International Journal of Hybrid Information Technology
Vol. 7, No. 6 (2014), pp. 315-324
http://dx.doi.org/10.14257/ijhit.2014.7.6.27

Estimation of Backward Perturbation Bounds for Linear Least Squares Problem

Xinxiu Li

School of Natural Sciences, Nanjing University of Posts and Telecommunications, 210023, China
lixinxiu410@163.com, lixx@njupt.edu.cn

Abstract

Waldén, Karlson, and Sun found an elegant explicit expression of backward error for the linear least squares problem. However, it is difficult to compute this quantity as it involves the minimal singular value of certain matrix. In this paper we present a simple estimation to this bound which can be easily computed especially for large problems. Numerical results demonstrate the validity of the estimation.

Keywords: *linear least squares problem; backward error; backward stable; residual error*

1. Introduction

Consider linear least squares (LS) problem

$$\min_{x\in C^n}\|Ax-b\|_2 \quad \text{with } A\in C^{m\times n}, b\in C^m. \qquad (1)$$

Let x be an approximate solution of (1), define the backward error sets

$$\mathcal{E}=\left\{E : E\in\mathcal{C}^{m\times n}, \|b-(A+E)\hat{x}\|_2=\min_x\|b-(A+E)x\|_2\right\},$$

$$\mathcal{G}=\left\{(E,e) : E\in\mathcal{C}^{m\times n}, e\in\mathcal{C}^m, \|(b+e)-(A+E)\hat{x}\|_2=\min_x\|(b+e)-(A+E)x\|_2\right\}.$$

and the smallest backward perturbation

$$\eta^{(0)}(\hat{x})=\min_{E\in\mathcal{E}}\|E\|_F, \qquad \eta(\hat{x})=\min_{(E,e)\in\mathcal{G}}\|(E,\omega e)\|_F. \qquad (2)$$

where ω is a real number, which balances the norms of the backward perturbations in A and b.The approximate solution x can be regarded as the exact solution to a perturbed problem of (1). A small $\eta(\hat{x})$ means that x is the exact solution to a nearby problem of (1), and x can be regarded as the computed result produced by a backward stable algorithm. Consequently, $\eta(\hat{x})$ can be used to test the stability of numerical algorithms. It is also used to monitor the convergence of iterative solution methods and to design reliable stopping criteria for these methods [13]. If the approximate solution x is an iterate result from any chosen iterative method, the iteration can be stopped and x is accepted as a valid computed solution when the backward error $\eta(\hat{x})$ is smaller than a chosen tolerance.

Backward error analysis of the linear least squares problem is a thirty years old problem suggested by Stewart and Wilkinson. Waldén, Karlson, and Sun[1] provided the following explicit expressions.

ISSN: 1738-9968 IJHIT

Theorem 1.1.[1] Let $x \neq 0$ and ω be a real number, then

$$\eta^{(0)}(x) = \begin{cases} \dfrac{\|\hat{r}\|_2}{\|x\|_2} & \lambda_* \geq 0 \\ \left[\left(\dfrac{\|\hat{r}\|_2}{\|x\|_2}\right)^2 + \lambda_*\right]^{1/2} & \lambda_* < 0 \end{cases}. \tag{3}$$

$$\eta(\hat{x}) = \begin{cases} \sqrt{\tau}\dfrac{\|\hat{r}\|_2}{\|\hat{x}\|_2}, & \lambda_\omega \geq 0 \\ \left[\left(\dfrac{\|\hat{r}\|_2}{\|\hat{x}\|_2}\right)^2 \tau + \lambda_\omega\right]^{1/2}, & \lambda_\omega < 0 \end{cases} \tag{4}$$

where

$$\hat{r} = b - A\hat{x}, \ \ \lambda_* = \lambda_{min}\left(AA^H - \tfrac{\hat{r}\hat{r}^H}{\|\hat{x}\|_2^2}\right),$$

$$\lambda_\omega = \lambda_{min}\left(AA^H - \tau\frac{\hat{r}\hat{r}^H}{\|\hat{x}\|_2^2}\right), \tau = \frac{\omega^2\|\hat{x}\|_2^2}{1+\omega^2\|\hat{x}\|_2^2}. \tag{5}$$

