# 第十讲　共轭梯度法

* 共轭向量
* 线搜索
* 共轭方向
* 算法收敛性
* 共轭方向法

使用导数的最优化方法 共轭梯度法

共轭方向法是介于最速下降法和*Newton*法之间的一种方法。克服了最速下降法的锯齿现象，从而提高了收敛速度；同时，共轭方向法的迭代公式比较简单，不必求目标函数的*Hessian*矩阵，比*Newton*法减少了计算量和存储量。是一种比较实用而且有效的方法。   
在讲共轭方向法和共轭梯度法之前，先对共轭向量进行说明。

<https://blog.csdn.net/u012430664/article/details/78551512>

## 共轭向量及其性质

定义1：(共轭方向) 是  对称正定矩阵，若  维向量空间中的非零向量满足

(1)

则称是的共轭向量或称向量是共轭的(简称共轭)，也称的方向是的共轭方向。

当（单位矩阵）时，公式(1)变为

(2)

即向量互相正交。由此可见，正交是共轭的一种特殊情况，共轭是正交的推广。

例：

则

定理2：设是阶对称正定矩阵，若是共轭的非零向量，则他们是线性无关的。

证明：设存在数,使得

左乘,

因为是共轭的非零向量，所以有

又因为正定，所以

所以线性无关。

推论3： 在维的向量空间中，非零的共轭向量个数不超过。   
定义4： 设是中线性无关向量，，那么由形式为

, 其中是任意实数

的向量构成的集合为由点和向量所生成的线性流行。记为。

## 二、共轭方向法

## 基本思想

在考虑普通函数之前，我们首先用二元正定二次函数进行讲解。首先考虑如下的正定二次函数

(3)

要求的目标函数的最优值，根据最速下降法的思想，我们首先选定一个初始点，然后沿着该点的最速下降方向做精确搜索，得到点，由最速下降法的性质可知

(4)

即与点处的等值线相切。   
在第二次迭代过程中我们不是用作为这次迭代的搜索方向，我们想只在第二次迭代之后能直接到达最优点，那么这次的迭代方向应该满足什么条件呢?   
首先根据迭代公式我们有

(5)

其中是最优步长因子,显然在未到达最优点之前，是不等于0的。对目标函数求梯度，有

(6)

对于极小点，我们有

即

在上式两边同乘以，由于公式(4)，并且我们可以得到

(7)

由公式(7)我们知，和为的共轭向量。

现在我们假设

(8)

在上式两侧同时乘以，得到

=0

解得

带入到公式(8)得到

## 三、算法收敛性

定理5：设有二次函数

其中是对称正定矩阵，是共轭的非零向量，从任意一点出发出发，依次沿这组向量进行一维搜索

则

证明： 因为,所以有

所以

两边同时右乘

由于及是共轭的非零向量，所以

。

证毕

定理6：设有二次函数

其中是对称正定矩阵，是共轭的非零向量，从任意一点出发出发，依次沿这组向量进行一维搜索

则至多经过步收敛，即是在上的极小点。

证明：

法一：设是线性流行的极小点，由定理5可知

对任意的，利用二阶*Taylor*展开式和的正定性得

因此是在上的极小点，即是在上的极小点。

证毕

法二：设是在上的极小点，则

因为为一组共轭的非零向量,所以线性无关，即构成 的一组基。

一方面 可由这组向量线性表出，令

两边左乘：

另一方面，由依次递推得

两边左乘

与(9)相比较得

所以。

证毕

综上所述，对于元正定二次函数，我们可以从任意点出发，然后沿着这个共轭方向最多做次直线搜索，就可以求的目标函数的最优点。

前面已经说过，共轭是正交的推广，对于维空间，我们可以把(不超过)个共轭向量作为维空间的基，只不过这个基不再是正交的，那么共轭向量法就是对于定义在空间中的函数，沿着每一个基的方向做直线搜索，那么最多做次搜索就能得到最优值。对于定理5，他所指出的就是最后一个迭代点处的梯度与之前的搜索方向垂直,因此根据定理5我们可以得到如下的推论。

