Lecture 3: The Generalised Minimum Residual Method (GMRES)

Dr Qianqian Yang

School of Mathematical Sciences

Queensland University of Technology

Outline

- GMRES: Introduction
- GMRES: Development
- GMRES: Examples
- GMRES: Convergence analysis
- Restarted GMRES

GMRES: Introduction

- So far we have looked at projection methods for the eigenvalue problem where our main tool was Arnold's method: a Krylov subspace projection method.
- Now, we turn to the other big problem in linear algebra: solving systems of equations
$$Ax = b.$$
- The standard algorithm of this kind is known as GMRES, which stands for "Generalised Minimum Residuals".

GMRES: Development

- To solve $Ax = b$, once again we build the Krylov subspace

$$\mathcal{K}_m(A, b) = \text{span}\{b, Ab, \ldots, A^{m-1}\, b\}.$$

- Here, the vector is not arbitrary; it is the RHS vector of the linear system.
- So, our approximate solution $x^{(m)}$, which we will choose (somehow) from the Krylov subspace $\mathcal{K}_m$ will be a linear combination of the vectors b, Ab, ..., $A^{m-1}b$.
- That is, $x^{(m)} = c_0\, b + c_1\, A\, b + \ldots c_{m-1}\, A^{m-1}\, b$ for some coefficients c_i. Or, in other words, $x^{(m)} = q_{m-1}(A)\, b$ for some degree $m-1$ polynomial q_{m-1}.
- According to the Arnoldi decomposition $AV_m = V_{m+1}\bar{H}_m$, our solution $x^{(m)}$ will be given by

$$x^{(m)} = V_m\, y_m$$

for some coordinate vector y_m.

GMRES: Development

- Question: how do we know a good approximate solution could be found in this Krylov space $\mathcal{K}_m$?
- The Cayley–Hamilton theorem: if $P_n(\lambda) = \det(\lambda I - A)$ is the characteristic polynomial of A, then $P_n(A) = 0$. That is,

$$a_0\, I + a_1\, A + a_2\, A^2 + \ldots + a_n\, A^n = 0, \text{ where } a_n = 1.$$

- Multiply A^{-1} on both sides

$$a_0\, A^{-1} + a_1\, I + a_2\, A + \ldots + A^{n-1} = 0$$

$$A^{-1} = -\frac{a_1}{a_0}\, I - \frac{a_2}{a_0}\, A - \ldots - \frac{1}{a_0}\, A^{n-1} = \tilde{P}_{n-1}(A)$$

- The exact solution $x = A^{-1}b = \tilde{P}_{n-1}(A)b \ \in \mathcal{K}_n(A, b)$
- The approximate solution $x^{(m)} = q_{m-1}(A)\, b = V_m\, y_m \ \in \mathcal{K}_m(A, b)$, note $m < n$

GMRES: Development

- Question: how do we choose y_m such that $x^{(m)}$ is a good approximation of x?
- This is what GMRES does: it minimises the residual norm

$$\|r^{(m)}\| = \|b - A\,x^{(m)}\|.$$

- Using the Arnoldi decomposition,

$$r^{(m)} = b - A\,x^{(m)} = b - A\,V_m\,y_m = b - V_{m+1}\,\bar{H}_m\,y_m$$

- Note the first column of V_{m+1} is $v_1 = b/\beta$, with $\beta = \|b\|$. So

$$\begin{aligned} r^{(m)} &= \beta\,v_1 - V_{m+1}\,\bar{H}_m\,y_m = \beta\,V_{m+1}\,e_1 - V_{m+1}\,\bar{H}_m\,y_m \\ &= V_{m+1}(\beta\,e_1 - \bar{H}_m\,y_m). \end{aligned}$$

- And hence,

$$\|r^{(m)}\| = \|V_{m+1}(\beta\,e_1 - \bar{H}_m\,y_m)\| = \|\beta\,e_1 - \bar{H}_m\,y_m\|$$

since ${V_{m+1}}^T\,V_{m+1} = I$.

