

Draft of the paper in: *International Journal of Computer Mathematics*, Vol. **89**, 8 (2012), 998-1016.

Lagged diffusivity fixed point iteration for solving steady state reaction diffusion problems

Emanuele Galligani

Dipartimento di Matematica Pura e Applicata "G. Vitali"

Università degli Studi di Modena e Reggio Emilia

Via Campi 213/b, I-41125, Modena, Italy

Abstract. The paper concerns with the computational algorithms for a steady state reaction diffusion problem. A lagged diffusivity iterative algorithm is proposed for solving resulting system of quasilinear equations from a finite difference discretization. The convergence of the algorithm is discussed and the numerical results show the efficiency of this algorithm.

Key Words: Nonlinear problems, lagging diffusivity, Newton method, Arithmetic Mean method.

2010 AMS Classification: 65H10, 65N06, 65N22

C.R. Categories: G.1.5., G.1.8.

1 Introduction

We consider the problem of finding a solution of the system of nonlinear equations

$$\mathbf{F}(\mathbf{u}) = \mathbf{0},$$

where $F : \Omega \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a continuously differentiable mapping.

We are interested in large scale systems for which the Jacobian of \mathbf{F} is not available or is difficult to compute.

Particularly, we consider the solution of the system of quasilinear equations

$$\mathbf{F}(\mathbf{u}) \equiv A(\mathbf{u})\mathbf{u} + \mathbf{G}(\mathbf{u}) - \mathbf{s} = \mathbf{0}, \quad (1)$$

where $A(\mathbf{u})$ is a real matrix of order n and $\mathbf{G} : \Omega \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a continuously differentiable mapping.

The systems of the form (1) appear in many problems of practical interest. For example, in problems related with the study of reaction and diffusion processes that can be described by nonlinear partial differential equations of elliptic type (e.g., [1], [11], [14]).

2 Statement of the problem

We are concerned with the numerical solution of the steady state reaction diffusion problem by the finite difference method.

In the rectangular domain R with boundary ∂R of the x, y plane, we consider the reaction diffusion equation

$$-\operatorname{div}(\sigma(x, y, u)\nabla u(x, y)) + g(x, y, u) = s(x, y), \quad (x, y) \in R, \quad (2)$$

supplemented by the Dirichlet boundary condition

$$u(x, y) = 0, \quad (x, y) \in \partial R. \quad (3)$$

The functions $\sigma(x, y, u)$, $g(x, y, u)$ and $s(x, y)$ are assumed to satisfy the following **Smoothness Conditions** for x and y in $R \cup \partial R$ and u in a neighbourhood of a solution of problem (2)–(3). We assume that this model problem has a solution.

- (i) The functions $\sigma(x, y, u)$ and $g(x, y, u)$ are piecewise continuous in x and y and continuous in u ; the source term $s(x, y)$ is piecewise continuous in x, y .
- (ii) There exist two positive constants σ_{\min} and σ_{\max} such that

$$0 < \sigma_{\min} \leq \sigma(x, y, u) \leq \sigma_{\max},$$

uniformly in x, y and u .

- (iii) For fixed x and y , the function $\sigma(x, y, u)$ is locally Lipschitz continuous at u (uniformly in x and y), with constant $\Lambda > 0$.
- (iv) For fixed x and y , the function $g(x, y, u)$ is uniformly monotone¹ at u (uniformly in x and y), with constant $c > 0$ and is continuously differentiable at u .

There exist various techniques for discretizing the problem (2)–(3). The use of the truncated Taylor series to represent the derivative in (2) provides more insight into the nature of the truncation error that arises when the continuous model equation (2) is replaced by a discrete set of finite difference equations. Using the Taylor series approach, equation (2) will be solved with the following standard finite difference scheme.

¹See, e.g., [16, p. 141].

If the domain R is a the unit square we superimpose on R a uniform grid of points R_h . More precisely, R_h is defined in this manner. In the plane Oxy we draw a system of straight lines parallel to the coordinate axes: $x = x_i$ and $y = y_j$ with mesh spacings uniform in each coordinate direction. We denote the intersection of these lines as mesh points or grid points. The coordinates of the interior mesh points (x_i, y_j) of R_h are given by $x_i = i \Delta x$ and $y_j = j \Delta y$ with $i = 1, \dots, N_x$ and $j = 1, \dots, N_y$ and $\Delta x = 1/(N_x + 1)$, $\Delta y = 1/(N_y + 1)$. For simplicity, we assume $\Delta x = \Delta y = h$ and consequently $N_x = N_y = N$. At the mesh points of $\bar{R}_h = R_h \cup \partial R_h$, the function $u(x_i, y_j)$ is to be approximated by a grid function u_{ij} . In order to approximate partial derivatives we shall make use of various difference quotients of grid functions. The forward, backward and centered difference quotients with respect to x and y of the grid function $\{u_{ij}\}$ at the mesh point (x_i, y_j) are, respectively

$$\begin{aligned}\Delta_x u_{ij} &= \frac{u_{i+1j} - u_{ij}}{h}, & \Delta_y u_{ij} &= \frac{u_{ij+1} - u_{ij}}{h}, \\ \nabla_x u_{ij} &= \frac{u_{ij} - u_{i-1j}}{h}, & \nabla_y u_{ij} &= \frac{u_{ij} - u_{ij-1}}{h}, \\ \delta_x u_{ij} &= \frac{1}{2}(\Delta_x u_{ij} + \nabla_x u_{ij}), & \delta_y u_{ij} &= \frac{1}{2}(\Delta_y u_{ij} + \nabla_y u_{ij}).\end{aligned}$$

This convenient notation was introduced by Courant, Friedrichs and Lewy ([3]). A particular important case is the centered second difference quotients which can be written as

$$\begin{aligned}\delta_x^2 u_{ij} &= \frac{1}{h^2}(u_{i-1j} - 2u_{ij} + u_{i+1j}) = \nabla_x \Delta_x u_{ij} = \Delta_x \nabla_x u_{ij}, \\ \delta_y^2 u_{ij} &= \frac{1}{h^2}(u_{ij-1} - 2u_{ij} + u_{ij+1}) = \nabla_y \Delta_y u_{ij} = \Delta_y \nabla_y u_{ij}.\end{aligned}$$

A centered finite difference approximation of (2) at the mesh point (x_i, y_j) may be written as

$$\begin{aligned}-\frac{1}{h} \left[\sigma_{i+1j}(u_{i+1j}) \frac{u_{i+1j} - u_{ij}}{h} - \sigma_{ij}(u_{ij}) \frac{u_{ij} - u_{i-1j}}{h} \right] - \\ -\frac{1}{h} \left[\sigma_{ij+1}(u_{ij+1}) \frac{u_{ij+1} - u_{ij}}{h} - \sigma_{ij}(u_{ij}) \frac{u_{ij} - u_{ij-1}}{h} \right] + \\ + g_{ij}(u_{ij}) = s(x_i, y_j),\end{aligned}\tag{4}$$

for $i, j = 1, \dots, N$.

The Dirichlet boundary condition (3) produces the following conditions

$$\begin{aligned}u_{i0} = u_{iN+1} = 0, & \quad i = 1, \dots, N, \\ u_{0j} = u_{N+1j} = 0, & \quad j = 1, \dots, N.\end{aligned}\tag{5}$$

The total number of difference equations is $n = N \times N$.

Here we have denoted

$$\sigma_{lp}(u_{ij}) \equiv \sigma(x_l, y_p u_{ij}), \quad g_{lp}(u_{ij}) \equiv g(x_l, y_p u_{ij}).$$

Here, $\|\mathbf{u}\|_h$ is the discrete $l_2(R_h)$ norm of grid functions $\{u_{ij}\}$:

$$\|\mathbf{u}\|_h = \left(\sum_{i=1}^N \sum_{j=1}^N h^2 |u_{ij}|^2 \right)^{1/2}. \quad (6)$$

- (II) At each mesh point (x_i, y_j) of \bar{R}_h the backward difference quotients with respect to x and to y of a grid function $\{u_{ij}\}$ belonging to Ω are bounded. The bound is independent of h , but depends on $\|\mathbf{s}\|_h$.

It is possible to prove the following propositions (see, e.g., [12], pp. 37–43).

Proposition 1. No solution of the matrix equation (1) generated by the discretization of the model problem (2)–(3) can lie outside the ball Ω , where

$$\rho = \frac{(\|\mathbf{G}(\mathbf{0})\|_h + \|\mathbf{s}\|_h)}{(\tilde{c} + c)},$$

with \tilde{c} and c positive constants independent of h for which

$$\begin{aligned} \langle A(\mathbf{w})\mathbf{u}, \mathbf{u} \rangle &\geq \tilde{c}\|\mathbf{u}\|_h^2, \\ \langle \mathbf{G}(\mathbf{u}) - \mathbf{G}(\mathbf{v}), \mathbf{u} - \mathbf{v} \rangle &\geq c\|\mathbf{u} - \mathbf{v}\|_h^2, \end{aligned}$$

for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \Omega$.

