COMPUTER GRAPHICS AND IMAGE PROCESSING 9, 56–71 (1979)

Fitting Conic Sections to Scattered Data

FRED L. BOOKSTEIN

Center for Human Growth and Development and Departments of Statistics and Biostatistics, The University of Michigan, Ann Arbor, Michigan 48109

Received January 6, 1978; revised March 28, 1978

The problem of fitting conic sections to scattered data has arisen in several applied literatures. The quadratic form $Ax^2 + Bxy + Cy^2 + Dx + Ey + F$ that is minimized in mean-square is proportional to the ratio of two squared distances along rays through the center of a conic. Considerations of invariance under translation, rotation, and scaling of the data configuration lead to a straightforward method of estimation somewhat different from earlier suggestions. The method permits an extension to conic splines around extended digitized curves, expediting a smooth reconstruction of their curvature. Some examples are presented indicating how the technique might be applied in morphometrics.

1. INTRODUCTION

In the course of investigations into the shape of plane curves, several authors, attempting to fit conic sections to scatters, have recently and independently discovered algorithms that are reducible to conventional linear computations. All minimize the sum-of-squares of a form $Q(x, y) = Ax^2 + Bxy + Cy^2 + Dx + Ey + F$ over a scatter in the (x, y)-plane. Paton [1–2] sets $A^2 + B^2 + C^2 + D^2 + E^2 + F^2 = 1$ as the constraint on this minimization, whereupon the solution vector (A, B, C, D, E, F) is a principal component of the cross-product matrix of the variables x^2, xy, y^2, x, y, 1. Biggerstaff [3], Albano [4], and Cooper and Yalabik [5] set $F = 1$ as the constraint, whereupon there arises a set of normal equations for A, B, C, D, E in terms of the same cross-product matrix. Gnanadesikan [6] sets $A^2 + B^2 + C^2 + D^2 + E^2 = 1$, yielding for extremum the last principal component of the variance–covariance matrix of x^2, xy, y^2, x, y. All these authors loosely characterize the optimal conic as "best-fitting" of all the conics of the plane, but the nature of the fit each is optimizing, whose computation is so surprisingly straightforward, has not been elucidated.

We seek an estimation rule which is *general*, *simple to compute*, and *invariant*. Generality requires a single algorithm by which we fit both ellipses and hyperbolas, with circles and parabolas included as constrained special cases. For instance, we cannot generalize the fitting of parabolas usual in quadratic regression, in which error-of-fit is measured parallel to the axis of the parabola, because the

0146-664X/79/010056-16$02.00/0

general conic has two axes, two directions through every point parallel to the axes, and in each direction zero, one, or two points of intersection with the conic.

Simplicity suggests that we eschew "nonlinear," iterative techniques whenever possible, and instead settle upon algorithms whose estimates can be computed from the data by finite arithmetic formulas or by the standard inversions and eigenextractions of matrix calculus. The solution proposed here is computable in this way; several alternatives, such as the minimization in mean square of distance perpendicular to the conic, involve expressions in roots of bilinear combinations in the data and the unknowns, and cannot be optimized in closed algebraic form.

Invariance is to be with respect to the equiform group of transformations of the Euclidean plane: rotations, translations, and changes of scale. If the coordinates of the data are transformed, point by point, according to an element of this group, then the best-fitting conic to the new scatter must be exactly the result of the same transformation applied to the conic which was best-fitting to the original scatter. The estimation rule proposed here minimizes a quadratic form in the unknown conic parameters subject to an exact constraint on the value of another quadratic form. It will be shown that both these forms are invariants under the action of the equiform group, and hence so is the net estimation procedure. To insist on this invariance is to alter our intuitive notion of curve-fitting "to a dependent variable." When abscissa and ordinate represent an arbitrary coordinate system, or arbitrary orientation, placed upon an underlying geometrical object, such as the outline of a skull, the axes separately have no geometric meaning at all; they are wholly commensurate and may be rotated freely. We must get the same geometric result in all cases.

2. GEOMETRIC INTERPRETATION OF $Q(x, y)$

Let a given central conic section have equation $Q(x, y) = Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$. Let (x_0, y_0) be an arbitrary point of the plane, and set (x_1, y_1) equal to the point on the conic on the ray from the center through (x_0, y_0).

Choose a new coordinate system, denoted by primes, which is centered at the center of the conic and aligned with its principal axes. Define the form $Q'(x', y')$ on the new coordinates so that $Q'(x', y') = Q(x, y)$ on the plane (x', y'). In this new coordinate system, $Q'(x' y') = A'x'^2 + B'y'^2 + F'$ for some A', B', F'.

In the primed system, the center is the origin, $(0', 0')$. By collinearity of (x_0, y_0), (x_1, y_1), and the center, expressed in the primed coordinate system, we have $x'_1/y'_1 = x'_0/y'_0$ or $x'_0 = x'_1(y'_0/y'_1)$. Then

$$\begin{aligned} Q(x_0, y_0) &= Q'(x'_0, y'_0) = A'x'^2_0 + B'y'^2_0 + F' \\ &= (y'_0/y'_1)^2(A'x'^2_1 + B'y'^2_1) + F' \\ &= (y'_0/y'_1)^2(-F') + F' \\ &= -F'[(y'_0/y'_1)^2 - 1], \end{aligned} \tag{1}$$

since (x'_1, y'_1) is on the conic $Q'(x', y') = 0$. But the term (y'_0/y'_1) is just the ratio of distances center-to-point and center-to-conic along the ray we are using—the

