CONJUGATE GRADIENT METHOD

MATH 524, SPRING 2008

Abstract. This short note is on the derivation and convergence of a popular algorithm for minimization of quadratic functionals (or solving linear systems), known as the method of *Conjugate Gradients (CG)*. To the best of the knowledge of the author of this short note, the CG algorithm has been first introduced in 1952 by M. R. Hestenes and E. Stiefel in [2]. The derivation of the CG algorithm, given here, follows lecture notes by D. N. Arnold [1].

1. Introduction

1.1. **Preliminaries and notation.** The conjugate gradient method is a method for minimizing the following quadratic functional:

$$x_* = \arg\min_{x\in\mathbb{R}^n} \varphi(x), \quad \varphi(x) = \frac{1}{2}x^TAx - b^Tx, \tag{1.1}$$

where $A \in \mathbb{R}^{n\times n}$ is symmetric positive definite (SPD) matrix and $b \in \mathbb{R}^n$ is a given vector. Clearly, we have

$$\nabla\varphi(x) = Ax - b, \quad \nabla^2\varphi = A, \quad \text{(the Hessian is independent on } x). \tag{1.2}$$

Since the Hessian A is SPD, from well known conditions for a minimum of a function, we may conclude that there is a unique minimizer x_* of $\varphi(\cdot)$. Moreover, (1.2) implies that x_* is also the solution to the system of linear equations:

$$Ax = b.$$

This is why CG method is oftentimes thought as a method for the solution of linear systems.

In what follows we will need the following preliminary settings

1. Since A is SPD, it defines an inner product x^TAy between two vectors x and y in $\mathbb{R}^n$, which we will refer to as A-inner product. The corresponding vector norm is defined by $\|x\|_A^2 = x^TAx$.
2. From the Taylor theorem for $g(t) = \varphi(y + tz)$ we obtain the following identity for all $t \in \mathbb{R}$, and all $y \in \mathbb{R}^n$ and $z \in \mathbb{R}^n$:

$$\varphi(y + tz) = \varphi(y) + t[\nabla\varphi(y)]^Tz + \frac{t^2}{2}z^TAz. \tag{1.3}$$

1.2. **Line search methods.** The CG method is nothing but a line search method with special choice of directions. Given a current approximation x_j to the minimum x_*, and a direction vector p_j, a line search method determines the next approximation x_{j+1} via the following two steps:

1. Find $\alpha_j = \arg\min \varphi(x_j + \alpha p_j)$,
2. Set $x_{j+1} = x_j + \alpha_j p_j$.

In the following, we assume that x_0 is a given vector (initial guess). Then, applying k steps of the line search method results in k-iterates $\{x_j\}_{j=0}^{k-1}$.

From the relation (1.3) and (1.2) we immediately find that

$$\alpha_j = \frac{-p_j^T r_j}{p_j^T A p_j}, \quad \text{where} \quad r_j = \nabla\varphi(x_j) = Ax_j - b. \tag{1.4}$$

Here r_j is usually referred to as the *residual vector.*

We introduce the following definition.

Definition 1.1. *We say that the set of directions $\{p_j\}_{j=0}^{k-1}$ is a conjugate set of directions, iff $p_j^T A p_i = 0$ for all $i = 1, \dots, (k-1)$, $j = 1, \dots, (k-1)$, $i \neq j$.*

By symmetry this definition can also be stated as: The set of directions $\{p_j\}_{j=0}^{k-1}$ is a conjugate set of directions, iff $p_j^T A p_i = 0$ for all i and j satisfying $0 \le i < j \le (k-1)$.

We introduce now the following vector spaces and afine spaces (for $k = 1, \dots$):

$$\begin{aligned} W_k &:= \operatorname{span}\{p_0, \dots, p_{k-1}\} \\ U_k &:= x_0 + W_k = \{z \in \mathbb{R}^n \ \mid \ z = x_0 + w_k, \quad w_k \in W_k\} \end{aligned} \tag{1.5}$$

For convenience we set $W_0 := \{0\}$ and $U_0 := \{x_0\}$.

We now prove a technical lemma which will be used later in the proof or Theorem 2.1.

