Galligani *Advances in Difference Equations* 2012, **2012**:30
http://www.advancesindifferenceequations.com/content/2012/1/30

Advances in Difference Equations
a SpringerOpen Journal

RESEARCH **Open Access**

Lagged diffusivity method for the solution of nonlinear diffusion convection problems with finite differences

Emanuele Galligani

Correspondence: emanuele.galligani@unimore.it
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

This article concerns with the analysis of an iterative procedure for the solution of a nonlinear nonstationary diffusion convection equation in a two-dimensional bounded domain supplemented by Dirichlet boundary conditions. This procedure, denoted Lagged Diffusivity method, computes the solution by lagging the diffusion term. A model problem is considered and a finite difference discretization for that model is described. Furthermore, properties of the finite difference operator are proved. Then, a sufficient condition for the convergence of the Lagged Diffusivity method is given. At each stage of the iterative procedure, a linear system has to be solved and the Arithmetic Mean method is used. Numerical experiments show the efficiency, for different test functions, of the Lagged Diffusivity method combined with the Arithmetic Mean method as inner solver. Better results are obtained when the convection term increases.

1 Statement of the problem

Consider as model problem the nonlinear diffusion convection equation

$$\frac{\partial u}{\partial t} = \mathrm{div}(\sigma \nabla u) - \tilde{v} \cdot \nabla u - \alpha u + s, \tag{1}$$

where $u = u(x, y, t)$ is the density function at the point (x, y) at the time t of a diffusion medium R, $\sigma = \sigma(x, y, u) > 0$ is the diffusion coefficient or diffusivity and is dependent on the solution u, $\alpha = \alpha(x, y) \geq 0$ is the absorption term, $\tilde{v} = \tilde{v}(x, y, t, u)$ is the velocity vector and the source term $s(x, y, t)$ is a real valued sufficiently smooth function.

Equation (1) can be supplemented by the initial condition $(t = 0)$

$$u(x, y, 0) = U_0(x, y), \tag{2}$$

in the closure $\bar{R}$ of R and by Dirichlet boundary condition on the contour ∂R of R of the form

$$u(x, y, t) = U_1(x, y, t). \tag{3}$$

In the following, we suppose R to be a rectangular domain with boundary ∂R and we assume that the functions σ, α, and s satisfy the "smoothness" conditions:

(i) the function $\sigma = \sigma(u)$ is continuous in u; the functions $\alpha(x, y)$ and $s(x, y, t)$ are continuous in x, y and in x, y, t respectively;

(ii) there exist two positive constants $\sigma_{\min}$ and $\sigma_{\max}$ such that

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

uniformly in u; in addition, $\alpha(x, y) \geq \alpha_{\min} \geq 0$;

(iii) for fixed $(x, y) \in R$, the function $\sigma(u)$ satisfies Lipschitz condition in u with constant Γ (uniformly in x, y), $\Gamma > 0$.

Here, the vector $\tilde{\boldsymbol{v}} = (\tilde{v}_1, \tilde{v}_2)^T$ is assumed to be constant.

The nonlinearity introduced by the u-dependence of the coefficient $\sigma(u)$ requires that, in general, the solution of equation (1) be approximated by numerical methods.

We superimpose on $R \cup \partial R$ a grid of points $R_h \cup \partial R_h$; the set of the internal points R_h of the grid are the mesh points (x_i, y_j), for $i = 1,\ldots, N$ and $j = 1,\ldots, M$, with uniform mesh size h along x and y directions, respectively, i.e., $x_{i+1} = x_i + h$ and $y_{j+1} = y_j + h$ for $i = 0,\ldots, N$, $j = 0,\ldots, M$.

Thus, at the mesh points of $R \cup \partial R$, (x_i, y_j), for $i = 0,\ldots, N + 1$, $j = 0,\ldots, M + 1$, the solution $u(x_i, y_j, t)$ is approximated by a grid function $u_{ij}(t)$ defined on $R_h \cup \partial R_h$ and satisfying the boundary condition (3) on ∂R_h for $t > 0$ and the initial condition (2) on $R_h \cup \partial R_h$ for $t = 0$.

By ordering in a row lexicographic order the mesh points $P_l = (x_i, y_j)$ (i.e., $l = (j - 1) \cdot N + i$ with $j = 1,\ldots, M$, and $i = 1,\ldots, N$), we can write the vector $\boldsymbol{u}(t)$ of components $u_{ij}(t)$ and approximate the right-hand side of (1) by

$$\boldsymbol{A}(\boldsymbol{u}(t))\boldsymbol{u}(t) + \boldsymbol{b}(\boldsymbol{u}(t)) + \boldsymbol{s}(t), \tag{4}$$

where the matrix $A(\boldsymbol{u}(t))$ is of order $\mu = M \times N$ and has the block tridiagonal form; the M diagonal blocks are tridiagonal matrices of order N and the $M - 1$ sub- and super-diagonal blocks are diagonal matrices of order N.

The five nonzero elements of $A(\boldsymbol{u}(t))$ corresponding to $u_{ij-1}(t)$, $u_{i-1j}(t)$, $u_{ij}(t)$, $u_{i+1j}(t)$ and $u_{ij+1}(t)$ respectively, are

$$\{(B_{ij} + \tilde{B}_{ij}), (L_{ij} + \tilde{L}_{ij}), -(D_{ij} + \hat{D}_{ij}), (R_{ij} + \tilde{R}_{ij}), (T_{ij} + \tilde{T}_{ij})\},$$

where

$$\begin{aligned}
L_{ij} \equiv L_{ij}(\boldsymbol{u}(t)) &= \frac{1}{h^2}\sigma\left(\frac{u_{ij}(t) + u_{i-1j}(t)}{2}\right),\\
B_{ij} \equiv B_{ij}(\boldsymbol{u}(t)) &= \frac{1}{h^2}\sigma\left(\frac{u_{ij}(t) + u_{ij-1}(t)}{2}\right),\\
R_{ij} \equiv R_{ij}(\boldsymbol{u}(t)) &= \frac{1}{h^2}\sigma\left(\frac{u_{i+1j}(t) + u_{ij}(t)}{2}\right),\\
T_{ij} \equiv T_{ij}(\boldsymbol{u}(t)) &= \frac{1}{h^2}\sigma\left(\frac{u_{ij+1}(t) + u_{ij}(t)}{2}\right),\\
\tilde{L}_{ij} &= \frac{\tilde{v}_1}{2h}, \tilde{B}_{ij} = \frac{\tilde{v}_2}{2h},\\
\tilde{R}_{ij} &= -\frac{\tilde{v}_1}{2h}, \tilde{T}_{ij} = -\frac{\tilde{v}_2}{2h},\\
D_{ij} \equiv D_{ij}(\boldsymbol{u}(t)) &= B_{ij} + L_{ij} + R_{ij} + T_{ij}, \hat{D}_{ij} = \alpha(x_i, y_j).
\end{aligned} \tag{5}$$

The matrix $A(\boldsymbol{u}(t))$ is an irreducible matrix [[1], p. 18].

Providing that the mesh spacing h is sufficiently small, i.e.,

$$h < \min\left\{\frac{2\sigma_{\min}}{|\tilde{v}_1|}, \frac{2\sigma_{\min}}{|\tilde{v}_2|}\right\}, \tag{6}$$

the matrix $A(\boldsymbol{u}(t))$ is strictly ($\alpha(x, y) > 0$) or irreducibly ($\alpha(x, y) = 0$) diagonally dominant ([[1], p. 23]) and has negative diagonal elements, $a_{ll}(\boldsymbol{u}(t)) < 0$ ($l = 1, ..., \mu$) and nonnegative off diagonal elements $a_{lp}(\boldsymbol{u}(t)) \geq 0$, $l \neq p$, with $l, p = 1, ..., \mu$; therefore, $-A(\boldsymbol{u}(t))$ is an M-matrix [[1], p. 91].

In the case of diffusion equation ($\tilde{\boldsymbol{v}} = \boldsymbol{0}$), the matrix $A(\boldsymbol{u}(t))$ is also symmetric; then $-A(\boldsymbol{u}(t))$ is a Stieltjes matrix and is symmetric positive definite [[1], p. 91].

The vector $\boldsymbol{b}(\boldsymbol{u}(t))$ in (4) is obtained imposing Dirichlet boundary conditions (3) and it depends on the function $U_1(x_i, y_j, t)$ at points (x_i, y_j) of ∂R and its lth component depends on the lth component of the solution $\boldsymbol{u}(t)$ for the mesh point P_l of R which is neighbor of points of ∂R.

The vector $\boldsymbol{s}(t)$ in (4) has components $s_l(t) = s(x_i, y_j, t)$ for $i = 1,..., N$, $j = 1,..., M$ and $l = (j - 1) N + i$.

We can apply a step-by-step method to the initial value problem of μ equations

$$\frac{d\boldsymbol{u}(t)}{dt} = A(\boldsymbol{u}(t))\boldsymbol{u}(t) + \boldsymbol{b}(\boldsymbol{u}(t)) + \boldsymbol{s}(t),$$

with the initial condition (2), $u(x_i, y_j, 0) = U_0(x_i, y_j)$, and computes the approximation $\boldsymbol{u}^{n+1}$ to $\boldsymbol{u}(t_{n+1})$ using the approximate solution $\boldsymbol{u}^n$ at the time level t_n, with $t_n = n\Delta t$ and Δt the time step.

Indicating $\boldsymbol{s}^n = \boldsymbol{s}(t_n)$, for $n = 0,1,...$, the well-known θ-method (see, e.g., [2]) is written

$$\frac{\boldsymbol{u}^{n+1} - \boldsymbol{u}^n}{\Delta t} = \theta(A(\boldsymbol{u}^{n+1})\boldsymbol{u}^{n+1} + \boldsymbol{b}(\boldsymbol{u}^{n+1}) + \boldsymbol{s}^{n+1}) + (1-\theta)(A(\boldsymbol{u}^n)\boldsymbol{u}^n + \boldsymbol{b}(\boldsymbol{u}^n) + \boldsymbol{s}^n),$$

where θ is a real parameter such that $0 \leq \theta \leq 1$; for any $\theta \neq 0$, the method is implicit. I is the $\mu \times \mu$ identity matrix.

Thus, at each time level $n = 0, 1,...$, the vector $\boldsymbol{u}^{n+1} \in \mathbb{R}^\mu$ is the solution of the nonlinear system,

$$\boldsymbol{F}(\boldsymbol{u}) \equiv (I - \Delta t\theta A(\boldsymbol{u}))\boldsymbol{u} - \Delta t\theta \boldsymbol{b}(\boldsymbol{u}) - \boldsymbol{w} = \boldsymbol{0}. \tag{7}$$

The vector $\boldsymbol{w} \in \mathbb{R}^\mu$ is given by

$$\boldsymbol{w} \equiv \boldsymbol{w}^n = (I + \Delta t(1-\theta)A(\boldsymbol{u}^n))\boldsymbol{u}^n + \Delta t(1-\theta)\boldsymbol{b}(\boldsymbol{u}^n) + \Delta t(\theta \boldsymbol{s}^{n+1} + (1-\theta)\boldsymbol{s}^n).$$

We set $\tau = \Delta t\theta$.

