Conjugate Gradient Method

Tsung-Ming Huang

Department of Mathematics
National Taiwan Normal University

October 10, 2011

Outline

1. Steepest Descent Method

2. Conjugate Gradient Method

3. Convergence of CG-method

A Variational Problem, Steepest Descent Method (Gradient Method)

Let $A \in \mathbb{R}^{n\times n}$ be a large and sparse symmetric positive definite (s.p.d.) matrix. Consider the linear system

$$Ax = b$$

and the functional $F : \mathbb{R}^n \to \mathbb{R}$ with

$$F(x) = \frac{1}{2}x^T Ax - b^T x. \tag{1}$$

Then it holds:

Theorem 1

For a vector x^* the following statements are equivalent:

$$(i) \quad F(x^*) < F(x), \ \textit{for all } x \neq x^*,$$
$$(ii) \quad Ax^* = b.$$

Proof: From assumption there exists $z_0 = A^{-1}b$ and $F(x)$ can be rewritten as

$$F(x) = \frac{1}{2}(x - z_0)^T A(x - z_0) - \frac{1}{2} z_0^T A z_0. \tag{2}$$

Since A is positive definite, $F(x)$ has a minimum at $x = z_0$ and only at $x = z_0$, it follows the assertion. ■

The solution of the linear system $Ax = b$ is equal to the solution of the minimization problem

$$\min F(x) \equiv \min \left(\frac{1}{2} x^T A x - b^T x \right).$$

Steepest Descent Method

Let x_k be an approximate of the exact solution x^* and p_k be a search direction. We want to find an α_k such that

$$F(x_k + \alpha_k p_k) < F(x_k).$$

Set $x_{k+1} := x_k + \alpha_k p_k$. This leads to the basic problem:

Given x, $p \neq 0$, find α_* such that

$$\Phi(\alpha_*) = F(x + \alpha_* p) = \min_{\alpha \in \mathbb{R}} F(x + \alpha p).$$

Solution: Since

$$\begin{aligned} F(x+\alpha p) &= \frac{1}{2}(x+\alpha p)^T A(x+\alpha p) - b^T(x+\alpha p) \\ &= \frac{1}{2}\alpha^2 p^T Ap + \alpha(p^T Ax - p^T b) + F(x), \end{aligned}$$

it follows that if we take

$$\alpha_* = \frac{(b-Ax)^T p}{p^T Ap} = \frac{r^T p}{p^T Ap}, \tag{3}$$

where $r = b - Ax = -\text{grad}F(x) =$ residual, then $x + \alpha_* p$ is the minimal solution. Moreover,

$$F(x+\alpha_* p) = F(x) - \frac{1}{2}\frac{(r^T p)^2}{p^T Ap}.$$

Lemma 2

Let

$$x_{k+1} = x_k + \frac{r_k^T p_k}{p_k^T A p_k} p_k, \; r_k = b - A x_k, \tag{4}$$

$$F(x_{k+1}) = F(x_k) - \frac{1}{2}\frac{(r_k^T p_k)^2}{p_k^T A p_k}, \; k = 0, 1, 2, \cdots . \tag{5}$$

Then, it holds

$$r_{k+1}^T p_k = 0. \tag{6}$$

Proof: Since

$$\frac{d}{d\alpha} F(x_k + \alpha p_k) = \text{grad} F(x_k + \alpha p_k)^T p_k,$$

it follows that $\text{grad} F(x_k + \alpha_{k+1} p_k)^T p_k = 0$ where $\alpha_{k+1} = \frac{r_k^T p_k}{p_k^T A p_k}$. Thus

$$(b - A x_{k+1})^T p_k = r_{k+1}^T p_k = 0,$$

hence (6) holds.

How to choose search direction p_k?

Let $\Phi : \mathbb{R}^n \to \mathbb{R}$ be a differential function on x. Then it holds

$$\frac{\Phi(x+\varepsilon p) - \Phi(x)}{\varepsilon} = \Phi^{'}(x)^T p + O(\varepsilon).$$

The right hand side takes minimum at $p = -\frac{\Phi^{'}(x)}{\|\Phi^{'}(x)\|}$ (i.e., the largest descent) for all p with $\|p\| = 1$ (neglect $O(\varepsilon)$). Hence, it suggests to choose

$$p_k = -\text{grad} F(x_k) = b - Ax_k = r_k.$$

Algorithm: Gradient Method

Give x_0 and set $k = 0$.
Compute $r_k = b - Ax_k$ and $\alpha_k = \frac{r_k^T r_k}{r_k^T A r_k}$;
repeat
    Compute $x_{k+1} = x_k + \alpha_k r_k$ and set $k := k + 1$;
    Compute $r_k = b - Ax_k$ and $\alpha_k = \frac{r_k^T r_k}{r_k^T A r_k}$.
until $r_k = 0$