Theoretically, these results have solved the backward error analysis for the LS problem, so they are valuable results. However, from the view point of practice, the main aim of backward perturbation theory is to analyze the accuracy of computed results, thus the backward error bound should be computed efficiently, but (3) and (4) don't have the property, which is disadvantage to practical computation. N. J. Higham pointed out that perhaps it is not numerically stable to compute (3) and (4) directly, and recommended the following formula that have been derived from the pre-publication manuscript [2]

$$\eta^{(0)}(\hat{x}) = \min\left\{\frac{\|\hat{r}\|_2}{\|\hat{x}\|_2}, \sigma_{min}([A,\ R_0])\right\}, \quad R_0 = \frac{\|\hat{r}\|_2}{\|\hat{x}\|_2}(I_m - \hat{r}\hat{r}^\dagger),$$

$$\eta(\hat{x}) = \min\left\{\sqrt{\tau}\frac{\|\hat{r}\|_2}{\|\hat{x}\|_2}, \sigma_{min}([A,\ R])\right\}, \quad R = \sqrt{\tau}\frac{\|\hat{r}\|_2}{\|\hat{x}\|_2}(I_m - \hat{r}\hat{r}^\dagger),$$

where τ is defined by (5). The two formulas can be computed with numerically stable method, but it requires to compute SVD of an $m \times (n+m)$ matrix, which can be probably expensive for large problems. So how to estimate $\eta^{(0)}(\hat{x})$ and $\eta(\hat{x})$ effectively is worth further researching. Many authors including Waldén, Karlson, and Sun have paid much attention on the problem to derive explicit approximation or find upper and lower bounds for $\eta^{(0)}(\hat{x})$ and $\eta(\hat{x})$ [3-13,16,17]. The estimates $\mu^{(0)}(\hat{x})$ and $\mu(\hat{x})$ derived by Karlson and Waldén[6] in particular have been studied by several authors. The estimates can be written as

$$\mu^{(0)}(\hat{x}) = \left\|\left(\|\hat{x}\|_2^2 A^H A + \|\hat{r}\|_2^2 I\right)^{-1/2} A^H \hat{r}\right\|_2,$$

$$\mu(\hat{x}) = \left\|\left(\|\hat{x}\|_2^2 A^H A + \tau\|\hat{r}\|_2^2 I\right)^{-1/2} A^H \hat{r}\right\|_2.$$

The quantity $\mu^{(0)}(\hat{x})$ and $\mu(\hat{x})$ can be computed more cheaply than $\eta^{(0)}(\hat{x})$ and $\eta(\hat{x})$[9-11], and they can be estimated very efficiently in the iterative method LSQR[13].

Karlson and Waldén [7] showed that

$$\frac{\mu^{(0)}(\hat{x})}{\eta^{(0)}(\hat{x})} \leq \frac{2+\sqrt{2}}{2} \approx 1.707.$$

When A is a full column rank matrix, Gu [5] proved that the estimation $\mu^{(0)}(\hat{x})$ differs from $\eta^{(0)}(\hat{x})$ by a factor $\frac{\sqrt{5}+1}{2}$, that is

$$\frac{\mu^{(0)}(\hat{x})}{\eta^{(0)}(\hat{x})} \leq \frac{\sqrt{5}+1}{2} \approx 1.618.$$

