Transactions of the Nigerian Association of Mathematical Physics
***Volume 11, (January - June, 2020 Issue*), pp173 – 180**
© Trans. of NAMP

INFORMATION CRITERIA ON REGULARIZED LEAST SQUARES PROBLEM

Stephen Ehidiamhen Uwamusi

Department of Mathematics, Faculty of Physical Sciences, University of Benin, Benin City, Edo State, Nigeria.

Abstract

The paper aims to give concise information for treating least squares problems. Useful estimates for computing optimal regularized low Rank inverse approximation to include the SVD and the probability measure encompassing the Wallis factor and Gamma density estimation for the subspace solution have been detailed. We also obtained the square root of the symmetric matrix from the normal equation from the least squares equation using the Lagrangre interpolation formula. Numerical example has been demonstrated with the described methods.

Keywords: Optimal low rank approximation, svd, chevbyshev semi-iterative method, QR –cholesky – factorization, ellipsoid, wallis factor, gamma density estimators, matrix square root.

1. Introduction

The first thing that comes to mind whenever we are given a set of data $d(x_i, y_i), i = 1,2,...,m$ which aims at constructing possibly, a line of best fit by the function $\varphi(x, t_i)$, for the parameters describing the data is the least squares. This is the simplest aspect of this type of problems. However, the aim of this paper hopes to give more insights than what is anticipated with precise useful information. The use of Singular values decomposition has gained prominence in recent years. Its main application areas are in systems requiring matrix rank determination, system identification, reliability and risk analysis, antenna beam formation and recently, mathematical biology and a number of many other methods [1], e.g.

We start off by giving general descriptive ideas on the least squares equation designated in the form:

$$\min_s \|F(x_k) - J(x_k)s\|_2 \qquad (1.1)$$

The solution to equation (1.1) is written as

$$x_{k+1} = x_k + s_k \text{ (k=1,2,..)}, \qquad (1.2)$$

Where, s_k is obtained by repeatedly solving system of normal equation

$$(A^T A)s_k = -A^T f(x_k) \qquad (1.3)$$

The matrix A is the Jacobian obtained from data which represents $J(x_k)$ and $A \in R^{m\times n}$, $m > n$.

As a special case, stringent conditions when $F \in R^{n\times n}$, is described in the form:

Definition 1.1[2]: The function $F \in R^{n\times n}$ satisfies a Lipschitz condition in a domain $D \subset X$, if a constant η is in existence, called a Lipschitz constant such that $\|F(x') - F(x'')\| \le \eta \|x' - x''\|, \forall x', x'' \in D$. If $\eta < 1$, F is called a contraction mapping provided Miranda's theorem holds verbatim.

The non smooth analysis for F is discussed in the context of continuity for a useful purpose. By semi-smoothness is implied the behaviour of singularity of Jacobian $F'(x^*)$ at the solution and Lipschtz continuity for F'. By a theorem due to Fowler and Kelly (2005) we have a definition for F in terms of smoothness.

Definition 1.2 [3]. F is semi-smooth at $x \in R^n$ if and only if $\lim\limits_{s\to 0, A\in\partial F(x+s)} \dfrac{\|F(x+s) - F(x) - As\|}{\|s\|} = 0$, where the directional derivatives imply

Correspondence Author: Stephen E.U., Email: Stephen_uwamusi@yahoo.com, Tel: +2348020741193

that $F^{/}(x:s)=\lim_{h\to 0}\frac{F(x+hs)-F(x)}{h}$ and, $x,s\in R^n$.

We link regularity space with directional derivatives for the function F.

Given that F is semi smooth, $F(x^*)=0$, assuming further all matrices in $\partial F(x^*)$ are nonsingular, there are matrices K and ε such that given the ball $B(x^*,\varepsilon)$ and $A\in\partial F(x)$, then $\|A^{-1}\|\le K$ remains valid. Thus if matrix A has an inverse that is bounded, the Jacobian matrices $A(x)$ corresponding to system of Equation (1.1) are not only $F-$Suslin and Polish, but also ultrabarrelled and possess Baire's second Category theorem.

The error estimate and order of convergence using approximate Jacobian matrices by the Finite difference methods with computed results for the solution by the Newton iteration to system of Equation (1.1) thereof are stated in the form:

Theorem 1.1 [3]. Let $F:D\subset R^n\to R^n$ be given, such that $F(x^*)=0$.

