Numerical Linear Algebra

Sheet 4 — MT20

Krylov methods and randomised algorithms

This sheet is due 9am two weekdays before the class (Thursday if class is on Monday).

1. If $A \in \mathbb{R}^{n\times n}$ is nonsingular, show that GMRES breaks down at the ℓ^{th} iteration (i.e. $h_{\ell+1,\ell} = 0$) if and only if $x_\ell = x$ (i.e. the solution of the linear system has been found).

2. (a) Let $Q_k R_k$ be a QR factorization of $\widehat{H}_k$ where $Q_k = J_1 J_2 \dots J_k$ with J_j being a Givens rotation matrix for each j. If $\widehat{H}_{k+1}$ is computed from $\widehat{H}_k$ by appending the one further column computed by the next step of the Arnoldi algorithm, show that only one further Givens rotation J_{k+1} gives the QR factorization of $\widehat{H}_{k+1}$.

   (b) If $s = \sin\theta$ in the Givens rotation in J_{k+1}, show that

   $$\|r_k\|_2 = |s|\|r_{k-1}\|_2.$$

   Hence for the sequence of succesive residuals $r_k, k = 0, 1, 2, \dots$ computed by the GMRES method, $\{\|r_k\|_2, k = 0, 1, 2, \dots\}$ must reduce monotonically. Are there any circumstances in which the convergence is not strictly monotonic?

3. Show how GMRES will converge on the linear system $Ax = b$ with $x_0 = 0$ when

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \qquad b = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

   Hand calculation (it is simple in this example to work out what the residual vectors must be!) is best here if you want to learn something!

Mathematical Institute, University of Oxford
Yuji Nakatsukasa: nakatsukasa@maths.ox.ac.uk

4. For $A \succ 0 \in \mathbb{R}^{n\times n}, b \in \mathbb{R}^n$ and a chosen x_0, let $r_0 = b - Ax_0$ and $p_0 = r_0$ and for $k = 0, 1, \dots$

$$\begin{aligned} & \alpha_k = \langle p_k, r_k\rangle / \langle p_k, Ap_k\rangle \\ (1) \quad & x_{k+1} = x_k + \alpha_k p_k \\ (2) \quad & r_{k+1} = b - Ax_{k+1} \\ & \beta_k = -\langle p_k, Ar_{k+1}\rangle / \langle p_k, Ap_k\rangle \\ (3) \quad & p_{k+1} = r_{k+1} + \beta_k p_k \end{aligned}$$

where $\langle\cdot,\cdot\rangle$ is the standard inner product $\langle x, y\rangle = x^T y$. Show that (2) and (1) imply

$$r_{k+1} = r_k - \alpha_k A p_k.$$

Prove that the definition of α_k implies $\langle r_{k+1}, p_j\rangle = 0$ for $j = k$ and that the definition of β_k implies $\langle p_{k+1}, Ap_j\rangle = 0$ for $j = k$. Prove also that $\langle r_{k+1}, r_j\rangle = 0$ for $j = k$. Now by employing induction in k for $k = 1, 2, \dots$, prove these three assertions for $j = 1, 2, \dots, k-1$. (The inductive assumption will be that

$$\langle r_k, p_j\rangle = 0, \quad \langle r_k, r_j\rangle = 0, \quad \langle p_k, Ap_j\rangle = 0, \quad j = 0, 1, \dots, k-1$$

and you may wish to tackle the assertions in this order.)

Note that these show x_k satisfies the CG condition $Q^T(Ax_k - b)$ where $\text{span}(Q) = \mathcal{K}_k(A, b)$.

5. Let $S \in \mathbb{R}^{n\times n}$ be a symmetric matrix with $\|S\|_2 \le 0.5$, and $A = I + S$.
   (a) Show that A is symmetric positive definite.
   (b) Consider a linear system $Ax = b$ for $b \in \mathbb{R}^n$ solved by CG. Give an upper bound for the number of CG iterations k required to get $\epsilon = 10^{-15}$ accuracy in the A-norm of the error $\|x - x^{(k)}\|_A / \|x\|_A \le 10^{-15}$.
   (c) Use MATLAB or Python to verify the above claim in experiments.

Yuji Nakatsukasa: nakatsukasa@maths.ox.ac.uk

6. (Randomised SVD.) Let $A \in \mathbb{R}^{m\times n}$, and $X \in \mathbb{R}^{n\times r}(r < n)$. Let $AX = QR$ be the thin QR factorisation, and let $\hat{A} = QQ^TA$ be a rank-r approximant to A.

   (a) Show that $A - \hat{A} = (I_m - QQ^T)A = (I_m - QQ^T)A(I_n - XM^T)$ for any $M \in \mathbb{R}^{n\times r}$.

