Mathematics, Pusan National University

Numerical Linear Algebra

Lecture 35. GMRES

Taehyeong Kim
th_kim@pusan.ac.kr

March 1, 2021

Content

Residual Minimization in $\mathcal{K}_n$

Last two lecture...

$A \in \mathbb{C}^{m \times m}$: square matrix
$b \in \mathbb{C}^{m}$: vector
$\mathcal{K}_n$: Krylov subspace $\langle b, Ab, \ldots, A^{n-1}b \rangle$

In this lecture...

$A \in \mathbb{C}^{m \times m}$: nonsingular matrix
Goal : solving $Ax = b$
Denote the exact solution $x_* = A^{-1}b$

Residual Minimization in $\mathcal{K}_n$

Idea of GMRES

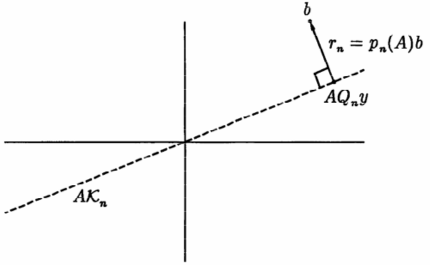


Figure 35.1. *The least squares polynomial approximation problem underlying GMRES: minimize the residual norm $\|r_n\|$. Compare Figure 34.1.*

Residual Minimization in $\mathcal{K}_n$

Let K_n be the $m \times n$ Krylov matrix that

$$K_n = \left[\begin{array}{c|c|c|c} b & Ab & \cdots & A^{n-1}b \end{array} \right]$$

In lecture 33, K_n must have a reduced QR factorization $K_n = Q_n R_n$.
And we have

$$AK_n = \left[\begin{array}{c|c|c|c} Ab & A^2b & \cdots & A^nb \end{array} \right] \tag{1}$$

The column space of AK_n is $A\mathcal{K}_n$.

Residual Minimization in $\mathcal{K}_n$

Problem

Find $c \in \mathbb{C}^n$ such that

$$\arg\min_c \|AK_n c - b\| \tag{2}$$

Once c is found, we would set $x_n = K_n c$.

And each K_n must have a reduced QR factorization

$$K_n = Q_n R_n$$

We can get Q_n after using Arnoldi iteration.

Problem

Find $y \in \mathbb{C}^n$ such that

$$\arg\min_y \|AQ_n y - b\| \tag{3}$$

We can write $x_n = Q_n y$ instead of $x_n = K_n c$

Residual Minimization in $\mathcal{K}_n$

Because of the special structure of Krylov subspaces, the dimension must be $(n+1) \times n$.

$$AQ_n = Q_{n+1}\tilde{H}_n$$

Problem

Find $y \in \mathbb{C}^n$ such that

$$\arg\min_y \|Q_{n+1}\tilde{H}_n y - b\| \tag{4}$$

Now both vectors inside the norm are in the column space of Q_{n+1}. Therefore, multiplying on the left by Q_{n+1}^* does not change that norm.

Problem

Find $y \in \mathbb{C}^n$ such that

$$\arg\min_y \|\tilde{H}_n y - Q_{n+1}^* b\| \tag{5}$$

Residual Minimization in $\mathcal{K}_n$

Finally, we note that by construction of the Krylov matrices $\{Q_n\}$, $Q_{n+1}^* b$ is equal to $\|b\|e_1$. Thus we reach at last the final form of the GMRES least squares problem:

Problem

Find $y \in \mathbb{C}^n$ such that

$$\arg\min_y \|A\tilde{H}_n y - \|b\|e1\| \quad (6)$$

Algorithm 35.1. GMRES

$q_1 = b/\|b\|$

for $n = 1, 2, 3, \ldots$

$\quad\langle$ *step n of Arnoldi iteration, Algorithm 33.1* $\rangle$

$\quad$ Find y to minimize $\|\tilde{H}_n y - \|b\| e_1\|$ $\;(= \|r_n\|)$

$\quad x_n = Q_n y.$

Mechanics of GMRES

Algorithm 33.1. Arnoldi Iteration

$b =$ arbitrary, $q_1 = b/\|b\|$

for $n = 1, 2, 3, \ldots$

$\quad v = Aq_n$

$\quad$ **for** $j = 1$ **to** n

$\quad\quad h_{jn} = q_j^* v$

$\quad\quad v = v - h_{jn} q_j$

$\quad h_{n+1,n} = \|v\|$ [see Exercise 33.2 concerning $h_{n+1,n} = 0$]

