Math. Program., Ser. A (2010) 123:339–343
DOI 10.1007/s10107-009-0267-8

SHORT COMMUNICATION

A short note on the Q-linear convergence of the steepest descent method

Ya-xiang Yuan

Received: 4 June 2007 / Accepted: 22 September 2008 / Published online: 3 February 2009

Abstract This short note gives the sharp bound for the Q-linear convergence rate of the iterates generated by the steepest descent method with exact line searches when the objective function is strictly convex quadratic.

Keywords Steepest descent · Exact line search · Q-linear · Rate of convergence

Mathematics Subject Classification (2000) 90C30 · 65K05

1 Introduction

Consider the steepest descent method with exact line searches for a strictly convex quadratic function in $\Re^n$:

$$f(x) = g^T x + \frac{1}{2} x^T H x, \tag{1.1}$$

where $H \in \Re^{n \times n}$ is a symmetric positive definite matrix. The method is defined by

$$x_{k+1} = x_k - \frac{g_k^T g_k}{g_k^T H g_k} g_k, \tag{1.2}$$

where $g_k = \nabla f(x_k) = g + Hx_k$. It is well known (for example, see [3,4]) that

This work is partially supported by Chinese NSF grant 10831006 and by CAS grant kjcx-yw-s7.

Y.-x. Yuan (✉)
State Key Laboratory of Scientific/Engineering Computing,
Institute of Computational Mathematics and Scientific/Engineering Computing, AMSS,
Chinese Academy of Sciences, 100190 Beijing, People's Republic of China
e-mail: yyx@lsec.cc.ac.cn

$$\frac{\|x_{k+1}-x^*\|_H}{\|x_k-x^*\|_H} \le \frac{\kappa-1}{\kappa+1} < 1, \tag{1.3}$$

where $x^* = -H^{-1}g$, $\|.\|_H$ is the H-norm defined by $\|v\|_H = \sqrt{v^T H v}$, and $\kappa = \lambda_1(H)/\lambda_n(H)$ is the condition number of H. From (1.3), one can show that the 2-norm of the error vector $\|x_k - x^*\|_2$ converges to zero R-linearly. A bound for the R-linear convergence rate was also given in Luenberger [2]. However, inequality (1.3) does not imply the Q-linear convergence of $\|x_k - x^*\|_2$. To the author's knowledge, there has not yet been any results on the rate of the Q-linear convergence of $\|x_k - x^*\|_2$. The aim of this short note is to find the sharp bound for the Q-linear convergence rate of $\|x_k - x^*\|_2$.

2 Reformulation

It is straightforward to see that

$$\|x_{k+1}-x^*\|_2 = \|(I-\alpha_k H)(x_k-x^*)||_2, \tag{2.1}$$

where $\alpha_k = (x_k-x^*)^T H^2 (x_k-x^*)/(x_k-x^*)^T H^3 (x_k-x^*)$. Thus, we have that

$$\frac{\|x_{k+1}-x^*\|_2}{\|x_k-x^*\|_2} \le \sqrt{\delta(H)}, \tag{2.2}$$

if we define

$$\delta(H) = \max_{y\in\Re^n, \alpha\in\Re} \quad \|(I-\alpha H)y\|_2^2 \tag{2.3}$$

$$\text{s. t.} \quad y^T H^2 y = \alpha y^T H^3 y, \quad y^T y = 1. \tag{2.4}$$

Let (y^*, α^*) be a solution of (2.3) and (2.4), there exist Lagrange multipliers t^* and u^* such that

$$\left(I-\alpha^* H\right)^2 y^* = t^* y^* + u^* \left(H^2 y^* - \alpha^* H^3 y^*\right), \tag{2.5}$$

$$-(y^*)^T H y^* + \alpha^* (y^*)^T H^2 = -\frac{u^*}{2}(y^*)^T H^3 y^*. \tag{2.6}$$

It follows from (2.5) that $Span\{y^*, Hy^*, H^2y^*\}$ 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 $\{y|y^T y = 1\}$ is invariant under orthogonal transformations, we can assume that H is a diagonal matrix $H = \mathrm{Diag}[\lambda_1, \lambda_2, \ldots, \lambda_n]$. Thus, we only need to study the following problem

$$\max \sum_{i=1}^{3}(1-\alpha\mu_i)^2 y_i^2 \tag{2.7}$$

$$\text{s.t.} \sum_{i=1}^{3}\mu_i^2 y_i^2 = \alpha\sum_{i=1}^{3}\mu_i^3 y_i^2, \quad \sum_{i=1}^{3} y_i^2 = 1. \tag{2.8}$$

where $\mu_i(i = 1, 2, 3)$ are 3 eigenvalues of H, namely $\mu_i(i = 1, 2, 3) \subset \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$. Without loss of generality, we assume that $\mu_1 > \mu_2 > \mu_3$. Let $y_i^*(i = 1, 2, 3)$ be the solution of (2.7) and (2.8). First, we prove that we can assume that one of $y_i^*(i = 1, 2, 3)$ is zero. Suppose 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

