

# An Application of the ABS Algorithm for Modeling Multiple Regression on Massive Data, Predicting the Most Influencing Factors

Soniya Lalwani<sup>1\*</sup>, M. Krishna Mohan<sup>1</sup>, Pooran Singh Solanki<sup>1</sup>, Sorabh Singhal<sup>1</sup>, Sandeep Mathur<sup>3</sup>, Emilio Spedicato<sup>2</sup>

<sup>1</sup>Advanced Bioinformatics Centre, Birla Institute of Scientific Research, Jaipur, India
 <sup>2</sup>Department of Operation Research, University of Bergamo, Bergamo, Italy
 <sup>3</sup>Department of Endocrinology, SMS Medical College and Hospital, Jaipur, India
 Email: \*slalwani.math@gmail.com

Received April 17, 2013; revised May 17, 2013; accepted May 27, 2013

Copyright © 2013 Soniya Lalwani *et al.* This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

# ABSTRACT

Linear Least Square (LLS) is an approach for modeling regression analysis, applied for prediction and quantification of the strength of relationship between dependent and independent variables. There are a number of methods for solving the LLS problem but as soon as the data size increases and system becomes ill conditioned, the classical methods become complex at time and space with decreasing level of accuracy. Proposed work is based on prediction and quantification of the strength of relationship between sugar fasting and Post-Prandial (PP) sugar with 73 factors that affect diabetes. Due to the large number of independent variables, presented problem of diabetes prediction also presented similar complexities. ABS method is an approach proven better than other classical approaches for LLS problems. ABS algorithm has been applied for solving LLS problem. Hence, separate regression equations were obtained for sugar fasting and PP severity.

Keywords: ABS Algorithm; Linear Least Square; Regression; Diabetes; Huang Algorithm

# **1. Introduction**

ABS methods were developed by Abaffy, Broyden and Spedicato in 1984. ABS methods are a large class of algorithms for solving continuous and discrete linear algebraic equations, and nonlinear continuous algebraic equations, with applications to optimization problems. ABS methods are being found pertinent in application areas due to their suitability for ill conditioned and rank deficient problems [1-3]. Presented work is based on one more application of ABS algorithm because of its proven efficiency towards solution of Linear Least Square (LLS) problems. This section shows the basic structure of the manuscript that starts with the information about Diabetes followed by explanation of LLS method, Regression analysis and ABS method, respectively.

## 1.1. Diabetes

Diabetes is a chronic disease causing high levels of sugar

\*Corresponding author.

in the blood due to under production or resistance to the production of insulin hormone in pancreas, as insulin control the blood sugar in the body.

This disease occurs when the food taken and sugar produced by it does not get stored into fat, liver or in muscles for energy production and they remain in the blood and come out of the body unused because either pancreas does not make enough insulin or cells do not respond to the insulin properly or both.

There are two main types of diabetes: Type 1 Diabetes that occur in children, teens and young adults, in this type the body either does not or makes very little insulin and so daily injections of insulin are needed; Type 2 Diabetes is characterized by high blood glucose in the context of insulin resistance and relative insulin deficiency mainly in adulthood. However, due to high obesity, teens and young adults are also now diagnosed with this disease [4].

According to World Health Statistics report 2012 global average prevalence of diabetes worldwide is 1 in each 10 people [5]. With reference to [6], Asian countries con-

tribute to more than 60% of the world's diabetic population. India, Nepal and China stand at the top three positions at increasing rural diabetic prevalence. These figures show the growing risk of diabetes for Asian and other countries. Present study has taken clinical data of type 2 diabetic patients with 73 different parameters as independent variables.

#### 1.2. Linear Least Square

LLS methods are used to compute estimations of parameters and to fit data where parameters are estimated by minimizing the sum of square deviations between data and model. LLS model can be used directly, or with an appropriate data. LLS regression can be used to fit it in the form with any function:

$$f(\mathbf{x};\boldsymbol{\beta}) = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \boldsymbol{x}_1 + \boldsymbol{\beta}_2 \boldsymbol{x}_2 + \dots + \boldsymbol{\beta}_n \boldsymbol{x}_n \tag{1}$$

in which,

1) Each explanatory variable in the function is multiplied by an unknown parameter,

2) At most one unknown parameter exists with no corresponding explanatory variable,

3) All of the individual terms are summed to produce the final function value.