Gratton[12] gave tight bounds on $\eta^{(0)}(\hat{x})$ which involve the estimate $\mu^{(0)}(\hat{x})$

$$\mu^{(0)}(\hat{x}) \leq \eta^{(0)}(\hat{x}) \leq \sqrt{2}\mu^{(0)}(\hat{x}).$$

Grcar[8] proved that $\mu^{(0)}(\hat{x})$ asymptotically equals $\eta^{(0)}(\hat{x})$ in the sense that

$$\lim_{\hat{x}\to x_*} \frac{\mu^{(0)}(\hat{x})}{\eta^{(0)}(\hat{x})} = 1,$$

where x_* is an exact solution of the LS problem (1). The above results showed that $\mu^{(0)}(\hat{x})$ is an excellent estimate of $\eta^{(0)}(\hat{x})$. Grcar[9] performed numerical experiments to examine the validity of $\mu^{(0)}(\hat{x})$ as an acceptable estimate for $\eta^{(0)}(\hat{x})$ in practice.

The purpose of this paper is to give a simple and cheaply computable estimate for $\eta^{(0)}(\hat{x})$ and $\eta(\hat{x})$.

The paper is organized as follows. In Section 2, we propose the estimation method for $\eta^{(0)}(\hat{x})$ and $\eta(\hat{x})$. Numerical examples are given to demonstrate the validity of the estimation in Section 3. Finally, some comments on the results are made in Section 4.

2. Estimation of $\eta^{(0)}(\hat{x})$ and $\eta(\hat{x})$.

Firstly, we analyze the estimate method for $\eta^{(0)}(\hat{x})$. Notice the fol lowing facts:

Fact 1. $\hat{r} = b - Ax$ is the residual error of x to the linear system $Ax=b$, and the residual error of x as the approximate solution to (1) is

$$r_0 = P_A(b - Ax), \tag{6}$$

where $P_A = AA^{\dagger}$.

The reason is as follows.

Let x be the solution to (1), then

$$x = A^{\dagger}b + (I_n - A^{\dagger}A)z, \quad z \in C^n . \tag{7}$$

Let $r=b-Ax$, then $\hat{r}$ is an approximation of r.

Therefore, the residual error of x as the approximate solution to (1) is

$$r_0 = \hat{r} - r = A(x - x).$$

Based on

$$x = A^{\dagger}Ax + (I_n - A^{\dagger}A)x$$

and (7), we have

$$x - x = A^{\dagger}(b - Ax) + (I_n - A^{\dagger}A)(z - x), \qquad z \in C^n \tag{8}$$

Consequently,

$$r_0 = A(x - x) = AA^{\dagger}\hat{r} = P_A\hat{r}.$$

Fact 2. Let *S* be the set of the solutions to (1), then

$$\|r_0\|_2 / \|A\|_2 \le \min_{x\in S}\|x - x\|_2 \le \|A^{\dagger}\|_2 \|r_0\|_2 . \tag{9}$$

This fact can be demonstrated as follows.
From (8), we have

$$\|x - x\|_2^2 = \|A^{\dagger}(b - Ax)\|_2^2 + \|(I_n - A^{\dagger}A)(z - x)\|_2^2 ,$$

therefore

$$\min_{x\in S}\|x - x\|_2 = \|A^{\dagger}P_A\hat{r}\|_2 = \|A^{\dagger}r_0\|_2 ,\ \|r_0\|_2 = \|A(x - x)\|_2 \le \|A\|_2 \|x - x\|_2 ,$$

hence (9) is obtained.

According to the application background of LS problem, we assume $m \ge n$, and *rank(A)=p*. Let

$$A = U\begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix}V^H$$

be the singular value decomposition of *A*,where $U \in C^{m\times m}$ and $V \in C^{n\times n}$ are unitary matrices, and $\Sigma_1 \in R^{p\times p}$ is a positive diagonal matrix. Then

$$U^H\left(AA^H - \hat{r}\hat{r}^H / \|x\|_2^2\right)U = \begin{pmatrix} \Sigma_1^2 & 0 \\ 0 & 0 \end{pmatrix} - \tilde{r}\tilde{r}^H / \|x\|_2^2 ,$$

where

$$\tilde{r} = U^H\hat{r} = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix}, \tag{10}$$

$r_1 \in C^p, r_2 \in C^{m-p}$. Let $\alpha = \|x\|_2^{-1}$, then $AA^H - \alpha^2\hat{r}\hat{r}^H$ is unitary similar to

$$G \equiv \begin{pmatrix} \Sigma_1^2 - \alpha^2 r_1 r_1^H & -\alpha^2 r_1 r_2^H \\ -\alpha^2 r_2 r_1^H & -\alpha^2 r_2 r_2^H \end{pmatrix}. \tag{11}$$

Therefore $AA^H - \alpha^2\hat{r}\hat{r}^H$ and *G* have the same eigenvalues.

