Chapter 4 - Experiments

Weber Jakob

October 22, 2020

Contents

1 Overview

A

Following the descriptions given in the previous chapters, we are now going to test the algorithm of structured additive regression using a priori knowledge on the problems given in Table 1. We will use two different, artificial functions with known behavior and one set of real life data, for which physical a priori knowledge is given.

The first function is given by

Exp.	Problem Definition	Function	n_{data}	k
1.1	Static function approx. using a priori knowledge	1	100	35
1.2	Static function approx. using a priori knowledge	2	500	35
2	Different noise levels	2	150	35
3	Different noise colors	1	200	35
4	Well-distributed data and knot placement	2	250	35
5.1	Skewed data distribution and knot placement	2	250	35
5.2	Skewed data distribution and knot placement	2	250	35
5.3	Skewed data distribution and knot placement	2	250	35
5.4	Skewed data distribution and knot placement	2	2500	35
6	Real life data	-	-	

Table 1

$$f(x) = \begin{cases} 0.1 & \text{if } x \leq 0.45 \\ 0.1 + 2\sin(\pi(x - 0.45)) & \text{else} \end{cases}., \quad \text{for } x \in [0, 1]. \tag{1}$$

It is a constant function till $x = 0.45$ and afterwards increasing. An example of the true function as well as noisy samples from it are given in Fig. 1.

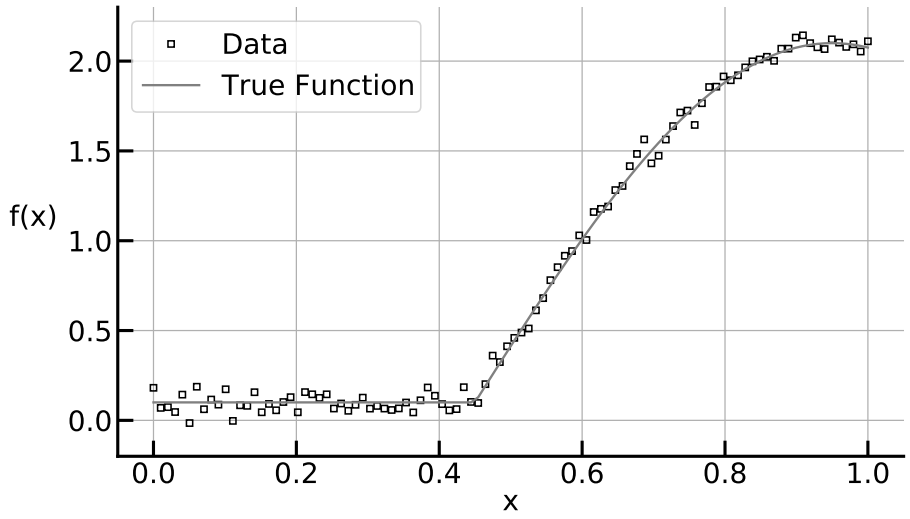

Figure 1: Test function 1

The second function that will be investigated is given by

$$f(x) = \begin{cases} 2\exp(. - \frac{(x-0.5)^2}{0.05}). + 3x & \text{, if } x \leq 0.85 \\ 2\exp(. - \frac{(0.85-0.5)^2}{0.05}). + 3 * 0.85 & \text{, else} \end{cases}., \quad \text{for } x \in [0, 1]. \tag{2}$$

The true functions depicts a unimodal behavior and is constant for $x > 0.85$. To test the algorithm, we deliberately generate "wrong" data using the following function

$$f(x) = 2\exp\left(. - \frac{(x-0.5)^2}{0.05}\right). + 3x \quad \text{for } x \in [0,1]. \tag{3}$$

An example of the true function 2 as well as noisy samples from function 3 are given in the figure 2.

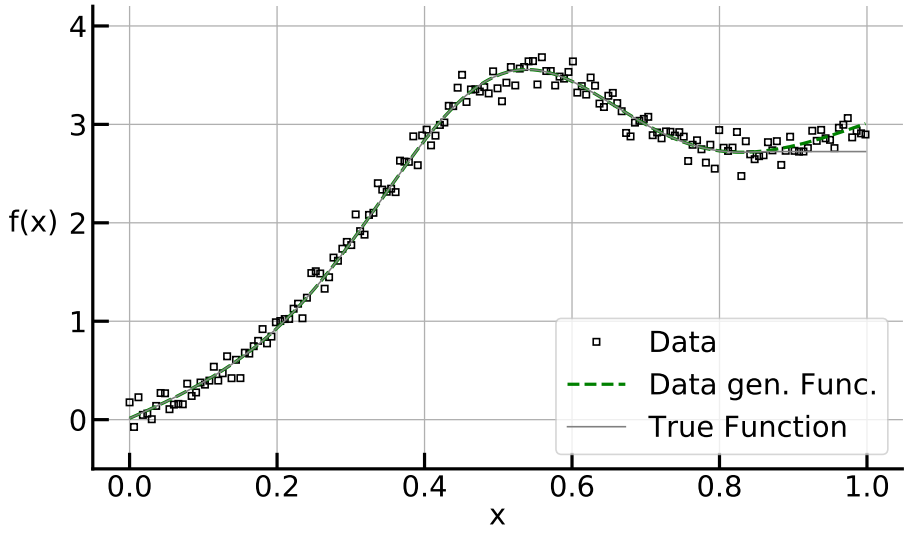

Figure 2: Test function 2

2 Experiments

2.1 Exp. 1.1: Static Function Approx. using a priori Knowledge on Function 1

For the experiment using function 1, the data set contains 100 points. We use $k = 35$ as number of splines. The smoothing parameter λ_s is optimized using cross-validation given in Chapter *CrossValidation*. The parameter λ_c regulating the influence of the constraint is set to the 1000-fold of the smoothing parameter λ_s. The resulting constrained fit, and for comparison, the unconstrained fit using the optimal smoothing parameter $\lambda_s = 2.3714$ are shown in figure 3.

