Fall, 2020   CIS 515

Fundamentals of Linear Algebra and Optimization
Jean Gallier

Homework 6

November 16, 2020; Due November 30, 2020

Problem B1 (30 pts). Let $\| \, \|$ be any operator norm. Given an invertible $n \times n$ real matrix A, if $c = 1/(2 \, \|A^{-1}\|)$, then for every $n \times n$ matrix H, prove that if $\|H\| \leq c$, then $A + H$ is invertible.

Furthermore, prove that if $\|H\| \leq c$, then $\|(A+H)^{-1}\| \leq 1/c$.

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 (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 $B \in \mathrm{M}_n(\mathbb{R})$.

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 B4 (150 pts). (1) Consider the determinant map, $f\colon \mathbf{M}_n(\mathbb{R}) \to \mathbb{R}$, given by

$$f(A) = \det(A), \quad A \in \mathbf{M}_n(\mathbb{R}).$$

For any matrix $B \in \mathrm{M}_n(\mathbb{R})$ (not necessarily invertible), let $\gamma\colon \mathbb{R} \to \mathbf{GL}(n, \mathbb{R})$ be the function given by

$$\gamma(t) = e^{tB}, \quad t \in \mathbb{R}.$$

Obviously, $\gamma(0) = I$. Geometrically, γ defines a curve in the group $\mathbf{GL}(n, \mathbb{R})$ passing through I at time $t = 0$. The function γ is differentiable, and by using the power series defining e^{tB} it is easily shown that

$$\gamma'(t) = Be^{tB},$$

so $\gamma'(0) = B$. In other words, the curve γ passes through I with velocity B. **You don't have to prove this fact** (Recall that when the domain space has dimension 1, we write $\gamma'(t) = d\gamma_1(t)$, the velocity vector at t.)

Let $g\colon \mathbb{R} \to \mathbb{R}$ be the function given by

$$g(t) = \det(\gamma(t)) = \det(e^{tB}), \quad t \in \mathbb{R}.$$

(1) Use the chain rule to prove that

$$d\det_I(B) = (\det \circ \gamma)'(0),$$

where $d\det_I$ is the derivative of the determinant function $\det\colon \mathrm{M}_n(\mathbb{R}) \to \mathbb{R}$ at I (the identity matrix).

(2) Prove that

$$d\det_I(B) = \mathrm{tr}(B),$$

the trace of B, for any matrix $B \in \mathrm{M}_n(\mathbb{R})$.

Hint. Use the fact that $\det(e^M) = e^{\mathrm{tr}(M)}$ for any matrix $M \in \mathrm{M}_n(\mathbb{R})$.

(3) Prove that

$$d\det_A(B) = \det(A)\mathrm{tr}(A^{-1}B),$$

for any $A \in \mathbf{GL}(n, \mathbb{R})$ and any matrix $B \in \mathrm{M}_n(\mathbb{R})$.

Hint. Find a curve $\gamma\colon \mathbb{R} \to \mathbf{GL}(n, \mathbb{R})$ such that $\gamma(0) = A$ and $\gamma'(0) = B$ and use the chain rule.

(4) Proposition 3.5 (Vol II) shows that for any continuous bilinear map $f\colon E_1 \times E_2 \to F$, for every $(a, b) \in E_1 \times E_2$, the derivative $\mathrm{D}f_{(a,b)}$ exists and is given by

$$\mathrm{D}f_{(a,b)}(u, v) = f(u, b) + f(a, v),$$

for all $(u, v) \in E_1 \times E_2$.

It can be shown (and you need not prove it, unless you decide to solve the extra credit problem) that for any continuous multilinear map $f\colon E_1 \times \cdots \times E_n \to F$, for any $(a_1, \ldots, a_n) \in$

$E_1 \times \cdots \times E_n$, the derivative $\mathrm{D}f_{(a_1,\ldots,a_n)}$ exists and is given by

$$\begin{aligned}\mathrm{D}f_{(a_1,\ldots,a_n)}(u_1,\ldots,u_n) &= f(u_1,a_2,a_3,\ldots,a_n) + f(a_1,u_2,a_3,\ldots,a_n) + \cdots \\ &\quad + f(a_1,a_2,a_3,\ldots,a_{n-1},u_n) \\ &= \sum_{k=1}^{n} f(a_1,\ldots,a_{k-1},u_k,a_{k+1},\ldots,a_n),\end{aligned}$$

for all $(u_1,\ldots,u_n) \in E_1 \times \cdots \times E_n$.

