Chin. Ann. of Math.
8B (1) 1987

ON ALGORITHMS INVARIANT TO NONLINEAR SCALING WITH INEXACT SEARCHES

DENG NAIYANG (邓乃扬)* CHEN ZHI (陈 志)*

Abstract

Among the researches of unconstrained optimization invariancy to nnolinear scaling is an interesting subject. But the discussions which have been so far made were all under the assumption of exact line searches. Hence there are some essential deficiency in theory and practice. In this paper, using more generalized concept of invariance, the invariant algorithms not depending on the accuracy of line searches are established for the model presented by Boland et al. in [2].

Differing from most unconstrained optimization methods derived from quadratic functions, it is an interesting subject to construct the algorithms derived from nonquadratic functions, in which the algorithms invariant to nonlinear scaling have been presented[1,2]. Because of the difficulty of calculating the factors ρ_k (see below) the invariant algorithms were all researched under the assumption of exact line searches up to now. The algorithms invariant to nonlinear scaling are fruitful if exact line searches are available, otherwise they are inefficient. In order to overcome these essential deficiency in theory and practice, the algorithms invariant to nonlinear scaling, which do not depend on the accuracy of line searches, are discussed in this paper.

Definition 1. *Suppose that some algorithm is used to minimize the functions*

$$f(x)=F(q(x)), \tag{1}$$

where function $q(x)$ is fixed, F is arbitrary but $\frac{dF}{dq}>0$. If the sequences of points generated by the above algorithm for all F with the same initial conditions are identical, then this algorithm is refered to an algorithm invariant to nonlinear scaling or invariant algorithm.

Definition 2. *Suppose that a line search method is applied to minimize the functions $\bar{f}(x)$ and $\tilde{f}(x)$ with the same contours, and denote the starting points as $\bar{x}_{k-1}$ and $\tilde{x}_{k-1}$, the search directions as $\bar{p}_{k-1}$ and $\tilde{p}_{k-1}$ respectively. If the terminal points $\bar{x}_k$ and $\tilde{x}_k$ obtained for $\bar{f}(x)$ and $\tilde{f}(x)$ respectively are the same when $\bar{x}_{k-1}=\tilde{x}_{k-1}$ and $\bar{p}_{k-1}=$*

Manuscript received September 10, 1984.

* Beijing Polytechnic University, Beijing, China.

$\tilde{p}_{k-1}$, *then this line search method is referred to be consistent.*

Obviously the consistent method is not necessarily exact. When the line search is inexact, we can regard the line search as point-to-set mapping. A consistent method will be restricted in selecting the same point from the image set provided both starting point and search direction are the same. Using the concept of consistent method, we can establish the invariant algorithm which does not depend on the accuracy of the line searches. In fact, a sufficient condition for *PRP* method and Beale's method to be invariant to nonlinear scaling is that vector $y_{k-1}=g_k-g_{k-1}$ is replaced by $\hat{y}_{k-1}=\rho_k g_k-g_{k-1}$, where

$$\rho_k=\frac{d}{dq}F(q(x_{k-1}))\Big/\frac{d}{dq}F(q(x_k)). \tag{2}$$

In addition to replacing y_{k-1} by $\hat{y}_{k-1}$, $\varepsilon_i=-\lambda_i\langle g_{i+1}, p_i\rangle/\langle y_i, p_i\rangle$ is replced by $\hat{\varepsilon}_i=-\lambda_i\rho_i\langle g_{i+1}, p_i\rangle/\langle\hat{y}_i, p_i\rangle$. Nazareth's method which constructs conjugate directions without line searches can also be invariant. Therefore, it is always possible to make the algorithm be invariant to nonlinear scaling no matter whether the line searches are exact. So the problem is reduced to calculating the values of ρ_k defined by formula (2). Because the known methods of calculating ρ_k are all under the condition of exact line searches, the exact values of ρ_k can not be obtained frequently when the line searches are inexact (for example, $\rho_k\equiv1$ for $f(x)=F(q(x))=q(x)$, but the computation values of ρ_k may be different from 1). Thus the invariancy to nonlinear scaling can be destroyed. This is why the invariant algorithms are inefficient with rough line searches. Now we derive a formula of calculating ρ_k not depending on the accuracy of line searches for the model considered by Boland et al[2]. Consider an objective function defined by (1), where $F(q)$ can be expressed by

