Principal Component Analysis:

Consider a n-dimensional random vector Y with expectation value $\eta := \mathbb{E}(Y)$ and covariance matrix $C := \operatorname{Cov}(Y)$. The main idea of the principal component analysis is to transform Y in an random vector Z such that is expectation values is zero and its covariance matrix is a diagonal matrix. The values of the diagonal then describe how the individual components of Z contribute to the total variance of Z.

We can use, e. g., a singular value decomposition of C for this transformation. Because C is symmetric, we get a singular value decomposition

$$C = V\Sigma V^T$$

where V is an orthogonal matrix and Σ is a diagonal matrix. We can then define

$$Z := V^T(Y - \eta)$$

which implies

$$\mathbb{E}(Z) = \mathbb{E}(V^T(Y - \eta)) = V^T(\mathbb{E}(Y) - \eta)) = V^T(\eta - \eta)) = 0$$

and

$$\operatorname{Cov}(Z) = \operatorname{Cov}(V^T(Y-\eta)) = \operatorname{Cov}(V^T Y) = V^T \operatorname{Cov}(Y)V = V^T \operatorname{Cov}(C)V = V^T V\Sigma V^T V = \Sigma.$$

Thus Z is our desired transformed random vector and Σ the corresponding diagonal covariance matrix.

Usually, however, η and C and thus V and Σ are not known and must be estimated from data. Let us assume for this that we have m data vectors $y_i \in \mathbb{R}^n$ for $i \in \{1, \dots, m\}$ which are realizations of Y. The vector

$$\bar{y} := \frac{1}{m}\sum_{i=1}^{m} y_i$$

is then an estimate of η. Furthermore

$$\bar{C} := \frac{1}{m-1}\sum_{i=1}^{m}(y_i - \bar{y})(y_i - \bar{y})^T$$

is an estimate of C.

The matrix $\bar{C}$ can also be written as matrix product using the data matrix,

$$A \in \mathbb{R}^{m\times n} \text{ with } A_{i*} = y_i \text{ for all } i \in \{1, \dots, m\}$$

i. e. each row contains one data vector, and the mean data matrix

$$\bar{A} \in \mathbb{R}^{m\times n} \text{ with } A_{i*} = \bar{y} \text{ for all } i \in \{1, \dots, m\},$$

i. e. each row contains the mean data vector $\bar{y}$. Then $\bar{C}$ can be written as

$$\bar{C} = \frac{1}{m-1}B^T B \text{ with } B := A - \bar{A}.$$

We can now use a singular value decomposition of B to estimate Σ and V. Let $\hat{U}\hat{\Sigma}\hat{V}^T$ be a singular value decomposition of B. Then

$$B = \hat{U}\hat{\Sigma}\hat{V}^T$$

where $\hat{U}$ and $\hat{V}$ are orthogonal matrices and $\hat{\Sigma}$ is a diagonal matrix. Furthermore

$$\bar{C} = \tfrac{1}{m-1} B^T B = \tfrac{1}{m-1} (\hat{U}\hat{\Sigma}\hat{V}^T)^T (\hat{U}\hat{\Sigma}\hat{V}^T) = \tfrac{1}{m-1} \hat{V}\hat{\Sigma}^T\hat{U}^T\hat{U}\hat{\Sigma}\hat{V}^T = \tfrac{1}{m-1} \hat{V}\hat{\Sigma}^T\hat{\Sigma}\hat{V}^T.$$

Because $\hat{\Sigma}$ is a diagonal matrix,

$$D := \tfrac{1}{m-1} \hat{\Sigma}^T\hat{\Sigma}$$

is a diagonal matrix as well and

$$\bar{C} = \hat{V}D\hat{V}^T.$$

Thus $\hat{V}$ is an estimation of V and D and estimation of Σ.

We can apply the transformation implied by Z to our data matrix A to get our transformed data matrix $\hat{A}$ by

$$\hat{A} := (A - \bar{A})\hat{V}.$$