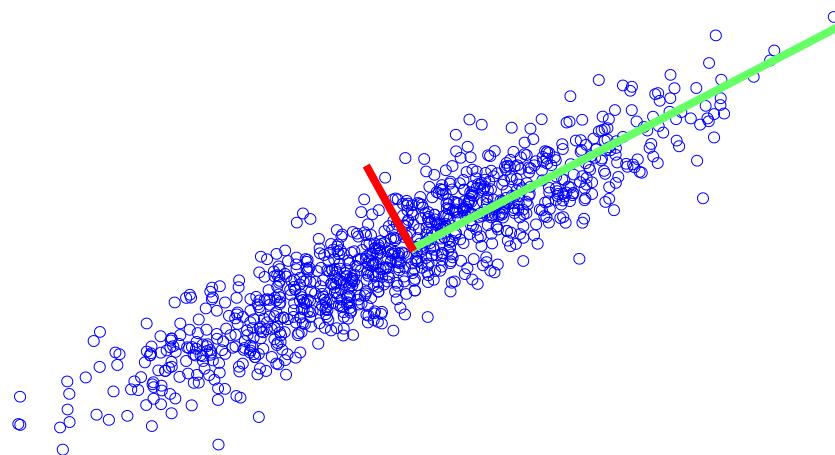


Analiza poglavitnih komponent

(Principal component analysis)

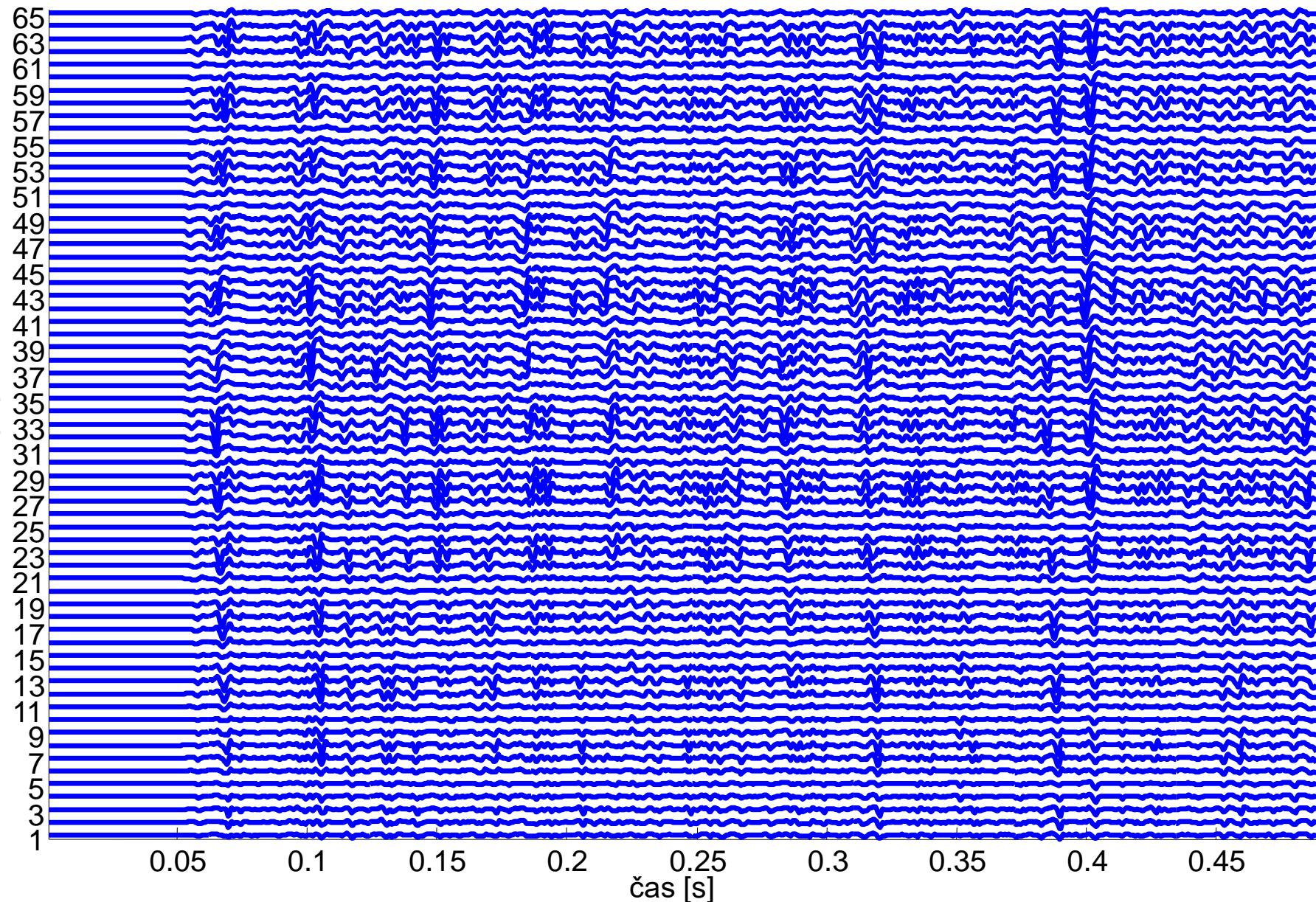


Motivacija

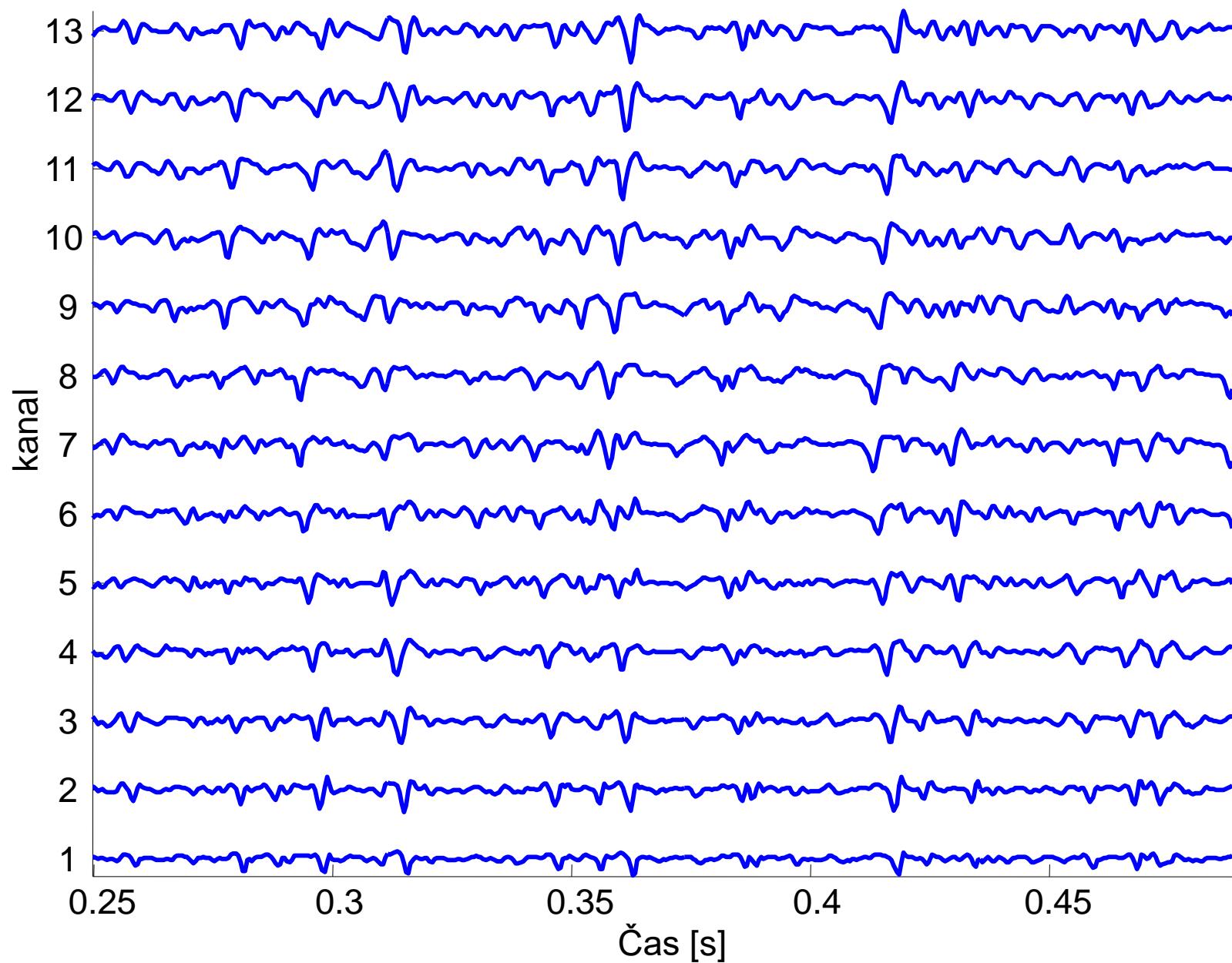
- Nedavni napredek na področju pridobivanja podatkov je povzročil ogromno povečanje števila in dolžine empiričnih signalov.
- Skrbno načrtovanje eksperimentov je nadomestila “off-line” analiza posnetih podatkov.
- Prekletstvo dimenzij:
 - velika računska zahtevnost
 - velika stopnja redundance
 - prisotnost šuma
- Pomembne izboljšave v obdelavi signalov lahko dosežemo tako, da najprej podatke preslikamo v prostor z manj dimenzijami.

Primer: 65-kanalni površinski EMG

(dvodimenzionalno polje elektrod 13x5)



Primer: površinski EMG (centralni stolpec elektrod)



Motivacija

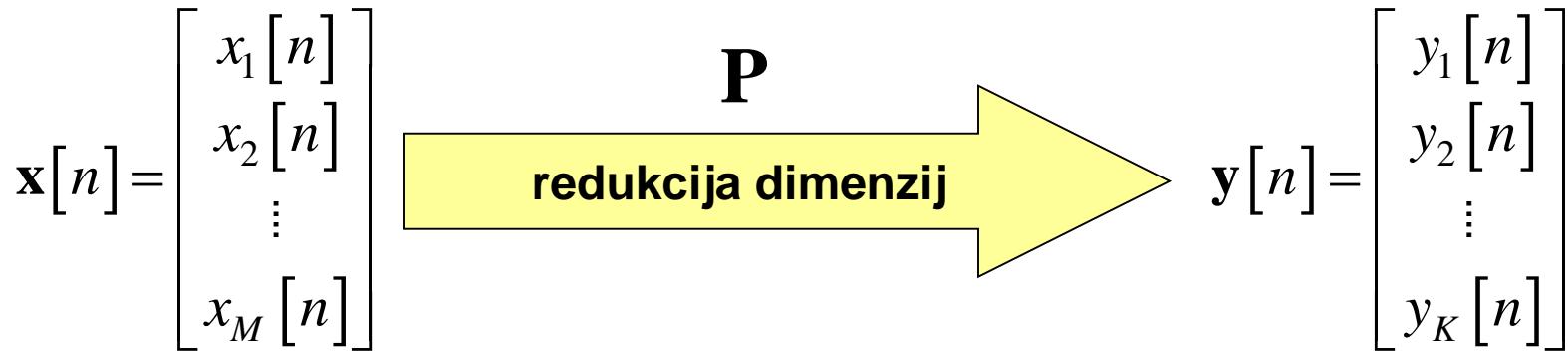
Dana naj bo množica meritev (naključni vektor),

$$\mathbf{x}[n] = \begin{bmatrix} x_1[n] \\ x_2[n] \\ \vdots \\ x_M[n] \end{bmatrix} \quad n = 1, \dots, N$$

n je vzorec signala (meritve) ali poskus (naključni vektorji)

- Kakšna je stopnja redundance v $\mathbf{x}[n]$?
- Ali lahko zmanjšamo dimenzionalnost podatkov na takšen način, da ohranimo čim več informacij o varianci (energiji) prvotnega nabora podatkov? Na kakšen način?

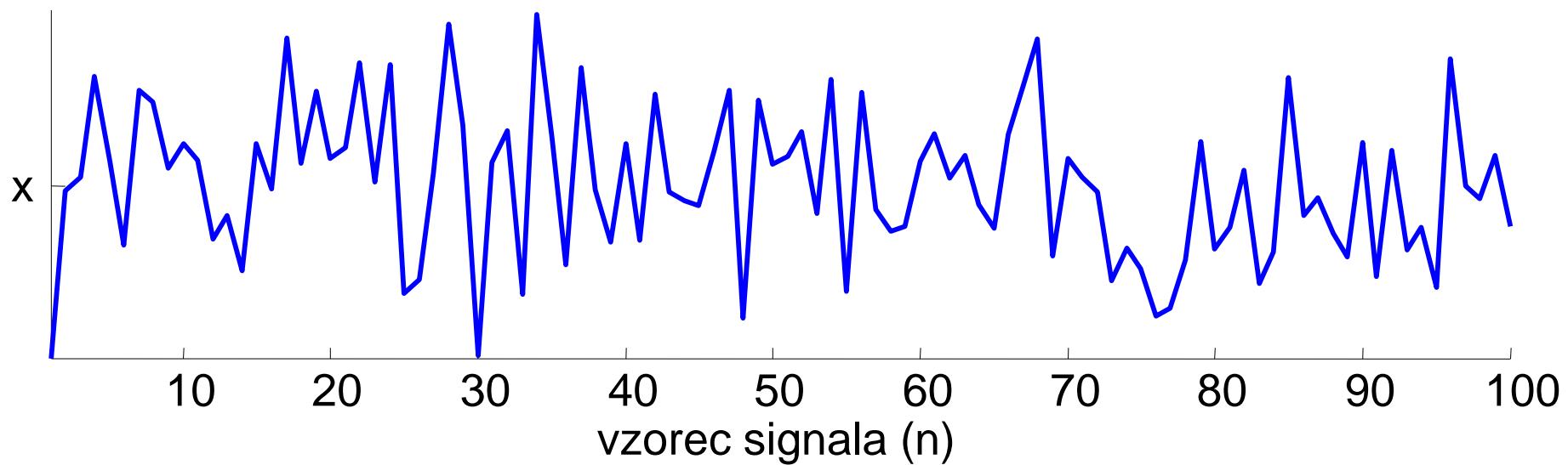
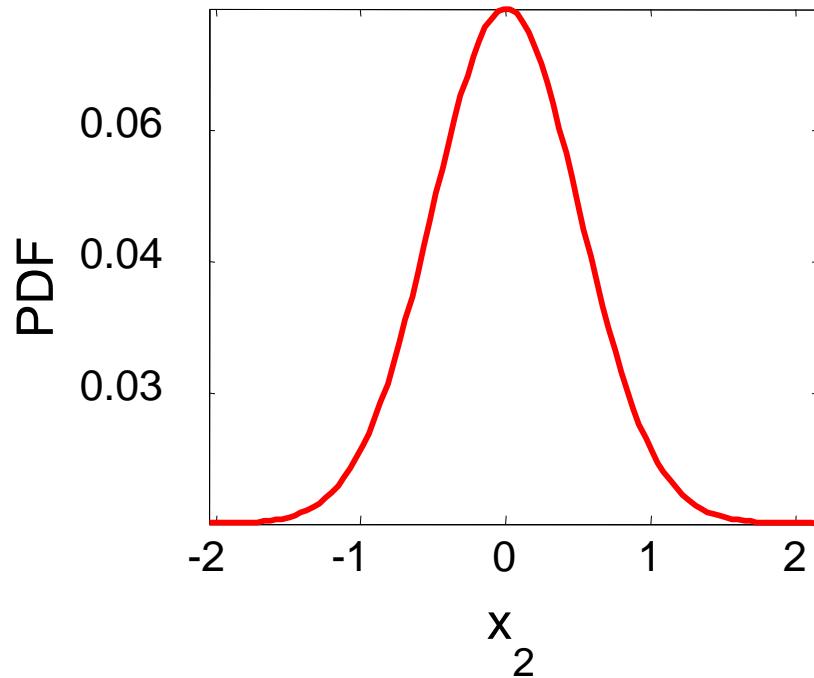
Motivacija



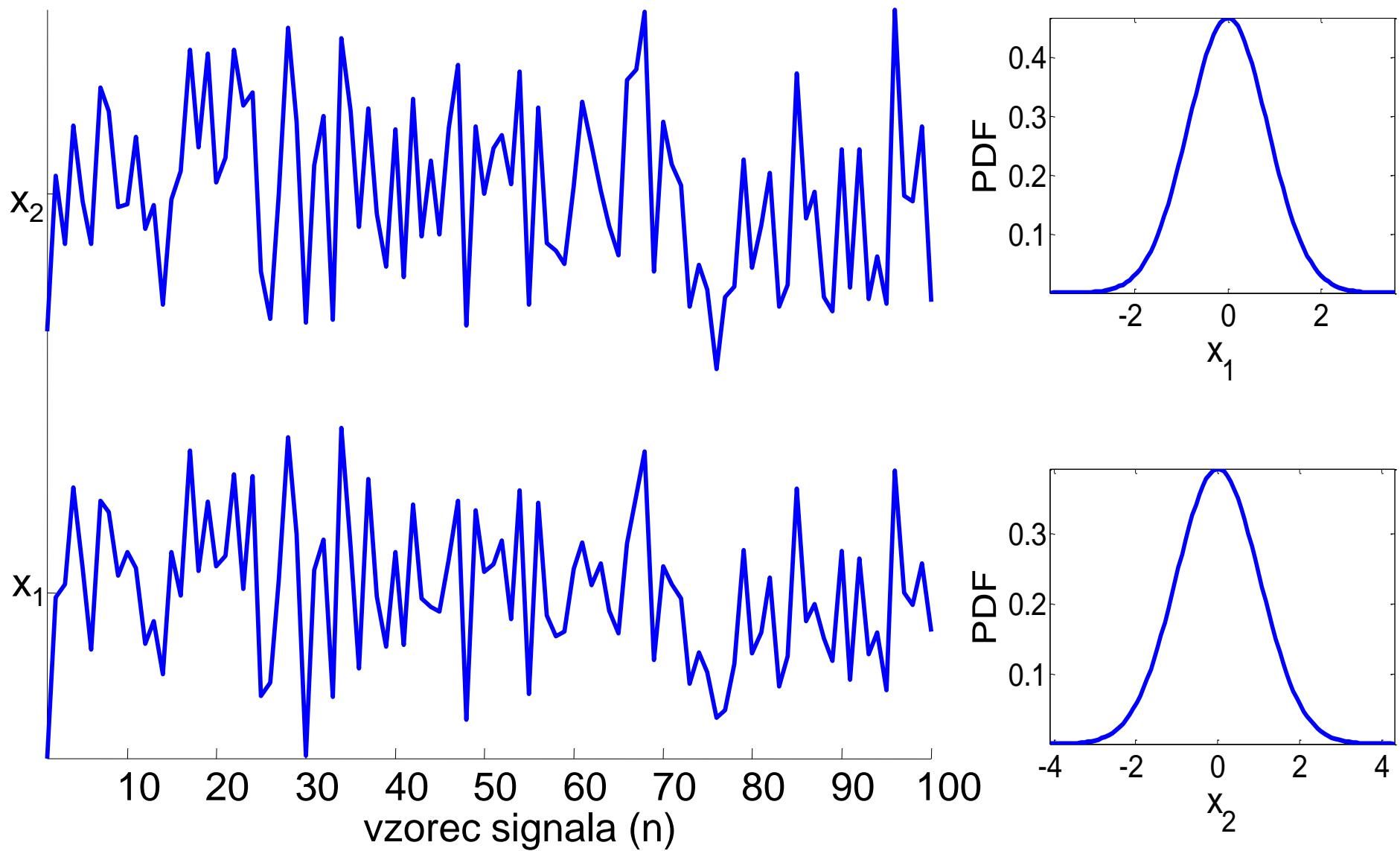
$$K \ll M$$

- Katera linearna transformacija \mathbf{P} je optimalna (v okviru omejitev ohranjanja variance)?
- Kakšno napako napravimo pri tem?

