

Evaluating Distance Approximation for Implicit Curve Fitting

Marina Goncharova, Alexei Uteshev
 Saint Petersburg State University
 St. Petersburg, Russia
 m.goncharova,a.uteshev@spbu.ru

Arthur Lazdin
 ITMO University
 St. Petersburg, Russia
 lazdin@mail.ru

Abstract—The curve or surface fitting problem to the given data set is treated as the one of comparison of the distance function with its suggested approximation. The comparison is performed in terms of statistical characteristics for the two sets of random variables. The approach is exemplified via the sets generated by single ellipse or a pair of ellipses.

I. INTRODUCTION

The curve or surface fitting problem is one of the nonlinear optimization challenges with multiple applications in clusterization, computer graphics and computer vision, image and video processing [1], [2], [3], [4]. Given data set

$$P = \{p_i\}, p_i \in \mathbb{R}^n, i = \overline{1, N}, n \in \{2, 3\},$$

the problem is stated as of selection the manifold best fitting to P in the meaning of minimization of an appropriate metric. This manifold is frequently selected in an implicit form

$$G(X) = 0, X \in \mathbb{R}^n, n \in \{2, 3\} \quad (1)$$

where G stands for a polynomial, whereas metric is taken as a sum of squares of the (geometric) distances from the points from P to (1):

$$\sum_{i=1}^N d^2(p_i, G) \rightarrow \min .$$

For practical reasons, the solution to this problem is sought in the set of polynomials of the lowest possible degree. However, even in the case $\deg G = 2$, i.e. quadric polynomials, solution is met several obstacles one of which is related to the problem of distance evaluation from a point to a quadric. For the aims of parametric synthesis, the distance is to be represented explicitly as a function of parameters involved into the problem, i.e. the coordinates of a point and coefficients of $G(X)$. Such a formula is hardly expected for the general case of $G(X)$ and, therefore, in modeling practice, the true distance function is replaced by its easier evaluated approximations. Dozens of such approximations have been suggested in current literature for the case of planar quadrics; for the qualitative comparison of some of them we refer to [5], [6], [7].

Extensions of these formulas to the 3D case or to the higher order planar curves is not always possible. An example of such a universal formula is given by the so-called Sampson's distance [8], [9], [10], [11]:

$$d_1(p_i, G) = \frac{|G(p_i)|}{\|\nabla G(p_i)\|}. \quad (2)$$

Here ∇G stands for the gradient column vector and $\|\cdot\|$ is the Euclidean norm.

A new point-to-algebraic-manifold distance approximation can be expressed by the formula

$$d_2(p_i, G) = \quad (3)$$

$$\frac{|G(p_i)|}{\|\nabla G(p_i)\|} \cdot \left(1 + \frac{\nabla G^T(p_i) \cdot \mathcal{H}(G(p_i)) \cdot \nabla G(p_i)}{2\|\nabla G(p_i)\|^4} G(p_i) \right).$$

Here T stands for transposition, $\mathcal{H}(G)$ is the Hessian of $G(X)$.

The validity of this approximation has been verified in [12] for the case of 2D and 3D quadrics [12] and for some planar algebraic curves in [13]; in the latter paper, this visual comparison of the level curves $d_j(X, G) = \text{const}$ with the offset (equidistant) curves to (1).

In the present paper we deal with the formula (3) validation for arbitrary algebraic curves via generation of data set with predefined estimations of the distance values. This approach should be treated as a generalization of analysis of the point-to-ellipse approximations carried out in [5], [6], [7].

