Lecture 1

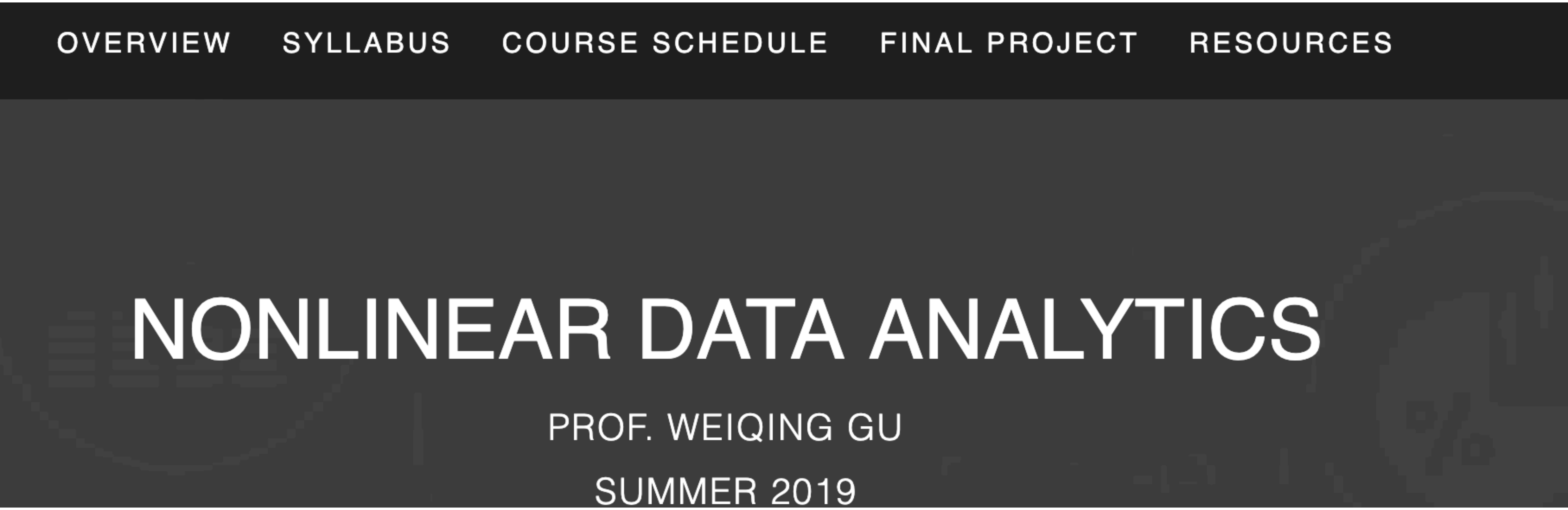
Math 178

Nonlinear Data Analytics

Prof. Weiqing Gu

Course Webpage

- https://math178su19.github.io/



Grading:

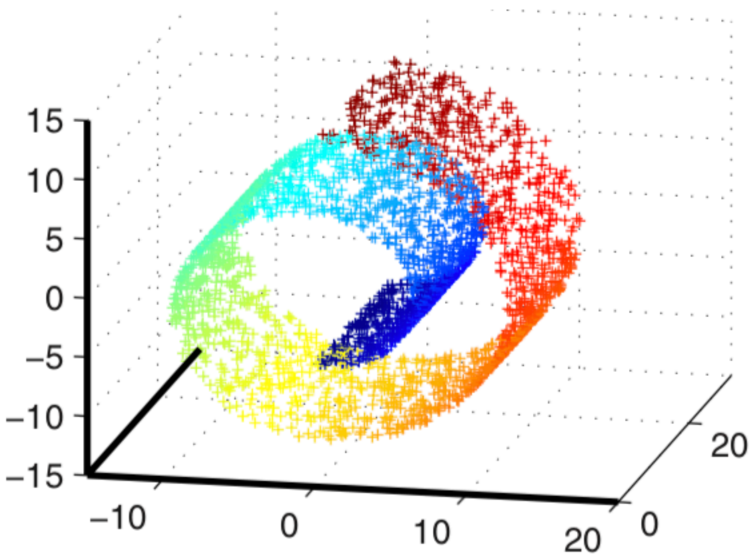
- 5% Reading Summary
- 35% Homework
- 20% Midterm Progress report
- 40% Final Project
- [Up to 5% Extra Credit]

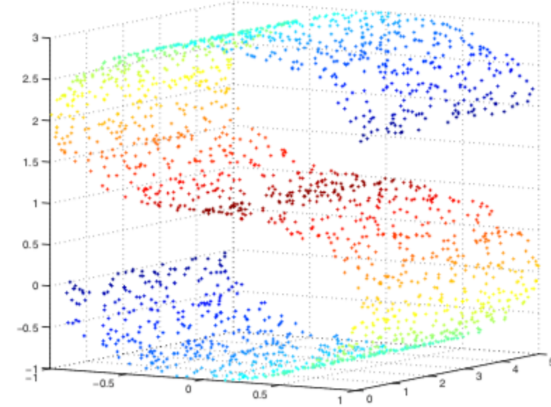
Overview of Lecture 1

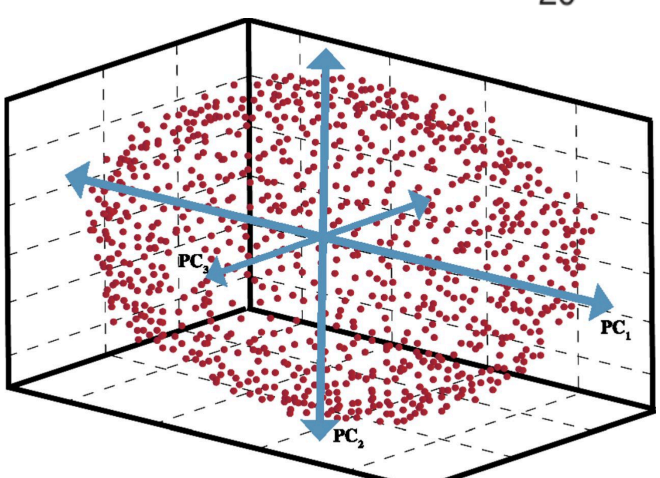
- Why we need nonlinear data analysis?
  - First starting with curves and their analysis
- Similarity measurements for nonlinear data
  - First a few examples: Arc-length, Geodesic length
- Introduction to cell phone data
- Introduction to rigid motion

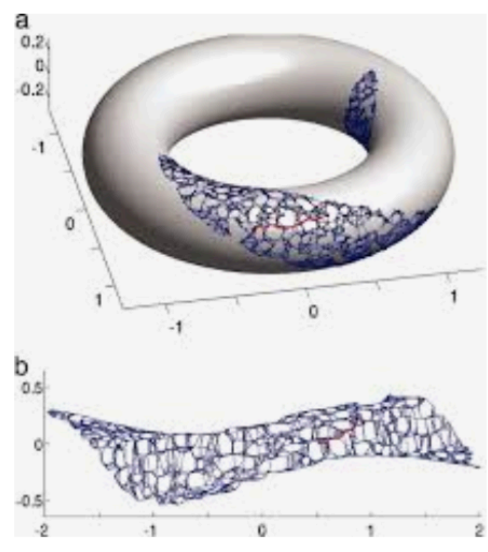
Why do we need nonlinear data analytics and why are they important?

- High dimensional data typically lives on or is near a low-dimensional manifold, but that manifold is not necessarily -- and usually not – linear!









Why Nonlinear data analysis or manifold learning can be important?

- High dimensional data typically lives on or is near a low-dimensional manifold, but that manifold is not necessarily -- and usually not – linear!
- Most of the new big data sets are coming generated by machine or people. They often have nonlinear relationships among them.
- Manifold Learning is relatively new and an exciting and important application of geometry to machine learning.
- There are a lot of theory behind the algorithm which can be developed for publications and for solving hard real world problems.
- Understanding nonlinear data analytics will benefit in applying algorithms more effectively.
  - E.g. Stock predicting/Algorithm trading
  - https://github.com/VivekPa/IntroNeuralNetworks

Many big data sets need to be analyzed by nonlinear data analysis, especially those generated by machines.

- *Where does big data come from?*

Organizations

Machines **E.g. Auto cars, UAVs, cell phone, other robots**

People

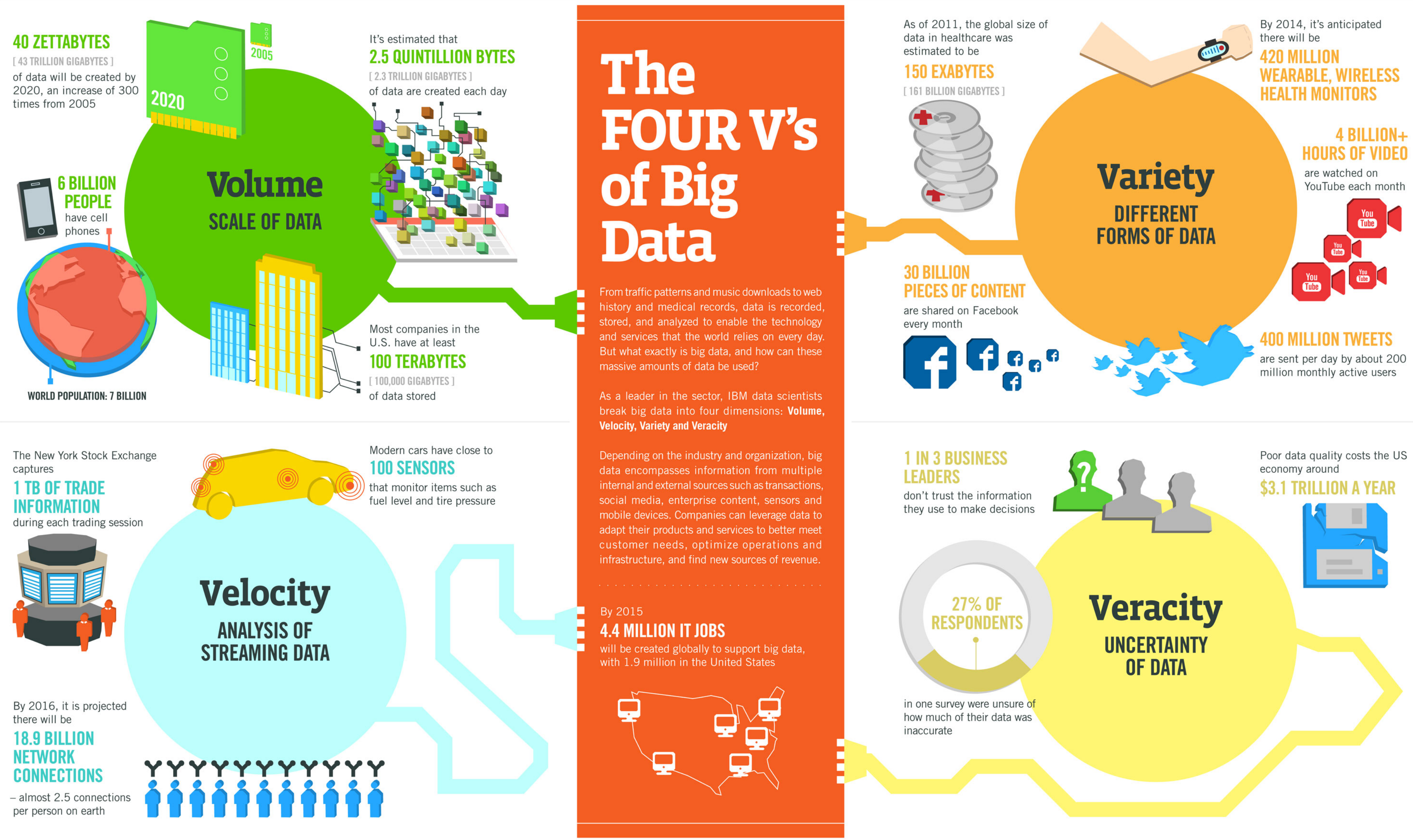
Data is not new. But the scale has been changed!
The way how people using data has been transformed!

Types of big data

1. Structured data (e.g. often Generated by organizations)
2. Semi-structured data (e.g. Generated by machine with manual records)
3. Unstructured data (often Generated by people)

- **What exactly is big data?**

- Does “big" here mean “big volume”?
- In fact, there are 5 “V”s to describe big data.
  - **Volume (Size)**
  - **Velocity (Speed)**
  - **Variety (Types)**
  - **Veracity (Quality)**
  - **Valence (Relationships)**



Sources: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPTEC, QAS

IBM

Data to Decision (D2D)

