MATRIX COMPUTATIONS: HOMEWORK 2 (October 10, 2024)

Krylov-based methods for eigenvalues computation and system solving

In this homework, we will study Krylov-based methods, which are numerical methods to approximate the eigenvalues of a matrix and the solution of a linear system. The idea is simple: given a large matrix A, find smaller matrices Q and H satisfying

$$Q^\top AQ = H.$$

Then, under some conditions on Q and H, the eigenvalues of A can be approximated by the eigenvalues of H and the solution of $Ax = b$ can be approximated by the solution of $Hy = Q^\top b$. In this homework, we will see how to construct efficiently H and Q satisfying the required conditions, and how to bound the approximation errors.

Notation We use a Matlab-style notation for the indexing of matrices. This means that given a matrix A, we denote by $A[a : b]$ the matrix formed by keeping in A only the columns with index from a to b included, and by $A[a : b, c : d]$ the matrix formed by keeping in A only the columns with index from c to d included and the rows with index from a to b included. Given $n \in \mathbb{N}_{>0}$ and $1 \leq k \leq n$, we denote by $e_{n,k}$ the vector in $\mathbb{R}^n$ whose elements are all equal to zero except the k^{th} element which is equal to one. For vectors, $\|\cdot\|$ is the 2-norm, and for matrices, it denotes the spectral norm.

Exercise A: Krylov subspaces

Definition 1. Given $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$ and $r \in \mathbb{N}_{>0}$, the *Krylov subspace* of order r generated by A and b, denoted by $\mathcal{K}_r(A, b)$, is defined by

$$\mathcal{K}_r(A, b) = \operatorname{span}\{b, Ab, \ldots, A^{r-1}b\}.$$

(A1) Let $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$ and $r \in \mathbb{N}_{>0}$. Assume that $\mathcal{K}_{r+1}(A, b) \subseteq \mathcal{K}_r(A, b)$. Show that for all $s \geq r$, it holds that $\mathcal{K}_s(A, b) = \mathcal{K}_r(A, b)$.

(A2) Let $A \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^n \setminus \{0\}$. Let $s = \dim(\mathcal{K}_n(A, b))$. From (A1), deduce that for all $1 \leq r \leq s$, it holds that $\dim(\mathcal{K}_r(A, b)) = r$.

Hint. Use a proof by induction for $r = 1, \ldots, s$.

Exercise B: Arnoldi's iteration

Definition 2. Given $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$ and $r \in \mathbb{N}_{>0}$, define the following $n \times r$ matrix:

$$K_r(A, b) = [b, Ab, \ldots, A^{r-1}b]. \tag{1}$$

(B1) Let $A \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^n \setminus \{0\}$. Let $s = \dim(\mathcal{K}_n(A, b))$. Let (Q_s, R_s) be a QR decomposition of $K_s(A, b)$, meaning that $K_s(A, b) = Q_sR_s$, $Q_s \in \mathbb{R}^{n\times s}$ is an isometry, i.e., $Q_s^\top Q_s = I_s$, and $R_s \in \mathbb{R}^{s\times s}$ is upper-triangular. Show that R_s has nonzero elements on its diagonal.

(B2) Let $A \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^n \setminus \{0\}$. Let $s = \dim(\mathcal{K}_n(A, b))$. Let (Q_s, R_s) be a QR decomposition of $K_s(A, b)$. 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 $r' = r + 1$.

Hint. Use the fact that for any $r \in \mathbb{N}_{>0}$, $K_{r+1}(A, b) = [b, AK_r(A, b)]$.

(B3) Let $A \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^n \setminus \{0\}$. Let $s = \dim(\mathcal{K}_n(A,b))$. Let (Q_s, R_s) be a QR decomposition of $K_s(A,b)$. Show that for each $1 \le r \le s-1$, there is a Hessenberg[1] matrix $H_r \in \mathbb{R}^{r'\times r}$ such that $AQ_s[1:r] = Q_s[1:r']H_r$.

