Principal Component Analysis

Acquisition and Analysis of Biosignals

Eero Lehtonen, 4.10.2022

eero.lennart.lehtonen@utu.fi

UNIVERSITY OF TURKU

Outline

- Motivation
- Demo
- Theory
- Demo revisited

Why Principal Component Analysis (PCA)?

Principal Component Analysis (PCA) is a dimensionality reduction method which can be used for data analysis and modeling. It is

- Computationally efficient (fast)
- Unique (in a sense)
  - Variances of the resulting *principle components* are unambiguous
  - In practice the resulting signals are often unique up to polarity

UNIVERSITY OF TURKU

Why Principal Component Analysis (PCA)?

Principal Component Analysis (PCA) is a dimensionality reduction method which can be used for data analysis and modeling. It is

- Computationally efficient (fast)
- Unique (in a sense)
    - Variances of the resulting *principle components* are unambiguous
    - In practice the resulting signals are often unique up to polarity

It can be used

- to simplify data analysis & visualization (incl. clustering!)
- to find underlying causes in the data
- to compress data
- as a first step in more complex signal processing methods (for example, Independent Component Analysis)

Why Principal Component Analysis (PCA)?

Principal Component Analysis (PCA) is a dimensionality reduction method which can be used for data analysis and modeling. It is

- Computationally efficient (fast)
- Unique (in a sense)
  - Variances of the resulting *principle components* are unambiguous
  - In practice the resulting signals are often unique up to polarity

It can be used

- to simplify data analysis & visualization (incl. clustering!)
- to find underlying causes in the data
- to compress data
- as a first step in more complex signal processing methods (for example, Independent Component Analysis)

PCA should always be considered when exploring and analyzing multidimensional data!

Why Principal Component Analysis (PCA)?

Principal Component Analysis (PCA) is a dimensionality reduction method which can be used for data analysis and modeling. It is

- Computationally efficient (fast)
- Unique (in a sense)
  - Variances of the resulting *principle components* are unambiguous
  - In practice the resulting signals are often unique up to polarity
    - PCA does not know ”which side is up”

It can be used

- to simplify data analysis & visualization (incl. clustering!)
- to find underlying causes in the data
- to compress data
- as a first step in more complex signal processing methods (for example, Independent Component Analysis)

PCA should always be considered when exploring and analyzing multidimensional data!

UNIVERSITY OF TURKU

History (from wikipedia)

PCA was invented in 1901 by Karl Pearson, as an analogue of the principal axis theorem in mechanics; it was later independently developed and named by Harold Hotelling in the 1930s. Depending on the field of application, it is also named the discrete Karhunen–Loève transform (KLT) in signal processing, the Hotelling transform in multivariate quality control, proper orthogonal decomposition (POD) in mechanical engineering, singular value decomposition (SVD) of X (Golub and Van Loan, 1983), eigenvalue decomposition (EVD) of X^TX in linear algebra, factor analysis, Eckart–Young theorem (Harman, 1960), or empirical orthogonal functions (EOF) in meteorological science, empirical eigenfunction decomposition (Sirovich, 1987), empirical component analysis (Lorenz, 1956), quasiharmonic modes (Brooks et al., 1988), spectral decomposition in noise and vibration, and empirical modal analysis in structural dynamics.

UNIVERSITY OF TURKU

History (from wikipedia)

PCA

is also named the discrete Karhunen–Loève transform the Hotelling transform proper orthogonal decomposition (POD) singular value decomposition eigenvalue decomposition (EVD) of $X^T X$ factor analysis, Eckart–Young theorem empirical orthogonal functions (EOF) empirical eigenfunction decomposition empirical component analysis quasiharmonic spectral decomposition empirical modal analysis

UNIVERSITY OF TURKU

Goals of this lecture

- To understand *why* certain signal processing steps are taken in PCA
- To understand *how* to use PCA for signal processing

UNIVERSITY OF TURKU

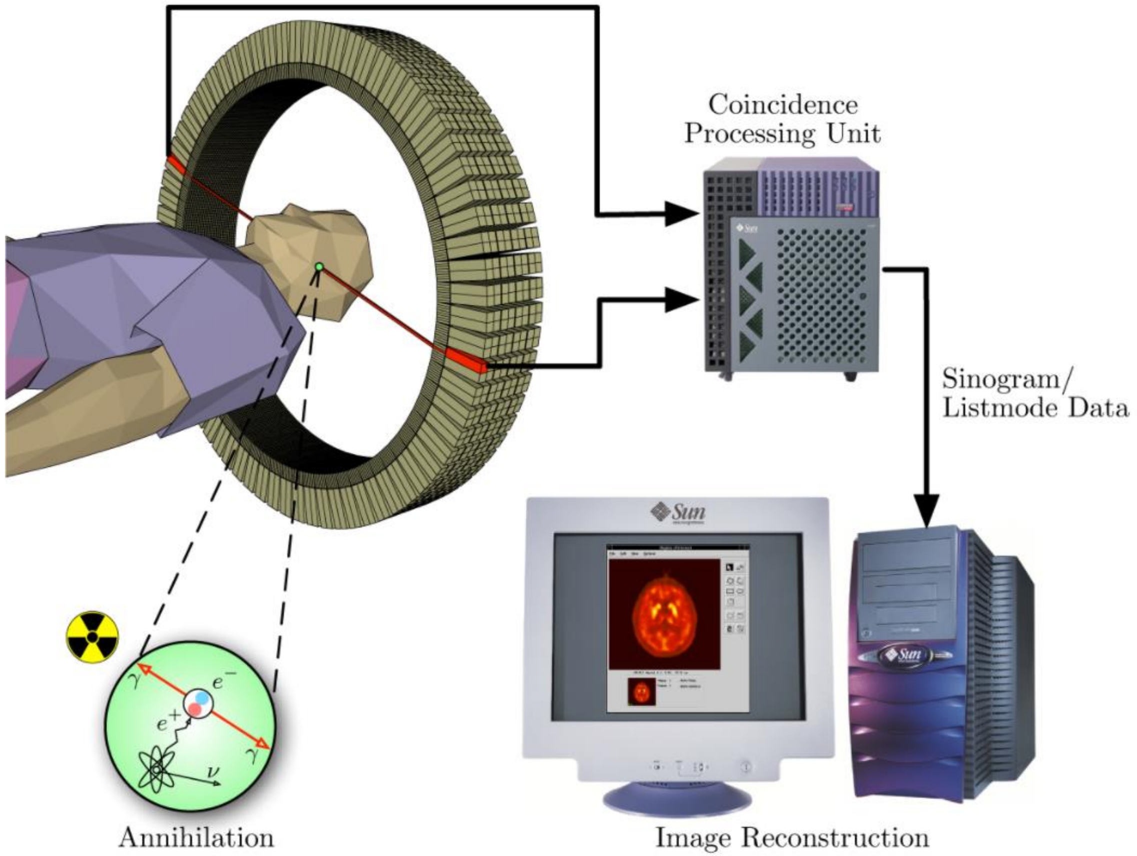
Goals of this lecture

- To understand *why* certain signal processing steps are taken in PCA
- To understand *how* to use PCA for signal processing

- Smaller probability of making mistakes!
    - Easier to catch mistakes and to debug your code!
- Easier to apply PCA in your own applications!

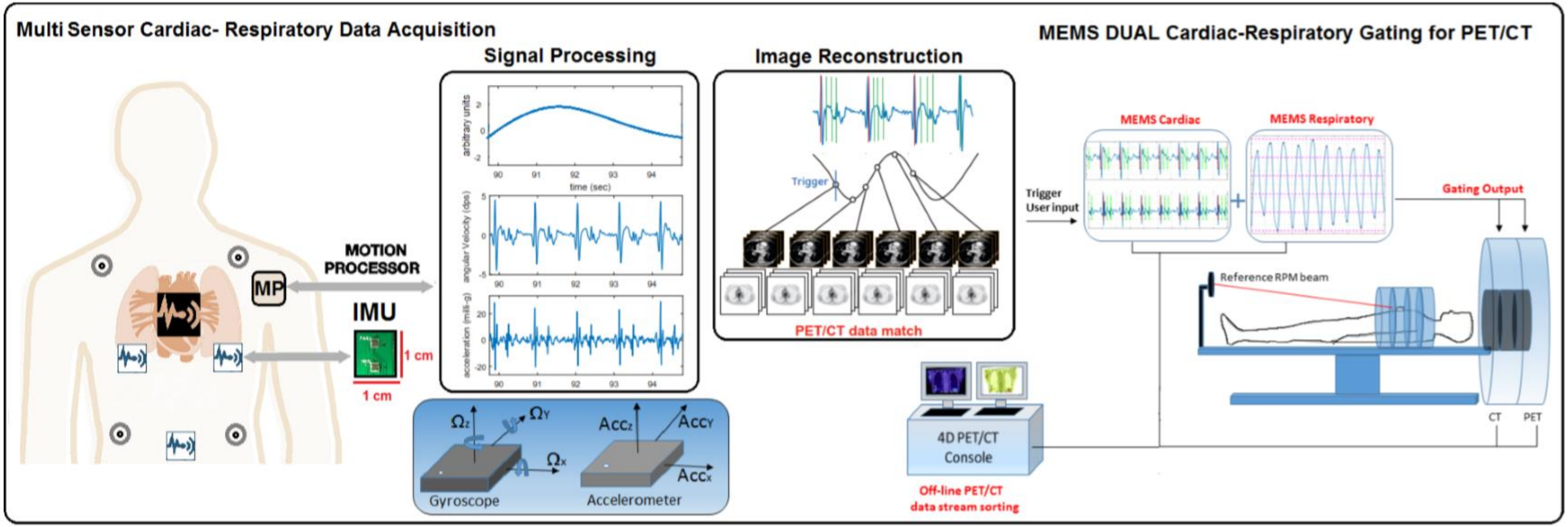
Use case:
respiration signal acquisition