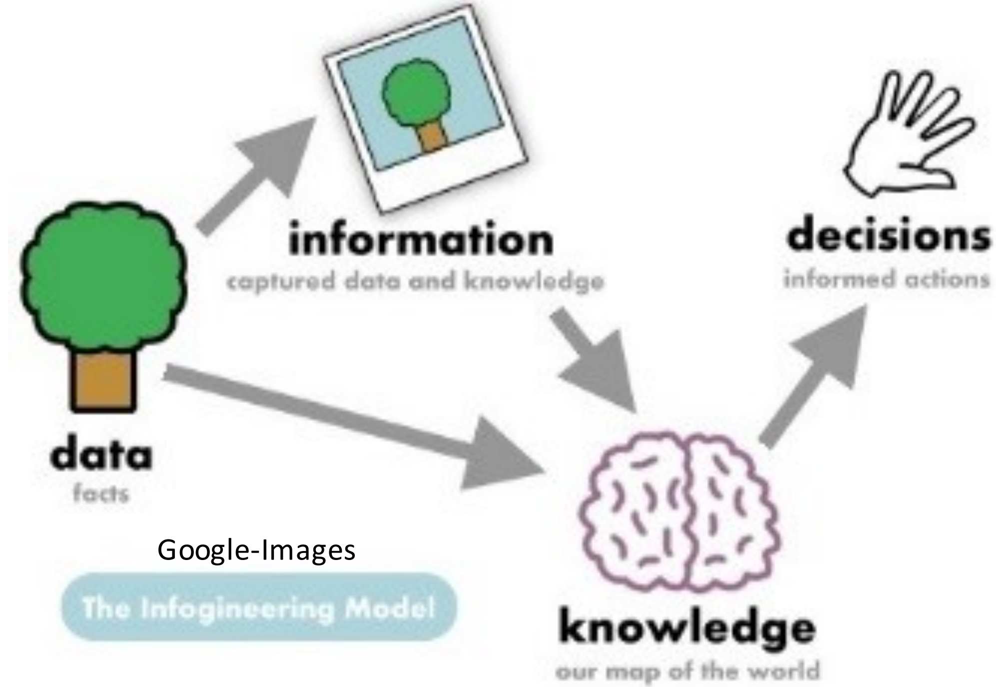
Volume

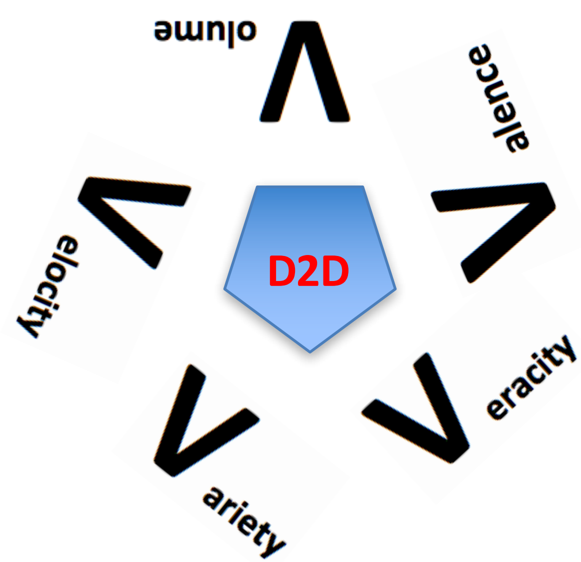
Velocity

Variety

Veracity

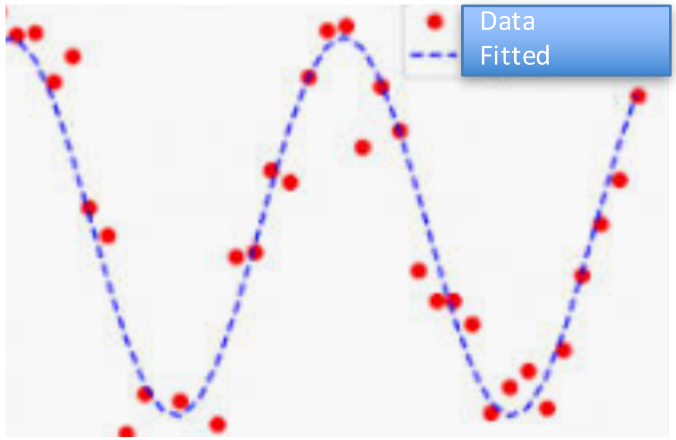
Valence





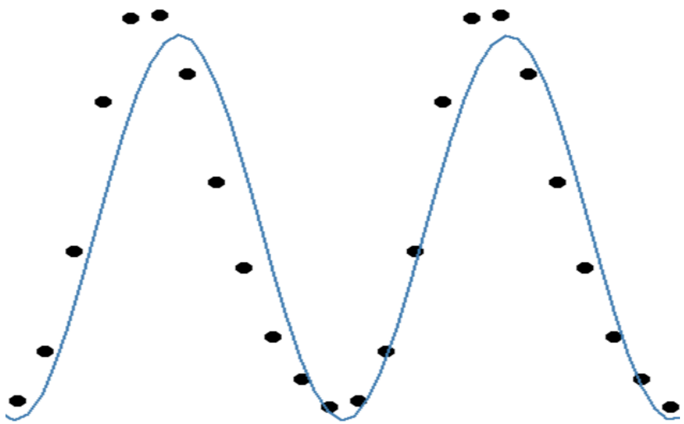
We need techniques of
Multivariate Data Analysis which analyzes and captures the nonlinear relationships among a given large data.

Toy Example of nonlinear relation among the data



← One of your team member get: Result of fitting **x**-accelerometer **data** of an auto car.

$$a_x(t) = 2\cos(t)$$



← Another your team member get: Result of fitting **y**-accelerometer **data** of the same auto car.

$$a_y(t) = 2\sin(t)$$

← You check: Both are on the same time scale.

You get a conclusion: the acceleration of the car is almost at constant 2.
Why?

- Recall: There are relations between $a_x(t)$ and $a_y(t)$.
- If we consider the set of all $(a_x(t), a_y(t))$, these are vectors live in $\mathbf{R}^2$.
- But in fact, the data lives on or close to the circle of radius 2.
- That is: there are nonlinear relations between $a_x(t)$ and $a_y(t)$:

$$(a_x(t))^2 + (a_y(t))^2 = 4$$

A circle is a simplest manifold.

Question: What if the data is so large that you can not see the nonlinear relationships among the data?

- There are relations between $a_x(t)$ and $a_y(t)$.
- If we consider the set of all $(a_x(t)\ ,\ a_y(t))$, these are vectors live in $\mathbf{R}^2$.
- But in fact, the data lives on or close to the circle of radius 2.
- That is there are nonlinear relations between

$a_x(t)$ and $a_y(t)$: $(a_x(t))^2 + (a_y(t))^2 = 4$

- A circle is a simplest manifold.

Question: Can we view as
$a_x(t)$ data as observations of a random variable X, and
$a_y(t)$ data as observations of a random variable Y? Then use the correlation of X and Y to detect the correlations between X and Y?

Answer: No!

Why? Because correlations only detect linear relations.

Recall: **Correlation**

- Correlation of two random variables are defined by "normalizing" the covariance of the two random variables.
- If we have a random vector, then we can define a covariance matrix.
- Covariance matrix is symmetric matrix, and in fact it is semi positive definite matrix.
- So the covariance matrix can be diagonalized, with eigenvalues being none negative; which is the base for PCA.

Covariance, and Covariance Matrix

- The **covariance** between two rv's X and Y measures the degree to which X and Y are (linearly) related; defined as

$$\text{cov}[X,Y] \triangleq \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])]$$

Exercise

$$= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

If **x** is a d-dimensional random vector, its **covariance matrix** is defined to be the following symmetric, semi positive definite matrix:

$$\text{cov}[\mathbf{x}] \triangleq \mathbb{E}\left[(\mathbf{x}-\mathbb{E}[\mathbf{x}])(\mathbf{x}-\mathbb{E}[\mathbf{x}])^T\right]$$

$$= \begin{pmatrix} \text{var}[X_1] & \text{cov}[X_1,X_2] & \cdots & \text{cov}[X_1,X_d] \\ \text{cov}[X_2,X_1] & \text{var}[X_2] & \cdots & \text{cov}[X_2,X_d] \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}[X_d,X_1] & \text{cov}[X_d,X_2] & \cdots & \text{var}[X_d] \end{pmatrix}$$

Often denoted by Σ

correlation coefficient & correlation matrix

- The (Pearson) **correlation coefficient** between two rvs X and Y is defined as

$$\operatorname{corr}[X, Y] \triangleq \frac{\operatorname{cov}[X, Y]}{\sqrt{\operatorname{var}[X] \operatorname{var}[Y]}}$$

- If X and Y are indep., then cov [X, Y] = 0; say X and Y are uncorrelated.

- A **correlation matrix** of a random vector has the form:

$$\mathbf{R} = \begin{pmatrix} \operatorname{corr}[X_1, X_1] & \operatorname{corr}[X_1, X_2] & \cdots & \operatorname{corr}[X_1, X_d] \\ \vdots & \vdots & \ddots & \vdots \\ \operatorname{corr}[X_d, X_1] & \operatorname{corr}[X_d, X_2] & \cdots & \operatorname{corr}[X_d, X_d] \end{pmatrix}$$

Exercise: show that $-1 \le \operatorname{corr}[X, Y] \le 1$ and

Show that $\operatorname{corr}[X,Y] = 1$ iff $Y = aX + b$ for some parameters a and b.

Example of Correlation Coefficients

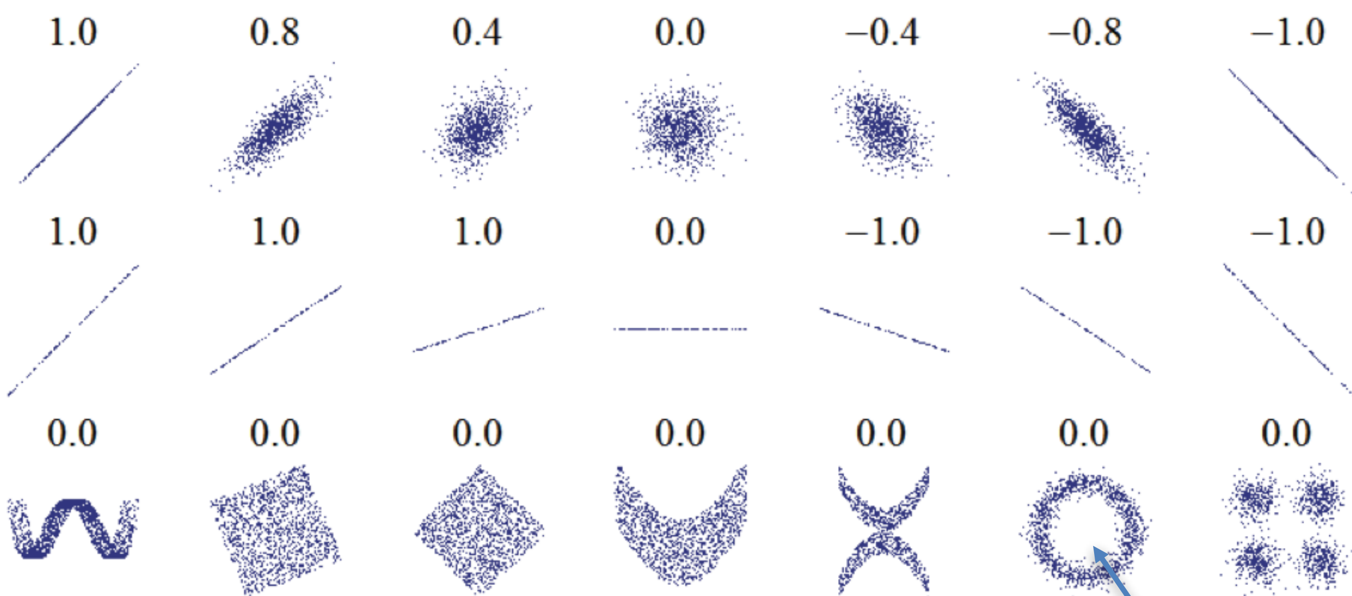


Figure 2.12 Several sets of (x, y) points, with the correlation coefficient of x and y for each set. Note that the correlation reflects the noisiness and direction of a linear relationship (top row), but not the slope of that relationship (middle), nor many aspects of nonlinear relationships (bottom). N.B.: the figure in the center has a slope of 0 but in that case the correlation coefficient is undefined because the variance of Y is zero. Source: `http://en.wikipedia.org/wiki/File:Correlation_examples.png`

E.g. It did not detect the data living close to a circle.

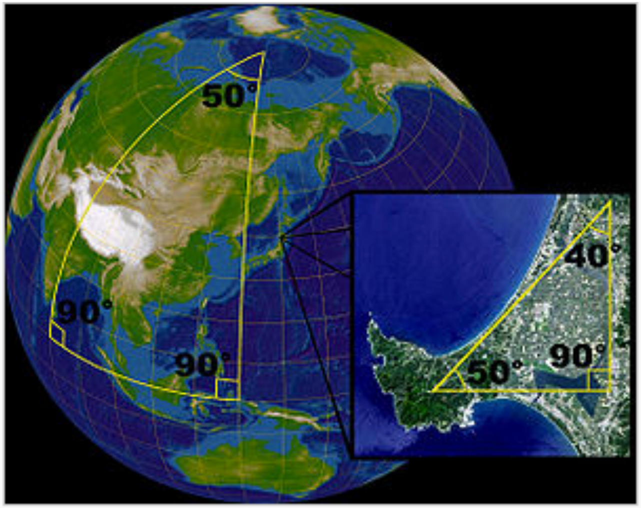
Multivariate Data Analysis

- When data is big, We have little visual guidance to help us identify any meaningful low- dimensional structure hidden in high-dimensional data.
- The linear PCA can be extremely useful in discovering low-dimensional structure when the data actually lie in a linear (or approximately linear) lower-dimensional subspace.
- But what if the data lives or nearly a nonlinear curved space (called a manifold) M in $\mathbf{R}^N$, whose structure and dimensionality are both assumed unknown?
- Our goal of dimensionality reduction then becomes one of identifying the nonlinear manifold in question. The problem of recovering that manifold is known as nonlinear manifold learning.

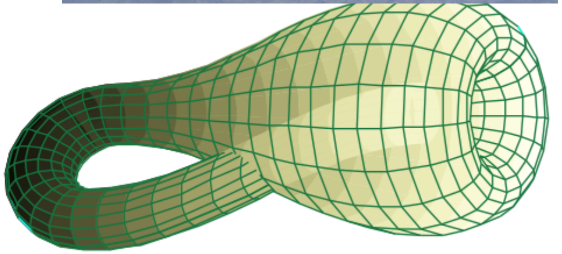
- Therefore it is crucial to understand nonlinear data analytics or manifold learning…

What is a manifold?

- An n-dimensional manifold locally “looks like” a piece of $\mathbf{R}^n$.
- For examples, sphere and torus.
- Key features of a **manifold**: **curved**

