Transactions of the Nigerian Association of Mathematical Physics
Volume 5, (September and November, 2017), pp73 – 78
© Trans. of NAMP

On De-Noising Solution Space to Least Squares Problems

Stephen Ehidiamhen UWAMUSI

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

Abstract

The task of de-noising solution space to least squares problem as a result of huge condition number occurring as a result of unwanted noise present in the data on the right hand side in the linear system is discussed. We filtered out noise from solution set with smallest singular values spaces based on Tikhonov regularization parameter using the numerical tool box of Singular value Decomposition (SVD), and compare results with Cholesky Factorization and preconditioned conjugate gradient method (PCG) for the over determined nonlinear least squares problem which was narrowed down to linear least squares problem. Bounds for the singular values of pseudo- inverse matrix A is constructed using ideas due to Rump.

Keyword: Nonlinear least squares problems, huge condition number, Tikhonov regularization, svd, backward error.

1. Introduction

Least squares problem deals with challenges found in mathematical formulations which are rank deficient largely due to nonlinear model having more free parameters than the model can describe. Quite often the Gauss-Newton iteration comes as a hand tool for approximating solution to nonlinear system of equation [1-3].

However, such resulting system of linear equation is already ill-posed due to noise present in the data. The nonlinear least squares problem is in the form:

$$x = \arg\min_{s_k} \left\| F'(x)s_k + F(x) \right\|_2, s_k \in R^n \quad , \quad (1.1)$$

Where, the current update of x is

$$x^{(k+1)} = x^{(k)} + s \quad (k = 0,1,2,\ldots,.) \quad (1.2)$$

The function $F: D \to R^n, x \in D \subset R^m$ is continuous on D and $m > n$.The vector s_k which is invariant under linear transformation of the independent variable x is computed as a minimiser of equation in the form:

$$s_k = -F'(x^{(k)})^+ F(x^{(k)}),\ (k = 0,1,2,\ldots,) \quad (1.3)$$

Where, it is understood that $F'(x^{(k)})^+$ is the Moore-Penrose inverse $F'(x^{(k)})$, [4].

Practical experience shows that iteration (1.3) does not always have solution for all s since the resulting linear system is singular or nearly singular [5], there is thus need for introduction of Tikhonov regularization parameter as defined in the equation (1.4) :

$$s = -[(\alpha_k^2 I + \left(F'(x^{(k)})\right)^T F'(x^{(k)}))]^{-1} F'(x^{(k)}) F(x^{(k)}),\ (k = 0,1,\ldots,.) \quad (1.4)$$

The α_k is the Levenberg-Marquardt parameter at the current point. Without loss of generality, we hereby make further adoption of notation as $A(x^{(k)})$ representing Jacobian matrix $F'(x^{(k)})$, and $b(x^{(k)}) = F'(x^{(k)})^T F(x^{(k)})$. Equation (1.4) is rewritten in the form:

$$s = -\left(\alpha_k^2 I + (A(x^{(k)})^T A(x^{(k)}))\right)^{-1} b(x^{(k)}) \quad , (k=0,1,2,\ldots,) \quad (1.5)$$

We give theoretical foundation necessary for discussion in the paper below:

The projection matrix $P_A = AA^+$ is Hermittian, P_A is idempotent and, the range space $R(P_A) = R(A)P_A$ is the orthogonal projection onto $R(A)$.We then have that $A^T\left(I - AA^+\right) = A^T - A^T AA^+ = A^T - A^T A^{+^T} A^T = A^T A^{T+} A^T = O$.

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

From analysis due to [6], it is inferred that:

$$A^{+}-A^{T}=\left(A^{T}A\right)^{-1}A^{T}-A^{T}=\left(I-A^{T}A\right)\left(A^{T}A\right)^{-1}A=\left(I-A^{T}A\right)A^{+}.$$

Wherefrom it is obtained that: $\|A\|_2\left\|\left(A^{T}A\right)^{-1}A^{T}\right\|_2=K_2(A),\ \|A\|_2^2\left\|\left(A^{T}A\right)^{-1}\right\|=K_2(A)^2$, so that $\sin\theta=\frac{\|r\|_2}{\|b\|_2}$ is well defined. In that case**,**

$\tan\theta=\varepsilon\left(K_2(A)+K_2(A)^2\|r\|_2\right)$*, where* $\theta\approx small$ **,** $\|r\|_2=\rho_{Ls(x)}$ is best possible**.**

For the perturbed data $\left\|(A+\Delta A)x-(b+\delta b)\right\|$=min!; we can verify the ill-conditioning in the sense of [7] that $\varepsilon=\max\left(\frac{\|\Delta A\|}{\|A\|},\frac{\|\delta b\|}{\|b\|}\right)<\frac{1}{K_2(A)}$, and that $K_2(A)=\frac{\sigma_1}{\sigma_n}$ the condition number of A .The relative error is bounded by

$$\frac{\left\|x-\bar{x}\right\|}{\|x\|}\le\varepsilon\frac{(2K_2(A))}{\cos(\theta)}+\tan(\theta).K_2^2(A)+O(\varepsilon^2), \qquad (1.6)$$

where θ =angle between b and its projection onto $R(A)$.

