Orthogonal and Spectral Decomposition Algorithms for the Optimal Estimation of Parameters from Indirect Measurements

O. E. Abiodun, J. O. Olusina, J. B. Olaleye
Department of Surveying and Geoinformatics, University of Lagos, Akoka, Yaba, Lagos.
Email: abiodunoludayo@yahoo.com, joolusina1@yahoo.com, jb_ola@yahoo.com

Abstract
The optimal estimation of the parameters of a physical system from a set of measurements is a task commonly performed in numerical analysis of data in engineering and all applied quantitative fields. In a statistical sense, experimental measurements are usually considered as a mixture of both useful signals and unwanted noise and so, often, the aim of measurement data processing is to separate these components of a measurement in an optimal way. In solving these problems, researchers often adopt the method of direct formation and inversion of normal equations for the solution of a least squares problem. This paper however presents the method of orthogonal decomposition as an easier, simpler and equally valid alternative approach to direct formation and inversion of normal equations. The basic principles of L_2-norm (a vector norm defined for a complex vector) optimization and its geometric structure are discussed and implemented through the techniques of orthogonal and spectral decompositions. A simple regression problem was used to demonstrate these algorithms from which conclusions are made that: the combined use of these decomposition algorithms produces both the parameter values and their variances simultaneously, making it possible to solve an optimization problem without an express formation of normal equations.

Keywords: Algorithm, Vectors, , Manifolds, Model, Orthogonality.

1.0 INTRODUCTION

THE scientific study of a real world phenomenon often starts from conceptualizing it as an entity having physical properties or attributes which may be classified as constants (properties whose values are known from common knowledge or previous experiments), observables (properties which can be directly observed in a measurement process) and parameters (properties which can only be derived indirectly as the outcome of the experiment (Bjiorck *et al*., 2000; Fu and Barlow, 2004; Markovsky and Huffel, 2005a; Chang and Paige 2003). Quite often in experimental work, the determination of the parameters is the main goal of the study, and the general technique is to associate properties to variables in a mathematical relation characterizing the phenomenon under study, measure the observable ones in sufficient quantities and compute the values for the unknown parameters using the model structure.

However, in measurement theory, it is common knowledge that the true value of a measured quantity is never obtainable due to imperfections of the measuring process and random effects. Therefore, measurements to be used as input into a system that estimates the parameters are considered as a mixture of both useful signals and unwanted noise and so the aim of the indirect measurement data processing is often to separate these components of a measurement in an optimal way, and this leads to the use of the term optimal estimation which suggests that only the most probable values of the parameters can be obtained. The optimal estimation concept is based on minimization of the noise component and, in the process, maximize the signal so that the most acceptable or probable values of the unknown parameters are obtained (Markovsky and Huffel, 2005a; Mikhail and Gracie, 1981; Wolf and Ghilani, 1997)

In this paper, the method of orthogonal decomposition is used to minimize the L_2 norm of the noise by breaking the observed data vector into two perpendicular components: the noiseless signal and the pure noise. The orthogonalization process generates a transfer matrix which symbolically filters out the noise from the data vector (Chang and Paige, 2003). The method of spectral decomposition symbolically decorrelates or mutually orthogonalizes the columns of the transition matrix and then finds the projection of the measured data vector along each column. These projections are the optimal values of the unknown parameters while the inverses of the square length of the orthogonal columns are the eigenvalues whose inverses provide the estimates of the variances. The combined use of these decomposition algorithms implicitly produces both the parameter values and their variances (Wolf and Ghilani, 1997; Markovsky and Huffel, 2005a).

The experimental model is assumed to have a non-zero derivatives at the point of observation. By ensuring that the noise is perpendicular to the tangent plane, the signal was recovered optimally. And for the spectral decomposition, the classical Grain-Schmit algorithm was used. The orthogonal and spectral decomposition algorithms were validated using a LS method of simple set of equally weighted measured line. Results show high level of compatibility.

2.0 METHODOLOGY

2.1 The Key Points of an Inner Product Space

The physical environment in which survey measurements are made is usually conceptualized to be a Cartesian coordinate system which may be characterized as a Euclidean vector space (Luenberger, 1969; Maddox, 1988). This permits us to use the axioms of that space to represent the sets of observed quantities, the parameters and the functional models as vectors and matrix elements and to derive computational algorithms for data processing. The important characteristics of the inner product space are summarized in what follows. The full discussion on the subject can be found in Luenberger (1969); Maddox (1988); Marcoux (2013):

The inner product space is often represented as $(X, <\cdot,\cdot>)$, where $<\cdot,\cdot>$ means inner product. For elements ***A***, ***B***, ***C*** in X, the following operations are allowed;

1) The Dot or Scalar product of two vectors ***A*** and ***B***, denoted by $\mathbf{A}\cdot\mathbf{B}$ (read A dot B), is defined as the product of the magnitudes of **A** and **B** and the cosine of the angle θ between them. In symbols,

$$< A, B > = A \cdot B = \boldsymbol{A^T B} = AB\cos\theta \qquad 0 \leqq \theta \leqq \pi \tag{1}$$

2) If $\mathbf{A}\cdot\mathbf{B} = 0$ and $\mathbf{A}$ and $\mathbf{B}$ are not null vectors, then $\mathbf{A}$ and $\mathbf{B}$ are perpendicular.
   If the dot product of two vectors is zero when none of the vectors is null, then they are orthogonal or perpendicular.

