# Homework 2
# Stat 215A, Fall 2024

**Due:** Submit a `homework2.pdf` file to **Gradescope** by Friday, October 04 at 23:59

## 0 Linear Algebra Review

Recall that the SVD of $\mathbf{X} \in \mathbb{R}^{n \times p}$ is a matrix decomposition such that $\mathbf{X} = \mathbf{U}\,\mathbf{D}\,\mathbf{V}^\top$, where $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_n] \in \mathbb{R}^{n \times n}$, $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_p] \in \mathbb{R}^{p \times p}$, and $\mathbf{D} = \operatorname{diag}(d_1, \dots, d_{\min\{n,p\}}) \in \mathbb{R}^{n \times p}$. In addition, $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices so that $\mathbf{U}^\top \mathbf{U} = \mathbf{U}\,\mathbf{U}^\top = \mathbf{I}$ and $\mathbf{V}^\top \mathbf{V} = \mathbf{V}\,\mathbf{V}^\top = \mathbf{I}$ (i.e., $\mathbf{u}_j^\top \mathbf{u}_i = \mathbf{v}_j^\top \mathbf{v}_i = 0$ for all $i \neq j$ and $\mathbf{u}_j^\top \mathbf{u}_j = \mathbf{v}_j^\top \mathbf{v}_j = 1$ for all $i$). Moreover, we can find a decomposition where $d_1 \geq \cdots \geq d_{\min\{n,p\}} \geq 0$.

Now, while the SVD can be used for any rectangular matrix, square matrices have an additional special property and can be decomposed via an eigendecomposition. Given a square matrix $\mathbf{A} \in \mathbb{R}^{p \times p}$, we say that $\mathbf{v} \in \mathbb{R}^p$ is an *eigenvector* of $\mathbf{A}$ if $\mathbf{v} \neq \mathbf{0}$ and $\mathbf{A}\,\mathbf{v} = \lambda\,\mathbf{v}$ for some $\lambda \in \mathbb{R}$. We also call $\lambda$ the *eigenvalue* of $\mathbf{A}$ corresponding to the eigenvector $\mathbf{v}$. For a more intuitive (geometric) interpretation of eigenvalues and eigenvectors, see this reference.

There is a close connection between the SVD and eigendecomposition. Namely, for any matrix $\mathbf{X} \in \mathbb{R}^{n \times p}$, $\mathbf{v} \in \mathbb{R}^p$ is a right singular vector of $\mathbf{X}$ with singular value $d$ if and only if $\mathbf{v} \in \mathbb{R}^p$ is an eigenvector of $\mathbf{X}^\top \mathbf{X}$ corresponding to the eigenvalue $d^2$. You may use this fact without proof.

## 1 Principal Components Analysis and SVD

Let $\mathbf{X}$ be an $n \times p$ data matrix, where $n$ is the number of observations and $p$ is the number of features. For simplicity, we will assume that $\mathbf{X}$ has been mean-centered (i.e., each column of $X$ has mean 0) and that $n \leq p$. The population version of PCA as can be seen as solving for each $j = 1, \dots, p$

$$\mathbf{v}_j^* = \underset{\mathbf{v} \in \mathbb{R}^p}{\operatorname{argmax}}\ \mathbf{v}^\top \operatorname{Var}(\mathbf{X})\,\mathbf{v} \qquad \text{subject to} \quad \|\mathbf{v}\|_2^2 = 1, \ \ \mathbf{v}^\top \mathbf{v}_i^* = 0 \ \forall\, i < j. \tag{1.1}$$

However, since $\operatorname{Var}(\mathbf{X})$ is almost always unknown in practice, we typically estimate $\operatorname{Var}(\mathbf{X})$ with the sample covariance $\frac{1}{n}\mathbf{X}^\top \mathbf{X}$. Thus, in practice, the principal component (PC) directions, $\hat{\mathbf{v}}_1, \dots, \hat{\mathbf{v}}_p$, are the solution to the following system of optimization problems:

$$\hat{\mathbf{v}}_j = \underset{\mathbf{v} \in \mathbb{R}^p}{\operatorname{argmax}}\ \mathbf{v}^\top \mathbf{X}^\top \mathbf{X}\,\mathbf{v} \qquad \text{subject to} \quad \|\mathbf{v}\|_2^2 = 1, \ \ \mathbf{v}^\top \hat{\mathbf{v}}_i = 0 \ \ \forall\, i < j. \tag{1.2}$$

In this problem, we will take small steps through the proof to show that the PC directions are precisely the right singular vectors of $\mathbf{X}$.