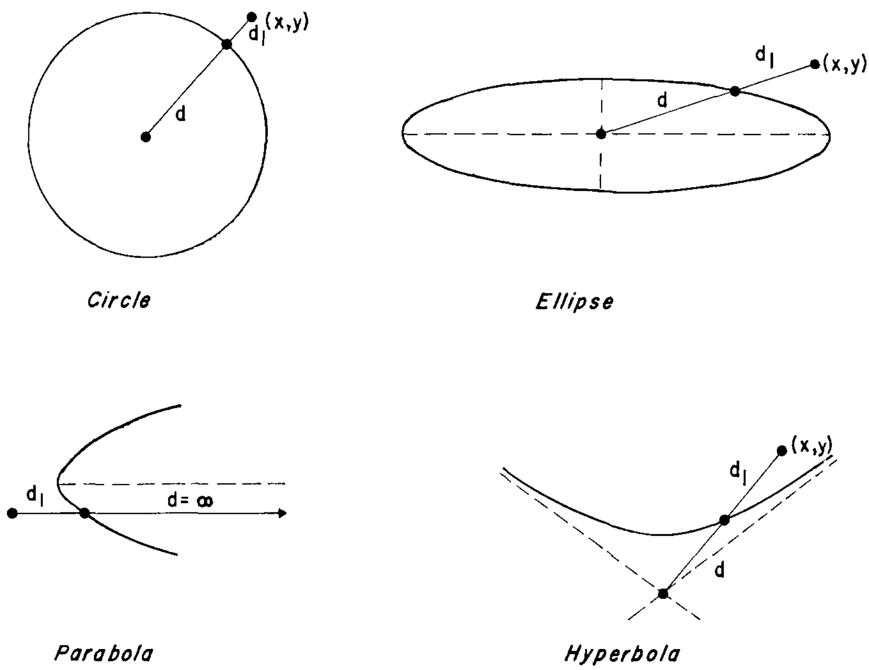

FIG. 1. The geometry of error-of-fit around a conic section, as measured by the equation of the conic. The error is always proportional to $[(d + d_1)^2/d^2] - 1$.

ratio of $d + d_1$ to d in Fig. 1; for projection onto any axis preserves distance along a line.

Hence $Q(x_0, y_0) \propto [(d + d_1)^2/d^2] - 1$, and as the equiform transformations do not alter distance ratios they preserve this proportionality (though they may alter F').

We have $[(d + d_1)^2 - d^2]/d^2 = d_1(d_1 + 2d)/d^2$. When the data are quite close to a conic, d_1 is small, $d_1 + 2d$ is approximately equal to $2d$, and $Q(x_0, y_0)$ is approximately proportional to d_1/d, the distance from the conic along a line through the conic's center, measured in units of distance from the center to the conic in that direction.

The fitting of a parabola is a limiting case, exactly transitional between ellipse and hyperbola. As the center of an ellipse moves off toward infinity while its major axis and the curvature of one end are held constant, the rays from its center become parallel, and the difference of squared distances which is the numerator of Q becomes linear in distance. The quantity being minimized in the fitting then approaches the usual deviation, Euclidean distance from the curve measured parallel to the axis.

3. ESTIMATION FOR A CIRCLE

For a circle, I propose using the instance of least mean-squared Q. Let there be given points $(x_1, y_1), \ldots, (x_n, y_n)$ and let the sample linear regression of $x^2 + y^2$ on $x, y, 1$ be $x^2 + y^2 \sim 2ax + 2by + c$, where a, b, c are the least-squares estimates. Then the circle centered at (a, b) of radius $(c + a^2 + b^2)^{\frac{1}{2}}$ is best-fitting when error is measured by $Q(x, y)$ as defined in the previous section.

Proof. For a circle, $Q(x, y) = (x - a)^2 + (y - b)^2 - R^2 = x^2 + y^2 - 2ax - 2by - (R^2 - a^2 - b^2)$, where R is its radius. As the relation between the parameters (a, b, R) of Q and the parameters $(2a, 2b, R^2 - a^2 - b^2)$ of the regression is one-to-one, we can minimize error sum-of-squares with respect to one set of parameters by minimizing it with respect to the other. But minimization with respect to the second set corresponds to the ordinary least-squares regression of $x^2 + y^2$ on x, y, 1, as asserted.

In this special case, the minimand in the fitting has a geometric meaning familiar from elementary circle geometry. For any circle C of center (a, b) and radius R, the *power* of the point (x_i, y_i) *with respect to* C is the value $(x_i - a)^2 + (y_i - b)^2 - R^2$, which is our Q. The circle itself is then the set of points of power zero. If (x_i, y_i) is outside the circle, its power is the squared length of the distance to the circle along either tangent to the circle through (x_i, y_i); if (x_i, y_i) is inside, the power still equals the squared tangent length, though the tangent is now an imaginary line and the power is negative. In both cases, the power is the product of the two extreme distances to the circle, the two distances along the diameter through (x_i, y_i). For any scatter of N points (x_i, y_i) we can compute the circle of least mean-squared power, that which minimizes the appropriate sum of squares of squared tangent distances. This is the same as the circle of least mean-squared Q discussed above.

4. INVARIANT FOR THE GENERAL CONIC

For an arbitrary conic, the generalization of the error-of-fit $(x - a)^2 + (y - b)^2 - R^2$ is clearly the value $Q(x, y) = Ax^2 + Bxy + Cy^2 + Dx + Ey + F$. We need a normalization of these coefficients, for the conic described by $Q(x, y) = 0$ is the same as that described by $kQ = kAx^2 + \ldots + kF = 0$ for any k, though the latter has error sum-of-squares k^2 times that of Q, and so can be as small as we wish. The norm we seek should be a quadratic form in A, B, C, D, E, F for simplicity of computation. It must be positive-definite, or else certain conics, those for which the form happened to be exactly zero, could never be fitted at all, even if the data lay exactly upon them. It cannot involve D, E, or F, for these are functions of the origin of coordinates and become arbitrarily large when the center is translated. The norm must therefore be a function of A, B, C. Now in the case of lines (cf. Pearson [7]), error-of-fit is characterized invariantly as distance perpendicular to the fitted line itself, and the line best-fitting a scatter is estimated by minimizing $\sum_{\text{data}} (ax + by + c)^2$ subject to the constraint $a^2 + b^2 = 1$.

For conics, we might try, by direct mimicry, the form $A^2 + B^2 + C^2$. Unfortunately, the hyperbola $2xy = 1$, with $A^2 + B^2 + C^2 = 4$, can be rotated into the form $x^2 - y^2 = 1$, with $A^2 + B^2 + C^2 = 2$, and so the first would automatically fit twice as badly if the data were rotated 45°; but the selected normalization must be invariant under rotation. Introductions to the algebraic theory of conics tell us that the forms $A + C$ and $B^2 - 4AC$ are invariant under the Euclidean group. The only positive-definite invariant that can be formed from these quantities is $(A + C)^2 + (B^2 - 4AC)/2 = A^2 + B^2/2 + C^2$. We may set a unique scale for Q, and thus proceed to actual minimization of ΣQ^2, by setting this norm equal to some conventional constant value. I suggest the value 2, so that the equation of a circle is in the usual form.

