On Solving Ill-Conditioned Linear Systems[1]

Craig C. Douglas[1,2], Long Lee[1], and Man-Chung Yeung[1]

University of Wyoming, Laramie, WY, USA
[1] Department of Mathematics and [2] School of Energy Resources

ICCS 2016, San Diego, June 7, 2016

[1]This research was supported in part by the National Science Foundation through grants ACI-1440610, ACI-1541392, and DMS-1413273.

Problem

- Consider the solution of

$$Ax = b, \qquad A \in \mathbb{C}^{n\times n}, b \in \mathbb{C}^n \tag{1}$$

  by a Krylov subspace method, where A has eigenvalues close to the origin.
- Examples.
  1. *bcsstm27* from a mass matrix buckling problem, 1224×1224 real symmetric and indefinite, $56,126$ nonzero entries.
  2. *mahindas* from an economic problem, 1258×1258 real unsymmetric, $7,682$ nonzero entries.

Krylov Subspace Methods

- Given an initial guess x_0, the exact solution of (1) is

$$x^* \in x_0 + span\{r_0,\ \ Ar_0,\ \ \cdots, A^{n-1}r_0\},$$

where $r_0 = b - Ax_0$. A space of the form

$$span\{v_0,\ \ Av_0,\ \ \cdots, A^{k-1}v_0\}$$

is called a Krylov subspace.

- At the k^{th} iteration, a Krylov subspace method searches an approximation x_k from

$$x_k \in x_0 + span\{r_0,\ \ Ar_0,\ \ \cdots, A^{k-1}r_0\}.$$

Hence, at the n^{th} iteration, a Krylov method must obtain x^*.

Conjugate Gradient (CG) Method

The CG method [Hestenes & Stiefel (1952) and Lanczos (1952)] was the first Krylov subspace method and works for symmetric, positive definite systems A.

Conjugate-Gradient method

1. Choose an initial guess x_0.
2. Compute $r_0 = b - Ax_0$ and set $p_0 = r_0$.
3. For $k = 0, 1, \cdots$, until convergence:
4. $\quad \alpha_k = r_k^H r_k / p_k^H A p_k$;
5. $\quad x_{k+1} = x_k + \alpha_k p_k$;
6. $\quad r_{k+1} = r_k - \alpha_k A p_k$;
7. $\quad \beta_k = r_{k+1}^H r_{k+1} / p_k A p_k$;
8. $\quad p_{k+1} = r_{k+1} + \beta_k p_k$;
9. End

Convergence properties:

1. At the k^{th} iteration, select x_k such that

$$x_k \in x_0 + span\{r_0,\ \ Ar_0,\ \ \cdots, A^{k-1} r_0\}.$$

and

$$r_k = b - Ax_k \perp span\{r_0,\ \ Ar_0,\ \ \cdots, A^{k-1} r_0\}.$$

2.

$$\|x_k - x^*\|_A \le \sqrt{2} \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^k \|x_0 - x^*\|_A,$$

where $\kappa = \lambda_{max}/\lambda_{min}$, the so-called condition number of A.

3. If A has l distinct eigenvalues, CG converges in l iterations.

Generalized Minimal Residual (GMRES) Method

Proposed in [Saad and Schultz (1986)] as a solver for a general system A.

GMRES method

1. Compute $r_0 = b - Ax_0$, $\beta = \|r_0\|_2$, and $v_1 = r_0/\beta$.
2. For $j = 1, 2, \cdots, m$, Do
3. Compute $w_j = Av_j$
4. For $i = 1, \ldots, j$, Do
5. $h_{ij} = (w_j, v_i)$
6. $w_j = w_j - h_{ij} v_i$
7. EndDo
8. $h_{j+1,j} = \|w_j\|_2$. If $h_{j+1,j} = 0$ set $m = j$ and go 11.
9. $v_{j+1} = w_j / h_{j+1,j}$
10. EndDo
11. Define the $(m+1) \times m$ Hessenberg matrix $\bar{H}_m = \{h_{ij}\}$.
12. Compute the minimizer y_m of $\|\beta e_1 - \bar{H}_m y\|_2$, and $x_m = x_0 + V_m y_m$.

