

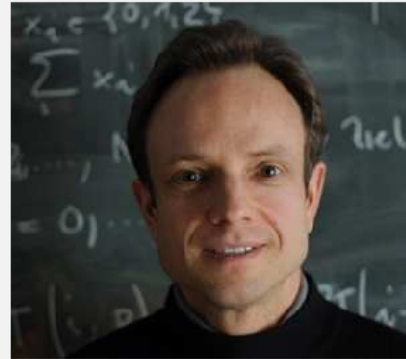
## Course Video

## Section 3.4.4.3: Principal Component Data Analysis (PCA)

Prof. R. Hiptmair, SAM, ETH Zurich

Date: September 29, 2020

(C) Seminar für Angewandte Mathematik, ETH Zürich



Dependencies. Section 3.4.1, Section 3.4.2

Duration: 54 minutes

## III. Direct Methods for Linear Least Squares Problems

## 3.4. Singular Value Decomposition (SVD)

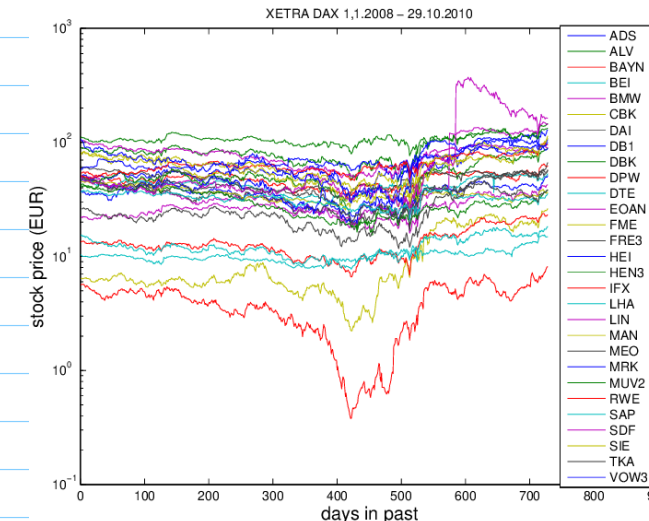
## 3.4.4. SVD-Based Optimization and Approximation

## 3.4.4.3. Principal Component Analysis (PCA)

Trend detection

Data classification

## Ex. 3.4.4.25 (Trend analysis)



Prices of  $n$  stocks for  $m$  days

$$a_k \in \mathbb{R}^m, k = 1, \dots, n$$

$$A = [a_1, \dots, a_n] \in \mathbb{R}^{m,n}$$

②

Task: Identify dominant/<sup>independent</sup> trends

LA phrasing: Find <sup>orthonormal</sup> vectors  $\underline{u}_1, \dots, \underline{u}_p \in \mathbb{R}^m$ ,  
 $p \ll m, n$  such that

$$\underline{a}_k \in \text{Span}\{\underline{u}_1, \dots, \underline{u}_p\}$$

not realistic, due to "random perturbation"

Realistic goal:

Find <sup>orthonormal</sup> vectors  $\underline{u}_1, \dots, \underline{u}_p \in \mathbb{R}^m$ ,  
 $p \ll m, n$  such that

$$\underline{a}_k \in \text{Span}\{\underline{u}_1, \dots, \underline{u}_p\} + \text{"small remainder"}$$

$$\text{SVD: } A = \sum_{l=1}^p \sigma_l \underline{u}_l \underline{v}_l^T + \sum_{l=p+1}^{\min\{n, m\}} \sigma_l \underline{u}_l \underline{v}_l^T$$

$$[A = U \Sigma V^T: \underline{u}_l := (U)_{:,l}, \underline{v}_l := (V)_{:,l}]$$

can be achieved, if SV decay fast, because

$$\left\| \sum_{l=p+1}^{\min\{n, m\}} \sigma_l \underline{u}_l \underline{v}_l^T \right\|_2 = \sigma_{p+1}$$