GMRES: Development

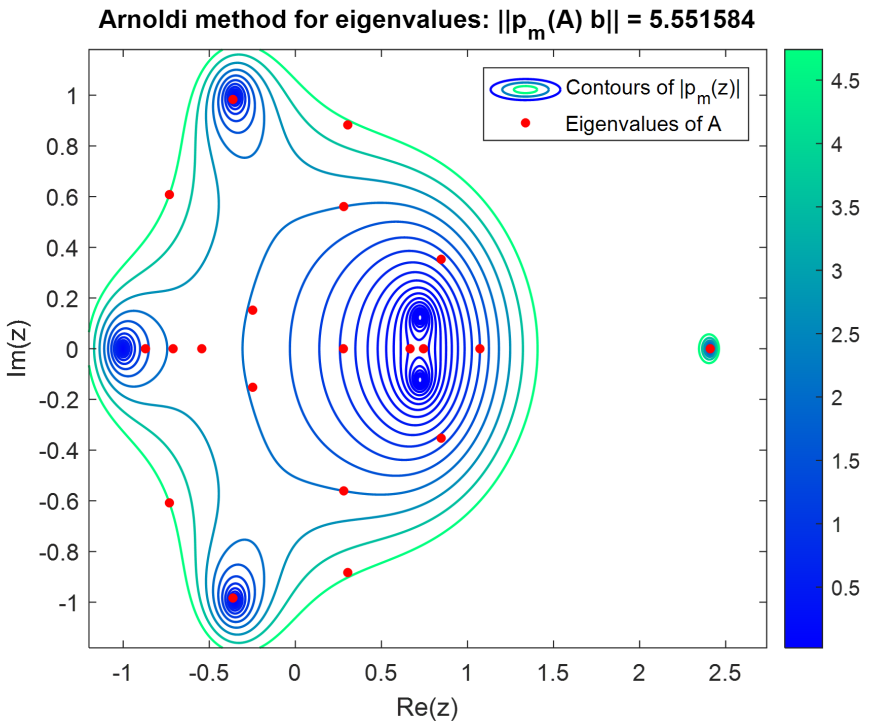
- So the n-dimensional minimisation reduces to the m-dimensional least squares problem:

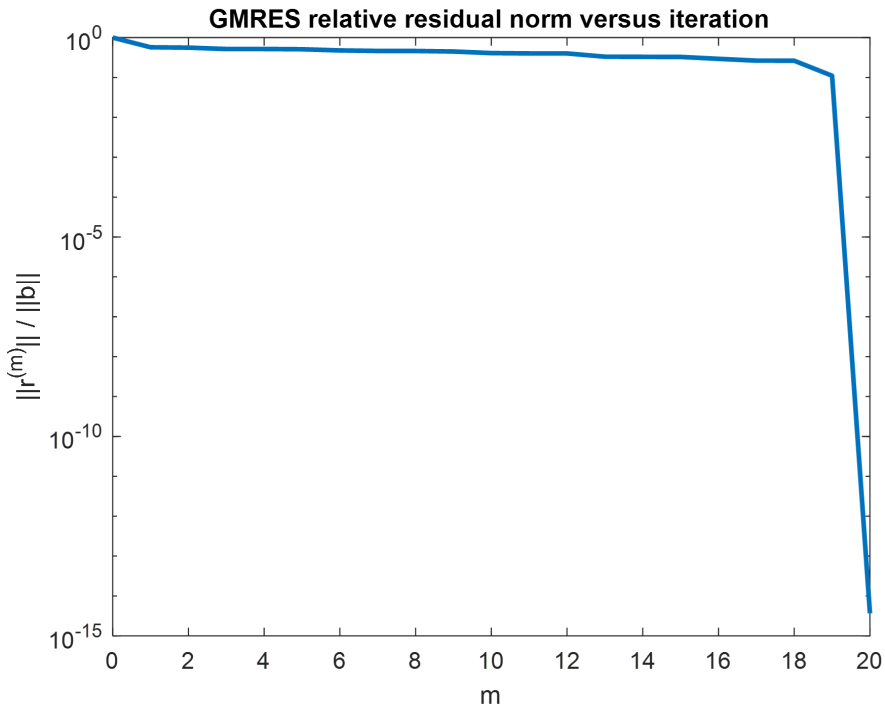
$$\min \|\beta\, e_1 - \bar{H}_m\, y_m\| \text{ over } y_m \in \mathbb{R}^m\,.$$

