Lecture Notes VII – Principal Component Analysis

Marina Meilă
mmp@uwaterloo.ca

With Thanks to Pascal Poupart & Gautam Kamath
Cheriton School of Computer Science
University of Waterloo

March 19, 2026

Eigendecompositions of Variance

Examples

Eigendecompositions of Variance

- The data matrix $\mathbf{X} \in \mathbb{R}^{n \times D}$, $x^{1:n} \in \mathbb{R}^D$

1. The covariance matrix eigendecomposition

$$\Sigma = \frac{1}{n}\mathbf{X}^T\mathbf{X} = V\tilde{\Sigma}V^T \quad \in \mathbb{R}^{D\times D} \text{with } \tilde{\Sigma} = \text{diag}\{\tilde{\sigma}_{1:D}\} \tag{1}$$

   and $V = [\, v_1 \ \ldots v_D \,]$ an orthogonal $D \times D$ matrix with columns $v_{1:D}$.

2. The **Gram matrix** eigendecomposition

$$G = \mathbf{X}\mathbf{X}^T = U\tilde{\Sigma}U^T \quad \in \mathbb{R}^{n\times n} \text{with } \tilde{\Sigma} = \text{diag}\{\tilde{\sigma}_{1:D}\} \tag{2}$$

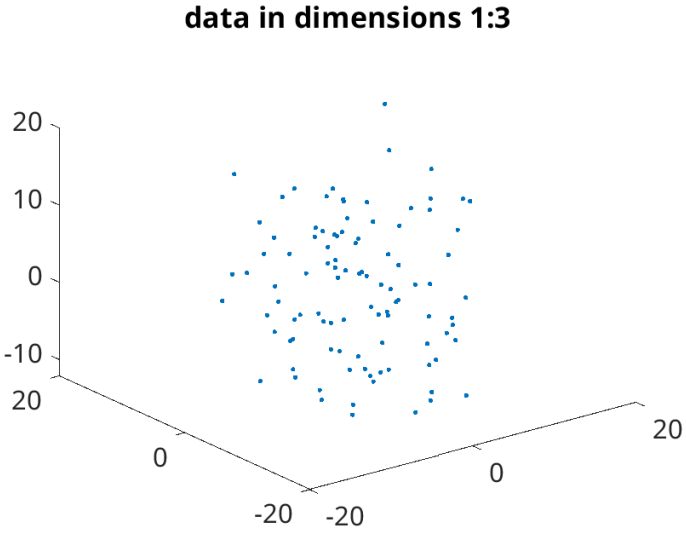
   and U an orthogonal $n \times D$ matrix with rows $u^{1:n}$.

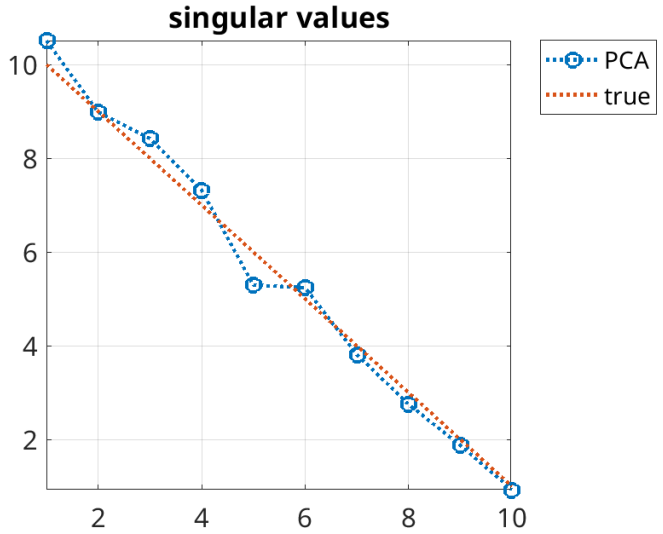
3. $\mathbf{X}$ Singular Value Decomposition and change of basis for $x^{1:n}$

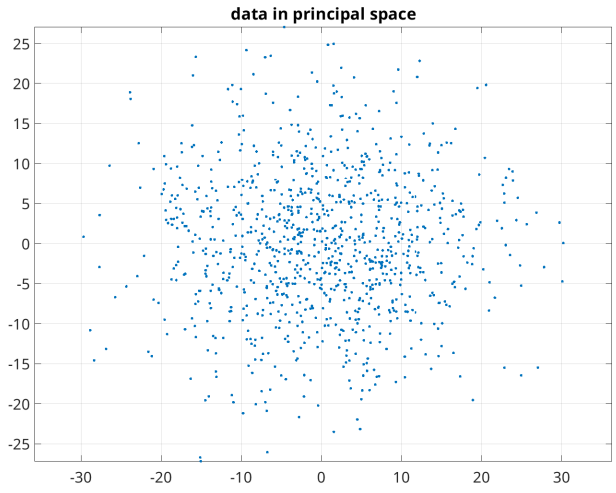
$$\mathbf{X} = U\tilde{\Sigma}V^T \quad \text{(SVD)} \quad \text{and } x^i = \sum_{j=1}^{D} u_j^i v_j \quad \text{with } u_j^i = \tilde{\sigma}_j u_{ij} \tag{3}$$