$\quad q_{n+1} = v/h_{n+1,n}$

Recall

Arnoldi/Lanczos Approximation Problem.
Find $p^n \in P^n$ such that

$$\arg\min_{p^n} \|p^n(A)b\|$$

where

$$P^n = \{\text{monic polynomials of degree } n\}$$

$$P_n = \{\text{polynomials } p \text{ of degree } \le n \text{ with } p(0) = 1\} \tag{7}$$

By using this, the iterate x_n can be written

$$x_n = q_n(A)b \tag{8}$$

where q is a polynomial of degree $n-1$. The corresponding residual $r_n = b - Ax_n$ is $r_n = (I - Aq_n(A))b$, where p_n is the polynomial defined by $p_n(z) = 1 - zq(z)$. Then we have

$$r_n = p_n(A)b$$

GMRES and Polynomial Approximation

Therefore, the GMRES process chooses the coefficients of p_n to minimize the norm of this residual.

GMRES Approximation Problem

Find $p_n \in P_n$ such that

$$\arg\min_{p_n} \|p_n(A)b\| \tag{9}$$

Theorem 3.1

Let the GMRES iteration be applied to a matrix $A \in \mathbb{C}^{m\times m}$ as described above.
Scale-invariance. *If A is changed to σA for some $\sigma \in \mathbb{C}$, and b is changed to σb, the residuals $\{r_n\}$ changed $\{\sigma r_n\}$.*
Invariance under unitary similarity transformations. *If A is changed to UAU^* for some unitary matrix U, and b is changed to Ub, the residuals $\{r_n\}$ changed $\{U^* r_n\}$.*

Convergence of GMRES

We use Monotone convergence theorem to determine convergence of most iterative methods.

Lemma 4.1

If a sequence of real numbers is increasing and bounded above, then its supremum is the limit.

Lemma 4.2

If a sequence of real numbers is decreasing and bounded below, then its infimum is the limit.

Theorem 4.3

If $\{a_n\}$ is a monotone sequence of real numbers, then this sequence has a finite limit if and only if the sequence is bounded.

Convergence of GMRES

But in GMRES cases, we use two observations instead of Monotone convergence theorem.
The first is that GMRES converges monotonically

$$\|r_{n+1}\| \leq \|r_n\| \tag{10}$$

The second is that after at most m steps the process must converge, at least in the absence of rounding errors:

$$\|r_m\| = 0 \tag{11}$$

The critical factor of the most of problems that determines the size of this quantity is usually $\|p_n(A)\|$.

$$\frac{\|r_n\|}{\|b\|} \leq \inf_{p_n \in P_n} \|p_n(A)\| \tag{12}$$

This inequality brings us to the mathematically elegant question: given a matrix A and a number n, how small can $\|p_n(A)\|$ be? This question is the basis of almost all analysis of convergence of Krylov subspace iterations for solving systems of equations.

Polynomials Small on the Spectrum

Given A and n, how small can $\|p_n(A)\|$? The standard way of obtaining estimates is to look for polynomials $p(z)$ that are as small as possible on the spectrum $\Lambda(A)$, while still satisfying $p(0) = 1$. If p is is a polynomial and S is a set in the complex plane, let us define the scalar $\|p\|_S$ by

$$\|p\|_S = \sup_{z \in S} |p(z)| \tag{13}$$

Suppose A is diagonalizable, satisfying $A = V\Lambda V^{-1}$. Since the *condition number* $\kappa(A) = \|A\|\|A^{-1}\|$, we have

$$\|p(A)\| \leq \|V\|\|p(\Lambda)\|\|V^{-1}\| = \kappa(V)\|p\|_{\Lambda(A)} \tag{14}$$

Polynomials Small on the Spectrum

Combining (14) with (12) gives the following basic theorem on convergence of GMRES.

