MATRIX COMPUTATIONS: HOMEWORK 2

Andre Theo, Dejean Maxime, Doat Damien, Nouidei Safiya

October 2024

1 Exercise A: Krylov subspaces

(A1)

As assumed in the assignement, $\mathcal{K}_{r+1}(A,b) \subseteq \mathcal{K}_r(A,b)$. If we can prove that we also have $\mathcal{K}_{r+1}(A,b) \supseteq \mathcal{K}_r(A,b)$, it means that we proved that $\mathcal{K}_{r+1}(A,b) = \mathcal{K}_r(A,b)$. Since, by definition of the krylov subspaces as $span\{b, Ab, ..., A^r b\}$, all elements of $\mathcal{K}_r(A,b)$ are in $\mathcal{K}_{r+1}(A,b)$, we know that, indeed, $\mathcal{K}_{r+1}(A,b) \supseteq \mathcal{K}_r(A,b)$. And thus, we have proved the equality between the two spaces. We then need to show that, for $s \geq r$, $\mathcal{K}_s(A,b) = \mathcal{K}_r(A,b)$. It can be done with a proof by induction. Indeed, since the eqaulity holds for all r, it means that $\mathcal{K}_{r+2}(A,b) = \mathcal{K}_{r+1}(A,b) = \mathcal{K}_r(A,b)$ etc. This means that for all i such that $s - r \leq i$, $\mathcal{K}_{r+i}(A,b) = \mathcal{K}_{r+(i-1)}(A,b) = ... = \mathcal{K}_r(A,b)$. This shows that $\mathcal{K}_s(A,b) = \mathcal{K}_r(A,b)$.

(A2)

Let's prove that $dim(\mathcal{K}_r(A,b)) = r$ by induction. Since $dim(\mathcal{K}_n(A,b)) = s$ and $r \leq s$, it means that, at each iteration i so that $i \leq r$, the dimension of $dim(\mathcal{K}_i(A,b)) = i$. Indeed, no combination will already be in the span since r is smaller than s. We thus have:

- $i = 1$: $\mathcal{K}_i(A,b) = \{b\}$ and $dim(\mathcal{K}_i(A,b)) = 1$
- $i = 2$: $\mathcal{K}_i(A,b) = \{b, Ab\}$ and $dim(\mathcal{K}_i(A,b)) = 2$
- ...
- $i = r$: $\mathcal{K}_i(A,b) = \mathcal{K}_r(A,b)$ and $dim(\mathcal{K}_i(A,b)) = r$

2 Exercise B: Arnoldi's iteration

(B1)

We want to show that R_s has non-zero elements on its diagonal. Fisrt, since $K_s(A, b) = [b, Ab, \dots, A^{s-1}b]$, it consists of s linearly independent columns. Thus, $K_s(A, b)$ has full column rank, so $rank(K_s(A, b)) = s$. In any QR decomposition $M = QR$ where $M \in R^{m\times s}$ and M has full column rank, R is invertible. Therefore, R is nonsingular, meaning all its diagonal elements are non-zero. Finally, in our case, since $K_s(A, b)$ has full column rank, R_s is invertible and hence has non-zero elements on its diagonal. Thus, we conclude that R_s has non-zero elements on its diagonal otherwise $rank(K_s(A, b)) \neq s$.

(B2)

The goal is to show that for each $1 \leq r \leq s - 1$,

$$AQ_s[1 : r]R_s[1 : r, 1 : r] = Q_s[1 : r']R_s[1 : r', 2 : r'],$$

where $r' = r + 1$. Let's do it step by step :

1. **Krylov Subspace Properties:**
   The Krylov subspace $K_r(A, b)$ is defined as:

   $$K_r(A, b) = span\{b, Ab, A^2b, \dots, A^{r-1}b\}.$$

   By construction, applying A to $K_r(A, b)$ generates $K_{r+1}(A, b)$:

   $$K_{r+1}(A, b) = span\{b, Ab, A^2b, \dots, A^r b\}.$$

   Therefore, $K_{r+1}(A, b)$ extends $K_r(A, b)$ by including the new vector $A^r b$, making it a shifted version of $K_r(A, b)$ with an additional dimension.

