

# ORTHOGONAL LINEAR REGRESSION ALGORITHM BASED ON AUGMENTED MATRIX FORMULATION

Andrzej Bargiela<sup>1</sup> and Joanna K. Hartley<sup>2</sup>

Department of Computing, Nottingham Polytechnic,  
Burton St., Nottingham NG1 4BU, U.K.

**Scope and Purpose :** In this paper, a new technique for solving a multivariate linear model using the orthogonal least absolute values regression is proposed. The orthogonal least absolute values (ORLAV) regression minimises the sum of the absolute, orthogonal distance from each data point to the resulting regression hyperplane. In a large set of equations where the variables are independent of each other, it is most convenient to measure the orthogonal distances. The other advantage of ORLAV regression is that this method is insensitive to outliers - if minimising the squared distances, for example, an extreme outlier can be most significant.

**Abstract :** The problem of n-dimensional orthogonal linear regression is a problem of finding an n-dimensional hyperplane minimising the sum of Euclidean distances between this hyperplane and a given set of m points, where  $m \geq n$ . This nonlinear programming problem has been re-cast in an augmented matrix form and solved as a sequence of iteratively re-weighted least squares problems. The proposed algorithm is seen as an alternative to the recently published algorithm by Cavalier and Melloy [1].

## 1. INTRODUCTION.

The results of many scientific investigations are sets of measurement data points which are then used to establish mathematical models (hyperplanes) of physical systems or phenomena. In situations where, for each data point in a set, one of the variables is a function of the remaining variables, a convenient way of calculating the relevant hyperplane is by means of ordinary least squares regression (OLS). These calculations minimise the sum of squares of distances to the hyperplane measured along the direction of the dependent variable (Fig.1a and 1b). However, if there is no obvious functional relationship between variables, it is more appropriate to minimise the sum of squares of Euclidean distances to the hyperplane measured along the normal to the hyperplane, as illustrated in Figure 1c. Such regression is referred to as orthogonal least squares (ORLS). Mathematically the ORLS regression is a significantly more difficult problem since the normal direction to the hyperplane is not known in advance, thus giving rise to inherent nonlinearity of the problem.

Both OLS and ORLS can be shown to provide optimal unbiased regression hyperplanes, if the measurement data is affected by the Gaussian measurement noise only. However, they are less well suited to deal with data containing gross measurement errors. Since the objective function in OLS and ORLS evaluates squares of distances to the hyperplane, the outlying measurements are unduly heavily weighted. Consequently, in situations where gross measurement errors do occur, a sum of absolute values of distances is taken in preference to the sum of squares [2]. The analogs of OLS and ORLS regressions derived by substituting the  $L_2$  with the  $L_1$  norm are referred to as ordinary least absolute values (OLAV) and orthogonal least absolute values (ORLAV) regressions. The latter is also referred to in the literature as Euclidean minimum sum of absolute errors (EMSAE).

---

1. Dr. Andrzej Bargiela is a Reader in Real Time Systems at Nottingham Polytechnic. His research interests include applied optimization, parallel and distributed programming and neural computing.

2. Miss Joanna Hartley is a Research Assistant at Nottingham Polytechnic. She is registered on a PhD degree project, researching into parallel algorithms for simulation of network systems.

