Fitting a Second Degree Curve if Both Coordinates Are Subject to Error

M. Werman and Z. Geyzel

Abstract— **This paper presents a statistically sound, simple and fast method to estimate the parameters of a second degree curve from a set of noisy points that originated from the curve.**

Keywords— **pattern analysis, low-level processing, perceptual grouping, curve fitting**

Introduction

Fitting a straight line or another figure that is consistent with and approximates a set of data points is a common problem encountered in computer vision, pattern recognition or statistics. The literature is full of different methods that find the "best" describing figure for a set of points, see [1] for a summary.

The method of least squares for computing the parameters is used very often even though for over 100 years [3] it has been known that in the case of noise in both the x and y axes, the least squares estimate is biased even for the case of straight line fitting.

There have been many suggestions as how to compute better estimates for the case of a straight line [6, 4, 5, 8, 2, 7] there have been a few treatments of conic section fitting [10] but they usually need a well defined model and are much more complex then our proposed method.

In this paper we describe a simple and easy to compute estimate of a fitting line or conic section that converges in probability to the correct solution. The algorithm is linear in the number of points and the convergence is up to a constant multiple optimal.

We will first present the solution for the straight line case as it has less parameters and is a little easier to describe. In the straight line case the method is somewhat similar to the method proposed in [9]. We will then present the general solution.

The problem

What we are modeling is the following: There is a plane curve γ that is a conic section; i.e. the curve is a set of points (x, y) s.t. $P(x, y) = 0$ where P is 2nd-degree bivariate polynomial with unknown coefficients. The "picture of γ" is a finite set of stochastic points $W_i = (X_i, Y_i)_{i=1,\dots,N}$ (N is "cardinality" of the picture) where

$$X_i = x_i + u_i \qquad Y_i = y_i + v_i$$

Dept. of Computer Science, The Hebrew University, 91904, Jerusalem, Israel, werman@cs.huji.ac.il

s.t.

- the points $w_i = (x_i, y_i)_{i=1,\dots,N}$ are on γ: $P(x_i, y_i) = 0$.
- $(u_i)_{i=1,\dots,N}, (v_i)_{i=1,\dots,N}$ are independent stochastic variables ("noise") with distributions that satisfy the following conditions:

$$E(u_i) = E(v_i) = 0, \;\; V(u_i) = V(v_i) = \sigma^2 < \infty, \;\;\; E(\mid u_i \mid^3), E(\mid v_i \mid^3$$

The problem is to compute the coefficients of P from a picture.

We also assume that:

- the picture is bounded: $\exists H > 0$ s.t. $\max_i(\mid x_i \mid), \max_i(\mid y_i \mid) < H$.
- the picture isn't "k-concentrated": $\exists \epsilon > 0 : \forall$ intervals $l_1, \dots, l_k$ on the γ with total length $< \epsilon$

  $$\overline{\lim_{N \to \infty}} \frac{\sum_{j=1}^{k} W(N, j)}{N} < 1$$

  where $W(N, j) = \#\{i \in \{1 \dots N\} \mid w_i \in l_j\}$.
  In another words the data points in the picture aren't concentrated around k points on γ, therefore intuitively there are at least $k + 1$ points on it; $k = 4$ is enough in our case because for 5 points a conic section is uniquely defined, otherwise there are more parameters to compute than data.
- what we see is mostly signal; namely, the following three facts are usually true (a positive fraction is more than enough);
  - $X_i > X_j$ *implies* $x_i > x_j$
  - $Y_i > Y_j$ *implies* $y_i > y_j$
  - the point (noise+signal) is in the picture *implies* that the point without noise should be in the picture.

The case of a line

We first present the case where the points are known to have originated from γ a straight line, $y = ax + b$, where a and b are unknown constants.

