SIAM J. NUMER. ANAL.
Vol. 48, No. 6, pp. 2247–2266

THE IMMERSED INTERFACE TECHNIQUE FOR PARABOLIC PROBLEMS WITH MIXED BOUNDARY CONDITIONS*

FRANÇOIS BOUCHON† AND GUNTHER H. PEICHL‡

Abstract. A finite difference scheme is presented for a parabolic problem with mixed boundary conditions. We use an immersed interface technique to discretize the Neumann condition, and we use the Shortley–Weller approximation for the Dirichlet condition. The proof of a discrete maximum principle is given as well as the proof of convergence of the scheme. This convergence is also validated on numerical examples.

Key words. finite differences, immersed interface, mixed boundary value problem

AMS subject classifications. 65M06, 35K20

DOI. 10.1137/09075384X

1. Introduction.

Immersed interface techniques have been recently developed for the numerical solution of partial differential equations in complex situations. This approach has been used, for example, to derive numerical schemes for elliptic problems with discontinuous coefficients (see [13], [15], [16], [26]) or boundary value problems in domains that do not fit the mesh (see [2], [9], [10], [11], [12], [18], [21]).

In [2], the authors consider the numerical solution of an elliptic problem with mixed boundary conditions and present a discretization of the Neumann boundary condition that does not require the knowledge of the tangential derivative of the Neumann data. This gives an alternative to the schemes proposed in [3] and [4], which were restricted to situations where such data are available.

In [9] and [10], the same problem with Dirichlet boundary conditions is investigated, in one- and two-space dimensions. The theoretical convergence analysis is achieved in one-space dimension in [10], and numerical tests validate this analysis in both one- and two-space dimensions. The approach in [9] differs slightly; an integration of the equation is done in a finite volume way on the cells located near the boundary. The convergence is observed numerically in two-space dimension; this work has been extended to three-space dimension more recently (see [21]) for both elliptic and parabolic cases with Dirichlet boundary conditions.

In [11] and [12], similar works are achieved for elliptic problems with mixed boundary conditions. The method is described for N-dimensional problems with $N \geq 3$, but the schemes developed here do not lead to linear systems described by M-matrices and then do not satisfy the discrete maximum principle.

In [18], the authors consider a parabolic problem on an evolving domain; their idea is to propose an immersed interface method with a fixed mesh, to avoid the remeshing process which would be necessary at each time-step if the mesh would have followed the domain. The numerical method is validated by numerical tests, but no proof of the convergence of the scheme is given.

*Received by the editors March 25, 2009; accepted for publication (in revised form) October 22, 2010; published electronically December 21, 2010.
http://www.siam.org/journals/sinum/48-6/75384.html

†Laboratoire de Mathématiques, Clermont Université, Université Blaise Pascal, BP10448, F-63000 Clermont-Ferrand, France and Laboratoire de Mathématiques, CNRS, UMR 6620, F-63177 Aubière, France (Francois.Bouchon@math.univ-bpclermont.fr).

‡Institute for Mathematics and Scientific Computing, Karl-Franzens-University Graz, A-8010 Graz, Austria (gunther.peichl@uni-graz.at).

Many efforts have been made to prove the convergence of discrete approximations of parabolic problems in the context of the finite elements method or finite differences. Although the finite element method naturally leads to convergence proofs in the L_2 norm, many authors have derived convergence estimates in the L_∞ norm. We also mention papers which consider semi-discrete approximations (see [14], [20], and more recently, [23] for problems with nonregular Neumann data) and one-dimensional problems (see [7]). In [19], an L_∞ optimal estimate is derived for a fully discrete approximation of a parabolic problem with Dirichlet boundary conditions. Note that dealing with a Neumann boundary condition is known to be more tedious, as mentioned in [25], even in one-space dimension.

In this work, we use an immersed interface technique to derive an implicit finite difference scheme in space and time for a parabolic problem with mixed boundary conditions. The space discretization is motivated by the scheme developed for the analogous problem in the elliptic context (see [2]), taking into account that the Laplacian of the solution is not available for parabolic problems. A proof of convergence is given, based on a maximum principle satisfied by the discrete operator. Since there are some positive off-diagonal entries in the matrix associated with the discrete operator, this matrix is not an L-matrix but a perturbation of an M-matrix. A technique similar to the one described in [1] is then used to show the monotonicity of the system. This step induces the condition $\delta t > Ch^2$. This condition is much less restrictive than the conventional conditions between the time-step and the mesh size (such as the CFL condition in the hyperbolic context, for example). Note that this condition is just opposite to $\delta t \ll h^2$ which is used to show the convergence for explicit schemes in the parabolic context.

This article is organized as follows: In section 2 we formulate the problem and derive the numerical scheme. Section 3 is dedicated to the proof of two theorems. Theorem 3.1 ensures the discrete maximum principle satisfied by the numerical scheme, and Theorem 3.2 establishes the convergence of the scheme. Section 4 is then devoted to numerical tests, including details about the fast solver used to run these tests. It is noted that the observed rate of convergence in space is better than the rate given by Theorem 3.2 for the first order Euler implicit scheme. Numerical tests for the Crank–Nicolson scheme are also given; the expected second order convergence in time is observed.

2. Problem formulation.

2.1. The continuous problem. We consider the following problem: Find the solution $u : \Omega \times (0, T) \to \mathbb{R}$ of the parabolic problem

$$\partial_t u(x,t) - \Delta u(x,t) = f(x,t), \qquad (x,t) \in \Omega \times (0,T), \tag{2.1}$$

$$u(x,0) = u_0(x), \qquad x \in \Omega, \tag{2.2}$$

$$u(x,t) = u_D(x,t), \qquad (x,t) \in \Gamma_D \times (0,T), \tag{2.3}$$

$$\partial_n u(x,t) = u_N(x,t), \qquad (x,t) \in \Gamma_N \times (0,T), \tag{2.4}$$

where the doubly connected bounded domain $\Omega \subset \mathbb{R}^2$ and the data f, u_0, u_D, and u_N are such that the solution is twice continuously differentiable in time and four times in space with bounded derivatives. Furthermore, we assume $\Gamma_D \cup \Gamma_N = \partial\Omega$, $\operatorname{dist}(\Gamma_D, \Gamma_N) > 0$, and that Ω satisfies the ball condition used in [2]:

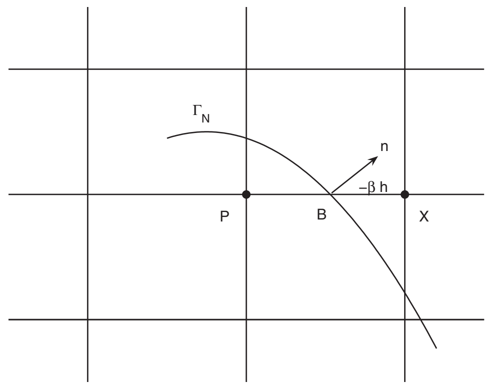

FIG. 1. *The generic configuration.*

BALL CONDITION 1. *There exists r_0 such that, for all $x \in \Gamma_N$, one can find points $\xi_x \in \Omega$ and $\eta_x \in \complement\bar{\Omega}$ such that the balls $B(\xi_x, r_0)$ and $B(\eta_x, r_0)$ satisfy*

$$B(\xi_x, r_0) \subset \Omega, \qquad B(\eta_x, r_0) \subset \complement\bar{\Omega},$$
$$\overline{B(\xi_x, r_0)} \cap \overline{B(\eta_x, r_0)} = \{x\}.$$

The condition for the interior ball ensures that Ω is not too thin, whereas the exterior ball condition guarantees that remote parts of Γ_N are not too close.

