THE EXPERIMENTAL DATA APPROXIMATION USING BEZIER CURVES

Tatyana Romanenko, Vladimir Kalmykov, Vitaliy Vishnevskey
Institute of Mathematical Machine and System Problems NAS of Ukraine
e-mail: romanenko@immsp.kiev.ua

I. INTRODUCTION

The experimental data usually represent the measurement process, which has been distorted by noise. The process is fully determined by some unknown function $y= f(x)$. Let the measured value corresponds to a single-valued function y = f (x), and as a result of measurement, we have a finite sequence of N pairs $\{x_i, y_i\}$; $i = 0,N$. In most cases experimental data are represented as the graphical curves. The graphical curves, i.e. graphs are, apparently, the simplest and a long ago in-use means of cognitive presentation of experimental data in the most different scopes of human activity which allow to estimate clearly the qualitative property of the process, in spite of clutters, measurement errors. Graphs, displaying the same processes or objects, can substantially differ from each other by scales, amount of the used measurements, level noises, and so on. At the same time the feature form of the graphical curve characterizes the parameters of the displayed object or process.

Automatic (automated) processing of the graphical curves supposes comparison of their forms to set, whether the different graphs characterize the same or different processes or ojects. The direct using the neuronet methods or statistical pattern recognition methods to solve this task, is impossible, because different graphs relating to the same object can differ from each other on such parameters as scale, noise, number of measurements, etc. Thus, the graphs should be approximated by functions that are invariant to affine transformations to make it possible to compare the graphs. Usually splines are used to approximate functions that are given in tabular form [1]. Among the various types of splines, Bezier curve (Fig.1) is most convenient to approximate the graphs. So, the experimental data – graphical curves may be replaced with the analytical curves that are congruent enough to the initial graph form, and, at the same time, invariant relatively to the scale change, amount of measurements, noises level. Such curves may be chosen as the polynomial functions $x(t)$, $y(t)$ of n power of the parameter t :

$$x(t) = a_0 t^n + a_1 t^{n-1} + \ldots + a_{n-1} t + a_n,$$
$$y(t) = b_0 t^n + b_1 t^{n-1} + \ldots + b_{n-1} t + b_n. \qquad (1)$$

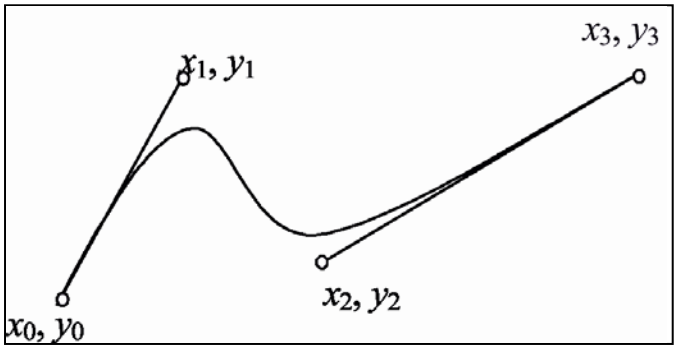

Fig. 1. Bezier curve

Choice of n – the power of polynomials depends on complication of the experimental curves to be approximated.

If such approach would be realized it is possible to hold all the substantial form features of the experimental curve and to eliminate the noises influencing. In addition, instead of the graph description in the space of initial signals it is possible to operate approximating polynomials in the space of their coefficients. The substantial advantage of such description is its invariance relatively the amount of the measured signals in every graph.

A method to approximate the graphics splines is known for a given amount of points is not large, tentatively – up to 10 [2]. It is believed that the spline curve must contain all of the given graphs points. In this case, the system of N equations must be solved to obtain the polynomial coefficients. But if the number of points in the graph is a few hundred, and in the

presence of noise, the problem ceases to be trivial and does not have the simple decision [3]. One of this task decisions is proposed in the article.

II. FORMAT PROBLEM DEFINITION

The graph can be represented by a polyline, the kinks of which form the sequence of N experimental values of some function $y(x)$. Let us consider the Bezier curve as an approximating curve because of the most often using in practical applications, parametrically defined polynomial of the third degree.

Under the Bezier curve we understand the arc of flat curve of the third degree [3]:

$$Bz(t) = \sum_{j=0}^{3} Br_j(t)\, Q_j \tag{2}$$

where $Br_j(t)$ are base Bernstein polynomials of the third degree,

Q_j are the coefficients of curve t is an real parameter which changes in an interval [0, 1].