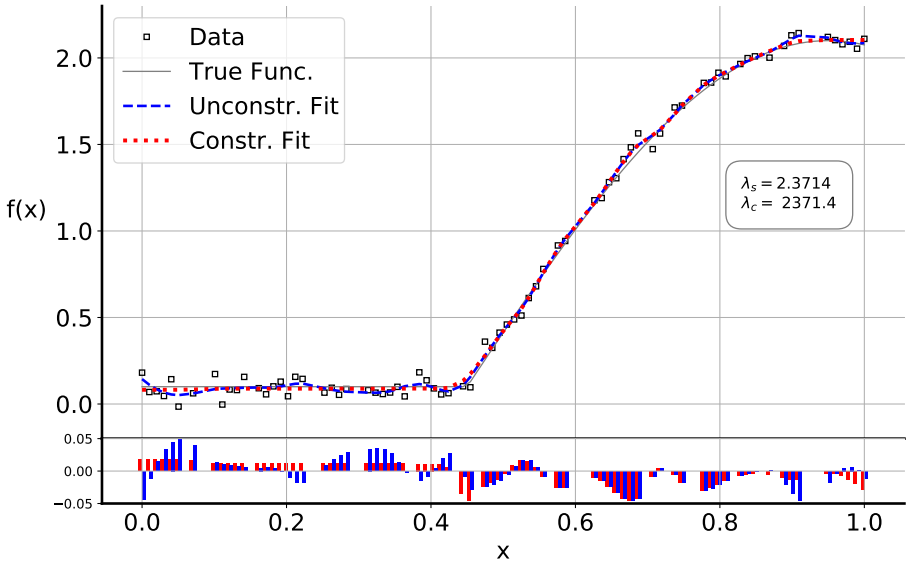

Figure 3: Constr. fit, unconstr. fit and residuals (lower plot) for function 1

The red line follows the a priori knowledge of increasing behavior better than the blue, unconstrained line. Especially in the constant area, e.g. $x \in [0, 0.45]$, the constraint improves the static function approximation and lowers the residual as seen in the lower part of figure 3. In the area of increasing behavior, the two fits are almost identical since the constraint is inactive. The mean squared errors MSE on train and test set as well as the AIC value are given in table 2. The mean squared errors for the train and test set are similar. Therefore, we can conclude that no overfitting is present. The AIC value for the constrained fit, as well as the MSEs, is/are fit are lower that the one for the unconstrained fit.

Model	**MSE**$_{train}$	**MSE**$_{test}$	**AIC**
Unconstraint	$4.913 * 10^{-4}$	$6.2156 * 10^{-4}$	-85.77
Constraint	$3.205 * 10^{-4}$	$2.6096 * 10^{-4}$	-157.5

Table 2: MSEs and AIC for Exp. 1.1

2.2 Exp. 1.2: Static Function Approx. using a priori Knowledge on Function 2

For the test using function 2 the data set contains 500 points. There are two possibilities to include the a priori knowledge of a unimodal behavior. First we can use the peak constraint matrix given in Chapter *Unimodality Constraint*. On the other hand, we can use the concavity constraint, since concave functions are likely to have a peak when the data shows a peak.

We examine both possibilities in figure 4. For each fit, we used $k = 35$ as number of splines and the smoothing parameter $\lambda_s = 510.87$ was again opti-

mized using cross-validation. The constraint parameter λ_c was set to 5109.

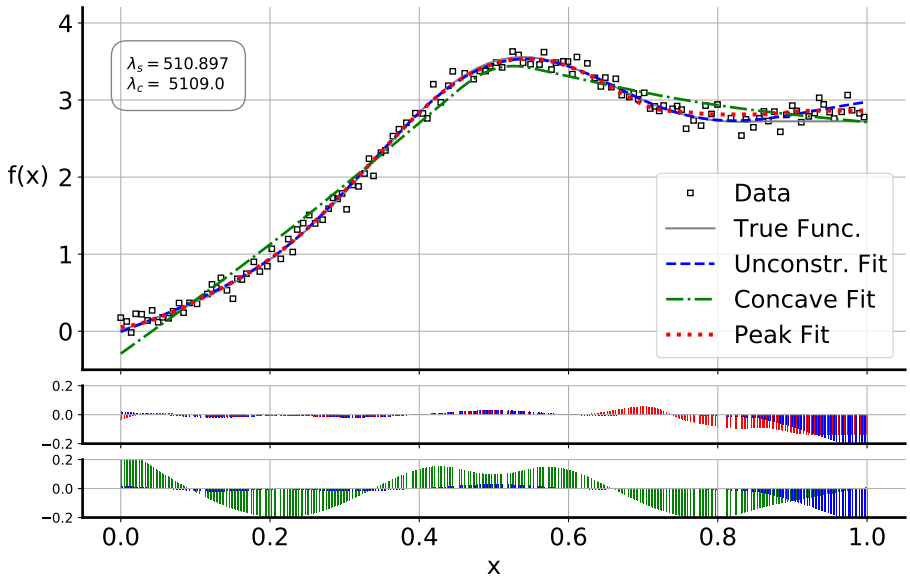

Figure 4: Constr. fit, unconstr. fit and residuals (lower plots) for function 2

The estimate for the concave constraint shows a the desired peak behavior, and is more smooth than the peak constraint estimate. This smoothness leads to changes in the sign of the residuals compared to the unconstrained fit as seen in the lower part of figure 4. The peak constraint identifies the peak correctly and produces a sufficiently smooth fit with low residual. Especially for $x > 0.8$, i.e. the region where the data is on purpose "wrong", the residual and the estimate of the peak constrained fit are almost constant while the concave constrained fit is decreasing.

The mean squared errors and the ACI value are given in the table 3. They show that the concave constraint fit is has higher MSE on train and test set, as well as a higher AIC value. The more accurate fit are therefore produced by the peak constraint. The AIC value implies to use the unconstrained model, since it possesses the lowest value. Nevertheless, as seen in figure 4, it violates the constraint for larger x. Therefore, the peak constrained model is the best model with respect to the a priori knowledge.