$$F(q)=\varepsilon_0+\varepsilon_1 q+\varepsilon_2 q^2\quad\left(\frac{dF}{dq}>0\right) \tag{3}$$

in some interval of q, $q(x)$ is any strictly convex quadratic function

$$q(x)=\frac{1}{2}x^TGx+r^Tx+\delta. \tag{4}$$

Suppose that the quantities

$$\begin{aligned}&f_{k-1}=f(x_{k-1}),\quad \alpha_{k-1}=\langle g_{k-1}, p_{k-1}\rangle<0,\\ &x_k=x_{k-1}+\lambda_{k-1}p_{k-1},\quad f_k=f(x_k)<f_{k-1},\quad \alpha_k=\langle g_k, p_{k-1}\rangle\end{aligned} \tag{5}$$

are known. Let us try to calculate ρ_k according to the data above.

Theorem 1. *Consider any function defined by* (1), *where* $q(x)$ *is defined by* (4), *and moreover* $F(q)$ *can be expressed by* (3) *when* $q\in[q(x_k), q(x_{k-1})]$. *The quadratic equation*

$$\alpha_k\rho^2+\left[\alpha_k+\alpha_{k-1}+\frac{4(f_{k-1}-f_k)}{\lambda_{k-1}}\right]\rho+\alpha_{k-1}=0 \tag{6}$$

is constructed according to the data in (5). *If* $\alpha_k\neq0$, *then* ρ_k *is one root of equation* (6) *and the other root is* $Z=\mu^*/(\mu^*-\lambda_{k-1})$, *where* μ^* *is the minimizer of* $\varphi(\lambda)=q(x_{k-1}+\lambda p_{k-1})$.

Proof At first we prove that ρ_k is the root of equation (6). According to (3), we have

$$f_k-f_{k-1}=(q(x_k)-q(x_{k-1}))\Big[\frac{d}{dq}F(q(x_{k-1}))+\varepsilon_2(q(x_k)-q(x_{k-1}))\Big], \tag{7}$$

$$\varepsilon_2(q(x_k)-q(x_{k-1}))=\frac{1}{2}\Big[\frac{d}{dq}F(q(x_k))-\frac{d}{dq}F(q(x_{k-1}))\Big]. \tag{8}$$

Additionally, $\varphi(\lambda)$ is strictly convex and quadratic, so that

$$q(x_k)-q(x_{k-1})=\frac{1}{2}\lambda_{k-1}\Big\{\Big[\frac{d}{dq}F(q(x_{k-1}))\Big]^{-1}\langle g_{k-1}, p_{k-1}\rangle +\Big[\frac{d}{dq}F(q(x_k))\Big]^{-1}\langle g_k, p_{k-1}\rangle\Big\}. \tag{9}$$

Substituting (8) and (9) into (7), we have

$$f_k-f_{k-1}=\frac{1}{4}\lambda_{k-1}\Big\{\Big[\frac{d}{dq}F(q(x_{k-1}))\Big]^{-1}\langle g_{k-1}, p_{k-1}\rangle+\Big[\frac{d}{dq}F(q(x_k))\Big]^{-1}\langle g_k, p_{k-1}\rangle\Big\} \cdot\Big[\frac{d}{dq}F(q(x_k))+\frac{d}{dq}F(q(x_{k-1}))\Big]. \tag{10}$$

It is easy to see that ρ_k satisfies equation (6), that is, ρ_k is a root of this equation.

Now we show that the other root of equation (6) is $Z=\mu^*/(\mu^*-\lambda_{k-1})$. Rewriting equation (6) into

$$4\frac{f_{k-1}-f_k}{\lambda_{k-1}}+\alpha_k+\alpha_{k-1}=-\alpha_{k-1}\frac{1}{\rho_k}-\alpha_k\rho_k \tag{11}$$

and using

$$q(\tilde{x}_k)=q(x_k),\quad f(\tilde{x}_k)=f_k,\quad \langle g(\tilde{x}_k), p_{k-1}\rangle=-\alpha_k, \tag{12}$$

