Available online at www.sciencedirect.com
ScienceDirect
Journal of Computational and Applied Mathematics 221 (2008) 76–97
JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS
www.elsevier.com/locate/cam

Fourth-order compact finite difference method for fourth-order nonlinear elliptic boundary value problems

Yuan-Ming Wang[a,b,*], Ben-Yu Guo[c,b]

[a] *Department of Mathematics, East China Normal University, Shanghai 200062, PR China*
[b] *Division of Computational Science, E-Institute of Shanghai Universities, Shanghai Normal University, Shanghai 200234, PR China*
[c] *Department of Mathematics, Shanghai Normal University, Shanghai 200234, PR China*

Received 12 February 2007; received in revised form 24 September 2007

Abstract

A compact finite difference method with non-isotropic mesh is proposed for a two-dimensional fourth-order nonlinear elliptic boundary value problem. The existence and uniqueness of its solutions are investigated by the method of upper and lower solutions, without any requirement of the monotonicity of the nonlinear term. Three monotone and convergent iterations are provided for resolving the resulting discrete systems efficiently. The convergence and the fourth-order accuracy of the proposed method are proved. Numerical results demonstrate the high efficiency and advantages of this new approach.

MSC: 65N06; 65N22; 35J40

Keywords: Fourth-order nonlinear elliptic boundary value problem; Compact finite difference method; Fourth-order accuracy; Monotone iterations

1. Introduction

Boundary value problems of fourth-order nonlinear differential equations have been paid considerable attention. Most of the existing results in this field were devoted to the existence, uniqueness and multiplicity of solutions for the following one-dimensional two-point boundary value problem:

$$\begin{cases} u'''' = f(x, u, u''), & 0 < x < 1, \\ u(0) = u(1) = 0, & u''(0) = u''(1) = 0, \end{cases} \tag{1.1}$$

where $f(x, u, v)$ is, in general, a nonlinear function of u and v (cf. [1,2,8,13,18,27]). This equation describes the static deflection of an elastic bending beam (with hinged ends) under a possible nonlinear loading (cf. [10,22]). It also describes the steady state of a prototype problem for phase transitions in condensed matter systems (cf. [9,23]).

Recently, much attention has been also paid to certain fourth-order elliptic boundary value problems in multiple dimensions (cf. [7,14–17,21]). In this paper, we focus on the following two-dimensional fourth-order nonlinear elliptic boundary value problem:

* Corresponding author at: Department of Mathematics, East China Normal University, Shanghai 200062, PR China.
E-mail address: ymwang@math.ecnu.edu.cn (Y.-M. Wang).

0377-0427/$ - see front matter
doi:10.1016/j.cam.2007.10.007

$$\begin{cases} \Delta(k(x,y)\Delta u) = f(x,y,u,\Delta u), & (x,y)\in\Omega, \\ u(x,y) = g(x,y),\ \Delta u(x,y) = g^*(x,y), & (x,y)\in\partial\Omega, \end{cases} \tag{1.2}$$

where Ω is a rectangular domain or a union of rectangular domains, and Δ is the Laplacian operator. Assume that $f(x,y,u,v)$, $g(x,y)$ and $g^*(x,y)$ are continuous functions, $k(x,y)$ is a strictly positive C^2-function on $\overline{\Omega}\equiv\Omega\cup\partial\Omega$. As a physical interpretation, (1.2) governs the static deflection of a plate under a lateral loading. In this case, $k(x,y)$ is the stiffness of plate, $g(x,y)$ and $g^*(x,y)$ are the possible boundary sources, and $f(x,y,u,\Delta u)$ stands for the loading function, which may depend on the deflection and the curvature of the plate (cf. [22]).

The existing work on (1.2) was also concerned with the existence, uniqueness, and multiplicity of solutions (cf. [7,14,15,21]). On the other hand, some authors developed numerical methods, but mainly for the linear cases (cf. [4,6,12]). Recently, a finite difference method was proposed in [16,17], for a class of nonlinear fourth-order elliptic boundary value problems including (1.2), by using the standard second-order finite difference approximation and three monotone iterations for resolving the resulting discrete systems with the quasimonotone function $f(\cdot,\cdot,u,v)$.

As we know, the standard second-order finite difference method fails to approach underlying problems effectively, unless a large number of mesh points are used. In other words, we have to take small mesh sizes for obtaining desirable accuracy. Thereby, in order to obtain satisfactory numerical results with reasonable computational cost, a reasonable approach is to develop a higher-order compact finite difference method, which not only provides accurate numerical results and saves computational work, but also is easier to treat boundary conditions. However, the usual fourth-order compact finite difference method is only available for second-order linear problems, see [19,20,28] and the references therein.

In this work, we propose a compact finite difference method for the fourth-order nonlinear problem (1.2), and three iteration processes for resolving the resulting nonlinear discrete system. These algorithms have several advantages. Firstly, the proposed compact finite difference scheme possesses the fourth-order accuracy and thus provides precise numerical results. Next, we do not need to deal with boundary conditions approximately. Moreover, the scheme preserves the monotonicity as the approximated continuous version and so the numerical solution fits the exact solution properly. Furthermore, the suggested iterations converge monotonically with geometric rates, which save a lot of work. Finally, these processes do not require any monotonicity of function $f(\cdot,\cdot,u,v)$ and so essentially enlarge their applications. Besides, we generalize in this work the method of upper and lower solutions to fourth-order elliptic problems in multiple dimensions, which is applicable to many other nonlinear elliptic problems of high order.

The outline of the paper is as follows. In the next section, we reformulate (1.2) to a coupled system of nonlinear second-order equations and then construct the compact finite difference scheme. In Section 3, we deal with the existence and uniqueness of its solutions, with nonmonotone function $f(\cdot,\cdot,u,v)$, by using the method of upper and lower solutions, which also leads to a monotone and convergent iteration for solving the difference scheme. In Section 4, we introduce two additional monotone iterations (block Jacobi and block Gauss–Seidel iterations), which simplify actual computation essentially and save a lot of computational cost. In Section 5, we prove the fourth-order accuracy of numerical solution. In Section 6, we present some numerical results demonstrating the monotone convergence of iterations and the fourth-order accuracy of numerical solution. We also compare our new method with other finite difference methods and show its advantages. The final section is for some concluding remarks.

2. Compact finite difference scheme

To design a proper finite difference scheme, we set $v=-k\Delta u$ and reform problem (1.2) to the coupled system of second-order elliptic equations:

$$\begin{cases} -\Delta u = v/k, \quad -\Delta v = f(x,y,u,-v/k), & (x,y)\in\Omega, \\ u(x,y) = g^{(1)}(x,y), \quad v(x,y) = g^{(2)}(x,y), & (x,y)\in\partial\Omega, \end{cases} \tag{2.1}$$

where $g^{(1)}(x,y)=g(x,y)$ and $g^{(2)}(x,y)=-k(x,y)g^*(x,y)$. Obviously, u is a solution of (1.2), if and only if (u,v) is a solution of (2.1).

For simplicity, we consider a rectangular domain $\Omega=[0,L_x]\times[0,L_y]$. We partition Ω with non-isotropic uniform mesh sizes h_x and h_y in the x and y directions, respectively. The integers $N_x=L_x/h_x$ and $N_y=L_y/h_y$. The mesh points $(x_i,y_j)=(ih_x,jh_y)$, $0\le i\le N_x$, $0\le j\le N_y$. For convenience, we also use the index pair (i,j) to represent

the mesh point (x_i, y_j), and the notations:

$$u_{i,j} = u(x_i, y_j), \qquad v_{i,j} = v(x_i, y_j), \qquad k_{i,j} = k(x_i, y_j),$$
$$g^{(1)}_{i,j} = g^{(1)}(x_i, y_j), \qquad g^{(2)}_{i,j} = g^{(2)}(x_i, y_j), \qquad F_{i,j}(u_{i,j}, v_{i,j}) = f(x_i, y_j, u_{i,j}, -v_{i,j}/k_{i,j}).$$

Our compact method for (1.2) is based on the alternative system (2.1). To do this, we set

$$\delta_x^2 u_{i,j} = h_x^{-2}(u_{i+1,j} - 2u_{i,j} + u_{i-1,j}), \qquad \delta_y^2 u_{i,j} = h_y^{-2}(u_{i,j+1} - 2u_{i,j} + u_{i,j-1}),$$

and introduce the finite difference operators

$$\bar{\delta}_\alpha^2 u_{i,j} = \left(1 + \frac{h_\alpha^2}{12}\delta_\alpha^2\right) u_{i,j}, \quad \alpha = x, y. \tag{2.2}$$

According to the Numerov's formula (cf. [3]),

$$\delta_\alpha^2 u_{i,j} = \bar{\delta}_\alpha^2 (u_{\alpha\alpha})_{i,j} + O(h_\alpha^4), \quad \alpha = x, y, \tag{2.3}$$

or symbolically,

$$\bar{\delta}_\alpha^{-2}\delta_\alpha^2 u_{i,j} = (u_{\alpha\alpha})_{i,j} + O(h_\alpha^4), \quad \alpha = x, y, \tag{2.4}$$

where $\bar{\delta}_\alpha^{-2} = (\bar{\delta}_\alpha^2)^{-1}$ denotes the inverse of $\bar{\delta}_\alpha^2$.

We now apply the above fourth-order compact approximations to the second-order derivatives involved in (2.1). This yields symbolically that

$$\begin{cases} -\left(\bar{\delta}_x^{-2}\delta_x^2 + \bar{\delta}_y^{-2}\delta_y^2\right) u_{i,j} = (v_{i,j}/k_{i,j}) + O(h^4), \\ -\left(\bar{\delta}_x^{-2}\delta_x^2 + \bar{\delta}_y^{-2}\delta_y^2\right) v_{i,j} = F_{i,j}(u_{i,j}, v_{i,j}) + O(h^4), \end{cases} \tag{2.5}$$

where $O(h^4)$ denotes the truncated term of the order $O(h_x^4 + h_y^4)$. Multiplying the above equations by the finite difference operator $\bar{\delta}_x^2\bar{\delta}_y^2$, we have

$$\begin{cases} -\left(\bar{\delta}_y^2\delta_x^2 + \bar{\delta}_x^2\delta_y^2\right) u_{i,j} = \bar{\delta}_x^2\bar{\delta}_y^2(v_{i,j}/k_{i,j}) + O(h^4), \\ -\left(\bar{\delta}_y^2\delta_x^2 + \bar{\delta}_x^2\delta_y^2\right) v_{i,j} = \bar{\delta}_x^2\bar{\delta}_y^2 F_{i,j}(u_{i,j}, v_{i,j}) + O(h^4). \end{cases} \tag{2.6}$$

After dropping the $O(h^4)$ terms, we derive a finite difference scheme as follows,

$$\begin{cases} -\left(\bar{\delta}_y^2\delta_x^2 + \bar{\delta}_x^2\delta_y^2\right) u^h_{i,j} = \bar{\delta}_x^2\bar{\delta}_y^2(v^h_{i,j}/k_{i,j}), \quad (i,j) \in \Omega, \\ -\left(\bar{\delta}_y^2\delta_x^2 + \bar{\delta}_x^2\delta_y^2\right) v^h_{i,j} = \bar{\delta}_x^2\bar{\delta}_y^2 F_{i,j}(u^h_{i,j}, v^h_{i,j}), \quad (i,j) \in \Omega, \\ u^h_{i,j} = g^{(1)}_{i,j}, \qquad v^h_{i,j} = g^{(2)}_{i,j}, \quad (i,j) \in \partial\Omega, \end{cases} \tag{2.7}$$

where $u^h_{i,j}$ and $v^h_{i,j}$ represent the approximations to u and v at the point (i, j), respectively. Furthermore, let $\sigma = h_x/h_y$ be the ratio of mesh sizes, and introduce the operators $\mathcal{L}_h = -6h_x^2(\bar{\delta}_y^2\delta_x^2 + \bar{\delta}_x^2\delta_y^2)$ and $\mathcal{P}_h = 6h_x^2\bar{\delta}_x^2\bar{\delta}_y^2$. Then a direct calculation shows that

$$\begin{aligned} \mathcal{L}_h u^h_{i,j} &= 10(1+\sigma^2)u^h_{i,j} + (\sigma^2 - 5)(u^h_{i+1,j} + u^h_{i-1,j}) + (1 - 5\sigma^2)(u^h_{i,j+1} + u^h_{i,j-1}) \\ &\quad - (1+\sigma^2)(u^h_{i+1,j+1} + u^h_{i+1,j-1} + u^h_{i-1,j+1} + u^h_{i-1,j-1})/2, \\ \mathcal{P}_h u^h_{i,j} &= 25h_x^2 u^h_{i,j}/6 + 5h_x^2(u^h_{i+1,j} + u^h_{i-1,j} + u^h_{i,j+1} + u^h_{i,j-1})/12 \\ &\quad + h_x^2(u^h_{i+1,j+1} + u^h_{i+1,j-1} + u^h_{i-1,j+1} + u^h_{i-1,j-1})/24. \end{aligned} \tag{2.8}$$

Accordingly, we can rewrite (2.7) as the following alternative form,

$$\begin{cases} \mathcal{L}_h u^h_{i,j} = \mathcal{P}_h(v^h_{i,j}/k_{i,j}), & (i,j) \in \Omega, \\ \mathcal{L}_h v^h_{i,j} = \mathcal{P}_h F_{i,j}(u^h_{i,j}, v^h_{i,j}), & (i,j) \in \Omega, \\ u^h_{i,j} = g^{(1)}_{i,j}, \quad v^h_{i,j} = g^{(2)}_{i,j}, & (i,j) \in \partial\Omega. \end{cases} \tag{2.9}$$

The operator $\mathcal{L}_h$ preserves the same maximum principle as the corresponding continuous operator, stated below.

