Defense Technical Information Center
Compilation Part Notice

ADP013727

TITLE: Iota [sub 1] and Iota [sub infinity] ODR Fitting of Geometric Elements

DISTRIBUTION: Approved for public release, distribution unlimited

This paper is part of the following report:

TITLE: Algorithms For Approximation IV. Proceedings of the 2001 International Symposium

To order the complete compilation report, use: ADA412833

The component part is provided here to allow users access to individually authored sections of proceedings, annals, symposia, etc. However, the component should be considered within the context of the overall compilation report and not as a stand-alone technical report.

The following component part numbers comprise the compilation report:
ADP013708 thru ADP013761

l_1 and l_∞ ODR fitting of geometric elements

Hans-Peter Helfrich

Mathematisches Seminar der Landwirtschaftlichen Fakultät der Universität Bonn
helfrich@uni-bonn.de

Daniel S. Zwick

Wilcox Associates, Inc.
dzwick@wilcoxassoc.com

Abstract

We consider the fitting of geometric elements, such as lines, planes, circles, cones, and cylinders, in such a way that the sum of distances or the maximal distance from the element to the data points is minimized. We refer to this kind of distance based fitting as orthogonal distance regression or ODR. We present a *separation of variables* algorithm for l_1 and l_∞ ODR fitting of geometric elements. The algorithm is iterative and allows the element to be given in either implicit form $f(x, \beta) = 0$ or in parametric form $x = g(t, \beta)$, where β is the vector of shape parameters, x is a 2- or 3-vector, and s is a vector of location parameters. The algorithm may even be applied in cases, such as with ellipses, in which a closed form expression for the distance is either not available or is difficult to compute. For l_1 and l_∞ fitting, the norm of the gradient is not available as a stopping criterion, as it is not continuous. We present a stopping criterion that handles both the l_1 and the l_∞ case, and is based on a suitable characterization of the stationary points.

1 Introduction

Let us be given N points $\{z_i\}_{i=1}^N \in \mathbb{R}^d$ and a geometric object S in

- *implicit form* $\{x : f(x, \beta) = 0\}$ with a scalar function f, or
- *parametric form* $x = g(t, \beta)$ with a vector function g,

where the shape parameter vector $\beta \in C$ lies within a closed, convex subset C of $\mathbb{R}^m$. Denote by

$$\phi_i(\beta) = \inf\{\|z_i - x_i\|_2 : x_i \text{ on } S\}$$

the distance of the point z_i to the geometric object S. Let

$$\phi(\beta) = (\phi_1(\beta), \ldots, \phi_N(\beta))^T$$

be the distance vector with norm

$$\Phi(\beta) = \|\phi(\beta)\|,$$

where $\|\phi(\beta)\|$ denotes either the l_∞-norm

$$\Phi(\beta) = \max(\phi_1(\beta), \ldots, \phi_N(\beta))$$

or the l_1-norm

$$\Phi(\beta) = \sum_{i=1}^{N} \phi_i(\beta).$$

We consider the problem:

Find $\beta \in C$ and points $\{x_i\}_{i=1}^N$ on S such that $\Phi(\beta) = \|\phi(\beta)\|$ is minimal.

If the minimum is attained, each function $\phi_i(\beta) = \|z_i - x_i\|_2$ is minimal for the point $x_i \in S$. Then $z_i - x_i$ is orthogonal to S for interior points of S, hence the term "orthogonal distance regression" or "ODR".

Nonlinear l_1 ODR problems are treated in WATSON [10, 12]. A survey for linear problems is given in ZWICK [13].

As stated, the problem has dimension $Nd+m$. In typical metrology applications, the data set is very large so that a direct approach to the problem becomes computationally expensive. We use a *separation of variables* algorithm that was used in [2, 4] and TURNER [9] for the l_2 ODR problem. Each iteration of our algorithm consists of two steps. In the first step, the *foot points* $\{x_i\}_{i=1}^N$ on S, i.e., the location parameters, are calculated for a fixed parameter vector β. These d-dimensional subproblems can be efficiently handled by trust region methods [3].

