Reading Report:
A Tutorial on Principal Component Analysis

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

September 19, 2025

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

Abstract

One paragraph (150–200 words): objective of PCA, main insights from the tutorial, and your MNIST takeaway. Mention the ED/SVD equivalence briefly. 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 Introduction

Briefly introduce PCA as a linear, non-parametric technique for variance-maximizing, decorrelating representation; follow the change-of-basis view in Shlens (2014).

2 Core Concepts and Questions (Q0–Q13)

Q0. Essence of a Matrix

Linear map / basis-change views; minimal example.

Q1. What Problems Does PCA Address?

Redundancy, meaningful axes of variation.

Q2. Assumptions and Limits

Linearity; variance→structure; orthogonality; failure cases.

Q3. Basis in Linear Algebra

Basis, orthonormality, projection.

Q4. Covariance and Redundancy

Cov, off-diagonal dependence, diagonal dominance.

Q5. SNR, Variance, Redundancy

Var, SNR; why diagonalize C_X.

Q6. Principal Component Meaning

Variance-maximizing directions under orthogonality.

Q7. PCA as Basis Transformation

$Y = PX$; geometric interpretation.

Q8. Re-expressing Inputs

Centering; dimensionality reduction by top-k rows of P.

Q9. PCA and Covariance

$C_Y = PC_X P^\top$; diagonalization.

Q10. PCA, ED, SVD

ED of C_X vs SVD route; equivalence and numerics.

Q11. Intuition: ED and SVD

Rotations + stretch interpretations.

Q12. Objective Function

$$\max_{\|w\|=1} w^\top C_X w \quad \Longleftrightarrow \quad \min_{\text{rank}(\tilde{X})\le k} \|X - \tilde{X}\|_F^2 \quad \text{(Eckart–Young)} \tag{1}$$

Q13. Why Dimension Reduction Works

Order by eigenvalues; preserve most variance.

3 Experiments on MNIST (Q14)

3.1 Dataset and Setup

Use 2,000 train + 2,000 test images. Preprocess: flatten, center, optionally scale.

3.2 Method

PCA via ED or SVD. Report explained variance ratio for top-2/top-50 as sanity check.

3.3 Results: 2D Visualization

Figure 1: PCA to 2D on MNIST. Distinct colors/markers per digit class.

3.4 Analysis

Comment on class overlap; limits of linear PCA; when kernel PCA/ICA might help.

4 Discussion and Conclusion

Summarize SNR/covariance $\rightarrow$ diagonalization $\rightarrow$ ED/SVD routes; practical gains and limits.

Compliance Checklist

- Self-contained definitions (matrix, basis, projection, variance, covariance, SNR, redundancy, PC).
- High logic, concise writing.
- Q0–Q13 covered; Q14 plots and analysis present.
- Notation table completed in **??**.

A Mathmatical Notations)

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.

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
$\text{SNR} = \sigma^2_{\text{signal}}/\sigma^2_{\text{noise}}$	Signal-to-noise ratio	measurement quality
$\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_r)$	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 (columns)	orthonormal basis of column space
V	Right singular vectors of X (columns)	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 $X^\top X$ (right singular vector)	input direction of mode i
$X \in \mathbb{R}^{m \times n}$	Data matrix (columns are centered samples)	stacked measurements
$x^{(j)}$	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}	Kronecker delta ($= 1$ if $i = j$, else 0)	orthogonality indicator
$\|\cdot\|$	Euclidean norm	vector length
$(\cdot)^\top$	Transpose	matrix/vector transpose
$\cdot$	Dot product	inner product