There are different methods for solving LLS problems such as Normal equations method, in which Cholesky factorization is used to obtain a solution to systems of normal equations, another method is QR factorization, in which LLS problem *i.e.* Ax = b is converted into traingular least square problem seeking an orthogonal matrix. Next method is Householder QR factorization that is used to compute the QR factorization of a matrix by using Householder transformations that annihilate the sub diagonal entries of each successive column. Next method is Modified Gram-Schmidt algorithm [7,8].

The LLS method is used in various fields like in Linear and Multiple Regression, ANOVA and Goodness of Fit, The coefficient of Determination, Modelling workhorse, Process modelling tool, Polynomial fitting and Numerical smoothing and differentiation. Here the LLS method is applied for Multiple Regression.

#### 1.3. ABS Algorithm

ABS algorithm is based on the initials of Abaffy, Broyden and Spedicato, introduced in 1984 for solving determined and underdetermined linear system, which was later used in LLS, nonlinear equations, optimization problems, integer (Diophantine) equations and LP problems. The ABS algorithm works faster on vector parallel machines and is more implementable in stable way, more accurate than the traditional algorithm and moreover it unifies most algorithms for solving the linear system [9, 10]. Basic ABS algorithm for solving the following linear system is defined by:

$$A^{\mathrm{T}}x = b, x \in \mathbb{R}^{n}, b \in \mathbb{R}^{m}, A^{\mathrm{T}} \in \mathbb{R}^{m,n}$$

$$\tag{2}$$

or equivalently

$$r(x) \equiv A^{\mathrm{T}} x - b = 0, r \in \mathbb{R}^{m}$$
(3)

is based on the following procedure:

a) Assign an arbitrary  $x_1 \in \mathbb{R}^n$ , an arbitrary positive definite matrix  $H_1 \in \mathbb{R}^{n,n}$ ; set k = 1;

b) Compute the residual  $r_i = r(x_i)$ ; if  $r_i = 0$  stop,  $x_i$  is a solution; otherwise proceed;

c) Compute the search vector by the following formula  $p_i = H_i^T z_i$  where  $z_i \in R^n$  is an arbitrary parameter vector, save for the well definiteness condition

$$z_i^{\mathrm{T}} H_i A v_i = p_i^{\mathrm{T}} A v_i \neq 0 \tag{4}$$

where  $v_i$  is the  $i^{\text{th}}$  column of an arbitrarily assigned nonsingular matrix  $\mathbf{V} \in \mathbf{R}^{m,m}$ .

d) Update the estimate of the solution by the following formula

$$x_{i+1} = x_i - \left(v_i^{\mathrm{T}} r_i\right) / \left(p_i^{\mathrm{T}} A v_i\right) p_i$$
(5)

e) Update the matrix  $H_i$  by the following formula

$$H_{i+1} = H_i - H_i A v_i w_i^{\mathrm{T}} H_i \tag{6}$$

where  $w_i \in \mathbb{R}^n$  is an arbitrary parameter save for the well-definiteness condition  $w_i^{\mathrm{T}} H_i A v_i = 1$ .

The basic form of ABS algorithm computes a solution  $x^+$  of the system Ax = b of *m* linear equations in *n* unknowns  $(m \le n)$  since the  $(m+1)^{\text{th}}$  iterate of a sequence of approximation  $x^i$  to  $x^+$  have following property as the approximation  $x^{i+1}$  obtained at the *i*<sup>th</sup> iteration in a solution of the *i* equations. Linear system were earlier solved by many approximation method and the oldest was the escalator method whose  $x^i$  iterates generated shows similarity with LU algorithm of ABS class when this is started with the zero vector as initial estimate for  $x^+$ .

#### 1.4. Regression and Multiple Regression

Regression analysis is used for modeling and analysis of the numerical data consisting of the values of one dependent variable and one or more independent variables. The dependent variable in the regression equation is modeled as a function of the independent variables, corresponding parameters "constants" and an "error term". The error term is treated as a "random variable". It represents unexplained variation in the dependent variable. The parameters are estimated to give a "best fit" of the data. Most commonly the best fit is evaluated by using the least squares method. Regression can be used for prediction (including forecasting of time-series data), inference and hypothesis testing, and modeling of causal relationships [11]. Multiple Regression equation has the form as earlier shown in Equation (1) for *n* number of independent variables. The coefficient  $\beta_0$  represents the constant coefficient, the predicted value of *y* when

 $x_1 = x_2 = x_3 = \dots = x_n = 0$ , which is calculated by:

$$\beta_0 = \overline{y} - \beta_1 \overline{x}_1 - \beta_2 \overline{x}_2 - \beta_3 \overline{x}_3 - \dots - \beta_n \overline{x}_n \tag{7}$$

 $\overline{x}$  and  $\overline{y}$  are the arithmetic means of x and y respectively.

