J. KSIAM **Vol.3**, No.1, 31-41, 1999

Blockwise analysis for solving linear systems of equations

Alicja Smoktunowicz

Abstract

We investigate some techniques of iterative refinement of solutions of a non-singular system $Ax = b$ with A partitioned into blocks using only single precision arithmetic.

We prove that iterative refinement improves a blockwise measure of backward stability. Some applications of the results for the least squares problem (LS) will be also considered.

Introduction In this paper we present various kinds of iterative refinement techniques for the solution of linear systems of the form

$$Ax = b, \tag{1}$$

where A is an $n \times n$ nonsingular matrix and has special block structure. We assume that the matrix A is partitioned into $s \times s$ blocks

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} & \dots & A_{1,s} \\ A_{2,1} & A_{2,2} & \dots & A_{2,s} \\ \dots & \dots & \dots & \dots \\ A_{s,1} & A_{s,2} & \dots & A_{s,s} \end{bmatrix} \tag{2}$$

where $A_{i,j} \in \mathbb{R}^{\mathrm{n_i x n_j}}$ is reffered to as the (i, j) block of A, $\{n_1, \dots, n_s\}$ is a given set of positive integers, $n_1 + \dots + n_s = n$. The vector x is partitioned conformally: $x = [{x_1}^T, \dots, {x_s}^T]^T$ where $x_i(n_i \times 1)$ and $\mu(x) = [\| x_1 \|, \dots, \| x_s \|]^T$.
Without loss of generality we assume that we consider only the spectral matrix norm and the second vector norm (length of x).

We would like to obtain algorithms that produce solutions y accurate to full machine precision, i.e. y is a solution of a slightly perturbed system $(A + E)y = b$ where $\| E_{i,j} \| \le \epsilon \| A_{i,j} \|$ and ϵ is small. We call such algorithms **blockwise backward stable**. Such algorithms are attractive because in some numerical applications it is important that the perturbed matrix $A + E$ has the same structure as A: $A_{i,j} = 0$ implies that $E_{i,j} = 0$.

Key Words :Iterative refinement, rounding error analysis, condition number, blockwise error bounds, least squares problem.
AMS subject classification: Primary: 65F05, 65G05.

We extend existing definitions of normwise and componentwise backward error to block matrices by introducing a *matricial norm* of A [18], [9]:

$$\mu(A) = \begin{bmatrix} \| A_{1,1} \| & \| A_{1,2} \| & \dots & \| A_{1,s} \| \\ \| A_{2,1} \| & \| A_{2,2} \| & \dots & \| A_{2,s} \| \\ \dots & \dots & \dots & \dots \\ \| A_{s,1} \| & \| A_{s,2} \| & \dots & \| A_{s,s} \| \end{bmatrix} \tag{3}$$

where $\| B \| = \| B \|_2$ denotes the spectral norm.

Some important cases of matricial norms are: $\mu(A) = | A |$ for $s = n$ and $\mu(A) = \| A \|_2$ for $s = 1$. The matrix $| A |$ is the matrix whose elements are $| a_{i,j} |$ and we write $| A | \leq | B |$ to mean that inequalities between matrices hold componentwise.

It is obvious that componentwise backward stability (for $s = n$) implies blockwise backward stability and that blockwise backward stability yields to normwise backward stability (for $s = 1$).

We investigate some techniques of iterative refinement of solutions of a nonsingular system $Ax = b$ with A partitioned into blocks using only single precision arithmetic. Our numerical analysis is similar in spirit to that of N.J.Higham [14], [15], and R.Skeel [22].

1 Linear least squares problem (LS)

We consider the solution of the linear least squares problem

$$\min_x \| b - Ax \|, \tag{4}$$

where $A(m \times n)$ and $m \geq n = rank(A)$.

The solution x of (4) is the solution of the normal equation system

$$A^T A x = A^T b. \tag{5}$$

If $r = b - Ax$ then r and x satisfy the augmented system $Mz = f$ where

$$M = \begin{bmatrix} I_m & A \\ A^T & 0 \end{bmatrix} \tag{6}$$

and $z = [r^T, x^T]^T$, $f = [b^T, 0^T]^T$. Here I_m denotes the identity matrix of order m.

