A Study of the Optimality of Approximate Maximum Likelihood Estimation

David N. R. M^{c}Kinnon and Brian C. Lovell
{mckinnon,lovell}@itee.uq.edu.au
Intelligent Real-Time Imaging and Sensing (IRIS) Group, EMI
School of Information Technology and Electrical Engineering
University of Queensland
St. Lucia QLD 4072, Australia

Abstract

Maximum Likelihood Estimation (MLE) is widely utilized in the computer vision literature as a means of solving parameter estimation problems assuming a Gaussian noise model for the measurement data. In order to solve a MLE problem it is necessary to have knowledge of the true parameters of the Gaussian noise model. Since this knowledge is unobtainable in practical setting approximate MLE has become a popular alternative. The theory behind the approximate MLE framework is presented and an analysis of the bias characteristics of the method for noisy data is performed. Several experiments are performed to ascertain the optimality of approximate MLE solutions and to determine whether or not there is a correlation between the degree and dimension of the algebraic hypersurface and optimality of the error metric.

1. Introduction

Parameter estimation is of central importance to a wide range of problems in computer vision such as line fitting, conic fitting and multiview constraint estimation. Parameter estimation is applicable in any situation where we wish to derive an unknown set of parameters from noisy measurement data by utilizing a functional relationship between the measurements (observations) and the parameters.

In this paper we will develop the basic theory underlying parameter estimation with the assumption that the measurement data is corrupted by Gaussian noise. Gaussian parameter estimation has received much attention in the computer vision literature due to favourable properties of the Maximum Likelihood Estimation (MLE) framework utilizing a Gaussian noise model for the measurement data, a selective chronology of the literature in this area can be found in [9, 10, 7].

We focus our attention on the approximate MLE framework utilizing a Gaussian noise model for the measurements. We perform a series of experiments on different estimation problems to determine the efficacy of this framework in determining approximations to the true values of the measurements (nuisance parameters) and more specifically how the accuracy of these approximations varies as the level Gaussian noise applied to measurements increases in addition to the degree and dimension of the hypersurface.

This information is of great importance to the implementor of parameter estimation software since the objective function minimized for such problems requires apriori knowledge of the true estimates of the nuisance parameters and consequentially the unknown parameters themselves.

2 Parameter Estimation

Parameter estimation is the process of calculating a set of variables (parameters) associated with a mathematical model, given a set of noisy measurements related to the model. As a form of convention we will denote the measurement data as a vector $\mathbf{x} \in \mathbb{R}^m$ and the parameters as a vector $\theta \in \mathbb{R}^n$. In our discussion we make the distinction between measured values, approximated values and true values of the measurements and the parameters, for this purpose we will use the notation $\mathbf{x}/\theta$, $\hat{\mathbf{x}}/\hat{\theta}$ and $\bar{\mathbf{x}}/\bar{\theta}$ respectively.

Assuming a standard measurement model for our data we have $\mathbf{x} = \bar{\mathbf{x}} + \mathcal{G}(\bar{\mu}_{\mathbf{x}}, \bar{\sigma}_{\mathbf{x}}^2 \bar{\Sigma}_{\mathbf{x}})$, where $\bar{\mathbf{x}}$ is the true value of the measurement and $\mathcal{G}(\bar{\mu}_{\mathbf{x}}, \bar{\sigma}_{\mathbf{x}}^2 \bar{\Sigma}_{\mathbf{x}})$ is an independently distributed Gaussian probability distribution function (pdf) with mean $\bar{\mu}_{\mathbf{x}}$, standard deviation $\bar{\sigma}_{\mathbf{x}}$ and covariance $\bar{\Sigma}_{\mathbf{x}}$.

2.1 The Functional & Bilinear Models

In this section we develop two different models for a parameter estimation problem. The models are referred to as the functional and bilinear parameter estimation frameworks and they cater for two distinct problem types. Both of these frameworks utilize noisy measurements ($\mathbf{x}$) (and possibly other known data) to determine a solution to a set of parameters (θ). Of interest in some situations is the calculation of the so called *nuisance* parameters, these are defined as the approximate values of the noisy measurement data (ie. $\hat{\mathbf{x}}$).

