最优化理论

Optimality Theory





目 录 (CONTENT)

PART TEN

10

使用导数的最优化方法

Optimization Method Using Derivative

■ 主要内容

- 最速下降法
- 牛顿法
- 共轭梯度法
- 拟牛顿法
- 信赖域法

■ 算法

算法描述

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

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

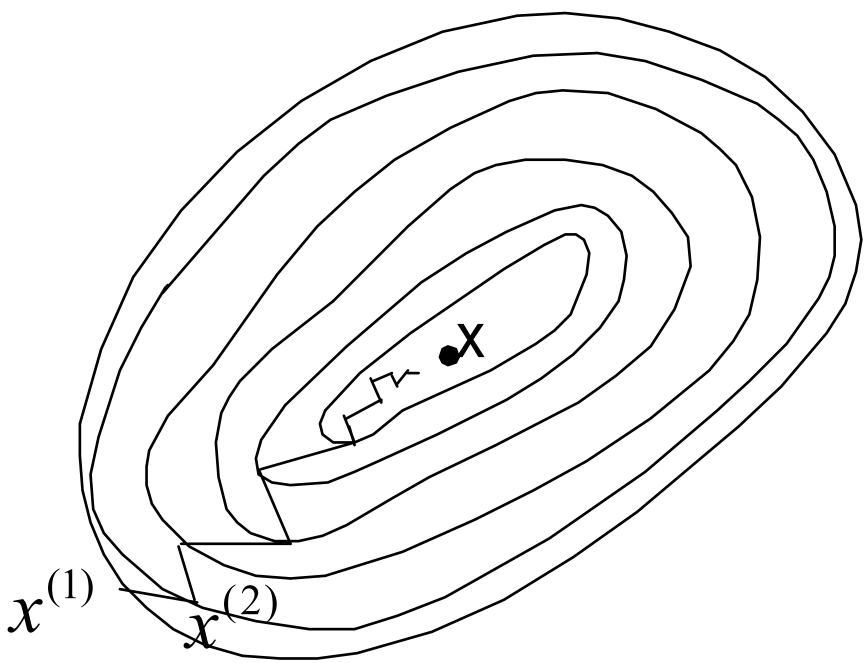
$$Step3, 若 \left\| d^{(k)} \right\| \le \varepsilon, 停止, 否则, 从 x^{(k)} 出发, 沿 d^{(k)} 进行一维搜索,$$

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

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

■ 性质1

最速下降法存在锯齿现象



■ 性质2

容易证明, 用最速下降法极小化目标函数时, **相邻两个搜索方向是正交的.**

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

则
$$\varphi'(\lambda)=\nabla f(x^{(k)}+\lambda_k d^{(k)})^{\mathrm{T}} d^{(k)}=0$$
$$\Rightarrow -\nabla f(x^{(k+1)})^{T}\nabla f(x^{(k)})=0$$
$$\Rightarrow d^{(k+1)}=-\nabla f(x^{(k+1)})\text{与}d^{(k)}=-\nabla f(x^{(k)})\text{正交}$$

■ 概念

设$\nabla^2 f(x^{(k)})$可逆, 则得牛顿法的迭代公式

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

算法(Newton法)

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

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

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

■ 局部收敛性

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

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

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

- 注意, 当初始点远离极小点时, 牛顿法可能不收敛

- **阻尼牛顿法**

基本思想:增加了沿牛顿方向的一维搜索.

迭代公式为

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

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

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

- 算法(阻尼牛顿法)

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

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

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

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

$Step 4$, 从 $x^{(k)}$ 出发, 沿方向 $d^{(k)}$ 作一维搜索:

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

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

$Step 5$, 置 $k := k + 1$, 转2

■ 存在问题

显然可能存在：

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

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

牛顿方向不一定是下降方向，算法失效！

■ 修正牛顿法

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

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

中用G_k取代矩阵$\nabla^2 f(x^{(k)})$:

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

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

再沿此方向作一维搜索

如何构造G_k?比如,可令

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

其中I是单位阵,ε_k是一个适当的正数.