Cost in each step: compute Ax_k (Ar_k does not need to compute).
To prove the convergence of Gradient method, we need the Kontorowitsch inequality:

Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0$, $\alpha_i \geq 0$, $\sum_{i=1}^{n} \alpha_i = 1$. Then it holds

$$\sum_{i=1}^{n} \alpha_i \lambda_i \sum_{j=1}^{n} \alpha_j \lambda_j^{-1} \leq \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1 \lambda_n} = \frac{1}{4} \left(\sqrt{\frac{\lambda_1}{\lambda_n}} + \sqrt{\frac{\lambda_n}{\lambda_1}} \right)^2 .$$

Theorem 3

If x_k, x_{k-1} are two approximations of Gradient Method Algorithm for solving $Ax = b$ and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0$ are the eigenvalues of A, then it holds:

$$F(x_k) + \frac{1}{2}b^T A^{-1} b \leq \left(\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}\right)^2 [F(x_{k-1}) + \frac{1}{2}b^T A^{-1} b], \tag{7a}$$

i.e.,

$$\|x_k - x^*\|_A \leq \left(\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}\right) \|x_{k-1} - x^*\|_A, \tag{7b}$$

where $\|x\|_A = \sqrt{x^T A x}$. Thus the gradient method is convergent.

It is favorable to choose that the search directions $\{p_i\}$ as mutually A-conjugate, where A is symmetric positive definite.

Definition 4

Two vectors p and q are called A-conjugate (A-orthogonal), if $p^T Aq = 0$.

Remark 1

Let A be symmetric positive definite. Then there exists a unique s.p.d. B such that $B^2 = A$. Denote $B = A^{1/2}$. Then $p^T Aq = (A^{1/2}p)^T(A^{1/2}q)$.

Lemma 5

Let $p_0, \ldots, p_r \neq 0$ be pairwisely A-conjugate. Then they are linearly independent.

Proof: From $0 = \sum_{j=0}^{r} c_j p_j$ follows that

$$0 = p_k^T A \left(\sum_{j=0}^{r} c_j p_j \right) = \sum_{j=0}^{r} c_j p_k^T A p_j = c_k p_k^T A p_k,$$

so $c_k = 0$, for $k = 1, \ldots, r$. ∎

Theorem 6

Let A *be s.p.d. and* $p_0, \ldots, p_{n-1}$ *be nonzero pairwisely* A*-conjugate vectors. Then*

$$A^{-1} = \sum_{j=0}^{n-1} \frac{p_j p_j^T}{p_j^T A p_j}. \tag{8}$$

Remark 2

$A = I$, $U = (p_0, \ldots, p_{n-1})$, $p_i^T p_i = 1$, $p_i^T p_j = 0$, $i \neq j$. $UU^T = I$ *and* $I = UU^T$. *Then*

$$I = (p_0, \ldots, p_{n-1}) \begin{bmatrix} p_0^T \\ \vdots \\ p_{n-1}^T \end{bmatrix} = p_0 p_0^T + \cdots + p_{n-1} p_{n-1}^T.$$

Proof of Theorem 6: Since $\tilde{p_j} = \frac{A^{1/2}p_j}{\sqrt{p_j^T A p_j}}$ are orthonormal, for $j = 0, 1, \ldots, n-1$, we have

$$\begin{aligned} I &= \tilde{p_0}\tilde{p_0}^T + \cdots + \tilde{p}_{n-1}\tilde{p}_{n-1}^T \\ &= \sum_{j=0}^{n-1} \frac{A^{1/2}p_j p_j^T A^{1/2}}{p_j^T A p_j} = A^{1/2}\left(\sum_{j=0}^{n-1} \frac{p_j p_j^T}{p_j^T A p_j}\right) A^{1/2}. \end{aligned}$$

Thus,

$$A^{-1/2} I A^{-1/2} = A^{-1} = \sum_{j=0}^{n-1} \frac{p_j p_j^T}{p_j^T A p_j}.$$

Remark 3

Let $Ax^* = b$ and x_0 be an arbitrary vector. Then from $x^* - x_0 = A^{-1}(b - Ax_0)$ and (8) follows that

$$x^* = x_0 + \sum_{j=0}^{n-1} \frac{p_j^T(b - Ax_0)}{(p_j^T A p_j)} p_j. \tag{9}$$

