Numerical Linear Algebra

Maya Neytcheva, TDB, February-March 2021

General framework – projection methods (short repetition)

Want to solve $A\mathbf{x} = \mathbf{b}, \mathbf{b}, \mathbf{x} \in R^n, A \in R^{n,n}$

Use the projection framework, i.e., we seek an approximate solution $\widetilde{\mathbf{x}} = \mathbf{x}^0 + \delta$, where $\delta \in K$, $dim(K) = m \ll n$, such that

$$\mathbf{b} - A\widetilde{\mathbf{x}} \perp L, dim(L) = m$$

$\mathbf{x}^0$ is arbitrary.

General framework – projection methods

Major results:

(A) The matrix $B = W^T AV$ is nonsingular for any W and V either if A is positive definite and $L = K$, or if A^{-1} exists and $L = AK$.

(B) Properties

- **(I)** $K = L$, A-spd $\Rightarrow$ $\|\mathbf{x}^* - \widetilde{\mathbf{x}}\|_A \leq \|\mathbf{x}^* - \mathbf{x}\|_A$ for any $\mathbf{x} = \mathbf{x}^0 + \mathbf{y}, \mathbf{y} \in K$
- **(II)** $L = AK$, $\Rightarrow$ $\|\mathbf{b} - A\widetilde{\mathbf{x}}\|_2 \leq \|\mathbf{b} - A\mathbf{x}\|_2$ for any $\mathbf{x} = \mathbf{x}^0 + \mathbf{y}, \mathbf{y} \in K$

General framework – projection methods

The important question is now how to choose K. We let

$$K \equiv \mathcal{K}^m(A, \mathbf{v}) = span\{\mathbf{v}, A\mathbf{v}, \cdots, A^{m-1}\mathbf{v}\}$$

for some vector $\mathbf{v}$.
Usual choices: $\mathbf{v} = \mathbf{b}$ or $\mathbf{v} = \mathbf{r}^0 \equiv \mathbf{b} - A\mathbf{x}^0$.

Relevant questions:

- Why is $\mathcal{K}(A, \mathbf{b})$ often a good space from which to construct an approximate solution?
- Why are eigenvalues important for Krylov methods?
- Why do Krylov methods often do so well for Hermitian matrices?

One can show that the solution of $A\mathbf{x} = \mathbf{b}$ has a natural representation in $\mathcal{K}_k(A, \mathbf{b})$ for some k.
If k happens to be small, we have a fast convergence.

The GMRES method

$L = AK$ + basic projection step

$$\|\mathbf{b} - A\widetilde{\mathbf{x}}\|_2 = \min_{\mathbf{x} \in \mathbf{x}^0 + K} \|\mathbf{b} - A\mathbf{x}\|_2$$

Basic GMRES

Choose $\mathbf{v}^{(1)}$ to be the normalized residual $\mathbf{r}^{(0)} = \mathbf{b} - A\mathbf{x}^{(0)}$.
Any vector $\mathbf{x} \in \mathbf{x}^{(0)} + K$ is of the form $\mathbf{x} = \mathbf{x}^{(0)} + V_m\mathbf{y}$. Then

$$\begin{aligned} \mathbf{b} - A\mathbf{x} &= \mathbf{b} - A(\mathbf{x}^{(0)} + V_m\mathbf{y}) \\ &= \mathbf{r}^{(0)} - AV_m\mathbf{y} \\ &= \beta\mathbf{v}^{(1)} - V_{m+1}\overline{H}_m\mathbf{y} \\ &= V_{m+1}(\beta\mathbf{e}_1 - \overline{H}_m\mathbf{y}). \end{aligned}$$

Since the columns of V_{m+1} are orthonormal, then

$$\|\mathbf{b} - A\mathbf{x}\|_2 = \|\beta\mathbf{e}_1 - \overline{H}_m\mathbf{y}\|_2.$$