II. PLANAR CURVES AND VISUALISATION

We start a visual comparison of the formulas (2) and (3) with the verified in [12] case of the quadric

$$G(X) := X^T \mathbf{A} X + 2B^T X - 1 = 0,$$

where $\mathbf{A} = \mathbf{A}^T \in \mathbb{R}^{n \times n}$, $n \in \{2, 3\}$, and $\{X, B\} \subset \mathbb{R}^n$ are the column vectors. Formula (2) is represented as

$$d_1 = \frac{1}{2} \cdot \frac{|G(p_i)|}{\sqrt{(\mathbf{A} p_i + B)^T (\mathbf{A} p_i + B)}}. \quad (4)$$

and formula (3) is represented as

$$d_2 = d_1 \left(1 + \frac{1}{4} \frac{(\mathbf{A} p_i + B)^T \mathbf{A} (\mathbf{A} p_i + B)}{[(\mathbf{A} p_i + B)^T (\mathbf{A} p_i + B)]^2} G(p_i) \right). \quad (5)$$

As an example we take the ellipse

$$G_1(x, y) := \frac{x^2}{16} + y^2 - 1 = 0$$

with the aspect ratio = 4 to better visualize distortions, which both formulas (4) and (5) produce. Here on Fig. 1 and Fig. 2 and further in this section we show level curves $d_j(X, G) =$

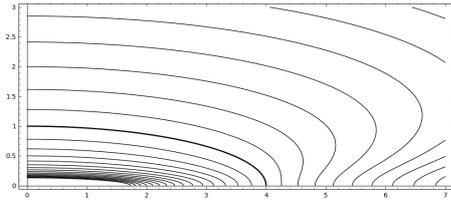


Fig. 1. Level curves $d_1 = const$ for ellipse $G_1(x, y) = 0$ (indicated bold)

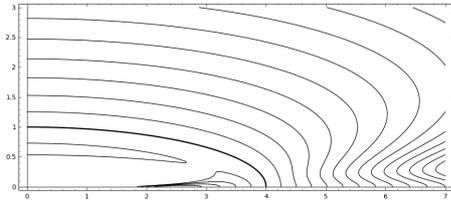


Fig. 2. Level curves $d_2 = const$ for ellipse $G_1(x, y) = 0$ (indicated bold)

$const$ from 0.25 to 3.00 in increments of 0.25. Analysis of the appearance of distortions in certain areas was in [12].

Fig. 3 and Fig. 4 illustrate how formulas (2) and (3) perform in the case of the complex high-order reference curve [2]

$$G_2(x, y) := \left(y^5 + x^3 - x^2 + \frac{4}{27} \right) \left(\frac{x}{2} + 1 \right) = 0.$$

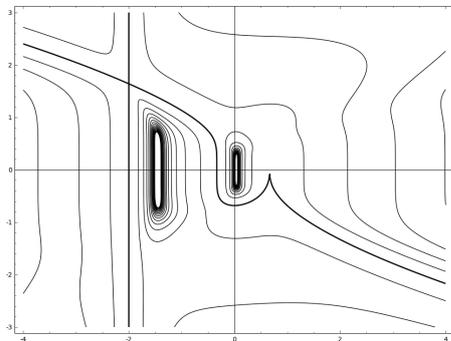


Fig. 3. Level curves $d_1 = const$ for high-order curve $G_2(x, y) = 0$ (indicated bold)

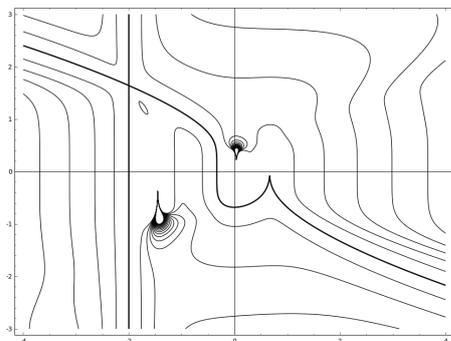


Fig. 4. Level curves $d_2 = const$ for high-order curve $G_2(x, y) = 0$ (indicated bold)

Despite the fact that the distortions seem huge, the usage of Sampson’s distance in the fitting process on noised random point cloud based on curve $G_2(x, y) = 0$ shows an appropriate result [3].

