

MS-E1651 - Numerical Matrix Computations

Exercise 6

Amirreza Akbari

October 15, 2023

1. (P70)

Solution: Base Case (n = 0): We know that

$$r_0 = b - Ax_0 = b$$

and

$$p_0 = r_0$$

so

$$p_0 \in \text{Span}\{b\}$$

Moreover, we have

$$x_1 = x_0 + \alpha_0 p_0 = \alpha_0 r_0$$

Since α_0 is just a scalar, we can conclude that $x_1 \in \text{Span}\{b\}$. Therefore, both x_1 and p_0 are in $\mathcal{K}_1(A, b)$.

Induction Step: Assume the statement holds for all $n < i$, then we want to prove it also holds for $n = i$.

Claim 1:

$$r_i \in \text{Span}\{b, Ab, \dots, A^i b\}$$

Proof of Claim 1: We have

$$r_i = b - Ax_i$$

and by the induction hypothesis, we know that $x_i \in \text{Span}\{b, Ab, \dots, A^{i-1} b\}$. Therefore,

$$Ax_i \in \text{Span}\{Ab, A^2 b, \dots, A^i b\}$$

and

$$r_i = b - Ax_i \in \text{Span}\{b, Ab, \dots, A^i b\}$$

Claim 2:

$$p_i \in \text{Span}\{b, Ab, \dots, A^i b\}$$

Proof of Claim 2: Using the expression for p_i ,

$$p_i = r_i - \sum_{k=0}^{i-1} \beta_{i,k} p_k$$

We already know that r_i is in $\text{Span}\{b, Ab, \dots, A^i b\}$, and each p_k is in $\text{Span}\{b, Ab, \dots, A^k b\}$ by the induction hypothesis. Since k is at most $i - 1$, we can conclude that p_i is in $\text{Span}\{b, Ab, \dots, A^i b\}$.

Claim 3:

$$x_{i+1} \in \text{Span}\{b, Ab, \dots, A^{i+1} b\}$$

Proof of Claim 3: We know that $x_i \in \text{Span}\{b, Ab, \dots, A^{i-1} b\}$ and $p_i \in \text{Span}\{b, Ab, \dots, A^i b\}$. Therefore,

$$x_{i+1} = x_i + \alpha_i p_i \in \text{Span}\{b, Ab, \dots, A^{i+1} b\}$$

This completes the induction step, and we have shown that if the statement holds for all $n < i$, then it also holds for $n = i$. Hence, by mathematical induction, the statement is true for all i .

2. (P71)

Solution:

Let's use induction for solving this problem.

Assume that x_i 's are solutions to first problem, and x'_i 's are solutions to the second problem.

Base induction: We have $x_1 = c_0 + x_0$, and $x'_1 = x_0 + \frac{r'_0}{p_0^T A p'_0} p'_0$.

$$r'_0 = b - Ax'_0 = b - Ax_0 \quad (1)$$

$$p'_0 = r'_0 \quad (2)$$

$$x'_1 = x'_0 + \frac{p_0^T r'_0}{p_0^T A p'_0} p'_0 \quad (3)$$

and for second one

$$r_0 = b - Ax_0 - Ac_0 = b - Ax_0 \quad (4)$$

$$p_0 = r_0 \quad (5)$$

$$c_1 = c_0 + \frac{p_0^T r_0}{p_0^T A p_0} p_0 = \frac{p_0^T r_0}{p_0^T A p_0} p_0 \quad (6)$$

$$x_1 = x_0 + c_1 = x_0 + \frac{p_0^T r_0}{p_0^T A p_0} p_0 \quad (7)$$

so we would have $x'_1 = x_1$, and $r'_0 = r_0$.

Induction step: Let's assume for all $n < i + 1$, we have $x'_n = x_n$, $r'_n = r_n$, $p'_n = p_n$.

$$r'_i = b - Ax'_i = b - Ax_i \quad (8)$$

$$p'_i = r'_i - \sum_{k=0}^i \frac{r_i^T A p'_k}{p_k^T A p'_k} p'_k \quad (9)$$

$$x'_{i+1} = x'_i + \frac{p_i^T r'_i}{p_i^T A p'_i} p'_i \quad (10)$$

and for the second one

$$r_i = b - Ax_0 - Ac_i = b - Ax_0 - A(x_i - x_0) = b - Ax_i \quad (11)$$

$$p_i = r_i - \sum_{k=0}^i \frac{r_i^T A p_k}{p_k^T A p_k} p_k \quad (12)$$

$$c_{i+1} = c_i + \frac{p_i^T r_i}{p_i^T A p_i} p_i = c_i + \frac{p_i^T r_i}{p_i^T A p_i} p_i \quad (13)$$

$$x_{i+1} = x_0 + c_{i+1} = x_0 + c_i + \frac{p_i^T r_i}{p_i^T A p_i} p_i = x_0 + (x_i - x_0) + \frac{p_i^T r_i}{p_i^T A p_i} p_i = x_i + \frac{p_i^T r_i}{p_i^T A p_i} p_i \quad (14)$$

based on the induction hypothesis, $x'_{i+1} = x_{i+1}$.