Proposition 2. Let \mathbf{u} be a solution of the system of 1-D finite difference equation (1). Then

$$\frac{|u_i - u_{i-1}|}{h} \leq \frac{2c_0 + \sqrt{c_0}(\sigma_{\max}\|\mathbf{u}\|_h)^{1/2}}{\sigma_{\min}}, \quad i = 1, \dots, N + 1,$$

where

$$c_0 = \|\mathbf{G}(\mathbf{u}) + \mathbf{s}\|_h.$$

Definition 1. We say that the *grid functions* $\{u_{ij}\}$ defined on \bar{R}_h and vanishing on ∂R_h , i.e., the vectors $\{\mathbf{u}\}$ representing in the interior mesh points R_h these grid functions, satisfy **Property S** on Ω if they are uniformly bounded in Ω and have uniformly bounded backward difference quotients $\nabla_x u_{ij}$ and $\nabla_y u_{ij}$ at each mesh point (x_i, y_j) of \bar{R}_h .

3 Monotonicity of the mapping F

We begin with two useful lemmas which establish the discrete analogs of two continuous $L_2(R)$ inner products used frequently in the study of elliptic partial differential equations.

Lemma 1. Let $\{u_{ij}\}, \{v_{ij}\}, \{w_{ij}\}$ be three grid functions defined at the mesh points (x_i, y_j) of \bar{R}_h , $i, j = 0, \dots, N + 1$. Then,

$$\begin{aligned} h \sum_{i=1}^N \left(-\sigma_{i+1j}(w_{i+1j}) \frac{u_{i+1j} - u_{ij}}{h^2} + \sigma_{ij}(w_{ij}) \frac{u_{ij} - u_{i-1j}}{h^2} \right) v_{ij} &= \\ &= \sum_{i=1}^N \sigma_{ij}(w_{ij}) \nabla_x u_{ij} \nabla_x v_{ij} h + \sigma_{1j}(w_{1j}) \frac{u_{1j} - u_{0j}}{h} v_{0j} - \\ &\quad - \sigma_{N+1j}(w_{N+1j}) \frac{u_{N+1j} - u_{Nj}}{h} v_{Nj}, \end{aligned} \quad (7)$$

and

$$\begin{aligned} h \sum_{i=1}^N \left(-\sigma_{ij+1}(w_{ij+1}) \frac{u_{ij+1} - u_{ij}}{h^2} + \sigma_{ij}(w_{ij}) \frac{u_{ij} - u_{i-1j}}{h^2} \right) v_{ij} &= \\ &= \sum_{i=1}^N \sigma_{ij}(w_{ij}) \nabla_y u_{ij} \nabla_y v_{ij} h + \sigma_{i1}(w_{i1}) \frac{u_{i1} - u_{i0}}{h} v_{i0} - \\ &\quad - \sigma_{iN+1}(w_{iN+1}) \frac{u_{iN+1} - u_{iN}}{h} v_{iN}. \end{aligned} \quad (8)$$

Proof. We have

$$\begin{aligned} h \sum_{i=1}^N \left(-\sigma_{i+1j}(w_{i+1j}) \frac{u_{i+1j} - u_{ij}}{h^2} + \sigma_{ij}(w_{ij}) \frac{u_{ij} - u_{i-1j}}{h^2} \right) v_{ij} &= \\ &= \sigma_{1j}(w_{1j}) \frac{u_{1j} - u_{0j}}{h} v_{1j} + \sum_{i=2}^N \sigma_{ij}(w_{ij}) \frac{u_{ij} - u_{i-1j}}{h} v_{ij} - \\ &\quad - \sum_{i=1}^{N-1} \sigma_{i+1j}(w_{i+1j}) \frac{u_{i+1j} - u_{ij}}{h} v_{ij} - \sigma_{N+1j}(w_{N+1j}) \frac{u_{N+1j} - u_{Nj}}{h} v_{Nj} \\ &= \sigma_{1j}(w_{1j}) \frac{u_{1j} - u_{0j}}{h} v_{1j} + \sum_{i=2}^N \sigma_{ij}(w_{ij}) \frac{u_{ij} - u_{i-1j}}{h} v_{ij} - \\ &\quad - \sum_{i=2}^N \sigma_{ij}(w_{ij}) \frac{u_{ij} - u_{i-1j}}{h} v_{i-1j} - \sigma_{N+1j}(w_{N+1j}) \frac{u_{N+1j} - u_{Nj}}{h} v_{Nj} \\ &= \sigma_{1j}(w_{1j}) \frac{u_{1j} - u_{0j}}{h} v_{1j} + \sum_{i=2}^N \sigma_{ij}(w_{ij}) \frac{u_{ij} - u_{i-1j}}{h} \frac{v_{ij} - v_{i-1j}}{h} h - \\ &\quad - \sigma_{N+1j}(w_{N+1j}) \frac{u_{N+1j} - u_{Nj}}{h} v_{Nj}, \end{aligned}$$

then, we have the result (7). Similarly we obtain formula (8). \sharp

Lemma 2. Let $\{u_{ij}\}, \{v_{ij}\}, \{w_{ij}\}$ be three grid functions defined at the mesh points (x_i, y_j) of \bar{R}_h , $i, j = 0, \dots, N + 1$, which are zero at the boundary ∂R_h of \bar{R}_h , i.e., satisfy the condition (5) on ∂R_h .

Then, we have the following expression for the discrete $l_2(R_h)$ inner product of the vectors $A(\mathbf{w})\mathbf{u}$ and \mathbf{v} :

$$\begin{aligned} \langle A(\mathbf{w})\mathbf{u}, \mathbf{v} \rangle &= h^2 \sum_{j=1}^N \sum_{i=1}^N \sigma_{ij}(w_{ij}) \nabla_h u_{ij} \nabla_h v_{ij} + \\ &+ \sum_{j=1}^N \sigma_{N+1j}(w_{N+1j}) u_{Nj} \overline{v_{Nj}} + \sum_{i=1}^N \sigma_{iN+1}(w_{iN+1}) u_{iN} v_{iN}, \end{aligned} \quad (9)$$

where $\nabla_h u_{ij}$ denotes the *gradient* of the grid function $\{u_{ij}\}$ at the mesh point (x_i, y_j) (see formulae (10) and (11)).

Proof. Since the discrete $l_2(R_h)$ inner product of two grid functions $\{u_{ij}\}$ and $\{v_{ij}\}$ which are zero on the boundary ∂R is defined by

$$\langle \mathbf{u}, \mathbf{v} \rangle = h^2 \sum_{i=1}^N \sum_{j=1}^N u_{ij} v_{ij}, \quad (10)$$

we have

$$\begin{aligned} \langle A(\mathbf{w})\mathbf{u}, \mathbf{v} \rangle &= h \sum_{j=1}^N \left[h \sum_{i=1}^N \left(-\sigma_{i+1j}(w_{i+1j}) \frac{u_{i+1j} - u_{ij}}{h^2} + \sigma_{ij}(w_{ij}) \frac{u_{ij} - u_{i-1j}}{h^2} \right) v_{ij} \right] + \\ &+ h \sum_{i=1}^N \left[h \sum_{j=1}^N \left(-\sigma_{ij+1}(w_{ij+1}) \frac{u_{ij+1} - u_{ij}}{h^2} + \sigma_{ij}(w_{ij}) \frac{u_{ij} - u_{ij-1}}{h^2} \right) v_{ij} \right] \\ &= h \sum_{j=1}^N \left[\sum_{i=1}^N \sigma_{ij}(w_{ij}) \nabla_x u_{ij} \nabla_x v_{ij} h + \frac{1}{h} \sigma_{N+1j}(w_{N+1j}) u_{Nj} v_{Nj} \right] + \\ &+ h \sum_{i=1}^N \left[\sum_{j=1}^N \sigma_{ij}(w_{ij}) \nabla_y u_{ij} \nabla_y v_{ij} h + \frac{1}{h} \sigma_{iN+1}(w_{iN+1}) u_{iN} v_{iN} \right]. \end{aligned}$$

We have used formulae (7) and (8) and the hypothesis that the grid functions $\{u_{ij}\}$ and $\{v_{ij}\}$ are zero at the boundary ∂R , i.e., satisfy the condition (5) on ∂R .

We define the gradient of the grid function $\{u_{ij}\}$ at (i, j) entry to be

$$\nabla_h u_{ij} \equiv (\nabla_x u_{ij}, \nabla_y u_{ij}), \quad i, j = 1, \dots, N. \quad (11)$$

Thus, we can write

$$\nabla_h u_{ij} \cdot \nabla_h v_{ij} = \nabla_x u_{ij} \nabla_x v_{ij} + \nabla_y u_{ij} \nabla_y v_{ij}.$$

Therefore, we obtain the result (9). \sharp

Remark. While the grid function $\{u_{ij}\}$ is defined on the whole mesh region \bar{R}_h , the vector $\mathbf{u} \in \mathbb{R}^n$ represents the grid function $\{u_{ij}\}$ defined only on the interior mesh points R_h .

Now we want to prove that under suitable conditions the mapping (1) is uniformly monotone on the closed ball Ω (see, Property (I)).

Let $\{u_{ij}\}$ and $\{v_{ij}\}$ be two grid functions belonging to Ω , i.e., two solutions of the nonlinear system (1). For the condition (5), they are represented by the vectors \mathbf{u} and \mathbf{v} of n components.