Lemma 2.1. *Let* $1/5 < \sigma^2 < 5$. *If* $\mathcal{L}_h u^h_{i,j} \le 0$ *for all* $(i,j) \in \Omega$, *then*

$$\max_{(i,j)\in\Omega} u^h_{i,j} \le \max_{(i,j)\in\partial\Omega} u^h_{i,j}. \tag{2.10}$$

Proof. We suppose that for some $(i_0, j_0) \in \overline{\Omega}$, $u^h_{i_0,j_0} = \max_{(i,j)\in\overline{\Omega}} u^h_{i,j} \equiv M_{\overline{\Omega}}$. If $(i_0, j_0) \in \partial\Omega$, then (2.10) follows immediately. Otherwise, we have from (2.8) that $\mathcal{L}_h u^h_{i_0,j_0} \ge 0$. This with the condition $\mathcal{L}_h u^h_{i,j} \le 0$ implies $\mathcal{L}_h u^h_{i_0,j_0} = 0$. This leads to $M_{\overline{\Omega}}$ that is attained also at all points which are the connected neighbors of (i_0, j_0). The same argument is valid at each of these points. Finally, all $u^h_{i,j}$ take the same value $M_{\overline{\Omega}}$. This proves (2.10). □

The above lemma leads to the following result which plays an important role in the forthcoming discussions. Its proof will be given in Appendix A.

Lemma 2.2. *Let* $1/5 < \sigma^2 < 5$, *and let* $u^h_{i,j}$ *be a discrete function defined on* $\overline{\Omega} = \Omega \cup \partial\Omega$. *Then*

$$\max_{(i,j)\in\Omega} |u^h_{i,j}| \le \max_{(i,j)\in\partial\Omega} |u^h_{i,j}| + \max_{(i,j)\in\Omega} |\mathcal{L}_h u^h_{i,j}|/(12h_x^2). \tag{2.11}$$

For analyzing the system (2.9), it is more convenient to consider its matrix form. To do this, we order the mesh points lexicographically. More precisely, we first arrange them from the left to the right in the x direction and then from the bottom to the top in the y direction. Corresponding to this ordering, we define the following $(N_x - 1)$-dimensional column vectors:

$$\begin{aligned} &U_{h,j} = (u^h_{1,j}, u^h_{2,j}, \ldots, u^h_{N_x-1,j})^T, \qquad V_{h,j} = (v^h_{1,j}, v^h_{2,j}, \ldots, v^h_{N_x-1,j})^T, \\ &F_j(U_{h,j}, V_{h,j}) = (F_{1,j}(u^h_{1,j}, v^h_{1,j}), \ldots, F_{N_x-1,j}(u^h_{N_x-1,j}, v^h_{N_x-1,j}))^T, \quad j = 1, 2, \ldots, N_y - 1. \end{aligned} \tag{2.12}$$

We also define the following $(N_x - 1)$-order diagonal and symmetric tridiagonal matrices:

$$\begin{aligned} &K_j = \mathrm{diag}(k^{-1}_{1,j}, k^{-1}_{2,j}, \ldots, k^{-1}_{N_x-1,j}), \quad j = 1, 2, \ldots, N_y - 1, \\ &A_0 = \mathrm{tridiag}(\sigma^2 - 5, 10(1+\sigma^2), \sigma^2 - 5), \qquad A_1 = \mathrm{tridiag}((1+\sigma^2)/2, 5\sigma^2 - 1, (1+\sigma^2)/2), \\ &B_0 = \mathrm{tridiag}(5/12, 25/6, 5/12), \qquad B_1 = \mathrm{tridiag}(1/24, 5/12, 1/24). \end{aligned} \tag{2.13}$$

Then the system (2.9) can be expressed in the matrix form as

$$\begin{cases} -A_1 U_{h,j-1} + A_0 U_{h,j} - A_1 U_{h,j+1} = h_x^2 \left(B_1 K_{j-1} V_{h,j-1} + B_0 K_j V_{h,j} + B_1 K_{j+1} V_{h,j+1}\right) + G^{(1)}_j, \\ -A_1 V_{h,j-1} + A_0 V_{h,j} - A_1 V_{h,j+1} = h_x^2 \left(B_1 F_{j-1}(U_{h,j-1}, V_{h,j-1}) + B_0 F_j(U_{h,j}, V_{h,j})\right. \\ \qquad \left. + B_1 F_{j+1}(U_{h,j+1}, V_{h,j+1})\right) + G^{(2)}_j, \quad j = 1, 2, \ldots, N_y - 1, \end{cases} \tag{2.14}$$

where $U_{h,0} = V_{h,0} = U_{h,N_y} = V_{h,N_y} = F_0(U_{h,0}, V_{h,0}) = F_{N_y}(U_{h,N_y}, V_{h,N_y}) = 0$, and $G^{(i)}_j$ are the $(N_x - 1)$-dimensional vectors associated with the boundary functions.

We can rewrite (2.14) in a more compact form. For this purpose, we set $\mathcal{N} = (N_x - 1) \times (N_y - 1)$, and define the $\mathcal{N}$-dimensional vectors:

$$\begin{aligned} &U_h = (U_{h,1}, U_{h,2}, \ldots, U_{h,N_y-1})^T, \qquad V_h = (V_{h,1}, V_{h,2}, \ldots, V_{h,N_y-1})^T, \\ &G^{(i)} = (G_1^{(i)}, G_2^{(i)}, \ldots, G_{N_y-1}^{(i)})^T \quad (i = 1, 2), \\ &F(U_h, V_h) = (F_1(U_{h,1}, V_{h,1}), \ldots, F_{N_y-1}(U_{h,N_y-1}, V_{h,N_y-1}))^T. \end{aligned} \tag{2.15}$$

We also introduce the $\mathcal{N}$-order block matrices A, B and K as

$$A = \text{tridiag}(-A_1, A_0, -A_1), \qquad B = \text{tridiag}(B_1, B_0, B_1), \qquad K = \text{diag}(K_1, K_2, \ldots, K_{N_y-1}). \tag{2.16}$$

Then, (2.14) reads

$$\begin{cases} AU_h = h_x^2 BKV_h + G^{(1)}, \\ AV_h = h_x^2 BF(U_h, V_h) + G^{(2)}. \end{cases} \tag{2.17}$$

If all entries of matrix S are positive (or nonnegative), then we say that S is positive (or nonnegative), also denoted by $S > 0$ (or $S \geq 0$) for simplicity. We define positive (or nonnegative) vectors similarly.

The matrices A and B have the following properties.

Lemma 2.3. *Let $1/5 < \sigma^2 < 5$, and λ_0 be the smallest eigenvalue of the generalized eigenvalue problem*

$$A\Phi = \lambda h_x^2 B\Phi. \tag{2.18}$$

Then λ_0 is real and positive. Moreover, its corresponding eigenvector could be chosen as a positive vector.

Proof. Since $1/5 < \sigma^2 < 5$, the matrix A is a real, symmetric and irreducibly diagonally dominant matrix with nonpositive off-diagonal entries. Thus, we have from Corollaries 1 and 2 of [24, p. 85] that $A^{-1} > 0$ and A is positive definite. On the other hand, along with Gerschgorin circle theorem, a direct calculation shows that the matrix B is also positive definite. Therefore, all eigenvalues of problem (2.18), including λ_0, are real and positive (cf. [11], pp. 176–177). Obviously, $1/\lambda_0$ is the largest eigenvalue of the matrix $h_x^2 A^{-1} B$, and $\rho(h_x^2 A^{-1} B) = 1/\lambda_0$ ($\rho(\cdot)$ denotes the spectral radius of the corresponding matrix). Since $h_x^2 A^{-1} B$ is a positive matrix, we know from Perron–Frobenius theorem (cf. [24], p. 30) that the eigenvector corresponding to its largest eigenvalue $1/\lambda_0$ may be chosen as a positive vector. □

Lemma 2.4. *Let M be a given constant. We have that*

(i) *if*

$$1/5 < \sigma^2 < 5, \qquad 0 \leq Mh_x^2 < 12 \min\{5 - \sigma^2, 5\sigma^2 - 1\}/5, \tag{2.19}$$

then $A_1 - Mh_x^2 B_1 \geq 0$, and the inverses $(A + Mh_x^2 B)^{-1}$ and $(A_0 + Mh_x^2 B_0)^{-1}$ exist and are positive;

(ii) *if*

$$1/5 < \sigma^2 < 5, \qquad -\lambda_0 < M < 0, \tag{2.20}$$

where $\lambda_0 > 0$ is the smallest eigenvalue of generalized eigenvalue problem (2.18), then the inverse $(A + Mh_x^2 B)^{-1}$ exists and is positive.

Proof. By (2.13) and (2.16),

$$\begin{aligned} &A + Mh_x^2 B = \text{tridiag}(-A_1 + Mh_x^2 B_1, A_0 + Mh_x^2 B_0, -A_1 + Mh_x^2 B_1), \\ &A_0 + Mh_x^2 B_0 = \text{tridiag}(\sigma^2 - 5 + 5Mh_x^2/12, 10(1 + \sigma^2) + 25Mh_x^2/6, \sigma^2 - 5 + 5Mh_x^2/12), \\ &-A_1 + Mh_x^2 B_1 = \text{tridiag}(-(1 + \sigma^2)/2 + Mh_x^2/24, 1 - 5\sigma^2 + 5Mh_x^2/12, -(1 + \sigma^2)/2 + Mh_x^2/24). \end{aligned}$$

The condition (2.19) ensures that

$$\begin{aligned} &\sigma^2 - 5 + 5Mh_x^2/12 < 0, \qquad 10(1 + \sigma^2) + 25Mh_x^2/6 > 0, \\ &-(1 + \sigma^2)/2 + Mh_x^2/24 < 0, \qquad 1 - 5\sigma^2 + 5Mh_x^2/12 < 0. \end{aligned}$$

Thus, the matrices $A + Mh_x^2 B$ and $A_0 + Mh_x^2 B_0$ are irreducibly diagonally dominant. Moreover, $A_1 - Mh_x^2 B_1 \geq 0$. Furthermore, we have from Corollary 1 of [24, p. 85] (also see [5]) that the inverses $(A + Mh_x^2 B)^{-1}$ and $(A_0 + Mh_x^2 B_0)^{-1}$ exist and are positive. This ends the proof of conclusion (i).

Next, we know from the proof of Lemma 2.3 that $A^{-1} > 0$ and the spectral radius $\rho(h_x^2 A^{-1} B) = 1/\lambda_0$. Therefore by (2.20), $-M\rho(h_x^2 A^{-1} B) < 1$ which implies that the inverse $(I + Mh_x^2 A^{-1} B)^{-1}$ exists and is positive (see [24, p. 83]). Finally, the conclusion (ii) follows from $A^{-1} > 0$ and $A + Mh_x^2 B = A(I + Mh_x^2 A^{-1} B)$. □

Remark 2.1. If the mesh sizes are isotropic, i.e., $h = h_x = h_y$, then the scheme (2.9) or (2.17) has a simple form, while the condition (2.19) is reduced to $0 \leq Mh^2 < 48/5$. However, the non-isotropic mesh sizes are more preferable, if the approximated solutions change more rapidly in one direction than in another direction.

3. Qualitative analysis of the scheme

To investigate the existence and uniqueness of the solution of (2.17) and derive an efficient algorithm, we have to generalize the method of upper and lower solutions.

Definition 3.1. A pair of vectors $(\widetilde{U}_h, \widetilde{V}_h)$ and $(\widehat{U}_h, \widehat{V}_h)$ is called a pair of coupled upper and lower solutions of (2.17), if $(\widetilde{U}_h, \widetilde{V}_h) \geq (\widehat{U}_h, \widehat{V}_h)$ and

$$\begin{cases} A\widetilde{U}_h \geq h_x^2 BK\widetilde{V}_h + G^{(1)}, \\ A\widetilde{V}_h \geq h_x^2 BF(Z_h, \widetilde{V}_h) + G^{(2)}, \\ A\widehat{U}_h \leq h_x^2 BK\widehat{V}_h + G^{(1)}, \\ A\widehat{V}_h \leq h_x^2 BF(Z_h, \widehat{V}_h) + G^{(2)}, \qquad \text{for all } \widehat{U}_h \leq Z_h \leq \widetilde{U}_h. \end{cases} \tag{3.1}$$

Hereafter, the inequalities between vectors are in the componentwise sense.

Clearly, the above definition does not depend on any monotonicity of function F. For a given pair of coupled upper and lower solutions $(\widetilde{U}_h, \widetilde{V}_h)$ and $(\widehat{U}_h, \widehat{V}_h)$, we set

$$\begin{aligned} \mathcal{S} &= \left\{ (U_h, V_h) \in (\mathbf{R}^{\mathcal{N}})^2; (\widehat{U}_h, \widehat{V}_h) \leq (U_h, V_h) \leq (\widetilde{U}_h, \widetilde{V}_h) \right\}, \\ \mathcal{S}_{i,j} &= \left\{ (u_{i,j}, v_{i,j}) \in \mathbf{R}^2; (\widehat{u}_{i,j}^h, \widehat{v}_{i,j}^h) \leq (u_{i,j}, v_{i,j}) \leq (\widetilde{u}_{i,j}^h, \widetilde{v}_{i,j}^h) \right\}, \end{aligned} \tag{3.2}$$

where $\widehat{u}_{i,j}^h$, $\widehat{v}_{i,j}^h$, $\widetilde{u}_{i,j}^h$ and $\widetilde{v}_{i,j}^h$ stand for the components of $\widehat{U}_h$, $\widehat{V}_h$, $\widetilde{U}_h$ and $\widetilde{V}_h$, respectively.

In what follows, without any more explanation, we always denote a pair of coupled upper and lower solutions of (2.17) by $(\widetilde{U}_h, \widetilde{V}_h)$ and $(\widehat{U}_h, \widehat{V}_h)$.

We now make the following basic hypothesis on F:

(H) There exists a constant $M \geq 0$ such that

$$F(U_h, V_h) - F(U_h, V_h') \geq -M(V_h - V_h') \tag{3.3}$$

whenever $(\widehat{U}_h, \widehat{V}_h) \leq (U_h, V_h') \leq (U_h, V_h) \leq (\widetilde{U}_h, \widetilde{V}_h)$.

