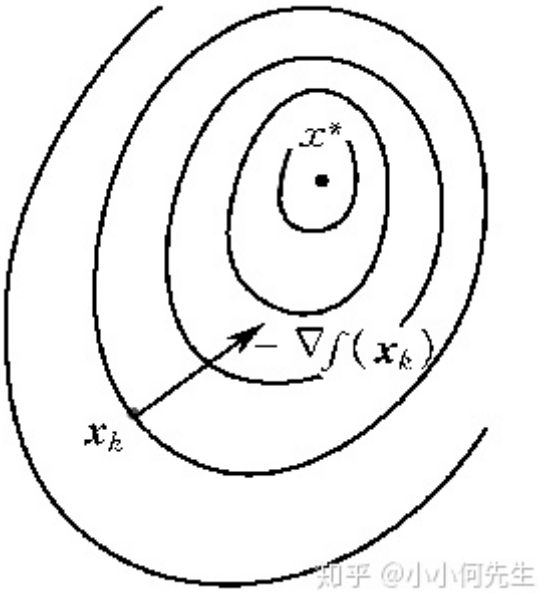
最速下降法利用目标函数一阶梯度进行下降求解，易产生锯齿现象，在快接近最小值时收敛速度慢。**Newton 法**利用了二阶梯度，收敛速度快，但是目标函数的`Hesse`矩阵不一定正定。于是出现了修正的`Newton`法，主要是对不同情况进行了分情况讨论。Newton 法的优缺点都很突出。优点：高收敛速度（二阶收敛）；缺点：对初始点、目标函数要求高，计算量、存储量大（需要计算、存储 Hesse 矩阵及其逆）。**拟 Newton** 法是模拟 Newton 法给出的一个保优去劣的算法。**共轭梯度法**是介于最速下降法和牛顿法之间的一个方法，相比最速下降法收敛速度快，并且不需要像牛顿法一样计算 Hesse 矩阵，只需计算一阶导数 (共轭梯度法是共轭方向法的一种，意思是搜索方向都互相共轭)。下文是详细解析：

最速下降法

最速下降法是最早的求解多元函数极值的数值方法。它直观、简单。它的缺点是，收敛速度较慢、实用性差。在点x_k处，沿什么方向寻找下一个迭代点呢？显然应该沿下降方向。一个非常直观的想法就是沿最速下降方向，即负梯度方向：$p_k = -\nabla f(x_k)$。



沿p_k方向进行直线搜索，由此确定下一个点的位置$x_{k+1} = x_k - t_k \nabla f(x_k)$，我们将$t_k$称为步长因子，满足以下等式：

$$f\left(x_k - t_k \nabla f\left(x_k\right)\right) = \min_t f\left(x_k - t \nabla f\left(x_k\right)\right)$$

简单合记为：

$$x_{k+1} = ls\left(x_k, -\nabla f\left(x_k\right)\right)$$

为了书写方便，以后记$g_k = g(x_k) = \nabla f(x_k)$。

到这里，我们已经大概知道最速下降法是怎么工作的，那这个步长因子t_k到底怎么求呢？，我们考虑特殊情况，假设我们的目标函数是正定二次函数：

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

目标函数对x的一阶梯度:

$$g(x) = Qx + b$$

这里引入一个定理，之后的求解就是依据这个定理的等式进行求解：

定理：设目标函数$f(x)$具有一阶连续偏导数，若$z = ls(x, p)$，则$\nabla f(z)^T p = 0$。

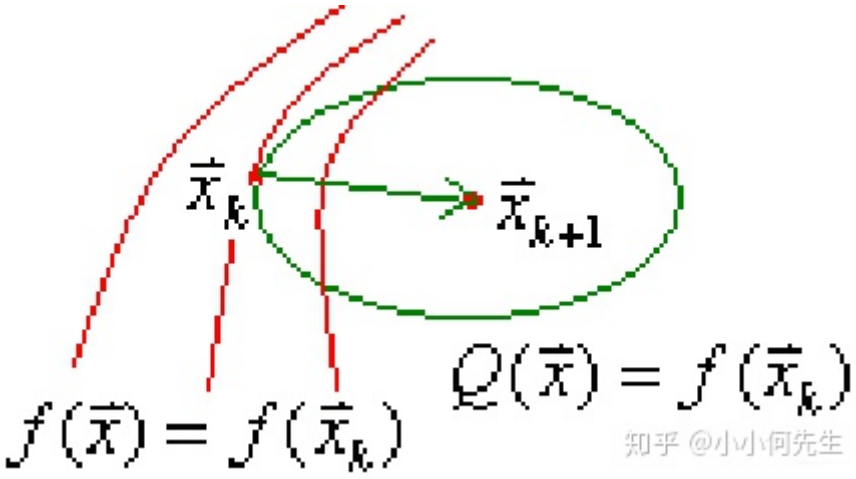
依据定理，我们可以得到$g_{k+1} \cdot g_k = 0$。由此有：

$$\begin{aligned} g_{k+1} \cdot g_k &= [Q(x_k - t_k g_k) + b]^T g(k) = 0 \\ &= [Qx_k + b - t_k Q g_k]^T g(k) = 0 \\ &= [g_k - t_k Q g_k]^T g(k) = 0 \end{aligned}$$

