Fall, 2024   CIS 515

Fundamentals of Linear Algebra and Optimization
Jean Gallier

Homework 5

November 13, 2024; Due November 27, 2024

Problem B7 requires knowledge of the generalization of Newton's method to matrices. You will need to read Chapter 5 of Vol II of my book (linalg-II.pdf). The goal is to find algorithms to find the square root of various real matrices. This is a challenging problem which states some open problems. It is probably wise to attempt some of the other easier problems rather than focusing on B7 alone.

Do any of the problems below so that the total number of points attempted is 300. Any additional points will be counted as extra credit.

Problem B1 (100 pts). (a) Let $\mathfrak{so}(3)$ be the space of 3×3 skew symmetric matrices

$$\mathfrak{so}(3) = \left\{ \begin{pmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{pmatrix} \;\middle|\; a, b, c \in \mathbb{R} \right\}.$$

For any matrix

$$A = \begin{pmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{pmatrix} \in \mathfrak{so}(3),$$

if we let $\theta = \sqrt{a^2 + b^2 + c^2}$ and

$$B = \begin{pmatrix} a^2 & ab & ac \\ ab & b^2 & bc \\ ac & bc & c^2 \end{pmatrix},$$

prove that

$$\begin{aligned} A^2 &= -\theta^2 I + B, \\ AB &= BA = 0. \end{aligned}$$

From the above, deduce that

$$A^3 = -\theta^2 A.$$

(b) Prove that the exponential map $\exp\colon \mathfrak{so}(3) \to \mathbf{SO}(3)$ is given by

$$\exp A = e^A = \cos\theta\, I_3 + \frac{\sin\theta}{\theta} A + \frac{(1-\cos\theta)}{\theta^2} B,$$

or, equivalently, by

$$e^A = I_3 + \frac{\sin\theta}{\theta} A + \frac{(1-\cos\theta)}{\theta^2} A^2, \quad \text{if } \theta \neq 0,$$

with $\exp(0_3) = I_3$.

(c) Prove that e^A is an orthogonal matrix of determinant $+1$, i.e., a rotation matrix.

(d) Prove that the exponential map $\exp\colon \mathfrak{so}(3) \to \mathbf{SO}(3)$ is surjective. For this, proceed as follows: Pick any rotation matrix $R \in \mathbf{SO}(3)$;

(1) The case $R = I$ is trivial.

(2) If $R \neq I$ and $\mathrm{tr}(R) \neq -1$, then

$$\exp^{-1}(R) = \left\{ \frac{\theta}{2\sin\theta}(R - R^T) \;\middle|\; 1 + 2\cos\theta = \mathrm{tr}(R) \right\}.$$

(Recall that $\mathrm{tr}(R) = r_{1\,1} + r_{2\,2} + r_{3\,3}$, the *trace* of the matrix R).

Show that there is a unique skew-symmetric B with corresponding θ satisfying $0 < \theta < \pi$ such that $e^B = R$.

(3) If $R \neq I$ and $\mathrm{tr}(R) = -1$, then prove that the eigenvalues of R are $1, -1, -1$, that $R = R^\top$, and that $R^2 = I$. Prove that the matrix

$$S = \frac{1}{2}(R - I)$$

is a symmetric matrix whose eigenvalues are $-1, -1, 0$. Thus, S can be diagonalized with respect to an orthogonal matrix Q as

$$S = Q \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} Q^\top.$$

Prove that there exists a skew symmetric matrix

$$U = \begin{pmatrix} 0 & -d & c \\ d & 0 & -b \\ -c & b & 0 \end{pmatrix}$$

so that

$$U^2 = S = \frac{1}{2}(R - I).$$

Observe that

$$U^2 = \begin{pmatrix} -(c^2+d^2) & bc & bd \\ bc & -(b^2+d^2) & cd \\ bd & cd & -(b^2+c^2) \end{pmatrix},$$

and use this to conclude that if $U^2 = S$, then $b^2 + c^2 + d^2 = 1$. Then, show that

$$\exp^{-1}(R) = \left\{ (2k+1)\pi \begin{pmatrix} 0 & -d & c \\ d & 0 & -b \\ -c & b & 0 \end{pmatrix},\ k \in \mathbb{Z} \right\},$$

where (b, c, d) is any unit vector such that for the corresponding skew symmetric matrix U, we have $U^2 = S$.