The existence of the constant M in (3.3) is trivial, if $F(U_h, V_h)$ is a C^1-function of (U_h, V_h) in $\mathcal{S}$. In fact, M may be taken as any nonnegative constant satisfying

$$M \geq \max \left\{ -\frac{\partial F_{i,j}}{\partial v}(u, v); (u, v) \in \mathcal{S}_{i,j}, (i, j) \in \Omega \right\}. \tag{3.4}$$

Theorem 3.1. *Let $(\widetilde{U}_h, \widetilde{V}_h)$ and $(\widehat{U}_h, \widehat{V}_h)$ be a pair of coupled upper and lower solutions of (2.17), and let hypothesis* (H) *and the condition (2.19) hold. Then problem (2.17) admits at least one solution (U_h^*, V_h^*) in $\mathcal{S}$.*

Proof. By Lemma 2.4, $(A + Mh_x^2 B)^{-1}$ and A^{-1} exist and are positive. Thus, for any $(U_h', V_h') \in \mathcal{S}$, the uncoupled linear problem

$$\begin{cases} AU_h = h_x^2 BKV_h' + G^{(1)}, \\ (A + Mh_x^2 B)V_h = h_x^2 B(MV_h' + F(U_h', V_h')) + G^{(2)} \end{cases} \tag{3.5}$$

has a unique solution (U_h, V_h) in $(\mathbf{R}^{\mathcal{N}})^2$. Now, we define the mapping $\mathcal{T} : \mathcal{S} \longrightarrow (\mathbf{R}^{\mathcal{N}})^2$ by

$$\mathcal{T}(U_h', V_h') = (U_h, V_h), \quad \forall (U_h', V_h') \in \mathcal{S}. \tag{3.6}$$

It follows from (3.1), (3.3) and (3.5) that for any $(U_h', V_h') \in \mathcal{S}$,

$$\begin{cases} A(\widetilde{U}_h - U_h) \geq h_x^2 BK(\widetilde{V}_h - V_h') \geq 0, \\ (A + Mh_x^2 B)(\widetilde{V}_h - V_h) \geq h_x^2 B(M(\widetilde{V}_h - V_h') + F(U_h', \widetilde{V}_h) - F(U_h', V_h')) \geq 0. \end{cases} \tag{3.7}$$

Because of $A^{-1} > 0$ and $(A + Mh_x^2 B)^{-1} > 0$, the above inequalities imply $(U_h, V_h) \leq (\widetilde{U}_h, \widetilde{V}_h)$. Similarly, we have $(U_h, V_h) \geq (\widehat{U}_h, \widehat{V}_h)$. Hence $\mathcal{T}$ maps $\mathcal{S}$ into itself. This with the continuity of F implies that $\mathcal{T}$ is a bounded continuous map on $\mathcal{S}$. Therefore, by virtue of the Brower's fixed point theorem, there exists (U_h^*, V_h^*) in $\mathcal{S}$ such that $\mathcal{T}(U_h^*, V_h^*) = (U_h^*, V_h^*)$, which is a solution of problem (2.17) in $\mathcal{S}$. □

By Theorem 3.1, (2.17) has at least one solution, provided that it possesses a pair of coupled upper and lower solutions, which also serve as the upper and lower bounds of this solution.

Next, we consider the uniqueness of solution by using a monotone iteration, which also improves the upper and lower bounds of the solution, step-by-step.

We construct two sequences $\{(\overline{U}_h^{(m)}, \overline{V}_h^{(m)})\}$ and $\{(\underline{U}_h^{(m)}, \underline{V}_h^{(m)})\}$, by the following Picard-type iteration:

$$\begin{cases} (\overline{U}_h^{(0)}, \overline{V}_h^{(0)}) = (\widetilde{U}_h, \widetilde{V}_h), \qquad (\underline{U}_h^{(0)}, \underline{V}_h^{(0)}) = (\widehat{U}_h, \widehat{V}_h), \\ A\overline{U}_h^{(m)} = h_x^2 BK\overline{V}_h^{(m-1)} + G^{(1)}, \quad m \geq 1, \\ (A + M^* h_x^2 B)\overline{V}_h^{(m)} = h_x^2 B(M^* \overline{V}_h^{(m-1)} + \max\limits_{Z_h \in \mathcal{S}^{(m-1)}} F(Z_h, \overline{V}_h^{(m-1)})) + G^{(2)}, \quad m \geq 1, \\ A\underline{U}_h^{(m)} = h_x^2 BK\underline{V}_h^{(m-1)} + G^{(1)}, \quad m \geq 1, \\ (A + M^* h_x^2 B)\underline{V}_h^{(m)} = h_x^2 B(M^* \underline{V}_h^{(m-1)} + \min\limits_{Z_h \in \mathcal{S}^{(m-1)}} F(Z_h, \underline{V}_h^{(m-1)})) + G^{(2)}, \quad m \geq 1, \end{cases} \tag{3.8}$$

where M^* is a nonnegative constant specified later, and

$$\mathcal{S}^{(m)} = \{Z_h \in \mathbf{R}^{\mathcal{N}}; \underline{U}_h^{(m)} \leq Z_h \leq \overline{U}_h^{(m)}\}. \tag{3.9}$$

In the above iteration, the maximum and the minimum of a vector function are in componentwise sense. The following lemma shows that these sequences are well-defined.

Lemma 3.1. *If hypothesis* (H) *and condition* (2.19) *hold, then the sequences defined by* (3.8) *and* (3.9) *with* $M^* = M$ *are well-defined, and for all* $m \geq 0$, $(\overline{U}_h^{(m)}, \overline{V}_h^{(m)}) \geq (\underline{U}_h^{(m)}, \underline{V}_h^{(m)})$.

Proof. Clearly, the set $\mathcal{S}^{(0)}$ is well-defined. Therefore, by the existence of A^{-1} and $(A + Mh_x^2 B)^{-1}$, the vectors $(\overline{U}_h^{(1)}, \overline{V}_h^{(1)})$ and $(\underline{U}_h^{(1)}, \underline{V}_h^{(1)})$ are determined uniquely. Further, by hypothesis (H),

$$M\overline{V}_h^{(0)} + \max_{Z_h \in \mathcal{S}^{(0)}} F(Z_h, \overline{V}_h^{(0)}) \geq M\underline{V}_h^{(0)} + \min_{Z_h \in \mathcal{S}^{(0)}} F(Z_h, \underline{V}_h^{(0)}).$$

Thus, we have from (3.8) with $m = 1$ that

$$A(\overline{U}_h^{(1)} - \underline{U}_h^{(1)}) \geq 0, \qquad (A + Mh_x^2 B)(\overline{V}_h^{(1)} - \underline{V}_h^{(1)}) \geq 0.$$

Then it follows from the positivity of A^{-1} and $(A + Mh_x^2 B)^{-1}$ that $(\overline{U}_h^{(1)}, \overline{V}_h^{(1)}) \geq (\underline{U}_h^{(1)}, \underline{V}_h^{(1)})$, while the set $\mathcal{S}^{(1)}$ is well-defined. Finally, the desired conclusion follows inductively. □

We next show that the sequences $\{(\overline{U}_h^{(m)}, \overline{V}_h^{(m)})\}$ and $\{(\underline{U}_h^{(m)}, \underline{V}_h^{(m)})\}$ converge monotonically to the limits $(\overline{U}_h, \overline{V}_h)$ and $(\underline{U}_h, \underline{V}_h)$ respectively, $(\overline{U}_h, \overline{V}_h) \geq (\underline{U}_h, \underline{V}_h)$ and

$$\begin{cases} A\overline{U}_h = h_x^2 BK\overline{V}_h + G^{(1)}, \\ A\overline{V}_h = h_x^2 B \max\limits_{\underline{U}_h \leq Z_h \leq \overline{U}_h} F(Z_h, \overline{V}_h) + G^{(2)}, \\ A\underline{U}_h = h_x^2 BK\underline{V}_h + G^{(1)}, \\ A\underline{V}_h = h_x^2 B \min\limits_{\underline{U}_h \leq Z_h \leq \overline{U}_h} F(Z_h, \underline{V}_h) + G^{(2)}. \end{cases} \tag{3.10}$$

Theorem 3.2. *Let the conditions in Lemma 3.1 hold. Then the sequences $\{(\overline{U}_h^{(m)}, \overline{V}_h^{(m)})\}$ and $\{(\underline{U}_h^{(m)}, \underline{V}_h^{(m)})\}$ given by (3.8) with $M^* = M$ converge monotonically to the limits $(\overline{U}_h, \overline{V}_h)$ and $(\underline{U}_h, \underline{V}_h)$, respectively. They satisfy (3.10), and for $m \geq 1$,*

$$(\underline{U}_h^{(m-1)}, \underline{V}_h^{(m-1)}) \leq (\underline{U}_h^{(m)}, \underline{V}_h^{(m)}) \leq (\underline{U}_h, \underline{V}_h) \leq (\overline{U}_h, \overline{V}_h) \leq (\overline{U}_h^{(m)}, \overline{V}_h^{(m)}) \leq (\overline{U}_h^{(m-1)}, \overline{V}_h^{(m-1)}). \tag{3.11}$$

Moreover, for any solution (U_h', V_h') of problem (2.17) in $\mathcal{S}$, we have $(\underline{U}_h, \underline{V}_h) \leq (U_h', V_h') \leq (\overline{U}_h, \overline{V}_h)$.

Proof. In fact, the second and the fourth inequalities in (3.1) are equivalent to

$$\begin{cases} (A + Mh_x^2 B)\widetilde{V}_h \geq h_x^2 B(M\widetilde{V}_h + \max\limits_{Z_h \in \mathcal{S}^{(0)}} F(Z_h, \widetilde{V}_h)) + G^{(2)}, \\ (A + Mh_x^2 B)\widehat{V}_h \leq h_x^2 B(M\widehat{V}_h + \min\limits_{Z_h \in \mathcal{S}^{(0)}} F(Z_h, \widehat{V}_h)) + G^{(2)}. \end{cases} \tag{3.12}$$

Thanks to (3.12), and the first and the third inequalities of (3.1), we derive from (3.8) with $M^* = M$ that

$$\begin{cases} A(\overline{U}_h^{(0)} - \overline{U}_h^{(1)}) \geq 0, & (A + Mh_x^2 B)(\overline{V}_h^{(0)} - \overline{V}_h^{(1)}) \geq 0, \\ A(\underline{U}_h^{(1)} - \underline{U}_h^{(0)}) \geq 0, & (A + Mh_x^2 B)(\underline{V}_h^{(1)} - \underline{V}_h^{(0)}) \geq 0. \end{cases}$$

By the positivity of A^{-1} and $(A + Mh_x^2 B)^{-1}$, we have from the above inequalities that $\overline{U}_h^{(0)} \geq \overline{U}_h^{(1)}$, $\overline{V}_h^{(0)} \geq \overline{V}_h^{(1)}$, $\underline{U}_h^{(1)} \geq \underline{U}_h^{(0)}$ and $\underline{V}_h^{(1)} \geq \underline{V}_h^{(0)}$. These facts with Lemma 3.1 lead to

$$(\underline{U}_h^{(0)}, \underline{V}_h^{(0)}) \leq (\underline{U}_h^{(1)}, \underline{V}_h^{(1)}) \leq (\overline{U}_h^{(1)}, \overline{V}_h^{(1)}) \leq (\overline{U}_h^{(0)}, \overline{V}_h^{(0)}).$$

We now assume inductively that for certain $m = m_0 \geq 1$,

$$(\underline{U}_h^{(m-1)}, \underline{V}_h^{(m-1)}) \leq (\underline{U}_h^{(m)}, \underline{V}_h^{(m)}) \leq (\overline{U}_h^{(m)}, \overline{V}_h^{(m)}) \leq (\overline{U}_h^{(m-1)}, \overline{V}_h^{(m-1)}). \tag{3.13}$$

Then by hypothesis (H),

$$\begin{aligned} &M\overline{V}_h^{(m_0-1)} + \max_{Z_h \in \mathcal{S}^{(m_0-1)}} F(Z_h, \overline{V}_h^{(m_0-1)}) \geq M\overline{V}_h^{(m_0)} + \max_{Z_h \in \mathcal{S}^{(m_0)}} F(Z_h, \overline{V}_h^{(m_0)}), \\ &M\underline{V}_h^{(m_0)} + \min_{Z_h \in \mathcal{S}^{(m_0)}} F(Z_h, \underline{V}_h^{(m_0)}) \geq M\underline{V}_h^{(m_0-1)} + \min_{Z_h \in \mathcal{S}^{(m_0-1)}} F(Z_h, \underline{V}_h^{(m_0-1)}). \end{aligned}$$

Accordingly, we use (3.8) to deduce that

$$(A + Mh_x^2 B)(\overline{V}_h^{(m_0)} - \overline{V}_h^{(m_0+1)}) \geq 0, \qquad (A + Mh_x^2 B)(\underline{V}_h^{(m_0+1)} - \underline{V}_h^{(m_0)}) \geq 0.$$

On the other hand, by (3.8) and (3.13) with $m = m_0$,

$$A(\overline{U}_h^{(m_0)} - \overline{U}_h^{(m_0+1)}) \geq 0, \qquad A(\underline{U}_h^{(m_0+1)} - \underline{U}_h^{(m_0)}) \geq 0.$$

In view of $(A + Mh_x^2 B)^{-1} > 0$, $A^{-1} > 0$ and Lemma 3.1, we assert that the relation (3.13) holds also for $m = m_0 + 1$. This completes the induction. The monotonicity of iteration also ensures the existence of the limits, namely,

$$\lim_{m \to \infty} (\overline{U}_h^{(m)}, \overline{V}_h^{(m)}) = (\overline{U}_h, \overline{V}_h), \qquad \lim_{m \to \infty} (\underline{U}_h^{(m)}, \underline{V}_h^{(m)}) = (\underline{U}_h, \underline{V}_h).$$

To prove (3.10), it suffices to show that

$$\begin{aligned}\lim_{m\to\infty}(M\overline{V}_h^{(m)}+\max_{Z_h\in\mathcal{S}^{(m)}}F(Z_h,\overline{V}_h^{(m)}))&=M\overline{V}_h+\max_{\underline{U}_h\le Z_h\le\overline{U}_h}F(Z_h,\overline{V}_h),\\ \lim_{m\to\infty}(M\underline{V}_h^{(m)}+\min_{Z_h\in\mathcal{S}^{(m)}}F(Z_h,\underline{V}_h^{(m)}))&=M\underline{V}_h+\min_{\underline{U}_h\le Z_h\le\overline{U}_h}F(Z_h,\underline{V}_h).\end{aligned}\tag{3.14}$$

Indeed, the above fact can be verified by exactly the same argument as that in proving Lemma A of the Appendix in [26].

