Lecture 4

Th^m The following are equivalent statements for a real $n \times n$ symmetric matrix A:

(i) A is strictly pos. definite

(ii) $\lambda_i(A) > 0$ for all eigenvalues of A, $i = 1, \dots, n$

(iii) The principle minors d_i, $i = 1, \dots, n$ are > 0

(iv) $A = C^T C$ for an upper triangular matrix C, $\det C \neq 0$

(v) $A = B^2$, $\det B \neq 0$, for some real symmetric matrix B

Proof: We have already proved (i) $\Leftrightarrow$ (ii). By Sylvester's theorem (i) $\Leftrightarrow$ (iii).

(v) $\Rightarrow$ (i): If $A = B^2$, $B = B^T$, $\det B \neq 0$

$$(u, Au) = (u, B^2 u) = (Bu, Bu) = \|Bu\|^2 \geq 0$$

Let $\gamma = \inf_{\|u\|=1} \|Bu\| \geq 0$. As $\{\|u\| = 1\} \ni u \to \|Bu\|$

is a continuous function on a compact set, it must

achieve its infimum at some point u_0, $\|u_0\| = 1$

Thus $(u, Au) \geq \|Bu_0\|^2 \quad \forall \; \|u\| = 1$

If $\|Bu_0\| = 0$, then as $\det B \neq 0$, $u_0 = 0$, which is a contradiction. Thus $\gamma > 0$ and $(u, Au) \geq \gamma^2 \|u\|^2$. Thus (i) is true.

(i) $\Rightarrow$ (v) $A = U \Lambda U^T$, U real orthogonal, $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$, $\lambda_i > 0$, $i = 1, \dots, n$

Set $B = U \Lambda^{\frac{1}{2}} U^T$ when $\Lambda^{\frac{1}{2}} = \operatorname{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n})$

Then $A = B^2$, $B = B^T$ and $\det B \neq 0$

Clearly B is real as U & $\Lambda^{1/2}$ are real.

(iv) $\Rightarrow$ (i) $(u, Au) = (u, C^T C u) = \|Cu\|^2$

As above, $\gamma = \inf_{\|u\|=1} \|Cu\| > 0$ and so $(u, Au) \geq \gamma^2 \|u\|^2$

(iii) $\Rightarrow$ (iv) Do Gaussian elimination on A as $d_1 = a_{11} > 0$

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ a_{21} & \dots & a_{2n} \\ \vdots & & \\ a_{n1} & \dots & a_{nn} \end{pmatrix} \rightarrow \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a'_{22} & \dots & a'_{2n} \\ \vdots & \vdots & & \\ 0 & a'_{n2} & \dots & a'_{nn} \end{pmatrix}$$

Now the sub-matrix $\begin{pmatrix} a_{11} & a_{12} \\ 0 & a'_{22} \end{pmatrix}$ is obtained by

adding a multiple of the row $(a_{11} \; a_{12})$ to $(a_{21} \; a_{22})$

Therefore, $a_{11} a'_{22} = \det \begin{pmatrix} a_{11} & a_{12} \\ 0 & a'_{22} \end{pmatrix} = \det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = d_2 > 0$

Hence $a'_{22} = d_2 / a_{11} > 0$

Thus we can continue the reduction

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a'_{22} & \dots & a'_{2n} \\ \vdots & \vdots & & \\ 0 & a'_{n2} & \dots & a'_{nn} \end{pmatrix} \rightarrow \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a'_{22} & \dots & a'_{2n} \\ 0 & 0 & a''_{33} \dots & a''_{3n} \\ \vdots & \vdots & \vdots & \\ 0 & 0 & a''_{n3} \dots & a''_{nn} \end{pmatrix}$$

Arguing as above, we have

$$a_{33}' = \frac{d_3}{a_{11} a_{22}'} > 0$$

etc. Thus by Gaussian elimination, we can reduce $A \to U$, where U is real and upper triangular with positive entries on the diagonal. Now Gaussian elimination is implemented by multiplying A at each step on the left by a lower triangular matrix with 1's on the diagonal

$$\begin{pmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22}' & a_{23}' \\ 0 & a_{32}' & a_{33}' \end{pmatrix}$$

where $l_{21} = -a_{21}/a_{11}$, $l_{31} = -a_{31}/a_{11}$

Thus we have

$$LA = U$$

for some lower triangular matrix with 1's on diagonal

Clearly $\det L = 1 \neq 0$ and so L^{-1} exists and must also be lower triangular (why?). Hence

