#### 1. 无约束优化问题

本章讨论下述无约束优化问题的求解方法



其中是二次可微凸函数（这意味着是开集）。我们假定该问题可解，即存在最优点。用表示最优值。

既然是可微凸函数，最优点应满足



上述的个方程含有个变量。在一些特殊情况下，我们可以解析求解方程确定优化问题的解，但在一般情况下，必须采用迭代算法求解方程，即计算点列使得时，这样的点列被称为优化问题的极小化点列。当时算法将终止，其中是设定的容许误差值。

初始点和下水平集

本章介绍的方法需要一个适当的初始点，该初始点必须属于，并且下水平集



必须是闭集。如果是闭函数，即它的下水平集是闭集，上述条件对所有的均能满足。定义域为或开集的连续函数都是闭函数。

* **强凸性及其含义**

本章大部分内容中我们都假设目标函数在上是**强凸的**，这是指存在使得



对于任意的都成立。

强凸性能够导致若干有意义的结果。对于，我们有



其中为线段上的某一点。利用强凸性假设，上式右边最后一项不会小于，因此不等式



对中任意的和都成立。当时，上式变回描述凸性的基本不等式；当时，对的下界我们可以得到比单独利用凸性更好的结果。

我们首先说明上述不等式可以用来界定，所得上界可表明是其目标值和最优目标值的偏差正比于的次优解。对任意固定的，不等式的右边是的二次凸函数，令其关于的导数为0，得到该二次函数的最优解。因此我们有



既然该式对任意的成立，我们又可得到



由此可见，任何梯度足够小的点都是近似最优解。因此上式是最优性条件的推广。由于



我们也可将其解释为次优性条件。

关于与最优解的距离上界

对于和最优解之间的距离，也可以建立正比于的上界



为了获得该不等式，我们首先将代入



其中导出第二个不等式用了Cauchy-Schwarz不等式。因为，必有



因此



从该式可以看出最优解是唯一的。

关于的上界

本身作为一个下水平集有界，又由于的最大特征值是在上的连续函数，所以它在上有界，即存在常数使得



对所有都成立。类似地可以得到



以及



下水平集的条件数

对于任意的都有



因此，比值是矩阵的条件数（其最大特征值和最小特征值之比）的上界。

强凸性常数

我们必须记住，只有在很少的情况下才可能知道常数和，因此



并不能用作算法停止准则，我们只能把它视为一个**概念上的**停止准则。但是这一概念是有启发作用的：如果我们在时终止算法，其中是选定的（非常可能）小于的充分小的数，那么我们就（非常可能）得到。

#### 2. 下降方法

本章描述的算法将产生一个优化点列，其中



并且（除非已经是最优点）。表示的一个向量，被称为**步径**或**搜索方向**（尽管它不需要具有单位范数），表示迭代次数。标量被称为第次迭代的**步进**或**步长**。

我们讨论的所有方法都是**下降方法**，只要不是最优点就有



由凸函数的一阶条件（），意味着，因此一个下降方法中的搜索方向必须满足



即它和负梯度方向的夹角必须是锐角。我们称这样的方向为**下降方向**。

下降方法的一般框架如下。

**给定** 初始点。

**重复进行**

1. 确定下降方向。
2. 直线搜索。选择步长。
3. 更新。

**直到** 满足停止准则。

适用的下降方法均有相同的结构，但组织方式可能不同。

* 步长的确定方法

精确直线搜索

步长通过沿着射线优化而确定：



当求解上式的成本同计算搜索方向的成本相比比较低时，适合进行精确直线搜索。

回溯直线搜索

实践中主要采用非精确直线搜索方法。目前已经提出了很多非精确直线搜索方法，其中回溯方法既简单又相当有效。回溯直线搜索算法如下。

**给定** 在处的下降方向，参数，。

。

**如果** ，令 

参数的正常取值在0.01和0.3之间，的正常取值在0.1和0.8之间。

#### 3. 梯度下降方法

用负梯度做搜索方向，即令是一种自然地选择。相应的方法被称为**梯度方法**或**梯度下降方法**。下面给出梯度下降方法。

**给定** 初始点。

**重复进行**

1. 。
2. 直线搜索。通过精确或回溯直线搜索方法确定步长。
3. 更新。

**直到** 满足停止准则。

* 收敛性分析

为书写方便，用代替，其中。我们假定是上的强凸函数，因此存在正数和使得对所有成立。 定义为。将代入可得的二次型上界：



采用精确直线搜索的分析

假定采用精确直线搜索方法，对上式右侧取极小，其最小解为，最小值为。左边极小值为，其中是使最小的步长。因此



上式两边同时减去，并考虑到，可以断定



重复应用以上不等式，可以看出



其中。该式表明，误差将至少像几何数列那样快的收敛于0。按照迭代数值方法的术语，这种情况被称为**线性收敛**，因为误差位于误差和迭代次数的对数线性坐标图中一根直线的下方。