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

$$\text{s.t.} \sum_{i=1}^{3}\mu_i^2 z_i = \alpha\sum_{i=1}^{3}\mu_i^3 z_i, \quad \sum_{i=1}^{3} z_i = 1, \quad z_i \geq 0,\ i = 1, 2, 3. \tag{2.10}$$

Our assumption indicates that inequalities $z_i \geq 0(i = 1, 2, 3)$ are inactive at the solution. Thus, there exist Lagrange multipliers t^* and u^* such that

$$(1-\alpha\mu_i)^2 = t^* + u^*\left(\mu_i^2 - \alpha\mu_i^3\right), \quad i = 1, 2, 3, \tag{2.11}$$

$$\sum_{i=1}^{3} 2(\alpha\mu_i - 1)\mu_i z_i^* = -u^*\sum_{i=1}^{3}\mu_i^3 z_i^*. \tag{2.12}$$

It follows from 2.11 that the determinant of the following matrix

$$\begin{pmatrix} 1 & (1-\alpha\mu_1)^2 & (\mu_1^2 - \alpha\mu_1^3) \\ 1 & (1-\alpha\mu_2)^2 & (\mu_2^2 - \alpha\mu_2^3) \\ 1 & (1-\alpha\mu_3)^2 & (\mu_3^2 - \alpha\mu_3^3) \end{pmatrix} \tag{2.13}$$

is zero, which gives that

$$(\mu_1\mu_2 + \mu_2\mu_3 + \mu_3\mu_1)\,\alpha^2 - 2\,(\mu_1 + \mu_2 + \mu_3)\,\alpha + 2 = 0. \tag{2.14}$$

Therefore, we have $\alpha = 2/(\mu_1 + \mu_2 + \mu_3 \pm \sqrt{\mu_1^2 + \mu_2^2 + \mu_3^2})$. This relation and (2.10) imply that $\alpha = 2/(\mu_1 + \mu_2 + \mu_3 - \sqrt{\mu_1^2 + \mu_2^2 + \mu_3^2})$. The fact that α is a constant (independent of z) tells us that (2.9) and (2.10) is a linear programming problem.

Hence there exists a solution $\hat{z}$ of (2.9) and (2.10) such that $\hat{z}_i = 0$ for some i. Hence, we have proved that it is sufficient for us to consider the 2-dimensional subproblem:

$$\max \sum_{i=1}^{2} (1-\alpha\mu_i)^2 z_i \tag{2.15}$$

$$\text{s. t.} \sum_{i=1}^{2} \mu_i^2 z_i = \alpha \sum_{i=1}^{2} \mu_i^3 z_i, \quad \sum_{i=1}^{2} z_i = 1, \quad z_i \geq 0,\ i = 1, 2. \tag{2.16}$$

We assume that $\mu_1 > \mu_2 > 0$. It follows from (2.16) that ${\mu_1}^{-1} < \alpha < {\mu_2}^{-1}$ and

$$\mu_1^2(\mu_1\lambda - 1)z_1 = \mu_2^2(1-\alpha\mu_2)z_2. \tag{2.17}$$

Denote $s_1 = \alpha\mu_1 - 1$, $s_2 = 1 - \alpha\mu_2$, then by (2.16) and (2.17) we have that $z_1 = \frac{\mu_2^2 s_2}{\mu_1^2 s_1 + \mu_2^2 s_2}$, $z_2 = \frac{\mu_1^2 s_1}{\mu_1^2 s_1 + \mu_2^2 s_2}$. Thus, the objective function in (2.15) can be written as

$$\hat{f}(z) = \frac{s_1 s_2 [s_1\mu_2^2 + s_2\mu_1^2]}{\mu_1^2 s_1 + \mu_2^2 s_2}. \tag{2.18}$$