 $\beta_1, \beta_2, \beta_3, \beta_4, \dots, \beta_n$  are the constants, called as the regression coefficients [12]. These coefficients are important in the sense that they act as the weight of each independent variable in the equation and in the prediction of dependent variable *i.e.* the strength of relation.

#### 2. Materials and Methods

LLS problems in this proposed method have been solved by ABS class using modified Huang Algorithm applied for solving the linear systems.

### 2.1. Huang Algorithm

This method belongs to both the basic ABS class and the optimally scaled subclass. It can be obtained by the parameter choices:  $H_1 = I$ ,  $z_i = w_i = a_i$ ,  $v_i = e_i$  in above algorithm. A mathematically equivalent, in the sense of generating the same iterates in exact arithmetic, but numerically more stable form is the modified Huang method, which is based upon formulas:

$$p_i = H_i (H_i a_i) \text{ and } H_{i+1} = H_i - p_i p_i^{\mathrm{T}} / p_i^{\mathrm{T}} p_i$$
 (8)

The method generates orthogonal search vectors and the algorithm can solve a system of linear inequalities where in a finite number of steps it either proves that no feasible point exists or it finds the solution of the least Euclidean norm. The algorithm works on symmetric algorithm which corresponds on various parameters in ABS class like  $H_1$  symmetric positive definite and

$$z_i = a_i, w_i = \frac{a_i}{a_i^{\mathrm{T}} H_i a_i} \tag{9}$$

and in Huang algorithm corresponds  $H_1 = I$  in symmetric algorithm. There are various versions of Huang algorithm according to various parameter choices and alternative formulations to compute *p* vectors and update *H*-matrices and the one used here takes:

$$H_i p_i = H_i H_i p_i = H_i a_i = p_i \tag{10}$$

which gives

$$H_{i} = \left(I - \frac{p_{i-1}p_{i-1}^{\mathrm{T}}}{p_{i-1}^{\mathrm{T}}a_{i-1}}\right)H_{i-1}.$$
 (11)

Therefore, by putting  $p_j^i = H_i a_j$ , the recurrence relation  $p_i = p_i^i$  to compute *p* vectors is obtained. Here, in

Copyright © 2013 SciRes.

the  $i^{th}$  stage, the *m*-*i* vectors must be stored. In particular, this version is adequate for pivoting [13]. The time and space complexities of all the methods implied with ABS algorithm is discussed in [14] whereas modified Huang algorithm that has been used here, is found performing well at time and memory requirements.

# 2.2. Why ABS Class for Solving Linear Least Square Problem?

ABS algorithm has shown their superiority while comparing modified Huang methods, Gram-Schmidt algorithm, the iterated stabilized Gram-Schmidt algorithm, QR algorithm [14]. In [15] a comparison with codes in the NAG, LINPACK and LAPACK libraries shows that ABS method are of comparable accuracy and are faster on vector/parallel machines. The testing had been done on 243 well-conditioned, 45 ill conditioned and 117 rank deficient problems. The superiority was evident for rank deficient and ill conditioned problems. Remarkable superiority had been obtained with respect to the relative error and accuracy [16]. A number of ABS method based approaches are discussed and used various ways of applying these algorithms for the solution of LLS problems in [17].

The testing was done on 4 families (A, B, C and D) of well-conditioned, ill conditioned, full rank, rank deficient and a few specific benchmark problems. The accuracy was specially obtained over NAG QR code [18] based upon QR factorization via Householder rotations. Hence due to the notable better computational performance, ABS algorithm with modified Huang method with LLS problem in [19] is found suitable for solving proposed problem. **Table 1** shows the comparison of ABS algorithm with NAG code where EX stands for relative error in the solution, ERES stands for relative error in the residual, M-flops stands for Mega flops and *m* and *n* show the number of rows and columns respectively [15].

These results are one of the examples of ABS algorithm efficiency in terms of time and accuracy. **Figure 1** shows the flowchart of the steps followed for modeling the regression equation using ABS class of algorithms.

