Geometrically designed, variable knot regression splines

Vladimir K. Kaishev*, Dimitrina S. Dimitrova, Steven Haberman and Richard J. Verrall

Cass Business School, City University London

August 4, 2015

Abstract

A new method of Geometrically Designed least squares (LS) splines with variable knots, named GeDS, is proposed. It is based on the property that the spline regression function, viewed as a parametric curve, has a control polygon and, due to the shape preserving and convex hull properties, it closely follows the shape of this control polygon. The latter has vertices whose x-coordinates are certain knot averages and whose y-coordinates are the regression coefficients. Thus, manipulation of the position of the control polygon may be interpreted as estimation of the spline curve knots and coefficients. These geometric ideas are implemented in the two stages of the GeDS estimation method. In stage A, a linear LS spline fit to the data is constructed, and viewed as the initial position of the control polygon of a higher order ($n > 2$) smooth spline curve. In stage B, the optimal set of knots of this higher order spline curve is found, so that its control polygon is as close to the initial polygon of stage A as possible and finally, the LS estimates of the regression coefficients of this curve are found. The GeDS method produces simultaneously linear, quadratic, cubic (and possibly higher order) spline fits with one and the same number of B-spline coefficients. Numerical examples are provided and further supplemental materials are available online.

Keywords: spline regression, B-splines, Greville abscissae, variable knot splines, control polygon

1 Introduction.

Consider a response variable y and an independent variable x, taking values within an interval $[a, b]$, and assume there is a relationship between x and y of the form

$$y = f(x) + \epsilon, \tag{1}$$

where $f(\cdot)$ is an unknown function and ϵ is a random error variable with zero mean and variance $E\epsilon^2 = \sigma_\epsilon^2 > 0$. We will consider the regression problem of estimating $f(\cdot)$, based on a sample of observations $\{x_i, y_i\}_{i=1}^N$, where the design points $\{x_i\}_{i=1}^N$ may be either deterministic or random.

*Corresponding author's address: Faculty of Actuarial Science and Insurance, Cass Business School, City University London, 106 Bunhill Row, London EC1Y 8TZ, UK. E-mail address: v.kaishev@city.ac.uk

One popular solution is to approximate f with an n-th order (degree $n-1$) spline function defined on $[a, b]$, by the number and location of its k internal knots and by the coefficients in front of the set of basis functions. Several approaches to constructing free-knot regression splines have been developed. One possibility is to assume that n and k are fixed (but unknown), and to find the knot locations which minimize the (non-linear) least squares criterion, or an appropriately penalized (see Lindstrom 1999) or bounded (see Molinari et al. 2004) version of it. For an extensive discussion of the (dis)advantages of non-linear free-knot spline estimation, we refer to Lindstrom (1999) and Jupp (1978). More recently, Beliakov (2004) proposed to apply the cutting angle deterministic global optimization method to the free-knot least squares spline approximation problem. As the author mentions, the method is effective for a reasonably small number of knots and data points.

In order to circumvent the difficulties related to the non-linear optimization approach, a number of authors have developed adaptive knot selection procedures, such as step-wise knot inclusion/deletion strategies. Among the latter are the early work of Smith (1982), the TURBO spline modelling technique of Friedman and Silverman (1989), the MARS method proposed by Friedman (1991), the POLYMARS of Stone et al. (1997), the spatially adaptive regression splines (SARS) of Zhou and Shen (2001) and the adaptive free-knot splines (AFKS) proposed by Miyata and Shen (2003). The properties of these greedy knot insertion/deletion schemes are well summarized by Hansen and Kooperberg (2002).

More recently, Kang et al. (2015) proposed a two stage, knot deletion and adjustment procedure. At the fist stage, it minimizes the number of knots of a dense initial knot vector by solving an appropriate convex sparse optimization problem. At the second stage, the number of knots and their positions are further adjusted, following specific knot manipulation rules. While this procedure generally leads to less number of knots compared to other methods, as noted by the authors, its major drawbacks are that it has a heavy computational cost and can only be applied to data with low noise level. Furthermore, the selection of the number and positions of the initial knots is not formally defined, and is somewhat arbitrary. Similar in spirit are the knot deletion strategies using sparse optimization proposed earlier by Van Loock et al. (2011) and Yuan et al. (2013) whose resulting knot vector is a subset of the initial knot vector. Other methods, such as the minimum description length (MDL) regression splines of Lee (2000), have been proposed. Multivariate spline regression and knot location has also been considered by

Kaishev (1984).

Another group of works applies reversible jump Markov chain Monte Carlo (RJMCMC) based methods to develop Bayesian adaptive splines, such as those of Smith and Kohn (1996), Denison et al. (1998) and Biller (2000), in the context of generalized linear models. Recently, Belitser and Serra (2014) use random splines as adaptive priors in Bayesian nonparametric regression. A stochastic optimization algorithm for free-knot splines, called adaptive genetic splines (AGS), has been proposed by Pittman (2002) but the related computational cost is a concern, as noted by the author. Genetic algorithms have also been used by Lee (2002a) in (free-knot) spline smoothing of discontinuous regression functions. See also Lee (2002b) for comparisons of genetic and knot inclusion/exclusion algorithms.

Smoothing spline fitting methods, involving a smoothing penalty in the objective function have also been proposed and thoroughly studied in the statistical literature. We will mention here the hybrid adaptive splines (HAS) of Luo and Wahba (1997) and the penalized splines, considered by Eubank (1988), Wahba (1990), Marx and Eilers (1996), Mammen and van der Geer (1997), Rupert and Carroll (2000), Rupert (2002), Wood (2003), Antoniadis et al. (2012).

Our aim in this paper is to give a new geometric perspective to variable knot spline regression estimation and to demonstrate how it can be used to construct a new adaptive spline smoothing method. It avoids many of the limitations of the knot optimization methods summarized above (see section 5), while at the same time it produces competitive fits with a low number of knots for a wide range of signal-to-noise ratios and for both sparse and dense data points at a very low computational cost. However, this extended flexibility is achieved at the cost of introducing two tuning parameters, which are accordingly adjusted, following a set of somewhat heuristic rules (see sections 3 and 4). We consider here the simple univariate regression case, but it is worth noting that our method can be extended to the more general framework of multivariate Extended Linear Models (EML), considered by Hansen and Kooperberg (2002). The novel variable knot spline regression estimation method presented here is based on ideas from the field of Computer Aided Geometric Design (CAGD) of curves and surfaces. It utilizes the close relationship of a spline regression curve to its *control polygon* (see section 2) which to the best of our knowledge has not been previously explored in the statistical literature.

The method includes two stages. In stage A, utilizing a "stick breaking" type of procedure, (see section 3), a least squares linear spline fit to the data is constructed and viewed as the

initial position of the control polygon of a higher order spline regression curve. In stage B, an optimal set of knots of a smoother, higher order ($n > 2$) least squares spline approximation is found so that the latter has also the characteristics of a Schoenberg's variation diminishing spline (VDS) approximation of the linear spline fit from stage A. Thus, the purpose of stage B is to produce a smoother, higher order spline model using the information about the unknown function already embedded in the linear spline fit from stage A.

For the purpose of constructing a linear spline fit in stage A, one can in principle use any of the adaptive knot selection procedures listed above. Similarly to these procedures, the proposed "stick breaking" approach in stage A, provides a heuristic search for the optimum solution of the non-linear knot-parameter estimation problem. As in the case of other methods, finding the global optimum is not guaranteed. At the same time based on our extensive experience in fitting real and simulated data samples, the method demonstrates remarkable spatial adaptivity, producing adequate piecewise-linear reconstructions of the unknown underlying function. Since it does not use the data points as knot candidates our stick breaking method is free of the so called "knot confounding" problem highlighted by Zhou and Shen (2001).

We show that this new spline regression estimation method has a direct Geometric Design interpretation. For this reason, we refer to the related estimator as a GeD spline estimator or simply GeDS. We demonstrate that the latter can be viewed as a competitive alternative to other existing spatially adaptive nonparametric spline methods.

The paper is organized as follows. In section 2, it is shown that a spline regression function can be interpreted as a special case of a parametric spline curve, with a (control) polygon closely related to it. This geometric characterization of the regression problem is used in section 3 to develop the GeD spline regression estimation method and in particular, to formulate its two stages as optimization problems. In section 3.1, the optimality properties of the knots of the higher order spline regression model, obtained in stage B, are established. The GeDS method is illustrated numerically and compared with other spline smoothing methods in section 4 and in the online supplement to the paper, based on simulated and real data examples. In section 5, we provide some discussion and conclusions. Proofs of the results of sections 3.1 are given in Appendix A.

2 The B-spline regression and its control polygon.

Denote by $S_{\mathbf{t}_{k,n}}$ the linear space of all n-th order spline functions defined on a set of non-decreasing knots $\mathbf{t}_{k,n} = \{t_i\}_{i=1}^{2n+k}$, where $t_n = a$, $t_{n+k+1} = b$. In this paper we will use splines with simple knots, except for the n left and right most knots which will be assumed coalescent, i.e.

$$\mathbf{t}_{k,n} = \{t_1 = \ldots = t_n < t_{n+1} < \ldots < t_{n+k} < t_{n+k+1} = \ldots = t_{2n+k}\}. \tag{2}$$

Following the Curry-Schoenberg theorem, a spline regression function $f \in S_{\mathbf{t}_{k,n}}$, can be expressed as

$$f(\mathbf{t}_{k,n}; x) = \boldsymbol{\theta}' \mathbf{N}_n(x) = \sum_{i=1}^{p} \theta_i N_{i,n}(x),$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)'$ is a vector of real valued regression coefficients and $\mathbf{N}_n(x) = (N_{1,n}(x), \ldots, N_{p,n}(x))'$, $p = n + k$, are B-splines of order n, defined on $\mathbf{t}_{k,n}$. It is well known that $\sum_{i=j-n+1}^{j} N_{i,n}(t) = 1$ for any $t \in [t_j, t_{j+1})$, $j = n, \ldots, n + k$, and $N_{i,n}(t) = 0$ for $t \notin [t_i, t_{i+n}]$. In the sequel, where necessary, we will emphasize the dependence of the spline $f(\mathbf{t}_{k,n}; x)$ on $\boldsymbol{\theta}$ by using the alternative notation $f(\mathbf{t}_{k,n}, \boldsymbol{\theta}; x)$.

The spline regression problem of section 1 can now be more precisely stated as follows. For a fixed order of the spline n, given a sample of observations $\{x_i, y_i\}_{i=1}^{N}$, estimate the number of knots k, their locations $\mathbf{t}_{k,n}$ and the regression coefficients, $\boldsymbol{\theta}$. In order to solve this estimation problem we will alternatively view $f(\mathbf{t}_{k,n}, \boldsymbol{\theta}; x)$ as a special case of a parametric spline curve $\mathbf{Q}(t)$, $t \in [a, b]$. A parametric spline curve $\mathbf{Q}(t)$ is given coordinate-wise as

$$\mathbf{Q}(t) = \{x(t), y(t)\} = \left\{ \sum_{i=1}^{p} \xi_i N_{i,n}(t), \sum_{i=1}^{p} \theta_i N_{i,n}(t) \right\}, \tag{3}$$

where t is a parameter, and $x(t)$ and $y(t)$ are spline functions, defined on one and the same set of knots $\mathbf{t}_{k,n}$, with coefficients ξ_i and θ_i, $i = 1, \ldots, p$, respectively. If the coefficients ξ_i in (3) are chosen to be the knot averages

$$\xi_i^* = (t_{i+1} + \cdots + t_{i+n-1})/(n-1), \quad i = 1, \ldots, p, \tag{4}$$

then it is possible to show that the identity

$$x(t) = \sum_{i=1}^{p} \xi_i^* N_{i,n}(t) = t, \tag{5}$$

referred to as the linear precision property of B-splines, holds (see e.g. De Boor 2001). The values ξ_i^* given by (4) are known as the Greville abscissae. We will alternatively use the notation $\boldsymbol{\xi}^*(\mathbf{t}_{k,n})$, to indicate the dependence of the set $\boldsymbol{\xi}^*$ on the knots $\mathbf{t}_{k,n}$. In view of (3) and (5), the spline regression function $f(\mathbf{t}_{k,n}, \boldsymbol{\theta}; x)$ can be expressed as a parametric spline curve

$$\mathbf{Q}^*(t) = \{t, f(\mathbf{t}_{k,n}, \boldsymbol{\theta}; t)\} = \left\{ \sum_{i=1}^{p} \xi_i^* N_{i,n}(t), \sum_{i=1}^{p} \theta_i N_{i,n}(t) \right\}, \tag{6}$$

where $t \in [a, b]$. In what follows, it will be convenient to use $\mathbf{Q}^*(t)$ and $f(\mathbf{t}_{k,n}, \boldsymbol{\theta}; t)$ interchangeably to denote a functional spline regression curve.

