ELLIPSE-SPECIFIC DIRECT LEAST-SQUARE FITTING

Maurizio Pilu   *Andrew W. Fitzgibbon*   *Robert B. Fisher*

Department of Artificial Intelligence,
The University of Edinburgh
5 Forrest Hill, Edinburgh EH1 2QL
SCOTLAND

ABSTRACT

This work presents the first direct method for specifically fitting ellipses in the least squares sense. Previous approaches used either generic conic fitting or relied on iterative methods to recover elliptic solutions. The proposed method is (i) ellipse-specific, (ii) directly solved by a generalised eigen-system, (iii) has a desirable low-eccentricity bias, and (iv) is robust to noise. We provide a theoretical demonstration, several examples and the Matlab coding of the algorithm.

1. INTRODUCTION

Ellipse fitting is one of the classic problems of pattern recognition and has been subject to considerable attention in the past ten years for its many application. Several techniques for fitting ellipses are based on mapping sets of points to the parameter space (notably the Hough transform).

In this paper we are concerned with the more fundamental problem of least squares (LSQ) fitting of ellipses to scattered data. Previous methods achieved ellipse fitting by using generic conic fitters that perform poorly, often yielding hyperbolic fits with noisy data, or by employing iterative methods, which are computationally expensive.

In this paper we presents and demonstrate the first ellipse-specific direct least squares fitting method that has the following desirable features: *i)* always yields elliptical fits *ii)* has low-eccentricity bias, and *iii)* is robust to noise.

2. THE LSQ ELLIPSE FITTING PROBLEM

Let us represent a generic conic as the zero set of an implicit second order polynomial:

$$F(\mathbf{a}, \mathbf{x}) = \mathbf{a}\mathbf{x} = ax^2 + bxy + cy^2 + dx + ey + f \quad (1)$$

where $\mathbf{a} = [\, a\; b\; c\; d\; e\; f \,]$ and $\mathbf{x} = [x^2\; xy\; y^2\; x\; y\; 1\,]^T$. $F(\mathbf{a}, \mathbf{x}_i) = d$ is called the "algebraic distance" of a point $\mathbf{x}_i$ to the conic $F(\mathbf{a}, \mathbf{x}) = 0$.

One way of fitting a conic is to minimise the algebraic distance over the set of N data points in the least squares sense, that is

$$\hat{\mathbf{a}} = \arg\min_{\mathbf{a}} \left\{ \sum_{i=1}^{N} F(\mathbf{a}, \mathbf{x}_i)^2 \right\} \quad (2)$$

Linear conic fitting methods have been investigated that used linear constraints that slightly bias conic fitting towards elliptical solutions. In particular Rosin [8] and Gander [4] investigated the constraint $a + c = 1$ and Rosin [7] $f = 1$.

In a seminal work, Bookstein [1] showed that if a quadratic constraint is set on the parameters (e.g., to avoid the trivial solution $\mathbf{a} = \mathbf{0}_6$) the minimisation (2) can be solved by the rank-deficient generalised eigenvalue system:

$$\mathbf{D}^T\mathbf{D}\mathbf{a} = \mathbf{S}\mathbf{a} = \lambda\mathbf{C}\mathbf{a} \quad (3)$$

where $\mathbf{D} = [\mathbf{x}_1\; \mathbf{x}_2 \cdots\; \mathbf{x}_n]^T$ is called *design matrix*, $\mathbf{S} = \mathbf{D}^T\mathbf{D}$ is called *scatter matrix* and $\mathbf{C}$ is the matrix that expresses the constraint.

A simple constraint is $\|\mathbf{a}\| = 1$ but Bookestein used the algebraic invariant constraint $a^2 + \frac{1}{2}b^2 + c^2 = 1$; Sampson [10] presented an *iterative* improvement to Bookstein method that replaces the algebraic distance (1) with a better approximation to the geometric distance, which was adapted by Taubin [11] to turn the problem again into a generalised eigen-system like (3).

Despite the amount of work, direct *specific* ellipse fitting, however, was left unsolved. If ellipses fitting was needed, one had to rely either on generic conic fitting or on iterative methods such as [6]. Recently Rosin [9] re-iterated this problem by stating that ellipse-specific fitting is essentially a non-linear problem and iterative methods must be employed for this purpose. In the following we show that this is no longer true.

