金属物探资料数据处理参考资料

赠阅

国家计委地质局150工程方法组
武汉地质学院物探、数学教研室

1974年12月

目 录

一种最小平方估计非线性参数的算法

D. W. MARQUARDT

引　言　　最小平方估计非线性参数的许多算法都是围绕着两条途径之一。一方面，模型可展成台劳级数，并且在局部线性的假设下，每次迭代中，参数的校正值都可以算出。另一方面，则是应用各种改进后的最速下降法。这两种方法常々行到搁浅。台劳级数方法是由于逐步迭代发散所致，最速下降法（或斜量法）则是由于开头几次迭代后就发生烦恼的缓慢收敛。

本文开展一种最大邻域的方法，它实际上就是在台劳级数法与斜量法之间进行内插，这种内插是基于最大邻域，在此邻域中，截断台劳级数给出非线性模型的一个合适的表示。

所有的结果都可以推广到非线性代数方程组的求解问题上去。

问题的陈述　　设拟合数据的模型为

$$(1)\qquad E(y)=f(x_1,x_2,\cdots,x_m;\beta_1,\beta_2,\cdots,\beta_k)=f(\underline{x},\underline{\beta})$$

其中 $x_1, x_2, \cdots, x_m$ 是自变量；$\beta_1, \beta_2, \cdots, \beta_k$ 是 k 个参数的母体值，$E(y)$ 是因变量的期望值。

数据点记为

$$(2)\qquad (Y_i, X_{1i}, X_{2i}, \cdots, X_{mi}),\qquad i=1,2,\cdots,n.$$

于是问题就是要计算这些参数的估计值使得

$$(3)\qquad \Phi=\sum_{i=1}^{n}\left[Y_i-\hat{Y}_i\right]^2=\|Y-\hat{Y}\|^2$$

为最小，其中 $\hat{Y}_i$ 表示在第 i 数据点，由（1）预测的 y 值。

人所熟知，当 f 是 β 的线性函数时，Φ 的等值面是椭球面，当 f 是 β 的非线性函数时，等值面随着非线性的严重程度发生畸变。不过，就是非线性模型的情况，等值面在 Φ 的极小点邻近仍然近似于椭球面。典型的情况是：Φ 的等值面在某一方向缩得很短，在另一方向又拉得很长以致极小点就位于一个长曲槽的底部。

本文只专注于得出最小平方估计的算法。至于非线性估计的更广泛的统计性态及其在特定问题中的应用，读者可阅参考文献中〔1〕、〔2〕、〔4〕、〔5〕、〔6〕各项。其它参考资料也都在所引用的文章中。

通用的方法　下面是基于把函数 f 展成台劳级数的方法（通常指为高斯方法〔1〕,〔7〕或高斯——牛顿方法〔4〕）把台劳级数只写到一次项

$$(4)\quad \langle Y(\underline{Z}_i, \underline{b}+\underline{\delta}_t)\rangle = f(\underline{Z}_i, \underline{b}) + \sum_{j=1}^{k} \left(\frac{\partial f_i}{\partial b_j}\right)(\underline{\delta}_t)_j$$

或

$$(4a)\quad \langle \underline{Y}\rangle = \underline{f}_0 + \underline{P}\,\underline{\delta}_t .$$

在(4)中，记号 $\underline{b}$ 代替 $\underline{\beta}$，$\underline{b}$ 的收敛值就是 β 的最小平方估计值。向量 $\underline{\delta}_t$ 表示 $\underline{b}$ 的校正值，下标 t 用以表示它是由台劳级数方法计算出来的。括号〈　〉用以区别它是基于线性化后模型的预测，而不是基于实际非线性模型的预测。于是由(4)预测到 Φ 的值为

$$(5)\quad \langle \Phi \rangle = \sum_{i=1}^{n} \left[Y_i - \langle Y_i \rangle \right]^2$$