- That is, to "solve" $\bar{H}_m\, y_m = \beta\, e_1 \;\rightarrow\; y_m = \bar{H}_m \setminus (\beta\, e_1)$.
- Our approximate solution $x^{(m)} = V_m\, y_m$.

GMRES: Example 1

```
rng('default'); n = 20; density = 0.3;
A = sprand(n, n, density); b = rand(n,1); m = 6;
[x, flag, relres, iter, resvec] = gmres(A, b, [], 0, n);
```





GMRES: Example 1

- This is discouraging!
- GMRES is behaving more like a direct method: it is no use at all until it "finishes" by reaching $m = n$ iterations, at which point it's found the exact solution.
- It's ok for this example with $n = 20$, but it's no use when n is large.
- But should we really expect we can do better than taking $m = n$ iterations? Perhaps this whole idea of projection methods for linear systems is no good?
- Before we throw it all away, let's take a moment to analyse what's happening, just in case we can salvage things somehow.

GMRES: Convergence analysis

Questions we have

- How quickly does GMRES converge?
- How many iterations m must be taken before $\|r^{(m)}\| / \|b\|$ is reduced to a satisfactory level such as 10^{-3} or 10^{-6}?
- What properties of A determine the size of $\|r^{(m)}\|$?

GMRES: Convergence analysis

Let's do some analysis:

- Recall that GMRES produces a solution at each iteration which is a polynomial in A times b: $x^{(m)} = q_{m-1}(A)\, b$.
- The residual vector is

$$r^{(m)} = b - A\, x^{(m)} = b - A\, q_{m-1}(A)\, b = (I - A\, q_{m-1}(A))\, b = \tilde{p}_m(A)\, b,$$

  where $\tilde{p}_m(z) = 1 - z\, q_{m-1}(z)$ is a polynomial of degree m with constant coefficient 1 (coefficient of z^0).
- The norm of $r^{(m)}$ is $\|r^{(m)}\| = \|\tilde{p}_m(A)\, b\|$.
- We are (again!) solving the problem of minimising $\|\tilde{p}_m(A)\, b\|$, but this time over a different space of polynomials.
- Previously, in the Arnoldi method, our polynomial $p_m(z) = \det(z\, I - H_m)$ has leading coefficient of 1 (coefficient of z^m).

GMRES: Convergence analysis

- Let's focus on GMRES. We have

$$\|r^{(m)}\| = \|\tilde{p}_m(A)\, b\| \leq \|\tilde{p}_m(A)\|\, \|b\|,$$

  or, in terms of the relative residual

$$\frac{\|r^{(m)}\|}{\|b\|} \leq \|\tilde{p}_m(A)\|.$$

- Since GMRES minimises the left hand side, we can write

$$\frac{\|r^{(m)}\|}{\|b\|} \leq \inf_{\tilde{p}_m \in P_m} \|\tilde{p}_m(A)\|$$

  where P_m denotes the space of degree m polynomials with constant coefficient 1, i.e. $\tilde{p}_m(z) = c_m z^m + \ldots + c_1 z + 1$.
- This inequality determines the convergence rate of GMRES.