3. ELLIPSE-SPECIFIC METHOD

Let us consider a different quadratic constraint that corresponds to the well known quadratic algebraic invariant of a conic

$$b^2 - 4ac = \mathbf{a}^T \begin{bmatrix} 0 & 0 & -2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ -2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \mathbf{a} = \mathbf{a}^T\mathbf{C}\mathbf{a} < 0 \tag{4}$$

This constraint was first introduced in [3] and it was shown to yield always elliptical solutions; the brief justification given was that because of the immateriality of the scale of $\mathbf{a}$, the inequality (4) can, w.l.o.g., turned into $\mathbf{a}^T\mathbf{C}\mathbf{a} = -1$ and hence the minimisation (2) subject to the constraint (4) can again be formulated like in (3).

In the following, we give theoretical account of the method by demonstrating its key feature of ellipse specificity, i.e. that it gives always *one and only one* elliptical solution. But before that, we need to state two Lemmas that will naturally lead to an uniqueness theorem.

Let $\mathbf{S} \in \Re_{n\times n}$ and $\mathbf{C} \in \Re_{n\times n}$ be symmetric matrices, with $\mathbf{S}$ positive definite. Let us define the *spectrum* $\sigma(\mathbf{S})$ as the set of eigenvalues of $\mathbf{S}$ and let $\sigma(\mathbf{S},\mathbf{C})$ analogously be the set of generalised eigenvalues of (5).

Lemma 1 *The signs of the generalised eigenvalues of*

$$\mathbf{S}\mathbf{u} = \lambda\mathbf{C}\mathbf{u} \tag{5}$$

are the same as those of the matrix $\mathbf{C}$*, up to permutation of the indices.*

Proof: Let the *inertia* $i(\mathbf{S})$ be defined as the set of signs of $\sigma(\mathbf{C})$, and let $i(\mathbf{S},\mathbf{C})$ analogously be the inertia of $\sigma(\mathbf{S},\mathbf{C})$. Then, the lemma is equivalent to proving that $i(\mathbf{S},\mathbf{C}) = i(\mathbf{C})$. As $\mathbf{S}$ is positive definite, it may be decomposed as $\mathbf{Q}^2$ for symmetric $\mathbf{Q}$, allowing us to write (5) as $\mathbf{Q}^2\mathbf{u} = \lambda\mathbf{C}\mathbf{u}$. Now, substituting $\mathbf{v} = \mathbf{Q}\mathbf{u}$ and pre-multiplying by $\mathbf{Q}^{-1}$ gives $\mathbf{v} = \lambda\mathbf{Q}^{-1}\mathbf{C}\mathbf{Q}^{-1}\mathbf{v}$ so that $\sigma(\mathbf{S},\mathbf{C}) = \sigma(\mathbf{Q}^{-1}\mathbf{C}\mathbf{Q}^{-1})^{-1}$ and thus $i(\mathbf{S},\mathbf{C}) = i(\mathbf{Q}^{-1}\mathbf{C}\mathbf{Q}^{-1})$. From Sylvester's Law of Inertia [12] we have that for any symmetric $\mathbf{S}$ and nonsingular $\mathbf{X}$, $i(\mathbf{S}) = i(\mathbf{X}^T\mathbf{S}\mathbf{X})$. Therefore, substituting $\mathbf{X} = \mathbf{X}^T = \mathbf{Q}^{-1}$ we have $i(\mathbf{C}) = i(\mathbf{Q}^{-1}\mathbf{C}\mathbf{Q}^{-1}) = i(\mathbf{S},\mathbf{C})$. □

Lemma 2 *If* $(\lambda_i,\ \mathbf{a}_i)$ *is a solution of the eigen-system (3), we have:* $sign(\lambda_i) = sign(\mathbf{a}_i^T\mathbf{C}\mathbf{a}_i)$.

