Pos Def Matrices

6.1 Calculus Detour – Minima, Maxima, Saddle pts are determined by a quadratic form.

Strang considers 2 fcns: a general non-linear $F: \mathbb{R}^2 \longrightarrow \mathbb{R}$
$(x,y) \longmapsto 7 + 2(x+y)^2 - y\sin y - x^3$

and a quadratic $f: \mathbb{R}^2 \longrightarrow \mathbb{R}$
$(x,y) \longmapsto \underset{a}{2}x^2 + \underset{2b}{4}xy + \underset{c}{}y^2 = [x\ y]\begin{bmatrix}2 & 2\\ 2 & 1\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix} = x^T A x$

Both of these have a critical pt at 0: $DF_{00} = 0 = Df_{00}$

And the 2^{nd} derivs are the same: $D^2F_{(00)} = \begin{bmatrix}4 & 4\\ 4 & 2\end{bmatrix} = D^2 f_{(00)}$

[Remark: $D^2 f = 2A$ see my sheets for M&T ch 4.2]

Strang asserts that the quadratic part of the Taylor series expansion dominates in the nhbd of a critical pt [cf M&T ch 4.2 Sheets and Morse Theory in G&P] and thus determines maxima and minima.

▷ so we are led to consider $f(x,y) := ax^2 + 2bxy + cy^2 = [x\ y]\begin{bmatrix}a & b\\ b & c\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix}$

What conds on a, b, c would guarantee f always pos when $\langle x,y\rangle \neq \langle 0,0\rangle$?

Note A is symm, but more on that later.

Just plug in $\langle 1,0\rangle$ and you see we requm $a>0$ [Likewise $\langle 0,1\rangle$ shows $c>0$ but we will get that for free momentarily]

Write f as the sum of 2 squares by completing the square:

$$f(x,y) = ax^2 + 2bxy + cy^2 = a\left(x^2 + \frac{2b}{a}xy\right) + cy^2$$
$$= a\left(x^2 + \frac{2b}{a}yx + \frac{b^2}{a^2}y^2\right) - \cancel{a}\frac{b^2 y^2}{a^{\cancel{2}}} + cy^2$$
$$= \underbrace{a\left(x + \frac{b}{a}y\right)^2}_{\text{This is always pos if } a>0} + \underbrace{\left(c - \frac{b^2}{a}\right)}_{\text{This is pos if } c - \frac{b^2}{a} > 0}y^2$$

$\Rightarrow c > \frac{b^2}{a}$

i.e. $ac - b^2 > 0$ $\det(A) > 0$

~~So~~ f is obviously a min at $(0,0)$ if f always pos away from Origin.

Note also that $g := -f$ would have a max at $(0,0)$ and the conds would become
'a' $= -a$ always neg
'$ac-b^2$' $= (-a)(-c) - (-b)^2 = ac - b^2 > 0$ same

Strang decides he will not consider the very degenerate case $D^2 f = \begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix}$ or even $\det A = 0$. But $\det A < 0$ is allowed and that corresponds to a saddle pt like for $f(x,y) = x^2 - y^2$ $\begin{bmatrix}1 & \\ & -1\end{bmatrix}$ More on this later.

▷ Now for the general quadratic $f: \mathbb{R}^n \longrightarrow \mathbb{R}$
$x \longmapsto x^T A x = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j$

[To have no mixed terms, only $x^2\ y^2\ z^2$ means A is diag.]

Def A is positive definite if $x^T A x > 0\ \forall x \neq 0$

Strang considers only symmetric matrices A when he says pos definite. There can be others, like $A = \begin{bmatrix}2 & 1\\ 0 & 3\end{bmatrix}$ but it is not nec to consider them because any quad form $x^T B x = x^T A x$ where B non-symm and A symm matrix. I will establish these claims on next page ⟶

▷ Show $\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}$ is pos def: $[x\ y]\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = 2x^2 + 1xy + 3y^2 > 0 \quad \forall xy \neq 0,0$

because $xy < 2x^2 + 3y^2$

To see this, for pos a, b $a \neq b$ $0 < (b-a)^2 = b^2 - 2ab + a^2$

$2ab < a^2 + b^2$

Thus $xy \leq |x||y| < 2|x||y| < x^2 + y^2 < 2x^2 + 3y^2$

Claim: For any non symm square B such that $f(x) = x^T B x$, $\exists$ symm A such that $x^T B x = x^T A x$. [Marsden & Tromba ch 4.2 sheets]

pf fix i, j. Then $x_i x_j = x_j x_i$ so $x_i x_j b_{ij} + x_j x_i b_{ji} = x_i x_j (b_{ij} + b_{ji})$

$$= x_i x_j \Big(\underbrace{\frac{(b_{ij} + b_{ji})}{2}}_{a_{ij}} + \underbrace{\frac{(b_{ij} + b_{ji})}{2}}_{a_{ji}} \Big) \quad \square$$

$\boxed{\text{ch 6.2}}$

Thm A symm pos def

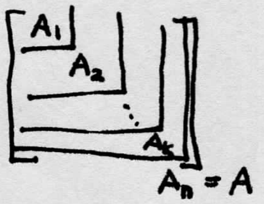
T.F.A.E. (I) $x^T A x > 0 \ \forall x \neq 0$

$\Leftrightarrow$ (II) All EWs λ of A satisfy $\lambda > 0$

$\Leftrightarrow$ (III) All upper left submatrices A_k have $\det(A_k) > 0$

$\Leftrightarrow$ (IV) All pivots satisfy $d_i > 0$ (with no row exchanges)

$\Leftrightarrow$ (V) $\exists$ matrix R with LI cols $\ni A = R^T R$



pf (I) $\Rightarrow$ (II) Let $Ax = \lambda x$ where $\|x\| = 1$. Then $0 < x^T A x = x^T \lambda x = \lambda$ $\square$

