

# Specialization Project

Ulrik Bernhardt Danielsen

September 29, 2022

---

## Contents

---

<b>1</b>	<b>Introduction</b>	<b>5</b>
1.1	Motivation . . . . .	5
1.2	Previous work . . . . .	5
<b>2</b>	<b>Theory</b>	<b>7</b>
2.1	Time series analysis . . . . .	7
2.1.1	Stationarity . . . . .	7
2.1.2	Detrending . . . . .	8
2.2	Fourier analysis . . . . .	9
2.2.1	Discrete-Time Fourier Transform . . . . .	10
2.3	Spectral density estimation . . . . .	11
2.3.1	Periodogram . . . . .	11
2.3.2	Sample spectrum . . . . .	12
2.3.3	Spectral Window . . . . .	12
2.3.4	Lag window . . . . .	13
2.3.5	Short-Time Fourier Transform . . . . .	14
2.3.6	Wavelet transformation . . . . .	14
2.4	Piecewise polynomials . . . . .	14
2.4.1	Splines . . . . .	14
2.4.2	Regression splines . . . . .	14
2.5	Dimensionality reduction . . . . .	15
2.5.1	Principal Component Analysis . . . . .	15
2.5.2	t-Stochastic Neighbor Embedding . . . . .	15
2.6	Watershed segmentation . . . . .	15

---

<b>3 Methodology</b>	<b>17</b>
3.1 Feature extraction . . . . .	17
3.2 Manifold embedding . . . . .	17
<b>Bibliography</b>	<b>19</b>

**Steps:**

1. z-scoring (standardization **although dividing by standard deviation is commented out**)
2. Smoothing by 3d splines
3. Detrending around the smoothed curve (**Why do we need to do this? Do we expect trends?**)
4. Morlet continuous wavelet transformation
  - a) Hypothesis testing against 1st order autoregressive process
  - b) Smoothing of power across scales
  - c) Scale-averaged wavelet power
  - d) Rescaling features
  - e) Concatenating trend data and power spectrum
5. Feature vectors were downsampled in time at 1 Hz and pooled across animals and conditions.  
**Where is this done in the code?** it
6. PCA, reducing to feature dimension explaining at least 95% of the variance
7. t-SNE
8. Watershed segmentation



---

## Chapter 1

# Introduction

---

### 1.1 Motivation

The human brain is an incredibly complex structures that researchers have been trying to understand for a long time. One way to gain information about how the brain operates is to study its neurons. Neurons are cells which can communicate with each other through synapses. This communication are electric signals and can be recorded. [Source?](#) At Kavli Institute for Systems Neuroscience at NTNU they are interested in relating these neural spike recordings to the behavior in rats. This in turn begs the question of how rats behave. Manually labelling video recordings of rats running around seems a tedious and unfruitful endeavor. Additionally it introduces bias in our prior assumptions of how the rats behave, and which activities they engage in. Thus, a methodology for automatically detecting distinct behaviours is needed.

### 1.2 Previous work

[Is this necessary?](#)



---

## Chapter 2

# Theory

---

### 2.1 Time series analysis

We define time series as a realization  $y_t = \{y_{t_1}, y_{t_2}, \dots, y_{t_n}\}$  of a stochastic process  $Y(w, t)$ , where  $w \in \Omega$ ,  $\Omega$  being the sample space, and  $t \in \mathbb{Z}$ ,  $\mathbb{Z}$  being the chosen index set Wei 2006. It is an ordered series of random variables which can be described completely by its joint probability function

$$F_{t_1, \dots, t_n}(x_1, \dots, x_n) = \Pr\{y_{t_1} \leq x_1, \dots, y_{t_n} \leq x_n\}.$$

The mean and variance function of a time series  $y$  are defined as

$$\mu_t = E(y_t) \tag{2.1}$$

and

$$\sigma_t^2 = E(y_t - \mu_t)^2.$$

Given two random variables in the series  $y_{t_1}$  and  $y_{t_2}$ , we define the covariance function and correlation function as

$$\gamma(t_1, t_2) = E[(y_{t_1} - \mu_{t_1})(y_{t_2} - \mu_{t_2})] \tag{2.2}$$

and

$$\rho(t_1, t_2) = \frac{\gamma(t_1, t_2)}{\sqrt{\sigma_{t_1}^2} \sqrt{\sigma_{t_2}^2}}. \tag{2.3}$$

