Ezhov, N., Neitzel, F., Petrovic, S. (2018): Spline approximation, Part 1: Basic methodology. - *Journal of Applied Geodesy*, *12*, 2, pp. 139—155.

DOI: http://doi.org/10.1515/jag-2017-0029

Spline Approximation
Part 1: Basic Methodology

Nikolaj Ezhov[1], Frank Neitzel[1] and Svetozar Petrovic[1,2]

[1] Technische Universität Berlin
Institute of Geodesy and Geoinformation Science
Str. des 17. Juni 135
10623 Berlin, Germany
nikolaj.ezhov@alumni.tu-berlin.de | frank.neitzel@tu-berlin.de

[2] GFZ German Research Centre for Geosciences
Section 1.2: Global Geomonitoring and Gravity Field
Telegrafenberg
14473 Potsdam, Germany
svetozar.petrovic@campus.tu-berlin.de

Abstract

In engineering geodesy point clouds derived from terrestrial laser scanning or from photogrammetric approaches are almost never used as final results. For further processing and analysis a curve or surface approximation with a continuous mathematical function is required. In this paper the approximation of 2D curves by means of splines is treated. Splines offer quite flexible and elegant solutions for interpolation or approximation of "irregularly" distributed data. Depending on the problem they can be expressed as a function or as a set of equations that depend on some parameter. Many different types of splines can be used for spline approximation and all of them have certain advantages and disadvantages depending on the approximation problem. In a series of three articles spline approximation is presented from a geodetic point of view. In this paper (Part 1) the basic methodology of spline approximation is demonstrated using splines constructed from ordinary polynomials and splines constructed from truncated polynomials. In the forthcoming Part 2 the notion of B-spline will be explained in a unique way, namely by using the concept of convex combinations. The numerical stability of all spline approximation approaches as well as the utilization of splines for deformation detection will be investigated on numerical examples in Part 3.

Keywords: Spline, B-Spline, curve, interpolation, approximation

1 Introduction

In engineering geodesy point clouds derived from areal (area-wise, in contrast to point-wise) measurement methods such as terrestrial laser scanning or photogrammetry are almost never used as final result. In computer-aided design (CAD) or deformation analyses a curve or surface approximation with a continuous mathematical function is required. An overview of curve and surface approximation of 3D point clouds, including polynomial curves, Bézier curves, B-Spline curves and NURBS curves, is given by Bureick et al. [2016b]. This paper will focus on the approximation of 2D curves by means of splines.

In the literature several different mathematical representations of splines can be found. The notion of spline, however, is almost always related to B-splines as pointed out by Lucas [2003]. In general, the goal of this article is to describe, in a most comprehensible way, different mathematical representations of a spline in 2D and its utilization in data approximation. In this article only univariate splines are used, in a form of a spline function or in a form of a parametric spline curve. For the approximation the minimization of L_2 norm (least-squares adjustment) is used.

When it comes to teaching or explaining splines, throughout the literature various approaches can be found. In general, they start with a brief or long historical introduction, sometimes with a quite elaborate explanation of their development as presented e.g. by Farin [1993]. However, the discussion of splines, in most of the literature, quickly turns into elaboration of Bézier curves and B-splines (basis splines), see e.g. the textbooks by De Boor [1978] or Farin [1993]. For a person not familiar with Bézier curves or B-splines this transition seems quite unnatural and confusing. In addition, one can get a false impression that splines can be represented only in those forms.

Another important point is that beside the complex and non-intuitive mathematical formulation of the B-splines, it is hard to find an explicit and comprehensive derivation in the literature. In that fashion, usually for the computation of a

B-spline, immediately the De-Boor's algorithm [De Boor 1986] is prescribed as a recipe, and the further computation is performed without any deeper understanding of the B-splines as such. Our impression is that in some articles the B-spline approximation is done even without any deeper understanding of least-squares adjustment. However, the B-splines and NURBS, see e.g. the textbook by Piegl and Tiller [2012], were mainly developed and widely used in the field of computer graphics. Their intensive use in this field significantly influenced the use of B-splines in many other fields as well. Therefore, it is natural that all methods for spline computation (interpolation, approximation, smoothing etc.) are mainly interpreted in the above mentioned ways.

When it comes to quantitative analysis, almost in all of the listed literature regarding the cubic spline and especially the B-spline approximation, there is no proper quality assessment in the sense that there is almost no interpretation of the residuals, no standard deviations, no reliability measures. The final result is lacking quantitative information about precision and reliability. Thus, there is no numerical justification for the quality of the approximation. In all examples results are interpreted mainly by visualization. It is completely understandable that the quantification is not of crucial importance in computer graphics, because the approximation is used mainly for getting a result which satisfies aesthetic requirements. On the contrary, in geodesy quality assessment is a process, which is as important as the data approximation.

In the field of geodesy, the notion of splines, particularly the spline approximation is considered in a different way. However, the existing literature in general does not clearly elaborate these notions, since there is a lack of simple and comprehensible publications on splines. Therefore, this article is a qualitative research on the topic of splines, from a geodetic point of view, which elaborates in details the univariate spline approximation in most of its forms.

The main goal of this contribution is a detailed description of how to derive and interpret different types of splines (without B-splines) from a geodetic point of view. This comprises, after the definition of a spline in section 2, the following objectives:

(i) Spline function constructed from ordinary cubic polynomials (Section 3).
(ii) Spline function constructed from truncated polynomials (Section 4).
(iii) Spline curve constructed from ordinary cubic polynomials (Section 5).

In all three cases a piecewise approximation with polynomials of degree three with constraints for continuity, smoothness and curvature is applied. In (i) and (ii) spline functions in the form $y = f(x)$ are used. In (i) the constraints for continuity, smoothness and curvature are formulated explicitly while the constraints in (ii) are implicitly taken into account. Thus (i) and (ii) are equivalent formulations for the same problem. In (iii) spline functions of the form $x = f_1(t)$ and $y = f_2(t)$ are applied. In contrast to (i) and (ii), x and y are considered independently of each other which enables the representation of arbitrary curves.

In Part 2 of a series of three articles a clear and comprehensive explanation of B-splines from a geodetic point of view will be elaborated. In Part 3 the considered approaches for spline approximation will be compared by using numerical examples.

2 Definition of a spline

Historically the word "spline" appears in the naval industry. It is a thin rod of some flexible material equipped with a groove and a set of weights (called ducks or rats) with attached arms designed to fit into the groove, see Figure 1. Such device was used by architects (particularly naval architects) for drawing smooth curves trough a set of given points. To accomplish this, the spline was forced to pass through the given points by adjusting the location of the ducks along the rod.

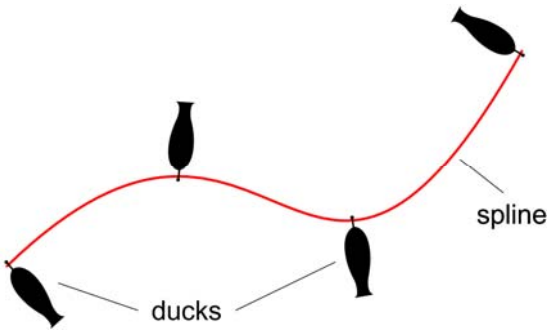

Figure 1: Spline and ducks

The situation in Figure 1 can be seen as a mechanical realization of a spline *interpolation* where one tries to interpolate given knots by means of piecewise continuous polynomials which meet at these knots. Considering a 2D Cartesian coordinate system, see Figure 2, the coordinates of knots are denoted by (x_{k_j}, y_{k_j}); hence j = 0, 1, …, n is the index (ordinal number) of the knot and n the number of segments. Character k stands for "knot" and is not an index; hence x_{k_j} means "x-value of the knot j".

Schumaker [2007, p. 6] explains that in the mid-1700s the following was discovered by Euler and the Bernoulli brothers. Knowing that the shape of such a bent rod has smallest strain energy, they found how to approximate its centerline with a function. To illustrate it, suppose that the points of contact of the ducks with the spline are located at the points (x_{k_j}, y_{k_j}) for j = 0, 1, ..., n in a Cartesian plane as depicted in Figure 2. Then the centerline of the spline is approximately given by the function $S(x)$ with the following properties:

1. $S(x)$ is a piecewise cubic polynomial with knots at x_{k_0}, x_{k_1}, …, x_{k_n},
2. $S(x)$ is a linear polynomial for $x \le x_{k_0}$ and $x \ge x_{k_n}$,
3. $S(x)$ has two continuous derivatives everywhere,
4. $S(x_{k_j}) = y_{k_j}$ for j = 0, 1, ..., n.

Schumaker [2007, p. 7] concludes that, defined like this, $S(x)$ (see Figure 2) represents, in a way, the best interpolating function. This is one of the reasons why the cubic spline is accepted as the most relevant of all splines. Consequently, later in mathematics the term "spline" refers to a smooth piecewise polynomial function. It is widely accepted that it is initially introduced by Schoenberg [1946]. The main motivation for introduction of splines is creation of a stable interpolating function through an arbitrary number of points, which would preserve the smoothness (curvature) as well as the shape of the dataset. The preservation of the curvature at the knots is of crucial importance, e.g. in engineering (railways, roads, etc.).

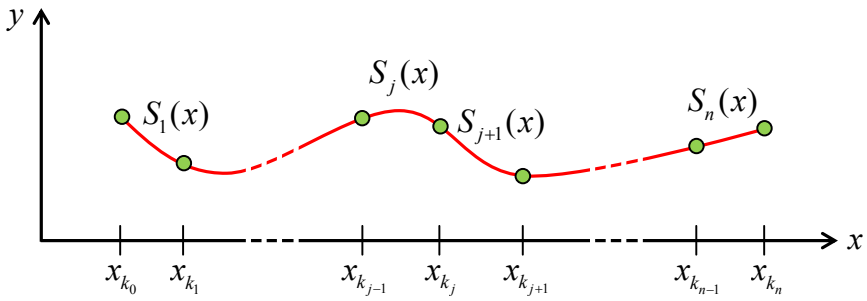

Figure 2: Spline interpolation

According to the above mentioned properties we can define a spline function as

$$S(x) = \begin{cases} S_1(x), & x_{k_0} \le x \le x_{k_1} \\ \vdots & \\ S_j(x), & x_{k_{j-1}} \le x \le x_{k_j} \\ \vdots & \\ S_n(x), & x_{k_{n-1}} \le x \le x_{k_n} \end{cases}, \tag{1}$$

where the function for a cubic polynomial in each segment j can be written as

$$S_j(x) = a_{0_j} + a_{1_j} x + a_{2_j} x^2 + a_{3_j} x^3, \quad j = 1, 2, \ldots, n. \tag{2}$$

The property that the spline has to meet exactly the data at the given knots yields the conditions

$$S_j(x_{k_{j-1}}) = y_{k_{j-1}} \text{ and } S_j(x_{k_j}) = y_{k_j}, \quad j = 1, 2, \ldots, n. \tag{3}$$

The property that $S(x)$ has two continuous derivatives everywhere yields the constraints

$$S'_j(x_{k_j}) = S'_{j+1}(x_{k_j}), \quad j = 1,2,\ldots,n-1, \tag{4}$$

$$S''_j(x_{k_j}) = S''_{j+1}(x_{k_j}), \quad j = 1,2,\ldots,n-1. \tag{5}$$

From equation (2) and Figure 2 we can notice n polynomial functions $S_j(x)$ and $n + 1$ knots. Since the cubic function has four parameters, the spline $S(x)$ has $4n$ unknowns. From (3) we obtain $2n$ and from (4) and (5) we get $2(n - 1)$ condition equations. This results in $4n - 2$ condition equations to solve for $4n$ unknowns and hence the equation system is underdetermined.

To solve this problem, arbitrary conditions can be introduced, but usually the missing constraints are introduced at the first and the last knot. Three types of conditions, also denoted as "boundary conditions" are commonly used, see e.g. the handbook of mathematics by Bronshtein and Semendyayev [2007, p. 931]:

a) $S''(x_{k_0}) = S''(x_{k_n}) = 0$, yielding a so-called natural spline,

b) $S'(x_{k_0}) = f_0'$, $S'(x_{k_n}) = f_n'$ where f_0' and f_n' are given values,

c) $S(x_{k_0}) = S(x_{k_n})$, in the case of $f_0 = f_n$, $S'(x_{k_0}) = S'(x_{k_n})$ and $S''(x_{k_0}) = S''(x_{k_n})$, yielding a periodic spline.