Lemma 1.2. *Assume that $p_i^T A p_j = 0$ for all $0 \le j < i$, where i is a fixed integer, and that $\{x_j\}_{j=0}^{i}$ are obtained via the line search algorithm. Then the following identity holds:*

$$p_i^T r_i = p_i^T [\nabla\varphi(y)], \quad \textit{for all} \quad y \in U_i. \tag{1.6}$$

Proof. We first note that since $\{x_j\}_{j=0}^{i}$ are obtained via the line search algorithm we have that $x_i \in U_i$. If we take $y \in U_i$, from the definition of U_i it follows that $x_i - y \in W_i$ and hence $p_i^T A(x_i - y) = 0$ (because $p_i^T A w = 0$ for all $w \in W_i = \operatorname{span}\{p_0, \dots, p_{i-1}\}$). The proof of the identity (1.6) then is as follows:

$$p_i^T (r_i - [\nabla\varphi(y)]) = p_i^T (Ax_i - b - Ay + b) = p_i^T A(x_i - y) = 0.$$

□

2. Properties of line search method with conjugate directions

Clearly on every step, the line search algorithm minimizes $\varphi(x)$ in a fixed direction only. However, if the directions are conjugate (see Definition 1.1), then much stronger result can be proved, as the Theorem 2.1 below states: a choice of conjugate directions in the line search method, results in obtaining a minimizer x_k for the whole space U_k. In some sense, one may say that the next theorem is the base for constructing the conjugate gradient method.

Theorem 2.1. *If the directions in the line search algorithm are conjugate, and $\{x_j\}_{j=0}^{k}$ are the iterates obtained after k steps of the line search algorithm then*

$$x_j = \arg\min_{x \in U_j} \varphi(x), \quad \textit{for all} \quad 1 \le j \le k \tag{2.1}$$

Proof. The proof is by induction. For $k = 1$, the result follows from the definition of x_1 as a minimizer on U_1. Assume that for $k = i$,

$$x_j = \arg\min_{y \in U_j} \varphi(y), \quad \text{for all} \quad 1 \le j \le i\,.$$

To prove the statement of the theorem then, we need to show that

$$\text{If} \quad x_{i+1} = x_i + \alpha_i p_i \quad \text{then} \quad x_{i+1} = \arg \min_{x \in U_{i+1}} \varphi(x)\,.$$

By the definition of U_{i+1}, any $x \in U_{i+1}$ can be written as $x = y + \alpha p_i$, where $\alpha \in \mathbb{R}$ and $y \in U_i$. Applying (1.3) and then Lemma 1.2 leads to

$$\begin{aligned} \varphi(x) &= \varphi(y + \alpha p_i) = \varphi(y) + \alpha p_i^T [\nabla \varphi(y)] + \frac{\alpha^2}{2} p_i^T A p_i \\ &= \varphi(y) + \left[\alpha p_i^T [\nabla \varphi(x_i)] + \frac{\alpha^2}{2} p_i^T A p_i \right]. \end{aligned}$$

Note that we have arrived at a decoupled functional, since the first term does not depend on α and the second term does not depend on y. Thus,

$$\min_{x \in U_{i+1}} \varphi(x) = \min_{y \in U_i} \varphi(y) + \min_{\alpha \in \mathbb{R}^n} \left[\alpha p_i^T r_i + \frac{\alpha^2}{2} p_i^T A p_i \right].$$

The right side is minimized when $y = x_i$ and $\alpha = \alpha_i = \dfrac{-p_i^T r_i}{p_i^T A p_i}$, and hence the left side is minimized exactly for $x_{i+1} = x_i + \alpha_i p_i$, which concludes the proof. □

3. The Conjugate Gradient algorithm

The conjugate gradient method is an algorithm that explores the result in Theorem 2.1 and constructs conjugate directions. Here is the rationale of what we plan to do in this section:

- We first give a general recurrence relation that generates a set of conjugate directions (Lemma 3.1).
- We then show that this recurrence relation can be reduced to a much simpler expression (see Lemma 3.2(iv)).
- As a result, we will get a line search method, which uses conjugate set of directions and is known as the CG method (see Algorithm 3.3).

We begin with a technical lemma, which could be obtained by Gramm-Schmidt orthogonalization with respect to the A-inner product of the residual vectors $\{r_j\}_{j=1}^k$.

Lemma 3.1. *Let $p_0 = -r_0$ and let for $k = 1, 2, \ldots$*

$$p_k = -r_k + \sum_{j=0}^{k-1} \frac{p_j^T A r_k}{p_j^T A p_j} p_j \tag{3.1}$$

Then $p_j^T A p_m = 0$ for all $0 \le m < j \le k$.