#### 2.1.1 Stationarity

A time series  $y_t$  is  $n$ th-order stationary if for any shift  $h$  and indexes  $t_1, t_2, \dots, t_n$  if

$$F_{y_{t_1}, \dots, y_{t_n}}(x_1, \dots, x_n) = F_{y_{t_1+h}, \dots, y_{t_n+h}}(x_1, \dots, x_n). \tag{2.4}$$

If (2.4) holds for all  $n$ , the time series is called *strictly* stationary. We also define a  $n$ th-order *weakly* stationary time series  $y_t$  if the first  $n$  joint moments are finite and time invariant. Specifically we define the second-order weakly stationary, i.e. with constant and time invariant mean function (2.1), and where the covariance function (2.2) is solely a function of the time difference,

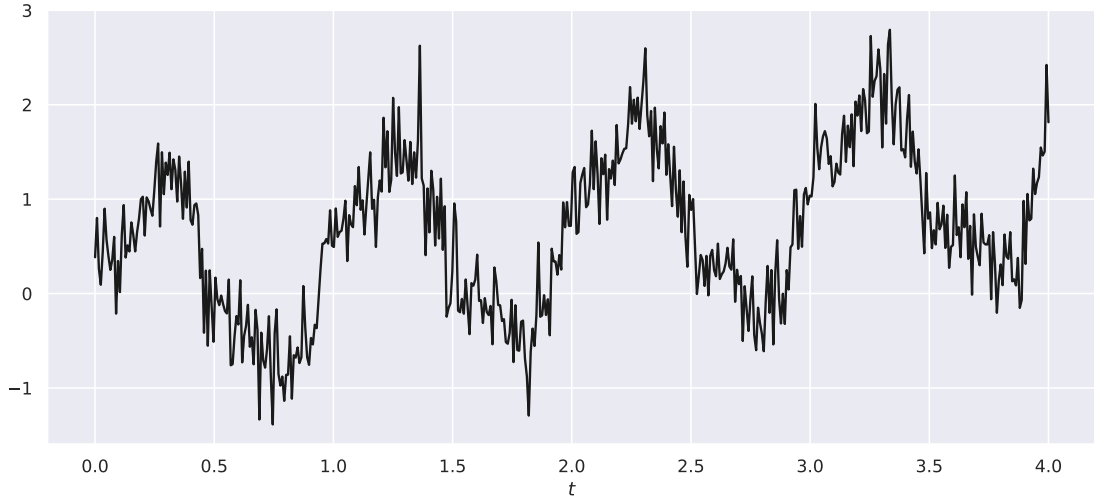


Figure 2.1: Example of a non-stationary time series with  $t_0 = 0, t_n = 4$ . The discrete data points are connected to better visualize the movement through time.

as *covariance* stationary. When the covariance function between  $t_1, t_2$  can be written as a function of the time difference  $h = |t_1 - t_2|$ , i.e.  $\gamma(t_1, t_2) = \gamma(h) = \gamma_h$ , we call it an *autocovariance* function. The same is true for the correlation function (2.3), which when is a function of the time difference is called an *autocorrelation* function (ACF). Figure 2.1 shows an example of a time series. As the mean seem to increase with  $t$  it is non-stationary.

We also define the common estimates for the mean and covariance functions. They are called the sample mean and sample covariance, and are written as

$$\bar{y}_t = \frac{1}{n} \sum_{i=1}^n y_{t_i}, \quad (2.5)$$

$$\hat{\gamma}_h = \frac{1}{n} \sum_{i=1}^{n-h} (y_{t_i} - \bar{y}_t)(y_{t_i+h} - \bar{y}_t), \quad (2.6)$$

respectively.

### 2.1.2 Detrending

Many methods for analysing and processing time series requires stationarity Shumway and Stoffer 2017. If the series is non-stationary, we can split the it into one stationary and one non-stationary part called the *trend*. Mathematically we write it as

$$y_t = \mu_t + x_t,$$

where  $x_t$  denotes the stationary part and  $\mu_t$  the trend. The process of finding  $\mu_t$  and then computing  $x_t = y_t - \mu_t$  is called *detrending*. Detecting the trend can be done in many ways, for instance using regression techniques or smoothing. The simplest way is to assume a linear trend,  $\mu_t = \beta_0 + \beta_1 t$  and estimate the parameters using least squares. In figure 2.2 the linear regression fit is shown, showing an upwards in the time series.



