $\|\nabla f(x_{k+1})\| \le \varepsilon$

两点步长梯度法

The Barzilai-Borwein gradient method

给定$x_0 \in R^n$, $0 \leq \varepsilon \ll 1$;

for $k = 0,\ 1,\ \cdots$

计算搜索方向$p_k = -\nabla f(x_k)$;

计算步长α_k，如果$k = 0$, 进行线搜索,

如果$k \neq 0$, $\alpha_k = \frac{s_{k-1}^T y_{k-1}}{\|y_{k-1}\|^2}$, 或者 $\alpha_k = \frac{\|s_{k-1}\|^2}{s_{k-1}^T y_{k-1}}$;

$x_{k+1} = x_k + \alpha_k p_k$;

if $\|\nabla f(x_{k+1})\| \leq \varepsilon$

stop;

两点步长梯度法

The Barzilai-Borwein gradient method

给定$x_0 \in R^n$, $0 \le \varepsilon \ll 1$;

for $k = 0,\ 1,\ \cdots$

计算搜索方向$p_k = -\nabla f(x_k)$;

计算步长α_k，如果$k = 0$, 进行线搜索,

如果$k \neq 0$, $\alpha_k = \frac{s_{k-1}^T y_{k-1}}{\|y_{k-1}\|^2}$, 或者 $\alpha_k = \frac{\|s_{k-1}\|^2}{s_{k-1}^T y_{k-1}}$;

$x_{k+1} = x_k + \alpha_k p_k$;

if $\|\nabla f(x_{k+1})\| \le \varepsilon$

stop;

end (if)

end (for)

两点步长梯度法

性质

两点步长梯度法

性质

- 不需要做线搜索(k=0除外)，

两点步长梯度法

性质

- 不需要做线搜索(k=0除外)，
- 没有矩阵乘以向量运算，因此计算量小，

两点步长梯度法

性质

- 不需要做线搜索(k=0除外)，
- 没有矩阵乘以向量运算，因此计算量小，
- 本质上是梯度法，但是收敛速度要更快.

两点步长梯度法

性质

- 不需要做线搜索(k=0除外)，
- 没有矩阵乘以向量运算，因此计算量小，
- 本质上是梯度法，但是收敛速度要更快.
- Barzilai and Borwein 证明了对于二次目标函数问题，该算法是 R-superlinearly收敛.

两点步长梯度法

性质

- 不需要做线搜索(k=0除外)，
- 没有矩阵乘以向量运算，因此计算量小，
- 本质上是梯度法，但是收敛速度要更快.
- Barzilai and Borwein 证明了对于二次目标函数问题，该算法是R-superlinearly收敛.
- 对于非二次目标函数，α_k有可能很大或很小，因此我们需要给α_k设置上下界.

牛顿法

$f(x)$二次可微, $x_k \in R^n$, Hessian$\nabla^2 f(x_k)$正定，有二次Taylor展开

$$f(x_k + s) \approx f(x_k) + \nabla f(x_k)^T s + \frac{1}{2} s^T \nabla^2 f(x_k) s. \tag{2.11}$$

牛顿法

$f(x)$二次可微，$x_k \in R^n$，Hessian$\nabla^2 f(x_k)$正定，有二次Taylor展开

$$f(x_k + s) \approx f(x_k) + \nabla f(x_k)^T s + \frac{1}{2} s^T \nabla^2 f(x_k) s. \tag{2.11}$$

极小化二次近似函数，可得到$s = -[\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$，于是

$$x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k) \tag{2.12}$$

记$p_k = -[\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$，$\alpha_k = 1$，则迭代格式可以写为$x_{k+1} = x_k + p_k$.

牛顿法

Remark

记$\nabla^2 f(x_k) = G_k$，$g_k = \nabla f(x_k)$.
迭代格式记为$x_{k+1} = x_k - G_k^{-1} g_k$

牛顿法

Remark

记$\nabla^2 f(x_k) = G_k$，$g_k = \nabla f(x_k)$.
迭代格式记为$x_{k+1} = x_k - G_k^{-1} g_k$

- 牛顿法可以看作在椭球范数$\|\cdot\|_{G_k}$下的最速下降法.

牛顿法

Remark

记$\nabla^2 f(x_k) = G_k$，$g_k = \nabla f(x_k)$.
迭代格式记为$x_{k+1} = x_k - G_k^{-1} g_k$

- 牛顿法可以看作在椭球范数$\|\cdot\|_{G_k}$下的最速下降法.
- 对于$f(x_k + s) \approx f(x_k) + g_k^T s$，$s$是如下极小化问题的解

$$\min_{s \in R^n} \frac{g_k^T s}{\|s\|}$$

  - 如果范数采用l_2，则$s_k = -g_k$.
  - 如果范数采用$\|\cdot\|_{G_k}$，
    则原问题等价于 $\min_{s \in R^n} \; g_k^T s$ where $\|s\|_{G_k} \leq 1$.
    由于$(g_k^T s)^2 \leq (g_k^T G_k^{-1} g_k)(s^T G_k s)$，则 $s_k = -G_k^{-1} g_k$.

牛顿法收敛性

如果f是二次函数，则牛顿迭代法一步就达到最优解.

牛顿法收敛性

如果f是二次函数，则牛顿迭代法一步就达到最优解.

牛顿法收敛定理

设f二次连续可微, x^*充分靠近x^*, $\nabla f(x^*)=0$, 如果$\nabla^2 f(x^*)$正定,且Hessian矩阵$G(x)$满足Lipschitz条件，即存在$L>0$, 对所有的(i,j),

$$|G_{ij}(x)-G_{ij}(y)| \leq L\|x-y\|. \tag{2.13}$$

则对一切k, 牛顿迭代法得到的序列x_k收敛到x^*, 并且具有二次收敛速度.

牛顿法

Remark G_k不正定

在牛顿法中可以增加步长因子的搜索过程，使得当G_k不正定的时候也收敛.

$$p_k = -G_k^{-1} g_k, \; x_{k+1} = x_k + \alpha_k p_k \tag{2.14}$$

牛顿法

Remark G_k不正定

在牛顿法中可以增加步长因子的搜索过程，使得当G_k不正定的时候也收敛.

$$p_k = -G_k^{-1} g_k, \ x_{k+1} = x_k + \alpha_k p_k \tag{2.14}$$

算法2(带步长因子的牛顿法)

给定$x_0 \in R^n$, $0 \le \varepsilon \ll 1$;

牛顿法

REMARK G_k不正定

在牛顿法中可以增加步长因子的搜索过程，使得当G_k不正定的时候也收敛.

$$p_k = -G_k^{-1} g_k,\ x_{k+1} = x_k + \alpha_k p_k \tag{2.14}$$

算法2(带步长因子的牛顿法)

给定$x_0 \in R^n$, $0 \le \varepsilon \ll 1$;

for $k = 0,\ 1,\ \cdots$

牛顿法

Remark G_k不正定

在牛顿法中可以增加步长因子的搜索过程，使得当G_k不正定的时候也收敛.

$$p_k = -G_k^{-1} g_k,\ x_{k+1} = x_k + \alpha_k p_k \tag{2.14}$$

算法2(带步长因子的牛顿法)