By definition, for every $a = (a_1,\ldots,a_n) \in E_1\times\cdots\times E_n$, the map $\mathrm{D}f_a$ is a continuous linear map from $E_1\times\cdots\times E_n$ to F, namely, $\mathrm{D}f_a \in \mathcal{L}(E_1\times\cdots\times E_n, F)$. The map $\mathrm{D}f \colon E_1\times\cdots\times E_n \to \mathcal{L}(E_1 \times \cdots \times E_n, F)$ given by $a \mapsto \mathrm{D}f_a$ is not multilinear, but it can be shown that it is continuous (you need not prove it, unless you decide to solve the extra credit problem).

Using the above facts, prove (quickly, this is easy) that for *any* matrix $A \in \mathrm{M}_n(\mathbb{R})$ and any matrix $B \in \mathrm{M}_n(\mathbb{R})$, the derivative $d\det_A$ exists and is given by

$$\begin{aligned}d\det_A(B) &= \det(B^1,A^2,A^3,\ldots,A^n) + \det(A^1,B^2,A^3,\ldots,A^n) + \cdots \\ &\quad + \det(A^1,A^2,A^3,\ldots,A^{n-1},B^n) \\ &= \sum_{k=1}^{n} \det(A^1,\ldots,A^{k-1},B^k,A^{k+1},\ldots,A^n),\end{aligned}$$

where $A^1,\ldots,A^n$ are the columns of A and $B^1,\ldots,B^n$ are the columns of B. Furthermore, the map $d\det \colon \mathrm{M}_n(\mathbb{R}) \to \mathcal{L}(\mathrm{M}_n(\mathbb{R}),\mathbb{R})$ given by $A \mapsto d\det_A$ is continuous.

Therefore, $d\det_A$ exists even if A is not invertible, but we would like to find a more "friendly" and more explicit expression for it. There such an explicit formula involving the adjugate matrix $\widetilde{A}$ of A from Section 6.4, Definition 6.9.

(5) (**Extra Credit 40 pts**) Prove that for any continuous multilinear map $f \colon E_1 \times \cdots \times E_n \to F$, for any $a = (a_1,\ldots,a_n) \in E_1 \times \cdots \times E_n$, the derivative $\mathrm{D}f_{(a_1,\ldots,a_n)}$ exists and is given by

$$\begin{aligned}\mathrm{D}f_{(a_1,\ldots,a_n)}(u_1,\ldots,u_n) &= f(u_1,a_2,a_3,\ldots,a_n) + f(a_1,u_2,a_3,\ldots,a_n) + \cdots \\ &\quad + f(a_1,a_2,a_3,\ldots,a_{n-1},u_n) \\ &= \sum_{k=1}^{n} f(a_1,\ldots,a_{k-1},u_k,a_{k+1},\ldots,a_n),\end{aligned}$$

for all $u = (u_1,\ldots,u_n) \in E_1 \times \cdots \times E_n$.

Hint. Generalize the proof of Proposition 3.5 (Vol II).

Prove that $\mathrm{D}f$ (a map from $E_1 \times \cdots \times E_n$ to $\mathcal{L}(E_1 \times \cdots \times E_n, F)$) is continuous.

Hint. To prove that $\mathrm{D}f$ is continuous, first observe that $\mathrm{D}f$ is the sum of the n functions $(\mathrm{D}f)^1, \ldots, (\mathrm{D}f)^n$, with $(\mathrm{D}f)^k$ from $E_1 \times \cdots \times E_n$ to $\mathcal{L}(E_1 \times \cdots \times E_n, F)$ given by

$$(\mathrm{D}f)^k_{(a_1,\ldots,a_n)}(u_1, \ldots, u_n) = f(a_1, \ldots, a_{k-1}, u_k, a_{k+1}, \ldots, a_n).$$

The function $(\mathrm{D}f)^k$ is independent of the variable u_k, so it is *not* multilinear, but its restriction to $E_1 \times \cdots \times E_{k-1} \times E_{k+1} \times \cdots \times E_n$ is $(n-1)$-multilinear, so if we can show that this restriction is continuous, then $(\mathrm{D}f)^k$ itself will be continuous. To simplify notation, write $\mathcal{E}_k = E_1 \times \cdots \times E_{k-1} \times E_{k+1} \times \cdots \times E_n$. We also use the notation $(\mathrm{D}f)^k$ to denote the restriction of $(\mathrm{D}f)^k$ to $\mathcal{E}_k$.

Show that the operator norm $\left\|(\mathrm{D}f)^k\right\|$ of the restriction of $(\mathrm{D}f)^k$ to $\mathcal{E}_k$ satisfies the inequality

$$\left\|(\mathrm{D}f)^k\right\| \leq \|f\|,$$

where $\|f\|$ is the norm of the multilinear map f (for norms of linear and multilinear maps, see Section 2.6, Vol. II).