Assuming further that F is semi-smooth at x^* and if all matrices in the Jacobian matrices $\partial F(x^*)$ exist. There are parameters $\bar{\eta},\bar{\delta},M>0$ such that if $x_0\in B\left(x^*,\bar{\delta}\right)$ and $\eta_k\le\bar{\eta}$ for which the inexact Newton iteration of Equation (1.2) converges to desired solution x^* with additional hypothesis given by the equation

$$\|e_{k+1}\|\le M\eta_k\|e_k\|+0\|e_k\|,$$

Where, the order of convergence p of F at x^* is

$$\|e_{k+1}\|\le M\left(\eta_k\|e_k\|+\|e_k\|^{p+1}\right).$$

Basic toolsin our favour are Singular values decomposition, QR-Cholesky-Factorization , Preconditioned Conjugate Gradient methods coupled with Tikhonov regularization and Chevbyshev –Semi iterative techniques in order to cope with a highly ill-conditioned system. Not too long ago, luckily enough , Chevbyshev-Semi-iterative method has gained several appeals in providing solution to Least squares problems.

The algorithmic form of Chevbyshev-Semi-iterative method is presented as follows in the sense of [4].

ALGORITHM

1) Input the matrix $A\in R^{m\times n}$, vector $b\in R^m$, tol. ε -order of accuracy.
2) Decompose $A=QR$ -Cholesky factorization for which $0<\sigma_L\le\sigma_U$ where all non zero singular values $\sigma_i\in A$ are contained in the interval $[\sigma_L,\sigma_U]$;
3) Set mid $\sigma=\frac{(\sigma_U^2+\sigma_L^2)}{2}$;rad $\sigma=\frac{(\sigma_U^2-\sigma_L^2)}{2}$;
4) Initialize $x=0, v=0$ and $r_0=b$;
5) For $k=0,1,\ldots,\left[\frac{(\log\varepsilon-\log 2)}{\log\frac{(\sigma_U-\sigma_L)}{(\sigma_U+\sigma_L)}}\right]$ do
6) $\beta=\begin{cases}0 & if\ \ k=0\\ \frac{1}{2}(rad\sigma/mid\sigma)^2, & if\ \ k=1\\ (\alpha(rad\sigma)/2)^2 & otherwise\end{cases}$;
7) $\alpha=\begin{cases}\frac{1}{mid\sigma} & ;\ if\ \ k=0\\ mid\sigma-(rad\sigma)^2/(2rad\sigma); & if\ k=1\\ \frac{1}{((mid\sigma)-\alpha(rad\sigma)^2/4)} & ;\ otherwise\end{cases}$;

   $v\leftarrow\beta v+A^T r_k$;
8) $x\leftarrow x+\alpha v$;
9) $r_k\leftarrow r_k-\alpha Av$
10) If solution found write $x=x_{k-1}$ and quit endif
11) Else repeat steps 5 to 9
12) end for

(13) end

The QR iterative refinement with residual vector is given below.
Algorithm:

1) Let $\hat{x}$ solve $R^T Rx = A^T b$
2) Compute $\bar{r} = b - A\bar{x}$
3) Solve $R^T Rw = A^T \bar{r}$
4) Compute $y = \bar{x} + w$

2. Computing Optimal Regularized Low Rank Inverse Approximation

Particularly, we are interested in application in the Low rank approximation of a matrix in the form:

$$\min_{rank(\bar{B}) \le r} \|(BA - I_n)\|_F^2 + \lambda^2 \|B\|_F^2 \quad (2.1)$$

In equation (2.1), the term B is a special matrix aimed at approximating the inverse of the matrix A while I is an identity matrix with a scalar λ being the Tikhonov regularization parameter aimed to penalize the equation (2.1).
Low rank matrix approximation is mainly used for data processing technique which helps to reduce noise present in the data and is applicable in such varied areas such as signal processing, compressed sensing, image and pattern recognition, and many other Engineering practices.
The singular value decomposition for A is given by $A = U \sum V^T$; where $U \in R^{m\times m}, V \in R^{n\times n}$.The matrix B is an approximate inverse to A. Practically, we use the SVD computation for the approximate inverse in place of B in the form:

$$B = V_A\left(\Sigma_A^T \Sigma_A + \lambda^2 I\right)^{-1} \Sigma_A U_A^T, \quad (2.2)$$

Where the global minimiser $\hat{B} \in R^{n\times m}$ is then given by the equation

$$\hat{B} = \underset{rank(\hat{B}) \le r}{\arg \min} \|(BA - I_n M)\|_F^2 + \lambda^2 \|B\|_F^2 \quad (2.3)$$

Using the above information given, the low rank approximation for the matrix A in terms of SVD is in the form:

$$[U, \Sigma, V] = \underset{U,\Sigma,V}{\arg\min} \left\{ \|A - U \Sigma V^T\|_2^2 + \frac{\lambda_u}{2}\|B_u U\|_F^2 + \frac{\lambda_v}{2}\|B_v V\|_F^2 \right\} \quad (2.4)$$

The matrices appearing in Equation (2.4) are compatible matrices which may be set to a second order finite difference matrices [5],[4],and [6].
By further setting that: $b = A^T f(x)$, and writing the ill-posed problem in the form:

$Ax = b + \varepsilon$, (2.5)

where $\varepsilon \sim N(0, \sigma^2)$, with unknown variance σ^2, and mean 0 has the introduced noise ε that is calibrated to a factor $\frac{\|\varepsilon\|_2^2}{\|Ax\|_2^2} = 0.01$.The corresponding standard Tikhonov regularization for the Pseudo-inverse matrix with respect to λ^2 is that matrix $A(\lambda^2) = A\left(A^T A + \lambda^2 I\right)^{-1} A^T$,it has negative log likelihood function:

$$\ell\left(x(\lambda^2), A, b\right) = -\log \prod_{i=1}^{m} \left(\frac{1}{\hat{\sigma}\sqrt{2\pi}} \exp\left(-\frac{\left(Ax(\lambda^2 - b)\right)_i^2}{2\hat{\sigma}^2} \right) \right) = \frac{m}{2}\log 2\pi \hat{\sigma}^2 + \frac{\|Ax(\lambda^2) - b\|^2}{2\hat{\sigma}^2}$$
$$= \frac{m}{2}\log\frac{2\pi}{m} + \frac{m}{2}\log\|Ax(\lambda^2) - b\|^2 + \frac{m}{2}. \quad (2.6)$$

Where, the bias correction term is given by $\eta(\lambda^2, A, b) = P^{eff}\left(\lambda^2\right) = TrA\left(\lambda^2\right)$.
As a special case, is the iterated unregularized Land webber method. The method is described in the form. Starting with the sequence

$$x_k = x_{k-1} + \lambda A^T (b - Ax_{k-1}), \qquad (x_0 = 0, k = 1,2,...,.) \quad (2.7)$$

With λ, a fixed parameter. Now at the kth step in the iteration we have that

$$x_k = \left(I - \lambda A^T A\right)^{Tk} x_0 + \sum_{i=0}^{k-1}\left(I - \lambda A^T A\right)^i \left(A^T b\right) \quad (2.8)$$
$$= \left(A^T A\right)^{-1}\left(I - \left(I - \lambda A^T A\right)\right)^k \left(A^T b\right)$$

We decompose the matrix $A^T A = V\left(\Sigma^T \Sigma\right)V^T$ where, $I = VV^T$. Therefore, using the information as above, we write that:

$$x_k = \sum_{i=1}^{k}\left(1 - \left(1 - \lambda\sigma_i^2\right)^k\right)\left(\frac{u_i^T b}{\sigma_i}\right) v_i \quad (2.9)$$

Thus for convergence, it is necessary that $\left|1-\lambda\sigma_i^2\right|<1$, which means that $0<\lambda\sigma_i^2<2$. This upper bound for $\lambda<\frac{2}{\sigma_i^2}\ \forall i$. Let us note that we often take the value of $\lambda=\sigma_1^{-2}$ in Equation (2.7).

For a value of $k\to\infty$, it would yield that $\lim_{k\to\infty}\frac{\left(\lambda+\sigma_i^2\right)^k-\lambda^k}{\left(\lambda+\sigma_i^2\right)^k}=\lim_{k\to\infty}\left(1-\frac{\lambda^k}{\left(\lambda+\sigma_i^2\right)^k}\right)=1$ for very small value of λ.

3. The Ellipsoid for the Data Problem and accompany metric topology.

It is important in providing ellipsoid for the described data. Before continuing, we give the expository basic set topology necessary for understanding the intention in the paper.

Definition 2.1. Let $D:X\to Y$ be a set valued map and $y_0\in D(x_0)$ be given with X and Y the metric space .Assuming further that $\delta:R_+\to R_+$ be a strictly monotone continuous function with $\delta(0)=0$,the statements following hold for adoption in our work:

(i) F is δ-open around (x_0,y_0) if there is a neighbourhood U of x_0 and a neighbourhood W of y_0 such that $B_{\delta(t)}(z)\subset F(B_t(x)),\forall x\in U,z\in w\cap F(x),t>0$, for which, $B_t(x_0)\subseteq U$.

(ii) F is approximately δ-open around (x_0,y_0) if there is some non-negative function

$\kappa:R_+\to R_+$ with $\lim_{t\downarrow 0}\left(\delta^{-1}(\kappa(t))/t\right)<1$,given U of x_0 and a neighbourhood W of y_o such that for all $x\in U,z\in w\bigcap F(x)$ and $t>0$ with $B_t(x_0)\subset U$ implies that

$$B_{\delta(t)}(z)\subset\overline{B_{\kappa(t)}[F(B_T(x))]}. \tag{3.1}$$

From the above preambles there follows:

F is δ-regular around (x_0,y_0) assuming one can find a constant $\kappa>0$, a neighbourhood U of x_0, a neighbourhood W of y_0 such that for all $z\in W$ and all $x\in U$ with $F(x)\bigcap w\neq 0$ induces metric topology

$$d\left(x,F^{-1}(z)\right)\le\kappa\delta^{-1}\left(d(z,F(x))\right). \tag{3.2}$$

Definition 2.2. A mapping $F:D\subset R^n\to R^n$ is a homeomorphism of D onto $F(D)$ if F is one-to-one on D and F and F^{-1} are continuous on D and $F(D)$, respectively.

