Investigation on Self-Scaling of Mixed Interior-Exterior Method for Non-Linear Optimization

Ghada M. Al-Naemi
College of Computer Sciences and Mathematics
University of Mosul, Iraq

Received on: 22/05/2005 **Accepted on: 29/06/2005**

ABSTRACT

In this paper, we have investigated self-scaling sequential unconstrained minimization techniques (SUMT). Our new modified version on CG-method and QN- method shows that it is too effective when compared with other established algorithms to solve standard constrained optimization problems.

Keywords: unconstrained optimization, self-scaling technique.

التقصي في التقييس الذاتي للربط الداخلي والخارجي لمسائل الامثلية غير الخطية

غادة مؤيد النعيمي
كلية علوم الحاسوب والرياضيات، جامعة الموصل

تاريخ الاستلام: 2005/05/22 **تاريخ القبول: 2005/06/29**

الملخص

في هذا البحث تم استحداث تقنية التقييس الذاتي لمعالجة المسائل العليلة التي قد تحدث في الأمثلية غير الخطية المشروطة و الغرض من الخوارزمية المقترحة هو العمل على التخلص من أخطاء البتر والتدوير، و إن التقنية الجديدة تمت برمجتها لحل بعض مسائل ألا مثلية القياسية المعروفة و توصلنا إلى أن نتائجها اكثر كفاية من الطرائق السابقة في هذا المجال.

الكلمات المفتاحية: الأمثلية غير الخطية، تقنية التقييس الذاتي.

1- General Introduction to Nonlinear Constrained

The general constrained minimization problem

$$\begin{aligned} &\text{minimize } f(x) \\ &\text{subject to } \begin{array}{ll} c_i(x) \le 0 & i = 1...m \\ h_i(x) = 0 & i = m+1...L \end{array} \end{aligned} \qquad \text{…….(1)}$$

where x is an n-dimensional vector and the functions $f(x)$, $c_i(x)$, $i = 1...m$ and $h_i(x) = 0$, $i = m+1,...,l$ are continuous and usually as-summed to possess continuous second partial derivatives. The constraints in eq.(1) are referred to as functional constraints.

There are basically two different kinds of constrained optimization approaches:

Indirect Method: changes the constrained optimization into unconstrained optimization to be solved.

(*S*equential *U*nconstrained *M*inimization *T*echnique, SUMT)

Direct Method: deals with the constraints directly in the search for the Optimum. (Kwon,2001)

2-Sequential Unconstrained Minimization Techniques (SUMT):

Main idea:

* Solve a constrained optimization problem by solving a sequence of unconstrained optimization problems, and in the limit, the solutions of the unconstrained problems will converge to the solution of the constrained problem.

* Use an auxiliary function that incorporates the objective function together with "penalty" terms that measure violations of the constraints. INT[2]

3-Classical SUMT:

Two groups of classical methods:

Barrier methods: impose a penalty for reaching the boundary of an inequality constraint.

Penalty methods: impose a penalty for violating a constraint.

4-Exterior Point Methods (Penalty function):

Definition: A function $p(x): R^n \to R$ is called a penalty function for eq.(1) satisfies

1- $p(x) = 0$ if $c(x) \le 0$, $\mathrm{h}(\mathrm{x}) = 0$ and

2- $p(x) \succ 0$ if $c(x) \le 0$ or $\mathrm{h}(\mathrm{x}) \ne 0$

Penalty function are typically defined by

$$p(x) = \sum_{i=1}^{m} \phi(c_i(x)) + \sum_{i=m+1}^{l} \varphi(h_i(x))$$

Where

$\phi(y) = 0$ if $y \le 0$ and $\varphi(y) \succ 0$ if $y \succ 0$

$\phi(y) = 0$ if $y = 0$ and $\varphi(y) \succ 0$ if $y \ne 0$

4-1 General Type of Penalty Function Methods

There are several types of penalty function method with the inequality constrained which has the following two terms:

1- $\phi(c_i(x)) = [\min(0, c_i(x))]^2$ (quadratic loss function)

2- $\phi(c_i(x)) = [\min(0, c_i(x))]$ (Zangwills,(1967) loss function) or with the equality constraint which has the following two forms