Convergence properties

1. At the k^{th} iteration, select x_k such that

$$x_k \in x_0 + span\{r_0, \ \ Ar_0, \ \ \cdots, A^{k-1} r_0\} \equiv x_0 + \mathcal{K}_k(A, r_0)$$

   and

$$\|r_k\|_2 = \min_{x \in x_0 + \mathcal{K}_k(A, r_0)} \|b - Ax\|_2$$

2. (Saad) Suppose that A has a diagonal decomposition

$$A = V \Lambda V^{-1}.$$

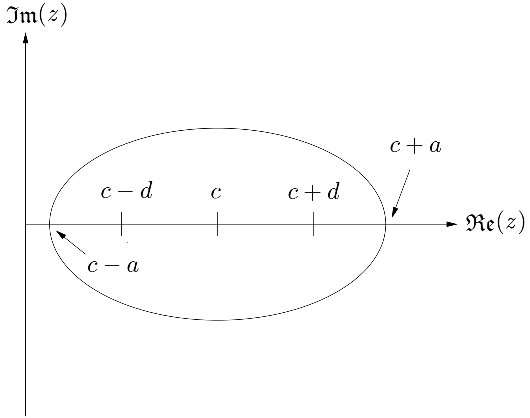
   Let $E(c, d, a)$ be an ellipse in the complex plane (see Fig. 1, next slide). If all of the eigenvalues of A are located inside $E(c, d, a)$ and excludes the origin of the complex plane, then

$$\|r_k\|_2 \leq \kappa_2(V) \delta^k \|r_0\|_2 \qquad (2)$$

   where $\kappa_2(V) = \|V\|_2 \|V^{-1}\|_2$ and $\delta = \dfrac{a + \sqrt{a^2 - d^2}}{c + \sqrt{c^2 - d^2}}$.

Fig. 1: A schematic ellipse in the complex plane

It has center c, focal distance d, and semi-major axis a.



Convergence properties

- The upper bound (2) contains two factors:
  - (i) the condition number $\kappa_2(V)$ of the eigenvector matrix V.
  - (ii) the scalar δ determined by the distribution of eigenvalues of A.
- For example, if A is nearly normal and has a spectrum clustering around 1, then $\kappa_2(V) \approx 1$ and $\delta < 1$. In this case, $\|r_k\|_2$ decays exponentially in a rate of power δ^k.

Deflated GMRES Method

- The ellipse $E(c, d, a)$ in the upper bound (2) is required to include all eigenvalues of A. Outlying eigenvalues may keep the ellipse large, implying a large δ. To reduce δ, we therefore want to remove outlying eigenvalues from $\sigma(A)$.
- To remove some eigenvalues, say, $\lambda_1, \ldots, \lambda_m$, from $\sigma(A)$, define

$$P = I - AZ(Z^H AZ)^{-1} Z^H,$$

where $Z = [v_1, \ldots, v_m]$ is the matrix of the corresponding eigenvectors. Then

$$\sigma(PA) = \{0, \ldots, 0, \lambda_{m+1}, \ldots, \lambda_n\}$$

Deflated GMRES method

1. Solve $PAx = Pb$ by GMRES to obtain a solution $x^{\#}$;
2. Compute $x_1^* = Z(Z^H AZ)^{-1} Z^H b$;
3. Compute $x_2^* = \widetilde{P} x^{\#}$ where $\widetilde{P} = I - Z(Z^H AZ)^{-1} Z^H A$;
4. Compute $x^* = x_1^* + x_2^*$ which is the solution of $Ax = b$.

Computation of Z

- Let Γ be a given simple closed curve in the complex plane that encloses the eigenvalues $\lambda_1, \ldots, \lambda_m$. Then, by the residue theorem in complex analysis,

$$P_\Gamma = \frac{1}{2\pi\sqrt{-1}} \oint_\Gamma (zI - A)^{-1} dz$$

is a projector onto the eigenspace $span\{v_1, \ldots, v_m\}$ (details in Y. Saad's book).

