

First Order Geometric Distance (The Myth of Sampsonus)

Matthew Harker and Paul O’Leary
Institute for Automation
University of Leoben
8700 Leoben, Austria
`matthew.harker@stud.unileoben.ac.at`
`http://automation.unileoben.ac.at`

Abstract

There is a prevalent myth in Computer Vision that the gradient weighted algebraic distance, the so-called “Sampson Error,” is a first order approximation of the distance from a point to a curve or surface. In truth, however, it is the exact geometric distance to the first order approximation of the curve. The linguistic difference is subtle, but mathematically, the two statements are at odds. In this paper, we derive the actual first order approximation of the Mahalanobis distance to a curve, a special case of which is the geometric distance. Furthermore, we show that it too, like the Sampson error, is a weighted algebraic distance. The first order distance introduces an increase in computational effort (total “flops”), which is the inevitable cost of a better approximation; however, since it too is an explicit expression, it has the same computational complexity as the Sampson error. Numerical testing shows that the first order distance performs an order of magnitude better than the Sampson error in terms of relative error with respect to the geometric and Mahalanobis distances. Our results suggest that the first order distance is an exceptional candidate cost function for approximate maximum likelihood fitting.

1 Introduction

Estimation of geometric objects from scattered data is a fundamental task in Computer Vision. When errors in measured coordinates behave according to Gaussian normal distributions, the maximum likelihood criterion for the estimated parameters is the sum of squared geometric distances from the points to the geometric object. More generally, if the errors in the measured coordinates behave according to anisotropic Gaussian distributions, the maximum likelihood criterion is the sum of squared Mahalanobis distances. The maximum likelihood estimate, however, is often computationally expensive and impractical for many applications. Frequently, an approximate solution is the only practical approach. Perhaps the most common of these approximations is the so-called Sampson Error [5], which is a generalized form of that proposed by Sampson for conics [7]. A literature review shows that the Sampson error is the basis of a popular myth, as it is

often touted as a first order approximation to the geometric distance. The source of the myth is perhaps the casual misplacement of the mathematical term “first order approximation.” Consider the problem of computing the distance from a point (p, q) to a general planar polynomial $f(x, y) = 0$. The first order approximation of the function $f(x, y)$ about the point (p, q) is found by expanding the function in a Taylor series, and truncating the higher order terms, i.e.

$$\tilde{f}(x, y) = f(p, q) + f_x(p, q)(x - p) + f_y(p, q)(y - q) = 0, \quad (1)$$

where the subscripts denote the partial derivatives

$$f_x(x, y) \triangleq \frac{\partial f(x, y)}{\partial x} \quad \text{and} \quad f_y(x, y) \triangleq \frac{\partial f(x, y)}{\partial y}. \quad (2)$$

The approximate function, $\tilde{f}(x, y)$, is a line; however, it is *not* the tangent line, since (p, q) is generally not on the curve. The normal distance from the point (p, q) to the line $\tilde{f}(x, y)$ is the Sampson error,

$$r_S^2 = \frac{f(p, q)^2}{f_x(p, q)^2 + f_y(p, q)^2}. \quad (3)$$

Hence, by definition, the Sampson error is *the distance to the first order approximation of the function*. At first glance, it may seem that the difference is purely linguistic, but the mathematical difference is significant. In the present work, we derive the first order approximation of the geometric distance to a curve. The derivation proceeds for the more general Mahalanobis distance, since the geometric distance is merely a special case of this.

Harker and O’Leary [4] demonstrated the derivation of a polynomial directly in the Mahalanobis distance from a point to a curve. The roots of the polynomial are the extrema of the distance function from a point to the curve, whereby the smallest positive root is the minimum distance to the curve. This polynomial is an implicit function of the curve parameters to be estimated, and the distance to be minimized. That is to say, there are no intermediate variables such as the orthogonal contact points or a Lagrange multiplier, which are typically referred to as “nuisance parameters.” The absence of intermediate variables in the polynomial has opened the door for the present work: the first order approximation, which was hitherto not possible with known methods. We first show an alternative means of deriving the polynomial in the distance. We then make use of the Taylor expansion to make an approximation to the *distance*, in contrast to the Sampson error, which makes an approximation to the *function*. What is interesting about the first order distance is that it too, like the Sampson error, is a weighted algebraic distance. Hence, it has the same computational complexity as the Sampson error.