(I) $\Leftarrow$ (II) Since A symm, Spectral Thm says there is a basis for $\mathbb{R}^n$ of O.N. EVs $\{x_1, \ldots, x_n\}$

For any vector u, $u = \sum^n c_i x_i$

$Au = c_1 A x_1 + \ldots + c_n A x_n = c_1 \lambda_1 x_1 + \ldots + c_n \lambda_n x_n$

$u^T A u = (c_1 x_1^T + \ldots + c_n x_n^T)(c_1 \lambda_1 x_1 + \ldots + c_n \lambda_n x_n) \overset{O.N.}{=} c_1^2 \lambda_1 + \ldots + c_n^2 \lambda_n > 0$ $\square$

(I) $\Rightarrow$ (III) we can isolate submatrices from the upper left corner (elt a_{11}) by using vectors with only the leading elts non-zero:

choose any $k \in \{1, \ldots, n-1\}$

$$\underbrace{[u^1\ u^2 \cdots u^k\ 0 \cdots 0]}_{\vec{u}_k} \left[\begin{array}{c|c} A_k & * \\ \hline * & * \end{array}\right] \begin{bmatrix} u^1 \\ \vdots \\ u^k \\ 0 \\ 0 \end{bmatrix} = u_k^T A_k u_k > 0 \text{ by (I)}$$

All EWs $\lambda_i^{(k)}$ of A_k are pos by (II)

det is product of EWs

$\Rightarrow \det A_k = \lambda_1^{(k)} \cdots \lambda_k^{(k)} > 0$

Remarks: There is nothing special about <u>upper left</u> submats.

Any chain of principal submats would work, starting at any a_{ii} on main diag and growing by adding a row and column pair each time.

$\Rightarrow$ this does show that a <u>nec cond</u> for pos def is all main diag elts pos.

cont'd →

(III) $\Rightarrow$ (IV) From ch 4 section 4 we know $\det A_k = d_1 \cdot d_2 \cdots d_k$ — a priori, a pivot is not zero

Then $$d_k = \frac{(d_1 \cdot d_2 \cdots d_{k-1}) \cdot d_k}{(d_1 \cdot d_2 \cdots d_{k-1})} = \frac{\det A_k > 0}{\det A_{k-1} > 0} \Rightarrow d_k \text{ pos } \forall k$$

(IV) $\Rightarrow$ (I) since A is symm, $A = LDL^T$ where D is diag matrix of pivots.

$$x^TAx = x^T(LDL^T)x = (x^TL)D(L^Tx)$$
$$= d_1\left(\left(L^Tx\right)^{(1)}\right)^2 + d_2\left(\left(L^Tx\right)^{(2)}\right)^2 + \dots + d_n\left(\left(L^Tx\right)^{(n)}\right)^2$$

This is always pos, since each d_i pos.

(I) $\Leftarrow$ (V) $A = R^TR$ where R has LI columns

$x^TAx = x^TR^TRx = \|Rx\|_2^2 > 0$ if $x \neq 0$

and we know $Rx \neq 0$ if $x \neq 0$ because R has LI cols $\Rightarrow \ker(R) = \{0\}$

(I) $\Rightarrow$ (V) There are many valid choices for R

$$A = LDL^T = \underbrace{(LD^{1/2})}_{R^T}\underbrace{(D^{1/2}L^T)}_{R} \quad \text{or} \quad A = Q\Lambda Q^T \text{ spectral Thm} = \underbrace{(Q\Lambda^{1/2})}_{R^T}\underbrace{(\Lambda^{1/2}Q^T)}_{R}$$

and R can be rectangular if we like:

Take any valid square R (as found above) and a matrix $Q = [\quad]$ with O.N. cols

Rectangular $\mathcal{R} := QR$

then $\mathcal{R}^T\mathcal{R} = R^T[\,Q^T\,][\,Q\,]R = R^TR = A$ □

In ch 3.4 we studied linear Least Sq and Normal eqs $A^TA\bar{x} = A^Tb$

This is now known to be symm pos def.

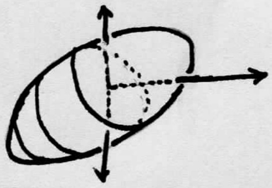
A^TA is symm $\xRightarrow{\text{Spectral}}$ $\exists$ full set of O.N. EVs $q_1, \dots q_n$

$A^TAq_j = \lambda_j q_j \Rightarrow q_j^TA^TAq_j = \lambda_j q_j^Tq_j \Rightarrow 0 \le \|Aq_j\|_2^2 = \lambda_j \Rightarrow$ □

▷ Ellipsoids in n-dim

Thm Let A be symm pos def
$\mathcal{E} := \{\text{all } x \in \mathbb{R}^n \mid x^TAx = 1\}$ $\Longrightarrow$

1. $\mathcal{E}$ is n-dim ellipsoid
2. The EVs of A correspond to the axes of $\mathcal{E}$ and the EWs define the lengths of these axes inside $\mathcal{E}$; the i^{th} axes of $\mathcal{E}$ has length $\frac{1}{\sqrt{\lambda_i}}$
3. a rotation matrix Q would rotate $\mathcal{E}$ to align with the co-ord axes [or phrase it as a COV $y = Q^Tx$ aligns $\mathcal{E}$]



Pf Since A is Symm, by Spectral Thm $\exists$ O.N. matrix Q such that $A = Q\Lambda Q^T$. Define COV $y = Q^Tx$, $x = Qy$

$1 = x^TAx = y^TQ^TAQy = y^T\Lambda y = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \dots + \lambda_n y_n^2$

A pos def $\Rightarrow$ all EWs λ are pos $\Rightarrow$ This is the cannonical eq of an ellipsoid.

Cont'd →

The cols of Q are EVs of A.