In the second step, a first order approximation of $\phi_i(\beta)$ is employed, that can be given without explicit knowledge of the dependence of the optimal points $x_i(\beta)$ on β. At this stage, the norm of the correction to the parameter vector β is limited by a trust region strategy. The correction can be computed by solving a linear programming problem. For general nonlinear minimax problems such methods were proposed in MADSEN AND SCHJÆR–JACOBSEN [6], HALD AND MADSEN [1] and JÓNASSON AND K. MADSEN [5].

Our convergence analysis follows the general approach given in POWELL [8] and MORÉ [7]. But in order to handle the l_1 and l_∞ case we cannot use the norm of the gradient as a stopping or convergence criterion, since the gradient is not continuous. Moreover, a neccessary condition for a minimum is that the subgradient contains the zero functional, see, e.g., WATSON [11]. In order to overcome this difficulty, we introduce a replacement for the norm of the gradient that serves both as a stopping criterion and as an essential tool in the convergence proof.

2 The trust region algorithm

At each iteration of our algorithm we solve the low-dimensional subproblems (P_i) for $\beta = \beta_k$ for each fixed i, $i = 1, \ldots, N$:

Minimize $\|z_i - x_i\|_2$ subject to $f(x_i, \beta) = 0$ or $x_i = g(t_i, \beta)$.

In order to apply the trust region method to l_1 and l_∞ ODR we need a first order approximation $\psi_i(\beta, \alpha)$ to $\phi_i(\beta)$. With appropriate regularity assumptions, this can be computed without knowledge of the dependence of the optimal points $x_i(\beta)$ on β ([2], [4]). This means that the iterative improvement in β is *uncoupled* from the calculations of $x_i(\beta)$, whereby a true first order approximation of the objective function is attained.

In the case of the implicit form $f(x,\beta)=0$, the first order approximation $\phi_i(\beta+\alpha) = \psi_i(\beta,\alpha)+o(\alpha)$ is given by

$$\psi_i(\beta,\alpha) = \frac{\nabla_x f(x_i,\beta)^T(z_i-x_i)+\nabla_\beta f(x_i,\beta)^T\alpha}{\|\nabla_x f(x_i,\beta)\|_2}, \tag{2.1}$$

as a first order approximation to the signed distance $\pm\phi_i(\beta+\alpha)$. For the parametric form $x=g(t,\beta)$, we have

$$\psi_i(\beta,\alpha) = \|z_i-x_i\|_2 - \frac{(z_i-x_i)^T}{\|z_i-x_i\|_2}D_\beta g(x_i,\beta)\alpha. \tag{2.2}$$

Note that (2.1) makes sense even for points on the surface. For an orientable hypersurface in parametric form, the expression $\frac{(z_i-x_i)^T}{\|z_i-x_i\|_2}$ in (2.2) should be replaced by the unit normal for points on the surface.

Denote by

$$\psi(\beta) = (\psi_1(\beta),\ldots,\psi_N(\beta))^T$$

the vector of the linearized distances and let

$$\Psi(\beta,\alpha) = \|\psi(\beta,\alpha)\| - \|\phi(\beta)\|.$$

The main algorithm:

- Step 0: An initial $\beta_0\in\mathbb{R}^m$, a trust region radius $\Delta_0>0$, and constants $0<\mu<1$ and $0<\gamma<1<\overline{\gamma}$, $\overline{\Delta}$ are given. Set $k=0$.
- Step 1: Minimize $\Psi(\beta_k,\alpha)$ subject to $\|\alpha\|_2\le\Delta_k$ and $\beta_k+\alpha\in C$. Let α_k denote the solution with minimal norm.
- Step 2: If $\alpha_k=0$, stop.
- Step 3: Compute
$$\rho_k = \frac{\Phi(\beta_k+\alpha_k)-\Phi(\beta_k)}{\Psi(\beta_k,\alpha_k)}.$$
- Step 4:
  (1) *Successful step.* If $\rho_k\ge\mu$ set
  $$\beta_{k+1}=\beta_k+\alpha_k$$
  and choose Δ_{k+1} such that
  $$\Delta_k\le\Delta_{k+1}\le\min(\overline{\gamma}\Delta_k,\overline{\Delta}). \tag{2.3}$$
  (2) *Unsuccessful step.* Otherwise, set
  $$\beta_{k+1}=\beta_k \text{ and } 0<\Delta_{k+1}\le\gamma\Delta_k.$$
- Step 5: Increment k by one and go to Step 1.