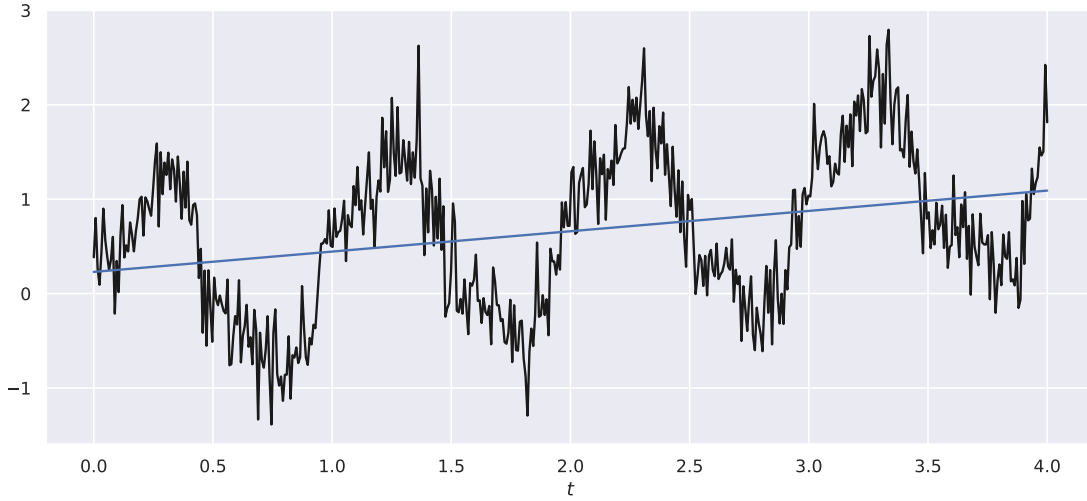


Figure 2.2: Time series from figure 2.1 with an estimated linear trend shown in blue.

It is clear that there exists an upward trend in the data.

## 2.2 Fourier analysis

Let  $Z_1, Z_2, \dots, Z_n$  be a sequence of numbers. For simplicity in notation we assume  $n$  to be an odd number. It can be shown that the sequence can be represented as a linear combination of complex exponentials

$$Z_t = \sum_{k=-\frac{n-1}{2}}^{\frac{n-1}{2}} c_k e^{\frac{i2\pi kt}{n}}. \quad (2.7)$$

This comes from the fact that the set

$$\left\{ e^{\frac{i2\pi kt}{n}} \mid k \in \left[ -\frac{n-1}{2}, \frac{n-1}{2} \right] \right\}$$

consists of  $n$  orthogonal functions Wei 2006. I.e., that

$$\sum_{i=1}^n e^{\frac{i2\pi kt}{n}} e^{-\frac{i2\pi jt}{n}} = \begin{cases} n, & k = j \\ 0, & k \neq j \end{cases}.$$

The coefficients  $c_k$  are given by

$$c_k = \frac{1}{n} \sum_{t=1}^n Z_t e^{-\frac{i2\pi kt}{n}}.$$

It is clear that (2.7) is periodic with period  $n$ , meaning  $Z_{t+jn} = Z_t, j = 0, \pm 1, \pm 2, \dots$ . Thus the Fourier series is able to capture periodic sequences. The smallest positive integer  $n$  for which  $Z_{t+n} = Z_t$  is called the fundamental period, with corresponding fundamental frequency  $2\pi/n$ . For the components  $k = \pm j, j = 1, 2, \dots, (n-1)/2$  the frequencies are multiples of the fundamental frequency,  $w_k = k(2\pi/n)$ . The set of frequencies making up the series is called the *spectrum*. As a consequence the coefficients  $c_k$  can be viewed as weighting the importance of the

contributions for the different frequencies making up the full sequence. This is formalized by the definitions of energy and power,

$$\text{energy} = \sum_{t=2}^n Z_t^2 = n \sum_{k=-\frac{n-1}{2}}^{\frac{n-1}{2}} |c_k|^2, \quad (2.8)$$

$$\text{power} = \frac{\text{energy}}{n} = \sum_{k=-\frac{n-1}{2}}^{\frac{n-1}{2}} |c_k|^2. \quad (2.9)$$

Let  $p_k$  be the contribution to the power from frequency  $k = 0, 1, \dots, (n-1)/2$ . As  $w_k$  and  $w_{-k}$  corresponds to the same frequency the contribution is given as  $p_0 = c_0^2, p_k = 2|c_k|^2, k = 1, \dots, (n-1)/2$ . The values  $p_k$  are called the power spectrum of the series.

