New Iterative Improvement of a Solution for an Ill-Conditioned System of Linear Equations Based on a Linear Dynamic System

XINYUAN WU, RONG SHAO AND YIRAN ZHU
Department of Mathematics
Nanjing University
Nanjing 210093, P.R. China

Abstract—In this paper, the analysis of the dynamic system for iterative improvement of a solution is discussed, and the new iterative improvement of a solution is proposed based on the dynamic system. We have proved that the new iterative improvement of the solution is convergent unconditionally. The numerical experiments illustrate that the new iterative improvement of the solution is more effective for an ill-conditioned system of linear equations. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords—Iterative improvement of solution, ill-conditioned system of linear equations, Dynamic system, ODE recursion, Iteration method, Preconditioner.

1. INTRODUCTION

It is well known that if matrix $A$ of coefficients is ill conditioned, then the computed solution $x_1$ of linear system

$$Ax = b$$

may not be sufficiently accurate. Therefore, the iterative improvement of the solution is employed in order to generate an iterative sequence $\{x_s\}$ of the approximate solution that converges to the solution of (1.1). A popular procedure of the iterative improvement is the famous Wilkinson’s iterative refinement [1,2] of solution in which the Cholesky decomposition is repeatedly used. It has been shown [3,4] that if $A$ is not “too ill conditioned”, then the $x_s$ will be convergent to the solution “to working accuracy”, provided the residual $r_s = b - Ax_s$ is computed to double-precision. But, from the point of view of numerical computation, this method may be impractical for the ill-conditioned system of linear equations. For example, consider an ill-conditioned system of equations with a $12 \times 12$ Hilbert matrix. If we use Wilkinson’s iterative improvement, it will provide the computed solution without any significant figures, no matter how many iterative
steps are performed. This is mainly because Wilkinson’s iterative improvement of solution can be viewed as an iterative procedure with starting vector $x_0 = 0$,

$$Ay_n = b - Ax_n,
\quad x_{n+1} = x_n + y_n,
\quad n = 0, 1, 2, \ldots, \quad (1.2)$$

or equivalently,

$$y_n = A^{-1}(b - Ax_n),
\quad x_{n+1} = x_n + y_n,
\quad n = 0, 1, 2, \ldots, \quad (1.3)$$

with $x_0 = 0$. This means that Wilkinson’s iterative improvement of solution can be viewed as an explicit Euler method with step size $h = 1$ for solving the following system of ordinary differential equations:

$$\frac{dy}{dt} = A^{-1}(b - Ay),
\quad x(0) = x_0 = 0, \quad (1.4)$$

where $x^* = A^{-1}b$ is a unique stationary point of system (1.4).

Since $A$ is ill conditioned, neither the numerical computed solution of (1.2) nor of (1.3) is sufficiently accurate, although they can be viewed as a procedure that generates numerical approximate sequence $\{x_n\}$ along the orbit defined by (1.4). Especially if $A$ is seriously ill conditioned, it may be unworkable to produce the sufficiently accurate $A^{-1}$ numerically, or equivalently, to generate sufficiently accurate $y_n$ step by step. Therefore, when $A$ is ill conditioned, in order to deal with this problem, a transformation should be considered in advance and the behaviour of the dynamic system relating to the continuation-like method (1.4) also should be analysed. Then some new iterative improvement of solution can be expected to be explored further. The principal strategy of us is to convert an ill-conditioned problem into a series of relatively “well-conditioned” problems but they have the same solution, and then solve numerically the “well-conditioned” problems by the method of successive approximation. So it also can be viewed as a new preconditioner. The discussion in detail will be presented in the next sections.