■ 修正牛顿法

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

■ 共轭方向与扩张子空间定理

定义10.3.1 设A是$n\times n$对称矩阵,若R^n 中的两个方向d^1 和d^2满足

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

则称这两个方向关于A共轭,或称它们关于A正交.

若$d^{(1)},d^{(2)},\dots,d^{(k)}$是$R^n$中$k$个方向, 它们两两关于$A$共轭, 即

$$d^{(i)T}Ad^{(j)}=0, i\neq j,\quad i,j=1,2,..k. \qquad (10.3.2)$$

则称这组方向是A共轭,或称它们为A的k个共轭方向

■ 几何意义

设有二次函数

$$f(x)=\frac{1}{2}(x-\bar{x})^T A(x-\bar{x}) \qquad (10.3.3)$$

其中A是$n\times n$**对称正定矩阵**, $\bar{x}$是一个定点.

$f(x)$的等值面

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

是以$\bar{x}$为中心的椭球面,

由于

$$\nabla f(\bar{x})=A(\bar{x}-\bar{x})=0$$

A正定,故$\bar{x}$是$f(x)$的极小值点.

共轭梯度法

设$x^{(1)}$是在某等值面上一点,此面在点$x^{(1)}$处的法向量

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

又设$d^{(1)}$是在该等值面在点$x^{(1)}$处的一个切向量.

