

TOTAL LEAST-SQUARES SPIRAL CURVE FITTING

By Thomas G. Davis¹

ABSTRACT: A rapidly convergent algorithm for fitting clothoids to measured points is developed and tested. The second-order, reduced Hessian method, broadly applicable to the class of scalable, C^2 parametrizations, is orthogonal distance regression with four-parameter similarity transformations. The local parameters, or state variables, are implicitly eliminated, and second-order solutions are rigorously computed in the model parameter space (rank ≤ 4). The algorithm is further distinguished from earlier works by the inclusion of approximation procedures that yield very good starting values. Additionally, a strong connection between the Helmert transformation and the total least-squares problem is established, and a fixed point method is suggested.

INTRODUCTION

The simplest example of ordinary least-squares (OLS) curve fitting is univariate linear regression in which one variable, the independent variable, is assumed to be error-free, whereas the other (dependent) variable is assumed to be the source of all errors. The first published inquiry into a method that would treat both variables equally was made by Adcock (1878). Pearson (1901) gives a solution for the line and plane fitting problems. Both authors define the residual as an orthogonal distance rather than difference of ordinates or abscissas.

The technique is variously known as errors-in-variables modeling, generalized least squares, orthogonal distance regression, and total least squares (TLS). The last of these rubrics (TLS) was first used by Golub and Van Loan (1980) who propose a singular value decomposition (SVD) method that easily generalizes to higher dimensional problems. SVD-based algorithms have been refined and extended, notably by Van Huffel and Vandewalle (1991), and research continues (Van Huffel 1997) with many new linear TLS algorithms.

Nonlinear TLS curve-fitting problems are broadly categorized as having either an explicit model or an implicit model. In the explicit problem, the curve is defined by an algebraic relation $y = f(x)$ or a parameterization $[x(t), y(t)]$. In the implicit problem, the curve is defined by $F(x, y) = 0$.

With few exceptions, investigators of nonlinear problems develop first-order algorithms that are, in one form or another, Gauss-Newton methods (Murray 1972). Gulliksson and Söderkvist (1995) develop first- and second-order methods for both explicit and implicit problems in a very general setting of arbitrarily large weights. Their solution is, however, considerably more sophisticated (and complicated) than is warranted for the types of curves we wish to consider here.

Specifically, we are interested in the primitive curves that arise in a branch of metrology called plane coordinate geometry, as used by civil engineers and land surveyors in route design applications. In the horizontal design

¹Visiting Prof., Dept. of Civ. and Envir. Engrg., Univ. of South Florida, Tampa, FL 33620.

Note. Discussion open until April 1, 2000. To extend the closing date one month, a written request must be filed with the ASCE Manager of Journals. The manuscript for this paper was submitted for review and possible publication on November 23, 1998. This paper is part of the *Journal of Surveying Engineering*, Vol. 125, No. 4, November, 1999. ©ASCE, ISSN 0733-9453/99/0004-0159-0176/\$8.00 + \$.50 per page. Paper No. 19682.

plane, path elements are lines, circles, and Cornu spirals, the components of a clothoidal spline (Walton and Meek 1990). The unweighted, unconstrained TLS line, circle, and clothoid fitting problems are the kernel problems of coordinate geometry.

For our purposes, the line fitting problem was solved by Pearson in 1901. Karimäki (1992) developed a high-precision approximation to the TLS circle fit, and Gander et al. (1994) presented a Gauss-Newton solution. Späth (1997) used fixed point methods to fit circles and other conic sections.

Our example is the TLS clothoid fit. Stoer (1982) fit n spiral arcs to n data points using a least-squares criterion, but the topic of TLS clothoid fitting is absent from the literature.

PROBLEM

The concept of minimizing some measure of misfit is common to most curve fitting procedures. Denote the i th residual or measure of difference between the i th observation and its expectation as ρ_i . Then $\boldsymbol{\rho} = (\rho_1, \dots, \rho_n)$ is a vector of residuals, and a measure of the magnitude or length of $\boldsymbol{\rho}$ is given by the p -norm (Hager 1988)

$$\|\boldsymbol{\rho}\|_p = \left[\sum_{i=1}^n |\rho_i|^p \right]^{1/p} \quad (1)$$

Choosing $p = 1$, the 1-norm, or L_1 curve fitting problem is

$$\min \sum_{i=1}^n |\rho_i| \quad (2)$$

Choosing $p = 2$, the L_2 curve fitting problem is

$$\min \left[\sum_{i=1}^n |\rho_i|^2 \right]^{1/2} \quad (3)$$

or, equivalently

$$\min \sum_{i=1}^n \rho_i^2 \quad (4)$$

This last formulation is the familiar least-squares problem and the one that we treat here. It should be clear, however, that other norms give rise to other problems. Note also that there is considerable latitude in the way that one defines both the expectation and the residual.

Consider a plane curve with C^2 parametric representation

$$\mathbf{z}(\boldsymbol{\alpha}, t) = \begin{bmatrix} x(\boldsymbol{\alpha}, t) \\ y(\boldsymbol{\alpha}, t) \end{bmatrix} \quad (5)$$

where $\boldsymbol{\alpha} \in \mathfrak{R}^m$ = vector of model parameters; and $t \in \mathfrak{R}$ = variable of parameterization or local parameter. For example, a circle has three model parameters [radius r and location of center (x_0, y_0)] with possible parameterizations in arc length s or turned angle θ . The model parameters $\alpha_1, \dots, \alpha_m$ govern geometric characteristics including the location, orientation, and size

of the curve. The local parameter t determines the location of points along the curve.