Computation of Z

- Pick a random $Y \in \mathbb{C}^{n \times m}$. Then P_Γ projects Y onto the eigenspace $span\{v_1, \ldots, v_m\}$, namely, each column of the $n \times m$ matrix

$$P_\Gamma Y = \frac{1}{2\pi\sqrt{-1}} \oint_\Gamma (zI - A)^{-1} Y dz$$

is a vector in $span\{v_1, \ldots, v_m\}$. In fact, $P_\Gamma Y$ is almost surely a basis matrix. Set

$$Z = P_\Gamma Y = \frac{1}{2\pi\sqrt{-1}} \oint_\Gamma (zI - A)^{-1} Y dz \tag{3}$$

Computation of Z

- The integral in (3) can be computed by any quadrature rule. In our experiments, we used the Legendre-Gauss quadrature rule as follows:
  (i) Suppose Γ is a circle with center c and radius r. Set
  $$z = c + re^{i\pi\theta}$$
  where $i = \sqrt{-1}$. Then (3) becomes
  $$Z = \frac{r}{2}\int_{-1}^{1} e^{i\pi\theta}((c + re^{i\pi\theta})I - A)^{-1} Y d\theta$$
  (ii) Apply Legendre-Gauss quadrature rule to the above integral
  $$Z \approx \frac{r}{2}\sum_{k=1}^{q} \omega_k e^{i\pi\theta_k}((c + re^{i\pi\theta_k})I - A)^{-1} Y. \qquad (4)$$
  where ω_k and θ_k are weights and notes.

Numerical Experiments

- We now compare the solution of

$$Ax = b, \tag{5}$$

  and the solution of the deflated system

$$PAx = Pb \tag{6}$$

  where $P = I - AZ(Z^H AZ)^{-1} Z^H$.
- We performed the following two experiments:

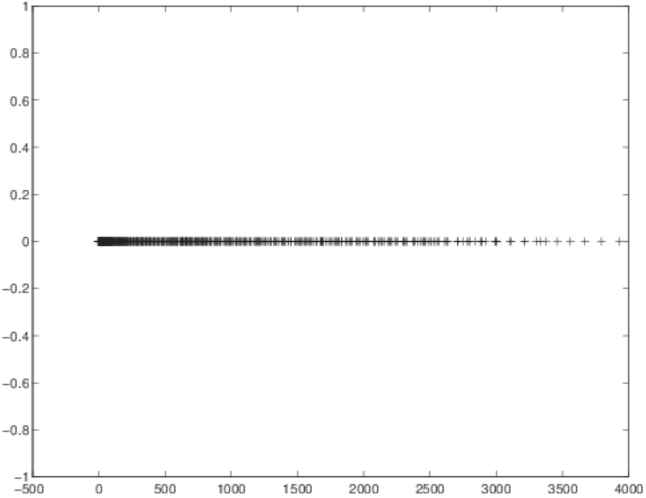
Exp#1: Solve (5) without deflation.

Exp#2: Compute Z through (4). Then perform QR factorization on the computed Z: $Z = QR$, where $Q \in \mathbb{C}^{n \times m}$ and $R \in \mathbb{C}^{m \times m}$. Then set $Z = Q$. Then solve (6).

Numerical Experiments

- Two test data downloaded from The University of Florida Sparse Matrix Collection:
  1. *bcsstm27* from a mass matrix buckling problem, size 1224×1224, real symmetric and indefinite, and has $56,126$ nonzero entries. As the right-hand side in (5), we set $b = A\mathbf{1}$, where $\mathbf{1} = [1, 1, \ldots, 1]^T$. No preconditioner was used in the solution of (5). A spectral plot for *bcsstm27* is in Fig. 2.

Fig. 2: Eigenvalue distribution of *bcsstm27*



2. *mahindas* from an economics problem, size 1258×1258, real unsymmetric, and has $7,682$ nonzero entries. Again, we set $b = A\mathbf{1}$ in (5). A ILU(0) preconditioner generated by the Matlab function $[L, U, P] = luinc(A'0')$ was used, namely, instead of solving (5), we solve

$$\widetilde{A}y = \widetilde{b},$$

where $\widetilde{A} = L^{-1}PAU^{-1}$, $\widetilde{b} = L^{-1}Pb$, and $y = Ux$. Spectral plots for *mahindas* and ILU(0)-preconditioned *mahindas* are in Figs. 3 and 4.