2. THE BEHAVIOUR OF A LINEAR DYNAMIC SYSTEM RELATING TO THE NEW ITERATIVE IMPROVEMENT OF THE SOLUTION UNDER CONSIDERATION

Consider a system of linear equations with an $m \times m$ matrix of coefficients

$$Ax = b, \quad (2.1)$$

where we restrict our attention to the matrix of coefficients which is normal positive definite. Wilkinson [1,2] offered iterative improvement (1.2) of solutions in which the Cholesky decomposition is employed. However, in Wilkinson’s iterative procedure, the improvement of the solution at every step still suffers from $\text{cond}(A)$. In other words, Wilkinson’s iteration is performed without improving the condition of the problem to be solved. This enlightens us to discuss the following iterative improvement of solution with a parameter $u \geq 0$:

$$(uI + A)x_{n+1} = ux_n + b, \quad (2.2)$$

or equivalently,

$$x_{n+1} = (uI + A)^{-1}(ux_n + b). \quad (2.3)$$

It is easy to see that (2.2) can be rewritten as

$$(uI + A)y_n = b - Ax_n, \quad (2.2')$$

$$x_{n+1} = x_n + y_n,$$
and accordingly (2.3) can be rewritten as

\[ y_n = (uI + A)^{-1}(b - A\xi_n), \]
\[ x_{n+1} = x_n + y_n. \]  

(2.3')

Explicitly, in the special case of \( u = 0 \), it leads to Wilkinson's iterative improvement of the solution, namely (1.2) or (1.3).

On one hand, many integrators for linear constant coefficient ODEs can be identified as an approximation of linear system (2.1), and there exists a connecting between ODE recursions and iterative solvers (see [5]); on the other hand, (2.3) or (2.3') is just an ODE recursion. Thus, we first consider the dynamic systems associated with (2.3) or (2.3').

The dynamic systems relating to the iterative improvement (2.3) or (2.3') are

\[ \frac{dx}{dt} = (uI + A)^{-1}(b - Ax), \]
\[ x(0) = x_0. \]  

(2.4)

where \( u \geq 0, x_0 \in \mathbb{R}^n \).

Obviously, letting \( u = 0 \) and \( x_0 = 0 \) in (2.4), we obtain (1.4), that is the dynamic system relating to Wilkinson's improvement of solution. So, in the following discussion, we only consider \( u > 0 \) in (2.4).

Now, let us study the behaviour of dynamic systems (2.4).

**Theorem 2.1.** The solution \( x^* \) of linear system (2.1) is a unique globally asymptotically stable equilibrium point of the dynamic systems (2.4).

**Proof.** Let

\[ f(x) = b - Ax \]

and

\[ v(x) = f^T(x)\frac{f(x)}{2}. \]  

(2.5)

Then we have

(i) \( v(x^*) = 0 \) and \( v(x) > 0 \), provided \( x \neq x^* \), and

(ii) \[ v'(x) = \frac{dv(x)}{dt} = f^T(x)(-A) \frac{dx}{dt} = -f^T(x)A(uI + A)^{-1}f(x) \]

\[ = -f^T(x)\left[(uI + A)^{-1}\right]^{-1}f(x) = -f^T(x)(I + uA^{-1})^{-1}f(x) < 0, \]

provided \( x \neq x^* \);

this is due to \( u > 0 \) and \( A \) is positive definite, so do \( A^{-1}, (I + uA^{-1}), \) and \( (I + uA^{-1})^{-1} \).

Thus, we conclude that \( v(x) \) defined by (2.5) is a strict Lyapunov function of the equilibrium point \( x^* \) of dynamic systems (2.4). Therefore, the unique solution \( x^* \) of linear system (2.1) is a unique globally asymptotically stable equilibrium point of dynamic systems (2.4).

**Corollary 2.1.** Suppose that the solution of systems (2.4) can be expressed

\[ x = x(t, x_0); \]  

(2.6)

then we have

\[ \lim_{t \to +\infty} x(t, x_0) = x^* = A^{-1}b. \]  

(2.7)

**Proof.** It follows immediately from Theorem 2.1 and the definition of the well-known Lyapunov's asymptotical stability.

From (2.7) we can see that the dynamic systems (2.4) present a continuation method for linear system (2.1). However, generally speaking, it is not easy to obtain the solution of linear system (2.1) directly from (2.7), because it is difficult to yield the analytic expression of \( x(t, x_0) \) in (2.7), when \( m \) is large enough. Consequently, we will employ numerical integration for ODEs (2.4).
3. THE NEW ITERATIVE IMPROVEMENT OF A SOLUTION BASED ON DYNAMIC SYSTEMS (2.4) AND THE ANALYSIS OF CONVERGENCE

One of the simple numerical integrations is the Euler method. For (2.4), the Euler method gives

\[ x_{n+1} = x_n + h(uI + A)^{-1}(b - Ax_n), \quad n = 0, 1, \ldots, \]

where \( h \) is a step size. In the following, the step size \( h = 1 \) is chosen for simplicity and convenience. In this case, we have the ODE recursion

\[ x_{n+1} = x_n + (uI + A)^{-1}(b - Ax_n), \quad n = 0, 1, \ldots, \]

This is (2.3').

