

On the Relationship Between Regression Analysis and Mathematical Programming

DONG QIAN WANG[†]
*School of Mathematical and Computing Sciences
Victoria University of Wellington, NZ*

dong.wang@mcs.vuw.ac.nz

STEFANKA CHUKOVA
*School of Mathematical and Computing Sciences
Victoria University of Wellington, NZ*

stefanka.chukova@mcs.vuw.ac.nz

C. D. LAI
*Institute of Information Sciences and Technology
Massey University, Palmerston North, NZ*

c.lai@massey.ac.nz

Abstract. The interaction between linear, quadratic programming and regression analysis are explored by both statistical and operations research methods. Estimation and optimization problems are formulated in two different ways: on one hand linear and quadratic programming problems are formulated and solved by statistical methods, and on the other hand the solution of the linear regression model with constraints makes use of the simplex methods of linear or quadratic programming. Examples are given to illustrate the ideas.

Keywords: Regression analysis, linear programming, simplex method, two-phase methods, least squares method, quadratic programming and artificial variable.

1. Introduction

We will discuss the interaction between linear, quadratic programming and regression analysis. These interactions are considered both from a statistical point of view and from an optimization point of view. We also examine the algorithms established by both statistical and operations research methods. Minimizing the sum of the absolute values of the regression has shown that it can be reduced to a general linear programming problem (Wager, 1969) and Wolfe (1959) hinted that his method can be applied to regression but no analysis is done. Estimation and optimization problems are formulated in two different ways: on one hand linear and quadratic

[†] Requests for reprints should be sent to Dong Qian Wang, School of Mathematical and Computing Sciences, Victoria University of Wellington, P.O.Box 600, Wellington, NZ.

programming problems are formulated and solved by statistical methods, and on the other hand the solution of the linear regression model with constraints makes use of the simplex methods of linear or quadratic programming. Examples are given to show some practical applications of the methods. It is our aim that students taking a linear or non-linear programming (NLP) courses as well as a course in linear models will now realize that there is a definite connection between these problems.

1.1. Regression models

Consider the linear regression (LR) model with nonnegative constraints

$$\begin{aligned} Y(\beta) &= X\beta + \epsilon \\ \text{subject to } \beta &\geq 0, \end{aligned} \quad (1)$$

where $Y \in \mathbf{R}^n$ represents the vector of responses, X is an $n \times p$ design matrix, $\beta \in \mathbf{R}^p$ represents the unknown parameters of the model and $\beta \geq 0$ means that all the elements in the vector are non-negative, and ϵ represents the random error term of the LR model.

A general linear regression model (1) with inequality constraints (LRWC) and nonnegative variables is given as follows:

$$\begin{aligned} Y(\beta) &= X\beta + \epsilon \\ \text{subject to } A\beta &\geq C \\ \beta &\geq 0, \end{aligned} \quad (2)$$

where $\beta \in \mathbf{R}^p$ is the unknown vector; $X_{n \times p}$ ($n \geq p$) and $A_{s \times p}$ ($s \leq p$) are constant matrices, $Y_{n \times 1}$, $C_{s \times 1}$, and $\epsilon_{n \times 1}$ are column vectors, $\epsilon \sim N(0, \sigma^2 I)$; $X^T X \geq 0$ and $\text{rank}(A) = s$. The solution of LRWC is a subject of the Karush-Kuhn-Tucker theorem. The subsequent algorithms for solving LRWC are discussed in Lawson and Hanson (1974) and Whittle (1971).

1.2. Mathematical programming

Due to the fact that $\max\{f(x)\} = \min\{-f(x)\}$, we will focus our attention only on *minimization* problems. A primal linear programming (LP) problem with nonnegative solution can be formulated as follows:

$$\begin{aligned} \text{minimize } L(\beta) &= b^T \beta \\ \text{subject to } G\beta &\geq f \\ \beta &\geq 0, \end{aligned} \quad (3)$$

where β is the unknown vector, $0 \neq b^T \in \mathbf{R}^p$, $f^T \in \mathbf{R}^m$ and $G_{m \times p}$ are known constant vectors and a matrix, respectively.