在(4)中，$\underline{\delta}_t$ 只线性地出现，所以可按标准的最小二乘法求

出，即对于一切 j，令 $\frac{\partial\langle\Phi\rangle}{\partial\delta_j}=0$，于是 $\underline{\delta_t}$ 可从下面方程组解出，

$$(6)\quad \underline{A}\,\underline{\delta_t}=\underline{g},$$

其中[注1]

$$(7)\quad \underline{A}^{[K\times K]}=\underline{P}^{T}\underline{P},$$

$$(8)\quad \underline{P}^{[n\times K]}=\left(\frac{\partial f_i}{\partial b_j}\right),\quad i=1,2,\cdots,n;\ j=1,2,\cdots,K$$

$$(9)\quad \underline{g}^{[K\times 1]}=\left(\sum_{i=1}^{n}(Y_i-f_i)\frac{\partial f_i}{\partial b_j}\right),\quad j=1,2,\cdots,K$$

$$=\underline{P}^{T}(\underline{Y}-\underline{f_0})$$

在实践中发现只用 $\underline{\delta_t}$ 的一小部分来校正 $\underline{b}$ 倒是很有帮助的，否则可能会外推到于能由（4）合适地表示的范围之外，以致迭代发散。因此一旦方向由 $\underline{\delta_t}$ 确定后，有各种方法〔1〕，〔4〕，〔6〕可以用来确定一个恰当的步长 $K\underline{\delta_t}$，$0<K\le 1$。但是，就是这样做，不收敛情形也是常有的。

与此对比，斜量法却是简单地从现有的试探值沿Φ的负梯度方向迈出步，所以

$$(10)\quad \underline{\delta_g}=-\left(\frac{\partial\Phi}{\partial b_1},\ \frac{\partial\Phi}{\partial b_2},\ \cdots\cdots,\ \frac{\partial\Phi}{\partial b_K}\right)^{T}$$

有各种改进的最速下降法〔5〕，是用来部分地弥补Φ的等值面特别坏，它使斜量法收敛得极其缓慢的情况。用这些改进的斜量法，也跟用台劳级数法一样，当校正向量的方向确定后，还

注1　上标T表示矩阵的转置

必须谨慎地控制步长。虽然这样做，缓慢的收敛性仍是常见，并非例外。

在各种关于 f 的性质的假设下，已推得台劳级数方法收敛性的多种证明〔4〕，〔7〕，也推得斜量法收敛性的各种证明〔3〕然而，在迭代过程中，假定 f 是很好的函数，并假定有无穷多恰当的算术运算过程，关于 Φ 的单调性的数学证明最好的也不过是一种关于收敛性的必要条件。这些证明，它们本身并不给出收敛速度的合适指示。我们必须从理论上说是收敛的一类方法中，找出这样的方法，它对于大多数问题，确实是只用有限次算术运算就能迅速收敛的。

问题的定性分析　　鉴于台劳级数法与斜量法都有不适用性，我们最好还是来检查一下有关的原理。首先，任何一个适用的方法应该使校正向量的方向与 Φ 的负梯度方向的夹角是在 $90°$ 之内。否则 Φ 沿着校正向量的各点处的值会变大而不是变小。其次，由于在许多问题中，Φ 曲面的严重拉长，δ_t 常常与 δ_g 成了几乎是 $90°$ 的角（事实上，我们已对各种各样问题调整过这个 γ 角，发现 γ 经常是在 $0<\gamma<90°$ 范围之内）。从这些来考虑，这样说似乎是合理的，即任何一种改进的方法从某种意义上说，就是在 δ_t 与 δ_g 之间进行内插。

上述这两种方法都暗含着这个意思，即在确定满意的步长之前，应选择好校正向量的方向。然而在将要描述的算法中方向与步长却是同时确定的。

算法的理论基础　　算法的理论基础包含在下面几个定理之中，定理 1 与定理 2 是 Morrison 的〔7〕。这里，对定理 1 的证明与 Morrison 的证明不同。这里的证明也许对所包含的

几何关系会提供较深的看法。

定理1　设 $\lambda \geq$ 是任意的，设 $\underline{\delta_0}$ 满足方程

$$(10)\qquad (\underline{A}+\lambda\underline{I})\underline{\delta_0}=\underline{g}\,,$$

则 $\underline{\delta_0}$ 在半径 $\|\underline{\delta}\|$ 满足 $\|\underline{\delta}\|^2=\|\underline{\delta_0}\|^2$ 的球面上使〈Φ〉为最小。

证：为了求出 δ 使

$$(11)\qquad \langle\Phi(\delta)\rangle=\|\underline{Y}-\underline{f_0}-\underline{P}\underline{\delta}\|^2$$

