Lecture 3 Linear Algebra

As noted earlier, not all matrices are diagonalizable. The question arises as to what forms a given matrix can always be reduced to by a simple change of basis. In other words what (preferably simple) form D can be achieved by conjugating a given matrix, $A \to V^{-1} A V = D$?

A useful result is the following.

Theorem (Schur form)

Given any square matrix A, there always exists a non-singular matrix $V = V(A)$ and an upper triangular matrix U such that $A = VUV^{-1}$. Moreover V can be chosen to be unitary, i.e. $U^* = U^{-1}$.

Proof: Suppose A is $n \times n$ and let λ be an eigenvalue, $Ax = \lambda x$, $\|x\| = 1$. Complete x to an orthonormal basis in $\mathbb{C}^n$, i.e. $v_1 = x, v_2, \ldots, v_n$ where $(v_i, v_j) = \delta_{ij}$. Let V' be the (clearly invertible matrix) whose columns are

are the v_i's, $V' = (v_1, \dots, v_n)$. Then a straight forward computation shows that

$$AV' = V' \left(\begin{array}{c|ccc} \lambda & a_2 & \cdots & a_n \\ \hline 0 & & & \\ \vdots & & B & \\ 0 & & & \end{array} \right)$$

for a suitble $(n-1)\times(n-1)$ matrix B. By the appropriate induction hypothesis $B = V''U''(V'')^{-1}$ for a suitable unitary matrix V'' and an upper triangular matrix U'', both of size $(n-1)\times(n-1)$. (The case $n=1$ is trivial).

Now set

$$V = V' \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & V'' & \\ 0 & & & \end{array} \right), \quad U = \left(\begin{array}{c|ccc} \lambda & a_2' & \cdots & a_n' \\ \hline 0 & & & \\ \vdots & & U'' & \\ 0 & & & \end{array} \right)$$

and check that $A = VUV^{-1}$ for suitable $(a_2', \dots, a_n')$.

Clearly $V^* = V^{-1}$ and U is upper. This completes the induction and the proof of the Theorem. $\square$

If A is normal, we can push Schur's result further to show that

<u>Theorem</u> (<u>the Spectral Theorem</u>) If A is normal then there exists a unitary matrix U and a diagonal matrix Λ such

that

$$A = U \Lambda U^* . \tag{30.1}$$

Necessarily $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ are the eigenvalues of A.

<u>Proof</u> By Schur, $A = U D U^*$ with D upper.

Thus $A^* = U D^* U^*$. From $A^*A = AA^*$, we

have $U D D^* U^* = U D^* D U^*$, so that $DD^* = D^*D$

Hence

$$\begin{pmatrix} d_{11} & d_{12} & d_{13} & \cdots \\ 0 & d_{22} & d_{23} & \\ & & d_{33} & \cdots \\ & 0 & & \end{pmatrix} \begin{pmatrix} \bar{d}_{11} & 0 & & \\ \bar{d}_{12} & \bar{d}_{22} & 0 & \\ \bar{d}_{13} & \bar{d}_{23} & \bar{d}_{33} & \\ & \vdots & & \end{pmatrix}$$

$$= \begin{pmatrix} \bar{d}_{11} & 0 & 0 & - \\ \bar{d}_{12} & \bar{d}_{22} & 0 & - \\ \bar{d}_{13} & \bar{d}_{23} & \bar{d}_{33} & 0 \\ & & & \ddots \end{pmatrix} \begin{pmatrix} d_{11} & d_{12} & d_{13} & \\ 0 & d_{22} & d_{23} & \\ 0 & 0 & d_{33} & \\ & 0 & & \ddots \end{pmatrix}$$