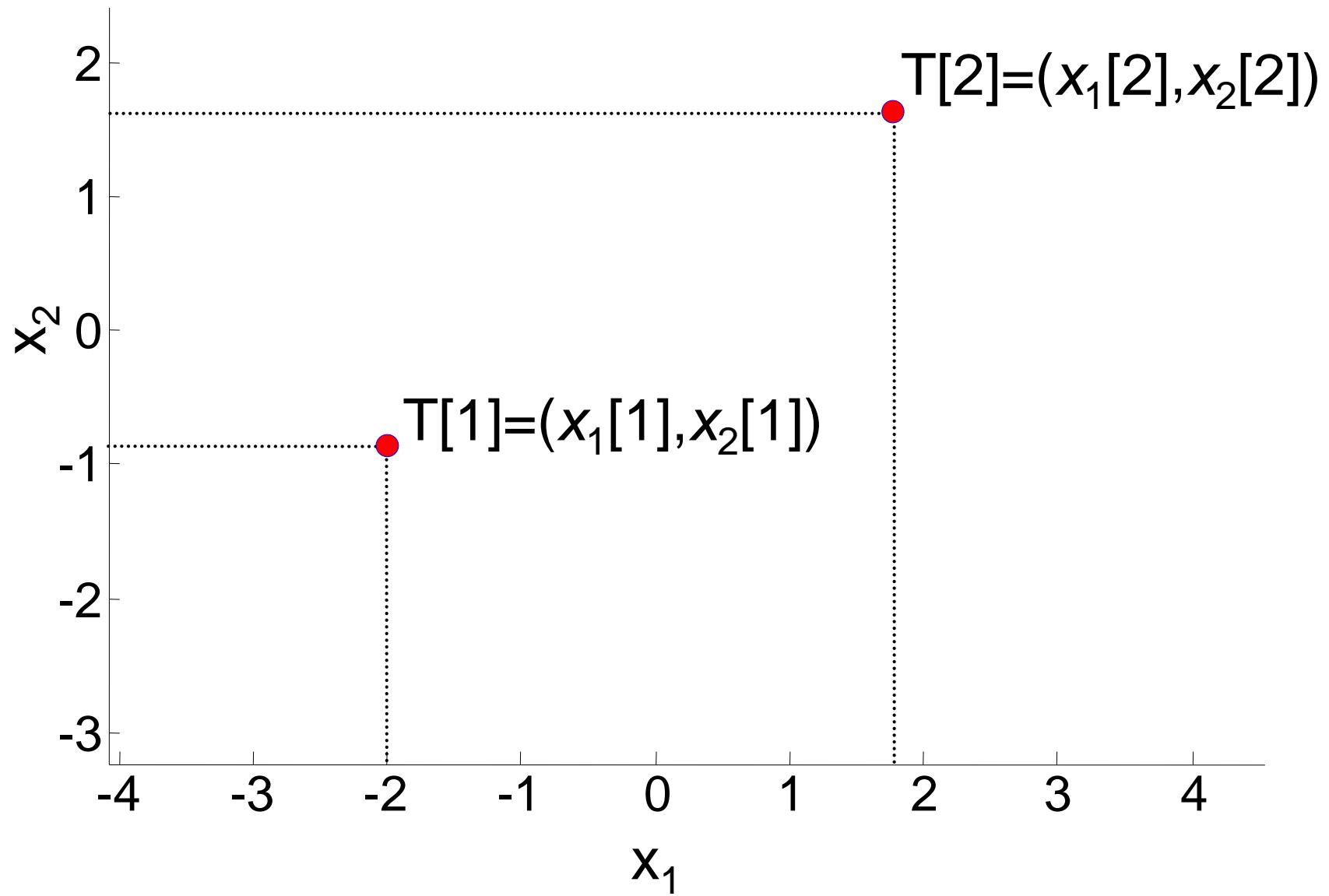
Primer 1: naključne meritve



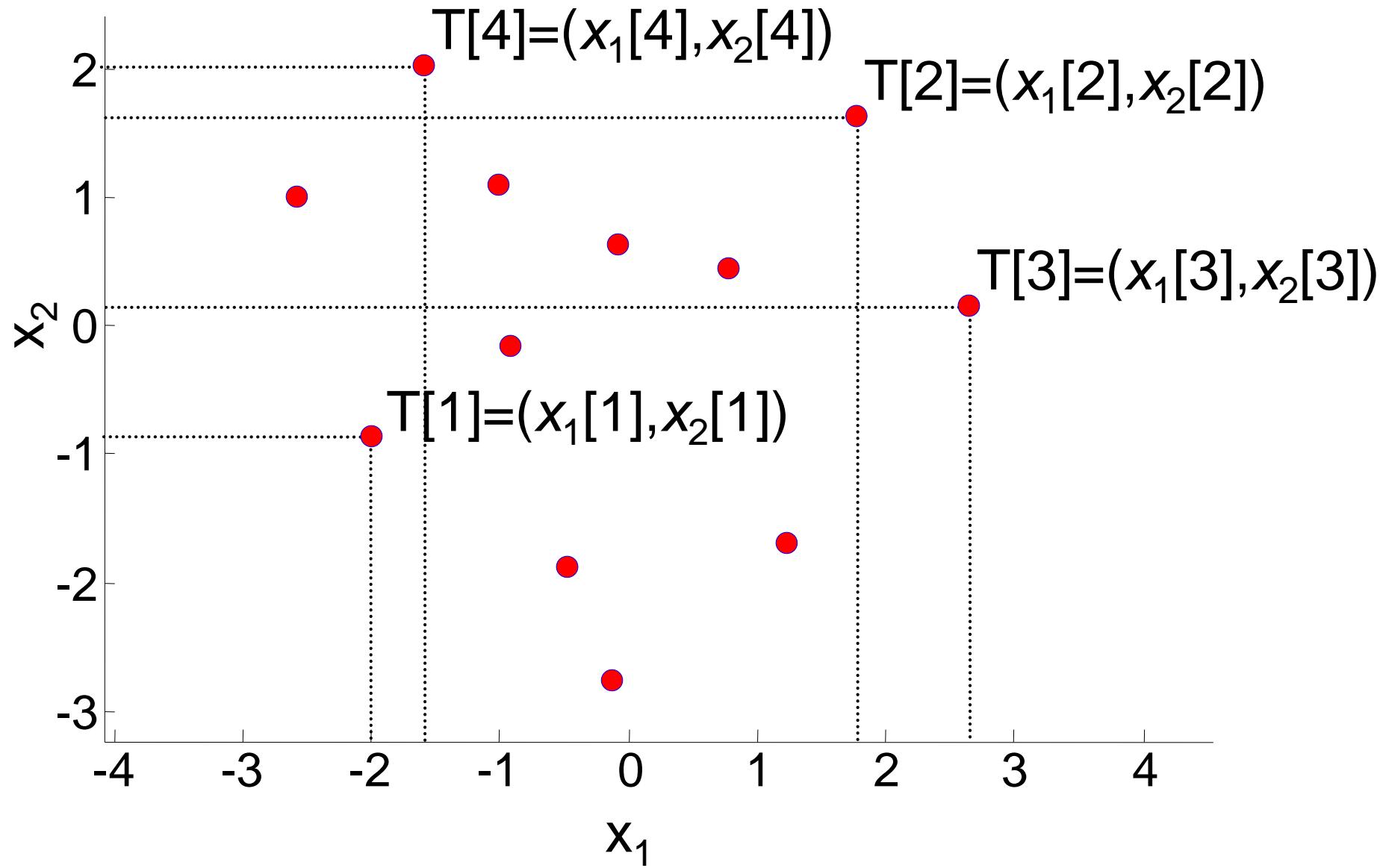
Primer 1: dve naključni meritvi



Upodobitev v skupnem vektorskem prostoru

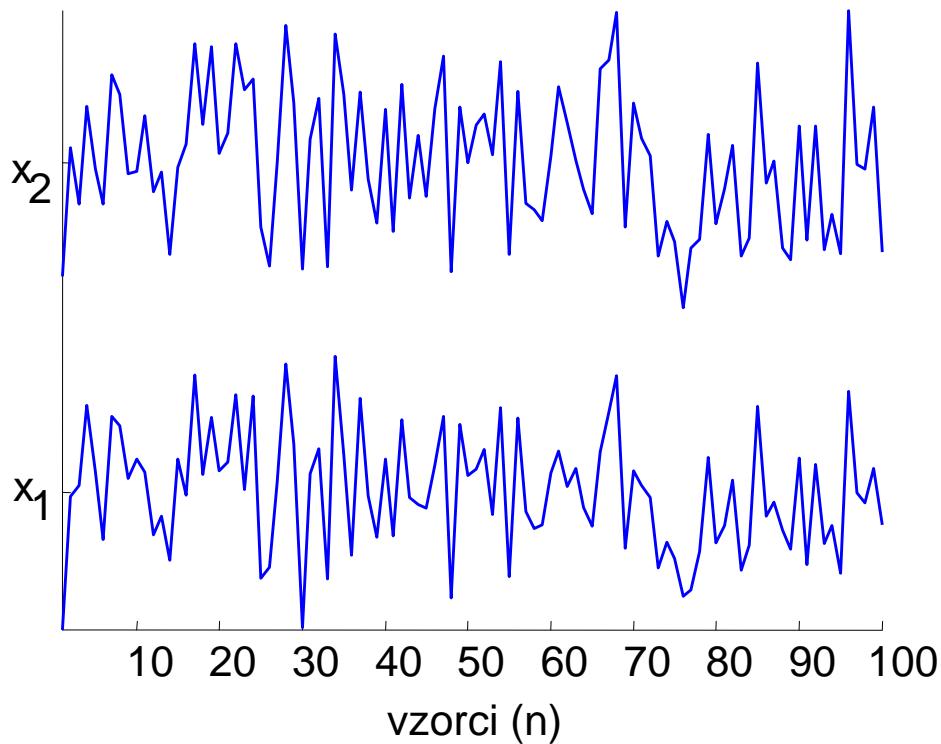


Upodobitev v skupnem vektorskem prostoru

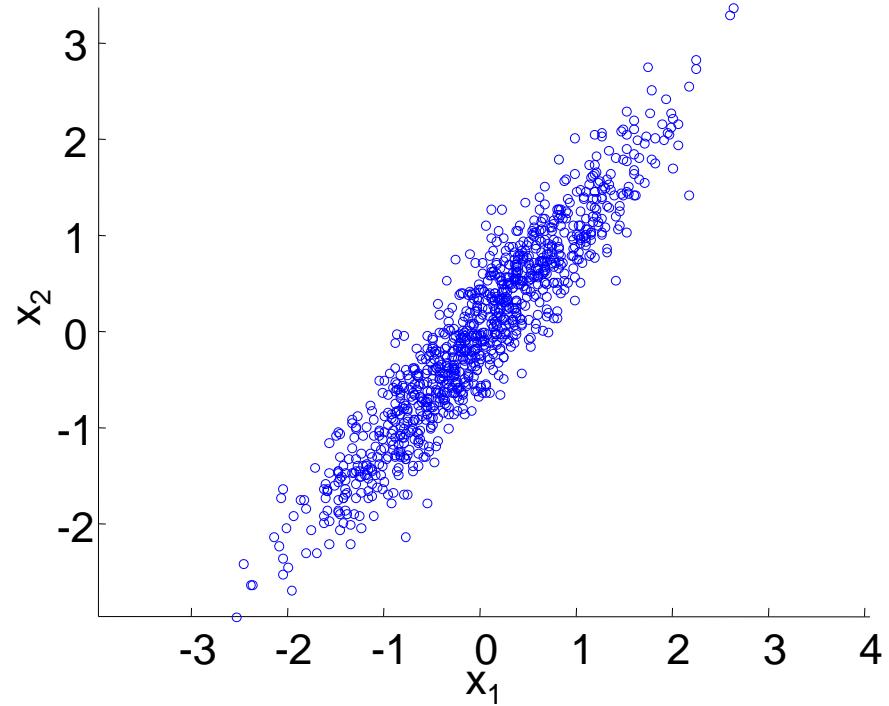


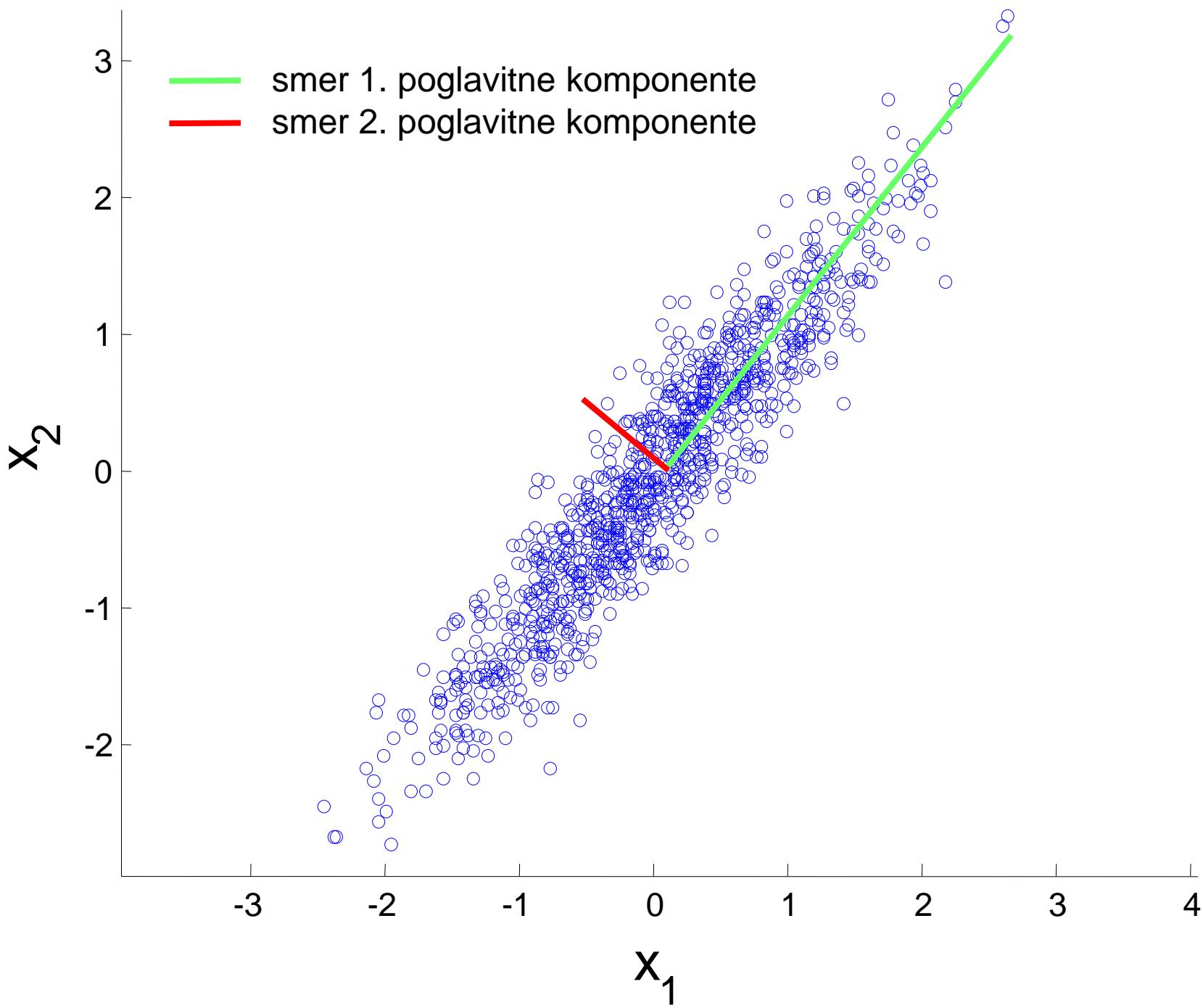
Primer 1: dve naključni meritvi

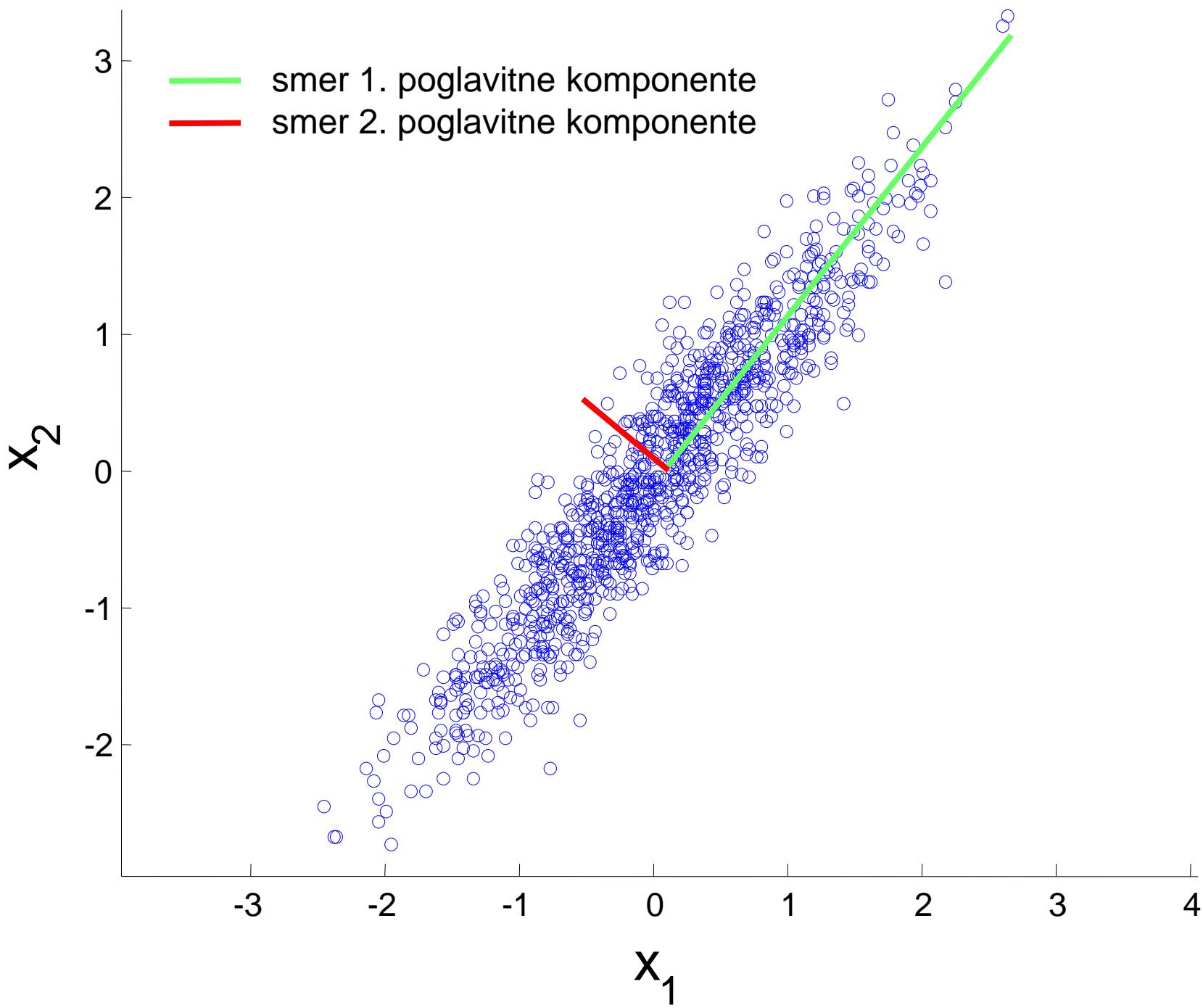
časovna domena



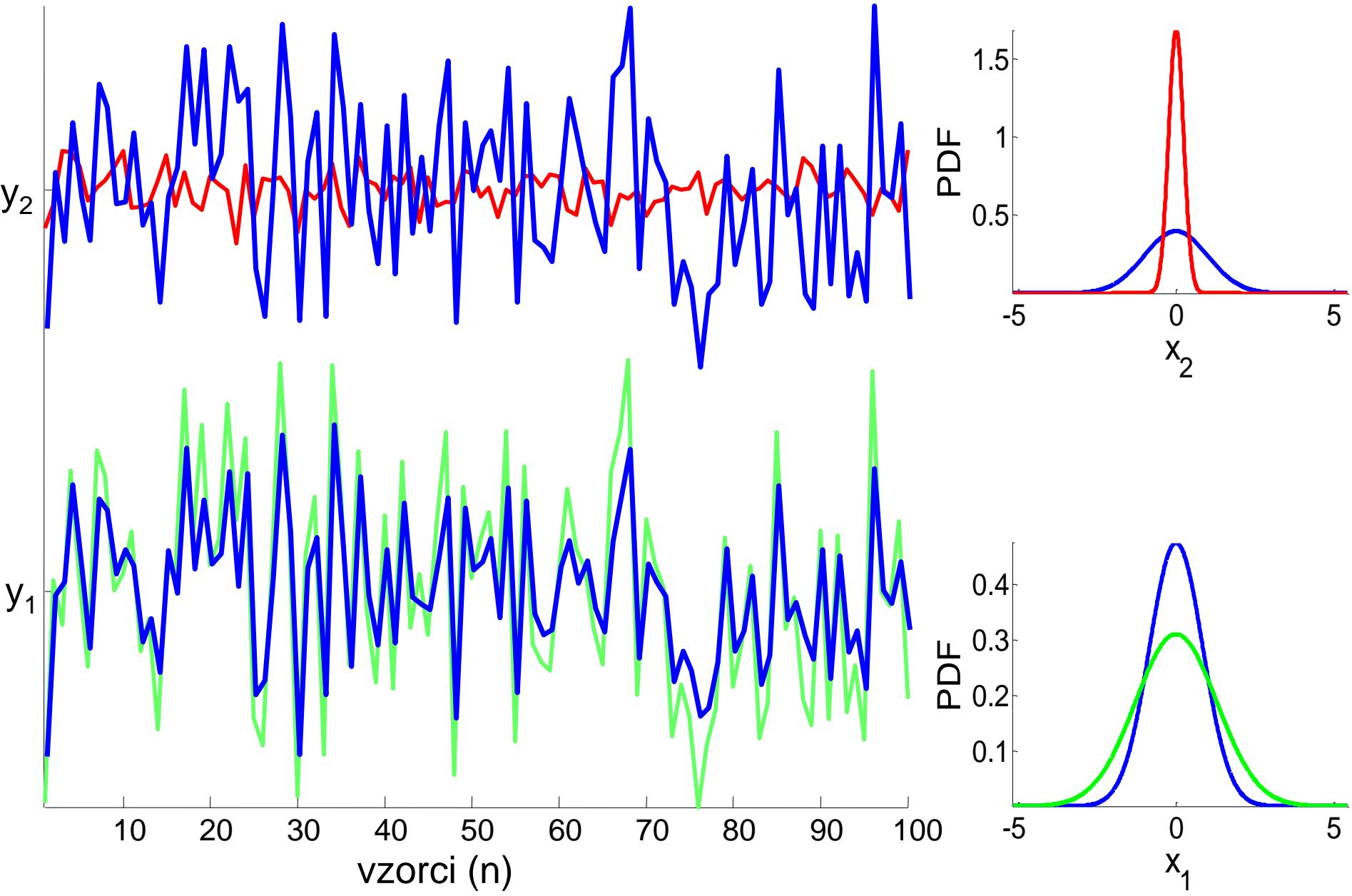
skupni vektorski prostor



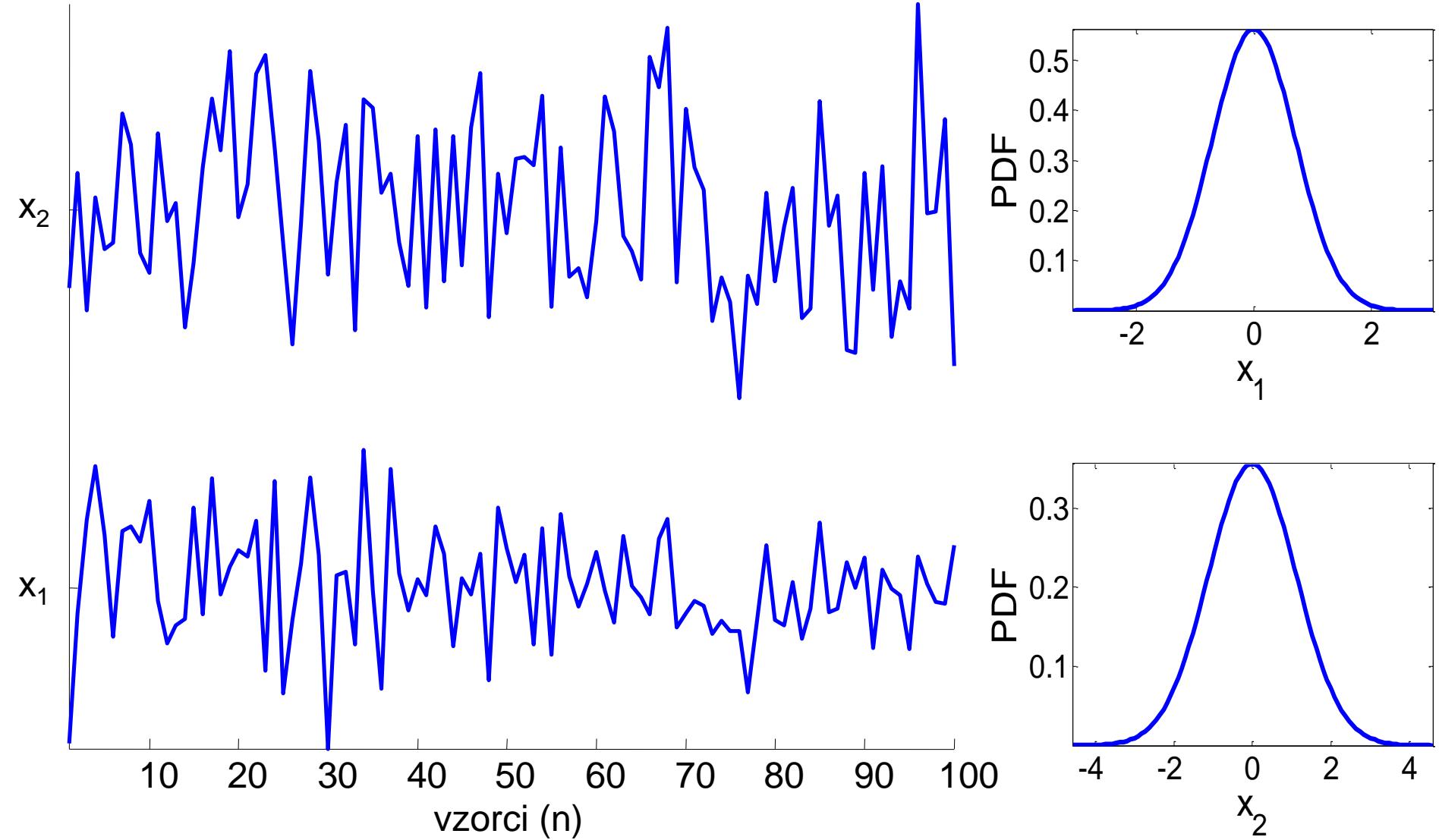




Primer 1: dve naključni meritvi

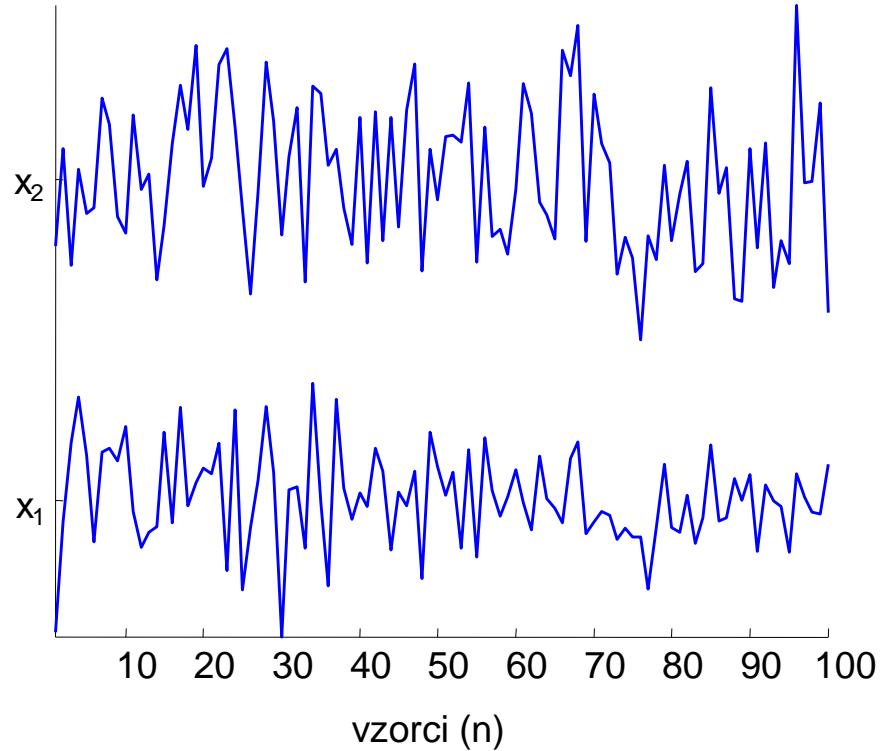


Primer 2 : drugi dve naključni meritvi

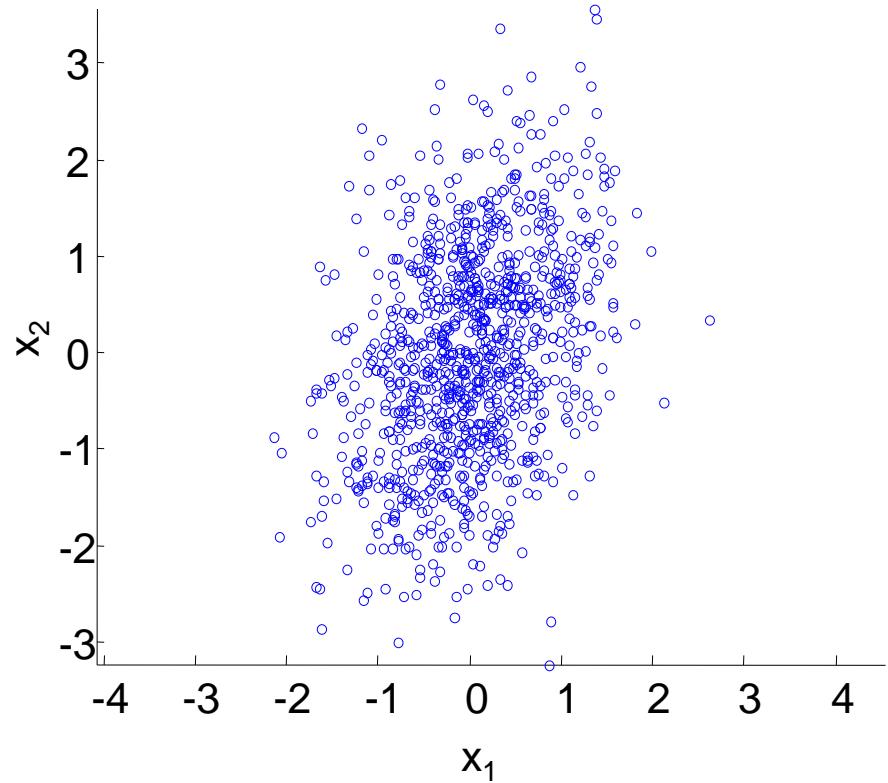


Primer 2: naključni meritvi

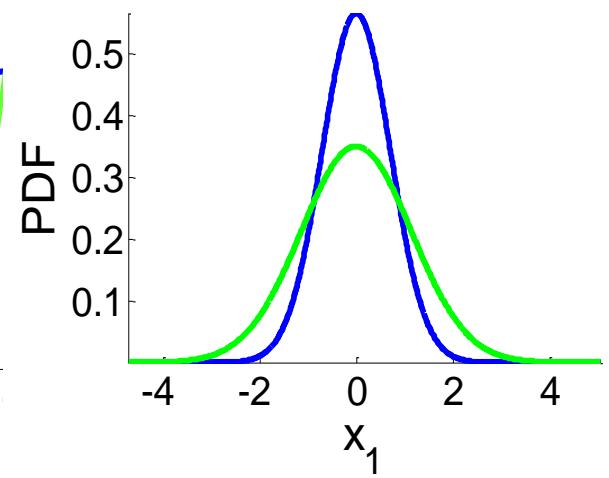
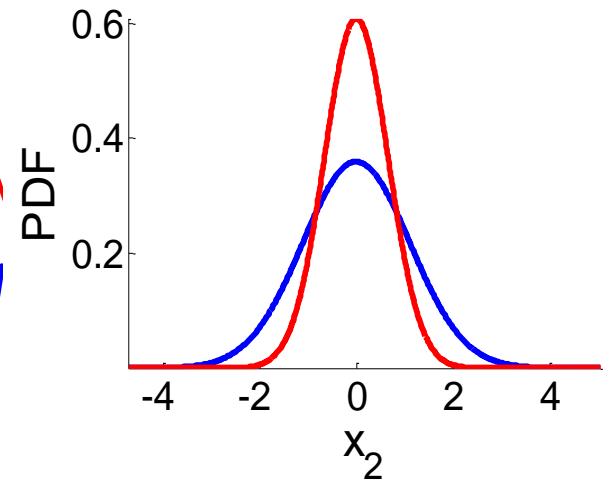
