**Optimization Methods Final Reports**

孙仲煜&3119307082

西安交通大学数学与统计学院 西安 邮编：710000

**摘要：**通过《最优化方法》的课程学习，我了解到，优化算法大体上可分为两类，无约束最优化算法与约束最优化算法。在设计优化算法时，我们需要用到一些基本的数值技术，包括线性方程组求解，矩阵分解与矩阵修正，线性搜索技术及信赖域子问题的求解。对于无约束优化问题，求解中小规模最优化问题适用拟牛顿法；求解大规模优化问题时，共轭梯度法以及有限内存拟牛顿法较为适用。对于约束优化问题，根据约束条件的形式又可分为线性约束与非线性约束，其中线性约束优化问题适合用以消去法为主体的可行点算法进行求解；而对于非线性约束优化问题，目前来说还不存在较为通用的方法对所有非线性约束问题进行求解，现存方法主要有罚函数法、乘子法、可行方向法以及SQP法。本文主要通过介绍最优化方法的上述关键技术与方法，结合学习过程中的优化问题求解经验，谈一谈自己的学习心得。

**关键词：**无约束优化、约束优化、线性搜索

# 1 优化的关键技术与方法

最优化问题的求解最后都可以归结为求一个目标函数的极值，而最优化方法的关键技术与方法则是寻找极值点的工具，其形式通常为通过不断迭代，使所选的初始点不断逼近极值点从而获取目标函数局部极小值的方法，那么解迭代后新解移动的方向与步长如何确定呢？这就要根据不同的优化问题选择不同的优化算法，也是我接下来要着重介绍的内容。

## 最优化基础

最优化问题的数学模型一般形式为

其中，，*x*称为优化变量或决策变量，*f(x)*称为目标函数，称为约束函数，称为等式约束，称为不等式约束。和称为约束条件。

定义1：设，，，若存在数，使得

则称向量是在处的下降方向。

定理1：设，在点处可微，若存在向量，使得

则向量是在处的下降方向。

定义2：设，且，如果存在，使得

成立，则称*d*是在处关于集合*D*的一个可行方向。在处关于集合*D*的所有可行方向构成的集合记为。

定义3：设，，若存在序列和，使得

且有和。则称*d*是在处关于集合*D*的一个序列可行方向。在处所有的序列可行方向构成的集合记为。

定义4：设，且，如果

则称是在处关于集合的线性可行方向。在处所有的线性化可行方向构成的集合记为即

引理1：设，如果所有的约束函数都在处一阶连续可微，则

定理2(K-T条件)：设是(NP)的一个局部最优解，和在的某一邻域内一阶连续可微，若

则必存在向量，使得

成立。

对于非线性规划(1.1)，若目标函数是凸函数，约束集合是凸集，则称非线性规划(1.1)是凸规划。对于凸规划，其最优解具有非常好的性质，其局部最优解必是全局最优解。

定理2(凸优化解判定)：设目标函数和约束函数一阶连续可微，若凸规划的可行点是K-T点，则必是最优解。

求解非线性规划(1.1)通常是通过求其K-T点的途径得到它的最优解，一般采用是的迭代法，其基本思想是，给定一个初始点，按照某一规则产生一个迭代序列，使得当是有限点列时，它的最后一个点是K-T点；当是无穷点列时，它的任意一个聚点就是K-T点。迭代序列使得评价函数序列充分单调减少的性质保证了K-T点是局部最优解。

求解非线性规划的算法基本迭代格式如下：

给定初始点

按某一规则构造搜索方向.

确定步长

取下一个迭代点

判别是否满足某些终止准则，若满足，停止迭代过程，是近似局部最优解；否则，转第步。

在上述迭代格式中，关键的两步是构造搜索方向和确定步长。事实上，构造搜索方向和确定步长有许多种方法，不同的构造方向与确定步长的方法构成了求解非线性规划的不同算法。

在有了迭代点和搜索方向后，迭代步长应使非线性规划的评价函数沿射线有所下降，即

由于和已知，评价函数是关于的一元函数，故将求的问题称为线性搜索。

评价函数用于评价迭代过程的进展，在无约束最优化问题中，一般选取评价函数为目标函数，在约束问题中，评价函数包含目标函数和约束函数的信息，以便能判断迭代点的可行性以及评价函数的下降性。对约束问题已经提出了许多不同的评价函数，如由Frishch，Carroll提出的两种障碍函数，由Zangwill提出的精确罚函数及Fletcher(1973)提出的可谓精确罚函数等等。