特别地，至多经过



次迭代，一定可以得到。对于较大的条件数上界，我们有



因此所需迭代次数的上界将随着条件数增大而近似线性的增长。

采用回溯直线搜索的分析

首先，只要就能满足回溯停止条件



因此回溯直线搜索将终止于或者，这位目标函数的减少提供了一个下界。将他们结合在一起，任何情况下总是成立



现在可以完全类似精确直线搜索的情况进行分析。最后可得



其中。

#### 4.最速下降方法

对在处进行一阶Taylor展开



令为上的任意范数。我们定义一个**规范化的最速下降方向**如下



一个规范化的最速下降方向是一个能使的线性近似下降最多的具有单位范数的步径。

我们也可以将规范化的最速下降方向乘以一个特殊的比例因子，考虑下述非规范化的最速下降方向：



其中表示对偶范数。对于这种最速下降步径，我们有



**最速下降方法**使用最速下降方向作为直线搜索方向。下面给出算法。

**给定** 初始点。

**重复进行**

1. 计算最速下降方向。
2. 直线搜索。通过精确或回溯直线搜索方法确定步长。
3. 更新。

**直到** 满足停止准则。

如果采用精确直线搜索方法，下降方向的比例因子不起作用，因此规范化或非规范化的方向都能用。

* **采用Euclid范数和二次范数的最速下降方法**

采用Euclid范数的最速下降方法

如果我们将取为Euclid范数，可以看出最速下降方向就是负梯度方向，即。因此，采用Euclid范数的最速下降方法就是梯度下降方法。

采用二次范数的最速下降方法

我们考虑二次范数



其中。规范化的最速下降方向由下式给出



其对偶范数为，非规范化的最速下降方向为



* 收敛性分析

对于任意范数的最速下降方法，都有



的形式，因此也具有线性收敛性。

* 讨论和例子

最速下降方法的范数选择

我们知道，梯度方法在下水平集（或最优点附近的Hessian矩阵）条件数适中时能很好地工作，但条件数很大时效果很差。由此可知，若进行坐标变换后条件数适中，最速下降方法将取得良好的效果。

例如，如果知道最优点处Hessian矩阵的近似矩阵，令是对的一种很好的选择，因为在最优点的Hessian矩阵将变成



因此很可能具有小的条件数。

#### 5.Newton方法

* **Newton步径**

对于，我们称向量



为（在处的）**Newton步径**。由的正定性可知，除非，否则就有



因此Newton步径是下降方向。我们可以用不同的方式解释和导出Newton步径。

二阶近似的最优解

函数在处的二阶Taylor近似（或模型）为



对于该二次凸函数，在处达到最小值。因此，将加上Newton步径能够极小化在处的二阶近似。

既然是二次可微的，当靠近时的二次模型应该非常准确。由此可以直观地认识到，当靠近时，点应该是很好的估计值。

Hessian范数下的最速下降方向

Newton步径也是处采用Hessian矩阵定义的二次范数，即



导出的最速下降方法。当靠近时我们有，矩阵的条件数很小，这就解释了为什么Newton步径是搜索方向的很好选择。

线性化最优性条件的解

如果我们在附近对最优性条件进行线性化，可以得到



这是的线性方程，其解为，因此在加上Newton步径能够满足线性化的最优化条件。这再一次表明当靠近时，点是很好的估计值。

Newton步径的仿射不变性

Newton步径的一个重要特点是对坐标的线性（或仿射）变换的独立性。假定是非奇异的，定义，我们有

，

其中。因此，在处的Newton步径是



其中是在处的Newton步径。于是，和的Newton步径由相同的线性变换相联系，并且



Newton减量

我们将



称为处的**Newton减量**。Newton减量在Newton方法的分析中有重要作用，并且也可用于设计停止准则。我们可以将Newton减量和按下式联系在一起：



其中是在处的二阶近似。因此，是基于在处的二阶近似对做出的估计值。

我们也可以将Newton减量表示为



该式表明是Newton步径的二次范数，该范数有Hessian矩阵定义。

Newton减量也出现在回溯直线搜索中，因为我们有



最后需要指出，同Newton步径一样，Newton减量也是仿射不变的。

* **Newton方法**

下面描述的Newton方法有时被称为**阻尼Newton方法**或者**谨慎Newton方法**，以区别于步长的纯Newton方法。

**给定** 初始点，误差阈值。

**重复进行**

1. 计算Newton步径和减量。

，

1. 停止准则。如果，退出。
2. 直线搜索。通过精确或回溯直线搜索方法确定步长。
3. 更新。

* 收敛性分析

主要结论如下。存在满足和的和使下面两式成立。

* 1. 如果，则



* 1. ，则回溯直线搜索产生，并且



重复应用第二个不等式，可以发现对任意的



于是



上述最后一个不等式表明，一旦第二个条件满足，收敛将会极其迅速。该现象被称为**二次收敛**。粗糙地说，在足够多次的迭代以后，每次迭代都能使正确数字的位数翻番。

对于收敛的两个阶段，我们将第一个阶段称为**阻尼Newton阶段**，第二个阶段称为**二次收敛阶段**或是**纯Newton阶段**。阻尼Newton阶段每次迭代至少减少，因此阻尼Newton阶段的迭代次数不可能超过次。而二次收敛阶段迭代次数不会超过次，其中。因此，总迭代次数不会超过上界



从实用的角度出发可以将视为常数，比如5或6。