Basic GMRES

1. Compute $\mathbf{r}^{(0)} = \mathbf{b} - A\mathbf{x}^{(0)}$, $\beta = \|\mathbf{r}^{(0)}\|_2$ and $\mathbf{v}^{(1)} = \mathbf{r}^{(0)}/\beta$
2. For $k = 1, 2, \cdots, m$
3. Compute $\mathbf{w}^{(k)} = A\mathbf{v}^{(k)}$
4. For $i = 1, 2, \cdots, k$
5. $h_{ik} = (\mathbf{w}^{(k)}, \mathbf{v}^{(i)})$
6. $\mathbf{w}^{(k)} = \mathbf{w}^{(k)} - h_{ik}\mathbf{v}^{(i)}$
7. End
8. $h_{k+1,k} = \|\mathbf{w}^{(k)}\|_2$; if $h_{k+1,k} = 0$, set $m = k$, goto 11
9. $\mathbf{v}^{(k+1)} = \mathbf{w}^{(k)}/h_{k+1,k}$
10. End
11. Define the $(m+1) \times m$ Hessenberg matrix $\overline{H}_m = \{h_{ik}\}$, $1 \le i \le m+1, 1 \le k \le m$
12. Compute $\mathbf{y}^{(m)}$ as the minimizer of $\|\beta\mathbf{e}_1 - \overline{H}_m\mathbf{y}\|_2$ and $\mathbf{x}^{(m)} = \mathbf{x}^{(0)} + V_m\mathbf{y}^{(m)}$

GMRES:

- No breakdown of GMRES
- As m increases, storage and work per iteration increase fast. Remedies:
  - Restart (keep m constant)
  - Truncate the orthogonalization process
- The norm of the residual in the GMRES method is monotonically decreasing. However, the convergence may stagnate. The rate of convergence of GMRES cannot be determined so easy as that of CG.
- The convergence history depends on the initial guess.

GMRES: convergence

Theorem: Let A be diagonalizable, $A = X^{-1}\Lambda X$ where $\Lambda = diag\{\lambda_1, \cdots, \lambda_n\}$ contains the eigenvalues of A. Define

$$\epsilon^m = \min_{p \in \Pi_m^1} \max_{i=1,\cdots n} |p(\lambda_i)|.$$

Then, the residual norm at the mth step of GMRES satisfies

$$\|\mathbf{r}^{(m)}\| \leq \kappa(X)\epsilon^m \|\mathbf{r}^{(0)}\|,$$

where $\kappa(X) = \|X\| \, \|X^{-1}\|$.

GMRES: convergence

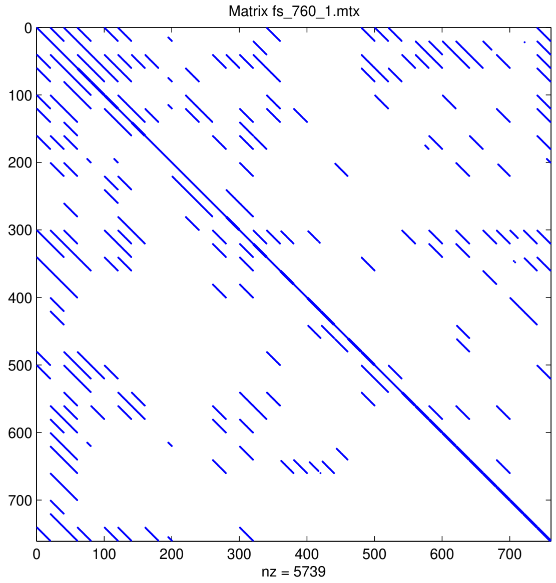
The rate of convergence of the GMRES method depends on the distribution of the eigenvalues of A in the complex plane. For fast convergence the eigenvalues need to be clustered away from the origin. Note that the eigenvalue distribution is much more important than the condition number of A, which is the main criterion for rapid convergence of the conjugate gradient method