Theorem 7

Let A be s.p.d. and $p_0, \ldots, p_{n-1} \in \mathbb{R}^n \backslash \{0\}$ be pairwisely A-orthogonal. Given x_0 and let $r_0 = b - Ax_0$. For $k = 0, \ldots, n-1$, let

$$\alpha_k = \frac{p_k^T r_k}{p_k^T A p_k}, \quad (10)$$

$$x_{k+1} = x_k + \alpha_k p_k, \quad (11)$$

$$r_{k+1} = r_k - \alpha_k A p_k. \quad (12)$$

Then the following statements hold:

- (i) $r_k = b - Ax_k$. *(By induction).*
- (ii) x_{k+1} *minimizes* $F(x)$ *on* $x = x_k + \alpha p_k$, $\alpha \in \mathbb{R}$.
- (iii) $x_n = A^{-1} b = x^*$.
- (iv) x_k *minimizes* $F(x)$ *on the affine subspace* $x_0 + S_k$, *where* $S_k =$ *Span*$\{p_0, \ldots, p_{k-1}\}$.

Proof: (i): By Induction and using (11), (12).
(ii): From (3) and (i).
(iii): It is enough to show that x_k corresponds with the partial sum in (9),

$$x_k = x_0 + \sum_{j=0}^{k-1} \frac{p_j^T(b - Ax_0)}{p_j^T A p_j} p_j.$$

Then it follows that $x_n = x^*$ from (9). From (10) and (11) we have

$$x_k = x_0 + \sum_{j=0}^{k-1} \alpha_j p_j = x_0 + \sum_{j=0}^{k-1} \frac{p_j^T(b - Ax_j)}{p_j^T A p_j} p_j.$$

To show that

$$p_j^T(b - Ax_j) = p_j^T(b - Ax_0). \qquad (13)$$

From $x_k - x_0 = \sum_{j=0}^{k-1} \alpha_j p_j$ we obtain

$$p_k^T A x_k - p_k^T A x_0 = \sum_{j=0}^{k-1} \alpha_j p_k^T A p_j = 0.$$

So (13) holds.

(iv): From (12) and (10) follows that

$$p_k^T r_{k+1} = p_k^T r_k - \alpha_k p_k^T A p_k = 0.$$

From (11), (12) and by the fact that $r_{k+s} - r_{k+s+1} = \alpha_{k+s} A p_{k+s}$ and p_k are A-orthogonal (for $s \geq 1$) follows that

$$p_k^T r_{k+1} = p_k^T r_{k+2} = \cdots = p_k^T r_n = 0.$$

Hence we have

$$p_i^T r_k = 0, \; i = 0, \ldots, k-1, \; k = 1, 2, \ldots, n. \text{ (i.e., } i < k). \tag{14}$$

We now consider $F(x)$ on $x_0 + S_k$:

$$F(x_0 + \sum_{j=0}^{k-1} \xi_j p_j) = \varphi(\xi_0, \cdots, \xi_{k-1}).$$

$F(x)$ is minimal on $x_0 + S_k$ if and only if all derivatives $\frac{\partial \varphi}{\partial \xi_s}$ vanish at x. But

$$\frac{\partial \varphi}{\partial \xi_s} = [\text{grad} F(x_0 + \sum_{j=0}^{k-1} \xi_j p_j)]^T p_s, \; s = 0, 1, \ldots, k-1.$$

If $x = x_k$, then $\text{grad} F(x) = -r_k$. From (14) follows that

$$\frac{\partial \varphi}{\partial \xi_s}(x_k) = 0, \; \text{for } s = 0, 1, \ldots, k-1.$$

■

Remark 4

The following conditions are equivalent:

(i) $p_i^T A p_j = 0,\ i \neq j,$

(ii) $p_i^T r_k = 0,\ i < k,$

(iii) $r_i^T r_j = 0,\ i \neq j.$

Proof of (iii): for $i < k$,

$$\begin{aligned} p_i^T r_k = 0 &\Leftrightarrow (r_i^T + \beta_{i-1} p_{i-1}^T) r_k = 0 \\ &\Leftrightarrow r_i^T r_k = 0 \\ &\Leftrightarrow r_i^T r_j = 0, \quad i \neq j. \end{aligned}$$

Remark 5

It holds

$$\langle p_0, p_1, \cdots, p_k \rangle = \langle r_0, r_1, \cdots, r_k \rangle = \left\langle r_0, Ar_0, \cdots, A^k r_0 \right\rangle.$$

