Krylov subspace methods

Convergence of CG, MINRES, and GMRES

Convergence bounds for CG

Consider the CG residual and from that the CG error.
CG: $r_m = b - A(x_0 + z_m) = r_0 - Az_m$ where $z_m \in K^m(A, r_0)$

This gives polynomial $r_m = r_0 - AP_{m-1}(A)r_0 = (I - AP_{m-1}(A))r_0$

Multiplying by A^{-1} gives $e_m = A^{-1}r_m = (I - AP_{m-1}(A))e_0$

Let $R_m(A) = (I - AP_{m-1}(A))$ be the residual polynomial.

Then we get a bound for the error: $\|e_m\|_A \le \|R_m(A)\|_2 \|e_0\|_A$

Let $A = V\Lambda V^H$ then $A^k = V\Lambda^k V^H$ and so $R_m(A) = VR_m(\Lambda)V^H$

Since Λ is a diagonal matrix, we have $R_m(\Lambda) = \mathrm{diag}(R_m(\lambda_i))$, and
$\|R_m(A)\|_2 = \|R_m(\Lambda)\|_2 = \max_{\lambda_i \in \lambda(A)} |R_m(\lambda_i)|$

As we do not know the eigenvalues we make a final simplification

$\max_{\lambda_i \in \lambda(A)} |R_m(\lambda_i)| \le \max_{a \le \lambda \le b} |R_m(\lambda)|$, where $\lambda(A) \subset [a, b]$

Convergence bounds for CG

Now we would like to find a bound on $\max_{a\le\lambda\le b}|R_m(\lambda)|$
From the optimality of CG we know that

$$\|R_m(A)A^{1/2}e_0\|_2 \le \|\tilde{R}_m(A)A^{1/2}e_0\|_2$$

for any other residual polynomial, $\tilde{R}_m(.)$, of the degree m.
One way to get a bound is to pick a particular polynomial for which we can easily compute the norm and which is know to be small.

One useful candidate is the (best) minimax polynomial over the interval that contains the eigenvalues. These are so-called Chebyshev polynomials. We consider other choices later.

What keeps us from taking zero polynomial, or what is residual polynomial? Our choices must satisfy a normalization:

We have $R_m(\lambda) = (1 - \lambda P_{m-1}(\lambda)) \Rightarrow R_m(0) = 1$

$$\min_{P\in\Pi_m^0} \max_{\lambda\in[a,b]} |P(\lambda)|$$

Convergence bounds for CG

We know from approximation theory that such a polynomial must be *equioscillating*. That is obtain alternatingly maxima and minima that are equal in absolute value. We also know it is unique.

One equioscillating function is $\cos m\theta$, but is it a polynomial?
The answer is yes; it is a polynomial in $\cos\theta$.

Let $\cos\theta = x$, for $0 \le \theta \le \pi$ and $\theta = \arccos x$ (pv), with $-1 \le x \le 1$
$T_m(x) = \cos(m \arccos x)$ for $-1 \le x \le 1$
Obviously $\max|T_m(x)| = 1$, attained at m interior points and ± 1.
Outside interval $-1 \le x \le 1$, we have $T_m(x) = \cosh(m \cosh^{-1} x)$.

In order to get minimax polynomial over required interval we use (linear polynomial) function

$$t(x) = \frac{2x-b-a}{b-a},$$

which maps $[a, b]$ to $[-1, 1]$ where Chebyshev pol. is defined.

Convergence bounds for CG

This gives the function $T_m(w(x))$, which is equioscillating but not normalized over the interval $[a,b]$. We normalize this function by dividing by $T_m(w(0))$ (a scalar constant):

$$\hat{T}_m(x) = \frac{T_m(w(x))}{T_m(w(0))} = \frac{T_m\left(\frac{2x-b-a}{b-a}\right)}{T_m\left(-\frac{b+a}{b-a}\right)}.$$

Since $\max_{x\in[a,b]}|T_m(x)| = 1$ we get $\max_{x\in[a,b]}\left|\hat{T}_m(x)\right| = |T_m(w(0))|^{-1}$

How large is $|T_m(w(0))|$?
Outside interval $-1 \le x \le 1$, we have $T_m(x) = \cosh(m \cosh^{-1} x)$.
Let $y = e^{\beta}$ and $x = \frac{1}{2}(y + y^{-1})$. Then $T_m(x) = \frac{1}{2}(y^m + y^{-m})$.