Theorem 5.1

At step n of the GMRES iteration, the residual r_n satisfies

$$\frac{\|r_n\|}{\|b\|} \leq \inf_{p_n \in P_n} \|p_n(A)\| \leq \kappa(V) \inf_{p_n \in P_n} \|p\|_{\Lambda(A)} \tag{15}$$

where $\Lambda(A)$ is the set of eigenvalues of A, V is a nonsingular matrix of eigenvectors, and $\|p_n\|_{\Lambda(A)}$ is defined by (13).

This theorem can be summarized in words as follows. If A is not too far from normal in the sense that $\kappa(V)$ is not too large, and if properly normalized degree n polynomials can be found whose size on the spectrum $\Lambda(A)$ decreases quickly with n, then GMRES converges quickly.

Example 5.2

Let A be a 200×200 matrix whose entries are independent samples from the real normal distribution of mean 2 and standard deviation $0.5/\sqrt{200}$. In MATLAB,

$$\texttt{m = 200;} \quad \texttt{A = 2 * eye(m) + 0.5 * randn(m)/sqrt(m);} \tag{16}$$

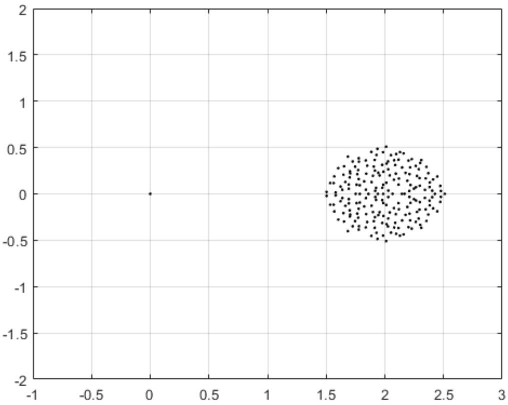
Our problem is

$$Ax = b \text{ where } b = (1, 1, \dots, 1)^*$$

The convergence in this case is extraordinarily steady at a rate approximately 4^{-n}. Since the spectrum of A approximately fills the disk indicated, $\|p(A)\|$ is approximately minimized by the choice $p(z) = (1 - z/2)^n$. Since $I - A/2$ is a random matrix scaled so that its spectrum approximately fills the disk of radius $1/4$ about 0, we have $\|p(A)\| = \|(I - A/2)^n\| \approx 4^{-n}$.

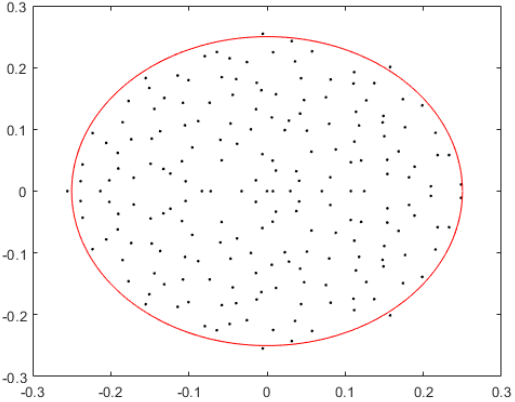
Polynomials Small on the Spectrum


```
%% example 1
clc; clear;
load('exampleseed.mat')
rng(s)
m = 200;
A = 2 * eye(m) + 0.5 * randn(m) / sqrt(m);
b = ones(m,1);

kappaA = norm(A,2) * norm(inv(A),2)
[V,D]=eig(A);
kappaV = norm(V,2) * norm(inv(V),2)

figure(1)
plot(D,'k.')
xlim([-1,3])
ylim([-2,2])
grid on

restart = 50;
maxit = 100;
tol = 1e-6;
[x0,fl0,rr0,it0,rv0] = gmres(A,b,restart,tol,maxit);

figure(2)
semilogy(0:length(rv0)-1,rv0/norm(b),'-o')
yline(tol,'r--');
grid on
```