给定$x_0 \in R^n$, $0 \le \varepsilon \ll 1$;
for $k = 0,\ 1,\ \cdots$
　　计算搜索方向$p_k = -[\nabla^2 f(x_k)]^{-1} f(x_k)$;

牛顿法

REMARK G_k不正定

在牛顿法中可以增加步长因子的搜索过程，使得当G_k不正定的时候也收敛.

$$p_k = -G_k^{-1} g_k,\ x_{k+1} = x_k + \alpha_k p_k \tag{2.14}$$

算法2(带步长因子的牛顿法)

给定$x_0 \in R^n$, $0 \leq \varepsilon \ll 1$;

for $k = 0,\ 1,\ \cdots$

计算搜索方向$p_k = -[\nabla^2 f(x_k)]^{-1} f(x_k)$;

计算步长α_k, 使得 $f(x_k + \alpha_k p_k) = \min\limits_{\alpha \geq 0} f(x_k + \alpha p_k)$;

牛顿法

Remark G_k不正定

在牛顿法中可以增加步长因子的搜索过程，使得当G_k不正定的时候也收敛.

$$p_k = -G_k^{-1} g_k, \ x_{k+1} = x_k + \alpha_k p_k \tag{2.14}$$

算法2(带步长因子的牛顿法)

给定$x_0 \in R^n$, $0 \le \varepsilon \ll 1$;

for $k = 0, \ 1, \ \cdots$

计算搜索方向$p_k = -[\nabla^2 f(x_k)]^{-1} f(x_k)$;

计算步长α_k, 使得 $f(x_k + \alpha_k p_k) = \min_{\alpha \ge 0} f(x_k + \alpha p_k)$;

$x_{k+1} = x_k + \alpha_k p_k$;

牛顿法

REMARK G_k不正定

在牛顿法中可以增加步长因子的搜索过程，使得当G_k不正定的时候也收敛.

$$p_k = -G_k^{-1} g_k,\ x_{k+1} = x_k + \alpha_k p_k \tag{2.14}$$

算法2(带步长因子的牛顿法)

给定$x_0 \in R^n$, $0 \leq \varepsilon \ll 1$;

for $k = 0,\ 1,\ \cdots$

  计算搜索方向$p_k = -[\nabla^2 f(x_k)]^{-1} f(x_k)$;

  计算步长α_k, 使得 $f(x_k + \alpha_k p_k) = \min\limits_{\alpha \geq 0} f(x_k + \alpha p_k)$;

  $x_{k+1} = x_k + \alpha_k p_k$;

  if $\|\nabla f(x_{k+1})\| \leq \varepsilon$

牛顿法

Remark G_k不正定

在牛顿法中可以增加步长因子的搜索过程，使得当G_k不正定的时候也收敛.

$$p_k = -G_k^{-1} g_k,\ x_{k+1} = x_k + \alpha_k p_k \tag{2.14}$$

算法2(带步长因子的牛顿法)

给定$x_0 \in R^n$, $0 \le \varepsilon \ll 1$;

for $k = 0,\ 1,\ \cdots$

  计算搜索方向$p_k = -[\nabla^2 f(x_k)]^{-1} f(x_k)$;

  计算步长α_k, 使得 $f(x_k + \alpha_k p_k) = \min_{\alpha \ge 0} f(x_k + \alpha p_k)$;

  $x_{k+1} = x_k + \alpha_k p_k$;

  if $\|\nabla f(x_{k+1})\| \le \varepsilon$

    stop;

牛顿法

Remark G_k不正定

在牛顿法中可以增加步长因子的搜索过程，使得当G_k不正定的时候也收敛.

$$p_k = -G_k^{-1} g_k,\ x_{k+1} = x_k + \alpha_k p_k \tag{2.14}$$

算法2(带步长因子的牛顿法)

给定$x_0 \in R^n$, $0 \le \varepsilon \ll 1$;
for $k = 0,\ 1,\ \cdots$
  计算搜索方向$p_k = -[\nabla^2 f(x_k)]^{-1} f(x_k)$;
  计算步长α_k, 使得 $f(x_k + \alpha_k p_k) = \min_{\alpha \ge 0} f(x_k + \alpha p_k)$;
  $x_{k+1} = x_k + \alpha_k p_k$;
  if $\|\nabla f(x_{k+1})\| \le \varepsilon$
    stop;
  end (if)
end (for)

带步长因子牛顿法收敛性

定理

设f在开凸集D中二阶连续可微，又设对任意的$x_0 \in D$, 存在常数$m > 0$, 使得$f(x)$在水平集$L(x_0) = \{x|f(x) \leq f(x_0)\}$ 上满足

$$u^T \nabla^2 f(x)u \geq m\|u\|^2, \forall u \in R^n, x \in L(x_0). \tag{2.15}$$

则在精确一维搜索条件下，带步长因子的牛顿法产生的迭代点列$\{x_k\}$满足

1. 当x_k为有限点列， 则对某个k， $\nabla f(x_k) = 0.$
2. 当x_k为无穷点列，$\{x_k\}$收敛到f的唯一极小值点.

GOLDSTEIN-PRICE修正牛顿法

- 如果G_k不正定，可以用$p_k = -\nabla f(x_k)$代替$p_k^N = -[\nabla^2 f(x_k)]^{-1}\nabla f(x_k)$，因此也会适当引入步长因子.

Goldstein-Price修正牛顿法

- 如果G_k不正定，可以用$p_k = -\nabla f(x_k)$代替$p_k^N = -[\nabla^2 f(x_k)]^{-1}\nabla f(x_k)$，因此也会适当引入步长因子.
- 给定阈值$\eta > 0$

$$p_k = \begin{cases} -G_k^{-1} g_k, & if \cos\theta \geq \eta \\ -g_k, & otherwise. \end{cases}$$

Goldfeld修正牛顿法

Algorithm 3 (修正牛顿法)

给定初始猜测点 x_0;

GOLDFELD修正牛顿法

ALGORITHM 3 (修正牛顿法)

给定初始猜测点 x_0;

For $k = 0, 1, 2, \cdots$

构造 矩阵 $\bar{G}_k = \nabla^2 f(x_k) + v_k I$,

其中 $v_k = 0$ 如 $\nabla^2 f(x_k)$ 正定;

otherwise, 选择 $v_k > 0$ 使得 $\bar{G}_k$ 正定;

Goldfeld修正牛顿法

Algorithm 3 (修正牛顿法)

给定初始猜测点 x_0;

For $k = 0, 1, 2, \cdots$

构造 矩阵 $\bar{G}_k = \nabla^2 f(x_k) + v_k I$,

其中 $v_k = 0$ 如 $\nabla^2 f(x_k)$ 正定;

otherwise, 选择 $v_k > 0$ 使得 $\bar{G}_k$ 正定;

求解 $\bar{G}_k p_k = -\nabla f(x_k)$;

GOLDFELD修正牛顿法

ALGORITHM 3 (修正牛顿法)

给定初始猜测点 x_0;

For $k = 0, 1, 2, \cdots$

  构造 矩阵 $\bar{G}_k = \nabla^2 f(x_k) + v_k I$,

   其中 $v_k = 0$ 如 $\nabla^2 f(x_k)$ 正定;

   otherwise, 选择 $v_k > 0$ 使得 $\bar{G}_k$ 正定;

  求解 $\bar{G}_k p_k = -\nabla f(x_k)$;

  令 $x_{k+1} \leftarrow x_k + \alpha_k p_k$,

   其中 α_k 满足 Wolfe, Goldstein, or strong Wolfe 条件;

End

GOLDFELD修正牛顿法

当$\nabla^2 f(x_k)$不定时，如何选取v_k?

