

# Use of $l_p$ norms in fitting curves and surfaces to data

A. Atieg      G. A. Watson\*

(Received 8 August 2003; revised 31 March 2004)

## Abstract

Given a family of curves or surfaces in  $\mathbb{R}^s$ , an important problem is that of finding a member of the family which gives a “best” fit to  $m$  given data points. A criterion which is relevant to many application areas is orthogonal distance regression, where the sum of squares of the orthogonal distances from the data points to the surface is minimized. For example, this is important in metrology, where measured data from a manufactured part may have to be modelled.

The least squares norm is not always suitable (for example, there may be wild points in the data, accept/reject decisions may be required, etc). So we use this to justify looking at the use of other  $l_p$  norms. There are different ways to formulate the problem, and we examine methods which generalize in a natural way those available for

---

\*Department of Mathematics, University of Dundee, Dundee DD1 4HN, SCOTLAND.  
<mailto:gawatson@maths.dundee.ac.uk>

See <http://anziamj.austms.org.au/V45/CTAC2003/Atie> for this article, © Austral. Mathematical Soc. 2004. Published 13 April 2004, amended April 20, 2004. ISSN 1446-8735

least squares. The emphasis is on the efficient numerical treatment of the resulting problems.

## Contents

<b>1 Introduction</b>	<b>C188</b>
<b>2 Jacobians and Gauss-Newton subproblems</b>	<b>C190</b>
<b>3 The cases <math>1 &lt; p &lt; \infty</math></b>	<b>C193</b>
<b>4 Concluding Remarks</b>	<b>C197</b>
<b>References</b>	<b>C198</b>

## 1 Introduction

Let data points  $\mathbf{x}_i \in \mathbb{R}^s$ ,  $i = 1, \dots, m$  be generated, and let a parametric curve or surface in  $\mathbb{R}^s$  be defined by  $\mathbf{a} \in \mathbb{R}^n$ , with a point on the surface given by  $\mathbf{x}(\mathbf{a}, \mathbf{t})$ , where  $\mathbf{t} \in \mathbb{R}^d$ , and  $d = s - 1$  normally. It is required to determine a value of  $\mathbf{a}$  which gives a “best” fit to the data, and we consider different criteria. For any  $\mathbf{a} \in \mathbb{R}^n$ ,  $\mathbf{t}_i \in \mathbb{R}^d$ ,  $i = 1, \dots, m$ , define

$$\mathbf{u}_i(\mathbf{a}, \mathbf{t}_i) = \mathbf{x}_i - \mathbf{x}(\mathbf{a}, \mathbf{t}_i), \quad i = 1, \dots, m,$$

$$\mathbf{w}(\mathbf{a}, \mathbf{t}_1, \dots, \mathbf{t}_m) = (\mathbf{u}_1(\mathbf{a}, \mathbf{t}_1)^T, \dots, \mathbf{u}_m(\mathbf{a}, \mathbf{t}_m)^T)^T \in \mathbb{R}^{ms}.$$

Two of the criteria are based on the idea of orthogonal distances: for every  $\mathbf{x}_i$  associate a point  $\mathbf{x}(\mathbf{a}, \mathbf{t}_i(\mathbf{a}))$  on the surface, where

$$\mathbf{t}_i(\mathbf{a}) = \arg \min_{\mathbf{t}} \|\mathbf{u}_i(\mathbf{a}, \mathbf{t})\|, \quad i = 1, \dots, m,$$

and where unadorned norms are  $l_2$  norms. Define

$$\begin{aligned}\mathbf{r}(\mathbf{a}) &= (\|\mathbf{u}_1(\mathbf{a}, \mathbf{t}_1(\mathbf{a}))\|, \dots, \|\mathbf{u}_m(\mathbf{a}, \mathbf{t}_m(\mathbf{a}))\|)^T \in \mathbb{R}^m, \\ \mathbf{s}(\mathbf{a}) &= (\mathbf{u}_1(\mathbf{a}, \mathbf{t}_1(\mathbf{a}))^T, \dots, \mathbf{u}_m(\mathbf{a}, \mathbf{t}_m(\mathbf{a}))^T)^T \in \mathbb{R}^{ms}.\end{aligned}$$

Let  $1 \leq p \leq \infty$ . Then two basic fitting problems which arise from this are to find

$$\min_{\mathbf{a} \in \mathbb{R}^n} \|\mathbf{r}(\mathbf{a})\|_p, \quad (1)$$

$l_p$  orthogonal distance regression, and

$$\min_{\mathbf{a} \in \mathbb{R}^n} \|\mathbf{s}(\mathbf{a})\|_p, \quad (2)$$

$l_p$  component wise orthogonal distance regression. A third problem is that of finding

$$\min_{\mathbf{a}, \mathbf{t}_1, \dots, \mathbf{t}_m} \|\mathbf{w}(\mathbf{a}, \mathbf{t}_1, \dots, \mathbf{t}_m)\|_p, \quad (3)$$

$l_p$  distance regression.