Bernstein polynomials are defined as:

$$Br_j(t) = C_3^j (1-t)^{3-j} t^j, \quad 0 \le j \le 3 \tag{3}$$

where C_3^j, $0 \le j \le 3$ are the binomial coefficients

$$C_3^j = \frac{3!}{j!(3-j)!} \tag{4}$$

Thus, the Bezier curve (fig. 1) can be set by two polynomials of the third degree depending on the parameter t:

$$\begin{aligned} X(t) &= x_0(1-t)^3 + x_1(1-t)^2 t + x_2(1-t)t^2 + x_3 t^3 \\ Y(t) &= y_0(1-t)^3 + y_1(1-t)^2 t + y_2(1-t)t^2 + y_3 t^3 \end{aligned} \tag{5}$$

The experimental data approximating as Bezier curve has many of advantages. Describing the wide enough class of curves, Bezier curves use reasonable quantity of coefficients are 8. Coefficients (x_0,y_0), (x_3,y_3) are the co-ordinates of initial and eventual points of approximating curve arc, accordingly, and coefficients (x_1,y_1), (x_2,y_2) are the co-ordinates of points which control its form (so-called "whiskers"). Changing coefficients or moving points proper to them on the screen of monitor (both initial and eventual and controls), an user can edit the form of approximating curve in the interactive mode, improving the results of automatic approximation.

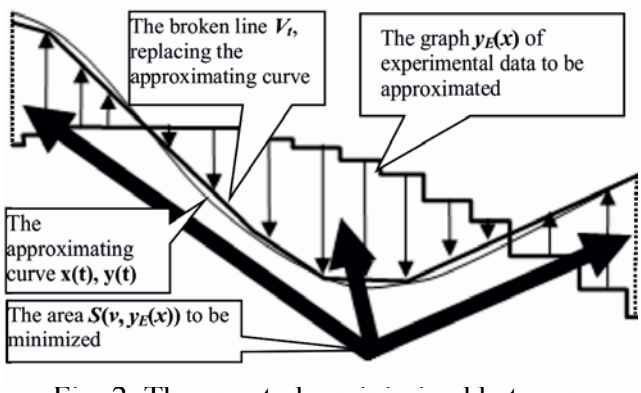

Fig. 2 The area to be minimized between the broken line $\boldsymbol{V_t}$, replacing the approximating curve **x(t), y(t)**, and the graph $\boldsymbol{y_E(x)}$ of experimental data to be approximated

Thus, every graph may be represented as a vector $v = \{x_0,y_0, x_1,y_1, x_2,y_2, x_3,y_3\}$, and its components are the coefficients of Bezier curve equations. The graph description as the vector v, always has the same amount of components regardless of number of measured experimental points, that makes solving of different tasks, in particular, tasks of the graphs recognition more simple.

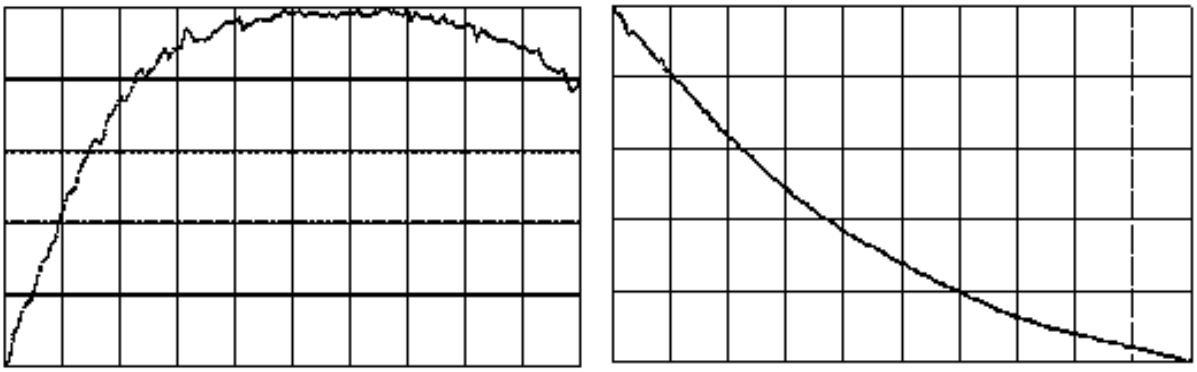

Fig.3. Examples of blood preparation spectrograms: on the left is spectrogram of sick, on the right - healthy man

It was noticed before, that determination of Bezier curve coefficients, most exactly approximating one or another graph, is not a trivial task. The iterative method is used to determine the coefficients of Bezier curve, approximating some experimental sequence of points, by minimizing the distances squares sum from every point of the graph up to the approximating curve [2]. Such a formulation of the problem follows from the assumption that the considered graphs are formed by mutually independent points, and do not reflect some of the functions. To calculate the distance from every point of the graph to the approximating curve, proper nearest point on the approximating curve must be found. Also, this task is iterative as the approximating curve is parametrically defined. When the quantity of points on the graph is large (more than 100), and the amount of iterations is not limited beforehand, such a method of task solving has resulted in substantial calculable difficulties. Also, the spline approximation in [3] uses the assumption of mutual independence of the experimental data points, but the computational problems do not arise when using the method due to the small number of points.