1. To begin, prove that the first PC direction $\hat{\mathbf{v}}_1$ is equal to the first right singular vector $\mathbf{v}_1$. To show this, use Lagrange multipliers to solve the PC1 optimization problem:

$$\hat{\mathbf{v}}_1 = \underset{\mathbf{v} \in \mathbb{R}^p}{\operatorname{argmax}}\ \mathbf{v}^\top \mathbf{X}^\top \mathbf{X}\,\mathbf{v} \qquad \text{subject to} \quad \|\mathbf{v}\|_2^2 = 1. \tag{1.3}$$

**Solution:**

Using the Lagrangian method, we can define the Lagrangian as:

$$\mathcal{L}(v, \lambda) = v^\top X^\top X v - \lambda(v^\top v - 1)$$

Taking the derivative of $\mathcal{L}$ with respect to $v$ and setting it to zero yields:

$$\nabla_v \mathcal{L} = 2X^\top X v - 2\lambda v = 0 \quad \Rightarrow \quad X^\top X v = \lambda v$$

This shows that $v$ is an eigenvector of $X^\top X$ corresponding to eigenvalue $\lambda$. Therefore, $\hat{v}_1$ is the first right singular vector of $X$, associated with the largest eigenvalue.

2. Next, let $j \in \{2, \ldots, p\}$ be given. Use the SVD and matrix multiplication to show that for all $\mathbf{v} \in \mathbb{R}^p$ satisfying $\mathbf{v}^\top \mathbf{v}_i = 0$ for each $i < j$, we have

$$\mathbf{v}^\top \mathbf{X}^\top \mathbf{X}\, \mathbf{v} = \sum_{k=j}^{p} \mathbf{v}^\top \left( d_k^2\, \mathbf{v}_k\, \mathbf{v}_k^\top \right) \mathbf{v}, \tag{1.4}$$

where we define $d_k = 0$ for $k = n+1, \ldots, p$.

**Solution:**

From the SVD decomposition of $X$:

$$X = UDV^\top$$

It follows that:

$$X^\top X = VD^2V^\top$$

Since $v$ is orthogonal to the first $j-1$ singular vectors, only the components corresponding to singular values $d_j, d_{j+1}, \ldots$ remain, resulting in the summation:

$$v^\top X^\top X v = \sum_{k=j} v^\top (d_k^2 v_k v_k^\top) v$$

3. Then, show that for each $j = 2, \ldots, p$, the original (sample) PCA formulation in (1.2) is equivalent to

$$\hat{\mathbf{v}}_j = \underset{\mathbf{v} \in \mathbb{R}^p}{\operatorname{argmax}}\ \mathbf{v}^\top \left( \mathbf{X}_{(j)}^\top\, \mathbf{X}_{(j)} \right) \mathbf{v} \qquad \text{subject to} \quad \|\mathbf{v}\|_2^2 = 1, \tag{1.5}$$

where $\mathbf{X}_{(j)} = \widetilde{\mathbf{U}}\widetilde{\mathbf{D}}\widetilde{\mathbf{V}}^\top$, $\widetilde{\mathbf{U}} = [\mathbf{u}_j, \ldots, \mathbf{u}_n, \mathbf{u}_1, \ldots, \mathbf{u}_{j-1}] \in \mathbb{R}^{n \times n}$, $\widetilde{\mathbf{D}} = \operatorname{diag}(d_j, \ldots, d_n, 0, \ldots, 0) \in \mathbb{R}^{n \times p}$, and $\widetilde{\mathbf{V}} = [\mathbf{v}_j, \ldots, \mathbf{v}_p, \mathbf{v}_1, \ldots, \mathbf{v}_{j-1}] \in \mathbb{R}^{p \times p}$.

**Solution:**

Using the truncated SVD $X^{(j)}$, we remove the first $j-1$ components from $X$. The corresponding optimization problem for the $j$-th singular vector remains equivalent because the SVD decomposition preserves orthogonality between components.

4. Conclude that for each $j = 1, \ldots, p$, the $j^{th}$ PC direction, $\hat{\mathbf{v}}_j$, is equal to the $j^{th}$ right singular vector $\mathbf{v}_j$. (Hint: Problem 1 may be useful).

**Solution:**

Since each principal component direction is the result of solving an eigenvalue problem corresponding to $X^\top X$, we conclude that each $j$-th PC direction is equal to the $j$-th right singular vector from the SVD of $X$.

# 2 Miscellaneous

- What are several concrete actions that could be taken to increase the ability of replication for data based scientific findings?

  **Solution:**

  1. Share datasets and preprocessing scripts openly.
  2. Provide detailed documentation of methods and results.
  3. Use standardized methodologies.
  4. Publish both positive and negative results to avoid bias.

- What are two relevant questions to keep in mind while designing algorithms for partitioning data?

  **Solution:**

  1. How well does the partitioning method scale with data size?
  2. How does the method handle imbalanced or noisy data?