All these problems are identical if  $p = 2$ , when we have a special case of *orthogonal distance regression*, for which methods of Gauss-Newton type have proved popular. Although the problems are identical, the direct treatment of the problems in the forms given here lead to different algorithms. For the Gauss-Newton method applied to (1), see for example [7, 11, 12, 15], and applied to (2), see for example [2, 3, 4]. For a unified treatment of these methods, and some comparisons, see [1]. The direct solution of (3) trades the subproblems for  $\mathbf{t}_i(\mathbf{a})$  for an increase in the number of variables, and generates orthogonal distances only in the limit: methods which exploit the structure of the Jacobian in Gauss-Newton (or Levenberg-Marquardt) methods are given in [6, 10]. Which approach should be used depends mainly on whether or not the points  $\mathbf{x}(\mathbf{a}, \mathbf{t}_i(\mathbf{a}))$  are easy to calculate.

Attention focuses on the case  $p = 2$  for two main reasons:

1. relative ease of computation,
2. an assumption that when the errors are normally distributed then least squares is appropriate.

However, the nature of the errors may be such that least squares is not particularly suitable. For example, there may be wild points in the data, in which case the choice  $p = 2$  may give too much weight to these. Another situation which can arise in the modelling of manufactured parts is that accept/reject decisions may be required, suggesting that  $p = 2$  can be too small. Therefore we here consider the situation when another criterion from this class is preferable, and emphasise the numerical treatment of the resulting problems.

The problems (1), (2) and (3) are all nonlinear  $l_p$  norm approximation problems, and a natural way to proceed is, as for  $p = 2$ , through a method of Gauss-Newton type, or one of its variants. In the next section, we look at the structure of the Jacobian matrices and their role in the definition of different Gauss-Newton steps. In Section 3, we consider values of  $p$  from the range  $1 < p < \infty$ : the intention is to show that steps which are asymptotically Gauss-Newton steps can be obtained with a level of complexity comparable with that for the  $p = 2$  case.

In some situations it may be desirable to work with rotated data in order to simplify the model. However, for ease of presentation, we will not consider that here: straightforward changes would be required. It is also possible to work with implicit models.

We use  $\nabla_1(\nabla_2)$  to denote the operation of taking partial derivatives of a function of two sets of variables with respect to the first (second) set.

## 2 Jacobians and Gauss-Newton subproblems

For (1), the Gauss-Newton step at  $\mathbf{a}$  is given by the minimizer of

$$\|\mathbf{r} + J_1 \mathbf{d}\|_p, \quad (4)$$

where  $J_1$  is the  $m \times n$  Jacobian matrix of  $\mathbf{r}$ . Various modifications of this (Levenberg-Marquardt, trust region) may be preferable, but for simplicity we concentrate on (4). Provided  $\mathbf{u}_i \neq \mathbf{0}$ , the  $i$ th row of  $J_1$  is

$$\rho_i(J_1) = \frac{\mathbf{u}_i(\mathbf{a}, \mathbf{t}_i(\mathbf{a}))^T}{\|\mathbf{u}_i(\mathbf{a}, \mathbf{t}_i(\mathbf{a}))\|} \nabla_{\mathbf{a}} \mathbf{u}_i(\mathbf{a}, \mathbf{t}_i(\mathbf{a})) = \frac{\mathbf{u}_i(\mathbf{a}, \mathbf{t}_i(\mathbf{a}))^T}{\|\mathbf{u}_i(\mathbf{a}, \mathbf{t}_i(\mathbf{a}))\|} \nabla_1 \mathbf{u}_i(\mathbf{a}, \mathbf{t}_i(\mathbf{a})), \quad (5)$$

because  $\mathbf{t}_i(\mathbf{a})$  minimizes  $\|\mathbf{u}_i(\mathbf{a}, \mathbf{t}_i)\|$  so that  $\nabla_2 \|\mathbf{u}_i(\mathbf{a}, \mathbf{t}_i(\mathbf{a}))\| = 0$ . If any  $\mathbf{u}_i$  is zero,  $J_1$  is not defined. Away from a limit point, we may regard that as a degenerate situation. More serious is such an occurrence at a limit point of the iteration. That may again be regarded as a degeneracy except when  $p = 1$ , as the interpolation characteristics of the  $l_1$  solution means that it is then an expected event. However, despite this, it is shown in [13] that the convergence property of the iteration is normally just as in the usual (smooth) case. Therefore the use of (4) does not present any particular difficulties here.