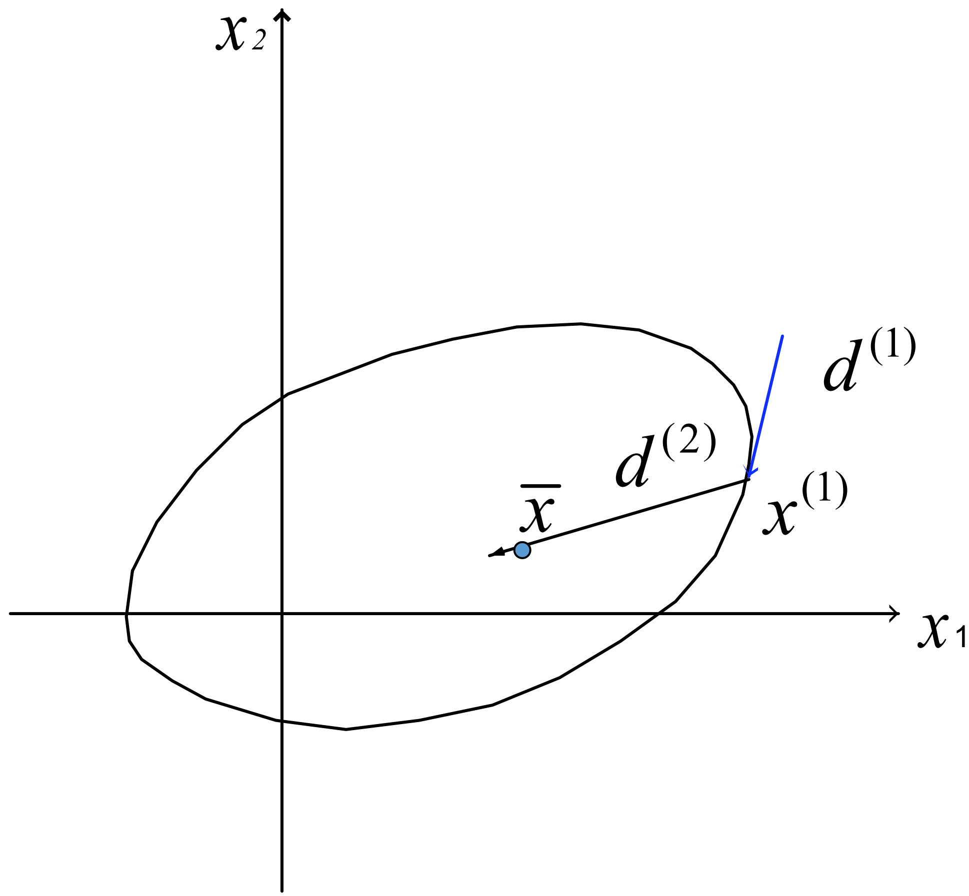
$$d^{(2)} = \bar{x} - x^{(1)}$$

显然,$d^{(1)}$与$\nabla f(x^{(1)})$正交.即$d^{(1)T}\nabla f(x^{(1)}) = 0$, 于是

$$d^{(1)T} A d^{(2)} = 0$$

即等值面上一点$x^{(1)}$处的切向量与由这点指向极小点的向量关于A共轭.

■ 几何意义



沿着$d^{(1)}$和$d^{(2)}$进行一维搜索，经两次迭代必达到极小点

算法1 共轭方向法

步1(初始化),给定初始点$x^0 \in R^n$, 计算$\nabla f(x^0)$, 给定一个搜索方向$d^0 \neq 0$, 使得$\nabla f(x^0)^T d^0 < 0$; 置$k=0$

步2(线搜索),求解一维极小化问题$\min\limits_{\alpha \in R} f(x^k + \alpha d^k)$;

$x^{k+1} = x^k + \alpha_k d^k$,若$\nabla f(x^{k+1}) = 0$或$k = n-1$, 停止, 否则转步3

步3(计算共轭方向),计算一个非零方向$d^{k+1} \in R^n$, 使得$(d^{k+1})^T G d^j = 0 (\forall j = 0,1,\ldots,k)$; 置$k+1 \to k$转步2

定理10.3.1 设A是n阶对称正定矩阵,$d^{(1)}, d^{(2)}, \ldots, d^{(k)}$是$k$个A共轭的非零向量,则这个向量组线性无关.

定理10.3.2(扩张子空间定理) 设有函数

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

其中A是n阶对称正定矩阵,$d^{(1)},d^{(2)},\dots,d^{(k)}$是A共轭的非零向量. 以任意的$x^{(1)}\in R^n$为初始点, 依次沿$d^{(1)},d^{(2)},\dots,d^{(k)}$进行一维搜索,得到点列$x^{(1)},x^{(2)},\dots,x^{(k+1)}$,则$x^{(k+1)}$是函数$f(x)$在线性流形$x^{(1)}+\Gamma_\kappa$上的唯一极小点.特别地,当$k=n$时,$x^{(k+1)}$是函数$f(x)$在$E^n$上的唯一极小点.其中

$$\Gamma_\kappa=\left\{x\,\middle|\,x=\sum_{i=1}^{k}\lambda_i d^{(i)},\lambda_i\in(-\infty,+\infty)\right\}$$

是$d^{(1)},d^{(2)},\dots,d^{(k)}$生成的子空间.

证明:由于$f(x)$严格凸,要证明$x^{(k+1)}$是函数$f(x)$在线性流形 $x^{(1)}+\Gamma_{\kappa}$ 上的唯一极小点,只要证在$x^{(k+1)}$点, $\nabla f(x^{(k+1)})$与子空间Γ_{κ}正交.

用归纳法证之,为方便,用g_j表示函数$f(x)$在$x^{(j)}$处的梯度,即

证明$g_{k+1}\perp\Gamma_k$,对k归纳

当$k=1$,由一维搜索的定义知$g_2\perp\Gamma_1$.

假设$k=$m<n时$g_{m+1}\perp\Gamma_m$,下证$g_{m+2}\perp\Gamma_{m+1}$.

