E11-2008-205

A. Matejčiková[1], Cs. Török[2]

PIECEWISE SMOOTHING WITH TWO-PART POLYNOMIAL SCHEME

[1] «Kybernetika», Košice, Slovakia;
E-mail: andrea.matejcikova@gmail.com
[2] P. J. Šafárik University, Košice, Slovakia;
E-mail: csaba.torok@upjs.sk

Матейчикова А., Торок Ч.
Е11-2008-205
Кусочное сглаживание с использованием полиномиальной модели с двухэтапной схемой вычислений

В работе усовершенствован недавно предложенный метод для сегментации и сглаживания зашумленных данных кусочно-кубической моделью с автоматическим слежением, использующий реперные точки и сплайны. Чтобы избежать трудностей, связанных с применением реперных точек в задачах аппроксимации, была разработана полиномиальная модель с двухэтапной схемой вычислений. Настоящая работа является продолжением этих результатов. Ее цель — показать, что благодаря правильному размещению и оцениванию реперных точек можно не только конструировать локальные аппроксимации для найденных сегментов без использования сплайнов, но и понижать их число. Метод проиллюстрирован на реальных данных.

Работа выполнена в Лаборатории информационных технологий ОИЯИ.

Сообщение Объединенного института ядерных исследований. Дубна, 2008

Matejčiková A., Török Cs.
E11-2008-205
Piecewise Smoothing with Two-Part Polynomial Scheme

The present work enhances a recently proposed method for segmentation and smoothing noisy data that used an autotracking piecewise reference points based on cubic model and splines. To overcome the issues with the application of reference points in approximation tasks a two-part polynomial scheme was elaborated. This work is a sequel of these results. Its goal is to show that due to the appropriate localization and assessment of the reference points, the local approximants of the detected segments can be not only constructed without the use of splines, but their number can also be decreased. The method is illustrated with real data.

The investigation has been performed at the Laboratory of Information Technologies, JINR.

Communication of the Joint Institute for Nuclear Research. Dubna, 2008

INTRODUCTION

A piecewise approximation is a common way to express complex relationships between variables [1, 2]. Any method of the piecewise approximation has to answer three base tasks of this approach:

• given the segments' knots how to handle the noisy data — local smoothing;

• how to guarantee a smooth transition between segments in the shared points — global smoothness (in knots);

• how to find the knots themselves — knot detection.

In [3] the approach to piecewise smoothing was based on a three-point cubic model derived from [4]. The suggested method found successfully the knots and segment approximants for data with moderate errors. Based on the persuasive results we were convinced that the three-point approach was correct and determined to enhance the approach for data with noise of any level. So the investigation has been continuing.

At the beginning a three-point polynomial model was derived [5] leveraging the forward and backward DPT (Discrete Projective Transformation) that use three reference points. By generalization of DPT first a two-point transformation was constructed and then in [6] proposed a general r-point transformation, $r \geqslant 2$. Based on this transformation, an IZA(p, r) representation (Interpolation Zero Approximation) of polynomials was derived, where p is the polynomial's degree and r the number of reference points. This representation, establishing a reparametrization by replacing the first r coefficients with r reference points lying on the polynomial, enabled to propose a general r-point polynomial approximation model that naturally includes the original three-point model.

However, these results have not solved yet the two main problems of the reference points based approach to the parametric approximation:

• how to overcome the issue connected with the erroneous ordinates of the reference points,

• how to ensure smooth transition between the neighboring approximants.

The work [5] overcame this problem using neural networks. In [7] the proposed method for noisy data using preliminary data smoothing and thinning partially solved the tasks of the piecewise approximation and the first problem of the new reference points based approach. However it used splines to guarantee smooth transition between approximants that resulted in double number of knots in some examples.

A theoretical answer to the two problems of the reference points based approach within the r-point model is given in [6,8,9], where a two-part *R-I* scheme (Regression-Iza) is proposed to build on two, three and any r ($\geqslant 2$) reference points, respectively. They proved such important properties of the estimates as consistency [8] and asymptotic normality [6], that were achieved by

• localization of the reference points between the first ordinary LS *R* model and the second IZA *I* model near their shared point,

• assessment of the second coordinates of the reference points in the *I* model using the estimator of the first *R* model.

Thanks to these two rules a C^{r-1} smoothness with accuracy $o(\tau)$ is ensured in the common point when the model uses r reference points. For given knots the operability of this quasi smooth idea was demonstrated by an *R-I-R-...-R-I-R* scheme that used four and six reference points: two/three on the left and two/three on the right of the *I* segments [9].