1- $\varphi(h_i(x)) = (h_i(x))^2$

2- $\varphi(h_i(x)) = |h_i(x)|$

Hence ,our objective function may be defined by

$$\theta(x_k, \mu_k) = f(x_k) + \mu \sum_{i=1}^{m} \phi[g_i(x)] + \frac{1}{\mu} \sum_{i=m+1}^{l} \varphi[h_i(x)]$$

5- Interior Point Methods (Barrier Function):

Definition: A barrier function for eq(1) is any function $B(x): R^n \to R$ that satisfies

- $B(x) \succ 0$ for all x that satisfy $c(x) \succ 0$

- $B(x) \to \infty$ as $\lim_{x} \max\{c_i(x)\} \to 0$

The idea in a barrier method is to dissuade points x from ever approaching the boundary of the feasible region. We consider solving

$$\theta(x_k, \mu_k) = \min \ f(x_k) + \mu_k B(x_k)$$
$$s.t. \quad c(x_k) > 0$$
$$x_k \in R^n$$

For a sequence of $\mu_k \to 0$. Note that the constraints $c(x_k) > 0$ are effectively unimportant in $\theta(\mathrm{x}_k, \mu_k)$, as they are never binding in $\theta(\mathrm{x}_k, \mu_k)$. INT[1]

5-1 General Types of Barrier Function Method:

There are several types of Barrier function method

1- $B(x) = \sum_{j=1}^{m} \frac{1}{c_j(x)}$

2- $B(x) = \sum_{j=1}^{m} \frac{1}{c_j(x)} \in, \quad \in > 0$ ……………………(Toint etal., 1997)

3- $B(x) = -\sum_{j=1}^{m} \ln(c_j(x))$

6-Mixed Exterior-Interior Methods:

we consider some method, which can be used to solve a general class (equality and inequality of problem) thus, the new problem can be converted into an unconstrained minimization problem by constructing a function of the form. (Fiacco & Mc Cormick, 1968a, 1968b)

$$\theta(x_k,\mu_k) = f(x_k) + \mu\sum_{i=1}^{m}\phi[c_i(x)] + \frac{1}{\mu}\sum_{i=m+1}^{l}\varphi[h_i(x)] \qquad (2)$$

Although both exterior and interior-point methods have many points of similarity, they represent two different points of view. In an exterior-point procedure, we start from an infeasible point and gradually approach feasibility. While doing so, we move away from the unconstrained optimum of the objective function. In an interior-point procedure, we start at a feasible point and gradually improve our objective function, while maintaining feasibility. The requirement that we begin at a feasible point and remain within the interior of the feasible inequality constrained region is the chief difficulty with interior-point methods. In many problems we have no easy way to determine a feasible starting point, and a separate initial computation may be needed. Also, if equality constraints are present, we do not have a feasible inequality constrained region in which to maneuver freely. Thus interior-point methods cannot handle equalities.

We many readily handle equalities by using a "mixed" method in which we use interior-point penalty functions for inequality constraints only. Thus, if the first *m* constraints are inequalities and constraints *(m+1)* to *n* are equalities, our problem becomes:

Minimize $$\theta(x,\mu_k) = f(x) + \mu_k B(x_k) + \frac{1}{\mu_k} p(x_k) \qquad (3)$$

The function $\theta(x,\mu)$ is then minimized for a sequence of monotonically decreasing $\mu \succ 0$.

We can solve the constrained problem given in eq.(1) construct a new objective function $\theta(x,\mu_k)$ which is defined in eq.(2). Now our aim is to minimize the function $\theta(x,\mu_k)$ by starting form a feasible point x_0 and with initial value $\mu_0 = 1$ and the method reducing μ_k is simple iterative method such :

$$\mu_{k+1} = \frac{\mu_k}{\hat{\omega}}, \qquad (4)$$

Where $\hat{\omega}$ is a constant equal to 10 and the search direction d_k in this case can be defined

$$d_k = -H_k g_k, \qquad (5)$$

Where H is a positive definite symmetric approximation matrix to the inverse Hessian matrix G^{-1} and g is the gradient vector of the function $\theta(x_k,\mu_k)$. The next iteration is set to further point

