

An Algorithm for Least Squares Computation Using the Manifold and Hilbert Space Dialectics

J. B. Olaleye¹, O. E. Abiodun² and J. O. Olusina³

Abstract

The method of least squares is widely used in numerical analysis of data in all applied quantitative fields. Although there is only one least squares criterion, several schemes have been used for its implementation. The manifold approach treats the entire least squares process, including the representation of the variables, the model formation and the computations, in terms of manifolds. A manifold is a group of variables or functions taken together and treated as an entity in the computation process. This paper presents the least squares optimization on the manifolds and shows that the express formation and solution of the usually formidable normal equations can be avoided by employing the Hilbert space axioms and methods in the Euclidean space generated by the axial manifolds. The sequential and systematic approach of the new scheme, the preservation of the group structure and the analytical insights it provides for understanding the

¹ University of Lagos, Nigeria, e-mail: jb_ola@yahoo.com

² University of Lagos, Nigeria, e-mail: abiodunoludayo@yahoo.com

³ University of Lagos, Nigeria, e-mail: joolusina1@yahoo.com

fundamental geometry of the least squares problem, all of which are demonstrated in the sample applications presented, support the conclusions that the manifold approach is less daunting, requires less core storage space and facilitate better understanding of the problem and the solution.

Mathematics Subject Classification: 93E24

Keywords: Manifolds, Inner Product Space, orthogonality, inner product, Basis Manifolds

1 Introduction

The scientific study of a real world phenomenon often starts from conceptualizing it as an entity having physical properties or attributes which are either directly observable or can be derived indirectly from a mathematical relation characterizing the phenomenon under study. The method of least squares (LS) is one of the many ways to derive values for the unknown properties (parameters) given the values for the observable properties in sufficient quantity. Conceptually, the method makes use of redundant but independent observations in the mathematical formulation of a given problem by seeking the minimum of the sum of the squares of the residual or measurement errors in the solution process. The concept, first used by Gauss to estimate the orbits of celestial bodies, has been popularized over the years and applied to numerical data analysis in engineering and the sciences.

Usually in traditional LS process, the variables are treated as individual elements or quantities in the governing equations and solution is sought using symbolic representation and manipulation in familiar long hand algebraic optimization procedure (Bjorck et al., 2000, Beck and Ben-Tal 2006, Fu and Barlow 2004, Markovsky and Huffel 2005 and Lemmerling et al., 2003). The

introduction of vectors and matrices as holders of groups of variables in more recent applications of the method brought a great relieve from the drudgery and boredom of long hand data manipulation. However, the vector-matrix approach has been used only to alleviate the mental effort needed to solve the problem (Eckl et al, 2002) the geometric ideas, which could facilitate better understanding of the problem and the solution, are often lost in the final stages as attention is focused on the huge task of pulling out the numerical values of the parameters from the normal equations. Not only that, the construction of the normal equations in matrix-vector form frequently involves storing and manipulating many, sometimes large, matrices too often difficult to handle and prone to instability of solution due to random computational errors (Lemmerlinget al., 2003, Arabelos and Tscherning 2009).

In this paper, we employ a geometric approach in which the LS variables and the equations are treated as composite variables called the *manifolds*. Conceptually, a manifold is a set which serves as a holder for quantities which are to be treated as an entity. Taken collectively, the observables, the parameters, the functional models and their derivatives may be regarded as geometric entities in a *Euclidean space*, and therefore can be manipulated using the axioms or dialects of the *Hilbert* or *inner product space* (Chang and Paige 2003). This approach has a number of advantages: (1) it avoids the express formation of the normal equations, but instead uses the concepts of *inner products*, *orthogonality* and *projections* in Euclidean spaces to directly compute the required parameters; (2) it enables us to preserve the group structure of the variables; (3) it carries the underlying geometry of the LS problem throughout the solution process; (4) it facilitates an easy understanding of the LS process; (5) its simplicity of operation (in terms of easy derivation of the computational algorithms); (6) economy of thought (in terms of the savings in mental effort); (7) ergonomics of the solution scheme (efficiency due to sequential data processing and direct computation of parameters); and (9) the geometric illumination it offers for understanding the structure of the problem

under study. In addition, the use of orthogonal method for LS has been known to provide more stable solutions than direct inversion method (Chang and Paige 2003, Zarowski 2004 and Eckl et al 2002). In this work, we refer to the set of observable (or observed) properties as *data manifold*; the set of unknown properties as *parameter manifold*; and the list of functional models (the mappings) as the *model manifold*. For ease of notation, we may refer to a manifold as a vector having a meaningful name and treated as an entity. Therefore, the terms manifold and vector will be used interchangeably.

