MAA 4212—Matrices, Power Series, and Functions of Matrices

Before the discussion gets serious, make sure you understand the difference between a linear transformation (a function) and a matrix (a bunch of numbers arranged in a rectangular array). A matrix can be used for many things, one of which is to **represent** a linear transformation (see exercise (1c) below), but the matrix does not **equal** the linear transformation. In the 1×1 case, the function $T(x) = 3x$ is a linear transformation, which, once we have agreed we're talking about linear transformations, can be represented simply by the number (1×1 matrix) 3. This representation distinguishes T from any other linear transformation, but there is still a difference between the number 3 and the function $x \mapsto 3x$.

Exercises.

1. Let $T : \mathbf{R}^n \to \mathbf{R}^m$ be a linear transformation.

(a) Prove that T is differentiable (and therefore continuous).

(b) Prove that "a linear transformation is its own derivative". More precisely (the statement in quotes is literal nonsense since the derivative of a function is a function of twice as many variables), prove that for all $\mathbf{a}, \mathbf{v} \in \mathbf{R}^n$,

$$(D_{\mathbf{a}}T)(\mathbf{v}) = T(\mathbf{v}).$$

(c) Let the matrix A represent T (i.e. $T(\mathbf{v}) = A\mathbf{v}$, where on the right-hand side we agree that elements of $\mathbf{R}^n$ are to be written as column vectors, i.e. $n \times 1$ matrices, and that $A\mathbf{v}$ means matrix multiplication of the $m \times n$ matrix by the $n \times 1$ matrix $\mathbf{v}$). Show that for all $\mathbf{a} \in \mathbf{R}^n$, the same matrix A represents $D_{\mathbf{a}}T$.

2. Reconcile problem (1b) with the following vague statement: "the exponential function $\mathbf{R} \to \mathbf{R}$ is its own derivative".

For the rest of these notes, let V_n denote the set of $n \times n$ matrices whose entries are real numbers. Recall that, under matrix addition, and multiplication of a matrix by a scalar (= real number), V_n becomes a vector space, whose zero element is the matrix all of whose entries are 0. The dimension of V_n is not n, but n^2, the number of entries. Be careful when reading the rest of these notes that you do not confuse the vector space V_n (which is isomorphic to $\mathbf{R}^{n^2}$) with the usually smaller space $\mathbf{R}^n$.

Recall the result of homework problem p. 93/23 (part of HW 1), which can be rephrased this way: for any two norms $\| \; \|_1, \| \; \|_2$, there exist positive constants c_1, c_2 so that for all $\mathbf{v} \in \mathbf{R}^n$, $\|\mathbf{v}\|_1 \leq c_1\|\mathbf{v}\|_2$, and $\|\mathbf{v}\|_2 \leq c_1\|\mathbf{v}\|_1$. In case you forgot to do that problem, here it is again, with a hint:

Exercise.

3. Let $n \in \mathbf{N}$. Prove that all norms on $\mathbf{R}^n$ are equivalent. (*Hint.* Show that "$\| \|_1$ is equivalent to $\| \|_2$" is an equivalence relation. Hence it suffices to show that any norm is equivalent to the Euclidean norm, $\| \|_{euc}$ Let S be the "unit sphere" $\{\mathbf{x} \in \mathbf{R}^n \mid \|\mathbf{x}\|_{euc} = 1\}$. Show S is compact. Then, given any norm $\| \|$, consider the function $f : S \to \mathbf{R}$ given by $f(\mathbf{x}) = \|\mathbf{x}\|/\|\mathbf{x}\|_{euc}$.)

The importance of the fact that all norms on $\mathbf{R}^n$ are equivalent is that it implies the following: a function $f : \mathbf{R}^n \to \mathbf{R}$ is continuous with respect to the metric determined by one norm iff it is continuous with respect to the metric determined by the other norm. (The same goes for functions $F : (U \subset \mathbf{R}^n) \to E$, or functions $f : E \to \mathbf{R}^n$ for any fixed metric space E.) Thus, for functions on a finite-dimensional vector space, the notion of which functions are continuous does not depend on which norm you choose. Below, you will show that the notion of "differentiable" also doesn't depend on which norm you choose.

Exercise.

