CAAM 335, Fall 2021, Homework 9 - Solutions

Problem 1 (8+6+6=20 points) Let $A \in \mathbb{R}^{n\times n}$ be invertible and $b \in \mathbb{R}^n$. This question analyzes the convergence of an iterative method to solve $Ax = b$.

Given a current iterate $x^{\text{old}} \in \mathbb{R}^n$, the new iterate $x^{\text{new}} \in \mathbb{R}^n$ is computed using

$$x^{\text{new}} = (I - \alpha A)x^{\text{old}} + \alpha b, \tag{1}$$

where $\alpha \in \mathbb{R}$ is a parameter to be determined later. (For the next iteration one sets $x^{\text{old}} \leftarrow x^{\text{new}}$ and repeats, but we only need to consider the single iteration (1).)

(a) Let $x^* \in \mathbb{R}^n$ denote the solution of $Ax = b$.

Show that the errors $e^{\text{old}} = x^{\text{old}} - x^*$, $e^{\text{new}} = x^{\text{new}} - x^*$ obey the iteration

$$e^{\text{new}} = (I - \alpha A)\, e^{\text{old}}. \tag{2}$$

(b) Assume that

$$A = V\Lambda V^{-1}, \tag{3}$$

where $V \in \mathbb{R}^{n\times n}$ is invertible and $\Lambda \in \mathbb{R}^{n\times n}$ is a diagonal matrix with positive diagonal entries $\lambda_1, \ldots, \lambda_n > 0$.

Define $\varepsilon^{\text{old}} = V^{-1}e^{\text{old}}$, $\varepsilon^{\text{new}} = V^{-1}e^{\text{new}}$ and derive a relation between ε^{new} and ε^{old} similar to (2) but with a diagonal matrix instead of $(I - \alpha A)$.

(c) In part (b) you have shown that (2) is equivalent to n scalar equations of the form

$$\varepsilon_j^{\text{new}} = d_j\, \varepsilon_j^{\text{old}}, \qquad j = 1, \ldots n, \tag{4}$$

with scalars d_j depending on $\alpha, \ldots$. What is d_j?

Assume $\varepsilon_j^{\text{old}} \neq 0$. The new error is smaller than the old error if and only if

$$|d_j| < 1. \tag{5}$$

Find the largest interval of all α for with (5) holds for all $j \in \{1, \ldots, n\}$.

Parts (a)-(c) show that the iterative method (1) converges for any starting value if and only if α is in the interval you have determined in (c).

Solution

(a) (7 pts)

$$\begin{aligned} e^{\text{new}} &= x^{\text{new}} - x^* = (I - \alpha A)x^{\text{old}} + \alpha b - x^* \\ &= (I - \alpha A)x^{\text{old}} + \alpha A x^* - x^* = (I - \alpha A)x^{\text{old}} - (I - \alpha A)x^* \\ &= (I - \alpha A)e^{\text{old}}. \end{aligned}$$

(b) (7 pts) We have

$$e^{\text{new}} = (I - \alpha A)e^{\text{old}} = (I - \alpha V \Lambda V^{-1})e^{\text{old}} = (VV^{-1} - \alpha V \Lambda V^{-1})e^{\text{old}}$$
$$= V\,(I - \alpha\Lambda)\,V^{-1}e^{\text{old}}.$$

Define $\varepsilon^{\text{old}} = V^{-1}e^{\text{old}}$, $\varepsilon^{\text{new}} = V^{-1}e^{\text{new}}$, to get

$$\varepsilon^{\text{new}} = (I - \alpha\Lambda)\,\varepsilon^{\text{old}}. \tag{6}$$

Since $(I - \alpha\Lambda)$ is a diagonal matrix, (6) is equivalent to the n scalar iterations

$$\varepsilon_j^{\text{new}} = (1 - \alpha\lambda_j)\,\varepsilon_j^{\text{old}} \tag{7}$$

for $j = 1, \ldots n$.

(c) (6 pts) $d_j = 1 - \alpha\lambda_j$ and $|1 - \alpha\lambda_j| < 1$, i.e., if and only if

$$-1 < 1 - \alpha\lambda_j < 1.$$

The latter inequalities hold for all $\alpha \in \mathbb{R}$ with (recall that $\lambda_1, \ldots, \lambda_n > 0$)

$$\alpha < 2/\lambda_j \text{ and } \alpha > 0.$$

The previous inequalities are satisfied for all $j \in \{1, \ldots, n\}$ if and only if

$$\alpha \in \left(0, \frac{2}{\max_{j\in\{1,\ldots,n\}} \lambda_j}\right).$$

Problem 2 (6+7+7= 20points) Let A be a matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$ and corresponding eigenvectors $v_1, \ldots, v_n$. Answer the following questions and justify your answer.