Assume that N (the number of points) is even and compute

$$A = \frac{\sum_{i=1}^{N/2}(Y_{2i} - Y_{2i-1})}{\sum_{i=1}^{N/2}(X_{2i} - X_{2i-1})}$$

Thus

$$\begin{aligned} A - a &= \frac{\sum_{i=1}^{N/2}((Y_{2i}-aX_{2i})-(Y_{2i-1}-aX_{2i-1}))}{\sum_{i=1}^{N/2}(X_{2i}-X_{2i-1})} \\ &= \frac{\sum_{i=1}^{N/2}((v_{2i}-v_{2i-1})-a(u_{2i}-u_{2i-1}))}{\sum_{i=1}^{N/2}(X_{2i}-X_{2i-1})} \\ &= \frac{\sum_{i=1}^{N/2}\theta_i}{\sum_{i=1}^{N/2}(X_{2i}-X_{2i-1})} \end{aligned}$$

where $\theta_i = (v_{2i} - v_{2i-1}) - a(u_{2i} - u_{2i-1})_{,i=1,\ldots,N/2}$

If we take pairs (W_{2i}, W_{2i-1}) s.t. $X_{2i} > X_{2i-1}$ (by reenumeration)

$$E(\sum_{i=1}^{N/2}(X_{2i} - X_{2i-1})) = \Omega(N)$$

(the picture isn't concentrated), but by the assumptions $E(\theta_i) = 0$

$$\sum_{i=1}^{N/2}\theta_i = O(\sqrt{N})$$

and therefore when $N \longrightarrow \infty$

$$A - a = O(\frac{1}{\sqrt{N}}) \longrightarrow_p 0 \ (converges\ \ in\ \ probability)$$

Remark: in this case for "unconcentration" it's enough that $k = 1$ because 2 points define a unique straight line.

Remark: θ_i depends on a, in order to assure that a is always bounded by 1, it is possible to use the algorithm twice by transposing the x and y axes.

For b we will use the expression

$$B = \frac{\sum_{i=1}^{N/2}(X_{2i}Y_{2i-1} - X_{2i-1}Y_{2i})}{\sum_{i=1}^{N/2}(X_{2i} - X_{2i-1})}$$

Thus

$$B - b = \frac{\sum_{i=1}^{N/2}(X_{2i}Y_{2i-1} - X_{2i-1}Y_{2i} - b(X_{2i} - X_{2i-1}))}{\sum_{i=1}^{N/2}(X_{2i} - X_{2i-1})}$$

but

$$\begin{aligned} &E(X_{2i}Y_{2i-1} - X_{2i-1}Y_{2i} - bX_{2i} + bX_{2i-1}) = \\ &E((x_{2i} + u_{2i})(y_{2i-1} + v_{2i-1}) - (x_{2i-1} + u_{2i-1})(y_{2i} + v_{2i}) \\ &-b(x_{2i} + u_{2i}) + b(x_{2i-1} + u_{2i-1})) \\ &= x_{2i}y_{2i-1} - x_{2i-1}y_{2i} - bx_{2i} + bx_{2i-1} \\ &= x_{2i}(y_{2i-1} - b) - x_{2i-1}(y_{2i} - b) \\ &= x_{2i}ax_{2i-1} - x_{2i-1}ax_{2i} = 0 \end{aligned}$$

and as for a when $N \longrightarrow \infty$ $B \longrightarrow_p b$ (converges in probability).

The general case

If γ is a conic section, $x^2 = axy + by^2 + cx + dy + e$, (as in the last case if some of the parameters are known they do not have to be estimated and the method is used only on the unknown parameters of the curve, for example if the curve is known to be a circle we know that a is 0 and b is 1.)

The algorithm is as follows: for a point $W = (X, Y)$ - we define the following vectors:

$$\begin{aligned} L(W) &= (XY, Y^2, X, Y, 1) \\ L_a(W) &= (X^2, Y^2, X, Y, 1) \\ L_b(W) &= (XY, X^2, X, Y, 1) \\ L_c(W) &= (XY, Y^2, X^2, Y, 1) \\ L_d(W) &= (XY, Y^2, X, X^2, 1) \end{aligned}$$

For a collection of 5 points $W_1, \ldots, W_5$ let

$$S(W_1, \ldots, W_5) = det \begin{pmatrix} L(W_1) \\ \vdots \\ L(W_5) \end{pmatrix}$$

and similarly the determinants S_a, S_b, S_c, S_d using L_a, L_b, L_c, L_d.