In practise we only have access to noisy measurement data ($\mathbf{x}$) from which we wish to approximate the true value of the parameters ($\bar{\theta}$). This problem is ill-posed since there is no means of determining the exact nature of the true noise model ($\mathcal{G}(\bar{\mu}_{\mathbf{x}}, \bar{\sigma}^2_{\mathbf{x}}\bar{\Sigma}_{\mathbf{x}})$) affecting the measurement data. Instead we can only approximate the noise model ($\mathcal{G}(\hat{\mu}_{\mathbf{x}}, \hat{\sigma}^2_{\mathbf{x}}\hat{\Sigma}_{\mathbf{x}})$) resulting in the eventual estimate of the parameters being only an approximate solution ($\hat{\theta}$).

The functional model for parameter estimation utilizes a mapping between the parameters (θ) and the measurements ($\mathbf{x}$).

$$\mathbf{x} = f(\theta) \tag{1}$$

We can view the relationship (1) as the basis for a least-squares estimation problem (either linear or non-linear) and define the following fundamental relationship between the noisy measurements and the approximate parameter values,

$$\mathbf{x} = f(\hat{\theta}) + \epsilon \tag{2}$$

where $\epsilon = \mathbf{x} - \hat{\mathbf{x}} = \mathcal{G}(\hat{\mu}_{\mathbf{x}}, \hat{\sigma}^2_{\mathbf{x}}\hat{\Sigma}_{\mathbf{x}})$ is the approximation to the additive noise obtained by utilizing the mapping (1). If the mapping (1) is linear then we can substitute $f(\hat{\theta})$ for $\mathbf{A}\hat{\theta}$, where $\mathbf{A}$ is a constraint matrix resulting in.

$$\mathbf{x} = \mathbf{A}\hat{\theta} + \epsilon \tag{3}$$

The other model that we will consider is the bilinear model for parameter estimation. This assumes that their exists a mapping $f(\mathbf{x})$ such that it is possible to form an equation linear in the coefficients of the parameters,

$$\epsilon = f(\mathbf{x})\hat{\theta} \tag{4}$$

in this case $\epsilon \equiv \mathbf{x} - \hat{\mathbf{x}} = \mathcal{G}(\hat{\mu}_{\mathbf{x}}, \hat{\sigma}^2_{\mathbf{x}}\hat{\Sigma}_{\mathbf{x}})$. It is not as obvious how we justify the same derivation of the noise model for this problem type however we will show in later sections that the nuisance parameter ($\hat{\mathbf{x}}$) can be determined in a non-specific fashion satisfactorily. The bilinear model can also be expressed as a linear mapping $\mathbf{A} \equiv f(\mathbf{x})$ resulting in an analogous linear form,

$$\epsilon = \mathbf{A}\hat{\theta} \tag{5}$$

the constraint matrix ($\mathbf{A}$) in this case is a linear function of the measurements.

We can generalize the two frameworks (2) and (4) in most instances by simply utilizing the objective function (which is a pdf) since ϵ retains the same definition,

$$\mathcal{R}(\mathbf{x}, \hat{\theta}) = \epsilon \tag{6}$$

this represents the relationship between the noisy measurements and the approximate of the parameters with the noise model. The solution to (6) corresponds with the parameter vector ($\hat{\theta}$) resulting in $\frac{\partial \mathcal{R}(\mathbf{x},\hat{\theta})}{\partial \hat{\theta}} = 0$ and $\frac{\partial^2 \mathcal{R}(\mathbf{x},\hat{\theta})}{\partial^2 \hat{\theta}} > 0$. A particular approach to parameter estimation is said to be asymptotically unbiased iff. $\lim_{m\to\infty} \mathrm{E}[\hat{\theta}] = \bar{\theta}$. An approach is said to be consistent iff. $\lim_{m\to\infty} \mathrm{E}[\mathcal{R}(\mathbf{x}, \hat{\theta})] = 0$ and efficient iff. $\mathrm{VAR}[\hat{\theta}] \geq \frac{\mathcal{F}^+}{m}$ where $\mathcal{F}$ is the Fisher information matrix [7].