We can introduce an iterative method of Lagged Diffusivity, computing the new iterate $\boldsymbol{u}^{(k+1)}$ keeping the diffusivity term at the previous iteration k. That is, since the matrix I-$\tau A(\boldsymbol{u})$ is nonsingular for all $\boldsymbol{u} \in \mathbb{R}^\mu$ then $\boldsymbol{u}^{(k+1)}$ is the solution of the linear system

$$(I - \tau A(\boldsymbol{u}^{(k)}))\boldsymbol{u} = \boldsymbol{w} + \tau \boldsymbol{b}(\boldsymbol{u}^{(k)}),$$

such that the residual

$$r^{(k+1)} = (I - \tau A(u^{(k)}))u^{(k+1)} - w - \tau b(u^{(k)}),$$

satisfies the stopping condition

$$\left\| r^{(k+1)} \right\| \le \varepsilon_{k+1},$$

where ε_k is a given tolerance such that $\varepsilon_k \to 0$ when $k \to \infty$ and $\|\cdot\|$ indicates the Euclidean norm. The initial iterate $u^{(0)}$ of this Lagged Diffusivity procedure can be set equal to u^n.

2 Uniform monotonicity and a convergence result

In this section, we consider the nonlinear system $F(u) = 0$ in (7) and we prove that $F(u)$ is continuously and uniformly monotone and then $F(u) = 0$ has a unique solution. Moreover, we prove that the sequence $\{u^{(k)}\}$ generated by the Lagged Diffusivity procedure is convergent to the solution. Before this, we have to prove three lemmas on some properties of finite difference operators.

In the following, we may consider the matrix $A(u)$ as

$$A(u) = \hat{A}(u) + \tilde{A} + \hat{D},$$

where $\hat{A}(u)$ and $\tilde{A}$ are the block tridiagonal matrices whose row elements are $\{B_{ij}, L_{ij}, -D_{ij}, R_{ij}, T_{ij}\}$ and $\{\tilde{B}_{ij}, \tilde{L}_{ij}, \tilde{R}_{ij}, \tilde{T}_{ij}\}$, respectively, while the matrix $\hat{D}$ is a diagonal matrix whose diagonal entries are $\{-\hat{D}_{ij}\}$. Furthermore, we denote

$$\hat{A}(u) = \hat{A}^x(u) + \hat{A}^y(u),$$

where $\hat{A}^x(u)$ is the block diagonal matrix whose row elements are $\{L_{ij}, -D^x_{ij}, R_{ij}\}$ with $D^x_{ij} = L_{ij} + R_{ij}$, and $\hat{A}^y(u)$ is the block tridiagonal matrix whose row elements are $\{B_{ij}, -D^y_{ij}, T_{ij}\}$ with $D^y_{ij} = B_{ij} + T_{ij}$. Analogously, we can define $b(u) = b^x(u) + b^y(u)$ where $b^x(u)$ contains the contributions of $U_1(x_0, y_j, t)$ and $U_1(x_{N+1}, y_j, t)$ for $j = 1, ...,M$ and $b^y(u)$ contains the contributions of $U_1(x_i, y_0, t)$ and $U_1(x_i, y_{M+1}, t)$ for $i = 1,...,N$.

For the sake of clarity, we set $u_{ij} \equiv u_{ij}(t)$ and $v_{ij} \equiv v_{ij}(t)$, ($i = 0,..., N + 1$, $j = 0,..., M + 1$), the grid functions defined on $R_h \cup \partial R_h$ and satisfying the Dirichlet boundary condition on ∂R_h for $t > 0$.

For grid functions $\{u_{ij}\}$ and $\{v_{ij}\}$ of this type, the discrete l_2 (R_h) inner product and norm are defined by the formulas

$$< u, v > = h^2 \sum_{i=1}^{N} \sum_{j=1}^{M} u_{ij} v_{ij},$$

$$\|u\|_h = (h^2 \sum_{i=1}^{N} \sum_{j=1}^{M} |u_{ij}|^2)^{1/2} = (< u, u >)^{1/2},$$

respectively

We denote by $\mathcal{B}$ the set of all grid functions defined on $R_h \cup \partial R_h$ and satisfying the Dirichlet boundary condition on ∂R_h for $t > 0$ for which there exist two positive constants ρ and β, both independent of h, such that

$$\|\boldsymbol{u}\|_h \leq \rho \tag{8}$$

$$\left|\nabla_x u_{ij}\right| \leq \beta, \qquad \left|\nabla_y u_{ij}\right| \leq \beta. \tag{9}$$

Here, $\nabla_x u_{ij}$ and $\nabla_y u_{ij}$ indicates the backward difference quotients

$$\nabla_x u_{ij} = \frac{u_{ij} - u_{i-1j}}{h}, \qquad \nabla_y u_{ij} = \frac{u_{ij} - u_{ij-1}}{h}. \tag{10}$$

Before to prove the main result, we summarize in three lemmas on some properties of finite difference operators.

Here, for a grid function $\{u_{ij}\}$, we denote

$$u_{i\pm 1/2j} = \frac{u_{i\pm 1j} + u_{ij}}{2}, \qquad u_{ij\pm 1/2} = \frac{u_{ij\pm 1} + u_{ij}}{2}.$$

Lemma 1. Let $\{u_{ij}\}$, $\{v_{ij}\}$, $\{z_{ij}\}$ be three grid functions defined at the mesh points (x_i, y_j) of a grid $R_h \cup \partial R_h$, $i = 0,\ldots,N + 1$, $j = 0,\ldots, M + 1$ which are equal to the prescribed function U_1 (x_i, y_j, t) at the point (x_i, y_j) of ∂R_h and $t > 0$; then

$$< -\hat{A}(\boldsymbol{z})\boldsymbol{u} - \boldsymbol{b}(\boldsymbol{z}), \boldsymbol{v} >=< -\hat{A}^x(\boldsymbol{z})\boldsymbol{u} - \boldsymbol{b}^x(\boldsymbol{z}), \boldsymbol{v} > + < -\hat{A}^y(\boldsymbol{z})\boldsymbol{u} - \boldsymbol{b}^y(\boldsymbol{z}), \boldsymbol{v} >,$$

where

$$\begin{aligned} < -\hat{A}^x(\boldsymbol{z})\boldsymbol{u} - \boldsymbol{b}^x(\boldsymbol{z}), \boldsymbol{v} > = \sum_{j=1}^{M} \Big[& \sigma(z_{1/2j})(u_{1j} - u_{0j})v_{1j} + \\ & + \sum_{i=2}^{N} \sigma(z_{i-1/2j})(u_{ij} - u_{i-1j})(v_{ij} - v_{i-1j}) + \\ & + \sigma(z_{N+1/2j})(u_{Nj} - u_{N+1j})v_{Nj} \Big], \end{aligned} \tag{11}$$

$$\begin{aligned} < -\hat{A}^y(\boldsymbol{z})\boldsymbol{u} - \boldsymbol{b}^y(\boldsymbol{z}), \boldsymbol{v} > = \sum_{i=1}^{M} \Big[& \sigma(z_{i1/2})(u_{i1} - u_{i0})v_{i1} + \\ & + \sum_{j=2}^{M} \sigma(z_{ij-1/2})(u_{ij} - u_{ij-1})(v_{ij} - v_{ij-1}) + \\ & + \sigma(z_{iM+1/2})(u_{iM} - u_{iM+1})v_{Mj} \Big]. \end{aligned} \tag{12}$$

Proof. Formulae (11) and (12) follow immediately from the definition of the coefficients in (5).

Lemma 2. Let $\{u_{ij}\}$ and $\{v_{ij}\}$ be two grid functions defined at the mesh points (x_i, y_j) of the grid $R_h \cup \partial R_h$, $i = 0,\ldots, N + 1$, $j = 0,\ldots, M + 1$ such that, at the point (x_i, y_j) of ∂R_h and $t > 0$, the grid function $\{u_{ij}\}$ is equal to the prescribed function $U_1(x_i, y_j, t)$ and the grid function $\{v_{ij}\}$ is equal to the null function, respectively.

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

$$< -\hat{A}(\boldsymbol{u})\boldsymbol{v},\boldsymbol{v} > = h^2 \sum_{j=1}^{M} \sum_{i=1}^{N} \sigma(u_{i-1/2j})(\nabla_x v_{ij})^2 + h^2 \sum_{i=1}^{N} \sum_{j=1}^{M} \sigma(u_{ij-1/2})(\nabla_y v_{ij})^2 + \\ + \sum_{j=1}^{M} \sigma(u_{N+1/2j}) v_{Nj} v_{Nj} + \sum_{i=1}^{N} \sigma(u_{iM+1/2}) v_{iM} v_{iM} \tag{13}$$

Proof. From (5) the expression of $< -\hat{A}(\boldsymbol{u})\boldsymbol{v},\boldsymbol{v} >$ is

$$< -\hat{A}(\boldsymbol{u})\boldsymbol{v},\boldsymbol{v} > = \sum_{j=1}^{M} \left[\sigma(u_{1/2j}) v_{1j} v_{1j} + \sum_{i=2}^{N} \sigma(u_{i-1/2j})(v_{ij} - v_{i-1j})(v_{ij} - v_{i-1j}) + \sigma(u_{N+1/2j}) v_{Nj} v_{Nj} \right] + \\ + \sum_{i=1}^{N} \left[\sigma(u_{i1/2}) v_{i1} v_{i1} + \sum_{j=2}^{M} \sigma(u_{ij-1/2})(v_{ij} - v_{ij-1})(v_{ij} - v_{ij-1}) + \sigma(u_{iM+1/2}) v_{iM} v_{iM} \right].$$

Since the grid function $\{v_{ij}\}$ is equal to zero for all the points of ∂R_h, we can write

$$< -\hat{A}(\boldsymbol{u})\boldsymbol{v},\boldsymbol{v} > = \sum_{j=1}^{M} \left[\sigma(u_{1/2j})(v_{1j} - v_{0j})(v_{1j} - v_{0j}) + \sum_{i=2}^{N} \sigma(u_{i-1/2j})(v_{ij} - v_{i-1j})(v_{ij} - v_{i-1j}) + \right. \\ \left. + \sigma(u_{N+1/2j}) v_{Nj} v_{Nj} \right] + \\ + \sum_{i=1}^{N} \left[\sigma(u_{i1/2})(v_{i1} - v_{i0})(v_{i1} - v_{i0}) + \sum_{j=2}^{M} \sigma(u_{ij-1/2})(v_{ij} - v_{ij-1})(v_{ij} - v_{ij-1}) + \right. \\ \left. + \sigma(u_{iM+1/2}) v_{iM} v_{iM} \right].$$

From the definition of backward difference quotients (10), we have formula (13).

Lemma 3. Let $\{u_{ij}\}$, $\{v_{ij}\}$ and $\{\tilde{v}_{ij}\}$ be three grid functions of $\mathcal{B}$ such that, at the points (x_i, y_j) of the boundary ∂R_h they are equal to the prescribed function $U_1(x_i, y_j, t)$ $(t > 0)$.