2.2. Derivation of the scheme. As in [2], we embed Ω into a rectangular domain which is discretized by a Cartesian grid $\{(ih, jh) : i, j \in \mathbb{N}_0\}$, where the grid size h is small enough. An interior/exterior node P is called an interior/exterior near boundary node if it has at least one exterior/interior neighbor. The neighbors of P are the four nodes adjacent to P on horizontal or vertical gridlines through P. We denote by Ω_h the set of interior nodes and by Γ_h the set of exterior near boundary nodes. In $\Gamma_{h,N}$ we collect the nodes in Γ_h which are close to the Neumann boundary; $\Gamma_{h,D}$ denotes the set of interior nodes which are close to the Dirichlet boundary. At nodes in $\Gamma_{h,D}$ we use the Shortley–Weller approximation for Δu (see [17], [22]). At all other interior nodes we discretize the Laplacian by standard central differences. At interior nodes near the Neumann boundary the stencil thus involves at least one exterior node X. We now describe how we derive an equation for such an exterior point X from the Neumann condition. First, we define the point B as the intersection point of Γ_N and the gridlines which is the closest to X. Let P be the neighbor point of X such that the gridline (PX) crosses Γ_N in B. By the invariance of the Laplacian with respect to rotations and symmetries, we may assume without loss of generality the configuration shown in Figure 1, for which the point B is on the left of X, and the normal vector $n = (n_1, n_2)$ to Γ_N on B is such that $n_2 \geq 0$. In the local coordinate system centered in X, the coordinates of P and B are $(-h, 0)$, $(-\beta h, 0)$, respectively, with $0 \leq \beta < 1$, since $P \in \Omega$ and $X \notin \Omega$. The case $\beta = 0$ corresponds to the case where the exterior point X is on the boundary Γ_N.

It is shown in [2] that $n_2 < 6n_1/5$ if h is small enough. We now select four points $P_\ell = (x_\ell, y_\ell)$, $\ell = 1, \ldots, 4$, according to Table 1 and define the four real numbers α_ℓ,

TABLE 1
Choice of the points for the discretization of the Neumann condition.

case			P_1	P_2	P_3	P_4
1	$\frac{1}{5} \le \frac{n_2}{n_1} \le \frac{6}{5}$		$(-h, 0)$	$(-11h, 0)$	$(-h, -5h)$	$(0, -5h)$
2	$\frac{n_2}{n_1} < \frac{1}{5}$,	$\frac{3}{4} < \beta < 1$	$(-h, 0)$	$(-3h, 0)$	$(-2h, h)$	$(-h, -h)$
3	$\frac{n_2}{n_1} < \frac{1}{5}$,	$0 \le \beta \le \frac{3}{4}$	$(0, h)$	$(0, -h)$	$(-2h, 0)$	$(-h, -h)$

$\ell = 1, \ldots, 4$, solution of the linear system:

$$\begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ x_1^2 - y_1^2 & x_2^2 - y_2^2 & x_3^2 - y_3^2 & x_4^2 - y_4^2 \\ y_1 & y_2 & y_3 & y_4 \\ x_1 y_1 & x_2 y_2 & x_3 y_3 & x_4 y_4 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{pmatrix} = \begin{pmatrix} -n_1 \\ 2\beta n_1 h \\ -n_2 \\ \beta n_2 h \end{pmatrix}. \tag{2.5}$$

It is shown in [2] that the matrix in (2.5) is nonsingular and that the solution $(\alpha_\ell)_{1 \le \ell \le 4}$ satisfies

$$\sum_{\ell=1}^{4} \alpha_\ell (u(X) - u(P_\ell)) = \frac{\partial u}{\partial n}(B) + k \Delta u(P) + \mathcal{O}(h^2). \tag{2.6}$$

It is also shown in [2] that $\alpha_\ell \ge 0$, $\alpha_\ell = \mathcal{O}(\frac{1}{h})$, $\ell = 1, \ldots 4$, $k \le 0$, and $k = \mathcal{O}(h)$. This discretization of the elliptic problem with mixed boundary conditions resulted in [2] in a linear system

$$AU = F$$

with

$$A = \begin{pmatrix} A_1 & A_2 \\ A_3 & A_4 \end{pmatrix}. \tag{2.7}$$

Above $A_1 \in \mathbb{R}^{N_i \times N_i}$, $A_4 \in \mathbb{R}^{N_e \times N_e}$, A_2 and $A_3^T \in \mathbb{R}^{N_i \times N_e}$, where N_i and N_e denote the number of nodes in Ω_h, respectively, $\Gamma_{h,N}$. The blocks A_1 and A_3 correspond to unknowns u_{ij} for which $x_{ij} \in \Omega_h$. A_2 and A_4 refer to unknowns u_{ij} for which $x_{ij} \in \Gamma_{h,N}$. The rows in A_1 and A_2 describe the discretization of the Laplacian; the rows of A_3 and A_4 describe the discretization of the Neumann boundary condition (2.6). Therefore the entries of A_1, A_2 are $\mathcal{O}(h^{-2})$, and the entries of A_3, A_4 are $\mathcal{O}(h^{-1})$. It was shown in [2] that A is an irreducible diagonally dominant M-matrix.

For the discretization of the parabolic problem (2.1)–(2.4) we use backward differences in time; the Laplacian is discretized as described above. However, special attention must be paid to the boundary condition (2.6) since, unlike in the elliptic problem, $-\Delta u(P)$ cannot be replaced by $f(P)$. Therefore we approximate $-\Delta u(P)$ by a central difference approximation which results in

$$\sum_{\ell=1}^{4} \alpha_\ell (u(X) - u(P_\ell)) + k \left(\frac{4u(P) - u(N) - u(S) - u(E) - u(W)}{h^2} \right) = \frac{\partial u}{\partial n}(B) + \mathcal{O}(h^2), \tag{2.8}$$

where N, S, E, and W denote the neighbors of P (one of them is X).

Note that the spatial discretization cannot be described by the matrix (2.7) anymore but by a matrix

$$B = \begin{pmatrix} B_1 & B_2 \\ B_3 & B_4 \end{pmatrix} = \begin{pmatrix} I & \mathbb{O} \\ M & I \end{pmatrix} A, \tag{2.9}$$

where M has at most one nonvanishing coefficient on each row which is given by the respective value of k in (2.8). As a consequence, $M \le 0$ and $|||M|||_\infty = \mathcal{O}(h)$.

Remark 1. We also note that $\left(\begin{smallmatrix} I & \mathbb{O} \\ M & I \end{smallmatrix}\right)$ is monotone since

$$\begin{pmatrix} I & \mathbb{O} \\ M & I \end{pmatrix}^{-1} = \begin{pmatrix} I & \mathbb{O} \\ -M & I \end{pmatrix} \ge 0.$$

Hence B, being the product of two monotone matrices by (2.9), is also monotone. The discretization of the parabolic problem (2.1)–(2.4) with the modified discretized Neumann boundary condition (2.8) is represented by the linear system

$$(B + \delta t^{-1} D)\mathbf{u}^k = \delta t^{-1} D \mathbf{u}^{k-1} + \mathbf{f}^k, \tag{2.10}$$

where

$$D = \begin{pmatrix} I & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{pmatrix},$$

and $\mathbf{u}^k$ collects the approximation of $u(x_{ij}, t_k)$ with $t_k = k\delta t$, $k = 0, \dots, N$, and $N\delta t = T$. The top block of components of the vector $\mathbf{f}^k$ in (2.10) are the values of $f(x_{ij}, t_k)$ for $x_{ij} \in \Omega_h$, with some modifications depending on u_D for the points in $\Gamma_{h,D}$ according to the Shortley–Weller approximation. The bottom block of components is given by the values of $u_N(B, t_k)$ with B as in (2.8).

3. Convergence analysis.

3.1. The truncation error. It was shown in [2] that the truncation error for the elliptic operator satisfies

$$\varepsilon_{i,j} = \begin{cases} \mathcal{O}(h^2) & \text{for} \quad x_{i,j} \in (\Omega_h \setminus \Gamma_{h,D}) \cup \Gamma_{h,N}, \\ \mathcal{O}(h) & \text{for} \quad x_{i,j} \in \Gamma_{h,D}. \end{cases}$$

Hence, for the parabolic problem the corresponding truncation error is given by

$$\varepsilon^k_{i,j} = \begin{cases} \mathcal{O}(h^2 + \delta t) & \text{for} \quad x_{i,j} \in (\Omega_h \setminus \Gamma_{h,D}), \\ \mathcal{O}(h + \delta t) & \text{for} \quad x_{i,j} \in \Gamma_{h,D}, \\ \mathcal{O}(h^2) & \text{for} \quad x_{i,j} \in \Gamma_{h,N}. \end{cases} \tag{3.1}$$

The local error $\mathbf{e}^k$ and the truncation error ε^k are related by

$$(B + \delta t^{-1} D)\mathbf{e}^k - \delta t^{-1} D \mathbf{e}^{k-1} = \varepsilon^k. \tag{3.2}$$

3.2. Monotonicity of the discrete operator. We emphasise that the matrix B in (2.9) is not an M-matrix anymore, since there are positive entries in the off-diagonal block B_3. Nevertheless, we will show in this section that $B + \delta t^{-1} D$ is monotone. The proof adapts some ideas and techniques from [1]. We start with a result which is of interest on its own and which corresponds to Lemma 3.2 in [1].

