Module 10

Metric Spaces, Continuity, and Differentiation
Solutions

Problem 1: 10 points

Let $f\colon \mathrm{M}_n(\mathbb{R}) \to \mathrm{M}_n(\mathbb{R})$ be the function defined on $n \times n$ matrices by

$$f(A) = A^2.$$

Prove that

$$\mathrm{D}f_A(H) = AH + HA,$$

for all $A, H \in \mathrm{M}_n(\mathbb{R})$.

Solution.
We have

$$\begin{aligned} f(A+H) - f(A) &= (A+H)^2 - A^2 \\ &= A^2 + AH + HA + H^2 - A^2 \\ &= AH + HA + H^2. \end{aligned}$$

If we write

$$\epsilon(H) = \frac{H^2}{\|H\|},$$

where $\| \ \|$ is any matrix norm, say the Frobenius norm, then

$$\|\epsilon(H)\| = \frac{\|H^2\|}{\|H\|} \leq \frac{\|H\|^2}{\|H\|} = \|H\|,$$

and so $\lim_{\|H\| \mapsto 0} \|\epsilon(H)\| = 0$, which proves that

$$\mathrm{D}f_A(H) = AH + HA.$$

Problem 2: 30 points total

Recall that $\mathfrak{so}(3)$ denotes the vector space of real skew-symmetric $n \times n$ matrices ($B^\top = -B$). Let $C\colon \mathfrak{so}(n) \to \mathrm{M}_n(\mathbb{R})$ be the function given by

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

(1) (5 points) Prove that if B is skew-symmetric, then $I - B$ and $I + B$ are invertible, and so C is well-defined. *Hint.* Consider the eigenvalues of B.

(2) (20 points) Prove that

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

Hint. Use the product rule.

(3) (5 points) Prove that $dC(B)$ is injective for every skew-symmetric matrix B.

Solution. (1) By the spectral theorem for skew-symmetric matrices (Week 7, Lesson 8, normal and other special matrices), we know that the eigenvalues of a (real) skew symmetric matrix B are of the form $i\mu$, with $\mu \in \mathbb{R}$. If $i\mu_1, \ldots, i\mu_n$ are the eigenvalues of B, then $1 + i\mu_1, \ldots, 1 + i\mu_n$ are the eigenvalues of $I + B$ (since if u is an eigenvector of B for $i\mu_j$, from $Bu = i\mu_j u$ we get

$$(I + B)u = u + Bu = u + i\mu_j u = (1 + i\mu_j)u,$$

so u is an eigenvector of $I + B$ for $1 + i\mu_j$). Similalry, $1 - i\mu_1, \ldots, 1 - i\mu_n$ are the eigenvalues of $I - B$. But since $\mu_j \in \mathbb{R}$, we have $1 + i\mu_j \neq 0$ and $1 - i\mu_j \neq 0$ for $j = 1, \ldots, n$, which implies that $I + B$ and $I - B$ are invertible.

(2) We will use the product rule (see Lesson 6 of Week 10),

$$d(fg)_A(B) = df_A(B)g(A) + f(A)dg_A(B),$$

for all $A, B \in \mathrm{M}_n(\mathbb{R})$, where $f\colon \mathrm{M}_n(\mathbb{R}) \to \mathrm{M}_n(\mathbb{R})$ and $g\colon \mathrm{M}_n(\mathbb{R}) \to \mathrm{M}_n(\mathbb{R})$ are differentiable matrix functions.

Let f and g be the maps defined on $n \times n$ matrices by $f(B) = I - B$ and $g(B) = I + B$. We would like to compute the derivative $d(fg^{-1})(B)$ of $C = fg^{-1}$ on the vector space of skew-symmetric matrices. We will use the product rule and the chain rule.

First we claim that

$$\begin{aligned} df(B) &= -\mathrm{id} \\ dg(B) &= \mathrm{id}, \end{aligned}$$

for all matrices B. Indeed

$$f(B + H) - f(B) + \mathrm{id}(H) = I - (B + H) - (I - B) + H = 0$$

and

$$g(B + H) - g(B) - \mathrm{id}(H) = I + (B + H) - (I + B) - H = 0,$$

which proves our claim.

