Lecture 7 — Linear Algebra — Fall 2019

We now give a proof of the Jordan canonical form theorem (p87 et seq). Our proof is taken from G. Strang, Linear Algebra and its Applications, 2nd Edition. Strang's proof is based in turn on an idea of Filippov, published in Vol. 26 of the Moscow University Vestnik

Consider the following example of a Jordan matrix

$$J = \begin{pmatrix} 8 & 1 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} = \left(\begin{array}{cc|cc|c} 8 & 1 & & & \\ 0 & 8 & & & \\ \hline & & 0 & 1 & \\ & & 0 & 0 & \\ \hline & & & & 0 \end{array}\right) = \left(\begin{array}{c|c|c} J_1 & & \\ \hline & J_2 & \\ \hline & & J_3 \end{array}\right)$$

The key question is this: If A is some other 5×5 matrix, under what conditions will its Jordan form be this same matrix J, i.e., $M^{-1}AM = J$ for some invertible matrix M?

As a first requirement, any similar matrix A must have the same eigenvalues, $8, 8, 0, 0, 0$ (Note that $\lambda = 8$ has only 1 eigenvector and $\lambda = 0$ has 2). But this is far from sufficient — the diagonal matrix diag $(8, 8, 0, 0, 0)$ is not similar to J — and our question really concerns eigenvectors.

To answer it, we rewrite $M^{-1}AM = J$ in the simpler form $AM = MJ$:

$$(90.1) \quad A\begin{bmatrix} x_1 & x_2 & \cdots & x_5 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_5 \end{bmatrix} \begin{bmatrix} 8 & 1 & & & \\ 0 & 8 & & & \\ & & 0 & 1 & \\ & & 0 & 0 & \\ & & & & 0 \end{bmatrix}$$

where x_i, $i=1,\dots,5$, are the columns of M. Carrying out the multiplications column by column, we find

$$(90.2) \quad Ax_1 = 8x_1 \quad \text{and} \quad Ax_2 = 8x_2 + x_1$$
$$(90.3) \quad Ax_3 = 0x_3 \quad \text{and} \quad Ax_4 = 0x_4 + x_3, \quad Ax_5 = 0x_5$$

Now we can recognize the conditions on A. It must have 3 genuine eigenvectors, just as J has. The one with $\lambda = 8$ will go into the first column of M, $Ax_1 = 8x_1$. The other 2 eigenvectors, which will be named x_3 and x_5, go into the 3rd and 5th columns of M: $Ax_3 = 0$, $Ax_5 = 0$. Finally there must be 2 other special vectors, the "generalized vectors" x_2 and x_4. We think of x_2 as belonging to a string of vectors, headed by the eigenvector x_1, and described by (90.2). In fact x_2 is the only other vector in the string, and the corresponding block J_1 is of order 2. Eqtn 2 describes, one in which x_4 follows the eigenvector x_3, and another in which x_5 is alone; the blocks J_2 and J_3 are 2×2 and 1×1 respectively.

The search for the Jordan form of A becomes a search for these strings of vectors, each one headed by an eigenvector: For every i

either $\quad A x_i = \lambda_i x_i \quad$ or $\quad A x_i = \lambda_i x_i + x_{i-1}$,

The vectors x_i go into the columns of M and each string produces a single block in J. Essentially we have to show how these strings can be constructed for every matrix A. Then if the strings match the particular eqtns (90.2) (90.3), our J will be the Jordan form of A

Fillipov's idea is to use induction on the order n of the matrix A. Clearly any 1×1 matrix is in Jordan form. So suppose and matrix of order $< n$ can be placed in Jordan form. We then show this is true for any $n \times n$ matrix A.

Step 1 Assume first that A is singular, i.e. $\dim N(A) > 0$

Now as

$$\dim R(A) + \dim N(A) = n$$

it follows that

$$r = \dim R(A) < n.$$

Now as A maps $R(A) \to R(A)$, we consider the restriction $\hat{A}$ of A to $V = R(A)$. Then by the induction hypothesis, $\hat{A}$ can be placed in Jordan form. Thus there are r independent vectors w_i in $R(A)$ such that

(92.1) $$\text{either} \quad Aw_i = \lambda_i w_i \quad \text{or} \quad Aw_i = \lambda_i w_i + w_{i-1}$$

Step 2 Suppose $R(A) \cap N(A)$ has dimension p. Of course every vector in the null space is an eigenvector corresponding to $\lambda = 0$. Therefore there must be p strings in Step 1 starting from this eigenvalue. Now consider the p vectors w_i at the end of these p strings

$$\vdots$$
$$Aw_{i-1} = 0\, w_{i-1} + w_{i-2}$$
$$Aw_i = 0\, w_i + w_{i-1}$$

But as $w_i \in R(A)$, we have $w_i = Ay_i$ for some y_i. Thus we can extend these strings as follows

$$\vdots$$
$$Aw_{i-1} = 0\, w_{i-1} + w_{i-2}$$
$$Aw_i = 0\, w_i + w_{i-1}$$
$$Ay_i = 0\, y_i + w_i$$

Thus in addition to the r independent vectors $\{w_i\}$, we have p vectors $\{y_i\}$. Each one of these vectors y_i lies at the end of a string headed by a null vector of A.