3 Global convergence

In an abstract setting our problem may be formulated as

Minimize $\Phi(\beta)=\|\phi(\beta)\|$ *on a closed, convex set* C.

To solve this problem, at each stage of the iteration we solve the following constrained, linearized problem:

Minimize $\Psi(\beta,\alpha)$ *subject to* $\beta+\alpha \in C$ *and* $\|\alpha\| \le \Delta$.

In order to get the linearization in our case, we solve the least distance subproblems (P_i), $i = 1,\dots,N$, with a shape parameter β, and use (2.2), or (2.1).

For the purpose of characterizing stationary points, we introduce the quantity

$$\nabla_1(\beta) = -\inf\{\Psi(\beta,\alpha) \mid \|\alpha\| \le 1,\ \beta+\alpha\ \in C\}.$$

Note that $\nabla_1(\beta) \ge 0$, since $\Psi(\beta,0) = 0$.

By convexity, $\nabla_1(\beta) = 0$ implies that $\alpha = 0$ is a solution of the linearized minimization problem. MADSEN AND SCHJÆR-JACOBSEN [6] have shown that the latter condition is equivalent to a condition given therein for the functional to have a stationary point. In order to prove Theorem 3.3 we prove a lemma that was given in a similar form for the l_∞ case in MADSEN AND SCHJÆR-JACOBSEN [6] and JÓNASSON AND MADSEN [5]). We give a different proof that is applicable to both the l_1 and l_∞ cases.

Lemma 3.1 *Let* $\nabla_1(\beta) \ge \epsilon$ *and* $\Delta \le \overline{\Delta}$. *For the solution of the linearized problem the estimate*

$$\Psi(\beta,\alpha) \le -C\epsilon\Delta \tag{3.1}$$

holds, with a constant that depends only on ϵ *and* $\overline{\Delta}$.

Proof: According to the definition of $\nabla_1(\beta)$ and the continuity of Ψ there exists a feasible α_1 with $\|\alpha_1\| \le 1$ such that

$$\Psi(\beta,\alpha_1) = -\epsilon.$$

Let $\alpha = t\alpha_1$, where $t = \min(1,\Delta)$. Since $\Psi(\beta,\alpha)$ is a convex function, we get

$$\Psi(\beta,\alpha) \le (1-t)\Psi(\beta,0) + t\Psi(\beta,\alpha_1) = -t\epsilon.$$

Since

$$t \ge \Delta\min(1, 1/\overline{\Delta})$$

we get the conclusion with $C = \min(1, 1/\overline{\Delta})$. □

Proposition 3.2 *For a minimum point,*

$$\nabla_1(\beta) = 0$$

holds.

Proof: Assume the contrary, then $\nabla_1(\beta) = \epsilon > 0$ holds. According to the definition of $\Psi(\beta,\alpha)$ we have

$$\Phi(\beta+\alpha) = \Phi(\beta) + \Psi(\beta,\alpha) + o(\alpha).$$

By Lemma 3.1, we can find an α with $\|\alpha\| \le \Delta$ such that (3.1) holds. As in the proof of the Lemma, we may conclude that

$$\Phi(\beta+t\alpha) \le \Phi(\beta) - C\epsilon t\Delta + o(t\alpha)$$

for $0 < t \le 1$. If we let $t \to 0$ we get a contradiction to the minimum property. □

Theorem 3.3 *Either the algorithm ends in a finite number of steps, or a sequence β_k is generated for which* $\liminf_{k\to\infty} \nabla_1(\beta_k) = 0$.

Proof: Assume the contrary. Then there exists $\epsilon > 0$ such that $\nabla_1(\beta_k) \geq \epsilon$ holds for all k. By the definition of ρ_k and the lemma, it follows that for a successful step

$$\phi(\beta_{k+1}) \leq \phi(\beta_k) - \mu C \epsilon \Delta_k$$

and by the updating rule for Δ_{k+1} we get

$$\Delta_{k+1} \leq c(\phi(\beta_{k+1}) - \phi(\beta_k)),$$

with $c = 1/(\mu C \epsilon)$. Combining this inequality with the updating rule for an unsuccessful step yields

$$\Delta_{k+1} \leq \gamma \Delta_k + c(\phi(\beta_{k+1}) - \phi(\beta_k)).$$

By summation and the monotonicity of $\phi(\beta_k)$ it follows that for all N

$$\sum_{k=0}^{N} \Delta_k \leq \frac{\Delta_0}{1-\gamma} + \frac{c}{1-\gamma}\phi(\beta_1).$$

