A First-Order Explicit-Implicit Splitting Method for a Convection-Diffusion Problem

Citation for the original published paper (version of record):
Pani, A., Thomee, V., Murthy, A. (2020). A First-Order Explicit-Implicit Splitting Method for a Convection-Diffusion Problem. Computational Methods in Applied Mathematics, 20(4): 769-782.
http://dx.doi.org/10.1515/cmam-2020-0009

(article starts on next page)

Research Article

Amiya K. Pani, Vidar Thomée* and A. S. Vasudeva Murthy

A First-Order Explicit-Implicit Splitting Method for a Convection-Diffusion Problem

https://doi.org/10.1515/cmam-2020-0009
Received January 19, 2020; revised June 10, 2020; accepted July 13, 2020

Abstract: We analyze a second-order in space, first-order in time accurate finite difference method for a spatially periodic convection-diffusion problem. This method is a time stepping method based on the first-order Lie splitting of the spatially semidiscrete solution. In each time step, on an interval of length k, of this solution, the method uses the backward Euler method for the diffusion part, and then applies a stabilized explicit forward Euler approximation on $m \geq 1$ intervals of length $\frac{k}{m}$ for the convection part. With h the mesh width in space, this results in an error bound of the form $C_0 h^2 + C_m k$ for appropriately smooth solutions, where $C_m \leq C' + \frac{C''}{m}$. This work complements the earlier study [V. Thomée and A. S. Vasudeva Murthy, An explicit-implicit splitting method for a convection-diffusion problem, *Comput. Methods Appl. Math.* **19** (2019), no. 2, 283–293] based on the second-order Strang splitting.

Keywords: Convection-Diffusion Problem, Time Stepping Method, Backward Euler Method, Lie Splitting

MSC 2010: 35K10, 65M06, 65M15

1 Introduction

In this paper, we shall consider the numerical solution by finite differences and splitting of the convection-diffusion equation

$$\frac{\partial U}{\partial t} = \nabla \cdot (a \nabla U) + b \cdot \nabla U + F \quad \text{in } \Omega, \quad \text{for } t \geq 0, \quad \text{with } U(0) = V, \tag{1.1}$$

in the cube $\Omega = (0, 2\pi)^d$, with $d = 1, 2, 3$, under periodic boundary conditions. With $x = (x_1, \dots, x_d)$, we assume that the positive definite $d \times d$ matrix $a = a(x) = (a_{ij}(x))$ and the vector $b = b(x) = (b_1(x), \dots, b_d(x))$ as well as the forcing term $F = F(x, t)$ and the initial data $V = V(x)$ are periodic and smooth.

Equation (1.1) is a special case of the initial-value problem for the operator equation

$$\frac{dU}{dt} = -\mathcal{A}U + \mathcal{B}U + F \quad \text{for } t \geq 0, \quad \text{with } U(0) = V, \tag{1.2}$$

where $\mathcal{A}$ and $\mathcal{B}$ represent different physical processes, in our case diffusion and convection, respectively, or

$$\mathcal{A}U = -\nabla \cdot (a \nabla U) \quad \text{and} \quad \mathcal{B}U = b \cdot \nabla U. \tag{1.3}$$

The solution of (1.2) may be formally expressed as

$$U(t) = \mathcal{E}(t)V + \int_0^t \mathcal{E}(t-y)F(y)\,dy, \quad \text{where } \mathcal{E}(t) = e^{-t(\mathcal{A}-\mathcal{B})} \quad \text{for } t \geq 0.$$

***Corresponding author: Vidar Thomée,** Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, SE–412 96 Gothenburg, Sweden, e-mail: thomee@chalmers.se
Amiya K. Pani, Department of Mathematics, IIT Bombay, Powai, Mumbai 400076, India, e-mail: akp@math.iitb.ac.in
A. S. Vasudeva Murthy, TIFR Centre for Applicable Mathematics, Yelahanka New Town, Bangalore 560 065, India, e-mail: vasu@math.tifrbng.res.in

To discretize (1.2) in time, let k be a time step, and set $t_n = nk$. We then have

$$U(t_n) = \mathcal{E}(k)U(t_{n-1}) + \int_{t_{n-1}}^{t_n} \mathcal{E}(t_n - y)F(y)\,dy \quad \text{for } n \ge 1.$$

We may then find an approximate solution $\breve{U}^n \approx U(t_n)$, $n \ge 1$, by approximating the integrand by its value at t_n, or $\breve{U}^n = \mathcal{E}(k)\breve{U}^{n-1} + kF^n$, where $F^n = F(t_n)$ for $n \ge 1$, with $\breve{U}^0 = V$.

In the next step, we shall approximate $\mathcal{E}(k) = e^{-k(\mathcal{A}-\mathcal{B})}$ by splitting $\mathcal{A} - \mathcal{B}$ into $\mathcal{A}$ and $-\mathcal{B}$, and replace $\mathcal{E}(k)$ by using the Lie splitting

$$\mathcal{E}_k = e^{k\mathcal{B}}e^{-k\mathcal{A}} \approx \mathcal{E}(k) = e^{-k(\mathcal{A}-\mathcal{B})} \tag{1.4}$$

so that the time discrete solution U^n, $n \ge 1$, is defined by

$$U^n = \mathcal{E}_k U^{n-1} + kF^n, \quad \text{where } F^n = F(t_n) \quad \text{for } n \ge 1, \quad \text{with } U^0 = V. \tag{1.5}$$

Applying $\mathcal{E}_k$ thus involves solving the parabolic equation $U_t = -\mathcal{A}U$ and the hyperbolic equation $U_t = \mathcal{B}U$ on (t_{n-1}, t_n); see e.g. Hundsdorfer and Verwer [5], Descombes [1] and references therein. The two exponential factors of $\mathcal{E}_k$ may then be further approximated by rational functions of $\mathcal{A}$ and $\mathcal{B}$, respectively, such as the backward or forward Euler approximations.

We note that if $\mathcal{A}$ and $\mathcal{B}$ commute, which holds for (1.1) when a and b are independent of x, then $e^{-k(\mathcal{A}-\mathcal{B})} = e^{k\mathcal{B}}e^{-k\mathcal{A}}$ so that the error in (1.4) is zero. When $\mathcal{A}$ and $\mathcal{B}$ do not commute, then formally, by Taylor expansion, $e^{k\mathcal{B}}e^{-k\mathcal{A}} - e^{-k(\mathcal{A}-\mathcal{B})} = O(k^2)$. Under the appropriate assumptions, this will lead to an $O(k)$ error bound in (1.5) for t_n bounded.

Another possible approximation of $\mathcal{E}(k)$ is the Strang splitting $\mathcal{E}(k) \approx \mathcal{E}_k = e^{\frac{1}{2}k\mathcal{B}}e^{-k\mathcal{A}}e^{\frac{1}{2}k\mathcal{B}}$ for which the symmetry gives local accuracy of order $O(k^3)$, resulting in an error estimate in (1.5) of order $O(k^2)$ for t_n bounded, cf. [8]. In [9], a finite difference method based on the Strang splitting was analyzed for the homogeneous case of (1.1). The present paper using the less accurate Lie splitting may be thought of as a complement to [9]. It is intended to provide background for future work on more general problems, considering for instance maximum-norm estimates, reaction-diffusion equations, problems with nonsmooth data and also the combination of splitting with the finite element method. As a first step, the case of nonhomogeneous equations is included here.

Error estimates for time splittings may be found in [2–4, 6] and references therein. For an extensive list of related work, see MacNamara and Strang [7].