(a) What are the eigenvalues and eigenvectors of $A + 2I$?

(b) Let T be invertible. What are the eigenvalues and eigenvectors of $T^{-1}AT$?

(c) Let A be invertible. What are the eigenvalues and eigenvectors of A^{-1}?

Solution

(a) The eigenvalues of $A+2I$ are $\lambda_1+2, \ldots, \lambda_n+2$ with the corresponding eigenvectors $v_1, \ldots, v_n$. This is true because $(A+2I)v_j = Av_j + 2v_j = (\lambda_j + 2)v_j$ for all $j = 1, \ldots, n$.

(b) The eigenvalues of $T^{-1}AT$ are $\lambda_1, \ldots, \lambda_n$ with the corresponding eigenvectors $T^{-1}v_1, \ldots, T^{-1}v_n$. This is true because $T^{-1}AT(T^{-1}v_j) = T^{-1}A(TT^{-1})v_j = T^{-1}Av_j = \lambda_j T^{-1}v_j$ for all $j = 1, \ldots, n$.

(c) Because A is invertible all eigenvalues of A are non-zero.

The eigenvalues of A^{-1} are $\frac{1}{\lambda_1}, \dots, \frac{1}{\lambda_n}$ with the corresponding eigenvectors $v_1, \dots, v_n$. This is true because $A^{-1}v_j = \frac{1}{\lambda_j}A^{-1}\lambda_j v_j = \frac{1}{\lambda_j}A^{-1}Av_j = \frac{1}{\lambda_j}v_j$ for all $j = 1, \dots, n$.

Problem 3 (10+10+10=30 points)

(a) Given a symmetric matrix $A \in \mathbb{R}^{n\times n}$ and a vector $b \in \mathbb{R}^n$ consider the minimization problem

$$\min_{x \in \mathbb{R}^n} \tfrac{1}{2}x^T Ax + b^T x. \tag{8}$$

Use the diagonalization

$$A = Q\Lambda Q^T,$$

where $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_n) \in \mathbb{R}^{n\times n}$ and $Q \in \mathbb{R}^{n\times n}$ is orthogonal, to transform (8) into

$$\min_{z \in \mathbb{R}^n} \tfrac{1}{2}z^T \Lambda z + c^T z. \tag{9}$$

How are x and z related? How are b and c related?

(b) Under what conditions on $\lambda_1, \dots, \lambda_n$ does (9) have a unique solution? What is the solution?

(**Hint:** If $g_j : \mathbb{R} \to \mathbb{R}$, $j = 1, \dots, n$, are given functions, then the minimizer $z = (z_1, \dots, z_n) \in \mathbb{R}^n$ of the function $g(z) \stackrel{\text{def}}{=} \sum_{j=1}^n g_j(z_j)$ is obtained by minimizing $g_j(z_j)$, $j = 1, \dots, n$, individually.)

(c) Let

$$\underbrace{\begin{pmatrix} 2 & -2 \\ -2 & 5 \end{pmatrix}}_{=A} = \underbrace{\begin{pmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ -2/\sqrt{5} & 1/\sqrt{5} \end{pmatrix}}_{=Q} \underbrace{\begin{pmatrix} 6 & 0 \\ 0 & 1 \end{pmatrix}}_{=\Lambda} \underbrace{\begin{pmatrix} 1/\sqrt{5} & -2/\sqrt{5} \\ 2/\sqrt{5} & 1/\sqrt{5} \end{pmatrix}}_{=Q^T}$$

and

$$b = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Compute the solution z of (9) and the solution x of (8).

Solution

(a) (10 pts) We insert $A = Q\Lambda Q^T$ into $\frac{1}{2}x^T Ax + b^T x$ and use the orthogonality of Q (i.e., $QQ^T = I$) to obtain

$$\tfrac{1}{2}x^T Ax + b^T x = \tfrac{1}{2}\underbrace{x^T Q}_{=z^T}\Lambda\underbrace{Q^T x}_{=z} + b^T x = \tfrac{1}{2}\underbrace{x^T Q}_{=z^T}\Lambda\underbrace{Q^T x}_{=z} + \underbrace{b^T Q}_{=c^T}\underbrace{Q^T x}_{=z}.$$

Thus the minimization problem (8) is equivalent to the minimization problem (9).

(b) (10 pts) Since Λ is diagonal, we find that

$$\tfrac{1}{2}z^T\Lambda z + c^T z = \sum_{j=1}^{n} \tfrac{1}{2}\lambda_j z_j^2 + c_j z_j$$

and the solution of

$$\min_{z\in\mathbb{R}^n} \tfrac{1}{2}z^T\Lambda z + c^T z = \min_{z_1,\ldots,z_n} \sum_{j=1}^{n} \tfrac{1}{2}\lambda_j z_j^2 + c_j z_j$$

can be obtained by minimizing the scalar functions $\frac{1}{2}\lambda_j z_j^2 + c_j z_j$ individually over z_j.