### 2.2.1 Discrete-Time Fourier Transform

We have seen that all sequences of length  $n$  can be viewed and parameterized as Fourier series with period  $n$ . Moving to non-periodic sequences essentially amounts to taking the limit of the series as  $n$  approaches infinity. Formally we now let  $Z_t$  be a finite discrete function of  $t$ , where  $Z_t = 0$  when  $|t| > M$  for some integer  $M$ . Choosing  $n = 2M + 1$  the function

$$Y_{t+jn} = Z_t, \quad t \in \left[-\frac{n-1}{2}, \frac{n-1}{2}\right], \quad j \in \mathbb{Z}$$

is periodic with period  $n$ . It's Fourier series is

$$Y_t = \sum_{k=-\frac{n-1}{2}}^{\frac{n-1}{2}} c_k e^{\frac{i2\pi kt}{n}}.$$

As  $Y_t = Z_t$  when  $t \in [-(n-1)/2, (n-1)/2]$ , and  $Z_t = 0$  when  $|t| > (n-1)/2$ , the coefficients  $c_k$  can be written as the infinite sum

$$\begin{aligned} c_k &= \frac{1}{n} \sum_{t=-\infty}^{\infty} Z_t e^{\frac{-i2\pi kt}{n}} \\ &= \frac{2\pi}{n} f\left(\frac{2\pi k}{n}\right), \end{aligned}$$

where

$$f(w) = \frac{1}{2\pi} \sum_{t=-\infty}^{\infty} Z_t e^{-iwt}.$$

If we now take the limit  $Z_t = \lim_{n \rightarrow \infty} Y_t$  the summation becomes an integral over the length  $2\pi$  Wei 2006. This gives the relation

$$Z_t = \int_{-\pi}^{\pi} f(w) e^{iwt} dw, \quad t \in \mathbb{Z} \quad (2.10)$$

$$f(w) = \frac{1}{2\pi} \sum_{t=-\infty}^{\infty} Z_t e^{-iwt}, \quad -\pi \leq w \leq \pi, \quad (2.11)$$

where  $f(w)$  in (2.11) is called the discrete-time Fourier transform of  $Z_t$ . As opposed to the periodic case (2.7) where the periodic sequence was made up by a finite number of frequencies, the non-periodic sequence is an integral over a continuum of frequencies  $w$ . We call  $|f(w)|$  the spectrum of the sequence, and the function  $g(w) = 2\pi|f(w)|^2$  the energy spectrum. The energy spectrum definition comes from Parseval's relation

$$\sum_{t=-\infty}^{\infty} |Z_t|^2 = 2\pi \int_{-\pi}^{\pi} |f(w)|^2 dw. \quad (2.12)$$

It is worth noting that the relation in (2.12) only holds when the sequence  $Z_t$  is absolutely summable, i.e., that

$$\sum_{t=-\infty}^{\infty} |Z_t| < \infty. \quad (2.13)$$

## 2.3 Spectral density estimation

Let  $y_t$  be a stationary time series where the autocovariance function  $\gamma_h$  from (2.2) is absolutely summable. Then we can write  $\gamma_h$  as a Fourier transform pair

$$f(w) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_h e^{-iwh}, \quad (2.14)$$

$$\gamma_h = \int_{-\pi}^{\pi} f(w) e^{iwh} dw. \quad (2.15)$$

It can be shown Shumway and Stoffer 2017 that the spectrum  $f(w)$  in (2.14) is real-valued and non-negative. Furthermore, as  $\text{Var}(y_t) = \gamma_0$ , we get the interpretation

$$\text{Var}(y_t) = \int_{-\pi}^{\pi} f(w) dw,$$

i.e., that  $f(w)$  is the contribution to the variance for frequency  $w$ . We often want to locate these important frequencies, and thus an important task is to estimate this spectrum.

### 2.3.1 Periodogram

Again we consider a time series sample  $y_1, y_2, \dots, y_n$  where  $n$  is chosen to be odd for simplicity. It can be written as a real Fourier representation

$$y_t = a_0 + \sum_{k=1}^{\frac{n-1}{2}} (a_k \cos(w_k t) + b_k \sin(w_k t)),$$

where  $w_k = 2\pi k/n, k = 0, 1, \dots, (n-1)/2$  and the coefficients are given as

$$a_0 = \bar{y}_t, \quad a_k = \frac{2}{n} \sum_{t=1}^n y_t \cos(w_k t), \quad b_k = \frac{2}{n} \sum_{t=1}^n y_t \sin(w_k t).$$