由此，可求解出t_k：

$$t_k = \frac{g_k^T g_k}{g_k^T Q g_k}$$

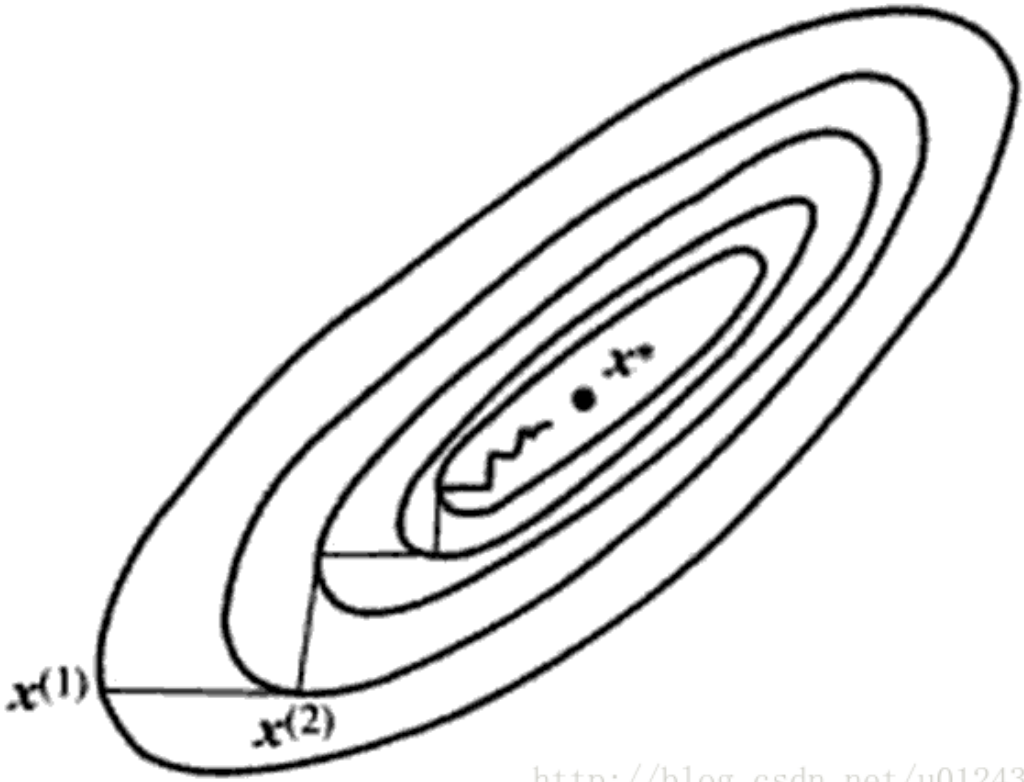
最速下降法的迭代点在向极小点靠近的过程中，走的是曲折的路线：后一次搜索方向p_{k+1}与前一次搜索方向p_k总是相互垂直的，称它为**锯齿现象**。除极特殊的目标函数（如等值面为球面的函数）和极特殊的初始点外，这种现象一般都要发生。



由于锯齿现象，在接近极小点的地方，每次迭代进行的距离变得越来越小，因而收敛速度不快，这正是最速下降法的缺点所在。

锯齿现象

在最速下降法中，迭代点向极小点靠近的过程中，走的是曲折的路线：后一次的搜索方向 $pk+1$与前一次的搜索方向 pk总是互相垂直的，称之为锯齿现象。如下图所示（在下图中一个圈表示等值线）。



在远离极小点的地方，最速下降法每次有较多的下降；但是在接近极小点的地方，由于锯齿现象的存在，使得每次迭代行进的距离缩短，收敛速度降低。造成锯齿现象的关键原因是$p_{k+1}p_k = 0$(或是$g_{k+1}g_k = 0$)，下面来解释一下具体原因。造成这种结果的具体原因是在点x_k沿p_k方向搜索的过程中，我们找到了该方向上的极小点，即x_{k+1}，也就是是说在x_{k+1}处p_k方向为该点的切线方向（如果不是切线方向，那么该点就不是极小点，还能找到更小的值），而在x_{k+1}处的搜索方向为负梯度方向，故$p_{k+1}p_k = 0$

> 证明：
> 最速下降法的迭代公式：$xk+1 = xk - \lambda\Delta f(x)$
> 其中，迭代步长λ 由一维搜索得到，λ 满足$\lambda = argmin\lambda f(xk - \lambda\Delta f(x))$
> 一阶导数为 0 时成立，即
> $\partial f(xk - \lambda\Delta f(x))\partial\lambda = -\Delta f(xk - \lambda\Delta f(x))\Delta f(x) = -\Delta f(xk+1)\Delta f(x) = 0$
> 最后一项中，前者就是下一步的迭代方向，后者是此时的迭代方向，因此，二者相互垂直，得证。

关于最速下降法的锯齿现象，文中的每一条等值线代表f(x)=const,其中x是向量形式的自变量，且x不唯一，等值线上的自变量都满足该等值线方程，其中在该等值线上取到最值的那个点的坐标是(x_k, C)，x_k必然属于x．也就是我们在p_k方向上取到了该方向上的极值点(x_k,C)，因此我们是沿着该方向（切线）来到的(x_k,C)极值点的，既然在这个方向取到了极值点，说明这一步梯度下降，在这个方向上只有这一个极值点．因此若要以(x_k,C)这个极值点作为新的梯度下降初始点，则梯度下降方向一定垂直于p_k这个梯度方向． 由图可见，在梯度下降初始点，因为迭代起始点距离局部最小点较远，因此迭代梯度下降收敛到局部最小点的速度还是较快的，但是随着迭代点越来越逼近最值点，由于锯齿效应的＂绕路＂行为，使得迭代点收敛于最值点所需要的时间比相同距离在边缘处收敛所用的时间加长，也就是收敛速度减慢．