The treatment of (2) rather than (1) means that the problem is one in  $\mathbb{R}^{ms}$ , and the Gauss-Newton step at  $\mathbf{a}$  is given by the minimizer of

$$\|\mathbf{s} + J_2 \mathbf{d}\|_p \quad (6)$$

where

$$J_2 = \begin{bmatrix} \nabla_{\mathbf{a}} \mathbf{u}_1(\mathbf{a}, \mathbf{t}_1(\mathbf{a})) \\ \nabla_{\mathbf{a}} \mathbf{u}_2(\mathbf{a}, \mathbf{t}_2(\mathbf{a})) \\ \vdots \\ \nabla_{\mathbf{a}} \mathbf{u}_m(\mathbf{a}, \mathbf{t}_m(\mathbf{a})) \end{bmatrix} \in \mathbb{R}^{ms \times n}.$$

No problems arise here from zero  $\mathbf{u}_i$ . However, there is no corresponding simplification to the derivative calculations, as in (5), so we need to use

$$\nabla_{\mathbf{a}} \mathbf{u}_i(\mathbf{a}, \mathbf{t}_i(\mathbf{a})) = \nabla_1 \mathbf{u}_i(\mathbf{a}, \mathbf{t}_i(\mathbf{a})) + \nabla_2 \mathbf{u}_i(\mathbf{a}, \mathbf{t}_i(\mathbf{a})) \nabla_{\mathbf{a}} \mathbf{t}_i(\mathbf{a}), \quad i = 1, \dots, m.$$

Therefore unless an explicit expression is available for  $\mathbf{t}_i(\mathbf{a})$ , it is necessary to use approximate derivatives or alternatively to compute

$$\nabla_{\mathbf{a}} \mathbf{t}_i(\mathbf{a}) = -(\nabla_{22} \|\mathbf{u}_i(\mathbf{a}, \mathbf{t}_i(\mathbf{a}))\|)^{-1} \nabla_{12} \|\mathbf{u}_i(\mathbf{a}, \mathbf{t}_i(\mathbf{a}))\|,$$

for each  $i$ , provided that the matrix being inverted is non-singular. This expression comes from the fact that  $\mathbf{t}_i(\mathbf{a})$  is defined by  $\nabla_2 \|\mathbf{u}_i(\mathbf{a}, \mathbf{t}_i(\mathbf{a}))\| = \mathbf{0}$ , for each  $i$ , which is therefore an identity in  $\mathbf{a}$ .

In addition to this extra computation, the linear subproblems are larger than before. However, (2) is equivalent to

$$\min_{\mathbf{a} \in \mathbb{R}^n} \left\| \left( \|\mathbf{u}_1(\mathbf{a}, \mathbf{t}_1(\mathbf{a}))\|_p, \dots, \|\mathbf{u}_m(\mathbf{a}, \mathbf{t}_m(\mathbf{a}))\|_p \right)^T \right\|_p \quad (7)$$

so that the problem could again be considered as one in  $\mathbb{R}^m$ . However, the Gauss-Newton method will only apply directly to (7) when  $1 < p < \infty$ , because the entries of the vector whose norm is being minimized are  $l_p$  norms, and these are not normally smooth functions when  $p = 1$  or  $p = \infty$ .

