Reading Report:
A Tutorial on Principal Component Analysis

Boyuan Du (2024151470021)
School of Software, **Sichuan University**
Email:2024151470021@stu.scu.edu.cn

September 20, 2025

Course / Assignment: Reading report on PCA .
Primary Source: J. Shlens, "A Tutorial on Principal Component Analysis," 2014.

Abstract

PCA is a powerful and fundamental technique for extracting useful information from complex(high-dimensional) data by lowering dimension onto basic features of the data. The essence of PCA lies in its ability to identify the directions of maximum variance in the data, which are represented by the principal components. By projecting the data onto these components, PCA effectively reduces redundancy and highlights the most informative aspects of the dataset. The tutorial emphasizes two main approaches to PCA: eigen-decomposition (ED) of the covariance matrix and singular value decomposition (SVD) of the data matrix. Both methods yield equivalent results with the help of eigenvector. However they has a common weakness that they rely on the assumption that principal components are orthogonal which limits the two techniques to extract non-vertical principal component, but Independent Component Analysis can handle this problem efficiently. "about the experiment of MNIST..."
Keywords: PCA, covariance, eigen-decomposition, SVD, variance.

Contents

1 Foundation & Settings

A matrix in Principal Component Analysis can serves as a linear operator that performs a change of basis: it re-expresses the recorded data X as Y via $PX = Y$. Geometrically,it represent the transformation of a rotation and stretch, and the rows of P serve as a new set of basis vectors for expressing the columns of X. In the tutorial's spring–mass example, three cameras yield six-dimensional measurements that are both noisy and redundant, whereas the true motion is essentially one-dimensional along the physical x-axis. Principal Component Analysis (PCA) aims to find the most meaningful basis that filters noise and reveals structure by finding directions of largest variance (high signal-to-noise ratio) and by transforming the covariance into a diagonal form; and the orthogonal martrix which realizes this contains the principal components.

2 Assumptions & Limits of PCA

***A*ssumptions**

I. Linearity PCA assumes a *linear change of basis.* With centered data $X \in \mathbb{R}^{m\times n}$ and covariance $C_X = \frac{1}{n}XX^\top$, it searches an orthonormal transform:

$$Y = PX, \qquad P \in O(m), \quad P^\top P = I_m.$$

II. Variance-as-Structure *High SNR*

$$X \text{ centered}, \qquad C_X \equiv \tfrac{1}{n}XX^T \quad \text{(sample covariance)}.$$

$$Y = PX, \qquad C_Y \equiv \tfrac{1}{n}YY^T = P\,C_X\,P^T = D \quad \text{(decorrelated / diagonal)}.$$

$$D = \operatorname{diag}(\lambda_1, \dots, \lambda_m), \quad \lambda_1 \geq \cdots \geq \lambda_m, \quad (C_Y)_{ii} = \lambda_i \quad \text{(variance along row } p_i \text{ of } P).$$

$$C_X = EDE^T, \quad P \equiv E^T \quad \text{(rows of } P \text{ are principal components)}.$$

Interpretation: larger $\lambda_i \;\Rightarrow$ "interesting structure" (high SNR); smaller $\lambda_i \;\Rightarrow$ "noise".

III. The principal components are orthogonal

$$C_X = \tfrac{1}{n}XX^T, \qquad C_X = EDE^T, \qquad D = \operatorname{diag}(\lambda_1, \dots, \lambda_m), \quad \lambda_1 \geq \cdots \geq \lambda_m.$$

$$P \equiv E^T, \qquad Y = PX, \qquad C_Y = PC_XP^T = D,$$
$$E^TE = I \iff PP^T = I, \qquad p_i^Tp_j = 0 \ (i \neq j), \quad p_i^Tp_i = 1.$$

3 Covariance, Redundancy, and SNR

3.1 Covariance

Sample covariance. For centered data $X \in \mathbb{R}^{m\times n}$ (columns are samples),

$$C_X = \tfrac{1}{n}\, XX^T, \qquad (C_X)_{ij} = \tfrac{1}{n}\sum_{\ell=1}^{n} X_{i\ell}\, X_{j\ell}.$$

Deeper quantities. Diagonal entries $(C_X)_{ii}$ are variances; off–diagonal entries $(C_X)_{ij}$ are covariances. The magnitude $|(C_X)_{ij}|$ quantifies linear redundancy between channels i and j which reveals the overlapping information between variables.