Given a set of data points in the plane

$$\tilde{\mathbf{z}}_i = \begin{bmatrix} \tilde{x}_i \\ \tilde{y}_i \end{bmatrix}, \quad i = 1, \dots, n \quad (6)$$

associate with each point a copy of \mathbf{z} denoted by $\mathbf{z}_i = \mathbf{z}(\boldsymbol{\alpha}, t_i)$. Define the “projection” (Fig. 1) of $\tilde{\mathbf{z}}_i$ onto the curve of $\mathbf{z}_i^* = \mathbf{z}(\boldsymbol{\alpha}, t_i^*)$, i.e., with each data point $\tilde{\mathbf{z}}_i$ associate a curve point \mathbf{z}_i^* with local parameter $t_i^* = t_i(\boldsymbol{\alpha}; \tilde{\mathbf{z}}_i)$. Each t_i is a free variable; the projection determines a particular local parameter t_i^* , and thus, a particular curve point for each data point. The $\tilde{\mathbf{z}}_i$ are observations with associated expectations \mathbf{z}_i^* . Denote the 2-norm or Euclidean norm supply by $\|\cdot\|$, and define the residual $\rho_i = \|\mathbf{z}_i^* - \tilde{\mathbf{z}}_i\|$. Then the least-squares curve fitting problem is

$$\min_{\boldsymbol{\alpha}} \sum_{i=1}^n \|\mathbf{z}_i^* - \tilde{\mathbf{z}}_i\|^2 \quad (7)$$

subject to (s.t.) the constraints imposed by the chosen projection (determination of the t_i^*).

When the $\tilde{\mathbf{x}}_i$ are considered error-free, a vertical projection Fig. 1(a) is required, and (7) becomes the unconstrained OLS problem

$$\min_{\boldsymbol{\alpha}} \sum_{i=1}^n [y(\boldsymbol{\alpha}, \tilde{x}_i) - \tilde{y}_i]^2 \quad (8)$$

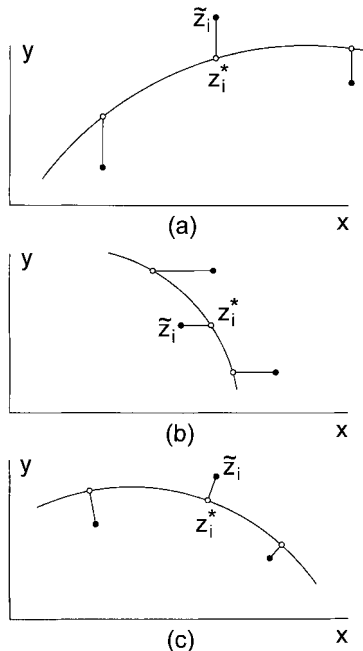


FIG. 1. Point Projections: (a) Vertical; (b) Horizontal; (c) Orthogonal

Similarly, when the \tilde{y}_i are considered error-free, a horizontal projection [Fig. 1(b)] is required.

When both the \tilde{x}_i and \tilde{y}_i are subject to errors that are independently and identically distributed with zero mean and common variance, an orthogonal projection [Fig. 1(c)] is required, and (7) becomes the TLS problem

$$\min_{\alpha} \sum_{i=1}^n \|\mathbf{z}_i^* - \tilde{\mathbf{z}}_i\|^2 \quad (9a)$$

$$\text{s.t. } \|\mathbf{z}_i^* - \tilde{\mathbf{z}}_i\|^2 = \min_{t_i} \|\mathbf{z}_i - \tilde{\mathbf{z}}_i\|^2, \quad i = 1, \dots, n \quad (9b)$$

When the constraints in (9) can be solved explicitly for t_i^* as a function of $\tilde{\mathbf{z}}_i$ and α , (9) may be solved as an unconstrained problem in m parameters $\alpha_1, \dots, \alpha_m$. The explicit elimination of t_i is not, however, always possible. Letting $f_i = \|\mathbf{z}_i - \tilde{\mathbf{z}}_i\|^2$, the constraints in (9) are a system of n minimization subproblems with first-order conditions

$$\frac{\partial f_i}{\partial t_i} = 0, \quad i = 1, \dots, n \quad (10)$$

and second-order conditions

$$\frac{\partial^2 f_i}{\partial t_i^2} > 0, \quad i = 1, \dots, n \quad (11)$$

at a solution $t_i = t_i^*$.

Now considered the unconstrained minimization problem

$$\min_{\alpha, \mathbf{t}} F \quad (12)$$

where $\mathbf{t} = (t_1, \dots, t_n)$, and

$$F = \sum_{i=1}^n f_i \quad (13)$$

Denoting the gradient of F by ∇F , the first-order condition for a solution of (12) is $\nabla F = \mathbf{0}$, which satisfies (10) because $\partial F / \partial t_i = \partial f_i / \partial t_i$. Denoting the Hessian matrix of F by $\nabla^2 F$, the second-order condition for a solution of (12) is that $\nabla^2 F$ be positive definite, which subsumes (11) because a symmetric, positive definite matrix must have positive diagonal elements and $\partial^2 F / \partial t_i^2 = \partial^2 f_i / \partial t_i^2$. So any t_i that solves (12) is equal to t_i^* , and $f_i(t_i = t_i^*) = f_i^*$. Thus (9) is equivalent to the unconstrained optimization problem in $m + n$ parameters (12).

In the next section the structure of a second-order method for the solution of (12) is examined, and through implicit variable elimination, a reduced Hessian method is developed.

STRUCTURE

The TLS problem [(12)] may be solved using Newton's method (Reklaitis et al. 1983)

$$\nabla^2 F(\xi^k) \Delta \xi = -\nabla F(\xi^k) \quad (14)$$

where F is defined by (13); ξ = column vector (α, \mathbf{t}) ; and $\Delta\xi = \xi^{k+1} - \xi^k$, so that $\xi^{k+1} = \xi^k + \Delta\xi$. This system has the form

$$\begin{bmatrix} \mathbf{H} & \mathbf{S} \\ \mathbf{S}' & \mathbf{T} \end{bmatrix}_{\xi^k} \Delta\xi = -\nabla F(\xi^k) \quad (15)$$

where \mathbf{S}' denotes the transpose of \mathbf{S} , and

$$\mathbf{H} = \frac{\partial^2 F}{\partial \alpha' \partial \alpha}; \quad \mathbf{S} = \frac{\partial^2 F}{\partial \alpha' \partial \mathbf{t}}; \quad \mathbf{T} = \frac{\partial^2 F}{\partial \mathbf{t}' \partial \mathbf{t}} \quad (16a-c)$$

The solution of (12) via (15) has a major drawback: (15) is potentially very large with rank = $m + n$. Furthermore, the convergence rate of such a system is generally slower than that of a smaller, but equivalent, system. Accordingly, we seek to implicitly eliminate the local parameters \mathbf{t} .