GOLDFELD修正牛顿法

当$\nabla^2 f(x_k)$不定时， 如何选取v_k?

- 应该大于G_k的最负特征值的绝对值.

GOLDFELD修正牛顿法

当$\nabla^2 f(x_k)$不定时，如何选取v_k?

- 应该大于G_k的最负特征值的绝对值.
- 按照Gill-Murray的修改Cholesky分解$(\bar{G}_k = \bar{L}_k \bar{D}_k \bar{L}_k^T)$算法确定$v_k$.

GOLDFELD修正牛顿法

当$\nabla^2 f(x_k)$不定时， 如何选取v_k?

- 应该大于G_k的最负特征值的绝对值.
- 按照Gill-Murray的修改Cholesky分解($\bar{G}_k = \bar{L}_k \bar{D}_k \bar{L}_k^T$)算法确定$v_k$.
- $v_k = \min\{b_1, b_2\}$其中

$$b_1 = \left| \min_{1 \le i \le n} \left\{ (G_k)_{ii} - \sum_{j \ne i} |(G_k)_{ij}| \right\} \right| \ge \left| \min_i \lambda_i \right| \tag{2.16}$$

$$b_2 = \max_i \{e_{ii}\}, \text{ 其中} e_{ii}\text{是}E\text{的第}i\text{个对角元}, \tag{2.17}$$

E是G_kGill-Murray修正Cholesky分解:$G_k + E = LDL^T$

Goldfeld修正牛顿法

当$\nabla^2 f(x_k)$不定时， 如何选取v_k?

- 应该大于G_k的最负特征值的绝对值.
- 按照Gill-Murray的修改Cholesky分解($\bar{G}_k = \bar{L}_k \bar{D}_k \bar{L}_k^T$)算法确定$v_k$.
- $v_k = \min\{b_1, b_2\}$其中

$$b_1 = \left| \min_{1 \le i \le n} \left\{ (G_k)_{ii} - \sum_{j \ne i} |(G_k)_{ij}| \right\} \right| \ge \left| \min_i \lambda_i \right| \tag{2.16}$$

$$b_2 = \max_i \{e_{ii}\}, \text{ 其中} e_{ii} \text{是} E \text{的第} i \text{个对角元,} \tag{2.17}$$

E是G_kGill-Murray修正Cholesky分解:$G_k + E = LDL^T$

- 如何对$G_k + E$做Cholesky分解?

Goldfeld修正牛顿法

当$\nabla^2 f(x_k)$不定时， 如何选取v_k?

- 应该大于G_k的最负特征值的绝对值.
- 按照Gill-Murray的修改Cholesky分解$(\bar{G}_k = \bar{L}_k \bar{D}_k \bar{L}_k^T)$算法确定$v_k$.
- $v_k = \min\{b_1, b_2\}$其中

$$b_1 = \left| \min_{1 \le i \le n} \left\{ (G_k)_{ii} - \sum_{j \ne i} |(G_k)_{ij}| \right\} \right| \ge \left| \min_i \lambda_i \right| \tag{2.16}$$

$$b_2 = \max_i \{e_{ii}\}, \text{ 其中} e_{ii}\text{是}E\text{的第}i\text{个对角元,} \tag{2.17}$$

E是G_kGill-Murray修正Cholesky分解:$G_k + E = LDL^T$

- 如何对$G_k + E$做Cholesky分解?
  - 先做G_k的Cholesky分解， 再做$\bar{G}_k = G_k + E$的分解 (不实用)

GOLDFELD修正牛顿法

当$\nabla^2 f(x_k)$不定时，如何选取v_k?

- 应该大于G_k的最负特征值的绝对值.
- 按照Gill-Murray的修改Cholesky分解$(\bar{G}_k = \bar{L}_k \bar{D}_k \bar{L}_k^T)$算法确定$v_k$.
- $v_k = \min\{b_1, b_2\}$其中

$$b_1 = \left| \min_{1\le i\le n} \left\{ (G_k)_{ii} - \sum_{j\neq i} |(G_k)_{ij}| \right\} \right| \ge \left| \min_i \lambda_i \right| \tag{2.16}$$

$$b_2 = \max_i \{e_{ii}\},\ \text{其中}e_{ii}\text{是}E\text{的第}i\text{个对角元,} \tag{2.17}$$

E是G_kGill-Murray修正Cholesky分解:$G_k + E = LDL^T$

- 如何对$G_k + E$做Cholesky分解?
  - 先做G_k的Cholesky分解，再做$\bar{G}_k = G_k + E$的分解 (不实用)
  - Gill-Murray修正Cholesky分解

修正牛顿法

Theorem

- 如果 f 在开集 $\mathcal{D}$上二次连续可微.

修正牛顿法

Theorem

- 如果 f 在开集 $\mathcal{D}$上二次连续可微.
- 如果起始点 x_0 使水平集$\mathcal{L} = \{x \in \mathcal{D} : f(x) \leq f(x_0)\}$ 是紧的

修正牛顿法

Theorem

- 如果 f 在开集 $\mathcal{D}$上二次连续可微.
- 如果起始点 x_0 使水平集$\mathcal{L} = \{x \in \mathcal{D} : f(x) \leq f(x_0)\}$ 是紧的
- 假设修正的Cholesky分解是有界的

$$\kappa(\bar{G}_k) = \|\bar{G}_k\| \|\bar{G}_k^{-1}\| \leq C, \text{ for some } C, \quad \forall k = 0, 1, \cdots$$

修正牛顿法

Theorem

- 如果 f 在开集 $\mathcal{D}$上二次连续可微.
- 如果起始点 x_0 使水平集$\mathcal{L} = \{x \in \mathcal{D} : f(x) \leq f(x_0)\}$ 是紧的
- 假设修正的Cholesky分解是有界的

$$\kappa(\bar{G}_k) = \|\bar{G}_k\|\|\bar{G}_k^{-1}\| \leq C, \text{ for some } C, \quad \forall k = 0, 1, \cdots$$

- 则有

$$\lim_{k\to\infty} \nabla f(x_k) = 0.$$

有限差分牛顿法

- 对于无约束优化问题，如果已有导数$\nabla f(x)$，那么可以通过前向差分或者中心差分来计算Hessian.

有限差分牛顿法

- 对于无约束优化问题，如果已有导数$\nabla f(x)$，那么可以通过前向差分或者中心差分来计算Hessian.
- 在迭代中的第k步，可以计算

$$
\begin{aligned}
(A)_{\cdot j} &= \frac{\nabla f(x_k + h_j e_j) - \nabla f(x_k)}{h_j}, j = 1, \cdots, n, \\
A_k &= \frac{A + A^T}{2}, \\
x_{k+1} &= x_k - A_k^{-1} \nabla f(x_k), \\
h_j &= \sqrt{\eta} \max\{|x_j|, \tilde{x}_j\} \mathsf{sign}(x_j).
\end{aligned}
$$

其中$\tilde{x}_j$通常是给定的估计值，η大于机器精度的很小的数.

