Minimizing Methods

There are several generalizations of CG (and CR) for nonsymmetric linear problems.

The most obvious approach is to apply CG to an equivalent linear systems, which is symmetric and positive definite, e.g.

$$A^T A x = A^T b$$

or

$$A A^T y = b, \quad x = A^T y.$$

These methods are called CGN method where the letter 'N' suggests that some sort of normal equations is considered.

For $A^T A x = A^T b$ the k-te CG iterate x^k minimizes the functional

$$\|x - x^*\|^2_{A^T A} = (x - x^*)^T A^T A (x - x^*) = \|A(x - x^*)\|_2^2 = \|r\|_2^2$$

in the Krylov space

$$\begin{aligned} & x^0 + \mathcal{K}_k(A^T r^0, A^T A) \\ = \quad & x^0 + \text{span}\{A^T r^0, (A^T A) A^T r^0, \ldots, (A^T A)^{k-1} A^T r^0\}. \end{aligned}$$

This method is called CGNR Verfahren where the letter 'R' indicates that the residuum is minimized.

The CGNR method was suggested by Hesteness und Stiefel (1953). The matrix $A^T A$ in general is more populated than A, however, the algorithm does not take advantage of its explicit form.

CGNR

$r^0 = b - Ax^0;$
$d^1 = A^T r^0;$
$\alpha_0 = \|d^1\|_2^2;$
for $k = 1, 2, \dots$ until convergence
  $s^k = Ad^k;$
  $\tau_k = \alpha_{k-1}/\|s^k\|_2^2;$
  $x^k = x^{k-1} + \tau_k d^k;$
  $r^k = r^{k-1} - \tau_k s^k;$
  $\beta_k = 1/\alpha_{k-1};$
  $t^k = A^T r^k;$
  $\alpha_k = \|t^k\|_2^2;$
  $\beta_k = \alpha_k \beta_k;$
  $d^{k+1} = t^k + \beta_k d^k;$
  end

Cost of CGNR: 2 matrix-vector products
2 scalar products
3 _axpy

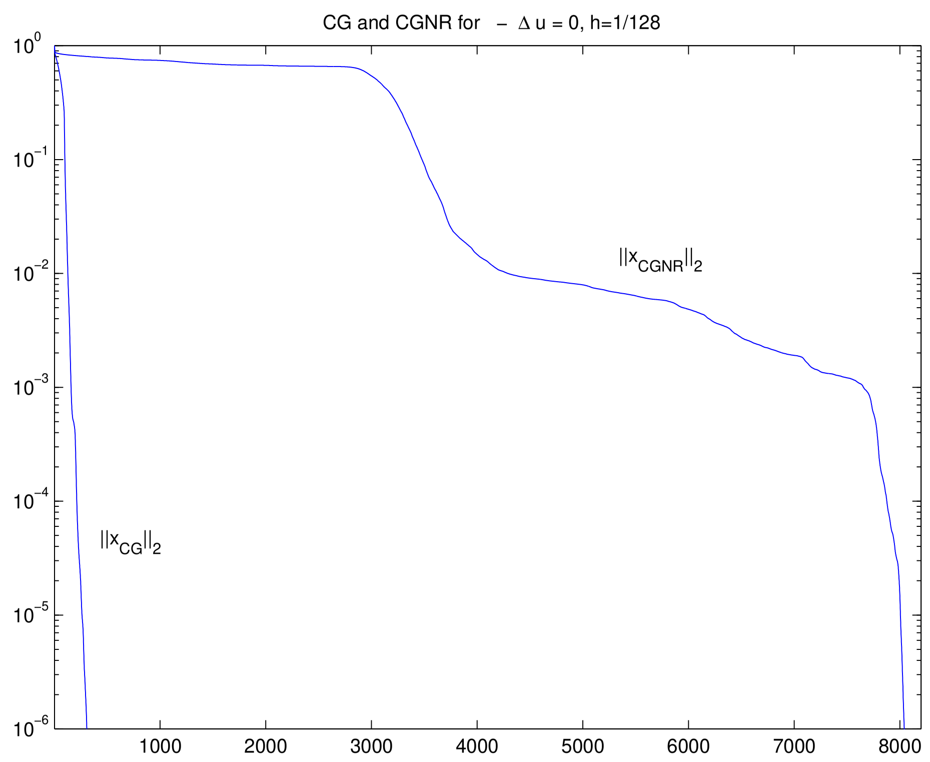
Storage requirements: 4 vectors

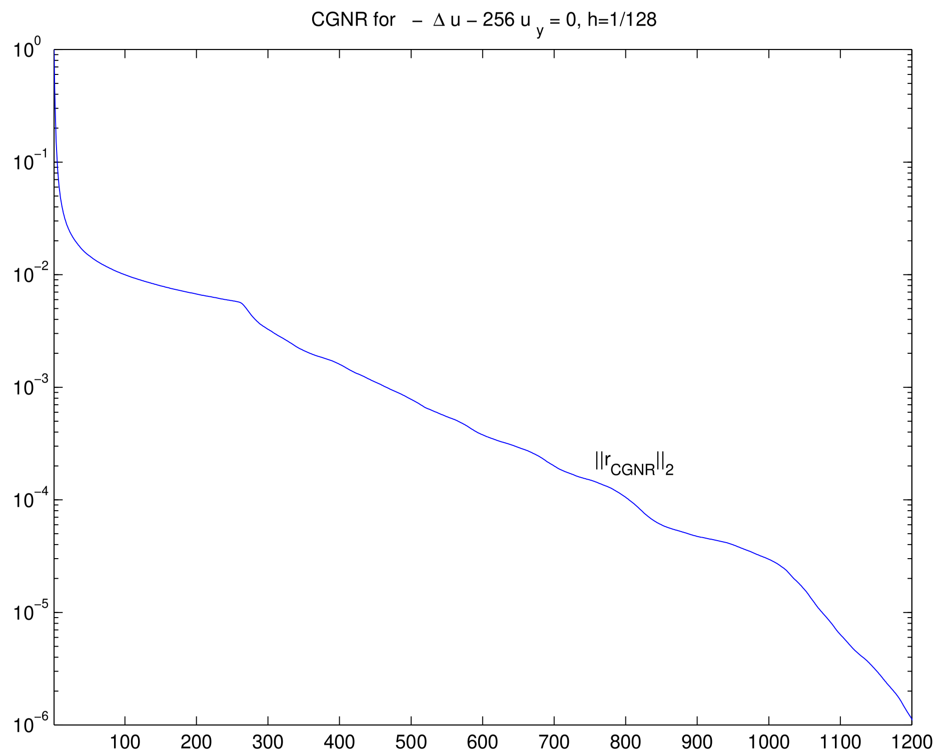
Theorem
Let A be nonsingular, σ_1 its largest and σ_n ist smallest singular value, and let

$$R := \frac{\sigma_1 - \sigma_n}{\sigma_1 + \sigma_n} = \frac{\mathrm{cond}_2(A) - 1}{\mathrm{cond}_2(A) + 1}.$$