To test the new result, we compare both the first order distance and the Sampson error to the geometric distance in terms of relative error; this gives a measure of how well the two approximate the geometric distance. The relative error plots indicate that when used for geometric fitting, the new approximation will typically perform an order of magnitude better than the Sampson error as an approximation to the “exact” distance. To confirm this, we test the algorithms numerically in the task of fitting ellipses by least-squares geometric distance and Mahalanobis distance. The results show that the first order distance is accurate to within 1% of the maximum likelihood solution, whereas the Sampson error is accurate on the order of 10%.

The first order distance is a fundamentally new approach to approximate maximum likelihood fitting since its computational complexity is commensurate with the Sampson error, yet it approximates the maximum likelihood solution to a fraction of a percent.

2 The Mahalanobis and Geometric Distances

An implicit polynomial in the Mahalanobis distance from a point to a curve was derived by Harker and O'Leary [4] using the concept of discriminants. Here we present a new approach using the Macaulay resultant. Consider a point (p, q) with a measured covariance matrix,

$$\Lambda = \begin{bmatrix} \sigma_{pp} & \sigma_{pq} \\ \sigma_{pq} & \sigma_{qq} \end{bmatrix}. \quad (4)$$

The covariance matrix defines a confidence envelope about the point, which satisfies,

$$g(x, y) - \chi^2 = 0, \quad (5)$$

where χ is the Mahalanobis distance. The value of χ^2 essentially determines the “size” of the confidence envelope, which is an ellipse. The function $g(x, y)$ is defined as,

$$\begin{aligned} g(x, y) &\triangleq \lambda_{pp}x^2 + 2\lambda_{pq}xy + \lambda_{qq}y^2 + (-2p\lambda_{pp} - 2q\lambda_{pq})x + (-2p\lambda_{pq} - 2q\lambda_{qq})y \\ &\quad - (-p\lambda_{pp} - q\lambda_{pq})p - (-p\lambda_{pq} - q\lambda_{qq})q \end{aligned} \quad (6)$$

where,

$$\begin{bmatrix} \lambda_{pp} & \lambda_{pq} \\ \lambda_{pq} & \lambda_{qq} \end{bmatrix} = \Lambda^{-1}. \quad (7)$$

We consider the general planar implicit function, $f(x, y) = 0$, of degree d . To solve for the minimum Mahalanobis distance from the point (p, q) to the curve, we must consider the following three conditions:

1. There is a “nearest point” on the curve, which amounts to the constraint $f(x, y) = 0$.
2. At the nearest point the curve and confidence envelope are tangent, i.e.

$$g_x(x, y)f_y(x, y) - g_y(x, y)f_x(x, y) = 0. \quad (8)$$

3. The only equation containing the quantity χ^2 is Equation (5), hence we must include it to derive a polynomial in the Mahalanobis distance.

Correspondingly, we must solve the three simultaneous equations,

$$f(x, y) = 0 \quad (9)$$

$$g_x(x, y)f_y(x, y) - g_y(x, y)f_x(x, y) = 0 \quad (10)$$

$$g(x, y) - \chi^2 = 0. \quad (11)$$

Equations (9) and (10) are both planar curves of degree d . Their intersection points are the points on the curve which satisfy the tangent condition. One possible approach would be to solve for these contact points and substitute them into Equation (11) to yield the corresponding Mahalanobis distances; however, this method is *indirect*, and does not facilitate

a first order approximation. An alternative and more fruitful approach is to consider the fact that Equations (9) and (10) define an affine variety. Specifically, they define the tangent contact points (x, y) where the Mahalanobis distance function has extrema. Equation (11), however, defines the Mahalanobis distance, χ , in terms of a general variable point (x, y) . We know from algebraic geometry, though, that if all three equations are consistent (i.e. they have a solution) then the multi-dimensional resultant eliminating the variables x and y vanishes [2]. We denote the Macaulay resultant¹ as the operator $\text{Res}(-, -)$. The Macaulay resultant [6] eliminating x and y from Equations (9), (10), and (11), is a polynomial in χ^2 , i.e.

$$\text{Res}([f(x, y), g_x(x, y)f_y(x, y) - g_y(x, y)f_x(x, y), g(x, y) - \chi^2], [x, y]) = P(\chi^2). \quad (12)$$

