

Piecewise-cubic Smoothing with Local Estimations and Automatic Knot Detection

N.D. Dikoussar

Laboratory of Information Technologies, JINR

Cs. Török

Technical University, Košice, Slovakia

Abstract

A smoother based on an adaptive cubic model [1, 2] and splines with free knots is proposed. The model uses three reference data points and two parameters of control for estimation a near optimal position of knots at the axis x in autotracking mode. The data points are prethinned and corrected by local linear fitting. A coefficient table is obtained by a standard spline procedure. The efficiency and the stability of the smoother with respect to random errors are shown on real noisy data.

Introduction. The aim of this work is *to enhance the robustness* of the algorithm for detection of local cubic segments (LOCUSD) [1, 2] intended for processing noisy data with complex dependency by piecewise cubic polynomials. This report is based mainly on papers [3, 4]. There are several ways to solve this problem leveraging the autotracking piecewise cubic approximation (APCA). We mention two of them.

The first one is that we can employ it to smoothed data. For smoothing (but not functionally describing) data, there are various methods, such as kernel smoothers [5] or Friedman's variable span smoother (*supersmoother*) [6].

As we provide for data description piecewise functions, it is not necessary to smooth every data point. It is sufficient to give local estimations for several data (trying not to lose any measurement) and to employ the autotracking piecewise approximation to the estimated data. We follow this second approach.

In [2] we studied approximation of data with complex dependence and no or moderate error, using a cubic model with a free parameter, in two stages: local and global approximation.

The model plays a three-fold role: *firstly* it is used on the local level for expressing the relation between x and y , *secondly* for the construction of an iterative scheme for the estimation of the model's parameter, and *lastly* it enables a global continuous and smooth approximation in an automatic mode by piecewise cubic polynomials. While in the case of data with no or moderate errors the proposed autotracking piecewise approximation gives satisfactory results, in the case of errors with any variance (noisy data), there are problems with both the quality of the local approximants and the global smoothness (but not continuity). APCA succeeds in smoothing noisy data due to its combination with neural networks [12]. This paper proposes a solution without using NN.

We develop a smoothing procedure that produces a cubic spline $S(x; \mathbf{k}; \mathbf{c}_j)$, $j = \overline{1, k}$, with $k \geq 1$ internal knots from a set of data points

$$\{(x_i, \tilde{y}_i)\}_{i=1}^n, n \gg 4, \quad (1)$$

where $\tilde{y}_i = y_i + \epsilon_i$, $\epsilon_i \in N(0, \sigma^2)$; $\mathbf{k} = [x_1^*, x_2^*, \dots, x_k^*]$, $x_j^* \in \{x_i\}_{i=1}^n$ is a set of knots detected automatically by the smoother, and $\mathbf{c}_j = [c_{0j}, c_{1j}, c_{2j}, c_{3j}]$ is a vector of coefficients of the

model's polynomial at interval $[x_{j-1}^*, x_j^*]$, $j = \overline{1, k}$. The smooth function $S(x; \dots)$ shows the association between x_i and \tilde{y}_i as follows

$$\tilde{y}_i = S(x_i; \dots) + r_i, \quad i = \overline{1, n}, \quad (2)$$

where $S(x_i; \dots) = \hat{y}_i$ is the estimation of \tilde{y}_i , and the r_i are residuals.

Smoothers have been using in many applications and are described in a number of references [5, 6, 7]. Recently we have proposed a new "4-point approximation" based on the four-point transformation methodology [8, 9, 10]. The method and the algorithm (LOCUSD) for approximation and smoothing data with no or moderate error have been described in [1, 2, 11].

Section 1 is a short introduction to the cubic model for piecewise approximation based on four points and provides the necessary formulas. The next section describes the way we correct and reduce the number of the reference points. Section three shows the results of smoothing real data.

A Cubic Model for Autotracking Piecewise Approximation. Consider an additive model

$$\tilde{f} = f(x) + e. \quad (3)$$

We present the standard cubic polynomial $C = a_0 + a_1x + a_2x^2 + a_3x^3$ in the parametric form $S(\tau; \alpha, \beta, \mathbf{r}, \theta)$ with three fixed (\mathbf{r}), one free (θ) and two control (α and β) parameters by Eq. (4). The curve $S(\tau; \alpha, \beta, \mathbf{r}, \theta)$ passes through four points $\{(x_*, f_*)\}$, $* = \tau, \alpha, \beta, 0$, where $\tau = x - x_0$, $\alpha = x_\alpha - x_0$, $\beta = x_\beta - x_0$, $f_* \equiv f(x_*)$. The vector $\mathbf{r} = [f_\alpha, f_\beta, f_0]^T$ is set up of the reference ordinates that are related to data points \tilde{f} . The abscissas $x_\tau, x_\alpha, x_\beta, x_0$ are used for evaluation of the vector of weight functions $\mathbf{w} = [w_1, w_2, w_3]^T$ defined by Eq. (5) and Q . θ is an unknown free parameter:

$$S(\tau; \alpha, \beta, \mathbf{r}, \theta) = \overbrace{f_\alpha w_1 + f_\beta w_2 + f_0 w_3}^{\mathbf{w}^T \mathbf{r}} + \overbrace{\tau(\tau - \alpha)(\tau - \beta)}^Q = \Pi(\tau; \alpha, \beta, \mathbf{r}) + \theta Q(\tau; \alpha, \beta), \quad (4)$$

where

$$w_1 = \frac{-\tau(\tau - \beta)}{\alpha\gamma}, w_2 = \frac{\tau(\tau - \alpha)}{\beta\gamma}, w_3 = \frac{(\tau - \alpha)(\tau - \beta)}{\alpha\beta}, \gamma = \beta - \alpha; \alpha\beta\gamma \neq 0; \sum_{i=1}^3 w_i = 1. \quad (5)$$

The quadratic parabola $\Pi(\tau; \alpha, \beta, \mathbf{r}) = \mathbf{w}^T \mathbf{r}$ passes via three reference points and the cubic parabola $Q = \tau(\tau - \alpha)(\tau - \beta)$ is a "zeroing" parabola. Fig. 1 explains how the cubic model S can be used in the dynamic mode for detection knots and approximation of function f via Π_n and $\theta_{n-1}Q_n$. As the shape of curve Eq. (4) depends on the selection of the reference points \mathbf{r} , we can use the parameters α, β for controlling the error $e(x) = \tilde{f}(x) - S(x)$. For example, using the model (4) in *dynamic mode* (Fig. 1) we fix the points $(x_\alpha, \tilde{f}_\alpha)$ and $(x_\beta, \tilde{f}_\beta)$, and move the other two points (x_τ, \tilde{f}_τ) , (x_0, \tilde{f}_0) with respect to the unmoved curve \tilde{f} . Minimization of $e^2(x)$ by the parameter θ leads to an iterative estimation of θ :

$$\hat{\theta}_n = \hat{\theta}_{n-1} + K_n \overbrace{(\tilde{f}_n - \tilde{\Pi}_n - \hat{\theta}_{n-1}Q_n)}^{\varepsilon_n}, \quad \hat{\theta}_0 = 0, \quad n = 1, 2, \dots, \quad (6)$$

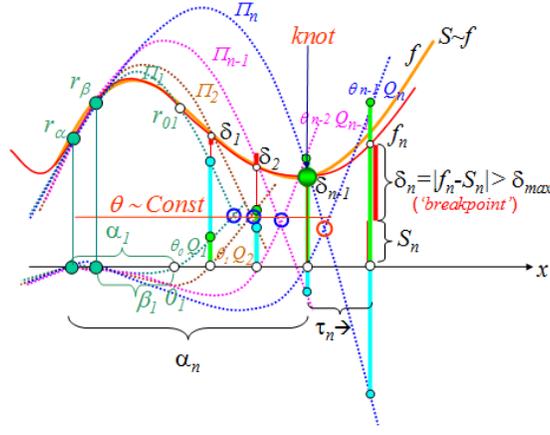


Fig. 1: A sketch of knot-detection process by using cubic model in the dynamic mode

where $K_n = Q_n / \sum_{k=1}^n Q_k^2$ is an amplification factor and $\tilde{\Pi}_n = \Pi(\tau_n; \alpha_n, \beta_n, \tilde{\mathbf{r}}_n)$, $\tau_n = x_n - x_{0n}$, $\tilde{\mathbf{r}} = [f_{\alpha n}, f_{\beta n}, f_{0n}]^T$.

Eq.(6) is a known adaptive procedure in which the output error is applied to input with the amplification factor $K_n(\alpha_n, \beta_n)$ that decreases as $\sim n^{-3}$, i.e. the errors e_n , $e_{\alpha n}$, $e_{\beta n}$ and e_{0n} from Eq. (3) are suppressed near to a cubic-low because of $|w_{in}| \rightarrow 1$ for the above described selection of α_n and β_n [2].