In the method that we study in this paper, we begin by discretizing (1.1) in the spatial variables. We let $h = \frac{2\pi}{M}$, where M is a positive integer, and define a corresponding uniform mesh

$$\Omega_h = \{x = x_\omega = h\omega = h(\omega_1, \dots, \omega_d),\ \omega_l = 1, 2, \dots, M,\ l = 1, \dots, d\}. \tag{1.6}$$

We denote by $\mathcal{P}_h$ the M-periodic vectors u with elements u_ω, corresponding to the mesh points x_ω, and with $u_{\omega+Me_l} = u_\omega$, $l = 1, \dots, d$, where e_l is the unit vector in the x_l direction. Note that we use capital letters for functions in Ω and lowercase letters for vectors in $\mathcal{P}_h$. We consider now the second-order spatially discrete, continuous in time, finite difference ODE system for $u(t) \in \mathcal{P}_h$,

$$\frac{du}{dt} = -Au + Bu + f \quad \text{for } t \ge 0, \quad \text{with } u(0) = v, \tag{1.7}$$

where the $M^d \times M^d$ matrices A and B corresponding to the differential operators $\mathcal{A}$ and $\mathcal{B}$ in (1.3) are defined by

$$Au = -\sum_{i,j=1}^{d} \bar{\partial}_j(\acute{a}_{ij}\partial_i u) \quad \text{and} \quad Bu = \sum_{j=1}^{d} \frac{1}{2} b_j(\partial_j + \bar{\partial}_j)u. \tag{1.8}$$

Here ∂_j and $\bar{\partial}_j$ are forward and backward finite difference quotients in the direction of x_j,

$$\acute{a}_{ij}(x_\omega) = a_{ij}\Big(x_\omega + \frac{1}{2}he_i\Big),$$

and v and f the restrictions of V and F to Ω_h. It is then to (1.7) that we will apply the Lie splitting.

The solution of (1.7), the spatially semidiscrete solution, is

$$u(t) = E(t)v + \int_0^t E(t-y)f(y)\,dy, \quad \text{where } E(t) = e^{-t(A-B)},$$

and we shall see that the error in this approximation is of order $O(h^2)$, under the appropriate regularity assumptions.

For the time discretization of (1.7), an obvious first choice would be the backward Euler method, defining the discrete solution $\breve{u}^n$ at t_n by

$$\breve{u}^n = \breve{E}_k \breve{u}^{n-1} + kf^n, \quad n \geq 1, \quad \text{with } \breve{u}^0 = v, \quad \text{where } \breve{E}_k = (I + kA - kB)^{-1}. \tag{1.9}$$

The basis of the method we study here, however, will be the splitting method analogous to (1.4), (1.5), defined by

$$\begin{gathered} u^n = E_k u^{n-1} + kf^n \quad \text{for } n \geq 1, \quad \text{with } u^0 = v, \\ \text{where } E_k = e^{kB}e^{-kA} \approx E(k) = e^{-k(A-B)}. \end{gathered} \tag{1.10}$$

Similarly to the continuous problem, when A and B commute, $E_k = E(k)$.

For a practical numerical method, we then need to approximate the two exponential factors in E_k by rational functions. For the approximation of e^{-kA}, we shall use the unconditionally stable backward Euler operator $Q_k = (I + kA)^{-1} \approx e^{-kA}$ for $k > 0$. To approximate e^{kB}, we would like to use the forward Euler method, or $e^{kB} \approx I + kB$. For stability reasons, we shall need to add an artificial viscosity term, so we set $Lu = -\sum_{j=1}^d \bar{\partial}_j \partial_j u$ and define $H_k = I + kB - \gamma k^2 L$ for $\frac{k}{h} \leq \rho_0$, where the positive constants γ and ρ_0 will be defined later.

In order to increase the limit ρ_0 for the mesh ratio, and to increase the accuracy of the approximation in the hyperbolic part, one may replace the operator H_k by $H_{k_m}^m$, with $k_m = \frac{k}{m}$ for some positive integer m, and the stability requirement will then be reduced to $\frac{k}{h} \leq m\rho_0$. This means that the approximate solution of the hyperbolic part of E_k is obtained in m steps of length $k_m = \frac{k}{m}$.

We consider thus the time discrete solution at time $t_n = nk$, defined by

$$\begin{gathered} \widetilde{u}_m^n = \widetilde{E}_{m,k} \widetilde{u}_m^{n-1} + kf^n, \quad \text{with } f^n = f(t_n) \quad \text{for } n \geq 1, \quad \widetilde{u}_m^0 = v, \\ \text{where } \widetilde{E}_{m,k} = H_{k_m}^m Q_k, \quad \text{with } k \leq m\rho_0 h. \end{gathered} \tag{1.11}$$

This method, which we will refer to as our fully discrete method, thus replaces at each time step the backward Euler solution of our nonsymmetric problem (1.7) by a backward Euler approximation of a symmetric parabolic problem, followed by an explicit finite difference solution of a hyperbolic problem.

After the introduction of notation and some preliminary observations in Section 2, the spatially semidiscrete problem (1.7), (1.8), the backward Euler method (1.9) and the corresponding basic time splitting method (1.10) will be analyzed in Section 3, where we will show $O(h^2)$ and $O(h^2 + k)$ error bounds for these methods, respectively. The analysis of the fully discrete method (1.11) will be carried out in Section 4, using discrete Sobolev norms. The time discretization error can be divided into an $O(k)$ part associated with the parabolic part of the equation and an $O(\frac{k}{m})$ term for the hyperbolic part. In Section 5, we illustrate our theoretical results with computations for examples with one and two spatial dimensions.

2 Notation and Preliminaries

With Ω_h as in (1.6), we introduce the discrete inner product and norm

$$(v, w)_h = h^d \sum_{x_\omega \in \Omega_h} v_\omega w_\omega \quad \text{and} \quad \|v\|_h = (v, v)_h^{\frac{1}{2}} \quad \text{for } v, w \in \mathcal{P}_h.$$

Further, for $x \in \Omega_h$, we set

$$\partial_j u(x) = \frac{u(x + he_j) - u(x)}{h} \quad \text{and} \quad \partial^\alpha u = \partial_1^{\alpha_1} \dots \partial_d^{\alpha_d} u \quad \text{for } \alpha = (\alpha_1, \dots, \alpha_d),$$

and define the discrete Sobolev inner product and norm, with $|\alpha| = \alpha_1 + \cdots + \alpha_d$, by

$$(v, w)_{h,s} = \sum_{|\alpha|\le s} (\partial^\alpha v, \partial^\alpha w)_h, \quad \|v\|_{h,s} = (v, v)_{h,s}^{\frac{1}{2}} \quad \text{for } s \ge 0.$$

We shall also use

$$\bar{\partial}_j u(x) = \frac{u(x) - u(x - he_j)}{h}.$$

For a domain $\mathcal{G} \subset R^d$, we define the norm on the Sobolev space $H^s(\mathcal{G})$ by

$$\|U\|_{s,\mathcal{G}} = \Big(\sum_{|\alpha|\le s} \|D^\alpha U\|^2_{L_2(\mathcal{G})}\Big)^{\frac{1}{2}}, \quad D^\alpha = \Big(\frac{\partial}{\partial x_1}\Big)^{\alpha_1} \cdots \Big(\frac{\partial}{\partial x_d}\Big)^{\alpha_d}.$$

For U periodic, we write $\|U\|_s$ for its norm on $H^s = H^s(\Omega)$. We note that, defining U_h to be the restriction of U to the mesh Ω_h, i.e., by $(U_h)_\omega = U(x_\omega)$, we have

$$\|U_h\|_{h,s} \le C\|U\|_s \quad \text{for } s > \frac{d}{2}. \tag{2.1}$$

In fact, let

$$Q_s(x) = \{y = (y_1, \dots, y_d) : |y_l - x_l| \le sh,\ l = 1, \dots, d\}.$$

Then, for $|\alpha| = s$, $\ell(U) = (\partial^\alpha U_h)_\omega$ is a bounded linear functional on $\mathbb{C}(Q_s(x_\omega))$, and therefore, by the Sobolev embedding theorem, if $s > \frac{d}{2}$, also on $H^s(Q_s(x_\omega))$, which vanishes for polynomials of degree $\le s - 1$. A simple argument using the Bramble–Hilbert lemma therefore shows

$$|(\partial^\alpha U_h)_\omega| \le Ch^{-\frac{d}{2}} \|U\|_{s,Q(x_\omega)}. \tag{2.2}$$

Hence $|U_h|_{h,s} \le C\|U\|_s$, where $|v|^2_{h,s} = \sum_{|\alpha|=s} \|\partial^\alpha v\|^2_h$. Since $\|U_h\|_{h,s} \le C(\|U_h\|_h + |U_h|_{h,s})$, this shows (2.1). In particular, $\|U_h\|_{h,2} \le C\|U\|_2$ for $d = 1, 2, 3$; it is our need for this inequality which is the reason for our restriction on d in (1.1).

