A. Definition for a Matrix $\mathbf{X} \in \mathbb{R}^{n \times p}$

First Definition (Maximizing Variance)[1]

The first principal component, defined by maximizing variance ($\mathbf{PC}_1^{\text{var}}$), is the unit vector $\mathbf{w}$ that maximizes the sum of squared projections of the data points ($\mathbf{x}_i$) onto it. This can be expressed as:

$$\begin{aligned}
\mathbf{PC}_1^{\text{var}} &= \underset{\substack{\mathbf{w}\in\mathbb{R}^p \\ \|\mathbf{w}\|_2=1}}{\arg\max} \sum_{i=1}^{n} (\mathbf{w}^T \mathbf{x}_i)^2 \\
&= \underset{\substack{\mathbf{w}\in\mathbb{R}^p \\ \|\mathbf{w}\|_2=1}}{\arg\max} \|\mathbf{X}\mathbf{w}\|_2^2 \\
&= \underset{\substack{\mathbf{w}\in\mathbb{R}^p \\ \|\mathbf{w}\|_2=1}}{\arg\max} \mathbf{w}^T \mathbf{X}^T \mathbf{X} \mathbf{w}
\end{aligned}$$

Note: The notation $||\mathbf{w}|| = 1$ typically implies the Euclidean norm, $||\mathbf{w}||_2 = 1$.

Second Definition (Minimizing Least Squares Error)

The first principal component, defined by minimizing the least squares reconstruction error ($\mathbf{PC}_1^{\text{LS}}$), is the unit vector $\mathbf{w}$ that minimizes the sum of squared distances between the data points ($\mathbf{x}_i$) and their projections onto the line defined by $\mathbf{w}$. This is given by:

$$\mathbf{PC}_1^{\text{LS}} = \underset{\substack{\mathbf{w}\in\mathbb{R}^p \\ \|\mathbf{w}\|_2=1}}{\arg\min} \sum_{i=1}^{n} \text{dist}(\mathbf{x}_i, \mathbf{w})^2$$

where:

$$\text{dist}(\mathbf{x}_i, \mathbf{w}) = \|\mathbf{x}_i - P_{\mathbf{w}}(\mathbf{x}_i)\|_2$$

and $P_{\mathbf{w}}(\mathbf{x}_i)$ is the orthonormal projection of the point $\mathbf{x}_i$ onto the one-dimensional subspace spanned by the vector $\mathbf{w}$. Specifically, $P_{\mathbf{w}}(\mathbf{x}_i) = (\mathbf{w}^T \mathbf{x}_i)\mathbf{w}$ since $\|\mathbf{w}\|_2 = 1$.

Exercise. *Assume that the eigenvalues of the matrix $\mathbf{X}^T\mathbf{X}$ satisfy $\lambda_1 > \lambda_2 > \cdots > \lambda_p \geq 0$. Show that $\mathbf{PC}_1^{var}$ corresponds to $\mathbf{v}_1$, the first column vector (eigenvector corresponding to λ_1) of $\mathbf{U}$ in the spectral decomposition of $\mathbf{X}^T\mathbf{X}$:*

$$\mathbf{X}^T\mathbf{X} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$$

(Where $\mathbf{U}$ is the orthogonal matrix whose columns are the eigenvectors $\mathbf{v}_i$, and $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues λ_i).

[1] The superscript '1' might refer to a footnote or citation in the original source document.