A Short Note on the Q-linear Convergence of the Steepest Descent Method*

Ya-xiang Yuan
State Key Laboratory of Scientific/Engineering Computing,
Institute of Computational Mathematics and Scientific/Engineering Computing,
The Academy of Mathematics and Systems Science,
Chinese Academy of Sciences, P.O.Box 2719, Beijing, 100080, P.R.China,
E-mail: yyx@lsec.cc.ac.cn

Abstract

This paper studies the Q-linear convergence properties of the steepest descent method. For strictly convex quadratic objective functions, we give a very accurate estimate to the Q-linear convergence rate of the steepest descent method with exact line searches.

Keywords: steepest descent, exact line search, q-linear convergence.

1 Introduction

The classic steepest descent method of Cauchy (1847) (see Curry, 1944) for unconstrained optimization

$$\min_{x \in R^n} f(x), \tag{1.1}$$

defines the iterations by

$$x_{k+1} = x_k - \alpha_k^* g_k, \tag{1.2}$$

where $g_k = \nabla f(x_k)$ is the gradient of the objective function and $\alpha_k^* > 0$ is a step-size satisfying

$$f(x_k - \alpha_k^* g_k) = \min_{\alpha > 0} f(x_k - \alpha g_k). \tag{1.3}$$

If $f(x)$ is a strictly convex quadratic function

$$f(x) = g^T x + \frac{1}{2} x^T H x, \tag{1.4}$$

where H is a symmetric positive definite matrix, it can be shown that the objective function value $f(x_k)$ converges Q-linearly to $f(x^*)$, where $x^* = -H^{-1} g$ is the unique minimizer of $f(x)$. In fact, the following inequality

$$\frac{f(x_{k+1}) - f(x^*)}{f(x_k) - f(x^*)} \le \left(\frac{\kappa - 1}{\kappa + 1} \right)^2 < 1 \tag{1.5}$$

*this work is partially supported by Chinese NSF grant 10231060 and by the Knowledge Innovation program of CAS

always holds, where κ is the condition number of H defined by

$$\kappa = \frac{\lambda_1(H)}{\lambda_n(H)} \tag{1.6}$$

with $\lambda_1(H)$ and $\lambda_n(H)$ being the largest and smallest eigenvalues of H respectively. The proof of inequality (1.5) can be found in many references, for example, see Akaike(1959), Greenstadt (1967), Forsythe(1968), Luenberger(1984), and Sun and Yuan(2006). Let us define the H-norm as follows,

$$\|v\|_H = \sqrt{v^T H v}\ , \qquad \forall v \in \Re^n\ . \tag{1.7}$$

A direct consequence of (1.5) is that the H-norm of the error vector $x_k - x^*$ converges Q-linearly:

$$\frac{\|x_{k+1} - x^*\|_H}{\|x_k - x^*\|_H} \leq \frac{\kappa - 1}{\kappa + 1} < 1\ . \tag{1.8}$$

Consider a simple example in $\Re^2$ with $g = 0$ and

$$H = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}, \tag{1.9}$$

and

$$x_1 = \begin{pmatrix} \frac{1}{\lambda_1} \\ \frac{1}{\lambda_2} \end{pmatrix}, \tag{1.10}$$

and $\lambda_1 > \lambda_2 > 0$. In this example, the iterate points zigzag very slowly, particularly when $\lambda_1 >> \lambda_2$. Though this example is a special problem in 2-dimension, it draws the general picture of the steepest descent method for all n. Akaike (1959) proved that the iterates x_k converge to the solution by asymptotically alternating between two directions - the "cage" of Stiefel (1952), unless the first search direction is an eigenvector of the Hessian H. Furthermore, Akaike (1959) showed that the two asymptotic directions are in a two-dimensional subspace spanned by two eigenvectors of H. Therefore, the steepest descent method converges only linearly and can be very slow if there is a very large ratio between the two eigenvalues whose corresponding eigenvectors span the two dimensional subspace containing the two asymptotic directions.

In a practical implementation, instead of exact line search (1.3), we can compute α_k by some line search conditions, such as Goldstein conditions or Wolfe conditions (see Fletcher, 1987). It is easy to show that the steepest descent method with such conditions is always convergent. That is, theoretically the method will not terminate unless a stationary point is found. However, as the exact line search step-size α_k^* normally satisfies such inexact line search conditions, we can see the zigzag phenomenon will also happen.