Given data points $(x_1, y_1), \ldots, (x_n, y_n)$ scattered about a conic, I propose to estimate the conic of best fit by minimizing $\sum_1^n [Q(x_i, y_i)]^2$ subject to the constraint $V\mathbf{D}V' = 2$, where $V = (A, B, C, D, E, F)$ is the vector of coefficients to be estimated, and $\mathbf{D}$ is the matrix diag $(1, \frac{1}{2}, 1, 0, 0, 0)$. The minimand may be written $V\mathbf{S}V'$, where $\mathbf{S}$ is the scatter matrix about 0 of the 6-vector $(x_i^2, x_iy_i, y_i^2, x_i, y_i, 1)$ as a function of the data points (x_i, y_i).

The invariance of this procedure is proved straightforwardly. If the coordinates of the data points are rotated or translated, the constraint $A^2 + B^2/2 + C^2$ is unchanged, likewise the term F' of Eq. (1), which is gotten by undoing the transformation; and the distance ratio, the other multiplicand of (1), is clearly unchanged. Then the fitting is invariant under translation and rotation. If the coordinates are rescaled by k, $(x, y) \to (x', y') = (kx, ky)$, let V_k be the vector of coefficients of the resulting best-fitting conic by this algorithm. The conic $Ax^2 + \ldots + F = 0$, after rescaling and renormalizing, becomes $Ax'^2 + Bx'y' + Cy'^2 + kDx' + kEy' + k^2F = 0$. The rescaling has replaced the matrix $\mathbf{S}$ by $\mathbf{S}_k = \mathbf{D}_k\mathbf{S}\mathbf{D}_k$ with $\mathbf{D}_k = \text{diag}\,(k^2, k^2, k^2, k, k, 1)$. Note that $\mathbf{D}_k\mathbf{D}\mathbf{D}_k = k^4\mathbf{D}$. Extremization of $V_k(\mathbf{D}_k\mathbf{S}\mathbf{D}_k)V_k'$ for $V_k\mathbf{D}V_k$ constant is equivalent to extremizing it for $V_k\mathbf{D}_k\mathbf{D}\mathbf{D}_kV_k'$ constant. Since $\mathbf{D}_k$ is nonsingular, the extremum is $V/\mathbf{D}_k$, where V is the extremum before rescaling. But $V/\mathbf{D}_k$ gives the coefficients of the old conic in the new system, for $k^2V/\mathbf{D}_k = (A, B, C, kD, kE, k^2F)$. Hence this method of fitting is invariant under equiform transformations.

As the authors cited in Section 1 all use D and E in the norms for their fitting, none of their solutions are invariant under the Euclidean group. Those which involve F, further, are not even invariant under translation: For by setting $F = 1$ it becomes impossible to fit conics through the origin ($F = 0$) at all. These authors must therefore forcibly standardize their data before conic-fitting in any of various ad hoc ways. Translation to center-of-mass and rotation to principal axes of scatter are the most common. This approach seems unsatisfactory. Data must be fit in a manner invariant under the Euclidean group, not arbitrarily constrained with respect to it.

5. ESTIMATION FOR THE GENERAL CONIC

We wish to minimize $V\mathbf{S}V'$ subject to $V\mathbf{D}V' = \text{constant}$. Partition V into $(V_1 | V_2)$, both components of length 3, and let $\mathbf{S}$ be partitioned correspondingly:

$$\mathbf{S} = \left(\begin{array}{c|c} S_{11} & S_{12} \\ \hline S_{21} & S_{22} \end{array}\right).$$

Then

$$V\mathbf{S}V' = V_1S_{11}V'_1 + 2V_1S_{12}V'_2 + V_2S_{22}V'_2.$$

We must minimize this subject to $V_1D_1V'_1 = \text{constant}$, where $D_1 = \text{diag}\,(1, \frac{1}{2}, 1)$.

For any fixed V_1, $V\mathbf{S}V'$ is minimal when

$$d(V\mathbf{S}V')/dV_2 = 2V_1S_{12} + 2V_2S_{22} = 0,$$

which implies

$$V_2 = -V_1S_{12}S_{22}^{-1}.$$

Then

$$V\mathbf{S}V' = V_1(S_{11} - S_{12}S_{22}^{-1}S_{21})V'_1 = V_1S_{11.2}V'_1.$$

To minimize this for $V_1D_1V'_1$ = constant, let λ be a Lagrangian multiplier for the constraint. Then we must set to zero the derivative with respect to V_1 of $V_1S_{11.2}V'_1 - \lambda V_1D_1V'_1$. This yields

$$2S_{11.2}V'_1 - 2\lambda D_1V'_1 = 0,$$

so that λ is a relative eigenvalue of $S_{11.2}$ with respect to D_1—that is to say, a solution of $|S_{11.2} - \lambda D_1| = 0$—and V_1 is the corresponding eigenvector. The eigenvector of best geometric fit is the one we want, usually (but not always) corresponding to smallest λ.

The matrix $S_{11.2}$ is the covariance matrix of the residuals of x^2, xy, y^2 after regression on x, y, 1; and the vector $V = (V_1|V_2) = V_1\ (I| - S_{12}S_{22}^{-1})$ corresponds to residualization of the basis of the V_1-space, e.g., subtraction of the part of each of x^2, xy, y^2 which is linearly predicted by x, y, 1. Geometrically, in the space of V we find the subspace completely orthogonal to the null space of $\mathbf{D}$—this is the subspace of residuals of the second-order terms on those of lower order—and perform our optimization within that subspace. For circles, with the coefficients of the quadratic part constrained to be proportional to $(1, 0, 1)$, we get just the residual itself of $x^2 + y^2$ on x, y, 1, as we derived previously. The solution for the arbitrary conic thus directly generalizes the analysis for circles.

