Krylov subspace method

Exercice 1: 1D projection process

For $\mathscr{S}_k, \mathscr{C}_k$ k-dimensional subspaces of $\mathbb{C}^n$, we recall that a projection process is defined by

$$\begin{cases} x^{(k)} = x^{(0)} + z^{(k)}, \quad z^{(k)} \in \mathscr{S}_k \\ r^{(k)} = b - Ax^{(k)} \perp \mathscr{C}_k. \end{cases}$$

We will consider in this exercise, 1D projection processes, *i.e.* $k = 1$ and $\mathscr{S}_1 = \operatorname{Span}(v)$ and $\mathscr{C}_1 = \operatorname{Span}(w)$, for some vectors $v, w \in \mathbb{C}^n$.

1. For $A \in \mathbb{R}^{n\times n}$ symmetric and positive-definite, show that the projection process defined with $v = w = r^{(0)} = b - Ax^{(0)}$ is equivalent to one step of the steepest gradient descent.

2. Let $A \in \mathbb{R}^{n\times n}$ be such that $A + A^T$ is positive-definite.
   (a) Let $\mathscr{S}_1 = \operatorname{Span}(r^{(0)})$ and $\mathscr{C}_1 = \operatorname{Span}(Ar^{(0)})$. Show that $x^{(1)} = x^{(0)} + \alpha r^{(0)}$ with $\alpha_0 = \frac{\langle r^{(0)}, Ar^{(0)}\rangle}{\langle Ar^{(0)}, Ar^{(0)}\rangle}$.
   (b) Consider the sequence defined by for all $k \geq 0$

$$\begin{cases} r^{(k)} = b - Ax^{(k)} \\ \alpha_k = \dfrac{\langle r^{(k)}, Ar^{(k)}\rangle}{\langle Ar^{(k)}, Ar^{(k)}\rangle} \\ x^{(k+1)} = x^{(k)} + \alpha_k r^{(k)}. \end{cases}$$

   (c) Show that $\|r^{(k+1)}\|^2 \leq \|r^{(k)}\|^2 - \frac{\langle r^{(k)}, Ar^{(k)}\rangle^2}{\|Ar^{(k)}\|^2}$.
   (d) Deduce that $\|r^{(k+1)}\| \leq \left(1 - \frac{\lambda^2}{\|A\|^2}\right)^{1/2} \|r^{(k)}\|$ where λ is the smallest eigenvalue of $\frac{1}{2}(A + A^T)$.

Exercise 2: a Krylov method is better than a stationary method

Let $A \in \mathbb{C}^{N\times N}$ be invertible and $A = M - N$ a splitting of the matrix such that M is invertible.

In this exercise, we want to compare the resolution of a linear system by the stationary iterative method

$$\begin{cases} Mx_{\text{stat}}^{(k+1)} = Nx_{\text{stat}}^{(k)} + b, \quad k \geq 0 \\ x_{\text{stat}}^{(0)} \in \mathbb{C}^N, \end{cases} \tag{1}$$

with the Krylov method for the preconditioned system

$$M^{-1}Ax_* = M^{-1}b. \tag{2}$$

The Krylov iterates $(x_{\text{kry}}^{(k)})$, initialised for some $x_{\text{kry}}^{(0)}$ are defined as

$$\|M^{-1}b - M^{-1}Ax_{\text{kry}}^{(k)}\| = \min_{z \in x_{\text{kry}}^{(0)} + \mathcal{K}_k(M^{-1}A, r_{\text{kry}}^{(0)})} \|M^{-1}b - M^{-1}Az\|, \tag{3}$$

where $r_{\text{kry}}^{(k)} = M^{-1}b - M^{-1}Ax_{\text{kry}}^{(k)}$.