The condition for which the equations are consistent is that the polynomial must vanish, i.e., we are interested in the roots, $P(\chi^2) = 0$. The polynomial $P(\chi^2)$ is of degree $N = d^2$ in χ^2 , and hence has the form,

$$P(\chi^2) = \sum_{k=0}^N \alpha_{2k} \chi^{2k} = \alpha_{2N} \chi^{2N} + \dots + \alpha_2 \chi^2 + \alpha_0 = 0. \quad (13)$$

Since the resultant has eliminated x and y , the coefficients of the polynomial $P(\chi^2)$ depend only on the curve coefficients, the point, and its covariance matrix. The roots of the polynomial are the extrema of the Mahalanobis distance from the point to the curve.

3 First Order Mahalanobis Distance

The minimum Mahalanobis distance from a point to a curve is the smallest positive root of the polynomial $P(\chi^2)$. In this section, we consider a means of approximating this root. The Newton-Raphson iteration [2] for polynomial root finding is a recurrence relation giving an estimate χ_{k+1}^2 based on the previous estimate χ_k^2 , i.e.

$$\chi_{k+1}^2 = \chi_k^2 - \frac{P(\chi_k^2)}{\left. \frac{dP(\chi^2)}{d\chi^2} \right|_{\chi^2 = \chi_k^2}}. \quad (14)$$

In estimation applications, points are typically on or near the curve in question, therefore, zero may be considered a good initial estimate for the minimal distance. The first Newton-Raphson step is then,

$$\chi_F^2 = - \frac{P(0)}{\left. \frac{dP(\chi^2)}{d\chi^2} \right|_{\chi=0}}. \quad (15)$$

Evaluating the expression, we have $P(0) = \alpha_0$ and $\left. \frac{dP(\chi^2)}{d\chi^2} \right|_{\chi=0} = \alpha_2$, yielding,

$$\chi_F^2 = - \frac{\alpha_0}{\alpha_2}, \quad (16)$$

¹More formally, the Macaulay resultant, $\text{Res}([f_1, \dots, f_n], [x_1, \dots, x_{n-1}])$, eliminates the set of $(n-1)$ variables $[x_1, \dots, x_{n-1}]$, from the set of n functions $[f_1, \dots, f_n]$. This is accomplished by augmenting the set of equations by monomials in $[x_1, \dots, x_{n-1}]$, taking the determinant of the coefficient matrix, and dividing out a redundant factor. The resultant is analogous to the condition $\det(A) = 0$ for the system of equations $Ax = 0$ to have a non-trivial solution.

which is the first order approximation of the Mahalanobis distance to a curve. Of note, is the fact that only the polynomial coefficients α_0 and α_2 are required. This is a significant reduction in computational complexity over the full Mahalanobis distance. To solve the full problem, a total of $d^2 + 1$ coefficients must be computed, followed by computing all d^2 roots of the polynomial. For example, for a cubic curve $d = 3$, and the polynomial has ten coefficients and nine roots.

4 First Order Geometric Distance for Conics

A general conic section is described by the implicit equation,

$$f(x, y) = a_1x^2 + a_2xy + a_3y^2 + a_4x + a_5y + a_6 = 0. \quad (17)$$

For the geometric distance, the confidence envelope becomes a circle, and the Mahalanobis distance is simply the radius r ,

$$g(x, y) - \chi^2 = (x - p)^2 + (y - q)^2 - r^2 = 0. \quad (18)$$

The tangent condition becomes,

$$(2x - 2p)(a_2x + 2a_3y + a_5) - (2y - 2q)(2a_1x + a_2y + a_4) = 0. \quad (19)$$

Eliminating x and y from Equations (17), (18), and (19) using the Macaulay resultant yields a fourth order polynomial in r^2 ,

$$\alpha_8 r^8 + \alpha_6 r^6 + \alpha_4 r^4 + \alpha_2 r^2 + \alpha_0 = 0. \quad (20)$$