PET imaging



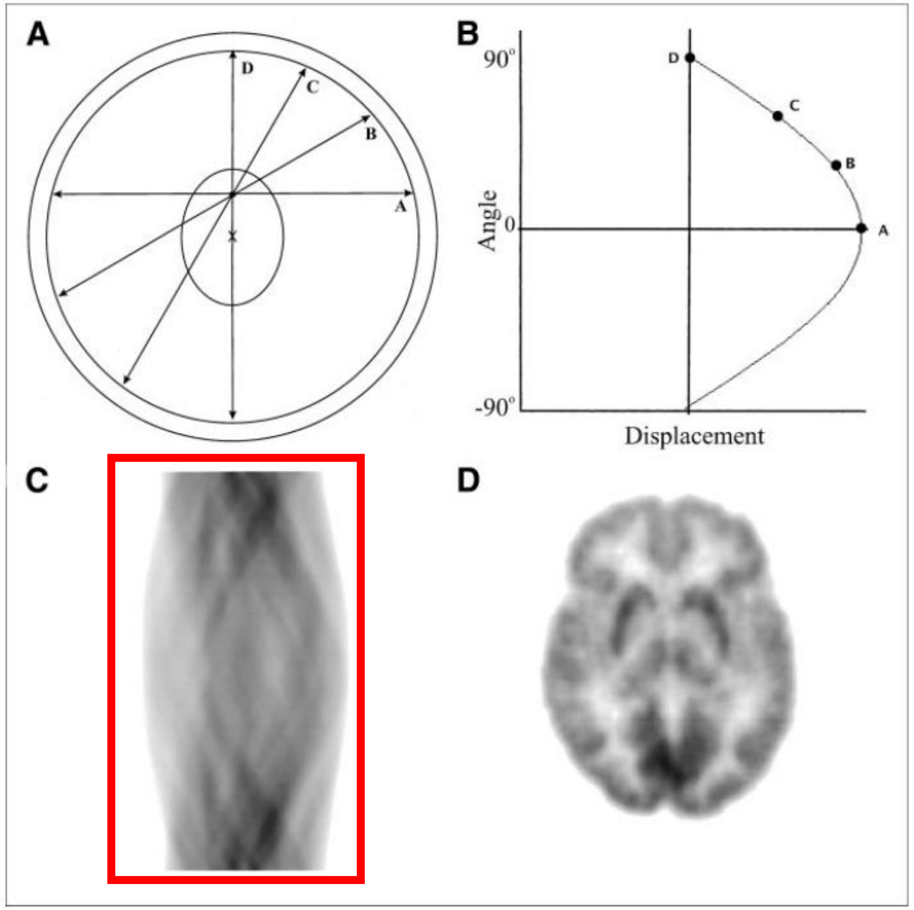
https://en.wikipedia.org/wiki/Positron_emission_tomography, Wikipedia, Wikimedia Foundation, 7.10.2020

Minimizing motion in medical PET imaging



UNIVERSITY OF TURKU

Sinograms



From: Fahey: ”Data Acquisition in PET Imaging”, Journal of Nuclear Medicine Technology 30(2):39-49

Data for respiration signal acquisition

We’ll consider the following three sources for respiration signal acquisition:

1. Signal from the Varian Real-time Position Management (RPM) optical sensor
2. Signal from a 3-axis MEMS accelerometer attached to the chest
3. Signal from summed sinograms (a 47 dimensional signal)
    - Each sinogram corresponds to one of 47 slices of the PET scanner
    - Each sinogram corresponds to a PET acquistion time of 50 ms
    - The summed sinogram corresponds to the total number of detected events in that slice (during a 50 ms time window)

Hypotheses for acquiring respiration signal

1. Respiration is the most significant effect in producing variance in the accelerometer and summed sinogram (PET) data
    - Respiration signal can thus be acquired as the first principal component of the data
2. Respiration frequency is approximately 30/min, or one respiratory cycle per 2 seconds
    - Low-pass filtering the outcome of PCA to frequencies near 2 Hz should bring out the respiration signal

UNIVERSITY OF TURKU

Demo

Theory

Vectors

Inner product

The inner product (dot product) of two real-valued vectors $\boldsymbol{v} = (v_1, v_2, \ldots, v_m)$ and $\boldsymbol{u} = (u_1, u_2, \ldots, u_m)$ is the sum

$$\boldsymbol{v} \circ \boldsymbol{u} = \sum_{i=1}^{m} v_i u_i$$

For example, the inner product of *(1,2,3)* and *(4,5,6)* equals *1 * 4 + 2 * 5 + 3 * 6 = 32.*

Orthonormality

Vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are orthonormal if their inner products satisfy

$$\boldsymbol{v}_i \circ \boldsymbol{v}_i = 1 \text{ for all } i, \text{and } \boldsymbol{v}_i \circ \boldsymbol{v}_j = 0 \text{ for all } i \neq j$$

For example, vectors (1,0,0), (0,1,0) and (0,0,1) are orthonormal. They all have length 1 and their pairwise inner products equal zero.

Pairwise inner products:
*1 * 0 + 0 * 1 + 0 * 0 = 0*
*1 * 0 + 0 * 0 + 0 * 1 = 0*
*0 * 0 + 1 * 0 + 0 * 1 = 0*

Orthonormality

Vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are orthonormal if their inner products satisfy

$$\boldsymbol{v}_i \circ \boldsymbol{v}_i = 1 \text{ for all } i, \text{and } \boldsymbol{v}_i \circ \boldsymbol{v}_j = 0 \text{ for all } i \neq j$$

For example, vectors $(1/\sqrt{2}, 1/\sqrt{2}, 0)$, $(1/\sqrt{2}, -1/\sqrt{2}, 0)$ and $(0,0,1)$ are orthonormal. They all have length 1 and their pairwise inner products equal zero.

Pairwise inner products:

$(1/\sqrt{2}) * (1/\sqrt{2}) + (1/\sqrt{2}) * (-1/\sqrt{2}) + 0 * 0 = 0$

$(1/\sqrt{2}) * 0 + (1/\sqrt{2}) * 0 + 0 * 1 = 0$

$(1/\sqrt{2}) * 0 + (-1/\sqrt{2}) * 0 + 0 * 1 = 0$

UNIVERSITY OF TURKU

Point of view

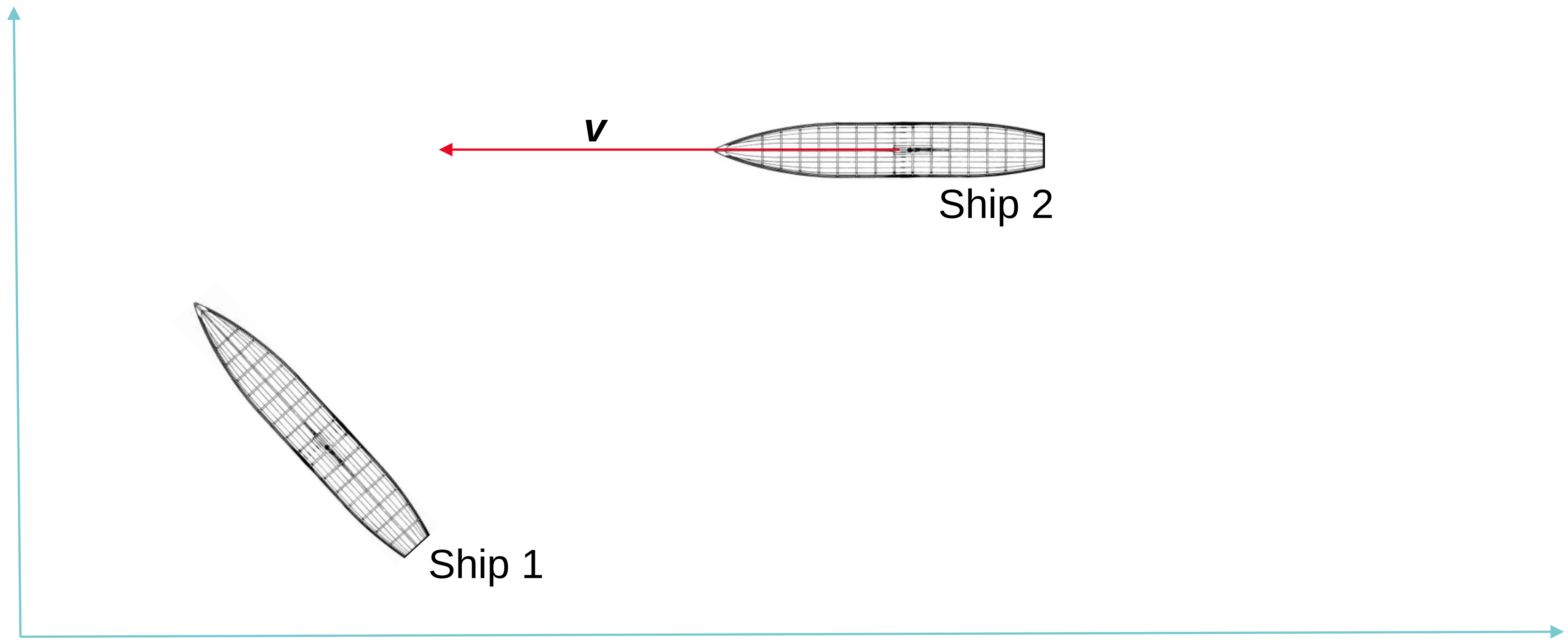
If the number of orthonormal vectors is the same as the dimensionality of the space, we say that these vectors offer us a **point of view**. The vectors are the directions of right-angled axes of a coordinate system.

For example, the sets {(1,0,0),(0,1,0),(0,0,1)} and {$(1/\sqrt{2},1/\sqrt{2},0)$, $(1/\sqrt{2},-1/\sqrt{2},0)$,(0,0,1)} offer us two different point of views of the 3-dimensional space.

Note: right-angled axes need not be right-handed axes.