(e) To find a skew symmetric matrix U so that $U^2 = S = \frac{1}{2}(R - I)$ as in (d), we can solve the system

$$\begin{pmatrix} b^2 - 1 & bc & bd \\ bc & c^2 - 1 & cd \\ bd & cd & d^2 - 1 \end{pmatrix} = S.$$

We immediately get b^2, c^2, d^2, and then, since one of b, c, d is nonzero, say b, if we choose the positive square root of b^2, we can determine c and d from bc and bd.

Implement a computer program to solve the above system.

Problem B2 (40 pts). Consider the 2×2 real matrices with zero trace,

$$A = \begin{pmatrix} a & b \\ c & -a \end{pmatrix}.$$

(1) If $a^2 + bc < 0$, let $\omega > 0$ be the real such that $\omega^2 = -(a^2 + bc)$. Prove that

$$e^A = \cos\omega I + \frac{\sin\omega}{\omega} A.$$

(2) Find two real 2×2 matrices A and B such that $AB \neq BA$, yet $e^{A+B} = e^A e^B$.

Problem B3 (20 pts). Let A be a real $n \times n$ matrix and consider the $2n \times 2n$ real symmetric matrix

$$S = \begin{pmatrix} 0 & A \\ A^\top & 0 \end{pmatrix}.$$

Suppose that A has rank r.

(1) If $A = V\Sigma U^\top$ is an SVD for A, with $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$ and $\sigma_1 \geq \cdots \geq \sigma_r > 0$, denoting the columns of U by u_k and the columns of V by v_k, prove that σ_k is an eigenvalue of S with corresponding eigenvector $\begin{pmatrix} v_k \\ u_k \end{pmatrix}$ for $k = 1, \ldots, n$, and that $-\sigma_k$ is an eigenvalue of S with corresponding eigenvector $\begin{pmatrix} v_k \\ -u_k \end{pmatrix}$ for $k = 1, \ldots, n$.

Hint. We have $Au_k = \sigma_k v_k$ for $k = 1, \ldots, n$. Prove that $A^\top v_k = \sigma_k u_k$ for $k = 1, \ldots, n$.

(2) Prove that the $2n$ eigenvectors of S in (1) are pairwise orthogonal.

Check that if A has rank r, then S has rank $2r$.

Problem B4 (30 pts). Let $f \colon \mathbb{C}^n \to \mathbb{C}^n$ be a linear map.

(1) Prove that if f is diagonalizable and if $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of f, then $\lambda_1^2, \ldots, \lambda_n^2$ are the eigenvalues of $f^2 = f \circ f$, and if $\lambda_i^2 = \lambda_j^2$ implies that $\lambda_i = \lambda_j$, then f and f^2 have the same eigenspaces.

   Hint. Consider the direct sum decomposition of the eigenspaces and a dimension argument.

(2) Let f and g be two real self-adjoint linear maps $f, g \colon \mathbb{R}^n \to \mathbb{R}^n$. Prove that if f and g have nonnegative eigenvalues (f and g are positive semidefinite) and if $f^2 = g^2$, then $f = g$.

Problem B5 (10 pts). Let A be an real $n \times n$ matrix. Assume A is invertible. Prove that if $A = Q_1 S_1 = Q_2 S_2$ are two polar decompositions of A, then $Q_1 = Q_2$ and $S_1 = S_2$.

Hint. $A^{\mathrm{T}} A = S_1^2 = S_2^2$, with S_1 and S_2 symmetric positive definite. Then use B4.