Each of the t_i is defined implicitly by (10) and (11) for fixed α . If $\mathbf{t}^k = \mathbf{t}(\alpha^k)$ can be computed for each Newton step by satisfying (10) s.t. (11), then the following formalisms are fully justified by the implicit (and attendant inverse) function theorem (Marsden and Tromba 1981).

Define $\xi(\alpha) = [\alpha, \mathbf{t}(\alpha)]$. Then the objective function may be written as $F(\alpha, \mathbf{t}) = F[\xi(\alpha)] = \hat{F}(\alpha)$, and implicit differentiation yields

$$\frac{\partial^2 \hat{F}}{\partial \alpha_\ell \partial \alpha_j} = \frac{\partial^2 F}{\partial \alpha_\ell \partial \alpha_j} - \sum_{i=1}^n \frac{\partial^2 F}{\partial \alpha_\ell \partial t_i} \frac{\partial^2 F}{\partial t_i \partial \alpha_j} / \frac{\partial^2 F}{\partial t_i^2}, \quad \ell = 1, \dots, m, \quad j = 1, \dots, m \quad (17)$$

and

$$\frac{\partial \hat{F}}{\partial \alpha_j} = \frac{\partial F}{\partial \alpha_j}, \quad j = 1, \dots, m \quad (18)$$

Eqs. (17) and (18) allow us to construct the second-order, reduced Hessian system

$$\hat{\mathbf{H}}(\alpha^k) \Delta\alpha = -\nabla \hat{F}(\alpha^k) \quad (19)$$

where

$$\hat{\mathbf{H}} = \frac{\partial^2 \hat{F}}{\partial \alpha' \partial \alpha} \quad (20)$$

The TLS problem (12) may be solved by iterative applications of (19). Unlike (15), the dimension of (19) does not depend on the number of data points. Moreover, the only extra work, which is more than adequately compensated by the reduced system size, is projecting the data points onto the curve at each Newton step. A geometric model suitable for the computation of (19) is presented in the next section.

MODEL

Consider a local coordinate system (u, v) embedded in the x, y -plane with origin (x_0, y_0) and orientation angle θ (Fig. 2). It is customary to simplify the expression for a plane curve by specifying the model parameters (x_0, y_0) and θ together with a parameterization in terms of u and v . Let (x, y) , (x_0, y_0) , and (u, v) be the column vectors \mathbf{z} , \mathbf{z}_0 , and \mathbf{w} , respectively. Then

$$\mathbf{z} = \mathbf{z}_0 + \mathbf{A}\mathbf{w}, \quad \mathbf{A} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \quad (21)$$

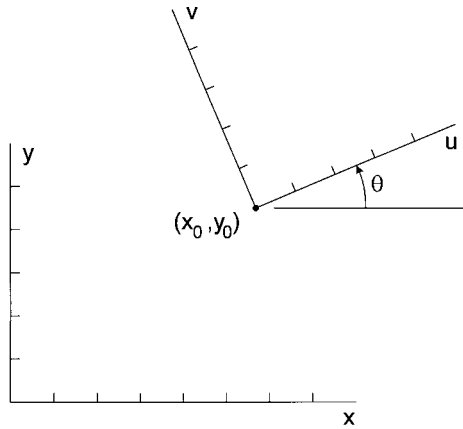


FIG. 2. Local Coordinate System

The transformation (21) is a Euclidean motion composed of a translation and a rotation. Denote the local coordinate curve or graph of \mathbf{w} by $G(\mathbf{w})$. Including a potential reflection through the u -axis and a dilation with ratio a , all possible curves similar to $G(\mathbf{w})$ are given by the graph of

$$\mathbf{z} = \mathbf{z}_0 + a\mathbf{A} \begin{bmatrix} 1 & 0 \\ 0 & \pm 1 \end{bmatrix} \mathbf{w} \quad (22)$$

At this point we restrict our attention to families of curves that can be described by a similarity transformation applied to some "unit" parameterization, i.e., families of curves described by (22) when $\mathbf{w} = [u(t), v(t)]$ is purely a function of the local parameter t . Examples include all lines, parabolas, ellipses with a fixed aspect ratio (e.g., circles), clothoids, and a number of other important curves. In short, we will treat the entire class of "scalable," C^2 curves.

Now, to evaluate (19), the data points must be projected onto the candidate curve at each Newton step. Because every candidate is similar to $G(\mathbf{w})$, it makes good sense to work in the local coordinate system and project transformed data points onto the invariant $G(\mathbf{w})$, rather than projecting invariant data points onto $G(\mathbf{z})$. Accordingly, define the inverse transformation

$$\tilde{\mathbf{w}}_i = \begin{bmatrix} \tilde{u}_i \\ \tilde{v}_i \end{bmatrix} = \frac{1}{a} \begin{bmatrix} 1 & 0 \\ 0 & \pm 1 \end{bmatrix} \mathbf{A}'(\tilde{\mathbf{z}}_i - \mathbf{z}_0) \quad (23)$$

where $\mathbf{A}' = \mathbf{A}^{-1}$ because \mathbf{A} is orthonormal. The transformation indicated by (23) is also a similarity. In this framework, we deal with a fixed curve $G(\mathbf{w})$ and a scaled, rotated, translated, and possibly reflected copy of the data points (Fig. 3).