Notice that the "boundary condition" a) is equivalent with the "property" 2. in the above definition of Schumaker [2007, p. 6] which describes the behavior of an elastic rod, which outside the region delimited by knots adopts the (natural) linear form ($S'' = 0$). The determination of the coefficients for the cubic interpolation spline is explained e.g. in the aforementioned handbook by Bronshtein and Semendyayev [2007, p. 932].

The main application of spline interpolation is designing curves through a sparse sequence of data points. For example in CAD the user wants to obtain an aesthetically pleasing curve even when the data points that are used as knots for the resulting spline are quite sparse.

In engineering geodesy this situation changes since from modern sensors such as terrestrial laser scanners very dense sequences of data points can be captured with high acquisition speed of up to one million points per second. Due to the high point density and the fact that measurements are affected by random measurement errors, it is no longer appropriate to apply a spline interpolation. Observation errors and other abrupt changes in the data points would be modeled, resulting in a strongly oscillating spline. The solution is to divide the sequence of points into not so many intervals determined by the predefined knots. Within these intervals we consider an overdetermined configuration and hence the parameters of a piecewise cubic polynomial can be computed via least-squares adjustment. Conditions for smoothness at the knots (where two polynomials meet) can be introduced as additional constraints.

This case, where the spline does not follow all the data points exactly but as close as possible under a predefined target function, e.g. sum of the weighted squared residuals, is called spline *approximation*, see Figure 3. In the following different approaches for spline approximation will be elaborated.

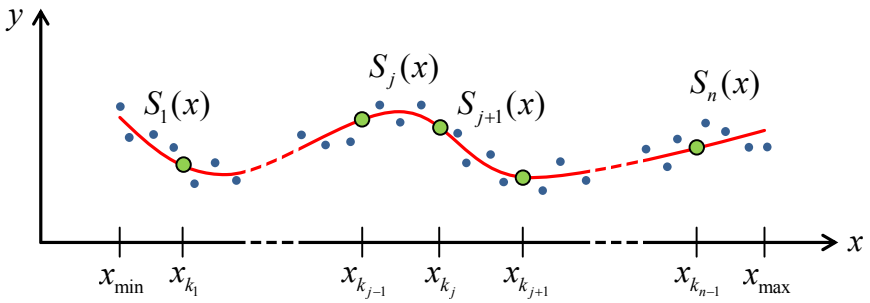

Figure 3: Spline approximation

3 Spline function constructed from ordinary cubic polynomials

The spline function constructed from ordinary cubic polynomials offers the most intuitive approach to the problem from a geodetic point of view. For formulating this approximation problem, cubic polynomials are used for each piece of the spline. Furthermore, three different equations are introduced as additional constraints at the knots. The values y_i are introduced as observations, the values x_i are introduced as fixed (error-free) values. Examples for this case are:

- Signal registration at certain time intervals, whereby the time measurement is much more precise than the signal registration and can therefore be regarded as error-free.
- Measurements at fixed sensor positions, e.g. measurement of the vertical component during a load test of a structure with displacement sensors located at fixed positions.
- Approximation of point clouds to obtain a result which satisfies aesthetic requirements and where the interpretation of residuals (e.g. for deformation analysis) is not the main focus.

In the following we always consider an overdetermined configuration that enables us to compute the parameters of the spline via least-squares adjustment (L_2 norm minimization). In contrast to spline interpolation, in spline approximation a solution can be found from minimization of the sum of the weighted squared residuals without imposing any extra conditions to the first and the last point of the spline. Conditions for smoothness at the "inner" knots can be introduced as additional constraints.

A cubic spline approximation as a special case of restricted least squares is proposed by Buse and Lim [1977]. However, they introduced constraints at the end nodes, and their solution for the system of equations is slightly different. There is also a solution without constraints at the end knots, proposed by Klaus and Van Ness [1967], but with a different mathematical definition of the spline, which was initially published by Walsh et al. [1962]. It should be emphasized that the spline approximation approach proposed in this section is not new and probably was already used before. However, the authors were not able to find a publication where exactly the same approach is presented.

3.1 Definition of the problem

Let:

- $y_{1_j}, y_{2_j}, \ldots, y_{i_j}, \ldots, y_{m_j}$ be a sequence of observed values within an interval j, with m_j being the number of observations within the interval j,
- $\mathbf{\Sigma}_{LL} = \sigma_0^2 \mathbf{Q}_{LL}$ be the variance-covariance matrix of all observations y_{i_j},
- $x_{1_j} < x_{2_j} < \ldots < x_{i_j} < \ldots < x_{m_j}$ be a non-decreasing sequence of error-free values referring to the observations within an interval j, and
- $x_{k_1} < x_{k_2} < \ldots < x_{k_j} < \ldots < x_{k_{n-1}}$ be a non-decreasing sequence of user-defined values for the knots on the x-axis which are regarded as error-free. As mentioned before, the first knot x_{k_0} and the last knot x_{k_n}, see Figure 2, have no particular influence on the adjustment, they are just $x_{k_0} = x_{\min}$ and $x_{k_n} = x_{\max}$. Therefore, when the positioning of the knots is discussed we refer only to the knots $x_{k_1}, \ldots, x_{k_j}, \ldots, x_{k_{n-1}}$, see Figure 3, on which some constraints are imposed. The positioning of the knots plays a key role in spline approximation problems. This problem is further discussed in section 3.4.

To determine:

- $(a_{0_1}, a_{1_1}, a_{2_1}, a_{3_1}), \ldots, (a_{0_j}, a_{1_j}, a_{2_j}, a_{3_j}), \ldots, (a_{0_n}, a_{1_n}, a_{2_n}, a_{3_n})$ which represent a set of parameters for the cubic polynomials $S_j(x)$ that has to be estimated for all intervals $[x_{k_{j-1}}, x_{k_j})$.

According to (2) the functional model for the cubic polynomials reads

$$y_{i_j} = S_j(x_{i_j}) = a_{0_j} + a_{1_j} x_{i_j} + a_{2_j} x_{i_j}^2 + a_{3_j} x_{i_j}^3 \quad , \tag{6}$$

with j = 1, 2, ..., n as index for an interval $[x_{k_{j-1}}, x_{k_j})$. Furthermore, i = 1, 2, ..., m_j represents an index for all pairs of observed and fixed values within an interval j.

Since we consider an overdetermined system of equations, a spline approximation is performed without putting any extra conditions on the first and the last knots of the spline. The dataset is subdivided into intervals w.r.t. the x-axis by predefined knots. For each interval $[x_{k_{j-1}}, x_{k_j})$ there is a set of at least two pairs of values (x_{i_j}, y_{i_j}) inside of it. According to (3), (4) and (5) for constructing a spline, the conditions

$$S_j(x_{k_j}) = S_{j+1}(x_{k_j}), \quad j = 1, \ldots, n-1 \quad , \tag{7}$$

$$S'_j(x_{k_j}) = S'_{j+1}(x_{k_j}), \quad j = 1, \ldots, n-1 \quad , \tag{8}$$

$$S''_j(x_{k_j}) = S''_{j+1}(x_{k_j}), \quad j = 1, \ldots, n-1 \tag{9}$$

are imposed on the knots. These knots x_{k_j} are also coordinates on the *x*-axis but their number and their values are freely specified, or can be determined by some knot placement strategy, see section 3.4.

After defining all these quantities, we can create a spline approximation from an overdetermined configuration using the Gauss-Markov model with additional constraints for the unknowns.

3.2 Formulation of the adjustment problem

Based on the functional model (6) considering y_{i_j} as observations and x_{i_j} as fixed values the observation equations

$$\begin{aligned} y_{i_1} + v_{i_1} &= \hat{a}_{0_1} + \hat{a}_{1_1} x_{i_1} + \hat{a}_{2_1} x_{i_1}^2 + \hat{a}_{3_1} x_{i_1}^3 \qquad i = 1,\ldots,m_1 \,, \\ &\vdots \\ y_{i_j} + v_{i_j} &= \hat{a}_{0_j} + \hat{a}_{1_j} x_{i_j} + \hat{a}_{2_j} x_{i_j}^2 + \hat{a}_{3_j} x_{i_j}^3 \qquad i = 1,\ldots,m_j \,, \\ &\vdots \\ y_{i_n} + v_{i_n} &= \hat{a}_{0_n} + \hat{a}_{1_n} x_{i_n} + \hat{a}_{2_n} x_{i_n}^2 + \hat{a}_{3_n} x_{i_n}^3 \qquad i = 1,\ldots,m_n \end{aligned} \tag{10}$$

for each segment $j = 1, \ldots, n$ of the spline can be set up. The construction of the observation vector is presented in two steps. First, all observation vectors for each segment of the spline are introduced as

$$\begin{aligned} \mathbf{L}_1 &= \begin{bmatrix} y_{1_1} & y_{2_1} & y_{3_1} & \cdots & y_{m_1} \end{bmatrix}^{\mathrm{T}} \,, \\ &\vdots \\ \mathbf{L}_j &= \begin{bmatrix} y_{1_j} & y_{2_j} & y_{3_j} & \cdots & y_{m_j} \end{bmatrix}^{\mathrm{T}} \,, \\ &\vdots \\ \mathbf{L}_n &= \begin{bmatrix} y_{1_n} & y_{2_n} & y_{3_n} & \cdots & y_{m_n} \end{bmatrix}^{\mathrm{T}} \,. \end{aligned} \tag{11}$$

Afterwards, these vectors are concatenated in one observation vector

$$\mathbf{L} = \begin{bmatrix} \mathbf{L}_1^{\mathrm{T}} & \cdots & \mathbf{L}_j^{\mathrm{T}} & \cdots & \mathbf{L}_n^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} . \tag{12}$$

The constraint equations

$$\begin{aligned} \hat{a}_{0_1} + \hat{a}_{1_1} x_{k_1} + \hat{a}_{2_1} x_{k_1}^2 + \hat{a}_{3_1} x_{k_1}^3 &= \hat{a}_{0_2} + \hat{a}_{1_2} x_{k_1} + \hat{a}_{2_2} x_{k_1}^2 + \hat{a}_{3_2} x_{k_1}^3 \,, \\ \hat{a}_{0_2} + \hat{a}_{1_2} x_{k_2} + \hat{a}_{2_2} x_{k_2}^2 + \hat{a}_{3_2} x_{k_2}^3 &= \hat{a}_{0_3} + \hat{a}_{1_3} x_{k_2} + \hat{a}_{2_3} x_{k_2}^2 + \hat{a}_{3_3} x_{k_2}^3 \,, \\ &\vdots \\ \hat{a}_{0_{n-1}} + \hat{a}_{1_{n-1}} x_{k_{n-1}} + \hat{a}_{2_{n-1}} x_{k_{n-1}}^2 + \hat{a}_{3_{n-1}} x_{k_{n-1}}^3 &= \hat{a}_{0_n} + \hat{a}_{1_n} x_{k_{n-1}} + \hat{a}_{2_n} x_{k_{n-1}}^2 + \hat{a}_{3_n} x_{k_{n-1}}^3 \end{aligned} \tag{13}$$

enforce the spline to be continuous at the knots. The next set of constraint equations

$$\begin{aligned} \hat{a}_{1_1} + 2\hat{a}_{2_1} x_{k_1} + 3\hat{a}_{3_1} x_{k_1}^2 &= \hat{a}_{1_2} + 2\hat{a}_{2_2} x_{k_1} + 3\hat{a}_{3_2} x_{k_1}^2 \,, \\ \hat{a}_{1_2} + 2\hat{a}_{2_2} x_{k_2} + 3\hat{a}_{3_2} x_{k_2}^2 &= \hat{a}_{1_3} + 2\hat{a}_{2_3} x_{k_2} + 3\hat{a}_{3_3} x_{k_2}^2 \,, \\ &\vdots \\ \hat{a}_{1_{n-1}} + 2\hat{a}_{2_{n-1}} x_{k_{n-1}} + 3\hat{a}_{3_{n-1}} x_{k_{n-1}}^2 &= \hat{a}_{1_n} + 2\hat{a}_{2_n} x_{k_{n-1}} + 3\hat{a}_{3_n} x_{k_{n-1}}^2 \end{aligned} \tag{14}$$