We then define the periodogram as

$$I(w_k) = \begin{cases} na_0^2, & k = 0 \\ \frac{n}{2}(a_k^2 + b_k^2), & k = 1, \dots, (n-1)/2. \end{cases} \quad (2.16)$$

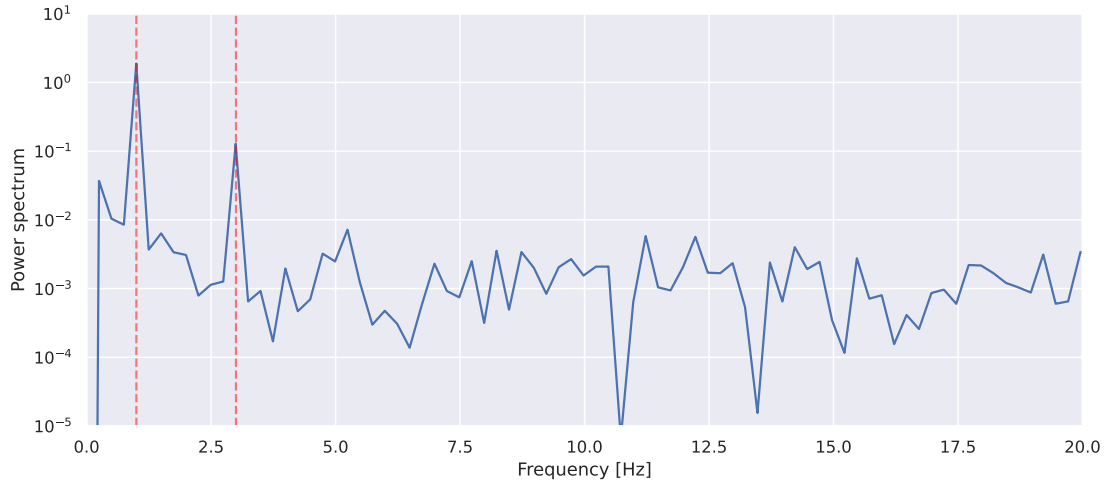


Figure 2.3: Periodogram for the time series plotted in figure 2.1. The time series is detrended assuming a linear trend. Observe that the two highest peaks in the periodogram are at the underlying frequencies 1Hz and 3Hz shown with dashed red lines.

The periodogram is of interest as it has a large value if the frequency  $w_k$  is of importance in the series. A scaled periodogram  $\frac{2}{n}I(w_k)$  estimates the sample variance of the sinusoid component at frequency  $w_k$  Shumway and Stoffer 2017. A periodogram for the time series example in figure 2.1 are shown in figure 2.3. Note that the time series is linearly detrended for the periodogram is computed. The two highest peaks match the underlying generating function which can be revealed to be  $\sin(2\pi t) + \frac{1}{4}\cos(6\pi t) + t/3$ .

### 2.3.2 Sample spectrum

One intuitive way of estimating the spectrum is to replace the autocovariance with the sample autocovariance from equation (2.6). I.e., we define the sample spectrum for a realization  $y_1, y_2, \dots, y_n$

$$\hat{f}(w) = \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} \hat{\gamma}_k e^{-i w k}. \quad (2.17)$$

At the Fourier frequencies  $w_k$  it is related to the periodogram through Wei 2006

$$\hat{f}(w_k) = \frac{I(w_k)}{4\pi}.$$

Although  $\hat{f}(w_k)$  is asymptotically unbiased, meaning  $\lim_{n \rightarrow \infty} E(\hat{f}(w)) = f(w)$ , it lacks consistency in the variance as  $n$  tends to infinity, i.e.,

$$\lim_{n \rightarrow \infty} \text{Var}(\hat{f}(w_k)) \neq 0.$$

### 2.3.3 Spectral Window

The fact that the variance does not decrease with the sample size produces quite jagged and noisy spectrum estimations Shumway and Stoffer 2017. To account for this we introduce the spectral

window which smooths the spectrum. Mathematically the smoothed spectrum is written as

$$\hat{f}_{\mathcal{W}}(w_k) = \sum_{j=-m}^m \mathcal{W}_n(w_j) \hat{f}(w_k - w_j), \quad (2.18)$$

where  $w_k = 2\pi k/n$ ,  $k = 0, 1, \dots, (n-1)/2$  and  $m$  is a function of  $n$ , typically  $m \ll n$ . The value of  $m$  decides how many points in the neighborhood of  $w_k$  should be included in the smoothing. Furthermore, the function  $\mathcal{W}_n(w_j)$  is chosen to have the following properties,

$$\begin{aligned} \sum_{j=-m}^m \mathcal{W}_n(w_j) &= 1, \\ \mathcal{W}_n(w_j) &= \mathcal{W}_n(-w_j), \\ \lim_{n \rightarrow \infty} \sum_{j=-m}^m \mathcal{W}_n^2(w_j) &= 0. \end{aligned} \quad (2.19)$$