The quadratic function $z_j \mapsto \frac{1}{2}\lambda_j z_j^2 + c_j z_j$ does not have a minimum if $\lambda_j < 0$. It has a unique minimum if $\lambda_j > 0$. If $\lambda_j = 0$ and $c_j \neq 0$ it does not have a minimum, and if $\lambda_j = 0$ and $c_j = 0$ it is the zeros functions which attains its minimum at every $z_j \in \mathbb{R}$. More formally, to find the minimum of the scalar function $f(z_j) = \frac{1}{2}\lambda_j z_j^2 + c_j z_j$ we need

$$f'(z_j) = \lambda_j z_j + c_j, \quad f''(z_j) = \lambda_j.$$

If z_j is a minimizer of f, then

$$f'(z_j) = \lambda_j z_j + c_j = 0 \quad \text{and} \quad f''(z_j) = \lambda_j \geq 0.$$

If $\lambda_j = 0$, then $f'(z_j) = \lambda_j z_j + c_j = 0$ only if $c_j = 0$ and in this case every $z_j \in \mathbb{R}$ is a minimizer. If $\lambda_j > 0$, then $f'(z_j) = \lambda_j z_j + c_j = 0$ implies $z_j = -c_j/\lambda_j$ and this point is the unique minimum.

Thus, (9) has a unique solution if and only if all eigenvalues of A are positive, $\lambda_1, \ldots, \lambda_n > 0$, and in this case the solution is given by

$$z_j = -c_j/\lambda_j, \quad j = 1, \ldots, n.$$

(c) (10 pts) If

$$\underbrace{\begin{pmatrix} 2 & -2 \\ -2 & 5 \end{pmatrix}}_{=A} = \underbrace{\begin{pmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ -2/\sqrt{5} & 1/\sqrt{5} \end{pmatrix}}_{=Q} \underbrace{\begin{pmatrix} 6 & 0 \\ 0 & 1 \end{pmatrix}}_{=\Lambda} \underbrace{\begin{pmatrix} 1/\sqrt{5} & -2/\sqrt{5} \\ 2/\sqrt{5} & 1/\sqrt{5} \end{pmatrix}}_{=Q^T}, \quad b = \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

then

$$c = Q^T b = \frac{1}{\sqrt{5}} \begin{pmatrix} -1 \\ 3 \end{pmatrix}.$$

The solution z of (9) is

$$z = \frac{1}{\sqrt{5}} \begin{pmatrix} 1/6 \\ -3 \end{pmatrix}$$

and the solution x of (8) is

$$x = Qz = \begin{pmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ -2/\sqrt{5} & 1/\sqrt{5} \end{pmatrix} \frac{1}{\sqrt{5}} \begin{pmatrix} 1/6 \\ -3 \end{pmatrix} = \frac{1}{5} \begin{pmatrix} 1/6 - 6 \\ -1/3 - 3 \end{pmatrix} = \begin{pmatrix} -7/6 \\ -2/3 \end{pmatrix}.$$

Problem 4 (10+5+5+5+5=30 points)

Let

$$A = \begin{pmatrix} 1 & \sqrt{2} & 1 \\ 1 & -\sqrt{2} & 1 \end{pmatrix}, \; b = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

(a) Diagonalize $A^T A$, i.e., find an orthogonal matrix $V \in \mathbb{R}^{3\times 3}$ and a diagonal matrix $\Lambda \in \mathbb{R}^{3\times 3}$ such that

$$A^T A = V \Lambda V^T.$$

Hint: the eigenvalues of $A^T A$ are $\lambda_1 = \lambda_2 = 4, \lambda_3 = 0$.

(b) Compute the SVD of A, i.e., find an orthogonal matrix $U \in \mathbb{R}^{2\times 2}$, and a diagonal matrix $\Sigma \in \mathbb{R}^{2\times 3}$ such that

$$A = U \Sigma V^T.$$

(c) Compute

$$x^\dagger = A^\dagger b = \sum_{j=1}^{2} \frac{1}{\sigma_j} u_j^T b \, v_j.$$

(d) Show that $x = x^\dagger$ solves the least squares problem

$$\min_{x \in \mathbb{R}^3} \|Ax - b\|_2 \tag{10}$$

(e) Are there any other solutions to (10) besides $x^\dagger$? Why or why not?