LEMMA 3.1. *Let $(z_k)_{k\geq 0}$ be a sequence of nonnegative real numbers such that $z_k \leq \gamma$ for some $\gamma > 0$ and $z_0 = 0$.*

If

$$z_k \leq \alpha \left(z_{k-1} + \beta \max_{0\leq i\leq k+1} z_i \right) \tag{3.3}$$

holds for some $\alpha \geq 1$, $\beta > 0$ for all $k \in \mathbb{N}$, then z_k satisfies the recursion

$$z_k \leq \alpha z_{k-1} + \gamma\beta\alpha^{k+2p} \frac{\beta^p}{p!} (k-1+2p)^p \tag{3.4}$$

for all $k \in \mathbb{N}$ and $p \in \mathbb{N}_0$.

We emphasize that due to the "advanced" term $\max_{0\leq i\leq k+1} z_i$, (3.3) is not a recursion. The following proof is taken from [1].

Proof of Lemma 3.1. Since $\alpha \geq 1$, (3.4) is true for $p = 0$ and $k \in \mathbb{N}$. Let us assume that (3.4) holds for some integer p. Since the recursion $z_k \leq \alpha z_{k-1} + \delta_k$ implies the estimate $z_k \leq \alpha^k z_0 + \sum_{\ell=1}^{k} \alpha^{k-\ell}\delta_\ell = \sum_{\ell=1}^{k} \alpha^{k-\ell}\delta_\ell$ (using $z_0 = 0$), the induction hypothesis implies

$$z_k \leq \gamma\alpha^{k+2p} \frac{\beta^{p+1}}{p!} \sum_{\ell=1}^{k} (\ell - 1 + 2p)^p \,.$$

The sum is estimated by

$$\sum_{\ell=1}^{k} (\ell-1+2p)^p \leq \int_p^{k+2p} x^p dx \leq \frac{(k+2p)^{p+1}}{p+1},$$

which entails the estimate

$$z_k \leq \gamma\alpha^{k+2p} \frac{\beta^{p+1}}{(p+1)!} (k+2p)^{p+1} \,.$$

Inserting this estimate into (3.3), one obtains

$$z_k \leq \alpha z_{k-1} + \beta\gamma\alpha^{k+2(p+1)} \frac{\beta^{p+1}}{(p+1)!} (k-1+2(1+p))^{p+1} \,.$$

This completes the induction step. ◻

Using (2.9), one finds

$$B + \delta t^{-1} D = \left(\begin{array}{c|c} I & \mathbb{O} \\ \hline M & I \end{array} \right) \left[\left(\begin{array}{c|c} A_1 & A_2 \\ \hline A_3 & A_4 \end{array} \right) + \delta t^{-1} \left(\begin{array}{c|c} I & \mathbb{O} \\ \hline -M & \mathbb{O} \end{array} \right) \right];$$

hence by Remark 1 it suffices to show that $A + \delta t^{-1} \left(\begin{array}{c|c} I & \mathbb{O} \\ \hline -M & \mathbb{O} \end{array} \right)$ is monotone. It turns out to be advantageous to balance the size of the entries in A_1, A_2 and A_3, A_4. Therefore, one considers the matrix

$$T = \left(\begin{array}{c|c} hI & \mathbb{O} \\ \hline \mathbb{O} & I \end{array} \right) \left(\begin{array}{c|c} A_1 & A_2 \\ \hline A_3 & A_4 \end{array} \right) + \delta t^{-1} \left(\begin{array}{c|c} hI & \mathbb{O} \\ \hline \mathbb{O} & I \end{array} \right) \left(\begin{array}{c|c} I & \mathbb{O} \\ \hline -M & \mathbb{O} \end{array} \right) = Q + R. \tag{3.5}$$

Note that T is a perturbation of the monotone matrix Q.

Let us recall some tools developed in [1].

DEFINITION 3.1. *An index i is* **directly connected** *to an index $j \neq i$ if there is a constant σ independent of i and h such that $|a_{ii}| < \sigma |a_{ij}|$ (the proof of Lemma 3.2 given in the appendix shows that σ can be chosen as 21). Analogously, we say that a point $Z \in \Omega_h \cup \Gamma_{h,N}$ is directly connected to a point $Y \in \Omega_h \cup \Gamma_{h,N}$ if the index i of Z is directly connected to the index j of Y.*

Remark 2. Let $Z \in \Omega_h \setminus \Gamma_{h,D}$ and $Y \in \Omega_h \cup \Gamma_{h,N}$ be two neighboring points. Then Z is directly connected to Y since $|a_{ii}| = 4|a_{ij}|$ holds by the discretization of the Laplacian for the indices i of Z and j of Y.

Given a fixed index i_0 we construct a sequence of subsets of indices by the following recursion:

$$\begin{aligned} &I_0 = \{i_0\}, \\ &I_k = \{j \notin \cup_{\ell<k} I_\ell, \exists i \in I_{k-1} \text{ which is directly connected to } j\}, \qquad k \in \mathbb{N}, \end{aligned} \tag{3.6}$$

I_k is the set of indices j that can be reached from i_0 via a chain $i_0, i_1, \ldots, i_k = j$ such that $i_{\ell-1}$ is directly connected to i_ℓ for all $\ell \in \{1, \ldots, k\}$, and k is the minimum length of such a chain.

Remark 3. As a consequence of this definition we have $I_j \cap I_k = \emptyset$ for $j \neq k$ and $I_k = \emptyset$ for k larger than the dimension of A.

The proof of the following technical result requires some details concerning the approximation of the elliptic operator given in [2] and is therefore deferred to the appendix.

LEMMA 3.2.
- (i) *Every node $X \in \Gamma_{h,N}$ can be directly connected to its interior neighbor $P \in \Omega_h$ by a chain of length not exceeding 2.*
- (ii) *The diagonal elements of Q can be bounded by below by*

$$Q_{ii} > Ch^{-1}$$

for some positive constant C.

THEOREM 3.1. *Assume that $h^2/\delta t$ is bounded by a constant ρ sufficiently small. Then, the matrix $B + \delta t^{-1} D$ is monotone for h and δt sufficiently small.*

Proof of Theorem 3.1. By the preceding discussion, it suffices to show that the matrix T defined in (3.5) is monotone. Let x be a vector such that $Tx \geq 0$; we want to show that $x \geq 0$. We consider an index i_0 such that $x_{i_0} = \min_i x_i$. Let us assume $x_{i_0} < 0$. We split T as suggested in (3.5), $T = \begin{pmatrix} T_1 \\ T_2 \end{pmatrix}$. Assume that i_0 corresponds to a row of T_1. Since $(Tx)_{i_0} \geq 0$,

$$\sum_j T_{i_0 j} x_j = \sum_j Q_{i_0 j} x_j + R_{i_0 i_0} x_{i_0} \geq 0$$

must hold, which entails

$$\sum_{j \neq i_0} Q_{i_0 j}(x_j - x_{i_0}) + R_{i_0 i_0} x_{i_0} + x_{i_0} \sum_j Q_{i_0 j} \geq 0.$$

This gives a contradiction since $x_j - x_{i_0} \geq 0$ for all j, $Q_{i_0,j} \leq 0$ for $j \neq i_0$ since i_0 corresponds to a row in the top block and A is an M-matrix, $R_{i_0 i_0} = \frac{h}{\delta t} > 0$, $x_{i_0} < 0$ and $\sum_j Q_{i_0 j} \geq 0$.

Hence i_0 corresponds to a row of the bottom block T_2.

For the index i_0 we construct the set of indices I_k by (3.6) and define for $k \in \mathbb{N}_0$ the real numbers

$$D_k = \begin{cases} \max_{j \in I_k} x_j - x_{i_0}, & I_k \neq \emptyset, \\ 0, & I_k = \emptyset. \end{cases} \tag{3.7}$$

Note:

- $D_0 = 0$,
- $0 \leq D_k \leq 2|x|_\infty$ for all k,
- and $D_k = 0$ for k larger than the dimension of A.