We view the smoothed spectrum as a weighted average of the sample spectrum (2.17) in a window around the target frequency  $w_k$ . How the weights are distributed in the window is governed by  $\mathcal{W}_n(w_j)$ , giving it the name spectral window. Because of the property in (2.19) we have

$$\begin{aligned} \text{Var}(\hat{f}_{\mathcal{W}}(w_k)) &\approx \sum_{j=-m}^m \mathcal{W}_n^2(w_j) (f(w_k))^2 \\ &= (f(w_k))^2 \sum_{j=-m}^m \mathcal{W}_n^2(w_j) \xrightarrow{n \rightarrow \infty} 0, \end{aligned}$$

assuming  $f(w)$  is approximately constant in the window. We can thus reduce the variance by increasing the points included in the window. However, when doing so we introduce bias.

### 2.3.4 Lag window

It can be shown Wei 2006 that the spectral window forms a Fourier transform pair with a lag window  $W_n(k)$ , i.e.,

$$W_n(k) = \int_{-\pi}^{\pi} \mathcal{W}_n(w) e^{i w k} dw, \quad k = 0, \pm 1, \dots, \pm M.$$

The lag window is a weighting function applied to the sample autocovariance

$$\hat{f}_{\mathcal{W}} = \frac{1}{2\pi} \sum_{k=-M}^M W_n(k) \hat{\gamma}_k e^{-i w k},$$

with  $W_n = W(k/M)$ ,  $W$  being a bounded even continuous function

$$\begin{aligned} |W(t)| &\leq 1, \\ W(0) &=, \\ W(t) &= W(-t), & W(t) &=, \quad t > 1. \end{aligned}$$

Figure [insert figure](#) shows the popular *hanning* window given by

$$W_n^H = \begin{cases} \frac{1}{2}(1 + \cos(\frac{\pi k}{M})), & |k| \leq M \\ 0, & |k| > M. \end{cases} \quad (2.20)$$

### 2.3.5 Short-Time Fourier Transform

#### 2.3.6 Wavelet transformation

Morlet wavelet—a sine wave that is "windowed" (i.e., multiplied point by point) by a Gaussian

## 2.4 Piecewise polynomials

Suppose we have an interval  $[a, b]$  divided into  $M$  contiguous subintervals. The connecting edges of the subintervals  $a = \xi_0, \xi_1, \dots, \xi_{M-1}, \xi_M = b$  are called knots. On each of the intervals  $[\xi_i, \xi_{i+1}]$ ,  $i = 0, \dots, M-1$  we define a polynomial  $p_i(t)$ . The function

$$f(t) = \begin{cases} p_0(t), & t \in [\xi_0, \xi_1) \\ p_1(t), & t \in [\xi_1, \xi_2) \\ \vdots & \\ p_{M-1}(t), & t \in [\xi_{M-1}, \xi_M] \end{cases}$$

is called a *piecewise polynomial*.

### 2.4.1 Splines

In the definition of piecewise polynomials no restrictions are made on the polynomials, they are allowed to take any form. As in Quarteroni, Sacco, and Saleri 2010 we define a *spline*  $s_k(t)$  of order  $k$  on the interval  $[a, b]$  as a piecewise polynomial where

$$\begin{aligned} s_k(t) &\in \mathcal{P}^k, \quad t \in [\xi_i, \xi_{i+1}], \quad i = 0, 1, \dots, M-1 \\ s_k(t) &\in \mathcal{C}^{k-1}[a, b]. \end{aligned}$$

I.e., the spline consists of piecewise polynomials of order  $k$  and has continuous derivatives up to order  $k-1$ . A common choice is letting  $k = 3$ , providing continuous second derivatives over the interval. This is called *cubic splines*, and are often considered sufficiently smooth for function approximations. It is also common to add curvature constraints at the endpoints,  $s_3''(a) = s_3''(b)$ , arriving at the *natural cubic splines*.