enforces the spline to have the first derivative at the knots. This implies that the spline will have smooth transition at the knots. The last set of constraint equations

$$\begin{aligned} 2\hat{a}_{2_1} + 6\hat{a}_{3_1} x_{k_1} &= 2\hat{a}_{2_2} + 6\hat{a}_{3_2} x_{k_1} \,, \\ 2\hat{a}_{2_2} + 6\hat{a}_{3_2} x_{k_2} &= 2\hat{a}_{2_3} + 6\hat{a}_{3_3} x_{k_2} \,, \\ &\vdots \\ 2\hat{a}_{2_{n-1}} + 6\hat{a}_{3_{n-1}} x_{k_{n-1}} &= 2\hat{a}_{2_n} + 6\hat{a}_{3_n} x_{k_{n-1}} \end{aligned} \tag{15}$$

enforces the spline to have a second derivative at the knots, which implies that the curvature at the knots will be continuous. Rearranging the equations (13) - (15) yields for continuity

$$\hat{a}_{0_j} + \hat{a}_{1_j} x_{k_j} + \hat{a}_{2_j} x_{k_j}^2 + \hat{a}_{3_j} x_{k_j}^3 - \hat{a}_{0_{j+1}} - \hat{a}_{1_{j+1}} x_{k_j} - \hat{a}_{2_{j+1}} x_{k_j}^2 - \hat{a}_{3_{j+1}} x_{k_j}^3 = 0\,, \tag{16}$$

for smoothness it is

$$\hat{a}_{1_j} + 2\hat{a}_{2_j} x_{k_j} + 3\hat{a}_{3_j} x_{k_j}^2 - \hat{a}_{1_{j+1}} - 2\hat{a}_{2_{j+1}} x_{k_j} - 3\hat{a}_{3_{j+1}} x_{k_j}^2 = 0\,, \tag{17}$$

and for the curvature we obtain

$$2\hat{a}_{2_j} + 6\hat{a}_{3_j} x_{k_j} - 2\hat{a}_{2_{j+1}} - 6\hat{a}_{3_{j+1}} x_{k_j} = 0\,. \tag{18}$$

The vector of unknown parameters for each cubic polynomial in the spline can be formulated separately for each segment as

$$\begin{aligned} \hat{\mathbf{X}}_1 &= \begin{bmatrix} \hat{a}_{0_1} & \hat{a}_{1_1} & \hat{a}_{2_1} & \hat{a}_{3_1} \end{bmatrix}^{\mathrm{T}}, \\ &\vdots \\ \hat{\mathbf{X}}_j &= \begin{bmatrix} \hat{a}_{0_j} & \hat{a}_{1_j} & \hat{a}_{2_j} & \hat{a}_{3_j} \end{bmatrix}^{\mathrm{T}}, \\ &\vdots \\ \hat{\mathbf{X}}_n &= \begin{bmatrix} \hat{a}_{0_n} & \hat{a}_{1_n} & \hat{a}_{2_n} & \hat{a}_{3_n} \end{bmatrix}^{\mathrm{T}}, \end{aligned} \tag{19}$$

and afterwards the entire vector of unknowns can be written as

$$\hat{\mathbf{X}} = \begin{bmatrix} \hat{\mathbf{X}}_1^{\mathrm{T}} & \cdots & \hat{\mathbf{X}}_j^{\mathrm{T}} & \cdots & \hat{\mathbf{X}}_n^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}. \tag{20}$$

From the stochastic model

$$\mathbf{\Sigma}_{LL} = \sigma_0^2 \mathbf{Q}_{LL}\,, \tag{21}$$

with σ_0^2 as theoretical variance factor the corresponding weight matrix is obtained from

$$\mathbf{P} = \mathbf{Q}_{LL}^{-1}\,, \tag{22}$$

supposing the cofactor matrix to be non-singular. Having a look at the observation equations (10) it is obvious that this spline approximation problem is linear and hence can easily be written in matrix notation

$$\mathbf{L} + \mathbf{v} = \mathbf{A}\hat{\mathbf{X}}\,, \tag{23}$$

where **A** is the design matrix that contains the coefficients of the unknowns. Equation (23) together with (21) is denoted Gauss-Markov model, see e.g. the textbook by Niemeier [2008, p. 137]. Considering the sequence of the unknowns in (19) and (20) the design matrix for the first interval reads

$$\mathbf{A}_1 = \begin{bmatrix} 1 & x_{1_1} & x_{1_1}^2 & x_{1_1}^3 \\ 1 & x_{2_1} & x_{2_1}^2 & x_{2_1}^3 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{m_1} & x_{m_1}^2 & x_{m_1}^3 \end{bmatrix}, \tag{24}$$

for the second interval we obtain

$$\mathbf{A}_2 = \begin{bmatrix} 1 & x_{1_2} & x_{1_2}^2 & x_{1_2}^3 \\ 1 & x_{2_2} & x_{2_2}^2 & x_{2_2}^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{m_2} & x_{m_2}^2 & x_{m_2}^3 \end{bmatrix}. \tag{25}$$

This construction of $\mathbf{A}_j$ matrices continues until the end of the last interval. Finally the entire design matrix has the form

$$\mathbf{A} = diag\begin{bmatrix}\mathbf{A}_1 & \cdots & \mathbf{A}_n\end{bmatrix}. \tag{26}$$

Having a look at the constraint equations (16) - (18) it is obvious that they are also linear and hence can easily be written in matrix notation

$$\mathbf{C}_q\hat{\mathbf{X}} = \mathbf{0}, \tag{27}$$

with q =1, 2, 3, where the matrices $\mathbf{C}_q$ contain the coefficients of the unknowns. Every row of these matrices refers only to two consecutive intervals, i.e.:

- Unknowns from the first row refer to the first two intervals, $[x_{k_0}, x_{k_1})$ and $[x_{k_1}, x_{k_2})$.
- Unknowns from the second row refer to the second and the third interval, $[x_{k_1}, x_{k_2})$ and $[x_{k_2}, x_{k_3})$.
- This continues until the last row where the unknowns refer to the last and its preceding interval, $[x_{k_{n-2}}, x_{k_{n-1}})$ and $[x_{k_{n-1}}, x_{k_n})$.

Considering the sequence of the unknowns in (19) and (20) the matrix $\mathbf{C}_1$ that imposes continuity at the knots is

$$\mathbf{C}_1 = \begin{bmatrix} 1 & x_{k_1} & x_{k_1}^2 & x_{k_1}^3 & -1 & -x_{k_1} & -x_{k_1}^2 & -x_{k_1}^3 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 1 & x_{k_2} & x_{k_2}^2 & x_{k_2}^3 & -1 & -x_{k_2} & -x_{k_2}^2 & -x_{k_2}^3 & \cdots & 0 \\ \vdots & & & & & & \vdots & & & & & & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & -x_{k_{n-1}}^3 \end{bmatrix}. \tag{28}$$

Matrix $\mathbf{C}_2$ contains the coefficients from the equations of constraints that preserve the smoothness of the spline at the knots and is expressed as

$$\mathbf{C}_2 = \begin{bmatrix} 0 & 1 & 2x_{k_1} & 3x_{k_1}^2 & 0 & -1 & -2x_{k_1} & -3x_{k_1}^2 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 2x_{k_2} & 3x_{k_2}^2 & 0 & -1 & -2x_{k_2} & -3x_{k_2}^2 & \cdots & 0 \\ \vdots & & & & & & \vdots & & & & & & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & -3x_{k_{n-1}}^2 \end{bmatrix}. \tag{29}$$

Matrix $\mathbf{C}_3$ contains the coefficients from the equations of constraints which enforce continuous curvature at the knots written as

$$\mathbf{C}_3 = \begin{bmatrix} 0 & 0 & 2 & 6x_{k_1} & 0 & 0 & -2 & -6x_{k_1} & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 6x_{k_2} & 0 & 0 & -2 & -6x_{k_2} & \cdots & 0 \\ \vdots & & & & & & \vdots & & & & & & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & -6x_{k_{n-1}} \end{bmatrix}. \tag{30}$$

3.3 Least-squares adjustment

We are facing a linear least-squares adjustment problem having a system of observation equations along with equations of constraints. For solution of such a system when constrained optimization is applied, the Lagrange multiplier method can be used. Therefore, we construct the Lagrange function and minimize it. First, the objective function is formulated as

$$\Omega(\hat{\mathbf{X}}) = \mathbf{v}^{\mathrm{T}}\mathbf{P}\mathbf{v} = (\mathbf{A}\hat{\mathbf{X}} - \mathbf{L})^{\mathrm{T}}\mathbf{P}(\mathbf{A}\hat{\mathbf{X}} - \mathbf{L}) \tag{31}$$

with the residuals

$$\mathbf{v} = \mathbf{A}\hat{\mathbf{X}} - \mathbf{L} \tag{32}$$

from (23). Afterwards, the constraint equations $\mathbf{C}_q\hat{\mathbf{X}} = \mathbf{0}$ are multiplied by the Lagrange multipliers $2\lambda_q$ for q =1, 2, 3 and we obtain

$$2\lambda_q\mathbf{C}_q\hat{\mathbf{X}} = \mathbf{0}. \tag{33}$$

The number 2 in $2\lambda_q$ does not change anything; it only makes the further minimization process easier. Furthermore, the Lagrange function is the sum of the objective function (31) and the function on the left side of constraint equations (33), and we obtain

$$\Phi(\hat{\mathbf{X}}, \lambda_1, \lambda_2, \lambda_3) = \mathbf{v}^{\mathrm{T}}\mathbf{P}\mathbf{v} + 2\lambda_1\mathbf{C}_1\hat{\mathbf{X}} + 2\lambda_2\mathbf{C}_2\hat{\mathbf{X}} + 2\lambda_3\mathbf{C}_3\hat{\mathbf{X}}. \tag{34}$$

For minimization of the objective function which is additionally constrained, we minimize the Lagrange function by taking the partial derivatives of (34) w.r.t. the unknowns $\hat{\mathbf{X}}$ and λ_q. Afterwards, every derivative is set to be equal to zero and we obtain

$$\begin{aligned} \mathbf{A}^{\mathrm{T}}\mathbf{P}\mathbf{A}\hat{\mathbf{X}} + \mathbf{C}_1{}^{\mathrm{T}}\lambda_1 + \mathbf{C}_2{}^{\mathrm{T}}\lambda_2 + \mathbf{C}_3{}^{\mathrm{T}}\lambda_3 &= \mathbf{A}^{\mathrm{T}}\mathbf{P}\mathbf{L}, \\ \mathbf{C}_1\hat{\mathbf{X}} &= \mathbf{0}, \\ \mathbf{C}_2\hat{\mathbf{X}} &= \mathbf{0}, \\ \mathbf{C}_3\hat{\mathbf{X}} &= \mathbf{0}. \end{aligned} \tag{35}$$

This system of matrix equations can be written in a matrix form

$$\begin{bmatrix} \mathbf{A}^{\mathrm{T}}\mathbf{P}\mathbf{A} & \mathbf{C}_1{}^{\mathrm{T}} & \mathbf{C}_2{}^{\mathrm{T}} & \mathbf{C}_3{}^{\mathrm{T}} \\ \mathbf{C}_1 & & & \\ \mathbf{C}_2 & & \mathbf{0} & \\ \mathbf{C}_3 & & & \end{bmatrix} \begin{bmatrix} \hat{\mathbf{X}} \\ \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} = \begin{bmatrix} \mathbf{A}^{\mathrm{T}}\mathbf{P}\mathbf{L} \\ \\ \mathbf{0} \\ \\ \end{bmatrix}, \tag{36}$$

and the solution for the unknowns can be derived by using matrix inversion. However, matrix inversion is prone to rounding errors and even the use of double-precision floating-point format for computation is in many cases not sufficient. An error analysis in matrix computations and alternative direct and iterative solutions for linear systems can be found e.g. in the comprehensive textbooks by Golub and Van Loan [1996] or Björck [2015].

