Automatic Knot Detection in the Piecewise-Cubic Approximation
(The Algorithm and MS.NET Components)

N.D. Dikoussar
Laboratory of Information Technologies, JINR
Cs. Török
Technical University in Košice, Slovakia

1. Motivation. The analysis of dependence between variables is one of the main tasks of technical and scientific research. The increasing use of the Internet and multimedia technologies, the transmission, storage and study of the massive volumes of records need new, high performance data compression, analysis and management tools. Methods of approximations are used every day in data analysis and information gain process, and the associated problems are of wide interest in theoretical and experimental sciences. One of the main problems in data/signal denoising, compression and forecasting is to find an optimal or good representation. Once it is achieved many things can be done about drawing information from data.

Piecewise polynomial methods and splines have been widely used. Various approaches and methods are proposed recently in this area. These include the *segment approximation problem* (or the free knots problem in the spline theory) [7], [8], smoothing spline methods and wavelet techniques [9], [10].

2. Problem Statement. The segment approximation problem is closely related to the piecewise and spline approximation problems. Spline continuity conditions at the breakpoints *are dropped* in the case of segment approximations. A search interval is divided into subintervals and an approximation problem is solved over each of these *subintervals.* It is clear that different subdivisions into subintervals lead to qualitatively different results. The main goal is to find a subdivision where the errors over the subintervals are as small as possible. The effectiveness of a spline representation of data *depends critically on their number and positions* [7], [8] and [10]. Notice that free knots optimization is a very hard non–linear problem.

3. Automatic knot detection using APCA. We suggested a new approach to the analysis of complex dependence with relatively small noise using the four point methodology [3]. The suggested algorithm LOCUSD [1] divides the interval/curve into subintervals/segments of various lengths, provides for every segment local cubic estimations and gives a technique for obtaining integral cubic approximants. Finding the breakpoints in an *auto-tracking mode and the iterative computation schemes* are the two main features of the proposed method that uses a *special* approximation model [1]. MS Visual C# components for **a**utotracking **p**iecewise **c**ubic **a**pproximation (APCA): a class library and a Windows-application have been developed too [2].

In our method neither the number of knots nor their placement are unknown. This is very important for applications in approximation and reduces real world data. The knots of the subintervals are detected *in auto–tracking mode* using a digitized curve (data points). A three–point cubic parametric model is used as a local approximant with three control (three different points at x axis), three fixed (ordinates on the curve) and one free (1/6 of a third derivative of the model) parameters. A free parameter θ is found in a line following mode, using either step–by–step averaging or the first order recursive least squares method (RLSM). A formula for expression of the free parameter via a length of the segment and values of a function and derivatives in the joining points is received. The C^1 — smoothness depends on the accuracy of the θ — estimate.

Let

$$\left\{P(x_m, \tilde{f}_m)\right\}_{m=\overline{1,N}},\ \tilde{f}_m = f(x_m) + e_m,\ x_m < x_{m+1},\ 4 \ll N\ ,$$

be a given set of data points, where $e_m \sim N(0, \sigma^2)$. We consider tetrads $T = \{P_\alpha, P_\beta, P_0, P_m\}$, $T \in \{P\}$, $P_i \neq P_j$; $i \neq j$, $i, j \in (\alpha, \beta, 0, m)$. Three points $R = (P_\alpha, P_\beta, P_0) \in T$ are called as *reference points* and P_m is a variable point. To approximate a piece of a curve f (the segment) at interval $[\alpha, \beta]$ we use a parametric cubic model (Fig. 1)