$$A = SU$$

S lower, $S_{ii} = 1$, U upper $U_{ii} > 0$. Set $U = DV$ where

$D = \mathrm{diag}(u_{11}, \ldots, u_{nn})$. Clearly V is upper with $V_{ii} = 1$, $i = 1, \ldots, n$. Have $A = SDV$. But $A = A^T$ and so $A = V^T D S^T$. Equating these 2 expressions for A we find $WD = DW^T$ where $W = V^{-T}S$ is lower triangular. Thus WD is lower, but DW^T is upper. Hence WD must be diagonal, and so

$$V^{-T}S = \Delta$$

for some diagonal matrix Δ ie $S = V^T\Delta$. But $S_{ii} = V_{ii} = 1$, $i = 1, \ldots, n$. Hence $\Delta = I$ and so $V = S^T$. Thus

$$A = V^T D V$$

Set $C = \mathrm{diag}(\sqrt{u_{11}}, \ldots, \sqrt{u_{nn}})\, V$ and so $A = C^T C$, which is (iv).

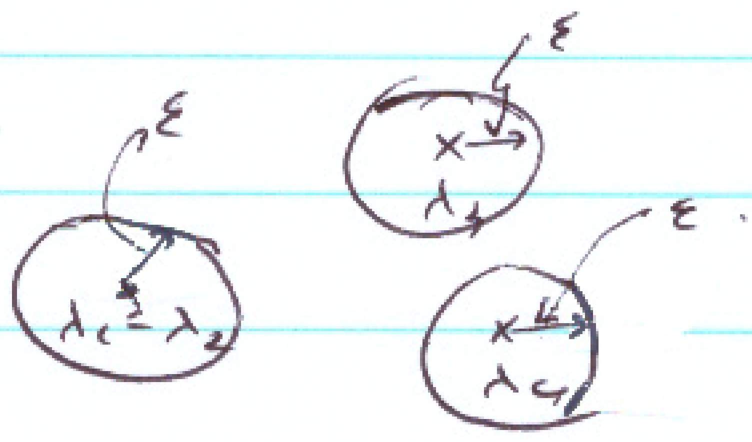
This proves the Proposition:

$$\begin{array}{ccccccc} (v) & \Leftrightarrow & (i) & \Leftrightarrow & (ii) & \Leftrightarrow & (iii) \\ & & & & & & \Downarrow \\ & & & & & & (iv) \end{array}$$

$\square$.

We now begin addressing the following basic questions: How do the eigenvalues of A depend on A? How do the eigenvectors of A depend on A?

Using Rouché's Theorem (see Problem #5, Problem set #7) the eigenvalues of $n \times n$ A are continuous functions of A in the following sense: Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of A, counting (algebraic) multiplicity. Draw disks $D_\varepsilon(\lambda_i)$ of radius $\varepsilon > 0$ around all the distinct eigenvalues λ_i of A, where ε is sufficiently small so that each disk contains only 1 (distinct) eigenvalue of A. For example if $n=4$ and λ_1, λ_3 and λ_4 are distinct, but $\lambda_1 = \lambda_2$, then we have



Then if $\|B-A\|$ is sufficiently small, the eigenvalues of B lie in these disks, and there are as many eigenvalues of B, counting multiplicities, in each of these disks as the multiplicity of the eigenvalue of A at the center of the disk.

So in the example above for $\|B-A\|$ sufficiently small, there are 2 eigenvalues of B, counting multiplicity, in the disk around $\lambda_1 = \lambda_2$, and one eigenvalue in the other 2 disks.

We want to ask finer questions: For example, are the eigenvalues and eigenvectors analytic functions of (the entries) of A?

Note that in general the eigenvalues are not analytic functions of A. For example, for $n=2$ with $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, then $\lambda^2 - t\lambda + \Delta = 0$ is the eigenvalue equation, where $t = a+d$, $\Delta = ad-bc$. Hence

$$\lambda_\pm = \frac{t \pm \sqrt{t^2-4\Delta}}{2}$$

But $t^2 - 4\Delta = (a+d)^2 - 4(ad-bc) = (a-d)^2 + 4bc$.

Now $\sqrt{t^2-4\Delta} = \sqrt{(a-d)^2+4bc}$, is an analytic function of (a,b,c,d) in the neighborhood of any point (a_0,b_0,c_0,d_0) where $(a_0-d_0)^2 + 4b_0c_0 \neq 0$. But if $(a_0-d_0)^2+4b_0c_0 = 0$ then λ is clearly not analytic.

in fact, its derivative blows up at this point. The takeaway is the following: trouble occurs at points at which $\lambda_+ = \lambda_-$, and only at such points.