For any $j \in I_k$, there exists an index $i \in I_{k-1}$ such that i is directly connected to j, which implies

$$|Q_{ii}| < \sigma |Q_{ij}|. \tag{3.8}$$

From

$$0 \leq (Tx)_i = \sum_\ell Q_{i\ell}(x_\ell - x_{i_0}) + \sum_\ell R_{i\ell}(x_\ell - x_{i_0}) + x_{i_0} \sum_\ell T_{i\ell}$$

and $x_{i_0} < 0$ and $\sum_\ell T_{i\ell} \geq 0$, we deduce

$$-\sum_{\ell \neq i} Q_{i\ell}(x_\ell - x_{i_0}) \leq Q_{ii}(x_i - x_{i_0}) + \sum_\ell R_{i\ell}(x_\ell - x_{i_0}).$$

On the left-hand side all terms in the sum are nonnegative, which gives

$$|Q_{ij}|(x_j - x_{i_0}) \leq Q_{ii}\left((x_i - x_{i_0}) + \frac{1}{Q_{ii}} \sum_\ell R_{i\ell}(x_\ell - x_{i_0})\right).$$

Using (3.8) and (3.7), we obtain

$$\begin{aligned} x_j - x_{i_0} &\leq \sigma \left((x_i - x_{i_0}) + \frac{1}{Q_{ii}} \sum_\ell R_{i\ell}(x_\ell - x_{i_0})\right) \\ &\leq \sigma \left(D_{k-1} + \frac{1}{Q_{ii}} \sum_\ell R_{i\ell}(x_\ell - x_{i_0})\right). \end{aligned}$$

If the index i refers to a row in the top block, then $R_{ik} = 0$ for all $k \neq i$ and the sum in the right-hand side reduces to $R_{ii}(x_i - x_{i_0})$, which is bounded by $R_{ii}D_{k-1}$. Then we obtain the estimate

$$x_j - x_{i_0} \leq \sigma \left(1 + \frac{R_{ii}}{Q_{ii}}\right) D_{k-1}. \tag{3.9}$$

Let us now consider the case where the index i refers to a row in the bottom block. Note that there exists at most one index ℓ such that $R_{i\ell} \neq 0$. The index i corresponds to a point $X \in \Gamma_{h,N}$. By Lemma 3.2, i can be connected to ℓ by a chain the length of which is bounded by 2. Since $i \in I_{k-1}$, we deduce $\ell \in I_q$ with $q \leq k+1$.

This results in

$$x_j - x_{i_0} \leq \sigma \left(D_{k-1} + \frac{|||R|||_\infty}{\min_i Q_{ii}} \max_{0 \leq \ell \leq k+1} D_\ell\right), \tag{3.10}$$

which holds also if the index i refers to a row in the top block due to (3.9). Since this estimate (3.10) holds for all $j \in I_k$, one obtains the relation

$$D_k \leq \sigma \left(D_{k-1} + \nu \max_{0 \leq \ell \leq k+1} D_\ell \right), \qquad k \in \mathbb{N},$$

with

$$\nu = \frac{|||R|||_\infty}{\min_i Q_{ii}}.$$

By Lemma 3.1, D_k satisfies the recursion

$$D_k \leq \sigma D_{k-1} + 2|x|_\infty \nu \sigma^{k+2p} \frac{\nu^p}{p!}(k-1+2p)^p, \qquad k \in \mathbb{N}.$$

In particular, we have

$$\begin{aligned} D_1 &\leq 2|x|_\infty \nu\sigma \left(2\sigma^2\nu\right)^p \frac{p^p}{p!} \\ &\leq 2|x|_\infty \nu\sigma \left(2\sigma^2\nu e\right)^p, \qquad p \in \mathbb{N}. \end{aligned} \tag{3.11}$$

Since by Lemma 3.2, $\min Q_{ii} > Ch^{-1}$, we obtain

$$\nu \leq C\frac{h}{\delta t} h \leq C\rho < \frac{1}{2\sigma^2 e}$$

if ρ has been chosen small enough, which by passing p to infinity in (3.11) implies

$$D_1 = 0.$$

This shows that $x_{i_1} = x_{i_0}$ for all indices i_1 such that i_0 is directly connected to i_1. By the first part of the proof we conclude that i_1 necessarily refers to a node in $\Gamma_{h,N}$. Lemma 3.2 on the other hand ensures that any point in $\Gamma_{h,N}$ is directly connected to at least one point in Ω_h. This contradiction shows that $x_{i_0} = \min_i x_i \geq 0$ and hence that T is monotone. ∎

Remark 4. We mention that Theorem 3.1 gives the proof of the following discrete maximum principle: If the data f, u_0, u_D, and u_N are nonnegative, then the approximate solution $\mathbf{u}^k$ given by (2.10) is nonnegative for all integers $k \geq 0$. This is a discrete version of the maximum principle satisfied by the solution u of the continuous problem (2.1)–(2.4).

3.3. Convergence. In this section, we prove the convergence of the scheme. The proof is based on the following two lemmas.

LEMMA 3.3. *Let w be a vector in $\mathbb{R}^{N_i+N_e}$ such that*

$$\begin{cases} w_i = 1 & \text{if } i \text{ is an index of a point in } \Omega_h, \\ w_i = 0 & \text{if } i \text{ is an index of a point in } \Gamma_{h,N}. \end{cases}$$

Then, the solution v of $(B + \delta t^{-1} D)v = w$ satisfies

$$|v|_\infty \leq \delta t$$

if $h^2/\delta t$ is bounded.

Proof. The proof of Lemma 3.3 is deduced from the evaluation of $(B+\delta t^{-1}D)\mathbf{1}$, where $\mathbf{1}=(1,\dots,1)\in\mathbb{R}^{N_i+N_e}$. This vector is nonnegative and has the following components:

$\delta t^{-1}+\mathcal{O}(h^{-2})$ if i is an index of a point in $\Gamma_{h,D}$,
δt^{-1} if i is an index of a point in $\Omega_h\setminus\Gamma_{h,D}$,
0 if i is an index of a point in $\Gamma_{h,N}$.

Then,

$$(B+\delta t^{-1}D)\mathbf{1}\geq \delta t^{-1}w,$$

where w is the vector in Lemma 3.3. The monotonicity of $(B+\delta t^{-1}D)$ completes the proof of Lemma 3.3. □

LEMMA 3.4. *Let w be a vector in $\mathbb{R}^{N_i+N_e}$ such that*

$$\begin{cases} w_i=0 & \text{if } i \text{ is an index of a point in } \Omega_h,\\ w_i=1 & \text{if } i \text{ is an index of a point in } \Gamma_{h,N}.\end{cases}$$

Then, the solution v of $(B+\delta t^{-1}D)v=w$ satisfies

$$|v|_\infty\leq C\frac{\delta t}{h}$$

if $h^2/\delta t$ is bounded, with a constant C independent of h and δt.

Proof. To prove Lemma 3.4, we first observe that:

$$(B+\delta t^{-1}D)^{-1}=\left[A+\delta t^{-1}\left(\begin{array}{c|c} I & \mathbb{O}\\ \hline -M & \mathbb{O}\end{array}\right)\right]^{-1}\left(\begin{array}{c|c} I & \mathbb{O}\\ \hline -M & I\end{array}\right).$$

Hence, the vectors v and w in Lemma 3.4 satisfy

$$v=(B+\delta t^{-1}D)^{-1}w=\left[A+\delta t^{-1}\left(\begin{array}{c|c} I & \mathbb{O}\\ \hline -M & \mathbb{O}\end{array}\right)\right]^{-1}w \tag{3.12}$$

since $\left(\begin{array}{c|c} I & \mathbb{O}\\ \hline -M & I\end{array}\right)w=w$.

