在多项式型变换下具有不变性的算法

邓乃扬　陈志

（应用数学系）

摘　要

本文讨论根据较一般的目标函数模型（3）建立具有非线性尺度不变性的算法的理论依据。文献[2]研究了 $F(q)$ 是二次函数的情形，我们可以把它看做是用 q 的二次函数逼近一般的 $F(q)$。本文研究逼近 $F(q)$ 的一般问题，对于用高阶多项式逼近和用低阶多项式分段逼近的两种策略，分别导出了计算 ρ_K 的解析表达式，并建立了相应的算法

On the Algorithms Invariant to Polynomial-Type Scaling

Deng Nai-yang and Chen Zhi

Abstract

This paper discusses the algorithms invariant to nonlinear scaling for more general objective function model(3). The case of algorithms was studied in [2] when F(q) is a quadratic function. Here we study a more general problem — approximation F(q) with high-order polynomial and with piecewise low-order polynomial. The analytic formulas of ρ_K for both cases are derived and the corresponding algorithms are established.

考虑正定二次函数

$$q(x) = \frac{1}{2}x^{T}Gx + r^{T}x + \delta \tag{1}$$

和非线性尺度变换

本文于1984年6月21日收到。

$$f=F(q),\qquad \frac{dF}{dq}>0 \tag{2}$$

生成的目标函数

$$f(x)=F(q(x)). \tag{3}$$

Spedicato[1]首次提出应建立对变换(2)具有不变性的算法，这无疑是一个很有前途的研究领域。我们已经知道[1],[2],[3],在许多共轭梯度法或拟牛顿法计算搜索方向时，用

$$\overset{+}{y}_{K-1}=\rho_K g_K-g_{K-1}\qquad \left(\rho_K=\frac{d}{dq}F(q(x_{K-1}))\Big/\frac{d}{dq}F(q(x_K)\right)$$

代替原公式中的 $y_{K-1}=g_K-g_{K-1}$，就能得到对非线性尺度变换（2）具有不变性的算法。例如文献[2]曾证明如下结论：

命题Ⅰ．考虑式（3）所示的目标函数。设在 PRP 算法中改用

$$\begin{aligned}\widehat{p}_1&=-\widehat{g}_1,\\ \widehat{p}_K&=-\widehat{g}_K+\widehat{\beta}_{K-1}p_{K-1},\ (k>1)\end{aligned} \tag{4}$$

确定搜索方向，其中 $\widehat{g}_K$是目标函数在 x_K点处（对 x）的梯度，而

$$\widehat{\beta}_{K-1}=\frac{\widehat{g}_K^{T}(\rho_K\widehat{g}_K-\widehat{g}_{K-1})}{\|\widehat{g}_{K-1}\|^2},\quad (k>1)$$

$$\rho_K=\frac{d}{dq}F(q(x_{K-1}))\Big/\frac{d}{dq}F(q(x_K))$$

若初始点 x_1相同，则把上述算法应用于不同的F生成的目标函数（3）时，得到的搜索方向和点列是相同的。

因此，导出计算 ρ_K 的表达式是建立具有不变性的算法的关键。对F为二次多项式的特殊情形，文献[2]通过较冗长的推算，得到了如下结论：

命题Ⅱ．设$F(q)$是二次多项式

$$F(q)=\varepsilon_1 q+\frac{\varepsilon_2}{2}q^2,\ \left(\frac{dF}{dq}>0,\ q>0\right)。$$

若 p_{K-1}是 x_{K-1}处的下降方向，μ^*是最优步长

$$f(x_K)=f(x_{K-1}+\mu^* p_{K-1})=min\{f(x_{K-1}+\lambda p_{K-1})\mid\lambda\geqslant 0\}$$

则

$$\rho_K=\begin{cases}\sqrt{f(x_{K-1})/f(x_K)}, & A=0;\\ 1 & B=0;\\ |\mu^*\langle g_{K-1},p_{K-1}\rangle/[\mu^*\langle g_{K-1},p_{K-1}\rangle+4(f(x_{K-1})-f(x_K))]|, & \text{其它,}\end{cases} \tag{5}$$

其中 g_{K-1}是 $f(x)=F(q(x))$在点 x_{K-1}处的梯度，而

$$A=\mu^*\langle g_{K-1},p_{K-1}\rangle^2+16f^2(x_{K-1})+8\mu^*\langle g_{K-1},p_{K-1}\rangle f(x_{K-1})-16f(x_{K-1})f(x_K)$$

$$B=8(f(x_{K-1})-f(x_K))+4\mu^*\langle g_{K-1},p_{K-1}\rangle$$

且A与B不同时为零.