We consider first the case where $A(\beta) = (A_{ij}(\beta))$ depends analytically on one complex variable β lying in some region $\Omega \subset \mathbb{C}$. The standard example is $A(\beta) = A + \beta B$

where A and B are given $n \times n$ matrices. To see what can happen, consider the following examples (taken from T. Kato, Perturbation Theory for Linear Operators, Chap. II — this is a basic reference for perturbation theory)

(a) $A(\beta) = \begin{pmatrix} 1 & \beta \\ \beta & -1 \end{pmatrix}$

Here the eigenvalues are $\lambda_\pm = \pm\sqrt{1+\beta^2}$
For $\beta = 0$, the eigenvalues are ± 1 so $\lambda_+(0) \neq \lambda_-(0)$,

i.e. the spectrum is simple, and for $|\beta|$ small, we see that $\lambda_+(\beta)$ and $\lambda_-(\beta)$ are analytic functions of β. The spectrum remains simple as long as $\beta \neq \pm i$. The eigenvectors are $v_\pm(\beta) = \begin{pmatrix} -\beta \\ 1 \mp \sqrt{1+\beta^2} \end{pmatrix}$

As long as $\beta \neq \pm i$, the eigenvectors are locally well-defined and analytic and independent. However if $\beta = \pm i$, then the eigenvalues and eigenvectors are no longer analytic at β, and moreover the eigenvectors are no longer independent. Note also that at $\beta = \pm i$, $A(\beta) = A(i) = \begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix}$, is

not diagonalizable: Indeed as $\lambda_+(\pm i) = \lambda_-(\pm i) = 0$, $A(\pm i)$

must be zero if it is diagonalizable, $A(\pm i) = U\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}U^{-1} = 0$,

which is a contradiction.

If β is real, the eigenvectors can be normalized

(126.1) $$\hat{v}_\pm(\beta) = \begin{pmatrix} \beta \Big/ \sqrt{2(1+\beta^2)(\sqrt{1+\beta^2} \mp 1)} \\ \pm\sqrt{(\sqrt{1+\beta^2} \mp 1) \big/ 2(1+\beta^2)} \end{pmatrix}$$

with $\|\hat{v}_\pm(\beta)\| = 1$, in such a way that $\hat{v}_\pm(\beta)$ depends

real analytically on β

(Exercise: Check that $\hat{v}_\pm(\beta)$ are indeed eigenvectors, $\|\hat{v}_\pm(\beta)\| = 1$ and that $\hat{v}_\pm(\beta)$ are real analytic $\forall \beta \in \mathbb{R}$, and $(\hat{v}_+(\beta), \hat{v}_-(\beta)) = 0$.)

Note that $\lambda_+(\beta) \neq \lambda_-(\beta)$ $\forall \beta \in \mathbb{R}$.

(b) $A(\beta) = \begin{pmatrix} 0 & \beta \\ \beta & 0 \end{pmatrix}$. Here the eigenvalues are $\lambda_\pm(\beta) = \pm\beta$.

For $\beta = 0$, both eigenvalues are 0 and the spectrum is

not simple. Nevertheless the eigenvectors are $v_\pm(\beta) = (1, \pm 1)^T$,

which can be normalized $v_\pm$ (trivially) for β real (in this case $\forall \beta$)

$\hat{v}_\pm(\beta) = \left(\frac{1}{\sqrt{2}}, \pm\frac{1}{\sqrt{2}}\right)^T$ such that $\hat{v}_\pm(\beta)$ is analytic for all real β (cf (126.1)) (in fact for all β in this case). Note that at $\beta = 0$, $A(0) = 0$, so that every vector is an eigenvector for $A(0)$. To ensure analytic behavior of $v_\pm(\beta)$ as $\beta \to 0$, however, we cannot choose the eigenvectors of $A(0)$ freely: we must choose special vectors in $N(A(0))$, viz,

(127.1) $$v_\pm(0) = \left(\frac{1}{\sqrt{2}}, \pm\frac{1}{\sqrt{2}}\right)^T$$

This is an illustrative example of singular perturbation theory: Here we are trying to perturb the eigenvalues of $A(0) = 0$ where $A(0) \to A(\beta)$. The problem is singular because the eigenvectors of $A(0)$ are not well-determined. It is the nature of the perturbation $A(0) \to A(\beta)$ that determines which eigenvectors $v_\pm(0)$ of $A(0)$ extend smoothly to $v_\pm(\beta)$.