Then, we have the following expression for the discrete $l_2(R_h)$ inner product

$$\left| < \left(-\hat{A}(\boldsymbol{u}) + \hat{A}(\boldsymbol{v})\right) \tilde{\boldsymbol{v}} - \boldsymbol{b}(\boldsymbol{u}) + \boldsymbol{b}(\boldsymbol{v}), \boldsymbol{u} - \tilde{\boldsymbol{v}} \right| \le \frac{\beta \Gamma}{\phi} \|\boldsymbol{u} - \boldsymbol{v}\|_h^2 + \frac{\beta \Gamma \phi}{2h^2} \|\boldsymbol{u} - \tilde{\boldsymbol{v}}\|_h^2 + \\ + \frac{\beta h^2 \Gamma \phi}{2} \sum_{j=1}^{M} \sum_{i=1}^{N} \left(\left|\nabla_x (u_{ij} - \tilde{v}_{ij})\right|^2 + \left|\nabla_y (u_{ij} - \tilde{v}_{ij})\right|^2 \right), \tag{14}$$

where $\Gamma > 0$ is the Lipschitz constant of condition (iii), $\beta > 0$ is a constant for which $|\nabla_x v_{ij}| \le \beta$ and $|\nabla_y v_{ij}| \le \beta$, and φ is an arbitrary positive number.

Proof. Using formulae (11) and (12) and since $u_{ij} - \tilde{v}_{ij}$ is equal to zero for all the points of ∂R_h, we have

$$< (-\hat{A}(\boldsymbol{u}) + \hat{A}(\boldsymbol{v}))\tilde{\boldsymbol{v}} - \boldsymbol{b}(\boldsymbol{u}) + \boldsymbol{b}(\boldsymbol{v}), \boldsymbol{u} - \tilde{\boldsymbol{v}} >= \sum_{j=1}^{M} [(\sigma(u_{1/2j}) - \sigma(v_{1/2j})) \times \\ \times (\tilde{v}_{1j} - \tilde{v}_{0j})((u_{1j} - \tilde{v}_{1j}) - (u_{0j} - \tilde{v}_{0j})) + \\ + (\sigma(u_{N+1/2j}) - \sigma(v_{N+1/2j}))(\tilde{v}_{Nj} - \tilde{v}_{N+1j})(u_{Nj} - \tilde{v}_{Nj}) - (u_{N+1j} - \tilde{v}_{N+1j})) + \\ + \sum_{i=2}^{N} (\sigma(u_{i-1/2j}) - \sigma(v_{i-1/2j}))(\tilde{v}_{ij} - \tilde{v}_{i-1j})((u_{ij} - \tilde{v}_{ij}) - (u_{i-1j} - \tilde{v}_{i-1j}))] + \\ + \sum_{i=1}^{N} [(\sigma(u_{i1/2}) - \sigma(v_{i1/2}))(\tilde{v}_{i1} - \tilde{v}_{i0})((u_{i1} - \tilde{v}_{i1}) - (u_{i0} - \tilde{v}_{i0})) + \\ + (\sigma(u_{iM+1/2}) - \sigma(v_{iM+1/2}))(\tilde{v}_{iM} - \tilde{v}_{iM+1})((u_{iM} - \tilde{v}_{iM}) - (u_{iM+1} - \tilde{v}_{iM+1})) + \\ + \sum_{j=2}^{M} (\sigma(u_{ij-1/2}) - \sigma(v_{ij-1/2}))(\tilde{v}_{ij} - \tilde{v}_{ij-1})((u_{ij} - \tilde{v}_{ij}) - (u_{ij-1} - \tilde{v}_{ij-1}))].$$

Writing the backward difference quotients (10) for the functions $\tilde{v}_{ij}$ and $u_{ij} - \tilde{v}_{ij}$ and by (9) for $\nabla_x \tilde{v}_{ij}$ and $\nabla_y \tilde{v}_{ij}$, then

$$\left| < (-\hat{A}(\boldsymbol{u}) + \hat{A}(\boldsymbol{v}))\tilde{\boldsymbol{v}} - \boldsymbol{b}(\boldsymbol{u}) + \boldsymbol{b}(\boldsymbol{v}), \boldsymbol{u} - \tilde{\boldsymbol{v}} > \right| \le \beta h^2 \sum_{j=1}^{M} \sum_{i=1}^{N+1} \left| \sigma\left(\frac{u_{ij} + u_{i-1j}}{2}\right) - \sigma\left(\frac{v_{ij} + v_{i-1j}}{2}\right) \right| \times \times \left| \nabla_x (u_{ij} - \tilde{v}_{ij}) \right| + \beta h^2 \sum_{i=1}^{N} \sum_{j=1}^{M+1} \left| \sigma\left(\frac{u_{ij} + u_{ij-1}}{2}\right) - \sigma\left(\frac{v_{ij} + v_{ij-1}}{2}\right) \right| \left| \nabla_y (u_{ij} - \tilde{v}_{ij}) \right| .$$

Since the property of Lipschitz continuity on the function σ with Lipschitz constant Γ (property (iii)), we have

$$\left| < (-\hat{A}(\boldsymbol{u}) + \hat{A}(\boldsymbol{v}))\tilde{\boldsymbol{v}} - \boldsymbol{b}(\boldsymbol{u}) + \boldsymbol{b}(\boldsymbol{v}), \boldsymbol{u} - \tilde{\boldsymbol{v}} > \right| \le \frac{\beta h^2 \Gamma}{2} \sum_{j=1}^{M} \sum_{i=1}^{N+1} \left(|u_{ij} - v_{ij}| + |u_{i-1j} - v_{i-1j}| \right) \times \times \left| \nabla_x (u_{ij} - \tilde{v}_{ij}) \right| + \frac{\beta h^2 \Gamma}{2} \sum_{i=1}^{N} \sum_{j=1}^{M+1} \left(|u_{ij} - v_{ij}| + |u_{ij-1} - v_{ij-1}| \right) \left| \nabla_y (u_{ij} - \tilde{v}_{ij}) \right| .$$

Keeping into account that $u_{ij} - \tilde{v}_{ij}$ and u_{ij} - v_{ij} are equal to zero for the points of ∂R, for an arbitrary positive number φ, we have

$$\left| < (-\hat{A}(\boldsymbol{u}) + \hat{A}(\boldsymbol{v}))\tilde{\boldsymbol{v}} - \boldsymbol{b}(\boldsymbol{u}) + \boldsymbol{b}(\boldsymbol{v}), \boldsymbol{u} - \tilde{\boldsymbol{v}} > \right| \le \frac{\beta h^2 \Gamma}{2} \sum_{j=1}^{M} \sum_{i=1}^{N} \frac{|u_{ij} - v_{ij}|}{\sqrt{\phi}} \left| \nabla_x (u_{ij} - \tilde{v}_{ij}) \right| \sqrt{\phi} + + \frac{\beta h^2 \Gamma}{2} \sum_{j=1}^{M} \sum_{i=1}^{N} \frac{|u_{ij} - v_{ij}|}{\sqrt{\phi}} \left| \nabla_y (u_{ij} - \tilde{v}_{ij}) \right| \sqrt{\phi} + \frac{\beta h^2 \Gamma}{2} \sum_{j=1}^{M} \sum_{i=1}^{N} \frac{|u_{ij} - v_{ij}|}{\sqrt{\phi}} \times \times \left| \nabla_x (u_{i+1j} - \tilde{v}_{i+1j}) \right| \sqrt{\phi} + \frac{\beta h^2 \Gamma}{2} \sum_{j=1}^{M} \sum_{i=1}^{N} \frac{|u_{ij} - v_{ij}|}{\sqrt{\phi}} \left| \nabla_y (u_{ij+1} - \tilde{v}_{ij+1}) \right| \sqrt{\phi}.$$

By the well-known technical trick (for a and b positive numbers: $ab < 1/2a^2 + 1/2b^2$), we can write

$$\begin{aligned} \left| < (-\hat{A}(\boldsymbol{u}) + \hat{A}(\boldsymbol{v}))\tilde{\boldsymbol{v}} - \boldsymbol{b}(\boldsymbol{u}) + \boldsymbol{b}(\boldsymbol{v}), \boldsymbol{u} - \tilde{\boldsymbol{v}} > \right| &\le \frac{\beta h^2 \Gamma}{2} \sum_{j=1}^{M} \sum_{i=1}^{N} \frac{|u_{ij} - v_{ij}|^2}{2\phi} + \\ &+ \frac{\beta h^2 \Gamma}{2} \sum_{j=1}^{M} \sum_{i=1}^{N} \frac{\left| \nabla_x (u_{ij} - \tilde{v}_{ij}) \right|^2}{2} \phi + \frac{\beta h^2 \Gamma}{2} \sum_{j=1}^{M} \sum_{i=1}^{N} \frac{|u_{ij} - v_{ij}|^2}{2\phi} + \\ &+ \frac{\beta h^2 \Gamma}{2} \sum_{j=1}^{M} \sum_{i=1}^{N} \frac{\left| \nabla_y (u_{ij} - \tilde{v}_{ij}) \right|^2}{2} \phi + \frac{\beta h^2 \Gamma}{2} \sum_{j=1}^{M} \sum_{i=1}^{N} \frac{|u_{ij} - v_{ij}|^2}{2\phi} + \\ &+ \frac{\beta h^2 \Gamma}{2} \sum_{j=1}^{M} \sum_{i=1}^{N} \frac{\left| \nabla_x (u_{i+1j} - \tilde{v}_{i+1j}) \right|^2}{2} \phi + \frac{\beta h^2 \Gamma}{2} \sum_{j=1}^{M} \sum_{i=1}^{N} \frac{|u_{ij} - v_{ij}|^2}{2\phi} + \\ &+ \frac{\beta h^2 \Gamma}{2} \sum_{j=1}^{M} \sum_{i=1}^{N} \frac{\left| \nabla_y (u_{ij+1} - \tilde{v}_{ij+1}) \right|^2}{2} \phi \\ &\le \frac{\beta \Gamma}{\phi} \|u - v\|_h^2 + \frac{\beta h^2 \Gamma \phi}{4} \sum_{j=1}^{M} \sum_{i=1}^{N} \left(\left| \nabla_x (u_{ij} - \tilde{v}_{ij}) \right|^2 + \left| \nabla_y (u_{ij} - \tilde{v}_{ij}) \right|^2 \right) + \\ &+ \frac{\beta h^2 \Gamma \phi}{4} \sum_{j=1}^{M} \sum_{i=1}^{N} \left(\left| \nabla_x (u_{i+1j} - \tilde{v}_{i+1j}) \right|^2 + \left| \nabla_y (u_{ij+1} - \tilde{v}_{ij+1}) \right|^2 \right) . \end{aligned}$$

Now, we analyze the term $\sum_{j=1}^{M} \sum_{i=1}^{N} \left(\left| \nabla_x (u_{i+1j} - \tilde{v}_{i+1j}) \right|^2 + \left| \nabla_y (u_{ij+1} - \tilde{v}_{ij+1}) \right|^2 \right)$. We have

$$\sum_{j=1}^{M}\sum_{i=1}^{N}\left(\left|\nabla_x(u_{i+1j}-\tilde{v}_{i+1j})\right|^2+\left|\nabla_y(u_{ij+1}-\tilde{v}_{ij+1})\right|^2\right)\le\sum_{j=1}^{M}\sum_{i=1}^{N+1}\left|\nabla_x(u_{ij}-\tilde{v}_{ij})\right|^2+$$
$$+\sum_{i=1}^{N}\sum_{j=1}^{M+1}\left|\nabla_y(u_{ij}-\tilde{v}_{ij})\right|^2$$
$$\le\sum_{j=1}^{M}\sum_{i=1}^{N}\left(\left|\nabla_x(u_{ij}-\tilde{v}_{ij})\right|^2+\left|\nabla_y(u_{ij}-\tilde{v}_{ij})\right|^2\right)+$$
$$+\sum_{j=1}^{M}\frac{\left|u_{Nj}-\tilde{v}_{Nj}\right|^2}{h^2}+\sum_{i=1}^{N}\frac{\left|u_{iM}-\tilde{v}_{iM}\right|^2}{h^2}$$
$$\le\sum_{j=1}^{M}\sum_{i=1}^{N}\left(\left|\nabla_x(u_{ij}-\tilde{v}_{ij})\right|^2+\left|\nabla_y(u_{ij}-\tilde{v}_{ij})\right|^2\right)+$$
$$+\frac{1}{h^2}\sum_{i=1}^{N}\sum_{j=1}^{M}\left|u_{ij}-\tilde{v}_{ij}\right|^2+\frac{1}{h^2}\sum_{j=1}^{M}\sum_{i=1}^{N}\left|u_{ij}-\tilde{v}_{ij}\right|^2$$
$$\le\sum_{j=1}^{M}\sum_{i=1}^{N}\left(\left|\nabla_x(u_{ij}-\tilde{v}_{ij})\right|^2+\left|\nabla_y(u_{ij}-\tilde{v}_{ij})\right|^2\right)+\frac{2}{h^4}\left\|u-\tilde{v}\right\|_h^2.$$

Then, we have the result (14).