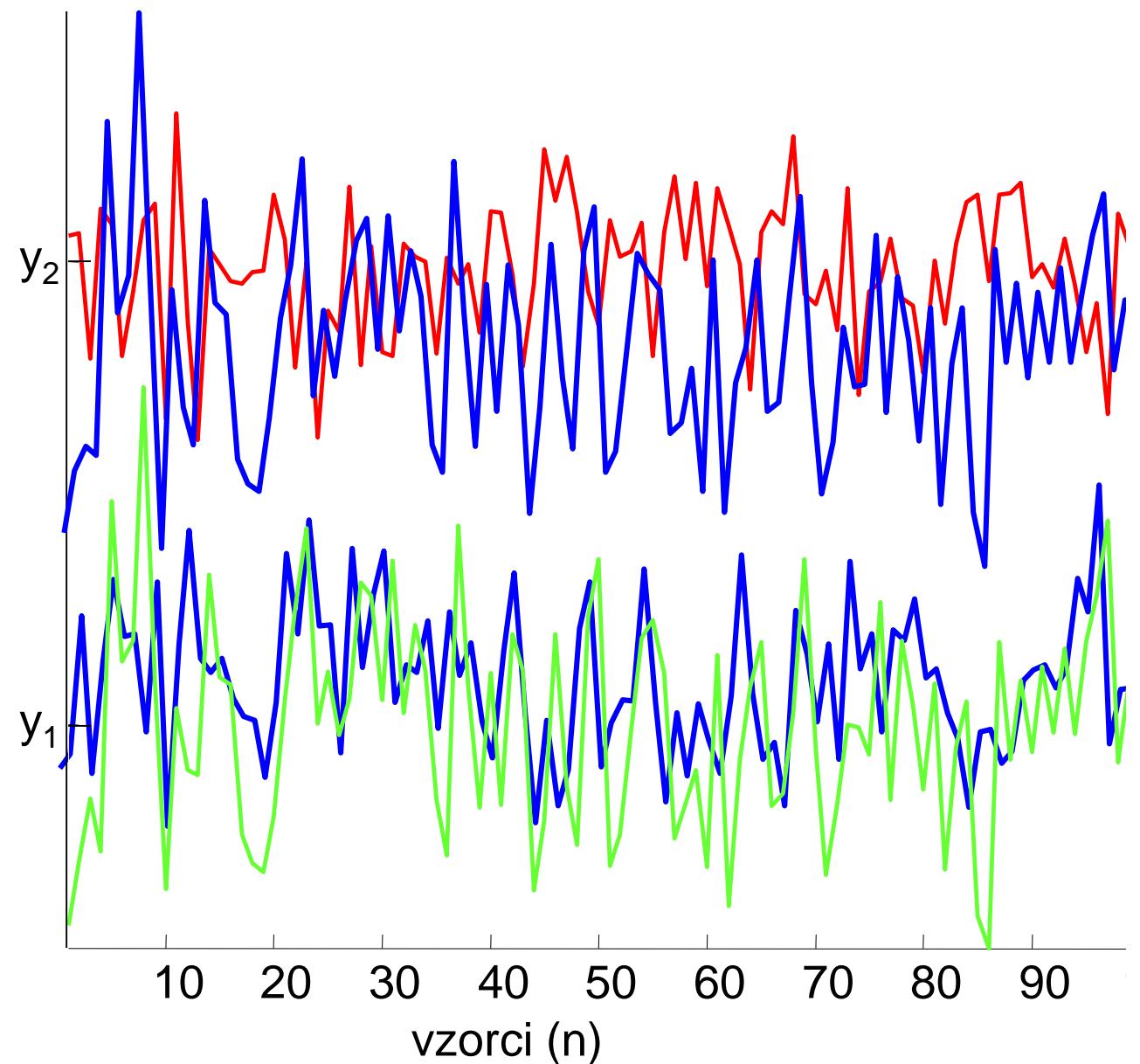
časovna domena



skupen vektorski prostor



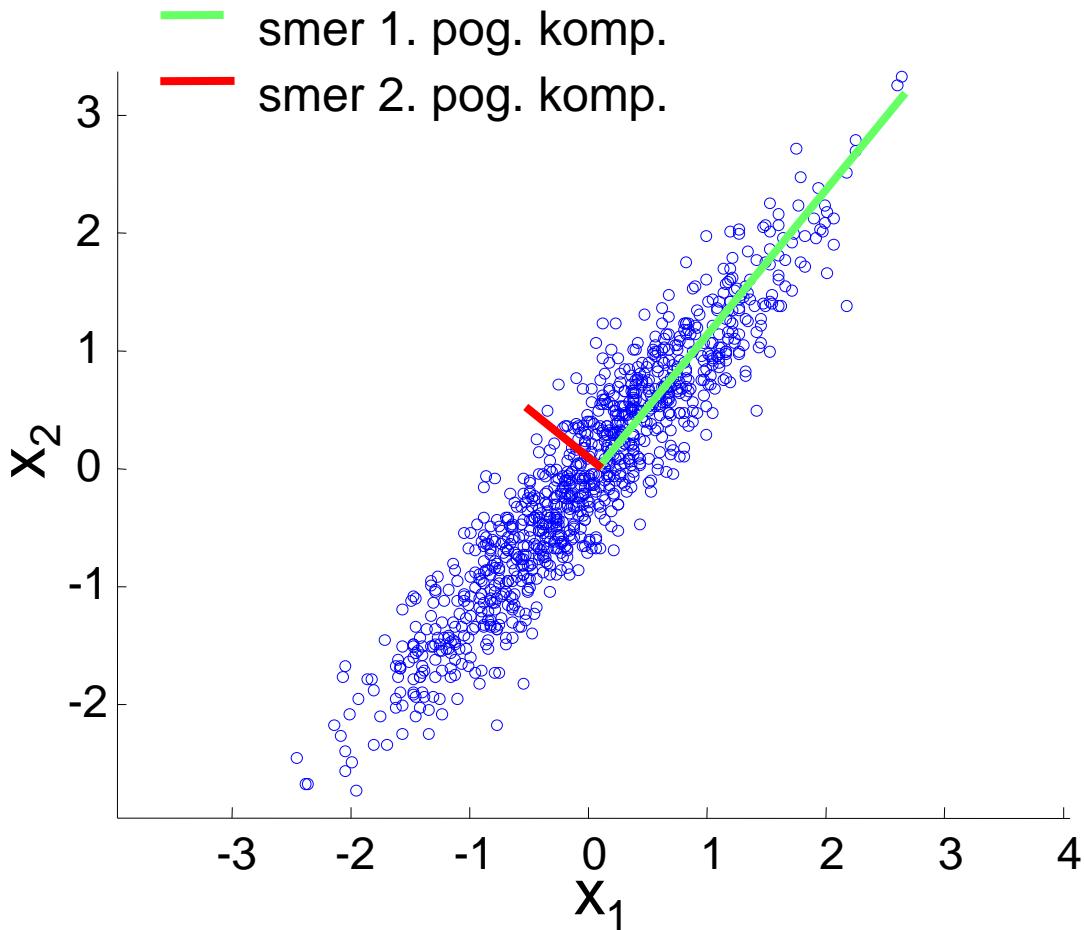
Primer 2: naključni spremenljivki



Naloge in cilji PCA

- Ali obstaja baza prostora, ki je linearna kombinacija prvotno podanih baznih vektorjev (podanih meritev), in je bolj primerna za opis danega nabora podatkov (ob omejitvah ohranjanja variance in pravokotnosti baznih vektorjev)?
- Kakšna je transformacija P , ki preslika dano bazo prostora v novo bazo prostora:
$$\mathbf{y}[n] = \mathbf{P}\mathbf{x}[n]$$
- Kako lahko določimo to transformacijo?

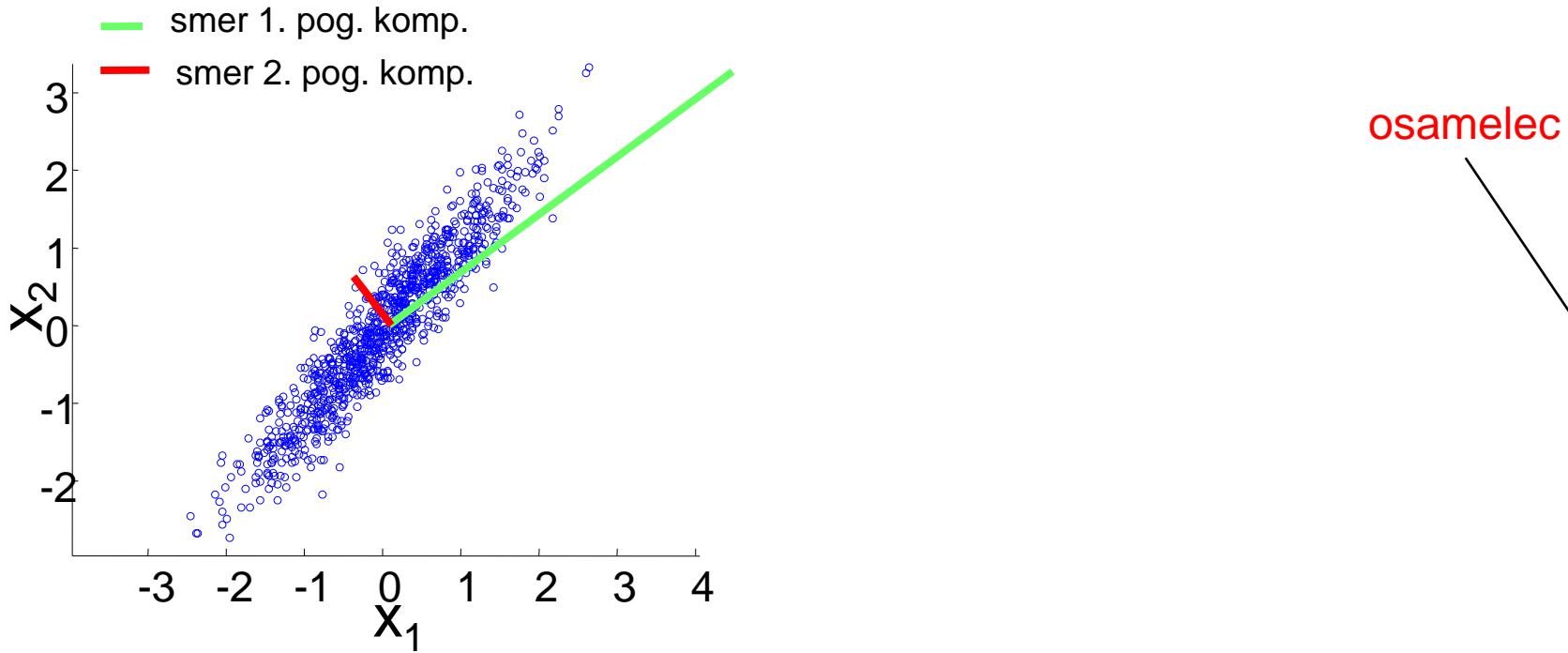
PCA z maksimizacijo variance



- 1. bazni vektor = smer največje variance
- Naj bodo vsi bazni vektorji ortogonalni
- 2. bazni vektor je smer druge največje variance
- itd.