Solution

(a) First, we compute $A^T A$:

$$A^T A = \begin{pmatrix} 2 & 0 & 2 \\ 0 & 4 & 0 \\ 2 & 0 & 2 \end{pmatrix}.$$

Given the hint, we must compute (for λ_1, λ_2)

$$\mathcal{N}(A^T A - 4I) = \mathcal{N}\begin{pmatrix} -2 & 0 & 2 \\ 0 & 0 & 0 \\ 2 & 0 & -2 \end{pmatrix} = \mathcal{N}\begin{pmatrix} -2 & 0 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \text{span}\left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\},$$

and (for λ_3),

$$\mathcal{N}(A^T A - 0I) = \mathcal{N}\begin{pmatrix} 2 & 0 & 2 \\ 0 & 4 & 0 \\ 2 & 0 & 2 \end{pmatrix} = \mathcal{N}\begin{pmatrix} 2 & 0 & 2 \\ 0 & 4 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \text{span}\left\{ \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

Note that this vector is already orthogonal to the basis vectors of the previous eigenspace. Therefore, $A^T A$ can be diagonalized by the matrices

$$V = \begin{pmatrix} 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 0 & 1 & 0 \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{pmatrix}, \; \Lambda = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

(b) The singular values are

$$\sigma_1 = \sqrt{\lambda_1} = 2,\ \sigma_2 = \sqrt{\lambda_2} = 2.$$

The left singular vectors are chosen so that

$$AV = U\Sigma,$$

i.e.,

$$u_1 = \frac{1}{\sigma_1} A v_1 = \frac{1}{2} A \begin{pmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

and

$$u_2 = \frac{1}{\sigma_2} A v_2 = \frac{1}{2} A \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Therefore,

$$U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},\ \Sigma = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \end{pmatrix}.$$

(c) This is

$$\begin{aligned} x^\dagger &= \frac{1}{2} v_1 (u_1^T b) + \frac{1}{2} v_2 (u_2^T b) \\ &= \frac{1}{2} \begin{pmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{pmatrix} \frac{3}{\sqrt{2}} + \frac{1}{2} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \frac{1}{\sqrt{2}} = \frac{3}{4} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + \frac{1}{2\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 3/4 \\ 1/(2\sqrt{2}) \\ 3/4 \end{pmatrix}. \end{aligned}$$

(d) In order to be a solution to the least squares problem, it must hold that

$$A^T (Ax^\dagger - b) = 0.$$

We first compute

$$Ax^\dagger = \begin{pmatrix} 2 \\ 1 \end{pmatrix} = b.$$

Therefore, the residual $Ax^\dagger - b$ is exactly zero, so it is a solution to the least squares problem. Of course, this was to be expected, as this is an underdetermined least squares problem involving a full-rank matrix.

(e) A has a non-trivial nullspace, i.e.,

$$\mathcal{N}(A) = \operatorname{span}\left(\underbrace{\begin{pmatrix} -1/\sqrt{2} \\ 0 \\ 1\sqrt{2} \end{pmatrix}}_{= v_3} \right).$$

Thus, $x^\dagger + \alpha v_3$ will also be a solution to the least-squares problem, for any scalar α. $x^\dagger$ is the minimum-norm solution.

Alternative Solution to get the SVD

(a) If we set $B = A^T$ and compute the SVD $B = \widetilde{U}\widetilde{\Sigma}\widetilde{V}^T$, then $A = B^T = \widetilde{V}\widetilde{\Sigma}^T\widetilde{U}^T$ is the SVD of A. Since

$$B^T B = \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}$$

its $\lambda_1 = \lambda_2 = 4$ is a double eigenvalue.

$$\mathcal{N}(4I - B^T B) = \mathcal{N}\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = \text{span}\left\{\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix}\right\},$$

Hence

$$\widetilde{\Sigma} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \\ 0 & 0 \end{pmatrix}, \qquad \widetilde{V} = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}.$$

$$\widetilde{u}_1 = \frac{1}{2}Bv_1 = \begin{pmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{pmatrix},$$

$$\widetilde{u}_2 = \frac{1}{2}Bv_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

This third column $\widetilde{u}_3$ is computed from

$$\mathcal{N}(B^T) = \mathcal{N}(\begin{pmatrix} 1 & \sqrt{2} & 1 \\ 1 & -\sqrt{2} & 1 \end{pmatrix}) = \mathcal{N}(\begin{pmatrix} 1 & \sqrt{2} & 1 \\ 0 & -2\sqrt{2} & 0 \end{pmatrix}) = \text{span}\Bigg(\underbrace{\begin{pmatrix} -1/\sqrt{2} \\ 0 \\ 1\sqrt{2} \end{pmatrix}}_{=\widetilde{u}_3}\Bigg).$$

Hence

$$\widetilde{U} = \begin{pmatrix} 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 0 & 1 & 0 \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{pmatrix}.$$

Note we could have also compute the following basis

$$\mathcal{N}(4I - B^T B) = \mathcal{N}\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = \text{span}\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right),$$

which gives $\widetilde{V} = I$ and different columns $\widetilde{u}_1, \widetilde{u}_2$.