On the Behavior of the Residual in Conjugate Gradient Method

Teruyoshi Washizawa
Analysis Technology Center, Canon Inc., Tokyo, Japan
E-mail: *washizawa.teruyoshi@canon.co.jp*

Abstract

In conjugate gradient method, it is well known that the recursively computed residual differs from true one as the iteration proceeds in finite arithmetic. Some work have been devoted to analyze this behavior and to evaluate the lower and the upper bounds of the difference. This paper focuses on the behavior of these two kinds of residuals, especially their lower bounds caused by the loss of trailing digit, respectively.

1. Introduction

Conjugate gradient (CG) method and its varieties are popular as one of the best unsteady iterative methods for solving the following linear system:

$$Ax = b \qquad (1)$$

In CG method, an approximate solution x_k is expected to approach the exact solution x^*. For the symmetric positive definite A, it is proved that the A-norm of the error monotonically decreases as the iteration proceeds in exact arithmetic. This will be called as A-norm monotonicity of the error in the remaining part of this article. It is obvious that we cannot calculate directly such a norm of the error without the solution. Therefore, almost algorithms employ the residual which is easily calculated as the difference between the left hand side (LHS) and the right hand side (RHS) of (1), $r_k := b - Ax_k$. In practice, the residual is calculated by the recursion formula because of the computational complexity of the matrix vector product Ax_k [1,2].

However, this recursion formula causes another problem in which the recursive residual differs from the true residual as the iteration proceeds. It can be also observed that the recursive residual decreases after the true one seams to reach its lower bound. We should terminate the CG steps just before the difference is too large to be neglected.

Ginsburg has proposed a simple criterion [1]:

For the true residual calculated as the difference between LHS and RHS of a linear system and the recursive residual calculated by using the recursion formula, the procedure is terminated when the 2-norm of their difference is greater than the 2-norm of the recursive residual:

$$\|r_k\| < \exp(k/n)^2 \|s_k - r_k\|$$

where n is dimensionarity of a linear system.

Several researchers have proposed the estimations of the lower and the upper bound of the norm of the error and the residual. Woźniakowski investigated the numerical stabilities and good-behaviors of three stationary iterative methods and CG method using the true residual $b - Ax_k$ [3,4]. Woźniakowski gave the upper bound of the ultimately attainable accuracies of the A-norm and the 2-norm of the error, and 2-norm of the true residual. Bollen gives the round-off error analysis of descent methods and lead a general result on the attainable accuracy of the approximate solution in finite arithmetic [2]. It has also shown that the general result is applied to the Gauss-Southwell method and the gradient method to obtain the decreasing rates of the A-norm of the error in finite arithmetic. Greenbaum have shown that for tiny perturbation ε_M, the eigenvalues and the A-norm of the error vectors generated over a fixed number of perturbed iterative steps are approximately the same as those quantities generated by the exact recurrences applied to a "nearby" matrix [5]. The lower bound of the true residual is pointed out in [6]. Two kinds of the estimates of the A-norm of the error at every step in CG algorithm has been proposed and verified that those estimates are the lower and the upper bound in [7]. The lower and the upper bounds of the A-norm of the error have been also given by Meurant [7] and Strakoš and Tichý [8]. Strakoš

and Tichý have proposed the tight estimate for the lower bound of both the A-norm and the 2-norm of the error in every step. This stepwise lower bound, however, keeps decreasing after the error reaches its 'global' lower bound. Therefore, the terminating criteria by using this stepwise lower bound cannot detect the global lower bound of the error. Calvetti *et al.* has proposed the estimates of the lower and upper bound of the A-norm of the error in CG method [9]. Those previous studies give the stepwise lower and the upper bound of the error and the residual but the global bounds.

In the remaining part of this article, we will first show that the true and the recursive residual almost monotonically decrease as the iteration proceeds. Then, these lower bounds will be shown.

2. Notations

We shall give the notations appeared throughout this article. A and b is, respectively, a coefficient matrix and a constant vector in a linear system. $\|\cdot\|$ in connection with a vector and a matrix, respectively, stands for the 2-norm and spectral norm, $\|\cdot\|_A$ in connection with a vector stands for the norm under the metric tensor A. The exact value of a variable x is denoted as $\bar{x}$. The floating point representation of a variable x is denoted simply as x. The computational error caused by the floating point representation is denoted as an operator $\varepsilon_M(x) := x - \bar{x}$.

