9. Calculus in vector spaces

Throughout, let X be a finite dimensional Euclidean space.

Let $x(t)$ be a vector-valued function, that is, $x: \mathbb{R} \longrightarrow X$.

Def: $x(t)$ is continuous at $t_0 \in \mathbb{R}$ if $\lim_{t \to t_0} \| x(t) - x(t_0) \| = 0$,
and it is differentiable at t_0 with derivative $\dot{x}(t_0)$ if

$$\lim_{h \to 0} \left\| \frac{x(t_0+h) - x(t_0)}{h} - \dot{x}(t_0) \right\| = 0.$$

Note: we usually write $\dot{x}(t)$ or $\dot{x}$ for $\frac{d}{dt} x(t)$.

We can define continuity and differentiability of matrix-valued functions $M(t)$, that is, $M: \mathbb{R} \longrightarrow \mathcal{L}(X, X)$.

Remark The domain can be e.g., $(0,1)$ instead of $\mathbb{R}$.

Basic facts:

(i) $\frac{d}{dt}$ is linear: $\frac{d}{dt}\left(a\,x(t) + b\,y(t)\right) = a\,\dot{x}(t) + b\,\dot{y}(t)$

(ii) If A is independent of t, then $\frac{d}{dt}\left(A\,B(t)\right) = A\,\dot{B}(t)$.

(iii) If If $l: \mathbb{R}^n \longrightarrow \mathbb{R}$ is linear, and $x(t)$ differentiable, then

$\frac{d}{dt}\, l(x(t)) = l(\dot{x}(t))$. An analogous result holds for matrix-valued functions.

Corollary: $\frac{d}{dt} \operatorname{tr}(A(t)) = \operatorname{tr}(\dot{A}(t))$ (since tr A is linear).

We have 5 product rules!

Suppose we have functions $k: \mathbb{R} \to \mathbb{R}$ (scalar), $x: \mathbb{R} \to X$ (vector-valued), and $\mathbb{R} \to \mathcal{L}(X, X)$ (matrix-valued). Then

(i) $\frac{d}{dt}[k(t)\, x(t)] = \frac{dk}{dt} x(t) + k(t)\, \dot{x}(t)$

(ii) $\frac{d}{dt}[A(t)\, x(t)] = \dot{A}(t) \cdot x(t) + A(t)\, \dot{x}(t).$

(iii) $\frac{d}{dt}[A(t)\, B(t)] = \dot{A}(t)\, B(t) + A(t)\, \dot{B}(t).$

(iv) $\frac{d}{dt}[k(t)\, A(t)] = \frac{dk}{dt} A(t) + k(t)\, \dot{A}(t)$

(v) $\frac{d}{dt}(y(t), x(t)) = (\dot{y}(t), x(t)) + (y(t), \dot{x}(t)).$

Theorem 9.1: Let $A(t)$ be a differentiable, invertible matrix-valued function. Then $A^{-1}(t)$ is differentiable, and $\frac{d}{dt} A^{-1}(t) = -A^{-1} \dot{A} A^{-1}$

Proof: Observe that

$$A^{-1}(t+h) - A^{-1}(t) = A^{-1}(t+h)[I - A(t+h) A^{-1}(t)] = A^{-1}(t+h)[A(t) - A(t+h)] A^{-1}(t).$$

Now divide through by h and take the limit as $h \to 0$. $\square$

Recall calculus (Chain rule): $\frac{d}{dt} f(g(t)) = f'(g(t))\, g'(t)$.

This fails for matrix-valued functions:

Example: Let $f(a) = a^2$. Compute $\frac{d}{dt} f(A(t)) = \frac{d}{dt}(A(t))^2$.

By the product rule, $\frac{d}{dt} A^2 = A\dot{A} + \dot{A}A \neq 2A\dot{A}$

Prop: $\frac{d}{dt} A^k = \dot{A}A^{k-1} + A\dot{A}A^{k-2} + \dots + A^{k-1}\dot{A}$. $\qquad (*)$

Pf Exercise (use induction).

Theorem 9.2: Let $p(t)$ be a polynomial and $A(t)$ a differentiable square matrix-valued function.

(a) If $A(t)$ and $\dot{A}(t)$ commute, then $\frac{d}{dt} p(A) = p'(A)\,\dot{A}$.