A quadratic programming (QP) problem in which all the variables must be nonnegative is formulated as follows:

$$\begin{aligned} \text{minimize } & Z_0(\beta) = b^T \beta + \beta^T D \beta \\ \text{subject to } & A \beta \geq C \\ & \beta \geq 0, \end{aligned} \quad (4)$$

where $A_{s \times p}$, ($s \leq p$), $D_{p \times p}$ are matrices; $C_{s \times 1}$, $\beta_{p \times 1}$ and $b_{p \times 1}$ are column vectors, $\text{rank}(A) = s \leq p$ and D is symmetric and positive definite matrix.

In the next section, we further explore the relationships between the above four models. The aim in this note is to provide a strong link and algorithms between these concepts. In fact they are equivalent in some cases. Parameter estimates of models (1) and (2) are obtained by the simplex method and the Karush-Kuhn-Tucker theorem. The optimization problems of models (3) and (4) are restated and solved by statistical methods.

2. Solving Linear Regression Model by Using Mathematical Programming

Consider given n observations $\{(x_i, y_i), i = 1, 2, \dots, n\}$. The linear regression model (1) with $p = 2$, can be rewritten as $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ and $\beta = (\beta_1, \beta_2)^T \geq 0$. Using the least squares method to estimate parameters β_0 and β_1 , we need to minimize $g(\beta)$, i.e

$$\begin{aligned} \text{minimize } & g(\beta) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \\ \text{subject to } & \beta \geq 0. \end{aligned}$$

The associated system of normal equations is given as follows:

$$\begin{aligned} n\beta_0 + \beta_1 \sum_{i=1}^n x_i - \sum_{i=1}^n y_i &= 0, \\ \beta_0 \sum_{i=1}^n x_i + \beta_1 \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i y_i &= 0. \end{aligned}$$

Therefore, model (1) is equivalent to the following mathematical programming problem:

$$\begin{aligned}
& \text{minimize} && g(\beta) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \\
& \text{subject to} && n\beta_0 + \beta_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \\
& && \beta_0 \sum_{i=1}^n x_i + \beta_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i \\
& && \beta \geq 0.
\end{aligned} \tag{5}$$

We will use the phase I in the two-phase version of the simplex method to solve (5). The problem to be solved by phase I is

$$\begin{aligned}
& \text{maximize} && Y_0^I = -R_1 - R_2 \\
& \text{subject to} && n\beta_0 + c_1\beta_1 + R_1 = c_2 \\
& && c_1\beta_0 + b_1\beta_1 + R_2 = b_2 \\
& && \beta \geq 0 \text{ and } R_1, R_2 \geq 0;
\end{aligned} \tag{6}$$

where R_1 and R_2 are artificial variables. The optimization problem (6) can be rewritten as

$$\begin{aligned}
& \text{maximize} && Y_0^I = (n + c_1)\beta_0 + (b_1 + c_1)\beta_1 - (b_2 + c_2) \\
& \text{subject to} && n\beta_0 + c_1\beta_1 + R_1 = c_2 \\
& && c_1\beta_0 + b_1\beta_1 + R_2 = b_2 \\
& && \beta_i \geq 0 \text{ and } R_i \geq 0, \quad i = 1, 2;
\end{aligned} \tag{7}$$

where $c_1 = \sum x_i$, $c_2 = \sum y_i$, $b_1 = \sum x_i^2$, and $b_2 = \sum x_i y_i$. The initial values are summarized in the Table 1.

Table 1. The initial values of problem (7).

BV	β_0	β_1	R_1	R_2	RHS
Y_0^I	$-n - c_1$	$-b_1 - c_1$	0	0	$-b_2 - c_2$
R_1	n	c_1	1	0	c_2
R_2	c_1	b_1	0	1	b_2

Bearing in mind that the solution of the LR problem is actually the solution of the corresponding system of normal equations, it is now easy

to see that problem (7) is equivalent to solving related the LR problem (1). Hence we can obtain the optimal solution for model (1) by using the simplex method for problem (7) with the initial values in Table 1. The above approach can be applied for solving linear regression model (1) with $p > 2$, i.e.,

$$Y_i = \beta_0 + \beta_1 x_i^{(1)} + \beta_2 x_i^{(2)} + \dots + \beta_p x_i^{(p)} + \epsilon_i \quad (8)$$

subject to $\beta \geq 0$.