对于无约束问题，当目标函数用它在处的一阶泰勒展开式来逼近时，得到的求解方法称为梯度法。当目标函数用它在处的二阶泰勒展开式来逼近时，得到的求解方法称为牛顿法。由于二阶泰勒展开式要用到目标函数的二阶导数矩阵，计算工作量大，若用一个由只包含的一阶导数信息构造的对称正定矩阵来代替二阶泰勒展开式中目标函数的二阶导数矩阵时，相应的二次逼近方法称为拟牛顿法。在处近似子问题的解具有目标函数的负梯度方向左乘一个对称正定矩阵的形式，即

当 时，为梯度法；当时，为牛顿法；当用布鲁丹族或黄族修正公式来构造时，便是拟牛顿法。

一维搜索方法的优势及收敛的快慢直接影响到最优化算法的收敛速度，最理想的情况是确定步长使目标函数下降尽可能的大，即要求步长满足

若记，则求满足上式的归结为求一元函数的极小化问题

由于这种搜索是以取得最佳步长为目的，因此称它为精确线性搜索，所得的步长称为最佳步长或最优步长。还有一类搜索方法为不精确的线性搜索法，它只利用一点处的目标函数值和梯度值，它在迭代的每一步使目标函数充分下降，具有计算量较小，算法易于实现等有点。

求解约束最优化问题的方法大致可分为三类。一类是将约束问题转化为一系列无约束问题来求解，用这一系列无约束问题的极小点取逼近原约束问题的最优解，故称这类方法为序列无约束极小化方法。该类具有代表性的方法由惩罚函数法和拉格朗日乘子法；第二类方法是在每个迭代点处构造一个二次函数去逼近目标函数，用线性函数逼近约束函数，在迭代的每一步构成一个二次规划子问题，将对原约束问题的求解转化为对一系列二次规划子问题的求解，以子问题的解作为本次迭代的搜索方向，沿寻优得到新的迭代点，迭代序列最终逼近原约束问题的最优解，故将这类算法称为序列二次规划法。各种方法的差异主要在于构造二次函数的海森矩阵的方法不同；第三类方法是可行方向法，它是一种直接处理约束的方法，它将无约束优化方法推广应用于约束问题，其特点为在迭代过程中目标函数逐次下降且迭代序列始终处于约束区域内，即在可行迭代点处确定一个搜索方向，使得

且是可行点，则称为可行下降方向。这类方法的关键是可行方向的构造，不同的构造的方法构成了不同的可行下降方法。像这种由构造可行下降方法来建立求解约束优化的方法统称为可行方向法，具有代表性的方法有Topkis-Veinott可行方向法、投影梯度法、简约梯度法等等。

误差函数：对于求解非线性规划(NP)的一个数值方法，评价这个方法优劣的标准之一是由该方法产生的迭代序列的收敛速度。设序列收敛于，我们用误差函数

来度量收敛速度。

## 1.2搜索策略

在优化问题中，的选取至关重要，求解的方法有精确线性搜索法，直接搜索法，插值法和不精确线性搜索法等几类，各类方法具有各自的特色，下面分别予以介绍。

### 1.2.1 精确线性搜索

精确线性搜索就是要求选取步长满足

以便获得最大可能的下降，可以通过一元函数的极小化

来求得最优步长，即准则：

在精确线性搜索中，是的精确极小点，而在实际应用中，除了是二次函数的特殊情况外，这是不可能实现的，因为我们不可能在有限步数中求得的精确解，我们智能得到满足一定精度的近似解，退一步讲，求精确解是低效且无必要的一件事情。然而，精确线性搜索还是具有一定的理论价值的，因为许多无约束优化算法的收敛性和收敛速度都是基于精确线性搜索的。

### 1.2.2 搜索区间的确定

在精确线性搜索中要求解方程需要用到目标函数梯度和二阶导数矩阵，计算工作量打，对于某些非光滑函数或导数表达式复杂的函数就不能利用精确线性搜索，对于这样的函数求，可利用直接搜索法和插值法。但是这两种方法都需要事先知道包含的搜索区间，然后在区间上求的近似极小点，下面我们介绍一种确定搜索区间的方法。