PCA z maksimizacijo variance

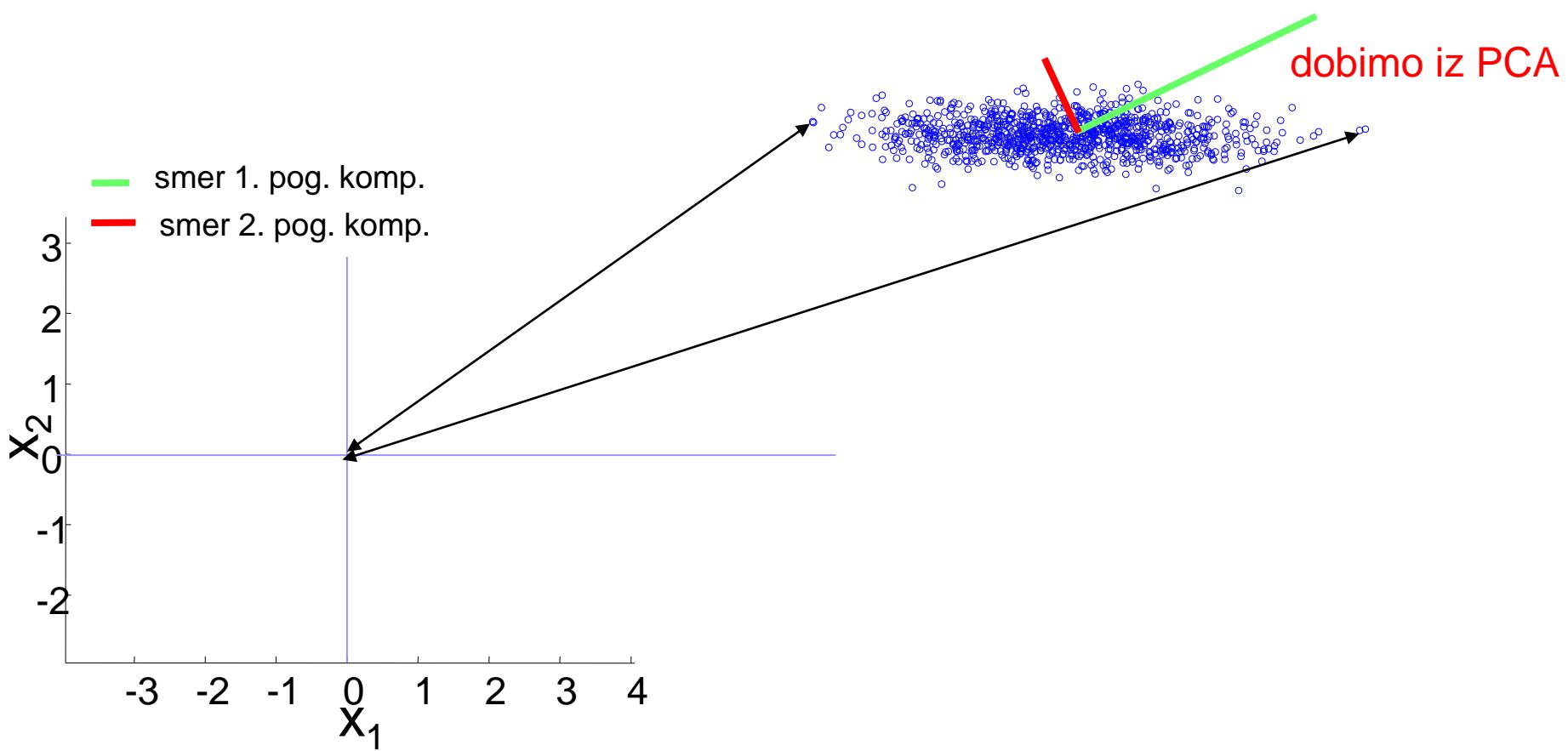
1. PCA je občutljiva na osamelce



PCA z maksimizacijo variance

2. PCA je občutljiva na neničelne povprečne vrednosti meritev

Ne dobimo iz PCA



Korelacijska matrika $\mathbf{x}[n]$

$$\mathbf{C}_{\mathbf{x}} = \frac{1}{N-1} \sum_{n=1}^N \mathbf{x}[n] \mathbf{x}^T[n] = \begin{bmatrix} c(x_1, x_1) & c(x_1, x_2) & \dots & c(x_1, x_M) \\ c(x_2, x_1) & c(x_2, x_2) & \dots & c(x_2, x_M) \\ \vdots & \vdots & \vdots & \vdots \\ c(x_M, x_1) & c(x_M, x_2) & \dots & c(x_M, x_M) \end{bmatrix}$$

kjer $c(x_i, x_j) = \frac{1}{N-1} \sum_{n=1}^N x_i[n] x_j[n]$

- $c(x_i, x_i)$ meri varianco $x_i \rightarrow$ maksimiziraj $c(y_i, y_i)$
- $c(x_i, x_j)$ meri korelacijo med x_i in $x_j \rightarrow$ minimiziraj $c(y_i, y_j)$

$$c(x_i, x_j) = \frac{1}{N-1} \sum_{n=1}^N x_i[n] x_j[n]$$

$x1 = [-1, 1, -1, 1, -1, 1, -1, 1]$
 $x2 = [-1, 1, -1, 1, -1, 1, -1, 1]$

$$c(x1, x2) = 1$$

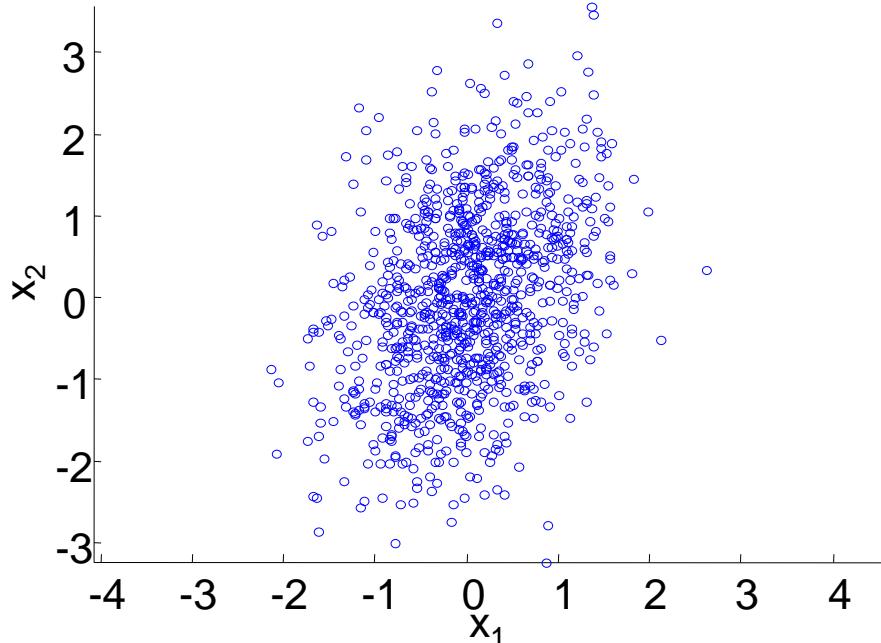
$x1 = [-1, 1, -1, 1, -1, 1, -1, 1]$
 $x2 = [1, -1, 1, -1, 1, -1, 1, -1]$

$$c(x1, x2) = -1$$

$x1 = [-1, 0, 1, 0, -1, 0, 1, 0]$
 $x2 = [0, -1, 0, 1, 0, -1, 0, 1]$

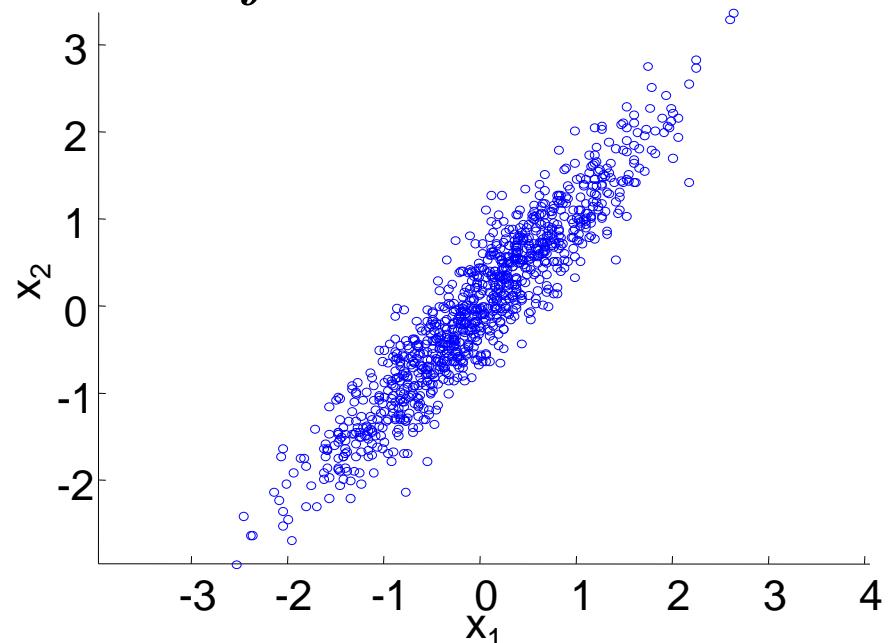
$$c(x1, x2) = 0$$

Prečna korelacija $c(x_i, x_j)$



Nizka redundanca

$$c(x_1, x_2) = 0.33$$



Visoka redundanca

$$c(x_1, x_2) = 0.90$$

Prečna korelacija meri redundanco!



minimiziraj $c(y_i, y_j)$!

Linearna transformacija \mathbf{P}

- Poiščimo ortogonalno linearno transformacijo \mathbf{P} , tako da bo $\mathbf{y}[n] = \mathbf{Px}[n]$ in bo

$$\mathbf{C}_y = \frac{1}{N-1} \sum_{n=1}^N \mathbf{y}^T[n] \mathbf{y}[n]$$

diagonalna matrika.

- Z drugimi besedami, diagonalizirajmo korelacijsko matriko meritev \mathbf{C}_x

Linearna transformacija \mathbf{P}

$$\begin{aligned}\mathbf{C}_y &= \frac{1}{N-1} \sum_{n=1}^N \mathbf{y}[n] \mathbf{y}^T[n] \\ &= \frac{1}{N-1} \sum_{n=1}^N \mathbf{Px}[n] (\mathbf{Px}[n])^T \\ &= \frac{1}{N-1} \sum_{n=1}^N \mathbf{Px}[n] \mathbf{x}^T[n] \mathbf{P}^T \\ &= \mathbf{P} \mathbf{C}_x \mathbf{P}^T\end{aligned}$$

Linearna transformacija \mathbf{P}

- \mathbf{C}_x je simetrična matrika. Diagonalizira jo ortonormalna matrika njenih lastnih vektorjev (\mathbf{E}):

$$\mathbf{C}_x = \mathbf{E} \mathbf{D} \mathbf{E}^T$$

kjer je:

- \mathbf{D} diagonalna matrika lastnih vrednosti λ_i matrike \mathbf{C}_x . Brez izgube splošnosti lahko privzamemo, da so lastne vrednosti $\lambda_i = \mathbf{D}[i,i]$ urejene v padajočem vrstnem redu.

$$\lambda_1 \geq \lambda_2 \geq \lambda_3, \dots, \geq \lambda_M$$

- \mathbf{E} je matrika ustreznih lastnih vektorjev matrike \mathbf{C}_x (lastni vektorji so stolpci matrike \mathbf{E}).

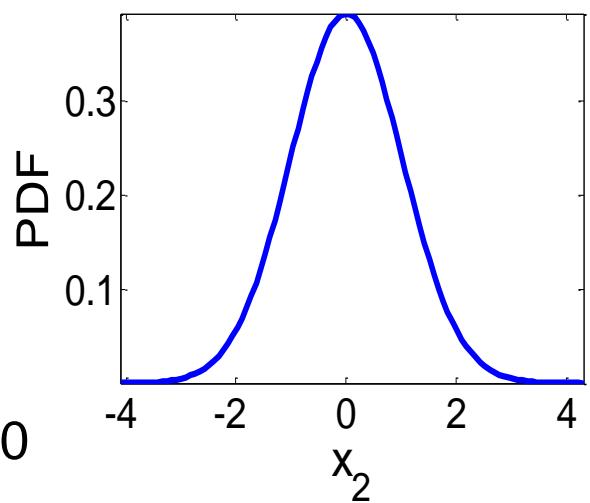
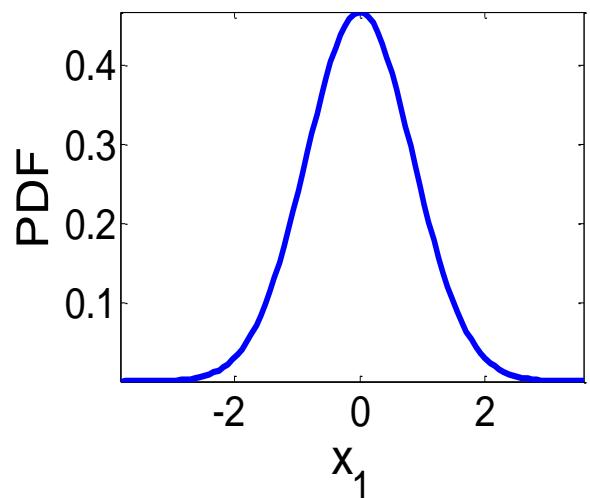
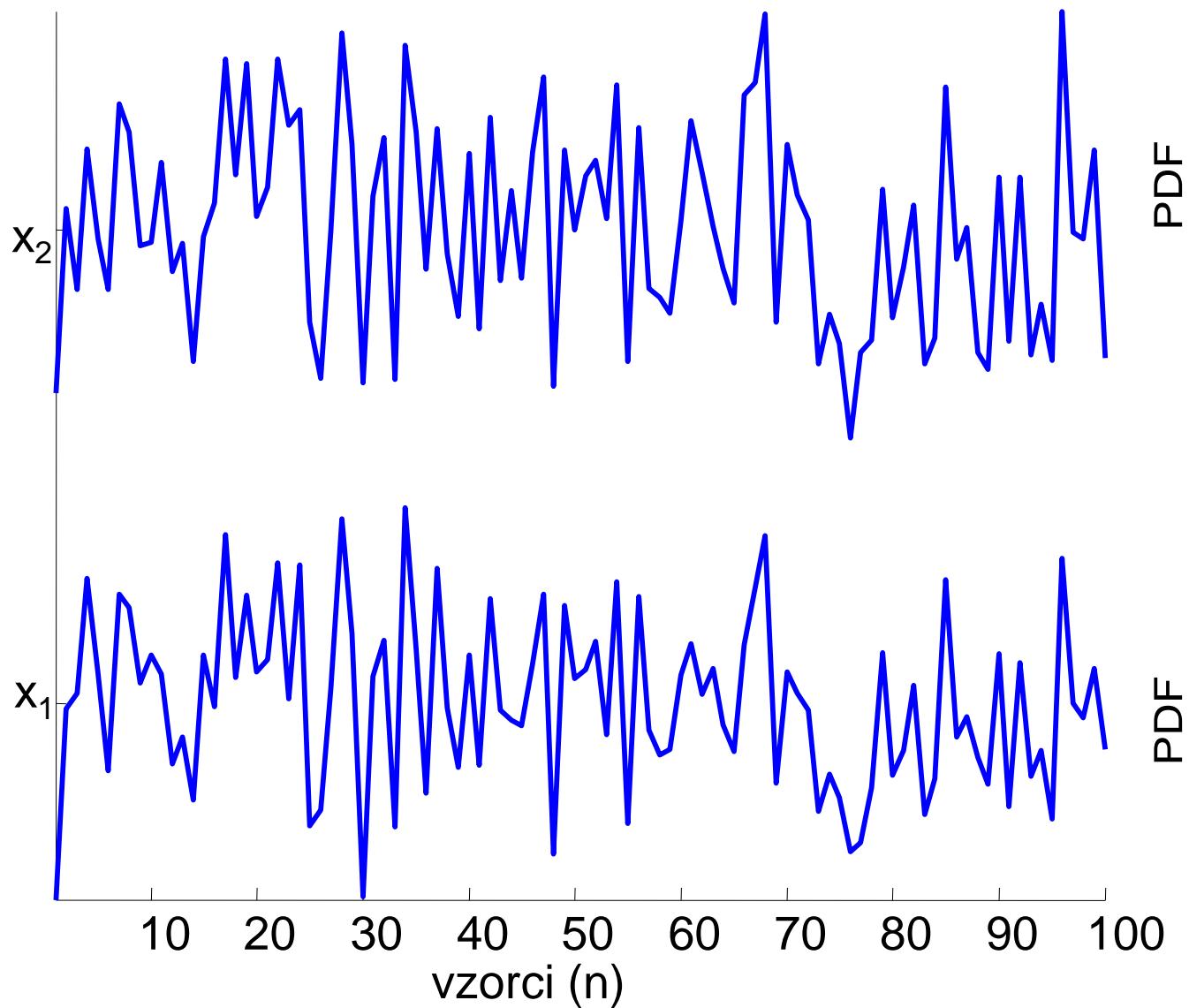
Linearna transformacija \mathbf{P}

- Matematičen trik: izberi $\mathbf{P} = \mathbf{E}^T$

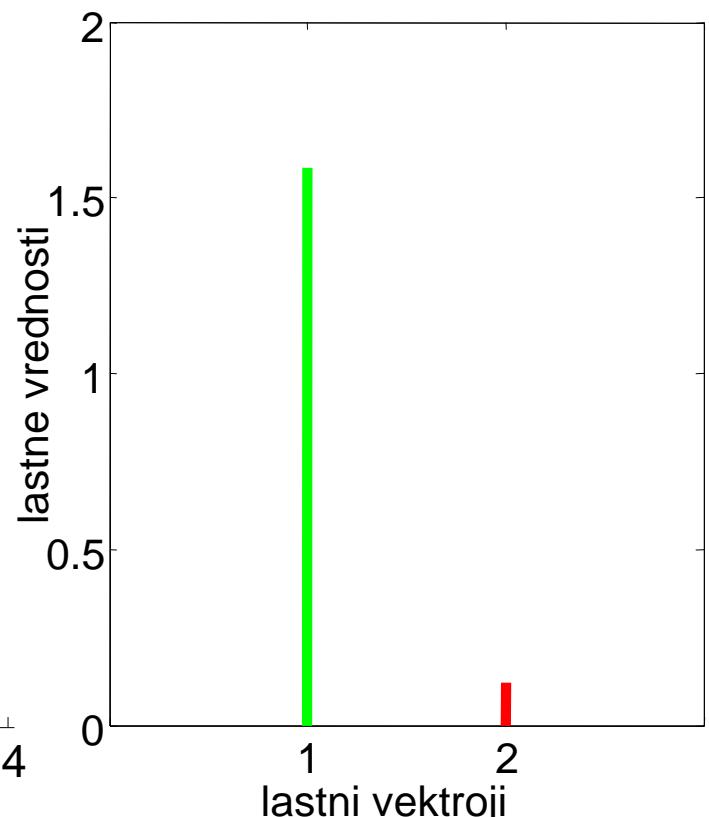
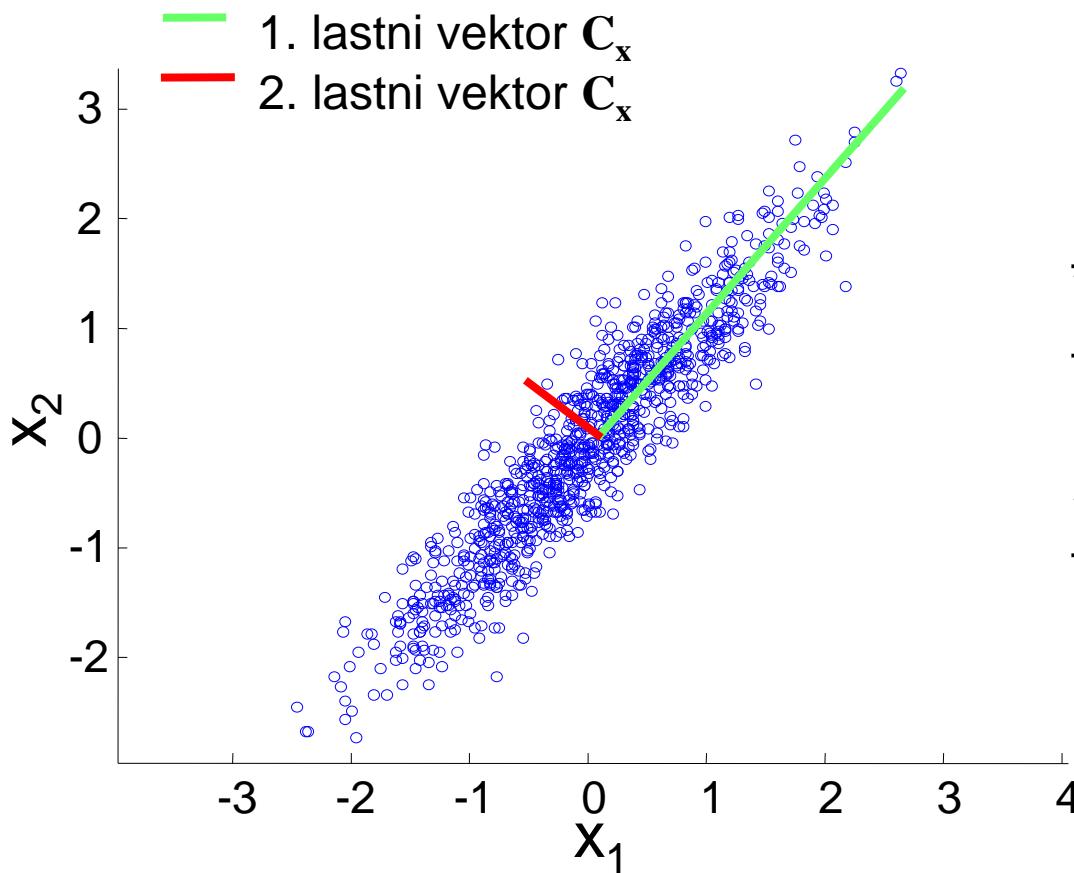
$$\begin{aligned}\mathbf{C}_y &= \mathbf{P} \mathbf{C}_x \mathbf{P}^T \\&= \mathbf{E}^T \mathbf{C}_x \mathbf{E} \\&= \mathbf{E}^T (\mathbf{E} \mathbf{D} \mathbf{E})^T \mathbf{E} \\&= (\mathbf{E}^T \mathbf{E}) \mathbf{D} (\mathbf{E}^T \mathbf{E}) \\&= \mathbf{D}\end{aligned}$$