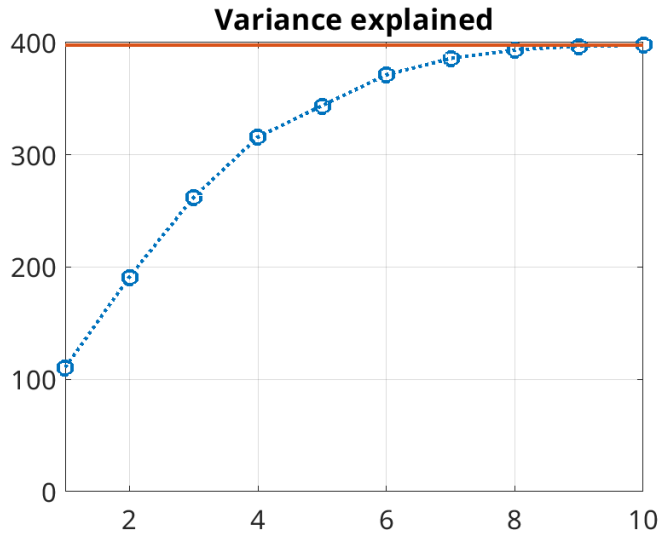
4. Dimension reduction – ignore the components $d+1, \ldots D$

Example – Gaussian data

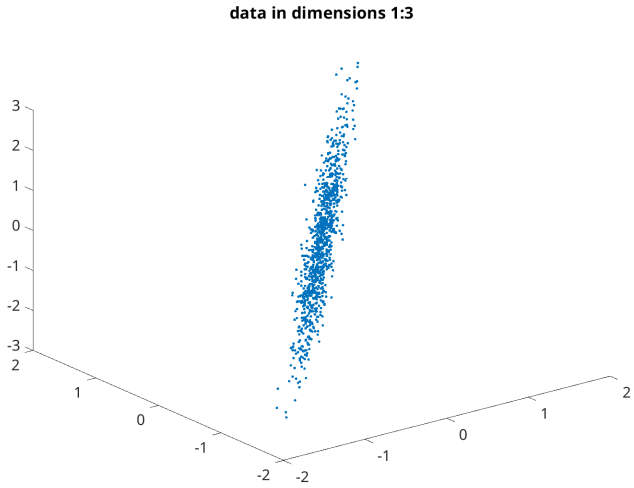


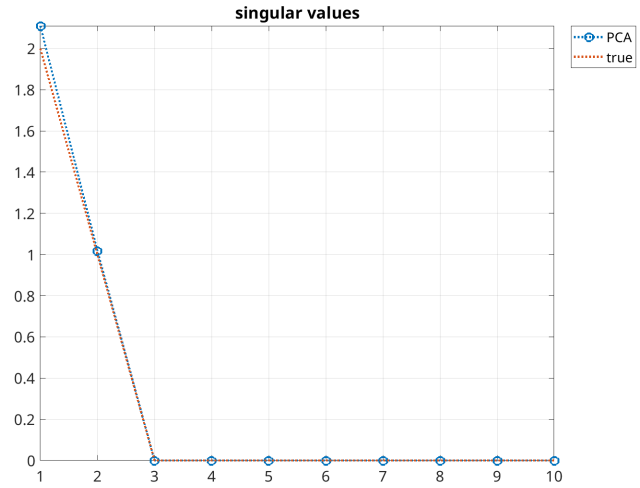


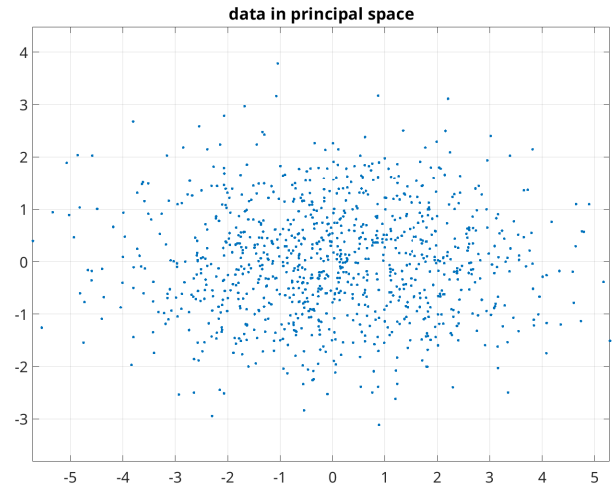


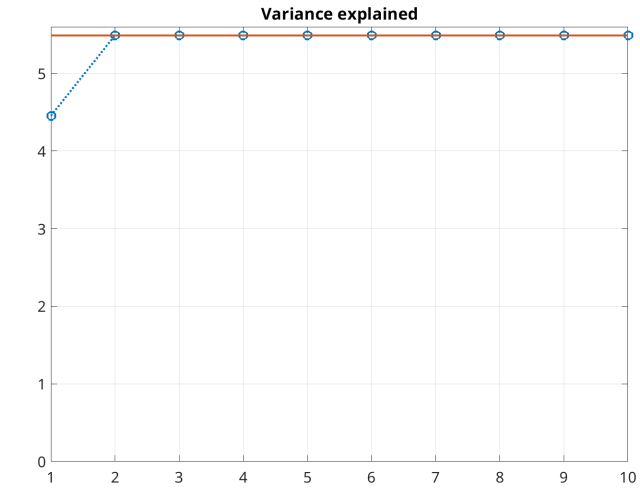


Example – Gaussian data 2D