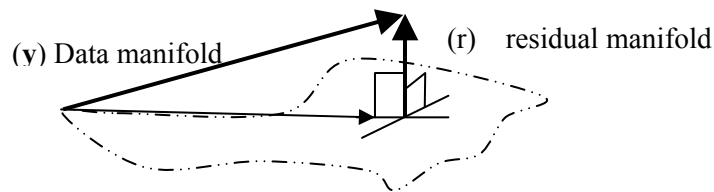


Figure 1: The LS manifold variables projection theorem

Although, the results of a LS adjustment of a given set of measurements must be the same regardless of which technique is applied, our preference for the manifold approach is premised its advantages as stated earlier. In the rest of the paper, the structure of the manifolds and their use in LS adjustment are presented. The main properties and rules of operation in an inner product space are listed. The algorithms for practical implementation of the technique, including the orthogonalization of the axial manifolds, the projection theorem and how they are used to compute the estimators for the parameters, are demonstrated with some computational examples. Some conclusions are presented at the end.

2 Preliminary Notes

2.1 Manifolds in Least Squares Adjustment

The manifolds in a LS adjustment can be grouped into the *primary manifolds* and the *secondary manifolds*. The primary manifolds are those mentioned earlier in this paper (see Figure 1).

They are: 1) the data manifold represented by $\mathbf{y} = [y_1, y_2, \dots, y_m]^T$,

2) the parameter manifold represented by $\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_n]^T$, and

3) the model manifold represented by $\mathbf{f} = [f_1, f_2, \dots, f_m]^T$.

The secondary manifolds are those derived in the LS process through the manipulation of the primary manifolds. They include: *expected data manifold* which is computed from the parameter and the model manifolds and may be represented as $\hat{\mathbf{y}} = [\hat{y}_1, \hat{y}_2, \dots, \hat{y}_m]^T$; the *residual manifold* which is introduced to account for the difference between the data manifold \mathbf{y} and the expected data manifold $\hat{\mathbf{y}}$ and represented by $\mathbf{r} = \mathbf{y} - \hat{\mathbf{y}} = [r_1, r_2, \dots, r_m]^T$; the *axes manifolds* which are the partial directional derivatives of the model manifold with respect to each of the parameters. They form the axes of the n -dimensional Euclidean space of the LS problem and will be represented by:

$$\mathbf{p}_i = \left[\frac{\partial f_1}{\partial \beta_i}, \frac{\partial f_2}{\partial \beta_i}, \dots, \frac{\partial f_m}{\partial \beta_i} \right]^T \quad i = 1, 2, \dots, n$$

where n is the number of parameters. Geometrically, each axial manifold has the same dimension (m) as the data manifold; and there are as many of them as there are unknowns (n) to be determined. In analytical geometry, these axial manifolds are called spanning or basis vectors. *Each one is in fact the stack of the slopes, gradients, scales, sensitivities, or rates of change of the functional manifold with respect to one parameter only*, and depending on the type of the LS problem being solved, can variously represent one direction (or axis) of change of a problem such

as: factor, category, coordinates, frequency channel, or length, breadth, and cross-sectional area of a structural member, etc. In general, they form the axes of the hyperspace in which the LS problem is formulated (Figure 2).

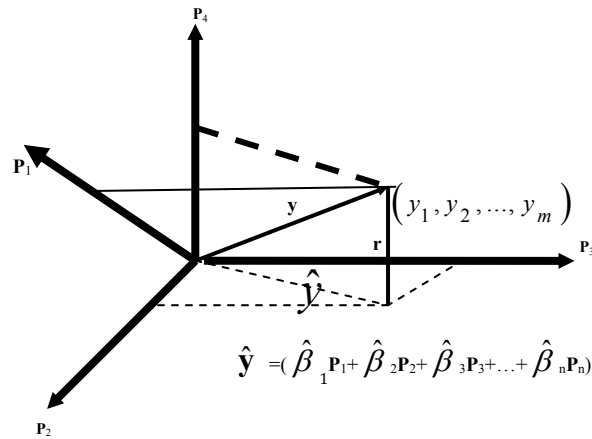


Figure 2: A Geometric representation of the LS problem space generated by the axial manifolds \mathbf{p}_1 , \mathbf{p}_2 , \mathbf{p}_3 , \mathbf{p}_4

We note that the common feature in all the manifolds listed is that they are column elements. They become variables in the LS process and may be arranged into larger groups called the *super manifolds*. A *super manifold* is one whose elements are manifolds arranged along a row. In other words, the super manifold is a row manifold whose elements are manifolds. Although, their use is limited in the sequel, the rules of the inner product space equally apply to super manifolds. In this paper, a super manifold is represented by capital boldfaced letter such as \mathbf{X} , and may be alternatively called a matrix.