Fig. 3: Eigenvalue distribution of *mahindas*

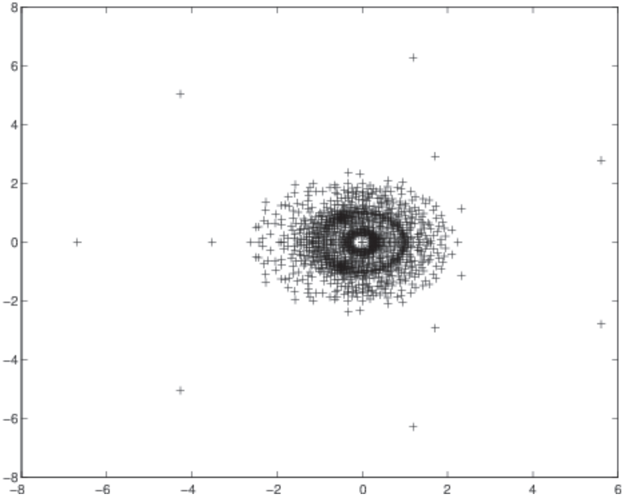
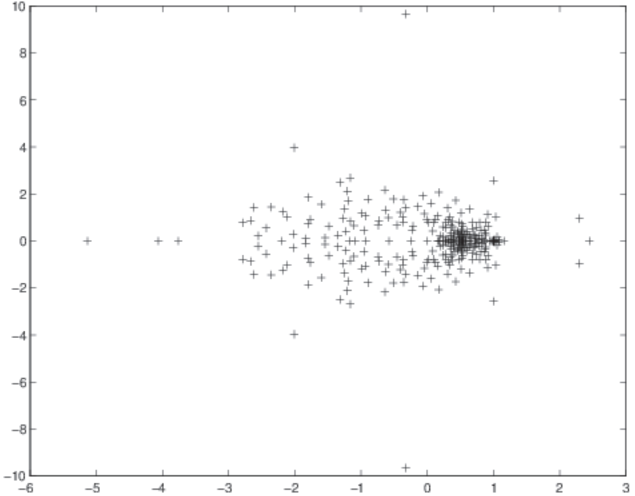


Fig. 4: Eigenvalues of ILU(0)-preconditioned *mahindas*



Numerical Experiments

- Some implementation details:
  1. All the computations were done in Matlab Version 7.1 on a Windows 7 machine with a Pentium 4 processor.
  2. Due to that full GMRES is too expensive in terms of time and storage, rather than use GMRES, we employed BiCG as the Krylov solver. The initial guess for BiCG was $x = 0$, and the stopping criteria were $\|b - Ax\|_2/\|b\|_2 < 10^{-7}$ for (5) and $\|Pb - PAx\|_2/\|Pb\|_2 < 10^{-7}$ for (6).
  3. We computed Z through (4). In (4), there are mq linear systems
  $$((c + re^{i\pi\theta_k})I - A)x = y_j$$
  to solve. We solved each of them by BiCG with stopping tolerance $tol = 10^{-10}$ and maximum number of iteration $maxit = n$.

Table 1

A comparison of solving (5) and (6) by BiCG. For *mahindas*, a ILU(0) preconditioner was used. Γ is a circle with center c and radius r. The q in (4) is $q = 2^7$. We remark that an efficient stochastic estimation method of the exact number of eigenvalues inside a given Γ has been developed by Futamura, Tadano, and Sakurai.

Matrix	$\Gamma : (c, r)$	#eig inside Γ	m	
bcsstm27	$(0, 5)$	363	400	
mahindas	$(-1, 1)$	31	50	
Matrix	Exp#1: #iter	Exp#1: Err	Exp#2: #iter	Exp#2: Err
bcsstm27	1224000	4.0×10^{-6}	763	8.9×10^{-8}
mahindas	1258000	1.3	3937	5.0×10^{-8}

Conclusion and Future Work

- We incorporate a deflation projector P into Krylov subspace methods to enhance the stability and accelerate the convergence for solving ill-conditioned systems.
- To our best knowledge, the construction of the deflation matrix Z in the literature are problem dependent. The method proposed here of constructing Z is problem independent.
- The most expensive part in the deflation method proposed here is the solution of mq linear systems in (4). We will implement robust and efficient parallel multigrid methods for solving these linear systems.