Interpretation (6), of the regression function $f(\mathbf{t}_{k,n}, \boldsymbol{\theta}; x)$ as a parametric spline curve $\mathbf{Q}^*(t)$, allows us to characterize the spline regression curve $\mathbf{Q}^*(t)$ by a polygon, with vertexes $\mathbf{c}_i = (\xi_i^*, \theta_i)$, $i = 1, \ldots, p$, which is closely related to $\mathbf{Q}^*(t)$, and is called the control polygon of $\mathbf{Q}^*(t)$, denoted by $\mathbf{C}_{\mathbf{Q}^*}(t)$, (see Figure 1). This relationship is due to the fact that both the x and y coordinates of the control points $\mathbf{c}_i$, $i = 1, \ldots, p$, are related to the spline regression curve $\mathbf{Q}^*(t)$. More precisely, the x-coordinates, ξ_i^*, are the Greville sites (4), obtained from the knots $\mathbf{t}_{k,n}$, and the y-coordinates, θ_i, are simply the spline regression coefficients. Due to the partition of unity property of B-splines, $\sum_{i=j-n+1}^{j} N_{i,n}(t) = 1$ for any $t \in [t_j, t_{j+1})$, $j = n, \ldots, n+k$, the curve $\mathbf{Q}^*(t)$ of order n is a convex combination of n of its control points, and its graph lies within the convex hull of its control polygon $\mathbf{C}_{\mathbf{Q}^*}$. The convex hull of $\mathbf{c}_1, \ldots, \mathbf{c}_p$ is the smallest convex polygon, enclosing these points. Thus, the higher is the degree, $n-1$, the stronger is the curve's deviation from its control polygon $\mathbf{C}_{\mathbf{Q}^*}$, but it still remains within the convex hull of $\mathbf{C}_{\mathbf{Q}^*}$. This suggests that a quadratic B-spline curve is very well suited as a compromise between smoothness and shape preservation. Hence, due to the convex hull property, the curve is in a close vicinity of its control polygon as illustrated in Figure 1 with respect to two adjacent polynomial segments of $\mathbf{Q}^*(t)$. The grey areas in Figure 1 are the two convex hulls, formed by $\mathbf{c}_3, \mathbf{c}_4, \mathbf{c}_5$ and $\mathbf{c}_4, \mathbf{c}_5, \mathbf{c}_6$, within which the two segments of $\mathbf{Q}^*(t)$, for $t \in [t_5, t_6]$ and $t \in [t_6, t_7]$, lie. In fact, the control polygon, $\mathbf{C}_{\mathbf{Q}^*}$, with vertices $\mathbf{c}_i = (\xi_i^*, \theta_i)$ is itself a linear spline function,

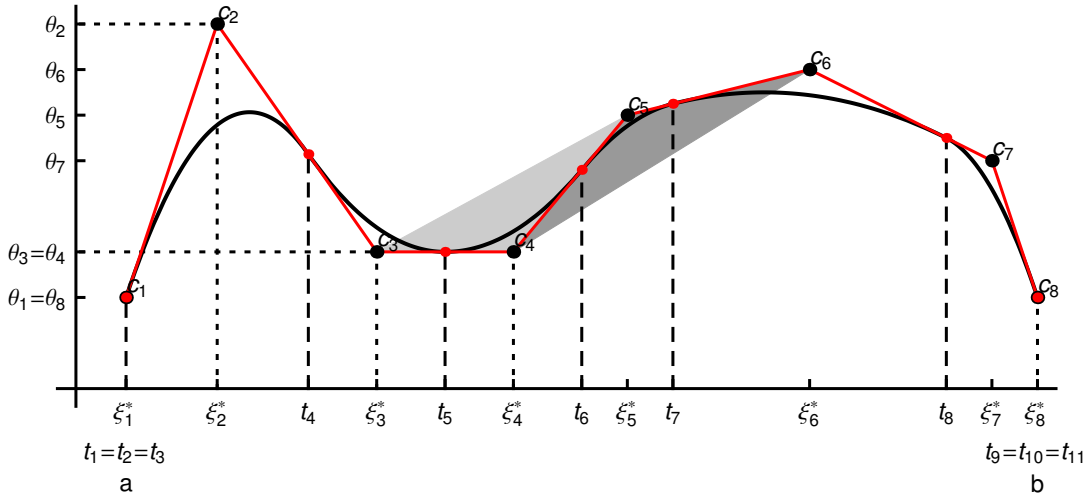


Figure 1: A quadratic ($n = 3$), functional spline regression curve $\mathbf{Q}^*(t)$, with $k = 5$ internal knots, and its control polygon $\mathbf{C}_{\mathbf{Q}^*}(t)$. The spline coefficients $\boldsymbol{\theta}$ and the set of knots $\boldsymbol{t}_{5,3}$ are arbitrarily chosen, and the Greville sites $\boldsymbol{\xi}^*(\mathbf{t}_{5,3})$ are evaluated following (4).

and hence can be expressed as

$$\mathbf{C}_{\mathbf{Q}^*}(t) = \left\{ \sum_{i=1}^{p} \xi_i^* N_{i,2}(t), \sum_{i=1}^{p} \theta_i N_{i,2}(t) \right\} = \left\{ t, \sum_{i=1}^{p} \theta_i N_{i,2}(t) \right\} \equiv \sum_{i=1}^{p} \theta_i N_{i,2}(t). \tag{7}$$

In (7), $\sum_{i=1}^{p} \xi_i^* N_{i,2}(t) = t$ since $N_{i,2}(t)$ are defined over the knots $\mathbf{t}_{p-2,2}$, where $t_1 \equiv \xi_1^*$, $t_{p+2} \equiv \xi_p^*$ and $t_{i+1} \equiv \xi_i^*$, $i = 1, \ldots, p$ and the linear precision property (5) applies. For further details related to geometric modelling with splines we refer to Cohen et al. (2001).

Next, we emphasize that the spline regression curve $\mathbf{Q}^*(t)$ not only lies within the convex hull of $\mathbf{C}_{\mathbf{Q}^*}(t)$ but it follows its shape. That is, because $\mathbf{Q}^*(t)$ is the Schoenberg's variation diminishing spline approximation of its control polygon $\mathbf{C}_{\mathbf{Q}^*}(t)$, i.e.

$$V\left[\mathbf{C}_{\mathbf{Q}^*}\right](t) = \sum_{i=1}^{p} \mathbf{C}_{\mathbf{Q}^*}(\xi_i^*) N_{i,n}(t) = \sum_{i=1}^{p} \theta_i N_{i,n}(t) \equiv \mathbf{Q}^*(t), \tag{8}$$

where ξ_i^*, $i = 1, 2, \ldots, p$, are the Greville abscissae, obtained from $\mathbf{t}_{k,n}$ and $\mathbf{C}_{\mathbf{Q}^*}(\xi_i^*) = \theta_i$ by the definition of the control polygon. Given a set of knots, $\mathbf{t}_{k,n}$, the spline approximation

$$V[g](x) = \sum_{i=1}^{p} g(\xi_i^*) N_{i,n}(x) \tag{9}$$

of any function g, defined on $[a, b]$, is known as the Schoenberg's variation diminishing spline

(VDS) approximation of order n to g, on the set of knots $\mathbf{t}_{k,n}$, see e.g. De Boor (2001). It is constructed by simply evaluating g at the Greville sites (4) and taking the values $g(\xi_i^*)$ as the B-spline coefficients of the VDS approximation.

It is important to recall a property of $V[g]$, which is crucial for developing the GeD estimator. That is, the VDS approximation, $V[g]$ is *shape preserving* since it preserves the shape of the function g it approximates. More precisely, if g is positive, then $V[g]$ is also positive, if g is monotone, then $V[g]$ is also monotone, and if g is convex, $V[g]$ is also convex. The *variation diminishing* character of $V[g]$ is due to the fact that it crosses any straight line at most as many times as does the function g itself. In view of the convex hull property and the shape preserving property of (8) it is now clear that $\mathbf{Q}^*(t)$ not only lies close to its control polygon $\mathbf{C}_{\mathbf{Q}^*}(t)$ but also preserves its shape. In fact, in the linear case, i.e., when $n = 2$, the parametric curve, $\mathbf{Q}^*(t)$ defined by (6) coincides with its control polygon $\mathbf{C}_{\mathbf{Q}^*}(t)$, i.e., $\mathbf{Q}^*(t) = \mathbf{C}_{\mathbf{Q}^*}(t)$, since the Greville sites in (4) coincide with the knots, i.e. $\xi_i^* = t_{i+1}$, $i = 1, \dots p$.

In summary, it has been established that the spline regression function $f(\mathbf{t}_{k,n}, \boldsymbol{\theta}; x)$, (alternatively denoted as $\mathbf{Q}^*(x)$, $x \in [a, b]$), can be expressed in the form (6) and that its control polygon, $\mathbf{C}_{f(\mathbf{t}_{k,n},\boldsymbol{\theta};x)}$, has vertices $\mathbf{c}_i = (\xi_i^*, \theta_i)$, $i = 1, \dots, p$, where ξ_i^* are the Greville sites (4), obtained from $\mathbf{t}_{k,n}$. The latter suggests that, given n and k, locating the knots $\mathbf{t}_{k,n}$ and finding the regression coefficients $\boldsymbol{\theta}$ of $f(\mathbf{t}_{k,n}, \boldsymbol{\theta}; x)$, based on the set of observations $\{y_i, x_i\}_{i=1}^N$, is equivalent to finding the location of the x- and y-coordinates of the vertices of $\mathbf{C}_{f(\mathbf{t}_{k,n},\boldsymbol{\theta};x)}$. This establishes the important fact that estimation of $\mathbf{t}_{k,n}$ and $\boldsymbol{\theta}$ affects the geometrical position of the control polygon $\mathbf{C}_{f(\mathbf{t}_{k,n},\boldsymbol{\theta};x)}$, which, due to the shape preserving and convex hull properties, defines the location of the spline curve $f(\mathbf{t}_{k,n}, \boldsymbol{\theta}; x)$. Inversely, locating the vertices $\mathbf{c}_i$ of $\mathbf{C}_{f(\mathbf{t}_{k,n},\boldsymbol{\theta};x)}$ affects the knots $\mathbf{t}_{k,n}$, through (4), and the values of $\boldsymbol{\theta}$, and hence affects the position of the regression curve $f(\mathbf{t}_{k,n}, \boldsymbol{\theta}; x)$. The latter conclusion motivates the construction, in stage A of GeDS, of a control polygon as a linear least squares spline fit to the data, whose knots determine the knots $\mathbf{t}_{k,n}$, and whose B-spline coefficients are viewed as an initial estimate of $\boldsymbol{\theta}$, which is improved further in stage B (see section 3). This is the basis of our approach to constructing the GeD variable knot spline approximation to the unknown function f in (1), and this is developed in the next section.

3 Geometrically designed spline regression.

In this section, we introduce the GeD spline regression method which is motivated by the ideas, outlined in section 2. The method "positions" first an initial control polygon (linear spline fit), which reproduces the "shape" of the data, applying least squares approximation. Secondly, an optimal set of knots of a higher order ($n > 2$) smooth spline curve is found, so that it preserves the shape of the initial control polygon and then this curve is fitted to the data, to adjust its position in the LS sense. In this way, it is ensured that the n-th order smooth LS fit follows the shape of the initial control polygon, and hence the shape of the data. This procedure simultaneously produces linear, quadratic, cubic, or higher order splines and the LS fit with the minimum residual sum of squares is chosen as the final fit which recovers best the underlying unknown function f. It should be mentioned that all the fits (linear, quadratic etc.) are constructed so that they have dimension p equal to the dimension of the linear fit constructed first. This is achieved by positioning the knots of the higher order, ($n > 2$) spline fits according to the averaging relation (20), as described in section 3.1. The two stages of the GeDS approach may be given a formal interpretation as certain optimization problems with respect to the variables k, $\mathbf{t}_{k,n}$, $\boldsymbol{\theta}$ and n.

Stage A. Starting from a straight line fit and adding one knot at a time, find the least squares linear spline fit $\hat{f}\left(\boldsymbol{\delta}_{l,2}, \hat{\boldsymbol{\alpha}}; x\right) = \sum_{i=1}^{p} \hat{\alpha}_i N_{i,2}(x)$ with a number of internal knots l, number of B-splines $p = l + 2$ and with a set of knots $\boldsymbol{\delta}_{l,2} = \{\delta_1 = \delta_2 < \delta_3 < \ldots < \delta_{l+2} < \delta_{l+3} = \delta_{l+4}\}$, such that the ratio of the residual sums of squares

$$\mathrm{RSS}(l+q)/\mathrm{RSS}(l) = \sum_{j=1}^{N} \left(y_j - \hat{f}(\boldsymbol{\delta}_{l+q,2}; x_j)\right)^2 \Big/ \sum_{j=1}^{N} \left(y_j - \hat{f}(\boldsymbol{\delta}_{l,2}; x_j)\right)^2 \geq \alpha_{\text{exit}} \qquad (10)$$

where $q \geq 1$ and α_{exit} is a certain threshold level. Testing the inequality in (10) serves as the stage A model selector. If the inequality in (10) is fulfilled, $\hat{f}\left(\boldsymbol{\delta}_{l,2}, \hat{\boldsymbol{\alpha}}; x\right)$ could not be significantly improved if q more knots are added and $\hat{f}\left(\boldsymbol{\delta}_{l,2}, \hat{\boldsymbol{\alpha}}; x\right)$ is the selected model which adequately reproduces the "shape" of the unknown, underlying function f.

In order to realize stage A, a locally-adaptive knot insertion scheme is proposed. It starts from an LS fit, in the form of a straight line segment, as described in step 1 below. The latter is then sequentially "broken" into a piece-wise linear LS fit, by adding knots, one at a time, at

those points, where the fit deviates most from the shape of the underlying function, (see Figure 2), according to a bias driven measure of appropriately defined clusters of residuals (see steps 2 - 8). A stopping rule is introduced, which serves as a model selector and allows us to determine the appropriate number and location of the knots of the linear spline fit $\hat{f}\left(\boldsymbol{\delta}_{l,2}, \hat{\boldsymbol{\alpha}}; x\right)$ (see steps 9 - 10). A formal description is given next.