Proof. We will show that the relation (3.1) gives conjugate directions is by induction. For $k = 1$ one directly checks that $p_1^T A p_0 = 0$. Assume that for $k = i$ the vectors $\{p_j\}_{j=0}^i$ are pairwise conjugate. We then need to show that $p_{i+1}^T A p_m = 0$ for all $m \le i$. Let $m \le i$. Then we have

$$\begin{aligned} p_{i+1}^T A p_m &= -r_{i+1}^T A p_m + \sum_{j=0}^{i} \frac{p_j^T A r_{i+1}}{p_j^T A p_j} p_j^T A p_m \\ &= -r_{i+1}^T A p_m + \frac{p_m^T A r_{i+1}}{p_m^T A p_m} p_m^T A p_m = 0 \end{aligned}$$

□

Next Lemma among other things, shows that the sum in (3.1) contains only one term.

Lemma 3.2. *Let* $\{p_j\}_{j=0}^k$ *are directions obtained via* (3.1). *Then*

(i) $W_k = \operatorname{span}\{r_0, \dots, r_{k-1}\}$
(ii) $r_m^T r_j = 0$, *for all* $0 \le j < m \le k$
(iii) $p_k^T r_j = -r_k^T r_k$, *for all* $0 \le j \le k$
(iv) *The direction vector* p_k *satisfies*

$$p_k = -r_k + \beta_{k-1} p_{k-1}, \quad \textit{where} \quad \beta_{k-1} = \frac{r_k^T r_k}{r_{k-1}^T r_{k-1}}.$$

Proof. The first item follows directly from (3.1) and a simple induction argument, since $p_0 = -r_0$.

To prove (ii), we first use (i) to conclude that for $0 \le j < m \le k$, and any $t \in \mathbb{R}$ we have that

$$r_j \in W_{j+1} \subset W_m \quad \text{and hence} \quad (x_m + t r_j) \in U_m.$$

Further, from Theorem 2.1, since x_m is the unique minimizer of $\varphi(\cdot)$ over U_m, it follows that $t = 0$ is the unique minimizer of $g(t) = \varphi(x_m + t r_j)$. Hence we have

$$0 = \left.\frac{d\varphi(x_m + t r_j)}{dt}\right|_{t=0} = [\nabla \varphi(x_m)]^T r_j = r_m^T r_j,$$

and this proves (ii).

To show that (iii) holds, we first show the identity in (iii) for $j = k$. Indeed, from (i) it follows that r_k is orthogonal to each p_l for $l < k$. Hence, if we take the inner product with r_k, the second term in the right side of (3.1) would vanish, and this is exactly the identity in (iii) for $j = k$. If $j < k$, then we have that $(x_k - x_j) \in W_k$, and hence $p_k^T A(x_k - x_j) = 0$. Therefore,

$$p_k^T (r_k - r_j) = p_k^T A (x_k - x_j) = 0.$$

To show (iv) we write $p_k \in W_{k+1}$ as linear combination of $\{r_j\}_{j=0}^k$ (which form an orthogonal basis), and then apply (iii). This leads to

$$\begin{aligned} p_k &= \sum_{j=0}^{k} \frac{p_k^T r_j}{r_j^T r_j} r_j = -\sum_{j=0}^{k} \frac{r_k^T r_k}{r_j^T r_j} r_j = -r_k - \frac{r_k^T r_k}{r_{k-1}^T r_{k-1}} \sum_{j=0}^{k-1} \frac{r_{k-1}^T r_{k-1}}{r_j^T r_j} r_j \\ &= -r_k + \beta_{k-1} \sum_{j=0}^{k-1} \frac{p_{k-1}^T r_j}{r_j^T r_j} r_j = -r_k + \beta_{k-1} p_{k-1}. \end{aligned}$$

□

We now can write the conjugate gradient algorithm, using the much shorter recurrence relation for the direction vectors p_k, which is provided by Lemma 3.2(iv). We denote below $\|y\|^2 = y^T y$ and $\|y\|_A^2 = y^T A y$ for a vector $y \in \mathbb{R}^n$.

Algorithm 3.3 (Conjugate Gradient)**.** *Let x_0 be given initial guess.*

Set $r_0 = Ax_0 - b$ and $p_0 = -r_0$, $k = 0$.