Finally, let (U_h', V_h') be any solution of (2.17) in $\mathcal{S}$. Then $(\underline{U}_h^{(0)}, \underline{V}_h^{(0)}) \le (U_h', V_h') \le (\overline{U}_h^{(0)}, \overline{V}_h^{0})$. We next assume inductively that for some $m = m_0 \ge 0$,

$$(\underline{U}_h^{(m)}, \underline{V}_h^{(m)}) \le (U_h', V_h') \le (\overline{U}_h^{(m)}, \overline{V}_h^{(m)}).\tag{3.15}$$

Then by hypothesis (H),

$$M\overline{V}_h^{(m_0)}+\max_{Z_h\in\mathcal{S}^{(m_0)}}F(Z_h,\overline{V}_h^{(m_0)})\ge MV_h'+F(U_h',V_h').$$

Therefore by (2.17) and (3.8),

$$A(\overline{U}_h^{(m_0+1)}-U_h')\ge 0,\qquad (A+Mh_x^2B)(\overline{V}_h^{(m_0+1)}-V_h')\ge 0.$$

This leads to $(U_h', V_h') \le (\overline{U}_h^{(m_0+1)}, \overline{V}_h^{(m_0+1)})$. In the same manner, we verify that $(U_h', V_h') \ge (\underline{U}_h^{(m_0+1)}, \underline{V}_h^{(m_0+1)})$. This completes the induction. Thus, the relation (3.15) is valid for all $m \ge 0$. Letting $m \to \infty$ in (3.15), we conclude that $(\underline{U}_h, \underline{V}_h) \le (U_h', V_h') \le (\overline{U}_h, \overline{V}_h)$. □

We know from Theorem 3.2 that if $(\overline{U}_h, \overline{V}_h) = (\underline{U}_h, \underline{V}_h)$, then it is the unique solution of (2.17) in $\mathcal{S}$.

To explore the conditions ensuring the uniqueness of solution, we assume that F is a C^1-function and introduce the following notations:

$$\begin{aligned}&\overline{M}_u=\max\left\{\left|\frac{\partial F_{i,j}}{\partial u}(u,v)\right|;\ (u,v)\in\mathcal{S}_{i,j},(i,j)\in\Omega\right\},\\ &\overline{M}_v=\max\left\{\frac{\partial F_{i,j}}{\partial v}(u,v);(u,v)\in\mathcal{S}_{i,j},(i,j)\in\Omega\right\},\\ &\underline{M}_u^{\pm}=\min\left\{\pm\frac{\partial F_{i,j}}{\partial u}(u,v);(u,v)\in\mathcal{S}_{i,j},(i,j)\in\Omega\right\},\qquad \underline{M}_u=\max\{0,\underline{M}_u^{+},\underline{M}_u^{-}\},\\ &\underline{M}_v=\min\left\{\frac{\partial F_{i,j}}{\partial v}(u,v);(u,v)\in\mathcal{S}_{i,j},(i,j)\in\Omega\right\},\\ &\overline{k}=\max\{k_{i,j}^{-1},(i,j)\in\Omega\},\qquad \underline{k}=\min\{k_{i,j}^{-1},(i,j)\in\Omega\}.\end{aligned}\tag{3.16}$$

The following result is for the uniqueness of solution in $\mathcal{S}$, as well as an efficient algorithm.

Theorem 3.3. *Let the conditions in Lemma 3.1 hold, and let λ_0 be the smallest eigenvalue of the generalized eigenvalue problem (2.18). If, in addition, either*

$$\lambda_0(\lambda_0-\overline{M}_v)>\overline{k}\overline{M}_u\quad or\quad \lambda_0(\lambda_0-\underline{M}_v)<\underline{k}\underline{M}_u,\tag{3.17}$$

then the sequences $\{(\overline{U}_h^{(m)}, \overline{V}_h^{(m)})\}$ and $\{(\underline{U}_h^{(m)}, \underline{V}_h^{(m)})\}$ given by (3.8) with $M^* = M$ converge monotonically to a unique solution (U_h^*, V_h^*) of (2.17) in $\mathcal{S}$.

Proof. It suffices to show that $(\overline{U}_h, \overline{V}_h) = (\underline{U}_h, \underline{V}_h)$, where $(\overline{U}_h, \overline{V}_h)$ and $(\underline{U}_h, \underline{V}_h)$ are the limits in Theorem 3.2. Let $W_h = \overline{U}_h - \underline{U}_h$ and $X_h = \overline{V}_h - \underline{V}_h$. Then $W_h \ge 0$, $X_h \ge 0$ and by (3.10),

$$AW_h=h_x^2BKX_h,\qquad AX_h=h_x^2B(\max_{\underline{U}_h\le Z_h\le\overline{U}_h}F(Z_h,\overline{V}_h)-\min_{\underline{U}_h\le Z_h\le\overline{U}_h}F(Z_h,\underline{V}_h)).\tag{3.18}$$

Applying the mean-value theorem to the second equality of (3.18), we observe that

$$AX_h \le h_x^2 B(\overline{M}_u W_h + \overline{M}_v X_h). \tag{3.19}$$

Let

$$V_h' = \begin{cases} \overline{U}_h, & \text{if } \underline{M}_u = \underline{M}_u^+ \text{ or } \underline{M}_u = 0, \\ \underline{U}_h, & \text{if } \underline{M}_u = \underline{M}_u^-, \end{cases} \qquad V_h'' = \begin{cases} \underline{U}_h, & \text{if } \underline{M}_u = \underline{M}_u^+, \\ \overline{U}_h, & \text{if } \underline{M}_u = \underline{M}_u^- \text{ or } \underline{M}_u = 0. \end{cases} \tag{3.20}$$

Obviously, (3.18) implies that

$$AX_h \ge h_x^2 B(F(V_h', \overline{V}_h) - F(V_h'', \underline{V}_h)). \tag{3.21}$$

Therefore, using the mean-value theorem with (3.20) yields that

$$AX_h \ge h_x^2 B(\underline{M}_u W_h + \underline{M}_v X_h). \tag{3.22}$$

A combination of (3.18), (3.19) and (3.22) leads to

$$AW_h = h_x^2 BKX_h, \qquad h_x^2 B(\underline{M}_u W_h + \underline{M}_v X_h) \le AX_h \le h_x^2 B(\overline{M}_u W_h + \overline{M}_v X_h). \tag{3.23}$$

Next, let Φ be the positive eigenvector corresponding to the eigenvalue λ_0. Multiplying the equations in (3.23) by Φ^T gives

$$\Phi^T AW_h = h_x^2 \Phi^T BKX_h, \qquad h_x^2 \Phi^T B(\underline{M}_u W_h + \underline{M}_v X_h) \le \Phi^T AX_h \le h_x^2 \Phi^T B(\overline{M}_u W_h + \overline{M}_v X_h). \tag{3.24}$$

Due to the symmetry of A and B, we have from (2.18) that $\Phi^T A = \lambda_0 h_x^2 \Phi^T B$. Thereby,

$$\begin{aligned} &\lambda_0 h_x^2 \Phi^T BW_h = h_x^2 \Phi^T BKX_h, \\ &h_x^2 \Phi^T B\left(\underline{M}_u W_h + \underline{M}_v X_h\right) \le \lambda_0 h_x^2 \Phi^T BX_h \le h_x^2 \Phi^T B(\overline{M}_u W_h + \overline{M}_v X_h). \end{aligned} \tag{3.25}$$

Substituting the equality in (3.25) into the inequality in (3.25), it follows that

$$(\underline{kM}_u + \lambda_0 \underline{M}_v)\Phi^T BX_h \le \lambda_0^2 \Phi^T BX_h \le (\overline{kM}_u + \lambda_0 \overline{M}_v)\Phi^T BX_h.$$

If BX_h is not zero, then the above relation with the positivity of Φ leads to

$$\underline{kM}_u + \lambda_0 \underline{M}_v \le \lambda_0^2 \le \overline{kM}_u + \lambda_0 \overline{M}_v. \tag{3.26}$$

This contradicts (3.17), and so $BX_h = 0$ which implies $X_h = 0$. By (3.18), we have also $W_h = 0$. This ends the proof. □

In general, it is difficult to evaluate λ_0 exactly. But we can estimate it by using Forbenius theorem, see Appendix B.

We now search other conditions ensuring the uniqueness of solution, which do not involve λ_0. For this purpose, we introduce the following discrete norms:

$$\begin{aligned} &\|u^h\|_\infty = \max_{(i,j)\in\overline{\Omega}} |u_{i,j}^h|, \qquad \|u^h\|^2 = \frac{h_x^2}{\sigma L_x L_y} \sum_{i=1}^{N_x-1} \sum_{j=1}^{N_y-1} (u_{i,j}^h)^2, \\ &|u^h|_1^2 = \frac{1}{\sigma L_x L_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y-1} (u_{i,j}^h - u_{i-1,j}^h)^2 + \frac{\sigma}{L_x L_y} \sum_{i=1}^{N_x-1} \sum_{j=1}^{N_y} (u_{i,j}^h - u_{i,j-1}^h)^2, \end{aligned} \tag{3.27}$$

where σ is the ratio of mesh sizes as before. We also set

$$\sigma^* = \min\{\sigma(5-\sigma^2), (5\sigma^2-1)/\sigma\}, \qquad L^* = \max\{L_x^2, L_y^2\}. \tag{3.28}$$

For the above norms, we have the following estimates which will be proved in Appendix A.

Lemma 3.2. *If $1/5 < \sigma^2 < 5$ and $u^h_{i,j} = 0$ on $\partial\Omega$, then*

$$\sigma^* L_x L_y |u^h|_1^2 \le \sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1} u^h_{i,j}\mathcal{L}_h u^h_{i,j}. \tag{3.29}$$

Lemma 3.3. *If $u^h_{i,j} = 0$ on $\partial\Omega$, then*

$$4\|u^h\|^2 \le L^*|u^h|_1^2. \tag{3.30}$$

Theorem 3.4. *Let the conditions of Theorem 3.3 hold except that the condition (3.17) is replaced by*

$$3\sigma\overline{k}L^* < 4\sigma^*, \qquad \frac{2\sigma^*}{4\sigma^* - 3\sigma\overline{k}L^*}\overline{M}_u + \overline{M}_v^+ < \frac{2\sigma^*}{3\sigma L^*}, \tag{3.31}$$

where $\overline{M}_v^+ = \max\{0, \overline{M}_v\}$. Then the conclusions of Theorem 3.3 are also valid.

Proof. Let $(\overline{U}_h, \overline{V}_h)$ and $(\underline{U}_h, \underline{V}_h)$ be the limits in Theorem 3.2. We set $W_h = \overline{U}_h - \underline{U}_h$ and $X_h = \overline{V}_h - \underline{V}_h$. Then $W_h \ge 0$ and $X_h \ge 0$. By (3.10) and the mean-value theorem,

$$AW_h = h_x^2 BKX_h, \qquad AX_h = h_x^2 B(F(Z_h^*, \overline{V}_h) - F(Z_h^{**}, \underline{V}_h)), \tag{3.32}$$

where Z_h^* and Z_h^{**} are two intermediate values between $\underline{U}_h$ and $\overline{U}_h$. Let $w^h_{i,j}$ and $x^h_{i,j}$ be the components of the vectors W_h and X_h, respectively. Then, in the componentwise form, the system (3.32) can be rewritten as

$$\begin{cases} \mathcal{L}_h w^h_{i,j} = \mathcal{P}_h(x^h_{i,j}/k_{i,j}), & (i,j) \in \Omega, \\ \mathcal{L}_h x^h_{i,j} = \mathcal{P}_h(F_{i,j}(z^{*h}_{i,j}, \overline{v}^h_{i,j}) - F_{i,j}(z^{**h}_{i,j}, \underline{v}^h_{i,j})), & (i,j) \in \Omega, \\ w^h_{i,j} = x^h_{i,j} = 0, & (i,j) \in \partial\Omega. \end{cases} \tag{3.33}$$

Multiplying the first equation and the second equation of (3.33) by $w^h_{i,j}$ and $x^h_{i,j}$ respectively, and then summing the results for all i, j, we obtain that

$$\begin{cases} \displaystyle\sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1} w^h_{i,j}\mathcal{L}_h w^h_{i,j} = \sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1} w^h_{i,j}\mathcal{P}_h(x^h_{i,j}/k_{i,j}), \\ \displaystyle\sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1} x^h_{i,j}\mathcal{L}_h x^h_{i,j} = \sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1} x^h_{i,j}\mathcal{P}_h(F_{i,j}(z^{*h}_{i,j}, \overline{v}^h_{i,j}) - F_{i,j}(z^{**h}_{i,j}, \underline{v}^h_{i,j})). \end{cases}$$

By Lemma 3.2 and (3.28), the above system reads

$$\begin{aligned} &\sigma^* L_x L_y |w^h|_1^2 \le \overline{k}\sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1} w^h_{i,j}\mathcal{P}_h x^h_{i,j}, \\ &\sigma^* L_x L_y |x^h|_1^2 \le \sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1} (\overline{M}_u x^h_{i,j}\mathcal{P}_h w^h_{i,j} + \overline{M}_v^+ x^h_{i,j}\mathcal{P}_h x^h_{i,j}). \end{aligned} \tag{3.34}$$

Since

$$\sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1} w^h_{i,j}\mathcal{P}_h x^h_{i,j} \le 3h_x^2\sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1}((w^h_{i,j})^2 + (x^h_{i,j})^2) = 3\sigma L_x L_y(\|w^h\|^2 + \|x^h\|^2), \tag{3.35}$$

we obtain from (3.34) that

$$\sigma^*|w^h|_1^2 \le 3\sigma\overline{k}(\|w^h\|^2 + \|x^h\|^2), \qquad \sigma^*|x^h|_1^2 \le 3\sigma\overline{M}_u\|w^h\|^2 + 3\sigma(\overline{M}_u + 2\overline{M}_v^+)\|x^h\|^2.$$

Applying Lemma 3.3 to the above inequalities, we have

$$\frac{4\sigma^*}{L^*}\|w^h\|^2 \le 3\sigma\overline{k}(\|w^h\|^2+\|x^h\|^2), \qquad \frac{4\sigma^*}{L^*}\|x^h\|^2 \le 3\sigma\overline{M}_u\|w^h\|^2+3\sigma(\overline{M}_u+2\overline{M}_v^+)\|x^h\|^2. \tag{3.36}$$

The first estimate of (3.36) implies $\|w^h\|^2 \le \frac{3\sigma\overline{k}L^*}{4\sigma^*-3\sigma\overline{k}L^*}\|x^h\|^2$, and so by the second estimate,

$$\frac{2\sigma^*}{3\sigma L^*}\|x^h\|^2 \le \left(\frac{2\sigma^*}{4\sigma^*-3\sigma\overline{k}L^*}\overline{M}_u+\overline{M}_v^+\right)\|x^h\|^2.$$

The condition (3.31) ensures $\|x^h\|^2=0$ which implies $(\overline{U}_h,\overline{V}_h)=(\underline{U}_h,\underline{V}_h)$. □

In what follows, we shall use the following notations:

$$\overline{M}_u^*=\overline{M}_u, \qquad \overline{M}_v^*=\max\left\{\left|\frac{\partial F_{i,j}}{\partial v}(u,v)\right|;\ (u,v)\in\mathcal{S}_{i,j},(i,j)\in\Omega\right\}. \tag{3.37}$$

Theorem 3.5. *Let the conditions of Theorem 3.3 hold except that the condition (3.17) is replaced by*

$$\overline{k}\overline{M}_u^*+2\overline{M}_v^*<4. \tag{3.38}$$

Then the conclusions of Theorem 3.3 are also valid.