Newton 法

由于最速下降法速度慢，Newton 引入二阶梯度，通过求其 Hesse 矩阵，一步到位直接求到极小点$\boldsymbol{x^*}$。

如果目标函数$\boldsymbol{f(x)}$在$\boldsymbol{R^n}$上具有连续的二阶偏导数，其 Hesse 矩阵$\boldsymbol{G(x) = \nabla^2 f(x)}$正定，那么就可以用 Newton 法对其进行求解了。原理如下：

考虑从x_k到x_{k+1}的迭代过程。在点x_k处，对$f(x)$按 Taylor 级数展开到第三项，即:

$$f(x) \approx Q(x) = f(x_k) + g(x_k)^T (x - x_k) + \frac{1}{2}(x - x_k)^T G(x_k)(x - x_k)$$

又因为 Hesse 矩阵$G(x)$正定，所以$Q(x)$是x的正定二次函数。令$Q(x)$其一阶导数等于 0，求出来的点就是极小点。

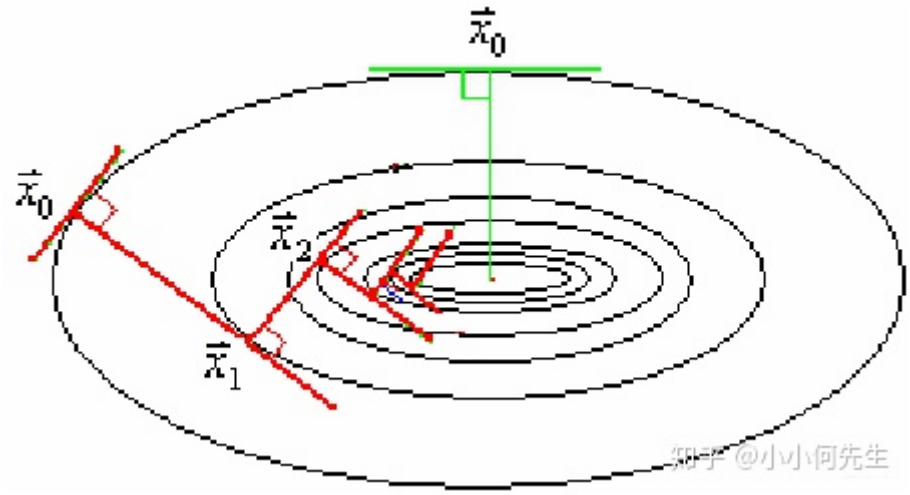
$$\nabla Q(x) = G(x_k)(x - x_k) + g(x_k) = 0$$

解出:

$$x_{k+1} = x_k - G(x_k)^{-1} g(x_k)$$

x_{k+1}是$f(x)$极小点x的新的近似点。上式称为 Newton 迭代公式，由该公式产生的算法称为**Newton 法**。当目标函数$f(x)$是正定二次函数时，有$f(x)$恒等于$Q(x)$。这说明：对于正定二次函数，Newton 法一次迭代就会得到最优解。

现在从几何上我们来直观理解一下，我们要求目标函数$f(x)$的极小值，函数$f(x)$过点x_k的等值面方程$f(x) = f(x_k)$, 在点x_k处，用一个与该曲面最密切的二次曲面来代替它 (Taylor 展开)，这个二次曲面的方程即是$Q(x) = f(x)$。当$G(x)$正定时，它是一个超椭球面，$Q(x)$的极小点x_{k+1}正是这个超椭球面的中心。我们就用x_{k+1}作为$f(x)$极小点x^* 的近似点。如下图所示:



对于具有正定 Hesse 矩阵的一般目标函数，由于在极小点附近，它近似地呈现为正定二次函数，所以可以想见，**Newton 法在最优点附近应该具有较高的收敛速度。**

修正 Newton 法

Newton 法的优点是收敛速度快、程序简单。但是**对于表达式很复杂的目标函数，由于其 Hesse 矩阵很难或不可能求出，这时显然不宜使用 Newton 法。**下面介绍修正 Newton 法:

以下讨论仅假定 Hesse 矩阵可以求到。

1. 在迭代点x_k处 Hesse 矩阵G_k变为**不可逆**，由线性方程组$G_k p_k = -g_k$无法解出搜索方向p_k。遇有此种情况，改取$p_k = -g_k$，然后作直线搜索:

   $$x_{k+1} = ls(x_k, p_k)$$

   即用最速下降法的迭代公式代替 Newton 法的迭代公式，从而完成这一次迭代。

2. 在迭代点x_k处 Hesse 矩阵G_k**非奇异**，即G_k^{-1}存在这时可解出$p_k = -G_k^{-1}g_k$ (称为 Newton 方向)。按 Newton 迭代公式，有:

$$x_{k+1} = x_k + p_k$$

上式可以理解为从点x_k出发沿p_k方向进行直线搜索，步长因子取为 1。上面这个公式是 Newton 法中假设目标函数为二次正定而推到出来的，但是现在这个目标函数并没有这一项约束，所以目标函数可能很复杂，因而不能总保证p_k的方向是下降方向，有时即使是下降方向，也会由于步长因子不加选择地取为 1，而不能保证$f(x_{k+1}) < f(x_k)$。对此又分情况进行处理：

a. 若$f(x_{k+1}) < f(x)$, 那函数是朝着下降方向去的，则该迭代有效。

b. 若$f(x_{k+1}) \leq f(x)$，则表明函数不是朝着下降方向去的，这里又分了两种情况进行了讨论

第一：当$\left|g_k^T p_k\right| \leq \varepsilon \left\|g_k\right\| \left\|p_k\right\|$，($\varepsilon$是某一很小的正数) 时 (说明$p_k$与$-g_k$几乎垂直，也就是说一阶梯度和假设目标函数为二次正定而求出的梯度方向完全不一致，也就是说明目标函数假设为二次正定函数错误，应该取一阶梯度方向)，故 Newton 方向p_k是不利方向。这时，改取$p_k = -g_k$，然后新进行直线搜索。