Let h be the matrix inverse function, namely,

$$h(B) = B^{-1}.$$

The map h is defined on $\mathbf{GL}(n, \mathbb{R})$, an open subset of $\mathbb{R}^{n^2}$, and we proved in Lesson 6 of Week 10 that

$$dh(B)(H) = -B^{-1}HB^{-1}.$$

Now, $g^{-1}(B) = (h \circ g)(B)$, so by the chain rule

$$dg^{-1}(B) = dh(g(B)) \circ dg(B)$$

and since $dg(B) = \mathrm{id}$ and $g(B) = I + B$, we get

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

for all skew-symmetric matrices B and for all H.

By the product rule,

$$dfg^{-1}(B)(H) = df(B)(H)g^{-1}(B) + f(B)dg^{-1}(B)(H),$$

and since $df(B)(H) = -H$, $f(B) = I - B$, and $g(B) = I + B$, we get

$$\begin{aligned}
dfg^{-1}(B)(H) &= -H(I + B)^{-1} + (I - B)\big(-(I + B)^{-1}H(I + B)^{-1}\big) \\
&= -[I + (I - B)(I + B)^{-1}]H(I + B)^{-1} \\
&= -[(I + B)(I + B)^{-1} + (I - B)(I + B)^{-1}]H(I + B)^{-1} \\
&= -[(I + B + I - B)(I + B)^{-1}]H(I + B)^{-1} \\
&= -2(I + B)^{-1}H(I + B)^{-1}
\end{aligned}$$

for all skew-symmetric matrices B and for all H, as claimed.

(3) Because B is a skew-symmetric matrix, we proved in (1) that $I + B$ is invertible. To prove that the linear map $dC(B)$ is given by

$$dC(B)(H) = -2(I + B)^{-1}H(I + B)^{-1}$$

is injective we show that its kernel is reduced to 0. Since $I + B$ is the inverse of $(I + B)^{-1}$, if

$$dC(B)(H) = -2(I + B)^{-1}H(I + B)^{-1} = 0,$$

then multiplying on the left and on the right by $I + B$, we get $H = 0$, which shows that the kernel of $dC(B)$ is indeed trivial.

Problem 3: 10 points

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

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

Solution. We would like to apply the theorem from Week 10, Lesson 6, Lagrange multipliers, to

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

and

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

with

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

The reason for the factor $1/2$ is that we obtain

$$d\varphi_v(h) = -h^\top B v,$$

instead of $-h^\top 2Bv$.

In order for the theorem to apply we need to check that the linear form $d\varphi_u$ is not the zero linear form for every $u \in U$. But

$$d\varphi_v(h) = -h^\top B v,$$

and by definition of U, we have $v \neq 0$ for all $v \in U$ (since $v^\top B v = 1$). Since B is symmetric positive definite, B is invertible and so $Bv \neq 0$ if $v \neq 0$. It follows that $d\varphi_v \neq 0$ for all $v \in U$, as desired.

The Lagrangian of this minimization problem is given by

$$L(v, \lambda) = J(v) + \lambda \varphi(v) = \frac{1}{2} v^\top A v - b^\top v + \lambda \left(\frac{1}{2} - \frac{1}{2} v^\top B v \right),$$

with $\lambda \in \mathbb{R}$. A necessary condition for J to have an extremum at $u \in U$ is that

$$dJ_u + \lambda d\varphi_u = 0.$$

Since

$$dJ_u(h) = h^\top (Au - b), \quad d\varphi_u(h) = -h^\top Bv,$$

we must have

$$h^\top (Au - b - \lambda Bu) = 0 \quad \text{for all } h \in \mathbb{R}^n,$$

so we must have

$$Au - b - \lambda Bu = 0,$$

that is,

$$Au - b = \lambda Bu.$$

Total: 50 points