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

To show that:

$$\mathbf{A} = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$$

given the singular value decomposition (SVD) of matrix $\mathbf{A} \in \mathbb{R}^{n \times d}$, $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T$, where:

- $\mathbf{U} \in \mathbb{R}^{n \times r}$ is the matrix of **left singular vectors**.
- $\mathbf{D} \in \mathbb{R}^{r \times r}$ is a diagonal matrix with positive diagonal entries $\sigma_i$ sorted in **descending order**: $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, where each $\sigma_i$ is a **singular value** of $\mathbf{A}$.
- $\mathbf{V} \in \mathbb{R}^{d \times r}$ is the matrix of **right singular vectors**.

**Proof:**

1. From the SVD, we have:

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T$$

2. Let's express $\mathbf{D}$ as a diagonal matrix with the singular values $\sigma_1, \sigma_2, \ldots, \sigma_r$ on its diagonal:

$$\mathbf{D} = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r)$$

Thus, $\mathbf{D}$ can be written as:

$$\mathbf{D} = \sum_{i=1}^{r} \sigma_i \mathbf{e}_i \mathbf{e}_i^T$$

where $\mathbf{e}_i$ is the $i$-th standard basis vector (a column vector with 1 at the $i$-th position and 0 elsewhere).

3. Substituting this into the SVD expression for $\mathbf{A}$, we get:

$$\mathbf{A} = \mathbf{U}\left(\sum_{i=1}^{r} \sigma_i \mathbf{e}_i \mathbf{e}_i^T\right)\mathbf{V}^T$$

4. Since matrix multiplication is distributive, we can distribute the sum:

$$\mathbf{A} = \sum_{i=1}^{r} \sigma_i \mathbf{U}\mathbf{e}_i \mathbf{e}_i^T \mathbf{V}^T$$

5. Notice that $\mathbf{U}\mathbf{e}_i$ gives the $i$-th column of $\mathbf{U}$, denoted as $\mathbf{u}_i$, and $\mathbf{V}\mathbf{e}_i$ gives the $i$-th column of $\mathbf{V}$, denoted as $\mathbf{v}_i$. Therefore:

$$\mathbf{A} = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$$

Thus, we have shown that the matrix $\mathbf{A}$ can be expressed as the sum of the outer products of the singular vectors, scaled by the corresponding singular values.

$$\boxed{\mathbf{A} = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^T}$$

# 4.b

We are given the equation to prove:

$$\mathbf{u}_i = \frac{1}{\sigma_i}\mathbf{A}\mathbf{v}_i$$

This equation shows that the $i$-th left singular vector $\mathbf{u}_i$ can be expressed as a scaled projection of the matrix $\mathbf{A}$ onto the $i$-th right singular vector $\mathbf{v}_i$.

**Proof:**

We know from the Singular Value Decomposition (SVD) that:

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T$$

Let's take the product $\mathbf{A}\mathbf{v}_i$:

$$\mathbf{A}\mathbf{v}_i = \mathbf{U}\mathbf{D}\mathbf{V}^T\mathbf{v}_i$$

Since the columns of $\mathbf{V}$ are orthonormal, we have $\mathbf{V}^T\mathbf{v}_i = \mathbf{e}_i$, where $\mathbf{e}_i$ is the $i$-th standard basis vector (a vector with 1 in the $i$-th position and 0 elsewhere).

Therefore:

$$\mathbf{A}\mathbf{v}_i = \mathbf{U}\mathbf{D}\mathbf{e}_i$$

Now, $\mathbf{D}\mathbf{e}_i = \sigma_i\mathbf{e}_i$, where $\sigma_i$ is the $i$-th singular value. Thus, we get:

$$\mathbf{A}\mathbf{v}_i = \mathbf{U}\sigma_i\mathbf{e}_i$$

Since $\mathbf{U}\mathbf{e}_i = \mathbf{u}_i$, we conclude that:

$$\mathbf{A}\mathbf{v}_i = \sigma_i\mathbf{u}_i$$

Solving for $\mathbf{u}_i$, we obtain:

$$\mathbf{u}_i = \frac{1}{\sigma_i}\mathbf{A}\mathbf{v}_i$$

This completes the proof.

**Conclusion:**

The components of $\mathbf{u}_i$ represent the size of the projection of the rows of $\mathbf{A}$ onto $\mathbf{v}_i$, scaled by the singular value $\sigma_i$.

# 4.c

We are given that the truncated SVD of $\mathbf{A}$, denoted as $\mathbf{A}_k$, is defined as:

$$\mathbf{A}_k = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$$

where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the left and right singular vectors corresponding to the singular values $\sigma_i$, and $k$ is the number of retained singular values. This is called the *truncated SVD*, where only the largest $k$ singular values and their corresponding singular vectors are kept.