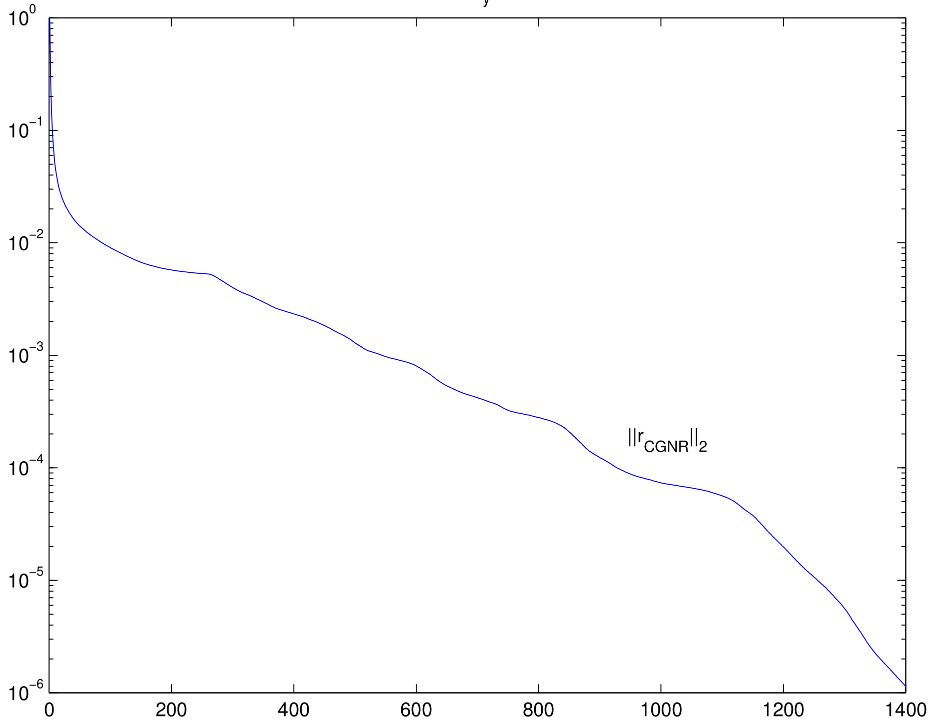
Then it holds

$$\|r^k\|_2 \le \frac{2R^k}{1 + R^{2k}} \|r^0\|_2 \le 2 \left(\frac{\mathrm{cond}_2(A) - 1}{\mathrm{cond}_2(A) + 1} \right)^k \|r^0\|_2.$$

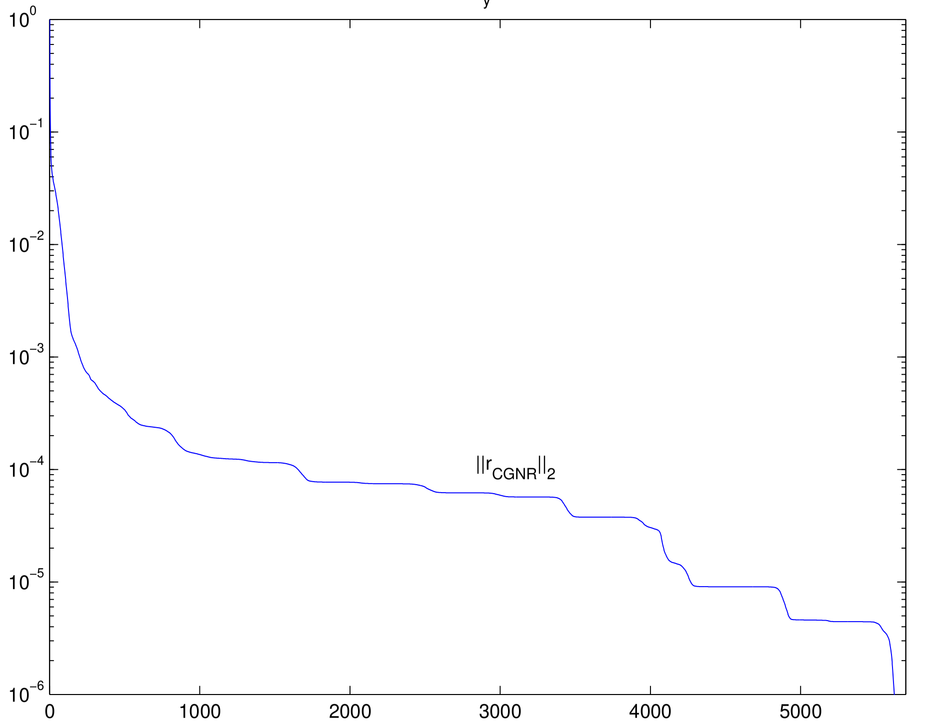

CG and CGNR for − Δ u = 0, h=1/128
10^0
10^−1
10^−2
10^−3
10^−4
10^−5
10^−6
1000
2000
3000
4000
5000
6000
7000
8000
||x_CGNR||_2
||x_CG||_2


CGNR for − Δ u − 256 u_y = 0, h=1/128
10^0
10^−1
10^−2
10^−3
10^−4
10^−5
10^−6
100
200
300
400
500
600
700
800
900
1000
1100
1200
||r_CGNR||_2

CGNR for $-\Delta u - 256 u_y - 163.84 u = 0$, h=1/128



CGNR for $-\Delta u - 256 u_y - 1638.4 u = 0$, h=1/128



For $AA^Ty = b$, $x = A^Ty$, the iterates y^k minimize the functional

$$\|y - y^*\|^2_{AA^T} = (y - y^*)AA^T(y - y^*) = \|(A^T(y - y^*)\|_2^2 = \|x - x^*\|_2^2$$

in the Krylov

$$\begin{aligned} &A^Tx^0 + A^T\mathcal{K}_k(r^0, A^TA) \\ &\quad = \quad A^Tx^0 + \text{span}\{A^Tr^0, A^T(A^TA)r^0, \ldots, A^T(A^TA)^{k-1}r^0\}. \end{aligned}$$

This method is called CGNE Verfahren where the letter 'E' indicates that the error is minimized. It has similar convergence properties as the CGNR method.