Proof. It suffices to show $W_h=X_h=0$, where the components of W_h and X_h satisfy (3.33). Applying Lemma 2.2 to (3.33) gives

$$\|w^h\|_\infty \le \overline{k}\|x^h\|_\infty/2, \qquad \|x^h\|_\infty \le \overline{M}_u^*\|w^h\|_\infty/2+\overline{M}_v^*\|x^h\|_\infty/2.$$

Consequently,

$$\|x^h\|_\infty \le (\overline{k}\overline{M}_u^*/4+\overline{M}_v^*/2)\|x^h\|_\infty.$$

The above with condition (3.38) implies $\|x^h\|_\infty=\|w^h\|_\infty=0$, and so $W_h=X_h=0$. □

Remark 3.1. Since we adopt the locally extreme values of F, at the right-hand sides of the iteration (3.8), the monotone convergence of the produced sequences follows without any requirement on the monotonicity of F. This fact essentially enlarges their applications.

Remark 3.2. If the function $F(U_h,V_h)$ is monotone in U_h, then the iteration (3.8) is reduced to a usual iteration with monotone function F. In this case, the computation of maximum and minimum values of nonlinear function in (3.8) is trivial. Moreover, $\underline{M}_u$ in (3.16) is defined by $\underline{M}_u=\min\{|\frac{\partial F_{i,j}}{\partial u}(u,v)|;\ (u,v)\in\mathcal{S}_{i,j},(i,j)\in\Omega\}$.

Remark 3.3. If $F(U_h,V_h)$ is nonmonotone in U_h, then the maximum and minimum values can be determined by $\frac{\partial F_{i,j}}{\partial u}=0,\ (i,j)\in\Omega$.

4. Two block monotone iterations

The Picard iteration (3.8) leads to a nine-diagonal linear system at each step of iteration. For resolving this system, we need another iteration procedure. This is expensive for large number of mesh points. To remedy this deficiency and maintain the monotonicity of convergence, we now propose two new iterations, called block Jacobi iteration and block Gauss–Seidel iteration, respectively. To do this, we split the matrices A and B as $A=\mathcal{D}-\mathcal{L}-\mathcal{U}$ and $B=\mathcal{D}^*+\mathcal{L}^*+\mathcal{U}^*$, where $\mathcal{D},\mathcal{L},\mathcal{U},\mathcal{D}^*,\mathcal{L}^*$ and $\mathcal{U}^*$ are $\mathcal{N}$-order block tridiagonal matrices, namely,

$$\mathcal{D}=\text{tridiag}(0,A_0,0), \qquad \mathcal{L}=\text{tridiag}(A_1,0,0), \qquad \mathcal{U}=\text{tridiag}(0,0,A_1),$$
$$\mathcal{D}^*=\text{tridiag}(0,B_0,0), \qquad \mathcal{L}^*=\text{tridiag}(B_1,0,0), \qquad \mathcal{U}^*=\text{tridiag}(0,0,B_1).$$

(a) Block Jacobi iteration

Let $(\widetilde{U}_h,\widetilde{V}_h)$ and $(\widehat{U}_h,\widehat{V}_h)$ be a pair of coupled upper and lower solutions of (2.17). The Block Jacobi iteration produces the sequences $\{(\overline{U}_{J,h}^{(m)},\overline{V}_{J,h}^{(m)})\}$ and $\{(\underline{U}_{J,h}^{(m)},\underline{V}_{J,h}^{(m)})\}$ as follows,

$$\begin{cases} (\overline{U}_{J,h}^{(0)}, \overline{V}_{J,h}^{(0)}) = (\widetilde{U}_h, \widetilde{V}_h), \qquad (\underline{U}_{J,h}^{(0)}, \underline{V}_{J,h}^{(0)}) = (\widehat{U}_h, \widehat{V}_h), \\ \mathcal{D}\overline{U}_{J,h}^{(m)} = (\mathcal{L}+\mathcal{U})\overline{U}_{J,h}^{(m-1)} + h_x^2 BK\overline{V}_{J,h}^{(m-1)} + G^{(1)}, \quad m \geq 1, \\ (\mathcal{D} + M^* h_x^2 \mathcal{D}^*)\overline{V}_{J,h}^{(m)} = (\mathcal{L}+\mathcal{U} - M^* h_x^2(\mathcal{L}^*+\mathcal{U}^*))\overline{V}_{J,h}^{(m-1)} \\ \quad + h_x^2 B(M^*\overline{V}_{J,h}^{(m-1)} + \max\limits_{Z_h \in \mathcal{S}^{(m-1)}} F(Z_h, \overline{V}_{J,h}^{(m-1)})) + G^{(2)}, \quad m \geq 1, \\ \mathcal{D}\underline{U}_{J,h}^{(m)} = (\mathcal{L}+\mathcal{U})\underline{U}_{J,h}^{(m-1)} + h_x^2 BK\underline{V}_{J,h}^{(m-1)} + G^{(1)}, \quad m \geq 1, \\ (\mathcal{D} + M^* h_x^2 \mathcal{D}^*)\underline{V}_{J,h}^{(m)} = (\mathcal{L}+\mathcal{U} - M^* h_x^2(\mathcal{L}^*+\mathcal{U}^*))\underline{V}_{J,h}^{(m-1)} \\ \quad + h_x^2 B(M^*\underline{V}_{J,h}^{(m-1)} + \min\limits_{Z_h \in \mathcal{S}^{(m-1)}} F(Z_h, \underline{V}_{J,h}^{(m-1)})) + G^{(2)}, \quad m \geq 1, \end{cases} \tag{4.1}$$

where M^* is a nonnegative constant specified later, and $\mathcal{S}^{(m)}$ is defined by (3.9) with respect to $\underline{U}_{J,h}^{(m)}$ and $\overline{U}_{J,h}^{(m)}$.

(b) Block Gauss–Seidel iteration

The block Gauss–Seidel iteration is designed by replacing the matrices $\mathcal{D}$, $\mathcal{D}^*$, $\mathcal{L}+\mathcal{U}$ and $\mathcal{L}^*+\mathcal{U}^*$ in (4.1) by the matrices $\mathcal{D}-\mathcal{L}$, $\mathcal{D}^*+\mathcal{L}^*$, $\mathcal{U}$ and $\mathcal{U}^*$, respectively. It produces the sequences $\{(\overline{U}_{G,h}^{(m)}, \overline{V}_{G,h}^{(m)})\}$ and $\{(\underline{U}_{G,h}^{(m)}, \underline{V}_{G,h}^{(m)})\}$.

Evidently, the matrices $\mathcal{D}$, $\mathcal{D} + M^* h_x^2 \mathcal{D}^*$, $\mathcal{D}-\mathcal{L}$ and $\mathcal{D}-\mathcal{L}+M^* h_x^2(\mathcal{D}^*+\mathcal{L}^*)$ are block diagonal or block lower-tridiagonal. Thereby, at each step of the block Jacobi and block Gauss–Seidel iterations, we could use certain explicit and efficient algorithms, such as Thomas algorithm. This simplifies the actual computation essentially and saves a lot of computational time.

Like the Picard iteration, we have the following results on the convergence of the above proposed iterations.

Theorem 4.1. *Let the conditions in Lemma 3.1 be satisfied. Then the conclusions in Lemma 3.1 and Theorem 3.2 are valid for the sequences produced by the block Jacobi and block Gauss–Seidel iterations with $M^* = M$. If, in addition, one of the conditions (3.17), (3.31) and (3.38) holds, then the conclusions in Theorem 3.3 also hold for these sequences.*

Proof. It is easy to verify that under the condition (2.19), the inverses $\mathcal{D}^{-1}$, $(\mathcal{D}+Mh_x^2\mathcal{D}^*)^{-1}$, $(\mathcal{D}-\mathcal{L})^{-1}$ and $(\mathcal{D}-\mathcal{L}+Mh_x^2(\mathcal{D}^*+\mathcal{L}^*))^{-1}$ exist and are positive. Moreover, the matrices $\mathcal{L}+\mathcal{U}$, $\mathcal{L}+\mathcal{U}-Mh_x^2(\mathcal{L}^*+\mathcal{U}^*)$, $\mathcal{U}$ and $\mathcal{U}-Mh_x^2\mathcal{U}^*$ are nonnegative. Accordingly, we follow the same line as in the proofs of Lemma 3.1, and Theorems 3.2 and 3.3 to reach the desired results. □

We next compare the Picard, block Jacobi and block Gauss–Seidel iterations.

Theorem 4.2. *Let the conditions in Lemma 3.1 hold, and let $\{\overline{U}_h^{(m)}, \overline{V}_h^{(m)}, \underline{U}_h^{(m)}, \underline{V}_h^{(m)}\}$, $\{\overline{U}_{J,h}^{(m)}, \overline{V}_{J,h}^{(m)}, \underline{U}_{J,h}^{(m)}, \underline{V}_{J,h}^{(m)}\}$ and $\{\overline{U}_{G,h}^{(m)}, \overline{V}_{G,h}^{(m)}, \underline{U}_{G,h}^{(m)}, \underline{V}_{G,h}^{(m)}\}$ be the sequences produced by the Picard, block Jacobi and block Gauss–Seidel iterations with $M^* = M$ and the same initial data, respectively. Then for all $m \geq 1$,*

$$(\underline{U}_{J,h}^{(m)}, \underline{V}_{J,h}^{(m)}) \leq (\underline{U}_{G,h}^{(m)}, \underline{V}_{G,h}^{(m)}) \leq (\underline{U}_h^{(m)}, \underline{V}_h^{(m)}) \leq (\overline{U}_h^{(m)}, \overline{V}_h^{(m)}) \leq (\overline{U}_{G,h}^{(m)}, \overline{V}_{G,h}^{(m)}) \leq (\overline{U}_{J,h}^{(m)}, \overline{V}_{J,h}^{(m)}). \tag{4.2}$$

Proof. Let $(\overline{W}_h^{(m)}, \overline{Z}_h^{(m)}) = (\overline{U}_{G,h}^{(m)} - \overline{U}_h^{(m)}, \overline{V}_{G,h}^{(m)} - \overline{V}_h^{(m)})$ and $(\underline{W}_h^{(m)}, \underline{Z}_h^{(m)}) = (\underline{U}_h^{(m)} - \underline{U}_{G,h}^{(m)}, \underline{V}_h^{(m)} - \underline{V}_{G,h}^{(m)})$. Then we have from (3.8) and the corresponding formulas of block Gauss–Seidel iteration that

$$\begin{cases} (\mathcal{D}-\mathcal{L})\overline{W}_h^{(m)} = \mathcal{U}(\overline{U}_{G,h}^{(m-1)} - \overline{U}_h^{(m)}) + h_x^2 BK\overline{Z}_h^{(m-1)}, \\ (\mathcal{D}-\mathcal{L}+Mh_x^2(\mathcal{D}^*+\mathcal{L}^*))\overline{Z}_h^{(m)} = (\mathcal{U} - Mh_x^2\mathcal{U}^*)(\overline{V}_{G,h}^{(m-1)} - \overline{V}_h^{(m)}) \\ \quad + h_x^2 B(M\overline{Z}_h^{(m-1)} + \max\limits_{Z_h \in \mathcal{S}_G^{(m-1)}} F(Z_h, \overline{V}_{G,h}^{(m-1)}) - \max\limits_{Z_h \in \mathcal{S}^{(m-1)}} F(Z_h, \overline{V}_h^{(m-1)})), \\ (\mathcal{D}-\mathcal{L})\underline{W}_h^{(m)} = \mathcal{U}(\underline{U}_h^{(m)} - \underline{U}_{G,h}^{(m-1)}) + h_x^2 BK\underline{Z}_h^{(m-1)}, \\ (\mathcal{D}-\mathcal{L}+Mh_x^2(\mathcal{D}^*+\mathcal{L}^*))\underline{Z}_h^{(m)} = (\mathcal{U} - Mh_x^2\mathcal{U}^*)(\underline{V}_h^{(m)} - \underline{V}_{G,h}^{(m-1)}) \\ \quad + h_x^2 B(M\underline{Z}_h^{(m-1)} + \min\limits_{Z_h \in \mathcal{S}^{(m-1)}} F(Z_h, \underline{V}_h^{(m-1)}) - \min\limits_{Z_h \in \mathcal{S}_G^{(m-1)}} F(Z_h, \underline{V}_{G,h}^{(m-1)})), \end{cases} \tag{4.3}$$

where $\mathcal{S}_G^{(m)}$ is defined by (3.9) with respect to the block Gauss–Seidel iteration. Since $\mathcal{U} \geq 0, \mathcal{U} - Mh_x^2\mathcal{U}^* \geq 0$, and

$$(\underline{U}_h^{(m-1)}, \underline{V}_h^{(m-1)}) \leq (\underline{U}_h^{(m)}, \underline{V}_h^{(m)}) \leq (\overline{U}_h^{(m)}, \overline{V}_h^{(m)}) \leq (\overline{U}_h^{(m-1)}, \overline{V}_h^{(m-1)}),$$

we obtain from (4.3) that

$$\begin{cases} (\mathcal{D}-\mathcal{L})\overline{W}_h^{(m)} \geq \mathcal{U}\overline{W}_h^{(m-1)} + h_x^2 BK\overline{Z}_h^{(m-1)}, \\ (\mathcal{D}-\mathcal{L}+Mh_x^2(\mathcal{D}^*+\mathcal{L}^*))\overline{Z}_h^{(m)} \geq (\mathcal{U}-Mh_x^2\mathcal{U}^*)\overline{Z}_h^{(m-1)} \\ \quad + h_x^2 B(M\overline{Z}_h^{(m-1)} + \max\limits_{Z_h\in\mathcal{S}_G^{(m-1)}} F(Z_h, \overline{V}_{G,h}^{(m-1)}) - \max\limits_{Z_h\in\mathcal{S}^{(m-1)}} F(Z_h, \overline{V}_h^{(m-1)})), \\ (\mathcal{D}-\mathcal{L})\underline{W}_h^{(m)} \geq \mathcal{U}\underline{W}_h^{(m-1)} + h_x^2 BK\underline{Z}_h^{(m-1)}, \\ (\mathcal{D}-\mathcal{L}+Mh_x^2(\mathcal{D}^*+\mathcal{L}^*))\underline{Z}_h^{(m)} \geq (\mathcal{U}-Mh_x^2\mathcal{U}^*)\underline{Z}_h^{(m-1)} \\ \quad + h_x^2 B(M\underline{Z}_h^{(m-1)} + \min\limits_{Z_h\in\mathcal{S}^{(m-1)}} F(Z_h, \underline{V}_h^{(m-1)}) - \min\limits_{Z_h\in\mathcal{S}_G^{(m-1)}} F(Z_h, \underline{V}_{G,h}^{(m-1)})). \end{cases} \tag{4.4}$$

