Foundation Of Data Science Assignment-I

Basis Functions for Regression

Submitted To
Dr. Navneet Goyal
BITS Pilani

Submitted By
Rishav Mishra
2017B3A70557P
Sumit Bisht
2017B4A70581P
BITS Pilani

April 2021

Table of Contents

1. Introduction

In data science, we usually come across regression problems where given a set of input vectors; we have to predict the target value. The most common approach to solve such a problem is to use a linear model to fit data, popularly known as linear regression. However, the problem with this approach is that the data we deal with in the real world is very complex. Therefore, it's usually difficult to fit it in a linear model [1].
The solution to this problem is resolved by making our model non-linear. This is done by using a set of functions known as basis functions.
Let us assume we have some n-dimensional real-valued data $\{x_n, y_n\}_{n=1}^{N}$ with $x_n \in \chi$ and $y_n \in \mathbb{R}$. We define our transformation as $\Phi : \chi \to \mathbb{R}^j$ where j is the complexity of the model. The regression model can be written as:

$$y = w^T \Phi(x) \tag{1}$$

where

$$\Phi(x) = \begin{bmatrix} \phi_0(x_1) & \phi_1(x_1) & \dots & \phi_{j-1}(x_1) \\ \phi_0(x_2) & \phi_1(x_2) & \dots & \phi_{j-1}(x_2) \\ \vdots & \vdots & \dots & \vdots \\ \phi_0(x_N) & \phi_1(x_N) & \dots & \phi_{j-1}(x_N) \end{bmatrix}, w = (w_0, w_1, \dots, w_{j-1}) \tag{2}$$

These ϕ_j are referred as **basis function** or derived features and w_0 is referred as the bias parameter. Now the problem translates to finding the value of these weights w to minimize the loss. Hence using the model, the objective function for least-squares regression is :

$$w* = \min_{w} \left\{ \frac{1}{N} (\Phi w - y)^T (\Phi w - y) \right\} \tag{3}$$

Here as we are fixing Φ we could treat this as a linear problem but since the solution is function of x hence we are able to model a non-linear function. In fact, we can find an analytic solution for our problem if $N \geq j$:

$$w* = (\Phi^T \Phi)^{-1} \Phi^T y \tag{4}$$

1.1 Types of Basis Functions

We have developed a non-linear model using the concept of basis functions. Now we can define different basis functions whose choice depends on our understanding of the domain of the problem. In this subsection, we will discuss a few of the available basis functions.

1.1.1 Polynomial Basis Functions

The easiest choice for the basis functions is a polynomial basis. Here the basis function takes the form of powers of x. An example of polynomial basis function on a one-dimensional data is:

$$\phi_j(x) = x^j \tag{5}$$

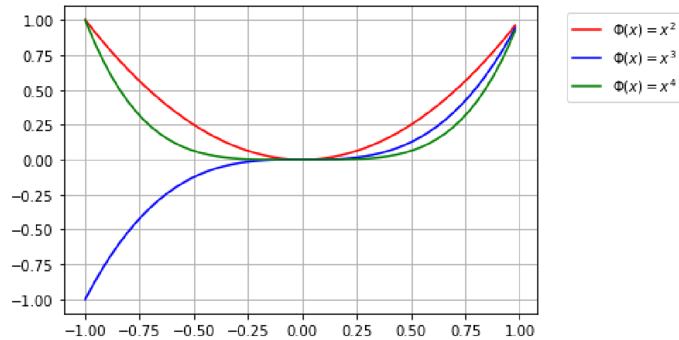


Figure 1: Example of polynomial basis functions.

Figure 1 shows the polynomial basis functions up to degree 4.

1.1.2 Sigmoidal Basis Functions

Another possible choice for the basis functions is the sigmoidal basis function. A sigmoidal function has a characteristic S-curve. An example of sigmoidal basis function on a one-dimensional data is:

$$\phi_j(x) = \sigma\left(\frac{x-\mu}{s}\right) \tag{6}$$

Here μ defines the centre of the curve and s defines the spread of the curve. The most common choice for $\sigma(a)$ is the logistic sigmoid function defined by:

$$\sigma(a) = \frac{1}{1+e^{-a}} \tag{7}$$

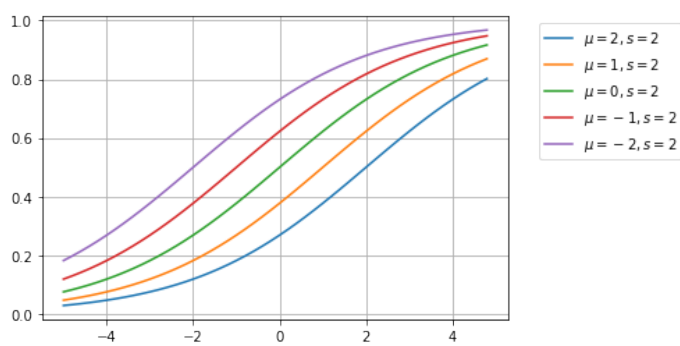


Figure 2: Example of sigmoidal basis functions.

Figure 2 shows five sigmoid functions with randomly chosen μ and s.

1.1.3 Gaussian Basis Functions

A common choice in practice for the basis functions is the gaussian basis function. A gaussian function is symmetric around a point and then typically decays to zero as one gets farther from the center. An example of gaussian basis function on a one-dimensional data is:

$$\phi_j(x) = e^{-\frac{(x-\mu_j)^2}{2s^2}} \tag{8}$$

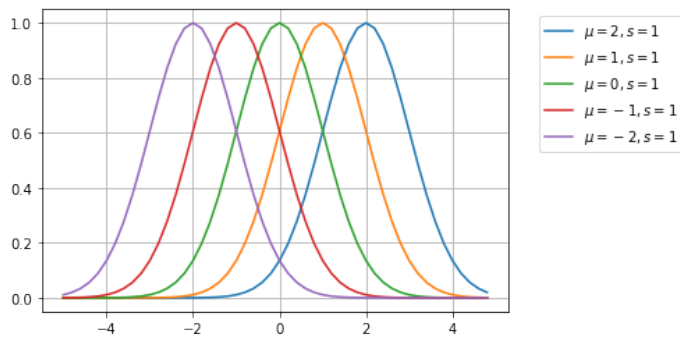


Figure 3: Example of gaussian basis functions.

Here μ defines the centre of the curve and s defines the spread of the curve. Figure 3 shows five gaussian functions with randomly chosen μ and s.

1.1.4 Fourier Basis Functions

A reasonable choice for periodic data is the fourier basis function [2]. One approach to fourier basis function [1] on a one-dimensional data is:

$$\phi_0(x) = 1, \phi_j(x) = \cos(\omega_j x + \psi_j), j > 0 \tag{9}$$

Here ω defines the frequency of the curve and ψ defines the phase of the curve. Figure 4 shows three fourier functions with randomly chosen ω and ψ.
For the region [-1,1], one can use truncated fourier series defined as follows:

$$\phi_j(x) = \begin{cases} 1, & \text{if } j = 0 \\ \sin(\pi x j), & \text{if } j \text{ is even} \\ \cos(\pi x j), & \text{if } j \text{ is odd} \end{cases} \tag{10}$$

1.1.5 Spline Basis Functions

The disadvantage with polynomial functions was they were global functions with the input variable, so change in one region affects the change in other [2]. To overcome this problem

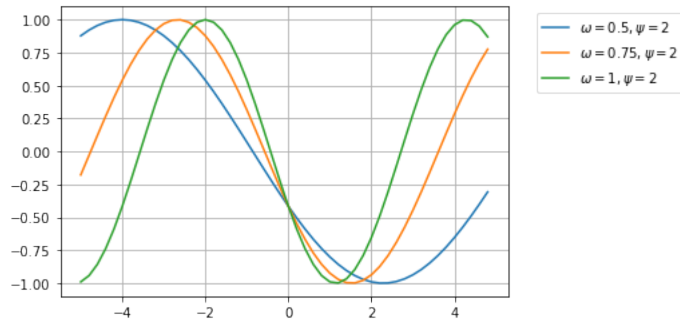


Figure 4: Example of fourier basis functions.