Model	$\mathbf{MSE}_{train}$	$\mathbf{MSE}_{test}$	**AIC**
Unconstraint	$0.3839 * 10^{-3}$	$5.6641 * 10^{-2}$	-705.1
Concave Constraint	$1.7563 * 10^{-3}$	$5.9008 * 10^{-2}$	-481.4
Peak Constraint	$0.3400 * 10^{-3}$	$5.7031 * 10^{-2}$	-696.2

Table 3: MSEs and AIC for Exp. 1.1

2.3 Constraint vs. Noise

In this section, we investigate the effect of different noise levels as well as noise types on the fitting process. In the first part, various levels of Gaussian noise are applied to function 2. In the second part, we will use different colors of noise acting on function 1. For both examples we use $k = 35$ splines and optimize the smoothing parameter λ_s using cross-validation.

2.3.1 Exp. 2: Different Noise Levels

Gaussian noise is characterized by the following two parameters:

- Location: The mean value μ of the Gaussian distribution.
- Scale: The variance value σ^2 of the Gaussian distribution.

We set the location equal to zero and vary the scale value. The effect of the noise on the fitting procedure is shown in the following figure. The chosen noise levels are $\sigma^2 = \{0.01, 0.05, 0.1\}$.

For these values of σ^2, the correspond optimized smoothing parameters are $\lambda_{0.01} = 320$, $\lambda_{0.05} = 1892$ and $\lambda_{0.1} = 1516$. The constraint parameter λ_c was set to $\lambda_c = 1000\lambda_s$.

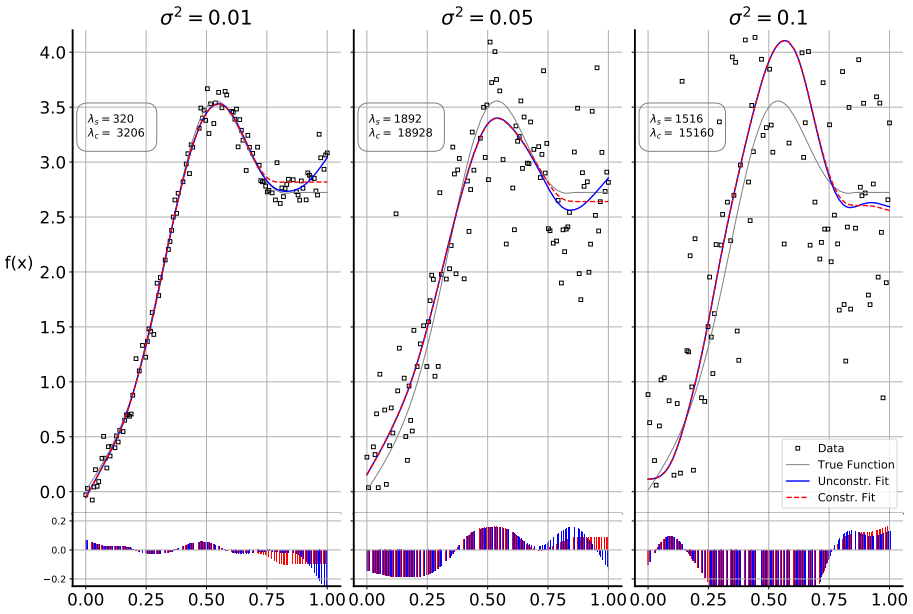

Figure 5: Constr. fit, unconstr. fit and residuals for various noise levels

The left plot shows the fit for $\sigma^2 = 0.01$. The noise level is moderate, which leads to a smooth and satisfying fit except for large x where the fit does not meet the constant function level. The peak constraint is well satisfied.

The noise level in the middle plot is already high. The unconstrained fit is nevertheless smooth, but it underestimates the peak and violates the constraint for $x > 0.8$. The constrained fit holds the peak constraint.

For the right plot and $\sigma^2 = 0.1$, both constrained and unconstrained fit are smooth but overestimate the fit especially in the peak region. The noise level is very high and the fits must be used carefully. Nevertheless, the peak constraint improves the estimate for large $x > 0.8$.

2.3.2 Exp. 3: Different Noise Colors

We investigate the effect of different noise color on the function fitting process. The tested colors are the following power-law noise types:

- White Noise
- Pink Noise
- Brownian Noise

The unconstrained as well as the increasing constrained fit for all noise colors are shown in Figure 6.

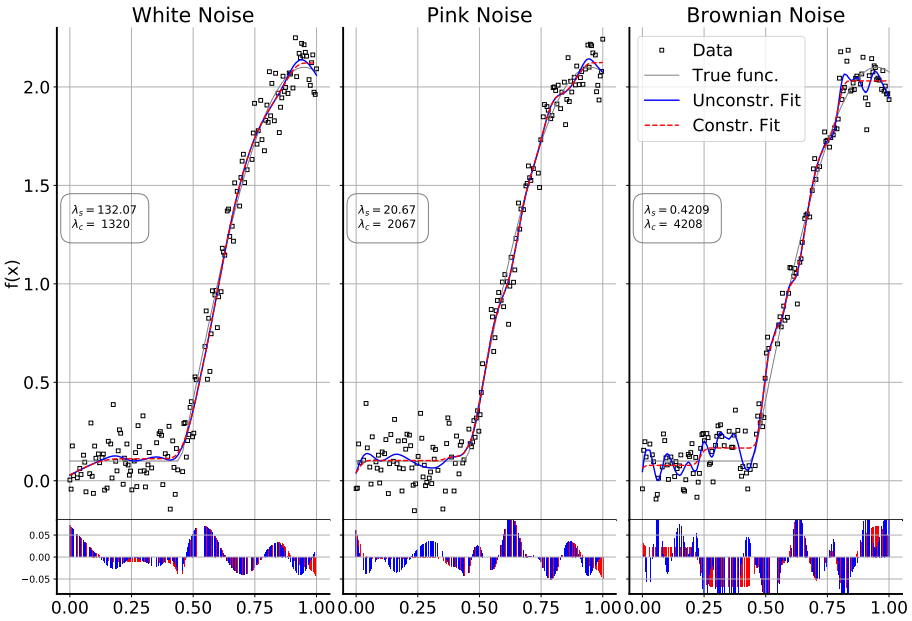

Figure 6: Constr. and unconstr. fit and residuals for noise colors

For the white noise data in the left plot, the increasing constraint produces a good fit according to the constraint and the residual except for low $x < 0.2$. The optimized smoothing parameter for the unconstrained fit was given by $\lambda_s = 132.07$ and the constraint parameter was set to $\lambda_c = 1320$. The fit recovers the true function quite well.