3. (P72)

Solution:

(a) Since $\hat{x} \in S$, so there is a $y \in R^k$ s.t. $Qa = \hat{x}$. Moreover consider cholskey decomposition of $A = LL^T$. So we have

$$\|\hat{x} - x\|_A = \sqrt{(\hat{x} - x)^T A (\hat{x} - x)} = \sqrt{(Qa - x)^T LL^T (Qa - x)} \quad (15)$$

$$= \sqrt{(Qa - x)^T (L^T)^T L^T (Qa - x)} \quad (16)$$

$$= \sqrt{(L^T (Qa - x))^T (L^T (Qa - x))} \quad (17)$$

$$= \sqrt{(L^T Qa - L^T x)^T (L^T Qa - L^T x)} \quad (18)$$

If we substitute $L^T Q = G$ and $L^T x = b$, we would have

$$\sqrt{(L^T Qa - L^T x)^T (L^T Qa - L^T x)} = \sqrt{(Ga - b)^T (Ga - b)} = \|Ga - b\|_2 \quad (19)$$

(b) So based on part a, $G = L^T Q$, $Qa = \hat{x}$, and $L^T x = b$.

So we have

$$G^T Ga = G^T b \quad (20)$$

$$a = (G^T G)^{-1} G^T b \quad (21)$$

$$Qa = Q(G^T G)^{-1} G^T b \quad (22)$$

$$\hat{x} = Q(Q^T A Q)^{-1} Q^T L L^T x \quad (23)$$

$$\hat{x} = Q(Q^T A Q)^{-1} Q^T A x \quad (24)$$

(c)

In the general case, we can have an arbitrary positive definite matrix D , the projection P_A is given by

$$P_A = A(A^T D A)^{-1} A^T D$$

This is exactly the same formula that we have in part b, if we substitute $A = Q$, and $D = A$.

4. (P74)

Solution:

(a)

$$Ax = \lambda x \quad (25)$$

$$QDQ^T x = \lambda x \quad (26)$$

$$Q^T QDQ^T x = Q^T \lambda x \quad (27)$$

$$DQ^T x = \lambda Q^T x \quad (28)$$

$$y = Q^T x \rightarrow Dy = \lambda y \quad (29)$$

Now if we calculate $\det(D - \lambda I) = (10 - \lambda)(0.5 - \lambda)(0.1 - \lambda)$, we see that $\lambda = 10, 0.5, 0.1$. So for $\lambda = 10$:

$$Dy = 10y \quad (30)$$

$$[10y_1, 0.5y_2, 0.1y_3] = 10[y_1, y_2, y_3] \quad (31)$$

$$y_1 = 1, y_2 = 0, y_3 = 0 \quad (32)$$

$$\rightarrow \lambda = 10, x = Qy \quad (33)$$

So for $\lambda = 0.5$:

$$Dy = 10y \quad (34)$$

$$[10y_1, 0.5y_2, 0.1y_3] = 0.5[y_1, y_2, y_3] \quad (35)$$

$$y_1 = 1, y_2 = 1, y_3 = 0 \quad (36)$$

$$\rightarrow \lambda = 0.5, x = Qy \quad (37)$$

So for $\lambda = 0.1$:

$$Dy = 10y \quad (38)$$

$$[10y_1, 0.5y_2, 0.1y_3] = 0.1[y_1, y_2, y_3] \quad (39)$$

$$y_1 = 0, y_2 = 0, y_3 = 1 \quad (40)$$

$$\rightarrow \lambda = 0.1, x = Qy \quad (41)$$

(b)

$$z_i = Az_{i-1} \quad (42)$$

$$\alpha_{i,1}q_1 + \alpha_{i,2}q_2 + \alpha_{i,3}q_3 = A\alpha_{i-1,1}q_1 + A\alpha_{i-1,2}q_2 + A\alpha_{i-1,3}q_3 \quad (43)$$

$$(44)$$

Let's substitute $q_i = Qe_i$, so we would have

$$\alpha_{i,1}Qe_1 + \alpha_{i,2}Qe_2 + \alpha_{i,3}Qe_3 = A\alpha_{i-1,1}Qe_1 + A\alpha_{i-1,2}Qe_2 + A\alpha_{i-1,3}Qe_3 \quad (45)$$

$$(\alpha_{i,1}Q - \alpha_{i-1,1}AQ)e_1 + (\alpha_{i,2}Q - \alpha_{i-1,2}AQ)e_2 + (\alpha_{i,3}Q - \alpha_{i-1,3}AQ)e_3 = 0 \quad (46)$$

Let's multiply a Q^T from right side, so we would have

$$(\alpha_{i,1}I - \alpha_{i-1,1}A)e_1 + (\alpha_{i,2}I - \alpha_{i-1,2}A)e_2 + (\alpha_{i,3}I - \alpha_{i-1,3}A)e_3 = 0$$

Then multiplying Q^T from left and Q from right gives us:

$$(\alpha_{i,1}I - \alpha_{i-1,1}D)e_1 + (\alpha_{i,2}I - \alpha_{i-1,2}D)e_2 + (\alpha_{i,3}I - \alpha_{i-1,3}D)e_3 = 0$$

Since e_i 's are basis of the space, and the right side is zero. The left side is a vector in R^3 . So if we analysis each coordinates of this vectore we should obtain $[(\alpha_{i,1}I - \alpha_{i-1,1}D)]_1 = 0$, $[(\alpha_{i,2}I - \alpha_{i-1,2}D)]_2 = 0$, and $[(\alpha_{i,3}I - \alpha_{i-1,3}D)]_3 = 0$.