As consequence of the identity

$$A(\mathbf{u})\mathbf{u} - A(\mathbf{v})\mathbf{v} = A(\mathbf{u})(\mathbf{u} - \mathbf{v}) + (A(\mathbf{u}) - A(\mathbf{v}))\mathbf{v}, \quad (12)$$

we have, by (1)

$$\mathbf{F}(\mathbf{u}) - \mathbf{F}(\mathbf{v}) = A(\mathbf{u})(\mathbf{u} - \mathbf{v}) + (A(\mathbf{u}) - A(\mathbf{v}))\mathbf{v} + \mathbf{G}(\mathbf{u}) - \mathbf{G}(\mathbf{v}). \quad (13)$$

Thus, from (13), the discrete $l_2(R_h)$ inner product

$$\begin{aligned} \langle \mathbf{F}(\mathbf{u}) - \mathbf{F}(\mathbf{v}), \mathbf{u} - \mathbf{v} \rangle &= \langle A(\mathbf{u})(\mathbf{u} - \mathbf{v}), \mathbf{u} - \mathbf{v} \rangle + \\ &+ \langle (A(\mathbf{u}) - A(\mathbf{v}))\mathbf{v}, \mathbf{u} - \mathbf{v} \rangle + \\ &+ \langle \mathbf{G}(\mathbf{u}) - \mathbf{G}(\mathbf{v}), \mathbf{u} - \mathbf{v} \rangle. \end{aligned} \quad (14)$$

Using (9), we have

$$\begin{aligned} \langle A(\mathbf{u})(\mathbf{u} - \mathbf{v}), \mathbf{u} - \mathbf{v} \rangle &= h^2 \sum_{j=1}^N \sum_{i=1}^N \sigma_{ij}(u_{ij}) |\nabla_h(u_{ij} - v_{ij})|^2 + \\ &+ \sum_{j=1}^N \sigma_{N+1j}(u_{N+1j}) |u_{Nj} - v_{Nj}|^2 + \\ &+ \sum_{i=1}^N \sigma_{iN+1}(u_{iN+1}) |u_{iN} - v_{iN}|^2. \end{aligned} \quad (15)$$

Assumption (iv) implies that, for all grid functions \mathbf{u} and \mathbf{v} belonging to Ω , there exists a positive constant c such that

$$(g_{ij}(u_{ij}) - g_{ij}(v_{ij}))(u_{ij} - v_{ij}) \geq c(u_{ij} - v_{ij})^2, \quad (16)$$

for all $i, j = 1, \dots, N$ (see [16, p. 141]). The constant c is independent of h .

Thus, for the discrete $l_2(R_h)$ inner product (see definition (10))

$$\langle \mathbf{G}(\mathbf{u}) - \mathbf{G}(\mathbf{v}), \mathbf{u} - \mathbf{v} \rangle = h^2 \sum_{j=1}^N \sum_{i=1}^N (g_{ij}(u_{ij}) - g_{ij}(v_{ij}))(u_{ij} - v_{ij}),$$

and from (16), we have the inequality (see, definition (6))

$$\langle \mathbf{G}(\mathbf{u}) - \mathbf{G}(\mathbf{v}), \mathbf{u} - \mathbf{v} \rangle \geq c \|\mathbf{u} - \mathbf{v}\|_h^2. \quad (17)$$

The next step is to find a bound for $|\langle (A(\mathbf{u}) - A(\mathbf{v}))\mathbf{v}, \mathbf{u} - \mathbf{v} \rangle|$. Using (9), we have

$$\begin{aligned}
|\langle (A(\mathbf{u}) - A(\mathbf{v}))\mathbf{v}, \mathbf{u} - \mathbf{v} \rangle| &= \left| \sum_{j=1}^N \sum_{i=1}^N (\sigma_{ij}(u_{ij}) - \sigma_{ij}(v_{ij})) \nabla_x v_{ij} \times \right. \\
&\quad \times \nabla_x (u_{ij} - v_{ij}) h^2 + \sum_{j=1}^N \sum_{i=1}^N (\sigma_{ij}(u_{ij}) - \sigma_{ij}(v_{ij})) \nabla_y v_{ij} \nabla_y (u_{ij} - v_{ij}) h^2 + \\
&\quad + \sum_{j=1}^N (\sigma_{N+1j}(u_{N+1j}) - \sigma_{N+1j}(v_{N+1j})) v_{Nj} (u_{Nj} - v_{Nj}) + \\
&\quad \left. + \sum_{i=1}^N (\sigma_{iN+1}(u_{iN+1}) - \sigma_{iN+1}(v_{iN+1})) v_{iN} (u_{iN} - v_{iN}) \right|. \tag{18}
\end{aligned}$$

Specifically, assumption (iii) says that, for given grid functions \mathbf{u} and \mathbf{v} belonging to Ω , there exists some \mathbf{w} on the line between \mathbf{u} and \mathbf{v} and a positive constant $\Lambda(w_{ij})$ such that

$$|\sigma_{ij}(u_{ij}) - \sigma_{ij}(v_{ij})| \leq \Lambda(w_{ij}) |u_{ij} - v_{ij}|, \tag{19}$$

for all $i, j = 1, \dots, N + 1$.

We define

$$\Lambda = \sup_{w_{ij}} \Lambda(w_{ij}); \tag{20}$$

here, Λ is independent of h and $\{w_{ij}\} \in \Omega$.

Property (II) assures that there exists a constant $\beta > 0$ such that

$$|\nabla_x v_{ij}| \leq \beta \quad \text{and} \quad |\nabla_y v_{ij}| \leq \beta, \tag{21}$$

for all $i, j = 1, \dots, N + 1$ and all grid function $\{v_{ij}\}$ belonging to Ω . β is independent of h .

Now, we may apply the inequalities (19) and (21) and the definition (20) to estimate the expression (18), then

$$\begin{aligned}
|\langle (A(\mathbf{u}) - A(\mathbf{v}))\mathbf{v}, \mathbf{u} - \mathbf{v} \rangle| &\leq \Lambda \beta \sum_{j=1}^N \sum_{i=1}^N |u_{ij} - v_{ij}| |\nabla_x (u_{ij} - v_{ij})| h^2 + \\
&\quad + \Lambda \beta \sum_{j=1}^N \sum_{i=1}^N |u_{ij} - v_{ij}| |\nabla_y (u_{ij} - v_{ij})| h^2 + \\
&\quad + \Lambda \sum_{j=1}^N |u_{N+1j} - v_{N+1j}| |v_{Nj}| |u_{Nj} - v_{Nj}| + \\
&\quad + \Lambda \sum_{i=1}^N |u_{iN+1} - v_{iN+1}| |v_{iN}| |u_{iN} - v_{iN}|.
\end{aligned}$$

The last two terms are zero because the grid functions $\{u_{ij}\}$ and $\{v_{ij}\}$ belonging to Ω are bounded and satisfy the boundary condition (5).

Using a well known technical trick involving the arithmetic mean and geometric mean inequality of positive numbers ($\sqrt{ab} \leq a/2 + b/2$), we obtain

$$\begin{aligned}
| \langle (A(\mathbf{u}) - A(\mathbf{v}))\mathbf{v}, \mathbf{u} - \mathbf{v} \rangle | &\leq \Lambda\beta \sum_{j=1}^N \sum_{i=1}^N h^2 \left(\frac{|u_{ij} - v_{ij}|}{\sqrt{\alpha}} \sqrt{\alpha} |\nabla_x(u_{ij} - v_{ij})| + \right. \\
&\quad \left. + \frac{|u_{ij} - v_{ij}|}{\sqrt{\alpha}} \sqrt{\alpha} |\nabla_y(u_{ij} - v_{ij})| \right) \\
&\leq \frac{\Lambda\beta}{2} \sum_{j=1}^N \sum_{i=1}^N h^2 \left(\frac{|u_{ij} - v_{ij}|^2}{\alpha} + \alpha |\nabla_x(u_{ij} - v_{ij})|^2 \right) + \\
&\quad + \frac{\Lambda\beta}{2} \sum_{j=1}^N \sum_{i=1}^N h^2 \left(\frac{|u_{ij} - v_{ij}|^2}{\alpha} + \alpha |\nabla_y(u_{ij} - v_{ij})|^2 \right), \tag{22}
\end{aligned}$$

where α is as yet an undetermined positive number.