A surprising result was given by Barzilai and Borwein (1988), which presented formulae for the step-size α_k which lead to superlinear convergence if the objective function is a convex quadratic function of two variables. The result of Barzilai and Borwein (1988) has triggered off many researches on the gradient method. For example, see Dai (2001), Dai and Fletcher (2003), Dai et al. (2002), Dai and Yuan (2003,2005), Dai and Zhang (2001), Fletcher (2001), Friedlander et al. (1999), Nocedal et al. (2000), Raydan (1993, 1997), Raydan and Svaiter(2002), Vrahatis et al. (2000), and Yuan (2004). In contrast to abundant convergence studies on BB type gradient methods, there are little works on the convergence properties of the steepest descent method.

In this paper, we study the convergence rate of the 2-norm of the error vector, namely we want to give an upper bound for

$$\frac{\|x_{k+1} - x^*\|_2}{\|x_k - x^*\|_2}. \tag{1.11}$$

The paper is organized as follows. In the next section, we reformulate the problem as a constrained optimization, and it is shown whose optimal value can only be reached in 3 dimensional subspaces. Thus, we consider 3 dimensional subspace subproblem in Section 3, where we show that the optimal value can be obtained with one variable reaching zero, implying that the 3 dimensional subspace problem can be reduced to a 2-dimensional subspace problem. In Section 4, we study the 2-dimensional subspace problem. The main results are presented in Section 5, and a brief discussion is given at the end of the paper.

2 Reformulation

The aim of this paper is to provide an upper bound for (1.11), assuming that $f(x)$ is given by (1.4) and x_k is generated by the steepest descent method (1.2)-(1.3).

It is easy to see that

$$\alpha_k = \frac{g_k^T g_k}{g_k^T H g_k} = \frac{(x_k - x^*)^T H^2 (x_k - x^*)}{(x_k - x^*)^T H^3 (x_k - x^*)}. \tag{2.1}$$

Thus the upper bound for (1.11) is the optimal value of

$$\max_{x \in \Re^n, x \neq x^*} \frac{\|(I - \alpha H)(x - x^*)\|_2}{\|x - x^*\|_2} \ , \tag{2.2}$$

subject to

$$\alpha = \frac{(x - x^*)^T H^2 (x - x^*)}{(x - x^*)^T H^3 (x - x^*)}. \tag{2.3}$$

We can rewrite (2.2)-(2.3) in the following equivalent form:

$$\max_{y \in \Re^n, \alpha \in \Re} \|(I - \alpha H)y\|_2^2 \tag{2.4}$$

subject to

$$y^T y = 1, \tag{2.5}$$

$$y^T H^2 y = \alpha y^T H^3 y. \tag{2.6}$$

Let (y^*, α^*) be a solution of (2.4)-(2.6), there exist Lagrange multipliers t^* and u^* such that

$$(I - \alpha^* H)^2 y^* = t^* y^* + u^* (H^2 y^* - \alpha^* H^3 y^*). \tag{2.7}$$

$$-(y^*)^T H y^* + \alpha^* (y^*)^T H^2 = -\frac{u^*}{2} (y^*)^T H^3 y^*. \tag{2.8}$$

It follows from (2.7) that $Span\{y^*, Hy^*, H^2 y^*\}$ is an invariance subspace with respect to H. Therefore it is sufficient for us to study the 3-dimensional subproblem. Furthermore, because the unite ball $\{y|y^Ty=1\}$ is invariance under orthogonal transformations, we can assume that H is a diagonal matrix $H = Diag[\lambda_1, \lambda_2, ..., \lambda_n]$. Thus, we only need to study the following problem

$$\max \sum_{i=1}^{3} (1-\alpha\mu_i)^2 y_i^2 \tag{2.9}$$

subject to

$$\sum_{i=1}^{3} y_i^2 = 1, \tag{2.10}$$

$$\sum_{i=1}^{3} \mu_i^2 y_i^2 = \alpha \sum_{i=1}^{3} \mu_i^3 y_i^2, \tag{2.11}$$

where $\mu_i (i = 1, 2, 3)$ are 3 eigenvalues of H, namely $\mu_i (i = 1, 2, 3) \subset \{\lambda_1, \lambda_2, ..., \lambda_n)$.