Problem B6 (100 pts). Recall that a matrix $B \in \mathrm{M}_n(\mathbb{R})$ is skew-symmetric if

$$B^\top = -B.$$

The set $\mathfrak{so}(n)$ of skew-symmetric matrices is a vector space of dimension $n(n-1)/2$, and thus is isomorphic to $\mathbb{R}^{n(n-1)/2}$.

(1) Given a rotation matrix

$$R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix},$$

where $0 < \theta < \pi$, prove that there is a skew symmetric matrix B such that

$$R = (I - B)(I + B)^{-1}.$$

(2)

Let $C\colon \mathfrak{so}(n) \to \mathrm{M}_n(\mathbb{R})$ be the function given by

$$C(B) = (I - B)(I + B)^{-1}.$$

Prove that if B is skew-symmetric, then $I - B$ and $I + B$ are invertible, and so C is well-defined.

Hint. The eigenvalues of a skew-symmetric matrix are either 0 or pure imaginary (that is, of the form $i\mu$ for $\mu \in \mathbb{R}$).

(3) Prove that

$$(I + B)(I - B) = (I - B)(I + B),$$

and that

$$(I + B)(I - B)^{-1} = (I - B)^{-1}(I + B).$$

Prove that

$$(C(B))^{\top} C(B) = I$$

and that

$$\det C(B) = +1,$$

so that $C(B)$ is a rotation matrix in $\mathbf{SO}(n)$. Furthermore, show that $C(B)$ does not admit -1 as an eigenvalue.

(4) Let $\mathbf{SO}(n)$ be the group of $n \times n$ rotation matrices. Prove that the map

$$C\colon \mathfrak{so}(n) \to \mathbf{SO}(n)$$

is bijective onto the subset of rotation matrices that do not admit -1 as an eigenvalue. Show that the inverse of this map is given by

$$B = (I + R)^{-1}(I - R) = (I - R)(I + R)^{-1},$$

where $R \in \mathbf{SO}(n)$ does not admit -1 as an eigenvalue.

(5) Prove that

$$dC_B(A) = -[I + (I - B)(I + B)^{-1}]A(I + B)^{-1} = -2(I + B)^{-1}A(I + B)^{-1},$$

for any $B \in \mathfrak{so}(n)$ and any $A \in \mathrm{M}_n(\mathbb{R})$.

Hint. Use the chain rule, the product rule, and the formula for the derivative of the map $A \mapsto A^{-1}$.

Prove that dC_B is injective for every skew-symmetric matrix B.

Problem B7 (300 +100 pts). (Newton's method to find the square root of a matrix).

First read Chapter 5 on Newton's method in linalg-II (Vol II).

Consider generalizing Problem 5.1 of linalg-II to the matrix function f given by $f(X) = X^2 - C$, where X and C are two real $n \times n$ matrices with C symmetric positive definite. The first step is to determine for which A does the inverse df_A^{-1} exist. Let g be the function given by $g(X) = X^2$.

Prove that that the derivative at A of the function g is $dg_A(X) = AX + XA$, and obviously $df_A = dg_A$.

Thus we are led to figure out when the linear matrix map $X \mapsto AX + XA$ is invertible. This can be done using the Kronecker product.

Given an $m \times n$ matrix $A = (a_{ij})$ and a $p \times q$ matrix $B = (b_{ij})$, the *Kronecker product* (or *tensor product*) $A \otimes B$ of A and B is the $mp \times nq$ matrix

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{pmatrix}.$$

It can be shown (and you may use these facts without proof) that $\otimes$ is associative and that

$$\begin{aligned} (A \otimes B)(C \otimes D) &= AC \otimes BD \\ (A \otimes B)^\top &= A^\top \otimes B^\top, \end{aligned}$$

whenever AC and BD are well defined.

Given any $n \times n$ matrix X, let $\mathrm{vec}(X)$ be the vector in $\mathbb{R}^{n^2}$ obtained by concatenating the *rows* of X.