The first order approximation of the geometric distance to a conic is therefore,

$$r_{\text{F}}^2 = -\frac{\alpha_0}{\alpha_2}. \quad (21)$$

At this point we come to a note on implementation: in their general form, the coefficients α_0 and α_2 can be rather large. If, however, we translate the conic such that the point (p, q) is the origin, then the size of α_0 and α_2 and the number of operations to compute them are significantly reduced. The first order distance then has the form,

$$r_{\text{F}}^2 = -\frac{w_{\text{n}}}{w_{\text{d}}} a_6^2 \quad (22)$$

with,

$$\begin{aligned} w_{\text{n}} = & 8a_4^2 a_3 a_6 + 16a_3^2 a_6^2 - 8a_3 a_6 a_5^2 + a_5^4 + 16a_2^2 a_6^2 + a_4^4 + 2a_4^2 a_5^2 \\ & - 32a_1 a_3 a_6^2 + 8a_1 a_5^2 a_6 - 16a_2 a_6 a_4 a_5 - 8a_1 a_6 a_4^2 + 16a_1^2 a_6^2 \end{aligned} \quad (23)$$

and

$$\begin{aligned} w_{\text{d}} = & 32a_1 a_2^2 a_6^3 + 32a_3 a_2^2 a_6^3 - 32a_1^2 a_3 a_6^3 - a_4^6 - 32a_1 a_3^2 a_6^3 + 10a_1 a_6 a_4^4 \\ & - 8a_1 a_5^4 a_6 - 8a_1^2 a_6^2 a_5^2 + 2a_1 a_4^2 a_5^2 a_6 + 18a_2 a_5^3 a_4 a_6 + 18a_2 a_6 a_4^3 a_5 \\ & + 40a_1 a_3 a_6^2 a_5^2 + 2a_3 a_4^2 a_5^2 a_6 - 32a_1^2 a_6^2 a_4^2 - 8a_4^4 a_3 a_6 - 20a_2^2 a_5^2 a_6^2 \\ & + 10a_3 a_6 a_5^4 - 20a_2^2 a_4^2 a_6^2 - 32a_3^2 a_6^2 a_5^2 - 8a_3^2 a_6^2 a_4^2 - 24a_1 a_2 a_6^2 a_4 a_5 \\ & - 24a_3 a_2 a_6^2 a_4 a_5 + 40a_1 a_4^2 a_3 a_6^2 + 32a_1^3 a_6^3 + 32a_3^3 a_6^3 - 3a_4^2 a_5^4 \\ & - 3a_4^4 a_5^2 - a_5^6. \end{aligned} \quad (24)$$

For clarity, the weighting factors have been left as is, however, to minimize the computational effort, they should be put into Horner form and coded such that the common terms are computed only once. With the aforementioned simplifications, the first order distance can be computed from the *general* coefficients and the point coordinates with 54 additions, 76 multiplications and 1 division (131 flops).

An interesting note, is that in the translated problem a_6 is the algebraic distance from the point to the conic, hence, the first order distance is nothing more than a weighted algebraic distance. Also of note is that w_n , w_d , and a_6 are invariant to Euclidean transformations of the data.

Figure 1 shows the polynomial in the geometric distance from a given point to a given conic. For purpose of comparison, the right-hand plot shows a magnification of the polynomial near the smallest root and the first order approximation to the polynomial at $r^2 = 0$. Both the minimum root and its first order approximation are identified.

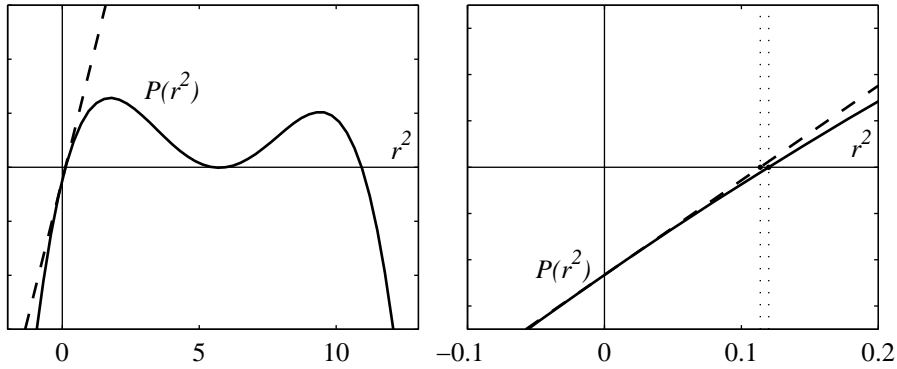


Figure 1: (LEFT) The polynomial (—) of geometric distances from a given point to a given conic and its first order approximation at $r^2 = 0$, the line (---). (RIGHT) Magnification of the polynomial near the smallest root. The smallest root and its first order approximation are marked with the vertical lines (· · ·).

