Geometric Foundations of Data Analysis I: Week 6

Joshua Maglione

13 October 2025

Contents

Question 3.5. Are the principal components $k+1$ through m useless?

3.4 Projections

The answer to Question 3.5 is essentially "Yes", and it can be helpful to consider an idealised example.

Recall that X is our $m \times n$ matrix whose columns are our n data points in $\mathbb{R}^m$. The matrix P is obtained from the eigenvectors of C_X, which are the rows of P, and $Y = PX$. Moreover, P is an orthogonal matrix. Suppose $k < m$ and

$$\lambda_{k+1} = \lambda_{k+2} = \cdots = \lambda_m = 0.$$

Therefore, we have

$$C_Y = \begin{pmatrix} \lambda_1 & & & & & \\ & \ddots & & & & \\ & & \lambda_k & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}.$$

Remark 3.6. In this situation, $y_{rj} = 0$ for all $k+1 \leqslant r \leqslant m$ and all $1 \leqslant j \leqslant n$. (Can you show this?!) In other words, the "new" data point y_j has a tail of zeroes.

Lemma 3.7. *Let Q be the first k rows of P, so that Q is $k \times m$. Then for all columns $x_i, x_j \in \mathbb{R}^m$ of X,*

$$d_{\mathbb{R}^m}(x_i, x_j) = d_{\mathbb{R}^k}(Qx_i, Qx_j).$$

Proof. For each column x_i of X, we have $Px_i = y_i$, where y_i is the ith column of Y. Since P is orthogonal, $d_{\mathbb{R}^m}(Px_i, Px_j) = d_{\mathbb{R}^m}(x_i, x_j)$. Thus, by 3.6

$$\begin{aligned} d_{\mathbb{R}^m}(Px_i, Px_j)^2 &= d_{\mathbb{R}^m}(y_i, y_j)^2 \\ &= \sum_{r=1}^{m} (y_{ri} - y_{rj})^2 \\ &= \sum_{r=1}^{k} (y_{ri} - y_{rj})^2 \\ &= d_{\mathbb{R}^k}(Qx_i, Qx_j)^2. \end{aligned}$$ $\square$

The key application of Lemma 3.7 is that ignoring rows $k+1$ through m in the matrix P does not change the geometry! That is, we can project the data to a smaller dimension and distances between the data points remain unchanged!

In practice the λ_i are strictly greater than 0, so this idealised situation does not occur. Using our rule in Equation (3.6), then the projected data would approximate the original geometry quite well. The larger the ratio, the better the approximation, so there is indeed a trade off. Thus, after constructing all of the principal components, one can take the first k, and project the original data (i.e. the matrix X) into a smaller dimension by constructing the matrix Q from the first k rows of P.

3.5 PCA is always possible—the Spectral Theorem

The real power of PCA is that we can *always* perform it. One does not need to input parameters; just the data. Note there is a pre-processing stage of normalising and rescaling, but this can be applied to all data. We address Question 3.4 which, at the time, we just assumed was true. In order to do this, we prove the Spectral Theorem.

Theorem 3.8 (Spectral Theorem). *Let M be a real symmetric $n \times n$ matrix. Then $\mathbb{R}^n$ has an orthonormal basis coming from eigenvectors of M.*

Corollary 3.9. *Every eigenvalue and eigenvector of a real symmetric matrix are real.*

In particular, the Spectral Theorem makes principal component analysis possible since the covariance matrix is always a real symmetric matrix.

Remark 3.10. Covariance matrices are examples of *positive semi-definite matrices*, which are real matrices whose eigenvalues are all real and nonnegative. We will not need this much, nor will we prove this, but it might be useful to know that covariance matrices are particularly nice.

Let's unpack what the Spectral Theorem says exactly. Recall that an orthonormal basis $\{b_1, \dots, b_n\}$ for a vector space V satisfies

$$b_i \cdot b_j = \begin{cases} 1 & i = j, \\ 0 & i \neq j. \end{cases}$$

There are two components to being an orthonormal basis: the basis vectors are pairwise orthogonal and each basis vector has unit length. The latter condition is not so particularly important (though useful); really, the magic is in the pairwise orthogonal condition.

We already saw in Section 3.3 that the Spectral Theorem does not hold if we drop 'symmetric'. Now we will build our way to prove the Spectral Theorem.

Lemma 3.11. *Let M be a real symmetric matrix. Then eigenvectors corresponding to distinct eigenvalues are orthogonal.*

Proof. Let u and v be eigenvectors of M corresponding to λ and μ with $\lambda \neq \mu$. Then

$$\mu u^{\mathsf{t}} v = u^{\mathsf{t}}(Mv) = (u^{\mathsf{t}} M)v = (Mu)^{\mathsf{t}} v = \lambda u^{\mathsf{t}} v.$$

Suppose via contradiction that $u^{\mathsf{t}} v \neq 0$, but then $\mu = \lambda$, which is a contradiction, so we must have $u^{\mathsf{t}} v = 0$. Hence, u and v are orthogonal. $\square$