For the pink noise data in the middle plot, the constraint helps regularizing the fit. The constant function part was fit very well, as seen in the residual part in figure 6. The increasing part of the function, i.e. $f(x)$ for $x \geq 0.45$, was fit well. The smoothing parameter was optimized using cross-validation and set to $\lambda_s = 20.67$. The constraint parameter was set to $\lambda_c = 2067$.

For the brownian noise data in the right plot, the increasing part of the function was recovered well. The increasing constraint is hold quite well. Nevertheless, the brownian noise leads to step function like artifacts in the estimate for the constrained fit. The optimized smoothing parameter used was $\lambda_s = 0.4209$ and the constraint parameter was set to $\lambda_c = 4208$.

2.4 Well-distributed vs. Skewed Data and Knot Placement

We will now investigate the effect of irregular data distributions on the fitting process. The main question is if any knot placement type is superior to the other. We therefore examine input data $x^{(i)}$ for $i = 1, \ldots, n$ generated on a grid vs. data generated from a beta distribution given by

$$f(x) = \frac{1}{B(a,b))} x^{a-1}(1-x)^{b-1} \tag{4}$$

with

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 u^{a-1}(1-u)^{b-1} du \tag{5}$$

for different values of the parameters a and b. We will investigate the effect of the knot placement, which can either be equidistant knot placement or quantile based knot placement, using this data. We again use noisy samples from function 2 and $k = 35$ as number of splines. The smoothing parameters were optimized using cross-validation. The constraint parameters were set to the $1000 -$ fold of λ_s or adjusted by hand if to low.

2.4.1 Exp. 4: Equidistant vs. Quantile Based Knot Placement for Grid Data

We will use 250 noisy samples from function (2) generated from a equidistant sequence of x. The smoothing parameters for the equidistant knot placement fit and for the quantile based knot placement fit were optimized using cross-validation and set to $\lambda_{s,equidistant} = 17.7828$ and $\lambda_{s,quantile} = 0.0056$. The constraint parameter for the peak constraint was set to $\lambda_c = 4000$ for both fits. We use 35 splines for both fits. The resulting fits are shown in figure 7.

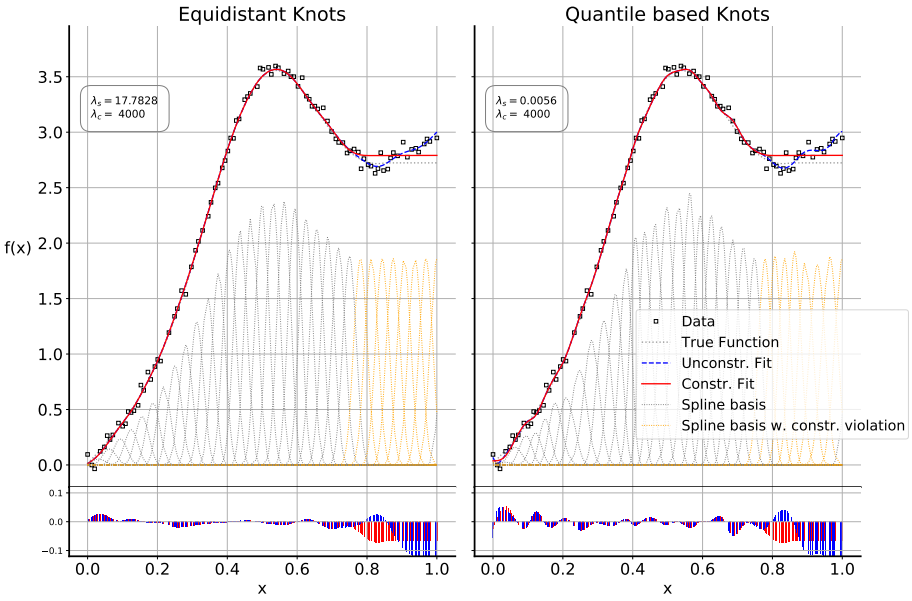

Figure 7: Constr. and unconstr. fit and residuals for grid data

Figure 7 shows not much difference between the fits for the knot placement types. The peak constraint is hold by both quite well. The residuals for equidistant knot placement are overall lower. The fit using equidistant knot placement is more smooth, since its optimal smoothing parameter is higher compared to the quantile based knot placement estimate. It is important to notice that the discontinuities in first derivative of the spline basis for both fits is just visual and comes from the number of used data points for the plot.

The mean squared errors and AIC values are given in table 4. In this table, "E" stands for equidistant knot placement and "Q" stands for quantile based knot placement. The equidistant model has lower MSEs for train and test set, as well as a lower AIC value, compared to the quantile model for both the constrained and unconstrained fit.