### 2.4.2 Regression splines

Suppose now we have data points  $y_{t_1}, y_{t_2}, \dots, y_{t_n}$  on  $[a = t_1, b = t_n]$ . A spline of order  $k$  with chosen knots at  $a = t_1 = \xi_0, \xi_1, \dots, \xi_M = t_n = b$  can be parameterized as

$$s_k(t) = \sum_{i=1}^{M+k} \beta_i h_i(t), \quad (2.21)$$

where the functions  $h_i$  are the truncated-power basis set

$$\begin{aligned} h_j(t) &= t^{j-1}, \quad j = 1, \dots, k+1, \\ h_{k+1+l}(t) &= (t - \xi_l)_+^k, \quad l = 1, \dots, M-1, \end{aligned}$$

with  $(t)_+ = \max\{t, 0\}$  Hastie, Friedman, and Tibshirani 2017. The parameters  $\beta_i$  can be found using least squares. An example of cubic spline regression are shown in figure 2.4.

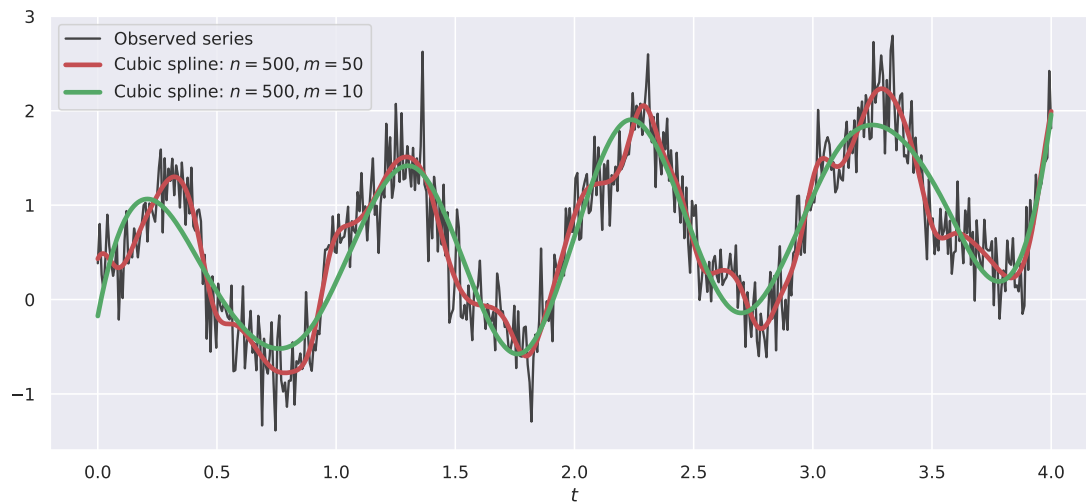


Figure 2.4: Two cubic splines fitted using least square regression on the time series from figure 2.1. Observe that the red spline with 50 knots (including endpoints) fits the data closer than the green spline with 10 knots.

## 2.5 Dimensionality reduction

### 2.5.1 Principal Component Analysis

### 2.5.2 t-Stochastic Neighbor Embedding

## 2.6 Watershed segmentation





---

## Chapter 3

# Methodology

---

The starting point of the analysis is a set  $\mathbf{Y} = \{Y_d : d = 1, 2, \dots, D\}$ , where each set  $Y_d$  is again a set of independent time series,  $Y_d = \{y_1^d(t), y_2^d(t), \dots, y_n^d(t) : t \in \{t_1, t_2, \dots, t_{m_d}\}\}$ . Each  $Y_d$  represents an animal for which  $n$  time series are collected tracking parts of its movements. Recording frequency is the same across animals. The goal of the analysis is to cluster these time points into distinct distinguishable actions.

Some summarization: "This is achieved by..."

### 3.1 Feature extraction

The first step is to analyse each time series separately, modifying them and thus creating  $D \times n$  features which we will extract the behavioral information from. Let  $y_t = \{y_1, y_2, \dots, y_m\} = y_i^d(t)$  for some  $d \in \{1, 2, \dots, D\}$  and  $i \in \{1, 2, \dots, n\}$  be one such series.

I don't see the point in centering the data around the mean before the detrending.

We detrend the data as  $y_t = s_3(t) + x_t$ ,  $s_3(t)$  being the non-linear trend and  $x_t$  the detrended time series. The cubic spline  $s_3(t)$  is found using least square regression as in equation (2.21) with equally spaced internal knots  $\xi_1, \xi_2, \dots, \xi_M$ . We choose the knots to have a fixed frequency, i.e., we choose  $\Delta\xi = \xi_{i+1} - \xi_i, i = 2, \dots, M$ . Figure 3.1 shows an example of a detrended time series using cubic splines with interior knots every 240th time point.