2.2 MLE for Gaussian Distributions

Maximum Likelihood Estimation (MLE) is a particular approach to parameter estimation. The goal of MLE is to increase the likelihood that the estimate of the parameters ($\hat{\theta}$) is correct assuming the relationship (6) between the parameters and measurements. The objective function for MLE is determined as the log of the objective function (6),

$$\mathcal{R}_{\mathrm{ML}}(\mathbf{x}, \hat{\theta}) = \log \mathcal{R}(\mathbf{x}, \hat{\theta}) \tag{7}$$

when dealing with exponentially defined noise models (such as a Gaussian distribution), it is much easier to maximize (7) than it is to minimize (6) due to simplification of the pdfs by the logarithm. MLE has the properties of being invariant to reparameterization, asymptotically unbiased, consistent and asymptotically efficient in the context stated above. However, a MLE solution can be heavily biased when the number of measurements (m) is small.

The pdf of (6) simplifies very conveniently when using MLE with a Gaussian noise model to the following objective function.

$$\mathcal{R}_{\mathrm{ML}}(\mathbf{x}, \hat{\theta}) \equiv \frac{1}{2}\sum_{i=1}^{m} (\mathbf{x}_i - \hat{\mathbf{x}}_i)^\top \hat{\Sigma}^+_{\mathbf{x}_i} (\mathbf{x}_i - \hat{\mathbf{x}}_i) \tag{8}$$

This expression for the objective function is equivalent to the square of the Mahalanobis distance of ϵ assuming a covariance matrix $\hat{\Sigma}_{\mathbf{x}_i}$ ($\|\epsilon\|^2_{\hat{\Sigma}_{\mathbf{x}}}$), in practise this is simple to compute.

3 Approximate MLE for Gaussian Distributions

MLE schemes seek to find the value of $\hat{\theta}$ that maximizes the pdf (7), which is equivalent to finding the value of $\hat{\theta}$ that minimizes the Mahalanobis distance (8),

$$\min_{\hat{\theta}} \|\epsilon\|^2_{\hat{\Sigma}_{\mathbf{x}}} \equiv \max_{\hat{\theta}} \mathcal{R}_{\mathrm{ML}}(\mathbf{x}, \hat{\theta}) \tag{9}$$

with the constraint that $\hat{\theta}$ must lie orthogonal to the null space of the least-squares constraint. We have already noted that MLE in a practical setting is intractable due to a lack of knowledge of the true noise distribution. We can however develop an approximation to the MLE residual ($\mathcal{R}_{\mathrm{AML}}(\mathbf{x}, \hat{\theta})$) allowing us to make affective use of the underlying principles.

3.1 Approximate MLE Residual Function

Returning to the fundamental statements of the MLE framework we can write the residual (8) of (7) as a Taylor series expansion to give us an alternative representation.

$$\mathcal{R}_{AML}(\mathbf{x}+\Delta\mathbf{x}, \hat{\theta}) \equiv \mathcal{R}_{ML}(\mathbf{x}, \hat{\theta}) + \frac{\delta \mathcal{R}_{ML}(\mathbf{x}, \hat{\theta})}{\delta \mathbf{x}} \Delta\mathbf{x} + \cdots + \frac{\delta^n \mathcal{R}_{ML}(\mathbf{x}, \hat{\theta})}{n! \delta \mathbf{x}^n} \Delta\mathbf{x}^n + \mathrm{R}_n \tag{10}$$

Where $\Delta\mathbf{x} = \hat{\mathbf{x}} - \mathbf{x}$ and $n+1$ is the number of times that the function $\mathcal{R}_{AML}(\mathbf{x}, \hat{\theta})$ is continuously differentiable. Also R_n is the remainder term which will converge to zero as n approaches infinity. From this point we can proceed by developing a residual function for approximate MLE. We start by rewriting (10) with just the first two terms of the RHS. This has the effect of making a first-order approximation to the proper MLE.

$$\mathcal{R}_{AML}(\mathbf{x}+\Delta\mathbf{x}, \hat{\theta}) \approx \mathcal{R}_{ML}(\mathbf{x}, \hat{\theta}) + \frac{\delta \mathcal{R}_{ML}(\mathbf{x}, \hat{\theta})}{\delta \mathbf{x}} \Delta\mathbf{x} \tag{11}$$

Making the substitution $\mathcal{R}_{ML}(\mathbf{x}, \hat{\theta}) = \epsilon$ and identifying $\mathbf{J}^{\epsilon}_{\mathbf{x}} = \frac{\delta \epsilon}{\delta \mathbf{x}}$ as the Jacobian of the residual function with respect to the measurements we have.