GMRES: Convergence analysis

- Question: Given a matrix A and a subspace size m, how small can $\|\tilde{p}_m(A)\|$ be?
- Suppose A is diagonalisable, with $A = X\,\mathrm{diag}(\lambda_i)\,X^{-1}$. Then

$$\tilde{p}_m(A) = X\,\mathrm{diag}(\tilde{p}_m(\lambda_i))\,X^{-1}$$

Take norms on both sides

$$\begin{aligned}\|\tilde{p}_m(A)\| &= \|X\,\mathrm{diag}(\tilde{p}_m(\lambda_i))\,X^{-1}\| \\ &\leq \|X\|\,\|\mathrm{diag}(\tilde{p}_m(\lambda_i))\|\,\|X^{-1}\| \\ &\leq \mathrm{cond}(X) \sup_{z\in\sigma(A)} |\tilde{p}_m(z)|\end{aligned}$$

where $\sigma(A) = \{\lambda_1, \lambda_2, \cdots, \lambda_n\}$ is the spectrum of A, and $\mathrm{cond}(X) = \|X\|\,\|X^{-1}\|$ is the condition number of the matrix of eigenvectors.

GMRES: Convergence analysis

- Combining with the earlier result, we obtain the following theorem on the convergence of GMRES

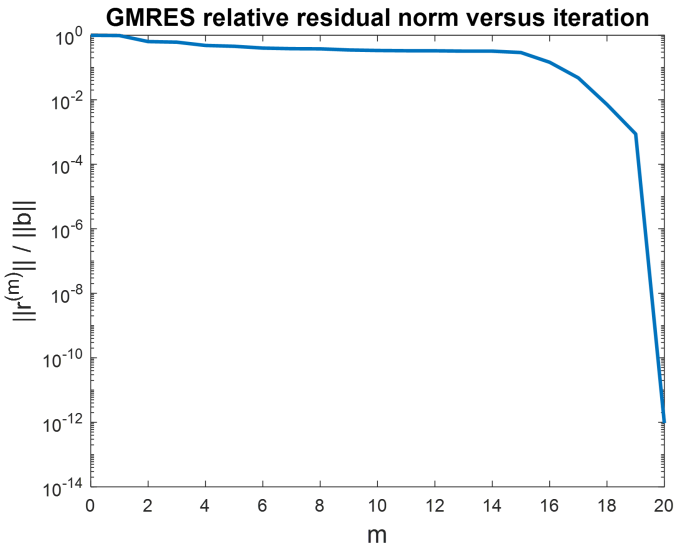
$$\frac{\|r^{(m)}\|}{\|b\|} \leq \mathrm{cond}(X) \inf_{\tilde{p}_m \in P_m} \sup_{z \in \sigma(A)} |\tilde{p}_m(z)|,$$

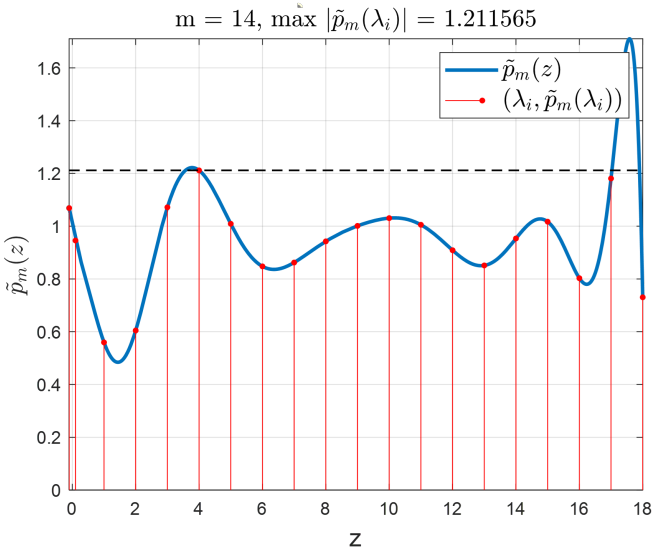
  where $\tilde{p}_m(z) = 1 - z\, q_{m-1}(z)$ is a polynomial of degree m with constant coefficient 1 ($\tilde{p}_m(0) = 1$).