有限差分牛顿法

- 对于无约束优化问题，如果已有导数$\nabla f(x)$，那么可以通过前向差分或者中心差分来计算Hessian.
- 在迭代中的第k步，可以计算

$$
\begin{aligned}
(A)_{\cdot j} &= \frac{\nabla f(x_k + h_j e_j) - \nabla f(x_k)}{h_j}, j = 1, \cdots, n, \\
A_k &= \frac{A + A^T}{2}, \\
x_{k+1} &= x_k - A_k^{-1} \nabla f(x_k), \\
h_j &= \sqrt{\eta} \max\{|x_j|, \tilde{x}_j\} \mathsf{sign}(x_j).
\end{aligned}
$$

其中$\tilde{x}_j$通常是给定的估计值，η大于机器精度的很小的数.

- 有时候需要计算 $\nabla^2 f(x_k) d$，也可以用差分法来代替，减少计算量

$$
\nabla^2 f(x_k) d \approx \frac{\nabla f(x_k + hd) - \nabla f(x_k)}{h}
$$

有限差分牛顿法

- 对于无约束优化问题，如果已有导数$\nabla f(x)$，那么可以通过前向差分或者中心差分来计算Hessian.
- 在迭代中的第k步，可以计算

$$\begin{aligned}(A)_{\cdot j} &= \frac{\nabla f(x_k + h_j e_j) - \nabla f(x_k)}{h_j}, j = 1, \cdots, n,\\ A_k &= \frac{A + A^T}{2},\\ x_{k+1} &= x_k - A_k^{-1} \nabla f(x_k),\\ h_j &= \sqrt{\eta} \max\{|x_j|, \tilde{x}_j\} \mathrm{sign}(x_j).\end{aligned}$$

  其中$\tilde{x}_j$通常是给定的估计值，η大于机器精度的很小的数.
- 有时候需要计算 $\nabla^2 f(x_k) d$，也可以用差分法来代替，减少计算量

$$\nabla^2 f(x_k) d \approx \frac{\nabla f(x_k + hd) - \nabla f(x_k)}{h}$$

- 在一定的条件下，这种有限差分牛顿法 仍然是二阶收敛的.

有限差分牛顿法

有限差分牛顿法

- 如果导数∇f未知，需要用差分来逼近梯度∇f以及Hessian $\nabla^2 f$.

有限差分牛顿法

- 如果导数∇f未知，需要用差分来逼近梯度∇f以及Hessian $\nabla^2 f$.
- 首先逼近梯度

$$(\hat{g}_k)_j = \frac{f(x_k + h_j e_j) - f(x_k)}{h_j}, \text{ or } (\hat{g}_k)_j = \frac{f(x_k + h_j e_j) - f(x_k - h_j e_j)}{2h_j}.$$

有限差分牛顿法

- 如果导数∇f未知，需要用差分来逼近梯度∇f以及Hessian $\nabla^2 f$.
- 首先逼近梯度

$$(\hat{g}_k)_j = \frac{f(x_k + h_j e_j) - f(x_k)}{h_j}, \text{ or } (\hat{g}_k)_j = \frac{f(x_k + h_j e_j) - f(x_k - h_j e_j)}{2h_j}.$$

- 逼近Hessian

$$(A_k)_{ij} = \frac{[f(x_k + h_i e_i + h_j e_j) - f(x_k + h_i e_i)] - [f(x_k + h_j e_j) - f(x_k)]}{h_i h_j}$$

$$h_j = \sqrt[3]{\eta} \max\{|x_j|, \tilde{x}_j\} \mathsf{sign}(x_j)$$

有限差分牛顿法

- 如果导数∇f未知，需要用差分来逼近梯度∇f以及Hessian $\nabla^2 f$.
- 首先逼近梯度

$$(\hat{g}_k)_j = \frac{f(x_k + h_j e_j) - f(x_k)}{h_j}, \text{ or } (\hat{g}_k)_j = \frac{f(x_k + h_j e_j) - f(x_k - h_j e_j)}{2h_j}.$$

- 逼近Hessian

$$(A_k)_{ij} = \frac{[f(x_k + h_i e_i + h_j e_j) - f(x_k + h_i e_i)] - [f(x_k + h_j e_j) - f(x_k)]}{h_i h_j}$$

$$h_j = \sqrt[3]{\eta}\max\{|x_j|, \tilde{x}_j\}\mathsf{sign}(x_j)$$

- 此时第k步牛顿迭代法为

$$x_{k+1} = x_k - A_k^{-1}\hat{g}_k$$

有限差分牛顿法

- 如果导数∇f未知，需要用差分来逼近梯度∇f以及Hessian $\nabla^2 f$.
- 首先逼近梯度

$$(\hat{g}_k)_j = \frac{f(x_k + h_j e_j) - f(x_k)}{h_j}, \text{ or } (\hat{g}_k)_j = \frac{f(x_k + h_j e_j) - f(x_k - h_j e_j)}{2h_j}.$$

- 逼近Hessian

$$(A_k)_{ij} = \frac{[f(x_k + h_i e_i + h_j e_j) - f(x_k + h_i e_i)] - [f(x_k + h_j e_j) - f(x_k)]}{h_i h_j}$$

$$h_j = \sqrt[3]{\eta} \max\{|x_j|, \tilde{x}_j\} \text{sign}(x_j)$$

- 此时第k步牛顿迭代法为

$$x_{k+1} = x_k - A_k^{-1} \hat{g}_k$$

- 可以推导出如下收敛性

$$\|x_{k+1} - x^*\| \le \|A_k^{-1}\| \Big(\frac{v}{2} \|x_k - x^*\|^2 + \|A_k - \nabla^2 f(x_k)\| \|x_k - x^*\| + \|\hat{g}_k - \nabla f(x_k)\| \Big).$$

负曲率方向

负曲率方向属于修正牛顿法，是处理在$\nabla^2 f(x)$不定时，如何寻找搜索方向

定义

设$f: R^n \to R$在开集D上二次连续可微

负曲率方向

负曲率方向属于修正牛顿法，是处理在$\nabla^2 f(x)$不定时，如何寻找搜索方向

定义

设$f : R^n \to R$在开集D上二次连续可微

1. 如果$\nabla^2 f(x)$至少有一个负特征值，则称x为不定点

负曲率方向

负曲率方向属于修正牛顿法，是处理在$\nabla^2 f(x)$不定时，如何寻找搜索方向

定义

设$f: R^n \to R$在开集D上二次连续可微

1. 如果$\nabla^2 f(x)$至少有一个负特征值，则称x为不定点
2. 如果x是一个不定点，若方向p满足 $p^T\nabla^2 f(x)p < 0$,则 称p为$f(x)$在x处的负曲率方向.

负曲率方向

负曲率方向属于修正牛顿法，是处理在$\nabla^2 f(x)$不定时，如何寻找搜索方向

定义

设$f: R^n \to R$在开集D上二次连续可微

1. 如果$\nabla^2 f(x)$至少有一个负特征值，则称x为不定点
2. 如果x是一个不定点，若方向p满足 $p^T\nabla^2 f(x)p < 0$,则 称p为$f(x)$在x处的负曲率方向.
3. 如果x为不定点，

$$s^T\nabla f(x) \le 0, p^T\nabla f(x) \le 0, p^T\nabla^2 f(x)p < 0$$

则称(s, p)为不定点x处的下降对.