Consider the algorithm in Algo. 1, called *Arnoldi's iteration* or *Arnoldi's method*. By construction, the output of Algo. 1 is an orthonormal sequence of vectors (you do not need to show this trivial fact). We will show that the matrix $[q_1, \ldots, q_s]$ corresponds to Q_s defined in (B1).

Algorithm 1: Arnoldi's method

Data: $A \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^n \setminus \{0\}$
Result: An orthonormal sequence of vectors $(q_1, \ldots, q_s)$
$q_1 \leftarrow b/\|b\|$;
$s \leftarrow 1$;
while *True* **do**
  $v \leftarrow Aq_s$;
  $w \leftarrow v - v'$ where v' is the orthogonal projection of v on span $\{q_1, \ldots, q_s\}$;
  if $w \neq 0$ **then**
    $s \leftarrow s+1$;
    $q_s = w/\|w\|$;
  else
    return $(q_1, \ldots, q_s)$;

(B4) Let $A \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^n \setminus \{0\}$, and let $(q_1, \ldots, q_s)$ be the output of Algo. 1 with input (A, b). Show that for each $1 \le r \le s$, span $\{q_1, \ldots, q_r\} \subseteq \mathcal{K}_r(A,b)$.
Hint. Use a proof by induction.

(B5) Let $A \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^n \setminus \{0\}$, and let $(q_1, \ldots, q_s)$ be the output of Algo. 1 with input (A, b). Using (A2) and (B4), show that for each $1 \le r \le s$, span $\{q_1, \ldots, q_r\} = \mathcal{K}_r(A,b)$.

(B6) Let $A \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^n \setminus \{0\}$, and let $(q_1, \ldots, q_s)$ be the output of Algo. 1 with input (A, b). Show that $s = \dim(\mathcal{K}_n(A,b))$.
Hint. Use (A2). What can you say about Aq_s?

(B7) Let $A \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^n \setminus \{0\}$, and let $(q_1, \ldots, q_s)$ be the output of Algo. 1 with input (A, b). Show that $Q_s \triangleq [q_1, \ldots, q_s]$ provides a QR decomposition of $K_s(A,b)$, i.e., $K_s(A,b) = Q_sR_s$ for some upper-triangular matrix $R_s \in \mathbb{R}^{s\times s}$.
Hint. Use (B5).

Summarizing, given $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n \setminus \{0\}$ and $1 \le r \le \dim(\mathcal{K}_n(A,b))$ with $r < n$, Algo. 1 provides a matrix $Q \in \mathbb{R}^{n\times r}$, a Hessenberg matrix $H \in \mathbb{R}^{r\times r}$ and a scalar $\beta \in \mathbb{R}$ such that there exists $q \in \mathbb{R}^n$ satisfying that (i) $[Q, q]$ is an isometry and (ii) $AQ = QH + \beta q e_{r,r}^\top$; see (B3).[2] Furthermore, the columns of Q form a basis of $\mathcal{K}_r(A,b)$; and it also holds that the first column of Q is $b/\|b\|$.

These are all the properties that we need to use Algo. 1 in problems of eigenvalue approximation and linear system solution approximation.

[1] A matrix H is Hessenberg if all elements below the first subdiagonal are zero, said otherwise, if $H[i,j] = 0$ whenever $i > j+1$.

[2] The Hessenberg matrix H can actually be generated as a by-product of Arnoldi's method.

Exercise C: GMRES for linear system solution approximation

Let us start with the approximation of the solution of a linear system. Given $A \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^n \setminus \{0\}$, consider the linear system

$$Ax = b. \tag{2}$$

When A is large, computing the exact solution of (2) can be prohibitive. The idea of GMRES (Generalized Minimal RESidual) is to approximate the solution of (2) with the solution of the following minimization problem:

$$\min_{x\in\mathcal{K}_r(A,b)} \|Ax - b\|, \tag{3}$$

where $1 \le r \le n$, and use the output of Algo. 1 to solve (3).