Fact 3. The sufficient and necessary condition of $\lambda_* \ge 0$ is $r_2 = 0$ and $\alpha\|\Sigma_1^{-1}r_1\|_2 \le 1.$

Proof. Notice that the sufficient and necessary condition of $\lambda_* \ge 0$ is $AA^H - \alpha^2\hat{r}\hat{r}^H$ be a positive semi-definite matrix. So if $\lambda_* \ge 0$, we have $-\alpha^2 r_2 r_2^H$, therefore $r_2 = 0$, and $\Sigma_1^2 - \alpha^2 r_1 r_1^H$ is positive semi-definite. Hence $\alpha\|\Sigma_1^{-1}r_1\|_2 \le 1.$

Conversely, if $r_2 = 0, \alpha\|\Sigma_1^{-1}r_1\|_2 \le 1$, then *G* is a positive semi-definite matrix. Therefore $\lambda_* \ge 0$.

Fact 4. $r_2 = 0$ if and only if *Ax=b* is consistent.

This fact can be demonstrated as follows.
If we split $U = [U_1, U_2]$, with $U_1 \in C^{m\times p}$, where *p=rank(A)*, then

$R(U_1) = R(A), R(U_1) = R^{\perp}(A)$.

From (10), we have

$r_2 = U_2^H (b - Ax) = U_2^H b.$

So

$r_2 = 0 \Leftrightarrow U_2^H b = 0 \Leftrightarrow b \in R(A) \Leftrightarrow Ax{=}b$ is consistent.

Since $\left\| A^\dagger P_A \hat{r} \right\|_2 = \left\| \Sigma_1^{-1} r_1 \right\|_2$, hence $\lambda_* \geq 0 \Leftrightarrow Ax{=}b$ is consistent and $\alpha \left\| A^\dagger P_A \hat{r} \right\|_2 \leq 1.$

We note that the method of least squares is mainly used in the case where $Ax{=}b$ is inconsistent, so the general case of (3) is

$$\eta^{(0)}(x) = \left[\left(\left\| \hat{r} \right\|_2 / \left\| x \right\|_2 \right)^2 + \lambda_* \right]^{1/2} \tag{12}$$

From (6) and (7),we know that $\|A^\dagger r_0\|_2$ is the distance from $\hat{x}$ to the set of the solution to (1), so generally, we have

$$\|r_0\|_2 << 1$$

On the other hand, from (10), we have

$$\|r_2\|_2 = \|(I - AA^\dagger)\hat{r}\|_2 = \|(I - AA^\dagger)b\|_2 = \min_x \|Ax - b\|_2$$

Obviously,$\|r_2\|_2$ evaluated the inconsistent degree of $Ax = b$,
so$\|r_2\|_2$ is not very small in many cases.

Theorem2.1 Let $A, b, \hat{r}, r_0, r_1, r_2, U, \Sigma_1, \alpha, \lambda_*$ be defined as above. Then

$$\eta^{(0)}(\hat{x}) \approx \frac{\|r_0\|_2}{\|\hat{x}\|_2} =: \eta_0 \tag{13}$$