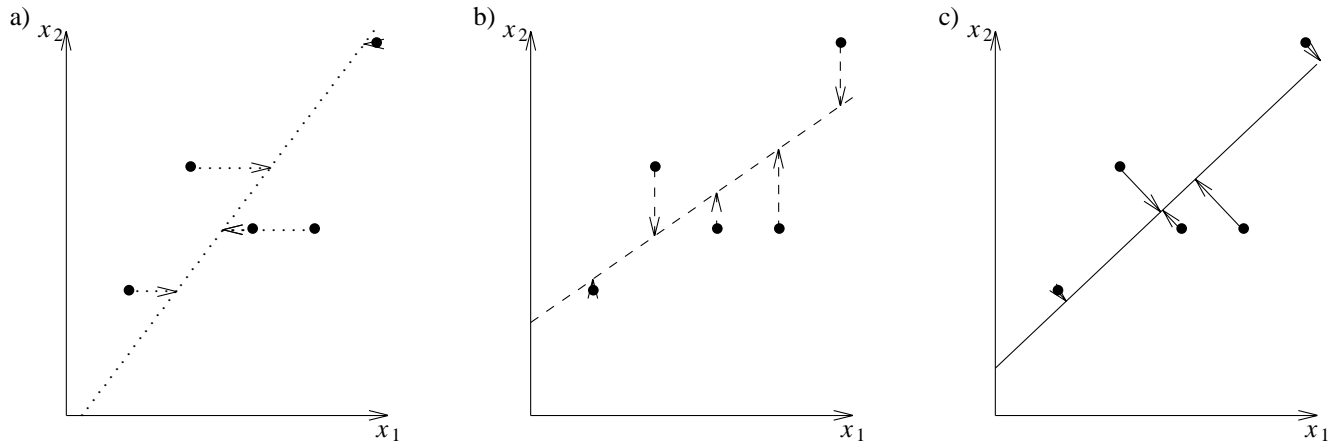


Fig. 1. Distances measured to appropriate regression line.

- a)  $\cdots$ , Least-squares regression with  $x_1$  dependent (OLS);
- b)  $- -$ , Least-squares regression with  $x_2$  dependent (OLS);
- c)  $-$ , Least-squares regression with  $x_1, x_2$  independent (ORLS).

An illustration of the error rejection capability of ORLAV regression compared to ORLS regression is given in Fig. 2. The  $L_1$  norm in the ORLAV regression ensures that, in  $n$ -dimensional Euclidean space, there are exactly  $n$  data points which are taken to span the hyperplane and the remaining  $(m-n)$  points are rejected. Therefore the outlier measurements find themselves among the rejected data points, thus having no effect on the calculated hyperplane. By contrast, the ORLS regression takes notice of the erroneous measurements, and the resulting hyperplane in Fig. 2b is different to the original one.

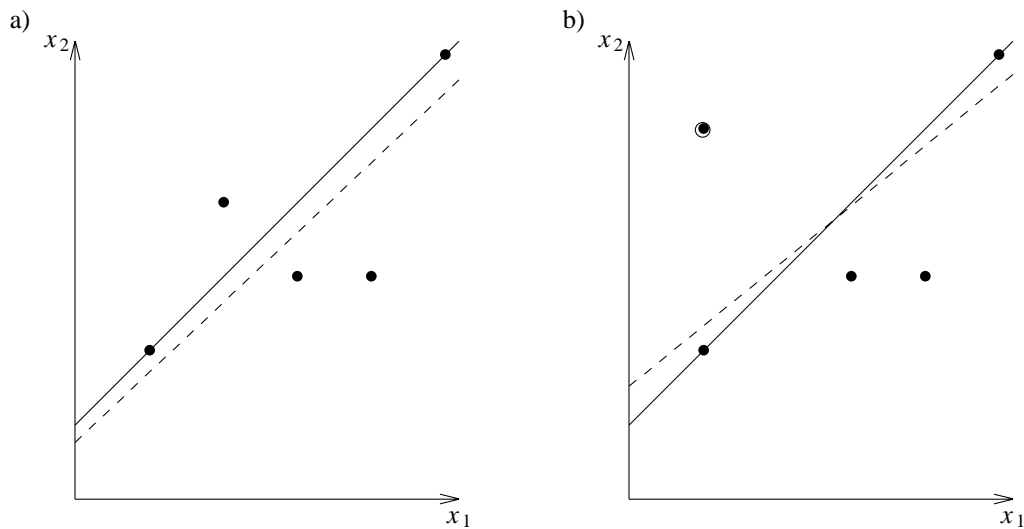


Fig. 2. Comparison of sensitivity to an outlier, of ORLAV and ORLS regression lines.

- $-$ , ORLAV regression line;
- $- -$ , ORLS regression line;
- , the outlier.

This paper is focussed on the orthogonal least absolute values (ORLAV) regression in n-dimensional Euclidean space. Our original contribution is the re-casting of this nonlinear programming problem into an augmented matrix form and solving it as a sequence of iteratively re-weighted least squares problems. The proposed algorithm is seen as an alternative to the algorithm recently proposed by Cavalier and Melloy [1].