Model	**MSE**$_{train}$	**MSE**$_{test}$	**AIC**
Unconstraint + E	$2.72 * 10^{-3}$	$4.43 * 10^{-3}$	-280.7
Unconstraint + Q	$3.02 * 10^{-3}$	$5.20 * 10^{-3}$	-208.2
Constraint + E	$1.07 * 10^{-3}$	$1.11 * 10^{-3}$	-379.9
Constraint + Q	$1.24 * 10^{-3}$	$1.31 * 10^{-3}$	-334.4

Table 4: MSEs and AIC for Exp. 4

2.4.2 Exp. 5.1: Equidistant vs. Quantile Based Knot Placement for Skewed Data

We will now investigate the effect of non-regular data distributions for function (2) given by a beta distribution with $a = 1$ and $b = 3$. Therefore the situation is that we have many data points in the increasing part for $x \leq 0.4$ and very little

points for the constant function part at $x > 0.8$. The question is if any knot placement type is superior to the other. The fits for equidistant and quantile based knot placement can be seen in figure 8.

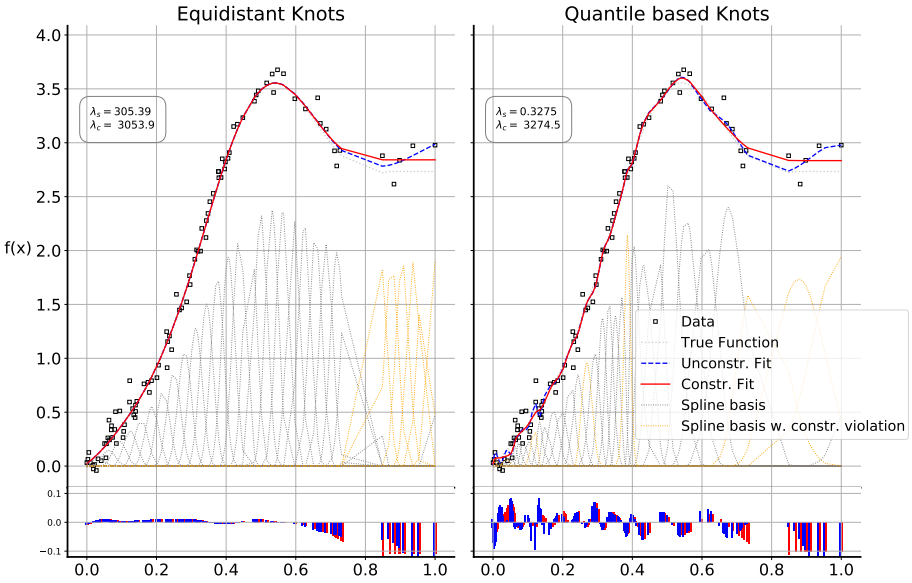

Figure 8: Constr. and unconstr. fit and residuals for left skewed data

The smoothing parameters were set to $\lambda_{s,equidistant} = 305.3856$ for equidistant knot placement and $\lambda_{s,quantile} = 0.3275$ for quantile based knot placement. This low optimal smoothing parameter leads to a wiggly estimate for the quantile based knot placement fit compared the the smooth estimate for equidistant knot placement. This can also be seen in the residual part of figure 8. The constraint parameters were set to $\lambda_{c,equidistant} = 3053.9$ and $\lambda_{c,quantile} = 3274.5$. The constraint is hold for both knot types. Especially for large $x > 0.9$ we see an improvement in the residual as well as the qualitative behavior for the constraint fits.

The mean squared errors on training and test set as well as the AIC values are given in table 5. Again, "E" stands for equidistant knot placement and "Q" stands for quantile based knot placement. The constraint, equidistant model produces the lowest MSE_{train}. The MSEs on the test set are of the same magnitude as for the train set, indication that there is no overfitting present. The lowest AIC values is present for the unconstraint equidistant model. Nevertheless, the constraint violation of this model is clearly visible in figure 8.

Model	$\textbf{MSE}_{train}$	$\textbf{MSE}_{test}$	**AIC**
Unconstraint + E	$0.98 * 10^{-3}$	$0.74 * 10^{-3}$	-421.3
Unconstraint + Q	$2.46 * 10^{-3}$	$2.08 * 10^{-3}$	-288.2
Constraint + E	$0.78 * 10^{-3}$	$1.09 * 10^{-3}$	-400.2
Constraint + Q	$1.41 * 10^{-3}$	$1.78 * 10^{-3}$	-311.5

Table 5: MSEs and AIC for Exp. 5.1

2.4.3 Exp. 5.2: Equidistant vs. Quantile Based Knot Placement for Skewed Data

We will now investigate the effect of non-regular data distributions for function (2) given by a beta distribution with $a = 3$ and $b = 3$. Therefore the situation is that we have many data points in the peak part for $x \in [0.3, 0.7]$ and less points for the constant function part at $x > 0.7$ and the increasing part at $x < 0.3$. The fits for equidistant and quantile based knot placement can be seen in figure 9.