Proof. Let $Q_1 r_2 = \left\| r_2 \right\|_2 e_1$ be the QR factorization of r_2, where Q_1 is a unitary matrix, and e_1 denotes the first column of I_{m-p}, we have

$$\begin{pmatrix} I_p & 0 \\ 0 & Q_1 \end{pmatrix} U^H \left(AA^H - \alpha^2 \hat{r}\hat{r}^H \right) U \begin{pmatrix} I_p & 0 \\ 0 & Q_1^H \end{pmatrix} = \begin{pmatrix} \Sigma_1^2 - \alpha^2 r_1 r_1^H & -\alpha^2 \left\| r_2 \right\|_2 r_1 e_1^T \\ -\alpha^2 \left\| r_2 \right\|_2 e_1 r_1^T & -\alpha^2 \left\| r_2 \right\|_2^2 e_1 e_1^T \end{pmatrix}.$$

For a Hermite matrix $M = (m_{ij})_{n \times n}$, it is well known that

$\lambda_{\min}(M) \leq \min_{1 \leq i \leq n} \{ m_{ii} \}.$

Hence

$$\lambda_* \leq -\alpha^2 \left\| r_2 \right\|_2^2 \tag{14}$$

Substituting (14) into (12), we get

$$\eta^{(0)}(x) \leq \left[\alpha^2 \left\| \hat{r} \right\|_2^2 - \alpha^2 \left\| r_2 \right\|_2^2 \right]^{1/2} = \alpha \left\| r_1 \right\|_2 = \left\| r_1 \right\|_2 / \left\| x \right\|_2 . \tag{15}$$

On the other hand, if $\hat{x}$ is computed by a suitable algorithm (for example, the QR factorization or the singular value decomposition method), then, generally,$\|r_1\|_2 << 1$.

From Weyl theorem [18] we have

$$|\lambda_* - (-\alpha^2 \|r_2\|_2^2)| \leq \alpha^2 \|r_2\|_2 \|r_1\|_2$$

and

$$\begin{aligned} \lambda_* &\geq -\alpha^2\|r_2\|_2^2 - \alpha^2\|r_2\|_2\|r_1\|_2 \\ &\approx -\alpha^2\|r_2\|_2^2 \quad (\text{ because } \ \|r_1\|_2 << \|r_2\|_2) \end{aligned}$$

Hence

$$\eta^{(0)}(\hat{x}) \gtrsim \left(\alpha^2\,\|\hat{r}\|_2^2 - \alpha^2\,\|r_2\|_2^2\right)^{1/2} = \alpha\,\|r_1\|_2 = \frac{\|r_1\|_2}{\|\hat{x}\|_2}. \qquad (16)$$

From (10) $r_1 = U_1^H \hat{r}$, hence

$$\|r_1\|_2 = \left\|U_1 U_1^H \hat{r}\right\|_2 = \|r_0\|_2 \qquad (17)$$

According to (15) - (17), the approximate estimation (13) is obtained.

From the above analysis, we can conclude that if we substitute α with $\sqrt{\tau}/\|\hat{x}\|_2$, the same analysis of $\eta^{(0)}(\hat{x})$ is suitable to $\eta(\hat{x})$, so we get the formula of estimating $\eta(\hat{x})$ as follows:

$$\eta(\hat{x}) \approx \sqrt{\tau}\frac{\|r_0\|_2}{\|\hat{x}\|_2} =: \hat{\eta},$$

where τ is given by (5).

3. Numerical Examples

In Section 2 we have derived an estimate for the backward error $\eta^{(0)}(x)$. In this section, we present two numerical examples to demonstrate the validity of our estimation for large problems.

Example 3.1. The coefficient matrix A is"well1850.mtx", a 1850×712 matrix from the Matrix Market [7] with 8758 non-zeros entries, $\kappa_2(A) = 1.1 \times 10^2$, the right hand side b is"well1850-rhs1.mtx".

Example 3.2. The coefficient matrix A is" illc1033.mtx", a 1033×320 matrix from the Matrix Market [7] with 4732 non-zeros entries, $\kappa_2(A) = 1.9 \times 10^4$, the right hand side b is" illc1033-rhs1.mtx".

Applying the iterative method LSQR proposed in [6] to the least square problems (1) generates a sequence of approximate solutions { x_k }.