(b) $\frac{d}{dt} \operatorname{tr} p(A) = \operatorname{tr}(p'(A)\,\dot{A})$ (even if $A(t)$, $\dot{A}(t)$ don't commute)

Proof: If $A\dot{A} = \dot{A}A$, then $(*)$ reduces to $\frac{d}{dt} A^k = k A^{k-1}\dot{A}$,

so (a) follows by linearity. ✓

Recall that $\operatorname{tr}(AB) = \operatorname{tr}(BA)$. Thus

$$\operatorname{tr}\left(\frac{d}{dt} A^k\right) = \operatorname{tr}\left(\sum_{i=0}^{k-1} A^i \dot{A} A^{k-i-1}\right) = \operatorname{tr}\left(\sum_{i=0}^{k-1} A^{k-1}\dot{A}\right),$$

so (b) follows from linearity. $\square$

4

<u>Theorem 9.3</u> (Product rule for multilinear functions):

Suppose $x_1(t), \dots, x_k(t)$ are differentiable vector-valued functions, and $M(a_1, \dots, a_k)$ multilinear. Then $M(x_1, \dots, x_k)$ is differentiable, with

$$\frac{d}{dt} M(x_1, \dots, x_k) = M(\dot{x}_1, x_2, \dots, x_k) + \dots + M(x_1, \dots, x_{k-1}, \dot{x}_k).$$

<u>Proof</u>: By multilinearity,

$$M(x_1(t+h), \dots, x_k(t+h)) - M(x_1(t), \dots, x_k(t))$$

$$= M(x_1(t+h) - x_1(t), \dots, x_k(t+h)) + \dots + M(x_1(t), \dots, x_k(t+h) - x_k(t)).$$

Divide by h and let $h \to 0$. $\square$

<u>Remark</u>: If D is the determinant of a linear map, then $\frac{d}{dt} D(x_1, \dots, x_n) = D(\dot{x}_1, x_2, \dots, x_n) + \dots + D(x_1, \dots, x_{n-1}, \dot{x}_n)$.

Now, suppose that $X(t) = (x_1(t), \dots, x_n(t))$ is a matrix with $X(0) = I$, (so, $x_i(0) = e_i$).

Note that

$$D(\dot{x}_1(0), e_2, \dots, e_n) = \dot{x}_{11}(0)$$

$$D(e_1, \dot{x}_2(0), e_3, \dots, e_n) = \dot{x}_{22}(0)$$

$$\vdots$$

$$D(e_1, \dots, e_{n-1}, \dot{x}_n(0)) = \dot{x}_{nn}(0)$$

Thus, if $X(t)$ is differentiable and $X(0) = I$, then

$$\frac{d}{dt} \det X(t) \Big|_{t=0} = \operatorname{tr} \dot{X}(0).$$

Now, we'll relax the condition that $X(0) = I$.

If $Y(t)$ is differentiable and invertible, then define

$$X(t) := Y(0)^{-1} Y(t) \implies X(0) = Y(0)^{-1} Y(0) = I.$$

Note that $X(0) = I$, and $\det X(t) = (\det Y(0))^{-1} \det Y(t)$.

Now, take $\frac{d}{dt}$ of both sides:

$$(\det Y(0))^{-1} \frac{d}{dt} (\det Y(t)) \Big|_{t=0} = \frac{d}{dt} X(t) \Big|_{t=0} = \operatorname{tr} \dot{X}(0) = \operatorname{tr}\left[Y(0)^{-1} \dot{Y}(0)\right]$$

Note that the LHS is just $\frac{d}{dt} \left(\log \det Y(t) \right) \Big|_{t=0}$.

However, $t=0$ is arbitrary (just do a change of variables).

Thus, we have proven:

<u>Theorem 9.4</u>: Let $Y(t)$ be a differentiable square matrix-valued function. Then for any t for which $Y(t)$ is invertible,

$$\frac{d}{dt} \left(\log \det Y \right) = \operatorname{tr}\left(Y^{-1} \dot{Y} \right).$$

<u>Question</u>: when and how can we define $f(A)$, for a (non-polynomial) analytic function $f(t)$?

<u>Example:</u> Consider $f(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!}$ (This is just e^x).

We claim that this "works": $f(A) = \sum_{k=0}^{\infty} \frac{A^k}{k!}$.

We must show convergence, that is, the difference between the partial sums tends to zero.

Let $e_m(A) = \sum_{k=0}^{m} \frac{A^k}{k!}$ (m^{th} partial sum)

$$\Rightarrow e_m(A) - e_\ell(A) = \sum_{k=\ell+1}^{m} \frac{A^k}{k!}.$$

Note that $\| e_m(A) - e_\ell(A) \| \leq \sum_{k=\ell+1}^{m} \frac{1}{k!} \|A\|^k \to 0$ as $k \to 0$.

However, the matrix exponential doesn't share every property of the scalar exponential function.

For example, if $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and $B = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, then

$A^2 = B^2 = 0$, so

$$e^A = I + A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad e^B = I + B = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

so, $e^A e^B = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ and $e^B e^A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$

Put $S = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = A + B$. $S^2 = I$, so $S^{2n} = S$, $S^{2n+1} = I$.

$$\Rightarrow e^{A+B} = \sum_{k=0}^{\infty} \frac{1}{k!} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}^k = \sum_{k=0}^{\infty} \frac{1}{(2k)!} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \sum_{k=0}^{\infty} \frac{1}{(2k+1)!} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

$$= \begin{pmatrix} \cosh x & \sinh x \\ \sinh x & \cosh x \end{pmatrix}$$

Theorem 9.5: Let A, B be square matrices.

