Journal of Computational Mathematics, Vol.24, No.3, 2006, 412–424.

ACCELERATION METHODS OF NONLINEAR ITERATION FOR NONLINEAR PARABOLIC EQUATIONS [*1)]

Guang-wei Yuan   Xu-deng Hang
(*Laboratory of Computational Physics, Institute of Applied Physics and Computational Mathematics, P.O.Box 8009, Beijing 100088, China*)

Dedicated to the 70th birthday of Professor Lin Qun

Abstract

This paper discusses the accelerating iterative methods for solving the implicit scheme of nonlinear parabolic equations. Two new nonlinear iterative methods named by the implicit-explicit quasi-Newton (IEQN) method and the derivative free implicit-explicit quasi-Newton (DFIEQN) method are introduced, in which the resulting linear equations from the linearization can preserve the parabolic characteristics of the original partial differential equations. It is proved that the iterative sequence of the iteration method can converge to the solution of the implicit scheme quadratically. Moreover, compared with the Jacobian Free Newton-Krylov (JFNK) method, the DFIEQN method has some advantages, e.g., its implementation is easy, and it gives a linear algebraic system with an explicit coefficient matrix, so that the linear (inner) iteration is not restricted to the Krylov method. Computational results by the IEQN, DFIEQN, JFNK and Picard iteration methods are presented in confirmation of the theory and comparison of the performance of these methods.

Mathematics subject classification: 65M06, 65M12.
Key words: Nonlinear parabolic equations, Difference scheme, Newton iterative methods.

1. Introduction

For solving the implicit scheme of nonlinear parabolic problems from various applications, iterative methods are used which adopt the inner-outer iteration mode. The inner iteration is the linear iterative methods for the linearized systems, and the outer cycle is the nonlinear iterative methods which will be discussed here. To a great extent the outer nonlinear iteration determines the accuracy and efficiency of the total solution procedure. In the energy conservative equation of the radiation hydrodynamics, the diffusion coefficients and the source term are nonlinear with respect to the temperature (the temperatures of radiation, ion or electron). During the construction of the linearization procedure, the key point is to preserve the characteristics of the original nonlinear parabolic equations so as to achieve high efficient solution. In [1]-[5], it is pointed out that the nonlinear convergence is tightly relevant to the selection of time step and the precision of solution. The efficient nonlinear iteration within one time step can speed up the convergence of the iteration solution greatly. So it is essential to find high efficient iterative methods in solving the nonlinear parabolic problems.

There are at least three reasons to prevent Newton methods applied in the nonlinear parabolic problems from some large scale scientific computations. The first is that the nonlinear

[*] Received March 1, 2006.
[1)] Supported by the The National Basic Research Program (No. 2005CB321703) and the National Natural Science Foundation of China (No.10476002, 60533020).

iteration methods with super-linear convergent order often have local convergent region. In this regard, a common approach is to reduce the time step to ensure the nonlinear iteration method convergent. Actually, Newton method is sensitive to the iterative initial value, and can be regarded as a measure of the nonlinearity. However, reducing frequently the time step will increase a lot of computational time. The second reason is that the Newton method may change the features of the original partial differential equations, which makes the iteration hard to be convergent. A iteration method of preserving the characteristics of the original PDEs during the iteration process is more valuable than the one that possesses such property only at the end of the iteration procedure. The iteration method of keeping the parabolic feature of the nonlinear parabolic equations not only ensures the efficiency of the computation, but also keeps the iteration solution to be positive (see [7] for detail). Keeping the positivity in the iterative procedure is the foundation of the correct simulation of the physical problem. The third reason is that the Newton iteration should form a Jacobian matrix, which is often time-consuming, and is even impossible for some applications. For this issue, some papers (e.g. [4]-[6]) suggest applying the JFNK (Jacobian Free Newton-Krylov) method to deal with such problems.

In this paper, we pay attention to the last two reasons due to their importance. The main objective of this paper is that two new nonlinear iteration methods, called as the implicit-explicit quasi-Newton (IEQN) method and the derivative free implicit-explicit quasi-Newton (DFIEQN) method, are proposed. In these methods we construct a iterative (linearized) difference scheme from the nonlinear implicit scheme, instead of simply applying the Newton method or JFNK method to the nonlinear algebraic system of equations. In other words, the device of IEQN and DFIEQN methods are based on the nonlinear implicit scheme for the nonlinear parabolic equations, and not on the corresponding nonlinear algebraic system of equations. Moreover the performance of the DFIEQN method is examined along with some existing iteration methods including the semi-implicit method (SI), the fully implicit Picard method (FIP), fully implicit partial Newton method (FIPN) and the JFNK (Jacobian Free Newton-Krylov) method. Like JFNK method, the DFIEQN method is derivative free. But, unlike JFNK method our DFIEQN method has the advantage of FIP, i.e., its implementation is simple, and it gives a linear algebraic system with an explicit coefficient matrix, so that the inner iteration is not restricted to be chosen as the Krylov method and it is more convenient and efficient to get a preconditioner. Moreover we will prove the DFIEQN method is convergent quadratically, while the SI, FIP and FIPN is convergent linearly (see [7]).

The paper is organized as follows. Some nonlinear iterative methods are constructed in following section 2. These include the known semi-implicit (SI) method, the fully implicit Picard (FIP) method, and the fully implicit partial Newton (FIPN) method. And then we describe the construction of the implicit-explicit quasi-Newton (IEQN) method and the derivative free implicit-explicit quasi-Newton (DFIEQN) method. In the section 3 some assumptions and auxiliary lemmas are introduced, and the main convergence theorems are stated. In the section 4, we study the convergence property of the constructed nonlinear iteration method, in particular we will prove the 2nd order convergence of the IEQN and DFIEQN methods. In the last section, numerical results are presented to show the performance of these methods.

2. Construction of the Iteration Sequences

2.1. The Problem and Some Notations

To present the idea of the construction of the nonlinear iteration, the following one dimensional nonlinear parabolic problem is considered for simplicity here

$$u_t - (A(x,t,u)u_x)_x = f(x,t,u), \;\; Q_T = \{0 < x < l, 0 < t \le T\} \tag{2.1}$$

$$u(x,0) = u^0(x), \;\; 0 \le x \le l \tag{2.2}$$

$$u(0,t) = u(l,t) = 0, \;\; 0 \le t \le T \tag{2.3}$$

where $A(x,t,u)$ and $f(x,t,u)$ are given functions of (x,t,u), $u^0(x)$ is a given function of x. For simplicity, we only consider the homogeneous boundary conditions (2.3) except in the section 5.