Numerical analysts often use the LU decomposition or Cholesky Factorization in the Least squares problems. But due to the huge presence of noise in the measured data, solutions obtained by either LU decomposition or Cholesky Factorization sometimes become misleading as a result of high condition number present in the Jacobian matrix. Experiences indicated that Pseudo-inverse of the matrix $A^{T}A$ by the aforementioned methods pushes the high condition number to the right hand side which is common in either the under- or over -determined system. Furthermore, in some cases this leads to cancellation of results due to dependency problems especially in the application of the methods in circular interval methods.

We may resort to the method of QR-Cholesky-**factorization in the form:** $A=QR,\left(Q\in R^{m\times m},R\in R^{n\times n}\right)$**, where Givens rotation matrix is**

$$Q(p,q,\theta)=\begin{pmatrix}1 & \cdots & & & & & 0\\ & \cdot & & & & & \\ & & \cdot & & & & \\ & & & 1 & & & \\ & & & \cos\theta & \sin\theta & \cdots & 0\\ & & & \cdots & \cdots & & \\ & & & 1 & & & \\ & & & & \cdot & & \\ & & & & \cdot & & \\ & & & & \cdot & & \\ \cdots & & & -\sin\theta & \cos\theta & \cdots & 0\\ & & & & 1 & & \\ & & & & & \cdot & \\ & & & & & \cdot & \\ \cdots & & & & & \cdot & \\ 0\cdots & & & & & & 1\end{pmatrix}-$$

- P - Q

The matrix $Q(p,q,\theta)$ **is a rotation matrix that is orthogonal and preserves length. We thus incorporate this Givens** QR **-Cholesky –factorization into** the linear least squares problem formulated in the form**:**

$$\|Ax-b\|_2=\left\|Q\left(Rx-Q^{T}b\right)\right\|_2=\left\|\begin{pmatrix}\bar{R}x-\beta_1\\ -\beta_2 & O\end{pmatrix}\right\|_2;Q^{T}b=\begin{bmatrix}\beta_1\\ \beta_2\end{bmatrix} \qquad \textbf{(1.7)}$$

This has solution

$$x^{*}=\left(\bar{R}\right)^{-1}\beta_1 \qquad \textbf{(1.8)}$$

We now have the case of unexpected occurrence in our discussion particularly, in the un-shifted QR algorithm, which holds that

$$R_k^{T}=Q_{k+1}^{T}R_{k+1},\quad (k=1,2,3,\ldots,.) \qquad (1.9)$$

There is gain to be made in using forward QR method as two successive iterations of R_k^{T} is equivalent to one iteration of basic QR algorithm. To this end, it is set that:

$R_k^T R_k = Q_{k+1}(R_{k+1}R_k)$, (1.10)

which always yield the QR factorization of A .This is further demonstrated with formation of product in reverse order:

$(R_{k+1}R_k)Q_{k+1} = R_{k+1}R_{k+1}^T Q_{k+1}^T Q_{k+1} = R_{k+1}R_{k+1}^T = R_{k+2}^T Q_{k+2}^T Q_{k+2}R_{k+2} = R_{k+2}^T R_{k+2}$.

Where is understood that:

$R_{k+1}^T R_{k+1} = R_k(Q_{k+1}Q_{k+1}^T)R_k^T = R_k R_k^T$, and that $R_{k+2}R_{k+2}^T = R_{k+2}R_{k+1}Q_{k+2} = Q_{k+2}^T(R_k R_k^T)Q_{k+2}$

The Givens QR-factorization method has several applications in Sciences and Engineering practices for example, in the eigenvalues based problems and inversion of matrices occurring in linear system solvers, particularly in the antenna beam formation and systolic matrix arrays resolution.

The remaining section in the paper is arranged as follows: Section 2 discusses the singular values decomposition for problem 1.1, the error perturbation analysis discussed in section 3.In section 4, construction bounds for Singular values of A^+ in algorithmic form following carefully the ideas in [6]. We filtered out noise for the solution set from singular values spaces.