Now y defined by $y^2 - 2xy + 1 = 0$. Take solution $|y| \ge 1$.

$y = -x - \sqrt{x^2 - 1}$ (note since $b > a > 0$, $-\frac{b+a}{b-a} < 0$)

Convergence bounds for CG

We have $x = w(0) = -\frac{b+a}{b-a}$. Take $b = \lambda_{\max}$ and $a = \lambda_{\min}$.

$$x = -\frac{b+a}{b-a} = -\left(\frac{b+a}{a}\right)\left(\frac{b-a}{a}\right)^{-1} = -\left(\frac{b}{a}+1\right)\left(\frac{b}{a}-1\right)^{-1} = -\frac{\kappa+1}{\kappa-1}$$

$\kappa = \frac{b}{a}$ is condition number of A. From $y = -x - \sqrt{x^2-1}$ we get

$$y = \frac{\kappa+1}{\kappa-1} - \left(\frac{\kappa^2+2\kappa+1}{\kappa^2-2\kappa+1} - \frac{\kappa^2-2\kappa+1}{\kappa^2-2\kappa+1}\right)^{1/2} = \frac{\kappa+1}{\kappa-1} - \left(\frac{4\sqrt{\kappa}^2}{(\kappa-1)^2}\right)^{1/2}$$

$$y = \frac{\kappa+1}{\kappa-1} - \frac{2\sqrt{\kappa}}{\kappa-1} = \frac{\sqrt{\kappa}^2 - 2\sqrt{\kappa}+1}{(\sqrt{\kappa}+1)(\sqrt{\kappa}-1)} = \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}, \text{ and so}$$

$$T_m\left(-\frac{b+a}{b-a}\right) = \frac{1}{2}\left(\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^m + \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^{-m}\right) \le \frac{1}{2}\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^{-m}$$

Result: $\max_{x\in[a,b]}\left|\hat{T}_m(x)\right| = |T_m(w(0))|^{-1} \le 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^m$

$$\|e_m\|_A \le \|R_m(A)e_0\|_A \le \|\hat{T}_m(A)e_0\|_A \le 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^m \|e_0\|_A$$

Convergence bounds for CG

Consider case where one eigenvalue λ_n much larger than others.

Construct better polynomial than $T_k\left(\frac{2\lambda-\lambda_n-\lambda_1}{\lambda_n-\lambda_1}\right)/T_k\left(\frac{-\lambda_n-\lambda_1}{\lambda_n-\lambda_1}\right)$ using this information.

For example, "polynomial that is zero at extreme eigenvalue and lower degree Chebyshev over other eigenvalues".

$p_k(z) = \left[T_k\left(\frac{2\lambda-\lambda_{n-1}-\lambda_1}{\lambda_{n-1}-\lambda_1}\right)/T_k\left(\frac{-\lambda_{n-1}-\lambda_1}{\lambda_{n-1}-\lambda_1}\right)\right]\left(\frac{\lambda_n-\lambda}{\lambda_n}\right)$

Clearly $p_k(\lambda_n) = 0$ and $|p_k(\lambda_i)| < \left|T_k\left(\frac{2\lambda_i-\lambda_{n-1}-\lambda_1}{\lambda_{n-1}-\lambda_1}\right)/T_k\left(\frac{-\lambda_{n-1}-\lambda_1}{\lambda_{n-1}-\lambda_1}\right)\right|$, $i < n$

So, new bound $\frac{\|e_k\|_A}{\|e_0\|_A} \leq 2\left(\frac{\sqrt{\kappa_{n-1}}-1}{\sqrt{\kappa_{n-1}}+1}\right)^{k-1}$, where $\kappa_{n-1} = \frac{\lambda_{n-1}}{\lambda_1}$,

versus old bound: $\frac{\|e_k\|_A}{\|e_0\|_A} \leq 2\left(\frac{\sqrt{\kappa_n}-1}{\sqrt{\kappa_n}+1}\right)^{k}$, where $\kappa_n = \frac{\lambda_n}{\lambda_1}$.

Convergence bounds for MINRES

Clearly, the trick can be applied if we have multiple outlying eigenvalues (large ones and small ones).