Divide Q_T by using parallel lines $x = x_j$ $(j = 0,1,\cdots,J)$ and $t = t^n$ $(n = 0,1,\cdots,N)$, where $x_j = jh$, $t^n = n\tau$, and $Jh = l, N\tau = T$, J and N are some positive integers, h and τ are the space and time step. For $0 \le n \le N$ denote the first order difference

$$\delta u^n_{j+\frac{1}{2}} = \frac{1}{h}\left(u^n_{j+1} - u^n_j\right), \quad (j = 0,1,\cdots,J-1)$$

and the second order difference

$$\delta^2 u^n_j = \frac{1}{h}\left(\delta u^n_{j+\frac{1}{2}} - \delta u^n_{j-\frac{1}{2}}\right), \quad (j = 1,\cdots,J-1).$$

For a discrete function $\{u_j | j = 0,1,\cdots,J\}$ (where $u_0 = u_J = 0$) define some discrete norms as follows

$$\|u_h\|_\infty = \max_{0\le j\le J} |u_j|, \quad \|\delta u_h\|_\infty = \max_{0\le j\le J-1} |\delta u_{j+\frac{1}{2}}|,$$

$$\|u_h\|_2^2 = \sum_{j=1}^{J-1} |u_j|^2 h, \quad \|\delta u_h\|_2^2 = \sum_{j=0}^{J-1} |\delta u_{j+\frac{1}{2}}|^2 h.$$

2.2. Fully Implicit Scheme (FIS)

A classical difference scheme for solving the problem (2.1)–(2.3) is the following implicit scheme

$$\frac{u^{n+1}_j - u^n_j}{\tau} = \frac{1}{h}\left(A^{n+1}_{j+\frac{1}{2}}\delta u^{n+1}_{j+\frac{1}{2}} - A^{n+1}_{j-\frac{1}{2}}\delta u^{n+1}_{j-\frac{1}{2}}\right) + f^{n+1}_j, \quad 1 \le j \le J-1, 0 \le n \le N-1 \tag{2.4}$$

$$u^0_j = u^0(x_j) \quad 0 \le j \le J \tag{2.5}$$

$$u^{n+1}_0 = u^{n+1}_J = 0, \quad 0 \le n \le N-1 \tag{2.6}$$

where

$$A^{n+1}_{j+\frac{1}{2}} = A(x_{j+\frac{1}{2}}, t^{n+1}, u^{n+1}_{j+\frac{1}{2}}), \quad f^{n+1}_j = f(x_j, t^{n+1}, u^{n+1}_j),$$

and $u^{n+1}_{j+\frac{1}{2}} = \frac{1}{2}(u^{n+1}_{j+1} + u^{n+1}_j), x_{j+\frac{1}{2}} = \frac{1}{2}(x_{j+1} + x_j)$. The basic properties of the implicit scheme for the nonlinear parabolic equations have been studied in [8].

2.3. Semi-Implicit scheme (SI)

In this paper the semi-implicit scheme is referred to linearize the nonlinear equations (2.4) by constructing the diffusion coefficients and the source term by the last iterative values on the previous time level, i.e.,

$$\frac{u^{n+1}_j - u^n_j}{\tau} = \frac{1}{h}\left(A^n_{j+\frac{1}{2}}\delta u^{n+1}_{j+\frac{1}{2}} - A^n_{j-\frac{1}{2}}\delta u^{n+1}_{j-\frac{1}{2}}\right) + f^n_j, \quad 1 \le j \le J-1, 0 \le n \le N-1 \tag{2.7}$$

with the initial and boundary condition (2.5)–(2.6). When $f = f(u) = u^k$, f^n_j is often replaced by $u^{n+1}_j(u^n_j)^{k-1}$ or $(ku^{n+1}_j - (k-1)u^n_j)(u^n_j)^{k-1}$ or some other forms.

2.4. Fully Implicit Picard Iteration (FIP, usually called as the simple iteration)

For a fixed non-negative integer n $(0 \le n \le N-1)$, define a sequence of discrete functions $\{\overset{(s)}{u}{}^{n+1}_j | j = 0,1,\cdots,J\}$ $(s = 0,1,\cdots)$ by the following way: $\{\overset{(s+1)}{u}{}^{n+1}_j | j = 0,1,\cdots,J\}$ is obtained by the solution of the following linear system of equations

$$\frac{\overset{(s+1)}{u}{}^{n+1}_j - u^n_j}{\tau} = \frac{1}{h}\left(\overset{(s)}{A}{}^{n+1}_{j+\frac{1}{2}}\delta \overset{(s+1)}{u}{}^{n+1}_{j+\frac{1}{2}} - \overset{(s)}{A}{}^{n+1}_{j-\frac{1}{2}}\delta \overset{(s+1)}{u}{}^{n+1}_{j-\frac{1}{2}}\right) + \overset{(s)}{f}{}^{n+1}_j \tag{2.8}$$

$$1 \le j \le J-1, 0 \le n \le N-1$$

$$\overset{(0)}{u}{}_j^{n+1} = u_j^n \quad 0 \le j \le J \tag{2.9}$$

$$\overset{(s+1)}{u}{}_0^{n+1} = \overset{(s+1)}{u}{}_J^{n+1} = 0, \quad 0 \le n \le N \tag{2.10}$$

where s is the number of the iterations, and

$$\overset{(s)}{A}{}_{j+\frac{1}{2}}^{n+1} = A(x_{j+\frac{1}{2}}, t^{n+1}, \overset{(s)}{u}{}_{j+\frac{1}{2}}^{n+1}), \quad \overset{(s)}{f}{}_j^{n+1} = f(x_j, t^{n+1}, \overset{(s)}{u}{}_j^{n+1}).$$

When $f = f(u) = u^k$, $\overset{(s)}{f}{}_j^{n+1}$ can be taken as $\overset{(s+1)}{u}{}_j^{n+1}(\overset{(s)}{u}{}_j^{n+1})^{k-1}$ or $(k\overset{(s+1)}{u}{}_j^{n+1} - (k-1)\overset{(s)}{u}{}_j^{n+1})(\overset{(s)}{u}{}_j^{n+1})^{k-1}$.

For the diffusion equation of non-divergence type, the convergence of FIP method and the contraction of the iteration sequence are proved in [9]. For more general iterative difference schemes, it has been proved in [10] that the discrete solution converges to the solution of the diffusion equation if the time step and space step tend to zero.

