Strang ch 3.4

Rectangular Matrices with O.N. Cols

want to solve $Qx = b$

$$\begin{array}{c} \\ n \end{array}\begin{array}{c} m \\ \begin{bmatrix} - & q_1^T & - \\ & \vdots & \\ - & q_n^T & - \end{bmatrix} \end{array} \begin{bmatrix} | & & | \\ q_1 & \cdots & q_n \\ | & & | \end{bmatrix} = \begin{array}{c} \\ n \end{array}\begin{array}{c} n \\ \begin{bmatrix} I \end{bmatrix} \end{array}$$

Here Q^T is only a left inv of Q.

[O.N. matrices are crucial for numerical LA as they don't change lengths and thus keep round off error under control.]

To solve $Qx = b$ we use least squares (it is simpler than usual)

$$Q^TQ\bar{x} = Q^Tb$$
$$\bar{x} = Q^Tb$$
$$p = Q\bar{x} = QQ^Tb$$

proj of b onto cols: $p = Pb = q_1^Tb\,\hat{q}_1 + \dots + q_n^Tb\,\hat{q}_n$
$= QQ^Tb$
$= [\hat{q}_1q_1^T + \hat{q}_2q_2^T + \dots + q_nq_n^T]\,b$

Proj Matrix

$$P = A(A^TA)^{-1}A^T$$
$$= Q(Q^TQ)^{-1}Q^T = QQ^T \quad \text{simple form for } P$$

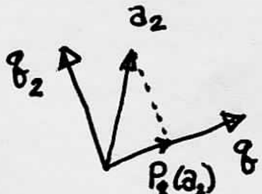
Gram-Schmidt (orthonormalization process)

Take an arb set of vectors $\{a_1, \dots, a_n\}$ and produce an O.N. set $\{\hat{q}_1, \dots, \hat{q}_n\}$ with same span.

Let $\hat{q}_1 := \frac{1}{\|a_1\|}a_1$

$a_2' := \vec{a}_2 - P_{q_1}(a_2)$
$= \vec{a}_2 - (q_1^Ta_2)\hat{q}_1$

$\hat{q}_2 := \frac{1}{\|a_2'\|}a_2'$



Actually if $\{a_i\}$ is not LI, then we end up with fewer than n q's

$a_3' := \vec{a}_3 - (q_1^Ta_3)\hat{q}_1 - (q_2^Ta_3)\hat{q}_2$

$\hat{q}_3 := \frac{1}{\|a_3'\|}a_3'$

etc...

Key Idea:
- Take the next vector a_j
- O.N. Proj it onto the subsp spanned by prev $\hat{q}_1 \dots \hat{q}_{j-1}$
- Subtract that off and get a new vector OG to that subsp, namely a_j'
- Normalize a_j' to become $\hat{q}_j$.

$$a_j' := a_j - \left[\langle a_j, q_1\rangle\hat{q}_1 + \dots + \langle a_j, q_{j-1}\rangle\hat{q}_{j-1}\right]$$
$$q_j := \frac{1}{\|a_j'\|}a_j'$$

[P. 181 #3.4.19 says it is more stable numerically to subtract these one at a time.]

A=QR factorization

The above eq gives us: $\|a_j'\|q_j = a_j - \sum_{i=1}^{j-1}\langle a_j, q_i\rangle q_i$

Apply q_j^T to both sides: $\|a_j'\|1 = q_j^Ta_j$

$\Rightarrow a_j = \sum^{j}\langle a_j, q_i\rangle q_i$

In matrix form:

Take $n=3$

Really $A = [\ ]$

$$\begin{array}{c} \\ m \\ \\ \end{array}\begin{array}{c} n \\ \begin{bmatrix} | & | & | \\ a_1 & a_2 & a_3 \\ | & | & | \end{bmatrix} \\ A \end{array} = \begin{array}{c} \\ m \\ \\ \end{array}\begin{array}{c} n \\ \begin{bmatrix} | & | & | \\ q_1 & q_2 & q_3 \\ | & | & | \end{bmatrix} \\ Q \end{array} \begin{array}{c} n \\ \begin{bmatrix} q_1^Ta_1 & q_1^Ta_2 & q_1^Ta_3 \\ 0 & q_2^Ta_2 & q_2^Ta_3 \\ 0 & 0 & q_3^Ta_3 \end{bmatrix} \\ R \end{array}\begin{array}{c} \\ n \\ \\ \end{array}$$

R — upper triang and invertible if A is.