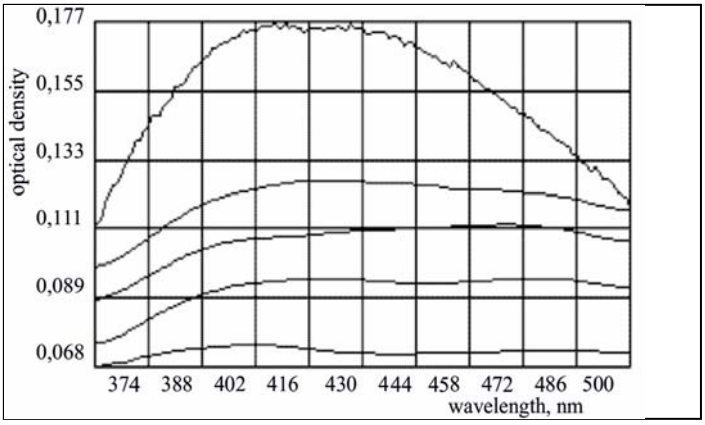

Fig. 4. The blood preparation spectrograms if the oncology diseases are of the same localization (stomach)

On the contrary we believe the right to consider all points of the experimental data related because they belong to some function. Consequently, the neighboring points of the graph must be connected by line segments. Thus, the graph of experimental data is a broken line.

We propose to accept the area $S(v,y(x))$of the figure, limited by the contour, formed with the approximating and experimental curves as the measure of likeness of this experimental graph with the approximating curve. So the values of Bezier curve coefficients v_{opt}, which approximates certain experimental graph in the best way, correspond to the minimum value $S(v,y_E(x))$:

$$v_{opt} = \arg\min_{v} S(v, y_E(x)) \qquad (6)$$

Algorithm

It is follows from (6) that computation of the area which is restricted by the contour $S(\boldsymbol{v},y_E(x))$, is basic operation of algorithm for searching the optimal coefficient of $\boldsymbol{v}_{\mathbf{opt}}$. The direct calculation $S(\boldsymbol{v},y_E(x))$ is impossible, since this contour is not simply connected.

The graph $y_E(x)$ and Bezier curve intersect each other many times. It is difficult to compute the co-ordinates of intersections because of Bezier curve is parametrically defined.

We propose to replace the Bezier curve to a polylineV_t. The fracture points of the polyline (x_t,y_t) belong to that Bezier curve and are calculated for the $t = 0, \square, 2\square, 3\square,...,1$. The value $\square$ is chosen sufficiently small, in order while replacing of the initial curve to the polyline Vt, the accuracy to be enough for the practical calculations. As a result, made to simplify the explicit value of the approximating polyline $V_t(x_n)$ directly corresponds to each value $y(x_n)$ of the graph (Fig.2). Then the area $S(\boldsymbol{v}, y_E(x))$ can be calculated as

$$S(\nu, y_E(x)) = \sum_{n=0}^{N} \left| V_t(x_n) - y_E(x_n) \right| \qquad (7)$$

The value $\boldsymbol{v}_{\mathbf{opt}}$, corresponds to the minimum value $S(\boldsymbol{v},y(x))$, can be got using the gradient method.

III. Experimental Verification of Algorithm

The proposed algorithm was experimentally tested on more than 1,100 real the graphs of the spectrograms, and preparations were obtained by the method of Oncotest [4]. The method to diagnose early oncology diseases allows to

define presence or absence of oncology disease using the spectrograms of blood preparations (Fig.4). At present researches are conducted also and for determination of localization of disease.

The spectrograms of blood preparations are represented in Fig.4. Their diagnoses are the oncology disease of identical localization (stomach). In spite of obvious distinctions of spectrograms on the levels of signals, hindrances, form of curves have undoubted likeness. The form of curves is specific property of the resulted graphs, which can correspond to the diagnosis. The identification task to form a diagnosis by spectrograms supposes additional researches which must be based on treatment of plenty of experimental information. In the same queue, the processing of large data arrays is impossible without the decision of tasks of their computer treatment, and including the tasks of automatic or automated classification of spectrograms on their form must be decided, using the methods of patterns recognition.