2.5. Fully Implicit Partial Newton Method (FIPN)

The following FIPN method differs from the FIP method in the way of linearization of the nonlinear source term. It is defined by modifying the equation (2.8) into

$$\frac{\overset{(s+1)}{u}{}_j^{n+1} - u_j^n}{\tau} = \frac{1}{h}\left(\overset{(s)}{A}{}_{j+\frac{1}{2}}^{n+1}\delta\overset{(s+1)}{u}{}_{j+\frac{1}{2}}^{n+1} - \overset{(s)}{A}{}_{j-\frac{1}{2}}^{n+1}\delta\overset{(s+1)}{u}{}_{j-\frac{1}{2}}^{n+1}\right) + \overset{(s)}{f}{}_j^{n+1} + \overset{(s)}{f'}{}_j^{n+1}(\overset{(s+1)}{u}{}_j^{n+1} - \overset{(s)}{u}{}_j^{n+1}), \tag{2.11}$$

$$(1 \le j \le J-1, \quad 0 \le n \le N-1)$$

where $\overset{(s)}{f'}{}_j^{n+1} = f'(x_j, t^{n+1}, \overset{(s)}{u}{}_j^{n+1})$. The motivation to construct the method is that the diffusive term and the source term should be managed separately for solving some practical problems. For example in the two-dimensional 3-temperature simulation of radiation hydrodynamics, the two nonlinear terms $(A(u)u_x)_x$ and $f(u)$ represent different physical peculiarities. In many cases the energy exchange term $f(u)$ has a strong stiffness, and the implicit method is required to resolve it, and a superlinear iterative method is needed to make it converge nonlinearity as fast as possible. When $A(u)$ is independent of u, the usual Newton method is reduced to be (2.11). Furthermore, in the Jacobian-Free method $\overset{(s)}{f'}{}_j^{n+1}v$ can be approximated by

$$\frac{f(x_j, t^{n+1}, \overset{(s)}{u}{}_j^{n+1} + \varepsilon v) - \overset{(s)}{f}{}_j^{n+1}}{\varepsilon}$$

where $\varepsilon > 0$ is a small parameter.

2.6. Implicit–Explicit Quasi–Newton Method (IEQN)

When the Newton iteration method and JFNK iteration method are used to solve the nonlinear implicit scheme (2.4)–(2.6), people always consider the system (2.4)–(2.6) as a nonlinear algebraic system os equations $\mathbf{F}(\mathbf{u}) = 0$, and then define the iterative sequence $\{\overset{(s)}{\mathbf{u}} : s = 0, 1, \cdots\}$ by

$$\overset{(s)}{\mathbf{J}}(\overset{(s+1)}{\mathbf{u}} - \overset{(s)}{\mathbf{u}}) + \mathbf{F}(\overset{(s)}{\mathbf{u}}) = 0,$$

where $\mathbf{J}$ is the Jacobian matrix for Newton method or replacing derivatives with difference quotients for JFNK method. As we know, the Newton-Krylov method and JFNK method (e.g., see [4]-[6]) never manage the term $(A(u)u_x)_x$ and $f(u)$ separately. When solving the nonlinear algebraic system equations by the Newton method, the parabolic properties of the original partial differential equations was seldom considered for the discretized and linearized system of equations. Following the argument method used in this paper, we can prove that FIPN and FIP methods are of the first order convergence. The convergence order is not higher than one since the term $(A(u)u_x)_x$ is linearized simply by $(A(\overset{(s)}{u})\overset{(s+1)}{u}_x)_x$. So how to construct a nonlinear iteration method which is of superlinear convergence as well as preserving the parabolic property is very interesting. The following IEQN method will serve as an example.

To show clearly the mechanism of the IEQN method, we consider the case $f(u) \equiv 0$ for simplicity. To emphasize the character of the method, we omit the superscripts $n+1$ and subscripts j if no confusion occurs. We still use δ to stand for the first order of difference and s stands for the number of iteration.

If $A(u)$ is linearized by one order Taylor expansion, that is to replace $\overset{(s)}{A}{}^{n+1}_{j+\frac{1}{2}}$ by $\overset{(s)}{A}{}^{n+1}_{j+\frac{1}{2}} + \overset{(s)}{A'}{}^{n+1}_{j}(\overset{(s+1)}{u}{}^{n+1}_{j} - \overset{(s)}{u}{}^{n+1}_{j})$, we get the following system

$$\frac{\overset{(s+1)}{u}_j - u^n_j}{\tau} = \delta\left((\overset{(s)}{A} + \overset{(s)}{A'}(\overset{(s+1)}{u} - \overset{(s)}{u}))\delta\overset{(s+1)}{u}\right), \quad 1 \le j \le J-1, 0 \le n \le N-1.$$

where $\overset{(s)}{A'}{}^{n+1}_{j+\frac{1}{2}} = f'(x_{j+\frac{1}{2}}, t^{n+1}, \overset{(s)}{u}{}^{n+1}_{j+\frac{1}{2}})$. Unfortunately, the system above is not linear, and can not be solved directly by linear solver. Under the assumption that $\overset{(s)}{A} + \overset{(s)}{A'}(\overset{(s+1)}{u} - \overset{(s)}{u})$ is always positive, we can prove the 2nd order convergence of the above iteration. But we can not ensure that $\overset{(s)}{A} + \overset{(s)}{A'}(\overset{(s+1)}{u} - \overset{(s)}{u})$ be always positive in practical computation, unless the time steps are chosen to be very small. Then we propose the following IEQN method:

$$\frac{\overset{(s+1)}{u}_j - u^n_j}{\tau} = \delta\left(\overset{(s)}{A}\delta\overset{(s+1)}{u}\right) + \delta\left(\overset{(s)}{A'}(\overset{(s+1)}{u} - \overset{(s)}{u}))\delta\overset{(s)}{u}\right), \quad 1 \le j \le J-1, 0 \le n \le N-1. \tag{2.12}$$

This method can be obtained from FIP (2.8) by adding the $\delta\left(\overset{(s)}{A'}(\overset{(s+1)}{u} - \overset{(s)}{u}))\delta\overset{(s)}{u}\right)$, which is linear and one order difference with respect to $\overset{(s+1)}{u}$. The resulting equations can preserve the parabolic property, and can be solved quickly since the quadratic convergence will be proved in the next section.

The IEQN method result from the nonlinear implicit scheme (2.4) instead of the corresponding nonlinear algebraic system of equations, i.e., it is different from the Newton method in that it gives a direct approach to form Jacobian matrix. Furthermore it enlighten us to propose the following derivative free implicit–explicit quasi–Newton method (DFIEQN), which can be applied in some scientific and engineering computations.

