On the next-to-last CG and MR iteration step

(with application to symmetric tridiagonal Toeplitz matrices)

Petr Tichý[*,†]

joint work with

Jörg Liesen[*]

[*]Technical University of Berlin,

[†]Institute of Computer Science AS CR.

March 31, 2005. GAMM 2005, Luxembourg.

A system of linear algebraic equations

Consider a system of linear algebraic equations

$$\mathbf{A}x = b$$

$\mathbf{A} \in \mathbb{R}^{n \times n}$ is nonsingular and normal, $b \in \mathbb{R}^n$.

- How to construct an approximation to the solution?

A system of linear algebraic equations

Consider a system of linear algebraic equations

$$\mathbf{A}x = b$$

$\mathbf{A} \in \mathbb{R}^{n\times n}$ is nonsingular and normal, $b \in \mathbb{R}^n$.

- How to construct an approximation to the solution?

Krylov subspace methods $\mapsto$ Given $x_0 \in \mathbb{R}^n$, $r_0 = b - \mathbf{A}x_0$. Find x_i,

$$x_i \in x_0 + \mathcal{K}_i(\mathbf{A}, r_0) \qquad \texttt{such that} \qquad r_i \perp \mathcal{C}_i,$$

where $r_i = b - \mathbf{A}x_i$, $\mathcal{K}_i(\mathbf{A}, r_0) \equiv \mathrm{span}\,\{r_0, \cdots, \mathbf{A}^{i-1}r_0\}$.

OR and MR Krylov subspace methods

Let $x_0 = 0$, i.e. $r_0 = b - \mathbf{A}x_0 = b$ (for simplicity).

Orthogonal Residual (OR) and Minimal Residual (MR) approach

(OR) `Find` $x_i \in \mathcal{K}_i(\mathbf{A}, b)$ `such that` $r_i \perp \mathcal{K}_i(\mathbf{A}, b)$.

(MR) `Find` $x_i \in \mathcal{K}_i(\mathbf{A}, b)$ `such that` $r_i \perp \mathbf{A}\mathcal{K}_i(\mathbf{A}, b)$.

OR and MR Krylov subspace methods

Let $x_0 = 0$, i.e. $r_0 = b - \mathbf{A}x_0 = b$ (for simplicity).

Orthogonal Residual (OR) and Minimal Residual (MR) approach

(OR) `Find` $x_i \in \mathcal{K}_i(\mathbf{A}, b)$ `such that` $r_i \perp \mathcal{K}_i(\mathbf{A}, b)$.

(MR) `Find` $x_i \in \mathcal{K}_i(\mathbf{A}, b)$ `such that` $r_i \perp \mathbf{A}\mathcal{K}_i(\mathbf{A}, b)$.

Optimality properties

(OR) $$\|e_i\|_{\mathbf{A}} = \min_{p \in \pi_i} \|p(\mathbf{A})x\|_{\mathbf{A}}$$ `(if A is SPD)`,

(MR) $$\|r_i\| = \min_{p \in \pi_i} \|p(\mathbf{A})b\|,$$

where $e_i \equiv x - x_i$, $\pi_i \equiv \{\, p$ is a polynomial; $\deg(p) \leq i;\ p(0) = 1 \}$.

The problem of convergence

MR constructs approximations $x_i \in \mathcal{K}_i(\mathbf{A}, b)$ to the solution x of the system $\mathbf{A}x = b$ such that

$$\|r_i\| = \min_{p \in \pi_i} \|p(\mathbf{A})\, b\| \,.$$

- **Our aim:**

  Description and understanding of this minimization process.

- **Considered classes of matrices in this talk:**

  Normal matrices, symmetric and positive definite matrices, symmetric positive definite tridiagonal Toeplitz matrices.

- We denote the OR method for SPD matrices as the CG method (Conjugate Gradient).