(C1) Let $A \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^n \setminus \{0\}$. Let $1 \le r \le n$. Let $Q \in \mathbb{R}^{n\times r}$, $q \in \mathbb{R}^n$, $H \in \mathbb{R}^{r\times r}$ and $\beta \in \mathbb{R}$ be such that $[Q, q]$ is an isometry and $AQ = QH + \beta q e_{r,r}^\top$. Furthermore, assume that the columns of Q form a basis of $\mathcal{K}_r(A, b)$, and the first column of Q is $b/\|b\|$. Show that

$$\min_{x\in\mathcal{K}_r(A,b)} \|Ax - b\| = \min_{y\in\mathbb{R}^r} \left\|\tilde{H}y - \|b\|e_{r+1,1}\right\|, \quad \tilde{H} = \begin{bmatrix} H \\ \beta e_{r,r}^\top \end{bmatrix}, \tag{4}$$

and that if y^* is a minimizer of the right-hand side of (4), then Qy^* is a minimizer of (3).

(C2) Based on (C1), explain why solving (3) is easier than solving (2), and give an estimate of the number of elementary operations (FLOPs) needed to solve (4) if H and β are given.

Now, let us study the approximation error of (3).

(C3) Let $A \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^n \setminus \{0\}$. Let $1 \le r \le n$. Let $\mathcal{P}_r^0$ be the set of polynomials $p \in \mathbb{R}[x]$ of degree at most r such that $p(0) = 1$. Show that $\{b - Ax : x \in \mathcal{K}_r(A, b)\} = \{p(A)b : p \in \mathcal{P}_r^0\}$. Deduce that

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

(C4) Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of A. Show that $\|A\| = \max_{1\le i\le n} |\lambda_i|$.

Hint. Use the eigenvalue decomposition of A and the unitary invariance of the spectral norm.

(C5) Let $A \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^n \setminus \{0\}$. Let $1 \le r \le n$. Let $\mathcal{P}_r^0$ be the set of polynomials p of degree at most r such that $p(0) = 1$. Assume that A is symmetric. Show that

$$\min_{x\in\mathcal{K}_r(A,b)} \|Ax - b\| \le \|b\| \min_{p\in\mathcal{P}_r^0} \max_{1\le i\le n} |p(\lambda_i)|,$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A.

Hint. Use (C3) and the eigenvalue decomposition of A.

Exercise D: Arnoldi's method for eigenvalue approximation

The initial idea of Arnoldi's method (Algo. 1) is that the eigenvalues of H (called *Ritz values*) produce a good approximation of r eigenvalues of A. To quantify this, we will quantify *how close* an eigenpair of $QHQ^\top$ is from being an eigenpair for A.

(D1) Let $A \in \mathbb{R}^{n\times n}$ and $1 \le r < n$. Let $Q \in \mathbb{R}^{n\times r}$, $q \in \mathbb{R}^n$, $H \in \mathbb{R}^{r\times r}$ and $\beta \in \mathbb{R}$ be such that $[Q, q]$ is an isometry and $AQ = QH + \beta q e_r^\top$. Let $(\lambda, y) \in \mathbb{C} \times \mathbb{C}^r$ be an eigenpair of H, meaning that $\|y\| = 1$ and $Hy = \lambda y$. Let $x = Qy$. Show that $\|Ax - \lambda x\| = |\beta e_r^\top y|$.

Remark 1. A bound in the same spirit of (C5) can be obtained for the distance between an eigenvector x of A and the Krylov space $\mathcal{K}_r(A, b)$. Even though the proof is not more difficult than the proof of (C5), we will not see or prove this result for the sake of brevity.