2.7. Derivative Free Implicit–Explicit Quasi–Newton Method (DFIEQN)

Now we describe the construction of DFIEQN iteration method. In the IEQN method (2.12) we replace the derivative $\overset{(s)}{\mathbf{A}'}$ with the difference quotient

$$\overset{(s)}{\mathbf{A}'} \approx \overset{(s)}{\mathbf{A}'_{\boldsymbol{\varepsilon}}} \approx \left\{\overset{(s)}{A'_\varepsilon}\right\} = \left\{\frac{1}{\overset{(s)}{\varepsilon}{}^{n+1}_{j+\frac{1}{2}}}\left(A(\overset{(s)}{u}{}^{n+1}_{j+\frac{1}{2}} + \overset{(s)}{\varepsilon}{}^{n+1}_{j+\frac{1}{2}}) - A(\overset{(s)}{u}{}^{n+1}_{j+\frac{1}{2}})\right)\right\},$$

where $\overset{(s)}{\varepsilon}{}^{n+1}_{j+\frac{1}{2}} > 0$ are small parameters. Then the DFIEQN iteration method is constructed as follows

$$\frac{\overset{(s+1)}{u}_j - u_j^n}{\tau} = \delta\left(\overset{(s)}{A}\delta\overset{(s+1)}{u}\right) + \delta\left(\overset{(s)}{A'_\varepsilon}(\overset{(s+1)}{u} - \overset{(s)}{u})\delta\overset{(s)}{u}\right), \quad 1 \le j \le J-1, 0 \le n \le N-1 \quad (2.13)$$

with the boundary condition (2.10). Note that the DFIEQN method (2.13) has same principal part as the FIP (2.8), but they differ in that the FIP method is of linear convergence while DFIEQN is of quadratic convergence for the parameters ε chosen properly. Moreover a linear system of equations is formed for DFIEQN (2.13) while it is not for JFNK.

3. Assumptions, Auxiliary Lemmas and Main Theorems

3.1. Assumptions

Introduce the assumptions:

(H1) $A \in C^2(R)$, and there exists a constant $\sigma > 0$ such that $A(v) \ge \sigma$ $(\forall v \in R)$.

(H2) The nonlinear implicit scheme (2.4)–(2.6) has one and only one solution $\{u_j^{n+1} | 1 \le j \le J-1, 0 \le n \le N-1\}$, and there exists a constant $M > 0$ such that

$$\max_{0 \le n \le N-1} \|\delta u_h^{n+1}\|_\infty \le M.$$

(H3) Let $\overset{(0)}{w}{}_j^{n+1} = u_j^n - u_j^{n+1}$. Assume $\tau \le \tau_0$ and $\|\delta \overset{(0)}{w}{}_h^{n+1}\|_2^2 < c_0$, where $\tau_0 > 0$ and $c_0 > 0$ are small constants to be determined in the following section.

3.2. Auxiliary Lemmas

We need some lemmas (see [8]) as follows:

Lemma 3.1. *(The discrete Green formula) Let u_j and v_j be the discrete function defined on $\{x_j | j = 0, 1, \cdots, J\}$, then*

$$\sum_{j=0}^{J-1} u_j(v_{j+1} - v_j) = -\sum_{j=1}^{J-1}(u_j - u_{j-1})v_j - u_0 v_0 + u_{J-1}v_J.$$

Lemma 3.2. *(The discrete Sobolev inequality) For any discrete function $u_h = \{u_j | j = 0, 1, \cdots, J\}$ $(Jh = l)$, the following assertions hold.*

(i) For all $\varepsilon > 0$, there are

$$\|u_h\|_\infty^2 \le \varepsilon \|\delta u_h\|_2^2 + \frac{C}{\varepsilon}\|u_h\|_2^2,$$

where C is a constant depending on l, and independent of ε, h and u_h;

(ii) If $u_0 = u_J = 0$, then

$$\|u_h\|_2 \le l\|\delta u_h\|_2, \quad \|u_h\|_\infty \le \|\delta u_h\|_2^{\frac{1}{2}} \|u_h\|_2^{\frac{1}{2}};$$

(iii) There exist a constant C independent of h and l, such that

$$\|\delta u_h\|_2 \le C\left(\|u_h\|_2^{\frac{1}{2}}\|\delta^2 u_h\|_2^{\frac{1}{2}} + l^{-1}\|u_h\|_2\right).$$

In this paper, C refers to a positive constant independent of h, τ, and s (the number of iteration), and may be different in different place.

3.3. Main Theorems

Let u_j^{n+1} be the solution of (2.4)–(2.6) with $f_j^{n+1} \equiv 0$ for simplicity.

Theorem 1. *If the assumptions (H1)–(H3) hold and τ is small enough, then for the sequence $\{\overset{(s)}{u}{}_j^{n+1}\}$ defined by the IEQN method (2.12) with (2.9) and (2.10) there hold*

$$\lim_{s\to\infty}\left(\|\overset{(s+1)}{w}{}_h^{n+1}\|_2 + \|\delta\overset{(s+1)}{w}{}_h^{n+1}\|_2\right) = 0, \quad \overline{\lim_{s\to\infty}} \frac{\|\delta\overset{(s+1)}{w}{}_h^{n+1}\|_2}{\|\delta\overset{(s)}{w}{}_h^{n+1}\|_2^2} \le C,$$

where $\overset{(s)}{w}{}_j^{n+1} = \overset{(s)}{u}{}_j^{n+1} - u_j^{n+1}$.

Theorem 2. *Assume (H1)–(H3) hold and* τ *is small enough. If the sequence* $\{\overset{(s)}{u}{}_j^{n+1}\}$ *is defined by the DFIEQN method (2.13) with (2.9) and (2.10), and let* $\overset{(s)}{w}{}_j^{n+1} = \overset{(s)}{u}{}_j^{n+1} - u_j^{n+1}$, *then we have*

(i) when

$$\lim_{s\to\infty} \max_{0\le j\le J-1} |\overset{(s)}{\varepsilon}{}_{j+\frac{1}{2}}^{n+1}| = 0$$

is satisfied then there holds

$$\lim_{s\to\infty} \frac{\|\delta \overset{(s+1)}{w}{}_h^{n+1}\|_2}{\|\delta \overset{(s)}{w}{}_h^{n+1}\|_2} = 0;$$

(ii) when

$$\max_{0\le j\le J-1} |\overset{(s)}{\varepsilon}{}_{j+\frac{1}{2}}^{n+1}| = O(\|\overset{(s)}{w}\|_2),$$

is satisfied then

$$\overline{\lim_{s\to\infty}} \frac{\|\delta \overset{(s+1)}{w}{}_h^{n+1}\|_2}{\|\delta \overset{(s)}{w}{}_h^{n+1}\|_2^2} \le C.$$

4. Proof of Iteration Convergence

In this section, we will prove the 2nd order convergence of IEQN and DFIEQN iteration methods, i.e., Theorem 1 and 2. For simplicity the discrete indices j and $n+1$ will be omitted if there is no confusion.