$$x_{k+1} = x_k + \lambda_k d_k \text{ , } \ldots\ldots(6)$$

where λ is a scalar chosen in such that $f_{k+1} < f_k$. We thus test $c_i(x_{k+1})$ to see that it is positive for all *i*. We find a feasible x_{k+1} and we can then proceed with the interpolation. Then a correction matrix to get updates the matrix H_k

$$H_{k+1} = H_k + \phi_k \quad \ldots\ldots (7)$$

where ϕ_k is a correction matrix which satisfies quasi-Newton condition namely $(H_{k+1} y_k = \sigma v_k)$ where v_k and y_k are difference vector between two successive points and gradients respectively and σ is any scalar.

The initial matrix H_0 chosen to be identity matrix *I*. H_k is updated through the formula of BFGS update. (Fletcher, 1970)

$$H_{k+1} = H_k^{(1)} + H_k^{(2)} \quad \ldots\ldots(8)$$

where

$$H_k^{(1)} = H_k - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} + ww^T \quad \ldots\ldots(9)$$

$$H_k^{(2)} = \frac{v_k v_k^T}{v_k^T y_k} \quad \ldots\ldots(10)$$

and

$$w = (y_k^T H_k y_k)^{0.5} \left(\frac{v_k}{v_k^T y_k} - \frac{H_k y_k}{y_k^T H_k y_k}\right) \ldots\ldots(11)$$

And terminate of the method if

$$|x_i - x_{i-1}| < \varepsilon \quad \ldots\ldots(12)$$

where $\varepsilon = 0.000001$, and

$$\mu_{k+1} = \frac{\mu_k}{10} \quad \ldots\ldots(13)$$

6-1 General Type of Mixed Interior and Exterior Point:
Methods

1- $\theta(x, \mu_k) = f(x) + \mu_k B(x_k) + \frac{1}{sqrt(\mu_k)} p(x_k)$ **……………….(Bigg,1983)**

2- $\theta(x, \mu_k) = f(x) + \mu_k B(x_k) + \frac{1}{\mu_k} p(x_k)$ **……………….(Rao,1994)**

3- $\theta(x,\mu_k) = f(x) - \mu_k \overline{B}(x_k) + \frac{1}{\mu_k} p(x_k)$ ……………**(Gettfried,1973)**

4- $\theta(x,\mu_k) = f(x) + \mu_k \hat{B}(x_k) + \frac{1}{\mu_k} p(x_k) \quad \in \succ 0$ …………..**(Toint,etal,1997)**

Where

$B(x)$: -Inverse Barrier function which handles the inequality.

$\hat{B}(x)$:- Inverse Barrier function which handles the equality.

$\overline{B}(x)$: -Log Barrier function which handles the inequality.

$p(x)$: -Penalty function which handles the equality.

6-2 Outline Mixed Interior-Exterior Point Methods:

Step1: Find an initial approximation x_0 in the interior of the feasible region for the inequality constraints i.e. $c_i(x_0)>0$.

Step2: Set $i=1$ and $\mu_0 = 1$ is the initial value of μ_0.

Step3: Set $d_i = -H_i g_i$

Step5: Set $x_{i+1} = x_i + \lambda_i d_i$, where λ is a scalar.

Step6: Update H by correction matrix defined in eq.(7)-(11).

Step7: Check for convergence i.e. if eq.(12) is satisfied then stop.

Step8: Otherwise, set $\mu_{i+1} = \frac{\mu_i}{10}$ and take $x=x^*$ and set $k=k+1$ and go to step5.

7- New Self-Scaling Variable Metric Methods:

In order to eliminate the truncation and rounding errors, the new scalar parameter σ is added to make the sequence and efficiency as problem dimension increase. The poor-scaling is an imbalance between the values of the function and change in *x*. The function values may be changed very little even though *x* is changing significantly. This difficulty can sometimes be removed by good scaling factor for the updating *H* and the performance of self-scaling methods is undoubtedly favorable in some cases especially when the number variables are large (Scales, 1985).

An idea is multiplying part of *BFGS* by scaling factor $\overline{\sigma}$ before the update takes place. The original motivation for self-scaling method arises from the analysis of quadratic objective function, and the main results also assume that exact line searches are performed.