3.2 Covariance, Redundancy, and Noise

Covariance decomposition. For centered data X, the sample covariance is

$$C_X = \tfrac{1}{n} XX^T.$$

Decompose it into signal and noise parts:

$$C_X = C_{\text{sig}} + C_{\text{noise}}, \qquad C_{\text{noise}} = \sigma^2 I$$

Redundancy encoded in covariance. Off–diagonal entries quantify linear redundancy between channels:

$$\text{redundancy}_{ij} \;\propto\; |(C_X)_{ij}| = |(C_{\text{sig}})_{ij} + (C_{\text{noise}})_{ij}| \quad (i \neq j).$$

In the direction of large varience, $(C_{\text{noise}})_{ij} = 0$ for $i \neq j$, so redundancy arises from C_{sig}.

Effect of change of basis. Let $Y = PX$ with P orthonormal chosen to diagonalize C_X:

$$C_Y = \tfrac{1}{n} YY^T = PC_XP^T = D, \qquad D = \text{diag}(\lambda_1, \ldots, \lambda_m).$$

Redundancy is removed in Y (off–diagonals $= 0$). The diagonal entries split signal and noise power along principal directions p_i (rows of P):

$$\lambda_i = p_i^T C_{\text{sig}}\, p_i \;+\; p_i^T C_{\text{noise}}\, p_i,$$

and in the direction of max varience,

$$p_i^T C_{\text{noise}}\, p_i = \sigma^2, \qquad D - \sigma^2 I = \text{diag}\left(p_1^T C_{\text{sig}} p_1, \ldots, p_m^T C_{\text{sig}} p_m\right).$$

4 SNR, Variance, and Redundancy

4.1 Definitions

Variance For any unit p, the projected variance is

$$\text{Var}(p^T x) = p^T C_X p.$$

For principal directions p_i with $C_X p_i = \lambda_i p_i$,

$$\text{Var}(p_i^T x) = \lambda_i.$$

Redundancy (pairwise). For two centered scalar matrix A, B,

$$\mathrm{Cov}(A,B) = \tfrac{1}{n}\sum_{\ell=1}^{n} a_\ell b_\ell,$$

and in matrix form $|(C_X)_{ij}|$ measures linear redundancy between row i and j.

Signal-to-Noise Ratio (SNR). With $C_X = C_{\mathrm{sig}} + \sigma^2 I$,

$$\mathrm{SNR}_i = \frac{\lambda_i - \sigma^2}{\sigma^2}, \qquad \lambda_i = p_i^T C_{\mathrm{sig}} p_i + \sigma^2.$$

4.2 Relations (Content Requirements)

Variance–SNR link. Larger λ_i implies larger SNR_i under a common noise floor σ^2; principal components are ranked by λ_i and equivalently by SNR_i.

Covariance–Redundancy link. Large $|(C_X)_{ij}|$ indicates strong redundancy; PCA finds P so that $C_Y = PC_XP^T = D$ becomes diagonal, eliminating pairwise redundancy in Y.

Combined PCA criterion. Let $C_X = EDE^T$ with $E^TE = I$ and $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_m)$. Choosing $P = E^T$ yields $C_Y = D$ (decorrelation) and orders coordinates by λ_i (high–SNR directions first). For k–dimensional retention with $E_k = [p_1, \ldots, p_k]$,

$$E_k^T C_X E_k = \mathrm{diag}(\lambda_1, \ldots, \lambda_k), \qquad \sum_{i=1}^{k} \lambda_i \text{ (retained power)}, \quad \sum_{i=k+1}^{m} \lambda_i \text{ (discarded power)}.$$

5 (PCA as a Change of Basis)

Change-of-Basis Formulation.

$$\mathbf{y} = \mathbf{P}\mathbf{x}, \qquad \mathbf{C}_Y = \mathbf{P}\mathbf{C}_X\mathbf{P}^\top \tag{1}$$

Projection and Re-expression.