Table 1. Comparison of ABS vs NAG code for classicaltechniques for solving LLS.

т	п	Method	EX	ERES	Time	M-flops
110	100	NAG	9E-3	1E-8	0.268	9.6
		ABS	1E-4	3E-7	0.268	9.8
500	100	NAG	7E-3	2E-8	0.880	11.9
		ABS	9E-4	2E-9	0.616	18.6
1000	100	NAG	1E-3	2E-8	1.67	12.4
		ABS	2E-3	2E-8	1.10	20.5
1500	300	NAG	6E-3	1E-7	26.00	10.8
		ABS	1E-3	4E-9	12.5	22.5



Hence obtained *x* and Equation (7) provides Regression coefficients to be substituted in Equation (1)

Figure 1. Flowchart of implementation of ABS algorithm for detection of sugar fastening level and PP.

## 2.3. Data Surveyed

A Survey on 80 patients has been done at an Endocrinologist's (one of the author's) patients for the number of factors affecting diabetes 73. All these 73 factors are taken as independent variables; sugar fasting and Post-Prandial (PP) sugar as dependent variables. The factors affecting diabetes include: Age; Height; Weight; BMI; Waist Circumference; Waist Hip ratio; Biceps; Triceps; Suprailiac; Sub scapular; Total lipids; Phospholipids; Triglycerides; Total Cholesterol; HDL; LDL; VLDL; HOMA-B; HOMA-R; Insulin; Hb1Ac; Hb1Ac (%); E:I Ratio; S:L Ratio; Pulse Rate; Fat in grams: L arm BMC; L arm Fat; L arm Lean; L arm Lean + BMC; L arm Total Mass; L arm % Fat; R arm BMC; R arm Fat; R arm Lean; R arm Lean + BMC; R arm Total Mass; R arm % Fat; Trunk BMC; Trunk Fat; Trunk Lean; Trunk Lean + BMC; Trunk Total Mass; Trunk % Fat; L Leg BMC; L Leg Fat; L Leg Lean; L Leg Lean + BMC; L Leg Total Mass; L Leg % Fat; R Leg BMC; R Leg Fat; R Leg Lean; R Leg Lean + BMC; R Leg Total Mass; R Leg % Fat; Subtotal BMC; Subtotal Fat; Subtotal Lean; Subtotal Lean + BMC; Subtotal Total Mass; Subtotal % Fat; Head BMC; Head Fat; Head Lean; Head Lean + BMC; Head Total Mass; Head % Fat; Total BMC; Total Fat; Total Lean; Total Lean + BMC; Total Mass; Total % Fat.

### **3. Numerical Solution**

An illustrative example of the solution of LLS through the ABS class of linear systems, which has been verified with the other methods on wikipedia [20] is described as follows:

$$x_1 + x_2 = 6; x_1 + 2x_2 = 5; x_1 + 3x_2 = 7; x_1 + 4x_2 = 10$$

Comparison with Ax = b gives

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}, b = \begin{bmatrix} 6 \\ 5 \\ 7 \\ 10 \end{bmatrix}, x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}$$
$$A^{T} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix},$$
$$A^{T} y = A^{T} b = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 6 \\ 5 \\ 7 \\ 10 \end{bmatrix}$$
$$A^{T} b = \begin{bmatrix} 28 \\ 77 \end{bmatrix} \text{ so } A^{T} y = \begin{bmatrix} 28 \\ 77 \end{bmatrix}$$
$$A^{T} b = \begin{bmatrix} 28 \\ 77 \end{bmatrix}$$
$$B = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$
$$B_i = H_i a_i = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$
$$B_i = H_i^{T} Z_i$$
$$For \ i = 1 \quad p_1 = H_1^{T} Z_1$$
$$P_i = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$
$$Condition \quad Z_i^{T} H_i a_i \neq 0$$