The present work is a sequel to [7] and the two-part *R-I* approach. It uses an *R-I-I-...-I* scheme with two reference points on the left of every *I* segment. Its goal is to show that due to the appropriate localization of the reference points and assessment of their ordinates using the preceding approximants as described above, it is possible to grant smooth transitions between the local approximants even without splines.

1. THE r-POINT SCHEME AND QUASI SMOOTHNESS

The general r-point polynomial model is a flexible technique: various schemes built on it can be constructed such as the *R-I*, *R-I-R-...-R-I-R* and *R-I-I-...-I* schemes with important properties. Here we first give the formula of the general **r-point polynomial model** for any fix degree $p \geqslant r \geqslant 2$ with r reference points that is based on the **IZA representation** (or IZA(p, r) representation) of polynomials [6] and then a particular form of the **two-part scheme** with a quasi smooth property for $r = 2$.

Theorem 1. IZA Representation of Polynomials. *Assume* $p \geqslant r \geqslant 2$. *Then any polynomial* $P_p(x)$ *can be expressed based on its any different* r *points* $\{[x_i,\, y_i],\ y_i = P_p(x_i),\ i = \overline{0, \mathrm{r}-1}\}$ *as*

$$\begin{aligned} P_p(x) &= I_{r-1}(x) + Z_r(x) A_{p-r,r}(x) \\ &= I_{r-1}(x) + \boldsymbol{w}_r(x)^T \cdot \boldsymbol{\alpha}, \end{aligned} \tag{1}$$

where $I_{r-1}(x) = \sum\limits_{i=0}^{r-1} \Pi_i(x) y_i$ *is an incomplete interpolating polynomial,*

$$\Pi_i(x) = \prod_{v \in V_i} \frac{x - v}{x_i - v}, \quad V_i = V \backslash \{v_i\},\ \ V = \{x_0, x_1, \ldots, x_{r-1}\},\ \ i = \overline{0, r-1},$$

$$Z_r(x) = \prod_{i=0}^{r-1}(x - x_i), \quad A_{p-r,r}(x) = \boldsymbol{S}^T \cdot \boldsymbol{\alpha}, \quad \boldsymbol{\alpha} = (a_r,\ a_{r+1}\ \ldots,\ a_p)^T,$$

$$\boldsymbol{S} = (S_{0,r},\ S_{1,r}\ \ldots,\ S_{p-r,r})^T, \qquad S_{1,r} = 1,\ r \geqslant 0, \quad S_{j,0} = x_0^j,\ j \geqslant 0,$$

$$S_{j,r} = S_{j,r-1} + x_r S_{j-1,r}, \quad j \geqslant 1,\ r \geqslant 1,$$

$$\boldsymbol{w}_r(x) = (w_{0,r}(x), \ldots, w_{p-r,r}(x))^T, \qquad w_{j,r}(x) = Z_r(x) S_{j,r}.$$

The general r-point polynomial *I* model for any fix degree $p \geqslant r \geqslant 2$ with a set of r different reference points

$$\mathcal{R} = \{[x_i,\ y_i],\ y_i = P_p(x_i),\ i = \overline{0, \mathrm{r}-1}\} \tag{2}$$

is given by

$$\tilde{P}_p(x) = I_{r-1}(x) + \boldsymbol{w}_r(x)^T \cdot \boldsymbol{\alpha} + \varepsilon.$$

Hence, e.g., for $p = 6$, $r = 2$ we get a model

$$\begin{aligned}
&\tilde{P}_6(x) = \varepsilon + \frac{(x-x_1)P_6(x_0)}{x_0 - x_1} + \frac{(x-x_0)P_6(x_1)}{x_1 - x_0} + (x - x_0)(x - x_1) \times \\
&\times[a_2 + (x + x_0 + x_1)\, a_3 + +(x^2 + (x_0 + x_1)x + {x_0}^2 + x_1(x_0 + x_1))\, a_4 + \\
&+(x^3 + (x_0 + x_1)x^2 + ({x_0}^2 + x_1(x_0 + x_1))x + {x_0}^3 + x_1({x_0}^2 + x_1(v_0 + x_1)))\, a_5 + \\
&+(x^4 + (x_0 + x_1)x^3 + ({x_0}^2 + x_1(x_0 + x_1))x^2 + ({x_0}^3 + x_1({x_0}^2 + x_1(x_0 + x_1)))x + \\
&+{x_0}^4 + x_1({x_0}^3 + x_1({x_0}^2 + x_1(x_0 + x_1))))\, a_6].
\end{aligned}$$

Models with lower degree are gained easily from it putting zero into the corresponding leading coefficients.