In view of spurious solutions of numerical methods for initial value problems [6–9], we will investigate the behaviour of the ODE recursion (3.2).

Equation (3.2) can be rewritten as

\[ x_{n+1} = (uI + A)^{-1}(ux_n + h), \quad n = 0, 1, \ldots, \]

since \((uI + A)(x_{n+1} - x_n) = b - Ax_n\).

Equation (3.3) can be regarded as a stationary iterative method with the iterative matrix

\[ M = u(uI + A)^{-1}. \]

Because \( A \) is a positive definite normal matrix and \( u > 0 \), \( uI + A \), \((uI + A)^{-1}\), and \( u(uI + A)^{-1}\) are all positive definite normal matrices.

THEOREM 3.1. Assume that \( \sigma_i \) \( (i = 1, 2, \ldots, m) \) are eigenvalues of \( m \times m \) positive definite normal matrix \( A \) in linear system (2.1) and

\[ 0 < \sigma_1 \leq \sigma_2 \leq \cdots \leq \sigma_m. \]

Then the spectral radius of the iterative method (3.3) is

\[ \rho(M) = \rho \left[ u(uI + A)^{-1} \right] = \frac{u}{u + \sigma_1}, \]

and the asymptotic rate of convergence of (3.2) or (3.3) is

\[ R(M) = -\ln \rho(M) = -\ln \left( \frac{u}{u + \sigma_1} \right), \]

for any \( u > 0 \).

PROOF. Because \( \sigma_i \) \( (i = 1, 2, \ldots, m) \) are the eigenvalues of \( A \), from (3.5) and \( u > 0 \), we have

\[ 0 < u + \sigma_1 \leq \sigma_2 \leq \cdots \leq u + \sigma_m \]

are the eigenvalues of \( uI + A \). Therefore,

\[ \frac{1}{u + \sigma_m} \leq \frac{1}{u + \sigma_{m-1}} \leq \cdots \leq \frac{1}{u + \sigma_1} \]

are the eigenvalues of \((uI + A)^{-1}\) and

\[ \frac{u}{u + \sigma_m} \leq \frac{u}{u + \sigma_{m-1}} \leq \cdots \leq \frac{u}{u + \sigma_1} \]

are the eigenvalues of \( u(uI + A)^{-1} \). Thus,

\[ \rho(M) = \rho \left[ u(uI + A)^{-1} \right] = \max_{1 \leq i \leq m} \frac{u}{u + \sigma_i} = \frac{u}{u + \sigma_1}, \]

and the asymptotic rate of convergence of (3.2) or (3.3) is

\[ R(M) = -\ln \rho(M) = -\ln \left( \frac{u}{u + \sigma_1} \right). \]
COROLLARY 3.1. The new iterative improvement (3.2) of solution is convergent unconditionally for any $u > 0$.

The result of Corollary 3.1 is straightforward from (3.7) of Theorem 3.1; in fact, we have $p(M) < 1$ for any $u > 0$.

THEOREM 3.2. Suppose that $0 < \sigma_1 \leq \sigma_2 \leq \cdots \leq \sigma_m$ are eigenvalues of the positive definite normal matrix $A$ in (2.1). Then for any $u > 0$ we always have

\[ \kappa(uI + A) < \kappa(A), \]  

(3.8)

where $\kappa(A)$ and $\kappa(uI + A)$ express the spectral condition numbers of $A$ and $(uI + A)$, respectively.