Since this implies the convergence of $\sum \Delta_k$, we get $\lim \Delta_k = 0$. From $\|\beta_k\| \leq \Delta_k$ we obtain the convergence of β_k. From the definition of ρ_k it then follows that $\lim \rho_k = 1$. But then the updating rule (2.3) implies that eventually $\Delta_{k+1} \geq \Delta_k$, which gives a contradiction. □

Theorem 3.4 *(Global Convergence, cf.* MORÉ *[7],* POWELL *[8]) Assume that* $\nabla_1(\beta)$ *is uniformly continuous. Then either the algorithm ends in a finite number of steps, or a sequence β_k is generated for which*

$$\lim_{k\to\infty} \nabla_1(\beta_k) = 0.$$

Proof: Assume the contrary. Then there exists an ϵ_1 such that for each k_0 there exists a $k \geq k_0$ with

$$\nabla_1(\beta_k) \geq \epsilon_1.$$

By Theorem 3.3 we can find an index $l > k$ such that

$$\nabla_1(\beta_l) \leq \epsilon_1/2$$

(k_0 will be determined later). We choose the smallest such l. As in the proof of Theorem 3.3, it follows that for that a successful step with $k \leq i < l$,

$$\|\beta_{i+1} - \beta_i\| \leq \Delta_k \leq 2c_1(\phi(\beta_i) - \phi(\beta_{i+1})).$$

Clearly, this also holds for an unsuccessful step. This yields

$$\|\beta_l - \beta_k\| \leq 2c_1(\phi(\beta_k) - \phi(\beta_l)).$$

Since $\phi(\beta_i)$ converges by monotonicity, we can make $\|\beta_l - \beta_k\|$ arbitrarily small for large enough k_0 . By the uniform continuity of $\nabla_1(\beta)$ we infer

$$|\nabla_1(\beta_k) - \nabla_1(\beta_l)| < \epsilon_1/2,$$

which is a contradiction. □

4 A numerical example

As an illustrative example, we fit an ellipse to data, given as coordinate pairs in $\mathbb{R}^2$. There are 24 data points and five components to the shape parameter vector (i.e., $n = 2, d = 2, m = 5, N = 24$). We used a standard parameterization involving a center (x_0, y_0), the axes (a, b), and a rotation angle θ.

The output is shown below. The initial values for the parameters and the obtained parameters in three different norms are given in Table 1. In the l_2 case, we give as the error the root mean square error, in the l_1 case the *mean absolute deviation*, and in the l_∞ case the *maximum deviation.*

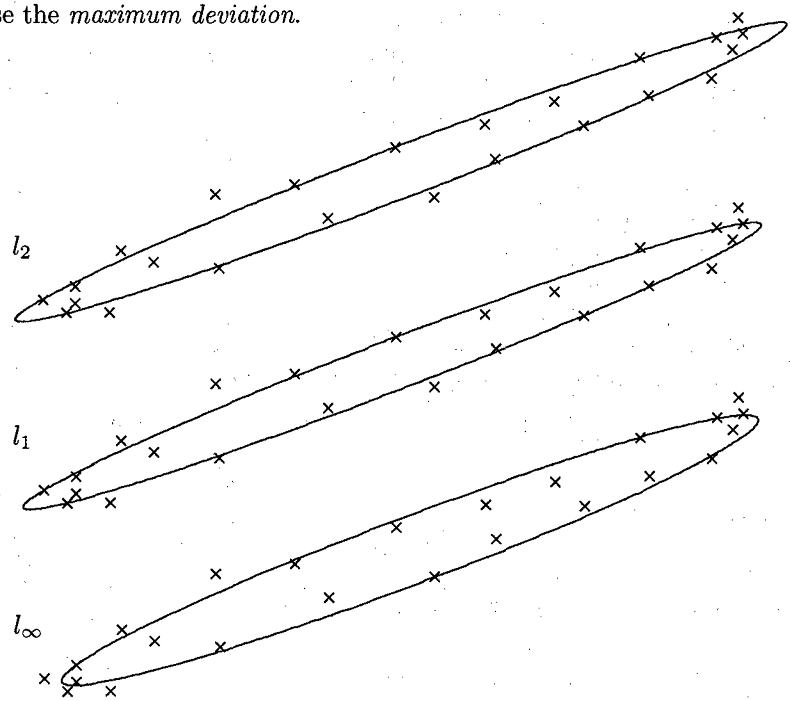

FIG. 1. l_2, l_1, and l_∞-Approximation.