Step 1. Let $n = 2$, $k = 0$. Find the LS spline fit $\hat{f}\left(\boldsymbol{\delta}_{0,2}, \hat{\boldsymbol{\alpha}}; x\right) = \hat{\alpha}_1 N_{1,2}(x) + \hat{\alpha}_2 N_{2,2}(x)$, where $\boldsymbol{\delta}_{0,2} = \{\delta_i\}_{i=1}^4$ with $a = \delta_1 = \delta_2 < \delta_3 = \delta_4 = b$. Calculate the residuals $r_i \equiv r(x_i) = y_i - \hat{f}\left(\boldsymbol{\delta}_{0,2}, \hat{\boldsymbol{\alpha}}; x_i\right)$, $i = 1, \ldots, N$ and the residual sum of squares $\mathrm{RSS}(k) = \sum_{i=1}^N r_i^2$ of the fit with k internal knots.

Step 2. Group the consecutive residuals r_i, $i = 1, \ldots, N$ into clusters by their sign, i.e. find a number u, $1 \le u \le N$ and a set of integer values $d_j > 0$, $j = 1, \ldots, u$ such that $\mathrm{sign}(r_1) = \ldots = \mathrm{sign}(r_{d_1}) \neq \mathrm{sign}(r_{d_1+1}) = \mathrm{sign}(r_{d_1+2}) = \ldots = \mathrm{sign}\,(r_{d_1+d_2}) \neq \ldots \neq \mathrm{sign}(r_{d_1+d_2+\cdots+d_{u-1}+1}) = \mathrm{sign}(r_{d_1+d_2+\cdots+d_{u-1}+2}) = \ldots = \mathrm{sign}(r_{d_1+d_2+\cdots+d_u})$, and $\sum_{j=1}^u d_j = N$.

Step 3. For each of the u clusters of residuals of identical signs, calculate the within-cluster mean residual value

$$\begin{aligned} m_j &= \left(\sum_{i=1}^{d_j} r_{d(j-1)+i}\right) \Big/ d_j = \sum_{i=1}^{d_j} \left(\left(\hat{f}_{d(j-1)+i} - E\hat{f}_{d(j-1)+i}\right) + \left(E\hat{f}_{d(j-1)+i} - f_{d(j-1)+i}\right) + \epsilon_{d(j-1)+i} \right) \Big/ d_j \\ &= \sum_{i=1}^{d_j} \left(\hat{f}_{d(j-1)+i} - E\hat{f}_{d(j-1)+i}\right)/d_j + \sum_{i=1}^{d_j} \left(E\hat{f}_{d(j-1)+i} - f_{d(j-1)+i}\right)/d_j + \sum_{i=1}^{d_j} \epsilon_{d(j-1)+i}/d_j, \end{aligned}$$

$j = 1, \ldots, u$, where $d(j) = d_1 + d_2 + \cdots + d_j$ and the three terms in the last decomposition can be interpreted as the within-cluster average variance, bias, and error, respectively. Calculate also the within-cluster range (with $d(0) = 0$)

$$\eta_j = x_{d(j)} - x_{d(j-1)+1}, \qquad j = 1, \ldots, u.$$

Step 4. Find $m_{\max} = \max_{1 \le j \le u}(m_j)$ and $\eta_{\max} = \max_{1 \le j \le u}(\eta_j)$ and calculate, correspondingly, the normalized within-cluster mean and range values $m'_j = m_j / m_{\max}$ and $\eta'_j = \eta_j / \eta_{\max}$, so that $0 < m'_j \le 1$, $0 < \eta'_j \le 1$.

Step 5. Calculate the cluster weights

$$w_j = \beta m'_j + (1-\beta)\eta'_j, \quad j = 1, \ldots, u, \tag{11}$$

where, β is a real valued parameter, $0 \leq \beta \leq 1$. The value w_j serves as a measure, attached to the j-th cluster of residuals of identical sign, which measures the deviation of the current least squares linear spline fit $\hat{f}\left(\boldsymbol{\delta}_{k,2}, \hat{\boldsymbol{\alpha}}; x\right)$ from f in the j-th cluster. The weight β is one of the parameters whose value needs to be chosen at the start of stage A.

Step 6. Order the weights of the clusters, i.e. $w_{j_1} \geq w_{j_2} \geq \ldots \geq w_{j_u}$. Thus, in order to improve $\hat{f}\left(\boldsymbol{\delta}_{k,2}, \hat{\boldsymbol{\alpha}}; x\right)$, in the next step a new knot is inserted, at an appropriate location, within the j_1-th cluster.

Step 7. Check whether there is already a knot within the j_1-th cluster, i.e. check whether $\delta_i \in \left[x_{d(j_1-1)+1}, x_{d(j_1-1)+d_{j_1}}\right]$, for each internal knot $\delta_i \in \boldsymbol{\delta}_{k,2}$, $i = 3, \ldots, k+2$. If there is already a knot within the j_1-th cluster, the check is repeated for the cluster with index j_2, and so on until the first cluster, say indexed j_s, is found which does not contain a knot. Then, insert a new knot δ^* at the site

$$\delta^* = \left(\sum_{i=d(j_s-1)+1}^{d(j_s-1)+d_{j_s}} r_i x_i\right) \Big/ \left(\sum_{i=d(j_s-1)+1}^{d(j_s-1)+d_{j_s}} r_i\right). \tag{12}$$

Note that $\delta^* \in \left[x_{d(j_s-1)+1}, x_{d(j_s-1)+d_{j_s}}\right]$. The new knot position (12) can be viewed as a weighted average of the x-coordinates of the residuals in the j_s-th cluster, the weights being the normalized values of the residuals. The set of knots $\boldsymbol{\delta}_{k,2}$ is being updated as $\boldsymbol{\delta}^*_{k+1,2} := \boldsymbol{\delta}_{k,2} \bigcup \{\delta^*\}$.

So, a new knot is placed where the cluster weight (11) is maximal. In view of the decomposition in step 3, the cluster weight (11) can be referred to as a bias dominated measure since the bias component is dominant in this cluster compared to the variance and error terms (at least at the initial iterations when there are small number of knots in the linear fit and the approximation error is large).

Step 8. Find the least squares linear spline fit

$$\hat{f}\left(\boldsymbol{\delta}^*_{k+1,2}, \hat{\boldsymbol{\alpha}}; x\right) = \sum_{i=1}^{p} \hat{\alpha}_i N_{i,2}(x).$$

Since $\boldsymbol{\delta}^*_{k+1,2}$ contains the new knot, the number of B-splines, p, will increase by one.

Step 9. Calculate the residuals r_i, $i = 1, \ldots, N$ and the RSS$(k+1)$ for $\hat{f}\left(\boldsymbol{\delta}^*_{k+1,2}, \hat{\boldsymbol{\alpha}}; x\right)$. Note that $\boldsymbol{\delta}_{k,2} \subset \boldsymbol{\delta}^*_{k+1,2}$ implies that $S_{\boldsymbol{\delta}_{k,2}} \subset S_{\boldsymbol{\delta}^*_{k+1,2}}$. Hence $\hat{f}\left(\boldsymbol{\delta}_{k,2}, \hat{\boldsymbol{\alpha}}; x\right) \in S_{\boldsymbol{\delta}^*_{k+1,2}}$ and applying the orthogonality property of least squares estimation it is easy to see that

$$\sum_{i=1}^{N}\left(y_i - \hat{f}\left(\boldsymbol{\delta}_{k,2}, \hat{\boldsymbol{\alpha}}; x_i\right)\right)^2 = \sum_{i=1}^{N}\left(y_i - \hat{f}\left(\boldsymbol{\delta}^*_{k+1,2}, \hat{\boldsymbol{\alpha}}; x_i\right)\right)^2 + \sum_{i=1}^{N}\left(\hat{f}\left(\boldsymbol{\delta}^*_{k+1,2}, \hat{\boldsymbol{\alpha}}; x_i\right) - \hat{f}\left(\boldsymbol{\delta}_{k,2}, \hat{\boldsymbol{\alpha}}; x_i\right)\right)^2. \tag{13}$$

Equation (13) implies that RSS$(k+1)<$RSS(k). It is obvious also that RSS(k) will converge to zero as $k + n \to N$ since when $k + n = N$ the fit interpolates the data. The greedy fashion of the new knot placement (12), combined with equation (13), gives rise to the rule for exit from stage A of the GeDS method, given next.

Step 10. Let $q \geq 1$ be a fixed integer, chosen at the beginning of stage A. If $k \leq q$ go back to step 2, otherwise calculate the ratio

$$\alpha = \text{RSS}(k+1)/\text{RSS}(k+1-q).$$

Note that from (13) it follows that $0 < \alpha < 1$. If $\alpha \geq \alpha_{\text{exit}}$, an exit from stage A is performed with the spline fit $\hat{f}\left(\boldsymbol{\delta}_{l,2}, \hat{\boldsymbol{\alpha}}; x\right)$, $l = k + 1 - q$. If $\alpha < \alpha_{\text{exit}}$ then $\hat{f}\left(\boldsymbol{\delta}^*_{k+1,2}, \hat{\boldsymbol{\alpha}}; x\right)$ is taken as the current fit and the procedure goes back to step 2. The value α_{exit} is chosen ex ante to be close to 1. This is because the ratio α will be close to zero if the fit has improved significantly by adding δ^* and will tend to 1 if no improvement has been achieved in the last $q + 1$ consecutive iterations, i.e the corresponding values of the RSS have stabilized. Our experience has shown that this stopping rule works well as a model selector with $q = 2$, i.e., stabilization of RSS in three consecutive iterations is sufficient to exit from stage A with the appropriate number of knots.

Remark 3.1 *Let us note that in the Normal error case, i.e. when $\epsilon \in \mathcal{N}\left(0, \sigma^2_\epsilon\right)$, the log-likelihood, logL$(\boldsymbol{\delta}_{k,2})$, of a current linear spline fit $\hat{f}\left(\boldsymbol{\delta}_{k,2}, \hat{\boldsymbol{\alpha}}; x\right)$ with k knots is proportional to its corresponding residual sum of squares, RSS(k), i.e. logL$(\boldsymbol{\delta}_{k,2}) \propto$-RSS(k) (see e.g. relation (6) of Hansen and Kooperberg 2002). Therefore, the ratio of RSS in the left-hand side of (10) can be interpreted as the ratio of the log-likelihoods of the two nested spline fits $\hat{f}\left(\boldsymbol{\delta}_{l+q,2}, \hat{\boldsymbol{\alpha}}; x\right)$ and $\hat{f}\left(\boldsymbol{\delta}_{l,2}, \hat{\boldsymbol{\alpha}}; x\right)$ or as the (normal) deviance in the broader GLM context.*

Remark 3.2 *The knot insertion scheme in stage A can be described as a "greedy" one (see Hastie 1989), since at each iteration it places a knot, δ^*, where a within-cluster bias dominated measure is maximal (see steps 3 and 5), which is very near to the site where placing a knot gives the largest reduction in the residual sum of squares. This can be quantified using the fact that given an LS fit $\hat{f}\left(\boldsymbol{\delta}_{k,2}; x\right)$, with $0 < k < l$ internal knots, if a knot, δ^*, is added in the interval $[\delta_{j^*}, \delta_{j^*+1}]$, $2 \leq j^* < k+2$, then the updated LS fit $\hat{f}\left(\boldsymbol{\delta}^*_{k+1,2}; x\right)$ adjusts best to the data in $[\delta_{j^*}, \delta_{j^*+1}]$, since $\left|\hat{f}\left(\boldsymbol{\delta}_{k,2}; x\right) - \hat{f}\left(\boldsymbol{\delta}^*_{k+1,2}; x\right)\right|$, $x \in [\delta_j, \delta_{j+1}]$ decreases exponentially in $|j^* - j|$, which is the number of knots between x and δ^*. This follows from Theorem 1 of Zhou and Shen (2001).*

As an alternative to the model selector in (10), (see also step 10 of stage A), we have implemented two additional model selection criteria based on the minimization of SURE and GCV. More precisely, the GeD linear spline fit with number of knots $k = k_{\min}$ which minimizes Stein's unbiased risk estimate (SURE)

$$R(\hat{f}) = \sum_{i=1}^{N} \left(y_i - \hat{f}\left(\boldsymbol{\delta}_{k,2}, \hat{\boldsymbol{\alpha}}; x_i\right)\right)^2 / N + D\frac{k+n}{N}\sigma^2 \tag{14}$$

or the generalized cross validation

$$\mathrm{GCV}(\hat{f}) = \left(\sum_{i=1}^{N} \left(y_i - \hat{f}\left(\boldsymbol{\delta}_{k,2}, \hat{\boldsymbol{\alpha}}; x_i\right)\right)^2 / N\right) \Big/ \left(1 - \frac{d(k)}{N}\right)^2 \tag{15}$$

criterion is selected as the outcome of stage A. We have assumed that the minimum is attained when SURE or GCV do not decrease in two consecutive iterations in stage A. Criteria (14) and (15) depend on the choice of the parameters D and $d(k)$, and when $D = 1.2$ and $d(k) = k + 1$ they behave roughly as the criterion in (10), as illustrated by Figure 5 (a) and the related results in the online supplement. The choice $D = 3$ and $d(k) = 3k + 1$, as noted by Zhou and Shen (2001) tends to yield a model underfitting the underlying function f. In section 4 and in the online supplement to the paper, we give a comparison of the three model selection approaches, (10), (14) and (15), and a thorough sensitivity study of the GeD spline estimation with respect to the choice of the free parameters, α_{exit} and β (see steps 10 and 5 of stage A).

If a smoother fit is needed, the linear LS spline fit $\hat{f}\left(\boldsymbol{\delta}_{l,2}, \hat{\boldsymbol{\alpha}}; x\right)$ is viewed as a control polygon with vertices $(\delta_{i+1}, \hat{\alpha}_i)$, $i = 1, \ldots, p$, (recall that $p = l + 2$) and a higher order GeD spline is

constructed in stage B, based on the geometrical form of $\hat{f}\left(\boldsymbol{\delta}_{l,2}, \hat{\boldsymbol{\alpha}}; x\right)$.