PROOF. The spectral condition number of $A$ and $(uI + A)$ are

\[ \kappa(A) = \frac{\sigma_m}{\sigma_1} \]

and

\[ \kappa(uI + A) = \frac{u + \sigma_m}{u + \sigma_1}, \]  

(3.9)

respectively. Since $u > 0$, and $\sigma_m > \sigma_1 > 0$, from the strictly monotone decrease of the function

\[ \varphi(x) = \frac{x + \sigma_m}{x + \sigma_1}, \quad x \geq 0 \quad \left( \text{since } \varphi'(x) = \frac{\sigma_1 - \sigma_m}{(x + \sigma_1)^2} < 0, \ x \geq 0 \right), \]

we deduce that $(u + \sigma_m)/(u + \sigma_1) < \sigma_m/\sigma_1$; i.e., $\kappa(uI + A) < \kappa(A)$.

Inequality (3.8) of Theorem 3.2 means that for any $u > 0$ the spectral condition number of the new iterative improvement (3.2) is less than the spectral condition number of the original problem (2.1) to be solved, that is the spectral condition number of Wilkinson’s iterative improvement (1.2) or (1.3). To illustrate this point, let us take Hilbert matrix $H_m$ as an example. The spectral condition numbers $\kappa(uI + H_m)$ change with parameter $u$; see Tables 1 and 2. From the tables, we can see the spectral condition numbers are improved efficiently.

Table 1. $\kappa(uI + H_m)$ with $u=10^{-5}$.

<table>
<thead>
<tr>
<th>$m$</th>
<th>10</th>
<th>20</th>
<th>50</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa(H_m)$</td>
<td>$1.6 \times 10^{13}$</td>
<td>$1.1 \times 10^{19}$</td>
<td>$1.0 \times 10^{19}$</td>
<td>$5.2 \times 10^{19}$</td>
</tr>
<tr>
<td>$\kappa(uI + H_m)$</td>
<td>$1.6 \times 10^6$</td>
<td>$1.9 \times 10^6$</td>
<td>$2.1 \times 10^6$</td>
<td>$2.2 \times 10^6$</td>
</tr>
</tbody>
</table>

Table 2. $\kappa(uI + H_m)$ with $u=10^{-4}$.

<table>
<thead>
<tr>
<th>$m$</th>
<th>10</th>
<th>20</th>
<th>50</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa(H_m)$</td>
<td>$1.6 \times 10^{13}$</td>
<td>$1.1 \times 10^{19}$</td>
<td>$1.0 \times 10^{19}$</td>
<td>$5.2 \times 10^{19}$</td>
</tr>
<tr>
<td>$\kappa(uI + H_m)$</td>
<td>$1.8 \times 10^4$</td>
<td>$1.9 \times 10^4$</td>
<td>$2.1 \times 10^4$</td>
<td>$2.2 \times 10^4$</td>
</tr>
</tbody>
</table>

As for the consistency, it is easy to see that, if iterative improvement (3.2) is convergent, then we have

\[ x = x + (uI + A)^{-1}(b - Ax) \]  

(3.10)

from (3.2). Consequently, linear system (3.10) and linear system (2.1) have the same solution $x^* = A^{-1}b$. That is to say, the iterative improvement (3.2) is consistent with system of linear equations (2.1).

The rate of convergence for the new iterative improvement (3.2) or (3.3) is changed with $u$. From (3.7), we can see that if $u$ is chosen as

\[ u \leq \sigma_1. \]
then we have
\[ \rho(M) \leq \frac{1}{2} \]
and the asymptotic rate of convergence is
\[ R(M) = -\ln \rho(M) \geq \ln 2. \]

On the other hand, from (3.9), if \( u \) is chosen too small, then the spectral condition number of the iterative matrix \( uI + A \) may not be improved significantly compared with Wilkinson's iterative improvement, for which the spectral condition number is just \( \kappa(A) \).

With regard to the estimate of error, we have the \textit{a priori} error estimate
\[ \|x_n - x^*\|_2 \leq \frac{\|M\|_2}{1 - \|M\|_2}\|x_1 - x_0\|_2 \]
and the \textit{a posteriori} error estimate
\[ \|x_n - x^*\|_2 \leq \frac{\|M\|_2}{1 - \|M\|_2}\|x_1 - x_0\|_2. \]

Remarks

**Remark 3.1.** For a nonnormal matrix \( A \) and the corresponding system of linear equations
\[ Ax = b, \tag{3.11} \]
we may consider
\[ \tilde{A}z = b, \tag{3.12} \]
where \( \tilde{A} = AA^T \), and
\[ x = A^Tz. \tag{3.13} \]
It is obvious that \( \tilde{A} \) is a positive definite normal matrix, provided \( \det(A) \neq 0 \).