由二次函数梯度的表达式和点$x^{(k+1)}$的定义, 有

$$g_{m+2}=Ax^{(m+2)}=A(x^{(m+1)}+\lambda_{m+1}d^{(m+1)})+b$$

$$=g_{m+1}+\lambda_{m+1}Ad^{(m+1)} \qquad (10.3.7)$$

利用上式可以将 g_{m+2} 和$d^{(i)}$的内积写成

$$d^{(i)}g_{m+2}=d^{(i)T}g_{m+1}+\lambda_{m+1}d^{(i)T}Ad^{(m+1)} \qquad (10.3.8)$$

当$i=m+1$时, 由一维搜索定义, 知

$$d^{(m+1)T}g_{m+2}=0 \qquad (10.3.9)$$

当$1\leq i<m+1$时, 由归纳假设知

$$d^{(i)T}g_{m+1}=0 \qquad (10.3.10)$$

由于$d^{(1)},d^{(2)},\ldots,d^{(m+1)}$关于A共轭, 则

$$d^{(i)T}Ad^{(m+1)}=0 \qquad (10.3.11)$$

由10.3.8-11, 知

$$d^{(i)T} g_{m+2} = 0$$

即

$$g_{m+2} \perp \Gamma_{m+1}.$$

根据上述证明, $x^{(k+1)}$是$f(x)$在$x^{(1)} + \Gamma_k$上的极小点.由于$f(x)$严格凸, 故必为此流形上的唯一极小点

当$k = n, d^{(1)}, d^{(2)}, \ldots, d^{(n)}$是$E^n$的一组基, 此时必有$g_{n+1} = 0$,
从而$x^{(n+1)}$是函数在E^n上的唯一极小点.

推论: 在$Th10.3.2$的条件下, $g_{k+1}^T d^{(j)} = 0, \quad \forall j \le k$

■ 线性共轭梯度法与二次终止性

上述定理表明,对于二次凸函数,若沿一组共轭方向(非零向量)搜索,经有限步迭代必到达极小点.

- 基本思想:把共轭性与最速下降法相结合, 利用已知点处的梯度构造一组共轭方向, 并沿着这组方向进行搜索, 求出目标函数的极小点

先讨论对于二次凸函数的共轭梯度法, 考虑问题

$$\min f(x)=\frac{1}{2}x^T Ax+b^T x+c \qquad (10.3.12)$$

$x\in E^n$, A对称正定, c是常数.

求解方法

首先, 任意给定一初始点$x^{(1)}$, 计算出目标函数在该点的梯度, 若$\|g_1\|=0$, 则停止计算, 否则, 令

$$d^{(1)}=-\nabla f(x^{(1)})=-g_1 \qquad (10.3.13)$$

沿$d^{(1)}$搜索, 得点$x^{(2)}$. 计算在$x^{(2)}$处的梯度, 若$\|g_2\|\neq 0$

则利用$-g_2$和$d^{(1)}$构造搜索方向$d^{(2)}$, 再沿$d^{(2)}$搜索.

一般地，若已知点$x^{(k)}$和搜索方向 $d^{(k)}$，则从$x^{(k)}$出发，沿$d^{(1)}$进行搜索，得

$$x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)} \qquad (10.3.14)$$

其中步长λ_k满足

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

下求λ_k的表达式

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

令 $\varphi'(\lambda) = \nabla f(x^{(k+1)})^{\mathrm{T}} d^{(k)} = 0 \qquad (10.3.15)$

即 $(Ax^{(k+1)} + b)^{\mathrm{T}} d^{(k)} = 0$

共轭梯度法

$$\Rightarrow (A(x^{(k)} + \lambda_k d^{(k)}) + b)^{\mathrm{T}} d^{(k)} = 0$$

$$\Rightarrow (g_k + \lambda_k A d^{(k)})^{\mathrm{T}} d^{(k)} = 0 \qquad (10.3.16)$$

$$\Rightarrow \lambda_k = -\frac{g_k{}^{\mathrm{T}} d^{(k)}}{d^{(k)\mathrm{T}} A d^{(k)}} \qquad (10.3.17)$$