推论7：在定理5中，的任意线性组合都与正交。

对于正定二次函数(3)，从任意点出发，顺次沿个共轭方向做直线搜索，最多经过次就可以到达极小点。

注意：不同求共轭方向的方法对应了不同的共轭方向法。

## 四、共轭梯度法

共轭梯度法：初始点的搜索方向为的负梯度方向；之后迭代点的搜索方向为该点的负梯度方向与已经得到的搜索方向的线性组合（即）。共轭梯度法是一种共轭方向法，在求每一个迭代点的搜索方向时，与该点的梯度有关，故叫做共轭梯度法。

首先对正定二次函数做说明。

### 用于正定二次函数的共轭梯度法

共轭梯度法不需要预先给定共轭方向，而是随着迭代的进行不断产生共轭方向。在每次的迭代中，利用上一个搜索方向和目标函数在当前迭代点的梯度向量 之间的线性组合构造一个新的方向，使其与前边已经产生的搜索方向组成共轭方向。对于一个维二次函数，沿着共轭方向进行搜索，经过次迭代，即可得到极小点。

(10)

选取初始点，取,从出发，沿做一维搜索，

当时，

若（否则迭代终止），则，所以与线性无关。

令

用左乘上式，并令，得

从出发，沿做一维搜索，得

以此类推，得

(11)

(12)

(13)

(14)

共轭梯度法的算法描述如下：

已知：二次函数(10)，终止限

⑴选定初始点；计算；置；

⑵作直线搜索

或者采用如下公式计算

⑶判断是否满足要求，满足则输出停止；否则转(4)

⑷计算

⑸置，转(2)

对于目标函数为(10)的正定二次函数来说。

\* 第一个迭代点

我们可以任意指定一个初始点，那么初始点处的搜索方向，从出发沿着方向做直线搜索

可以求得(可以参照公式(14)的推导)

故

由此我们便得到了第一个迭代点。   
\* 第二个迭代点

由直线搜索的特性我们可以知道(可以想一下最速下降法中的锯齿现象)，故和是线性无关的。根据共轭梯度法的特性我们可以得到在点处的搜索方向为

再由共轭方向法的特性我们知道与是的共轭方向。故

便可以得到

由此我们便确定了迭代点处的搜索方向，之后便是从开始沿搜索方向做直线搜索，根据公式(14)，可以得到

由此我们便得到了第二个迭代点   
\* 第三个迭代点   
同理，我们有,可以得到点处的搜索方向为

同理，和是共轭方向，能够得到

便可得到

至此，我们便得到了点处的搜索方向，然后从开始沿方向做直线搜索，与上相同便可得到，求得

至此我们便得到了第三个迭代点。   
但是上面在求第三个迭代点时有一个小小的问题便是，以上求得的方法能保证和共轭吗？即是共轭向量(根据共轭向量定义我们知道共轭向量之间需要两两共轭)吗？答案是肯定的，接下来推导如下：

由于与都可以表示成和的线性组合，故有 (因为可以把和看作是由和确定的超平面上的线，由于,故有)。   
关于为什么，可以参考公式(14)的推导过程，有

，由此得到。

 \* 第个迭代点   
与前面相似，我们可以得到

可以得到

我们便能得到处的搜索方向，在处沿方向做直线搜索，同理我们能得到

同理我们能够证出是共轭向量。

所以可以按照上述方法，依次构造出共轭向量，最多经过次迭代就能找到最优点。

定理8：对于正定二次函数，共轭梯度法在次一维搜索后终止，且对，下列关系式成立：

⑴;

⑵;

⑶

证明：（用归纳法）

当时，，所以(3)成立。

当时，由构造法，，即(1)成立。

由一维搜索性质，，而

(2)成立。

因为

所以

因此得出(3)成立。

假设对某个，(1)，(2)，(3)均成立，下证对也成立。先证(2)

(2)

即,其中

所以

(1)

若，由。

以下假设

因为

所以

因此

(3)