After the solution is obtained, we can compute the residuals from (32) and the a posteriori estimate of the reference standard deviation from

$$s_0 = \sqrt{\frac{\mathbf{v}^{\mathrm{T}}\mathbf{P}\mathbf{v}}{r}}, \tag{37}$$

where r is the redundancy of the adjustment problem. For quality assessment of the adjustment results the cofactor matrix $\mathbf{Q}_{xx}$ for precision measures and $\mathbf{Q}_{vv}$ for reliability measures can be computed with the formulas that can be found e.g. in the textbook by Niemeier [2008]. The value s_0 can be applied in knot placement strategies described in the next section.

3.4 Knot placement strategies

The distribution of the knots plays a crucial role in spline approximation problems and can be considered as a separate scientific topic. The basic problem comprises two fundamental challenges:

1. systematic or heuristic choice of the knots,
2. selection of suitable quality criteria according to which the resulting spline approximation can be assessed.

According to these specifications, the nodes are rearranged in an iterative process until the quality criteria are met.

As an example for a knot placement strategy let us take the one presented by Gálvez et al. (2015) where a sophisticated evolutionary approach is applied for the choice of the knots. As quality criteria they applied the root mean square error (RMSE) which is equivalent to the reference standard deviation (37). Noting that the best RMSE might be obtained at the expense of a large number of variables, possibly leading to overfitting, Gálvez et al. (2015) also computed AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) functions to provide an adequate trade-off between the quality of approximation and the complexity of the model.

In this article we do not focus on the development of new knot placement strategies, so a simple approach in form of a uniformly distributed knot sequence is chosen, proposed by De Boor [1986], also called equidistant arrangement by Yanagihara and Ohtaki [2003]. The term uniformly distributed knot sequence describes that all knots have equal spacings within the dataset $[x_{\min}, x_{\max}]$ to which the spline is fitted.

One possible way of assessing the resulting spline approximation is to use the well-known "global test" to check the functional and the stochastic models of the adjustment, see e.g. the textbook by Niemeier [2008, p. 167]. Alternatively the term "overall model test" is used, e.g. by Teunissen [2000, p. 93]. Ghilani [2010, p. 315, p. 528 ff] uses the term "goodness-of-fit test". A hypothesis is to be tested, whether s_0^2 coincides with the a priori value σ_0^2 or not. For testing such a hypothesis, the χ^2-test is applied. After formulating the null hypothesis

$$H_0: \qquad E\{s_0^2\} = \sigma_0^2 \tag{38}$$

and the corresponding alternative hypothesis

$$H_A: \qquad E\{s_0^2\} \neq \sigma_0^2 \tag{39}$$

a test statistic can be computed and compared with the quantile of the χ^2-distribution $\chi^2_{f,1-\alpha}$, where f represents the degree of freedom that coincides with the redundancy r and α represents the probability of error, e.g. $\alpha = 5\%$. If the test decision is that we fail to reject the null hypothesis, the functional and the stochastic models of the adjustment are regarded as appropriate. If the test decision is that we have to reject the null hypothesis H_0 in favor of the alternative H_A, two cases can occur:

- $s_0^2 < \sigma_0^2$: The stochastic model was chosen too pessimistic or too many knots were introduced to represent the spline that led to an overfitting.
- $s_0^2 > \sigma_0^2$: The stochastic model was chosen too optimistic or outliers are present. If the stochastic model is appropriate and no outliers are present, this case can occur if the knots are not chosen in an appropriate way. This problem will be discussed in the following.

The initial assumption is that we have no previous knowledge about the distribution of the observed points. Hence, we cannot immediately specify the number of knots and their positions for obtaining an optimal result. With the help of the overall model test the knot placement can be performed as follows:

- A least-squares adjustment is performed with three knots, thus one knot is in the middle of the dataset, while the first and the last knots are at the beginning and the end of the interval.
- Afterwards the overall model test is performed.
- If we fail to reject the null hypothesis (38), the final result is obtained.
- If the null hypothesis is rejected, a new adjustment is performed with four uniformly distributed knots.
- This increasing of the number of knots continues until we fail to reject the null hypothesis (38).

Figure 4 shows a flowchart of this iterative procedure.

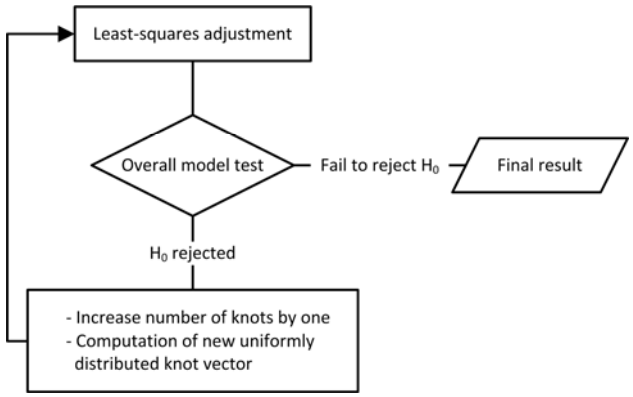

Figure 4: Flowchart of the iterative procedure for the spline approximation

3.5 A posteriori analysis

After a solution is obtained, the well-known tools for an a posteriori analysis of adjustment results can be applied. From the wide range of analysis tools we select some of them for significance testing of parameters, assessment of reliability, and a more in-depth analysis of the goodness-of-fit.

As proposed by Buse and Lim [1977] and shown in a numerical example, the *t*-test for the significance of an individual coefficient allows testing the hypothesis $a_{3_j} = 0$ to determine whether the cubic term in the j^{th} interval is significant. In the same contribution it is shown how "structural changes" in the data can be detected as a jump discontinuity in the third derivative at a knot x_{k_j}, again with the help of the *t*-test. Furthermore it is shown how the *F*-test can be applied to test the validity of the imposed constraints.

In addition to the investigations proposed by Buse and Lim [1977], further analyses can be carried out. As a measure of reliability, the partial redundancy numbers

$$r_i = (q_{vv})_{ii} \cdot p_i \tag{40}$$

can be computed for each observation, where $(q_{vv})_{ii}$ is the i^{th} diagonal element of the cofactor matrix of the residuals $\mathbf{Q}_{vv}$ and p_i the i^{th} diagonal element of the weight matrix of the observations $\mathbf{P}$. Low redundancy numbers, e.g. $r_i < 30\%$, indicate observations that lack sufficient checks to detect blunders, and thus the chance that undetected blunders affect the shape of the spline is high. Using the partial redundancy numbers it is furthermore possible, to compute individual variance components for each interval *j*. From the residual vector $\mathbf{v}_j$, the weight matrix $\mathbf{P}_j$ and the redundancy

$$r_{g_j} = \sum_{i=1}^{m_j} r_i \tag{41}$$

for the group of observations in an interval *j*, the variance component

$$s_{0_j}^2 = \frac{\mathbf{v}_j^{\mathrm{T}} \mathbf{P}_j \mathbf{v}_j}{r_{g_j}} \tag{42}$$

can be computed for the corresponding interval. The square root of this value provides a reference variance for each interval, which can help to check whether the knot placement has yielded a meaningful result for all intervals.

For the analysis of systematic deviations between the given data and the adjusted spline, an investigation of a residual plot for each interval is recommended. Furthermore the following test, based on the explanations of Helmert [1924, p. 334 ff], can be applied. Considering the residuals for the observations in an interval *j*, the expected value for the sum of +/- signs for purely random deviations is zero with a standard deviation of $\sqrt{m_j}$, where m_j is the number of observations in the corresponding interval. If the absolute value of the sum of +/- signs is greater than $\sqrt{m_j}$, there is a first indication that systematic deviations can be assumed. This test can also help to check whether the knot placement has yielded a meaningful result for all intervals.

3.6 Discussion

The spline approximation presented in section 3 is general and the most intuitive. The functional model is simple and the constraints are explicitly defined. The approximation is applied using the constrained L_2 norm minimization. Hypothesis testing is applied to check the mathematical model in combination with a variation of the knots. A uniformly distributed knot sequence is maybe not the optimal solution for a spline approximation; however, it is the simplest one to be applied. Other knot placement and knot removal strategies are using the notions of fixed and free knots, and some of them also include the concept of regularization which substantially complicates the adjustment. The knot placement strategies are a separate scientific topic and many authors, e.g. Cox et al. [1989], Schwetlick and Schütze [1995], Park [2011] have addressed this problem. An overview of strategies to determine the knot vector is offered by Bureick et al. [2016a] and [2016b]. Current research by Gálvez et al. (2015) or Harmening and Neuner [2016] focusses on choosing the optimal number of B-spline control points using the Akaike Information Criterion and the Bayesian Information Criterion. Bureick et al. [2016b] developed an algorithm which determines the knot vector of a B-Spline with a mixture of Monte-Carlo methods and an evolutionary algorithm.

4 Spline function constructed from truncated polynomials

As in the previous section, spline functions in the form $y = f(x)$ are used, but in contrast to an explicit formulation of the constraints for continuity, smoothness and curvature, a functional model is used where these constraints are taken into account implicitly. A functional model of this kind is the model that uses truncated polynomials. This type of spline function is created in such a way that the constraint is not given explicitly as an equation, but is "hidden" in the function. This approach was used by Smith [1979] for statistical data smoothing.

The idea behind the truncated polynomials is quite simple, even though in the literature, e.g. in the technical report by De Boor and Rice [1968], it is not always described in a comprehensive way. On the contrary, Fuller [1969] explains the idea in a nice and easily understandable way. In place of "truncated polynomials" he calls them "grafted polynomials". The considered class of functions is more general than splines and includes them as a special case. The polynomials for different intervals might be of different degrees and not all of the usual constraints have to be imposed in knots. Here we only consider the special case of cubic splines. The function can be easily derived. In the first step, the spline function with *n* segments can be written as

$$\begin{aligned} y = {} & b_{0_1} + b_{1_1} x + b_{2_1} x^2 + b_{3_1} x^3 + \\ & + b_{1_2} (x - x_{k_1}) + b_{2_2} (x - x_{k_1})^2 + b_{3_2} (x - x_{k_1})^3 + \dots \\ & + b_{1_j} (x - x_{k_j}) + b_{2_j} (x - x_{k_j})^2 + b_{3_j} (x - x_{k_j})^3 + \dots \\ & + b_{1_n} (x - x_{k_n}) + b_{2_n} (x - x_{k_n})^2 + b_{3_n} (x - x_{k_n})^3 , \end{aligned} \tag{43}$$

with $b_{1_j} = b_{2_j} = b_{3_j} = 0$ for $x < x_{k_j}$. In the next step, in each polynomial after the first interval only the last part, $b_{3_j} (x_i - x_{k_j})^3$, is taken into consideration. This part is in the literature called the "truncated part" of the polynomial for the given interval. This, at first sight strange truncation is nicely explained by Fuller [1969]. Each new interval brings a new cubic polynomial, i.e. 4 additional parameters. Imposing 3 conditions (7), (8) and (9) reduces the number of new additional parameters to one. That just b_{3_j} remains is explained by Smith [1979]. Hence, the whole expression can be rewritten as

$$\begin{aligned} y = {} & b_{0_1} + b_{1_1} x + b_{2_1} x^2 + b_{3_1} x^3 + b_{3_2} (x - x_{k_1})^3 + \dots \\ & + b_{3_j} (x - x_{k_j})^3 + \dots + b_{3_n} (x - x_{k_n})^3 , \end{aligned} \tag{44}$$

with $b_{3_j} = 0$ for $x < x_{k_j}$. However, this notation appears to be unorganized, because b_{3_j} repeats in each interval. For solving this notation issue, the unknowns for the first interval are changed from $b_{0_1}, b_{1_1}, b_{2_1}, b_{3_1}$ to a_0, a_1, a_2, a_3, and the index for the unknown b_{3_j} from the truncated part of the polynomials is changed into b_j. Finally, the function is given as

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \sum_{j=2}^{n} b_j (x - x_{k_j})_+^3 , \tag{45}$$

where the small "plus" (+), after the brackets from $b_j (x - x_{k_j})_+^3$, represents an additional indicator for the truncated part which is accepted as a standard notation in the literature [Smith 1979]. This truncated part is defined as

$$(x - x_{k_j})_+ = \begin{cases} (x - x_{k_j})_+ & \text{if} \quad x \geq x_{k_j} \\ 0 & \text{if} \quad x < x_{k_j} \end{cases} . \tag{46}$$

This additional restriction indicates that the full polynomial, along with the truncated parts, can be developed up to the last interval $[x_{k_{n-1}}, x_{k_n})$, see Figure 2.