For automatic tracking of a cubic segment of a curve the criterion of constancy of the third derivative of the cubic model is used [2].

Correction of the Reference Ordinates. The critical part of the piecewise approximation by Eq. (6) in the case of noisy data is $\tilde{\Pi}_n = \Pi(\tau_n; \alpha_n, \beta_n, \tilde{\mathbf{r}}_n)$. It is clear that the reference ordinates $\tilde{\mathbf{r}}_n$ of the local approximants must be adjusted in some way. Here we propose a process that does not need the correction of every data point.

Consider M data points $(x_i, \hat{f}_i)_{i=\overline{1, M}}$. We describe the algorithm of the piecewise functional smoothing in four steps with remarks.

1. *Thinning of data points* by selection of $N \ll M$ points. As we will see the process can be applied to data with both equidistant and non-equidistant step.

2. *Local estimation* of the ordinates \hat{f} of the selected N points. There are many ways how to get good local point estimations. They have to be effective and take into account every M data points.

3. *Reduction* of the estimated N selected points using Eqs. (4)-(6) to K points. The detection of K knots by the first stage of the autotracking piecewise cubic approximation is executed on the N estimated points, so the reference ordinates in Eq. (6) have been corrected.

4. *Construction of integral approximants* based on the reduced K number of estimated points. To get continuous integral estimation the methods and formulas from the second stage of the autotracking piecewise cubic approximation can be leveraged, see [1, 2, 11]. To get not only continuous approximants but also approximants with continuous first and second derivatives, the spline table can be computed based on the reduced K estimated points from the third step.

Examples. In the previous section we described shortly in four steps the smoothing process based on local estimation (step 2) and the autotracking piecewise cubic approximation (step 3). To demonstrate the process we considered real noisy data with both

equidistant and non equidistant step. From the three data sets the first one shows the most complex relation. The figures contain the original data of length M denoted by little squares, the continuous spline smoothers, the residuals at the bottom of the pictures, the histograms constructed from the residuals, and the verticals denote the endpoints of the K segments. We also provide the number N of selected and estimated data points.

Fig. 2 illustrates smoothing data with equidistant step, the cross sections for π^-p collision [13]. Although the left and right splines were evaluated based roughly on every second and fifth locally corrected data, thanks to the autotracking knot detection from the third step their number was reduced approximately five and three times, to 24 and 20, respectively.

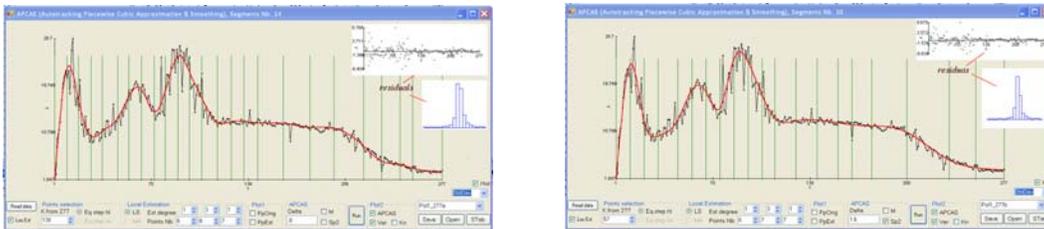


Fig. 2: Cross section for π^-p collision;
left: $M = 277$, $N = 130$, $K = 24$, right: $M = 277$, $N = 57$, $K = 20$

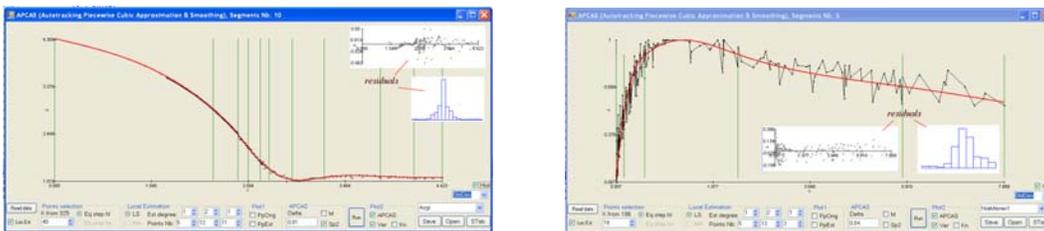


Fig. 3: Data with non equidistant step: cross section for np collision [14] (left) and concrete characteristic (right);
left: $M = 325$, $N = 40$, $K = 10$, right: $M = 196$, $N = 18$, $K = 5$

In fig. 3 we give the smoothing results of two data sets with non equidistant step: the left plot illustrates the cross sections for np collision and the right the resistance ratio.