Step 3 Now complement the p null vectors of A in $R(A) \cap N(A)$ with $\ell = \dim N(A) - p$ vectors $\{z_i\}$, $Az_i = 0$, $i = 1, \dots, \ell$, to form a basis of $N(A)$. Note that the span $\langle z_1, \dots, z_\ell \rangle \cap R(A) = \{0\}$.

Indeed if $z = \sum_1^l c_i z_i \in R(A)$, then z must be a linear combination of the p null vectors of A in $R(A)$. But this contradicts the construction of $(3c)$. Now we put these steps together to give Jordan's theorem:

The r vectors w_i, the p vectors y_i, and the $l = \dim N(A) - p$ vectors form Jordan strings for the matrix A,

$$r + p + l = \dim R(A) + \dim N(A) = n$$

and, as we will show these vectors are independent, they go into the columns of M and $J = M^{-1} A M$ is in Jordan form.

To show that these n vectors are independent,

Suppose

$$\sum c_i w_i + \sum d_i y_i + \sum e_i z_i = 0 \tag{93.1}$$

Multiplying by A and using (92.1) we obtain

$$\sum c_i \begin{bmatrix} \lambda_i w_i \\ \text{or} \\ \lambda_i w_i + w_{i-1} \end{bmatrix} + \sum d_i A y_i = 0$$

as $A z_i = 0$. Now the $A y_i$ are the special w_i at the end of the strings corresponding to $\lambda_i = 0$, & so they clearly cannot appear in the first sum. But as these w_i are independent, we must have each $d_i = 0$. Returning to (93.1) this leaves

$$\sum c_i w_i = -\sum e_i z_i$$

but the LHS lies in $R(A)$ and the RHS does not. Hence each $e_i = 0$. But then $\sum c_i w_i = 0 \Rightarrow$ each $c_i = 0$

as the w_i's are independent. Thus $c_i = d_j = e_k = 0$

and the w_i's, y_i's and z_i's give n independent vectors.

This proves the induction for any $n \times n$ matrix A with

$N(A) \neq \{0\}$. If A is $n \times n$ and $N(A) = \{0\}$,

let c be any eigenvalue of A. Then $A - c$ has

$N(A-c) \neq \{0\}$, and so by the preceding construction, $\exists M, J_c$, M invertible, J_c Jordan, st

$$A - c = M J_c M^{-1}$$

But then

$$A = A - c + cI = M J_c M^{-1} + cI$$
$$= M J M^{-1}$$

where $J = J_c + cI$ is Jordan. This completes the proof of the Jordan form Theorem. $\square$.

Example

$$A = \begin{pmatrix} 0 & 1 & 2 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$ has a triple eigenvalue $\lambda = 0$,

with only 1 eigenvector: $A \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = 0$, $R(A)$ is

spanned by the vectors $w_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ and $w_2 = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}$

Now

$$Aw_1 = 0 \quad \text{and} \quad Aw_2 = w_1 = 0w_2 + w_1$$

Here $p = 1$ and the third vector y comes from A

$$Ay = w_2 = 0y + w_2$$

Thus $y = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ and so $M = (w_1\ w_2\ y)$

$= \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

Checking:

$$AM = \begin{pmatrix} 0 & 1 & 2 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 2 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

$$MJ = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 2 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \checkmark .$$

Min - max Theorems

In what follows A is taken to be a real $n \times n$ symmetric matrix. Most of the theorems that follow hold for Hermitian matrices: just use the complex inner product instead of the real one.

Given a (real, symmetric) $n \times n$ matrix A, define,

for each k, $1 \le k \le n$, the numbers

$$\lambda_k = \lambda_k(A) = \sup_{V_{k-1}} \Big(\inf_{\substack{u \perp V_{k-1} \\ \|u\|=1}} (u, Au) \Big)$$

where the sup is taken over all $(k-1)$-dimensional subspaces V_k (by convention $V_0 = \{0\}$ so $\lambda_1 = \inf_{\|u\|=1} (u, Au)$)

Theorem 96.1 (Rayleigh-Ritz: Courant) The #'s λ_k are increasing

$$\lambda_1 \le \lambda_2 \le \dots \le \lambda_n$$

and are precisely the eigenvalues of A, with appropriate (geometric = algebraic) multiplicities.