For example, consider $\mathring{A}(\beta) = \begin{pmatrix} \beta & \beta \\ \beta & 0 \end{pmatrix}$ with eigenvalues

$\lambda_\pm(\beta) = \beta\left(\frac{1 \pm \sqrt{5}}{2}\right)$ and associated eigenvectors

$v_\pm(\beta) = \left(1, \frac{-1 \pm \sqrt{5}}{2}\right)^T$ $\forall \beta$. So again

we are dealing with a singular problem $\mathring{A}(0) = 0$,

and the eigenvectors of $\mathring{A}(0)$ that extend analytically are

$$v_\pm(0) = \left(1, \frac{-1 \pm \sqrt{5}}{2}\right)^T. \tag{128.1}$$

This should be compared with (127.1). The takeaway is the following: the ambiguity in the eigenvectors at $\beta = 0$ is resolved by the nature of the perturbation $A(0) \to A(\beta)$. <u>Singular perturbation problems</u> are very common in physics and applied mathematics: the unperturbed problem is singular in some sense, and the ambiguity in the perturbation problem is resolved by the specific nature of the perturbation.

(c) $A(\beta) = \begin{pmatrix} 0 & \beta \\ 0 & 0 \end{pmatrix}$. Again the problem is singular at $\beta = 0$; $A(0) = 0$ and so the eigenvectors of $A(0)$ are not uniquely determined. However, for $\beta \neq 0$, we still have $\lambda_\pm(\beta) = 0$, but $N(A(\beta)) = \langle \binom{1}{0} \rangle$ and there is no choice of a basis $(v_+(0), v_-(0))$ for $N(A(0))$ which continues analytically $v_\pm(0) \to v_\pm(\beta)$ as $A(0) \to A(\beta)$.

Only $v(\beta) = \binom{1}{0}$ continues (and is analytic). We note that the difference between (b) and (c) is that in case (b), $A(\beta)$ is Hermitian for β real, but not in case (c).

(d) $A(\beta) = \begin{pmatrix} 0 & 1 \\ \beta & 0 \end{pmatrix}$. Here $\lambda_\pm(\beta) = \pm\sqrt{\beta}$. We see that in this case there is no way to choose the branches of $\sqrt{\beta}$ so that $\lambda_\pm(\beta)$ are analytic in β in a neighborhood of $\beta = 0$. Note that here the problem is singular in the sense that $\lambda_\pm(0) = 0$ but $A(0)$ has only 1 eigenvector $v(0) = (1,0)^T$. For $\beta \neq 0$, $v_\pm(\beta) = (1, \pm\sqrt{\beta})^T$

Thus the single eigenvector $v(0) = (1,0)^T$ <u>bifurcates</u> to <u>two</u> eigenvectors $v_\pm(\beta) = (1, \pm\sqrt{\beta})^T$ as $0 \to \beta$. Again we note that even for β real, $\beta \neq 1$, $A(\beta)$ is not Hermitian.

Contrasting the singular perturbation problems (c) and (d), we see that in case (c), A(0) has two eigenvectors, but only one of the eigenvectors continues as $A(0) \to A(\beta)$. However, in case (d), $A(0)$ has only 1 eigenvector, which then bifurcates, <u>non-analytically</u>, to two eigenvectors as $A(0) \to A(\beta)$.

(e) $A(\beta) = \begin{pmatrix} 1 & \beta \\ 0 & 0 \end{pmatrix}$. Here $\lambda_+(\beta) = 1$ and $\lambda_-(\beta) = 0$, which are clearly analytic (in fact constant) functions of β. The eigenvectors are $v_+(\beta) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $v_-(\beta) = \begin{pmatrix} -\beta \\ 1 \end{pmatrix}$, which can be normalized as $\hat{v}_+(\beta) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\hat{v}_-(\beta) = \begin{pmatrix} -\beta/\sqrt{1+\beta^2} \\ 1/\sqrt{1+\beta^2} \end{pmatrix}$ for β real. Note that this is true even as $A(\beta)$ is not Hermitian, $\beta \in \mathbb{R}$.