Finally, consider the  $l_p$  distance regression problem (3). Then the Gauss-Newton step is the minimizer of

$$\|\mathbf{w} + J_3 \mathbf{d}\|_p \quad (8)$$

where  $J_3$  is the matrix of partial derivatives of  $\mathbf{w}$  with respect to the variables  $\mathbf{a}, \mathbf{t}_i, i = 1, \dots, m$ . Clearly this matrix, although much larger than  $J_1$  or  $J_2$ , is highly structured, because we have

$$J_3 = \begin{bmatrix} \nabla_1 \mathbf{u}_1(\mathbf{a}, \mathbf{t}_1) & \nabla_2 \mathbf{u}_1(\mathbf{a}, \mathbf{t}_1) & 0 & \cdots & 0 \\ \nabla_1 \mathbf{u}_2(\mathbf{a}, \mathbf{t}_2) & 0 & \nabla_2 \mathbf{u}_1(\mathbf{a}, \mathbf{t}_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \nabla_1 \mathbf{u}_m(\mathbf{a}, \mathbf{t}_m) & 0 & 0 & \cdots & \nabla_2 \mathbf{u}_1(\mathbf{a}, \mathbf{t}_m) \end{bmatrix}. \quad (9)$$

TABLE 1: Properties of different formulations,  $1 < p < \infty$ .

Problem	formulation	usual derivative level needed	dimensions of Jacobian
(1)	(1)	1	$m \times n$
(2)	(2)	2	$ms \times n$
(2)	(7)	2	$m \times n$
(3)	(3)	1	$ms \times (n + md)$
(3)	(10)	1	$m \times n$

Note that as a final equivalence, the problem (3) for  $1 < p < \infty$  corresponds to finding

$$\min_{\mathbf{a} \in \mathbb{R}^n} \left\| \left( \|\mathbf{u}_1(\mathbf{a}, \hat{\mathbf{t}}_1(\mathbf{a}))\|_p, \dots, \|\mathbf{u}_m(\mathbf{a}, \hat{\mathbf{t}}_m(\mathbf{a}))\|_p \right)^T \right\|_p, \quad (10)$$

where

$$\hat{\mathbf{t}}_i(\mathbf{a}) = \arg \min_{\mathbf{t}} \|\mathbf{u}_i(\mathbf{a}, \mathbf{t})\|_p, \quad i = 1, \dots, m.$$

Although the treatment of the problem in this form would result in a Jacobian matrix which is  $m \times n$  and which can be formed using first derivatives only, analogous to (5), the subproblems for  $\hat{\mathbf{t}}_i(\mathbf{a})$  are never likely to be worth solving.

### 3 The cases $1 < p < \infty$

Table 1 summarizes the equivalences identified in the previous section and some properties of the different formulations when  $1 < p < \infty$ .

When  $p = 1, 2, \infty$ , then the problems of minimizing (4), (6) and (8) are all finite. However, for other values of  $p$ , this not the case, and the Gauss-Newton method becomes a doubly infinite process. It is clearly desirable

to replace these problems by simpler problems, and we can do that in the following way.

Consider the general nonlinear  $l_p$  problem of minimizing  $\|\mathbf{f}(\mathbf{a})\|_p$ , where  $\mathbf{f} \in \mathbb{R}^m$ ,  $\mathbf{a} \in \mathbb{R}^n$  and  $\mathbf{f}$  is differentiable. Let  $A = \nabla_{\mathbf{a}}\mathbf{f}$ . As indicated, the conventional Gauss-Newton method involves the subproblem:

$$\text{minimize } \|\mathbf{A}\mathbf{d} + \mathbf{f}\|_p,$$

to identify a search direction  $\mathbf{d}$ . For  $1 < p < \infty$ , if  $\mathbf{d}_1$  is a solution, and  $\mathbf{c} = \mathbf{A}\mathbf{d}_1 + \mathbf{f}$ , then  $\mathbf{d}_1$  satisfies the zero derivative condition

$$A^T D_{\mathbf{c}} \mathbf{A} \mathbf{d}_1 + A^T D_{\mathbf{c}} \mathbf{f} = \mathbf{0}, \quad (11)$$

where

$$D_{\mathbf{c}} = \text{diag} \{ |c_j|^{p-2}, j = 1, \dots, m \},$$

if this matrix is defined. This definition will be used with other vector subscripts on  $D$ . Note that  $\mathbf{d}_1$  is always defined, although when  $1 < p < 2$ ,  $D_{\mathbf{c}}$  is only defined if  $c_j \neq 0$ , for all  $j$ .<sup>1</sup>

The above system of equations (11) is a nonlinear system of equations for  $\mathbf{d}_1$ . Now consider Newton's method for minimizing

$$F(\mathbf{a}) = \|\mathbf{f}(\mathbf{a})\|_p^p = \sum_{i=1}^m |f_i(\mathbf{a})|^p.$$