6. LINEAR CONSTRAINTS

This computation remains tractable when we place arbitrary linear constraints upon the parameters. If V is used to denote the 6-vector of coefficients we are estimating, and a set of constraints is written together in the form $V\mathbf{M} = 0$, where $\mathbf{M}$ is a matrix, then Rao [8, Section 1c.6] instructs us in this formula for the extremum we seek: The vector V of smallest $V\mathbf{S}V'$ subject to the constraints $V\mathbf{D}V'$ = constant, $V\mathbf{M} = 0$ is the eigenvector of largest eigenvalue of the matrix $(I - \mathbf{M}(\mathbf{M}'\mathbf{S}^{-1}\mathbf{M})^{-}\mathbf{M}'\mathbf{S}^{-1})\ \mathbf{D}$ with respect to $\mathbf{S}$. Here the superscript minus denotes any generalized inverse. The matrix $\mathbf{S}$ must be positive definite, e.g., it must have an ordinary inverse, whereas $\mathbf{D}$ may be singular and the constraints may be redundant.

Setting linear constraints erodes the available degrees of freedom in various ways, some quite useful. Here are some possibilities:

(1) $B = 0$. The conic's principal axes are parallel to the coordinate axes.
(2) $A = C$. The conic's principal axes are parallel to $y = \pm x$.
(3) $A = C$, $B = 0$ (two constraints). The conic is a circle.
(4) $B = 0$, $C = 0$. The conic is a parabola with axis vertical.

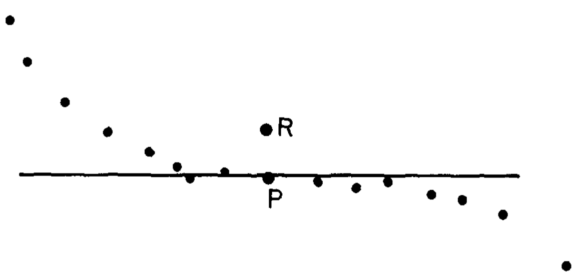

FIG. 2. Two conics abutting at a knot where the tangent crosses the curve.

(5) $2Ax_0 + By_0 + D = 0$, $Bx_0 + 2Cy_0 + E = 0$. The conic has its center at (x_0, y_0).

(6) $Ax_1^2 + Bx_1y_1 + Cy_1^2 + Dx_1 + Ey_1 + F = 0$. The conic goes through (x_1, y_1).

(7) $(2Ax_1 + By_1 + D) \cos \alpha + (Bx_1 + 2Cy_1 + E) \sin \alpha = 0$. If the point (x_1, y_1) is on the conic, this is the criterion that the tangent there make angle α with the positive x axis.

(8) $A(-x_1 \sin \alpha + R \cos^2 \alpha) + B(-0.5y_1 \sin^2 \alpha + 0.5x_1 \cos^2 \alpha + R \cos \alpha \sin \alpha) + C(y_1 \cos \alpha + R \sin^2 \alpha) + D(-0.5 \sin \alpha) + E(0.5 \cos \alpha) = 0$. The conic through (x_1, y_1) with tangent making an angle α with the positive x axis is by this constraint required to have radius of curvature there equal to R.

7. CONIC SPLINING

There is a rapidly growing literature (cf. Poirier [9], Wold [10]) on polynomial spline regression. In this technique, the expected value of a dependent variable is fitted by different functions—usually polynomials—over distinct ranges of an independent variable. At "knots," where these ranges abut, the predictions are required to agree in value and in their first few derivatives. After the knots are set, all the polynomials are computed at once, by minimizing the total squared error-of-fit about all of them in one clever dummy-variable regression. The method becomes much more complex when error-of-fit is not always measured along a fixed axis. The constraints which force the curves to line up become nonlinear in the coefficients being estimated, and linear methods fail (cf. Thomas [11] and references therein). In particular, one cannot spline a closed curve linearly by polynomials. But just as parabolas (and higher-order polynomial graphs) generalize straight-line regression, so do conics (and higher-order algebraic curves) generalize straight-line curve-fitting in general orientation. It is then of interest to explore the possibility of fitting multiple conics to separate parts of a curve simultaneously in a single linear computation generalizing the previous discussion for fitting conics singly.

Suppose we have a data set in the plane which looks like two conics, in the manner of Fig. 2, impinging upon each other at some visible nexus P. It would be nice to fit two conic segments simultaneously, one to the left of P, one to the right, which minimize the net error-of-fit. We would like the conics to connect at P without any corner, that is, to have the same tangent there. The inclination of this mutual tangent is not given in the data, however; we must estimate it, too, as that position of the tangent line at P for which the best two conics through P with that tangent, one for each side, together have least mean-squared error-of-fit. To execute this plan we must link the normalizations of the two conics. The best way to do this is by insisting that a given error-of-fit near P be the same whichever conic we consider it on. To a linear approximation, the error for a point such as R relative to a conic $Q(x, y) = 0$ through P is equal to the gradient of the scalar field $Q(x, y)$ times the distance from P to R normal to the curve. These will be equal for the two conics impinging at P if and only if the gradients of the two scalar fields are equal at P.

But to specify that these two gradients be equal is just a pair of linear constraints, for the gradients are linear forms in the coefficients of the conics. Specifically: If the two conics are $Q_i(x, y) = A_ix^2 + B_ixy + C_iy^2 + D_ix + E_iy + F = 0$, $i = 1, 2$, and if P have the coordinates (r, s), then the two consistency criteria are

$$2A_1r + B_1s + D_1 - (2A_2r + B_2s + D_2) = 0,$$

$$2C_1s + B_1r + E_1 - (2C_2s + B_2r + E_2) = 0.$$

We must also insist that both conics actually pass through the point (r, s) at which we are specifying their tangents, and this is two more constraints:

$$A_ir^2 + B_irs + C_is^2 + D_ir + E_is + F_i = 0, \qquad i = 1, 2.$$

We can estimate both conics, all 12 coefficients, simultaneously by minimizing their total squared error-of-fit subject to the normalization $A_1^2 + B_1^2/2 + C_1^2 = 2$ and the four constraints above.

