Strang LAAIA
Appendix A

Singular Value Decomposition

Strang's discussion had to be rearranged and extended so it made sense.

Thm SVD: General matrix $A : \mathbb{R}^n \to \mathbb{R}^m$ $\Rightarrow$ $A = Q_1 \Sigma Q_2^T$

Here e.g. $n < m$

$$\begin{array}{cc} & n \\ m & \begin{bmatrix} \quad \\ \quad \\ \quad \end{bmatrix} \end{array}$$

where

Q_1 O.N. — cols are EVs of AA^T

Q_2 O.N. — " " " " A^TA

$$\Sigma = \left[\begin{array}{c|c} \Delta & 0 \\ \hline 0 & 0 \end{array}\right] \qquad \Delta = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}$$

$r = \text{rank}(A)$

For $i = 1, \dots, r$ λ_i is pos EW of both AA^T & A^TA

Define $\sigma_i = \sqrt{\lambda_i}$

A Complex $\Rightarrow A = \underbrace{U_1 \Sigma U_2^H}_{\text{unitary}}$

Key ideas which will become clearer as we procede:

$$\underset{A}{\begin{bmatrix} \quad \\ \quad \\ \quad \end{bmatrix}} = \underset{Q_1}{\begin{bmatrix} | \\ q_j \\ | \end{bmatrix}} \underset{\Sigma}{\left[\begin{array}{c|c} \begin{matrix}\sigma_1 & \\ & \sigma_r\end{matrix} & 0 \\ \hline 0 & 0 \end{array}\right]} \underset{Q_2^T}{\begin{bmatrix} - x_i^T - \end{bmatrix}}$$

$$\begin{tikzcd} {[v]_e} \arrow[d, bend right] & \mathbb{R}^n \arrow[r, "A"] \arrow[d, "Q_2^{-1} = Q_2^T"'] & \mathbb{R}^m \\ {[v]_X = v_r + v_n} & \mathrm{Im}(A^T) \oplus \ker(A) \arrow[r, "\Sigma"] & \mathrm{Im}(A) \oplus \ker(A^T) \arrow[u, "Q_1"'] \end{tikzcd}$$

Change of basis matrix Q_2^T takes a co-ord vector $[v]_e$ in orig basis to a new co-ord vector in terms of the basis $\{x_1 \dots x_r, x_{r+1} \dots x_n\}$

pf

Step 1 Define matrix Q_2 as EVs of A^TA

we know the $n \times n$ matrix A^TA is Symm $\xRightarrow{\text{Spectral Thm}}$ $\exists$ a full set of O.N. EVs x_j:

$A^TAx_j = \lambda_j x_j$ and

$x_j^T x_k = 0$ if $j \neq k$

Put these in cols of $Q_2 = \begin{bmatrix} | & & | \\ x_1 & \cdots & x_n \\ | & & | \end{bmatrix}$

Now observe that all λ_j are in fact non-neg:

$$x_j^T A^T A x_j = x_j^T \lambda_j x_j$$
$$(Ax_j)^T A x_j = \lambda_j$$
$$0 \leq \|Ax_j\|_2^2 = \lambda_j$$

Lets order these so $\lambda_1 \dots \lambda_r$ are pos and $\lambda_j = 0$ for $j = r+1, \dots, n$

Then $Ax_j = 0$ for $j = r+1, \dots, n$ i.e. $Ax_j = 0x_j$

This tells us $r = \text{rank}(A)$ because we show $\dim(\ker(A)) = n - r$ [rank+nullity Thm]

Lemma: $\ker(A) = \text{Span}\{x_{r+1}, \dots, x_n\}$

Pf. by contrad. Suppose there was $v \in \text{Span}\{x_1 \dots x_r\} \ni Av \overset{!}{=} 0$. $v = \sum_1^r \alpha_i x_i$ and some $\alpha_i \neq 0$