Stage B.1. For each of the values of $n = 3, \ldots, n_{\max}$, find the optimal set of knots $\tilde{\mathbf{t}}_{l-(n-2),n}$, as a solution of the constrained minimization problem

$$\min_{\substack{\mathbf{t}_{l-(n-2),n}, \\ \delta_{i+2}<t_{i+n}<\delta_{i+n}, \\ i=1,\ldots,(l-(n-2))}} \left\| \hat{f}\left(\boldsymbol{\delta}_{l,2}, \hat{\boldsymbol{\alpha}}; x\right) - \mathbf{C}_{f\left(\mathbf{t}_{l-(n-2),n}, \hat{\boldsymbol{\alpha}}; x\right)} \right\|_{\infty}, \tag{16}$$

where $\|g\|_{\infty} := \max_{a \le x \le b} |g(x)|$ is the uniform (L_{∞}) norm of a function $g(x)$. In fact, minimization in (16) gives the polygon $\mathbf{C}_{f\left(\tilde{\mathbf{t}}_{l-(n-2),n}, \hat{\boldsymbol{\alpha}}; x\right)}$ with vertices $(\xi_i^*, \hat{\alpha}_i)$, so that

$$\xi_i^*(\tilde{\mathbf{t}}_{l-(n-2),n}) \simeq \delta_{i+1}, i = 1, \ldots, p \tag{17}$$

or equivalently, so that

$$t = \sum_{i=1}^{p} \xi_i^* N_{i,n}(t) \simeq \sum_{i=1}^{p} \delta_{i+1} N_{i,n}(t). \tag{18}$$

Remark 3.3 *Achieving exact equality in (17), respectively (18), is impossible in general since expressing* $\xi_i^*(\mathbf{t}_{l-(n-2),n})$ *following (4), we have that,* $\xi_1^* = \delta_2 = a$, $\xi_p^* = \delta_{p+1} = b$ *and*

$$\delta_{i+1} = \left(t_{i+1} + \ldots + t_{i+n-1}\right) / \left(n-1\right), i = 2, \ldots, p-1, \tag{19}$$

which is an over-determined system of $p-2 = l$ *equations and* $l-(n-2)$ *ordered, unknown knots,* $\mathbf{t}_{l-(n-2),n}$, $(n > 2)$. *Instead, what is achieved by solving (16) is that* $\mathbf{C}_{f\left(\tilde{\mathbf{t}}_{l-(n-2),n}, \hat{\boldsymbol{\alpha}}; x\right)}$ *is as close to* $\hat{f}\left(\boldsymbol{\delta}_{l,2}, \hat{\boldsymbol{\alpha}}; x\right)$ *as possible because* $\boldsymbol{\xi}^*(\tilde{\mathbf{t}}_{l-(n-2),n})$ *is as close to* $\boldsymbol{\delta}_{l,2}$ *as possible. Note that, since we view* δ_{i+1}, $i = 1, \ldots, p$, *as Greville sites of a higher order spline function* $f\left(\mathbf{t}_{l-(n-2),n}, \hat{\boldsymbol{\alpha}}; x\right)$, *the constraints* $\delta_{i+2} < t_{i+n} < \delta_{i+n}$, $i = 1, \ldots, (l-(n-2))$ *in (16), follow from (19).*

In summary, the objective of stage B is to produce a set of optimal knots $\tilde{\mathbf{t}}_{l-(n-2),n}$, which ensures that $\hat{f}\left(\boldsymbol{\delta}_{l,2}, \hat{\boldsymbol{\alpha}}; x\right)$ becomes (nearly) the control polygon of the spline function $f\left(\tilde{\mathbf{t}}_{l-(n-2),n}, \hat{\boldsymbol{\alpha}}; x\right)$, i.e. that $\mathbf{C}_{f\left(\tilde{\mathbf{t}}_{l-(n-2),n}, \hat{\boldsymbol{\alpha}}; x\right)} \simeq \hat{f}\left(\boldsymbol{\delta}_{l,2}, \hat{\boldsymbol{\alpha}}; x\right)$. Or equivalently, $\tilde{\mathbf{t}}_{l-(n-2),n}$ is placed so that $f\left(\tilde{\mathbf{t}}_{l-(n-2),n}, \hat{\boldsymbol{\alpha}}; x\right)$ becomes (nearly) the Schoenberg's VDS approximation of $\hat{f}\left(\boldsymbol{\delta}_{l,2}, \hat{\boldsymbol{\alpha}}; x\right)$. Therefore, due to its convex hull and shape preserving properties (see section 2), $f\left(\tilde{\mathbf{t}}_{l-(n-2),n}, \hat{\boldsymbol{\alpha}}; x\right)$ lies very close to $\hat{f}\left(\boldsymbol{\delta}_{l,2}, \hat{\boldsymbol{\alpha}}; x\right)$, and hence to the "shape" of the data for which the linear LS approximation is $\hat{f}\left(\boldsymbol{\delta}_{l,2}, \hat{\boldsymbol{\alpha}}; x\right)$ (according to stage A). This is the fundamental concept of optimal knot place-

ment in Stage B of GeDS. The fact that $f\left(\tilde{\mathbf{t}}_{l-(n-2),n}, \hat{\boldsymbol{\alpha}}; x\right)$ is nearly a VDS approximation of $\hat{f}\left(\boldsymbol{\delta}_{l,2}, \hat{\boldsymbol{\alpha}}; x\right)$ has been quantified with appropriate error bounds but due to volume limitation these are omitted, and bounds are only given for the knot location $\bar{\mathbf{t}}_{l-(n-2),n}$ as described in the next section 3.1.

However, we note that $f\left(\tilde{\mathbf{t}}_{l-(n-2),n}, \hat{\boldsymbol{\alpha}}; x\right)$ is not a least squares approximation to the data set. In order to obtain a least squares fit to the data and at the same time to preserve the shape of $f\left(\tilde{\mathbf{t}}_{l-(n-2),n}, \hat{\boldsymbol{\alpha}}; x\right)$, its optimal knots $\tilde{\mathbf{t}}_{l-(n-2),n}$ are preserved, whereas its B-spline coefficients, $\hat{\boldsymbol{\alpha}}$, are released and are assumed to be unknown parameters, $\boldsymbol{\theta}$, which are estimated in the least squares sense, using $\{x_i, y_i\}_{i=1}^N$, as follows.

Stage B.2. For a fixed $n = 3, \ldots, n_{\max}$, find the least squares fit $\hat{f}\left(\tilde{\mathbf{t}}_{l-(n-2),n}, \hat{\boldsymbol{\theta}}; x\right)$ which solves

$$\min_{\boldsymbol{\theta}} \sum_{j=1}^{N} \left(y_j - f\left(\tilde{\mathbf{t}}_{l-(n-2),n}, \boldsymbol{\theta}; x_j\right)\right)^2.$$

Among all fits $\hat{f}\left(\tilde{\mathbf{t}}_{l-(n-2),n}, \hat{\boldsymbol{\theta}}; x\right)$, of order $n = 2, \ldots, n_{\max}$, i.e. including the linear fit, $\hat{f}\left(\boldsymbol{\delta}_{l,2}, \hat{\boldsymbol{\alpha}}; x\right)$ from stage A, choose the one of order $\hat{n}$, for which the RSS is minimal.

In this way in stage B.2, along with the number of knots and their locations, the degree of the spline is also estimated. This is an important feature of the proposed estimation method which is rarely offered by other spline estimation procedures. One alternative that we are aware of is the MDL method of Lee (2000) (see also Lee 2002a,b). Of course, any of the produced final fits of order $n \neq \hat{n}$ could be used, if other features were more desirable, for example if better smoothness were required.

3.1 Stage B.1 - Approximate solution: the averaging knot location.

The minimization problem (16), of stage B.1, is a constrained non-linear optimization problem with respect to the knots and although it is related to linear splines, it is still computationally involved. In addition, as with any other non-linear optimization problem, finding the globally optimal solution is not guaranteed. The knots $\tilde{\mathbf{t}}_{l-(n-2),n}$, which are the optimal solution, may also be (almost) coalescent and this may cause edges and corners in the final LS fit in stage B.2. In order to avoid these undesirable features, but to preserve the optimality properties of the knots, as described in stage B.1, we propose to place the knots in stage B.1 according to (20), which we call the *averaging knot location*, and which gives an easy to evaluate, approximate

solution to the minimization problem (16), respectively (19). Thus, the final GeD spline fit from stage B.2 is $\hat{f}\left(\bar{\mathbf{t}}_{l-(\hat{n}-2),\hat{n}}, \hat{\boldsymbol{\theta}}; x\right)$, where $\bar{\mathbf{t}}_{l-(\hat{n}-2),\hat{n}}$ is given by (20) with $n = \hat{n}$.

The averaging knot location: Given the control polygon $\hat{f}\left(\boldsymbol{\delta}_{l,2}, \hat{\boldsymbol{\alpha}}; x\right)$ of stage A, for each $n = 3, \ldots, n_{\max}$, calculate the knot placement $\bar{\mathbf{t}}_{l-(n-2),n}$ with internal knots defined as the averages of the knots, $\boldsymbol{\delta}_{l,2}$, i.e.,

$$\bar{t}_{i+n} = (\delta_{i+2} + \cdots + \delta_{i+n})/(n-1), \quad i = 1, \ldots, l-(n-2). \tag{20}$$

The choice of the knots $\bar{\mathbf{t}}_{l-(n-2),n}$ according to (20) makes it possible to establish appropriate error bounds given by Corollaries 3.5 and 3.6 of Theorem 3.4, the proofs of which are provided in Appendix A.

Theorem 3.4 *Let $\{\xi_i\}_{i=1}^{p}$ be an ordered set, $a = \xi_1 < \xi_2 < \ldots < \xi_p = b$, and let $\mathbf{t}_{k,n}$, $(p \geq n \geq 2,\ k = p - n)$, be a set of knots, defined as in (2), with*

$$t_{i+n} = (\xi_{i+1} + \cdots + \xi_{i+n-1})/(n-1), \quad i = 1, \ldots, k. \tag{21}$$

Then, for the n-th order spline approximation $V^a[g]$, defined on $\mathbf{t}_{k,n}$, of a continuous function $g \in C[a,b]$, given by

$$V^a[g](x) = \sum_{i=1}^{p} g(\xi_i) N_{i,n}(x), \tag{22}$$

we have

$$\|V[g] - V^a[g]\|_\infty \leq \left\lceil \frac{(n-2)^2}{2(n-1)} \right\rceil \omega(g; max_{j\in\{1,\ldots,p-1\}}(\xi_{j+1} - \xi_j)), \tag{23}$$

where $\lceil \nu \rceil := min\{z \in \mathbb{Z} : \nu \leq z\}$, $V[g]$ is the Schoenberg's VDS approximation, defined on $\mathbf{t}_{k,n}$ and

$$\omega(g; h) := max\{|g(x) - g(y)| : |x - y| \leq h, \quad x, y \in [a,b]\} \tag{24}$$

is the modulus of continuity of the function g at h.

Corollary 3.5 *Under the assumptions of Theorem 3.4, with $n = 3$, ξ_i replaced by δ_{i+1}, $i = 1, \ldots, p$, $p = l + 2$, $k = l - (n-2)$ and if g coincides with the straight line t, i.e. $g \equiv t$, then*

$$\|V[t] - V^a[t]\|_\infty = \left\| t - \sum_{i=1}^{p} \delta_{i+1} N_{i,n}(t) \right\|_\infty \leq \frac{(n-2)^2}{2(n-1)} max_{j\in\{1,\ldots,p-1\}}(\delta_{j+2} - \delta_{j+1}). \tag{25}$$

Corollary 3.6 *Under the assumptions of Theorem 3.4, with ξ_i replaced by δ_{i+1}, $i = 1, \ldots, p$, $p = l+2$, $k = l-1$, $\boldsymbol{\xi}^*(\bar{\boldsymbol{t}}_{l-1,3})$, and assuming that g is the linear spline function $\hat{f}\left(\boldsymbol{\delta}_{l,2}, \hat{\boldsymbol{\alpha}}; x\right) = \sum_{i=1}^{p} \hat{\alpha}_i N_{i,2}(t)$ with vertices $(\delta_{i+1}, \hat{\alpha}_i)$, we have*

$$\begin{aligned} \|V[\hat{f}] - V^a[\hat{f}]\|_\infty &= \left\| \sum_{i=1}^{p} \hat{f}\left(\boldsymbol{\delta}_{l,2}, \hat{\boldsymbol{\alpha}}; \xi_i^*\right) N_{i,3}(x) - \sum_{i=1}^{p} \hat{f}\left(\boldsymbol{\delta}_{l,2}, \hat{\boldsymbol{\alpha}}; \delta_{i+1}\right) N_{i,3}(x) \right\|_\infty \\ &\leq \frac{1}{4} max_{j \in \{2,\ldots,p-1\}} \left|\hat{\alpha}_j - \hat{\alpha}_{j-1}\right|. \end{aligned} \tag{26}$$

We have performed a special simulation study, which illustrates that bound (25) is quite tight (best for $n = 3$) and that the averaging knot location (20) gives a reasonably accurate approximation to the solution of system (19). Due to volume limitations, details of this study are relegated to the online supplement of the paper (see section 1.2 therein). Furthermore, following Huang (2003) one can also study the local asymptotic properties of the proposed GeD spline estimator, but this is beyond the scope of this paper.