(6) Prove that for *any* matrix $A \in \mathrm{M}_n(\mathbb{R})$, not necessarily invertible, there is a convergent sequence $(A_k)_{k\geq 1}$ of *invertible* matrices $A_k \in \mathrm{GL}(n, \mathbb{R})$ whose limit is A. To prove this, it is convenient to use the Frobenius norm or the operator 2-norm (the spectral norm). You need to construct a sequence of invertible matrices A_k such that

$$\lim_{k\mapsto\infty} \|A - A_k\| = 0.$$

Hint. Use a convenient factorization of A.

(7) Recall the definition of the *adjugate matrix* $\widetilde{A}$ of an $n \times n$ matrix A and the fact that if A is invertible, then by Proposition 6.7 (see Vol I),

$$A^{-1} = (\det(A))^{-1}\widetilde{A}.$$

Using the above, (3) is rewritten as

$$d\det{}_A(B) = \mathrm{tr}(\widetilde{A}B),$$

for any $A \in \mathbf{GL}(n, \mathbb{R})$ and any matrix $B \in \mathrm{M}_n(\mathbb{R})$. Use (6) to prove that

$$d\det{}_A(B) = \mathrm{tr}(\widetilde{A}B),$$

for *any* matrix $A \in \mathrm{M}_n(\mathbb{R})$ (not necessarily invertible) and any matrix $B \in \mathrm{M}_n(\mathbb{R})$.

(8) Let $\mathrm{GL}^+(n, \mathbb{R})$ be the subgroup of $\mathrm{GL}(n, \mathbb{R})$ consisting of all matrices A such that $\det(A) > 0$. It can be shown that this subgroup is open in $\mathrm{M}_n(\mathbb{R})$. Consider the function $\ell\colon \mathrm{GL}^+(n, \mathbb{R}) \to \mathbb{R}$ given by

$$\ell(A) = \log\det(A).$$

Prove that

$$d\ell_A(B) = \mathrm{tr}(A^{-1}B)$$

for all $A \in \mathrm{GL}^+(n, \mathbb{R})$ and all $B \in \mathrm{M}_n(\mathbb{R})$.

Remark: The function $\log\det$ is a *barrier* function used in convex optimization.

(9) Let J be the $(n+1) \times (n+1)$ diagonal matrix

$$J = \begin{pmatrix} I_n & 0 \\ 0 & -1 \end{pmatrix}.$$

We denote by $\mathbf{SO}(n, 1)$ the group of real $(n+1) \times (n+1)$ matrices

$$\mathbf{SO}(n, 1) = \{A \in \mathbf{GL}(n+1, \mathbb{R}) \mid A^\top J A = J \quad \text{and} \quad \det(A) = 1\}.$$

Check that $\mathbf{SO}(n, 1)$ is indeed a group with the inverse of A given by $A^{-1} = JA^\top J$ (this is the *special Lorentz group*).

(10) Consider the function $h\colon \mathbf{GL}^+(n+1) \to \mathbf{S}(n+1)$, given by

$$h(A) = A^\top J A - J,$$

where $\mathbf{S}(n+1)$ denotes the space of $(n+1) \times (n+1)$ symmetric matrices. Prove that

$$dh_A(H) = A^\top J H + H^\top J A$$

for any matrix $H \in \mathrm{M}_{n+1}(\mathbb{R})$.

Prove that dh_A is surjective for all $A \in \mathbf{SO}(n, 1)$.

Remark: Parts (9) and (10) can be used to prove that $\mathbf{SO}(n, 1)$ is a smooth manifold of dimension $\frac{n(n+1)}{2}$.

Problem B5 (20). Let A be an $n \times n$ real symmetric matrix, B an $n \times n$ symmetric positive definite matrix, and let $b \in \mathbb{R}^n$.

Prove that a necessary condition for the function J given by

$$J(v) = \frac{1}{2} v^\top A v - b^\top v$$

to have an extremum in $u \in U$, with U defined by

$$U = \{v \in \mathbb{R}^n \mid v^\top B v = 1\},$$

is that there is some $\lambda \in \mathbb{R}$ such that

$$Au - b = \lambda B u.$$

Hint. Express the definition of U as

$$U = \{v \in \mathbb{R}^n \mid \varphi(v) = 0\},$$

with

$$\varphi(v) = \frac{1}{2} - \frac{1}{2} v^\top B v.$$

Extra credit (20 points). Prove that there is a symmetric positive definite matrix S such that $B = S^2$. Prove that if $b = 0$, then λ is an eigenvalue of the symmetric matrix $S^{-1} A S^{-1}$.

Remark: If $b \neq 0$, solving for λ is a lot harder.

TOTAL: 340 points + (60 extra credit)