Trends  $\leftrightarrow$   $p$  leftmost columns of  $U$

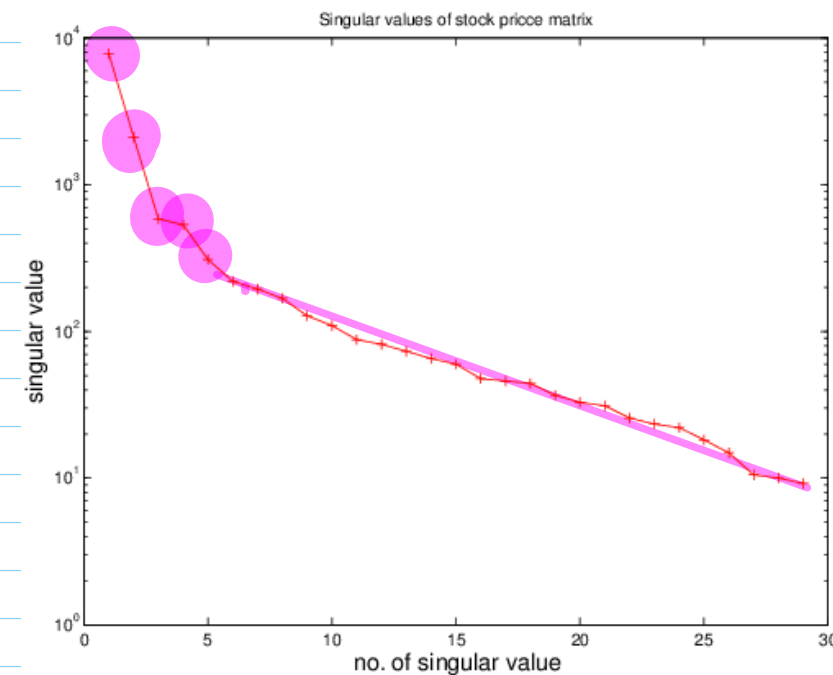
- $\sigma_l \leftrightarrow$  importance of  $l$ -th trend

$$\begin{bmatrix} \text{A} \end{bmatrix} = \sigma_1 \begin{bmatrix} \underline{u}_1 \end{bmatrix} \begin{bmatrix} \underline{v}_1^T \end{bmatrix} + \sigma_2 \begin{bmatrix} \underline{u}_2 \end{bmatrix} \begin{bmatrix} \underline{v}_2^T \end{bmatrix} + \dots$$

↑  
k

$(\underline{v}_l)_k \leftrightarrow$  strength of the  $l$ -th trend in  $\underline{a}_k$

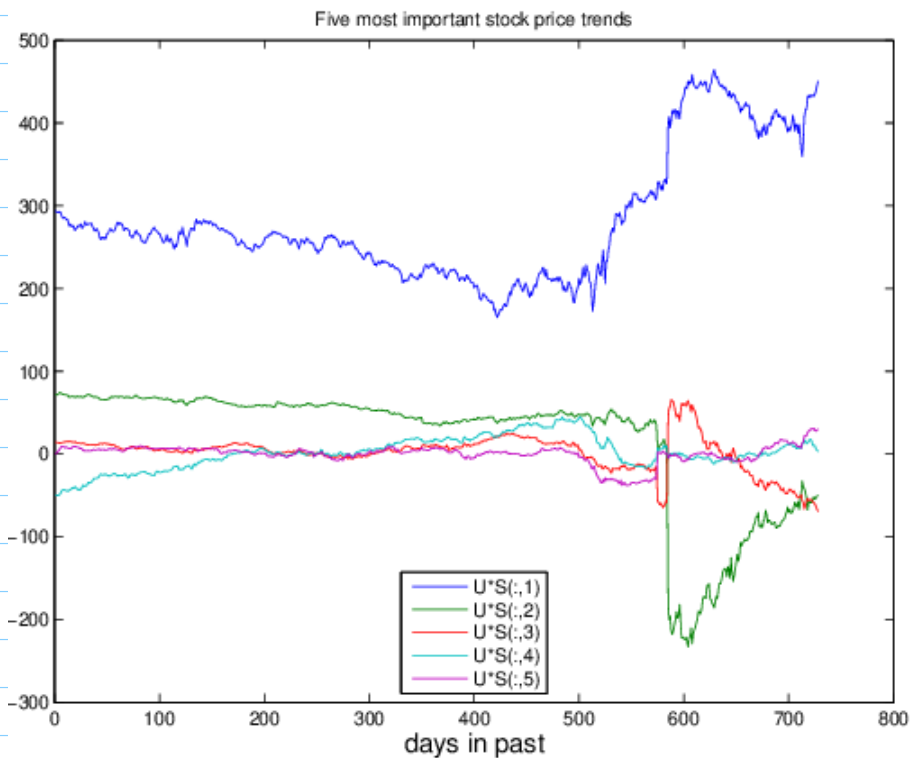
Ex. 3.4.4.25/3.4.4.30: PCA of stock prices