4 Numerical Study.

In this section, and in the online supplement to the paper, we illustrate the numerical performance of the GeDS method and its properties, established in section 3, both on simulated and real data examples. We have performed a thorough simulation study of the impact on the GeDS knot location and related MSE (mean square error), of different assumptions and choices made in constructing the GeDS estimate, namely, different levels of the signal to noise ratio (SNR from 2 to 7), sample sizes ($N = 90, 150, 256, 512, 2048$) and levels of smoothness of the underlying function (smooth, medium smooth and wiggly functions), different choices of the GeDS parameters, α_{exit} and β, ($\alpha_{\text{exit}} = 0.8, 0.9, 0.95, 0.99, 0.995, 0.999$, $\beta = 0.3, 0.5, 0.7$) different choice of the model selection criterion (GeDS criterion, GCV and SURE), and different degree of the GeD spline estimate (linear, quadratic and cubic). We have also tested the GeDS knot selection strategy versus equally spaced knots and other established approaches (see the real data example of section 4 and the test examples in the online supplement). We have tested GeDS on a series of simulated examples, used in many other studies on variable knot spline methods (cf. Schwetlick and Schütze 1995, Fan and Gijbels 1995, Donoho and Johnstone 1994 and Luo and Wahba 1997). Due to volume limitations, we present here the results for one of the

simulated test functions, given in Table 1, first considered by Schwetlick and Schütze (1995), and one real data example from materials science, due to Kimber et al. (2009). The detailed results related to the other test functions are given in the online supplement to this paper, see Tables 1 and 2 therein.

As specified in section 3, there are two parameters, $\alpha_{\text{exit}} \in (0,1)$ and $\beta \in (0,1)$, defined in steps 10 and 5 of stage A of GeDS, by means of which the model selection can be controlled. The parameter α_{exit} is related to the model selection rule which determines when to exit from stage A, i.e. it determines the number of knots, l, in the set $\boldsymbol{\delta}_{l,2}$ of the linear spline fit $\hat{f}\left(\boldsymbol{\delta}_{l,2}, \hat{\boldsymbol{\alpha}}; x\right)$ and hence, the number of knots of the final higher order LS spline fit $\hat{f}\left(\overline{\mathbf{t}}_{l-(n-2),n}, \hat{\boldsymbol{\theta}}; x\right)$. The parameter β determines the weight put on the cluster range and the mean cluster size within each cluster of residuals of same sign, as specified in step 5 of stage A. It therefore affects to some extent the ordering of the cluster weights and hence, the knot placement.

As can be presumed, the choice of β depends on the level of the signal-to-noise ratio (SNR), SNR$= (\text{var}(f))^{0.5}/\sigma_\epsilon$ and on the degree of smoothness/wiggliness of f. As can be concluded from the performed numerical tests, when the SNR is high and f is smooth, recommended values are $\beta \in [0.5, 0.7]$, $\alpha_{\text{exit}} \in [0.9, 0.95]$. If the SNR is high and f is a wiggly function then the recommended choice is $\beta \in [0.5, 0.7]$, $\alpha_{\text{exit}} \in [0.99, 0.995]$, since otherwise underfitting may result. In the case when SNR is low and f is smooth, one may use $\beta \in [0.3, 0.5]$, $\alpha_{\text{exit}} \in [0.9, 0.95]$. Finally, it is known that when the SNR is low and the underlying function is very unsmoothy recovering f is very difficult and different choices of β and α_{exit} may need to be attempted.

In summary, in most numerical examples (see section 4.1 and the online supplement), GeDS produces very good results avoiding over and under fitting with values $\alpha_{\text{exit}} \geq 0.9$, $\beta = 0.5$. Recall, $\beta = 0.5$ means that the within-cluster mean residual value and the cluster range are considered equally important components of a cluster weight. Our experience shows that choices of $\alpha_{\text{exit}} \in (0, 0.8)$ generally lead to underfitting, especially for less smooth functions. Thus, in order to obtain a GeDS estimate it is necessary to input only the set of data $\{x_i, y_i\}_{i=1}^N$ and use the default preassigned values of $\alpha_{\text{exit}} = 0.9$ and $\beta = 0.5$, which in general need not be re-set.

4.1 A simulated example.

In this section, we aim at illustrating the method and its properties on the following test example, which appears in Schwetlick and Schütze (1995). The proposed GeDS method has been implemented using *Mathematica* 9.0 and a standard PC (Intel core i7 CPU, 2.93 Ghz, 8GB RAM) has been used for all test examples. For a simulated data set, graphs of the linear

Table 1: Test function.

Test function	Interval	Sample size, N	x_i, $i = 1, \ldots, N$	Noise level, σ_ϵ	SNR
$f(x) = \frac{10x}{1+100x^2}$	[-2,2]	90	$x_i = -2 + \frac{2-(-2)}{89}(i-1)$	$U(-0.05, 0.05)$	7
		180		$U(-0.05, 0.05)$	7
		180		$U(-0.065, 0.065)$	5

spline fits, produced at each consecutive iteration in stage A of GeDS, preceding the final one, are given in Figure 2. As can be seen, the initial straight line fit, presented in Figure 2 (a), is sequentially improved by adding knots, one at each iteration, to reach the fit $\hat{f}\left(\boldsymbol{\delta}_{8,2}; x\right)$, plotted in Figure 3 (a), which can not be further significantly improved by adding more knots. Applying the averaging knot location (20) to the knots $\boldsymbol{\delta}_{8,2}$, the set of knots $\overline{\mathbf{t}}_{8-(n-2),n}$ of the quadratic, $n = 3$, and cubic, $n = 4$, fits, $\hat{f}\left(\overline{\mathbf{t}}_{8-(n-2),n}; x\right)$, are defined as follows: $\bar{t}_4 = (-1.1 - 0.33)/2 = -0.69$, $\bar{t}_5 = (-0.33 - 0.12)/2 = -0.22$, ... in $\overline{\mathbf{t}}_{7,3}$, or $\bar{t}_5 = (-1.1 - 0.33 - 0.12)/3 = -0.51$, $\bar{t}_6 = (-0.33 - 0.12 - 0.05)/3 = -0.17$, ... in $\overline{\mathbf{t}}_{6,4}$. The LS spline fits $\hat{f}\left(\overline{\mathbf{t}}_{8-(n-2),n}, \hat{\boldsymbol{\theta}}; x\right)$, resulting from stage B of GeDS, are plotted in Figure 3 (b) and (c) for $n = 3$ and $n = 4$, respectively. The polygons $\mathbf{C}_{f\left(\tilde{\mathbf{t}}_{7,3}, \hat{\boldsymbol{\alpha}}; x\right)}$ and $\mathbf{C}_{f(\overline{\mathbf{t}}_{7,3}, \hat{\boldsymbol{\alpha}}; x)}$, plotted in Figure 3 (d), using dot-dashed and dashed lines, and obtained with $\tilde{\mathbf{t}}_{7,3}$ as the solution of (16) and with $\overline{\mathbf{t}}_{7,3}$, calculated using (20), are seen to be very close to each other and also close to the initial control polygon $\hat{f}\left(\boldsymbol{\delta}_{8,2}, \hat{\boldsymbol{\alpha}}; x\right)$. The LS fits, $\hat{f}\left(\tilde{\mathbf{t}}_{7,3}; x\right)$ and $\hat{f}\left(\overline{\mathbf{t}}_{7,3}; x\right)$, have close L_2-errors, respectively 0.2798 and 0.2944, which confirms that $\overline{\mathbf{t}}_{7,3}$ approximates very well the optimal set of knots $\tilde{\mathbf{t}}_{7,3}$. The L_2-error is defined as the square root of the sum of squared residuals. The details of the final linear fit, and its corresponding quadratic and cubic spline fits are presented in Table 2. As can be seen, the function f is symmetric and GeDS places, symmetrically around the origin, 8, 7 and 6 knots, respectively for the linear, quadratic and cubic LS fits. All the fits are of a very good quality with respect to the mean square error (MSE), defined as MSE$=\left\{\sum_{i=1}^{N}\left(f(x_i) - \hat{f}(x_i)\right)^2\right\}/N$. Based on the L_2-errors for fits 1-3 given in Table 2, the best GeDS fit for this particular data set is the cubic one. We have compared it (No 3, Table 2) with the optimal cubic spline fits

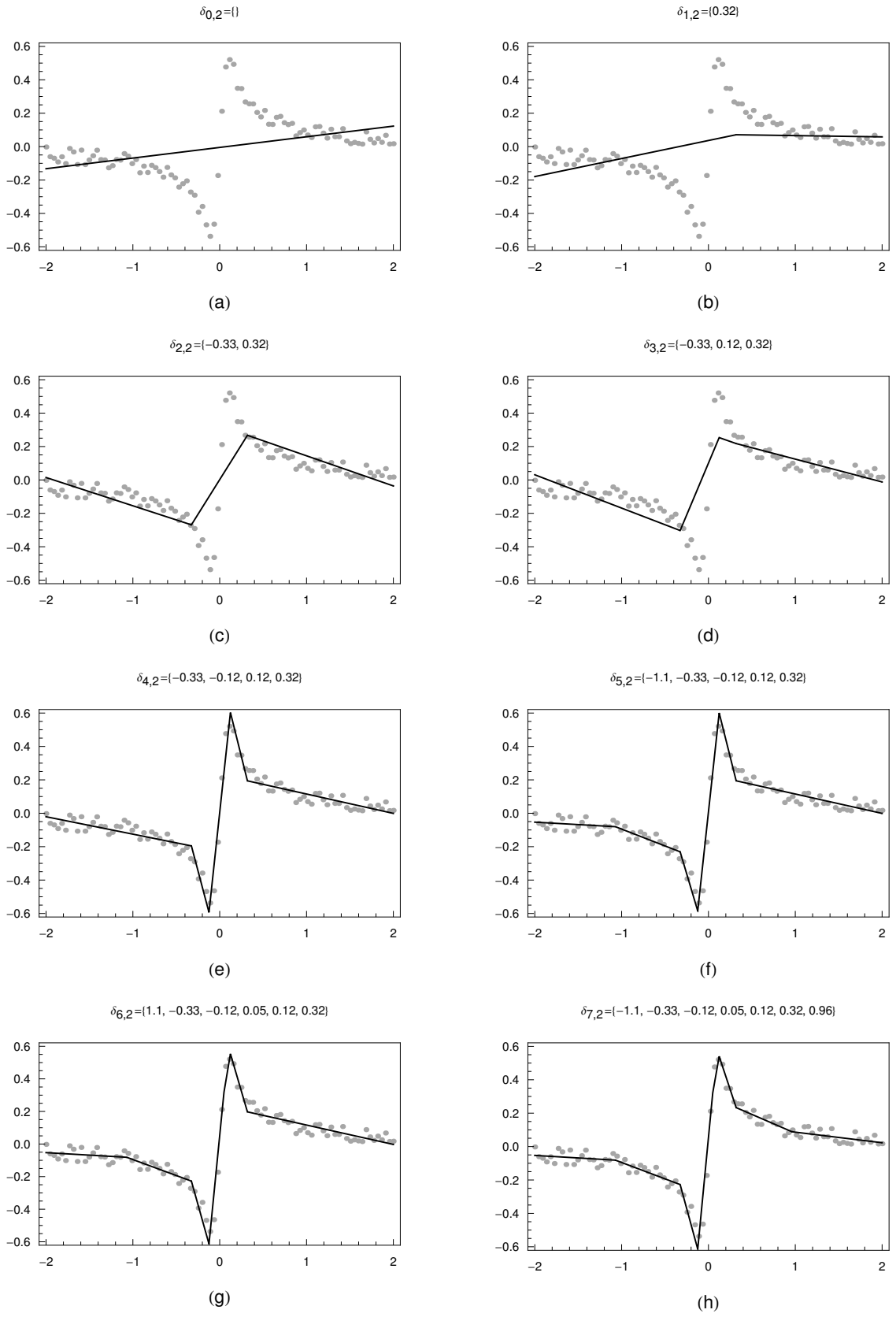


Figure 2: The linear spline fits, obtained at each consecutive iteration in stage A, except the final one (given in Figure 3 (a)). $N = 90$, SNR= 7.