Definition 2.3. A mapping $F:D\subset R^m\to R^n$ is Holder continuous on $D_0\subset D$ if there exist constants $\omega\ge 0$ and $p\in(0,1)$ such that, for all $x_1,x_2\in D$, we have that

$\|F(x_2)-F(x_1)\|\le\omega\|x_2-x_1\|^p$. If $p=1$, F is Lipschitz-continuous on D_0.

Motivated by the above details, the homeomorphism it carries is stated below:

- Let X and Y be metric spaces and $F:X\to Y$ be a set valued map. Suppose $\delta:R_+\to R_+$ is strictly monotone i.e., continuous function and $y_0\in F(x_0)$ .We say that F^{-1} is δ^{-1} Lipschitz continuous (LSC) around (x_0,y_0) if there exist a neighbourhood W of y_0 and a neighbourhood U of x_0 such that (i) $F^{-1}(y)\bigcap U\neq\varphi$,for all $y\in w$;(ii) there exists $\kappa>0$ with the property that for every $y_1,y_2\in w$ the inclusion holds verbatim such that:

$$F^{-1}(y_1)\bigcap U\subset B_{k\delta^{-1}(d(y_1,y_2))}\left(F^{-1}(y_2)\bigcap U\right), \tag{3.3}$$

where $\delta(t)=ct^r$ for some $c>0,r>0$. The map δ^{-1}-LSC around $(x_0,y_0)$is pseudo-Holder at the rate r around (y_0,x_0).We noted that when $r=1$, it is pseudo- Lipschitz on (y_0,x_0).

After all these we are now in a position to consider the closed graph implied by the algorithm as stated earlier. Take $|x|=\|x\|+\|Ax\|$. If $|x_n|$ is Cauchy, then $x_n\to x\in X$ as $Ax_n\to y\in Y$ .The closed graph induces that $Ax=y$ and $|x_n-x|\to O$. By introducing uniform boundedness theorem, that $F_M=\{x:|F(x)|\le M\}$ and taking into existence of Baire category for some F_M containing a ball $B(s,r)$, there is $\|x\|\le r$ which produces $|F(x)|=|F(s+x)-F(s)|\le M+M_s$. Thus $\|F\|$ is uniformly bounded by the factor $\frac{(M+M_s)}{r}$.

Having taken into consideration that X,Y are separable metric spaces and, Y complete, the point wise mapping $F:X\to Y$ is computable provided there exists a strongly continuous tightening. The ellipsoid described by the minimum generator X for the least squares problem 1.1 is obtained using p-norms. Consider $p>1$ and q which satisfy the unit ball $\frac{1}{p}+\frac{1}{q}=1$ .By recalling that $\alpha\beta\le\frac{\alpha^p}{p}+\frac{\beta^q}{q}$ it is easy to verify the convexity $e^{\lambda x+(1-\lambda)y} \le \lambda e^x + (1-\lambda)e^y$. The result is proven by setting $\lambda = \frac{1}{p}, x = p\log\alpha,$ and $y = q\log\beta$. Therefore Holder inequality gives $\left|x^T y\right| \le \|x\|_p + \|y\|_q$, for $\frac{1}{p}+\frac{1}{q}=1, p\ge 1$. The unit ball $\{x : \|x\| \le 1\}$ gives the elliptic norm. Using the above it holds that $\|x\|_q \le \|x\|_p \le n^{\left(\frac{1}{p}-\frac{1}{q}\right)}\|x\|_q, 1\le p\le q\le\infty$. By taking $p=2$ and $q=\infty$ it would yield that $\|x\|_\infty \le \|x\|_2 \le \sqrt{n}\|x\|_\infty$ bounded by a factor $\sqrt{n}$. Thus for a vector valued function $f : |t_0, t_k| \to R^n$ the ℓ_p-norm for $p \ge 1$ is given by

$$\|f\|_{\ell_p} = \left(\int_{t_0}^{t_k} \|f(x)\|_p^p dx\right)^{\frac{1}{p}}, \quad (3.4)$$

Where ,

$$\|f(x)\|_p = \left(\sum_{i=1}^{n} |f_i(x)|^p\right)^{\frac{1}{p}}, \ \|f\|_{\ell_1} = \int_{t_0}^{t_k}\sum_{i=1}^{n} |f_i(x)|dx, \ \|f\|_{\ell_\infty} = esp\sup\left(\max_i |f_i(x)|\right).$$

