CONSTRAINED APPROXIMATION BY SPLINES WITH FREE KNOTS

TORSTEN SCHÜTZE AND HUBERT SCHWETLICK

ABSTRACT. In this paper, a method that combines shape preservation and least squares approximation by splines with free knots is developed. Besides the coefficients of the spline a subset of the knot sequence, the so-called *free knots*, is included in the optimization process resulting in a nonlinear least squares problem in both the coefficients and the knots. The original problem, a special case of a *constrained semi-linear least squares problem*, is reduced to a problem that has only the knots of the spline as variables. The reduced problem is solved by a generalized Gauss-Newton method. Special emphasize is given to the efficient computation of the residual function and its Jacobian.

1. INTRODUCTION

Let $\{x_i, y_i\}$ $(i = 1, \ldots, m)$ be given data with monotonously increasing abscissae $x_i \in [a, b] \subset \mathbb{R}$ and noisy measurements $\{y_i\}$ of values of an unknown smooth function $g \in \mathbb{C}^q[a, b]$. We want to approximate these data by a function s from the n-dimensional spline space $\mathcal{S}_{k,\mathbf{t}}$ consisting of all polynomial splines of order $k \geq 1$ with knot sequence $\mathbf{t} = \{t_j\}$ where

$$t_1 = \cdots = t_k = a < t_{k+1} \leq \cdots \leq t_n < b = t_{n+1} = \cdots = t_{n+k}$$

and $m \geq n$. The parameters of the spline s have to be chosen in such a way that the *smoothing functional* which is known as Schoenberg functional

$$\mathfrak{f}(s) := \frac{1}{2} \sum_{i=1}^{m} [y_i - s(x_i)]^2 + \mu \frac{1}{2} \int_a^b \left[s^{(r)}(x) \right]^2 dx$$

with the *smoothing parameter* $\mu > 0$ and fixed $r \in \{0, \ldots, q\}$ becomes minimal.

Suppose that besides the data $\{x_i, y_i\}$ additional information on the shape of the function g is known, e.g., $g^{(p)}(x) \geq 0$ for all $x \in [a, b]$ with a prescribed order $p \in \{0, \ldots, q\}$ of derivative. We then require that the spline s is shape preserving, i.e., $s^{(p)}(x) \geq 0$ for all $x \in [a, b]$ if $g^{(p)}(x) \geq 0$ for all $x \in [a, b]$, $p \in \{0, \ldots, q\}$. Later on even more general shape constraints will be allowed.

The number and location of the knots $\{t_j\}$ is of vital importance for the definition of the spline space $\mathcal{S}_{k,\mathbf{t}}$ and, therefore, for the quality of the approximation.

There exists a tremendous number of papers and efficient numerical methods on shape preserving approximation by splines with fixed knots: In Micchelli/Utreras [MU88] existence and uniqueness of spline interpolation and smoothing in a convex subset of a Hilbert space is investigated. Elfving/Andersson [EA88] consider the case $r = 2$ and convexity conditions $s''(x) \geq \delta(x)$, δ given. Schmidt/Scholz [SS90] develop an efficient method for $r = 2$ and the generalized convexity conditions $\delta(x) \leq s''(x) \leq \epsilon(x)$ (δ, ϵ linear $\mathbb{C}^0$ splines) by solving the unconstrained dual problem instead of the partially separable primal problem. In Schwetlick/Kunert

---

1991 *Mathematics Subject Classification.* Primary 65D10, 65D07; Secondary 41A15, 41A29.

Key words and phrases. Data fitting, shape preservation, splines with free knots, constrained approximation, semi-linear least squares problems.

Research of the first author was supported by Deutsche Forschungsgemeinschaft under grant Schm 968/2-1,2-2.

[SK93] the general problem $\delta(x) \leq s^{(p)}(x) \leq \epsilon(x)$ (δ, ϵ linear $\mathbb{C}^0$ splines; $p, r \in \{0, \ldots, q\}$) is treated using orthogonalization techniques.

If we regard the knots of the spline as free parameters then the approximation can significantly be improved, in general. There are several methods for the computation of splines with free knots in the least squares context, but all without considering shape constraints, see [SS95]. Almost all methods for computing such "optimal" splines separate the linear and nonlinear aspects of the problem and solve an optimization problem in the spline knots only. While in [dBR68], [Jup78], and [Die79] the approximation case ($\mu = 0$) is investigated, in [SS95] the regularizing smoothing term $\mu \frac{1}{2} \int_a^b [s^{(r)}(x)]^2 \, dx$ widely used in approximation theory is added. For the equivalence of the original and the reduced problem, a certain regularity condition must be fulfilled. In the case of smoothing splines with free knots this condition is satisfied for all knot sequences, see again [SS95].

In this paper we want to combine shape preserving approximation and least squares approximation by splines with free knots. A subset of the knot sequence, the so-called *free knots*, is included in the optimization process resulting in a nonlinear least squares problem in both the coefficients and the knots of the spline. One obtains linear inequality constraints on the free knots and nonlinear inequality constraints on the free knots and the coefficients. The latter constraints are linear in the coefficients if the knot sequence is fixed. Hence, the original problem is a special case of a constrained semi-linear least squares problem (CSLS), a generalization of the well known separable least squares problems. Applying results of Parks [Par85], we derive a reduced problem in the free knots only and show under which conditions the original and the reduced problems are equivalent.

The reduced problem is solved by a generalized Gauss-Newton method. Since the structure of the Jacobian is rather complicated we use an approximation to this Jacobian similar to the proposal of Kaufman [Kau75] for separable least squares problems. We develop an algorithm that, if the number and initial position of knots is given, seeks for the optimal placement of the knots depending on the data $\{x_i, y_i\}$ with respect to the shape constraints.

The paper is organized as follows: In Section 2, the smoothing functional $\mathfrak{f}$, the constraints on the derivative of the spline s, and the constraints on the free knots are expressed as functions of the spline coefficients and the knot sequence. We then define the original and the reduced problem. In Section 3, we summarize the results from the unpublished PhD Thesis [Par85] for reducible nonlinear programming and state the conditions for the equivalence of these problems in the general case. The existence of solutions to the reduced problem and the correspondence between the problems in our special context of spline smoothing is shown in Section 4. In Section 5, a generalized Gauss-Newton method for the reduced problem is developed. We show how an approximation to the Jacobian can be computed in an efficient and numerically stable way. Special emphasize is given to the exploitation of the inherent sparsity structure. Finally, in Section 6 some results of numerical tests are given. Moreover, the knot placement delivered by our algorithm is compared with an adaptive strategy for the placement of knots taken from the literature.

2. Formulation of the problem

2.1. **Representation of the smoothing functional.** Let $\mathcal{S}_{k,\mathbf{t}}$ denote the space of polynomial splines of order $k \geq 1$ with respect to the knot sequence $\mathbf{t} = (t_1, \ldots, t_{n+k})^T \in \mathbb{R}^{n+k}$ where

$$t_1 = \cdots = t_k = a < t_{k+1} \leq \cdots \leq t_n < b = t_{n+1} = \cdots = t_{n+k}$$

represented by its B-spline basis as

$$\mathcal{S}_{k,\mathbf{t}} := \left\{ s \in \mathcal{S}_{k,\mathbf{t}} \;:\; s = \sum_{j=1}^{n} B_{j,k,\mathbf{t}}\alpha_j,\ \alpha_j \in \mathbb{R} \right\}. \tag{2.1}$$

The B-splines $B_{j,k,\mathbf{t}}$ are recursively defined by

$$B_{j,1,\mathbf{t}}(t) := \begin{cases} 1 & \text{if} \quad t_j \le t < t_{j+1} \\ 0 & \text{otherwise} \end{cases} \tag{2.2}$$

$$B_{j,k,\mathbf{t}}(t) := \omega_{j,k}(t) \cdot B_{j,k-1,\mathbf{t}}(t) + (1 - \omega_{j+1,k}(t)) \cdot B_{j+1,k-1,\mathbf{t}}(t) \quad \text{for} \quad k > 1 \tag{2.3}$$

where

$$\omega_{j,k}(t) := \begin{cases} \frac{t-t_j}{t_{j+k-1}-t_j} & \text{if} \quad t_j < t_{j+k-1} \\ 0 & \text{otherwise.} \end{cases} \tag{2.4}$$

The notation $B_{j,k,\mathbf{t}}$ indicates that the B-splines depend (nonlinearly) on the knot sequence $\mathbf{t}$.

Under the assumption

$$t_j < t_{j+k-q} \qquad (j = q+1, \ldots, n) \tag{2.5}$$

that implies $q+1 \le n$ and $k > q$ we have $\mathcal{S}_{k,\mathbf{t}} \subset \mathbb{C}^q[a,b]$, $\dim \mathcal{S}_{k,\mathbf{t}} = n$, and $\mathcal{S}_{k,\mathbf{t}}$ is the linear space of all piecewise polynomials of order k with breakpoints t_j which are $k-1-\#t_j$ times continuously differentiable at t_j, where $\#t_j$ is the multiplicity of the node t_j.

Let

$$s(x) = \sum_{j=1}^{n} B_{j,k,\mathbf{t}}(x)\alpha_j = \boldsymbol{\beta}^T(x,\mathbf{t})\boldsymbol{\alpha} \tag{2.6}$$

be the unique representation of a spline $s \in \mathcal{S}_{k,\mathbf{t}}$ where

$$\boldsymbol{\beta}(x,\mathbf{t}) := (B_{1,k,\mathbf{t}}(x), \ldots, B_{n,k,\mathbf{t}}(x))^T \in \mathbb{R}^n, \quad \boldsymbol{\alpha} := (\alpha_1, \ldots, \alpha_n)^T \in \mathbb{R}^n$$

is the vector of B-splines and the vector of spline coefficients, respectively. With the observation matrix

$$\mathbf{B}(\mathbf{t}) := (B_{j,k,\mathbf{t}}(x_i))_{i=1,\ldots,m;\, j=1,\ldots,n} \in \mathbb{R}^{m,n}$$

and the vector $\mathbf{y} := (y_1, \ldots, y_m)^T \in \mathbb{R}^m$ of data the *approximation term* which describes the least squares error is given by

$$\varphi := \frac{1}{2}\sum_{i=1}^{m}[y_i - s(x_i)]^2 = \frac{1}{2}\sum_{i=1}^{m}\Big[y_i - \sum_{j=1}^{n} B_{j,k,\mathbf{t}}(x_i)\alpha_j\Big]^2 = \frac{1}{2}\left\|\mathbf{y} - \mathbf{B}(\mathbf{t})\boldsymbol{\alpha}\right\|^2; \tag{2.7}$$

if not otherwise stated, $\|.\|$ denotes the Euclidean vector norm.

Next we derive a similar representation of the *smoothing term*

$$\bar{\rho} := \frac{1}{2}\int_a^b [s^{(r)}(x)]^2\,dx = \frac{1}{2}\left\|s^{(r)}\right\|_{L_2}^2 \quad \text{with fixed } r \in \{0, \ldots, q\}. \tag{2.8}$$

At first, the derivative of a spline is expressed in terms of its coefficients and the knot sequence.

Lemma 2.1 (Derivative of a spline with respect to its argument).
Let $s \in \mathcal{S}_{k,\mathbf{t}}$ and $t_j < t_{j+k-r}$ $(j = r+1, \ldots, n)$. Then the rth derivative of s with respect to its argument exists and is a spline of order $k-r$ to the same knot sequence. If $s(x) = \sum_{j=1}^{n} B_{j,k,\mathbf{t}}\alpha_j$, then $s^{(r)}$ is given by

$$s^{(r)}(x) = \sum_{j=r+1}^{n} B_{j,k-r,\mathbf{t}}(x)\alpha_j^{(r)} = \boldsymbol{\beta}_r^T(x,\mathbf{t})\boldsymbol{\alpha}^{(r)} \tag{2.9}$$

where

$$\boldsymbol{\beta}_r(x,\mathbf{t}) := \left(B_{r+1,k-r,\mathbf{t}}(x),\dots,B_{n,k-r,\mathbf{t}}(x)\right)^T \in \mathbb{R}^{n-r},$$
$$\boldsymbol{\alpha}^{(r)} := \left(\alpha_{r+1}^{(r)},\dots,\alpha_n^{(r)}\right)^T \in \mathbb{R}^{n-r}.$$

The coefficients $\boldsymbol{\alpha}^{(r)}$ *are related to the coefficients* $\boldsymbol{\alpha}$ *by* $\boldsymbol{\alpha}^{(r)} = \mathbf{D}_r\boldsymbol{\alpha}$ *with* $\mathbf{D}_0 := \mathbf{I} \in \mathbb{R}^{n,n}$, $\mathbf{D}_r := \mathbf{H}_r\mathbf{L}_r\dots\mathbf{H}_1\mathbf{L}_1 \in \mathbb{R}^{n-r,n}$, *and* $\mathbf{H}_\nu$ *and* $\mathbf{L}_\nu$ *are defined by*

$$\mathbf{H}_\nu := (k-\nu)\,\mathrm{diag}\left(\frac{1}{t_{k+j-\nu}-t_j}\right)_{j=\nu+1,\dots,n} \in \mathbb{R}^{n-\nu,n-\nu}$$

$$\mathbf{L}_\nu := \begin{bmatrix} -1 & 1 & & & & \\ & -1 & 1 & & & \\ & & \ddots & \ddots & & \\ & & & -1 & 1 & \\ & & & & -1 & 1 \end{bmatrix} \in \mathbb{R}^{n-\nu,n-\nu+1}$$

for $\nu = 1,\dots,r$.

Note that the matrix $\mathbf{D}_r = \mathbf{D}_r(\mathbf{t})$ depends nonlinearly on the knots.

Since the computation of $\bar\rho$ requires the integration over products of B-splines, we look for a cheaper approximation $\tilde\rho$. As in [SK93] and [SS95], we replace the L_2-semi-norm by its discrete analogue in $\mathcal{S}_{k-r,\mathbf{t}}$, i.e., by

$$\tilde\rho := \frac{1}{2}\left\|s^{(r)}\right\|_{l_2}^2 := \frac{1}{2}\sum_{j=r+1}^{n}\left(\alpha_j^{(r)}\right)^2\frac{t_{j+k-r}-t_j}{k-r}. \tag{2.10}$$

By inserting expression (2.9) into (2.10) we obtain

$$\tilde\rho = \frac{1}{2}\boldsymbol{\alpha}^T\mathbf{D}_r^T(\mathbf{t})\tilde{\mathbf{F}}_r^T(\mathbf{t})\tilde{\mathbf{F}}_r(\mathbf{t})\mathbf{D}_r(\mathbf{t})\boldsymbol{\alpha} = \frac{1}{2}\left\|\tilde{\mathbf{S}}_r(\mathbf{t})\boldsymbol{\alpha}\right\|^2$$

where

$$\tilde{\mathbf{F}}_r(\mathbf{t}) := \mathrm{diag}\left(\sqrt{\frac{t_{j+k-r}-t_j}{k-r}}\right)_{j=r+1,\dots,n} \in \mathbb{R}^{n-r,n-r}.$$

The matrix $\tilde{\mathbf{S}}_r(\mathbf{t}) := \tilde{\mathbf{F}}_r(\mathbf{t})\mathbf{D}_r(\mathbf{t})$ is upper triangular with bandwidth $r+1$. Analogously, the exact smoothing term $\bar\rho$ can be expressed by $\bar\rho = \frac{1}{2}\left\|\bar{\mathbf{S}}_r(\mathbf{t})\boldsymbol{\alpha}\right\|^2$ with $\bar{\mathbf{S}}_r(\mathbf{t}) := \bar{\mathbf{F}}_r(\mathbf{t})\mathbf{D}_r(\mathbf{t})$ where $\bar{\mathbf{F}}_r$ is the Cholesky factor of the Gram matrix defined by the r-th derivatives of the B-splines, see [SK93] for details.