This tactic is easily generalized for the fitting of conic splines to extended chains, even closed curves. Let there be I arcs, and on the ith arc data points (x_{ij}, y_{ij}), $j = 1, \ldots, n_i$, $i = 1, \ldots, I$. Let the arcs abut at knots (x_i, y_i), $i = 1, \ldots, I$, and set $(x_{I+1}, y_{I+1}) = (x_1, y_1)$. Let S_i be the 6×6 scatter matrix about 0 for the points $(x_{ij}^2, x_{ij}y_{ij}, y_{ij}^2, x_{ij}, y_{ij}, 1)$ derived from the data points (x_{ij}, y_{ij}) of the ith arc. Define $\mathbf{S}$, $(6I) \times (6I)$, equal to blockdiag $(S_1, \ldots, S_I)$. Consider a conic spline which is equal to $A_ix^2 + B_ixy + C_iy^2 + D_ix + E_iy + F_i$ on the ith arc, $i = 1, \ldots, I$, and let $A_{I+1} = A_1, \ldots, F_{I+1} = F_1$. The error sum-of-squares about the ith arc for the data of the ith arc is $(A_i, B_i, C_i, D_i, E_i, F_i)S_i(A_i, B_i, C_i, D_i, E_i, F_i)'$. Define $V = (A_1, \ldots, F_1, A_2, \ldots, F_{I-1}, A_I, \ldots, F_I)$, a $(6I)$-vector. Then the total sum-of-squares for error-of-fit about all the arcs is $V\mathbf{S}V'$. This may be minimized subject to a set of homogeneous linear constraints $V\mathbf{M} = 0$ by the formula of Rao quoted above. In this application, $\mathbf{S}$ must be nonsingular; then each S_i must be nonsingular, which implies that $n_i > 5$, all i, and that the points of all arcs must lie in "general position" (not all upon one conic).

We may minimize $V\mathbf{S}V'$ subject to up to $4I$ constraints of incidence and tangency:

$$A_ix_j^2 + B_ix_jy_j + C_iy_j^2 + D_ix_j + E_iy_j + F_i = 0, \qquad i = 1, \ldots, I, j = i, i + 1;$$

$$2A_kx_k + B_ky_k + D_k - (2A_{k-1}x_k + B_{k-1}y_k + D_{k-1}) = 0, \qquad k = 1, \ldots, I;$$

$$2C_ky_k + B_kx_k + E_k - (2C_{k-1}y_k + B_{k-1}x_k + E_{k-1}) = 0, \qquad k = 1, \ldots, I;$$

and also subject to a normalization. As the tangency constraints explicitly link normalizations of one conic to the next, we can only assign one normalizing constraint for any subarc chain of continuously turning tangent. In the usual case where this subarc is the whole spline, as in Figs. 3, 5, 6, and 7, this means that any constraint $A_i^2 + B_i^2/2 + C_i^2$ = constant determines the normalization of all the arcs all the way around.

8. WORKED EXAMPLE

The vaguely biological outline, Fig. 3a, derives from a freehand sketch on graph paper. I chose three "landmarks": No. 1, the "corner" in lower center; No. 2, the extreme right point; and No. 3, the extreme top point. The word "landmark" is used here, in preference to the "knot" of the splining literature, to suggest determination by anatomical criteria rather than asymptotic optimization tactics. If this were a skull X ray, No. 1 might be the auditory meatus, No. 2 the bridge of the nose, and No. 3 the bregma. Points between the landmarks have been digitized crudely by eye.

The spline arcs are computed by a program executing the following algorithm.

(a) The data are read in arc by arc and the scatter matrix **S** accumulated and stored. A vector of 18 unknown coefficients $V = (A_1, \ldots, F_1, \ldots, A_3, \ldots, F_3)$ is allocated, bearing the conic coefficients arc-wise.

(b) The constraints are written out in full. Arcs 3 and 1 pass through point 1 of coordinates $(0, -3)$. This requires

$$A_i \cdot 0 + B_i \cdot 0 + C_i \cdot 9 + D_i \cdot 0 + E_i \cdot (-3) + F_i = 0, \qquad i = 3, 1,$$

or

$$9C_3 - 3E_3 + F_3 = 0, \tag{1}$$

$$9C_1 - 3E_1 + F_1 = 0. \tag{2}$$

Similarly, arcs 1 and 2 pass through point 2 = (8.5, 0), requiring

$$72.25A_1 + 8.5D_1 + F_1 = 0, \tag{3}$$

$$72.25A_2 + 8.5D_2 + F_2 = 0; \tag{4}$$

and for point 3 = $(-2, 5.7)$,

$$4A_2 - 11.4B_2 + 32.49C_2 - 2D_2 + 5.7E_2 + F_2 = 0, \tag{5}$$

$$4A_3 - 11.4B_3 + 32.49C_3 - 2D_3 + 5.7E_3 + F_3 = 0. \tag{6}$$

At landmark 1 we allow the arcs to intersect at an arbitrary angle, and there are no further constraints to be had here. At landmark 2, we pick up a seventh constraint and an eighth from componentwise equality of the gradients of the first conic and the second:

$$17A_1 + D_1 - 17A_2 - D_2 = 0, \tag{7}$$

$$8.5B_1 + E_1 - 8.5B_2 - E_2 = 0. \tag{8}$$

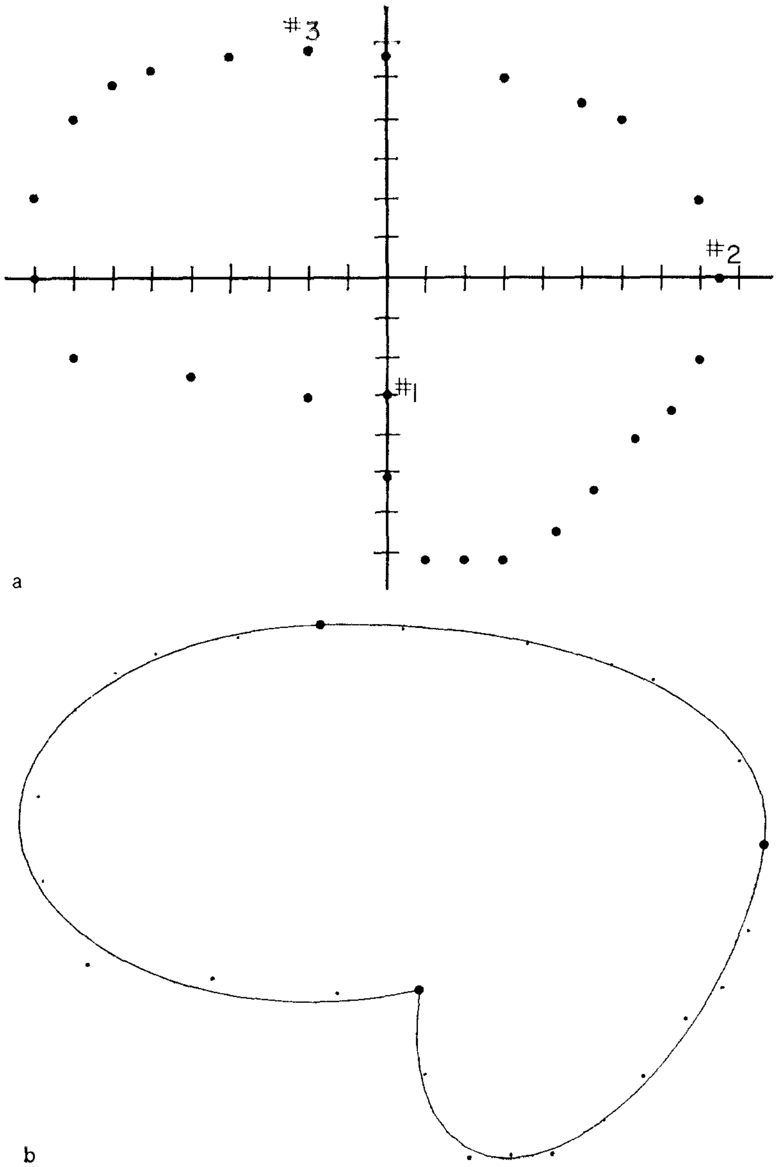

Fig. 3. (a) Hypothetical data set. (b) A three-arc conic spline approximating this scatter, with knots at the points numbered in (a).

The final two constraints derive from point 3 similarly:

$$-4A_2 + 5.7B_2 + D_2 + 4A_3 - 5.7B_3 - D_3 = 0, \tag{9}$$

$$11.4C_2 - 2B_2 + E_2 - 11.4C_3 + 2B_3 - E_3 = 0. \tag{10}$$

(c) For any vector V of coefficients for any spline, the total error-of-fit is the scalar $V\mathbf{S}V'$, where $\mathbf{S}$ is the cross-product matrix accumulated in (a). We need to minimize this for fixed invariant norm $A_1^2 + B_1^2/2 + C_1^2 = V\mathbf{D}V'$, $D = \text{diag}\ (1, \frac{1}{2}, 1, 0, \ldots, 0)$, subject to constraints (1) through (10). Conversely, we can instead maximize $V\mathbf{D}V'$ for fixed error $V\mathbf{S}V'$ subject to the same 10 constraints. The theorem of Rao quoted previously applied verbatim here to provide a computing formula for the solution we seek. A single eigenvector of a certain large matrix manifests the coefficients of all three conics, in blocks of six, for its loadings. The spline corresponding to this eigenvector exactly passes through all the knots with well-defined tangents there and minimizes the total error-of-fit from the conics in between. This curve is conic between every pair of consecutive landmarks, and its tangent turns continuously throughout, except where we have expressly omitted the constraint that it do so.

(d) These arcs are drawn out between appropriate endpoints (by short segmental approximations) and then superimposed over the data in Fig. 3b. The spline fits the data quite well except on the far left, where the conic arc has only one curvature maximum whereas the data apparently have two (the northwest and southwest "corners"). This systematic deviation can, if we wish, be declared a "local feature" of the outline relative to the spline, as the nose is a local feature relative to the plane of the face; it is not necessarily a flaw in the fitting. The splining-with-landmarks replaces the observed (digitized) boundary with smooth conventional arcs. Its tangent and normal, turning smoothly, provide a moving coordinate system to relate deviations from the "standard" in different regions all around the curve. The spline is more elegant and more easily read than a detailed tracing, as local features do not disrupt the global aspects of the shape, the ebb and flow of curvature around the form.

Should a landmark, say the second, be itself known only with error, one can free the conics from passing exactly through it. The revised equations would require, instead of Eqs. (3), (4), (7), and (8), only that the landmark have a common distance from the two conics and the same polar with respect to them. The appropriate constraint equation replaces (3) and (4) with the equality of their left-hand sides:

$$72.25A_1 + 8.5D_1 + F_1 - 72.25A_2 - 8.5D_2 - F_2 = 0;$$

Eqs. (7) and (8) persist unchanged. With this revision, the landmarks serve only a housekeeping function as knots, and do not serve any special role as data themselves.

9. RECTIFIABLE CURVES AND MORPHOMETRIC DATA ANALYSIS

The technique just presented applies directly to a nagging problem in morphometrics: the function representation of outline. In morphometrics today it is the custom to measure shape by methods which are indirect. Landmarks (reliably locatable points, "places to point to") may be located in images, and distances and angles among them measured as variables; or a center of coordinates may be set and Cartesian or polar coordinates for the curve extracted and condensed into moments or Fourier coefficients. In these procedures the underlying physical object, which I presume to be a biological outline, is never actually given a

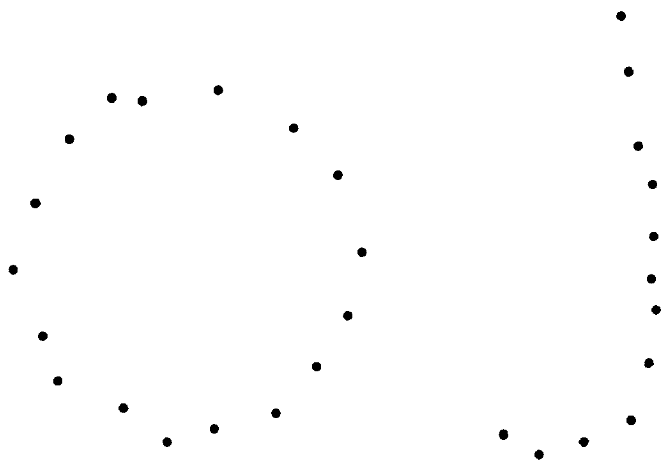

FIG. 4. Typical data for outline-fitting: functional relationships of unknown structure. Empirical arc-length does not exist.

geometric expression free of an imposed coordinate system. Then all these methods study a measurement vector, not the outline. A critical review of the scattered literature on this subject may be found in Bookstein [12, Chaps. II, III].