All main diag elts pos because $q_i^Ta_i = \|a_i'\|$ $\Rightarrow$ R is invertible

Only pos elts on main diag $\Rightarrow \det R > 0$

The main point of orthogonalization is to simplify the least sq problem:

$$A^TA = R^TQ^TQR = R^TR$$

$A^TA\bar{x} = A^Tb$ becomes $R^TR\bar{x} = R^TQ^Tb$

or just $\boxed{R\bar{x} = Q^Tb}$

since R invertible and can be cancelled.

Ch 6 (Symmetric) Pos Def Matrices

For <u>any</u> $n \times n$ matrix A, define quadratic form $f: \mathbb{R}^n \longrightarrow \mathbb{R}$

$$x \longmapsto x^T A x = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$$

If A is <u>anti</u>-Symm ($A^T = -A$), this reduces to pure quadratic $a_{11} x_1^2 + a_{22} x_2^2 + \dots + a_{nn} x_n^2$

$$[x\ y\ z]\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = ax^2 + (b+d)xy + ey^2 + (c+g)xz + (f+g)yz + iz^2$$

mixed terms vanish if A anti-symm.

Motivation
- Consider a smooth fcn $F: \mathbb{R}^n \to \mathbb{R}$. Taylor expand: $F(x) = F(0) + DF_0(x) + \frac{1}{2!} D^2F_0(x,x) + \mathcal{O}(\|x\|^3)$
- F has a local minima at 0 if $DF_0 = 0$ and $D^2F_0(x,x) > 0 \ \forall x \neq 0$
- Let $A := D^2 F_0$. Then $DF_0(x,x) = x^T A x$ and $x^T A x > 0 \ \forall x \neq 0$ if A pos def.

<u>Def</u> A is <u>pos def</u> if $x^T A x > 0$ for all x, $x \neq 0$

Strang only considers <u>Symm</u> pos def matrices, so that the following thm holds.
It certainly is possible for a non-Symm matrix to satisfy $x^T A x > 0$
For example $A = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}$ $[x\ y]\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = 2x^2 + 1xy + 3y^2 > 0$ for $x, y \neq 0$ because $xy < 2x^2 + 3y^2$

To see that $xy < 2x^2 + 3y^2$: For pos numbers a, b $a \neq b$ $0 < (b-a)^2 = b^2 - 2ab + a^2$, $2ab < a^2 + b^2$

Thus $xy \le |x||y| < 2|x||y| < x^2 + y^2 < 2x^2 + 3y^2$ (If $x=y$ the result is even easier).

<u>Another way</u>: $f(x,y) := 2x^2 + xy + 3y^2$ $D^2 f_x = \begin{bmatrix} 4 & 1 \\ 1 & 6 \end{bmatrix}$ This is <u>not</u> A, but it <u>is</u> Symm pos def.

$Df_x = [4x+y,\ x+6y]$

$Df_0 = [0\ 0]$ critical pt.

Thus, from calculus we would know $f(0,0) = 0$ is a local min (in fact global) and thus $f(x,y) > 0 \ \forall x, y \neq 0$ $\square$

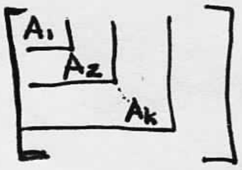
Thm A symm pos def matrix T.F.A.E. (I) $x^T A x > 0 \ \forall x \neq 0$

$\Leftrightarrow$ (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 $\|x\| = 1$

$0 < x^T A x = x^T \lambda x = \lambda$ $\square$

(I) $\Leftarrow$ (II) Since A symm, Spectral Thm says A has full set of O.N. EVs $\{x_1, \dots, x_n\}$ (basis of span)

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

$Ax = c_1 A x_1 + \dots + c_n A x_n = c_1 \lambda_1 x_1 + \dots + c_n \lambda_n x_n$

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

(I) $\Rightarrow$ (III) $x^T A x > 0 \ \forall x$, in particular if we fix $k \in \{1, \dots, n-1\}$ let $x^T = [x_k^T \,\vdots\, 0 \cdots 0]$ (last $n-k$ components)

$\Rightarrow [x_k^T \,\vdots\, 0]\left[\begin{array}{c:c} A_k & * \\ \hdashline * & * \end{array}\right]\left[\begin{array}{c} x_k \\ \hdashline 0 \end{array}\right] = x_k^T A_k x_k > 0 \Rightarrow$ All EWs $\lambda_i^{(k)}$ of A_k are pos by II

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

Thus a nec cond for A is each $a_{ii} > 0$. Not suff.