Consider now the matrices A and B in (1.8), and note that, for $|\alpha| \le s$,

$$\|\partial^\alpha Au - A\partial^\alpha u\|_h \le C\|u\|_{h,s+1} \quad \text{and} \quad \|\partial^\alpha Bu - B\partial^\alpha u\|_h \le C\|u\|_{h,s}. \tag{2.3}$$

In fact, $\partial^\alpha A$ and $A\partial^\alpha$ are linear combinations of difference quotients of orders $\le s + 2$, and in the difference, the highest-order terms cancel so that the orders are $\le s + 1$, which shows the first inequality. The second inequality is shown analogously. Further, with C independent of h,

$$\|Au\|_{h,s} \le C\|u\|_{h,s+2} \quad \text{and} \quad \|Bu\|_{h,s} \le C\|u\|_{h,s+1}. \tag{2.4}$$

The matrix A is positive semidefinite, with

$$\|v\|^2_{h,1} \le C((Av, v)_h + \|v\|^2_h) \quad \text{and} \quad \|v\|^2_{h,2} \le C(\|Av\|^2_h + \|v\|^2_h). \tag{2.5}$$

Since the terms in AU_h are symmetric difference quotients of U at the mesh points x_ω, and the terms in $(\mathcal{A}U)_h$ are the corresponding derivatives of U at x_ω, we find, similarly to (2.2),

$$|(AU_h)(x_\omega) - (\mathcal{A}U)(x_\omega)| \le Ch^{2-\frac{d}{2}} \|U\|_{4,Q_1(x_\omega)}, \tag{2.6}$$

and similarly for B,

$$|(BU_h)(x_\omega) - (\mathcal{B}U)(x_\omega)| \le Ch^{2-\frac{d}{2}} \|U\|_{3,Q_1(x_\omega)}, \tag{2.7}$$

expressing, in particular, that (1.7) is a second-order approximation of (1.1).

For $B = (b_{\omega,\chi})$, we have

$$b_{\omega,\chi} = \begin{cases} \pm\frac{1}{2h} b_j(x_\omega) & \text{if } \chi = \omega \pm e_j,\ j = 1, \dots, d, \\ 0 & \text{for other } \chi, \end{cases}$$

and that then $b_{\omega+e_j,\omega} = -\frac{1}{2h} b_j(x_{\omega+e_j})$. We may write

$$B = B_0 + B_1, \quad \text{with } B_0 = \frac{1}{2}(B - B^T),\ B_1 = \frac{1}{2}(B + B^T).$$

Here B_0 is skew-symmetric and B_1 symmetric, with

$$\|B_1 v\|_h \le \beta_1 \|v\|_h, \quad \text{where } \beta_1 = \frac{1}{2}\sum_{j=1}^{d}\left\|\frac{\partial b_j}{\partial x_j}\right\|_{\mathcal{C}}.$$

In fact, for $d = 1$, with $l = \omega_1$, we have $b_{l,l\pm1} = \pm\frac{1}{2}\frac{b(x_l)}{h}$, and hence

$$\left|\frac{1}{2}(b_{l,l+1} + b_{l+1,l})\right| = \frac{1}{4}\left|\frac{b(x_{l+1}) - b(x_l)}{h}\right| \le \frac{1}{4}\|b'\|_{\mathcal{C}}$$

so that $\|B_1 v\|_h \le \frac{1}{2}\|b'\|_{\mathcal{C}}\|v\|_h$. The case $d > 1$ is treated analogously.

We note that $(B_0 v, v) = 0$ for all v so that

$$|(Bv, v)_h| = |(B_1 v, v)_h| \le \beta_1 \|v\|_h^2 \quad \text{for all } v \in \mathcal{P}_h. \tag{2.8}$$

We shall also use

$$\|Bv\|_h^2 \le \tilde{\beta}(Lv, v)_h \quad \text{for all } v \in \mathcal{P}_h, \quad \text{with } \tilde{\beta} = \left\|\sum_{j=1}^{d} b_j^2\right\|_{\mathcal{C}}, \tag{2.9}$$

In fact, since $(Bv)_\omega = \frac{1}{2}\big(\sum_{j=1}^d b_j(\partial_j + \bar\partial_j)v\big)_\omega$, we find

$$(Bv)_\omega^2 \le \frac{1}{4}\sum_{j=1}^{d} b_j(x_\omega)^2 \sum_{j=1}^{d}((\partial_j + \bar\partial_j)v_\omega)^2 \le \frac{1}{2}\tilde{\beta}\sum_{j=1}^{d}((\partial_j v_\omega)^2 + (\partial_j v_{\omega-e_j})^2).$$

Thus $\|Bv\|_h^2 \le \tilde{\beta}\sum_{j=1}^d \|\partial_j v\|_h^2 = \tilde{\beta}(Lv, v)_h$.

3 The Semidiscrete Problem, the Backward Euler Method and the Basic Splitting

We begin with the straightforward standard error analysis of the spatially semidiscrete problem (1.7), which we include for completeness. We first show stability and a smoothing property of the solution operator of (1.7) for $f = 0$, in discrete Sobolev norms.

Lemma 3.1. *Let $E(t) = e^{t(B-A)}$. Then, for any $s \ge 0$ and $T > 0$, we have, with $C = C_{T,s}$ independent of h,*

$$\|E(t)v\|_{h,s} \le Ct^{-\frac{j}{2}}\|v\|_{h,s-j} \quad \text{for } j = 0, \dots, s, \quad 0 < t \le T. \tag{3.1}$$

Proof. Let $s \ge 0$, and let $|\alpha| = s$. From (1.7), we find, for $u(t) = E(t)v$,

$$(\partial^\alpha u_t, \partial^\alpha u)_h + (\partial^\alpha Au, \partial^\alpha u)_h = (\partial^\alpha Bu, \partial^\alpha u)_h.$$

Hence, by (2.3) and (2.8),

$$\frac{1}{2}\frac{d}{dt}\|\partial^\alpha u\|_h^2 + (A\partial^\alpha u, \partial^\alpha u)_h + \|\partial^\alpha u\|_h^2 \le C\|u\|_{h,s+1}\|u\|_{h,s}.$$

Using (2.5) and summing over $|\alpha| \le s$, we find, with $c > 0$,

$$\frac{d}{dt}\|u\|_{h,s}^2 + 2c\|u\|_{h,s+1}^2 \le C\|u\|_{h,s+1}\|u\|_{h,s} \le c\|u\|_{h,s+1}^2 + C\|u\|_{h,s}^2,$$

or, by Gronwall's lemma, with $C = C_{T,s}$,

$$\|u(t)\|_{h,s}^2 + \int_0^t \|u(y)\|_{h,s+1}^2\,dy \le C\|v\|_{h,s}^2 \quad \text{for } t \le T, \tag{3.2}$$

from which, in particular, (3.1) with $j = 0$ follows.

Similarly, we have, for $s \ge 1$,

$$\|u_t\|^2_{h,s-1} + (Au, u_t)_{h,s-1} = (Bu, u_t)_{h,s-1} \le C\|u\|^2_{h,s} + \|u_t\|^2_{h,s-1},$$

and hence

$$\frac{d}{dt}(t(Au,u)_{h,s-1}) = (Au,u)_{h,s-1} + 2t(Au,u_t)_{h,s-1} \le C\|u\|^2_{h,s} \quad \text{for } t \le T,$$

or, by (3.2),

$$t(Au,u)_{h,s-1} \le C \int_0^t \|u(y)\|^2_{h,s}\,dy \le C\|v\|^2_{h,s-1}. \tag{3.3}$$

By (2.3) and (2.5), we have, for $|\alpha| \le s-1$,

$$c\|\partial^\alpha u\|^2_{h,1} \le (A\partial^\alpha u, \partial^\alpha u)_h + \|\partial^\alpha u\|^2_h \le (\partial^\alpha Au, \partial^\alpha u)_h + C\|u\|_{h,s}\,\|u\|_{h,s-1}.$$

By summation over $|\alpha| \le s-1$, this shows

$$\|u\|^2_{h,s} \le C(Au,u)_{h,s-1} + C\|u\|_{h,s}, \|u\|_{h,s-1},$$

and hence, after kicking $\|u\|_{h,s}$ back to the left, and using (3.3),

$$t\|u\|^2_{h,s} \le Ct(Au,u)_{h,s-1} + Ct\|u\|^2_{h,s-1} \le C\|v\|^2_{h,s-1} \quad \text{for } t \le T,$$

i.e., (3.1) for $j = 1$. For $j > 1$, (3.1) now follows from $E(t)v = E(\frac{t}{j})^j v$. □

Note that the special case of $E(t) = e^{-tA}$ is included for $B = 0$.