For the last example in this section we choose a well-known curve, so-called Folium of Descartes

$$G_3(x, y) := x^3 + y^3 - 3xy = 0.$$

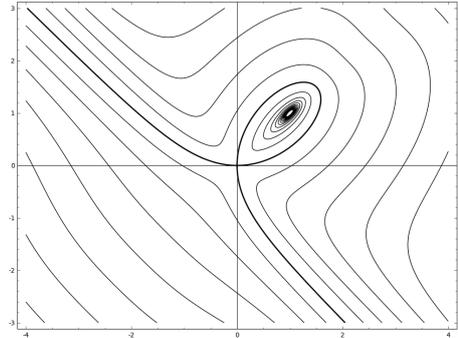


Fig. 5. Level curves $d_1 = const$ for Folium of Descartes $G_3(x, y) = 0$ (indicated bold)

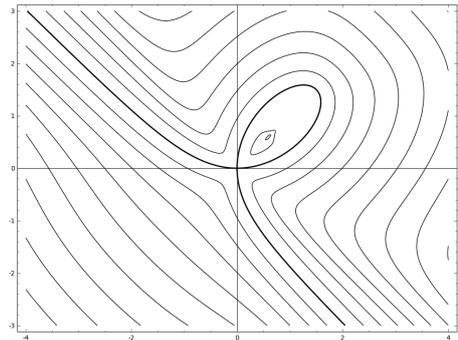


Fig. 6. Level curves $d_2 = const$ for Folium of Descartes $G_3(x, y) = 0$ (indicated bold)

The level curves for both formulas (2) and (3) have neither visible significant distortions nor excessive curvature bias.

In the following sections there will be more quantitative analysis.

III. QUANTITATIVE ASSESSMENTS

The best ellipse fitting problem has been intensively studied in the literature. Since we are able to treat the general type of algebraic curve, we concern ourselves with the case of a pairs of ellipses which is of importance to the **multiple** ellipse fitting problem [14], [15], [16]. We consider both cases, namely when the ellipses are separated or, on the contrary, have a nonempty intersection.

The test data set is composed in the following way. We utilize the analytical expressions for the offset curves and compose a grid from these curves on varying the distance from 0.1 to 3.0 in increments of 0.1. We then find the intersections of these curves with lines passing through the center of one of the ellipses. In this way, we generate about 3500 of points

with nearly equally distributed distance values in the interval $[0.1, 3.0]$.

The quality of distance approximations by the formulas (2) and (3) will be estimated with the aid of characteristics similar to those utilized in [6].

A. Linearity

To find a linear correlation between the true distance value and its approximates, we compute the Pearson correlation coefficient. To compute it, we first find the mean value μ_i for distance approximation for the points of the set lying at the distance z_i from the curve.

$$\hat{L} = \frac{\sum_i (z_i - \bar{z}_i)(\mu_i - \bar{\mu}_i)}{\sqrt{\sum_i (z_i - \bar{z}_i)^2 \sum_i (\mu_i - \bar{\mu}_i)^2}}. \quad (6)$$

Together with \hat{L} we also compute its counterparts \hat{L}^+ and \hat{L}^- for the equidistant curves lying outside or inside the treated curve (assuming distances inside and outside are signed, or can be made so). The closer to ± 1 is the value (6), the closer to linear is the dependency of true value on the distance to it approximation.

B. Curvature bias

To investigate the deviation ranges of the approximations from the fixed Euclidean distance, we calculate the variance σ_i^2 for the points of each offset (equidistant) curve lying at the distance z_i from the initial curve. Similar to those utilized in [6] we combine local measures σ_i^2 to give a global measure

$$\hat{C} = \sum_i \sigma_i^2. \quad (7)$$

Together with \hat{C} we also compute its counterparts \hat{C}^+ and \hat{C}^- for the equidistant curves lying outside or inside the treated curve.

The closer to 0 is the value (7), the better fitting quality can provide considered distance approximation.