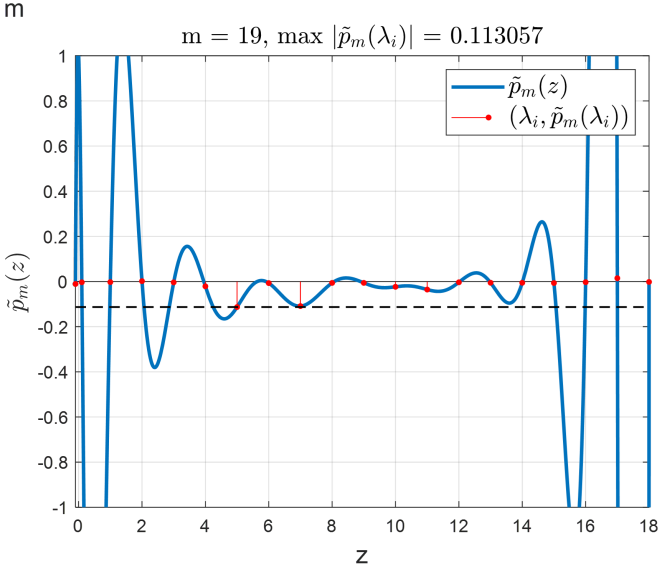
- Summary of this theorem:
  If the condition number $\mathrm{cond}(X)$ is not too large, and if degree m polynomials $\tilde{p}_m(z)$ can be found whose magnitude on the spectrum of A, $\sigma(A)$, decreases quickly with m, then GMRES converges quickly.

GMRES: Example 2 (small eigenvalues)

Let's build a 20x20 matrix with real eigenvalues [-0.1, 0.1, 1, 2, ..., 18] and random eigenvectors, and run GMRES for different choices of m.

