Principal Component Analysis Algorithm Derivation

1. Motivation

The need to reduce the dimensionality of a high feature vector to a lower feature vector with minimal loss of information. This can be used in compression (e.g., high resolution photographs to lower resolution while still being able to understand what the photograph is about) and de-noising technologies.

Mathematically, the set-up can be defined as follows:

Let $X = \{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N\}$ with $\bar{x}_i = \begin{bmatrix} x_{i,1} \\ x_{i,2} \\ \vdots \\ x_{i,D} \end{bmatrix}$ for $i \in \{1,2,\dots,N\}$, and $x_{i,j} \in \mathbb{R}$ for $j \in \{1,2,\dots,D\}$.

The objective is to find $\tilde{X} = \{\bar{\tilde{x}}_1, \bar{\tilde{x}}_2, \dots, \bar{\tilde{x}}_N\}$ with $\bar{\tilde{x}}_i = \begin{bmatrix} \tilde{x}_{i,1} \\ \tilde{x}_{i,2} \\ \vdots \\ \tilde{x}_{i,M} \end{bmatrix}$ for $i \in \{1,2,\dots,M\}$, and $\tilde{x}_{i,j} \in \mathbb{R}$ for $j \in \{1,2,\dots,D\}$, where $M < D$ and $\tilde{X}$ as closely as possible resembles X. Mathematically speaking, we wish to project X onto a lower dimensional ecosystem yielding $\tilde{X}$ such that $\tilde{X}$ is as close to X as possible. A measure of closeness can be defined by a distance function $\mathcal{D}$. Thus, the idea is to find $\bar{\tilde{x}}_i$'s such that $\sum_{i=1}^{N} \mathcal{D}(\bar{x}_i, \bar{\tilde{x}}_i)$ is minimized.

2. Orthogonal Projections

Let $\mathcal{D}(\bar{x}_i, \bar{\tilde{x}}_i) = \|\bar{x}_i - \bar{\tilde{x}}_i\| = \sqrt{(\bar{x}_i - \bar{\tilde{x}}_i)^T (\bar{x}_i - \bar{\tilde{x}}_i)} = \sum_{j=1}^{D} \left(x_{i,j} - \tilde{x}_{i,j}\right)^2$ (where $\tilde{x}_{i,j} = 0$ for $j \in \{M+1, \dots, D\}$), which is known as the Euclidean distance. This distance is minimized if each $\bar{\tilde{x}}_i$ is an orthogonal projection of $\bar{x}_i$. Mathematically, this is defined as follows:

Let $\vec{b}_1, \vec{b}_2, \dots, \vec{b}_D$ form the basis set that spans the D dimensional subspace. Let the first M terms denote the basis vectors responsible for spanning M dimensional space. Thus, using the definition of what a basis is, there exists $\beta_{i,j} \in \mathbb{R}$, for $i \in \{1,2,\dots,N\}$ and $j \in \{1,2,\dots,M\}$ such that $\bar{\tilde{x}}_i = \sum_{j=1}^{M} \beta_{i,j} \vec{b}_j$ (note $\vec{b}_j = \begin{bmatrix} b_{1,j} \\ b_{2,j} \\ \vdots \\ b_{M,j} \\ \vdots \\ b_{D,j} \end{bmatrix}$)

If $\bar{\tilde{x}}_i$ is an orthogonal projection of $\bar{x}_i$, then the vector $\bar{x}_i - \bar{\tilde{x}}_i$ is orthogonal to each $\beta_{i,j}$. In other words, $< \bar{x}_i - \bar{\tilde{x}}_i, \beta_{i,j} > = 0$, for $j \in \{1,2,\dots,M\}$.

3. Defining a Loss Function

Given that an orthogonal projection minimizes the Euclidean distance, we require $\tilde{X}$ to be a matrix consisting of column vectors orthogonal to the column vectors captured by X. We add another restriction: column vectors of $\tilde{X}$ are such that they minimize the following loss function $\mathcal{J}(X,\tilde{X}) = \frac{1}{N}\sum_{i=1}^{N}\|\bar{x}_i - \tilde{\bar{x}}_i\|^2$, (where $\tilde{x}_{i,j} = 0$ for $j \in \{M+1,\dots,D\}$).
The task then becomes finding $\beta_{i,j}$ and $\vec{b}_j$, $j = 1,2,\dots,M$.