Define $t = s_1/s_2$, which gives $s_1 = \frac{(\mu_1-\mu_2)t}{\mu_1+\mu_2 t}$ and $s_2 = \frac{(\mu_1-\mu_2)}{\mu_1+\mu_2 t}$. Thus, (2.18) implies that

$$\hat{f}(z) = \frac{(\mu_1-\mu_2)^2 t(t\mu_2^2+\mu_1^2)}{(\mu_1+\mu_2 t)^2(\mu_1^2 t + \mu_2^2)} = \frac{(\beta-1)^2 t(t+\beta^2)}{(\beta+t)^2(\beta^2 t+1)} = \phi(t), \tag{2.19}$$

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

$$\psi(t) = \beta t^3 + (2\beta^3 - \beta^2)t^2 + (\beta-2)t - \beta^2 = 0. \tag{2.20}$$

Let $t(\beta)$ be the unique root of $\psi(t) = 0$ in $(0, +\infty)$, we see that the maximum value of (2.15) is $\phi(t(\beta))$. What we need is to get an accurate estimate of $\phi(t(\beta))$. It can be shown that [5] $\psi\left(\frac{1}{\sqrt{2\beta-1}}\right) > 0$ and $\psi(\frac{1}{\sqrt{2\beta}}) < 0$. Thus, we have

$$\frac{1}{\sqrt{2\beta-1}} > t(\beta) > \frac{1}{\sqrt{2\beta}}. \tag{2.21}$$

Consequently, we have the following estimate

$$\begin{aligned}\max\phi(t) = \phi(t(\beta)) &= \frac{(\beta-1)^2 t(\beta)(t(\beta)+\beta^2)}{(\beta+t(\beta))^2(\beta^2 t(\beta)+1)} \\ &\leq \frac{(\beta-1)^2}{(\beta+t(\beta))^2} \leq \frac{(\beta-1)^2}{(\beta+1/\sqrt{2\beta})^2}.\end{aligned} \tag{2.22}$$

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

$$\max \phi(t) = \phi(t(\beta)) > \phi(1) = \frac{(\beta - 1)^2}{(\beta + 1)^2}. \tag{2.23}$$

3 Q-linear convergence of the steepest descent method

From the results in the previous section, we can see that $\delta(H)$ equals to $\max \phi(t(\beta))$ for all $\beta = \lambda_i/\lambda_j$ with $\lambda_i > \lambda_j$, $i, j \in \{1, 2, \ldots, n\}$. Because the last term in the equality (2.22) is a monotonically increasing function of β, and because the maximal possible value of β is κ, (2.22) implies that

$$\delta(H) = \max \phi(t(\beta)) = \phi(t(\kappa)) < \frac{(\kappa - 1)^2}{(\kappa + 1/\sqrt{2\kappa})^2}. \tag{3.1}$$

Now we can establish our convergence result as follows.

Theorem 31 *Let $f(x)$ be the convex quadratic function 1.1, $\{x_k, k = 1, 2, \ldots\}$ generated by (1.2), and $\kappa = \lambda_1(H)/\lambda_n(H) > 1$, then for any $x_1 \in \Re^n$ either $x_2 = x^* = -H^{-1}g$ or*

$$\frac{\|x_{k+1} - x^*\|_2}{\|x_k - x^*\|_2} \le \sqrt{\phi(t(\kappa))} < \frac{\kappa - 1}{\kappa + 1/\sqrt{2\kappa}}, \tag{3.2}$$

for all k.

Proof If $x_2 \neq x^*$, it follows from Forsythe [1] that $x_k \neq x^*$ for all k. Therefore it can be seen that (3.2) follows from (2.2) and (3.1). □

Due to (2.23), the upper bound given in the righthand side of (3.2) can not be improved to $(\kappa - 1)/(\kappa + 1)$. In fact, for any H with $\kappa > 1$ we can construct an example [5] which gives $\frac{\|x_{k+1}-x^*\|_2}{\|x_k-x^*\|_2} > \frac{\kappa-1}{\kappa+1}$ for all odd k. This indicates that the inequality (3.2) we established is very close to the best possible that can be obtained.

References

1. Forsythe, G.E.: On the asymptotic directions of the s-dimensional optimum gradient method. Numer. Math. **11**, 57–76 (1968)
2. Luenberge, D.G.: Optimization by Vector Space Methods. Wiley, New York (1968)
3. Nocedal, J., Wright, S.: Numerical Optimization. Springer, New York (1999)
4. Sun, W.Y., Yuan, Y.: Optimization Theory and Methods: Nonlinear Programming. Springer, New York (2006)
5. Yuan, Y.: A Short Note on the Q-linear convergence of the steepest descent method. Research report, Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and Systems Sciences, Chinese Academy of Sciences (2007)