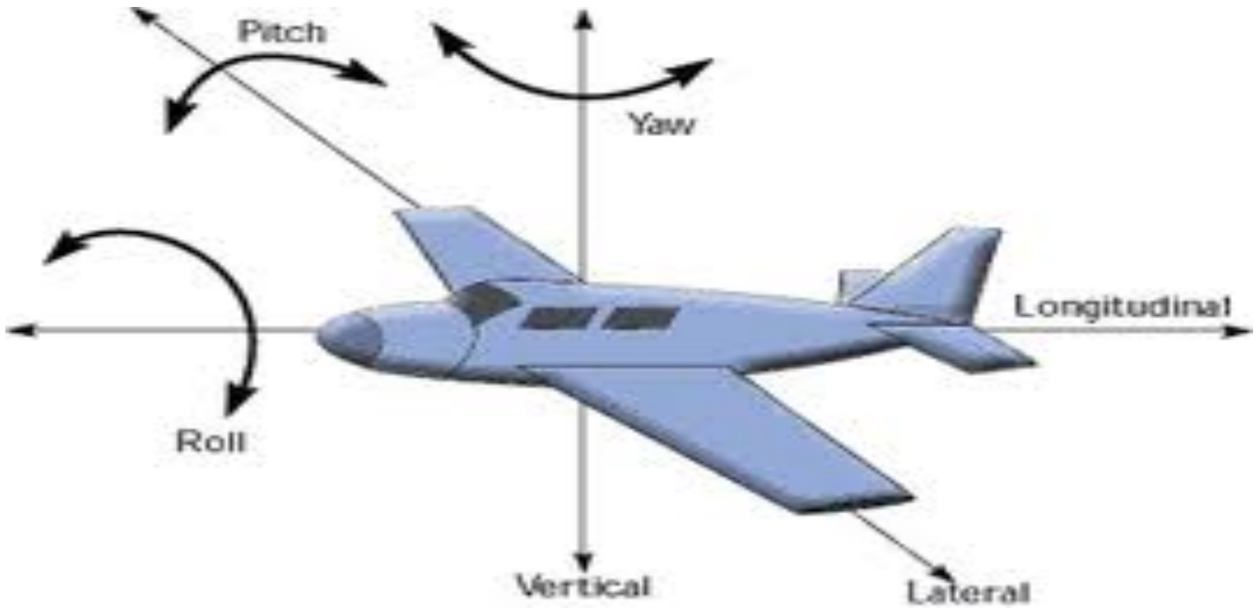


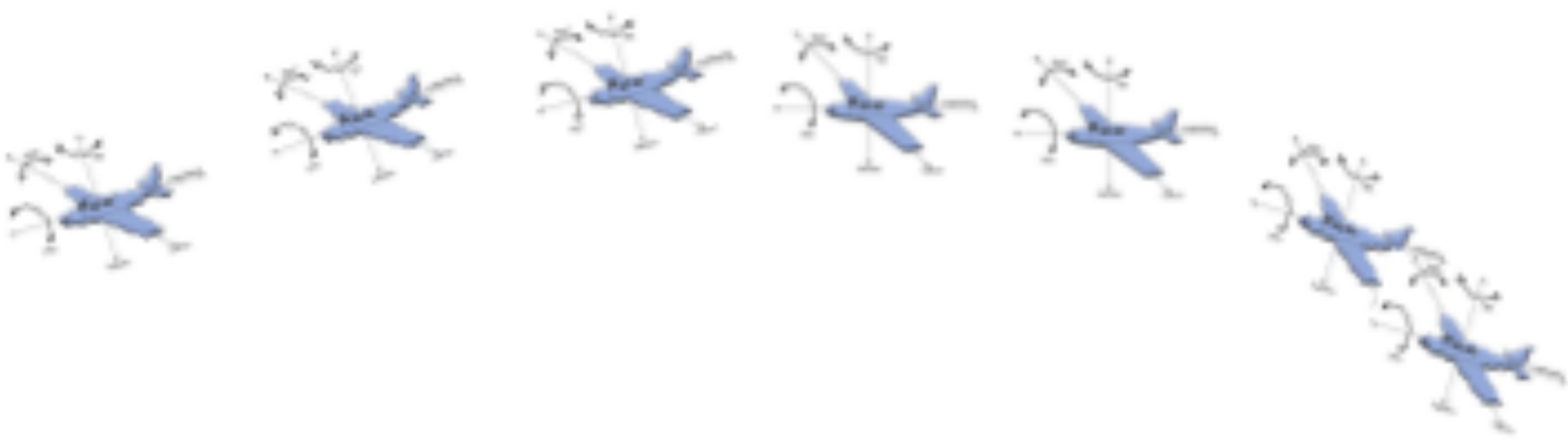
The sphere (surface of a ball) is a two-dimensional manifold since it can be represented by a collection of two-dimensional maps.



- Only manifolds can capture UAV’s dynamical behaviors

- **How to model and capture the dynamics and kinematics of an UAV?**





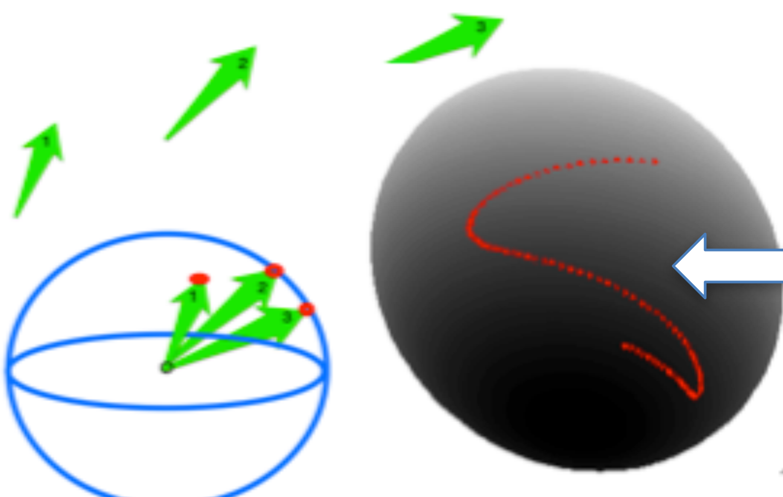
You may wonder: How to use manifold to study UAV data?
Simplest case: drawing a curve on a sphere

Try to capture characteristics of flight controls

- **For example: Only look at UAV "headings"**
- **All possible headings for all UAVs form a sphere.**

Only consider UAV heading directions here, but works for any other UAV characteristics

- ***Key: Developed a dimension-reduction technique for large nonlinear data.***



Just recording the heading while a UAV is flying gives a heading-behavior curve.

Overview of nonlinear analytic techniques

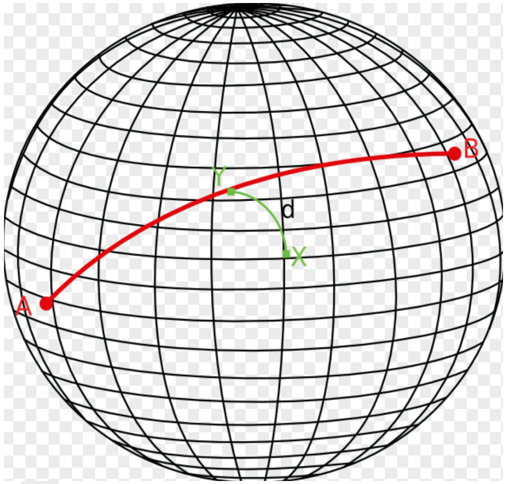
- One of the existing powerful dimension reduction methods is the Principal Component Analysis (PCA). (In fact, it is a Linear PCA).
- Later in this course, we will extend linear PCA to None Linear PCA.
- We will transform nonlinear items to linear items, and the use Machine Learning methods in linear space and them map them back. (e.g. Log and Exponential Maps)
- Kernel methods
- ISOMAP
- …..

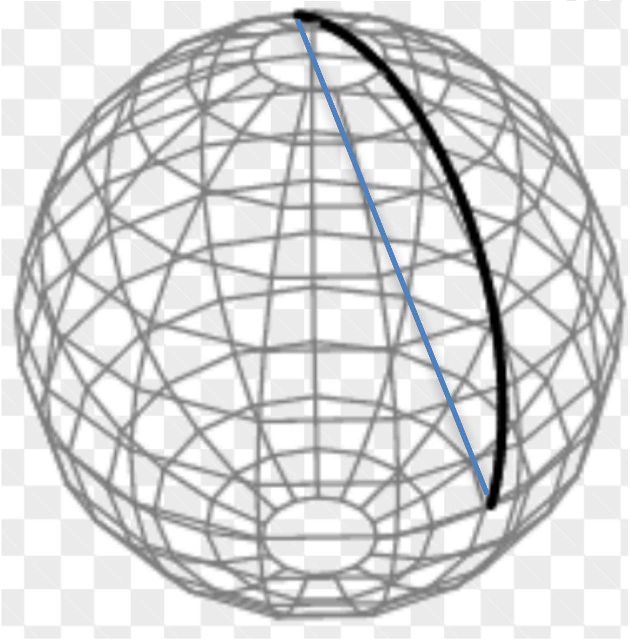
Overview of Lecture 1

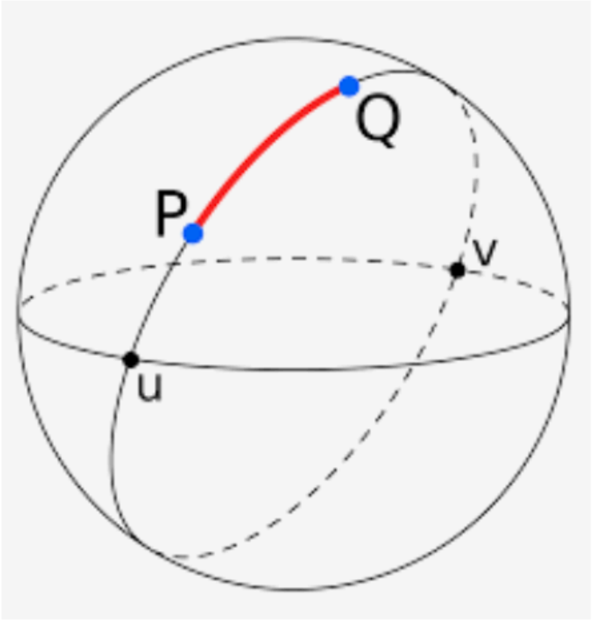
- Why we need nonlinear data analysis?
  - First starting with curves and their analysis
- **Similarity measurements for nonlinear data**
  - **First a few examples: Arc-length, Geodesic length**
- Introduction to cell phone data
- Introduction to rigid motion

Similarity measurements for nonlinear data

Concept of manifold and nonlinear Euclidean distance







Another examples

- UAV Mishap Analysis
- Anomaly Detections in UAV systems

Example:

Identified various causes affecting UAV behaviors for anomaly detection in UAV systems

Human behaviors | *UAV-Health & Status* | *Lost GPS or Communications* | *Environmental conditions* | *. . .* | *Cyber attacks*

Example: The causes of this mishap

1) Engine overheat*: coolant line leaking*

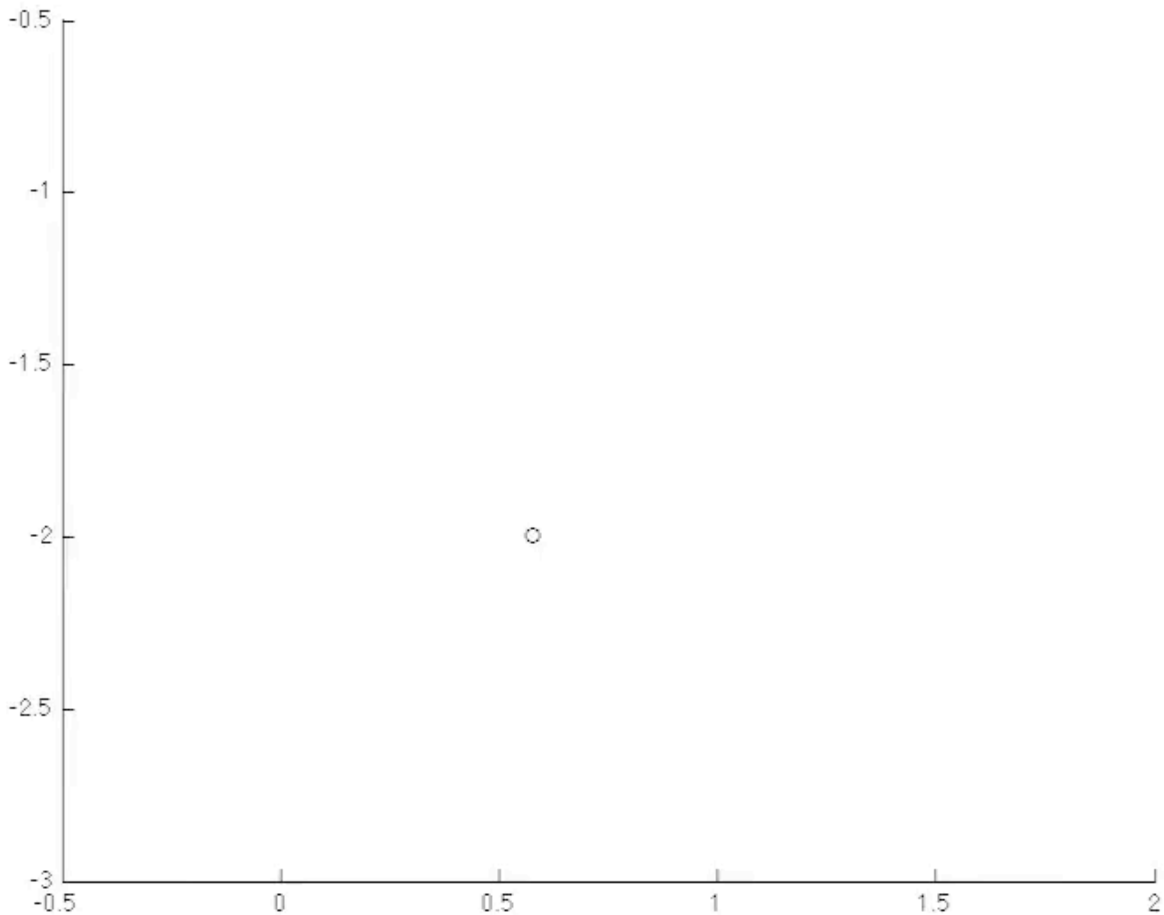
2) Lost control: *human error*

Lessen Learned: Many mishaps resulted from **combined causes** but ***no metric*** *for a combination of anomaly behaviors!*

How could nonlinear data analysis be useful here?

A simple example

We use math to model behaviors of an UAV:



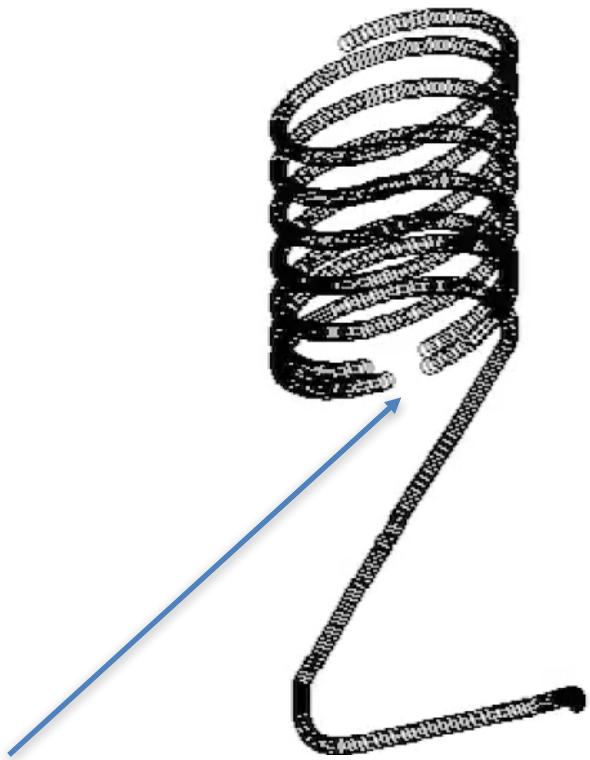
Imagine an UAV is just a point as in the video.

This example uses the true data from an UAV instructor on how to control an UAV climb up.

The curve represents a trajectory of the UAV the student is controlling.

Our method could fix issues such as missing data.