Outline

1. Introduction
2. Convergence bounds
3. Formulas for the next-to-last CG and MR iteration step
4. Application to symmetric tridiagonal Toeplitz matrices
5. Model problem: 1D Poisson equation
6. Conclusions

Convergence of the MR method

Let $\mathbf{A}$ be normal, $L \equiv \{\lambda_1, \ldots, \lambda_n\}$, $\|b\| = 1$. Then

$$\|r_i\| \quad = \quad \min_{p \in \pi_i} \|p(\mathbf{A})b\|$$

Convergence of the MR method

Let $\mathbf{A}$ be normal, $L \equiv \{\lambda_1, \ldots, \lambda_n\}$, $\|b\| = 1$. Then

$$\begin{aligned} \|r_i\| &= \min_{p \in \pi_i} \|p(\mathbf{A})b\| \\ &\leq \max_{\|b\|=1} \min_{p \in \pi_i} \|p(\mathbf{A})b\| \qquad \text{(worst-case)} \\ &= \min_{p \in \pi_i} \|p(\mathbf{A})\| \\ &= \min_{p \in \pi_i} \max_{\lambda_j \in L} |p(\lambda_j)| \end{aligned}$$

Convergence of the MR method

Let $\mathbf{A}$ be normal, $L \equiv \{\lambda_1, \ldots, \lambda_n\}$, $\|b\| = 1$. Then

$$\begin{aligned} \|r_i\| &= \min_{p \in \pi_i} \|p(\mathbf{A})b\| \\ &\leq \max_{\|b\|=1} \min_{p \in \pi_i} \|p(\mathbf{A})b\| \qquad \text{(worst-case)} \\ &= \min_{p \in \pi_i} \|p(\mathbf{A})\| \\ &= \min_{p \in \pi_i} \max_{\lambda_j \in L} |p(\lambda_j)| \end{aligned}$$

- In this sense we understand the MR-CG worst-case behaviour.
- How to describe $\|r_i\|$ or the worst-case bound in terms of input data?

General formula for the MR residual

Krylov matrix

$$\mathbf{K}_{i+1} \equiv [b, \mathbf{A}b, \ldots, \mathbf{A}^i b] \, .$$

Residual r_i can be written as (Assumption: $\mathbf{K}_{i+1}$ has full column rank)

$$r_i = \|r_i\|^2 \, (\mathbf{K}_{i+1}^{+})^H \, e_1 \quad \Rightarrow \quad \|r_i\| = \frac{1}{\|(\mathbf{K}_{i+1}^{+})^H \, e_1\|} \, .$$

[Ipsen ’00, Liesen & Rozložník & Strakoš ’02]

General formula for the MR residual

Krylov matrix

$$\mathbf{K}_{i+1} \;\equiv\; [b, \mathbf{A}b, \ldots, \mathbf{A}^i b]\,.$$

Residual r_i can be written as (Assumption: $\mathbf{K}_{i+1}$ has full column rank)

$$r_i \;=\; \|r_i\|^2 \, (\mathbf{K}_{i+1}^{+})^H \, e_1 \quad \Rightarrow \quad \|r_i\| \;=\; \frac{1}{\|(\mathbf{K}_{i+1}^{+})^H \, e_1\|}\,.$$

[Ipsen ’00, Liesen & Rozložník & Strakoš ’02]

We consider $\mathbf{A}$ and b in the form

$$\mathbf{A} \;=\; \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^H, \qquad b \;=\; \mathbf{Q}\,[\varrho_1, \ldots, \varrho_n]^T\,.$$

We will assume that all eigenvalues of $\mathbf{A}$ are **distinct**.