$$\begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = 4.$$

Hence satisfied, Now,

$$\begin{aligned} x_{2} &= x_{1} - \left[\frac{\left(a_{1}^{T}x_{1} - b_{1}\right)}{\left(a_{1}^{T}P_{1}\right)}\right]P_{1} \\ x_{2} &= \begin{bmatrix} 0\\0\\0\\0\\-\end{bmatrix} - \left[\frac{\left[1 & 1 & 1 & 1\right] \begin{bmatrix} 0\\0\\0\\-\end{bmatrix} - 28\\0\\-\end{bmatrix} \\ \left[\frac{1}{1}\right] \\ \left[1 & 1 & 1 & 1\right] \begin{bmatrix} 1\\1\\1\\1\\-\end{bmatrix} \\ \left[\frac{1}{1}\right] \\ x_{2} &= \begin{bmatrix} 7\\7\\7\\7\\7\\7\\7\\-\end{bmatrix}, H_{2} &= H_{1} - \frac{\left(P_{1}P_{1}^{T}\right)}{\left(P_{1}^{T}P_{1}\right)} \\ P_{1} &= \begin{bmatrix} 1\\1\\1\\1\\-\end{bmatrix}, P_{1}^{T} &= \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \\ P_{1} &= \begin{bmatrix} 1\\1\\1\\-\end{bmatrix} \\ \left[1 & 1 & 1 & 1 \end{bmatrix} \\ H_{2} &= \begin{bmatrix} 1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1 \end{bmatrix} - \left[ \begin{bmatrix} 1\\1\\1\\1\\-\\1\\-\\-\\-\\-\\-\\0.25&-0.25&-0.25\\-0.25&-0.25\\-\\0.25&-0.25\\-\\0.25&-0.25&-0.25\\-\\0.25&-0.25&-0.25\\-\\0.25&-0.25&-0.25\\-\\0.25&-0.25&-0.25\\-\\0.25&-0.25&-0.25\\-\\0.25&-0.25&-0.25\\-\\0.25&-0.25&-0.25\\-\\0.25&-0.25&-0.25\\-\\0.25&-0.25\\-\\0.25&-0.25\\-\\0.25&-0.25\\-\\0.25&-0.25\\-\\0.25&-0.25\\-\\0.25&-0.25\\-\\0.25&-0.25\\-\\0.25&-\\0.25$$

Now,

$$p_{2} = H_{2}^{1}Z_{2}$$

$$= \begin{bmatrix} 0.75 & -0.25 & -0.25 & -0.25 \\ -0.25 & 0.75 & -0.25 & -0.25 \\ -0.25 & -0.25 & 0.75 & -0.25 \\ -0.25 & -0.25 & -0.25 & 0.75 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} -1.5 \\ -0.5 \\ 0.5 \\ 1.5 \end{bmatrix}.$$

Copyright © 2013 SciRes.

Now,

$$x_{3} = x_{2} - \left[\frac{a_{2}^{T}x_{2} - b_{2}}{a_{2}^{T}p_{2}}\right]p_{2}$$

$$x_{3} = \begin{bmatrix}7\\7\\7\\7\\\end{bmatrix} - \left[\frac{(1 \ 2 \ 3 \ 4)}{(1 \ 2 \ 3 \ 4)}\begin{bmatrix}7\\7\\7\\7\\-7\\7\end{bmatrix} - 77\\7\\-0.5\\0.5\\1.5\end{bmatrix} = \begin{bmatrix}-1.5\\-0.5\\0.5\\1.5\end{bmatrix}$$

$$x_{3} = \begin{bmatrix}4.9\\6.3\\7.7\\9.1\end{bmatrix}.$$

The above equation is obtained for ABS linear problem and applying it to finding solution of Linear Least Square Problem,

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}, b = \begin{bmatrix} 4.9 \\ 6.3 \\ 7.7 \\ 9.1 \end{bmatrix}, x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Let for the equation Ax = b

 $x_2$ 

$$x_{1} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, H_{1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$S_{k} = H_{k}a_{k}$$

$$S_{1} = H_{1}a_{1}, H_{1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$p_{i} = H_{i}^{T}z_{i}$$

$$p_{1} = H_{1}^{T}z_{1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$x_{2} = x_{1} - \begin{bmatrix} \frac{a_{1}^{T}x_{1} - b_{1}}{a_{1}^{T}p_{1}} \end{bmatrix} p_{1}$$

$$= \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \begin{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} - 4.9 \\ \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2.45 \\ 2.45 \end{bmatrix}$$

$$H_{2} = H_{1} - \begin{bmatrix} \frac{p_{1}p_{1}^{T}}{p_{1}^{T}p_{1}} \end{bmatrix}$$

$$p_{1}^{T} = \begin{bmatrix} 1\\1 \end{bmatrix}$$

$$H_{2} = \begin{bmatrix} 1 & 0\\0 & 1 \end{bmatrix} - \begin{bmatrix} \begin{bmatrix} 1\\1 \end{bmatrix} \begin{bmatrix} 1 & 1\\1 \end{bmatrix} \begin{bmatrix} 1 & 1\\1 \end{bmatrix} = \begin{bmatrix} 0.5 & -0.5\\-0.5 & 0.5 \end{bmatrix}$$