CGNE was introduced by Craig (1955).

Desirable are generalizations of CG (or CR) methods which have similar properties as CG without squaring the condition number of the systemmatrix.

Outstanding properties of the CG method:
— based on a short recurrence (low storage requirements)
— minimizes error on ascending sequence of Krylov spaces

Theorem of Faber & Manteuffel
Methods with these two properties exist only for a very limited class of matrices. For 3-term-recurrences only for matrices

$$A = e^{i\theta}(H + \sigma I), \; H = H^*, \; \theta \in \mathbb{R}, \; \sigma \in \mathbb{C}.$$

Hence, one has to decide which one of the two properties to keep.

Generalized Minimal Residual Method (GMRES) Saad & Schultz (1986)

GMRES constructs a sequence

$$x^k \in x^0 + \mathcal{K}_k(r^0, A)$$

of vectors such that the Euclidean norm of the residuum becomes minimal.

To this end an orthonormal basis $q^1, \dots, q^k$ of $\mathcal{K}_k(r^0, A)$ is determined using Arnoldi's method (i.e. the modified Gram-Schmidt process), such that the problem

$$\|r^k\|_2 = \|b - A(x^0 + Q_k y)\|_2 = \text{min!}, \quad y \in \mathrm{IR}^k,$$

can be solved easily.

Arnoldi's method

Let q^1 be an initial vector such that $\|q^1\|_2 = 1$.
for $k = 1, 2, \dots$
  $q^{k+1} = Aq^k$;
  for $j = 1 : k$
    $h_{jk} = (q^{k+1})^T q^j$;
    $q^{k+1} = q^{k+1} - h_{jk} q^j$;
    end
  $h_{k+1,k} = \|q^{k+1}\|_2$;
  $q^{k+1} = q^{k+1} / h_{k+1,k}$;
  end

With $Q_k = (q^1, \dots, q^k)$ and

$$\tilde{H}_k = (h_{j\ell}) \in \mathrm{IR}^{(k+1,k)} \quad \text{such that } h_{j\ell} = 0 \text{ for } j - \ell > 2$$

it holds

$$AQ_k = Q_k \tilde{H}_k,$$

and for $x = x^0 + Q_k y$, $r = b - Ax$ and $q^1 = r^0/\|r^0\|_2$

$$\begin{aligned} \|r\|_2 &= \|b - A(x^0 + Q_k y)\|_2 = \|r^0 - AQ_k y\|_2 = \|r^0 - Q_k \tilde{H}_k y\|_2 \\ &= \left\| Q_k(\|r^0\|_2 e^1 - \tilde{H}_k y) \right\|_2 = \left\| \|r^0\|_2 e^1 - \tilde{H}_k y \right\|_2 . \end{aligned}$$

Hence, the problem to minimize the Euclidean norm of the residuum on a Krylov space yields a least squares problem where the system matrix $\tilde{H}_k$ has Hessenberg form .

If $U_k \in \mathrm{IR}^{(k+1,k+1)}$ is an orthogonal matrix such that

$$U_k \tilde{H}_k = \begin{pmatrix} R_k \\ 0^T \end{pmatrix}, \quad R_k \text{ upper triangular,}$$

and if u^k is the first column of U_k then it holds

$$\begin{aligned} \left\| \|r^0\|_2 e^1 - \tilde{H}_k y \right\|_2^2 &= \left\| U_k(\|r^0\|_2 e^1 - \tilde{H}_k y) \right\|_2^2 \\ &= \left\| \|r^0\|_2 u^k - \begin{pmatrix} R_k \\ 0^T \end{pmatrix} y \right\|_2^2 \\ &= \left\| \|r^0\|_2 (u_1^k, \dots, u_k^k)^T - R_k y \right\|_2^2 \\ &\quad + \|r^0\|_2 \cdot |u_{k+1}^k|. \end{aligned}$$

Hence, the least squares problem is equivalent to the linear system

$$R_k y = \|r^0\|_2 \begin{pmatrix} u_1^k \\ \vdots \\ u_k^k \end{pmatrix}$$

where R_k is an upper triangular matrix, and the norm of the residuum satisfies

$$\|b - Ax^k\|_2 = \|r^0\| \cdot |u_{k+1}^k|,$$

i.e. the accuracy of the method, can be monitored without explicitly forming x^k or y^k.

If the residual norm is sufficiently small, GMRES is terminated, and only after this final step of the Arnoldi method, the triangular system is solved and $x^k = x^0 + Q_k y^k$ is determined.

Since $\tilde{H}_k$ is an upper Hessenberg matrix, the matrix U_k can be calculated efficiently using Givens reflections. Let

$$G_j = \begin{pmatrix} I_{j-2} & 0 & 0 \\ 0^T & c_j & s_j \\ 0^T & s_j & -c_j \end{pmatrix} \in \mathrm{IR}^{(j,j)}, \; c_j^2 + s_j^2 = 1$$

denote a Givens reflection acting on the $(j-1)$-st and j-th component of a vector..

Let $U_1 = G_2$ be such that

$$U_1 \tilde{H}_1 = \begin{pmatrix} c_2 & s_2 \\ s_2 & -c_2 \end{pmatrix} \begin{pmatrix} h_{11} \\ h_{21} \end{pmatrix} = \begin{pmatrix} \sqrt{h_{11}^2 + h_{21}^2} \\ 0 \end{pmatrix} =: \begin{pmatrix} \hat{h}_{11} \\ 0 \end{pmatrix}.$$

Assume that the orthogonal matrix

$$U_{k-1} = G_k \begin{pmatrix} G_{k-1} & 0 \\ 0^T & 1 \end{pmatrix} \cdot \ldots \cdot \begin{pmatrix} G_2 & O \\ O & I_{k-1} \end{pmatrix} =: G_k \begin{pmatrix} U_{k-2} & 0 \\ 0^T & 1 \end{pmatrix}$$

is chosen such that

$$U_{k-1}\tilde{H}_{k-1} =: \begin{pmatrix} \hat{h}_{11} & \hat{h}_{12} & \ldots & \hat{h}_{1,k-1} \\ 0 & \hat{h}_{22} & \ldots & \hat{h}_{2,k-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \hat{h}_{k-1,k-1} \\ 0 & 0 & \ldots & 0 \end{pmatrix} =: \begin{pmatrix} R_{k-1} \\ 0^T \end{pmatrix} \in \mathbb{R}^{(k,k-1)}$$

is an upper triangular matrix.