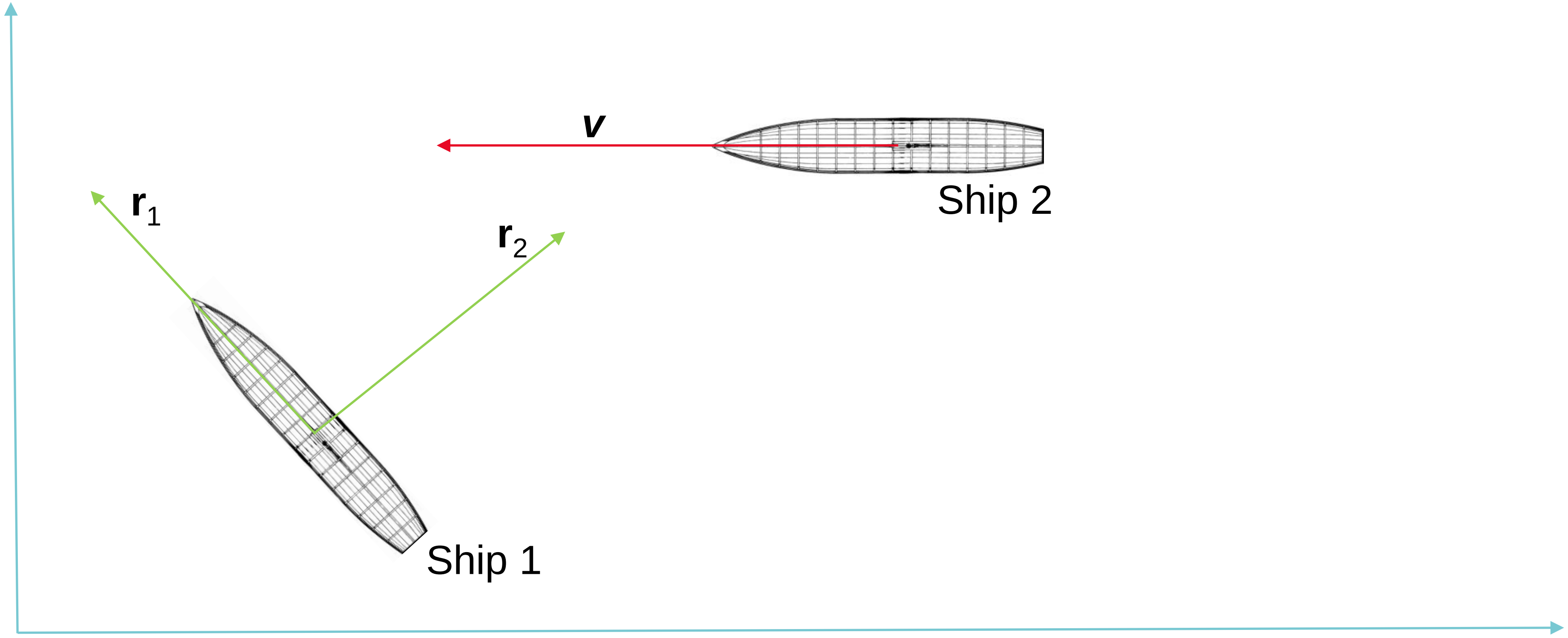
Using the point of view

Problem 1: From the point of view of Ship 1, what is the velocity vector ***v*** of Ship 2?



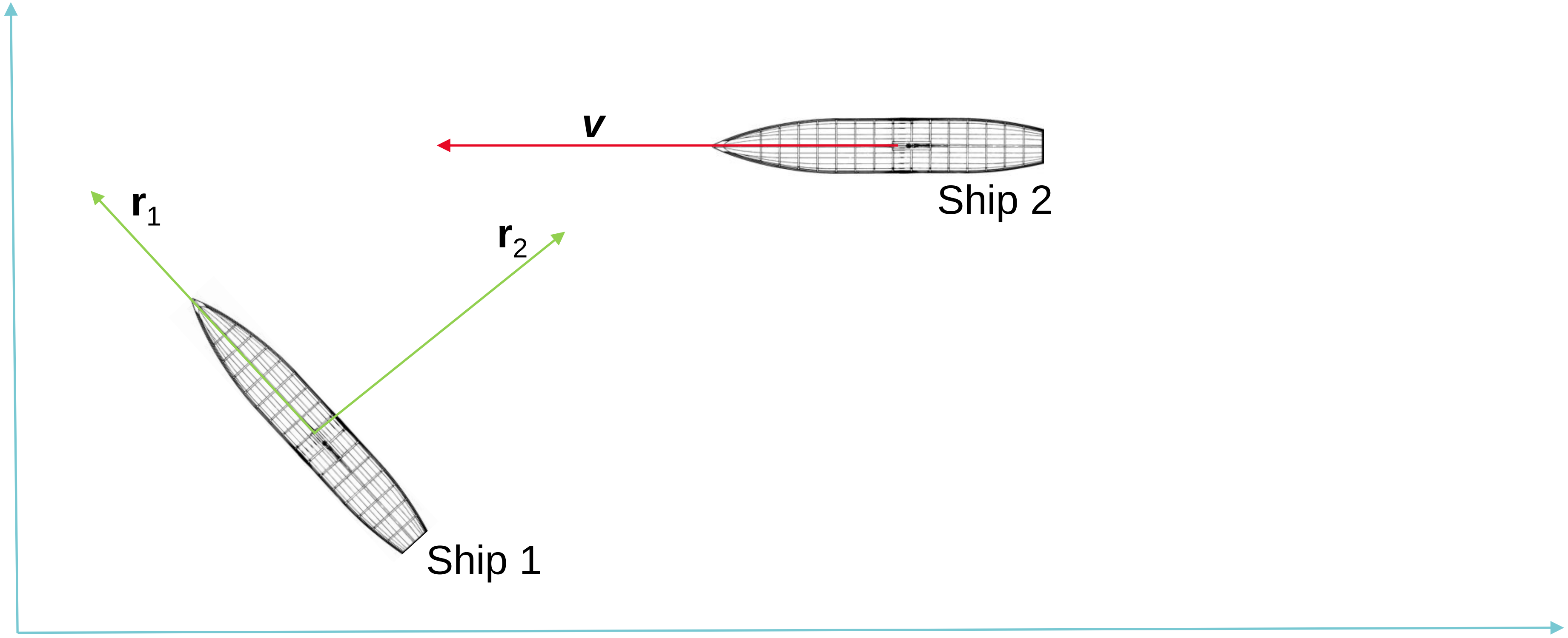
Using the point of view

Solution: First, let us find the unit vectors $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ in world coordinates pointing towards the heading and the port-to-starboard direction, respectively, for Ship 1:



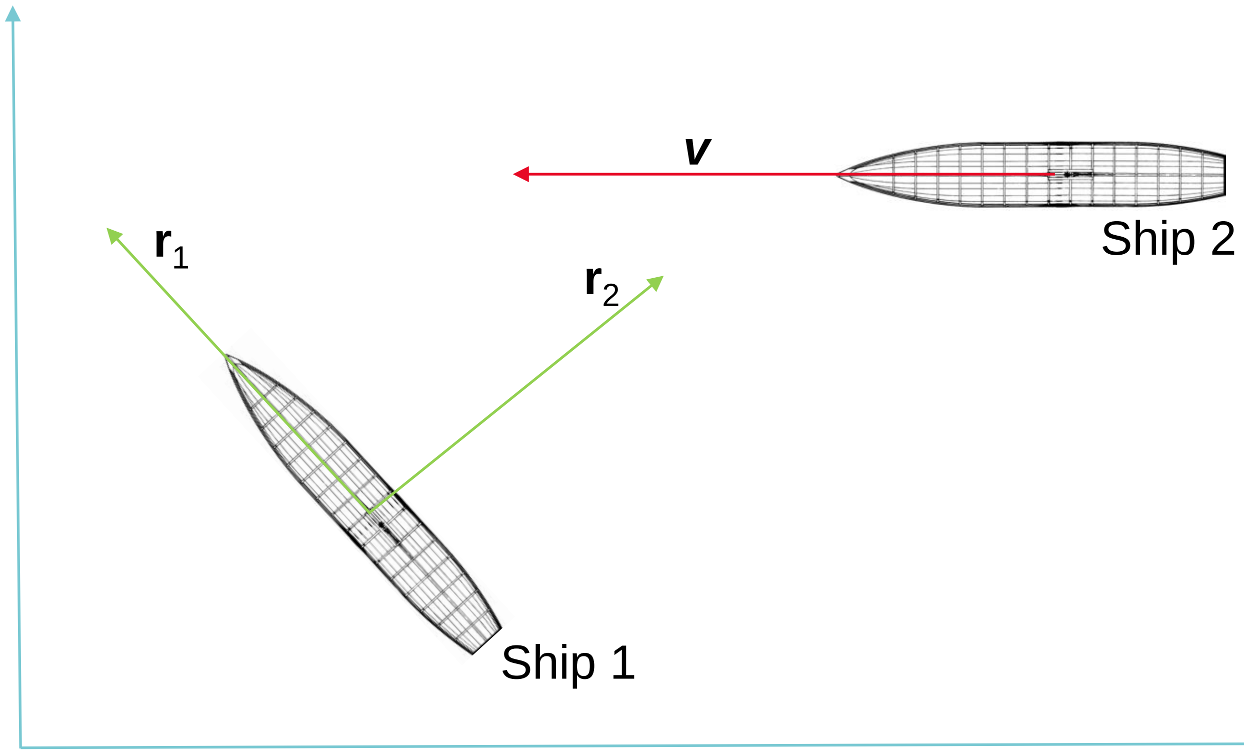
Using the point of view

Solution: Then, the velocity of Ship 2 from the point of view of Ship 1 is the vector $\boldsymbol{v}_{s1} = (\boldsymbol{r}_1 \circ \boldsymbol{v}, \boldsymbol{r}_2 \circ \boldsymbol{v})$.



Using the point of view

Solution: Then, the velocity of Ship 2 from the point of view of Ship 1 is the vector $\boldsymbol{v}_{s1} = (\boldsymbol{r}_1 \circ \boldsymbol{v}, \boldsymbol{r}_2 \circ \boldsymbol{v})$.