As a consequence of Lemmas 1, 2, and 3, we prove the result of the uniform monotonicity of the mapping $\boldsymbol{F}(\boldsymbol{u})$; thus the nonlinear system (7) has a unique solution in $\mathcal{B}$.

Theorem 3. If

$$\alpha_{\min}+\frac{1}{\tau}>\frac{\sigma_{\min}}{h^2}+\frac{\beta^2\Gamma^2}{2\sigma_{\min}},\qquad(15)$$

then the nonlinear system $\boldsymbol{F}(\boldsymbol{u})=\boldsymbol{0}$, with $\boldsymbol{F}(\boldsymbol{u})$ as in (7), has a unique solution in $\mathcal{B}$.

Proof. We show that the mapping $\boldsymbol{F}(\boldsymbol{u})$ is continuous and uniformly monotone, i.e., there exists a positive scalar γ such that

$$<\boldsymbol{F}(\boldsymbol{u})-\boldsymbol{F}(\boldsymbol{v}),\boldsymbol{u}-\boldsymbol{v}>\quad\ge\quad\gamma<\boldsymbol{u}-\boldsymbol{v},\boldsymbol{u}-\boldsymbol{v}>,\qquad(16)$$

for all $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathcal{B}$. Then, the nonlinear system in (7) has a unique solution [[3], p. 143, 167].

From

$$\frac{1}{\tau}(\boldsymbol{F}(\boldsymbol{u})-\boldsymbol{F}(\boldsymbol{v}))=(-A(\boldsymbol{u})+\frac{1}{\tau}I)(\boldsymbol{u}-\boldsymbol{v})+(-A(\boldsymbol{u})+A(\boldsymbol{v}))\boldsymbol{v}-\boldsymbol{b}(\boldsymbol{u})+\boldsymbol{b}(\boldsymbol{v}),$$

we have

$$\frac{1}{\tau}<\boldsymbol{F}(\boldsymbol{u})-\boldsymbol{F}(\boldsymbol{v}),\boldsymbol{u}-\boldsymbol{v}>=<-\hat{A}(\boldsymbol{u})(\boldsymbol{u}-\boldsymbol{v}),\boldsymbol{u}-\boldsymbol{v}>+<-\tilde{A}(\boldsymbol{u}-\boldsymbol{v}),\boldsymbol{u}-\boldsymbol{v}>+$$
$$+<(-\hat{D}+\frac{1}{\tau}I)(\boldsymbol{u}-\boldsymbol{v}),\boldsymbol{u}-\boldsymbol{v}>+$$
$$+<(-\hat{A}(\boldsymbol{u})+\hat{A}(\boldsymbol{v}))\boldsymbol{v}-\boldsymbol{b}(\boldsymbol{u})+\boldsymbol{b}(\boldsymbol{v}),\boldsymbol{u}-\boldsymbol{v}>.$$

Since $\tilde{A}$ is the skew-symmetric part of $A(\boldsymbol{u})$, it follows that $<-\tilde{A}(\boldsymbol{u}-\boldsymbol{v}),\boldsymbol{u}-\boldsymbol{v}>=0$.

We separately examine the terms $< -\hat{A}(\boldsymbol{u})(\boldsymbol{u}-\boldsymbol{v}), \boldsymbol{u}-\boldsymbol{v} >$, $< (-\hat{D}+\frac{1}{\tau}I)(\boldsymbol{u}-\boldsymbol{v}), \boldsymbol{u}-\boldsymbol{v} >$ and $< (-\hat{A}(\boldsymbol{u})+\hat{A}(\boldsymbol{v}))\boldsymbol{v}-\boldsymbol{b}(\boldsymbol{u})+\boldsymbol{b}(\boldsymbol{v}), \boldsymbol{u}-\boldsymbol{v} >$.

Lemma 2 with $\boldsymbol{u}$-$\boldsymbol{v}$ instead of $\boldsymbol{v}$ and the assumption (ii) on the uniform lower boundedness of σ respect to the variable u permit to write

$$< -\hat{A}(\boldsymbol{u})(\boldsymbol{u}-\boldsymbol{v}), \boldsymbol{u}-\boldsymbol{v} > \geq h^2\sigma_{\min}\sum_{j=1}^{M}\sum_{i=1}^{N}(\left|\nabla_x(u_{ij}-v_{ij})\right|^2+\left|\nabla_y(u_{ij}-v_{ij})\right|^2)+ + \sigma_{\min}\sum_{j=1}^{M}\left|u_{Nj}-v_{Nj}\right|^2+\sigma_{\min}\sum_{i=1}^{N}|u_{iM}-v_{iM}|^2.$$

Since $\alpha(x, y) \geq \alpha_{\min} \geq 0$ we have

$$< (-\hat{D}+\frac{1}{r}I)(\boldsymbol{u}-\boldsymbol{v}), \boldsymbol{u}-\boldsymbol{v} > \geq (\alpha_{\min}+\frac{1}{r})\ \|\boldsymbol{u}-\boldsymbol{v}\|_h^2 .$$

By using Lemma 3 with $\boldsymbol{v}$ instead of $\tilde{\boldsymbol{v}}$, we can write

$$< (-\hat{A}(\boldsymbol{u})+\hat{A}(\boldsymbol{v}))\boldsymbol{v}-\boldsymbol{b}(\boldsymbol{u})+\boldsymbol{b}(\boldsymbol{v}), \boldsymbol{u}-\boldsymbol{v} > \geq -\beta\Gamma(\frac{1}{\phi}+\frac{\phi}{2h^2})\ \|\boldsymbol{u}-\boldsymbol{v}\|_h^2 - -\frac{\beta h^2\Gamma\phi}{2}\sum_{j=1}^{M}\sum_{i=1}^{N}(\left|\nabla_x(u_{ij}-v_{ij})\right|^2+\left|\nabla_y(u_{ij}-v_{ij})\right|^2).$$

Then, collecting the last three inequalities we obtain

$$\frac{1}{\tau} < \boldsymbol{F}(\boldsymbol{u})-\boldsymbol{F}(\boldsymbol{v}), \boldsymbol{u}-\boldsymbol{v} > \geq \left(\alpha_{\min}+\frac{1}{\tau}-\beta\Gamma\left(\frac{1}{\phi}+\frac{\phi}{2h^2}\right)\right)\|\boldsymbol{u}-\boldsymbol{v}\|_h^2 + + h^2\left(\sigma_{\min}-\frac{\beta\Gamma\phi}{2}\right)\sum_{j=1}^{M}\sum_{i=1}^{N}(\left|\nabla_x(u_{ij}-v_{ij})\right|^2+\left|\nabla_y(u_{ij}-v_{ij})\right|^2).$$

If we set

$$\phi = \frac{2\sigma_{\min}}{\beta\Gamma},$$

we obtain

$$\frac{1}{\tau} < \boldsymbol{F}(\boldsymbol{u})-\boldsymbol{F}(\boldsymbol{v}), \boldsymbol{u}-\boldsymbol{v} > \geq \left(\alpha_{\min}+\frac{1}{\tau}-\frac{\beta^2\Gamma^2}{2\sigma_{\min}}-\frac{\sigma_{\min}}{h^2}\right)\|\boldsymbol{u}-\boldsymbol{v}\|_h^2 .$$

When condition (15) holds, then the mapping $\boldsymbol{F}(\boldsymbol{u})$ is uniformly monotone on $\mathcal{B}$ where the constant γ in (16) is

$$\gamma = \tau\left(\alpha_{\min}-\frac{\sigma_{\min}}{h^2}-\frac{\beta^2\Gamma^2}{2\sigma_{\min}}\right)+1.$$

Now, we can state a result for the convergence of the Lagged Diffusivity method where the vector $\boldsymbol{u}^{(k+1)}$ is the approximate solution of the linear system

$$(I-\tau A(\boldsymbol{u}^{(k)}))\boldsymbol{u} = \boldsymbol{w}+\tau\boldsymbol{b}(\boldsymbol{u}^{(k)}), \tag{17}$$

such that the residual

$$r^{(k+1)} = (I - \tau A(u^{(k)}))u^{(k+1)} - w - \tau b(u^{(k)}),$$

satisfies the stopping condition

$$\left\| r^{(k+1)} \right\| \leq \varepsilon_{k+1}, \tag{18}$$

where ε_k is a given tolerance such that $\varepsilon_k \to 0$ when $k \to \infty$.

Thus, the iterate $u^{(k+1)}$ is the solution of the system (7) whose diffusivity term σ in $A(u)$ and $b(u)$ depends on the iterate $u^{(k)}$ and the inhomogeneous term now depends by $u^{(k+1)}$.

We suppose that the grid functions $\{u_{ij}^{(k)}\}, k = 0, 1, \dots$, are belonging to the set $\mathcal{B}$. In particular, the backward difference quotients of each grid function $\{u_{ij}^{(k)}\}$ are bounded. Since this bound depends on the inhomogeneous term, we have that there exist two constants $\beta > 0$ and $\beta_0 > 0$ such that

$$\left| \nabla_x u_{ij}^{(k)} \right| \leq \tilde{\beta}_k \quad \text{and} \quad \left| \nabla_y u_{ij}^{(k)} \right| \leq \tilde{\beta}_k, \tag{19}$$

with $\tilde{\beta}_k = \beta + \varepsilon_k \beta_0$. (Formula (19) replaces formula (9).)

Theorem 4. Let $u^* \in \mathcal{B}$ be the solution of the nonlinear system $F(u) = \mathbf{0}$ in (7). Let $u^{(k+1)}$ be the solution of the linear system in (17) with condition (18). If condition (15) is satisfied, and, in particular

$$\alpha_{\min} + \frac{1}{\tau} > \frac{\sigma_{\min}}{h^2}, \tag{20}$$

then, the sequence $\{u^{(k)}\}$ converges to u^*.