Now,

$$p_{2} = H_{2}^{T} z_{2} \quad p_{2} = \begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -0.5 \\ 0.5 \end{bmatrix}$$
$$x_{3} = x_{2} - \begin{bmatrix} \frac{a_{2}^{T} x_{2} - b_{2}}{a_{2}^{T} p_{2}} \end{bmatrix} p_{2}$$
$$x_{3} = \begin{bmatrix} 2.45 \\ 2.45 \end{bmatrix} - \begin{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 2.45 \\ 2.45 \end{bmatrix} - 6.3 \\ \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} -0.5 \\ 0.5 \end{bmatrix} \end{bmatrix} \begin{bmatrix} -0.5 \\ 0.5 \end{bmatrix} \quad \mathbf{x}_{3} = \begin{bmatrix} \mathbf{3.5} \\ \mathbf{1.4} \end{bmatrix}$$

Hence, the solution has been verified with the other classical methods.

## 4. Results and Conclusion

Proposed work contains two aims: First to find out the regression equation relating all the factors affecting diabetes with the sugar fasting level and PP; Second is to find out the strength of relationship among sugar fasting level and PP with these factors, that is clear from the below mentioned Equations (A) and (B) by their weights in the equation. General available softwares for regression analysis do have their limits for the number of independent variables and the data size *i.e.* for Microsoft Excel it is limited to 16 variables, this is not a case for presented program. The programming has been done is C language. The Regression equations obtained are:

$$\begin{split} y_1 &= 8.540 + 0.4103t_1 + 1.8127t_2 + 0.291t_3 + 0.8983t_4 \\ &+ 1.2169t_5 - 1.1023t_6 + 0.298t_7 + 0.6106t_8 - 0.020t_9 \\ &+ 0.0699t_{10} - 0.0168t_{11} - 0.0401t_{12} - 0.04t_{13} - 0.0213t_{14} \\ &- 0.4012t_{15} + 0.06t_{16} - 0.172t_{17} + 1.4998t_{18} + 71.091t_{19} \\ &- 42.08t_{20} + 13.163t_{21} - 0.8t_{22} + 3.016t_{23} - 4.9329t_{24} \\ &+ 0.8919t_{25} - 0.2969t_{26} + 0.001t_{27} + 0.0051t_{28} + 0.0051t_{29} \\ &+ 0.0017t_{30} + 0.1722t_{31} - 0.0168t_{32} - 0.0018t_{33} - 0.0086t_{34} \\ &- 0.0088t_{35} - 0.0041t_{36} + 0.1774t_{37} + 0.031t_{38} - 0.0002t_{39} \\ &+ 0.002t_{40} + 0.0021t_{41} + 0.0006t_{42} + 0.038t_{43} - 0.051t_{44} \\ &- 0.0008t_{45} + 0.0015t_{46} + 0.0009t_{47} + 0.0034t_{48} + 0.3142t_{49} \\ &+ 0.0128t_{50} + 0.0021t_{51} - 0.0008t_{52} - 0.0002t_{53} + 0.0004t_{54} \\ &+ 0.2034t_{55} + 0.002t_{56} + 0.0001t_{57} - 0.0002t_{58} - 0.0001t_{59} \\ &+ 0t_{60} + 0.201t_{61} + 0.0129t_{62} + 0.0009t_{63} + 0.0019t_{64} \\ &+ 0.0028t_{65} + 0.0036t_{66} + 0.616t_{67} - 0.0063t_{68} + 0t_{69} \\ &+ 0.0002t_{70} + 0.0002t_{71} + 0.0001t_{72} + 0.0799t_{73} \\ \end{split}$$