The matrix M is nonsingular and symmetric. It is interesting that the inverse of the augmented matrix M can be expressed in the terms of the pseudoinverse matrix of A,

$$A^+ = (A^T A)^{-1} A^T.$$

We have (see [1], [6], [7]):

$$M^{-1} = \begin{bmatrix} P & (A^{+})^{T} \\ A^{+} & -(A^{T}A)^{-1}, \end{bmatrix} \tag{7}$$

where $P = I_m - AA^{+}$.

If $s = n_1 + n_2$, $n_1 = m$ and $n_2 = n$ then we get

$$\mu(M) = \begin{bmatrix} 1 & \sigma_1 \\ \sigma_1 & 0 \end{bmatrix} \tag{8}$$

and

$$\mu(M^{-1}) = \begin{bmatrix} 1 & \frac{1}{\sigma_n} \\ \frac{1}{\sigma_n} & \frac{1}{\sigma_n{}^2} \end{bmatrix} \tag{9}$$

σ_1, σ_n being, respectively, the biggest and the smallest singular values of A.

We see that we can study the property of the algorithms for solving LS problem using the general blockwise approach.

2 Blockwise perturbation analysis

In this section we derive perturbation results and condition numbers in a blockwise sense. We extend the Bauer-Skeel analysis [3], [22] to a linear system of equation (1) with A partitioned into blocks.

We review of the main facts on the matricial norms; see [18] and [9].

Theorem 2.1 *Let μ be a matricial norm on $\in \mathbb{R}^{n\times n}$. For matrices A and B partitioned as in (2), (3) we have*

(1) $\mu(cA) = \mid c \mid \mu(A)$ *for* $c \in \mathbb{R}$,

(2) $\mu(A+B) \leq \mu(A) + \mu(B)$,

(3) $\mu(AB) \leq \mu(A)\mu(B)$,

(4) $\mu(A) \neq 0$ *if* $A \neq 0$,

(5) $\mu(x+y) \leq \mu(x) + \mu(y)$ *for* $x, y \in \mathbb{R}^n$,

(6) $\mu(A\,x) \leq \mu(A)\,\mu(x)$,

(7) $\rho(A) \leq \rho(\mu(A))$,

(8) $\parallel A \parallel \leq \parallel \mu(A) \parallel$.

Here $\rho(A) = \max\{\lambda : \lambda \in spect(A)\}$ *denotes the spectral radius of* A.

The property (7) is a generalization of the Perron-Frobenius inequality and was first proved by Ostrowski [18]; see also [9].

The property (8) is an immediate consequence of $\| A \|^2 = \rho(A^T A)$.

We can now state the analogues of the theorems proved by Skeel [22] for the componentwise case ($s = n$).

How sensitive is the solution α of $Ax = b$ to perturbations in A ?

Theorem 2.2 *Let $A \in \mathrm{I\!R}^{n\times n}$ be nonsingular, $A\alpha = b$ and $(A+E)y = b$, where $\mu(E) \le \epsilon\mu(A)$. Assume that $\epsilon\, cond_\mu(A) < 1$ where*

$$C = \mu(A^{-1})\mu(A)$$

and

$$cond_\mu(A) = \| C \|$$

is called a blockwise condition number of A. Then

$$\mu(y-\alpha) \le \epsilon\ (I - \epsilon C)^{-1}\ C\ \mu(\alpha) \tag{10}$$

and

$$\frac{\| y-\alpha \|}{\| \alpha \|} \le \epsilon \frac{cond_\mu(A,\alpha)}{1 - \epsilon\, cond_\mu(A)}, \tag{11}$$

where

$$cond_\mu(A,\alpha) = \frac{\| \mu(A^{-1})\mu(A)\mu(\alpha) \|}{\| \alpha \|}.$$

Proof. Since $\alpha = A^{-1}b$ and $y - \alpha = -(I + A^{-1}E)^{-1}\, A^{-1}E\alpha$, we have

$$y - \alpha = -(A^{-1}E - (A^{-1}E)^2 + \ldots)\ \alpha.$$

Taking norms we get $\mu(y-\alpha) \le (\mu(A^{-1}E) + \mu((A^{-1}E)^2) + \ldots)\ \mu(\alpha)$.
Since $\mu(A^{-1}E) \le \epsilon C$ hence $\mu(y-\alpha) \le (\epsilon C + \epsilon^2 C^2 + \ldots)\ \mu(\alpha)$ which leads to the inequalities (10) and (11).

