Princeton University Computer Science
Spring 2025

# COS 324: Introduction to Machine Learning
# Proofs for Principal Component Analysis (PCA)

Ruth Fong

**Last updated: Wednesday 12th February, 2025, 5:38pm.**

## 1. Proof for best projection

**Goal:** Given a centered dataset $\{\vec{\mathbf{v}}_i \in \mathbb{R}^d\}_{i=1}^N$ (i.e. mean of dataset is $\vec{\mathbf{0}}$), an orthonormal basis of $k$ vectors $\mathcal{U} = \{\vec{\mathbf{u}}_j \in \mathbb{R}^d\}_{j=1}^k$, and a new datapoint $\vec{\mathbf{v}} \in \mathbb{R}^d$, find its best low-rank approximation (i.e. best projection) $\widehat{\vec{\mathbf{v}}} \in \text{span}(\mathcal{U})$ that minimizes its reconstruction error:

$$\min \|\vec{\mathbf{v}} - \widehat{\vec{\mathbf{v}}}\|_2^2 \tag{1}$$

where

$$\widehat{\vec{\mathbf{v}}} = \alpha_1\vec{\mathbf{u}}_1 + \alpha_2\vec{\mathbf{u}}_2 + \ldots + \alpha_k\vec{\mathbf{u}}_k = \sum j = 1^k \alpha_j\vec{\mathbf{u}}_j, \text{ using } \widehat{\vec{\mathbf{v}}} \in \text{span}(\mathcal{U}) \tag{2}$$

This amounts to finding the best solution for the $\alpha_i$ terms.
**Claim:** The best projection $\widehat{\vec{\mathbf{v}}}$ for $\vec{\mathbf{v}}$ that minimizes the above reconstruction error is defined as follows:

$$\widehat{\vec{v}} = \sum_{j=1}^{k}(\vec{\mathbf{v}} \cdot \vec{\mathbf{u}}_j)\vec{\mathbf{u}}_j, \text{ where } \alpha_j = \vec{\mathbf{v}} \cdot \vec{\mathbf{u}}_j \tag{3}$$

**Approach:** We'll prove this by computing the partial derivatives of the reconstruction loss with respect to (w.r.t.) $\alpha_i$, setting it to 0, and solving for $\alpha_i$ (Problem 7.1.3 in course notes).

### 1.1. Re-write reconstruction error.

We'll first re-write the reconstruction error $L$ for datapoint $\vec{\mathbf{v}}$:

$$L = \|\vec{\mathbf{v}} - \widehat{\vec{\mathbf{v}}}\|_2^2 \tag{4}$$

$$= \|\vec{\mathbf{v}} - \sum_{j=1}^{k}\alpha_j\vec{\mathbf{u}}_j\|_2^2, \text{ using eq. (2)} \tag{5}$$

$$= (\vec{\mathbf{v}} - \sum_{j=1}^{k}\alpha_j\vec{\mathbf{u}}_j) \cdot (\vec{\mathbf{v}} - \sum_{j=1}^{k}\alpha_j\vec{\mathbf{u}}_j), \text{ using } \|\vec{\mathbf{v}}\|_2^2 = \vec{\mathbf{v}} \cdot \vec{\mathbf{v}} \tag{6}$$

$$= \vec{\mathbf{v}} \cdot \vec{\mathbf{v}} - 2\vec{\mathbf{v}} \cdot (\sum_{j=1}^{k}\alpha_j\vec{\mathbf{u}}_j) + \sum_{i=1}^{k}\sum_{j=1}^{k}\alpha_i\alpha_j(\vec{\mathbf{u}}_i \cdot \vec{\mathbf{u}}_j) \tag{7}$$

$$= \|\vec{\mathbf{v}}\|_2^2 - 2\vec{\mathbf{v}} \cdot (\sum_{j=1}^{k}\alpha_j\vec{\mathbf{u}}_j) + \sum_{j=1}^{k}\alpha_j^2(\vec{\mathbf{u}}_j \cdot \vec{\mathbf{u}}_j), \text{ using } \vec{\mathbf{u}}_i \cdot \vec{\mathbf{u}}_j = 0 \text{ if } i \neq j \text{ (orthonormal prop.)} \tag{8}$$

$$= \|\vec{\mathbf{v}}\|_2^2 - 2\vec{\mathbf{v}} \cdot (\sum_{j=1}^{k}\alpha_j\vec{\mathbf{u}}_j) + \sum_{j=1}^{k}\alpha_j^2, \text{ using } \vec{\mathbf{u}}_j \cdot \vec{\mathbf{u}}_j = 1 \text{ (orthonormal property)} \tag{9}$$