1. Show that $x_{\text{kry}}^{(k)} = x_{\text{kry}}^{(0)} + P(M^{-1}A)r_{\text{kry}}^{(0)}$ where P is a polynomial of degree $k-1$.
2. Deduce that $r_{\text{kry}}^{(k)} = \phi(M^{-1}A)r_{\text{kry}}^{(0)}$ where ϕ is a degree k polynomial such that $\phi(0) = 1$.
3. For the stationary iterative method, let $r_{\text{stat}}^{(k)} = M^{-1}b - M^{-1}Ax_{\text{stat}}^{(k)}$. Show that $r_{\text{stat}}^{(k)} = (\text{id} - M^{-1}A)^k r_{\text{stat}}^{(0)}$.
4. Show that if $x_{\text{stat}}^{(0)} = x_{\text{kry}}^{(0)}$ then $\|r_{\text{kry}}^{(k)}\| \leq \|r_{\text{stat}}^{(k)}\|$ for all $k \geq 1$.

Exercise 3: another proof of CG convergence rate

Let A be a HPD matrix with eigenvalues $0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ and associated eigenvectors $y_1, \ldots, y_n$ such that $A = Y\Lambda Y^*$ with $Y = [y_1, \ldots, y_n]$, $Y^*Y = \text{id}$ and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$.

Let $x_* \in \mathbb{C}^n$ be the solution to $Ax_* = b$. For $x_0 \in \mathbb{C}^n$, let $(\zeta_j)_{1\leq j\leq n}$ be the coefficients of

$$x_* - x_0 = \sum_{j=1}^{n} \zeta_j y_j. \tag{4}$$

Let $\widetilde{x}_0$ be defined by

$$x_* - \widetilde{x}_0 = \sum_{j=1}^{n-\ell} \zeta_j y_j, \tag{5}$$

i.e. removing the contributions of the largest eigenvalues.

The goal of this exercise is to compare the CG iterates (x_k) and $(\widetilde{x}_k)$ starting respectively from x_0 and $\widetilde{x}_0$. We are going to show that for all $k \geq \ell$

$$\|x_* - \widetilde{x}_k\|_A \leq \|x_* - x_k\|_A \leq \|x_* - \widetilde{x}_{k-\ell}\|_A. \tag{6}$$

1. Show that for any polynomial ϕ and vector $v = \sum_{j=1}^n \eta_j y_j$, we have

$$\|\phi(A)v\|_A^2 = \sum_{j=1}^n \lambda_j |\phi(\lambda_j)\eta_j|^2.$$

2. Let ϕ_k^{CG} (resp. $\widetilde{\phi}_k^{\mathrm{CG}}$) be the CG iteration polynomial associated to x_0 (resp. $\widetilde{x}_0$). Deduce that

$$\|\widetilde{\phi}_k^{\mathrm{CG}}(A)(x_* - \widetilde{x}_0)\|_A \leq \|\phi_k^{\mathrm{CG}}(A)(x_* - x_0)\|_A,$$

   and

$$\|x_* - \widetilde{x}_k\|_A \leq \|x_* - x_k\|_A.$$

3. Prove that for $k \geq \ell$, $\|x_* - x_k\|_A \leq \|x_* - \widetilde{x}_{k-\ell}\|_A$.

   Hint: recall that $\|x_* - x_k\|_A = \min\limits_{\phi \in \mathbb{C}^k[X], \phi(0)=1} \|\phi(A)(x_* - x_0)\|_A$.

4. Let $\kappa_\ell = \frac{\lambda_{n-\ell}}{\lambda_1}$. Show that

$$\|x_* - \widetilde{x}_{k-\ell}\|_A \leq 2\Big(\frac{\sqrt{\kappa_\ell} - 1}{\sqrt{\kappa_\ell} + 1}\Big)^{k-\ell} \|x_* - \widetilde{x}_0\|_A,$$

   and deduce

$$\|x_* - x_k\|_A \leq 2\Big(\frac{\sqrt{\kappa_\ell} - 1}{\sqrt{\kappa_\ell} + 1}\Big)^{k-\ell} \|x_* - x_0\|_A.$$

Exercise 4: MINRES algorithm

The MINRES algorithm is deduced from GMRES by considering an invertible Hermitian matrix A.