计算$f(x)$在$x^{(k+1)}$处的梯度，若$\|g_{k+1}\|=0$，则停止计算，否则，利用$-g_{k+1}$和$d^{(k)}$构造下一搜索方向$d^{(k+1)}$,并使$d^{(k+1)}$和$d^{(k)}$关于A共轭，按此设想. 令

$$d^{(k+1)} = -g_{k+1} + \beta_k d^{(k)} \qquad (10.3.18)$$

上式两端左乘$d^{(k)T}A$,并令

$$d^{(k)T}Ad^{(k+1)}=-d^{(k)T}Ag_{k+1}+\beta_k d^{(k)T}Ad^{(k)}=0$$

$$\Rightarrow \beta_k=\frac{d^{(k)T}Ag_{k+1}}{d^{(k)T}Ad^{(k)}} \qquad (10.3.19)$$

再从$x^{(k+1)}$出发,沿方向$d^{(k+1)}$搜索.

综上分析,在第一个搜索方向取负梯度的前提下,重复使用公式3.14,3.17-3.19就能伴随计算点的增加,构造出一组搜索方向. (Fletcher-Reeves法)

定理10.3.3 对于正定二次函数(10.3.12),具有精确一维搜索的Fletcher-Reeves法在$m\le n$次一维搜索后即终止,并且对所有$i(1\le i\le m)$,下列关系成立:

$$1, d^{(i)T}Ad^{(j)}=0, \quad j=1,2,\ldots,i-1$$

$$2, g_i^T g_j=0, \quad j=1,2,\ldots,i-1$$

$$3, g_i^T d^{(i)}=-g_i^T g_i \quad (\text{蕴涵} d^{(i)}\ne 0)$$

证明: 显然m≥1,下用归纳法(对i)证之.

当$i=1$时,由于$d^{(1)}=-g_1$,从而3)成立,对$i=2$时,

关系1)和2)成立,从而3)也成立.

设对某个$i<m$,这些关系均成立,我们证明对于$i+1$也成立.先证2),

$$x^{(i+1)}=x^{(i)}+\lambda_i d^{(i)}$$

由迭代公式两端左乘A,再加上b,得

$$g_{i+1}=g_i+\lambda_i A d^{(i)} \qquad (10.3.20)$$

其中λ_i 由式(10.3.17)确定,即

$$\lambda_i=-\frac{g_i^T d^{(i)}}{d^{(i)T} A d^{(i)}}=\frac{g_i^T g_i}{d^{(i)T} A d^{(i)}}\neq 0 \qquad (10.3.21)$$

考虑到(10.3.20)和(10.3.18),则

$$g_{i+1}^T g_j = \left[g_i + \lambda_i A d^{(i)} \right]^T g_j$$

$$= g_i^T g_j + \lambda_i d^{(i)T} A(-d^{(j)} + \beta_{j-1} d^{(j-1)}) \qquad (10.3.22)$$

(注: j=1时上式为$g_{i+1}^T g_1 = g_i^T g_1 - \lambda_i d^{(i)T} A d^{(1)}$)

当$j = i$时,由归纳假设$d^{(i)T} A d^{(i-1)} = 0$, 根据(10.3.21)

$$-\lambda_i d^{(i)T} A d^{(i)} = -g_i^T g_i$$

$$\Rightarrow g_{i+1}^T g_i = 0$$

当$j<i$时,根据归纳假设,式(10.3.22)等号右端各项均为0

故 $$g_{i+1}^T g_j = 0$$

再证1),运用(10.3.18)和(10.3.20),则

$$d^{(i+1)T}Ad^{(j)} = \left[-g_{i+1} + \beta_i d^{(i)}\right]^T Ad^{(j)}$$

$$= -g_{i+1}^T \frac{g_{j+1} - g_j}{\lambda_j} + \beta_i d^{(i)T} A d^{(j)}$$

当$j=i$时,把(10.3.19)代入上式第一个等号的右端,立得

当j<i时,由前面已经证明的结论和归纳假设,式中第2个等号右端显然为0,因此

$$d^{(i+1)T}Ad^{(j)}=0$$

最后证3), 易知

$$g_{i+1}^{\;\;T}d^{(i+1)}=g_{i+1}^{T}\left[-g_{i+1}+\beta_i d^{(i)}\right]=-g_{i+1}^{T}g_{i+1}$$

综上,对i+1,上述三种关系成立

由上可知,$Fletcher-\mathrm{Reeves}$共轭梯度法所产生的搜索方向$d^{(1)}.d^{(2)},\ldots,d^{(m)}$是A共轭的, 根据$Th10.3.2$, 经有限步迭代必达极小点.