Then

$$\mathbf{g}(\mathbf{a}) = \nabla_{\mathbf{a}} F(\mathbf{a})^T = p \sum_{i=1}^m |f_i|^{p-1} \text{sign}(f_i) \nabla_{\mathbf{a}} f_i^T = p A^T D_{\mathbf{f}} \mathbf{f},$$

and Newton's method involves the subproblem

$$\mathbf{g} + \nabla_{\mathbf{a}} \mathbf{g} \mathbf{d}_N = \mathbf{0},$$

---

<sup>1</sup>A referee made the interesting suggestion that a way around potential difficulties with  $c_j = 0$  in the cases  $1 < p < 2$  could be to reformulate (11) to use  $V$ -invariant solution methods [8, 9], with  $V = D_{\mathbf{c}}^{-1}$ .



for the Newton step  $\mathbf{d}_N$ . Now

$$\begin{aligned}\nabla_{\mathbf{a}}\mathbf{g} &= p \sum_{i=1}^m |f_i|^{p-2} \nabla_{\mathbf{a}} f_i^T \nabla_{\mathbf{a}} f_i \\ &+ p(p-2) \sum_{i=1}^m |f_i|^{p-3} \text{sign}(f_i) f_i \nabla_{\mathbf{a}} f_i^T \nabla_{\mathbf{a}} f_i + p \sum_{i=1}^m |f_i|^{p-2} f_i \nabla_{\mathbf{a}}^2 f_i \\ &= p(p-1) A^T D_{\mathbf{f}} A + p \sum_{i=1}^m |f_i|^{p-2} f_i \nabla_{\mathbf{a}}^2 f_i.\end{aligned}$$

Ignoring the second derivative term, then an approximate Newton step  $\mathbf{d}_2$  is given by

$$(p-1)A^T D_{\mathbf{f}} A \mathbf{d}_2 + A^T D_{\mathbf{f}} \mathbf{f} = \mathbf{0}.$$

This is a descent step for  $F(\mathbf{a})$  if  $A^T D_{\mathbf{f}} A$  is positive definite, unless  $A^T D_{\mathbf{f}} \mathbf{f} = \mathbf{0}$ , but of course then we have a zero derivative of  $F$ .

Note that  $\mathbf{d}_2$  can be obtained in two steps:

1. Find  $\mathbf{d}$  to minimize  $\|D_{\mathbf{f}}^{1/2} A \mathbf{d} + D_{\mathbf{f}}^{1/2} \mathbf{f}\|$ ;
2. Set  $\mathbf{d}_2 = \frac{1}{p-1} \mathbf{d}$ .

In other words the step  $\mathbf{d}_2$  is readily obtained from a linear least squares problem. Note that when  $p < 2$ , no component of  $f_i$  can become zero, otherwise  $D_{\mathbf{f}}$  is not defined, and so neither is  $\mathbf{d}_2$ . There is no corresponding restriction when  $p \geq 2$ .

**Theorem 1** *Let  $\mathbf{a}$  be such that  $D_{\mathbf{f}}$  and  $D_{\mathbf{c}}$  are both defined. Then  $\mathbf{d}_2 = \mathbf{d}_1 + \mathcal{O}(\|\mathbf{d}_1\|^2)$ .*

**Proof:** If  $\mathbf{c} = A\mathbf{d}_1 + \mathbf{f}$ , then by Taylor expansion, for each  $i$ ,

$$|c_i|^{p-2} = |f_i|^{p-2} + (p-2)|f_i|^{p-4}f_i(\mathbf{A}\mathbf{d}_1)_i + \mathcal{O}(\|\mathbf{d}_1\|^2).$$

Thus

$$D_{\mathbf{c}} = D_{\mathbf{f}} + (p-2)\text{diag}\{|f_i|^{p-4}f_i(\mathbf{A}\mathbf{d}_1)_i\} + \mathcal{O}(\|\mathbf{d}_1\|^2).$$

By definition of  $\mathbf{d}_1$ ,

$$\begin{aligned} \mathbf{0} &= A^T D_{\mathbf{c}} \mathbf{A} \mathbf{d}_1 + A^T D_{\mathbf{c}} \mathbf{f} \\ &= A^T D_{\mathbf{f}} \mathbf{A} \mathbf{d}_1 + A^T D_{\mathbf{f}} \mathbf{f} + (p-2)A^T D_{\mathbf{f}} \mathbf{A} \mathbf{d}_1 + \mathcal{O}(\|\mathbf{d}_1\|^2) \\ &= (p-1)A^T D_{\mathbf{f}} \mathbf{A} \mathbf{d}_1 + A^T D_{\mathbf{f}} \mathbf{f} + \mathcal{O}(\|\mathbf{d}_1\|^2). \end{aligned}$$

Now

$$\mathbf{d}_2 = -\frac{1}{p-1}(A^T D_{\mathbf{f}} A)^{-1} A^T D_{\mathbf{f}} \mathbf{f} = \mathbf{d}_1 + \mathcal{O}(\|\mathbf{d}_1\|^2),$$

and so the result follows. 