Since the multiplication of a matrix by G_{k+1} from the left leaves invariant the first $k-1$ rows one obtains the upper triangular matrix in the k-th step of the method by

$$\begin{aligned} U_k\tilde{H}_k &= G_{k+1} \begin{pmatrix} U_{k-1} & 0 \\ 0^T & 1 \end{pmatrix} \begin{pmatrix} \tilde{H}_{k-1} & h^k \\ 0^T & h_{k+1,k} \end{pmatrix} \\ &= \begin{pmatrix} I_{k-1} & 0 & 0 \\ 0^T & c_{k+1} & s_{k+1} \\ 0^T & s_{k+1} & -c_{k+1} \end{pmatrix} \begin{pmatrix} R_{k-1} & \tilde{h}^k \\ 0^T & g_k \\ 0^T & h_{k+1,k} \end{pmatrix} = \begin{pmatrix} R_k \\ 0^T \end{pmatrix} \end{aligned}$$

where

$$U_{k-1} \begin{pmatrix} h_{1k} \\ \vdots \\ h_{k-1,k} \\ h_{kk} \end{pmatrix} =: U_{k-1}h^k =: \begin{pmatrix} \hat{h}_{1k} \\ \vdots \\ \hat{h}_{k-1,k} \\ g_k \end{pmatrix} =: \begin{pmatrix} \hat{h}^k \\ g_k \end{pmatrix}$$

and

$$s_{k+1} := \frac{h_{k+1,k}}{\sqrt{g_k^2 + h_{k+1,k}^2}}, \quad c_{k+1} := \frac{g_k}{\sqrt{g_k^2 + h_{k+1,k}^2}}.$$

In a similar manner as in the SYMMLQ method the elements $\hat{h}_{j\ell}$ can be obtained along with the Arnoldi process, and the reflections can be applied to the right hand side $\|r^0\|_2 e^1$ currently. Observe however, that differently from the SYMMLQ method the nontrivial entries c_j and s_j of G_j have to be stored since they are needed in the transformation of the k-th column h^k of $\tilde{H}_k$.

The diagonal elements of the triangular matrix R_k are given by $\hat{h}_{jj} = \sqrt{g_j^2 + h_{j+1,j}^2}$. Hence, R_k is nonsingular, if the Arnoldi process constructs k orthogonal vectors q^j. GMRES breaks down if and only if $h_{j+1,j} = 0$ for some j, and this happens if and only if $r^j = 0$, i.e. $x^j = x^*$. (lucky breakdown)

GMRES

$q^1 = b - Ax^0$;
$z_1 = \|q^1\|_2$;
$q^1 = q^1/z_1$;
$k = 1$;
while $|z_k|$ >tol
  $q^{k+1} = Aq^k$;
  for $i = 1 : k$
    $h_{ik} = (q^i)^T q^{k+1}$;;
    $q^{k+1} = q^{k+1} - h_{ik} q^i$; end
  $h_{k+1,k} = \|q^{k+1}\|_2$;
  $q^{k+1} = q^{k+1}/h_{k+1,k}$;
  for $i = 1 : k-1$
    $\begin{pmatrix} h_{i,k} \\ h_{i+1,k} \end{pmatrix} = \begin{pmatrix} c_{i+1} & s_{i+1} \\ s_{i+1} & -c_{i+1} \end{pmatrix} \begin{pmatrix} h_{i,k} \\ h_{i+1,k} \end{pmatrix}$ end

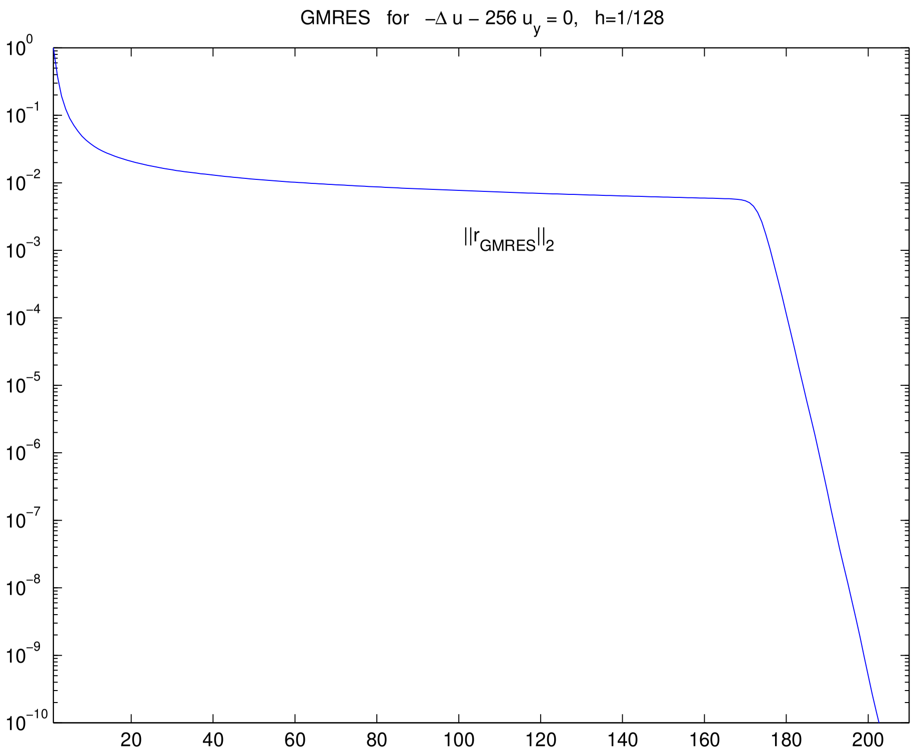
$\alpha = \sqrt{h_{kk}^2 + h_{k+1,k}^2};$
$s_{k+1} = h_{k+1,k}/\alpha;$
$c_{k+1} = h_{kk}/\alpha;$
$h_{kk} = \alpha;$
$z_{k+1} = s_{k+1} z_k;$
$z_k = c_{k+1} z_k;$
$k = k + 1;$
end
$y_k = z_k / h_{kk};$
for $i = k-1 : -1 : 1$
$y_i = (z_i - \sum_{j=i+1}^k h_{ij} y_j)/h_{ii};$ end
$x^k = x^0 + \sum_{j=1}^k y_j q^j;$