该方法的基本思想是：假定是连续函数，逐步确定三点(其中)，使得满足

则在与之间必有的一个极小点，我们便取为搜索区间，具体过程如下：

给定，计算，取步长，令，计算，若，则令，计算；若，则取，则得到搜索区间；否则，令，计算，如此操作，由于是局部凸函数，故存在某个使得

取，则得到一个搜索区间，满足。

若在一开始有，则将两个点的标注互换，则有，。再向左寻找第三个点，令，计算，若，则取，否则，令，计算，如此操作，直到对某个，有，这时取，则得到一个搜索区间，满足。

### 1.2.3 黄金分割法

黄金分割法又称0.618法，是一种简单易行、在实际应用中十分广泛的求在区间上极小点的直接搜索法，是一种线性搜索方法。

该方法的基本思想是：在搜索区间上选取两个对称点，通过比较这两点处的函数值的大小来决定删除左半区间，还是删除右半区间。删除后的新区见的长度是原区间长度的0.618倍，新区间包含原区间中两个对称点中的一个，关于这一点再选取一个对称点，根据新的两对称点处的函数值来确定新区间的删除，不断进行上述过程，直到新区间的长度小于预先指定的精度为止，最后根据最终区间中两个对称点的函数值来确定极小点。

综上所述，0.618法的算法步骤如下：

|  |
| --- |
| 0.618法 |
| 1. 确定，记，，令，给定，. 2. 计算：   ，，  计算：，    ，，，  计算：，   1. ，        1. 输出，停机； |

### 1.2.4 Golden-Stein法

上面介绍的精确线性搜索法和0.618法都是线性搜索法，求得的是的精确极小点或近似极小点，用这些方法求得的的极小点，虽然是的目标函数在每次迭代中下降较多，但计算工作量较大；另一方面来说，在求目标函数的最优解时，没有必要将线性搜索搞得非常精确，尤其是在计算的初始阶段。因此，我们放松对的要求，只要求目标函数在迭代的每一步都有充分下降即可，这一大类方法被称为不精确线性搜索法，具有简单、计算量小的特点，其中应用较为广泛的是Golden-Stein算法。

下面给出Golden-Stein不精确线性搜索法的计算步骤：

|  |
| --- |
| Golden-Stein |
| 1. 给定初始搜索区间，及，，， 2. 计算，，令，，取， 3. 计算   若，则转4；否则，，转5.   1. 若，则，输出；否则，   若，则转5；否则，，转3.   1. ，，转3. |

## 1.3无约束优化方法

无约束优化方法是优化技术中极其重要和基本的内容，它不仅可以直接用来求解无约束优化问题，也可以求解被转化为无约束优化问题的约束优化问题。最速下降法和牛顿法是较为常见的求解无约束问题的最优化方法，这两种算法作为基本算法，在最优化方法中占有重要地位。

### 1.3.1 最速下降法

最速下降法(Steepest Descent)是求解无约束问题最简单的方法，其优点是工作量少，存储变量较少，对初始点要求不高；缺点是收敛慢，效率低。

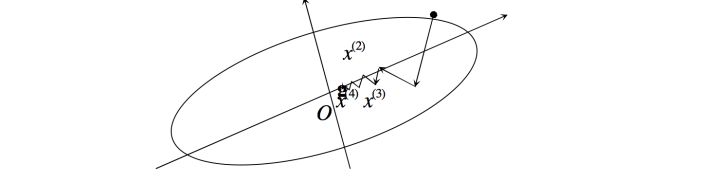
在基本的迭代公式中，每次迭代搜索方向取为目标函数的负梯度方向，即，而每次迭代的步长由线性搜索策略决定，由此确定的算法称为最速下降法。

定理6：设一阶连续可微，是由采用精确线性搜索的最速下降法产生的迭代序列，则的每一个聚点都是的平稳点。

定理7：设是二阶连续可微函数且，对任意给定的初始点，采用精确线性搜索的最速下降法或有限步终止，会有或。

上述两个定理阐述了最速下降法在在选取初始点后迭代序列的性质，并且由于最速下降法的搜索方向是函数在点处的负梯度方向，在点处沿下降的最快，这是函数的局部性质，但对整个求最优解的过程来说不一定是最快的，例如，对于采用精确线性搜索的最速下降法，步长是的极小点，故满足

从而有，这说明最速下降法中相邻两次的搜索方向是互相正交的。对于目标函数是二元函数的极小化问题，从几何上看最速下降法产生的迭代点列移向极小点的路径呈“锯齿状”如下图所示，在迭代开始的最初几步，下降的幅度比较大，但在以后迭代中，随迭代点越来越接近极小点，下降的越来越慢，因而的收敛速度较慢。因此，在计算的开始阶段使用最速下降法较为适合。