There is an alternative to be had from the general theory of plane curves. The curvature of a smooth plane curve is the rate of change of tangent direction with respect to arc-length, the reciprocal of the radius of the "osculating circle," the circle making closest contact with the curve at a point. From curvature as an explicit function of arc-length, measured from any origin on the curve, the whole curve can be reconstructed without error. Then a curve is wholly represented by

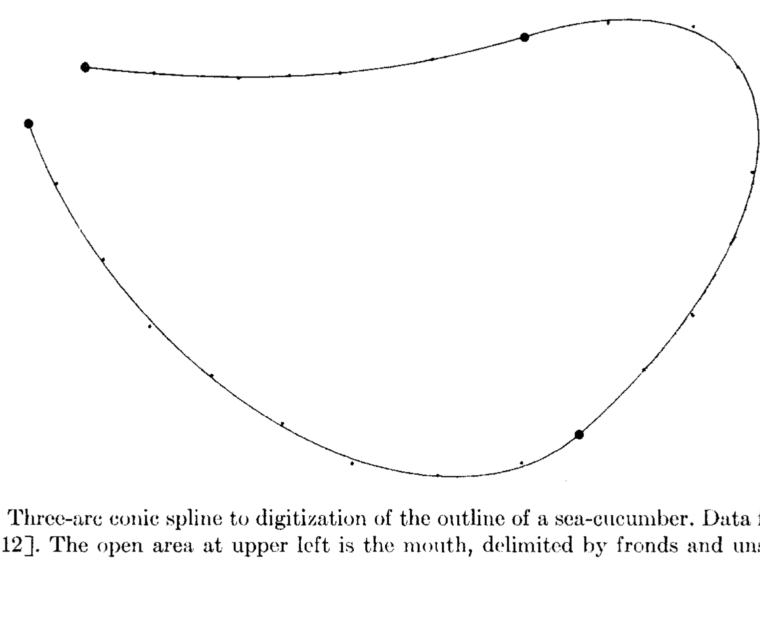

FIG. 5. Three-arc conic spline to digitization of the outline of a sea-cucumber. Data from Storer [16, p. 412]. The open area at upper left is the mouth, delimited by fronds and unsuitable for splining.

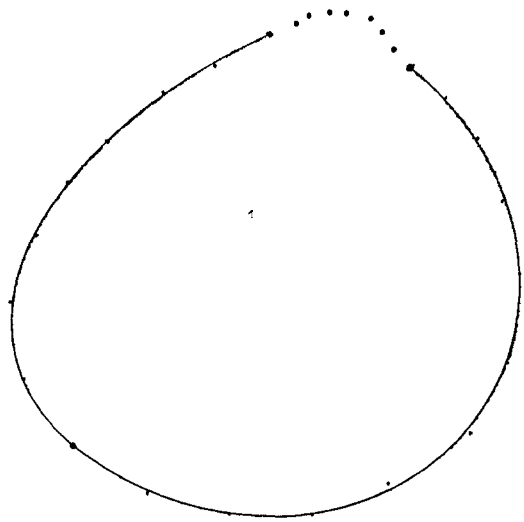

FIG. 6. Two-arc conic spline to digitization of the outline of a pelecypod, *Pisidium compressum*. Data from Burch [17, p. 14]. The open area at the top is a projection of the hinge obscuring the line of closure.

its curvature as a function of arc-length, the *intrinsic* representation (Stoker [13, Chap. II]). It does not matter at which point of the Cartesian plane the integration begins or in which direction the tangent line initially points; vary these and the intrinsic function draws out curves all perfectly congruent.

Unfortunately, such a function is never given to us in the course of empirical research. We have, rather, selected points upon such a curve, in the manner of Fig. 4. Before we can proceed at all, these isolated points must be combined into a smooth curve somehow. It is clear that we cannot connect the given points with straight lines, for the resulting function has its tangent angle undefined at exactly the points, now vertices of a polygon, at which we need to sample it. In case the data were chain-coded, the situation is even worse, for every azimuth is one of eight values. Nor can we say that the data approximate to a "true" curve, that with greater and greater density of points sampled we would converge to a limiting "true" locus. At the limit, the tangent angle is not a continuous function of arc-length, for arc-length does not exist. The greater the detail in which one examines any real locus, natural or fabricated, the more irregular it appears, the longer the arc-length, and the denser its local extrema of curvature. See Steinhaus [14] and Mandelbrot [15] for a discussion of this remarkably general phenomenon, that arc-length generally increases exponentially with the scale of observation. This paradox is not peculiar to biometrics; it complicates also the study of seacoasts, galaxies, turbulence, evolutionary histories, and soap.

Then to utilize the intrinsic representation theorem in data analysis it is necessary to construct a rectifiable curve fitting the data before we proceed. We could, for instance, pass a suitably smooth curve through every point of the digitized outline, thus enshrining all measurement error on a par with form itself. Such curves would tend to be a good deal longer than what we intuitively feel is