It now follows from (14) that

$$\begin{aligned}
\langle \mathbf{F}(\mathbf{u}) - \mathbf{F}(\mathbf{v}), \mathbf{u} - \mathbf{v} \rangle &\geq \langle A(\mathbf{u})(\mathbf{u} - \mathbf{v}), \mathbf{u} - \mathbf{v} \rangle + \\
&\quad + \langle \mathbf{G}(\mathbf{u}) - \mathbf{G}(\mathbf{v}), \mathbf{u} - \mathbf{v} \rangle - \\
&\quad - | \langle (A(\mathbf{u}) - A(\mathbf{v}))\mathbf{v}, \mathbf{u} - \mathbf{v} \rangle |, \tag{23}
\end{aligned}$$

and from (15), (17) and (22) and assumption (ii), formula (23) becomes

$$\begin{aligned}
\langle \mathbf{F}(\mathbf{u}) - \mathbf{F}(\mathbf{v}), \mathbf{u} - \mathbf{v} \rangle &\geq \sum_{j=1}^N \sum_{i=1}^N h^2 \sigma_{\min} (|\nabla_x(u_{ij} - v_{ij})|^2 + |\nabla_y(u_{ij} - v_{ij})|^2) + \\
&\quad + \sigma_{\min} \left(\sum_{j=1}^N |u_{Nj} - v_{Nj}|^2 + \sum_{i=1}^N |u_{iN} - v_{iN}|^2 \right) + \\
&\quad + c \|\mathbf{u} - \mathbf{v}\|_h^2 - \Lambda\beta \sum_{j=1}^N \sum_{i=1}^N h^2 \frac{|u_{ij} - v_{ij}|^2}{\alpha} - \\
&\quad - \frac{\Lambda\beta}{2} \sum_{j=1}^N \sum_{i=1}^N h^2 \alpha (|\nabla_x(u_{ij} - v_{ij})|^2 + |\nabla_y(u_{ij} - v_{ij})|^2) \\
&= \sigma_{\min} \left(\sum_{j=1}^N |u_{Nj} - v_{Nj}|^2 + \sum_{i=1}^N |u_{iN} - v_{iN}|^2 \right) + \\
&\quad + \sum_{j=1}^N \sum_{i=1}^N h^2 \left(\sigma_{\min} - \frac{\Lambda\beta\alpha}{2} \right) (|\nabla_x(u_{ij} - v_{ij})|^2 + \\
&\quad + |\nabla_y(u_{ij} - v_{ij})|^2) + \left(c - \frac{\Lambda\beta}{\alpha} \right) \|\mathbf{u} - \mathbf{v}\|_h^2. \tag{24}
\end{aligned}$$

If we set

$$\alpha = \frac{2\sigma_{\min}}{\Lambda\beta},$$

the inequality (24) can be written

$$\langle \mathbf{F}(\mathbf{u}) - \mathbf{F}(\mathbf{v}), \mathbf{u} - \mathbf{v} \rangle \geq \left(c - \frac{\Lambda\beta}{\alpha}\right) \|\mathbf{u} - \mathbf{v}\|_h^2. \quad (25)$$

When

$$c - \frac{\Lambda^2\beta^2}{2\sigma_{\min}} > 0,$$

or

$$\frac{\Lambda^2\beta^2}{2c\sigma_{\min}} < 1, \quad (26)$$

inequality (25) shows that the mapping $\mathbf{F}(\mathbf{u})$ is uniformly monotone on Ω .

When inequality (25) holds, there exists only one solution \mathbf{u}^* of the system $\mathbf{F}(\mathbf{u}) = \mathbf{0}$ in Ω .

Indeed, suppose that $\tilde{\mathbf{u}}^*$ is some other solution of $\mathbf{F}(\mathbf{u}) = \mathbf{0}$ in Ω , $\tilde{\mathbf{u}}^* \neq \mathbf{u}^*$. Then, since the mapping $\mathbf{F}(\mathbf{u})$ is uniformly monotone on Ω and (25) holds with $c - (\Lambda^2\beta^2)/(2\sigma_{\min}) > 0$, we have

$$0 = \langle \mathbf{F}(\mathbf{u}^*) - \mathbf{F}(\tilde{\mathbf{u}}^*), \mathbf{u}^* - \tilde{\mathbf{u}}^* \rangle \geq \left(c - \frac{\Lambda^2\beta^2}{2\sigma_{\min}}\right) \|\mathbf{u}^* - \tilde{\mathbf{u}}^*\|_h^2 > 0,$$

which is a contradiction. Hence \mathbf{u}^* is the only solution of $\mathbf{F}(\mathbf{u}) = \mathbf{0}$ in Ω .

Inequality (26) says that the best chances for the applicability of the algorithm to be defined in next section, are assured when the local Lipschitz constant of $\sigma(x, y, u)$ is small and the constant c (in condition (iv)) on the monotonicity of $g(x, y, u)$ is large.

4 Lagged diffusivity fixed point iteration

We will investigate the solvability of the system of nonlinear difference equations (1) under the assumption that the computation of the Jacobian matrix of $\mathbf{F}(\mathbf{u})$ is not convenient.

For solving this system the easiest and may be the most common method is *to lag* part of the nonlinear term in (1) (see [20]).

We will show that under weak restrictions (i)–(iv) imposed on the functions $\sigma(x, y, u)$ and $g(x, y, u)$ the problem (2)–(3) can be solved via a sequence of systems of weakly nonlinear difference equations where only g but not σ depends on the approximate solution \mathbf{u} of $u(x, y)$.

That is, we will apply the so called *lagged diffusivity fixed point iteration* procedure.

Many numerical experiments have shown the following facts.

1. In the special case in which the diffusion term in (2) dominates the reaction term, the lagged diffusivity procedure is an efficient and robust method, even if only linearly convergent, for solving problems with a highly nonlinear differential operator $-\text{div}(\sigma\nabla u)$. (For example, this may happen for nonlinear diffusion models in image denoising [22]).

When this operator is discretized, the matrix $A(\mathbf{u})$ has highly varying coefficients and Newton's method for solving the nonlinear system (1) does not work satisfactorily, in the sense that its domain of convergence is very small.

2. In the general case, it may be noted that the steeper the slope of g (in equation (2)) and the flatter that of σ relative to that of g , the better are the chances for the applicability of the lagged diffusivity procedure.

Specifically, let $\mathbf{u}^{(0)} \in \Omega$ be an initial approximation to the solution of \mathbf{u}^* of the system (1). The mapping

$$\mathbf{F}_0(\mathbf{u}) \equiv A(\mathbf{u}^{(0)})\mathbf{u} + \mathbf{G}(\mathbf{u}) - \mathbf{s} = \mathbf{0},$$

represents the discretized weakly nonlinear difference system which can be solved *inexactly* with a Newton iterative method. This solution $\mathbf{u}^{(1)}$ produces a residual $\mathbf{F}_0(\mathbf{u}^{(1)})$ with bound $\|\mathbf{F}_0(\mathbf{u}^{(1)})\| \leq \varepsilon_1$. Here $\|\cdot\|$ is the Euclidean norm.

Thus, $\mathbf{u}^{(1)}$ is a new approximation to \mathbf{u}^* . Now, $\mathbf{u}^{(1)}$ becomes the current iterate to generate the new iterate $\mathbf{u}^{(2)}$, and so on.

In general, if $\mathbf{u}^{(\nu)}$ is an estimate of the solution \mathbf{u}^* of (1), we will determine a new estimate of \mathbf{u}^* by solving the weakly nonlinear system

$$\mathbf{F}_\nu(\mathbf{u}) \equiv A(\mathbf{u}^{(\nu)})\mathbf{u} + \mathbf{G}(\mathbf{u}) - \mathbf{s} = \mathbf{0}. \quad (27)$$

An approximate solution of the weakly nonlinear system (27) is computed by a Newton iterative method in such a way that its solution $\mathbf{u}^{(\nu+1)}$ will be accepted if the residual $\mathbf{F}_\nu(\mathbf{u}^{(\nu+1)})$ satisfies the condition

$$\|\mathbf{F}_\nu(\mathbf{u}^{(\nu+1)})\| \leq \varepsilon_{\nu+1}, \quad (28)$$

where $\varepsilon_{\nu+1}$ is a given tolerance such that $\varepsilon_{\nu+1} \rightarrow 0$ for $\nu \rightarrow \infty$.

If such suitable solution $\mathbf{u}^{(\nu+1)}$ is found, we say that *the algorithm does not break down*.

The special form of the equations in (1) and the large dimension of the system (1) suggest to use an iterative method for solving at each *outer* iteration $\nu + 1$ the system (27). (See, e.g., [16, §§7.4, 10.3], [4], [5], [8]).

Now, it is important to note that the iterate $\mathbf{u}^{(\nu+1)}$ is the solution of a weakly nonlinear reaction diffusion equation (whose diffusivity σ depends on the previous iterate) with inhomogeneous term $-\mathbf{s} - \mathbf{F}_\nu(\mathbf{u}^{(\nu+1)})$.

We assume that all the iterates $\mathbf{u}^{(\nu)}$, $\nu = 0, 1, \dots$, belong to Ω and satisfy Property S in Ω (see Definition 1).

Since the bound of the backward difference quotients depends on the inhomogeneous term (see Proposition 2), we have that there exist two constants β and $\tilde{\beta}$ such that

$$|\nabla_x u_{ij}^{(\nu)}| \leq \beta + \varepsilon_\nu \tilde{\beta} \quad \text{and} \quad |\nabla_y u_{ij}^{(\nu)}| \leq \beta + \varepsilon_\nu \tilde{\beta}, \quad (29)$$

instead of (21), $i, j = 1, \dots, N + 1$.

A result concerning with the convergence of the lagged diffusivity procedure is the following theorem.

Theorem 1. Let be given the nonlinear system (1)

$$\mathbf{F}(\mathbf{u}) \equiv A(\mathbf{u})\mathbf{u} + \mathbf{G}(\mathbf{u}) - \mathbf{s} = \mathbf{0},$$

arising from the discretization of the problem (2)–(3) subject to the Smoothness Conditions (i)–(ii)–(iii)–(iv), and characterized by the Properties (I) and (II), which define the closed ball Ω and the bound β (see formula (21)).