We give the probability measure for the computed errors in the solution. Let $\hat{x} \in R^n$, and suppose S is selected uniformly and randomly from the unit sphere S_{n-1} in n-dimension, the expected value of $\left|\hat{x} s\right|$ is defined to be

$$E\left(\left|\hat{x}^T s\right|\right) = \left\|\hat{x}\right\| E_n. \quad (3.5)$$

The Wallis factor E_n is given by

$E_1 = 1, E_2 = \frac{2}{\pi},$ and for $n > 2$, we have that

$$E_n = \frac{1.3.5.7....(n-2)}{2.4.6.8...(n-1)}, \quad \text{fornodd;} \quad (3.6)$$

$$E_n = \frac{2}{\pi}.\frac{2.4.6.8...(n-2)}{1.3.5.7.(n-1)}, \text{ for n even.} \quad (3.7)$$

The estimate for E_n is given by the equation

$$E_n = \sqrt{\frac{2}{\pi(n-0.5)}}. \quad (3.8)$$

By further setting $\tau = \left|\hat{x}^T s\right| * E_n$ as an approximation for $\left\|\hat{x}\right\|$, and for $\upsilon > 1$, the probability measure is given by

$$\Pr\left(\frac{\left\|\hat{x}\right\|}{E_n} \le \tau \le E_n\left\|\hat{x}\right\|\right) \ge 1 - \frac{2}{\pi E_n} + 0(inv(E_n)). \quad (3.9)$$

We move further to discuss the gamma density and the moment estimation of parameters and construct $75\%, 95\%, 98\%$ confidence intervals associated with the set of computed vector solution for purpose of reliability in terms of survival and risk analysis.

Given that a random variable $X = x_1, x_2, \dots, x_n \ (n \ge 3)$ be drawn from a sample population with i.i.d random variable, with a probability density function $f(x)$. If the independence of sample mean $\bar{X}_n$ and sample coefficient of variation $V_n = \frac{S_n}{\bar{X}_n}$ can be calculated, the corresponding gamma density for the rate parameter and is defined in the form

$$gamma(x,\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \ x > 0, \alpha > 0, \beta > 0. \quad (3.10)$$

where S_n is the standard deviation .

Drawn from standard Statistics Texts, it can be obtained that:

$E(\bar{X}_n^2) = \frac{(n\alpha+1)}{n^2\beta^2} n\alpha$; $E(S_n^2) = \frac{\alpha}{\beta^2}$, $E(X^k) = \frac{(\alpha+k-1)...(\alpha+1)\alpha}{\beta^k}$, for $k \ge 1$.

To compute α and β , the following procedure is adopted in the sense of [7]: $\bar{\alpha}_k = \bar{X}_n^2 / S_n^2,\ \bar{\beta}_k = \bar{X}_n / S_n^2.$ $E\left(\frac{S_n^2}{\bar{X}_n^2}\right) = \frac{n}{1+n\alpha}$ and $\lim_{n\to\infty}\left(\frac{n}{1+n\alpha}\right) = \lim_{n\to\infty}\frac{1}{\frac{1}{n}+\frac{n.\alpha}{n}} = \frac{1}{\alpha}$. The limiting value $\frac{1}{\alpha}$ is the square of the coefficient of variation which is an asymptotically unbiased estimator of the square of variation.

Assuming instead, we estimate the values of $\bar{\alpha}_i$ and $\bar{\beta}_i$ by the equation:

$\bar{\alpha}_i = \frac{1}{V_n^2} - \frac{1}{n}$, and $\bar{\beta}_i = \frac{\bar{\alpha}_i}{X_n} = \frac{1}{\bar{X}_n}\left(\frac{1}{V_n^2} - \frac{1}{n}\right)$, then it could be deduced that:

$$\operatorname{var}(S_n^2) = \frac{\alpha}{\beta^4}\left(\frac{2n\alpha}{(n-1)^2} + \frac{6}{n}\right),\quad \operatorname{var}\left(\frac{S_n^2}{\bar{X}_n^2}\right) = \frac{2\alpha(\alpha+1)}{(n-1)\left(\alpha+\frac{1}{n}\right)^2\left(\alpha+\frac{2}{n}\right)\left(\alpha+\frac{3}{n}\right)}. \tag{3.11}$$

Therefore for n large enough, the terms ($\bar{\alpha}_i$,respectively $\bar{\beta}_i$) $\to (\alpha, \beta)$ in that order.