在约束条件

$$(12)\qquad \|\underline{\delta}\|^2=\|\underline{\delta_0}\|^2$$

下为最小，根据拉格朗日方法，产生驻点的必要条件为

$$(13)\qquad \frac{\partial u}{\partial\delta_1}=\frac{\partial u}{\partial\delta_2}=\cdots=\frac{\partial u}{\partial\delta_K}=0,\quad \frac{\partial u}{\partial\lambda}=0,$$

其中

$$(14)\qquad u(\underline{\delta},\lambda)=\|\underline{Y}-\underline{f_0}-\underline{P}\underline{\delta}\|^2+\lambda(\|\underline{\delta}\|^2-\|\underline{\delta_0}\|^2)$$

λ 是拉格朗日乘数。求上列各导数，得

$$(15)\qquad 0=-[\underline{P}^T(\underline{Y}-\underline{f_0})-\underline{P}^T\underline{P}\underline{\delta}]+\lambda\underline{\delta},$$

$$(16)\qquad 0=\|\delta\|^2-\|\delta_0\|^2$$

对于给定的 λ，方程

$$(17)\qquad (\underline{P}^T\underline{P}+\lambda\underline{I})\underline{\delta}=\underline{P}^T(\underline{Y}-\underline{f_0})$$

的解 $\underline{\delta}$ 满足(15)。将(17)左乘以 $(\underline{P}^T\underline{P})^{-1}$，并写成这样形状

$$(18)\qquad \underline{\delta}=(\underline{P}^T\underline{P})^{-1}\underline{P}^T(\underline{Y}-\underline{f_0})-(\underline{P}^T\underline{P})^{-1}\lambda\underline{\delta},$$

然后代入(15)就可验证这一点[译者注]。(10)与(17)是同一个方

[译者注] 还可简单地验证：从(17)式移项得 $\underline{P}^T\underline{P}+\lambda\underline{I}\underline{\delta}-\underline{P}^T(\underline{Y}-\underline{f_0})=0$，$-[\underline{P}^T(\underline{Y}-\underline{f_0})-\underline{P}^T\underline{P}]+\lambda\underline{\delta}=0$，可见方程(17)的解必满足(15)。

程。从 $(\underline{A}+\lambda\underline{I})$ 是正定的这个事实，显见这个驻点实际上就是极小点。

定理2　设 $\underline{\delta}(\lambda)$ 是方程组(10)对应于给定 λ 的解，则 $\|\underline{\delta}(\lambda)\|^2$ 是 λ 的这样连续递减函数，当 $\lambda\to\infty$ 时，$\|\delta(\lambda)\|^2\to 0$。

证：　因为对称矩阵 A 是正定的，可用一个正交规范转轴变换把 A 化成对角线矩阵 D，并保持两点间距离不变，记此变换为 $\underline{S}^T\underline{A}\underline{S}=\underline{D}$，这里 $\underline{S}^T\underline{S}=\underline{I}$，又 $\underline{D}$ 的对角线上元素均为正。这样，(10)就取这样形状：$(\underline{D}+\lambda\underline{I})\underline{S}^{-1}\underline{\delta_0}=\underline{S}^T\underline{g}$，从而得

$$\underline{\delta_0}=\underline{S}(\underline{D}+\lambda\underline{I})^{-1}\underline{S}^T\underline{g} \tag{19}$$

于是，令 $\underline{V}=\underline{S}^T\underline{g}$，得

$$\begin{aligned}\|\delta_0(\lambda)\|^2&=\underline{g}^T\underline{S}(\underline{D}+\lambda\underline{I})^{-1}\underline{S}^T\underline{S}(\underline{D}+\lambda\underline{I})^{-1}\underline{S}^T\underline{g}\\&=\underline{V}^T\left[(\underline{D}+\lambda\underline{I})^2\right]^{-1}\underline{V}\\&=\sum_{j=1}^{k}\frac{V_j^2}{(D_j+\lambda)^2}\end{aligned} \tag{20}$$

显然这是 λ 的递减函数($\lambda\geq 0$)，故当 $\lambda\to\infty$ 时，$\|\delta_0(\lambda)\|^2\to 0$。