	x_0	x_1	a	b	θ (degrees)	Error
Initial values	0.4989881	-1.4262126	4.6719913	0.4364267	20.75913	
l_2	0.6637511	-1.3987826	5.5124671	0.3376480	20.90124	0.11520
l_1	0.5368646	-1.4465520	5.2778061	0.3358224	20.88869	0.09047
l_∞	0.7694412	-1.3829474	4.9731226	0.4491259	20.66893	0.23489

TAB. 1. Parameters for different norms.

The number of iterations in each case was five or six. We note that the deviations for the best fit l_1 and l_∞ ellipses exhibit behavior typical to these norms: five of the data points lie on the best fit l_1 ellipse and there are six deviations of largest magnitude in the l_∞ case.

Bibliography

1. J. Hald and K. Madsen. Combined LP and Quasi-Newton methods for minimax optimization. *Mathematical Programming*, 20:49–62, 1981.
2. H.-P. Helfrich and D. Zwick. A trust region method for implicit orthogonal distance regression. *Numerical Algorithms*, 5:535–545, 1993.
3. H.-P. Helfrich and D. Zwick. Trust region algorithms for the nonlinear least distance problem. *Numerical Algorithms*, 9:171–179, 1995.
4. H.-P. Helfrich and D. Zwick. A trust region algorithm for parametric curve and surface fitting. *J. Comput. Appl. Math.*, 73:119–134, 1996.
5. K. Jónasson and K. Madsen. Corrected sequential linear programming for sparse minimax optimization. *BIT*, 34:372–387, 1994.
6. K. Madsen and H. Schjær–Jacobson. Linearly constrained minimax optimization. *Mathematical Programming*, 14:208–223, 1978.
7. J. J. Moré. Recent developments in algorithms and software for trust region methods. In A. Bachem, M. Grötschel, and B. Korte, editors, *Mathematical Programming Bonn 1982–The State of the Art*, pages 259–287. Springer, 1983.
8. M. J. D. Powell. Convergence properties of a class of minimization algorithms. In O. L. Mangasarian, R. R. Meyer, and S. M. Robinson, editors, *Nonlinear Programming 2*, pages 1–27. Academic Press, 1975.
9. D. A. Turner, I. J. Anderson, J. C. Mason, M. G. Cox, and A. B. Forbes. An efficient separation-of-variables approach to parametric orthogonal distance regression. In P. Ciarlini, M. G. Cox, F. Pavese, and D. Richter, editors, *Advanced Mathematical and Computational Tools in Metrology IV*, pages 246–255, Singapore, 2000. World Scientific.
10. G. A. Watson. The use of the l_1 norm in nonlinear errors-in-variables problems. In S. Van Huffel, editor, *Recent Advances in Total Least Squares Techniques and Errors-in-Variables Modeling*, pages 183–192, Philadelphia, 1997. SIAM.
11. G. A. Watson. Choice of norms for data fitting and function approximation. *Acta Numerica*, pages 337–377, 1998.
12. G. A. Watson. Some robust methods for fitting parametrically defined curves and surfaces to measured data. In F. Pavese P. Ciarlini, A. B. Forbes and D. Richter, editors, *Advanced Mathematical and Computational Tools in Metrology IV*, volume 53 of *Series on Advances in Mathematics for Applied Sciences*, pages 256–272. World Scientific, 2000.
13. D. Zwick. Algorithms for orthogonal fitting of lines and planes: a survey. In P. Ciarlini, M. G. Cox, F. Pavese, D. Richter, editors, *Advanced Mathematical Tools in Metrology II*, pages 272–283. World Scientific, 1996.