Theorem 2.3 *We have the following inequalities:*

(i) $cond_\mu(A) \ge 1$,

(ii) $cond_\mu(A,\alpha) \le cond_\mu(A)$,

(iii) $cond_\mu(A) \le s^2\, cond(A)$ *where* $cond(A) = \| A^{-1} \| \| A \|$ *denotes the normwise condition number of A.*

(iv) $\| |A^{-1}|\, |A| \| \le s^{\frac{1}{2}}\ cond_\mu(A)$ *where* $\| |A^{-1}|\, |A| \|$ *denotes the componentwise Bauer-Skeel condition number of A.*

Proof. The proof of (i) is straightforward. We have $I = A^{-1}A$. Thus $\mu(I) \leq \mu(A^{-1})\mu(A)$ and by Theorem 1.1 (8) we obtain that $1 \leq cond_{\mu}(A)$.

The proof of (ii) is a consequence of the fact that the spectral norm is consistent: $\| Cx \| \leq \| C \| \| x \|$.

In order to prove (iii) and (iv) we use the following inequalities:

$$\| \mu(A) \| \leq s \max_{i,j} \| A_{i,j} \| \leq s \| A \|$$

and

$$\| \mu(| A |) \| \leq s^{\frac{1}{2}} \| \mu(A) \| .$$

Theorem 2.4 (Rigal and Gaches) *The blockwise relative error*

$$\eta_{\mu}(y) = \min\{\epsilon : (A+E)y = b,\ \mu(E) \leq \epsilon\mu(A)\} \tag{12}$$

is given by

$$\eta_{\mu}(y) = \max_{i} \frac{\| r_i \|}{g_i},$$

where $r = b - Ay$ and $g = \mu(A)\mu(y)$ are partitioned as $r = [{r_1}^T, \ldots, {r_s}^T]^T$, $g = [{g_1}^T, \ldots, {g_s}^T]^T$, where $r_i,\ g_i \in \mathbf{R}^{n_i}$ for $i = 1, \ldots, s$. Here $\xi/0$ is interpreted as zero if $\xi = 0$ and infinity otherwise.

Proof. It is easily seen that this bound is attained for the perturbation E where

$$E_{i,j} = \frac{\| A_{i,j} \| r_i {y_j}^T}{g_i \| y_j \|}.$$

Then

$$\| E_{i,j} \| = \frac{\| A_{i,j} \| \| r_i \|}{g_i}.$$

We can use the blockwise relative error as an easy way to terminate process. We have to check if $\mu(b - Ay) \leq \mu(A)\ \mu(y)\ 10^{-10}$ (say).

We prove that the speed of the convergence of iterative refinement depends mainly on the blockwise condition number of $Ax = b$ which measures the sensivity of the solution z to perturbations in the data and on the accuracy of computing residual vector $r = b - Az$. Notice that if a system $Ax = b$ is ill-conditioned then usually we can't find the solution x to very high accuracy in a blockwise sense.

3 Stability of the iterative refinement algorithm

The solution of the nonsingular linear system $Ax = b$ by some algorithm can be denoted by $W(b)$; that is, W is a mapping that approximates A^{-1} but is nonlinear due to floating point arithmetic.

We say that W is **forward stable** if there is some modest constant K_1 depending only on n such that

$$\| W(b) - A^{-1}b \| \leq \epsilon K_1 \ \ cond \| A^{-1}b \|, \tag{13}$$

whenever $\epsilon K_1 \ \ cond \leq 0.1$ where

$$cond = cond_{\mu}(A) = \| \mu(A^{-1})\mu(A) \| \tag{14}$$

is the blockwise condition number of A and ϵ is the precision.