4.1. IEQN

For $\{\overset{(s)}{u}{}_j^{n+1} | j = 0, 1, \cdots, J\}$ $(s = 0, 1, \cdots)$ defined by the IEQN method (2.12) with (2.9)–(2.10) denote $\overset{(s)}{w}{}_j^{n+1} = \overset{(s)}{u}{}_j^{n+1} - u_j^{n+1}$. From (2.12) and (2.4) we have

$$\frac{\overset{(s+1)}{w}{}_j}{\tau} = \delta\left(\overset{(s)}{A}\delta\overset{(s+1)}{w}\right) + \delta\left[(\overset{(s)}{A} - A + \overset{(s)}{A'}(\overset{(s+1)}{u} - \overset{(s)}{u}))\delta u + \overset{(s)}{A'}(\overset{(s+1)}{u} - \overset{(s)}{u}))\delta(\overset{(s)}{u} - u)\right], \quad (4.1)$$

where the term in [ ] at the right of (4.1) is equal to the following

$$\begin{aligned}&\left(\overset{(s)}{A^*}(\overset{(s)}{u} - u) + \overset{(s)}{A'}(\overset{(s+1)}{u} - u - (\overset{(s)}{u} - u))\right)\delta u + \overset{(s)}{A'}(\overset{(s+1)}{u} - u - (\overset{(s)}{u} - u))\delta(\overset{(s)}{u} - u)\\ &= \left(\overset{(s)}{A''}(\overset{(s)}{w})^2 + \overset{(s)}{A'}\overset{(s+1)}{w}\right)\delta u + \overset{(s)}{A'}(\overset{(s+1)}{w} - \overset{(s)}{w})\delta\overset{(s)}{w},\end{aligned}$$

where the following abbreviations are used

$$\begin{aligned}\overset{(s)}{A} - A &\equiv A(\overset{(s)}{u}{}_{j+\frac{1}{2}}^{n+1}) - A(u_{j+\frac{1}{2}}^{n+1}) = \int_0^1 A'(r\overset{(s)}{u}{}_{j+\frac{1}{2}}^{n+1} + (1-r)u_{j+\frac{1}{2}}^{n+1})dr(\overset{(s)}{u}{}_{j+\frac{1}{2}}^{n+1} - u_{j+\frac{1}{2}}^{n+1})\\ &\equiv \overset{(s)}{A^*}(\overset{(s)}{u}{}_{j+\frac{1}{2}}^{n+1} - u_{j+\frac{1}{2}}^{n+1}), \ \text{ where } \overset{(s)}{A^*} = \int_0^1 A'(r\overset{(s)}{w}{}_{j+\frac{1}{2}}^{n+1} + u_{j+\frac{1}{2}}^{n+1})dr,\end{aligned}$$

$$\overset{(s)}{A'} = A'(\overset{(s)}{u}{}_{j+\frac{1}{2}}^{n+1})), \quad \overset{(s)}{A''} = \int_0^1 dr\int_0^1 d\bar{r}\left[A''(\bar{r}(r-1)\overset{(s)}{w}{}_{j+\frac{1}{2}}^{n+1} + \overset{(s)}{u}{}_{j+\frac{1}{2}}^{n+1})(r-1)\right].$$

At the moment we assume

$$\left|A'(r\overset{(s)}{w}{}^{n+1}_{j+\frac12}+u^{n+1}_{j+\frac12})\right|\le C',\quad \left|A''(\bar r(r-1)\overset{(s)}{w}{}^{n+1}_{j+\frac12}+\overset{(s)}{u}{}^{n+1}_{j+\frac12})\right|\le C''.$$

Here C' and C'' are positive constants to be determined later.

Multiplying (4.1) by $\overset{(s+1)}{w}{}_j h$, and summing up the products for $j=1,\cdots,J-1$, we obtain

$$\begin{aligned}&\frac{1}{\tau}\|\overset{(s+1)}{w}\|_2^2+\sigma\|\delta\overset{(s+1)}{w}\|_2^2\\
\le{}& C\sum_{j=0}^{J-1}\left[(|\overset{(s)}{w}|^2+|\overset{(s+1)}{w}|^2)|\delta u|+|\overset{(s+1)}{w}-\overset{(s)}{w}||\delta\overset{(s)}{w}|\right]|\delta\overset{(s+1)}{w}|h\\
\le{}& \frac{\sigma}{2}\|\delta\overset{(s+1)}{w}\|_2^2+C\left(\|\overset{(s)}{w}\|_\infty^4+\|\overset{(s+1)}{w}\|_2^2+\|\overset{(s+1)}{w}-\overset{(s)}{w}\|_\infty^2\|\delta\overset{(s)}{w}\|_2^2\right),\end{aligned}$$

where $C\ge 1$ is a constant dependent on M. So we have

$$\begin{aligned}&\frac{1-C\tau}{\tau}\|\overset{(s+1)}{w}\|_2^2+\frac{\sigma}{2}\|\delta\overset{(s+1)}{w}\|_2^2\\
\le{}& C\left[\|\delta\overset{(s)}{w}\|_2^4+\left(\|\overset{(s+1)}{w}\|_\infty^2+\|\overset{(s)}{w}\|_\infty^2\right)\|\overset{(s)}{w}\|_2^2\right]\\
\le{}& C\left(\|\delta\overset{(s)}{w}\|_2^4+\|\overset{(s+1)}{w}\|_2\|\delta\overset{(s+1)}{w}\|_2\|\overset{(s)}{w}\|_2^2\right)\\
\le{}& \frac{\sigma}{4}\|\delta\overset{(s+1)}{w}\|_2^2+C\|\delta\overset{(s)}{w}\|_2^4\left(1+\|\overset{(s+1)}{w}\|_2^2\right).\end{aligned}$$