**Remark 3.2.** For a given problem (2.1), the choice of the parameter \( u \) depends on a compromise between accuracy and stability. An efficient strategy to achieve this can be performed by the discrepancy principle (see [10,11], etc.).

### 4. NUMERICAL ILLUSTRATIONS

**Problem 1.** Let us consider a well-known ill-conditioned system of linear equations with \( m \times m \) Hilbert matrix \( H_m = (h_{ij})_{m \times m} \) of coefficients
\[ H_m x = b \tag{4.1} \]
where \( b_i = \sum_{k=1}^{m} h_{ik}, \ i = 1,2,\ldots,m. \) So the accuracy solution is \( [1,1,\ldots,1]^T \) and stopping criteria is \( \|x_{n+1} - x_n\|_2 < \text{eps}, \text{eps} = 0.5 \times 10^{-5}, \) and maximum iterative number of step is 500,000. The numerical results obtained by the new iterative improvement (3.2) are listed in Table 3.

| Table 3. Number of significant digits of computed solutions for Problem 1. |
|-----------------|---|---|---|---|
| \( m \)       | 12 | 20 | 50 | 90 |
| Wilkinson's Solver | No | No | No | No |
| Our Solver \((u = 10^{-5})\) | 6  | 6  | 5  | 5  |
Problem 2. The matrix of coefficients is the same as Problem 1. Now the difference is $b$ with $b_i = \sum_{k=1}^{m} h_{ik} \times k$, $i = 1, 2, \ldots, m$. Thus the corresponding accuracy solution is $[1, 2, \ldots, m]^T$. The computed results are listed in Table 4.

Table 4. Number of significant digits of computed solutions for Problem 2.

<table>
<thead>
<tr>
<th>$m$</th>
<th>12</th>
<th>20</th>
<th>50</th>
<th>90</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wilkinson’s Solver</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Our Solver ($u = 10^{-5}$)</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

Problem 3. As another example we use iterative improvement solver (3.2) with $u = 10^{-6}$ to deal with the ill-conditioned system of linear equations as follows:

$$Ax = b,$$

where $A = (a_{ij})_{90 \times 90}$, $b = [b_1, b_2, \ldots, b_{90}]^T$,

$$a_{ij} = \begin{cases} 1, & i \neq j, \\ 1 + p^2, & i = j, \end{cases} \quad i, j = 1, 2, \ldots, 90, \quad \text{and} \quad b_i = \sum_{k=1}^{90} a_{ik}, \quad i = 1, 2, \ldots, 90.$$

The theoretical solution of the problem is $[1, 1, \ldots, 1]^T$. Letting $p = 0.5 \times 10^{-5}$, the spectral condition number of $A$ is $\kappa(A) = (90 + p^2)/p^2 \approx 3.6 \times 10^{12}$ (see [12]). After seven steps, the new iterative improvement yields the approximate solution with six significant figures. However, after 100,000 steps, Wilkinson’s iterative method gives a computed solution with only two significant figures (see Table 5).

Table 5. Number of significant digits and iterative steps of computed solutions for Problem 3.

<table>
<thead>
<tr>
<th>Method</th>
<th>Iterative Steps</th>
<th>Number of Significant Digits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wilkinson’s Method</td>
<td>100,000</td>
<td>2</td>
</tr>
<tr>
<td>Our Method</td>
<td>7</td>
<td>6</td>
</tr>
</tbody>
</table>

All the numerical results in this paper are provided with double precision arithmetic.

5. Conclusion

Our intention in this paper is to establish a new iterative improvement of the solution for the ill-conditioned system of linear equations; to present our ideas in a particularly beneficial framework, we chose to concentrate on (3.2) in Section 3. The convergence of (3.2) and its corresponding dynamic system are analyzed in Sections 3 and 2. Some numerical experiments are provided for illustrating the new method.

REFERENCES