Approximation and smoothing term together define the final *smoothing functional*

$$\begin{aligned} \mathsf{f}(\boldsymbol{\alpha},\mathbf{t}) &:= \varphi + \mu\rho = \frac{1}{2}\left\|\mathbf{y}-\mathbf{B}(\mathbf{t})\boldsymbol{\alpha}\right\|^2 + \mu\frac{1}{2}\left\|\mathbf{S}_r(\mathbf{t})\boldsymbol{\alpha}\right\|^2 \\ &= \frac{1}{2}\left\|\begin{pmatrix}\mathbf{y}\\ \mathbf{0}\end{pmatrix} - \begin{bmatrix}\mathbf{B}(\mathbf{t})\\ \sqrt{\mu}\mathbf{S}_r(\mathbf{t})\end{bmatrix}\boldsymbol{\alpha}\right\|^2 \end{aligned}$$

where either $\rho = \bar\rho = \frac{1}{2}\left\|\bar{\mathbf{S}}_r(\mathbf{t})\boldsymbol{\alpha}\right\|^2$, $\mathbf{S}_r(\mathbf{t}) = \bar{\mathbf{S}}_r(\mathbf{t})$ (exact smoothing term), or $\rho = \tilde\rho = \frac{1}{2}\left\|\tilde{\mathbf{S}}_r(\mathbf{t})\boldsymbol{\alpha}\right\|^2$, $\mathbf{S}_r(\mathbf{t}) = \tilde{\mathbf{S}}_r(\mathbf{t})$ (approximate smoothing term).

Lemma 2.2 (Full rank property of the system matrix).
If the regularity condition $m \geq r$ *and* $\mu > 0$ *is satisfied, then the regularized observation matrix*

$$\mathbf{B}_\mu(\mathbf{t}) := \begin{bmatrix}\mathbf{B}(\mathbf{t})\\ \sqrt{\mu}\mathbf{S}_r(\mathbf{t})\end{bmatrix} \in \mathbb{R}^{m+n-r,n}$$

has full rank n.

2.2. **Constraints on derivatives.** In order to obtain shape preserving splines, we impose constraints on the derivatives of the spline. One approach is to prescribe lower and upper bounds for certain derivatives of the spline on the subintervals $[t_i, t_{i+1})$ $(i = k, \ldots, n)$. Let the knot sequence $\mathbf{t}$ fulfill condition (2.5), and let $p \in \{0, \ldots, q\}$ be a fixed order of derivative. Then the derivative $s^{(p)}$ exists and has the representation

$$s^{(p)}(x) = \sum_{j \in K_i} B_{j,k-p,\mathbf{t}}(x) \alpha_j^{(p)} \tag{2.11}$$

where $K_i := \{i-k+p+1, \ldots, i\}$ for $x \in [t_i, t_{i+1})$. We consider the shape constraints

$$l_i^{(p)} \le s^{(p)}(x) \le u_i^{(p)} \quad \forall x \in [t_i, t_{i+1}),\ i = k, \ldots, n \tag{2.12}$$

with $2(n-k+1)$ constants $\mathbf{l} := \left(l_k^{(p)}, \ldots, l_n^{(p)}\right)^T \in \mathbb{R}^{n-k+1}$, $\mathbf{u} := \left(u_k^{(p)}, \ldots, u_n^{(p)}\right)^T \in \mathbb{R}^{n-k+1}$. Since the B-splines form a nonnegative partition of unity, we have

$$\min\left\{\alpha_j^{(p)} \ : \ j \in K_i\right\} \le s^{(p)}(x) = \sum_{j \in K_i} B_{j,k-p}(x) \alpha_j^{(p)} \le \max\left\{\alpha_j^{(p)} \ : \ j \in K_i\right\}.$$

Therefore, the condition

$$l_i^{(p)} \le \min\left\{\alpha_j^{(p)} \ : \ j \in K_i\right\} \text{ and } \max\left\{\alpha_j^{(p)} \ : \ j \in K_i\right\} \le u_i^{(p)} \quad i = k, \ldots, n \tag{2.13}$$

is sufficient for (2.12) to be satisfied. The sufficient condition (2.13) can equivalently be written as box constraints

$$L_j^{(p)} \le \alpha_j^{(p)} \le U_j^{(p)} \qquad j = p+1, \ldots, n$$

with $2(n-p)$ constants

$$L_j^{(p)} := \max\left\{l_i^{(p)} \ : \ i \in W_j\right\}, \quad U_j^{(p)} := \min\left\{u_i^{(p)} \ : \ i \in W_j\right\}$$

and the index set $W_j := \Big\{\max\{j, k\}, \ldots, \min\{j+k-p-1, n\}\Big\}$. Using Lemma 2.1, in vector notation and in component-wise partial ordering we obtain

$$\mathbf{L} \le \mathbf{D}_p(\mathbf{t}) \boldsymbol{\alpha} \le \mathbf{U} \tag{2.14}$$

where

$$\mathbf{L} := \left(L_{p+1}^{(p)}, \ldots, L_n^{(p)}\right)^T \in \mathbb{R}^{n-p}, \quad \mathbf{U} := \left(U_{p+1}^{(p)}, \ldots, U_n^{(p)}\right)^T \in \mathbb{R}^{n-p}.$$

In the following we formally allow $-\infty$ and $+\infty$ as upper and lower bounds. Thus condition (2.12) covers also the practical important case of one-sided constraints on $s^{(p)}$, e.g., $s^{(p)} \ge 0$. The algorithms described later on are able to handle these special cases, too.

We call the shape constraints (2.12) *consistent*, if $L_j^{(p)} \le U_j^{(p)}$, $j = p+1, \ldots, n$, $(\mathbf{L} \le \mathbf{U})$, and *strictly consistent*, if $L_j^{(p)} < U_j^{(p)}$, $j = p+1, \ldots, n$, $(\mathbf{L} < \mathbf{U})$.

Example 2.1. $k = 4$, $n = 9$, $p = 2$, $l_i^{(2)} \le s''(x) \le u_i^{(2)}$ $\forall x \in [t_i, t_{i+1})$ $i = 4, \ldots, 9$

$$\begin{matrix} \mathbf{l} = (0,0,0,0,-\infty,-\infty) \\ \mathbf{u} = (+\infty,+\infty,+\infty,+\infty,-1,-1) \end{matrix} \implies \begin{matrix} \mathbf{L} = (0,0,0,0,0,-\infty,-\infty) \\ \mathbf{U} = (+\infty,+\infty,+\infty,+\infty,-1,-1,-1) \end{matrix}$$

The consistency relation $\mathbf{L} \le \mathbf{U}$ is violated $(L_7 > U_7)$ although $\mathbf{l} < \mathbf{u}$.

Example 2.2. $k = 4$, $n = 9$, $p = 1$, $s'(x) \ge 0$ $\forall x \in [t_4, t_8)$, $s'(x) \le 0$ $\forall x \in [t_9, t_{10})$

$$\begin{matrix} \mathbf{l} = (0,0,0,0,-\infty,-\infty) \\ \mathbf{u} = (+\infty,+\infty,+\infty,+\infty,+\infty,0) \end{matrix} \implies \begin{matrix} \mathbf{L} = (0,0,0,0,0,0,-\infty,-\infty) \\ \mathbf{U} = (+\infty,+\infty,+\infty,+\infty,+\infty,0,0,0) \end{matrix}$$

The shape constraints are consistent, but not strictly consistent.

Example 2.3. $k = 4$, $n = 9$, $p = 2$, $s''(x) \geq 0\ \forall x \in [t_4, t_8)$, $s''(x) \leq 0\ \forall x \in [t_9, t_{10})$

$$\begin{matrix} \mathbf{l} = (0,0,0,0,-\infty,-\infty) \\ \mathbf{u} = (+\infty,+\infty,+\infty,+\infty,+\infty,0) \end{matrix} \quad \Longrightarrow \quad \begin{matrix} \mathbf{L} = (0,0,0,0,0,-\infty,-\infty) \\ \mathbf{U} = (+\infty,+\infty,+\infty,+\infty,+\infty,0,0) \end{matrix}$$

The shape constraints are strictly consistent.

2.3. **Constraints on the free knots.** We include a subset $\tilde{\mathbf{t}}$ of the knot sequence $\mathbf{t}$, the so-called *free knots*, into the optimization process. The number of free knots is denoted by l, and $\tilde{\mathbf{t}} = \left(t_{p(1)}, \ldots, t_{p(l)}\right)^T \in \mathbb{R}^l$ is the vector of free knots where $\mathbf{p} = (p(1), \ldots, p(l))^T \in \mathbb{Z}^l$ contains the indices of these free knots. We require that only inner knots $t_{k+1}, \ldots, t_n$ can be free, i.e., $k < p(1) < \cdots < p(l) < n+1$.

Algorithms for the computation of splines with free knots have to avoid the coalescing of the knots. The methods from the literature differ mainly in the way these constraints are included. As in [SS95] we bound the relative distance of two consecutive knots from below according to

$$t_{p(j)-1} + \epsilon\left(t_{p(j)+1} - t_{p(j)-1}\right) \leq t_{p(j)} \leq t_{p(j)+1} - \epsilon\left(t_{p(j)+1} - t_{p(j)-1}\right), \tag{2.15}$$

$j = 1, \ldots, l$, with the relative knot separation parameter $\epsilon > 0$. Condition (2.15) can equivalently be written in matrix notation as $\mathbf{C}\tilde{\mathbf{t}} - \mathbf{h} \geq \mathbf{0}$ with $\mathbf{C} \in \mathbb{R}^{2l,l}$, $\mathbf{h} \in \mathbb{R}^{2l}$ where $\mathbf{C}$ contains at most three non-zero elements per row. The matrix $\mathbf{C}$ and the vector $\mathbf{h}$ depend on $\mathbf{t}\backslash\{\tilde{\mathbf{t}}\}$ and the parameter ϵ.

2.4. **Formulation of the complete problem.** Finally, we can formulate the complete problem. Let $t_j < t_{j+k-q}$ $(j = q+1, \ldots, n)$ and $p, r \in \{0, \ldots, q\}$. Then the problem

$$\mathfrak{f}(\boldsymbol{\alpha}, \tilde{\mathbf{t}}) := \frac{1}{2}\left\| \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} - \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix} \boldsymbol{\alpha} \right\|^2 \longrightarrow \min_{\boldsymbol{\alpha}, \tilde{\mathbf{t}}} \tag{2.16}$$

subject to

$$\mathbf{C}\tilde{\mathbf{t}} - \mathbf{h} \geq \mathbf{0} \tag{2.17}$$

and

$$\mathbf{L} \leq \mathbf{D}_p(\tilde{\mathbf{t}})\boldsymbol{\alpha} \leq \mathbf{U} \tag{2.18}$$

is called *full constrained smoothing problem* (**FCSP**). The vectors, matrices, and matrix functions occurring have the following dimensions: $\mathbf{y} \in \mathbb{R}^m$; $\boldsymbol{\alpha} \in \mathbb{R}^n$; $\tilde{\mathbf{t}} \in \mathbb{R}^l$; $\mathbf{h} \in \mathbb{R}^{2l}$; $\mathbf{U}, \mathbf{L} \in \mathbb{R}^{n-p}$; $\mathbf{C} \in \mathbb{R}^{2l,l}$; $\mathbf{B}(.) : \tilde{\mathbf{t}} \in \mathbb{R}^l \to \mathbf{B}(\tilde{\mathbf{t}}) \in \mathbb{R}^{m,n}$; $\mathbf{S}_r(.) : \tilde{\mathbf{t}} \in \mathbb{R}^l \to \mathbf{S}_r(\tilde{\mathbf{t}}) \in \mathbb{R}^{n-r,n}$; $\mathbf{D}_p(.) : \tilde{\mathbf{t}} \in \mathbb{R}^l \to \mathbf{D}_p(\tilde{\mathbf{t}}) \in \mathbb{R}^{n-p,n}$.

3. Constrained semi-linear least squares problems

The problem **FCSP** is a nonlinear least squares problem where the variable $\boldsymbol{\alpha}$ occurs linearly. In this section we consider *general* problems of such type and begin with the

Full problem

$$\mathfrak{f}(\boldsymbol{\alpha}, \tilde{\mathbf{t}}) := \frac{1}{2}\left\|\mathfrak{F}(\tilde{\mathbf{t}})\right\|^2 = \frac{1}{2}\left\|\mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right\|^2 \longrightarrow \min_{\boldsymbol{\alpha} \in \mathbb{R}^n, \tilde{\mathbf{t}} \in \mathbb{R}^l} \tag{3.1}$$

subject to

$$\mathbf{C}\tilde{\mathbf{t}} - \mathbf{h} \geq \mathbf{0} \quad \text{and} \quad \begin{bmatrix} \mathbf{D}_p(\tilde{\mathbf{t}}) \\ -\mathbf{D}_p(\tilde{\mathbf{t}}) \end{bmatrix} \boldsymbol{\alpha} - \begin{pmatrix} \mathbf{L} \\ -\mathbf{U} \end{pmatrix} \geq \mathbf{0}. \tag{3.2}$$

Here $\mathbf{B}$ and $\mathbf{D}_p$ are arbitrary smooth matrix functions, and the remaining quantities $\mathbf{y}, \mathbf{h}, \mathbf{U}, \mathbf{L}$, and $\mathbf{C}$ are constant vectors and matrices.

If the variable $\tilde{\mathbf{t}}$ is fixed we obtain a linear least squares problem called *Subproblem (A)* whose solution shall be denoted by $\boldsymbol{\alpha}(\tilde{\mathbf{t}})$. By replacing the variable $\boldsymbol{\alpha}$ in the full problem by its optimal value $\boldsymbol{\alpha}(\tilde{\mathbf{t}})$ we obtain a *reduced problem* in the variable $\tilde{\mathbf{t}}$ only. This reduction technique is a generalization of the variable projection method of Golub/Pereyra for the unconstrained case. Following Parks [Par85], we call problems of this type *constrained semi-linear least squares problems* (**CSLS**), and define the related optimization problems

Reduced problem

$$f(\tilde{\mathbf{t}}) := \frac{1}{2}\left\|\mathbf{F}(\tilde{\mathbf{t}})\right\|^2 = \frac{1}{2}\left\|\mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}(\tilde{\mathbf{t}})\right\|^2 \longrightarrow \min_{\tilde{\mathbf{t}}\in\mathbb{R}^l} \tag{3.3}$$

subject to

$$\mathbf{C}\tilde{\mathbf{t}} - \mathbf{h} \geq \mathbf{0} \tag{3.4}$$

where $\mathbf{F}(\tilde{\mathbf{t}}) := \mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}(\tilde{\mathbf{t}})$, and $\boldsymbol{\alpha}(\tilde{\mathbf{t}})$ solves

Subproblem (A)

$$\frac{1}{2}\left\|\mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right\|^2 \longrightarrow \min_{\boldsymbol{\alpha}\in\mathbb{R}^n} \tag{3.5}$$

subject to

$$\begin{bmatrix} \mathbf{D}_p(\tilde{\mathbf{t}}) \\ -\mathbf{D}_p(\tilde{\mathbf{t}}) \end{bmatrix} \boldsymbol{\alpha} - \begin{pmatrix} \mathbf{L} \\ -\mathbf{U} \end{pmatrix} \geq \mathbf{0}. \tag{3.6}$$

Note that the full problem (3.1), (3.2) is not the general form of a **CSLS** since, e.g., the equality constraints are missing. For sake of simplicity of notation we restrict to this special case.