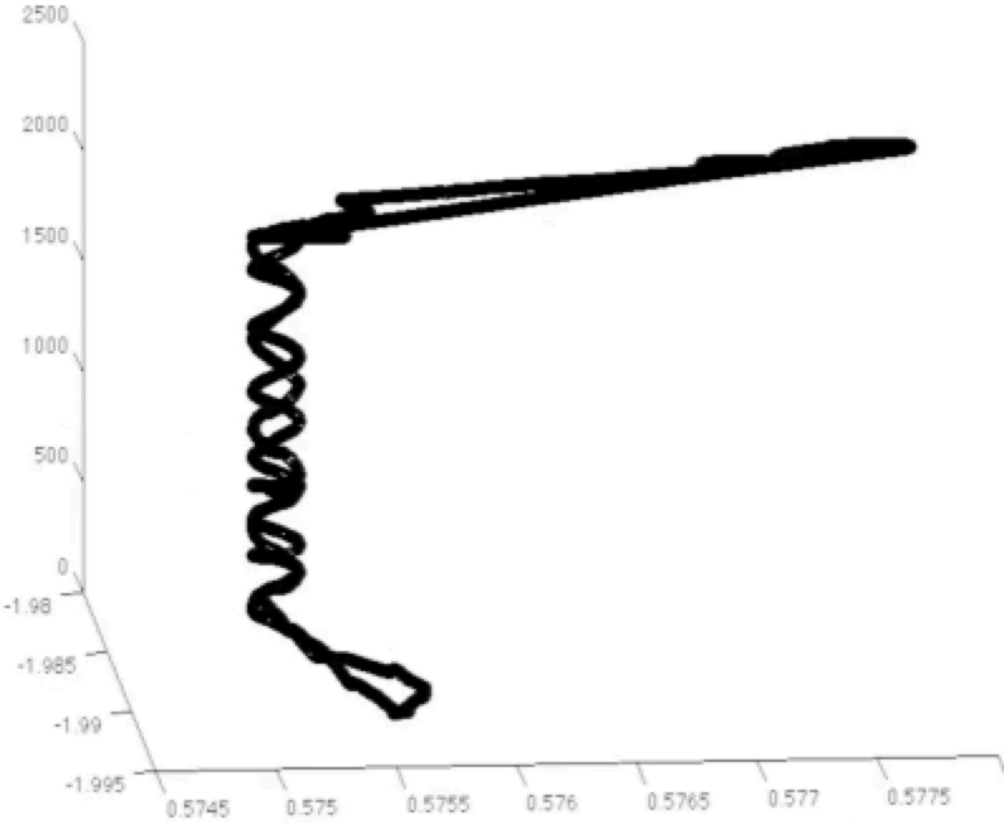
How to fill the missing data here?

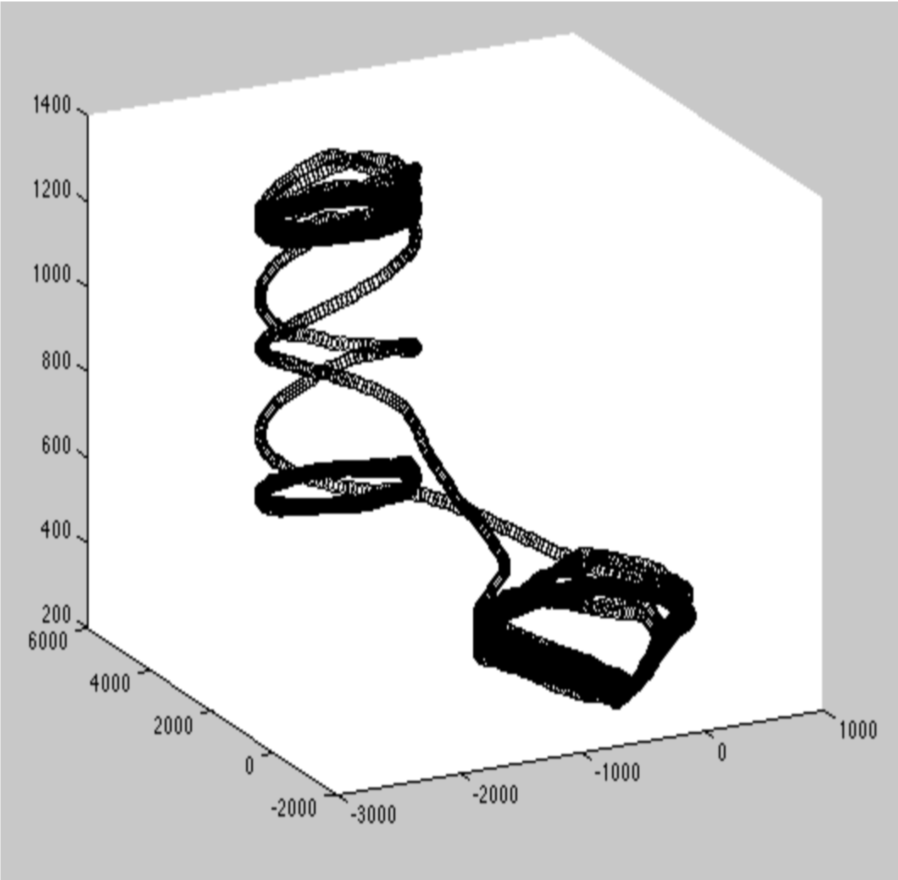


- This is just because of missing data.
- We can confirm it by the dynamics and kinematics of an UAV.

- Linear interpolation here may not make sense.
- We need to use other parts of the trajectory to predict how this missing part should look like.
- We need mathematically describe the trajectory.
- What kind of curve best describe the trajectory?

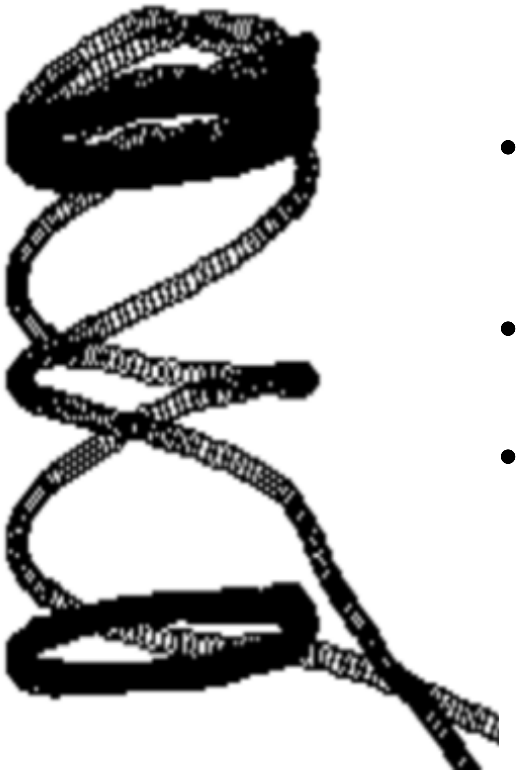
Here are the trajectory of Student1 and Student2





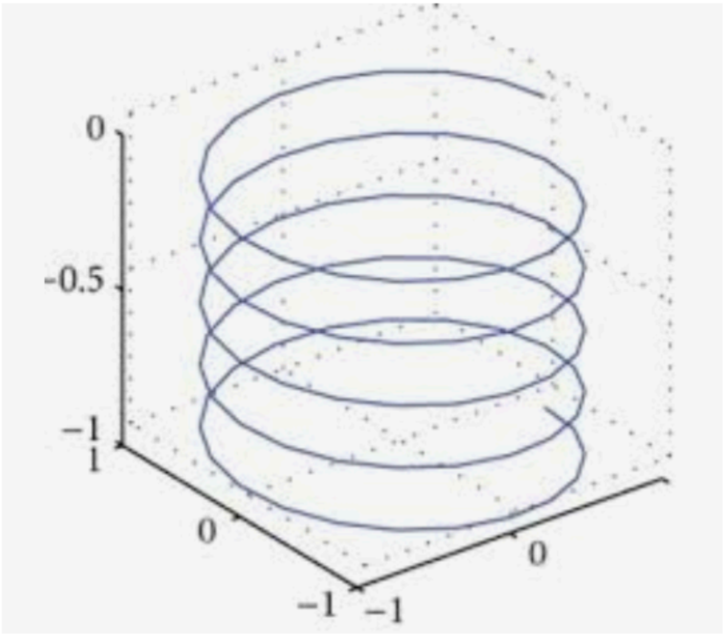
Q: What are the differences between the instructor's trajectory and that of the students?

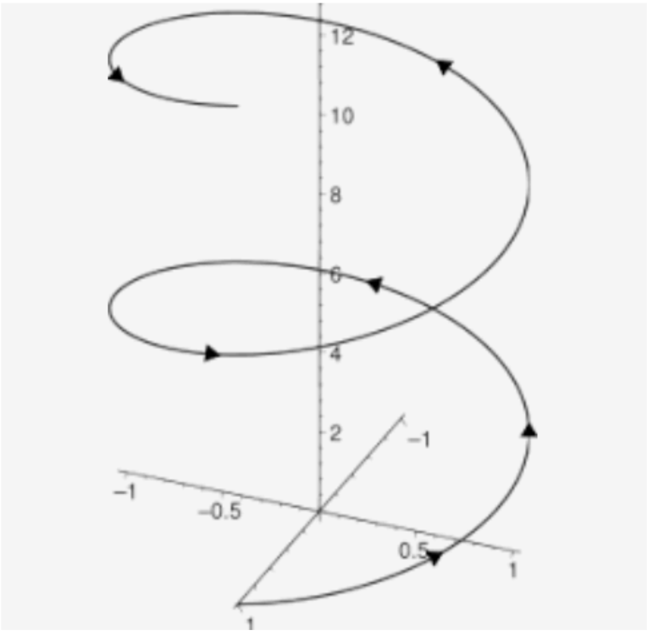
Compare with the instructor's trajectory with that of the student



- What kind of differences you have seen?
- How to describe the dissimilarity?
- Need non Euclidean metrics.

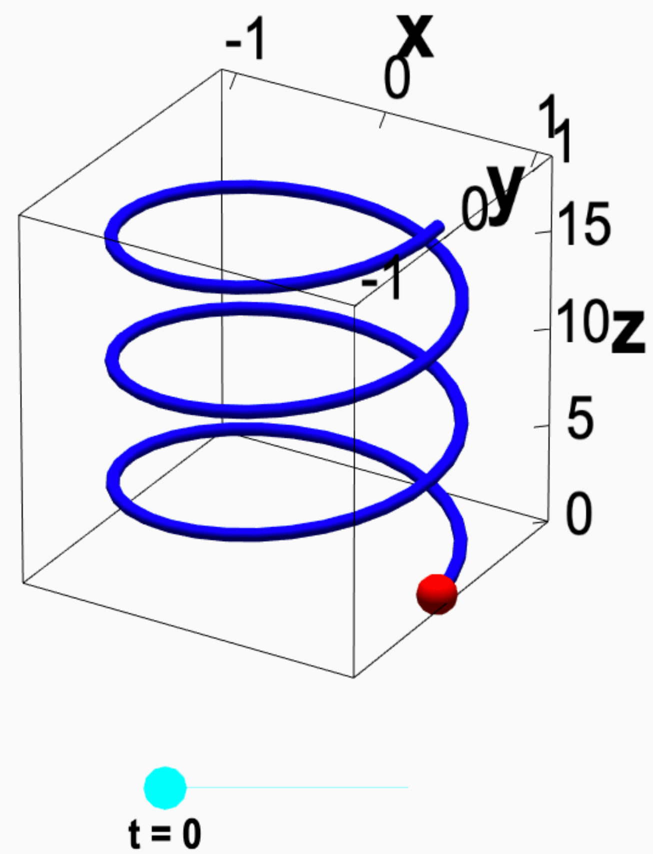
Recall: **Helix Curve**





- How to describe a helix curve?
- They could have different orientations!

You could think of a curve $\mathbf{c} : \mathbf{R} \to \mathbf{R}^3$ as being a wire. For example, $\mathbf{c}(t) = (\cos t, \sin t, t)$, for $0 \le t \le 6\pi$, is the parametrization of a helix. You can view it as a slinky or a spring.



Parametrized helix. The vector-valued function $\mathbf{c}(t) = (\cos t, \sin t, t)$ parametrizes a helix, shown in blue. This helix is the image of the interval $[0, 6\pi]$ (shown in cyan) under the mapping of $\mathbf{c}$. For each value of t, the red point represents the vector $\mathbf{c}(t)$. As you change t by moving the cyan point along the interval $[0, 6\pi]$, the red point traces out the helix.

In general: Parametrized Curve

Parametrized and Regular Curves

Definition

A *parametrized differentiable curve* is a differentiable map $\alpha : I \to \mathbb{R}^3$ of an open interval $I = (a, b)$ of the real line $\mathbb{R}$ into $\mathbb{R}^3$.

Definition

A parametrized differentiable curve $\alpha : I \to \mathbb{R}^3$ is said to be *regular* if $\alpha'(t) \neq 0$ for all $t \in I$.

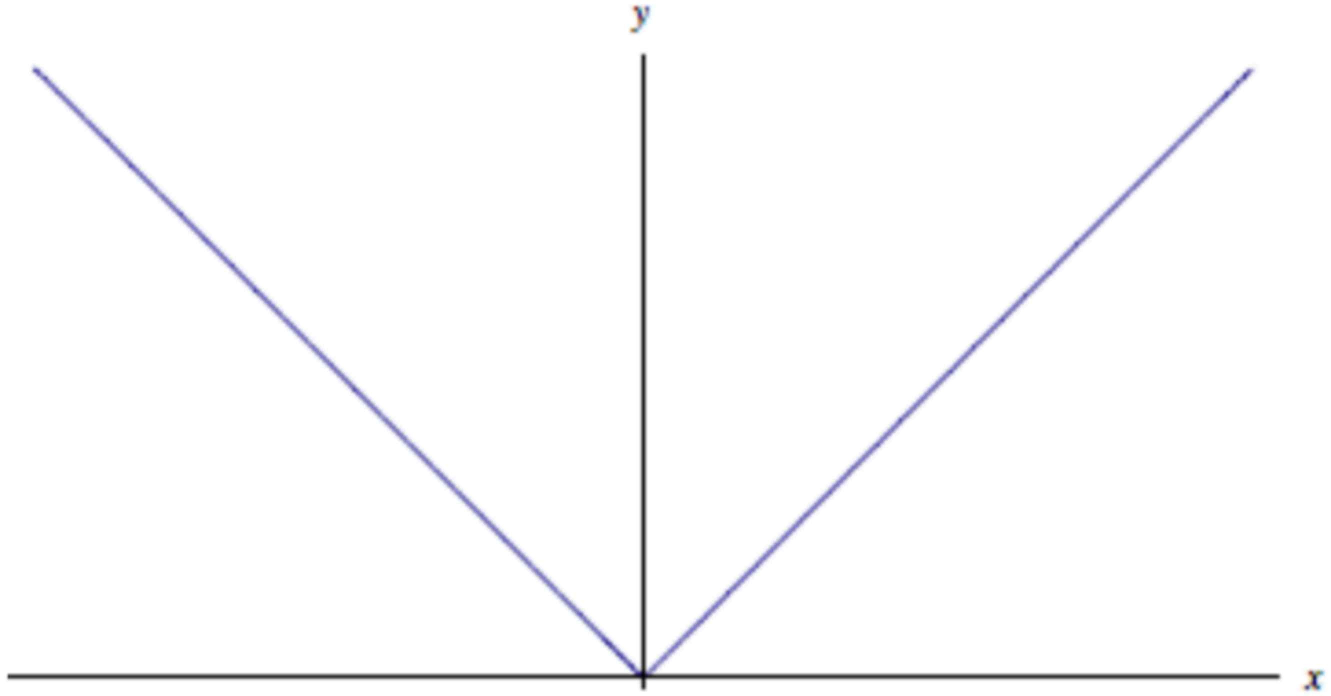
Definition

We say that $s \in I$ is a *singular point of order 1* if $\alpha''(s) = 0$ (in this context, the points where $\alpha'(s) = 0$ are called singular points of order 0).