△ Lin-log plot of  $\sigma_l$ 's  
 $\sigma_l$  decay exponentially  
 $\sigma_l = O(q^l)$   
 for some  $0 < q < 1$

$p \approx 5$  trends capture  
 the time-dependence  
 of stock prices.

③



$$\Delta \sigma_j U_j$$

### § 3.4.4.33 Proper Orthogonal Decomposition (POD)

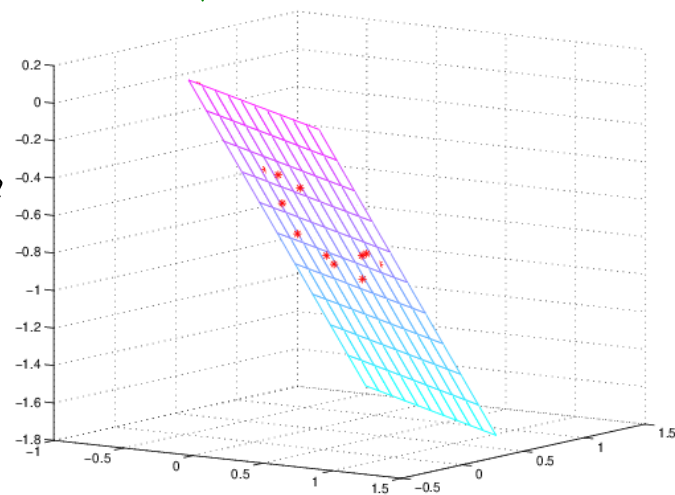
Ex. 3.4.4.32 : (Detect data points  $\hat{=}$  subspace)

Data points  $\bullet \leftrightarrow \mathbf{a}_i \in \mathbb{R}^3$  "almost" located on a plane  $\hat{=}$  2D subspace

Non-zero singular values of  $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_n] \in \mathbb{R}^{3,n}$