1. Suppose that A has eigenvalues $\lambda_1 \leq \cdots \leq \lambda_s < 0 < \lambda_{s+1} \leq \cdots \leq \lambda_n$. Show that

$$\|r^{(2k)}\| \leq \min_{\phi \in \mathbb{C}^{2k}[X], \phi(0)=1} \max_{1 \leq i \leq n} |\phi(\lambda_i)| \|r^{(0)}\|.$$

2. Deduce that

$$\|r^{(2k)}\| \leq \min_{\phi \in \mathbb{C}^{2k}[X], \phi(0)=1} \max_{\lambda \in [\lambda_1, \lambda_s] \cup [\lambda_{s+1}, \lambda_n]} |\phi(\lambda)| \|r^{(0)}\|.$$

3. Suppose that $\lambda_n + \lambda_1 - \lambda_s - \lambda_{s+1} = 0$. Let $q(x) = 1 + 2\frac{(x-\lambda_s)(x-\lambda_{s+1})}{\lambda_1\lambda_n - \lambda_s\lambda_{s+1}}$. Prove that $q([\lambda_1, \lambda_s] \cup [\lambda_{s+1}, \lambda_n]) = [-1, 1]$ and deduce that

$$\min_{\phi \in \mathbb{C}^{2k}[X], \phi(0)=1} \max_{\lambda \in [\lambda_1, \lambda_s] \cup [\lambda_{s+1}, \lambda_n]} |\phi(\lambda)| \leq \frac{1}{|T_k(q(0))|},$$

where T_k is the k-th Chebyshev polynomial.

4. Show that $q(0) = \frac{1}{2}\Big(\frac{\sqrt{|\lambda_1\lambda_n|} - \sqrt{|\lambda_s\lambda_{s+1}|}}{\sqrt{|\lambda_1\lambda_n|} + \sqrt{|\lambda_s\lambda_{s+1}|}} + \frac{\sqrt{|\lambda_1\lambda_n|} + \sqrt{|\lambda_s\lambda_{s+1}|}}{\sqrt{|\lambda_1\lambda_n|} - \sqrt{|\lambda_s\lambda_{s+1}|}}\Big)$ and deduce that the convergence rate of the MINRES algorithm is bounded by

$$\|r^{(2k)}\| \leq 2\Big(\frac{\sqrt{|\lambda_1\lambda_n|} - \sqrt{|\lambda_s\lambda_{s+1}|}}{\sqrt{|\lambda_1\lambda_n|} + \sqrt{|\lambda_s\lambda_{s+1}|}}\Big)^k \|r^{(0)}\|.$$

Hint: you can use without proof that for all $x \neq 0$, $T_k\big(\frac{1}{2}(x + \frac{1}{x})\big) = \frac{1}{2}\big(x^k + \frac{1}{x^k}\big)$.

5. Suppose from now on that $\lambda_n = -\lambda_1$ and $\lambda_{s+1} = -\lambda_{s+1}$. Denote by κ the 2-norm conditioning number of A. Show that

$$\|r^{(2k)}\| \leq 2\Big(\frac{\kappa - 1}{\kappa + 1}\Big)^k \|r^{(0)}\|.$$

6. Equivalently, it is possible to solve $A^*Ax = A^*b$ if we know A^* (or the matrix-vector product with A^*). Give the rate of convergence of the conjugate-gradient method in that case, and compare to the rate obtained in the previous question.

Exercise 5: a GMRES example

Let $A \in \mathbb{R}^{n \times n}$ be defined by

$$A = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ 0 & 1 & -1 & \ddots & \vdots \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{7}$$

.

1. Show that for every $1 \leq k \leq n$, there exists $r^{(0)} \in \mathbb{R}^n$ such that GMRES stops in exactly k steps.

Exercise 6: stagnation of GMRES algorithm

The goal of this exercise is to show that for any prescribed eigenvalues $\lambda_1, \dots, \lambda_n \in \mathbb{C}$, there exist a matrix $A \in \mathbb{C}^{n\times n}$ with eigenvalues (λ_j) and a starting vector $x^{(0)}$ such that the residuals $r^{(k)}$ of the GMRES algorithm stagnates for $0 \leq k \leq n-1$, *i.e.* $\|r^{(k)}\| = \|r^{(0)}\|$ for any $0 \leq k \leq n-1$.