这里已显明地揭示正交规范变换把矩阵变成对角线形状，这是为了便于下面定理的证明。

定理3.　设 γ 是 $\underline{\delta_0}$ 与 $\underline{\delta_g}$ 之间的夹角，则 γ 是 λ 的这样单调递减函数，当 $\lambda\to\infty$ 时，$\gamma\to 0$。由于 δ_g 不依赖于 λ，所以当 $\lambda\to\infty$ 时 δ_0 转向 $\underline{\delta_g}$。

证：　首先观察

$$\underline{\delta_g}=\left(\sum_{i=1}^{n}(Y_i-f_i)\frac{\partial f_i}{\partial b_j}\right)^T,\quad j=1,2,\cdots,k。 \tag{21}$$

因此（除差一个比例因子，不过这是无关紧要的，因为有关的仅是向量的方向）

$$(22)\qquad \underline{\delta}_{g} = \underline{g}\ .$$

根据 γ 角的定义

$$\cos\gamma = \frac{\underline{\delta}^{T}\underline{g}}{(\|\underline{\delta}\|)(\|\underline{g}\|)}$$

$$= \frac{\underline{V}^{T}(\underline{D}+\lambda\underline{I})^{-1}\underline{V}}{(\underline{V}^{T}[(\underline{D}+\lambda\underline{I})^{2}]^{-1}\underline{V})^{1/2}(\underline{g}^{T}\underline{g})^{1/2}}$$

$$= \frac{\sum_{j=1}^{K}\frac{v_j^2}{D_j+\lambda}}{\left[\sum_{j=1}^{K}\frac{v_j^2}{(D_j+\lambda)^2}\right]^{1/2}(\underline{g}^{T}\underline{g})^{1/2}}$$

求导并化简，得

$$(24)\qquad \frac{d}{d\lambda}\cos\gamma = \frac{\left[\sum_{j=1}^{K}\frac{v_j^2}{D_j+\lambda}\right]\left[\sum_{j=1}^{K}\frac{v_j^2}{(D_j+\lambda)^3}\right]-\left[\sum_{j=1}^{K}\frac{v_j^2}{(D_j+\lambda)^2}\right]^2}{\left[\sum_{j=1}^{K}\frac{v_j^2}{(D_j+\lambda)^2}\right]^{3/2}(\underline{g}^{T}\underline{g})^{1/2}}$$

$$(25)\qquad = \frac{\left[\sum_{j=1}^{K}v_j^2\Pi_{1j}\right]\left[\sum_{j=1}^{K}v_j^2\Pi_{3j}\right]-\left[\sum_{j=1}^{K}v_j^2\Pi_{2j}\right]^2}{\left[\sum_{j=1}^{K}\frac{v_j^2}{(D_j+\lambda)^2}\right]^{3/2}\left[\prod_{j=1}^{K}(D_j+\lambda)^2\right]^2(\underline{g}^{T}\underline{g})^{1/2}}$$

这里 $\Pi_{1j}=\prod_{\substack{j'=1\\ j'\neq j}}^{K}(D_{j'}+\lambda)$，$\Pi_{2j}=\prod_{\substack{j'=1\\ j'\neq j}}^{K}(D_{j'}+\lambda)^2$，$\Pi_{3j}=\prod_{\substack{j'=1\\ j'\neq j}}^{K}(D_{j'}+\lambda)^3$ 。

(25)的分母是正的，因为每个因子都是正的。因此 $\frac{d\cos\gamma}{d\lambda}$ 的符号决定于其分子的符号。注意到 $\Pi_{1j}\Pi_{3j}=(\Pi_{2j})^2$ ，上式的

分子可写成

$$(26)\quad \left[\sum_{j=1}^{K}(v_j\Pi_{1j}^{1/2})^2\right]\left[\sum_{j=1}^{K}(v_j\Pi_{3j}^{1/2})^2\right]-\left[\sum_{j=1}^{K}(v_j\Pi_{1j}^{1/2})(v_j\Pi_{3j}^{1/2})\right]^2.$$

由许瓦兹不等式，(26)为正，故 $\frac{d\cos\gamma}{d\lambda}$ 恒为正（$\lambda>0$）。于是 γ 是 λ 的单调递减函数。