3 Three Dimensional Subspace Case

In this section, we study problem (2.9)-(2.11). Without loss of generality, we assume that $\mu_1 > \mu_2 > \mu_3$. Let $y_i^*(i = 1, 2, 3)$ be the solution of (2.9)-(2.11). If $y_3^* = 0$, $(y_1^*,\ y_2^*)$ is the solution of a two dimensional subproblem, which will be studied in the next section. Thus, for the rest of this section, we assume that $y_i^* \neq 0 (i = 1, 2, 3)$. It is obviously that $z_i^* = (y_i^*)^2 (i = 1, 2, 3)$ is a solution of

$$\max \sum_{i=1}^{3} (1-\alpha\mu_i)^2 z_i \tag{3.1}$$

subject to

$$\begin{aligned} \sum_{i=1}^{3} z_i &= 1, & (3.2)\\ \sum_{i=1}^{3} \mu_i^2 z_i &= \alpha \sum_{i=1}^{3} \mu_i^3 z_i, & (3.3)\\ z_i &\geq 0, \qquad i = 1, 2, 3. & (3.4) \end{aligned}$$

Our assumption indicates that inequalities (3.4) are inactive at the solution. Thus, there exist Lagrange multipliers t^* and u^* such that

$$(1-\alpha\mu_i)^2 = t^* + u^*(\mu_i^2 - \alpha\mu_i^3), \quad i = 1, 2, 3, \tag{3.5}$$

$$\sum_{i=1}^{3} 2(\alpha\mu_i - 1)\mu_i z_i^* = -u^* \sum_{i=1}^{3} \mu_i^3 z_i^*. \tag{3.6}$$

It follows from (3.5) that the determinant of the following matrix

$$\begin{pmatrix} 1 & (1-\alpha\mu_1)^2 & (\mu_1^2 - \alpha\mu_1^3) \\ 1 & (1-\alpha\mu_2)^2 & (\mu_2^2 - \alpha\mu_2^3) \\ 1 & (1-\alpha\mu_3)^2 & (\mu_3^2 - \alpha\mu_3^3) \end{pmatrix} \tag{3.7}$$

is zero, which gives that

$$(\mu_1\mu_2 + \mu_2\mu_3 + \mu_3\mu_1)\alpha^2 - 2(\mu_1 + \mu_2 + \mu_3)\alpha + 2 = 0. \tag{3.8}$$

Therefore, we have

$$\alpha = \frac{2}{\mu_1 + \mu_2 + \mu_3 \pm \sqrt{\mu_1^2 + \mu_2^2 + \mu_3^2}}. \tag{3.9}$$

The above relation and (3.3) imply that

$$\alpha = \frac{2}{\mu_1 + \mu_2 + \mu_3 - \sqrt{\mu_1^2 + \mu_2^2 + \mu_3^2}}. \tag{3.10}$$

The fact that α is a constant (independent of z) tells us that (3.1)-(3.4) is a linear programming problem. Hence there exists a solution $\hat{z}$ of (3.1)-(3.4) such that $\hat{z}_i = 0$ for some i. Hence we conclude that it is sufficient for us to consider the 2-dimensional subproblem, which will be discussed in the following section.

4 Two Dimensional Subspace Case

As discussed in the previous section, we only need to study the following 2-dimensional problem:

$$\max \sum_{i=1}^{2} (1 - \alpha\mu_i)^2 z_i \tag{4.1}$$

subject to

$$\sum_{i=1}^{2} z_i = 1, \tag{4.2}$$

$$\sum_{i=1}^{2} \mu_i^2 z_i = \alpha \sum_{i=1}^{2} \mu_i^3 z_i, \tag{4.3}$$

$$z_i \geq 0, \qquad i = 1, 2. \tag{4.4}$$

We assume that $\mu_1 > \mu_2 > 0$. It is obviously that

$$\frac{1}{\mu_1} < \alpha < \frac{1}{\mu_2}. \tag{4.5}$$

(4.3) gives

$$\mu_1^2(\mu_1\lambda - 1)z_1 = \mu_2^2(1 - \alpha\mu_2)z_2. \tag{4.6}$$

Denote $s_1 = \alpha\mu_1 - 1$, $s_2 = 1 - \alpha\mu_2$, then by (4.2) and (4.6) we have that

$$z_1 = \frac{\mu_2^2 s_2}{\mu_1^2 s_1 + \mu_2^2 s_2}, \qquad z_2 = \frac{\mu_1^2 s_1}{\mu_1^2 s_1 + \mu_2^2 s_2}. \tag{4.7}$$