The same convergence bounds obtained for the error in CG can be obtained for the residual in MINRES if A is HPD, since we bound the same polynomial.
MINRES: $\frac{\|r_k\|_2}{\|r_0\|_2} \leq 2\left(\frac{\sqrt{\kappa_n}-1}{\sqrt{\kappa_n}+1}\right)^{k}$, where $\kappa_n = \frac{\lambda_n}{\lambda_1}$.

However, if A is Hermitian but not definite (MINRES) we need to find a (Chebyshev) polynomial that is small on both sides of the origin. This is much harder, which has a significant effect on the convergence (bound).

Convergence bounds for MINRES

Let A be Hermitian and let $\lambda(A) \subset [a,b] \cup [c,d]$, where $a < b < 0 < c < d$ and $b - a = d - c$.

We need a polynomial that is small over both these intervals.
We proceed more or less the same way as for CG: we construct a polynomial q that maps both intervals into $[-1,1]$ and define the Chebyshev polynomial in terms of q.

We take $q(z) = 1 + \frac{2(z-b)(z-c)}{ad-bc}$ (2nd degree polynomial)
Check that $q(z)$ maps $[a,b] \cup [c,d]$ into $[-1,1]$ (draw q).
How would you compute $q(z)$ for more general $[a,b] \cup [c,d]$?

Now we take $p_k(z) = T_l(q(z))/T_l(q(0))$, where $l = [k/2]$ (integral part).

Note that we have Chebyshev polynomials of half the degree we had in the definite case.

Convergence bounds for MINRES

To compute $\max|p_k(z)|$ we need to compute $T_l(q(0))$.

We have $q(0) = 1 + \frac{2bc}{ad-bc} = \frac{ad+bc}{ad-bc}$.

Set $\mu = \frac{ad+bc}{ad-bc} = \frac{1}{2}(y + y^{-1})$ then $T_l(\mu) = \frac{1}{2}(y^l + y^{-l})$.

Solve $\mu = \frac{1}{2}(y + y^{-1}) \Leftrightarrow \frac{1}{2}y^2 - \mu y + \frac{1}{2} = 0 \qquad (y \neq 0)$

$y = \mu \pm \sqrt{\mu^2 - 1}$ (solution are each other's inverse, so same result)

$$y = \frac{ad+bc}{ad-bc} + \sqrt{\frac{(ad+bc)^2}{(ad-bc)^2} - \frac{(ad-bc)^2}{(ad-bc)^2}} = \frac{ad+bc}{ad-bc} + \sqrt{\frac{4adbc}{(ad-bc)^2}} = \frac{ad+bc}{ad-bc} + \frac{2\sqrt{adbc}}{ad-bc} \Leftrightarrow$$

$$y = \frac{(\sqrt{ad}+\sqrt{bc})^2}{(\sqrt{ad}+\sqrt{bc})(\sqrt{ad}-\sqrt{bc})} = \frac{(\sqrt{ad}+\sqrt{bc})}{(\sqrt{ad}-\sqrt{bc})}$$

Bound: $\frac{\|r_k\|_2}{\|r_0\|_2} \leq \left(\frac{\sqrt{|ad|}-\sqrt{|bc|}}{\sqrt{|ad|}+\sqrt{|bc|}} \right)^{[k/2]}$.

Convergence bounds for MINRES

Let A be Hermitian and let $\lambda(A) \subset [a,b] \cup [c,d]$, where $a < b < 0 < c < d$ and $b - a = d - c$.

Bound for MINRES: $\frac{\|r_k\|_2}{\|r_0\|_2} \leq 2\left(\frac{\sqrt{|ad|} - \sqrt{|bc|}}{\sqrt{|ad|} + \sqrt{|bc|}}\right)^{[k/2]}$

In the case that $a = -d$ and $b = -c$ (symmetric w.r.t. the origin), we can simplify bound further (but bound does not get better):

$$\frac{\|r_k\|_2}{\|r_0\|_2} \leq 2\left(\frac{d-c}{d+c}\right)^{[k/2]} = 2\left(\frac{d/c-1}{d/c+1}\right)^{[k/2]} \quad \text{(note } \kappa = \tfrac{d}{c}\text{)}$$

In HPD case: $\frac{\|r_k\|_2}{\|r_0\|_2} \leq 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^{k}$.