所以

因此。

证毕

定理9：

证明：由得

证毕

注：共轭梯度法中的取值不同又可细分FR、PRP等方法，上述定理的取值所对应的方法被称为FR共轭梯度法。

根据上面定理，我们可以重新给出算法框架：

**FR共轭梯度法**

1. 给定初始点，置。
2. 计算，若，则停止计算，得点；否则进行下一步。
3. 令，其中

当时，，当时，

1. 令，其中
2. 若，则停止计算，得点；否则，置，返回2。

## 四、一般函数的共轭梯度法

上面所讲的迭代公式要想适用于非二次函数，就要将迭代公式(12)中的去掉，那么根据的推导我们可以得到

由此，我们能得到迭代公式

(15)

在公式(11)两边同时右乘，得

由直线搜索的性质，我们知道

(16)

(17)

另外

(18)

将公式(16)(17)(18)带入到公式(15)之后，我们得到

(19)

这个公式称为Fletcher-Reeves公式。   
将公式(16)(17)带入到公式(15)后得到

(20)

这个公式称为Dixon-Myers公式。   
将公式(16)(18)带入到公式(15)后得到

(21)

这个公式称为Polak-Ribiere公式。   
将公式(19)(20)(21)替换公式(12)对应了不同的共轭梯度法。共轭梯度法不要求精确的直线搜索，但是不精确的直线搜索可能会造成之后迭代出来的向量不再是共轭的，这将会降低共轭梯度法的效能，解决方法就是重设初始点，即经过次迭代后得到的作为初始点，开始新一轮的迭代。

一般函数的共轭梯度法：

1. 步长不能用计算，必须用其他一维搜索方法来确定。
2. 凡用到矩阵之处，需用现行点处的*Hessian*矩阵代替。
3. 有限步迭代达不到极小点。

迭代的延续方法：

1. 直接搜索：即总用公式构造搜索方向。
2. 重新开始，把步作为一轮，没搜索一轮之后，取一次最速下降方向，开始

下一轮。

用于一般函数的*Fletcher-Reeves*共轭梯度法描述如下。

已知：目标函数以及梯度函数，问题的维数以及终止准则的终止限以及终止限

1. 选定初始点，计算；置
2. 进行直线搜索

计算，

1. 判断终止准则是否满足，若满足：输出，停止；否则：转(4)
2. 判断是否成立，即已经迭代了次，若是：重置，（）然后转(2)；不是：转(5)
3. 按照Fletcher-Reeves公式计算
4. 判断是否成立（检查是不是下降方向，由于实际计算中无法精确求解，可能会出现不是下降方向的情况；在精确求解中，一定是下降方向），做三种情况处理：

i)若，这时则改取作为搜索方向，并置，转(2)；

ii)若，则重置，，转(2)；

iii)若，则就作为搜索方向，并置，转(2)。

FR共轭梯度法

1. 给定初始点，允许误差，置，。
2. 若，则停止计算；否则做一维搜索，求满足

令。

3．若，则转4，否则转5。

4. 令，置，转2。

5.令，返回2。

<http://www.codelast.com/?p=8095>

共轭方向法是介于最速下降法和牛顿法之间的一种存在——它的收敛速度（二阶收敛）比最速下降法（线性收敛）快，同时它的计算量又比牛顿法要小，因此它的存在是有意义的。共轭梯度法不需要预先给定共轭方向，而是随着迭代的进行不断产生共轭方向。在每次的迭代中，利用上一个搜索方向和目标函数在当前迭代点的梯度向量 之间的线性组合构造一个新的方向，使其与前边已经产生的搜索方向组成共轭方向。对于一个维二次型函数，沿着共轭方向进行搜索，经过次迭代，即可得到极小点。

需要注意，共轭方向法可以不使用目标函数的一阶导数信息（当然也可以使用）。所以，如果目标函数的一阶导数不容易求的话，共轭方向法可能就可以派上用场了。

共轭方向法的显著特征就是：两次搜索方向之间是有关联的，这种关联就是“共轭”。

参考文献

1. 《Convex Analysis and Optimization》 Bertsekas, Nedic, and Ozdaglar, 2015
2. 《Numerical Optimization》Jorge Nocedal, Stephen Wright， 1999(2006)
3. 《Convex Optimization》 Stephen Boyd, Lieven Vandenberghe， 2004
4. 《最优化理论与算法》陈宝林，1989(2005)编著