We thus construct confidence intervals in the form of $75\%, 95\%, 98\%$ for values of $k = 1,2,3$ as shown below.

$$\Pr\left(\frac{\bar{X}_n^2}{S_n^2}\left(1 - \frac{k}{\sqrt{n-1}}\right) - \frac{1}{n} \le \alpha \le \frac{\bar{X}_n^2}{S_n^2}\left(1 - \frac{k}{\sqrt{n-1}} + \frac{1}{n}\right)\right) \tag{3.12}$$

An alternate form of expression given in Equation (3.12) in the form of ordinary interval is

$$\left(\frac{\bar{X}_n^2}{S_n^2}\left(1 - \frac{k}{\sqrt{n-1}}\right) - \frac{1}{n}, \frac{\bar{X}_n^2}{S_n^2}\left(1 - \frac{k}{\sqrt{n-1}}\right) + \frac{1}{n}\right), (k = 1,2,3) \tag{1.13}$$

Using the fact that $E(\ln(X)) = \psi(\alpha) - \ln(\beta)$, where ψ is a digamma function defined as $\psi(\alpha) \approx \log(\alpha) - \frac{1}{2\alpha}$, $\log\Gamma(\alpha) \approx \alpha\log(\alpha) - \alpha - \frac{1}{2}\log\alpha + const$ (Stirling), we give the information entropy in the form:

$$\begin{aligned} H(X) &= E[-\ln(P(X))] = E[-\alpha\ln(\beta) + \ln(\Gamma(\alpha)) - (\alpha-1)\ln(X) + \beta X] \\ &= \alpha - \ln(\beta) \end{aligned} \tag{3.14}$$

The Kullback-Leibler divergence is the expression given by

$$\begin{aligned} D_{KL}(\alpha_p, \beta_p, \alpha_q, \beta_q) &= (\alpha_p - \alpha_q)\psi(\alpha_p) - \log\Gamma(\alpha_p) + \log\Gamma(\alpha_q) + \alpha_q(\log\beta_p - \log\beta_q) + \\ &\quad \alpha_p\frac{\beta_q - \beta_p}{\beta_q} \end{aligned} \tag{3.15}$$

4. NUMERICAL ILLUSTRATION

Consider the set of data taken at secondary source in [8] as means of numerical weather computing. The experimental data were taking at a primary source from TRODAN at Anyigba, Kogi State University. The period ranges from 2010 to 2013 with average monthly records of Temperatures and relative Humidity as showed in Table 1.

Table 1

S/N	DATE/TIME	TEMPERATURE (O^0), Y	RELATIVE HUMIDITY $(\%)$, X
1	2011-04	32.002906	67.961866
2	2012-04	31.331634	68.431059
3	2013-04	29.887573	74.096862
4	2010-08	27.060463	82.752606
5	2011-08	26.438872	83.023268
6	2012-08	26.191286	83.452398
7	2010-12	30.039266	47.560099
8	2011-12	28.355457	37.122311
9	2012-12	28.638782	50.906093
10	2011-02	32.267793	60.666083
11	2012-02	31.043251	64.010414
12	2013-02	31.063575	55.738473
13	2011-01	29.062744	32.955172
14	2012-01	28.816789	41.495453
15	2013-01	29.158118	47.972735
16	2010-07	26.048773	83.517390
17	2011-07	27.902184	79.944623
18	2012-07	26.945129	82.449950
19	2011-06	28.900817	78.000009
20	2012-06	27.873762	79.053775
21	2011-03	33.950121	60.316848
22	2012-03	33.401677	54.788059
23	2013-03	33.303313	65.349271
24	2011-05	30.639803	73.495656
25	2012-05	29.369926	75.301975
26	2010-11	29.471345	76.122503
27	2011-11	28.170606	66.909216
28	2012-11	28.891984	76.298347
29	2010-10	28.214158	80.737601
30	2011-10	27.204551	80.169298
31	2012-10	27.401785	81.294369
32	2010-09	27.053905	82.342825
33	2011-09	26.616421	81.995541
34	2012-09	26.590674	83.569152

By using polynomial fit of degree 4, the following system of normal equation is constructed

$$Bx = b \qquad (4.1)$$

Where $B = A^T A,\ b = A^T y$ such that

$$1.0e+16*\begin{pmatrix} 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0006 \\ 0.0000 & 0.0000 & 0.0000 & 0.0006 & 0.0446 \\ 0.0000 & 0.0000 & 0.0000 & 0.0444 & 3.5129 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = 1.0e+10*\begin{pmatrix} 0.0000 \\ 0.0000 \\ 0.0005 \\ 0.0353 \\ 2.6598 \end{pmatrix} \qquad (4.2)$$

With solution

$$x = B \backslash b = \begin{pmatrix} 156.9696 \\ -9.6952 \\ 0.2595 \\ -0.0029 \\ 0.0000 \end{pmatrix},$$

$$eig(B) = 1.0e+16*\begin{pmatrix} 0.0000 \\ 0.0000 \\ 0.0000 \\ 0.0000 \\ 3.5134 \end{pmatrix}.$$

