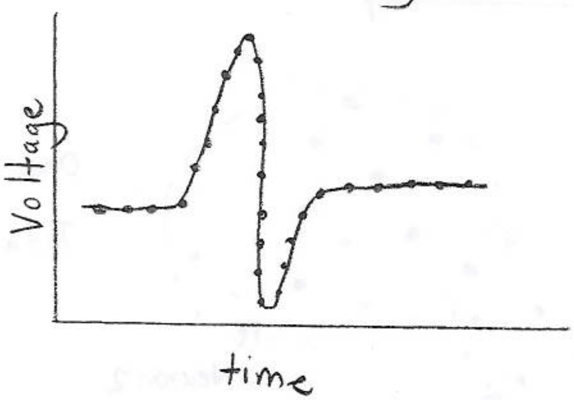
Dimensionality Reduction

(Part 1)

A) Motivation

A.1) Spike sorting



$\underline{x}_n \in \mathbb{R}^{31}$ is one spike snippet $(n=1,\ldots,N)$

- How do the snippets differ?

  It may be that:

$$\underline{x}_n = \alpha_n \begin{bmatrix} \\ \underline{v} \\ \\ \end{bmatrix} + \beta_n \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \qquad (1)$$

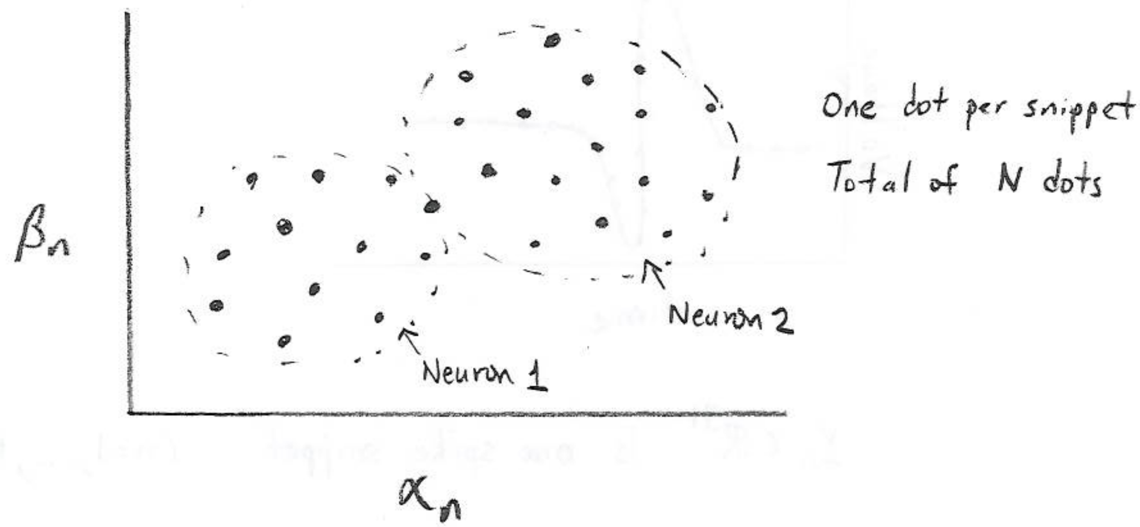
where $\underline{v}$ is a canonical waveform

α_n is the amplitude

β_n is constant offset

If this were true, then across the N snippets, there would only be two <u>degrees of freedom</u> of variability, corresponding to α_n and β_n.

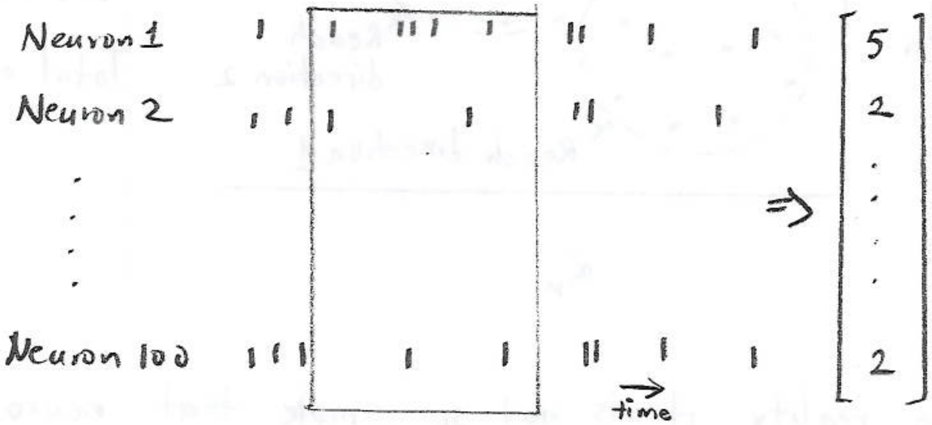
The data points would therefore live on a subspace of the data space whose <u>intrinsic dimensionality</u> is two.