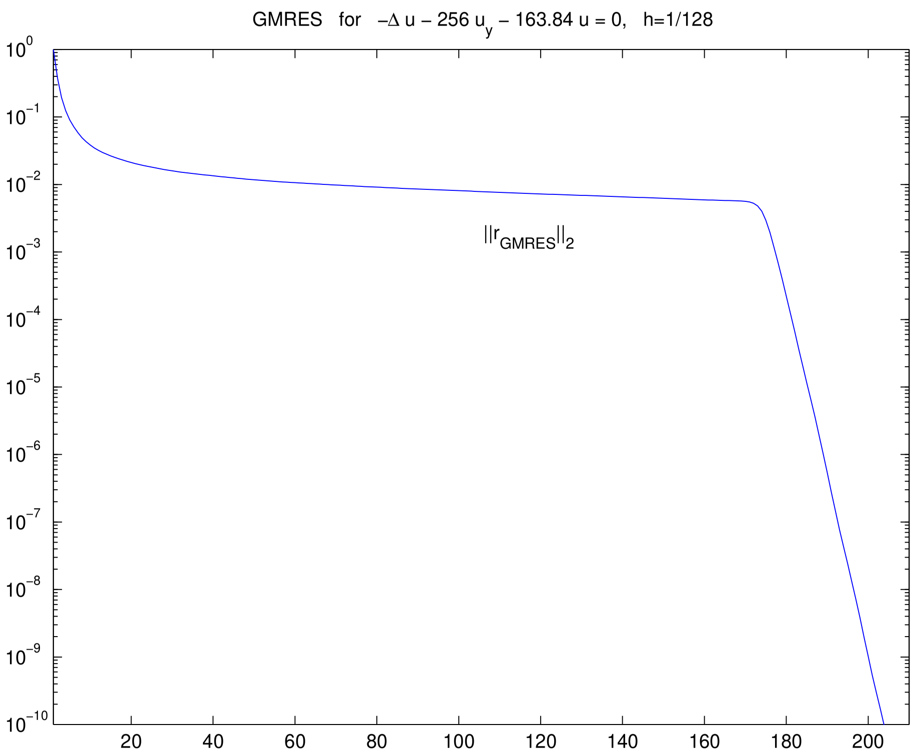
Cost of the k-th step (neglecting the transformation of $\tilde{H}_k$ to triangular form)

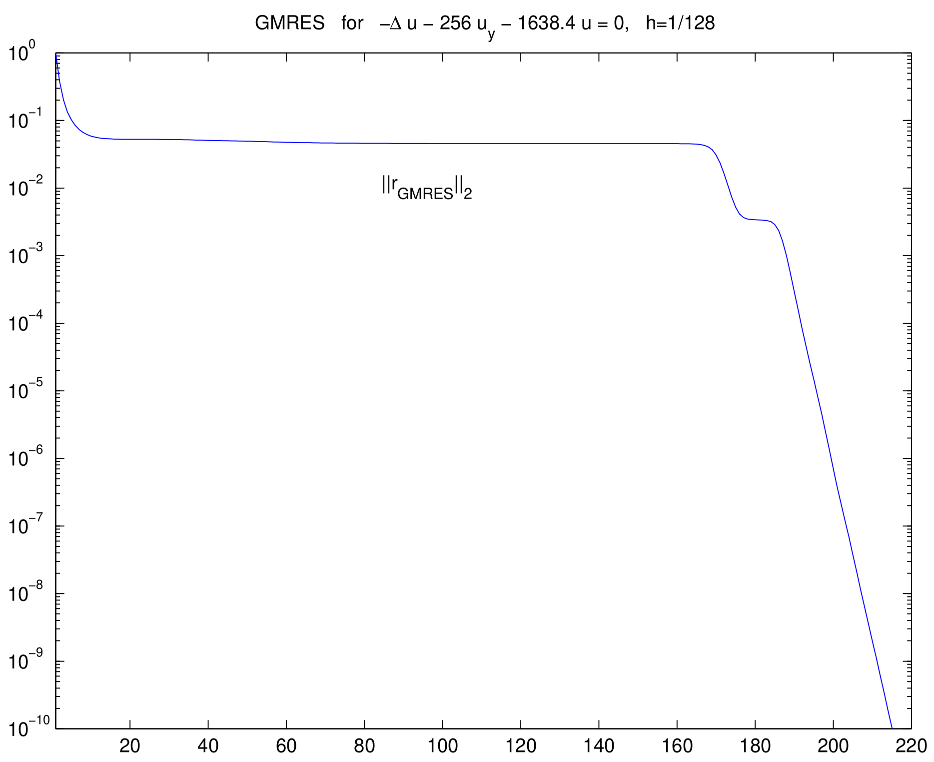
1	matrix-vector product
k+1	scalar products
k+1	_axpy

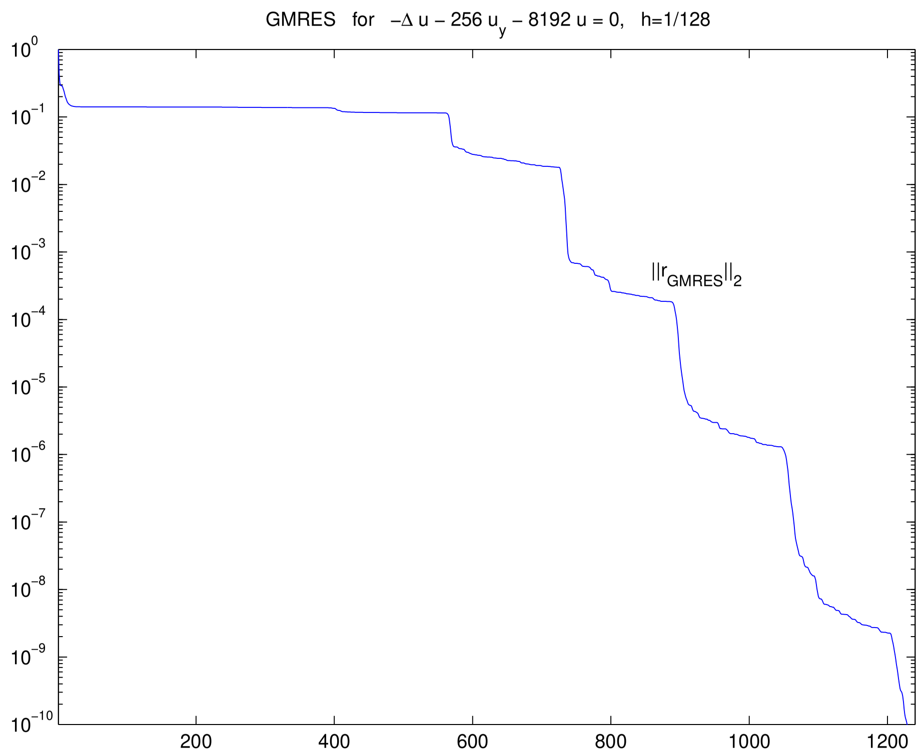
Determining $x^m \in x^0 + \mathcal{K}_m(r^0, A)$ costs

m+1	matrix-vector products
(m+1)(m+2)+m	level-1-operations
m+1	vectors to store


GMRES for −Δ u − 256 u_y = 0, h=1/128
||r_GMRES||_2
10^0
10^−1
10^−2
10^−3
10^−4
10^−5
10^−6
10^−7
10^−8
10^−9
10^−10
20 40 60 80 100 120 140 160 180 200