2. **QR Decomposition of $K_r(A, b)$:**
   Given the QR decomposition of $K_s(A, b)$, we have:

   $$K_s(A, b) = Q_s R_s,$$

   where $Q_s \in R^{n\times s}$ is an orthonormal basis for $K_s(A, b)$ and $R_s \in R^{s\times s}$ is upper-triangular. The submatrices $Q_s[1 : r]$ and $R_s[1 : r, 1 : r]$ provide a basis and the associated coefficients for the subspace $K_r(A, b)$.

3. **Applying A to $Q_s[1 : r]R_s[1 : r, 1 : r]$:**
   When we apply A to $Q_s[1 : r]R_s[1 : r, 1 : r]$, we obtain a matrix that spans the extended subspace $K_{r+1}(A, b)$. This new subspace can be represented by:

   $$AQ_s[1 : r]R_s[1 : r, 1 : r] = Q_s[1 : r']R_s[1 : r', 2 : r'],$$

   where $r' = r + 1$.

(B3)

(B4)

We want to show that $span\{q_1, ..., q_r\} \subseteq \mathcal{K}_r(A, b)$. This is equivalent, by definition, to proving that $span\{q_1, ..., q_r\} \subseteq span\{b, Ab, ..., A^{r-1}b\}$. Let's proceed by induction:

- $q_1 = \frac{b}{||b||}$: this is indeed included in $\mathcal{K}_r(A, b)$ as it is just the normalisation of b and it can be express only in terms of b.

- by definition of υ and ω in algorithm 1, $q_2 = \frac{\omega}{||\omega||} = \frac{Aq_1 - (q_1^T Aq_1)q_1}{||Aq_1 - (q_1^T Aq_1)q_1||}$. q_2 is also included in the krylov space because it is the normalisation of Ab with regards to $q_1 = \frac{b}{||b||}$.

- we use the same arguments for each q_i to prove that it belongs to the subsapce since it can be expressed with only A and q_{i-1}.

(B5)

From (B4), we know that $span\{q_1, ..., q_r\} \subseteq \mathcal{K}_r(A, b)$. Since $dim(span\{q_1, ..., q_r\}) = r$ (as it is formed of orthoromals vectors) and $dim(\mathcal{K}_r(A, b)) = r$ (as we proved in (A2)), it means that the two spans have the same dimension. Thus, as one is include or equal to the other, the two spans are actually equal.

(B6)

Arnoldi's iteration constructs an orthonormal basis $\{q_1, q_2, \ldots, q_s\}$ for $K_s(A, b)$. During the iteration, each new vector q_{r+1} is obtained as a linear combination of Aq_r, with orthogonalization against previous q_i's to ensure orthonormality. The algorithm halts when the next vector Aq_s lies within the span of the existing vectors $\{q_1, q_2, \ldots, q_s\}$, indicating that no new dimension is added. Since here s is the dimension of $dim(span\{q_1, ..., q_s\})$, we know that any krylov subsapces with $n > s$ will still be of dimesion s because the vectors that we want to add already lie in the span.

(B7)

3 Exercise C: GMRES for linear system solution approximation

(C1)

Let's set $x = Qy$ with $y \in R^r$. We now have

$$AQy = QHy + \beta q e_r^T y$$

$$||Ax - b|| = ||QHy + \beta q e_r^T y - b||$$

We thus have :

$$\min_{x \in \mathcal{K}_r(A,b)} ||Ax-b|| = \min_{y \in R^r} ||[Q\ q][Hy\ \beta e_r^T y]^T - b|| = \min_{y \in R^r} ||[Q\ q]([H\ \beta e_r^T]^T y - ||b|| e_{r+1,1})||$$

As $[Q\ q]$ is an isometry, the norm is not impacted by it and it gives us exactly what we wanted to find, with $\tilde{H} = [H\ \beta e_r^T]^T$. And, as $x = Qy$, if $y*$ is a minimiser of (4), then $x* = Qy*$ is a minimiser of (3).

(C2)

To solve the minimization problem in Equation (3),

$$\min_{x \in K_r(A,b)} \|Ax - b\|,$$

instead of directly solving the linear system $Ax = b$, we use the GMRES (Generalized Minimal Residual) method, which approximates the solution within a Krylov subspace $K_r(A, b)$.