Thus, the objective function in (4.1) can be written as

$$\hat{f}(z) = \frac{s_1 s_2 [s_1 \mu_2^2 + s_2 \mu_1^2]}{\mu_1^2 s_1 + \mu_2^2 s_2}. \tag{4.8}$$

Define $t = s_1/s_2$, which gives

$$s_1 = \frac{(\mu_1 - \mu_2)t}{\mu_1 + \mu_2 t}, \qquad s_2 = \frac{(\mu_1 - \mu_2)}{\mu_1 + \mu_2 t}. \tag{4.9}$$

The above two relations and (4.8) imply that

$$\begin{aligned} \hat{f}(z) &= \frac{(\mu_1 - \mu_2)^2 t(t\mu_2^2 + \mu_1^2)}{(\mu_1 + \mu_2 t)^2(\mu_1^2 t + \mu_2^2)} \\ &= \frac{(\beta - 1)^2 t(t + \beta^2)}{(\beta + t)^2(\beta^2 t + 1)} = \phi(t), \end{aligned} \tag{4.10}$$

where $\beta = \mu_1/\mu_2 > 1$. Maximizing $\phi(t)$ over $(0, +\infty)$, we obtain that $\phi'(t) = 0$, which gives

$$\psi(t) = \beta t^3 + (2\beta^3 - \beta^2)t^2 + (\beta - 2)t - \beta^2 = 0. \tag{4.11}$$

Let $t(\beta)$ be the unique root of $\psi(t) = 0$ in $(0, +\infty)$, we see that the maximum value of (4.1) is $\phi(t(\beta))$. What we need is to get an accurate estimate of $\phi(t(\beta))$. Direct calculations show that

$$\psi(\frac{1}{\sqrt{2\beta - 1}}) = \frac{2(\beta - 1)^2}{(2\beta - 1)^{\frac{3}{2}}}, \tag{4.12}$$

and

$$\begin{aligned} \psi(\frac{1}{\sqrt{2\beta}}) &= -\frac{1}{2}\beta + \frac{1}{\sqrt{2}}\sqrt{\beta} - \frac{3\sqrt{2}}{4}\frac{1}{\sqrt{\beta}} \\ &= -\frac{1}{2}\sqrt{\beta}(\sqrt{\beta} - \sqrt{2}) - \frac{3\sqrt{2}}{4}\frac{1}{\sqrt{\beta}} \\ &= -\frac{1}{2}(\beta - \sqrt{\beta}) - \frac{\sqrt{2} - 1}{2\sqrt{\beta}}(2 - \beta) - \frac{1}{\sqrt{\beta}}(1 - \frac{\sqrt{2}}{4}). \end{aligned} \tag{4.13}$$

From (4.12), (4.13) and $\beta > 1$, it is easy to see that

$$\psi(\frac{1}{\sqrt{2\beta - 1}}) > 0, \qquad \psi(\frac{1}{\sqrt{2\beta}}) < 0. \tag{4.14}$$

Thus, we have

$$\frac{1}{\sqrt{2\beta - 1}} > t(\beta) > \frac{1}{\sqrt{2\beta}}. \tag{4.15}$$

Consequently, we have the following estimate

$$\max \phi(t) = \phi(t(\beta)) = \frac{(\beta - 1)^2 t(\beta)(t(\beta) + \beta^2)}{(\beta + t(\beta))^2(\beta^2 t(\beta) + 1)} \leq \frac{(\beta - 1)^2}{(\beta + t(\beta))^2} \leq \frac{(\beta - 1)^2}{(\beta + 1/\sqrt{2\beta})^2}. \tag{4.16}$$

On the other hand, it follows from $t(\beta) < 1/\sqrt{2\beta - 1} < 1$ that

$$\max \phi(t) = \phi(t(\beta)) > \phi(1) = \frac{(\beta - 1)^2}{(\beta + 1)^2}. \tag{4.17}$$