3) The following laws are also valid:

$\mathbf{A}\cdot\mathbf{B} = \mathbf{B}\cdot\mathbf{A}$ Commutative Law for Dot Products

$\mathbf{A}\cdot(\mathbf{B}+\mathbf{C}) = \mathbf{A}\cdot\mathbf{B} + \mathbf{A}\cdot\mathbf{C}$ Distributive Law

$$m(\mathbf{A}\cdot\mathbf{B}) = (m\mathbf{A})\cdot\mathbf{B} = \mathbf{A}\cdot(m\boldsymbol{B}) = (A\cdot B)m \tag{2}$$

where m is a scalar

Other types of operations which may be performed on vector elements of an inner product space include:

4) Distance between two position vectors:

$$d(\boldsymbol{A},\boldsymbol{B}) = \sqrt{(\boldsymbol{A}-\boldsymbol{B})^T(\boldsymbol{A}-\boldsymbol{B})} \tag{3}$$

5) Length (magnitude) of a position vector

$$\|\boldsymbol{A}\| = d(\boldsymbol{A},0) = \sqrt{\boldsymbol{A}^T\boldsymbol{A}} \tag{4}$$

6) The projection of position vector ***A*** on position vector ***B*** (i.e. length of **A** along **B)**

$$\Pr oj_A B = \frac{A^T B}{B^T B} \tag{5}$$

7) Minimum length of a vector ***A***

$$\text{min } \|\boldsymbol{A}\|^2 \equiv \boldsymbol{min}\ \boldsymbol{A}^T\boldsymbol{A} \tag{6}$$

However, in order to take account of the stochastic nature of the variables of a vector space involved in optimal estimation, recourse is often made to the concept of symmetric operator ***(W)*** in an inner product space (Maddox, 1988). The use of a symmetric operator makes the inner product space to become also a probability space (Olaleye *et al.,* 2012; Zarowski, 2004). By this, the inner product definition is adjusted to include the covariances of the observed (or estimated) elements of the space. The inner product becomes a weighted inner product. The following are the list of axioms using the weight matrix *(W)*:

(8) $$\mathbf{A}\cdot\mathbf{B} = \mathbf{A}^T\mathbf{W}\mathbf{B} \tag{7}$$

This form of the inner product is used in weighted optimization process (Lista *et al.*, 2004).

Nonetheless, a more practical and useful approach of including probability measure in the inner product definition associates a weight factor to every element of the space before computing the inner product. This is done by finding the square root of the weight matrix ***W*** and using this to transform every vector element into the equivalent weighted metric space, reflecting the confidence in the data. This transformation naturally incorporates the reliability measure into the inner product. For example, if we define a weight metric ***W*** and presume that there is a factorization of the weight $\mathbf{W} = \mathbf{S}^T\mathbf{S}$, then, we can weight each vector as:

(9) $$\mathbf{A}' \approx \mathbf{SA} \tag{8}$$

where ***S*** is a symmetric matrix root of ***W,*** thus the inner product in the weighted space can be given as:

(10) $$\mathbf{A}'\cdot\mathbf{A}' = \mathbf{A}'^T\mathbf{A}' = (\mathbf{SA})^T(\mathbf{SA}) = \mathbf{A}^T\mathbf{S}^T\mathbf{SA} = \mathbf{A}^T\mathbf{WA} \tag{9}$$

With this transformation into a probability vector space, a weighted operation can be performed using the above formulas and axioms. Nonetheless, a direct use of the weight matrix can be used as an alternative; and the formulas are given as below:

(11) Distance between two position vectors:

$$d(\boldsymbol{A},\boldsymbol{B}) = \sqrt{(\boldsymbol{A}-\boldsymbol{B})^T\boldsymbol{W}(\boldsymbol{A}-\boldsymbol{B})} \tag{10}$$

(12) Length of a position vector:

$$\|\boldsymbol{A}\| = d(\boldsymbol{A},0) = \sqrt{\boldsymbol{A}^T\boldsymbol{A}} \tag{11}$$

(13) The projection of position vector ***A*** on position vector ***B*** **(**i.e. length of ***A*** along ***B*****)**

$$\Pr oj_A B = \frac{A^T WB}{B^T WB} \tag{12}$$

(14) Minimum length of a vector

$$\text{mini } \|\boldsymbol{r}\|^2 \equiv \boldsymbol{min}\, \boldsymbol{r}^T\boldsymbol{W}\boldsymbol{r} \tag{13}$$

which all involve the weight matrix in the inner product. Thus all kinds of statistically meaningful questions about the goodness of observations, blunders, goodness of estimated parameters, etc., can be asked and answered in the space of our least squares problem (Lista *et al.*, 2004; Vanicek and Krakiwsky, 1986).

2.2 Functional Representation of Indirect Measurements in a Vector Space

In experimental work, the functional model linking a vector of measured quantities to a vector of parameters is often postulated based on some theoretical or empirical concepts, analytical geometry and topology. The model can be linear or nonlinear and can include many variables as desired. It is expressed in a mathematical form as follows:

Let an observable vector $\boldsymbol{y}$ be linked to a parameter vector $\boldsymbol{\beta}$ by a specified function$\boldsymbol{f}$, then we can write the equation as:

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{\beta}) \tag{14}$$

where ***f*** indicates the system function or process, ***y*** is the vector of observed data, $\boldsymbol{\beta}$ is the vector of desired parameters.