4.1 Definition of the problem

Let:

- $y_{1_j}, y_{2_j}, \dots, y_{i_j}, \dots, y_{m_j}$ be a sequence of observed values within an interval *j*, with m_j being the number of observations within the interval *j*,
- $\boldsymbol{\Sigma}_{LL} = \sigma_0^2 \mathbf{Q}_{LL}$ be the variance-covariance matrix of all observations y_{i_j},
- $x_{1_j} < x_{2_j} < \dots < x_{i_j} < \dots < x_{m_j}$ be a non-decreasing sequence of error-free values referring to the observations within an interval *j*, and
- $x_{k_1} < x_{k_2} < \dots < x_{k_j} < \dots < x_{k_{n-1}}$ be a non-decreasing sequence of user-defined values for the knots on the *x*-axis which are error-free. Their placement is already described in section 3.1, as well as in section 3.4. Also in this functional model, the first knot x_{k_0} and the last knot x_{k_n} are not taken into consideration.

To determine:

- $a_0, a_1, a_2, a_3, b_2, \ldots, b_j, \ldots, b_n$ which represent a set of parameters for a cubic polynomial that has to be estimated w.r.t. each interval $[x_{k_{j-1}}, x_{k_j})$.

According to (45) and (46) the functional model for the truncated polynomials reads

$$y_i = a_0 + a_1 x_i + a_2 x_i^2 + a_3 x_i^3 + \sum_{j=2}^{n} b_j (x_i - x_{k_j})_+^3 \,, \tag{47}$$

and

$$(x_i - x_{k_j})_+ = \begin{cases} (x_i - x_{k_j})_+ & \text{if} \quad x_i \geq x_{k_j} \\ 0 & \text{if} \quad x_{i_j} < x_{k_j} \end{cases}, \tag{48}$$

with i as an index for all pairs of observed and fixed values within an interval. In the first interval $[x_{k_0}, x_{k_1})$ index i ranges from $i = 1, 2, ..., m_1$, in the subsequent intervals $[x_{k_{j-1}}, x_{k_j})$, with $j = 2, 3, ..., n$, index i ranges from $i = 1, 2, ..., m_j$.

Having defined all these quantities, we can create a spline approximation from an overdetermined configuration using the Gauss-Markov model.

4.2 Formulation of the adjustment problem

Based on the functional model (47) together with (48) and considering y_{i_j} as observations and x_{i_j} as fixed values the observation equations

$$\begin{aligned}
y_{i_1} + v_{i_1} &= \hat{a}_0 + \hat{a}_1 x_{i_1} + \hat{a}_2 x_{i_1}^2 + \hat{a}_3 x_{i_1}^3, && i = 1, 2, \ldots, m_1 \,, \\
y_{i_2} + v_{i_2} &= \hat{a}_0 + \hat{a}_1 x_{i_2} + \hat{a}_2 x_{i_2}^2 + \hat{a}_3 x_{i_2}^3 + \hat{b}_2 (x_{i_2} - x_{k_1})_+^3, && i = 1, 2, \ldots, m_2 \,, \\
y_{i_3} + v_{i_3} &= \hat{a}_0 + \hat{a}_1 x_{i_3} + \hat{a}_2 x_{i_3}^2 + \hat{a}_3 x_{i_3}^3 + \hat{b}_2 (x_{i_3} - x_{k_1})_+^3 + \hat{b}_3 (x_{i_3} - x_{k_2})_+^3, && i = 1, 2, \ldots, m_3 \,, \\
&\vdots \\
y_{i_n} + v_{i_n} &= \hat{a}_0 + \hat{a}_1 x_{i_n} + \hat{a}_2 x_{i_n}^2 + \hat{a}_3 x_{i_n}^3 + \hat{b}_2 (x_{i_n} - x_{k_1})_+^3 + \hat{b}_3 (x_{i_n} - x_{k_2})_+^3 + \cdots + \hat{b}_n (x_{i_n} - x_{k_{n-1}})_+^3, && i = 1, 2, \ldots, m_n \,,
\end{aligned} \tag{49}$$

that are seemingly different for each cubic polynomial within the spline, can be set up. If we consider an individual set of observations for each segment of the spline, the observation vector can be set up in the same way as in (11) and (12). The vector of unknowns simply contains all unknown parameters from the last set of observation equations

$$\hat{\mathbf{X}} = \begin{bmatrix} \hat{a}_0 & \hat{a}_1 & \hat{a}_2 & \hat{a}_3 & \hat{b}_2 & \hat{b}_3 & \cdots & \hat{b}_j & \cdots & \hat{b}_n \end{bmatrix}^{\mathrm{T}}. \tag{50}$$

The stochastic model is derived from (21) and (22). Since the problem is linear the design matrix contains the coefficients of the unknowns from the observation equations. Thus we obtain

$$\mathbf{A} = \left[\begin{array}{cccccccc}
1 & x_{1_1} & x_{1_1}^2 & x_{1_1}^3 & 0 & 0 & \cdots & 0 \\
1 & x_{2_1} & x_{2_1}^2 & x_{2_1}^3 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
1 & x_{m_1} & x_{m_1}^2 & x_{m_1}^3 & 0 & 0 & \cdots & 0 \\
\hline
1 & x_{1_2} & x_{1_2}^2 & x_{1_2}^3 & (x_{1_2} - x_{k_1})_+^3 & 0 & \cdots & 0 \\
1 & x_{2_2} & x_{2_2}^2 & x_{2_2}^3 & (x_{2_2} - x_{k_1})_+^3 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
1 & x_{m_2} & x_{m_2}^2 & x_{m_2}^3 & (x_{m_2} - x_{k_1})_+^3 & 0 & \cdots & 0 \\
\hline
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\hline
1 & x_{1_n} & x_{1_n}^2 & x_{1_n}^3 & (x_{1_n} - x_{k_1})_+^3 & (x_{1_n} - x_{k_2})_+^3 & \cdots & (x_{1_n} - x_{k_{n-1}})_+^3 \\
1 & x_{2_n} & x_{2_n}^2 & x_{2_n}^3 & (x_{2_n} - x_{k_1})_+^3 & (x_{2_n} - x_{k_2})_+^3 & \cdots & (x_{2_n} - x_{k_{n-1}})_+^3 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
1 & x_{m_n} & x_{m_n}^2 & x_{m_n}^3 & (x_{m_n} - x_{k_1})_+^3 & (x_{m_n} - x_{k_2})_+^3 & \cdots & (x_{m_n} - x_{k_{n-1}})_+^3
\end{array}\right]. \tag{51}$$

4.3 Least-squares adjustment

Since the constraints are not given explicitly as equations, but "hidden" in the functional model, we don't have to introduce constraint equations into the minimization. Thus the least-squares solution is obtained from

$$\hat{\mathbf{X}} = (\mathbf{A}^{\mathrm{T}}\mathbf{P}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{P}\mathbf{L}\,. \tag{52}$$

The residuals and the reference standard deviation are computed from (32) respectively (37). The knots can be selected using the strategy described in section 3.4.

4.4 Discussion

The spline approximation using truncated polynomials represents an equivalent formulation of the spline approximation using ordinary polynomials and explicitly formulated constraints in section 3. This spline function is more convenient for handling and, depending on the adjustment problem, it can be interpreted in a form of condition equations or as a parametric spline curve. In many publications this approach is used for statistical analysis. However, De Boor and Rice [1968] are stating that this approach is ill conditioned and they propose other approaches as numerically more stable. An investigation of the numerical stability of the approaches presented in this article will be given in the forthcoming Part 3 of a series of three articles.

5 Spline curve constructed from ordinary cubic polynomials

In sections 3 and 4 spline functions in the form $y = f(x)$ are used for which Bronshtein and Semendyayev [2007, p. 47] give the following definition: "If x and y are two variable quantities, and if there is a rule which assigns a unique value of y to a given value of x, then we call y a function of x, and we use the notation $y = f(x)$." Consequently, all other relations that do not have this property cannot be considered as functions. In order to be able to display arbitrary curves, a parametric representation in the form $x = f_1(t)$ and $y = f_2(t)$ is now applied in this section. In contrast to a spline function of the form $y = f(x)$, x and y are now considered independently of each other which enables the representation of arbitrary curves.

In this article the term "curve" refers to a graph of a spline curve that is represented as a set of parametric equations, and is not regarded as a function. The spline curve approximation is quite useful for a dataset that "moves" in every possible direction, in a two, three or multidimensional space. This section considers curves in 2D. Examples are open but self-intersecting curves, see Figure 5, closed curves with irregular shape, or as the simplest form, a circle.
In sections 3 and 4 the values y_i are introduced as observations and the corresponding x_i as fixed values. Now, using a spline curve, we introduce both y_i and x_i as two sequences of observed values that can be expressed as functions of some error-free parameter t. In practice this form can be used for an approximation of a one-dimensional substructure or superstructure (pipelines, railways, roller coasters etc.) in two or three-dimensional space.

In the literature most of the approaches, regarding the spline approximation, are actually using spline curves; mainly because most of the literature is from the field of computer graphics, see e.g. the textbook by Piegl and Tiller [2012]. In those publications B-spline curves are used as a functional model, see e.g. the articles by Wang et al. [2006], Zheng et al. [2012] and the textbook by Piegl and Tiller [2012, pp. 410-419]. The spline curve approximation is quite handy. However, the interpretation of the residuals is not intuitive, because y_i and x_i are introduced as two separate sequences of observations and hence the residuals v_y and v_x vary independently from each other.

The approach that is proposed here is very similar to the one in section 3. The difference is that both y_i and x_i are represented as functions of t. Having this property, the further implication is that these two functions $x(t)$ and $y(t)$ do not have a common unknown which implies that these two functions are treated in two separate systems of equations.

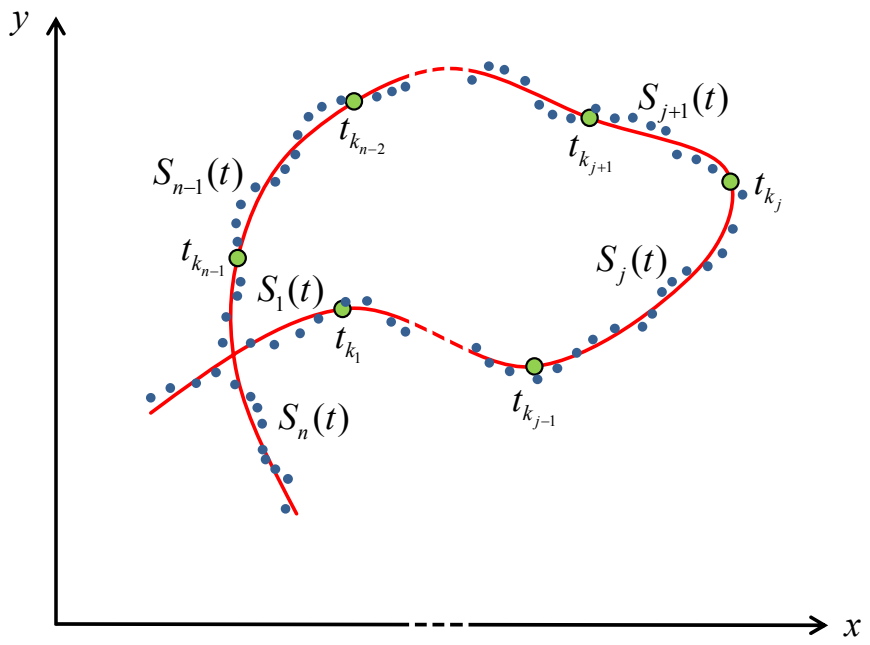

Figure 5: Spline curve approximation

5.1 Definition of the problem

Let:

- $x_{1_j}, x_{2_j}, \ldots, x_{i_j}, \ldots x_{m_j}$ and $y_{1_j}, y_{2_j}, \ldots, y_{i_j}, \ldots, y_{m_j}$ be two sequences of observed values within an interval j, with m_j being the number of observations x_{i_j} respectively y_{i_j} within the interval j,
- $\mathbf{\Sigma}_{LL} = \sigma_0^2 \mathbf{Q}_{LL}$ be the variance-covariance matrix of all observations x_{i_j}, y_{i_j},
- $t_{1_j} < t_{2_j} < \ldots < t_{i_j} < \ldots t_{m_j}$ be a non-decreasing sequence of error-free values referring to the observations within an interval j that is usually interpreted as time. These values can be freely chosen. However, for interpretable results they should be proportional to the distances between the observations. In this case the Euclidean distances between the observed points are chosen to represent these values,
- $t_{k_1} < t_{k_2} < \ldots < t_{k_j} < \ldots t_{k_{n-1}}$ be a non-decreasing sequence of user-defined values for the knots, which are regarded as error-free. Again, character k stands for "knot" and is not an index; hence t_{k_j} means "t-value of the knot j". As for the spline function in section 3, the first knot t_{k_0} and the last knot t_{k_n} are not taken into account, see Figure 5.