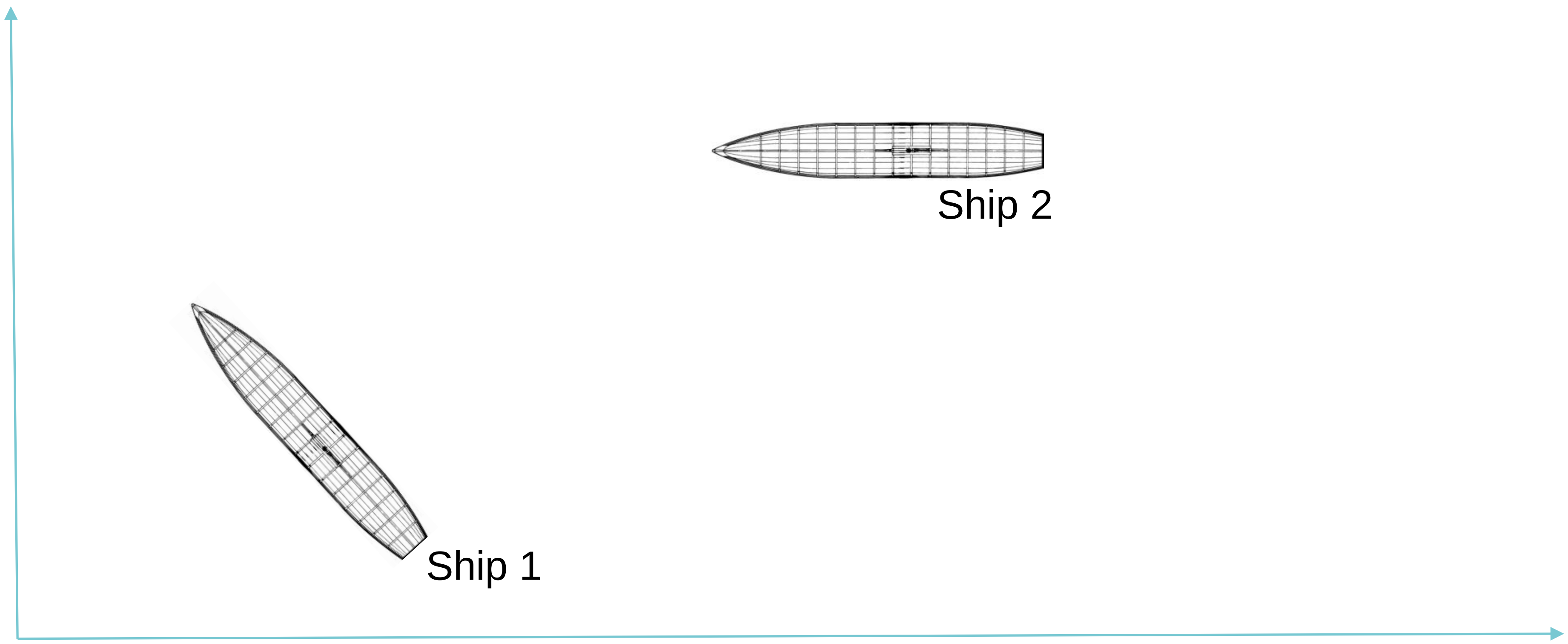


For example, if $\boldsymbol{v} = (-10.0, 0.0)$ m/s,
$\boldsymbol{r}_1 = (-0.707, 0.707)$ and $\boldsymbol{r}_2 = (0.707, 0.707)$,
then $\boldsymbol{v}_{s1} = (-7.07, 7.07)$ m/s.

From the point of view of Ship 1, Ship 2 moves approx. 7 m/s to the left and 7 m/s towards the direction of its heading.

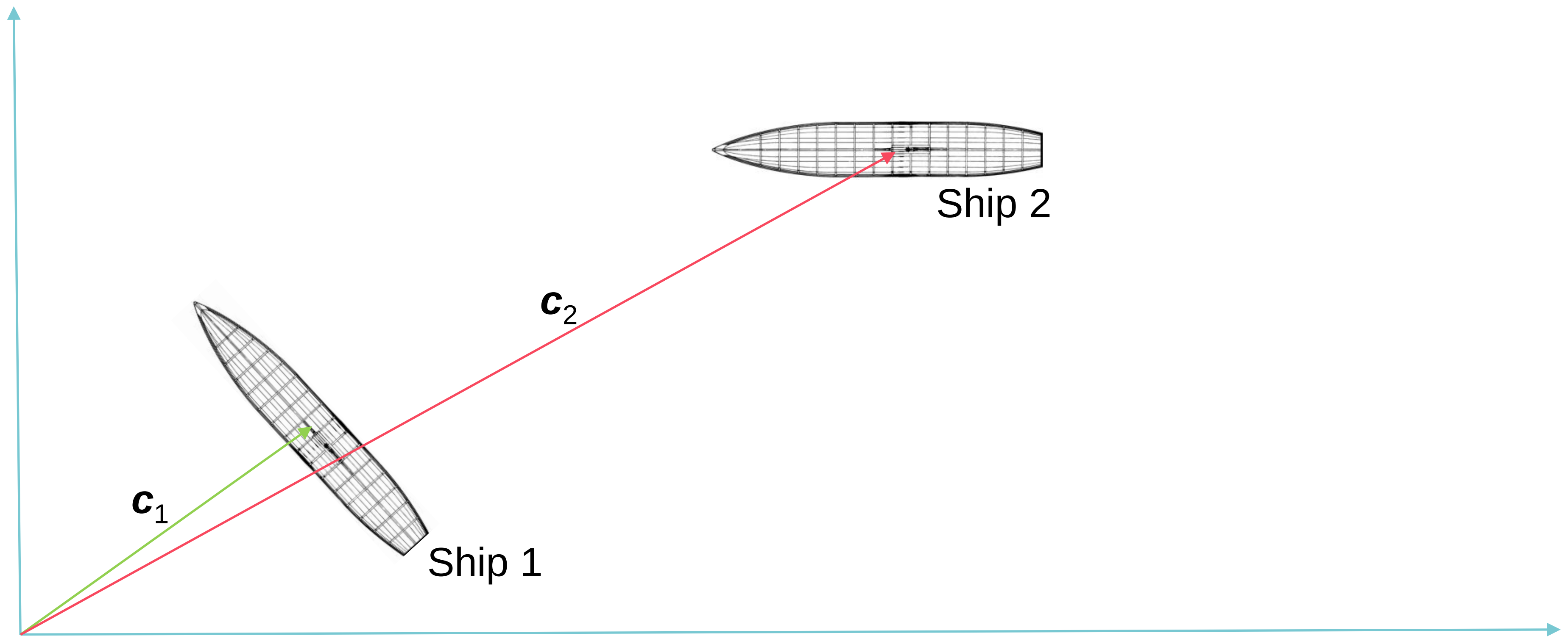
Using the point of view

Problem 2: From the point of view of Ship 1, what is the position vector of Ship 2?



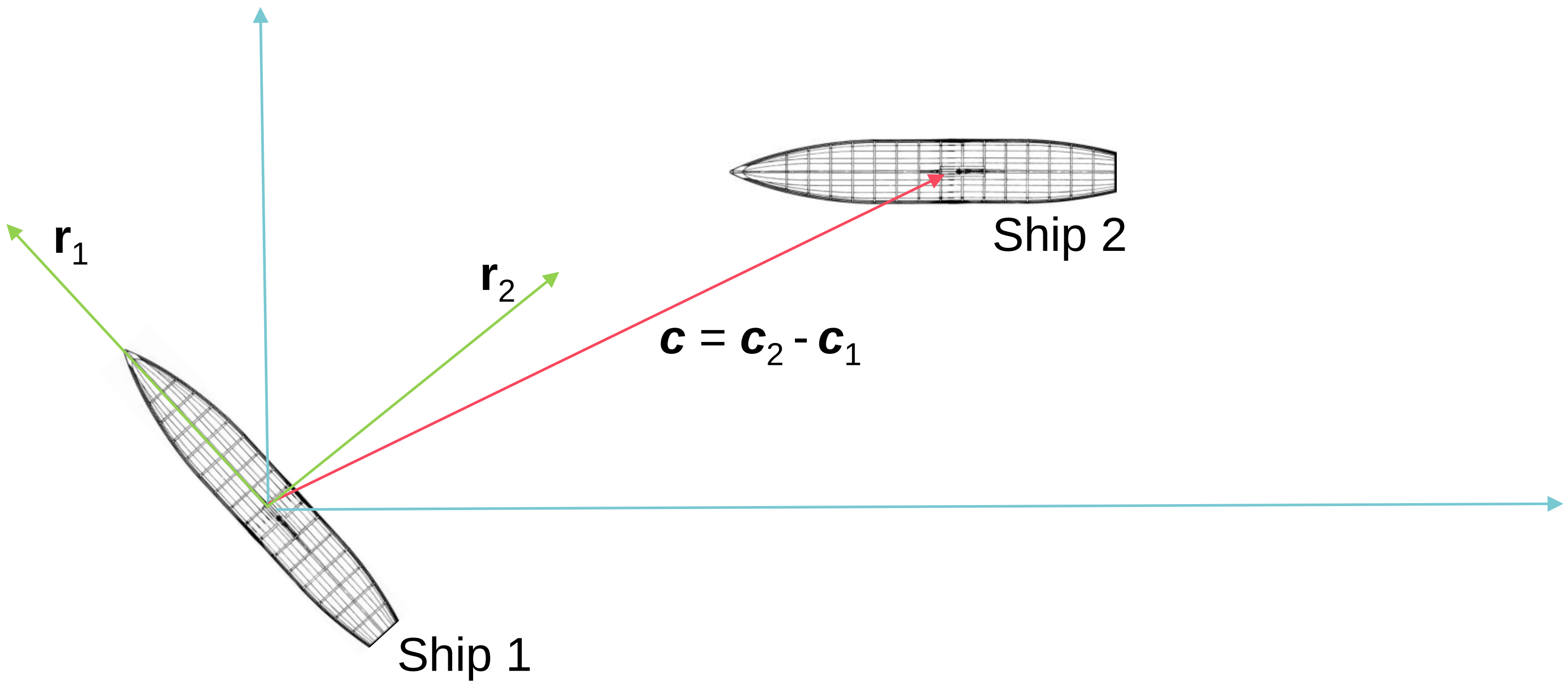
Using the point of view

Solution: First we subtract the position vector $\boldsymbol{c}_1$ of Ship 1 from the position vector $\boldsymbol{c}_2$ of Ship 2. These position vectors are represented in world coordinates.