As a consequence of Lemma 3.1, we have the following second-order error estimate for the spatially semidiscrete solution $u(t)$ of (1.7).

Theorem 3.1. *For the solution u of* (1.7)*, with $v = V_h$, $f(t) = F_h(t)$ and $U(t)$ the exact solution of* (1.1)*, we have, for any $\varepsilon > 0$,*

$$\|u(t) - U_h(t)\|_h \le C_{\varepsilon,T}h^2\bigg(\|V\|_{2+\varepsilon} + \int_0^t \|F(\sigma)\|_{2+\varepsilon}\,d\sigma\bigg) \quad \textit{for } t \le T. \tag{3.4}$$

Proof. Setting $\omega = u - U_h$, we find

$$\omega_t = -A\omega + B\omega + \rho \quad \text{in } \Omega_h, \quad \text{for } t > 0, \quad \text{with } \omega(0) = 0,$$

where $\rho = ((\mathcal{A}U)_h - AU_h) - ((\mathcal{B}U)_h - BU_h)$. Here, by (2.6) and (2.7), $\|\rho(t)\|_h \le Ch^2\|U(t)\|_4$, and hence, since $E(t)$ is bounded by Lemma 3.1,

$$\|\omega(t)\|_h = \bigg\|\int_0^t E(t-y)\rho(y)\,dy\bigg\|_h \le C_T h^2 \int_0^t \|U(y)\|_4\,dy \quad \text{for } t \le T. \tag{3.5}$$

For the homogeneous equation, we recall the smoothing estimate

$$\|U(y)\|_4 = \|\mathcal{E}(y)V\|_4 \le C_\varepsilon y^{-1+\frac{\varepsilon}{2}}\|V\|_{2+\varepsilon} \quad \text{for } 0 < y \le T. \tag{3.6}$$

In the present context, this may be shown for $\varepsilon = 0$ and 1 in the same way as in Lemma 3.1, and for $\varepsilon \in (0,1)$ by interpolation. This estimate implies (3.4) when $F = 0$, by (3.5). To complete the proof of (3.4), we need to use linearity and add the estimate for the solution of the nonhomogeneous equation with $V = 0$, which is $U(y) = \int_0^y \mathcal{E}(y-\sigma)F(\sigma)\,d\sigma$. Using (3.6), we find

$$\|U(y)\|_4 \le C_\varepsilon \int_0^y (y-\sigma)^{-1+\frac{\varepsilon}{2}}\|F(\sigma)\|_{2+\varepsilon}\,d\sigma.$$

Hence (3.4) follows from

$$\int_0^t \|U(y)\|_4\,dy \le C_\varepsilon \varepsilon^{-1} \int_0^t (t-\sigma)^{\frac{\varepsilon}{2}}\|F(\sigma)\|_{2+\varepsilon}\,d\sigma \le C_{\varepsilon,T}\int_0^t \|F(\sigma)\|_{2+\varepsilon}\,d\sigma.$$ □

We now turn to the backward Euler method (1.9), the analysis of which will be the basis for the analysis of our splitting method (1.11). We begin with stability.

Lemma 3.2. *We have* $\|\breve{E}_k^n v\|_h \le e^{2\beta_1 T}\|v\|_h$, *for* $t_n \le T$ *and* $k \le \frac{1}{2\beta_1}$.

Proof. For $\breve{u}^1 = \breve{E}_k v$, we have $(I + kA - kB)\breve{u}^1 = v$, and hence

$$\|\breve{u}^1\|_h^2 + k(A\breve{u}^1, \breve{u}^1)_h = k(B\breve{u}^1, \breve{u}^1)_h + (v, \breve{u}^1)_h \le k\beta_1\|\breve{u}^1\|_h^2 + \|v\|_h\|\breve{u}^1\|_h$$

so that $\|\breve{u}^1\|_h \le k\beta_1\|\breve{u}^1\|_h + \|v\|_h$ or $\|\breve{E}_k v\|_h \le (1 - k\beta_1)^{-1}\|v\|_h \le e^{2\beta_1 k}\|v\|_h$ for $k \le \frac{1}{2\beta_1}$, from which the result follows. □

Lemma 3.3. *We have* $\|\breve{E}_k v - E(k)v\|_h \le Ck^j\|v\|_{h,2j}$ *for* $j = 1, 2$.

Proof. Setting $G(y) = (I + yA - yB)^{-1} - e^{y(B-A)}$ and noting that $G(0) = G'(0) = 0$, we obtain by Taylor's formula

$$\|\breve{E}_k v - E(k)v\|_h = \|G(k)v\|_h = \|(G(k) - G(0) - kG'(0))v\|_h \le \frac{1}{2}k^2 \sup_{y\le k}\|G''(y)v\|_h.$$

Here, for $y \le k$, using (2.3) and Lemmas 3.1 and 3.2, with $M = A - B$,

$$\|G''(y)v\|_h \le 2\|(I + yM)^{-3}M^2 v\|_h + \|e^{-yM}M^2 v\|_h \le C\|v\|_{h,4},$$

which shows our claim for $j = 2$. Similarly, for $j = 1$, the result follows from

$$\|G(k)v\|_h \le k \sup_{y\le k}\|G'(y)v\|_h \le Ck\|v\|_{h,2}.$$ □

We can now prove the following error estimate for the time discretization of the homogeneous equation.

Lemma 3.4. *With* $v = V_h$, *we have, for any* $\varepsilon > 0$,

$$\|\breve{E}_k^n v - E(t_n)v\|_h \le C_{\varepsilon,T}k\|V\|_{2+\varepsilon} \quad \textit{for } t_n \le T.$$

Proof. We write

$$\breve{E}_k^n v - E(t_n)v = \sum_{j=0}^{n-1} \breve{E}_k^{n-1-j}(\breve{E}_k - E(k))E(t_j)v.$$

Using Lemmas 3.2 and 3.3, we obtain, for $t_n \le T$,

$$\|\breve{E}_k^n v - E(t_n)v\|_h \le Ck\|v\|_{h,2} + Ck^2\sum_{j=1}^{n-1}\|E(t_j)v\|_{h,4}.$$

By Lemma 3.1 and (2.1), we have $\|E(t)v\|_{h,4} \le Ct^{-j}\|V_h\|_{h,4-2j} \le Ct^{-j}\|V\|_{4-2j}$ for $j = 0, 1$, and hence, by interpolation between H^2 and H^4,

$$\|E(t)v\|_{h,4} \le C_\varepsilon t^{-1+\frac{\varepsilon}{2}}\|V\|_{2+\varepsilon} \quad \text{for } \varepsilon > 0,\ t > 0.$$

Therefore,

$$\|\breve{E}_k^n v - E(t_n)v\|_h \le C_\varepsilon k\Big(1 + k\sum_{j=1}^{n-1} t_j^{-1+\frac{\varepsilon}{2}}\Big)\|V\|_{2+\varepsilon} \le C_{\varepsilon,T}k\|V\|_{2+\varepsilon}.$$ □

We now show the corresponding error estimate for the nonhomogeneous equation with vanishing initial values.