The following theorem [Par85, Theorem 4.7] shows under which conditions the transformation of the full problem into the reduced problem is feasible, and it explains the relationship between the solutions to these two problems. It asserts that the change from minimizing the full functional to minimizing the reduced functional does not add any critical points and does not exclude the solution of the original problem.

Theorem 3.1 (Correspondence between full and reduced problem).

1. *Let the function $\mathfrak{f}$ be twice continuously differentiable in $\boldsymbol{\alpha}$, and assume that its gradient with respect to $\boldsymbol{\alpha}$ is continuously differentiable in $\tilde{\mathbf{t}}$.*
2. *Let each of the constraints present in the problem be continuously differentiable in its arguments.*
3. *Assume that, for every $\tilde{\mathbf{t}}$, the subproblem (A) has a solution $\boldsymbol{\alpha}(\tilde{\mathbf{t}})$ such that*
   (a) *the second-order sufficiency conditions for a local minimizer of Subproblem (A) hold at $\boldsymbol{\alpha}(\tilde{\mathbf{t}})$ (with appropriate Lagrange multipliers),*
   (b) *the gradients (with respect to $\boldsymbol{\alpha}$) of those constraints of Subproblem (A) which are binding at $\boldsymbol{\alpha}(\tilde{\mathbf{t}})$ are linearly independent,*
   (c) *strict complementarity holds for Subproblem (A) at $\boldsymbol{\alpha}(\tilde{\mathbf{t}})$.*

Then the full and the reduced problem are related in the following way

(i) *Let $(\boldsymbol{\alpha}^*, \tilde{\mathbf{t}}^*)$ be a global minimizer of the full problem. Then $\boldsymbol{\alpha}^*$ satisfies the first-order conditions for Subproblem (A), $\tilde{\mathbf{t}}^*$ is a global minimizer of the reduced problem, and*

$$f(\tilde{\mathbf{t}}^*) = \mathfrak{f}(\boldsymbol{\alpha}^*, \tilde{\mathbf{t}}^*).$$

Furthermore, if there is a unique $\boldsymbol{\alpha}^*$ among the pairs $(\boldsymbol{\alpha}^*, \tilde{\mathbf{t}}^*)$ yielding the (same) minimal value of $\mathfrak{f}$, then

$$\boldsymbol{\alpha}^* = \boldsymbol{\alpha}\left(\tilde{\mathbf{t}}^*\right).$$

(ii) *Let $\tilde{\mathbf{t}}^*$ satisfy the first-order conditions for the reduced problem. Then the pair $\left(\boldsymbol{\alpha}(\tilde{\mathbf{t}}^*), \tilde{\mathbf{t}}^*\right)$ satisfies the first-order conditions for the full problem.*

The above theorem was proved by Parks for general nonlinear optimization problems of the form

$$\mathfrak{f}(\boldsymbol{\alpha},\tilde{\mathbf{t}}) \longrightarrow \min_{\boldsymbol{\alpha}\in\mathbb{R}^n,\tilde{\mathbf{t}}\in\mathbb{R}^l} \quad \text{subject to}$$

$$\begin{array}{lll} g_i(\boldsymbol{\alpha},\tilde{\mathbf{t}}) \geq 0,\ i=1,\dots,p_1, & c_i(\boldsymbol{\alpha}) \geq 0,\ i=1,\dots,p_3, & r_i(\tilde{\mathbf{t}}) \geq 0\ i=1,\dots,p_5, \\ h_i(\boldsymbol{\alpha},\tilde{\mathbf{t}}) = 0,\ i=1,\dots,p_2, & d_i(\boldsymbol{\alpha}) = 0,\ i=1,\dots,p_4, & s_i(\tilde{\mathbf{t}}) = 0\ i=1,\dots,p_6. \end{array}$$

The key in the proof of Theorem 3.1 lies in the application of the basic sensitivity theorem (first-order sensitivity analysis for a second order solution) of Fiacco [Fia76], see also [Fia83], to Subproblem (A). In fact, conditions 3(a)–(c) are essentially the assumptions of this sensitivity theorem.

Theorem 3.1 is of similar importance as the corresponding theorem for separable least squares problems (see [GP73, Theorem 2.1]) and is a direct generalization of this theorem. The case of semi-linear equality constraints $\mathbf{H}(\tilde{\mathbf{t}})\boldsymbol{\alpha} - \boldsymbol{\delta}(\tilde{\mathbf{t}}) = \mathbf{0}$ was considered in [KP78] and [Cor81].

For describing our solution method later on, we need a quantitative analysis of the reduced problem. The Lagrangian l to Subproblem (A) is

$$l(\boldsymbol{\alpha},\mathbf{u};\tilde{\mathbf{t}}) := \frac{1}{2}\left\|\mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right\|^2 - \mathbf{u}^T\mathbf{g}(\boldsymbol{\alpha};\tilde{\mathbf{t}})$$

with the Lagrange parameters $\mathbf{u} \in \mathbb{R}_+^{2(n-p)}$ and the vector of constraints

$$\mathbf{g} = \mathbf{g}(\boldsymbol{\alpha};\tilde{\mathbf{t}}) := \begin{bmatrix} \mathbf{D}_p(\tilde{\mathbf{t}}) \\ -\mathbf{D}_p(\tilde{\mathbf{t}}) \end{bmatrix}\boldsymbol{\alpha} - \begin{pmatrix} \mathbf{L} \\ -\mathbf{U} \end{pmatrix} \in \mathbb{R}^{2(n-p)}.$$

The signed gradients of these constraints are

$$\mathbf{R} := -\left(\nabla_{\boldsymbol{\alpha}}\mathbf{g}\right)^T = -\begin{bmatrix} \mathbf{D}_p(\tilde{\mathbf{t}}) \\ -\mathbf{D}_p(\tilde{\mathbf{t}}) \end{bmatrix} \in \mathbb{R}^{2(n-p),n}$$

$$\boldsymbol{\Gamma} := -\left(\nabla_{\tilde{\mathbf{t}}}\mathbf{g}\right)^T = -\left(\nabla_{\tilde{\mathbf{t}}}\begin{bmatrix} \mathbf{D}_p(\tilde{\mathbf{t}}) \\ -\mathbf{D}_p(\tilde{\mathbf{t}}) \end{bmatrix}\boldsymbol{\alpha}\right)^T \in \mathbb{R}^{2(n-p),l}.$$

Quantities related to binding constraints of Subproblem (A) shall be denoted by a bar, e.g., $\bar{\mathbf{R}} := -\left(\nabla_{\boldsymbol{\alpha}}\mathbf{g}_i\right)_{i\in\mathcal{I}}^T \in \mathbb{R}^{nactiv,n}$, $\bar{\boldsymbol{\Gamma}} := -\left(\nabla_{\tilde{\mathbf{t}}}\bar{\mathbf{g}}_i\right)_{i\in\mathcal{I}}^T \in \mathbb{R}^{nactiv,l}$, where $\mathcal{I} := \{i \in \{1,\dots,2(n-p)\} \mid g_i(\boldsymbol{\alpha};\tilde{\mathbf{t}}) = 0\}$ and $nactiv := \#\mathcal{I}$ are the index set and the number of active constraints, respectively.

Let $\boldsymbol{\partial} = \nabla_{\tilde{\mathbf{t}}}^T$ be the operator of Fréchet differentiation with respect to $\tilde{\mathbf{t}}$. Due to the regularity assumption 3(b) the matrix $\bar{\mathbf{R}}$ has full row rank $nactiv$. Let $\mathbf{N} \in \mathbb{R}^{n,n-nactiv}$ be a basis of the null space of $\bar{\mathbf{R}}$, and denote the Moore-Penrose-inverse of $\bar{\mathbf{R}}$ by $\bar{\mathbf{R}}^+$. Parks shows for general reducible nonlinear programming problems that

$$\begin{bmatrix} \nabla^2_{\boldsymbol{\alpha}} l & \bar{\mathbf{R}}^T \\ \bar{\mathbf{R}} & \mathbf{0} \end{bmatrix}\begin{pmatrix} \boldsymbol{\partial}\boldsymbol{\alpha} \\ \boldsymbol{\partial}\bar{\mathbf{u}} \end{pmatrix} = -\begin{pmatrix} \nabla^2_{\boldsymbol{\alpha}\tilde{\mathbf{t}}} l \\ \bar{\boldsymbol{\Gamma}} \end{pmatrix} \tag{3.7}$$

with

$$\begin{bmatrix} \nabla^2_{\boldsymbol{\alpha}} l & \bar{\mathbf{R}}^T \\ \bar{\mathbf{R}} & \mathbf{0} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{12}^T & \mathbf{W}_{22} \end{bmatrix} \tag{3.8}$$

and

$$\begin{aligned}
\mathbf{W}_{11} &= \mathbf{N}\left[\mathbf{N}^T\left(\nabla^2_{\boldsymbol{\alpha}} l\right)\mathbf{N}\right]^{-1}\mathbf{N}^T \in \mathbb{R}^{n,n}\\
\mathbf{W}_{12} &= \left[\mathbf{I} - \mathbf{W}_{11}\left(\nabla^2_{\boldsymbol{\alpha}} l\right)\right]\bar{\mathbf{R}}^+ \in \mathbb{R}^{n,nactiv}\\
\mathbf{W}_{22} &= -\mathbf{W}_{12}^T\left(\nabla^2_{\boldsymbol{\alpha}} l\right)\mathbf{W}_{12} \in \mathbb{R}^{nactiv,nactiv}.
\end{aligned}$$

For the Hessian of l we obtain in our special case

$$\begin{aligned}
\nabla^2_{\boldsymbol{\alpha}} l &= \mathbf{B}^T\mathbf{B} \in \mathbb{R}^{n,n} & \nabla^2_{\boldsymbol{\alpha}\tilde{\mathbf{t}}} l &= -\mathbf{B}^T\mathfrak{J}_{\tilde{\mathbf{t}}} + \mathbf{K} \in \mathbb{R}^{n,l}\\
\nabla^2_{\tilde{\mathbf{t}}\boldsymbol{\alpha}} l &= -\mathfrak{J}_{\tilde{\mathbf{t}}}^T\mathbf{B} + \mathbf{K}^T \in \mathbb{R}^{l,n} & \nabla^2_{\tilde{\mathbf{t}}} l &= \mathfrak{J}_{\tilde{\mathbf{t}}}^T\mathfrak{J}_{\tilde{\mathbf{t}}} + \mathbf{S}_{\tilde{\mathbf{t}}} \in \mathbb{R}^{l,l}
\end{aligned}$$

with

$$\begin{aligned}
\mathfrak{J}_{\tilde{\mathbf{t}}} &:= \partial\mathfrak{F} = -\partial\mathbf{B}\boldsymbol{\alpha} \in \mathbb{R}^{m,l}\\
\mathbf{K} &:= -\partial\mathbf{B}^T\left(\mathbf{y} - \mathbf{B}\boldsymbol{\alpha}\right) + \partial\mathbf{R}^T\mathbf{u} \in \mathbb{R}^{n,l}\\
\mathbf{S}_{\tilde{\mathbf{t}}} &:= \mathbf{u}^T\partial^2\mathbf{R}\boldsymbol{\alpha} - \left(\partial^2\mathbf{B}\boldsymbol{\alpha}\right)^T\left(\mathbf{y} - \mathbf{B}\boldsymbol{\alpha}\right) \in \mathbb{R}^{l,l}.
\end{aligned}$$

Note that the term $\left(\partial^2\mathbf{B}\boldsymbol{\alpha}\right)^T\left(\mathbf{y} - \mathbf{B}\boldsymbol{\alpha}\right)$ in $\mathbf{S}_{\tilde{\mathbf{t}}}$ is erroneously missing in [Par85]. But this does not affect further results, since $\mathbf{S}_{\tilde{\mathbf{t}}}$ as the only term which contains second derivatives is anyway dropped later on.

From (3.7) and (3.8) we have

$$\begin{aligned}
\partial\boldsymbol{\alpha} &= -\mathbf{W}_{11}\left(\nabla^2_{\boldsymbol{\alpha}\tilde{\mathbf{t}}} l\right) - \mathbf{W}_{12}\bar{\mathbf{\Gamma}}\\
&= \mathbf{W}_{11}\mathbf{B}^T\mathfrak{J}_{\tilde{\mathbf{t}}} - \mathbf{W}_{11}\mathbf{K} - \left(\mathbf{I} - \mathbf{W}_{11}\mathbf{B}^T\mathbf{B}\right)\bar{\mathbf{R}}^+\bar{\mathbf{\Gamma}}.
\end{aligned}$$

Consider the orthogonal projectors $\mathbf{P}_{BN} := (\mathbf{BN})(\mathbf{BN})^+ \in \mathbb{R}^{m,m}$ and $\mathbf{P}_{BN}^\perp := \mathbf{I} - \mathbf{P}_{BN} \in \mathbb{R}^{m,m}$. Obviously it holds $\mathbf{B}\mathbf{W}_{11}\mathbf{B}^T = \mathbf{P}_{BN}$, $\mathbf{B}\left(\mathbf{I} - \mathbf{W}_{11}\mathbf{B}^T\mathbf{B}\right) = \mathbf{P}_{BN}^\perp\mathbf{B}$, and $\mathbf{B}\mathbf{W}_{11}\mathbf{K} = \left[(\mathbf{BN})^+\right]^T\mathbf{N}^T\mathbf{K}$. For the Jacobian of the reduced functional we get $\mathbf{J}(\tilde{\mathbf{t}}) := \partial\mathbf{F}(\tilde{\mathbf{t}}) = \partial(\mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}(\tilde{\mathbf{t}})) = \mathfrak{J}_{\tilde{\mathbf{t}}} - \mathbf{B}\partial\boldsymbol{\alpha}$. Substituting $\mathbf{B}\partial\boldsymbol{\alpha} = \mathbf{P}_{BN}\mathfrak{J}_{\tilde{\mathbf{t}}} - \left[(\mathbf{BN})^+\right]^T\mathbf{N}^T\mathbf{K} - \mathbf{P}_{BN}^\perp\mathbf{B}\bar{\mathbf{R}}^+\bar{\mathbf{\Gamma}}$ we obtain

Lemma 3.1 (Jacobian of the reduced functional, [Par85, Lemma 6.2]**).** *The Jacobian* $\mathbf{J}(\tilde{\mathbf{t}}) = \partial\mathbf{F}(\tilde{\mathbf{t}})$ *of* $\mathbf{F}(\tilde{\mathbf{t}}) = \mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}(\tilde{\mathbf{t}})$ *is given by*

$$\mathbf{J}(\tilde{\mathbf{t}}) = \mathbf{P}_{BN}^\perp(\tilde{\mathbf{t}})\Big(\mathfrak{J}_{\tilde{\mathbf{t}}}(\tilde{\mathbf{t}}) + \mathbf{B}(\tilde{\mathbf{t}})\bar{\mathbf{R}}^+(\tilde{\mathbf{t}})\bar{\mathbf{\Gamma}}(\tilde{\mathbf{t}})\Big) + \mathbf{P}_{BN}(\tilde{\mathbf{t}})\left[\left(\mathbf{B}(\tilde{\mathbf{t}})\mathbf{N}(\tilde{\mathbf{t}})\right)^+\right]^T\mathbf{N}^T(\tilde{\mathbf{t}})\mathbf{K}(\tilde{\mathbf{t}})$$

where $\mathbf{K}(\tilde{\mathbf{t}}) = \mathbf{K}(\boldsymbol{\alpha}(\tilde{\mathbf{t}}), \tilde{\mathbf{t}})$ *and* $\mathbf{K}(\boldsymbol{\alpha}, \tilde{\mathbf{t}}) := -\partial\mathbf{B}^T(\tilde{\mathbf{t}})\left(\mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right) + \partial\mathbf{R}^T(\tilde{\mathbf{t}})\mathbf{u}$.