To determine:

- a set of parameters $(a_{0_1}, a_{1_1}, a_{2_1}, a_{3_1}), \ldots, (a_{0_j}, a_{1_j}, a_{2_j}, a_{3_j}), \ldots, (a_{0_n}, a_{1_n}, a_{2_n}, a_{3_n})$ of the cubic polynomials $S_{x_j}(t)$ that has to be estimated for all intervals $[t_{k_{j-1}}, t_{k_j})$.
- a different set of parameters $(b_{0_1}, b_{1_1}, b_{2_1}, b_{3_1}), \ldots, (b_{0_j}, b_{1_j}, b_{2_j}, b_{3_j}), \ldots, (b_{0_n}, b_{1_n}, b_{2_n}, b_{3_n})$ of the cubic polynomials $S_{y_j}(t)$ that has to be estimated for all intervals same as for $S_{x_j}(t)$.

The functional model is similar to (6), just with t as parameter, yielding in the considered case two separate functional models

$$x_{i_j} = S_{x_j}(t_{i_j}) = a_{0_j} + a_{1_j} t_{i_j} + a_{2_j} t_{i_j}^2 + a_{3_j} t_{i_j}^3, \tag{53}$$

$$y_{i_j} = S_{y_j}(t_{i_j}) = b_{0_j} + b_{1_j} t_{i_j} + b_{2_j} t_{i_j}^2 + b_{3_j} t_{i_j}^3, \tag{54}$$

with j = 1, 2, …, n as index for an interval $[t_{k_{j-1}}, t_{k_j})$. Furthermore, i = 1, 2, …, m_j represents an index for all pairs of observed and fixed values within an interval.

Since we consider an overdetermined system of equations, a spline approximation is performed without imposing any extra conditions on the first and the last knot of the spline. The dataset is subdivided in intervals w.r.t. the parameter *t*. For each interval $[t_{k_{j-1}}, t_{k_j})$ there exists a set of minimum two (x_{i_j}, t_{i_j}) pairs of values for the *x*-axis, and two (y_{i_j}, t_{i_j}) pairs of values for the *y*-axis. For constructing a spline curve, the following conditions, according to (7) - (9), are imposed at the knots:

for *x*

$$S_{x_j}(t_{k_j}) = S_{x_{j+1}}(t_{k_j}), \quad j = 1,\ldots,n-1, \tag{55}$$

$$S'_{x_j}(t_{k_j}) = S'_{x_{j+1}}(t_{k_j}), \quad j = 1,\ldots,n-1, \tag{56}$$

$$S''_{x_j}(t_{k_j}) = S''_{x_{j+1}}(t_{k_j}), \quad j = 1,\ldots,n-1, \tag{57}$$

for *y*

$$S_{y_j}(t_{k_j}) = S_{y_{j+1}}(t_{k_j}), \quad j = 1,\ldots,n-1, \tag{58}$$

$$S'_{y_j}(t_{k_j}) = S'_{y_{j+1}}(t_{k_j}), \quad j = 1,\ldots,n-1, \tag{59}$$

$$S''_{y_j}(t_{k_j}) = S''_{y_{j+1}}(t_{k_j}), \quad j = 1,\ldots,n-1. \tag{60}$$

Introduced knots t_{k_j} are freely specified or predefined by some knot placement strategy.

Having defined all these quantities, we can create a spline approximation from an overdetermined configuration using the Gauss-Markov model with additional constraints for the unknowns.

5.2 Formulation of the adjustment problem

The observation equations are similar to those in section 3.2. Here they refer to the parameter *t* for both coordinates. Their structuring depends on the interval $[t_{k_{j-1}}, t_{k_j})$ which the observations belong to with *j* = 1, 2, …, *n*. Based on the functional model (53) considering x_i as observations and t_i as error-free values the observation equations

$$\begin{aligned}
x_{i_1} + v_{x_{i_1}} &= \hat{a}_{0_1} + \hat{a}_{1_1} t_{i_1} + \hat{a}_{2_1} t_{i_1}^2 + \hat{a}_{3_1} t_{i_1}^3 \qquad i = 1,2,\ldots,m_1, \\
&\vdots \\
x_{i_j} + v_{x_{i_j}} &= \hat{a}_{0_j} + \hat{a}_{1_j} t_{i_j} + \hat{a}_{2_j} t_{i_j}^2 + \hat{a}_{3_j} t_{i_j}^3 \qquad i = 1,2,\ldots,m_j, \\
&\vdots \\
x_{i_n} + v_{x_{i_n}} &= \hat{a}_{0_n} + \hat{a}_{1_n} t_{i_n} + \hat{a}_{2_n} t_{i_n}^2 + \hat{a}_{3_n} t_{i_n}^3 \qquad i = 1,2,\ldots,m_n
\end{aligned} \tag{61}$$

can be set up. Based on (54) with y_i as observations and t_i as error-free values we obtain

$$\begin{aligned}
y_{i_1} + v_{y_{i_1}} &= \hat{b}_{0_1} + \hat{b}_{1_1} t_{i_1} + \hat{b}_{2_1} t_{i_1}^2 + \hat{b}_{3_1} t_{i_1}^3 \qquad i = 1,2,\ldots,m_1, \\
&\vdots \\
y_{i_j} + v_{y_{i_j}} &= \hat{b}_{0_j} + \hat{b}_{1_j} t_{i_j} + \hat{b}_{2_j} t_{i_j}^2 + \hat{b}_{3_j} t_{i_j}^3 \qquad i = 1,2,\ldots,m_j, \\
&\vdots \\
y_{i_n} + v_{y_{i_n}} &= \hat{b}_{0_n} + \hat{b}_{1_n} t_{i_n} + \hat{b}_{2_n} t_{i_n}^2 + \hat{b}_{3_n} t_{i_n}^3 \qquad i = 1,2,\ldots,m_n.
\end{aligned} \tag{62}$$

To construct the observation vectors, first all vectors for the *x*-observations for each segment of the spline are introduced as

$$\begin{aligned}
\mathbf{L}_{x_1} &= \begin{bmatrix} x_{1_1} & x_{2_1} & x_{3_1} & \cdots & x_{m_1} \end{bmatrix}^{\mathrm{T}}, \\
&\vdots \\
\mathbf{L}_{x_j} &= \begin{bmatrix} x_{1_j} & x_{2_j} & x_{3_j} & \cdots & x_{m_j} \end{bmatrix}^{\mathrm{T}}, \\
&\vdots \\
\mathbf{L}_{x_n} &= \begin{bmatrix} x_{1_n} & x_{2_n} & x_{3_n} & \cdots & x_{m_n} \end{bmatrix}^{\mathrm{T}},
\end{aligned} \tag{63}$$

afterwards, the vectors for the y-observations

$$\begin{aligned}
\mathbf{L}_{y_1} &= \begin{bmatrix} y_{1_1} & y_{2_1} & y_{3_1} & \cdots & y_{m_1} \end{bmatrix}^{\mathrm{T}}, \\
&\vdots \\
\mathbf{L}_{y_j} &= \begin{bmatrix} y_{1_j} & y_{2_j} & y_{3_j} & \cdots & y_{m_j} \end{bmatrix}^{\mathrm{T}}, \\
&\vdots \\
\mathbf{L}_{y_n} &= \begin{bmatrix} y_{1_n} & y_{2_n} & y_{3_n} & \cdots & y_{m_n} \end{bmatrix}^{\mathrm{T}}
\end{aligned} \tag{64}$$

are set up. Finally, these vectors are concatenated in two observation vectors

$$\mathbf{L}_x = \begin{bmatrix} \mathbf{L}_{x_1}^{\mathrm{T}} & \cdots & \mathbf{L}_{x_j}^{\mathrm{T}} & \cdots & \mathbf{L}_{x_n}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}, \tag{65}$$

$$\mathbf{L}_y = \begin{bmatrix} \mathbf{L}_{y_1}^{\mathrm{T}} & \cdots & \mathbf{L}_{y_j}^{\mathrm{T}} & \cdots & \mathbf{L}_{y_n}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}. \tag{66}$$

The constraint equations are similar to those in section 3.2. The difference is that here they refer to each sequence of observations x_i respectively y_i separately. Hence, for continuity at the knots we set up the following equations:

for x

$$\begin{aligned}
\hat{a}_{0_1} + \hat{a}_{1_1} t_{k_1} + \hat{a}_{2_1} t_{k_1}^2 + \hat{a}_{3_1} t_{k_1}^3 &= \hat{a}_{0_2} + \hat{a}_{1_2} t_{k_1} + \hat{a}_{2_2} t_{k_1}^2 + \hat{a}_{3_2} t_{k_1}^3, \\
\hat{a}_{0_2} + \hat{a}_{1_2} t_{k_2} + \hat{a}_{2_2} t_{k_2}^2 + \hat{a}_{3_2} t_{k_2}^3 &= \hat{a}_{0_3} + \hat{a}_{1_3} t_{k_2} + \hat{a}_{2_3} t_{k_2}^2 + \hat{a}_{3_3} t_{k_2}^3, \\
&\vdots \\
\hat{a}_{0_{n-1}} + \hat{a}_{1_{n-1}} t_{k_{n-1}} + \hat{a}_{2_{n-1}} t_{k_{n-1}}^2 + \hat{a}_{3_{n-1}} t_{k_{n-1}}^3 &= \hat{a}_{0_n} + \hat{a}_{1_n} t_{k_{n-1}} + \hat{a}_{2_n} t_{k_{n-1}}^2 + \hat{a}_{3_n} t_{k_{n-1}}^3,
\end{aligned} \tag{67}$$

for y

$$\begin{aligned}
\hat{b}_{0_1} + \hat{b}_{1_1} t_{k_1} + \hat{b}_{2_1} t_{k_1}^2 + \hat{b}_{3_1} t_{k_1}^3 &= \hat{b}_{0_2} + \hat{b}_{1_2} t_{k_1} + \hat{b}_{2_2} t_{k_1}^2 + \hat{b}_{3_2} t_{k_1}^3, \\
\hat{b}_{0_2} + \hat{b}_{1_2} t_{k_2} + \hat{b}_{2_2} t_{k_2}^2 + \hat{b}_{3_2} t_{k_2}^3 &= \hat{b}_{0_3} + \hat{b}_{1_3} t_{k_2} + \hat{b}_{2_3} t_{k_2}^2 + \hat{b}_{3_3} t_{k_2}^3, \\
&\vdots \\
\hat{b}_{0_{n-1}} + \hat{b}_{1_{n-1}} t_{k_{n-1}} + \hat{b}_{2_{n-1}} t_{k_{n-1}}^2 + \hat{b}_{3_{n-1}} t_{k_{n-1}}^3 &= \hat{b}_{0_n} + \hat{b}_{1_n} t_{k_{n-1}} + \hat{b}_{2_n} t_{k_{n-1}}^2 + \hat{b}_{3_n} t_{k_{n-1}}^3.
\end{aligned} \tag{68}$$