4. Setting up the Problem

We impose two restrictions to facilitate derivations:

(1) $\bar{x}_{i,j}$'s across i's are centralized (i.e., they have an average of zero) for $i \in \{1,2,\dots,N\}$, $j \in \{1,2,\dots,D\}$

(2) $\vec{b}_j$'s form an orthonormal basis (i.e., $\vec{b}_j^T\vec{b}_j = 1$) for $j \in \{1,2,\dots,M\}$

The goal now becomes finding β_j and $\vec{b}_j$, $j = 1,2,\dots,M$ that minimize $\mathcal{J}(X,\tilde{X})$.
Since $\mathcal{J}(X,\tilde{X}) = \frac{1}{N}\sum_{i=1}^{N}\|\bar{x}_i - \tilde{\bar{x}}_i\|^2$ and $\tilde{\bar{x}}_i$ is the variable influenced by varying $\beta_{i,j}$ and $\vec{b}_j$, therefore, $\mathcal{J}(X,\tilde{X})$ is a convex function with respect to $\beta_{i,j}$ and $\vec{b}_j$, and thus, we can minimize this function by taking partial derivatives and setting them to zero. We begin by applying the chain rule as follows:

$$\frac{\partial\mathcal{J}(X,\tilde{X})}{\partial\beta_{i,j}} = \frac{\partial\mathcal{J}(X,\tilde{X})}{\partial\tilde{\bar{x}}_i} \times \frac{\partial\tilde{\bar{x}}_i}{\partial\beta_{i,j}}$$

$$\frac{\partial\mathcal{J}(X,\tilde{X})}{\partial b_j} = \frac{\partial\mathcal{J}(X,\tilde{X})}{\partial\tilde{\bar{x}}_i} \times \frac{\partial\tilde{\bar{x}}_i}{\partial b_j}$$

We can calculate $\frac{\partial\mathcal{J}(X,\tilde{X})}{\partial\bar{x}_i}$ as follows:

$$\frac{\partial\mathcal{J}(X,\tilde{X})}{\partial\bar{x}_i} = \frac{\partial\left(\frac{1}{N}\sum_{i'=1}^{N}\|\bar{x}_{i'} - \tilde{\bar{x}}_{i'}\|^2\right)}{\partial\tilde{\bar{x}}_i} = \frac{1}{N} \times \frac{\partial\left(\sum_{i'=1}^{N}(\bar{x}_{i'} - \tilde{\bar{x}}_{i'})^T \times (\bar{x}_{i'} - \tilde{\bar{x}}_{i'})\right)}{\partial\tilde{\bar{x}}_i}$$

$$= \frac{1}{N} \times \frac{\partial\left((\bar{x}_i - \tilde{\bar{x}}_i)^T \times (\bar{x}_i - \tilde{\bar{x}}_i)\right)}{\partial\tilde{\bar{x}}_i}$$

$$= \frac{1}{N}\left[\frac{\partial\left(\sum_{j=1}^{D}\left(x_{i,j} - \tilde{x}_{i,j}\right)^2\right)}{\partial\tilde{x}_{i,1}} \quad \frac{\partial\left(\sum_{j=1}^{D}\left(x_{i,j} - \tilde{x}_{i,j}\right)^2\right)}{\partial\tilde{x}_{i,2}} \quad \cdots \quad \frac{\partial\left(\sum_{j=1}^{D}\left(x_{i,j} - \tilde{x}_{i,j}\right)^2\right)}{\partial\tilde{x}_{i,D}}\right]$$

$$= \frac{1}{N}\left[-2\left(x_{i,1} - \tilde{x}_{i,1}\right) \quad -2\left(x_{i,2} - \tilde{x}_{i,2}\right) \quad \cdots \quad -2\left(x_{i,D} - \tilde{x}_{i,D}\right)\right]$$

$$= -\frac{2}{N}\left[x_{i,1} - \tilde{x}_{i,1} \quad x_{i,2} - \tilde{x}_{i,2} \quad \cdots \quad x_{i,D} - \tilde{x}_{i,D}\right]$$

$$= -\frac{2}{N}(\bar{x}_i - \tilde{\bar{x}}_i)^T$$

The next two sections aim to derive $\frac{\partial\tilde{\bar{x}}_i}{\partial\beta_{i,j}}$ and $\frac{\partial\tilde{\bar{x}}_i}{\partial\vec{b}_j}$, and thus consequently $\frac{\partial\mathcal{J}(X,\tilde{X})}{\partial\beta_{i,j}}$ and $\frac{\partial\mathcal{J}(X,\tilde{X})}{\partial\vec{b}_j}$