While $r_k \neq 0$ ***do***

$$\begin{aligned}
\alpha_k &= \frac{\|r_k\|^2}{\|p_k\|_A^2} && \text{[from Lemma 3.2(iii)]} \\
x_{k+1} &= x_k + \alpha_k p_k \\
r_{k+1} &= r_k + \alpha_k A p_k && \text{[because } Ax_{k+1} - b = Ax_k - b + \alpha_k A p_k\text{]} \\
\beta_k &= \frac{\|r_{k+1}\|^2}{\|r_k\|^2} \\
p_{k+1} &= -r_k + \beta_k p_k && \text{[from Lemma 3.2(iv)]}
\end{aligned}$$

Set $k = k+1$

endWhile

4. Convergence rate of the Conjugate Gradient method

In this section we will present an estimate for the convergence rate of the CG algorithm. The convergence rate estimate, given here is rather general and does not take into account knowledge of the distribution of the eigenvalues of A. There are estimates that are more refined in this regard. We refer to Luenberger [3] for further reading.

4.1. **Krylov subspaces and error reduction.** To analyze the error we first prove the following result:

Lemma 4.1. *The following relation holds:*

$$W_l = \operatorname{span}\{r_0, \dots, A^{l-1} r_0\}. \tag{4.1}$$

Proof. The case $l = 1$, being clear, we assume that the relation holds for $l = i$, and we would like to show that the same relation holds for $l = (i+1)$. From Lemma 3.2(i), this would be equivalent to showing that $r_i \in \operatorname{span}\{r_0, \dots, A^i r_0\}$. By the induction assumption, we can write

$$W_i \ni r_{i-1} = R_{i-1}(A) r_0, \quad \text{and} \quad W_i \ni p_{i-1} = P_{i-1}(A) r_0,$$

where $R_{i-1}(\cdot)$ and $P_{i-1}(\cdot)$ are polynomials of degree less that or equal to $(i-1)$. We then have

$$\begin{aligned}
r_i &= r_{i-1} + \alpha_{i-1} A p_{i-1} \\
&= R_{i-1}(A) r_0 + \alpha_{i-1} A P_{i-1}(A) r_0 \in \operatorname{span}\{r_0, \dots, A^i r_0\},
\end{aligned}$$

which concludes the proof. □

We now present a general error estimate relating $\|x_* - x_l\|_A$ and $\|x_* - x_0\|_A$.

Lemma 4.2. *The following estimate holds:*

$$\|x_* - x_l\|_A = \inf_{P \in \mathcal{P}_l;\, P(0)=1} \|P(A)(x_* - x_0)\|_A.$$

Proof. Since r_l is orthogonal to W_l, we have

$$(x_* - x_l)^T A y = r_l^T y = 0, \quad \text{for all} \quad y \in W_l.$$

Denoting for a moment $w_l = (x_l - x_0) \in W_l$ and $e_0 = x_* - x_0$, the relation above implies that

$$0 = (x_* - x_l)^T A y = (e_0 - w_l)^T A y \quad \text{for all} \quad y \in W_l.$$

Therefore, $w_l = (x_l - x_0)$ is an A-orthogonal projection of $e_0 = (x_* - x_0)$ on W_l. Thus,

$$\|e_0 - w_l\|_A = \min_{w \in W_l} \|e_0 - w\|_A$$

But from Lemma 4.1 we know that $w = Q_{l-1}(A) r_0$, for a polynomial $Q_{l-1} \in \mathcal{P}_{l-1}$. Also, $A e_0 = -r_0$ and $e_0 - w = (I - Q_{l-1}(A)A) e_0$ and hence

$$\|x_* - x_l\|_A = \|e_0 - w_l\|_A = \min_{P_l \in \mathcal{P}_l;\ P_l(0)=1} \|P_l(A) e_0\|_A. \tag{4.2}$$

This completes the proof. □

To obtain a qualitative estimate on the right hand side of (4.2), we observe that for any polynomial $P_l(\lambda)$ we have

$$\|x_* - x_l\|_A = \min_{P_l \in \mathcal{P}_l;\ P_l(0)=1} \|P_l(A) e_0\|_A \leq \min_{P_l \in \mathcal{P}_l;\ P_l(0)=1} \rho(P_l(A)) \|e_0\|_A,$$

where $\rho(P_l(A))$ is the spectral radius of $P_l(A)$. Since both A and $P_l(A)$ have the same eigenvectors, we may conclude that

$$\|x_* - x_l\|_A \leq \min_{P_l \in \mathcal{P}_l;\ P_l(0)=1} \max_{1 \leq j \leq n} |P_l(\lambda_j)| \|e_0\|_A = c_l(\lambda_1, \ldots, \lambda_n) \|e_0\|$$

where $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ are the eigenvalues of A.