When condition (26) holds, there exists a unique solution $\mathbf{u}^* \in \Omega$ of $\mathbf{F}(\mathbf{u}) = \mathbf{0}$. Suppose that $\{\varepsilon_\nu\}$ is a sequence of positive numbers such that $\varepsilon_\nu \rightarrow 0$ for $\nu \rightarrow \infty$.

Let $\mathbf{u}^{(0)} \in \Omega$ be arbitrary and let $\mathbf{u}^{(\nu+1)}$ be the solution of $\mathbf{F}_\nu(\mathbf{u}) = \mathbf{0}$ satisfying the condition (28) with $\mathbf{F}_\nu(\mathbf{u})$ as in (27).

If all the vectors $\{\mathbf{u}^{(\nu)}\}$ belong to Ω and satisfy the Property S in Ω with the bound (29) instead of (21), then the sequence $\{\mathbf{u}^{(\nu)}\}$ converges to the unique solution \mathbf{u}^* of $\mathbf{F}(\mathbf{u}) = \mathbf{0}$.

Proof. The unique solution \mathbf{u}^* in Ω of (1) satisfies the equation

$$A(\mathbf{u}^*)\mathbf{u}^* + \mathbf{G}(\mathbf{u}^*) - \mathbf{s} = \mathbf{0}, \quad (30)$$

and the iterate $\mathbf{u}^{(\nu+1)}$ satisfies the equation

$$A(\mathbf{u}^{(\nu)})\mathbf{u}^{(\nu+1)} + \mathbf{G}(\mathbf{u}^{(\nu+1)}) = \mathbf{F}_\nu(\mathbf{u}^{(\nu+1)}) + \mathbf{s}, \quad (31)$$

where the Euclidean norm of $\mathbf{F}_\nu(\mathbf{u}^{(\nu+1)})$ satisfies the inequality (28).

Subtracting (31) from (30) and taking into account of the identity (12), we can write

$$A(\mathbf{u}^*)\mathbf{u}^* + \mathbf{G}(\mathbf{u}^*) - A(\mathbf{u}^{(\nu)})\mathbf{u}^{(\nu+1)} - \mathbf{G}(\mathbf{u}^{(\nu+1)}) = -\mathbf{F}_\nu(\mathbf{u}^{(\nu+1)}),$$

as

$$A(\mathbf{u}^*)(\mathbf{u}^* - \mathbf{u}^{(\nu+1)}) + (A(\mathbf{u}^*) - A(\mathbf{u}^{(\nu)}))\mathbf{u}^{(\nu+1)} + \mathbf{G}(\mathbf{u}^*) - \mathbf{G}(\mathbf{u}^{(\nu+1)}) = -\mathbf{F}_\nu(\mathbf{u}^{(\nu+1)}).$$

Thus, we have

$$\begin{aligned} &< A(\mathbf{u}^*)(\mathbf{u}^* - \mathbf{u}^{(\nu+1)}), \mathbf{u}^* - \mathbf{u}^{(\nu+1)} > + < (A(\mathbf{u}^*) - A(\mathbf{u}^{(\nu)}))\mathbf{u}^{(\nu+1)}, \mathbf{u}^* - \mathbf{u}^{(\nu+1)} > + \\ &+ < \mathbf{G}(\mathbf{u}^*) - \mathbf{G}(\mathbf{u}^{(\nu+1)}), \mathbf{u}^* - \mathbf{u}^{(\nu+1)} > = < -\mathbf{F}_\nu(\mathbf{u}^{(\nu+1)}), \mathbf{u}^* - \mathbf{u}^{(\nu+1)} >. \end{aligned}$$

The estimates (15), (17) and (18) in the previous section can be used to yield

$$\begin{aligned} &< -\mathbf{F}_\nu(\mathbf{u}^{(\nu+1)}), \mathbf{u}^* - \mathbf{u}^{(\nu+1)} > \geq h^2 \sum_{j=1}^N \sum_{i=1}^N \sigma_{ij}(u_{ij}^*) |\nabla_h(u_{ij}^* - u_{ij}^{(\nu+1)})|^2 + \\ &+ \sum_{j=1}^N \sigma_{N+1j}(u_{N+1j}^*) |u_{Nj}^* - u_{Nj}^{(\nu+1)}|^2 + \end{aligned}$$

$$\begin{aligned}
& + \sum_{i=1}^N \sigma_{iN+1}(u_{iN+1}^*) |u_{iN}^* - u_{iN}^{(\nu+1)}|^2 - \\
& - h^2 \sum_{j=1}^N \sum_{i=1}^N |\sigma_{ij}(u_{ij}^*) - \sigma_{ij}(u_{ij}^{(\nu)})| |\nabla_h(u_{ij}^{(\nu+1)})| |\nabla_h(u_{ij}^* - u_{ij}^{(\nu+1)})| - \\
& - \sum_{j=1}^N |\sigma_{N+1j}(u_{N+1j}^*) - \sigma_{N+1j}(u_{N+1j}^{(\nu)})| |u_{Nj}^{(\nu+1)}| |u_{Nj}^* - u_{Nj}^{(\nu+1)}| - \\
& - \sum_{i=1}^N |\sigma_{iN+1}(u_{iN+1}^*) - \sigma_{iN+1}(u_{iN+1}^{(\nu)})| |u_{iN}^{(\nu+1)}| |u_{iN}^* - u_{iN}^{(\nu+1)}| + \\
& + c \|\mathbf{u}^* - \mathbf{u}^{(\nu+1)}\|_h^2.
\end{aligned}$$

Inequalities (19) and (20) and Property (II) and, hence, inequality (29), imply (see formula (22))

$$\begin{aligned}
\langle -\mathbf{F}_\nu(\mathbf{u}^{(\nu+1)}), \mathbf{u}^* - \mathbf{u}^{(\nu+1)} \rangle & \geq h^2 \sum_{j=1}^N \sum_{i=1}^N \sigma_{ij}(u_{ij}^*) \left(|\nabla_x(u_{ij}^* - u_{ij}^{(\nu+1)})|^2 + \right. \\
& \left. + |\nabla_y(u_{ij}^* - u_{ij}^{(\nu+1)})|^2 \right) + \sum_{j=1}^N \sigma_{N+1j}(u_{N+1j}^*) |u_{Nj}^* - u_{Nj}^{(\nu+1)}|^2 + \\
& + \sum_{i=1}^N \sigma_{iN+1}(u_{iN+1}^*) |u_{iN}^* - u_{iN}^{(\nu+1)}|^2 - \\
& - \frac{\Lambda(\beta + \varepsilon_{\nu+1}\tilde{\beta})}{2} h^2 \sum_{j=1}^N \sum_{i=1}^N \left(\frac{|u_{ij}^* - u_{ij}^{(\nu)}|^2}{\alpha} + \alpha |\nabla_x(u_{ij}^* - u_{ij}^{(\nu+1)})|^2 \right) - \\
& - \frac{\Lambda(\beta + \varepsilon_{\nu+1}\tilde{\beta})}{2} h^2 \sum_{j=1}^N \sum_{i=1}^N \left(\frac{|u_{ij}^* - u_{ij}^{(\nu)}|^2}{\alpha} + \alpha |\nabla_y(u_{ij}^* - u_{ij}^{(\nu+1)})|^2 \right) + \\
& + c \|\mathbf{u}^* - \mathbf{u}^{(\nu+1)}\|_h^2. \tag{32}
\end{aligned}$$

where α is as yet an undetermined positive number.

Taking into account of assumption (ii), (32) becomes

$$\begin{aligned}
\varepsilon_{\nu+1} \|\mathbf{u}^* - \mathbf{u}^{(\nu+1)}\|_h & \geq \langle -\mathbf{F}_\nu(\mathbf{u}^{(\nu+1)}), \mathbf{u}^* - \mathbf{u}^{(\nu+1)} \rangle \\
& \geq h^2 \sigma_{\min} \sum_{j=1}^N \sum_{i=1}^N (|\nabla_x(u_{ij}^* - u_{ij}^{(\nu+1)})|^2 + |\nabla_y(u_{ij}^* - u_{ij}^{(\nu+1)})|^2) + \\
& + \sigma_{\min} \sum_{j=1}^N |u_{Nj}^* - u_{Nj}^{(\nu+1)}|^2 + \\
& + \sigma_{\min} \sum_{i=1}^N |u_{iN}^* - u_{iN}^{(\nu+1)}|^2 -
\end{aligned}$$

$$\begin{aligned}
& -\frac{\Lambda(\beta + \varepsilon_{\nu+1}\tilde{\beta})}{\alpha}h^2 \sum_{j=1}^N \sum_{i=1}^N |u_{ij}^* - u_{ij}^{(\nu)}|^2 - \\
& -\frac{\Lambda(\beta + \varepsilon_{\nu+1}\tilde{\beta})\alpha}{2}h^2 \sum_{j=1}^N \sum_{i=1}^N \left(|\nabla_x(u_{ij}^* - u_{ij}^{(\nu+1)})|^2 + \right. \\
& \left. + |\nabla_y(u_{ij}^* - u_{ij}^{(\nu+1)})|^2 \right) + c\|\mathbf{u}^* - \mathbf{u}^{(\nu+1)}\|_h^2.
\end{aligned}$$