## 2. THE ORLAV REGRESSION.

Consider the hyperplane  $\mathbf{p}'\mathbf{x}=\alpha$  along with  $\mathbf{x}_i, i=1,\dots, m$ , being the given set of points in  $E_n$ , Euclidean n-space (with  $m \geq n$ ). The normal Euclidean distance from a point  $\mathbf{x}_i$  to the hyperplane  $\mathbf{p}'\mathbf{x}=\alpha$  is given by

$$s_i = |\mathbf{p}'\mathbf{x}_i - \alpha|/\|\mathbf{p}\|. \quad (1)$$

So, the required hyperplane can be found by solving the task:

$$(\mathbf{P}) \quad \text{minimise } \sum_{i=1}^m |\mathbf{p}'\mathbf{x}_i - \alpha|/\|\mathbf{p}\| \quad (2)$$

$$\mathbf{p}, \alpha \text{ unrestricted.} \quad (3)$$

Assuming that the normal vector to the hyperplane is of unit length, the above can be rewritten as

$$(\mathbf{P1}) \quad \text{minimise } \sum_{i=1}^m |\mathbf{p}'\mathbf{x}_i - \alpha| \quad (4)$$

$$\text{subject to } \|\mathbf{p}\|=1 \quad (5)$$

$$\mathbf{p}, \alpha \text{ unrestricted.} \quad (6)$$

The nonlinearity in the objective function, due to the absolute values operator, can be dealt with by two simple alterations:

Let  $d_i = \mathbf{p}'\mathbf{x}_i - \alpha$ , so the objective function becomes  $\sum_{i=1}^m |d_i|$ , which is equivalent to  $\sum_{i=1}^m d_i^2/|d_i|$ .

Now let  $W$  be the  $m \times m$  diagonal matrix with entries  $1/|d_i|, i=1,\dots, m$ .

Also, let  $A$  be an  $m \times n$  matrix which consists of rows of  $\mathbf{x}_i$ 's, and let  $\mathbf{e}=(1, 1,\dots, 1)' \in E_m$ . Now (P1) can be rewritten as:

$$(\mathbf{P2}) \quad \text{minimise } \mathbf{d}'W\mathbf{d} \quad (7)$$

$$\text{subject to } A\mathbf{p} - \mathbf{e}\alpha - \mathbf{d} = \mathbf{O} \quad (8)$$

$$\|\mathbf{p}\|=1 \quad (9)$$

$$\mathbf{p}, \alpha, \mathbf{d} \text{ unrestricted.} \quad (10)$$

As  $\|\mathbf{p}\|=1$  in (9), then  $\mathbf{p}'\mathbf{p} = \|\mathbf{p}\|^2 = 1$  also holds. So (9) can be exchanged for  $\mathbf{p}'\mathbf{p} = 1$ , to give

$$(\mathbf{P3}) \quad \text{minimise } \mathbf{d}'W\mathbf{d} \quad (11)$$

$$\text{subject to } A\mathbf{p} - \mathbf{e}\alpha - \mathbf{d} = \mathbf{O} \quad (12)$$

$$\mathbf{p}'\mathbf{p} = 1 \quad (13)$$

$$\mathbf{p}, \alpha, \mathbf{d} \text{ unrestricted.} \quad (14)$$

Equation (13) is the only non-linear constraint. In order to linearise this constraint, we will introduce additional variable  $\mathbf{q}$  which will be iteratively updated and will represent our estimate of  $\mathbf{p}$ .