本文讨论根据一般的目标函数模型建立相应无约束算法的理论依据. 某些基于二次模型

构造的算法，可以看做是用q的一次函数逼近而得到的。而基于命题Ⅱ构造算法，可以看做是用q的二次函数逼近$F(q)$。本文研究逼近$F(q)$的一般问题。事实上，在一次迭代中（特别是在执行算法的前期），除了命题Ⅱ计算ρ_K时使用的数据外，进行一维搜索时，往往还能得到其它某些点处的若干信息。我们将讨论如何充分利用这些信息，从而达到在基本上不增加计算量的条件下更好地逼近$F(q)$的目的。为此，以下先对$F(q)$是一般多项式的情形，

$$F(q)=\varepsilon_0+\varepsilon_1 q+\varepsilon_2 q^2+\cdots+\varepsilon_t q^t$$

导出计算ρ_K的解析表达式，然后对一般非线性变换（2）提出两种逼近策略，为建立一类更合理的算法奠定理论基础。

一、当F(q)为多项式时ρ_K的计算

定理1. 考虑式（3）所示的目标函数，其中$F(q)$是$2m+p$次多项式

$$F(q)=\varepsilon_0+\varepsilon_1 q+\varepsilon_2 q^2+\cdots+\varepsilon_{2m+p}q^{2m+p}$$

设p_{K-1}是点x_{K-1}处的下降方向，记

$$\psi(\lambda)=f(x_{K-1}+\lambda p_{K-1})=F(q(x_{K-1}+\lambda p_{K-1})) \tag{6}$$

并记$\psi(\lambda)$的极小点为μ^*，$x_K=x_{K-1}+\mu^* p_{K-1}$. 设点

$$0=\mu_1,\ \cdots,\ \mu_m,\ \mu_{m+1},\ \cdots,\ \mu_{m+p},\ \mu^*$$

满足

$$|\mu_i-\mu^*|\neq|\mu_j-\mu^*|\neq 0,\quad (i\neq j). \tag{7}$$

若已知

$$\psi(\mu^*)=\psi^*,\ \psi(\mu_j)=\psi_i,\quad (j=1,\ \cdots,\ m+p) \tag{8}$$

$$\psi'(\mu_1)=\psi_1'<0,\ \psi'(\mu_j)=\psi_j',\quad (j=2,\ \cdots,\ m) \tag{9}$$

则

$$\rho_K=-\frac{\psi_1'}{2}\mu^*\cdot\frac{1}{\widetilde{H}} \tag{10}$$

其中

$$\widetilde{H}=\sum_{j=1}^{m}\left[\left(\psi_i-\psi^*\right)u_{1j}b_j-\frac{\mu^*}{2}\left|\ \psi_j'\ \right|\sqrt{u_{1i}}\right]L_i(0)$$
$$+\sum_{j=m+1}^{m+p}\psi_j u_{1j}L_j(0) \tag{11}$$

$$L_i(0)=\begin{cases}\displaystyle\prod_{\substack{i=r\\ i\neq j}}^{m}\frac{1}{(u_{ji}-1)^2}\prod_{l=m+1}^{m+p}\frac{1}{1-u_{jl}}, & j=1,\ \cdots,\ m;\\[2ex] \displaystyle\prod_{i=1}^{m}\frac{1}{(u_{ji}-1)^2}\prod_{\substack{l=m+1\\ l\neq j}}^{m+p}\frac{1}{1-u_{jl}}, & j=m+1,\ \cdots,m+p\end{cases} \tag{12}$$

$$b_j=2+2\sum_{\substack{i=1\\ i\neq j}}^{m}\frac{1}{1-u_{ij}}+\sum_{l=m+1}^{m+p}\frac{1}{1-u_{1j}} \tag{13}$$