Then, when $\frac{1-C\tau}{\tau}\ge 2$, it can be deduced to

$$2\|\overset{(s+1)}{w}\|_2^2+\|\delta\overset{(s+1)}{w}\|_2^2\le C\|\delta\overset{(s)}{w}\|_2^4\left(1+\|\overset{(s+1)}{w}\|_2^2\right).$$

Assume by induction $C\|\delta\overset{(s)}{w}\|_2^2<1$. Then

$$2\|\overset{(s+1)}{w}\|_2^2+\|\delta\overset{(s+1)}{w}\|_2^2\le\frac{1}{C}\left(C\|\delta\overset{(s)}{w}\|_2^2\right)^2\left(1+\|\overset{(s+1)}{w}\|_2^2\right)<\frac{1}{C}\left(1+\|\overset{(s+1)}{w}\|_2^2\right).$$

It follows $C\|\delta\overset{(s+1)}{w}\|_2^2<1$. So we conclude that, if $C\|\delta\overset{(0)}{w}\|_2^2<1$, then $C\|\delta\overset{(s)}{w}\|_2^2<1$ $(\forall s\ge 0)$. Therefore

$$\|\overset{(s+1)}{w}\|_2^2+\|\delta\overset{(s+1)}{w}\|_2^2\le C\|\delta\overset{(s)}{w}\|_2^4\le\cdots\le\frac{1}{C}(C\|\delta\overset{(0)}{w}\|_2^2)^{2^{s+1}}.$$

Under the conditions of Theorem 1 it is showed the following results hold

$$\lim_{s\to\infty}\left(\|\overset{(s+1)}{w}\|_2+\|\delta\overset{(s+1)}{w}\|_2\right)=0,\quad \overline{\lim_{s\to\infty}}\frac{\|\delta\overset{(s+1)}{w}\|_2}{\|\delta\overset{(s)}{w}\|_2^2}\le C^{\frac12}.$$

The proof of Theorem 1 is completed.

4.2. DFIEQN

Now we give the proof of Theorem 2. Since it is similar to the argument of the above subsection, we only state the difference between them. Now (4.1) should be changed into

$$\begin{aligned}\frac{\overset{(s+1)}{w}{}_j}{\tau}={}&\delta\left(\overset{(s)}{A}\delta\overset{(s+1)}{w}\right)+\delta\left[(\overset{(s)}{A}-A+\overset{(s)}{A'}(\overset{(s+1)}{u}-\overset{(s)}{u}))\delta u+\overset{(s)}{A'_\varepsilon}(\overset{(s+1)}{u}-\overset{(s)}{u})\delta(\overset{(s)}{u}-u)\right]\\
&+\delta\left[(\overset{(s)}{A'_\varepsilon}-\overset{(s)}{A'})(\overset{(s+1)}{u}-\overset{(s)}{u}))\delta u\right],\end{aligned}\tag{4.2}$$

where

$$\overset{(s)}{A'_\varepsilon} = \frac{1}{\overset{(s)}{\varepsilon}{}^{n+1}_{j+\frac{1}{2}}} \left(A(\overset{(s)}{u}{}^{n+1}_{j+\frac{1}{2}} + \overset{(s)}{\varepsilon}{}^{n+1}_{j+\frac{1}{2}}) - A(\overset{(s)}{u}{}^{n+1}_{j+\frac{1}{2}}) \right) = \int_0^1 A'(\overset{(s)}{u}{}^{n+1}_{j+\frac{1}{2}} + r\overset{(s)}{\varepsilon}{}^{n+1}_{j+\frac{1}{2}})dr,$$

$$\overset{(s)}{A'_\varepsilon} - \overset{(s)}{A'} = \int_0^1 \left(A'(\overset{(s)}{u}{}^{n+1}_{j+\frac{1}{2}} + r\overset{(s)}{\varepsilon}{}^{n+1}_{j+\frac{1}{2}}) - A'(\overset{(s)}{u}{}^{n+1}_{j+\frac{1}{2}}) \right) dr$$

$$= \int_0^1 dr \int_0^1 d\bar{r} \left(A''(\bar{r}(\overset{(s)}{u}{}^{n+1}_{j+\frac{1}{2}} + r\overset{(s)}{\varepsilon}{}^{n+1}_{j+\frac{1}{2}}) + (1-\bar{r})\overset{(s)}{u}{}^{n+1}_{j+\frac{1}{2}})r\overset{(s)}{\varepsilon}{}^{n+1}_{j+\frac{1}{2}} \right),$$

$$\overset{(s)}{A^*} - \overset{(s)}{A'_\varepsilon} = \int_0^1 \left(A'(r\overset{(s)}{w}{}^{n+1}_{j+\frac{1}{2}} + u^{n+1}_{j+\frac{1}{2}}) - \overset{(s)}{A'_\varepsilon}(\overset{(s)}{u}{}^{n+1}_{j+\frac{1}{2}} + r\overset{(s)}{\varepsilon}{}^{n+1}_{j+\frac{1}{2}}) \right) dr,$$

$$= \int_0^1 dr \int_0^1 d\bar{r} \left[A''(\bar{r}(r-1)\overset{(s)}{w}{}^{n+1}_{j+\frac{1}{2}} + \overset{(s)}{u}{}^{n+1}_{j+\frac{1}{2}}) \left((r-1)\overset{(s)}{w}{}^{n+1}_{j+\frac{1}{2}} + r\overset{(s)}{\varepsilon}{}^{n+1}_{j+\frac{1}{2}} \right) \right].$$

From the proof in **4.1** we can see that as long as

$$\lim_{s\to\infty} \max_{0\le j\le J-1} |\overset{(s)}{\varepsilon}{}^{n+1}_{j+\frac{1}{2}}| = 0, \tag{4.3}$$

then we can obtain

$$\lim_{s\to\infty} \frac{\|\delta \overset{(s+1)}{w}{}^{n+1}_{h}\|_2}{\|\delta \overset{(s)}{w}{}^{n+1}_{h}\|_2} = 0. \tag{4.4}$$

Furthermore, if

$$\max_{0\le j\le J-1} |\overset{(s)}{\varepsilon}{}^{n+1}_{j+\frac{1}{2}}| = O(\|\overset{(s)}{w}{}^{n+1}_{h}\|_2), \tag{4.5}$$

then it follows

$$\overline{\lim_{s\to\infty}} \frac{\|\delta \overset{(s+1)}{w}{}^{n+1}_{h}\|_2}{\|\delta \overset{(s)}{w}{}^{n+1}_{h}\|_2^2} \le C^{\frac{1}{2}}. \tag{4.6}$$

Therefore we have proved the Theorem 2.

