Principal Component Analysis

- Known by different names: Karhunen-Loeve transform, Hotelling transform etc
- Given a dataset, find a projection that maximizes the variance and that gives the best reconstruction

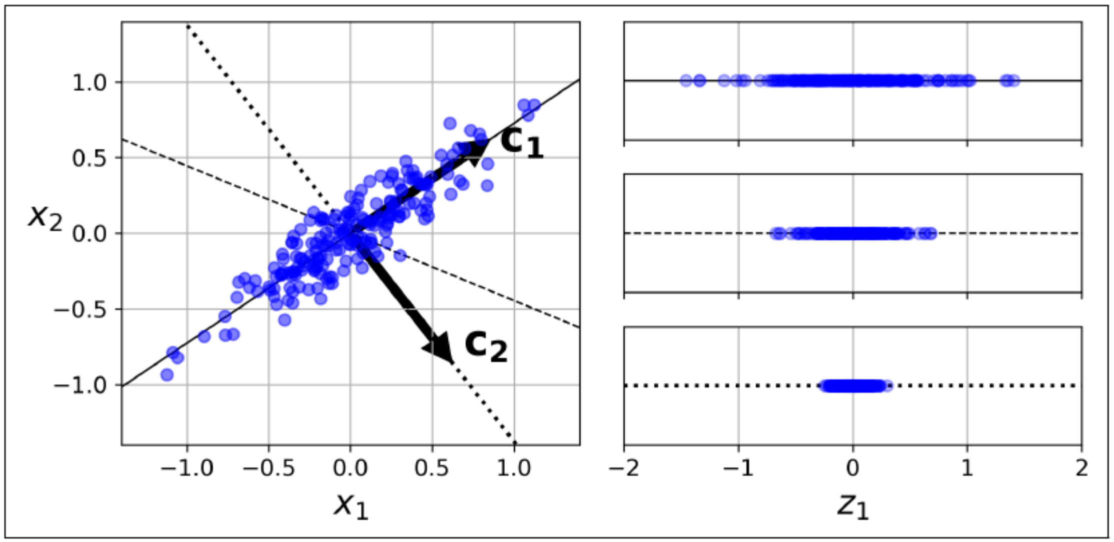


Figure 8-7. Selecting the subspace onto which to project

- PCA decorrelates data, i.e., identifies the subspaces where the data is found

- Let $\mathbf{x}$ be D-dimensional vector written as a linear combination of orthonormal basis vector

  $$\mathbf{x} = \sum_{i=1}^{D} a_i \boldsymbol{\phi}_i$$

  where $\boldsymbol{\phi}_i^T \boldsymbol{\phi}_j = \begin{cases} 1 & if\ i = j \\ 0 & otherwise \end{cases}$

- Now represent using only C of the basis vector where $C < D$

  $$\hat{\mathbf{x}}_C = \sum_{i=1}^{C} a_i \boldsymbol{\phi}_i + \sum_{i=C+1}^{D} b_i \boldsymbol{\phi}_i$$

- Difference $\Delta\mathbf{x} = \mathbf{x} - \hat{\mathbf{x}}_c = \sum_{i=C+1}^{D} (a_i - b_i)\boldsymbol{\phi}_i$

- Measure this difference as error by the residual sum of squares. The goal is to find the basis vectors and their coefficients that minimize the overall mean squared error

  $$\varepsilon^2 = E\left[|\Delta x|^2\right] = \sum_{i=C+1}^{D} E[(a_i - b_i)^2]$$

- Resorting to eigen-decomposition, $\varepsilon^2 = \sum_{i=C+1}^{D} \lambda_i$
  - To minimize ε^2 requires small eigenvalues
  - This RSS represents the error due to the discarded dimensions
- Thus to represent $\mathbf{x}$ with minimum RSS, choose eigenvectors corresponding to the largest eigenvalues, i.e., maximal spread

Steps with Eigen-decomposition

- Zero out the mean of the data

$$\boldsymbol{\mu} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i$$

$$XM = X - M$$

  - M : mean matrix

- Compute the covariance matrix

$$\Sigma = \frac{1}{N-1}\left(XM^T.XM\right)$$

- Perform eigen-decomposition on Σ

$$\Phi = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \dots, \boldsymbol{\phi}_D] \text{ and}$$

$$\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_D) \text{ where } \lambda_1 > \lambda_2 > \ \dots > \lambda_D$$

- Project XM onto Φ

$$Y = XM.\Phi$$

Reconstruction

- For reconstruction → $\hat{X} = Y\,\Phi^T + M$
- Using all principal components → $\hat{X} = X$
  - Total variance is preserved
- Using the largest C principal components $(C < D)$ → $\hat{X} \approx X$
  - $\hat{X} = \hat{Y}\hat{\Phi}^T + M$
  - Dimensionality reduction
  - Some information is lost

- *Normalized* reconstruction error:

$$error = \frac{\left\|X - \hat{X}\right\|_F}{\|X\|_F}$$

  - Frobenious norm is the generalized Eucledian-norm

Steps with SVD

- Zero out the mean of the data

  $$\boldsymbol{\mu} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i$$

  $$XM = X - M$$

  - M : mean matrix
- Compute SVD

  $$U, S, V^T = \text{svd}\ (XM)$$

- where
  - $S = \text{singular values}$

    $$= diag\ (\sigma_1, \sigma_2, \dots, \sigma_D)\ where\ \sigma_1 \geq \sigma_2 \geq\ \dots\ \geq \sigma_D$$

  - $V = right\ singular\ vectors = [\boldsymbol{V}_1, \boldsymbol{V}_2, \dots, \boldsymbol{V}_D]$
- Project XM onto V