5. Calculation of Basis Coefficients

Since, $\bar{\bar{x}}_i = \sum_{i=1}^{M} \beta_{i,j} \vec{b}_j$, we thus have $\frac{\partial \bar{\bar{x}}_i}{\partial \beta_{i,j}} = \frac{\partial\left(\sum_{i'=1}^{M} \beta_{i',j} \vec{b}_j\right)}{\partial \beta_{i,j}} = \frac{\partial(\beta_{i,j} \vec{b}_j)}{\partial \beta_{i,j}} = \vec{b}_j$. Therefore, we have the following result: $\frac{\partial \bar{\bar{x}}_i}{\partial \beta_{i,j}} = -\frac{2}{N}(\bar{x}_i - \bar{\bar{x}}_i)^T \vec{b}_j$

We now proceed to expand $\bar{\bar{x}}_i$ and conduct algebra to arrive at a final expression:

$$\begin{aligned}\frac{\partial \bar{\bar{x}}_i}{\partial \beta_{i,j}} &= -\frac{2}{N}(\bar{x}_i - \bar{\bar{x}}_i)^T \vec{b}_j \\ &= -\frac{2}{N}\left(\bar{x}_i - \sum_{j'=1}^{M} \beta_{i,j'} \vec{b}_{j'}\right)^T \vec{b}_j \\ &= -\frac{2}{N}\left(\bar{x}_i^T b_j - \sum_{j'=1}^{M} \beta_{i,j'} \vec{b}_{j'}^T \vec{b}_j\right) \\ &= -\frac{2}{N}\left(\bar{x}_i^T \vec{b}_j - \beta_{i,j}\right)\end{aligned}$$

Note that $\frac{\partial \mathcal{J}^2(X,\tilde{X}\,)}{\partial \beta_{i,j}^2} = \frac{2}{N} > 0$ confirming that the loss function is convex with respect to $\beta_{i,j}$

Now, setting the derivative to zero, we have $-\frac{2}{N}\left(\bar{x}_i^T \vec{b}_j - \beta_{i,j}\right) = 0$, and thus,

$$\boxed{\beta_{i,j} = \bar{x}_i^T \vec{b}_j}$$

6. Calculation of Basis Vectors

We begin by rewriting $\bar{\bar{x}}_i$ and $\bar{x}_i$:

$$\bar{\bar{x}}_i = \sum_{j=1}^{M} \beta_{i,j} \vec{b}_j = \sum_{j=1}^{M} \left(\bar{x}_i^T \vec{b}_j\right) \vec{b}_j = \sum_{j=1}^{M} \vec{b}_j \times \bar{x}_i^T \vec{b}_j = \sum_{j=1}^{M} \vec{b}_j \times \left(\vec{b}_j^T \bar{x}_i\right) = \sum_{j=1}^{M} \left(\vec{b}_j \vec{b}_j^T\right) \bar{x}_i$$

$\bar{x}_i$ can be written as $\sum_{j=1}^{D} \beta_{i,j} \vec{b}_j$ since $\vec{b}_j$'s span the vector space of dimension D (we let the $\beta_{i,j}$'s be the same as those used in expressing $\bar{\bar{x}}_i$ and thus for $j > M$, the $\beta_{i,j}$'s are set such that when multiplied by $\vec{b}_j$ and added, we end up with $\bar{x}_i$) Thus,

$$\bar{x}_i = \sum_{j=1}^{D} \left(\vec{b}_j \vec{b}_j^T\right) \bar{x}_i = \sum_{j=1}^{M} \left(\vec{b}_j \vec{b}_j^T\right) \bar{x}_i + \sum_{j=M+1}^{D} \left(\vec{b}_j \vec{b}_j^T\right) \bar{x}_i$$

Hence, $\bar{x}_i - \bar{\bar{x}}_i = \sum_{j=M+1}^{D} \left(\vec{b}_j \vec{b}_j^T\right) \bar{x}_i$