So bound in indefinite case at iteration k is that of the definite case at iteration $k/2$ for matrix with condition number d^2/c^2.

Dramatic loss of convergence compared with definite case.

Convergence bounds for MINRES

If we know that A has only a few negative (positive) eigenvalues, we again can improve the bound significantly by taking product $p_s(z)T_{k-s}(z)$, where $p_s = 0$ on negative eigenvalues and T_{k-s} is scaled and shifted Chebyshev polynomial over positive eigenvalues. Product must also satisfy our normalization: $p_s(0)T_{k-s}(0) = 1$.

Let $\lambda_1 < \lambda_2 < \lambda_3 < 0 < \lambda_4 < \cdots < \lambda_n$.

Possibility: $\tilde{p}_k(z) = (z-\lambda_1)(z-\lambda_2)(z-\lambda_3)T_{k-3}\left(\frac{2z-\lambda_n-\lambda_4}{\lambda_n-\lambda_4}\right)$

Normalize: $p_k(z) = \tilde{p}_k(z)/\tilde{p}_k(0)$

$$p_k(z) = \frac{(z-\lambda_1)(z-\lambda_2)(z-\lambda_3)}{-\lambda_1\lambda_2\lambda_3}\left[T_{k-3}\left(\frac{2z-\lambda_n-\lambda_4}{\lambda_n-\lambda_4}\right)/T_{k-3}\left(\frac{-\lambda_n-\lambda_4}{\lambda_n-\lambda_4}\right)\right]$$

$p_k(z) \leq 2C_3\left(\frac{\sqrt{\kappa_4}-1}{\sqrt{\kappa_4}+1}\right)^{k-3}$, where $C_3 = \frac{(\lambda_n-\lambda_1)(\lambda_n-\lambda_2)(\lambda_n-\lambda_3)}{-\lambda_1\lambda_2\lambda_3}$ and $\kappa_4 = \frac{\lambda_n}{\lambda_4}$.

Note that $p_k(z)$ may not be good for small k.

Convergence bounds for GMRES

We will now consider convergence bounds for non-Hermitian problems solved by GMRES. This brings some important changes. First of all, A may not be diagonalizable. In this case we have to take ploynomials over Jordan blocks into account.
Second, the eigenvectors (and proper vectors) of A may not be orthogonal.

Let's assume A is diagonalizable: $A = V\Lambda V^{-1}$
We still have

$$\|r_k\|_2 \le \min_{p_k} \|Vp_k(\Lambda)V^{-1}r_0\|_2 \le \kappa(V)\min_{p_k}\|p_k(\Lambda)\|_2\|r_0\|_2 \Rightarrow$$

$$\|r_k\|_2/\|r_0\|_2 \le \kappa(V)\min_{p_k}\max_i |p_k(\lambda_i)|$$

Clearly, usefulness of bounding $\min_{p_k}\max_i |p_k(\lambda_i)|$ depends on $\kappa(V)$.
Sharp for normal A, approach still useful if A almost normal (V unitary).

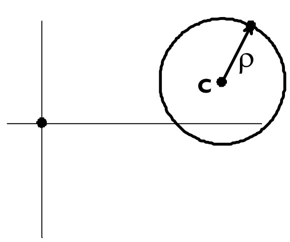
Convergence bounds for GMRES

Now we must find polynomials that are small over a region in the complex plane. More complicated than Hermitian case.
Generally we try to find 'simple' regions containing the eigenvalues, and devise polynomial over such a region (e.g. circle or ellipse).

Eigenvalues in circle $C(c,\rho)$ not containing the origin with center c and radius ρ:

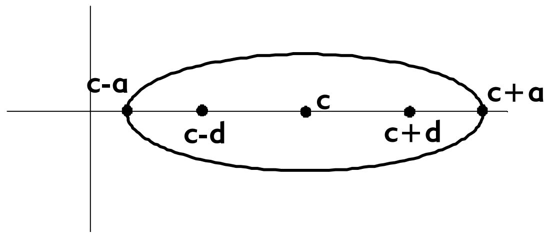
$$\min_{p_k(0)=1} \max_{z\in C(c,\rho)} |p_k(z)| = \left(\frac{\rho}{|c|}\right)^k$$

Obtained for polynomial $p_k(z) = \left(\frac{z-c}{0-c}\right)^k = (1 - z/c)^k$



Convergence bounds for GMRES

Eigenvalues in ellipse $E(c,d,a)$ not containing the origin with center c, major semi-axis a, and focal distance d.