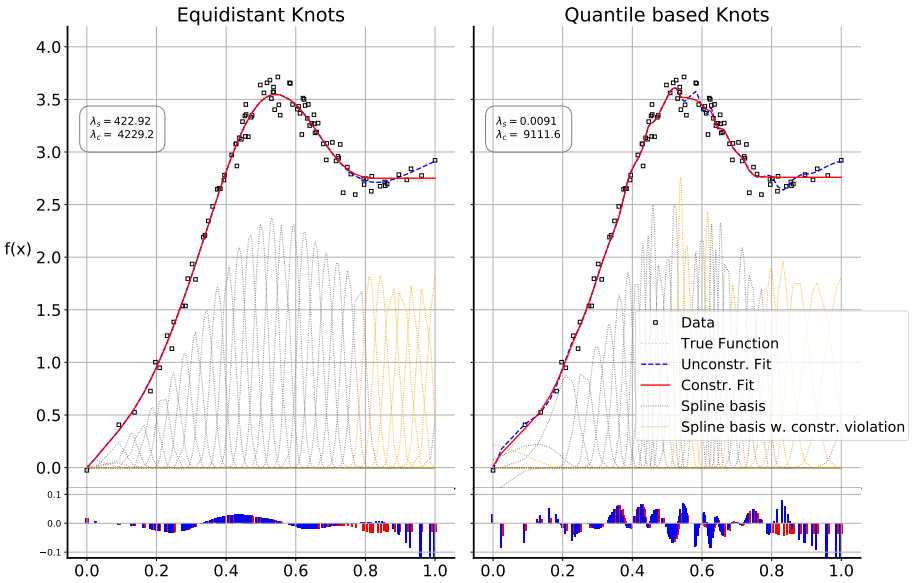

Figure 9: Constr. and unconstr. fit and residuals for middle skewed data

The smoothing parameters were set to $\lambda_{s,equidistant} = 422.92$ for equidistant knot placement and $\lambda_{s,quantile} = 0.0091$ for quantile based knot placement. The low optimal smoothing parameter again leads to a wiggly estimate in the peak region for the quantile based knot placement fit compared the the smooth estimate for equidistant knot placement. This can also be seen in the residual part of figure 9. The constraint parameters were set to $\lambda_{c,equidistant} = 4229$ and $\lambda_{c,quantile} = 9111.6$. The constraint is hold for both knot types. The equidistant knot placement produces a nearly optimal fit especially for the peak region. The quantile based knot placements fit is also quite good compared to the true function, but the estimate is quite wiggly.

The mean squared errors on training and test set as well as the AIC values are given in table 6. The situation is quite similar to the left skewed data experiment given in Chapter 2.4.2. The constraint equidistant model is the overall best model producing the lowest MSE_{train} and MSE_{test} as well as AIC value.

Model	$\mathbf{MSE}_{train}$	$\mathbf{MSE}_{test}$	AIC
Unconstraint + E	$0.96 * 10^{-3}$	$0.41 * 10^{-3}$	-459.2
Unconstraint + Q	$2.14 * 10^{-3}$	$1.85 * 10^{-3}$	-273.4
Constraint + E	$0.44 * 10^{-3}$	$0.37 * 10^{-3}$	-469.8
Constraint + Q	$0.94 * 10^{-3}$	$0.79 * 10^{-3}$	-366.1

Table 6: MSEs and AIC for Exp. 5.2

2.4.4 Exp. 5.3: Equidistant vs. Quantile Based Knot Placement for Skewed Data

We will now investigate the effect of non-regular data distributions for function (2) given by a beta distribution with $a = 3$ and $b = 1$. Therefore the situation is that we have many data points in the constant part for $x > 0.7$ and less points for the increasing and peak part $x < 0.7$. The fits for equidistant and quantile based knot placement can be seen in figure 10.

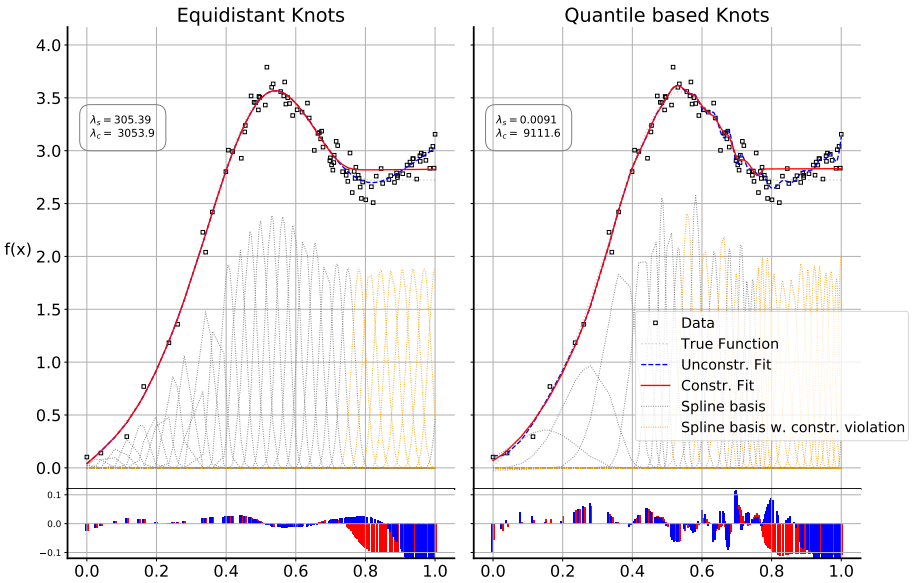

Figure 10: Constr. and unconstr. fit and residuals for right skewed data

The smoothing parameters were set to $\lambda_{s,equidistant} = 303.39$ for equidistant knot placement and $\lambda_{s,quantile} = 0.0091$ for quantile based knot placement. The low optimal smoothing parameter again leads to a wiggly estimate especially for the decreasing part of the quantile based knot placement fit compared the the smooth estimate for equidistant knot placement. This can also be seen in the residual part of figure 10. The constraint parameters were set to $\lambda_{c,equidistant} = 3033.9$ and $\lambda_{c,quantile} = 9111.6$. The constraint is hold for both knot types, except for $x \approx 0.73$ for quantile based knot placement. Both fits overestimate the level of the constant function part.

The mean squared errors on training and test set as well as the AIC values are given in table 7. The situation is quite similar to the left skewed data

experiment and middle skewed data experiment given in the previous chapters. The constraint equidistant model is the overall best model producing the lowest MSE_{train} and MSE_{test} as well as AIC value.

Model	**MSE**$_{train}$	**MSE**$_{test}$	**AIC**
Unconstraint + E	$0.89 * 10^{-2}$	$1.06 * 10^{-2}$	-253.2
Unconstraint + Q	$1.05 * 10^{-2}$	$0.99 * 10^{-2}$	-167.8
Constraint + E	$0.36 * 10^{-2}$	$0.50 * 10^{-2}$	-308.3
Constraint + Q	$0.45 * 10^{-2}$	$0.61 * 10^{-2}$	-268.2

Table 7: MSEs and AIC for Exp. 5.3

2.4.5 Exp. 5.4: Equidistant vs. Quantile Based Knot Placement for Skewed Data

We will now investigate the effect of non-regular data distributions for function (2) given by a beta distribution with $a = 1$ and $b = 3$, this time using $n = 2500$ data points. The fits for equidistant and quantile based knot placement can be seen in figure 11.