The following operations and methods are available in a Euclidean space having a weight metric (Ng et al., 2002 and Chang and Paige 2003).

Distance between two position vectors:

$$d(\mathbf{a}, \mathbf{b}) = \sqrt{(\mathbf{a} - \mathbf{b})^T \mathbf{W}(\mathbf{a} - \mathbf{b})} \quad (1)$$

a) Length (L_2 Norm) of a position vector \mathbf{a} :

$$\|\mathbf{a}\|_2 = \mathbf{d}(\mathbf{a}, \mathbf{0}) = \sqrt{\mathbf{a}^T \mathbf{W} \mathbf{a}} \quad (2)$$

b) The projection of position vector \mathbf{a} on position vector \mathbf{b} (i.e. length of \mathbf{a} along \mathbf{b})

$$\frac{\mathbf{a}^T \mathbf{W} \mathbf{b}}{\mathbf{b}^T \mathbf{W} \mathbf{b}} \quad (3)$$

c) Minimum length of a vector \mathbf{r} can be expressed as

$$\min \|\mathbf{r}\| \equiv \min \mathbf{r}^T \mathbf{W} \mathbf{r} \quad (4)$$

It is a quadratic form which when \mathbf{W} is diagonal is the equation of a hyper ellipsoid.

d) The Outer Product \mathbf{C} of a manifold \mathbf{a} is a super manifold which depicts the covariation or correlation among its elements.

$$\mathbf{C} = \mathbf{a} \mathbf{W} \mathbf{a}^T \quad (5)$$

The trace of the outer product is equal to the inner product.

3 Main Results

To demonstrate the advantages of the manifold approach of LS optimization, a linear regression problem is solved. The algorithm is applied to the regression problem when the observations are of equal weight. In the field of Geomatics Engineering, equipment used for position determination is calibrated before use. This is done by the use of the same equipment for repeated independent observations which are later solved using the least squares method. The manifold approach presented above was used to determine the operational model of an equipment and the results are as presented.

The problem

In order to determine the operational model of an equipment, five measurements (x_i, y_i) , $i = 1, 2, \dots, 5$, of its input and output were made and recorded as follows:

$$(1, 2.5) \quad (3, 3.5) \quad (6, 5) \quad (5, 3) \quad (3, 4).$$

From these five measurements, determine the parameters of a straight line model $y = ax + b$ that best fits the data by assuming they are of equal weight.

Result

The primary manifolds are as stated below:

$$\mathbf{y} = \begin{bmatrix} 2.5 \\ 3.5 \\ 5 \\ 3 \\ 4 \end{bmatrix}, \quad \mathbf{f} = \begin{bmatrix} a + b \\ 3a + b \\ 6a + b \\ 5a + b \\ 3a + b \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} a \\ b \end{bmatrix}$$

The dimension m and n of the problem is:

$$M = 5, \quad n = 2$$

The axial manifolds are as stated below:

For $I = 1$ to n compute the axial manifolds

$$\mathbf{p}_1 = [1 \ 3 \ 6 \ 5 \ 3]^T, \quad \mathbf{p}_2 = [1 \ 1 \ 1 \ 1 \ 1]^T$$

The computed parameters are:

$$b = \frac{\mathbf{v}_2^T \mathbf{y}}{\mathbf{v}_2^T \mathbf{v}_2} = \frac{180}{80} \cdot \frac{160}{152} = 2.3684$$

$$a = \frac{\mathbf{p}_1^T \mathbf{y} - (\mathbf{p}_1^T \mathbf{p}_2) \beta_2}{\mathbf{p}_1^T \mathbf{p}_1} = \frac{3 - 3(-1)}{2} = \frac{70 - 18 \cdot 2.3684}{80} = 0.3421$$

$$\boldsymbol{\beta} = \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0.3421 \\ 2.3684 \end{bmatrix}$$