Lemma 3.5. *For the backward Euler solution* (1.9) *and the solution of* (1.7), *with* $v = 0$, *we have, for any* $\varepsilon > 0$,

$$\|\breve{u}^n - u(t_n)\|_h \le C_{\varepsilon,T}k\int_0^{t_n}(\|F(\sigma)\|_{2+\varepsilon} + \|F'(\sigma)\|_{\mathbb{C}})\,d\sigma \quad \textit{for } t_n \le T.$$

Proof. With $I_j = (t_{j-1}, t_j)$, we may write the error $e^n = \breve{u}^n - u(t_n)$ as

$$\begin{aligned} e^n &= k\sum_{j=1}^{n} \breve{E}_k^{n-j} f^j - \int_0^{t_n} E(t_n - y)f(y)\,dy \\ &= k\sum_{j=1}^{n} (\breve{E}_k^{n-j} - E(t_{n-j}))f^j + \sum_{j=1}^{n}\Big(kE(t_{n-j})f^j - \int_{I_j} E(t_n - y)f(y)\,dy\Big) = J' + J''. \end{aligned}$$

In J', we write

$$kf^j = \int_{I_j} f(y)\,dy + \int_{I_j}\int_y^{t_j} f'(\sigma)\,d\sigma\,dy.$$

Hence $J' = J_1' + J_2'$, where, using Lemma 3.4, for $t_n \le T$,

$$\|J_1'\|_h \le \sum_{j=1}^{n} \int_{I_j} \|(\breve{E}_k^{n-j} - E(t_{n-j}))f(y)\|_h\,dy \le C_{\varepsilon,T}k\int_0^{t_n} \|F(y)\|_{2+\varepsilon}\,dy.$$

Further, since $\breve{E}_k^{n-j} - E(t_{n-j})$ is bounded for $t_n \le T$,

$$\|J_2'\|_h \le C_T\sum_{j=1}^{n}\int_{I_j}\int_s^{t_j} \|F_h'(\sigma)\|_h\,d\sigma\,ds \le C_T k\int_0^{t_n} \|F'(\sigma)\|_{\mathbb{C}}\,d\sigma.$$

Similarly, $J'' = \sum_{j=1}^{n} J_j''$, where

$$\begin{aligned} J_j'' &= \int_{I_j} (E(t_{n-j})f^j - E(t_n - y)f(y))\,dy \\ &= \int_{I_j}\int_s^{t_j} \frac{d}{d\sigma}(E(t_n - \sigma)f(\sigma))\,d\sigma\,ds \\ &= \int_{I_j}\int_y^{t_j} E(t_n - \sigma)((A - B)f(\sigma) + f'(\sigma))\,d\sigma\,dy. \end{aligned}$$

Thus, again using (2.1),

$$\|J_j''\|_h \le C_T k\int_{I_j} (\|f(\sigma)\|_{h,2} + \|f'(\sigma)\|_h)\,d\sigma \le C_T k\int_{I_j} (\|F(\sigma)\|_2 + \|F'(\sigma)\|_{\mathbb{C}})\,d\sigma,$$

and thus

$$\|J''\|_h \le \sum_{j=1}^{n} \|J_j''\|_h \le C_T k\int_0^{t_n} (\|F(\sigma)\|_2 + \|F'(\sigma)\|_{\mathbb{C}})\,d\sigma \quad \text{for } t_n \le T.$$

Since $\|e^n\|_h \le \|J_1'\|_h + \|J_2'\|_h + \|J''\|_h$, this completes the proof. □

Together, Theorem 3.1 and Lemmas 3.4 and 3.5 show the following complete error estimate for the backward Euler method.

Theorem 3.2. *Let $\breve{u}^n$ be the solution of* (1.9) *and $U(t_n)$ the exact solution of* (1.1) *at $t = t_n$. Then, for any $\varepsilon > 0$ and $t_n \le T$,*

$$\|\breve{u}^n - U_h(t_n)\|_h \le C_{\varepsilon,T}(h^2 + k)\Big(\|V\|_{2+\varepsilon} + \int_0^{t_n} (\|F(\sigma)\|_{2+\varepsilon} + \|F'(\sigma)\|_{\mathbb{C}})\,d\sigma\Big).$$

We now turn to our basic splitting method. For small t, we shall need a stability estimate for the hyperbolic part of $E_k = e^{kB}e^{-kA}$.

Lemma 3.6. *For any $s \ge 0$, there is a constant $\bar\beta_s \ge \beta_1$, with $\bar\beta_0 = \beta_1$, such that*

$$\|e^{tB}v\|_{h,s} \le e^{\bar\beta_s t}\|v\|_{h,s} \quad \text{for } t \ge 0. \tag{3.7}$$

Also, with $|||v|||_h^2 = \|Av\|_h^2 + \|v\|_h^2$ and $\kappa > 0$,

$$|||e^{tB}v|||_h \le e^{\kappa t}|||v|||_h \quad \text{for } t \ge 0. \tag{3.8}$$

Proof. Let $s \ge 0$, and let $|\alpha| \le s$. Then, for $w(t) = e^{tB}v$, using (2.3),

$$(\partial^\alpha w_t, \partial^\alpha w)_h = (\partial^\alpha Bw, \partial^\alpha w)_h \le (B\partial^\alpha w, \partial^\alpha w)_h + C\|w\|_{h,s}^2.$$

By (2.8) and summation, we conclude that

$$\frac{d}{dt}\|w\|_{h,s}^2 \le (2\beta_1 + C)\|w\|_{h,s}^2, \tag{3.9}$$

which shows (3.7). For $s = 0$, this holds with $C = 0$, and hence $\bar\beta_0 = \beta_1$.

Using $\|(AB - BA)w\|_h \le C\|w\|_{h,2}$, cf. (2.3), (2.4) and (2.8), we have

$$(Aw_t, Aw)_h = (ABw, Aw)_h \le (BAw, Aw)_h + C\|w\|_{h,2}^2 \le C\|w\|_{h,2}^2 \le C|||w|||_h^2.$$

Using also (3.9) with $s = 0$, we find, by addition, $\frac{d}{dt}|||w|||_h^2 \le 2\kappa|||w|||_h^2$, with $\kappa > 0$, which shows (3.8). □

In particular, this implies the stability of the basic time stepping operator.

Lemma 3.7. *We have*

$$\|E_k^n v\|_h \le e^{\beta_1 T}\|v\|_h \quad \text{for } t_n \le T. \tag{3.10}$$

Proof. Since obviously $\|e^{-kA}v\|_h \le \|v\|_h$ for all $v \in \mathcal{P}_h$ and $k \ge 0$, we have, by (3.7) with $s = 0$,

$$\|E_k v\|_h = \|e^{kB}e^{-kA}v\|_h \le e^{\beta_1 k}\|e^{-kA}v\|_h \le e^{\beta_1 k}\|v\|_h,$$

which implies (3.10). □

We now show the following local-in-time error estimate for E_k.

Lemma 3.8. *We have $\|E_k v - E(k)v\|_h \le Ck^j\|v\|_{h,2j}$ for $j = 1, 2$.*

Proof. The proof is analogous to that of Lemma 3.3, now with $G(y) = e^{yB}e^{-yA} - e^{y(B-A)}$, in which case,

$$\|G''(y)v\|_h \le \sum_{i_1+i_2=2}\|e^{yB}B^{i_1}A^{i_2}e^{-yA}v\|_h + \|(B-A)^2e^{y(B-A)}v\|_h \le C\|v\|_{h,4}. \qquad \square$$

We now show the following error estimate for the basic splitting method.

Theorem 3.3. *Let u^n be the solution of* (1.10) *and $U(t_n)$ the exact solution of* (1.1) *at $t = t_n$. Then, for any $\varepsilon > 0$ and $t_n \le T$,*

$$\|u^n - U_h(t_n)\|_h \le C_{\varepsilon,T}(h^2 + k)\biggl(\|V\|_{2+\varepsilon} + \int_0^{t_n}(\|F(\sigma)\|_{2+\varepsilon} + \|F'(\sigma)\|_{\mathbb{C}})\,d\sigma\biggr).$$

Proof. The proof is analogous to that of Theorem 3.2 for the backward Euler method. For the homogeneous equation, we obtain, using Lemma 3.8 instead of Lemma 3.3, for any $\varepsilon > 0$,

$$\|E_k^n v - E(t_n)v\|_h \le C_{\varepsilon,T}k\|V\|_{2+\varepsilon} \quad \text{for } t_n \le T. \tag{3.11}$$

Using this then shows, as in Lemma 3.5, for $v = 0$,

$$\|u^n - u(t_n)\|_h \le C_{\varepsilon,T}k\int_0^{t_n}(\|F(\sigma)\|_{2+\varepsilon} + \|F'(\sigma)\|_{\mathbb{C}})\,d\sigma \quad \text{for } t_n \le T. \tag{3.12}$$

The proof is now completed by Theorem 3.1. □

We note that, when A and B commute, the error in (3.11) vanishes, and thus the term in $k\|V\|_{2+\varepsilon}$ in Theorem 3.3 may be removed. However, as is easily checked, (3.12) remains unchanged.