$$y_{2} = 11.653 - 3.912t_{1} + 0t_{2} + 12.014t_{3} + 2.171t_{4} \\ -20.0103t_{5} + 6.4121t_{6} + 8.11t_{7} + 0t_{8} - 10.988t_{9} + 0t_{10} \\ -0.061t_{11} - 7.098t_{12} + 1.251t_{13} + 5.2113t_{14} + 5.002t_{15} + 0t_{16} \\ -0.5027t_{17} + 0.1831t_{18} + 90.0121t_{19} - 33.5166t_{20} \\ +161.1213t_{21} + 0t_{22} - 341.9048t_{23} + 0t_{24} - 0.411t_{25} \\ -2.0051t_{26} + 0t_{27} - 0.089t_{28} + 0t_{29} + 0.019t_{30} + 0t_{31} \\ -291.113t_{32} + 0t_{33} - 291.76t_{34} + 291.531t_{35} + 0t_{36} \\ -1.0056t_{37} + 194.1214t_{38} + 0t_{39} + 193.535t_{40} - 193.314t_{41} \\ +0t_{42} + 0t_{43} + 0t_{44} + 0t_{45} + 0t_{46} + 0t_{47} + 0t_{48} + 0t_{49} \\ +0.9012t_{50} + 0t_{51} + 0.6655t_{52} - 0.6655t_{53} - 0.0052t_{54} + 0t_{55} \\ +0t_{56} + 0.0007t_{57} + 0t_{58} + 0t_{59} + 0t_{60} - 2.9648t_{61} + 0t_{62} \\ +0.0931t_{63} - 0.4101t_{64} + 0t_{65} + 0.201t_{66} + 0t_{67} + 0t_{68} + 0t_{69} \\ +0.0011t_{70} + 0t_{71} + 0t_{72} + 0t_{73} \\ (B)$$

 $y_1$  and  $y_2$  present the sugar fasting and PP levels respectively, whereas  $t_1, t_2, t_3, \dots, t_{73}$  are the independent variables in the same order as discussed in 2.3. However, merely the regression coefficients cannot conclude about the significance of relationship between the parameters. Since, the scope of proposed study is limited towards finding the regression equation via ABS algorithm, not the p values and the significant impact of independent variables on the dependent one. The weights in regression equation have been used for further discussion for the proposed study and are summarized in **Table 2**.

For sugar fasting level:

1) It can be observed that the regression coefficients of HOMA-R, Hb1Ac, E:I ratio, HOMA-B, height, waist circumference, pulse rate, BMI, insulin (negative value), S:L ratio (negative value) were the top 10 positive/negative parameters (largest regression coefficient values), showing their observable impact on sugar fasting  $(y_1)$ .

Table 2. Top 10 most influencing factors affecting sugar fasting and PP.

Parameters	Regression coeff. for sugar fasting	Regression coeff. for PP	
Height	1.8127	-	
S:L ratio	-4.9329	-	
Pulse rate	0.8919	-	
BMI	0.8983	-	
Waist circumference	1.2169	-	
HOMA-B	1.4998	-	
HOMA-R	71.091	90.0121	
Insulin	-42.08	-33.5166	
Hb1Ac	13.163	161.1213	
E:I ratio	3.016	-341.9048	
R arm Fat	-	-291.76	
R arm Lean	-	291.531	
R arm % fat	-	194.1214	
Trunk Fat	-	193.535	
Trunk Lean	-	-193.314	
Weight	-	12.014	

2) The regression coefficients of subtotal lean, subtotal lean + BMC, subtotal total mass, total fat, total mass have the smallest values; hence they are the parameters having least effect on either increase or decrease in sugar fasting level  $(y_1)$ .

For PP level:

1) It can be observed that the regression coefficients of biceps, weight, HOMA-R, Hb1Ac, trunk lean, trunk BMC, R arm lean + BMC, E:I Ratio (negative value), trunk lean + BMC (negative value), insulin (negative value) were the top 10 positive/negative parameters (largest regression coefficient values), showing their observable impact on PP ( $y_2$ ).

2) The regression coefficients of a large number of parameters have the small values depicting no impact on PP  $(y_2)$ .

Further development of the work could be in the direction of application of LLS via ABS algorithm for more applications *i.e.* ANOVA and Chi-square, finding out the coefficient of determination for the significance of strength of correlation, polynomial fitting, numerical smoothing & differentiation and modelling workhorse & process modelling. Besides more ABS algorithm variants can be used for LLS, linear programming and non-linear programming problems due to their vast applicability in real life problems. The binary diabetes parameters can be solved by logistic regression using integer programming applications of ABS algorithms.