$\|r_k\|_2/\|r_0\|_2 \le \kappa(V)\frac{T_k(a/d)}{T_k(c/d)}$, where $T_k(z) = \cosh k \cosh^{-1} z$.

$$\frac{T_k(a/d)}{T_k(c/d)} = \frac{\left(a/d+\sqrt{(a/d)^2-1}\right)^k + \left(a/d+\sqrt{(a/d)^2-1}\right)^{-k}}{\left(c/d+\sqrt{(c/d)^2-1}\right)^k + \left(c/d+\sqrt{(c/d)^2-1}\right)^{-k}} \approx \frac{\left(a+\sqrt{a^2-d^2}\right)^k}{\left(c+\sqrt{c^2-d^2}\right)^k}$$

Convergence bounds for GMRES

Although the eigenvalues often give important information about the convergence of GMRES, we have the following theorem that states this is not generally the case.

Theorem:
Given any set of eigenvalues and any non-increasing convergence curve, a matrix with those eigenvalues and a right hand side can be constructed for which GMRES will display the prescribed convergence curve.

So even with a 'nice' spectrum the convergence can be arbitrarily poor.
This does not have to the case. Nonnormal matrices are not inherently bad.

Convergence bounds for GMRES

Consider matrix

$$A = \begin{bmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & & 1 \\ c_0 & c_1 & c_2 & \cdots & c_{n-1} \end{bmatrix}$$

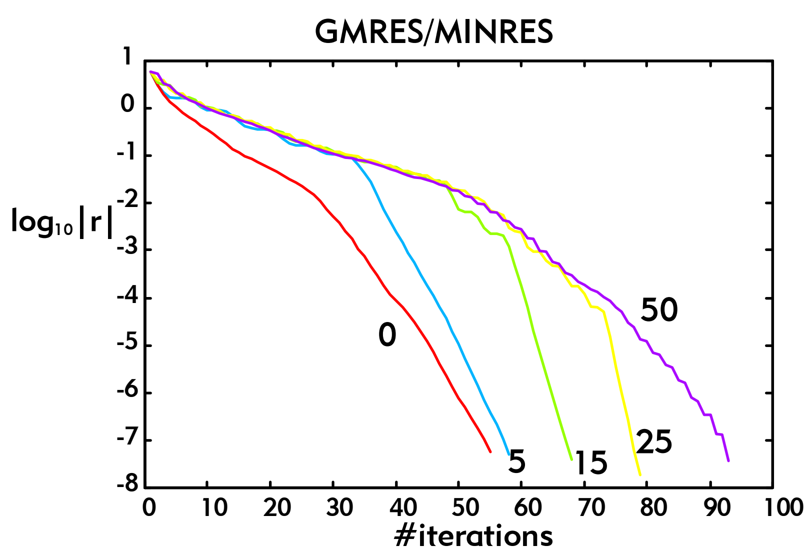
Eigenvalues are zeros of $p(\lambda) = \lambda^n - c_{n-1}\lambda^{n-1} - c_{n-2}\lambda^{n-2} - \cdots - c_0 = 0$.

Solving $Ax = e_1$ gives no convergence till last step.
Taking appropriate initial residual yields any convergence curve.

This matrix is rather special, but we get same behavior for any unitarily similar matrix. Note that this matrix is reordered lower triangular matrix, and any matrix is unitarily similar to some lower triangular matrix (Schur decomposition).