4 The Fully Discrete Splitting Method

We now turn to our proposed time stepping operator $\widetilde{E}_{m,k} = H_{k_m}^m Q_k$ defined in (1.11) and begin with the following stability result.

Lemma 4.1. *Let β_1 and $\widetilde{\beta}$ be as in (2.8) and (2.9), and let $\gamma > \widetilde{\beta}$. Then*

$$\|H_k v\|_h \le (1 + \beta_1 k)\|v\|_h \quad \text{if } \frac{k}{h} \le \rho_0, \quad \text{where } \rho_0^2 = \frac{\gamma - \widetilde{\beta}}{4d\gamma^2}. \tag{4.1}$$

Further, with $m \ge 1$, we have

$$\|\widetilde{E}_{m,k}^n v\|_h \le e^{\beta_1 T}\|v\|_h \quad \text{for } \frac{k}{h} \le m\rho_0,\ t_n \le T. \tag{4.2}$$

Proof. We have

$$\|H_k v\|_h^2 = \|v + kBv - \gamma k^2 Lv\|_h^2 = \|v\|_h^2 + 2k(Bv, v)_h - 2\gamma k^2 (Lv, v)_h + \|kBv - \gamma k^2 Lv\|_h^2.$$

We note that $\|Lv\|_h^2 \le \lambda_{max}(L)(Lv, v)_h \le 4dh^{-2}(Lv, v)_h$. We thus find

$$\begin{aligned}\|kBv - \gamma k^2 Av\|_h^2 \le 2k^2\|Bv\|_h^2 + 2\gamma^2 k^4\|Lv\|_h^2 &\le 2k^2\Big(\widetilde{\beta} + 4d\gamma^2\Big(\frac{k}{h}\Big)^2\Big)(Lv, v)_h \\ &\le 2k^2(\widetilde{\beta} + 4d\gamma^2\rho_0^2)(Lv, v)_h = 2k^2\gamma(Lv, v)_h.\end{aligned}$$

Hence, using (2.8),

$$\|H_k v\|_h^2 \le \|v\|_h^2 + 2k|(Bv, v)_h| \le \|v\|_h^2 + 2\beta_1 k\|v\|_h^2 \le (1 + \beta_1 k)^2\|v\|_h^2,$$

which shows (4.1). Since $\|Q_k v\|_h \le \|v\|_h$, we have

$$\|\widetilde{E}_{m,k} v\|_h = \|H_{k_m}^m Q_k v\|_h \le \Big(1 + \beta_1\frac{k}{m}\Big)^m \|Q_k v\|_h \le e^{\beta_1 k}\|v\|_h.$$

Hence

$$\|\widetilde{E}_{m,k}^n v\|_h \le e^{n\beta_1 k}\|v\|_h \le e^{\beta_1 T}\|v\|_h \quad \text{for } t_n \le T,$$

which completes the proof of (4.2). □

We note that the choice of γ which makes ρ_0 as large as possible is $\gamma = 2\widetilde{\beta}$, in which case $\rho_0 = 1/\sqrt{16d\widetilde{\beta}}$.

We start the analysis of the time discretization error with the following lemma concerning the error in the hyperbolic and parabolic parts of E_k, where we assume that α_1, α_2 and β_2 satisfy

$$\|A^j v\|_h \le \alpha_j\|v\|_{h,2j}, \quad j = 1, 2, \quad \text{and} \quad \|B^2 v\|_h \le \beta_2\|v\|_{h,2} \quad \text{for all } v \in \mathcal{P}_h.$$

Lemma 4.2. *With ρ_0 as in Lemma 4.1 and $\bar{\beta}_2$ as in (3.7), we have, with $C_0 = \frac{1}{2}\beta_2 + \alpha_1\gamma$,*

$$\|(e^{kB} - H_{k_m}^m)v\|_h \le C_0 e^{\bar{\beta}_2 k} m^{-1} k^2\|v\|_{h,2} \quad \text{for } \frac{k}{h} \le m\rho_0. \tag{4.3}$$

Further,

$$\|(e^{-kA} - Q_k)v\|_h \le C_j k^j\|v\|_{h,2j} \quad \text{for } j = 1, 2. \tag{4.4}$$

Proof. By Taylor expansion and Lemma 3.6, we have

$$\begin{aligned}\|(e^{kB} - H_k)v\|_h \le \|(e^{kB} - I - kB)v\|_h + \gamma k^2\|Av\|_h &\le \Big\|\int_0^k (k - y)e^{yB}B^2 v\,dy\Big\|_h + \gamma k^2\|Av\|_h \\ &\le \frac{1}{2}e^{\beta_1 k}k^2\|B^2 v\|_h + \gamma k^2\|Av\|_h \le e^{\beta_1 k}\Big(\frac{1}{2}\beta_2 + \alpha_1\gamma\Big)k^2\|v\|_{h,2},\end{aligned} \tag{4.5}$$

which shows (4.3) for $m = 1$.

To show (4.3) for $m > 1$, we write

$$H_{k_m}^m v - e^{kB}v = \sum_{j=0}^{m-1} H_{k_m}^{m-j-1}(H_{k_m} - e^{k_m B})e^{jk_m B}v. \tag{4.6}$$

By (4.5), we have

$$\|(e^{k_m B} - H_{k_m})v\|_h \le C_0 e^{\beta_1 k_m} k_m^2 \|v\|_{h,2} \quad \text{for } \frac{k}{h} \le m\rho_0. \tag{4.7}$$

Using (4.6), (4.1), (4.7) and (3.7), we conclude

$$\begin{aligned}\|H_{k_m}^m v - e^{kB}v\|_h &\le \sum_{j=0}^{m-1} e^{(m-1-j)k_m\beta_1}\|(e^{k_m B} - H_{k_m})e^{jk_m B}v\|_h\\ &\le C_0 k_m^2 \sum_{j=0}^{m-1} e^{(m-j)k_m\beta_1}\|e^{jk_m B}v\|_{h,2}\\ &\le C_0 e^{mk_m\bar{\beta}_2} mk_m^2 \|v\|_{h,2} = C_0 e^{\bar{\beta}_2 k} m^{-1}k^2\|v\|_{h,2}.\end{aligned}$$

For (4.4), it is easily verified that $0 \le (1+\lambda)^{-1} - e^{-\lambda} \le c_j\lambda^j$ for $\lambda > 0$, $j = 1, 2$, and hence, by eigenfunction expansion,

$$\|Q_k v - e^{-kA}v\|_h \le c_j k^j \|A^j v\|_h \le c_j\alpha_j k^j \|v\|_{h,2j}. \qquad \square$$

We continue the analysis by defining the one-step operator $\mathring{E}_k = e^{kB}Q_k$ approximating only the factor e^{-kA} in the product $E_k = e^{kB}e^{-kA}$ and leaving the factor e^{kB} exact. We then define the corresponding solution of the nonhomogeneous equation by

$$\mathring{u}^n = \mathring{E}_k \mathring{u}^{n-1} + kf^n \quad \text{for } n \ge 1. \tag{4.8}$$

For the homogeneous equation, we write the error in the full discretization of the basic splitting method after n steps as

$$\widetilde{E}_{m,k}^n v - E_k^n v = (\mathring{E}_k^n v - E_k^n v) + (\widetilde{E}_{m,k}^n v - \mathring{E}_k^n v). \tag{4.9}$$

We note that the first term on the right is independent of m. We begin by estimating this term.