The exact solution of (1) is denoted as x^* which is described formally as $x^* = A^{-1}b$. At the k-th step, an approximate solution, the error, the true residual, and the recursive residual is, respectively, described as x_k, e_k, s_k, and r_k. They are computed in CG method as follows:

$$x_{k+1} := x_k + \alpha_k p_k, e_k := x^* - x_k,$$
$$s_k := b - Ax_k, r_{k+1} := r_k - \alpha A p_k$$

Since A is a constant matrix and $\|\varepsilon_M(A)\|/\|A\|$ is almost equal to ε_M without the dependence on the number of iterations, $\varepsilon_M(A)$ is out of our concern as well as $\varepsilon_M(b)$.

3. Almost Monotonicity of Residuals in Finite Arithmetic

In this section, we will see the true and the recursive residual has the 2-norm almost monotonicity in finite arithmetic, respectively.

3.1. The 2-Norm Almost Monotonicity of True Residual

The true residual is calculated as the difference between LHS and RHS. This is equivalent to multiplication of A

to the error in finite arithmetic,

$$s_k = b - Ax_k = b - A(x_k - x^*) - Ax^* = -Ae_k$$

The behavior of true residual s_k is, therefore, equivalent to Ae_k.

The A-norm monotonicity of the error in finite arithmetic has been proved in theorem-3.1 of [2]. The following theorem shows the error has the 2-norm almost monotonicity.

Theorem-1. If e_k has the A-norm monotonicity, e_k has the 2-norm almost monotonicity for a regular matrix A,

$$\exists k > j, \|e_k\| < \|e_j\| \tag{2}$$

Proof. The relationship between 2-norm and A-norm of error is

$$\|e_k\| = \|A^{-1/2}e_k\|_A \leq \|A^{-1/2}\|\|e_k\|_A \tag{3}$$

Similarly,

$$\|e_k\|_A = \|A^{1/2}e_k\| \leq \|A^{1/2}\|\|e_k\| \tag{4}$$

Substituting (3) and (4) into (2) and we yield

$$\|A^{-1/2}\|\|e_k\|_A < \|A^{1/2}\|^{-1}\|e_j\|_A \tag{5}$$

Equation (5) holds if k exists to satisfy the following relation

$$\|e_k\|_A < \kappa\left(A^{1/2}\right)^{-1}\|e_j\|_A \tag{6}$$

where $\kappa(A)$ is the condition number of a matrix A. From the A-norm monotonicity of the error e_k, since the following relation holds for any positive value a,

$$\exists k > j, \|e_k\|_A < a\|e_j\|_A$$

there exists $k > j$ that satisfies (6) and consequently (2). We have to notice that (2) does not hold when $a = 0$, *i.e.*, the inverse of the coefficient matrix A is singular.

Theorem-1 leads to the 2-norm almost monotonicity of the true residual using the relationship $\|s_k\| = \|Ae_k\|$.

Theorem-2. If e_k has the 2-norm almost monotonicity, s_k has the 2-norm almost monotonicity for a regular matrix A,

$$\exists k > j, \|s_k\| < \|s_j\| \tag{7}$$

Proof. The relationship between the 2-norm of the error and that of the residual is

$$\|s_k\| = \|Ae_k\| \leq \|A\|\|e_k\| \tag{8}$$

Similarly,

$$\|e_j\| = \|A^{-1}s_j\| \leq \|A^{-1}\|\|s_j\| \tag{9}$$

Substituting (8) and (9) into (7) and we yield

$$\|A\|\|e_k\| < \|A^{-1}\|^{-1}\|e_j\| \qquad (10)$$

Equation (10) holds if k exists to satisfy the following relation

$$\|e_k\| < \kappa(A)^{-1}\|e_j\| \qquad (11)$$

From the 2-norm almost monotonicity of the error e_k, since the following relation holds for any positive value a,

$$\exists k > j, \|e_k\| < a\|e_j\| \qquad (12)$$

there exists $k > j$ that satisfies (11) and consequently (7).

3.2. The 2-Norm Almost Monotonicity of Recursive Residual

Before the proof of almost monotonicity of recursive residual in finite arithmetic, we first give the proof of almost monotonicity of recursive residual in exact arithmetic.

From the A-norm monotonicity of the error in exact arithmetic, the 2-norm almost monotonicity of the residual in exact arithmetic can be proved. We have to notice that the recursive residual $\bar{r}_j$ is identical to the true residual $\bar{s}_j$ in exact arithmetic.