Proof. The solution $u^* \in \mathcal{B}$ of (7) satisfies the equation

$$u^* - \tau A(u^*)u^* - w - \tau b(u^*) = \mathbf{0}, \tag{21}$$

and the iterate $u^{(k+1)}$ satisfies the equation

$$u^{(k+1)} - \tau A(u^{(k)})u^{(k+1)} - w - \tau b(u^{(k)}) = r^{(k+1)}. \tag{22}$$

Subtracting (22) from (21), we obtain

$$-A(u^*)u^* + A(u^{(k)})u^{(k+1)} + \frac{1}{\tau}u^* - \frac{1}{\tau}u^{(k+1)} - b(u^*) + b(u^{(k)}) = -\frac{1}{\tau}r^{(k+1)}.$$

Taking into account of the identity

$$-A(u)u + A(w)v = -A(u)(u - v) + (-A(u) + A(w))v,$$

for all u, v and w belonging to $\mathcal{B}$, we can write

$$\left(-A(u^*) + \frac{1}{\tau}I\right)(u^* - u^{(k+1)}) + (-A(u^*) + A(u^{(k)}))u^{(k+1)} - b(u^*) + b(u^{(k)}) = \frac{1}{\tau}r^{(k+1)}.$$

Thus, we have

$$< \left(-A(\boldsymbol{u}^*) + \frac{1}{\tau}I\right)(\boldsymbol{u}^* - \boldsymbol{u}^{(k+1)}), \boldsymbol{u}^* - \boldsymbol{u}^{(k+1)} > + \\ + < (-A(\boldsymbol{u}^*) + A(\boldsymbol{u}^{(k)}))\boldsymbol{u}^{(k+1)} - \boldsymbol{b}(\boldsymbol{u}^*) + \boldsymbol{b}(\boldsymbol{u}^{(k)}), \boldsymbol{u}^* - \boldsymbol{u}^{(k+1)} > \\ =< -\frac{1}{\tau}\boldsymbol{r}^{(k+1)}, \boldsymbol{u}^* - \boldsymbol{u}^{(k+1)} > .$$

Since $\tilde{A}$ is the skew-symmetric part of $A(\boldsymbol{u})$, for any vector $\boldsymbol{z} \in \mathbb{R}^\mu$ we have

$$< (\hat{A}(\boldsymbol{u}) + \tilde{A} + \hat{D})\boldsymbol{z}, \boldsymbol{z} >=< (\hat{A}(\boldsymbol{u}) + \hat{D})\boldsymbol{z}, \boldsymbol{z} >,$$

and then, we can write

$$< -\frac{1}{\tau}\boldsymbol{r}^{(k+1)}, \boldsymbol{u}^* - \boldsymbol{u}^{(k+1)} > \quad \geq \\ \geq < -\hat{A}(\boldsymbol{u}^*)(\boldsymbol{u}^* - \boldsymbol{u}^{(k+1)}), \boldsymbol{u}^* - \boldsymbol{u}^{(k+1)} > + < (-\hat{D} + \frac{1}{\tau}I)(\boldsymbol{u}^* - \boldsymbol{u}^{(k+1)}), \boldsymbol{u}^* - \boldsymbol{u}^{(k+1)} > - \\ - \left| < (-\hat{A}(\boldsymbol{u}^*) + \hat{A}(\boldsymbol{u}^{(k)}))\boldsymbol{u}^{(k+1)} - \boldsymbol{b}(\boldsymbol{u}^*) + \boldsymbol{b}(\boldsymbol{u}^{(k)}), \boldsymbol{u}^* - \boldsymbol{u}^{(k+1)} > \right| .$$

From Lemmas 2 and 3 we obtain

$$\begin{aligned} & < -\frac{1}{\tau}r^{(k+1)}, u^* - u^{(k+1)} > \quad \geq \\ \geq \quad & h^2\sigma_{\min}\sum_{i=1}^{N}\sum_{j=1}^{M}\left(\left|\nabla_x\left(u_{ij}^* - u_{ij}^{(k+1)}\right)\right|^2 + \left|\nabla_y(u_{ij}^* - u_{ij}^{(k+1)})\right|^2\right) + \\ & + \sigma_{\min}\sum_{j=1}^{M}\left|u_{Nj}^* - u_{Nj}^{(k+1)}\right|^2 + \sigma_{\min}\sum_{i=1}^{N}\left|u_{iM}^* - u_{iM}^{(k+1)}\right|^2 + \\ & + \left(\alpha_{\min} + \frac{1}{\tau}\right)\left\|u^* - u^{(k+1)}\right\|_h^2 - \\ & - \frac{\tilde{\beta}_{k+1}\Gamma}{\phi}\left\|u^* - u^{(k)}\right\|_h^2 - \frac{\tilde{\beta}_{k+1}\Gamma\phi}{2h^2}\left\|u^* - u^{(k+1)}\right\|_h^2 - \\ & - \frac{\tilde{\beta}_{k+1}h^2\Gamma\phi}{2}\sum_{j=1}^{M}\sum_{i=1}^{N}\left(\left|\nabla_x(u_{ij}^* - u_{ij}^{(k+1)})\right|^2 + \left|\nabla_y(u_{ij}^* - u_{ij}^{(k+1)})\right|^2\right) \geq \\ \geq & h^2\left(\sigma_{\min} - \frac{\tilde{\beta}_{k+1}\Gamma\phi}{2}\right)\sum_{i=1}^{N}\sum_{j=1}^{M}\left(\left|\nabla_x(u_{ij}^* - u_{ij}^{(k+1)})\right|^2 + \left|\nabla_y(u_{ij}^* - u_{ij}^{(k+1)})\right|^2\right) + \\ & + \left(\alpha_{\min} + \frac{1}{\tau}\right)\left\|u^* - u^{(k+1)}\right\|_h^2 - \\ & - \frac{\tilde{\beta}_{k+1}\Gamma}{\phi}\left\|u^* - u^{(k)}\right\|_h^2 - \frac{\tilde{\beta}_{k+1}\Gamma\phi}{2h^2}\left\|u^* - u^{(k+1)}\right\|_h^2 . \end{aligned} \tag{23}$$

Keeping into account of the stopping condition (18) we can write

$$\frac{1}{\tau}\varepsilon_{k+1}\left\|\boldsymbol{u}^* - \boldsymbol{u}^{(k+1)}\right\|_h \geq \frac{1}{\tau}\left\|\boldsymbol{r}^{(k+1)}\right\|\left\|\boldsymbol{u}^* - \boldsymbol{u}^{(k+1)}\right\|_h \geq < -\frac{1}{\tau}\boldsymbol{r}^{(k+1)}, \boldsymbol{u}^* - \boldsymbol{u}^{(k+1)} > . \tag{24}$$

Since φ is an arbitrary positive number, we can choose φ such that in (23)

$$\sigma_{\min} - \frac{\tilde{\beta}_{k+1}\Gamma\phi}{2} = 0,$$

that is

$$\phi = \frac{2\sigma_{\min}}{\tilde{\beta}_{k+1}\Gamma}.$$

Thus, from (23) and (24) we have

$$\frac{1}{\tau}\varepsilon_{k+1}\left\|\boldsymbol{u}^* - \boldsymbol{u}^{(k+1)}\right\|_h \geq \left(\alpha_{\min} + \frac{1}{\tau} - \frac{\sigma_{\min}}{h^2}\right)\left\|\boldsymbol{u}^* - \boldsymbol{u}^{(k+1)}\right\|_h^2 - \frac{\tilde{\beta}_{k+1}^2\Gamma^2}{2\sigma_{\min}}\left\|\boldsymbol{u}^* - \boldsymbol{u}^{(k)}\right\|_h^2.$$

Since the grid function $\{u_{ij}^{(k+1)}\}$ belongs to $\mathcal{B}$, it then satisfies the inequality (8) and so we have

$$\left\|\boldsymbol{u}^* - \boldsymbol{u}^{(k+1)}\right\|_h \leq 2\rho,$$

then

$$\frac{2\rho}{\tau}\varepsilon_{k+1} \geq \left(\alpha_{\min} + \frac{1}{\tau} - \frac{\sigma_{\min}}{h^2}\right)\left\|\boldsymbol{u}^* - \boldsymbol{u}^{(k+1)}\right\|_h^2 - \frac{\tilde{\beta}_{k+1}^2\Gamma^2}{2\sigma_{\min}}\left\|\boldsymbol{u}^* - \boldsymbol{u}^{(k)}\right\|_h^2.$$

We assume that condition (15) holds. Then, keeping into account of the expression of $\tilde{\beta}_k$, we have the inequality

$$\left\|\boldsymbol{u}^* - \boldsymbol{u}^{(k+1)}\right\|_h^2 \leq \hat{\gamma}\left\|\boldsymbol{u}^* - \boldsymbol{u}^{(k)}\right\|_h^2 + \hat{a}\varepsilon_{k+1}, \tag{25}$$

where

$$a = \alpha_{\min} + \frac{1}{\tau} - \frac{\sigma_{\min}}{h^2} > 0, \quad \hat{\gamma} = \frac{(\beta + \varepsilon_{k+1}\beta_0)^2\Gamma^2}{2\sigma_{\min}a}, \quad \hat{a} = \frac{2\rho}{\tau a}.$$

Now, since there exists an integer k_0 such that $\hat{\gamma} < 1$ for all $k \geq k_0$, we can write formula (25) as

$$\left\|\boldsymbol{u}^* - \boldsymbol{u}^{(k_0+r)}\right\|_h^2 \leq \hat{\gamma}^r\left\|\boldsymbol{u}^* - \boldsymbol{u}^{(k_0)}\right\|_h^2 + \hat{a}\sum_{j=1}^{r}\hat{\gamma}^{r-j}\varepsilon_{k_0+j},$$

$r = 1, 2, \ldots$, and since $\varepsilon_k \to 0$ for $k \to \infty$, it follows from the general Toeplitz Lemma (e.g., see [[3], p. 399]) that

$$\lim_{k\to\infty}\left\|\boldsymbol{u}^* - \boldsymbol{u}^{(k)}\right\|_h^2 = 0.$$

Therefore, the sequence $\{\boldsymbol{u}^{(k)}\}$ of approximate solutions of the Lagged Diffusivity method converges to the solution $\boldsymbol{u}^*$ of the system (7).

3 Numerical experiments

In this section, we consider a numerical experimentation of the Lagged Diffusivity method for the solution of the nonlinear system generated by the θ-method applied on the model problem (1) in a rectangular domain with Dirichlet boundary conditions. Indeed we solve with the Lagged Diffusivity procedure the nonlinear system

$$(I - \tau A(\boldsymbol{u}))\boldsymbol{u} - \boldsymbol{w} - \boldsymbol{b}(\boldsymbol{u}) = \boldsymbol{0}.$$

The Lagged Diffusivity procedure is an efficient and robust method, even if only linearly convergent, for solving the digital image restoration problem using diffusive filters.

One of the most used diffusive filter is defined (in equation (1) with $(\tilde{\boldsymbol{v}} = \mathbf{0})$) by a diffusivity $\sigma(u)$ chosen as a rapidly decreasing function of the gradient magnitude $|\nabla u|^2$. Specifically, $\sigma(s) : [0, \infty) \rightarrow (0, \infty)$ is a decreasing function satisfying $\sigma(0) = 1$ and $\lim_{s\rightarrow\infty}\sigma(s) = 0$ and $s = |\nabla u|^2 = (\partial u/\partial x)^2 + (\partial u/\partial y)^2$.