$$f \approx S = \Pi + \theta Q, \tag{1}$$

where Π is a quadratic parabola passing via reference points R and Q is a cubic parabola: $Q = \tau(\tau - \alpha)(\tau - \beta)$; θ is a free parameter. The ordinates of the reference points are used as fixed parameters: $\mathbf{r}_0 = [r_\alpha, r_\beta, r_0]^T$, where $r_* \approx f_*$ or $r_* \equiv f_*$ if $e_* = 0$. Parameters $\mathbf{a} = (\alpha, \beta, x_0)$ and τ_m are defined as $\alpha = x_\alpha - x_0, \beta = x_\beta - x_0$ and $\tau_m = x_m - x_0$. These parameters are used to evaluate weight homographic functions $\mathbf{d}_0 = [d_1, d_2, d_3]^T$ and Q. Fig. 1 shows a cubic arc S approximating a piece of a curve f on the subinterval $[\alpha, \beta]$. To obtain $\{d_i\}_{i=1}^3$ we use a single–purpose cross–ratio functions [3]. In these terms the parabola equation for the nth segment is written as $\Pi(\tau; \mathbf{a}_n, \mathbf{r}_{0n}) = (\mathbf{r}_{0n}, \mathbf{d}_n)$. θ is the free parameter which is related to a third derivative of the model: $\theta = S'''/6$. Eq. (1) and weight functions construction yield some advantages in development of algorithms: *flexibility, control, stability, simple computations* and so on.

The stability of the method w.r.t. input errors is shown as well. The factors of error suppression (K_n and $d_{in}K_n$) are shown in Fig. 2. The key parameters of the approximation are: *the parameters of the weight functions, the variance of the input errors, and a sampling step.*

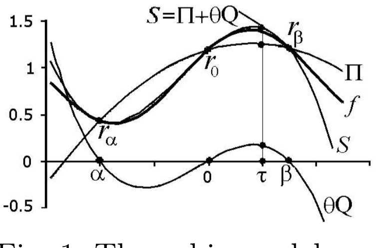

Fig. 1: The cubic model

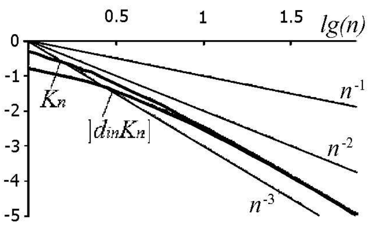

Fig. 2: $\log 10 K_n$ and $\log 10 \mid d_m K_n \mid$

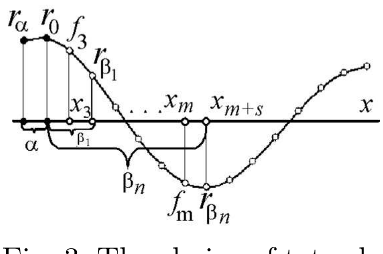

Fig. 3: The choice of tetrads

The constancy of the third derivative of the cubic model ($S''' = const$) is used as a criterion for knot detection in the dynamic mode. This value can be estimated via a local $\hat{\theta}$ using four points on the curve and Eq. 1. A choice of tetrads at every segment uses two *fixed* points (P_α, P_0) and two *variable* points $(P_m, P_{m+1\equiv\beta_n})$ (Fig. 3). To get the global estimate $\hat{\theta}$ at the whole segment we use a recurrent formula using *the first-order recursive* *least–squares method* (RLSM):

$$\hat{\theta}_n = \hat{\theta}_{n-1} + K_n\Big(\tilde{f}_m - \Pi_{nm} - \hat{\theta}_{n-1}Q_{nm}\Big), \hat{\theta}_0 = 0, m \in \{1, ..., N\}, n = 1, 2, ..., n_*, \qquad (2)$$

where $K_n = Q_{nm}/\sum_{j=1}^{n} Q_j^2$ (see Fig. 2). The number n_*, is defined as $n_* = n$ under condition $\mid \delta_m \mid > \delta$, where $\delta_m = f_m - \Pi(\tau_m; \mathbf{a}_n, \mathbf{r}_{0n}) - \hat{\theta}_n Q(\tau_m, \alpha, \beta_n)$ and δ is the given control parameter. The efficiency of the method is shown by numerical calculations on test examples (see Figs. 5 — 9) and real measurements.

To perform this analysis and construct data approximants a Windows–application was built based on class components (Fig. 4). We introduced in the .NET framework within a name space LinAlg [5] special vector and matrix types with a wide range of object and static numerical, statistical, database and visualization methods, properties and components that enable to perform in Windows and Web environment not only exploratory data analysis, but also our approximation and compression techniques, and that are extensible and manageable. LinAlg is a set of types that enables vectorial programming and incorporates a wide range of numerical, statistical and graphical methods.