[3] introduced spline functions which uses a different polynomial in each region. A simple example of spline basis function in disjoint intervals $\chi_{1,2}, \ldots, \chi_K$ is:

$$\phi_{2i-1}(x) = 1_{\chi_i}(x), \phi_{2i}(x) = x1_{\chi_i}(x) \tag{11}$$

A special type of spline basis function in practice is B-spline. The B-spline basis functions are defined as:

$$\phi_i(x) = N_{i,p}(x) \tag{12}$$

Here p defines the degree of the polynomial which is chosen on the basis of domain knowledge. $N_{i,p}(u)$ is the ith B-spline basis function of degree p. We first define $m+1$ non-decreasing numbers $u_0, u_1, \ldots, u_m$ in the range $[u_0, u_m]$. These u_i's are known as knots. To define j B-spline functions with degree p we need $j+p+1$ knots. The function $N_{i,p}(u)$ is defined by Cox-deBoor recursion formula [4]:

$$N_{i,0}(u) = \begin{cases} 1, & \text{if } u_i \leq u < u_{i+1} \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

$$N_{i,p}(u) = \frac{u - u_i}{u_{i+p} - u_i} N_{i,p-1}(u) + \frac{u_{i+p+1} - u}{u_{i+p+1} - u_{i+1}} N_{i+1,p-1}(u) \tag{14}$$

1.1.6 Wavelet Basis Functions

The disadvantage with Fourier functions was they were only localized in frequency. To overcome this problem, we can use wavelet functions which are localized in both space and frequency. The wavelet basis functions are defined as:

$$\phi_j(x) = \psi(a_j x + b_j) \tag{15}$$

Here a_j defines the frequency of the wavelet and b_j defines the phase of the wavelet. $\psi(t)$ denotes the wavelet to be chosen which depends on the domain knowledge. [5] Some of the famous wavelets are :

1. Haar wavelet

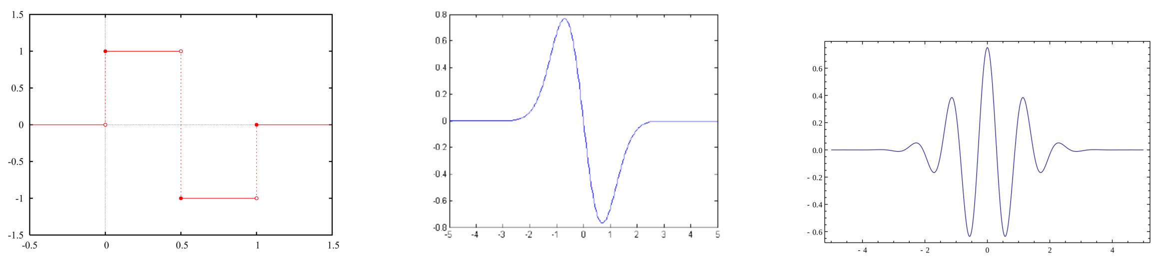
$$\psi(t) = \begin{cases} 1, & \text{if } 0 \leq t < \frac{1}{2} \\ -1, & \text{if } \frac{1}{2} \leq t < 1 \\ 0, & \text{otherwise} \end{cases} \tag{16}$$

2. Gaussian wavelet

$$\psi(t,a,b) = -(t-b)e^{-\frac{(t-b)^2}{2\sqrt{2}a^5}} \tag{17}$$

3. Morlet wavelet

$$\psi(t) = e^{-\frac{t^2}{2}}\cos(t) \tag{18}$$



(a) Haar Wavelet (b) Gaussian Wavelet (c) Morlet Wavelet

Figure 5: Example of Wavelet basis functions.

Figure 5 shows different wavelet functions as defined above.

2. Implementation

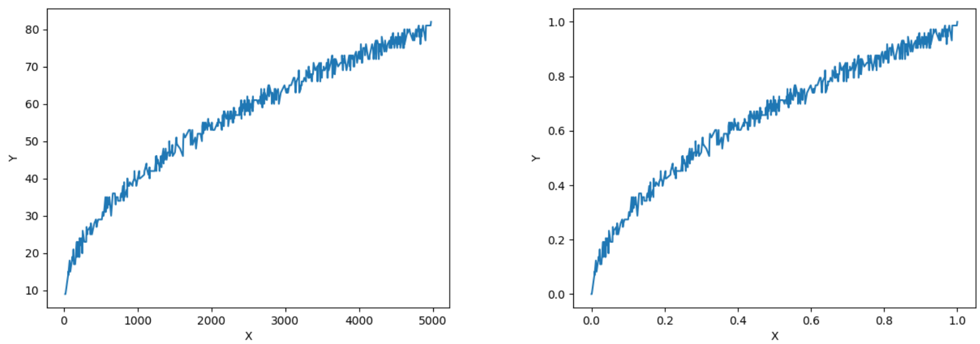
This section details the implementation of different basis functions for performing non-linear regression of data. The basis functions are implemented on C++ using Eigen library (for linear algebra) and Matplotlib (for plots).