$$\ker \mathbf{E}^{-1} = \mathbf{E}^T$$

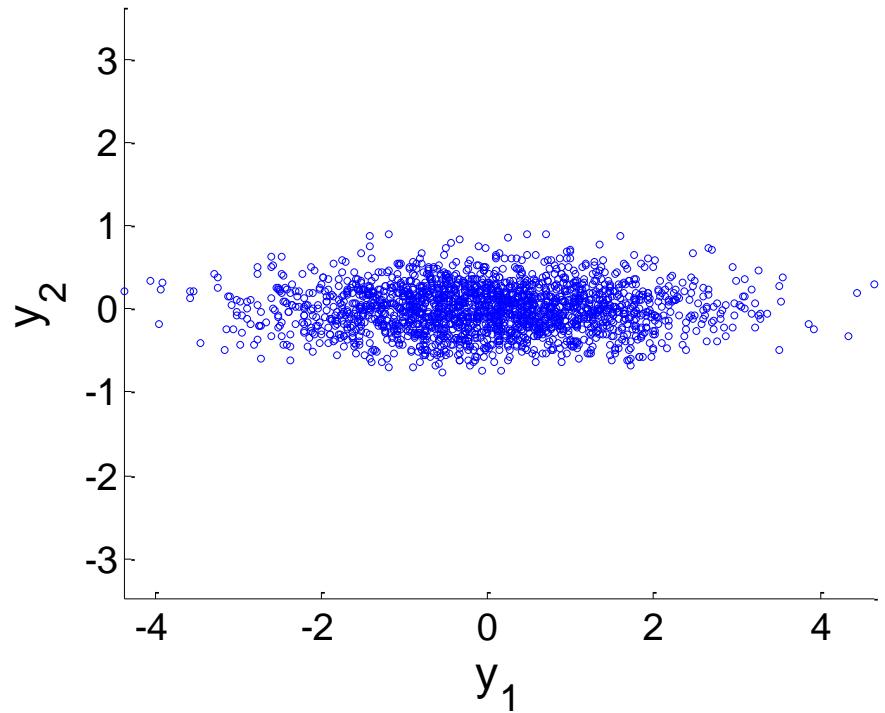
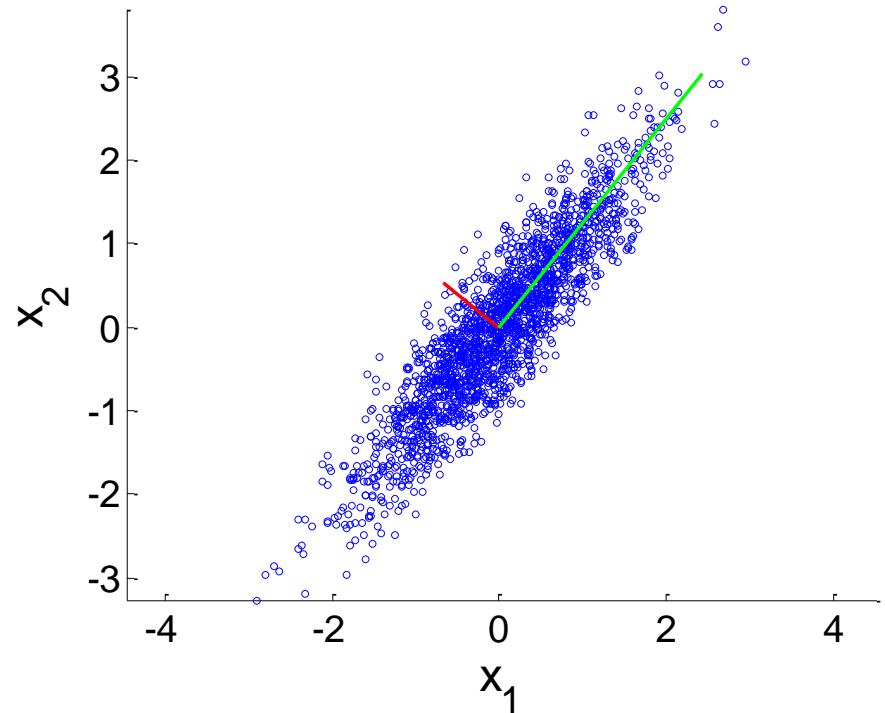
Grafični primer: Primer 1



Grafični primer: Primer 1

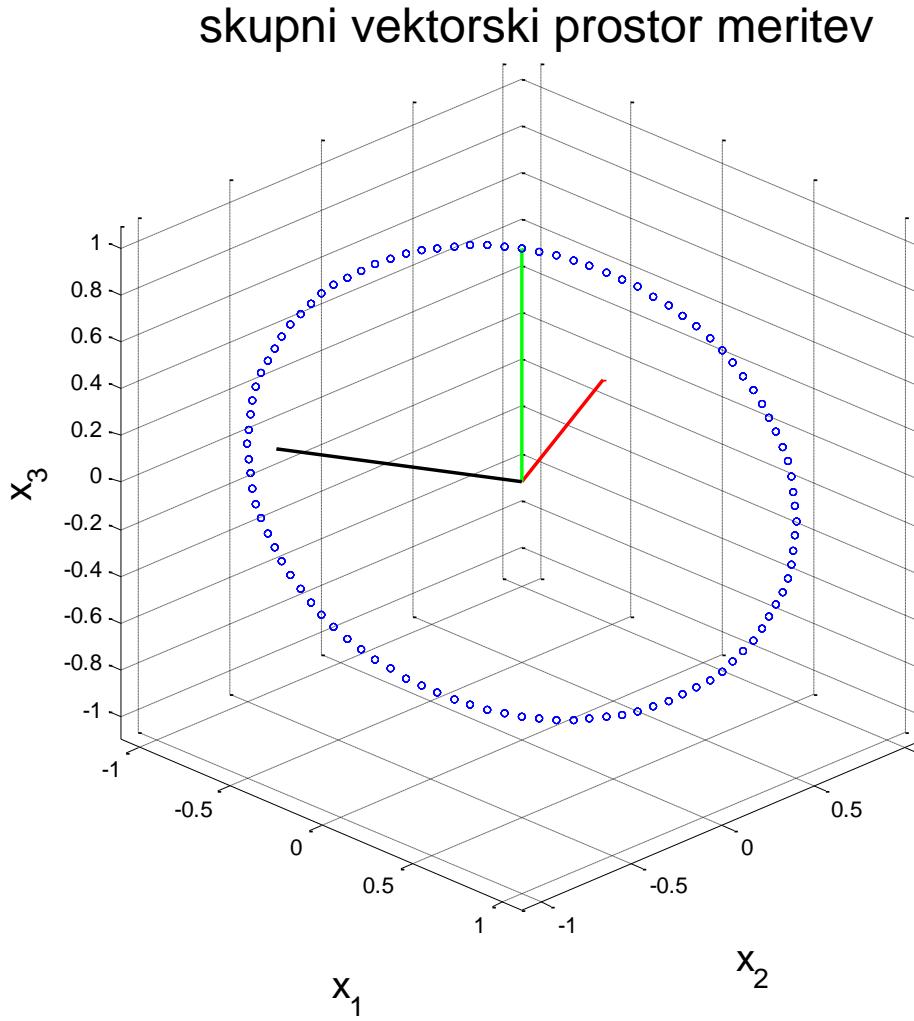
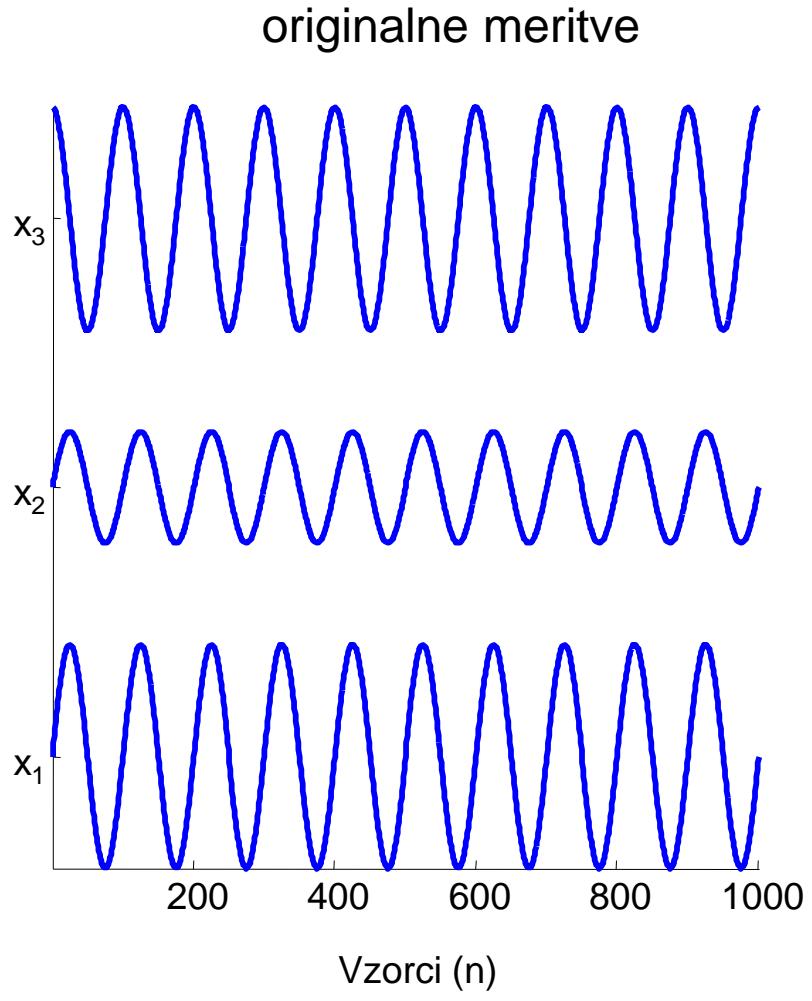


Grafični primer: Primer 1

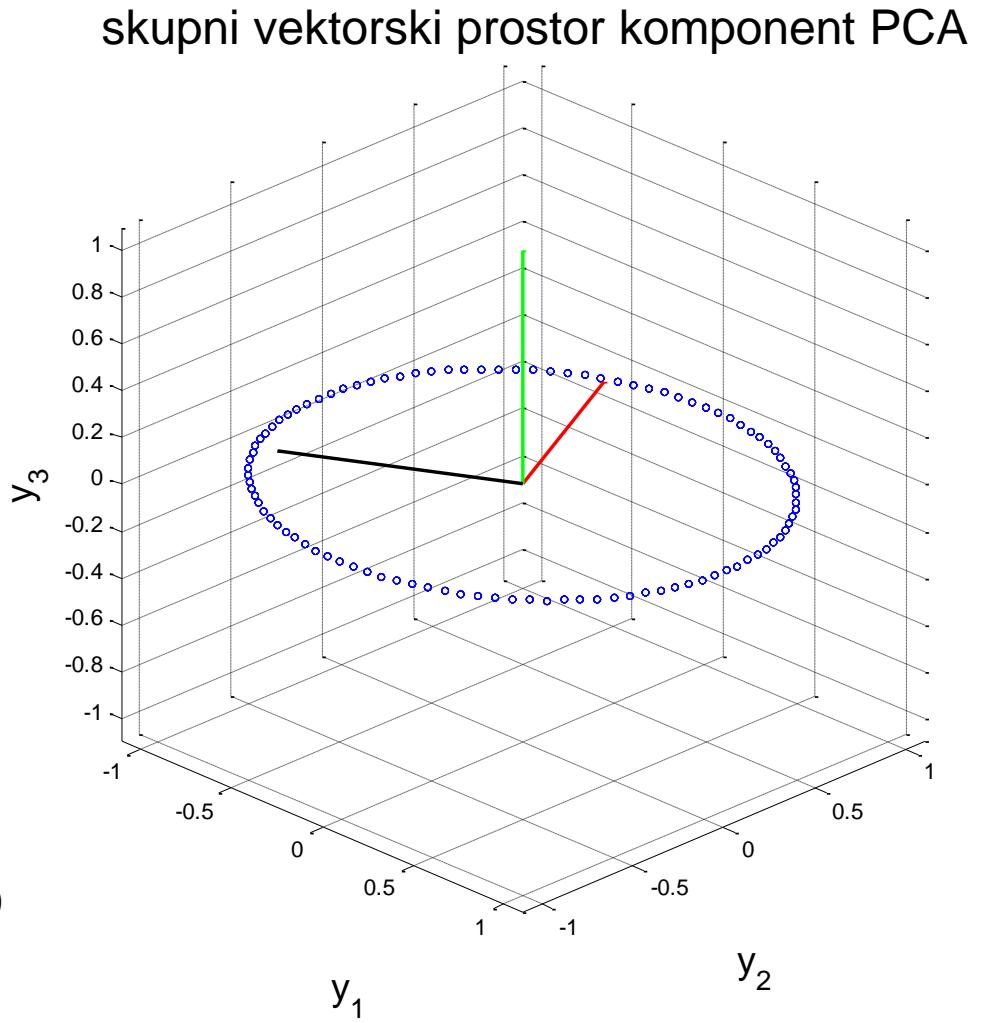
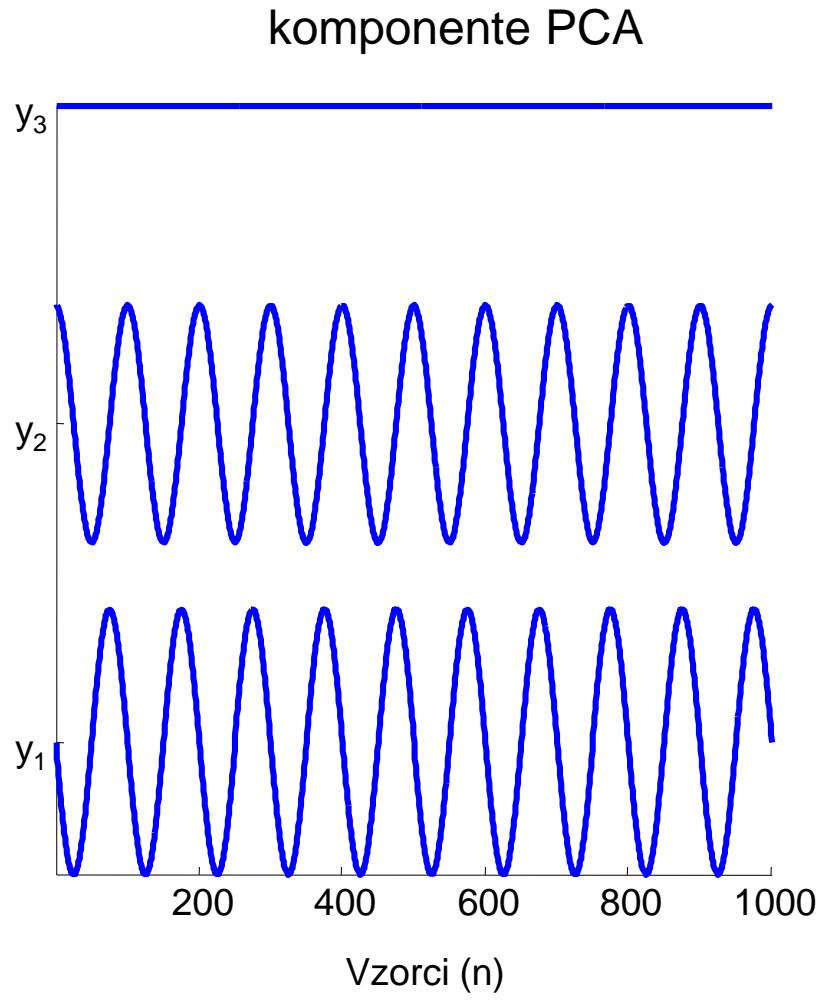


$$E\left\{y_m^2[n]\right\} = E\left\{\mathbf{e}_m^T \mathbf{x}[n] \mathbf{x}^T[n] \mathbf{e}_m\right\} = \mathbf{e}_m^T \mathbf{C}_{\mathbf{x}} \mathbf{e}_m = \lambda_m$$

Primer 3: PCA je lahko zavajajoča



Primer 3: PCA je lahko zavajajoča



Vmesni povzetek

- Smeri poglavitnih komponent meritev $\mathbf{x}[n]$ so določene z lastnimi vektorji korelacijske matrike meritev \mathbf{C}_x .
- i -ta lastna vrednost matrike \mathbf{C}_x določa varianco i -te poglavitne komponente meritev $\mathbf{x}[n]$
- Linearna transformacija (projekcija) meritev na njihove poglavitne komponente:

$$\mathbf{y}[n] = \mathbf{E}^T \mathbf{x}[n]$$

kjer je \mathbf{E} matrika lastnih vrednosti \mathbf{C}_x

Primer 4: sintetični površinski EMG

- simulator površinskega EMG [1]:

- koža: 1 mm
 - podkožno tkivo: 1.5 mm
 - prečni prerez mišice: 16 x 10 mm

- 120 motoričnih enot:

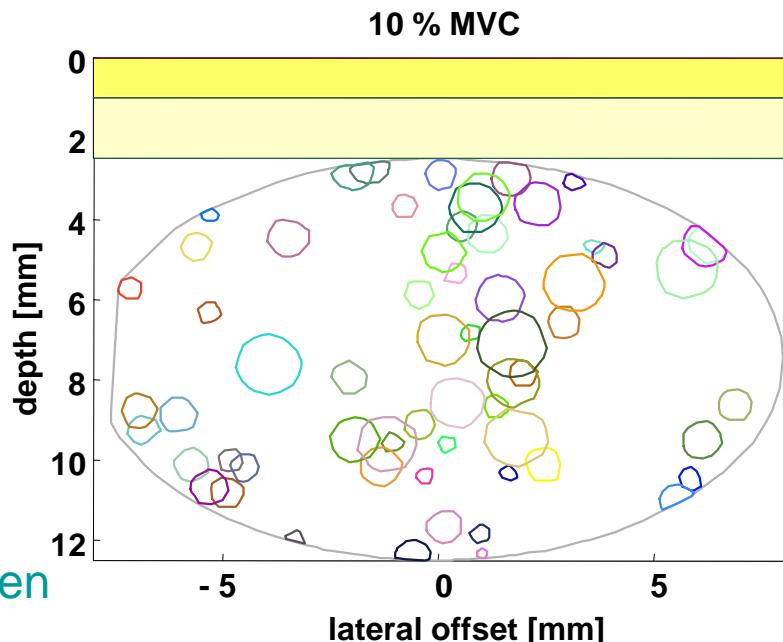
- dolžina vlakna: 50 mm
 - hitrost prevajanja AP: 4 ± 0.3 m/s
 - inervacijsko število: $25 \sim 2500$ vlaken

- Proženja ME [2]:

- vse ME rekrutirane pri 80 % maksimalne prostovoljne skrčitve (MVC)
 - simulirana stopnja skrčitve: 50 % MVC
 - hitrosti proženja simuliranih ME: 8 ~ 35 Hz

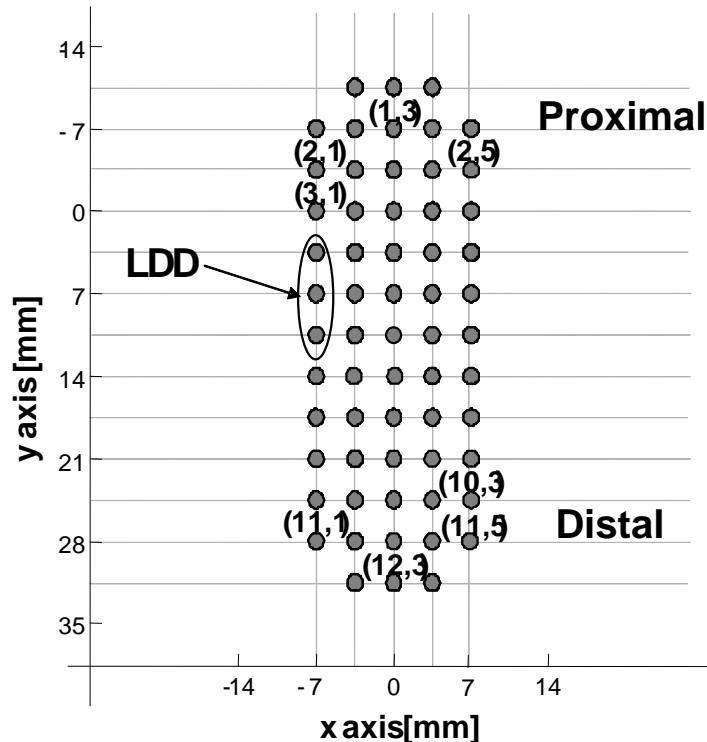
[1] Farina et al. A surface EMG generation model with multilayer cylindrical description of the volume conductor. IEEE Trans Biomed Eng. 2004; 51(3):415-26.