这里 u_{ij} 是由点 μ_1，…，μ_{m+p} 和 μ^* 确定的量

$$u_{ij}=\left(\frac{\mu_i-\mu^*}{\mu_j-\mu^*}\right)^2 \tag{14}$$

证明：记 $\varphi(\lambda)=q(x_{K-1}+\lambda p_{K-1})$，则有

$$\psi(\lambda)=F(\varphi(\lambda)) \tag{15}$$

而且 φ 是严格凸二次函数

$$\varphi(\lambda)=q(x_{K-1}+\lambda p_{K-1})=a(\lambda-\mu^*)^2+b,\quad (a>0) \tag{16}$$

再注意到 $\varphi'(\mu^*)=0$，可知

$$\psi'(0)=\frac{dF}{d\varphi}\bigg|_{\varphi=\varphi((0)}\cdot\frac{d\varphi}{d\lambda}\bigg|_{\lambda=0} \tag{17}$$

$$\psi''(\mu^*)=\frac{dF}{d\varphi}\bigg|_{\varphi=\varphi(\mu^*)}\cdot\frac{d^2\varphi}{d\lambda^2}\bigg|_{\lambda=\mu^*} \tag{18}$$

利用以上两式推得

$$\frac{\left.\frac{dF}{d\varphi}\right|_{\varphi=\varphi(0)}}{\left.\frac{dF}{d\varphi}\right|_{\varphi=\varphi(\mu^*)}}=\frac{\psi'(0)}{\psi''(\mu^*)}\cdot\frac{\varphi''(\mu^*)}{\varphi'(0)}$$

由于 $\varphi''(\mu^*)=2a$，$\varphi'(0)=-2a\mu^*$，故有 $\varphi''(\mu^*)/\varphi'(0)=-1/\mu^*$.

再由 ρ_K 的定义 $\rho_K=\frac{dF}{dq}\Big(q(x_{K-1})\Big)\Big/\frac{d}{dq}F\Big(q(x_K)\Big)$ 知

$$\rho_K=\frac{dF}{d\varphi}\bigg|_{\varphi=\varphi(0)}\bigg/\frac{dF}{d\varphi}\bigg|_{\varphi=\varphi(\mu^*)}=-\frac{\psi'(0)}{\mu^*}\cdot\frac{1}{\psi''(\mu^*)}.$$

根据 $\mu_1=0$，即可把上式写为

$$\rho_K=-\frac{\psi_1'}{\mu^*}\cdot\frac{1}{\psi''(\mu^*)} \tag{19}$$

由此可见，只要算出 $\psi''(\mu^*)$，就能解析地求得 ρ_K.

现在考虑如何用式（8）、（9）中的数据计算 $\psi''(\mu^*)$。由定理条件及式（15）、(16) 得知 $\psi(\lambda)$ 是 $(\lambda-\mu^*)^2$ 的 $2m+p$ 次多项式

$$\psi(\lambda)=H_0+H_1(\lambda-\mu^*)^2+H_2(\lambda-\mu^*)^4+\cdots+H_{2m+p}(\lambda-\mu^*)^{2(2m+p)} \tag{20}$$

故

$$\psi''(\mu^*)=2H_1,\quad \rho_K=-\frac{\psi_1'\mu^*}{2}\cdot\frac{1}{(\mu^*)^2H_1} \tag{21}$$

作变换 $\xi=(\lambda-\mu^*)^2$，并记

$$\xi_j=(\mu_j-\mu^*)^2 \tag{22}$$

则 $2m+p$ 次多项式

$$h(\xi)=H_1\xi+H_2\xi^2+\cdots+H_{2m+p}\xi^{2m+p} \tag{23}$$

满足

$$\begin{aligned} &h(0)=\psi(\mu^*)-\psi^*=0\\ &h_j\equiv h(\xi_j)=\psi(\mu_j)-\psi^*=\psi_j-\psi^*,\ (j=1,\ \cdots,m+p)\\ &h'_j\equiv h'(\xi_j)=\frac{\psi'(\mu_j)}{2(\mu_j-\mu^*)}=\frac{\psi'_j}{2(\mu_j-\mu^*)},\ (j=1,\cdots,m). \end{aligned} \tag{24}$$