对于非常大的 λ 值，矩阵（$\underline{A}+\lambda\underline{I}$）由其对角线矩阵 $\lambda\underline{I}$ 所控制。因此，从(10)可看出当 $\lambda\to\infty$ 时，$\delta_0\to g/\lambda$，从而在极限状态时，δ_0 与 g 成比例，所以它们之间的夹角趋于零。另一方面，若在(10)中，$\lambda=0$ 则（除 A 是对角线矩阵这种明显的情况而外）向量 δ_g 与 g 交成一个有限的角，$0<\gamma<\pi/2$。因此 γ 是 λ 的这样连续单调递减函数，当 $\lambda\to\infty$ 时，$\gamma\to 0$。

量测的变尺度　方程组（6）的解 $\underline{\delta_t}$ 的有关性质对于 $\underline{b}$ — 空间中线性变换是不变的。可是人所熟知，〔3〕斜量法不具有尺度不变性，故有必要按某种方式改变 $\underline{b}$ — 空间的尺度。我们选取导数 $\frac{\partial f_i}{\partial b_j}$ 在所有数据点 $i=1,2,\cdots,n$ 的标准差作为 b — 空间的尺度单位，由于这些导数一般是依赖于 b_j 本身，因而 b_j 的现有试探值就得用来计算导数。这样选择的尺度会使矩阵 A 变换成 $\frac{\partial f_i}{\partial b_j}$ 的相关系数矩阵。事实上，这样选择尺度在线性最小平方问题中已经广泛地使用，作为改进数值计算过程的一种手段。

因此，我们定义变尺度后的矩阵 $\underline{A}^*$ 与变尺度后的向量 $\underline{g}^*$：

$$(27)\qquad \underline{A}^*=(a_{jj'}^*)=\left(\frac{a_{jj'}}{\sqrt{a_{jj}}\sqrt{a_{j'j'}}}\right)$$

$$(28)\qquad \underline{g}^*=(g_j^*)=\left(\frac{g_j}{\sqrt{a_{jj}}}\right)$$

并由

$$(29) \qquad \underline{A}^* \underline{\delta_t}^* = \underline{g}^*$$

解出台劳级数方法的校正向量，于是得

$$(30) \qquad \underline{\delta_j} = \underline{\delta_j}^* / \sqrt{a_{jj}} \quad .$$

算法的构造　适用的算法的大概轮廓现在已经明显了。确定地说，对于第r次迭代，造方程

$$(31) \qquad (\underline{A}^{*(r)} + \lambda^{(r)} \underline{I}) \underline{\delta}^{*(r)} = \underline{g}^{*(r)}$$

由此方程，解出$\underline{\delta}^{*(r)}$。方程(30)可用来求$\underline{\delta}^{(r)}$，新试探向量

$$(32) \qquad \underline{b}^{(r+1)} = \underline{b}^{(r)} + \underline{\delta}^{(r)}$$

给出新平方和$\Phi^{(r+1)}$。要紧的是要这样选择$\lambda^{(r)}$使得

$$(33) \qquad \Phi^{(r+1)} < \Phi^{(r)} \quad .$$

从前面理论，显见恒存在一个充分大的$\lambda^{(r)}$使得(33)成立，除非$b^{(r)}$已经到达Φ的极小点。因此需要某种形式的试错方法求出$\lambda^{(r)}$，使得(33)成立并使算法迅速地收敛于最小平方值。

在每次迭代中，我们都要求在（近似的）最大邻域中Φ为最小，在此最大邻域中，线性函数就给出非线性函数的合适的表示。因此，选择$\lambda^{(r)}$的策略总是要求用很小的$\lambda^{(r)}$值，每当未改造的台劳级数方法收敛得很好的这种情况时。这在收敛过程的最后阶段是特别有关系的，当猜测是在极小点的极邻近处进行时，这里Φ的等值面渐近于椭球面，模型的线性展开必须在非常小的区域内才是一个好的近似。因此只有当必须满足(33)时才用大的$\lambda^{(r)}$值。假使$\Phi^{(r+1)}$作为λ的函数，的确有一

个极小值，而且在第 r 次迭代中，选择与这极小值相应的 λ 值使（$\Phi^{(r)}-\Phi^{(r+1)}$）为极大，那末这个局部最优选择对于大范围来说并不是好的策略，因为这实质上要求 λ 比为了满足 (3.3) 所必要的大的值还要大。这样的策略会保留最速下降法的许多性质，比如，开头很快地前进，跟着就越来越慢了。