C. Asymmetry

We compute the mean approximate signed distance μ_i^+ and μ_i^- along corresponding offset contours inside and outside the treated curve. Then we calculate asymmetry at each contour pair as the normalized difference in their mean approximate distances

$$a_i = \frac{|\mu_i^+ - \mu_i^-|}{\mu_i^+ + \mu_i^-}.$$

Again, similar we produce a global measure

$$\hat{A} = \sum_i a_i$$

which would be better to be close to zero to produce the better fitting by the considered distance approximation.

IV. RESULTS

We treated two pairs of random ellipses, one pair is separated, another one has nonempty intersection. The assessment measures discussed in the previous section was computed for every pair. In addition, in each case a level curves $d_j = \text{const}$ visualization similar to the one discussed above is provided.

A. Ellipses with nonempty intersection

To produce the quantitative assessments we take the pair of ellipses

$$G_4(x, y) := (x^2 + 9(y - 20)^2 - 3600) \times \\ (x^2 + xy + y^2 - 20x - 60y - 1000) = 0,$$

scaled in comparison to those who were visualized on Fig. 7 and Fig. 8.

Distance approximation (3) takes the form

$$d_2 := \frac{M \cdot |G_4|}{2\sqrt{k}N^2}$$

here

$$M := 814x^{12} + 6644x^{11}y - 137200x^{11} + \\ 76796x^{10}y^2 - 3201920x^{10}y + 24152400x^{10} + \\ 374388x^9y^3 - 23294560x^9y^2 + 410609200x^9y - \\ \dots \\ + 16436843481600000000y^5 + 352168819200000000y^4 - \\ 202128721920000000000y^3 - 2603404800000000000y^2 + \\ 1058158080000000000000y + 335923200000000000000, \\ N := 17x^6 + 64x^5y - 1320x^5 + 623x^4y^2 - 26040x^4y + \\ 178400x^4 + 1344x^3y^3 - 83040x^3y^2 + 1392000x^3y - \\ \dots \\ + 1377y^6 - 204120y^5 + 9525600y^4 - 117936000y^3 - \\ 1257120000y^2 + 18144000000y + 129600000000, \\ k := ((x^2 + 9(y - 20)^2 - 3600)(2x + y - 20) + \\ 2x(x^2 + xy + y^2 - 20x - 60y - 1000))^2 + \\ ((x^2 + 9(y - 20)^2 - 3600)(x + 2y - 60) + \\ (x^2 + xy + y^2 - 20x - 60y - 1000)(18y - 360))^2.$$

TABLE I. NORMALIZED ASSESSMENT RESULTS \hat{L} , \hat{L}^+ AND \hat{L}^- FOR ELLIPSES WITH NONEMPTY INTERSECTION

	\hat{L}	\hat{L}^+	\hat{L}^-
d_1	1.0000000	1.0000000	1.0000000
d_2	0.9997189	0.9999210	0.9997744

Looking at the Fig. 9 and the Table I we can note that both approximations (2) and (3) exhibit linear relationship with Euclidean distance at least close to curve bounds. Formula (2) outperforms formula (3) not only inside initial curve as we

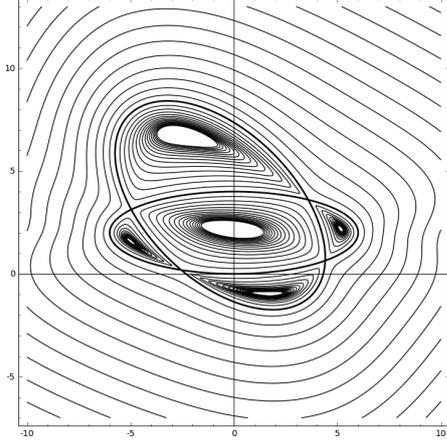


Fig. 7. Level curves $d_1 = const$ for ellipses $G_4(x, y) = 0$ (indicated bold)