[2] Fuglevand et al. Models of recruitment and rate coding organization in motor unit pools. J. Neurophysiol 1993; 70:2470–88.

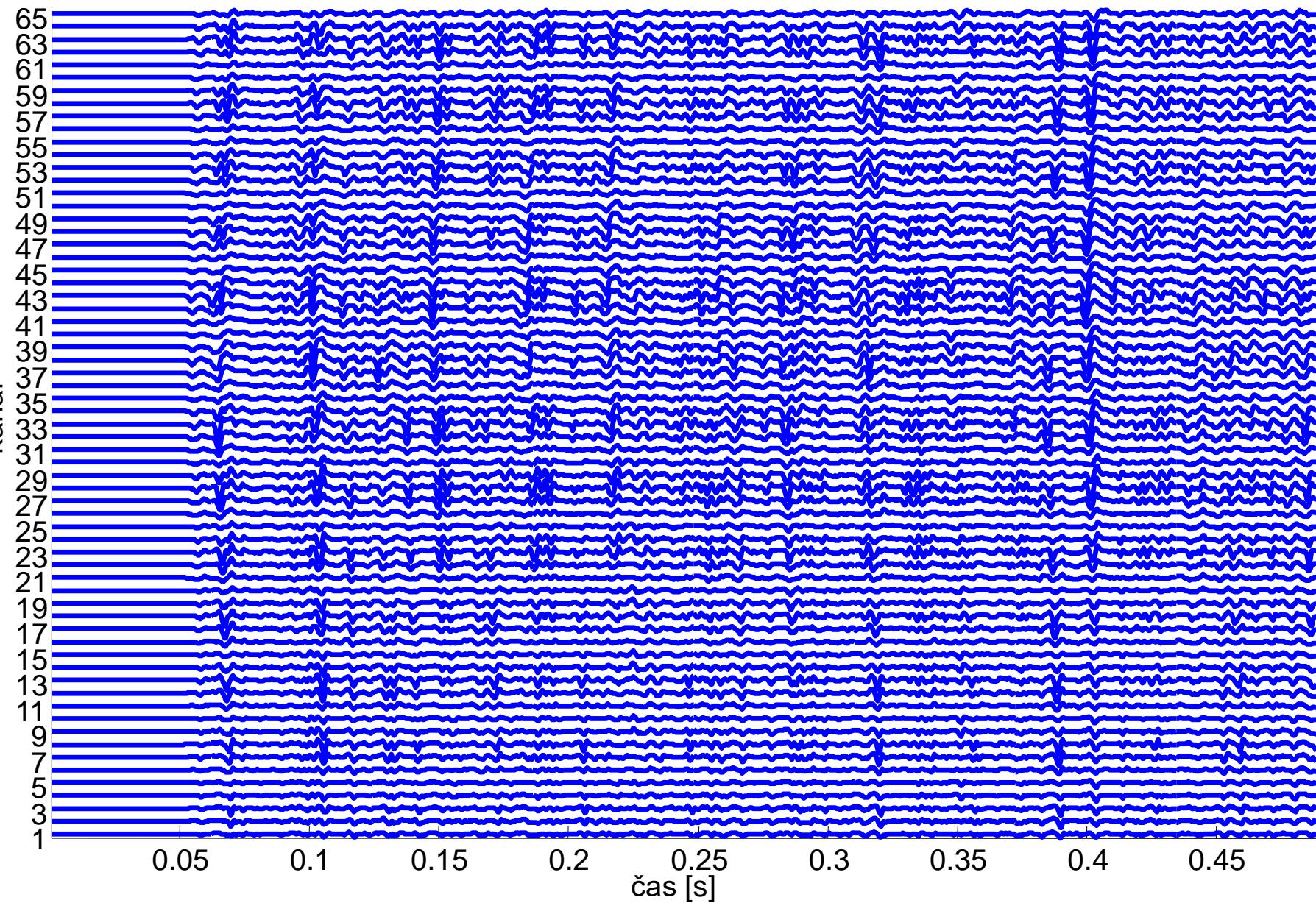


Primer 4: polje elektrod

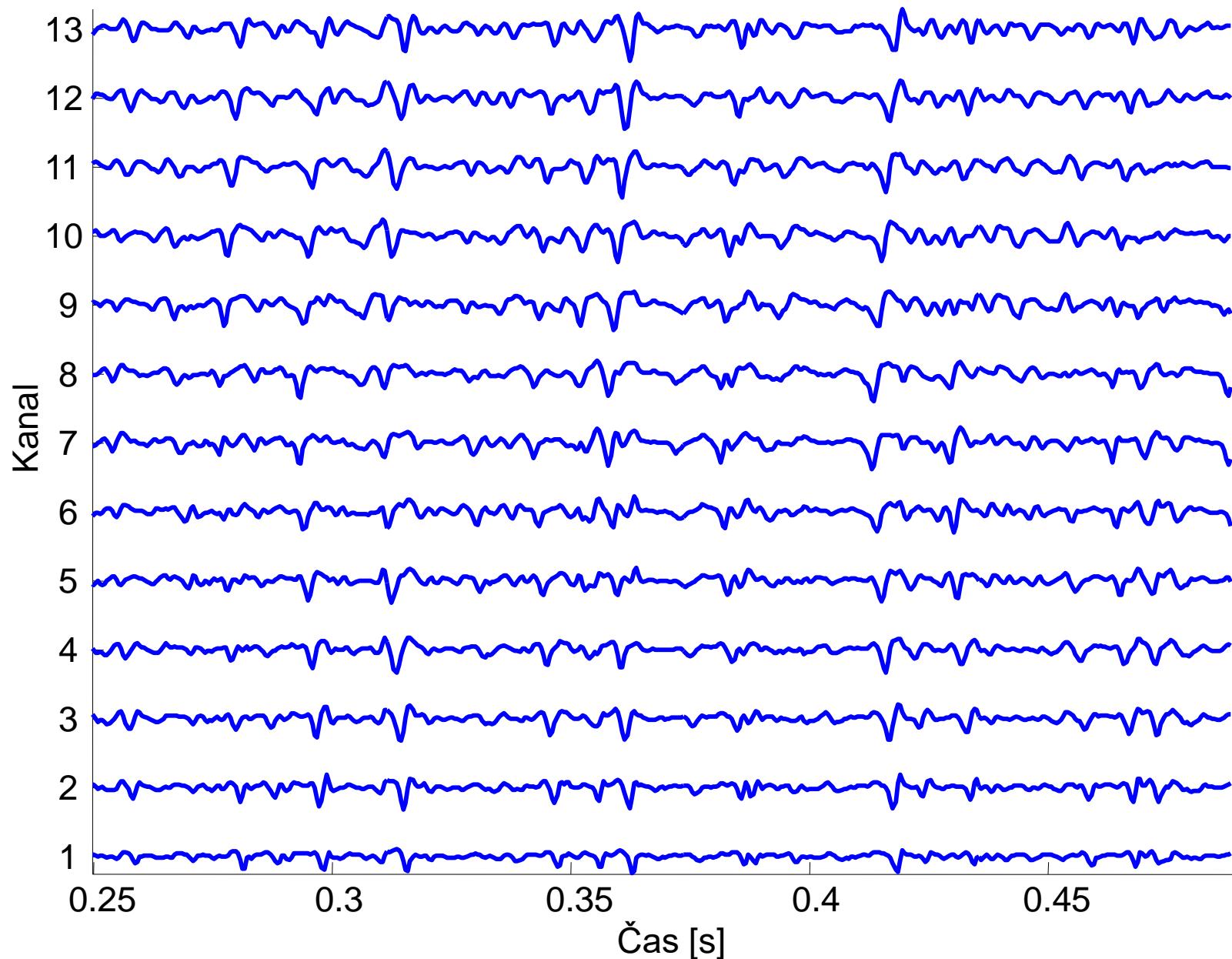
- simulirano polje elektrod:
 - medelektrodná razdalja: 3.5 mm
 - premer elektrod: 1.0 mm
 - št. kanalov: 61
- Snemalna tehnika:
 - vzdolžno dvojno diferenciranje (longitudinal double differential)
 - frekvenca vzorčenja: 1024 Hz
 - dolžina signalov: 20 s
- SNR: brez šuma



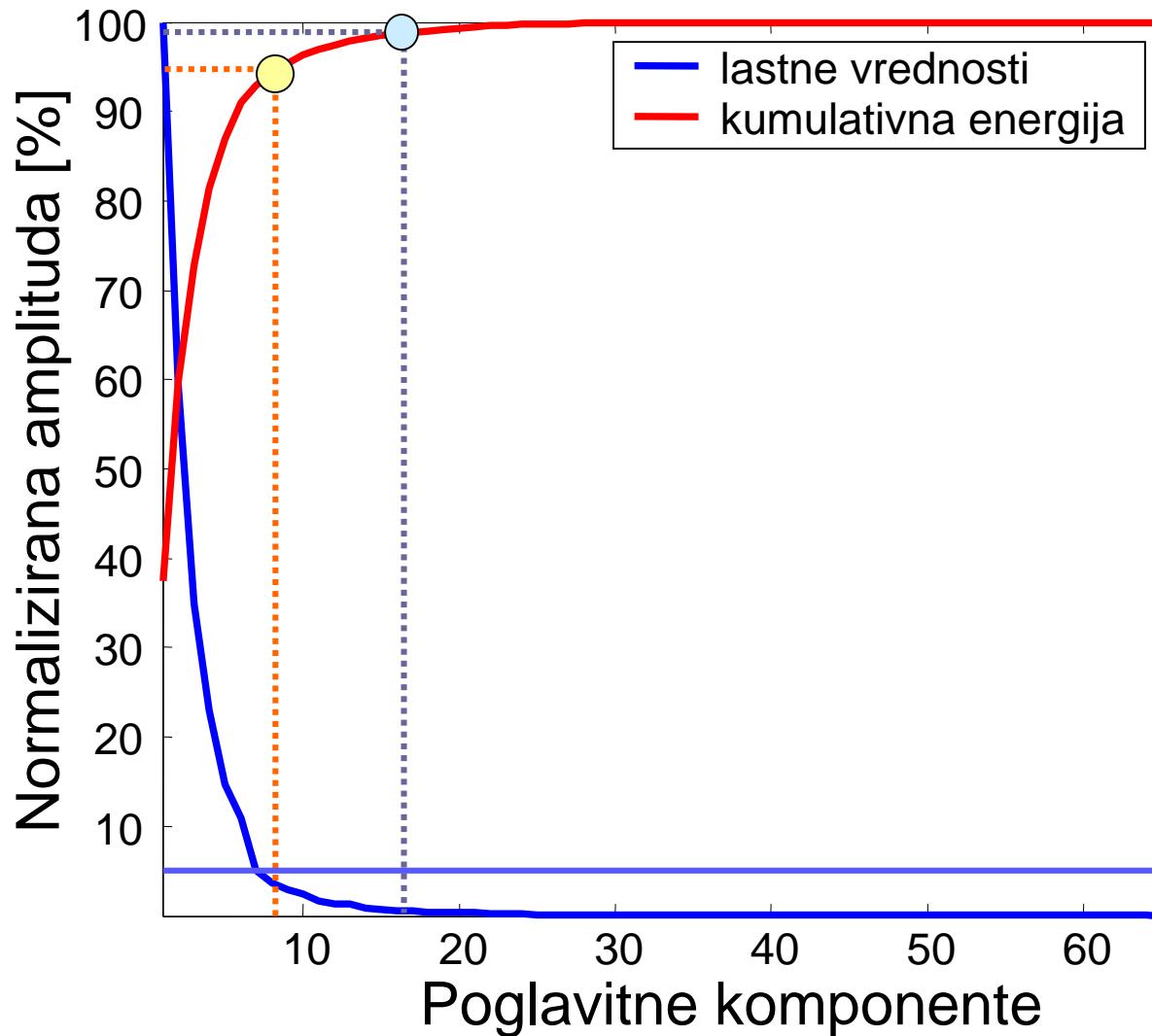
Primer 4: sintetični površinski EMG



Primer 4: sintetični EMG (centralni stolpec elektrod)



Redukcija dimenzij



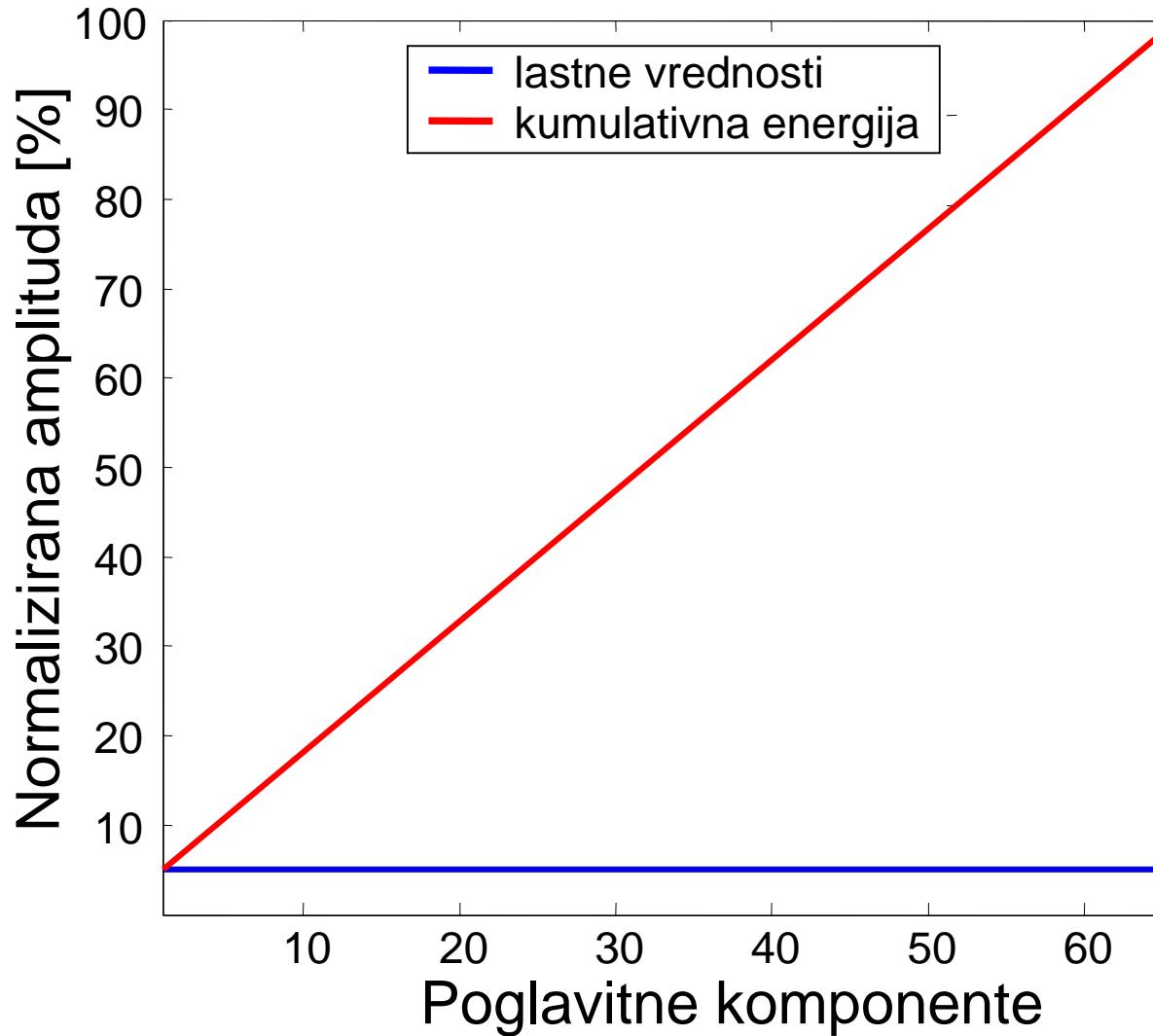
kumulativna energija:

$$E(K) = \frac{\sum_{k=1}^K \lambda_k}{\sum_{m=1}^M \lambda_m}$$

95 % energije je vsebovane v prvih 8 poglavitnih komponentah

99 % energije je vsebovane v prvih 16 poglavitnih komponentah

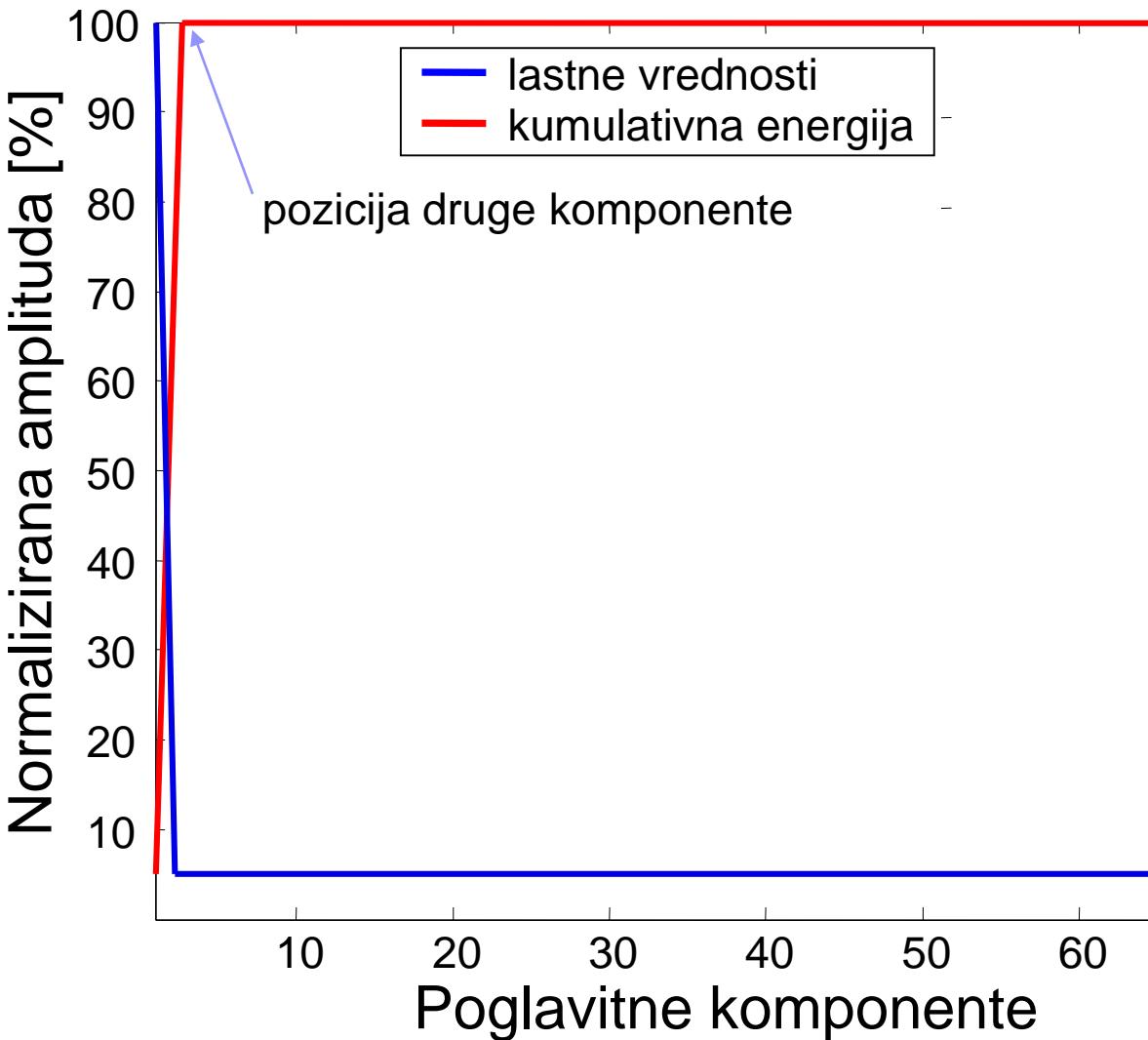
Redukcija dimenzij – najslabši primer za PCA



kumulativna energija:

$$E(K) = \frac{\sum_{k=1}^K \lambda_k}{\sum_{m=1}^M \lambda_m}$$

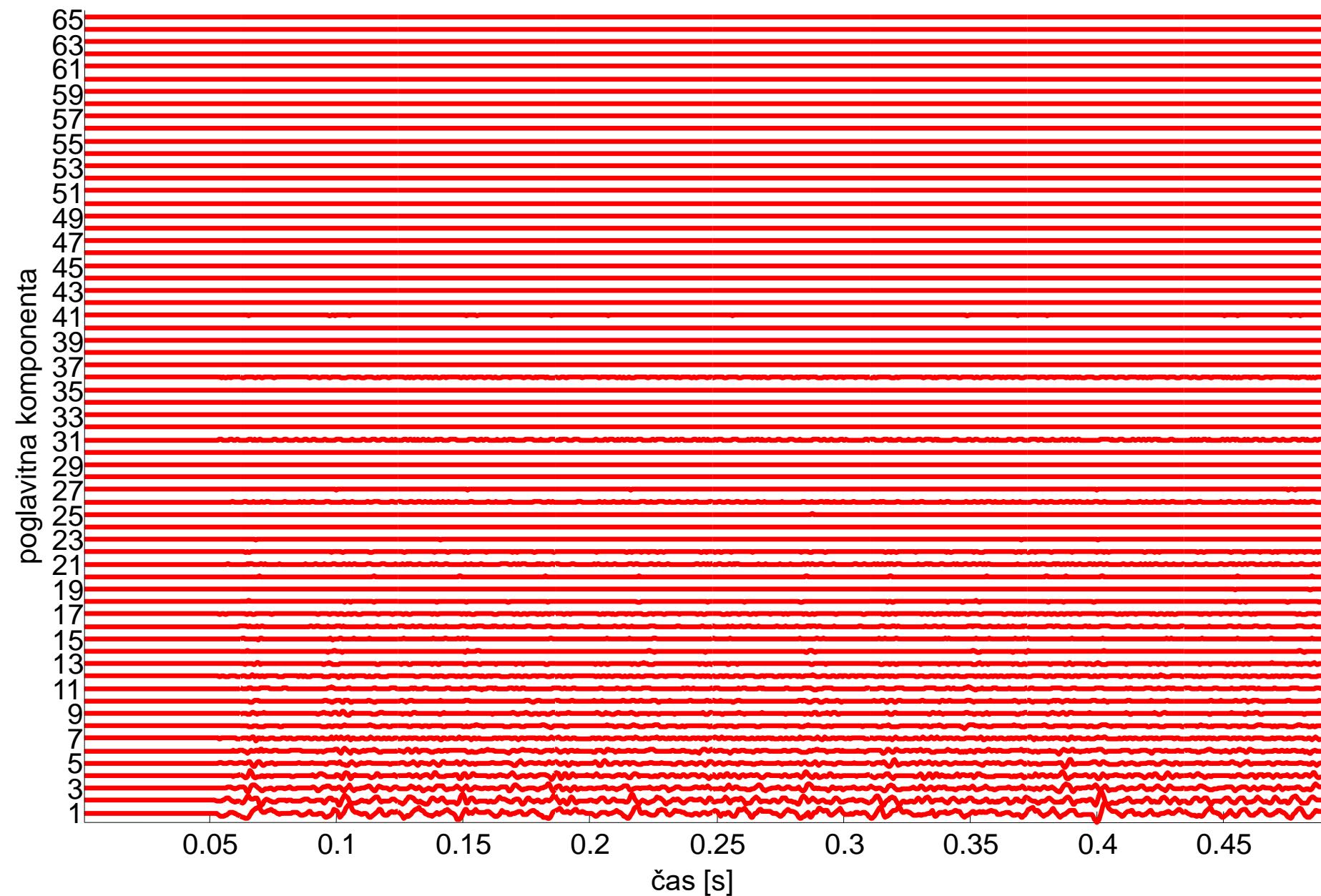
Redukcija dimenzij – najboljši primer za PCA