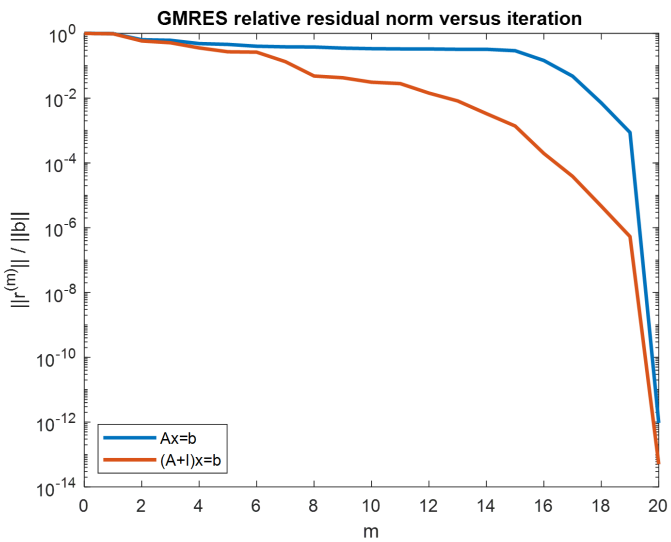


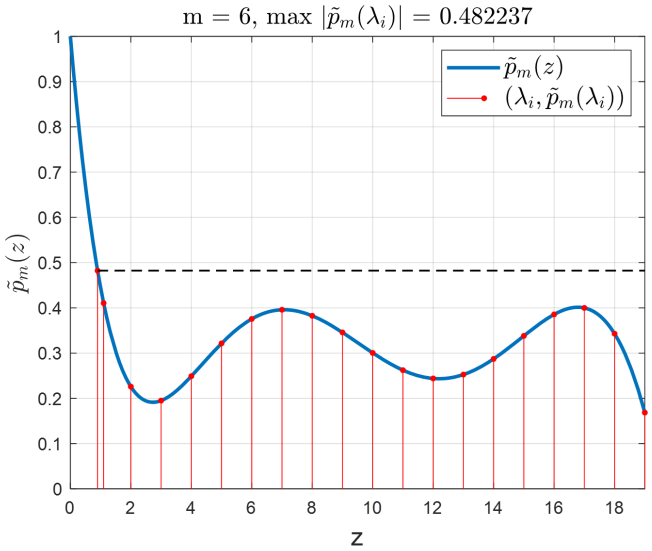


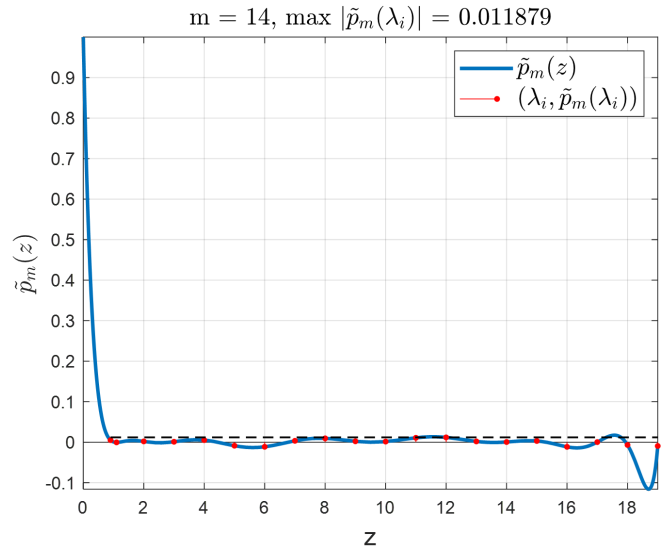


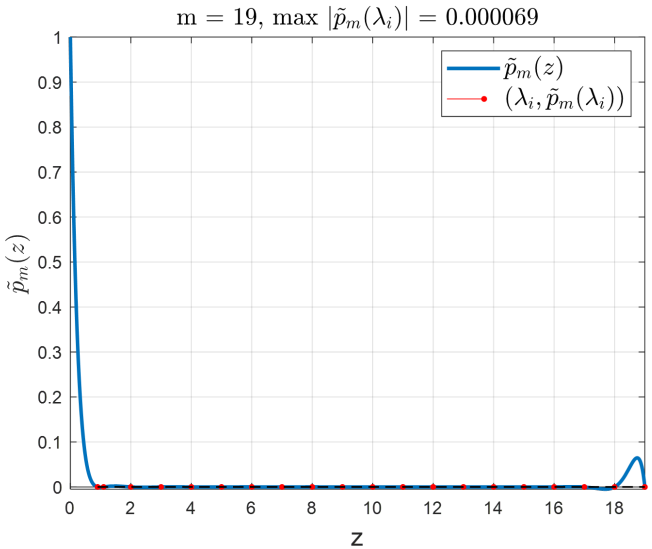
GMRES: Example 3 (small eigenvalues shifted)

Take the matrix from Example 2, and solve $(A + I)\, x = b$, instead of solving $A\, x = b$.