Hence the model is

$$\hat{y} = 0.3421x + 2.3684$$

Results Validation

The residual manifold is computed as

$$\mathbf{r} = \mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} 2.5 \\ 3.5 \\ 5 \\ 3 \\ 4 \end{bmatrix} - \begin{bmatrix} 2.7105 \\ 3.3947 \\ 4.4210 \\ 4.0789 \\ 3.3947 \end{bmatrix} = \begin{bmatrix} -0.2105 \\ 0.1053 \\ 0.5790 \\ -1.0789 \\ 0.6053 \end{bmatrix}$$

$$\mathbf{r}^T \hat{\mathbf{y}} = \begin{bmatrix} -0.2105 \\ 0.1053 \\ 0.5790 \\ -1.0789 \\ 0.6053 \end{bmatrix}^T \begin{bmatrix} 2.7105 \\ 3.3947 \\ 4.4210 \\ 4.0789 \\ 3.3947 \end{bmatrix} = \begin{bmatrix} -0.57056025 \\ +0.35746191 \\ +2.559759 \\ -4.40072521 \\ +2.05481191 \end{bmatrix} = 0.0007$$

$$\mathbf{r}^T \mathbf{p}_1 = \begin{bmatrix} -0.2105 \\ 0.1053 \\ 0.5790 \\ -1.0789 \\ 0.6053 \end{bmatrix}^T \begin{bmatrix} 1 \\ 3 \\ 6 \\ 5 \\ 3 \end{bmatrix} = \begin{bmatrix} -0.2105 \\ +0.3159 \\ +3.4740 \\ -5.3945 \\ +1.8159 \end{bmatrix} = 0.0008$$

$$\mathbf{r}^T \mathbf{p}_2 = \begin{bmatrix} -0.2105 \\ 0.1053 \\ 0.5790 \\ -1.0789 \\ 0.6053 \end{bmatrix}^T \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -0.2105 \\ +0.1053 \\ +0.5790 \\ -1.0789 \\ +0.6053 \end{bmatrix} = 0.0002$$

Discussion of Results

An examination of the computational steps in the above examples shows a series of inner product calculations and arithmetic operations of addition, multiplication and division of inner products. The form of the formulas makes it easy to understand the geometric basis of the LS problem. Comparing the results, we see that the effect of the weighting is to increase the projection or inclination of the data manifold on the first axial manifold (which is influenced more by the data

points) and reduce the same on the second axial manifold (which is a constant axis). The increase in the slope and the decrease in the intercept of the fitted line is as a result of the heavy reliance on the first three elements of the data manifold than the last two. The check on the numerical values of the parameters confirms that the adjustment satisfies the projection theorems.

3.1 Properties

3.1.1 The LS Problem in Geometric Terms

The geometric structure of the LS process can be established by considering the definition of the axial manifold given in the introduction section above. Since it is in fact the stack of the rates of change of the functional manifold with respect to one parameter only, it follows that for a LS problem with n unknowns, there are n such axial manifolds representing the n -dimensional rates of change of a phenomenon. Therefore, the total change observed in a phenomenon over a given time or space interval (or duration) is logically the sum of the products of individual axial manifolds and its corresponding space-time interval. Thus, intuitively, we can understand that what we call observed manifold is the totality of the changes contributed by all factors of change in a phenomenon. And along the same line of reasoning, since the observed total changes have been contributed by each direction or unit of change, what we call unknown parameter manifold is the collection of the numerical values of the time or space interval corresponding to the observed change. Accordingly, in analytical geometry, the observed manifold is represented as a linear (or linearised) combination of the axial manifolds, and is written as:

$$\mathbf{y} = \beta_1 \mathbf{p}_1 + \beta_2 \mathbf{p}_2 + \beta_3 \mathbf{p}_3 + \dots + \beta_n \mathbf{p}_n + \mathbf{r} \quad (6)$$

Note that \mathbf{r} is introduced to account for the errors in the measurement of \mathbf{y} . Equation 6 can be transposed in terms of residues as:

$$\mathbf{r} = \mathbf{y} - (\beta_1 \mathbf{p}_1 + \beta_2 \mathbf{p}_2 + \dots + \beta_n \mathbf{p}_n) \quad (7)$$

$$\mathbf{r} = \mathbf{y} - \hat{\mathbf{y}}$$

where $\hat{\mathbf{y}} = (\beta_1 \mathbf{p}_1 + \beta_2 \mathbf{p}_2 + \dots + \beta_n \mathbf{p}_n)$ is the expected or predicted manifold of the observed quantities. We observe in equation 6 that the observed manifold is a linear combination of the n axial manifolds $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n$ which are composed of the partial derivatives of the functional manifold \mathbf{f} with respect to a particular parameter in the parameter manifold. It follows that these axial manifolds form the axes of an n -dimensional space in which all other elements for the LS problem exist (see Figure 3).