Let $C \in \mathbb{C}^{n\times n}$ be given by

$$C = \begin{bmatrix} 0 & \dots & 0 & \alpha_0 \\ 1 & \ddots & \vdots & \vdots \\ & \ddots & 0 & \alpha_{n-2} \\ & & 1 & \alpha_{n-1} \end{bmatrix}. \tag{8}$$

1. Show that the characteristic polynomial of C is given by $\chi(\lambda) = \det(\lambda \mathrm{id} - C) = \lambda^n - \sum_{j=0}^{n-1} \alpha_j \lambda^j$.
2. Explain how to set (α_j) so that C has eigenvalues $(\lambda_j)_{1\leq j\leq n}$.
3. Let $(e_k)_{0\leq k\leq n-1}$ be the canonical vectors of $\mathbb{C}^n$. Show that for $1 \leq k \leq n-2$, $(e_1, \dots, e_k)$ is an orthonormal basis of $C\mathcal{K}_k(C, e_0) = \mathrm{Span}(Ce_0, C^2e_0, \dots, C^ke_0)$.
4. Let $b \in \mathbb{C}^n$. Assume that the GMRES algorithm is applied to $Cx_* = b$ with $x^{(0)} = C^{-1}(b - e_0)$. Show that $r^{(k)} = e_0$ for all $0 \leq k \leq n-1$.

Exercise 7: quasi-minimal residual method

Let $A \in \mathbb{C}^{n\times n}$, $x^{(0)}, b \in \mathbb{C}^n$ and $r^{(0)} = b - Ax^{(0)}$.

Let $V_{k+1} = \left[v_1, \dots, v_{k+1}\right]$ be a basis of the Krylov space $\mathcal{K}_{k+1}(A, r^{(0)})$ such that

$$AV_k = V_{k+1}T_{k+1},$$

where $T_{k+1} \in \mathbb{C}^{(k+1)\times k}$ is a tridiagonal matrix.

Such V_{k+1} and T_{k+1} can be constructed (under some assumption), using the nonhermitian Lanczos algorithm. Note that since T_{k+1} is tridiagonal, in general, V_{k+1} does not have orthogonal columns.

We consider the following algorithm

$$x^{(k+1)} = x^{(0)} + V_k t_k, \text{where } t_k = \operatorname*{arg\,min}_{t\in\mathbb{C}^k} \|\|r^{(0)}\|e_1 - T_{k+1}t\|.$$

This iteration scheme is called the *quasi-minimal residual (QMR) method.*

1. What is the advantage of the QMR method compared to GMRES? What can we say about QMR and GMRES in the case where V_{k+1} has orthonormal columns?

2. Show that $\|r_k^{\mathrm{QMR}}\| \leq \sigma_{\max}(V_{k+1}) \big\| \|r^{(0)}\| e_1 - T_{k+1} t_k \big\|$, where $\sigma_{\max}(V_{k+1})$ is the smallest singular value of V_{k+1}.

   Hint: one can first prove that $\|V_{k+1} t\| \leq \sigma_{\max}(V_{k+1}) \|t\|$ for all $t \in \mathbb{C}^k$.

3. Show that the GMRES residual r_k^{GMRES} can be written as $r_k^{\mathrm{GMRES}} = V_{k+1}\big(\|r^{(0)}\| e_1 - T_{k+1} \widehat{t}_k\big)$ for some $\widehat{t}_k \in \mathbb{C}^k$.

4. Deduce that $\|r_k^{\mathrm{GMRES}}\| \geq \sigma_{\min}(V_{k+1}) \|r_k^{\mathrm{QMR}}\|$, where $\sigma_{\min}(V_{k+1})$ is the smallest singular value of V_{k+1}.

5. Show that $\|r_k^{\mathrm{QMR}}\| \leq \frac{\sigma_{\max}(V_{k+1})}{\sigma_{\min}(V_{k+1})} \|r_k^{\mathrm{GMRES}}\|$.