Thus, we can rewrite the loss function equation as follows:

$$
\begin{aligned}
\mathcal{J}(X,\tilde{X}\,) &= \frac{1}{N}\sum_{i=1}^{N}\|\bar{x}_i - \bar{\tilde{x}}_i\|^{2} = \frac{1}{N}\sum_{i=1}^{N}\left\|\sum_{j=M+1}^{D}\left(\vec{b}_j\vec{b}_j^T\right)\bar{x}_i\right\|^{2} \\
&= \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{j=M+1}^{D}\left(\vec{b}_j\vec{b}_j^T\right)\bar{x}_i\right)^{T}\left(\sum_{j=M+1}^{D}\left(\vec{b}_j\vec{b}_j^T\right)\bar{x}_i\right) \\
&= \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{j=M+1}^{D}\left(\vec{b}_j^T\bar{x}_i\right)\vec{b}_j^T\right)\left(\sum_{j=M+1}^{D}\vec{b}_j\left(\vec{b}_j^T\bar{x}_i\right)\right) \\
&= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=M+1}^{D}\sum_{j'=M+1}^{D}\left(\vec{b}_j^T\bar{x}_i\right)\vec{b}_j^T\vec{b}_{j'}\left(\vec{b}_{j'}^T\bar{x}_i\right) \\
&= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=M+1}^{D}\left(\vec{b}_j^T\bar{x}_i\right)^{2} \\
&= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=M+1}^{D}\vec{b}_j^T\bar{x}_i\bar{x}_i^T\vec{b}_j \\
&= \frac{1}{N}\sum_{j=M+1}^{D}\sum_{i=1}^{N}\vec{b}_j^T\bar{x}_i\bar{x}_i^T\vec{b}_j \\
&= \sum_{j=M+1}^{D}\vec{b}_j^T \times \left(\frac{1}{N}\sum_{i=1}^{N}\bar{x}_i\bar{x}_i^T\right) \times \vec{b}_j \\
&= \sum_{j=M+1}^{D}\vec{b}_j^T \times S \times \vec{b}_j
\end{aligned}
$$

where $S = \frac{1}{N}\sum_{i=1}^{N}\bar{x}_i\bar{x}_i^T$, which is the covariance matrix of X (whose features each have a zero).

We use the Lagrangian method to find b_j that minimizes $\mathcal{J}(X,\tilde{X}\,)$, for $j \in \{1,2,\ldots,D\}$:

Let $\mathcal{L} = \mathcal{J}(X,\tilde{X}\,) + \sum_{j=M+1}^{D}\lambda_j\left(\vec{b}_j^T\vec{b}_j - 1\right)$, with the constraint $\vec{b}_j^T\vec{b}_j = 1$:

$$
\begin{aligned}
\frac{\partial\mathcal{L}}{\partial\lambda_j} &= \frac{\partial\mathcal{J}(X,\tilde{X}\,)}{\partial\lambda_j} + \frac{\partial\left(\sum_{j'=M+1}^{D}\lambda_{j'}\left(\vec{b}_{j'}^T\vec{b}_{j'} - 1\right)\right)}{\partial\lambda_j} \\
&= 0 + \frac{\partial\left(\lambda_j\left(\vec{b}_j^T\vec{b}_j - 1\right)\right)}{\partial\lambda_j} \\
&= \vec{b}_j^T\vec{b}_j - 1
\end{aligned}
$$

Setting the derivative to zero, we get $\vec{b}_j^T\vec{b}_j = 1$ and letting $s_{i,j}$ represent the ith row and jth column element of the matrix S, we proceed to derive $\frac{\partial\mathcal{L}}{\partial\vec{b}_j}$:

$$\frac{\partial\mathcal{L}}{\partial\vec{b}_j} = \frac{\partial\mathcal{J}(X,\widetilde{X}\,)}{\partial\vec{b}_j} + \frac{\partial\left(\sum_{j^{'}=M+1}^{D}\lambda_{j^{'}}\left(\vec{b}_{j^{'}}^{T}\vec{b}_{j^{'}}-1\right)\right)}{\partial\vec{b}_j}$$

$$= \frac{\partial\left(\sum_{j'=M+1}^{D}\vec{b}_{j'}^{T}\times S\times\vec{b}_{j'}\right)}{\partial\vec{b}_j} + \frac{\partial\left(\sum_{j^{'}=M+1}^{D}\lambda_{j^{'}}\left(\vec{b}_{j^{'}}^{T}\vec{b}_{j^{'}}-1\right)\right)}{\partial\vec{b}_j}$$