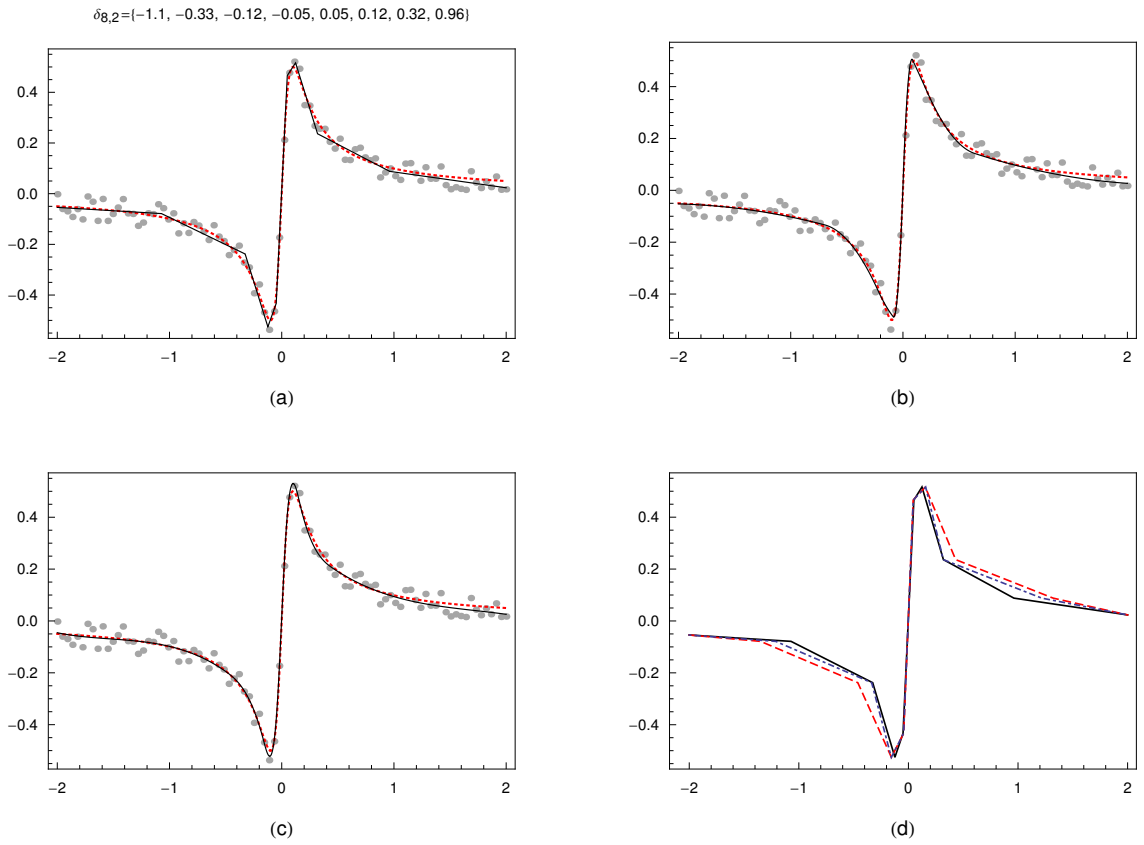


Figure 3: The final GeD spline fits: (a) linear, $\hat{f}\left(\boldsymbol{\delta}_{8,2}, \hat{\boldsymbol{\alpha}}; x\right)$; (b) quadratic, $\hat{f}\left(\bar{\mathbf{t}}_{7,3}, \hat{\boldsymbol{\theta}}; x\right)$; (c) cubic, $\hat{f}\left(\bar{\mathbf{t}}_{6,4}, \hat{\boldsymbol{\theta}}; x\right)$; (d) graphs of $\hat{f}\left(\boldsymbol{\delta}_{8,2}, \hat{\boldsymbol{\alpha}}; x\right)$ - the solid line, $\mathbf{C}_{f(\tilde{\mathbf{t}}_{7,3}, \hat{\boldsymbol{\alpha}}; x)}$ - the dot-dashed line and $\mathbf{C}_{f(\bar{\mathbf{t}}_{7,3}, \hat{\boldsymbol{\alpha}}; x)}$ - the dashed line; The dotted curve in (a), (b), (c) is the true function.

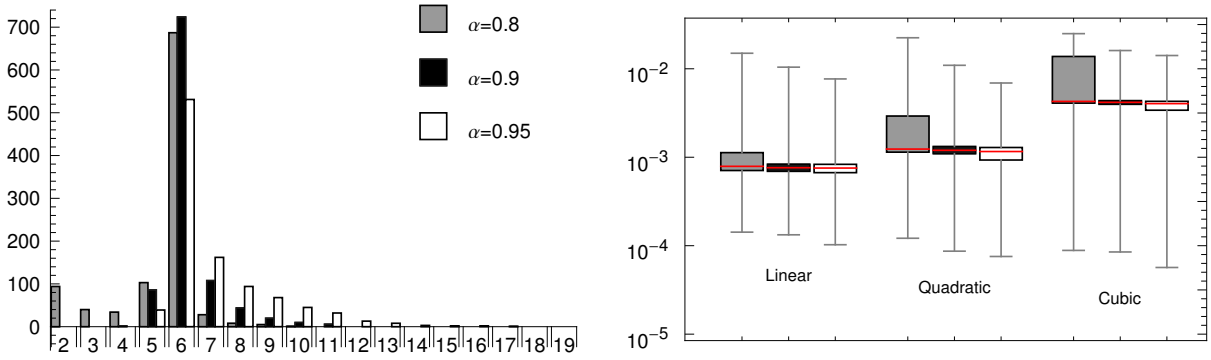
Table 2: Summary of GeD (No 1-3) and PNOM and NOM spline fits. $N = 90$, SNR= 7.

No	Fit	n	k	Internal knots	$\alpha_{\text{exit}}, \beta$	L_2-error, MSE
1	Fig. 4,(a)	2	8	{-1.1,-0.33,-0.12,-0.05,0.05,0.12,0.32,0.96}	0.9,0.5	0.2699,0.000189
2	Fig. 4,(b)	3	7	{-0.69,-0.22,-0.09,0.00,0.09,0.22,0.64}	0.9,0.5	0.2944,0.000127
3	Fig. 4,(c)	4	6	{-0.51,-0.17,-0.04,0.04,0.16,0.47}	0.9,0.5	0.2631,0.000119
4	PNOM	4	6	{-0.53,-0.16,-0.06,0.05,0.17,0.51}	-	0.2623,0.000131
5	NOM	4	6	{-0.48,-0.15,-0.07,0.05,0.18,0.40}	-	0.2614,0.000154

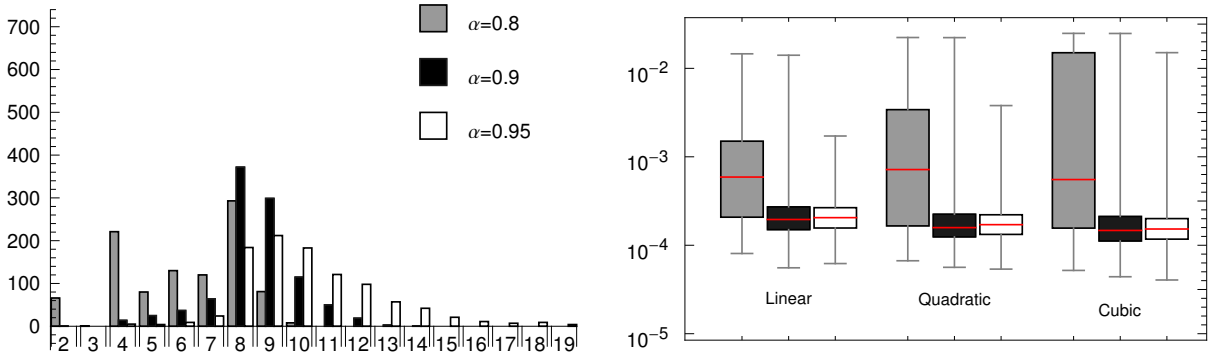
obtained applying the LS non-linear optimization method (NOM) and its penalized version (PNOM), due to Lindstrom (1999). As can be seen, the three fits are very close, comparing the L_2-errors, the MSE and the location of the knots. However, the GeD fit recovers best the original function as indicated by the corresponding MSE values. The computation time for the three GeDS fits is 0.87 seconds whereas for PNOM and NOM it is respectively, 4.5 hours and 1.4 hours, using the *Mathematica* function NMinimize.

On this example, we test GeDS by fitting 1000 simulated data sets. Frequency plots of the number of internal knots of the 1000 linear GeD spline fits and box plots of the MSE values of the linear, quadratic and cubic GeDS fits are presented in Figure 4, for choices of values of the parameter $\alpha_{\text{exit}} = 0.8, 0.9, 0.95$, and choices for parameter $\beta = 0.3, 0.5, 0.7$. The frequency plots given on the left panels of Figure 4 show that for this test example a choice of $\alpha_{\text{exit}} = 0.8$ leads to a relative underfitting, whereas setting $\alpha_{\text{exit}} = 0.95$ results in overfitting, regardless of the level of the parameter β. The choice $\alpha_{\text{exit}} = 0.9$ provides the best distribution of the number of knots chosen for the linear fit at the end of Stage A, in particular when $\beta = 0.5$ or $\beta = 0.7$. Recall that a higher value of β assigns more weight to the within-cluster mean relative to the within-cluster range for clusters of residuals of same sign. The box plots presented in the right panels of Figure 4 illustrate that, for the particular level of smoothness of the test function and the chosen SNR=7, the results for $\beta = 0.5$ and $\beta = 0.7$ are substantially better than those for $\beta = 0.3$, with the cubic and the quadratic GeDS fits being very close for both $\alpha_{\text{exit}} = 0.9$ and $\alpha_{\text{exit}} = 0.95$. For levels of $\beta = 0.7$ and $\alpha_{\text{exit}} = 0.9$ for instance, the 1000 linear, quadratic and cubic GeD spline fits, with median number of regression functions $n + k = 10$, have median L_2-errors 0.2604, 0.2679, 0.2637 (MSE values 0.000186, 0.000151, 0.000137) respectively, which are slightly lower than 0.277, obtained by Schwetlick and Schütze (1995) for a quartic (of order 5) fit with the same number of regression parameters and optimally located knots. The optimal quartic fit of Schwetlick and Schütze (1995) is obtained starting from 15 knots and after three time-consuming knot generation, removal and relocation stages.

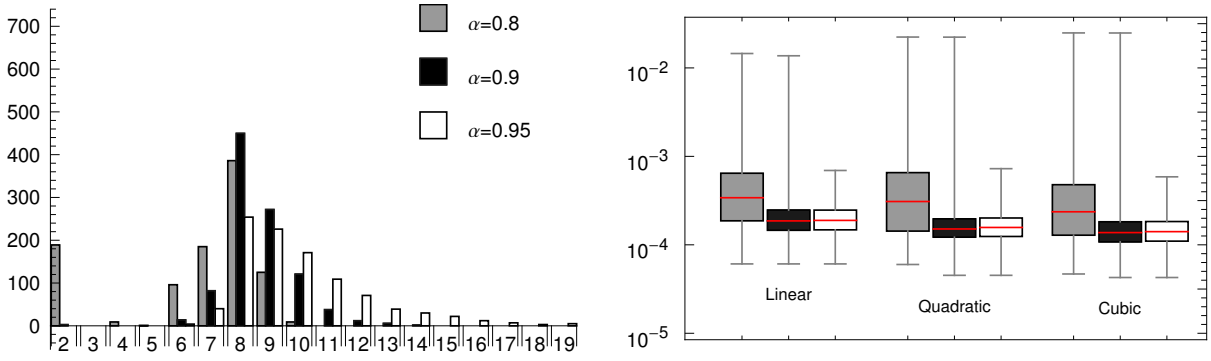
Continuing with the same test example, for $N = 90$ and $SNR = 7$, in Figure 5, the number of knots and the MSE values for the 1000 GeD spline fits obtained with $\beta = 0.7$ and $\alpha_{\text{exit}} = 0.9$, and illustrated in black on Figure 4 (c), are compared with the results obtained applying the GeDS methodology to the same 1000 data sets but with two alternative model selection criteria, namely GCV and SURE given by (15) and (14). As can be seen, the GCV, with a choice of



(a) β=0.3

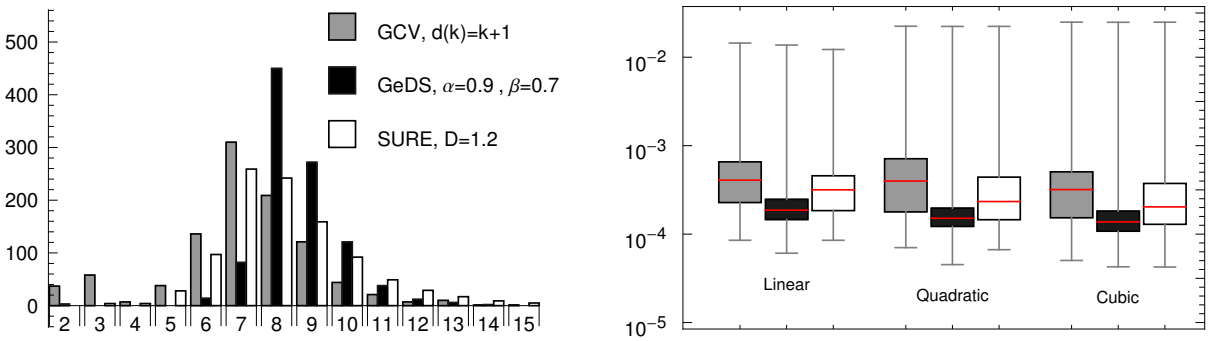


(b) β=0.5



(c) β=0.7

Figure 4: Frequency plots of the number of knots, l, (left panels) and box plots of the MSE values (right panels) of the 1000 GeD spline fits obtained with three different choices of values of the exit parameter α, and three choices of parameter β illustrated on (a), (b) and (c) respectively. $N = 90$, SNR= 7.



(a) N=90 , SNR=7

Figure 5: Frequency plots of the number of knots, l, (left panels) and box plots of the MSE values (right panels) of the 1000 linear GeD spline fits for three different choices of model selection criterion, GCV, GeDS and SURE

$d(k) = k + 1$, leads to slight underfitting, and the SURE, with choice of $D = 1.2$, has a more dispersed distribution of the number of knots, both resulting in somewhat worse MSE values given on the top right panel of Figure 5. We have tried a few different choices for the parameters $d(k)$ and D, e.g. $D = 3$ and $d(k) = 3k + 1$, and have chosen those giving the best results.

Using the same test example, the performance of GeDS has also been tested when the number of simulated data points increases from $N = 90$ to $N = 180$, and when the SNR worsens from SNR=7 to SNR=5 approximately. Results of this test suggest very robust performance of our estimator. For details we refer to the online supplement.