(1) Prove that $AX = Y$ iff

$$(A \otimes I_n)\mathrm{vec}(X) = \mathrm{vec}(Y)$$

and $XA = Y$ iff

$$(I_n \otimes A^\top)\mathrm{vec}(X) = \mathrm{vec}(Y).$$

Deduce that $AX + XA = Y$ iff

$$((A \otimes I_n) + (I_n \otimes A^\top))\mathrm{vec}(X) = \mathrm{vec}(Y).$$

In the case where $n = 2$ and if we write

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

check that

$$A \otimes I_2 + I_2 \otimes A^\top = \begin{pmatrix} 2a & c & b & 0 \\ b & a+d & 0 & b \\ c & 0 & a+d & c \\ 0 & c & b & 2d \end{pmatrix}.$$

The problem is to determine when the matrix $(A \otimes I_n) + (I_n \otimes A^\top)$ is invertible.

Remark: The equation $AX + XA = Y$ is a special case of the equation $AX + XB = C$ (sometimes written $AX - XB = C$), called the *Sylvester equation*, where A is an $m \times m$ matrix, B is an $n \times n$ matrix, and X, C are $m \times n$ matrices; see Higham [1] (Appendix B).

(2) In the case where $n = 2$, prove that

$$\det(A \otimes I_2 + I_2 \otimes A^\top) = 4(a+d)^2(ad - bc).$$

(3) Let A and B be any two $n \times n$ complex matrices. Use Schur factorizations $A = UT_1U^*$ and $B = VT_2V^*$ (where U and V are unitary and T_1, T_2 are upper-triangular) to prove that if $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A and $\mu_1, \ldots, \mu_n$ are the eigenvalues of B, then the scalars $\lambda_i \mu_j$ are the eigenvalues of $A \otimes B$, for $i, j = 1, \ldots, n$.

Hint. Check that $U \otimes V$ is unitary and that $T_1 \otimes T_2$ is upper triangular.

(4) Prove that the eigenvalues of $(A \otimes I_n) + (I_n \otimes B)$ are the scalars $\lambda_i + \mu_j$, for $i, j = 1, \ldots, n$. Deduce that the eigenvalues of $(A \otimes I_n) + (I_n \otimes A^\top)$ are $\lambda_i + \lambda_j$, for $i, j = 1, \ldots, n$. Thus $(A \otimes I_n) + (I_n \otimes A^\top)$ is invertible iff $\lambda_i + \lambda_j \neq 0$, for $i, j = 1, \ldots, n$. In particular, prove that if A is symmetric positive definite, then so is $(A \otimes I_n) + (I_n \otimes A^\top)$.

Hint. Use (3).

(5) Prove that if A and B are symmetric and $(A \otimes I_n) + (I_n \otimes A^\top)$ is invertible, then the unique solution X of the equation $AX + XA = B$ is symmetric.

(6) Starting with a symmetric positive definite matrix X_0, the general step of Newton's method is

$$X_{k+1} = X_k - (f'_{X_k})^{-1}(X_k^2 - C) = X_k - (g'_{X_k})^{-1}(X_k^2 - C),$$

and since g'_{X_k} is linear, this is equivalent to

$$X_{k+1} = X_k - (g'_{X_k})^{-1}(X_k^2) + (g'_{X_k})^{-1}(C).$$

But since X_k is SPD, $(g'_{X_k})^{-1}(X_k^2)$ is the unique solution of

$$X_k Y + Y X_k = X_k^2$$

whose solution is obviously $Y = (1/2)X_k$. Therefore the Newton step is

$$X_{k+1} = X_k - (g'_{X_k})^{-1}(X_k^2) + (g'_{X_k})^{-1}(C) = X_k - \frac{1}{2}X_k + (g'_{X_k})^{-1}(C) = \frac{1}{2}X_k + (g'_{X_k})^{-1}(C),$$

so we have

$$X_{k+1} = \frac{1}{2}X_k + (g'_{X_k})^{-1}(C) = (g'_{X_k})^{-1}(X_k^2 + C).$$

Prove that if X_k and C are symmetric positive definite, then $(g'_{X_k})^{-1}(C)$ is symmetric positive definite, and if C is symmetric positive semidefinite, then $(g'_{X_k})^{-1}(C)$ is symmetric positive semidefinite.