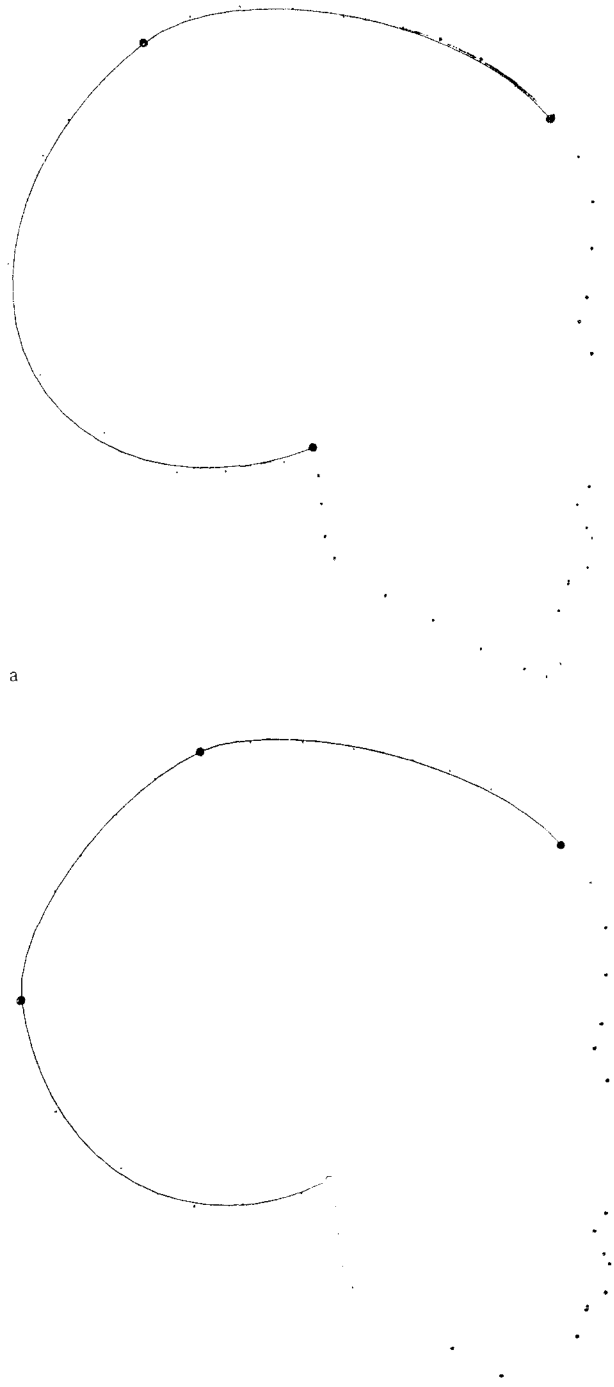

FIG. 7. (a) Two-arc, and (b) three-arc splines to the calvarium of a 16-year-old girl. The data points were selected from the digitization of a cephalogram in the files of the Biometrics Laboratory, Dental Research Institute, University of Michigan.

the appropriate "length" of the outline. We could contrariwise try to capture the form by a "best-fitting" specimen of some family of simple mathematical curves. These families, however, would have to have a number of parameters which is some multiple of the count of "pieces" of outline (arcs between landmarks) to be satisfactorily fit. They are most simply generated as splines such as we have been discussing: linked series of curves each of which best fits the data in its vicinity and passes smoothly into its neighbors.

Then the value of conic splining is that it supplies a *substitute curve*, easily fitted closely to the data, with a curvature computable simply from the coefficients of the arcs and with arc-length equal to a familiar elliptic integral offered in the various collections of scientific subroutines. Regular local features of shape—indentations inside the spline or bumps outside—become new shapes which can be analyzed further. The spline captures the "general shape," and provides an arc-length argument for the local features, in a wholly coordinate-free way. Measures of this general shape can then be derived from inspection of the splined intrinsic function and its variation over sets of shapes, without any ratios, distances, or other "shape variables" having been specified in advance.

I forsee two statistical methods by which these functions might be analyzed. In one, a multivariate measurement vector may be constructed by sampling the intrinsic function directly at finitely many points, corresponding to either actual landmark locations or aliquots of arc-length. Then the full information about the curving outline is preserved until the last stages of the analysis, whereas it is lost early if one samples isolated points from the outline. I suggest a specific statistical algorithm in Bookstein [12, Chap. IV]. In the other, the spline coefficient vector V is used itself as the measurement vector. In an I-arc spline with one "corner," V exists in projective $(2I)$-space. Means, group contrasts, and the like can be examined there by multivariate analyses upon the cross-product matrix of direction cosines of V.

These techniques should apply whenever the derivative of curvature has few changes of sign over the outline of a form. Figures 5, 6, and 7 are examples of fits to suitable biological data. The first of these is a sea-cucumber outline that I digitized from a 3- by 5-cm image in an old zoology text. It can be seen how closely a three-arc conic can be made to follow this creature of very odd shape, a shape, in fact, describable only by reference to the smooth spline. The next figure is of a clam shell. The fit of two elliptical arcs to this natural form is close enough that the intrinsic function should differentiate among growth stages of this organism and among congeneric adults, by shell outline alone (should anyone care to digitize more of the images).

Figure 7 is of more general interest. Of particular irksomeness in craniometrics is the absence of identifiable landmarks over the vault of the human skull. It is very difficult to describe that form having only quantities (distances, angles, ratios) of rather large reach. Figure 7 shows that an arbitrary specimen of this form is quite closely fit by either two or three conic arcs splined together—the optional third arc makes it possible to represent the little bump above the inion, at the far left. There is then no need for landmarks at all in this part of the skull, but the whole shape itself, as smoothly reconstructed by the spline, may be used as a function-valued or coefficient-valued measure in analysis of growth patterns, racial differences, evolution of the hominid form, and the like.

REFERENCES

1. K. A. Paton, Conic sections in chromosome analysis, *Pattern Recog.* **2**, 1970, 39.
2. K. A. Paton, Conic sections in automatic chromosome analysis, *Machine Intelligence* **5**, 1970, 411.
3. R. H. Biggerstaff, Three variations in dental arch form estimated by a quadratic equation, *J. Dent. Res.* **51**, 1972, 1509.
4. A. Albano, Representation of digitized contours in terms of conic arcs and straight-line segments, *Computer Graphics and Image Processing* **3**, 1974, 23.
5. D. R. Cooper and N. Yalabik, On the computational cost of approximating and recognizing noise-perturbed straight lines and quadratic arcs in the plane, *IEEE Trans. Computers* **C-25**, 1976, 1020.
6. R. Gnanadesikan, *Methods for Statistical Data Analysis of Multivariate Observations*, Wiley, New York, 1977.
7. K. Pearson, On lines and planes of closest fit to systems of points in space, *Philos. Mag. Ser. 6* **2**, 1901, 559.
8. C. R. Rao, *Linear Statistical Inference and Its Applications*, 2nd ed., Wiley, New York, 1973.
9. D. J. Poirier, Piecewise regression using cubic splines, *J. Amer. Statist. Assoc.* **68**, 1973, 515.
10. S. Wold, Spline functions in data analysis, *Technomet.* **16**, 1974, 1.
11. D. T. Thomas, Pseudospline interpolation for space curves, *Math. Comp.* **30**, 1976, 58.
12. F. L. Bookstein, *The Measurement of Biological Shape and Shape Change*, Springer-Verlag, New York, 1978.
13. J. J. Stoker, *Differential Geometry*, Wiley, New York, 1969.
14. H. Steinhaus, Length, shape, and area, *Colloq. Math.* **3**, 1954, 1.
15. B. Mandelbrot, *Fractals: Form, Chance, Dimension*, Freeman, San Francisco, 1977.
16. R. Storer, *General Zoology*, McGraw–Hill, New York, 1951.
17. J. Burch, *Freshwater Sphaeriacean Clams of North America*, Malacological Publications, Hamburg, Mich., 1975.