GMRES history

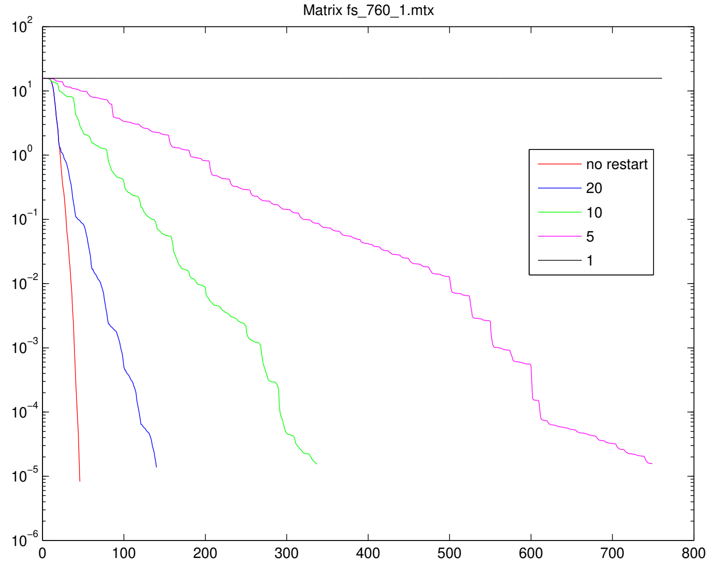
- Full GMRES, Saad and Schultz, 1986, costly
- Restarted GMRES - GMRES(m), Saad and Schultz, 1986
  Use $x^{(m)}$ as the starting guess for a fresh run of GMRES.
  Moral: Restarting compromises global optimality.
  GMRES with no restarts converges exactly in no more than n steps; restarted GMRES may fail to converge.

GMRES convergence history

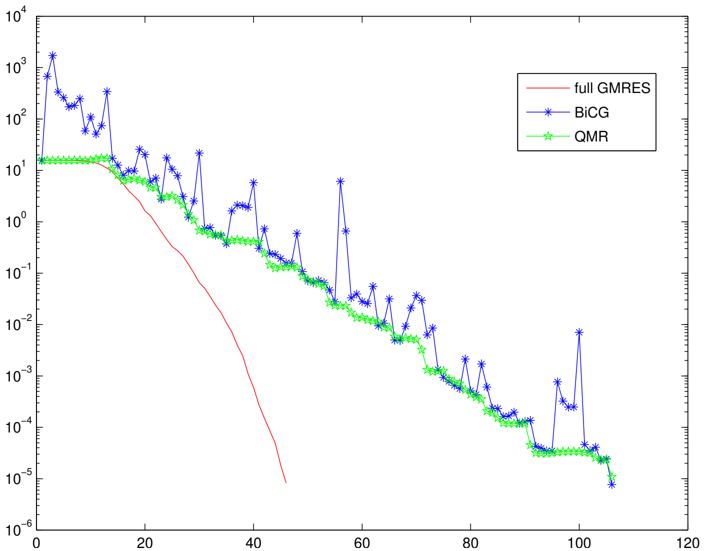
The matrix `fs_760_1.mtx` (Matrix Market)



GMRES convergence history



Other methods



Result:

A. Greenbaum, Estimating the attainable accuracy of recursively computed residual methods, *SIAM Journal on Matrix Analysis and Applications*, 18 (3), 1997, 535-551.

Estimate:

$$\frac{\mathbf{b} - A\mathbf{x}^k - \mathbf{r}^k}{\|A\|\|\mathbf{x}} = \varepsilon O(k)(1 + \max_{j\leq k} \|\mathbf{x}^j\|/\|\mathbf{x}\|$$

where ε is machine accuracy.

Earlier, it has been noted that an increase in the 2-norm of the residual at intermediate steps leads to a corresponding increase in the size of the final residual.

The above inequality shows that it is not really the size of intermediate residuals that is of importance but the size of the iterates.

In the paper, there is an example in which the residual remains small but intermediate iterates grow, causing a loss of accuracy in the final solution.

O. Axelsson, *Iterative solution methods*, Canbridge Univ. Press, 1994.

G. Golub and C.F. van Loan, *Matrix computations*, The Johns Hopkins University Press, 1996 (Third edition).

A. Greenbaum, *Iterative methods for solving linear systems*, Frontiers in Applied Mathematics, SIAM, 1997.

A. Greenbaum, Estimating the attainable accuracy of recursively computed residual methods, *SIAM Journal on Matrix Analysis and Applications*, 18 (3), 1997, 535-551.

Y. Saad, *Iterative methods for sparse linear systems*, PWS Publishing Company, 1996.

L.N. Trefethen and D. Bau, *Numerical Linear Algebra*, SIAM. Philadelphia, 1997.