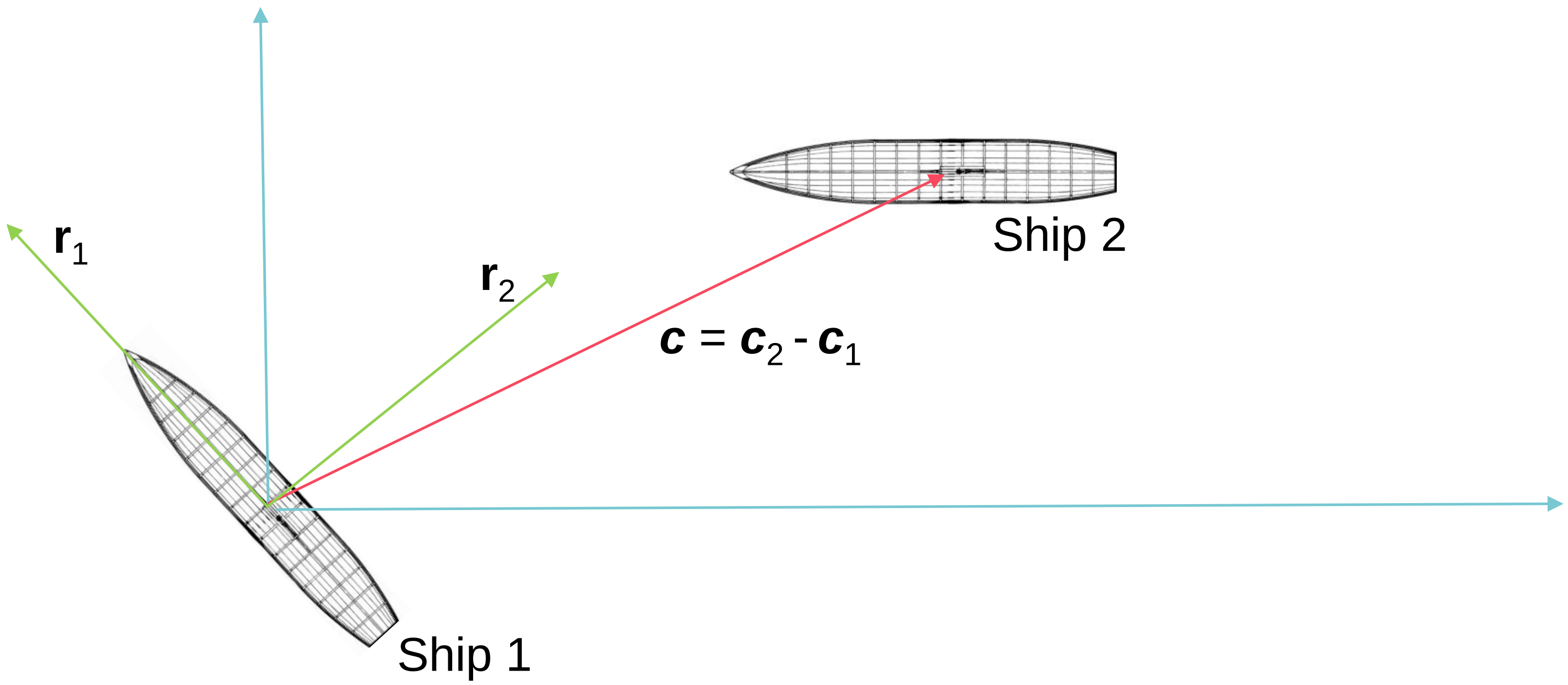
Using the point of view

Solution: First we subtract the position vector $\boldsymbol{c}_1$ of Ship 1 from the position vector $\boldsymbol{c}_2$ of Ship 2. This operation is called *zero-centering*.



Using the point of view

Solution: Then the position of Ship 2 from the point of view of Ship 1 is the vector $\boldsymbol{p} = (\boldsymbol{r_1} \circ \boldsymbol{c}, \boldsymbol{r_2} \circ \boldsymbol{c})$,
where $\boldsymbol{c} = \boldsymbol{c}_2 - \boldsymbol{c}_1$ is the zero-centered position of Ship 2,
and $\boldsymbol{r_1}$ and $\boldsymbol{r_2}$ are the heading and starboard-to-port unit vectors of Ship 1.



Using the point of view (matrix notation)

Solution: Then the position of Ship 2 from the point of view of Ship 1 is the vector $\boldsymbol{p} = (\boldsymbol{r}_1 \circ \boldsymbol{c}, \boldsymbol{r}_2 \circ \boldsymbol{c})$,
where $\boldsymbol{c} = \boldsymbol{c}_2 - \boldsymbol{c}_1$ is the zero-centered position of Ship 2,
and $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are the heading and starboard-to-port unit vectors of Ship 1.

If we write the vectors $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ as the *rows* of matrix **R**, and the zero-centered position vector $\boldsymbol{c}$ as a *column vector*, we can write the above formula for $\boldsymbol{p}$ using matrix multiplication as

$$\boldsymbol{p} = \mathbf{R}\boldsymbol{c}.$$

Matrix notation

Solution: The position of Ship 2 from the point of view of Ship 1 is the vector $\boldsymbol{p} = (\boldsymbol{r}_1 \circ \boldsymbol{c}, \boldsymbol{r}_2 \circ \boldsymbol{c})$, where $\boldsymbol{c} = \boldsymbol{c}_2 - \boldsymbol{c}_1$ is the zero-centered position of Ship 2, and $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are the heading and starboard-to-port unit vectors of Ship 1.

If we write the vectors $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ as the *rows* of a matrix **R**, and the zero-centered position vector $\boldsymbol{c}$ as a *column vector*, we can write the above formula for $\boldsymbol{p}$ using matrix multiplication as

$$\boldsymbol{p} = \mathbf{R}\boldsymbol{c}.$$

NOTE! Since $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are orthonormal vectors, the matrix **R** multiplied by its *transpose* $\mathbf{R}^T$ equals

$$\mathbf{R}\,\mathbf{R}^T = \mathbf{R}^T\,\mathbf{R} = \begin{bmatrix} \boldsymbol{r}_1 \circ \boldsymbol{r}_1 & \boldsymbol{r}_1 \circ \boldsymbol{r}_2 \\ \boldsymbol{r}_2 \circ \boldsymbol{r}_1 & \boldsymbol{r}_2 \circ \boldsymbol{r}_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I} \text{ (the identity matrix)}$$

Matrix notation

If we write the point of view vectors $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ as the *rows* of a matrix **R**, and the zero-centered vector $\boldsymbol{c}$ as a *column vector*, then the vector $\boldsymbol{c}$ from the point of view of $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ equals

$$\boxed{\boldsymbol{p} = \mathbf{R}\boldsymbol{c}.}$$

NOTE! Since $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are orthonormal vectors, the matrix **R** multiplied by its *transpose* $\mathbf{R}^\mathrm{T}$ equals

$$\boxed{\mathbf{R}\,\mathbf{R}^\mathrm{T} = \mathbf{R}^\mathrm{T}\,\mathbf{R} = \begin{bmatrix} \boldsymbol{r}_1 \circ \boldsymbol{r}_1 & \boldsymbol{r}_1 \circ \boldsymbol{r}_2 \\ \boldsymbol{r}_2 \circ \boldsymbol{r}_1 & \boldsymbol{r}_2 \circ \boldsymbol{r}_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I} \text{ (the identity matrix)}}$$

Statistics

Measurement vector

Let $\boldsymbol{x}$ be an n-dimensional vector $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$, whose elements correspond to subsequent measurements of some quantity.

The mean value of $\boldsymbol{x}$ is

$$\text{mean}(\boldsymbol{x}) = \frac{1}{n}\sum x_i$$

Zero-centered measurement vector

Let $\boldsymbol{x}$ be an *n*-dimensional vector $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$, whose elements correspond to subsequent measurements of some quantity.

The mean value of $\boldsymbol{x}$ is

$$\text{mean}(\boldsymbol{x}) = \frac{1}{n}\sum x_i$$

The vector $\boldsymbol{x}$ is *zero-centered,* if $\text{mean}(\boldsymbol{x}) = 0$.

Zero-centering a measurement vector

To zero-center an *n*-dimensional vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, subtract mean($\boldsymbol{x}$) from each of its elements:

$$\widehat{\boldsymbol{x}} = \boldsymbol{x} - \mathrm{mean}(\boldsymbol{x}) = \left(\mathrm{x}_1 - \mathrm{mean}(\boldsymbol{x}), \mathrm{x}_2 - \mathrm{mean}(\boldsymbol{x}), \ldots, \mathrm{xn} - \mathrm{mean}(\boldsymbol{x})\right)$$

Now,

$$\mathrm{mean}(\widehat{\boldsymbol{x}}) = \frac{1}{n}\sum \widehat{\boldsymbol{x}}_i = \frac{1}{n}\sum \boldsymbol{x}_\mathrm{i} - \mathrm{mean}(\boldsymbol{x}) = \mathrm{mean}(\boldsymbol{x}) - \frac{n}{n}\,\mathrm{mean}(\boldsymbol{x}) = 0.$$

Variance of a measurement vector

Let ***x*** be an *n*-dimensional vector ***x*** = (x_1, x_2, … , x_n), whose elements correspond to subsequent measurements of some quantity.

The variance of ***x*** is

$$\mathrm{var}(\boldsymbol{x}) = \frac{1}{n}\sum\left(x_i - \mathrm{mean}(\boldsymbol{x})\right)^2.$$

For a zero-centered vector, mean(***x***) = 0, and thus

$$\mathrm{var}(\boldsymbol{x}) = \frac{1}{n}\sum x_{\mathrm{i}}^2.$$

UNIVERSITY OF TURKU

Variance is signal power

Variance of ***x*** represents the *signal power* in ***x***. For example, the signal-to-noise ratio (SNR) can be defined as

$$\mathrm{SNR} = \frac{\mathrm{var(signal)}}{\mathrm{var(noise)}}$$

Signal power is independent of the mean of the signal. In other words, zero-centering a vector does not affect its power.

Variance as an inner product

Let $\boldsymbol{x}$ be a *zero-centered* n-dimensional vector $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$, whose elements correspond to subsequent measurements of some quantity.