Reorder the points so that, $S(W_{5i+1}, \ldots, W_{5i+5}) \geq 0$ then the approximation for a is:

$$\frac{\sum_{i=0}^{\frac{N}{5}-1} S_a(W_{5i+1}, \ldots, W_{5i+5})}{\sum_{i=0}^{\frac{N}{5}-1} S(W_{5i+1}, \ldots, W_{5i+5})}$$

similarly for b,c, and d, proof in Appendix 1, (the reordering only entails changing the sign of the determinant).

Remark: The rate of convergence for the coefficients is $O(N^{-1/2})$ which is optimal.

Remark: This algorithm is, similar to the algorithm in the case of straight line.

Remark: The algorithm takes $O(N)$ time and $O(1)$ storage.

Remark: The algorithm is more stable if the quituplets are not taken to be points that are very close together.

The free coefficient

The last problem is to find the coefficient e, which is essentially a scale or size parameter. This is possible only with additional assumptions: we will assume that the noise is Gaussian (the same type of analysis can be carried out for other families of noise, such as, uniform). This analysis is much more dependent on the noise characteristics of the points and on the type of conic section under scrutiny, the type is known either from prior knowledge or from the coefficients already computed. The derivations are in Appendix II. To see intuitively why this parameter is different, look at the problem of approximating the radius of a circle with known center, the average distance of the points of the noisy circle do not center around the radius as the curvature gets larger more and more points will fall outside of the circle.

By translations and rotations (they keep a Gaussian distribution) - we can arrive at one of two kinds of equation:

$$x^2 + ky^2 = M$$

or

$$x^2 + ky = M$$

where k is known and M is unknown constant.

The first case includes ellipses and hyperbolas; the second case is parabolas.

Circle, ellipse, hyperbola $x^2 + ky^2 = M$:
Define:

$$A_N = \frac{1}{N}\sum_{i=1}^{N}(X_i^2 + kY_i^2), \quad B_N = \frac{1}{N}\sum_{i=1}^{N}(X_i^2 + kY_i^2)^2$$

$$S_N = \frac{1}{N}\sum_{i=1}^{N}(X_i^2 + k^2Y_i^2)$$

$$Z = \frac{2S_N - \sqrt{4S_N^2 + 2(k^2+1)(B_N - A_N^2)}}{2(k^2+1)} \longrightarrow_p \sigma^2$$

$$A_N - (1+k)Z \longrightarrow_p M$$

Parabola, $x^2 + ky = M$:
Define:

$$A_N = \frac{1}{N}\sum_{i=1}^{N}(X_i^2 + kY_i), \quad B_N = \frac{1}{N}\sum_{i=1}^{N}(X_i^2 + kY_i)^2$$

$$S_N = \frac{1}{N}\sum_{i=1}^{N}(X_i^2 + k^2)$$

$$Z = S - k^2 - \sqrt{4\;(S_N - k)^2 - 2(B_N - A_N^2)} \longrightarrow_p \sigma^2$$

$$A_N - Z \longrightarrow_p M$$

Results

These methods result in curves that are visually almost indistinguishable from the original curves. Results of modeling on 100 points for circle with center $(0,0)$ and radius 20 when there is a $(0,5)$-Gaussian noise.

Method	x_{center}	y_{center}	radius
LSQ	-0.21	-0.41	21.43
This	+0.08	-0.04	20.23
True	0	0	20

Some typical results of fitting points with this method:

Appendix I

Proposition 1.

$$\begin{aligned}
E(S(W_1 \dots W_5)) &= S(w_1 \dots w_5) \\
E(S_a(W_1 \dots W_5)) &= S_a(w_1 \dots w_5) \\
E(S_b(W_1 \dots W_5)) &= S_b(w_1 \dots w_5) \\
E(S_c(W_1 \dots W_5)) &= S_c(w_1 \dots w_5) \\
E(S_d(W_1 \dots W_5)) &= S_d(w_1 \dots w_5)
\end{aligned}$$