In reality, we don't know that $\underline{v}$ and $\underline{1}$ in (1) are the important vectors, nor how many vectors there should be.

The methods we will soon learn will allow us to identify these vectors and how many there should be from the data $\underline{x}_1, \ldots, \underline{x}_N$.

A.2) Visualization of high-dimensional neural activity



spike count vector $\underline{X}_n \in \mathbb{R}^{100}$

- How do different spike count vectors differ?

It may be that:

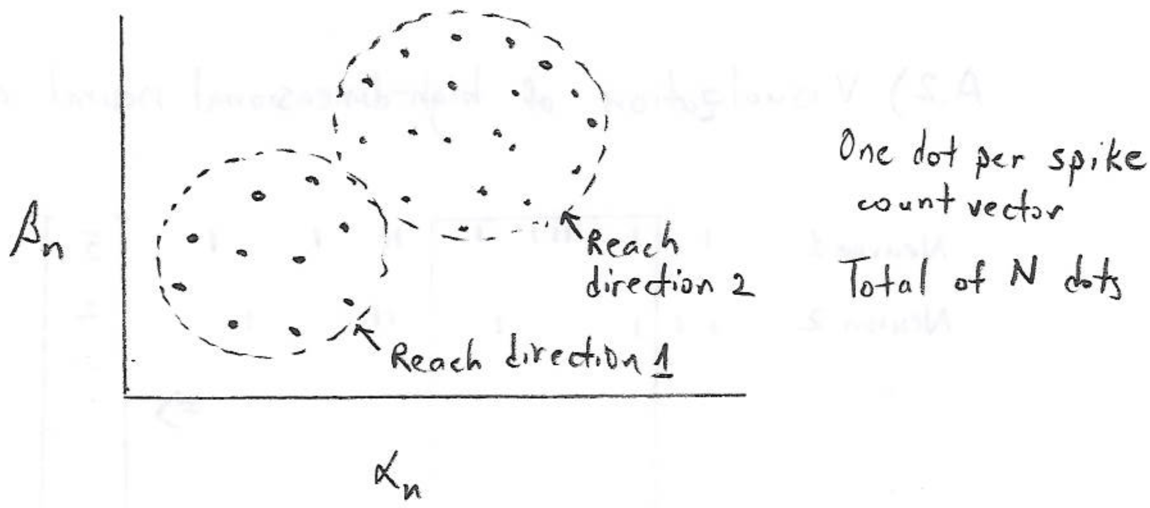
$$\underline{X}_n = \alpha_n \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ \hline 0 \\ 0 \\ 0 \\ \vdots \end{bmatrix} + \beta_n \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ \hline 1 \\ 1 \\ 1 \\ \vdots \end{bmatrix} \qquad (2)$$

where α_n describes the activity of neurons $1, \ldots, 50$

β_n describes the activity of neurons $51, \ldots, 100$

If this were true, the there would be two degrees of freedom of variability (α_n and β_n).

Intrinsic dimensionality would be two.



In reality, it is not so simple that neurons can be grouped as shown in (2).

The methods we will soon learn will allow us to identify the "groupings" and how many there should be from the data $\underline{x}_1, \dots, \underline{x}_N$.

B) Principal Components Analysis (PCA)

Data set $\underline{x}_n \in \mathbb{R}^D$, $n=1,\dots,N$

Goal: Project data into a space with dimensionality $M < D$ while maximizing variance of projected data.