GMRES for −Δ u − 256 u_y − 163.84 u = 0, h=1/128
||r_GMRES||_2
10^0
10^−1
10^−2
10^−3
10^−4
10^−5
10^−6
10^−7
10^−8
10^−9
10^−10
20 40 60 80 100 120 140 160 180 200


GMRES for –Δ u – 256 u_y – 1638.4 u = 0, h=1/128
10^0
10^-1
10^-2
10^-3
10^-4
10^-5
10^-6
10^-7
10^-8
10^-9
10^-10
20
40
60
80
100
120
140
160
180
200
220
||r_GMRES||_2


GMRES for –Δ u – 256 u_y – 8192 u = 0, h=1/128
10^0
10^-1
10^-2
10^-3
10^-4
10^-5
10^-6
10^-7
10^-8
10^-9
10^-10
200
400
600
800
1000
1200
||r_GMRES||_2

Convergence Theorem 1

If A is positive real, i.e. the symmetric part $M := 0.5(A + A^T)$ of A is positive definite, then it holds

$$\|r^k\|_2 \leq \left(1 - \frac{\lambda_{\min}(M)^2}{\lambda_{\max}(A^T A)}\right)^{k/2} \|r^0\|_2.$$

Here $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ denotes the minimal and maximal eigenvalue, respectively.

Proof: By construction

$$\|r^k\|_2 = \min_{q \in \Pi_k,\ q(0)=1} \|q(A)r^0\|_2,$$

and for every $\alpha \in \mathrm{IR}$ we get with the polynomial $\tilde{q}(\lambda) = (1 - \alpha\lambda)^k$

$$\min_{q \in \Pi_k,\ q(0)=1} \|q(A)\|_2 \leq \|\tilde{q}(A)^k\|_2 \leq \|\tilde{q}(A)\|_2^k.$$

From $x^T M x \geq \lambda_{\min}(M) x^T x$ and $x^T A^T A x \leq \lambda_{\max}(A^T A) x^T x$ for every $x \neq 0$ we obtain for every $\alpha \in \mathrm{IR}$

$$\begin{aligned} \|\tilde{q}(A)\|_2^2 &= \max_{\|x\|_2=1} x^T (I + \alpha A)^T (I + \alpha A) x \\ &= \max_{\|x\|_2=1} (1 + 2\alpha x^T (A + A^T) x + \alpha^2 x^T A^T A x) \end{aligned}$$

and for $\alpha < 0$

$$\|\tilde{q}(A)\|_2^2 \le 1 + 2\alpha\lambda_{\min}(M) + \alpha^2\lambda_{\max}(A^TA).$$

The quadratic expression on the right hand side attains its minimum for $\alpha = -\lambda_{\min}(M)/\lambda_{\max}(A^TA)$. Hence,

$$\|\tilde{q}(A)\|_2 \le \left(1 - \frac{\lambda_{\min}(M)}{\lambda_{\max}(A^TA)}\right)^{1/2},$$

from which we get the statement.

Convergence Theorem 2

If $A = T\Lambda T^{-1}$ is diagonalizable with a regular matrix T and a diagonal matrix Λ, then it holds

$$\|r^k\|_2 \le \mathrm{cond}_2(T)\|r^0\|_2 \cdot \min_{q\in\Pi_k,\ q(0)=1}\ \max_{\lambda\in\sigma(A)} |q(\lambda)|.$$

If A is normal, then

$$\|r^k\|_2 \le \|r^0\|_2 \cdot \min_{q\in\Pi_k,\ q(0)=1}\ \max_{\lambda\in\sigma(A)} |q(\lambda)|.$$

Here $\sigma(A) = \{\lambda_1, \dots, \lambda_n\}$ denotes the spectrum of A.

Proof: With $\tilde{\Pi}_k := \{q \in \Pi_k \,:\, q(0) = 1\}$ it holds

$$\begin{aligned}
\|r^k\|_2 &= \min_{q\in\tilde{\Pi}_k} \|q(A)r^0\|_2 \\
&= \min_{q\in\tilde{\Pi}_k} \|q(T\Lambda T^{-1})r^0\|_2 \\
&= \min_{q\in\tilde{\Pi}_k} \|T\,q(\Lambda)T^{-1}r^0\|_2 \\
&\leq \|T\|_2 \cdot \|T^{-1}\|_2 \min_{q\in\tilde{\Pi}_k} \|q(\Lambda)\|_2 \|r^0\|_2 \\
&= \mathrm{cond}_2(T)\|r^0\|_2 \min_{q\in\tilde{\Pi}_k} \max_{\lambda\in\sigma(A)} |q(\lambda)|.
\end{aligned}$$

If A is normal, then A can be transformed to diagonal form unitarily, and $\mathrm{cond}_2(T) = 1$.

If A is normal, then (cf. Greenbaum & Gurvits (1994)) the bound of the speed of convergence is sharp, i.e. for A there exists an initial residual r^0 such that the inequality holds with '$=$'. In this case the speed of convergence depends on the spectrum of the matrix.

For nonnormal matrices it was shown by Greenbaum, Ptak and Strakos (1996), that for every monotonely decreasing sequence of positive real numbers there exists a matrix and an initial vector such that the residuals in the GMRES method attain the elements of the given sequence. The matrix can be determined such that it has predetermined eigenvalues. Hence, the speed of convergence of the GMRES method is completely independent of the spectrum of the system matrix.