5. Numerical Experiments

To compare the performance of the IEQN and DFIEQN with other methods, we give some numerical experiments on a model problem. Consider a model problem of the following form

$$\frac{\partial u}{\partial t} - \frac{\partial}{\partial x}\left(\kappa \frac{\partial u}{\partial x}\right) - \frac{\partial}{\partial y}\left(\kappa \frac{\partial u}{\partial y}\right) = S - u^4, \tag{5.1}$$
$$for \quad (x,y) \in [0,10]\times[0,10], t\in[0,\infty)$$

where

$$\kappa = z^3u^3; z = z_0 = 1, u(x,y,0) = 1.d-5$$

with the boundary conditions

$$\begin{aligned}
&\tfrac{1}{4}u - \tfrac{1}{6}F = f,\ \text{for } x=0, 0<y<8;\\
&u_x = 0,\ \text{for } x=0, 8<y<10; u_x=0,\ \text{for } x=10, 8<y<10\\
&u_y = 0,\ \text{for } 0<x<10, y=10; u_y=0,\ \text{for } 0<x<10, y=0\\
&\tfrac{1}{4}u + \tfrac{1}{6}F = 0,\ \text{for } x=10, 0<y<8\\
&where \quad F = z^3u^3u_x
\end{aligned}$$

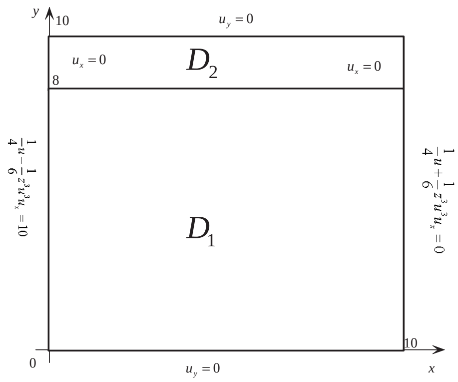

Figure 1: The computational domain with boundary conditions

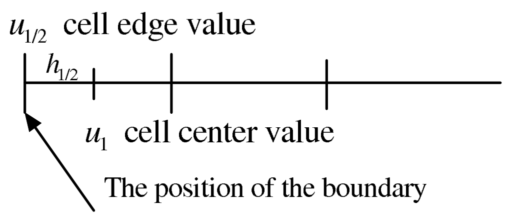

Figure 2: The Robin boundary condition

The detail of the computational model is depicted in Figure 1.

For the boundary conditions, the following discretization is adopted. For simplicity, we may consider a one dimensional case, the notions are listed in Figure 2. Since the boundary condition is valid on the boundary, then it is equivalent to

$$\frac{1}{4}(u_{1/2} - u_1) + \frac{1}{4}u_1 - \frac{1}{6}F_{1/2} = f$$

where $u_{1/2}$ denotes the value on the boundary.

And we approximate the term $\frac{1}{4}(u_{1/2} - u_1)$ by

$$\frac{1}{4}(u_{1/2} - u_1) \approx -\frac{h_{1/2}}{4}u_x = -\frac{h_{1/2}}{4z^3u_{1/2}^3}F_{1/2}$$

which is substituted into the discretized form of the boundary condition, and obtain

$$-(\frac{h_{1/2}}{4z^3u_{1/2}^3} + \frac{1}{6})F_{1/2} = f - \frac{1}{4}u_1.$$

And replace the $u_{1/2}$ by u_1 on the left side of the equation, we obtain the discretized approximation to $F_{1/2}$ as

$$F_{1/2} = \frac{f - \frac{1}{4}u_1}{-(\frac{h_{1/2}}{4z^3u_{1/2}^3} + \frac{1}{6})} \tag{5.2}$$

The equation (5.1) is integrated on a cell and discretized by finite volume methods, in which the normal flux $F_{\vec{n}}$ can be replaced by (5.2) if the side of the cell is on the boundary of Ω.

In the present numerical experiments, we select

$$S = 1.D0; f = 10, z_0 = 1$$

and in each of the above two subdomains, the mesh size is 50×50. The problem is solved by different nonlinear solvers different time steps.

This model shows the conduction of heat flow which is introduced from the left side, and flows out on the right side. The process is shown by the following pictures.

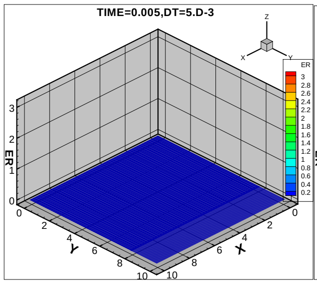

Figure 3. Energy distribution

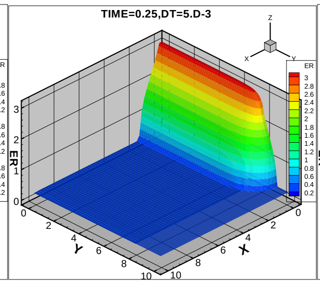

Figure 4. Energy distribution

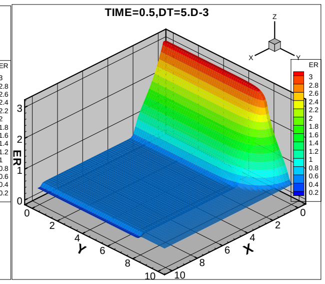

Figure 5. Energy distribution

GMRES(10) is used as the linear solver, with the preconditioner ILUTP$(10, 1.d-10)$. For the JFNK test examples, FGMRES(10) is used instead of the standard GMRES(10). The program is written in FORTRAN, and run on a windows system. Two results of different time steps are presented below.

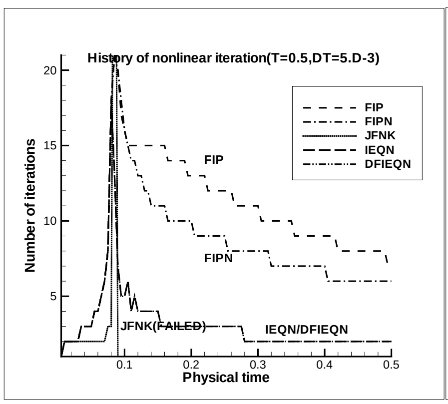

Figure 6. Iterative history of ex. 1

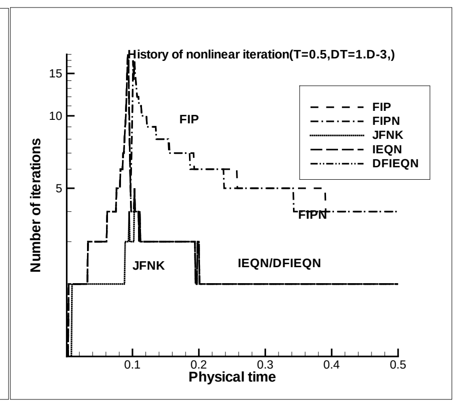

Figure 7. Iterative history of ex. 2

The total time cost of each iterative method are listed in the following tablets.