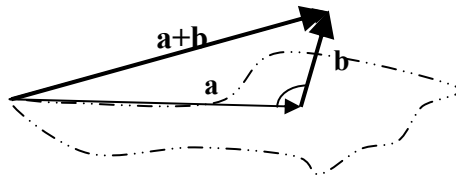


Figure 3: The projection theorem

Geometrically, we note that the elements \mathbf{y} , $\hat{\mathbf{y}}$ and \mathbf{r} are in triangular formation and each casts its shadows along the directions of the axial manifolds. The task of the LS process is to adjust them so that equation 8 (general principle of orthogonality condition) is satisfied such that \mathbf{r} has zero components along the axial directions.

$$\mathbf{r} \perp \mathbf{p}_i, \quad i = 1, 2, \dots, n, \quad (8)$$

As a matter of fact, this requirement of orthogonality of the residual error manifold (Equation 7) to each of the n -axial manifolds (Equation 8) results in a set of n -equations usually called “normal equations” in conventional LS adjustment. In geometric terms, these equations are called *orthogonality equations*.

3.1.2 Construction of the Axial Manifolds

As said earlier, each of the axial manifolds is the vector of the partial derivatives of the functional manifold with respect to one variable. The expressions for the construction of these column elements are given below:

$$\mathbf{p}_1 = \begin{bmatrix} \frac{\partial f_1}{\partial \beta_1} \\ \frac{\partial f_2}{\partial \beta_1} \\ \vdots \\ \frac{\partial f_m}{\partial \beta_1} \end{bmatrix}, \quad \mathbf{p}_2 = \begin{bmatrix} \frac{\partial f_1}{\partial \beta_2} \\ \frac{\partial f_2}{\partial \beta_2} \\ \vdots \\ \frac{\partial f_m}{\partial \beta_2} \end{bmatrix}, \quad \mathbf{p}_3 = \begin{bmatrix} \frac{\partial f_1}{\partial \beta_3} \\ \frac{\partial f_2}{\partial \beta_3} \\ \vdots \\ \frac{\partial f_m}{\partial \beta_3} \end{bmatrix}, \quad \dots, \quad \mathbf{p}_n = \begin{bmatrix} \frac{\partial f_1}{\partial \beta_n} \\ \frac{\partial f_2}{\partial \beta_n} \\ \vdots \\ \frac{\partial f_m}{\partial \beta_n} \end{bmatrix} \quad (9)$$

3.1.3 Construction of the Axial Manifolds

The orthogonality of the n -axial manifolds $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n$ to themselves is an important requirement in the method of the manifolds, i.e. $\mathbf{p}_i^T \mathbf{p}_j = 0 : i, j = 1$ to $n : i \neq j$. To achieve this, the classical Gram-Schmidt algorithm is used (Keerthi and Shevade 2003, Zarowski 2004 and Chang and Paige 2003). The steps are as follows:

Given a set of n axes manifolds $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n$, find an equivalent set $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ which are orthogonal. We take \mathbf{p}_1 and equate it to \mathbf{v}_1 . We now remove from \mathbf{p}_2 the component of \mathbf{p}_2 lying in the direction of \mathbf{v}_1 and the remainder will be orthogonal to \mathbf{v}_1 . This remainder manifold gives us \mathbf{v}_2 . Also, by subtracting off the components of \mathbf{p}_3 in the directions of \mathbf{v}_1 and \mathbf{v}_2 , we obtain \mathbf{v}_3 . This process can be repeated for all axes manifolds using the general formula (Zarowski 2004 and Chang and Paige 2003):

$$\mathbf{v}_i = \mathbf{p}_i - \sum_{j=1}^{i-1} \frac{(\mathbf{v}_j^T \mathbf{p}_i)}{(\mathbf{v}_j^T \mathbf{v}_j)} \mathbf{v}_j \quad (10)$$

For instance, when $i = 1$ equation 10 gives: $\mathbf{v}_1 = \mathbf{p}_1$

For $i = 2$, we have:
$$\mathbf{v}_2 = \mathbf{p}_2 - \frac{\mathbf{v}_1^T \mathbf{p}_2}{\mathbf{v}_1^T \mathbf{v}_1} \mathbf{v}_1$$

For $i = 3$, we have:
$$\mathbf{v}_3 = \mathbf{p}_3 - \frac{\mathbf{v}_1^T \mathbf{p}_3}{\mathbf{v}_1^T \mathbf{v}_1} \mathbf{v}_1 - \frac{\mathbf{v}_2^T \mathbf{p}_3}{\mathbf{v}_2^T \mathbf{v}_2} \mathbf{v}_2$$

Thus, the axial manifolds $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ are orthogonal and may be used in place of the original axes $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n$.