Since

$$\begin{aligned} p_1 &= r_1 + \beta_0 p_0 = r_1 + \beta_0 r_0, \\ r_1 &= r_0 - \alpha_0 A r_0, \end{aligned}$$

by induction, we have

$$\begin{aligned} r_2 &= r_1 - \alpha_0 A p_1 \\ &= r_1 - \alpha_0 A (r_1 + \beta_0 r_0) \\ &= r_0 - \alpha_0 A r_0 - \alpha_0 A (r_0 - \alpha_0 A r_0 + \beta_0 r_0). \end{aligned}$$

Method of conjugate directions

Input: Let A be s.p.d., b and $x_0 \in \mathbb{R}^n$. Given $p_0, \ldots, p_{n-1} \in \mathbb{R}^n \backslash \{0\}$ pairwisely A-orthogonal.

1: Compute $r_0 = b - Ax_0$;
2: **for** $k = 0, \ldots, n-1$ **do**
3: Compute $\alpha_k = \frac{p_k r_k}{p_k^T A p_k}$, $x_{k+1} = x_k + \alpha_k p_k$,;
4: Compute $r_{k+1} = r_k - \alpha_k A p_k = b - A x_{k+1}$.
5: **end for**

From Theorem 7 we get $x_n = A^{-1} b$.

Practical Implementation

Algorithm: Conjugate Gradient method (CG-method)

Input: Given s.p.d. A, $b \in \mathbb{R}^n$ and $x_0 \in \mathbb{R}^n$ and $r_0 = b - Ax_0 = p_0$.

Set $k = 0$.
repeat
    Compute $\alpha_k = \frac{p_k^T r_k}{p_k^T A p_k}$;
    Compute $x_{k+1} = x_k + \alpha_k p_k$;
    Compute $r_{k+1} = r_k - \alpha_k A p_k = b - A x_{k+1}$;
    Compute $\beta_k = \frac{-r_{k+1}^T A p_k}{p_k^T A p_k}$;
    Compute $p_{k+1} = r_{k+1} + \beta_k p_k$;
    Set $k = k + 1$;
until $r_k = 0$

Theorem 8

The CG-method holds

(i) *If k steps of CG-method are executable, i.e., $r_i \neq 0$, for $i = 0, \ldots, k$, then $p_i \neq 0$, $i \leq k$ and*

$$p_i^T A p_j = 0 \quad \text{for } i, j \leq k, \; i \neq j.$$

(ii) *The CG-method breaks down after N steps for $r_N = 0$ and $N \leq n$.*

(iii) $x_N = A^{-1} b$.

Proof:

(i): By induction on k, it is trivial for $k = 0$. Suppose that (i) is true until k and $r_{k+1} \neq 0$. Then p_{k+1} is well-defined. We want to verify

$$p_{k+1}^T A p_j = 0, \quad \text{for } j = 0, 1, \ldots, k.$$

From Lines 6 and 7 in CG Algorithm, we have

$$p_{k+1}^T A p_k = r_{k+1}^T A p_k + \beta_k p_k^T A p_k = 0.$$

Let $j < k$, from Line 7 we have

$$p_{k+1}^T A p_j = r_{k+1}^T A p_j + \beta_k p_k^T A p_j = r_{k+1}^T A p_j.$$

It is enough to show that

$$A p_j \in \text{span}\{p_0, ..., p_{j+1}\}, \; j < k. \tag{15}$$

Then from the relation $p_i^T r_j = 0$, $i < j \leq k+1$, which has been proved in (14), follows assention.

Claim (15): For $r_j \neq 0$, it holds that $\alpha_j \neq 0$. Line 5shows that

$$Ap_j = \frac{1}{\alpha_j}(r_j - r_{j+1}) \in \text{span}\{r_0, ..., r_{j+1}\}.$$

Line 7 shows that $\text{span}\{r_0, ..., r_{j+1}\} = \text{span}\{p_0, ..., p_{j+1}\}$ with $r_0 = p_0$, so is (15).
(ii): Since $\{p_i\}_{i=0}^{k+1} \neq 0$ and are mutually A-orthogonal, $p_0, ..., p_{k+1}$ are linearly independent. Hence there exists a $N \leq n$ with $r_N = 0$. This follows $x_N = A^{-1}b$. ■

Advantage

1. Break-down in finite steps.
2. Less cost in each step: one matrix $\times$ vector.

Convergence of CG-method

Consider the A-norm with A being s.p.d.

