Lecture #13: The Singular Value Decomp.

Today we will cover ~~arguably~~ the most powerful tool in linear algebra

SVD: Given an $n \times m$ matrix A, there is always a way to write it as

$$\underset{}{A} = \underset{n\times n}{U}\,\underset{n\times m}{\Sigma}\,\underset{m\times m}{V^T}$$

where (1) U and V have orthonormal columns

(2) and Σ is nonnegative and diagonal

$$\Sigma = \begin{bmatrix} \sigma_1 & & 0 \\ & \sigma_2 & \\ 0 & & \ddots \end{bmatrix}$$

and $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > \sigma_{r+1} = \sigma_{r+2} = \dots = 0$

First, there is an alternative expression that is sometimes more convenient:

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T$$

Some nomenclature:

① the σ_i's are called singular values

② the u_i's and v_i's are called the left and right singular vectors, respectively

We'll spend this lecture digesting what this decomposition means, i.e.

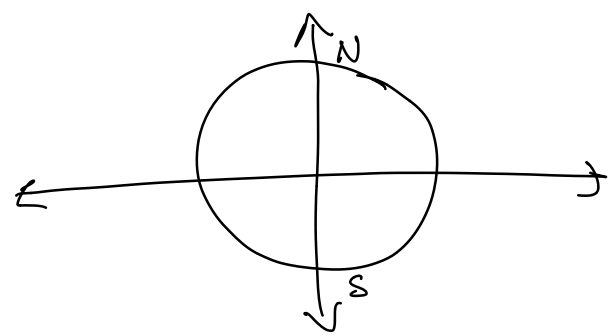
How can we read off important properties of A from it?

Let's start from the geometry and understand what happens to the unit ball:

$$B = \{ x \mid \|x\| \leq 1, x \in \mathbb{R}^n \}$$

Let's visualize what's happening as we apply V^T, Σ, then U

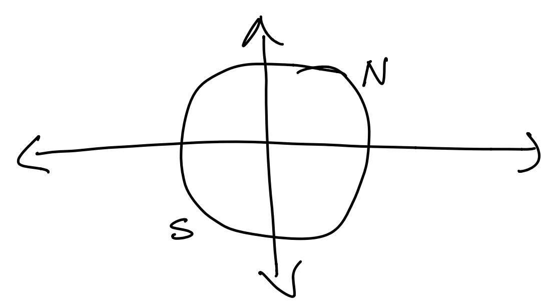
In particular the unit ball looks like



Q: what happens when we apply V^T?

Multiplying by V^T does not change the length of a vector

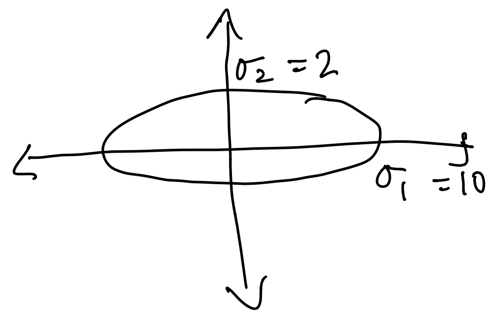
$$B = V^T B = \{ V^T x \mid \|x\| \leq 1, x \in \mathbb{R}^m \}$$



It's just the coordinates that are different

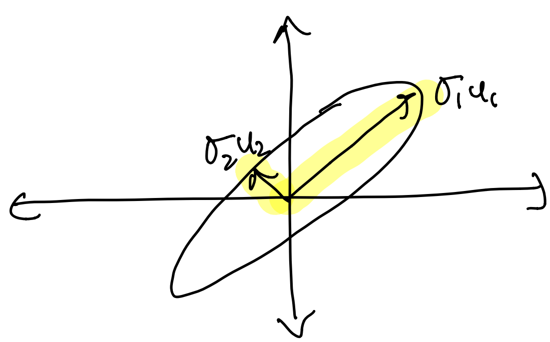
Q2: what happens when we multiply by Σ?

we get something like



This just changes the ball into ellipsoid

And finally, multiplying by U we get



It's principal axes are $\sigma_1 u_1, \dots$

Let's see it in action

<u>Application:</u> Uncertainty regions

In many applications like MRI, we get linear measurements of some unknown x

e.g. a picture we would like to reconstruct

$$y = Ax + z$$

noise, suppose $\|z\| \le \delta$

How can we reconstruct x approximately?