Hence

$$\begin{aligned}
\varepsilon_{\nu+1}\|\mathbf{u}^* - \mathbf{u}^{(\nu+1)}\|_h & \geq c\|\mathbf{u}^* - \mathbf{u}^{(\nu+1)}\|_h^2 - \\
& -\frac{\Lambda(\beta + \varepsilon_{\nu+1}\tilde{\beta})}{\alpha}h^2 \sum_{j=1}^N \sum_{i=1}^N |u_{ij}^* - u_{ij}^{(\nu)}|^2 + \\
& +(\sigma_{\min} - \frac{\Lambda(\beta + \varepsilon_{\nu+1}\tilde{\beta})\alpha}{2})h^2 \sum_{j=1}^N \sum_{i=1}^N \left(|\nabla_x(u_{ij}^* - u_{ij}^{(\nu+1)})|^2 + \right. \\
& \left. + |\nabla_y(u_{ij}^* - u_{ij}^{(\nu+1)})|^2 \right).
\end{aligned}$$

Choosing

$$\alpha = \frac{2\sigma_{\min}}{\Lambda(\beta + \varepsilon_{\nu+1}\tilde{\beta})},$$

we obtain

$$\|\mathbf{u}^* - \mathbf{u}^{(\nu+1)}\|_h^2 - \frac{\Lambda^2(\beta + \varepsilon_{\nu+1}\tilde{\beta})^2}{2\sigma_{\min}c}\|\mathbf{u}^* - \mathbf{u}^{(\nu)}\|_h^2 \leq \frac{\varepsilon_{\nu+1}}{c}\|\mathbf{u}^* - \mathbf{u}^{(\nu+1)}\|_h.$$

Suppose that $\{\varepsilon_\nu\}$ is a sequence of positive numbers such that $\varepsilon_\nu \rightarrow 0$ for $\nu \rightarrow \infty$ and that condition (26) holds.

Then (see [12]), we may assume that there exists an integer ν_0 such that

$$\frac{\Lambda^2(\beta + \varepsilon_{\nu+1}\tilde{\beta})^2}{2\sigma_{\min}c} \leq \gamma < 1, \quad \text{for all } \nu > \nu_0.$$

Moreover, since the grid function $\{u_{ij}^{(\nu+1)}\}$ belongs to Ω , we may assume that

$$\frac{1}{c}\|\mathbf{u}^* - \mathbf{u}^{(\nu+1)}\|_h \leq \frac{1}{c}2\rho \equiv a.$$

Thus, we obtain the inequality

$$\|\mathbf{u}^* - \mathbf{u}^{(\nu+1)}\|_h^2 \leq \gamma\|\mathbf{u}^* - \mathbf{u}^{(\nu)}\|_h^2 + a\varepsilon_{\nu+1}, \quad \nu > \nu_0,$$

and iteratively

$$\|\mathbf{u}^* - \mathbf{u}^{(\nu_0+\mu)}\|_h^2 \leq \gamma^\mu\|\mathbf{u}^* - \mathbf{u}^{(\nu_0)}\|_h^2 + a\sum_{j=1}^{\mu} \gamma^{\mu-j}\varepsilon_{\nu_0+j}, \quad \mu = 1, 2, \dots$$

Since $\varepsilon_\nu \rightarrow 0$ it follows from the general Toeplitz lemma ([16, p. 399]) that

$$\lim_{\nu \rightarrow \infty} \|\mathbf{u}^* - \mathbf{u}^{(\nu)}\|_h^2 = 0.$$

Hence, the sequence $\{\mathbf{u}^{(\nu)}\}$ of approximate solutions converges to the solution \mathbf{u}^* of the system (1). $\#$

Let us now turn to the numerical solution of the weakly nonlinear system (27) with a Newton iterative method.

For the solution of systems of nonlinear equations of the form (1) with a Newton iterative method a common set of **Standard Assumptions** on $\mathbf{F}(\mathbf{u})$ is the following.

- Equation (1) has a solution \mathbf{u}^* in an open domain \mathcal{K} of \mathbb{R}^n .

For any vector \mathbf{u} and $\mathbf{u}^{(\nu)}$ belonging to \mathcal{K} :

- $A(\mathbf{u}^{(\nu)})$ is a block tridiagonal matrix of order n for any iterate $\mathbf{u}^{(\nu)}$.
The diagonal blocks are square (although not necessarily all of the same order) tridiagonal submatrices, and the off-diagonal blocks are diagonal submatrices.
- The matrix $A(\mathbf{u}^{(\nu)})$ is irreducibly diagonally dominant ([21, pp. 18, 23]) and has positive diagonal entries and nonpositive off-diagonal entries for all the mesh spacings sufficiently small and for all the iterates $\mathbf{u}^{(\nu)} \in \mathcal{K}$.
- $\mathbf{G}(\mathbf{u})$ is a continuously differentiable diagonal mapping on \mathcal{K} , i.e., the k -th component G_k of \mathbf{G} is a function of only the k -th variable u_k , $k = 1, \dots, n$; furthermore $G'(\mathbf{u}) \geq 0$ for all $\mathbf{u} \in \mathcal{K}$.

Thus, $A(\mathbf{u}^{(\nu)})$ is an irreducible nonsingular M-matrix ([15, p. 108]) and $F'_\nu(\mathbf{u}) = A(\mathbf{u}^{(\nu)}) + G'(\mathbf{u})$ is also an irreducible M-matrix with $F'_\nu(\mathbf{u})^{-1} \leq A(\mathbf{u}^{(\nu)})^{-1}$ for all \mathbf{u} and for all the iterates $\mathbf{u}^{(\nu)}$ (see, e.g., [15, p. 109]).

A less restrictive assumption on $\mathbf{G}(\mathbf{u})$ may be considered (see, e.g., [6]).

However, with the assumption of $-\mathbf{G}(\mathbf{u})$ being diagonal, it is possible to express the rate of change due to a reaction of many reaction diffusion processes in realistic applications. (See e.g., [1], [2], [10], [11], [13], [14], [17], [18]).

As a specific example of a solver of (27), we consider the Newton–Arithmetic Mean (AM) method and its simplified version developed in [6], [7], [8] and [9]. The Newton–AM method belongs to the general class of Newton–Iterative methods, as, for example, the well known Newton–SOR method (see [16, §§7.4, 10.3]). The Newton–AM method incorporates at each stage of Newton’s method, the Arithmetic Mean method as inner iterative solver for the linear Newton equation ([8]). When we apply this method to the weakly nonlinear system (27), we can reduce it to a two-stage iterative method as that described in [6] and [9].

The above Standard Assumptions on $\mathbf{F}_\nu(\mathbf{u})$ permit to analyse the convergence of the Newton–AM method and of the simplified Newton–AM method ([8], [6], [9]).

In [8], [9], [6] and [7] we have discussed also the questions concerning with the *global convergence* of the Newton–AM method, using the line search strategy, and the *monotone convergence* of the simplified Newton–AM method.

These methods are well suited for implementation on parallel computers and are significantly efficient when the nonlinear differential operator in equation (2) is *general*, i.e., it is composed of a diffusion term and an advection term (see [8]). The advection term may be present in a reaction diffusion equation if the reactions are taking place in a flowing fluid.

5 Numerical experiments

In this section we consider numerical experiments of the lagged diffusivity fixed point iteration for the solution on a square domain $[0, 1] \times [0, 1]$ of the problem (2)–(3).

Different functions for the nonlinearity factors $\sigma(\mathbf{x}, u)$ and $g(\mathbf{x}, u)$ have been considered.

We are concerned with some simple steady state problems (2)–(3) in which reaction kinetics and diffusion are coupled.

This coupling gives rise to reaction diffusion equations which, in a one dimensional scalar case, can look like

$$\frac{\partial u}{\partial t} - \sigma(u) \frac{\partial^2 u}{\partial x^2} + g(u) = 0,$$

where $u \equiv u(x, t)$ is the concentration, $-g(u)$ represents the kinetics and $\sigma(u)$ is the diffusion coefficient, here taken to be concentration dependent.

(When $g(u) = -ru(1 - u/k)$, this equation is known as the Fisher–Kolmogoroff equation [14, Vol. I, p. 400]).

More specifically, we will consider reaction diffusion problems concerning the heat transfer caused by a chemical reaction (see, e.g., [17]) and the growth of spatially distributed communities (see, e.g., [2]).

The aim of this numerical study is to make a computational verification of the effectiveness of the lagged diffusivity procedure for solving nonlinear difference systems of the type (1) in a extended domain of convergence.

The source function $s(\mathbf{x})$ is chosen in order to satisfy a prespecified exact solution $\mathbf{u}^* = \{u^*(x_i, y_j)\}$ of the nonlinear system (1), $i, j = 1, \dots, N$. The solution $u^*(x, y)$ is chosen equal to $\sin(\pi x) \sin(\pi y)$.