Moreover, the residual of the reduced functional can be expressed as follows

Lemma 3.2 (Residual of the reduced functional, [Par85, Lemma 6.1]**).**

$$\mathbf{F}(\tilde{\mathbf{t}}) = \mathbf{P}_{BN}^\perp(\tilde{\mathbf{t}})\Big(\mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\bar{\mathbf{R}}^+(\tilde{\mathbf{t}})\bar{\boldsymbol{\xi}}\Big) \quad \textit{with} \quad \bar{\boldsymbol{\xi}} := -\begin{pmatrix}\mathbf{L}\\ -\mathbf{U}\end{pmatrix}_{i\in\mathcal{I}} \in \mathbb{R}^{nactiv}.$$

To clarify the structure of the Jacobian and the residual, we rewrite them by defining

$$\boldsymbol{\psi} := \mathfrak{J}_{\tilde{\mathbf{t}}} + \mathbf{B}\bar{\mathbf{R}}^+\bar{\mathbf{\Gamma}}, \qquad \boldsymbol{\phi} := \Big((\mathbf{BN})^+\Big)^T\mathbf{N}^T\mathbf{K}, \qquad \mathbf{v} := \mathbf{y} - \mathbf{B}\bar{\mathbf{R}}^+\bar{\boldsymbol{\xi}}.$$

With $\mathbf{P} = \mathbf{P}_{BN}$, Lemma 3.1 and 3.2 read as $\mathbf{F} = \mathbf{P}^\perp\mathbf{v}$, $\mathbf{J} = \mathbf{P}^\perp\boldsymbol{\psi} + \mathbf{P}\boldsymbol{\phi}$. The occurrence of the term $\mathbf{P}\boldsymbol{\phi} = \mathbf{P}_{BN}\Big((\mathbf{BN})^+\Big)^T\mathbf{N}^T\mathbf{K}$ complicates the computation of the Jacobian, especially any exploitation of sparsity. Fortunately, it holds

$$\mathbf{J}^T\mathbf{F} = \boldsymbol{\psi}^T\left(\mathbf{P}^\perp\right)^T\mathbf{P}^\perp\mathbf{v}$$

and

$$\mathbf{J}^T\mathbf{J} = \boldsymbol{\psi}^T \left(\mathbf{P}^{\perp}\right)^T \mathbf{P}^{\perp}\boldsymbol{\psi} + \boldsymbol{\phi}^T\mathbf{P}^T\mathbf{P}\boldsymbol{\phi},$$

i.e., the term $\mathbf{P}\boldsymbol{\phi}$ does not contribute to $\mathbf{J}^T\mathbf{F}$, and its contribution to $\mathbf{J}^T\mathbf{J}$ is $\boldsymbol{\phi}^T\mathbf{P}^T\mathbf{P}\boldsymbol{\phi}$ as Parks recognized in her Thesis [Par85]. In the unconstrained case, these properties were first observed by Kaufman. Moreover, in the unconstrained case there even holds $\boldsymbol{\phi}^T\mathbf{P}^T\mathbf{P}\boldsymbol{\phi} = \mathcal{O}(\|\mathbf{y}-\mathbf{B}\boldsymbol{\alpha}\|^2)$ so that this term can be neglected within the framework of Gauss-Newton methods what Kaufman did in her paper [Kau75]. A direct derivation of this property and the proof that the fast convergence for small residual problems is retained can be found in [RW80], see also [Sch91] and [Sch92].

Generalizing this philosophy to the constrained case, we define

Definition 1 (Kaufman-Approximation)**.** The approximation

$$\mathbf{J}_K := \mathbf{P}^{\perp}\boldsymbol{\psi} = \mathbf{P}^{\perp}\left(\mathbf{J}_{\tilde{\mathbf{t}}} + \mathbf{B}\bar{\mathbf{R}}^{+}\bar{\mathbf{\Gamma}}\right) \in \mathbb{R}^{m,l}$$

to the Jacobian

$$\mathbf{J} = \mathbf{J}^{\perp}\boldsymbol{\psi} + \mathbf{P}\boldsymbol{\phi} = \mathbf{P}^{\perp}_{BN}\left(\mathbf{J}_{\tilde{\mathbf{t}}} + \mathbf{B}\bar{\mathbf{R}}^{+}\bar{\mathbf{\Gamma}}\right) + \mathbf{P}^{\perp}_{BN}\left((\mathbf{BN})^{+}\right)^T\mathbf{N}^T\mathbf{K}$$

of the reduced functional is called *Kaufman-Approximation.*

Let us remark that in the constrained case considered here the term $\boldsymbol{\phi}^T\mathbf{P}^T\mathbf{P}\boldsymbol{\phi}$ contains besides a part of order $\mathcal{O}(\|\mathbf{y}-\mathbf{B}\boldsymbol{\alpha}\|^2)$ a further part of order $\mathcal{O}(\|\boldsymbol{\partial}\mathbf{R}^T\mathbf{u}\|^2)$. It can be shown, however, that the Lagrange multipliers $\mathbf{u}$ are of order $\mathcal{O}(\|\mathbf{y}-\mathbf{B}\boldsymbol{\alpha}\|)$ if strict complementarity holds. Hence, under this assumption, we have $\mathbf{J}_K^T\mathbf{J}_K = \mathbf{J}^T\mathbf{J} + \mathcal{O}(\|\mathbf{y}-\mathbf{B}\boldsymbol{\alpha}\|^2)$ and, therefore, qualitatively the same local convergence properties as in the unconstrained case, see [SS96] for details.

4. Smoothing by Splines with Free Knots under Constraints on Derivatives

We can immediately apply the results of Section 3 to the full constrained smoothing problem **FCSP** if we substitute the pair $\{\mathbf{B}(\tilde{\mathbf{t}}), \mathbf{y}\}$ by the quantities

$$\left\{ \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix}, \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} \right\}.$$

The corresponding reduced problem is called *reduced constrained smoothing problem* (**RCSP**) and defined by

$$f(\tilde{\mathbf{t}}) := \frac{1}{2}\|\mathbf{F}(\tilde{\mathbf{t}})\|^2 = \frac{1}{2}\left\| \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} - \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix} \boldsymbol{\alpha}(\tilde{\mathbf{t}}) \right\|^2 \longrightarrow \min_{\tilde{\mathbf{t}}\in\mathbb{R}^l} \tag{4.1}$$

subject to

$$\mathbf{C}\tilde{\mathbf{t}} - \mathbf{h} \geq \mathbf{0}, \tag{4.2}$$

where $\mathbf{F}(\tilde{\mathbf{t}}) := \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} - \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix} \boldsymbol{\alpha}(\tilde{\mathbf{t}})$, and $\boldsymbol{\alpha}(\tilde{\mathbf{t}})$ solves *Subproblem (A)*

$$\frac{1}{2}\left\| \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} - \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix} \boldsymbol{\alpha} \right\|^2 \longrightarrow \min_{\boldsymbol{\alpha}\in\mathbb{R}^n} \tag{4.3}$$

subject to

$$\mathbf{L} \leq \mathbf{D}_p(\tilde{\mathbf{t}})\boldsymbol{\alpha} \leq \mathbf{U}. \tag{4.4}$$

In this section we show that problem **RCSP** has always a solution, and we investigate the correspondence between solutions of **FCSP** and **RCSP**. To prove the existence of solutions to the reduced problem we need a weaker perturbation theorem than the basic sensitivity theorem of Fiacco. One can show, see [Dan73], that for the solution of the quadratic programs

$$\mathbf{x}^* = \operatorname{argmin}\left\{\mathbf{b}^T\mathbf{x} + \tfrac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} \,:\, \mathbf{H}^T\mathbf{x} + \mathbf{h}^0 = \mathbf{0}, \mathbf{G}^T\mathbf{x} + \mathbf{g}^0 \geq \mathbf{0}, \mathbf{x} \in \mathbb{R}^n\right\}$$
$$\tilde{\mathbf{x}}^* = \operatorname{argmin}\left\{\tilde{\mathbf{b}}^T\mathbf{x} + \tfrac{1}{2}\mathbf{x}^T\tilde{\mathbf{A}}\mathbf{x} \,:\, \tilde{\mathbf{H}}^T\mathbf{x} + \tilde{\mathbf{h}}^0 = \mathbf{0}, \tilde{\mathbf{G}}^T\mathbf{x} + \tilde{\mathbf{g}}^0 \geq \mathbf{0}, \mathbf{x} \in \mathbb{R}^n\right\}$$

the relation

$$\|\tilde{\mathbf{x}}^* - \mathbf{x}^*\| \leq C \max\left\{\|\mathbf{A} - \tilde{\mathbf{A}}\|, \|\mathbf{H} - \tilde{\mathbf{H}}\|, \|\mathbf{G} - \tilde{\mathbf{G}}\|, \|\mathbf{b} - \tilde{\mathbf{b}}\|, \|\mathbf{h}^0 - \tilde{\mathbf{h}}^0\|, \|\mathbf{g}^0 - \tilde{\mathbf{g}}^0\|\right\}$$

holds, whenever the perturbations $\|\mathbf{A} - \tilde{\mathbf{A}}\|, \ldots$ are sufficiently small ($\mathbf{A}$ symmetric, positive definite, $\mathbf{H}$ full column rank, and $\exists\, \mathbf{x}^0 : \mathbf{G}^T\mathbf{x}^0 + \mathbf{g}^0 > \mathbf{0}$ (Slater condition)).

Proposition 4.1 (Linear independence of gradients of constraints). *Let $t_j < t_{j+k-p}$ $(j = p+1, \ldots, n)$, and let*

$$\mathbf{g} = \mathbf{g}(\boldsymbol{\alpha}; \tilde{\mathbf{t}}) := \begin{bmatrix} \mathbf{D}_p(\tilde{\mathbf{t}}) \\ -\mathbf{D}_p(\tilde{\mathbf{t}}) \end{bmatrix} \boldsymbol{\alpha} - \begin{pmatrix} \mathbf{L} \\ -\mathbf{U} \end{pmatrix} \geq \mathbf{0}$$

be the constraints of Subproblem (A). Then the gradients of active constraints are linearly independent, i.e.,

$$\bar{\mathbf{R}} := -\left(\nabla_{\boldsymbol{\alpha}} \mathbf{g}_i^T\right)_{i \in \mathcal{I}} = -\begin{bmatrix} \mathbf{D}_p(\tilde{\mathbf{t}}) \\ -\mathbf{D}_p(\tilde{\mathbf{t}}) \end{bmatrix}_{i \in \mathcal{I}} \in \mathbb{R}^{nactiv, n}$$

has $\operatorname{rank} \bar{\mathbf{R}} = nactiv = \#\mathcal{I}$ *if and only if the* strict consistency condition $L_i < U_i$ $(i = 1, \ldots, n-p)$ $(\mathbf{L} < \mathbf{U})$ *holds.*

Proof. In a first step we consider the constraints $\mathbf{g} = \mathbf{D}_p(\tilde{\mathbf{t}})\boldsymbol{\alpha} - \mathbf{L} \geq \mathbf{0}$. The gradients of active constraints are linearly independent if *all* rows of $\mathbf{D}_p$ are linearly independent, i.e., $\operatorname{rank} \mathbf{D}_p = n - p$. Since $\mathbf{D}_p$ is upper triangular, it holds

$$\operatorname{rank} \mathbf{D}_p = n - p \quad \Longleftrightarrow \quad (\mathbf{D}_p)_{ii} \neq 0 \quad (i = 1, \ldots, n-p).$$

By induction one can prove

$$(\mathbf{D}_p)_{ii} = (-1)^p \prod_{\nu=1}^{p} (k - \nu) \frac{1}{t_{k+i} - t_{i+\nu}} \quad \text{for} \quad i = 1, \ldots, n-p \quad \text{and} \quad p \geq 1.$$

Therefore, we have that the gradients of the active parts to the constraints to $\mathbf{g} = \mathbf{D}_p(\tilde{\mathbf{t}})\boldsymbol{\alpha} - \mathbf{L} \geq \mathbf{0}$ are linearly independent if $t_j < t_{j+k-p}$ $(j = p+1, \ldots, n)$.

Let us now consider the original constraints. Obviously, if $L_i < U_i$, then a constraint can become active either at L_i or at U_i, but never simultaneously. Thus, the matrix $\bar{\mathbf{R}}$ does not contain two identical rows of $\mathbf{D}_p$ (except a factor of -1). Hence, the rows of $\bar{\mathbf{R}}$ are linear independent if $t_j < t_{j+k-p}$ $(j = p+1, \ldots, n)$. Conversely, if $L_i = U_i$, the matrix $\bar{\mathbf{R}}$ contains two identical rows of $\mathbf{D}_p$ (except a factor of -1), i.e., $\bar{\mathbf{R}}$ does not have full rank. ☐

Remark 1. The condition $l_i^{(p)} < u_i^{(p)}$ for the constraints $l_i^{(p)} \leq s^{(p)}(x) \leq u_i^{(p)}$ for all $x \in [t_i, t_{i+1})$ $(i = k, \ldots, n)$ is *not sufficient* for strict consistency, see Examples 2.1–2.3. In our case the strict consistency condition is equivalent to the Slater condition, i.e., to the existence of $\boldsymbol{\alpha}^0 \in \mathbb{R}^n$ with $\mathbf{g}(\boldsymbol{\alpha}^0, \tilde{\mathbf{t}}) > \mathbf{0}$.

Theorem 4.1 (Existence of solutions to the reduced problem). *Consider the set of feasible knot sequences $\left\{\tilde{\mathbf{t}} \in \mathbb{R}^l \,:\, \mathbf{C}\tilde{\mathbf{t}} - \mathbf{h} \geq \mathbf{0}\right\}$, and let the following conditions be fulfilled for fixed $p, r \in \{0, \ldots, q\}$, $q < k$:*

(C1) *The knots satisfy $t_j < t_{j+k-q}$ $(j = q+1, \ldots, n)$.*
(C3) *The regularity condition $m \geq r$ and $\mu > 0$ is met.*

(C4) *The shape constraints* $\mathbf{L} \le \mathbf{D}_p(\tilde{\mathbf{t}})\boldsymbol{\alpha} \le \mathbf{U}$ *satisfy the strict consistency condition* $\mathbf{L} < \mathbf{U}$.

Then the reduced constrained smoothing problem ***RCSP*** *(4.1), (4.2), (4.3), (4.4) has a solution* $\tilde{\mathbf{t}}^*$.