In the next section, we will derive a somewhat pessimistic upper bound on c_l by first estimating

$$c_l(\lambda_1, \ldots, \lambda_n) \leq \min_{P_l \in \mathcal{P}_l;\ P_l(0)=1} \|P_l\|_{\infty, [\lambda_1, \lambda_n]}$$

and then with the help of a construction based on the Chebyshev polynomials, we will find the value of the right side of the above inequality in terms of λ_1 and λ_n.

4.2. **Chebyshev polynomials and a convergence rate estimate.** The Chebyshev polynomials of first kind on $[-1, 1]$ are defined as

$$T_l(\xi) = \cos(l \arccos(\xi)), \quad l = 0, 1, \ldots$$

Using a simple trigonometric identity (with $\theta = \arccos(\xi)$) shows that

$$T_{l+1}(\xi) + T_{l-1}(\xi) = \cos(l+1)\theta + \cos(l-1)\theta = 2(\cos\theta)\cos l\theta.$$

Hence,

$$T_{l+1}(\xi) = 2\xi T_l(\xi) - T_{l-1}(\xi). \tag{4.3}$$

This proves that T_l are indeed polynomials, because $T_0(\xi) = 1$ and $T_1(\xi) = \xi$. The form (4.3) defnes $T_l(\xi)$ for all $\xi \in \mathbb{R}$. Another form of the Chebyshev polynomals, which will be useful in the convergence rate estimate given below in Theorem 4.4 is derived as follows: From the relation (4.3) for fixed ξ we observe that

$$T_l(\xi) = c_1 [\eta_1(\xi)]^l + c_2 [\eta_2(\xi)]^l, \quad l = 0, 1, \ldots,$$

where $\eta_1(\xi)$ and $\eta_2(\xi)$ are the roots of the characteristic equation

$$\eta^2 - 2\xi\eta + 1 = 0.$$

The constants c_1 and c_2 are easily computed from the initial conditions $T_0(\xi) = 1$ and $T_1(\xi) = \xi$ and hence

$$T_l(\xi) = \frac{1}{2}[(\xi + \sqrt{\xi^2 - 1})^l + (\xi - \sqrt{\xi^2 - 1})^l]. \tag{4.4}$$

We further have that $|T_l(\xi)| \leq 1$ for all $\xi \in [1, 1]$ and that

$$\text{If} \quad \xi_m = \cos(\frac{l\pi}{l}), \quad \text{then} \quad T_l(\xi_m) = (-1)^l, \quad m = 0, 1, \ldots, l. \tag{4.5}$$

Define now

$$S_l(\lambda) = \left[T_l\left(\frac{\lambda_n + \lambda_1}{\lambda_n - \lambda_1}\right)\right]^{-1} T_l\left(\frac{\lambda_n + \lambda_1 - 2\lambda}{\lambda_n - \lambda_1}\right)$$

Note that

$$\|S_l\|_{\infty,[\lambda_1,\lambda_n]} = \left|T_l\left(\frac{\lambda_n + \lambda_1}{\lambda_n - \lambda_1}\right)\right|^{-1}.$$

Next Lemma shows that S_l is a polynomial with minimum "max"-norm, that is,

$$\|S_l\|_{\infty,[\lambda_1,\lambda_n]} = \min_{P_l \in \mathcal{P}_l;\ P_l(0)=1} \|P_l\|_{\infty,[\lambda_1,\lambda_n]}.$$

Lemma 4.3. *For any $P_l \in \mathcal{P}_l$ with $P_l(0) = 1$,*

$$\|S_l\|_{\infty,[\lambda_1,\lambda_n]} \leq \|P_l\|_{\infty,[\lambda_1,\lambda_n]}$$

Proof. Denote

$$t_* = \left[T_l\left(\frac{\lambda_n + \lambda_1}{\lambda_n - \lambda_1}\right)\right]^{-1}.$$