## 5. Acknowledgements

We wish to thank the Executive Director, Birla Institute of Scientific Research for the support given during this work. We gratefully acknowledge financial support by BTIS-sub DIC (supported by DBT, Govt. of India) and Advanced Bioinformatics Centre (supported by Govt. of Rajasthan) at Birla Institute of Scientific Research for infrastructure facilities for carrying out this work.

#### REFERENCES

- E. Spedicato, E. Bodon, Z. Xia and N. Mahdavi-Amiri, "ABS Method for Continuous and Integer Linear Equations and Optimization," *Central European Journal of Operations Research*, Vol. 18, No. 1, 2010, pp. 73-95. doi:0.1007/s10100-009-0128-9
- [2] S. Lalwani, R. Kumar, E. Spedicato and N. Gupta, "An Application of the ABS LX Algorithm to Multiple Sequence Alignment," *Iranian Journal of Operations Research*, Vol. 3, 2012, pp. 31-45.
- [3] S. Lalwani, R. Kumar, V. Rastogi and E. Spedicato, "An Application of the ABS LX Algorithm to Schedule Medical Residents," *Journal of Computer and Information Technology*, Vol. 1, 2011, pp. 95-118.

- [4] S. S. Rich, "Genetics of Diabetes and Its Complications," *Journal of the American Society of Nephrology*, Vol. 17, No. 2, 2006, pp, 353-360. doi:0.1681/ASN.2005070770
- [5] World Health Statistics Report, 2012. http://www.diabetes24-7.com/?p=1272
- [6] A. Ramachandran, C. Snehalatha, A. S. Shetty and A. Nanditha, "Trends in Prevalence of Diabetes in Asian Countries," *World Journal of Diabetes*, Vol. 3, No. 6, 2012, 110-117. doi:0.4239/wjd.v3.i6.110
- [7] C. L. Lawson and R. J. Hanson, "Solving Least Squares Problems," *Society for Industrial and Applied Mathematics*, 1995. doi:0.1137/1.9781611971217
- [8] A. Björck, "Numerical Methods for Least Squares Problems," Society for Industrial and Applied Mathematics, 1996. doi:0.1137/1.9781611971484
- [9] J. Abaffy, C. G. Broyden and E. Spedicato, "A Class of Direct Methods for Linear Equations," *Numerische Mathematic*, Vol. 45, 1984, pp. 361-376.
- [10] J. Abaffy and E. Spedicato, "Numerical Experiments with the Symmetric Algorithm in the ABS Class for Linear Systems," *Optimization*, Vol. 18, No. 2, 1987, pp. 197-212. doi:0.1080/02331938708843232
- [11] J. O. Rawlings, S. G. Pantula and D. A. Dickey, "Applied Regression Analysis: A Research Tool," 2nd Edition, Springer, Berlin, 1998. <u>doi:0.1007/b98890</u>
- [12] S. Chatterjee and A. A. S. Hadi, "Regression Analysis by Example," 4th Edition, Wiley Interscience, 2006. doi:0.1002/0470055464
- [13] H. Y. Huang, "A Direct Method for the General Solution of a System of Linear Equations," *Optimization Theory* and Application, Vol. 16, No. 5, 1975, pp. 429-445.
- [14] E. Spedicato and E. Bodon, "ABSPACK 1: A Package of ABS Algorithms for Solving Linear Determined, Underdetermined and Overdetermined Systems." www.unibg.it/dati/persone/636/404.pdf
- [15] E. Spedicato, "ABS Algorithm for Linear Systems and Linear Least Squares: Theoretical Results and Computational Performance," *Scientia Iranica*, Vol. 1, 1995, pp. 289-303.
- [16] J. Abaffy and E. Spedicato, "ABS Projection Algorithms: Mathematical Techniques for Linear and Nonlinear Equations," Ellis Horwood Ltd., Chichester, 1989.
- [17] E. Spedicato and E. Bodon, "Solution of Linear Least Squares via the ABS Algorithms," *Mathematical Programming*, Vol. 58, No. 1-3, 1993, pp. 111-136. doi:0.1007/BF01581261
- [18] Numerical Algorithms Group (NAG) Codes. http://www.nag.co.uk/
- [19] E. Spedicato, "On the Solution of Linear Least Squares through the ABS Class for Linear Systems," AIRO Conference, 1985, pp. 89-98.
- [20] Linear Least Squares (Mathematics), Motivational Example on Wikipedia. http://en.wikipedia.org/wiki/Linear\_least\_squares\_%28m athematics%29#Motivational\_example