Many authors have proposed a special scaling as follows:

1- $\bar{\sigma}_{1k} = \dfrac{v_k^T y_k}{4v_k^T g_{k+1} + 2v_k^T g_k - 6(f_{k+1} - f_k)}$ (Bigg ,1973)……………(14)

2- $\bar{\sigma}_{k2} = \dfrac{v_k^T y_k}{y_k^T H_k y_k}$ (Oren,1974)…………….. (15)

3- $\bar{\sigma}_{k3} = \dfrac{y_k^T H_k y_k}{v_k^T y_k}$ (Al-Bayati,1991) ………………….(16)

4- $\bar{\sigma}_{k4} = \dfrac{v_k^T y_k}{2v_k^T g_k - 6(f_{k+1} - f_k)}$ (Al-Assady,1991) ……………...(17)

In This paper, we have suggested a new parameter say:

$$\bar{\sigma}_{new} = (1 - \frac{v_i^T g_i}{y_i^T H y_i})^m$$

Where
m= number of constrained

From the above, one can suggeste a more general family of the form:

$H^*_{New1} = \gamma H^1 + \bar{\sigma}_{new} H^2$ ……………………………………(18)

Which satisfied the QN-like condition

$H^*_{New1} y = \bar{\sigma}_{new} v$ …………………..…………………………………...(19)

In fact, this relaxation of the QN-condition is of particular interest in deriving algorithm for non-quadratic objective function. Several choices of $\gamma, \bar{\sigma}_{new}$ have been investigated, but the most effective one (in our numerical computation and for the class of the constrained optimization problem) presented here is readily interpreted in term of the earlier algorithms and their property. We defined

$H^*_{New1} = H^1 + (1 - \dfrac{v^T g}{y^T H y})^m H^2$ …………………………………...(20)

Comparing with

$$\gamma = 1 \qquad \bar{\sigma}_{\text{new}} = (1 - \frac{v^T g}{y^T H y})^m$$

The above suggestion will be true if we prove that:

$H^*_{New1} y = \bar{\sigma}_{new} v$ is true i.e. $$H^*_{New1}y=(H^1+\overline{\sigma}_{new}H^2)y$$

$$H^*_{New1}y=H^1y+\overline{\sigma}_{new}H^2y$$

$$Hy=Hy-\frac{H\ y\ y^THy}{y^TH\ y}+ww^T\ +\overline{\sigma}_{new}\ \frac{v\ v^T}{v^T y}$$

$$Hy=Hy-Hy+(y^THy)\left[\frac{v^Ty}{v^Ty}-\frac{y^TH\ y}{y^TH\ y}\right]\left[\frac{v^Ty}{v^Ty}-\frac{y^TH\ y}{y^TH\ y}\right]^T+\overline{\sigma}_{new}v=\overline{\sigma}_{new}v$$

Hence, the new formula (20)satisfied the QN-condition. Our last inquiry: Does formula (20) generates mutually conjugate gradient search direction?To answer this equation, follow this new theorem

3-1 New theorem:

The new formula (18) generates mutually conjugate gradient search direction

Proof:

Let $F(x)=(1/2)x^T\ G\,x+b^Tx$ be quadratic function. Choose an initial approximation matrix $H_i=H$. We have to prove that for ELS,the search direction d must satisfy :-

$$H_ig_{k+1}=H\ g_{k+1}\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..\ldots.(19)$$

Now, proceed by induction let $i=0$ this implies

$$H_1g_{k+1}=H\ g_{k+1}\quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..(20)$$

From (18) we have

$$H_i\ g_{k+1}=\left[H_i-\frac{H_iy_iy_i^TH_i}{y_i^TH_iy_i}+ww^T\ +\overline{\sigma}_{new}\ \frac{v_iv_i^T}{v_i^T y_i}\right]g_{k+1}\ \ldots\ldots\ldots\ldots\ldots\ldots..(21)$$

$$H_i\ g_{k+1}=H_ig_{k+1}-\left(\frac{y_i^TH_ig_{k+1}}{y_i^TH_iy_i}\right)H_iy_i+w_i^T\mathrm{g}_{k+1}w\ \ +\overline{\sigma}_{new}\frac{v_i^Tg_{k+1}}{v_i^T y_i}v_i^T$$