Let $\mu_m = \frac{\lambda_1 - \lambda_n}{2}\xi_m + \frac{\lambda_n + \lambda_1}{2}$, where ξ_m are defined in (4.5). Note that

$$S_l(\mu_m) = (-1)^m t_*, \quad m = 0, \ldots, l,$$

and also that $\mu_m \in [\lambda_1, \lambda_n]$. Assume that there exists $P_l \in \mathcal{P}_l$ with $P_l(0) = 1$, such that

$$|P_l(\lambda)| < |t_*|, \quad \text{for all} \quad \lambda \in [\lambda_1, \lambda_n].$$

This in particular implies that

$$-|t_*| < P_l(\mu_m) < |t_*|, \quad m = 0, 1, \ldots, l.$$

If $\text{sign}(t_*) > 0$ then

$$P_l(\mu_m) - S_l(\mu_m) < 0, \quad \text{for } m \text{ even, and} \quad P_l(\mu_m) - S_l(\mu_m) > 0, \quad \text{for } m \text{ odd.}$$

On the other hand, the case $\text{sign}(t_*) < 0$, just switches "odd" with "even" and "even" with "odd" in the above inequalities. Hence, regardless of the sign of t_*, the difference $P_l - S_l$ has a zero in every interval (μ_m, μ_{m+1}). There are l such intervals. But we also have that $P_l(0) - S_l(0) = 0$. Since $P_l - S_l$ is a polynomial of degree at most l, it follows that $P_l \equiv S_l$, which is a contradiction. □

Clearly, from this lemma it follows that

$$\|x_* - x_l\|_A \le \|S_l\|_{\infty,[\lambda_1,\lambda_n]} \|x_* - x_0\|_A. \tag{4.6}$$

In the next Theorem 4.4 we obtain this estimate in terms of the condition number of A, by calculating $\|S_l\|_{\infty,[\lambda_1,\lambda_n]}$.

Theorem 4.4. *The error after l iterations of the CG algorithm can be bounded as follows:*

$$\|x_* - x_l\|_A \le \frac{2}{\left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}\right)^l + \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^l} \|x_* - x_0\|_A \le 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^l \|x_* - x_0\|_A, \tag{4.7}$$

where $\kappa = \kappa(A) = \lambda_n/\lambda_1$ is the condition number of A.

Proof. We aim to calculate $\|S_l\|_{\infty,[\lambda_1,\lambda_n]} = \left|T_l\left(\frac{\lambda_n+\lambda_1}{\lambda_n-\lambda_1}\right)\right|^{-1}$. From (4.4), for $\xi = \frac{\lambda_n+\lambda_1}{\lambda_n-\lambda_1} = \frac{\kappa+1}{\kappa-1}$, we obtain

$$\xi \pm \sqrt{\xi^2-1} = \frac{\kappa+1}{\kappa-1} \pm \frac{2\sqrt{\kappa}}{\kappa-1} = \frac{\kappa+1\pm 2\sqrt{\kappa}}{\kappa-1} = \frac{(\sqrt{\kappa}\pm 1)^2}{(\sqrt{\kappa}-1)(\sqrt{\kappa}+1)} = \frac{\sqrt{\kappa}\pm 1}{\sqrt{\kappa}\mp 1}.$$

Thus,

$$T_l\left(\frac{\lambda_n+\lambda_1}{\lambda_n-\lambda_1}\right) = \frac{1}{2}\left[\left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}\right)^l + \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^l\right].$$

Finally

$$\|S_l\|_{\infty,[\lambda_1,\lambda_n]} = \left|T_l\left(\frac{\lambda_n+\lambda_1}{\lambda_n-\lambda_1}\right)\right|^{-1} = \frac{2}{\left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}\right)^l + \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^l} \le 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^l.$$

The proof is completed by substituting the above expression in (4.6). □

References

[1] Douglas N. Arnold. A concise introduction to numerical analysis. Lecture Notes, Penn State, MATH 597I–Numerical Analysis, Fall 2001.

[2] Magnus R. Hestenes and Eduard Stiefel. Methods of conjugate gradients for solving linear systems. *J. Research Nat. Bur. Standards*, 49:409–436 (1953), 1952.

[3] David G. Luenberger. *Linear and nonlinear programming*. Kluwer Academic Publishers, Boston, MA, second edition, 2003.