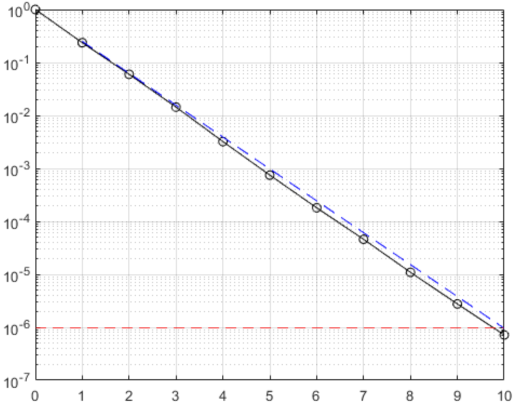
Polynomials Small on the Spectrum


```
%% example 1
clc; clear;
load('exampleseed.mat')
rng(s)
m = 200;
A = 2 * eye(m) + 0.5 * randn(m) / sqrt(m);
b = ones(m,1);

kappaA = norm(A,2) * norm(inv(A),2)
[V,D]=eig(A);
kappaV = norm(V,2) * norm(inv(V),2)

figure(1)
plot(D,'k.')
xlim([-1,3])
ylim([-2,2])
grid on

restart = 50;
maxit = 100;
tol = 1e-6;
[x0,fl0,rr0,it0,rv0] = gmres(A,b,restart,tol,maxit);

figure(2)
semilogy(0:length(rv0)-1,rv0/norm(b),'-o')
yline(tol,'r--');
grid on
```



Polynomials Small on the Spectrum


```
%% example 1
clc; clear;
load('exampleseed.mat')
rng(s)
m = 200;
A = 2 * eye(m) + 0.5 * randn(m) / sqrt(m);
b = ones(m,1);

kappaA = norm(A,2) * norm(inv(A),2)
[V,D]=eig(A);
kappaV = norm(V,2) * norm(inv(V),2)

figure(1)
plot(D,'k.')
xlim([-1,3])
ylim([-2,2])
grid on

restart = 50;
maxit = 100;
tol = 1e-6;
[x0,fl0,rr0,it0,rv0] = gmres(A,b,restart,tol,maxit);

figure(2)
semilogy(0:length(rv0)-1,rv0/norm(b),'-o')
yline(tol,'r--');
grid on
```



Example 5.3

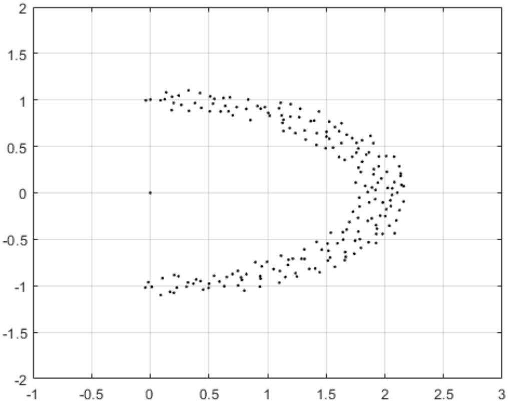
If the eigenvalues of a matrix "surround the origin," on the other hand, such rapid convergence cannot be expected. The matrix is now $A' = A + D$, where A is the matrix of (16) and D is the diagonal matrix with complex entries

$$d_k = (-2 + 2\sin\theta_k) + i\cos\theta_k, \quad \theta_k = \frac{k\pi}{m-1}, 0 \le k \le m-1$$

After this, the eigenvalues now lie in a semicircular cloud that bends around the origin. The convergence rate is much worse than before, making the iterative computation no better than Gaussian elimination for this problem. The condition numbers are now $\kappa(A) \approx 3.7790$ and $\kappa(V) \approx 78.3663$, so the deterioration in convergence cannot be explained by conditioning alone; it is the locations of the eigenvalues, not their magnitudes (or those of the singular values) that are causing the trouble. If the arc extended much further around the spectrum, the convergence would worsen further.