Proof. Condition (C1) implies existence of the matrices for all feasible knot sequences. Moreover, the matrix functions $\mathbf{B}(.)$, $\mathbf{S}_r(.)$, and $\mathbf{D}_p(.)$ are continuous functions of the knots. Due to Lemma 2.2, the regularity condition (C3) assures the full rank property of the system matrix $\mathbf{B}_\mu$ independent of the position of knots, i.e., the Hessian of Subproblem (A) is positive definite. Finally, the strict consistency condition (C4) yields the existence of a parameter $\boldsymbol{\alpha}^0$ so that $\mathbf{L} < \mathbf{D}_p(\tilde{\mathbf{t}})\boldsymbol{\alpha}^0 < \mathbf{U}$, i.e., the Slater condition on the shape constraints is met, see Remark 1. With

$$\mathbf{A}(\tilde{\mathbf{t}}) := \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix}^T \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix}, \qquad \mathbf{b}(\tilde{\mathbf{t}}) := -\begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix} \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix},$$

$$\mathbf{G}(\tilde{\mathbf{t}}) := \begin{bmatrix} \mathbf{D}_p(\tilde{\mathbf{t}}) \\ -\mathbf{D}_p(\tilde{\mathbf{t}}) \end{bmatrix}^T, \qquad \mathbf{g}^0 := -\begin{pmatrix} \mathbf{L} \\ -\mathbf{U} \end{pmatrix},$$

Subproblem (A) is equivalent to

$$\underset{\boldsymbol{\alpha}\in\mathbb{R}^n}{\text{minimize}} \;\; \mathbf{b}(\tilde{\mathbf{t}})^T\boldsymbol{\alpha} + \frac{1}{2}\boldsymbol{\alpha}^T\mathbf{A}(\tilde{\mathbf{t}})\boldsymbol{\alpha} \quad \text{s.t.} \quad \mathbf{G}(\tilde{\mathbf{t}})^T\boldsymbol{\alpha} + \mathbf{g}^0 \ge \mathbf{0}.$$

We obtain a perturbed quadratic optimization problem by replacing the parameter $\tilde{\mathbf{t}}$ by $\tilde{\mathbf{t}} + \delta\tilde{\mathbf{t}}$. Because of the continuity of the matrix functions $\mathbf{B}(.)$, $\mathbf{S}_r(.)$, and $\mathbf{D}_p(.)$ the variation of $\mathbf{b}$, $\mathbf{A}$, $\mathbf{G}$, and $\mathbf{g}^0$ is small for small perturbations $\delta\tilde{\mathbf{t}}$ of the parameter. Thus, we can apply the perturbation theorem for quadratic programs [Dan73]. It shows that the solution $\boldsymbol{\alpha}(.)$ of Subproblem (A) is a Lipschitz-continuous function of the parameter $\tilde{\mathbf{t}}$, i.e., the reduced functional f is itself Lipschitz-continuous. Therefore, the continuous functional f attains its minimum on the closed (because of $\mathbf{C}\tilde{\mathbf{t}} - \mathbf{h} \ge \mathbf{0}$) and bounded (because of $a \le t_{p(1)}$, $t_{p(l)} \le b$) set of feasible knot sequences $\{\tilde{\mathbf{t}} \in \mathbb{R}^l : \mathbf{C}\tilde{\mathbf{t}} - \mathbf{h} \ge \mathbf{0}\}$. □

Now we show the correspondence between the original problem **FCSP** and the reduced problem **RCSP**. For doing so we need a differentiable dependence of the matrix functions on the knots. From the definition of $\mathbf{D}_p(.)$ and $\mathbf{S}_r(.)$ it is clear that these matrix functions are continuously differentiable with respect to the knots whenever $t_j < t_{j+k-q}$ $(j = q+1, \ldots, n)$, $p, r \in \{0, \ldots, q\}$. For the B-spline matrix $\mathbf{B}$ we have

Lemma 4.2 (Derivative of a spline with respect to the knots).
Let $s(x) = \sum_{j=1}^n B_{j,k,\mathbf{t}}(x)\alpha_j$ *be a spline of order* $k \ge 3$ *with knot sequence* $\mathbf{t} \in \mathbb{R}^{n+k}$ *where* $t_1 = \cdots = t_k = a < t_{k+1} \le \cdots \le t_n < b = t_{n+1} = \cdots = t_{n+k}$. *Let* t_{j_0} *be a knot with multiplicity* $\#t_{j_0} = 1$ *and* $k < j_0 < n+1$. *Then the derivative of* s *with respect to* t_{j_0} *exists for all* $x \in [a, b]$ *and it holds*

$$\frac{\partial s(x)}{\partial t_{j_0}} = \sum_{j=j_0-k+1}^{j_0} \frac{\alpha_{j-1} - \alpha_j}{t'_{j+k} - t'_j} \times B_{j,k,\mathbf{t}'}(x)$$

with the knot sequence $\mathbf{t}' = (t_1, \ldots, t_{j_0}, t_{j_0}, \ldots t_{n+k})^T \in \mathbb{R}^{n+k+1}$, *i.e.,*

$$\begin{aligned} t'_j &= t_j & j &= 1, \ldots, j_0 \\ t'_j &= t_{j-1} & j &= j_0+1, \ldots, n+k+1. \end{aligned}$$

Theorem 4.2 (Correspondence between original and reduced problem). *Let $\tilde{\mathbf{t}}^*$ be a feasible knot sequence, i.e., $\tilde{\mathbf{t}}^* \in \{\tilde{\mathbf{t}} \in \mathbb{R}^l : \mathbf{C}\tilde{\mathbf{t}} - \mathbf{h} \geq \mathbf{0}\}$. Moreover, let the following conditions be fulfilled for fixed $p, r \in \{0, \ldots, q\}$, $q < k$:*

(C1) *The knots satisfy $t_j < t_{j+k-q}$ $(j = q+1, \ldots, n)$.*

(C2) *The free knots $\tilde{\mathbf{t}}^*$ are simple knots, i.e., $\#t^*_{p(j)} = 1$ $(j = 1, \ldots, l)$, and it holds $k \geq 3$.*

(C3) *The regularity condition $m \geq r$ and $\mu > 0$ is met.*

(C4) *The shape constraints $\mathbf{L} \leq \mathbf{D}_p(\tilde{\mathbf{t}})\boldsymbol{\alpha} \leq \mathbf{U}$ satisfy the strict consistency condition $\mathbf{L} < \mathbf{U}$.*

(C5) *The Lagrange parameters $\mathbf{u}^*$ of Subproblem (A) at $\boldsymbol{\alpha}(\tilde{\mathbf{t}}^*)$ are strictly complementary.*

*Then the full constrained smoothing problem **FCSP** (2.16), (2.17), (2.18) and the reduced constrained smoothing problem **RCSP** (4.1), (4.2), (4.3), (4.4) are related in the following way:*

(i) *If $(\boldsymbol{\alpha}^*, \tilde{\mathbf{t}}^*)$ is a global minimizer of problem **FCSP**, then $\boldsymbol{\alpha}^*$ satisfies the necessary (and for quadratic definite problems also sufficient) first order optimality conditions for Subproblem (A), $\tilde{\mathbf{t}}^*$ is a global minimizer of the reduced problem **RCSP**, and $f(\tilde{\mathbf{t}}^*) = \mathfrak{f}(\boldsymbol{\alpha}^*, \tilde{\mathbf{t}}^*)$, $\boldsymbol{\alpha}^* = \boldsymbol{\alpha}(\tilde{\mathbf{t}}^*)$.*

(ii) *If $\tilde{\mathbf{t}}^*$ satisfies the necessary first order optimality conditions for the reduced problem **RCSP**, then $(\boldsymbol{\alpha}(\tilde{\mathbf{t}}^*), \tilde{\mathbf{t}}^*)$ satisfies the necessary first order optimality conditions for the original problem **RCSP**.*

Proof. Let $\tilde{\mathbf{t}}^*$ be a feasible knot sequence, i.e., $\mathbf{C}\tilde{\mathbf{t}}^* - \mathbf{h} \geq \mathbf{0}$.

1. Differentiability of the problem functions

From conditions (C1) and (C2) it follows that

$$\mathfrak{f}(\boldsymbol{\alpha}, \tilde{\mathbf{t}}) = \frac{1}{2} \left\| \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} - \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix} \boldsymbol{\alpha} \right\|^2$$

is twice continuously differentiable with respect to $\boldsymbol{\alpha}$, and that the gradient

$$\nabla_{\boldsymbol{\alpha}} \mathfrak{f}(\boldsymbol{\alpha}, \tilde{\mathbf{t}}) = - \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix}^T \left(\begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} - \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix} \boldsymbol{\alpha} \right)$$

is continuously differentiable with respect to $\tilde{\mathbf{t}}$ in a neighborhood $\mathcal{U}_1^*$ of $\tilde{\mathbf{t}}^*$.

2. Differentiability of the constraints

The knot constraints $\mathbf{C}\tilde{\mathbf{t}} - \mathbf{h} \geq \mathbf{0}$ are continuously differentiable in $\tilde{\mathbf{t}}$. From (C1) and (C2) we have that the shape constraints $\mathbf{L} \leq \mathbf{D}_p(\tilde{\mathbf{t}})\boldsymbol{\alpha} \leq \mathbf{U}$ are continuously differentiable with respect to $\boldsymbol{\alpha}$ and $\tilde{\mathbf{t}}$ in a neighborhood $\mathcal{U}_2^*$ of $\tilde{\mathbf{t}}^*$.

3. Conditions on Subproblem (A)

(a) Under condition (C3), the system matrix $\mathbf{B}_\mu$ of Subproblem (A) has full rank n for all feasible knot sequences $\tilde{\mathbf{t}}$. Hence, the quadratic optimization problem has a positive definite Hessian and a unique solution $\boldsymbol{\alpha}(\tilde{\mathbf{t}})$ on the non-empty feasible set $\mathbf{L} \leq \mathbf{D}_p(\tilde{\mathbf{t}})\boldsymbol{\alpha} \leq \mathbf{U}$. The second-order sufficiency conditions hold at $\boldsymbol{\alpha}(\tilde{\mathbf{t}})$.

(b) Proposition 4.1 states that the gradients of the constraints of Subproblem (A) are linearly independent if condition (C4) holds.

(c) If the unique Lagrange multipliers $\mathbf{u}^*$ of Subproblem (A) at $\tilde{\mathbf{t}}^*$ are strictly complementary, cf. (C5), then there exists a neighborhood $\mathcal{U}_4^*$ of $\tilde{\mathbf{t}}^*$ such that the Lagrange multipliers $\mathbf{u}$ of Subproblem (A) at $\tilde{\mathbf{t}} \in \mathcal{U}_4^*$ (respectively $\boldsymbol{\alpha}(\tilde{\mathbf{t}})$) are strictly complementary, too.

In the non-empty intersection of the sets $\mathcal{U}_1^*, \ldots, \mathcal{U}_4^*$ all conditions of Theorem 3.1 are satisfied. The statements immediately follow by applying Theorem 3.1. Furthermore, we obtain for part (i):

original problem $(\boldsymbol{\alpha}^*, \tilde{\mathbf{t}}^*)$ global minimizer	$\xrightarrow[\text{(i)}]{\text{Theorem 4.2}}$	**reduced problem** $\tilde{\mathbf{t}}^*$ global minimizer **Subproblem (A)** $\boldsymbol{\alpha}^*$ satisfies necessary (and sufficient) optimality condition
original problem $(\boldsymbol{\alpha}(\tilde{\mathbf{t}}^*), \tilde{\mathbf{t}}^*)$ satisfies necessary optimality condition	$\xleftarrow[\text{(ii)}]{\text{Theorem 4.2}}$	**reduced problem** $\tilde{\mathbf{t}}^*$ satisfies necessary optimality condition
original problem $(\boldsymbol{\alpha}(\tilde{\mathbf{t}}^*), \tilde{\mathbf{t}}^*)$ global minimizer for $\tilde{\mathbf{t}} \in \Omega$	$\xleftarrow{\text{Corollary 4.3}}$	**reduced problem** $\tilde{\mathbf{t}}^*$ global minimizer in Ω

FIGURE 1. Correspondence between original and reduced problem

Since Subproblem (A) is a quadratic, definite optimization problem, there exists a unique $\boldsymbol{\alpha}^*$ among all pairs $(\boldsymbol{\alpha}^*, \tilde{\mathbf{t}}^*)$ that minimize $\mathfrak{f}$ and yield the same minimal value, and there holds $\boldsymbol{\alpha}^* = \boldsymbol{\alpha}(\tilde{\mathbf{t}}^*)$. ☐

Corollary 4.3.
Let the conditions of Theorem 4.2 be satisfied in a neighborhood Ω of the feasible knot sequence $\tilde{\mathbf{t}}^*$. If $\tilde{\mathbf{t}}^*$ is a global minimizer of the reduced functional $f(\tilde{\mathbf{t}})$ in Ω, then $(\boldsymbol{\alpha}^*, \tilde{\mathbf{t}}^*)$ is a global minimizer of $\mathfrak{f}(\boldsymbol{\alpha}, \tilde{\mathbf{t}})$ for $\tilde{\mathbf{t}} \in \Omega$.

Proof. Let $\tilde{\mathbf{t}}^*$ be a global minimizer of $f(\tilde{\mathbf{t}})$ in Ω, and let $\boldsymbol{\alpha}(\tilde{\mathbf{t}}^*)$ be the corresponding unique solution to Subproblem (A). Obviously, there holds $\mathfrak{f}\left(\boldsymbol{\alpha}(\tilde{\mathbf{t}}^*), \tilde{\mathbf{t}}^*\right) = f(\tilde{\mathbf{t}}^*)$. Suppose there exist a knot sequence $\tilde{\mathbf{t}}^\dagger \in \Omega$ and coefficients $\boldsymbol{\alpha}^\dagger$ with $\mathfrak{f}\left(\boldsymbol{\alpha}^\dagger, \tilde{\mathbf{t}}^\dagger\right) < \mathfrak{f}\left(\boldsymbol{\alpha}(\tilde{\mathbf{t}}^*), \tilde{\mathbf{t}}^*\right)$. From the definition of the reduced problem we have $\mathfrak{f}(\boldsymbol{\alpha}, \tilde{\mathbf{t}}) \geq f(\tilde{\mathbf{t}})$ for all $\tilde{\mathbf{t}} \in \Omega$, because of $\mathfrak{f}(\boldsymbol{\alpha}, \tilde{\mathbf{t}}) := \frac{1}{2}\left\|\mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right\|^2 + \frac{1}{2}\mu\left\|\mathbf{S}_r(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right\|^2$ with *arbitrary* $\boldsymbol{\alpha}$ and $f(\tilde{\mathbf{t}}) := \frac{1}{2}\left\|\mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}(\tilde{\mathbf{t}})\right\|^2 + \frac{1}{2}\mu\left\|\mathbf{S}_r(\tilde{\mathbf{t}})\boldsymbol{\alpha}(\tilde{\mathbf{t}})\right\|^2$ with an *optimal* $\boldsymbol{\alpha}$.