4. Prove that the definition of differentiability of a function $f : U \to \mathbf{R}$, where U is an open subset of $\mathbf{R}^n$, does not depend on which norm is used. I.e. prove that if you use one norm, the functions that you end up calling differentiable are exactly the same functions you'd have called differentiable using any other norm. Generalize this fact to $\mathbf{R}^m$-valued functions.

The reason exercise 4 is included in this set of notes is that for functions on V_n, there is a more convenient norm to use than the Euclidean norm ($\|A\|_{euc} = (\sum_{i,j} A_{ij}^2)^{1/2}$, called by some the Frobenius norm).

Definition. The *operator norm* on V_n is defined by

$$\|A\|_{op} = \sup_{\mathbf{x} \in S}(\|A\mathbf{x}\|_{euc}),$$

where S is the unit sphere as in exercise 3 above.

Exercises.

5. Prove that $\| \|_{op} : V_n \to \mathbf{R}$ is, in fact, a norm.

6. Let $A, B \in V_n$.

(a) Prove that

$$\|A\|_{op} = \sup_{\mathbf{x} \neq \mathbf{0}} \frac{\|A\mathbf{x}\|_{euc}}{\|\mathbf{x}\|_{euc}}.$$

(b) Prove that for all $\mathbf{x} \in \mathbf{R}^n$, $\|A\mathbf{x}\|_{euc} \leq \|A\|_{op}\|\mathbf{x}\|_{euc}$.

(c) Prove that $\|AB\|_{op} \leq \|A\|_{op}\|B\|_{op}$. (It is this property that makes the operator norm more convenient than the Euclidean norm.)

7. For $A \in V_n$, define $L_A : V_n \to V_n$ and $R_A : V_n \to V_n$ by $L_A(B) = AB,\ R_A(B) = BA$ (the "L" and "R" stand for "left" and "right"). Check that, for all A, the maps L_A and R_A are linear, and find the directional derivatives

$$(D_B L_A)(C), \quad (D_B R_A)(C).$$

Make sure you understand that A is not "the matrix of L_A" (or the matrix of R_A). Given a linear transformation T of an m-dimensional vector space V to itself, *and a basis* $\{v_j\}$, one defines the matrix of T *with respect to that basis* using the coefficients that are needed to express $T(v_i)$, for each i, as a linear combinations of the $\{v_j\}$. For $V = V_n$, the dimension is n^2. Were we to choose a basis for V (the most obvious one being $\{\mathbf{e}_{i,j}\}_{i,j=1}^n$, where $\mathbf{e}_{i,j}$ is the matrix whose $(i,j)^{\text{th}}$ entry is 1 and whose other entries are all 0), the matrices of L_A and R_A with respect to that basis would be $n^2 \times n^2$ matrices, not $n \times n$ matrices, and the matrices of L_A and R_A would be different from each other. You should work out these matrices by hand for the 2×2 case to make sure you understand what I just said.

******

Before proceeding further, you should do the problem p. 162/17, which I forgot to assign earlier. You're going to be applying the results of this problem to $V = V_n$.

******

The fact that products and sums of $n \times n$ matrices are again $n \times n$ matrices enables us to make sense out of polynomials and power series whose variables are matrices. For example, if $p(x) = x^3 + 4x + 4$, we can define an analogous function $\tilde{p} : V_n \to V_n$ by

$$\tilde{p}(A) = A^3 + 4A + 4I,$$

where I is the $n \times n$ identity matrix, and where $A^3 = AAA$. More generally, given any power series $\sum c_k x^k$, we can consider the power series $\sum c_k A^k$, where A is a V_n-valued variable; by convention we define $A^0 = I$ (the identity matrix again). If the real-valued power series converges to a function $f(x)$, and the matrix-valued power series converges, it is customary to denote the value of the matrix-valued series by $f(A)$ (i.e. not to bother with the tilde used in the example above).

Exercises.

8. (a) Define $s : V_n \to V_n$ by $s(A) = A^2$. Compute the directional derivatives $(D_A s)(B)$ (warning: remember that matrix multiplication is non-commutative), and prove that s is differentiable. Generalize to higher exponents.

(b) Let $p : \mathbf{R} \to \mathbf{R}$ be any polynomial function. Prove that the associated function $p : V_n \to V_n$ is differentiable.

9. (a) Suppose the real-valued series $\sum c_k x^k$ has positive radius of convergence ρ. Prove that for any n the associated matrix-valued series has radius of convergence at least ρ (i.e. that $\sum c_k A^k$ converges if $\|A\|_{op} < \rho$). (I'm not using our usual "R" for radius of convergence because I've already used "R" in "R_A".)