Let S be the sample covariance

$$S = \frac{1}{N}\sum_{n=1}^{N}(\underline{x}_n - \underline{\mu})(\underline{x}_n - \underline{\mu})^T,$$

where $\underline{\mu} = \frac{1}{N}\sum_{n=1}^{N}\underline{x}_n$

B.1) Diagonalization

(also known as "eigendecomposition")

Any covariance matrix S can be expressed as:

$$S = U \Lambda U^T,$$

where the columns of U are orthonormal and Λ is diagonal.

$$U = \begin{bmatrix} | & | & & | \\ \underline{u}_1 & \underline{u}_2 & \cdots & \underline{u}_D \\ | & | & & | \end{bmatrix} \qquad \Lambda = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_D \end{bmatrix}$$

($D \times D$ matrix) ($D \times D$ matrix)

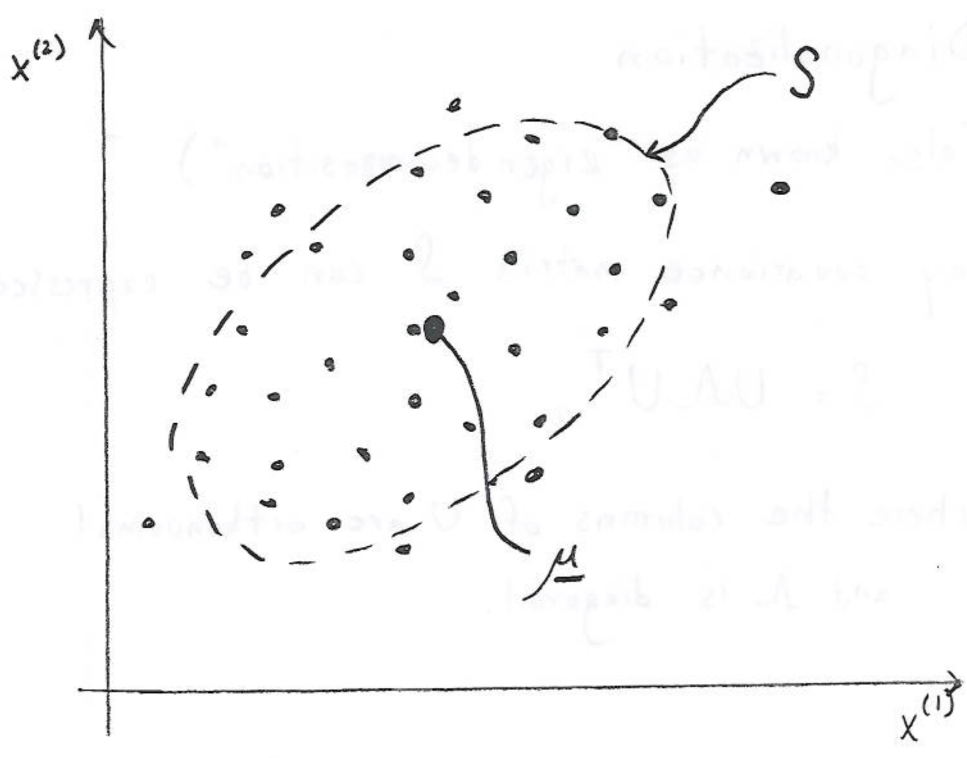
$\underline{u}_i$ is i^{th} eigenvector

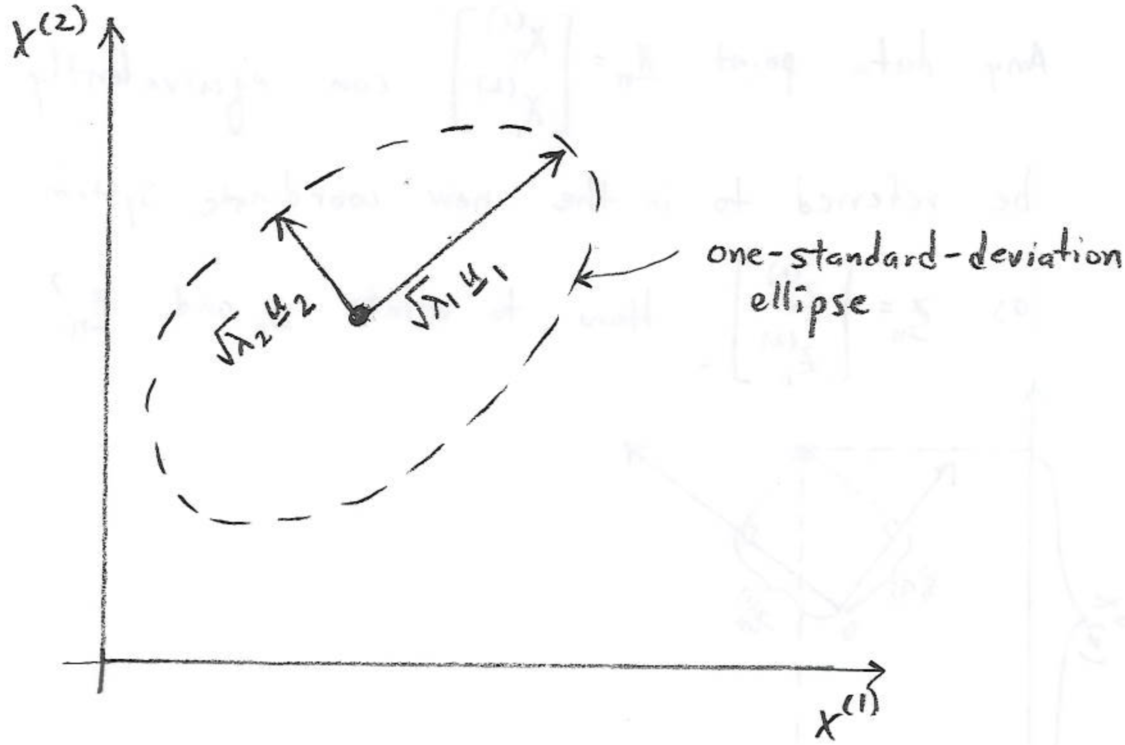
λ_i is i^{th} eigenvalue