Eq. 14 is usually referred to as an observation equation. The optimal estimation process determines the solution $\boldsymbol{\beta}$ only when the problem formulation is overdetermined, that is, when more than enough observations have been obtained so that there is a redundancy in the formulation. The use of an overdetermined system with the mathematical model explicit in ***y*** and dimension of ***y*** = *m* > dim $\boldsymbol{\beta}$ = *n* brings into existence the residual vector ***r*** whose values are yet unknown. Therefore, if sufficient and independent measurements are made such that the presence of noise can be detected, the functional model can be represented as:

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{\beta}) + \boldsymbol{r} \tag{15}$$

r is the vector of noise in the measurements

In the general application of Eq. 15, the function ***f***, may be nonlinear in the unknown parameters $\boldsymbol{\beta}$**.** In such cases, the function ***f*** is usually replaced with an *m* x *n* matrix ***X,*** variously called the tangent matrix, design matrix, system matrix, kernel etc. whose *n* columns are the m-dimensional vectors each of which contains the derivatives of the function ***f*** with respect to one of the unknown parameters in $\boldsymbol{\beta}$. For a multi-parameter problem, Eq. 15 in a linearized form will have the following vector representations (Olaleye *et al*., 2012).

$$\boldsymbol{y} = [\boldsymbol{P_1}, \boldsymbol{P_2}, \dots, \boldsymbol{P_n}]\begin{pmatrix} \boldsymbol{\beta_1} \\ \boldsymbol{\beta_2} \\ \vdots \\ \boldsymbol{\beta_n} \end{pmatrix} + \boldsymbol{r} \quad (16)$$

where

$$\boldsymbol{P_1} = \left[\frac{\partial f_1}{\partial \beta_1} \quad \frac{\partial f_2}{\partial \beta_1}, \quad \dots \quad \frac{\partial f_m}{\partial \beta_1}\right]^T$$

$$\boldsymbol{P_2} = \left[\frac{\partial f_1}{\partial \beta_2} \quad \frac{\partial f_2}{\partial \beta_2}, \quad \dots \quad \frac{\partial f_m}{\partial \beta_2}\right]^T$$

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

n is the number of parameters. ***X*** is a matrix of *n* column vectors $\boldsymbol{P_i}$ and is written as:

$$\boldsymbol{X} = [\boldsymbol{P_1}, \boldsymbol{P_2}, \dots, \boldsymbol{P_n}] \quad (17)$$

The matrix in Eq. 16 will be referred to as the *tangent plane or tangent matrix* **(**see **Figure 1).**

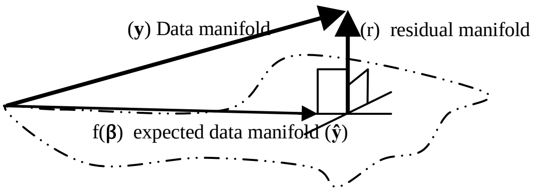

Figure 1: The LS manifold variables projection theorem

In order to simplify matters in this paper, we choose a simple function ***f*** which is a linear transformation represented by a matrix ***X*** whose columns serve as a basis for the transformation. For example, for a regression application, the ***f*** becomes a matrix ***X*** whose columns are the known regressors, causal values or specified base functions representing the derivatives (or tangents) of the regression function. Then our functional model Eq. 15 can be written as:

$$\boldsymbol{y} = \boldsymbol{X}(\boldsymbol{\beta}) + \boldsymbol{r} \quad (18)$$

Equation (15) can also be written as:

$$\boldsymbol{y} = \hat{\boldsymbol{y}} + \boldsymbol{r} \equiv \boldsymbol{y} - \hat{\boldsymbol{y}} = \boldsymbol{r}$$

where $\hat{\boldsymbol{y}} = \boldsymbol{X\beta}$ is the estimated observation or the noiseless signal. From Eq. 18, we can express the noise component as:

$$\boldsymbol{y} - \boldsymbol{X}(\boldsymbol{\beta}) = \boldsymbol{r} \quad (19)$$

2.3 Orthogonal Decomposition of Indirect Measurement

In mathematics, given a vector at a point on a curve, that vector can be decomposed uniquely as a sum of two vectors: one tangent to the curve, called the *tangential* *component* of the vector, and another one perpendicular to the curve, called the *normal component* of the vector. Similarly, a vector at a point on a surface can be broken down the same way. More generally, given a set of functions representing a group of curves and a vector at the point of tangency to the curves, the vector can be decomposed into the component tangent to the curves and a component normal to the tangent plane provided the derivatives exist at the position of tangency. This concept of orthogonal decomposition is applied in optimal estimation to separate a measured data vector into noiseless signal and noise components. The measured data vector represents the given vector at the point of tangency and the collection of derivatives of the experimental models with respect to some parameters provides the tangent plane. The experimental model or collection of models is assumed to have non-zero derivatives at the point of observation. The concept is that the measured data vector is a mixture of signal and noise. The signal component is along the tangent plane while the noise is along the normal to the tangent plane. Hence, by ensuring that the noise is perpendicular to the tangent plane, its L2-norm length is minimum and the signal can be recovered optimally. In practice, the tangent plane is represented by the matrix of model derivatives. The process can be treated mathematically as follows: If the directions of the tangent plane is known, then using the functional model in Eq. 18 and the orthogonality condition (Axioms 1 & 2) in section 2, the decomposition process is as follows (see **Figure 2**) (Fu and Barlow, 2004; Markovsky and Huffel, 2005b; Olaleye *et al*., 2012).

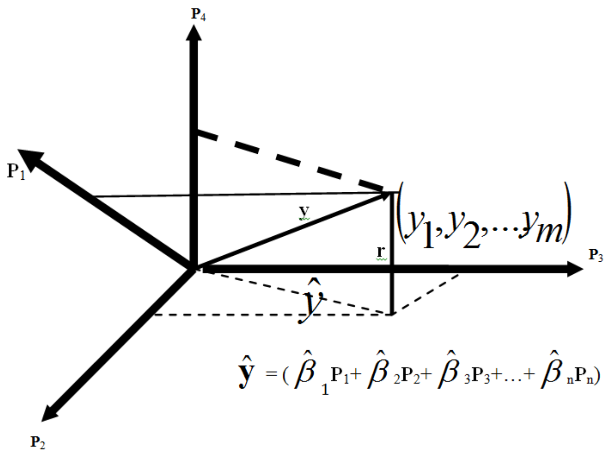

Figure 2: A Geometric representation of the LS problem space generated by the axial manifolds