$$\mathbf{J}^{\epsilon}_{\mathbf{x}} \Delta\mathbf{x} = -\epsilon \tag{12}$$

We wish to solve for $\Delta\mathbf{x}$ subject to the equation above, the standard method to solve problems of this type is Lagrange multipliers [6]. After an application of Lagrange multipliers we find that the first-order approximation of $\Delta\mathbf{x}$ is,

$$\Delta\mathbf{x} = \hat{\mathbf{x}} - \mathbf{x} \approx -\hat{\Sigma}_{\mathbf{x}} \mathbf{J}^{\epsilon\top}_{\mathbf{x}} \left(\mathbf{J}^{\epsilon}_{\mathbf{x}} \hat{\Sigma}_{\mathbf{x}} \mathbf{J}^{\epsilon\top}_{\mathbf{x}}\right)^{+} \epsilon \tag{13}$$

making this equation negative and applying the Mahalanobis distance we find,

$$\mathcal{R}_{AML}(\mathbf{x}, \hat{\theta}) \equiv \|\mathbf{x} - \hat{\mathbf{x}}\|^2_{\hat{\Sigma}_{\mathbf{x}}} \approx \epsilon^{\top} (\mathbf{J}^{\epsilon}_{\mathbf{x}} \hat{\Sigma}_{\mathbf{x}} \mathbf{J}^{\epsilon\top}_{\mathbf{x}})^{+} \epsilon \tag{14}$$

which is the approximate to the proper MLE residual function (8).

4 Experiments with Error Metrics

Error metrics allow us to determine the approximate distance between hyperplanes and embedded features, as well as providing approximate corrections to a hyperplane position that is not coincident with an embedded feature. In this section we present the formulae for the error metrics corresponding to curves in $\mathbb{P}^2$ and $\mathbb{P}^3$ and surfaces in $\mathbb{P}^3$, these are all instances of approximate MLE [9, 6].

Of greatest interest is the performance of the approximate MLE framework in determining the error metrics and the associated corrections in situations involving high levels of noise and configurations that involve singular points on the feature. This information will be useful in assessing the efficacy of approximate MLE for practical purposes where we desire the error metrics to perform gracefully in the presence of large amounts of error and singular points, a similar analysis is performed in [10]. The analysis in this case differs since we wish to quantify the results through many random trials using embedded features of varying degree and dimension to establish whether or not these variables play a role in the optimality of the ensuing estimates.

4.1 Nuisance Parameters and Error Metrics

In [8] the embedded hypersurface representation for curves and surfaces in $\mathbb{P}^2$ and $\mathbb{P}^3$ was introduced utilizing tensor algebra. These features can be expressed by tensor algebra as codimension-1 hypersurfaces $\mathbf{h}_{\beta(d)} \mathbf{x}^{\beta(d)} = 0$ using the symmetrization operator. The coefficients of a hypersurface of degree-d embedded in $\mathbb{P}^n$ will have (generically) either $\nu^d_n = \binom{d+n}{d} - 1$ **DOF** if it is a curve(ν^d_2)/surface(ν^d_3) or $\xi^d_5 = \binom{d+5}{d} - \binom{d-2+5}{d-2} - 1$ **DOF** if it is a Chow polynomial.

We are interested in determining the distance of a hyperplane $\mathbf{x}^{\beta} \in \mathbb{P}^n$ from a hypersurface - where the hyperplane is not exactly incident with the hypersurface - using the bilinear parameter estimation framework (4). The parameters of the model are the coefficients of the hypersurface $\mathbf{h}_{\beta(d)} \in \mathbb{P}^{\nu^d_2}$ and we will assume those to be fixed, the measurements correspond with the hyperplane $\mathbf{x}^{\beta} \in \mathbb{P}^2$ lying on or near the hypersurface. The noise model (ϵ) in this case is associated with the contraction of the embedded hyperplane $\mathbf{x}^{\beta^{(d)}}$ with the coefficients of the hypersurface $\mathbf{h}_{\beta^{(d)}}$ (this will be 0 for a hyperplane incident with the hypersurface). With these specializations equation (4) becomes,

$$\epsilon = \mathbf{h}_{\beta(d)}\mathbf{x}^{\beta^{(d)}} \tag{15}$$

where ϵ is a 1-dimensional Gaussian pdf $\mathcal{G}(0, \sigma_{\mathbf{x}}^2 \Sigma_{\mathbf{x}})$. In order to determine an approximation to the nuisance parameter (the unperturbed position of the hyperplane) $\hat{\mathbf{x}}^{\beta}$, we utilize equation (13) resulting in,