The next-to-last MR iteration step

Let $\varrho_j \neq 0$ for all j. Then

$$\|r_{n-1}\| = \left(\sum_{j=1}^{n} \left| \frac{l_j}{\varrho_j} \right|^2 \right)^{-1/2}, \qquad l_j \equiv \prod_{\substack{k=1 \\ k \neq j}}^{n} \frac{\lambda_k}{\lambda_k - \lambda_j}.$$

[Liesen & T. '04, Ipsen '00]

The next-to-last MR iteration step

Let $\varrho_j \neq 0$ for all j. Then

$$\|r_{n-1}\| = \left(\sum_{j=1}^{n} \left| \frac{l_j}{\varrho_j} \right|^2 \right)^{-1/2}, \qquad l_j \equiv \prod_{\substack{k=1 \\ k \neq j}}^{n} \frac{\lambda_k}{\lambda_k - \lambda_j}.$$

[Liesen & T. ’04, Ipsen ’00]

Using Cauchy’s inequality,

$$\frac{\|r_{n-1}^w\|}{\|b_{MR}^w\|} = \left(\sum_{j=1}^{n} |l_j| \right)^{-1},$$

where

$$b_{MR}^w = \mathbf{Q} \, [\varrho_1^w, \ldots, \varrho_n^w]^T, \qquad |\varrho_k^w|^2 = \gamma \, |l_k|, \quad k = 1, \ldots, n \, ,$$

$\gamma > 0$ is any scaling factor. [Liesen & T. ’04]

The next-to-last CG iteration step

CG can be seen as MR for a special right hand side $\tilde{b}$,

$$\min_{p\in\pi_i} \|p(\mathbf{A})x\|_{\mathbf{A}} = \min_{p\in\pi_i} \|p(\mathbf{A})\mathbf{A}^{1/2}x\| = \min_{p\in\pi_i} \|p(\mathbf{A})\tilde{b}\| \,.$$

The next-to-last CG iteration step

CG can be seen as MR for a special right hand side $\tilde{b}$,

$$\min_{p \in \pi_i} \|p(\mathbf{A})x\|_{\mathbf{A}} = \min_{p \in \pi_i} \|p(\mathbf{A})\mathbf{A}^{1/2}x\| = \min_{p \in \pi_i} \|p(\mathbf{A})\tilde{b}\| \,.$$

Then

$$\|e_{n-1}\|_{\mathbf{A}} \;=\; \left(\sum_{j=1}^{n} \left| \frac{\lambda_j^{1/2}\, l_j}{\varrho_j} \right|^2 \right)^{-1/2}, \qquad \frac{\|e_{n-1}^w\|_{\mathbf{A}}}{\|x_{CG}^w\|_{\mathbf{A}}} \;=\; \left(\sum_{j=1}^{n} |l_j| \right)^{-1},$$

$$b_{CG}^w \;=\; \mathbf{Q}\,[\varrho_1^w, \ldots, \varrho_n^w]^T, \qquad |\varrho_k^w|^2 = \gamma\, |\lambda_k\, l_k|, \quad k = 1, \ldots, n\,,$$

$\gamma > 0$ is any scaling factor.

[Liesen & T. '05]

Summary (the next-to-last iteration step)

The next-to-last step of CG and MR is completely understood!

We know

- the convergence quantities,
- the worst-case convergence quantities and corresponding b^w.

Summary (the next-to-last iteration step)

The next-to-last step of CG and MR is completely understood!

We know

- the convergence quantities,
- the worst-case convergence quantities and corresponding b^w.

How to use this information?

- We can study the influence of the right hand side,
- we can compare the worst-case bound with classical bounds,
- we can determine right hand sides leading to the slowest convergence and identify the worst input data of our original problem.

Symmetric tridiagonal Toeplitz matrices

Consider linear algebraic systems $\mathbf{A}x = b$, where

$$\mathbf{A} = \begin{bmatrix} \alpha & \beta & & \\ \beta & \ddots & \ddots & \\ & \ddots & \ddots & \beta \\ & & \beta & \alpha \end{bmatrix} \in \mathbb{R}^{n \times n}.$$

Let α and β be such that $\mathbf{A}$ is **symmetric and positive definite** matrix.