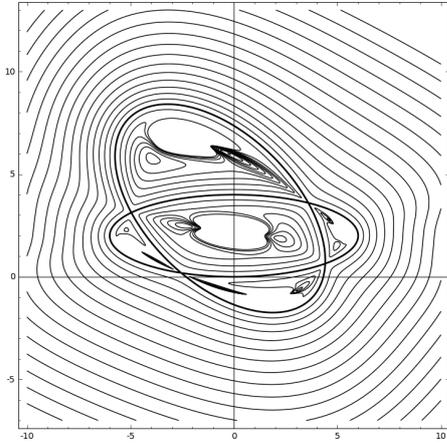


Fig. 8. Level curves $d_2 = const$ for ellipses $G_4(x, y) = 0$ (indicated bold)

could suppose due to the visualization analysis, but the outside too. Global measure also is better for d_1 .

As we might assume after viewing some visualizations that emphasize the unstable behavior of formula (3) inside the curve, the best performed formula in the asymmetry category should not be easily determined, but Fig. 10 shows that new point-to-algebraic-manifold distance approximation performed well.

Fig. 11 shows the variance of the distance approximations for each level contour. Variance increases fast with Euclidean distance from the curve, especially for the internal points.

In addition, looking at the Table II, we can note, that outside the initial curve formula (3) shows considerably better results in curvature and asymmetry assessments.

TABLE II. NORMALIZED ASSESSMENT RESULTS \hat{A} , \hat{C} , \hat{C}^+ AND \hat{C}^- FOR ELLIPSES WITH NONEMPTY INTERSECTION

	\hat{A}	\hat{C}	\hat{C}^+	\hat{C}^-
d_1	1.0000000	1.0000000	1.0000000	1.0000000
d_2	0.1401102	0.5829589	0.2283491	0.8366350

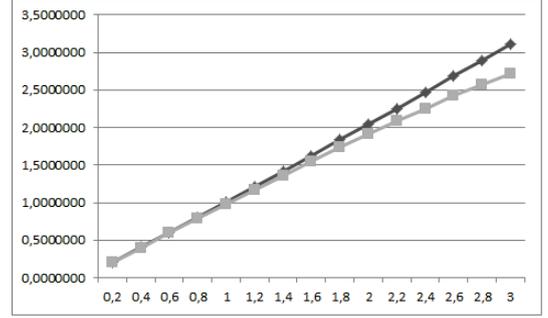


Fig. 9. Mean distances μ_i of approximated distances d_1 (black) and d_2 (grey) along offset curves plotted against the corresponding Euclidean distances

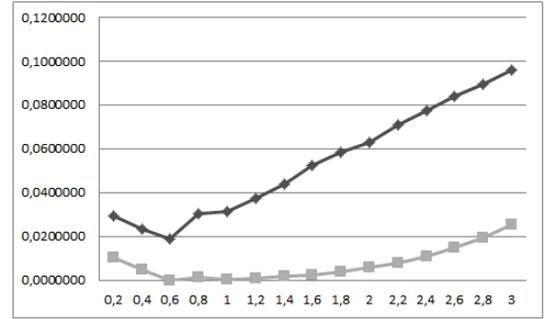


Fig. 10. Asymmetry \hat{A} of approximated distances d_1 (black) and d_2 (grey) along offset curves plotted against the corresponding Euclidean distances

B. Separated ellipses

Again, we take the pair of ellipses

$$G_5(x, y) := (x^2 + 16y^2 - 1600) \times (x^2 + xy + y^2 - 100x - 130y + 3600) = 0$$

scaled in comparison to those who were visualized on Fig. 12 and Fig. 13.

