Computing Principal Components with the Eigen-decomposition of a Low Dimensional Matrix

Tony X. Han

Electrical & Computer Engineering Department
University of Missouri

When we do PCA, we need to do an eigen-decomposition of the covariance matrix. Suppose the m sample data we have are: $\tilde{\mathbf{x}}_i$, $i = 1, \dots, m$. $\tilde{\mathbf{x}}_i$'s are usually vectorized image with dimension of $N \times 1$. For example, if you want to convert an image with resolution of 300×200 to a vector, the $N = 300 \times 200 = 60\,000$, which is much larger than the number of data sample m. So, usually $N \gg m$. The procedure of PCA is as follows:

1. Compute the mean:
$$\bar{\mathbf{x}} = \frac{1}{m} \sum_{i=1}^{m} \tilde{\mathbf{x}}_i. \tag{1}$$

2. Generate the zero-mean data matrix:
$$\mathbf{x}_i = \tilde{\mathbf{x}}_i - \bar{\mathbf{x}}. \tag{2}$$
$$\mathbf{A} = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots \mathbf{x}_m \end{pmatrix} \tag{3}$$

3. Construct the covariance matrix:
$$\mathbf{C} = \mathbf{A}\mathbf{A}^T \tag{4}$$
The covariance matrix $\mathbf{C}$ is symmetric and positive definite. So the eigenvalues of $\mathbf{C}$ is real and non-negative.

4. Eigen-decomposition:

   The eigenvalues λ_i's and the eigenvectors $\mathbf{v}_i$'s of $\mathbf{C}$ satisfy
$$\mathbf{C}\mathbf{v}_i = \lambda_i \mathbf{v}_i. \tag{5}$$
Here we assume that the eigenvector is normalized, i.e., $\|\mathbf{v}_i\|_2 = 1$. Writing Eq. (5) in the matrix form, we have
$$\mathbf{C}\mathbf{V} = \mathbf{V}\mathbf{\Lambda}, \tag{6}$$
where
$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & 0 & \dots \\ 0 & \lambda_2 & 0 & \dots \\ & & \ddots & \\ 0 & \dots & 0 & \lambda_N \end{pmatrix} \tag{7}$$

is the diagonal matrix with the egenvlues λ_i's in the diagonal. In the PCA procedure, we put λ_i's in $\mathbf{\Lambda}$ with the descent order, i.e., $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$. $\mathbf{V}$ is the matrix consists of the egenvectors corresponding to λ_i's.

$$\mathbf{V} = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \ldots \mathbf{v}_N \end{pmatrix}. \tag{8}$$

So we have the eigen-decomposition of the covariance matrix:

$$\mathbf{C} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T. \tag{9}$$

The reason that $\mathbf{V}^T = \mathbf{V}^{-1}$ is due to the fact that $\mathbf{C}$ is symmetric and positive definite [1].

5. Construct the projection matrix $\mathbf{P}$:

   In order to apply PCA for dimension reduction, we can pick a number $k < N$ to form the projection matrix from the first k principal componets as

$$\mathbf{P} = \begin{pmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_k^T \end{pmatrix}. \tag{10}$$

   To represent a sample $\mathbf{x}_i$ of high dimension $(N \times 1)$, we use its low dimensional projection $\mathbf{z}_i$ $(k \times 1)$, i.e.

$$\mathbf{z}_i = \mathbf{P}\mathbf{x}_i. \tag{11}$$

If N is not very big, we can use the procedure mentioned above to compute the principal components. But for our eigen-background modeling case [2], N is usually more than 40 000. It is infeasible for us to construct the covariance matrix with the dimension $N \times N$. In our eigen-background modeling scenario, the number of data sample m is the number of image frames of the training sequence. m usually ranges from 20 to 300. So $m \ll N$ here.

Let us consider the matrix $\mathbf{D} = \mathbf{A}^T\mathbf{A}$, which is a $m \times m$ matrix. This is a very small matrix comparing with $\mathbf{C}$.

The eigenvalues $\tilde{\lambda}_i$'s and the eigenvectors $\tilde{\mathbf{v}}_i$'s of $\mathbf{D}$ satisfy

$$\mathbf{A}^T\mathbf{A}\tilde{\mathbf{v}}_i = \tilde{\lambda}_i\tilde{\mathbf{v}}_i. \tag{12}$$

We want to explore the relationships among the eigenvalues and eigenvectors of $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}\mathbf{A}^T$. Left multiply $\mathbf{A}$ at the both side of Eq. (12), we have

$$\mathbf{A}\mathbf{A}^T\mathbf{A}\tilde{\mathbf{v}}_i = \tilde{\lambda}_i\mathbf{A}\tilde{\mathbf{v}}_i. \tag{13}$$

i.e.,

$$\mathbf{C}\mathbf{A}\tilde{\mathbf{v}}_i = \tilde{\lambda}_i\mathbf{A}\tilde{\mathbf{v}}_i. \tag{14}$$

We conclude that $\tilde{\lambda}_i$ is an eigenvaule of $\mathbf{C}$, with the corresponding eigenvector $A\tilde{\mathbf{v}}_i$.

Similarly, left multiply $\mathbf{A}^T$ at the both side of Eq. (5), we have

$$\mathbf{A}^T\mathbf{A}\mathbf{A}^T\mathbf{v}_i = \lambda_i\mathbf{A}^T\mathbf{v}_i. \tag{15}$$

i.e.,

$$\mathbf{D}\mathbf{A}^T\mathbf{v}_i = \lambda_i \mathbf{A}^T\mathbf{v}_i. \tag{16}$$

So λ_i is an eigenvalue of $\mathbf{D}$, with the corresponding eigenvector $A^T\mathbf{v}_i$. Both $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$ are symmetric and postive definite matrix. So Eq.s (14) and (16) shows that $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$ have exactly same set of eigenvalues. For the corresponding eigenvectors, we can establish a bijection between $\mathbf{v}_i$s and $\tilde{\mathbf{v}}_i$s using Eq.s (14) and (16). They are mapped to each other by:

$$\mathbf{v}_i = \mathbf{A}\tilde{\mathbf{v}}_i \tag{17}$$

and

$$\tilde{\mathbf{v}}_i = \mathbf{A}^T\mathbf{v}_i. \tag{18}$$

There is an interesting question. Since the dimension of $\mathbf{A}\mathbf{A}^T$ ($N \times N$is much larger than the one of $\mathbf{A}^T\mathbf{A}$ ($m \times m$), how can they have the same amount of eigenvalues and eigenvectors?

That is because rank($\mathbf{A}\mathbf{A}^T$) = rank($\mathbf{A}^T\mathbf{A}$) = rank($\mathbf{A}$). So they have exactly same amount of non-zero eigenvalues and the eigenvectors corresponding to the non-zero eigenvalues. The principoal components are eigenvectors corresponding to the largest k eigenvectors. So we can compute the principal components of $\mathbf{A}\mathbf{A}^T$ using the eigen-decomposition of $\mathbf{A}^T\mathbf{A}$. The eigen-decomposition of $\mathbf{A}^T\mathbf{A}$ can be further simplified by Singular Value Decomposition (SVD). We can simply compute the SVD of $\mathbf{A}$. Refer to page 369-372 of [3] for details.

References

[1] G. Strang, *Linear Algebra and its Applications.* San Diego: Harcourt Brace Jonanovich, 3rd ed., 1988.

[2] N. Oliver, B. Rosario, and A. Pentland, "A bayesian computer vision system for modeling human interactions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, pp. 831–843, Aug 2000.

[3] T. K. Moon and W. C. Stirling, *Mathematical methods and algorithms for signal processing.* 2000.