Equating $(1,1)$ elements, we get

$$\sum_{j=1}^{n} |d_{1j}|^2 = |d_{11}|^2$$

which implies $d_{12} = d_{13} = \cdots = d_{1n} = 0$

Equating $(2,2)$ elements, we get

$$\sum_{j=2}^{n} |d_{2j}|^2 = |d_{12}|^2 + |d_{22}|^2 = |d_{22}|^2$$

and so $d_{23} = d_{24} = \cdots = d_{2n} = 0$, etc

We find that $D = \mathrm{diag}(d_{11}, \ldots, d_{nn})$, as desired. $\square$

Theorem

The eigenvalues $\{\lambda\}$ of a Hermitian (or real symmetric) A matrix are real, $\lambda = \bar{\lambda}$. The eigenvalues $\{\lambda\}$ a unitary matrix A have modulus 1, $|\lambda| = 1$. If A is real, then the eigenvalues $\{\lambda\}$ of A come in conjugate pairs $(\lambda, \bar{\lambda})$. If A is real and orthogonal, then the eigenvalues $\{\lambda\}$ of A have modulus 1, $|\lambda| = 1$, and come in conjugate pairs, $(\lambda, \bar{\lambda} = \frac{1}{\lambda})$. If u, v are eigenvectors of a Hermitian matrix corresponding to distinct eigenvalues, $Au = \lambda u$, $Av = \mu v$, $\lambda \neq \mu$, then u and v are orthogonal, $(u, v) = 0$.

<u>Proof</u> If $A = A^{+}$, then $(Au, u) = (u, Au)$ $\forall u$, and so if $Au = \lambda u$, $u \neq 0$,

$$\bar{\lambda}(u,u) = (\lambda u, u) = (Au, u) = (u, Au) = \lambda (u, u)$$

$\Rightarrow \lambda = \bar{\lambda}$. If A is unitary, then $(Au, Av) = (u, A^{+}Au)$ $= (u, u)$. Hence if $Au = \lambda u$, $u \neq 0$,

$$|\lambda|^2 (u,u) = (\lambda u, \lambda u) = (Au, Au) = (u, u)$$

$\Rightarrow |\lambda| = 1$

If A is real and λ is an eigenvalue of A, $\det(A - \lambda I) = 0$

But then taking conjugates,

$$0 = \overline{\det(A-\lambda)} = \det(\bar{A}-\bar{\lambda}) = \det(A-\bar{\lambda})$$

and so $\bar{\lambda}$ is an eigenvalue: if $Au = \lambda u$, then clearly $\bar{u}$ is an eigenvector for A corresponding to $\bar{\lambda}$. The real orthogonal case is now clear.

If $A = A^*$ and $Au = \lambda u$, $Av = \mu v$, $\lambda \neq \mu$, then from

$$(Au, v) = (u, Av)$$

we have as λ, μ are real,

$$\lambda(u, v) = \mu(u, v)$$

$\Rightarrow$ $(u, v) = 0$.

Suppose A is unitary and $Au = \lambda u$, $Av = \mu v$, $\lambda \neq \mu$.

Now $Au = \lambda u \Rightarrow u = \lambda A^* u$ as $A^* A = I$.

But $|\lambda| = 1$ so $\lambda^{-1} = \bar{\lambda}$: Thus $A^* u = \bar{\lambda} u$. But then from $(A^* u, v) = (u, Av)$ we again find

$$\lambda(u, v) = \mu(u, v)$$

$\Rightarrow$ $(u, v) = 0$. $\square$

<u>Note</u> The fact that for unitary A, $Au = \lambda u$ $\Rightarrow$ $A^* u = \bar{\lambda} u$, is also directly clear from (30.1).

The spectral Theorem is of fundamental importance in linear algebra, and it is instructive to give different proofs using different methods: we give two additional proofs, one using differential equations, the other using variational methods.

We consider the case where $A = A_0$ is real symmetric.

First we note some basic facts about ordinary differential equations (ODE's) in $\mathbb{R}^n$. Let $f: \mathbb{R}^{n+1} \to \mathbb{R}^n$ taking $\mathbb{R}^{n+1} = \mathbb{R}^n \times \mathbb{R} \ni (x,t) \mapsto f(x,t) \in \mathbb{R}^n$.

We seek a solution $x(t) \in \mathbb{R}^n$, $t \geq t_0$ of the <u>initial value problem</u>

(33.1) $$\dot{x}(t) = \frac{dx}{dt} = f(x(t), t)$$

with

(33.2) $$x(t = t_0) = x_0 \in \mathbb{R}^n,$$

A solution of (33.1)(33.2) $x(t)$, i.e. a differentiable function $x(t)$ satisfying (33.1)(33.2), may or may not exist. If the solution exists for all $t \geq 0$, we say

$x(t)$ is a global solution of (33.1)(.2): if we only exists for a finite time, $0 \le t \le T < \infty$, we say it is a local solution. The solution may or may not be unique. If $f(x,t)$ is independent of t,

$$f(x,t) = f(x) : \mathbb{R}^n \to \mathbb{R}^n$$

we say that (33.1) is autonomous; otherwise we say it is not autonomous.