Moreover,

$$A+\delta t^{-1}\left(\begin{array}{c|c} I & \mathbb{O}\\ \hline \mathbb{O} & \mathbb{O}\end{array}\right)\leq A+\delta t^{-1}\left(\begin{array}{c|c} I & \mathbb{O}\\ \hline -M & \mathbb{O}\end{array}\right), \tag{3.13}$$

and we know from Theorem 3.1 that both matrices in this inequality are monotone. We thus deduce

$$\left[A+\delta t^{-1}\left(\begin{array}{c|c} I & \mathbb{O}\\ \hline -M & \mathbb{O}\end{array}\right)\right]^{-1}w\leq\left[A+\delta t^{-1}\left(\begin{array}{c|c} I & \mathbb{O}\\ \hline \mathbb{O} & \mathbb{O}\end{array}\right)\right]^{-1}w. \tag{3.14}$$

Let $\tilde v=(\tilde v_1,\tilde v_2)=[A+\delta t^{-1}(\begin{array}{c|c} I & \mathbb{O}\\ \hline \mathbb{O} & \mathbb{O}\end{array})]^{-1}w\in\mathbb{R}^{N_i+N_e}$ with $\tilde v_1\in\mathbb{R}^{N_i}$ and $\tilde v_2\in\mathbb{R}^{N_e}$. Since the first block of lines of the linear system satisfied by $\tilde v$ is strictly diagonally dominant, $\|\tilde v_1\|_\infty\leq\|\tilde v_2\|_\infty$ follows, which implies

$$0\leq\tilde v\leq\|\tilde v_2\|_\infty\mathbf{1}. \tag{3.15}$$

Moreover, $\tilde v_2$ satisfies

$$(A_4-A_3(A_1+\delta t^{-1}I)^{-1}A_2)\tilde v_2=\mathbf{1}_{N_e}=(1,\dots,1)\in\mathbb{R}^{N_e}.$$

We define the function $\varphi : \mathbb{R}_+ \to \mathbb{R}^{N_e}$ by

$$\varphi(\lambda) = (A_4 - A_3(A_1 + \lambda I)^{-1} A_2)\mathbf{1}_{N_e}.$$

Let $\lambda \in \mathbb{R}_+$ and $z \in \mathbb{R}^{N_i}$ be the solution vector of

$$(A_1 + \lambda I)z = -A_2 \mathbf{1}_{N_e}. \tag{3.16}$$

Since the top block in the matrix A in (2.7) corresponds to the discretized version of the Laplace equation for the interior points, we first observe that

$$0 \le z \le \mathbf{1}_{N_i}. \tag{3.17}$$

Moreover, for an index $k \in \Omega_h \setminus \Gamma_{h,D}$, we deduce from (3.16) and (3.17)

$$\left(\frac{4}{h^2} + \lambda\right) z_k \le \frac{4}{h^2},$$

and since an even sharper estimate can be obtained for indices $k \in \Gamma_{h,D}$,

$$0 \le z \le \frac{4}{4 + \lambda h^2} \mathbf{1}_{N_i} \tag{3.18}$$

follows. In view of

$$\varphi(\lambda) = A_4 \mathbf{1}_{N_e} + A_3 z, \tag{3.19}$$

(3.18) and $A_3 \le 0$ yield the estimate

$$\varphi(\lambda) \ge A_4 \mathbf{1}_{N_e} + \frac{4}{4 + \lambda h^2} A_3 \mathbf{1}_{N_i}. \tag{3.20}$$

Observing that the sum of the entries of each row in the bottom block of A in (2.7) vanishes, which is equivalent to

$$A_3 \mathbf{1}_{N_i} + A_4 \mathbf{1}_{N_e} = 0 \in \mathbb{R}^{N_e}, \tag{3.21}$$

then leads to

$$\varphi(\lambda) \ge \frac{\lambda h^2}{4 + \lambda h^2} A_4 \mathbf{1}_{N_e}.$$

Since it is shown in the appendix that $A_4 \mathbf{1} \ge \frac{C}{h} \mathbf{1}_{N_e}$ for some positive constant C, we eventually obtain

$$\varphi(\lambda) \ge \frac{\lambda h^2}{4 + \lambda h^2} \frac{C}{h} \mathbf{1}_{N_e}. \tag{3.22}$$

The matrix $(A_4 - A_3 \left(A_1 + \delta t^{-1} I)^{-1} A_2\right)^{-1}$ is the bottom right block of the nonnegative matrix $\left(\begin{array}{c|c} A_1 + \delta t^{-1} I & A_2 \\ A_3 & A_4 \end{array}\right)^{-1}$ and hence nonnegative by itself. Multiplying the estimate (3.22) for $\lambda = \delta t^{-1}$ by this matrix leads to

$$\tilde{v}_2 \le C \left(\frac{\delta t}{h} + h\right) \mathbf{1}_{N_e},$$

which completes the proof of Lemma 3.4 using (3.12), (3.14), (3.15) and the boundedness of $h^2/\delta t$. ◻

We are now ready to state and prove the convergence theorem.

THEOREM 3.2. *Assume $h^2/\delta t$ bounded. Then the local error satisfies the bound*

$$|e^k|_\infty = \mathcal{O}(\delta t + h),$$

which shows the convergence of the scheme.

Proof of Theorem 3.2. From (3.2) and Lemmas 3.3 and 3.4, we now show by induction on k the following inequality:

$$|e^k|_\infty \le Ck(\delta t^2 + h\delta t), \qquad k = 0, \dots, N. \tag{3.23}$$

This is true for $k = 0$; let us consider that this inequality holds for $k - 1$.

Then,

$$|e^k|_\infty \le \left|(B + \delta t^{-1} D)^{-1} \delta t^{-1} D e^{k-1}\right|_\infty + \left|(B + \delta t^{-1} D)^{-1} \varepsilon^k\right|_\infty . \tag{3.24}$$

The last term in the right-hand side of (3.24) is bounded by $\mathcal{O}(\delta t(\delta t + h) + \delta t(\delta t + h^2) + h^2 \frac{\delta t}{h}) = \mathcal{O}(\delta t^2 + h\delta t)$ due to (3.1) and Lemmas 3.3 and 3.4.

The vector $\delta t^{-1} D e^{k-1}$ has vanishing entries in the bottom block of components and is bounded by $|\delta t^{-1} D e^{k-1}|_\infty \le |\delta t^{-1} e^{k-1}|_\infty \le C(k-1)(\delta t + h)$. Using Lemma 3.3 again, the first term in the right-hand side of (3.24) is bounded by $C(k-1)(\delta t^2 + h\delta t)$, which completes the induction proof of (3.23).

The proof of Theorem 3.2 is then completed by taking $k = N$ with $N\delta t = T$ in (3.23). ◻

4. Numerical tests.

4.1. Computational considerations. In this section we describe the fast solver used to run the simulation. This method is inspired by [5] and [6]. The domain Ω is embedded into a square, which without loss of generality we assume to be the unit square $(0,1) \times (0,1)$. We consider the linear system described by (2.10), and complete this linear system with the following set of equations:

$$\frac{u_{i,j}^k}{\delta t} + \frac{4u_{i,j}^k - u_{i-1,j}^k - u_{i+1,j}^k - u_{i,j-1}^k - u_{i,j+1}^k}{h^2} = 0 \tag{4.1}$$

for points $x_{i,j} \notin \Omega_h \cup \Gamma_{h,N}$ *(where $h = \frac{1}{n+1}$ and the unknowns $u_{\ell,m}$ are removed whenever $\ell = 0$ or $n+1$ or $m = 0$ or $n+1$).*

Equations (2.10) and (4.1) can be described by a linear system of n^2 equations with n^2 unknowns, where the unknowns inside/outside $\Omega_h \cup \Gamma_{h,N}$ are decoupled (thus, the solution inside the domain $\Omega_h \cup \Gamma_{h,N}$ is the original solution of (2.10)). Let us consider the matrix $G \in \mathcal{M}_{n^2,n^2}(\mathbb{R})$ of the discretization of the parabolic operator on the square $(0,1) \times (0,1)$ with homogeneous Dirichlet boundary conditions on $\partial\left((0,1) \times (0,1)\right)$. If the matrix formulation of this linear system (2.10) and (4.1) is

$$Mx = z, \tag{4.2}$$

then M and G are identical, except on the rows corresponding to the boundary condition on $\partial\Omega$. We denote n_1 the number of rows where $M - G$ has nonvanishing coefficients.

Remark 5. Since Γ is a one-dimensional smooth curve in $\mathbb{R}^2$, we have $n_1 = \mathcal{O}(n)$. Moreover, the number of nonvanishing entries on each of the rows of $M - G$ is bounded.

We now use these observations to propose a fast solver for (4.2). First, note that (4.2) is equivalent with the following system:

$$Gx = \tilde{z}, \tag{4.3}$$

$$\tilde{z} = z - Py, \tag{4.4}$$

$$Py = (M - G)x, \tag{4.5}$$

where $y \in \mathbb{R}^{n_1}$ collects the possible nonvanishing values of $(M - G)x$ and P is a matrix of dimensions $n^2 \times n_1$ with one nonvanishing coefficient (equal to one) on each column, which then satisfies the following properties:

$$P^t P = I_{n_1} \tag{4.6}$$

and

$$PP^t(M - G) = M - G. \tag{4.7}$$

Now, (4.5) reads

$$y - P^t(M - G)x = 0. \tag{4.8}$$

Inserting (4.3) and (4.4) into (4.8) leads to

$$\left(I_{n_1} + P^t(M - G)G^{-1}P\right) y = P^t(M - G)G^{-1}z. \tag{4.9}$$

THEOREM 4.1. *The matrix* $(I_{n_1} + P^t(M - G)G^{-1}P)$ *is a nonsingular matrix of dimension* n_1.