$$\underline{u}_i^T \underline{u}_j = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{else} \end{cases}$$

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_D \geq 0$$

Note that $\underline{u}_1, \ldots, \underline{u}_D$ form an orthonormal basis for $\mathbb{R}^D$. In other words, any point in $\mathbb{R}^D$ can be expressed as a linear combination of $\underline{u}_1, \ldots, \underline{u}_D$





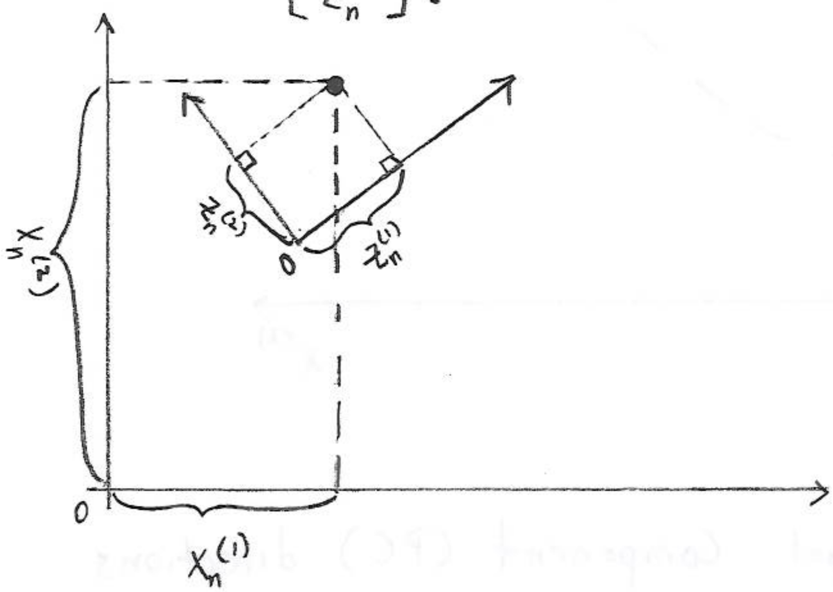
B.2) Principal Component (PC) directions

- The 1st PC direction captures the greatest data variance $\Rightarrow \underline{u}_1$
- The 2nd PC direction captures the 2nd most variance <u>and</u> is orthogonal to the 1st PC direction $\Rightarrow \underline{u}_2$

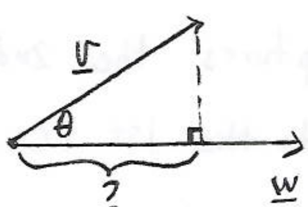
$\vdots$

The PC directions define a new set of coordinate axes.