For example

(34.1) $$\dot{x} = x, \qquad x(0) = x_0$$

has a global solution $x(t) = e^t x_0$, $t \ge 0$. Also the (non) autonomous equation

(34.2) $\dot{x} = t x$, $x(0) = x_0$, has a global solution $x(t) = e^{\frac{1}{2}t^2} x_0$

On the other hand

(34.3) $$\dot{x} = x^2, \qquad x(0) = x_0 > 0$$

has a local solution

$$x(t) = \frac{1}{x_0^{-1} - t}, \qquad 0 \le t < T = x_0^{-1} < \infty$$

(Notice, however, that if $x_0 < 0$, the solution is global.) In all the cases (34.1)(34.2)(34.3), the solutions are unique as long as they exist (prove this?) On the other hand, the initial value problem

(34.4) $$\dot{x} = \sqrt{|x(t)|}, \qquad x(0) = 0$$

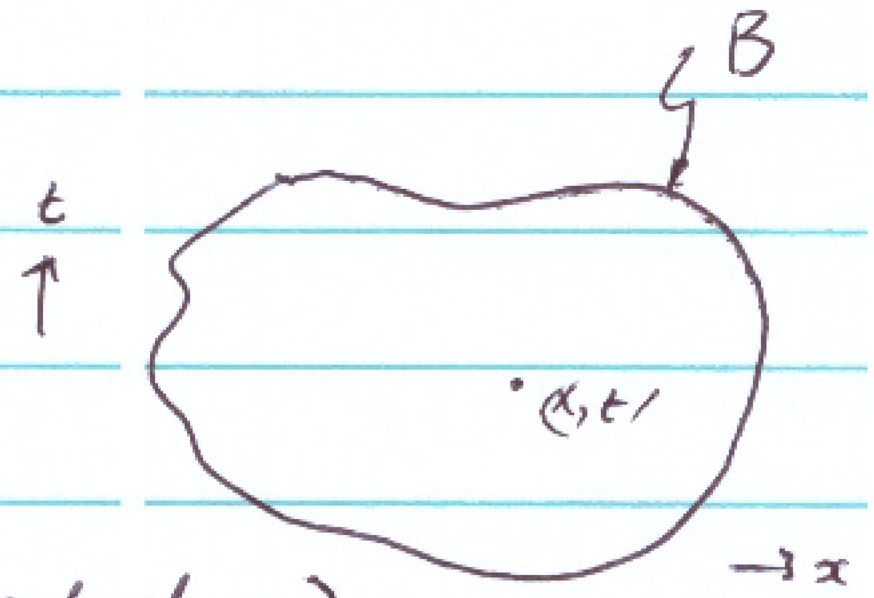
has many solutions

$$x(t) = 0, \qquad 0 < t < t_0$$
$$= \tfrac{1}{4}(t-t_0)^2, \qquad t \geq t_0$$

for any choice of $t_0 \geq 0$.

There is a convenient condition on f that ensures that (33.1) (33.2) has a unique, local solution, viz., the <u>Lipschitz condition</u>. For an open subset $B \subset \mathbb{R}^{n+1}$, we say a function $f: B \to \mathbb{R}^n$ is <u>Lipschitz continuous</u> on B if it is continuous on B and if there exists L (called the Lipschitz constant of f on B) such that

(35.1) $$\|f(x,t) - f(y,t)\| \leq L\,\|x-y\|$$



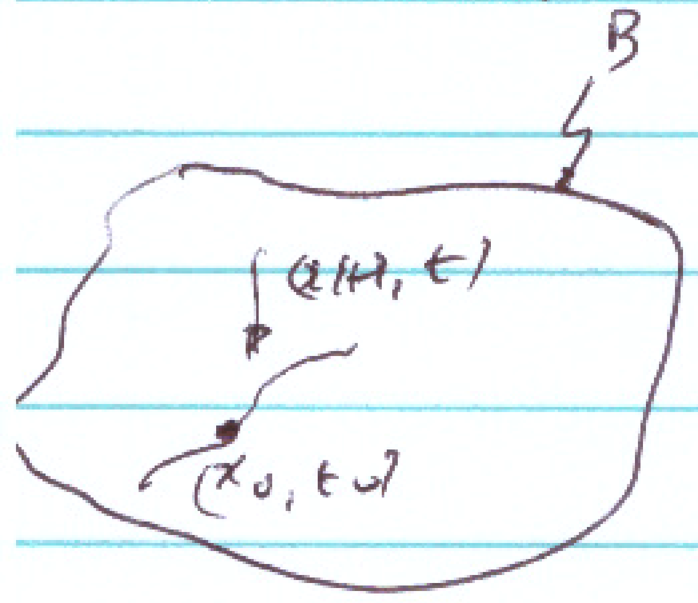
for all (x,t) and $(y,t) \in B$.
(Here $\|\cdot\|$ denotes any norm on $\mathbb{R}^{n+1}$ and $\mathbb{R}^n$ respectively.)
Clearly if f is Lipschitz continuous, then it is continuous.