Given the functional model linking measurement vector quantities containing error components to a set of parameters as in Eq. 18, and the noise component given by Eq. 19, the orthogonal projection theorem provides that the optimal solution (minimum L_2-norm of the noise component) can be obtained by projecting the error component so that it is orthogonal to the plane containing the optimal signal (the tangent plane) of the operating function. In other words, the L_2-norm of the residual error $\|\mathbf{r}\|$ is minimized when $\boldsymbol{r} \perp \boldsymbol{X}$ i.e the inner product of **r** and each of the spanning base vectors $\mathbf{P}_i$ of the tangent plane ***X*** is zero. This is clearly expressible as inner product operation or the orthogonality conditions as:

$$< (\boldsymbol{y} - \boldsymbol{X\beta}, \boldsymbol{X}) >= \boldsymbol{X}^T(\boldsymbol{y} - \boldsymbol{X\beta}) = \mathbf{0} \qquad (20)$$

Performing the indicated inner product and using the distributive Axiom 2 and 3 above to Eq. 20, results in what is called the normal equations:

$$\boldsymbol{X}^T\boldsymbol{y} - \boldsymbol{X}^T\boldsymbol{X\beta} = \mathbf{0} \qquad (21)$$

From where we get the expression for $\boldsymbol{\beta}$ as:

$$X^T y = X^T X \beta \tag{22}$$

$$\beta = (X^T X)^{-1} X^T y \tag{23}$$

If the data vector $\boldsymbol{y}$ is assumed to have uniform and independent variability matrix

$$C_y = \sigma^2 I \tag{24}$$

The parameter vector $\boldsymbol{\beta}$ will also have a variability matrix given by Gauss Error Law as:

$$C_\beta = ((X^T X)^{-1} X^T) C_y ((X^T X)^{-1} X^T)^T$$

$$C_\beta = \sigma^2 ((X^T X)^{-1} X^T)((X^T X)^{-1} X^T)^T$$

$$C_\beta = \sigma^2 (X^T X)^{-1} X^T X (X^T X)^{-1}$$

$$C_\beta = \sigma^2 (X^T X)^{-1} \tag{25}$$

And then the signal projection expression becomes

$$\hat{y} = X\beta = X(X^T X)^{-1} X^T y \tag{26}$$

If we let

$$\Phi = X(X^T X)^{-1} X^T \tag{27}$$

Then we can compute the estimates of the optimal signal and its variances from:

$$\hat{y} = \Phi y \tag{28}$$

The estimates of the signal variances can be computed from:

$$C_{\hat{y}} = \Phi C_y \Phi^T$$

$$C_{\hat{y}} = \sigma^2 \Phi \Phi^T$$

$$C_{\hat{y}} = \sigma^2 \Phi \tag{29}$$

The symbol $\boldsymbol{\Phi}$ therefore may be called the transfer operator or the signal filter or the signal projector which serves to project the observation vector on the tangent plane. It is a symmetric idempotent operator since the multiplications by itself leaves the matrix unchanged (Mikhail and Gracie, 1981; Kreyszig, 1978). This can be seen from the following:

$$\Phi\Phi = X(X^T X)^{-1} X^T X (X^T X)^{-1} X^T = X(X^T X)^{-1} X^T = \Phi$$

$$\Phi\Phi^T = X(X^T X)^{-1} X^T X (X^T X)^{-1} X^T = X(X^T X)^{-1} X^T = \Phi$$

$$\Phi^T\Phi = X(X^T X)^{-1} X^T X (X^T X)^{-1} X^T = X(X^T X)^{-1} X^T = \Phi$$

By substituting Eq. 28 into Eq. 19, we obtain the equation for the noise component as:

$$r = y - \Phi y = (I - \Phi) y \tag{30}$$

It is seen that the projector for the noise component of the data vector is given by:

$$(I - \Phi) \tag{31}$$

This is the operator which puts the error component in a direction normal to the tangent plane. It is also a *symmetric idempotent operator.*

Some times in experimental data analysis, it is necessary to compute estimates of the variances of the noise component perhaps for blunder detection, the formulas for this are derived as below: Using the error propagation calculus on Eq. 30, the estimates of the noise variances may be computed as follows:

$$\begin{aligned}
C_r &= (I-\Phi)C_y(I-\Phi)^T \\
C_r &= \sigma^2(I-\Phi)(I-\Phi)^T \\
C_r &= \sigma^2(I-\Phi-\Phi^T+\Phi\Phi^T) \\
C_r &= \sigma^2(I-\Phi-\Phi^T+\Phi^T) \\
C_r &= \sigma^2(I-\Phi)
\end{aligned} \tag{32}$$

It is important to note that if all sources of biases are removed from the input data, this orthogonal noise component becomes random and can be referred to as **pure white noise** as opposed to a **coloured noise** which may still contain traces of the signal (Mikhail and Gracie, 1981; Kreyszig, 1978; Kavanagh, 2014).

Nonetheless, the principle of the projection theorem states that these two projectors in Eq. 27 and Eq. 31 must be perpendicular, i.e. their inner product must be zero. We show that these two projectors meet this condition by computing their inner product as follows:

$$\begin{aligned}
&(I-\Phi),\Phi \geq \Phi^T(I-\Phi) = \\
&\Phi^T-\Phi^T\Phi = \Phi^T-\Phi^T = 0
\end{aligned} \tag{33}$$

This confirms that the orthogonality condition is satisfied. Hence, the optimal or orthogonal projectors or filters can be constructed directly from the tangent plane matrix ***X*** as:

$$\Phi = X(X^TX)^{-1}X^T,\ (I-\Phi) \tag{34}$$

Furthermore, the integrity of these variance formulas may be checked by computing the variances of the raw data vector from the model expression as follows:

$$\begin{aligned}
y &= X\beta + r \\
y &= (X \quad I)\begin{pmatrix}\beta \\ r\end{pmatrix} \\
C_y &= (X \quad I)\begin{bmatrix}C_\beta & 0 \\ 0 & C_r\end{bmatrix}\begin{pmatrix}X \\ I\end{pmatrix} \\
C_y &= XC_\beta X^T + C_r \\
C_y &= \sigma^2\Phi + \sigma^2(I-\Phi) \\
C_y &= \sigma^2[\Phi + I - \Phi] \\
C_y &= \sigma^2 I
\end{aligned}$$

which is the same as the variances we started with.