1.2. **Set partial derivative to 0 and solve.** Now, let's compute the partial derivative $\frac{\partial L}{\partial \alpha_i}$, set it to 0, and solve for $\alpha_i$:

$$\frac{\partial L}{\partial \alpha_i} = \frac{\partial}{\partial \alpha_i}\left(\|\vec{\mathbf{v}}\|_2^2 - 2\vec{\mathbf{v}} \cdot (\sum_{j=1}^{k} \alpha_j \vec{\mathbf{u}}_j) + \sum_{j=1}^{k} \alpha_j^2\right) \tag{10}$$

$$= -2\vec{\mathbf{v}} \cdot \vec{\mathbf{u}}_i + 2\alpha_i = 0 \tag{11}$$

$$\rightarrow \alpha_i = \vec{\mathbf{v}} \cdot \vec{\mathbf{u}}_i \tag{12}$$

Finally, plugging $\alpha_i = \vec{\mathbf{v}} \cdot \vec{\mathbf{u}}_i$ (eq. (12)) back into $\widehat{\vec{\mathbf{v}}}$:

$$\widehat{\vec{\mathbf{v}}} = \sum_{j=1}^{k} \alpha_j \vec{\mathbf{u}}_j = \sum_{j=1}^{k} (\vec{\mathbf{v}} \cdot \vec{\mathbf{u}}_j)\vec{\mathbf{u}}_j \tag{13}$$

This is what we wanted to prove.

## 2. Proof for best orthonormal basis

**Goal:** Given a centered dataset $\{\vec{\mathbf{v}}_i \in \mathbb{R}^d\}_{i=1}^N$ (i.e. mean of dataset is $\vec{\mathbf{0}}$), find an orthonormal basis $\{\vec{\mathbf{u}}_j \in \mathbb{R}^d\}_{j=1}^k$ that minimizes the following reconstruction error:

$$\min \frac{1}{N} \sum_{i=1}^{N} \|\vec{\mathbf{v}}_i - \widehat{\vec{\mathbf{v}}}_i\|_2^2 \tag{14}$$

**Claim:** The best orthonormal basis $\{\vec{\mathbf{u}}_j \in \mathbb{R}^d\}_{j=1}^k$ that minimizes the above reconstruction error is the $k$ eigenvectors with the largest eigenvalues of the following symmetric matrix $\mathbf{M} \in \mathbb{R}^{N \times N}$:

$$\mathbf{M} = \frac{1}{N}\mathbf{A}\mathbf{A}^T \tag{15}$$

where $\mathbf{A} \in \mathbb{R}^{d \times N}$ and contains $\vec{\mathbf{v}}_i$ as its column vectors, and $\mathbf{A}^T \in \mathbb{R}^{N \times d}$ and contains $\vec{\mathbf{v}}_i^T$ as its row vectors:

$$\mathbf{A} = \begin{bmatrix} | & | & \vdots & | \\ \vec{\mathbf{v}}_1 & \vec{\mathbf{v}}_2 & \vdots & \vec{\mathbf{v}}_N \\ | & | & \vdots & | \end{bmatrix}, \mathbf{A}^T = \begin{bmatrix} - & \vec{\mathbf{v}}_1^T & - \\ - & \vec{\mathbf{v}}_2^T & - \\ - & \cdots & - \\ - & \vec{\mathbf{v}}_N^T & - \end{bmatrix} \tag{16}$$

2.1. **Re-write average reconstruction error objective.** Suppose we have an orthonormal basis $\{\vec{\mathbf{u}}_j \in \mathbb{R}^d\}_{j=1}^d$.

Then, we can write our datapoints $\vec{\mathbf{v}}_i$ in terms of the basis:

$$\vec{\mathbf{v}}_i = \alpha_1 \vec{\mathbf{u}}_1 + \alpha_2 \vec{\mathbf{u}}_2 + \cdots + \alpha_d \vec{\mathbf{u}}_d = \sum_{j=1}^{d} \alpha_j \vec{\mathbf{u}}_j \tag{17}$$

Now, let's use only the first $k$ orthonormal vectors $\vec{\mathbf{u}}_1, \vec{\mathbf{u}}_2, \ldots, \vec{\mathbf{u}}_k$ to approximate a datapoint $\vec{\mathbf{v}}_i$:

$$\vec{\mathbf{v}}_i \approx \widehat{\vec{\mathbf{v}}}_i = \alpha_1 \vec{\mathbf{u}}_1 + \alpha_2 \vec{\mathbf{u}}_2 + \cdots + \alpha_k \vec{\mathbf{u}}_k = \sum_{j=1}^{k} \alpha_j \vec{\mathbf{u}}_j \tag{18}$$