   (b) Show that choosing $M^T = (V^TX)^\dagger V^T$ where $V \in \mathbb{R}^{n\times r}$ is orthonormal, $XM^T = \mathcal{P}_{X,V}$ becomes a projector $\mathcal{P}_{X,V}^2 = \mathcal{P}_{X,V}$. Show further that if V^TX is nonsingular, then $A(I - \mathcal{P}_{X,V}) = A(I - VV^T)(I - \mathcal{P}_{X,V})$.

   (Note: The pseudoinverse $W^\dagger$ is defined via the (economical) SVD $W = U_W\Sigma_W V_W^T$ (where $\Sigma_W \succ 0$) by $W^\dagger = V_W\Sigma_W^{-1}U_W^T$.)

   (c) By choosing V appropriately, show that $\|A - \hat{A}\|_2 \le \|\Sigma_2\|_2\|I - \mathcal{P}_{X,V}\|_2$ where $\Sigma_2 = \mathrm{diag}(\sigma_{r+1}(A), \sigma_{r+2}(A), \ldots, \sigma_n(A))$.

   (Note: we further have $\|I - \mathcal{P}_{X,V}\|_2 = \|\mathcal{P}_{X,V}\|_2$, which holds for any projector that is not 0 or I. $\|\mathcal{P}_{X,V}\|_2 =$ "$O(1)$" can be shown with high probability when X is a Gaussian random matrix with i.i.d. entries $X_{ij} \sim N(0, 1)$. The above bound then means $\hat{A}$ is a near-optimal rank-r approximation).

7. (Randomised least-squares.) Consider a least-squares problem $\min_x \|Ax - b\|_2$, where $A \in \mathbb{R}^{m\times n}(m > n)$ has full column rank.

   (a) Let $A = QR$ be the thin QR factorisation. Show that $\kappa_2(AR^{-1}) = 1$.

   (b) Let $X \in \mathbb{R}^{m\times\tilde{n}}$ where $\tilde{n} \ge n$. Show that if X^TQ is well-conditioned $\kappa_2(X^TQ) = O(1)$, then with the QR factorisation $X^TA = Q_1R_1$, AR_1^{-1} is well-conditioned.

   (c) Hence show that once such R_1 becomes available, the least-squares problem $\min_x \|Ax - b\|_2$ can be solved efficiently in $O(mn)$ operations (or $O(mn \log \frac{1}{\epsilon})$ operations for ϵ accuracy).

8. Explore question 6 with experiments: take A to be a matrix with singualr values geometrically distributed between 1 and 10^{-20}. Take $X \in \mathbb{R}^{n\times r}(r < n)$ to be a Gaussian random matrix (`X=randn(m,r)` in MATLAB), and form $\hat{A} = QQ^TA$, where $AX = QR$. Then examine the error $\|A - \hat{A}\|_2$ as you vary r and do an r-vs.-error plot, comparing the error with the optimal value $\sigma_{r+1}(A)$ (by the truncated SVD). Verify that error is a modest constant multiple of optimal.

Yuji Nakatsukasa: nakatsukasa@maths.ox.ac.uk

9. (Optional) Use matlab (`[x,flag,relres,iter,resvec]=gmres(A,b,[],1.e-6,size(A,1))`) with suitably chosen matrices A and b as below to investigate the behaviour of GMRES.

   Note in the form above matlab will use unrestarted GMRES, `flag=0` will indicate successful convergence (the relative residual norm - `relres` - less than 10^{-6} in less than dimension(A) `=size(A,1)` iterations), `iter` is the number of restarts (should be 1 with no restarting) and iterations taken and `resvec` is the vector of residual norms at each iteration (hence `semilogy(resvec)` will plot the convergence curve). See `help gmres` if you want to read more or change any of the defaults.

   (i) `A=randn(n); b=ones(n,1);` for `n=7,47,...` as you choose (and have patience for! (note `ctrl C` will interrupt a computation). These are dense matrices!

   (ii) `A=sprandn(100,100,0.1); b=ones(100,1);`. This is a sparse 100×100 matrix with approximately 10 non-zero entries per row (ie. `0.1` of the 10,000 entries non-zero).

   (iii) `A=sprandn(100,100,0.1) +2*eye(100,100); b=ones(100,1);`

   (iv) `A=sprandn(100,100,0.1) +4*eye(100,100); b=ones(100,1);`

   (v) a diagonalisable matrix that has few distinct eigenvalues
   eg. `X=randn(9,9); A=X*diag([1,1,-4,3,3,-4,-4,-4,3])/X`
   (note `/X` is an more efficient way of computing `*inv(X)` and that it is possible that an X generated with random entries is singular, but is rarely so!)

   (vi) any matrix

Mathematical Institute, University of Oxford
Yuji Nakatsukasa: nakatsukasa@maths.ox.ac.uk