It is also the practice to replace the measurement variance often called the a priori variance factor with an estimated variance often called the a posteriori variance in the final analysis of the estimated quantities. The a-posteriori variance factor is computed as:

$$\hat{\sigma}_o^2 = \frac{r^Tr}{m-n} \tag{35}$$

where m is the number of rows and n is the number of columns.

This is used in all the computational models listed below.
It is appropriate to remark here that for a weighted optimization, every vector element in the vector space will be transformed to a weighted equivalent using the square root of the weight matrix. All the above formulations are therefore applicable to weighted optimization. The weighted approach is not covered in this paper.

Putting it together, the end result of orthogonal decomposition of measurement vector is to obtain values for the parameter vector and its variances in the first place and then other results purely for completeness and extended statistical analysis of the experimental results. Thus, the following is a listing of the orthogonal formulas for the complete outputs of an optimal estimation process:

- Parameter estimates and variances:

$$\boldsymbol{\beta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$
$$\boldsymbol{C}_{\boldsymbol{\beta}} = \hat{\boldsymbol{\sigma}}_{\boldsymbol{o}}^2(\boldsymbol{X}^T\boldsymbol{X})^{-1} \tag{36a}$$

- Idempotent orthogonal transfer operator

$$\boldsymbol{\Phi} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T \tag{36b}$$

- Signal estimates and variances:

$$\hat{\boldsymbol{y}} = \boldsymbol{\Phi}\boldsymbol{y}$$
$$\boldsymbol{C}_{\hat{\boldsymbol{y}}} = \hat{\boldsymbol{\sigma}}_{\boldsymbol{o}}^2\boldsymbol{\Phi} \tag{36c}$$

- Noise estimates and variances:

$$\boldsymbol{r} = (\boldsymbol{I} - \boldsymbol{\Phi})\boldsymbol{y}$$
$$\boldsymbol{C}_{\boldsymbol{r}} = \hat{\boldsymbol{\sigma}}_{\boldsymbol{o}}^2(\boldsymbol{I} - \boldsymbol{\Phi})$$
$$\hat{\boldsymbol{\sigma}}_{\boldsymbol{o}}^2 = \frac{\boldsymbol{r}^T\boldsymbol{r}}{m-n} \tag{36d}$$

Another objective of this paper is to explore the effect of orthogonalizing the columns of the tangent matrix $\boldsymbol{X}$ on all the optimal estimation output formulas listed above. In other words, to see the geometric implications and the simplification that can be expected from spectral decomposition of the tangent matrix. This is the subject of the next section.

2.4 Spectral Decomposition of the Tangent Matrix

Spectral decomposition is the process of decorrelating the column vectors of the tangent matrix so that they are orthogonal to themselves. In effect, this process reduces the coefficient matrices in the optimal output formulas to simpler forms by sweeping out the inner product of any two of the column vectors whenever such is indicated in the solution process. The orthogonalization of the n-column vectors of the tangent plane $\boldsymbol{X} = [\boldsymbol{P}_1, \boldsymbol{P}_2, \ldots, \boldsymbol{P}_n]$ to themselves is, in effect, to make their inner products zero, i.e $\boldsymbol{P}_i^T\boldsymbol{P}_j = 0 : i, j = 1 \ to \ n : i \neq j$. This process is a *De-Corellation of the column vectors of the tangent matrix***.** To achieve this, the classical Gram-Schmidt algorithm is used (Keerthi and Shevade, 2003; Zarowski, 2004; Chang and Paige, 2003). The steps are as discussed below:

Given a set of *n* column vectors $\boldsymbol{P}_1, \boldsymbol{P}_2, \ldots, \boldsymbol{P}_n$, find an equivalent set $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ which are orthogonal. We take $\boldsymbol{P}_1$ and equate it to $\boldsymbol{v}_1$. We now remove from $\boldsymbol{P}_2$ the component of $\boldsymbol{P}_2$ lying in the direction of $\boldsymbol{v}_1$ and the remainder will be orthogonal to $\boldsymbol{v}_1$. This remainder vector then gives us $\boldsymbol{v}_2$. Also, by subtracting the components of $\boldsymbol{P}_3$ in the directions of $\boldsymbol{v_1}$ and $\boldsymbol{v_2}$, we obtain $\boldsymbol{v_3}$. This process can be repeated for all columns using the general formula in Eq. 37 (Zarowski, 2004 and Chang and Paige, 2003):

$$\boldsymbol{v_i = P_i - \sum_{j=1}^{i-1} \frac{(v_j^T P_i)}{(v_j^T v_j)} v_j} \tag{37}$$

For instance, when i = 1 Eq. 18 gives: $\boldsymbol{v_1 = P_1}$

For i = 2, we have: $\boldsymbol{v_2 = P_2 - \frac{v_1^T P_2}{v_1^T v_1} v_1}$

For i = 3, we have: $\boldsymbol{v_3 = P_3 - \frac{v_1^T P_3}{v_1^T v_1} v_1 - \frac{v_2^T P_3}{v_2^T v_2} v_2}$