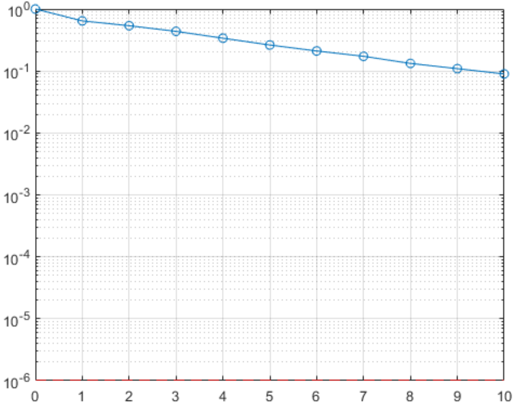
Polynomials Small on the Spectrum

```
%% example 2
for k = 0:m-1
    theta = k*pi / (m-1);
    d(k+1) = (-2 + 2*sin(theta)) + 1i * cos(theta);
end
dp = diag(d);
Ap = A+dp;

kappaAp = norm(Ap,2) * norm(inv(Ap),2)
[Vp,Dp]=eig(Ap);
kappaVp = norm(Vp,2) * norm(inv(Vp),2)

figure(3)
plot(Dp,'k.')
xlim([-1,3])
ylim([-2,2])
grid on

restart = 50;
maxit1 = it0(2);
maxit2 = 100;
tol = 1e-6;
[x1,fl1,rr1,it1,rv1] = gmres(Ap,b,[],tol,maxit1);
[x2,fl2,rr2,it2,rv2] = gmres(Ap,b,[],tol,maxit2);
[L,U]= lu(Ap);
L(abs(L)<0.01)=0;
U(abs(U)<0.01)=0;
[x3,fl3,rr3,it3,rv3] = gmres(Ap,b,[],tol,maxit2,L,U);

figure(4)
semilogy(0:length(rv1)-1,rv1/norm(b),'-o')
yline(tol,'r--');
grid on

figure(5)
semilogy(0:length(rv2)-1,rv2/norm(b),'-o')
hold on
semilogy(0:length(rv3)-1,rv3/norm(U₩(L₩b)),'-o')
yline(tol,'r--');
grid on
```



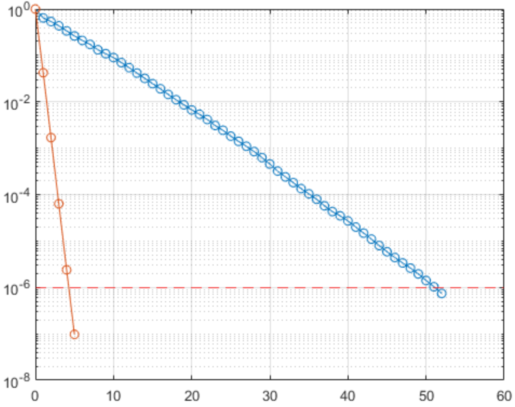
Polynomials Small on the Spectrum

```
%% example 2
for k = 0:m-1
    theta = k*pi / (m-1);
    d(k+1) = (-2 + 2*sin(theta)) + 1i * cos(theta);
end
dp = diag(d);
Ap = A+dp;

kappaAp = norm(Ap,2) * norm(inv(Ap),2)
[Vp,Dp]=eig(Ap);
kappaVp = norm(Vp,2) * norm(inv(Vp),2)

figure(3)
plot(Dp,'k.')
xlim([-1,3])
ylim([-2,2])
grid on

restart = 50;
maxit1 = it0(2);
maxit2 = 100;
tol = 1e-6;
[x1,fl1,rr1,it1,rv1] = gmres(Ap,b,[],tol,maxit1);
[x2,fl2,rr2,it2,rv2] = gmres(Ap,b,[],tol,maxit2);
[L,U]= lu(Ap);
L(abs(L)<0.01)=0;
U(abs(U)<0.01)=0;
[x3,fl3,rr3,it3,rv3] = gmres(Ap,b,[],tol,maxit2,L,U);

figure(4)
semilogy(0:length(rv1)-1,rv1/norm(b),'-o')
yline(tol,'r--');
grid on

figure(5)
semilogy(0:length(rv2)-1,rv2/norm(b),'-o')
hold on
semilogy(0:length(rv3)-1,rv3/norm(U\(L\b)),'-o')
yline(tol,'r--');
grid on
```