3.1378  
1.8092  
0.1792

} 2 dominant SVs



▷ SVD detects, if data approximately  $\in$  subspace

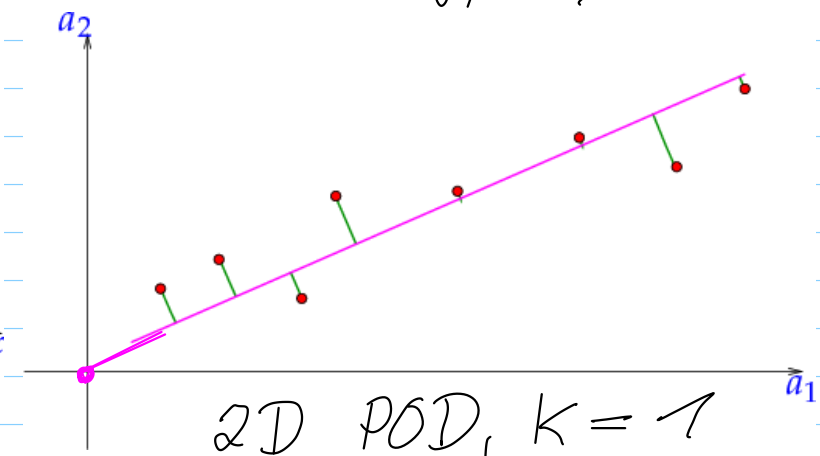
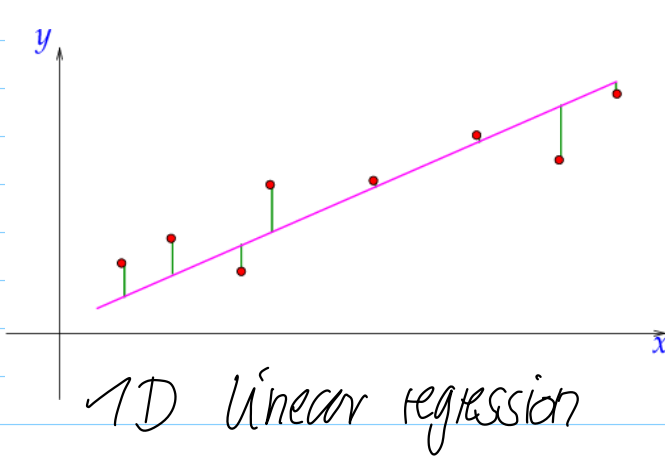
Problem of proper orthogonal decomposition (POD):

Given: Data points  $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{R}^m$ ,  $m, n \in \mathbb{N}$  (often  $n \gg m$ )

Sought: For  $k \leq \min\{m, n\}$ , find a subspace  $U_k \subset \mathbb{R}^m$  such that

$$U_k = \underset{W \subset \mathbb{R}^m, \dim W = k}{\operatorname{argmin}} \sum_{j=1}^n \inf_{\mathbf{w} \in W} \|\mathbf{a}_j - \mathbf{w}\|_2^2, \quad (3.4.4.34)$$

$\hat{=}$  provide ONB  
dist( $\mathbf{a}_j, W$ )



Subspace  $W \subset \mathbb{R}^m$ ,  $\dim W = k$ , ONB:  $W = \operatorname{Span}\{\mathbf{w}_1, \dots, \mathbf{w}_k\}$ ,

$$W = [\mathbf{w}_1, \dots, \mathbf{w}_k] \in \mathbb{R}^{m,k}, \quad k \leq m$$

$$\operatorname{dist}(\mathbf{x}, W) = \min_{\mathbf{z} \in W} \|\mathbf{x} - \mathbf{z}\|_2$$

$\hat{=}$  A linear least squares problem, minimizer  $\mathbf{y} = W^T \mathbf{x}$

(4)

$$\text{dist}(x, W) = \|x - WW^T x\|_2$$

$$\sum_{j=1}^n \inf_{w \in W} \|a_j - w\|_2^2 = \sum_{j=1}^n \|a_j - WW^T a_j\|_2^2 = \|A - WW^T A\|_F^2$$

$$\text{POD} \leftrightarrow \arg \min_{W \in \mathbb{R}^{m,k}, W^T W = I} \|A - \underbrace{WW^T A}_{\text{rank} \leq k}\|_F^2$$

**Definition 3.4.4.16. Frobenius norm**

The **Frobenius norm** of  $A \in \mathbb{K}^{m,n}$  is defined as

$$\|A\|_F^2 := \sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2.$$

**Theorem 3.4.4.18. Best low rank approximation** →

Let  $A = U\Sigma V^H$  be the SVD of  $A \in \mathbb{K}^{m,n}$  ( $\rightarrow$  Thm. 3.4.1.1). For  $1 \leq k \leq \text{rank}(A)$  set

$$A_k := U_k \Sigma_k V_k^H = \sum_{\ell=1}^k \sigma_\ell (U)_{:, \ell} (V)_{:, \ell}^H \quad \text{with} \quad \begin{aligned} U_k &:= \begin{bmatrix} (U)_{:,1}, \dots, (U)_{:,k} \end{bmatrix} \in \mathbb{K}^{m,k}, \\ V_k &:= \begin{bmatrix} (V)_{:,1}, \dots, (V)_{:,k} \end{bmatrix} \in \mathbb{K}^{n,k}, \\ \Sigma_k &:= \text{diag}(\sigma_1, \dots, \sigma_k) \in \mathbb{K}^{k,k}. \end{aligned}$$