Examples

Example 1

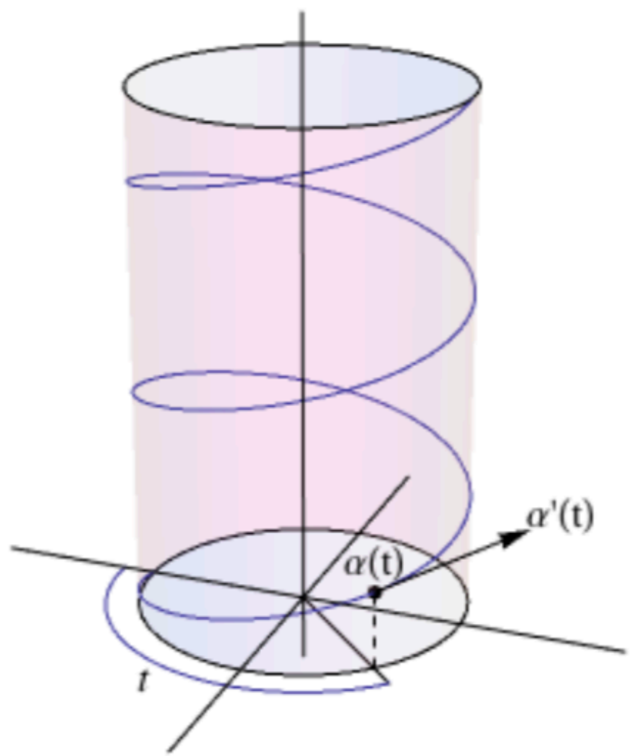
The map $\alpha : \mathbb{R} \to \mathbb{R}^2$ given by $\alpha(t) = (t, |t|), t \in \mathbb{R}$ (not differentiable).



Examples

Example 2

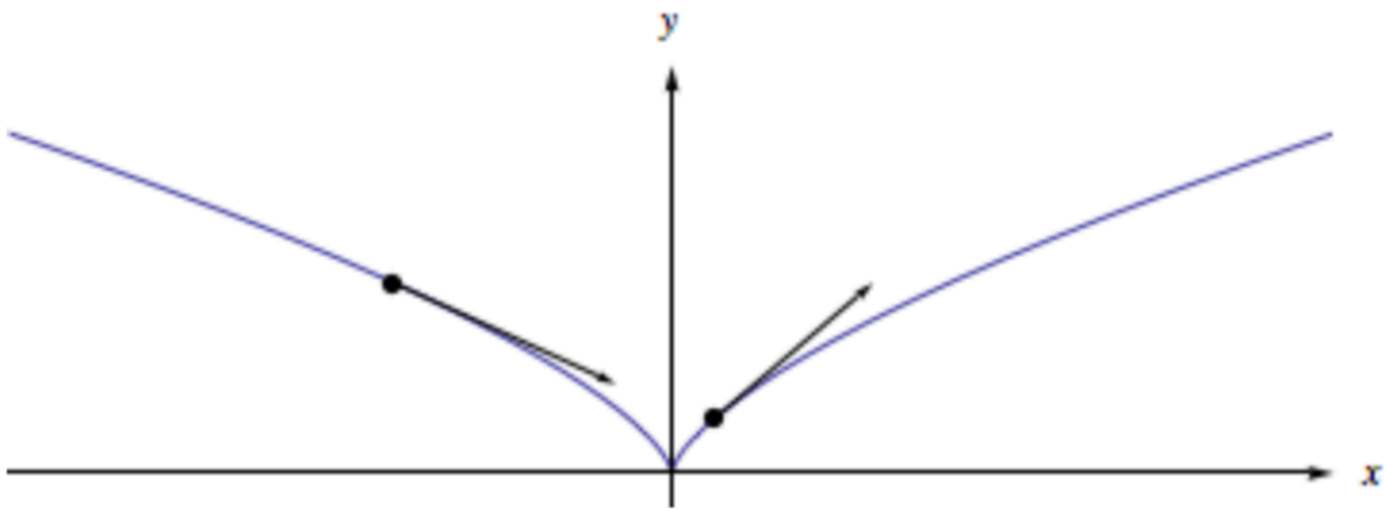
A helix of pitch $2\pi b$ on the cylinder $x^2 + y^2 = a^2$.



Examples

Example 3

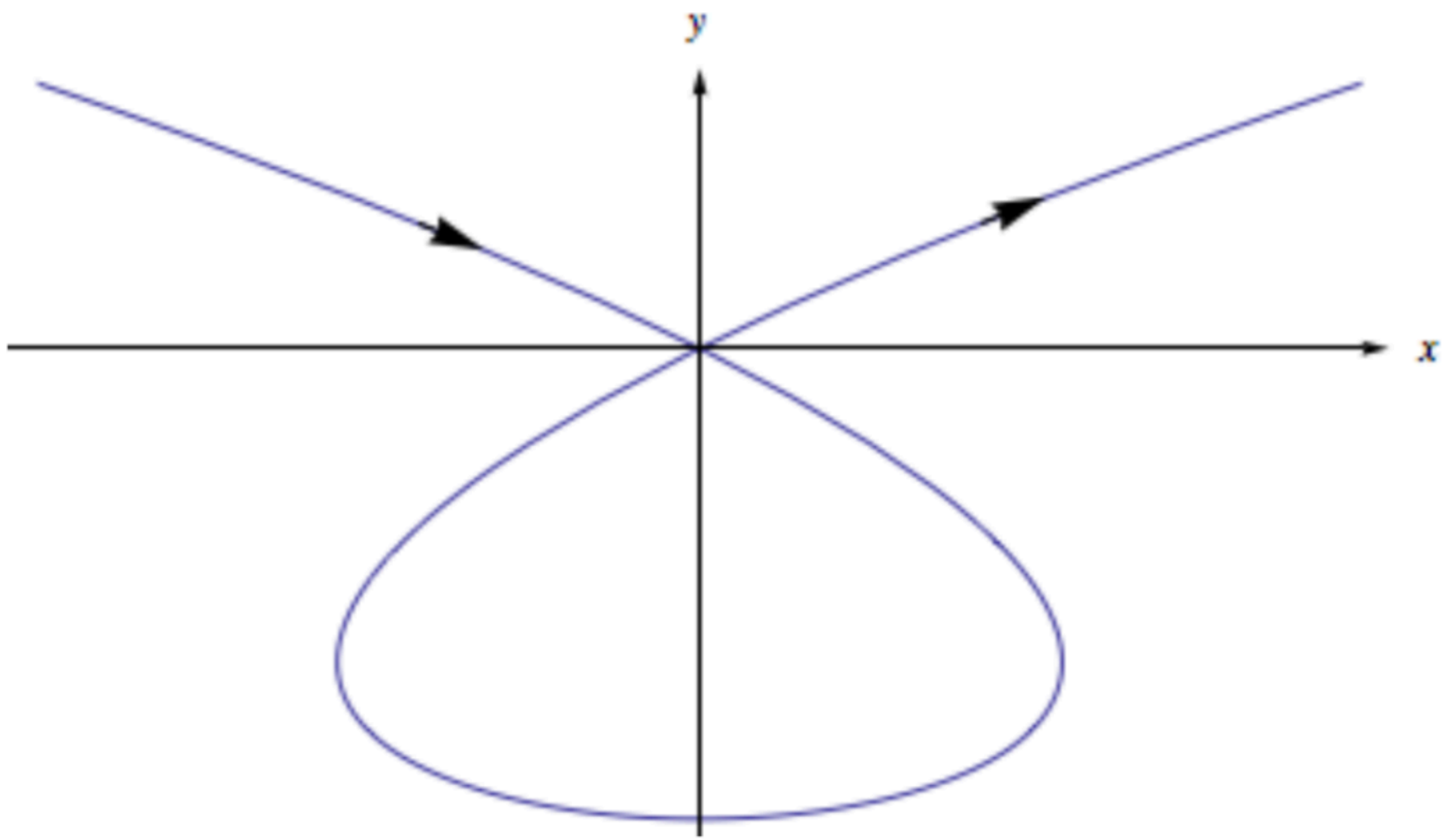
The map $\alpha : \mathbb{R} \to \mathbb{R}^2$ given by $\alpha(t) = (t^3, t^2), t \in \mathbb{R}$.



Examples

Example 4

The map $\alpha : \mathbb{R} \to \mathbb{R}^2$ given by $\alpha(t) = (t^3 - 4t, t^2 - 4), t \in \mathbb{R}$.



Arc Length of a Curve

Definition

Given $t \in I$, the *arc length* of a regular parametrized curve $\alpha : I \to \mathbb{R}^3$, from the point t_0, is by definition

$$s(t) = \int_{t_0}^{t} \|\alpha'(t)\| \, dt,$$

where

$$\|\alpha'(t)\| = \sqrt{(x'(t))^2 + (y'(t))^2 + (z'(t))^2}$$

is the length of the vector $\alpha'(t)$.

Definition

A parametrized curve $\alpha : I \to \mathbb{R}^3$ is said to be parametrized by arc length if $\|\alpha'(t)\| = 1$ (that is, if α has unit speed) for all $t \in I$.

Parametrization by Arc Length

Proposition (Geometric meaning of above definition)

A curve $\alpha : I \to \mathbb{R}^3$ is parametrized by arc length if and only if the parameter t is the arc length of α measured from some point.

Proof.

□

Proposition (Advantages of $\|\alpha'(s)\| = 1$)

Let $\alpha : I \to \mathbb{R}^3$ be a curve parametrized by arc length. Then $\alpha''(s)$ is orthogonal to $\alpha'(s)$ for all $s \in I$.

Proof.

□

Reparametrization by Arc Length

Example

Consider the helix $\alpha : \mathbb{R} \to \mathbb{R}^3$ given by $\alpha(t) = (\cos t, \sin t, t)$.

- From now on, we are going to assume curves are parametrized by arc length.

Curvature

Geometric Meaning

Let $\alpha : I = (a, b) \to \mathbb{R}^3$ be a curve parametrized by arc length s. Since the tangent vector $\alpha'(s)$ has unit length, the norm $\|\alpha''(s)\|$ of the second derivative measures the rate of change of the angle which neighboring tangents make with the tangent at s. $\|\alpha''(s)\|$ gives, therefore, a measure of how rapidly the curve pulls away from the tangent line at s, in a neighborhood of s.

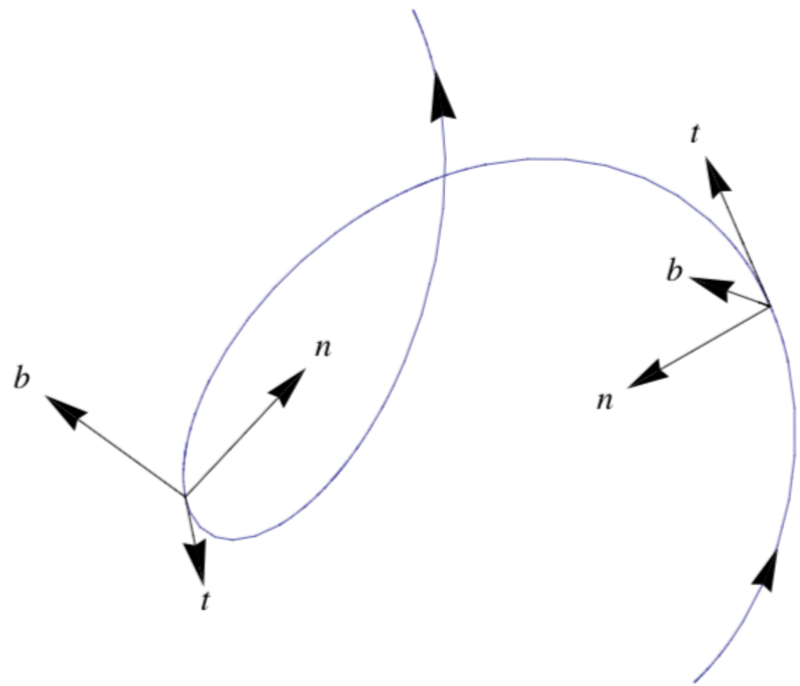
Definition

Let $\alpha : I \to \mathbb{R}^3$ be a curve parametrized by arc length $s \in I$. The number $\|\alpha''(s)\| = k(s)$ is called the *curvature* of α at s.

Torsion

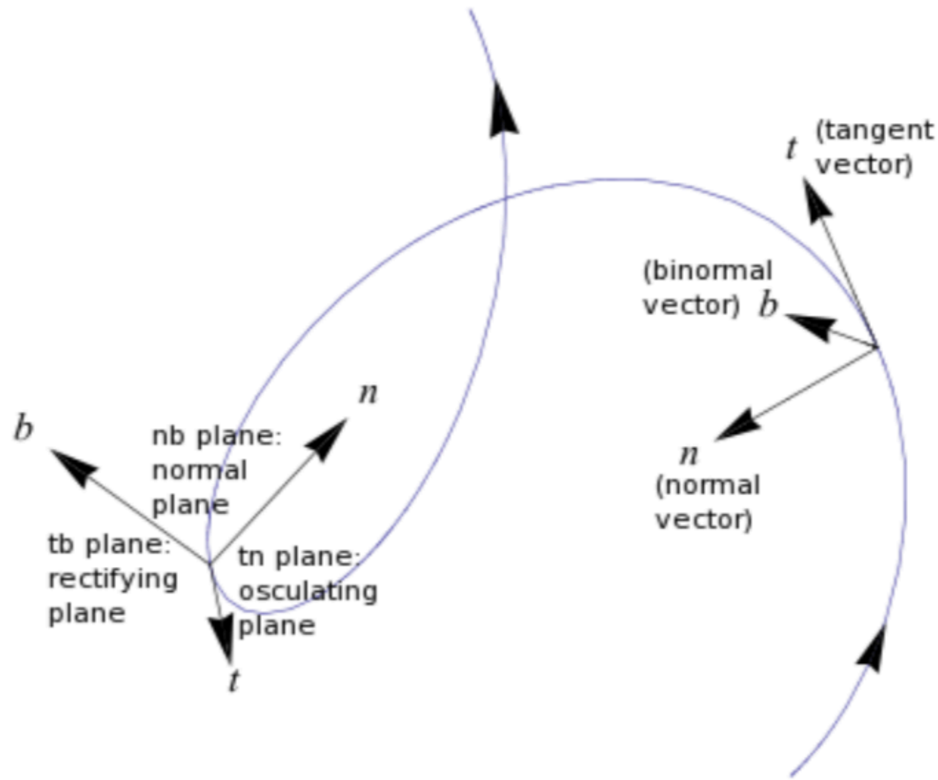
Geometric Meaning

Since $b(s)$ is a unit vector, the length $\|b'(s)\|$ measures the rate of change of the neighboring osculating planes with the osculating plane at s; that is $b'(s)$ measures how rapidly the curve pulls away from the osculating plane at s, in a neighborhood of s.



