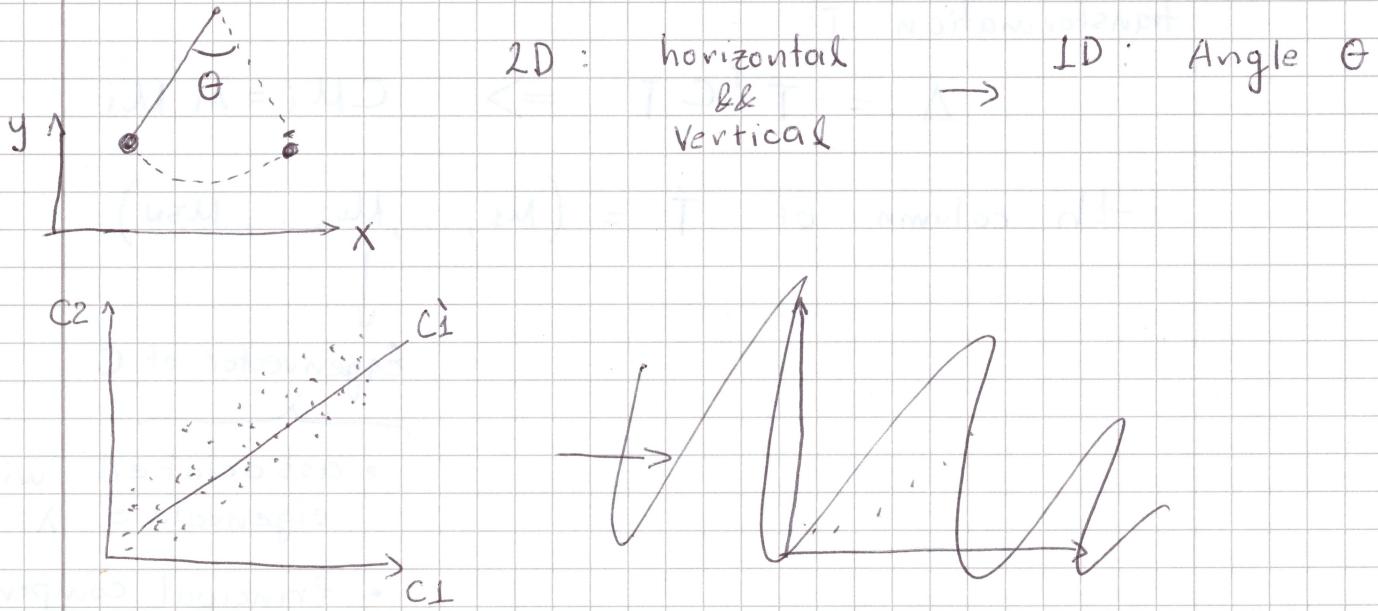


L3 Principal Component Analysis (PCA)

Biomolecular systems → • highly complex
 • How do we detect relevant degrees of freedom?

Example : Pendulum



PCA : multidimensional fit.

In proteins : Large fraction of positional fluctuations occur along few collective degrees of freedom.

Separate $\begin{cases} \text{- Large amplitude collective motions} \\ \text{- Rest} \end{cases}$ $\begin{cases} \text{- Few DF} \\ \text{- Functional DF} \\ \text{- Essential space (dynamics)} \\ \text{- Cover the most of PF.} \end{cases}$

N atoms $\vec{x} = (x_1, \dots, x_{3N}) \rightarrow \text{Coordinates}$
 $\vec{x}(t) \longrightarrow \text{Trajectory.}$

Covariance matrix of positions :

$$C_{ij} = \langle (x_i - \langle x_i \rangle) \cdot (x_j - \langle x_j \rangle)^T \rangle; \quad \langle \rangle = \text{ensemble average}$$

$C \Rightarrow$ Generalized standard deviation.

- Deviation of all positions with respect to their average

C is symmetric: $C_{ij} = C_{ji}$

- Let us diagonalize C applying an orthogonal transformation T :

$$\Delta = T^T C T \Rightarrow C M_i = \lambda_i M_i$$

i-th column of $T = (M_1, \dots, M_i, \dots, M_{3N})$

\downarrow
Eigenvector of C

- associated with eigenvalue λ_i
- Principal component.

- λ_i : mean square positional fluctuation along M_i .

- 3N eigenvectors:

RB rotational or translational (6 DF) \Rightarrow
3N - 6 eigenvectors with $\lambda_i \neq 0$

- Projection of the trajectory: $x(t)$

$$P_i(t) = M_i \cdot (\underline{x}(t) - \langle \underline{x} \rangle)$$

↑
scalar ↓
 dot product

- Projection back into cartesian space

$$x_i^{\dagger}(t) = \pi_i(t) \mu_i + \langle x \rangle$$

↑
Vector

Other methods:

- FMA , Hub , PLoS Comp Biol 5: e1000480 (2009)
- PLS , Krivobokova , Biophys. J. 103: 786 (2013)
- Markov Models : Noe , Curr Opin Struct Biol 18 : 154 (2008)
- Normal mode Analysis . (Diagonalize Hessian matrix , instead)
quasi - harmonic approx.