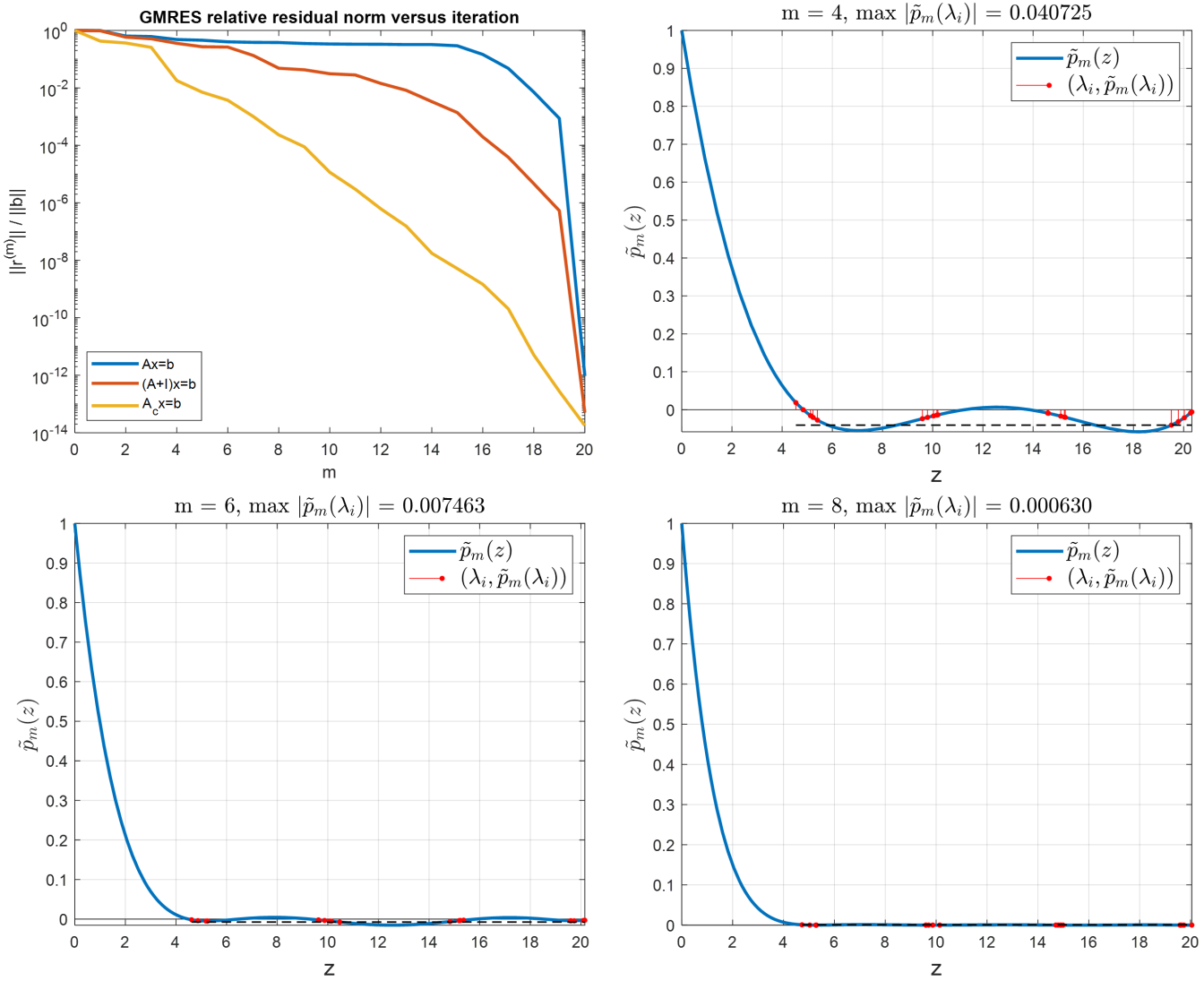






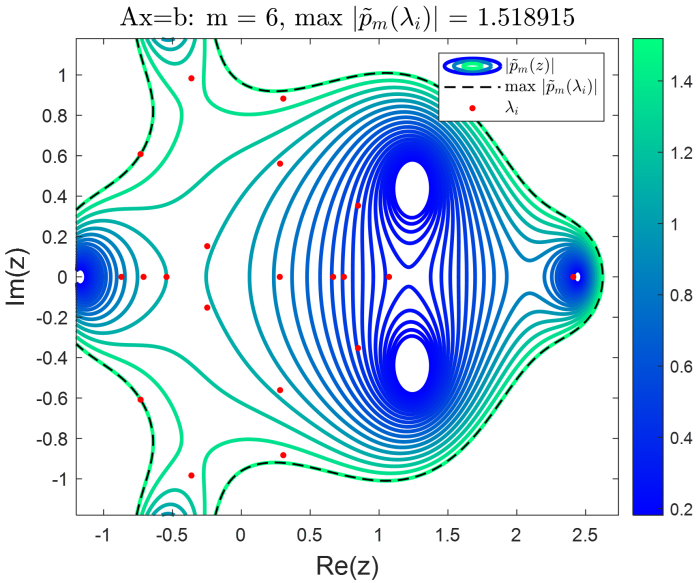
GMRES: Example 4 (clustered eigenvalues)

