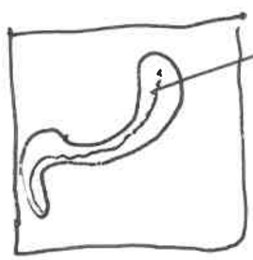
Boulder Summer School

Lecture #2: Representing Behavior II

(Beyond the worm)
(An abridged collection of Gordon's mistakes)
(Mistakes were made)

Last Time:

Worm



1) Reduced to a centerline
2) Found low-d representation (1-d!) of posture
3) Fit SDE to map posture to dynamics
4) Predicted stuff

Great! Now look at a fly / mouse / rat / human / crazy thing

Q: Is the worm the only organism where we have a chance?

To succeed, need to answer

1) How to represent posture?
2) How to map posture $\longrightarrow$ dynamics?
3) What can we do once we've solved this?

This is too vague! What do we want to know?

Me: 1) How does an animal spend its day?
(repertoire + dynamics on multiple scales)
2) What are the internal processes driving these choices?

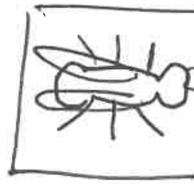
Assumptions: 1) Infinite amount of perfect data
2) Any algorithm must scale $\sim N$ or $\sim N \log N$

$\Longrightarrow$ Largely, assume all technical hurdles are gone...

Berman et al, RSI, 2014

Start with a movie (say, of a fly)

How should we represent it?



1) Should be translationally & rotationally invariant
→ Alignment

2) Any quantity should be readily extractable from a given image

3) should be continuous in time

First idea: Track stuff (have good logic for these degrees of freedom)

→ This used to be difficult,

Now CNNs make this surprisingly easy

→ Not always good though (gelatenous blob) even w/ technology exists hydra

Second idea: Use raw data → compress

1) extract ~~image~~ animal from background

2) align image

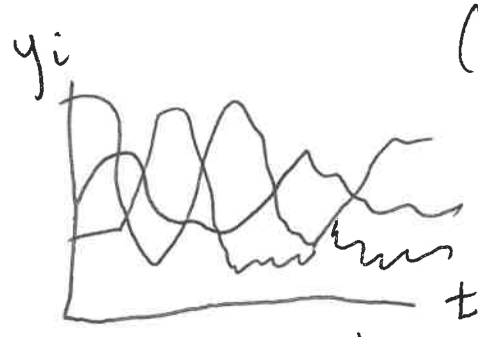
3) Reduce dimensionality (PCA / NMF)

4) Project onto bases to generate time series

"Eigenflies"

Outcome:

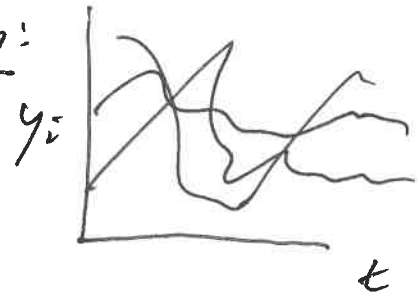
images → (time series)



Should be consistent, continuous, and accurate
(ideally parsimonious)

So I have a time series, now what?

<u>Problem:</u>



Postural Representation $\longrightarrow$ Dynamical Representation

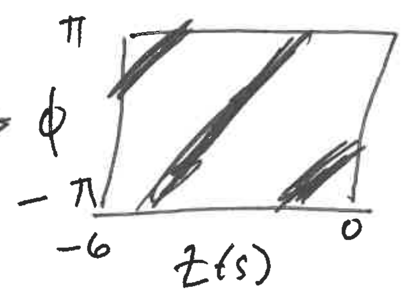
<u>Remember:</u> <u>Worm:</u> $\frac{d\phi}{dt} = \omega$, $\frac{d\omega}{dt} = F(\phi,\omega) + \sigma(\phi,\omega)\eta(t)$

Way too many variables if $d > 1$!