$$\Delta\mathbf{x} = \hat{\mathbf{x}} - \mathbf{x} \approx -\hat{\Sigma}_{\mathbf{x}} \mathbf{J}_{\mathbf{x}}^{\epsilon\top} \left(\mathbf{J}_{\mathbf{x}}^{\epsilon} \hat{\Sigma}_{\mathbf{x}} \mathbf{J}_{\mathbf{x}}^{\epsilon\top}\right)^{+} \epsilon$$

allowing the approximation to be calculated as $\hat{\mathbf{x}} = \mathbf{x} + \Delta\mathbf{x}$. The corresponding error metric for embedded hypersurfaces in $\mathbb{P}^n$ can be defined according to equation (14).

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_{\hat{\Sigma}_{\mathbf{x}}}^2 \approx \epsilon^{\top} \left(\mathbf{J}_{\mathbf{x}}^{\epsilon} \hat{\Sigma}_{\mathbf{x}} \mathbf{J}_{\mathbf{y}}^{\epsilon\top}\right)^{+} \epsilon$$

An example is shown in Figure 1 of the correction of a series of points in $\mathbb{P}^2$ to their approximate location on a conic. The error applied to the points is synthetic and made to lie normal to the tangent of the conic. The bias exhibited by this approximate form of correction is evident in the location of the black crosses being consistently perturbed from their true location (the green circles).

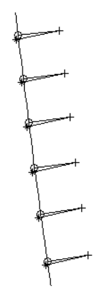

Figure 1. A section of a conic with the true points (o), approximated points (*) and noisy points (+). The approximations are consistently perturbed from the location of the true points.

4.2 Curves and Surfaces

Having outlined the general form for the approximate error metric and nuisance parameters associated with features of codimension-1, we can now specialize this formulation for curves in $\mathbb{P}^2$ and $\mathbb{P}^3$ and surfaces. Definitions of the noise models and the incident hyperplanes are presented for the different features in Table 1.

The definition of the noise model and accompanying error metric for a planar curve and a surface are very similar. Analytically, the major difference between planar curves and surfaces is the fact that surfaces lie in $\mathbb{P}^3$ and planar curve lie in $\mathbb{P}^2$, both sets of parameters are of codimension-1 and (generically) have no additional constraints unless we are estimating a special form of the hypersurface (eg. a parabola for degree-2).

The case for space curves embedded in a Chow polynomial is somewhat different (see [8]). The coefficients of the Chow polynomial of a curve are subject to a set of ancillary constraints generated by a simple relationship between a subset of the polynomials coefficients [1, 2, 8]. The noise models presented in Table 1 are geometrically valid iff. the coefficients of Chow polynomial satisfy the ancillary constraints.

4.3 Experiments

In order to assess the optimality of the approximate error metric (14), we have performed a series of random experiments where we compare the approximate values determined for the nuisance parameters ($\hat{\mathbf{x}}$) with the true values ($\bar{\mathbf{x}}$) using the pythagorean equality ($\|\mathbf{x} - \bar{\mathbf{x}}\|^2 = \|\mathbf{x} - \hat{\mathbf{x}}\|^2 + \|\bar{\mathbf{x}} - \hat{\mathbf{x}}\|^2$, see [5]). We expect there to be a bias in the estimates of the nuisance parameters but we are most interested in the extent of the bias as a function of the noise applied to the measurements ($\mathbf{x}$) as well as the degree (d) of the embedding.

The process used to test the nuisance parameters is to generate random degree-d planar Bezier curves and degree-d triangular Bezier surfaces and via the process of approximate implicitization (see [3]) determine the corresponding implicit equations ($\mathbf{c}_{A^{(d)}}/\mathbf{S}_{a^{(d)}}$). Since we now possess a parametric and implicit form of the Bezier we can accurately generate noisy measurements ($\mathbf{x}$) normal to the curve/surface at regular intervals - using the deCasteljau algorithm [4] to calculate tangents and then antisymmetric algebra to determine normals - whilst also retaining the true value of these points ($\bar{\mathbf{x}}$).