Therefore, $f(\tilde{\mathbf{t}}^\dagger) \leq \mathfrak{f}\left(\boldsymbol{\alpha}^\dagger, \tilde{\mathbf{t}}^\dagger\right) < \mathfrak{f}\left(\boldsymbol{\alpha}(\tilde{\mathbf{t}}^*), \tilde{\mathbf{t}}^*\right) = f(\tilde{\mathbf{t}}^*)$, which is a contradiction to the assumption that $\tilde{\mathbf{t}}^*$ is a global minimizer of $f(\tilde{\mathbf{t}})$ in Ω. Hence, $\left(\boldsymbol{\alpha}(\tilde{\mathbf{t}}^*), \tilde{\mathbf{t}}^*\right)$ is a global minimizer of $\mathfrak{f}(\boldsymbol{\alpha}, \tilde{\mathbf{t}})$ for $\tilde{\mathbf{t}} \in \Omega$. ☐

Figure 1 illustrates the correspondence between original and reduced problem.

5. NUMERICAL SOLUTION OF THE REDUCED CONSTRAINED SMOOTHING PROBLEM

After having proved the equivalence of the original and the reduced problem in the sense of Theorem 4.2, we will treat the efficient numerical solution of the reduced problem.

The solution of the reduced problem **RCSP** has several advantages over the solution of the original problem **FCSP**:

- The number of independent variables is l instead of $n + l$.
- The constraints of **RCSP** are linear whereas **FCSP** has nonlinear constraints.
- Existing software for the solution of Subproblem (A), see [SK93], can directly be used.
- The inherent band structure of the matrices involved can fully be exploited.

But it should be mentioned that there are the following disadvantages:

- The structure of gradient, Hessian, and Jacobian of the reduced functional f is rather complicated.
- The coefficients $\boldsymbol{\alpha}$ can only be shown to be Lipschitz-continuous but not continuously differentiable with respect to $\tilde{\mathbf{t}}$ in case of nonstrict complementarity.

We treat the reduced problem, which is a nonlinear least squares problem with linear inequality constraints, by a generalized Gauss-Newton method.

5.1. **A generalized Gauss-Newton method.** Let $\mathbf{F} \in \mathbb{R}^{m+n-r}$ be the residual vector of the reduced functional $f = \frac{1}{2}\mathbf{F}^T\mathbf{F}$ and $\tilde{\mathbf{t}}^\nu$ be the current iterate. In the ν-th step of a generalized Gauss-Newton method we have to solve

$$\min\left\{\mu(\tilde{\mathbf{t}}^\nu + \mathbf{s}) = \frac{1}{2}\left\|\mathbf{F}(\tilde{\mathbf{t}}^\nu) + \mathbf{J}(\tilde{\mathbf{t}}^\nu)\mathbf{s}\right\|^2 \,:\, \mathbf{Cs} \geq \mathbf{h} - \mathbf{C}\tilde{\mathbf{t}}^\nu \,:\, \mathbf{s} \in \mathbb{R}^l\right\} \tag{5.1}$$

where $\mu(\tilde{\mathbf{t}}^\nu + \mathbf{s}) \approx f(\tilde{\mathbf{t}}^\nu + \mathbf{s})$ is a quadratic model to the functional f. We call $\mu = \mu_{GP}$ with $\mathbf{J}(\tilde{\mathbf{t}}^\nu) = \boldsymbol{\partial}\mathbf{F}(\tilde{\mathbf{t}}^\nu)$ Golub/Pereyra model and $\mu = \mu_K$ with $\mathbf{J}(\tilde{\mathbf{t}}^\nu) = \mathbf{J}_K(\tilde{\mathbf{t}}^\nu)$ Kaufman model, cf. Section 3. If the Jacobian or its approximation has full rank l, then (5.1) has a unique solution $\mathbf{s}^\nu$ which defines the next iterate $\tilde{\mathbf{t}}^{\nu+1} := \tilde{\mathbf{t}}^\nu + \mathbf{s}^\nu$ ($\nu = 0, 1, \ldots$). In order to obtain global convergence, we apply a safeguarded Armijo-Goldstein line search.

When using the Golub-Pereyra model, the Jacobian $\boldsymbol{\partial}\mathbf{F}(\tilde{\mathbf{t}}^\nu)$ is approximated column-wise according to

$$\boldsymbol{\partial}\mathbf{F}(\tilde{\mathbf{t}}^\nu)\mathbf{e}^\tau \approx \frac{\mathbf{F}(\tilde{\mathbf{t}}^\nu + h_\tau\mathbf{e}^\tau) - \mathbf{F}(\tilde{\mathbf{t}}^\nu)}{h_\tau} \quad (\tau = 1, \ldots, l)$$

with appropriate step-size h_τ. We call this *outer discretization* and denote the corresponding model by **RCSP-GP-OD** (reduced constrained smoothing problem, Golub-Pereyra model, outer discretization).

The overall structure of the algorithm and especially the linear algebra involved is the same as in the Gauss-Newton method of [SS95]. The crucial part is the efficient evaluation of the residual function $\mathbf{F}$ and its Jacobian $\mathbf{J}$ both of which are more complicated in the constrained case considered here.

Algorithm 5.1 (Damped Generalized Gauss-Newton Method).

S1: Choose a feasible initial knot sequence $\tilde{\mathbf{t}}^0 \in \mathbb{R}^l$
$\delta \in (0, \frac{1}{4})$
$\nu := 0$

S2: Repeat

S2.1: $\mathbf{F} := \mathbf{F}(\tilde{\mathbf{t}}^\nu) = \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} - \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}^\nu) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}^\nu) \end{bmatrix} \boldsymbol{\alpha}(\tilde{\mathbf{t}}^\nu)$, where $\boldsymbol{\alpha}(\tilde{\mathbf{t}}^\nu)$ solves Subproblem (A)
{ Compute residual function of reduced functional }
$\mathbf{J} :\approx \boldsymbol{\partial}\mathbf{F}(\tilde{\mathbf{t}}^\nu)$
{ Compute an approximation to the Jacobian (Kaufman model or finite differences) }

S2.2: Compute a descent direction s^ν from

$$\min\left\{\tfrac{1}{2}\left\|\mathbf{F} + \mathbf{J}\mathrm{s}\right\|^2 \,:\, \mathbf{C}\mathrm{s} \geq \mathbf{h} - \mathbf{C}\tilde{\mathbf{t}}^\nu, \, \mathrm{s} \in \mathbb{R}^l\right\}$$

If the problem is ill-conditioned, compute s^ν from the regularized problem

$$\min\left\{\tfrac{1}{2}\left\|\begin{pmatrix} \mathbf{F} \\ \mathbf{0} \end{pmatrix} + \begin{bmatrix} \mathbf{J} \\ \sqrt{\lambda}\mathbf{I} \end{bmatrix}\mathrm{s}\right\|^2 \,:\, \mathbf{C}\mathrm{s} \geq \mathbf{h} - \mathbf{C}\tilde{\mathbf{t}}^\nu, \, \mathrm{s} \in \mathbb{R}^l\right\}$$

with $\lambda = \sqrt{l \times \text{macheps}}\,\left\|\mathbf{J}^T\mathbf{J}\right\|_1$

S2.3: Safeguarded line search
$\gamma := 1.0$
while $f(\tilde{\mathbf{t}}^\nu) - f(\tilde{\mathbf{t}}^\nu + \gamma \mathbf{s}^\nu) < -\gamma\delta\nabla f(\tilde{\mathbf{t}}^\nu)^T \mathbf{s}^\nu$ do
$\gamma := \gamma * \alpha$
{ Here α is chosen so that $\gamma*\alpha$ is the minimizer of a quadratic or cubic polynomial that models $f(\tilde{\mathbf{t}} + \gamma \mathbf{s}^\nu)$ provided that α is not to close to the boundary points 0 or 1 }
S2.4: $\tilde{\mathbf{t}}^{\nu+1} := \tilde{\mathbf{t}}^\nu + \gamma \mathbf{s}^\nu$
$\nu := \nu + 1$
until $\nu > \nu max$ or *convergence*

5.2. **The Kaufman-Approximation in the case of smoothing splines.** In order to simplify the computations we use the Kaufman-Approximation to the Jacobian which can be evaluated much more efficiently. The Kaufman-Approximation $\mathbf{J}_K \in \mathbb{R}^{m+n-r,l}$ to the Jacobian $\boldsymbol{\partial}\mathbf{F}$ is given by

$$\mathbf{J}_K := \mathbf{P}^\perp_{\left[\begin{smallmatrix} B \\ \sqrt{\mu}S \end{smallmatrix}\right] N} \left(\mathbf{J}_{\tilde{\mathbf{t}}} + \begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix} \bar{\mathbf{R}}^+ \bar{\mathbf{\Gamma}} \right) \in \mathbb{R}^{m+n-r,l} \tag{5.2}$$

with

$$\mathbf{J}_{\tilde{\mathbf{t}}} := -\boldsymbol{\partial} \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix} \boldsymbol{\alpha}(\tilde{\mathbf{t}}) \in \mathbb{R}^{m+n-r,l}, \tag{5.3}$$

(5.4)

$$\mathbf{P}^\perp_{\left[\begin{smallmatrix} B \\ \sqrt{\mu}S \end{smallmatrix}\right] N} := \mathbf{I}_{m+n-r} - \left(\begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix} \mathbf{N} \right) \left(\begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix} \mathbf{N} \right)^+ \in \mathbb{R}^{m+n-r,m+n-r}$$

where $\bar{\mathbf{R}} \in \mathbb{R}^{nactiv,n}$ and $\bar{\mathbf{\Gamma}} \in \mathbb{R}^{nactiv,l}$ are defined as in Section 3. As stated for the unconstrained case in [RW80], methods using this approximate Jacobian perform not worse than methods with the full Jacobian. As it can be expected from the theoretical properties of the Kaufman-Approximation, cf. Section 3, our numerical tests confirm that this remains true for the constrained case, too.

In the remaining part of this section it will be shown how the Kaufman-Approximation (5.2) can be computed in an efficient and numerically stable way utilizing the sparsity structure of the matrices involved. The matrix $\mathbf{J}_K \in \mathbb{R}^{m+n-r,l}$ will be computed column-wise, i.e., $\mathbf{J}_K\mathbf{e}^\tau \in \mathbb{R}^{m+n-r}$ $(\tau = 1, \ldots, l)$.

Let us first consider the matrices $\mathbf{J}_{\tilde{\mathbf{t}}}$ and $\bar{\mathbf{\Gamma}}$ that contain derivatives with respect to $\tilde{\mathbf{t}}$. Obviously, we have

$$(\mathbf{J}_{\tilde{\mathbf{t}}}\mathbf{e}^\tau)_{i=1,\ldots,m} = -\left(\sum_{j=1}^n \frac{\partial B_{j,k,\tilde{\mathbf{t}}}(x_i)}{\partial t_{p(\tau)}} \alpha_j(\tilde{\mathbf{t}}) \right)_{i=1,\ldots,m} = -\left(\frac{\partial s(x_i)}{\partial t_{p(\tau)}} \right)_{i=1,\ldots,m}.$$

Thus, we can apply Lemma 4.2 to compute the first m components of $\mathbf{J}_{\tilde{\mathbf{t}}}\mathbf{e}^\tau$. For the smoothing matrix $\mathbf{S}_r(\tilde{\mathbf{t}}) = \mathbf{F}_r(\tilde{\mathbf{t}})\mathbf{D}_r(\tilde{\mathbf{t}})$ we have $\boldsymbol{\partial}\mathbf{S}_r = (\boldsymbol{\partial}\mathbf{F}_r)\,\mathbf{D}_r + \mathbf{F}_r\,(\boldsymbol{\partial}\mathbf{D}_r)$. Whereas $\boldsymbol{\partial}\mathbf{F}_r = \boldsymbol{\partial}\tilde{\mathbf{F}}_r$ can easily be computed in case of the approximate smoothing matrix, it seems very difficult to give an explicit expression for $\boldsymbol{\partial}\mathbf{F}_r = \boldsymbol{\partial}\bar{\mathbf{F}}_r$ in case of the exact smoothing matrix. By induction one can derive the recursion

$$\boldsymbol{\partial}\mathbf{D}_r = \begin{cases} \mathbf{0} & \text{for } r = 0, \\ (\boldsymbol{\partial}\mathbf{H}_r)\,\mathbf{L}_r\mathbf{D}_{r-1} + \mathbf{H}_r\mathbf{L}_r\,(\boldsymbol{\partial}\mathbf{D}_{r-1}) & \text{for } r \geq 1. \end{cases}$$

Using the defining equations for $\mathbf{H}_r$ and $\mathbf{F}_r$, one can finally compute $\boldsymbol{\partial}\mathbf{D}_r\,[\mathbf{e}^\tau]\,\boldsymbol{\alpha}(\tilde{\mathbf{t}}) \in \mathbb{R}^{n-r}$ and $\boldsymbol{\partial}\mathbf{S}_r\,[\mathbf{e}^\tau]\,\boldsymbol{\alpha}(\tilde{\mathbf{t}}) \in \mathbb{R}^{n-r}$. The vector $\bar{\mathbf{\Gamma}}\mathbf{e}^\tau = -\boldsymbol{\partial}\bar{\mathbf{R}}\,[\mathbf{e}^\tau]\,\boldsymbol{\alpha}(\tilde{\mathbf{t}}) \in \mathbb{R}^{nactiv}$ can be computed by identifying the active constraints and choosing the corresponding component of $\boldsymbol{\partial}\mathbf{D}_p\,[\mathbf{e}^\tau]\,\boldsymbol{\alpha}(\tilde{\mathbf{t}})$ with the correct sign. If the occurring derivatives are evaluated in the above way then the model is denoted by **RCSP-Ka-ED** (reduced constrained smoothing problem, Kaufman model, exact derivatives).

In the next step we consider the computation of the matrix $\mathbf{N} \in \mathbb{R}^{n,n-nactiv}$ and of vectors $\mathbf{x} := \bar{\mathbf{R}}^{+}\mathbf{y} \in \mathbb{R}^{n}$, $\mathbf{y} \in \mathbb{R}^{nactiv}$ arbitrary. Note that $\bar{\mathbf{R}} \in \mathbb{R}^{nactiv,n}$ has $p+1$ non-zero elements per row, and that rank $\bar{\mathbf{R}} = nactiv$. We choose the method of semi-normal equations that, unlike in the over-determined case is numerically stable in case of under-determined equations, see [Pai73] and [Bjö90].

Algorithm 5.2 (Computation of $\mathbf{x} = \bar{\mathbf{R}}^{+}\mathbf{y}$, $\mathbf{N}$ for given $\bar{\mathbf{R}}$, $\mathbf{y}$; semi-normal equations).

S1: Compute QR-factorization by means of Householder transformations

$$\begin{bmatrix} \mathbf{Q}_{11}^T \\ \mathbf{Q}_{12}^T \end{bmatrix} \bar{\mathbf{R}}^T = \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{0} \end{bmatrix}, \qquad \bar{\mathbf{R}}^T \in \mathbb{R}^{n,nactiv}$$

Save the regular, upper triangular factor $\mathbf{R}_1 \in \mathbb{R}^{nactiv,nactiv}$

S2: Set $\mathbf{N} := \mathbf{Q}_{12} \in \mathbb{R}^{n,n-nactiv}$ (orthonormal basis of the null space of $\bar{\mathbf{R}}$

S3: For $\mathbf{y} \in \mathbb{R}^{nactiv}$

S3.1: Solve $\mathbf{R}_1^T\mathbf{R}_1\mathbf{w} = \mathbf{y}$ for $\mathbf{w} \in \mathbb{R}^{nactiv}$

S3.2: Compute $\mathbf{x} := \bar{\mathbf{R}}^T\mathbf{w} \in \mathbb{R}^{n}$

Let us point out that steps S1 and S2 have to be performed only once for all right-hand sides $\mathbf{y}$, and that only the "small" matrices $\mathbf{R}_1$ and $\mathbf{Q}_{12}$ have to be stored.