Symmetric tridiagonal Toeplitz matrices

Consider linear algebraic systems $\mathbf{A}x = b$, where

$$\mathbf{A} = \begin{bmatrix} \alpha & \beta & & \\ \beta & \ddots & \ddots & \\ & \ddots & \ddots & \beta \\ & & \beta & \alpha \end{bmatrix} \in \mathbb{R}^{n \times n}.$$

Let α and β be such that $\mathbf{A}$ is **symmetric and positive definite** matrix.

Eigenvalues and eigenvectors of $\mathbf{A}$ are known

$$\begin{aligned} \lambda_k &= \alpha + 2\,\beta\,\cos(k\pi h)\,, \\ q_k &= (2h)^{1/2}\ \left[\sin\left(k\pi h\right), \sin\left(2k\pi h\right), \ldots, \sin\left(nk\pi h\right)\right]^T, \end{aligned}$$

where $h \equiv (n+1)^{-1}$.

Worst-case bound versus classical κ-bound

Now we are able to determine l_j and the worst-case bound

$$\frac{\|e_{n-1}^w\|_A}{\|e_0^w\|_A} = \left(\sum_{j=1}^{n} |l_j| \right)^{-1} \approx \frac{2\,\nu^{n-1}}{1+\nu^2+\cdots+\nu^{2(n-1)}+\nu^{2n}},$$

where

$$\nu \equiv \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}, \quad \kappa = \frac{\lambda_{max}}{\lambda_{min}}.$$

Worst-case bound versus classical κ-bound

Now we are able to determine l_j and the worst-case bound

$$\frac{\|e_{n-1}^w\|_A}{\|e_0^w\|_A} = \left(\sum_{j=1}^{n} |l_j|\right)^{-1} \approx \frac{2\,\nu^{n-1}}{1+\nu^2+\cdots+\nu^{2(n-1)}+\nu^{2n}},$$

where

$$\nu \;\equiv\; \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}, \quad \kappa = \frac{\lambda_{max}}{\lambda_{min}}.$$

The classical κ-bound is given by

$$\frac{\|e_{n-1}^w\|_A}{\|e_0^w\|_A} \;\leq\; \frac{2\nu^{n-1}}{1+\nu^{2(n-1)}} \;\leq\; 2\nu^{n-1}.$$

Model problem: Poisson equation

$$-u''(z) \;=\; f(z), \qquad z \in (0,1)\,, \qquad u(0) = u_0, \;\; u(1) = u_1\,.$$

The central finite difference approximation on the uniform grid kh, $k = 1, \ldots, n$, $h = 1/(n+1)$, leads to a system $\mathbf{A}x = b$

$$\overbrace{\begin{bmatrix} 2 & -1 & & \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{bmatrix}}^{\mathbf{A}} x \;=\; \overbrace{h^2 \begin{bmatrix} f(h) \\ \vdots \\ \vdots \\ f(nh) \end{bmatrix} + \begin{bmatrix} u_0 \\ \\ \\ u_1 \end{bmatrix}}^{b}.$$

The eigenvalues λ_k and the eigenvectors q_k of $\mathbf{A}$ are known,

$$\lambda_k = 4\sin^2\left(\frac{k\pi h}{2}\right) \quad \Rightarrow \quad l_j = 2\cos^2\left(\frac{j\pi h}{2}\right).$$

[Liesen & T. '05]

Various right hand sides

Formulas for the next-to-last step

$$\|r_{n-1}\| = \left(\sum_{j=1}^{n} \left| \frac{l_j}{\varrho_j} \right|^2 \right)^{-1/2}, \qquad \|e_{n-1}\|_{\mathbf{A}} = \left(\sum_{j=1}^{n} \left| \frac{\lambda_j^{1/2}\, l_j}{\varrho_j} \right|^2 \right)^{-1/2}.$$

We consider two types of right hand sides:

- worst-case b's: right hand sides leading to maximal relative convergence quantities in the next-to-last step $\rightarrow b^w_{MR}$, b^w_{CG}.
- unbiased $b \rightarrow b^u$, all ϱ_j are of equal size.