Fix i. Plug in 0 for all y values except y_i: $0+0+\dots\lambda_i y_i^2+0=1$

$\Rightarrow y_i = \pm\frac{1}{\sqrt{\lambda_i}}$ $\lambda_i>0$

length of i "semi axis" is $\frac{1}{\sqrt{\lambda_i}}$

Finally, to show Q is a rotation:
we already know Q is O.N. so $\det Q=\pm 1$.
If $+1$ we are done since that is def of rotation. If $\det Q=-1$, replace the first col $\hat{q}_1 \rightsquigarrow -\hat{q}_1$ $\tilde{Q}=\begin{bmatrix} | & | & & | \\ -q_1 & q_2 & \cdots & q_n \\ | & | & & | \end{bmatrix}$ then $\det\tilde{Q}=+1$ and since $A(-q_1)=\lambda(-q_1)$,
this new matrix $\tilde{Q}$ works for all the other steps $\square$

Remark: If A were not pos def, we could get other quadric surfaces like paraboloids, hyperboloids, etc depending on signs of EWs and if any would be 0.

6.2.4 If A symm pos def, so is A^2, A^{-1} (in fact A^p) because $\lambda \rightsquigarrow \lambda^p$ always pos if λ is.

6.2.5 If A, B symm pos def, so is $A+B$ because $x^T(A+B)x = x^TAx + x^TBx > 0$

6.2.8 A symm pos def, C nsing $\Rightarrow B=C^TAC$ is symm pos def (special case of Sylvester's Law of Inertia)

- $B^T=(C^TAC)^T = C^TA^TC = C^TAC = B$ ✓
- For any $y\neq 0$, $y^TBy = y^TC^TACy = (Cy)^TA(Cy)>0$ because $Cy\neq 0$ nsing.

6.2.9 If A symm pos def, Cauchy-Schwartz generalizes to $|\langle x, Ay\rangle|^2 \le \langle x, Ax\rangle\langle y, Ay\rangle$

i.e. $|x^TAy|^2 \le x^TAx\, y^TAy$

We know $A=R^TR$

$$|x^TAy|^2 = |x^TR^TRy|^2 = |(Rx)^TRy|^2 \overset{\text{regular CS}}{\le} (Rx)^TRx(Ry)^TRy = x^TR^TRx\,y^TR^TRy = x^TAx\,y^TAy \quad \square$$

Ch 6.3 Semi def and Indefinite matrices Also $Ax=\lambda Mx$

- Now we want to generalize prev Thm on pos def matrices to pos <u>semi</u> def
- Prove Sylvester's Law of Inertia: Signs of EWs match signs of pivots
- Introduce generalized EW problem: $Ax=\lambda Mx$

Thm 6E A symm pos semidef [cf Thm ch 6.2]

TFAE (I′) $x^TAx \geq 0 \quad \forall x$

(II′) All EWs satisfy $\lambda \geq 0$

(III′) All principal submatrices have $\det(A_k) \geq 0$

(IV′) All pivots satisfy $d_i \geq 0$

(V′) $\exists$ matrix R, possibly with LD cols, $\ni A = R^TR$

Pf (I′)→(II′) same as before: If $Ax = \lambda x$, take $\|x\| = 1$ $\quad 0 \leq x^TAx = x^T\lambda x = \lambda$

(I′)←(II′) Same as before: A symm $\overset{\text{Spectre}}{\Longrightarrow} A = Q\Lambda Q^T \qquad y := Q^Tx$

$$x^TAx = x^TQ\Lambda Q^Tx = y^T\Lambda y = \sum^n \lambda_i y_i^2 \geq 0 \quad \text{since each } \lambda_i \geq 0$$

[If A nsing, no $\lambda = 0$ so to get semi-def A must be sing.]

(I′)⇐(V′) $A = R^TR$ but cols of R not nec LI

$x^TAx = x^TR^TRx = \|Rx\|_2^2 \geq 0$ since $Rx = 0$ possibly even when $x \neq 0$

(I′)⇒(V′) same as before, I'll just give one possibility $A = Q\Lambda Q^T$ spectral

$$= (Q\Lambda^{1/2})(\Lambda^{1/2}Q^T)$$
$$= R^T \; R$$

Strang sketches an arg for (III′) and (IV′) similar to the prev thm, but it involves row exchanges and is not entirely clear. He also gives a homotopy arg to derive II′, III′, IV′ from prev thm:

Let $A(\varepsilon) := A + \varepsilon I$ $\qquad$ $A(\varepsilon)$ is in fact symm pos def because $x^TA_\varepsilon x = \underset{\geq 0}{x^TAx} + \underset{>0}{\varepsilon^2\|x\|^2}$

(II′) $A(\varepsilon)$ has pos $\lambda_i(\varepsilon)$ for all $\varepsilon > 0$ and $\lambda(\varepsilon)$ is a smooth fcn of ε since $\det\left[(A+\varepsilon I) - \lambda I\right] = 0$ is smooth poly in elts of $A(\varepsilon)$

$$\lim_{\varepsilon \searrow 0} \lambda_i(\varepsilon) = \lambda_i \geq 0$$

(III′) $\det(A_k(\varepsilon)) > 0 \;\vdots\; \det(A_k) = \lim_{\varepsilon \searrow 0} \left[\det(A_k(\varepsilon))\right] \geq 0$

(IV′) Same idea for pivots: $d_i(\varepsilon) > 0 \quad \lim \Rightarrow d_i \geq 0$

Thm 6E A symm pos semidef [cf Thm ch 6.2]

TFAE (I′) $x^T A x \ge 0 \quad \forall x$

(II′) All EWs satisfy $\lambda \ge 0$

(III′) All principal submatrices have $\det(A_k) \ge 0$

(IV′) All pivots satisfy $d_i \ge 0$

(V′) $\exists$ matrix R, possibly with LD cols, $\ni A = R^T R$

Pf (I′)$\Rightarrow$(II′) same as before: If $Ax = \lambda x$, take $\|x\| = 1$ $\quad 0 \le x^T A x = x^T \lambda x = \lambda$

(I′)$\Leftarrow$(II′) Same as before: A symm $\overset{\text{Spectre}}{\Longrightarrow}$ $A = Q\Lambda Q^T$ $\quad y := Q^T x$

$$x^T A x = x^T Q \Lambda Q^T x = y^T \Lambda y = \sum^n \lambda_i y_i^2 \ge 0 \text{ since each } \lambda_i \ge 0$$

[If A nsing, no $\lambda = 0$ so to get semi-def A must be sing.]