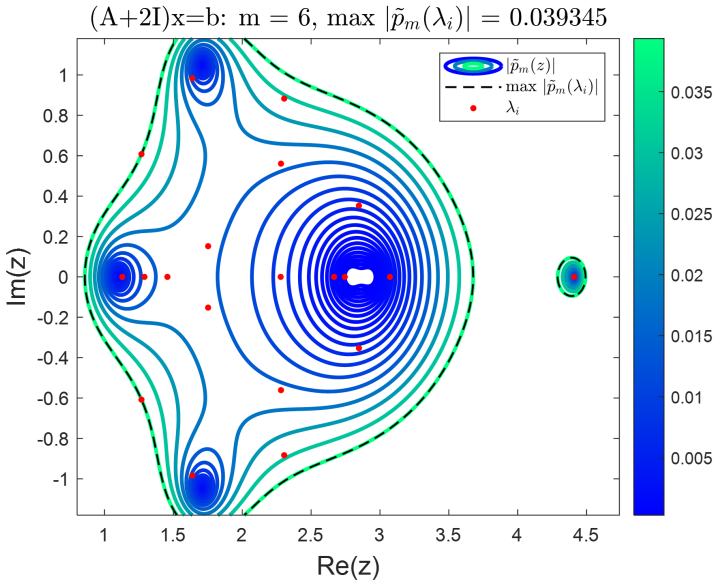
Let's build another 20×20 matrix A_c with eigenvalues clustered around 5, 10, 15, and 20.

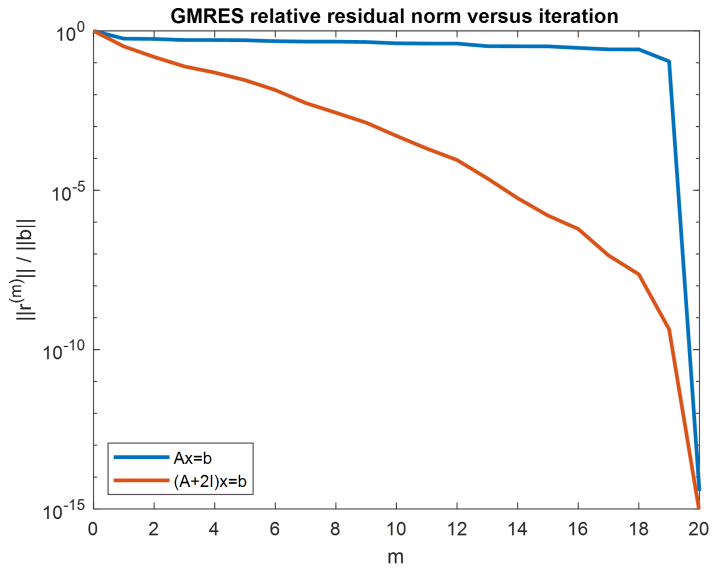


GMRES: revisit Example 1

Let's solve $A\,x = b$ and $(A + 2\,I)\,x = b$ using GMRES with $m = 6$.







GMRES: Key points on the convergence

In summary, for GMRES to solve $Ax = b$:

- the rate of convergence depends on the condition number of the eigenvector matrix, $\mathrm{cond}(X)$, and how quick $|\tilde{p}_m(\lambda_i)|$ goes down with m (note $\tilde{p}_m(0) = 1$);
- small eigenvalues are not desirable;
- clustering of eigenvalues is desirable.