Then $A^T(Av) \overset{!}{=} A^T 0 = 0 \Rightarrow (A^TA)(v) = (A^TA)(\sum \alpha_i x_i) \overset{!}{=} 0$

$$\sum \alpha_i (A^TAx_i) = \sum \alpha_i \lambda_i x_i \overset{!}{=} 0$$

but $\lambda_i > 0$ and $\{x_i\}$ is LI $\Rightarrow$ each $\alpha_i = 0$ $\boxed{\Rightarrow\Leftarrow}$

contd →

Step 2 Define cols of Q_1

For $i=1,\dots,r$ define $\begin{cases} \sigma_j = \sqrt{\lambda_j} & \text{the "singular values"} \\ q_j = \frac{1}{\sigma_j} A x_j \end{cases} \implies \boxed{A x_j = \sigma_j q_j}$ Key relation

we see $\{q_j\}$ is ON:

$$q_i^T q_j = \frac{1}{\sigma_i}\frac{1}{\sigma_j}(x_i^T A^T A x_j) = \frac{\lambda_j}{\sigma_i \sigma_j} x_i^T x_j = \begin{cases} 0 & i \neq j \\ \frac{\lambda_j}{\sigma_j^2} = 1 & i = j \end{cases}$$

Extend $\{q_1 \dots q_r\}$ using Gram-Schmidt to add ON $\{q_{r+1} \dots q_m\}$

$$Q_1 := \begin{bmatrix} | & & | \\ q_1 & \cdots & q_m \\ | & & | \end{bmatrix}$$

Step 3 Now define $\Sigma := Q_1^T A Q_2$

Elt $\Sigma_{ij} = (Q_1^T A Q_2)_i^{(j)} = q_i^T A x_j = \begin{cases} q_i^T \sigma_j q_j & \text{if } j \leq r \text{ because } Ax_j = \sigma_j q_j \\ & \text{BUT } q_i^T q_j = 0 \text{ for } i \neq j \\ & \text{so we really get} \begin{cases} \sigma_i & i=j \\ 0 & i \neq j \end{cases} \\ 0 & \text{if } j > r \text{ because} \\ & Ax_j = \lambda_j x_j = 0 x_j = 0 \end{cases}$

$$\begin{bmatrix} - q_i^T - \end{bmatrix} \begin{bmatrix} | \\ Ax_j \\ | \end{bmatrix}$$

The only <u>nonzero</u> elts in Σ are $\Sigma_{ii} = \sigma_i$ for $i = 1,\dots,r$

$\triangleright$ so $\Sigma = Q_1^T A Q_2 \implies A = Q_1 \Sigma Q_2^T$ $\boxed{\text{QED}}$

COR
1. $\{x_1 \dots x_r\}$ is ON basis for RowSpace(A) = Col(A^T) = Im(A^T)
2. $\{x_{r+1} \dots x_n\}$ ON basis ker(A)
3. $\{q_1 \dots q_r\}$ ON basis Col(A) = Im(A) = $A(\mathbb{R}^n)$
4. $\{q_{r+1} \dots q_m\}$ ON basis ker(A^T)

(2) we proved as <u>Lemma</u> in Step 1

(1) follows from Strang's Fund Thm LA II : $\mathbb{R}^n = \overbrace{\text{Im}(A^T)}^{\ker(A)^\perp} \oplus \ker(A)$

(3) choose any $w \in$ Col(A). we must show it can be expressed as LC $\{q_i\}_1^r$

By def $\exists\, u \in \mathbb{R}^n \ni Au = w$

By <u>step 1 Lemma</u> $u = \sum \alpha_i x_i$ thus $Au = A(\sum^r \alpha_i x_i) = \sum^r \alpha_i A x_i \overset{\text{step 2 key relation}}{=} \sum^r \alpha_i \sigma_i q_i$

so $w = $ LC $\{q_1 \dots q_r\}$

(4) Since $\mathbb{R}^m = \text{Im}(A) \oplus \underbrace{\ker(A^T)}_{\text{Im}(A)^\perp}$ we see $\{q_{r+1} \dots q_m\}$ must span $\ker(A^T)$

$\square$

Strang gives some remarks:

① For A symm pos def, SVD is just $A = Q\Lambda Q^T$

②
$$Q_1 = \begin{bmatrix} | & & | & | & & | \\ q_1 & \cdots & q_r & q_{r+1} & \cdots & q_m \\ | & & | & | & & | \end{bmatrix} \qquad Q_2 = \begin{bmatrix} | & & | & | & & | \\ x_1 & \cdots & x_r & x_{r+1} & \cdots & x_n \\ | & & | & | & & | \end{bmatrix}$$

$$\underbrace{q_1 \cdots q_r}_{\substack{r \\ Col(A)}} \quad \underbrace{q_{r+1} \cdots q_m}_{\substack{m-r \\ ker(A^T)}} \qquad \underbrace{x_1 \cdots x_r}_{\substack{r \\ Col(A^T)}} \quad \underbrace{x_{r+1} \cdots x_n}_{\substack{n-r \\ ker(A)}}$$

③ $AQ_2 = Q_1\Sigma$ i.e. $Ax_j = \sigma_j q_j$ when $\sigma_j = 0$ if $j > r$

④ By construction in the pf, $Q_2 = [x_1 \cdots x_n]$ is EV matrix of A^TA w/ EWs λ_i

What about AA^T? $AA^T = Q_1\Sigma \cancel{Q_2^T Q_2}\Sigma^T Q_1^T = Q_1\Sigma\Sigma^T Q_1^T \Rightarrow (AA^T)Q_1 = Q_1(\Sigma\Sigma^T)$

Using my block multiplying notation in ch 2 sheets B_k^ℓ (ℓ ← cols, k ← rows) $\quad (B_k^\ell)^T = B_\ell^k$

Let $r+k=n$
$r+\ell=m$

$$\underbrace{\left[\begin{array}{c:c} \Delta_r^r & 0_r^k \\ \hdashline 0_\ell^r & 0_\ell^k \end{array}\right]}_{\Sigma} \underbrace{\left[\begin{array}{c:c} \Delta_r^r & 0_r^\ell \\ \hdashline 0_k^r & 0_k^\ell \end{array}\right]}_{\Sigma^T} = \left[\begin{array}{c:c} \Delta_r^r\Delta_r^r + 0_r^k 0_k^r & \Delta_r^r 0_r^\ell + 0_r^k 0_k^\ell \\ \hdashline 0_\ell^r \Delta_r^r + 0_\ell^k 0_k^r & 0_\ell^r 0_r^\ell + 0_\ell^k 0_k^\ell \end{array}\right]$$

$$= \left[\begin{array}{c:c} (\Delta^2)_r^r & 0_r^\ell \\ \hdashline 0_\ell^r & 0_\ell^\ell \end{array}\right] \qquad \Delta^2 = \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_r^2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_r \end{bmatrix}$$

$$\begin{array}{c c} & n \\ m & \begin{bmatrix} A \end{bmatrix} \end{array} \begin{array}{c c} m & \\ \begin{bmatrix} A^T \end{bmatrix} & n \end{array} \begin{array}{c} m \\ \begin{bmatrix} Q_1 \end{bmatrix} \end{array} = \begin{bmatrix} | & & | \\ q_1 & \cdots & q_m \\ | & & | \end{bmatrix} \left[\begin{array}{c:c} \begin{matrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_r \end{matrix} & \\ \hdashline & \begin{matrix} 0 & & \\ & \ddots & \\ & & 0 \end{matrix} \end{array}\right]$$

⑤ For a numerically stable def of rank(A), we can't count pivots, so we count the non-zero σ_i values [A^TA has EWs σ_i^2] we decree $\sigma_i = 0$ if say $\sigma_i < 10^{-6}$. EFFECTIVE RANK

▷ Polar Decomposition

$$\begin{tikzcd} \mathbb{R}^n \arrow[r, "A"] \arrow[d, "Q_2^T"'] & \mathbb{R}^m \\ Im(A^T) \oplus ker(A) \arrow[r, "\Sigma"] & Im(A) \oplus ker(A^T) \arrow[u, "Q_1"'] \end{tikzcd}$$

we can see how the action of any matrix is broken up into rotation/reflection Q_2, Q_1 and stretching/compressing Σ