4.2 Real data example.

In this section, we use the GeDS method to fit a high pressure neutron barium-iron-arsenide ($BaFe_2As_2$) powder diffraction data from Kimber et al. (2009), with number of observations $N = 1151$. These authors explore the superconductivity properties of $BaFe_2As_2$ by conducting a neutron diffraction experiment and using the Rietveld method (see e.g. Will 2006) to fit the observed $BaFe_2As_2$ neutron diffraction profile. The latter profile includes the observed neutron diffraction intensity, viewed as a response variable, dependent on the angle of dispersion of the neutrons, known as the Bragg angle, denoted 2θ. As can be seen from Figure 6, this dependence is highly non-linear with numerous intensity peaks occurring at certain angle values. Important information about the structural properties of the $BaFe_2As_2$ compound are retrieved by analyzing the position of the peaks (the 2θ values at which they occur) and their shape

(heights and widths). Although the Rietveld method is extensively used to fit X-ray and neutron diffraction patterns in structural physics, chemistry and materials science, it has some recognized disadvantages. The non-linear least squares Rietveld procedure is complex, time consuming and not easy to implement, there are numerous unknown parameters (about 12 per peak), and there is a sophisticated goodness-of-fit refinement stage involving manual adjustment of the estimated parameters through a visual inspection of the residuals.

In Fig. 6 (a), the quadratic GeD spline fit obtained with $\beta = 0.6$ and $\alpha_{\text{exit}} = 0.99$ in 1581 secs, and with 227 knots and L_2-error of 10847, is given together with its corresponding residual plot. It can be compared with the Rietveld fit obtained by Kimber et al. (2009), as presented in Figure 6 (b), whose L_2-error is 26936. In Figure 6 (c), a quadratic least squares spline fit with 227 uniform knots is illustrated and the residual plot shows that it provides a significantly less accurate approximation to the data, confirmed by its L_2-error of 64790. More importantly, it can be seen that the heights and the widths of the peaks are poorly fitted. Clearly, the GeDS fit captures best the positions, the heights and the widths of the peaks and therefore enables a more accurate estimation of the structural parameters of the $BaFe_2As_2$ compound.

5 Discussion and conclusions.

In summary, the GeDS method is motivated by a novel geometric interpretation of spline regression parameter estimation as location of control points i.e., control polygons (see section 2 for details). The latter can serve as an important bridge between non-parametric spline estimation and computer aided geometric design, that can further enrich both areas with new ideas and insights. GeDS is based on the core novel idea of first fitting a polygon in stage A, capturing the shape of the data and then approximating it in stage B, by a variation-diminishing higher order spline function which is also an LS fit to the data (see section 3). For the purpose of constructing a linear spline fit in stage A, one can in principle use one of the alternative spatially adaptive methods, proposed in the literature, with an appropriately chosen model selector (e.g AIC, GCV or SURE).

Many free-knot spline smoothers have been proposed in the statistical literature and it is difficult to exhaustively compare GeDS with all of them. Instead we have used results from other comparative studies for spline smoothers, e.g. those due to Lee (2002a,b) and Luo and

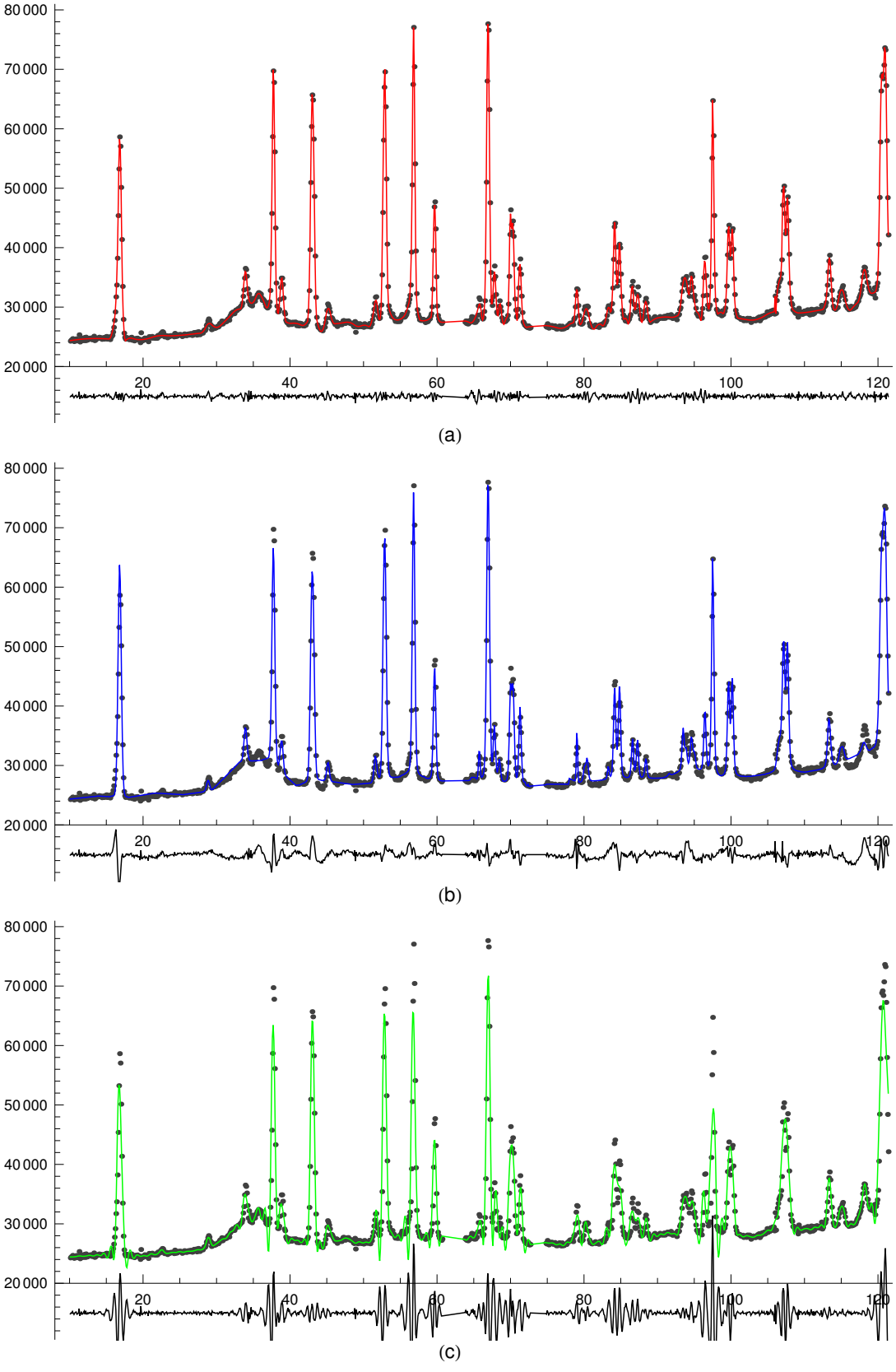


Figure 6: Estimated fits to $BaFe_2As_2$ data and their corresponding residuals: (a) quadratic GeD spline fit with 227 knots; (b) the Rietveld fit obtained by Kimber et al. (2009); (c) quadratic LS spline fit with 227 uniform knots.

Wahba (1997) to compare and contrast the performance of GeDS. In summary, GeDS has performed favorably in comparison with most methods implemented in these papers (see the online supplement for details of this comparison). Our method exhibits the following features which make it an attractive alternative to existing methods (cf. section 4 and the online supplement to the paper). GeDS:

- has an intuitive geometric interpretation which allows to follow the entire fitting process. Its shape preserving (VDS) knot positioning avoids "knot confounding" and "lethargy" problems which occur with other methods (see e.g., Zhou and Shen 2001 or Jupp 1978);
- does not rely on costly non-linear optimization as is the case for other free-knot methods and model selection criteria, which is problematic in high dimensions, i.e. for highly spatially inhomogeneous functions requiring many knots;
- produces simultaneously linear, quadratic, cubic, and possibly higher order spline fits and offers the flexibility to choose the degree of the final fit providing best compromise between smoothness and accuracy;
- can be flexibly tuned by the choice of the two parameters, α_{exit} and β, taking into consideration the prior information (if available or otherwise, estimated) about the smoothness of the underlying function and related SNR;
- is fast and produces very accurate fits for a reasonably wide range of SNR values and smooth or wiggly functions;
- has strong practical appeal especially in fast and accurate (real time) smoothing of spatially inhomogeneous functions with possible discontinuities, as demonstrated in section 4.2, based on the $BaFe_2As_2$ data.

In conclusion, the proposed univariate GeDS method provides a novel solution to the spline regression problem and in particular, to the problem of estimating the number and position of the knots. It is motivated by geometric arguments and can be extended to multivariate non-parametric smoothing as well as to generalized linear models. Details of how this may be done are outside the scope of this paper and are the subject of ongoing research.

6 Supplemental materials.

As mentioned and referenced in the body of this paper, supplemental materials are available online. It contains the following files.

Online supplement to: Geometrically designed, variable knot regression splines This is a supporting document with two sections. Section 1.1 includes results of a simulation study illustrating the theoretical bounds of section 3.1 given by Theorem 3.4 and its corollaries 3.5 and 3.6. This section also contains Remarks 3.2 and 3.3 which provide further insight into the optimality of the knot positioning in stages A and B. Section 1.2 contains further results of the sensitivity test of GeDS based on Example 1 from section 4.1 of the paper. Section 2 provides further detailed sensitivity tests and comparison of GeDS with other existing free-knot spline methods from the literature, given e.g. in Lee (2002a,b) and Luo and Wahba (1997), based on stylized examples of both smooth and highly spatially inhomogeneous functions introduced by Donoho and Johnstone (1994). (Kaishev et al.(2015)GeDS online supplement.pdf)

Data file File containing the $BaFe_2As_2$ data used in section 4.2. (data.xls)

Acknowledgements

The authors would like to acknowledge support received through a research grant from the UK Institute of Actuaries. The authors would also like to thank Simon Kimber for providing them with the $BaFe_2As_2$ dataset and the results from the Rietveld fit given in Kimber et al. (2009). The sincere encouragement received by David van Dyk, and his help in discussing and providing invaluable advice on ways to improve the paper are greatly appreciated.

Appendices

A Proofs of the results of section 3.1

Proof of Theorem 3.4. Note that, for $n = 2$, $\xi_i \equiv \xi_i^*$, $i = 1, \ldots, p$, hence $V^a[g] \equiv V[g]$ and the bound in (23), which is zero, is sharp. For $n > 2$, from (4) it follows that $\xi_1^* \equiv a \equiv \xi_1$ and

$\xi_p^* \equiv b \equiv \xi_p$, and from the definitions of $V[g]$ and $V^a[g]$, (9) and (22) respectively, we have

$$\begin{aligned}
\|V[g] - V^a[g]\|_\infty &= \max_{t\in[a,b]} \left| \sum_{i=1}^{p} (g(\xi_i^*) - g(\xi_i)) N_{i,n}(t) \right| \\
&\leq \max_{t\in[a,b]} \sum_{i=1}^{p} \left| (g(\xi_i^*) - g(\xi_i)) \right| N_{i,n}(t) \\
&\leq \max_{t\in[a,b]} \sum_{i=1}^{p} \left\{ \max_{j\in\{2,\ldots,p-1\}} \left| \left(g\left(\xi_j^*\right) - g(\xi_j)\right) \right| \right\} N_{i,n}(t) \\
&\leq \max_{j\in\{2,\ldots,p-1\}} \left| \left(g\left(\xi_j^*\right) - g(\xi_j)\right) \right| \max_{t\in[a,b]} \sum_{i=1}^{p} N_{i,n}(t) \\
&= \max_{j\in\{2,\ldots,p-1\}} \left| \left(g\left(\xi_j^*\right) - g(\xi_j)\right) \right|, \qquad (27)
\end{aligned}$$

where the last equality follows from the partition of unity property of B-splines (see section 2). Applying the definition of the modulus of continuity to (27) we have

$$\|V[g] - V^a[g]\|_\infty \leq \max_{j\in\{2,\ldots,p-1\}} \left| \left(g\left(\xi_j^*\right) - g(\xi_j)\right) \right| \leq \omega\left(g; \max_{j\in\{2,\ldots,p-1\}} \left|\xi_j^* - \xi_j\right|\right) \quad (28)$$

From the definition (4) of the Greville sites ξ_i^* we have $\xi_j^* = (t_{j+1} + \ldots + t_{j+n-1})/(n-1)$, $j = 2, \ldots, p-1$. From (21), it follows that $t_{j+1} = \left(\xi_{j-(n-2)} + \ldots + \xi_j\right)/(n-1), \ldots, t_{j+n-1} = \left(\xi_j + \ldots + \xi_{j+(n-2)}\right)/(n-1)$, where we have defined $\xi_{1-l} := a$ and $\xi_{p+l} := b$, $l = 1, 2, \ldots$. Consider the $\max_{j\in\{2,\ldots,p-1\}} \left|\xi_j - \xi_j^*\right|$ and assume it is achieved for some j^m, $2 \leq j^m < p-1$. Expressing $\xi_{j^m}^*$ in terms of ξ_{j^m}, using the above equalities, after some algebra, it is not difficult to see that

$$\left|\xi_{j^m} - \xi_{j^m}^*\right| = \frac{1}{(n-1)^2} \left| \sum_{i=1}^{n-2} i \left(\xi_{j^m+(n-1-i)} + \xi_{j^m-(n-1-i)}\right) - (n-1)(n-2)\xi_{j^m} \right| \qquad (29)$$

and if we now rearrange the terms in the sum in (29), we obtain

$$\left|\xi_{j^m} - \xi_{j^m}^*\right| = \frac{1}{(n-1)^2} \left| \sum_{i=1}^{n-2} i \left(\left(\xi_{j^m+(n-1-i)} - \xi_{j^m}\right) - \left(\xi_{j^m} - \xi_{j^m-(n-1-i)}\right) \right) \right|. \qquad (30)$$