In the following we list the functions σ and g and how they are referred.

$$\begin{aligned} \sigma 1 & : \quad \sigma(u) = 0.5 + 0.5u; \\ \sigma 2 & : \quad \sigma(u) = 0.02 + 0.5u^2; \\ \sigma 3 & : \quad \sigma(u) = 1/(0.02 + 0.5u); \end{aligned}$$

$$g1 : \quad g(u) = 100e^{0.5u};$$

$$\begin{aligned}
g2 & : g(u) = \frac{10^3 u}{(1 + 10u)}; \\
g3 & : g(u) = 5u \log(1 + u).
\end{aligned}$$

We observe that

for $g1$: $g > 0$, $g' > 0$ and $g'' > 0$ for any value of u ;

for $g2$: $g \geq 0$ and $g' > 0$ when $u \geq 0$;

for $g3$: $g \geq 0$, $g' \geq 0$ and $g'' > 0$ when $u \geq 0$;

and then, the functions $g1$, $g2$ and $g3$ satisfy the Standard Assumptions on g for $u \geq 0$.

The lagged diffusivity fixed point iteration has been implemented in a Fortran code with machine precision 2.2×10^{-16} .

This method stops when

$$\varepsilon_{\nu+1} \leq \underline{\varepsilon}, \quad (33)$$

with $\underline{\varepsilon} = 10^{-4}$. An experiment has been carried out with different values of the threshold $\underline{\varepsilon}$.

At each iteration of the lagged diffusivity fixed point method, the approximate solution computed by the simplified Newton-AM method satisfies the stopping rule

$$\|\mathbf{F}_\nu(\mathbf{u}^{(\nu+1)})\| \leq \varepsilon_{\nu+1},$$

with $\varepsilon_1 = 0.1 \|\mathbf{F}(\mathbf{u}^{(0)})\|$ and

$$\varepsilon_{\nu+1} = 0.5 \varepsilon_\nu, \quad \nu = 1, 2, \dots \quad (34)$$

The starting vector of the lagged diffusivity fixed point iteration $\mathbf{u}^{(0)}$ is the null vector ($\mathbf{u}^{(0)} = \mathbf{0}$) or the vector whose all components are equal to 1 ($\mathbf{u}^{(0)} = \mathbf{e}$). In the tables, ν^* indicates the iteration of the lagged diffusivity fixed point method for which condition (33) is satisfied. The number *ktot* is the sum of the simplified Newton method's iterations and it is included in brackets; j_k indicates the number of iterations of the Arithmetic Mean method for the solution of the linear system that occurs at each iteration of the Newton method. This number of iterations of the Arithmetic Mean method is chosen to be fixed at each simplified Newton iteration. For the details of the Arithmetic Mean method used here, see [19].

Here *err* denotes the computed relative error in $l_2(R_h)$ norm,

$$err = \|\mathbf{u}^{(\nu^*)} - \mathbf{u}^*\|_h,$$

and *diff* indicates the last difference of iterations in $l_2(R_h)$ norm,

$$diff = \|\mathbf{u}^{(\nu^*)} - \mathbf{u}^{(\nu^*-1)}\|_h.$$

While, with *res* and *res0* we indicate the residual and the initial residual in the Euclidean norm:

$$res = \|\mathbf{F}(\mathbf{u}^{(\nu^*)})\|, \quad res0 = \|\mathbf{F}(\mathbf{u}^{(0)})\|.$$

The term 4.81(-9) indicates $4.81 \cdot 10^{-9}$.

From the numerical experiments we can draw the conclusions below.

- We observe that, since $\varepsilon_{\nu+1}$ decreases, for ν increasing, as (34) and the lagged diffusivity iteration method stops at the iteration ν^* when the criterium for ε_{ν^*+1} in (33) is satisfied, we have

$$\varepsilon_{\nu^*+1} = \frac{1}{2}\varepsilon_{\nu^*} = \frac{1}{2^2}\varepsilon_{\nu^*-1} = \dots = \frac{1}{2^{\nu^*}}\varepsilon_1 \leq \underline{\varepsilon},$$

where we set $\varepsilon_1 = 0.1\|\mathbf{F}(\mathbf{u}^{(0)})\|$. Then

$$\nu^* > \log_2 \left(\frac{\varepsilon_1}{\underline{\varepsilon}} \right).$$

In the experiments we obtain

$$\nu^* = \lceil \log_2 \left(\frac{\varepsilon_1}{\underline{\varepsilon}} \right) \rceil.$$

- We also observe that, generally, in the experiments the outer residual $\|\mathbf{F}(\mathbf{u}^{(\nu^*)})\|$ has the same order of $\underline{\varepsilon}$ and the error in the discrete $l_2(R_h)$ norm has order of $h\underline{\varepsilon}$.
- When in equation (2) $\sigma(u) = \sigma 2$ and $g(u) = g 3$, the diffusion term and the reaction term are small (Degenerate Equation). If the initial iterate $\mathbf{u}^{(0)}$ is zero (Table 1), we have a nonmonotone decreasing of $\{\|\mathbf{F}(\mathbf{u}^{(\nu)})\|\}$ that produces a large number of iterations of the simplified Newton method; e.g., we have $\|\mathbf{F}(\mathbf{u}^{(0)})\| = 933.26$ and $\|\mathbf{F}(\mathbf{u}^{(1)})\| = 8514.10$. The suggestion is to change the initial vector; the same case with $\mathbf{u}^{(0)} = \mathbf{e}$ produces better results (see Table 2) and a monotone decreasing of $\{\|\mathbf{F}(\mathbf{u}^{(\nu)})\|\}$.
- When in equation (2) $\sigma(u) = \sigma 1$ or $\sigma(u) = \sigma 3$ and $g(u) = g 3$, the diffusion term dominates the reaction term (Equation of Diffusion-type).

The matrix $A(\mathbf{u})$ with $\sigma(u) = \sigma 3$ ($g(u) = g 3$) is more ill conditioned than $A(\mathbf{u})$ with $\sigma(u) = \sigma 1$

When in equation (2) $\sigma(u) = \sigma 1$ or $\sigma(u) = \sigma 2$ and $g(u) = g 2$, the slope of $\sigma(u)$ is quasi-flat and the slope of $g(u)$ is steep.

When in equation (2) $\sigma(u) = \sigma 1$ or $\sigma(u) = \sigma 2$ and $g(u) = g 1$, the slope of $\sigma(u)$ is quasi-flat and the slope of $g(u)$ is quasi-flat with $\|G'(\mathbf{u})\|$ large.

When in equation (2) $\sigma(u) = \sigma 3$ and $g(u) = g 1$ or $g(u) = g 2$, the diffusion term is comparable to the reaction term.

The numerical results on these test problems (Tables 1 and 2) confirm the statements 1. and 2. at the beginning of the previous section.

- In the experiment in Table 4 the function σ is chosen dependent on u and on x and y . The range of values of this function σ is larger than that of σ_1 and a comparison with Tables 1 and 2 confirms that we could expect a larger number of the simplified Newton iterations.
- The results in Tables 3 and 5 show, in some cases, the behaviour of the Lagged Diffusivity procedure for different values of the threshold $\underline{\varepsilon}$ and of the number N of grid points per dimension. From Table 3, we see that the choice $\underline{\varepsilon} = 10^{-4}$ is appropriate.

Acknowledgements

The author is very grateful to the anonymous referees for their valuable comments and suggestions, which have improved the paper.

$N = 256; \underline{\varepsilon} = 10^{-4}; j_k = 20; \mathbf{u}^{(0)} = \mathbf{0};$					
$\sigma(u) = \sigma_1$					
$g(u)$	$\nu^*(ktot)$	err	res	$res0$	$diff$
g_1	24 (528)	4.81(-9)	1.07(-4)	9545.60	4.93(-9)
g_2	25 (1283)	1.13(-8)	1.06(-4)	20077.85	1.14(-8)
g_3	22 (2135)	1.65(-8)	9.18(-5)	2234.89	1.66(-8)
$\sigma(u) = \sigma_2$					
$g(u)$	$\nu^*(ktot)$	err	res	$res0$	$diff$
g_1	23 (208)	7.62(-9)	1.72(-4)	8102.34	7.28(-9)
g_2	25 (661)	1.38(-8)	8.64(-5)	18815.92	1.35(-8)
g_3	20 (6606)	1.48(-6)	1.25(-2)	933.26	3.36(-6)
$\sigma(u) = \sigma_3$					
$g(u)$	$\nu^*(ktot)$	err	res	$res0$	$diff$
g_1	28 (1793)	8.32(-9)	3.05(-4)	233897.67	8.27(-9)
g_2	28 (3077)	2.28(-8)	4.70(-4)	237686.05	2.27(-8)
g_3	28 (3987)	3.41(-8)	6.66(-4)	233112.89	3.41(-8)

Table 1: Results for different functions $g(u)$ and $\sigma(u)$ with $\mathbf{u}^{(0)} = \mathbf{0}$.

References

- [1] Aris R.: The Mathematical Theory of Diffusion in Permeable Catalysts, vol. I, II, Clarendon Press, Oxford, 1975.
- [2] Buffoni G., Griffa A., Li Z., de Mottoni P.: *Spatially distributed communities: the resource-consumer system*, Journal of Mathematical Biology, 33 (1995), 723–743.
- [3] Courant R., Friedrichs K.O., Lewy H.: *Über die partiellen Differenzgleichungen der mathematischen Physik*, Mathematische Annalen, 100 (1928), 32–74.