5 Approximation Error

To demonstrate the proposed approximation, we define the relative error of the first order distance, r_F , with respect to the exact geometric distance, r_G , and the same quantity for the Sampson error, r_S , i.e.

$$\varepsilon_F = 100 \times \frac{|r_F - r_G|}{r_G} \% \quad \text{and} \quad \varepsilon_S = 100 \times \frac{|r_S - r_G|}{r_G} \% . \quad (25)$$

Figure 2 shows plots of the relative errors ε_F and ε_S as functions of the points in the plane and specific conic sections.

For hypothetical scattered data near the curves, the first order distance is accurate to 1% relative error, whereas the Sampson error is accurate to 10%. Note that the only time the Sampson error performs well is for the asymptotes of the hyperbola, where the function is almost linear.

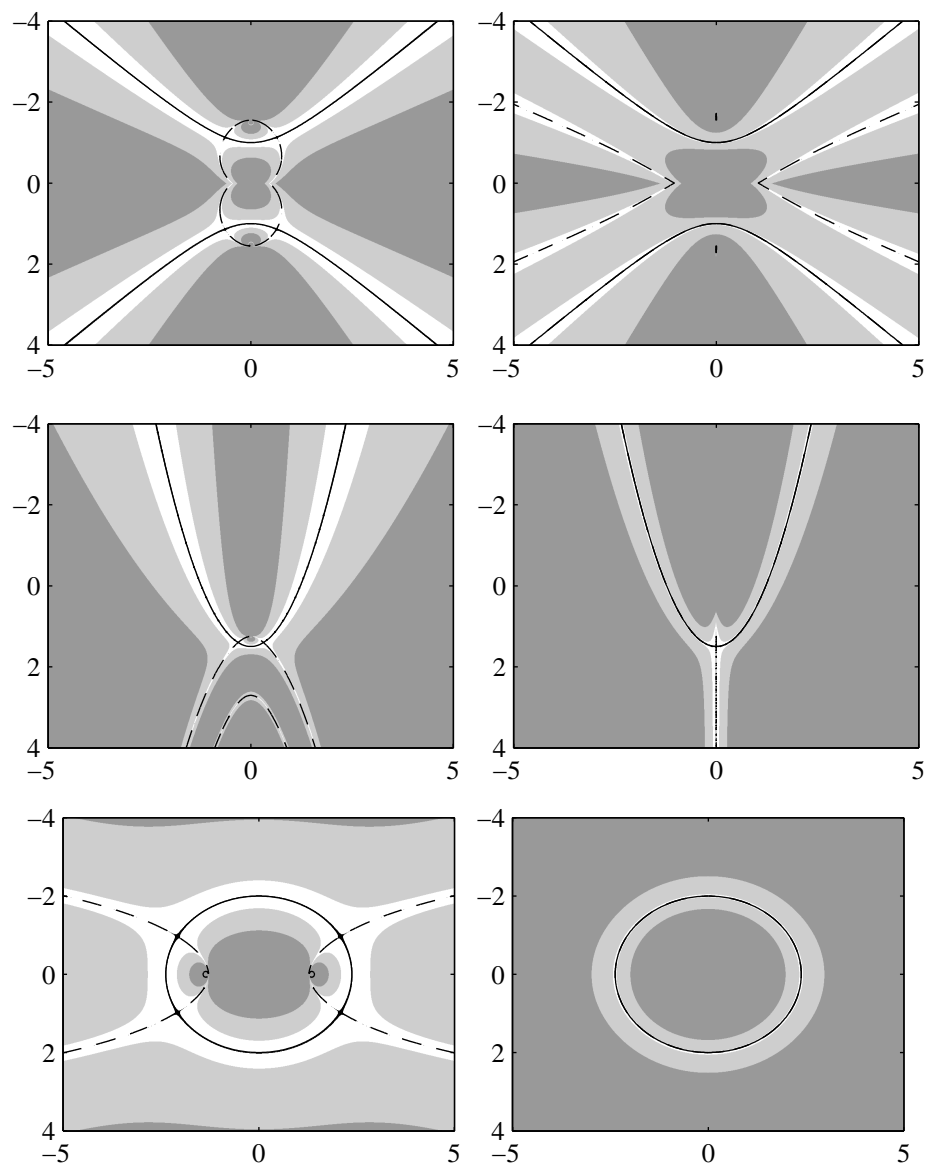


Figure 2: Relative error with respect to the actual geometric distance of (LEFT) the first order distance ϵ_F , and (RIGHT) the Sampson error ϵ_S . Row-wise are a hyperbola, a parabola and an ellipse. Colour coding of the relative error is by logarithmic scale: white \rightarrow (0 – 1%), light-grey \rightarrow (1 – 10%), dark-grey \rightarrow (> 10%). Dashed lines (---) are the additional loci where the relative error is zero.