We now use induction. Consider the case $m = 1$. Since the considered iterations possess the same initial data, the relation (4.4) with $m = 1$ is reduced to

$$(\mathcal{D}-\mathcal{L})\overline{W}_h^{(1)} \geq 0, \qquad (\mathcal{D}-\mathcal{L}+Mh_x^2(\mathcal{D}^*+\mathcal{L}^*))\overline{Z}_h^{(1)} \geq 0,$$
$$(\mathcal{D}-\mathcal{L})\underline{W}_h^{(1)} \geq 0, \qquad (\mathcal{D}-\mathcal{L}+Mh_x^2(\mathcal{D}^*+\mathcal{L}^*))\underline{Z}_h^{(1)} \geq 0.$$

The above with the positivity of $(\mathcal{D}-\mathcal{L})^{-1}$ and $(\mathcal{D}-\mathcal{L}+Mh_x^2(\mathcal{D}^*+\mathcal{L}^*))^{-1}$ leads to $(\overline{W}_h^{(1)}, \overline{Z}_h^{(1)}) \geq (0,0)$ and $(\underline{W}_h^{(1)}, \underline{Z}_h^{(1)}) \geq (0,0)$. Finally, with the aid of (3.3) and (4.4), we show inductively that $(\overline{W}_h^{(m)}, \overline{Z}_h^{(m)}) \geq (0,0)$ and $(\underline{W}_h^{(m)}, \underline{Z}_h^{(m)}) \geq (0,0)$ for all $m \geq 1$, i.e.,

$$(\underline{U}_{G,h}^{(m)}, \underline{V}_{G,h}^{(m)}) \leq (\underline{U}_h^{(m)}, \underline{V}_h^{(m)}) \leq (\overline{U}_h^{(m)}, \overline{V}_h^{(m)}) \leq (\overline{U}_{G,h}^{(m)}, \overline{V}_{G,h}^{(m)}), \quad m \geq 1.$$

We can prove the other inequalities in (4.2) in the same manner. □

Remark 4.1. According to (4.2), the Picard iteration might converge faster than the block Gauss–Seidel iteration. The latter in turn might converge faster than the block Jacobi iteration. However, the block Jacobi and block Gauss–Seidel iterations are more preferable, since they are much easier to be carried out, and do not require any additional iteration at each step.

Remark 4.2. Using the same argument as in [25], we can show that the Picard, block Jacobi and block Gauss–Seidel iterations given in this paper have the geometric convergence rates.

5. Convergence of compact scheme

In this section, we deal with the convergence of finite difference scheme (2.9) (or (2.17)), and show its fourth-order accuracy.

Let $(u_{i,j}, v_{i,j})$ be the value at the point (i, j) of solution of (2.1), and $(u_{i,j}^h, v_{i,j}^h)$ stands for the solution of (2.9). We consider the errors $e_{i,j}^h = u_{i,j} - u_{i,j}^h$ and $e_{i,j}'^h = v_{i,j} - v_{i,j}^h$. In fact, we have from (2.6) and (2.9) that

$$\begin{cases} \mathcal{L}_h e_{i,j}^h = \mathcal{P}_h(e_{i,j}'^h/k_{i,j}) + O(h^6), & (i,j) \in \Omega, \\ \mathcal{L}_h e_{i,j}'^h = \mathcal{P}_h(F_{i,j}(u_{i,j}, v_{i,j}) - F_{i,j}(u_{i,j}^h, v_{i,j}^h)) + O(h^6), & (i,j) \in \Omega, \\ e_{i,j}^h = e_{i,j}'^h = 0, & (i,j) \in \partial\Omega. \end{cases} \tag{5.1}$$

Multiplying the first and the second equations of (5.1) by $e^h_{i,j}$ and $e'^h_{i,j}$ respectively, and then summing the results for all i, j, we obtain that

$$\begin{cases} \displaystyle\sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1} e^h_{i,j}\mathcal{L}_h e^h_{i,j} = \sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1} e^h_{i,j}\mathcal{P}_h(e'^h_{i,j}/k_{i,j}) + O(h^6)\sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1} e^h_{i,j}, \\ \displaystyle\sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1} e'^h_{i,j}\mathcal{L}_h e'^h_{i,j} = \sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1} e'^h_{i,j}\mathcal{P}_h(F_{i,j}(u_{i,j},v_{i,j}) - F_{i,j}(u^h_{i,j},v^h_{i,j})) + O(h^6)\sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1} e'^h_{i,j}. \end{cases} \tag{5.2}$$

Theorem 5.1. *Let $\mathcal{S}^*_{i,j}$ be the set in $\mathbf{R}^2$ such that $(u_{i,j},v_{i,j}),(u^h_{i,j},v^h_{i,j})\in\mathcal{S}^*_{i,j}$, and let $\overline{M}^*_u$ and $\overline{M}^*_v$ be the constants defined by (3.37) with respect to $\mathcal{S}^*_{i,j}$. If $u,v\in C^6([0,L_x]\times[0,L_y])$, $1/5<\sigma^2<5$ and*

$$3\sigma\bar{k}L^*<4\sigma^*,\qquad \frac{2\sigma^*}{4\sigma^*-3\sigma\bar{k}L^*}\overline{M}^*_u+\overline{M}^*_v<\frac{2\sigma^*}{3\sigma L^*}, \tag{5.3}$$

then

$$\|u-u^h\|\le\frac{c^*}{\sqrt{\lambda^*}}h^4,\qquad \|v-v^h\|\le\frac{c^*}{\sqrt{\lambda^*}}h^4,$$

where c^* is a positive constant independent of h, and $\lambda^*=\frac{1}{2}(\frac{2\sigma^*}{3\sigma L^*}-\frac{2\sigma^*}{4\sigma^*-3\sigma\bar{k}L^*}\overline{M}^*_u-\overline{M}^*_v)>0$.

Proof. By using Lemma 3.2, we have from (5.2) that

$$\begin{cases} \sigma^*L_xL_y|e^h|_1^2\le\bar{k}\displaystyle\sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1}|e^h_{i,j}|\mathcal{P}_h(|e'^h_{i,j}|)+O(h^6)\sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1}e^h_{i,j}, \\ \sigma^*L_xL_y|e'^h|_1^2\le\displaystyle\sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1}(\overline{M}^*_u|e'^h_{i,j}|\mathcal{P}_h(|e^h_{i,j}|)+\overline{M}^*_v|e'^h_{i,j}|\mathcal{P}_h(|e'^h_{i,j}|))+O(h^6)\sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1}e'^h_{i,j}. \end{cases} \tag{5.4}$$

Let ε be a suitably small positive constant determined below. It is easy to see that

$$O(h^6)\sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1}e^h_{i,j}=\sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1}O(h^5)h_x\sqrt{\frac{4\varepsilon}{\sigma L^*}}\,e^h_{i,j}\le\frac{2\varepsilon L_xL_y}{L^*}\|e^h\|^2+\frac{1}{\varepsilon}O(h^8),\quad\text{etc.}$$

Thus, by virtue of Lemma 3.3 and (3.35), we derive from (5.4) that

$$\begin{cases} \dfrac{4\sigma^*}{L^*}\|e^h\|^2\le3\sigma\bar{k}\left(\|e^h\|^2+\|e'^h\|^2\right)+\dfrac{2\varepsilon}{L^*}\|e^h\|^2+\dfrac{1}{\varepsilon}O(h^8), \\ \dfrac{4\sigma^*}{L^*}\|e'^h\|^2\le3\sigma\overline{M}^*_u\|e^h\|^2+3\sigma(\overline{M}^*_u+2\overline{M}^*_v)\|e'^h\|^2+\dfrac{2\varepsilon}{L^*}\|e'^h\|^2+\dfrac{1}{\varepsilon}O(h^8). \end{cases} \tag{5.5}$$

For any positive $\varepsilon<(4\sigma^*-3\sigma\bar{k}L^*)/2$, the first estimate of (5.5) implies that

$$\|e^h\|^2\le\frac{3\sigma\bar{k}L^*}{4\sigma^*-3\sigma\bar{k}L^*-2\varepsilon}\|e'^h\|^2+\frac{1}{\varepsilon}O(h^8). \tag{5.6}$$

Along with the second estimate of (5.5), we use (5.6) to verify that

$$\frac{2\sigma^*-\varepsilon}{3\sigma L^*}\|e'^h\|^2\le\left(\frac{2\sigma^*-\varepsilon}{4\sigma^*-3\sigma\bar{k}L^*-2\varepsilon}\overline{M}^*_u+\overline{M}^*_v\right)\|e'^h\|^2+\frac{1}{\varepsilon}O(h^8). \tag{5.7}$$

Due to the condition (5.3), we may take ε to be sufficiently small so that $\varepsilon<(4\sigma^*-3\sigma\bar{k}L^*)/2$ and

$$\frac{2\sigma^*-\varepsilon}{3\sigma L^*}-\frac{2\sigma^*-\varepsilon}{4\sigma^*-3\sigma\bar{k}L^*-2\varepsilon}\overline{M}^*_u-\overline{M}^*_v\ge\lambda^*>0. \tag{5.8}$$

Hence, there exists a positive constant c_v independent of h such that $\|e'^h\|^2 \le \frac{c_v}{\lambda^*}h^8$. This with (5.6) in turn implies $\|e^h\|^2 \le \frac{c_u}{\lambda^*}h^8$, c_u being a positive constant independent of h. The proof is completed. □

We next give an error estimate of numerical solution in the discrete L_∞-norm, which has the same order as the error estimate in the discrete L_2-norm.

Theorem 5.2. *Let the hypotheses in Theorem 5.1 be satisfied except that the condition (5.3) is replaced by (3.38). Then there exists a positive constant c^*_∞ independent of h such that*

$$\|u-u^h\|_\infty \le \frac{c^*_\infty}{\lambda^*_\infty}h^4, \qquad \|v-v^h\|_\infty \le \frac{c^*_\infty}{\lambda^*_\infty}h^4, \tag{5.9}$$

where $\lambda^*_\infty = \frac{1}{4}(4-\overline{kM}^*_u - 2\overline{M}^*_v) > 0$.

Proof. Let $e^h_{i,j}$ and $e'^h_{i,j}$ be the same as before. An application of the mean-value theorem and Lemma 2.2 to (5.1) yields that

$$\|e^h\|_\infty \le \bar{k}\|e'^h\|_\infty/2 + O(h^4), \qquad \|e'^h\|_\infty \le \overline{M}^*_u\|e^h\|_\infty/2 + \overline{M}^*_v\|e'^h\|_\infty/2 + O(h^4). \tag{5.10}$$

This implies

$$\|e'^h\|_\infty \le \left(\overline{kM}^*_u/4 + \overline{M}^*_v/2\right)\|e'^h\|_\infty + O(h^4). \tag{5.11}$$

By using (3.38), we reach the estimate (5.9). □

6. Numerical results

We now present some numerical results demonstrating the monotone convergence of iterations and the fourth-order accuracy of numerical solution, as predicted in the analysis. As mentioned in the previous sections, we have to find certain pairs of coupled upper and lower solutions for ensuring the monotone convergence of iterations, which, in turn, depend mainly on the function $F(U,V)$. Our example below also illustrates a technique for constructing such pairs.

Let $\Omega = \{(x,y); 0<x<1, 0<y<1\}$, and consider the boundary value problem

$$\begin{cases} \Delta^2 u = p(x,y)\dfrac{\Delta u}{1+u} + q(x,y), & (x,y)\in\Omega, \\ u = \Delta u = 0, & (x,y)\in\partial\Omega, \end{cases} \tag{6.1}$$

where $p(x,y)$ is a sign-changing continuous function and $q(x,y)$ is a nonnegative continuous function. Problem (6.1) is a special case of (1.2) with

$$k(x,y)=1, \quad f(x,y,u,v) = p(x,y)\frac{v}{1+u} + q(x,y), \quad g(x,y) = g^*(x,y) = 0, \quad L_x = L_y = 1. \tag{6.2}$$

To obtain an explicit solution of (6.1), we take a positive constant κ, and

$$q(x,y) = 2\pi^2\kappa\left(2\pi^2 + \frac{p(x,y)}{1+\kappa\sin(\pi x)\sin(\pi y)}\right)\sin(\pi x)\sin(\pi y). \tag{6.3}$$

The true solution of (6.1) is given by $u(x,y) = \kappa\sin(\pi x)\sin(\pi y)$ for any $p(x,y)$. Moreover, $q(x,y)\ge 0$, if $p(x,y)\ge -2\pi^2$ in Ω.

Let $h = h_x = h_y$. The finite difference scheme (2.17) for problem (6.1) is now reduced to

$$\begin{cases} AU_h = h^2 BV_h, \\ AV_h = h^2 BF(U_h, V_h), \end{cases} \tag{6.4}$$

where A and B are the same as in (2.16) with $\sigma = 1$, and $F(U_h,V_h)$ is defined by (2.15) and (6.2). Since $p(x,y)$ changes sign, $F(U_h,V_h)$ is not monotone in U_h.