因此我们应规定如下的策略：

设 $\nu > 1$

令 $\lambda^{(r-1)}$ 表示由前一次迭代算出的 λ 值

计算 $\Phi(\lambda^{(r-1)})$ 与[注2] $\Phi(\lambda^{(r-1)}/\nu)$

i) 如果 $\Phi(\lambda^{(r-1)}/\nu) \leq \Phi^{(r)}$，则令 $\lambda^{(r)}=\lambda^{(r-1)}/\nu$

ii) 如果 $\Phi(\lambda^{(r-1)}/\nu) > \Phi^{(r)}$，$\Phi(\lambda^{(r-1)}) \leq \Phi^{(r)}$，则令 $\lambda^{(r)}=\lambda^{(r-1)}$。

iii) 如果 $\Phi(\lambda^{(r-1)}/\nu) > \Phi^{(r)}$，$\Phi(\lambda^{(r-1)}) > \Phi^{(r)}$，则继续[注3]乘以 ν 来加大 λ 一直到对于某个最小的正整数 w 使得 $\Phi(\lambda^{(r-1)}\nu^{w}) \leq \Phi^{(r)}$ 为止，这时，令 $\lambda^{(r)}=\lambda^{(r-1)}\nu^{w}$。

由这种算法，我们常常得到一个可使用的邻域，而且除一个由 ν 决定的因子外，我们几乎经常得到最大邻域，在其中台劳级数就给出符合我们目的的一个合适的表示。当 $\frac{|\delta_j^{(r)}|}{\tau+|b_j^{(r)}|} < \varepsilon$，对一切 j，某个恰当小的 $\varepsilon > 0$，比如说 10^{-5}，与某个恰当的 τ，比如说 10^{-3} 成立时，迭代收敛。ν 的选择是任意的；取 $\nu=10$ 经实践证实是一个好的选择。

注 2　如果 $\lambda^{(r-1)}$ 按有效数字个数与 1.0 比较已可忽略的话，则可立即进行检验 ii，或 iii，不必计算 $\Phi(\lambda^{(r-1)}/\nu)$，并省去含有 $\Phi(\lambda^{(r-1)}/\nu)$ 的比较。

注 3　遇到这样的问题，参数估计的相关系数是极高的（>0.99）。

确定地说，情况 iii，只是少见的。因此，在每次迭代中，常々要用两个 $\lambda^{(r)}$ 的值来解(31)。其中一个解是为了标准的台劳级数方法所需要的，多余的一个解一般所需要的计算量比计算 A^* 的计算量少，所以每次迭代中，小比例的增加计算量会在迭代功效方面得到更加的补偿。

可能还要注意，当用浮点算术时，在单精度计算 $\hat{Y}_i$ 后，常々可望双精度累积(3)式的平方和（为了检验 i，与 ii，之用）做不到这一点则在极小点附近，单凭舍入误差就会导致Φ的错误性质。将 λ 加到 A^* 的对角线元素上去，这在数值计算上的必然便利就是合成矩阵（A^*+IA）常々比 A^* 本身具有较好的制约性［译者注：即从矩阵的病态的角度来说］，这会出现 λ 增大到不合理的大值。在这种情况，修改检验 iii，会有帮助的。这样修改检验：

设 $\underline{b}^{(r+1)}=\underline{b}^{(r)}+K^{(r)}\underline{\delta}^{(r)}$，$K^{(r)}\leqslant 1$

注意到角 $\gamma^{(r)}$ 是 $\lambda^{(r)}$ 的递减函数，选择一个准则角 $\gamma_0<\pi/2$，并取

$$K^{(r)}=1,\ 若\ \gamma^{(r)}\geqslant\gamma_0$$

如果 $\lambda^{(r)}$ 已增大到使 $\gamma^{(r)}<\gamma_0$，而检验 iii 还不能通过的话，则不再加大 $\lambda^{(r)}$，而是选取充分小的 $K^{(r)}$ 使 $\Phi^{(r+1)}<\Phi^{(r)}$。这是可以做到的，因为 $\gamma^{(r)}<\gamma_0<\pi/2$。合适的准则角选为 $\gamma_0=\pi/4$。