Let's suppose $\underbrace{\text{A has full column rank}}_{①}$

Why? otherwise it would not be possible to recover x even w/o noise

Property ① is equivalent to:

"A has a left inverse, i.e. a matrix N so that $NA = I$" (lecture 8)

So if we estimate x using

$$\hat{x} = Ny$$

what can we say about the reconstruction error $\hat{x} - x$?

$$\hat{x} - x = N(Ax+z) - x = \cancel{x} + Nz - \cancel{x}$$

Hence the error is in an <u>uncertainty ellipsoid:</u>

$$\{ Nz \mid \|z\| \le \delta \}$$

<u>side note:</u> Is the left inverse unique?

No, consider

$$A = \begin{bmatrix} M \\ 3M \end{bmatrix}$$

<u>Q3:</u> Does using the left inverse of the top/bottom dominate the other?

Now let's see how we can read off facts about A from its SVD

<u>Property 1:</u> The <u>rank</u> of A is the # of non-zero singular values

Property 2: The vectors $u_1, u_2 \dots u_r$ are an orthonormal basis for $C(A)$

Let's get some intuition for this

Q4: How should I choose x so that Ax is in the direction of u_1?

Recall the second expression for the SVD

$$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \dots + \sigma_r u_r v_r^T$$

Now if I choose $x = v_1$ then

$$Ax = \sigma_1 u_1 \underbrace{v_1^T v_1}_{=1} + \cancelto{0}{\sigma_2 u_2 v_2^T v_1} + \cancelto{0}{\dots}$$

This follows from the v_i's being orthonormal

Similarly $A v_i = \sigma_i u_i$ for all i, so I know that every $u_i \in C(A)$

Conversely for <u>any</u> x I get

$$Ax = \sigma_1 u_1 v_1^T x + \sigma_2 u_2 v_2^T x + \dots$$

which is a linear combination of u_i's

This proves $C(A) = \text{span}(u_1, u_2 \dots u_r)$, and by assumption the u_i's are orthonormal

Similarly we have:

<u>Property 3</u>: The vectors $v_{r+1}, v_{r+2} \dots v_m$ are an orthonormal basis for $N(A)$

In particular consider Ax for $x = v_{r+1}$

$$Ax = \sigma_1 u_1 \cancel{v_1^T v_{r+1}}^{\,0} + \sigma_2 u_2 \cancel{v_2^T v_{r+1}}^{\,0} + \dots$$

again by orthonormality

In fact the SVD also contains powerful theorems as a corollary:

Recall:

<u>Rank-Nullity Theorem</u>: For any $n \times m$ A we have that $\text{rank}(A) + \dim(N(A)) = m$

<u>Q5</u>: How can we see that from the SVD?

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

$$\dim(N(A)) = m - r$$

Not only that but we can directly compute A^{-1} (if it exists) from the SVD too!

<u>Fact</u>: Suppose A is square and invertible and has SVD $A = U\Sigma V^T$. Then

$$A^{-1} = V\Sigma^{-1}U^T$$

First let's see why this is natural

$$A^{-1} = (U\Sigma V^T)^{-1} = (V^T)^{-1}\Sigma^{-1}U^{-1}$$

But since U and V^T are orthogonal we have that $U^{-1} = U^T$ and $(V^T)^{-1} = V$, hence:

$$A^{-1} = V \Sigma^{-1} U^T$$

It's useful to double-check that this indeed works

$$V \Sigma^{-1} U^T A = V \Sigma^{-1} \underbrace{U^T U}_{I} \Sigma V^T$$

$$= V \Sigma^{-1} \Sigma V^T$$

$$= V V^T = I$$

In fact even when A is not invertible (or maybe not even square) we can still do the next best thing:

definition: The pseudoinverse, denoted by A^+, of A is

$$A^+ = \sum_{i=1}^{r} \sigma_i^{-1} v_i u_i^T$$

What does this do? Let's try AA^+:

$$AA^+ = \left(\sum_{i=1}^{r} \sigma_i u_i v_i^T\right)\left(\sum_{i=1}^{r} \sigma_i^{-1} v_i u_i^T\right)$$

$$= \sum_{i=1}^{r} \sigma_i \sigma_i^{-1} u_i u_i^T = \sum_{i=1}^{r} u_i u_i^T$$

Hence $AA^+ = \sum_{i=1}^{r} u_i u_i^T =$ projection onto $C(A)$

Similarly we have that

$$A^+A = \sum_{i=1}^{r} v_i v_i^T = \text{projection onto } N(A)^\perp$$

Now returning to our application in estimation

we know that the left inverse of A is not always unique, but

Q6: What is the best left inverse to use, in terms of minimizing the uncertainty ellipsoid?

It turns out that it is A^+, i.e.

<u>Fact:</u> The uncertainty ellipsoid for A^+ is contained in those of any other left inverse