METHODS	FIP	FIPN	JFNK	IEQN	DFIEQN	IEQN (high precision)
TIME(s)	51.8	37.0	-	16.4	16.5	20.1

Table 1. The time costs of the five methods (DT=$\tau = 5.d-3$)

METHODS	FIP	FIPN	JFNK	IEQN	DFIEQN	IEQN (high precision)
TIME(s)	113.6	110.8	208.8	62.9	63.1	90.0

Table 2. The time costs of the five methods (DT=$\tau = 1.d-3$)

Since the exact solution of the problem is unavailable, we solve the problem with a higher precision. It is converged if the residual of each nonlinear iteration is less than $1.d-11$. It is solved by IEQN, and the time consumption is listed in the last column of the above two tablets. And the solution is used as the approximate exact solution. Then we can get the following results. The relative error is the maximum norm of the relative error vectors.

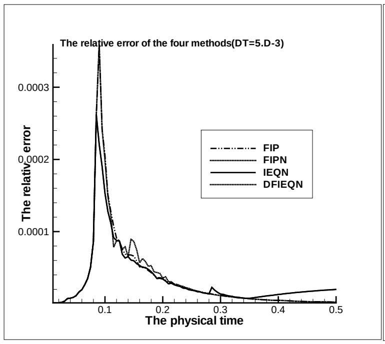

Figure 8. Relative error of ex. 1

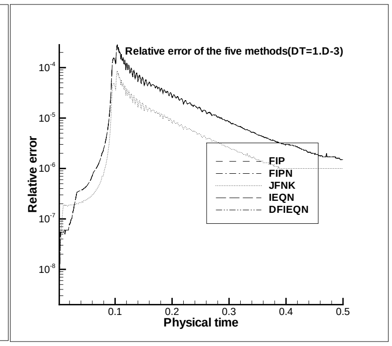

Figure 9. Relative error of ex. 2

The first experiment adopts time step $5.d-3$. The nonlinear convergence judgement is the residual norm is less than $1.d-6$. For this time step, we notice that JFNK failed, while the others succeeded. JFNK seems very sensitive to the time step. However, if JFNK converges, it converges fastest. And from the time and number of iterations, we find IEQN as well as DFIEQN the best. On most time, IEQN (DFIEQN) converges the fastest and costs the least time.

For the second experiment, we use a smaller time step. The nonlinear convergence judgement is the residual norm is less than $1.d-6$. In this experiment, JFNK converges the fastest. IEQN in most place cost as many iterations as JFNK. But JFNK costs the most time. This is because the equation is nonlinear, and the discretization costs much more time than a matrix vector multiplication operation. For this point, though JFNK may be the fastest in convergence, it is not the fastest in time of solution.

A third result is that DFIEQN and IEQN performs almost the same. In the two examples, we approximate the derivative of the diffusion coefficients by

$$\frac{\partial \kappa}{\partial u} \approx \frac{\kappa(u+\varepsilon)-\kappa(u)}{\varepsilon}, where \quad \varepsilon = 1.d-6$$

and the results show that it almost doesn't change the convergence property of the IEQN method. And if we solve the practical problem, whose diffusion coefficients are often provided by libraries, then DFIEQN can be used instead of IEQN to get good performance.

From the relative error of the problem (Fig 8 and Fig 9), we find for the large time step problem, the relative error of IEQN and DFIEQN are better in the former part of the simulation ($T < 0.25$) and a little larger in the latter part (where the IEQN only need two nonlinear iteration to converge). And in the small time step problem, the relative error of IEQN and DFIEQN is as good as FIP and FIPN, while worse than JFNK. But it takes still less time for IEQN to reach the same precision as JFNK. It takes about 90 seconds for the second example.

In all, from the numerical examples, we find that IEQN (DFIEQN) are shown to yield good convergence rate, stability and efficiency.

References

[1] V.A. Mousseau, D.A. Knoll, W.J. Rider, Physics-Based Preconditioning and the Newton–Krylov

Method for Non-equilibrium Radiation Diffusion, *Journal of Computational Physics,* **160** (2000), 743–765.

[2] D.A. Knoll, L. Chacon, L.G. Margolin, V.A. Mousseau, On balanced approximations for time integration of multiple time scale systems, *Journal of Computational Physics,* **185** (2003), 583–611.

[3] D.A. Knoll, J. Morel, L.G. Margolin, M. Shashkov, Physically Motivated Discretization Methods–A Strategy for Increased Predictiveness, *Los Alamos Science,* **29** (2005), 188–212.

[4] D.A. Knoll, W.J. Rider, G.L. Olson, An efficient nonlinear solution method for non-equilibrium radiation diffusion, *J. Quant. Spectrosc. Radiat. Transfer,* **63** (1999), 15–29.

[5] D.A. Knoll, W.J. Rider, G.L. Olson, Nonlinear convergence, accuracy, and time step control in non-equilibrium radiation diffusion, *J. Quant. Spectrosc. Radiat. Transfer,* **70** (2001), 25–36.

[6] D.A. Knoll, D.E. Keyes, Jacobian–free Newton–Krylov methods: a survey of approaches and applications, *Journal of Computational Physics,* **193** (2004), 357–397.

[7] G.W. Yuan, Acceleration techniques of iterative solution methods for nonlinear parabolic equations, Annual Report of LCP, Beijing, 366–413, 2004.

[8] Y.L. Zhou, Discrete Functional Analysis and Its Application to Finite Difference Method, Inter. Pub., Beijing, 1991.

[9] Y.L. Zhou, L.J. Shen, G.W. Yuan, Finite Difference Method of First Boundary Problem for quasi-linear Parabolic System (IV) Convergence of Iteration, Science in China (A), **40**:5 (1997), 469–474.

[10] Y. L. Zhou, L. J. Shen, G. W. Yuan, Finite Difference Method of First Boundary Problem for Quasilinear Parabolic System (V) Convergence of Iterative Schemes, Science in China (A), **40**:11, 1152–1157.