The dataset for the implementation is generated using a Python script which takes a combination of sine, logarithmic and polynomial functions with random noise added to it. Figure 6a shows the plot for the data. The values for X are chosen randomly between 0 and 5000. The dataset consists of 500 datum points.

For the purpose of performing regression upon the data, we normalize the data using Min-Max normalization. Figure 6b shows our normalized dataset. The data is then divided into training and test data in a 70-30 split.

2.1 Polynomial Basis Functions

The number of basis functions define the complexity of the polynomial model. By iterating through different values of complexity of the model, we find out that the best accuracy on



(a) Data (b) Normalized Data

Figure 6: Dataset for non-linear regression.

our data is identified using 11 basis functions.
The basis functions used in the polynomial model are:

$$\phi_i(x) = x^i \tag{19}$$

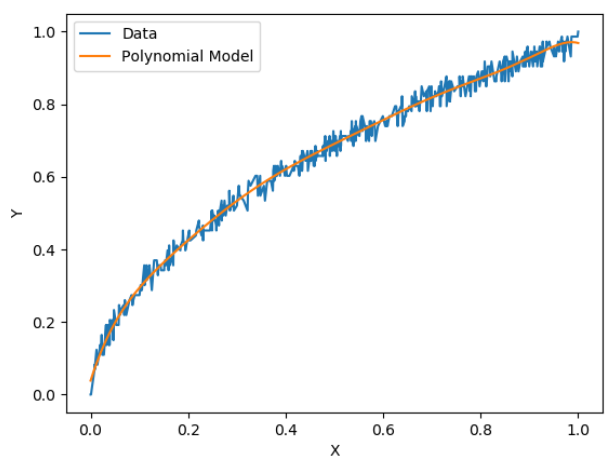


Figure 7: Model using polynomial basis functions.

The accuracy achieved by the polynomial model is 0.0004612. Figure 7 compares the actual data with the prediction data on the basis of polynomial model.

2.2 Sigmoidal Basis Functions

For the implementation of the sigmoidal model, we use 50 sigmoidal basis functions. The parameters defining the model are μ_i and σ_i for each basis function ϕ_i. The values of μ_i's are calculated by taking 50 equal-distant points in the range [0,1] as the normalised data is spread over the range [0,1]. By iterating over different values of σ_i, we find out that the best accuracy on our data is identified by setting $\sigma_i = 0.02$. The basis functions used in the sigmoidal model are:

$$\phi_i(x) = \sigma\left(\frac{x-\mu_i}{0.02}\right) \tag{20}$$

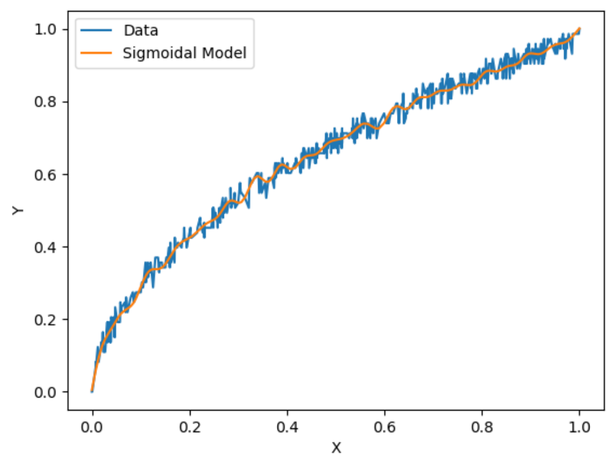


Figure 8: Model using sigmoidal basis functions.

The accuracy achieved by the sigmoidal model is 0.0004285. Figure 8 compares the actual data with the prediction data on the basis of sigmoidal model.