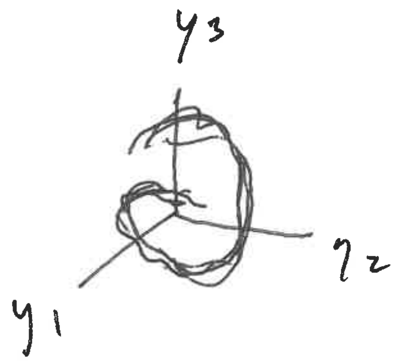
(Also, low Re makes $(\phi,\omega) \to \dot{\omega}$ to be more deterministic)

<u>Idea:</u> If we can't find the dynamical system, find its outputs $\longrightarrow$ <u>stereotyped behaviors</u>

Worm $\longrightarrow$



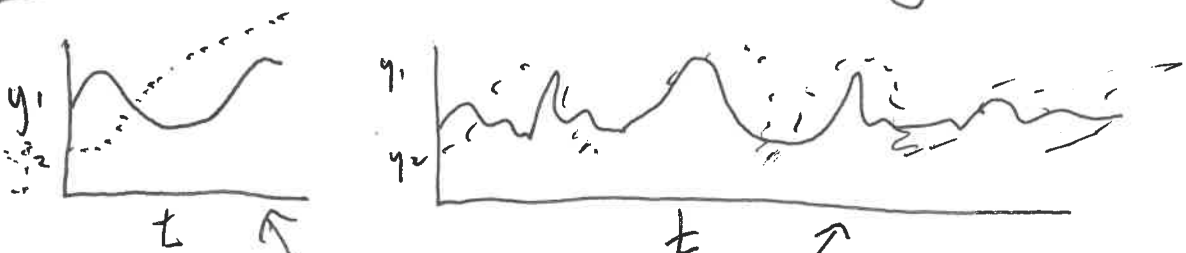
<u>or:</u>



$\longrightarrow$ Can we find these from a data set?

$\longrightarrow$ More challenging than in a 1-d case!

Idea #1: Explicitly find repeating motifs



convolve this with this

$$X_i^{(t)} = \int_{-\infty}^{\infty} y_{1i}(t'-t)\,\hat{y}_{i2}(t')\,dt' \longrightarrow \text{find } X_{TOT}^{(t)} = \sum_{i=1}^{d} X_i(t)$$

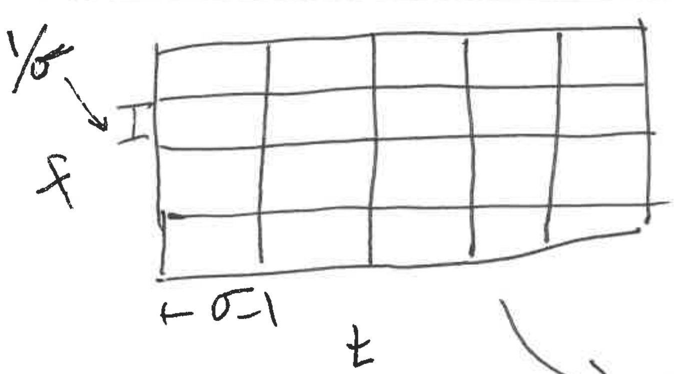
Problems: 1) Have to choose ~~length~~ length
2) Will miss if slightly out of phase
3) Will miss if shape same, but frequency slightly different

(Can do Dynamic Time Warping, but computationally difficult over many variables)

Idea #2: Spectral Analysis

$$S_\sigma^{[X_i]}(t_0 f) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} X_i(t') \exp\left[-\frac{(t'-t_0)^2}{2\sigma^2}\right] e^{-2\pi i f t'}\,dt'$$

(Short time Fourier Transform)



$\underline{\Delta t \Delta f \sim 1}$

$\downarrow$

$\approx \sigma$

Look at $|S_\sigma[X_i](t,f)|$

$\rightarrow$ No phasing

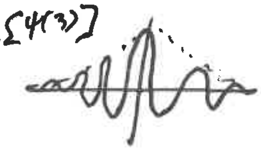
Only a single time scale!

④

To include multiple time scales:

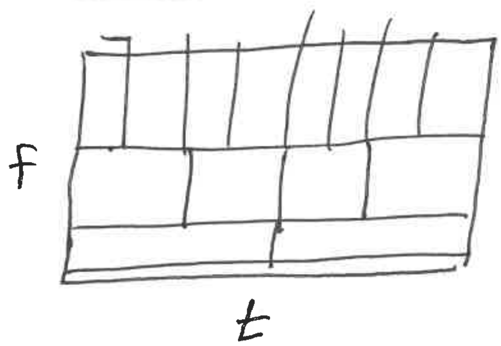
Continuous Wavelet Transform $\rightarrow$ $W[X_i(t)](s,t) = \frac{1}{\sqrt{s}}\int_{-\infty}^{\infty} X_i(t')\,\psi^*\!\left(\frac{t'-t}{s}\right)dt'$

Morlet Wavelet: $\psi(\eta) = \pi^{-1/4} e^{-i\omega_0\eta}\, e^{-\eta^2/2}$



$$\rightarrow W[X_i(t)]_{(s,t)} = \frac{\pi^{-1/4}}{\sqrt{s}}\int_{-\infty}^{\infty} X_i(t')\exp\left[\frac{-(t'-t)^2}{2s^2}\right] e^{-i\omega_0\frac{(t'-t)}{s}}\,dt'$$

Factor of $1/s$ $\rightarrow$ Trade-off between frequency & time res.