Here we give the implicit equation of equidistant curves in the form

$$Eqv(x, y, z) = 0,$$

where

$$Eqv := (x^8 + 34x^6y^2 + 28x^6z - 6200x^6 + 321x^4y^4 - 486x^4y^2z - 150600x^4y^2 + 166x^4z^2 - 138200x^4z + 14410000x^4 + 544x^2y^6 - 1506x^2y^4z - 777600x^2y^4 + \dots - 10752000000y^2 + 225z^4 - 765000z^3 + 722250000z^2 - 12240000000z + 576000000000) \times (9x^8 + 18x^7y - 2640x^7 + 45x^6y^2 - 7740x^6y - 27x^6z + 403400x^6 + 54x^5y^3 - 14940x^5y^2 - 72x^5yz + 1369600x^5y + 7620x^5z - 42540000x^5 + 72x^4y^4 - \dots + 251965000000yz - 26983656000000y + 9z^4 - 185900z^3 + 1331320000z^2 -$$

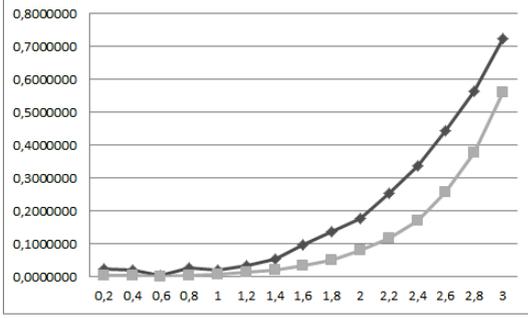


Fig. 11. Variance \widehat{C} of approximated distances d_1 (black) and d_2 (grey) along offset curves plotted against the corresponding Euclidean distances

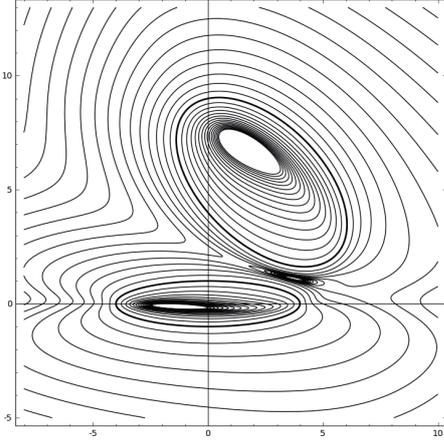


Fig. 12. Level curves $d_1 = const$ for ellipses $G_5(x, y) = 0$ (indicated bold)

$$3668093000000z + 236092320000000).$$

All the quantitative measures were computed and represented at Fig. 14, Fig. 15, Fig. 16, Table III and Table IV. The description of the obtained results in general repeats the above for the case of intersecting ellipses. We just emphasize slight difference among the variance values inside and outside the initial curve.

TABLE III. NORMALIZED ASSESSMENT RESULTS \widehat{L} , \widehat{L}^+ AND \widehat{L}^- FOR SEPARATED ELLIPSES

	\widehat{L}	\widehat{L}^+	\widehat{L}^-
d_1	1.0000000	1.0000000	1.0000000
d_2	0.9993255	0.9994992	0.9993132

TABLE IV. NORMALIZED ASSESSMENT RESULTS \widehat{A} , \widehat{C} , \widehat{C}^+ AND \widehat{C}^- FOR SEPARATED ELLIPSES

	\widehat{A}	\widehat{C}	\widehat{C}^+	\widehat{C}^-
d_1	1.0000000	1.0000000	1.0000000	1.0000000
d_2	0.1895881	0.3343521	0.4503979	0.3347797

C. Overview of results

Based on the visualization results, we outline the fact that in the both cases of multiple ellipses the features of level curves $d_j = const$ do not change in comparison with a single ellipse

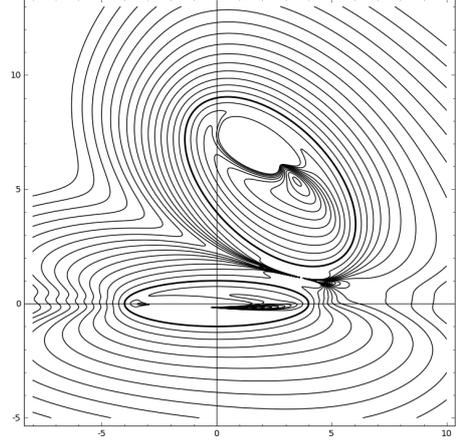


Fig. 13. Level curves $d_2 = const$ for ellipses $G_5(x, y) = 0$ (indicated bold)