Now, we'll show that the average reconstruction error objective can be re-written as follows:

$$L = \frac{1}{N} \sum_{i=1}^{N} \|\vec{\mathbf{v}}_i - \widehat{\vec{\mathbf{v}}}_i\|_2^2 = \sum_{j=k+1}^{d} \vec{\mathbf{u}}_j^T \mathbf{M} \vec{\mathbf{u}}_j \tag{19}$$

First, we'll focus on re-writing the squared Euclidean distance term (omitting $i$ for now). Plugging in eq. (17) and eq. (18) for $\vec{\mathbf{v}}$ and $\widehat{\vec{\mathbf{v}}}$:

$$\|\vec{\mathbf{v}} - \widehat{\vec{\mathbf{v}}}\|_2^2 = \|\alpha_{k+1} \vec{\mathbf{u}}_{k+1} + \cdots + \alpha_d \vec{\mathbf{u}}_d\|_2^2 = \|\sum_{j=k+1}^{d} \alpha_j \vec{\mathbf{u}}_j\|_2^2 \tag{20}$$

$$= (\sum_{j=k+1}^{d} \alpha_j \vec{\mathbf{u}}_j)^T (\sum_{j=k+1}^{d} \alpha_j \vec{\mathbf{u}}_j), \text{ using } \|\vec{\mathbf{v}}\|_2^2 = \vec{\mathbf{v}}^T \vec{\mathbf{v}} \tag{21}$$

$$= (\sum_{j=k+1}^{d} \alpha_j \vec{\mathbf{u}}_j^T)(\sum_{j=k+1}^{d} \alpha_j \vec{\mathbf{u}}_j) = \sum_{i=k+1}^{d} \sum_{j=k+1}^{d} \alpha_i \alpha_j \vec{\mathbf{u}}_i^T \vec{\mathbf{u}}_j \tag{22}$$

$$= \sum_{j=k+1}^{d} \alpha_j^2 \vec{\mathbf{u}}_j^T \vec{\mathbf{u}}_j, \text{ using } \vec{\mathbf{u}}_i^T \vec{\mathbf{u}}_j = 0 \text{ if } i \neq j \text{ (orthonormal property)} \tag{23}$$

$$= \sum_{j=k+1}^{d} \alpha_j^2, \text{ using } \vec{\mathbf{u}}_i^T \vec{\mathbf{u}}_j = 1 \text{ if } i = j \text{ (orthonormal property)} \tag{24}$$

$$= \sum_{j=k+1}^{d} (\vec{\mathbf{v}} \cdot \vec{\mathbf{u}}_j)^2, \text{ using } \alpha_j = \vec{\mathbf{v}} \cdot \vec{\mathbf{u}}_j \text{ (best projection)} \tag{25}$$

$$= \sum_{j=k+1}^{d} (\vec{\mathbf{u}}_j^T \vec{\mathbf{v}})(\vec{\mathbf{v}}^T \vec{\mathbf{u}}_j), \text{ using } \vec{\mathbf{u}} \cdot \vec{\mathbf{v}} = \vec{\mathbf{u}}^T \vec{\mathbf{v}} = \vec{\mathbf{v}}^T \vec{\mathbf{u}} = \vec{\mathbf{v}} \cdot \vec{\mathbf{u}} \tag{26}$$

Now, plugging eq. (26) back into our objective function (and adding $i$ back in), we get the following:

$$L = \frac{1}{N} \sum_{i=1}^{N} \|\vec{\mathbf{v}}_i - \widehat{\vec{\mathbf{v}}}_i\|_2^2 = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=k+1}^{d} (\vec{\mathbf{u}}_j^T \vec{\mathbf{v}}_i)(\vec{\mathbf{v}}_i^T \vec{\mathbf{u}}_j) \tag{27}$$

$$= \sum_{j=k+1}^{d} \vec{\mathbf{u}}_j^T (\frac{1}{N} \sum_{i=1}^{N} \vec{\mathbf{v}}_i \vec{\mathbf{v}}^T) \vec{\mathbf{u}}_j \tag{28}$$

$$= \sum_{j=k+1}^{d} \vec{\mathbf{u}}_j^T (\frac{1}{N} \mathbf{A} \mathbf{A}^T) \vec{\mathbf{u}}_j, \text{ using matrix multiplication definition} \tag{29}$$

$$= \sum_{j=k+1}^{d} \vec{\mathbf{u}}_j^T \mathbf{M} \vec{\mathbf{u}}_j, \text{ using } \mathbf{M} = \frac{1}{N}\mathbf{A}\mathbf{A}^T \tag{30}$$

Thus, we've shown eq. (19), namely that our objective $L$ can be re-written as eq. (30).