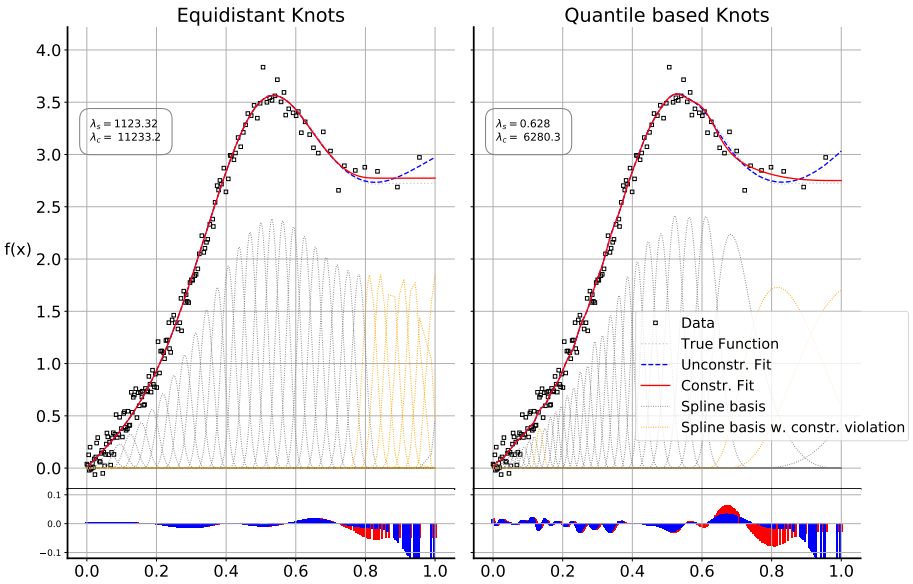

Figure 11: Constr. and unconstr. fit and residuals for left skewed data

The smoothing parameters were set to $\lambda_{s,equidistant} = 1123.32$ for equidistant knot placement and $\lambda_{s,quantile} = 0.628$ for quantile based knot placement. The constraint parameters were set to $\lambda_{c,equidistant} = 11233.2$ and $\lambda_{c,quantile} = 6280.3$. The constraint is hold for both knot types. Both fits produces a nearly optimal fit especially for the peak region. Using a higher number of data points relaxes the problem of a wiggly estimate for the quantile based knot placement fit. Nevertheless, its residual is higher than for equidistant knot placement, which is not affected by the number of data point.

The mean squared errors on training and test set as well as the AIC values are given in table 8. The situation is similar to the experiment given in the previous chapters. The constraint equidistant model is the overall best model producing the lowest MSE_{train} and second lowest MSE_{test} as well as AIC value, while holding the constraint. Nevertheless, using a larger number of data points, the differences between equidistant and quantile based knot placement in terms of accuracy and constraint fidelity vanish.

Model	**MSE**$_{train}$	**MSE**$_{test}$	**AIC**
Unconstraint + E	$2.25 * 10^{-4}$	$1.20 * 10^{-4}$	-5605.4
Unconstraint + Q	$3.76 * 10^{-4}$	$2.27 * 10^{-4}$	-5135.1
Constraint + E	$1.29 * 10^{-4}$	$1.31 * 10^{-4}$	-5553.9
Constraint + Q	$3.30 * 10^{-4}$	$3.33 * 10^{-4}$	-4908.5

Table 8: MSEs and AIC for Exp. 5.4

In summary, equidistant knot placement produces better fits in terms of smoothness and accuracy for a low number of data points. Its residuals are overall lower than the ones from quantile based knot placement. This comes mostly from the fact that the optimal smoothing parameter for quantile based knot placement is quite low for all experiments compared to the equidistant equivalent. We can therefore conclude that quantile based knot placement using the weighting approach given in **ferziger2008numerische** for a lower number of data points is not beneficial in terms of fit quality compared to equidistant knot placement. Further, we can conclude that the incorporation of a priori knowledge improves the qualitative and quantitative behavior of the fit even in situations where the data is not well distributed.

2.5 Exp. 6: Real life data

We will now try to incorporate a priori knowledge in the fitting process using real-life data. The data is generated from a heat treatment process of aluminum. The aluminum is heated to a specific temperature and then cooled using water jets.

We try to estimate the heat transfer coefficient $\alpha := \alpha(T, \dot{m})$ as a function of the temperature T and the mass flow $\dot{m}$. We know beforehand, that the heat transfer coefficient α may only increase when increasing the mass flow $\dot{m}$ and that it shows a peak behavior for increasing temperature T.

A further problem is the data situation given in figure 12. Here we show how many data points are given in a area of approximately $50K$ and $0.9[kg/s]$. We have some regions, where no data is available, while the majority of data points is located in small areas.