Finally, we treat the computation of the orthogonal projector (5.4). It should be noted that the usual way, namely forming the matrix $\left[\begin{smallmatrix} B \\ \sqrt{\mu}S \end{smallmatrix}\right] N$ explicitly and doing an orthogonal decomposition, is very inefficient, as the band structure of $\mathbf{B}$ and $\mathbf{S}$ is destroyed. Instead we make use of two other orthogonal decompositions that preserve the band structure and have been computed already when solving Subproblem (A), see [SK93] for details.

Lemma 5.1 (Computation of the projector P).

Let $\mathbf{B} \in \mathbb{R}^{m,n}$ and $\mathbf{S}_r \in \mathbb{R}^{n-r,n}$ be given matrices so that $\begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix} \in \mathbb{R}^{m+n-r,n}$ *has full rank n for $\mu > 0$ and $m \geq r$. Further, let the following QR-factorizations be known*

$$\mathbf{Q}_0^T\mathbf{B} = \begin{bmatrix} \mathbf{R}_0 \\ \mathbf{0} \end{bmatrix}, \quad \tilde{\mathbf{Q}}^T \begin{bmatrix} \mathbf{R}_0 \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{R}} \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{Q}_2^T\left(\tilde{\mathbf{R}}\mathbf{N}\right) = \begin{bmatrix} \mathbf{R}_2 \\ \mathbf{0} \end{bmatrix}$$

where $\mathbf{N} \in \mathbb{R}^{n,n-nactiv}$ has linear independent columns. Then, for an arbitrary vector $\mathbf{y} \in \mathbb{R}^{m+n-r}$, we have

$$\mathbf{P}^{\perp}_{\left[\begin{smallmatrix} B \\ \sqrt{\mu}S \end{smallmatrix}\right]N}\mathbf{y} = \left[\mathbf{I}_{m+n-r} - \begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix} \tilde{\mathbf{R}}^{-1}\mathbf{Q}_2 \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{Q}_2^T\tilde{\mathbf{R}}^{-T} \begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix}^T\right]\mathbf{y}.$$

Summarizing the results of this section we can formulate the algorithm for the computation of the Kaufman-Approximation:

Algorithm 5.3 (Computation of the Jacobian, Kaufman-Approximation).

S1: Compute QR-factorization by means of row-by-row Givens rotations

$$\mathbf{Q}_0^T\mathbf{B} = \begin{bmatrix} \mathbf{R}_0 \\ \mathbf{0} \end{bmatrix}, \qquad \mathbf{B} \in \mathbb{R}^{m,n}$$

S2: Compute QR-factorization by means of row-by-row Givens rotations

$$\tilde{\mathbf{Q}}^T \begin{bmatrix} \mathbf{R}_0 \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{R}} \\ \mathbf{0} \end{bmatrix}, \qquad \begin{bmatrix} \mathbf{R}_0 \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix} \in \mathbb{R}^{2n-r,n}$$

S3: Compute $\boldsymbol{\alpha}(\tilde{\mathbf{t}})$ as solution of

$$\min\left\{\frac{1}{2}\left\|\begin{pmatrix}\mathbf{y}\\ \mathbf{0}\end{pmatrix}-\begin{bmatrix}\mathbf{B}(\tilde{\mathbf{t}})\\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}})\end{bmatrix}\boldsymbol{\alpha}\right\|^2 : \mathbf{L}\le\mathbf{D}_p(\tilde{\mathbf{t}})\boldsymbol{\alpha}\le\mathbf{U} : \boldsymbol{\alpha}\in\mathbb{R}^n\right\}$$

using the QR-factorizations of S1 and S2, see [SK93]

S4: Let

$$\bar{\mathbf{R}} := -\begin{bmatrix}\mathbf{D}_p(\tilde{\mathbf{t}})\\ -\mathbf{D}_p(\tilde{\mathbf{t}})\end{bmatrix}_{i\in\mathcal{I}} \in\mathbb{R}^{nactiv,n}$$

S5: Compute QR-factorization by means of Householder transformations

$$\mathbf{Q}_1^T\bar{\mathbf{R}}^T=\begin{bmatrix}\mathbf{Q}_{11}^T\\ \mathbf{Q}_{12}^T\end{bmatrix}\bar{\mathbf{R}}^T=\begin{bmatrix}\mathbf{R}_1\\ \mathbf{0}\end{bmatrix},\qquad \bar{\mathbf{R}}^T\in\mathbb{R}^{n,nactiv}$$

S6: Null space basis $\mathbf{N}\in\mathbb{R}^{n,n-nactiv}$ of $\bar{\mathbf{R}}$

$$\mathbf{N} := \mathbf{Q}_{12}$$

S7: Compute QR-factorization by means of Householder transformations

$$\mathbf{Q}_2^T\left(\tilde{\mathbf{R}}\mathbf{N}\right)=\begin{bmatrix}\mathbf{R}_2\\ \mathbf{0}\end{bmatrix},\qquad \tilde{\mathbf{R}}\mathbf{N}\in\mathbb{R}^{n,n-nactiv}$$

S8: for $\tau := 1$ to l do

S8.1: Compute $\mathbf{v}^1 := \mathbf{J}_{\tilde{\mathbf{t}}}\mathbf{e}^\tau = -\boldsymbol{\partial}\begin{bmatrix}\mathbf{B}(\tilde{\mathbf{t}})\\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}})\end{bmatrix}[\mathbf{e}^\tau]\,\boldsymbol{\alpha}(\tilde{\mathbf{t}})\in\mathbb{R}^{m+n-r}$;

S8.2: Compute $\mathbf{v}^2 := \bar{\boldsymbol{\Gamma}}\mathbf{e}^\tau = -\boldsymbol{\partial}\bar{\mathbf{R}}(\tilde{\mathbf{t}})\,[\mathbf{e}^\tau]\,\boldsymbol{\alpha}(\tilde{\mathbf{t}})\in\mathbb{R}^{nactiv}$;

S8.3: Solve the system $\mathbf{R}_1^T\mathbf{R}_1\mathbf{v}^3 = \mathbf{v}^2$; ($\mathbf{R}_1$ upper triangular, $\mathbf{v}^3\in\mathbb{R}^{nactiv}$)

S8.4: $\mathbf{v}^4 := \bar{\mathbf{R}}^T\mathbf{v}^3\in\mathbb{R}^n$;

S8.5: $\mathbf{v}^5 := \begin{bmatrix}\mathbf{B}\\ \sqrt{\mu}\mathbf{S}_r\end{bmatrix}\mathbf{v}^4\in\mathbb{R}^{m+n-r}$;

S8.6: $\mathbf{v}^6 := \mathbf{v}^1+\mathbf{v}^5\in\mathbb{R}^{m+n-r}$;

S8.7: Compute $\mathbf{v}^7\in\mathbb{R}^{m+n-r}$

$$\mathbf{v}^7 := \left[\mathbf{I}_{m+n-r}-\begin{bmatrix}\mathbf{B}\\ \sqrt{\mu}\mathbf{S}_r\end{bmatrix}\tilde{\mathbf{R}}^{-1}\mathbf{Q}_2\begin{bmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{Q}_2^T\tilde{\mathbf{R}}^{-T}\begin{bmatrix}\mathbf{B}\\ \sqrt{\mu}\mathbf{S}_r\end{bmatrix}^T\right]\mathbf{v}^6;$$

S8.8: $\mathbf{J}_K\mathbf{e}^\tau := \mathbf{v}^7$;

6. Numerical tests

In this section, the capabilities of our method are demonstrated using some examples from the literature. Algorithm 5.1 has been implemented in PASCAL with IEEE double arithmetic ($macheps = 2.2$ E-16). We have used the termination criteria of Table 1 with $\varepsilon_0^t = \varepsilon_1^t = \varepsilon_2^t = \varepsilon_5^t = 1.0$ E-10, $\varepsilon_3^t = 1.0$ E-06, $\varepsilon_4^t = 1.0$ E-03. Note that these rather stringent values, e.g., $\varepsilon_3^t = 1.0$ E-06, have been chosen only for testing purposes. In real applications one would take $10^{-2}\ldots 10^{-3}$. All tests have been run with the relative knot separation parameter $\epsilon = 0.0625$.

6.1. **Titanium Heat Data.** Our first example is the well known *Titanium Heat Data* set from de Boor. We want to approximate these $m = 49$ data points by $n = 11$ B-splines of order $k = 4$ using the smoothing parameter $\mu = 1.0$ and $r = 2$. We fix the inner knots $t_7 = 835$ and $t_{10} = 955$, i.e., $l = 5$, and distribute the remaining free knots t_5, t_6 and t_8, t_9 and t_{11} equidistantly in the intervals $[595, 835]$, $[835, 955]$ and $[955, 1075]$. Imposing convexity conditions, i.e., $p = 2$, on the spline in the intervals $[595, 835)$ and $[955, 1075)$ we obtain

$$\mathbf{L} = (0,0,0,0,-\infty,-\infty,0,0,0)^T$$

$$\mathbf{U} = (+\infty,+\infty,+\infty,+\infty,+\infty,+\infty,+\infty,+\infty,+\infty)^T.$$

TABLE 1. Return codes and termination criteria

1	$\left\|\mathbf{F}^{\nu+1}\right\| \le \varepsilon_0^t$
2	$\left\|\mathbf{J}^{\nu T}\mathbf{F}^{\nu}\right\| \le \varepsilon_1^t$
3	$\left\|\mathbf{F}^{\nu T}\mathbf{J}^{\nu}\mathbf{s}^{\nu}\right\| \le \varepsilon_2^t$
4	$\left\|\mathbf{x}^{\nu+1} - \mathbf{x}^{\nu}\right\| \le \varepsilon_3^t \left(\|\mathbf{x}^{\nu}\| + \varepsilon_4^t\right)$
5	$\left\|\left\|\mathbf{F}^{\nu+1}\right\| - \|\mathbf{F}^{\nu}\|\right\| \le \varepsilon_5^t \|\mathbf{F}^{\nu}\|$
6	$\nu > \nu max$
7	failed otherwise

TABLE 2. Titanium Heat Data: Spline Smoothing

	$\tilde{\mathbf{t}}^0$	**RCSP-Ka-ED**	**RCSP-GP-OD**	**RSP-Ka-ED**
t_5	675.0	7.975133 E+02	7.822991 E+02	5.959958 E+02
t_6	755.0	8.110142 E+02	7.947857 E+02	6.109336 E+02
t_8	875.0	8.751572 E+02	8.755310 E+02	8.767428 E+02
t_9	915.0	8.810366 E+02	8.804978 E+02	8.816339 E+02
t_{11}	1015.0	9.625000 E+02	9.625000 E+02	9.625000 E+02
$\|\mathbf{F}\|$	1.027722 E+00	3.469246 E-01	3.460394 E-01	3.544604 E-01
steps		7	13	9
time (ms)		149	280	143
$\|\mathbf{F}^T\mathbf{J}\mathbf{s}\|$		2.491557 E-03	3.592591 E-11	5.363720 E-10
$\|\mathbf{J}^T\mathbf{F}\|$		2.797501 E-03	2.988107 E-03	1.749176 E-03
Ret. Code		4	3	4

Table 2 shows the results of spline smoothing with free knots using different methods. The name **RSP-Ka-ED** (reduced smoothing problem, Kaufman model, exact derivatives) denotes a method from [SS95] for unconstrained spline smoothing with free knots. Here the location of knots was first optimized disregarding the shape constraints. Afterwards a shape preserving spline to this fixed knot sequence was computed.

For comparison we have repeated the tests by using unsmoothed spline approximation, i.e., by setting $\mu = 0$. Note that in this case there are some simplifications which have been implemented in a special algorithm **RCAP** (reduced constrained approximation problem). Table 3 displays the results of these tests.

Figure 2 and 3 show the spline curve before and after the optimization.[1] The improvement in the shape of the spline and the approximation error is clearly recognizable.

6.2. **Arctan Data.** In a second example we illustrate smoothing by monotone splines with free knots. We consider data generated by sampling the function $g(x) = \arctan(10x)$ at $m = 41$ equidistant points x_i in $[-10, 10]$. Using pseudo random numbers ϵ_i, $-0.075 \le \epsilon_i \le 0.075$ we obtain the perturbed values $y_i = g(x_i)(1 + \epsilon_i)$, $i = 1, \dots, m$.

[1] In all figures, a small circle or a cross mark a data point or a knot, respectively.

TABLE 3. Titanium Heat Data: Spline Approximation ($\mu = 0$)

	$\tilde{\mathbf{t}}^0$	**RCAP-Ka-ED**	**RCAP-GP-OD**	**RAP-Ka-ED**
t_5	675.0	6.047707 E+02	7.821307 E+02	5.959958 E+02
t_6	755.0	7.352655 E+02	7.946061 E+02	6.109336 E+02
t_8	875.0	8.756307 E+02	8.755297 E+02	8.767300 E+02
t_9	915.0	8.805912 E+02	8.804966 E+02	8.816219 E+02
t_{11}	1015.0	9.625000 E+02	9.625000 E+02	9.625000 E+02
$\\|\mathbf{F}\\|$	1.027678 E+00	3.457712 E-01	3.449610 E-01	3.532900 E-01
steps		6	13	9
time (ms)		149	253	116
$\|\mathbf{F}^T\mathbf{J}\mathbf{s}\|$		1.621106 E-04	2.981807 E-11	5.518354 E-12
$\\|\mathbf{J}^T\mathbf{F}\\|$		3.015662 E-03	2.998139 E-03	1.757170 E-03
Ret. Code		4	3	4