Lemma 4.3. *Let $\varepsilon > 0$. Then, for $v = V_h$ and $t_n \le T$,*

$$\|\mathring{E}_k^n v - E_k^n v\|_h \le C_{\varepsilon,T} k\|V\|_{2+\varepsilon}.$$

Proof. In view of (3.11) in the proof of Theorem 3.3, it suffices to show the same bound for $\mathring{E}_k^n v - E(t_n)v$. Using Lemma 3.6, we have $\|\mathring{E}_k v\|_h \le e^{\beta_1 k}\|Q_h v\|_h \le e^{\beta_1 k}\|v\|_h$ so that $\mathring{E}_k$ is stable, and hence

$$\|\mathring{E}_k^n v - E(t_n)v\|_h \le e^{\beta_1 t_n}\sum_{j=0}^{n-1}\|(\mathring{E}_k - E(k))E(t_j)v\|_h.$$

Here, by Lemmas 3.6, 4.2 and 3.8,

$$\begin{aligned}\|\mathring{E}_k v - E(k)v\|_h &\le \|\mathring{E}_k v - E_k v\|_h + \|E_k v - E(k)v\|_h\\ &\le \|e^{kB}(Q_k - e^{-kA})v\|_h + Ck^j\|v\|_{h,2j} \le Ck^j\|v\|_{h,2j}, \quad j = 1, 2.\end{aligned}$$

Hence, as in Lemma 3.4, for $t_n \le T$,

$$\|\mathring{E}_k^n v - E(t_n)v\|_h \le C\Big(k\|v\|_{h,2} + k^2\sum_{j=1}^{n-1}\|E(t_j)v\|_{h,4}\Big) \le C_{\varepsilon,T}k\|V\|_{2+\varepsilon}. \qquad \square$$

For the second term on the right in (4.9), we have the following lemma.

Lemma 4.4. *With ρ_0 and m as in Lemma 4.1, we have, with $v = V_h$,*

$$\|\widetilde{E}_{m,k}^n v - \mathring{E}_k^n v\|_h \le C_T k m^{-1}\|v\|_{h,2} \le C_T k m^{-1}\|V\|_2 \quad \textit{for } t_n \le T.$$

Proof. By Lemma 4.2, we have

$$\|(\widetilde{E}_{m,k} - \mathring{E}_k)v\|_h = \|(H^m_{k_m} - e^{kB})Q_k v\|_h \le Ck^2 m^{-1}\|Q_k v\|_{h,2} \le Ck^2 m^{-1}\|v\|_{h,2},$$

where the last step follows since $\|\cdot\|_{h,2}$ is equivalent to $|||\cdot|||_h$, and hence

$$\|Q_h v\|_{h,2} \le C|||Q_h v|||_h \le C|||v|||_h \le C\|v\|_{h,2}.$$

Since $\widetilde{E}_{m,k}$ is stable in $\|\cdot\|_h$ by Lemma 4.1, we find, for $t_n \le T$,

$$\|\widetilde{E}^n_{m,k}v - \mathring{E}^n_k v\|_h \le e^{\beta_1 t_n}\sum_{j=0}^{n-1}\|(\widetilde{E}_{m,k} - \mathring{E}_k)\mathring{E}^j_k v\|_h \le C_T\sum_{j=0}^{n-1}k^2 m^{-1}\|\mathring{E}^j_k v\|_{h,2}.$$

Further, $\mathring{E}_k$ is stable in $\|\cdot\|_{h,2}$ since, by Lemma 3.6,

$$|||\mathring{E}_k v|||_h = |||e^{kB}Q_k v|||_h \le e^{Ck}|||Q_k v|||_h \le e^{Ck}|||v|||_h.$$

Hence, for $t_n \le T$,

$$\|(\widetilde{E}^n_{m,k} - \mathring{E}^n_k)v\|_h \le C_T nk^2 m^{-1}\|v\|_{h,2} \le C_T km^{-1}\|V\|_2. \qquad \square$$

The results corresponding to Lemmas 4.3 and 4.4 for the nonhomogeneous equation are the following.

Lemma 4.5. *Let $\mathring{u}^n$ and u^n be defined by* (4.8) *and* (1.10)*, respectively, with $v = 0$. Then, for $\varepsilon > 0$, we have*

$$\|\mathring{u}^n - u^n\|_h \le C_{\varepsilon,T}k\int_0^{t_n}(\|F(\sigma)\|_{2+\varepsilon} + \|F'(\sigma)\|_{\mathbb{C}})\,d\sigma \quad \textit{for } t_n \le T.$$

Lemma 4.6. *With ρ_0 as in Lemma 4.1 and $\varepsilon > 0$, let $\tilde{u}^n_m$ and $\mathring{u}^n$ be defined by* (1.11) *and* (4.8)*, respectively, with $v = 0$. Then we have, for $\frac{k}{h} \le m\rho_0$,*

$$\|\tilde{u}^n_m - \mathring{u}^n\|_h \le C_T\frac{k}{m}\int_0^{t_n}(\|F(\sigma)\|_2 + \|F'(\sigma)\|_{\mathbb{C}})\,d\sigma \quad \textit{for } t_n \le T.$$

The proofs are analogous to that of Lemma 3.5, using Lemmas 4.3 and 4.4 instead of Lemma 3.4, respectively.

Using Theorem 3.1 and Lemmas 4.3–4.6, we can now state our main result.

Theorem 4.1. *Let $\tilde{u}^n_m$ be the solution of* (1.11) *and $U(t_n)$ that of* (1.1) *at $t = t_n$. Then, with ρ_0 as in Lemma 4.1 and $\varepsilon > 0$, we have, for $t_n \le T$,*

$$\|\tilde{u}^n_m - U_h(t_n)\|_h \le (C_{\varepsilon,T}h^2 + (C'_{\varepsilon,T}m^{-1} + C''_{\varepsilon,T})k)\Big(\|V\|_{2+\varepsilon} + \int_0^{t_n}(\|F(\sigma)\|_{2+\varepsilon} + \|F'(\sigma)\|_{\mathbb{C}})\,d\sigma\Big).$$

5 Numerical Illustrations

In this section, we present some numerical computations to illustrate our error estimates. We begin with the one-dimensional version of (1.1),

$$U_t = (aU_x)_x + bU_x + F \quad \text{for } x \in (0, 2\pi),\ t > 0,$$

with $a(x) = 1 + \frac{1}{2}\cos x$, $b(x) = 1 + \frac{1}{2}\sin x$, and consider first the inhomogeneous term

$$F(x, t) = U_t - aU_{xx} - (a_x + b)U_x$$

with $U(x, t) = \sin(x - t) + \frac{1}{2}\cos 2(x + t)$ thus giving $V(x) = \sin x + \frac{1}{2}\cos 2x$. For the above choice of b, we obtain $\beta_1 = \frac{1}{2}\|b'\|_{\mathbb{C}} = \frac{1}{4}$ and $\tilde{\beta} = \|b^2\|_{\mathbb{C}} = 2.25$ giving $\gamma = 4.5$ and $\rho_0 = \frac{1}{6} \approx 0.17$. We take $h = \frac{2\pi}{M}$, $k = \frac{1}{N}$, with $N = M$, which ensures $\frac{k}{h} = \frac{1}{2\pi} \approx 0.159 < \rho_0$ thus satisfying the hypothesis of Theorem 4.1.

M	$\|u - U_h\|_h$	$\|\breve{u}^N - U_h\|_h$	$\|u^N - U_h\|_h$	$\|P^N\|_h$	$\|H_1^N\|_h$	$\|H_4^N\|_h$	$\|\widetilde{u}_4^N - U_h\|_h$
10	0.2061	0.4212	0.3379	0.0726	0.3082	0.0854	0.3626
20	0.0513	0.1718	0.1279	0.0389	0.1604	0.0425	0.1471
40	0.0128	0.0760	0.0545	0.0199	0.0825	0.0212	0.0660
80	0.0032	0.0355	0.0251	0.0101	0.0419	0.0106	0.0313

Table 1: Partial and complete errors for the first example ($d = 1$), with $m = 4$.

m	$\|u^N - U_h\|_h$	$\|u^N - u\|_h$	$\|P^N\|_h$	$\|H_m^N\|_h$	$\|T_m^N\|_h$	$\|\widetilde{u}_m^N - U_h\|_h$
1	0.0251	0.0232	0.0101	0.0419	0.03578	0.05055
2	"	"	"	0.0212	0.01605	0.03595
3	"	"	"	0.0142	0.01033	0.03255
4	"	"	"	0.0106	0.00824	0.03129
5	"	"	"	0.0085	0.00752	0.03071
8	"	"	"	0.0053	0.00751	0.03009
9	"	"	"	0.0047	0.00766	0.03001
10	"	"	"	0.0043	0.00781	0.02995
12	"	"	"	0.0036	0.00807	0.02988

Table 2: Partial and complete errors for the first example ($d = 1$), with m varying and $M = 80$.