Theorem-3. If $\forall n, \|\bar{e}_{n+1}\|_A < \|\bar{e}_n\|_A$, then the following proposition holds for a regular matrix A:

$$\exists k > j, \|\bar{r}_k\| < \|\bar{r}_j\| \qquad (13)$$

Proof. The relationship between the error and the residual gives:

$$\|\bar{r}_k\| = \|A\bar{e}_k\| = \|A^{1/2}\bar{e}_k\|_A \qquad (14)$$

Then we yield the lower and the upper bound of the 2-norm of the true residual:

$$\|A^{-1/2}\|^{-1}\|\bar{e}_k\|_A \le \|\bar{r}_k\| \le \|A^{1/2}\|\|\bar{e}_k\|_A \qquad (15)$$

From above equation, the sufficient condition for (13) can be given as follows:

$$\|A^{1/2}\|\|\bar{e}_k\|_A < \|A^{-1/2}\|^{-1}\|\bar{e}_j\|_A \qquad (16)$$

that is, the equation holds if there exists $k > j$ so that

$$\|\bar{e}_k\|_A < \kappa(A^{1/2})^{-1}\|\bar{e}_j\|_A \qquad (17)$$

From the A-norm monotonicity of the error $\bar{e}_k$, the following equation holds for any positive value a,

$$\exists k > j, \|\bar{e}_k\|_A < a\|\bar{e}_j\|_A \qquad (18)$$

and (13) holds.

Now we show the almost monotonicity of the recur-

sive residual in finite arithmetic.

Theorem-4. If the recursive residual has 2-norm almost monotonicity in exact arithmetic, then the recursive residual has the 2-norm almost monotonicity in finite arithmetic:

$$\exists k > j, \|r_k\| < \|r_j\| \qquad (19)$$

Proof. Equation (19) is rewritten as

$$\exists k > j, \|\bar{r}_k + \varepsilon_M(\bar{r}_k)\| < \|\bar{r}_j + \varepsilon_M(\bar{r}_j)\| \qquad (20)$$

The following relationship is one of its sufficient conditions

$$\max\left[\|\bar{r}_k + \varepsilon_M(\bar{r}_k)\|\right] < \min\left[\|\bar{r}_j + \varepsilon_M(\bar{r}_j)\|\right]$$

The evaluation of the maximum value of LHS is

$$\max\left[\|r_k\|\right] = (1 + \varepsilon_M)\|\bar{r}_k\| \qquad (21)$$

Similarly, the minimum value of RHS is evaluated as

$$\min\left[\|r_j\|\right] = (1 - \varepsilon_M)\|\bar{r}_j\| \qquad (22)$$

Substituting (21) and (22) into (20), the sufficient condition of (20) is given as

$$\exists k > j, \|\bar{r}_k\| < (1 - \varepsilon_M)/(1 + \varepsilon_M)\|\bar{r}_j\| \qquad (23)$$

There exists $k > j$ for $(1 - \varepsilon_M)/(1 + \varepsilon_M) > 0$ from theorem-3 and (23) holds.

4. Lower Bounds of Error and Residual in Finite Arithmetic

It has been shown that the 2-norm of two kinds of residuals, respectively, decreases almost monotonically in finite arithmetic in the previous section.

Now we consider whether if the 2-norm of each variable stops decreasing before the approximate x_k does not reach its target x^*.

4.1. Lower Bound of Error

Theoem-1 shows the approximate solution x_j approaches the exact solution almost monotonically in finite arithmetic. The correction of the recursion formula of x_k, however, vanishes by the loss of trailing digits so that the error stops changing, *i.e.*,

$$\left|\Delta x_{stop}(n)\right|/\left|x_{stop}(n)\right| < \varepsilon_M \Rightarrow x_{stop+1} = x_{stop}$$

where $x_k(n)$ is the n-th component of x_k.

On the other hand, the solution in finite arithmetic x^* is not always identical to that in exact arithmetic $\bar{x}^*$.

Therefore, the target for iterative algorithms in finite

arithmetic should not be $\overline{x}^*$ but x^*. The error caused by the loss of trailing digits is described formally as $x^* - x_{stop}$. Since the true residual is given by multiplying A to the error, the lower bound of the true residual is given as $A\left(x^* - x_{stop}\right)$.

4.2. Lower Bound of Recursive Residual

Theorem-4 shows the recursive residual reduces its 2-norm almost monotonically. The next theorem proves that the change of the recursive residual stops only when $\|r_{k+1}\| < \varepsilon_M \|r_k\|$. It decreases almost monotonically until then.

Theorem-5. The recursive residual never have a lower bound caused by the loss of trailing digits.