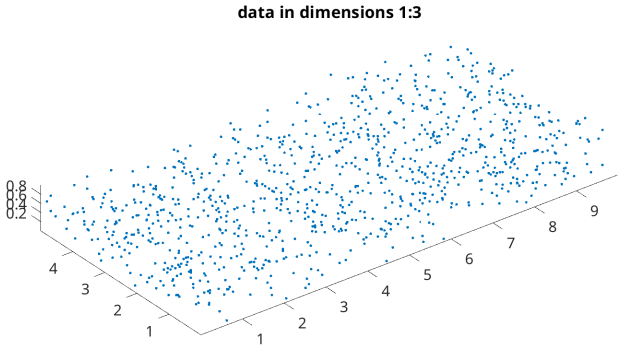


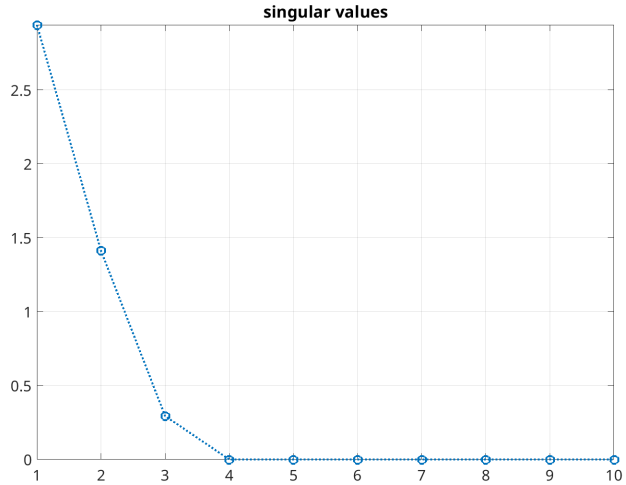


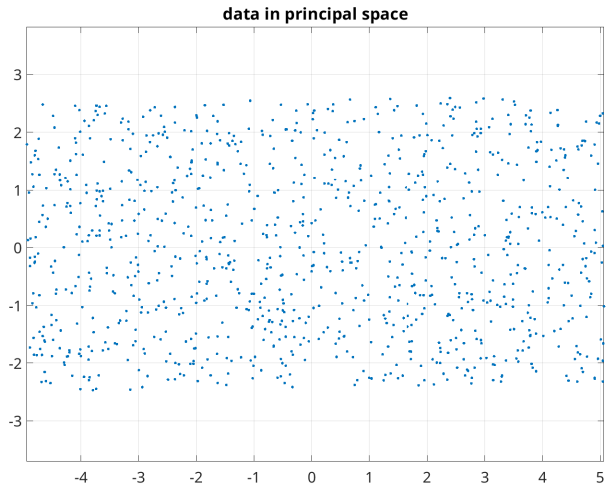


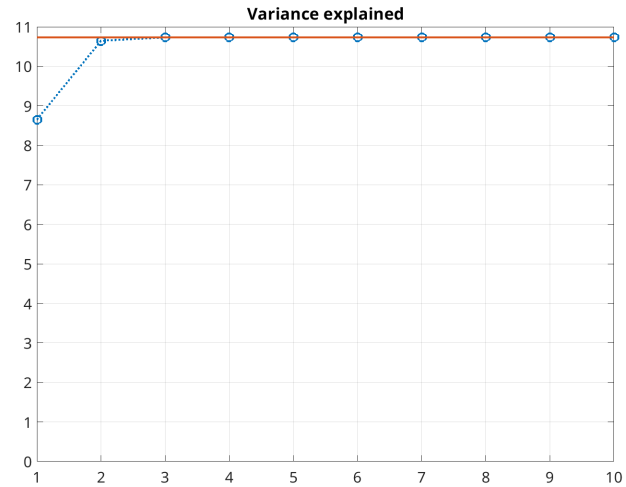


Example – Brick

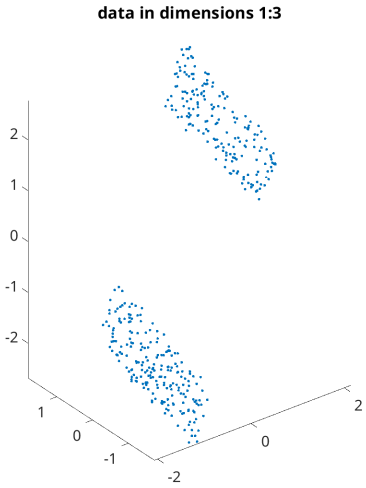


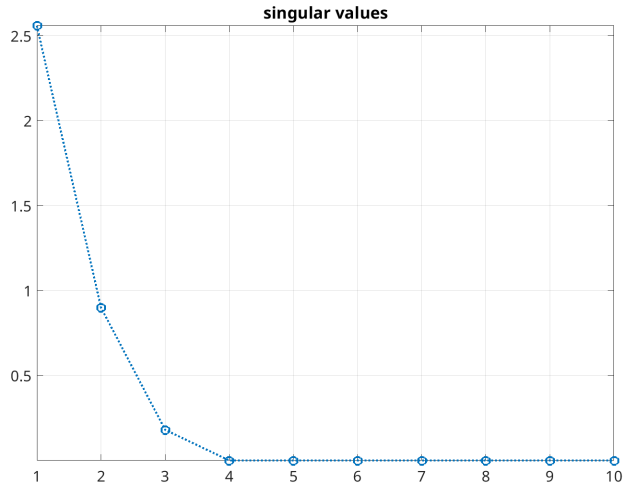


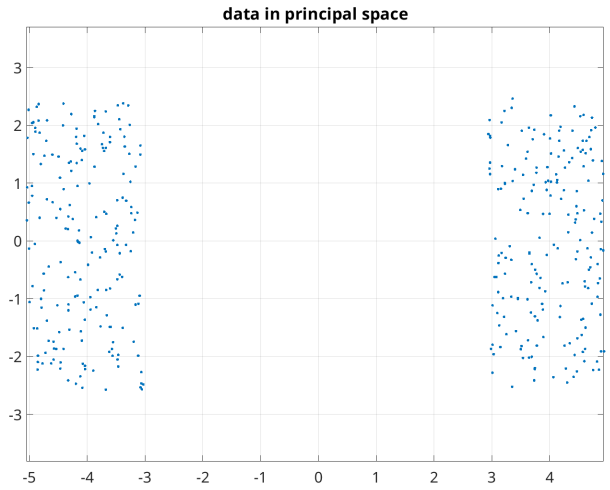


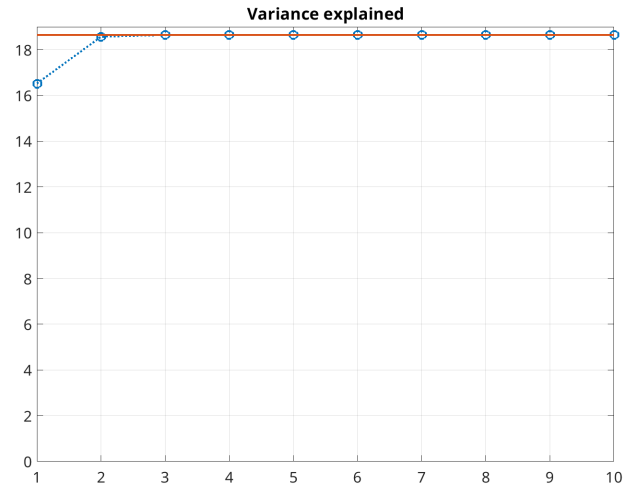


Example – clusters









PCA Summary

- Reduces data dimensiont from D to d
- Linear operation (projection)
- "Optimal" linear method to reduce dimension
- Can discover if data is low-dimensional
- For clustering – recommended pre-processing: PCA in $K-1$ dimensions
- Limitation: fails to discover non-linear low dimensional structure