Proof: Fix $V_{k-1} = \langle v_1, \dots, v_{k-1} \rangle$. Adjoin any vector v_k ($\notin V_{k-1}$) to obtain the k-dim space $V_k = \{v_1, \dots, v_k\}$

Clearly, $\lambda_{k+1} \ge \inf_{u \perp V_k, \|u\|=1} (u, Au)$

But

$$\inf_{\substack{u \perp V_k \\ \|u\|=1}} (u, Au) \geq \inf_{\substack{u \perp V_{k-1} \\ \|u\|=1}} (u, Au)$$

Hence

$$\lambda_{k+1} \geq \inf_{\substack{u \perp V_{k-1} \\ \|u\|=1}} (u, Au)$$

Taking sup's over V_{k-1}, we obtain $\lambda_{k+1} \geq \lambda_k$. Now let $\beta_1 \leq \beta_2 \leq \dots \leq \beta_n$ be the eigenvalues of A with associated orthonormal eigenvectors $v_1, \dots, v_n$. The spectral th^m tells us that

(97.1) $$A = \sum_{i=1}^{n} \beta_i (v_i, \cdot) v_i$$

and so $$(u, Au) = \sum_{i=1}^{n} \beta_i (v_i, u)^2$$

Fix k and consider a fixed $(k-1)$-dim space V_{k-1}. Now $v_1, \dots, v_k$ span a k-dim space, and so $\exists$ $a_1, \dots, a_k$ such that $u = \sum_{i=1}^{k} a_i v_i$ is $\perp V_{k-1}$ and $\|u\|^2 = \sum_{i=1}^{k} a_i^2 = 1$. For this u, $(u, Au) = \sum_{i=1}^{k} \beta_i a_i^2$ $\leq \beta_k \sum_1^k a_i^2 = \beta_k$. It follows that

$$\inf_{\substack{w \perp V_{k-1} \\ \|w\|=1}} (w, Aw) \leq \beta_k$$

As V_{k-1} is arbitrary, taking sup's we conclude that

$$\lambda_k \leq \beta_k .$$

Conversely, let $V_{k-1} = \langle v_1, \ldots, v_{k-1} \rangle$. Then if $u \perp V_{k-1}$,

$\|u\| = 1$, $u = \sum_{i=k}^{n} a_i v_i$, $\sum_{i=k}^{n} a_i^2 = 1$ and so

$$\lambda_k \geq \inf_{\substack{u \perp V_{k-1} \\ \|u\|=1}} (u, Au) = \inf_{\sum a_i^2 = 1} \sum_{i=k}^{n} \beta_i a_i^2 \geq \beta_k .$$

We conclude that $\lambda_k = \beta_k$. $\square$

To understand these formulae, consider the following

problem:

Find the stationary points of

(98.1) $$f(u) = (u, Au) \quad \text{subject to} \quad \|u\| = 1$$

Using Lagrange multipliers, we minimize

(98.2) $$g(u) = (u, Au) - \lambda \left((u,u) - 1 \right)$$

with respect to u and λ

Stationary points: $\frac{\partial g}{\partial u_i} = 0$, $i = 1, \ldots, n$

$\frac{\partial g}{\partial \lambda} = 0$

Thus $2 (Au)_i - 2 \lambda u_i = 0$

$(u,u) - 1 = 0$

Thus $Au = \lambda u$ and $\lambda = \lambda (u,u) = (u, Au)$.

i.e. the stationary points of $f(u)$ subject to $\|u\| = 1$

occur at u's which are the eigenvectors of A and

$\lambda = (u, Au) = f(u)$. Thus, in particular, the value of

the Lagrange multiplier at the stationary point is the

eigenvalue. Thus from Rayleigh Ritz

$$\lambda_1 = \min_{\|u\|=1} (u, Au)$$

is an eigenvalue.

A geometrical picture is useful. Suppose $n = 3$ and

V is positive definite, ie $(u, Vu) > c\|u\|^2$ for some $c > 0$.