Thm A $n\times n$ square, Real $\Rightarrow A = QS$
Q = O.N
S = Symm, pos semi def
[A nsing $\Rightarrow S$ pos def]

Also can have Q unitary, S Hermitian

pf.
$$\begin{aligned} A &= Q_1\Sigma Q_2^T \\ &= Q_1(Q_2^TQ_2)\Sigma Q_2^T \\ &= \underbrace{(Q_1Q_2^T)}_{\substack{Q \\ \text{ON rotation,} \\ \text{reflection}}} \underbrace{(Q_2\Sigma Q_2^T)}_{\substack{S \\ \text{stretching}}} \end{aligned}$$

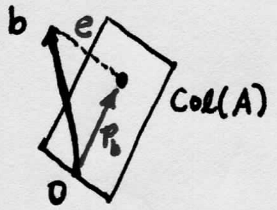
□

Least Squares and SVD

From ch 3.3 we want to solve $Ax = b$ for $A: \mathbb{R}^n \to \mathbb{R}^m$ $\quad {}_m\begin{bmatrix} & {}^n & \\ & A & \\ & & \end{bmatrix} \begin{bmatrix} x \end{bmatrix}^n = \begin{bmatrix} b \end{bmatrix}^m$

3 Things can go wrong in general
1) sys of eqs inconsistent, $b \notin Col(A)$
2) rows of A are LD (must happen if $m > n$)
3) Cols of A are LD ← This was not encountered in ch 3.3

Let's recap what we did in ch 3.3:



we OG proj b into $Col(A)$ and get p_b and by def of $Col(A)$
There must be some $\bar{x} \ni A\bar{x} = p_b$. $\quad e := b - A\bar{x}$
$e \in \ker(A^T)$ because $\mathbb{R}^m = Col(A) \oplus \underbrace{\ker(A^T)}_{Col(A)^\perp}$
Then $A^T e = 0 \Rightarrow A^T(b - A\bar{x}) = 0 \Rightarrow A^T A\bar{x} = A^T b$
we showed if A has LI cols, $A^T A$ is invertible and $\bar{x} = (A^T A)^{-1} A^T b$

That took care of cases (1) and (2). Now we must also handle (3) cols of A <u>not</u> LI:

If cols of A are LD $\Rightarrow \ker(A) \neq \{0\}$ so $A^T A\bar{x}$ is not one-to-one

$\mathbb{R}^n = Col(A^T) \oplus \ker(A)$ so $\bar{x} = x_r + x_n$. Any $x_n \in \ker(A)$ gives same value for $A^T A(\overset{\bar{x}}{x_r + x_n})$

<u>Key Idea</u>: choose $\bar{x}$ of minimum length [i.e. $x_n = 0$]. Then $\bar{x} = x_r = x^+$

Thm the minimum length soln to $Ax = b$ is $x^+ = A^+ b$, where we use the SVD to find the 'Generalized Inverse' $A^+ := Q_2 \Sigma^+ Q_1^T$

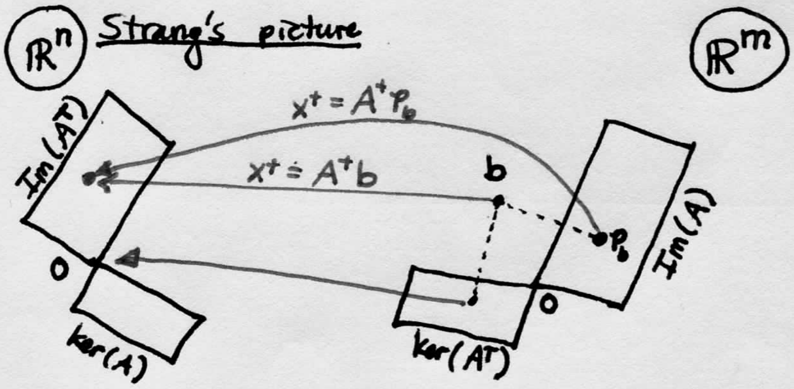
$\|\bar{x}\|^2 = \|x_r\|^2 + \|x_n\|^2$

and from $\Sigma = \left[\begin{array}{c|c} \Delta & 0 \\ \hline 0 & 0 \end{array}\right]$ $\quad \Delta = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}$