Proof: By pre-multiplying by $\mathbf{a}_i^T$ both sides of (3) we have $\mathbf{a}_i^T\mathbf{S}\mathbf{a}_i = \lambda_i\mathbf{a}_i^T\mathbf{C}\mathbf{a}_i$. Since $\mathbf{S}$ is positive-definite, $\mathbf{a}_i^T\mathbf{S}\mathbf{a}_i > 0$ and therefore λ_i and the scalar $\mathbf{a}_i^T\mathbf{C}\mathbf{a}_i$ must have the same sign. □

Now we can state the following uniqueness theorem:

Theorem 1 *The solutions to the conic fitting problem given by the generalised eigen-system (3) subject to the constraint (4) include one and only one elliptical solution corresponding to the single negative generalised eigenvalue of (3). The solution is also invariant to rotation and translation of the data points.*[1]

Proof: Since the non-zero eigenvalues of $\mathbf{C}$ are $\sigma(\mathbf{C}) = \{\ -2,\ 1,\ 2\ \}$, from Lemma 1 we have that $\sigma(\mathbf{S},\mathbf{C})$ has one and only one negative eigenvalue $\lambda_i < 0$, associated with a solution $\mathbf{a}_i$; then, by applying Lemma 2, the constraint $\mathbf{a}_i^T\mathbf{C}\mathbf{a}_i = b^2 - 4ac$ is negative and therefore $\mathbf{a}_i$ is a set of coefficients representing an ellipse. The constraint (4) is a conic invariant to Euclidean transformations and so is the solution (see [1]) □

Theorem 1 does not state anything about the *quality* of the unique elliptical solution, since classical optimisation theory states that it might not be the global minimum of (2) under our *non-positive definite* inequality constraint. However, the physical solution (the actual ellipse) does not change under linear scaling of the coefficients and therefore it can be easily shown that the minimisation with the inequality constraint (4) can be equivalently turned to a minimisation with an equality constraint $\mathbf{a}^T\mathbf{C}\mathbf{a} = -1$. By doing so, as illustrated in [2], we can say that:

Corollary 1 *The unique elliptical solution is the one that minimises (2) subject to the constraint* $\mathbf{a}^T\mathbf{C}\mathbf{a} = -1$.

A more practical interpretation of this corollary is that the unique elliptical solution is a local minimiser of the *Rayleigh quotient* $\frac{\mathbf{a}^T\mathbf{S}\mathbf{a}}{\mathbf{a}^T\mathbf{C}\mathbf{a}}$ and thus the solution can also be seen as the *best least squares ellipse under a re-normalisation of the coefficients by* $b^2 - 4ac$. Although experimental evidence would suggest that this statement could be valid, a formal demonstration is currently not known to the authors. This implicit normalisation turns singular for $b^2 - 4ac = 0$ and, following the observations in [7], we can say that the minimisation tends to "pull" the solution away from singularities; in our case the singularity is a parabola and so the unique elliptical solution tends to be biased towards low eccentricity, which explains many of the following results, such as those in Figure 2.

[1] Since $\mathbf{C}$ is rank deficient, the eigen-system (3) should be solved by block decomposition like in [1]; however most numerical packages will handle this detail.

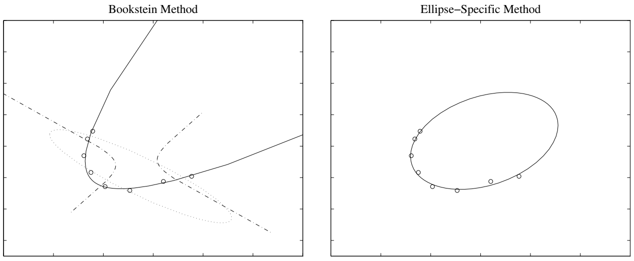

Figure 1: Specificity to ellipses. The three eigensolution obtained by the Bookstein algorithm (left) and the best LSQ *elliptical* solution obtained by our ellipse-specific method (right).

Figure 2: Low-eccentricity bias of the ellipse-specific method when fitting to noisy parabolic data. Encoding is Bookstein: dotted; Gander: dashed; Taubin: dash-dot; Ellipse-specific: solid.