The basic theorem is the following (see, for example, Coddington and Levinson, Theory of ordinary differential equations

A good reference for the following 2 theorems is F. John, ODE's, Courant Lecture Notes. Also Coddington & Levinson, Theory of ODE's.

<u>Theorem 36.1</u> (<u>Picard's Theorem</u>)

Suppose f is Lipschitz continuous on an open subset $B \subset \mathbb{R}^{n+1}$. Then the initial value problem (33.1)(33.2) with $(x(t_0)=x_0, t_0) \in B$ has a unique local solution $x(t)$ for $|t-t_0| < T$, for some $T < \infty$.



<u>Theorem 36.2</u> (Global existence)

Suppose that f is Lipschitz continuous on $B \subset \mathbb{R}^n$ as in Theorem 36.1. Let $x(t)$ be a (local) solution of (33.1)(33.2) that exists for $t_0 \le t < T$ for some T, and let β be the supremum of such T's, i.e.

(36.1) $$\beta = \sup\{T : \text{a solution } x(t) \text{ of } (33.1)(33.2) \ \exists \text{ for } t_0 \le t < T\}$$

(Such T's $\exists$ by Theorem 36.1 so the set is non-empty.)

Let B^* be any closed and bounded subset of B. Then if $\beta < \infty$, the trajectory $z(t) = (x(t), t)$ lies outside of D^* for all t sufficiently close to β.

local solutions for any $T > 0$, $x(t)$ is a global solution.

In particular if $z(t) \in B^*$ some some closed and bounded set in B for all $t_0 \le t < \beta$, then $x(t)$ is a global solution of (33.1) (33.2).

Note that in (34.1), $B = \infty$ so the solution is global. In (34.2), for any $B = (-R, R)$, the Lipschitz constant $L = 2R$, and the solution $x(t)$ escapes from B for any R, so the solution is not global. In (34.3), f is not Lipschitz continuous in a neighborhood of $x(0) = 0$, so uniqueness fails.

Returning to the spectral Theorem for the $n \times n$ real symmetric matrix A_0, consider the autonomous ODE in $M(n, \mathbb{R}) = \{ n \times n \text{ real matrices} \} \cong \mathbb{R}^{n^2}$,

(37.1) $$\frac{dA}{dt} = [B(A), A] = B(A)A - AB(A), \quad A(t=0) = A_0,$$

where

(37.2) $$B(A) = A_+ - A_+^T = -(B(A))^T$$

and A_+ is the strict upper part of A,

$$A_+ = \begin{pmatrix} 0 & a_{12} & \cdots & a_{1n} \\ & 0 & a_{23} \cdots & a_{2n} \\ & & \ddots & \\ & 0 & & a_{n-1,n} \\ & & & 0 \end{pmatrix}$$

Equations (37.1) are called the <u>Toda equations</u>.

The inner product on $M(n,\mathbb{R}) \cong \mathbb{R}^{n^2}$ is

given by

(38.0) $$(A,B) = \sum_{1\le i,j\le n} A_{ij} B_{ij}$$

In particular

(38.1) $$\|A\| = \sqrt{(A,A)} = \Big(\sum_{1\le i,j\le n} A_{ij}^2\Big)^{\frac{1}{2}}.$$