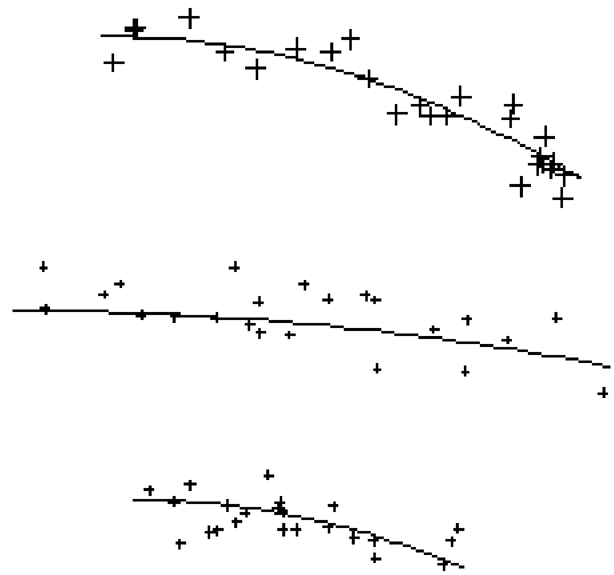

Figure 1: Some example sets:

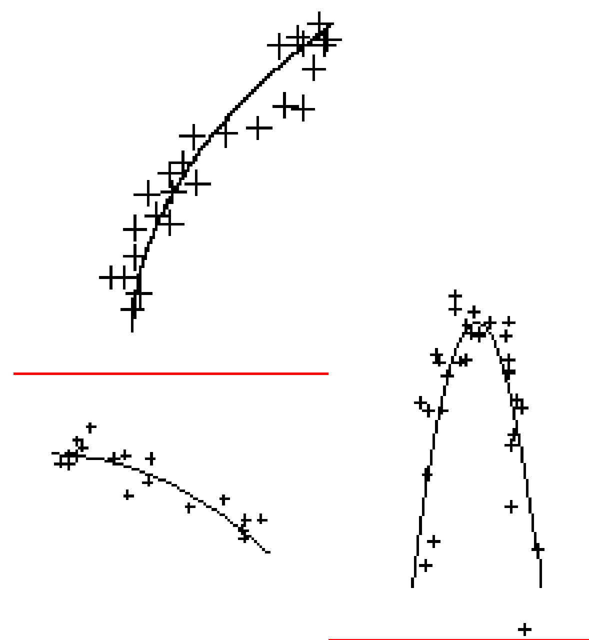

Figure 2: Some example sets:

(but the analogical equation for S_e is wrong).

Proof: $\lhd$ First,

$$\begin{aligned}
&E(XY) = E(X)E(Y) &&= xy \\
&E(Y^2) = E(Y)^2 + V(Y) &&= y^2 + \sigma^2 \\
&E(X) = x \\
&E(Y) = y
\end{aligned}$$

and so

$$E(L(W)) = (xy, y^2 + \sigma^2, x, y, 1)$$

Second, by one of the basic definitions a determinant

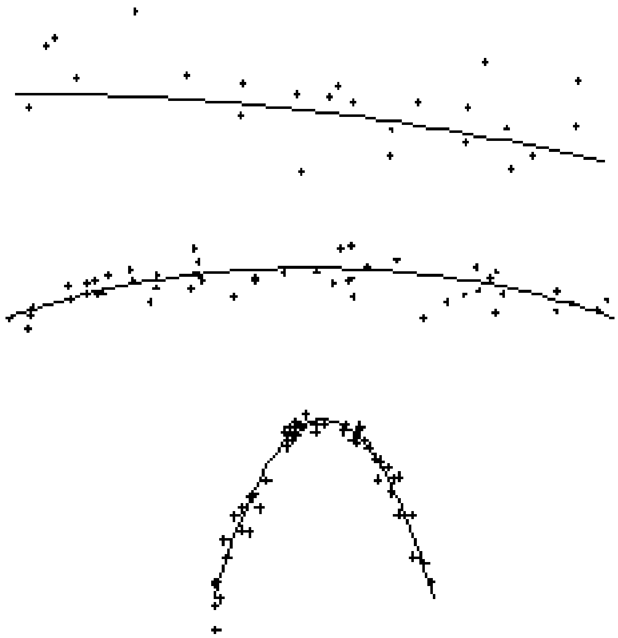

Figure 3: Some example sets:

is a sum of $\pm$ products of matrix elements from *different* rows. But the members in *different* rows are functions of *different* points and therefore they (matrix' elements) are independent $\Longrightarrow$

$$E(S(W_1 \dots W_5)) = det \begin{pmatrix} E(L(W_1)) \\ \vdots \\ E(L(W_5)) \end{pmatrix}$$

$$= det \begin{pmatrix} x_1 y_1 & y_1^2 + \sigma^2 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_5 y_5 & y_5^2 + \sigma^2 & x_5 & y_5 & 1 \end{pmatrix}$$

By the properties of the determinant we can subtract the 5^{th} (column of 1-s) multiplied by σ^2 from the 2^{nd} column $(y_i^2 + \sigma^2)$ - and the value of the determinant will be the same:

$$E(S(W_1 \dots W_5)) = det \begin{pmatrix} x_1 y_1 & y_1^2 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_5 y_5 & y_5^2 & x_5 & y_5 & 1 \end{pmatrix}$$

$$= S(w_1 \dots w_5)$$

(the last sign "=" by definition of S).

Similar proofs work for S_a, S_b, S_c, S_d. $\triangleright$

Proposition 2. If $w_1 = (x_1, y_1), \dots, w_5 = (x_5, y_5)$ are on the curve γ its polynomial equation is

$$x^2 = axy + by^2 + cx + dy + e$$

then

$$S_a = Sa, \;\; S_b = Sb, \;\; S_c = Sc, \;\; S_d = Sd$$

where $S, S_a, \dots, S_d$ are defined as above.

Proof

$\triangleleft$ Write the polynomial equation of γ for $w_1 = (x_1, y_1), \dots, w_5 = (x_5, y_5)$ as a matrix equation:

$$\begin{pmatrix} x_1 y_1 & y_1^2 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_5 y_5 & y_5^2 & x_5 & y_5 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \\ e \end{pmatrix} = \begin{pmatrix} x_1^2 \\ \vdots \\ x_5^2 \end{pmatrix}$$

The proposition follows from Cramer's rule for this linear system. $\triangleright$

Proposition 3: If $(W_i)_{i=1,\dots,N}$ is a "good" picture then it is possible to change the numbering of a pictures points s.t. in probability

$$\lim_{N\to\infty} A_N = a, \;\; \lim_{N\to\infty} B_N = b, \;\; \lim_{N\to\infty} C_N = c, \;\; \lim_{N\to\infty} D_N = d$$

where

$$A_N = \frac{\sum_{i=0}^{\frac{N}{5}-1} S_a(W_{5i+1}, \dots, W_{5i+5})}{\sum_{i=0}^{\frac{N}{5}-1} S(W_{5i+1}, \dots, W_{5i+5})}$$

and B_N, C_N, D_N are defined by the same way (we suppose that $N/5$ is integer).

Proof: $\triangleleft$ Change the order of points s.t.

$$S(W_{5i+1}, \dots, W_{5i+5}) \geq 0$$

So

$$A_N - a = \frac{\sum_{i=0}^{\frac{N}{5}-1} (S_a(W_{5i+1}, \dots, W_{5i+5}) - aS(W_{5i+1}, \dots, W_{5i+5}))}{\sum_{i=0}^{\frac{N}{5}-1} S(W_{5i+1}, \dots, W_{5i+5})}$$

Now

$$\begin{aligned} & E((S_a(W_{5i+1}, \dots, W_{5i+5}) - aS(W_{5i+1}, \dots, W_{5i+5})) = \\ & = E((S_a(W_{5i+1}, \dots, W_{5i+5})) - aE(S(W_{5i+1}, \dots, W_{5i+5})) \\ & = S_a(w_{5i+1}, \dots, w_{5i+5})) - aS(w_{5i+1}, \dots, w_{5i+5}) \text{ (by proposition 1)} \\ & = 0 \text{ (by proposition 2)} \end{aligned}$$

and therefore by the Law of Large Numbers

$$\sum_{i=0}^{\frac{N}{5}-1} (S_a(W_{5i+1}, \dots, W_{5i+5}) - aS(W_{5i+1}, \dots, W_{5i+5}) = O(\sqrt{N})$$

in probability. But

$$\sum_{i=0}^{\frac{N}{5}-1} S(W_{5i+1}, \dots, W_{5i+5}) = \Omega(N)$$

(here all summed members are positive)

$\Longrightarrow A_N - a = O(N^{-1/2})$ and finally $\lim_{N\to\infty}(A_N - a) = 0$ or in another words

$$\lim_{N\to\infty} A_N = a$$

in probability.