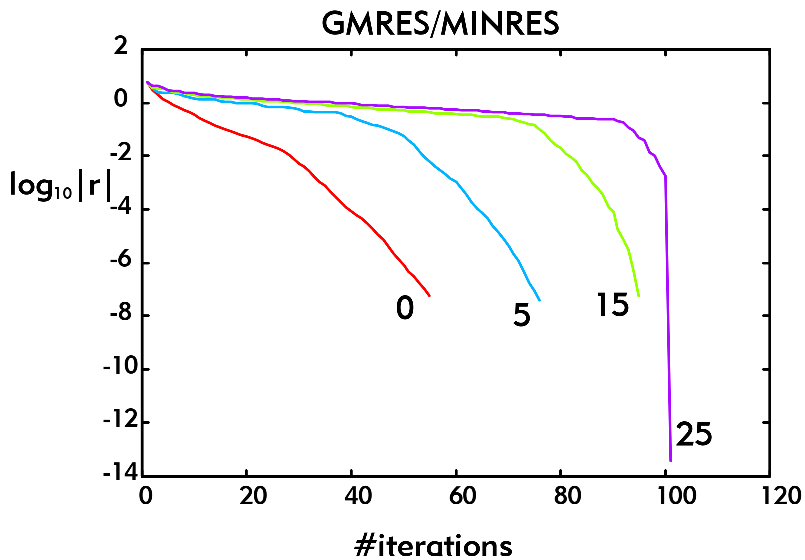
Convergence for Indefinite Problems

A=diag(-1,-2,...,-s,s+1,s+2,...,100)



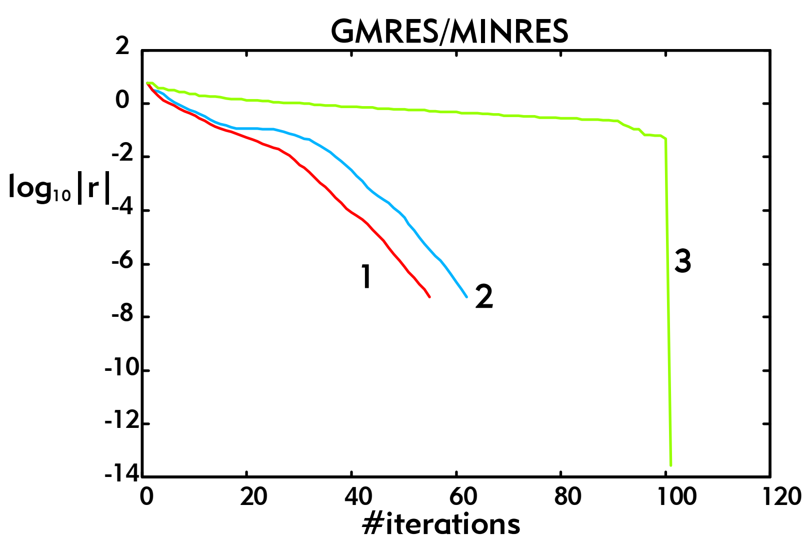
Convergence for Indefinite Problems

A=diag(1,2,...,25,-1,-2,...,-s,26+s,...,100)



Convergence for Indefinite Problems

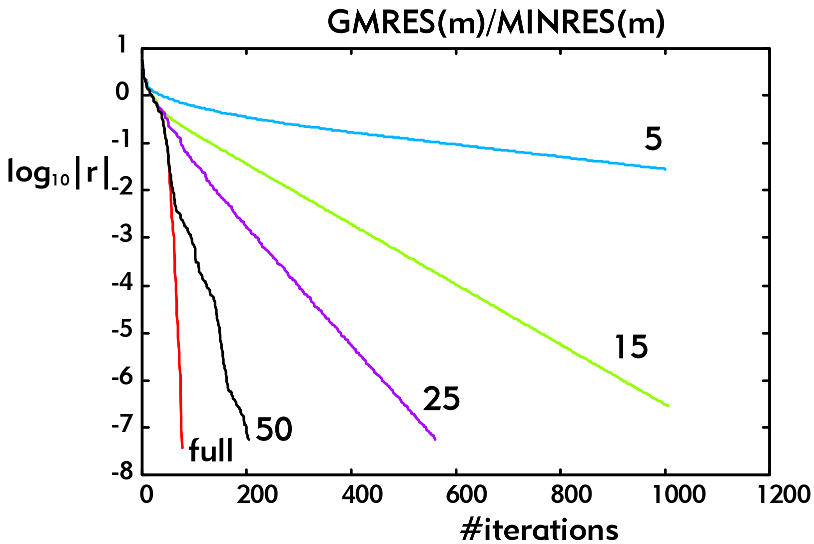
1: A=diag(1,2,3,...,100)
2: A=diag(-1,-100,1,2,...,49,52,53,...,100)
3: A=diag(-99,-97,...,-1,1,3,...,99)



Convergence for Indefinite Problems

A=diag(1,2,...,25,-1,-2,...,-5,31,32,...,100)