The TLS problem (12) may be restated as

$$\min_{a, \mathbf{z}_0, \theta, t} \sum_{i=1}^n \frac{1}{2} a^2 \|\mathbf{w}_i - \tilde{\mathbf{w}}_i\|^2 \quad (24)$$

where the factor of 1/2 of the objective function is included for computational reasons. A detailed algorithm for the solution of (24) is developed in the next section.

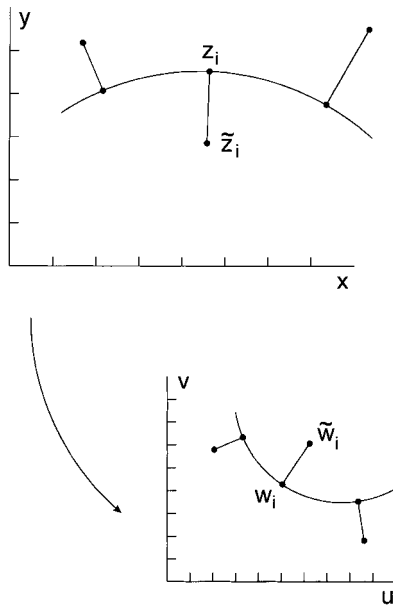


FIG. 3. Similarity Transformation

ALGORITHM

We seek an explicit recipe for the solution of (24), given an arbitrary C^2 parameterization of the archetypal curve $G(\mathbf{w})$. Define $\Delta \mathbf{z}_i = \tilde{\mathbf{z}}_i - \mathbf{z}_0$, and initially, consider the group of direct similarities

$$\tilde{\mathbf{w}}_i = \frac{1}{a} \mathbf{A}' \Delta \mathbf{z}_i \quad (25)$$

i.e., the group of similarities that can be represented without a reflection. Further define $\Delta \mathbf{w}_i = \mathbf{w}_i - \tilde{\mathbf{w}}_i$

$$F = \sum_{i=1}^n f_i = \sum_{i=1}^n \frac{a^2}{2} \|\Delta \mathbf{w}_i\|^2 = \sum_{i=1}^n \frac{a^2}{2} \Delta \mathbf{w}_i' \Delta \mathbf{w}_i \quad (26)$$

and identify a , x_0 , y_0 , and θ with α_1 , α_2 , α_3 , and α_4 , respectively. Then the solution of (24) may be affected by iterative applications of (19). In the sequel, the index i may be omitted to simplify notation.

For the purpose of program coding, define

$$\tau_1 = \tilde{u}\dot{u} + \tilde{v}\dot{v}; \quad \tau_4 = \tilde{u}\dot{v} - \tilde{v}\dot{u} \quad (27a,b)$$

$$\tau_2 = \dot{u} \cos \theta - \dot{v} \sin \theta; \quad \tau_5 = (u - \tilde{u})\dot{u} + (v - \tilde{v})\dot{v} \quad (28a,b)$$

$$\tau_3 = \dot{u} \sin \theta + \dot{v} \cos \theta; \quad \tau_6 = (u - \tilde{u})\ddot{u} + (v - \tilde{v})\ddot{v} + \dot{u}^2 + \dot{v}^2 \quad (29a,b)$$

$$\eta_{11} = \sum u^2 + v^2; \quad \eta_{21} = \sum u \cos \theta - v \sin \theta \quad (30a,b)$$

$$\eta_{41} = \sum \tilde{u}v - \tilde{v}u; \quad \eta_{31} = \sum u \sin \theta + v \cos \theta \quad (31a,b)$$

$$\eta_{44} = \sum u\tilde{u} + v\tilde{v} \quad (32)$$

In the next section we investigate a related problem that reveals yet another way to solve the TLS problem (24).

HELMERT TRANSFORMATION

Given a set of data points $\tilde{\mathbf{z}}_i, i = 1, \dots, n$, in the global coordinate system (x, y) and a corresponding set of data points $\mathbf{w}_i, i = 1, \dots, n$, in a local coordinate system (u, v) , we seek a similarity transformation that maps the local data points as closely as possible, in the least-squares sense, to the corresponding global data points. Express the similarity transportation as $\mathbf{z}_i = \mathbf{z}_0 + a\mathbf{A}\mathbf{w}_i$ or

$$\tilde{\mathbf{w}}_i = \frac{1}{a} \mathbf{A}'(\tilde{\mathbf{z}}_i - \mathbf{z}_0), \quad \mathbf{A} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \quad (45)$$

Then the Helmert problem may be stated as

$$\min_{a, \mathbf{z}_0, \theta} \sum_{i=1}^n \frac{1}{2} a^2 \|\mathbf{w}_i - \tilde{\mathbf{w}}_i\|^2 \quad (46)$$

with solution (Wolf 1968; Kahmen and Faig 1988)

$$\mathbf{z}_0 = \underline{\tilde{\mathbf{z}}} - a\mathbf{A}\underline{\mathbf{w}}; \quad \tan \theta = C/D; \quad a = \sqrt{(C^2 + D^2)} / \sum \|\underline{\mathbf{w}} - \underline{\mathbf{w}}\|^2 \quad (47a-c)$$

where the underscored variables are centroids, and

$$C = \sum (u - \underline{u})(\tilde{y} - \underline{y}) - (v - \underline{v})(\tilde{x} - \underline{x}) \quad (48a)$$

$$D = \sum (u - \underline{u})(\tilde{x} - \underline{x}) + (v - \underline{v})(\tilde{y} - \underline{y}) \quad (48b)$$

The connection between the TLS problem (24) and the Helmert problem (46) is immediately apparent. In the TLS problem, the \mathbf{w}_i are functions of the data $\tilde{\mathbf{z}}_i$ and the parameters a, \mathbf{z}_0 , and θ . In the Helmert problem, the \mathbf{w}_i are fixed. Otherwise, the problems are identical. It should be clear that a solution of the TLS problem is also a solution of the Helmert problem. Indeed, the TLS problem (24) might be called an orthogonal or continuous Helmert problem where the unit curve provides a continuum of local control points. The Helmert problem (46), in turn, might be called a discrete TLS problem where the local control points play the part of a unit "curve."