The same style of proof works for B_N, C_N, D_N. $\triangleright$

Appendix II

$$\begin{aligned}
E(Y_i) &= y_i\\
E(X_i^2) &= x_i^2+\sigma^2\\
E(Y_i^2) &= y_i^2+\sigma^2\\
E(X_i^4) &= x_i^4+6x_i^2\sigma^2+3\sigma^4\\
E(Y_i^4) &= y_i^4+6y_i^2\sigma^2+3\sigma^4\\
E(X_i^2Y_i^2) &= x_i^2y_i^2+(x_i^2+y_i^2)\sigma^2+\sigma^4\\
E(X_i^2Y_i) &= x_i^2y_i+y_i\sigma^2
\end{aligned}$$

$\underline{x^2+ky^2}$:

$$\begin{aligned}
E(A_N) &= M+(1+k)\sigma^2\\
E(B_N) &= M^2+\tfrac{1}{N}\textstyle\sum_{i=1}^N(6x_i^2+6y_i^2+2k(x_i^2+y_i^2))\sigma^2\\
&+(3+3k^2+2k)\sigma^4\\
&= (E(A_N))^2+\tfrac{4\sigma^2}{N}\textstyle\sum_{i=1}^N(x_i^2+k^2y_i^2)+2(1+k^2)\sigma^4
\end{aligned}$$

From the last equation we see that $E(B_N)-(E(A_N))^2$ is positive. Second, we can continue the calculations for $E(B_N)$:

$$E(B_N)-4\sigma^2S_N=E^2(A_N)-2(1+k^2)\sigma^4$$

But $A_N \longrightarrow_p E(A_N)$ and therefore

$$2(1+k^2)Z^2-4S_NZ+(B_N-A_N^2)\longrightarrow 0$$

where $Z=\sigma^2$. Solving the quadratric in σ^2 and taking the positive square root as σ^2 must be positive gives the result.

$\underline{x^2+ky=M}$:

$$\begin{aligned}
E(X_i^2+kY_i) &= M+\sigma^2\\
E(X_i^2+kY_i)^2 &= (E(A_N))^2+2\,\sigma^4+4\,x^2\sigma^2
\end{aligned}$$

Continuing as in the last case we get the final result.

References

[1] R. Duda and P. Hart. *Pattern Classification and Scene Analysis*. Wiley, 1973.

[2] H. Imai, K. Kato, and P. Yamamoto. A linear time algorithm for linear l_1 approximation of points. *Algorithmica*, 4:77–96, 1989.

[3] C.H. Kummel. Reduction of observed equations which contain more than one observed quantity. *The Analyst*, 6:97–105, 1879.

[4] A. Madansky. The fitting of straight lines when both variables are subject to error. *American Statistical Association Journal*, 54:173–205, 1959.

[5] P.A.P. Moran. Estimating structural and functional relationships. *Journal of Multivariate Analysis*, 1:232–255, 1971.

[6] F. P. Preparata and I. M. Shmaos. *Comptational Geometry*. Springer-Verlag, New York, 1985.

[7] A. Stein and M. Werman. Finding the repeated median regression line. In *3'rd Symposium on Discrete Algorithms*, 1992.

[8] A. Stein and M. Werman. Robust statistics in shape fitting. In *CVPR*, pages 540–546, Champaign, 1992.

[9] A. Wald. Fitting of straight lines if both variables are subject to error. *Annals of Mathematical Statistics*, 11:284–300, 1940.

[10] M. Werman, A.Y. Wu, and R.A. Melter. Recognition and characterization of digitized curves. *Pattern Recognition Letters*, 5:207–213, 1987.