Proof. The recursion formula of the residual is in general described as

$$r_{k+1} = r_k - \Delta r_k \qquad (24)$$

where $\Delta r_k := \alpha_k A p_k$. The residual reaches its lower bound r_k if the following condition is satisfied:

$$\Delta r_k < \varepsilon_M \left(r_k\right) \qquad (25)$$

We will show the condition of (25) never be satisfied in not only exact but also finite arithmetic.
In exact arithmetic, (24) satisfies the following relationship:

$$\begin{aligned}
\left\|\overline{r}_{k+1}\right\|^2 &= \left\|\overline{r}_k - \Delta \overline{r}_k\right\|^2 = \left\|\overline{r}_k\right\|^2 - 2\left(\overline{r}_k, \Delta \overline{r}_k\right) + \left\|\Delta \overline{r}_k\right\|^2 \\
&= \left\|\overline{r}_k\right\|^2 - 2\left(\overline{r}_k, \overline{r}_k - \overline{r}_{k+1}\right) + \left\|\Delta \overline{r}_k\right\|^2 \\
&= \left\|\overline{r}_k\right\|^2 - 2\left\|\overline{r}_k\right\|^2 + \left\|\Delta \overline{r}_k\right\|^2 = -\left\|\overline{r}_k\right\|^2 + \left\|\Delta \overline{r}_k\right\|^2
\end{aligned}$$

where using $\left(\overline{r}_k, \overline{r}_{k+1}\right) = 0$.

Therefore, three vectors in above recursion formula forms the right triangle so that the following relationship holds in exact arithmetic :

$$\left\|\Delta \overline{r}_k\right\| \geq \left\|\overline{r}_k\right\| \qquad (26)$$

This shows that (25) never be satisfied in exact arithmetic.
Now we evaluate above in finite arithmetic.

$$\begin{aligned}
\frac{\left\|\Delta r_k\right\|}{\left\|r_k\right\|} &\geq \frac{\min\left[\left\|\Delta \overline{r}_k + \varepsilon_M\left(\Delta \overline{r}_k\right)\right\|\right]}{\max\left[\left\|\overline{r}_k + \varepsilon_M\left(\overline{r}_k\right)\right\|\right]} \\
&= \frac{\left(1-\varepsilon_M\right)\left\|\Delta \overline{r}_k\right\|}{\left(1+\varepsilon_M\right)\left\|\overline{r}_k\right\|} = \left(1 - \frac{2\varepsilon_M}{1+\varepsilon_M}\right)\frac{\left\|\Delta \overline{r}_k\right\|}{\left\|\overline{r}_k\right\|}
\end{aligned}$$

According to (26), we yield

$$\left\|\Delta r_k\right\|/\left\|r_k\right\| \geq \left\{1 - 2\varepsilon_M/\left(1+\varepsilon_M\right)\right\}$$

and prove that the recursive residual never have a lower bound caused by the loss of trailing digits.
The termination of the iterations is caused only when $\|r_{k+1}\| \leq \varepsilon_M \|r_k\|$ by the significant decrease of the recursive residual for $\alpha_k A p_k \approx r_k$.

5. Conclusions

In this article, the convergence behaviors of true and recursive residual have been analyzed.
The results obtained are summarized below:
1) In finite arithmetic, the 2-norm of the error and the residual, respectively, almost monotonically decreases.
2) 2-norm of the error has the lower bound in finite arithmetic as well as the true residual.
3) 2-norm of the recursive residual never has a non-zero lower bound caused by the loss of trailing digits in finite arithmetic.

6. References

[1] T. Ginsburg, "The Conjugate Gradient Method," *Numerische Mathematik*, Vol. 5, No. 1, 1963, pp. 191-200.

[2] J. A. M. Bollen, "Numerical Stability of Descenet Methods for Solving Linear Equations," *Numerische Mathematik*, Vol. 43, No. 3, 1984, pp. 361-377.

[3] H. Woźniakowski, "Round-off Error Analysis of Iteratinos for Large Linear Systems," *Numeriche Mathematik*, Vol. 30, No. 3, 1978, pp. 301-314.

[4] H. Woźniakowski, "Roundoff Error Analysis of New Class of Conjugate-Gradient Algorithms," *Linear Algebra and its Applications*, Vol. 29, 1980, pp. 507-529.

[5] A. Greenbaum, "Behavior of Slightly Perturbed Lanczos and Conjugate-Gradient Recurrences," *Linear Algebra and its Applications*, Vol. 113, 1989, pp. 7-63.

[6] A. Greenbaum, "Estimating the Attainable Accuracy of Recursively Computed Residual Methods," *SIAM Journal on Matrix Analysis and Applications*, Vol. 18, No. 3, 1997, pp. 535- 551.

[7] G. Meurant, "The Computation of Bounds for the Norm of the Error in the Conjugate Gradient Algorithm," *Numerical Algorithms*, Vol. 16, No. 3-4, 1997, pp. 77-87.

[8] Z. Strakoš and P. Tichý, "On Error Estimation in the Conjugate Gradient Method and Why it Works in Finite Precision Computations," *Electronic Transactions on Numerical Analysis*, Vol. 13, 2002, pp. 56-80.

[9] D. Calvetti, S. Morigi, L. Reichel and F. Sgallari, "Computable Error Bounds and Estimates for the Conjugate Gradient Method," *Numerical Algorithms*, Vol. 25, No. 1-4, 2000, pp. 75-88.