5 Q-linear convergence of the steepest descent method

From the results in the previous sections, the upper bound for $\|x_{k+1} - x^*\|_2^2 / \|x_k - x^*\|_2^2$ is $\max \phi(t(\beta))$ for all $\beta = \lambda_i/\lambda_j$ with $\lambda_i > \lambda_j,\ i, j \in \{1, 2, ..., n\}$. Because the last term in the equality (4.16) is a monotonically increasing function of β, and because the maximal possible value of β is κ, (4.16) implies that

$$\frac{\|x_{k+1} - x^*\|_2^2}{\|x_k - x^*\|_2^2} \leq \phi(t(\kappa)) < \frac{(\kappa - 1)^2}{(\kappa + 1/\sqrt{2\kappa})^2}, \tag{5.1}$$

as long as $x_k \neq x^*$.

Theorem 5.1. *Let $f(x)$ be the convex quadratic function (1.4), $\{x_k, k = 1, 2, ...\}$ be the sequence generated by the steepest descent method (1.2)-(1.3), and $\kappa = \lambda_1(H)/\lambda_n(H) > 1$, then for any starting point $x_1 \in \Re^n$ either $x_2 = x^* = -H^{-1}g$ or*

$$\frac{\|x_{k+1} - x^*\|_2}{\|x_k - x^*\|_2} \leq \sqrt{\phi(t(\kappa))} < \frac{\kappa - 1}{\kappa + 1/\sqrt{2\kappa}}, \tag{5.2}$$

for all k. Furthermore, there exists $x_1 \in \Re^n$ such that

$$\frac{\|x_{k+1} - x^*\|_2}{\|x_k - x^*\|_2} > \frac{\kappa - 1}{\kappa + 1}, \tag{5.3}$$

for all odd k.

Proof. If $x_2 \neq x^*$, it follows from Forsythe(1968) that $x_k \neq x^*$ for all k. Therefore we can see that (5.2) holds due to (5.1).

Since $t(\kappa) \leq 1/\sqrt{2\kappa - 1} < 1$, there exists a $\hat{t} \in (t(\kappa), 1)$. Hence we have

$$\phi(\ \hat{t}\ ) > \phi(1) = \frac{(\kappa - 1)^2}{(\kappa + 1)^2}. \tag{5.4}$$

Now, we can define

$$\hat{s}_1 = \frac{(\kappa - 1)\hat{t}}{\kappa + \hat{t}}\ , \qquad \hat{s}_2 = \frac{\kappa - 1}{\kappa + \hat{t}}\ , \tag{5.5}$$

and

$$\hat{z}_1 = \frac{s_2}{\kappa^2 s_1 + s_2}\ , \qquad \hat{z}_2 = \frac{\kappa^2 s_1}{\kappa^2 s_1 + s_2}\ . \tag{5.6}$$

Let v_1 and v_2 be the two unit-norm eigenvalues of H corresponding to $\lambda_1(H)$ and $\lambda_n(H)$:

$$Hv_1 = \lambda_1(H)v_1, \qquad \|v_1\|_2 = 1, \tag{5.7}$$

$$Hv_2 = \lambda_n(H)v_2, \qquad \|v_2\|_2 = 1. \tag{5.8}$$

Now we can choose the initial vector x_1 by

$$x_1 = x^* + (\sqrt{\hat{z}_1}\,)v_1 + (\sqrt{\hat{z}_2}\,)v_2, \tag{5.9}$$

which, from the analysis in section 3, implies that

$$\frac{\|x_2 - x^*\|_2^2}{\|x_1 - x^*\|_2^2} = \phi(\hat{t}). \tag{5.10}$$

(5.9) shows that all iterate points $\{x_k\}$ are in the 2-dimensional subspace $x^* + Span\{v_1, v_2\}$. Thus, exact line search conditions imply that there exists a constant $c \in (0, 1)$ such that $x_k - x^* = c^{k-1}(x_1 - x^*)$ for all odd k. Hence,

$$\frac{\|x_{2k} - x^*\|_2^2}{\|x_{2k-1} - x^*\|_2^2} = \frac{\|x_2 - x^*\|_2^2}{\|x_1 - x^*\|_2^2} = \phi(\hat{t}) \tag{5.11}$$

for all k. Consequently, (5.11) and (5.4) imply that (5.3) holds for all odd k. □

6 Discussion

In this paper we have proven that the steepest descent method implies that

$$\frac{\|x_{k+1} - x^*\|_2}{\|x_k - x^*\|_2} < \frac{\kappa - 1}{\kappa + 1/\sqrt{2\kappa}}, \tag{6.12}$$

for all k. And we have also shown that the upper bound given in the righthand side of the above inequality can not be improved to $(\kappa - 1)/(\kappa + 1)$. In fact, we have constructed an example which gives

$$\frac{\|x_{k+1} - x^*\|_2}{\|x_k - x^*\|_2} > \frac{\kappa - 1}{\kappa + 1}, \tag{6.13}$$

for all odd k. This indicates that the inequality (6.12) we established is very close to the best possible that can be obtained.