Thus, the column vectors $\boldsymbol{v_1, v_2, \dots, v_n}$ are orthogonal and may be used in place of the original axes $\boldsymbol{P_1, P_2, \dots, P_n}$. When the orthogonal vectors $\boldsymbol{v_1, v_2, \dots, v_n}$ are substituted in each of Eq. 36a-d and simplifying, we have:

$$\boldsymbol{\begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{bmatrix} \frac{1}{v_1^T v_1} & 0 & 0 \\ 0 & \frac{1}{v_2^T v_2} & 0 \\ 0 & 0 & \frac{1}{v_3^T v_3} \end{bmatrix} \begin{pmatrix} v_1^T \\ v_2^T \\ v_3^T \end{pmatrix} y} \tag{38a}$$

And the variance matrix as:

$$\boldsymbol{C_\beta = \begin{bmatrix} \sigma_{\beta 1}^2 & 0 & 0 \\ 0 & \sigma_{\beta 2}^2 & 0 \\ 0 & 0 & \sigma_{\beta 3}^2 \end{bmatrix} = \begin{bmatrix} \frac{\hat{\sigma}_o^2}{v_1^T v_1} & 0 & 0 \\ 0 & \frac{\hat{\sigma}_o^2}{v_2^T v_2} & 0 \\ 0 & 0 & \frac{\hat{\sigma}_o^2}{v_3^T v_3} \end{bmatrix}} \tag{38b}$$

The idempotent transfer operator becomes:

$$\boldsymbol{\Phi = \left[\frac{v_1 v_1^T}{v_1^T v_1} + \frac{v_2 v_2^T}{v_2^T v_2} + \frac{v_3 v_3^T}{v_3^T v_3}\right]} \tag{38c}$$

- Signal estimates and variances:

$$\boldsymbol{\hat{y} = \Phi y = \left[\frac{v_1 v_1^T}{v_1^T v_1} + \frac{v_2 v_2^T}{v_2^T v_2} + \frac{v_3 v_3^T}{v_3^T v_3}\right] y}$$

$$\boldsymbol{C_{\hat{y}} = \hat{\sigma}_o^2 \Phi = \hat{\sigma}_o^2 \left[\frac{v_1 v_1^T}{v_1^T v_1} + \frac{v_2 v_2^T}{v_2^T v_2} + \frac{v_3 v_3^T}{v_3^T v_3}\right]} \tag{38d}$$

- Noise estimates and variances:

$$\boldsymbol{r = \left(I - \left[\frac{v_1 v_1^T}{v_1^T v_1} + \frac{v_2 v_2^T}{v_2^T v_2} + \frac{v_3 v_3^T}{v_3^T v_3}\right]\right) y}$$

$$\boldsymbol{C_r = \hat{\sigma}_o^2 \left(I - \left[\frac{v_1 v_1^T}{v_1^T v_1} + \frac{v_2 v_2^T}{v_2^T v_2} + \frac{v_3 v_3^T}{v_3^T v_3}\right]\right)}$$

$$\boldsymbol{\hat{\sigma}_o^2 = \frac{r^T r}{m-n}} \tag{38e}$$

It is noted that the orthogonalization of the column vectors has greatly simplified the computational equations. In fact, by carrying out the inversion of the now diagonal coefficient matrix of Eq. 36a, we obtain Eq. 39, from which the simple formulas in Eq. 30 emerge:
Carrying out the indicated multiplications we have:

$$\begin{pmatrix} \beta_2 \\ \beta_3 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} \frac{v_2^T y}{v_2^T v_2} \\ \frac{v_3^T y}{v_3^T v_3} \\ \frac{v_1^T y - \sum_{j=2}^{n}(v_1^T P_j)\beta_j}{v_1^T v_1} \end{pmatrix} \tag{39}$$

$$\beta_1 = \frac{v_1^T y - \sum_{j=2}^{n}(v_1^T P_j)\beta_j}{v_1^T v_1} \tag{40}$$

$$\sigma_{\beta 1}^2 = \frac{\hat{\sigma}_o^2}{v_1^T v_1}, \sigma_{\beta 2}^2 = \frac{\hat{\sigma}_o^2}{v_2^T v_2}, \sigma_{\beta 3}^2 = \frac{\hat{\sigma}_o^2}{v_3^T v_3}$$

$$\Phi = \left[\frac{v_1 v_1^T}{v_1^T v_1} + \frac{v_2 v_2^T}{v_2^T v_2} + \frac{v_3 v_3^T}{v_3^T v_3}\right] \tag{41}$$

Eq. 39 shows that the solution to the unknown parameters of an optimization 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 correct estimate of the first parameter in the manifold is achieved as a linear combination of the other parameters as shown in Eq. 40. Most significantly, the spectral decomposition of the tangent operator has actually simplified all the computational formulas and in fact led to a simultaneous extraction of both the parameter estimates and the variances. Furthermore, the idempotent filter matrix has been made geometrically transparent as the sum of the ratios of outer product and inner product of the orthogonal columns of the tangent matrix as in Eq. 41. These are demonstrated in the example applications below.

2.5 Summary of the Computational Steps

a) State the primary vectors

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}, \quad \boldsymbol{f} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix}$$

Specify the dimension *m* and *n* of the problem *m* must be greater *n.*
n is the number of parameters. ***X*** is a matrix of column vectors $\boldsymbol{P}_i$ obtained as derivatives of the function vector ***f*** with respect to each parameter up to n parameters and is written as:

$$\boldsymbol{X} = [\boldsymbol{P}_1, \boldsymbol{P}_2, \dots, \boldsymbol{P}_n] \tag{42}$$

b) Generate the columns of the tangent operator $\boldsymbol{X} = [\boldsymbol{P}_1, \boldsymbol{P}_2, \dots, \boldsymbol{P}_n]$

- If nonlinear model then linearize
- For i = 1 to *n compute the n columns of* ***X***