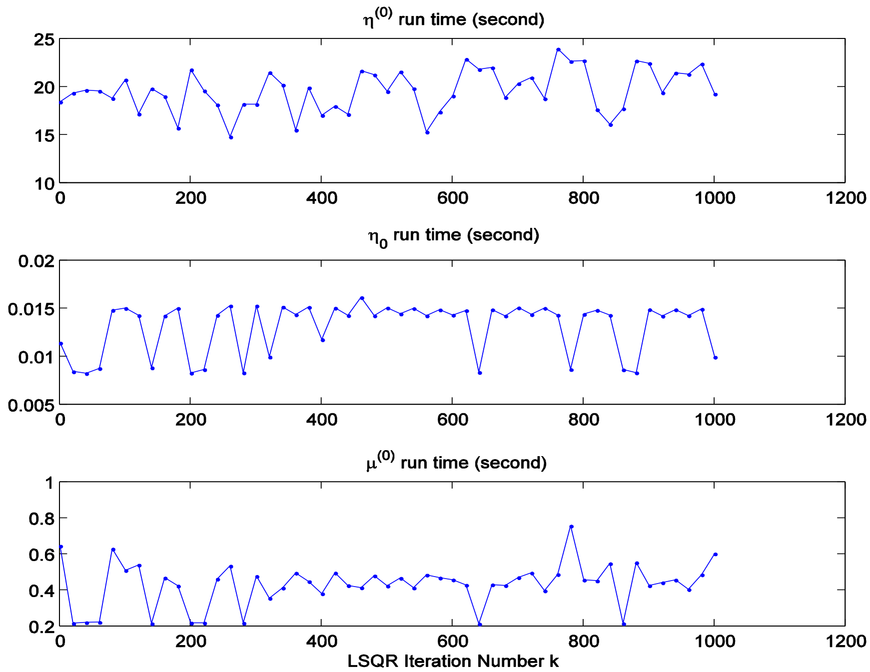


Figure 1. Absolute Errors of η_0 and $\mu^{(0)}(x_k)$ for LSQR Iterate x_k in Example 3.1.

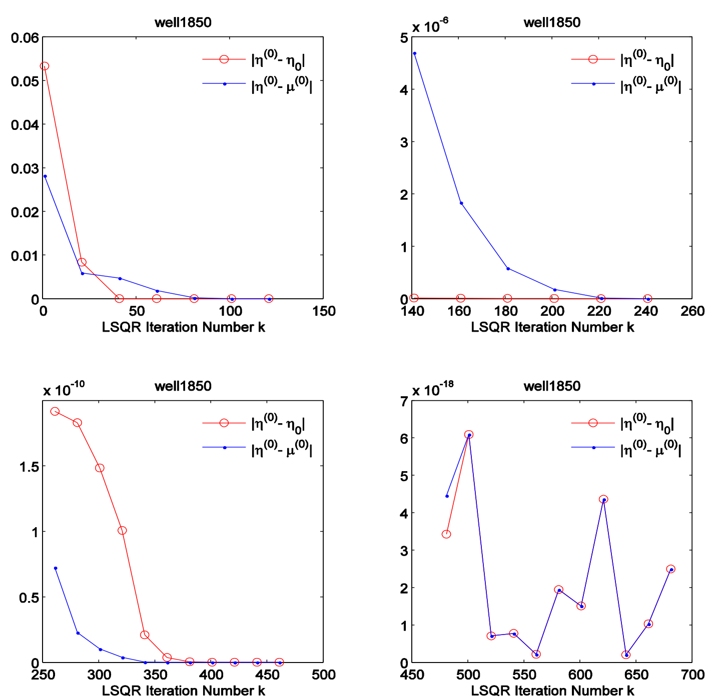


Figure 2. Run Time in Computing $\eta^{(0)}(x_k)$ and their Estimations in Example3.1

International Journal of Hybrid Information Technology
Vol. 7, No. 6 (2014), pp. 315-324
http://dx.doi.org/10.14257/ijhit.2014.7.6.27