Some results for MR

Let $\|b^w_{MR}\| = \|b^u\| = 1.$ [Liesen & T. '05]

Worst-case $\times$ unbiased case (MR)

$$\|r^w_{n-1}\| \;=\; \frac{1}{n}\,, \qquad \|r^u_{n-1}\| \;>\; \sqrt{\frac{2}{3}}\,\frac{1}{n}\,.$$

Some results for MR

Let $\|b^w_{MR}\| = \|b^u\| = 1$. [Liesen & T. '05]

Worst-case $\times$ unbiased case (MR)

$$\|r^w_{n-1}\| = \frac{1}{n}, \qquad \|r^u_{n-1}\| > \sqrt{\frac{2}{3}}\,\frac{1}{n}.$$

Worst data for MR

$$u(0) = 0,\ u(1) = 0,\ f(z) \approx \cot\left(\frac{\pi z}{2}\right)$$

yield a worst right-hand side for MR.

Worst data for CG

We are able to determine worst-case ϱ_{CG}^{w},

$$\varrho_{CG}^{w} \qquad b_{CG}^{w} = \mathbf{Q}\varrho_{CG}^{w} \qquad \text{diff. eq.}$$

$$\begin{bmatrix} \sin(\pi h) \\ \vdots \\ \sin(n\pi h) \end{bmatrix} \quad \leftrightarrow \quad \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \leftrightarrow \quad \begin{aligned} f(z) &= 0\,, \\ u(0) &= 1\,, \\ u(1) &= 0\,. \end{aligned}$$

CG started with $x_0 = 0$ and b_{CG}^{w} attains the worst-case relative A-norm of the error in the $(n-1)$st iteration step.

Worst data for CG

We are able to determine worst-case ϱ_{CG}^{w},

$$\varrho_{CG}^{w} \qquad b_{CG}^{w} = \mathbf{Q}\varrho_{CG}^{w} \qquad \text{diff. eq.}$$

$$\begin{bmatrix} \sin(\pi h) \\ \vdots \\ \sin(n\pi h) \end{bmatrix} \leftrightarrow \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \leftrightarrow \begin{aligned} f(z) &= 0\,, \\ u(0) &= 1\,, \\ u(1) &= 0\,. \end{aligned}$$