We say that W is **backward stable** (normwise backward stable) if there is some modest constant K_2 depending only on n such that

$$\| AW(b) - b \| \leq \epsilon K_2 \| A \| \| A^{-1}b \| . \tag{15}$$

We investigate **recurrent iterative refinement (RIR)** for solving nonsingular linear systems $Ax = b$ using only **single precision arithmetic (fl).**

Recurrent iterative refinement was proposed by Woźniakowski; see eg. [16] . Kielbasiński [17], Sokolnicka and Smoktunowicz [24] applied this algorithm in increasing precision arithmetics (BCIR– binary cascades iterative refinement). Smoktunowicz [23], [25] developed results for RIR using only single precision arithmetic.

For recurrent iterative refinement we need a basic (direct or iterative) linear equation solver S_0 such that

$$\| S_0(b) - A^{-1}b \| \leq q \| A^{-1}b \|,$$

where $q \leq 1$.

A single iteration of iterative refinement is given by **1-fold iterative refinement :**

$$\begin{aligned} x &= S_0(b) \\ r &= fl(b - Ax) \\ p &= S_0(r) \\ y &= fl(x + p). \end{aligned}$$

Let us use $S_1(b)$ to denote the result y of this computation. We call this 1-fold iterative refinement.

The idea of $(k+1)$**- fold iterative refinement** is to replace S_0 in the above algorithm by S_k. Thus S_{k+1} is defined to be the result y of the computation:

$$\begin{aligned} x &= S_k(f) \\ r &= fl(f - Ax) \\ p &= S_k(r) \\ y &= fl(x + p). \end{aligned}$$

If $p = S_k(r)$ is replaced by $p = S_0(r)$ then the algorithm is k iterations of classical **iterative refinement (IR).**

Recurrent iterative refinement requires additional storage proportional to the depth of the recursion, which will not be too great because the computation time is proportional to 2^k.

In this section we consider only the case $s = n$: $\mu(A) = \mid A \mid$.

The explicit floating-point computations are assumed to satisfy

$$r = (I + D)(f - (A + G)u), \ \ \mu(D) \leq \epsilon\mu(I), \ \ \mu(G) \leq \epsilon L\mu(A), \tag{16}$$

where $L \geq 1$ depends on n alone, and

$$y = (I + F)(u + d), \ \ \mu(F) \leq \epsilon\mu(I). \tag{17}$$

We can use different kinds of algorithms for computing Au, because $v = Au$ can be written in the form $v_i = \sum_{j=1}^{s} A_{i,j}u_j$, $i = 1, \ldots, s$ and computed in parallel by different processors.

Theorem 3.1 **(1)** *Suppose that* $\epsilon \leq 0.01$ *and*

$$\| S_k(f) - A^{-1}f \| \leq \gamma_k \| A^{-1}f \|,$$

where $\gamma_k \leq 1$. *Then (1) holds for* $k + 1$ *with*

$$\gamma_{k+1} = {\gamma_k}^2 + \epsilon 8.11L \ cond$$

and

$$cond = cond_\mu(A) = \| \mu(A^{-1})\mu(A) \| . \tag{18}$$

(2) *Suppose in addition that* $\epsilon cond \leq 0.01$ *and*

$$\| AS_k(f) - f \| \leq \Delta_k \| A \| \| A^{-1}f \| .$$

Then (2) holds for $k + 1$ *with*

$$\Delta_{k+1} = (\gamma_k + \epsilon 4.02L \ cond) \, \Delta_k + \epsilon 4.09L.$$