The direct use of statistical pattern recognition methods for the automated diagnosis forming in this case is not possible owing to large changeability of scale (dozens of one times), and, also, configuration of function of realizing the same appearance. Therefore it is appeared expedient to find some suitable Bezier curve for every graph, which polynomial coefficients, after adduction to one scale, can be used as pattern parameters, invariant to the changes of scale, amount of measures, noises level.

Examples of spectrograms and approximating them Bezier curves are represented on a fig. 5. A result was achieved in most cases for 15 - 40 iterations (in especially difficult cases - to 300). As it is obvious from figures, approximating the Bezier curves give the satisfactory approaching.

V. Conclusion

A new method for approximating the graphs that reflect experimental data is proposed. From the known methods of the proposed method differs in that it is based on the assumption that points of the graph represent a function, and therefore are connected and form a connected path.

We propose to evaluate the closeness of the approximating function to the graph, determining the area of the closed contour, which they form together.

To simplify the calculations, we propose to replace the approximating function to broken line, the nodes are the points of this function, and the nodes are chosen so that the accuracy of the approximation corresponds to specified requirements

The offered algorithm allows finding the parameters of Bezier curve, which approximates the experimental graph. The examples of approximation of spectrograms are resulted for implementation of the automated diagnostics. It will enable in the future to automatize the process of screening examination of population with the purpose of early detection of malignant tumors.

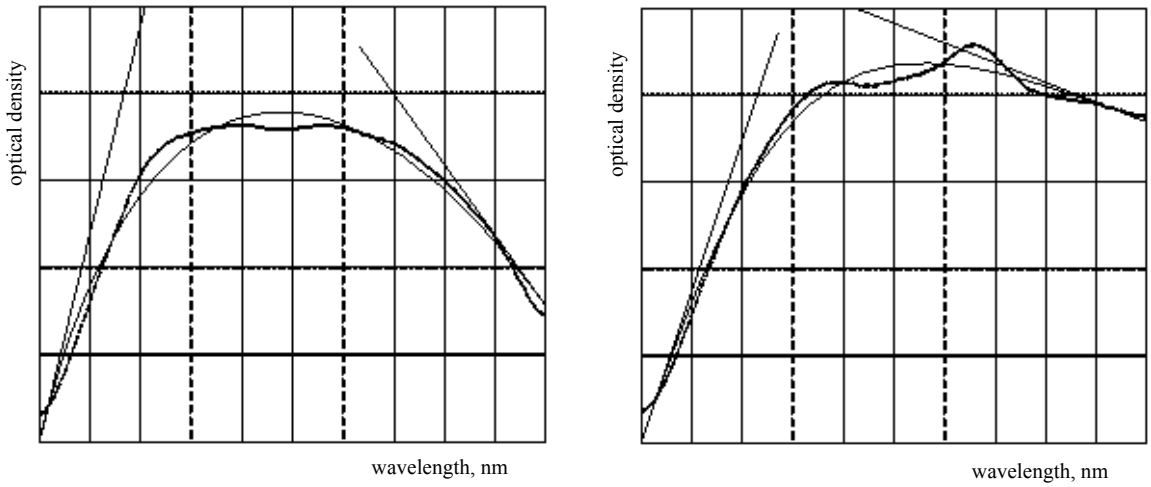

Figure 5. Spectrograms of blood preparations and their approximation by Bezier curves; localization is a mammary gland

REFERENCES

1. J.H. Ahlberg, E.N. Nilson and J.L. Wash, The Theory of Splines and Their Applications, Academic Press, New York, 1967.

2. D.F. Rogers, J.A. Adams, Mathematical Elements for Computer Graphics, McGraw-Hill, New York, 1976.

3. V.V. Vishnevsky, I.K. Rystsov, and M.V. "Voljeva, Iteratsionny algoritm postroyeniya krivoy Bezie po zadannymi tochkami", Matematicheskie mashiny i sistemy, no. 4, pp. 108-116, 2004 (in Russian).

4. V.V. Vishnevsky, V.A. Vladimirov and T.N. Romanenko, "Programmno-apparatny kompleks Oncotest –WM-01", Ukrainian Journal of Telemedicine and Medical Telematics, vol. 4, no.2 pp. 182-185. 2006 (in Russian).