$$= \frac{\partial\left(\vec{b}_j^{T}\times S\times\vec{b}_j\right)}{\partial\vec{b}_j} + \frac{\partial\left(\lambda_j\left(\vec{b}_j^{T}\vec{b}_j-1\right)\right)}{\partial\vec{b}_j}$$

$$= \left[\frac{\partial\left(\sum_{j'=1}^{D}b_{j',j}\sum_{j''=1}^{D}s_{j',j''}b_{j'',j}\right)}{\partial b_{1,j}} \quad \frac{\partial\left(\sum_{j'=1}^{D}b_{j',j}\sum_{j''=1}^{D}s_{j',j''}b_{j'',j}\right)}{\partial b_{2,j}} \quad \dots \quad \frac{\partial\left(\sum_{j'=1}^{D}b_{j',j}\sum_{j''=1}^{D}s_{j',j''}b_{j'',j}\right)}{\partial b_{2,j}}\right]$$

$$+ \left[\frac{\partial\left(\lambda_j\left(\sum_{j'=1}^{D}b_{j',j}^{2}-1\right)\right)}{\partial b_{1,j}} \quad \frac{\partial\left(\lambda_j\left(\sum_{j'=1}^{D}b_{j',j}^{2}-1\right)\right)}{\partial b_{2,j}} \quad \dots \quad \frac{\partial\left(\lambda_j\left(\sum_{j'=1}^{D}b_{j',j}^{2}-1\right)\right)}{\partial b_{D,j}}\right]$$

Note that for a given k, $\frac{\partial\left(\sum_{j'=1}^{D}b_{j',j}\sum_{j''=1}^{D}s_{j',j''}b_{j'',j}\right)}{\partial b_{k,j}}$ can be written as follows

$$\frac{\partial\left(\sum_{j'=1}^{D}b_{j',j}\sum_{j''=1}^{D}s_{j',j''}b_{j'',j}\right)}{\partial b_{k,j}} = \frac{\partial\left(b_{j',k}\sum_{j''=1}^{D}s_{j',j''}b_{j'',j}\right)}{\partial b_{k,j}} + \frac{\partial\left(\sum_{j'=1}^{D}b_{j',j}s_{j',k}b_{k,j}\right)}{\partial b_{k,j}}$$

$$= \left(\sum_{j^{''}=1}^{D}s_{k,j^{''}}b_{j^{''},j}\right) + \left(\sum_{j^{'}=1}^{D}b_{j^{'},k}s_{j^{'},k}\right)$$

Noting S is symmetric, and thus $s_{n,m} = s_{m,n}$ the above summation can be combined into one:

$$\left(\sum_{j''=1}^{D}s_{k,j''}b_{j'',j}\right) + \left(\sum_{j'=1}^{D}b_{j',k}s_{j',k}\right) = 2\times\sum_{j''=1}^{D}s_{k,j''}b_{j'',j}$$

Thus, we have

$$\frac{\partial\mathcal{L}}{\partial\vec{b}_j} = \left[2\left(\sum_{j^{''}=1}^{D}s_{1,j^{''}}b_{j^{''},j}\right) \quad 2\left(\sum_{j^{''}=1}^{D}s_{2,j^{''}}b_{j^{''},j}\right) \quad \dots \quad 2\left(\sum_{j^{''}=1}^{D}s_{D,j^{''}}b_{j^{''},j}\right)\right]$$
$$+[2\times\lambda_j\times b_{1,j} \quad 2\times\lambda_j\times b_{2,j} \quad \dots \quad 2\times\lambda_j\times b_{D,j}]$$

$$= 2\vec{b}_j^{T}\times S + 2\lambda_j\vec{b}_j^{T}$$
$$= 2\left(\vec{b}_j^{T}\times S - \lambda_j\vec{b}_j^{T}\right)$$

Setting the derivative to zero, and noting that $S^T = S$, we have the following result

$$\vec{b}_j^{T}\times S = \lambda_j\vec{b}_j^{T} \Leftrightarrow S^T\vec{b}_j = \lambda_j\vec{b}_j \Leftrightarrow S\vec{b}_j = \lambda_j\vec{b}_j$$

Thus, the problem is reduced to an eigenvector/eigenvalue problem.