For all  $p$  there is therefore a finite calculation which gives either the Gauss-Newton step or a step which is the Gauss-Newton step asymptotically. Note that properties of the Jacobian will be the same as in the  $l_2$  case, so similar conclusions can be drawn about performance.

The results for some examples, in particular the fitting of paraboloids and spheres in 3 dimensions and circles in 2 and 3 dimensions, are given in Table 2. In each case we fit to 100 points using (1), (2) and (7). The number of iterations to satisfy  $\|\mathbf{d}\|_{\infty} < 10^{-4}$  is shown, starting from common starting points. In all cases, it was possible to take full steps. Similar results hold for other examples. Generally, there are advantages in treating the problems as ones in  $\mathbb{R}^m$ ; however, for planar data in 3 dimensions, this problem is ill-conditioned (see [1] for a treatment of this in the  $l_2$  case).

The results suggest that these  $l_p$  problems can often be solved in a small multiple of the cost of solving  $l_2$  problems. Note that this is not necessarily true for linear problems.

TABLE 2: Iteration counts for some examples

$p$	paraboloid			sphere			circle in 2-D			circle in 3-D		
	(1)	(7)	(2)	(1)	(7)	(2)	(1)	(7)	(2)	(1)	(7)	(2)
1.1	28	29	25	17	19	24	23	21	34	23	40	19
1.2	13	32	29	14	14	9	24	10	20	25	167	22
1.5	10	13	19	11	7	30	7	7	20	38	18	16
1.8	10	10	16	7	7	6	7	7	17	16	18	7
2.0	7	8	18	5	5	10	5	5	16	15	15	6
2.2	9	10	21	6	6	9	7	7	9	15	15	7
2.7	10	10	11	8	8	8	9	9	21	20	55	8
3.6	12	13	28	11	11	12	12	12	14	67	19	13

Now consider the  $l_p$  distance regression problem (3). For the case  $p = 2$  the structure of the Jacobian matrix in (9) can be exploited as in ODRPACK or FUNKE [5, 6, 10]. But in this more general case, this structure can also be exploited in the least squares subproblem which, as shown, gives the Gauss-Newton step asymptotically. We will not go into further details here.

## 4 Concluding Remarks

We have been concerned with the use of  $l_p$  norms in the fitting of curves and surfaces to discrete data. Three fitting criteria have been considered, along with formulations and methods which are natural generalizations of those for the  $l_2$  case. The main aim has been to show how variants of the Gauss-Newton method can be used for all of these in an efficient way. Lack of space has meant that it is not possible to consider in any detail the important limiting values  $p = 1$  and  $p = \infty$ , although all the forms (1), (2) and (3) are also relevant for both these values.

For the case  $p = 1$ , as already indicated, the analysis of [13] shows that difficulties are not caused in solving (1) by the Gauss-Newton method if any  $\mathbf{u}_i$  is zero at a limit point of the iteration, and the convergence property of the iteration is normally just as in the usual (smooth) case. However, for some geometric fitting problems, there can be a distinct difference between the use of (1) and the use of (2). The reason for this is that second order convergence requires  $n$  zero components at a solution, and this can be more easily achieved for the problem (2). For example the problem of fitting a line in 3 dimensions to data gives difficulties for (1) because here  $n = 4$  but there are generally at most 2 interpolation points. Therefore second order convergence is not possible. However, (2) does not have the same disadvantage.

For the  $l_1$  distance regression problem (the problem (3) with  $p = 1$ ), the Gauss-Newton subproblem is again highly structured. For the case  $s = 2$ , it is possible to incorporate bounds on the variables (making a trust region variant), and efficiently solve the resulting subproblem by a version of the simplex method for linear programming. This can be formulated in such a way that it corresponds to a bounded variable linear programming problem with  $m$  equality constraints and  $4m + n$  variables. The idea uses the analogue of a method for a class of data fitting problems developed in [14].