综上所述，最速下降法的算法步骤如下：

|  |
| --- |
| 最速下降法(Steepest Descent) |
| 1) 令*k = 0*, 取初始点 迭代精度*,* 根据下述方式进行迭代：  2)  3)  4)  5)  6) 迭代停止条件：< 或 *k+1 = n*, 则停止迭代；否则继续进行如上步骤。 |

### 1.3.2 共轭梯度法

定理3：(求解线性方程组与求解凸二次函数极小点的等价性)：设，其中是阶对称正定矩阵则是线性方程组的解的充要条件是是的极小点，即

共轭梯度法最初是由计算数学家Hestenes和几何学家Stiefel于1952年为求正定系数矩阵线性方程组而独立提出，他们合作的文章*Method of conjugate gradients for solving linear systems*被认为是共轭梯度法的奠基性文章。共轭梯度法的基本思想是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，并沿着此方向进行搜索，求出目标函数的极小点。它仅需利用一阶导数信息，但克服了最速下降法收敛慢的缺点，又避免了牛顿法需要存储和计算海森矩阵并求逆的缺点，共轭梯度法不仅仅是解决大型线性方程组最有用的方法之一，也是解大型非线性优化最有效的算法之一。

定义5(共轭方向)：设是个非零变量，是阶对称正定矩阵，如果对于任意的，有

则称是关于共轭的向量组，或简称是关于共轭的向量。

定理4：关于共轭的向量组中的向量是线性无关的。

定理5：设是阶对称正定矩阵，是一组关于共轭的向量，求解以下二次函数的极小问题：

若从任意的初始点出发，依次沿方向进行精确线性搜索，则至多经过次迭代，可达到上述二次问题的极小点。迭代公式如下：

定义6(残向量)：记，称为在处的残向量。

则的求法如下：

而共轭向量的构造会使用到如下定理。

定理6：设向量线性无关，则这组向量可以构造个关于共轭的向量。

则使用上述定理可选取：

从而得出共轭向量求法如下：

综上所述，共轭梯度法的计算步骤如下:

|  |
| --- |
| 共轭梯度法(Gradient Descent) |
| 1) 任意给定初始点及精度  2)  3)  4)  5)  6) 迭代停止条件：< 或 *k+1 = n*, 则停止迭代；否则继续进行如下步骤。  7)  8) |

### 1.3.3牛顿法

牛顿法的基本思想是在目标函数的极小点的近似点附近将二阶泰勒展开，用展开的二次函数取逼近，将这个二次函数的极小值点作为的一个新的近似点。用一系列二次函数的极小点去逼近的极小点。

设二阶连续可微，是的一个近似点，有泰勒公式有：

令，即

若正定，则存在，由此解出就是二次函数的极小点，即

我们将作为的一个新的近似点，我们称上式为牛顿迭代公式，相应的算法称为牛顿法。

综上所述，牛顿法的计算步骤如下：

|  |
| --- |
| 牛顿法(Newton Method) |
| 1) 令，任意给定初始点及精度  2)  3)  4)  5) 迭代停止条件：< 或 *k+1 = n,* 则停止迭代；否则继续进行如上步骤。 |

### 1.3.4 拟牛顿法

拟牛顿法是求解非线性优化问题最有效的方法之一，在牛顿法的迭代中，我们需要计算海森矩阵的逆矩阵，这一计算比较复杂，换种方式考虑用一个阶矩阵来近似替代逆矩阵，这就是拟牛顿法的基本思想。

首先，我们需要构造目标函数在当前迭代的二次模型：

这里是一个对称正定矩阵，于是我们取上述二次模型的最优解作为搜索方向，并且得到新的迭代点，其中我们要求步长满足条件。这样的迭代与牛顿法类似，其区别在于用近似的矩阵取代替海森矩阵。所以拟牛顿法最关键的地方就在于每一步迭代中的更新。现在假设我们得到一个新的迭代，并得到一个新的二次模型：

我们尽可能利用上一步的信息来选取。具体地，我们要求，从而得到

上述公式被称为割线方程,为保证能较好的近似，构造矩阵的原则，除了只能利用目标函数及其一阶导数信息外，还要满足以下三点：

1. 是对称正定矩阵，从而是下降方向。
2. 是由对的修正得到的，即
3. 必须满足拟牛顿条件

其中，.