Assume that this property is true for i namely

$$H_ig_{k+1}=H\ g_{k+1}\qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..(22)$$

We have to prove that this property is true for $i+1$, realizing that for quadratic function, it is well-know that: -

$$\left.\begin{array}{ll} v_i^Tg_{k+1}=0 & \text{for j}=1,2,...\text{k} \\ g_j^TH_jg_{k+1}=0 & \text{for j}=1,2,...\text{k}\end{array}\right\}\ \ldots\ldots\ldots\ldots\ldots(23)$$

use (23) in (21) to get

$H_{i+1}g_{k+1} = H_i g_{k+1}$ ……………………………………(24)

Since we have: -

$y_j^T H_j g_{k+1} = 0$ for i⟨k ………………….. (25)

and

$$w^T g_{k+1} = (y_i^T H_i y_i)^{0.5}(\frac{v_i g_{k+1}}{v_i^T y_i} - \frac{(H_i y_i)^T g_{k+1}}{y_i^T H_i y_i}) = 0 \qquad \text{……………(26)}$$

Thus new formula (18) generates mutually conjugate gradient search direction because

$v_i^T g_{k+1} = 0$ ……………………………(27)

7-2 Outlines of the New Self-Scaling method:

Step 1: Find an initial approximation x_0 in the interior of the feasible region for the inequality constraints i.e. $c_i(x) \succ 0$.

Step 2: Set $i = 1$ and $\mu_0 = 1$ is the initial value of μ_0.

Step 3: Set $d_i = -H_i g_i$

Step 4: Set $x_{i+1} = x_i + \lambda_i d_i$, where λ is a scalar.

Step 5:Update H by correction matrix which is defined in eq.(8-11) where σ_{new} is defined in eq.(18)

Step 6: Check for convergence if $|x_i - x_{i-1}| \prec \in$ where $\in = 1E-5$ satisfied then stop.

Step 7: Otherwise, set $\mu_{i+1} = \frac{\mu_i}{10}$ and take x=x* and set $i = i+1$ and go to Step 5.

8- Numerical Results:

Several standard non-linear constrained test functions were minimized to compare the new algorithms with standard algorithm see (Appendix), with $1 \le n \le 10$ and $1 \le c_i(x) \le 10$ and $1 \le h_i(x) \le 10$.

All the results are obtained using Pentium 4. All programs are written in FORTRAN language and for all cases the stopping criterion taken to be

$$|x_i - x_{i-1}| < \delta, \qquad \delta = 10^{-5}$$

All the algorithms in this paper use the same ELS strategy which is the quadratic interpolation technique directly adapted from (Bunday, 1984).

The comparative performance for all of these algorithms are evaluated by considering NOF, NOI, NOG and NOC, where NOF is the number of function evaluations and NOI is the number of iterations and NOG is the number of gradient evaluations and NOC number of constrained evaluations.

In table (1) we have compared our new algorithm with the standard algorithm

Table (1)
Comparison of the BFGS algorithm with The new Self-Scaling algorithm

Test Fn.	BFGS- algorithm NOF(NOI)NOG(NOC)	Self-Scaling BFGS- algorithm NOF(NOI)NOG(NOC)
1-	2630(244)5(3)	1510(202)5(3)
2-	767(131)10(19)	712(130)10(19)
3-	109(38)7(11)	109(38)7(11)
4-	2153(263)8(13)	1717(242)8(13)
5-	749(124)10(19)	689(120)10(19)
6-	146(53)2(1)	72(35)2(1)
7-	734(123)10(19)	761(137)10(19)
8-	28345(295)15(29)	1512(189)10(29)
9-	2719(282)15(29)	2725(310)15(29)
Total	38352(1553)82(143)	9807(1368)77(143)

Appendix:

1. $\min f(x) = x_1 x_4 (x_1 + x_2 + x_3) + x_3$

s.t.