据此可以用构造基函数的方法计算出 H_1。事实上，先对 $j=1,\cdots,m$ 定义

$$L_j(\xi)=\prod_{\substack{i=1\\ i\neq j}}^{m}\left(\frac{\xi-\xi_i}{\xi_j-\xi_i}\right)^2\cdot\prod_{l=m+1}^{m+p}\frac{\xi-\xi_l}{\xi_j-\xi_l}, \tag{25}$$

$$D_j(\xi)=(a_j\xi+b_j)\frac{\xi}{\xi_j}L_j(\xi), \tag{26}$$

$$E_j(\xi)=(c_j\xi+d_j)\frac{\xi}{\xi_j}L_j(\xi). \tag{27}$$

其中

$$\begin{aligned} &a_j=-\left[\frac{1}{\xi_j}+2\sum_{\substack{i=1\\ i\neq j}}^{m}\frac{1}{\xi_j-\xi_i}+\sum_{l=m+1}^{m+p}\frac{1}{\xi_j-\xi_l}\right]\\ &b_j=2+2\sum_{\substack{i=1\\ i\neq j}}^{m}\frac{\xi_j}{\xi_j-\xi_i}+\sum_{l=m+1}^{m+p}\frac{\xi_j}{\xi_j-\xi_l},\\ &c_j=1,\qquad d_j=-\xi_j. \end{aligned} \tag{28}$$

再对 $j=m+1,\cdots,m+p$ 定义

$$L_j(\xi)=\prod_{i=1}^{m}\left(\frac{\xi-\xi_i}{\xi_j-\xi_i}\right)^2\cdot\prod_{\substack{l=m+1\\ l\neq j}}^{m+p}\frac{\xi-\xi_l}{\xi_j-\xi_l}, \tag{29}$$

则有

$$\begin{aligned} h(\xi)&=\sum_{j=1}^{m}h_jD_j(\xi)+\sum_{j=1}^{m}h'_jE_j(\xi)+\sum_{j=m+1}^{m+p}h_j\frac{\xi}{\xi_j}L_j(\xi)\\ &=\sum_{j=1}^{m}h_j(a_j\xi+b_j)\frac{\xi}{\xi_j}L_j(\xi)+\sum_{j=1}^{m}h'_j(c_i\xi+d_i)\frac{\xi}{\xi_j}L_j(\xi)\\ &\quad+\sum_{j=m+1}^{m+p}h_j\frac{\xi}{\xi_j}L_j(\xi) \end{aligned}$$

因此

$$H_1=\sum_{j=1}^{m}\left[h_jb_j\frac{1}{\xi_j}-h'_j\right]L_j(0)+\sum_{j=m+1}^{m+p}h_j\frac{1}{\xi_j}L_j(0) \tag{30}$$

其中

$$L_j(0)=\begin{cases}\prod\limits_{\substack{i=1\\i\neq j}}^{m}\dfrac{1}{(u_{ji}-1)^2}\cdot\prod\limits_{l=m+1}^{m+p}\dfrac{1}{1-u_{jl}}, & j=1,\cdots,m\\[2ex] \prod\limits_{i=1}^{m}\dfrac{1}{(u_{ji}-1)^2}\cdot\prod\limits_{\substack{l=m+1\\l\neq j}}^{m+p}\dfrac{1}{1-u_{jl}}, & j=m+1,\cdots,m+p\end{cases}\tag{31}$$

$$b_j=2+2\sum_{\substack{i=1\\i\neq j}}^{m}\frac{1}{1-u_{ij}}+\sum_{l=m+1}^{m+p}\frac{1}{1-u_{lj}}\tag{32}$$

$$u_{ij}=\left(\frac{\mu_i-\mu^*}{\mu_j-\mu^*}\right)^2\tag{33}$$

把式（24）代入式（30），便有

$$H_1=\sum_{j=1}^{m}\left[\left(\psi_j-\psi^*\right)\frac{b_j}{(\mu_j-\mu^*)^2}-\frac{\psi_j'}{2(\mu_j-\mu^*)}\right]L_j(0)$$

$$+\sum_{j=m+1}^{m+p}(\psi_j-\psi^*)\frac{1}{(\mu_j-\mu^*)^2}L_j(0)$$

再记$\widetilde{H}=\mu^{*2}H_1=(\mu_1-\mu^*)^2H_1$，即知(10)式成立．定理证毕．

根据定理1对实际上常用的几种情形进行计算，所得结果如下。

当$m=1$，$p=0$时的情形．此时假定已知一维搜索初始点处的函数值和导数值以及终点处的函数值，是命题Ⅱ要解决的问题．设$F(q)$是二次函数，有

$$b_1=2,\ L_1(0)=0,\ u_{11}=1$$

$$\widetilde{H}=(\psi_1-\psi^*)\cdot 1\cdot 2+\frac{\mu^*}{2}\psi_1'$$

$$\rho_K=-\frac{\psi_1'\mu^*}{2}\cdot\frac{1}{\widetilde{H}}=\frac{-\psi_1'\mu^*}{4(\psi_1-\psi^*)+\psi_1'\mu^*}\tag{34}$$