负曲率方向

负曲率方向属于修正牛顿法，是处理在$\nabla^2 f(x)$不定时，如何寻找搜索方向

定义

设$f: R^n \to R$在开集D上二次连续可微

1. 如果$\nabla^2 f(x)$至少有一个负特征值，则称x为不定点
2. 如果x是一个不定点，若方向p满足 $p^T \nabla^2 f(x) p < 0$,则 称p为$f(x)$在x处的负曲率方向.
3. 如果x为不定点，

   $$s^T \nabla f(x) \leq 0, p^T \nabla f(x) \leq 0, p^T \nabla^2 f(x) p < 0$$

   则称(s,p)为不定点x处的下降对.
4. 如果x不是一个不定点，(s,p)满足

   $$s^T \nabla f(x) < 0, p^T \nabla f(x) \leq 0, p^T \nabla^2 f(x) p = 0$$

   则称(s,p)为x处的下降对.

负曲率方向

Remark

- 一个下降对例子

$$s = -\nabla f(x), p = \begin{cases} 0, & \text{如果} \nabla^2 f(x) \geq 0, \\ -\mathsf{sign}(u^T \nabla f(x))u, & otherwise, \end{cases}$$

其中u是$\nabla^2 f(x)$的负特征值对应的特征向量.

负曲率方向

Remark

- 一个下降对例子

$$s = -\nabla f(x), p = \begin{cases} 0, & 如果\nabla^2 f(x) \geq 0, \\ -\mathsf{sign}(u^T \nabla f(x))u, & otherwise, \end{cases}$$

其中u是$\nabla^2 f(x)$的负特征值对应的特征向量.

- 显然，当且仅当$\nabla f(x) = 0$且$\nabla^2 f(x)$半正定时，下降对不存在.

负曲率方向

Remark

- 一个下降对例子

$$s = -\nabla f(x), p = \begin{cases} 0, & \text{如果}\nabla^2 f(x) \geq 0, \\ -\mathsf{sign}(u^T \nabla f(x))u, & otherwise, \end{cases}$$

其中u是$\nabla^2 f(x)$的负特征值对应的特征向量.

- 显然，当且仅当$\nabla f(x) = 0$且$\nabla^2 f(x)$半正定时，下降对不存在.
- 如果$\nabla f(x) = 0$, 负曲率方向是下降方向.

负曲率方向

Remark

- 一个下降对例子

$$s = -\nabla f(x), p = \begin{cases} 0, & \text{如果}\nabla^2 f(x) \geq 0, \\ -\mathsf{sign}(u^T \nabla f(x))u, & otherwise, \end{cases}$$

其中u是$\nabla^2 f(x)$的负特征值对应的特征向量.

- 显然，当且仅当$\nabla f(x) = 0$且$\nabla^2 f(x)$半正定时，下降对不存在.
- 如果$\nabla f(x) = 0$, 负曲率方向是下降方向.
- 在一般点处($\nabla f(x) \neq 0$),
  - 如果负曲率方向p满足$p^T \nabla f = 0$， 则p和$-p$都是下降方向.
  - 如果负曲率方向p满足$p^T \nabla f \leq 0$， 则p是下降方向.
  - 如果负曲率方向p满足$p^T \nabla f \geq 0$， 则$-p$是下降方向.

Fiacco-MaCormick法

基本思想：当G_k阵有负特征值时，沿着负曲率方向，可以使目标函数值下降.

$$f(x_k + p_k) \approx f(x_k) + g_k^T p_k + \frac{1}{2} p_k^T G_k p_k,$$

$$\text{如果 } g_k^T p_k \leq 0, \;\; p_k^T G_k p_k < 0.$$

Fiacco-MaCormick法

基本思想： 当G_k阵有负特征值时，沿着负曲率方向，可以使目标函数值下降.

$$f(x_k + p_k) \approx f(x_k) + g_k^T p_k + \frac{1}{2} p_k^T G_k p_k,$$

$$\text{如果 } g_k^T p_k \leq 0, \ \ p_k^T G_k p_k < 0.$$

- 首先对G_k做Cholesky分解$G_k = LDL^T$，其中L是单位下三角矩阵，D为对角矩阵.

Fiacco-MaCormick法

基本思想： 当G_k阵有负特征值时，沿着负曲率方向，可以使目标函数值下降.

$$f(x_k+p_k)\approx f(x_k)+g_k^Tp_k+\frac{1}{2}p_k^TG_kp_k,$$

$$\text{如果 } g_k^Tp_k\le 0,\ \ p_k^TG_kp_k<0.$$

- 首先对G_k做Cholesky分解$G_k=LDL^T$，其中L是单位下三角矩阵，D为对角矩阵.
- 如果$d_{ii}<0$，求解$L^Tt=a$，其中

$$a=\begin{cases}1, & d_{ii}\le 0,\\ 0, & d_{ii}>0.\end{cases}$$

Fiacco-MaCormick法

基本思想： 当G_k阵有负特征值时，沿着负曲率方向，可以使目标函数值下降.

$$f(x_k + p_k) \approx f(x_k) + g_k^T p_k + \frac{1}{2} p_k^T G_k p_k,$$

$$\text{如果 } g_k^T p_k \leq 0, \quad p_k^T G_k p_k < 0.$$

- 首先对G_k做Cholesky分解$G_k = LDL^T$，其中L是单位下三角矩阵，D为对角矩阵.
- 如果$d_{ii} < 0$，求解$L^T t = a$，其中

$$a = \begin{cases} 1, & d_{ii} \leq 0, \\ 0, & d_{ii} > 0. \end{cases}$$

- 则如下的p_k是负曲率下降方向

$$p_k = \begin{cases} t, & g_k^T t \leq 0, \\ -t, & g_k^T t > 0. \end{cases}$$

FIACCO-MACORMICK法

基本思想： 当G_k阵有负特征值时，沿着负曲率方向，可以使目标函数值下降.

$$f(x_k+p_k)\approx f(x_k)+g_k^Tp_k+\frac{1}{2}p_k^TG_kp_k,$$

$$\text{如果 } g_k^Tp_k\le 0,\ \ p_k^TG_kp_k<0.$$

- 首先对G_k做Cholesky分解$G_k=LDL^T$，其中L是单位下三角矩阵，D为对角矩阵.
- 如果$d_{ii}<0$，求解$L^Tt=a$，其中

$$a=\begin{cases}1, & d_{ii}\le 0,\\ 0, & d_{ii}>0.\end{cases}$$

- 则如下的p_k是负曲率下降方向

$$p_k=\begin{cases}t, & g_k^Tt\le 0,\\ -t, & g_k^Tt>0.\end{cases}$$

- 缺点： $d_{ii}\approx 0$ 或者 $d_{ii}=0$.

Fletcher-Freeman方法

如何稳定地分解对称不定矩阵?

FLETCHER-FREEMAN方法

如何稳定地分解对称不定矩阵?