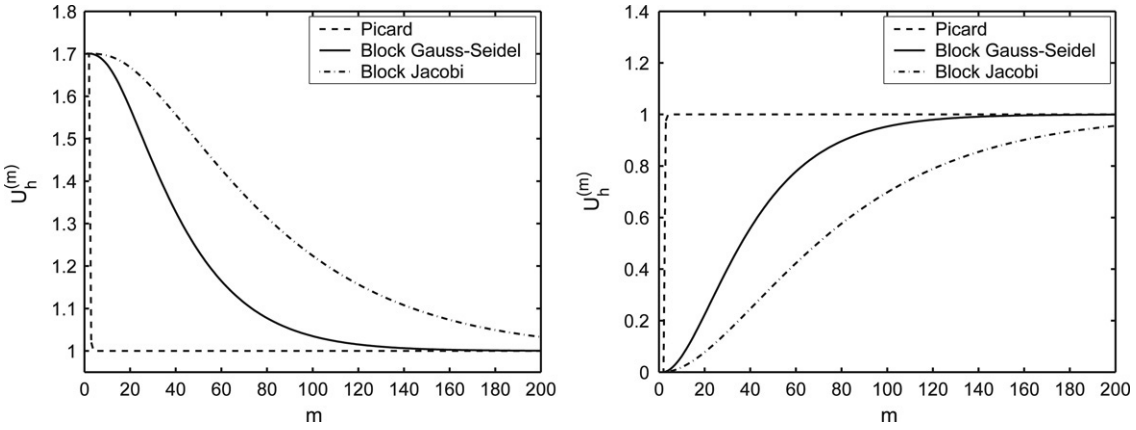

Fig. 6.1. The monotone property of $\{\overline{U}_h^{(m)}, \underline{U}_h^{(m)}\}$ at (0.5, 0.5) by different iterations (left: $\overline{U}_h^{(m)}$; right: $\underline{U}_h^{(m)}$).

To find a pair of coupled upper and lower solutions of (6.4), we consider the following auxiliary linear system

$$\begin{cases} A\widetilde{Z}_h = \overline{p}h^2 B\widetilde{Z}_h + \overline{q}h^2 BE, \\ A\widetilde{W}_h = h^2 B\widetilde{Z}_h, \end{cases} \tag{6.5}$$

where $\overline{p}$ and $\overline{q}$ are sufficiently large so that $|p(x, y)| \le \overline{p}$ and $q(x, y) \le \overline{q}$ in Ω, $E \in \mathbf{R}^{\mathcal{N}}$ is a vector whose components are all one. The system (6.5) admits a unique solution $(\widetilde{W}_h, \widetilde{Z}_h)$, and $(\widetilde{W}_h, \widetilde{Z}_h) \ge (0, 0)$ if $\overline{p} < \lambda_0$, where λ_0 is the smallest eigenvalue of the generalized eigenvalue problem (2.18) (see Lemma 2.4). Notice that $\lambda_0 \ge 8/3$ (see Theorem B.1 in Appendix B). It is easy to verify that $F(U_h, V_h) \le \overline{p}V_h + \overline{q}E$ for all $(U_h, V_h) \ge (0, 0)$. Consequently, $(\widetilde{W}_h, \widetilde{Z}_h)$ and $(0, 0)$ form a pair of coupled upper and lower solutions of (6.4).

Since $\partial F_{i,j}/\partial v = -p/(1+u) \ge -\overline{p}$ for all $u \ge 0$, the nonnegative constant M^* in the iteration processes (3.8) and (4.1) may be chosen as $M^* = \overline{p}$. Let $\kappa = 1$, $p(x, y) = 0.5\cos(\pi x)\cos(\pi y)$, $\overline{p} = 1/2$ and $\overline{q} = 2\pi^2\kappa(2\pi^2 + 1)$. Taking $(\overline{U}_h^{(0)}, \overline{V}_h^{(0)}) = (\widetilde{W}_h, \widetilde{Z}_h)$ and $(\underline{U}_h^{(0)}, \underline{V}_h^{(0)}) = (0, 0)$, we produce the corresponding sequences $\{(\overline{U}_h^{(m)}, \overline{V}_h^{(m)})\}$ and $\{(\underline{U}_h^{(m)}, \underline{V}_h^{(m)})\}$ by the Picard iteration, block Jacobi iteration and block Gauss–Seidel iteration. All computations are carried out on a Pentium-4 computer with 512 MB memory and a MATLAB subroutine. The termination criterion of iterations is

$$\|\overline{U}_h^{(m)} - \underline{U}_h^{(m)}\|_\infty + \|\overline{V}_h^{(m)} - \underline{V}_h^{(m)}\|_\infty < 10^{-12}. \tag{6.6}$$

In Fig. 6.1, we plot the values of sequences $\{\overline{U}_h^{(m)}\}$ and $\{\underline{U}_h^{(m)}\}$ at the point $(x_i, y_j) = (0.5, 0.5)$, with $h = 1/20$. As expected from our analysis, the sequence $\{\overline{U}_h^{(m)}\}$ is nonincreasing, while the sequence $\{\underline{U}_h^{(m)}\}$ is nondecreasing. Besides, the comparison result (4.2) is also confirmed. In numerical experiments, the sequences $\{(\overline{U}_h^{(m)}, \overline{V}_h^{(m)})\}$ and $\{(\underline{U}_h^{(m)}, \underline{V}_h^{(m)})\}$ tend to the same limit (U_h^*, V_h^*) as $m \to \infty$. It indicates that the limit (U_h^*, V_h^*) is the unique solution of (6.4) in $\mathcal{S} = \{(U_h, V_h) \in (\mathbf{R}^{\mathcal{N}})^2; (0, 0) \le (U_h, V_h) \le (\widetilde{W}_h, \widetilde{Z}_h)\}$. Therefore, for illustrating the accuracy of iterations, we may take $(\overline{U}_h^{(m^*)}, \overline{V}_h^{(m^*)})$ as the computed solution (U_h^*, V_h^*) where m^* is the number of required steps of iterations for the tolerance in (6.6).

In Table 6.1, we list the values of computed solution U_h^* at the point (x_i, y_j), with $y_j = 0.5$ and the mesh size $h = 1/10, 1/20, 1/40$. We also list the relative error (Relat. err.) between the computed solution U_h^* and the true solution u at every point, the values of the true solution u, the number of required steps of iterations (Number of iter.) and CPU time (in seconds). We see that the Picard iteration converges faster than the block Gauss–Seidel and block Jacobi iterations. However, the block Gauss–Seidel and block Jacobi iterations cost much less computational time, especially for small mesh size.

In Fig. 6.2, we sketch the absolute maximum errors (i.e., L_∞-errors, see (6.7)) between the true solution u and the computed solution U_h^* produced by the block Gauss–Seidel iteration with $h = 1/N$ for different N. Clearly, the computed solution meets the true solution closely, and the numerical error decays as the mesh size decreases.

Table 6.1
The computed solution U_h^* and the true solution u

h	(x_i, y_j)	(0.1, 0.5)	(0.2, 0.5)	(0.3, 0.5)	(0.4, 0.5)	(0.5, 0.5)	Number of iter.	CPU time (s)
(a) Picard iteration								
1/10	Comp. sol.	0.30904218	0.58783315	0.80908293	0.95113402	1.00008150		
	Relat. err.	8.15165e−5	8.14923e−5	8.15063e−5	8.14883e−5	8.15000e−5	9	0.094
1/20	Comp. sol.	0.30901856	0.58778824	0.80902110	0.95106135	1.00000508		
	Relat. err.	5.08063e−6	5.08689e−6	5.08024e−6	5.07856e−5	5.08000e−6	9	4.344
1/40	Comp. sol.	0.30901709	0.58778544	0.80901725	0.95105682	1.00000032		
	Relat. err.	3.23607e−7	3.23247e−7	3.21378e−7	3.15439e−7	3.20000e−7	9	304.141
(b) Block Gauss–Seidel iteration								
1/10	Comp. sol.	0.30904218	0.58783315	0.80908293	0.95113402	1.00008150		
	Relat. err.	8.15165e−5	8.14923e−5	8.15063e−5	8.14883e−5	8.15000e−5	167	0.344
1/20	Comp. sol.	0.30901856	0.58778824	0.80902110	0.95106135	1.00000508		
	Relat. err.	5.08063e−6	5.08689e−6	5.08024e−6	5.07856e−5	5.08000e−6	660	2.984
1/40	Comp. sol.	0.30901709	0.58778544	0.80901725	0.95105682	1.00000032		
	Relat. err.	3.23607e−7	3.23247e−7	3.21378e−7	3.15439e−7	3.20000e−7	2637	56.016
(c) Block Jacobi iteration								
1/10	Comp. sol.	0.30904218	0.58783315	0.80908293	0.95113402	1.00008150		
	Relat. err.	8.15165e−5	8.14923e−5	8.15063e−5	8.14883e−5	8.15000e−5	330	0.641
1/20	Comp. sol.	0.30901856	0.58778824	0.80902110	0.95106135	1.00000508		
	Relat. err.	5.08063e−6	5.08689e−6	5.08024e−6	5.07856e−5	5.08000e−6	1317	7.281
1/40	Comp. sol.	0.30901709	0.58778544	0.80901725	0.95105682	1.00000032		
	Relat. err.	3.23607e−7	3.23247e−7	3.21378e−7	3.15439e−7	3.20000e−7	5270	166.313
	True sol.	0.30901699	0.58778525	0.80901699	0.95105652	1		

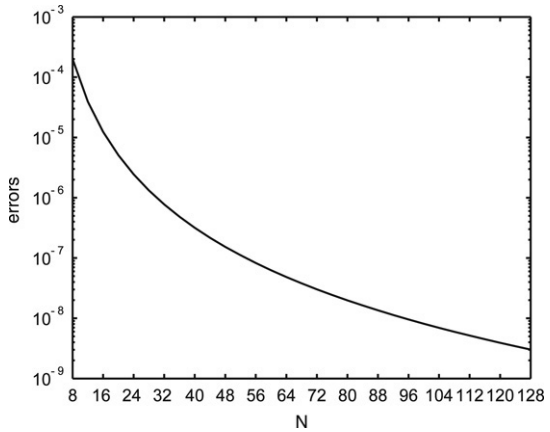

Fig. 6.2. The absolute maximum errors between U_h^* and u for $h = 1/N$.

To further demonstrate the accuracy of scheme (6.4), we calculate the L_α-error and the order of L_α-error of numerical solution U_h^* ($\alpha = 2, \infty$), which are defined by

$$\begin{aligned} &\mathrm{error}_2(h) = \|U_h^* - U\|, \qquad \mathrm{error}_\infty(h) = \|U_h^* - U\|_\infty, \\ &\mathrm{order}_\alpha(h) = \log_2\left(\frac{\mathrm{error}_\alpha(h)}{\mathrm{error}_\alpha(h/2)}\right), \quad \alpha = 2, \infty, \end{aligned} \tag{6.7}$$

where U denotes the true solution vector. In Tables 6.2 and 6.3, we list the L_α-error and the order of L_α-error, where the numerical solution is given by the block Gauss–Seidel iteration with the tolerance in (6.6). We see that the

Table 6.2
The comparison between scheme (6.4) and SFD in L_2-norm

h	Scheme (6.4)		SFD	
	$\text{error}_2(h)$	$\text{order}_2(h)$	$\text{error}_2(h)$	$\text{order}_2(h)$
1/4	1.62641965e−3	4.0279	5.44353829e−2	2.0622
1/8	9.97043905e−5	4.0067	1.30346208e−2	2.0154
1/16	6.20261613e−6	4.0017	3.22414850e−3	2.0038
1/32	3.87217570e−7	4.0004	8.03901337e−4	2.0010
1/64	2.41942020e−8	4.0002	2.00842171e−4	2.0002
1/128	1.51192645e−9		5.02022250e−5	

Table 6.3
The comparison between scheme (6.4) and SFD in L_∞-norm

h	Scheme (6.4)		SFD	
	$\text{error}_\infty(h)$	$\text{order}_\infty(h)$	$\text{error}_\infty(h)$	$\text{order}_\infty(h)$
1/4	3.25283892e−3	4.0279	1.08870753e−1	2.0622
1/8	1.99408765e−4	4.0067	2.60692390e−2	2.0154
1/16	1.24052313e−5	4.0017	6.44829644e−3	2.0038
1/32	7.74435076e−7	4.0004	1.60780254e−3	2.0010
1/64	4.83883977e−8	4.0002	4.01684310e−4	2.0002
1/128	3.02384962e−9		1.00404442e−4	

numerical solution U_h^* has the fourth-order accuracy in both L_2-norm and L_∞-norm. This coincides well with the analysis.

For comparison, we also solve (6.1) by the standard finite difference method (SFD) as in [16,17]. This method leads to a system of nonlinear algebraic equations of the form (6.4) with the matrices

$$A = \text{tridiag}(-I, A_1, -I), \qquad A_1 = \text{tridiag}(-1, 4, -1), \qquad B = I.$$

Thus, a similar block Gauss–Seidel iteration can be used in actual computation. The corresponding L_α-error and the order of L_α-error of numerical solution U_h^* ($\alpha = 2, \infty$) are also given in Tables 6.2 and 6.3. We see that the standard method possesses only the second-order accuracy.

To compare the convergence rates of iterations and time consumption, the number of required steps of iterations (Number of iter.) and CPU time for scheme (6.4) and SFD scheme are listed in Table 6.4. We see that with the same mesh size, the iteration for resolving scheme (6.4) converges slightly faster than that for SFD scheme. But scheme (6.4) costs more computational time than SFD scheme (except the case $h = 1/4$). This is reasonable, since more number of arithmetic operations are involved in scheme (6.4). However, we see from Tables 6.2–6.4 that for obtaining the numerical solution of the SFD scheme, with the L_2-error around 5.02022×10^{-5} or the L_∞-error around 1.00404×10^{-4}, we need to take $h = 1/128$, and so cost 10291.766 CPU seconds. In contrast, a more accurate numerical solution is provided by scheme (6.4) with $h = 1/16$. In this case, the L_2-error is 6.20262×10^{-6} and the L_∞-error is 1.24052×10^{-5}. But the corresponding cost is only 1.406 CPU seconds. The above comparisons clearly indicate that the presented scheme (6.4) is much more efficient than the standard finite difference method.