where $\tilde{x}_k=x_{k-1}+(2\mu^*-\lambda_{k-1})p_{k-1}$, we obtain

$$4\frac{f_{k-1}-f_k}{\lambda_{k-1}}u-\alpha_k+\alpha_{k-1}=-\alpha_{k-1}\frac{1}{\rho_k}+\alpha_k\rho_k, \tag{13}$$

where

$$u=\Big(2\frac{\mu^*}{\lambda_{k-1}}-1\Big)^{-1}. \tag{14}$$

According to (11), (13) and (14)

$$Z=\mu^*/(\mu^*-\lambda_{k-1})=(1+u)/(1-u)=\frac{\alpha_{k-1}}{\alpha_k}\cdot\frac{1}{\rho_k}$$

is the other root of equation (6).

The next problem is to discuss how to choose ρ_k from the two roots of equation (6). We will first ask whether ρ_k can be determined uniquely by the data (5). The answer is negative from the following example.

Example. Consider the family of one dimensional objective functions

$$f(x)=F(q(x)), \tag{15}$$

where

$$F(q)=\varepsilon_1 q+\varepsilon_2 q^2 \quad \left(\frac{dF}{dq}>0\right), \quad q(x)=a(x-\mu^*)^2+b \quad (a>0). \tag{16}$$

Suppose that the data in (5) are given by

$$\begin{aligned} &x_{k-1}=-2, \quad p_{k-1}=1, \quad x_k=1, \quad \lambda_{k-1}=1, \\ &f_{k-1}=20, \quad \alpha_{k-1}=-36, \quad f_k=2, \quad \alpha_k=-6. \end{aligned} \tag{17}$$

Let

$$\bar{\mu}^*=0, \quad \bar{a}=1, \quad \bar{b}=0, \quad \bar{\varepsilon}_1=1, \quad \bar{\varepsilon}_2=1,$$

$$\tilde{\mu}^*=-\frac{1}{2}, \quad \tilde{a}=1, \quad \tilde{b}=\frac{7}{4}-\frac{2}{3}\sqrt{6}, \quad \tilde{\varepsilon}_1=2\sqrt{6}, \quad \tilde{\varepsilon}_2=\frac{3}{2}$$

in (16), It then follows that

$$\bar{f}(x)=x^4+x^2, \tag{18}$$

$$\tilde{f}(x)=\frac{3}{2}\left(x+\frac{1}{2}\right)^4+\frac{21}{4}\left(x+\frac{1}{2}\right)^2+\frac{19}{32}. \tag{19}$$

Both the functions satisfy the condition (17), but

$$\bar{\rho}_k=3\neq 2=\tilde{\rho}_k.$$

The following theorem indicates the case in which ρ_k can be determined uniquely.

Theorem 2. *If $\alpha_k=0$, then ρ_k can be determined uniquely from equation* (6) *as*

$$\rho_k=-\alpha_{k-1}/[\alpha_{k-1}+4(f_{k-1}-f_k)/\lambda_{k-1}].$$

If $\alpha_k>0$, then the two roots of equation (6) *have different sign, the positive root is ρ_k; If $\alpha_k<0$, then at least one root of equation* (6) *is greater than* 1 *and the other root in case which is not greater than* 1 *is ρ_k.*

Proof If $\alpha_k=0$, the conclusion holds obviously. From this, it is very easy to obtain the main result of [2]. For other cases, the conclusion can also be proved by Theorem 1, according to the relationship between the two roots and the relationship between ρ_k and μ^*. The proof is omitted.

Theorem 2 indicates that ρ_k cannot be determined uniquely only when the two roots of equation (6) are all greater than 1. In this case we can use other information of objective function to determine ρ_k, such as a value of objective function at another point $\hat{x}=x_{k-1}+\mu p_{k-1}$ $(\mu>0)$,

$$\hat{f}=f(\hat{x})=f(x_{k-1}+\mu p_{k-1}) \quad (\mu\neq\lambda_{k-1}), \tag{20}$$

which may have been calculated in the process of line search. ρ_k can be determined by the data in (5) and (20). In fact, suppose that the objective function is defined by (1), (3) and (4). We can prove that $f(x_{k-1}+\lambda p_{k-1})$ can be expressed by