Next, we will illustrate the above ideas by an example.

Example 2.1: (Rencher, 2000, p113-114) The exam scores y and homework scores x (average value) for 18 students in a statistics class were as follows

\mathbf{x}	96	77	0	0	78	64	89	47	90
y	95	80	0	0	79	77	72	66	98
\mathbf{x}	93	18	86	0	30	59	77	74	67
y	90	0	95	35	50	72	55	75	66

From the given data set, we obtain: $c_1 = \sum x_i = 1,045$, $c_2 = \sum y_i = 1,105$, $b_2 = \sum x_i y_i = 81,195$, $b_1 = \sum x_i^2 = 80,199$ and $n = 18$. Hence, Table 1 becomes

BV	β_0	β_1	R_1	R_2	RHS	Ratio
Y_0^I	-1063	-81244	0	0	-82300	–
R_1	18	1045	1	0	1105	1.057
R_2	1045	80199	0	1	81195	1.012

Using the simplex method, the optimal table is obtained in two iterations as:

BV	β_0	β_1	RHS	Ratio
Y_0^I	0	0	0	–
β_0	1	0	10.727	–
β_1	0	1	0.873	–

Therefore, $\hat{\beta}_0 = 10.727$, $\hat{\beta}_1 = 0.873$ and a fitted linear regression line is given by $\hat{y} = 10.727 + 0.873x$.

3. Solving the Least Squares Problem With Constraints Using NLP Methods

Using the least squares method to model (2) we obtain a general regression problem:

$$\begin{aligned} \min \quad & Z(\beta) = (Y - X\beta)^T(Y - X\beta) \\ \text{subject to} \quad & A\beta \geq C \\ & \beta \geq 0. \end{aligned} \quad (9)$$

Problem (9) is a simultaneous quadratic programming problem and thus it can be solved by using Wolfe's method based on Karush-Kuhn-Tucker conditions. Rewriting problem (9) as a quadratic programming problem leads to:

$$\begin{aligned} \min \quad & Z_0(\beta) = a + n\beta_0^2 + b_1\beta_1^2 - 2c_2\beta_0 - 2b_2\beta_1 + 2c_1\beta_0\beta_1 \\ \text{subject to} \quad & A\beta \geq C \\ & \beta \geq 0, \end{aligned} \quad (10)$$

where, as in the previous example, $a = \sum y_i^2$, $c_1 = \sum x_i$, $c_2 = \sum y_i$, $b_2 = \sum x_i y_i$ and $b_1 = \sum x_i^2$.

Example 3.1: Use the given set of data to evaluate the parameters of a simple linear regression model with additional restrictions imposed on the parameters of the model, i.e.,

$$\begin{aligned} 2\beta_0 + \beta_1 &\geq 650 \\ -2\beta_0 + \beta_1 &\geq 500 \\ \beta_i &\geq 0 \quad i = 1, 2. \end{aligned}$$

x	.055	.091	.138	.167	.182	.211	.232	.248	.284	.351
y	90	97	107	124	142	150	172	189	209	253

From the given data set, we can calculate the values of $a = 259993$, $c_1 = \sum x_i = 1.959$, $c_2 = \sum y_i = 1533$, $b_2 = \sum x_i y_i = 341.68$, $b_1 = \sum x_i^2 = .4551$ and $n = 10$. Parameter estimates of the model obtained by the standard regression technique are $\beta_0 = 39.6484$ and $\beta_1 = 580.151$.