Proof of Theorem 4.1. Let $y \in \mathbb{R}^{n_1}$ such that

$$\left(I_{n_1} + P^t(M - G)G^{-1}P\right) y = 0 \in \mathbb{R}^{n_1}. \tag{4.10}$$

We want to prove that $y = 0$. Multiplying (4.10) by P and using (4.7) gives

$$\left(I_{n^2} + (M - G)G^{-1}\right) Py = 0 \in \mathbb{R}^{n^2},$$

and then

$$\left(MG^{-1}\right) Py = 0 \in \mathbb{R}^{n^2}. \tag{4.11}$$

Since M and G^{-1} are nonsingular, (4.11) shows that $Py = 0$ and then $y = 0$ by (4.6), which completes the proof of Theorem 4.1. ∎

We now explain how one can efficiently compute x using (4.9), (4.4), and (4.3). First, we need to compute $I_{n_1} + P^t(M - G)G^{-1}P$. Let D_1 stand for the matrix

$$D_1 = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & \ddots & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & -1 \\ 0 & \cdots & 0 & -1 & 2 \end{pmatrix} \in \mathcal{M}_{n,n}(\mathbb{R}).$$

This matrix corresponds up to a coefficient $-h^2$ to the discretized Laplace operator in one-space dimension. Let $V \in \mathcal{M}_{n,n}(\mathbb{R})$ denote the matrix made of an orthonormal basis of eigenvectors of D_1:

$$
\begin{aligned}
V_{ij} &= \sqrt{\frac{2}{n}}\sin(ij\pi/(n+1)) \ \forall 1 \le i,j \le n, \\
V^{-1} &= V^t = V, \\
VD_1V &= \Lambda,
\end{aligned}
\tag{4.12}
$$

where the diagonal matrix $\Lambda \in \mathcal{M}_{n,n}(\mathbb{R})$ collects the eigenvalues of D_1. Let $W \in \mathcal{M}_{n^2,n^2}(\mathbb{R})$ be the block diagonal matrix, with n diagonal blocks equal to V:

$$
W = \left(\begin{array}{c|c|c|c|c}
V & \mathbb{O} & \cdots & \cdots & \mathbb{O} \\ \hline
\mathbb{O} & \ddots & & & \vdots \\ \hline
\vdots & & \ddots & & \vdots \\ \hline
\vdots & & & \ddots & \mathbb{O} \\ \hline
\mathbb{O} & \cdots & \cdots & \mathbb{O} & V
\end{array}\right) \in \mathcal{M}_{n^2,n^2}(\mathbb{R}).
$$

One can see that

$$
WGW = D_2,
$$

where

$$
D_2 = \frac{1}{h^2}\left(\begin{array}{c|c|c|c|c}
\tilde{\Lambda} & -I_n & \mathbb{O} & \cdots & \mathbb{O} \\ \hline
-\,I_n & \ddots & \ddots & & \vdots \\ \hline
\mathbb{O} & \ddots & \ddots & \ddots & \mathbb{O} \\ \hline
\vdots & & \ddots & \ddots & -I_n \\ \hline
\mathbb{O} & \cdots & \mathbb{O} & -I_n & \tilde{\Lambda}
\end{array}\right) \in \mathcal{M}_{n^2,n^2}(\mathbb{R})
$$

with $\tilde{\Lambda} = \left(\frac{h^2}{\delta t} + 2\right) I_n + \Lambda \in \mathcal{M}_n(\mathbb{R})$. Thus, we have

$$
I_{n_1} + P^t(M-G)G^{-1}P = I_{n_1} + P^t(M-G)WD_2^{-1}WP.
$$

We now show how one can compute this matrix in $\mathcal{O}(n^3)$ floating point operations:

- The computation of $WP \in \mathcal{M}_{n^2,n_1}(\mathbb{R})$ requires the computation of the eigenvectors of D_1 given by (4.12) (this gives $\mathcal{O}(n^2)$ floating point operations). Note that each column of WP is a vector in $\mathbb{R}^{n^2}$ with n nonvanishing coefficients at most.
- For each column w of WP, due to the block tridiagonal structure of D_2, the computation of $D_2^{-1}w$ is equivalent to the solution of n linear systems; each of them can be rewritten as a tridiagonal linear system of n equations of n unknowns. For each w, we then need $\mathcal{O}(n^2)$ floating point operations to compute $D_2^{-1}w$; this gives $\mathcal{O}(n^3)$ floating point operations for the computation of $D_2^{-1}WP \in \mathcal{M}_{n^2,n_1}$.

- By Remark 5, we see that $(M-G)W$ can be computed in $\mathcal{O}(n^2)$ floating point operations, and so the matrix $P^t(M-G)W \in \mathcal{M}_{n_1,n^2}$ can be computed in $\mathcal{O}(n^2)$ operations and has at most $\mathcal{O}(n)$ nonvanishing coefficients on each of its rows.
- Finally, the product of $P^t(M-G)W$ by $D_2^{-1}WP$ can be computed in $\mathcal{O}(n^3)$ floating point operations, and since this matrix is in $\mathcal{M}_{n_1,n_1}(\mathbb{R})$, its LU-factorization can also be computed in $\mathcal{O}(n^3)$ floating point operations.

We then need $\mathcal{O}(n^3)$ floating point operations to compute the LU-factorization of the matrix on the left-hand side of (4.9). This preprocessing step is done once (for all) at the beginning of the code.

Each time-step then requires to compute the right-hand side of (4.9) and then to solve the linear system in (4.9) using the LU-factorization of $(I_{n_1} + P^t(M-G)G^{-1}P)$ computed in the preprocessing step. The computation of $G^{-1}z$ requires $\mathcal{O}(n^2 \log n)$ floating point operations using fast Fourier transforms (see [8], [24]), and the product by $P^t(M-G)$ requires $\mathcal{O}(n)$ operations (using again Remark 5). We then need $\mathcal{O}(n^2 \log n)$ floating point operations to compute y in (4.9); it then takes another $\mathcal{O}(n^2 \log n)$ operations to compute x in (4.3) (the computation of (4.4) being neglectible).

Computing the solution at final time T requires $\mathcal{O}(h^{-3})$ floating point operations for the preprocessing step and $\mathcal{O}(\delta t^{-1} h^{-2} |\log h|)$ floating point operations for $N = T\delta t^{-1}$ time-steps.

4.2. Numerical results. For a numerical test, we consider the system (2.1)–(2.4), where $T = 1$ and $\Omega \subset \mathbb{R}^2$ (see Figure 2) is specified by the level set function

$$\psi(x,y) = 16(x-0.5)^4 + 16(y-0.5)^4 - 2.8(x-0.5)^2 + 0.2(x-0.5) - 0.8(y-0.5)^2 + 0.2(y-0.5).$$

We have chosen

$$\begin{aligned} \Omega &= \{(x,y) \in (0,1)\times(0,1) \colon \psi(x,y) < 0, (x,y) \notin B(M_1,r_1) \cup B(M_2,r_2)\},\\ \Gamma_N &= \{(x,y) \in (0,1)\times(0,1) \colon \psi(x,y) = 0\},\\ \Gamma_D &= \partial B(M_1,r_1) \cup \partial B(M_2,r_2) \end{aligned}$$

with

$$\begin{aligned} M_1 &= (0.7, 0.425) \quad \text{and} \quad r_1 = 0.1,\\ M_2 &= (0.3, 0.4) \quad \text{and} \quad r_2 = 0.125. \end{aligned}$$

The data f, u_0, u_N, and u_D (2.1)–(2.4) have been chosen so that the solution u satisfies

$$u(x,y,t) = (x^3 - 4xy^2 + 2y^4)\cos t + \sin t + e^{t^2}(\cos x + \sin y).$$

4.2.1. Euler implicit scheme. Table 2 presents the local error between the computed and the exact solution: $\max_{k,i,j} \left| u_{i,j}^k - u(x_{ij}, t_k) \right|$. This table shows an error of order $\mathcal{O}(\delta t + h^2)$ since this error is (approximately) divided by 4 when h is divided by 2 and δt is divided by 4. The observed convergence rates are better than the theoretical ones given by Theorem 3.2.