2.3 Gaussian Basis Functions

For the implementation of the gaussian model, we use 50 gaussian basis functions. The parameters defining the model are μ_i and σ_i for each basis function ϕ_i. The values of μ_i's are calculated by taking 50 equal-distant points in the range [0,1] as the normalised data is spread over the range [0,1]. By iterating over different values of σ_i, we find out that the best accuracy on our data is identified by setting $\sigma_i = 0.02$. The basis functions used in the gaussian model are:

$$\phi_i(x) = e^{-\frac{(x-\mu_i)^2}{2(0.02)^2}} \tag{21}$$

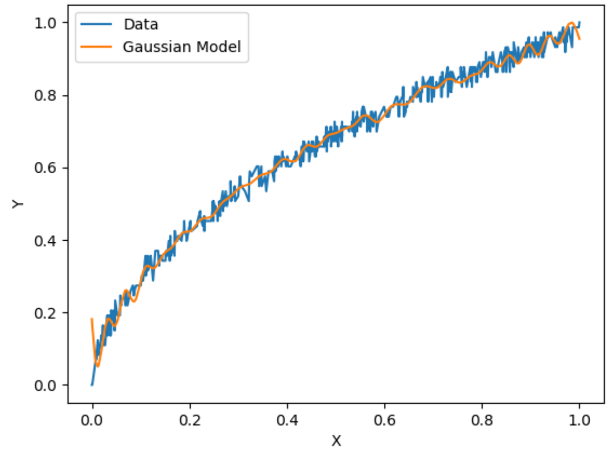


Figure 9: Model using gaussian basis functions.

The accuracy achieved by the gaussian model is 0.0005997. Figure 9 compares the actual data with the prediction data on the basis of gaussian model.

2.4 Fourier Basis Functions

For the implementation of the fourier model, we simplify the fourier basis functions as our data is spread over the domain [0,1]. We use 50 fourier basis functions for our model.
The basis functions used in the fourier model are:

$$\phi_i(x) = \begin{cases} 1, & \text{if } i = 0 \\ \sin(0.4\pi x i), & \text{if } i \text{ is even} \\ \cos(0.4\pi x i), & \text{if } i \text{ is odd} \end{cases} \tag{22}$$

The accuracy achieved by the fourier model is 0.0004418. Figure 10 compares the actual data with the prediction data on the basis of fourier model.

2.5 Spline Basis Functions

For the sake of simplicity of the spline model, we divide the range [0,1] into 50 regions and define a linear function for each region thus defining a linear spline which can fit the data. Therefore this model has 100 spline basis functions. Let the regions in the data be $\chi_0, \chi_1, \ldots, \chi_{49}$.
The basis functions used in the spline model are:

$$\phi_{2i-1}(x) = 1_{\chi_i}(x), \phi_{2i}(x) = x 1_{\chi_i}(x) \tag{23}$$

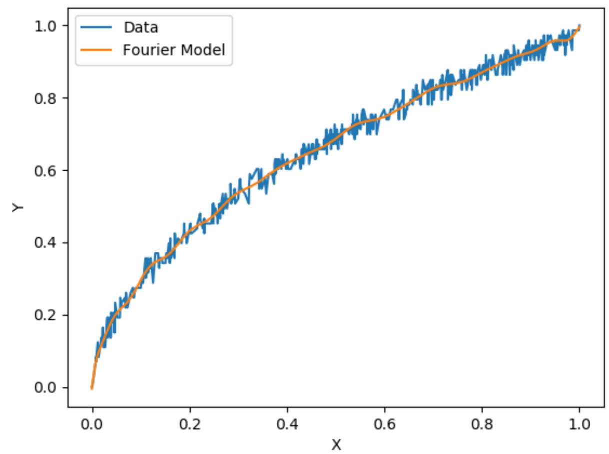


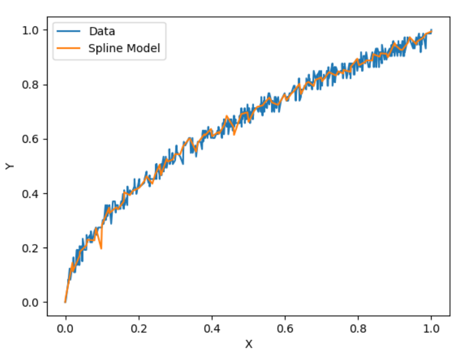
Figure 10: Model using fourier basis functions.

The accuracy achieved by the spline model is 0.0004774. Figure 11 compares the actual data with the prediction data on the basis of spline model.