与命题Ⅱ所得结果（5）相比较可见，此处ρ_K的表达式是以统一的形式给出的，比较完整；另外还放松了命题Ⅱ的某些限制条件，这无异于扩大了定理的应用范围，在实际上是有意义的。

当$m=1$，$p=1$时的情形．这相当于在一维搜索的初始点和终点之间，还曾计算过一点处的函数值．此时假定$F(q)$是三次多项式，有

$$u_{12}=\frac{\mu^{*2}}{(\mu_2-\mu^*)^2},\ u_{21}=\frac{(\mu_2-\mu^*)^2}{\mu^{*2}},\ b_1=2,$$

$$L_1(0)=\frac{1}{1-u_{12}},\ L_2(0)=\frac{1}{(u_{21}-1)^2}$$

由此可得

$$\rho_K = -\frac{\psi_1'}{2}\mu^*\widetilde{H}^{-1} \tag{35}$$

其中

$$\widetilde{H}=2(\psi_1-\psi^*)\frac{(\mu_2-\mu^*)^2}{(\mu_2-\mu^*)^2-\mu^{*2}}-\frac{\mu^*}{2}\psi_1'\frac{(\mu_2-\mu^*)^2}{(\mu_2-\mu^*)^2-\mu^{*2}}$$

$$+(\psi_2-\psi^*)\frac{\mu^{*2}}{(\mu_2-\mu^*)^2}\cdot\left[\frac{\mu^{*2}}{(\mu_2-\mu^*)^2-\mu^{*2}}\right]^2 \tag{36}$$

当 $m=2$，$p=0$ 时的情形. 这相当于在一维搜索的初始点和终点之间，还曾计算过一点处的函数值和导数值. 此时假定 $F(q)$ 是四次多项式，有

$$u_{12}=\frac{\mu^{*2}}{(\mu_2-\mu^*)^2},\ u_{21}=\frac{(\mu_2-\mu^*)^2}{\mu^{*2}},\ b_1=2\left(1+\frac{1}{1-u_{21}}\right),$$

$$b_2=2\left(1+\frac{1}{1-u_{12}}\right),\ L_1(0)=\frac{1}{(u_{12}-1)^2},\ L_2(0)=\frac{1}{(u_{21}-1)^2}$$

由此可得

$$\rho_K = -\frac{\psi_1'}{2}\mu^*\widetilde{H}^{-1} \tag{37}$$

其中

$$\widetilde{H}=2(\psi_1-\psi^*)\left(1+\frac{\mu^{*2}}{\mu^{*2}-(\mu_2-\mu^*)^2}\right)\left[\frac{(\mu_2-\mu^*)^2}{\mu^{*2}-(\mu_2-\mu^*)^2}\right]^2+$$

$$+\frac{\mu^*}{2}\psi_1'\left[\frac{(\mu_2-\mu^*)^2}{\mu^{*2}-(\mu_2-\mu^*)^2}\right]^2+2(\psi_2-\psi^*)\left(1+\frac{(\mu_2-\mu^*)}{(\mu_2-\mu^*)^2-\mu^{*2}}\right)\frac{(\mu_2-\mu^*)\mu^{*2}}{(\mu_2-\mu^*)^2-\mu^{*2}}$$

$$-\frac{\mu^*}{2}\psi_2'\frac{\mu^*}{\mu_2-\mu^*}\left[\frac{\mu^{*2}}{(\mu_2-\mu^*)^2-\mu^{*2}}\right]^2 \tag{38}$$

二、对一般F(q)的两种逼近策略

对于满足条件（2）的一般的$F(q)$来说，在某次迭代中有时能用低次多项式近似，有时则不能. 可以想见，当不能用低次多项式近似时，进行相应一维搜索的过程中往往也需要计算较多点处的函数值（或导数值）. 这自然为较精确地逼近 $F(q)$ 提供了条件. 具体地说，我们可以采用以下两种策略.