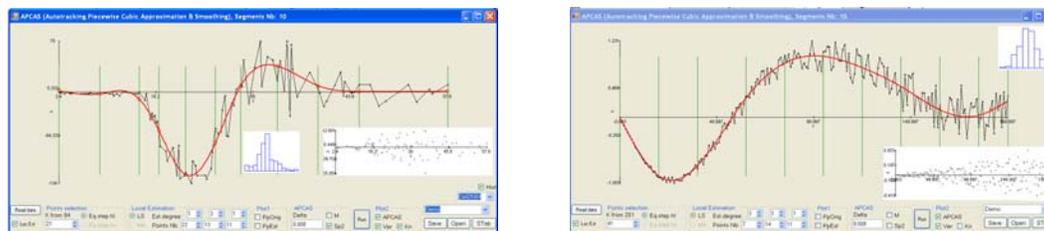


Fig. 4: APCA-smoothing of noisy data: a) “motorcycle” data [5]; b) the test function [6]

Fig. 4 illustrates new APCAS - smoothing results of the so-called motorcycle data from the book [5] and simulated data $\tilde{y}_i = \sin(2\pi(1 - x_i)^2 + x_i\epsilon_i)$, $x \in [0, 1]$; $i = \overline{1, n}$, $n = 200$; $\epsilon_i \sim N(0, 1)$ [6]. The distributions of residuals and histograms are also shown. The Fig. 4(b) shows the stability of the method to “a situation in which the curvature of f decreases and the variance of the random component increases with increases x ” [6]. Vertical lines show the knot-positions detected by APCAS.

The reduced autotracked number of segments and the quality of the piecewise approximants are adequate and acceptable in all cases.

Conclusions. The paper describes a smoothing process with local estimations and automatic knot detection for describing noisy data with complex dependence by piecewise continuous cubic polynomials. The resulting spline tables are slim and the splines provide for both simulated and real noisy data satisfactory approximation.

References

- [1] N.D. Dikoussar. Kusochno-kubicheskoe priblizhenije i sglazhivanie krivykh v rezhime adaptacii. Communication JINR, P10-99-168, Dubna, (1999), (in russian).
- [2] N.D. Dikoussar, Cs. Török. Automatic Knot Finding for Piecewise-Cubic Approximation. *Mathem. Model.*, **18**, 3, (2006), 23-40 (in russian).
- [3] N.D. Dikoussar, Cs. Török. Data smoothing by splines with free knots. *Particles and Nuclei, Letters*, 2007 (in print).
- [4] N.D. Dikoussar. Four-point transformation methods in approximation and the smoothing problems. *Particles and Nuclei, Letters*, 2007 (in print).
- [5] W. Härdle. Applied nonparametric regression, Cambridge University Press, (1990).
- [6] J. Friedman. A Variable span smoother. *SLAC PUB-3477*, Stanford, (1984).
- [7] B.W. Silverman, A fast and efficient cross-validation method for smoothing parameter choice in spline regression. *J. Amer. Staist. Assn.* **19**, (1984).
- [8] N.D. Dikoussar. Discrete Projective Transformations on the Coordinate Plane, *Mathem. Model.*, **10**, 3, (1991), 50-64 (in russian).
- [9] N.D. Dikoussar. Adaptive Projective Filters for Track Finding, *Comp. Phys. Commun.*, **79**, (1994), 39-51.
- [10] N.D. Dikoussar. Function parametrization by using 4-point transforms, *Comput. Phys. Commun.* **99** (1997), 235-254.
- [11] Cs. Török, N.D. Dikoussar. MS.NET Components for Piecewise-Cubic Approximation, *Commun. of JINR*, P10-2004-202, Dubna, (2004).
- [12] M. Révayová , Cs. Török. Piecewise Approximation and Neural Networks, *Kibernetika*, accepted for print, (2006).
- [13] The European Physical Journal, C. Review of Particle Physics, Springer, (2000), p.235.
- [14] L.S. Azhgirey, V.P. Ladygin, V.N. Zhmyrov, and L.S. Zolin. Proposal of the Experiment on the Investigation of the (d, p) Reaction at the Extended Deuteron Beam of the U-70 Accelerator, *Physical of Particles and Nuclei Letters*, (2007), vol. 4, N6, pp.497-502.