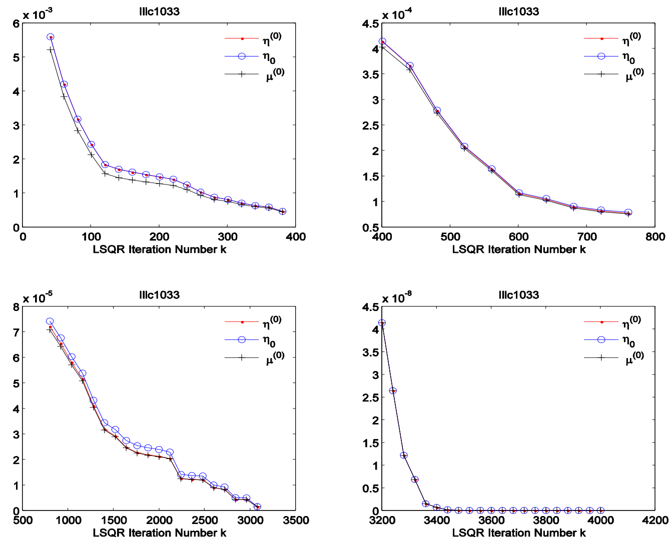


Figure 3. The Comparison between $\eta^{(0)}(x_k)$ and η_0, $\mu^{(0)}(x_k)$ for LSQR Iterate x_k in Example 3.2

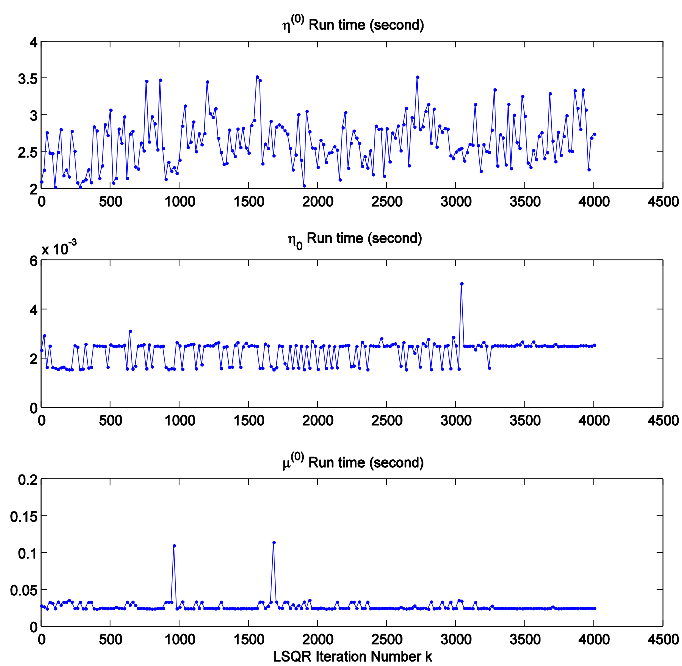


Figure 4. Run Time in Computing $\eta^{(0)}(x_k)$ and Therir Estimations in Example3.2.

The estimation η_0 are compared with $\eta^{(0)}(x_k)$ and $\mu^{(0)}(x_k)$ in Figure 1 and Figure 3, which show that for every purported solution of LS (1), η_0 is a respectively good and efficient

ISSN: 1738-9968 IJHIT

estimate of $\eta^{(0)}(x_k)$. Only when x_k is a good approximation to the exact solution of (1), the asymptotic property of $\mu^{(0)}(x_k)$ is displayed. Running time in computing $\eta^{(0)}(x_k)$ and $\mu^{(0)}(x_k)$, η_0 are given in Figure 2 and Figure 4, which show that computational complexity of η_0 is much smaller than $\eta^{(0)}(x)$ and $\mu^{(0)}(x_k)$. Therefore, η_0 is a practical and cheaply computable estimate of the backward error for the LS problem.