References

[1] H. Akaike, *On a successive transformation of probability distribution and its application to the analysis of the optimum gradient method*, Ann. Inst. Statist. Math. Tokyo, 11 (1959) 1-16.

[2] J. Barzilai and J. M. Borwein, *Two point step size gradient methods*, IMA J. Numer. Anal., 8 (1988) 141-148.

[3] A. Cauchy, *Méthode générale pour la résolution des systèms d'equations simultanées*, Comp. Rend. Sci. Paris, 25 (1847), pp. 46-89.

[4] H.B. Curry, *The method of steepest descent for nonlinear minimization problems*, Quart. Appl. Math., 2 (1944) 258-261.

[5] Y.H. Dai, *Alternate step gradient method*, Report AMSS-2001-041, Academy of Mathematics and Systems Sciences, Chinese Academy of Sciences, 2001.

[6] Y.H. Dai and R. Fletcher, *On the asymptotic behaviour of some new gradient methods*, Numerical Analysis Report, NA/212, Dept. of Math. University of Dundee, Scotland, UK (2003).

[7] Y.H. Dai and Y. Yuan, *Alternate minimization gradient method*, IMA Journal of Numerical Analysis, 23 (2003) 377-393.

[8] Y.H. Dai and Y. Yuan, *Analysis of monotone gradient methods*, J. Industrial and Management Optimization, 1 (2005) 181-192.

[9] Y. H. Dai and H. Zhang, *An Adaptive Two-Point Step-size Gradient Method*, Research report, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, 2001.

[10] R. Fletcher, *Practical Methods of Optimization*(second Edition), John Wiley and Sons, Chichester, 1987.

[11] R. Fletcher, *On the Barzilar-Borwein method*, Research Report, University of Dundee, UK, 2001.

[12] G. E. Forsythe, *On the asymptotic directions of the s-dimensional optimum gradient method*, Numerische Mathematik, 11 (1968), pp. 57-76.

[13] A. Friedlander, J. M. Martínez, B. Molina, and M. Raydan, *Gradient method with retards and generalizations*, SIAM J. Numer. Anal., 36 (1999), 275-289.

[14] J. Greenstadt, *On the relative efficiencies of gradient methods*, Math. Comp. 21 (1967) 360-367.

[15] D. Luenberger, *Introduction to Linear and Nonlinear Programming, 2nd Edition*, (Addison Wesley, 1984).

[16] J. Nocedal, A. Sartenaer and C. Zhu, *On the behavior of the gradient norm in the steepest descent method*, Research report, Northwestern University, USA, 2000

[17] M. Raydan, *On the Barzilai and Borwein choice of steplength for the gradient method*, IMA J. Numer. Anal. 13 (1993) 321-326.

[18] M. Raydan, *The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem*, SIAM J. Optim., 7 (1997) 26-33.

[19] M. Raydan and B. F. Svaiter, *Relaxed Steepest Descent and Cauchy-Barzilai-Borwein Method*, Computational Optimization and Applications, 21 (2002), pp. 155-167.

[20] W.Y. Sun and Y. Yuan, *Optimization Theory and Methods: Nonlinear Programming*, (Springer, New York, 2006).

[21] E. Stiefel, *Über einige Methoden der Relaxationsrechnung*, Z. Angew. Math. Physik, 3 (1952) 1-33.

[22] M.N. Vrahatis, G.S. Androulakis, J.N. Lambrinos and G.D. Magoulas, *A class of gradient unconstrained minimization algorithms with adaptive step-size*, J. Comp. and Appl. Math. 114 (2000) 367-386.

[23] Y. Yuan, *A new stepsize for the steepest descent method*, Research report, Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and Systems Sciences, Chinese Academy of Sciences, 2004.