(I′)$\Leftarrow$(V′) $A = R^T R$ but cols of R not nec LI

$x^T A x = x^T R^T R x = \|Rx\|_2^2 \ge 0$ since $Rx = 0$ possibly even when $x \ne 0$

(I′)$\Rightarrow$(V′) same as before, I'll just give one possibility $A = Q\Lambda Q^T$ spectral

$$= (Q\Lambda^{1/2})(\Lambda^{1/2} Q^T) = R^T R$$

Strang sketches an arg for (III′) and (IV′) similar to the prev thm, but it involves row exchanges and is not entirely clear. He also gives a homotopy arg to derive II′, III′, IV′ from prev thm:

Let $A(\varepsilon) := A + \varepsilon I$ $\quad$ $A(\varepsilon)$ is in fact symm pos def because $\underbrace{x^T A_\varepsilon x}_{} = \underbrace{x^T A x}_{\ge 0} + \underbrace{\varepsilon^2 \|x\|^2}_{> 0}$

(II′) $A(\varepsilon)$ has pos $\lambda_i(\varepsilon)$ for all $\varepsilon > 0$ and $\lambda(\varepsilon)$ is a smooth fcn of ε since $\det\left[(A + \varepsilon I) - \lambda I\right] = 0$ is smooth poly in elts of $A(\varepsilon)$

$$\lim_{\varepsilon \searrow 0} \lambda_i(\varepsilon) = \lambda_i \ge 0$$

(III′) $\det(A_k(\varepsilon)) > 0$ $\vdots$ $\det(A_k) = \lim_{\varepsilon \searrow 0} \left[\det(A_k(\varepsilon))\right] \ge 0$

(IV′) Same idea for pivots: $d_i(\varepsilon) > 0$ $\quad$ $\lim \Rightarrow d_i \ge 0$

▷ what are the elementary operations and invariants associated with a quadratic form $q(x) = x^TAx$?

Given a nsing matrix C, we consider cov $x = Cy$ which gives us the

<u>Congruence transform</u> $A \longmapsto C^TAC$

Symmetry is preserved: $(C^TAC)^T = C^TA^TC = C^TAC$

Thm $\boxed{\text{Sylvesters Law of Inertia}}$

$n \times n$ matrices: A symm, C nsing $\Longrightarrow$ C^TAC has the same 'signature' as A — same number of pos, neg, and 0 EWs

pf

$\boxed{\text{Case 1}}$ <u>A is nsing</u> (no zero EWs)

<u>IDEA</u>: We will define a homotopy from A to C^TAC. But for this to work, we actually need the homotopy $Q^TAQ \rightsquigarrow C^TAC$ where Q is an ON matrix we will find.

<u>Step 1</u> By Gram-Schmidt we can factorize $C = QR$ — Q ON, R upper triang with strictly pos elts on main diag

Define $C_t := (1-t)Q + tQR \quad t \in [0,1]$

$$= Q\underbrace{\left[(1-t)I + tR\right]}$$

This matrix is nsing because it is upper triang with strictly pos elts on main diag $\left[(1-t)1 + t\|a_j\|^2 \text{ is always pos}\right]$ $\Rightarrow \det > 0 \Rightarrow$ nsing

Thus $\forall t \in [0,1]$ C_t is nsing. $C_0 = Q \qquad C_1 = QR = C$

<u>Step 2</u> Define $B_t := C_t^TAC_t$ then $B_0 = Q^TAQ$ and $B_1 = C^TAC$

For each value of t, B_t is nsing. EWs are smooth fcns of t since $\det[B_t - \lambda_i I] = 0$ is a poly

<u>Step 3</u> A has p pos and r neg EWs

$Q^{-1}AQ = Q^TAQ$ has exactly the <u>same EWs</u>, since similiar matrices have same EWs.

$B_0 = Q^TAQ$

For all $t \in [0,1]$ B_t is nsing, thus even though each EW $\lambda_i(t)$ of B_t may move with t, none can touch 0, much less cross over it.

$B_1 = C^TAC \longleftarrow$ so this has same <u>number</u> p of pos EWs and r neg EWs □

$\boxed{\text{Case 2}}$ A singular — Strang says to work with nsing $A + \varepsilon I$ and $A - \varepsilon I$ and at the end let $\varepsilon \searrow 0$

[I will give Strang's other pf on sheet (8) which does not use homotopy, just algebra.]

□

COR A symm $\Rightarrow$ Signature of pivots matches signature of EWs
$A = LDL^T$ $\qquad$ $A = Q\Lambda Q^T$

Pf we know Gaussian elimination gives us $A = LDU$ where L is lower triang and has all 1's on main diag $\Rightarrow$ det $=1$ $\Rightarrow$ invertible.
To get $A = LDL^T$ (which is not exactly Cholesky) observe $A^T = U^T D^T L^T$ and Symm $A = A^T$
$\Rightarrow U = L^T$

step 2 $A = LDL^T \Leftrightarrow L^{-1} A L^{-T} = D$
Define $C := (L^{-1})^T$ then $C^T A C = D$ and we can apply Sylvester LOI to conclude A and D have same sig. D is matrix of pivots $\square$

$\triangleright$ Now Strang gives a partial example of finding EWs by Given's method using these ideas. Actually Givens does rotations to make a matrix tridiagonal and here we start with A already tri-diag.