4. EXPERIMENTAL RESULTS

First, let us now have a glimpse at what this ellipse-specificity means. Figure 1-left shows the three eigensolutions yielded by the Bookstein algorithm on a small set of hand-input points; the best LSQ fit is a hyperbola and the (incidentally) elliptical one is extremely poor. With the proposed ellipse-specific algorithm, the only solution satisfying the constraint is the best LSQ *elliptical* solution, shown in Figure 1-right.

Figure 2 shows three experiments designed after [10] that consist of the same *parabolic* data but with different realizations of added isotropic Gaussian noise ($\sigma = 10\%$ of data spread). In his paper, Sampson refined the poor initial fitting obtained with Bookstein algorithm using an iterative Kalman filter to minimise his approximate geometrical distance [10]. The final results were ellipses with low eccentricity that are qualitatively similar to those produced by our ellipse-specific direct method (solid lines) but at the *same* computational cost of producing Sampson's initial estimate.

The low-eccentricity bias of our method discussed in Section. 3 is most evident in Figure 2 when comparing the results to other methods, namely Bookstein (dotted), Taubin (dash-dots) and Gander (dashed); these results are not surprising, since those methods are non-ellipse specific whereas the one presented here is.

Let us now qualitatively illustrate the robustness of the ellipse-specific method as compared to Gander's and Taubin's. A number of experiments have been carried out, of which here we present a couple, shown in Figures 3 and 4. They have been conducted by adding isotropic Gaussian noise to a synthetic elliptical arc; note that in both sets each column has the *same* set of points. More quantitative results can be found in [2] and are not reported here for reasons of space.

Figure 3 shows the performance with respect to increasing noise level (see [3] for more experiments). The standard deviation of the noise varies from 3% in the leftmost column to 20% of data spread in the rightmost column; the noise has been set to relatively high level because the performance of the three algorithm is substantially the same at low noise level of *precise* elliptical data. The top row shows the results for the method proposed here. Although, as expected, the fitted ellipses shrink with increasing levels of high noise (as a limit the elliptical arc will look like a noisy line), it can be noticed that the ellipse dimension decreases smoothly with the increase of noise level: this is an indication of well-behaved fitting. This shrinking phenomenon is evident also with the other two methods but presents itself more erratically: in the case of Taubin's algorithm, the fitted ellipses are on average somewhat closer to the original one [3], but they are rather unpredictable and its ellipse non-specificity, as it happens in the Gander's case, sometimes yields unbounded hyperbolic fits.

The second set, shown in Figure 4, is concerned with assessing stability to different realizations of noise with the *same variance* ($\sigma = 0.1$). (It is very desirable that an algorithm's performance be affected only by the noise level, and not by a particular realization of the noise). This and similar experiments (see [2, 3]) showed that our method has a remarkably greater stability to noise with respect to Gander's and Taubin's.

5. CONCLUSION

In this paper we have presented an ellipse least squares fitting method which for is specific to ellipses and direct at the same time; other previous method were either not ellipse-specific or iterative.

We argue that our method is possibly the best trade-off between speed and accuracy for ellipse fitting and its uniqueness property makes it also extremely robust to

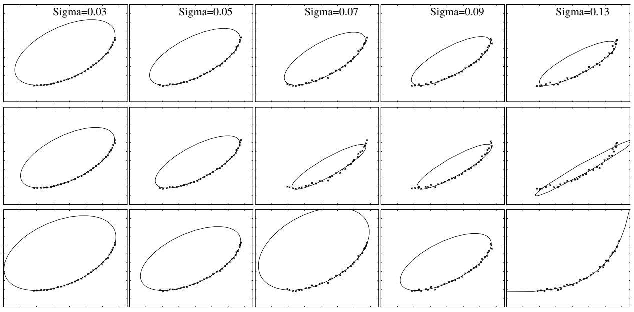

Figure 3: Stability experiments with increasing noise level. Top row: ellipse-specific method; Mid Row: Gander; Bottom Row: Taubin. The ellipse-specific method shows a much smoother and predictable decrease in quality than the other two methods.