GMRES/MINRES without restart takes 76 iterations

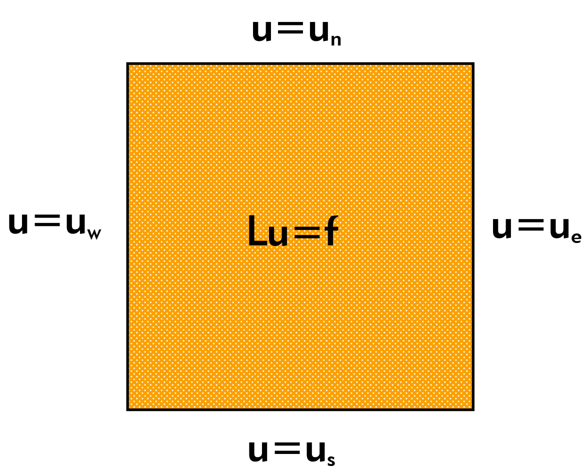


Model Problem

Convection-Diffusion(-Reaction) Equation

Dirichlet boundary conditions

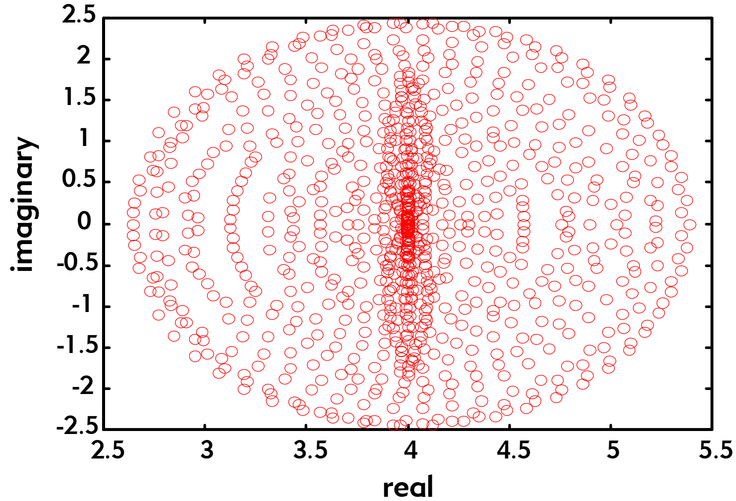
$$Lu = -(pu_x)_x - (qu_y)_y + ru_x + su_y + tu = f$$



Eigenvalues

p=q=1; r=s=70; h=1/31;

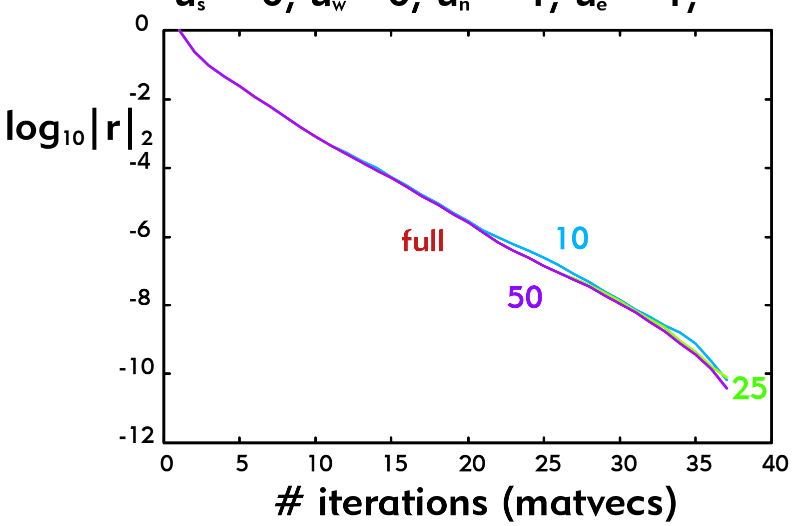
u_s=0; u_w=0; u_n=1; u_e=1;



GMRES(m)

p=q=1; r=s=70; h=1/ 31;

u_s = 0; u_w =0; u_n = 1; u_e = 1;



GMRES(m) after shifting spectrum

A=A-3.65*I
p=q=1; r=s=70; h=1/ 31;
u_s = 0; u_w =0; u_n = 1; u_e = 1;