$$A = \begin{bmatrix} 3 & 3 & 0 \\ 3 & 10 & 7 \\ 0 & 7 & 8 \end{bmatrix}$$

Find pivots by Gaussian Elimination
$R_2 \rightsquigarrow R_2 - R_1 = R_2'$
$R_3 \rightsquigarrow R_3 - R_2'$

$$\begin{bmatrix} 3 & 3 & 0 \\ 0 & 7 & 7 \\ 0 & 0 & 1 \end{bmatrix}$$

so $D = \begin{bmatrix} 3 & & \\ & 7 & \\ & & 1 \end{bmatrix}$ all pos elts

Guess:
Is $\lambda = 2$ an EW? we would form $\det[A - 2I] \stackrel{!}{=} 0$ to check, but Strang wants to compute pivots:

$$A - 2I = \begin{bmatrix} 1 & 3 & 0 \\ 3 & 8 & 7 \\ 0 & 7 & 6 \end{bmatrix}$$

$R_2 \rightsquigarrow R_2 - 3R_1$
$R_3 \rightsquigarrow R_3 - 7R_2'$

$$\begin{bmatrix} 1 & 3 & 0 \\ 0 & -1 & 7 \\ 0 & 6 & 54 \end{bmatrix}$$

1 neg pivot
This violates Sylvester
so rule out $\lambda = 2$

Try $\lambda = 1 \rightsquigarrow$ again neg pivot

$\lambda = \frac{1}{2} \xrightarrow{\text{Prob 6.3.5}}$ pivots $= \{2.5, 5.9, -0.81\}$ again neg pivot

Keep going by bisection method to approx the EW

$A = A - 0I$ so we know it is between $(0, \frac{1}{2})$... (between endpts of this interval)

algebraic pf of Sylvester's Law of Inertia (problems 6.3.6 and 6.3.7 out of order)

pf.

Lemma If domain of A has a basis of EVs $\{u_1 \dots u_n\}$ and $u_1, \dots, u_r$ are associated with EW $\lambda = 0$ $\Rightarrow$ $\{u_1 \dots u_r\}$ are a basis for $\ker(A)$. So $\dim(\ker(A)) = r$. Any $w = \sum_{i=r+1}^{n} \alpha_i u_i$ is NOT in $\ker(A)$

pf. For $u = \sum_{i=1}^{r} \alpha_i u_i$ $\quad A(u) = \sum \alpha_i A(u_i) = \sum \alpha_i 0 u_i = 0$

For $w = \sum_{i=r+1}^{n} \alpha_i u_i$, all $\alpha_i \neq 0$ $\quad A(w) = \sum \alpha_i A(u_i) = \sum \alpha_i \lambda_i u_i \neq 0$ because $\{u_i\}_{r+1}^{n}$ is LI $\square$

Lemma 2 $\operatorname{rank}(A) = \operatorname{rank}(C^TAC)$ C nsing, A symm

Equivalently A has the same number of 0 EWs (multiplicity) as C^TAC

pf. we will show $\dim(\ker(A)) = \dim(\ker(C^TAC))$ $\qquad$ rank + nullity $= n$

Domain of A has EV basis $\{q_1 \dots q_n\}$ and since C is nsing, there is a basis $\{s_1 \dots s_n\}$ $\ni$ $Cs_i = q_i$. If the basis for $\ker(A)$ is $q_1 \dots q_r$, then $C^TACs_i = C^TAq_i = C^T0 = 0$ $\forall i = 1, \dots, r$

so $\dim(\ker(C^TAC)) \geq \dim(\ker(A))$.

Can any other vector w not in span $\{s_1 \dots s_r\}$ be in $\ker(C^TAC)$? $w = \sum_{r+1}^{n} \alpha_i s_i$ $\quad Cw = \sum_{r+1}^{n} \alpha_i q_i$

and from first lemma $A\left(\sum_{r+1}^{n} \alpha_i q_i\right) \neq 0$ $\square$

So now we have only to show that A and C^TAC have same numbers of pos and neg EWs. Both matrices are symm, so both have ON bases.

Let $\{x_1 \dots x_p\}$ be the ON EVs of A assoc with λ_i pos

$\{y_1 \dots y_q\}$ " " " " " C^TAC " " μ_i neg

Step 1 show $\{x_1, \dots x_p, Cy_1 \dots Cy_q\}$ are LI.

Suppose not: If LD, then $\sum^{p} a_i x_i + \sum^{q} (-b_j) Cy_j = 0 \Rightarrow \sum^{p} a_i x_i = \sum^{q} b_j Cy_j$ $\quad$ Call this common vector z

$z^TAz \geq 0$ because $Az = \sum a_i (\lambda_i x_i)$ and $z^TAz = \sum^{p} \lambda_i a_i^2 \geq 0$ since $x_i \cdot x_j = 0$ O.N. since λ_i pos

AND

$z^TAz \leq 0$ because $Az = \sum^{q} b_i ACy_i = \sum b_i \mu_i C^{-T} y_i$ $\quad$ because $C^TACy = \mu y \Rightarrow ACy = \mu C^{-T} y$

$z^TAz = \mu_1 b_1^2 y_1^T \underbrace{C^TC^{-T}}_{I} y_1 + \mu_1 \mu_2 b_1 b_2 y_1^T C^TC^{-T} y_2 + \dots$

$= \mu_1 b_1^2 1 + \mu_2 b_2^2 + \dots + \mu_q b_q^2 \leq 0$ because each μ_i neg.

$\Downarrow$

$z^TAz = 0$ but this requires every $a_i = 0$, $b_j = 0 \Rightarrow \{x_1 \dots x_p\ Cy_1 \dots Cy_q\}$ is LI.

step 2 $\{x_1 \dots x_p, Cy_1 \dots Cy_q\}$ LI in $\mathbb{R}^n \Rightarrow p + q \leq n$

step 3 Apply same arg to $(n-p)$ neg λ_i's and $(n-q)$ pos μ_j's. $\Rightarrow (n-p) + (n-q) \leq n$
$\Rightarrow n \leq p + q$

This combined with prev step shows $p + q = n \Rightarrow$ $n - p = q$ so same # pos EWs!
$n - q = p$ " " " neg "

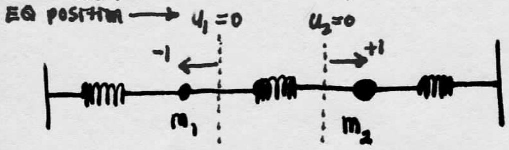
Schuams LA Thm 12.4, 12.5 uses basically this same pf, but without EVs since they don't have Spectral Thm yet.