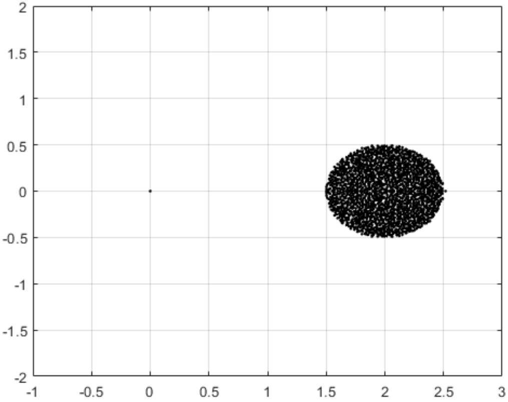
Polynomials Small on the Spectrum


```
%% example 2
for k = 0:m-1
    theta = k*pi / (m-1);
    d(k+1) = (-2 + 2*sin(theta)) + 1i * cos(theta);
end
dp = diag(d);
Ap = A+dp;

kappaAp = norm(Ap,2) * norm(inv(Ap),2)
[Vp,Dp]=eig(Ap);
kappaVp = norm(Vp,2) * norm(inv(Vp),2)

figure(3)
plot(Dp,'k.')
xlim([-1,3])
ylim([-2,2])
grid on

restart = 50;
maxit1 = it0(2);
maxit2 = 100;
tol = 1e-6;
[x1,fl1,rr1,it1,rv1] = gmres(Ap,b,[],tol,maxit1);
[x2,fl2,rr2,it2,rv2] = gmres(Ap,b,[],tol,maxit2);
[L,U]= lu(Ap);
L(abs(L)<0.01)=0;
U(abs(U)<0.01)=0;
[x3,fl3,rr3,it3,rv3] = gmres(Ap,b,[],tol,maxit2,L,U);

figure(4)
semilogy(0:length(rv1)-1,rv1/norm(b),'-o')
yline(tol,'r--');
grid on

figure(5)
semilogy(0:length(rv2)-1,rv2/norm(b),'-o')
hold on
semilogy(0:length(rv3)-1,rv3/norm(U\(L\b)),'-o')
yline(tol,'r--');
grid on
```

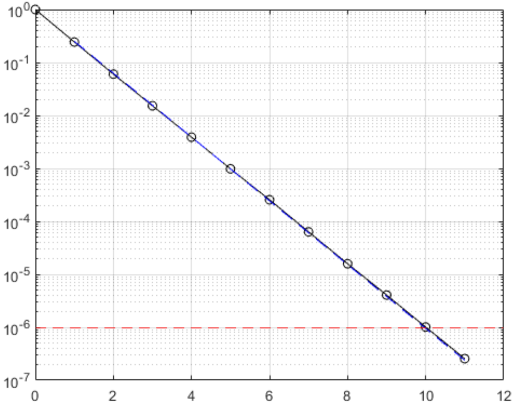


Polynomials Small on the Spectrum

```
%% example 1
clc; clear;
load('exampleseed.mat')
rng(s)
m = 200;
A = 2 * eye(m) + 0.5 * randn(m) / sqrt(m);
b = ones(m,1);

kappaA = norm(A,2) * norm(inv(A),2)
[V,D]=eig(A);
kappaV = norm(V,2) * norm(inv(V),2)

figure(1)
plot(D,'k.')
xlim([-1,3])
ylim([-2,2])
grid on

restart = 50;
maxit = 100;
tol = 1e-6;
[x0,fl0,rr0,it0,rv0] = gmres(A,b,restart,tol,maxit);

figure(2)
semilogy(0:length(rv0)-1,rv0/norm(b),'-o')
yline(tol,'r--');
grid on
```