$N = 256; \varepsilon = 10^{-4}; j_k = 20; \mathbf{u}^{(0)} = \mathbf{e};$					
$\sigma(u) = \sigma 1$					
$g(u)$	$\nu^*(ktot)$	err	res	$res0$	$diff$
$g1$	31 (553)	6.64(-9)	1.48(-4)	1678846.53	6.38(-9)
$g2$	31 (1291)	1.48(-8)	1.38(-4)	1679411.24	1.48(-8)
$g3$	31 (1987)	2.43(-8)	1.34(-4)	1676964.70	2.44(-8)
$\sigma(u) = \sigma 2$					
$g(u)$	$\nu^*(ktot)$	err	res	$res0$	$diff$
$g1$	30 (231)	5.52(-9)	1.24(-4)	780956.25	6.09(-9)
$g2$	30 (684)	1.82(-8)	1.13(-4)	781361.58	1.77(-8)
$g3$	30 (928)	9.14(-8)	1.48(-4)	779441.66	5.32(-8)
$\sigma(u) = \sigma 3$					
$g(u)$	$\nu^*(ktot)$	err	res	$res0$	$diff$
$g1$	37 (1487)	5.17(-9)	1.91(-4)	74825897.36	5.24(-9)
$g2$	37 (2878)	1.39(-8)	2.89(-4)	74826351.39	1.40(-8)
$g3$	37 (3600)	2.14(-8)	4.18(-4)	74824482.23	2.14(-8)

Table 2: Results for different functions $g(u)$ and $\sigma(u)$ with $\mathbf{u}^{(0)} = \mathbf{e}$.

- [4] Dembo R.S., Eisenstat S.C., Steihaug T.: *Inexact Newton methods*, SIAM Journal on Numerical Analysis, 19 (1982), 400–408.
- [5] Eisenstat S.C., Walker H.F.: *Globally convergent inexact Newton methods*, SIAM Journal on Optimization, 4 (1994), 393–422.
- [6] Galligani E.: *A two-stage iterative method for solving a weakly nonlinear system*, Atti del Seminario Matematico e Fisico dell’Università di Modena, L (2002), 195–215.
- [7] Galligani E.: *A two-stage iterative method for solving a weakly nonlinear parametrized system*, International Journal of Computer Mathematics, 79 (2002), 1211–1224.
- [8] Galligani E.: *The Newton-arithmetic mean method for the solution of systems of nonlinear equations*, Applied Mathematics and Computation, 134 (2003), 9–34.
- [9] Galligani E.: *On solving a special class of weakly nonlinear finite-difference systems*, International Journal of Computer Mathematics, 86 (2009), 503–522.
- [10] Keller H.B., Cohen D.S.: *Some positive problems suggested by nonlinear heat generation*, Journal of Mathematics and Mechanics, 16 (1967), 1361–1376.
- [11] Kernevez J.P.: *Enzyme Mathematics*, North–Holland, Amsterdam, 1980.

$N = 256; \sigma(u) = \sigma 2; j_k = 20; \mathbf{u}^{(0)} = \mathbf{0};$				
$g(u) = g1; res0=8102.34;$				
$\underline{\varepsilon}$	$\nu^*(ktot)$	err	res	$diff$
10^{-2}	17 (152)	4.91(-7)	1.11(-2)	4.68(-7)
10^{-3}	20 (180)	6.12(-8)	1.38(-3)	5.83(-8)
10^{-4}	23 (208)	7.62(-9)	1.72(-4)	7.28(-9)
10^{-5}	27 (245)	4.86(-10)	1.10(-5)	4.63(-10)
$g(u) = g2; res0=18815.92;$				
$\underline{\varepsilon}$	$\nu^*(ktot)$	err	res	$diff$
10^{-2}	18 (462)	1.75(-6)	1.09(-2)	1.87(-6)
10^{-3}	21 (547)	2.21(-7)	1.37(-3)	2.17(-7)
10^{-4}	25 (661)	1.38(-8)	8.64(-5)	1.35(-8)
10^{-5}	28 (747)	1.72(-9)	1.07(-5)	1.75(-9)
$g(u) = g3; res0=933.26;$				
$\underline{\varepsilon}$	$\nu^*(ktot)$	err	res	$diff$
10^{-2}	14 (4405)	6.54(-5)	0.53	1.49(-4)
10^{-3}	17 (5540)	1.10(-5)	8.20(-2)	2.14(-5)
10^{-4}	20 (6606)	1.48(-6)	1.25(-2)	3.36(-6)
10^{-5}	24 (8169)	1.16(-7)	9.90(-4)	2.60(-7)

Table 3: Results for different values of $\underline{\varepsilon}$.

- [12] Meyer G.H.: *The numerical solution of quasilinear elliptic equations*, in: Numerical Solution of Systems of Nonlinear Algebraic Equations (G. Byrne, C.A. Hall eds.), Academic Press, New York, 1973, 27–61.
- [13] Moré J.J.: *A collection of nonlinear model problems*, in: Computational Solution of Nonlinear Systems of Equations (E.L. Allgower, K. Georg eds.), Lectures in Applied Mathematics, vol. 26, American Mathematical Society, Providence RI, 1990, 723–762.
- [14] Murray J.D.: *Mathematical Biology*, vol. I, II, Springer–Verlag, Berlin, 2003.
- [15] Ortega J.M.: *Numerical Analysis: A Second Course*, Academic Press, New York, 1972. (SIAM, Philadelphia, 1990).
- [16] Ortega J.M., Rheinboldt W.C.: *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970. (SIAM, Philadelphia, 2000).
- [17] Pao C.V.: *Monotone iterative methods for finite difference system of reaction diffusion equations*, Numerische Mathematik, 46 (1985), 571–586.
- [18] Pohozaev S.T.: *The Dirichlet problem for the equation $\Delta u = u^2$* , Soviet Mathematics, 1 (1960), 1143–1146.

$$N = 256; \sigma(u, x, y) = 0.5 + 0.5u + (x + y)^2; \underline{\varepsilon} = 10^{-4}; j_k = 20;$$

$\mathbf{u}^{(0)} = \mathbf{0}$					
$g(u)$	$\nu^*(ktot)$	err	res	$res0$	$diff$
$g1$	25 (1052)	4.63(-9)	1.49(-4)	25787.21	4.59(-9)
$g2$	25 (1788)	1.14(-8)	1.80(-4)	32151.45	1.15(-8)
$g3$	25 (2457)	1.16(-8)	1.27(-4)	23142.53	1.16(-8)
$\mathbf{u}^{(0)} = \mathbf{e}$					
$g(u)$	$\nu^*(ktot)$	err	res	$res0$	$diff$
$g1$	32 (1058)	3.89(-9)	1.25(-4)	2777386.07	3.86(-9)
$g2$	32 (1844)	7.70(-9)	1.21(-4)	2777975.09	7.63(-9)
$g3$	32 (2428)	1.08(-8)	1.19(-4)	2775500.98	1.08(-8)

Table 4: Results for different functions $g(u)$ and σ dependent on x, y and u .

- [19] Ruggiero V., Galligani E.: *A parallel algorithm for solving block tridiagonal linear systems*, Computers Mathematics with Applications, 24 (1992), 15–21.
- [20] Thomas J.: *Numerical Partial Differential Equations: Finite Difference Methods*, Springer, New York, 1995.
- [21] Varga R.S.: *Matrix Iterative Analysis*, Second Edition, Springer, Berlin, 2000.
- [22] Vogel C.R., Oman M.E.: *Iterative methods for total variation denoising*, SIAM Journal on Scientific Computing, 17 (1996), 227–238.

$\underline{\varepsilon} = 10^{-4}; j_k = 20; \mathbf{u}^{(0)} = \mathbf{0};$

$\sigma(u) = \sigma 1; g(u) = g 1$					
N	$\nu^*(ktot)$	err	res	$res0$	$diff$
512	25 (2193)	2.41(-9)	1.07(-4)	19054.35	2.45(-9)
256	24 (528)	4.81(-9)	1.07(-4)	9545.60	4.93(-9)
128	23 (128)	9.04(-9)	1.01(-4)	4791.20	1.03(-8)
64	22 (29)	1.52(-8)	8.60(-5)	2413.97	9.80(-9)
$\sigma(u) = \sigma 2; g(u) = g 2$					
n	$\nu^*(ktot)$	err	res	$res0$	$diff$
512	26 (2698)	6.96(-9)	8.64(-5)	37558.59	6.99(-9)
256	25 (661)	1.38(-8)	8.64(-5)	18815.92	1.35(-8)
128	24 (172)	2.57(-8)	8.04(-5)	9444.59	2.98(-8)
64	23 (53)	4.61(-8)	7.28(-5)	4758.82	5.44(-8)
$\sigma(u) = \sigma 3; g(u) = g 2$					
n	$\nu^*(ktot)$	err	res	$res0$	$diff$
512	29 (12641)	1.21(-8)	5.04(-4)	504777.77	1.21(-8)
256	28 (3077)	2.28(-8)	4.70(-4)	237686.05	2.27(-8)
128	27 (759)	4.14(-8)	4.14(-4)	109421.02	4.08(-8)
64	26 (197)	7.28(-8)	3.48(-4)	49538.06	7.33(-8)

Table 5: Results for different values of N .