Frenet Frame

$$\alpha'(s) \stackrel{\triangle}{=} t(s)$$
$$\alpha''(s) = k(s)n(s)$$
$$t(s) \wedge n(s) = b(s)$$



Fundamental Theorem of the Local Theory of Curves

Theorem

Given differentiable functions $k(s) > 0$ *and* $\tau(s), s \in I$*, there exists a regular parametrized curve* $\alpha : I \to \mathbb{R}^3$ *such that* s *is the arc length,* $k(s)$ *is the curvature, and* $\tau(s)$ *is the torsion of* α *Moreover, any other curve* $\overline{\alpha}$ *satisfying the same conditions differs from* α *by a rigid motion; that is, there exists an orthogonal map* ρ *of* $\mathbb{R}^3$*, with positive determinant, and a vector* c *such that* $\overline{\alpha} = \rho \circ \alpha + c$*.*

Proof of uniqueness.

<u>Claim:</u> arc length, curvature, and torsion are invariant under the rigid motion. □

Techniques in Geometric Analysis:

Example

Assume that all normals of a parametrized curve pass through a fixed point. Prove that the trace of the curve is contained in a circle.

Homework: Rewrite all the proofs in the example.

Note: Homework will be given in the lecture.

Homework problems

- Problem A

Let $\alpha(t)$ be a parametrized curve which does not pass through the origin. If $\alpha(t_0)$ is the point of the trace of α closest to the origin and $\alpha'(t_0) \neq 0$, show that the position vector $\alpha(t_0)$ is orthogonal to $\alpha'(t_0)$.

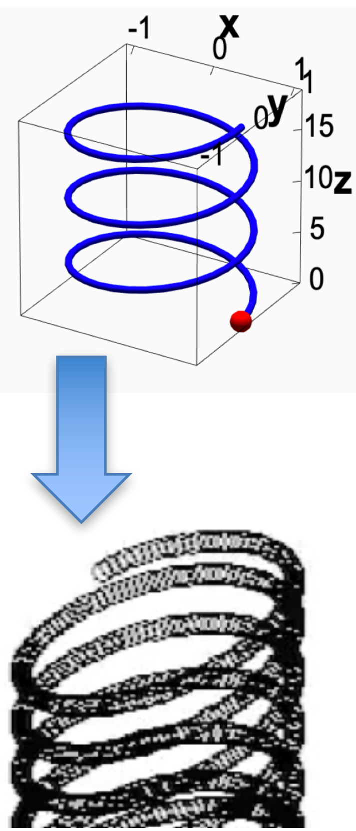
Homework problems

- Problem B

Show that the set of rigid motions forms a group.

Creative activity - Extra Credit

- How to create a transformation from the data on some helix to the data of the instructor’s trajectory?
- Review different operators in R^2, e.g. we have shear map below. Here we want to shear a curve! For more info:



https://en.wikipedia.org/wiki/Transformation_matrix

For shear mapping (visually similar to slanting), there are two possibilities.

A shear parallel to the *x* axis has $x' = x + ky$ and $y' = y$. Written in matrix form, this becomes:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

A shear parallel to the *y* axis has $x' = x$ and $y' = y + kx$, which has matrix form:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

Overview of Lecture 1

- Why we need nonlinear data analysis?
  - First starting with curves and their analysis
- Similarity measurements for nonlinear data
  - First a few examples: Arc-length, Geodesic length
- **Introduction to cell phone data**
- Introduction to rigid motion

Introduction of Cell Phone Data

- There are a lot of data sets available online
- For examples:
- 1. HMOG data set: http://www.cs.wm.edu/~qyang/hmog.html

Rotation Data

Rotation data is returned as a Euler angle, representing the number of degrees of difference between the device coordinate frame and the Earth coordinate frame.

Alpha

The rotation around the z axis. The `alpha` value is 0° when the top of the device is pointed directly north. As the device is rotated counter-clockwise, the `alpha` value increases.

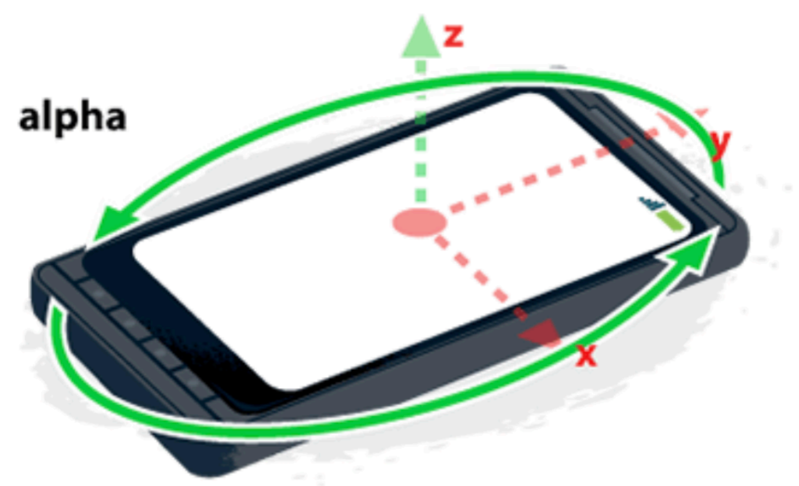


Illustration of alpha in the device coordinate frame

Beta

The rotation around the x axis. The `beta` value is 0° when the top and bottom of the device are equidistant from the surface of the earth. The value increases as the top of the device is tipped toward the surface of the earth.

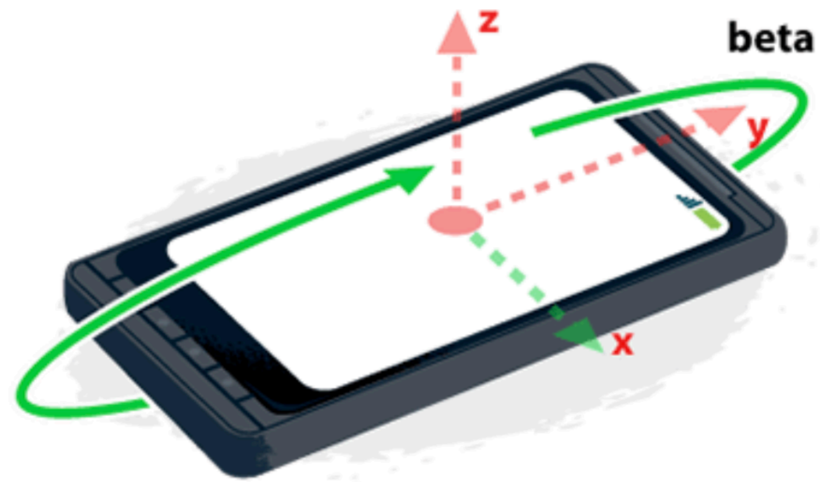


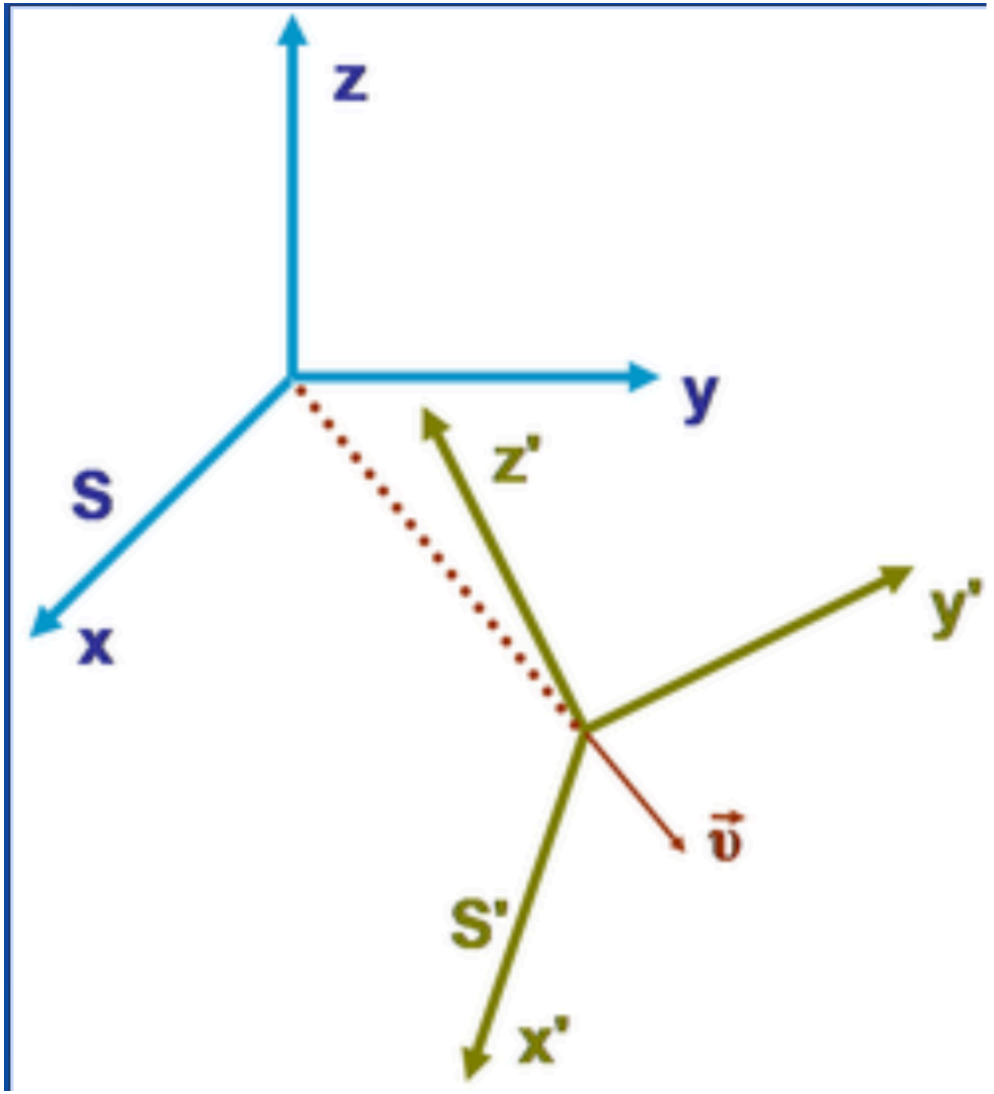
Illustration of beta in the device coordinate frame

Gamma

The rotation around the y axis. The `gamma` value is 0° when the left and right edges of the device are equidistant from the surface of the earth. The value increases as the right side is tipped towards the surface of the earth.

Illustration of gamma in the device coordinate frame

Concept of Moving Frame



Real world Application

- **Using cell phone data to authenticate users.**
- Very hard problem and lots of math involved

H-MOG Data Set: A Multimodal Data Set for Evaluating Continuous Authentication Performance in Smartphones

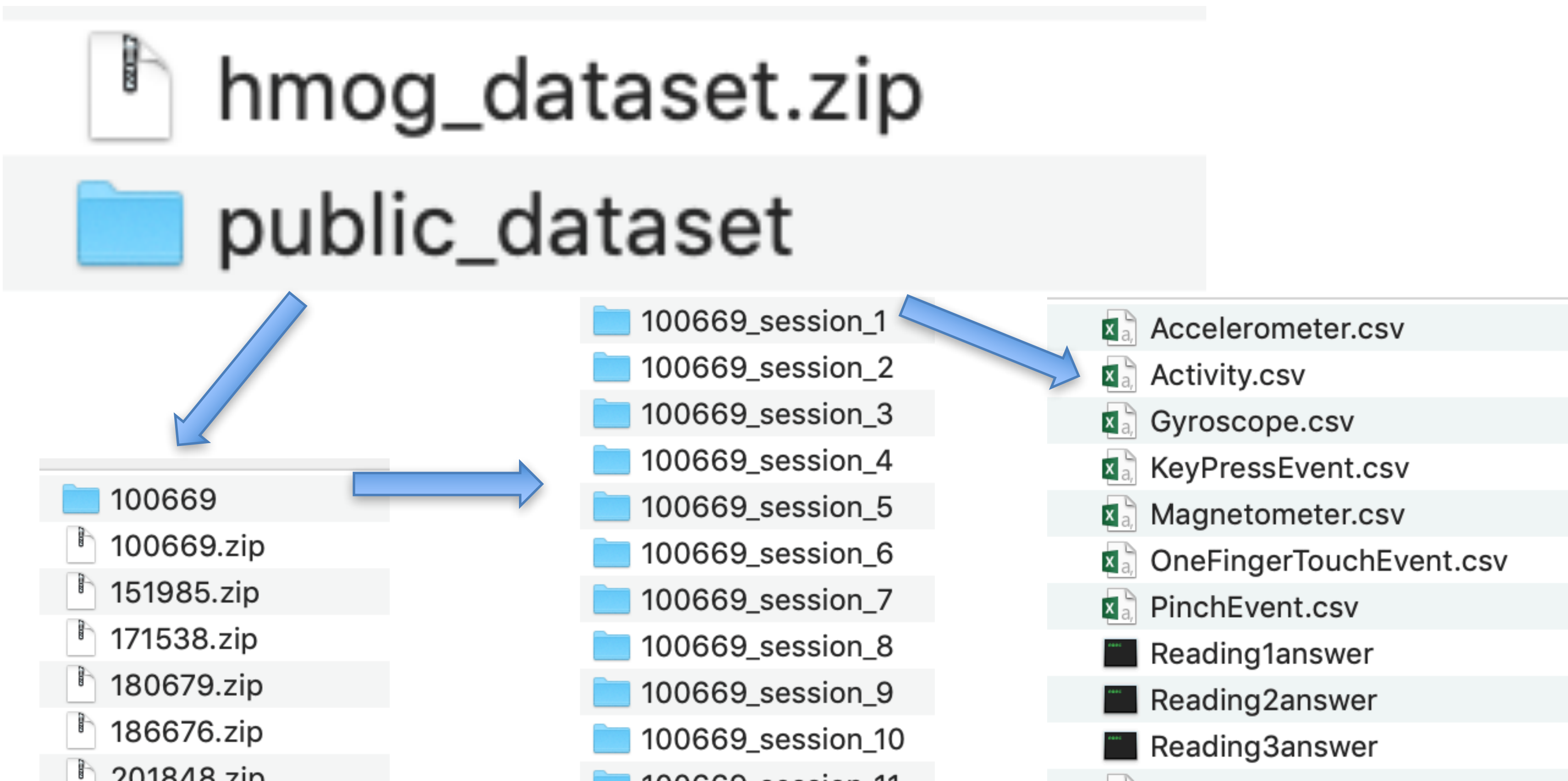
Qing Yang, Ge Peng, David T. Nguyen, Xin Qi, Gang Zhou (*Colleget of William and Mary*)
Zdeňka Sitová (*New York Institute of Technology*; *Masaryk University*)
Paolo Gasti, Kiran S. Balagani (*New York Institute of Technology*)

1. Introduction

We performed a large-scale user study to collect a wide spectrum of signals about user behaviors on smartphones, including touch, gesture, and pausality of the user, as well as movement and orientation of the phone. This dataset has been used to evaluate a continuous authentication modality named H-MOG in smartphones. A detailed description of this dataset and its application is in our poster paper (PDF) in ACM SenSys'14. The H-MOG paper using this dataset is published on IEEE Transactions on Information Forensics and Security (link on IEEE Xplore).