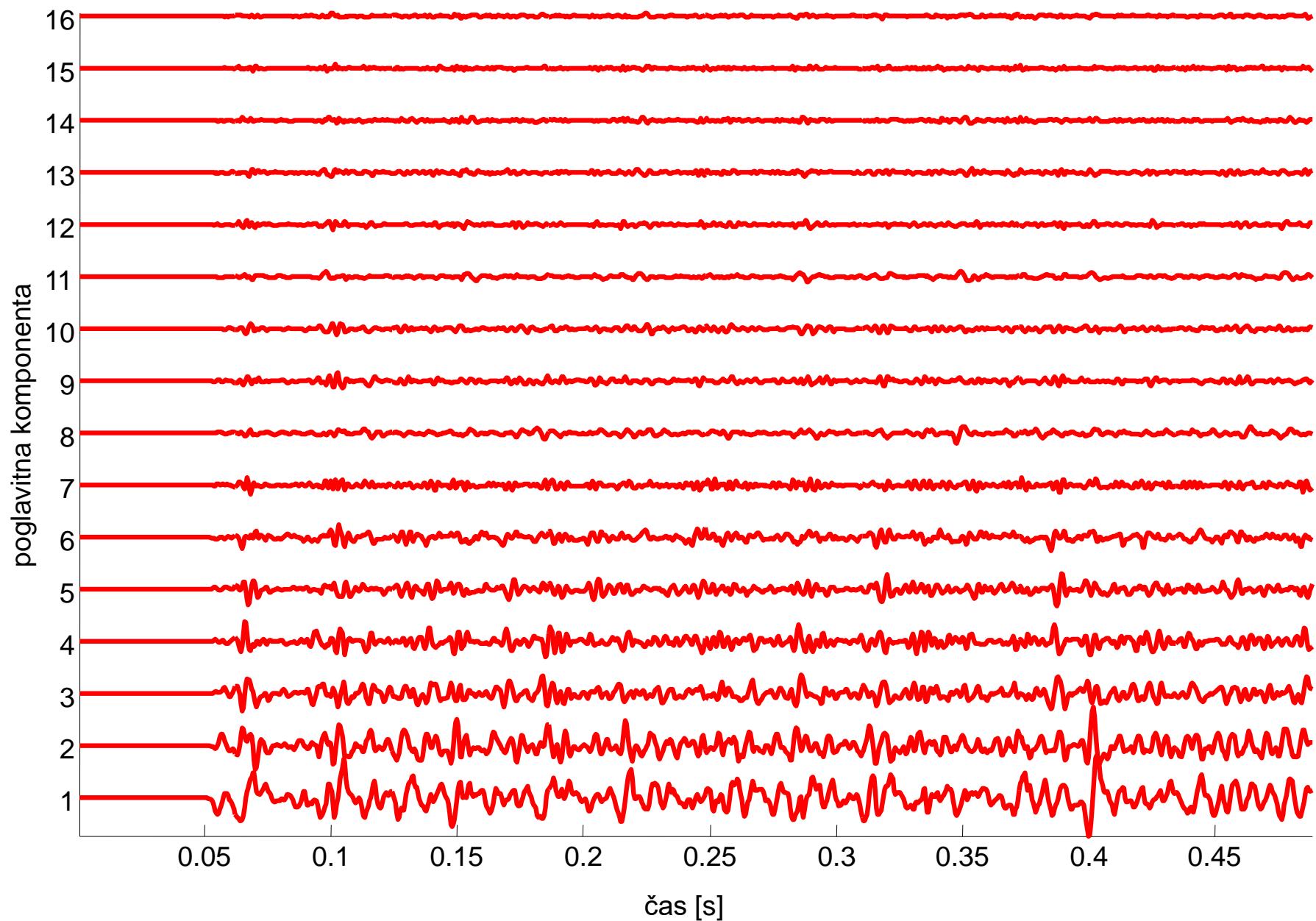
kumulativna energija:

$$E(K) = \frac{\sum_{k=1}^K \lambda_k}{\sum_{m=1}^M \lambda_m}$$

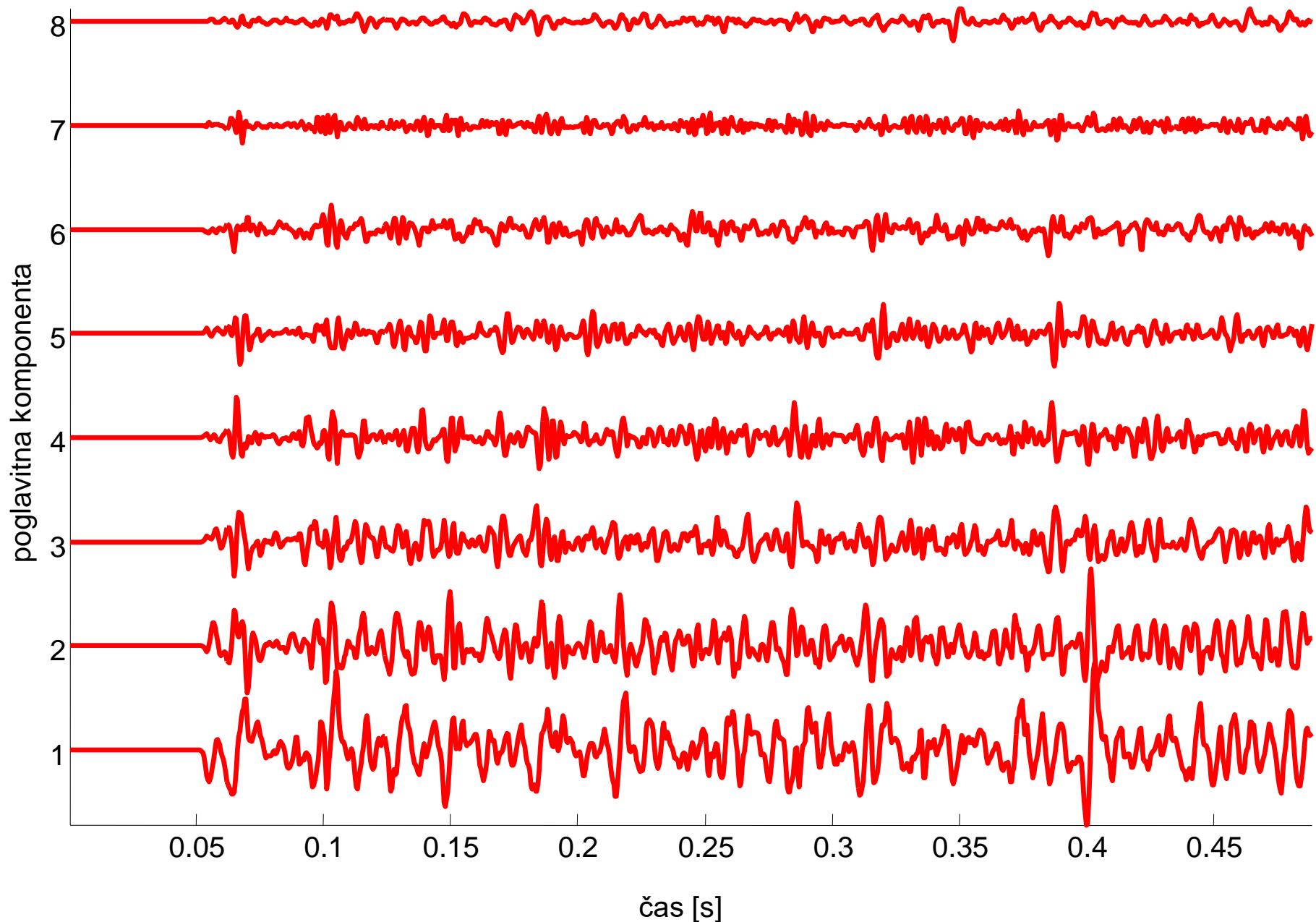
Primer 4: poglavitne komponente $y[n]$



Primer 4: poglavitne komponente $y[n]$



Primer 4: poglavitne komponente $y[n]$



Rekonstrukcija meritev

Po definiciji: $\mathbf{x}[n] = \mathbf{E}\mathbf{y}[n] = \sum_{m=1}^M \mathbf{e}_m y_m[n]$

kjer je \mathbf{e}_m m -ti lastni vektor matrike $\mathbf{C}_{\mathbf{x}}$

Ker je

$$E\{y_m^2[n]\} = E\{\mathbf{e}_m^T \mathbf{x}[n] \mathbf{x}^T[n] \mathbf{e}_m\} = \mathbf{e}_m^T \mathbf{C}_{\mathbf{x}} \mathbf{e}_m = \lambda_m$$

lahko $\mathbf{x}[n]$ aproksimiramo z

$$\hat{\mathbf{x}}[n] = \sum_{k=1}^K \mathbf{e}_k y_k[n]$$

kjer je $K < M$

Redukcija dimenzij: ocena napake

- ocenimo kvadratični pogrešek takšne aproksimacije:

$$E \left\{ \| \mathbf{x}[n] - \hat{\mathbf{x}}[n] \|^2 \right\}$$

kjer $\mathbf{x}[n] = \sum_{k=1}^M \mathbf{e}_k y_k[n]$ in $\hat{\mathbf{x}}[n] = \sum_{k=1}^K \mathbf{e}_k y_k[n]$.

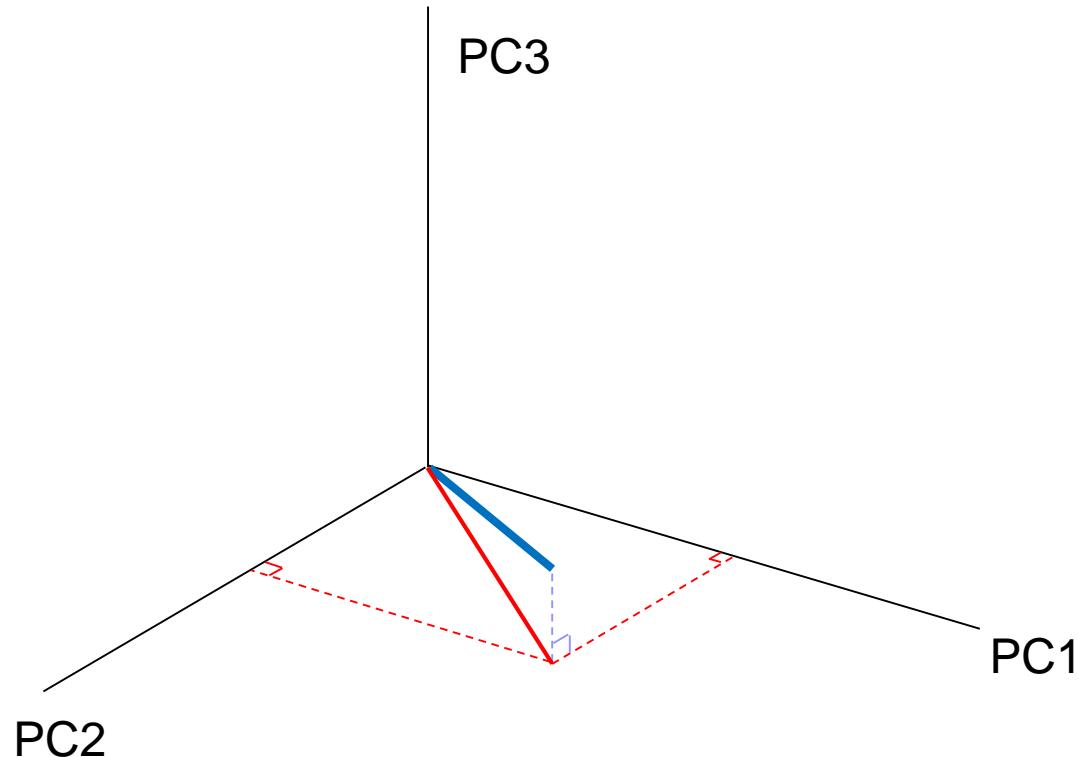
- krajša matematična manipulacija:

$$\begin{aligned} E \left\{ \| \mathbf{x}[n] - \hat{\mathbf{x}}[n] \|^2 \right\} &= E \left\{ \left\| \sum_{m=K+1}^M \mathbf{e}_m y_m[n] \right\|^2 \right\} = E \left\{ \sum_{m=K+1}^M y_m[n] \mathbf{e}_m^T \mathbf{e}_m y_m[n] \right\} \\ &= \sum_{m=K+1}^M E \left\{ y_m^2[n] \right\} = \sum_{m=K+1}^M \lambda_m \end{aligned}$$

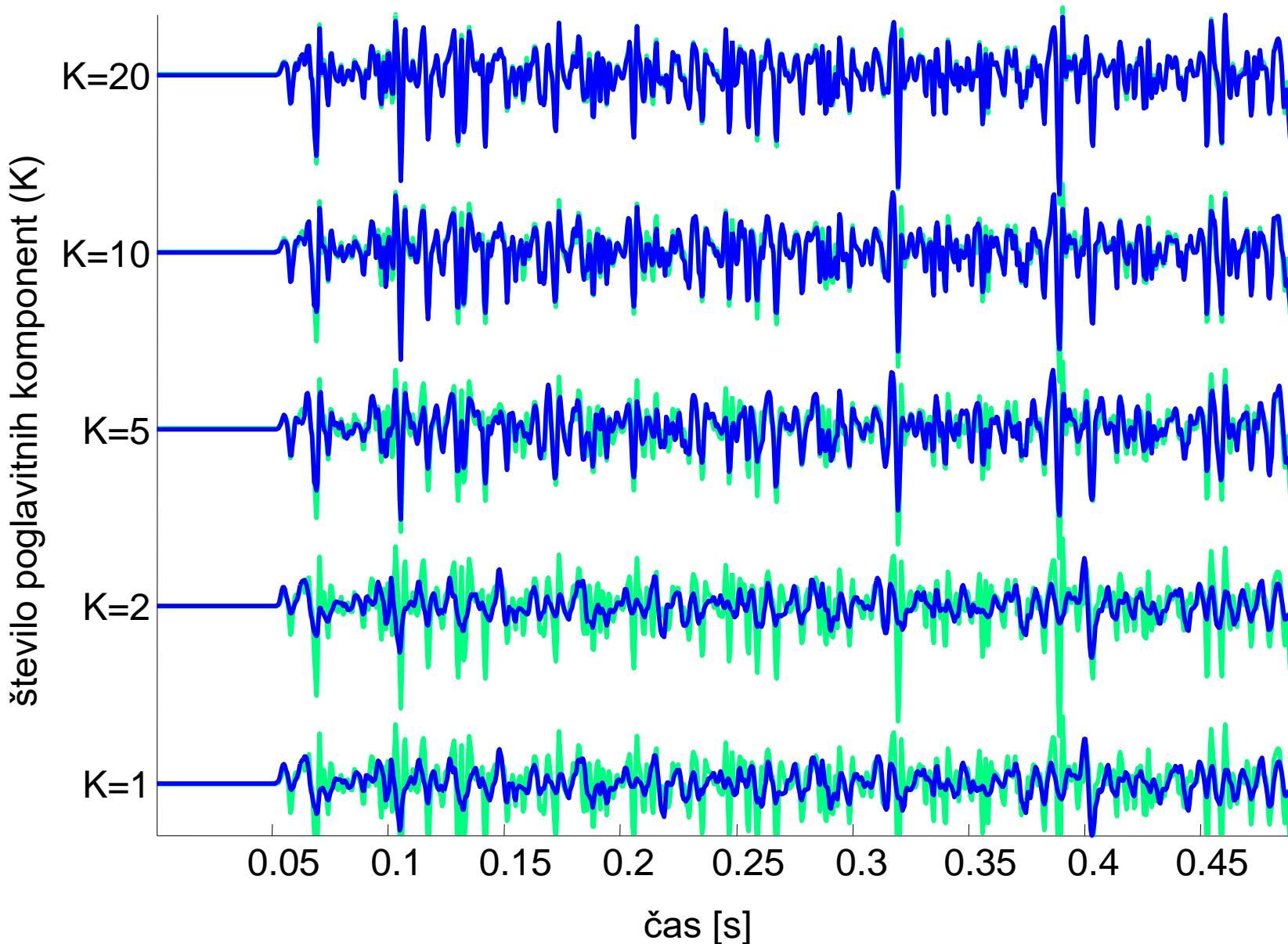
Redukcija dimenzij: ocena napake – grafični zgled

- stara dobra tabla...

... in zapiski



Primer 4: Napaka rekonstrukcije $\hat{\mathbf{x}}[n] = \sum_{k=1}^K \mathbf{e}_k y_k[n]$



Kako uporabiti PCA na sliki?