Due to the presence of this term $\sigma(u)$, the operator $\mathrm{div}(\sigma\nabla u)$ is highly nonlinear and, when linearized by lagging the diffusivity σ, it has highly varying coefficients [4-8].

Often happens in real images that $|\nabla u| = 0$. This makes necessary the use of numerical regularization, consisting in replacing the term $|\nabla u|$ by $(|\nabla u|^2 + \beta)^{1/2}$ for a "small enough" positive artificial parameter β. Due to the presence of the highly nonlinear operator $\mathrm{div}(\sigma\nabla u)$, Newton's method for solving the nonlinear system (7) does not work satisfactory, in the sense that its domain of convergence is very small.

These diffusion problems for image restoration are also solved using operator splitting methods (e.g., [9-11]). Operator splitting is a powerful concept used in Computational Mathematics for the design of effective numerical methods. These splitting methods are essentially based on certain special relaxation processes which allow one, to reduce the complicated problem into a sequence of simpler problems which can be effectively solved with a computer.

At present, there exists a lot of interest in the applications of operator splitting methods to problems of Financial Mathematics, and, in particular, to diffusion convection equations with mixed derivatives terms (e.g., [12-14]).

Operator splitting is also a very common tool for solving nonlinear evolution equations which include hyperbolic conservation laws and degenerate diffusion convection equations with nonsmooth solutions. In these evolution equations, the convection dominates diffusion (e.g., [15-19]).

An alternative case has been considered in our computational experiments. Indeed, we consider diffusion convection problems which are of diffusion dominated nature, as those concerning the groundwater transport of a contaminant in an aquifer or the control of heating processes of industrial kilns.

The convergence of the "outer" iteration $\{\boldsymbol{u}^{(k)}\}$ to $\boldsymbol{u}^*$ of the Lagged Diffusivity procedure involves solution for an unknown vector $\boldsymbol{u}$ of the matrix equation (17). This linear system may be solved by an operator splitting method. The effect of this "inner" iteration on convergence of the "outer" iteration to $\boldsymbol{u}^*$ must be analyzed in order to define a good strategy for the convergence of the Lagged Diffusivity procedure. Indeed, a significant reduction in total effort can often be achieved by proper coordination of inner and outer iteration.

In the experiments, the vector solution $\boldsymbol{u}^*$ is prefixed and is composed by the values of prescribed functions $u(x, y, t)$ defined on $[a, b] \times [a, b] \times [0, \infty)$. In all the experiments, we have $a = 0$ and $b = 1$. We choose different solution functions where the time value t is set equal to 1:

$$u1 : u(x, y, t) = xyt,$$

$$u2 : u(x, y, t) = \sum_{i=1}^{3} 15e^{-8((x-\xi_i)^2+(y-\eta_i)^2)t},$$

$$u3 : u(x, y, t) = (x + y)t,$$

$$u4 : u(x, y, t) = (1 + x - y)^3 t.$$

For the function $u2$, we have $\xi_1 = \eta_1 = 0.2$, $\xi_2 = 0.2$, $\eta_2 = 0.8$ and $\xi_3 = \eta_3 = 0.8$.

The chosen functions $\sigma(u)$ are:

$$\sigma 1 : \sigma(u) = 0.01 + 0.5u,$$
$$\sigma 2 : \sigma(u) = 0.01 + 0.5u^2,$$
$$\sigma 3 : \sigma(u) = \frac{100}{1 + 500u}.$$

The vector $\boldsymbol{w}$ is computed as

$$\boldsymbol{w} \equiv \boldsymbol{w}^* = (I - \tau A(\boldsymbol{u}^*))\boldsymbol{u}^* - \boldsymbol{b}(\boldsymbol{u}^*),$$

where the matrix $A(\boldsymbol{u}^*)$ and the vector $\boldsymbol{b}(\boldsymbol{u}^*)$ have elements as in formula (5) with $N = M$ and $\alpha(x, y) = 0$. We set $\theta = 1$ and $\Delta t = 10^{-3}$ or $\Delta t = 10^{-4}$; $\tau = \theta\, \Delta t$. Here, the condition (20) holds.

At each iteration k of the Lagged Diffusivity procedure, we have to solve the linear system of order $\mu = N \times N$:

$$(I - \tau A(\boldsymbol{u}^{(k)}))\boldsymbol{u} = \boldsymbol{w}^* + \boldsymbol{b}(\boldsymbol{u}^{(k)}).$$

We consider the case that the coefficient matrix is an M-matrix ($\tilde{\boldsymbol{v}} \neq 0$).

In these experiments, the iterative method of the Arithmetic Mean is used as linear solver in the form introduced in [20]. This method is convergent when the coefficient matrix is a nonsingular M-matrix. This form of the Arithmetic Mean method is a variant of the Alterating Group Explicit (AGE) decomposition introduced by Evans (e.g., see [21-23]). The effectiveness of the Arithmetic Mean method, even on parallel architectures, is highlighted in [20,24-27]. Some recent papers on the Arithmetic Mean method on linear systems are in [28-31].

We call $\boldsymbol{u}^{(k+1)}$ the solution of the linear system above computed with j_k iterations of the Arithmetic Mean solver where the inner residual

$$\boldsymbol{r}^{(k+1)} = (I - \tau A(\boldsymbol{u}^{(k)}))\boldsymbol{u}^{(k+1)} - \boldsymbol{w}^* - \boldsymbol{b}(\boldsymbol{u}^{(k)}),$$

satisfies the condition

$$\left\|\boldsymbol{r}^{(k+1)}\right\| \le \varepsilon_{k+1},$$

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

$$\varepsilon_{k+1} = 0.5\varepsilon_k. \tag{26}$$

The vector $\boldsymbol{F}(\boldsymbol{u}^{(0)}) = (I - \tau A(\boldsymbol{u}^{(0)}))\boldsymbol{u}^{(0)} - \boldsymbol{w}^* - \boldsymbol{b}(\boldsymbol{u}^{(0)})$ is the initial outer residual and its Euclidean norm is called *res0*.

The initial vector $\boldsymbol{u}^{(0)}$ is taken as the null vector ($\boldsymbol{u}^{(0)} = \boldsymbol{0}$) or as the vector $\boldsymbol{e}$ which is the vector with all the components equal to 1 ($\boldsymbol{u}^{(0)} = \boldsymbol{e}$).

The Lagged Diffusivity procedure has been implemented in a Fortran code with machine precision 2.2×10^{-16} and stops when

$$\varepsilon_{k+1} \le \underline{\varepsilon}, \tag{27}$$

with $\underline{\varepsilon} = 10^{-4}$.

An experiment evaluates the effectiveness of the Lagged Diffusivity method for different values of $\underline{\varepsilon}$; another experiment shows the behavior of the method for different choices of ε_{k+1} respect to the one in (26) with $\underline{\varepsilon} = 10^{-4}$.

We call k^* the iteration of the Lagged Diffusivity procedure for which condition (27) is satisfied.

In the tables, we report the number of iterations k^*, the total number of iterations of the Arithmetic Mean method j_T, the discrete $l_2(R_h)$ norm of the error, $err = \|\boldsymbol{u}^* - \boldsymbol{u}^{(k^*)}\|_h$, the Euclidean norm of the outer residual

$$res = \left\|F(\boldsymbol{u}^{(k^*)})\right\| = \left\|(I - \tau A(\boldsymbol{u}^{(k^*)}))\boldsymbol{u}^{(k^*)} - \boldsymbol{w}^* - \boldsymbol{b}(\boldsymbol{u}^{(k^*)})\right\|,$$

and *res0*.

The symbol * close to the value of *res* indicates that the behavior of the norm of the outer residual ($\|\boldsymbol{F}(\boldsymbol{u}^{(k)})\|$) is not monotone decreasing. In addition, $\bar{\sigma}^*$ is an approximation of $\sigma_{\max}$.

Furthermore, in the tables, writing 5.26(-7) means 5.26×10^{-7}.

From the numerical experiments, we can drawn the below conclusions (See Tables 1, 2 and 3).

- We observe that, since ε_{k+1} decreases, for k increasing, as (26) and the Lagged Diffusivity method stops at the iteration k^* when the criterium for ε_{k^*+1} in (27) is satisfied, we have

$$\varepsilon_{k^*+1} = \frac{1}{2}\varepsilon_{k^*} = \frac{1}{2^2}\varepsilon_{k^*-1} = \cdots = \frac{1}{2^{k^*}}\varepsilon_1 \le \underline{\varepsilon},$$

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

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

In the experiments, we obtain

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

- We observe that in all the experiments with the rule (26) the outer residual $\|\boldsymbol{F}(\boldsymbol{u}^{(k^*)})\|$ has the same order of $\underline{\varepsilon}$ with an error in the discrete $l_2(R_h)$ norm of order of $h\underline{\varepsilon}$.
- From the experiments, the choice for ε_{k+1} as in (26) gives satisfactory results in terms of iterations of the inner solver and in relation to the outer residual and the error.

Table 1 Results with $N = 256$, $\tilde{\nu}_1 = \tilde{\nu}_2 = 1$, $u^{(0)} = 0$

u(*x, y, t*)	$\bar{\sigma}^*/\tilde{\nu}_1$	*k**	j_T	*err*	*res*	*res 0*
$\tau = 10^{-3}$						
$\sigma(u) = \sigma 1$						
*u*1	0.51	18	260	5.26(-7)	1.51(-4)	205.61
*u*2	8.52	26	5972	4.81(-7)	1.48(-4)	51451.65
*u*3	1.01	20	585	7.00(-7)	1.97(-4)	1041.69
*u*4	4.01	24	1607	3.46(-7)	1.13(-4)	9617.82
$\sigma(u) = \sigma 2$						
*u*1	0.51	17	196	5.96(-7)	1.80(-4)	123.15
✔ *u*2	145.03	29	35427	2.19(-7)	9.73(-5)*	276108.92
*u*3	2.01	20	899	5.59(-7)	1.71(-4)	898.18
✔ *u*4	32.01	25	3960	4.70(-7)	1.85(-4)*	31481.01
$\tau = 10^{-4}$						
$\sigma(u) = \sigma 1$						
*u*1	0.51	17	33	4.22(-7)	1.12(-4)	89.44
*u*2	8.52	23	676	5.40(-7)	1.45(-4)	6203.62
*u*3	1.01	19	70	3.75(-7)	9.94(-5)	304.43
*u*4	4.01	21	216	3.80(-7)	1.10(-4)	1170.79
$\sigma(u) = \sigma 2$						
*u*1	0.51	17	30	3.54(-7)	9.63(-5)	86.61
*u*2	145.03	25	7160	4.65(-7)	1.55(-4)	28054.93
*u*3	2.01	19	115	4.00(-7)	1.10(-4)	297.76
*u*4	32.01	22	852	4.43(-7)	1.53(-4)	3266.07
$\sigma(u) = \sigma 3$						
*u*2	1.50	22	1534	4.19(-7)	1.30(-4)	3061.65
*u*3	100.00	19	1515	3.16(-7)	1.10(-4)	297.96
*u*4	100.00	20	1986	2.67(-7)	1.06(-4)	558.51