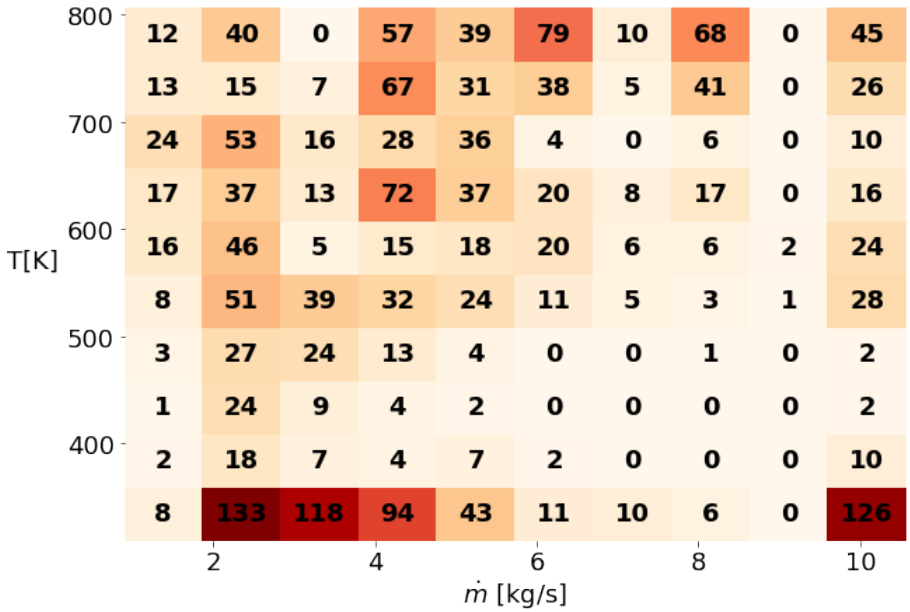

Figure 12: Data Situation

We use the additive model ansatz using two spline models, $s_1(T)$ and $s_2(\dot{m})$, and constrain $s_1(T)$ by the unimodality constraint and $s_2(\dot{m})$ by the increasing constraint. We use $k = 50$ splines for both. The smoothing parameters were set to $\lambda_{s,s_1} = 908.52$ and $\lambda_{s,s_2} = 383.12$. The constraint parameters were set to $\lambda_{c,s_1} = 9085.2$ and $\lambda_{c,s_2} = 3831.2$.

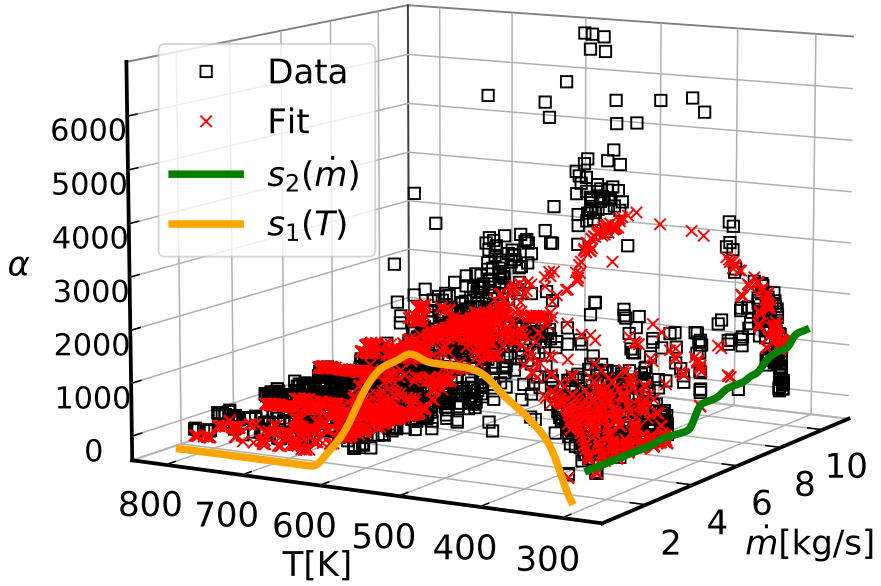

Figure 13: Constr. fit for real-life data

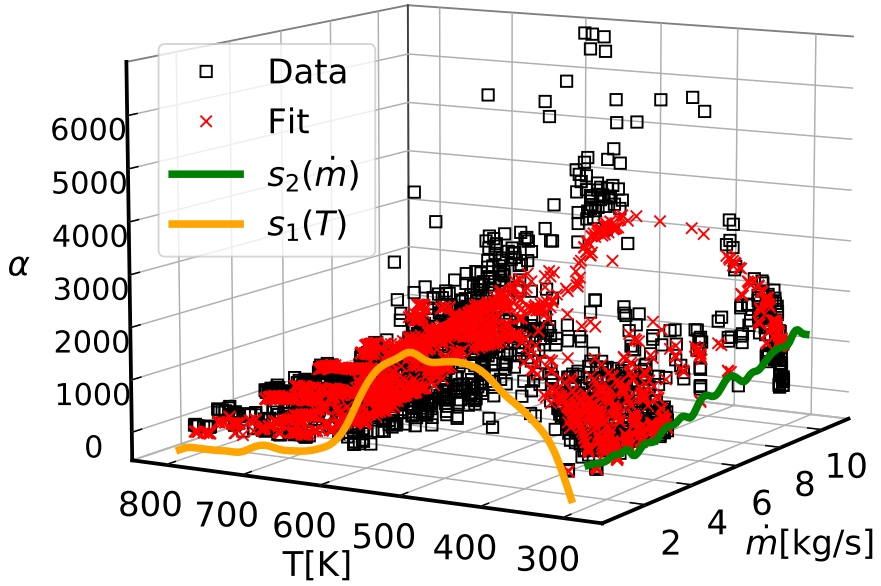

Figure 14: Unconstr. fit for real-life data

The a priori knowledge is hold quite well for the constraint fit given in figure 13. In comparison, the unconstraint fit using the same smoothing parameters is given in figure 14. Here, especially the mass flow dependent function $s_2(\dot{m})$ does not hold the constraint of being increasing. The temperature dependent function $s_1(T)$ is quite smooth, but also violates the unimodality constraint for $T \approx 700$.

The mean squared errors and AIC values for both fits are shown in table 9. The constraint fit has lower MSE$_{train}$ and MSE$_t est$ as well as AIC value.

Model	**MSE**$_{train}$	**MSE**$_{test}$	**AIC**
Unconstraint	$1.5178 * 10^6$	$1.3843 * 10^6$	2748.5
Constraint	$1.5161 * 10^6$	$1.3808 * 10^6$	2723.9

Table 9: MSEs and AIC for Exp. 6

3 Summary

The incorporation of a priori knowledge in the fitting process using structured additive regression, B-splines, a smoothness penalty and a user-defined con-

straint penalty is possible for simulated and real-life data. The use of a priori knowledge improves the quality of the fit and makes the estimation process more robust against noise, data outliers and measurements errors.