2.0 The Singular values Decomposition for Problem 1.1

We recollect that $\hat{x} \in R^n$ is a solution to nonlinear least squares problem 1.1 whenever $A^T\left(A\hat{x}-b\right)=O$. The existence of solution to such system is heavily dependent on the nature of spectral radius of the system matrix. Based on this fact, we often use the power iteration method to compute dominant eigenvalue and corresponding eigenvector- the fundamental mode of

$Bx = \lambda x$, (2.1)

where the matrix B is the Tikhonov regularized matrix to equation(1.4) with the eigenvalue-eigenvector pair (eigen pair) being the Ritz pair [8].

Motivated by this idea we are led to describing the Singular value decomposition of the matrix $A(x) \in R^{m\times n}, m > n$ as follows. The matrix A **is decomposed in the form**

$A(x) = U\sum V^T$ (2.2)

Where

$\sum = diag(\sigma_1, \sigma_2, ..., \sigma_r, \sigma_{r+1}, ..., \sigma_n)$.

The rank of $A(x)$ is determined as $\hat{\sum} = diag(\sigma_1, \sigma_2, ..., \sigma_r, 0 ..., 0)$, wherefrom,

$A(x) = U\hat{\sum} V^T$ (2.3)

Because of expression given in equation (2.3) we then rewrite equation (1.5) in the form:

$s_\alpha = -\sum_{i=1}^{n}\left[(v_i\sigma_i^T u_i^T)(u_i\sigma_i v_i^T) + \alpha^2 I\right]^{-1}(v_i\sigma_i^T u_i^T)b$ (2.4)

Further simplifications of equation (2.4) would yield that:

$s_\alpha = -\sum_{i=1}^{n}(\sigma_i^2 + \alpha^2)^{-1}(v_i^2\sigma_i u_i^T)b$ **(2.5)**

By further adoption of ideas from existing literature [9-10] leads to the following result:

$s_\alpha = -\sum_{i=1}^{n}\frac{\sigma_i}{\alpha^2 + \sigma_i^2}(u_i^T)bv_i$ (2.6)

Further exposition of equation (2.6) for discussion shows that for the matrix of rank r **, it is that**

$x = \sum_{i=1}^{r}\frac{u_i^T b}{\sigma_i}v_i + \sum_{i=r+1}^{n}\tau_i v_i, \tau \in R$ **(2.7).**

The terms $v_{r+1}, v_{r+2}, \ldots, v_n$ span the kernel, the Null space $N(A)$ of A . We give the solution in the form:

$\bar{x} = x_{Ls} + N(A)$, **(2.8)**

Where, as usual, the term $x_{Ls} = \sum_{j=1}^{r}\frac{u_i^T b}{\sigma_i}v_i$ **, and** $N(A) \to O$ **as** $r \to n$ **.**

The Picard condition is that $|u_i^T b| \to 0$ at a faster rate compared to σ_i .

The contributory factor of regulation parameter α shows that for $\alpha \approx 0$, *the term* $\frac{\sigma_i}{\alpha^2 + \sigma_i^2} \to \frac{1}{\sigma_i}$ and for large values of α , that is for $\alpha \to \infty$,the sum $\sigma_i^2 + \alpha \cong \alpha$ **.**Therefore when the value of $\alpha >> 0$, there will be an unwanted noise in the solution with high condition number. A similar analysis holds for the discrete least squares problems where only the right hand side is perturbed by noise.

We construct the filter function for the Tikhonov regularization with sole aim to dampen out noise from the data:

$$x_\varphi = A^+ \varphi(A) b \qquad \textbf{(2.9)}$$

The term $\varphi(A)$ in equation (2.9), a filter function, is given in the form:

$$\varphi_\alpha(\sigma) = 1 - \frac{1}{1+\left(\frac{\sigma}{\alpha}\right)^2} = \frac{\sigma^2}{\sigma^2+\alpha^2} \qquad \textbf{(2.10)}$$

Practically, the definitive role of the filter function is to remove or dampen all eigenvalues close to zero from the right hand side. This can be illustrated for two different matrices [11-13] with diagonals $\varphi(\sum)$ and $f(\sum)$ whose entries are $\varphi(\delta_j)$ and $f(\delta_j)$. By defining the filtered solution as well as filtered right hand side by the equations

$$x_\varphi = Vf(\sum)U^T b \ , \qquad \textbf{(2.11)}$$

$$Ax_\varphi = U\varphi(\sum)U^T b , \qquad \textbf{(2.12)}$$

the residual is obtained as

$$b - Ax_\varphi = U[I - \varphi(\sum)]U^T b \qquad (2.13)$$

The use of regularization [14] indicated that any noise component in the direction u_j will often be amplified by the factor $\varphi(\delta_j)/\delta_j$ whereas, for small $\delta_{j's}$ the amplification of the noise caused by $\varphi(\delta_j)/\delta_j \to 0$.