$\boxed{\text{QED}}$

Generalized EW Problem

$Ax = \lambda x$ becomes $Ax = \lambda M x$ and we want A & M both symm.

Consider mass pts in a lattice of springs governed by Hooke's Law $F = -kx$, $m\ddot{x} = -kx$



unequal masses lead to M matrix
here $m_1 = 1$ $m_2 = 2$

$$\begin{aligned} m_1 \ddot{u}_1 &= \\ m_2 \ddot{u}_2 &= \end{aligned} \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

$$\begin{bmatrix} m_1 & \\ & m_2 \end{bmatrix} \begin{bmatrix} \ddot{u}_1 \\ \ddot{u}_2 \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

Assume soln $\vec{u} = e^{i\omega t} \vec{x}$
like p. 284 ch 5.4

$$\dot{u} = i\omega e^{i\omega t} x$$
$$\ddot{u} = \underbrace{(i\omega)^2}_{-\omega^2} e^{i\omega t} x$$

$$M\ddot{u} = Au$$
$$M(-\omega^2)\cancel{e^{i\omega t}} x = A\cancel{e^{i\omega t}} x$$
$$M\lambda x = Ax \Rightarrow Ax = \lambda M x$$
$$Ax - \lambda M x = 0$$
$$(A - \lambda M)x = 0$$

So we solve $\det(A - \lambda M) = 0$

Here $\det \begin{bmatrix} -2-\lambda & 1 \\ 1 & -2-2\lambda \end{bmatrix} = 0$ EVs give normal modes of oscillation.

Now if M is assumed <u>Symm Pos def</u> $M = R^T R \Rightarrow Ax = \lambda R^T R x$

$C := R^{-1} \Rightarrow C^T A C y = \lambda y$ with COV $y = Rx$

The properties of symm C^TAC lead directly to corresp props for $Ax = \lambda M x$

1. EWs are Real
2. Same signs for EWs by Sylvester LOI.
3. EVs of C^TAC are ON $\Rightarrow$ EVs of A are 'M-O.N.' $x_i^T M x_j = \delta_{ij}$
4. $x_i^T A x_j = \lambda_j x_i^T M x_j = \begin{cases} \lambda_j \\ 0 \end{cases}$

   Matrices A & M are being simultaneously diagonalized.

$$S = \begin{bmatrix} | & & | \\ x_1 & \cdots & x_n \\ | & & | \end{bmatrix} \Rightarrow S^TAS = \Lambda \qquad S^TMS = I$$

Not S^{-1}; not similiarity transform

Strang says hard to visualize in general. In pos def case, $x^TAx = 1$ $x^TMx = 1$ are ellipsoids

$x = Sz$ gives COV that makes them correctly aligned (not a pure rotation)
S not ON

$$x^TAx = z^TS^TASz = \sum \lambda_i z_i^2 = 1$$
$$x^TMx = \sum z_i^2 = 1 \quad \text{Sphere.}$$

Minimum Principles and Rayleigh Quotient

Find a minimum principle that yields $Ax=b$ and then one that yields $Ax=\lambda x$. ch 1-dim $p(x)=\frac{1}{2}ax^2-bx$ parabola $p'(x)=ax-b\overset{!}{=}0$ $p''(x)=a$ a pos $\Rightarrow$ concave up $\Rightarrow$ minimum

n-dim: we want to find a paraboloid $\mathcal{P}$ which has a minimizing pt x that is found by $Ax=b$

Given A,b define $P_{A,b}:\mathbb{R}^n \longrightarrow \mathbb{R}$
$x \longmapsto \frac{1}{2}x^TAx-x^Tb = \frac{1}{2}x^T(Ax-2b)$ Paraboloid in $\mathbb{R}^n\times\mathbb{R}$

Thm If A symm pos def $\Rightarrow$ P attains its min at x^* where $Ax^*=b$
The min value is $P(x^*)=-\frac{1}{2}b^TA^{-1}b$

Pf. Let x^* be soln to $Ax=b$. Let $y\in\mathbb{R}^n$ arb pt. not x^*

$$P(y)-P(x^*)=\frac{1}{2}y^TAy-y^T\underset{Ax^*}{\underset{\downarrow}{b}}-\left[\frac{1}{2}x^TAx-x^T\underset{Ax^*}{\underset{\downarrow}{b}}\right]$$

$$=\frac{1}{2}y^TAy-y^TAx\underbrace{-\frac{1}{2}x^TAx+x^TAx}_{\frac{1}{2}x^TAx} = \frac{1}{2}(y-x)^TA(y-x)>0$$

since A pos def and $y\neq x$

$\Rightarrow P(y)-P(x^*)>0$ for all $y\neq x^*$

$\Rightarrow P(y)>P(x^*)$ and thus x^* is the minimizer. $\square$

6.4.4 Another quadratic that has minimum for $Ax=b$ is $q(x)=\frac{1}{2}\|Ax-b\|_2^2$
$=\frac{1}{2}x^TA^TAx-x^TA^Tb+\frac{1}{2}b^Tb$

We discussed this in ch 3.2 sheets ['Linear Alg with OG' sheet 2]

$Dq_x(h)=2x^T(A^TA)h-h^TA^Tb-b^TAh\overset{!}{=}0$ for extrema

$x^T(A^TA)h=b^TAh$ for any h

$\Rightarrow A^TAx=A^Tb$ Normal Eqs

Remark In my sheets on Quadrics, I have 2 examples of paraboloids

elliptic: $\frac{1}{\alpha^2}x^2+\frac{1}{\beta^2}y^2=cz$ $c>0$

hyperbolic: $\frac{1}{\alpha^2}x^2-\frac{1}{\beta^2}y^2=cz$ this is saddle point

Now lets consider $Ax = \lambda x$

Define Rayleigh Quotent $R_A : \mathbb{R}^n - \{0\} \longrightarrow \mathbb{R}$, $\quad x \longmapsto \dfrac{x^T A x}{x^T x}$

This is really $S^{n-1}_{(0,1)} \longrightarrow \mathbb{R}$, $\quad \hat{x} \longmapsto \hat{x}^T A \hat{x}$

unit vectors – pts on unit sphere.