(f) $A(\beta) = \begin{pmatrix} \beta & 1 \\ 0 & 0 \end{pmatrix}$. Here $\lambda_+(\beta) = \beta$, $\lambda_-(\beta) = 0$, which are clearly analytic in β.
Again the problem is singular at $\beta = 0$, as $\lambda_\pm(0) = 0$.
Here the eigenvectors are $v_+(\beta) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $v_-(\beta) = \begin{pmatrix} 1 \\ -\beta \end{pmatrix}$, which

are analytic in β. Note that, as opposed to case (d), which also has a single eigenvector at $\beta(0)$, and for which the eigenvector bifurcates non-analytically, as $0 \to \beta$, in this case the single eigenvector at $\beta = 0$, bifurcates <u>analytically</u> as $0 \to \beta$. Note also that for β real, the eigenvectors can be normalized in an analytic fashion

$$\hat{v}_+(\beta) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \hat{v}_-(\beta) = \begin{pmatrix} 1/\sqrt{1+\beta^2} \\ -\beta/\sqrt{1+\beta^2} \end{pmatrix},$$

In summary, what we observe / guess from these examples is that everything is "nice", i.e., all the eigenvalues and eigenvectors depend nicely on β in a neighborhood of a simple eigenvalue. When the eigenvalue is multiple, then things are still "nice" in the self-adjoint case $A(\beta) = A^*(\beta)$, $\beta \in \mathbb{R}$, but, <u>in general</u>, not in the non-self-adjoint case (but there are exceptions (see e.g. case (f) above)).

In order to develop perturbation theory in full for $n \times n$ matrices, we need to introduce some concepts/definitions.

References for Perturbation Theory

① T. Kato, as above, p125 ② Reed-Simon, Methods of Modern Math. Physics Vol IV ③ F. Rellich, Perturbation Theory of Eigenvalue Problems

---

We say that a square matrix P is a <u>projection</u> if $P^2 = P$. For a projection P set

$$M_P = R(P) \quad , \quad N_P = N(P).$$

Clearly $R(P)$ and $N(P)$ are subspaces.

Note that if $x \in R(P)$, then $x = Py$ for some y and so $Px = P^2 y = Py = x$. Thus

$$R(P) = \{x : x = Px\}$$

<u>Proposition</u> Let $V = \mathbb{R}^n$ or $\mathbb{C}^n$. and let $P: V \to V$ be a projection. Then V has a <u>direct</u> <u>decomposition</u>

$$V = R(P) \oplus N(P)$$

ie each $x \in V$ has a <u>unique</u> representation

$$x = r + n$$

where $r \in R(P)$ and $n \in N(P)$.

Conversely, if
$$V = X \oplus Y$$
is a direct decomposition of V, then
$$X = R(P) \quad \text{and} \quad Y = N(P)$$
for some projection $P : V \to V$.

Proof: If $x \in V$, then
$$x = Px + (1-P)x$$
Clearly $Px \in R(P)$ and as $P(1-P)x = (P-P^2)x$
$$= (P-P)x = 0,$$
we see that $(1-P)x \in N(P)$.

Now suppose $r + n = r' + n'$ where $r, r' \in R(P)$ and $n, n' \in N(P)$. Then $r - r' = n' - n$

so $Pr - Pr' = P(n') - P(n) = 0$. Hence $r = Pr = Pr' = r'$. But then $n = n'$. Thus $R(P) \oplus N(P)$ is indeed a direct decomposition.

Conversely, suppose
$$V = X \oplus Y$$
is a direct decomposition of V into subspaces X & Y.

Set
$$Px = x \quad \text{if } x \in X$$
$$Py = 0 \quad \text{if } y \in Y.$$

Then $P^2(x+y) = P(Px + Py) = PPx = Px$ as $Px \in X$ & so $P(Px) = Px$. But $Py = 0$ and so

$P^2(x+y) = P(x+y)$ as $P^2 = P$.

Now suppose $x \in X$. Then $x = Px \in R(P)$, so $X \subset R(P)$. On the other hand, if $x \in R(P)$, then $x = Pu$ for some $u \in V$. But $u = x' + y'$ for some $x' \in X$, $y' \in Y$. Hence $x = Pu = Px' + Py' = Px' = x' \in X$. Thus $R(P) \subset X$ and so $X = R(P)$.

If $y \in Y$, then $Py = 0$ and so $y \in N(P)$. Thus $Y \subset N(P)$. On the other hand, if for some y $Py = 0$, then $y = x' + y'$, $x' \in X$, $y' \in Y$.

Hence $0 = Py = Px' + Py' = Px' = x'$. So $x' = 0$ and hence $y = y' \in Y$. Thus $Y = N(P)$, which proves the result. $\square$.