第二：$g_k^T p_k < \varepsilon \left\|g_k\right\| \left\|p_k\right\|$时说明 Newton 方向$p_k = -G_k^{-1}g_k$是下降方向 (一阶梯度和假设目标函数为二次正定而求出的梯度方向之间的夹角是小于 90 度的，大体方向一致)，这时重新进行直线搜索。否则，有$g_k^T p_k > \varepsilon \left\|g_k\right\| \left\|p_k\right\|$时说明 Newton 方向$p_k = -G_k^{-1}g_k$是上升方向 (一阶梯度和假设目标函数为二次正定而求出的梯度方向之间的夹角是大于 90 度的，大体方向相反) 改取 Newton 方向的反方向$p_k = G_k^{-1}g_k$为搜索方向，然后重新进行直线搜索。

拟 Newton 法

拟 Newton 法是效果很好的一大类方法。它当中的 DFP 算法和 BFGS 算法是直到目前为止在不用 Hesse 矩阵的方法中的最好方法。

基本思想

考虑 Newton 迭代公式

$$x_{k+1} = x_k - G_k^{-1}g_k, \quad k = 0, 1, \cdots$$

这里搜索方向为$p_k = -G_k^{-1}g_k$，步长因子为 1。

我们从以下两点考虑对 Newton 迭代公式的改进：

(1) 为避免求逆矩阵，设想用某种近似矩阵$H_k = H(x_k)$替换$-G_k^{-1}$，上式则变为

$$x_{k+1} = x_k - H_k g_k, \quad k = 0, 1, \cdots$$

这时搜索方向为$p_k = -H_k g_k$，步长因子仍为 1。

(2) 为了取得更大的灵活性，考虑更一般的公式

$$x_{k+1} = x_k - t_k H_k g_k, \quad k = 0, 1, \cdots$$

这时搜索方向仍为$p_k = -H_k g_k$，但步长因子取为最优步长因子。上式是代表面很广的一类迭代公式。例如，当$H_k = E$时，它是最速下降法公式。当$H_k = G_k^{-1}$时，它是阻尼 Newton 法公式。

这样的H_k存在吗？又如何来求呢？假如存在，那么为使H_k确实近似G_k^{-1}并易于计算，我们要对H_k人为地附加某些条件。主要是三点：

第一，为保证搜索方向$p_k = -H_k g_k$是下降方向，如果$\nabla f(x_0)^T p < 0$，则p方向是函数$f(x)$在点x_0处的下降方向得到：

$$p_k^T g_k < 0 \Rightarrow -g_k^T H_k g_k < 0 \Rightarrow g_k^T H_k g_k > 0$$

要求每一个H_k都是对称正定矩阵，可以保证该式成立。

第二，为易于计算，要求H_k到H_{k+1}之间具有简单的迭代形式。最简单的迭代关系为

$$H_{k+1} = H_k + E_k$$

E_k称为**校正矩阵**，上式称为**校正公式。**

第三，为使H_k确实近似G_k^{-1}要求每一个H_k必须满足拟 Newton 条件。那所谓的拟 Newton 条件是啥呢？

我们假设目标函数$f(x)$具有连续的二阶偏导数，将$f(x)$在点x_{k+1}处作 Taylor 级数展开：

$$f(x) \approx f\left(x_{k+1}\right) + \nabla f(x_{k+1})^T \left(x - x_{k+1}\right) + \frac{1}{2}(x - x_{k+1})^T G_{k+1} \left(x - x_{k+1}\right)$$

对上式两端求梯度有：

$$\nabla f(x) \approx g_{k+1} + G_{k+1} \left(x - x_{k+1}\right)$$

令$x = x_k$，则有：

$$g_k \approx g_{k+1} + G_{k+1} \left(x_k - x_{k+1}\right)$$

所以，当G_{k+1}正定时，有

$$x_{k+1} - x_k \approx G_{k+1}^{-1} \left(g_{k+1} - g_k\right)$$

对于正定二次函数, 上式近似将变为等式。

因此，如果迫使H_{k+1}满足类似于上式的等式

$$x_{k+1} - x_k = H_{k+1} \left(g_{k+1} - g_k\right)$$

那么H_{k+1}就有可能很好地近似于G_{k+1}^{-1}上式称为**拟 Newton 条件**或**拟 Newton 方程。**

记：

$$s_k = x_{k+1} - x_k$$
$$y_k = g_{k+1} - g_k$$

那么拟 Newton 条件可简记为：

$$H_{k+1} y_k = S_k$$

带入迭代关系式$H_{k+1} = H_k + E_k$有：

$$E_k y_k = S_k - H_k y_k$$

算法中的校正矩阵E_k可由确定的公式来计算，不同的公式对应不同的拟 Newton 算法。满足条件的E_k有无穷多个，因此上述的拟 Newton 算法是一簇算法。

DFP 算法

DFP 法是首先由 Davidon(1959 年) 提出，后由 Fletcher 和 Powell（1963 年）改进的算法。它是无约束优化方法中最有效的方法之一。DFP 法虽说比共轭梯度法有效，但它对直线搜索有很高的精度要求。