For  $p = \infty$ , zero distances are not normally an issue, so there are no difficulties using (1). The performance of the methods for (1) and (2) depends primarily on the number of components where the vectors  $\mathbf{r}$  or  $\mathbf{s}$  attain the norm at a limit point of the iteration: a second order rate of convergence is possible if that is  $n + 1$ . For the problem (3), the Gauss-Newton subproblem (8) is usually solved by going to the dual formulation of this problem posed as a linear programming problem. Again the structure present in the matrix  $J_3$ , given by (9), should be exploited.

## References

- [1] A. Atieg and G. A. Watson, A class of methods for fitting a curve or surface to data by minimizing the sum of squares of orthogonal distances, *J. Comp. Appl. Math.* 158, pp. 277–296 (2003). [C189](#), [C196](#)
- [2] S. J. Ahn, W. Rauh and H.-J. Warnecke, Best-fit of implicit surfaces and plane curves, in *Mathematical Methods for Curves and Surfaces: Oslo 2000*, T. Lyche and L. L. Schumacker (eds), Vanderbilt University Press (2001). [C189](#)
- [3] S. J. Ahn, W. Rauh and H.-J. Warnecke, Least-squares orthogonal distances fitting of circle, sphere, ellipse, hyperbola, and parabola, *Pattern Recognition* 34, pp. 2283–2303 (2001). [C189](#)
- [4] S. J. Ahn, E. Westkaemper and W. Rauth, Orthogonal distance fitting of parametric curves and surfaces, in *Algorithms for Approximation IV: Huddersfield 2001*, eds J. Levesley, I. Anderson and J. C. Mason, University of Huddersfield, pp. 122–129 (2002). [C189](#)
- [5] Å. Björck,  $QR$  factorization of the Jacobian in some structured nonlinear least squares problems, in *Total Least Squares and Errors-in-Variables Modeling*, eds. S. van Huffel and P. Lemmerling, Kluwer, Dordrecht, pp. 225–234 (2002). [C197](#)
- [6] P. T. Boggs, R. H. Byrd and R. B. Schnabel, A stable and efficient algorithm for nonlinear orthogonal distance regression, *SIAM J. Sci. Stat. Comp.* 8, pp. 1052–1078 (1987). [C189](#), [C197](#)
- [7] H.-P. Helfrich and D. Zwick, A trust region algorithm for parametric curve and surface fitting, *J. Comp. Appl. Math.* 73, pp. 119–134 (1996). [C189](#)

- [8] M. R. Osborne,  $V$ -invariant methods for generalised least squares problems, Research Report MRR01.024, School of Mathematical Sciences, ANU (2001). [C194](#)
- [9] I. Söderkvist, On algorithms for generalized least squares problems with ill-conditioned covariance matrices, *Comp. Stat.* 11, pp. 303–313 (1996). [C194](#)
- [10] R. Strebler, D. Sourlier and W. Gander, A comparison of orthogonal least squares fitting in coordinate metrology, in *Recent Advances in Total Least Squares and Errors-in-Variables Techniques*, ed. S. Van Huffel, SIAM, Philadelphia, pp. 249–258 (1997). [C189](#), [C197](#)
- [11] D. A. Turner, *The Approximation of Cartesian Co-ordinate Data by Parametric Orthogonal Distance Regression*, PhD Thesis, University of Huddersfield (1999). [C189](#)
- [12] 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 *Advanced Mathematical and Computational Tools in Metrology IV*, eds P. Ciarlini, A. B. Forbes, F. Pavese and D. Richter, Series on Advances in Mathematics for Applied Sciences, Volume 53, World Scientific, Singapore, pp. 246–255 (2000). [C189](#)
- [13] G. A. Watson, On the Gauss-Newton method for  $l_1$  orthogonal distance regression, *IMA J. of Num. Anal.* 22, pp. 345–357 (2002). [C191](#), [C197](#)
- [14] G. A. Watson and K. F. C. Yiu, On the solution of the errors in variables problem using the  $l_1$  norm, *BIT* 31, pp. 697–710 (1991). [C198](#)
- [15] D. S. Zwick, Applications of orthogonal distance regression in metrology, in *Recent Advances in Total Least Squares and Errors-in-Variables Techniques*, ed S. Van Huffel (SIAM, Philadelphia), pp. 265–272 (1997). [C189](#)