The Helmert transformation suggests a simple, fixed point method for the solution of (24):

1. Estimate the parameters a, \mathbf{z}_0 , and θ .
2. Project the transformed data $\tilde{\mathbf{w}}_i$ onto the unit curve. Perform a Helmert transformation using the data $\tilde{\mathbf{z}}_i$ as global control and the projections \mathbf{w}_i as local control. Replace the parameters with the transformation results.
3. Repeat Step 2 until (hopefully) convergence is achieved.

Unfortunately, the resulting sequence converges slowly, often requiring thousands of iterations for large residual problems and hundreds of thousands of iterations for small residual problems. One application of the Helmert transformation can, however, significantly improve the initial parameter estimate.

CLOTHOID

As an example of curve fitting that uses a parameterization in normalized arc length t , consider the clothoid or Cornu spiral. A parameterization of the unit spiral (Fig. 4) is given by (Davis and Lin 1996)

$$\mathbf{w} = \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} C(t) \\ S(t) \end{bmatrix} \quad (49)$$

where $C(t)$ and $S(t)$ are the Fresnel integrals

$$C(t) = \int_0^t \cos \varphi(\zeta) d\zeta = t \sum_{i=0}^{\infty} \frac{(-1)^i \varphi^{2i}}{(4i+1)(2i)!} \quad (50a)$$

$$S(t) = \int_0^t \sin \varphi(\zeta) d\zeta = t \sum_{i=0}^{\infty} \frac{(-1)^i \varphi^{2i+1}}{(4i+3)(2i+1)!} \quad (50b)$$

$\varphi(t) = t^2/2$ being the angle that the tangent line makes with the u -axis.

The first derivatives of \mathbf{w} are the integrands of (50)

$$\dot{u} = \cos \varphi; \quad \dot{v} = \sin \varphi \quad (51a,b)$$

and the second derivatives are

$$\ddot{u} = -t \sin \varphi; \quad \ddot{v} = t \cos \varphi \quad (52a,b)$$

Eqs. (27)–(29) may now be fully evaluated as

$$\tau_1 = \tilde{u} \cos \varphi + \tilde{v} \sin \varphi; \quad \tau_4 = \tilde{u} \sin \varphi - \tilde{v} \cos \varphi \quad (53a,b)$$

$$\tau_2 = \cos \varphi \cos \theta - \sin \varphi \sin \theta = \cos(\varphi + \theta) \quad (54)$$

$$\tau_3 = \cos \varphi \sin \theta + \sin \varphi \cos \theta = \sin(\varphi + \theta) \quad (55)$$

$$\tau_5 = (u - \tilde{u})\cos \varphi + (v - \tilde{v})\sin \varphi \quad (56)$$

and

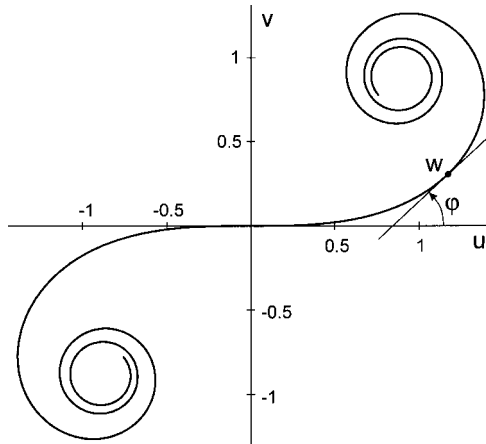


FIG. 4. Unit Spiral

$$\tau_0 = 1 + t[(v - \tilde{v})\cos \varphi - (u - \tilde{u})\sin \varphi] \quad (57)$$

Solutions may be computed by Marquardt's method (44).

Starting Values

It remains to devise estimates α^0 and t_i^0 . Assume an ordered data set $\tilde{z}_1, \dots, \tilde{z}_n, n \geq 5$. Compute two approximate circle fits (Appendix I) over the data (Fig. 5). The first circle is fitted with $\tilde{z}_1, \dots, \tilde{z}_{n/2+1}$ and yields a center C_0 and radius r_0 . The second curve is fitted with $\tilde{z}_{n/2+1}, \dots, \tilde{z}_n$ and yields a center C_1 and radius r_1 . Angle γ_0 is turned through C_0 from \tilde{z}_1 to $\tilde{z}_{n/2+1}$, and γ_1 is the angle turned through C_1 from $\tilde{z}_{n/2+1}$ to \tilde{z}_n . It is assumed that the curvature regime is compound ($\gamma_0\gamma_1 > 0$). If $r_0 < r_1$, reverse the data order and relabel the circles so that $r_0 > r_1$.

Construct the circular curves K_0 and K_1 with delta angles $\gamma_0/2$ and $\gamma_1/2$, respectively, as indicated in Fig. 6. The radii r_0 and r_1 are taken to be approximate radii of curvature at the beginning of K_0 and the end of K_1 , respectively.

The approximate spiral arc length from the beginning of K_0 to the end of K_1 is

$$\Delta s = \frac{|r_0\gamma_0| + |r_1\gamma_1|}{2} \quad (58)$$

The approximate arc length from the spiral origin to the beginning of K_0 is

$$s_0 = \frac{\Delta s r_1}{r_0 - r_1} \quad (59)$$

And the approximate arc length from the spiral origin to the end of K_1 is $s_1 = \Delta s + s_0$. An approximation of the scale factor is given by the weighted average

$$a = \frac{|r_0\gamma_0|\sqrt{(s_0r_0)} + |r_1\gamma_1|\sqrt{(s_1r_1)}}{2\Delta s} \quad (60)$$