For preservation of smoothness at the knots we set:

for x

$$\begin{aligned}
\hat{a}_{1_1} + 2\hat{a}_{2_1} t_{k_1} + 3\hat{a}_{3_1} t_{k_1}^2 &= \hat{a}_{1_2} + 2\hat{a}_{2_2} t_{k_1} + 3\hat{a}_{3_2} t_{k_1}^2, \\
\hat{a}_{1_2} + 2\hat{a}_{2_2} t_{k_2} + 3\hat{a}_{3_2} t_{k_2}^2 &= \hat{a}_{1_3} + 2\hat{a}_{2_3} t_{k_2} + 3\hat{a}_{3_3} t_{k_2}^2, \\
&\vdots \\
\hat{a}_{1_{n-1}} + 2\hat{a}_{2_{n-1}} t_{k_{n-1}} + 3\hat{a}_{3_{n-1}} t_{k_{n-1}}^2 &= \hat{a}_{1_n} + 2\hat{a}_{2_n} t_{k_{n-1}} + 3\hat{a}_{3_n} t_{k_{n-1}}^2,
\end{aligned} \tag{69}$$

for y

$$\begin{aligned}
\hat{b}_{1_1} + 2\hat{b}_{2_1} t_{k_1} + 3\hat{b}_{3_1} t_{k_1}^2 &= \hat{b}_{1_2} + 2\hat{b}_{2_2} t_{k_1} + 3\hat{b}_{3_2} t_{k_1}^2, \\
\hat{b}_{1_2} + 2\hat{b}_{2_2} t_{k_2} + 3\hat{b}_{3_2} t_{k_2}^2 &= \hat{b}_{1_3} + 2\hat{b}_{2_3} t_{k_2} + 3\hat{b}_{3_3} t_{k_2}^2, \\
&\vdots \\
\hat{b}_{1_{n-1}} + 2\hat{b}_{2_{n-1}} t_{k_{n-1}} + 3\hat{b}_{3_{n-1}} t_{k_{n-1}}^2 &= \hat{b}_{1_n} + 2\hat{b}_{2_n} t_{k_{n-1}} + 3\hat{b}_{3_n} t_{k_{n-1}}^2.
\end{aligned} \tag{70}$$

The last set of equations is used to enforce continuity of the second derivative at the knots:

for x

$$\begin{aligned} 2\hat{a}_{2_1} + 6\hat{a}_{3_1} t_{k_1} &= 2\hat{a}_{2_2} + 6\hat{a}_{3_2} t_{k_1} , \\ 2\hat{a}_{2_2} + 6\hat{a}_{3_2} t_{k_2} &= 2\hat{a}_{2_3} + 6\hat{a}_{3_3} t_{k_2} , \\ &\vdots \\ 2\hat{a}_{2_{n-1}} + 6\hat{a}_{3_{n-1}} t_{k_{n-1}} &= 2\hat{a}_{2_n} + 6\hat{a}_{3_n} t_{k_{n-1}} , \end{aligned} \tag{71}$$

for y

$$\begin{aligned} 2\hat{b}_{2_1} + 6\hat{b}_{3_1} t_{k_1} &= 2\hat{b}_{2_2} + 6\hat{b}_{3_2} t_{k_1} , \\ 2\hat{b}_{2_2} + 6\hat{b}_{3_2} t_{k_2} &= 2\hat{b}_{2_3} + 6\hat{b}_{3_3} t_{k_2} , \\ &\vdots \\ 2\hat{b}_{2_{n-1}} + 6\hat{b}_{3_{n-1}} t_{k_{n-1}} &= 2\hat{b}_{2_n} + 6\hat{b}_{3_n} t_{k_{n-1}} . \end{aligned} \tag{72}$$

There are two groups of unknowns. One group is related to the observations w.r.t. the x-axis (a_0, a_1, a_2, a_3), and the other to the y-axis (b_0, b_1, b_2, b_3). Therefore, the formulation of vector of unknowns follows the previous pattern:

for x

$$\begin{aligned} \hat{\mathbf{a}}_1 &= \begin{bmatrix} \hat{a}_{0_1} & \hat{a}_{1_1} & \hat{a}_{2_1} & \hat{a}_{3_1} \end{bmatrix}^{\mathrm{T}} , \\ &\vdots \\ \hat{\mathbf{a}}_j &= \begin{bmatrix} \hat{a}_{0_j} & \hat{a}_{1_j} & \hat{a}_{2_j} & \hat{a}_{3_j} \end{bmatrix}^{\mathrm{T}} , \\ &\vdots \\ \hat{\mathbf{a}}_n &= \begin{bmatrix} \hat{a}_{0_n} & \hat{a}_{1_n} & \hat{a}_{2_n} & \hat{a}_{3_n} \end{bmatrix}^{\mathrm{T}} , \end{aligned} \tag{73}$$

for y

$$\begin{aligned} \hat{\mathbf{b}}_1 &= \begin{bmatrix} \hat{b}_{0_1} & \hat{b}_{1_1} & \hat{b}_{2_1} & \hat{b}_{3_1} \end{bmatrix}^{\mathrm{T}} , \\ &\vdots \\ \hat{\mathbf{b}}_j &= \begin{bmatrix} \hat{b}_{0_j} & \hat{b}_{1_j} & \hat{b}_{2_j} & \hat{b}_{3_j} \end{bmatrix}^{\mathrm{T}} , \\ &\vdots \\ \hat{\mathbf{b}}_n &= \begin{bmatrix} \hat{b}_{0_n} & \hat{b}_{1_n} & \hat{b}_{2_n} & \hat{b}_{3_n} \end{bmatrix}^{\mathrm{T}} . \end{aligned} \tag{74}$$

Finally, the vectors of unknowns are constructed as

$$\hat{\mathbf{X}}_x = \begin{bmatrix} \hat{\mathbf{a}}_1^{\mathrm{T}} & \cdots & \hat{\mathbf{a}}_j^{\mathrm{T}} & \cdots & \hat{\mathbf{a}}_n^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} , \tag{75}$$

$$\hat{\mathbf{X}}_y = \begin{bmatrix} \hat{\mathbf{b}}_1^{\mathrm{T}} & \cdots & \hat{\mathbf{b}}_j^{\mathrm{T}} & \cdots & \hat{\mathbf{b}}_n^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} . \tag{76}$$

From (21) and (22) we can compute the weight matrix $\mathbf{P}$ of the observations x_i, y_i. Since the observation equations are linear, the design matrix is constructed by combining the observation equations for each axis separately, similarly as in section 3.2. Therefore, for the first interval the matrix

$$\mathbf{A}_{1_x} = \mathbf{A}_{1_y} = \mathbf{A}_1 = \begin{bmatrix} 1 & t_{1_1} & t_{1_1}^2 & t_{1_1}^3 \\ 1 & t_{2_1} & t_{2_1}^2 & t_{2_1}^3 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & t_{m_1} & t_{m_1}^2 & t_{m_1}^3 \end{bmatrix} \tag{77}$$

is constructed and for the second interval the matrix

$$\mathbf{A}_{2_x} = \mathbf{A}_{2_y} = \mathbf{A}_2 = \begin{bmatrix} 1 & t_{1_2} & t_{1_2}^2 & t_{1_2}^3 \\ 1 & t_{2_2} & t_{2_2}^2 & t_{2_2}^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t_{m_2} & t_{m_2}^2 & t_{m_2}^3 \end{bmatrix}. \tag{78}$$

By following the pattern for $\mathbf{A}_1$ and $\mathbf{A}_2$ all matrices $\mathbf{A}_j$ are constructed in the same way until the end of the last interval. Thus for $\mathbf{A}_x$ and $\mathbf{A}_y$ we obtain

$$\mathbf{A}_x = \mathbf{A}_y = diag\begin{bmatrix} \mathbf{A}_1 & \cdots & \mathbf{A}_n \end{bmatrix}. \tag{79}$$

The formulation of the matrices of constraints is similar to section 3.2. The only difference is that two sets of matrices, $\mathbf{C}_{q_x}$ related to the *x*-axis and $\mathbf{C}_{q_y}$ related to the *y*-axis, are introduced, where q = 1, 2, 3. The matrices $\mathbf{C}_{1_x}$ and $\mathbf{C}_{1_y}$ refer to the equations of constraints that enforce the continuity of the curve:

$$\mathbf{C}_{1_x} = \mathbf{C}_{1_y} = \mathbf{C}_1 = \begin{bmatrix} 1 & t_{k_1} & t_{k_1}^2 & t_{k_1}^3 & -1 & -t_{k_1} & -t_{k_1}^2 & -t_{k_1}^3 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 1 & t_{k_2} & t_{k_2}^2 & t_{k_2}^3 & -1 & -t_{k_2} & -t_{k_2}^2 & -t_{k_2}^3 & \cdots & 0 \\ \vdots & & & & & & \vdots & & & & & & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & -t_{k_{n-1}}^3 \end{bmatrix}. \tag{80}$$

The matrices $\mathbf{C}_{2_x}$ and $\mathbf{C}_{2_y}$ enforce the smoothness of the spline:

$$\mathbf{C}_{2_x} = \mathbf{C}_{2_y} = \mathbf{C}_2 = \begin{bmatrix} 0 & 1 & 2t_{k_1} & 3t_{k_1}^2 & 0 & -1 & -2t_{k_1} & -3t_{k_1}^2 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 2t_{k_2} & 3t_{k_2}^2 & 0 & -1 & -2t_{k_2} & -3t_{k_2}^2 & \cdots & 0 \\ \vdots & & & & & & \vdots & & & & & & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & -3t_{k_{n-1}}^2 \end{bmatrix}. \tag{81}$$

The matrices $\mathbf{C}_{3_x}$ and $\mathbf{C}_{3_y}$ contain the coefficients from the equation of constraints that preserve the curvature at the knots:

$$\mathbf{C}_{3_x} = \mathbf{C}_{3_y} = \mathbf{C}_3 = \begin{bmatrix} 0 & 0 & 2 & 6t_{k_1} & 0 & 0 & -2 & -6t_{k_1} & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 6t_{k_2} & 0 & 0 & -2 & -6t_{k_2} & \cdots & 0 \\ \vdots & & & & & & \vdots & & & & & & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & -6t_{k_{n-1}} \end{bmatrix}. \tag{82}$$

Now we can set up $\mathbf{C}_x$ and $\mathbf{C}_y$ as

$$\mathbf{C}_x = \mathbf{C}_y = \begin{bmatrix} \mathbf{C}_1 & \mathbf{C}_2 & \mathbf{C}_3 \end{bmatrix}^{\mathrm{T}}. \tag{83}$$

5.3 Least-squares adjustment

The solution of the adjustment problem with constraints is obtained as already described in section 3.3 by introduction of Lagrange multipliers. In this case the vectors

$$\boldsymbol{\lambda}_x = \begin{bmatrix} \lambda_{1_x} & \lambda_{2_x} & \lambda_{3_x} \end{bmatrix}^{\mathrm{T}}, \tag{84}$$

$$\boldsymbol{\lambda}_y = \begin{bmatrix} \lambda_{1_y} & \lambda_{2_y} & \lambda_{3_y} \end{bmatrix}^{\mathrm{T}} \tag{85}$$

and finally

$$\boldsymbol{\lambda} = \begin{bmatrix} \boldsymbol{\lambda}_x & \boldsymbol{\lambda}_y \end{bmatrix}^{\mathrm{T}} \tag{86}$$

are introduced. With

$$\mathbf{L} = \begin{bmatrix} \mathbf{L}_x & \mathbf{L}_y \end{bmatrix}^{\mathrm{T}}, \tag{87}$$

$$\hat{\mathbf{X}} = \begin{bmatrix} \hat{\mathbf{X}}_x & \hat{\mathbf{X}}_y \end{bmatrix}^{\mathrm{T}}, \tag{88}$$

$$\mathbf{A} = diag\begin{bmatrix} \mathbf{A}_x & \mathbf{A}_y \end{bmatrix}, \tag{89}$$