E.E. Tyrtyshnikov, *A brief introduction to Numerical Analysis*, Birkhäuser, 1997.

Why Krylov subspaces are so much used?

Alexei Nikolaevich Krylov

1863-1945, Maritime Engineer

- 300 papers and books on: shipbuilding, magnetism, artillery, math, astronomy, geodesy
- 1890: theory of oscillating motions of the ship
- 1904: he built the first machine in Russia for integrating ODEs
- 1931: Krylov subspace methods

Properties of the Krylov subspaces

$\mathcal{K}^m(A, \mathbf{v}) = span\{\mathbf{v}, A\mathbf{v}, \cdots, A^{m-1}\mathbf{v}\}$
The dimension of $\mathcal{K}^m$ increases with each iteration.

- Theorem [Cayley-Hamilton]: $d \leq n$
- $\mathcal{K}^d$ is invariant under A, thus, $\mathcal{K}^m = \mathcal{K}^d$ for $m > d$, thus,

$$dim(\mathcal{K}^m) = \min(m, d)$$

Presentation, based on the paper

The Idea Behind Krylov Methods

Ilse C. F. Ipsen and Carl D. Meyer
The American Mathematical Monthly, Vol. 105, No. 10, Dec., 1998

Summary:

Why Krylov methods make sense, and why it is natural to represent a solution to a linear system as a member of a Krylov space?

The authors show that the solution to a nonsingular linear system $Ax = b$ lies in a Krylov space whose dimension is the degree of the minimal polynomial of A.

Therefore, if the minimal polynomial of A has low degree then the space in which a Krylov method searches for the solution can be small. In this case a Krylov method has the opportunity to converge fast.

When the matrix is singular, however, Krylov methods can fail.

Even if the linear system does have a solution, it may not lie in a Krylov space. In this case one describes a class of right-hand sides for which a solution lies in a Krylov space. As it happens, there is only a single solution that lies in a Krylov space, and it can be obtained from the so-called Drazin inverse.

Assume that A is nonsingular.

Idea: express A^{-1} in terms of powers of A.

The minimal polynomial of A, $q_d(t)$ of degree d, is the unique monic polynomial of minimal degree, for which

$$q(A) = 0.$$

It has the form

$$q_d(t) = \prod_{j=1}^{d} (t - \lambda_j)^{m_j},$$

where
- $\lambda_1, \cdots, \lambda_d$ are distinct eigenvalues of A,
- $m_1, \cdots, m_d$ are the corresponding indeces of λ_j (the sizes of the largest Jordan block, associated with λ_j).

Idea: express A^{-1} in terms of powers of A.

$$q_d(t) = \prod_{j=1}^{d} (t - \lambda_j)^{m_j} = \sum_{s=0}^{m} \alpha_s t^s, \qquad (1)$$

where $m = \sum_{j=1}^{d} m_j$.

Example: $A = \begin{bmatrix} 3 & 1 & & \\ & 3 & & \\ & & 4 & \\ & & & 4 \end{bmatrix}$. Then we have $\lambda_1 = 3,\ m_1 = 2,$ $\lambda_2 = 4,\ m_2 = 1.$

Note that, since we have assumed that A is nonsingular, in (1), the coefficient $\alpha_0 = \prod_{j=1}^{d} (-\lambda_j)^{m_j} \neq 0$.

Idea: express A^{-1} in terms of powers of A.

$$q(A) = \alpha_0 I_n + \alpha_1 A + \alpha_2 A^2 + \cdots + \alpha_m A^m = 0,\ \alpha_0 \neq 0$$

Then $A^{-1} q(A) = 0$, thus,

$$A^{-1} = \frac{1}{\alpha_0} \sum_{j=0}^{m-1} \alpha_{j+1} A^j$$

However, $\mathbf{x} = A^{-1}\mathbf{b}$!

If the minimal polynomial of A ($A^{-1}\exists$) has degree m, then $\mathbf{x} = A^{-1}\mathbf{b} \in \mathcal{K}^m(A, \mathbf{b})$.
($\mathbf{b} = \mathbf{r}^{(0)}$ for $\mathbf{x}^{(0)} = \mathbf{0}$)

Idea: express A^{-1} in terms of powers of A.