The experiments are structured such that each random planar curve and surface is tested at 100 positions ($\bar{\mathbf{x}}$) along it's domain with a 1-dimensional zero-mean Gaussian noise of varying standard deviation (σ) applied to the true measurement of each point in the direction of the normal. The results from tests on 100 randomly generated degree-d ($d = 2, \ldots, 4$) planar curves (Left) and surfaces (Right) are presented in Figure 3. The values on the vertical axis of Figure 3 are the average of $\|\mathbf{x} - \bar{\mathbf{x}}\|^2 - \|\mathbf{x} - \hat{\mathbf{x}}\|^2 - \|\bar{\mathbf{x}} - \hat{\mathbf{x}}\|^2$ (for an optimal estimator this value should be ~ 0 [5]), the horizontal axis is in terms of the standard deviation

Feature	Parameters	Measurements	Noise Model
Planar Curve	$\mathbf{c}_{A^{(d)}}$	$\mathbf{x}^A \in \mathbb{P}^2$	$\epsilon = \mathbf{c}_{A^{(d)}}\mathbf{x}^{A^{(d)}}$
Surface	$\mathbf{S}_{a^{(d)}}$	$\mathbf{x}^a \in \mathbb{P}^3$	$\epsilon = \mathbf{S}_{a^{(d)}}\mathbf{x}^{a^{(d)}}$
Space Curve (1)	$\mathbf{C}_{\omega^{(d)}}$	$\mathbf{x}^\omega \in \mathbb{P}^5$	$\epsilon = \mathbf{C}_{\omega^{(d)}}\mathbf{x}^{\omega^{(d)}}$
Space Curve (2)	$\mathbf{C}_{\omega^{(d)}}\mathbf{P}^{\omega^{(d)}}_{A^{(d)}}$	$\mathbf{x}^A \in \mathbb{P}^2$	$\epsilon = \mathbf{P}^{\omega^{(d)}}_{A^{(d)}}\mathbf{C}_{\omega^{(d)}}\mathbf{x}^{A^{(d)}}$

Table 1. Degree-d feature types (hypersurfaces) and their associated noise models in $\mathbb{P}^2$ & $\mathbb{P}^3$

(σ) of the noise applied normal to true measurements.

The results in Figure 3 indicate that as the standard deviation of the noise is increased, the approximate MLE of the nuisance parameters becomes increasingly less reliable. The relationship in these trials between the optimality of the approximate MLE error metric and the standard deviation can be observed to be approximately linear. Interestingly there is no correlation between the degree of the hypersurface and the optimality of the estimate. The addition of another dimension in the case of surfaces results in a slightly improved performance associated with a decrease in the gradient.

Also of interest is the quality of the estimate from the point to the curve in the presence of a singularity on the curve. A singular point on a planar curve $f(\mathbf{x}) \equiv \mathbf{c}_{A^{(d)}}\mathbf{x}^{A^{(d)}}$ is defined as any point $\mathbf{x} = [x_0, x_1, x_2]$ upon the domain of the curve where the partial derivatives $\frac{\partial f(\mathbf{x})}{\partial x_1}$ and $\frac{\partial f(\mathbf{x})}{\partial x_2}$ both equal 0 (assuming that x_0 is the homogenizing coefficient).

Singular points on plane curves (of degree > 2) can appear as either cusps, inflexion points or a multiple point of the curve. Figure 2 demonstrates the degeneration of the approximate MLE of a cubic plane curve in the presence of a singular point (cuspoidal). We can study the effect of a singular point on the approximate MLE of the nuisance parameters by observing the behaviour of the approximate MLE error metric as $\mathbf{J}^\epsilon_\mathbf{x}$ approaches the singular point. An example of this type of analysis is presented in Figure 4, where clearly the approximate MLE of the error metric increases as the L2-norm of the gradient approaches 0 (ie. the singularity). This implies that some care can be taken in practise to discount approximations of the error from portions of the algebraic hypersurface where the L2-norm of the gradient approaches 0, this strategy results in more reliable determination of the error metric.

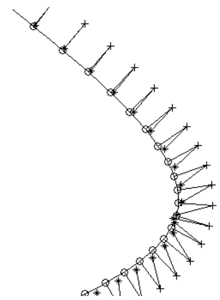

Figure 2. A cuspoidal section of a degree-3 curve with the true points (o) and the noisy points (+) adjusted to lie closer to the curve via a first-order approximation (*). The accuracy of the approximation decreases as the cusp is approached.