Then, for  $\|\cdot\| = \|\cdot\|_F$  and  $\|\cdot\| = \|\cdot\|_2$ , holds true

$$\|A - A_k\| \leq \|A - F\| \quad \forall F \in \mathcal{R}_k(m, n),$$

that is,  $A_k$  is the rank- $k$  best approximation of  $A$  in the matrix norms  $\|\cdot\|_F$  and  $\|\cdot\|_2$ .

$$\|A - WW^T A\|_F \geq \|A - U_k \Sigma_k V_k^T\|_F \quad \forall W \in \mathbb{R}^{m,k}, \quad W^T W = I.$$

Can we find  $W \in \mathbb{R}^{m,k}$  with

$$(i) \quad W^T W = I$$

$$(ii) \quad WW^T A = U_k \Sigma_k V_k^T$$

Lucky:  $W = U_k$

$$WW^T A = U_k U_k^T U \Sigma V^T = U_k [I_k \quad 0] \Sigma V^T = U_k \Sigma_k V_k^T.$$

$U_k$  has orthonormal columns

**Theorem 3.4.4.37. Solution of POD problem**

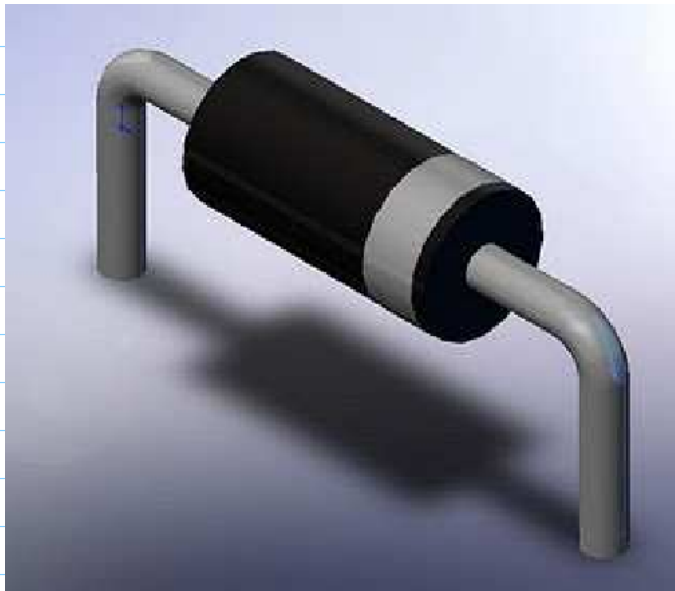
The subspace  $U_k$  spanned by the first  $k$  left singular vectors of  $A = [a_1, \dots, a_n] \in \mathbb{R}^{m,n}$  solves the **POD problem** (3.4.4.34):

$$U_k = \mathcal{R}((U)_{:,1:k}) \quad \text{with} \quad A = U \Sigma V^T \quad \text{the SVD of } A.$$

POD estimate:

$$\sum_{j=1}^n \inf_{w \in U_k} \|a_j - w\|_2^2 = \sum_{\ell=k+1}^p \sigma_\ell^2. \quad (3.4.4.38)$$