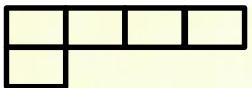




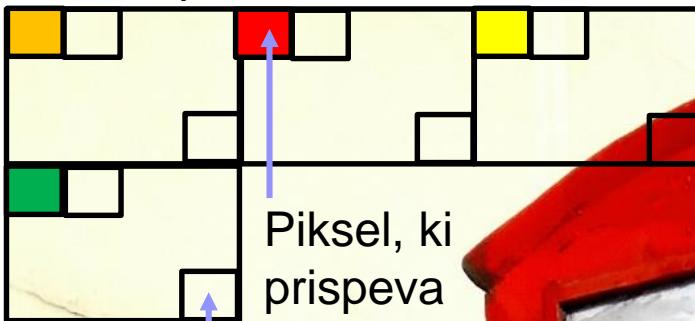
NE BO NAJBOLJE



4x4 = 16 pikslov -> 16 meritev, s toliko vzorci, kot je blokov na sliki



$4 \times 4 = 16$ pikslov -> 16 meritv, s toliko vzorci, kot je blokov na sliki



Piksel, ki
prispeva
k prvi
meritvi

Piksel, ki
prispeva k 16.
meritvi

1. Meritev $x_1[n]$

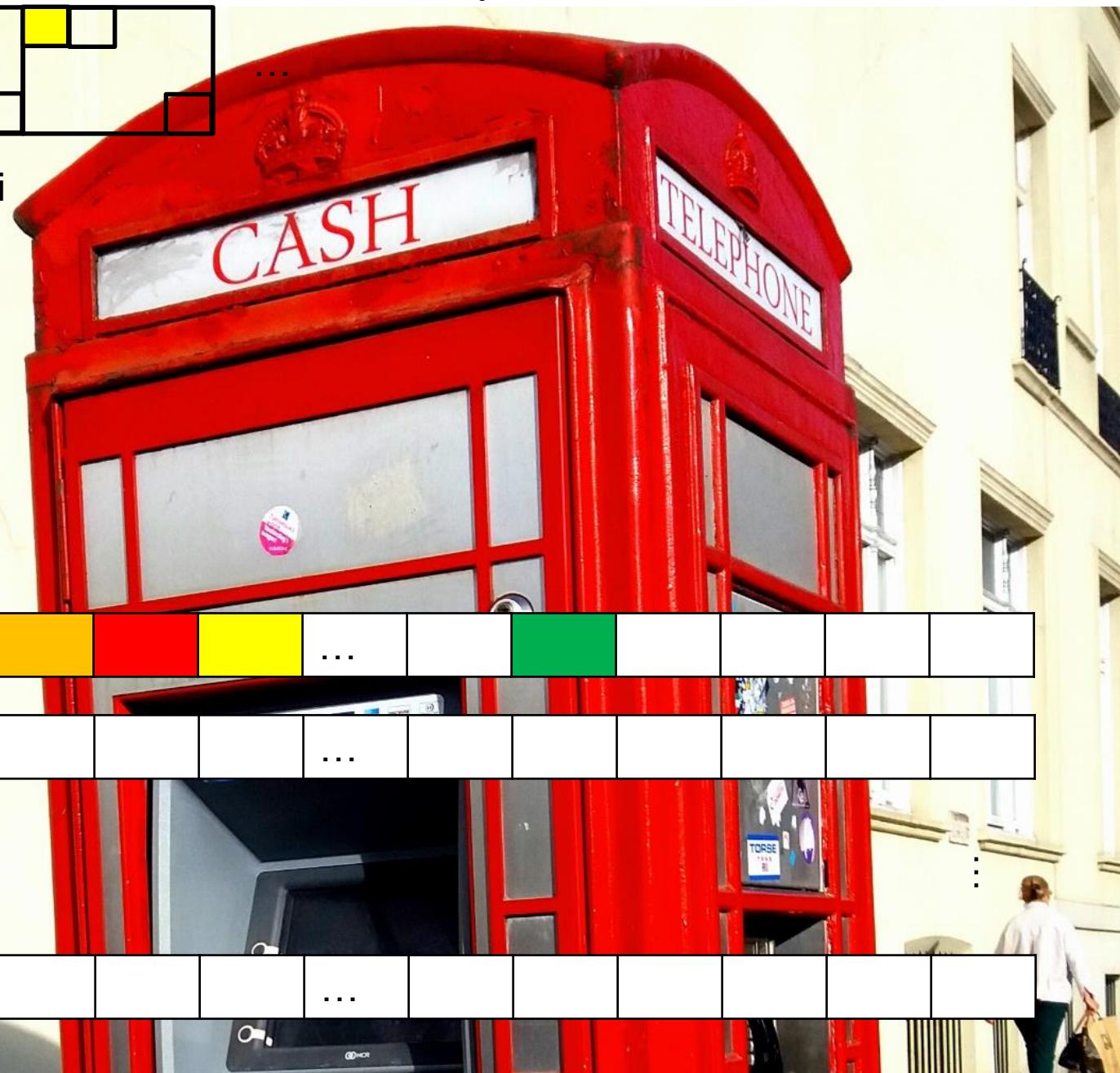


2. Meritev $x_2[n]$

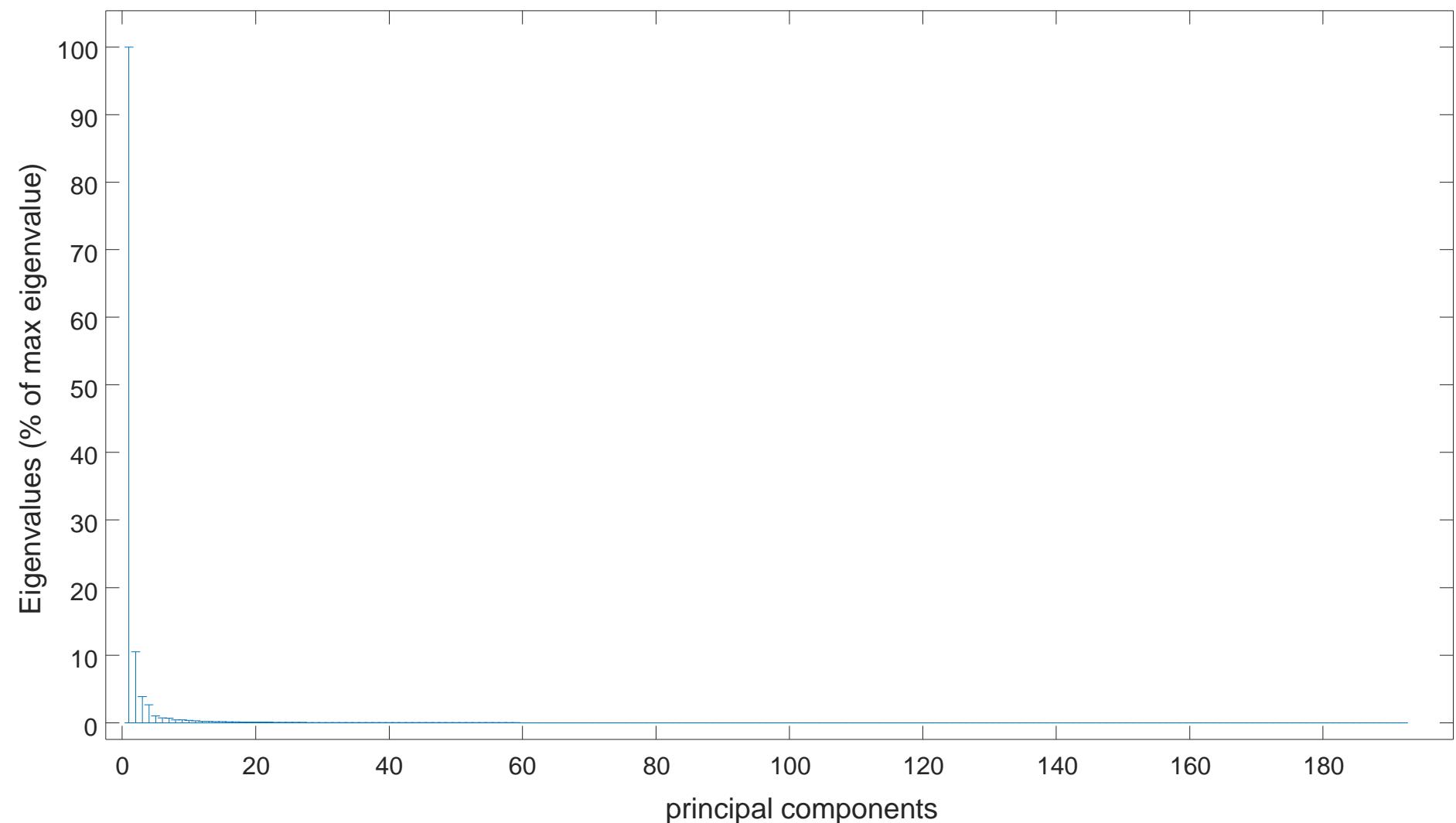


⋮ ⋮

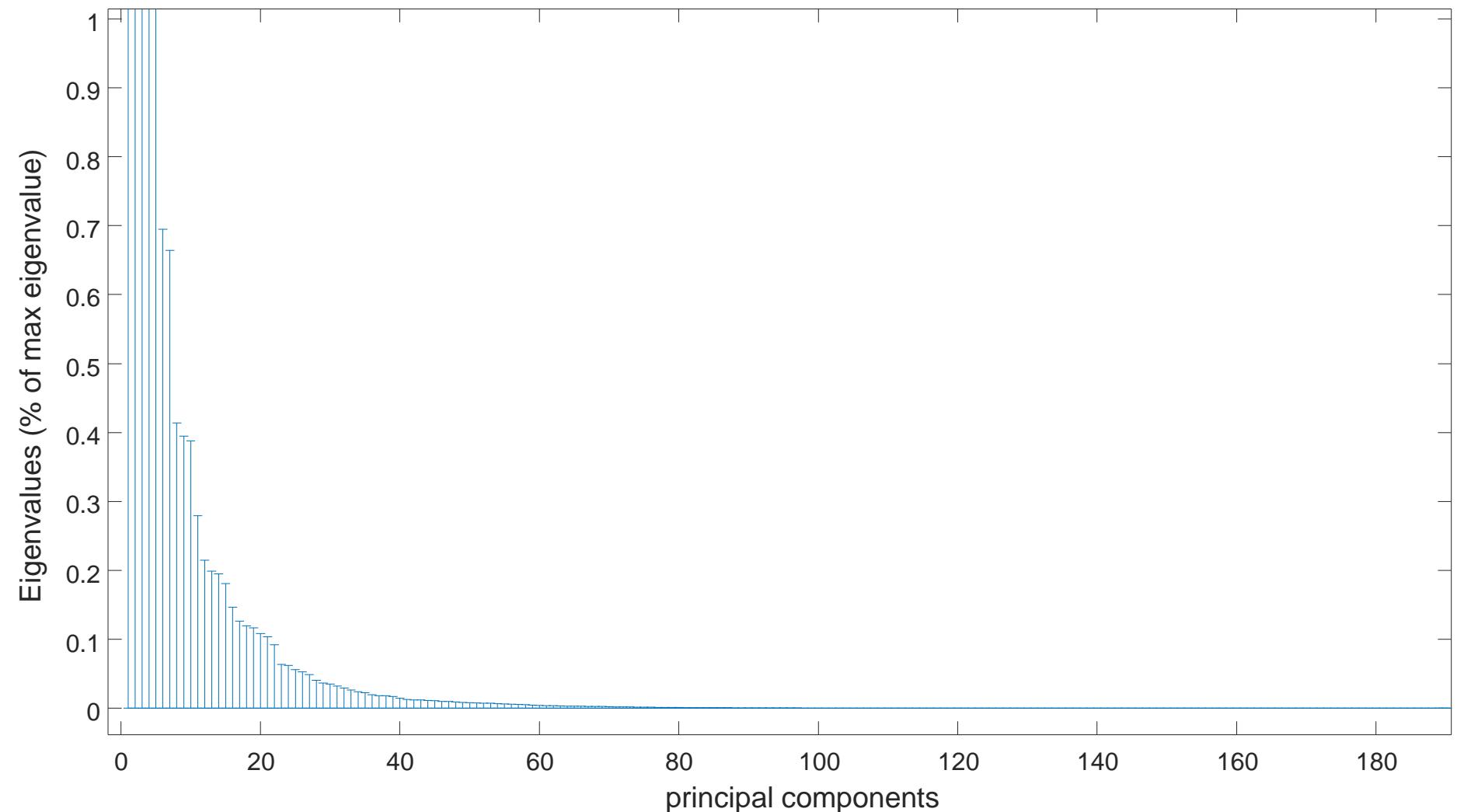
16. Meritev $x_{16}[n]$



Velikost bloka 8×8 ($64^*3 = 192$ PCA komponent)



Velikost bloka 8×8 ($64^*3 = 192$ PCA komponent)



Velikost bloka 8×8 ($64 \times 3 = 192$ PCA komponent)

original image



Image reconstructed by 122 PCs



Image reconstructed by 62 PCs



Image reconstructed by 22 PCs



Velikost bloka 8×8 ($64 \times 3 = 192$ PCA komponent)

Image reconstructed by 8 PCs



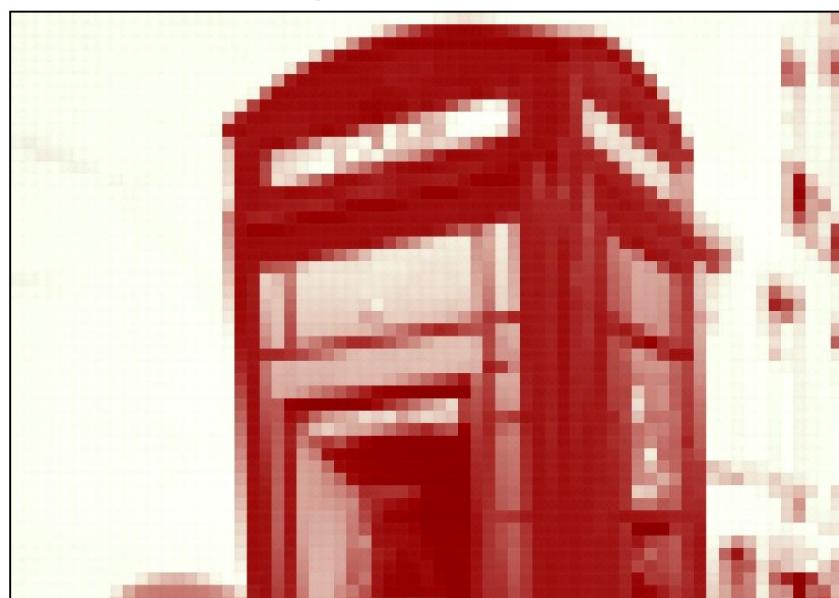
Image reconstructed by 5 PCs



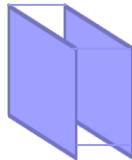
Image reconstructed by 3 PCs



Image reconstructed by 1 PCs



širina_slikovnega_okvirja x višina _slikovnega_okvirja x časovni_okvirji

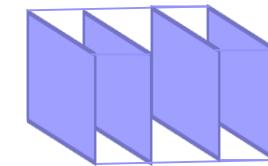


4x4x2

dinamična
scena



čas (v slikovnih okvirjih)



4x4x4

bolj statična
scena

1. PRIMER: 3D koordinate ptic v gibanju jate z 200 pticami, ko ta jata lovi jato mušic:

Meritev 1: x koordinata ptice 1 v času

Meritev 2: y koordinata ptice 1 v času

Meritev 3: z koordinata ptice 1 v času

Meritev 4: x koordinata ptice 2 v času

...

Meritev 600:

ali 3 x PCA ločeno po x, y in z koordinatah

Meritev 1: x koordinata ptice 1 v času

Meritev 2: x koordinata ptice 2 v času

Meritev 3: x koordinata ptice 3 v času

...

Meritev 200: x koordinata ptice 3 v času

potem pa nov PCA po:

Meritev 1: y koordinata ptice 1 v času

Meritev 2: y koordinata ptice 2 v času

Meritev 3: y koordinata ptice 3 v času

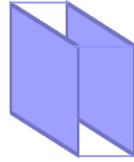
...

Meritev 200: y koordinata ptice 3 v času

2. PRIMER: Jutranja (ob 6. uri) temperatura krajev z > 5000 prebivalci v Sloveniji skozi celo leto.

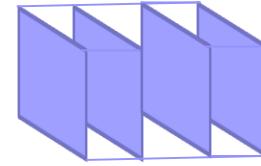
geografska in časovna lokalnost:

Geografska širina x geografska višina x čas (dnevi)



$4 \times 4 \times 2$

dinamična scena



$4 \times 4 \times 4$

bolj statična scena

matlab/octave: statistics toolbox

■ pca() method

eigenvalues of C_x
the eigenvalues of the covariance matrix of X

matrix of measurements
(each column corresponds to different measurement)

```
[E, Y, Lambda, tsquare] = pca (X);
```

principal components e_i
(the 1st pc is in the 1st column, the 2nd pc in the 2nd column, etc.)

matrix of projections y_i
(projection to the 1st pc is in the 1st column, projection to the 2nd column in the 2nd column, etc.)

Hotelling's T2 statistic
a measure of the multivariate distance of each observation from the center of the data set

matlab/octave: statistics toolbox

```
>> for k = 1:size(X,2) % meritve morajo imeti ničelno  
    % povprečje  
    X(:,k) = X(:,k) - mean(X(:,k));  
end
```



```
>> [E,Y,Lambda,tsquare] = pca(X); % PCA of X
```



```
>> X = Y * E'; % rekonstrukcija originalnih meritv
```



```
>> k = 1:K; rX = Y(:,k) * E(:,k)'; % rekonstrukcija  
    % iz zmanjšanega  
    % števila poglavitnih  
    % komponent
```

$$\hat{\mathbf{x}}[n] = \sum_{k=1}^K \mathbf{e}_k y_k[n]$$

Povzetek lastnosti analize PCA

- Linearnost.
 - zamenjava baze prostora je linearна operacija
- Potrebujemo samo ocene povprečja in variance:
 - $\mathbf{x}[n]$ morajo imeti ničelno povprečje,
 - porazdelitev meritev $\mathbf{x}[n]$ ni pomembna,
 - mešalni model (tvorba meritev) $\mathbf{x}[n]$ ni pomemben,
- **Informacija = energija (varianca): velike variance v $\mathbf{x}[n]$ so nosilke glavne dinamike**
- Smeri poglavitnih komponent so ortogonalne.
- Tudi poglavitne komponente so ortogonalne

Študijski material

- Jon Shlens: A TUTORIAL ON PRINCIPAL COMPONENT ANALYSIS: Derivation, Discussion and Singular Value Decomposition
http://www.cs.princeton.edu/picasso/mats/PCA-Tutorial-Intuition_jp.pdf
(nazorno učno gradivo z veliko zgledi)
- Smith, L.I. (2002), “A tutorial on Principal Components Analysis,
http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf
(zelo priljubljeno učno gradivo z neformalnimi razlagami)
- A. Daffertshofer et al., PCA in studying coordination and variability: a tutorial, in *Clinical Biomechanics*, vol. 19, pp. 415–428, 2004
(Nazoren članek z lepimi prikazi rezultatov analize PCA)

PRIMERI UPORABE ANALIZE PCA

A. Daffertshofer, C. J. C. Lamothe, O. G. Meijer, P. J. Beek: **PCA in studying coordination and variability: a tutorial**, Clin Biomech (Bristol, Avon), 2004 May;19(4):415-28. DOI: [10.1016/j.clinbiomech.2004.01.005](https://doi.org/10.1016/j.clinbiomech.2004.01.005)

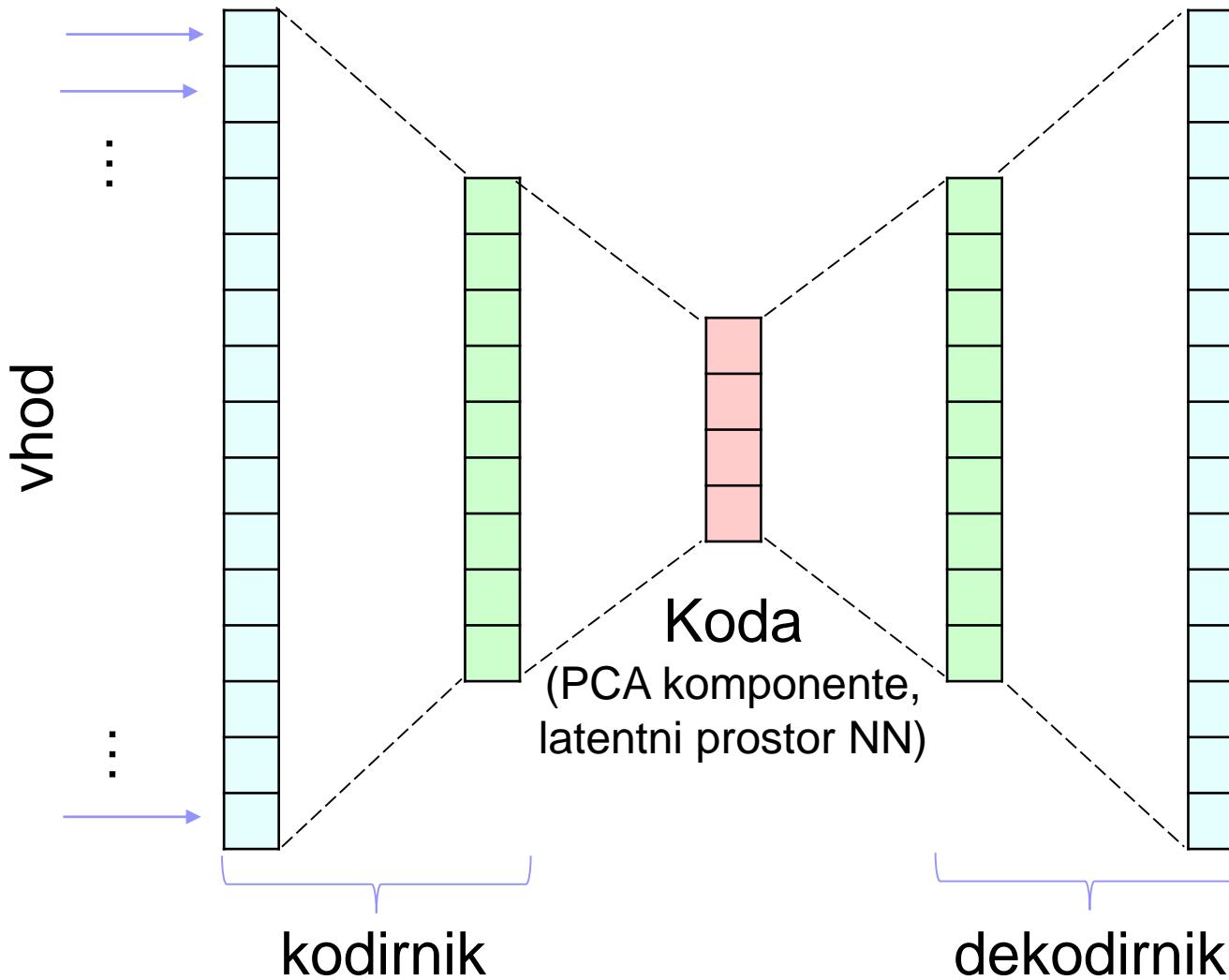
B. J. Stetter, M. Herzog, F. Möhler, S. Sell, T. Stein: **Modularity in Motor Control: Similarities in Kinematic Synergies Across Varying Locomotion Tasks**, Front Sports Act Living 2020 Nov 13, DOI: [10.3389/fspor.2020.596063](https://doi.org/10.3389/fspor.2020.596063)

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7739575/>

R. Z. U. Rehman, Y. Zhou, S. Del Din, L. Alcock, C. Hansen, Y. Guan, T. Hortobágyi, W. Maetzler, L. Rochester, C. J. C. Lamothe: **Gait Analysis with Wearables Can Accurately Classify Fallers from Non-Fallers: A Step toward Better Management of Neurological Disorders**, Sensors (Basel) 2020 Dec 7;20(23):6992
DOI: [10.3390/s20236992](https://doi.org/10.3390/s20236992) , <https://pubmed.ncbi.nlm.nih.gov/33297395/>

Kodirnik in dekodirnik

(angl. autoencoder)



Brez dodatnih pravil o generiranih kodi bo avtoenkoder najverjetneje konvergiral k rešitvi, ki je podobna rezultatu PCA