Table 3 presents the structure of the matrices M described in section 4.1. The total number of points in the linear system is n^2 (first column). In the second column the number of points is reported for which the discretization of the parabolic operator

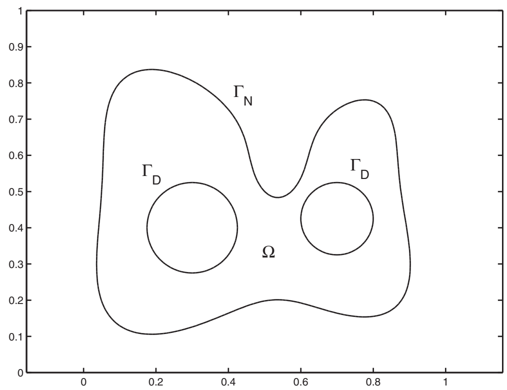

Fig. 2. *The domain* Ω.

Table 2
Local error.

	$\delta t = 4 \times 10^{-4}$	$\delta t = 2 \times 10^{-4}$	$\delta t = 10^{-4}$	$\delta t = 5 \times 10^{-5}$
$h = 8 \times 10^{-3}$	1.61×10^{-3}	1.62×10^{-3}	1.63×10^{-3}	1.63×10^{-3}
$h = 4 \times 10^{-3}$	3.81×10^{-4}	3.98×10^{-4}	4.06×10^{-4}	4.11×10^{-4}
$h = 2 \times 10^{-3}$	2.50×10^{-4}	1.03×10^{-4}	9.24×10^{-5}	9.65×10^{-5}
$h = 10^{-3}$	2.84×10^{-4}	1.36×10^{-4}	6.27×10^{-5}	2.58×10^{-5}

Table 3
Matrix structure.

	n^2	$n^2 - n_1$	n_1
$n = 124$ $h = 8 \times 10^{-3}$	15376	14841	535
$n = 249$ $h = 4 \times 10^{-3}$	62001	60937	1064
$n = 499$ $h = 2 \times 10^{-3}$	249001	246883	2118
$n = 999$ $h = 10^{-3}$	998001	993775	4226

is done, and we present in the last column the number of points (denoted n_1 in section 4) where a boundary condition is discretized. This table confirms that $n_1 = \mathcal{O}(n)$.

In Table 4 we record some CPU times when running these tests on an HP Proliant DL 145 computer with an ADM Opteron 2214 processor. The second column presents the CPU time for the preprocessing step which consists of computing the LU-factorization of $\left(I_{n_1} + P^t(M - G)G^{-1}P\right)$.

The third column shows the average time for one time-step, which should be of order $\mathcal{O}(n^2 \log n)$ according to the analysis in section 4.1.

4.2.2. Crank–Nicolson scheme. In this section we present some numerical results, where we replace the first order implicit scheme in time by a Crank–Nicolson scheme.

This time discretization requires at each time-step the solution of the following linear system:

$$(B + 2\delta t^{-1} D)\mathbf{u}^k = (-\bar{B} + 2\delta t^{-1} D)\mathbf{u}^{k-1} + \mathbf{F}^k, \tag{4.13}$$

Table 4
CPU times.

	Preproc. time	one time-step
$n = 124$, $h = 8 \times 10^{-3}$	$4.5s$	$0.21s$
$n = 249$, $h = 4 \times 10^{-3}$	$49s$	$0.89s$
$n = 499$, $h = 2 \times 10^{-3}$	$408s$	$3.74s$
$n = 999$, $h = 10^{-3}$	$4428s$	$15.06s$

Table 5
Local error.

	$\delta t = 4 \times 10^{-2}$	$\delta t = 2 \times 10^{-2}$	$\delta t = 10^{-2}$	$\delta t = 5 \times 10^{-3}$
$h = 8 \times 10^{-3}$	1.63×10^{-3}	1.64×10^{-3}	1.64×10^{-3}	1.64×10^{-3}
$h = 4 \times 10^{-3}$	4.72×10^{-4}	4.11×10^{-4}	4.14×10^{-4}	4.15×10^{-4}
$h = 2 \times 10^{-3}$	6.17×10^{-4}	1.21×10^{-4}	9.98×10^{-5}	1.00×10^{-4}
$h = 10^{-3}$	6.52×10^{-4}	1.55×10^{-4}	3.04×10^{-5}	2.45×10^{-5}

where

$$\bar{B} = \begin{pmatrix} B_1 & B_2 \\ \mathbb{O} & \mathbb{O} \end{pmatrix}, \tag{4.14}$$

and B_i, $i = 1, 2$ are given by (2.9). The top block of components of the vector $\mathbf{F}^k$ in (4.13) are the values of $(f(x_{ij}, t_{k-1}) + f(x_{ij}, t_k))$ for $x_{ij} \in \Omega_h$, with some modifications depending on u_D for the points in $\Gamma_{h,D}$ according to the Shortley–Weller approximation. The bottom block of components is given as in (2.10) by the values of $u_N(B, t_{k+1})$ with B as in (2.8).

The structure of the matrices is the same as for the Euler implicit scheme and can therefore be found in Table 3. The only additional operation required for the Crank–Nicolson scheme is the product $\bar{B}\mathbf{u}^{k-1}$ in the right-hand side of (4.13). This extra effort can be neglected compared to the time required for the solution of the system. The observed CPU times are less than 10% larger than those of Table 4.

The local error is reported in Table 5 and shows a second order convergence in space and time. As expected, the Crank–Nicolson scheme allows use of larger time-steps to obtain the same accuracy as the Euler explicit scheme.

Note that the matrix $-\bar{B} + 2\delta t^{-1} D$ on the right-hand side of (4.13) is not a non-negative matrix; it has negative diagonal entries if $2\delta t > h^2$. Therefore the technique developed in this paper cannot be directly appplied to show the convergence of the Crank–Nicolson scheme.

5. Conclusion. We have presented a finite difference scheme for a parabolic problem with mixed boundary conditions. The convergence analysis of the Euler implicit scheme shows that the error is bounded by $\mathcal{O}(\delta t + h)$, but the numerical experiments indicate the better $\mathcal{O}(\delta t + h^2)$ convergence rate. We mention that this difference between the theoretical analysis and the experiments has already been observed by several authors for parabolic problems with Neumann boundary conditions (see [25]). Numerical results also show the second order convergence in space and time of the Crank–Nicolson version of the method.

The presented scheme can be adapted to moving boundary problems (see [18]). We mention that the preprocessing step presented in section 4.1 would then need to be achieved at each time-step. Each of these time-steps would then require the sum of the times in columns 2 and 3 of Table 4. Numerical results for this problem will be the subject of future research.

Appendix A. Proof of Lemma 3.2. We show how the exterior point X can be connected to its interior neighbor P in each of the three cases discussed in [2]. Replacing α_ℓ by $\frac{a_\ell}{h}$ in (2.6), the diagonal coefficient of A in the row corresponding to X is given by $\frac{1}{h}\sum_{i=1}^4 a_i$ and the off-diagonal nonvanishing coefficients are $-\frac{a_\ell}{h}$ for $\ell = 1,\dots,4$. In cases 1 and 2 in [2] (see Table 1), we show that X is directly connected to $P_1 = P$. In case 1, using $1/5 \le n_2/n_1 \le 6/5$, we eventually find

$$\sum_{i=1}^4 a_i = (a_1 + a_2) + (a_3 + a_4) \le \left(n_1 - \beta\frac{n_2}{5}\right) + \frac{n_2}{5} \le n_1 + \frac{n_2}{5} \le \frac{31}{25}n_1$$

and

$$a_1 = \frac{1}{10}\left((11-2\beta)n_1 - (5+2\beta)n_2\right) \ge \frac{1}{10}\left((11-2\beta)n_1 - (6+\frac{12}{5}\beta)n_1\right) \ge \frac{3n_1}{50}.$$