可能还要注意，当 $\lambda=0$ 时，$\cos\gamma$ 为正只有当 $\underline{A}$ 是正定时才能保证到。在出现极高的相关系数时，矩阵 $\underline{A}$ 的正定性单凭舍入误差就能被破坏。在这种情况，无论用什么值 $K^{(r)}$，台劳级数方法都会发散。现在描述的算法，由于用了（$\underline{A}+\lambda\underline{I}$），

所以纵然 $\underline{A}$ 可能不是十分正定的，也能保证 $\cos\gamma$ 为正。虽然为非线性模型配数据时，不一定就能预见到参数估计具有很高的相关系数，不过较好的实验设计常能真正地减小这些极端的相关系数，当它们出现的时候。

对于其它问题的应用　　虽然所描述的方法也能用到其它类型的问题上去。例如，可以在解非线性代数方程组的斜量法与牛顿——拉普孙方法之间进行内插，应用隐方法解非线性边值问题所关联的非线性差分联立方程。总之，这样的问题都是最小平方估计问题在 $n=k$ 时的特殊情况。设待解的方程组能写成

$$0=f_i(\underline{X}) \qquad i=1,2,---,k, \tag{34}$$

这里 $\underline{X}$ 是未知量 x_j 的 k 维向量

对于斜量法，习惯上定义一个误差准则

$$\Phi=\sum_{i=1}^{k}\left[f_i(\underline{X})\right]^2 \tag{35}$$

并沿着由

$$-\left(\frac{\partial\Phi}{\partial x_j}\right)=-\left(2\sum_{i=1}^{k}f_i\frac{\partial f_i}{\partial x_j}\right) \tag{36}$$

定义的Φ的负梯度方向上移动来校正试探向量 $\underline{X}^{(r)}$。

于是

$$x_j^{(r+1)}=x_j^{(r)}+\delta_j^{(r)}, \qquad j=1,2,---,k, \tag{37}$$

这里

$$\delta_j^{(r)}=-a\frac{\partial\Phi}{\partial x_j}, \qquad j=1,2,---,k, \tag{38}$$

a 是一个恰当定义的常数。

用矩阵记号，则有

$$(39) \qquad \left(\frac{1}{a}\underline{I}\right)\underline{\delta}_g^{(r)} = \underline{g}^{(r)}。$$

另一方面，牛顿——拉夫森方法，它包含 f_i 的台劳级数展开式的一次项，给出方程

$$(40) \qquad \sum_{j=1}^{k}\left(\frac{\partial f_i}{\partial x_j}\right)^{(r)} \delta_j^{(r)} = -f_i^{(r)}, \quad i=1,2,\cdots,k.$$

用矩阵记号，则有

$$(41) \qquad \underline{B}\,\underline{\delta}_t = \underline{f},$$

这里矩阵 $\underline{B} = -\left(\frac{\partial f_i}{\partial x_j}\right)$ 一般不是对称的。将(41)左乘以 $\underline{B}^T$ 形成 $\underline{B}^T\underline{B} = \underline{A}$，$\underline{B}^T\underline{f} = \underline{g}$。矩阵 $\underline{A}$ 是对称的。于是矩阵 $\underline{A}$ 与向量 $\underline{g}$ 分别规范化为 $\underline{A}^*$ 与 $\underline{g}^*$

应用这新算法，则形成方程组

$$(42) \qquad (\underline{A}^* + \lambda^{(r)}\underline{I})\underline{\delta}^{*(r)} = \underline{g}^*$$

这与非线性最小平方估计问题所形成的方程组完全一样。求解的计算方法也完全一样。不过，除舍入误差而外，至的最小值已知是零。

结　论　　所描述的算法分担着斜量法的能力，即能从一个初始值收敛，而这个初始值可能是在其它方法的收敛域之外。这个方法又分担着台劳级数方法的能力，即当进入收敛值的邻

域后，就能迅速地接近收敛值。因此这个方法兼併以前两个方法的最优特点，同时又避免了它们的严重的限制性。

计算机程序　　Fortrain 程序" Least - Squares Estimation of Nonlinear Parameters" 实现了本文中的算法，作为 IBM 分程序 No. 1428 采用。