Note that

$$\sum_{1\le i,j\le n} A_{ij} B_{ij} = \sum_{1\le i,j\le n} A_{ij} (B^T)_{ji}$$

$$= \sum_{i=1}^{n} ((AB)^T)_{ii}$$

ie

(38.2) $$(A,B) = \operatorname{tr} AB^T = \operatorname{tr} BA^T = \operatorname{tr} B^T A$$

and

(38.3) $$\|A\| = \sqrt{\operatorname{tr} AA^T}, \text{ and by Cauchy Schwarz}$$

(38.4) $$|(A,B)| \le \|A\| \, \|B\|$$

Now observe that for $X, Y \in M(n,\mathbb{R})$

(38.5) $$\| [B(X), X] - [B(Y), Y] \|$$

$$\le \| [B(X), X-Y] \| + \| [B(Y)-B(X), Y] \|$$

$$\le \| B(X)(X-Y) \| + \| (X-Y) B(X) \|$$
$$+ \| (B(Y)-B(X)) Y \| + \| Y (B(Y)-B(X)) \|$$

Now observe that ~~for~~ any C, D in $M(n,\mathbb{R})$

$$\|CD\|^2 = \sum_{1\le i,j\le n} (CD)_{ij}^2$$

$$= \sum_{1\le i,j\le n} \Big(\sum_{k=1}^{n} C_{ik} D_{kj}\Big)^2$$

$$\le \sum_{1\le i,j\le n} \Big(\sum_{k=1}^{n} C_{ik}^2\Big)\Big(\sum_{k=1}^{n} D_{kj}^2\Big), \quad \text{by Cauchy-Schwarz},$$

$$= \Big(\sum_{1\le i,k\le n} C_{ik}^2\Big)\Big(\sum_{1\le j,k} D_{kj}^2\Big)$$

$$= \|C\|^2 \, \|D\|^2$$

Thus

(39.1) $$\|CD\| \le \|C\| \, \|D\|$$

<u>Remark</u> (39.1) shows in particular that $M(n,\mathbb{R})$ with norm (38.1) is a <u>Banach algebra</u>.

Inserting (39.1) into (38.5) we find

$$\|[B(X),X] - [B(Y),Y]\| \le 2\,\|B(X)\|\,\|X-Y\| + \|B(Y)-B(X)\|\,\|Y\|.$$

But $\|B(Y)-B(X)\|^2 = \sum_{ij} (B(Y)-B(X))_{ij}^2$

$$= \sum_{1\le i<j\le n} (B(Y)-B(X))_{ij}^2, \text{ as } B(X) \text{ is skew,}$$

$$= \sum_{i<j} (Y_{ij}-X_{ij})^2 \le \|Y-X\|^2$$

Also $\|B(X)\| = \|B(X)-B(0)\| \le \|X-0\| = \|X\|$

We conclude that $f(X) = [B(X), X]$ is Lipschitz continuous and

(40.1) $$\|f(X)-f(Y)\| \le 2(\|X\|+\|Y\|)\,\|X-Y\|$$

It follows that (37.1) has a (unique) local solution $A(t)$ for $0\le t<T$, $T<\infty$. Now

$$\frac{d\,AA^T}{dt} = (B(A)A - AB(A))A^T + A(A^T B(A)^T - B(A)^T A^T)$$

$$= [B(A), AA^T], \text{ as } B(A)^T = -B(A)$$

Thus as $\operatorname{tr}[X,Y] = 0$ $\forall X, Y$,

$$\frac{d}{dt}\operatorname{tr} AA^T = \operatorname{tr}[B(A), AA^T] = 0$$

Thus

(40.2) $$\|A(t)\| = \sqrt{\operatorname{tr} A(t)A(t)^T} = \text{const} = \|A_0\|$$

It now follows from Theorem (36.2) that $A(t)$ is in fact global.

Now

$$\frac{dA^T}{dt} = [B(A), A]^T = A^T B(A)^T - B(A)^T A^T$$
$$= [B(A), A^T]$$

and $A^T(0) = A_0^T = A_0$. Hence $A^T(t)$ is a second solution of (37.1) and so by uniqueness

$$A^T(t) = A(t)$$

ie (37.1) preserves symmetry.

Let $Q(t)$ be the solution of the ODE

(42.2) $$\frac{dQ}{dt} = -Q(t)\, B(A(t)), \qquad Q(0) = I$$

where $A(t)$ solves (37.1).