- $$P_i = \left[\frac{\partial f_1}{\partial \beta_i} \quad \frac{\partial f_2}{\partial \beta_i}, \quad \dots \quad \frac{\partial f_m}{\partial \beta_i}\right]^T, i = 1, 2, \dots, n \tag{43}$$

c) Generate the weight matrix operator **S**

If weighted optimization, then transform data and the axial manifolds using the square root of the weight super manifold as:

$$W = S^T S, \qquad S = W^{1/2} \tag{44}$$

$$y^* = Sy, \qquad P_i^* = SP_i \tag{45}$$

d) Initialize the computations

Compute certain quantities for later use

$$P_1^T y, P_1^T P_1, P_1^T P_j : j = 2 \; to \; n \tag{46}$$

e) Generate Orthogonal columns of the tangent matrix

For i=1 to n generate the orthogonal columns

$$v_i = P_i - \sum_{j=1}^{i-1} \frac{(v_j^T P_i)}{(v_j^T v_j)} v_j \tag{47}$$

f) Compute parameters and variances

g) $$\begin{pmatrix}\beta_1\\ \beta_2\\ \beta_3\end{pmatrix} = \begin{pmatrix}\frac{v_1^T y}{v_1^T v_1}\\ \frac{v_2^T y}{v_2^T v_2}\\ \frac{v_3^T y}{v_3^T v_3}\end{pmatrix} \tag{48}$$

$$\beta_1 = \frac{P_1^T y - \sum_{j=2}^{n} (P_1^T P_j)\beta_j}{P_1^T P_1} \tag{49}$$

$$\sigma_{\beta 1}^2 = \frac{\hat{\sigma}_o^2}{v_1^T v_1}, \sigma_{\beta 2}^2 = \frac{\hat{\sigma}_o^2}{v_2^T v_2}, \sigma_{\beta 3}^2 = \frac{\hat{\sigma}_o^2}{v_3^T v_3} \tag{50}$$

h) Compute a posteriori variance factor, noise vector and optimal signal vector

Idempotent orthogonal transfer operator

$$\Phi = \left[\frac{v_1 v_1^T}{v_1^T v_1} + \frac{v_2 v_2^T}{v_2^T v_2} + \frac{v_3 v_3^T}{v_3^T v_3}\right] \tag{51}$$

Signal estimates and variances:

$$\hat{y} = \Phi y$$

$$C_{\hat{y}} = \hat{\sigma}_o^2 \Phi$$

Noise estimates and variances:

$$r = (I - \Phi) y$$

$$C_r = \hat{\sigma}_o^2 (I - \Phi)$$

$$\hat{\sigma}_o^2 = \frac{r^T r}{m - n}$$

i) Validate result.

3.0 RESULTS AND DISCUSSION

3.1 Application problem

To demonstrate the effectiveness of the orthogonal and spectral decomposition approach of LS optimization, a simple set of equally weighted line measurements listed in Wolf and Ghilani (1997) are adjusted to obtain the optimal estimates. The three distances were taken between points A, B, and C on a line. The problem is to determine the most probable values of the distances AB and BC. The measured values are: AC = 431.71 ft, AB=211.52 ft and BC=220.10 ft

3.2 The solution

This problem can be formulated as an indirect observation of the required line lengths as the parameters.

3.3 The Algorithm

a) State the primary manifolds

$$y = \begin{pmatrix} 431.71 \\ 211.52 \\ 220.10 \end{pmatrix}; \quad f = \begin{pmatrix} x + y \\ x \\ y \end{pmatrix}, P_1 = \frac{df_i}{x} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \qquad P_2 = \frac{df}{dy} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

Specify the dimension *m* and *n* of the problem

$$m = 3, n = 2$$

Generate the Weight Transformation Operator S

Equally weighted, so identity transformation

b) Initialize the computations

$$\boldsymbol{P_1^T y} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} 431.71 \\ 211.52 \\ 220.10 \end{pmatrix} = \begin{pmatrix} 431.71 \\ 211.52 \\ 0 \end{pmatrix} = 643.23$$

$$\boldsymbol{P_1^T P_1} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = 2$$

$$\boldsymbol{P_1^T P_2} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = 1$$

$$\boldsymbol{V_1} = \boldsymbol{P_1} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, V_1^T V_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = 2$$

$$\boldsymbol{V_1^T P_2} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = 1$$

When $i = 2$ we have V_2 and it is determined as follows:

$$\boldsymbol{V_2} = \boldsymbol{P_2} - \frac{\boldsymbol{V_1^T V_2}}{\boldsymbol{V_1^T V_1}}\boldsymbol{V_1} = \begin{pmatrix}1\\0\\1\end{pmatrix} - \frac{1}{2}\begin{pmatrix}1\\1\\0\end{pmatrix} = \frac{1}{2}\left[\begin{pmatrix}2\\0\\2\end{pmatrix} - \begin{pmatrix}1\\1\\0\end{pmatrix}\right] = \frac{1}{2}\begin{pmatrix}1\\-1\\2\end{pmatrix}$$

$$\boldsymbol{V_2^T V_2} = \frac{1}{2}\begin{pmatrix}1\\-1\\2\end{pmatrix}\frac{1}{2}\begin{pmatrix}1\\-1\\2\end{pmatrix} = \frac{1}{4}\begin{pmatrix}1\\1\\4\end{pmatrix} = \frac{6}{4}$$

$$\boldsymbol{V_2^T y} = \frac{1}{2}\begin{pmatrix}1\\-1\\2\end{pmatrix}\begin{pmatrix}431.71\\211.52\\220.10\end{pmatrix} = \frac{1}{2}\begin{pmatrix}431.71\\-211.52\\440.20\end{pmatrix} = \frac{660}{2} = 330.195$$

c) Compute parameters

$$b = \frac{\boldsymbol{V_2^T y}}{\boldsymbol{V_2^T V_2}} = \frac{330.195}{6/4} = 220.13$$

$$a = \frac{\boldsymbol{P_1^T y} - \left(\boldsymbol{P_1^T P_2}\right)}{\boldsymbol{P_1^T P_1}}\beta_2 = \frac{643.23 - (1) \times 220.13}{2}$$

$$a = \frac{643.23 - 220.13}{2}$$

$$a = 211.55$$

$$\sigma_b^2 = \frac{\mathbf{1}}{\boldsymbol{V_2^T V_2}} = \frac{1}{6/4} = \frac{4}{6} = \frac{2}{3}$$

$$\boldsymbol{\beta} = \begin{bmatrix}a\\b\end{bmatrix} = \begin{bmatrix}211.55\\220.13\end{bmatrix};\ C = \begin{bmatrix}\sigma_a^2\\\sigma_b^2\end{bmatrix} = \begin{bmatrix}2/3\\2/3\end{bmatrix}$$

d) Validation: *These results agree with the values computed by Wolf and Ghilani, (1997) through regular least squares method as shown in **Table 1**.*