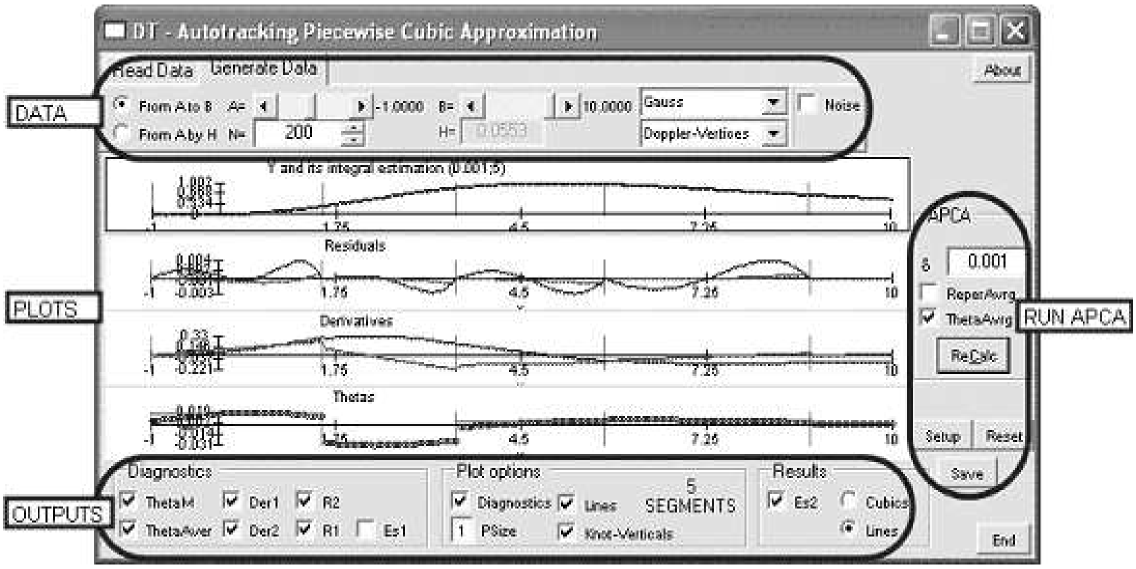

Fig. 4: The Windows application GUI

The three main objects in APCA are the data points, segments and the interval/set of all segments. Based on them we designed an object–oriented implementation of APCA in MS Visual C# [11] with three classes/components: *Point4, Segment, SegmentsAll.* Due to this architecture one can easily access a given segment and gain information about it. This feature may play a key role in the componentŸs extension connected with the generalization and improvement of APCA in the future. Although due to LinAlg one can easily view, plot or save the results and diagnostics of APCA, to simplify these tasks we created a forth class *FormShowPlotSave.*

4. Examples. Figure 5 shows the result of *knot detection* and piecewise–cubic approximation for 134 points situated on the test curve $f = 25/(x^2 + 25) - 0.55Sin(x + 2)/(x + 2)$ using LOCUSD.