One immediately would ask what the gain is achievable in the described methods. As it were, the gain in the discussed methods for the solutions between the use of pseudo inverse and regularized solution differ largely by the quantity

$$x^+ - x_\varphi = \sum_{\delta_{j>0}} \left[\frac{1-\varphi(\delta_j)}{\delta_j}\right] \xi_j v_i \qquad \textbf{(2.14)}$$

The $\xi_j = u_j^T b;\ j = 1,2,\dots,m$ **,** $x^+ = A^+ b = \sum_{\delta_{j>0}} \frac{1}{\delta_j} \xi_j v_j$ **and** $x_\varphi = \sum_{\delta_j>0} \frac{\varphi_j}{\delta_j} \xi_j v_j$ **.**

3.0 Error Perturbation analysis for the least squares problem

Firstly, we start with backward error analysis in order to show how data perturbations affect the solution. In other word, it is an interpretation of approximation result to exact solution of a nearby system

$$Bx = b \qquad (3.1)$$

for the matrix $B = \left(\alpha^2 I + (A(x)^T A(x)\right)$*, the vector* $b = A(x)^T F(x)$.

If we solve system (3.1) correctly see e.g.,[15-17] as well as[10] by Cholesky method for $\bar{B} = R^T R$ where R is upper triangular matrix, the relative error for system (3.1) is at most ε per component such that $\left|B - \bar{B}\right| \le \varepsilon |B|$ and $\left|\bar{b} - b\right| < \varepsilon |b|$. The computed approximate solution $\hat{x}$ to $B^{-1}b$ will be an acceptable solution whenever

$$\left|b - B\hat{x}\right| \le \varepsilon\left(|b| + |B|.\left|\hat{x}\right|\right) \qquad (3.2)$$

Practically, it is that for $\varepsilon > \varepsilon_0$, there holds that

$$\varepsilon_0 = \max_i \frac{\left|b_i - \sum_j B_{ij}\hat{x}_j\right|}{|b_i| + \sum_j \left|B_{ij}\hat{x}\right|},\ 0 \le \varepsilon_0 \le 1 , \qquad (3.3)$$

4.0 Construction bounds for Singular values of A^+

The construction bounds for singular values of A^+ is in the form of algorithm using ideas due to [6].

The Algorithm

Given a matrix $A \in R^{m \times n}$ and $p \in \{1,2,\infty\}$,

1) If $\left\|I - A^T A\right\|_p \le \beta < 1$ and $\left\|A^+ - A^T\right\|_2 \le \frac{\beta}{\sqrt{1-\beta}}$, endif and stop.

Else perform the following operations.

2) For $k \geq 1$ find a matrix $C \in R^{m\times k}$ such that the error estimate $\left\|\left(A^{+} - A^{T}C\right)\right\|_p \leq \frac{\beta\left\|A^{T}C\right\|_p}{1-\beta}; p \in \{1,2,\infty\}$ holds.

3) If $\left\|A^{+}C\right\|_p \leq \frac{\left\|A^{T}C\right\|_p}{1-\beta}$ or $\left\|A^{+}\right\|_p \leq \frac{\left\|A^{T}\right\|_p}{1-\beta};\ for\ p \in \{1,2,\infty\}$ Go to step 6.

4) Compute $r = A^{T}\left(A\hat{x} - b\right)$;

5) Compute $\left\|A^{+}b - \hat{x}\right\|_2 = \left\|A^{+}(A^{T})^{+}r\right\|_2 \leq \left[\frac{\left\|A^{T}\right\|_2}{1-\beta}\right]^2 \left\|r\right\|_2$;

6) Compute $\frac{1}{(1+\beta)^{\frac{1}{2}}} \leq \sigma_1\left(A^{+}\right) \leq \frac{1}{(1-\beta)^{\frac{1}{2}}}$.

endif

end

5.0 NUMERICAL EXPERIMENTS

We illustrate with the aid of linear over determined system since nonlinear over-determined system can always be transformed to linear system.