Hence,

$$\frac{a_1}{\sum_{i=1}^4 a_i} \ge \frac{3}{62},$$

which implies that X is directly connected to P_1. We also note the estimate

$$\sum_{i=1}^4 a_i = (a_1+a_2)+(a_3+a_4) = \frac{1}{11}\left((12-2\beta)n_1 - (5+\frac{11\beta}{5})n_2\right) + \frac{n_2}{5} \ge \frac{10n_1}{11} - \frac{5n_2}{11},$$

which leads to

$$\sum_{i=1}^4 a_i \ge \frac{10n_1}{11} - \frac{6n_1}{11} \ge \frac{4n_1}{11} \ge \frac{5}{22},$$

since $n_1^2 = 1 - n_2^2 \ge 1 - \frac{36}{25}n_1^2$, and then $n_1 \ge \frac{\sqrt{25}}{\sqrt{61}} \ge \frac{5}{8}$ in this case. In the second case we observe that

$$a_1 + 3a_2 + 2a_3 + a_4 = n_1,$$

which implies

$$\frac{n_1}{3} \le \sum_{i=1}^4 a_i \le n_1.$$

In this case, a_1 is given by

$$a_1 = \frac{3-2\beta}{2}\left(n_1 - 3n_2\right) \ge \frac{n_1}{5},$$

which proves that

$$\frac{a_1}{\sum_{i=1}^4 a_i} \ge \frac{1}{5}.$$

Hence X is directly connected to P. In this case we have $\frac{n_2}{n_1} \leq \frac{1}{5}$; hence one can show $n_1 \geq \frac{5}{6}$ (using $n_1 + n_2 \geq 1$), which implies

$$\sum_{i=1}^{4} a_i \geq \frac{n_1}{3} \geq \frac{5}{18}.$$

In case 3, using $\beta \leq \frac{3}{4}$ and again $\frac{n_2}{n_1} \leq \frac{1}{5}$, one can derive the estimates

$$\begin{aligned} a_3 &\quad \frac{1}{2}(n_1 - \beta n_2) \geq \frac{17}{40} n_1 \geq \frac{2}{5} n_1, \\ \sum_{i=1}^{4} a_i &= -2\beta n_1 + \frac{5}{2}(n_1 - \beta n_2) + \beta n_2 \leq \frac{5}{2} n_1, \end{aligned}$$

which results in

$$\frac{a_3}{\sum_{i=1}^{4} a_i} \geq \frac{4}{25},$$

proving that X is directly connected to P_3. Using $n_1 \geq \frac{5}{6}$ again, one obtains the bound

$$\sum_{i=1}^{4} a_i \geq a_3 \geq \frac{1}{3}.$$

Hence in all cases the point X is directly connected to at least one point $P_i \in \Omega_h$, which is either P or a neighbor of P. By Remark 2 we conclude that X can be connected to its interior neighbor by a chain the length of which is bounded by 2 (the constant σ in Definition 3.1 can be chosen as 21). The bound below for $\sum_{i=1}^{4} a_i$ shows that the diagonal entries in A_4 can be bounded below by $\frac{5}{22h}$. Since the diagonal entries in A_1 are $\frac{4}{h^2}$ (or even larger in the row where a Shortley–Weller approximation is made), we deduce from (3.5)

$$\min_i Q_{ii} \geq C h^{-1}$$

for some positive constant C.

Moreover, since X is directly connected to at least one point $P_i \in \Omega_h$, then there is at least one coefficient on each row in A_3 whose modulus by Definition 3.1 and Lemma 3.2 is larger than C/h for some positive constant C. Using (3.21), this implies

$$A_4 \mathbf{1}_{N_e} = -A_3 \mathbf{1}_{N_i} \geq \frac{C}{h} \mathbf{1}_{N_e}$$

REFERENCES

[1] F. Bouchon, *Monotonicity of some perturbations of irreducibly diagonally dominant M-matrices*, Numer. Math., 105 (2007), pp. 591–601.

[2] F. Bouchon and G. H. Peichl, *A second order immersed interface technique for an elliptic Neumann problem*, Numer. Methods Partial Differential Equations, 23 (2007), pp. 400–420.

[3] J. H. Bramble and B. E. Hubbard, *A finite difference analog of the Neumann problem for Poisson's equation*, SIAM J. Soc. Indus. Appl. Math. Ser. B, 2 (1965), pp. 1–14.

[4] J. H. Bramble and B. E. Hubbard, *Approximation of solutions of mixed boundary value problems for Poisson's equation by finite differences*, J. ACM, 12 (1965), pp. 114–123.

[5] B. L. Buzbee and F. W. Dorr, *The direct solution of the biharmonic equation on rectangular regions and the Poisson equation on irregular regions*, SIAM J. Numer. Anal., 11 (1974), pp. 753–763.
[6] B. L. Buzbee, F. W. Dorr, J. A. George, and G. H. Golub, *The direct solution of the discrete Poisson equation on irregular regions*, SIAM J. Numer. Anal., 8 (1971), pp. 722–736.
[7] J. R. Cash, *Two new finite difference schemes for parabolic equations*, SIAM J. Numer. Anal., 21 (1984), pp. 433–446.
[8] R. W. Hockney, *A fast direct solution of Poisson's equation using Fourier analysis*, J. ACM, 12 (1965), pp. 95–113.
[9] H. Johansen and P. Colella, *A Cartesian grid embedded boundary method for Poisson's equation on irregular domains*, J. Comput. Phys., 147 (1998), pp. 60–85.
[10] Z. Jomaa and C. Macaskill, *The embedded finite difference method for the Poisson equation in a domain with an irregular boundary and Dirichlet boundary conditions*, J. Comput. Phys., 202 (2005), pp. 488–506.
[11] Z. Jomaa and C. Macaskill, *Numerical solution of the 2D Poisson equation on an irregular domain with Robin boundary conditions*, ANZIAM J., 50 (2008), pp. 413–428.
[12] Z. Jomaa and C. Macaskill, *The Shortley–Weller embedded finite-difference method for the 3D Poisson equation with mixed boundary conditions*, J. Comput. Phys., 229 (2010), pp. 3675–3690.
[13] R. J. LeVeque and Z. Li, *The immersed interface method for elliptic equations with discontinuous coefficients and singular sources*, SIAM J. Numer. Anal., 31 (1994), pp. 1019–1044.
[14] D. Leykekhman, *Pointwise localized error estimates for parabolic finite element equations*, Numer. Math., 96 (2004), pp. 583–600.
[15] Z. Li, *The immersed interface method—A numerical approach for partial differential equations with interfaces*, Ph.D. thesis, University of Washington, 1994.
[16] Z. Li and K. Ito, *The Immersed Interface Method: Numerical Solutions of PDEs Involving Interfaces and Irregular Domains*, SIAM, Philadelphia, 2007.
[17] N. Matsunaga and T. Yamamoto, *Superconvergence of the Shortley–Weller approximation for Dirichlet problems*, J. Comput. Appl. Math., 116 (2000), pp. 263–273.
[18] P. McCorquodale, P. Colella, and H. Johansen, *A Cartesian grid embedded boundary method for the heat equation on irregular domains*, J. Comput. Phys., 173 (2001), pp. 620–635.
[19] C. Palencia, *Maximum norm analysis of completely discrete finite element methods for parabolic problems*, SIAM J. Numer. Anal., 33 (1996), pp. 1654–1668.
[20] A. H. Schatz, V. Thomée, and L. B. Wahlbin, *Stability, analycity, and almost best approximation in maximum norm for parabolic finite element equations*, Comm. Pure Appl. Math., 51 (1998), pp. 1349–1385.
[21] P. Schwartz, M. Barad, P. Colella, and T. Ligocki, *A Cartesian grid embedded boundary method for the heat equation and Poisson's equation in three dimensions*, J. Comput. Phys., 211 (2006), pp. 531–550.
[22] G. H. Shortley and R. Weller, *The numerical solution of Laplace's equation*, J. Appl. Phys., 9 (1938), pp. 334–344.
[23] A. Solo, *Sharp estimates for finite element approximations to parabolic problems with Neumann boundary data of low regularity*, BIT, 48 (2008), pp. 117–137.
[24] P. N. Swarztrauber, *The methods of cyclic reduction, Fourier analysis and the FACR algorithm for the discrete solution of Poisson's equation on a rectangle*, SIAM Rev., 19 (1977), pp. 490–501.
[25] J. W. Thomas, *Numerical Methods in Partial Differential Equations*, Springer, New York, 1995.
[26] A. Wiegmann and K. P. Bube, *The explicit-jump immersed interface method: Finite difference methods for PDEs with piecewise smooth solutions*, SIAM J. Numer. Anal., 37 (2000), pp. 827–862.