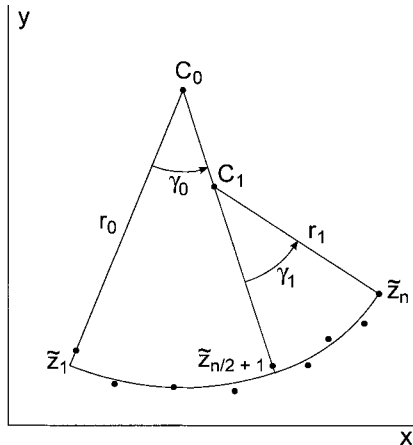


FIG. 5. Circular Approximation

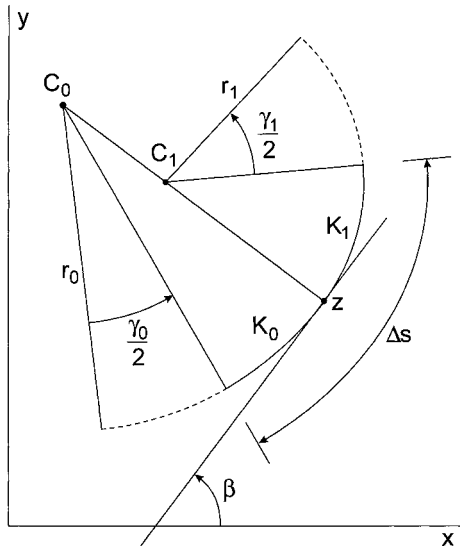


FIG. 6. Approximate Curvature and Arc Length

Now compute the average of the end point of K_0 and the start point of K_1 ; label it z . Compute the average of the counterclockwise tangent directions at the end point of K_0 and the start point of K_1 ; label it β . If $\gamma_0 < 0$, reflect the data, including z and β , through the x -axis.

Take z as a point on the approximate spiral (Fig. 7) with coordinates $z = z_0 + aAw(s/a)$ and local tangent direction $\phi = (s/a)^2/2$. Then an approxi-

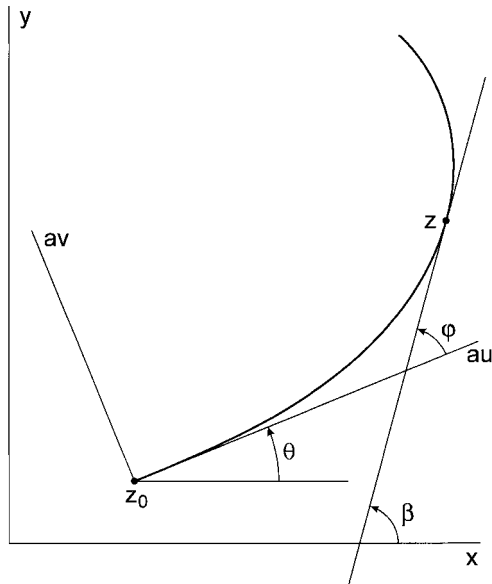


FIG. 7. Approximate Spiral

mation of the orientation angle is $\theta = \beta - \varphi$, and an approximation of the spiral origin is $\mathbf{z}_0 = \mathbf{z} - a\mathbf{A}\mathbf{w}(s/a)$, where $s = s_0 + r_0\gamma_0/2$. We now have approximate values for the model parameters a , \mathbf{z}_0 , and θ .

The first local parameter is approximated by the length of the local data vector with the sign of the local data ordinate, i.e., $t_1^0 = \text{sign}(\tilde{u}_1)\|\tilde{\mathbf{w}}_1\|$. Subsequent local parameters are approximated by $t_i^0 = t_{i-1}^0 + \|\tilde{\mathbf{w}}_i - \mathbf{w}_{i-1}\|$, $i > 1$.

To refine the model parameter estimates we project the transformed data points $\tilde{\mathbf{w}}_i$ onto the unit spiral and then compute a Helmert transformation using the data $\tilde{\mathbf{z}}_i$ as global control and the projections \mathbf{w}_i as local control. The resulting parameters are a^0 , \mathbf{z}_0^0 , and θ^0 .

Numerical Examples

The data depicted in Fig. 8 are used in numerical examples 1–3. Eleven evenly spaced points are located along a Cornu spiral with $a = 100$ m, $\mathbf{z}_0 = (0,0)$ m, $\theta = 0$ rad, and initial and terminal angles $\varphi = \pi/8$ and $\pi/2$ rad, respectively. The exact spiral coordinates are used for Example 1. For Examples 2 and 3, each of the points is distributed by a vector orthogonal to the spiral with a random magnitude from $N(0, \sigma^2)$, the normal distribution with zero mean and standard deviation σ .

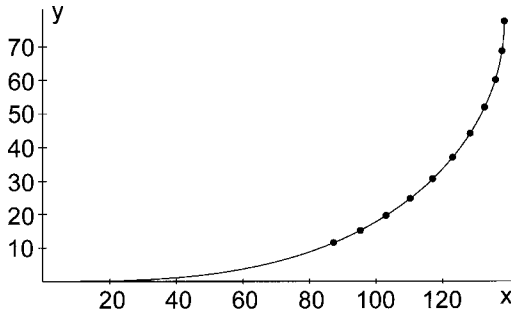


FIG. 8. Data for Numerical Examples 1–3

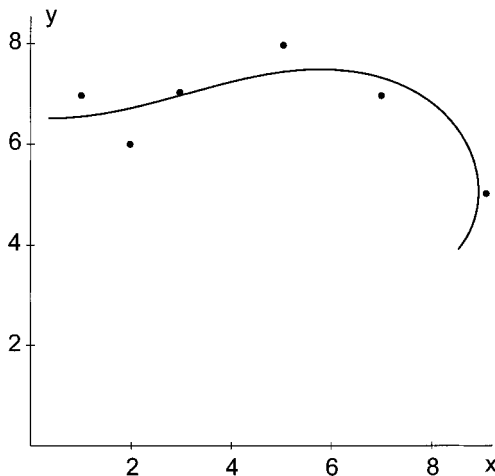


FIG. 9. Numerical Example 4

TABLE 1. Numerical Examples 1–4

Number of iterations k (1)	Scale factor a (m) (2)	Origin		Angle θ (rad) (5)	Objective function F (m^2) (6)
		x_0 (m) (3)	y_0 (m) (4)		
(a) Numerical example 1: $\sigma = 0.0$ m					
0	99.94539	-0.042569	0.153823	-0.001682	0.001225
7	100.00000	-0.000002	0.000001	0.000000	0.000000
(b) Numerical example 2: $\sigma = 0.5$ m					
0	79.17683	45.528059	-0.906662	0.246353	1.772223
11	82.71979	45.849335	-3.615592	0.295484	0.873454
(c) Numerical example 3: $\sigma = 1.0$ m					
0	67.82617	65.111738	3.869819	0.342128	7.121067
13	72.00527	68.851158	1.810261	0.441884	3.487289
(d) Numerical example 4					
0	4.59671	4.629824	6.764230	0.029989	1.681458
10	3.72732	3.104647	7.012556	0.272679	0.553825