The variance of $\boldsymbol{x}$ is

$$\mathrm{var}(\boldsymbol{x}) = \frac{1}{n}\sum x_{\mathrm{i}}^2,$$

which can be computed as an inner product as

$$\mathrm{var}(\boldsymbol{x}) = \frac{1}{n}(\boldsymbol{x} \circ \boldsymbol{x}).$$

Variance as an inner product

Let $\boldsymbol{x}$ be a *zero-centered* n-dimensional vector $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$, whose elements correspond to subsequent measurements of some quantity.

The variance of $\boldsymbol{x}$ is

$$\mathrm{var}(\boldsymbol{x}) = \frac{1}{n}\sum x_i^2,$$

which can be computed as an inner product as

$$\boxed{\mathrm{var}(\boldsymbol{x}) = \frac{1}{n}(\boldsymbol{x} \circ \boldsymbol{x}).}$$

UNIVERSITY OF TURKU

Inner product of different measurements

Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be *zero-centered* n-dimensional vectors, whose elements correspond to subsequent measurements of some quantities.

Their inner product $\boldsymbol{x} \circ \boldsymbol{y}$ represents how much there is same signal in $\boldsymbol{x}$ and $\boldsymbol{y}$.

To see this, we can write $\boldsymbol{y}$ as

$$\boldsymbol{y} = \frac{\boldsymbol{x} \circ \boldsymbol{y}}{\boldsymbol{x} \circ \boldsymbol{x}} \boldsymbol{x} + \boldsymbol{z} = \boxed{\alpha\, \boldsymbol{x} + \boldsymbol{z}\ ,}$$

where $\boldsymbol{z}$ is orthogonal to $\boldsymbol{x}$,

$$\boldsymbol{x} \circ \boldsymbol{z} = 0.$$

Inner product of different measurements

Their inner product $\boldsymbol{x} \circ \boldsymbol{y}$ represents how much there is same signal in $\boldsymbol{x}$ and $\boldsymbol{y}$.

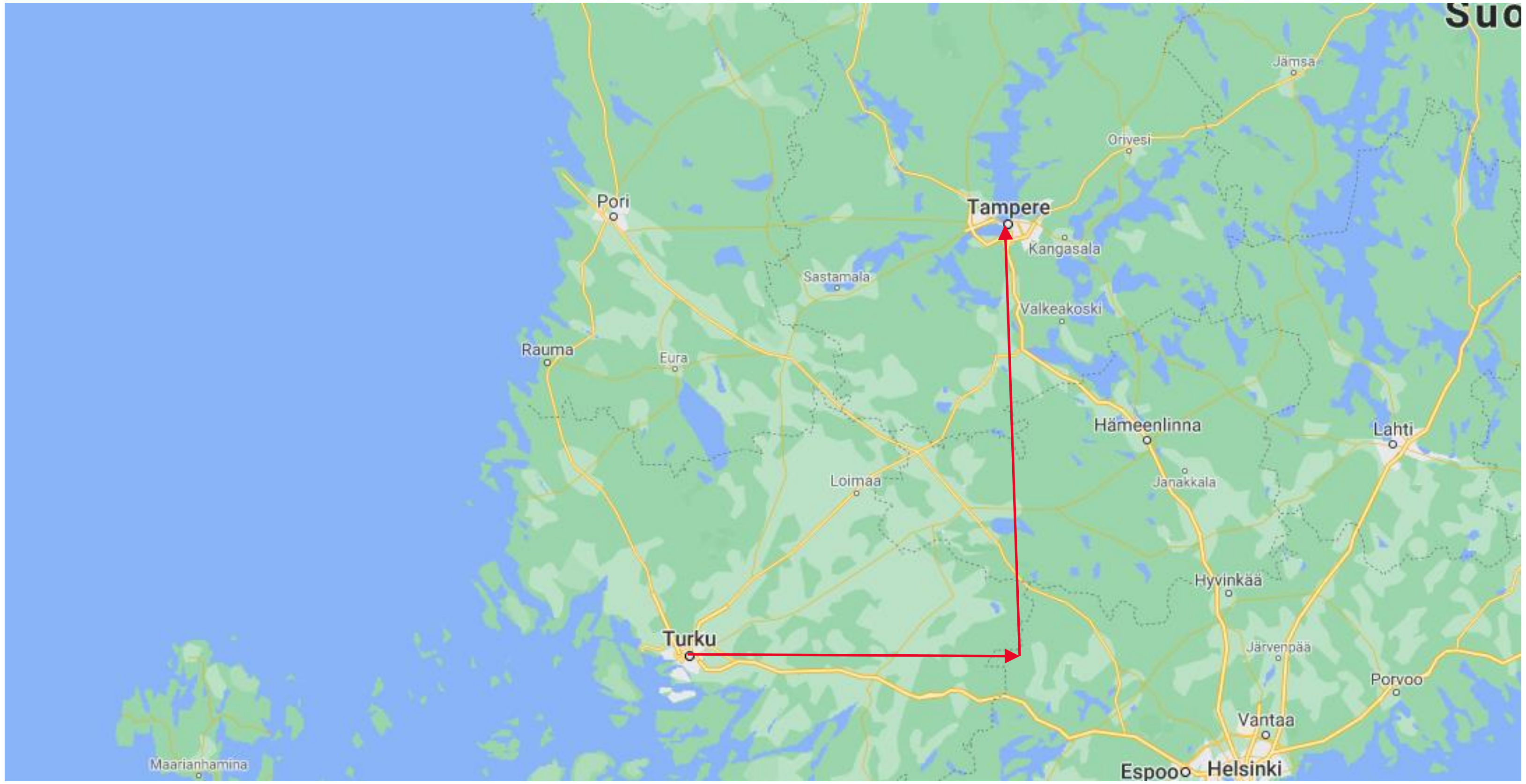
To have as informative measurements as possible, we would like to have

$$\boxed{\boldsymbol{x} \circ \boldsymbol{y} = 0}$$

for all pairs of different zero-centered measurements $(\boldsymbol{x},\boldsymbol{y})$.

Orthogonal measurements

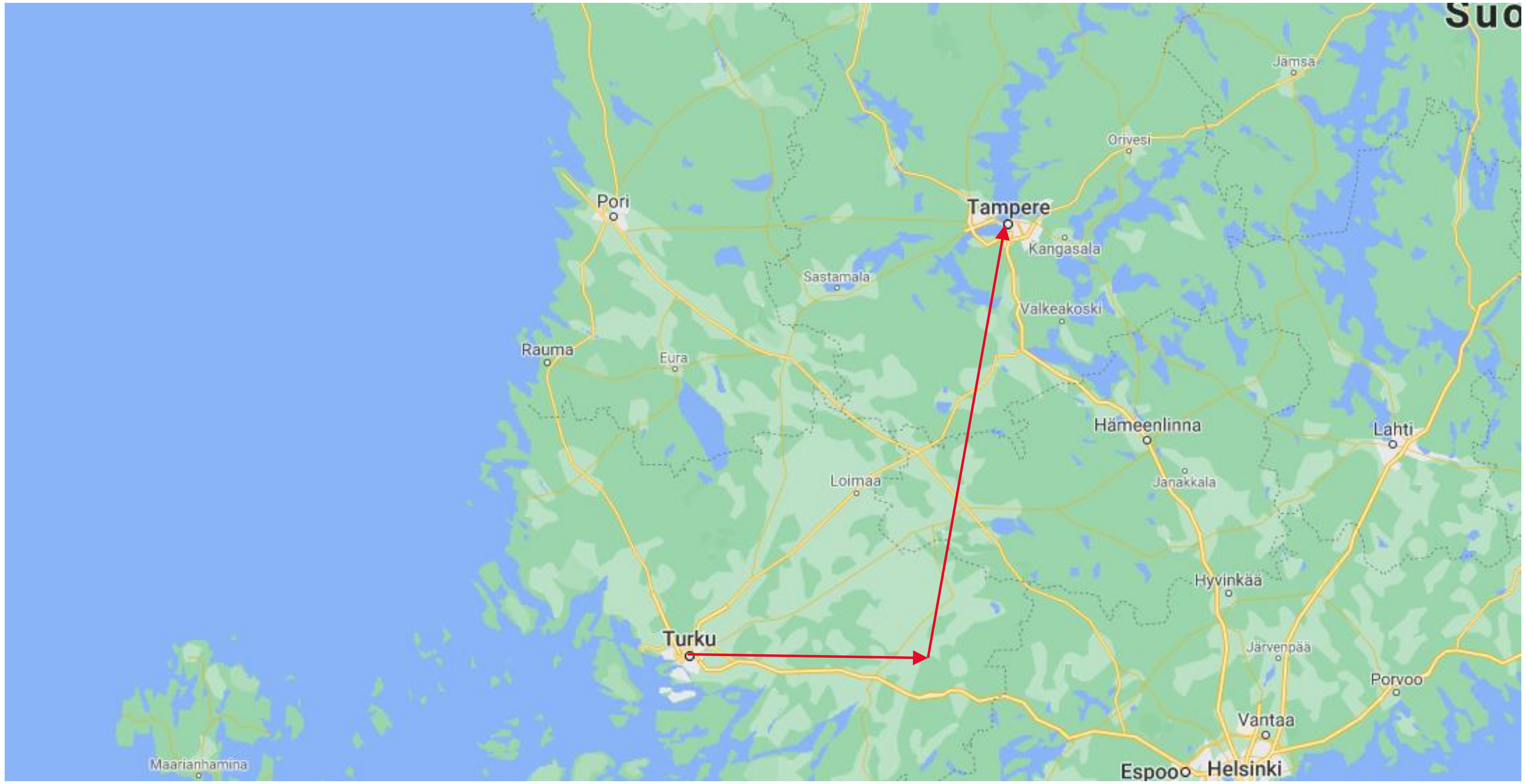
For example, if you need to describe the location of Tampere w.r.t Turku, you could say how much it is towards north and towards east of Turku:
These directions are orthogonal to each other.