(a) If $AB = BA$, then $e^{A+B} = e^A e^B$.

(b) If $A(t)$ is differentiable, then so is $e^{A(t)}$.

(c) If $A(t)$ and $\dot{A}(t)$ commute for a particular value of t, then $\frac{d}{dt} e^{A(t)} = e^{A(t)} \dot{A}(t)$ (for that t).

(d) If A is anti-self-adjoint ($A^* = -A$), then e^A is unitary.

Proof: (a) If $AB = BA$, then $(A+B)^k = \sum_{i=0}^{k} \binom{k}{i} A^i B^{k-i}$, so

$$e^{A+B} = \sum_{k=0}^{\infty} \frac{(A+B)^k}{k!} = \sum_{k=0}^{\infty} \frac{1}{k!}\left(\sum_{i=1}^{k} \binom{k}{i} A^i B^{k-i}\right) = \left(\sum_{k=0}^{\infty} \frac{1}{k!} A^k\right)\left(\sum_{k=0}^{\infty} \frac{1}{k!} B^k\right) = e^A e^B. \checkmark$$

(b) We need to following analysis lemma:

Lemma: Let $\{E_m(t)\}$ be a sequence of differentiable matrix-valued functions such that:

(i) $E_m(t) \to E(t)$ uniformly

(ii) $\dot{E}_m(t) \to F(t)$ uniformly (for some $F(t)$).

Then $E(t)$ is differentiable, and $\dot{E}(t) = F(t)$.

Pf: Exercise. $\square$

Now, put $E_m(t) = e_m(A(t)) = \sum_{k=0}^{m} \frac{1}{k!} A^k$.

We've shown that $E_m(t) \to e^{A(t)}$.

Similarly, it's elementary (Exercise!) to show that

$\dot{E}_m(t) = \sum_{k=0}^{m} \frac{1}{k!} \sum_{i=1}^{k} A^i \dot{A} A^{k-i-1}$ converges. (Show $\|E_m(t) - E_\ell(t)\| \to 0$.) $\checkmark$

(c) $\frac{d}{dt} e^A = \sum_{k=0}^{\infty} \frac{1}{k!} \frac{d}{dt} A^k = \sum_{k=0}^{\infty} \frac{1}{k!} k A^{k-1} \dot{A} = \left(\sum_{k=0}^{\infty} \frac{1}{k!} A^k \right) \dot{A} = e^A \dot{A}$ ✓

if $A\dot{A} = \dot{A}A$

(d) $(e^A)^* = \sum_{k=0}^{\infty} \left(\frac{A^k}{k!}\right)^* = \sum_{k=0}^{\infty} \frac{1}{k!}(A^*)^k = e^{A^*} = e^{-A}$.

Now, $(e^A)^* e^A = e^{-A} e^A = e^0 = I \Rightarrow e^A$ is unitary. ✓

□

Goal: How do the eigenvalues of a matrix depend on the matrix? (Throughout, let $K = \mathbb{C}$.)

The following result shows how the eigenvalues depend continuously on $A(t)$:

Theorem 9.6: Suppose $A_m \longrightarrow A$ (i.e, the corresponding entries converge) Then for every $\varepsilon > 0$, there is some N such that $m > N \Rightarrow$ the eigenvalues of A_m are contained in $B_\varepsilon(\lambda_1) \cup \dots \cup B_\varepsilon(\lambda_n)$ - i.e, within distance ε of the eigenvalues of A.

Proof: The eigenvalues of A_m are the roots of $p_m(s) = \det(sI - A_m)$.

So the entries of $A_m \longrightarrow$ entries of A

$\Rightarrow$ the coefficients of $p_m(s) \longrightarrow$ coeffs. of $p(s)$.

□

Question: Now that we know that the eigenvalues of $A(t)$ depend on t continuously, is this dependence differentiable as well?

Theorem 9.7: Let $A(t)$ be differentiable and say $A(0)$ has an eigenvalue λ_0 of algebraic multiplicity 1. Then for some $\varepsilon > 0$, if $|t| < \varepsilon$, $A(t)$ has an eigenvalue $\lambda(t)$ where $\lambda(t)$ is differentiable on $(-\varepsilon, \varepsilon)$ and $\lambda(0) = \lambda_0$.

Proof: The characteristic polynomial of $A(t)$ is $\det(sI - A(t)) = p(s,t)$, a polynomial of degree n in s, whose coefficients are differentiable functions of t.