- 注意,初始搜索方向选择最速下降方向十分重要,如果选择别的方向作为初始方向,其余方向均按FR方法构造,则极小化正定二次函数时,这样构造出来的一组方向并不能保证共轭性.

例 考虑下列问题

$$\min \quad x_1^2 + \frac{1}{2}x_2^2 + x_3^2$$

取初始点和初始搜索方向分别为

$$x^{(1)} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, d^{(1)} = \begin{pmatrix} -1 \\ -2 \\ 0 \end{pmatrix}$$

显然, $d^{(1)}$不是目标函数在$x^{(1)}$处的最速下降方向.

下面,我们用FR法构造两个搜索方向.

从$x^{(1)}$出发,沿方向$d^{(1)}$进行搜索, 求步长λ_1, 使满足:

$$f(x^{(1)}+\lambda_1 d^{(1)})=\min_{\lambda\geq 0} f(x^{(1)}+\lambda d^{(1)})$$

得$\lambda_1 = 2/3$

$$\Rightarrow x^{(2)}=x^{(1)}+\lambda_1 d^{(1)}=\begin{pmatrix}1/3\\-1/3\\1\end{pmatrix}, g_2=\begin{pmatrix}2/3\\-1/3\\1\end{pmatrix}$$

$$d^{(2)}=-g_2+\beta_1 d^{(1)}$$

根据公式(10.3.19),有

$$\beta_1 = \frac{d^{(1)T} A g_2}{d^{(1)T} A d^{(1)}} = \frac{-2/3}{6} = -\frac{1}{9}$$

因此

$$d^{(2)} = -\begin{pmatrix} 2/3 \\ -1/3 \\ 1 \end{pmatrix} - \frac{1}{9}\begin{pmatrix} -1 \\ -2 \\ 0 \end{pmatrix} = \begin{pmatrix} -5/9 \\ 5/9 \\ -1 \end{pmatrix}$$

从$x^{(2)}$出发,沿方向$d^{(2)}$进行搜索, 求步长λ_2, 使满足:

$$f(x^{(1)} + \lambda_2 d^{(1)}) = \min_{\lambda \ge 0} f(x^{(2)} + \lambda d^{(2)})$$

得 $\lambda_2 = 21/26$

$$\Rightarrow x^{(3)} = x^{(2)} + \lambda_2 d^{(2)} = \begin{pmatrix} -9/78 \\ 9/78 \\ 5/26 \end{pmatrix}, g_3 = \begin{pmatrix} -18/78 \\ 9/78 \\ 5/26 \end{pmatrix}$$

令

$$d^{(3)} = -g_3 + \beta_2 d^{(2)}$$

根据公式(10.3.19),有

$$\beta_2 = \frac{d^{(2)T} A g_3}{d^{(2)T} A d^{(2)}} = \frac{45}{676}$$

因此

$$d^{(3)} = -\begin{pmatrix} -18/78 \\ 9/78 \\ 5/26 \end{pmatrix} + \frac{45}{676}\begin{pmatrix} -5/9 \\ 5/9 \\ -1 \end{pmatrix} = \frac{1}{676}\begin{pmatrix} 131 \\ -53 \\ -175 \end{pmatrix}$$

可以验证,$d^{(1)}$与$d^{(2)}$关于A共轭, $d^{(3)}$与$d^{(2)}$关于A 共轭, 但$d^{(1)}$与$d^{(3)}$不关于A共轭, 于是$d^{(1)}, d^{(2)}, d^{(3)}$不关于A 共轭.