3.1.4 Formation of the Orthogonality Equations

The least squares solution seeks to minimize the L_2 -norm of the residual error-manifold $\|\mathbf{r}\|_2^2 \equiv \|\mathbf{y} - (\beta_0 \mathbf{P}_0 + \beta_1 \mathbf{P}_1 + \beta_2 \mathbf{P}_2 + \beta_3 \mathbf{P}_3)\|_2^2$. This is achieved by inserting equation 18 into equation 15 to obtain:

$$\langle (\mathbf{y} - (\beta_1 \mathbf{p}_1 + \beta_2 \mathbf{p}_2 + \beta_3 \mathbf{p}_3 + \dots + \beta_n \mathbf{p}_n)), \mathbf{p}_i \rangle = 0, \quad i = 1, 2, \dots, n \quad (11)$$

For n axial manifolds, this equation is clearly a set of n linear equations which result from the orthogonality condition. For instance, for $n = 4$, there are 4 such equations as listed below:

$$\begin{aligned} \langle [\mathbf{y} - (\beta_1 \mathbf{p}_1 + \beta_2 \mathbf{p}_2 + \beta_3 \mathbf{p}_3 + \beta_4 \mathbf{p}_4)], \mathbf{p}_1 \rangle &= 0 \\ \langle [\mathbf{y} - (\beta_1 \mathbf{p}_1 + \beta_2 \mathbf{p}_2 + \beta_3 \mathbf{p}_3 + \beta_4 \mathbf{p}_4)], \mathbf{p}_2 \rangle &= 0 \\ \langle [\mathbf{y} - (\beta_1 \mathbf{p}_1 + \beta_2 \mathbf{p}_2 + \beta_3 \mathbf{p}_3 + \beta_4 \mathbf{p}_4)], \mathbf{p}_3 \rangle &= 0 \\ \langle [\mathbf{y} - (\beta_1 \mathbf{p}_1 + \beta_2 \mathbf{p}_2 + \beta_3 \mathbf{p}_3 + \beta_4 \mathbf{p}_4)], \mathbf{p}_4 \rangle &= 0 \end{aligned} \quad (12)$$

which, by carrying out the indicated inner product expansions, may be re-arranged in matrix-vector form as:

$$\begin{bmatrix} \langle \mathbf{p}_1, \mathbf{p}_1 \rangle & \langle \mathbf{p}_2, \mathbf{p}_1 \rangle & \langle \mathbf{p}_3, \mathbf{p}_1 \rangle & \langle \mathbf{p}_4, \mathbf{p}_1 \rangle \\ \langle \mathbf{p}_1, \mathbf{p}_2 \rangle & \langle \mathbf{p}_2, \mathbf{p}_2 \rangle & \langle \mathbf{p}_3, \mathbf{p}_2 \rangle & \langle \mathbf{p}_4, \mathbf{p}_2 \rangle \\ \langle \mathbf{p}_1, \mathbf{p}_3 \rangle & \langle \mathbf{p}_2, \mathbf{p}_3 \rangle & \langle \mathbf{p}_3, \mathbf{p}_3 \rangle & \langle \mathbf{p}_4, \mathbf{p}_3 \rangle \\ \langle \mathbf{p}_1, \mathbf{p}_4 \rangle & \langle \mathbf{p}_2, \mathbf{p}_4 \rangle & \langle \mathbf{p}_3, \mathbf{p}_4 \rangle & \langle \mathbf{p}_4, \mathbf{p}_4 \rangle \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \begin{bmatrix} \langle y, \mathbf{p}_1 \rangle \\ \langle y, \mathbf{p}_2 \rangle \\ \langle y, \mathbf{p}_3 \rangle \\ \langle y, \mathbf{p}_4 \rangle \end{bmatrix}. \quad (13)$$

Equation 13 is referred to as the “normal equation” in conventional LS method and many algorithms have been developed for the direct solution of these equations (Bjorck et al., 2000, Beck and Ben-Tal 2006, Fu and Barlow 2004). The approach here is to avoid a direct solution by imposing the orthogonality of the axial manifolds as shown in the next section.