Table 1: Comparison of Results of Regular Matrix and Orthogonal Decomposition Approaches

S/N	Parameter	Regular Matrix Approach	Orthogonal Decomposition Approach
1	a	211.55	211.55
2	b	220.13	220.13

3.4 Discussion of Results

A comparison of the results from the spectral decomposition with those of a full matrix method of least squares computation shows that the orthogonal-spectral decomposition methods yields the same results. An examination of the computational steps in the above example shows a series of inner product calculations and arithmetic operations of addition, multiplication and division of inner products. Considering the mechanics of the method, it is seen that the processes of generating the design (or tangent) matrices, idempotent matrices orthogonal decomposition and the spectral decomposition and the spectral decomposition of the tangent matrix have effectively reduced the usually computational intensive least squares problem into simple systematic sequence of scalar value computations as can be seen from Eqs. 38 and 40. In addition, the method also provides an insight into the geometrical structure of a least squares process especially by reducing the problem into a filtering process as seen in Eqs. 28 and 30 respectively. The transfer matrix is also simplified by the spectral decomposition as seen in Eq. 41. The check on the numerical values of the parameters as computed in section 3.3 above confirms that the orthogonal-spectral decomposition process is effective. The form of the formulas makes it easy to understand the geometric basis of the LS problem. The orthogonalization followed by the spectral decomposition of the kernel matrix expounds the inner geometry of the optimal estimation process.

4.0 CONCLUSION

The findings in this paper can be summarized as follows.

i) The method of direct orthogonal and spectral decompositions of the measurement functions tangent matrix 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 vector-valued functions is the sum of the products of their corresponding elements.

ii) The sequence of the solution steps holds the promise of huge savings in the calculation efforts when additional parameters and thus more columns are to be included in the solution such as when trying out different models on the same dataset.

iii) 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. This could form a basis for further investigation in the near future. We however recognize that savings in memory poses no serious challenge in modern times of small but powerful computer systems.

iv) 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.

REFERENCES

Bjiorck, A., Heggernes, P., and Matstoms, P. (2000). Methods for Large Scale Total Least Squares Problems. *SIAM Journal of Matrix and Analytical Applications*, 22(**2**): 413–429.

Chang, X., and Paige C. C. (2003). An Orthogonal Transformation Algorithm for GPS Positioning. *SIAM J. SCI. COMPUT*., 24(**5**): 1710–1732

Fu H. and Barlow J. (2004). A regularized structured total least squares algorithm for high-resolution image reconstruction. *Linear Algebra Applications*, 391(**1**): 75–98.

Ghilani, C.D., and Wolf P.R., (2012), Elementary surveying: An introduction to geomatics, 13th ed. London: Pearson Education, 2012, 1-931

Kavanagh, B.F., (2014), Surveying Principles and Applications. Prentice Hall, New Jersey (9th Edition) 2014, 1 - 234

Keerthi S. S. and Shevade S. K. (2003). SMO Algorithm for Least-Squares SVM Formulations. *Neural Computations,* 15(**2**): 487 – 507.

Kreyszig, E. (1978). *Introductory Functional Analysis with Applications*. New York: John Wiley & Sons, Inc., 1978, 60-102

Liska R., Shashkov M. and Ganzha V. (2004), Analysis and optimization of inner products for mimetic finite difference methods on a triangular grid. *Mathematics and Computers in Simulation* 67: 55–66

Luenberger, David G. (1969). Optimization by Vector Space Methods. New York: John Wiley & Sons Inc., 1969, 1- 342.

Marcoux L. W. (2013), An Introduction to Functional Analysis, University of Waterloo press, Ontario, 2013 1-146

Maddox I. J. (1988). Elements of Functional Analysis. Cambridge: Cambridge University Press, 1988 1-242

Markovsky K. I. and Huffel S. V. (2005a). Consistency of the Structured Total Least Squares Estimator in a Multivariate Errors-in-Variables model. *Journal of Statistical Planning Inference*, 133 (**2**): 315–358.

Markovsky I. and Huffel S. V. (2005b). High-performance Numerical Algorithms and Software for Structured Total Least Squares. *Journal of Computer and Applied Mathematics*, 180(**2**): 311–331.

Mikhail, E. M., and G. Gracie, (1981). *Analysis and Adjustment of Survey Measurements*. New York: Van Nostrand Reinhold Co., 1981, 1-55.

Olaleye J. B., O. E. Abiodun and J. O. Olusina (2012). An Algorithm for Least Squares Computation Using the Manifold and Hilbert Space Dialectics. *Theoretical Mathematics and Applications*, 2(**1**): 143 – 160.

Wolf P. R. and C. D. Ghilani (1997). Survey Measurement and Adjustment by Least Squares in *The Surveying Handbook*, Russel C. Brinker & Roy Minnick Eds, New Delhi: CBS Publishers & Distributors Pvt. Ltd.,

Zarowski, C. J. (2004). *An Introduction to Numerical Analysis for Electrical Engineers.* John Wiley and Sons Inc., New Jersey.