1. 用高次多项式逼近F(q)

可依据定理1进行，在各次迭代中所用多项式的次数由当前迭代一维搜索得到的信息确定，它们可以随着迭代次数的变化而改变. 倘若进行第 $k-1$ 次迭代的一维搜索时曾对 $\psi(\lambda)=f(x_{K-1}+\lambda p_{K-1})$计算出了式（8)、(9）所示的数据(粗略地说，即除 $\mu^*=0$处的函数值 ψ_1 和导数值 ψ_1'以及在极小点 μ^*处的函数值 ψ^*外，还计算出了其它 $m-1$ 个点处的函数值和导数值以及另外 p个点处的函数值)，则可用 $2m+p$ 次多项式逼近真正的 $F(q)$（即实际计算时假设$F(q)$ 为 $2m+p$ 次多项式). 这样，依据定理1就可以对已有的许多算法进行改造. 例如PRP 算法可改造成如下的算法.

改进PRP算法Ⅰ

(1) 取初始点 x_1，置 $k=1$.

(2) 计算 $g_K=g(x_K)$.

(3) 若 $g_K=0$，则停止计算；否则：当 $k=1$ 时，置 $\beta_1=0$ 转5；当 $k\neq1$ 时转4.

(4) 计算 ρ_K. 根据前次（第 $k-1$ 次）迭代的一维搜索所得数据：

$$\begin{aligned}&\psi(\mu^*)=\psi^*,\ \psi(\mu_j)=\psi_j,\ (j=1,\cdots,m+p)\\&\psi'(\mu_j)=\psi'_j,\ (j=1,\cdots,m)\end{aligned}\tag{39}$$

按式（10）—（14）计算 ρ_K. 当以上数据不满足式（7）时，需进行一些改造. 例如倘若存在着 i，j $(m<i<j\leqslant m+p)$ 使得 $\mu_i-\mu^*=\mu^*-\mu_j$，则可把 μ_i和 μ_j看成为一个点 μ_i，改取$\psi(\mu_i)=\frac{1}{2}(\psi(\mu_i)+\psi(\mu_j))$.

(5) 计算 p_K. 令

$$p_K=-g_K+\beta_{K-1}p_{K-1},\quad \beta_{K-1}=g_K^T(\rho_K g_K-g_{K-1})/\|g_{K-1}\|^2.$$

(6) 一维搜索. 求 λ_K使

$$f(x_K+\lambda_K p_K)=min\{f(x_{K-1}+\lambda p_{K-1})|\lambda\geqslant0\}.$$

(7) 置 $x_{K+1}=x_K+\lambda_K p_K$，置 $k=k+1$，转2.

对上述算法有下列定理.

定理2. 若在每次执行改进 PRP 算法第4步时，所用数据（39）中的 $2m+p$ 都不小于某个正整数 t，则算法用于七次多项式 $F(q)$ 生成的目标函数(3)时，具有 n 步终止性.

证明：根据定理1和命题Ⅰ证明. 详论从略.

说明：上述改进的 PRP 算法只是许多可能形式中的一种，例如还可以考虑增加某种重新开始策略，或者采用某种变度量形式确定搜索方向等等.

2. 用二次多项式分段逼近F(q)

假设在某次迭代的一维搜索中得到了如下的数据：

$$\psi(\mu_j)=\psi_j,\ \psi'(\mu_j)=\psi'_j,\ (j=1,2,\cdots,m,m\geqslant3)$$

不妨假定

$$\begin{aligned}&0=\mu_1<\mu_2<\cdots<\mu_{m-1}<\mu_m\\&\psi'_j<0\quad(j=1,2,\cdots,m-1)\end{aligned}$$

记 $x_K=x_{K-1}+\mu_m p_{K-1}$，则有

$$\rho_K=\frac{\frac{d}{dq}F(q(x_{K-1}))}{\frac{d}{dq}F(q(x_K))}=\gamma_{1m}=\gamma_{12}\cdot\gamma_{23}\cdot\cdots\cdot\gamma_{m-1m}\tag{40}$$

其中

$$\begin{aligned}&\gamma_{kl}=\left(\frac{dF}{dq}\right)_K\Big/\left(\frac{dF}{dq}\right)_l\\&\left(\frac{dF}{dq}\right)_j=\frac{d}{dq}F(q(x_{K-1}+\mu_j p_{K-1})),\ j=1,\cdots,m\end{aligned}\tag{41}$$

因此，如果能计算出所有的 $\gamma_{j-1,j}$，就能由式（40）得到 ρ_K.

