

# INTRODUCTION TO PRINCIPAL COMPONENT ANALYSIS I(DA\_2022)

Andrea Giansanti

Dipartimento di Fisica, Sapienza Università di Roma

[Andrea.Giansanti@roma1.infn.it](mailto:Andrea.Giansanti@roma1.infn.it)

Lecture n. 23, Rome wed May 25<sup>th</sup> 2022

DIPARTIMENTO DI FISICA



SAPIENZA  
UNIVERSITÀ DI ROMA

## Outline L 23

- Principal Component analysis PCA
- MULTIVARIATE CORRELATIONS
- [geometric data analysis/dimensional reduction / classification/clustering)]
- The introductory material can be found in a Chapter 3 of the textbook by
- Higgs and Attwood Bioinformatics and Molecular Evolution
- Together with basic ideas of reference frame, geometric idea of a vector and
- numerical representation

The discussion has been mainly elaborated at the white-board in the classroom see the recording of this lecture.

## **INTRODUCTION TO PRINCIPAL COMPONENT ANALYSIS**

**Search for multivariate correlations in an object-descriptor data table**

**The language of linear mathematics: symbolic operators and numerical representatives**

**What is a vector? What is a linear transformations of a vector?**

**What is a change of reference frame (basis)?**

**The eigenvalue problem**

**Z-transform of an object-descriptor table**

**Find eigenvalues and eigenvectors of the covariance matrix such as to maximize variance**

**The (ordered) eigenvalues of the covariance matrix encode the variance contained in the original data.**