The reference points from $P_p(x)$ in (2) are arbitrary. Their location are so far undefined and ordinates are generally unknown or noisy. We define now a two-part scheme with a quasi smooth transition between its two components due to a proper selection of the reference points. This approach, its generalization to many components, will be used for the piecewise approximation in the next section.

Consider two polynomials $P_{\boldsymbol{a}}(x)$ and $P_{\boldsymbol{b}}(x)$ over the intervals $[-1,\, 0]$ and $[0,\, 1]$, respectively, where $\boldsymbol{a} = (a_0, a_1, \ldots, a_p)^T$, $\boldsymbol{b} = (b_0, b_1, \ldots, b_q)^T$, the degrees p, q are finite and $p,\ q \geqslant 2$. The two-part *R-I* polynomial scheme is given by [8, 9, 6]

$$\begin{aligned}
\tilde{\boldsymbol{Y}}^* &= \boldsymbol{X}^* \boldsymbol{a} + \boldsymbol{\varepsilon}^*, \\
\tilde{\boldsymbol{Y}} &= \boldsymbol{I}_{\mathcal{R}_{\boldsymbol{a}}} + \boldsymbol{W}\boldsymbol{\beta} + \boldsymbol{\varepsilon},
\end{aligned}$$

where ($r = 2$)

$$\mathcal{R}_{\boldsymbol{a}} = \{\ [x_j,\ P_{\boldsymbol{a}}(x_j)] :\ \ x_j = -j\tau,\ \ j = 0, 1\ \}.$$

The model of the first part is a classical polynomial regression. The model of the second part corresponds to the second form (1) of the IZA representation of the polynomial $P_{\boldsymbol{b}}(x)$ with $\mathcal{R}_{\boldsymbol{a}}$. The key moment in the two-part scheme is $\mathcal{R}_{\boldsymbol{a}}$, or as we will see in a while, $\mathcal{R}_{\hat{\boldsymbol{a}}}$ in the estimation of $\boldsymbol{\beta}$.

It can be shown [8, 6] that due to $\mathcal{R}_{\boldsymbol{a}}$ the quasi smooth/spline conditions

$$P_{\boldsymbol{a}}(0) = P_{\boldsymbol{b}}(0), \quad P_{\boldsymbol{a}}'(0) = P_{\boldsymbol{b}}'(0) + o(\tau)$$

hold, where τ is a small positive real number, and under the assumptions

$$\begin{aligned} \widehat{\boldsymbol{a}} &\equiv \widehat{\boldsymbol{a}}_M = (\boldsymbol{X}^{*T}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*T}\tilde{\boldsymbol{Y}}^*, \\ \mathcal{R}_{\hat{\boldsymbol{a}}} &= \{\,[x_j,\, P_{\hat{\boldsymbol{a}}}(x_j)]:\;\; x_j = -j\tau,\;\; j = \overline{0, r-1}\,\}, \\ \widehat{\boldsymbol{\beta}}_N &= (\boldsymbol{W}^T\boldsymbol{W})^{-1}\boldsymbol{W}^T(\tilde{\boldsymbol{Y}} - \hat{\boldsymbol{I}}), \quad \hat{\boldsymbol{I}} = \boldsymbol{I}_{\mathcal{R}_{\hat{\boldsymbol{a}}}}, \end{aligned}$$

$\widehat{\boldsymbol{\beta}}_N$ is consistent and $\sqrt{N}(\widehat{\boldsymbol{\beta}}_N - \boldsymbol{\beta})$ — asymptotically normal.

2. QUASI SMOOTHNESS AND PIECEWISE APPROXIMATION

We speak about a quasi spline property, quasi smoothness in the presence of at least two segments — three knots. Figure 1 depicts the essence of the quasi smoothness of the *R-I-I* scheme with two *I* models over $[u_1,\, u_2]$ and $[u_2$, U]: one reference point is localized in the left knot of the *I* segment, here in u_1 or u_2, and the other at a distance of τ to the left, and the ordinates of both are estimated using the previous (*R* and then *I*) approximant.

The *R-I-I* scheme can be easily generalized to an *R-I-I-...-I* scheme with a number of *I* segments. The most crucial thing is that the first model must not be an *I* model to prevent the issue of estimating the reference points' ordinates of the first *I* model. So on the first *I* segment the estimation of the reference points is insured using the *R* approximant and on the further *I* segments using the previous *I* approximant.