Any data point $\underline{x}_n = \begin{bmatrix} x_n^{(1)} \\ x_n^{(2)} \end{bmatrix}$ can equivalently be referred to in the new coordinate system as $\underline{z}_n = \begin{bmatrix} z_n^{(1)} \\ z_n^{(2)} \end{bmatrix}$. How to relate $\underline{x}_n$ and $\underline{z}_n$?



B.3) Projections



What is the projection of $\underline{v}$ onto $\underline{w}$?

$$\|\underline{v}\| \cos\theta = \frac{\|\underline{v}\| \, \|\underline{w}\| \cos\theta}{\|\underline{w}\|} = \frac{\underline{v}^T \underline{w}}{\|\underline{w}\|}$$

B.4) PCA: Projecting from high-dimensional space ($\underline{x}$) down into low-dimensional space ($\underline{z}$)

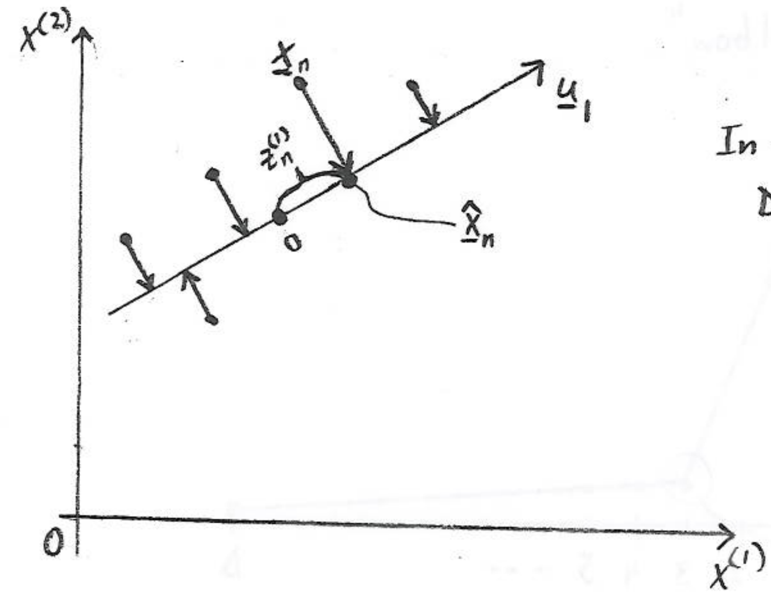
new coordinates

"PC score"

$$\boxed{z^{(i)} = (\underline{x} - \underline{\mu})^T \underline{u}_i} \qquad i = 1, \ldots, M \quad (M < D) \tag{3}$$

In words: center high-dimensional data, then project onto axis defined by $\underline{u}_i$.

Note that $\|\underline{u}_i\| = 1$ for all i.



In this illustration, $D=2$ and $M=1$.

In low-dimensional coordinates, projected point is $\underline{z}_n$.
What is the same point in high-dimensional coordinates?

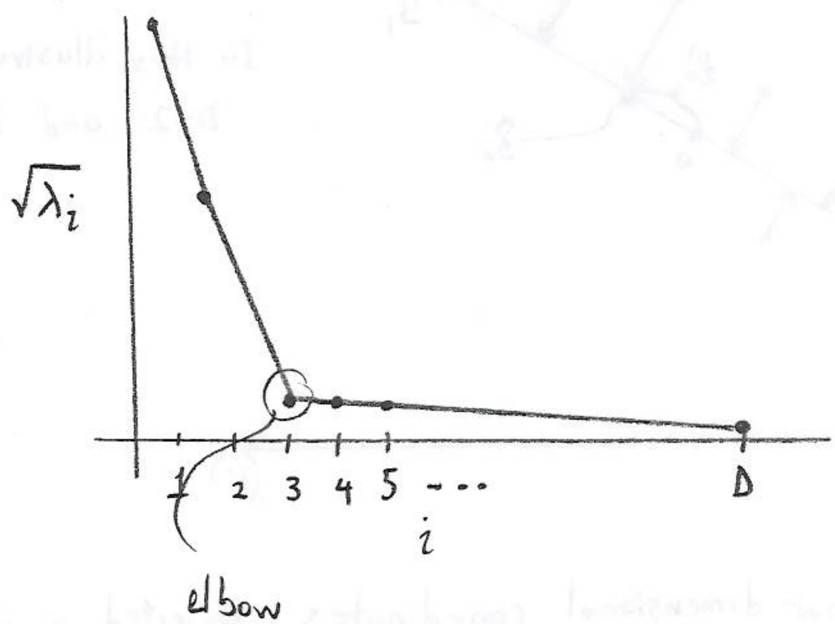
$$\boxed{\hat{\underline{x}}_n = \sum_{i=1}^{M} z_n^{(i)} \underline{u}_i + \underline{\mu}} \tag{4}$$