- 对于任何对称矩阵，存在排列矩阵P使得，$P^T G_k P = LDL^T$，其中L是单位下三角矩阵，D是块对角矩阵，阶数是1或者2.

Fletcher-Freeman方法

如何稳定地分解对称不定矩阵?

- 对于任何对称矩阵，存在排列矩阵P使得，$P^TG_kP = LDL^T$，其中L是单位下三角矩阵，D是块对角矩阵，阶数是1或者2.

$$D = \begin{bmatrix} * & & & & & \\ & * & * & & & \\ & * & * & & & \\ & & & * & * & \\ & & & * & * & \\ & & & & & * \end{bmatrix}, L = \begin{bmatrix} 1 & & & & & \\ * & 1 & & & & \\ * & 0 & 1 & & & \\ * & * & * & 1 & & \\ * & * & * & 0 & 1 & \\ * & * & * & * & * & 1 \end{bmatrix}$$

Fletcher-Freeman方法

如何稳定地分解对称不定矩阵?

- 对于任何对称矩阵，存在排列矩阵P使得，$P^TG_kP = LDL^T$，其中L是单位下三角矩阵，D是块对角矩阵，阶数是1或者2.

$$D = \begin{bmatrix} * & & & & & \\ & * & * & & & \\ & * & * & & & \\ & & & * & * & \\ & & & * & * & \\ & & & & & * \end{bmatrix}, L = \begin{bmatrix} 1 & & & & & \\ * & 1 & & & & \\ * & 0 & 1 & & & \\ * & * & * & 1 & & \\ * & * & * & 0 & 1 & \\ * & * & * & * & * & 1 \end{bmatrix}$$

- 求解系统$L^Tt = a$,

$$a_i = \begin{cases} 1, & d_{ii} \le 0, \\ 0, & d_{ii} > 0; \end{cases} \text{或者} \begin{pmatrix} a_i \\ a_{i+1} \end{pmatrix} \text{是} \begin{bmatrix} d_{ii}, & d_{i,i+1} \\ d_{i+1,i} & d_{i+1,i+1} \end{bmatrix} \text{的单位特征向量}$$

FLETCHER-FREEMAN方法

如何稳定地分解对称不定矩阵?

- 对于任何对称矩阵，存在排列矩阵P使得，$P^T G_k P = LDL^T$，其中L是单位下三角矩阵，D是块对角矩阵，阶数是1或者2.

$$D = \begin{bmatrix} * & & & & & \\ & * & * & & & \\ & * & * & & & \\ & & & * & * & \\ & & & * & * & \\ & & & & & * \end{bmatrix}, L = \begin{bmatrix} 1 & & & & & \\ * & 1 & & & & \\ * & 0 & 1 & & & \\ * & * & * & 1 & & \\ * & * & * & 0 & 1 & \\ * & * & * & * & * & 1 \end{bmatrix}$$

- 求解系统$L^T t = a$,

$$a_i = \begin{cases} 1, & d_{ii} \le 0, \\ 0, & d_{ii} > 0; \end{cases} \text{或者} \begin{pmatrix} a_i \\ a_{i+1} \end{pmatrix} \text{是} \begin{bmatrix} d_{ii}, & d_{i,i+1} \\ d_{i+1,i} & d_{i+1,i+1} \end{bmatrix} \text{的单位特征向量}$$

- 令$p_k = \begin{cases} t, & \text{如果} g_k^T t \le 0; \\ -t, & \text{如果} g_k^T t > 0. \end{cases}$ 是下降的负曲率方向. 可以验证

$$p_k^T LDL^T p_k = a^T Da = \sum \lambda_i < 0, \text{且} p_k^T g_k \le 0.$$

Fletcher-Freeman 方法

- 当D有负特征值时，搜索方向还可以由其广义逆矩阵构造

$$p_k = -L^{-T}\tilde{D}^{\dagger}L^{-1}g_k$$

其中$\tilde{D}^{\dagger}$是$\tilde{D}$的广义逆，$\tilde{D}_i = \begin{cases} d_{ii}, & \text{当} d_{ii} > 0, \\ 0, & \text{其他}. \end{cases}$

Fletcher-Freeman 方法

- 当D有负特征值时，搜索方向还可以由其广义逆矩阵构造

$$p_k = -L^{-T}\tilde{D}^{\dagger}L^{-1}g_k$$

$$\text{其中}\tilde{D}^{\dagger}\text{是}\tilde{D}\text{的广义逆，}\quad \tilde{D}_i = \begin{cases} d_{ii}, & \text{当} d_{ii} > 0, \\ 0, & \text{其他}. \end{cases}$$

- 当D有零特征值时，搜索方向p_k可以由

$$LDL^T p_k = 0, \quad g_k^T p_k < 0.$$

Fletcher-Freeman 方法

- 当D有负特征值时，搜索方向还可以由其广义逆矩阵构造

$$p_k = -L^{-T} \tilde{D}^{\dagger} L^{-1} g_k$$

$$\text{其中}\tilde{D}^{\dagger}\text{是}\tilde{D}\text{的广义逆,} \quad \tilde{D}_i = \begin{cases} d_{ii}, & \text{当} d_{ii} > 0, \\ 0, & \text{其他}. \end{cases}$$

- 当D有零特征值时，搜索方向p_k可以由

$$LDL^T p_k = 0, \quad g_k^T p_k < 0.$$

- 当D的特征值为正数时，D中全是一阶块. 这时就是正常的牛顿法.

Gill-Murray修正Cholesky分解

- 传统Cholesky分解$G = LDL^T$，有缺点$d_{jj} \approx 0$

$$d_{jj} = g_{jj} - \sum_{s=1}^{j-1} d_{ss} l_{js}^2, \qquad l_{ij} = \frac{1}{d_{jj}}\left(g_{ij} - \sum_{s=1}^{j-1} d_{ss} l_{js} l_{is}\right), i \geq j+1.$$

Gill-Murray修正Cholesky分解

- 传统Cholesky分解$G = LDL^T$，有缺点$d_{jj} \approx 0$

$$d_{jj} = g_{jj} - \sum_{s=1}^{j-1} d_{ss} l_{js}^2, \qquad l_{ij} = \frac{1}{d_{jj}} \Big(g_{ij} - \sum_{s=1}^{j-1} d_{ss} l_{js} l_{is} \Big), i \geq j+1.$$

- 额外要求： 存在某个正数$\delta > 0$ and $\beta > 0$，使得

$$d_{jj} \geq \delta, \quad |r_{ij} = l_{ij} \sqrt{d_{jj}}| \leq \beta, \quad i \geq j.$$

Gill-Murray修正Cholesky分解

- 传统Cholesky分解$G = LDL^T$，有缺点$d_{jj} \approx 0$

$$d_{jj} = g_{jj} - \sum_{s=1}^{j-1} d_{ss} l_{js}^2, \qquad l_{ij} = \frac{1}{d_{jj}}\left(g_{ij} - \sum_{s=1}^{j-1} d_{ss} l_{js} l_{is}\right), i \geq j+1.$$

- 额外要求： 存在某个正数$\delta > 0$ and $\beta > 0$，使得

$$d_{jj} \geq \delta, \quad |r_{ij} = l_{ij}\sqrt{d_{jj}}| \leq \beta, \quad i \geq j.$$

- 假设前$j-1$步都满足以上额外要求，我们看第j步(分成两步)

第一步:计算$\gamma_j = \left|\xi_j - \sum_{s=1}^{j-1} d_{ss} l_{js}^2\right|$，其中$\xi_j = g_{jj}$，然后取$\bar{d} = \max\{\gamma_j, \delta\}$.