Proof. We have $S_{k+1}(f) = y$ where

$$u = S_k(f),$$

$$r = (I + D)(f - (A + G)u),$$

$$d = S_k(r),$$

$$y = (I + F)(u + d).$$

(1) It is sufficient to show that

$$\| y - A^{-1}f \| \leq \gamma_{k+1} \| A^{-1}f \| .$$

We have

$$y - A^{-1}f = u + d - A^{-1}f + F(u+d)$$

and so

$$\| y - A^{-1}f \| \leq \| u + d - A^{-1}f \| + \epsilon \| u + d \|$$

with

$$\| u + d \| \leq \| A^{-1}f \| + \| u + d - A^{-1}f \| .$$

Making use of (1), we have

$$\| u + d - A^{-1}f \| \leq \gamma_k \ \| A^{-1}r \| + \| A^{-1}r - A^{-1}f + u \|$$

and, again, we get

$$\| A^{-1}r \| \leq \gamma_k \ \| A^{-1}f \| + \| A^{-1}r - A^{-1}f + u \| .$$

For the last term we obtain

$$A^{-1}r - A^{-1}f + u = -A^{-1}DA(u - A^{-1}f) - A^{-1}(I+D)Gu,$$

from which we get

$$\| A^{-1}r - A^{-1}f + u \| \leq \epsilon cond \gamma_k \| A^{-1}f \| + \epsilon 1.01 \ Lcond \| u \|$$

with

$$\| u \| \leq (1+\gamma_k) \| A^{-1}f \| \leq 2 \ \| A^{-1}f \| .$$

Working backwards on the chain of inequalities we have

$$\| A^{-1}r - A^{-1}f + u \| \leq \epsilon 3.02 \ Lcond \ \| A^{-1}f \|,$$

$$\| A^{-1}r \| \leq (\gamma_k + \epsilon 3.02 \ Lcond) \ \| A^{-1}f \|,$$

$$\| u + d - A^{-1}f \| \leq ({\gamma_k}^2 + \epsilon 6.04 \ Lcond) \ \| A^{-1}f \|,$$

$$\| u + d \| \leq (2 + \epsilon \ 6.04L \ cond) \| A^{-1}f \|$$

where we have used $\gamma_k \leq 1$.

(2) It is enough to show that

$$\| Ay - f \| \leq \Delta_{k+1} \| A \| \| A^{-1}f \| .$$

We obtain

$$\| Ay - f \| \leq \| A(u+d) - f \| + \epsilon \| A \| \| u + d \| .$$

Using (2), we have

$$\| A(u+d) - f \| \leq \| r - f + Au \| + \Delta_k \| A \| \| A^{-1} r \| .$$

From this it follows that

$$\| A(u+d) - f \| \leq ((\gamma_k + \epsilon\ 4.02 L cond)\Delta_k + \epsilon 2.02\ L) \| A \| \| A^{-1} f \| .$$

Combining this with $1 \leq L\ cond$ and $\epsilon\ cond \leq 0.01$ establishes (2).

Theorem 3.2 **(1)** *Assume that $\gamma_0 \leq 0.9$ (say) and $\epsilon\ Lcond \leq 0.01$ (say). Then there exists k_1 depending only on n such that*

$$\| S_k(f) - A^{-1} f \| \leq \epsilon 10 L\ cond \| A^{-1} f \|$$

whenever $k \geq k_1$ and

$$cond = cond_{\mu}(A) = \| \mu(A^{-1})\mu(A) \| . \tag{19}$$

(2) *Also there exists k_2 depending only on n such that*

$$\| AS_k(f) - f \| \leq \epsilon 5\ L \| A \| \| A^{-1} f \|$$

whenever $k \geq k_2$.

Proof. Using the previous theorem, we show by induction on k that the pair of conditions

$$\gamma_k \leq 0.9,$$

$$\gamma_{k+1} = {\gamma_k}^2 + \epsilon 8.11\ Lcond$$

holds for all k. It can be shown that the limit of the sequence $\{\gamma_k\}$ is less than $\epsilon 9 L\ cond$, which establishes the first result.

(2) Clearly we can choose $\Delta_0 \leq \gamma_0$. Then it is easy to show that $\Delta_k \leq \gamma_k$ for all k so that for $k \geq k^*$ we have

$$\Delta_k \leq \epsilon 10\ L\ cond \leq 0.1,$$

$$\Delta_{k+1} \leq 0.15\Delta_k + \epsilon\ 4.09 L$$

from which the second result easily follows.