Thm Rayleigh Prince

$\left.\begin{array}{l} A \text{ symm (not nec pos def)} \\ \text{EWs (Real) } \lambda_1 \le \lambda_2 \le \dots \le \lambda_n \\ \text{corresponding EVs } x_1 \dots x_n \end{array}\right\} \Longrightarrow$ R_A is minimized by x_1, $R_A(x_1) = \lambda_1$ (smallest EW)

[maximized by x_n $\;R(x_n) = \lambda_n$]

Pf. By Spectral Thm $\Lambda = Q^T A Q$. and COV $x = Qy$ (does not change lengths)

Then $R(x) = \dfrac{(Qy)^T A Q y}{(Qy)^T Q y} = \dfrac{y^T \Lambda y}{y^T y} = \dfrac{\lambda_1 y_1^2 + \lambda_2 y_2^2 + \dots + \lambda_n y_n^2}{\sum^n y_i^2}$

Since $\lambda_1 \le \lambda_i$ for $i = 2, \dots, n$ $\quad \lambda_1 (y_1^2 + y_2^2 + \dots + y_n^2) \le \lambda_1 y_1^2 + \lambda_2 y_2^2 + \dots + \lambda_n y_n^2$

$\Rightarrow \lambda_1 = \dfrac{\lambda_1 \cancel{(\sum y_i^2)}}{\cancel{\sum y_i^2}} \le \dfrac{\sum^n \lambda_i y_i^2}{\sum y_i^2} = R(x)$ and $R(x_1) = x_1^T \lambda_1 x_1 = \lambda_1$

$\Rightarrow \lambda_1 = R(x_1) \le R(x)$

$\triangleright$ By same arg $R(x) \le R(x_n) = \lambda_n$ $\quad \square$

Thus we could numerically seach for EWs by minimizing or maximizing $\dfrac{x^T A x}{x^T x}$; I don't know if this is ever done in practice.

Lemma Some computations we will need:

$x^T \Lambda x = \sum \lambda_i x_i^2$

$x^T Q \Lambda Q^T x = \sum \lambda_i (q_i^T x)^2$

This refers to the case $x \in \mathbb{C}^n$, A Hermetian, U unitary $\longrightarrow$ $x^H U \Lambda U^H x = \sum \lambda_i |u_i^H x|^2$

$x^T \Lambda x = [x_1\, x_2\, x_3] \begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

$= [x_1\, x_2\, x_3] \begin{bmatrix} \lambda_1 x_1 \\ \lambda_2 x_2 \\ \lambda_3 x_3 \end{bmatrix}$

$= \lambda_1 x_1^2 + \lambda_2 x_2^2 + \lambda_3 x_3^2$

$x^T Q \underbrace{\Lambda Q^T x}$

$\begin{bmatrix} -\lambda_1 q_1- \\ -\lambda_2 q_2- \\ -\lambda_3 q_3- \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

$\begin{bmatrix} \lambda_1 q_1^T x \\ \lambda_2 q_2^T x \\ \lambda_3 q_3^T x \end{bmatrix}$

and $x^T Q = (Q^T x)^T = \begin{bmatrix} q_1^T x \\ q_2^T x \\ q_3^T x \end{bmatrix}^T$

$\Rightarrow [q_1^T x\; q_2^T x\; q_3^T x] \begin{bmatrix} \lambda_1 q_1^T x \\ \lambda_2 q_2^T x \\ \lambda_3 q_3^T x \end{bmatrix}$

$= \sum \lambda_i (q_i^T x)^2$ $\quad \square$

$U \Lambda U^H$

$\begin{bmatrix} | & | & | \\ u_1 & u_2 & u_3 \\ | & | & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{bmatrix} \begin{bmatrix} -\bar{u}_1- \\ -\bar{u}_2- \\ -\bar{u}_3- \end{bmatrix} = \begin{bmatrix} | & | & | \\ u_1 & u_2 & u_3 \\ | & | & | \end{bmatrix} \begin{bmatrix} -\lambda_1 \bar{u}_1- \\ -\lambda_2 \bar{u}_2- \\ -\lambda_3 \bar{u}_3- \end{bmatrix}$

Thm A Symm, EWs ordered $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$

$$\Rightarrow \min_{\substack{\|x\|=1 \\ x \perp \operatorname{span}(q_1, \dots q_{k-1})}} \{x^T A x\} = \lambda_k$$

$$\max_{\substack{\|x\|=1 \\ x \perp \operatorname{span}(q_n, \dots q_{n-k+1})}} \{x^T A x\} = \lambda_{n-k}$$

Horn & Johnson MA p.177–178

This requires us to know EVs already, so not too useful.