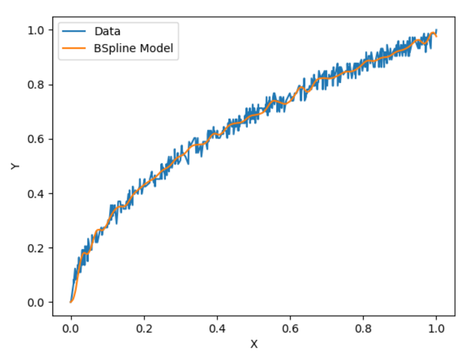
For the implementation of the B-spline model, we use 50 B-spline basis functions of degree 2. Therefore, the region [0,1] is divided into 50 equal-sized regions with a quadratic function defined for each region.
The basis functions used in the B-spline model are:

$$\phi_i(x) = N_{i,2}(x) \tag{24}$$

Here $N_{i,p}$ define the recurrence formula for a B-spline as defined in Equation 14.



(a) Spline basis functions



(b) B-Spline basis functions

Figure 11: Model using spline basis functions.

The accuracy achieved by the bspline model is 0.0004798. Figure 11b compares the actual data with the prediction data on the basis of bspline model.

2.6 Wavelet Basis Functions

For the implementation of the wavelet model, we use 50 morlet wavelet functions as defined in Equation 18. The parameters defining the model are μ_i which define the centre of the function. The values of μ_i's are calculated by taking 50 equal-distant points in the range [0,1] as the normalised data is spread over the range [0,1].
The basis functions used in the morlet model are:

$$\phi_i(x) = e^{-\frac{(10x-10\mu_i)^2}{2}} \cos(10x - 10\mu_i) \tag{25}$$

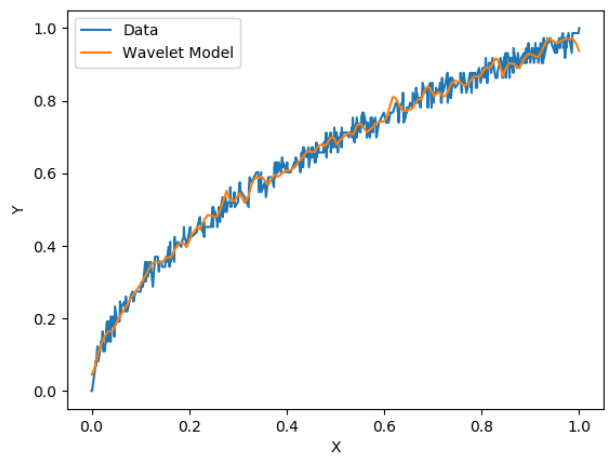


Figure 12: Model using wavelet basis functions.

The accuracy achieved by the morlet model is 0.0005148. Figure 12 compares the actual data with the prediction data on the basis of morlet model.

3. Results

This section illustrate the results obtained by performing non-linear regression using different basis functions. Figure 13 compares the Sum-of-Square Error(SSE) for different basis functions model. The minimum loss is achieved using sigmoidal basis functions.

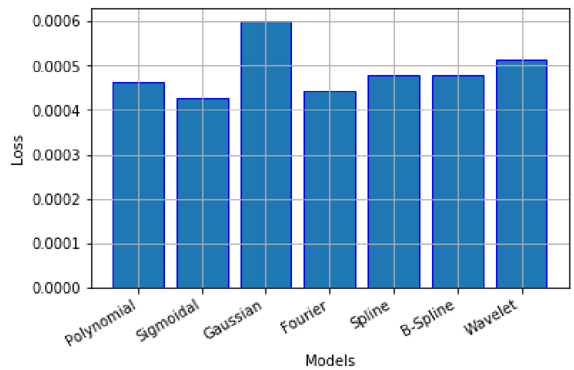


Figure 13: Sum-of-Square loss using different basis functions.

References

[1] Ryan (2018) Features and basis functions Available at https://www.cs.princeton.edu/courses/archive/fall18/cos324/files/basis-functions.pdf.

[2] Bishop CM (2016) *Pattern Recognition and Machine Learning* (Springer New York), .

[3] Hastie T, Stuetzle W (1989) Principal curves. *Journal of the American Statistical Association* 84(406):502–516.

[4] Boor CD (2001) *A practical guide to splines* (Springer), .

[5] Meyer Y, Salinger DH (2004) *Wavelets and operators* (Cambridge University Press), .