Now, note that since $S\vec{b}_j = \lambda_j\vec{b}_j$, therefore, $\mathcal{J}\left(X,\tilde{X}\,\right)$ at a minimum value can be written as follows:

$$\mathcal{J}(X, \tilde{X}) = \sum_{j=M+1}^{D} \vec{b}_j^T \times S \times \vec{b}_j = \sum_{j=M+1}^{D} \vec{b}_j^T \times \lambda_j \vec{b}_j = \sum_{j=M+1}^{D} \lambda_j \left(\vec{b}_j^T \times \vec{b}_j\right) = \sum_{j=M+1}^{D} \lambda_j$$

Therefore, for $\mathcal{J}(X, \tilde{X})$ to be minimized, $\lambda_{M+1}, \lambda_{M+2}, \dots, \lambda_D$ need to be the smallest eigenvalues. This means that $\lambda_1, \lambda_2, \dots, \lambda_M$ need to be the first M largest eigenvalues.

Since M can vary depending on how much we would like to implement dimensionality reduction, we choose the eigenvalues in the following order: $\lambda_1 > \lambda_2 > \cdots > \lambda_M$.

And so, the b_j that minimizes $\mathcal{J}(X, \tilde{X})$ corresponds to the eigenvector corresponding to the jth largest eigenvalue.

7. Handling Case When D > N

Solving the eigenvector/eigenvalue problem $S\vec{b}_j = \lambda_j \vec{b}_j$ can be numerically cumbersome if D is very large. There is a trick we can implement if $D > N$. Noting that $S = \frac{1}{N}\sum_{i=1}^{N} \vec{x}_i^T \vec{x}_i = \frac{1}{N} X^T X$

$$S\vec{b}_j = \lambda_j \vec{b}_j \Leftrightarrow \frac{1}{N} X^T X \vec{b}_j = \lambda_j \vec{b}_j$$
$$\Leftrightarrow \frac{1}{N} X X^T X \vec{b}_j = \lambda_j X \vec{b}_j$$
$$\Leftrightarrow \frac{1}{N} (X X^T)\left(X \vec{b}_j\right) = \lambda_j \left(X \vec{b}_j\right)$$

Let $\frac{1}{N}(XX^T) = S'$ which is $N \times N$ and $X\vec{b}_j = \vec{b}_j'$
Thus, $S'\vec{b}_j' = \lambda_j \vec{b}_j'$

Thus, we have an eigenvector/eigenvalue problem. Finding $\vec{b}_j$ can be obtained as follows:

$$S'\vec{b}_j' = \lambda_j \vec{b}_j' \Leftrightarrow \frac{1}{N}(XX^T)\vec{b}_j' = \lambda_j \vec{b}_j'$$
$$\Leftrightarrow \frac{1}{N} X^T (XX^T)\vec{b}_j' = \lambda_j X^T \vec{b}_j'$$
$$\Leftrightarrow \frac{1}{N} (X^T X) X^T \vec{b}_j' = \lambda_j X^T \vec{b}_j'$$
$$\Leftrightarrow S\left(X^T \vec{b}_j'\right) = \lambda_j \left(X^T \vec{b}_j'\right)$$

Since $S\vec{b}_j = \lambda_j \vec{b}_j$, and $\vec{b}_j$ is orthonormal, therefore $\vec{b}_j = \frac{X^T \vec{b}_j'}{\left\|X^T \vec{b}_j'\right\|}$ is a viable solution (note that normalization has no effect on λ_j since $S\left(X^T \vec{b}_j'\right) = \lambda_j \left(X^T \vec{b}_j'\right) \Leftrightarrow S\left(\frac{X^T \vec{b}_j'}{\left\|X^T \vec{b}_j'\right\|}\right) = \lambda \left(\frac{X^T \vec{b}_j'}{\left\|X^T \vec{b}_j'\right\|}\right)$).

Thus, if we have a very large dimension D, in estimating the $\beta_{i,j}$'s and b_j's that minimize the loss function $\mathcal{J}\left(X, \tilde{X}\ \right)$, the greatest pain point, which is obtaining M eigenvalues from solving a system with a $D \times D$ matrix, is reduced to solving a system with an $N \times N$ matrix.