We also know that:

$$\mathbf{A}_k = \mathbf{U}_k \mathbf{D}_k \mathbf{V}_k^T$$

**Claim:**

The rows of $\mathbf{A}_k$ are the projections of the rows of $\mathbf{A}$ onto the subspace of $\mathbf{V}_k$ spanned by the first $k$ right singular vectors.

**Proof:**

From the SVD of $\mathbf{A}$, we know that:

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T$$

Each row of $\mathbf{A}$ can be written as a linear combination of the right singular vectors (columns of $\mathbf{V}$). Now, consider the projection of a row of $\mathbf{A}$ onto the subspace spanned by the first $k$ right singular vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$.

Using the formula for projecting a vector $\mathbf{a}$ onto a subspace $W$ spanned by $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$, the projection is given by the sum of the individual projections of $\mathbf{a}$ onto each $\mathbf{v}_i$:

$$\text{Proj}_{\mathbf{V}_k}(\mathbf{a}) = \sum_{i=1}^{k} \frac{\mathbf{a}^T \mathbf{v}_i}{\|\mathbf{v}_i\|^2} \mathbf{v}_i$$

Since the right singular vectors $\mathbf{v}_i$ are orthonormal, $\|\mathbf{v}_i\| = 1$, so the projection simplifies to:

$$\text{Proj}_{\mathbf{V}_k}(\mathbf{a}) = \sum_{i=1}^{k} (\mathbf{a}^T \mathbf{v}_i) \mathbf{v}_i$$

In the context of SVD, this projection is precisely captured by the truncated SVD, which retains only the first $k$ singular vectors and values. Thus, $\mathbf{A}_k$ represents the matrix where each row of $\mathbf{A}$ is projected onto the subspace spanned by $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$.

Since each row of $\mathbf{A}_k$ is constructed using the right singular vectors $\mathbf{v}_i$ and scaled by the corresponding singular values $\sigma_i$, this proves that the rows of $\mathbf{A}_k$ are indeed the projections of the rows of $\mathbf{A}$ onto the subspace spanned by the first $k$ right singular vectors.

# 4.d

We are asked to prove that:

$$\mathbf{A}_k = \arg \min_{\mathsf{rank}(\mathbf{B})=k} \|\mathbf{A} - \mathbf{B}\|_F$$

where the matrix $\mathbf{A}_k$ is the truncated SVD of $\mathbf{A}$, and the Frobenius norm is defined as:

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

**Best-Fit Subspace:**

The best-fit subspace for a matrix $\mathbf{A}$ is given by the subspace spanned by the top $k$ right singular vectors. This is known as $\mathbf{V}_k = \mathrm{span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k)$. The truncated SVD $\mathbf{A}_k = \mathbf{U}_k \mathbf{D}_k \mathbf{V}_k^T$ retains only the largest $k$ singular values and their corresponding singular vectors.

Thus, $\mathbf{A}_k$ provides the best rank-$k$ approximation of $\mathbf{A}$, capturing the most significant components of $\mathbf{A}$.

**Proof:**

1. Frobenius Norm:

The Frobenius norm measures the distance between two matrices by summing the squared differences between their elements. Minimizing $\|\mathbf{A} - \mathbf{B}\|_F$ ensures that $\mathbf{B}$ is the closest approximation of $\mathbf{A}$ in terms of this norm.

2. Best-Fit Subspace:

From the information about best-fit subspaces, we know that the subspace spanned by the top $k$ right singular vectors of $\mathbf{A}$, i.e., $\mathbf{V}_k$, is the $k$-dimensional subspace that minimizes the distance between the rows of $\mathbf{A}$ and their projections onto this subspace.

3. Truncated SVD:

The truncated SVD, $\mathbf{A}_k = \mathbf{U}_k \mathbf{D}_k \mathbf{V}_k^T$, projects the rows of $\mathbf{A}$ onto the subspace spanned by the first $k$ right singular vectors. This means that $\mathbf{A}_k$ is the matrix that best fits the rows of $\mathbf{A}$ by minimizing the projection error.

4. Minimizing the Frobenius Norm:

Since $\mathbf{A}_k$ is constructed by projecting the rows of $\mathbf{A}$ onto the subspace spanned by the top $k$ singular vectors, it minimizes the Frobenius norm $\|\mathbf{A} - \mathbf{B}\|_F$ for any matrix $\mathbf{B}$ of rank $k$.

Thus, we conclude that $\mathbf{A}_k$ is the matrix that minimizes the Frobenius norm and provides the best rank-$k$ approximation of $\mathbf{A}$:

$$\mathbf{A}_k = \arg \min_{\mathsf{rank}(\mathbf{B})=k} \|\mathbf{A} - \mathbf{B}\|_F$$

**Conclusion:**

Using the best-fit subspace $\mathbf{V}_k$ and the properties of the truncated SVD, we have shown that $\mathbf{A}_k$ minimizes the Frobenius norm between $\mathbf{A}$ and any rank-$k$ matrix, providing the best approximation.