# Ex. 3.4.4.26 : Data classification



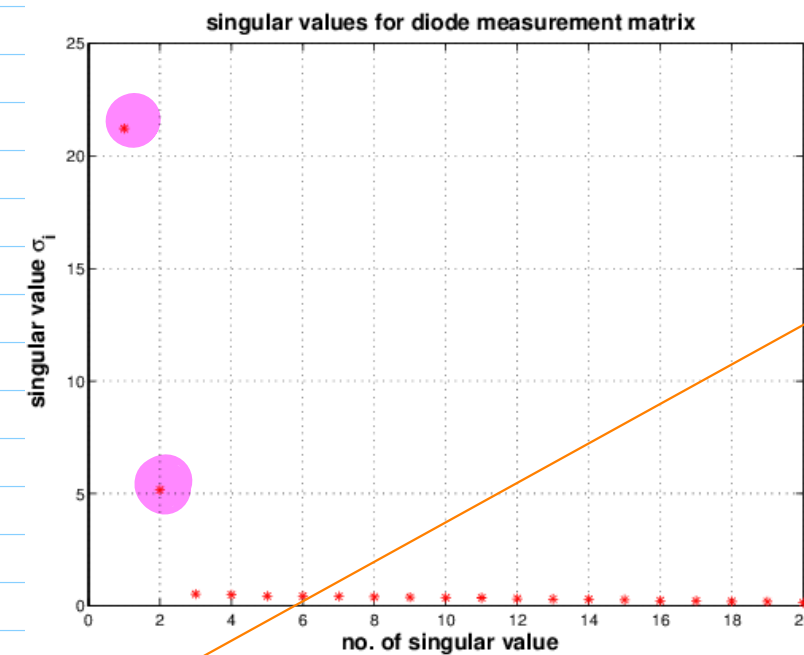
Given: measured  $U$ - $I$  characteristics of  $n$  diodes in a box  
(data points  $(U_j, I_j^{(k)})$ ,  $j = 1, \dots, m$ ,  $k = 1, \dots, n$ )

**Classification problem:** find out

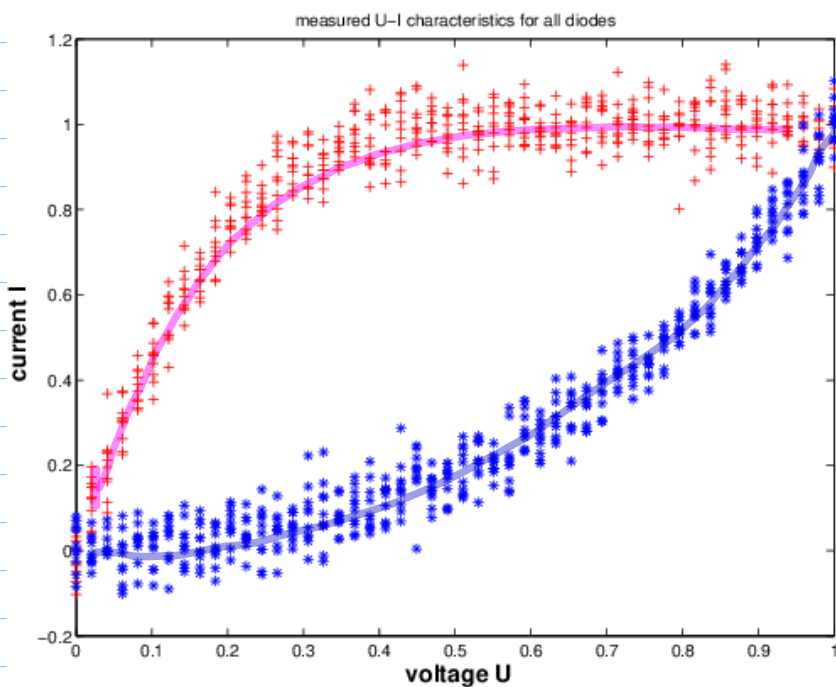
- how many different types of diodes in box,
- the  $U$ - $I$  characteristic of each type.