The solution $Q(t)$ of (42.2) exists globally and is unique (why? prove this). We have

$$\frac{d}{dt} QQ^T = \dot{Q}Q^T + Q\dot{Q}^T = -QBQ^T + QBQ^T = 0$$

Hence $Q(t)Q(t)^T = \text{const} = I$. Thus $Q(t)$ is a (real) orthogonal matrix. Finally, set

(42.3) $$S(t) \equiv Q^T(t)\, A_0\, Q(t)$$

Then

$$\frac{dS}{dt} = B\,Q^T A_0 Q - Q^T A_0 Q B$$

$$= [B, S]$$

and $S(t=0) = A_0$. It follows again by uniqueness that $S(t) = A(t)$, the solution of (37.1) (37.2). Thus

(42.1) $$A(t) = Q^T(t)\, A_0\, Q(t)$$

It follows, in particular, that the Toda equations (37.1) (37.2) generate an isospectral deformation of A_0

ie.

(42.2) $$\{\text{eigenvalues of } A(t)\} = \{\text{eigenvalues of } A_0\}$$

Now note that

$$\frac{d}{dt} a_{11} = (BA - AB)_{11} = \sum_j b_{1j} a_{j1} - a_{1j} b_{j1}$$

$$= 2 \sum b_{1j} a_{1j}$$

$$= 2 \sum_{j=2}^{n} a_{1j}^2$$

as $b_{11} = 0$ and $b_{1j} = a_{1j}$ for $j > 1$.

Thus

(42.3) $$a_{11}(t) = a_{11}(0) + 2 \sum_{j=2}^{n} \int_0^t a_{1j}^2(s)\, ds$$

As $|a_{11}(t)| \le \left(\sum_{i,j} a_{ij}^2(t)\right)^{\frac{1}{2}} = \|A(t)\| = \|A_0\|$, $a_{11}(t)$ is bounded, and as it is monotone by (42.3), we see that

$$\gamma_{11} = \lim_{t \to \infty} a_{11}(t)$$

exists and is finite and satisfies

(43.1) $$\gamma_{11} = a_{11}(0) + 2 \sum_{j=2}^{n} \int_0^\infty a_{1j}^2(s)\,ds$$

In particular

$$\int_0^\infty a_{1j}^2(s)\,ds < \infty$$

But as $(a_{ij}(t))_{1 \le i,j \le n}$ is bounded, $([B(t), A])_{ij}$

is bounded, it follows that $\frac{d a_{ij}(t)}{dt}$ is bounded.

In particular $a_{ij}(t)$ is uniformly continuous for all $t \ge 0$.

This uniform continuity together with $\int_0^\infty a_{1j}^2(s)\,ds < \infty$

$\Rightarrow$ (Exercise)

$$\lim_{t \to \infty} a_{1j}^2(t) = 0 \qquad , \quad 2 \le \; \le n.$$

Now show by induction (Exercise) that

$$\frac{d}{dt} \sum_{i=1}^{k} a_{ii} = 2 \sum_{j=1}^{k} \left(\sum_{i > k} a_{ji}^2 \right)$$

and conclude that $\gamma_{ij} = \lim_{t \to \infty} a_{ij}(t)$ exists for

all i, j, where $\gamma_{ij} = 0$ for $i \ne j$. In other words we see that $\lim_{t \to \infty} A(t)$ exists and is a diagonal matrix

$\Lambda = \operatorname{diag}(\gamma_{11}, \ldots, \gamma_{nn})$. But

$$A(t) = Q(t)^T A_0 Q(t) \quad , \text{ or}$$

$$A_0 = Q(t) A(t) Q(t)^T$$

Now the orthogonal (real) matrices form a compact

set (why?). Thus $\exists$ a sequence t_n, $t_n \to \infty$, and

an orthogonal matrix Q, such that

$$Q(t_n) \to Q$$

Consequently,

$$A_0 = \lim_{n \to \infty} Q(t_n) A(t_n) Q(t_n)^T$$

$$= Q \Lambda Q^T$$

which is the spectral theorem for A_0. Necessarily $\Lambda = \operatorname{diag}(\gamma_{11}, \ldots, \gamma_{nn})$ are the eigenvalues of A_0. $\square$

Remark The length of this proof is connected with the fact that we proved the existence of eigenvalues en route. Said differently, we did not assume the fundamental theorem of algebra which ensures the existence of the eigenvalues in the previous proof.

Important exercise

Try to understand how the above proof gets around the need for the fundamental thm of algebra.

Exercise Redo the above proof for Hermitian matrices A. In place of

$$\frac{dA}{dt} = [B(A), A]$$

we now consider

$$\frac{dA}{dt} = [\hat{B}(A), A]$$

where $\hat{B}(A) = A_+ - A_+^* = -(\hat{B}(A))^*$