考虑如下校正公式

$$H_{k+1} = H_k + \alpha_k u_k u_k^T + \beta_k v_k v_k^T$$

其中u_k，v_k是待定列向量，α_k，β_k是待定常数。校正矩阵

$$E_k = \alpha_k u_k u_k^T + \beta_k v_k v_k^T$$

根据拟 Newton 条件，有

$$\alpha_k u_k u_k^T y_k + \beta_k v_k v_k^T y_k = s_k - H_k y_k$$

上式u_k和v_k有很多种取法

$$\alpha_k u_k u_k^T y_k = s_k$$
$$\beta_k v_k v_k^T y_k = -H_k y_k$$

如果取$u_k = s_k$，$v_k = H_k y_k$那么有

$$\alpha_k = \frac{1}{s_k^T y_k},\ \beta_k = -\frac{1}{y_k^T H_k y_k}$$

$$H_{k+1} = H_k + \frac{s_k s_k^T}{s_k^T y_k} - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k}$$

DFP 算法的性质

定理 1：在 DFP 算法中，若初始矩阵H_0对称正定，则H_k中每一个都对称正定。

定理 2：设将 DFP 算法施用于具有对称正定矩阵Q的二次函数$f(x)=\frac{1}{2}x^TQx+b^Tx+c$，如果:

(i) 初始矩阵H_0对称正定；

(ii) 迭代点互异，产生的搜索方向向量依次为$p_0,p_1,\quad\cdots,\quad P_k\ (k\leq n-1)$，则有

$$p_i^TQp_j=0,\ i,j=0,1,\cdots,\quad k(i\neq j)$$
$$H_{k+1}Qp_j=p_j,\ j=0,1,\cdots,\quad k.$$

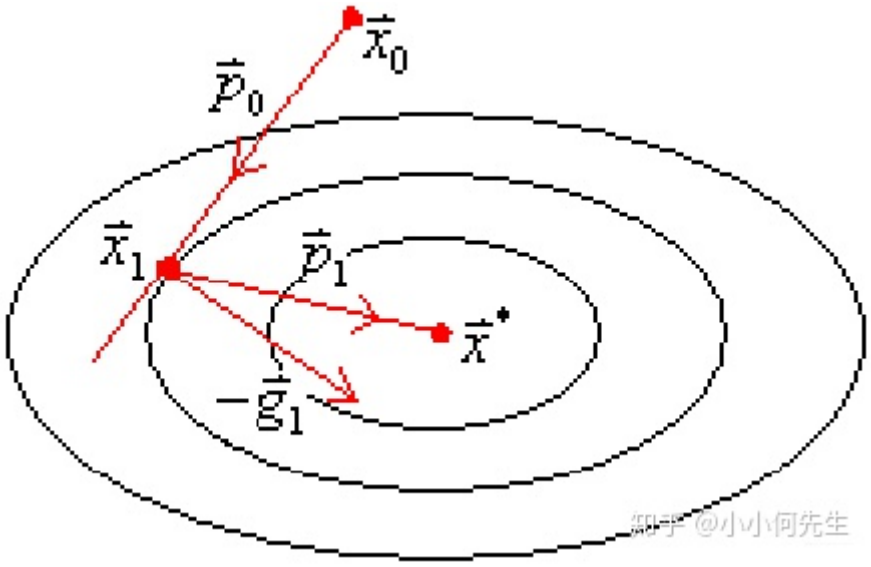
定理 3：若定理 2 的条件都满足，并且经过n次迭代才求到极小点，则$H_n=Q^{-1}$。

共轭方向法与共轭梯度法

基本思想

最速下降法存在锯齿现象，Newton 法需要计算目标函数的二阶导数。接下来介绍的**共轭方向法**是介于最速下降法和 Newton 法之间的一种方法，它克服了最速下降法的锯齿现象，从而提高了收敛速度；它的迭代公式也比较简单，不必计算目标函数的二阶导数，与 Newton 法相比，减少了计算量和存储量。它是比较实用而有效的最优化方法。

我们先将其在正定二次函数$f(x)=\frac{1}{2}x^TQx+b^Tx+c$上研究，然后再把算法用到更一般的目标函数上。首先考虑二维的情形。



任选初始点x_0，沿它的某个下降方向，例如向量p_0的方向，作直线搜索，如上图所示。由下面这个定理：

定理：设目标函数$f(x)$具有一阶连续偏导数，若$z=ls(x,p)$，则$\nabla f(z)^Tp=0$。

知$\nabla f(x_1)^Tp_0=0$。如果按照最速下降法选择的就是负梯度方向为搜索方向 (也就是$-g_1$方向)，那么将要发生锯齿现象。于是一个设想是，干脆选择下一个迭代的搜索方向p_1就从x_1直指极小点x^*，也就是找到上图所示的p_1方向。