Then decompose $B = U \sum V^T$, with

$$U = \begin{pmatrix} 0.0000 & -0.0000 & -0.0018 & -0.0728 & -0.9993 \\ -0.0000 & -0.0007 & -0.0506 & -0.9961 & -0.0728 \\ -0.0002 & -0.0292 & -0.9983 & -0.0506 & -0.0019 \\ -0.0126 & -0.9995 & 0.0292 & 0.0008 & 0.0000 \\ -0.9999 & 0.0126 & -0.0002 & -0.0000 & -0.0000 \end{pmatrix},$$

$$D = 1.0e+16*\begin{pmatrix} 3.5134 & 0 & 0 & 0 & 0 \\ 0 & 0.0000 & 0 & 0 & 0 \\ 0 & 0 & 0.0000 & 0 & 0 \\ 0 & 0 & 0 & 0.0000 & 0 \\ 0 & 0 & 0 & 0 & 0.0000 \end{pmatrix};$$

$$V^T = \begin{pmatrix} 0.0000 & -0.0000 & -0.0018 & -0.0728 & -0.9993 \\ -0.0000 & -0.0007 & -0.0506 & -0.9961 & -0.0728 \\ -0.0002 & -0.0292 & -0.9983 & -0.0506 & -0.0019 \\ -0.0126 & -0.9995 & 0.0292 & 0.0008 & 0.0000 \\ -0.9999 & 0.0126 & -0.0002 & -0.0000 & -0.0000 \end{pmatrix}.$$

We compute the Wallis factor, and probability measure for the variable Temperature θ^0 (Y) in the form:

$$E\left|Y^T s\right| = \|Y\| E_n = 23.3965,\ \tau = \left|Y^T s\right| / E_n = 7.1598e+03;$$

$$pr\left(\frac{\|Y\|}{E_n} \le \tau \le E_n \|Y\|\right) = pr(1.2312e+03 \le 7.1598E+03 \le 23.3965) \ge 1 - \frac{2}{\pi E_n} = -3.6181$$

We also computed that

$$Cov(A) = 1.0e+14*\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0 & 0.0000 & 0.0000 & 0.0000 & 0.0003 \\ 0 & 0.0000 & 0.0000 & 0.0003 & 0.0300 \\ 0 & 0.0000 & 0.0003 & 0.0300 & 2.7224 \end{pmatrix};$$

Using Komolgorov Test Statistic for the gamma density function $gamma(x,\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$, we computed the following characteristics for the Relative Humidity

$$mean(X) = 68.7353,\ \mathrm{var}(X) = 220.3824,\ \alpha = \frac{\bar{X}^2}{S_n^2} = 0.0973,\ \beta = \frac{\bar{X}}{S_n^2} = 0.0014,\ E\left(\frac{S_n^2}{\bar{X}^2}\right) = \frac{n}{1+n\alpha} = 7.8934,\ E\left(\bar{X}_n^2\right) = \frac{(n\alpha+1)}{n^2\beta^2} n\alpha = 6.1530e+03,$$

$$E\left(S_n^2\right) = \frac{\alpha}{\beta^2} = 4.8568e+04,\ \mathrm{var}\left(\frac{S_n^2}{\bar{X}^2}\right) = 1.7641.$$

We now construct the probability confidence interval for the Relative Humidity data in the form:

$$pr\left(\frac{\bar{X}_n^2}{S_n^2}\left(1 - \frac{k}{\sqrt{n-1}}\right) - \frac{1}{n} \le \alpha \le \frac{\bar{X}_n^2}{S_n^2}\left(1 - \frac{k}{\sqrt{n-1}}\right) + \frac{1}{n}\right),\quad k = 1,2,3,...,. \qquad (4.3)$$

That is,

$$\Pr\left(0.0973\left(0.9706 - \frac{k}{33}\right) \le \alpha \le 0.0973\left(1.0294 + \frac{k}{33}\right)\right). \qquad (4.4)$$

On the other hand, we also can present in terms of intervals in the form of 75%,95%, and 98% confidence intervals for α as

$$\left(\left(\frac{\bar{X}_n^2}{S_n^2}\left(1-\frac{k}{\sqrt{33}}\right)-\frac{1}{34}\right),\frac{\bar{X}_n^2}{S_n^2}\left(1-\frac{k}{\sqrt{33}}\right)+\frac{1}{34}\right),\quad k=1,2,3,....\qquad (4.5)$$

As point of remark, we could have also constructed both probability measure and confidence interval for the computed solution for the least squares problems. Thus the same procedures apply verbatim.

