This handout includes space for every question that requires a written response. Please feel free to use it to handwrite your solutions (legibly, please). If you choose to typeset your solutions, the `README.md` for this assignment includes instructions to regenerate this handout with your typeset LaTeX solutions.

4.a

Question: Suppose we have a matrix $A \epsilon \mathbb{R}^{n \times d}$ with SVD $A = UDV^T$, where $U \epsilon \mathbb{R}^{n \times r}$, $D \epsilon \mathbb{R}^{r \times r}$ and $V \epsilon \mathbb{R}^{d \times r}$

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

ANSWER:
u_i is the i^{th} column of matrix U where $i = [1...r]$ and r represents the rank of matrix A.

D is the diagonal matrix with the values σ_i representing the non zero value of the i^{th} column.

$UD = [\sigma_1 u_1 \sigma_2 u_2 .... \sigma_r u_r]$

$V = [v_1 v_2 .... v_r]$ where v_i is the i^{th} column vector

$UDV^T = [\sigma_1 u_1 \sigma_2 u_2 .... \sigma_r u_r][v_1 v_2 .... v_r]^T$

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

Substituting A

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

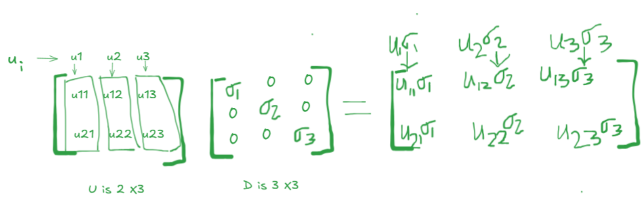

Figure 1: UD multiplication illustration

4.b

Question: Show that
$u_i = \frac{1}{\sigma_i} A v_i$
In particular, the components u_i represent the size of the projection of the rows of A onto v_i scaled by σ_i

$V^T V = I$
$[v_1 v_2 .... v_r]^T [v_1 v_2 .... v_r] = I$
$A = UDV^T$
Multiplying both sides by v_i $Av_i = UDV^T v_i$
$Av_i = UD(V^T v_i)$
$V^T v_i = i^{th}$ column of vector I represented as e_i

$$e_i = \begin{bmatrix} 0 \\ 0 \\ \dots \\ 1 \\ 0 \end{bmatrix}$$

$Av_i = UDV^T v_i$
$Av_i = UDe_i$
$Av_i = U(De_i)$

D is the diagonal matrix with

$$D = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ 0 & 0 & \dots & \sigma_r \end{bmatrix}$$

$Av_i = U(\sigma_i e_i)$
$Av_i = \sigma_i (Ue_i)$
$Av_i = \sigma_i u_i$

$u_i = \frac{1}{\sigma_i} A v_i$

4.c

One way of finding a reduced rank approximation of A is by hard-setting all but the k largest σ_i to 0. This approximation is called truncated SVD, and by (a) we see it can be written as

$A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T$

From (a) we see the truncated SVD can also be written as $A_k = U_k D_k V_k^T$, where $U_k \, \epsilon \, \mathbb{R}^{n\times k}$ and $V \, \epsilon \, \mathbb{R}^{d\times k}$ are the first k columns of U, V and $D \, \epsilon \, \mathbb{R}^{k\times k}$ has the first k singular values.

Show the the rows of A_k are the projections of the rows A onto the subspace V_k spanned by the first k right singular vectors.

Hint: Recall that the projection of a vector a onto a subspace spanned by v_1, ... , v_k where the v_i are pairwise orthogonal is given by the sum of projections of a onto the individual v_i.

$A = UDV^T$ ... (1)
$AV_k = UDV^TV_k$ ... (2)
$AV_k = UD(V^TV_k)$ ... (3)

$V^T \, \epsilon \, \mathbb{R}^{k\times d}$ $V_k \, \epsilon \, \mathbb{R}^{d\times k}$

$$V^TV_k = \begin{bmatrix} I_k \\ 0 \end{bmatrix}$$

$AV_k = U_k D_k$

Now,
$A_k = U_k D_k V_k^T$ Substituting $U_k D_k with AV_k$.

$A_k = A(V_k V_k^T)$ ... (4)

this proves that each row of A_k is a projection of the corresponding row of A onto the v_1, . . . , v_k

4.d

The Frobenium norm of a matrix $M \epsilon \mathbb{R}^{m \times n}$ is defined as

$\|M\|_F = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{m} M_{ij}^2}$

Show that

$A_k = argmin_{rank(B)=k} \|A - B\|_F$

where the arg min is taken over matrices of rank k. Hint: Use the fact that V_k is the best-fit k-dimensional subspace for the rows of A.

Rank of matrix = Number of independent direction the matrix has.
Lets assume A is a $10X8$ matrix. There are 8 columns but the rows truly depend on 3 underlying directions / patters i.e. 3 columns. That is the Rank / True Rank / Actual Rank. We determined that the $r = rank(A)$ by counting the number of non-zero σ_i values in the following equation:
$A = UDV^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T$
$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T$
$r = rank(A)$
$\sigma_1 > \sigma_2 \ldots \sigma_r > 0$

Low Rank Matrix has few main directions that are most important.
In low Rank approximation (Truncated SVD), we keep only the top k columns such that $k < r$
$A_k = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_k u_k v_k^T$

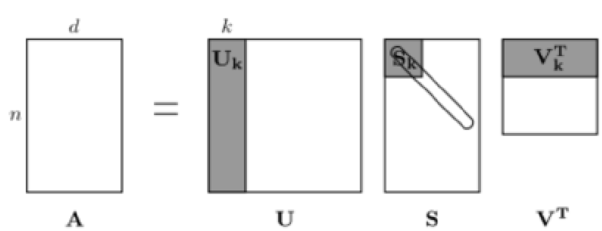

Figure 1: Dimensionality reduction by truncated SVD

Figure 2: From Assignment A1

$A_k = U_k D_k V_k^T$
Using Frobenium norm
$A_F^2 = \sum_{i=1}^{r} \sigma_i^2$
$A_k{}_F^2 = \sum_{i=1}^{k} \sigma_i^2$
$A - A_k{}_F^2 = \sum_{i=k+1}^{r} \sigma_i^2$

Given $\sigma_1 > \sigma_2 \ldots \sigma_r > 0$, the norm will be smallest if the top k eigen values are chosen and hence satisfies the equation:

$A_k = argmin_{rank(B)=k} \|A - B\|_F$