Relation to Covariance.

6 (Connections: PCA–ED–SVD)

ED Route (Eigendecomposition).

$$\mathbf{C}_X = \mathbf{E}\mathbf{D}\mathbf{E}^\top, \quad \text{PCs} \equiv \text{columns of } \mathbf{E}, \text{ variance} \equiv \mathrm{diag}(\mathbf{D}) \tag{2}$$

SVD Route (Singular Value Decomposition).

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top, \quad \text{PCs} \equiv \text{columns of } \mathbf{V}, \text{ scores} \equiv \mathbf{\Sigma}\mathbf{V}^\top \tag{3}$$

7 (Objective & Why DR Works)

Reconstruction-Error Minimization (MSE).

$$\min_{\mathrm{rank}(\hat{\mathbf{X}})\le k} \left\|\mathbf{X} - \hat{\mathbf{X}}\right\|_F^2 \iff \text{take the top } k \text{ principal components with largest variance.} \tag{4}$$

A Mathmatical Notations

Table 1: Alphabetical summary of mathematical notations used in the PCA tutorial

Notation	Definition	Corresponds to
A, B	Two general matrix used to define or explain other definitions below	tool matrix
a_i, b_i	i-th samples of A and B	scalar observations in demonstrating example
$C_X = \frac{1}{n}XX^\top$	Covariance matrix of X	reveals the redundancy of dataset X
$C_Y = \frac{1}{n}YY^\top$	Covariance of $Y = PX$ under a new basis	covariance matrix after change of basis
D	Diagonal matrix of eigenvalues in eigendecomposition	variances along principal components
E	Matrix whose columns are eigenvectors of C_X	principal directions
I	Identity matrix	orthonormal basis in $\mathbb{R}^m$
k	Target dimension for reduction($Xa = kb$in the explaination of SVD)	
m	Number of features (measurement types)	dimensions of dataset
n	Number of samples (trials)	scale of training set
$P = [p_1^\top \cdots p_m^\top]$	rotation and stretch to transforms$X$into$Y$	projection matrix
p_i	i-th principal component (row of P)	principal axis
r	Rank of X (or $X^\top X$)	intrinsic dimensionality
$\mathrm{SNR} = \sigma^2_{\text{signal}}/\sigma^2_{\text{noise}}$	Signal-to-noise ratio	measurement quality
$\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$	Diagonal matrix of singular values	covarience in new basis
σ_i	i-th singular value of X; $\lambda_i = \sigma_i^2/n$ (if $C_X = \frac{1}{n}XX^\top$)	scale of mode i
σ^2	Variance of a scalar variable/sequence	spread/energy
U	Left singular vectors of X	orthonormal basis of column space
V	Right singular vectors of X	orthonormal basis of row space
$\hat{u}_i$	i-th left singular vector; $\hat{u}_i = \frac{1}{\sigma_i}X\hat{v}_i$	output direction of mode i
$\hat{v}_i$	i-th eigenvector of $X^\top X$	input direction of mode i
$X \in \mathbb{R}^{m\times n}$	Data matrix	stacked measurements dataset
$x^{(j)}$	j-th sample vector (a column of X)	per-sample measurement
$Y = PX$	Data expressed in PCA coordinates	projections onto PCs
$Z = U^\top X$	Coordinates in the left-singular basis	transformed data
λ_i	i-th eigenvalue of C_X	variance along the i-th PC
δ_{ij}	element $U(= 1$ if $i = j$, else $0)$	orthogonality indicator
$\Vert \cdot \Vert$	Euclidean norm	vector length
$(\cdot)^\top$	Transpose	matrix transpose
$\cdot$	Dot product	inner product

B Derivations (Optional)

Sketch why eigenvectors of C_X diagonalize C_Y; Eckart–Young link.

C Reproducibility (Optional)

OS, Python/Matlab version, libs, seed, commands to regenerate figures.

References

[1] J. Shlens, *A Tutorial on Principal Component Analysis*, 2014.