We compute the square root for the matrix $B = A^T A$

Writing as the Taylor Series,

$$P(t) = f(\lambda_i) + f^{i}(\lambda_i)(t-\lambda_i) + f^{//}(\lambda_i)\frac{(t-\lambda_i)^2}{2!} + ... + f^{(m-1)}(\lambda_i)\frac{(t-\lambda_i)^{m_i-1}}{(m_i-1)!} \qquad (4.6)$$

And Hermite formula,

$$H(t) = \sum_{i=1}^{s}\left[\sum_{k=0}^{n-1}\frac{1}{k!}\varphi_i^{(k)}(\lambda_i)(t-\lambda_i)^k\prod_k(t-\lambda_k)\right], \qquad (4.7)$$

where,

$$\varphi_i(t) = \frac{f(t)}{\prod_{\substack{k\neq i\\k=1}}(t_i-\lambda_k)^k}$$

We define a function f on the spectrum of $A \in C^{n\times n}$ for the Jordan canonical form, then $f(A) = Qf(J)Q^T$ such that

$$f(J_k) = \begin{bmatrix} f(_i) & f^{/}(\lambda_i)... & & & \frac{f^{(m_i-1)}(\lambda_i)}{(m_i-1)!} \\ & f(\lambda_i) & ... & & \\ & & . \quad . & & \\ & & & . \quad . & \\ & & & . & \\ & & & & f^{/}(\lambda_k) \\ & & & & f(\lambda_k) \end{bmatrix} \qquad (4.8)$$

Furthermore, we substitute $J_k \in C^{m_k\times m_k}$ to have

$$f(J_k) = f(\lambda_k)I + f^{/}(\lambda_k)P_k + ... + \frac{f^{(m_k-1)}(\lambda_k)}{(m_k-1)!}P_k^{m_k-1} \qquad (4.9)$$

It must be noted that Equation (4.6) coincides with that of Equation (4.8) in the long run based on the fact that higher powers of $P_k \to o$.The Lagrange polynomial function being adopted is given by

$$P(\lambda) = f(\lambda_i)\prod_{\substack{j\neq i\\j=1}}^{n}\frac{(\lambda-\lambda_i)}{(\lambda_j-\lambda_i)} \qquad (4.10)$$

Where $f(\lambda) = \sqrt{\lambda}$.

Therefore we obtain the matrix square root for $B = A^T A$ by the Lagrange interpolation formula of Equation (4.10) with $P(t) = 1.2300e-58*t^4$, therefore it follows that $f(B) = p(B) = 1.2300e-58*B^4 = 1.2300e-58*B^4$

$$= (1.0e+08)*\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.0000 & 0.0003 \\ 0 & 0 & 0 & 0.0003 & 0.0238 \\ 0 & 0 & 0 & 0.0237 & 1.8740 \end{pmatrix}.$$

5.0 Conclusion

Methods for solving the least equations were described and in particular the normal equation approach was used to illustrate the theoretical example based on the data collected [8] as a primary source from Lower Atmospheric Studies at Kogi State University, Kogi State, Anyigba. We reported the probability confidence interval for the Gamma density function calculated for the Relative Humidity as well as the interval for the 75%, 95% and 98& confidence interval for the data. We also calculated the Wallis Factor for Probability confidence interval for the Atmospheric Temperature. Finally, using the Lagrange interpolation, we were able to obtain the result for the square root of the symmetric matrix appearing in the normal equation for the least squares problem provided its eigenvalues are not located in the negative real line,a quite significant advantage in Engineering and Scientific practices in numerical weather Computing.

References

[1] Markovsky I., Low Rank Approximation: Algorithms, Implementation, Applications, Springer Verlag, New York, 2012.

[2] Bjorck A; Numerical methods in Scientific Computing, Volume II, SIAM,(2009)

[3] Fowler KR and Kelly CT, Pseudo-transient continuation for Nonsmooth nonlinear equations, SIAM J. Numer.Anal., Vol. 43, No. 4, pp. 1385-1406, 2005.

[4] Meng X, Saunders MA and Mahoney MW.; LSRN: A parallel iterative solver for strongly Over- or Underdetermined systems, SIAM J. Sci.Comput. Vol. 36,No. 2, pp. C95-C118, 2014.

[5] Chung J and Chung M; Optimal regularized inverse matrices for inverse problems, SIAM J. Matrix Anal. Appl. Vol. 38,No. 2, pp. 458-477, 2017.

[6] Martin D and Reichel L; Projected Tikhonov regularization of large scale discrete ill-posed problems, Journal of Scientific Computing, Vol. 56, Issue 3, pp. 471-493, 2013.

[7] Hwang T and Huang P, On New moment estimation of parameters of the Gamma distribution using its characterization. Ann Inst. Staist.Math, Vol. 54, No.4 , 840-847, 2002.

[8] Uwamusi SE; Jacobi Similarity Transformation for SVD and Tikhonov Regularization for Least squares Problem: The Theoretical Foundation. IOSR Journal of Mathematics (IOSR-JM) e-ISSN:2278-5728,P-ISSN: 2319-765X. Vol 12, Issue 5 Ver. V11, PP. 76-85, 2016.