Numerical Example 4 (Fig. 9) uses data from Gander et al. (1994). The data points are (1,7), (2,6), (3,7), (5,8), (7,7), and (9,5). The points clearly exhibit reverse curvature and are not candidates for the estimation procedure employed in Examples 1–3. We instead use the approximate values $a = 5$ m, $\mathbf{z}_0 = (5,8)$ m, and $\theta = 0$ rad; reflect the data; and perform a Helmert transformation for the starting values.

Numerical Examples 1–4 are presented in Table 1. Initial values ($k = 0$) are shown on the first line of each example. The solution and number of iterations are indicated on the second line. Results are iterated until $\|\nabla F\|/a < 10^{-8}$.

CONCLUSIONS

We have developed and tested an algorithm for the TLS spiral fitting problem. Perhaps as importantly, we have developed an exact formulation of the reduced Hessian matrix (20) as a function of the model parameters. This Hessian reduction procedure is applicable to any unconstrained TLS problem with an explicit model.

Our algorithm is essentially orthogonal distance regression with similarity transformations. The method is applicable to any C^2 curve primitive and yields the “best” fit that can be obtained through a four-parameter similarity transformation. The algorithm is a rigorous second-order method that is solved in the model parameter space and is further distinguished from more general works by the depth to which the problem is resolved. The unstudied clothoid fit completes a suite of civil engineering tools for TLS curve fitting.

The algorithm presented here has been found to be reliable and rapidly convergent for typical (small residual) problems encountered in route design applications. It is not globally convergent, nor does it anticipate every pathological data case. A more robust algorithm could be obtained by including a linesearch, but we believe that seeking better starting values would be more rewarding.

APPENDIX I. APPROXIMATE CIRCLE FIT

Assume an ordered data set $\tilde{\mathbf{z}}_1, \dots, \tilde{\mathbf{z}}_n$, $n \geq 3$, and set $\mathbf{z}_k = \tilde{\mathbf{z}}_{n/3+1}$. Karimäki (1992) describes a high-precision approximation of the center point

$$\mathbf{z}_0^0 = \mathbf{z}_k + \left(d + \frac{1}{\rho} \right) \begin{bmatrix} \sin \varphi \\ -\cos \varphi \end{bmatrix} \quad (61)$$

where, in keeping with Karimäki’s notation

$$\rho = 2\kappa/\sqrt{(1 - 4\delta\kappa)}; \quad \tan 2\varphi = 2q_1/q_2 \quad (62a,b)$$

$$d = 2\delta/[1 + \sqrt{(1 - 4\delta\kappa)}] \quad (63)$$

$$\kappa = (C_{xr^2} \sin \varphi - C_{yr^2} \cos \varphi)/C_{r^2r^2} \quad (64)$$

$$\delta = -\kappa\langle r^2 \rangle + \langle x \rangle \sin \varphi - \langle y \rangle \cos \varphi \quad (65)$$

$$q_1 = C_{r^2r^2}C_{xy} - C_{xr^2}C_{yr^2} \quad (66)$$

$$q_2 = C_{r^2r^2}(C_{xx} - C_{yy}) - C_{xr^2}^2 + C_{yr^2}^2 \quad (67)$$

Here the measurements $x_i = \tilde{x}_i - x_k$, $y_i = \tilde{y}_i - y_k$, and $r_i^2 = x_i^2 + y_i^2$ are taken in an auxiliary coordinate system with origin \mathbf{z}_k . The coefficients $C_{xx} = \langle x^2 \rangle - \langle x \rangle^2$, $C_{yy} = \langle y^2 \rangle - \langle y \rangle^2$, $C_{xy} = \langle xy \rangle - \langle x \rangle \langle y \rangle$, etc., are statistical covariances. The notation $\langle \cdot \rangle$ indicates an average: $\langle x \rangle = \sum_i x_i/n$, $\langle r^4 \rangle = \sum_i r_i^4/n$, etc.

In Karimäki's development, ρ is the approximate radius. A slightly more accurate approximation, consistent with the exact solution, is

$$a^0 = \frac{1}{n} \sum \|\tilde{\mathbf{z}}_i - \mathbf{z}^0\| \quad (68)$$

ACKNOWLEDGMENTS

The research reported in this paper was conducted as part of a dissertation submitted to the University of South Florida in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Civil Engineering. A full report of this study can be found in Davis (1998).