ie $(u, Vu) \geq c > 0$ for all $\|u\| = 1$. Then it follows from

min max that the eigenvalues α_i of A satisfy

(99.1) $$\alpha_1 \geq \alpha_2 \geq \alpha_3 > c$$

Suppose v_1, v_2, v_3 are the associated orthonormal

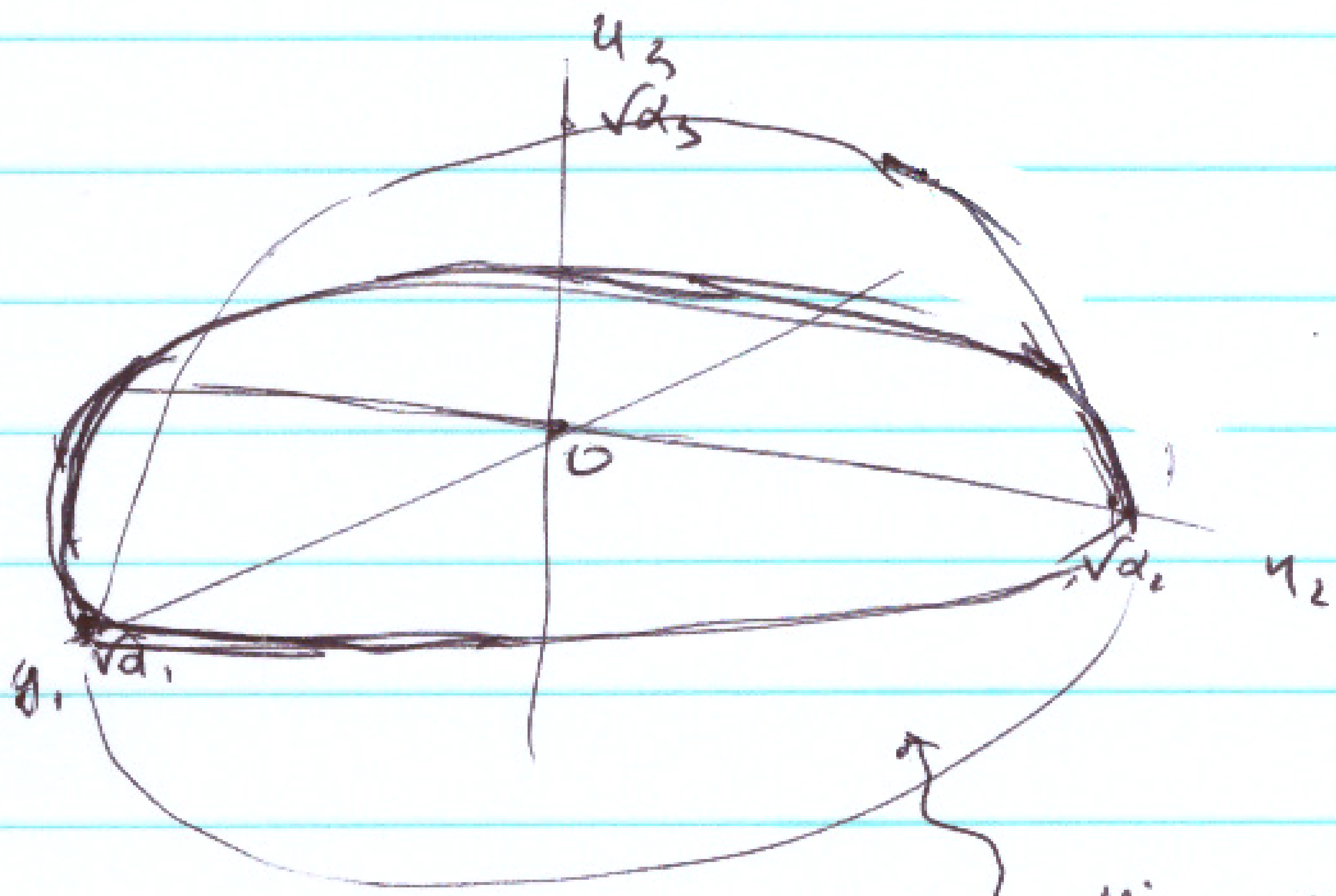
eigenvectors. Then writing $u = \sum_{i=1}^{3} x_i v_i$ we have

(100.1) $$\lambda_1 = \min_{\sum_1^3 x_i^2 = 1} \sum_1^3 \alpha_i x_i^2$$

Or setting $y_i = \sqrt{\alpha_i}\, x_i$

$$\lambda_1 = \min_{\sum_1^3 \frac{y_i^2}{\alpha_i} = 1} \sum y_i^2$$

Geometrically, the problem is the following:



ellipsoid: $\frac{u_1^2}{\alpha_1} + \frac{u_2^2}{\alpha_2} + \frac{u_3^2}{\alpha_3} = 1$

Problem: Find the square of the minimal distance of a point on the ellipsoid to the origin.

Clearly this distance must be α_3!

Thus $\lambda_1 = \alpha_3$ = smallest eigenvalue

Now consider

$$\lambda_2 = \sup_v \min_{\substack{u \perp v \\ \|u\| = 1}} (u, Au)$$

or changing $u \to (y_1, u_2, u_3)$

$$\lambda_2 = \sup_v \min_{y \perp v,\ \sum y_i^2/\alpha_i = 1} \sum_{i=1}^3 y_i^2$$

Thus for every fixed v, we must find the point on the ellipsoid and in the plane $\perp v$, which is at a minimum distance to the origin.

Now convince yourself that if you maximize the minimal distance over all hyperplanes, i.e. over all v, you obtain λ_2.