We end this section by giving a simple comment on the computational cost in Tables 6.1 and 6.4. At each step of the Picard iteration, we need to resolve a linear algebraic system of order $s = O(\frac{1}{h^2})$. When we use the Gauss elimination method to solve such system, the number of total operations at each step is about $s^3/3$. Therefore, if h is halved, then the number of total operations becomes $(4s)^3/3$. Accordingly, the CPU time with the mesh size $h/2$ is approximately 64 times the CPU time with mesh size h. Clearly, the CPU time of the Picard iteration presented in Table 6.1 agrees with the above analysis. In fact, the numbers of the Picard iterations for $h = \frac{1}{10}, \frac{1}{20}, \frac{1}{40}$ are the same. Therefore, the total CPU times for these three cases are nearly proportional to $\frac{1}{3h^6}, \frac{64}{3h^6}, \frac{4096}{3h^6}$, respectively. We can analyze the CPU time of block Gauss–Seidel iteration and block Jacobi iteration similarly. In these cases, the numbers of total operations at each step is about $O(s^2)$.

Table 6.4
The number of iterations and CPU time for scheme (6.4) and SFD

h	Scheme (6.4)		SFD	
	Number of iter.	CPU time (s)	Number of iter.	CPU time (s)
1/4	28	0.093	29	0.141
1/8	105	0.234	107	0.172
1/16	417	1.406	418	1.140
1/32	1 661	21.687	1 663	15.953
1/64	6 601	504.86	6 610	370.86
1/128	26 303	18 798.328	26 351	10 291.766

7. Concluding remarks

In this paper, we proposed a compact finite difference method for a fourth-order nonlinear elliptic problem, with the fourth-order accuracy. It also preserves monotonicity as the underlying continuous version and so fits the exact solution properly. We provided three monotone iterations with geometric convergence rates, for resolving the resulting nonlinear discrete systems. Moreover, all procedures do not require any monotonicity of involved nonlinear function and so essentially enlarge their applications. The numerical results coincide with the analysis very well.

In this work, we generalized the method of upper and lower solutions to multi-dimensional partial differential equations of high order. We also developed a technique for designing and analyzing compact and monotone finite difference schemes with high accuracy. They are very useful for accurate numerical simulations of many other nonlinear problems, such as numerical solution of the stream function form of the Navier–Stokes equations and so on.

Acknowledgements

Yuan-Ming Wang's work was supported in part by the NSF of China No. 10571059, the E-Institutes of Shanghai Municipal Education Commission No. E03004, and the Shanghai Priority Academic Discipline. Ben-Yu Guo's work was supported in part by the NSF of China No. 10471095, the SF of Shanghai No. 04JC14062, the fund of Chinese Education Ministry No. 20040270002, the Shanghai Leading Academic Discipline Project No. T0401, and the fund for E-Institutes of Shanghai Municipal Education Commission No. E03004.

Appendix A

In this appendix, we prove Lemmas 2.2, 3.2 and 3.3.

Proof of Lemma 2.2. We set

$$\phi^h_{i,j} = (x_i^2 + y_j^2)/24 = (i^2h_x^2 + j^2h_y^2)/24, \quad (i, j) \in \overline{\Omega}.$$

Clearly, $0 \le \phi^h_{i,j} \le 1/12$ for all $(i, j) \in \overline{\Omega}$. Moreover, a calculation shows that

$$\mathcal{L}_h\phi^h_{i,j} = -h_x^2, \quad (i, j) \in \Omega.$$

Next, let $M = \max_{(i,j)\in\Omega} |\mathcal{L}_h u^h_{i,j}|/h_x^2$, and $w^{\pm}_{i,j} = \pm u^h_{i,j} + M\phi^h_{i,j}$. We observe that

$$\mathcal{L}_h w^+_{i,j} = \mathcal{L}_h u^h_{i,j} - Mh_x^2, \qquad \mathcal{L}_h w^-_{i,j} = -\mathcal{L}_h u^h_{i,j} - Mh_x^2.$$

This implies $\mathcal{L}_h w^{\pm}_{i,j} \le 0$ for all $(i, j) \in \Omega$. Therefore, by virtue of Lemma 2.1,

$$\max_{(i,j)\in\Omega} w^{\pm}_{i,j} \le \max_{(i,j)\in\partial\Omega} w^{\pm}_{i,j} \le \max_{(i,j)\in\partial\Omega} (\pm u^h_{i,j}) + \max_{(i,j)\in\partial\Omega} (M\phi^h_{i,j}) \le \max_{(i,j)\in\partial\Omega} (\pm u^h_{i,j}) + M/12.$$

Since $\pm u^h_{i,j} \le w^{\pm}_{i,j}$, we have $\max_{(i,j)\in\Omega}(\pm u^h_{i,j}) \le \max_{(i,j)\in\partial\Omega}(\pm u^h_{i,j}) + M/12$, which implies the desired result (2.11). □

Proof of Lemma 3.2. Thanks to (2.8), we can check directly that

$$\sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1} u_{i,j}^h \mathcal{L}_h u_{i,j}^h = (5-\sigma^2)\sum_{i=1}^{N_x}\sum_{j=1}^{N_y-1}(u_{i,j}^h - u_{i-1,j}^h)^2 + (5\sigma^2-1)\sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y}(u_{i,j}^h - u_{i,j-1}^h)^2$$
$$+\frac{1+\sigma^2}{2}\sum_{i=1}^{N_x}\sum_{j=1}^{N_y}(u_{i,j}^h - u_{i-1,j-1}^h)^2 + \frac{1+\sigma^2}{2}\sum_{i=1}^{N_x}\sum_{j=0}^{N_y-1}(u_{i,j}^h - u_{i-1,j+1}^h)^2,$$

from which (3.29) follows immediately. □

Proof of Lemma 3.3. Clearly, $u_{i,j}^h = \sum_{p=1}^{i}(u_{p,j}^h - u_{p-1,j}^h)$, and so

$$(u_{i,j}^h)^2 \le \sum_{p=1}^{i} 1 \cdot \sum_{p=1}^{i}(u_{p,j}^h - u_{p-1,j}^h)^2 \le \frac{L_x}{h_x}\sum_{p=1}^{i}(u_{p,j}^h - u_{p-1,j}^h)^2.$$

Similarly,

$$(u_{i,j}^h)^2 \le \frac{L_x}{h_x}\sum_{p=i+1}^{N_x}(u_{p,j}^h - u_{p-1,j}^h)^2.$$

Putting the above two results together yields that

$$(u_{i,j}^h)^2 \le \frac{L_x}{2h_x}\sum_{p=1}^{N_x}(u_{p,j}^h - u_{p-1,j}^h)^2.$$

Therefore,

$$\|u^h\|^2 = \frac{h_x^2}{\sigma L_x L_y}\sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1}(u_{i,j}^h)^2 \le \frac{L_x}{2\sigma L_y}\sum_{i=1}^{N_x}\sum_{j=1}^{N_y-1}(u_{i,j}^h - u_{i-1,j}^h)^2.$$

In the same manner, we deduce that

$$\|u^h\|^2 \le \frac{\sigma L_y}{2L_x}\sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y}(u_{i,j}^h - u_{i,j-1}^h)^2.$$

Then the desired result (3.30) follows from the above two estimates. □

Appendix B

In this appendix, we estimate the smallest eigenvalue λ_0 of the generalized eigenvalue problem (2.18).

Theorem B.1. *If* $1/5 < \sigma^2 < 5$, *then* $\frac{2\sigma^*}{3\sigma L^*} \le \lambda_0 \le \frac{132(1+\sigma^2)}{121h_x^2}$, *where* σ^* *and* L^* *are given by* (3.28).

Proof. Let $A^{-1}B = (t_{i,j})$. Since $1/\lambda_0$ is the largest eigenvalue of $h_x^2 A^{-1}B$, we have from Forbenius theorem (see [24], p. 31) that

$$\lambda_0^{-1} \ge h_x^2 \min_i \sum_{j=1}^{\mathcal{N}} t_{i,j}. \tag{B.1}$$

Let $E = (1, 1, \ldots, 1)^T$ and $S = A^{-1}BE$. Then $S = (\sum_{j=1}^{\mathcal{N}} t_{1,j}, \sum_{j=1}^{\mathcal{N}} t_{2,j}, \ldots, \sum_{j=1}^{\mathcal{N}} t_{\mathcal{N},j})^T$ and by (2.13),

$$AS = BE \ge \frac{121}{24}E. \tag{B.2}$$

Let $S_{i_0} = \sum_{j=1}^{\mathcal{N}} t_{i_0,j} = \min_i \sum_{j=1}^{\mathcal{N}} t_{i,j}$, and let $(AS)_{i_0}$ stand for the i_0th component of AS. Then by (2.13), $(AS)_{i_0} \le \frac{11}{2}(1+\sigma^2)S_{i_0}$. This with (B.2) implies that $\frac{11}{2}(1+\sigma^2)S_{i_0} \ge \frac{121}{24}$. Thus, $\min_i \sum_{j=1}^{\mathcal{N}} t_{i,j} \ge \frac{121}{132(1+\sigma^2)}$, which with (B.1) leads to $\lambda_0 \le \frac{132(1+\sigma^2)}{121h_x^2}$.

Next, let Φ be the positive eigenvector corresponding to λ_0. Then $\Phi^T A\Phi = \lambda_0 h_x^2 \Phi^T B\Phi$. Furthermore by (3.35) and Lemmas 3.2 and 3.3,

$$h_x^2 \Phi^T B\Phi \le 6\sigma L_x L_y \|\Phi\|^2 \le \frac{3\sigma L^*}{2} L_x L_y |\Phi|_1^2 \le \frac{3\sigma L^*}{2\sigma^*} \Phi^T A\Phi.$$

Therefore, $\lambda_0 = \frac{\Phi^T A\Phi}{h_x^2 \Phi^T B\Phi} \ge \frac{2\sigma^*}{3\sigma L^*}$. This completes the proof. □

References

[1] A.R. Aftabizadeh, Existence and uniqueness theorems for fourth-order boundary value problems, J. Math. Anal. Appl. 116 (1986) 416–426.
[2] R. Agarwal, On fourth-order boundary value problems arising in beam analysis, Differential Integral Equations 2 (1989) 91–110.
[3] R.P. Agarwal, Y.-M. Wang, Some recent developments of the Numerov's method, Comput. Math. Appl. 42 (2001) 561–592.
[4] I. Babuska, J. Osborn, J. Pitkäranta, Analysis of mixed methods using mesh-dependent norms, Math. Comp. 35 (1980) 1039–1062.
[5] A. Berman, R. Plemmons, Nonnegative Matrix in the Mathematical Science, Academic Press, New York, 1979.
[6] P.G. Ciarlet, P.A. Raviart, A mixed finite element method for the biharmonic equation, in: C. de Boor (Ed.), Mathematical Aspect of Finite Elements in Partial Differential Equations, Academic Press, New York, 1974, pp. 125–145.
[7] Q.H. Choi, T. Jung, A fourth order nonlinear elliptic equation with jumping nonlinearity, Houston J. Math. 24 (1998) 735–756.
[8] M.A. Del Pino, R.F. Manasevich, Existence for a fourth-order nonlinear boundary problem under a two-parameter nonresonance condition, Proc. Amer. Math. Soc. 112 (1991) 81–86.
[9] G. Grinstein, A. Luther, Application of the renormalization group to phase transitions in disordered systems, Phys. Rev. B 13 (1976) 1329–1343.
[10] C.P. Gupta, Existence and uniqueness theorem for the bending of an elastic beam equation, Appl. Anal. 26 (1988) 289–304.
[11] A. Jennings, J.J. McKeown, Matrix Computations, John Wiley & Sons, New York, 1992, pp. 176–177.
[12] J. Li, Full-order convergence of a mixed finite element method for fourth-order elliptic equations, J. Math. Anal. Appl. 230 (1999) 329–349.
[13] R.Y. Ma, J.H. Zhang, S.M. Fu, The method of lower and upper solutions for fourth-order two-point boundary value problems, J. Math. Anal. Appl. 216 (1997) 416–422.
[14] A.M. Micheletti, A. Pistoia, Nontrivial solutions for some fourth-order semilinear elliptic problems, Nonlinear Anal. 34 (1998) 509–523.
[15] C.V. Pao, On fourth-order elliptic boundary value problems, Proc. Amer. Math. Soc. 128 (2000) 1023–1030.
[16] C.V. Pao, Numerical methods for fourth order nonlinear elliptic boundary value problems, Numer. Methods Partial Differential Equations 17 (2001) 347–368.
[17] C.V. Pao, X. Lu, Block monotone iterations for numerical solutions of fourth order nonlinear elliptic boundary value problems, SIAM J. Sci. Comput. 25 (2003) 164–185.
[18] J. Schroder, Fourth-order two-point boundary value problems: Estimate by two side bounds, Nonlinear Anal. 8 (1984) 107–114.
[19] W.F. Spotz, G.F. Carey, A high-order compact formulation for the 3d Poisson equation, Numer. Methods Partial Differential Equations 12 (1996) 235–243.
[20] G. Sutmann, B. Steffen, High-order compact solvers for the three-dimensional Poisson equation, J. Comput. Appl. Math. 187 (2006) 142–170.
[21] G. Tarantello, A note on a semilinear elliptic value problem, Differential Integral Equations 5 (1992) 561–565.
[22] S.P. Timoshenko, J.M. Gere, Theory of Elastic Stability, McGraw-Hill, New York, 1961.
[23] D. Uzunov, Theory of Critical Phenomena, World Scientific, Singapore, 1993.
[24] R.S. Varga, Matrix Iterative Analysis, Prentice-Hall, Englewood Cliffs, NJ, 1962.
[25] Y.-M. Wang, B.-Y. Guo, Monotone finite difference schemes for nonlinear systems with mixed quasi-monotonicity, J. Math. Anal. Appl. 267 (2002) 599–625.
[26] Y.-M. Wang, C.V. Pao, Time-delayed finite difference reaction-diffusion systems with nonquasimonotone functions, Numer. Math. 103 (2006) 485–513.
[27] Y. Yang, Fourth-order two-point boundary value problem, Proc. Amer. Math. Soc. 104 (1988) 175–180.
[28] J. Zhang, Multigrid method and fourth-order compact scheme for 2D Poisson equation with unequal mesh-size discretization, J. Comput. Phys. 179 (2002) 170–179.