Remarks:

- If d is small, then the convergence is fast.
- We also see that the eigenvalues of A, not its singular values, are important, because the dimension of the solution space is determined by the degree of the minimal polynomial.

What happens if A^{-1} does not exist?

Suppose that A is singular. One can show that even if a solution exists, it may not lie in the Krylov space $\mathcal{K}^m(A, \mathbf{b})$.
Example: Consider a consistent linear system $N\mathbf{x} = \mathbf{c}$, where N is a nilpotent matrix, i.e., there exists some integer ℓ, such that $N^\ell = 0$ but $N^{\ell-1} \neq 0$.
Suppose that the solution $\mathbf{x}$ is a linear combination of Krylov vectors, i.e.,

$$\mathbf{x} = \beta_0\mathbf{c} + \beta_1 N\mathbf{c} + \beta_2 N^2\mathbf{c} + \cdots + \beta_{\ell-1}N^{\ell-1}\mathbf{c}$$

Then, $\mathbf{c} = N\mathbf{x} = \beta_0 N\mathbf{c} + \beta_1 N^2\mathbf{c} + \cdots + \beta_{\ell-2}N^{\ell-1}\mathbf{c}$ and
$(I - \beta_0 N - \beta_1 N^2 - \cdots - \beta_{\ell-2}N^{\ell-1})\mathbf{c} = 0$.
The matrix $Q = I - \beta_0 N - \beta_1 N^2 - \cdots - \beta_{\ell-2}N^{\ell-1}$ is nonsingular, because of the following reasons. The eigenvalues of any nilpotent matrix are all equal to zero, thus, the eigenvalues of Q are all equal to 1. Therefore, $\mathbf{c}$ must be zero.

Moral: the solution of a system with a nilpotent matrix and a nonzero right hand side cannot lie in the Krylov subspace, generated by the matrix and the rhs.

What happens if A^{-1} does not exist?

Apply the following trick: Decompose the space $C^n = \mathcal{R}(A^\ell) \oplus \mathcal{N}(A^\ell)$, where ℓ is the index of the zero eigenvalue of $A \in C^{n\times n}$ and $\mathcal{R}(\cdot)$ and $\mathcal{N}(\cdot)$ denote range and nullspace. Then

$$A = \begin{bmatrix} R & 0 \\ 0 & N \end{bmatrix},$$

where R is nonsingular and N is nilpotent of index ℓ.
Suppose now that $A\mathbf{x} = \mathbf{b}$ has a Krylov solution

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \sum_{j=1}^{d} \alpha_j A\mathbf{b} = \sum_{j=0}^{d} \alpha_j \begin{bmatrix} R^j & 0 \\ 0 & N^j \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix}$$

thus

$$\mathbf{x}_1 = \sum_{j=0}^{d} \alpha_j R^j \mathbf{b}_1 \quad \text{and} \quad \mathbf{x}_2 = \sum_{j=0}^{d} \alpha_j N^j \mathbf{b}_2.$$

What happens if A^{-1} does not exist?

From $A\mathbf{x} = \mathbf{b}$ we have that $N\mathbf{x}_2 = \mathbf{b}_2$, so $\sum\limits_{j=0}^{d-1} \alpha_j N^{j+1}\mathbf{b}_2 = \mathbf{b}_2$ and

$$(I - \sum_{j=0}^{d-1} \alpha_j N^{j+1})\mathbf{b}_2 = \mathbf{0}$$

The matrix in parentheses is nonsingular, thus $\mathbf{b}_2 = \mathbf{0}$.
In other words, the existence of a Krylov solution requires that $\mathbf{b} \in \mathcal{R}(A^\ell)$.
The converse statement is also true.

Theorem

A square linear system $A\mathbf{x} = \mathbf{b}$ has a Krylov solution if and only if $\mathbf{b} \in \mathcal{R}(A^\ell)$, where ℓ is the index of the zero eigenvalue of A.

The index of an eigenvalue λ for J is defined as the dimension of the largest Jordan block associated to that eigenvalue.