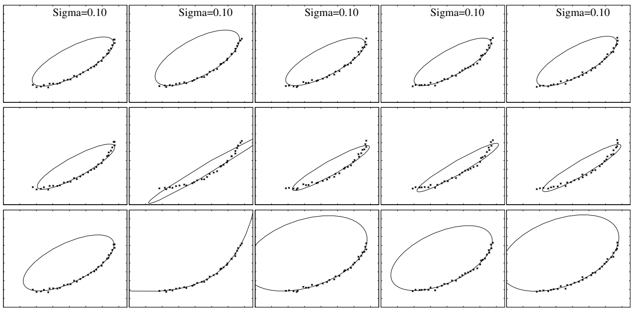

Figure 4: Stability experiments for different runs with same noise variance (10% of data spread). Top row: ellipse-specific method; Mid Row: Gander; Bottom Row: Taubin. The ellipse-specific method shows a remarkable stability.

noise and usable in many applications, especially in industrial vision. In order for other researchers to quickly assess the validity of the method, Figure 5 gives a Matlab implementation of the proposed algorithm and an interactive JAVA demonstration is available at *http://vision.dai.ed.ac.uk/maurizp/ElliFitDemo/demo.html*.

In the near future, a method for correcting the bias to the noise for incomplete elliptical arcs will be explored that is inspired by [5]. Moreover, the proposed ellipse-specific method could be used to produce excellent initial estimates for iterative methods, thus significantly increasing their speed; we are currently investigating this possibility.

Acknowledgements: Maurizio Pilu was partially sponsored by SGS-THOMSON Microelectronics. This work was partially funded by UK EPSRC Grant GR/H/86905.

```
% x,y are lists of coordinates
function a = fit_ellipse(x,y)
% Build design matrix
D = [ x.*x x.*y y.*y x y ones(size(x)) ];
% Build scatter matrix
S = D'*D;
% Build 6x6 constraint matrix
C(6,6) = 0; C(1,3) = -2; C(2,2) = 1; C(3,1) = -2;
% Solve eigensystem
[gevec, geval] = eig(inv(S)*C);
% Find the negative eigenvalue
[NegR, NegC] = find(geval < 0 & ~isinf(geval));
% Extract eigenvector corresponding to positive eigenvalue
a = gevec(:,NegC);
```

Figure 5: Complete 6-line Matlab implementation of the proposed algorithm.

6. REFERENCES

[1] F.L. Bookstein. Fitting conic sections to scattered data. *Computer Graphics and Image Processing*, (9):56–71, 1979.

[2] A. W. Fitzgibbon, M. Pilu, and R.B. Fisher. Direct least squares fitting of ellipses. In *International Conference on Pattern Recognition*, August 1996. To Appear.

[3] A.W. Fitzgibbon and R.B. Fisher. A buyer's guide to conic fitting. In *Proceedings of British Machine Vision Conference*, Birmingam, 1995.

[4] W. Gander, G.H. Golub, and R. Strebel. Least-square fitting of circles and ellipses. *BIT*, (43):558–578, 1994.

[5] K. Kanatani. Statistical bias of conic fitting and renormalization. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 16(3):320–326, 1994.

[6] J. Porrill. Fitting ellipses and predicting comfidence envelopes using a bias corrected Kalman filter. *Image and Vision Computing*, 8(1), February 1990.

[7] P.L. Rosin. A note on the least squares fitting of ellipses. *Pattern Recognition Letters*, (14):799–808, October 1993.

[8] P.L. Rosin and G.A. West. Segmenting curves into lines and arcs. In *Proceedings of the Third International Conference on Computer Vision*, pages 74–78, Osaka, Japan, December 1990.

[9] P.L. Rosin and G.A. West. Non-parametric segmentation of curves into various representations. *PAMI*, 17(12):1140–1153, December 1995.

[10] P.D. Sampson. Fitting conic sections to very scattered data: an iterative refinement of the Bookstein algorithm. *Computer Graphics and Image Processing*, (18):97–108, 1982.

[11] G. Taubin. Estimation of planar curves, surfaces and non-planar space curves defined by implicit equations, with applications to edge and range image segmentation. *IEEE PAMI*, 13(11):1115–1138, November 1991.

[12] J. H. Wilkinson. *The algebraic eigenvalue problem*. Claredon Press, Oxford, England, 1965.