Given a large matrix $A \in R^{n \times n}$ and a vector $b \in R^n$, solving $Ax = b$ directly can be computationally intensive, typically requiring $O(n^3)$ operations for exact solutions (for example, Gaussian elimination). In contrast, GMRES aims to solve the minimization problem within a subspace of dimension r, which is often much smaller than n, making it less costly.

In (C1), we have rewritten the minimisation problem as follows:

$$\min_{y \in R^r} \|Hy - \|b\| e_1\| ,$$

where H is of size $r \times r$, smaller than $n \times n$.

The computational cost to solve this reduced problem involves:

- Constructing H and β through Arnoldi iteration, which requires $O(r^2 n)$ FLOPs.
- Solving the least-squares problem $\min_{y \in R^r} \|Hy - \|b\| e_1\|$ using QR decomposition or another method, which has a complexity of $O(r^3)$.

These complexities are much smaller than $O(n^3)$ for direct solutions, particularly for large n and much smaller r values. Moreover, here we consider that H and β are given.

(C3)

By definition of krylov subspaces, since $x \in \{b, Ab, ..., A^{r-1}b\}$, we can write $x = p(A)b$. We then have:

$$b - Ax = b - Ap(A)b = (I - Ap(A))b$$

Let's set $q(A) = I - Ap(A)$. Actually, $q(A)$ also belongs to the $\mathcal{P}_r^0$ set. Indeed, if $A = 0$, $q(A) = I$. This means that, interverting the symbols p and q, we get what we wanted:

$$\min_{x \in K_r(A,b)} \|Ax - b\| = \min_{p \in \mathcal{P}_r^0} \|p(A)b\|$$

(C4)

Since A is symmetric, it can be diagonalized as $A = Q\Lambda Q^T$, where Q is an orthogonal matrix , and $\Lambda = diag(\lambda_1, \ldots, \lambda_n)$ contains the eigenvalues of A on the diagonal. We use the spectral norm of A which is defined as:

$$\|A\| = \max_{\|x\|=1} \|Ax\|.$$

For any vector $x \in R^n$ with $\|x\| = 1$, let $y = Q^T x$. Since Q is orthogonal, we have $\|y\| = \|x\| = 1$, and we can write:

$$\|Ax\| = \|Q\Lambda Q^\top x\| = \|\Lambda y\| = \left\| [\begin{matrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{matrix}]y \right\|.$$

Finally, since $\|\Lambda y\| = \sqrt{\sum_{i=1}^n (\lambda_i y_i)^2} \leq \max_i |\lambda_i| \cdot \|y\| = \max_i |\lambda_i|$, the spectral norm $\|A\|$ is indeed $\max_{1 \leq i \leq n} |\lambda_i|$.

(C5)

By the result in (C3), we know that:

$$\min_{x \in K_r(A,b)} \|Ax - b\| = \min_{p \in P_r^0} \|p(A)b\|.$$

then, since A is symmetric, it can be diagonalized as $A = Q\Lambda Q^T$ where Q is orthogonal and $\Lambda = diag(\lambda_1, \ldots, \lambda_n)$. Therefore, for any polynomial $p \in P_r^0$, we have:

$$p(A) = Qp(\Lambda)Q^T,$$

where $p(\Lambda) = diag(p(\lambda_1), \ldots, p(\lambda_n))$. Thus,

$$\|p(A)b\| = \|Qp(\Lambda)Q^T b\| = \|p(\Lambda)Q^T b\| \leq \max_{1 \leq i \leq n} |p(\lambda_i)| \cdot \|Q^T b\| = \max_{1 \leq i \leq n} |p(\lambda_i)| \cdot \|b\|.$$

Finally, taking the minimum over $p \in P_r^0$, we obtain:

$$\min_{x \in K_r(A,b)} \|Ax - b\| = \min_{p \in P_r^0} \|p(A)b\| \leq \|b\| \min_{p \in P_r^0} \max_{1 \leq i \leq n} |p(\lambda_i)|.$$

4 Exercise D: Arnoldi's method for eigenvalue approximation

(D1)

5 Exercise E: Implementation