2.2. **Solve constrained optimization problem.** Recall our goal: find $k$ orthonormal vectors $\{\vec{\mathbf{u}}_j \in \mathbb{R}^d\}_{j=1}^k$ that minimizes our objective $L$. We can write this explicitly as a constrained optimization problem ($\forall$ = "for all"):

$$\min \sum_{j=k+1}^{d} \vec{\mathbf{u}}_j^T \mathbf{M} \vec{\mathbf{u}}_j \text{ subject to (s.t.) } \vec{\mathbf{u}}_j^T \vec{\mathbf{u}}_j = 1, \forall j \text{ (unit vector)} \tag{31}$$

Using Lagrange multipliers (see section 3), we can solve this constrained optimization problem by solving the following Lagrange function $\mathcal{L}$:

$$\mathcal{L} = \sum_{j=k+1}^{d} \left( \vec{\mathbf{u}}_j^T \mathbf{M} \vec{\mathbf{u}}_j + \lambda_j (1 - \vec{\mathbf{u}}_j^T \vec{\mathbf{u}}_j) \right) \tag{32}$$

To solve this, we set the partial derivatives $\frac{\partial \mathcal{L}}{\partial \lambda_i} = 0, \frac{\partial \mathcal{L}}{\vec{\mathbf{u}}_i} = \vec{\mathbf{0}}$ and solving for the variables.

$$\frac{\partial \mathcal{L}}{\partial \lambda_i} = 1 - \vec{\mathbf{u}}_i^T \vec{\mathbf{u}}_i = 0 \tag{33}$$

$$\rightarrow \vec{\mathbf{u}}_i^T \vec{\mathbf{u}}_i = 1 \tag{34}$$

$$\frac{\partial \mathcal{L}}{\vec{\mathbf{u}}_i} = \vec{\mathbf{u}}_i (\mathbf{M} + \mathbf{M}^T) - \lambda_i \vec{\mathbf{u}}_i^T (\mathbf{I} + \mathbf{I}^T), \text{ using } \frac{\partial \vec{\mathbf{x}}^T \mathbf{B} \vec{\mathbf{x}}}{\partial \vec{\mathbf{x}}} = \vec{\mathbf{x}}^T (\mathbf{B} + \mathbf{B}^T) \text{ (MML 5.107)} \tag{35}$$

$$= 2\vec{\mathbf{u}}_i^T \mathbf{M} - 2\lambda_i \vec{\mathbf{u}}_i^T \mathbf{I}, \text{ using symmetry: } \mathbf{M} = \mathbf{M}^T, \mathbf{I} = \mathbf{I}^T \tag{36}$$

$$= 2\vec{\mathbf{u}}_i^T \mathbf{M} - 2\lambda_i \vec{\mathbf{u}}_i^T, \text{ using } \mathbf{BI} = \mathbf{B} \text{ (MML 2.20)} \tag{37}$$

Now, by setting $\frac{\partial \mathcal{L}}{\vec{\mathbf{u}}_i} = 0$, we get the following:

$$2\vec{\mathbf{u}}_i^T \mathbf{M} = 2\lambda_i \vec{\mathbf{u}}_i^T \tag{38}$$

$$\vec{\mathbf{u}}_i^T \mathbf{M} = \lambda_i \vec{\mathbf{u}}_i^T \tag{39}$$

$$(\vec{\mathbf{u}}_i^T \mathbf{M})^T = (\lambda_i \vec{\mathbf{u}}_i^T)^T, \text{ taking the transpose on both sides} \tag{40}$$

$$\mathbf{M}^T \vec{\mathbf{u}}_i = \lambda_i \vec{\mathbf{u}}_i, \text{ using } (\mathbf{B}^T)^T = \mathbf{B}, (\mathbf{BC})^T = \mathbf{C}^T \mathbf{B}^T \text{ (MML 2.29, 2.31)} \tag{41}$$

$$\mathbf{M} \vec{\mathbf{u}}_i = \lambda_i \vec{\mathbf{u}}_i, \text{ using symmetry: } \mathbf{M} = \mathbf{M}^T \tag{42}$$

From eq. (42), we get that the solution to $\mathcal{L}$ are the eigenvectors $\vec{\mathbf{u}}_i$ of $\mathbf{M}$ with eigenvalues $\lambda_i$. Because $\mathbf{M}$ is symmetric, its eigenvectors form an orthonormal basis, thereby satisfying eq. (34), and its eigenvalues are real-valued numbers, i.e. $\lambda_i \in \mathbb{R}$. Because $\mathbf{M}$ is positive semidefinite (MML Theorem 4.14), its eigenvalues are positive, i.e. $\lambda_i > 0$ (MML pg. 106).