6 Numerical Testing

To demonstrate the first order distance, we have used the common estimation task of ellipse fitting, both by geometric distance as well as by Mahalanobis distance. For comparison purposes, we consider the following cost functions:

- GD/MD: The full maximum likelihood functional for geometric distance (GD) and Mahalanobis distance (MD).
- FOGD/FOMD: The first order geometric distance (FOGD) and Mahalanobis distance (FOMD).
- SE/SEM: The Sampson error (SE) and the Sampson error for the Mahalanobis distance (SEM).

To minimize the cost functions, we used a simple Gauss-Newton minimization with the Jacobians computed numerically. In this vein, all initial conditions, iteration parameters and stopping conditions were identical for all cost functions. The data are standard “worst case” data sets for testing non-linear fitting algorithms, which can be found in [1]. For the Mahalanobis fitting we generated covariance matrices from a small number of points from a random isotropic distribution, which mimics a real measurement.

Note that for the maximum likelihood cost functions (GD and MD), we have used the full polynomials derived in Section 2 for the functional evaluations. Since they avoid computing the nuisance parameters, our solution for MD is much simpler than any found to date in the literature (cf. [5]).

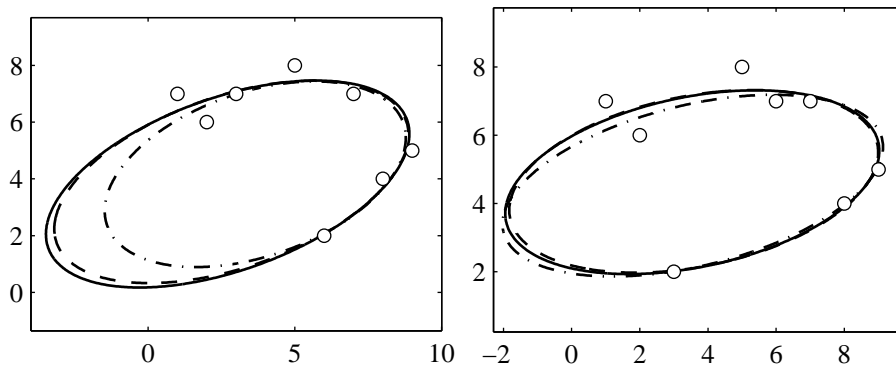


Figure 3: Ellipse fitting by geometric distance: Results for the cost functions GD (—), FOGD (---), SE (-·-), to Data A (LEFT) and Data B (RIGHT).

Table 1: Non-Linear Ellipse Fitting: Geometric Distance

Metric	Complexity	Data A		Data B	
		$\sum r_G^2$	$\epsilon_{rel.}$	$\sum r_G^2$	$\epsilon_{rel.}$
GD	Implicit	1.3735	0.00 %	1.4519	0.00 %
FOGD	Explicit	1.3740	0.04 %	1.4548	0.20 %
SE	Explicit	1.4840	8.05 %	1.6389	12.88 %

Table 2: Non-Linear Ellipse Fitting: Mahalanobis Distance

Metric	Complexity	Data A		Data B	
		$\sum \chi_M^2$	ϵ_{rel}	$\sum \chi_M^2$	ϵ_{rel}
MD	Implicit	2.1881	0.00 %	1.9675	0.00 %
FOMD	Explicit	2.1925	0.20 %	1.9717	0.21 %
SEM	Explicit	2.2512	2.88 %	2.0465	4.02 %

With respect to the maximum likelihood cost functions, the first order distance is accurate up to three significant digits (i.e. when rounded to three digits, they are identical). In fact, the relative error is much less than 1%, and almost negligible. The relative error for the Sampson error, however, is on the order of 10% with respect to the maximum likelihood cost function. These results are in agreement with the relative error plots presented in Section 5. That is to say, the first order distance remains a good approximation to the exact distance even when the points are further from the curve. This however, cannot be said for the Sampson error. In sum, the results show that the first order distance is more robust as an approximate maximum likelihood estimator than the Sampson error.