- About the initial vectors, we can say that, generally, the null vector is a good choice, in terms of total number of inner iterations, for $\sigma(u) = \sigma 1$ and $\sigma(u) = \sigma 2$; while for the functions $u\,(x, y, t)$ and $\sigma(u) = \sigma 3$ we have better results when $\boldsymbol{u}^{(0)} = \boldsymbol{e}$. A detection of that the null vector is not a good initial vector for the problems with $\sigma(u) = \sigma 3$, is the number of inner iterations at the first outer iteration. Indeed, for the functions $u2$, $u3$, and $u4$, most of the inner iterations happen at the first iteration, $k = 1$; for instance, in Table 1 when $\sigma(u) = \sigma 3$ and $\tau = 10^{-4}$, at $k = 1$, for $u2$ we have 1,484 inner iterations on the total number 1,534, for $u3$ we have 1,421 inner iterations on the total number 1,515, for $u4$ we have 1,407 inner iterations on the total number 1,986.
- When the behavior of the norm of the outer residual $\boldsymbol{F}(\boldsymbol{u}^{(k)})$ is not monotone decreasing (i.e., $\sigma(u) = \sigma 2$, $u(x, y, t) = u2$, $u4$, $\tau = 10^{-3}$) we can have a large total number of inner iterations at an outer iteration. We suggest to change the initial vector to obtain a monotone decreasing of the norm of the outer residual that implies a reduction of the total number of inner iterations. Changing the initial vector $\boldsymbol{u}^{(0)}$, the rows marked with "✔" in Table 1 become the results of Table 4.
- From Table 1, we observe that the Arithmetic Mean method gives better performances when the ratio $\overline{\sigma}^*/\tilde{\nu}_1(\tilde{\nu}_1 = \tilde{\nu}_2)$ is small, that is the coefficient matrix of the inner linear system is strongly asymmetric (see [24]).

Table 2 Results with N = 256, $\tilde{\nu}_1 = \tilde{\nu}_2 = 1$, $u^{(0)} = e$

u(*x, y, t*)	*k**	j_T	*Err*	*res*	*res0*
$\tau = 10^{-3}$					
$\sigma(u) = \sigma 1$					
*u*1	20	302	3.69(-7)	1.03(-4)	546.04
*u*2	26	6066	4.76(-7)	1.47(-4)	50940.66
*u*3	20	633	4.63(-7)	1.31(-4)	692.97
*u*4	24	1631	3.37(-7)	1.10(-4)	9323.23
$\sigma(u) = \sigma 2$					
*u*1	19	249	4.51(-7)	1.32(-4)	362.86
*u*2	29	38837	2.36(-7)	1.06(-4)*	298830.38
*u*3	20	932	5.45(-7)	1.67(-4)	881.87
*u*4	26	4122	2.63(-7)	1.03(-4)*	35275.82
$\tau = 10^{-4}$					
$\sigma(u) = \sigma 1$					
*u*1	18	37	5.17(-7)	1.34(-4)	211.75
*u*2	23	737	5.23(-7)	1.40(-4)	6026.72
*u*3	18	76	3.73(-7)	9.89(-5)	132.60
*u*4	21	240	3.46(-7)	1.00(-4)	1074.69
$\sigma(u) = \sigma 2$					
*u*1	18	33	4.50(-7)	1.19(-4)	205.39
*u*2	25	7615	5.02(-7)	1.68(-4)*	30266.34
*u*3	18	121	3.82(-7)	1.05(-4)	144.41
*u*4	22	910	4.89(-7)	1.69(-4)	3611.96
$\sigma(u) = \sigma 3$					
*u*2	22	48	3.98(-7)	1.24(-4)	2825.73
*u*3	17	102	4.76(-7)	1.66(-4)	113.34
*u*4	19	604	3.99(-7)	1.58(-4)	416.74

- A very general conclusion is that, at each time step t_n, the Lagged Diffusivity method (17)-(18)-(26)-(27) allows to obtain an approximate solution of the nonlinear system (7) with sufficiently high accuracy and low computational complexity, when the initial vector $\boldsymbol{u}^{(0)}$ has been chosen properly.

Remark. Given the exact solution $u(x, y, t)$ of equations (1), (3) at time $t = t_{n+1}$ (for example, $t_{n+1} = 1$), in the above numerical experiments, we have considered in system (7) the "source" vector $\boldsymbol{w}$ as $\boldsymbol{w} \equiv \boldsymbol{w}^* = (I\text{-}\tau A(\boldsymbol{u}^*))\ \boldsymbol{u}^*\ \text{-}\boldsymbol{b}(\boldsymbol{u}^*)$ with $\boldsymbol{u}^* = \{u(x_i, y_j, t_{n+1})\}$.

With this definition of $\boldsymbol{w}$, it has been possible to obtain, for our test problems, the behavior of the error associated with the numerical computation of the solution of (7), i.e., an indication of the accuracy on the solution of the nonlinear equations (7). We have denoted this error by *err*.

In order to analyze the behavior of the *effective one-step error* (in the discrete l_2 (R_h) norm) of the θ-method, denoted by E_{step}, we must consider in system (7) the "source" vector $\boldsymbol{w}$ as

$$\boldsymbol{w} \equiv \boldsymbol{w}^n = \left(I + \Delta t(1-\theta)A(\boldsymbol{u}^n)\right)\boldsymbol{u}^n + \Delta t(1-\theta)\boldsymbol{b}(\boldsymbol{u}^n) + \Delta t(\theta \boldsymbol{s}^{n+1} + (1-\theta)\boldsymbol{s}^n)$$

with $\boldsymbol{u}^n = \{u(x_i, y_j, t_n)\}$ and $t_n = t_{n+1} - \Delta t$.

From Table 5, we observe that the values of E_{step} are comparable with those of *err*, with the exception of the test problems $\sigma(u) = \sigma 1, \sigma 2$, $u(x, y, t) = u2$ for $\tau = 10^{-3}, 10^{-4}$,

Table 3 Results for different $\underline{\varepsilon}$ and ε_{k+1} ($u^{(0)}$ = 0)

$N = 256; u = u1; \sigma(u) = \sigma 1; \tilde{v}_1 = \tilde{v}_2 = 1; \tau = 10^{-3}; res0 = 205.61$				
$\underline{\varepsilon}$	k^*	j_T	err	res
10^{-3}	15	210	4.21(-6)	1.20(-3)
10^{-4}	18	260	5.26(-7)	1.51(-4)
10^{-5}	21	310	6.59(-8)	1.90(-5)
10^{-6}	25	377	4.09(-9)	1.18(-6)
$N = 256; u = u1; \sigma(u) = \sigma 1; \tilde{v}_1 = \tilde{v}_2 = 1; \tau = 10^{-3}; res0 = 205.61$				
ε_{k+1}	k^*	j_T	err	res
$0.7\varepsilon_k$	35	268	3.85(-7)	1.10(-4)
$0.5\varepsilon_k$	18	260	5.26(-7)	1.51(-4)
$0.1\ \varepsilon_k$	6	250	7.08(-7)	2.01(-4)
$0.05\varepsilon_k$	5	260	4.36(-7)	1.23(-4)
$0.01\varepsilon_k$	3	189	7.67(-6)	2.31(-3)
$N = 256; u = u4; \sigma(u) = \sigma 1; \tilde{v}_1 = \tilde{v}_2 = 1; \tau = 10^{-3}; res0 = 9617.82$				
ε_{k+1}	k^*	j_T	err	res
$0.7\varepsilon_k$	46	1625	3.10(-7)	1.02(-4)
$0.5\varepsilon_k$	24	1607	3.46(-7)	1.13(-4)
$0.1\ \varepsilon$	7	1327	3.06(-6)	9.34(-4)
$0.05\varepsilon_k$	6	1408	1.02(-6)	3.09(-4)
$0.01\ \varepsilon_k$	4	1384	4.12(-5)	2.46(-2)
$N = 256; u = u4; \sigma(u) = \sigma 2; \tilde{v}_1 = \tilde{v}_2 = 1; \tau = 10^{-3}; res0 = 31481.01$				
ε_{k+1}	k^*	j_T	err	res
$0.7\varepsilon_k$	49	4146	2.78(-7)	1.14(-4)*
$0.5\varepsilon_k$	25	3960	4.70(-7)	1.85(-4)*
$0.1\ \varepsilon_k$	8	4930	1.75(-5)	1.42(-2)*
$0.05\varepsilon_k$	6	4827	3.16(-4)	0.27*
$0.01\ \varepsilon_k$	4	3800	4.70(-3)	4.63*

where the discrepancy between *err* and E_{step} depends on the fact that the matrix $A(\boldsymbol{u})$ is ill conditioned because the Lipschitz constant is large.

Furthermore, the global error for solving the problem (1)-(2)-(3) with the θ-method combined with the Lagged Diffusivity procedure is computed for the cases

$$\begin{aligned} c1 &: \theta = 1, \quad u(x, y, t) = u2, \sigma(u) = \sigma 1, \\ c2 &: \theta = 0.5, u(x, y, t) = u2, \sigma(u) = \sigma 1, \\ c3 &: \theta = 1, \quad u(x, y, t) = u4, \sigma(u) = \sigma 1, \\ c4 &: \theta = 0.5, u(x, y, t) = u4, \sigma(u) = \sigma 1, \end{aligned}$$

and it is denoted with $E(c1)$, $E(c2)$, $E(c3)$, and $E(c4)$.

The behavior of this global error, step-by-step, from $1 \le t \le 1.2$ and $\Delta t = 10^{-3}$ is highlighted in Figure 1; the numerical results are seen to be largely in keeping with the theory.