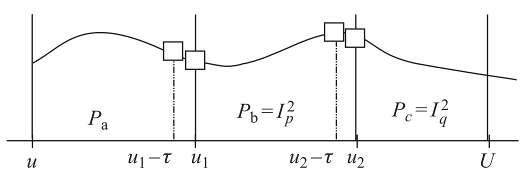

Fig. 1. The *R-I-I* scheme with $r = 2$ reference points

An autotracking piecewise approximation method is described in [7] with a recursive parameter estimation [3]. Here we give concisely the knot detection algorithm's base steps. The enhanced step is the last one, where the approximants based on the thinned smoothed points of the found segments are computed instead of splines in accordance with the *R-I-I-...-I* scheme that ensures quasi smooth transition between the approximants.

Algorithm

Consider M data points $\{[x_i, \tilde{y}_i],\ i = \overline{1, M}\}$:

1) Thinning of data by selection $N \ll M$ points and their ordinates' local estimation $\breve{y}_i,\ i = \overline{1, N}$ using the original M measurements.

2) Detection of $K + 1\,(< N)$ knots using the criterion

$$|\breve{y}_n^k - \hat{y}_n^k| < \delta, \qquad n = \overline{1, N_k},\ k = \overline{1, K}$$

from the selected and estimated $N = \sum_{k=1}^{K} N_k - K + 1$ points $\{[x_i, \breve{y}_i],\ i = \overline{1, N}\}$, where N_k is the count of these points from the k^{th} segment and

$$\hat{y}_n^k = I_n^k + \hat{\theta}_n^k W_n^k, \quad n = \overline{1, N_k},\ k = \overline{1, K},$$

$$\theta \equiv a_3,\ \hat{\theta}_0^k = 0, \quad \hat{\theta}_n^k = \hat{\theta}_{n-1}^k + \mathcal{K}_n^k(\hat{y}_n^k - I_n^k - \hat{\theta}_{n-1}^k W_n^k),$$

$$\mathcal{K}_n^k = W_n^k / \sum\nolimits_{i=1}^{n} (W_i^k)^2.$$

We mention that $\hat{y}_1^k \equiv \hat{y}_{N_{k-1}}^{k-1} = \breve{y}_1^k$ for every k and $\breve{y}_1^1 \equiv \breve{y}_1$.

3) Construction of K local approximants based on the *quasi smooth R-I-I-...-I* scheme and the estimated N selected points.

3. EXAMPLES

In the previous sections we described the essence of the local two-part *R-I* scheme and its quasi-smoothness, the scheme's generalization to a global piecewise *R-I-I-...-I* scheme and its application for autotracking piecewise approximation.

To demonstrate the effect of the two-part polynomial scheme on the autotracking piecewise approximation, we consider the real noisy data from [7] with both equidistant and nonequidistant steps. From the three data sets the first one shows the most complex relation. Figures 2–4 contain the original data of length M and their piecewise approximants. The verticals denote the boundary of the K segments and the squares — the reference points. We also provide in brackets the number of segments gained in [7], and the number of the selected and estimated data points denoted by N. In all pictures an IZA model was used for the *I* segments with $\tau = 0.01$.

Figure 2 illustrates smoothing data with an equidistant step, the cross sections for $\pi - p$ collision [10]. In Figures 3, 4 we give the smoothing results of two data sets with nonequidistant step. Figure 3 illustrates the cross sections for np collision [11] and Fig. 4 — the concrete resistance ratio.

As we can see, thanks to the enhanced autotracking knot detection the number of knots was reduced two times in the first two examples. The quality of the piecewise approximants are adequate and acceptable in all cases.