- **注意,在FR法中,初始搜索方向必须取最速下降方向**

- 可以证明,对于正定二次函数,运用FR法时不作矩阵运算就能求出因子β_i

定理10.3.4 对于正定二次函数,FR法中因子β_i具有下述表达式

$$\beta_i = \frac{\|g_{i+1}\|^2}{\|g_i\|^2}, \qquad (i \ge 1, g_i \ne 0) \qquad (10.3.24)$$

证明:

$$\beta_i = \frac{d^{(i)T} A g_{i+1}}{d^{(i)T} A d^{(i)}} = \frac{g_{i+1}^T A(x^{(i+1)} - x^{(i)})/\lambda_i}{d^{(i)T} A(x^{(i+1)} - x^{(i)})/\lambda_i}$$

$$= \frac{g_{i+1}^T (g_{i+1} - g_i)}{d^{(i)T} (g_{i+1} - g_i)} = \frac{\left\| g_{i+1} \right\|^2}{-d^{(i)T} g_i} \qquad (10.3.23)$$

根据定理10.3.3, $d^{(i)T} g_i = -\left\| g_i \right\|^2$.因此

$$\beta_i = \frac{\left\| g_{i+1} \right\|^2}{\left\| g_i \right\|^2}, \qquad (10.3.24)$$

共轭梯度法

- FR法(对二次凸函数)

1, 给定初点$x^{(1)}$, 置$k=1$.

2, 计算$g_k=\nabla f(x^{(k)})$. 若$\|g_k\|=0$, 停止计算, 得点

$\bar{x}=x^{(k)}$; 否则, 进行下一步.

3, 构造搜索方向. 令

$$d^{(k)}=-g_k+\beta_{k-1}d^{(k-1)}$$

其中, 当$k=1$时, $\beta_{k-1}=0$,

当$k>1$时按公式(10.3.24)计算β_{k-1}.

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

其中按公式(10.3.17)计算步长λ_k.

5, 若$k = n$, 则停止计算, 得点$\bar{x} = x^{(k+1)}$

否则, 置$k := k + 1$, 转2

共轭梯度法

例3.2 用FR法求解下列问题

$$\min \quad f(x) = x_1^2 + 2x_2^2 \qquad \text{初点} x^{(1)} = (5,5)^T$$

目标函数$f(x)$在点x处的梯度

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

第一次迭代，

令

$$d^{(1)} = -g_1 = \begin{bmatrix} -10 \\ -20 \end{bmatrix}$$

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

$$\lambda_1 = -\frac{g_1^T d^{(1)}}{d^{(1)T} A d^{(1)}} = \frac{(-10,-20)\begin{pmatrix}-10\\-20\end{pmatrix}}{(-10,-20)\begin{pmatrix}2 & 0\\0 & 4\end{pmatrix}\begin{pmatrix}-10\\-20\end{pmatrix}} = \frac{5}{18}$$

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

第2次迭代

目标函数在点x $^{(2)}$ 处的梯度 $g_2 = \begin{pmatrix}40/9\\-20/9\end{pmatrix}$

构造搜索方向$d^{(2)}$.先计算因子β_1

$$\beta_1=\frac{\|g_2\|^2}{\|g_1\|^2}=\frac{(40/9)^2+(-20/9)^2}{10^2+20^2}=\frac{4}{81}$$

令

$$d^{(2)}=-g_2+\beta_1 d^{(1)}=\frac{100}{81}\begin{bmatrix}-4\\1\end{bmatrix}$$

从$x^{(2)}$出发,沿方向$d^{(2)}$作一维搜索,求λ_2:

$$\lambda_2=-\frac{g_2^T d^{(2)}}{d^{(2)T}Ad^{(2)}}=\frac{-\frac{20}{9}\cdot\frac{100}{81}(2,-1)\begin{pmatrix}-4\\1\end{pmatrix}}{\left(\frac{100}{81}\right)^2(-4,1)\begin{pmatrix}2&0\\0&4\end{pmatrix}\begin{pmatrix}-4\\1\end{pmatrix}}=\frac{9}{20}$$

$$x^{(3)} = x^{(2)} + \lambda_2 d^{(2)} = \begin{bmatrix} 20/9 \\ -5/9 \end{bmatrix} + \frac{9}{20} \cdot \frac{100}{81} \begin{pmatrix} -4 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