$$(\mathbf{P4}) \quad \text{minimise } \mathbf{d}'W\mathbf{d} \quad (15)$$

$$\text{subject to } A\mathbf{p} - \mathbf{e}\alpha - \mathbf{d} = \mathbf{O} \quad (16)$$

$$\mathbf{q}'\mathbf{p} = 1 \quad (17)$$

$$\mathbf{p}, \alpha, \mathbf{d} \text{ unrestricted.} \quad (18)$$

These constraints can be written in matrix form:

$$\begin{bmatrix} A & -\mathbf{e} \\ \mathbf{q}^t & 0 \end{bmatrix} \begin{bmatrix} \mathbf{p} \\ \alpha \end{bmatrix} = \begin{bmatrix} \mathbf{d} \\ 1 \end{bmatrix},$$

$$\begin{aligned} \text{let matrix } F &= \begin{bmatrix} A & -\mathbf{e} \\ \mathbf{q}^t & 0 \end{bmatrix}, \underline{\theta} = (\mathbf{p}^t, \alpha)^t, \underline{\psi} = (\mathbf{d}^t, 1)^t, \\ &\text{resulting in } F\underline{\theta} = \underline{\psi}. \end{aligned} \quad (19)$$

Equation (19), which represents constraints (16) and (17), introduces the variable  $\underline{\theta}$ ; alongside this the weight matrix  $W$  in (15) needs to be modified to give  $G = \begin{bmatrix} W & \mathbf{O} \\ \mathbf{O}^t & \frac{1}{\varepsilon} \end{bmatrix}$  ( $\frac{1}{\varepsilon}$  is the reciprocal of the computational accuracy  $\varepsilon$ ). This results in high weighing of the equation (17), which needs to be satisfied to preserve the integrity of the Euclidean distances measurements. So (P4) can be rewritten as:

$$\text{(P5)} \quad \text{minimise } \underline{\psi}^t G \underline{\psi} \quad (20)$$

$$\text{subject to } F\underline{\theta} - \underline{\psi} = \mathbf{O} \quad (21)$$

$$\underline{\theta}, \underline{\psi} \text{ unrestricted.} \quad (22)$$

Now, to find the minimum of  $\underline{\psi}^t G \underline{\psi}$ , substitute  $F\underline{\theta}$  for  $\underline{\psi}$  and then differentiate by  $\underline{\theta}$  and put the result equal to zero:

$$\text{Let } \mathbf{J} = \underline{\psi}^t G \underline{\psi} \quad (23)$$

$$= \underline{\theta}^t F^t G F \underline{\theta}, \quad (24)$$

$$\frac{d\mathbf{J}}{d\underline{\theta}} = 2F^t G F \underline{\theta} = \mathbf{O}. \quad (25)$$

Introducing temporary variables  $\underline{\psi}$  and  $\underline{\phi}$ , representing the partial vector-matrix products in (25), this minimality condition can be written as a set of simultaneous matrix equations

$$F\underline{\theta} = \underline{\psi} \quad (26)$$

$$G\underline{\psi} = \underline{\phi} \quad (27)$$

$$F^t \underline{\phi} = \mathbf{O} \quad (28)$$

or using the augmented matrix formulation:

$$\begin{bmatrix} 0 & -I & F \\ -I & G & 0 \\ F^t & 0 & 0 \end{bmatrix} \begin{bmatrix} \underline{\phi} \\ \underline{\psi} \\ \underline{\theta} \end{bmatrix} = \begin{bmatrix} \mathbf{O} \\ \mathbf{O} \\ \mathbf{O} \end{bmatrix}. \quad (29)$$

If  $\mathbf{q}$  is calculated to be, within a given computational accuracy, equal to  $\mathbf{p}$  then the problems (P3) and (P4) are identical. The initial vector for  $\mathbf{q}$  can be selected quite arbitrarily but, to enhance computational efficiency, it is a good idea to evaluate it by calculating an OLS (ordinary least squares) equation,  $\mathbf{q}^t \mathbf{x} = \beta$ , where one of the  $x_j$  is (arbitrarily) chosen as the dependent variable.

$$\begin{aligned} \text{Let } \mathbf{q}^t &= (\hat{\mathbf{q}}^t, 1) \text{ and } \mathbf{x}^t = (\hat{\mathbf{x}}^t, x_j), \\ \text{so } \mathbf{q}^t \mathbf{x} = \beta &\text{ can also be written as } \hat{\mathbf{q}}^t \hat{\mathbf{x}} - \beta = -x_j. \end{aligned}$$

The algorithm can thus be summarised as follows:

1. Calculate  $\hat{\mathbf{q}}$  of  $\hat{\mathbf{q}}^t \hat{\mathbf{x}} - \beta = -x_j$ , for a given set of points  $x_i \in E_n$ ,  $i=1, \dots, m$ . Normalise  $\mathbf{q}$ , and continue to 2.
2. Find the  $L_1$  solution to (P5) by iteratively modifying the weight matrix,  $G$ , as follows:
  - 2.1 Solve the matrix equation (29) which gives the  $L_2$  solution for  $\underline{\theta}$ ,  $\underline{\psi}$ .
  - 2.2 Using the calculated discrepancies  $\psi_i$ 's, modify the corresponding  $G_{ii}$ 's, with  $G_{ii} = \frac{1}{\psi_i}$ .
  - 2.3 If at least  $n$  of the original equations are satisfied (corresponding  $\psi_i=0$ ), continue to 3. Otherwise repeat from 2.1.
3. If  $\|\tilde{\mathbf{p}}\|=1$  and  $\tilde{\mathbf{p}}=\mathbf{q}$  stop, with  $\tilde{\mathbf{p}}$ ,  $\tilde{\alpha}$  as the optimal solution to problem (P). Otherwise, go to 4.
4. Replace  $\mathbf{q}$  by  $\tilde{\mathbf{p}}/\|\tilde{\mathbf{p}}\|$  and reset the weights to 1; then return to 2.

The algorithm proved to be very robust, and with sets of randomly generated data, with up to  $n=5$  and  $m=100$ , the algorithm was found to need a maximum of 10 iterations; with a typical convergence achieved in 2 iterations. The maximum time needed for convergence was 10.3 seconds on the SUN Sparc SLC workstation.

### 3. THE DUAL.

The dual of problem (P5) is derived using the Lagrangian:

$$L(\underline{\psi}, \underline{\theta}; \underline{\eta}) = \underline{\psi}^t G \underline{\psi} + \underline{\eta}^t (F \underline{\theta} - \underline{\psi}) \quad (30)$$

$$\frac{\partial L}{\partial \underline{\psi}} = 2\underline{\psi}^t G - \underline{\eta}^t, \quad \frac{\partial L}{\partial \underline{\theta}} = \underline{\eta}^t F \quad (31)$$

$$\frac{\partial L}{\partial \underline{\psi}} = \frac{\partial L}{\partial \underline{\theta}} = \mathbf{0}^t \Rightarrow G \underline{\psi} = \frac{1}{2} \underline{\eta} \quad (32)$$

$$F^t \underline{\eta} = \mathbf{0} \quad (33)$$

along with the constraint

$$F \underline{\theta} = \underline{\psi}. \quad (34)$$

$$\frac{\partial^2 L}{\partial \underline{\psi} \partial \underline{\theta}} = H(\underline{\theta}, \underline{\psi}) = \begin{bmatrix} 2G & 0 \\ 0 & 0 \end{bmatrix} \quad (35)$$

So  $H$  is always positive semi-definite; therefore there will be a minimum at any stationary point.

Let  $\underline{\eta} = 2\underline{\phi}$ .

So the dual of (P5) can be written as:

$$(D5) \quad \text{maximise } \underline{\phi} \quad (36)$$

$$\text{subject to } F\underline{\theta} = \underline{\Psi} \quad (37)$$

$$F' \underline{\phi} = \mathbf{0} \quad (38)$$

$$G \underline{\Psi} = \underline{\phi} \quad (39)$$

$$\underline{\theta}, \underline{\Psi} \text{ unrestricted.} \quad (40)$$

The solution of the dual problem (D5) involves, therefore, solving the same set of equations as (P5), with the variable  $\underline{\phi}$  being the dual variable. Consequently our algorithm gives simultaneous solution to the primal and dual problem.

#### 4. EXAMPLES.

Example 1.

Using the same example as Cavalier and Melloy [1] seems appropriate for this paper.