$\Delta t\,\Delta f \sim 1$

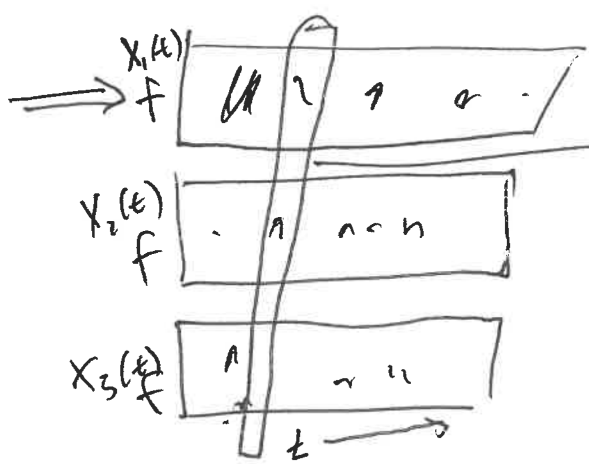
$\rightarrow$ better time resolution at high f, better freq. resolution at low f

$$s(f) = \frac{\omega_0 + \sqrt{2+\omega_0^2}}{4\pi f}$$

(Maximize responce for $X(t) = e^{-i\omega t}$)

Pick $f_s \Rightarrow f_i = f_{max}\cdot 2^{-\frac{(i-1)}{N_f-1}\log_2\frac{f_{max}}{f_{min}}}$

Use $|W(t,f)|$ for all $X_i(t)$ $\rightarrow$ N_f frequency channels, d time series

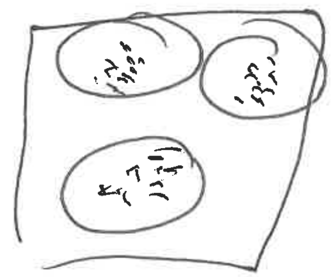


$\rightarrow$ $N_f \times d$-dimensional vector describing which body parts are moving at which speeds. (flies $\in \mathbb{R}_+^{1250}$)

$\rightarrow$ Say I have a lot of these, how to make sense? $\rightarrow$ Dimensionality Reduction

Goal: To use these high-d vectors to find stereotyped behaviors

Idea #1: Clustering (Zoology of methods)



→ Each cluster is a stereotyped behavior

Problems:

1) Have to pick a number of clusters (or a parameter that does it)
2) All points have to belong to a "behavior" (no non-stereotyped behaviors)
3) We don't know if discrete behaviors exist!

Idea #2: Make a "behavioral map"

$f: \mathbb{R}^{1250} \to \mathbb{R}^{2,3,4,\ldots}$ (something small)

f should ⇝

1) preserve local clustering structure
2) keep far-away points from being near each other

Note: PCA, NMF, Isomap, LLE all are terrible at this

Try t-Distributed Stochastic Neighbor Embedding (t-SNE) t-SNE Let $p_{ij} \sim \exp\left(\frac{-D_{ij}^2}{2\sigma_i^2}\right)$ $\left(H_i = -\sum_j p_{ij} \log p_{ij} = \text{const}\right)$

$D_{ij} \to$ distance in high-d space

$$q_{ij} \sim \frac{1}{1+\Delta_{ij}^2}$$ ($\Delta_{ij} \to$ distance in 2/3-d space)

$$\vec{y}^* = \underset{\vec{y}}{\arg\min}\ D_{KL}\left(\frac{p_{ij}}{N} \,\Big\|\, \frac{q_{ij}(\vec{y})}{N}\right) = \underset{\vec{y}}{\arg\min} \sum_{ij} \frac{p_{ij}}{N} \log \frac{p_{ij}}{q_{ij}(\vec{y})}$$

Note: cost function only matters if $p_{ij} \neq 0$
$\Rightarrow$ pay attention to local points!

---

for us: $D_{ij} = \frac{1}{2}\left[D_{KL}(\vec{s}_i \| \vec{s}_j) + D_{KL}(\vec{s}_j \| \vec{s}_i)\right]$

Solve through iterative gradient descent

7