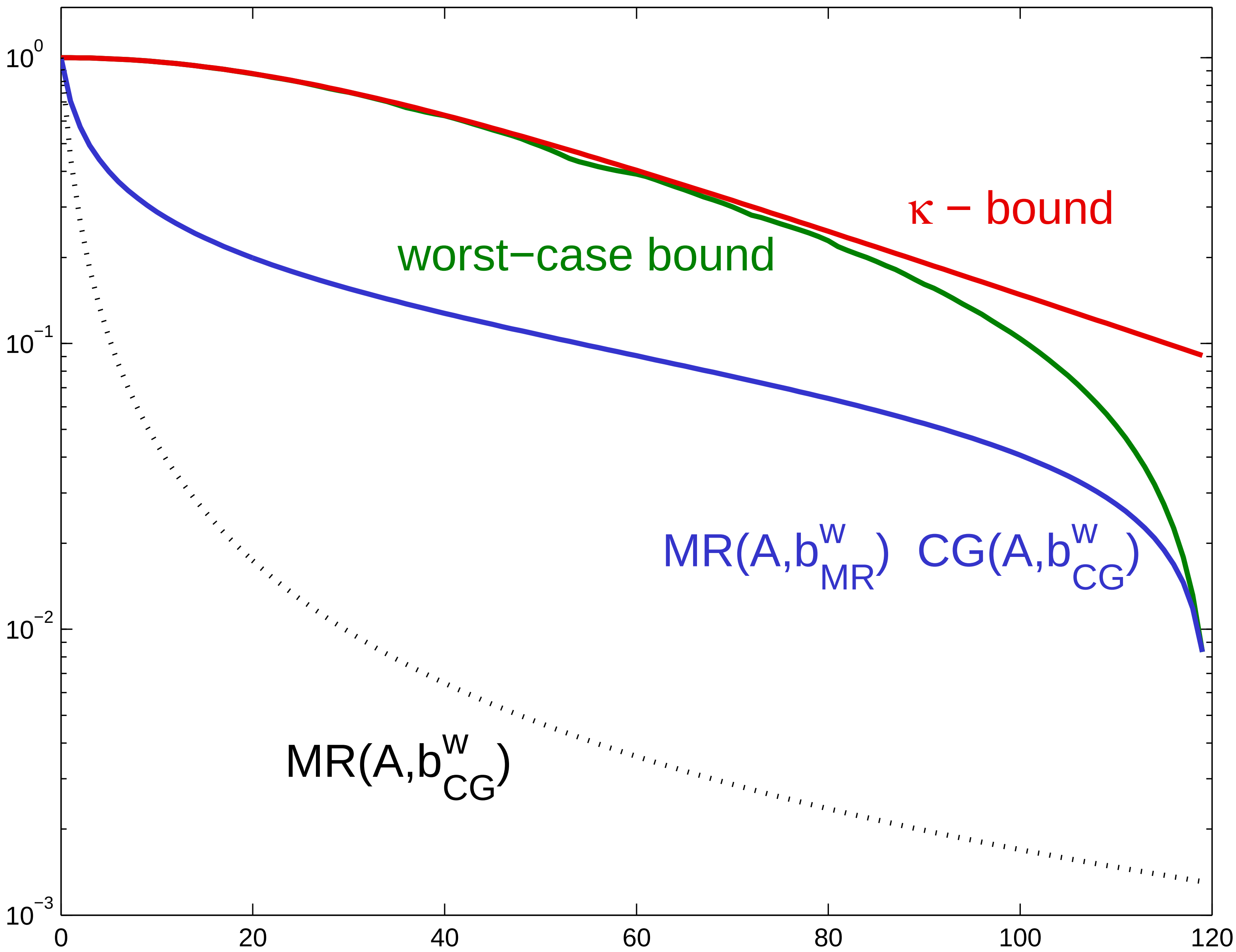
CG started with $x_0 = 0$ and b_{CG}^{w} attains the worst-case relative A-norm of the error in the $(n-1)$st iteration step.

Another example: Let n be even.

$$u''(z) = 0\,,\ u(0) = 1\,,\ u(1) = 1 \quad \Longrightarrow \quad b = [1, 0, \ldots, 0, 1]^T.$$

Then $\|x - x_{n/2}\|_A / \|x\|_A$ is the worst possible one and CG finds the solution in the following step. [Liesen & T. '05]

Numerical experiment



Conclusions

[Liesen & T. '05]

For MR ($\mathbf{A}$ is normal) and for CG ($\mathbf{A}$ is SPD):

- **The next-to-last iteration step is completely understood!**
- Our results allow to study **model problems with known eigenvalues** .
- For **1-D Poison equation** we obtained interesting results:

  $\longrightarrow$ particular worst-case quantities in the next-to-last step,

  $\longrightarrow$ implications for the connection between the differential equation and the linear solver for the discretized problem.

Thank you for your attention!

Thank you for your attention!

More details can be found in

Liesen, J. and Tichý, P., **The worst-case GMRES for normal matrices**, BIT Numerical Mathematics, Volume 44, pp. 79-98, 2004.

Liesen, J. and Tichý, P., **On the next-to-last CG and MR iteration step**, submitted to ETNA, January 2005.

See also `http://www.math.tu-berlin.de/~tichy`