Let us solve the same problem by employing nonlinear programming ideas. Firstly, we have to rewrite the problem in the form of quadratic programming by using the previously calculated values of b_1, b_2, c_1 and c_2 . We have:

$$\begin{aligned}
\min Z_0 &= 259993 + 10\beta_0^2 + .4551\beta_1^2 - 3066\beta_0 - 683.36\beta_1 + 3.918\beta_0\beta_1 \\
\text{subject to} \quad & 2\beta_0 + \beta_1 \geq 650 \\
& -2\beta_0 + \beta_1 \geq 500 \\
& \beta_i \geq 0 \quad i = 1, 2.
\end{aligned}$$

Then, solving the above quadratic programming problem with Wolfe's method confirms the optimal values of $\beta_0 = 39.6484$ and $\beta_1 = 580.151$.

4. Solving QP Problem Using the Least Squares Method

The relationship between the quadratic programming (4) and the least squares method (9) is studied by Wang, Chukova and Lai (2003),

Theorem 1: The relationship of QP (4) and LS (9) is given by

$$Z_0(\beta) = \frac{1}{4}b^T D^{-1}b + Z(\beta),$$

where X is a real upper triangular matrix with positive diagonal elements satisfying $X^T X = D$ and $Y = -\frac{1}{2}(X^T)^{-1}b$.

Hence, minimizing Z_0 is equivalent to minimizing $Z(\beta)$. Moreover, when $b = 0$, we have

$$Z_0(\beta) = Z(\beta) = (Y - X\beta)^T(Y - X\beta).$$

Let us consider the least squares problem similar to (9) where all the constraints are in the form of equality, i.e $A\beta = C$. Using the Lagrangian method, we obtain the corresponding normal equation:

$$\begin{aligned}
A\beta &= C \\
A^T\lambda + X^T X\beta &= X^T Y.
\end{aligned} \tag{11}$$

Theorem 2: Let $\hat{\beta}^*$ be the solution of (11) and $\hat{\beta}_0$ be the solution of linear regression model with no constraints. Then, the relationship between $\hat{\beta}^*$ and $\hat{\beta}_0$ is:

$$\hat{\beta}^* = [I - (X^T X)^{-1}A^T H^{-1}A]\hat{\beta}_0 + (X^T X)^{-1}A^T H^{-1}C$$

where $H = A(X^T X)^{-1}A^T$ is a hat matrix (Sen 1990).

Based on Theorem 1 and Theorem 2, Wang, Chukova and Lai (2003) developed a stepwise algorithm for reducing and solving QP problem (4)

with regression analysis. The following example illustrates the algorithm.

Example 4.1: Consider

$$\begin{aligned} \min Z_0 = & -2x_1 - 3x_2 + x_1^2 + x_2^2 + x_1x_2 \\ \text{subject to} & \quad 2x_1 + 2x_2 \leq 2 \\ & \quad 3x_1 - 2x_2 \geq 1 \\ & \quad x_i \geq 0 \quad i = 1, 2. \end{aligned}$$

The solution to this QP by using Wolfe's method is found to be $Z^T = (x_1, x_2) = (\frac{3}{4}, \frac{5}{8})$. Let us apply our algorithm for reducing the above QP to LS. We have

$$\min Z_0 = (-2 \ -3) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + (x_1 \ x_2) \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$\text{subject to} \begin{pmatrix} -1 & -2 \\ 3 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \geq \begin{pmatrix} -2 \\ 1 \end{pmatrix}$$

$$\text{and} \quad x_i \geq 0, \quad i = 1, 2.$$

The above is a matrix representation of model (4). Let

$$A = \begin{pmatrix} -1 & -2 \\ 3 & -2 \end{pmatrix} = \begin{pmatrix} a_1^T \\ a_2^T \end{pmatrix}, \quad C = \begin{pmatrix} -2 \\ 1 \end{pmatrix},$$

and

$$D = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}, \quad b = \begin{pmatrix} -2 \\ -3 \end{pmatrix}.$$

1. Find the matrices X and Y and convert a QP problem to a LS problem.

$$X = \text{Choleski}(D) = \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & .866 \end{pmatrix}$$

and

$$Y = -\frac{1}{2}(X^T)^{-1}b = (1, 1.1547)^T.$$

2. Solve LS $\min Q(\beta) = (Y - X\beta)^T(Y - X\beta)$ over R_+^2 .
The solution is $\beta^* = (\frac{1}{3}, \frac{4}{3})^T$.