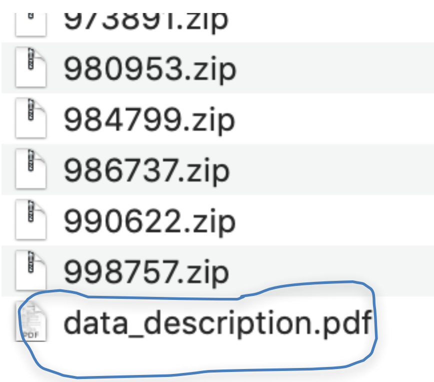
Abstract:
We introduce hand movement, orientation, and grasp (HMOG), a set of behavioral features to continuously authenticate smartphone users. HMOG features unobtrusively capture subtle micro-movement and orientation dynamics resulting from how a user grasps, holds, and taps on the smartphone. We evaluated authentication and biometric key generation (BKG) performance of HMOG features on data collected from 100 subjects typing on a virtual keyboard. Data were collected under two conditions: 1) sitting and 2) walking. We achieved authentication equal error rates (EERs) as low as 7.16% (walking) and 10.05% (sitting) when we combined HMOG, tap, and keystroke features. We performed experiments to investigate why HMOG features perform well during walking. Our results suggest that this is due to the ability of HMOG features to capture distinctive body movements caused by walking, in addition to the hand-movement dynamics from taps. With BKG, we achieved the EERs of 15.1% using HMOG combined with taps. In comparison, BKG using tap, key hold, and swipe features had EERs between 25.7% and 34.2%. We also analyzed the energy consumption of HMOG feature extraction and computation. Our analysis shows that HMOG features extracted at a 16-Hz sensor sampling rate incurred a minor overhead of 7.9% without sacrificing authentication accuracy. Two points distinguish our work from current literature: 1) we present the results of a comprehensive evaluation of three types of features (HMOG, keystroke, and tap) and their combinations under the same experimental conditions and 2) we analyze the features from three perspectives (authentication, BKG, and energy consumption on smartphones).

- Please Download from the webpage
- You will get a zip file



What are those data sets? For example, what is gyroscope data?

- There is a read me at the end of the data set with all zip files of all user IDs.



Data Description

1. Activity.csv

Name	Description
ID	Composed as: SubjectID + Session_number + ContentID + Run-time determined Counter value
SubjectID	6 digits: ID of current subject
Session_number	1-24: session number for current subject
Start_time	Start time of current activity, in absolute timestamps
End_time	End time of current activity, in absolute timestamps
Relative_Start_time	Start time of current activity, relative to system boot
Relative_End_time	End time of current activity, relative to system boot

Gesture_scenario	1: Sit 2:Walk
TaskID	1, 7, 13, 19: Reading + Sitting 2, 8, 14, 20: Reading + Walking 3, 9, 15, 21: Writing + Sitting 4, 10, 16, 22: Writing + Walking 5, 11, 17, 23: Map + Sitting
ContentID	1: first sub-task 2: second sub-task 3: third sub-task

2. Accelerometer.csv

Name	Description
Systime	Absolute time-stamp
EventTime	Sensor event relative time-stamp
ActivityID	Belonged activity
X	Acceleration minus Gx on the x-axis
Y	Acceleration minus Gy on the y-axis
Z	Acceleration minus Gz on the z-axis
Phone_orientation	0: Portrait and no rotate 1: device rotated 90 degrees counter-clockwise 3: device rotated 90 degrees clockwise

3. Gyroscrope.csv

Name	Description
Systime	Absolute time-stamp
EventTime	Sensor event relative time-stamp
ActivityID	Belonged activity
X	Angular speed around the x-axis
Y	Angular speed around the y-axis
Z	Angular speed around the z-axis
Phone_orientation	0: Portrait and no rotate 1: device rotated 90 degrees counter-clockwise 3: device rotated 90 degrees clockwise

4. Magnetometer.csv

Name	Description
Systime	Absolute time-stamp
EventTime	Sensor event relative time-stamp
ActivityID	Belonged activity
X	Ambient magnetic field in the X axis in micro-Tesla (uT)
Y	Ambient magnetic field in the Y axis in micro-Tesla (uT)
Z	Ambient magnetic field in the Z axis in micro-Tesla (uT)
Phone_orientation	0: Portrait and no rotate 1: device rotated 90 degrees counter-clockwise 3: device rotated 90 degrees clockwise

5. TouchEvent.csv

Name	Description
Systime	Absolute time-stamp
EventTime	Sensor event relative time-stamp
ActivityID	Belonged activity
Pointer_count	1: Single touch 2: Multi-touch
PointerID	0: Single touch; or first pointer in multi-touch 1: Second pointer in multi-touch
ActionID	0 or 5: DOWN 1 or 6: UP 2: MOVE
X	Touch location in X coordination
Y	Touch location in Y coordination
Pressure	Touch pressure
Contact_size	Touch contact size
Phone_orientation	0: Portrait and no rotate 1: device rotated 90 degrees counter-clockwise 3: device rotated 90 degrees clockwise

Homework

- Please read the following paper:

Journals & Magazines > IEEE Transactions on Informat... > Volume: 11 Issue: 5

HMOG: New Behavioral Biometric Features for Continuous Authentication of Smartphone Users

- https://ieeexplore.ieee.org/document/7349202?arnumber=7349202

Overview of Lecture 1

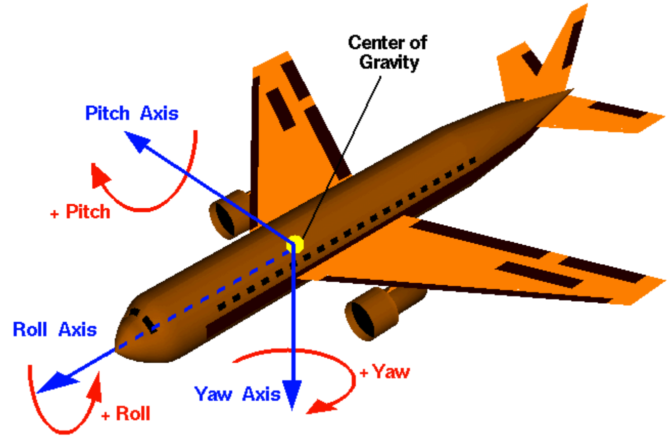
- Why we need nonlinear data analysis?
  - First starting with curves and their analysis
- Similarity measurements for nonlinear data
  - First a few examples: Arc-length, Geodesic length
- Introduction to cell phone data
- **Introduction to rigid motion**

Introduction to Rigid Motion

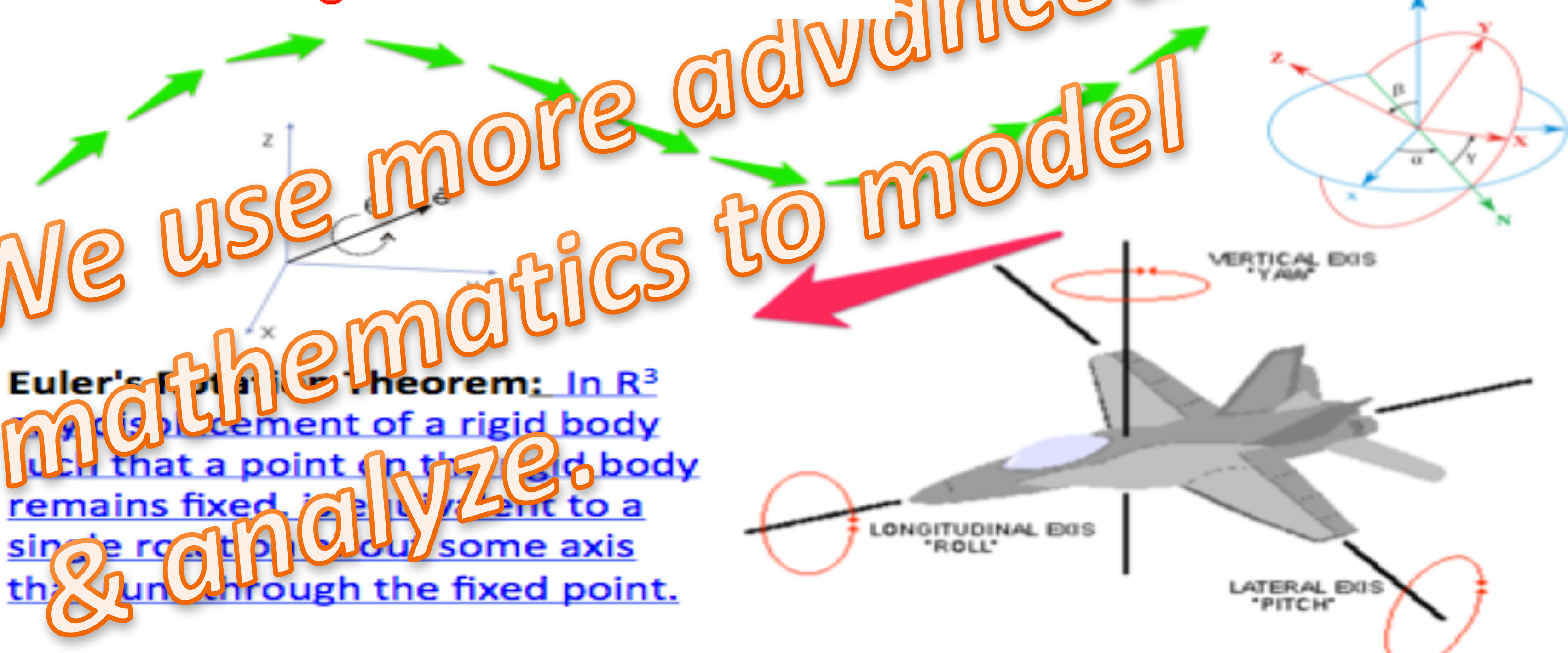
Details of Hard Math Behind UAV Data
(similar for cell phone data
or auto vehicle data)

- **Moving frames**
- **The set of orthonormal matrices**
- **The set of rotations in R^3**
- **Lie group SO(3)**
- Work out details with students on the board.

Viewing an UAV as a point is not enough since it has more complicated dynamics such as pitch, roll, yaw and their angular velocities

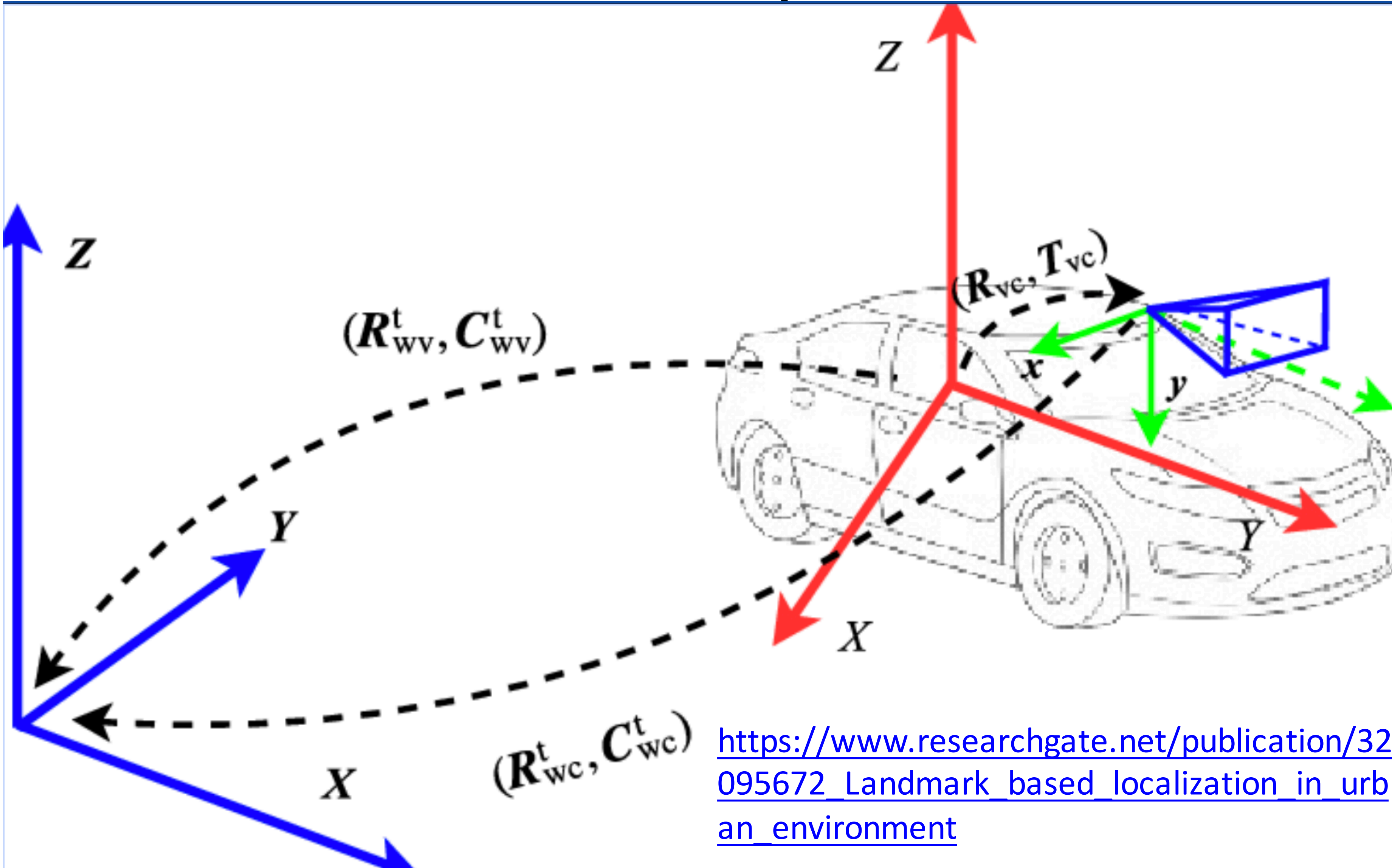


Details later!