APPENDIX II. REFERENCES

- Adcock, R. J. (1978). "A problem in least squares." *Analyst*, 5, 53–54.
- Davis, T. G. (1998). "Total least squares curve fitting," PhD dissertation, University of South Florida, Tampa, Fla.
- Davis, T. G., and Lin, S. (1996). "Discussion of 'Intersection of spiral curve with circle,' by Olcay Öztan, Orhan Baykal, Oguz Müftüoğlu, and Muhammed Sahin." *J. Surv. Engrg.*, ASCE, 122(4), 181–183.
- Gander, W., Golub, G. H., and Strelbel, R. (1994). "Least-squares fitting of circles and ellipses." *BIT*, 34, 558–578.
- Golub, G. H., and Van Loan, C. F. (1980). "An analysis of the total least squares problem." *SIAM J. Numer. Anal.*, 17(6), 883–393.
- Gulliksson, M., and Söderkvist, I. (1995). "Surface fitting and parameter estimation with nonlinear least squares." *Opt. Meth. and Soft.*, 5, 247–269.
- Hager, W. W. (1988). *Applied numerical linear algebra*. Prentice-Hall, Englewood Cliffs, N.J.
- Kahmen, H., and Faig, W. (1988). *Surveying*. Walter de Gruyter, Hawthorne, N.Y.
- Karimäki, V. (1992). "Fast code to fit circular arcs." *Comput. Phys. Comm.*, 69, 133–141.
- Marquardt, D. W. (1963). "An algorithm for least-squares estimation of nonlinear parameters." *J. Soc. Indust. Appl. Math.*, 11(2), 431–441.
- Marsden, J. E., and Tromba, A. J. (1981). *Vector calculus*. Freeman, San Francisco.
- Murray, W., ed. (1972). *Numerical methods for unconstrained optimization*. Academic, New York.
- Pearson, K. (1901). "On lines and planes of closest fit to systems of points in space." *Philosophical Mag.*, 2, 559–572.
- Reklaitis, G. V., Ravindran, A., and Ragsdell, K. M. (1983). *Engineering optimization*. Wiley, New York.
- Späth, H. (1997). "Orthogonal least squares fitting by conic sections." *Recent advances in total least squares techniques and errors-in-variables modeling*, S. Van Huffel, ed., SIAM, Philadelphia, 259–264.
- Stoer, J. (1982). "Curve fitting with clothoidal splines." *J. Res. Natl. Bur. Stand.*, 87(4), 317–346.
- Van Huffel, S., ed. (1997). *Recent advances in total least squares techniques and errors-in-variables modeling*. SIAM, Philadelphia.
- Van Huffel, S., and Vandewalle, J. (1991). *The total least squares problem: Computational aspects and analysis*. SIAM, Philadelphia.
- Walton, D. J., and Meek, D. S. (1990). "Clothoidal splines." *Comp. and Graphics*, 14(1), 95–100.
- Wolf, H. (1968). *Ausgleichsrechnung nach der Methode der kleinsten Quadrate*. Dümmlers Verlag, Bonn, Germany (in German).

APPENDIX III. NOTATION

The following symbols are used in this paper:

- \mathbf{A} = rotation matrix;
- \mathbf{A}' = transpose of \mathbf{A} ;
- a = scale factor;
- $C(t)$ = Fresnel cosine integral;
- C^2 = twice continuously differentiable;
- F = objective function;
- f = squared residual;
- $G()$ = graph of argument;
- \mathbf{I} = identity matrix;
- m = number of model parameters;
- n = number of data points;
- \Re = real number line;
- r = radius of curvature;
- $S(t)$ = Fresnel sine integral;
- s = arc length;
- \mathbf{t} = vector of local parameters;
- u = abscissa of \mathbf{w} ;
- \tilde{u} = abscissa of $\tilde{\mathbf{w}}$;
- v = ordinate of \mathbf{w} ;
- \tilde{v} = ordinate of $\tilde{\mathbf{w}}$;
- \mathbf{w} = point on unit curve;
- $\tilde{\mathbf{w}}$ = transformed data point;
- x = abscissa of \mathbf{z} ;
- \tilde{x} = abscissa of $\tilde{\mathbf{z}}$;
- x_0 = abscissa of \mathbf{z}_0 ;
- y = ordinate of \mathbf{z} ;
- \tilde{y} = ordinate of $\tilde{\mathbf{z}}$;
- y_0 = ordinate of \mathbf{z}_0 ;
- \mathbf{z} = point on curve;
- $\tilde{\mathbf{z}}$ = data point;
- \mathbf{z}_0 = local coordinate origin;
- $\boldsymbol{\alpha}$ = vector of model parameters;
- $\Delta\mathbf{w} = \mathbf{w} - \tilde{\mathbf{w}}$;
- $\Delta x = \tilde{x} - x_0$;
- $\Delta y = \tilde{y} - y_0$;
- $\Delta\mathbf{z} = \tilde{\mathbf{z}} - \mathbf{z}_0$;
- $\Delta\boldsymbol{\alpha} = \boldsymbol{\alpha}^{k+1} - \boldsymbol{\alpha}^k$;
- $\Delta\boldsymbol{\xi} = \boldsymbol{\xi}^{k+1} - \boldsymbol{\xi}^k$;
- θ = orientation angle;
- $\boldsymbol{\xi} = (\boldsymbol{\alpha}, \mathbf{t})$;
- φ = spiral angle;
- ∇ = gradient operator;
- ∇^2 = Hessian operator; and
- $\| \|$ = Euclidean norm.

Subscripts

i = summation index.

Superscripts

k = iteration index; and

0 = starting value.

Errata

Total Least-Squares Spiral Curve Fitting^a

The following corrections should be made to the original paper:

Page 161, line 5: Should read “onto the curve by $\mathbf{z}_i^* = \mathbf{z}(\boldsymbol{\alpha}, t_i^*)$ ”

Page 161, line 10: Should read “simply” instead of “supply”

Page 161, line 14: Should read “the \tilde{x}_i are” instead of “the $\tilde{\mathbf{x}}_i$ are”

Page 161, line 14: Should read “projection [Fig. 1(a)] is”

Page 162, line 13: Should read “Now consider the”

Page 164, line 22: Should read “factor of 1/2 in the objective function”

Page 166, line 2: Should read “data points, (\hat{u}, \hat{v}) ”

Page 166, line 11: Should read “such that $F(\boldsymbol{\alpha}^{k+1}) < F(\boldsymbol{\alpha}^k)$ ”

Page 167, line 8: Should read “transformation” instead of “transportation”

Page 173, line 2: Should read “and (9,5) m.”

^a November, 1999, Vol. 125, No. 4, by Thomas G. Davis (Paper 19682).