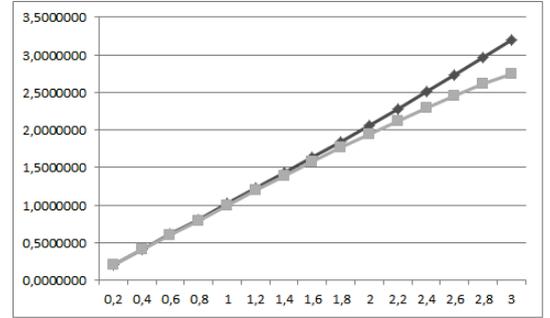


Fig. 14. Mean distances μ_i of approximated distances d_1 (black) and d_2 (grey) along offset curves plotted against the corresponding Euclidean distances

case. This may indicate that increasing number of treated ellipses will not lead to appearance of huge distortions as in the case of curve $G_2(x, y) = 0$.

The Sampson's distance approximation linearity assessment twice was better, but the difference was not huge. We can consider both formulas (2) and (3) as linear with respect to the Euclidean distance.

In general, new point-to-algebraic-manifold distance approximation have good properties close to ellipses and may be used in the curve fitting problems including multiple ellipse fitting.

V. CONCLUSION

We discuss here the problem of evaluating the validity of a new point-to-algebraic-manifold distance approximation formula. We represent the qualitative and quantitative comparison of the new point-to-algebraic-manifold distance approximation formula (3) with the well-known and widely applicable Sampson's distance formula (2). We concentrate to the potential benefits of application of the formula (3) in the implicit algebraic curves fitting problem and in the multiple-ellipse fitting. Comparison is performed not only for a single ellipse, which is usually done in other studies, but also for more complex curves - a pair of intersecting and non-intersecting

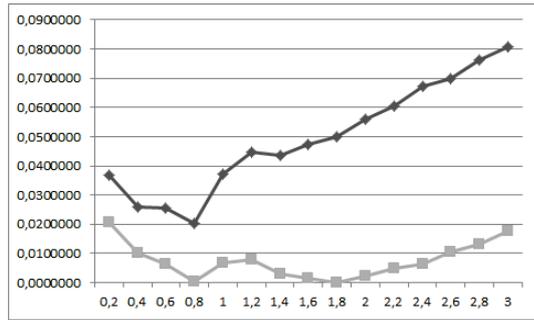


Fig. 15. Asymmetry \hat{A} of approximated distances d_1 (black) and d_2 (grey) along offset curves plotted against the corresponding Euclidean distances

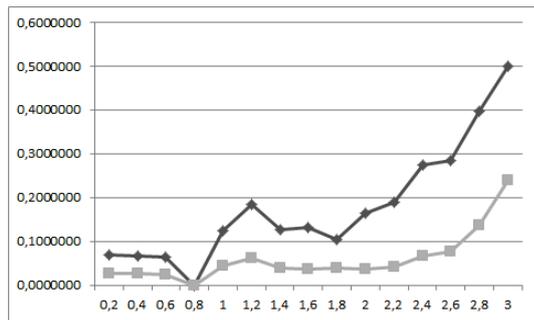


Fig. 16. Variance \hat{C} of approximated distances d_1 (black) and d_2 (grey) along offset curves plotted against the corresponding Euclidean distances

ellipses and several higher-order algebraic curves with different properties.

A qualitative comparison of approximate distance formulas (2) and (3) is performed by constructing level curves $d_i = const$ for all the treated curves. The results are shown in Fig. 1 – 8, 12, 13. For distant from the initial curves level contours we detected that formula (2) have larger deviations outside, and formula (3) inside the initial curves.