因为p_1从x_1直指极小点x^*，所以x^*可以表示为：

$$x^* = x_1 + t_1 p_1$$

其中t_1是最优步长因子。显然，当$x^* \neq x_1$时，$t_1 \neq 0$。到这里，我们还有一个已知条件没用，就是目标函数为二次正定，所以我们对目标函数求导，得到：

$$\nabla f(x) = Qx + b$$

因为x^*是极小点，所以有:

$$\nabla f(x^*) = Qx^* + b = 0$$

将$x^* = x_1 + t_1 p_1$带入上述方程式，有：

$$\nabla f(x_1) + t_1 Q p_1 = 0$$

上式两边同时左乘p_0^T，并注意到$p_0^T \nabla f(x_1) = 0$和$t \neq 0$，得到$p_0^T Q p_1 = 0$。这就是为使p_1直指极小点x^*，p_1所必须满足的条件。并且我们将两个向量p_0和p_1称为Q**共轭向量**或称p_0和p_1是Q**共轭方向**。

由上面共轭梯度法那张图可以设：

$$p_1 = -\nabla f(x_1) + \alpha_0 p_0$$

上式两边同时左乘$p_0^T Q$，得：

$$-p_0^T Q \nabla f(x_1) + \alpha_0 p_0^T Q p_0 = 0$$

由此解出：

$$\alpha_0 = \frac{p_0^T Q \nabla f(x_1)}{p_0^T Q p_0}$$

代回$p_1 = -\nabla f(x_1) + \alpha_0 p_0$得：

$$p_1 = -\nabla f(x_1) + \frac{p_0^T Q \nabla f(x_1)}{p_0^T Q p_0} p_0$$

从而求到了p_1的方向。

归纳一下，对于正定二元二次函数，从任意初始点x_0出发，沿任意下降方向p_0做直线搜索得到x_1再从x_1出发，沿p_0的共轭方向p_1作直线搜索，所得到的x_2必是极小点x^*。到目前为止的共轭梯度法依旧是假设了目标函数是二次正定矩阵。

上面的结果可以推广到n维空间中，即在n维空间中，可以找出n个互相共轭的方向，对于n元正定二次函数从任意初始点出发，顺次沿着这n个共轭方向最多作n次直线搜索，就可以求到目标函数的极小点。

对于n元正定二次目标函数，如果从任意初始点出发经过**有限次迭代**就能够求到极小点，那么称这种算法具有**二次终止性**。例如，Newton 法对于二次函数只须经过一次迭代就可以求到极小点，因此是二次终止的；而最速下降法就不具有二次终止性。共轭方向法（如共轭梯度法、拟 Newton 法等）也是二次终止的。

一般说来，具有二次终止性的算法，在用于一般函数时，收敛速度是较快的。

共轭向量及其性质

定义：设Q是$n \times n$对称正定矩阵。若n维向量空间中的非零向量$p_0, p_1, \cdots, \ p_{m-1}$满足$p_i^T Q p_j = 0, \ i, j = 0, 1, \cdots, \ m-1 (i \neq j)$则称$p_0, p_1, \ \cdots, \ p_{m-1}$是$Q$共轭向量或称向量$p_0, p_1, \cdots, \ p_{m-1}$是$Q$共轭的（简称共轭）。

当$Q = E$ (单位矩阵) 时$p_i^T Q p_j = 0$变为$p_i^T p_j = 0$，
$i, j = 0, 1, \cdots, \ m-1 (i \neq j)$。即向量$i, j = 0, 1, \cdots, \ m-1 (i \neq j)$**互相正交**。由此看到，“正交”是 “共轭” 的一种特殊情形，或说，“共轭”是 “正交” 的推广。

下面介绍几个定理：

定理：若非零向量$p_0, p_1, \ \cdots, \ p_{m-1}$是$Q$共轭的，则线性无关。

推论：在n维向量空间中，R^n非零的共轭向量的个数不超过n。

定义 设$p_0, p_1, \ \cdots, \ p_{m-1}$是$R$中的线性无关向量，$x_0 \in R$。那么形式为：

$$z = x_0 + \sum_{i=0}^{m-1} \alpha_i p_i, \forall \alpha_1, \alpha_2, \cdots, \alpha_{m-1} \in R$$

的向量构成的集合，记为$L \left[x_0; p_0, p_1, \cdots, \ p_{m-1} \right]$。称为由点$x_0$和向量$p_0, p_1, \ \cdots, \ p_{m-1}$所生成的**线性流形**。

共轭方向法

共轭方向法的理论基础是下面的定理。

定理 假设

(1) Q 为$n \times n$对称正定矩阵;

(2) 非零向量$p_0, p_1, \ \cdots, \ p_{m-1}$是$Q$共轭向量;

(3) 对二次目标函数$f(x) = \frac{1}{2} x^T Q x + b^T x + c$顺次进行$m$次直线搜索:

$$x_{i_1} = ls(x_i, p_i), i = 0, 1, \ \cdots, \ m-1$$

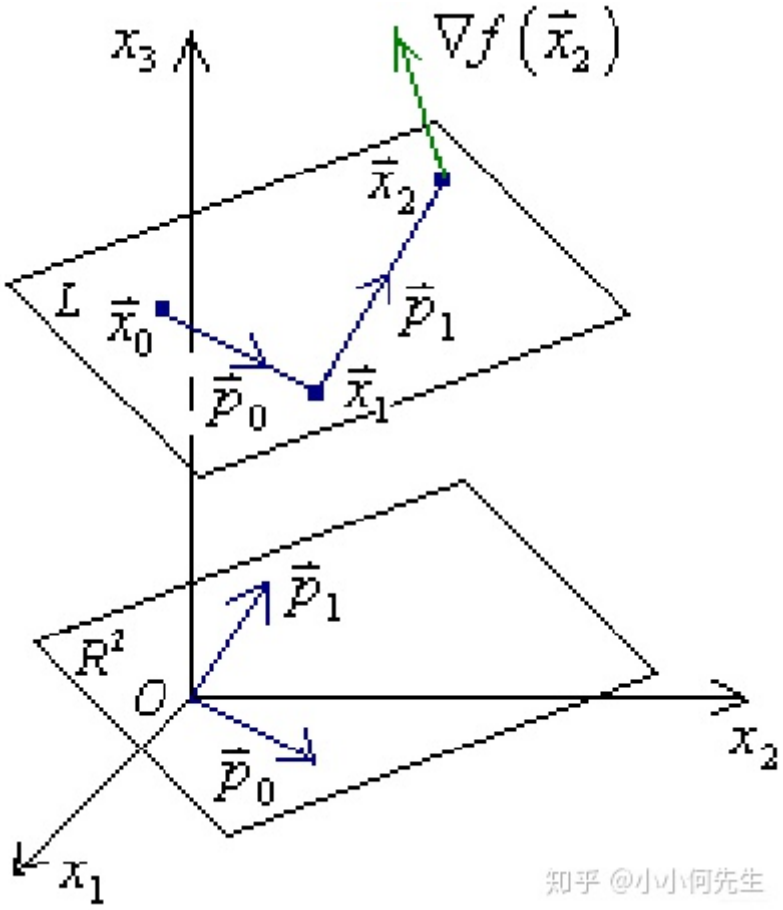
其中$x_0 \in R$是任意选定的初始点，则有：

i) $p_j^T \nabla f(x_m) = 0, \ 0 \leqslant j \leqslant m$;

ii) x_m是二次函数$f(x)=\frac{1}{2}x^TQx+b^Tx+c$在线性流形

$L\left[x_0;p_0,p_1,\cdots,\ p_{m-1}\right]$上的极小点。

这个定理看来较繁，但可借用直观的几何图形来帮助理解。$n=3$，$m=2$的情形为例，如图示。



p_0和p_1是 Q 共轭向量，张成了二维空间R^2，这是过坐标原点的一个平面。 现在，过点x_0沿p_0方向作直线搜索得到x_1，再过点x_1沿p_1方向作直线搜索得到x_2过点x_0由向量p_0和p_1张成的平面就是线性流形$L\left[x_0;p_0,p_1\right]$。它是$R^2$的平行平面。

定理的论断是，最后一个迭代点x_2处的梯度$\nabla f(x_2)$必与p_0和p_1垂直。并且x_2是三元二次目标函数$f(x)$在线性流形$L\left[x_0;p_0,p_1\right]$(即过$x_0$由$p_0$和$p_1$张成的平面) 上的极小点。

共轭方向法算法的大体流程就是：选定初始点x_0和下降方向向量p_0，做直线搜索$x_{k+1}=ls(x_k,p_k)$。提供的梯度方向p_{k+1}使得$p_j^TQp_{k+1}=0$，$j=0,1,\cdots,\ k$。提供共轭方向的方法有多种。不同的提供方法将对应不同的共轭方法。每种方法也因产生共轭方向的特点而得名。

那么这里做直线搜索$x_{k+1}=x_k+tp_k$中的t是如何确定的呢？这里我们先回顾一下在最速下降法中是如何计算这个t的。最速下降法：

依据定理 设目标函数$f(x)$具有一阶连续偏导数，若$z=ls(x,p)$，则$\nabla f(z)^Tp=0$。，我们可以得到$g_{k+1}\cdot g_k=0$。由此有：

$$\begin{aligned} g_{k+1} \cdot g_k &= [Q(x_k - t_k g_k) + b]^T g(k) = 0 \\ &= [Qx_k + b - t_k Q g_k]^T g(k) = 0 \\ &= [g_k - t_k Q g_k]^T g(k) = 0 \end{aligned}$$

由此，可求解出t_k：

$$t_k = \frac{g_k^T g_k}{g_k^T Q g_k}$$

这里还可以采用另外一种种方式计算t_k，下面对另外一种方式进行公式推导：

由$x_{k+1} = x_k + tp_k$，用Q左乘上式两边，然后再同时加上b，利用$\nabla f(x) = Qx + b$能够得到：

$$\nabla f(x_{k+1}) = \nabla f(x_k) + tQp_k$$

左乘p_k有

$$p_k^T \nabla f(x_k + tp_k) = p_k^T \nabla f(x_k) + tp_k^T Q p_k = 0$$

由此解出：

$$t = -\frac{p_k^T \nabla f(x_k)}{p_k^T Q p_k}$$

在最速下降法中$x_{k+1} = x_k - t_k g_k$，在共轭方向法中$x_{k+1} = x_k + t_k g_k$。

共轭梯度法

在共轭方向法中，如果初始共轭向量p_0恰好取为初始点x_0处的负梯度$-g_0$，而其余共轭向量p_k $(k = 1, 2, \cdots, \ n-1)$由第$k$个迭代点$x_k$处的负梯度$-g_k$与已经得到的共轭向量$p_{k-1}$的线性组合来确定，那么这个共轭方向法就称为**共轭梯度法**。