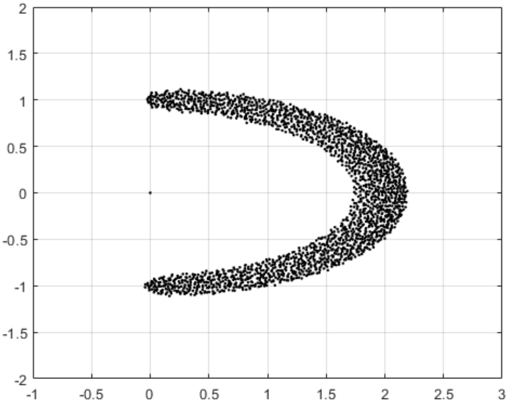


```
%% example 1
clc; clear;
load('exampleseed.mat')
rng(s)
m = 200;
A = 2 * eye(m) + 0.5 * randn(m) / sqrt(m);
b = ones(m,1);

kappaA = norm(A,2) * norm(inv(A),2)
[V,D]=eig(A);
kappaV = norm(V,2) * norm(inv(V),2)

figure(1)
plot(D,'k.')
xlim([-1,3])
ylim([-2,2])
grid on

restart = 50;
maxit = 100;
tol = 1e-6;
[x0,fl0,rr0,it0,rv0] = gmres(A,b,restart,tol,maxit);

figure(2)
semilogy(0:length(rv0)-1,rv0/norm(b),'-o')
yline(tol,'r--');
grid on
```



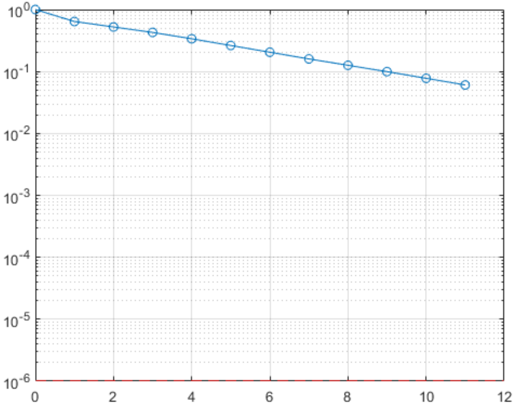
Polynomials Small on the Spectrum


```
%% example 2
for k = 0:m-1
    theta = k*pi / (m-1);
    d(k+1) = (-2 + 2*sin(theta)) + 1i * cos(theta);
end
dp = diag(d);
Ap = A+dp;

kappaAp = norm(Ap,2) * norm(inv(Ap),2)
[Vp,Dp]=eig(Ap);
kappaVp = norm(Vp,2) * norm(inv(Vp),2)

figure(3)
plot(Dp,'k.')
xlim([-1,3])
ylim([-2,2])
grid on

restart = 50;
maxit1 = it0(2);
maxit2 = 100;
tol = 1e-6;
[x1,fl1,rr1,it1,rv1] = gmres(Ap,b,[],tol,maxit1);
[x2,fl2,rr2,it2,rv2] = gmres(Ap,b,[],tol,maxit2);
[L,U]= lu(Ap);
L(abs(L)<0.01)=0;
U(abs(U)<0.01)=0;
[x3,fl3,rr3,it3,rv3] = gmres(Ap,b,[],tol,maxit2,L,U);

figure(4)
semilogy(0:length(rv1)-1,rv1/norm(b),'-o')
yline(tol,'r--');
grid on

figure(5)
semilogy(0:length(rv2)-1,rv2/norm(b),'-o')
hold on
semilogy(0:length(rv3)-1,rv3/norm(U₩(L₩b)),'-o')
yline(tol,'r--');
grid on
```