Exercise E: Implementation

(E1) Consider Poisson's equation in 1D:

$$u''(\xi) = f(\xi), \quad \xi \in [0,1], \quad u(0) = u(1) = 0, \tag{5}$$

where $u : \mathbb{R} \to \mathbb{R}$ is unknown. Consider the *force term* $f : [0,1] \to \mathbb{R}$ defined by

$$f(\xi) = \begin{cases} -1 & \text{if } \xi \le 0.2 \\ 5\xi - 2 & \text{if } 0.2 < \xi < 0.8 \\ 2 & \text{if } \xi \ge 0.8. \end{cases}$$

We will solve this equation approximately by using the finite-difference method. For that, let $n = 101$ be a number of *nodes*, and define

$$A = \frac{1}{h^2}\begin{bmatrix} 1 & & & & & & \\ 1 & -2 & 1 & & & & \\ & 1 & -2 & 1 & & & \\ & & 1 & -2 & 1 & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & 1 & -2 & 1 \\ & & & & & & 1 \end{bmatrix} \in \mathbb{R}^{n\times n}, \qquad b = \begin{bmatrix} 0 \\ f(h) \\ f(2h) \\ \vdots \\ f((n-3)h) \\ f((n-2)h) \\ 0 \end{bmatrix} \in \mathbb{R}^n,$$

where $h = 1/(n-1)$. For $r = 10, 20, 30, \dots, 50$, plot the solution of (3).

(E2) Consider Schrödinger's equation in 1D:

$$-u''(\xi) + V(\xi)u(\xi) = Eu(\xi), \quad \xi \in \mathbb{R}, \quad u(-\infty) = u(+\infty) = 0, \tag{6}$$

where $u : \mathbb{R} \to \mathbb{R}$ and $E \in \mathbb{R}$ are unknown. Consider the *potential* $V : \mathbb{R} \to \mathbb{R}$ defined by

$$V(\xi) = \begin{cases} 10 & \text{if } \xi \le -1 \\ 0 & \text{if } -1 < \xi < 1 \\ 10 & \text{if } \xi \ge 1. \end{cases}$$

We will solve this equation approximately by using the finite-difference method. For that, let $n = 101$ be a number of *nodes*, and define

$$\hat{A} = \frac{1}{h^2}\begin{bmatrix} -2 & 1 & & & & & \\ 1 & -2 & 1 & & & & \\ & 1 & -2 & 1 & & & \\ & & 1 & -2 & 1 & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & 1 & -2 & 1 \\ & & & & & 1 & -2 \end{bmatrix} \in \mathbb{R}^{n\times n}, \qquad v = \begin{bmatrix} V(-2+0h) \\ V(-2+h) \\ V(-2+2h) \\ \vdots \\ V(-2+(n-3)h) \\ V(-2+(n-2)h) \\ V(-2+(n-1)h) \end{bmatrix} \in \mathbb{R}^n,$$

where $h = 4/(n-1)$. Let $A = -\hat{A} + \mathrm{diag}(v)$, and let $b = [1, 2, \ldots, n] \in \mathbb{R}^n$. For $r = 50$, plot the approximate eigenvectors of A obtained by using the method in (D1) where the columns of Q form an orthonormal basis of $\mathcal{K}_r(A, b)$. Plot only the approximate eigenvectors associated to the largest 5 eigenvalues of H.

Practical information

The homework solution should be written in English.
Please submit your solution in a `pdf` file named `Group_XX.pdf`.[3]
As you are in master, we strongly recommend you to write your report in latex.
Deadline for turning in the homework: Wednesday October 30, 2024 (11:59 pm).
It is expected that each group makes the homework individually.
If your group has problems or questions, you are welcome to contact the teaching assistants: julien.calbert@uclouvain.be, guillaume.berger@uclouvain.be.

[3] E.g., `Group_01.pdf`, `Group_02.pdf`, `Group_12.pdf`. Failure to comply with these requirements may result in point penalties.