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_x & \mathbf{C}_y \end{bmatrix}^{\mathrm{T}} \tag{90}$$

we obtain

$$\left[\begin{array}{c|c} \mathbf{A}^{\mathrm{T}}\mathbf{P}\mathbf{A} & \mathbf{C}^{\mathrm{T}} \\ \hline \mathbf{C} & \mathbf{0} \end{array}\right] \left[\begin{array}{c} \hat{\mathbf{X}} \\ \hline \boldsymbol{\lambda} \end{array}\right] = \left[\begin{array}{c} \mathbf{A}^{\mathrm{T}}\mathbf{P}\mathbf{A} \\ \hline \mathbf{0} \end{array}\right] \tag{91}$$

and the solution for the unknowns can be derived e.g. by using matrix inversion or other techniques as described in section 3.3. Remark: If there are no correlations between the sequences of observations x_i and y_i, the weight matrix is a block diagonal matrix

$$\mathbf{P} = \left[\begin{array}{c|c} \mathbf{P}_{xx} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{P}_{yy} \end{array}\right] \tag{92}$$

and hence the observation equations (61) and (62) do not share any common unknowns, the parameter estimation can be performed in two individual adjustments with

$$\left[\begin{array}{c|c} \mathbf{A}_x^{\mathrm{T}}\mathbf{P}_{xx}\mathbf{A}_x & \mathbf{C}_x^{\mathrm{T}} \\ \hline \mathbf{C}_x & \mathbf{0} \end{array}\right] \left[\begin{array}{c} \hat{\mathbf{X}}_x \\ \hline \boldsymbol{\lambda}_x \end{array}\right] = \left[\begin{array}{c} \mathbf{A}_x^{\mathrm{T}}\mathbf{P}_{xx}\mathbf{A}_x \\ \hline \mathbf{0} \end{array}\right] \tag{93}$$

and

$$\left[\begin{array}{c|c} \mathbf{A}_y^{\mathrm{T}}\mathbf{P}_{yy}\mathbf{A}_y & \mathbf{C}_y^{\mathrm{T}} \\ \hline \mathbf{C}_y & \mathbf{0} \end{array}\right] \left[\begin{array}{c} \hat{\mathbf{X}}_y \\ \hline \boldsymbol{\lambda}_y \end{array}\right] = \left[\begin{array}{c} \mathbf{A}_y^{\mathrm{T}}\mathbf{P}_{yy}\mathbf{A}_y \\ \hline \mathbf{0} \end{array}\right]. \tag{94}$$

With the solution either from (91) or from (93) and (94) the residual vectors

$$\mathbf{v}_x = \mathbf{A}_x\hat{\mathbf{X}}_x - \mathbf{L}_x, \quad \mathbf{v}_y = \mathbf{A}_y\hat{\mathbf{X}}_y - \mathbf{L}_y \tag{95}$$

can be obtained according to (32). The actual magnitude of the vectors of the residuals is computed by

$$v_i = \sqrt{v_{x_i}^2 + v_{y_i}^2}\ . \tag{96}$$

The direction of the vector for each residual varies throughout the curve. Hence, the magnitude of the residual does not resemble neither the perpendicular nor the axial distance from the measured point to the estimated point on the graph of the curve. It is possible for the residuals to be perpendicular to the graph of the curve by imposing additional constraints. However, the solution of that problem goes beyond the scope of this article.

With the vector of residuals

$$\mathbf{v} = \begin{bmatrix} \mathbf{v}_x & \mathbf{v}_y \end{bmatrix}^{\mathrm{T}} \tag{97}$$

the reference standard deviation can be computed from (37). This value can be applied in a knot placement strategy as described in section 3.4.

5.4 Discussion

The spline approximation presented in this section uses both the values y_i and the values x_i as observations. However, instead of introducing them and the corresponding residuals into (6), which would lead to nonlinear condition equations, here they are introduced as observations in an adjustment using the Gauss-Markov model. This results in the disadvantage of an "irregular" distribution of the residual vector for each point w.r.t. the graph of the curve. This limits the informational value of the residuals, e.g. for engineering geodesy tasks such as deformation analysis. Another

problem with this approach is that the values of the parameter (in this case t) that refer to the observations, must be chosen in advance. An important issue is that the values of t have to be proportional with the distances between the observed points. The advantage of using spline curves instead of spline functions is that one-dimensional problems in a multidimensional space can be approximated without any restrictions on the direction in which the curve is heading. Hence, this approach can be useful for approximation of railways, pipelines or similar structures. The spline curve approximation is mainly used in computer graphics for getting a result which satisfies aesthetic requirements.

6 Conclusions and outlook

Splines are very applicable for approximation and interpolation of arbitrary shapes. They offer a lot of possibilities in many fields including the field of engineering geodesy. In our opinion, in spite of the many textbooks that are written about splines, e.g. by De Boor [1978], Farin [1993], Schumaker [2007], Micula and Micula [2012], Piegl and Tiller [2012], the notion of spline is not always comprehensively explained. In many of those publications, especially the articles and the textbooks on B-splines, the explanations are written in a style that is understandable only to a narrow group of experts.

In this article different approaches for spline approximation were elaborated. The simplest and the most intuitive approach for understanding the spline approximation is the approximation using a spline function constructed from ordinary cubic polynomials. With the values y_i as observations and x_i as error-free values the resulting adjustment problem can be solved within the Gauss-Markov model with constraints for the unknowns. What makes it intuitive are the piecewise polynomials represented in their simplest form, and the clearly stated constraints imposed on each knot.

The spline approximation using truncated polynomials has an advantage over the previous type of approximation, because it does not require explicit constraints. With the values y_i as observations and x_i as error-free values the resulting adjustment problem can be solved within the Gauss-Markov model. It is also possible to use the truncated polynomials instead of a spline function to approximate a spline curve.

The so called spline curve, in this article, refers to a spline that is represented as a set of parametric equations. Both values y_i and x_i are introduced as two separate sequences of observations in an adjustment using the Gauss-Markov model with constraints for the unknowns. This approach results in an "irregular" distribution of the residual vector for each point w.r.t. the graph of the curve, which limits the interpretation of the residuals, e.g. for engineering geodesy tasks. It is possible for the residuals to become perpendicular to the graph of the curve by imposing additional constraints. The advantage of spline curves is that a problem can be approximated without any restrictions to the direction in which the curve is heading.

Using cubic polynomials as functional model the basic methodology of spline approximation as an adjustment problem from a geodetic point of view is finally available. This enables us now to consider the case that both the values y_i and x_i are introduced as observations in (6) which leads to nonlinear condition equations with additional constraints for the unknowns. The fact that y_i and x_i are introduced as observations makes this approach the most important one for applications in engineering geodesy, e.g. for approximation of profiles measured with terrestrial laser scanners. If the observations are equally weighted the residuals are perpendicular to the graph of the spline function and hence can be utilized for deformation monitoring purposes. In Neitzel et al. [2017] a solution for the resulting adjustment problem will be presented.

Due to the fact that in many publications B-splines are used, the question arises whether it is possible to elaborate a clear and comprehensive explanation from a geodetic point of view also for the B-spline approach. This question is dealt with in the forthcoming Part 2 of a series of three articles. In Part 3 the numerical stability of the considered approaches for spline approximation as well as the utilization of splines for deformation detection will be investigated using numerical examples.

References

Björck, Å. (2015). Numerical Methods in Matrix Computations, Texts in Applied Mathematics 59, Springer International Publishing.

Bronshtein, I. N. and Semendyayev, K. A. (2007). Handbook of Mathematics. Springer-Verlag, Berlin Heidelberg.

Bureick, J., Alkhatib, H. and Neumann, I. (2016a). Robust Spatial Approximation of Laser Scanner Point Clouds by Means of Free-form Curve Approaches in Deformation Analysis. Journal of Applied Geodesy 10: 27–35.

Bureick, J., Neuner, H., Harmening, C., Neumann, I. (2016b). Curve and Surface Approximation of 3D Point Clouds. Allgemeine Vermessungs-Nachrichten (AVN), 123: 315-327.

Buse, A. and Lim, L. (1977). Cubic splines as a special case of restricted least squares. Journal of the American Statistical Association, 72(357):64-68.

Cox, M., Harris, P. and Jones, H. M. (1989). Algorithms for Approximation, II. Chapman and Hall, London.

De Boor, C. (1978). A practical guide to splines, volume 27. Springer-Verlag, New York.

De Boor, C. (1986). B(asic)-spline basics. Technical report, No. MRC-TSR-2952. Wisconsin Univ-Madison Mathematics Research Center.

De Boor, C. and Rice, J. R. (1968). Least squares cubic spline approximation, I -Fixed knots, II -variable knots. Purdue University, Department of Computer Science, Technical Reports, 20 and 21.

Farin, G. (1993). Curves and surfaces for computer-aided geometric design: a practical guide. Academic press Inc., San Diego.

Fuller, W.A. (1969). Grafted Polynomials as Approximating Functions. Australian Journal of Agricultural Economics, 13: 35-46.

Gálveza, A., Iglesiasa, A., Avilaa, A., Oteroc, C., Ariasc, R. and Manchado, C. (2015). Elitist clonal selection algorithm for optimal choice of free knots in B-spline data fitting. Applied Soft Computing 26: 90-106.

Ghilani, C. D. (2010). Adjustment Computations: Spatial Data Analysis, 5th Edition. John Wiley & Sons, Inc.

Golub, G. H. and Van Loan, C. F. (1996). Matrix computations, volume 3. John Hopkins University Press, Baltimore.

Harmening, C. and Neuner, H. (2016). Choosing the Optimal Number of B-spline Control Points (Part 1: Methodology and Approximation of Curves), Journal of Applied Geodesy 10: 139-157.

Helmert, F. R. (1924). Die Ausgleichungsrechnung nach der Methode der kleinsten Quadrate. 3. Auflage, Teubner Verlag, Leipzig.

Klaus, R. and Van Ness, H. (1967). An extension of the spline fit technique and applications to thermodynamic data. American Institute of Chemical Engineers (AIChE) Journal, 13(6):1132-1136.

Lucas, S. (2003). Least squares cubic splines without B-splines. Gazette of the Australian Mathematical Society, (30):207-2011.

Micula, G. and Micula, S. (2012). Handbook of splines, volume 462. Springer Science & Business Media.

Neitzel, F., Ezhov, N. and Petrovic, S. (2018). Total Least-Squares Spline Approximation. *Under preparation for Survey Review*.

Niemeier, W. (2008). Ausgleichungsrechnung: Statistische Auswertemethoden. Walter de Gruyter.

Park, H. (2011). B-spline surface fitting based on adaptive knot placement using dominant columns. Computer-Aided Design, 43(3):258-264.

Piegl, L. and Tiller, W. (2012). The NURBS book. Springer Science & Business Media.

Schoenberg, I. J. (1946). Contributions to the problem of approximation of equidistant data by analytic functions, part b: On the problem of osculatory interpolation, a second class of analytic approximation formulae. Quarterly of Applied Mathematics, 4(2):112-141.

Schumaker, L. (2007). Spline functions: basic theory. Cambridge University Press.

Schwetlick, H. and Schütze, T. (1995). Least squares approximation by splines with free knots. BIT Numerical mathematics, 35(3):361-384.

Smith, P. L. (1979). Splines as a useful and convenient statistical tool. The American Statistician, 33(2):57–62.

Teunissen, P.J.G. (2000). Testing Theory. An Introduction. VSSD Publishing, Delft, The Netherlands.

Walsh, J., Ahlberg, J., and Nilson, E. (1962). Best approximation properties of the spline fit. Journal of Applied Mathematics and Mechanics, 11(2):225-234.

Wang, W., Pottmann, H., and Liu, Y. (2006). Fitting b-spline curves to point clouds by curvature-based squared distance minimization. ACM Transactions on Graphics (ToG), 25(2):214-238.

Yanagihara, H. and Ohtaki, M. (2003). Knot-placement to avoid over fitting in b-spline scedastic smoothing. Communications in Statistics-Simulation and Computation, 32(3):771-785.

Zheng, W., Bo, P., Liu, Y., and Wang, W. (2012). Fast b-spline curve fitting by l-bfgs. Computer Aided Geometric Design, 29(7):448-462.