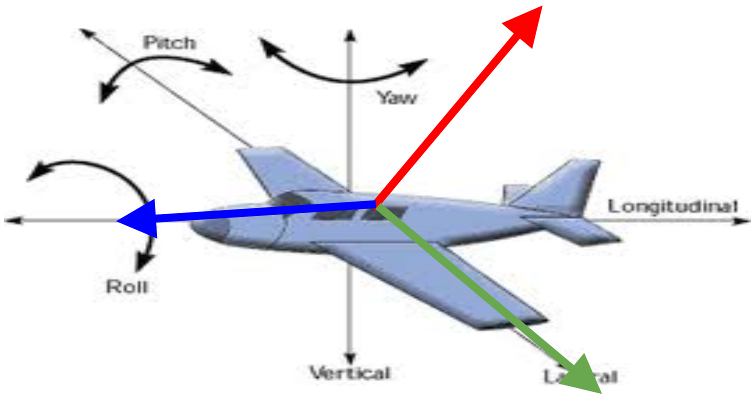
We use more advanced mathematics to model & analyze.

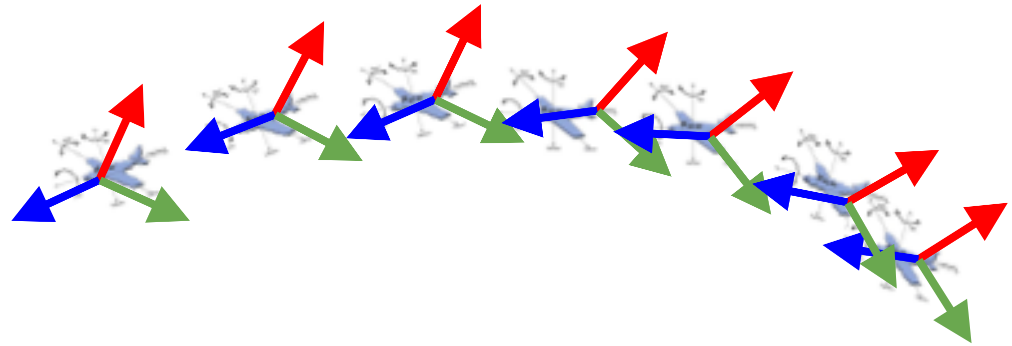
How to set a good coordinate system to model a problem?



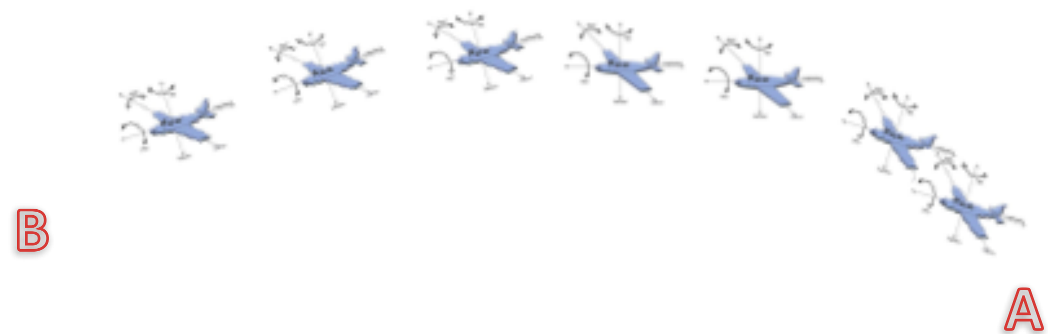
https://www.researchgate.net/publication/320095672_Landmark_based_localization_in_urban_environment

For Example: we want a computer to mathematically understand a pilot's manual flight control skill. Then we can compare between good controls and poor controls.





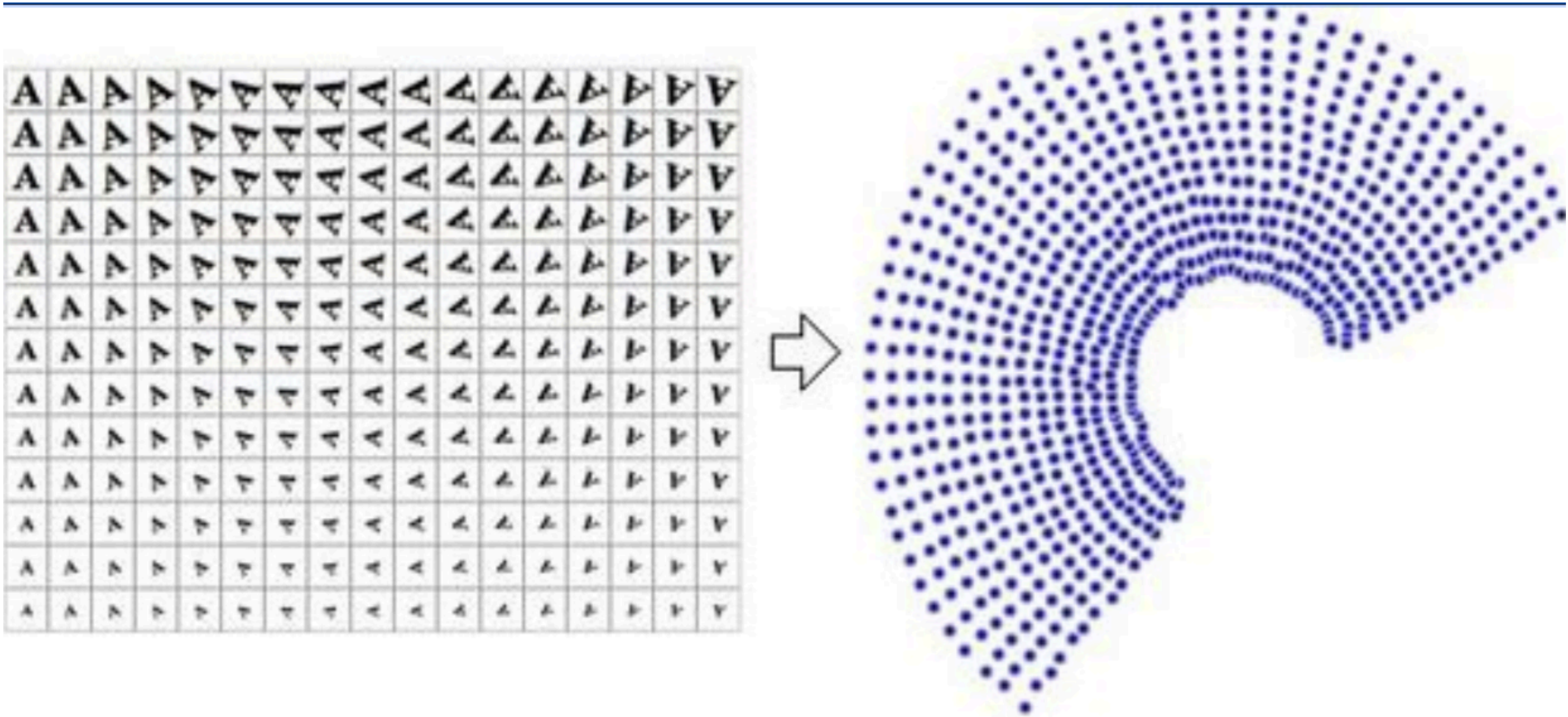
Key: This kind of mathematics captures dynamical behaviors of any UAVs



References

- https://datascience.stackexchange.com/questions/5694/dimensionality-and-manifold
- https://github.com/VivekPa/IntroNeuralNetworks
- https://developers.google.com/web/fundamentals/native-hardware/device-orientation/
- https://ieeexplore.ieee.org/document/7349202?arnumber=7349202

Back up slides

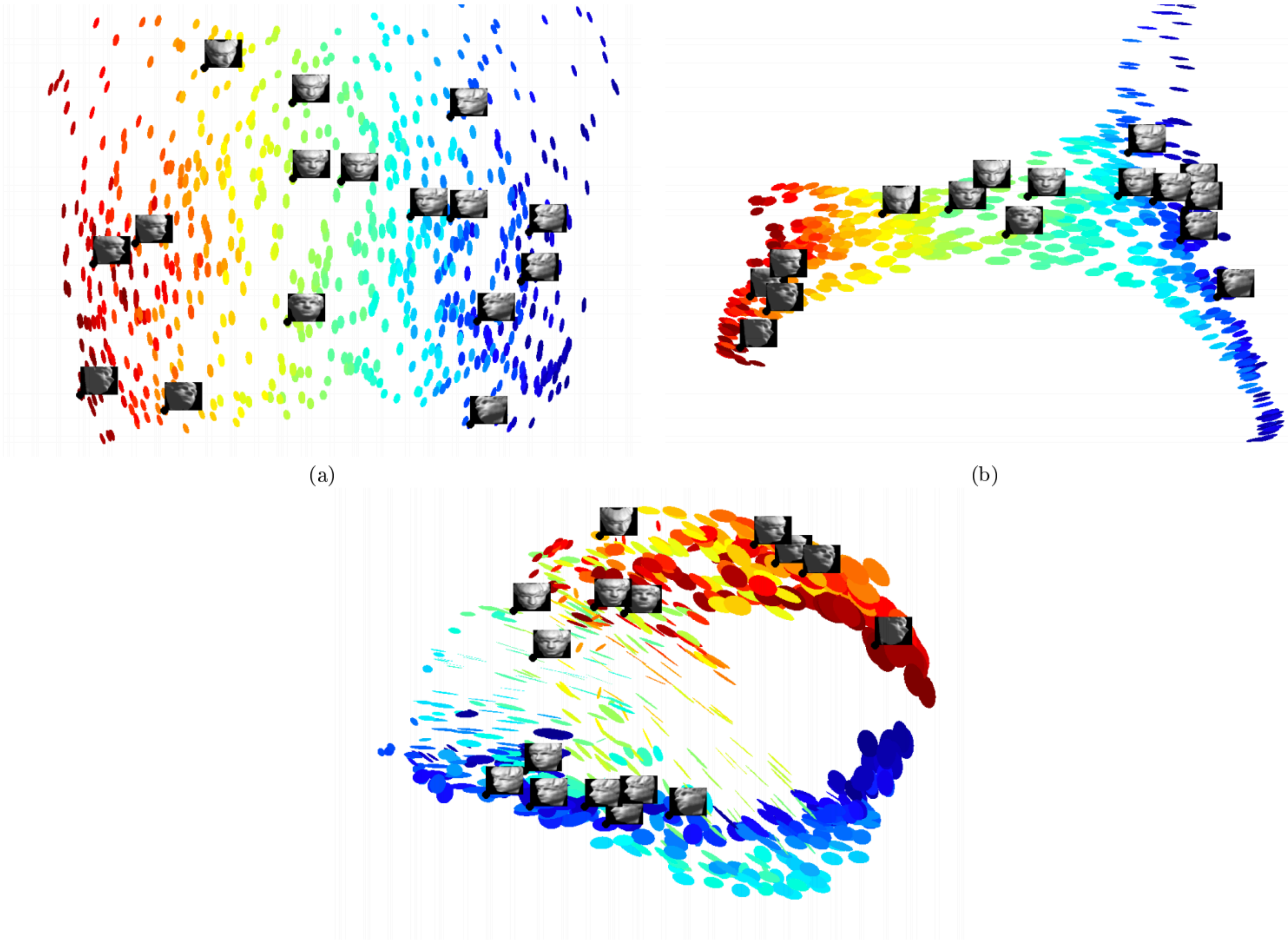


- Metric Learning and Manifolds: Preserving the Intrinsic Geometry
- https://www.stat.washington.edu/mmp/geometry/reading-group17/html/RMetric.pdf

Abstract

A variety of algorithms exist for performing non-linear dimension reduction, but these algorithms do not preserve the original geometry of the data except in special cases. In general, in the low-dimensional representations obtained, distances are distorted, as well as angles, areas, etc. This paper proposes a generic method to estimate the distortion incurred at each point of an embedding, and subsequently to "correct" distances and other intrinsic geometric quantities back to their original values (up to sampling noise).

Our approach is based on augmenting the output of an embedding algorithm with geometric information embodied in the Riemannian metric of the manifold. The Riemannian metric allows one to compute geometric quantities (such as angle, length, or volume) for any coordinate system or embedding of the manifold. In this work, we provide an algorithm for estimating the Riemannian metric from data, consider its consistency, and demonstrate the uses of our approach in a variety of examples.



(a) (b) (c)

Figure 5: Two-dimensional visualization of the faces manifold, along with embedding. The color corresponds to the left-right motion of the faces. The embeddings shown are: (a) Isomap, (b) LTSA, and Diffusion Maps ($\lambda = 1$) (c) . Note the very elongated ellipses at the top and bottom of the LTSA embedding, indicating the distortions that occurred there.

Find a paper to read which does the analysis using HMOG data

- Read
- Give a 1-2 page summary

HMOG: New Behavioral Biometric Features for Continuous Authentication of Smartphone Users

Zdeňka Sitová, Jaroslav Šeděnka, Qing Yang, Ge Peng, Gang Zhou, *Senior Member, IEEE*, Paolo Gasti, *Member, IEEE*, and Kiran S. Balagani, *Member, IEEE*

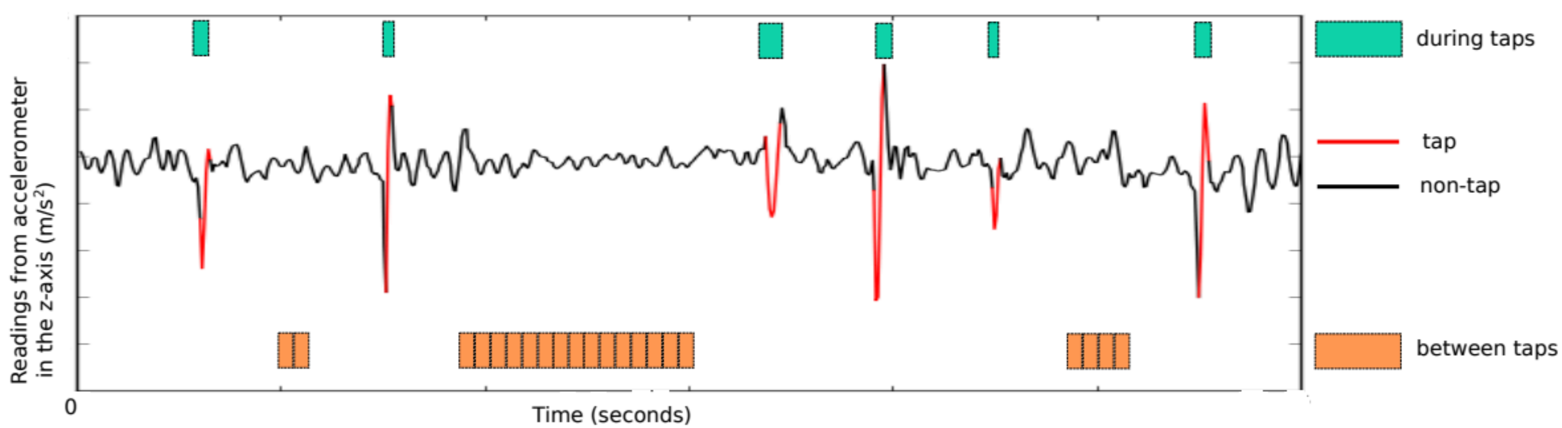


Fig. 8. HMOG features extracted *during* and *between* taps. The figure shows a sample of readings from the z-axis of accelerometer in sitting condition.