Measurement errors !  
Manufacturing tolerances !



two dominant SVs



Two types  
"two heads"

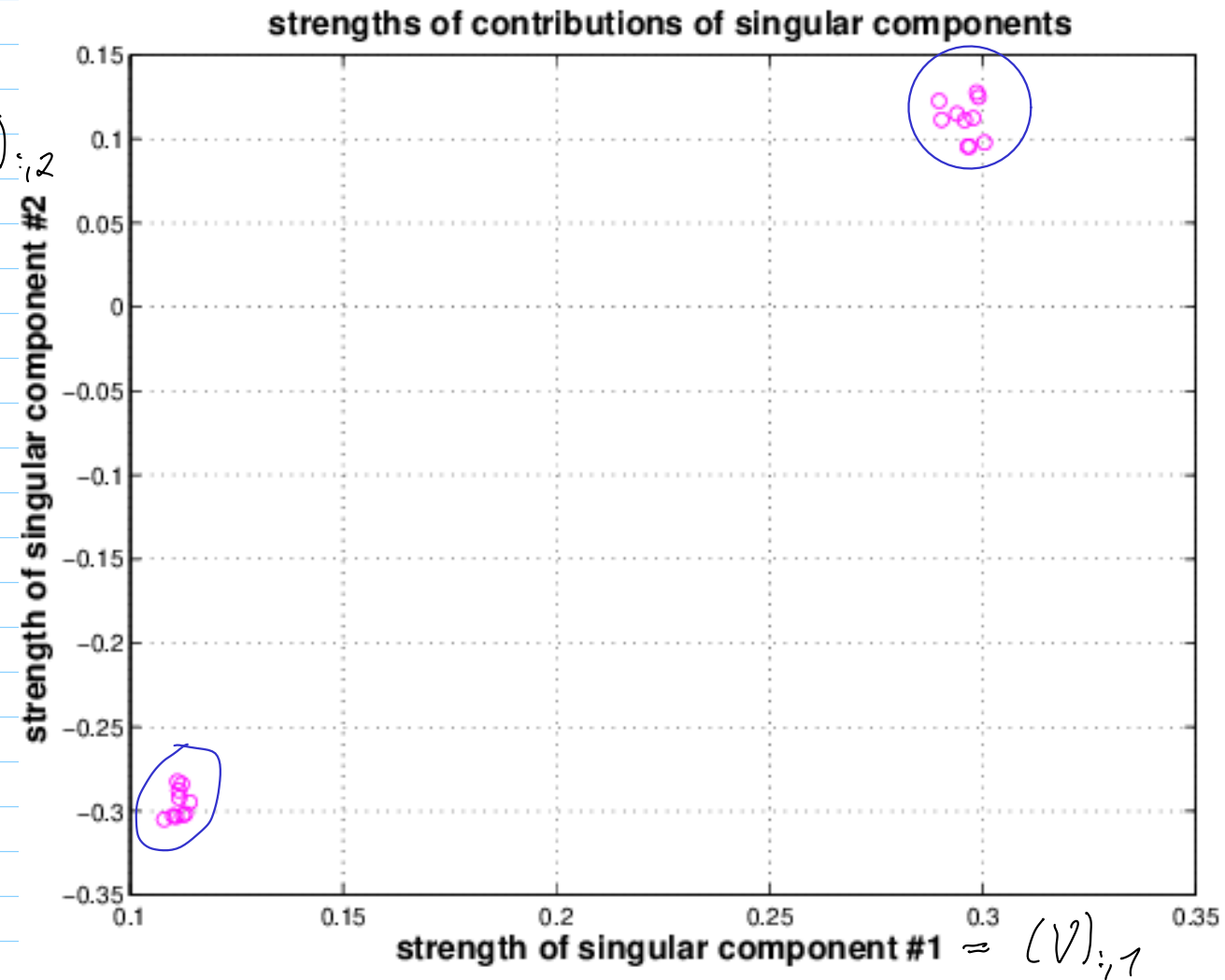
Data vectors  $\underline{a}_k \in \mathbb{R}^m$   
[m voltages]

$$A = [\underline{a}_1, \dots, \underline{a}_n] \in \mathbb{R}^{m,n}$$

$n \hat{=}$  no. of diodes

Idea : Classification by comparing components  
of  $(V)_{:,1}$  and  $(V)_{:,2}$   
↑ ↑  
strength of presence of  $(U)_{:,1} / (U)_{:,2}$   
in a particular measurement

6



Next : cluster analysis  $\rightarrow$  lecture document

Review questions :

Try to answer review questions 3.4.4.51

7