This is referred to as "projecting a low-dimensional point back out into high-dimensional space."

What is $\sum_{i=1}^{D} z_n^{(i)} \underline{u}_i + \underline{\mu}$?

B.5) How to choose M, the dimensionality of low-dimensional space?

Plot eigenvalue spectrum of S and look for "elbow".



M is typically taken to be the number of dominant eigenvalues above the elbow.

In this case, $M=2$.

In practice, there is often not a clear cut elbow.

This motivates the probabilistic formulation of PCA, where we can use cross-validated likelihoods to determine M.

$$\frac{\sum_{i=1}^{M} \lambda_i}{\sum_{i=1}^{D} \lambda_i}$$ is the fraction (or percentage) of variance explained by the first M principal components.

B.6) Applying PCA to motivating examples

- Spike sorting

In (4), $z^{(i)}$ is the i^{th} feature value

(recall: without PCA, we would choose ad-hoc features like waveform max, min, width, etc.)

$\underline{u}_i$ is the i^{th} eigenvector waveform

$\underline{\mu}$ is mean of all recorded waveforms

Each spike is represented as a linear combination of eigenvector waveforms (weighted by feature value), plus $\underline{\mu}$.

- Visualization of high-dimensional neural activity

In (4), $z^{(i)}$ indicates how strongly the i^{th} "group" of neurons is firing

$\underline{u}_i$ indicates to what extent each neuron belongs to the i^{th} "group"

$\underline{\mu}$ is the mean spike count for each neuron.

<u>Note</u>: We have formulated PCA in terms of <u>maximizing</u> the variance of the projected data. Equivalently, we could have formulated PCA in terms of <u>minimizing</u> the projection error.

B.7) Summary of PCA

Data set $\underline{x}_n \in \mathbb{R}^D$, $n = 1, \ldots, N$

1) Find the sample covariance S and sample mean $\underline{\mu}$:

$$S = \frac{1}{N} \sum_{n=1}^{N} (\underline{x}_n - \underline{\mu})(\underline{x}_n - \underline{\mu})^T$$

$$\underline{\mu} = \frac{1}{N} \sum_{n=1}^{N} \underline{x}_n$$

2) Diagonalize S

$$S = U \Lambda U^T$$ (do this in Matlab using 'eig')

Let $U_M = \begin{bmatrix} | & | & & | \\ \underline{u}_1 & \underline{u}_2 & \cdots & \underline{u}_M \\ | & | & & | \end{bmatrix}$ ($D \times M$ matrix)

where $\underline{u}_i$ is the eigenvector corresponding to the i^{th} largest eigenvalue.

3) <u>PC directions</u> are columns of U_M.

4) PC scores are:

$$\underline{z}_n = U_M^T (\underline{x}_n - \underline{\mu}) \qquad \underline{z}_n \in \mathbb{R}^{M \times 1} \tag{5}$$

$\underline{z}_n$ is the low-dimensional projection of $\underline{x}_n$.

5) Location of projected point in high-dimensional space:

$$\begin{aligned} \hat{\underline{x}}_n &= U_M \underline{z}_n + \underline{\mu} \\ &= U_M U_M^T (\underline{x}_n - \underline{\mu}) + \underline{\mu} \end{aligned} \tag{6}$$

Note that the PC directions are only unique up to a sign difference. In other words, the i^{th} PC direction can be $\underline{u}_i$ or $-\underline{u}_i$. This will determine the sign of the i^{th} PC score (i.e., the i^{th} element of $\underline{z}$).