5 Discussion

We have presented the theory as well as an analysis of the approximate MLE framework using Gaussian noise models. We showed that the approximate MLE framework can be applied in a generic fashion to a suite a range of parameter estimation problem types and can also be used as an error metric.

In our analysis we focused upon the determination of nuisance parameters and through a series of experiments we show that optimality of this framework decreases linearly as a function of the Gaussian noise applied to the measurements. We also established that there is very little correlation between the degree and dimension of the hypersurface and optimality of the estimator. The challenge posed by algebraic singularities in the parameter space was also analyzed and a simple scheme nominated to identify singularities in a practical setting.

References

[1] A. Cayley. On a new analytical representation of curves in space. *Quart. J. of Pure and Appl. Math.*, 3, 1860.

[2] A. Cayley. On a new analytical representation of curves in space ii. *Quart. J. Math.*, 5, 1862.

[3] T. Dokken. *Aspects of Intersection Algorithms and Approximation.* PhD thesis, 1997.

[4] G. Farin. *Curves and Surfaces for Computer Aided Geometric Design: A Practical Guide.* Academic Press, San Diego, California, 1991.

[5] R. I. Hartley. Tutorial on optimization. In *ACCV (2)*, 2004.

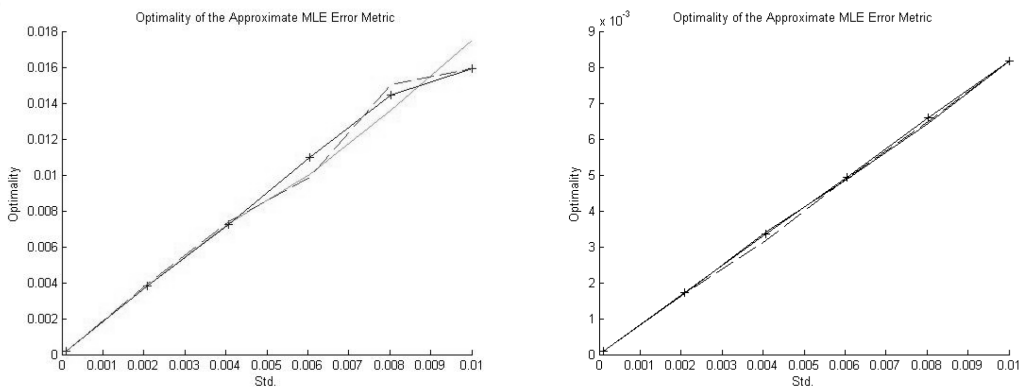

Figure 3. The optimality of the approximate MLE error metric over a range of standard deviations for a series of 1000 degree 2 (–), 3 (+-) and 4 (-) randomly generated planar curves (Left) and randomly generated surfaces (Right).

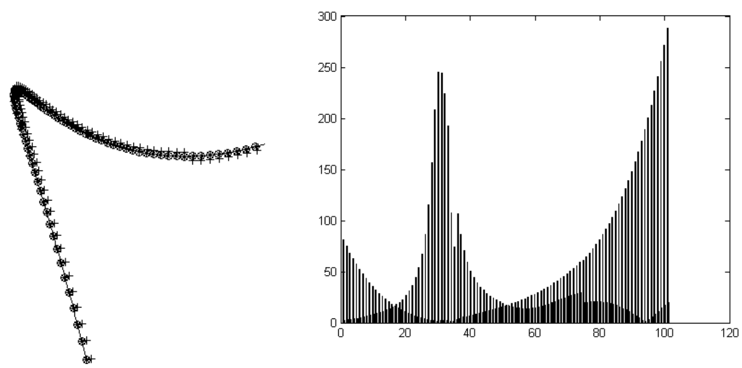

Figure 4. Left - A section of a cubic curve with noisy points and approximate corrections. Right - The corresponding L2-norm of the gradient (black) and AMLE residual values (gray) evaluated at points along the curve bottom-to-top.

[6] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[7] K. Kanatani. *Statistical Optimization for Geometric Computation: Theory and Practice*. Elsevier science, Amsterdam, 1996.

[8] D. N. McKinnon and B. C. Lovell. Tensor algebra: A combinatorial approach to the projective geometry of figures. *IWCIA04*, pages 558–568, 2004.

[9] 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.

[10] 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 Analysis and Machine Intelligence*, 13(11):1115–1138, 1991.