That the speed of convergence of GMRES for nonnormal matrices is independent of the spectrum of the system matrix can be seen in the following way. Let

$$A = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ -a_0 & -a_1 & -a_2 & \dots & -a_{n-1} \end{pmatrix}$$

be a Frobenius matrix. Assume that the initial vector is chosen such that $r^0 = e^1$. Then Ar^0 is a multiple of e^n, $A^2 r^0$ is a linear combination of e^n and e^{n-1}, etc. All $A^k r^0$, $k = 1, \dots, n-1$ are orthogonal to r^0. Hence, the optimal approximation in $x^0 + \mathcal{K}_k(r^0, A)$ is $x^k = x^0$, $k = 1, \dots, n-1$, and only in the n-th step the solution is arrived. By choosing the a_js suitably every spectrum of A can be realized.

Remark: With this result in mind it is easy to construct an example for which CGNR clearly outperforms GMRES:

For $A \in \mathrm{IR}^{(10,10)}$ as on the last slide where the roots of the polynomial

$$p(\lambda) = \lambda^{10} + \sum_{j=0}^{9} a_j \lambda^j$$

are

$$\lambda_j = 1 + 0.1 \exp(0.2\pi j i), \; j = 0, \dots, 9,$$

$b = e^1 = (1, 0, \dots, 0)^T$ and $x0 = 0$ GMRES stagnates with $x_1 = \dots = x_9 = 0$, $x_{10} = e^2$, whereas CGNR reduces the Euclidean norm of the residual to $4.1e - 11$ within 3 steps.

The main disadvantage of GMRES is, that the cost and the storage requirements grow linearly with the number of iterations. The usual way to cope with this problem is restarting. After a given number m of steps the current least squares problem is solved, and the solution x^m is chosen to be the initial vector for another m GMRES steps. This method called GMRES(m) is repeated until convergence.

There are no rules for choosing m. If m is chosen too large the method can become too expensive, and if m is chosen too small the convergence can become very slow, or the method can even stagnate without convergence.

Example

For

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \; b = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \; x^0 = 0,$$

$x^1 = x^0$ solves the least squares problem

$$\|Ax - b\|_2 = \min!, \quad x = x^0 + \alpha b.$$

GMRES(1) stagnates.

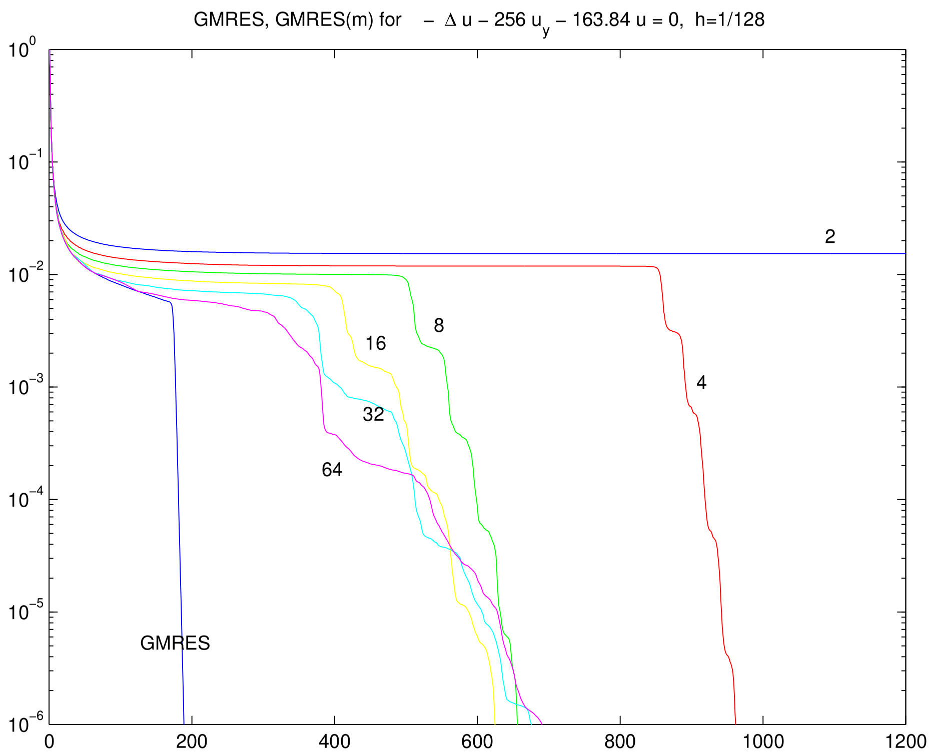
For positive real A we get from the first convergence theorem for GMRES(m) die residual bound

$$\|r^{jm+i}\|_2 \le \left(1 - \frac{\lambda_{\min}(M)^2}{\lambda_{\max}(A^T A)}\right)^{i/2} \|r^{jm}\|_2,$$

$$\|r^{jm}\|_2 \le \left(1 - \frac{\lambda_{\min}(M)^2}{\lambda_{\max}(A^T A)}\right)^{m/2} \|r^{(j-1)m}\|_2,$$

and therefore

$$\|r^{k}\|_2 \le \left(1 - \frac{\lambda_{\min}(M)^2}{\lambda_{\max}(A^T A)}\right)^{k/2} \|r^{0}\|_2.$$