Hint. By (5) the unique solution Z of the equation $X_kZ + ZX_k = C$ (where C is symmetric) is symmetric so it can be diagonalized as $Z = QDQ^\top$ with Q orthogonal and D a real diagonal matrix. Prove that

$$Q^\top X_k QD + DQ^\top X_k Q = Q^\top CQ,$$

and solve the system using the diagonal elements.

Deduce that if X_k and C are SPD, then X_{k+1} is SPD.

Since $C = P\Sigma P^\top$ is SPD, it has an SPD square root (in fact unique) $C^{1/2} = P\Sigma^{1/2}P^\top$. Prove that

$$X_{k+1} - C^{1/2} = (g'_{X_k})^{-1}(X_k - C^{1/2})^2.$$

Prove that

$$\left\|(g'_{X_k})^{-1}\right\|_2 \geq \frac{1}{2\left\|X_k\right\|_2}.$$

Open problem: Does Theorem 5.1 apply for some suitable r, M, β?

(7) Prove that if C and X_0 commute, provided that the equation $X_kZ + ZX_k = C$ has a unique solution for all k, then X_k and C commute for all k and Z is given by

$$Z = \frac{1}{2}X_k^{-1}C = \frac{1}{2}CX_k^{-1}.$$

Deduce that

$$X_{k+1} = \frac{1}{2}(X_k + X_k^{-1}C) = \frac{1}{2}(X_k + CX_k^{-1}).$$

This is the matrix analog of the formula given in Problem 5.1 of linalg-II (Vol II).

Prove that if C and X_0 have positive eigenvalues and C and X_0 commute, then X_{k+1} has positive eigenvalues for all $k \geq 0$ and thus the sequence (X_k) is defined.

Hint. Because X_k and C commute, X_k^{-1} and C commute, and obviously X_k and X_k^{-1} commute. By Proposition 22.15 of Vol. I, X_k, X_k^{-1}, and C are triangulable in a common basis, so there is some orthogonal matrix P and some upper-triangular matrices T_1, T_2 such that

$$X_k = PT_1P^\top, \quad X_k^{-1} = PT_1^{-1}P^\top, \quad C = PT_2P^\top.$$

It follows that

$$X_{k+1} = \frac{1}{2}P(T_1 + T_1^{-1}T_2)P^\top.$$

Also recall that the diagonal entries of an upper-triangular matrix are the eigenvalues of that matrix.

(8) **Extra Credit (100pts).** Prove that if X_0 and C commmute, are both diagonalizable, and have positive eigenvalues, then the sequence (X_k) given by

$$X_{k+1} = \frac{1}{2}(X_k + X_k^{-1}C) = \frac{1}{2}(X_k + CX_k^{-1})$$

converges to a square root of C.

(9) Implement the above method in `Matlab` (there is a command `kron(A, B)` to form the Kronecker product of A and B). Test your program on diagonalizable and nondiagonalizable matrices, including

$$W = \begin{pmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{pmatrix}, \quad A_1 = \begin{pmatrix} 5 & 4 & 1 & 1 \\ 4 & 5 & 1 & 1 \\ 1 & 1 & 4 & 2 \\ 1 & 1 & 2 & 4 \end{pmatrix},$$

and

$$A_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 0.01 & 0 & 0 \\ -1 & -1 & 100 & 100 \\ -1 & -1 & -100 & 100 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad A_4 = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 1 \end{pmatrix}.$$

What happens with

$$C = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad X_0 = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

The problem of determining when square roots of matrices exist and procedures for finding them are thoroughly investigated in Higham [1] (Chapter 6).

TOTAL: 300 points + 100 extra credit

References

[1] Nicholas J. Higham. *Functions of Matrices. Theory and Computation.* SIAM, first edition, 2008.