Pf Lets just consider $x \perp \operatorname{span}\{q_1\}$ i.e. $\langle x, q_1\rangle = 0$

$$x^T A x = x^T Q \Lambda Q^T x \overset{\text{lemma}}{=} \sum_{i=1}^{n} \lambda_i (q_i^T x)^2 = \sum_2 \lambda_i (q_i^T x)^2 \quad \text{because } \langle x, q_1\rangle = 0$$

For pos $\alpha, \beta, \gamma > 0$
$\lambda_2 \le \lambda_3 \le \lambda_4$
$\alpha\lambda_2 + \beta\lambda_3 + \gamma\lambda_4 \ge \lambda_2\alpha + \lambda_2\beta + \lambda_2\gamma$

$$\ge \lambda_2 \sum_2 (q_i^T x)^2$$

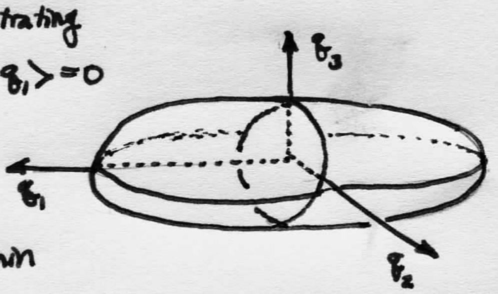
also

$$= \lambda_2 \sum_1 \left[(Q^T x)_i\right]^2 = \lambda_2 \|Q^T x\|_2^2 = \text{Q preserves lengths}$$

$$= \lambda_2 \|x\|_2^2$$

$$= \lambda_2 \, 1 \qquad \square$$

Strang p.351 is illustrating how restricting to $\langle x, q_1\rangle = 0$ is cutting the original ellipsoid with a plane and now we are looking at the max and min axes of that.



▷ We want to recast the above thm to avoid needing to know the EVs. I am taking this argument from Horn & Johnson MA p.178–180 so I am switching to $\mathbb{C}^n$ and $A = A^H$

Fix any $\vec{w} \in \mathbb{C}^n$.

$$\sup_{\substack{\text{all } x \text{ satisfying} \\ \|x\|=1 \\ \langle x, w\rangle = 0}} \{x^H A x\} = \sup_{\substack{|x|=1 \\ \langle x, w\rangle = 0}} x^H U \Lambda U^H x$$

let $z := U^H x \dashrightarrow x = Uz$
then $x^H x = (Uz)^H Uz = z^H U^H U z = z^H z$
$\Rightarrow \|x\| = 1 = \|z\|$

$$= \sup \sum \lambda_i |(U^H x)_i|^2$$

$0 = \langle x, w\rangle = \langle Uz, w\rangle = \langle z, U^H w\rangle$ def of IP

$$= \sup_{\substack{|z|=1 \\ \langle z, U^H w\rangle = 0}} \sum \lambda_i |z_i|^2$$

$$\ge \sup_{\substack{|z|=1 \\ \langle z, U^H w\rangle = 0 \\ z^{(1)} = \dots z^{(n-2)} = 0}} \left\{\sum \lambda_i |z_i|^2\right\} = \sup_{\substack{|z_{n-1}|^2 + |z_n|^2 = 1 \\ \langle z, U^H w\rangle = 0}} \left\{\lambda_{n-1}|z_{n-1}|^2 + \lambda_n |z_n|^2\right\}$$

restrict to subset of $\vec{z}$ with first $n-2$ components $=0$
[Is this set empty??]

$$\ge \lambda_{n-1}$$

Now $|z_{n-1}|^2 + |z_n|^2 = 1$

$$\lambda_{n-1}|z_{n-1}|^2 + \lambda_n |z_n|^2 \ge \lambda_{n-1}|z_{n-1}|^2 + \lambda_{n-1}|z_n|^2 \text{ since } \lambda_{n-1} \le \lambda_n$$

$$= \lambda_{n-1}(1)$$

Thus we have shown, for any given $\vec{w}$ $\quad \sup_{\substack{|x|=1 \\ \langle x,w\rangle=0}} \{x^H A x\} \geq \lambda_{n-1}$

By $\boxed{\text{Spectral Thm}}$ there is Unitary EV matrix U where $Au_k = \lambda_k u_k$

Thus $u_{n-1}^H A u_{n-1} = u_{n-1}^H \lambda_{n-1} u_{n-1} = \lambda_{n-1}$ so equality is obtained

'inf' can become 'min'.

$$\inf_{\text{all } w} \left[\sup_{\substack{|x|=1 \\ \langle x,w\rangle=0}} x^H A x \right] = \lambda_{n-1}$$

This arg generalizes to

Courant-Fischer min-max Thm

A Herm

$\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$

$$\min_{\substack{\text{all tuples} \\ \vec{w}_1,\dots,\vec{w}_{n-k}}} \left[\max_{x \perp W} \frac{x^H A x}{x^H x} \right] = \lambda_k$$

$W = \text{Span}\{w_1 \dots w_{n-k}\}$

$$\max_{\substack{\text{all} \\ \vec{w}_1 \dots \vec{w}_{n-k}}} \left[\min_{x \perp W} \frac{x^H A x}{x^H x} \right] = \lambda_k$$

Horn p.181

Thm (Weyl) A, B Hermetian

order EWs: $\lambda_1(A) \leq \lambda_2(A) \leq \dots \leq \lambda_n(A)$

$\lambda_1(B) \leq \dots \leq \lambda_n(B)$

$\lambda_1(A+B) \leq \dots \leq \lambda_n(A+B)$

$\Longrightarrow \lambda_k(A) + \lambda_1(B) \leq \lambda_k(A+B) \leq \lambda_k(A) + \lambda_n(B)$

Pf. For any $x \in \mathbb{C}^n$ we have $\lambda_1(B) \leq \frac{x^H B x}{x^H x} \leq \lambda_n(B)$

For any $k = 1, \dots, n$ we have

$$\lambda_k(A+B) = \min_{\vec{w}_1 \dots \vec{w}_{n-k}} \left[\max_{x \perp W} \frac{x^H (A+B) x}{x^H x} \right] = \min \left[\max \frac{x^H A x}{x^H x} + \frac{x^H B x}{x^H x} \right]$$

$$\geq \min \left[\max \frac{x^H A x}{x^H x} + \lambda_1(B) \right]$$

$$= \lambda_k(A) + \lambda_1(B) \qquad \square$$

Cor If B pos def $\lambda_k(A+B) \geq \lambda_k(A)$

Just subtracting a pos quantity from smaller side.

This solves Strang 6.4.8 6.4.7

NEXT: Strang has Finite Element Method