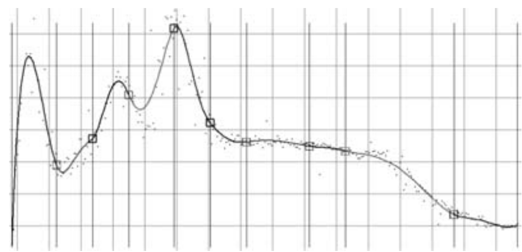

Fig. 2. Cross section for $\pi - p$ collision: $M = 277$, $N = 57$, $K = 10(20)$

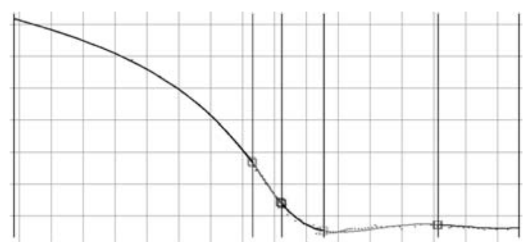

Fig. 3. Cross section for np collision: $M = 325$, $N = 40$, $K = 5(10)$

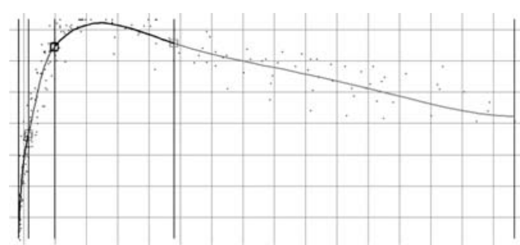

Fig. 4. Concrete resistance ratio: $M = 196$, $N = 18$, $K = 4(5)$

4. DISCUSSION

We have succeeded in solving the above described first two tasks of the piecewise approximation within the r-point polynomial model: the quasi smooth

R-I-I-...-I scheme guarantees both the local and global smoothness. A question arises: Is it possible within the framework of the r-point polynomial model to solve the third task of the piecewise approximation, knot selection, too?

Before answering it, let us discuss briefly the APCA method [3]. Today we know how to enhance it for noisy data due to the quasy smoothness. When looking for the next knot in APCA, we held (intuitively and rightly) two reference points together and on the left, and the third moved (just this, holding two reference points together and on the left, justifies us to leverage the quasi smooth *R-I-I-...-I* scheme instead of splines in APCAS [7]). However, after finding the knot, the reference points for computing the integral estimate were allocated uniformly that assured in general only continuity but not smoothness. This and the nonadequate estimate of the reference points' ordinates caused that APCA was not suited for approximating noisy data.

While the *R-I-R-...-R-I-R* scheme from [9] enables parallelization (first compute independently the *R* approximants and then the *I* approximants, each of them based on the two neighboring *R* approximants), the scheme *R-I-I-...-I* allows recursive estimation that is necessary for autotracking piecewise approximation. So we expect that leveraging the quasi smoothness of the two-part polynomial *R-I* scheme, laid in the base of the underlying *R-I-R-...-R-I-R* and *R-I-I-...-I* schemes for local approximants, we can enhance the APCA algorithm for knot detection further to make it applicable straight to noisy data with no preliminary thinning and smoothing. In regard to the theoretical justification, there are ideas, which require further consideration.

Acknowledgement. We thank A. Ďuricová and M. Rovňák for providing the data on the concrete resistance ratio. The work was partially supported by VEGA Grant 1/4003/07 of MŠ SR.

REFERENCES

1. *Seber G. A. F.* Linear Regression Analysis. N. Y.: J. Wiley and Sons, 1977.

2. *Eubank R. L.* Nonparametric Regression and Spline Smoothing. Marcel Dekker, Inc., 1999.

3. *Dikoussar N. D., Török Cs.* Automatic Knot Finding for Piecewise-Cubic Approximation // Mat. Model. 2006. V. 17, No. 3.

4. *Dikoussar N. D.* Function Parametrization by Using 4-Point Transforms // Comp. Phys. Commun. 1997. V. 99. P. 235–254.

5. *Révayová M., Török Cs.* Piecewise Approximation and Neural Networks // Kybernetika. 2007. V. 43, No. 4. P. 547–559.

6. *Török Cs.* Reference Points Based Transformation and Approximation. UPJŠ Košice, 2008 (to be published).

7. *Dikoussar N. D., Török Cs.* Data Smoothing by Splines with Free Knots // Part. Nucl., Lett. 2008. V. 5, No. 3. P. 324–327.

8. *Matejčiková A., Török Cs.* Recursive Estimation in a Quadratic Model // VIII Conf. of CeF TU, Košice, Slovakia, 2007.

9. *Révayová M., Török Cs.* Reference Points Based Approximation // VIII Conf. of CeF TU, Košice, Slovakia, 2007.

10. The European Physical Journal, C. Review of Particle Physics. Springer, 2000. P. 235.

11. *Azhgirey L. S. et al.* Proposal of the Experiment on the Investigation of the (d, p) Reaction at the Extended Deuteron Beam of the U-70 Accelerator // Part. Nucl., Lett. 2007. V. 4, No. 6. P. 497–502.

Received on December 30, 2008.

Корректор *Т. Е. Попеко*

Подписано в печать 5.03.2009.
Формат $60 \times 90/16$. Бумага офсетная. Печать офсетная.
Усл. печ. л. 0,68. Уч.-изд. л. 0,96. Тираж 310 экз. Заказ № 56528.

Издательский отдел Объединенного института ядерных исследований
141980, г. Дубна, Московская обл., ул. Жолио-Кюри, 6.
E-mail: publish@jinr.ru
www.jinr.ru/publish/