Let's find the hyperplane  $\mathbf{p}' \mathbf{x} = p_1 x_1 + p_2 x_2 = \alpha$ , which minimises the sum of the normal Euclidean distances to the points  $\mathbf{x}_1 = (1,2)^t$ ,  $\mathbf{x}_2 = (2,4)^t$ ,  $\mathbf{x}_3 = (3,3)^t$ ,  $\mathbf{x}_4 = (4,3)^t$ ,  $\mathbf{x}_5 = (5,6)^t$ . The problem (P5) is written as :

$$\text{minimise } \sum_{i=1}^5 d_i^2 / |d_i| \quad (41)$$

$$\text{subject to } p_1 + 2p_2 - \alpha - d_1 = 0 \quad (42)$$

$$2p_1 + 4p_2 - \alpha - d_2 = 0 \quad (43)$$

$$3p_1 + 3p_2 - \alpha - d_3 = 0 \quad (44)$$

$$4p_1 + 3p_2 - \alpha - d_4 = 0 \quad (45)$$

$$5p_1 + 6p_2 - \alpha - d_5 = 0 \quad (46)$$

$$q_1 p_1 + q_2 p_2 = 1 \quad (47)$$

$$\mathbf{p}, \alpha, \mathbf{d} \text{ unrestricted.} \quad (48)$$

Choosing  $x_1$  as the dependent variable, then calculating the OLS equation results in the line  $x_1 - 0.76x_2 = 0.26$ ; a choice of  $x_2$  as the dependent variable, gives  $-0.7x_1 + x_2 = 1.5$ . This system is solved using equation (29):

$$\text{with } \mathbf{q} = \begin{bmatrix} -.7 \\ 1 \end{bmatrix} \text{ being used as the initial vector,}$$

$$\text{which normalises to } \mathbf{q} = \begin{bmatrix} -.573 \\ .819 \end{bmatrix}$$

$$\text{and } F = \begin{bmatrix} 1 & 2 & -1 \\ 2 & 4 & -1 \\ 3 & 3 & -1 \\ 4 & 3 & -1 \\ 5 & 6 & -1 \\ -.573 & .819 & 0.0 \end{bmatrix} \quad \underline{\theta} = \begin{bmatrix} p_1 \\ p_2 \\ \alpha \end{bmatrix} \quad \underline{\Psi} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \\ 1 \end{bmatrix}$$

The outcome is:

$$\underline{\theta} = \begin{bmatrix} 0.7071 \\ -0.7071 \\ -0.7071 \end{bmatrix}, \quad \underline{\phi} = \begin{bmatrix} -0.5 \\ -1.0 \\ 1.0 \\ 1.0 \\ -0.5 \\ -2.8284 \end{bmatrix} \quad \text{and} \quad \underline{\psi} = \begin{bmatrix} 0.0 \\ 0.707 \\ -0.707 \\ -1.414 \\ 0.0 \\ 1.0 \end{bmatrix}$$

With either  $\mathbf{q}=(1,-0.76)^t$  or  $\mathbf{q}=(-0.7,1)^t$  as the initial vector in (47), the algorithm generates the Euclidean regression line,  $0.7071x_1 - 0.7071x_2 = -0.7071$ . Alongside this result, the algorithm determines the dual vector  $\underline{\phi}$  which can be interpreted as: the weights associated with each of the data points  $\underline{\pi}=(-0.5,-1.0,1.0,1.0,-0.5)^t$  and the sum of the normal Euclidean distances,  $\lambda=2.8284$ . So, the results confirm those of Cavalier and Melloy [1]. The algorithm took 2 iterations to achieve the solution.