The time stepping is carried out till $t_N = Nk = 1$, and the errors are presented in Tables 1 and 2. The first column of Table 1 contains the number M of spatial points, columns 2, 3 and 4 show bounds for the semidiscrete error (Theorem 3.1), the error using the backward Euler method (BE) (Theorem 3.2), and the basic splitting error (Theorem 3.3). We split the total remaining error in the time discretization error of the inhomogeneous equation as

$$T_m^N = \widetilde{u}_m^N - u^N = (\mathring{u}^N - u^N) + (\widetilde{u}_m^N - \mathring{u}^N) = P^N + H_m^N,$$

where P^N and H_m^N may be thought of as the parabolic and hyperbolic parts of the total error T_m^N and are given in columns 5 and 6. By Lemmas 4.3 and 4.5 for U appropriately smooth, $\|P^N\|_h \le C(U)k$, independently of m, and by Lemmas 4.4 and 4.6, $\|H_m^N\|_h \le C(U)\frac{k}{m}$. Columns 5 and 6 show that $\|P^N\|_h$ and $\|H_1^N\|_h$ are essentially proportional to k. Table 2 then shows that, for $M = 80$, $\|H_1^N\|_h$ is essentially proportional to $\frac{1}{m}$.

We want to discuss the choice of m for our problem. We recall the constants $C_0 = \frac{1}{2}\beta_2 + \alpha_1\gamma$ and $C_2 = c_2\alpha_2$ in Lemma 4.2. By simple calculations using our definitions, we may now take $\alpha_1 = \frac{\sqrt{10}}{2}$, $\alpha_2 = \frac{19}{2}$, $\beta_2 = \frac{5}{8}$, $c_2 = \frac{1}{2}$, and thus $C_0 \approx 7.4$ and $C_2 \approx 4.7$. Since $C_0 > C_2$, we have reason to believe that $\|H_1^N\|_h > \|P^N\|_h$, and this is confirmed by Table 1. A reasonable approach is to choose m in such a way that the parabolic and parabolic parts of the error balance. In our example, we have $\frac{\|H_1^N\|_h}{\|P^N\|_h} \approx 4.2$ for $M = 10$, and since this ratio is essentially independent of M, $m = 4$ would be a reasonable choice also for other M. This choice is used in columns 7 and 8 of Table 1.

We recall that $\|H_m^N\|_h \to 0$ as $m \to \infty$. Hence $\|T_m^N\|_h \to \|P^N\|_h$ as $m \to \infty$. We also observe that the sum of the norms in columns 4 and 5 of Table 2 would be a pessimistic estimate of $\|T_m^N\|_h$ in column 6. Furthermore, this error decreases for small m with a factor greater than m and, in fact, goes below its limit value $\|P^N\|_h$ and then starts to increase.

We next consider a case where the hyperbolic error is already smaller than the parabolic error for $m = 1$. For this, we take $a \equiv 4$, $b \equiv 1$, $U(x, t) = e^{-t}\sin(x + t)$ giving us $V(x) = \sin x$, $F = U_t - 4U_{xx} - U_x$. The results are presented in Table 3 with $N = 2M$. Here $\|H_1^N\|_h < \|P^N\|_h$, and thus $m = 1$ is a reasonable choice.

Finally, we consider the problem in two space dimensions for the nonhomogeneous equation

$$U_t = \Delta U + b_1 U_{x_1} + b_2 U_{x_2} + F \quad \text{for } x = (x_1, x_2) \in (0, 2\pi)^2,$$

with $b_1(x) = 1 + \frac{1}{2}\sin x_1 \cos x_2$, $b_2(x) = 1 + \frac{1}{2}\cos x_1 \sin x_2$. We choose F and V to be such that the exact solution is $U(x, t) = e^{-t}\sin(x_1 + t)\sin(x_2 + t)$, which gives

$$F(x, t) = U(x, t) + (1 - b_1(x))U_{x_1}(x, t) + (1 - b_2(x))U_{x_2}(x, t) \quad \text{and} \quad V(x) = U(x, 0) = \sin x_1 \sin x_2.$$

M	$\|u - U_h\|_h$	$\|\breve{u}^N - U_h\|_h$	$\|u^N - U_h\|_h$	$\|P^N v\|_h$	$\|H_1^N v\|_h$	$\|\widetilde{u}_1^N - U_h\|_h$
10	0.0311	0.1582	0.0761	0.0802	0.0494	0.1066
20	0.0076	0.0721	0.0305	0.0407	0.0225	0.0486
40	0.0019	0.0344	0.0134	0.0205	0.0108	0.0231
80	0.0005	0.0167	0.0062	0.0103	0.0053	0.0113

Table 3: Partial and complete errors for the second example ($d = 1$), with $m = 1$.

M	$\|u - U_h\|_h$	$\|\breve{u}^N - U_h\|_h$	$\|u^N - U_h\|_h$	$\|P^N\|_h$	$\|H_1^N\|_h$	$\|H_6^N\|_h$	$\|\widetilde{u}_6^N - U_h\|_h$
10	0.0870	0.1953	0.1304	0.0696	0.3852	0.0770	0.1004
20	0.0216	0.0804	0.0565	0.0358	0.2044	0.0376	0.0320
40	0.0054	0.0358	0.0281	0.0181	0.1062	0.0186	0.0117
80	0.0013	0.0168	0.0161	0.0091	0.0542	0.0093	0.0049

Table 4: Partial and complete errors for the third example ($d = 2$), with $m = 6$.

In this case, we obtain $\tilde{\beta} = \|b_1^2 + b_2^2\|_{\mathbb{C}} = 3.25.$ and $\gamma = 2\tilde{\beta} = 6.5$, thus

$$\rho_0 = \frac{1}{(4 \cdot 8 \cdot 3.25)^{\frac{1}{2}}} \approx 0.098 > \frac{k}{h} = \frac{1}{4\pi} \approx 0.080,$$

the condition needed for ρ_0 as required by Theorem 4.1. Table 4 contains the corresponding results for the same h, k, M and m as earlier, with $N = 2M$. The error behavior is similar to that in the one-dimensional cases given above. It was found that the time taken for BE is larger for $M = 80$ in this case. This could be due to our replacing an antisymmetric problem (1.9) with a symmetric version (1.11) thus making the linear algebra efficient. We conclude that for two-dimensional case with large number of mesh points, our splitting method is less time consuming and more accurate than the BE. The optimal choice for m for $M = 80$ is now found to be $m = 6$, and the estimate by a calculation for $M = 10$ is $m \approx 5.5$.

References

[1] S. Descombes, Convergence of a splitting method of high order for reaction-diffusion systems, *Math. Comp.* **70** (2001), no. 236, 1481–1501.

[2] E. Faou, A. Ostermann and K. Schratz, Analysis of exponential splitting methods for inhomogeneous parabolic equations, *IMA J. Numer. Anal.* **35** (2015), no. 1, 161–178.

[3] E. Hansen and A. Ostermann, Exponential splitting for unbounded operators, *Math. Comp.* **78** (2009), no. 267, 1485–1496.

[4] E. Hansen, A. Ostermann and K. Schratz, The error structure of the Douglas–Rachford splitting method for stiff linear problems, *J. Comput. Appl. Math.* **303** (2016), 140–145.

[5] W. Hundsdorfer and J. Verwer, *Numerical Solution of Time-dependent Advection-diffusion-reaction Equations*, Springer Ser. Comput. Math. 33, Springer, Berlin, 2003.

[6] T. Jahnke and C. Lubich, Error bounds for exponential operator splittings, *BIT* **40** (2000), no. 4, 735–744.

[7] S. MacNamara and G. Strang, Operator splitting, in: *Splitting Methods in Communication, Imaging, Science, and Engineering*, Sci. Comput., Springer, Cham (2016), 95–114.

[8] G. Strang, On the construction and comparison of difference schemes, *SIAM J. Numer. Anal.* **5** (1968), 506–517.

[9] V. Thomée and A. S. Vasudeva Murthy, An explicit-implicit splitting method for a convection-diffusion problem, *Comput. Methods Appl. Math.* **19** (2019), no. 2, 283–293.