$$\|x\|_A = (x^T Ax)^{1/2}.$$

Let $x^* = A^{-1}b$. Then from (2) we have

$$F(x) - F(x^*) = \frac{1}{2}(x - x^*)^T A(x - x^*) = \frac{1}{2}\|x - x^*\|_A^2,$$

where x_k is the k-th iterate of CG-method. From Theorem 7 x_k minimizes the functional F on x_0+ span$\{p_0, ..., p_{k-1}\}$. Hence it holds

$$\|x_k - x^*\|_A \leq \|y - x^*\|_A, \ \ y \in x_0 + \text{span}\{p_0, .., p_{k-1}\}. \tag{16}$$

From Lines 5 and 7 in CG Algorithm it is easily seen that both p_k and r_k can be written as linear combination of r_0, $Ar_0, \ldots, A^{k-1}r_0$.
If $y \in x_0 + \text{span}\{p_0, \ldots, p_{k-1}\}$, then

$$y = x_0 + c_1 r_0 + c_2 A r_0 + \ldots.. + c_k A^{k-1} r_0 = x_0 + \tilde{\mathcal{P}}_{k-1}(A) r_0,$$

where $\tilde{\mathcal{P}}_{k-1}$ is a polynomial of degree $\leq k-1$. But $r_0 = b - Ax_0 = A(x^* - x_0)$, thus

$$\begin{aligned} y - x^* &= (x_0 - x^*) + \tilde{\mathcal{P}}_{k-1}(A) A (x^* - x_0) \\ &= \left[I - A \tilde{\mathcal{P}}_{k-1}(A) \right] (x_0 - x^*) = \mathcal{P}_k(A)(x_0 - x^*), \end{aligned} \tag{17}$$

where degree $\mathcal{P}_k \leq k$ and

$$\mathcal{P}_k(0) = 1. \tag{18}$$

Conversely, if $\mathcal{P}_k$ is a polynomial of degree $\leq k$ and satisfies (18), then

$$x^* + \mathcal{P}_k(A)(x_0 - x^*) \in x_0 + S_k.$$

Hence (16) means that if $\mathcal{P}_k$ is a polynomial of degree $\leq k$ with $\mathcal{P}_k(0) = 1$, then

$$\|x_k - x^*\|_A \leq \|\mathcal{P}_k(A)(x_0 - x^*)\|_A.$$

Lemma 9

Let A be s.p.d. It holds for every polynominal Q_k of degree k that

$$\max_{x \neq 0} \frac{\|Q_k(A)x\|_A}{\|x\|_A} = \rho(Q_k(A)) = \max\{|Q_k(\lambda)| : \lambda \text{ eigenvalue of } A\}. \tag{19}$$

▸ Proof

From (19) we have that

$$\|x_k - x^*\|_A \leq \rho(\mathcal{P}_k(A))\|x_0 - x^*\|_A, \tag{20}$$

where degree $\mathcal{P}_k \leq k$ and $\mathcal{P}_k(0) = 1$.

Replacement problem for (20): For $0 < a < b$,

$$\min\{\max\{|\mathcal{P}_k(\lambda)| : a \leq \lambda \leq b, \forall \deg(\mathcal{P}_k(\lambda)) \leq k \text{ with } \mathcal{P}_k(0) = 1\}\} \tag{21}$$

We use Chebychev poly. of the first kind for the solution. They are defined by

$$\begin{cases} T_0(t) = 1, T_1(t) = t, \\ T_{k+1}(t) = 2tT_k(t) - T_{k-1}(t). \end{cases}$$

it holds $T_k(\cos\phi) = \cos(k\phi)$ by using
$\cos((k+1)\phi) + \cos((k-1)\phi) = 2\cos\phi\cos k\phi$. Especially,

$$T_k(\cos\frac{j\pi}{k}) = \cos(j\pi) = (-1)^j, \text{ for } j = 0, \ldots, k,$$

i.e. T_k takes maximal value "one" at $k+1$ positions in $[-1,1]$ with alternating sign. In addition (Exercise!), we have

$$T_k(t) = \frac{1}{2}\left[(t+\sqrt{t^2-1})^k + (t-\sqrt{t^2-1})^k\right]. \tag{22}$$

Lemma 10

The solution of the problem (21) is given by

$$Q_k(t) = T_k\left(\frac{2t-a-b}{b-a}\right) \Big/ T_k\left(\frac{a+b}{a-b}\right),$$

i.e., for all $\mathcal{P}_k$ of degree $\leq k$ with $\mathcal{P}_k(0) = 1$ it holds

$$\max_{\lambda\in[a,b]} |Q_k(\lambda)| \leq \max_{\lambda\in[a,b]} |\mathcal{P}_k(\lambda)|.$$

Proof: $Q_k(0) = 1$. If t runs through the interval $[a, b]$, then $(2t - a - b)/(b - a)$ runs through the interval $[-1, 1]$. Hence, in $[a, b]$, $Q_k(t)$ has $k + 1$ extreme with alternating sign and absolute value $\delta = |T_k(\frac{a+b}{a-b})^{-1}|$.