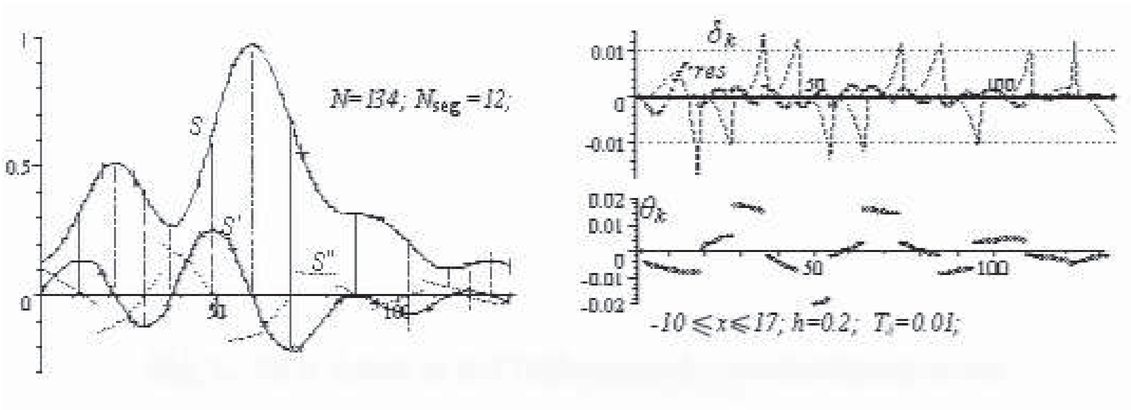

Fig. 5: Knot detection and cubic approximants for the test curve

The data of the following example were gained by numerical modeling of the electron thermal capacity (ETC) for D–acetone molecule and so they are practically without errors [6]. Fig. 6 display the data and their APCA approximants, fig. 7 the estimations of the derivations and fig. 8 the residuals. The quality of the automatically detected 25 approximants ($\delta = 0.005$) is satisfactory. The subintervals are shorter in the left part of the figures, where the graph is more dynamic.

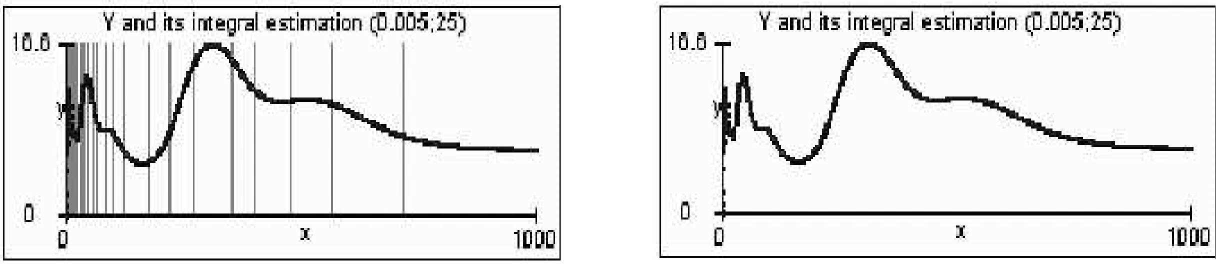

Fig. 6: ETC and its approximations [6]