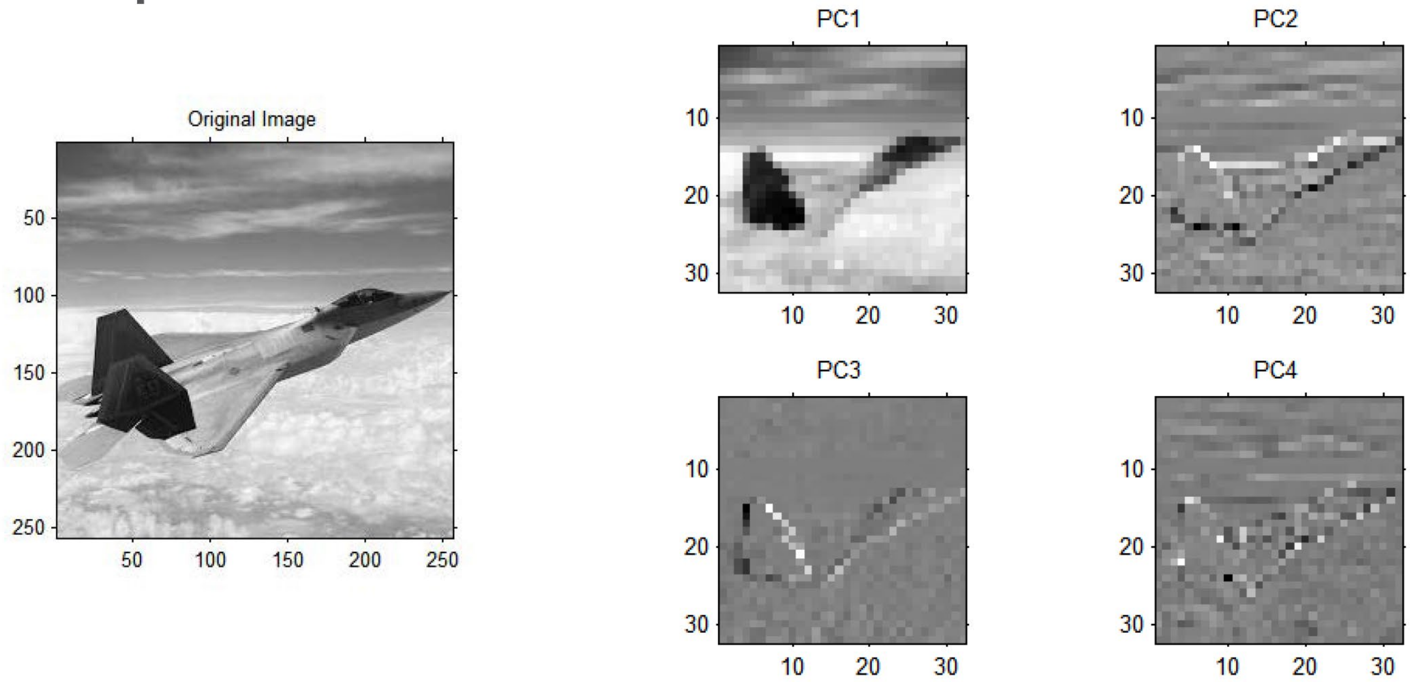
  $$Y = XM.V$$

Reconstruction

- Same as before, just replace Φ with V
- For reconstruction → $\hat{X} = Y\, V^T + M$

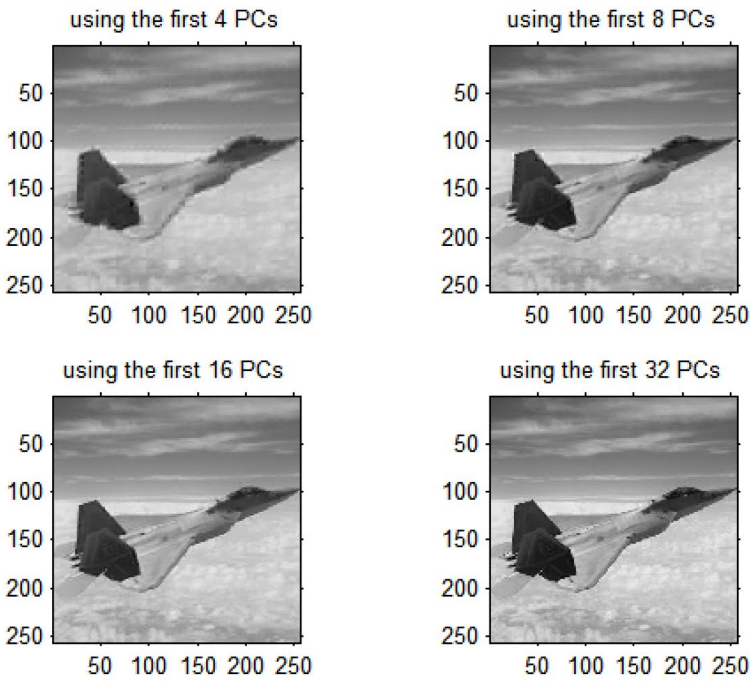
- How many singular vectors to keep ?
  - Rule of thumb: ~90% of variance (energy) = $\frac{\sum_i \sigma_i}{Total\ \sigma' s}$

Example



Recovered images

# PCs	Total variance preserved
4	0.9595
8	0.9785
16	0.9897
32	0.9973



Pan-sharpening

- Technique to generate high-resolution color image from low-resolution color image and high-resolution gray-scale image
  - Replace the 1st PC of low-resolution color image with the high-resolution gray-scale image
  - Recovered image will be high-resolution color image