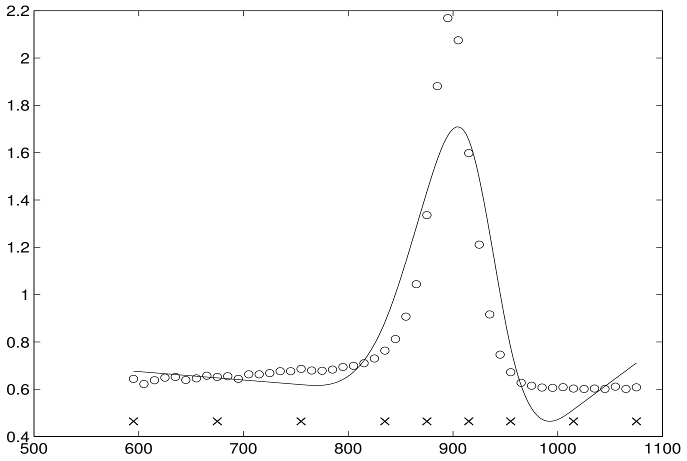

FIGURE 2. Titanium Heat Data: Spline s, initial knot sequence

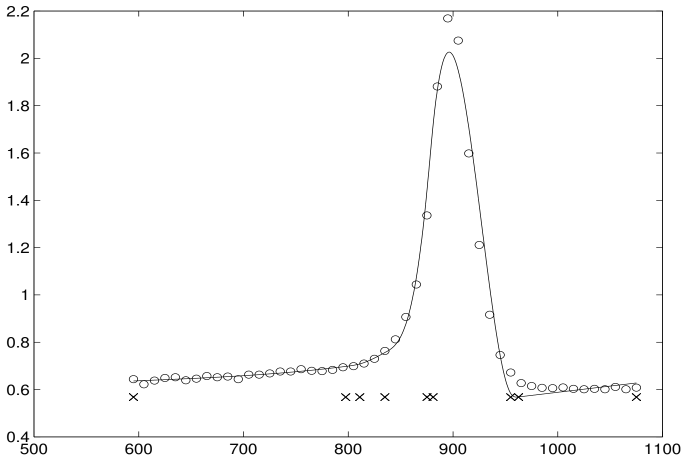

FIGURE 3. Titanium Heat Data: Spline s, optimized knot sequence, **RCSP-Ka-ED**

TABLE 4. Arctan Data: Spline Smoothing

	$\tilde{\mathbf{t}}^0$	**RCSP-Ka-ED**	**RCSP-GP-OD**	**RSP-Ka-ED**
t_5	-6.0	-8.760175 E-01	-9.652265 E-01	-8.838273 E-01
t_6	-2.0	-2.301281 E-01	-3.258263 E-01	-2.760825 E-01
t_7	2.0	2.743868 E-01	3.454588 E-01	2.779473 E-01
t_8	6.0	8.822404 E-01	9.809856 E-01	1.090103 E-01
$\|\mathbf{F}\|$	2.359790 E+00	5.230920 E-01	5.098921 E-01	5.283729 E-01
steps		4	100	13
time (ms)		66	1741	148
$\lvert\mathbf{F}^T\mathbf{J}\mathbf{s}\rvert$		8.407664 E-04	4.345827 E-06	1.0626114 E-10
$\|\mathbf{J}^T\mathbf{F}\|$		4.336670 E-02	1.270197 E-03	7.8463565 E-03
Ret. Code		4	6	5

TABLE 5. Arctan Data: Spline Approximation ($\mu = 0$)

	$\tilde{\mathbf{t}}^0$	**RCAP-Ka-ED**	**RCAP-GP-OD**	**RAP-Ka-ED**
t_5	-6.0	-8.100201 E-01	-8.074899 E-01	-6.794899 E-01
t_6	-2.0	-1.689546 E-01	-1.946541 E-01	-5.812258 E-02
t_7	2.0	1.598707 E-01	1.855516 E-01	5.200732 E-03
t_8	6.0	8.773101 E-01	8.745770 E-01	9.550505 E-01
$\|\mathbf{F}\|$	2.359717 E+00	4.268960 E-01	4.268960 E-01	4.423301 E-01
steps		6	8	8
time (ms)		77	126	77
$\lvert\mathbf{F}^T\mathbf{J}\mathbf{s}\rvert$		1.664365 E-14	7.873466 E-09	1.571400 E-12
$\|\mathbf{J}^T\mathbf{F}\|$		1.727962 E-07	7.822950 E-07	1.278808 E-02
Ret. Code		3	4	3

We approximate these data by $n = 8$ B-splines of order $k = 4$. The spline s is required to be monotone in $[-10, 10]$, i.e.,

$$\mathbf{L} = (0, 0, 0, 0, 0, 0, 0)^T$$

$$\mathbf{U} = (+\infty, +\infty, +\infty, +\infty, +\infty, +\infty, +\infty)^T$$

We choose $l = 4$ equidistant inner knots as initial knot sequence.

Tables 4 and 5 show the results of spline smoothing and approximation, respectively. We have used $\mu = 1.0E - 3$ and $r = 2$ for the smoothing case.

As it can be observed in Figure 4, the rapid change of curvature near zero is not adequately represented by equidistant inner knots. This can clearly be seen in Figure 5 where the first derivative of the spline s and the underlying function g is shown. By optimizing the location of the inner knots the residual was reduced by 80%. Although the residuals of the initial knot sequence are almost identical for spline smoothing and approximation, a significant difference is observed after the optimization. This fact is caused by the greater influence of the smoothing term for non-equidistant knots.

6.3. **Volumetric Moisture Content Data.** In the last example we compare our algorithm with an adaptive strategy for knot placement under convexity conditions that has been implemented in the routine **CONCON** of the well-known **FIT-PACK** package by Dierckx [Die87], [Die89]. In this routine, starting from a coarse knot sequence, additional knots will be inserted until the shape constraints are met

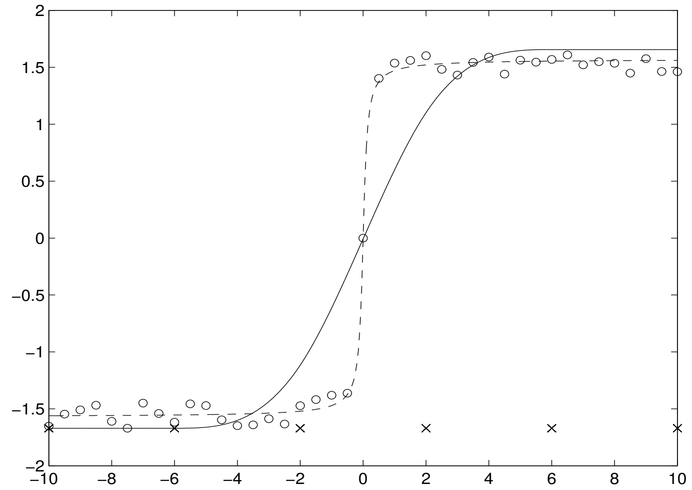

FIGURE 4. Arctan Data: Spline s (solid) and function g (dashed), initial knot sequence

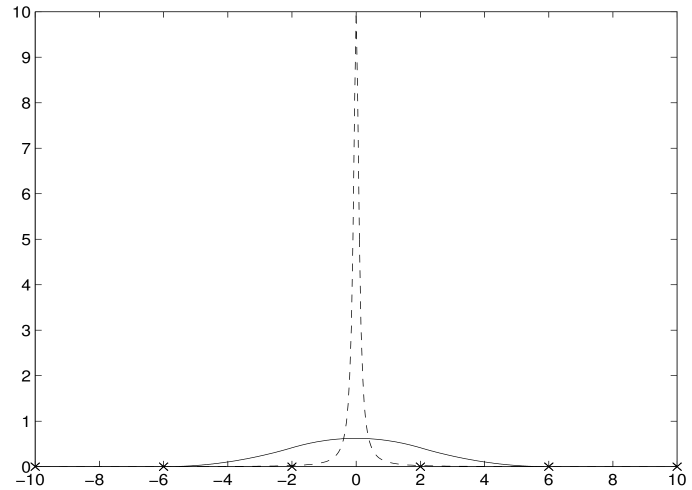

FIGURE 5. Arctan Data: First derivatives s' (solid) and g' (dashed), initial knot sequence

and the residual is below a prescribed threshold S. Note that the intention of **CONCON** is *not* to find an "optimal" location of knots rather than to provide a fast and "good" automatic knot placement.

We have used the volumetric moisture content data of the test program accompanying the **FITPACK** package, see [Die87, p. 79] and [Die93, p. 129]. Given the bound $S = ||\mathbf{F}||^2 = 0.0002$ for the residual sum, **CONCON** computes an concave approximation with $n = 7$ cubic B-splines resulting in $||\mathbf{F}|| = 0.012097$. Our algorithm **RCAP-Ka-ED** was started from an equidistant knot sequence and gives $||\mathbf{F}|| = 0.010675$, see Table 6. The quality of both approximations is comparable as it can be observed in Figure 7. It should be mentioned that we were not able to get an satisfactory result with **CONCON** in the Titanium Heat Data example. **CONCON** stops after inserting two inner knots since "adding one or more knots will not further reduce the value of SQ (sum of squared residuals)."

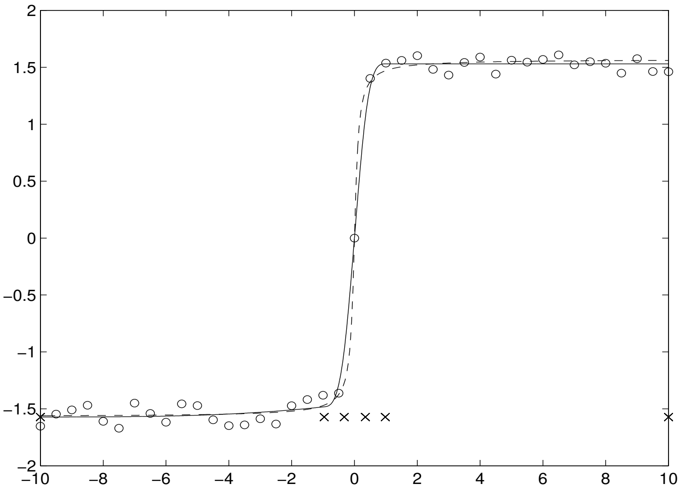

FIGURE 6. Arctan Data: Spline s (solid) and function g (dashed), optimized knot sequence, **RCAP-GP-OD**

TABLE 6. Volumetric Moisture Content Data

	$\tilde{\mathbf{t}}^0$	**CONCON**	**RCAP-Ka-ED**
t_5	2.45	0.30	0.14
t_6	4.80	0.70	0.83
t_8	7.15	2.25	4.01
$\|\mathbf{F}\|$	0.064072	0.012709	0.010675

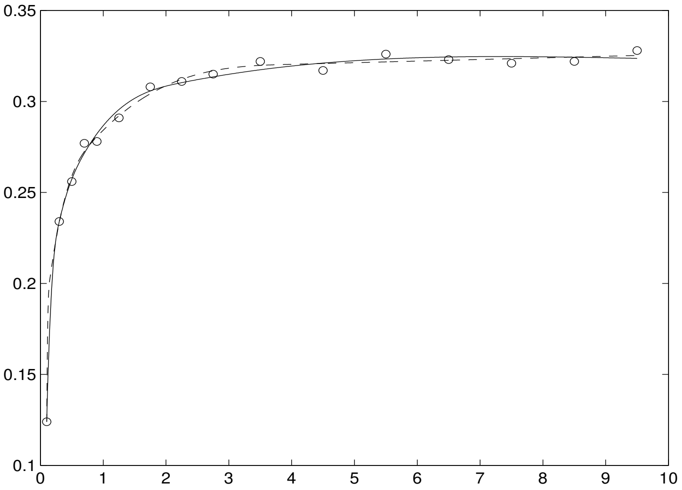

FIGURE 7. Volumetric Moisture Content Data: **CONCON** (solid), **RCAP-Ka-ED** (dashed)

The numerical tests show that our method is a competitive algorithm for the computation of free knot splines under shape constraints. The algorithm gives comparable results as the commonly used **CONCON** routine from the **FITPACK** package and extends its functionality, i.e., other constraints than convexity-concavity conditions, arbitrary spline order, and the incorporation of a smoothing term. Since the sparsity structure of the matrices involved is exploited to a great extent, the algorithm is efficient even for larger data sets.

References

[Bjö90] Å. Björck, *Least squares methods*, Handbook of Numerical Analysis, volume I, Solution of equations in R^n (P. G. Ciarlet and J. L. Lions, eds.), North Holland, 1990, pp. 589–652.

[Cor81] C. Corradi, *A note on the solution of separable nonlinear least squares problems with separable nonlinear equality constraints*, SIAM J. Numer. Anal. **18** (1981), 1134–1138.

[Dan73] J. W. Daniel, *Stability of definite quadratic programs*, Math. Prog. **5** (1973), 41–53.

[dBR68] C. de Boor and J. R. Rice, *Least squares cubic spline approximation II – variable knots*, Technical Report CSD TR 21, Computer Science Department, Purdue University, 1968.

[Die79] P. Dierckx, *Het aanpassen van krommen en oppervlakken aan meetpunten met behulp van spline funkties*, Ph.D. thesis, Katholieke Universiteit Leuven, 1979.

[Die87] P. Dierckx, *FITPACK user guide, part 1: Curve fitting routines*, Tech. Report TW Report 89, Department of Computer Science, Katholieke Universiteit Leuven, Belgium, 1987.

[Die89] P. Dierckx, *FITPACK user guide, part 2: Surface fitting routines*, Tech. Report TW Report 122, Department of Computer Science, Katholieke Universiteit Leuven, Belgium, 1989.

[Die93] P. Dierckx, *Curve and surface fitting with splines*, Oxford University Press, 1993.

[EA88] T. Elfving and L.-E. Andersson, *An algorithm for computing constrained smoothing spline functions*, Numer. Math. **52** (1988), 583–595.

[Fia76] A. V. Fiacco, *Sensitivity analysis for nonlinear programming using penalty methods*, Math. Prog. **10** (1976), 287–311.

[Fia83] A. V. Fiacco, *Introduction to sensitivity and stability analysis in nonlinear programming*, Academic Press, 1983.

[GP73] G. H. Golub and V. Pereyra, *The differentiation of pseudoinverses and nonlinear least squares problems whose variables separate*, SIAM J. Numer. Anal. **10** (1973), 413–432.

[Jup78] D. L. B. Jupp, *Approximation to data by splines with free knots*, SIAM J. Numer. Anal. **15** (1978), no. 2, 328–343.

[Kau75] L. Kaufman, *A variable projection method for solving separable nonlinear least squares problems*, BIT **15** (1975), no. 4, 49–57.

[KP78] L. Kaufman and V. Pereyra, *A method for separable nonlinear least squares problems with separably nonlinear equality constraints*, SIAM J. Numer. Anal. **15** (1978), 12–20.

[MU88] C. A. Micchelli and F. I. Utreras, *Smoothing and interpolation in a convex subset of a hilbert space*, SIAM J. Sci. Stat. Comput. **9** (1988), no. 4, 728–746.

[Pai73] C. C. Paige, *An error analysis of a method for solving matrix equations*, Math. Comp. **27** (1973), 355–359.

[Par85] T. A. Parks, *Reducible nonlinear programming problems*, Ph.D. thesis, Houston Univ., Dept. of Mathematics, Houston, 1985.

[RW80] A. Ruhe and P. Å. Wedin, *Algorithms for separable nonlinear least squares problems*, SIAM Rev. **22** (1980), no. 3, 318–337.

[Sch91] H. Schwetlick, *Nichtlineare Parameterschätzung: Modelle, Schätzkriterien und numerische Algorithmen*, GAMM-Mitteilungen **2/91** (1991), 13–51.

[Sch92] H. Schwetlick, *Nonlinear parameter estimation: Models, criteria, and algorithms*, Numerical Analysis 1991. Proceedings of the 14th Dundee Conference on Numerical Analysis (New York) (D. F. Griffiths and G. A. Watson, eds.), J. Wiley, 1992, pp. 164–193.

[SK93] H. Schwetlick and V. Kunert, *Spline smoothing under constraints on derivatives*, BIT **33** (1993), 512–528.

[SS90] J. W. Schmidt and I. Scholz, *A dual algorithm for convex-concave data smoothing by cubic C^2-splines*, Numer. Math. **57** (1990), 333–350.

[SS95] H. Schwetlick and T. Schütze, *Least squares approximation by splines with free knots*, BIT **35** (1995), no. 3, 361–384.

[SS96] T. Schütze and H. Schwetlick, *On the convergence of Kaufman-like methods for semilinear least squares problems*, in preparation, 1996.

(T. Schütze) Department of Mathematics, Technical University of Dresden, D-01062 Dresden, Germany

E-mail address: `schuetze@math.tu-dresden.de`

(H. Schwetlick) Department of Mathematics, Technical University of Dresden, D-01062 Dresden, Germany

E-mail address: `schwetlick@math.tu-dresden.de`