Table 4 Results for the cases in Table 1 marked with "✓" with different initial vectors

$u^{(0)}$	u (x, y, t)	k*	j_T	err	res	res 0
10$\boldsymbol{e}$	$u2$	28	26839	3.01(-7)	1.33(-4)	189726.62
4$\boldsymbol{e}$	$u\ 4$	26	3669	2.67(-7)	1.05(-4)	35807.59

Table 5 Results with $N = 256,\ \tilde{\nu}_1 = \tilde{\nu}_2 = 1,\ \boldsymbol{u}^{(0)} = \boldsymbol{u}^n,\ \boldsymbol{w} = \boldsymbol{w}^n$

u(x, y, t)	$\bar{\sigma}^*/\tilde{\nu}_1$	*k**	j_T	E_{step}	*res*	*res0*	‖*w**-*w*^{n}‖
$\tau = 10^{-3}$							
$\sigma(u) = \sigma 1$							
*u*1	0.51	9	143	4.84(-7)	1.37(-4)	0.36	1.16(-12)
*u*2	8.52	17	3850	7.25(-5)	1.01(-4)	68.40	3.58(-2)
*u*3	1.01	11	341	6.78(-7)	1.89(-4)	1.97	3.51(-12)
*u*4	4.01	15	967	2.65(-7)	1.15(-4)	19.05	8.34(-5)
$\sigma(u) = \sigma 2$							
*u*1	0.51	9	120	3.97(-7)	1.18(-4)	0.31	9.75(-13)
*u*2	145.03	20	17950	1.94(-4)	1.32(-4)	734.79	7.13(-1)
*u*3	2.01	12	538	5.24(-7)	1.58(-4)	3.27	6.79(-12)
*u*4	32.01	17	2452	2.46(-6)	1.88(-4)	125.38	1.50(-3)
$\tau = 10^{-4}$							
$\sigma(u) = \sigma 1$							
*u*1	0.51	4	12	4.06(-7)	1.07(-4)	9.63(-3)	1.17(-13)
*u*2	8.52	10	341	1.24(-5)	1.39(-4)	0.73	3.56(-3)
*u*3	1.01	6	30	4.05(-7)	1.06(-4)	3.56(-2)	3.52(-13)
*u*4	4.01	8	95	5.35(-7)	1.58(-4)	0.20	8.34(-6)
$\sigma(u) = \sigma 2$							
*u*1	0.51	4	10	4.29(-7)	1.15(-4)	9.37(-3)	9.86(-14)
*u*2	145.03	13	3585	1.01(-4)	1.67(-4)	7.36	7.13(-2)
*u*3	2.01	6	46	5.12(-7)	1.39(-4)	4.49(-2)	6.81(-13)
*u*4	32.01	11	413	2.84(-7)	1.23(-4)	1.26	1.50(-4)
$\sigma(u) = \sigma 3$							
*u*2	1.50	8	15	5.68(-7)	1.43(-4)	0.21	7.58(-5)
*u*3	100.00	6	19	1.08(-6)	1.00(-4)	3.38(-2)	1.86(-2)
*u*4	100.00	6	24	5.22(-7)	1.66(-4)	5.49(-2)	8.45(-5)

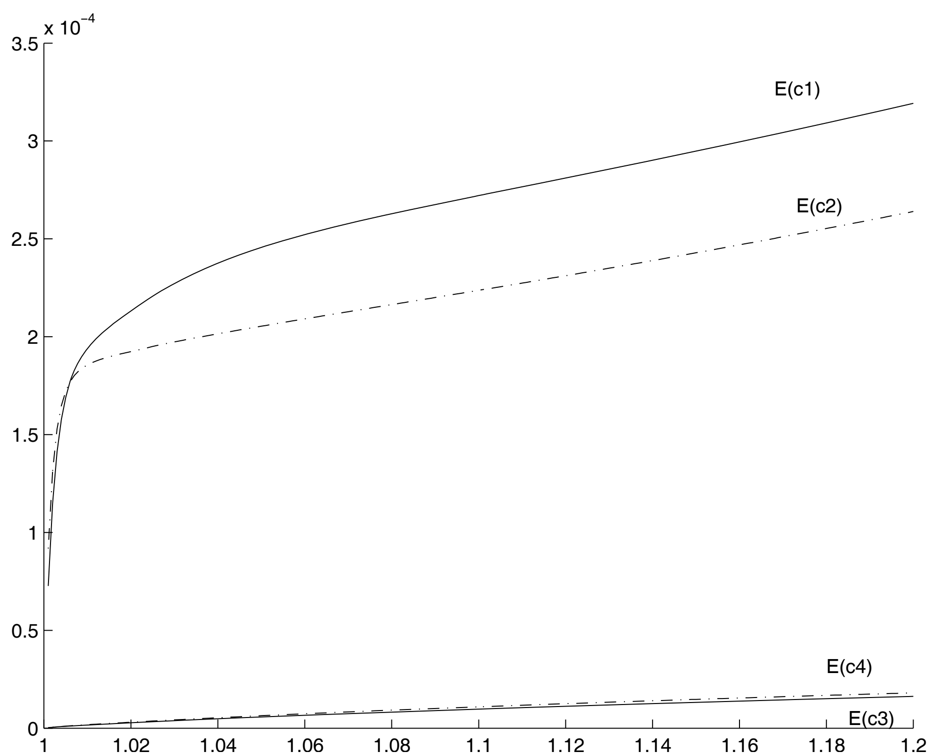

Figure 1 Global error for cases *c*1-*c*4.

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

Competing interests
The author declares that he has no competing interests.

Received: 6 September 2011 Accepted: 12 March 2012 Published: 12 March 2012

References

1. Varga, R: Matrix Iterative Analysis, 2nd edn.Berlin: Springer (2000)
2. Isaacson, E, Keller, HB: Analysis of Numerical Methods. New York: John Wiley & Sons (1966)
3. Ortega, JM, Rheinboldt, WC: Iterative Solution of Nonlinear Equations in Several Variables. New York: Academic Press (1970)
4. Vogel, CR, Oman, ME: Iterative methods for total variation denoising. SIAM Journal on Scientific Computing. **17**, 227–238 (1996). doi:10.1137/0917016
5. Dobson, DO, Vogel, CR: Convergence of an iterative method for total variation denoising. SIAM Journal on Numerical Analysis. **34**, 1779–1791 (1997). doi:10.1137/S003614299528701X
6. Chan, TF, Mulet, P: On the convergence of the lagged diffusivity fixed point method in total variation image restoration. SIAM Journal on Numerical Analysis. **36**, 354–367 (1999). doi:10.1137/S0036142997327075
7. Chang, Q, Chern, IL: Acceleration methods for total variation-based image denoising. SIAM Journal on Scientific Computing. **25**, 982–994 (2003). doi:10.1137/S106482750241534X
8. Shi, Y, Chang, Q, Xu, J: Convergence of fixed point iteration for deblurring and denoising problem. Applied Mathematics and Computation. **189**, 1178–1185 (2007). doi:10.1016/j.amc.2006.12.004
9. Weickert, J: Anisotropic Diffusion in Image Processing. Stuttgart: B. G. Teubner (1998)
10. Weickert, J, ter Haar Romeny, BM, Viergever, MA: Efficient and reliable schemes for nonlinear diffusion filtering. IEEE Transaction on Image Processing. **7**, 398–410 (1998). doi:10.1109/83.661190
11. Barash, D, Schlick, T, Israeli, M, Kimmel, R: Multiplicative operator splitting in nonlinear diffusion: from spatial splitting to multiple timesteps. Journal of Mathematical Image and Vision. **19**, 33–48 (2003). doi:10.1023/A:1024484920022
12. in 't Hout, KJ, Welfert, BD: Stability of ADI schemes applied to convection-diffusion equations with mixed derivative terms. Applied Numerical Mathematics. **57**, 19–35 (2007). doi:10.1016/j.apnum.2005.11.011
13. in 't Hout, KJ, Welfert, BD: Unconditional stability of second-order ADI schemes applied to multi-dimensional diffusion equations with mixed derivative terms. Applied Numerical Mathematics. **59**, 677–692 (2009). doi:10.1016/j.apnum.2008.03.016
14. in 't Hout, KJ, Foulon, S: ADI finite difference schemes for option pricing in the Heston model with correlation. International Journal of Numerical Analysis and Modeling. **7**, 303–320 (2010)
15. Karlsen, KH, Risebro, NH: An operator splitting method for nonlinear convection-diffusion equations. Numerische Mathematik. **77**, 365–382 (1997). doi:10.1007/s002110050291
16. Karlsen, KH, Lie, KA, Natvig, JR, Nordhaug, HF, Dahle, HK: Operator splitting methods for systems of convection-diffusion equations: nonlinear error mechanism and correction strategies. Journal of Computational Physics. **173**, 636–663 (2001). doi:10.1006/jcph.2001.6901
17. Holden, H, Karlsen, KH, Lie, KA, Risebro, NH: Splitting Methods for Partial Differential Equations with Rough Solutions. Zurich: European Mathematical Society Publishing House (2010). [Series of Lectures in Mathematics]
18. Geiser, J: Modified Jacobian Newton iterative method: theory and applications. Mathematical Problems in Engineering 24 (2009). Article ID 307298
19. Geiser, J: Consistency of iterative operator-splitting methods. theory and applications. Numerical Methods for Partial Differential Equations. **26**, 135–158 (2010). doi:10.1002/num.20422
20. Ruggiero, V, Galligani, E: A parallel algorithm for solving block tridiagonal linear systems. Computers Mathematics with Applications. **24**, 15–21 (1992). doi:10.1016/0898-1221(92)90003-Z
21. Evans, DJ: Group explicit iterative methods for solving large linear systems. International Journal of Computer Mathematics. **17**, 81–108 (1985). doi:10.1080/00207168508803452
22. Evans, DJ, Yousif, WS: The solution of two-point boundary value problems by the alternating group explicit (AGE) method. SIAM Journal on Scientific and Statistic Computing. **9**, 474–484 (1988). doi:10.1137/0909031
23. Tavakoli, R, Davami, P: 2D parallel and stable group explicit finite difference method for solution of diffusion equation. Applied Mathematics and Computation. **188**, 1184–1192 (2007). doi:10.1016/j.amc.2006.10.057
24. Ruggiero, V, Galligani, E: An iterative method for large sparse linear systems on a vector computer. Computers Mathematics with Applications. **20**, 25–28 (1990)
25. Galligani, E, Ruggiero, V: Analysis of splitting parallel methods for solving block tridiagonal linear systems. In: Ivannikov VP, Serebriakov VA (eds.) Proceedings of 2nd International Conference on Software for Multiprocessors and Supercomputers, Theory, Practice, Experience - SMS TPE'94: 19-23 September 1994; Moscow (Russia). pp. 406–416. Moscow: Russian Academy of Sciences (1994)
26. Galligani, E, Ruggiero, V: Implementation of splitting methods for solving block tridiagonal linear systems on transputers. In: Valero M, Gonzalez A, Los Alamitos CA (eds.) Proceedings of 3rd EuroMicro Workshop on Parallel and Distributed Processing - Euromicro PDP'95: 25-27 January 1995; Sanremo (Italy). pp. 409–415. IEEE Computer Society Press (1995)
27. Galligani, E, Ruggiero, V: The two-stage arithmetic mean method. Applied Mathematics and Computation. **85**, 245–264 (1997). doi:10.1016/S0096-3003(96)00139-7
28. Sulaiman, J, Othman, M, Hasan, MK: A new half-sweep arithmetic mean (HSAM) algorithm for two-point boundary value problems. Proceedings of the International Conference on Statistics and Mathematics and Its Application in the Development of Science and Technology: 4-6 October 2004; Bandung (Indonesia). 169–173 (2004)

29. Sulaiman, J, Othman, M, Hasan, MK: A new quarter sweep arithmetic mean (QSAM) method to solve diffusion equations. Chamchuri Journal of Mathematics. **1**, 93–103 (2009)
30. Sulaiman, J, Hasan, MK, Othman, M, Yaacob, Z: Quarter-sweep arithmetic mean iterative algorithm to solve fourth-order parabolic equations. Proceedings of the 2009 International Conference on Electrical Engineering and Informatics - ICEEI 2009, vol. 1: 5-7 August 2009; Bangi Selangor (Malaysia). pp. 194–198.Institute of Electrical and Electronics Engineers (IEEE) (2009)
31. Hasan, MK, Sulaiman, J, Karim, SAA, Othman, M: Development of some numerical methods applying complexity reduction approach for solving scientific problems. Journal of Applied Sciences. **11**, 1255–1260 (2011)

doi:10.1186/1687-1847-2012-30
Cite this article as: Galligani: **Lagged diffusivity method for the solution of nonlinear diffusion convection problems with finite differences.** *Advances in Difference Equations* 2012 **2012**:30.