If there are a $\mathcal{P}_k$ with

$$\max\{|\mathcal{P}_k(\lambda)| : \lambda \in [a, b]\} < \delta,$$

then $Q_k - \mathcal{P}_k$ has the same sign as Q_k of the extremal values, so $Q_k - \mathcal{P}_k$ changes sign at $k + 1$ positions. Hence $Q_k - \mathcal{P}_k$ has k roots, in addition a root zero. This contradicts that degree $(Q_k - \mathcal{P}_k) \leq k$. ■

Lemma 11

It holds

$$\delta = \left| T_k \left(\frac{b+a}{a-b} \right)^{-1} \right| = \frac{1}{T_k\left(\frac{b+a}{b-a}\right)} = \frac{2c^k}{1+c^{2k}} \leq 2c^k,$$

where $c = \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}$ *and* $\kappa = b/a$.

▸ Proof

Theorem 12

CG-method satisfies the following error estimate

$$\|x_k - x^*\|_A \leq 2c^k \|x_0 - x^*\|_A,$$

where $c = \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}$, $\kappa = \frac{\lambda_1}{\lambda_n}$ *and* $\lambda_1 \geq \cdots \geq \lambda_n > 0$ *are the eigenvalues of* A.

▸ Proof

Remark 6

To compare with Gradient method (see (7b)): Let x_k^G be the kth iterate of Gradient method. Then

$$\|x_k^G - x^*\|_A \le \left|\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}\right|^k \|x_0 - x^*\|_A.$$

But

$$\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n} = \frac{\kappa - 1}{\kappa + 1} > \frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} = c,$$

because in general $\sqrt{\kappa} \ll \kappa$. Therefore the CG-method is much better than Gradient method.

Appendix

Proof:

$$
\begin{aligned}
\frac{\|Q_k(A)x\|_A^2}{\|x\|_A^2} &= \frac{x^T Q_k(A) A Q_k(A) x}{x^T A x} \\
&= \frac{(A^{1/2}x)^T Q_k(A) Q_k(A)(A^{1/2}x)}{(A^{1/2}x)(A^{1/2}x)} \qquad (\text{Let } z := A^{1/2}x) \\
&= \frac{z^T Q_k(A)^2 z}{z^T z} \leq \rho(Q_k(A)^2) = \rho^2(Q_k(A)).
\end{aligned}
$$

Equality holds for suitable x, hence the first equality is shown. The second equality holds by the fact that $Q_k(\lambda)$ is an eigenvalue of $Q_k(A)$, where λ is an eigenvalue of A. ■

return

Proof: For $t = \frac{b+a}{b-a} = \frac{\kappa+1}{\kappa-1}$, we compute

$$t + \sqrt{t^2 - 1} = \frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1} = c^{-1}$$

and

$$t - \sqrt{t^2 - 1} = \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1} = c.$$

Hence from (22) follows

$$\delta = \frac{2}{c^k + c^{-k}} = \frac{2c^k}{1 + c^{2k}} \leq 2c^k.$$

■

return

Proof: From (20) we have

$$\begin{aligned}\|x_k - x^*\|_A &\leq \rho\left(\mathcal{P}_k(A)\right)\|x_0 - x^*\|_A \\ &\leq \max\left\{|\mathcal{P}_k(\lambda)| : \lambda_1 \geq \lambda \geq \lambda_n\right\}\|x_0 - x^*\|_A,\end{aligned}$$

for all $\mathcal{P}_k$ of degree $\leq k$ with $\mathcal{P}_k(0) = 1$. From Lemma 10 and Lemma 11 follows that

$$\begin{aligned}\|x_k - x^*\|_A &\leq \max\left\{|Q_k(\lambda)| : \lambda_1 \geq \lambda \geq \lambda_n\right\}\|x_0 - x^*\|_A \\ &\leq 2c^k\|x_0 - x^*\|_A.\end{aligned}$$

■

return