4. Conclusion

In this paper, we propose an efficient estimation for the backward error of least square problem. Numerical experiments show that in terms of the calculation accuracy and computational complexity, our result is good.

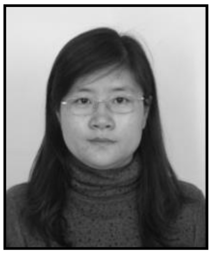
Acknowledgements

This work is supported by the Natural Science Foundation of China under Grant No. 11201240.

References

[1] B. Waldén, R. Karlson, and J. G. Sun, “Optimal backward perturbation bounds for the least squares problem”, Numer. Lin. Alg. Appl., vol. 2, no. 3, **(1995)**, pp. 271-286.
[2] N. J. Higham, “Accuracy and stability of numerical algorithms”, SIAM, Philadelphia, PA, USA, **(1996)**.
[3] J. G. Sun, “Optimal backward perturbation bounds for linear systems and linear least squares problems”, UMINF 96.15, Umå Unibersity.
[4] M. Gu, “Backward perturbation bounds for linear least squares problems”, SIAM J. Mat. Anal. Appl. vol. 20 **(1999)**, pp. 363 - 372.
[5] A. J. Cox and N. J. Higham, “Backward error bounds for constrained least squares problem”, BIT, vol. 39, no. 2, **(1999)**, pp. 210 -227.
[6] R. Karlson and B. Waldén, “Estimation of optimal backward pertutbation bounds for linear least squares problem”, BIT, vol. 37, **(1997)**, pp. 862-869.
[7] J. F. Grcar, “Optimal sensitivity analysis of linear least squares”, Technical Report LBNL-52434, Lawrence Berkeley National Laboratory, **(2003)**.
[8] J. F. Grcar, M. A. Saunders, and Z. Su, “Estimates of optimal backward perturbations for linear least squares problems”, Technical Report SOL, **(2007)**.
[9] S. Zheng, “Computational methods for least squares problems and clinical trials”, PhD thesis, Stanford University, vol. 2, no. 5, **(2005)**.
[10] S. Zheng, “Bounds for an estimate of the optimal backward error for linear least squares problems”, International Journal of Computer Mathematics, vol. 88, **(2011)**, pp. 2743-2751.
[11] S. Gratton, P. Jiranek and T. P. David, “Simple backward error bounds for linear least squares problems”, Lin. Alg. Appl., vol. 439, **(2013)**, pp. 78-89.
[12] S. Gratton, P. Jiranek and T. P. David, “On the accuracy of the Karlson-Waldén's estimate of the backward for linear least squares problem”, Cerfacs technical report TR/PA/11/19, **(2011)**.
[13] P. Jiranek and T. P. David, “Estimating the backward error in LSQR”, SIAM J. Matrix Anal. Appl . vol. 31, no. 4, **(2010)**, pp. 2055-2074.
[14] C. C. Paige and M. A. Saunders, “ALGORITHM 583; LSQR: Sparse linear equations and least-squares problems”, ACM Trans. Math. Software, vol. 8, no. 2, **(1982)**, pp. 195-209.
[15] R. F. Boisvert, R. Pozo, K. Remington, R. Barrett and J. Dongarra, “The Matrix Market: a web repository for test matrix data”, http://math.nist.gov/MatrixMarket/.
[16] G. W. Stewart, “An inverse perturbation theorem for the linear least squares problem”, SIGNUM Newsletter, vol. 10, **(1975)**, pp. 39-40.
[17] G. W. Stewart, “On the perturbations of pseudo-inverses, projections, and linear least squares problems”, SIAM Review, vol. 19, no. 4, **(1977)**, pp. 634-662.
[18] R. Horn and C. Johnson, ”Matrix analysis“, New York: Cambridge University Press, **(1985)**.

Author



Xinxiu Li, she is a lecturer of School of Natural Sciences, Nanjing University of Posts and Telecommunications, Her main research fields include numerical computation, numerical analysis and signal processing.