(b) Prove that if $\|B\|_{op} < 1$, then (i) $\sum_{k=0}^{\infty}(-1)^k B^k$ converges to some function $f(B)$, (ii) $(I+B)f(B) = I$, and hence (iii) $I+B$ is invertible and $(I+B)^{-1}$ equals the infinite series in (i).

10. Let $V_{n,*} \subset V_n$ denote the subset of invertible matrices, and let $\iota : V_{n,*} \to V_{n,*}$ denote the inversion map ($\iota(A) = A^{-1}$).

(a) Prove that $(D_I\, \iota)(B) = -B$.

(b) More generally, if $A \in V_{n,*}$, prove that

$$(D_A \iota)(B) = -A^{-1}BA^{-1}.$$

(Hint: $(A+tB)^{-1} = A^{-1}(I+tBA^{-1})^{-1}$.) How does this fact fit in with the one-dimensional case?

(c) Prove that ι is differentiable.

Definition. The *Hessian* of a function $f : \mathbf{R}^n \to \mathbf{R}^m$ at a point $\mathbf{a} \in \mathbf{R}^n$ is the function $H_{\mathbf{a}}f : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^m$ defined by

$$(H_{\mathbf{a}}f)(\mathbf{v}, \mathbf{w}) = D_{\mathbf{a}}(\mathbf{x} \mapsto (D_{\mathbf{x}}f)(\mathbf{v}))(\mathbf{w}),$$

provided all these directional second derivatives exist. (If the formula above is confusing, this is what it says: Fix a vector $\mathbf{v}$. The directional derivative $(D_{\mathbf{x}}f)(\mathbf{v})$ is then a function of the base point $\mathbf{x}$. Take the directional derivative of this new function in the direction $\mathbf{w}$, at the point $\mathbf{a}$. The result is defined to be $H_{\mathbf{a}}f(\mathbf{v}, \mathbf{w})$. If we were to restrict $\mathbf{v}, \mathbf{w}$ to be the standard basis unit vectors, the Hessian would just be the collection of second partials.)

(d) Prove that $H_A \iota$ exists for all $A \in V_{n,*}$, and compute $(H_A \iota)(B, C)$. As a check on your answer, see what your formula reduces to when $n = 1$.

11. (a) For $A \in V_n$, define

$$\exp(A) = e^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k$$

and explain why this sum converges for all A.

(b) Prove that *if A and B commute* (i.e. if $AB = BA$), then $e^{A+B} = e^A e^B$ (hint: homework problem D3(e)), but that if A and B do not commute this equation may fail.

(c) Compute $(D_A \exp)(B)$ (for arbitrary A and B). As a check, see what your answer reduces to in the 1-dimensional case.

(d) Let $n = 2$, and let $J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Check that $J^2 = -I$. Using this fact to simplify your work, compute e^{tJ} and e^{xI+yJ}, where $t, x, y \in \mathbf{R}$. Then say something deep.

(e) Let $t, a, b, c \in \mathbf{R}$, and let $A = \begin{pmatrix} 0 & t \\ 0 & 0 \end{pmatrix}$, $B = \begin{pmatrix} 0 & a & b \\ 0 & 0 & c \\ 0 & 0 & 0 \end{pmatrix}$. Compute e^A and e^B. Speculate about how much computation is involved in exponentiating strictly upper-triangular $n \times n$ matrices for general n.

(f) Define $C = \begin{pmatrix} 0 & a & b \\ -a & 0 & c \\ -b & -c & 0 \end{pmatrix}$. Check that $C^3 = -\Delta^2 C$, where $\Delta = \sqrt{a^2 + b^2 + c^2}$. Using this to simplify your work, compute e^C. In your computation, you should see some familiar-looking series of real numbers coming up. Replace these familiar series by the functions they converge to, so that your final answer has no infinite series left in it.

(g) Using part (e) above and exercise 9(a) as your guide, name at least two conditions on a nonzero matrix A, either of which guarantees that the series

$$\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} A^k$$

converges. When it converges, call the sum $\log(I + A)$, since that's what it reduces to in the 1×1 case. For the specific matrix A in part (e), compute $\log(I + A)$, $\exp(\log(I + A))$, and $\log(e^A)$.