we get $\Sigma^+ := \left[\begin{array}{c|c} \Delta^{-1} & 0 \\ \hline 0 & 0 \end{array}\right]$ 1. Transpose 2. Take reciprocals in Δ

We need a warm-up example and a Lemma before giving the pf, but first a picture of what is going on:

$$\begin{tikzcd}
\mathbb{R}^n \arrow[r, "A"] & \mathbb{R}^m \arrow[l, bend left=70, "A^+"'] \\
\mathrm{Im}(A^T)\oplus\ker(A) \arrow[u, "Q_2"] \arrow[r, shift left, "\Sigma"] & \mathrm{Im}(A)\oplus\ker(A^T) \arrow[u, "Q_1"'] \arrow[l, shift left, "\Sigma^+"] \arrow[l, bend left=60, "\Delta"] \arrow[l, bend left=80, "\Delta^{-1}"]
\end{tikzcd}$$



What is the relationship between A^T and A^{-1}?
For A^{-1} to exist, we need $n = m$ and $\ker(A) = \{0\} \Rightarrow$ then $\Sigma = \Delta$
But $A^T = A^{-1}$ only if $A = Q$ a O.N. matrix
because only then is $\Delta = I$ no stretching in any direction.

cont'd →

warm-up example

$$A = \Sigma = \left[\begin{array}{cc|cc} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \end{array}\right] \qquad b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

Here $A : \mathbb{R}^4 \longrightarrow \mathbb{R}^3$

usually we think of $m > n$ but here we don't just to show it works in all cases.

we are showing $\Sigma x = b$ has soln $x^+ = \Sigma^+ b$

$\text{col}(A) = \left\{ \begin{array}{l} \text{all} \\ (y_1 y_2 0) \end{array} \right\}$ so $\underset{p_b}{\text{Proj}}(b) = \begin{bmatrix} b_1 \\ b_2 \\ 0 \end{bmatrix}$ and what $\bar{x}$ would map to this?

we can see

$$\begin{bmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} b_1/\sigma_1 \\ b_2/\sigma_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ 0 \end{bmatrix}$$

so x_3 and x_4 are arb and $\bar{x}$ is not unique.
we choose x_3, x_4 so $\bar{x}$ has min length: $x_3 = 0$, $x_4 = 0$

Then $x^+ = \begin{bmatrix} b_1/\sigma_1 \\ b_2/\sigma_2 \\ 0 \\ 0 \end{bmatrix}$ and what is A^+ such that $A^+ b = x^+$?

By inspection $A^+ = \left[\begin{array}{cc|c} 1/\sigma_1 & 0 & 0 \\ 0 & 1/\sigma_2 & 0 \\ \hline 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right]$ we had to ~~transpose~~ reciprocal non-zero diag elts, and transpose matrix.

Lemma The least sq soln to $\Sigma x = b$ is $x^+ = \Sigma^+ b$

Pf

$$x^+ = \Sigma^+ b = \left[\begin{array}{c|c} (\Delta^{-1})^r_r & 0^\ell_r \\ \hline 0^r_k & 0^\ell_k \end{array}\right] \left[\begin{array}{c} b^1_r \\ \hline b^1_\ell \end{array}\right] = \left[\begin{array}{c} \Delta^{-1} b^1_r \\ \hline 0^1_k \end{array}\right] \begin{array}{l} \} \text{ co-ords for basis vectors } x_1 \dots x_r \Rightarrow \text{row space} \\ \} \text{ co-ords for } x_{r+1} \dots x_n \Rightarrow \ker(A) \end{array}$$