第二步:计算$\left|r_{ij} = l_{ij}\sqrt{\bar{d}}\right|$，如果$|r_{ij}| < \beta$，接受$\bar{d} = d_{jj}, l_{ij} = r_{ij}/\sqrt{d_{jj}}$

否则$d_{jj} = \left|d_{jj} + e_{jj} - \sum_{s=1}^{j-1} d_{ss} l_{js}^2\right|$，选择正数$e_{jj}$ 使得 $\max|r_{ij}| = \beta$.

Gill-Murray修正Cholesky分解

- 传统Cholesky分解$G = LDL^T$，有缺点$d_{jj} \approx 0$

$$d_{jj} = g_{jj} - \sum_{s=1}^{j-1} d_{ss} l_{js}^2, \qquad l_{ij} = \frac{1}{d_{jj}}\left(g_{ij} - \sum_{s=1}^{j-1} d_{ss} l_{js} l_{is}\right), i \geq j+1.$$

- 额外要求：存在某个正数$\delta > 0$ and $\beta > 0$，使得

$$d_{jj} \geq \delta, \quad |r_{ij} = l_{ij}\sqrt{d_{jj}}| \leq \beta, \quad i \geq j.$$

- 假设前$j-1$步都满足以上额外要求，我们看第j步(分成两步)

第一步:计算$\gamma_j = \left|\xi_j - \sum_{s=1}^{j-1} d_{ss} l_{js}^2\right|$，其中$\xi_j = g_{jj}$，然后取$\bar{d} = \max\{\gamma_j, \delta\}$.

第二步:计算$\left|r_{ij} = l_{ij}\sqrt{\bar{d}}\right|$，如果$|r_{ij}| < \beta$，接受$\bar{d} = d_{jj}$，$l_{ij} = r_{ij}/\sqrt{d_{jj}}$

否则$d_{jj} = \left|d_{jj} + e_{jj} - \sum_{s=1}^{j-1} d_{ss} l_{js}^2\right|$，选择正数$e_{jj}$ 使得 $\max|r_{ij}| = \beta$.

- 本质上完成了$\bar{G}_k = G_k + E = LDL^T$的Cholesky分解.

Gill-Murray修正Cholesky分解

Remark

- E的选择与β有关. Gill-Murray证明

$$\|E(\beta)\|_\infty \le \left(\frac{\xi}{\beta} + (n-1)\beta\right)^2 + 2(\gamma + (n-1)\beta^2) + \delta$$

其中ξ是G_k的非对角元最大模，γ是G_k的对角元最大模.

Gill-Murray修正Cholesky分解

Remark

- E的选择与β有关. Gill-Murray证明

$$\|E(\beta)\|_\infty \le \left(\frac{\xi}{\beta} + (n-1)\beta\right)^2 + 2(\gamma + (n-1)\beta^2) + \delta$$

  其中ξ是G_k的非对角元最大模，γ是G_k的对角元最大模.
- 当$\beta^2 = \xi/\sqrt{n^2-1}$时，上面的界是最小的. 因此取$\beta$满足

$$\beta^2 = \max\{\gamma, \xi/\sqrt{n^2-1}, \varepsilon_M(\text{机器精度})\}$$

GILL-MURRAY修正CHOLESKY分解

REMARK

- E的选择与β有关. Gill-Murray证明
$$\|E(\beta)\|_\infty \le \left(\frac{\xi}{\beta} + (n-1)\beta\right)^2 + 2(\gamma + (n-1)\beta^2) + \delta$$
其中ξ是G_k的非对角元最大模，γ是G_k的对角元最大模.
- 当$\beta^2 = \xi/\sqrt{n^2-1}$时，上面的界是最小的. 因此取$\beta$满足
$$\beta^2 = \max\{\gamma, \xi/\sqrt{n^2-1}, \varepsilon_M(\text{机器精度})\}$$
- Gill-Murray修正Cholesky分解，保证了$d_{jj} \ge \delta$, $\bar{G}_k$是正定的，其条件数$\|\bar{G}_k\|\|\bar{G}_k^{-1}\| \le \kappa$有一致的上界.

Gill-Murray修正Cholesky分解

Gill-Murray修正Cholesky分解算法

第一步：确定分解因子元素的界$\beta^2 = \max\{\gamma, \xi/\sqrt{n^2-1}, \varepsilon_M\}$,
这里γ和ξ分别是G_k的对角元和非对角元最大值.

第二步：初始化. 令$j=1$, $c_{ii} = g_{ii}$ for $i = 1, \cdots, n$.

第三步：求最小指标q, 使得$|c_{qq}| = \max_{j\le i\le n} |c_{ii}|$,
交换G_k的q行和j行, q列和j列的信息.

第四步：计算L的第j行, 并求出$l_{ij}d_{jj}$的最大值:
令$l_{js} = c_{js}/d_{ss}$, $s = 1, \cdots, j-1$;
计算$c_{ij} = g_{ij} - \sum_{s=1}^{j-1} l_{sj}c_{is}$, $i = j+1, \cdots, n$;
令$\theta_j = \max_{j+1\le i\le n} |c_{ij}|$, $(\theta_j = 0$若$j = n)$

第五步: 计算D的第j个对角元: $d_{jj} = \max\{\delta, |c_{jj}|, \theta_j^2/\beta^2\}$;
E的对角元修改为$E_j = d_{jj} - c_{jj}$, 如果$j = n$, 则停止.

第六步：校正对角元和列指标：
令$c_{ii} = c_{ii} - c_{ij}^2/d_{jj}$, $i = j+1, \cdots, n$;
令$j = j+1$,转第三步.

Gill-Murray稳定牛顿法

基本思想：

(i) 当G_k不定时，采用Gill-Murray修正Cholesky分解强制Hessian正定,

Gill-Murray稳定牛顿法

基本思想：

(i) 当G_k不定时，采用Gill-Murray修正Cholesky分解强制Hessian正定，

$$设\ \bar{G}_k = G_k + E_k = L_k D_k L_k^T,\ D_k = diag(d_{ii}),\ E_k = diag(e_{ii}).$$

(ii) 当$\nabla f \to 0$时，采用负曲率方向使目标函数值下降.

构造负曲率方向算法

第一步：令$\psi_j = d_{jj} - e_{jj},\ j = 1, \cdots, n$

第二步：求下标t，使得$\psi_t = \min\{\psi_j | j = 1, \cdots, n\}$.

第三步：如果$\psi_t \geq 0$，则停止；否则，求解方程组 $L_k^T p = e_t$，得到负曲率方向p，其中e_t为第t个元素为1的单位向量.

GILL-MURRAY稳定牛顿法

基本思想：

(i) 当G_k不定时，采用Gill-Murray修正Cholesky分解强制Hessian正定,

$$设\ \bar{G}_k = G_k + E_k = L_k D_k L_k^T,\ D_k = diag(d_{ii}),\ E_k = diag(e_{ii}).$$

(ii) 当$\nabla f \to 0$时, 采用负曲率方向使目标函数值下降.