Since λ_0 has algebraic multiplicity 1,

$$p(\lambda_0, 0) = 0 \qquad \text{and} \qquad \frac{\partial}{\partial s} p(s,0)\Big|_{s=\lambda_0} \neq 0.$$

By the implicit function theorem, $p(s,t) = 0$ has a solution $s = \lambda(t)$ in some neighborhood $(-\varepsilon, \varepsilon)$ in which $\lambda(t)$ is differentiable.

$\square$

Next question: Does the correspond eigenvector depend on $A(t)$ differentiably?

Answer: If we choose carefully.

Theorem 9.8: Let $A(t)$ be differentiable with an eigenvalue $\lambda(t)$ of algebraic multiplicity 1. Then there exists an eigenvector $v(t)$ for $\lambda(t)$ that is differentiable.

First, we need a lemma:

Lemma 9.9: Let A be an $n \times n$ matrix with an eigenvalue λ of algebraic multiplicity 1. Then at least one $(n-1)\times(n-1)$ minor of $A - \lambda I$ has non-zero determinant.

Proof: Assume WLOG that $\lambda = 0$ is a simple eigenvalue (λ is an eigenvalue of $A \Leftrightarrow 0$ is an eigenvalue of $A - \lambda I$).

If $p(s)$ is the char. poly of A, then by assumption,

$$p(0) = 0 \quad \text{but} \quad p'(0) \neq 0.$$

Let $A = (c_1, \dots, c_n)$ $\quad c_i$ is the i^{th} column vector

$$\Rightarrow sI - A = (se_1 - c_1, \dots, se_n - c_n)$$

$$\Rightarrow p(s) = \det(sI - A).$$

Recall that $p'(0) = \frac{d}{ds} \det(sI - A)\Big|_{s=0} \neq 0$

$$= \det(e_1, -c_2, \dots, -c_n) + \dots + \det(-c_1, \dots, -c_{n-1}, e_n).$$

One of these must be non-zero! $\square$

Now, let A be an $n \times n$ matrix with a simple eigenvalue (algebraic multiplicity 1) $\lambda = 0$.

WLOG, suppose that A_{nn}, the n^{th} principle minor (remove the n^{th} row and column) has nonzero determinant.

<u>Claim</u>: If $v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \neq 0$ is an eigenvector for λ, then $v_n \neq 0$.

To see why, let $v^{(n)} = \begin{pmatrix} v_1 \\ \vdots \\ v_{n-1} \end{pmatrix}$ and note that

$$A_{nn} v^{(n)} = 0, \quad \det A_{nn} \neq 0 \Rightarrow v^{(n)} = 0 \Rightarrow v = 0 \text{ ↯}$$

Thus, we may assume that $v_n = 1$.

We get the following system, $Av = 0$:

$$\begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1,n-1} & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2,n-1} & c_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ c_{n-1,1} & c_{n-1,2} & \cdots & c_{n-1,n-1} & c_{n-1,n} \\ c_{n1} & c_{n2} & \cdots & c_{n,n-1} & c_{nn} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_{n-1} \\ 1 \end{pmatrix} = \begin{pmatrix} h_1 c_{11} + h_2 c_{12} + \cdots + h_{n-1} c_{1,n-1} + c_{1n} \\ h_1 c_{21} + h_2 c_{22} + \cdots + h_{n-1} c_{2,n-1} + c_{2n} \\ \\ h_1 c_{n-1,1} + \cdots + h_{n-1} c_{n-1,n-1} + c_{n-1,n} \\ h_1 c_{n1} + \cdots + h_{n-1} c_{n,n-1} + c_{nn} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}$$

Ignoring the last row, we get a system of $n-1$ equations

$$A_{nn} v^{(i)} = -C_n^{(n)} \Rightarrow v^{(i)} = -A_{nn}^{-1} C_n^{(n)}$$

Now let $A(t)$ be as in Theorem 9.8: differentiable with a simple eigenvalue $\lambda(t)$.

Apply Lemma 9.9: Some $(n-1)\times(n-1)$ minor, WLOG $A_{nn}(0)$, is invertible, so by Theorem 9.6, $A_{nn}(0)-\lambda(t)I$ is invertible for $t\in(-\varepsilon,\varepsilon)$.

For any such t, put $v_n=1$ (where $v=\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}\neq 0$ is an eigenvector), and as we just saw,

$$v^{(n)} = \begin{pmatrix} v_1 \\ \vdots \\ v_{n-1} \end{pmatrix} = -A_{nn}^{-1}(t)\, c^{(n)}(t).$$

Each $v_i(t)$ is differentiable, thus so is $v(t)$. $\square$

There are similar results for the continuity and differentiability of repeated eigenvalues and the corresponding (possibly generalized) eigenvectors, but the analysis is <u>much</u> more involved.