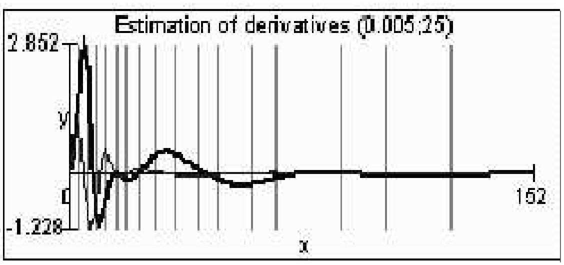

Fig. 7: First and second derivations

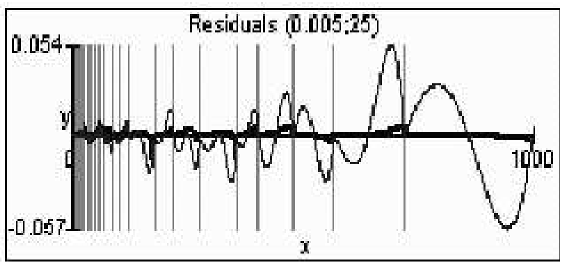

Fig. 8: Local and interval residuals

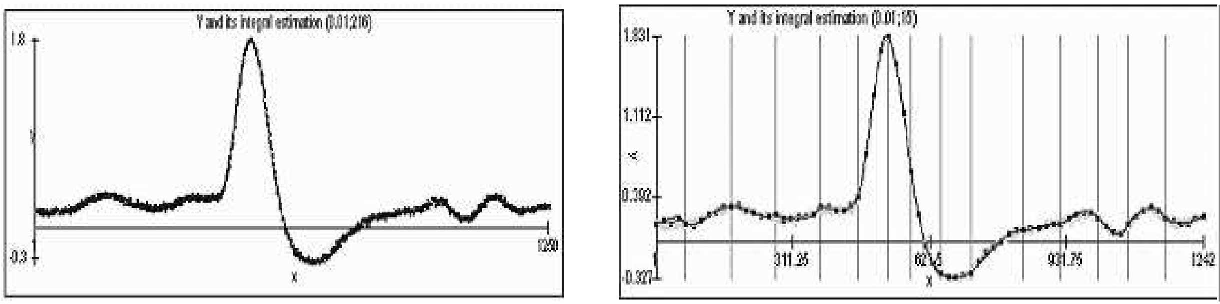

Fig. 9: Knot detection and cubic approximation for data with small noise

5. Conclusion. Let us summing up the results:

- *an automatic knot detection and a piecewise–cubic approximation method* are proposed;
- *algorithm LOCUSD, MS.NET components* and *Windows–application APCA* for segment approximation are developed;
- the continuity of the first derivatives of the approximants for functions presented by data without errors are acceptable;
- a smaller δ results in more segments with more precise approximants;
- for noisy data it is advisable to choose a greater *sampling step* and δ;
- the goal is to find such δ that yields desirable approximation quality and an acceptable count of segments.

Our plan is to develop methods, algorithms and tools for smoothing data point with a low signal to noise ratio.

References

[1] *Дикусар Н.Д. Ч. Торок.* Кусочно–кубическая аппроксимация в режиме автоматического слежения, Сообщение ОИЯИ, Р11–2004–187, Дубна, 2004; http://www.jinr.ru/publish/Preprints/2004/187/(P11–2004–187).pdf.
Дикусар Н.Д. Ч. Торок. Автоматический поиск узлов для кусочно–кубической аппроксимации, 2005. Статья принята к публикации в журнал Математическое Моделирование.
Dikoussar N.D., Cs. Török. Automatic Knot Detection in the Piecewise–Cubic Approximation. The report on the Conference PRASTAN 2005, Spoločna Slovensko–Česka Štatistika Konferencija, Banska Bystrica, 2005.

[2] *Торок Ч., Дикусар Н.Д.* MS.NET компоненты для кусочно–кубической аппроксимации, Сообщение ОИЯИ, Р10–2004–202, Дубна, 2004; http://www.jinr.ru/publish/Preprints/2004/187(P10–2004–202).pdf.

[3] *Дикусар Н.Д.* Методы 4–точечных преобразований в задачах аппроксимации и сглаживания кривых и поверхностей, 10–2002–155. Автореферат диссертации на

соискание ученой степени доктора физико–математических наук, Дубна 2002;
http://www.jinr.ru/publish/Preprints/2004/155(10–2002–155).pdf.

[4] *Kepič T., Török Cs., Dikoussar N.D.* Wavelet compression, 13. International Workshop on computational statistics, Bratislava (2004) (49–52).

[5] *Török Cs.* Visualization and Data analysis in the MS.NET Framework, JINR Commun., E10–2004–136, Dubna, 2004;
http://www.jinr.ru/publish/Preprints/2004/136(e10–2004–136).pdf.

[6] *Костенко Б.Ф., Прибыш Я.* Теплоемкость электронов в плазме, образующей-ся при схлопывании кавитационного пузырька в D–ацетоне, Сообщение ОИЯИ, P11–2004–139, Дубна, 2004;
http://www.jinr.ru/publish/Preprints/2004/139(P11–2004–139).pdf.

[7] *Wolters H.J.* A Newton–type Method for Computing Best Segment Approximations, Communications on Pure and Applied Analysis, 3(1) (2004) 133–149.

[8] *C. Conti at all.* Cubic spline data reduction choosing the knots from a third derivative criterion, Numerical Algorithms, 28 (2001) 45–61.

[9] *G. Farin.* NURBS from Projective Geometry to Practical Use (Second Edition). A K Peters Ltd, Natick, 1999.

[10] *H. Schwetlick and T. Schütze.* Least squares approximation by splines with free knots, BIT 35(1) (1995) 361–384.

[11] *Э. Гуннерсон.* Введение в C#, Издательский дом Питер, Санкт–Петербург, 2001.