Consider problem 1

$$A = \begin{pmatrix} 1 & 2 & 0 & 0 \\ 3 & 5 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 3 & 5 \\ 0 & 0 & 1 & 1 \\ -2 & -4 & 3 & 6 \\ -6 & -10 & 9 & 15 \\ -2 & -2 & 3 & 3 \end{pmatrix}, \quad b = \begin{pmatrix} 7 \\ 18 \\ 4 \\ 10 \\ 26 \\ 6 \\ 16 \\ 42 \\ 10 \end{pmatrix}$$

Using Filter function defined in equation (2.8) for the Tikhonov regularization method, we compute the following results as a comparison with those obtained from SVD, the PCG method and Cholesky factorization .We use MATLAB windows 2007 and computed results are displayed in Table **1.**

Table 1 showing results computed by the described methods for Problem 1

Theoretical Solution with tikhonov regularization parameter from Normal Equation Approach by Cholesky method for Problem 1	**SVD Least squares approach method (2.7) for Problem1**	**PCG method Without tikhonov regularization parameter method Problem1**	**The filtered Noise obtained from smallest singular value for Tikhonov method (2.8) for Problem1**
1.0000	**1.0031**	**1.0000**	**0.0162**
2.0000	**1.9998**	**2.0000**	**0.2045**
3.0000	**2.9978**	**3.0000**	**0.1090**
4.0000	**4.0000**	**4.0000**	**0.2566**

As expected the computed solutions given by the use of PCG and traditional Normal equation are the same. Whereas the result computed by the backward stable SVD method increased slightly higher than the results obtained from the use of Normal equation method and that of PCG.

6.0 CONCLUSSION

In the paper, after preliminary exposition for solving nonlinear least squares problems by Gauss-Markov's method,an attention was paid to singular value decomposition for the resulting Overdetermined linear system of equation. We constructed bounds for the singular values of the rectangular matrix using closely the ideas due to [6].Sample numerical illustration was carried out on Overdetermined linear system.

We used the numerical tool box of Cholesky factorization, the singular value decomposition (SVD) applied on Tikhonov regularization method and compared results with preconditioned conjugate gradient (PCG) method for unregularized Tikhonov method. The computed results are quite in agreement compared to the method of SVD for solving overdetermined systems of equation since SVD is numerically backward stable. In addition we used a technique of Tikhonov regularization method to filter out noise associated with smallest singular value space in the solution set of linear least squares problem.

References

[1] Chang XW., Golub GH, and Paige CC; Towards a backward perturbation analysis for data least squares problems, SIAM J.Matrix Anal. Appl., 30(4),1281-1301, (2008).

[2] O'Leary DP.; Near-Optimal parameters for Tikhonov and other regularization methods, SIAM J. Sci. Compt., 33(4),1101-1171,(2001).

[3] Uwamusi SE; Graph Completion inclusion isotone for Interval Least Squares Equation, American Journal of Mathematics and Statistics, 5(1),24-31,(2015).

[4] Ortega JM, and Rheinboldt WC; Iterative solution of nonlinear equations in several variables, SIAM, Philadelphia, (2000).

[5] Uwamusi SE; Regularization of Non-smooth function whose Jacobian is nearly singular or singular, America International Journal of Contemporary Scientific Research-180,ISSN 2349-4425, (2014).Available in www.americanij.com

[6] Rump SM., Verified bounds for least squares problems and underdetermined linear systems, SIMAX (33),130-148,(2012).

[7] Voss H; Regularization of least squares problems, Hamburg University of Technology, Institute of Numerical Simulation, Germany, (2014) .

[8] Zhan X; Extreme eigenvalues of real symmetric matrices with entries in an interval. SIAM Matrix Anal. Appl 27(3), 851-860, (2006).

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

[10] Golub GH., and Van-Loan C.; Matrix Computations. Johns Hopkins University Press, Baltimore, third edition,(1996) .

[11] Lin W.; Lecture Notes in Matrix Computations, Department of Mathematics National Tsing Hua University Hsinchu Taiwan 30043, R.O.C., (2008).

[12] EL Ghaoui L., and Lebret H.; Robust solution to least squares problems with uncertain data, SIAM J. Matrix Anal. Appl 18(4), (1997), PP. 1035-1064.

[13] Vogel C.; Computational methods for inverse problem, SIAM, Philadelphia, (2002).

[14] Burger M.; Inverse problems, Lecture Notes, Winter 2007/08 Instistut fur Numerische und Angewandte Mathematik West Falische Wilhelms-Universitat Munster(2007) .

[15] Rump SM.; Computable backward error bounds for basic algorithms in linear algebra, Nonlinear theory and its applications, IEICE, 6(3), (2016), PP. 1-4.

[16] Skeel DR.; Scaling for Numerical stability in Gaussian elimination. Journal of the Association for Computing Machinery 26(3), 494-526,(1979).

[17] Chang XW., Paige CC., and Titley-Peloquin D.; Stopping criterion for the iterative solution of linear least squares problem. SIAM J. Matrix Anal. Appl 31(2),831-852(2009).