There is nothing special about using upper left submatrices. Any chain of principal submats would work, starting from a diag elt a_{ii}. Principal Submatrix A_k is starting with a_{ii} and adding new col and row pair each time.

$(\mathrm{III}) \Rightarrow (\mathrm{IV})$ 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 $d_k = \dfrac{d_1 \cdot d_2 \cdots d_{k-1} d_k}{d_1 \cdots d_{k-1}} = \dfrac{\det A_k > 0}{\det A_{k-1} > 0}$

$(\mathrm{IV}) \Rightarrow (\mathrm{I})$ 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 if each $d_i > 0$ □

$(\mathrm{I}) \Leftarrow (\mathrm{V})$ $A = R^TR$ where R has LI cols

$x^TAx = x^TR^TRx = \|Rx\|_2^2 > 0$ if $x \neq 0$ and $Rx \neq 0$ if $x \neq 0$ because R has LI cols $\Rightarrow \ker(R) = \{0\}$

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

$A = LDL^T \Rightarrow \underbrace{(L\sqrt{D})}_{R^T}\underbrace{(\sqrt{D}L^T)}_{R}$ or $A = Q\Lambda Q^T$ Spectral Thm
$= (Q\sqrt{\Lambda})(\sqrt{\Lambda}Q^T)$
$= R^T \ R$

And R can be rectangular if we like.
Take any valid sq R and a matrix $Q = \begin{bmatrix} & \end{bmatrix}$ with O.N. cols.
Rectangular $R := QR$
Then $R^TR = R^T\begin{bmatrix} Q^T \end{bmatrix}\begin{bmatrix} Q \end{bmatrix}R = R^TR = A$

□

In least squares, the normal eqs are $A^TA\bar{x} = A^Tb$
This is now known to be Symm pos def!

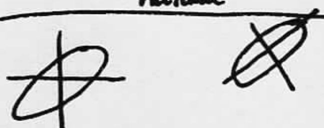
Ellipsoids in n-dim

$x^TAx = 1$ defines an ellipsoid for ~~arb A (exclude extreme degenerate cases)~~.
A symm pos def (or else we might get hyperboloids, etc...)

$\sum\sum a_{ij}x_ix_j = 1$

Since A is Symm, pos def $A = Q\Lambda Q^T$ where Q is O.N. and Λ has only pos elts.

Thm $\exists$ rotation $Q \ni$ with the new co-ords $y = Q^Tx$ ellipsoid is aligned with the co-ord axes. Thus the EVs of A are the ~~co-ord~~ natural axes of the ellipsoid and the length of axis i is $\frac{1}{\sqrt{\lambda_i}}$



Pf We already found $Q \ni A = Q\Lambda Q^T$ from Spectral thm
If $\det Q = +1$, Q is rotation. We know $\det Q = \pm 1$, if $\det Q = -1$ replace Q with $\begin{bmatrix} -q_1 & q_2 & \cdots & q_n \end{bmatrix}$ Still EVs with same EWs.

$1 = x^TAx = x^TQ\Lambda Q^T = y^T\Lambda y = \lambda_1 y_1^2 + \dots \lambda_n y_n^2$ each $\lambda_i > 0$

For axis i, plug in $y_j = 0 \ \forall j \neq i \Rightarrow 1 = \lambda_i y_i^2 \Rightarrow y_i = \pm\frac{1}{\sqrt{\lambda_i}}$. □

Say we have $q(x) = x^TAx$ for A Symm, pos def
We want to make a linear COV $x = Cy$ for C nsing matrix
$q(x) = x^TAx = y^TC^TACy =: \tilde{q}(y)$

$\boxed{B = C^TAC \text{ is still Symm pos def!}}$

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