### 3.2 Manifold embedding

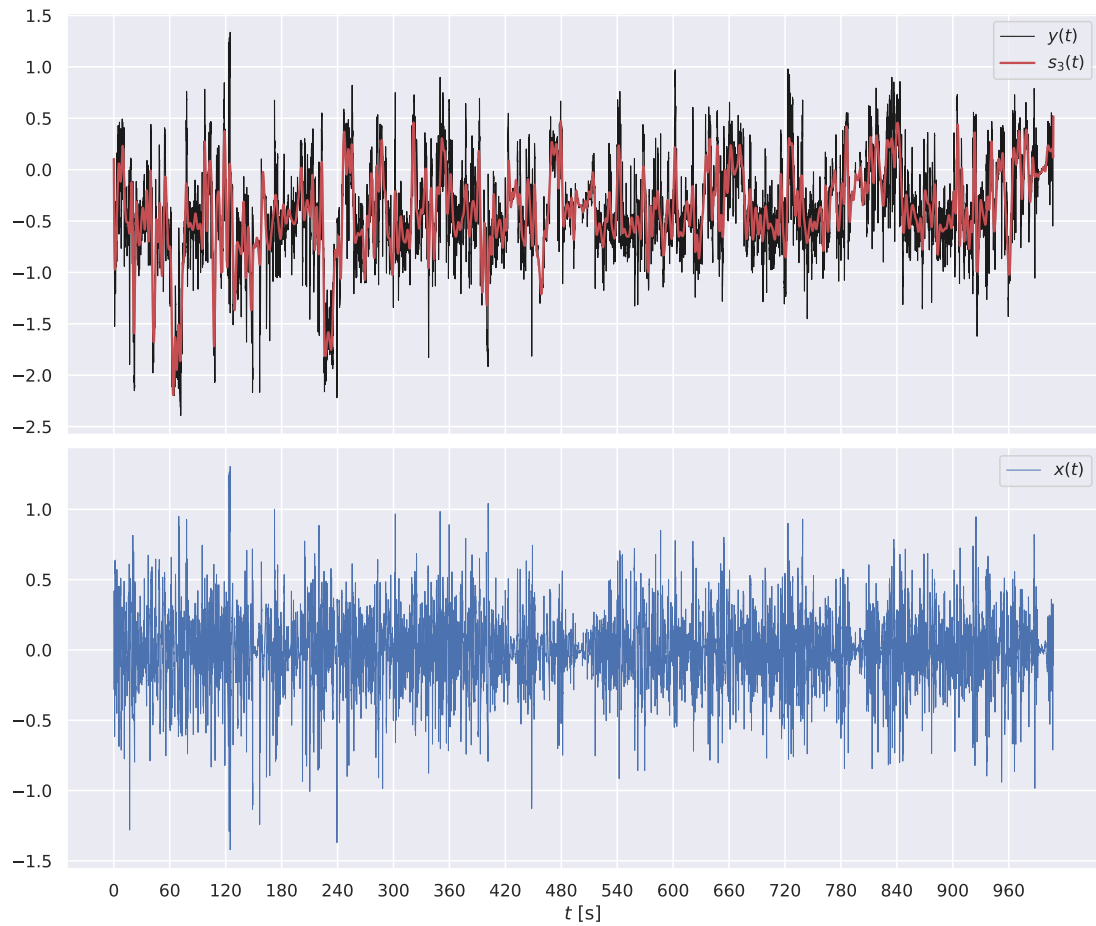


Figure 3.1: Detrended time series using cubic spline regression. The top figure shows the original time series in black with the fitted trend in red. The bottom figure shows the detrended time series.

---

# Bibliography

---

- Hastie, Trevor, Jerome Friedman, and Robert Tibshirani (2017). *The elements of Statistical Learning: Data Mining, Inference, and prediction*. 2nd ed. Springer.
- Quarteroni, Alfio, Riccardo Sacco, and Fausto Saleri (2010). *Numerical mathematics*. Springer.
- Shumway, Robert H. and David S. Stoffer (2017). *Time series analysis and its applications: With R examples*. 4th ed. Springer International Publishing.
- Wei, William W.S. (2006). *Time series analysis univariate and multivariate methods*. 2nd ed. Pearson.