3. Verify whether $A\beta^* \geq C$. In this case both conditions in model (4) are not satisfied. Thus we have to solve the following two problems:

- First, we solve

$$\begin{array}{ll} \min & Q(\beta^{(1)}) = (Y - X\beta^{(1)})^T(Y - X\beta^{(1)}) \\ \text{subject to} & -\beta_1^{(1)} - 2\beta_2^{(1)} = -2 \\ & \beta_i^{(1)} \geq 0, i = 1, 2. \end{array}$$

and obtain

$$\beta^{(1)} = \left(\frac{1}{3}, \frac{5}{6} \right)^T.$$

- Then solve

$$\begin{array}{ll} \min & Q(\beta^{(2)}) = (Y - X\beta^{(2)})^T(Y - X\beta^{(2)}) \\ \text{subject to} & 3\beta_1^{(2)} - 2\beta_2^{(2)} = 1 \\ & \beta_i^{(2)} \geq 0, i = 1, 2. \end{array}$$

and obtain

$$\beta^{(2)} = (.894739, .8421085)^T.$$

4. Verify whether $A\beta^{(1)} \geq C$ or $A\beta^{(2)} \geq C$.

It is easy to check that the constraint $A\beta^{(i)} \geq C$, for $i = 1, 2$ is not satisfied. Hence, we solve for

$$\begin{array}{ll} \min & Q(\beta^{(1,2)}) = (Y - X\beta^{(1,2)})^T(Y - X\beta^{(1,2)}) \\ \text{subject to} & -\beta_1^{(1,2)} - 2\beta_2^{(1,2)} = -2 \\ & 3\beta_1^{(1,2)} - 2\beta_2^{(1,2)} = 1 \\ & \beta_i^{(1,2)} \geq 0, i = 1, 2, \end{array}$$

which gives

$$\beta^{(1,2)} = \left(\frac{3}{4}, \frac{5}{8} \right)^T.$$

5. Verify whether $A\beta^{(1,2)} \geq C$.

The constraint $A\beta^{(1,2)} \geq C$ is satisfied. Thus, the the optimal solution is

$$\hat{\beta} = \beta^{(1,2)} = \left(\frac{3}{4}, \frac{5}{8} \right)^T.$$

The above solution confirms the previous result obtained by Wolfe's method.

5. Conclusions

A linear regression model is solved by two-phase version of the simplex method. A statistical algorithm to solve the quadratic programming problem is proposed. In comparison with the nonlinear programming methods for solving QP, our algorithm has the following advantages:

- (a) Statistical courses often form the core portion for most degree programs at bachelor level. The algorithm based on basic statistical concepts is easy to understand, learn and apply.
- (b) Some of the steps of the algorithm are included as built-in functions or procedures in many of the commonly used software packages like MAPLE, MATHEMATICA and so on.
- (c) The algorithm avoids the usage of slack and artificial variables.

References

1. W. Kaplan. *Maximum and minima with applications, practical optimization and duality*. John Wiley and Sons Ltd, 1999.
2. C. L. Lawson and R. J. Hanson. *Solving least squares problems*. John Wiley and Sons Ltd, 1974.
3. A. C. Rencher. *Linear models in statistics*. Wiley-Interscience Publications, John Wiley and Sons Ltd, 2000.
4. A. N. Sadoski. *L1-norm fit of a straight line: algorithm AS74*. Applied Statistics, 23: 244-248, 1974.
5. A. Sen and M. Srivastava. *Regression Analysis*. Springer - Verlag, 1990.
6. D. Q. Wang, S. Chukova and C. D. Lai. *Reducing quadratic programming problem to regression analysis: stepwise algorithm*. To appear in European Journal of Operational Research, 2003.
7. P. Whittle. *Optimization under constraints*. Wiley-Interscience Publications, John Wiley and Sons Ltd, 1971.
8. W. L. Winston. *Operations research, with applications and algorithms*. Duxbury Press, 1994.