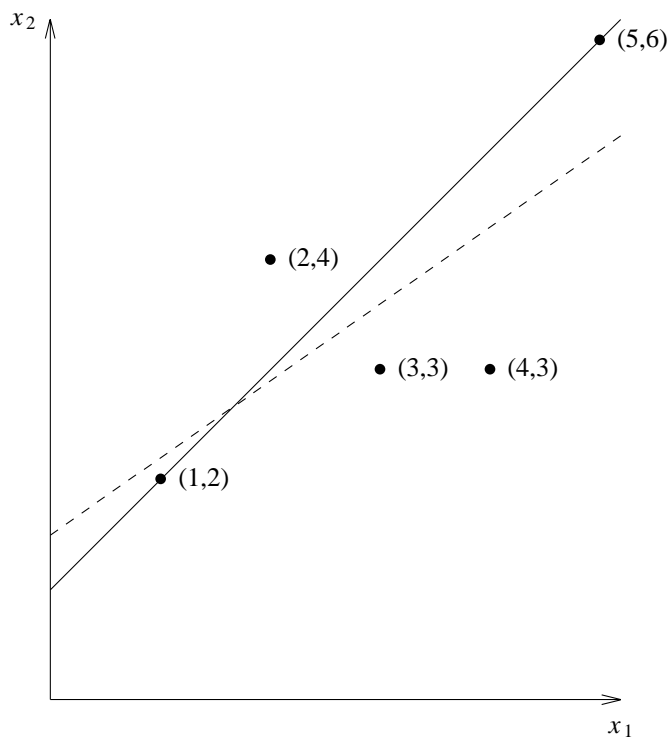


Fig. 3. Lines described in Example 1. —, Euclidean regression;  
 - - -, OLS regression with  $x_2$  dependent.

Example 2.

To illustrate the robustness of the algorithm, the starting solution was chosen to be the vector normal to the final solution vector. The solution was found in 5 iterations:

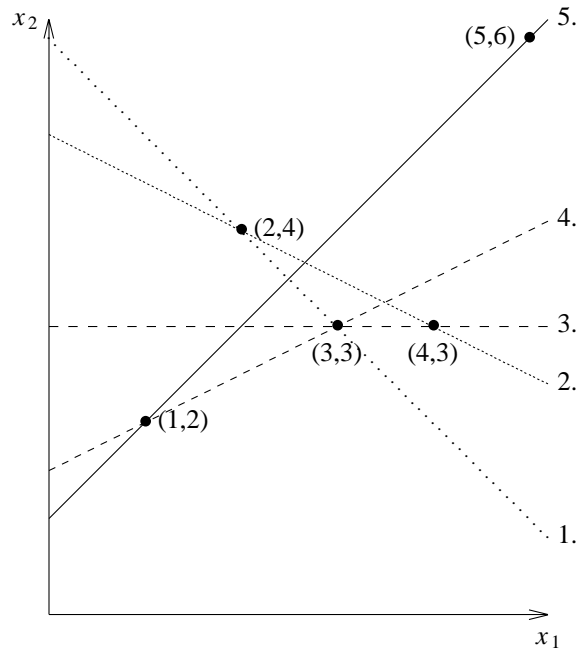


Fig. 4. Demonstrates rotation of the hyperplane until eventual solution is found.

Figure 4 illustrates the progression of the solution hyperplane from 1. through to 5. and emphasises the robustness of the algorithm.

The 1st. iteration determines the initial line through (2,4) and (3,3).

The 2nd. iteration is a rotation of the hyperplane around (2,4) and passing through (4,3).

The 3rd. iteration is a rotation of the hyperplane around (4,3) and passing through (3,3).

The 4th. iteration is a rotation of the hyperplane around (3,3) and passing through (1,2).

The 5th. iteration is a rotation of the hyperplane around (1,2) and passing through (5,6).

## 5. CONCLUSIONS.

A new technique for solving n-dimensional orthogonal linear regression (ORLAV) has been proposed in this paper. The ORLAV regression is a nonlinear, nonconvex programming problem. It has been re-cast in an augmented matrix form and solved as a sequence of iteratively re-weighted least squares problems. The algorithm is numerically robust (the augmented matrix preserves the condition number of the constraints matrix) and it gives a simultaneous solution to the primal and dual problems.

## ACKNOWLEDGEMENTS

The financial support of the Science and Engineering Research Council for Miss Joanna Hartley is gratefully acknowledged.

The authors wish to acknowledge the contribution of the East Worcestershire Waterworks Company in form of a loan of computing hardware.

## REFERENCES

1. T.M.Cavalier and B.J.Melloy, An Iterative Linear Programming solution to the Euclidean Regression Model. *Computers Ops Res.* **18**, 655-661 (1991).
2. R.L.Branham Jr, Alternatives to least-squares. *Astr. J.* **87**, 928-937 (1982).