3.1.5 Derivation of the Solution Equations

When the orthogonal axial manifolds $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ are substituted in Equation 13, we have:

$$\begin{bmatrix} \mathbf{v}_1^T \mathbf{v}_1 & 0 & 0 & 0 \\ 0 & \mathbf{v}_2^T \mathbf{v}_2 & 0 & 0 \\ 0 & 0 & \mathbf{v}_3^T \mathbf{v}_3 & 0 \\ 0 & 0 & 0 & \mathbf{v}_4^T \mathbf{v}_4 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1^T \mathbf{y} \\ \mathbf{v}_2^T \mathbf{y} \\ \mathbf{v}_3^T \mathbf{y} \\ \mathbf{v}_4^T \mathbf{y} \end{bmatrix}. \quad (14)$$

It is noted that the orthogonalization of the axial manifolds has greatly simplified normal equations. In fact, by carrying out the inversion of the coefficient matrix of equation 14, we obtain equation 15, from which the simple formulas in Equation 16 emerge:

$$\boldsymbol{\beta} = \begin{bmatrix} \frac{1}{\mathbf{v}_1^T \mathbf{v}_1} & 0 & 0 & 0 \\ 0 & \frac{1}{\mathbf{v}_2^T \mathbf{v}_2} & 0 & 0 \\ 0 & 0 & \frac{1}{\mathbf{v}_3^T \mathbf{v}_3} & 0 \\ 0 & 0 & 0 & \frac{1}{\mathbf{v}_4^T \mathbf{v}_4} \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \mathbf{y} \\ \mathbf{v}_2^T \mathbf{y} \\ \mathbf{v}_3^T \mathbf{y} \\ \mathbf{v}_4^T \mathbf{y} \end{bmatrix} \quad (15)$$

$$\boldsymbol{\beta} = [\beta_1, \beta_2, \beta_3, \beta_4]^T = \left[\frac{\mathbf{v}_1^T \mathbf{y}}{\mathbf{v}_1^T \mathbf{v}_1}, \frac{\mathbf{v}_2^T \mathbf{y}}{\mathbf{v}_2^T \mathbf{v}_2}, \frac{\mathbf{v}_3^T \mathbf{y}}{\mathbf{v}_3^T \mathbf{v}_3}, \frac{\mathbf{v}_4^T \mathbf{y}}{\mathbf{v}_4^T \mathbf{v}_4} \right]^T \quad (16)$$

Equation 16 shows that the solution to the unknown parameters of a LS problem can be computed directly without an explicit formation of the familiar normal equations.

However, while they give the correct values for all parameters, the estimate computed for the first parameter in the manifold is always wrong. Our investigation showed that because the de-correlation carried out on the axial manifolds (Equation 10) is not absolute but relative to the first axial manifold, the correct solution for the first parameter can be computed only by normalizing the first axial manifold to the residual vector \mathbf{r} (Equation 8). Thus, by inserting equation 7 into equation 8 for $i = 1$ and expanding the indicated inner product, solving the resulting equation will yield the correct solution for β_1 . Thus, the correct formula for β_1 in equation 16 is:

$$\beta_1 = \frac{\mathbf{p}_1^T \mathbf{y} - \sum_{j=2}^n (\mathbf{p}_1^T \mathbf{p}_j) \beta_j}{\mathbf{p}_1^T \mathbf{p}_1} \quad (17)$$

3.1.6 Summary of Computational Steps

a. State the primary manifolds

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_m \end{bmatrix} \quad \mathbf{f} = \begin{bmatrix} f_1 \\ f_2 \\ \cdot \\ \cdot \\ f_m \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \beta_n \end{bmatrix}$$

Specify the dimension m and n of the problem

b. Generate the axial manifolds

- If nonlinear model then linearize
- For $i = 1$ to n compute the n axial manifolds

$$\mathbf{p}_i = \left[\frac{\partial f_1}{\partial \beta_i}, \frac{\partial f_2}{\partial \beta_i}, \dots, \frac{\partial f_m}{\partial \beta_i} \right]^T \quad i = 1, 2, \dots, n$$

c. Initialize the computations

Compute certain quantities for later use

$$\mathbf{p}_1^T \mathbf{y}, \mathbf{p}_1^T \mathbf{p}_1, \mathbf{p}_1^T \mathbf{p}_j: j = 2, \dots, n$$

d. Generate Orthogonal Axes manifolds

For $i=1$ to n generate the orthogonal axial manifolds

$$\mathbf{v}_i = \mathbf{p}_i - \sum_{j=1}^{i-1} \frac{(\mathbf{v}_j^T \mathbf{p}_i)}{(\mathbf{v}_j^T \mathbf{v}_j)} \mathbf{v}_j$$