So we would have $\alpha_{i,1} = 10\alpha_{i-1,1}$, $\alpha_{i,2} = 0.5\alpha_{i-1,2}$, and $\alpha_{i,3} = 0.1\alpha_{i-1,3}$.

(c) We can see after some iterations, as it is shown in the plot, the two lines indicating the two small eigenvalues start to converge and grow, near iteration 10. The method can struggle when dealing with matrices containing poorly-conditioned eigenvalues. In such cases, "well-conditioned" means that the eigenvalues have significant differences in magnitudes. If the eigenvalues are closely matched in magnitude, the method may have difficulty converging or may converge very slowly. This can lead to instability in calculations and hinder the ability to separate meaningful information from noisy data.

One potential reason for this is that when the eigenvalues are not well-separated, the power iteration method may have trouble identifying the dominant eigenvalues. As a result, the $\alpha_{i,j}$ values may fluctuate unpredictably or grow slowly. The method relies on dominant eigenvalues for rapid convergence, and without clear dominance, numerical instability and slow convergence can become significant issues.

5. (P75)

Solution:

(a)

1. Form the matrix P , where the columns are the orthonormal eigenvectors of A :

$$P = [v_1, v_2, \dots, v_n]$$

Since the eigenvectors are orthonormal, P is an orthogonal matrix, meaning P^T (the transpose of P) is the same as its inverse (P^{-1}).

2. Create a diagonal matrix Λ , where the diagonal entries are the eigenvalues of A :

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

3. Now, you can express A in terms of its orthonormal eigenbasis:

$$A = P\Lambda P^T$$

$$\begin{aligned} \|b\|_A^2 &= b^T A b = b^T P \Lambda P^T b = b^T [v_1, v_2, \dots, v_n] \Lambda [v_1, v_2, \dots, v_n]^T b \\ &= [b^T v_1, b^T v_2, \dots, b^T v_n] \Lambda [b^T v_1, b^T v_2, \dots, b^T v_n]^T = \begin{bmatrix} \lambda_1 b^T v_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 b^T v_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 b^T v_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_n b^T v_n \end{bmatrix} [b^T v_1, b^T v_2, \dots, b^T v_n]^T \\ &= \sum \lambda_i (b^T v_i)^2 \end{aligned}$$

(b) Assume that $q(A) = c_0 + c_1 A + c_2 A^2 + \dots + c_n A^n$.

$$q(A) = c_0 + c_1 A + c_2 A^2 + \dots + c_n A^n \quad (47)$$

$$= c_0 + c_1 P \Lambda P^T + c_2 P \Lambda P^T P \Lambda P^T + \dots + c_n P \Lambda^n P^T \quad (48)$$

and it is clear that

$$\Lambda^i = \begin{bmatrix} \lambda_1^i & 0 & 0 & \dots & 0 \\ 0 & \lambda_2^i & 0 & \dots & 0 \\ 0 & 0 & \lambda_3^i & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_n^i \end{bmatrix} \quad (49)$$

Following same formula iton as last part we have:

$$\begin{aligned} \|q(A)b\|_A^2 &= (q(A)b)^T A (q(A)b) \\ &= (q(A)b)^T P \Lambda P^T (q(A)b) = (q(A)b)^T [v_1, v_2, \dots, v_n] \Lambda [v_1, v_2, \dots, v_n]^T (q(A)b) \\ &= ((c_0 + c_1 P \Lambda P^T + \dots + c_n P \Lambda^n P^T) b)^T [v_1, v_2, \dots, v_n] \Lambda [v_1, v_2, \dots, v_n]^T ((c_0 + c_1 P \Lambda P^T + \dots + c_n P \Lambda^n P^T) b) \end{aligned}$$

Now let's consider term i in the polynomial so we would have

$$(c_i P \Lambda^i P^T b)^T P \Lambda P^T (c_i P \Lambda^i P^T b) = c_i b^T P (\Lambda^i)^T P^T P \Lambda P^T c_i P \Lambda^i P^T b \quad (50)$$

$$= c_i^2 b^T P (\Lambda^i)^T \Lambda \Lambda^i P^T b \quad (51)$$

$$(52)$$

As you can see exactly we remove A from polynomials and we just keep eigenvalues. As you can see it is obvious that there is $q(\lambda_i)^2$ is above equations more than the equation that we had in part a.

6. (P76)

Solution:

(b) As i grows the error converge to 0. However, as L become larger convexity of plot become less and less.