构造负曲率方向算法

第一步：令$\psi_j = d_{jj} - e_{jj},\ j = 1, \cdots, n$

第二步：求下标t, 使得$\psi_t = \min\{\psi_j | j = 1, \cdots, n\}$.

第三步：如果$\psi_t \geq 0$, 则停止；否则，求解方程组 $L_k^T p = e_t$, 得到负曲率方向p, 其中e_t为第t个元素为1的单位向量.

定理

设G_k为f在x_k处的Hessian矩阵, $\bar{G}_k = G_k + E_k = L_k D_k L_k^T$. 则上述算法产生的$p_k$是在$x_k$处的负曲率方向, 并且$p_k$和$-p_k$中至少有一个是$x_k$处的下降方向.

Gill-Murray稳定牛顿法

Gill-Murray稳定牛顿法

第一步：给定初始点x_0, $\varepsilon > 0$, $k = 1$
第二步：计算g_k和G_k.
第三步：使用Gill-Murray修正Cholesky分解,得到$L_k D_k L_k^T = G_k + E_k$
第四步：如果$\|g_k\| \geq \varepsilon$,
求解方程$L_k D_k L_k^T p_k = -g_k$, 得到搜索方向p_k，转第六步；
否则(即$\|g_k\| \leq \varepsilon$)转第五步
第五步: 调用构造负曲率方向算法，如果无负曲率方向,则停止;
否则, 求出负曲率方向p_k满足,

$$p_k = \begin{cases} -p_k, & g_k^T p_k > 0, \\ p_k, & otherwise. \end{cases}$$

第六步：一维搜索求α_k满足

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

令$x_{k+1} = x_k + \alpha_k p_k$;
第七步：若$f(x_{k+1}) \geq f(x_k)$, 则停止计算， 否则$k = k + 1$, 转第二步.

Gill-Murray稳定牛顿法收敛性

定理：Gill-Murray稳定牛顿法收敛性

- 设f二次连续可微
- 存在$\bar{x} \in R^n$, 使得$L(\bar{x}) = \{x|f(x) \le f(\bar{x})\}$ 为有界闭凸集
- 以上算法中$\varepsilon = 0$
- 则序列$\{x_k\}$满足
  1. 当$\{x_k\}$为有限序列， 最后一个点是稳定点
  2. 当$\{x_k\}$为无限序列，则必有聚点, 且所有聚点都是稳定点.

不精确的牛顿法

考虑非线性方程组 $F(x) = 0$, 其中 $F: R^n \to R^n$, 满足

- (1)存在解 x^*, $F(x^*) = 0$,
- (2)F 在 x^* 的领域中连续可微,
- (3)$F'(x^*)$ 非奇异.

不精确的牛顿法

考虑非线性方程组 $F(x)=0$，其中 $F:R^n\to R^n$，满足

- (1)存在解x^*, $F(x^*)=0$,
- (2)F在x^*的领域中连续可微,
- (3)$F'(x^*)$非奇异.

- 经典牛顿法: $F(x_k+s)\approx F(x_k)+F'(x_k)s=0$, $s=-(F'(x_k))^{-1}F(x_k)$

不精确的牛顿法

考虑非线性方程组 $F(x) = 0$，其中$F : R^n \to R^n$，满足

- (1)存在解x^*，$F(x^*) = 0$，
- (2)F在x^*的领域中连续可微，
- (3)$F'(x^*)$非奇异.

- 经典牛顿法：$F(x_k + s) \approx F(x_k) + F'(x_k)s = 0$，$s = -(F'(x_k))^{-1}F(x_k)$
- 不精确牛顿法：$F'(x_k)s = -F(x_k) + r_k$，其中$\frac{r_k}{\|F(x_k)\|} \le \eta_k$.
  $r_k = F'(x_k)s + F(x_k)$表示残量，η_k是一个控制不精确程度的序列.

引理：F'在x^*附近非奇异

设$F : R^n \to R^n$在$x^* \in D$连续可微，$F'(x^*)$非奇异，则存在$\delta > 0, \gamma > 0, \epsilon > 0$，使得当$\|y - x^*\| < \delta$，$y \in D$时，$F'$非奇异，且

$$\|F'(y)^{-1}\| \le \gamma, \quad \|F'(y)^{-1} - F'(x^*)^{-1}\| \le \epsilon.$$

不精确的牛顿法收敛性

定理：线性收敛

设$F: R^n \to R^n$满足假设条件(1)(2)(3). 设η_k满足$0 \leq \eta_k \leq \eta_{\max} < t < 1$. 那么，对于某个$\epsilon > 0$, 如果$\|x_0 - x^*\| < \epsilon$, 则有不精确牛顿法产生的迭代序列$\{x_k\}$线性收敛到$x^*$,

$$\|x_{k+1} - x^*\|_* \leq t\|x_k - x^*\|_*,$$

其中$\|y\|_* = \|F'(x^*)^{-1}y\|$.

不精确的牛顿法收敛性

定理：线性收敛

设$F: R^n \to R^n$满足假设条件(1)(2)(3). 设η_k满足$0 \le \eta_k \le \eta_{\max} < t < 1$. 那么，对于某个$\epsilon > 0$, 如果$\|x_0 - x^*\| < \epsilon$, 则有不精确牛顿法产生的迭代序列$\{x_k\}$线性收敛到$x^*$,

$$\|x_{k+1} - x^*\|_* \le t\|x_k - x^*\|_*,$$

其中$\|y\|_* = \|F'(x^*)^{-1}y\|$.

引理

设$\alpha = \max\{\|F'(x^*) + \frac{1}{2\beta}\|, 2\beta\}$, $\beta = \|F'(x^*)^{-1}\|$, 则当$\|y - x^*\|$充分小时，

$$\frac{1}{\alpha}\|y - x^*\| \le \|F(y)\| \le \alpha\|y - x^*\|.$$

不精确的牛顿法收敛性

定理：线性收敛

设$F : R^n \to R^n$满足假设条件(1)(2)(3). 设η_k满足$0 \le \eta_k \le \eta_{\max} < t < 1$. 那么，对于某个$\epsilon > 0$, 如果$\|x_0 - x^*\| < \epsilon$, 则有不精确牛顿法产生的迭代序列$\{x_k\}$线性收敛到$x^*$,

$$\|x_{k+1} - x^*\|_* \le t\|x_k - x^*\|_*,$$

其中$\|y\|_* = \|F'(x^*)^{-1}y\|$.

引理

设$\alpha = \max\{\|F'(x^*) + \frac{1}{2\beta}\|, 2\beta\}$, $\beta = \|F'(x^*)^{-1}\|$, 则当$\|y - x^*\|$充分小时,

$$\frac{1}{\alpha}\|y - x^*\| \le \|F(y)\| \le \alpha\|y - x^*\|.$$

定理: 超线性收敛

假设上述定理条件都成立，不精确牛顿法产生的序列$\{x_k\}$是收敛到x^*, 那么当且仅当$\|r_k\| = o(\|F(x_k)\|)$, $k \to \infty$时, $\{x_k\}$超线性收敛到x^*.

Thanks For Your Attention