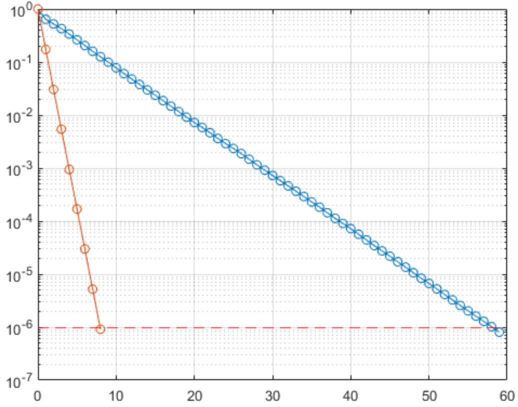
Polynomials Small on the Spectrum


```
%% example 2
for k = 0:m-1
    theta = k*pi / (m-1);
    d(k+1) = (-2 + 2*sin(theta)) + 1i * cos(theta);
end
dp = diag(d);
Ap = A+dp;

kappaAp = norm(Ap,2) * norm(inv(Ap),2)
[Vp,Dp]=eig(Ap);
kappaVp = norm(Vp,2) * norm(inv(Vp),2)

figure(3)
plot(Dp,'k.')
xlim([-1,3])
ylim([-2,2])
grid on

restart = 50;
maxit1 = it0(2);
maxit2 = 100;
tol = 1e-6;
[x1,fl1,rr1,it1,rv1] = gmres(Ap,b,[],tol,maxit1);
[x2,fl2,rr2,it2,rv2] = gmres(Ap,b,[],tol,maxit2);
[L,U]= lu(Ap);
L(abs(L)<0.01)=0;
U(abs(U)<0.01)=0;
[x3,fl3,rr3,it3,rv3] = gmres(Ap,b,[],tol,maxit2,L,U);

figure(4)
semilogy(0:length(rv1)-1,rv1/norm(b),'-o')
yline(tol,'r--');
grid on

figure(5)
semilogy(0:length(rv2)-1,rv2/norm(b),'-o')
hold on
semilogy(0:length(rv3)-1,rv3/norm(U\(L\b)),'-o')
yline(tol,'r--');
grid on
```



Polynomials Small on the Spectrum


```
%% example 2
for k = 0:m-1
    theta = k*pi / (m-1);
    d(k+1) = (-2 + 2*sin(theta)) + 1i * cos(theta);
end
dp = diag(d);
Ap = A+dp;

kappaAp = norm(Ap,2) * norm(inv(Ap),2)
[Vp,Dp]=eig(Ap);
kappaVp = norm(Vp,2) * norm(inv(Vp),2)

figure(3)
plot(Dp,'k.')
xlim([-1,3])
ylim([-2,2])
grid on

restart = 50;
maxit1 = it0(2);
maxit2 = 100;
tol = 1e-6;
[x1,fl1,rr1,it1,rv1] = gmres(Ap,b,[],tol,maxit1);
[x2,fl2,rr2,it2,rv2] = gmres(Ap,b,[],tol,maxit2);
[L,U]= lu(Ap);
L(abs(L)<0.01)=0;
U(abs(U)<0.01)=0;
[x3,fl3,rr3,it3,rv3] = gmres(Ap,b,[],tol,maxit2,L,U);

figure(4)
semilogy(0:length(rv1)-1,rv1/norm(b),'-o')
yline(tol,'r--');
grid on

figure(5)
semilogy(0:length(rv2)-1,rv2/norm(b),'-o')
hold on
semilogy(0:length(rv3)-1,rv3/norm(U\(L\b)),'-o')
yline(tol,'r--');
grid on
```



Reference

[1] Walter Edwin Arnoldi. "The principle of minimized iterations in the solution of the matrix eigenvalue problem". In: *Quarterly of applied mathematics* 9.1 (1951), pp. 17–29.

[2] Nicholas J Higham. *Functions of matrices: theory and computation*. SIAM, 2008.

[3] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.

[4] Gilbert Strang. *Linear algebra and learning from data*. Wellesley-Cambridge Press, 2019.

[5] Lloyd N Trefethen and David Bau III. *Numerical linear algebra*. Vol. 50. Siam, 1997.

Thank you!