我们试从文献[3]导出的 γ_{j-1j}应满足的二次方程出发计算 γ_{j-1j}. 鉴于有时不能由此方程唯一地确定 γ_{j-1j}，所以把相邻的两个值 γ_{j-1j}和 γ_{jj+1}联系起来讨论，即设法同时求出它们来. 下列定理表明，利用 γ_{jj-1}和 γ_{jj+1}分别满足的方程

$$\psi'_{j-1}\rho^2+\left[\psi'_{j-1}+\psi'_j+\frac{4(\psi_j-\psi_{j-1})}{\mu_{j-1}-\mu_j}\right]\rho+\psi'_j=0 \tag{42}$$

和

$$\psi'_{j+1}\rho^2+\left[\psi'_{j+1}+\psi'_j+\frac{4(\psi_j-\psi_{j+1})}{\mu_{j+1}-\mu_j}\right]\rho+\psi'_j=0 \tag{43}$$

借助于函数

$$s(u_1 v)=\left|\left(\frac{1-u}{\mu_{j-1}-\mu_j}\right)\Big/\frac{1-v}{\mu_{j+1}-\mu_j}-1\right| \tag{44}$$

就能确定 γ_{i-1i}和γ_{jj+1}.

定理3. 设目标函数如式（3）所示，并设当 q 在 $q(x_{K-1}+\mu_{j-1}p_{K-1})$，$q(x_{K-1}+\mu_j p_{K-1})$ 和 $q(x_{K-1}+\mu_{j+1}p_{K-1})$所在区间取值时，$F(q)$是 q 的二次多项式. 设方程（42）的根为 $\bar\rho_{jj-1}$和 $\tilde\rho_{jj-1}$，方程（43）的根为 $\bar\rho_{jj+1}$和 $\tilde\rho_{jj+1}$. 若这些根中有一个为1，则 $\gamma_{j-1j}=\gamma_{jj+1}=1$；否则比较 $s(\bar\rho_{jj-1},\ \bar\rho^{-1}_{jj+1})$，$s(\bar\rho_{jj-1},\ \tilde\rho^{-1}_{jj+1})$，$s(\tilde\rho_{jj-1},\ \bar\rho^{-1}_{jj+1})$和$s(\tilde\rho_{jj-1},\ \tilde\rho^{-1}_{jj+1})$，记其最小值者相应的自变量为$(\theta_{jj-1},\ \theta_{jj+1})$；当 $\theta_{jj-1}=\bar\rho_{jj-1}$时，$\gamma_{j-1j}=\tilde\rho^{-1}_{jj-1}$；当 $\theta_{jj-1}=\tilde\rho_{jj-1}$时，$\gamma_{j-1j}=\bar\rho^{-1}_{jj-1}$；当 $\theta_{jj+1}=\bar\rho^{-1}_{jj+1}$时，$\gamma_{jj+1}=\tilde\rho_{jj+1}$；当 $\theta_{jj+1}=\tilde\rho^{-1}_{jj+1}$时，$\gamma_{jj+1}=\bar\rho_{jj+1}$.

证明：用文献[3]的方法可以证明，“方程（42）或（43）有其值为1的解”与“$e_2=0$”等价，所以前者成立时必有 $\gamma_{jj-1j}=\gamma_{jj+1}=1$. 而且，据此还可断言，以下证明定理其余部份时可以假定$e_2\neq 0$.

考虑二次方程

$$\psi'(\lambda)\rho^2+\left[\psi'(\lambda)+\psi'_j+\frac{4(\psi_j-\psi(\lambda))}{\lambda-\mu_j}\right]\rho+\psi'_j=0 \tag{45}$$

由文献[3]知

$$\gamma(\lambda)=\frac{d}{dq}F(q(x_{K-1}+\mu_j p_{K-1}))\Big/\frac{d}{dq}F(q(x_{K-1}+\lambda p_{K-1})\text{和}\tau(\lambda)=\frac{\mu^*}{\mu^*-\lambda}$$