Assume that $\sum_{i=1}^{n-2} i\left(\xi_{j^m+(n-1-i)} - \xi_{j^m}\right) > \sum_{i=1}^{n-2} i\left(\xi_{j^m} - \xi_{j^m-(n-1-i)}\right)$. In this case, it is not difficult to see that (30) is bounded by

$$\begin{aligned}
\left|\xi_{j^m} - \xi^*_{j^m}\right| &\leq \frac{1}{(n-1)^2}\sum_{i=1}^{n-2} i\left(\xi_{j^m+(n-1-i)} - \xi_{j^m}\right) \\
&\leq \frac{1}{(n-1)^2}\frac{(n-2)(n-1)}{2}\left(\xi_{j^m+(n-2)} - \xi_{j^m}\right) \\
&\leq \frac{(n-2)}{2(n-1)}\left(\xi_{j^m+(n-2)} - \xi_{j^m}\right) \\
&\leq \frac{(n-2)^2}{2(n-1)}\max_{j\in\{1,\ldots,p-1\}}(\xi_{j+1} - \xi_j). \qquad (31)
\end{aligned}$$

Similarly, it can be shown that if $\sum_{i=1}^{n-2} i\left(\xi_{j^m+(n-1-i)} - \xi_{j^m}\right) \leq \sum_{i=1}^{n-2} i\left(\xi_{j^m} - \xi_{j^m-(n-1-i)}\right)$ the bound in (31) also holds. Thus, from (31) and (28) we have

$$\|V[g] - V^a[g]\|_\infty \leq \omega\left(g; \frac{(n-2)^2}{2(n-1)}\max_{j\in\{1,\ldots,p-1\}}(\xi_{j+1} - \xi_j)\right). \qquad (32)$$

Using the monotonicity and subadditivity of $\omega(g;h)$ in h, from (32) we finally obtain

$$\|V[g] - V^a[g]\|_\infty \leq \left\lceil \frac{(n-2)^2}{2(n-1)} \right\rceil \omega\left(g; \max_{j\in\{1,\ldots,p-1\}}(\xi_{j+1} - \xi_j)\right)$$

where $\lceil\nu\rceil := \min\{z \in \mathbb{Z} : \nu \leq z\}$. This completes the proof of Theorem 3.4. □

Proof of Corollary 3.5. This follows directly from (32) and from the definition, (24) of $\omega(g;h)$, i.e.

$$\|V[t] - V^a[t]\|_\infty = \left\| t - \sum_{i=1}^{p} \delta_{i+1} N_{i,n}(t) \right\|_\infty \leq \frac{(n-2)^2}{2(n-1)}\max_{j\in\{1,\ldots,p-1\}}(\delta_{j+2} - \delta_{j+1}).$$

□

Proof of Corollary 3.6. From (27), for $n = 3$ and $g = \hat{f}$, we have

$$\begin{aligned}
\|V[\hat{f}] - V^a[\hat{f}]\|_\infty &\leq \max_{j\in\{2,\ldots,p-1\}} \left| \hat{f}\left(\boldsymbol{\delta}_{l,2}, \hat{\boldsymbol{\alpha}}; \xi_j^*\right) - \hat{f}\left(\boldsymbol{\delta}_{l,2}, \hat{\boldsymbol{\alpha}}; \delta_{j+1}\right) \right| \qquad (33)\\
&= \max_{j\in\{2,\ldots,p-1\}} \left| \sum_{i=1}^{p} \hat{\alpha}_i N_{i,2}\left(\xi_j^*\right) - \sum_{i=1}^{p} \hat{\alpha}_i N_{i,2}(\delta_{j+1}) \right|\\
&= \max_{j\in\{2,\ldots,p-1\}} \left| \sum_{i=1}^{p} \hat{\alpha}_i N_{i,2}\left(\xi_j^*\right) - \hat{\alpha}_j \right|\\
&= \max_{j\in\{2,\ldots,p-1\}} \left| \sum_{i=j-1}^{j+1} \hat{\alpha}_i N_{i,2}\left(\xi_j^*\right) - \hat{\alpha}_j \right| \qquad (34)
\end{aligned}$$

Recall that $n = 3$ and hence, $\xi_j^* = (t_{j+1} + t_{j+2})/2$, and $t_{j+1} = (\delta_j + \delta_{j+1})/2$, $t_{j+2} = (\delta_{j+1} + \delta_{j+2})/2$. Therefore, we need to consider the cases when $\delta_j < \xi_j^* \leq \delta_{j+1}$, or $\delta_{j+1} \leq \xi_j^* < \delta_{j+2}$, $2 \leq j \leq p-1$. In the first case, applying the Mansfield-De Boor-Cox recurrence formula we know that if $\delta_j < \xi_j^* < \delta_{j+1}$, then $\sum_{i=j-1}^{j+1} \hat{\alpha}_i N_{i,2}\left(\xi_j^*\right) = \hat{\alpha}_{j-1} N_{j-1,2}\left(\xi_j^*\right) + \hat{\alpha}_j N_{j,2}\left(\xi_j^*\right)$, which is a convex combination of only two B-spline coefficients. Thus, (34) becomes

$$\begin{aligned}
& \max_{j\in\{2,\ldots,p-1\}} \left| \hat{\alpha}_{j-1} N_{j-1,2}\left(\xi_j^*\right) + \hat{\alpha}_j N_{j,2}\left(\xi_j^*\right) - \hat{\alpha}_j \right|\\
= \; & \max_{j\in\{2,\ldots,p-1\}} \left| \hat{\alpha}_{j-1} \frac{\delta_{j+1} - \xi_j^*}{\delta_{j+1} - \delta_j} + \hat{\alpha}_j \frac{\xi_j^* - \delta_j}{\delta_{j+1} - \delta_j} - \hat{\alpha}_j \frac{\delta_{j+1} - \delta_j}{\delta_{j+1} - \delta_j} \right|\\
= \; & \max_{j\in\{2,\ldots,p-1\}} \left| (\hat{\alpha}_{j-1} - \hat{\alpha}_j) \right| \left(\frac{\delta_{j+1} - \xi_j^*}{\delta_{j+1} - \delta_j} \right)\\
< \; & \max_{j\in\{2,\ldots,p-1\}} \left| (\hat{\alpha}_{j-1} - \hat{\alpha}_j) \right| \left(\frac{\frac{1}{4}(\delta_{j+1} - \delta_j)}{\delta_{j+1} - \delta_j} \right)\\
= \; & \frac{1}{4} \max_{j\in\{2,\ldots,p-1\}} \left| (\hat{\alpha}_{j-1} - \hat{\alpha}_j) \right|, \qquad (35)
\end{aligned}$$

where we have used the fact that $\delta_{j+2} - \delta_{j+1} > 0$ to arrive at the last inequality. Similarly, it is not difficult to see that the same bound as in (35) holds in the case when $\delta_{j+1} \leq \xi_j^* \leq \delta_{j+2}$. This completes the proof of Corollary 3.6. □

References

Antoniadis, A., Gijbels, I. and Verhasselt, A. (2012). Variable Selection in Additive Models Using P-Splines, *Technometrics*, **54** (4), 425–438.

Belitser, E. and Serra, P. (2014). Adaptive priors based on splines with random knots. *Bayesian Analysis*, **9** (4), 859–882.

Beliakov, G. (2004). Least squares splines with free knots:global optimization approach. *Applied Mathematics and Computation*, **149**, 783–798.

Biller, C. (2000). Adaptive Bayesian regression splines in semiparametric generalized linear models. *J. Comput. and Graph. Stat.*, **9**, 122–140.

Cohen, E., Riesenfeld, R. F. and Elber, G. (2001). *Geometric Modelling with Splines: An Introduction*. Natick, Massachusetts: A K Peters.

De Boor, C. (2001). *A practical Guide to Splines*, Revised Edition, New York: Springer.

Denison, D., Mallick, B., and Smith, A. (1998). Automatic Bayesian curve fitting, *J. R. Statist. Soc.*, B, **60**, 333–350.

Donoho, D. and Johnstone, I. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, **81**, 425–455.

Eubank, R. (1988). *Spline smoothing and Nonparametric Regression*. Dekker, New York.

Fan, J. and Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting: Variable bandwidth and spatial adaptation. *J. R. Statist. Soc.*, B, **57**, 371–394.

Farin, G. (2002). *Curves and Surfaces for CAGD*, Fifth Edition, San Francisco: Morgan Kaufmann.

Friedman, J. H. (1991). Multivariate adaptive regression splines (with discussion). *Ann. Statist.*, **19**, 1–141.

Friedman, J. H. and Silverman, B. W. (1989). Flexible Parsimonious smoothing and additive modeling (with discussion). *Technometrics*, **31**, 3–39.

Hastie, T. (1989). Discussion of "Flexible Parsimonious smoothing and additive modeling (with discussion)" by Friedman, J. H. and Silverman, B. W. *Technometrics*, **31**, 23–29.

Hansen, M. H. and Kooperberg, C. (2002). Spline Adaptation in Extended Linear Models (with comments and a rejoinder by the authors). *Statistical Science*, **17** (1), 2–51.

Huang, J. Z. (2003). Local assymptotics for polynomial spline regression. *Ann. Statist.*, **31**, 1600–1635.

Jupp, D. (1978). Approximation to data by splines with free knots. *SIAM J. Num. Analysis.*, **15**, 328–343.

Kaishev, V. K. (1984). A computer program package for solving spline regression problems, In: *Proceedings in Computational Statistics, COMPSTAT* (eds T. Havranek, Z. Sidak and M. Novak), pp. 409–414, Wien: Physica-Verlag.

Kang, H., Chen, F., Li, Y., Deng. J., and Yang, Z. (2015). Knot calculation for spline fitting via sparse optimization. *Computer-Aided Design*, **58**, 179–188.

Kimber, S. A. J., Kreyssig, A., Zhang, Y. Z., Jeschke, H. O., Valenti, R., Yokaichiya, F., Colombier, E., Yan, J., Hansen, T. C., Chatterji, T., McQueeney, R. J., Canfield, P. C., Goldman, A. I. and Argyriou, D. N. (2009). Similarities between structural distortions under pressure and chemical doping in superconducting $BaFe_2As_2$. *Nature Materials*, **8**, 471–475.

Lee, T. C. M. (2000). Regression spline smoothing using the minimum description length principle. *Stat. & Prob. Letters*, **48**, 71–82.

Lee, T. C. M. (2002a). Automatic smoothing for discontinuous regression functions. *Stat. Sinica*, **12**, 823–842.

Lee, T. C. M. (2002b). On algorithms for ordinary least squares regression spline fitting: A comparative study. *J. of Stat. Comp. and Simulation*, **72**, 647–663.

Lindstrom, M. J. (1999). Penalized estimation of free-knot splines. *J. Comput. and Graph. Stat.*, **8**, 2, 333–352.

Luo, Z., and Wahba, G. (1997). Hybrid adaptive splines. *J. Am. Statist. Ass.*, **92**, 107–115.

Mammen, E. and Van der Geer, S. (1997). Locally adaptive regression splines. *Ann. Statist.*, **25**, 1, 387–413.

Marx, B. D. and Eilers, P. H. C. (1996). Flexible Smoothing with B-splines and Penalties. *Stat. Science*, **11**, 2, 89–121.

Micchelli, C. A., Rivlin, T.J. and Winograd, S. (1976). The optimal recovery of smooth functions. *Numer. Math.*, **26**, 191–200.

Miyata, S. and Shen, X. (2003). Adaptive free-knot splines. *J. Comput. and Graph. Stat.* 12(1), 197-231.

Molinari, N., Durand, J-F., Sabatier, R. (2004). Bounded optimal knots for regression splines. *Comput Statist Data Anal*, 45(2), 159-178.

Pittman, J. (2002). Adaptive Splines and Genetic Algorithms. *J. Comput. and Graph. Stat.*, **11**, 3, 1–24.

Rupert, D. (2002). Selecting the number of knots for penalized splines. *J. Comput. and Graph. Stat.*, **11**, 4, 735–757.

Rupert, D. and Carroll, R. J. (2000). Spatially-Adaptive penalties for spline fitting. *Australian and New Zealand Journal of Statistics*, **42**, 205–223.

Rupert, D. M.P.Wand and Carroll, R. J. (2003). *Semiparametric Regression.*, Cambridge Series in Statistics and Probabilistic Mathematics, Cambridge University Press, New York USA.

Schwetlick, H. and Schütze, T. (1995). Least squares approximation by splines with free knots. *BIT. Numerical Math.*, **35**, 854–866.

Smith, P. L. (1982). Curve fitting and modeling with splines using statistical variable selection techniques. *Report NASA 166034*, Langley Research Center, Hampton, VA.

Smith, M. and Kohn, R. (1996). Nonparametric regression using Bayesian variable selection. *J. Econometrics*, **75**, 317–344.

Stone, C. J., Hansen, M. H., Kooperberg, C. and Truong, Y. K. (1997). Polynomial Splines and their tensor products in extended linear modeling. *Ann. Statist.*, **25**, 1371–1470.

Zhou, S. and Shen, X. (2001). Spatially adaptive regression splines and accurate knot selection schemes. *J. Am. Statist. Ass.*, **96**, 247–259.

Van Loock, W. Pipeleers, G. De Schutter, J. and Swevers, J. (2011). A convex optimization approach to curve fitting with B-splines,in: *Preprints of the 18th international federation of automatic control (IFAC), Milano (Italy)*, 2290–2295.

Wahba, G. (1990). *Spline Models for Observational Data*. SIAM, Philadelphia.

Will, G. (2006). *Powder Diffraction: The Rietveld Method and the Two Stage Method*. Springer Berlin Heidelberg.

Wood, S. N. (2003). Thin plate regression splines. *J. R. Statist. Soc.*, B, **65**, Part 1, 95–114.

Yuan, Y., Chen, N., and Zhou, S. (2013). Adaptive B-spline knots selection using multi-resolution basis set. *IIE Trans*, 45(12), 1263-1277.