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 = 40$$
$$x_1 x_2 x_3 \geq 25$$
$$5 \geq x_i \geq 1$$

2- $\min f(x) = (x_1 - 2)^2 + \frac{1}{4} x_2^2$

s.t.

$$2x_1 + 3x_2 = 4$$
$$x_1 - \frac{7}{2} + x_2 \leq 1$$

3-. $\min f(x) = x_1^2 + x_2^2$

s.t.

$$x_1 + 2x_2 = 4$$
$$x_1^2 + x_2^2 \le 5$$
$$x_i \ge 0$$

4- $\min f(x) = x_1 x_2$

s.t.

$$25 - x_1^2 - x_2^2 = 0$$

5- $\min f(x) = (x_1 - 2)^2 + (x_2 - 1)^2$

s.t..

$$x_1 - 2x_2 = -1$$
$$\frac{-x_1^2}{4} + x_2^2 + 1 \ge 0$$

6- $\min f(x) = x_1^2 x_2$

s.t.

$$x_1 x_2 - (\frac{x_1^2}{2}) = 6$$
$$x_1 + x_2 \ge 0$$

7- $\min f(x) = (x_1 - 3)^2 + (x_2 - 2)^2$

s.t.

$$x_1 + 2x_2 = 4$$
$$x_1^2 + x_2^2 \le 5$$
$$x_i \ge 0$$

8- $\min f(x) = x_1^2 - x_1 x_2 + x_2^2$

s.t.

$$x_1^2 + x_2^2 = 4$$
$$2x_1 + x_2 \le 2$$

9- $\min f(x) = x_1^2 - x_1 x_2 + x_2^2$

s.t.

$$x_1^2 + x_2^2 = 4$$
$$2x_1 + x_2 \le 2$$

<u>REFERENCES</u>

[1] Al-Bayati, A.Y. (1991) "A New Family of Self Scaling Variable Metric Algorithms for Unconstrained Optimization" **J.Ed. and Sc.**,12

[2] Biggs, M.C. (1973) "A Note On Minimization Algorithms which Make Use of Non-Quadratic Properties of the Objective Function", **Journal of Institute of Mathematics and Its Application**, No. 12, pp. 337-338.

[3] Biggs, M.C. (1983) **Computational with Sequential Unconstrained Minimization Technique for Nonlinear Programming** Numerical Optimization Center T.R.N.

[4] Bunday, B.D.(1984) **Basic Optimization Methods**, London: Edward Arnold.

[5] Fletcher, R. (1970) " A New Approach to Variable Metric Algorithms" **The Computer Journal**, No. 13, pp. 317-322.

[6] Fiacco, A.V. and McCormick, G.P., (1968a) " Extensions of SUMT for Nonlinear Programming Nonlinear programming Equality Constraints and Extrapolation " ,**Management science**, Vol 12,No.111,pp.816-828

[7] Fiacco, A.V. and McCormick, G.P., (1968b) **Nonlinear Programming Sequential Unconstrained Minimization Techniques,** New York: Wiley

[8] Frish,K. R.(1955) "**The Logarithmic Potential Method of ConvexProgramming**" Memorandum may 13,195,University Institute of Economics Oslo.

[9] Kwon,T.H.(2001) "**Constrained optimization** " University of Western Ontario London, Ontario, Canada

[10] Oren, S.S. 1974,and Luenbereger D.G. "Self-Scaling Variable Metric Algorithm, Part II", **Management Science**, No. 20, pp. 863-874.

[11] Rao, S.S., (1994) "**Optimization Theory and Applications**" ,Wiley Eastern limited.

[12] Scales L.E.(1985) "**Introduction to Non-Linear Optimization**", London: Macillan.

[13] Toint, Ph. L and Nicholas I.M. (1997) "**Anote on The Second – Order Convergence of Optimization Algorithm by using Barrier Function**" : A current survey, Report 94/1, IBM T. J. Watson Research Center, U.S.A.

[14] Zangwill, w G.(1967)"**Non-linear Programming Via Penalty Function**" Mgnt csi.13 pp.344-358

[15] **Continuous Optimization Method Introduction to Penalty and Barrier Method for Unconstrained Optimization** (2001), http:// www. personal .engin. umich. edu/ ~mepelman /teaching /IOE11 /section 9.pdf

[16] "**Penalty and Barrier Method**"(2001) http://www .cityu.edu. hk /ma/staff/zhang/MA66-11-16.pdf.