Now, plugging eq. (42) back into our objective eq. (30), we get the following:

$$L = \sum_{j=k+1}^{d} \vec{\mathbf{u}}_j^T \mathbf{M} \vec{\mathbf{u}}_j = \sum_{j=k+1}^{d} \vec{\mathbf{u}}_j^T (\lambda_j \vec{\mathbf{u}}_j) \tag{43}$$

$$= \sum_{j=k+1}^{d} \lambda_j (\vec{\mathbf{u}}_j^T \vec{\mathbf{u}}_j) = \sum_{j=k+1}^{d} \lambda_j, \text{ using } \vec{\mathbf{u}}_j^T \vec{\mathbf{u}}_j = 1 \text{ (orthonormal property)} \tag{44}$$

Then, to minimize $L$, we need $\vec{\mathbf{u}}_{k+1}, \ldots, \vec{\mathbf{u}}_d$ to be the $(d-k)$ eigenvectors of $\mathbf{M}$ with the smallest eigenvalues $\lambda_j$'s. Recall our original goal, which was to find the $k$ orthonormal vectors $\{\vec{\mathbf{u}}_j\}_{j=1}^{k}$ that minimizes $L$. This set of vectors then must be the $k$ eigenvectors of $\mathbf{M}$ with the largest eigenvalues $\lambda_j$'s.

From eq. (44), the average reconstruction error $L$ is the sum of the $(d-k)$ smallest eigenvalues of $\mathbf{M}$. Then, when choosing $k$, we can use the eigenvalues sorted by magnitude to decide how much error we're willing to allow in order for a small choice of $k$.

## 3. Constrained optimization

So far, we've primarily focused on unconstrained optimization problems of the following form:

$$\min_{\vec{\mathbf{x}}} f(\vec{\mathbf{x}}) \text{ or } \max_{\vec{\mathbf{x}}} f(\vec{\mathbf{x}}) \tag{45}$$

A constrained optimization adds on several constraints or equations $e_1, e_2, \ldots$ that need to be satisfied:

$$\min_{\vec{\mathbf{x}}} f(\vec{\mathbf{x}}) \text{ subject to (s.t.) } e_1, e_2, \ldots \tag{46}$$

For instance, here's a constrained problem with one constraint:

$$\min_{\vec{\mathbf{x}}} f(\vec{\mathbf{x}}) \text{ s.t. } g(\vec{\mathbf{x}}) = 0 \tag{47}$$

One way to solve constrained problems like the above is to use its analogous Lagrange function $\mathcal{L}$:

$$\mathcal{L}(\vec{\mathbf{x}}, \lambda) = f(\vec{\mathbf{x}}) + \lambda g(\vec{\mathbf{x}}) \tag{48}$$

It turns out the solution to the original constrained problem (eq. (47)) is a saddle point of $\mathcal{L}$ (eq. (48)).

A saddle point for a function occurs where all partial derivatives for that function equal 0. For instance, for $\mathcal{L}$, a saddle point occurs where

$$\frac{\partial \mathcal{L}}{\partial \vec{\mathbf{x}}} = \vec{\mathbf{0}} \text{ and } \frac{\partial \mathcal{L}}{\partial \lambda} = 0 \tag{49}$$

Consider a constrained problem with multiple constraints:

$$\min_{\vec{\mathbf{x}}} f(\vec{\mathbf{x}}) \text{ s.t. } g_j(\vec{\mathbf{x}}) = 0 \text{ for } j = 1, \ldots, M \tag{50}$$

Then, its analogous Lagrange function is given as follows:

$$\mathcal{L}(\vec{\mathbf{x}}, \lambda) = f(\vec{\mathbf{x}}) + \sum_{j=1}^{M} \lambda_j g_j(\vec{\mathbf{x}}) \tag{51}$$

and a saddle point occurs when

$$\frac{\partial \mathcal{L}}{\partial \vec{\mathbf{x}}} = \vec{\mathbf{0}} \text{ and } \frac{\partial \mathcal{L}}{\partial \lambda_j} = 0 \text{ for all } j \tag{52}$$

To find a saddle point of $\mathcal{L}$ (and a solution to the original constrained optimization problem), compute the partial derivatives of $\mathcal{L}$, set them to 0, and solve for the variables.