7 Cubics

Figure 4 shows relative error plots for a cubic curve. This shows how the Sampson error generally degrades with the higher curvature as compared to the conics. The performance of the first order distance, however, remains more consistent in spite of the higher curve degree.

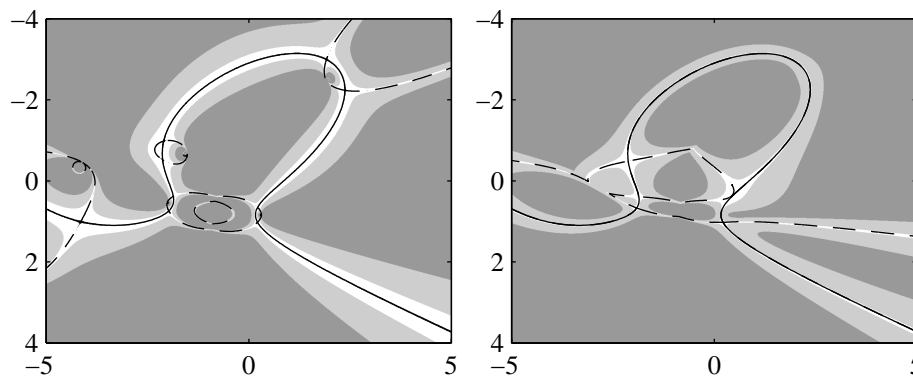


Figure 4: Relative error with respect to the actual geometric distance to a cubic curve for the (LEFT) the first order distance, and (RIGHT) the Sampson error. See Figure 2 for the legend.

For higher order curves such as the cubics, a computationally more practical approach to curve fitting would be to use re-weighted least squares, as per Sampson [7]. Since the first order distance is also a weighted algebraic distance, this method would have the advantage of simplicity. Although this method often works well, the disadvantage is that

the re-weighted least squares problem does not minimize the desired cost function [3]. This is a topic of future research.

8 Conclusions

In this paper, we derived the first order approximation to the Mahalanobis distance from a point to a curve. We showed that it is quite different to the Sampson error, and in so doing, we hope to dispel a myth that is all too common to the literature. Through numerical testing, we showed that the first order distance is commensurate with the Sampson error in terms of computational complexity, yet approaches the exact distance in terms of accuracy. The first order distance is therefore an excellent candidate cost function for approximate maximum likelihood fitting. The authors have used a similar approach to that presented here to derive the first order geometric distance for surfaces in space, as well as a slight variation to derive the first order distance for two-view geometries. The approach hence applies to a vast array of approximate maximum likelihood estimation problems commonly encountered in Computer Vision. Future research will involve making higher order approximations to the polynomial roots, an analysis of convergence and stability of the polynomial roots, as well as using the first order distance for re-weighted least-squares estimation. A further application of the first order distance is in robust estimation, since it provides a fast and accurate estimate of how well a data point fits an instance of a model.

References

- [1] S.J. Ahn, W. Rauh, and H.-J. Warnecke. Least-squares orthogonal distances fitting of circle, sphere, ellipse, hyperbola, and parabola. *Pattern Recognition*, 34:2283–2303, 2001.
- [2] D. Cox, J. Little, and D. O’Shea. *Using Algebraic Geometry*. Springer-Verlag, New York, Second edition, 2005.
- [3] W. Gander, G. Golub, and R. Strebler. Fitting of circles and ellipses least squares solution. Technical Report doc/tech-reports/1994/217.ps, ETH Zurich, Department Informatik, June 1994.
- [4] M. Harker and P. O’Leary. Using discriminants to derive direct polynomials in the Mahalanobis and Euclidean distances. *Submitted to IEEE Trans. Pattern Analysis and Machine Intelligence*, 2006.
- [5] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, Second edition, 2003.
- [6] M. Minimair. MR: Macaulay Resultant Package for Maple, <http://minimair.org/mr>, 2005.
- [7] P.D. Sampson. Fitting conic sections to ‘very scattered’ data: An iterative refinement of the Bookstein algorithm. *Computer Vision, Graphics, and Image Processing*, 18:97–108, 1982.