A quantitative comparison of formulas was made for pairs of ellipses according to the important practical measures - linearity and curvature bias (both calculated inside and outside the treated curve, and overall), asymmetry. This is carried out via the generation of the data set of test points with the known distance values. The calculated values of measures are shown in Tables I – IV. Fig. 9 – 11, 14 – 16 demonstrates the dependence of the studied measures on the corresponding Euclidean distance. The results of the comparison demonstrate that formula (3) exceeds formula (2) in a number of criteria.

For further research remains the inversion of this approach, i.e. the best manifold fitting problem for the given data set.

ACKNOWLEDGMENT

This research was supported by the RFBR according to the projects No **17-29-04288** (Alexei Uteshev) and No **18-31-00413** (Marina Goncharova).

REFERENCES

- [1] T.Hastie and W. Stuetzle, "Principal Curves", *J. of the American Statistical Association*, vol. 84, no. 406, 1989, pp. 502-516.
- [2] M. Aigner and B. Juttler, "Robust computation of foot points on implicitly defined curves", *Mathematical Methods for Curves and Surfaces 2004*, 2005, pp. 1-10.
- [3] M. Hu, Y. Zhou and X.Li, "Robust and accurate computation of geometric distance for Lipschitz continuous implicit curves", *Vis. Comput.*, vol. 33, 2017, pp. 937-947.
- [4] C. Lu, S. Xia, M. Shao and Y.Fu, "Arc-support line segments revisited: an efficient high-quality ellipse detection", *IEEE Trans. on Image Process.*, vol. 29, 2020, pp. 768-781.
- [5] P.L. Rosin, "Analysing error of fit functions for ellipses", *Pattern Recognition Lett.*, vol. 17, 1996, pp. 1461-1470.
- [6] P.L. Rosin, "Assessing error of fit functions for ellipses", *Graphical Models Image Process.*, vol. 58, 1996, pp. 494-502.
- [7] P.L. Rosin, "Evaluating Harker and O'Leary's distance approximation for ellipse fitting", *Pattern Recognition Lett.*, vol. 28, 2007, pp. 1804-1807.
- [8] P. Sampson, "Fitting conic sections to very scattered data: an iterative refinement of the Bookstein algorithm", *Comput. Gr. Image Process.*, vol. 18, 1982, pp. 97-108.
- [9] G. Taubin, "Estimation of planar curves, surfaces, and nonplanar space curves defined by implicit equations with applications to edge and range image segmentation", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 11, 2002, pp. 1115-1138.
- [10] G. Taubin, "Distance approximations for rasterizing implicit curves", *ACM Trans. Graph.*, vol. 13, no. 1, 1994, pp. 3-42.
- [11] G. Taubin, F. Cukierman, S. Sullivan, J. Ponce, and D. Kreigman, "Parameterized families of polynomials for bounded algebraic curve and surface fitting", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, 1994, pp. 287-303.
- [12] A. Uteshev and M. Goncharova, "Point-to-ellipse and point-to-ellipsoid distance equation analysis", *J.Comput. Appl. Math.*, vol. 328, 2018, pp. 232-251.
- [13] A. Uteshev and M. Goncharova, "Approximation of the distance from a point to an algebraic manifold", in *Proc. of the 8th International Conf. on Pattern Recognition Applications and Methods – Vol. 1: ICPRAM*, 2019, pp. 715-720.
- [14] Q. Zhang, C.H. Jin, H.T. Xu, L.Y. Zhang, X.B. Ren et al., "Multiple-ellipse fitting method to precisely measure the positions of atomic columns in a transmission electron microscope image", *Micron*, vol. 113, 2018, pp. 99-104.
- [15] C. Panagiotakis and A. Argyros, "Region-based Fitting of Overlapping Ellipses and its application to cells segmentation", *Image and Vision Computing*, vol. 93, 2020, 103810.
- [16] Y. Wu, H. Wang, F. Tang and Z. Wang, "Efficient conic fitting with an analytical Polar-N-Direction geometric distance", *Pattern Recognition*, vol. 90, 2019, pp. 415-423.