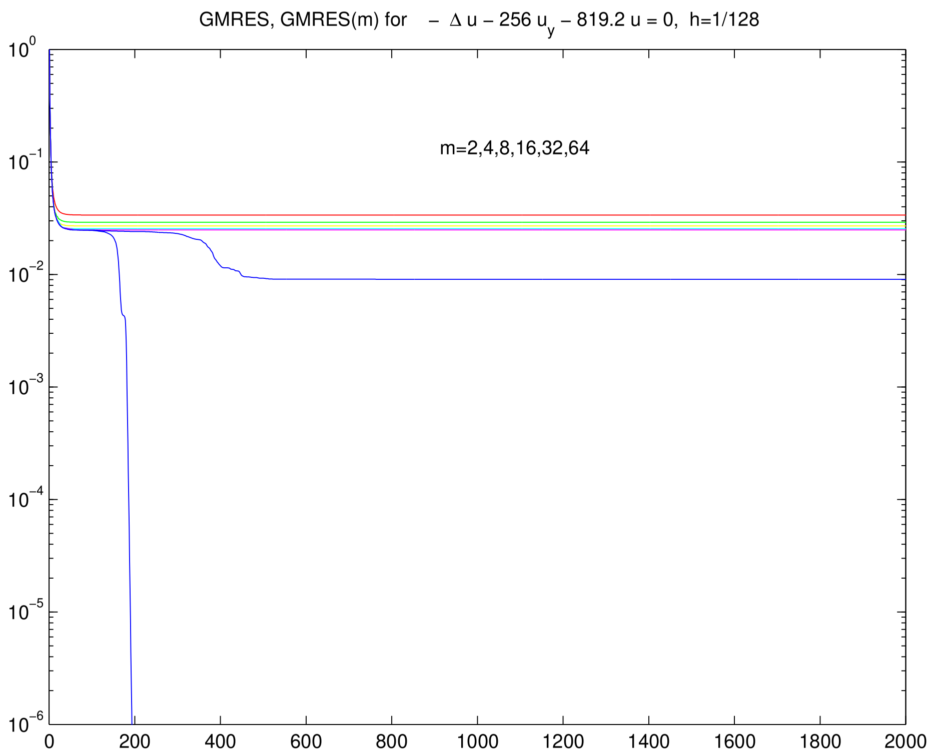


For

$$-\Delta u - 128 u_y - 163.84 u = 0, \quad h = 1/128,$$

we get

Type	iterations	flops
GMRES	190	1.20e9
GMRES(2)	stationär	
GMRES(4)	964	4.39e8
GMRES(8)	656	3.66e8
GMRES(16)	624	5.01e8
GMRES(32)	704	9.24e8
GMRES(64)	704	1.65e9



GMRES and GMRES(m) can be precondioned applying the methods to the system

$$M^{-1}Ax = M^{-1}b \qquad \text{left preconditioning}$$

or to

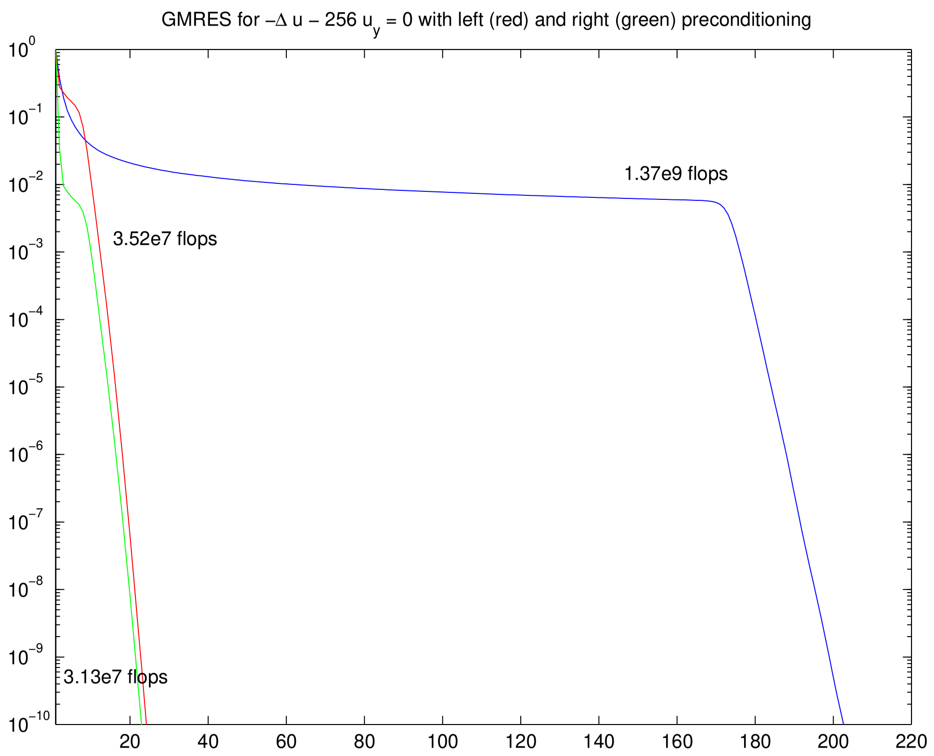
$$AM^{-1}y = b, \; y = Mx \qquad \text{right preconditioning}$$

where in both cases M is an approximation to A such that linear systems $Mu = c$ can be solved easily.

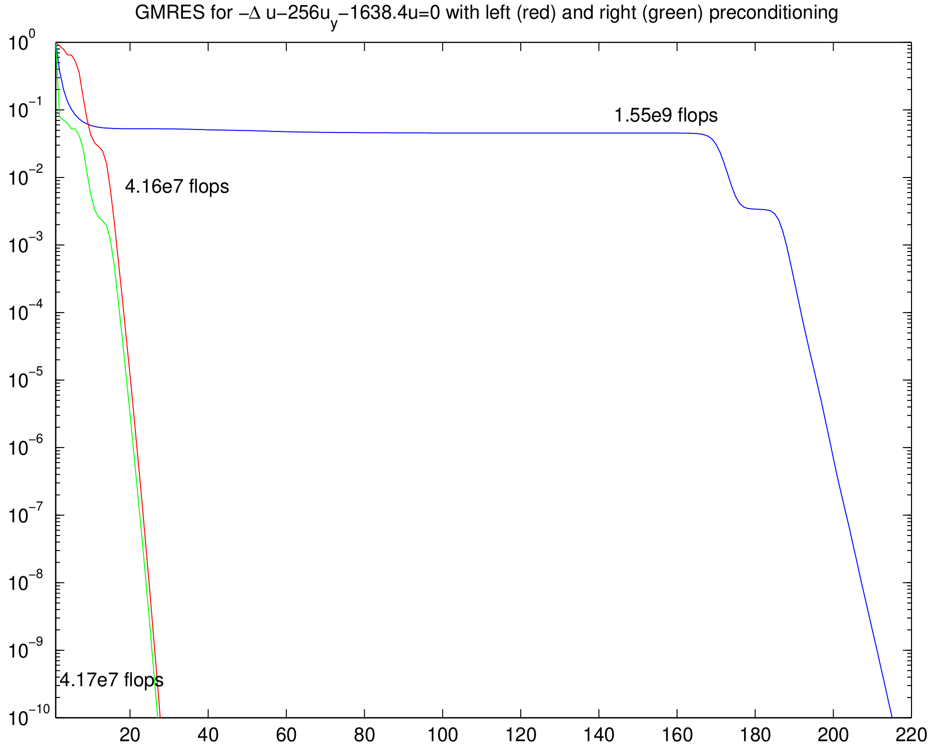
In the following examples we used $M = LU$ where

$$[L, U] = \text{luinc}(A, 0.1)$$

is an incomplete LU factorization of A with threshold 0.1.



GMRES for $-\Delta u - 256u_y - 1638.4u = 0$ with left (red) and right (green) preconditioning



GMRES for $-\Delta u - 256u_y - 8192u = 0$ with left (red) and right (green) preconditioning