e. Compute parameters

$$\boldsymbol{\beta} = [\beta_1, \beta_2, \beta_3, \beta_4]^T = \left[\frac{\mathbf{v}_1^T \mathbf{y}}{\mathbf{v}_1^T \mathbf{v}_1}, \frac{\mathbf{v}_2^T \mathbf{y}}{\mathbf{v}_2^T \mathbf{v}_2}, \frac{\mathbf{v}_3^T \mathbf{y}}{\mathbf{v}_3^T \mathbf{v}_3}, \frac{\mathbf{v}_4^T \mathbf{y}}{\mathbf{v}_4^T \mathbf{v}_4} \right]^T$$

f. Validate the results

$$\beta_1 = \frac{\mathbf{p}_1^T \mathbf{y} - \sum_{j=2}^n (\mathbf{p}_1^T \mathbf{p}_j) \beta_j}{\mathbf{p}_1^T \mathbf{p}_1}$$

Compute $\mathbf{r}, \mathbf{r}^T \hat{\mathbf{y}}, \mathbf{r}^T \mathbf{p}_i$ These inner products must be zero.

4 Conclusion

The findings in this paper can be summarized as follows:

- 1) The method of the manifolds is a potent alternative to the ubiquitous method of direct formation and inversion of normal equations for the solution of a least squares problem. The computation formulas are simple, the steps are routine

and the only mental exercise required is to remember that the inner product of two manifolds is the sum of the products of their corresponding elements.

- 2) The sequence of the solution steps holds the promise of huge savings in the calculation efforts when additional parameters and thus axial manifolds are to be included in the solution such as when trying out different models on the same data manifold.
- 3) Since only one parameter is treated at a time, huge savings in memory is possible as not all the data needs to be loaded into core memory at the same time.
- 4) The method provides an insight into the fundamental geometry of the least squares problem by preserving the group structure of the variables in the solution process.
- 5) It is also concluded that the correct formulae for the computation of the first parameter in the parameter list is that obtained from the explicit imposition of the projection theorem on the first axial manifold and the residual manifold.

References

- [1] D.N. Arabelos and C.C. Tscherning, Error-covariances of the estimates of spherical harmonic coefficients computed by LSC, using second-order radial derivative functionals associated with realistic GOCE orbits, *Journal of Geodesy*, Springer Publishers, **83**,(5), (2009), 419-430.
- [2] A. Beck and A. Ben-Tal, On the Solution of the Tikhonov Regularization of the Total Least Squares, *SIAM Journal of Optimization*, **17**(1), (2006), 98-118.
- [3] A. Bjorck, P. Heggernes and P. Matstoms, Methods for Large Scale Total Least Squares Problems, *SIAM Journal of Matrix and Analytical Applications*, **22**(2), (2000), 413-429.

- [4] X. Chang and C.C. Paige, An Orthogonal Transformation Algorithm for GPS Positioning, *SIAM J. Sci. Comput.*, **24**(5), (2003), 1710-1732.
- [5] M.C. Eckl, R.A. Snay, T. Soler, M.W. Cline and G.L. Mader, Accuracy of GPS Derived Relative Positions as a Function of Interstation Distance and Observing-Session Duration, *Journal of Geodesy*, **75**(12), (2002), 633-640.
- [6] H. Fu and J. Barlow, A Regularized Structured Total Least Squares Algorithm for High-resolution Image Reconstruction, *Linear Algebra Applications*, **391**(1), (2004), 75-98.
- [7] S.S. Keerthi and S. K. Shevade, SMO Algorithm for Least-Squares SVM Formulations, *Neural Computations*, **15**(2), (2003), 487-507.
- [8] P. Lemmerling, N. Mastronardi and S. V. Huffel, Fast Algorithm for Solving the Hankel/Toeplitz Structured Total Least Squares Problem, *Numerical Algorithms*, **23**, (2000), 371-392.
- [9] I. Markovsky and S.V. Huffel, High-performance Numerical Algorithms and Software for Structured Total Least Squares, *Journal of Computer and Applied Mathematics*, **180**(2), (2005), 311-331.
- [10] K.I. Markovsky and S.V. Huffel, Consistency of the Structured Total Least Squares Estimator in a Multivariate Errors-in-Variables model, *Journal of Statistical Planning Inference*, **133**(2), (2005), 315-358.
- [11] M. Ng, J. Koo and N. Bose, Constrained Total Least Squares Computations for High Resolution Image Reconstruction with Multisensors, *International Journal of Imaging Systems Technology*, **12**, (2002), 35-42.
- [12] C.J. Zarowski, *An Introduction to Numerical Analysis for Electrical Engineers*, John Wiley and Sons Inc., New Jersey, 2004.