$\left[\begin{array}{l} r+\ell = m \\ r+k = n \end{array}\right.$ as before$\left.\right]$

Thus we see x^+ in row space

Now we want to show $p_b = \Sigma \Sigma^+ b$ and it will be of interest to do it by multiplying $\Sigma\Sigma^+$ first:

$$\underset{\Sigma}{\left[\begin{array}{c|c} \Delta^r_r & 0^k_r \\ \hline 0^r_\ell & 0^k_\ell \end{array}\right]} \underset{\Sigma^+}{\left[\begin{array}{c|c} (\Delta^{-1})^r_k & 0^\ell_r \\ \hline 0^r_k & 0^\ell_k \end{array}\right]} \left[\begin{array}{c} b^1_r \\ \hline b^1_\ell \end{array}\right] = \left[\begin{array}{c|c} \Delta^r_r (\Delta^{-1})^r_r + 0^k_r 0^r_k & \Lambda^r_r 0^\ell_r + 0^k_r 0^\ell_k \\ \hline 0^r_\ell (\Delta^{-1})^r_r + 0^k_\ell 0^r_k & 0^r_\ell 0^\ell_r + 0^k_\ell 0^\ell_k \end{array}\right] \left[\begin{array}{c} b^1_r \\ \hline b^1_\ell \end{array}\right]$$

$$m\left\{\underset{\Sigma\Sigma^+}{\overbrace{\left[\begin{array}{c|c} I^r_r & 0^\ell_r \\ \hline 0^r_\ell & 0^\ell_\ell \end{array}\right]}^{m}}\right. \left[\begin{array}{c} b^1_r \\ \hline b^1_\ell \end{array}\right] = \left[\begin{array}{c} b^1_r \\ \hline 0^1_\ell \end{array}\right] = p_b \checkmark$$

$\square$

Now we need to prove the thm $\longrightarrow$

Repeat thm: The min length least sq soln to $Ax=b$ is $x^+ = A^+ b$

Pf. want to minimize
$$\|Ax-b\| = \|Q_1 \Sigma Q_2^T x - b\|$$
$$= \|Q_1^T(Q_1 \Sigma Q_2^T x - b)\|$$ we can mult by ON matrix w/out changing length
$$= \|\Sigma (Q_2^T x) - Q_1^T b\|$$ $y := Q_2^T x$
$$= \|\Sigma y - Q_1^T b\|$$

So we want min length soln to $\Sigma y = Q_1^T b$ and from the Lemma we know this is $y^+ = \Sigma^+ (Q_1^T b)$

Transform back to x: $x^+ = Q_2 y^+ = (Q_2 \Sigma^+ Q_1^T) b = A^+ b$ $\square$

Strang asks us to validate:
- x^+ is in the rowspace of A $[\exists z \ni A^T z = x^+]$
- $Ax^+ = p_b$

For $Ax^+ = p_b$, we are showing
$$AA^+ b = Q_1 \Sigma Q_2^T (Q_2 \Sigma^+ Q_1^T) b$$
$$= Q_1 \Sigma \Sigma^+ Q_1^T b$$
$$= Q_1 \left[\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right] Q_1^T b$$ from prev lemma
$$= \begin{bmatrix} | & & | & | & & | \\ q_1 & \cdots & q_r & 0 & \cdots & 0 \\ | & & | & | & & | \end{bmatrix} \begin{bmatrix} - q_1 - \\ \vdots \\ - q_m - \end{bmatrix} \begin{bmatrix} b \end{bmatrix}$$

really it is
$$Q_1 \begin{bmatrix} - q_1 - \\ \vdots \\ - q_r - \\ 0 \\ \vdots \\ 0 \end{bmatrix} \begin{bmatrix} b \end{bmatrix} = Q_1 \begin{bmatrix} q_1^T b \\ \vdots \\ q_r^T b \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$
← This vector is the proj of original b into Col(A) [i.e. p_b] and Q_1 is COB to put it in terms of orig basis for $\mathbb{R}^m$.