UNIVERSITY OF TURKU

Non-orthogonal measurements

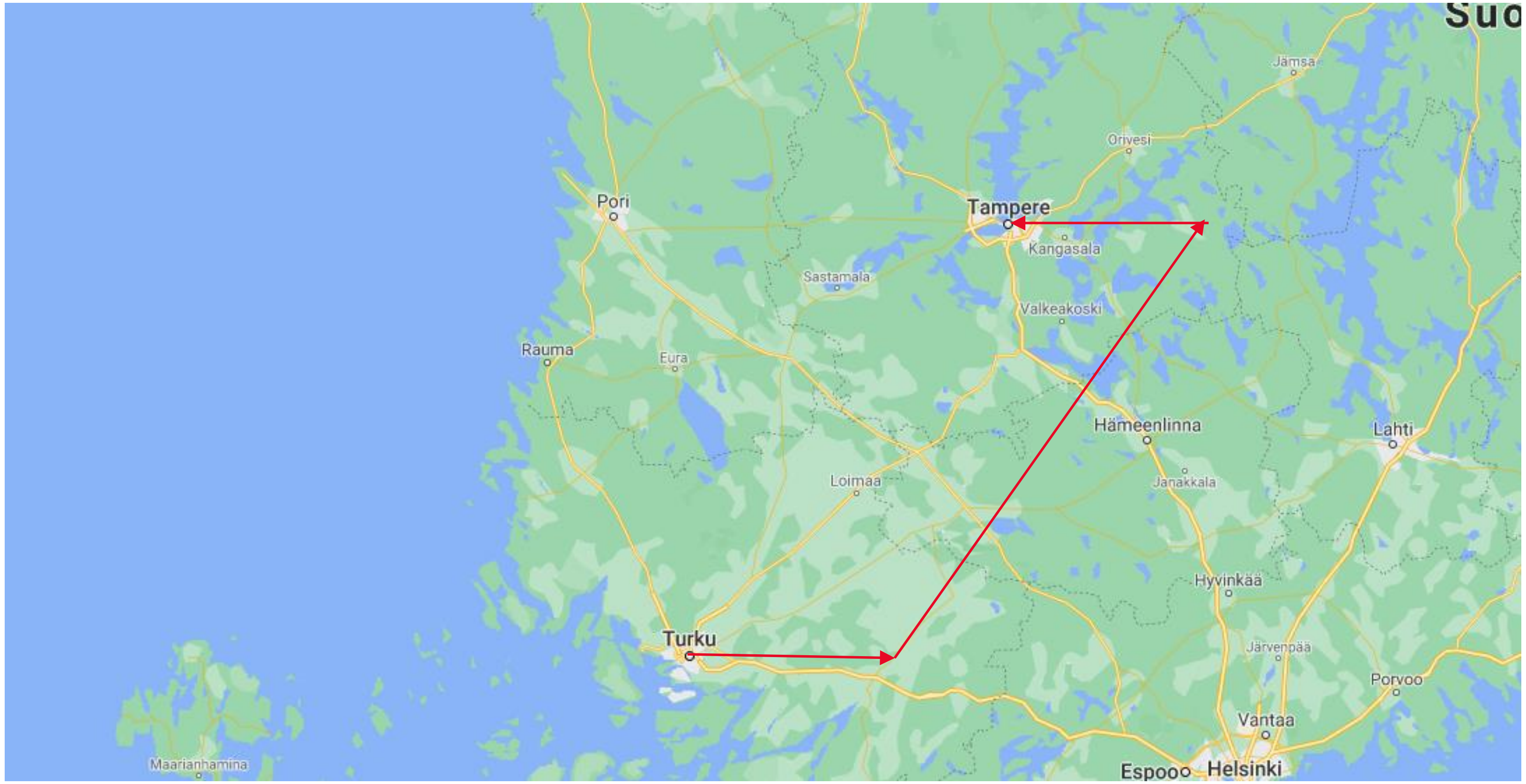
It would make less sense to describe how much east, and then north by northeast Tampere is:



UNIVERSITY OF TURKU

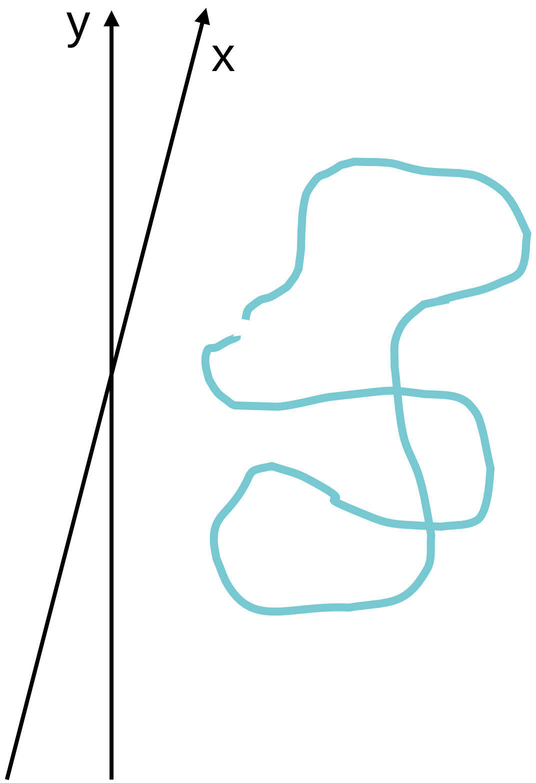
Non-orthogonal measurements

It would make even less sense to describe how much east, then northeast and finally west Tampere is:



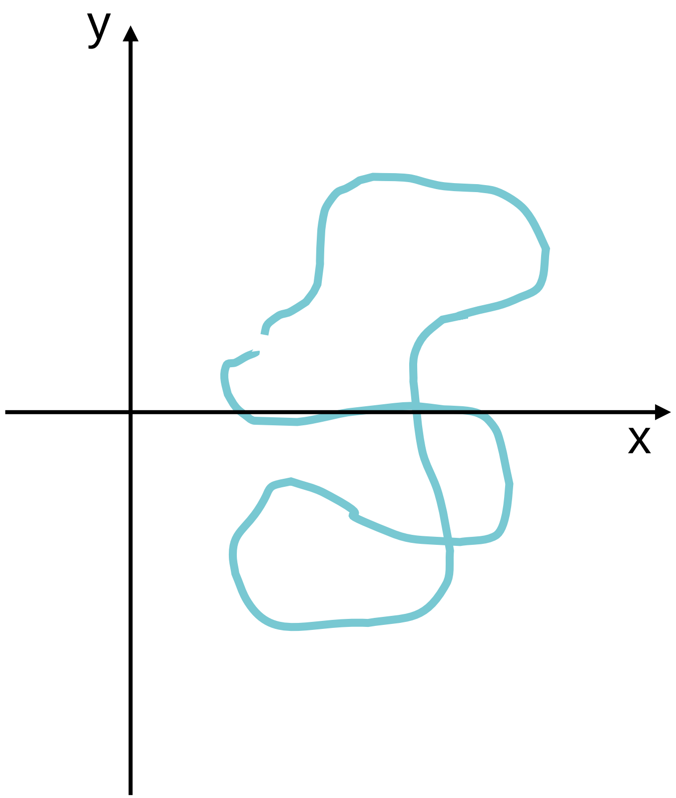
UNIVERSITY OF TURKU

Finite precision (finite number of bits)



Non-orthogonal measurements (redundancy)

vs.



Orthogonal measurements (no redundancy)

Inner product of different measurements

Their inner product $\boldsymbol{x} \circ \boldsymbol{y}$ represents how much there is same signal in $\boldsymbol{x}$ and $\boldsymbol{y}$.

To have as informative (and non-redundant) measurements as possible, we would like to have

$$\boxed{\boldsymbol{x} \circ \boldsymbol{y} = 0}$$

for all pairs of different zero-centered measurements ($\boldsymbol{x}$,$\boldsymbol{y}$).

Such measurements are called *uncorrelated.*

Uncorrelated measurements

In an ideal world, we would like to have uncorrelated measurements with high SNR.

In reality, measurements are typically correlated. This redundancy may be beneficial in correcting errors and filtering out noise, but the pure sensor description of the world may not be the most informative one.

Principal component analysis uncorrelates the measurements, and gives us a point of view which is typically much more informative than the pure sensor point of view.

Principal Component Analysis

UNIVERSITY
OF TURKU

Uncorrelated measurements

Let ***X*** be an m x n matrix (a matrix with *m* rows and *n* columns), whose rows represent subsequent *zero-centered* measurements using a specific sensors. In other words, there are *m* different sensors each having *n* different measurements. We assume that *n >> m.*

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix}$$

Uncorrelated measurements

Let $\boldsymbol{X}$ be an m x n matrix (a matrix with *m* rows and *n* columns), whose rows represent subsequent *zero-centered* measurements using a specific sensors. Different measurements are denoted by $\boldsymbol{x}_i$, where i = 1, …, *m.*

In the ideal case, the inner products of measurements performed with different sensors equal to 0, while the signal powers (variances) of each sensor are large in comparison to the noise power:

$$\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} = \begin{bmatrix} \boldsymbol{x}_1 \circ \boldsymbol{x}_1 & \boldsymbol{x}_1 \circ \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_1 \circ \boldsymbol{x}_m \\ \boldsymbol{x}_2 \circ \boldsymbol{x}_1 & \boldsymbol{x}_2 \circ \boldsymbol{x}_2 & & \boldsymbol{x}_2 \circ \boldsymbol{x}_m \\ \vdots & & \ddots & \vdots \\ \boldsymbol{x}_m \circ \boldsymbol{x}_1 & \boldsymbol{x}_m \circ \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_m \circ \boldsymbol{x}_m \end{bmatrix} = n \cdot \begin{bmatrix} \mathrm{var}(\boldsymbol{x}_1) & 0 & 0 & \cdots & 0 \\ 0 & \mathrm{var}(\boldsymbol{x}_2) & 0 & & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & & & \cdots & \mathrm{var}(\boldsymbol{x}_{\mathrm{m}}) \end{bmatrix}$$

Uncorrelated measurements

In the ideal case of uncorrelated measurements,

$$\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} = \begin{bmatrix} d_1 & 0 & 0 & \dots & 0 \\ 0 & d_2 & 0 & & 0 \\ & \vdots & & \ddots & \vdots \\ & 0 & & \dots & d_m \end{bmatrix} = \boldsymbol{D},$$

where $\boldsymbol{D}$ is a *diagonal* matrix. If the measurements $\boldsymbol{x}_i$ have been ordered in a decreasing order in terms of their signal powers, then $d_1 \geq d_2 \geq \ldots \geq d_m$.