REFERENCES

1. M.Arioli, J.W.Demmel and I.S.Duff, *Solving sparse linear systems with sparse backward error*, SIAM J.Matrix Anal.Appl., 10 (1989), pp. 165-190.

2. M.Arioli, I.S.Duff and P.P.M. de Rijk, *On the augmented system approach to sparse least-squares problems*, Numer.Math., 55(1989), 667-684.

3. F.L.Bauer, *Genauigkeitsfragen bei der Lösung linearer Gleichungssysteme*, ZAMM, 46 (1966), 667-684.

4. Å.Björck, *Iterative refinement of linear least squares solutions I*, BIT, 7 (1967), 257–278.

5. Å.Björck, *Iterative refinement and reliable computing,* M.G.Cox and S.J.Hammarling, eds.., Oxford Univ. Press, 1990, 249-266.

6. Å.Björck, *Component-wise perturbation analysis and error bounds for linear least squares solutions,* BIT, 31(1991), 238-244.

7. Å. Björck, Numerical Methods for Least Squares Problems, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1996.

8. J.Demmel, *The componentwise distance to the nearest singular matrix*, SIAM J.Matrix Anal. Appl., 13 (1992), 10-19.

9. E.Deutsch, *Matricial norms*, Numer.Math. 16 (1970), 73–84.

10. J.Głuchowska and A.Smoktunowicz, *Solving the linear least squares problem with very relative accuracy*, Computing, 45 (1990), 345-354.

11. G.H.Golub and J.H.Wilkinson, *Note on the iterative refinement of least squares solution,* Numer.Math., 9 (1966), 139-704.

12. D.J.Higham, *Condition numbers and their condition numbers*, Linear Algebra Appl., 214 (1995), 193-213.

13. D.J.Higham and N.J.Higham, *Backward error and condition of structured linear systems,* SIAM J.Matrix Anal.Appl., 13 (1992), 162-175.

14. N.J.Higham, *Iterative refinement enhances the stability of QR factorization methods for solving linear equations,* BIT 31 (1991), 441-468.

15. N.J.Higham, *Iterative refinement for linear systems and LAPACK*, Numerical Analysis Report No. 277, September 1995, Univ. of Manchester.

16. M.Jankowski and H.Woźniakowski, *Iterative refinement implies numerical stability*, BIT, 17 (1977), 303-311.

17. A.Kielbasiński, *Iterative refinement for linear systems in variable–precision arithmetics*, BIT, 21 (1981), 97-103.

18. A.M.Ostrowski, *On some metrical properties of operator matrices and matrices partitioned into blocks*, Journal of Mathematical Analysis and Applications, 2 (1961), 161-209.

19. M.Pankowski, *Iterative refinement for solving linear system of equations* (in Polish), M.Sc.Thesis, Warsaw Univ. of Techn., 1995.

20. J.L.Rigal and J.Gaches, *On the compatability of a given solution with the data of a linear system*, J.Assoc.Comput.Mach., 14 (1967), 543-548.

21. J.Rohn, *New condition numbers for matrices and linear systems*, Computing, 41(1989), 167-169.

22. R.D.Skeel, *Iterative refinement implies numerical stability for Gaussian elimination*, Math.Comp. , 35 (1980), 817-832.

23. A.Smoktunowicz, *Stability issues for special algebraic problems*, Ph.D. Thesis, Univ. of Warsaw, 1981 (in Polish).

24. A.Smoktunowicz and J.Sokolnicka, *Binary cascades iterative refinement in doubled–mantissa aritmetics*, BIT, 24 (1984), 123-127.

25. A.Smoktunowicz, *Iterative refinement of solutions of linear system in single precision*, Univ. of Warsaw, 1988, Grant No. RP.0.09.

26. J.H.Wilkinson, *Rounding errors in algebraic processes*, Clarendon Press, Oxford, 1965.

Institute of Mathematics, Warsaw University of Technology,
Pl.Politechniki 1, 00-661 Warsaw, Poland
(smok@im.pw.edu.pl)