针对目标函数是正定二次函数来讨论：

(1) **第一个迭代点的获得**：

选定初始点x_0，设$x_0 \neq x^*$ (否则迭代终止)，因此$\nabla f(x_0) \neq 0$。（以下用g_k表示$\nabla f(x_k)$）从x_0出发沿p_0方向做直线搜索，得到第 1 个迭代点$x_1 = x_0 + t_0 p_0$，其中t_0可由下式确定：

$$t_0 = -\frac{p_0^T g_0}{p_0^T Q p_0} = \frac{g_0^T g_0}{p_0^T Q p_0}$$

显然$t_0 \neq 0$

(2) **第二个迭代点的获得**：

设$x_1 \neq x^*$，因此$g_1 \neq 0$。由$p_0^T g_1 = 0$知p_0与g_1线性无关。取$p_1 = -g_1 + \alpha_0 p_0$其中$\alpha_0$是使$p_1$与$p_0$共轭的待定系数，令：

$$p_1^T Q p_0 = -g_1^T Q p_0 + \alpha_0 p_0^T Q p_0 = 0$$

由此解出

$$\alpha_0 = \frac{g_1^T Q p_0}{p_0^T Q p_0}$$

并代回确定p_1，并获得第 2 个迭代点。

$$x_2 = x_1 + t_1 p_1$$

由公式$t = -\dfrac{p_k^T \nabla f(x_k)}{p_k^T Q p_k}$可以求得$t_1$，带入公式$p_1 = -g_1 + \alpha_0 p_0$可进一步优化得到：

$$t_1 = -\frac{p_1^T g_1}{p_1^T Q p_1} = \frac{g_1^T g_1}{p_1^T Q p_1} \neq 0$$

(3) **第三个迭代点的获得**：

设$x_2 \neq x^*$，因此$g_2 \neq 0$。由$p_1^T g_2 = 0$知p_1与g_2线性无关。取$p_2 = -g_2 + \alpha_1 p_1$
其中α_1是使p_2与p_1共轭的待定系数，令：

$$p_2^T Q p_1 = -g_2^T Q p_1 + \alpha_1 p_1^T Q p_1 = 0$$

由此解出

$$\alpha_1 = \frac{g_2^T Q p_1}{p_1^T Q p_1}$$

并代回确定p_2，并获得第 3 个迭代点。

$$x_3 = x_2 + t_2 p_2$$

其中

$$t_2 = -\frac{p_2^T g_2}{p_2^T Q p_2} = \frac{g_2^T g_2}{p_2^T Q p_2} \neq 0$$

上述过程仅表明p_0与p_1，p_1与p_2共轭，现在问，p_0与p_2也共轭吗？

$$\begin{aligned}
p_2^T Q p_0 &= \left(-g_2 + \alpha_1 p_1\right)^T Q p_0 \\
&= -g_2^T Q p_0 + \alpha_1 p_1^T Q p_0 \\
&= -g_2^T Q p_0 \ \left[\mathbb{L} | p_1^T Q p_0 = 0\right] \\
&= -g_2^T \left(g_1 - g_0\right) / t_0 \ \left(\mathrm{Hg}_{1+1} = g_i + t_i Q p_i, t_i \neq 0\right) \\
&= -\left(g_2^T g_1 - g_2^T g_0\right) / t_0
\end{aligned}$$

(4) **第k个迭代点的获得**：

由$p_{k-1}^T g_k = 0$知p_{k-1}与g_k线性无关。取$p_k = -g_k + \alpha_{k-1} p_{k-1}$其中$\alpha_{k-1}$是使$p_k$与$p_{k-1}$共轭的待定系数，令：

$$p_k^T Q p_{k-1} = -g_k^T Q p_{k-1} + \alpha_{k-1} p_{k-1}^T Q p_{k-1} = 0$$

由此解出

$$\alpha_{k-1} = \frac{g_k^T Q p_{k-1}}{p_{k-1}^T Q p_{k-1}}$$

并代回确定p_k，并获得第 k+1 个迭代点。

$$x_{k+1} = x_k + t_k p_k$$

其中

$$t_k = -\frac{p_k^T g_k}{p_k^T Q p_k} = \frac{g_k^T g_k}{p_k^T Q p_k} \neq 0$$

以上就是共轭梯度法得核心内容。

Fletcher-Reeves 共轭梯度法

为使共轭梯度算法也适用于非二次函数，需要消去算法中的Q对于正定二次函数，有$Q p_k = \frac{1}{t_k}(g_{k+1} - g_k)$代入到$\alpha_k$中，得：

$$\alpha_k = \frac{g_{k+1}^T Q p_k}{p_k^T Q p_k} = \frac{g_{k+1}^T (g_{k+1} - g_k)}{p_k^T (g_{k+1} - g_k)}$$

此式中已不再出现矩阵Q，将$p_k = -g_k + \alpha_{k-1} p_{k-1}$两端转置运算，并同时右乘$g_{k+1}$得：

$$p_k^T g_{k+1} = -g_k^T g_{k+1} + \alpha_{k-1} p_{k-1}^T g_{k+1}$$

将共轭方向法中的定理带入得到$p_{k-1}^T g_{k+1} = 0$，由直线搜索的性质有$p_k^T g_{k+1} = 0$，带入上式有$g_{k+1}^T g_k = 0$。此外：

$$p_k^T g_k = -g_k^T g_k + \alpha_{k-1} p_{k-1}^T g_k = -g_k^T g_k$$

带入α_k，得到：

$$\alpha_k = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2}$$

此式称为 Fletcher - Reeves 公式 (1964 年)。