The signal power of $\boldsymbol{x}_i$ is $\mathrm{var}(\boldsymbol{x}_i) = d_i / n$.

Recap: Matrix notation

If we write the point of view vectors $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ as the *rows* of a matrix **R**, and the zero-centered vector $\boldsymbol{c}$ as a *column vector*, then the vector $\boldsymbol{c}$ from the point of view of $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ equals

$$\boxed{\boldsymbol{p} = \mathbf{R}\boldsymbol{c}.}$$

NOTE! Since $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are orthonormal vectors, the matrix **R** multiplied by its *transpose* $\mathbf{R}^T$ equals

$$\boxed{\mathbf{R}\,\mathbf{R}^T = \mathbf{R}^T\,\mathbf{R} = \begin{bmatrix} \boldsymbol{r}_1 \circ \boldsymbol{r}_1 & \boldsymbol{r}_1 \circ \boldsymbol{r}_2 \\ \boldsymbol{r}_2 \circ \boldsymbol{r}_1 & \boldsymbol{r}_2 \circ \boldsymbol{r}_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I} \text{ (the identity matrix)}}$$

Principal component analysis

Typically, $\boldsymbol{XX}^{\mathrm{T}}$ is not a diagonal matrix, since measurements in the real world are often correlated.

Let's find a point of view ***R***, from which the zero-centered measurements do look uncorrelated. This is called *Principal Component Analysis.*

From this point of view, the measurement matrix can be written as

$$\mathbf{Y} = \mathbf{RX},$$

and for it,

$$\boldsymbol{YY}^{\mathrm{T}} = \begin{bmatrix} d_1 & 0 & 0 & \dots & 0 \\ 0 & d_2 & 0 & & 0 \\ & \vdots & & \ddots & \vdots \\ & 0 & & \cdots & d_m \end{bmatrix} = \boldsymbol{D},$$

where $d_1 \geq d_2 \geq \ldots \geq d_m$. In other words, $\boldsymbol{y}_1$ has the largest signal power, $\boldsymbol{y}_2$ has the second-largest signal power, and so on, and the signals $\boldsymbol{y}_i$ are uncorrelated.

UNIVERSITY OF TURKU

Principal component analysis

Principal Component Analysis (PCA) means finding a point of view matrix ***R*** for which

$$\boldsymbol{Y} = \boldsymbol{R}\boldsymbol{X},$$

and

$$\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}} = \begin{bmatrix} d_1 & 0 & 0 & \dots & 0 \\ 0 & d_2 & 0 & & 0 \\ & \vdots & & \ddots & \vdots \\ & 0 & & \cdots & d_m \end{bmatrix} = \boldsymbol{D},$$

where $d_1 \geq d_2 \geq \ldots \geq d_m$.

The rows of ***Y*** are the principal components of ***X***.

Finding the correct point of view

We try to find a point of view matrix $\boldsymbol{R}$ for which

$$\boldsymbol{Y} = \boldsymbol{R}\boldsymbol{X}$$

and

$$\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}} = \begin{bmatrix} d_1 & 0 & 0 & \dots & 0 \\ 0 & d_2 & 0 & & 0 \\ & \vdots & & \ddots & \vdots \\ & 0 & & \cdots & d_m \end{bmatrix} = \boldsymbol{D}.$$

Therefore,

$$\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}} = \boldsymbol{R}\boldsymbol{X}(\boldsymbol{R}\boldsymbol{X})^{\mathrm{T}} = \boldsymbol{R}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{R}^{\mathrm{T}} = \boldsymbol{D}.$$

Finding the correct point of view

We try to find a point of view matrix ***R*** for which

$$\boldsymbol{Y} = \boldsymbol{R}\boldsymbol{X}$$

and

$$\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}} = \begin{bmatrix} d_1 & 0 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & & 0 \\ & \vdots & & \ddots & \vdots \\ & 0 & & \cdots & d_m \end{bmatrix} = \boldsymbol{D}.$$

Therefore,

$$\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}} = \boldsymbol{R}\boldsymbol{X}(\boldsymbol{R}\boldsymbol{X})^{\mathrm{T}} = \boldsymbol{R}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{R}^{\mathrm{T}} = \boldsymbol{D}.$$

$\boxed{(\boldsymbol{AB})^{\mathrm{T}} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}}$

UNIVERSITY OF TURKU

Finding the correct point of view

From

$$\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}} = \boldsymbol{R}\boldsymbol{X}(\boldsymbol{R}\boldsymbol{X})^{\mathrm{T}} = \boldsymbol{R}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{R}^{\mathrm{T}} = \boldsymbol{D}$$

it follows that

$$\boxed{\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{R}.}$$

Finding the correct point of view

From

$$\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}} = \boldsymbol{R}\boldsymbol{X}(\boldsymbol{R}\boldsymbol{X})^{\mathrm{T}} = \boldsymbol{R}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{R}^{\mathrm{T}} = \boldsymbol{D}$$

it follows that

$$\boxed{\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{R}.}$$

$$\mathbf{R}\,\mathbf{R}^{\mathrm{T}} = \mathbf{R}^{\mathrm{T}}\,\mathbf{R} = \begin{bmatrix} \boldsymbol{r}_1 \circ \boldsymbol{r}_1 & \boldsymbol{r}_1 \circ \boldsymbol{r}_2 \\ \boldsymbol{r}_2 \circ \boldsymbol{r}_1 & \boldsymbol{r}_2 \circ \boldsymbol{r}_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I} \text{ (the identity matrix)}$$

UNIVERSITY OF TURKU

Finding the correct point of view

The correct point of view is the one represented by the point of view matrix (orthonormal matrix) $\boldsymbol{R}$, which satisfies

$$\boxed{\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{R}.}$$

Luckily, there are several methods to find ***R*** and ***D*** from this final equation, for example the *eigenvalue decomposition* and the *singular value decomposition (SVD)*, which are readily available in Matlab and Numpy. SVD is a numerically stable and computationally efficient method, so we will use that.

Singular value decomposition

Singular value decomposition (SVD) decomposes any matrix ***M*** into a product of three matrices ***Q***, ***Δ*** and ***P*** :

$$\boldsymbol{M} = \boldsymbol{Q}\boldsymbol{\Delta}\boldsymbol{P}.$$

Of these matrices, ***Q*** and ***P*** are point of view matrices (orthonormal matrices), and ***Δ*** is a (rectangular) diagonal matrix.

UNIVERSITY OF TURKU

Singular value decomposition

Singular value decomposition (SVD) decomposes the matrix $\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$ into a product of three matrices $\boldsymbol{R}^{\mathrm{T}}$, $\boldsymbol{D}$ and $\boldsymbol{R}$:

$$\boxed{\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{R}.}$$

.

Recipe for PCA

Let $\boldsymbol{X}$ be an m x n matrix (a matrix with *m* rows and *n* columns), whose rows represent subsequent measurements using a specific sensor. In other words, there are *m* different sensors each having *n* different measurements. We assume that *n* >> *m.*

Principal Component Analysis of $\boldsymbol{X}$ consists of the following steps:

1. Zero-centering of $\boldsymbol{X}$: Remove from each row its mean.

2. Compute the SVD of the matrix $\boldsymbol{X}\,\boldsymbol{X}^{\mathrm{T}}$. As the output of SVD you will get three matrices $\boldsymbol{Q}$, $\boldsymbol{D}$ and $\boldsymbol{R}$, for which

$$\boldsymbol{X}\,\boldsymbol{X}^{\mathrm{T}} = \boldsymbol{QDR}$$

3. The principal components of $\boldsymbol{X}$ are the rows of $\boldsymbol{Y} = \boldsymbol{RX}$

4. The signal powers of the principal components are the diagonal elements of the matrix
   $\boldsymbol{M} = \boldsymbol{D} / n$

UNIVERSITY OF TURKU

Recipe for PCA

Let ***X*** be an m x n matrix (a matrix with *m* rows and *n* columns), whose rows represent subsequent measurements using a specific sensor. In other words, there are *m* different sensors each having *n* different measurements. We assume that *n* >> *m.*

Principal Component Analysis of ***X*** consists of the following steps:

1. Zero-centering of ***X***: Remove from each row its mean.

2. Compute the SVD of the matrix $\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$. As the output of SVD you will get three matrices ***Q***, ***D*** and ***R***, for which

$$\boldsymbol{X}\,\boldsymbol{X}^{\mathrm{T}} = \boldsymbol{Q}\boldsymbol{D}\boldsymbol{R}$$

For some reason, Matlab gives you **Q**, **D** and $\mathbf{R}^{\mathrm{T}}$

3. The principal components of ***X*** are the rows of ***Y*** = ***RX***

4. The signal powers of the principal components are the diagonal elements of the matrix
M = ***D*** / *n*

UNIVERSITY OF TURKU

Demo revisited

Custom PCA function in Python:

```
def get_first_principal_component(X):
    (m,n) = X.shape
    for i in range(m):
        X[i,:] -= np.mean(X[i,:]) #zero-centering

    Q,D,R = np.linalg.svd(X @ X.T) #Singular Value Decomposition of the matrix X X^T (@ means matrix multiplication)

    Y = R @ X #principal components, whose signal powers are the diagonal elements of D/n

    return Y[0,:] #first principal component
```

UNIVERSITY OF TURKU

Summary

PCA is

- Computationally efficient (fast)
- Unique (in a sense)
    - Variances of the resulting *principle components* are unambiguous
    - In practice the resulting signals are often unique up to polarity
        - PCA does not know ”which side is up”

Caution: If in the original data there is lot of noise in some channel / sensor, PCA will pick it up (PCA treats all variance equally).

PCA is used

- to simplify data analysis & visualization (incl. clustering!)
- to find underlying causes in the data
- to compress data
- as a first step in more complex signal processing methods (for example, Independent Component Analysis)

PCA should always be considered when exploring and analyzing multidimensional data!