$$f(\lambda, \mu^*)=A(\mu^*)(\lambda-\mu^*)^4+B(\mu^*)(\lambda-\mu^*)^2+C(\mu^*), \tag{21}$$

where μ^* is the minimizer of $\varphi(\lambda)=q(x_{k-1}+\lambda p_{k-1})$ and

$$A(\mu^*)=\frac{\alpha_{k-1}(\lambda_{k-1}-\mu^*)+\mu^*\alpha_k}{-4\mu^*\lambda_{k-1}(\lambda_{k-1}-\mu^*)(2\mu^*-\lambda_{k-1})}, \tag{22}$$

$$B(\mu^*)=\frac{\mu^{*3}\alpha_k+(\lambda_{k-1}-\mu^*)^3\alpha_{k-1}}{2\mu^*\lambda_{k-1}(\lambda_{k-1}-\mu^*)(2\mu^*-\lambda_{k-1})}, \tag{23}$$

$$C(\mu^*)=f_{k-1}-A(\mu^*)\mu^{*4}-B(\mu^*)\mu^{*2}. \tag{24}$$

Therefore, according to Theorem 1, if we denote the two roots of equation (6) as $\bar{\rho}$ and $\tilde{\rho}$, the minimizer of $\varphi(\lambda)=q(x_{k-1}+\lambda p_{k-1})$ is either $\lambda_{k-1}\bar{\rho}/(\bar{\rho}-1)$ or $\lambda_{k-1}\tilde{\rho}/(\tilde{\rho}-1)$. In the light of above discussion, it is not difficult to establish a computational method defining ρ_k.

As another more general approach, ρ_k can be defined by the additional gradient of objective function at another point $\hat{x}=x_{k-1}+\mu p_{k+1}$ $(\mu>0)$,

$$\hat{g}=\nabla f(\hat{x}). \tag{25}$$

In fact, we have

$$\rho_k=\frac{(1-\lambda)\langle g_{k-1}, h\rangle}{\langle g_k, h\rangle}, \tag{26}$$

where

$$\lambda=\|x_k-x_{k-1}\|/\|\hat{x}-x_{k-1}\|, \tag{27}$$

$$h=g_{k-1}-\langle g_{k-1}, \hat{g}\rangle\hat{g}/\|\hat{g}\|^2. \tag{28}$$

Thus we can construct algorithms invariant to nonlinear scaling, no matter whether the line searches are exact or inexact. To do this, it is sufficient to replace $y_{k-1}=g_k-g_{k-1}$ with $\hat{y}_{k-1}=\rho_k g_k-g_{k-1}$, where ρ_k is defined acording to above discussion. As an example, the following conjugate gradient algorithm is briefly described.

Algorithm

Step 1. Select an x_1 and set $k=1$.

Step 2. Compute $f_k=f(x_k)$ and $g_k=\nabla f(x_k)$. If $g_k=0$, stop.

Step 3. If $k=1$, set $\beta_{k-1}=0$, go to Step 11.

Step 4. Compute $\alpha_{k-1}=\langle g_{k-1}, p_{k-1}\rangle$ and $\alpha_k=\langle g_k, p_{k-1}\rangle$.

Step 5. If $\alpha_k=0$, set ρ_k equals the unique root of (6), go to Step 10.

Step 6. Compute two roots $\bar{\rho}$ and $\tilde{\rho}$ of (6).

Step 7. If $\alpha_k>0$, set $\rho_k=\max\{\bar{\rho}, \tilde{\rho}\}$, go to Step 10.

Step 8. If $\min\{\bar{\rho}, \tilde{\rho}\}\leqslant 1$, set $\rho_k=\mathrm{mid}\{\bar{\rho}, \tilde{\rho}\}$, go to Step 10.

Step 9. Compute ρ_k according to (26).

Step 10. Compute $\beta_{k-1}=\langle g_k, \rho_k g_k-g_{k-1}\rangle/\|g_{k-1}\|^2$.

Step 11. Set $p_k=-g_k+\beta_{k-1}p_{k-1}$.

Step 12. Starting from x_k, line search is executed along p_k and x_{k+1} is obtained.

Step 13. Set $k=k+1$, go to Step 2.

References

[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] Nazareth, L., A conjugate gradient algorithm without line searches, *J. O. T. A.*, **23**(1977), 373—387.