是方程（45）的两个根，特别地，$\gamma(\mu_{j-1})=\gamma_{jj-1}$和$\tau(\mu_{j-1})=\mu^*/(\mu^*-\mu_{j-1})$是方程（42）的两个根，它们与$\bar\rho_{jj-1}$和$\tilde\rho_{jj-1}$对应. 而$\gamma(\mu_{j+1})=\gamma_{jj+1}$和$\tau(\mu_{j+1})=\mu^*/(\mu^*-\mu_{j+1})$是方程（43）的两个根，它们与$\bar\rho_{jj+1}$和$\tilde\rho_{jj+1}$相对应. 现在的问题是找出这些对应量之间的具体对应关系. 考虑函数$y=[\tau(\lambda)]^{-1}=1-\dfrac{\lambda}{\mu^*}$，它的图象是一条直线，而且点$(\mu_j,1)$，$\left(\mu_{j-1},\ 1-\dfrac{\mu_{j-1}}{\mu^*}\right)$和$\left(\mu_{j+1},\ 1-\dfrac{\mu_{j+1}}{\mu^*}\right)$都在该直线上. 另外考虑函数 $y=[\gamma(\lambda)]^{-1}$，由

$$\frac{1}{\gamma(\lambda)}=\frac{\varepsilon_1+\varepsilon_2 q(x_{K-1}+\lambda p_{K-1})}{\varepsilon_1+\varepsilon_2 q(x_{K-1}+\mu_j p_{K-1})}=\frac{\varepsilon_1+\varepsilon_2[a(\lambda-\mu^*)^2+b]}{\varepsilon_1+\varepsilon_2 q(x_{K-1}+\mu_j p_{K-1})}$$

可见，其图象是一条二次曲线，而且点$(\mu_i, 1)$，$(\mu_{j-1}, \gamma_{j j-1}^{-1})$和$(\mu_{j+1}, \gamma_{j j+1}^{-1})$都在该曲线上。由此不难推知

$$s\left(1-\frac{\mu_{j-1}}{\mu^*}, 1-\frac{\mu_{j+1}}{\mu^*}\right)=0$$

而当$1-\frac{\mu_{j-1}}{\mu^*}\neq\gamma_{j j-1}^{-1}, 1-\frac{\mu_{j+1}}{\mu^*}\neq\gamma_{j j+1}^{-1}$时，有

$$s\left(1-\frac{\mu_{j-1}}{\mu^*}, \gamma_{j j+1}^{-1}\right)>0,\quad s\left(\gamma_{j j-1}^{-1}, 1-\frac{\mu_{j+1}}{\mu^*}\right)>0,$$

$$s(\gamma_{j j-1}^{-1}, \gamma_{j j+1}^{-1})>0$$

进而便可推得定理结论．定理证毕．

根据定理3，可以建立各种形式的具有不变性的算法，例如与改进PRP算法Ⅰ对应，可得下列算法．

改进PRP算法Ⅱ．同于改进PRP算法Ⅰ，只是其中的第4步改为

4′计算ρ_K．设前次（第$k-1$次）迭代的一维搜索所得数据可表为

$$\psi(\mu_j)=\psi_j,\ \psi'(\mu_j)=\psi'_j,\ (j=1,\cdots,m,\ m\geqslant 3)$$

其中

$$0=\mu_1<\mu_2<\cdots<\mu_{m-1}<\mu_m$$

$$\psi'_j<0,\ (j=1,\cdots,m-1).$$

地点μ_j分为若干组

$$\mu_1<\mu_2<\mu_3,\ \mu_3<\mu_4<\mu_5,\ \cdots,\ \mu_{m-4}<\mu_{m-3}<\mu_{m-2},\ \mu_{m-2}<\mu_{m-1}<\mu_m$$

或

$\mu_1<\mu_2<\mu_3$，$\mu_3<\mu_4<\mu_5$，…，$\mu_{m-3}<\mu_{m-2}<\mu_{m-1}$，$\mu_{m-2}<\mu_{m-1}<\mu_m$，利用定理3计算出$\gamma_{j-1\,j}(j=2, \cdots, m)$用式（40）算得$\rho_K$．

该算法可以看作是对文献[2]中所提方法的改进．

参考文献

[1] Spedicato, E., A variable metric method for function minimization derived from invariancy to nonlinear scaling, J. O. T. A., 20(1976), 315—329.

[2] Boland, W. R., Kamgnia, E. R. and Kowalik, J. S., A Conjugate-gradient optimization method invariant to nonlinear scaling, J. O. T. A., 27(1979), 221—230.

[3] 邓乃扬、陈 志：不精确一维搜索条件下具有非线性尺度不变性的算法，待发表。