显然点$x^{(2)}$处目标函数的梯度$g_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, 已达极小点$\begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

■ 一般函数的共轭梯度法－非线性共轭梯度法

一般迭代格式

$$\begin{cases} x^{k+1} = x^k + \alpha_k d^k \\ d^{k+1} = -g^{k+1} + \beta_k d^k \end{cases} \quad k=0,1,\ldots \qquad (10.3.3.1)$$

其中初始方向$d^0 = -g^0$,步长参数α_k由一维搜索得到，β_k的计算公式通常有如下几种：

1, $$\beta_k = \frac{(g^{k+1})^T g^{k+1}}{(g^k)^T g^k} \qquad \text{(Fletcher-Reeves(FR))}$$

2, $$\beta_k = \frac{g_{k+1}^T (g_{k+1} - g_k)}{g_k^T g_k}$$ ----PRP(Polak-Ribiere-Polyar)

3, $$\beta_{\mathrm{k}} = \frac{g_{k+1}^T (g_{k+1} - g_k)}{(d^k)^T (g_{k+1} - g_k)}$$ -----SW(Sorenson-Wolfe)

4, $$\beta_k = \frac{(d^k)^T \nabla^2 f(x^{k+1}) g_{k+1}}{(d^k)^T \nabla^2 f(x^{k+1}) d^k}$$ ----Daniel

5, $$\beta_{\mathrm{k}} = \frac{g_{k+1}^T g_{k+1}}{(d^k)^T g_k}$$ -----Dixon

FR共轭梯度法

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

$y^{(1)} = x^{(1)}, d^{(1)} = -\nabla f(y^{(1)}), k = j = 0.$

2, 若$\left\| \nabla f(y^{(j)}) \right\| < \varepsilon$, 则停止计算, 否则作一维搜索,

求λ_j满足 $f(y^{(j)} + \lambda_j d^{(j)}) = \min_{\lambda} f(y^{(j)} + \lambda d^{(j)})$

令 $y^{(j+1)} = y^{(j)} + \lambda_j d^{(j)}$

3,如果$j < n$,转步4,否则,转5

4,令 $d^{(j+1)} = -\nabla f(y^{(j+1)}) + \beta_j d^{(j)}$

其中 $$\beta_j = \frac{\left\|\nabla f(y^{(j+1)})\right\|^2}{\left\|\nabla f(y^{(j)})\right\|^2}$$

置$j := j+1$,转步2.

5,令 $x^{(j+1)} = y^{(n+1)}, y^{(1)} = x^{(k+1)}, d^{(1)} = -\nabla f(y^{(1)})$

置$j = 1, k := k+1$,转步2.

可以证明,对一般函数,共轭梯度法在一定条件下是收敛的,

FR算法中使用精确线搜索，我们有如下收敛性结果

定理 假设函数$f: R^n \to R$有下界，梯度$\nabla f(x)$是Lipschitz连续的.在FR共轭梯度法中，步长参数α_k是由精确线搜索确定的，并且满足充分下降条件(即Armijo条件). 若$\forall k \geq 0, g^k \neq 0$, 则

$$\liminf_{k\to+\infty} \left\| g^k \right\| = 0$$

(Armijo条件:选择步长λ_k满足

$$f(x^k + \lambda_k d^k) \leq f(x^k) + c_1 \lambda_k \nabla f(x^k)^T d^k \quad)$$

小结

- 使用导数基本概念
- 最速下降法
- 牛顿法
- 共轭梯度法
- 拟牛顿法
- 信赖域方法

■ 作业

- 习题 3、4、5、10、14、17、18、19